Chapter 9
Defect Engineering in Silicon Materials

Werner Bergholz

Abstract Point Defects (i.e. vacancies and selfinterstials are zero-dimensional defects) in silicon. In solids in general, such defects are unavoidable, their presence follows from the Second Law of Thermodynamics; for a given temperature and for each species of intrinsic point defects (vacancy, self-interstitial). In contrast, impurities for which there is an equilibrium solubility, can be avoided by suitable countermeasures. Likewise, one-dimensional defects (dislocations), two-dimensional defects (e.g. stacking faults) and three-dimensional defects (e.g. voids or precipitates of impurities) can be avoided by defect engineering; they are not inevitable by the laws of thermodynamics. To turn silicon into practical devices (e.g. integrated circuits, photovoltaic cells), defect engineering is part and parcel of the overall process and production technology. First, the fundamentals will be explained, e.g. the analogy between the chemistry of ions in water as a "substrate" and the "chemistry" of point defects in silicon, the most perfect, purest solid material available. Secondly, the application of these fundamental defect engineering principles will be described for various technology areas. The focus will be on microelectronics, photovoltaics will also be mentioned. Topics in microelectronics will include the impact of oxidation, ion implantation, reactive ion etching and thermal processes for diffusion of dopant atoms. Gettering of metal contamination will also be considered. It will be elucidated that there is a close interaction of metal impurities and extended defects, and that an uncontrolled metal contamination is a source of unacceptable quality risks in the production of microelectronic devices and photovoltaic cells.

Keywords Point defects • Dislocations • Stacking faults • Metal contamination • Device failure • Quality engineering • Key control characteristic

---

W. Bergholz (✉)
Jacobs University Bremen, Bremen, Germany
e-mail: w.bergholz@jacobs-university.de

© Springer Japan 2015

Y. Yoshida, G. Langouche (eds.), *Defects and Impurities in Silicon Materials*,
Lecture Notes in Physics 916, DOI 10.1007/978-4-431-55800-2_9

9.1 Introduction

In this section the scope and the objectives of defect engineering, i.e. the application of this knowledge in the semiconductor industry, will be outlined against the backdrop of the remarkable perfection of electronic grade silicon material for the microelectronics industry. Also, the transfer of the material knowledge gained in the microelectronics industry to the photovoltaics (PV) industry which in terms of silicon consumption has overtaken microelectronics will be briefly dealt with in this chapter:

Before going into any detail, the way how a good connection between science (in the preceding chapters) and application has to be explained: The term "Engineering" implies that in this chapter the scope has to be shifted from an emphasis on science to generate new knowledge about materials to how to gain new insights with regards to the application of silicon materials science to a corporate semiconductor technology R&D environment or in a medium and large scale production environment. This has two implications:

(a) In terms of the analytical methods many of the scientifically oriented state-of-the-art detection and analytical methods have to be replaced by high throughput/short cycle time methods.
(b) In R&D it is possible to create clearcut "boundary conditions" for experiments and systematic studies. In a production environment one is faced with complexity imposed by the production process.

Also, new concepts beyond Si materials have to be included, such as Statistical Process Control (SPC) of semiconductor manufacturing processes, with implications. Finally, the description of complex "real-word defects in semiconductor and photovoltaic production scenarios" necessitates the use of engineering concepts which are less rigid than "pure" science concepts but enable the engineer to understand the rationale behind practical defect engineering and enable him/her to apply such concepts to new technologies and unprecedented defect formation mechanisms.

Important sources for this chapter are books and key papers on silicon and semiconductor technologies, training materials from internal training for engineers in Siemens Semiconductors and last but not least real and generally important defect scenarios that have occurred in 30 years of microelectronics and photovoltaics production. In the following we strive to keep a good balance between academic rigor and a pragmatic approach needed to cope with the complexity of semiconductor processes. We are led by the objective that the reader will be enabled to monitor and analyze defect formation phenomena to identify the root causes of defect formation and develop countermeasures by a profound understanding of the defect scenarios.

It is not an exaggeration to state that microelectronics-grade silicon is the most perfect and clean material ever made [1, 2]. The mainstream production of integrated circuits, such as central processing units (CPUs), dynamic random access memories (DRAMs), microprocessors, chip cards etc. is carried out on single-crystal silicon

wafers of 150–300 mm diameter, with 450 mm diameter wafers at the R&D and sooner or later at the pilot production stage [3]. Advanced microelectronic devices would simply not work on Si wafers which are not a single-crystal and dislocation free [4], or which do not have the same mechanical (macroscopic shape) and structural (microscopic) perfection as wafers for integrated circuit production, or which are not as pure in volume and clean at the surface. This is not true for the photovoltaics industry, where relatively impure and unpolished multicrystalline silicon has the largest share in the production of PV cells and modules so far.

As the first step to a profound understanding of defect engineering, it is instructive to briefly review some of the most remarkable and relevant properties of silicon wafers [2]:

1. They are single crystals, the wafer surface is parallel to one of the {100} planes, a flat or a notch as the fiducial mark for alignment of the wafers indicates one of the <110> directions, as shown in Fig. 9.1a and b (recently, notchless wafers with a laser mark to indicate the crystal direction have been under discussion in the SEMI standards development organization, also see [5]). A single crystal means, that the atomic rows and planes are perfect from wafer surface to wafer surface in any direction, see also Fig. 9.3. For the sake of completeness it has to be mentioned that some power devices are also made on {111} oriented wafers.
2. The deviation of the local flatness of the wafer corresponds to about the thickness of a human hair (80 μm) on a football pitch [2]. Also the other dimensions, such as thickness, diameter, edge profile and the shape and size of the notch are well defined by tight geometric specifications.
3. The purity of the material with respect to metal contamination in the bulk is by far a smaller fraction than one person of the whole world population. The most abundant impurities are carbon and oxygen, and appropriate dopant atoms, which

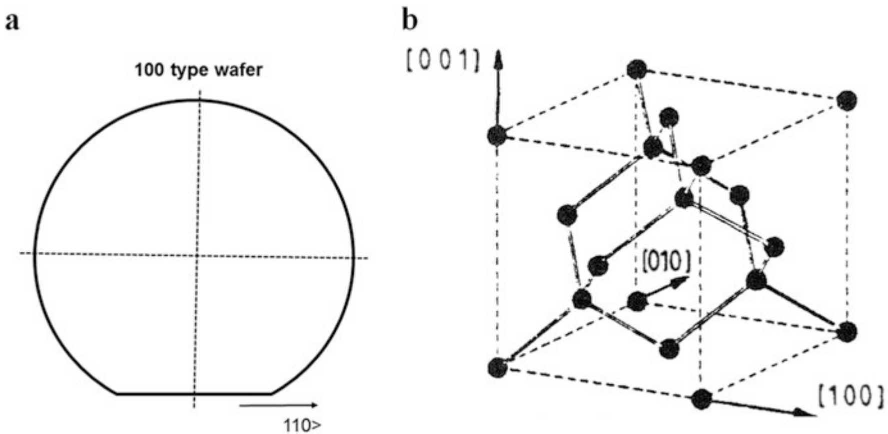

Fig. 9.1 (**a**) Sketch of a Si wafer with 100 surfaces. (**b**) Sketch of an elementary cell of the silicon lattice

are incorporated in well-defined and specified concentrations during the crystal growth process (to engineer the resilience of the wafer to thermal stress and to enable well defined intrinsic gettering, as will be described later). Although the concentrations of oxygen and carbon are only of the order of parts per million, they are critical for the robustness of the wafers against stress and for defect engineering, as will be described later in detail [7, 8].

A good source for the actual state of the art for manufacturing silicon wafers for the microelectronics industry can be found in the SEMI-standard M1 [2] and information about the future technical requirements are described in the International Roadmap for Semiconductors ITRS [3].

4. As briefly alluded to before, in the PV industry the wafers are by far not as clean and as perfect as Si wafers for microelectronics. The raw wafers used to produce cells are just sawn from a silicon ingot or brick, they are not edge rounded, and they are as thin as compatible with processes, currently around 150 μm, i.e. twice the thickness of a human hair (wafers for microelectronics are typically 625–925 μm) [9]. The reason for this drastically reduced technical sophistication is that the cost of the silicon wafers must be as low as possible, because ultimately the cost to produce electric energy by means of silicon PV cells must be low enough to compete with the standard ways to produce electric energy by burning fossil fuels or by using nuclear power [10]. Since the purification of silicon material is expensive (even under reduced purity requirements), the wafers are made "paper – thin" to use as little of the expensive purified material as possible and to avoid any technical specification which is not really needed for the performance and durability of the end product. All the same, also a PV wafer has to comply with the relevant technical specifications, which however as mentioned, are much less demanding.

So, the demanding properties according to the technical specification for micro-electronics silicon wafers are necessary in most instances to warrant the full functionality and reliability/durability of the integrated circuits produced, in contrast to the situation in the PV industry, where the technical properties are much less demanding.

To apply the knowledge presented in the preceding chapters of this book to practical problems in technology development and to mass manufacturing, i.e. in defect engineering, needs the introduction of additional concepts which are rooted in general process and technology management and presents the challenge to apply the silicon science and technology presented so far in a comprehensive and concatenated manner, with many potential interactions of different phenomena.

So, the purpose of defect engineering is to establish and maintain the properties (1)–(4) as well as possible throughout the complete integrated circuit or PV-cell manufacturing processes, including the procurement of the wafers, since defects can be already "built-in" in the raw silicon wafer before any device process, likewise the robustness of the wafers to cope with e.g. mechanical stresses can be determined already by the silicon wafer manufacturing process as will be explained in later

sections. Such defect engineering is indispensable in order to prevent the detrimental consequences of defects (such as leakage currents in pn-junctions, gate oxide shorts, drift of transistor parameters etc. for microelectronics and below par electrical conversion efficiency for PV cells), and other problems in the performance of the reliability/durability of these products [11–15]. In spite of the much less stringent requirements for PV silicon wafers, defects do play a role to detract the PV cell performance, and it can be stated that improvements of the conversion efficiency of PV cells that have been achieved over the last decades are both due to improvements in the manufacturing process stability in terms of reproducibly achieving the critical device parameters, improvement in the design and last but not least have been achieved by reducing defect densities. More detailed information for these facts can be found in Refs. [16, 17].

Remarkably, in addition to the prevention of defects in a number of instances defect engineering serves to intentionally introduce defects in a controlled fashion to support or impart certain device functions.

Thus the scope of defect engineering is the prevention of those defects which harm the integrated circuits and to create defects which support or enhance the functionality and/or reliability/durability of the device.

This chapter will mainly focus on defects that are relevant for device processing, it will not cover in detail defect formation during the manufacture of silicon ingots or wafers at the suppliers of silicon wafers, this is the subject of other chapters of this book.

In microelectronics, defect engineering includes both defects in the silicon material and defects in the technology layers above the silicon wafer (such as metal lines). The latter are mainly, but not exclusively due to particles that cannot be avoided entirely in spite of a cleanroom [18], and in spite of equipment that is low in particle shedding and materials which are ultraclean, both with respect to particles and impurities.

In this chapter we will only deal with the defects in the silicon wafer, not the defects which are defects in the structure of the technology layers (such as short or open circuits in metal lines in one of the metallization layers above the silicon).

9.2 What Does Defect Engineering Mean in Detail?

As mentioned in the introduction to this chapter, defect engineering means that the defect formation has to be prevented and/or promoted in a “controlled” fashion, in an industrial manufacturing environment. It is one of the desired learning outcomes of this chapter that the difference in “boundary conditions” between a research laboratory situation and an R&D pilot production line and/or a large volume manufacturing is clearly understood, with respect to the defect formation mechanisms and the control of defects, which includes assessment of the individual unit process performance and the integrated process performance in terms of defect types and densities. It will turn out that often, defect formation/prevention is not a

question of optimizing a unit process but to understand and modify the integrated process so as to remove undesired detrimental interactions between processes.

To be able to implement state-of-the art defect engineering, the following three areas of competence are required:

(a) Scientific understanding and modelling of defect formation and prevention. The preceding chapters of this book have been instrumental to this purpose. Very often, defects are avoided by trial and error through variation of the process parameters. Time and again, experience has shown, that without a profound understanding of the defect formation mechanism, this is neither a safe nor an efficient method, and recurrence of defects when some other process parameters are changed, is frequently observed in such situations, due to interactions between different process steps that have been neglected or not understood. A significant part of this chapter will be dedicated to teaching by the introduction of concepts and examples how such interactions can be detected or even anticipated at the stage of process/equipment design.

