

Semiconductor Manufacturing—Clean Rooms, Wafer Cleaning, and Gettering

4

4.1 Introduction

In describing the CMOS process in Chapter 2, the sentence “The wafers are first cleaned in a combination of chemical baths that remove any impurities from the surface” was used many times. As we have seen in previous chapters, semiconductor devices are fabricated by introducing dopants, often at concentrations of parts per million, and by depositing and patterning thin films on the wafer surface, often with a thickness control of a few nm. Such processes are manufacturable only if stray contaminants can be held to levels below those that affect device characteristics or chip yield. In practice, this means that unwanted impurities must be kept below the ppm or ppb range and stray particles must be essentially eliminated.

Modern IC factories employ a three-tiered approach to control unwanted impurities—clean rooms, wafer cleaning, and gettering. The first level is implemented by building the chips in a clean environment. The air is highly filtered. Machines are designed to minimize particle production. Ultra-pure chemicals and gases are used in wafer processing. The second line of defense is to chemically clean the wafers often and thoroughly. This removes particles and contaminant films from the wafer surface before they can get into the thin films on the wafer or into the silicon itself. The final line of defense is called gettering and is a method whereby those unwanted impurities that do get into the wafer are made to collect in noncritical parts of the wafer, typically the wafer backside or the wafer bulk, far away from the active devices on the top wafer surface.

We described in Chapter 1 the great strides that have been made in shrinking device geometries and in improving manufacturing so that larger chips can be economically built. This progress requires that defect control associated with the manufacturing process also dramatically improve. Consider the SIA roadmap data summarized in Table 4-1 [4.1]. The entries in bold are new entries from those we previously discussed in Chapter 1.

The critical defect size in the table is an indication of how large a random particle needs to be in order to create a malfunctioning chip. For example, if a particle is present on the wafer surface when photoresist is being exposed, then that particle will likely prevent exposure beneath it. If such a defect is a significant fraction of the feature size being printed and it is located in a “critical area” of the chip, it will likely lead to a malfunctioning device. What constitutes a “critical area” on the chip depends on the particular manufacturing step. For example, at the gate oxidation stage, a particle located in the gate oxide region of a MOSFET will likely cause a malfunctioning device (gate short). It has been estimated that about 75% of the yield loss in a modern silicon IC manufacturing plant is due directly to defects caused by particles on the wafer. Note that the projected critical defect size in 2012 is only $0.025\text{ }\mu\text{m}$ (25 nm) or about 100 atoms across.

The defect density numbers in Table 4-1 arise from the fact that economic manufacturing of complex chips requires that most of the ICs on a wafer work at the end of the process. The roadmap numbers are driven by the desire to achieve chip yields on the order of 90% at the end of the manufacturing process. This implies that the yield on each individual step must be far higher (>99%) since so many steps are performed

Table 4-1 Semiconductor industry projected progress in chip size and feature size and the implications of this progress for defect size, density and contamination levels [4.1]

Year of First DRAM Shipment	1997	1999	2003	2006	2009	2012
Minimum Feature Size	250 nm	180 nm	130 nm	100 nm	70 nm	50 nm
Wafer Diameter (mm)	200	300	300	300	450	450
DRAM Bits/Chip	256M	1G	4G	16G	64G	256G
DRAM Chip Size (mm ²)	280	400	560	790	1120	1580
Microprocessor Transistors/chip	11M	21M	76M	200M	520M	1.40B
Maximum Wiring Levels	6	6-7	7	7-8	8-9	9
Minimum Mask Count	22	22/24	24	24/26	26/28	28
Critical Defect Size	125 nm	90 nm	65 nm	50 nm	35 nm	25 nm
Starting Wafer Total LLS (cm ⁻²)	0.60	0.29	0.14	0.06	0.03	0.015
DRAM GOI Defect Density (cm ⁻²)	0.06	0.03	0.014	0.006	0.003	0.001
Logic GOI Defect Density (cm ⁻²)	0.15	0.15	0.08	0.05	0.04	0.03
Starting Wafer Total Bulk Fe (cm ⁻³)	3×10^{10}	1×10^{10}	Under 1×10^{10}	Under 1×10^{10}	Under 1×10^{10}	Under 1×10^{10}
Critical Metals on Wafer Surface After Cleaning (cm ⁻²)	5×10^9	4×10^9	2×10^9	1×10^9	$< 10^9$	$< 10^9$
Starting Material Recombination Lifetime (μsec)	≥ 300	≥ 325	≥ 325	≥ 325	≥ 450	≥ 450

in series in a manufacturing process. Fortunately, not all steps are critical and so optimizing the yield of chips requires identifying the critical steps, understanding the types of defects or contaminants that affect the yield in each step, and then setting upper bounds on those types of defects for each step. We will discuss yield issues more fully in Section 4.5.

The critical particle size is on the order of half of the minimum feature size. Particles larger than this have a high probability of causing a manufacturing defect. In starting wafers, the particle density is defined in terms of Localized Light Scatterers (LLS). Laser-based systems are often used to detect such particles and can produce automated counts (Section 4.4). At the gate insulator stage of the process, the issue is the electrical integrity of the gate dielectric and so the specification is in terms of Gate Oxide Integrity (GOI). This measurement is usually made by applying electric fields to test MOS structures and measuring the maximum field they can withstand before dielectric breakdown. The amount of charge trapping that occurs in the dielectric when current flows under high fields is also often measured as part of a GOI evaluation (Section 4.4). Note that the allowable particle density is much higher on starting wafers than it is at the critical gate insulator step. This is because the wafers are cleaned by the wafer manufacturer immediately before gate insulator processing. This step, when done properly, removes many particles as we will see later in this chapter. Notice also in Table 4-1 that the requirements are different for DRAM and logic chips, because of the greater gate insulator area on DRAM chips.

Table 4-1 also contains representative data from the SIA NTRS regarding contamination levels. The numbers for starting wafers are expressed in terms of volume concentrations in the wafer and are the contamination levels resulting from the crystal growth and wafer preparation processes. These numbers (on the order of $1 \times 10^{10} \text{ cm}^{-3}$) are well below 1 ppb. While the numbers are specifically given for Fe contamination, similar specifications apply to other metals and to elements like Na and Ca. We will see why these elements are important later in this chapter. Following a wafer cleaning step during chip manufacturing, the specification of metal contamination is expressed as a surface “dose” per cm^2 because the chemical cleaning procedures are surface cleaning steps. These specifications (metal contamination levels on the order of 10^9 cm^{-2}) represent only about 0.0001% of a monolayer. We will see later in this chapter how modern cleaning procedures can consistently achieve such remarkably clean surfaces.

It is obvious that great care must be taken in making sure that the factories in which chips are manufactured are as clean as possible. Even with such ultra-clean environments, and even with procedures which clean wafers thoroughly and often, it is not realistic to expect that all impurities can be kept out of silicon wafers. There is simply too much processing and handling of the wafers during IC fabrication. As a result, some means of removing these impurities from active device regions—gettering—is normally incorporated into the manufacturing process. We will discuss in this chapter how this three-tiered approach—clean factories, wafer cleaning, and gettering—combine to make possible the economical manufacturing of today’s integrated circuits.

4.2 Historical Development and Basic Concepts

If workers from one of today's modern IC manufacturing facilities were suddenly transported to a 1960s semiconductor plant, they would probably be amazed that chips could be successfully manufactured in such a place. Such factories were "dirty" by today's standards, wafer cleaning procedures were poorly understood, and gettering was an art rather than a science if it was used at all. Of course chips were manufacturable even in those days, but they were very small and contained very few components by today's standards. Since defects on a chip tend to reduce yields exponentially as chip size increases, small chips can be manufactured even in quite dirty environments. (We will discuss yield issues in more detail in Section 4.5.) However all of the progress that has been made in the past three decades in shrinking device sizes and designing very complex chips would have been for naught if similar advances had not been made in manufacturing capability, specifically in defect density.

Figure 4-1 illustrates the problem. Silicon wafers ready for cleaning may have a variety of surface films (SiO_2 , polysilicon, silicides, metals, etc.) and doped regions, which generally must not be significantly attacked by the cleaning chemicals. Photoresists are commonly on the surface since photolithography is such a frequent step in IC manufacturing. Particles may well be present from the room air or from process steps the wafers have been exposed to; these must be removed. Finally, small concentrations of undesirable elements such as metals (Fe, Au, Cu, etc.) and alkali ions (Na, K, etc.) may be present from machines, chemicals, or the people the wafers have been exposed to; these must also be removed. Several examples may help to illustrate both the importance of removing these particles and undesirable contaminants, and also the deleterious effects they can have on device performance.

In Section 1.4.2 in Chapter 1 we described the sensitivity of modern MOS devices to trace levels of alkali ion contamination in gate oxides (Na^+ , K^+ in particular). The threshold voltage of an MOS transistor (see Section 2.2.6 in Chapter 2) is to first order given by

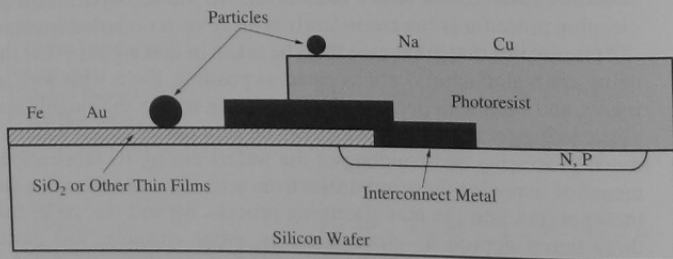


Figure 4-1 Example of a silicon wafer ready for cleaning. Contaminants include photoresist films, particles (usually $0.01 - 100 \mu\text{m}$ in size) and small concentrations of atoms or small clusters of metals, alkali ions, and other species.

$$V_{TH} = V_{FB} + 2\phi_f + \frac{\sqrt{2\epsilon_s q N_A (2\phi_f)}}{C_{OX}} - \frac{qQ_M}{C_{OX}} \quad (4.1)$$

where Q_M is the mobile charge density (number of charges per cm^2) associated with Na^+ or K^+ in the gate oxide. The other terms are defined in Chapters 1 and 2. If we consider a typical gate oxide thickness of 10 nm, then a Q_M value of 6.5×10^{11} ions cm^{-2} is sufficient to cause a 0.1-volt shift in V_{TH} . Such a shift is an upper bound on what is acceptable today. 6.5×10^{11} ions cm^{-2} corresponds to less than 0.1% of a monolayer of surface Na^+ or K^+ contamination. Equivalently, it corresponds to a volume concentration of $6.7 \times 10^{17} \text{ cm}^{-3}$ or ≈ 10 ppm contamination in the gate oxide. Identifying the sources of Na^+ or K^+ contamination and learning to control them dominated research on the MOS system in the 1960s and largely prevented MOS technologies from being important commercially until toward the end of that decade. Wafer cleaning procedures today effectively remove alkali ions as we will see, and clean facilities largely prevent their introduction into wafers.

As a second example, consider the Dynamic Random Access Memory or DRAM. DRAMs are the highest-volume product manufactured by the semiconductor industry. Chips containing up to 256 million memory bits are now commercially available (256-MBit DRAMs). The basic memory cell structure is shown in Figure 4-2. Charge stored on an MOS capacitor represents the “1” or “0” state. The MOS transistor is used to write information onto the capacitor and to access the stored information. Most of the time, information stored on any particular capacitor simply waits to be read. During this time, the stored charge will decay because of leakage currents that discharge the capacitor. (The drain-substrate junction of the MOS transistor is electrically in parallel with the capacitor and the junction leakage will discharge it.) Periodically, circuitry on the DRAM chip must refresh the stored charge so that the data are not permanently lost. This typically happens automatically every few msec (the “refresh” time in a DRAM).

In silicon devices, junction leakage currents are often dominated by SRH recombination as described in Chapter 1. To provide a refresh time of a few msec, the minority carrier lifetime in the silicon must typically be greater than approximately 25 μsec . In fact numbers much larger than this are specified in the SIA NTRS (Table 4-1). From Chapter 1,

$$\tau_R = \frac{1}{\sigma v_{th} N_t} \quad (4.2)$$

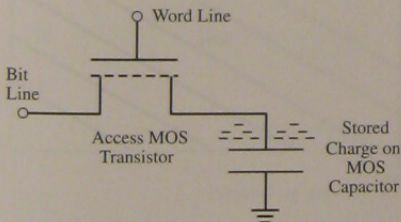


Figure 4-2 Equivalent circuit of a DRAM memory cell. Charge is stored on a MOS capacitor and the stored information is accessed through the MOS transistor switch.

where τ_R is the recombination lifetime, σ is the capture cross section of the trap (on the order of the atom cross section or about 10^{-15} cm^2), v_{th} is the minority carrier thermal velocity (about 10^7 cm sec^{-1}) and N_t is the density of traps per cm^3 . If $\tau_R = 100 \text{ } \mu\text{sec}$, then using typical values for σ and v_{th} , gives $N_t \approx 10^{12} \text{ cm}^{-3}$. We saw in Chapter 1 that in silicon, these traps are normally associated with deep-level impurities like Au, Cu, Fe, and other elements. This simple calculation therefore suggests that proper DRAM operation requires impurity levels of these elements in the parts per billion range or lower. In fact the requirements may be even more severe than this calculation would suggest because deep level traps like Au, Cu, and Fe preferentially tend to accumulate at wafer surfaces and in junction regions of devices. (These elements segregate to defects or imperfections or simply regions of the crystal that are different from the normal perfect periodic lattice.) It is really quite difficult to keep the levels of these impurities below the ppb range simply by purifying the starting wafers and by excluding metal contaminants from manufacturing environments. The cost of doing this is usually simply too high. Fortunately, however, an alternative exists. Gettering, a process we will discuss in this chapter, provides a means of removing these unwanted elements from the regions of wafers where active devices are located. Normal manufacturing practice then simply recognizes that small amounts of deep-level impurities will inevitably be present and includes a gettering operation as part of the IC fabrication process to eliminate them.

A final example will illustrate a very different effect of “contaminants” and the importance of cleaning procedures. In an interesting set of experiments in 1978, Schwettman and colleagues cleaned silicon wafers with several different chemical cleaning procedures and then oxidized the wafers together in an oxidation furnace [4.2]. They measured the resulting oxide growth kinetics, with the results shown in Figure 4-3. The cleaning procedures that were used were chosen because they are variations on the industry standard “RCA clean” that we will discuss in detail later in this chapter. It is clear from this data that the cleaning procedure has a significant effect on the oxide

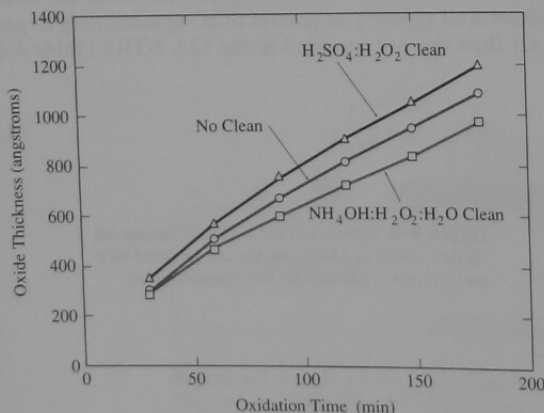


Figure 4-3 Effect of preoxidation cleaning procedure on oxide growth kinetics. (After [4.2]).

growth rate, presumably because of residual chemicals or surface oxides left on the wafers following cleaning. At the time these experiments were done, there was no understanding of the basic mechanism responsible for the results. Only in recent years with careful studies of surface chemistry have the reasons become clear.

We now turn to a more specific discussion of the three levels of contamination control used in modern IC manufacturing—clean factories, wafer cleaning, and gettering.

4.2.1 Level 1 Contamination Reduction: Clean Factories

The laboratories in which ICs are manufactured today clearly must be clean facilities. Particles that might deposit on a silicon wafer and cause a defect may originate from many sources including people, machines, chemicals, and process gases. Such particles may be airborne or may be suspended in liquids or gases. It is common to characterize the cleanliness of the air in IC facilities by the designation “class 10” or “class 100.” Figure 4-4 illustrates the meaning of these terms. Class X simply means that in each cubic foot of air in the factory, there are less than X total particles greater than $0.5\text{ }\mu\text{m}$ in size. A typical office building is about class 100,000, while room air in state-of-the-art manufacturing facilities today is typically class 1 in critical areas. This level of cleanliness is obtained through a combination of air filtration and circulation, clean room design, and through careful elimination of particulate sources.