(b) In manufacturing, there is a strict requirement for effective (i.e. high success rate, high yield and good quality [i.e. within the specified defect density targets]) and efficient production (achieve a high throughput with i.e. with as low cost for consumables [such as Si wafers and other materials] and equipment as possible). In the context of defect engineering this means that the methods to monitor the defect densities must be fast and cover a large enough area of the processed wafers not to interfere too much with the speed and sequences of the production process (i.e. the cycle time targets for the unit processes and time coupling requirements which exist between some processes should not be jeopardized) [15], and to cover a representative number of samples. Also, they must satisfy the needs of statistical process control (SPC), which is a pivotal quality management tool that is indispensable in microelectronics to ensure that processes are predictably under control [16, 19]. SPC is also becoming more and more relevant in the photovoltaics industry (although initially this was not the case see [16]). Therefore, the defect monitoring methods in manufacturing are frequently different from sophisticated but low throughput, time-consuming analytical methods in a science laboratory, which would be more frequently used to understand and model defect formation in the context of research. Figure 9.2 illustrates one "production-oriented" method to detect, e.g., contamination by Co, Ni and Cu, and with limitations also Fe (these are ubiquitous in a technical environment and extremely detrimental to devices) on a large wafer area in a short time [6, 15, 18, 20]. The method is called haze test, it consists of driving in any surface impurities by e.g. a 30s rapid anneal step at 1200 °C and delineating the defects that have formed in the process by a suitable defect etch, which converts invisible surface near precipitates of Ni and Cu (and other extended defects) into etch pits which can be visualized on the whole wafer in one photograph by dark field illumination, see also [6]. With this method it is possible to test a process (mechanical transport as in the illustrated example, or chemical or dry etch processes) in 15 min over the surface of a

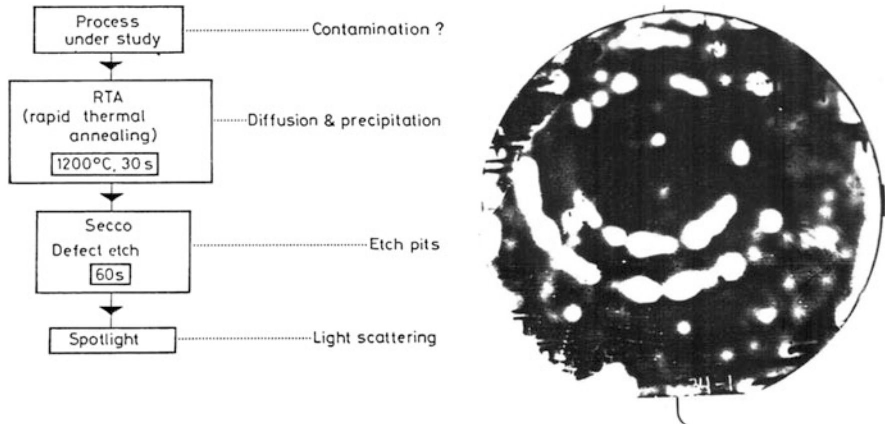

Fig. 9.2 Illustration of the haze test for fast diffusing metal impurities. *Left hand side*: Process steps for the rapid haze test to detect any Co, Ni or Co impurities (and with limitations, Fe) on or in the wafer (total duration as short as 15 min, this preparation step can also be used if the detection is electrically, e.g. by microwave PCD method mentioned in Chap. 3); *right hand side*: result of the test: Ni contaminated areas are seen as bright spots (light scattering from the etch pits), the sensitivity of the method better than 10^{11} cm^{-3}. Note that in this example the original contamination was on the wafer back surface, the haze picture has been taken of the front side. That implies that in 30s the Ni diffused right through the thickness of the silicon within 30s (!) and also spread laterally over several mm, as indicated by the circular bright areas (The wafer is a 100 mm diameter wafer) [6]

complete wafer. So the method is a large area fast detection method with the additional benefit that it shows up the contamination pattern on the wafer, which often is a direct indicator of the contamination source. If parallel processing of several test wafers is done, the time per wafer is reduced to below 5 min. Other methods for fast large-area detection of metal impurities and other defects exist and examples will be given at appropriate places in this chapter. Also, in a later section, an example for the fast large-area electrical detection of these impurities will be presented, the method is equivalent in sensitivity to the DLTS method presented in Chap. 3, and less specific, but has become the mainstream method for the detection of metal contamination both in microelectronics and PV technology. As will be explained in detail in Sect. 9.5 of this chapter, the prevention of metal contamination is one of the key factors for the suppression of the formation of extended defects. Already at this stage it is pointed out that for Fe, Co and Mn an interesting scientific method exists, namely Mössbauer spectroscopy. The method has the unique advantage that it can detect all Fe (in absorber experiments) and all Co/Mn (in source experiments, see Chap. 8 for details regarding source and absorber experiments), and at the same time it has spectroscopic properties, namely that different species (isolated impurities on different sites in the crystal, agglomerates, precipitates) can be separated. The method and its applications and the additional information

obtained and resulting open questions are explained in Chap. 8 and have contributed significantly to the understanding of the behavior of transition metals in silicon, although it has to be conceded that to date no overall consistent model to explain all empirical observations has been arrived at so far.

(c) A profound knowledge of the critical parameters in unit processes which are relevant for defect formation or prevention, since as pointed out earlier a controlled defect density is a pre-requisite for consistent performance and reliability/durability of the final products [4, 12–15]. Obviously, the profound knowledge mentioned in a. is a pre-requisite for such endeavors. In the microelectronics industry, the automotive quality management system standard ISO TS 16949 (see http://www.iso.org/iso/catalogue_detail?csnumber=52844) is a "must" (this is because a significant part of microelectronic products is for automotive applications), and one requirement of the standard is to define and monitor parameters critical to quality. By standard, such a parameter is called "special characteristic". This is why these parameters are also called "key control characteristics" (KCCs). Since defects are obviously critical to quality (in the case of the defects introduced on purpose their absence can also affect the performance and/or the reliability/durability negatively!), all defect densities and/or the associated process parameters are by definition KCCs and should ideally be monitored by SPC.

In the following chapters all three aspects (a)–(c) will repeatedly be mentioned where appropriate for a clear logical connection of this abstract concept with manufacturing reality and to create a more profound understanding of how quality cannot tested into the end product but "manufactured into the product", so the manufacturing process must predictably create this quality. This is the basic philosophy behind defect engineering, as a foundation it needs the knowledge presented in the preceding chapters.

9.3 Classification of Defects

The starting point in defect engineering is modelling and understanding defects, i.e. the aspect a. from Sect. 9.2. As the first step, the crystal structure of silicon and how the crystal structure is "embedded" in the external shape of silicon wafers used in microelectronics manufacturing has to be described.

Silicon crystallizes in the diamond lattice, i.e. the each Si atom is bonded to four nearest neighbors (compare Fig. 9.1b), the bonds are directional covalent bonds in a tetrahedral arrangement. A Si wafer is a single crystal without any extended crystal defects, i.e. the elementary cell is reproduced faithfully throughout the wafer, if the lattice constant would be about 50 million times larger as it actually is, a <100> wafer would look like in Fig. 9.3.

As the next step towards a scientific and technical understanding for defects we introduce a commonly used classification of crystal defects. Crystal defects are

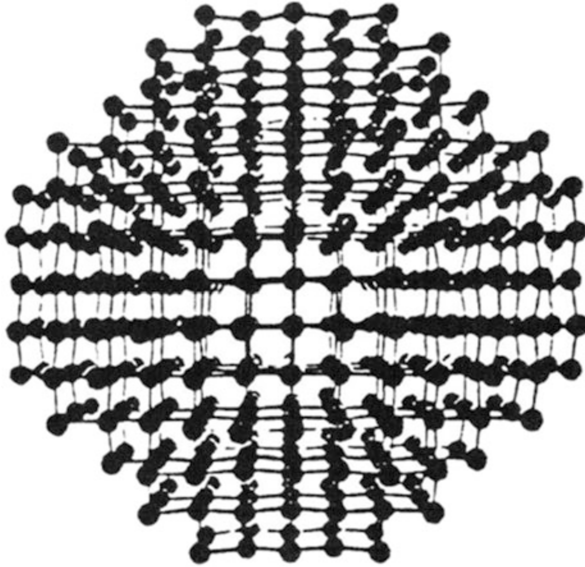

Fig. 9.3 View of a <100> wafer if the lattice constant would be 50 million times larger than it is in reality

any deviations from the ideal single crystal. The usual classification of defects is according to their dimensions.

9.3.1 Zero-Dimensional Defects

These defects are called point defects. They can be either the "wrong" atoms on a regular site of the crystal lattice, i.e. an impurity atom (larger or smaller than the host lattice atoms, with or without a different number of valence electrons), or they can be intrinsic defects. That means that there is either a missing silicon atom where there should be one or there is an extra Si atom not on a regular lattice site. The first type of defect is called a vacancy, the second one is termed self-interstitial, since an extra Si atom is "squeezed" in between regular lattice sites as there is no room on a regular lattice site. This is possible since the diamond lattice is a relatively open structure, compare Fig. 9.1b. Likewise, impurity atoms can be on such interstitial sites, and as will be explained in detail later, such impurity atoms can diffuse up to ten orders of magnitude faster than substitutional impurity atoms, the haze test explained above is an illustration of such fast diffusion.

As a rule, these defects introduce mechanical distortions to the crystal lattice in the immediate neighborhood as indicated in the 2D sketch of Fig. 9.4. It is to be noted that point defects can also form pairs or clusters of more than two point

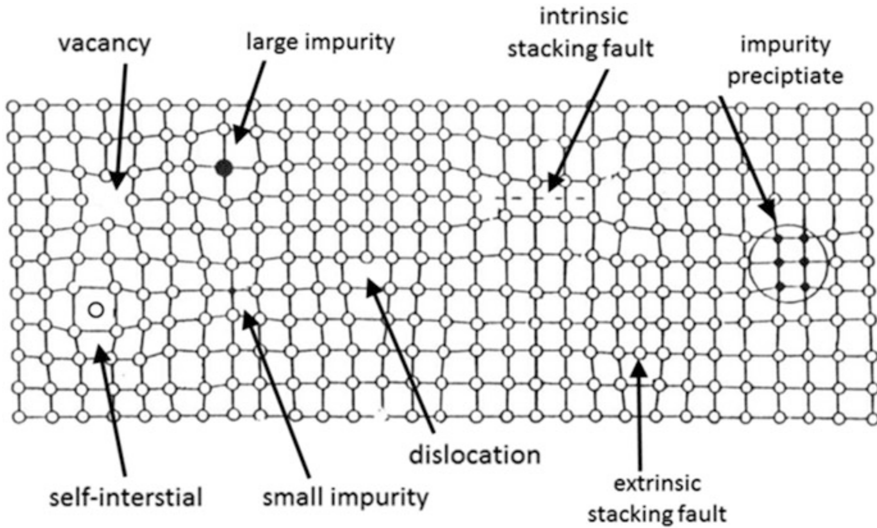

Fig. 9.4 Sketch of 0-D, 1-D, 2-D and 3-D crystal defects as a projection into a plane, using a *square* lattice

defects, consisting of equal or different point defects. This is also the case in metals, because the additional energy spent on the mechanical distortion can be lowered that way.

In silicon, there is second way to lower the energy of the system, which we pragmatically call a second "driving force" for pairing and/or clustering: In a semiconductor, the point defects can be neutral or singly or multiply charged, so there is an additional degree of freedom for an energy lowering of the system by coulomb interaction between the defects or hybridization of the electronic wave functions of the defects. Furthermore, it has to be realized that depending on the position of the Fermi level and on the donor/acceptor levels in the band gap, more than one charge state can exist at the same time, for a detailed presentation the reader is referred to the book by Tuck [21] or [18, 22]. The consequences of this fact will be followed up below and in a later section.

The most important and striking property of intrinsic point defects (vacancies, interstitials) is that they cannot be avoided. Point defects have a certain thermal equilibrium concentration which depends on the temperature following an Arrhenius (i.e. exponential) law. The reason for their existence is founded in the second law of thermodynamics since the free energy of a solid can be lowered by the entropy term -TS, and point defects can be generated in principle at any site of the solid, leading to a large entropy. Furthermore, the solubility can in addition depend on the doping level, due to the fact that other than neutral charge states can exist in addition to the neutral impurity or point defect, since this is an extra thermodynamic species, it adds to the solubility [21, 22]. This is a very significant factor for manufacturing of electronic devices and photovoltaic cells, especially in connection with gettering

of unwanted metal impurities (i.e. the removal of detrimental metal impurities from active device regions to place in the wafer where they are not harmful).

Intrinsic point defects are always present in thermal equilibrium. Under normal circumstances the concentration of intrinsic point defects is so low that their effects on the product are negligible. However, under certain process conditions, vacancies and/or interstitial can agglomerate to form clusters or larger extended defects, and then they can become very detrimental to the device and their formation has to be prevented. If there is enough time given for a judicious thermal equilibration program, the agglomeration of intrinsic defects to extended defects can, in principle, be prevented (see Chaps. 4 and 5, and [1]). Also, the presence of impurities can be suppressed to a large degree by suitable counter measures, i.e. ultra clean materials and ultraclean processing, such strategies and technological countermeasures are essential constituents of the defect engineering toolbox [11–15].

So, remarkably, even a perfect silicon single crystal MUST contain a certain number of vacancies and self-interstitials at a given temperature, in thermal equilibrium. The thermal equilibrium concentrations depend strongly on temperature and other factors, such as the doping level. A point of particular importance is, as mentioned above, that point defect concentrations in excess of the thermal equilibrium concentration can coalesce into extended two-dimensional or three-dimensional defects, e.g., during crystal growth (compare Chaps. 4 and 5) or during any high temperature device processing steps, under some circumstances even at room temperatures, as detailed in later sections.

9.3.2 One-Dimensional Defects (1D Defects)

Contrary to point defects, the existence of extended 1D, 2D or 3D extended defects is not "mandated" by the principles of thermodynamics; they can be avoided altogether by suitable defect engineering strategies, i.e. the right process conditions and suitable precautions.

For the further understanding of geometry of extended defects, it is important to get some insight into the geometric structure of such extended defects in silicon, in particular to understand how extended defects are drawn often drawn in a 2D projected view in textbooks or scientific publications, whereas in reality they are three-dimensional objects in the silicon lattice.

The most important and frequent one-dimensional defects are dislocations, i.e. line – type defects. By definition, a dislocation is the edge of a crystal lattice plane that ends inside the crystal, as sketched for a 3D cubic lattice in Fig. 9.5.

The sketch in Fig. 9.5 is a simplified representation of a real geometry of dislocations in a diamond lattice. Figure 9.6 shows a ball and stick model of the most common dislocation type, the 60° dislocation. For most situations in defect engineering it is not necessary to consider the detailed geometry of a dislocation in the diamond lattice, it is however relevant to understand the orientation of the glide plane if slip in a wafer occurs, which is sketched in Fig. 9.7.

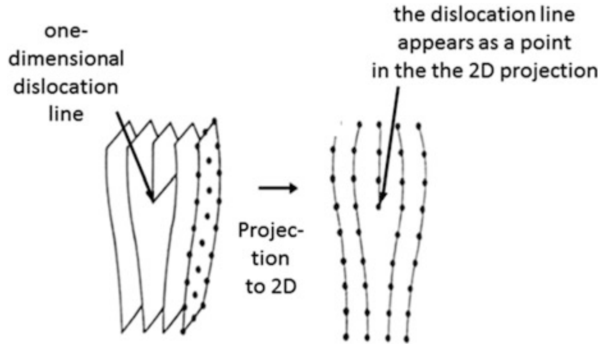

Fig. 9.5 Schematic drawing of a dislocation in 3D and how it is projected into 2D for easier drawing. The lattice planes are indicated as lines in the 2D projections, lattice points are only sketched for one of the planes in the 3D sketch. In the 2D projection the dislocation appears as a point, which is an "end-on" view of the dislocations. Note that dislocations can be and are as a rule curved rather than straight as in this illustration. Dislocations are also represented in Fig. 9.4

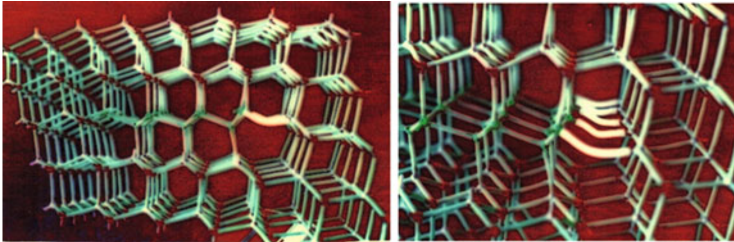

Fig. 9.6 Ball-and stick-model of a 60° dislocation in silicon. *Left* is an overview, *right* a close-up photograph of the core of the dislocation. The normal covalent bonds between the silicon atoms are shown in *green*, the "dangling" bonds at the end of the extra half plane are indicated by *white pipes* (which also point to the end of the extra half plane)

9.3.3 Two-Dimensional Defects

If an extra half plane is inserted into a crystal which nowhere intersects the surface of the wafer, as sketched in Fig. 9.8, (or if a plane is removed), a two-dimensional defect results. Such defects are called extrinsic or intrinsic stacking faults (SFs), respectively. It is obvious, that a stacking fault is bounded by a (partial) dislocation, which lies entirely inside the crystal. Both types of stacking faults are also represented in Fig. 9.4.

An important property of stacking faults and dislocations in connection with device processing is that they can absorb or emit intrinsic defects, so that they can grow or shrink (i.e. the dislocation around it "climbs", as opposed to glides if it is not an in-lattice-plane movement), so they can move into active device regions even if they were originally formed away from them. In addition, for defect engineering it is important to note that they can act as sources or sinks for intrinsic defects and

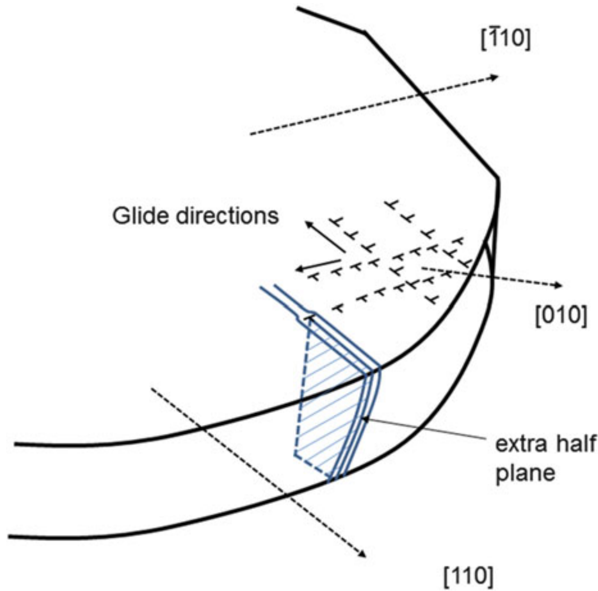

Fig. 9.7 Schematic drawing of slip via (111) glide planes in a (100) wafer by excessive thermal stress and the orientation of the extra half plane relative to the wafer external surfaces and also to the crystal directions in the wafer plane for a (100) wafer. As indicated in the figure, the dislocations arrays typically appear in lines, the <010> directions. Such dislocation formation is typical for too harsh thermal processing or unsuitable combinations of layers deposited on a silicon wafer before thermal processing (After [4])

impurities and can thus influence the concentration of intrinsic point defects. This can both be detrimental and useful, depending on the situation. Stacking fault growth is particularly significant in connection with e.g. oxidation (see later sections).

9.3.4 Three-Dimensional Defects

Three-dimensional defects in silicon can be introduced unintentionally (e.g. the aggregates of vacancies, also called crystal originated pits (COPs), see also Chaps. 4 and 5, or precipitates of unwanted impurities), or on purpose, e.g. precipitates of oxygen, which have mostly the structure of amorphous silicon dioxide [7, 8]. Oxygen precipitates and defects that can be co-created (e.g. stacking faults) are instrumental to remove unwanted metal precipitates from the device layer near the surface of the wafer to the bulk, a process called intrinsic gettering (see a later subsection of this chapter and Chap. 6). As mentioned before, extended defects are not necessary but can form unintentionally under unsuitable process conditions for their prevention (or can be engineered to be formed by using a process which provokes their formation, relevant detailed information to be found in Chaps. 4, 5 and 6).

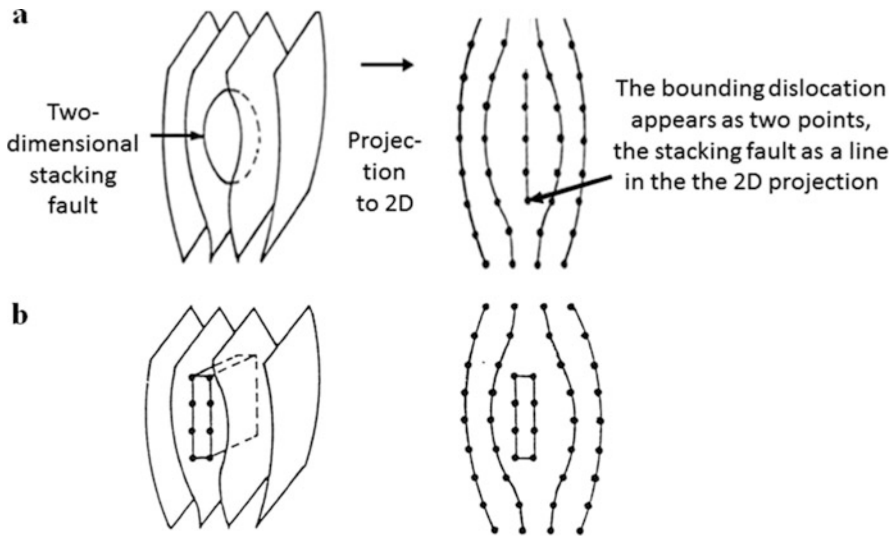

Fig. 9.8 (**a**) Schematic drawing of an extrinsic stacking fault in 3-D and in 2-D projection. (**b**) Schematic drawing of a precipitate in 3D (*left hand side*) and in the 2D projection (*right hand side*)

Once formed, all defects can interact and react with each other, which can give rise to quite complex defect reaction and interaction scenarios. This is at the core of defect engineering, and is ultimately the chief cause that metal impurities, if not controlled, constitute an unacceptable risk to the performance and the reliability/durability of microelectronic and photovoltaic devices and products [12–15, 23].

The task of defect engineering is to control such reactions and interactions. The interaction can be direct (e.g. by exchange, absorption or emission of point defects), but it can also be indirect by elastic interaction or, most important, by electrical interaction, since most defects can assume different charge states, depending on the Fermi level in the active device regions and in the rest of the wafer (in a photovoltaic cell or a power semiconductor, the entire wafer is active device region, in microelectronic products other than power semiconductors only a very thin surface layer of a few μm thickness is the active device area).

9.4 The Electrical Activity of Defects in Silicon and Some Consequences

In defect engineering, the fact that in silicon most defects are electrically active is of paramount importance, both in terms of direct electrical effects on the recombination and generation lifetimes (compare Chap. 3), and in terms of indirect

effects by influencing the solubility and diffusivity of defects [21, 22] and reactions between defects. The effects are further complicated by the fact that such defects can act as single or multiple donors/acceptors or even both, i.e. an "amphoteric" behavior, a term used in chemistry for elements that can act as acids and as bases. In fact, it is insightful to note that there is an interesting analogy between the chemistry in water and the "defect chemistry" in silicon. This idea has been put forward by Hannay [24] and is a very useful concept for defect engineering, because well-known principles from wet chemistry that many engineers are familiar with, such as the law of mass action, can be applied in an analogue fashion, see also [21] for more detailed explanations.

In this analogy, silicon assumes the role as a host, as water does for liquid chemistry. Donor impurities such as phosphorus correspond to a base, since they add an electron to the conduction band, in analogy to the fact that ammonium hydroxide NH_4OH adds an OH^- to water. The two corresponding "chemical" reactions are:

$$P \rightarrow P^+ + e^-$$

$$NH_4OH \rightarrow NH_4{}^+ + OH^-$$

Likewise, an acceptor impurity such as boron corresponds to an acid.