Particles are always present in a distribution of sizes and shapes. Those that are of most concern in semiconductor plants are between about 10 nm and $10\text{ }\mu\text{m}$ in size. In room air, particles smaller than 10 nm tend to coagulate into larger sizes; those larger than $10\text{ }\mu\text{m}$ tend to be heavy enough to precipitate fairly quickly. Particles between 10 nm and $10\text{ }\mu\text{m}$ can remain suspended in the air for very long periods of time. Such particles deposit on surfaces primarily through two mechanisms: Brownian motion and

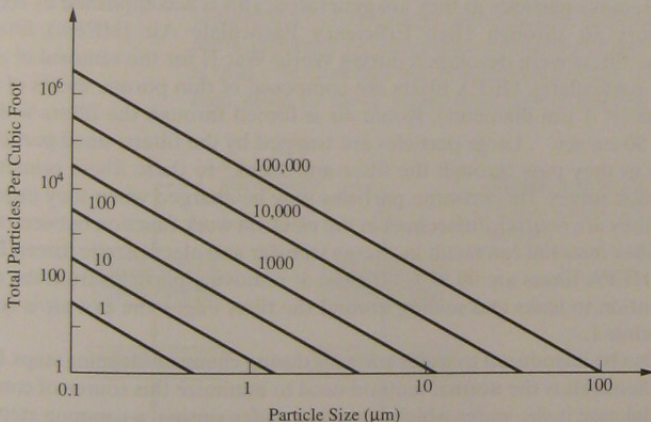


Figure 4-4 Particle size distribution curves for various classes of clean rooms (Federal Standard 209D, U.S. GSA 1989). The vertical axis is the total number of particles larger than a given particle size.

gravitational sedimentation. The first is the random motion of the particles in the air that occasionally brings them into contact with surfaces; the second is the gravitational force acting on the particles. Assuming a clean room environment in which the airflow is laminar at 50 cm sec^{-1} , Hu has estimated the rate of particle deposition on horizontal surfaces [4.3]. He finds that Brownian motion dominates for particles smaller than about $0.5 \text{ }\mu\text{m}$ and gravitational effects dominate for larger particles. The net result is that in a class 100 clean room, about 5 particles larger than $0.1 \text{ }\mu\text{m}$ will deposit on each square cm of surface area per hour. Comparing this estimate with the requirements in Table 4-1 illustrates why class 10 or class 1 clean rooms are now required.

Particles in the air in a manufacturing plant generally come from several main sources. These include the people who work in the plant, machines that operate in the plant, and supplies that are brought into the plant. Many studies have been done to identify particle sources and the relative importance of various sources. For example, people typically emit several hundred particles per minute from each cm^2 of surface area. The actual rate is different for clothing versus skin versus hair, but the net result is that a typical person emits 5 – 10 million particles per minute. This rate also varies with activity level, ranging from less than 10^6 to more than 10^7 particles min^{-1} for sitting versus running. Most modern IC manufacturing plants make extensive use of robots for wafer handling in an effort to minimize human handling and therefore particle contamination.

The first step in reducing particles is to minimize these sources. People in the plant wear “bunny suits” which cover their bodies and clothing and which block particle emission from these sources. Often face masks and individual air filters are worn to prevent exhaling particles into the room air. Air showers at the entrance to the clean room blow loose particles off people before they enter and clean room protocols are enforced to minimize particle generation. Machines that handle the wafers in the plant are specifically designed to minimize particle generation and materials are chosen for use inside the plant which minimize particle emission.

Since the sources of particles can never be completely eliminated, constant air filtration is used to remove particles as they are generated. This is accomplished by recirculating the factory air through High Efficiency Particulate Air (HEPA) filters. Interestingly, these filters were developed during World War II for the removal of airborne fissionable particulates. HEPA filters are composed of thin porous sheets of ultrafine glass fibers ($< 1 \text{ }\mu\text{m}$ diameter). Room air is forced through the filters with a velocity of about 50 cm sec^{-1} . Large particles are trapped by the filters; small particles impact the fibers as they pass through the filter and “stick” to these fibers primarily through electrostatic forces. The airborne particles may be charged when they impact the filter. Even if they are neutral, differences in the electron work function between the particle and the filter material can result in charge transfer and electrostatic forces. The net result is that HEPA filters are 99.97% efficient at removing particles from the air. With careful attention to leaks and sealing around the filter edges, the exit air is typically better than class 1.

Particles can also be introduced to wafer surfaces during chemical cleaning steps. Filtration of these chemicals is the normal method used to minimize this source of contamination. A special case is the water which is used in wafer rinsing, a common step in wafer cleaning procedures. Most IC manufacturing facilities produce their own clean

water on site, starting with water from the local water supply. This water is filtered to remove dissolved particles and organics. Dissolved ionic species are removed by ion exchange or reverse osmosis. The result is high-purity water that is used in large quantities in the plant.

4.2.2 Level 2 Contamination Reduction: Wafer Cleaning

Contamination on wafers consists of particles, which we discussed in the last section, films such as photoresist which must be removed after they serve their purpose in lithography, and trace levels of any element which has not been purposely introduced. The latter category is perhaps the most difficult to measure and to control. The second level of contamination reduction is wafer cleaning. There is probably no step that is used more often in IC manufacturing. Cleaning must remove particles, films such as photoresist, and it must remove trace concentrations of any other elements present on the wafer surface.

Figure 4-5 illustrates the general strategy used in cleaning wafers. Most cleaning procedures begin with photoresist removal since photolithography normally precedes each processing step. Photoresists are organic compounds and can be stripped in a variety of chemicals. For front-end processing (no metals on the wafers), two methods are common. The first uses an acid (usually H_2SO_4) and a strong oxidant (usually H_2O_2) to decompose the resist into CO_2 and H_2O . The second method uses an oxygen plasma to convert the resist to gaseous byproducts (again CO_2 and H_2O). The oxygen plasma method offers the advantages of reduced pollution problems and very good selectivity to almost all underlying materials. For removal of photoresists in back-end processes the oxygen plasma approach can be used as can a variety of phenol-based organic strippers that do not significantly attack metals.

The most important aspects of wafer cleaning are the steps that are used following photoresist removal for front-end processes. Front-end processes often involve high temperatures (oxidations, anneals, film deposition etc.) and it is these steps which can allow diffusion of contaminants into thin films on the silicon or into the silicon itself. It is thus essential that the cleaning procedure remove contaminants from the wafer surface prior to these steps. Back-end processes are normally low temperature, so the

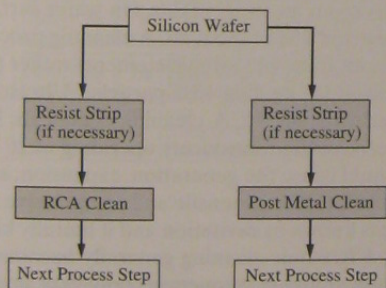


Figure 4-5 Typical cleaning strategies in a modern IC fabrication process. Left, front-end cleaning; right, back-end cleaning.

wafer cleaning, while important, is not as critical. (Both front-end and back-end cleaning procedures must remove particles from the wafers however since these translate directly into defects.)

Prior to about 1970, a variety of front-end chemical cleaning procedures were used in the semiconductor industry. In that year, however, a classic paper was published [4.4] which provided both a “standard” cleaning procedure and the beginnings of a scientific basis for wafer cleaning. This paper by Kern and Puotinen described work done at RCA over the previous several years in which detailed measurements were made of the effectiveness of various cleaning procedures in removing contaminants from silicon wafers. The resulting cleaning procedure became known as the “RCA clean” and has been the mainstay of the industry since that time. An excellent general reference on wafer cleaning issues is [4.5].

The original RCA clean is based on a two-step oxidizing and complexing treatment using hydrogen peroxide solutions. The first solution (called SC-1 in the literature) is a high-pH solution consisting of 5 H_2O :1 H_2O_2 :1 NH_4OH in which the wafers are placed at 70–80°C for 10 minutes. This solution oxidizes organic films and complexes Group IB and IIB metals as well as other metals like Au, Ag, Cu, Ni, Zn, Cd, Co, and Cr. The SC-1 solution slowly dissolves the thin native oxide layer on silicon and continuously grows a new oxide layer by oxidation. This combination of etching and reoxidation helps to dislodge particles from the wafer surface. NH_4OH etches silicon and this solution can produce microroughening of the silicon. In recent years, the concentration of NH_4OH in the SC-1 solution has generally been reduced to minimize these effects. This improves the quality of very thin gate oxides grown in modern devices on these cleaned surfaces.

The second solution (called SC-2 in the literature) is a low-pH solution consisting of 6 H_2O :1 H_2O_2 :1 HCl in which the wafers are placed at 70–80°C for 10 minutes. This solution removes alkali ions and cations like Al^{+3} , Fe^{+3} , and Mg^{+2} that form NH_4OH insoluble hydroxides in basic solutions like SC-1. These metals precipitate onto the wafer surface in the SC-1 solution. In the SC-2 solution, they form soluble complexes. The SC-2 solution also completes the removal of metallic contaminants such as Au that may not have been completely removed by the SC-1 step. The scientific basis for the SC-1 and SC-2 solutions comes from reaction chemistry and oxidation potentials; we will discuss these issues more carefully in Section 4.5.

Particles may sometimes adhere tenaciously to surfaces primarily due to electrostatic forces, so additional steps may be necessary to remove them. We depend on these forces in HEPA filters which clean the room air in IC plants. On wafer surfaces, however, we need to fully remove such particles. The original RCA cleaning procedure was really designed for removing contaminant films and atomic elements rather than particles, so additional strategies have evolved for dealing with particles. Ultrasonic scrubbing works well and can be combined with the RCA cleaning chemicals. Ultrasonic agitation is produced usually with piezoelectric transducers operating at 20–50 KHz. The sound waves in an appropriate liquid cause the generation, expansion, and violent collapse of tiny vapor bubbles under the alternating tensile and compressive stresses of the ultrasound. The bubble collapsing is known as cavitation and it literally knocks particles off the wafers or loosens them. Ultrasonic cleaning generally becomes less efficient as the frequency increases because at higher frequencies only very small bubbles can collapse in the compressive portion of the sound cycle. However, it has been found

that very high frequencies (≈ 1 MHz) are also quite effective for wafer cleaning and some commercial systems use these frequencies (megasonic cleaning).

The second approach to particle removal literally involves scrubbing the wafers with a brush in a liquid. The liquid allows the brush to hydroplane over the wafer surface, minimizing scratching while the rapidly rotating brush loosens particles from the surface.

4.2.3 Level 3 Contamination Reduction: Gettering

The term gettering has its origins in the vacuum tube industry where it referred to the use of materials like cesium which were used to remove trace gases from vacuum tubes. The basic idea of gettering in silicon manufacturing is simple. It is very difficult to reduce the concentration of undesirable elements such as metals (Fe, Au, Cu, etc.) and alkali ions (Na^+ , K^+ , etc.) in silicon materials to levels where they would have no effect on device performance. Clean factories and wafer cleaning are crucial as we have seen, but they are generally not sufficient. Gettering is the third tier defense and is a means of collecting these unwanted elements in regions of the chip where they do minimal harm. This is not as difficult as it might seem because the active devices in chips occupy a very small fraction of the wafer volume. In addition, the undesirable elements tend to have very high diffusivities (which is part of the reason why they are so “bad”) and they tend to be easily captured either in regions with mechanical defects or in regions which chemically trap them. The elements we are most concerned about are highlighted on the periodic table in Figure 4-6.

Period																	Noble Gases	
1	1	2															3	4
	¹ H	² He															4.003	20.18
2	3	4															5	6
	³ Li	⁴ Be															12.01	16.00
3	11	12															13	14
	¹¹ Na	¹² Mg															26.98	28.09
4	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
	¹⁹ K	²⁰ Ca	²¹ Sc	²² Ti	²³ V	²⁴ Cr	²⁵ Mn	²⁶ Fe	²⁷ Co	²⁸ Ni	²⁹ Cu	³⁰ Zn	³¹ Ga	³² Ge	³³ As	³⁴ Se	³⁵ Br	³⁶ Kr
5	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
	³⁷ Rb	³⁸ Sr	³⁹ Y	⁴⁰ Zr	⁴¹ Nb	⁴² Mo	⁴³ Tc	⁴⁴ Ru	⁴⁵ Rh	⁴⁶ Pd	⁴⁷ Ag	⁴⁸ Cd	⁴⁹ In	⁵⁰ Sn	⁵¹ Sb	⁵² Te	⁵³ I	⁵⁴ Xe
6	55	56	57	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86
	⁵⁵ Cs	⁵⁶ Ba	⁵⁷ La	⁷² Hf	⁷³ Ta	⁷⁴ W	⁷⁵ Re	⁷⁶ Os	⁷⁷ Ir	⁷⁸ Pt	⁷⁹ Au	⁸⁰ Hg	⁸¹ Tl	⁸² Pb	⁸³ Bi	⁸⁴ Po	⁸⁵ At	⁸⁶ Rn
7	87	88	89	104	105	106	107											
	⁸⁷ Fr	⁸⁸ Ra	⁸⁹ Ac	¹⁰⁴ Unq	¹⁰⁵ Unp	¹⁰⁶ Unh	¹⁰⁷ Uns											
	223	226	227.0	261	262	263	262											

Alkali Ions: ¹H, ³Li, ¹¹Na, ¹⁹K, ³⁷Rb, ⁵⁵Cs, ⁸⁷Fr

Deep Level Impurities in Silicon: ²³V, ²⁴Cr, ²⁵Mn, ²⁶Fe, ²⁷Co, ²⁸Ni, ²⁹Cu, ³⁰Zn, ³¹Ga, ³²Ge, ³³As, ³⁴Se, ³⁵Br, ³⁶Kr, ⁴¹Nb, ⁴²Mo, ⁴³Tc, ⁴⁴Ru, ⁴⁵Rh, ⁴⁶Pd, ⁴⁷Ag, ⁴⁸Cd, ⁴⁹In, ⁵⁰Sn, ⁵¹Sb, ⁵²Te, ⁵³I, ⁵⁴Xe

Shallow Acceptors: ¹³Al, ¹⁴Si, ¹⁵P, ¹⁶S, ¹⁷Cl, ¹⁸Ar, ³¹Ga, ³²Ge, ³³As, ³⁴Se, ³⁵Br, ³⁶Kr, ⁴⁹In, ⁵⁰Sn, ⁵¹Sb, ⁵²Te, ⁵³I, ⁵⁴Xe

Elemental Semiconductors: ¹⁴Si, ¹⁵P, ¹⁶S, ¹⁷Cl, ¹⁸Ar, ³²Ge, ³³As, ³⁴Se, ³⁵Br, ³⁶Kr, ⁵⁰Sn, ⁵¹Sb, ⁵²Te, ⁵³I, ⁵⁴Xe

Shallow Donors: ¹⁵P, ¹⁶S, ¹⁷Cl, ¹⁸Ar, ³³As, ³⁴Se, ³⁵Br, ³⁶Kr, ⁵¹Sb, ⁵²Te, ⁵³I, ⁵⁴Xe

Figure 4-6 Periodic table indicating the elements that are of most concern in gettering. The alkali ions are in column 1 of the table; the metal ions that form deep level SRH centers are largely in the central portion of the table.

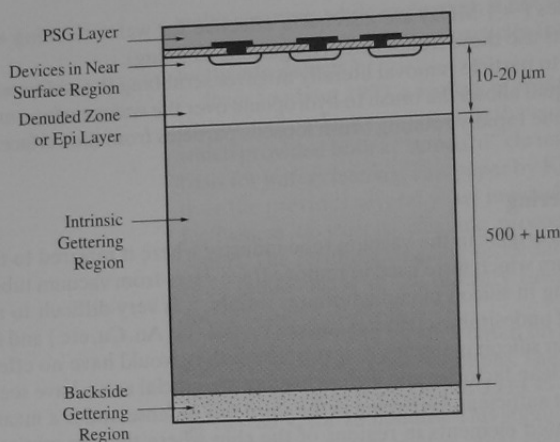


Figure 4-7 Wafer gettering strategies showing a surface denuded zone, intrinsic or bulk gettering regions and backside gettering.

Gettering therefore consists of three steps. First, the elements to be gettered must be “freed” from any trapping sites they may currently occupy and made mobile. Second, they must diffuse to the gettering site. Finally, they must be trapped. Figure 4-7 illustrates some of the key points.

Gettering strategies differ for the alkali ions (Na^+ , K^+ , etc.) and metals. The alkali ions are of most concern in the thin dielectrics on the top surface of ICs as the example earlier on MOS V_{TH} stability showed. Several approaches to gettering such elements have been explored, beginning in the 1960s [4.6–4.9]. The PSG (phosphosilicate glass) layer on the top of the wafer in Figure 4-7 is a common example. PSG is a $\text{P}_2\text{O}_5/\text{SiO}_2$ glass that is normally deposited by CVD or LPCVD. Usually it is about 5% by weight phosphorus. This glass has been found to be a very efficient trap for alkali ions, forming a stable complex that binds Na^+ or K^+ ions. Depositing such a glass as one of the upper thin film layers on a chip can thus effectively keep such ions from penetrating down to sensitive gate oxides or field oxides which are adjacent to the silicon surface. PSG layers are also effective getters for such ions. Na^+ and K^+ are very mobile at temperatures above room temperature and they will easily diffuse to and be trapped by a PSG layer. Effective gettering of these ions from dielectric layers below the PSG can thus take place. We used such a layer in the CMOS process in Chapter 2 (see Figure 2-38).

There are some drawbacks to PSG layers however. These layers contain charge dipoles that can affect surface electric fields [4.9], they are susceptible to absorbing water vapor and they can cause Al corrosion problems. These effects can be minimized by keeping the P concentration to a few percent. An alternative is to use a deposited Si_3N_4 layer as the top dielectric layer on a chip. These layers are impermeable to alkali ions and can form effective barriers to indiffusion. Using a barrier layer like this requires that earlier processing steps be free from significant alkali ion contamination since if Na^+ or K^+ ions were introduced earlier in the process, they would simply be locked into

the IC structure by the surface nitride layer and could therefore cause instability problems. This strategy—clean processing plus a nitride cap layer—has been found to be workable by many VLSI manufacturers.

Trace metal atoms in the silicon bulk are gettered in a different fashion. Two basic methods are widely used. Extrinsic gettering makes use of a layer, usually on the wafer backside, which is created explicitly to trap such atoms. Intrinsic gettering makes use of a property of CZ silicon, namely that it contains oxygen. The oxygen can be made to precipitate and create traps for metal atoms within the wafer bulk.

Two basic properties of metal atoms, which act as deep level traps, have been well established for many years. The first is that these atoms have very high diffusion coefficients in silicon. Figure 4-8 shows some representative data [4.10, 4.11]. Note that the metals have diffusivities many orders of magnitude larger than the common dopants like P, B and As in silicon. Most of these metals diffuse as interstitials in silicon whereas dopants are substitutional and diffuse by interacting with point defects (see Chapter 7). Because of this, in the same time that standard dopants are diffusing to produce a PN junction $1\text{ }\mu\text{m}$ below the silicon surface, metals such as Au can easily diffuse completely through a silicon wafer. From a positive perspective, this makes it easy to getter these elements to regions away from active devices. From a negative perspective, metal contamination anywhere on a wafer can redistribute throughout an entire wafer.

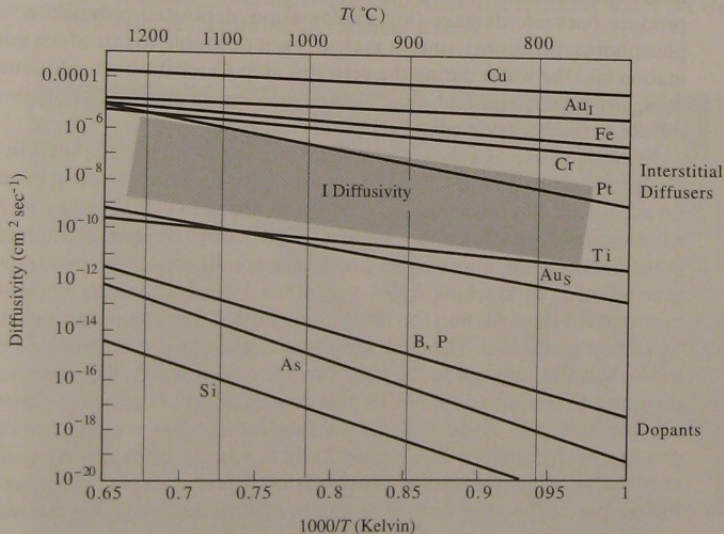


Figure 4-8 Diffusivities of various species in silicon. Au_s refers to gold in substitutional form (on a lattice site); Au_i to gold in an interstitial site. The silicon interstitial (I) diffusivity is also shown and will be discussed later. The gray area representing the I diffusivity indicates the uncertainty in this parameter. (After [4.10, 4.11].)

The second important property of these metals is the fact that they preferentially reside at sites in the silicon lattice where “imperfections” exist. Simply stated, these elements do not “fit” in the silicon lattice easily because of their very different atomic size. Thus if there are regions where faults or disorder exist, it is easier for these elements to find a suitable home. If we can create such regions in noncritical areas of the wafer, then we can getter these elements to those regions. Figure 3-4 illustrated some of the simple types of defects that can exist in a silicon crystal. Dislocations and stacking faults are well known sites for metals like Au, Cu, and Fe to “decorate.” If such regions exist near or in device structures, metals can decorate them and this will result in significant degradation in device performance. The trick is to form such defects in the silicon, but away from the devices.

Extrinsic gettering sites have been formed in silicon by a wide variety of methods. Normally they are created on the wafer backside, which is an accessible surface far away from the devices on the top surface. Methods that have been used range from grinding and sandpaper abrasion, to ion implantation, laser melting, depositing amorphous or polycrystalline films (usually polysilicon), or high-concentration backside diffusions (usually phosphorous). Whatever the method, the objective is to create extended defects—dislocation loops, grain boundaries, precipitates—or other traps that are stable through subsequent high-temperature processing and hence can trap and hold metal atoms. Modern processes tend to use the “cleaner” methods from this list to produce backside damage (ion implantation, deposited polysilicon films, or backside phosphorus diffusions), since it is clearly not desirable to introduce additional contamination into the wafer during the gettering step. Once the backside damage sites are created, any subsequent high-temperature processing step will allow the metal atoms to diffuse to the backside where they can be trapped.

Intrinsic gettering relies on the ideas we described in Chapter 3 in connection with Figure 3-25. Oxygen in CZ silicon is normally supersaturated at processing temperatures because it is incorporated during growth at 1417°C. As we saw in Chapter 3, CZ silicon normally contains 10–20 ppm ($\approx 10^{18} \text{ cm}^{-3}$) oxygen. SiO_2 precipitates can form in the silicon under appropriate processing conditions. These precipitates are normally accompanied by stacking faults and other extended defects because of the volume mismatch of the SiO_2 and the host silicon lattice. Such sites are ideal trapping sites for unwanted metal ions. The trick here is to cause this precipitation to occur in the bulk of the Si wafer, but not in the near surface region where the devices are located. This turns out to be quite possible. In many cases, an epitaxial layer is grown on the silicon substrate as part of normal device fabrication (this was an option in the CMOS process we considered in Chapter 2). If this is the case, epitaxial silicon is normally very low in oxygen and so no precipitation would occur in the epilayer. Alternatively, the oxygen in the near surface region can be outdiffused from the wafer as one of the first steps in wafer processing, creating the denuded zone shown in Figure 4-7. In either case, SiO_2 precipitation can then take place only in the wafer bulk, away from the active devices.

4.3 Manufacturing Methods and Equipment

In this section we will discuss how the three levels of contamination control are actually implemented in modern IC manufacturing.

4.3.1 Level 1 Contamination Reduction: Clean Factories

Modern IC manufacturing plants are designed to continuously recirculate the room air through HEPA filters to maintain a class 10 or class 1 environment. A typical clean room is shown in Figure 4-9. All of the mechanical support equipment (pumps, etc.) is located beneath the clean room to minimize contamination from those sources. The HEPA filters are located in the ceiling of the clean room and the fans that recirculate the air are normally above these filters. Within the clean room, fingerwalls or chases provide a path for air return as well as a means to bring in electrical power, DI water, and process gases. Notice in the photo in Figure 4-9 the clean room clothing being worn. These “bunny suits” are designed to minimize particle emission from the people in the clean room.

The chemicals, gases, and DI water used in the plant are also potential sources of contamination. Normally electronic grade chemicals and gases are supplied by vendors to the plant. These may be further filtered on-site. DI water is usually generated at the plant site using filtering and deionization. The resulting water is extremely pure and is used in large quantities for cleaning wafers.



Figure 4-9 Configuration of typical modern cleanroom for IC fabrication. Photo courtesy of Stanford Nanofabrication Facility.

4.3.2 Level 2 Contamination Reduction: Wafer Cleaning

Wafer cleaning is usually accomplished today either using immersion of cassettes of wafers into cleaning baths or through chemical sprays. A typical wet bench for immersion cleaning is shown in Figure 4-10. Such baths may use recirculating filter systems and ultrasonic cleaning is shown in Figure 4-10. Note the robot arm that is used to transfer the wafer cassettes from bath to bath. In spray systems, the wafers are generally mounted on a spinning drum and rotated past nozzles that spray the cleaning chemicals.

The standard implementation of the RCA cleaning procedure is shown in Figure 4-11. Options include the addition of ultrasonic agitation to the SC-1 or SC-2 solutions, using heated DI water for the rinses, the addition of a 50:1 H_2O/HF step between the SC-1 and SC-2 solutions (with DI H_2O rinses before and after the HF step), or the addition of a 50:1 H_2O/HF step at the end of the cleaning procedure (followed by another DI H_2O rinse). The ultrasonic agitation is designed to help particle removal as we have previously discussed. The additional 50:1 H_2O/HF steps are designed to remove the chemical oxide layers that are grown during the RCA cleaning steps. Removing these oxides remains a controversial option. The advantage is that any impurities contained in the oxide are stripped by the 50:1 H_2O/HF step. However this step leaves the silicon surface hydrogen terminated which is not particularly stable, which makes it more susceptible to absorbing additional contaminants during subsequent processing.

Other cleaning solutions are used during back-end processing as described earlier in this chapter. Generally the equipment used is similar to that described here for front-end cleaning. The cleaning solutions themselves are usually purchased and used as is from suppliers.



Figure 4-10 A typical cleaning station in a modern IC manufacturing facility. Wet chemical cleaning baths are in the sinks. The robot is loading a cassette of wafers into a spin dryer. Photo courtesy of Ruth Carranza.

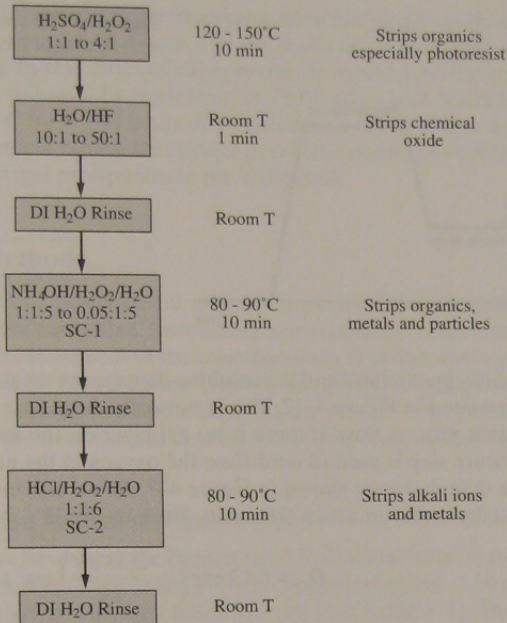


Figure 4-11 Standard implementation of the RCA cleaning procedure.

4.3.3 Level 3 Contamination Reduction: Gettering

Both extrinsic and intrinsic gettering processes are common in the semiconductor industry and sometimes both are used for added insurance. Often an extrinsic gettering process is used in which backside damage or a backside polysilicon layer is created at some convenient time during the normal process flow and then simply left in place during subsequent high-temperature steps. Gettering then takes place automatically without any special high-temperature step. This approach requires that the backside damage or the polysilicon layer be thermally stable during all the remaining high-temperature steps. If the damage anneals out, or the polysilicon recrystallizes, the gettered metal ions can be released and recontaminate the device regions on the wafer topside.

Intrinsic gettering is more popular because it tends to be a better controlled process. The damaged region is much larger than simply the wafer backside, is closer to the devices than the backside, and is thermally stable once it is created. As described conceptually earlier, intrinsic gettering consists of creating SiO₂ precipitates in the wafer bulk by appropriate thermal cycling of the wafer. Because of the volume mismatch between these precipitates and the silicon lattice, defects such as dislocations are generated as a result of the precipitation. These defects provide trapping sites for the gettered species, well away from the device region on the wafer surface.

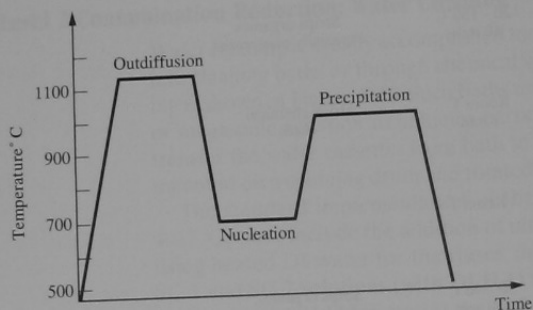


Figure 4-12 Processes and time/temperature cycle for a typical intrinsic gettering process.

The SiO_2 precipitates and the resulting damage are created using a thermal cycle like that illustrated in Figure 4-12. This is normally done near the beginning of the wafer fabrication process flow. If there is no epi layer on the wafer surface, an initial high-temperature step is used to outdiffuse the oxygen in the near surface region. This creates the denuded zone shown in Figure 4-7. As we saw in Chapter 3, oxygen diffuses interstitially in silicon with a diffusivity given by [4.12]

$$D_o = 0.13 \exp\left(-\frac{2.53 \text{ eV}}{kT}\right) \text{ cm}^2 \text{ sec}^{-1} \quad (4.3)$$

The time required to create an “oxygen free” surface region can thus be easily calculated. The surface region does not actually have to be oxygen free. What is required is that the oxygen concentration is reduced below the level at which precipitation will occur in the next part of the cycle. Normal oxygen concentrations in CZ silicon are 15–20 ppm ($\approx 10^{18} \text{ cm}^{-3}$). If the concentration is much higher than 20 ppm, too much precipitation will take place and there may be problems with wafer warpage or excessive defect formation. Precipitation does not readily occur below concentrations of about 10 ppm. A temperature cycle of several hours at 1100–1200°C is sufficient to create a denuded zone several tens of μm deep.

The next step in the process is precipitation of the oxygen to form small SiO_2 precipitates or embryos. We described this process in some detail in Chapter 3 (Section 3.5.5). The optimum temperature for precipitation is around 700°C [4.13]. Precipitation usually takes place over several hours. The embryos must be grown to a minimum critical size during this step so that during the subsequent higher temperature growth step they will grow and not shrink. This critical size depends on process conditions (especially temperature), but is on the order of 1–3 nm. The density of embryos created also depends strongly on process conditions, but is on the order of 10^{11} cm^{-3} .