The law of mass action holds for all cases, also for the absence of any acids in water, or the absence of significant impurities (in comparison to the intrinsic concentration of holes and electrons) in silicon. In such a special situation (normally on speaks of undoped intrinsic silicon) the product of the concentrations of the two types of ions/charge carriers electrons e and holes h (as denoted by square brackets) is constant for a given temperature, and the equilibrium constants K for the reaction depends exponentially on temperature, which reflects the energy needed to break the respective bonds, as illustrated in Fig. 9.9.

$$\left[H^+\right]\ [OH^-] = K_{water}$$

$$\left[e^+\right]\ [h^-] = K_{silicon}$$

In complete analogy to the fact that the solubility of, e.g., Fe^+ ions in water depends on the pH-value, the solubility of impurities can be influenced significantly be changing the concentration of boron acceptor impurities in Si, the mathematics to quantify the dependence has to take into account the Fermi level (not relevant for water), but is in principle analogous:

Figure 9.10a shows experimental values from Hannay for the impurity Li, which is a shallow donor, similar to phosphorus:

The phenomenon, that the solubility of Li can be increased by doping can be qualitatively understood considering the reaction, and remembering the law of mass action from chemistry:

$$Li \rightarrow Li^+ + e^-$$

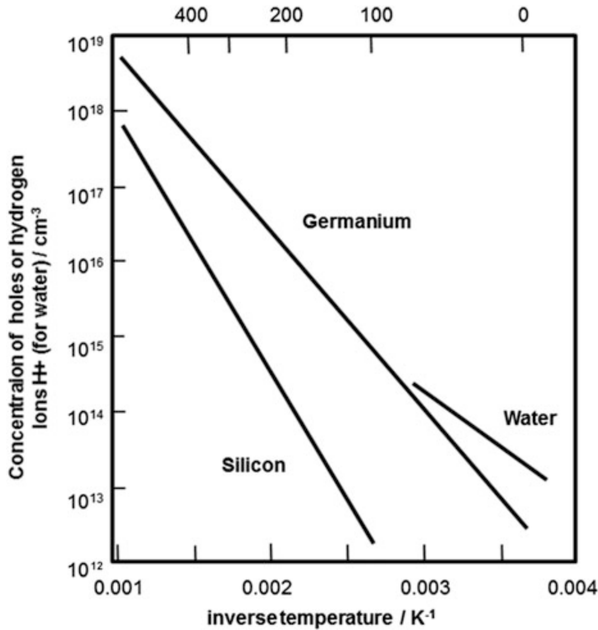

Fig. 9.9 Intrinsic carrier concentration in Si and Ge, and in water, and the concentration of OH− and H+ in neutral pH7 water (which corresponds to intrinsic silicon), after Hannay [24]

The "product" e^- of the ionization reaction of the Li impurity (which is a donor) is removed from the reaction by the addition of acceptors, which contribute holes h^+, which serve to neutralize (i.e. eliminate) the electrons e^-. Hence, by the law of mass action, the equation is driven to the right. This means, that in addition to the "normal" solubility of Li in intrinsic undoped silicon an additional solubility results from the additional thermodynamic entity Li^+. In technology, this defect engineering "tool" has been used in the production of Li drifted silicon nuclear detectors at times when it was not possible to manufacture silicon crystals pure enough to use intrinsic silicon for nuclear detectors (which is possible nowadays, that technology has been obsolete for more than 30 years) [25].

Similar phenomena have been observed by Gilles et al. [22] for the solubility of the transition metals Mn, Fe and Co in silicon as a function of doping level, as shown in Fig. 9.10b for Fe in Si at 700 °C:

The solubility increases by orders of magnitude for both heavy n-type and p-type doping. For n-type doping there are strong indications that this must be due to an immobile i.e. non-interstitial Fe species. This finding is especially important for Chap. 8 in which the results for Fe, Co and Mn in silicon from nuclear methods found strong evidence for substitutional Mn, Fe, Co, which is somewhat at variance with the other data on the interstitial and substitutional solubilities of these elements in silicon.

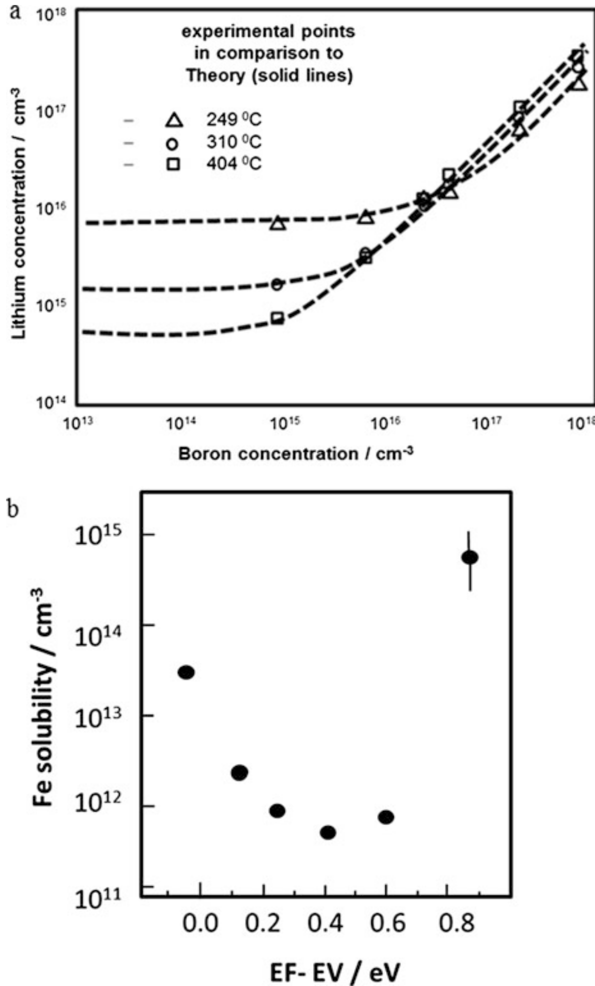

Fig. 9.10 (**a**) Li solubility as a function of the acceptor concentration for three different temperatures, data from [24] (**b**) The Fe solubility in silicon at 70 °C has been measured by radioactive tracers as function of the Fermi level for different doping levels, and is plotted as a function of the Fermi level at 700 °C, after Gilles et al. [22]

The strong doping dependence of transition metals is one component of the scientific basis for modelling phosphorus diffusion gettering and the larger resilience of heavily boron doped epitaxial wafers against metal contamination. Also, in photovoltaics the effect of the heavy aluminium doping by the Al paste on the back surface of standard photovoltaic cells has most likely led to a high resilience against metal contamination through the gettering action of the highly Al doped back surface layer, in combination with the phosphorus diffusion gettering during the n^+ emitter diffusion for the basic n^+p photovoltaic cell [17].

There are many more phenomena associated with the fact that defects can occur in different charge states in silicon, some of which will be mentioned in the following sections. In all defect engineering efforts, the potential impact of charge states on the solubility, the diffusion constant and potential interaction with other defects has to be kept in mind.

9.5 Formation Mechanisms of Extended Defects and How to Suppress Their Formation in Defect Engineering

As mentioned before, extended defects are not required by thermodynamics. Whether extended defects form or whether their formation can be prevented both during the crystal growth or during the processing in microelectronics/photovoltaic manufacturing depends on a kind of "competition" between two "forces" (not mechanical but proverbial forces). In strict science term it is the tendency of the system in question to lower its energy by the formation of defects for A, and the activation energy or kinetic barrier for defect formation for B, which help to understand different pathways for extended defect formation and the problem caused in electronic devices [6, 15]:

(A) Driving forces: These are the super-saturations of intrinsic defects and/or impurities and/or thermal/mechanical stresses due to temperature gradients or different thermal expansion coefficient of silicon, silicon dioxide, nitride etc. (circles on the left in Fig. 9.11).
(B) Impeding forces: These are the nucleation barrier (activation barrier, see Fig. 9.12) for the formation of extended defects and the fact the diffusion of point defects can be slow (if nothing moves, nothing can happen even though it would be energetically favorable), i.e. a kinetic barrier.

With respect to the principles (a)–(c) described in Sect. 9.2, Fig. 9.11 is a practical "map" or guideline how to prevent the formation of extended defects by defect engineering interventions. The defect engineering tools developed on this basis can be categorized according to two strategies, strategy I is to reduce the "driving forces", and strategy II is to strengthen the "impeding forces", usually both types of tools are applied simultaneously.

Strategy I Reduce the driving forces, such as thermal/ layer stress or the super saturation of intrinsic defects

In thermal processing the thermal stress can be done by suitably slow temperature ramps from the furnace stand-by temperature to the processing temperature and back after the end of the process, and slow enough move-in and move-out of the wafers into and out of the furnaces. Both too fast ramp-up and too fast move-in (or ramp down and move-out) lead to significantly higher temperatures of the outer

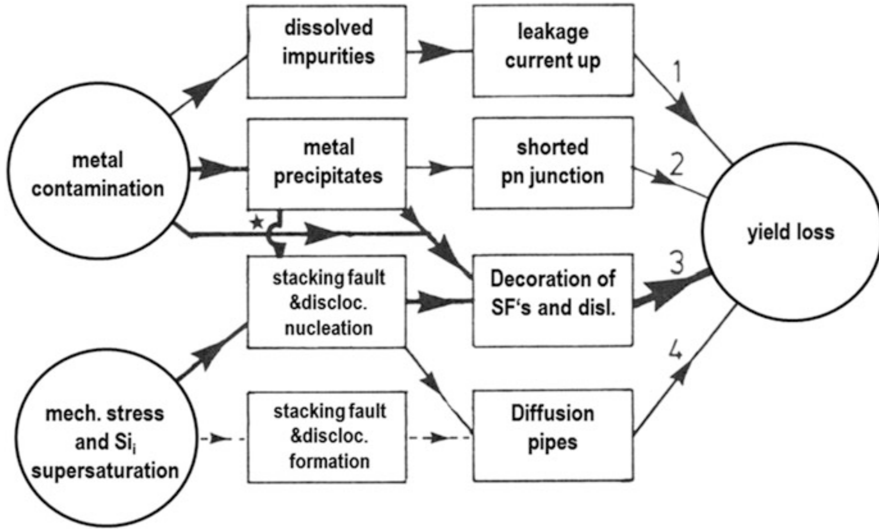

Fig. 9.11 Different pathways from driving forces to the formation of extended defects and or electrically active point defect impurities to electrically detrimental defects [6]

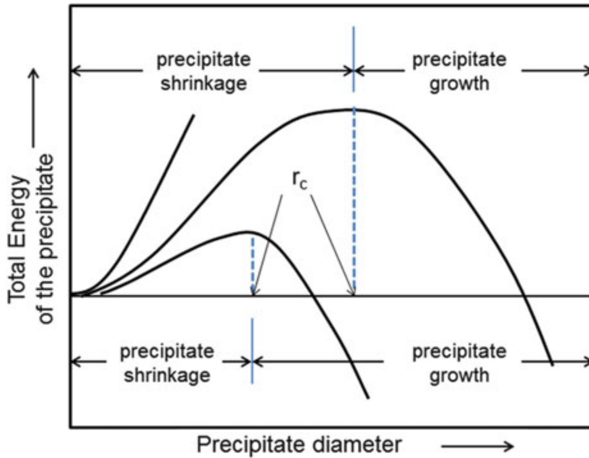

Fig. 9.12 Schematic representation of the energy barrier for the nucleation of extended defects. The chemical energy gained by reduction of the supersaturation of intrinsic or point defects is proportional to the volume (i.e. proportional to radius r cubed) whereas the additional energy needed to create the additional surface (proportional to the square of r). Therefore, for small radii, there is barrier which has to be surmounted (spending energy for the additional surface energy without much gain energy from reducing the supersaturation), which is an "impeding" force for the formation of extended defects

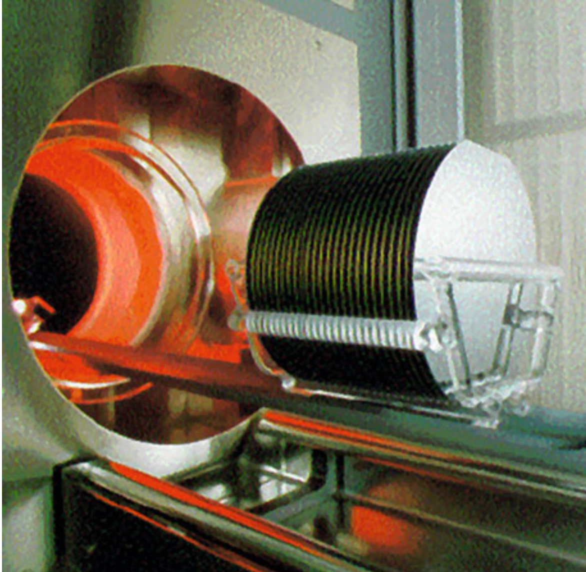

Fig. 9.13 Photograph of wafers moving into a furnace at a stand-by temperature of 800 °C. Note the close spacing of the wafers in the quartz wafer carrier which is needed to reach a reasonable productivity, to accommodate as many wafers as possible in the limited space of a furnace tube. The quartz has been manufactured from high purity quartz to reduce metal contamination

rim/center of the wafers (respectively) which are relatively closely spaced in a boat, see Fig. 9.13.