The final step in creating the intrinsic gettering region is growth of the embryo precipitates. This is accomplished by raising the temperature to approximately 1000°C for several hours, which allows them to grow to 50–100 nm in size. It may be necessary to carefully control the temperature ramp rate at the start of the growth cycle, because the critical embryo size increases with temperature. If the temperature is ramped up too quickly, the embryos may be below the critical size at the growth temperature and

therefore shrink rather than grow. Slow ramping gives them a chance to grow during the ramp up in temperature. The net result of these process steps is shown in Figure 4–13 [4.14]. In this series of experiments, an initial outdiffusion for 10 hours at 1100°C was used, followed by nucleation at 750°C for 6 – 24 hours and growth (precipitation) at 1000°C for 8 – 48 hours. A denuded zone about 25 μm deep is visible at the surface of the wafers. Notice that longer precipitation and growth times result in a higher density and larger precipitates in the wafer bulk.

4.4 Measurement Methods

Accurate measurement methods are critical if contamination sources are to be identified and controlled. Instruments are needed for particle detection and for identification and measurement of the concentrations of contaminating species. As IC manufacturing processes have become cleaner, sensitivity requirements on measurement methods have become correspondingly higher. We will describe in this section some of the common methods used to detect and measure such contaminants.

4.4.1 Level 1 Contamination Reduction: Clean Factories

As we have seen, the biggest issue in clean factories is particle control. We need methods to accurately detect particle concentrations of < 10 particles per wafer, with a critical particle size well below 0.1 μm (see Table 4–1). The simplest approach is optical microscopy, which can routinely detect particles down to about 1 μm in size. Manual inspection can be used to search for particles but this is a very time consuming and inefficient process.

Particle detection is simplest on blank wafers since there is no pattern present to scatter the light. One common method of particle detection involves running a blank wafer through a process and then counting the particles to assess the cleanliness of that process. Automated systems generally rely on lasers and the scattering of laser light that takes place when the light strikes a particle on the wafer. The laser beam is scanned over the surface of the wafer. The light is reflected at a predictable angle unless the light strikes a particle. Light that is scattered off particles is collected by an optical system, amplified by a photodetector, and used to produce a wafer map of the particles present. This process can reliably detect particles down to $\approx 0.2 \mu\text{m}$ in size and can produce a wafer map in a few seconds. Shorter wavelength lasers can reduce the detected particle size. The process is simple and reliable, but it is restricted to unpatterned wafers. Several commercial systems are available to make this measurement.

While the data from blank wafers are useful, the real interest of course is in defect densities on product wafers. Since these wafers are necessarily patterned, simple laser scanning systems are not a solution. Two approaches to this problem are used. The first uses an optical system and comparison of the image with a “known good reference.” The reference can be either an adjacent die or a database from the chip design of what the pattern is supposed to look like. When an adjacent die is used as the reference, any differences in the images from the two die are “interesting events” and are usually shown to a technician for evaluation. In more advanced systems, image processing and

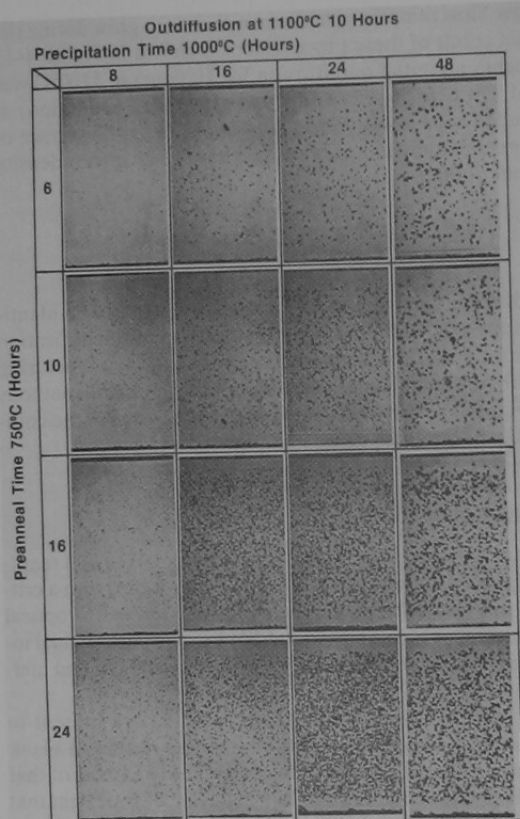


Figure 4-13 Cross-sectional SEM images of a wafer utilizing intrinsic gettering under various process conditions. All samples were initially outdiffused at 1100°C for 10 hrs. The SiO_2 precipitates are clearly visible in the wafer bulk, as is the denuded zone at the surface. Reprinted with permission of Solid State Technology [4.14].

pattern recognition software can be used to help identify defects. This approach is much more time consuming than the simple laser scanning systems (5 – 10 minutes per wafer), but it does provide information on real product wafers. The particle sizes that can be observed in this way are limited by the optical system used, but the principle of image comparison can be extended to very small particle sizes through the use of Scanning Electron Beam (SEM) systems. There are, however, throughput issues associated with the SEM systems in a manufacturing environment.

The second approach to measuring defect densities on product wafers involves test structures specifically designed to detect defects. These test structures may occupy a few test sites on a product wafer or they may occupy a full wafer processed alongside the product wafer. A simple example is shown in Figure 4-14 in which a mask level for an interconnect layer is shown (metal or polysilicon layer for example). Simple electrical

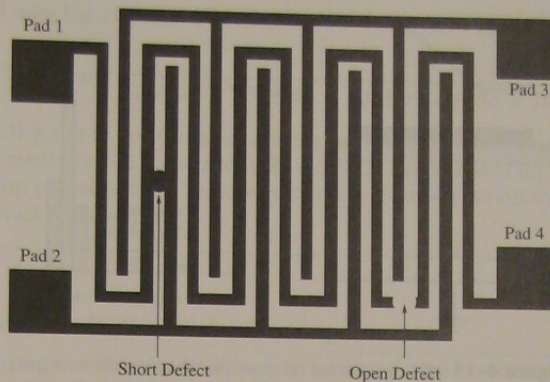


Figure 4-14 Layout (top view) of a typical electrical test structure designed to detect defects in an interconnect layer. Typical open and short defects are illustrated.

measurements of connectivity at the end of the process can give information on defect densities. For example, an open between pads 1 and 4 would indicate a break in what should be a continuous line, probably caused by a particle defect during the deposition or etching of that line. A short between pads 1 and 2 would indicate bridging between two metal lines, again, probably caused by a particle during deposition or etching of the metal lines. A significant amount of research in recent years has investigated the design of such defect monitors [4.15 – 4.17]. They are widely used, especially during the development of a new process or technology. Obviously, the lower the defect density, the larger such structures must be to reliably detect the defects. Once a product is in high-volume manufacturing, it is generally undesirable to “waste” wafer area on such monitors since a chip that is used for monitoring could also be a product chip.

It is also common to use electrical methods to detect defects in insulating layers, particularly gate insulators. Here the simplest test structure is a MOS capacitor, illustrated in Figure 4-15. In the simplest case a thin insulator is grown on a silicon wafer and then metal gates are patterned. A voltage is then applied to each capacitor structure and ramped up until dielectric breakdown occurs. The field at which this occurs is recorded and a histogram plot generated as shown. Ideally, all the capacitors would break down at the same field, corresponding to the maximum field the insulator can sustain. In practice, some capacitors break down prematurely, because of defects in the insulator, often caused by particles present on the surface when the dielectric was grown.

More sophisticated measurements are also possible with this simple MOS structure. For example, a high field (below breakdown) can be applied to the capacitor for a fixed time period. Some current will flow through the insulator due to tunneling or other conduction mechanisms. Traps or defects in the insulator can capture some of the electrons or holes while this current is flowing, resulting in charge accumulation in the insulator. This charge trapping can be measured through a change in the threshold voltage of the

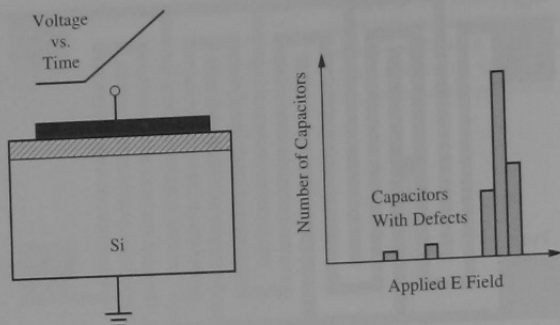


Figure 4-15 MOS capacitor test structure. The voltage applied to the gate is ramped up until significant current flows through the gate dielectric. The electric field at which this breakdown occurs is plotted on the right.

capacitor [trapped charge Q_T replaces mobile charge Q_M in Eq. (4.1)]. The amount of trapped charge is a measure of the quality of the dielectric layer (generally the smaller the better). Capacitors with defects due to particles or contaminants would show up as outliers as in Figure 4-15.

The most common chemical used in IC manufacturing is water. As we have seen, this water is filtered and deionized and is usually referred to simply as DI water. The most common impurities in DI water are ions. A simple measurement of water resistivity is often used to assess the “quality” of DI water. In pure water the resistivity is not infinite because the following reaction, at equilibrium, provides finite concentrations of H^+ and OH^- ions



Example At room temperature in equilibrium, $[H^+] = [OH^-] \approx 6 \times 10^{13} \text{ cm}^{-3}$ in water. The diffusivities of the H^+ and OH^- ions are $9.3 \times 10^{-5} \text{ cm}^2 \text{ sec}^{-1}$ and $5.3 \times 10^{-5} \text{ cm}^2 \text{ sec}^{-1}$, respectively. Using the Nernst-Einstein relationship $\mu = zqD/kT$, where z is the charge on the ion, calculate the ion mobilities, and the DI water resistivity.

Answer

$$\mu_{H^+} = \frac{qD}{kT} = \frac{9.3 \times 10^{-5} \text{ cm}^2 \text{ sec}^{-1}}{25.9 \times 10^{-3} \text{ V}} = 3.59 \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}$$

$$\mu_{OH^-} = \frac{qD}{kT} = \frac{5.3 \times 10^{-5} \text{ cm}^2 \text{ sec}^{-1}}{25.9 \times 10^{-3} \text{ V}} = 2.04 \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}$$

Using Eq. (1.1) for the resistivity, we finally have

$$\rho = \frac{1}{1.6 \times 10^{-19}[(3.59)(6 \times 10^{13}) + (2.04)(6 \times 10^{13})]} \cong 18.5 \text{ M}\Omega\text{cm}$$

DI water in an IC manufacturing facility is normally monitored to make certain its resistivity is at least 18 M Ω cm. When DI water is used for rinsing at the end of cleaning procedures, its resistivity can also be easily monitored to determine when the rinsing is complete.

4.4.2 Level 2 Contamination Reduction: Wafer Cleaning

The purpose of wafer cleaning is to remove particles and contaminants from the surface of wafers. Measurement techniques associated with cleaning thus focus on identifying and measuring the concentration of any residual contaminants left on the wafer after the cleaning process. Particles can be measured with the methods described in the previous section, so we will focus here on measuring and quantifying the concentrations of impurities left on the surface or introduced onto the surface by cleaning procedures. These techniques are basically surface analysis techniques. We will describe the more useful techniques here. More details on these and other methods are available in a number of excellent texts on this topic such as [4.18].

Surface analysis generally involves exciting the surface atoms of the material (using electron beams, X-rays, or ions) and then measuring a characteristic emission (of electrons, X-rays, or ions) from the surface atoms. Identification relies on the fact that the emitted species carry a unique atomic signature. Quantitative measurement relies on the fact that the magnitude of the signal is generally proportional to the concentration of the particular atomic species present in the sample. Figure 4-16 illustrates many of the commonly used measurement techniques.

When a high-energy electron beam strikes a solid, the incoming electrons give up their energy through a variety of physical processes, some of which result in electrons being emitted from the solid. Figure 4-17 illustrates the energy spectrum of these emitted electrons. Secondary electrons have the lowest energies (typically peaking at about 5 eV). They result from processes in which the incoming electrons collide inelastically with electrons in the solid and transfer enough energy to the target electrons to allow them to escape from the solid. These are the electrons that are normally collected and used to form the images in Scanning Electron Micrographs (SEMs) as described in Section 3.4.2.3 in Chapter 3. At the highest energies we find backscattered electrons which result from elastic collisions between the incoming electrons and atoms in the solid. (The incoming electrons “bounce” off atoms and are reflected back out of the sample, with energies similar to their initial energy.)

At intermediate energies, we find the Auger electrons, which are released through a process illustrated in Figure 4-18. This figure extends the band diagrams we introduced in Chapter 1 over a wider energy spectrum. The core electron energy levels (E_K , E_{L1} and

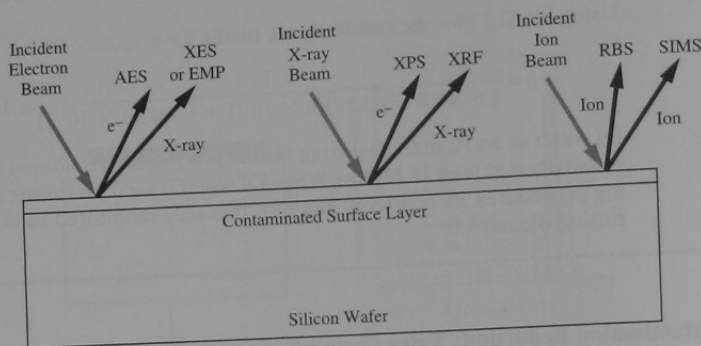


Figure 4-16 Surface analysis techniques used to identify and quantify contamination in IC manufacturing.

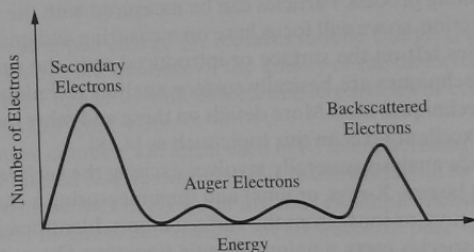


Figure 4-17 Energy distribution of electrons emitted from a solid under electron bombardment.

E_{L2}) represent those electrons nearer the nucleus which are more tightly bound. In this example, the incident (primary) electron with an energy of several keV enters the sample and “kicks out” a core electron from the E_K level (the secondary electron). A higher energy electron from the E_{L1} level drops into the vacant E_K level. The energy released in this process may be given off as an X-ray as illustrated on the right side of the figure. In this case the analysis method is called XES (X-ray Electron Spectroscopy) or EMP (Electron Microprobe). The released energy from the E_{L1} to E_K transition may also be given to a third electron as shown on the left of the figure. In this case this electron is ejected from the E_{L2} level and is known as an Auger electron. The analysis method in this case is AES (Auger Electron Spectroscopy). The transition in this example is called a KLL transition and involves three electrons as do all Auger processes. The Auger electron in AES or the X-ray in XES carry a unique signature of the element from which they were emitted because transitions between specific atomic energy levels are involved. Thus these techniques can be used for compositional analysis. Since the ejection of an Auger electron or an X-ray are competitive processes, the relative probabilities of the two events determine which is more likely. In general the lighter elements produce

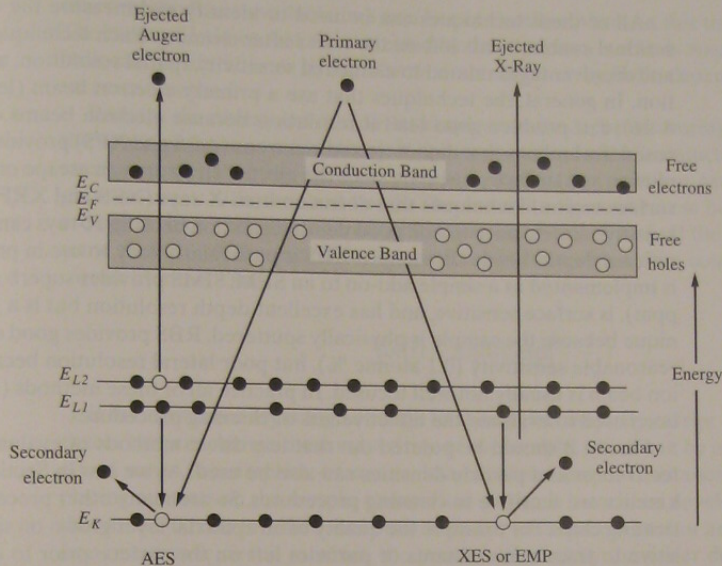


Figure 4-18 Band diagram representation of some of the processes used in surface analytical methods.

Auger electrons; the heavier elements X-rays. The crossover in probability between the two processes occurs around $Z = 33$ (arsenic).

The incident energy can also be provided by an X-ray as illustrated in Figure 4-16. The emitted particle can be either an electron or an X-ray, providing atom specific information as do the electron beam techniques described above. With an X-ray input, the analysis methods are known as XPS (X-ray Photoelectron Spectroscopy) when an electron is emitted or XRF (X-ray Fluorescence) when an X-ray is emitted.

Two other analysis methods are also conceptually shown in Figure 4-16. RBS (Rutherford Backscattering) detects ions which are elastically scattered by collisions with atomic nuclei in the substrate. Typically He ions with an energy of 1 to 3 MeV are used. This technique also provides compositional information because the backscattered ions have an energy equal to their incident energy minus an energy characteristic of the atom with which they collided and the depth in the sample at which the collision occurred.

The final analytic technique illustrated in Figure 4-16 is SIMS (Secondary Ion Mass Spectrometry). In this case, ions (usually O^+ or Cs^+) are used to bombard the sample. Collisions of these ions with sample atoms cause ejection (sputtering) of the sample atoms. The sputtered atoms are mass analyzed to provide compositional information on the sample. SIMS is the dominant method used today for dopant profiling and will be discussed in greater detail in Chapter 7 on dopant diffusion.

All of these techniques can be used to identify and measure the concentration of residual contaminants left on the wafer after cleaning. Each technique has advantages and disadvantages related to elemental sensitivity, spatial resolution, and depth resolution. In general, the techniques that use a primary electron beam (left side of Figure 4-16), can produce good lateral resolution because electron beams can be easily focused. Techniques that detect ejected electrons (AES and XPS) provide good depth resolution and surface sensitivity since the ejected electrons can escape only from the near surface region. Techniques that detect emitted X-rays (XES and XRF) are not surface sensitive and do not provide good depth resolution because X-rays can escape from significant depths inside the sample. XES is particularly easy to use in practice because it is implemented as a simple add-on to an SEM. SIMS provides superb sensitivity (often ppm), is surface sensitive, and has excellent depth resolution but is a destructive technique because the sample is physically sputtered. RBS provides good depth resolution, reasonable sensitivity (0.1 atomic %), but poor lateral resolution because the incident ion beam is usually not well focused. In practice all of these methods (and others) have been used to evaluate the effectiveness of cleaning procedures.

Finally, it should be pointed out that less direct methods of evaluating cleaning effectiveness and particle densities can also be used. As we saw in Section 4.2, oxidation kinetics are sensitive to cleaning procedures. So are many other processes used in fabricating chips. For example the quality of an epitaxial layer grown on silicon is very sensitive to trace contaminants or particles left on the surface prior to epitaxial growth. Thus the next process step after cleaning may often serve as a very real measure of how well the wafers were cleaned.

4.4.3 Level 3 Contamination Reduction: Gettering

Measuring the concentration and location of the metal atoms that gettering steps try to remove is a difficult task because the metal concentrations are normally very low. IC manufacturers, therefore, usually focus on measuring what they are really interested in—device properties and carrier lifetime.

Many studies of gettering have measured device characteristics directly, usually with large sample sizes to improve the statistical significance. Histograms of diode leakage currents, refresh time in DRAM cells, breakdown voltage in PN diodes, breakdown voltage in thin dielectric layers (as in Figure 4-15), low current BJT current gain, and other similar parameters have been widely used. Typically the mean value of a particular parameter and its standard deviation are measured in comparing various gettering procedures.

In some studies, more fundamental material parameters are measured, the most common of which is simply carrier lifetime. The hypothesis is that better carrier lifetimes correspond to better gettering treatments. Both carrier generation and recombination processes are important in semiconductor devices and so both the generation lifetime and the recombination lifetime are of interest. As we saw in Chapter 1, the two lifetimes are not equal in general. τ_G is usually much longer than τ_R . Different experimental approaches measure one or the other. We will describe in this section several of the more common methods for measuring carrier lifetime. More details on these and

other methods are available in a number of excellent texts on this topic such as [4.18]. The carrier lifetimes that are extracted by these various methods may be dominated by bulk carrier lifetimes or by surface recombination or by both, depending on the measurement method.

Perhaps the oldest and one of the simplest methods for measuring lifetime involves the use of photoconductive decay. If light is shined on a semiconductor sample with an energy greater than the bandgap, excess hole electron pairs are generated when the photons break Si-Si covalent bonds. The concentrations of excess holes and electrons are equal since each photon interaction generates one of each. If the light is shut off at some time $t = 0$, the excess carriers will recombine with a time constant equal to the carrier lifetime. Thus

$$\Delta n(t) = \Delta n(0) \exp\left(-\frac{t}{\tau_R}\right) \quad (4.6)$$

If the light intensity is high enough to change the majority carrier concentration, the excess carriers change the conductivity of the sample, which can be measured in a variety of ways. The simplest, of course, is a measurement of resistance, but this involves placing contacts on the sample. Contactless methods have been developed which use either an RF circuit in which the semiconductor resistance is part of an RF bridge [4.19], or a microwave circuit in which the power reflected by the wafer depends on its conductivity [4.20]. Eq. (4.6) is not strictly correct under these high level injection conditions, but it can be modified to extract an effective carrier lifetime [4.18]. The extracted τ_R usually reflects both bulk and surface recombination although it is possible to design the experiment to maximize one or the other. For example, the depth at which the photons are absorbed depends on their wavelength with higher energy photons being absorbed closer to the surface. Thus choosing a high photon energy would maximize surface recombination effects.

One of the most common methods for extracting carrier lifetime is the use of MOS Capacitance-Voltage measurements (MOS CV). The attraction of this technique is that it measures exactly what is important in MOS devices and these devices, of course, dominate modern ICs. We will discuss this method in much more detail in Chapter 6 (Section 6.4), because it is widely used to measure the properties of the Si/SiO₂ interface. Readers interested in a more complete description of the technique may wish to read that material at this point. Here we will focus only on a specific application, the extraction of carrier lifetime.

Figure 4-19 again shows the basic MOS capacitor structure—a metal or doped polysilicon electrode on an insulator (usually thermally grown SiO₂) on a silicon substrate. This is of course the basic gate structure of the MOS transistor and the capacitor is fabricated identically to the transistor. If the gate electrode is stepped to $+V_G$ (typically +5 volts) at $t = 0$, the vertical electric field produced in the silicon will push the mobile carriers (holes since this example is P type), away from the surface. This creates a depleted region near the surface in which the depletion region charge (due to the acceptor doping ions) exactly balances the positive charge on the gate electrode. $Q_G = Q_D$. This is not the final equilibrium state of the structure, however. After some time, an inversion layer of electrons forms. We described this process qualitatively in Chapter 1 in

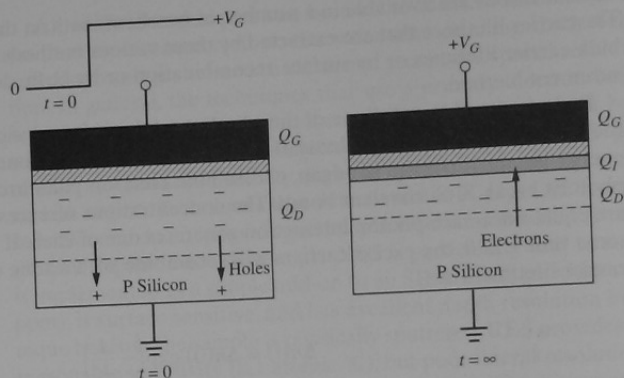


Figure 4-19 MOS CV measurement to extract the carrier generation lifetime. At $t = 0$ the gate voltage is stepped to $+V_G$. The capacitor goes into deep depletion. As carriers are generated to form the inversion layer, the depletion region shrinks.

Section 1.4.2 in connection with MOS transistors. Once the inversion layer forms, the gate charge is balanced by the sum of the depletion region charge and the inversion layer charge. $Q_G = Q_D + Q_I$. To satisfy this balance, the depletion layer shrinks as the inversion layer forms.

The electrons needed for the inversion layer are generated in the silicon through SRH generation. Thus if we can measure the time dependence of this process, we can extract the carrier generation lifetime. The simplest way to do this is to measure the capacitance of the structure. The MOS structure behaves very much like a simple parallel plate capacitor with the upper plate being the metal electrode and the lower plate being the undepleted silicon substrate. As the depletion layer shrinks, the distance between these plates decreases and therefore the capacitance rises. The time required for the capacitance to reach its final equilibrium value is given by [4.21]

$$t = 10 \left(\frac{N_A}{n_i} \right) \tau_G \quad (4.7)$$

where τ_G is the generation lifetime and N_A is the doping concentration in the substrate. For typical lifetime values of 10 μsec to 1 msec, the measured capacitor response time is thus in the range of seconds to minutes. The extracted τ_G value is an effective value that reflects the contributions of bulk and surface generation.

Another very simple electrical measurement that can be made to estimate carrier lifetime makes use of a PN diode and is called the open circuit voltage decay method. Figure 4-20 illustrates this method. A power supply and resistor are used to establish a DC forward current through the diode. At $t = 0$, the switch is opened, forcing the circuit current to 0. We saw in Chapter 1 (Section 1.4.1), that when a diode is forward biased ($V_D \approx 0.7$ volts), minority carriers are injected across the diode. When the switch

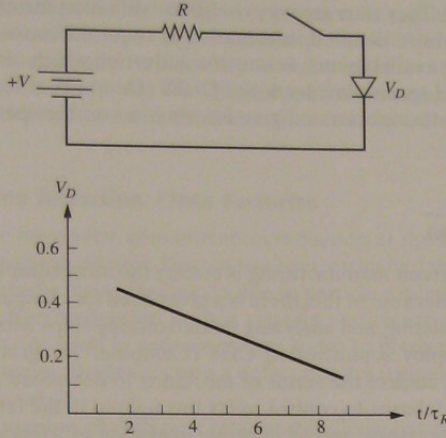


Figure 4-20 Diode open circuit voltage decay to extract the carrier recombination lifetime. The switch is opened at $t = 0$. The diode voltage decays due to carrier recombination.

is opened, these carriers are gradually eliminated in the diode by recombination and the diode voltage gradually returns to 0. The simplest analysis of this process leads to the following equation:

$$V_D(t) = V_D(0) - \frac{kT}{q} \ln \left(\operatorname{erfc} \sqrt{\frac{t}{\tau_R}} \right) \quad (4.8)$$

where erfc is the complementary error function (tabulated in the appendix). This equation can be derived from the ideal diode current voltage relationship described in Chapter 1 [Eq. (1.23)] [4.18]. For t/τ_R larger than ≈ 4 , this produces a straight line as shown in Figure 4-20. The slope of the line gives the recombination lifetime from

$$\tau_R = \frac{kT/q}{dV_D/dt} \quad (4.9)$$

Since PN diodes are always part of any integrated circuit, this simple measurement is usually straightforward to make. As was the case in the MOS CV method described above, the extracted τ_R value is usually an effective value that reflects the contributions of bulk and surface recombination.

The experimental methods we have described for extracting carrier lifetime do not give detailed information on the spatial distribution of the SRH centers or on the species (Au, Cu, Fe, etc.) responsible for the lifetime degradation. Nonetheless, they are simple measurements and do provide a clear indication of whether carrier lifetime is

good or bad. They thus are very useful in evaluating the effectiveness of gettering procedures. If more detailed information is required, more sophisticated measurement methods are available such as neutron activation which can provide information on the specific SRH species involved, and DLTS (Deep Level Transient Spectroscopy) which can identify the species and give information on the spatial location of the contaminants [4.18].

4.5 Models and Simulation

Integrated circuit manufacturing is among the most complex industrial processes practiced today. Because of this, there is a great need for computer tools for monitoring, controlling, simulating, and analyzing manufacturing steps. Most modern IC manufacturing facilities employ sophisticated CIM (Computer Integrated Manufacturing) tools to monitor and control the status of machines, to download process recipes from central computer databases, to control wafer throughput in the fab, to store information about factory performance, and to improve operating efficiency. These topics are generally outside the scope of this book, but there are extensive references in the literature on these topics. See for example [4.22, 4.23].

One issue with respect to these kinds of CIM tools that is insightful is the use of these tools and the information they provide to quickly ramp up high yield manufacturing on new products in new factories. Table 4-2, taken from the 1997 NTRS, illustrates this point.

The last two rows (in bold) are the new information in this table, compared to Table 4-1 which we previously discussed. Notice that today it is common to introduce a new product into manufacturing with a yield less than 50%. Projections for the future suggest that it will be necessary for new products to achieve much higher yields right from the start. The reason for this is that it is often the case that the largest profits on a new product are made when the product is new (higher prices can often be charged for new, better products and there is usually less competition when a product is new). Today it takes several years before a new product achieves a mature level. That time is also projected to decrease dramatically in the future. The only way in which these dramatic improvements will be achieved is through better understanding and control of the manufacturing process.

Table 4-2 Semiconductor industry projected progress and the implications of this progress for chip yield and manufacturing yield ramp-up [4.1].

Year of First DRAM Shipment	1997	1999	2003	2006	2009	2012
Minimum Feature Size	250 nm	180 nm	130 nm	100 nm	70 nm	50 nm
Wafer Diameter (mm)	200	300	300	300	450	450
DRAM Bits/Chip	256M	1G	4G	16G	64G	256G
Initial Yield Level (%)	25	50	80	85	88	90
Time to Mature Yield Level (years)	4	3	1	0.8	0.6	0.5

Simulation and modeling have begun to be applied to the manufacturing related topics discussed in this chapter, although the level of sophistication of these simulation tools is generally rudimentary in comparison to the TCAD tools discussed in other chapters of this book. We will describe in this section some of the current modeling and simulation efforts and also point out where future work is needed.

4.5.1 Level 1 Contamination Reduction: Clean Factories

As we have seen, contamination reduction at the factory level is primarily associated with particle control. Since it has been estimated that 75% of the yield loss in modern IC manufacturing facilities is due to particle contamination, there has been a considerable effort to develop models relating particles to wafer yield. There are no computer simulation tools in widespread use today to predict particle deposition rates on wafers in factories. However given a defect density (usually measured), yield models are commonly used.

A number of empirical relationships have been used to describe the distribution of particle sizes in air. As we saw earlier, we are mostly concerned with particles with sizes between 10 nm and 10 μm . The actual size distribution of particles has been described with a power law function where $N(d_p)$, the number of particles larger than some diameter d_p , is given by

$$N(d_p) = K(d_p)^{-3} \quad (4.10)$$

where K is a constant [4.24]. This formula must fail for very small particles and so in practice, the distribution function, which cannot be easily measured at small sizes, is assumed to also fall off at small sizes as illustrated in Figure 4-21.

The second issue with respect to particles and yield is how large a particle needs to be to cause a defect. Generally, particles on the order of the technology minimum feature size or larger will cause defects unless they happen to occur in a noncritical area of

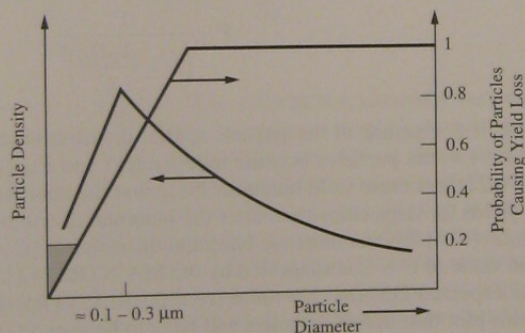


Figure 4-21 Left axis: Empirical particle size distribution in room air.
Right axis: Probability of a particle causing a chip defect.

the chip. For example, an unintended isolated area of metal due to a photolithography defect during the metal etching may not cause a yield problem if it is well away from other metal lines. If the particle is in a critical area, then the probability it will cause a defect behaves empirically as shown in Figure 4-21. The probability is 1 for large particles and falls off as the particle or defect size decreases. Note that the probability may not be zero for very small particles, depending on the type of process involved (gray area near the origin in Figure 4-21). For example, a pinhole in a gate oxide is a catastrophic failure no matter how small the pinhole is.