It is a general experience that a temperature difference of 100 K between the edge and the center of the wafers can occur during move-in and ramp-up which is close to the process window limit, beyond which plastic deformation will be induced since the critical shear stress for dislocation formation is exceeded [4, 12, 13]. Such dislocations result in wafer warpage to form a saddle or propeller-like shape (since the perimeter of the wafer is larger than the center) which can lead to problems with photolithography and which can cause excessive leakage current in devices via the dislocations generated. During move-out the opposite temperature gradient develops, experience (and simulation) shows that the dislocations will develop in the center of the wafers, which is larger than the perimeter, deformation will be like a bow. Figure 9.14 shows an example of a untypically strong formation of dislocation formation by excessive thermal stress. To prevent such defects by very slow ramps and move-in and move-out would be easy, however the resulting loss in productivity and the corresponding increase in cost would not be acceptable. So defect engineering has to identify a good compromise between the two requirements high productivity and the prevention of defects.

More specifically, defect engineering has to determine the process limits and a sufficient safety margin for the process parameters move-in/out and speeds and ramp rates so that the process is robust against small variations of the material

Fig. 9.14 Dark field image of part of a wafer which had undergone excessive temperature stress (during rapid thermal anneal as part of a haze test), where the dislocation arrays in the [100] direction are well visible via the etch pits of the dislocations (additionally visible: haze from metal contamination, which is not present in the dislocated areas due to gettering of the metal impurities by the dislocations). The wafer is has been defect etched to show develop etch pits where the dislocation penetrate the surface, the light scattering from the etch pits shows up the intersection of the glide planes with the surface, where a high density of etch pits shows up as lines in the <110> directions, also compare Fig. 9.7

(which depends, among other things on the oxygen content, doping level and thermal history before the thermal process in question [4]), but at the same time does not waste equipment capacity. It is also obvious from first principles in radiation physics that a larger distance between the wafers would decrease the tendency for dislocation formation by thermal stress; however this implies again lower productivity since furnace tubes have a limited length.

Likewise, the layer stress between silicon and a silicon nitride layer due to different thermal expansion coefficients can be reduced by adding a thin silicon dioxide layer between the silicon and the nitride [4]. The silicon dioxide gets viscous above a certain temperature and allows the glide of the nitride on top of the silicon, very similar to a lubricating layer between two plates.

In terms of the topic (b) for the requirements of defect engineering mentioned in Sect. 9.2 for manufacturing, an efficient and effective method to make sure that there is no formation of dislocations is needed. Such a fast low-cost method is the etching method for test wafers or for partially or fully processed production wafers. In order to satisfy the requirements of a quality management system, there must be a standard operating procedure in place for defect etching, in combination with a standard procedure to inspect the wafers by dark field illumination and evaluated the length of glide lines (if any), as shown in Fig. 9.14, and additional inspection

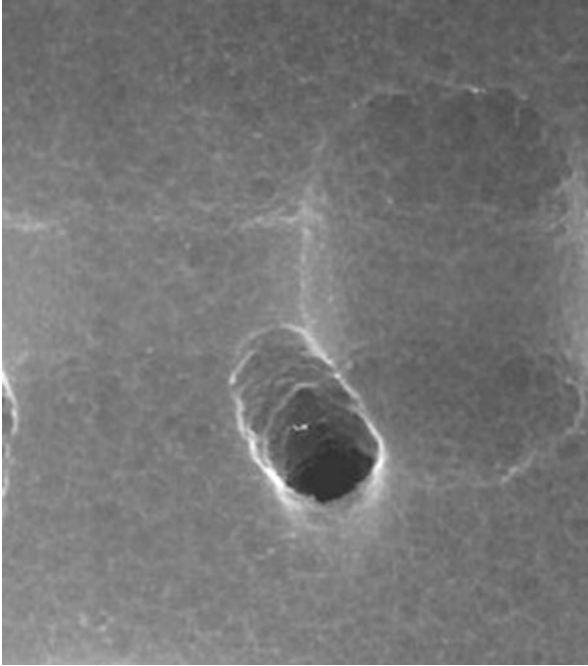

Fig. 9.15 Etch pit of a dislocation in the center of the photograph. The slight etch depression next to it is where the direct contact for a 4 Mbit DRAM was located (The etching depression is due to the residual damage from the implantation of the contact). The field of view in the micrograph is approximately 1.5×2 µm

by either optical microscopy according to a fixed inspection plan (i.e. the number of fields of view at which magnification) or by SEM with a similar inspection plan. Figure 9.15 shows the etch pit of a dislocation after Secco defect etching, the funnel-like etch pit is very characteristic, so an unambiguous attribution of that type of etch pit to a dislocation is possible (for details and reference on defect etching, see Chap. 6, and [26]).

Another very efficient way to check for dislocations is the use of X-ray topography [26], in which the dislocation arrays inside the whole volume of the wafer show up as contrast, an example of an X-ray topograph is shown in Fig. 9.16, which images approximately a quarter of a wafer of a 150 mm wafer (equipment which can image a whole wafer can be purchased to-day, but was not available at the time when the X-ray topograph was recorded). The dislocations on the glide planes in the <100> direction are clearly visible. In addition, additional dislocations have emerged near the wafer flat. The nucleation of these dislocations was most likely due to mechanical damage from the laser marking near the wafer edge.

In terms of topic (c) for defect engineering under production conditions, such regular process inspection results then have to be entered in to a run chart (process results plotted vs. the run number or date/time) and, what is better, developed into

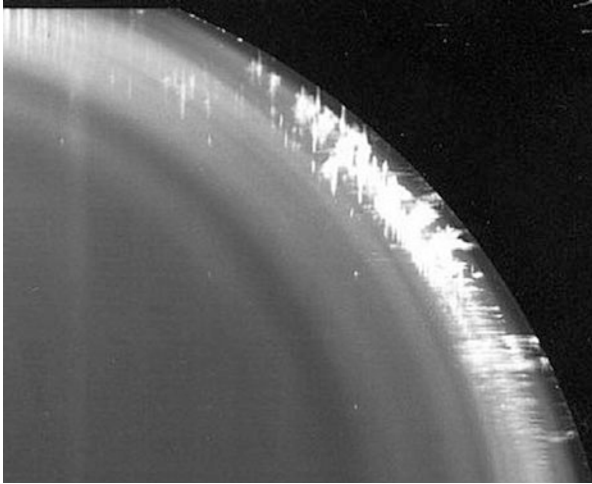

Fig. 9.16 X-ray topograph of part of a 150 mm wafer. The extensive dislocation formation in the <100> direction is clearly visible by a crisscross pattern of bright lines (the imaging condition is set such that the lattice distortion around the dislocation core gives rise to additional reflected X-ray intensity). In addition one can see many dislocations near the wafer flat and a ring-like contrast, the latter stems from the lattice distortion of oxygen precipitates and additional secondary bulk microdefects triggered by oxygen precipitation

an SPC (statistical process control) control chart [19]. If according to the definition of SPC (i.e. cpk-value larger than 1.67), the process is under control, there is a probability lower than 3.4 in 1 million that that process will generate dislocations in the next production runs, i.e. quality is "predictably" manufactured [19].

The second type of "driving force" frequently encountered in microelectronic and photovoltaic production is the super saturation of point defects and impurities.

The most frequent process for which this applies is oxidation. Oxidation leads to a super saturation of self-interstitials, which can result in the formation of oxidation induced stacking faults (OISFs) [4], if the super saturation is high enough, and/or if there are nuclei which significantly lower the nucleation barrier. The driving force can be reduced e.g. by reducing the oxidation rate, so the formation of OISFs can be suppressed, via avoiding exceeding the nucleation barrier. To kind of visualize the super saturation via the growth rate of extrinsic stacking faults, Fig. 9.17 shows the lengths of stacking faults which were intentionally nucleated as a function of temperature for dry oxidation for 3 h under conditions where the reduced nucleation barrier was easily surmounted. The different growth rates at different temperatures and crystal orientation demonstrate that the production rate and the resulting super saturation of self-interstitials during oxidation depend on the processing temperature and the oxidation rate (influenced via the oxygen concentration in the gas fed through the furnace).

It terms of defect engineering for production control (b) and (c), similar principles apply as for the routine checks for dislocations, namely defect etching

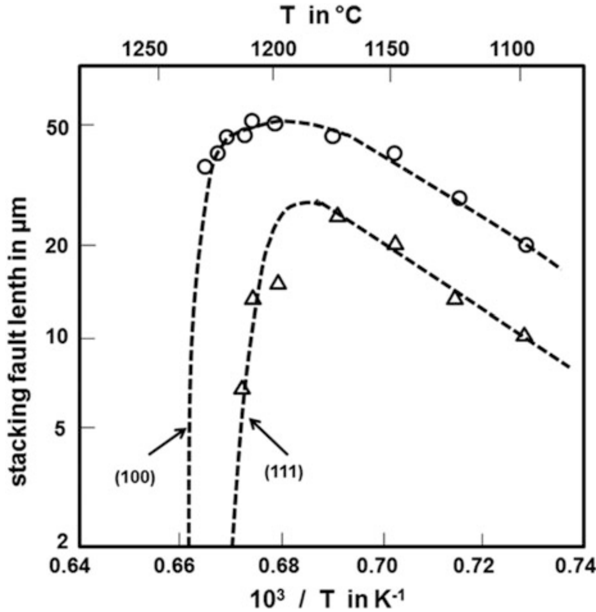

Fig. 9.17 Length of stacking faults, as measured by defect etching and optical microscopy. The length of the stacking fault reflects the super saturation with self-interstitials caused by oxidation at different temperatures and different crystal surfaces, after data from [4]

in combination with a test plan and a run chart of an SPC chart. Also, a good compromise between high productivity (= high oxidation rate) and robustness of the process against spurious formation of stacking faults has to be determined from process window experiments.

So, to reduce the driving forces most of the time does come at a cost (usually in terms of lost productivity, but it can also result in higher cost for better wafers or more expensive equipment). Therefore, it is clear that in parallel, the second defect prevention strategy has to be employed.

Strategy II Strengthen the "impeding forces"

Quite frequently, the cost for reducing the driving forces is simply too high, and/or the process is difficult to control, such as that the nucleation barrier for the formation of extended defects is lowered by the presence of metal impurities. Such metal impurities have an extremely low nucleation barrier themselves, and diffuse very fast at the same time, which is a most "unfortunate" combination, as will be shown later [15], the nucleation barrier of the formation of dislocations and stacking faults can be dramatically reduced by formation of metal impurity precipitates (which often have a small nucleation barrier and a negligible kinetic barrier, see several pathways in Fig. 9.11). Figure 9.18 is a striking example how the presence of metal impurities can dramatically enhance the formation of OISFs: For

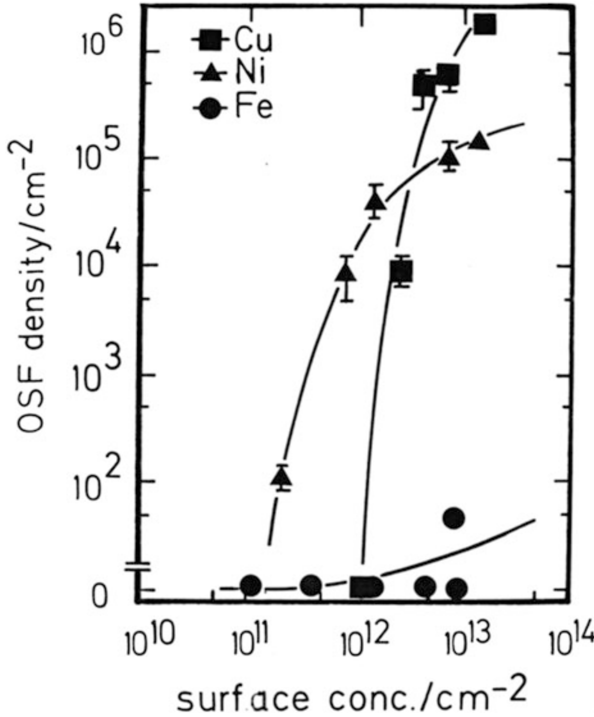

Fig. 9.18 Enhancement of the formation of oxidation induced stacking faults by surface metal impurities introduced intentionally before oxidation, after Hourai et al. [27]

a given process with negligible metal contamination the process does not generate oxidation-induced stacking faults. However, for Cu or Ni impurity concentrations which correspond to less than one person in the world's population (this means an impurity concentration of 10^{12} cm^{-3}, one cubic centimeter of Si contains 5×10^{22} atoms) the stacking fault density increases by up to a factor of 100,000! [27].