A wide variety of yield models have been developed to describe IC manufacturing [4.24–4.26]. The simplest of these assumes Poisson statistics and independent randomly distributed defects each large enough to cause a chip defect. Thus,

$$Y = \exp^{-A_c D_o} \quad (4.11)$$

where Y is the yield, A_c is the critical area of the chip and D_o is the defect density. A_c may be less than the total chip area because not all of the area may be sensitive to particular defects. If some fraction of the wafer always produces zero yield (around the edges for example), or if some die sites are occupied by test devices, then

$$Y = (1 - G) \exp^{-A_c D_o} \quad (4.12)$$

where G is the fraction of the wafer where all the ICs fail. If information on particle size distribution and/or the probability of a particular size defect causing a fault (i.e., the information in Figure 4-21) is available, it can be used in defining A_c and/or D_o .

A variety of more sophisticated yield models have been investigated in recent years mainly driven by the inadequate predictions of the simple Poisson model. This model tends to underpredict the yield of large chips because of its assumptions about independent and randomly distributed defects. A somewhat better model is the negative binomial model that recognizes the fact that defects may not be completely randomly distributed and independent. In this model,

$$Y = \frac{1}{\left(1 + \frac{A_c D_o}{C}\right)^C} \quad (4.13)$$

where C is a measure of the particle spatial distribution called the clustering factor. When $C \rightarrow \infty$, the particles become independent and Eq. (4.13) reduces to Eq. (4.11). Figure 4-22 plots these yield functions. Note that the Poisson distribution predicts much lower yields for large chips than does the binomial distribution.

Figure 4-23 plots the negative binomial distribution ($C = 2$) for various defect levels. The value of $C = 2$ is suggested by the SIA NTRS [4.1]. Superimposed on this plot are the expected DRAM chip areas from the SIA NTRS through 2012. It is apparent from this plot that defect densities will have to be around 0.01 cm^{-2} to achieve manufacturable yields after the turn of the century. This number is consistent with the projections in Table 4-1 earlier in this chapter.

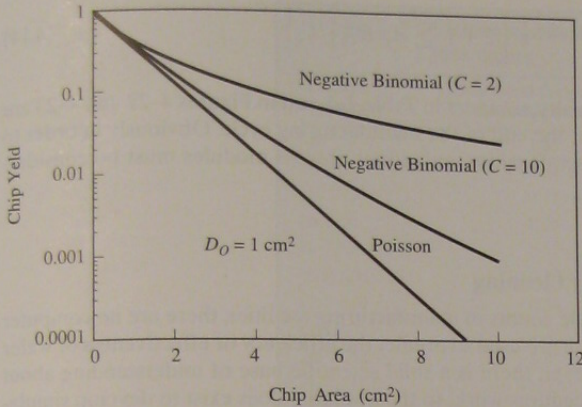


Figure 4-22 Plots of various yield functions assuming $D_0 = 1 \text{ defect/cm}^2$.

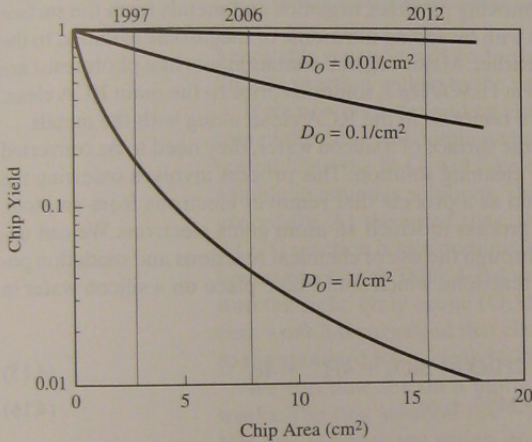


Figure 4-23 Plots of the negative binomial yield function versus chip area for various defect densities, $C = 2$ in all plots. The vertical lines correspond to DRAM chip areas in specific years from the SIA NTRS [4.1].

In manufacturing environments, it is generally very useful to monitor defect levels for as many of the individual process steps or process modules as possible. This can be done as described in Section 4.4.1 by running blank wafers through a particular process step or series of process steps and then using a particle counter to estimate the particles introduced by that particular step. If data like this are available, then the overall yield associated with a manufacturing process flow is simply the product of the individual module yields, assuming that the individual yields are independent. Any of the yield distribution functions described above can be used to describe the overall yield. For example, for the negative binomial distribution, the overall yield would be given by

$$Y = \prod_{i=1}^{\text{levels}} \left(1 + \frac{A_{Ci} D_{0i}}{C_i} \right)^{-C_i}$$

$$\equiv \left\{ 1 + \left(\sum_{i=1}^{\text{levels}} A_{Ci} D_{Oi} \right) / C_i \right\}^{-C_i} \quad (4.14)$$

Note that the defect density numbers in Table 4-1 and in Figures 4-22 and 4-23 are the total defect densities at the end of the manufacturing cycle. Obviously in order to achieve these overall defect densities, individual process modules must be considerably better.

4.5.2 Level 2 Contamination Reduction: Wafer Cleaning

As was the case with particle counts in manufacturing facilities, there are no computer simulation tools today which are used to predict the efficiency or effectiveness of wafer cleaning procedures. However, there is a solid scientific base of understanding about how and why cleaning procedures work, so the potential does exist to develop simulation tools should they be needed in the future.

Wafer cleaning is aimed at removing particles, organics, and metals from the surface of wafers. Particles are best dealt with by adding ultrasonic or megasonic agitation to the cleaning baths as was discussed earlier. Major organic contaminants like photoresist are removed in an oxygen plasma or in $\text{H}_2\text{SO}_4/\text{H}_2\text{O}_2$ solutions prior to the main RCA clean. Any remaining trace organics are removed in the RCA clean along with the metals.

To remove metal atoms from the surface of a silicon wafer, they need to be converted into ions that are soluble in the cleaning solution. This process involves oxidizing the metal atoms. Oxidation is defined as a process that removes electrons from an atom, while reduction is the opposite process in which an atom gains electrons. We can understand the various processes through the use of chemical reactions and oxidation potentials. Consider the following reactions, which could take place on a silicon wafer in an aqueous solution:



where M represents a metal atom and z is the charge on the metal ion. Oxidation reactions go to the right, reduction reactions to the left. If electrons are available to drive reactions, the stronger oxidant will consume them. Table 4-3 summarizes the standard oxidation potentials of a number of reactions of interest in silicon cleaning (Cu, Ni, and Fe are usually the dominant impurities in silicon). These potentials are the open circuit voltages that would be measured in an electrochemical cell operating with the given reaction, referenced to a standard hydrogen cell. In this table, stronger oxidants have more negative oxidation potentials. In the chemical reactions, oxidants are on the right side of the equations, reductants are on the left side. Standard chemistry references discuss these concepts in more detail. (See for example [4.27].)

Suppose we now have a water solution containing a silicon wafer and metal atoms (dissolved in the liquid or on the wafer surface). Consider the SiO_2/Si and Fe^{3+}/Fe reactions as examples. Here the Fe^{3+}/Fe reaction has the stronger oxidation potential and

Table 4-3 Oxidation-reduction reactions for a number of species of interest in silicon wafer cleaning

Oxidant/ Reductant	Standard Oxidation Potential (volts)	Oxidation-Reduction Reaction
Mn^{2+}/Mn	1.05	$\text{Mn} \leftrightarrow \text{Mn}^{2+} + 2\text{e}^-$
SiO_2/Si	0.84	$\text{Si} + 2\text{H}_2\text{O} \leftrightarrow \text{SiO}_2 + 4\text{H}^+ + 4\text{e}^-$
Cr^{3+}/Cr	0.71	$\text{Cr} \leftrightarrow \text{Cr}^{3+} + 3\text{e}^-$
Ni^{2+}/Ni	0.25	$\text{Ni} \leftrightarrow \text{Ni}^{2+} + 2\text{e}^-$
Fe^{3+}/Fe	0.17	$\text{Fe} \leftrightarrow \text{Fe}^{3+} + 3\text{e}^-$
$\text{H}_2\text{SO}_4/\text{H}_2\text{SO}_3$	-0.20	$\text{H}_2\text{O} + \text{H}_2\text{SO}_3 \leftrightarrow \text{H}_2\text{SO}_4 + 2\text{H}^+ + 4\text{e}^-$
Cu^{2+}/Cu	-0.34	$\text{Cu} \leftrightarrow \text{Cu}^{2+} + 2\text{e}^-$
$\text{O}_3/\text{H}_2\text{O}$	-1.23	$2\text{H}_2\text{O} \leftrightarrow \text{O}_2 + 4\text{H}^+ + 4\text{e}^-$
Au^{3+}/Au	-1.42	$\text{Au} \leftrightarrow \text{Au}^{3+} + 3\text{e}^-$
$\text{H}_2\text{O}_2/\text{H}_2\text{O}$	-1.77	$2\text{H}_2\text{O} \leftrightarrow \text{H}_2\text{O}_2 + 2\text{H}^+ + 2\text{e}^-$
O_3/O_2	-2.07	$\text{O}_2 + \text{H}_2\text{O} \leftrightarrow \text{O}_3 + 2\text{H}^+ + 2\text{e}^-$

hence this reaction will go to the left, plating out Fe atoms on the wafer surface. The SiO_2/Si reaction will go to the right, oxidizing the silicon.

If we now consider adding H_2O_2 to the solution, the $\text{H}_2\text{O}_2/\text{H}_2\text{O}$ reaction near the bottom of the table will dominate the Fe^{3+}/Fe reaction because of its stronger oxidation potential. Thus the H_2O_2 will take electrons from the metal atoms, creating ions which are soluble in aqueous solutions. This is the basis for the use of H_2O_2 in the RCA cleaning procedure. At the same time, the silicon will also be oxidized by the H_2O_2 since the SiO_2/Si reaction is also dominated by the $\text{H}_2\text{O}_2/\text{H}_2\text{O}$ reaction. The general rule is that the lowest reaction in Table 4-3 dominates, going to the left and driving all reactions above it to the right. Only ozone (O_3) has a stronger oxidation potential than H_2O_2 . Some recent work has suggested that cleaning procedures based on O_3 may actually be superior to the standard RCA clean [4.28]. We will discuss these possibilities in Section 4.6.

We can now see on a more fundamental level how the RCA cleaning procedure works. The first solution (SC-1) is a high-pH solution consisting of 5 H_2O :1 H_2O_2 :1 NH_4OH in which the wafers are placed at 70–80°C for 10 minutes. This solution oxidizes organic films into water soluble compounds (CO_2 , H_2O , etc.) and complexes Group IB and IIB metals as well as other metals like Au, Ag, Cu, Ni, Zn, Cd, Zn, Co, and Cr. For example, Cu forms a soluble $\text{Cu}(\text{NH}_3)_4^{+2}$ amino-complex.

The second solution (SC-2) is a low-pH solution consisting of 6 H_2O :1 H_2O_2 :1 HCl in which the wafers are placed at 70–80°C for 10 minutes. This solution removes alkali ions and cations like Al^{+3} , Fe^{+3} , and Mg^{+2} that form NH_4OH insoluble hydroxides in basic solutions like SC-1. The SC-2 solution also completes the removal of metallic contaminants such as Au through reactions like Eq. (4.16) that may not have been completely removed by the SC-1 step. Both SC-1 and SC-2 solutions depend on the strong oxidation potential of H_2O_2 to remove metal atoms from the wafer surface. The kinetics of the cleaning process depend of course on the reaction rate constants associated with reactions like Eq. (4.16). These are temperature dependent so that the necessary cleaning time is reduced at elevated temperatures.

4.5.3 Level 3 Contamination Reduction: Gettering

Many models have been developed in recent years for metal gettering. The process seems conceptually simple as illustrated in Figure 4-24. If metal atoms are present throughout the wafer, three steps need to occur to getter them. First they need to be freed from whatever sites they presently occupy. Second, they need to migrate to the gettering sites, either the wafer bulk or the wafer backside. Third, they need to be trapped at the gettering sites. Models that have been developed for these processes attempt to treat all three parts of this process. There is still considerable uncertainty about the exact mechanisms responsible for gettering and as a result, the models that we will discuss are not universally accepted. Nevertheless, they do provide a coherent physical picture of how gettering probably operates.

4.5.3.1 Step 1: Making the Metal Atoms Mobile

Most of the metals of interest in gettering (those that create deep levels in silicon), exist in two states in the silicon crystal, either occupying a substitutional lattice site M_s , or an interstitial site M_i . These metals are pure interstitials in the latter case, and generally do not occupy interstitial sites (see Section 3.2.2 in Chapter 3). Metals can diffuse in silicon in either state, but their diffusivities are generally orders of magnitude higher in the interstitial form. We will discuss diffusion mechanisms in much more detail in Chapter 7, but it should be easy to imagine that substitutional diffusion involves hopping from lattice site to lattice site, a process that involves breaking bonds. Interstitial diffusion simply involves moving through the interstitial spaces between lattice atoms with no bond breaking. In general the interstitial process is much faster. The data in Figure 4-8 illustrate this. Interstitial diffusers like Au and Cu have orders of magnitude higher diffusivities than substitutional dopants like P or even Si itself. Thus step 1—

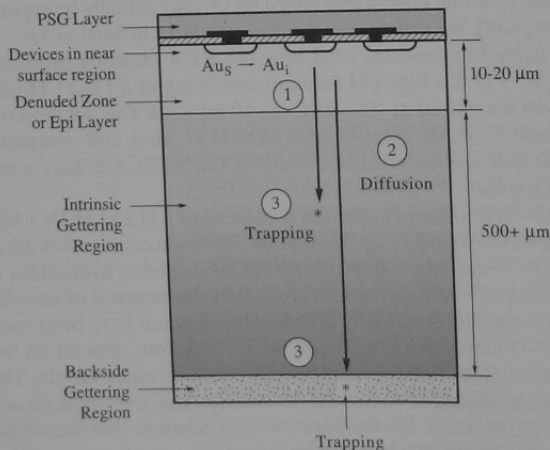
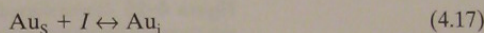


Figure 4-24 Three-step gettering process: release, diffusion, and trapping. Au is used as a specific example.

making the metal atoms mobile—generally involves getting the metal atoms into their interstitial form.

The importance of this first step depends on the specific metal atom being considered. Some metals (Cu and Ni, for example) have much higher solubilities in their interstitial form and so most such atoms are in the interstitial form naturally. The release step is therefore unnecessary unless these atoms have formed precipitates due to very high concentrations. A second category of metals (Au and Pt are examples) have much higher solubilities in their substitutional form but diffuse rapidly once they become interstitials. For these elements, the release process is very important. Finally, there are some elements (Ti and Mo, for example) that are primarily substitutional, but which have relative slow diffusion rates in either interstitial or substitutional form (see Figure 4–8). These elements are the most difficult to getter effectively because they will not diffuse over long distances (to the wafer backside for example), unless very long thermal cycles are used.

Metals such as Au and Pt are the most interesting case, because they need to be converted from substitutional to interstitial sites to be gettered effectively. This can happen through one of two reactions, where we have taken Au as a specific example:



In Eq. (4.17), a silicon interstitial exchanges places with the Au atom on a lattice site, creating a Au interstitial. This is called the “kick-out” mechanism. In Eq. (4.18), the Au atom jumps off its lattice site, creating a silicon vacancy. This is known as the dissociative or Frank-Turnbull mechanism [4.29]. Process conditions that create excess concentrations of I should be useful in driving Au_s to Au_i and thus help the gettering process. Similarly, processes that create excess V should hinder the gettering process by driving Au_i back on to substitutional sites. It has been proposed that many of the effective gettering procedures inject excess concentrations of I and thus drive Eq. (4.17) to the right [4.30]. For example, we will see in Chapter 7 that high-concentration phosphorus diffusions produce very large supersaturations of I . Backside P diffusions are known to be among the most effective gettering treatments and such diffusions are thought to inject large quantities of I . Similarly, ion implantation damage creates excess I as we will see in Chapter 8. Such damage is also effective at backside gettering. Internal gettering involves oxidation (SiO_2 formation) and this process also injects I (see Figure 3–25).