Therefore it is clear that a central requirement to strengthen the nucleation barrier is to prevent metal contamination as best as possible. As mentioned in Sect. 9.2, it is absolutely necessary to have fast detection methods so that the up to 1000 process steps in microelectronics manufacturing can be regularly screened for metal contamination. Figure 9.19 [28] gives an overview of typical metal contamination levels in microelectronic manufacturing. It is obvious that plasma etching process and implantation are particularly prone to metal contamination, and that also handling via wafer chucks made from stainless steel or other materials which contain dangerous metals can contribute significantly to metal contamination, the published results are congruent with the practical experience of the author in 20 years industry experience [11–15].

The permissible metal contamination levels in photovoltaic production are much higher [17], and therefore it would not be economical to dramatically reduce these

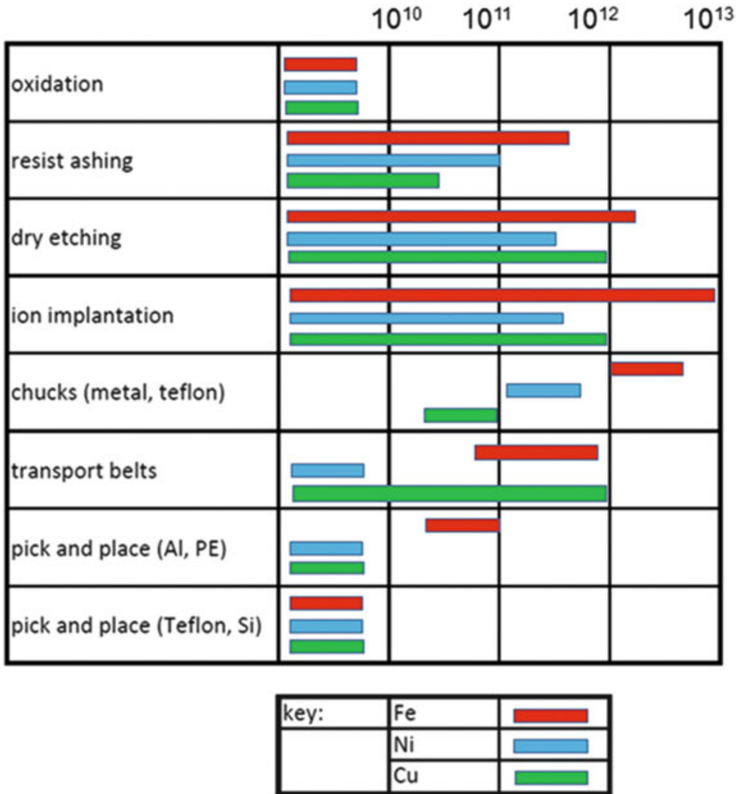

Fig. 9.19 Typical metal contamination levels in microelectronics production. Depending on the particular device process, the typical requirements may be even lower than the concentration limits indicated by the traffic light symbols shown here. Data according from [28], the units given on top of the diagram are cm^{-3}, i.e. volume concentrations

levels by using e.g. cleaner but significantly more expensive chemicals. However, it is predictable as designs of PV cells get more sophisticated, the allowable metal contamination levels will go down appreciably.

So, for effective and efficient process control of extended defects, control of metal contamination levels, according to (b) and (c) of defect engineering principles under mass production is absolute necessity, which implies the needs fast and efficient methods to detect metal contamination, such as the haze method described earlier, or the microwave PCD method described in Chap. 3. While most electrical and optical detection methods, such as DLTS, FDLTS, FTIR or electro and photoluminescence methods (see Chap. 3) are very good methods to identify and in-depth characterize the metal impurities, they are relatively slow and some of them very expensive methods with a sampling area of the order of mm^2. On the other hands, electrical methods based on measurement of the minority carrier lifetime (e.g. microwave PCD) are normally not impurity specific, but very fast

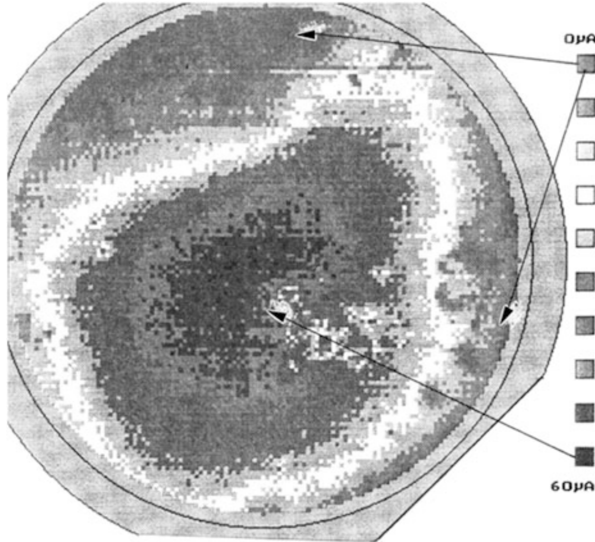

Fig. 9.20 Diffusion length wafer map of a reactive ion etched and rapid thermal annealed wafer to drive in any process-induced metal contamination into the wafer volume. The measure pattern is the magnitude of the photo induced current on the wafer back surface which results from a laser beam which is scanned over the wafer front surface. The technique has been named as Elymat technique. The whole volume is sampled because the minority carriers generated on the wafer front surface have to diffuse to the back surface where they are collected in a space charge region of the an electrolyte semiconductor contact. The contact with a low surface recombination velocity is implemented by a dilute HF electrolyte. The Elymat technique is not as popular as the microwave PCD method or the SPV method, due to the complications which arise from the safety issues that have to be controlled when using HF

and can sample large wafer areas (essentially the complete wafer surface), such as SPV (surface photo voltage, [15]) or they can even sample the complete wafer volume, if the diffusion current through the wafer is measured, as in the Elymat method (compare also Fig. 9.25, [29, 31]). Figure 9.20 shows the result of such a whole volume measurement of the spatially resolve photocurrent measurement of a dry etched test wafer. The measurement principle is explained in the captions of Figs. 9.20 and 9.25 and in Ref. [29]. Since a reference wafer would have a diffusion length at least ten times as high and almost homogenous over the wafer, it is demonstrated that the contamination is considerable and not homogenous over the wafer. With the help of such wafer maps it is possible to

- identify the localized sources of the metal contamination in the plasma etching equipment
- to test whether equipment and/or process modifications can lower the metal contamination level
- whether different cleaning steps can or cannot remove the impurities

A similar wafer map could have been generated with the microwave photoconductivity method explained in Chap. 3, with the proviso that only a surface layer of a few micrometer thickness is sampled. Commercial equipment to generate wafer maps of this kind is also available for the SPV method. In the case of SPV, the top 5–50 μm layer is sampled for metal contamination, depending on the illumination conditions, which can be chosen within certain limits [32].

Another important way to strengthen the "impeding forces" is obviously to go to lower temperatures in processing, since this slows down diffusion processes, also for some metals (if "nothing moves" or does not move fast enough, nucleation of defects with the help of metal contamination is prevented kinetically) This constitutes increasing the kinetic barrier to defect formation.

This will be illustrated in more detail in the next section, in which more complex defect formation scenarios will be described, which are typically encountered in microelectronic and/or photovoltaic production. A lower processing temperature may of course at the same time slow down the actual diffusion or annealing process needed in the device production, so here similar productivity issues may arise.

9.6 Selected Practically Important Examples for Defect Formation Mechanisms and Prevention Measures from Microelectronics Manufacturing

In this section the knowledge and understanding about fundamental principles of defect formation (as outlined in the previous sections) will be put into the context of common defect formation and defect prevention scenarios in microelectronics production. In this way, we will further explore the defect mechanism paths of Fig. 9.11 and in particular give examples for the electrical effects of defects on devices, i.e. up to when the defects affect not only the performance but also the functionality and/or the reliability of the electric devices. We will always start in one or both of the two circles representing the primary "driving forces" of defect formation on the left of the pathway map, namely super saturation of intrinsic defects, mechanical stress and metal contamination.

Due to the ubiquitous and widely varying metal contamination in microelectronic and PV production, which together with the "unfortunate" properties of these impurity atoms [15] makes metal contamination the highest-ranking risk factor for defect engineering and defect control for defects in silicon, we will first describe the general properties of the practically most important 3d-transition metal impurities in some detail. Subsequently, we will consider examples for more complex interaction of extended defects and metal impurities, and finally we will treat gettering by intentionally introduced defects in the bulk to remove metal impurities from active device areas, which is the beneficial rather than detrimental effect of the complex interaction of metal impurities and extended defects.

9.6.1 3d-Metal Impurity Diffusion and Precipitation

In Sect. 9.5 and in Chap. 3 the detrimental effect of metal contamination on the minority carrier diffusion length in silicon was already mentioned, and how this effect can be used to scan a wafer or a process in a fast and efficient manner for the presence of metal contamination. In this subsection, we will turn our primary attention to the most critical and detrimental effects associated with an uncontrolled metal contamination, the formation of metal precipitates, which can lead to a drastic increase in the formation of extended defects via lowering the nucleation barrier for dislocations and stacking faults, with potentially dramatic consequences for the manufactured product [4, 13, 23].

Many of the transition metal impurities (in particular Fe, Co, Ni, Cu and Pd) in silicon combine two rather "unfortunate" properties, which make it extremely easy for them to generate precipitates (i.e. three-dimensional defects), where Co, Ni and Cu are in particular dangerous because of a third "unfortunate" property.

The first property is that these elements diffuse extremely fast for solid state diffusion, up to ten orders of magnitude faster than doping elements like boron or phosphorus, as illustrated in Fig. 9.21. This means that at common diffusion temperatures these elements can distribute almost homogeneously throughout the wafer. Remarkably, the diffusion constants at those device processing temperatures

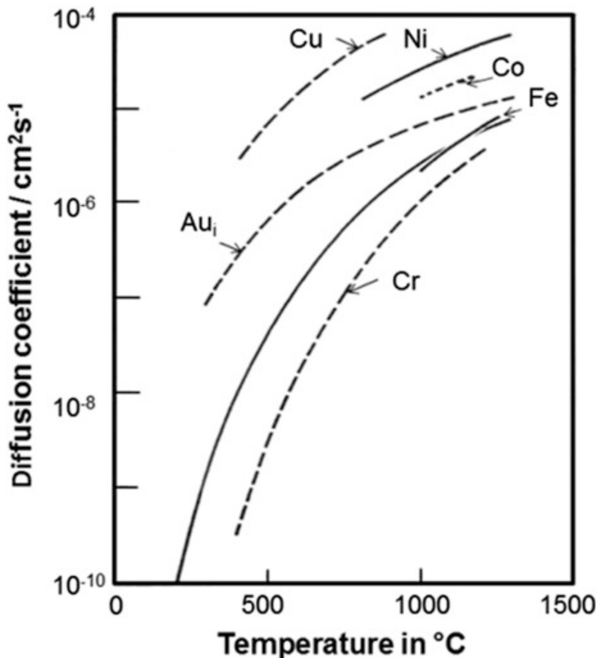

Fig. 9.21 Diffusion coefficient of some transition elements in silicon (After Graff [20])

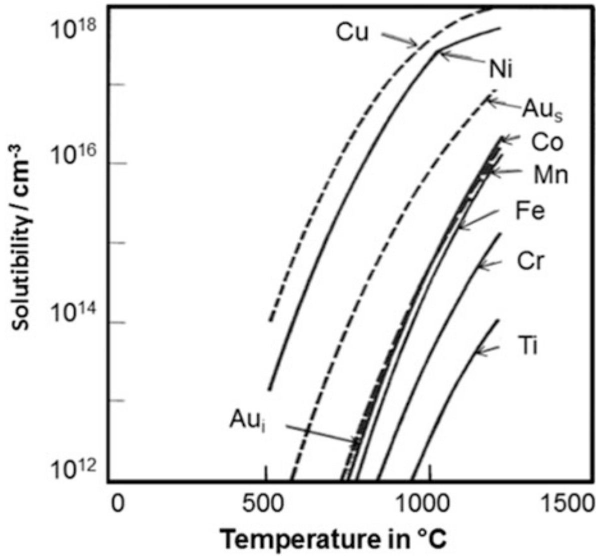

Fig. 9.22 Solubility of transition elements in silicon (After Graff [20])

are of the order of 10^{-6}–10^{-4} cm^2s^{-1}, which is a range typical for the diffusion of e.g. ink molecules in water.

The second "problematic" property is that the solubility is very low in absolute terms (a few parts per million at diffusion temperature) and that it drops rapidly with temperature (see Fig. 9.22), which means that during cooling a high super saturation develops, so a driving force for precipitation even for µg of contamination is extremely high. In addition, the precipitation process is further supported by the high diffusion coefficient which "unfortunately" falls very slowly with temperature, so the kinetic barrier is very low.

The third undesirable additional property of Co, Cu and Ni is that the initial stages of the precipitates (e.g. $NiSi_2$) are isomorphous to the silicon lattice with a very small lattice mismatch [15]. So the nucleation barrier is almost negligible, since the surface energy and lattice distortion energies are small, which are the main factors which determine the energy barrier for extended defect nucleation. So the energy barriers is unusually low as well.

As for dislocation and stacking fault detection, the "work horse" in defect engineering for the routine detection of metal precipitation is defect etching. As an example and illustration of this extremely fast nucleation Fig. 9.23 shows etch figures from Cu precipitates, which have formed after a haze test with intentional Cu contamination from the wafer back surface, metal precipitates can be found mainly at the two surfaces (and very few in the bulk, unless intrinsic gettering is active). Further details about defect etching, in particular how to replace the environmentally undesirable chromium containing defect etchants (mainly used in the examples) by alternative acid mixtures which have the same sensitivity can be found in Chap. 6.

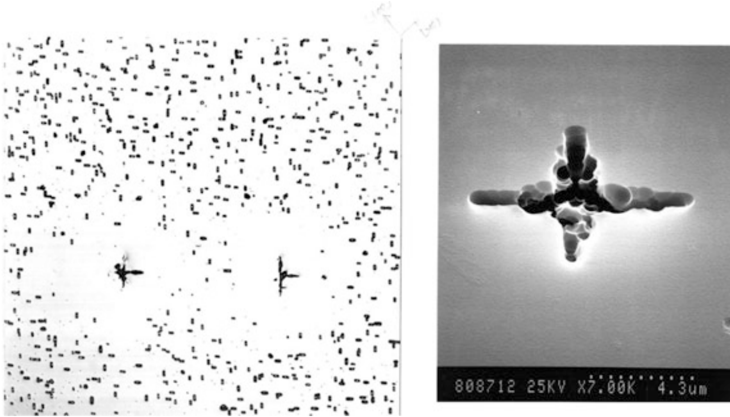

Fig. 9.23 Microscope image of the etch pits of Cu precipitates after a haze test, as described in Sect. 9.2. On the left hand side in the optical microscope image, there is a high density of etch pits of $CuSi_2$ precipitates, which consist of several layers of $CuSi_2$ on {111} crystallographic planes, therefore the etch pits resemble those of stacking faults. The two larger more complex etch pits in the center are secondary Cu precipitates which involve the formation of dislocations nucleated on the initial precipitates on {111} planes. A typical etch pit like the two complex pits in the center of the optical microscope image on the left is enlarged in an SEM image on the right hand side. Note that the large secondary precipitates have obviously emptied their vicinity from Cu in a very effective manner, since there are no $CuSi_2$ precipitates in an area around the larger precipitates. In other words, a re-dissolution of the original precipitates and a kind of gettering of the Cu to the secondary defects has happened. Quite remarkably, all this has occurred during the 10 s or so of cooling from 1200 °C to near room temperature in the rapid thermal annealing equipment used. So primary nucleation, re-dissolution, diffusion over a distance of approximately 10 μm and re-precipitation have all occurred during this time

As explained in the caption of Fig. 9.23, there is an "internal" competition between two types of Cu precipitates, the ones which resemble stacking faults are easier to nucleate (they are almost perfectly isomorphous to the silicon lattice), and the larger precipitates, which are colonies of small Cu precipitates and associated dislocation complexes. They compete for the Cu in super saturation, obviously the second more complex Cu precipitates win.

Similar to the competition between different Cu precipitates, there can be a competition and interaction and mutual enhancement of precipitation between different metals. Figure 9.24 demonstrates this interaction on a haze tested wafer which has been intentionally contaminated by slight scratching of different metal wires on the wafer back surface and a haze test described in Sect. 9.2.

Metal contamination can practically occur in any process, and in many instances there will be several of the hazardous metals present simultaneously, therefore the interaction demonstrated in this artificial contamination experiment is relevant for production. As already briefly mentioned in the preceding section, experience has shown that high temperature diffusion, dry etching of silicon surfaces and ion implantation all have a particularly high risk of contamination [14], and even

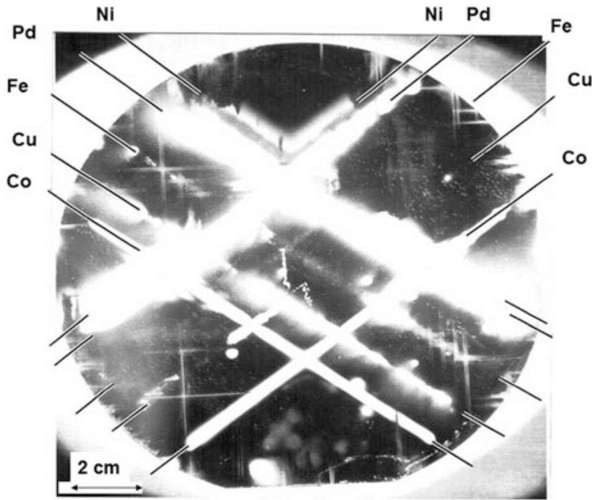

Fig. 9.24 Haze test of a wafer after intentional contamination by Ni, Pd, Fe, Cu and Co along the lines indicated. *Bright* areas represent regions of a high density of precipitates. Pd has the strongest tendency for precipitation. Fe displays the weakest precipitation. Note that there are obviously interactions between some of the metals at the crossing points of the lines. Also visible: the etch pits of dislocations on their respective *glide lines*, as *thin bright lines* in <110> direction, after [33]

cleaning (!) can, under certain circumstances lead to additional metal contamination (see also Fig. 9.27), [31].

Thus, in the sense of item (c) of Sect. 9.2, these are the processes to be primarily screened by defect engineering, because of the high and hazardous defect formation potential of the metals themselves, and because of the interaction with other extended lattice defects (see Fig. 9.11) which will be explained and illustrated further in Sect. 9.6.3 of this chapter. So, the task of defect engineering is to investigate the main factors which potentially increase metal contamination and counter measures to take to prevent it, and how to monitor the "state of health" of the suspicious process regarding metal contamination at an acceptable cycle time and for acceptable cost.

Such a screening concept is typically implemented by a combination of the methods to detect metal contamination via their effect on the minority carrier diffusion length (SPV, Elymat, Microwave photoconductivity) and defect etching/X-ray topography.

The least common of the electrical methods, the Elymat technique [29] pioneered was by Lehmann and Föll. It is at the same time the most "versatile" method (also see Fig. 9.25), since it is effectively a way to build an "instantaneous" test diode, both on the front and back sides of the wafer using a semiconductor electrolyte contact by HF. As pointed out before, the HF electrolytic contact ensures a low surface recombination, so that the effect on the minority carriers generated by a

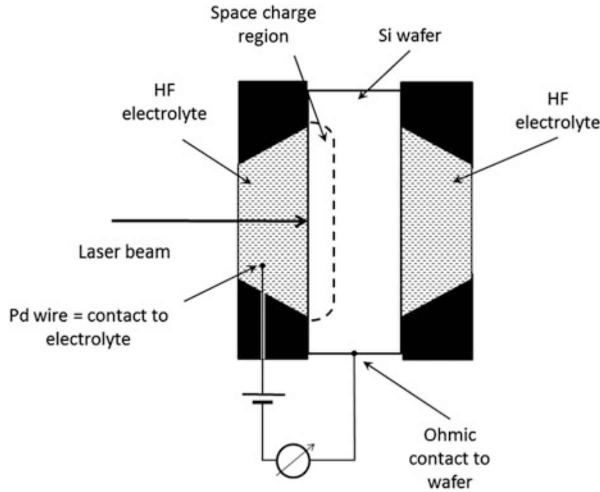

Fig. 9.25 Schematic cross section of the Elymat measurement set up, after [29]. The wafer is immersed from both sides by dilute HF, Pt wires in the electrolyte and tungsten wires on the edge of the wafer enable that an electrolyte-semiconductor junction can be formed, and at reverse bias, a space charge region can be made larger or smaller by an externally applied voltage, which will collect the minority carriers generated by a laser beam scanned across the front surface of the wafer. For routine monitoring it is usual not to use the wafer map itself, but the Key Control Characteristics are average values across the wafer and/or the standard deviation of the several hundred or thousand (depending on the chosen resolution) measured points across the wafer. Both the average value and the standard deviation in run charts or SPC charts are good indicators whether a process is in its normal state or whether untypically high contamination levels have occurred either over the whole wafer or locally (as detected by an increased standard deviation). In such a case, the pattern visible on the wafer map will be conducive to identifying the root cause of the contamination, almost like a fingerprint of the contamination source [29, 30]

laser beam scanned over the front surface can be studied (1) via the photocurrent from the front surface to the back surface (2) via the photocurrent collected at the front surface itself. The former is a measure of the diffusion length in the whole wafer volume, the latter is an indicator of surface near defects, since the diffusion distance is negligible, but the quality of the depleted zone and the junction to collect the carries is affected by defects, if these are present. In addition the dark leakage currents of both the back and front surface electrolytic contacts provide additional information on defects either caused by metals and/or by extended defects (however, these are naturally not spatially resolved but give a “fudged” quality indicator of the front or back surface).

In the routine monitoring of metal contamination, some metal contamination will almost always be found, so it is important to know for each process, what is

(i) the typical metal contamination level, and what is
(ii) the maximum contamination level which can be tolerated by the process

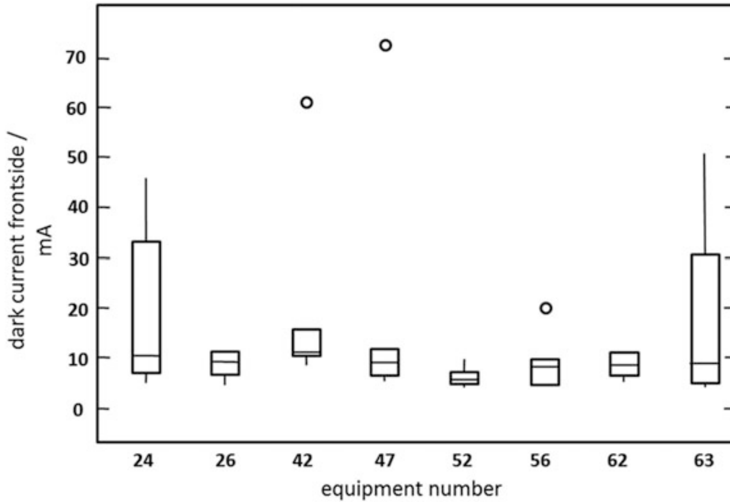

Fig. 9.26 Comparison of eight gate oxide furnaces in an Elymat cross sectional metal contamination study, the data are represented in box plot form. The quantity to indicate the metal contamination level is not the minority carrier lifetime, but the leakage current of the front site electrolyte-semiconductor contact, which in a way is a test device to study the effect of metal precipitates on a semiconductor space charge layer, in this case caused be a reverse biased electrolyte-semiconductor junction, so high is "bad", low is "good"

In order to get a representative data set for i., one can do a time series study ("longitudinal" study) on one equipment for a number of consecutive days or weeks, or a comparison of the equipment of one kind that are in use in that particular processing line ("cross sectional" study).