4.5.3.2 Step 2: Metal Diffusion to the Gettering Site

Once the metal atoms are in their interstitial (fast diffusing) form, they can diffuse through the wafer to the backside sink, or to the trap sites associated with SiO_2 precipitates if internal gettering is being used. Obviously the diffusion distance is a lot smaller for the intrinsic gettering case, which is one of the reasons why it is more popular today.

The expected general shape of the Au profiles as they outdiffuse is illustrated in Figure 4–25. The Au concentration goes to zero at the wafer backside where the Au atoms are trapped. We will see in Chapter 7 how to calculate the actual detailed shape of such profiles, but the general shape should be intuitive. As time increases, the depth over

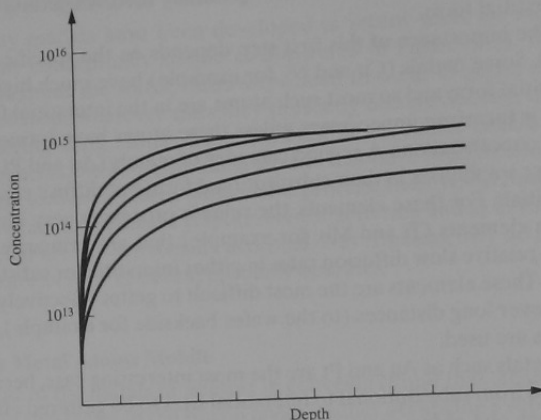


Figure 4-25 General shape of Au outdiffusion profiles if simple outdiffusion dominates the gettering process. Starting with a flat profile (10^{15} cm^{-3}), the concentration drops off with time through the wafer as the Au diffuses to the back surface (left side). This simulation was done with Silvaco's SSUPREM IV, a process simulator we will discuss in more detail in later chapters.

which the Au is gettered increases away from the wafer backside, until at long enough times, the Au is removed entirely from the wafer.

Some actual experimental measurements of Au profiles are shown in Figure 4-26 [4.31]. Note that the shape is quite different from the expected simple outdiffusion profile in Figure 4-25. The experimental data can be understood in connection with the model in Figure 4-27 and the diffusivity of the I species indicated in Figure 4-8 [4.30]. There is still great uncertainty about the value of the diffusion coefficient of the silicon interstitial (see Chapter 7), so we have indicated a region in which it likely lies in Figure 4-8. Note that this diffusion coefficient is much faster than dopant diffusivities, but significantly slower than metal atom diffusivities. Thus if the gettering process depends on the backside gettering region for injection of I to drive reaction (4.17), then the slowest (rate limiting) part of the process will be the indiffusion of the I from the backside. As quickly as silicon interstitials reach a particular depth from the backside, the Au in that region are driven to Au_i which then quickly diffuse to the wafer backside where they are trapped. This qualitative explanation leads to Au profiles like those in Figure 4-26. The sharp drop in the Au profile which moves away from the backside as time goes on, occurs at the depth corresponding to how far the silicon interstitials have diffused in from the backside at that particular time.

Other types of gettering processes also often lead to Au profiles like Figure 4-26, so it is quite possible that the process described here operates in many common gettering methods. When gettering is performed at higher temperatures, the equilibrium concentration of I may be high enough that reaction (4.17) can proceed without the excess I

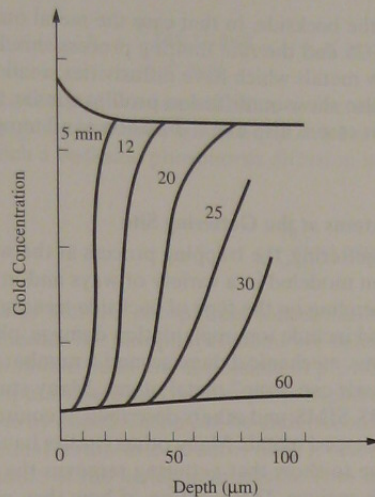


Figure 4-26 Experimental 1000°C Au outdiffusion profiles during gettering to the wafer backside. (After [4.31].)

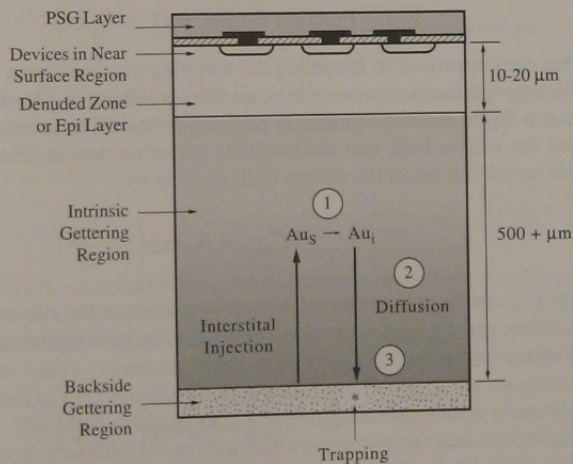


Figure 4-27 Schematic representation of the flux of silicon interstitials and M_1 during back surface gettering. (After [4.30].)

provided by the backside. In that case the metal outdiffusion profiles should be more like Figure 4-25 and the rate limiting process should be simply the metal atom diffusivity. The few metals which have diffusivities smaller than I (Ti for example in Figure 4-8) should also show outdiffusion profiles similar to Figure 4-25 since the rate limiting step in this case is also metal diffusion at all temperatures.

4.5.3.3 Step 3: Trapping the Metal Atoms at the Gettering Site

For extrinsic gettering, the trapping process at the wafer backside (process 3 in Figure 4-27) has been modeled in a variety of ways and in fact, different models may be appropriate depending on the type of backside treatment used. Backside treatments that have been used include ion implantation damage, phosphorus diffusion, laser damage, polysilicon films, mechanical damage, and a number of others. It is quite clear that any of these methods can “trap” metal atoms. Many studies have used physical measurements like RBS, SIMS, and others described in connection with Figure 4-16, to directly measure the trapped metals. Many other studies have used measurements of bulk generation lifetime to show that gettering removes the metal atoms (SRH centers) from the wafer bulk regions. The question of how they are trapped, however, is less clear. Physical damage to the wafer backside (ion implantation, laser or mechanical damage, or polysilicon films containing many grain boundaries), is generally believed to trap the metal atoms at defect sites. This trapping process is most simply characterized by some binding energy E_B of the metal atom to the trap site. In that case, the metal atoms are kept trapped at low temperatures. As T increases, more and more of them are able to escape the traps. Clearly we would want E_B as large as possible.

$$\text{Fraction Bound} = (1 - K_1 \exp^{-E_B/kT}) \quad (4.19)$$

Another approach to modeling the trapping process uses a segregation model [4.32, 4.33]. Segregation is a process that we used in Chapter 3 to describe dopant behavior between liquid and solid phases in growing crystals. The same concept applies here, between the silicon bulk and the backside gettering region. The solubility of a substitutional metal like Au in the silicon bulk is given by

$$C_{\text{Au,Si}} = N_{\text{Si}} \exp\left(-\frac{E_{A1}}{kT}\right) \quad (4.20)$$

where $C_{\text{Au,Si}}$ is the substitutional Au concentration in the silicon, N_{Si} is the density of silicon atoms and E_A is an activation energy. Similarly, the solubility of Au in the gettering region is given by

$$C_{\text{Au,G}} = N_G \exp\left(-\frac{E_{A2}}{kT}\right) \quad (4.21)$$

where N_G is the density of gettering sites. The segregation coefficient of the Au between the two regions is now defined as

$$k_O = \frac{C_{Au,G} + C_{Au,Si}}{C_{Au,Si}} = 1 + K_2 \exp\left(-\frac{E_{A1} - E_{A2}}{kT}\right) \quad (4.22)$$

If $E_{A1} - E_{A2}$ is a positive number, then k_O decreases as temperature goes up, providing the same qualitative behavior as Eq. (4.19). For the specific case of phosphorus gettering in which a backside phosphorus diffusion is used, Baldi [4.32] estimate that k_O is given by

$$k_O = 1 + \frac{N_G}{5 \times 10^{22}} \exp\left(\frac{0.82 \text{ eV}}{kT}\right) \quad (4.23)$$

which indicates that the phosphorus concentration should be as high as possible (to maximize N_G since the phosphorus atoms are believed to provide the N_G sites) and the temperature should be as low as possible to provide the best trapping.

Other proposals have also been made to explain the trapping. Enhanced solubility of metals in heavily doped silicon can be understood by taking Au again as an example. In its substitutional form, Au introduces deep levels (both acceptor and donor) which are near the middle of the bandgap in silicon. Thus in N-type silicon, in which the Fermi level is close to the conduction band, Au tends to be an acceptor. (The acceptor level is below E_F and hence is filled.) In P-type silicon, Au tends to be a donor. If we consider the reaction



then we can write from the definition of the equilibrium constant that

$$K_{eq} = \frac{[\text{Au}^-]}{[\text{Au}][e^-]} = \text{constant} \quad (4.25)$$

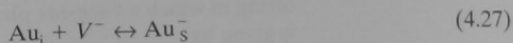
where $[]$ denotes equilibrium concentrations of the various species. Since Eq. (4.25) holds both in intrinsic and in doped silicon where $[e^-] = n_i$ and $[e^-] = n$, respectively, we have that

$$\frac{[\text{Au}_i^-]}{[\text{Au}]n_i} = \frac{[\text{Au}_n^-]}{[\text{Au}]n} \quad \text{or} \quad \frac{[\text{Au}_n^-]}{[\text{Au}_i^-]} = \frac{n}{n_i} \quad (4.26)$$

Thus the solubility of the Au acceptor is much higher in N-type silicon. For example, we saw in Chapter 1 that $n_i = 7.14 \times 10^{18} \text{ cm}^{-3}$ at 1000°C . If the gettering step is carried out at this temperature, with an N-type doping level of 10^{21} cm^{-3} , the solubility is over 100 times greater in the N-doped region than in the wafer bulk. This model has been used to explain why phosphorus-doped regions are effective getters.

The ion pairing model [4.34] suggests that a large atom like Au prefers to pair with a small atom like P, forming AuP complexes, because this minimizes the strain in the crystal. There would also be coulombic interactions between the Au and P because $\text{Au}^- \text{P}^+$ complexes could form. The solubility of such complexes would be higher in N-type Si than in lightly doped silicon because of the higher P concentration.

Finally, we saw in Chapter 3 that the point defect concentrations are much higher in doped silicon than in intrinsic silicon. Au diffuses primarily by an interstitial mechanism as we have seen, so when it arrives at the gettering site at the wafer backside, it needs to find a lattice site on which to sit since in equilibrium, the Au prefers to be predominantly on substitutional sites. One possible reaction is



The much higher V^- population in N^+ silicon would drive this reaction to the right, with the vacancies essentially providing lattice sites for the Au atoms.

Intrinsic gettering relies on trapping the metal atoms at defects in the wafer bulk. These defects are generally dislocations and/or stacking faults that are caused by the precipitation of oxygen while forming SiO_2 precipitates in CZ silicon. We discussed the process of precipitate formation in some detail in Chapter 3 (Figure 3–25) and again briefly earlier in this chapter. As the SiO_2 precipitates form, a net volume expansion takes place because the oxide occupies more volume than the silicon crystal. We modeled this process in Chapter 3 with the following equation:

$$(1 + 2\gamma)\text{Si} + 2\text{O}_i + 2\beta V \leftrightarrow \text{SiO}_2 + 2\delta I + \text{stress} \quad (4.28)$$

where γ is the number of I that contribute to the precipitation process per O_i atom joining the SiO_2 , and β is a similar fraction for V . O_i represents the interstitial oxygen atom and Si is simply a lattice site in the crystal. The terms involving point defects model the fact that the volume needed for formation of the SiO_2 can come partly by consuming vacancies or ejecting interstitials. We have included a stress term on the right-hand side because it is unlikely that the point defects could provide all the necessary volume and so it is likely that the region surrounding the SiO_2 precipitates will be under stress with the precipitate and the surrounding silicon under compression. The dislocations and stacking faults that form around the precipitates are natural sites for metal atoms to be trapped. Dislocations have regions of both compressive and tensile stress, so they provide “homes” for metal atoms that are smaller and larger, respectively, than silicon atoms [4.33]. Many studies have shown the metal atoms located around these dislocations. These trapping processes can be qualitatively modeled by expressions similar to Eqs. (4.19) or (4.22) since the metal trapping at dislocations in intrinsic gettering is similar to what occurs at the wafer backside in extrinsic gettering processes that produce similar damage.

It is interesting to note that the reaction given by Eq. (4.28) injects I . In fact, surface oxidation of silicon does the same thing (Chapter 6). These injected I coalesce to form the dislocations and stacking faults around the precipitates. They can also help to drive reaction (4.17) to the right, freeing up substitutional metal atoms so they can diffuse to the gettering sites. Finally, we would expect that intrinsic gettering would be a better approach for gettering metals which have relatively small diffusion coefficients (like Ni), because they would not have to diffuse nearly as far as they would if extrinsic gettering were being used.

The variety of models described in this section should give some indication of the ways in which gettering is understood today. There is still considerable controversy

about many of the ideas described in this section. However, it is likely that the qualitative concepts described here do play important roles in the gettering process.

4.6 Limits and Future Trends in Technologies and Models

The goals for the future are fairly clear—cleaner environments in which to process wafers, better cleaning procedures that remove particles and trace contaminants more reliably, and better gettering procedures for eliminating contaminants from device regions of wafers. In each of these areas, there is considerable ongoing research that we will briefly describe in this section.

Traditionally, semiconductor manufacturing facilities have used the approach of providing clean rooms in which to manufacture the chips. While this has been very successful, it is in some sense a wasteful approach because the volume of the clean room is far larger than the wafers that are being processed. There are approaches to minimizing the actual volume of “clean room” which must be maintained. These approaches keep the wafers in sealed minienvironments in which they are transported from one machine or process to another. This approach was proposed many years ago and has been implemented by providing sealed cassettes to transport the wafers from machine to machine. This approach is becoming popular today and some newer factories employ these SMIF (standard mechanical interface) boxes extensively to provide a cleaner environment for the wafers as shown in Figure 4–28. This alternative approach to “mini-clean rooms” may be more common in the future than it is today [4.1]. Economics may play a significant role in determining how extensively such minienvironments are used in the future, since the costs associated with constructing and operating large clean rooms are very high.

Wafer cleaning has remained largely unchanged for the past 25 years, because of the success of the RCA clean. Yet there are reasons to investigate improved methods. The RCA cleaning method uses large quantities of chemicals and DI water at high

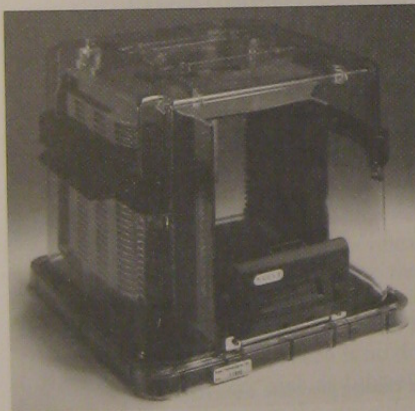


Figure 4–28 Example of a SMIF box or pod for transporting wafers in an IC manufacturing facility. These carriers connect directly to manufacturing machines so that the wafers are not exposed to room air. This figure was taken directly from the Asyst Technologies Corp. web site at <http://www.asyst.com/>.

temperatures. This generates large amounts of chemical vapors that must be exhausted from the clean rooms and then treated. The chemicals themselves must also be treated before disposal. These environmental issues are becoming increasingly important as the scale of IC manufacturing plants increases. One simple change which is being used by some wafer manufacturers is to simply use SC-1 and SC-2 solutions that are more dilute. These have been found to be as effective as the original concentrations and they reduce the environmental issues associated with chemical disposal. In addition, H_2O_2 is unstable at the temperatures used in the SC-1 and SC-2 solutions, especially in acidic solutions, decomposing into H_2O and O_2 . As a result, the composition of the solutions can change with time, resulting in less effective cleaning. In the SC-1 solution, if the concentration of H_2O_2 drops too much, the NH_4OH will attack silicon.

A number of replacement cleaning procedures have been proposed [4.35]. A particularly interesting procedure is that proposed by Ohmi [4.28]. In Table 4-3 the only oxidant stronger than H_2O_2 is O_3 . Ohmi's process uses ozonized ultrapure water as shown in Figure 4-29. This process is all at room temperature, uses far fewer chemicals than the standard RCA clean, and is claimed to be as effective in cleaning wafers. The principles behind Ohmi's cleaning procedure are similar to the RCA clean. A strong oxidant is used to oxidize organics and metals. Megasonic agitation is used to loosen particles that may have adhered to the wafers. The final steps in Ohmi's process involve an HF etch of the surface. This strips away the chemical oxide that naturally grows as part of the

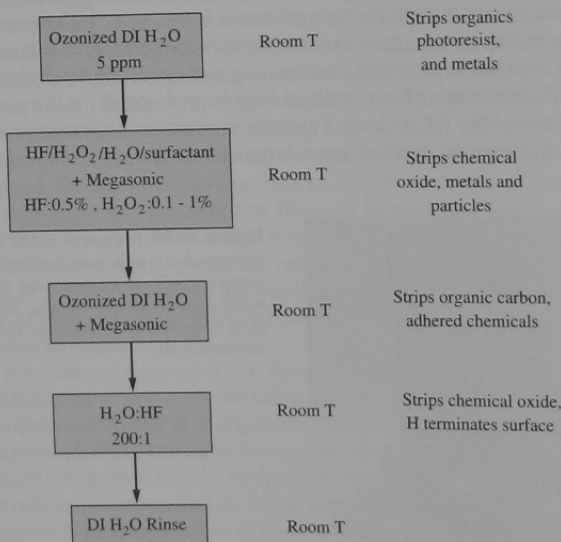


Figure 4-29 Possible room temperature cleaning procedure replacement for the RCA clean [4.28].

cleaning procedure. The silicon surface atoms are left with dangling (incomplete bonds) which are highly reactive and which form hydrogen bonds with H atoms from the HF solution [4.36]. The hydrogen-passivated surface that results is ready for immediate further processing, such as epitaxial layer growth. There are some concerns with using a final HF step for wafers whose next processing step is oxidation. The hydrogen passivated Si (100) surface is roughened by the H passivation process and this may lead to poorer quality thin oxides grown on these surfaces [4.37]. In addition, the hydrogen passivated surface is very sensitive to ambient effects between the time the cleaning ends and the next process step begins, so special care may have to be taken in transporting and handling such wafers.

There has also been considerable recent work on “dry” or vapor phase cleaning procedures [4.5]. This work has been motivated both by the environmental issues associated with wet chemistries and by the increasing use of cluster tools in IC manufacturing. In these tools, multiple-process steps, often on single wafers, are performed in a single machine by moving wafers from one chamber to another. There is great interest in accomplishing the cleaning process in one of these chambers because the wafers would then move directly to the process step under well-controlled ambient conditions, minimizing contamination between cleaning and the next step. “Dry” cleaning methods which are under investigation include the use of gas vapors, plasmas, low-energy physical processes like sputtering, and photochemically enhanced cleaning in which UV energy is used to help break surface bonds and release contaminants on the wafer surface. None of these techniques is currently in widespread use in manufacturing.

Gettering methods have remained relatively unchanged for the past decade. The scientific knowledge providing our understanding of how gettering actually works has certainly improved in recent years, primarily because of improved understanding of point defect properties in silicon. It is likely that intrinsic gettering will dominate in the future because with smaller devices, thermal budgets for IC processing will have to decrease. The closeness of the gettering traps to the active devices in intrinsic gettering gives this process a distinct advantage over extrinsic gettering. It is likely that tighter controls on oxygen concentration in CZ wafers will be required in the future, because this is an essential element of the intrinsic gettering process and is also required in order to control warpage and the mechanical strength of the wafers. Better understanding is also needed of the role of carbon in the oxygen precipitation process. The SIA NTRS [4.1] lists two primary future needs in the gettering area. The first of these emphasizes low-temperature gettering, for the reasons just described. Intrinsic gettering is the best hope at this point, but alternatives may be needed for very low thermal budget processes. The second priority is a high-quality simulation tool for gettering. Such a tool will have to be based on the ideas we have discussed in this chapter—the role of point defects and the three gettering steps of release, diffusion, and trapping.

Some workers believe that gettering will be unnecessary in the future, because the level of cleanliness in factories and wafer cleaning procedures will be sufficient to keep metal atom concentrations to levels that will not affect device performance and chip yield. This approach is no doubt possible but the cost of such manufacturing practices may be very high and the alternative of providing some safety margin through gettering will remain the cost-effective solution in many cases.

4.7 Summary of Key Ideas

In this chapter, we have considered some of the “nuts and bolts” of IC manufacturing. Particle control, wafer cleaning, and gettinger sometimes do not receive the attention in the scientific community that more “high-tech” processes such as ion implantation, etching, and lithography do. However, the topics we have discussed in this chapter are crucial for the successful manufacturing of complex chips. In fact it is likely the case that the economic success of any company in this business depends as much on its attention to particles, cleaning, and gettinger as it does on the design features of a particular product. High yields, which come from attention to these issues, are the mark of successful manufacturers.

We have seen that a three-tiered approach is used to minimize contaminants in silicon wafer processing. The first step is to clean the environment around the wafers. Most commonly this is done through advanced air filtration systems to remove particles, careful protocols for workers in the plant, and the use of highly purified, filtered chemicals and gases. The second tier of contamination control is careful wafer cleaning. This is probably the most common step performed in wafer manufacturing. Cleaning needs to remove particulates that adhere to wafer surfaces, and it needs to strip trace levels of organic and metal contaminants. The RCA clean, which is most commonly used, is based on a solid scientific understanding of the processes occurring in cleaning. Finally, gettinger is the third level of contamination control. This process collects metal atoms in regions of the wafer away from the active devices. Gettering occurs by releasing the metal atoms from whatever their current position is, allowing them to diffuse to the gettinger site, and then trapping them at those sites. While both extrinsic and intrinsic gettinger are used in wafer manufacturing today, it is likely that intrinsic gettinger will be the most common in the future because the gettinger sites are physically closer to the devices, minimizing the thermal cycling needed for gettinger.

The bottom line is yield. A difference of a few % in good die from each wafer can make a tremendous difference in the profitability of a company. Because of the batch fabrication processes used in IC manufacturing, with hundreds of die per wafer, the bad die are inherently manufactured along with the good die. There is generally no additional manufacturing cost when the yield goes up, so the incremental good die are essentially pure profit for the manufacturer. This fact should provide strong motivation for careful attention to the topics discussed in this chapter.

4.8 References

- [4.1]. “National Technology Roadmap for Semiconductors,” SIA, 1997.
- [4.2]. F. Schwettman, K. Chiang, and W. Brown, “Variation of Silicon Dioxide Growth Rate with Pre-Oxidation Clean,” *Electrochem. Soc. Abstracts*, vol. 78-1, p. 689, 1978.
- [4.3]. S. M. Hu, private communication.
- [4.4]. W. Kern and D. A. Puotinen, “Cleaning Solutions Based on Hydrogen Peroxide for Use in Silicon Semiconductor Technology,” *RCA Review*, vol. 31, p. 187, 1970.
- [4.5]. W. Kern, editor, *Handbook of Semiconductor Wafer Cleaning Technology*, Noyes Publications, 1993.

- [4.6]. C. W. Pearce, J. L. Moore, and F. A. Stevie, "Removal of Alkaline Impurities in a Polysilicon Gate Structure by Phosphorus Diffusion," *J. Electrochem. Soc.*, vol. 140, p. 1409, 1993.
- [4.7]. P. Balk and J. M. Eldridge, "Phosphosilicate Glass Stabilization of FET Devices," *Proc. IEEE*, vol. 57, p. 1558, 1969.
- [4.8]. G. Masetti and M. Severi, "Dependence of Flat-band Voltage of MOS Structures on Phosphosilicate-glass Growing Conditions," *Appl. Phys. Lett.*, vol. 37, p. 226, 1980.
- [4.9]. E. Snow and B. E. Deal, "Polarization Effects in Insulating Films on Silicon—a Review," *Trans. Metall. Soc. of AIME*, vol. 242, p. 512, 1966.
- [4.10]. E. R. Weber, "Transition Metals in Silicon," *Appl. Phys.*, vol. A30, p. 1, 1983.
- [4.11]. W. E. Beadle, J. C. C. Tsai and R. D. Plummer, *Quick Reference Manual for Silicon Integrated Circuit Technologies*, John Wiley & Sons, 1985.
- [4.12]. J. C. Mikkelsen Jr., "The Diffusivity and Solubility of Oxygen in Silicon," *Mat. Res. Symp. Proc.*, vol. 59, p. 19, 1986.
- [4.13]. H. W. Kennel, "Physical Modeling of Oxygen Precipitation, Defect Formation and Diffusion in Silicon," Ph.D. Thesis, Stanford University, TR. No. G710-2, Feb. 1991.
- [4.14]. D. Huber and J. Reffle, "Precipitation Process Design for Denuded Zone Formation in CZ-Silicon Wafers," *Solid State Tech.*, p. 137, Aug., 1983.
- [4.15]. M. Buehler, "Microelectronic Test Chips for VLSI Electronics," in *VLSI Electronics, Microstructure Science*, vol. 6, p. 529, Academic Press, 1983.
- [4.16]. W. Lukaszek, K. G. Grambow, and W. J. Yarbrough, "Test Chip Based Approach to Automatic Diagnosis of CMOS Yield Problems," *IEEE Trans. on Semi. Manufac.*, vol. 3, 1990.
- [4.17]. C. Weber, "Generic Test Chip Formats for ASIC-Oriented Semiconductor Process Development," *Proc. IEEE Int. Conf. on Microelectronic Test Structures*, p. 247, March 1993.
- [4.18]. D. K. Schroder, *Semiconductor Material and Device Characterization*, John Wiley & Sons, 1990.
- [4.19]. T. Tiedje, J. I. Haberman, R. W. Francis, and A. K. Ghosh, "An RF Bridge Technique for Contactless Measurement of the Carrier Lifetime in Silicon Wafers," *J. Appl. Phys.*, vol. 54, p. 2499, 1983.
- [4.20]. Y. Mada, "A Nondestructive Method for Measuring the Spatial Distribution on Minority Carrier Lifetime in Si Wafer," *Jpn. J. Appl. Phys.*, vol. 18, p. 2171, 1979.
- [4.21]. D. K. Schroder and J. Guldberg, "Interpretation of Surface and Bulk Effects Using the Pulsed MIS Capacitor," *Solid State Electronics*, vol. 14, p. 1285, 1971.
- [4.22]. Y. Mizokami, "The Total CIM System for Semiconductor Manufacturing," *Proc. 1990 IEEE International Semiconductor Manufacturing Science Symposium*, p. 24, 1990.
- [4.23]. J. McGehee, J. Hebley, and J. Mahaffey, "The MMST Computer-Integrated Manufacturing Framework," *IEEE Trans. Semi. Manufac.*, vol. 7, p. 107, 1994.
- [4.24]. C. H. Stapper, "Fact and Fiction in Yield Modeling," *Microelectronics Journal*, vol. 20, 1989.
- [4.25]. D. Dance and K. Gildersleeve, "Estimating Semiconductor Yield From Equipment particle Measurements," *1992 IEEE/SEMI International Semiconductor Manufacturing Science Symposium*.
- [4.26]. G. Cheek and G. O'Donoghue, "Yield Models in a Design for Manufacturability Environment: A Bibliography," *1993 IEEE/SEMI International Semiconductor Manufacturing Science Symposium*.

- [4.27]. J.W. Hill and R.H. Petrucci, *General Chemistry, An Integrated Approach*, Prentice Hall, 1999.
- [4.28]. T. Ohmi, "Total Room Temperature Wet Cleaning of Silicon Surfaces," *Semiconductor International*, p. 323, July 1996.
- [4.29]. F. C. Frank and D. Turnbull, "Mechanism of Diffusion of Copper in Germanium," *Phys. Rev.*, vol. 104, p. 617, 1956.
- [4.30]. G. B. Bronner and J. D. Plummer, "Gettering of Gold in Silicon: a Tool for Understanding the Properties of Silicon Interstitials," *J. Appl. Phys.*, vol. 61, p. 5286, 1987.
- [4.31]. H. Higuchi and S. Nakamura, "Gettering of Electrically Active Gold in Silicon," *Extended Abstracts of the Electrochem. Soc. Meeting, Spring 1975*, vol. 75-1, p. 412, 1975.
- [4.32]. L. Baldi, G. F. Cerofolini, G. Ferla, and G. Frigerio, "Gold Solubility in Silicon and Gettering by Phosphorus," *Phys. Status Solidi A*, vol. 48, p. 523, 1978.
- [4.33]. J. S. Kang and D. K. Schroder, "Gettering in Silicon," *J. Appl. Phys.*, vol. 65, p. 2974, 1989.
- [4.34]. R. L. Meek and T. E. Seidel, "Enhanced Solubility and Ion Pairing of Cu and Au in Heavily Doped Silicon at high Temperatures," *J. Phys. Chem. Solids*, vol. 36, p. 731, 1975.
- [4.35]. W. A. Cady and M. Varadarajan, "RCA Clean Replacement," *J. Electrochem. Soc.*, vol. 143, p. 2064, 1996.
- [4.36]. Y. J. Chabal, G. S. Higashi and K. Raghavachari, "Infrared Spectroscopy of Si (111) and Si (100) Surfaces After HF Treatment: Hydrogen Termination and Surface Morphology," *J. Vac. Science and Tech. a*, vol. 7, p. 2104, 1989.
- [4.37]. M. Offenbergh, M. Liehr, and G. W. Rubloff, "Ultraclean Integrated Processing of Thermal Silicon (111) and (100) Surfaces," *Appl. Phys. Lett.*, vol. 57, p. 1254, 1990.

4.9 Problems

- 4.1. An IC manufacturing plant produces 1000 wafers per week. Assume that each wafer contains 100 die, each of which can be sold for \$50 if it works. The yield on these chips is currently running at 50%. If the yield can be increased, the incremental income is almost pure profit because all 100 chips on each wafer are manufactured whether they work or not. How much would the yield have to be increased to produce an annual profit increase of \$10,000,000?
- 4.2. Calculate and plot the total number of allowable LLS (Localized Light Scatterers) on an 8" wafer, through 2012, according to the NTRS. Generate a similar plot of the total number of allowed gate oxide defects in DRAMs. What conclusions do you draw about these requirements?
- 4.3. As MOS devices are scaled to smaller dimensions, gate oxides must be reduced in thickness.
 - a. As the gate oxide thickness decreases, do MOS devices become more or less sensitive to sodium contamination? Explain.
 - b. As the gate oxide thickness decreases, what must be done to the substrate doping (or alternatively the channel V_{TH} implant, to maintain the same V_{TH} ? Explain.

- 4.4. A new cleaning procedure has been proposed which is based on H_2O saturated with O_2 as an oxidant. This has been suggested as a replacement for the H_2O_2 oxidizing solution used in the RCA clean. Suppose a Si wafer, contaminated with trace amounts of Au, Fe, and Cu is cleaned in the new $\text{H}_2\text{O}/\text{O}_2$ solution. Will this clean the wafer effectively? Why or why not? Explain.
- 4.5. Explain why it is important that the generation lifetime measurement illustrated in Figure 4-19 is done in the dark.
- 4.6. If you wanted to look for trace levels of contaminants left on a wafer surface after cleaning, would XRF or XPS be a better choice? Explain.
- 4.7. In the discussion in the text related to Figure 4-12 the importance of the temperature ramp rate between the nucleation and precipitation stages is described. Is the ramp rate coming back down to room temperature after the precipitation stage important in terms of gettering? Explain.