Figure 9.26 shows an example of a cross-sectional study and is, at the same time, an example of one of several ways how measurements using the Elymat method can be used. The equipment monitored were oxidation furnaces. The furnace tubes obviously differ considerably in their contamination levels, as indicated by the different levels of the electrolyte-silicon leakage currents measured without laser illumination (presumably due to different precipitate densities or sizes in the surface areas). By using in parallel a method to detect the type and concentration of impurities in terms of atoms/cm^{-3} one could in principle try to calibrate the leakage current vs. concentration. Since defect formation is a complex non-linear process, this would probably not work very well, but is not really needed: The important information for the defect engineer is already there: The furnace tubes 24 and 63 have an untypical high leakage current (i.e. contamination level) and need to be improved, the baseline level is represented by the other furnace tubes.

Figure 9.27 shows another example of a "cross sectional" monitoring study, in which different cleaning equipment has been characterized by the backside photo current, which samples the whole wafer volume. Contrary to the previous example, "good" is a high value of photocurrent, since defects decrease the photocurrent.

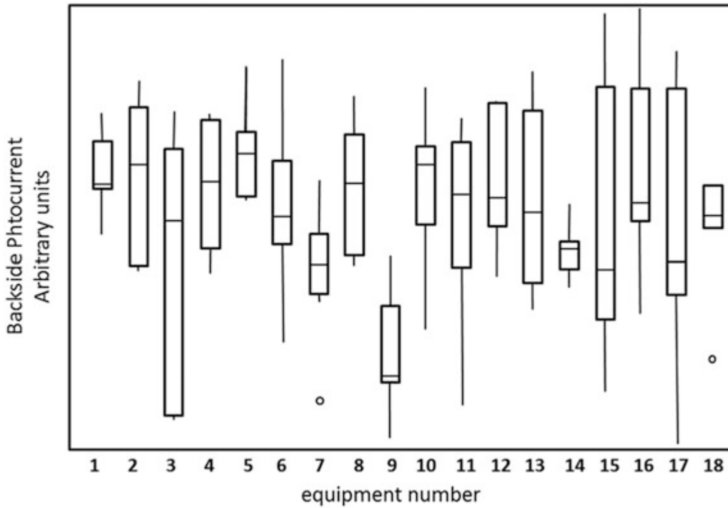

Fig. 9.27 Example for the weekly monitoring of contamination levels for wet cleaning equipment via the backside photocurrent represented in box plot form. By parallel investigations with DLTS it was established that the diffusion current in this case is mainly affected by the Fe contamination level. It is obvious that in equipment 3 and 9 the contamination is significantly higher in than in the other cleaners. However, the level was still within the permissible specification, but indicated a "special cause" (in the sense of statistical process control) for the contamination level which had to be eliminated by the responsible wet chemical process engineer before it could reach critical levels. Note that a higher level of contamination leads to lower current values, since it reduces the minority carrier diffusion length, so here in contrast to the previous graph high is "good" and "low" is bad

In the previous case, "good" was a low value, since defects cause higher leakage currents.

To demonstrate that the backside photo current measurement is indeed correlated to the Fe contamination level, a systematic correlation study on test wafers was performed, in which the BPC Elymat values were compared to the Fe concentration levels determined by the SPV technique on the same wafers. Figure 9.28 shows the results. There is a clear correlation and the expected dependence over several orders of magnitude. The scatter is due to the fact, that potentially other impurities also influence the BPC photocurrent marginally, and probably more significantly, that the two methods average over different areas, which causes some variability of the results, since the contamination level is not homogenous across the wafer, which is the case in such wafers.

To gain insight into question ii concerning the permissible contamination level, one needs to know HOW does a particular type of metal contamination cause leakage currents or other device failure, and to investigate the critical level of contamination level by process window or intentional contamination experiments [27].

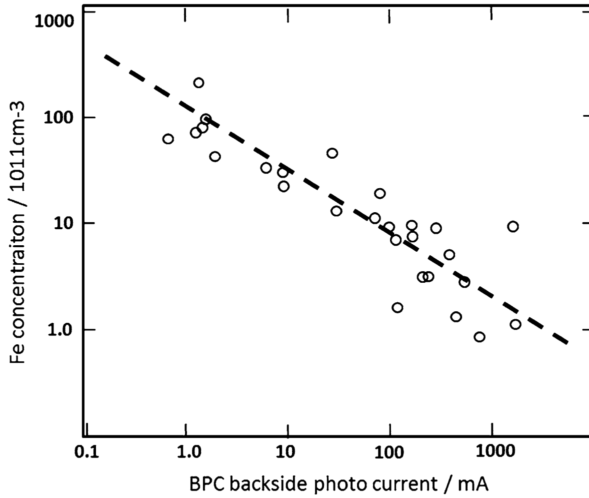

Fig. 9.28 Correlation between the Fe concentration in test wafers, as determined by SPV, and the Elymat photocurrent values, all values measured on the same test wafers

For Fe, which is typically the most important and at the same time, ubiquitous contamination element, the main degradation effect is the reduction of the minority carrier lifetime. By contrast, the main degradation mechanism of the Ni, Cu and Co is via their strong tendency for the formation of precipitates (as illustrated above), which can lead to several different device failure mechanisms, such as direct causing shorts in pn – junctions or thinning gate oxides [15, 34].

For Fe, an empirical correlation between the Fe concentration for gate-oxide test wafers and the yield in a gate oxide test method is shown in Fig. 9.29: For a threshold level of 5×10^{11} cm^{-2} surface concentration, the gate-oxide yield starts to degrade, and the variability of the results increases simultaneously [31]. Since empirical correlations have shown that the results of such tests correlate with the burn-in failure rate for DRAMs [23], this contamination level is absolutely critical for the production of CMOS and BiCMOS devices with a large gate-oxide area. In particular this is true for DRAM production since DRAMs have the largest gate-oxide area due to the high packing density of transistors in the memory cells arrays.

The degradation mechanism for Fe has been identified as thinning of gate oxide by $FeSi_2$ precipitates [6], Fig. 9.30a and b demonstrate that similar mechanisms are also operative for copper. In Fig. 9.30a an SEM picture of a gate-oxide surface shows a mushroom like protrusion, where something has obviously lifted the gate oxide from underneath. A TEM study with energy dispersive X-ray analysis on the same type of wafer demonstrated that the "mushrooming" of the gate oxide was caused by a Cu precipitate under the gate oxide, which led a local thinning of the gate oxide, as seen in the cross sectional TEM picture in Fig. 9.30b.

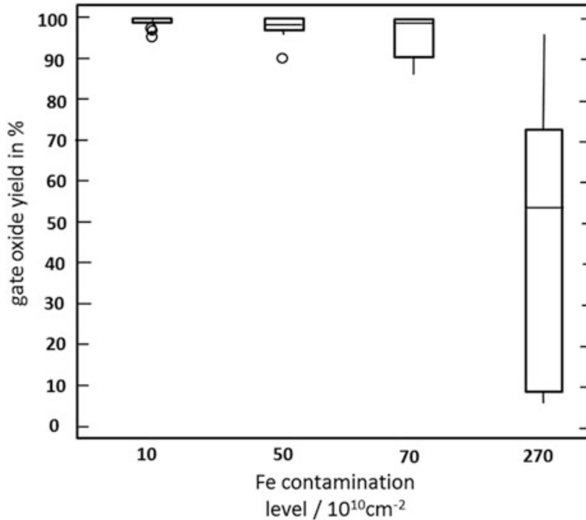

Fig. 9.29 Correlation between the surface Fe contamination of gate oxide test wafers and the yield of a standard gate oxide quality test on special test wafers

The detrimental effect of local thinning of the gate oxide on the gate oxide breakdown can be explained in the following straightforward manner: The normal operating field strength across the gate oxide is usually about 5MV/cm, which can be calculated from the applied voltage, e.g. 5 V and the average gate oxide thickness, e.g. 10 nm. If above the Cu silicide, the thickness is only 7.5 nm, the field strength is 7.5 MV/cm, which is close to the Fowler Nordheim tunneling threshold at 8–9 V [34]. Due to the excessive injection of charge carriers into the gate oxide at the thinned location, the normal wear-out of the gate oxide is dramatically accelerated.

In order to investigate what level of contamination is tolerable for a given contamination element, data have to be gathered for a correlation of the metal impurity concentration and a gate oxide integrity test for a particular device process. The critical threshold can indeed depend on the particular device process and device type, and the wafer type. Fortunately, as a rule of thumb, degradation thresholds for contamination levels are fairly consistently around concentrations of 10^{11} cm^{-3}.

9.6.2 Extended Defect Formation by Thermal or Layer Stress or Silicon Interstitial Super Saturation

The phenomenon of OISF formation by oxidation has already been mentioned above. Stacking faults can also form via a similar mechanism after ion implantation, since ion implantation creates a large number of intrinsic defects, so in both cases the super saturation of mainly self-interstitials results in the formation of extrinsic

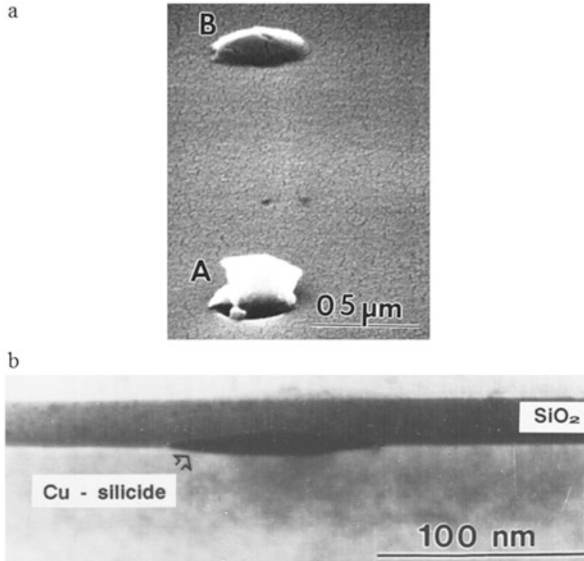

Fig. 9.30 (**a**) SEM picture of a gate oxide surface on a silicon wafer which has been intentionally contaminated. At both sites *A* and *B* the gate-oxide is lifted by a Cu precipitate underneath the oxide. These are particularly drastic examples. The precipitate in cross section Fig. 9.30b hardly lifts the gate-oxide (but thins it!). (**b**) TEM cross section of a Cu-silicide precipitate (as indicated by the *arrow*). It is clearly visible that the lens-shape silicide precipitate locally thins the gate oxide, which implies an increased field strength, which dramatically accelerates the wear out of the gate oxide by local Fowler-Nordheim tunneling currents which ultimately lead to a local gate oxide breakdown

stacking faults. Similar arguments apply to reactive ion etching, which among other effects results in the formation of crystal defects.

The formation of thermally induced dislocations is visible in Figs. 9.14 and 9.16 where the rapid thermal anneal step of the haze test introduces thermal stress which forms dislocations visible as thin lines in the <100> directions, which are glide lines caused by slip through the glide of dislocation [6]. This phenomenon can occur both in contaminated and in uncontaminated areas and is caused by pure thermal stress, not by the contamination alone (although contamination can facilitate the process). For the geometry of the dislocations, also consider Fig. 9.7.

Figure 9.31 provides more evidence that dislocation formation by thermal stress is not only possible in the untypical high stress situation of a rough rapid thermal anneal process, but also occurs in a regular device process, if the silicon material is and/or the process are not optimized. Dislocation arrays on {111} glide planes are visible In the critical <100> directions, near the wafer flat further dislocation arrays have formed with the additional "assistance" of the damage caused by the laser ID marking of the wafer and/or slight mechanical damage of the wafer edge at the wafer flat (this part of the wafer edge is handled particularly often during the alignment process of the wafer in equipment, and therefore some light

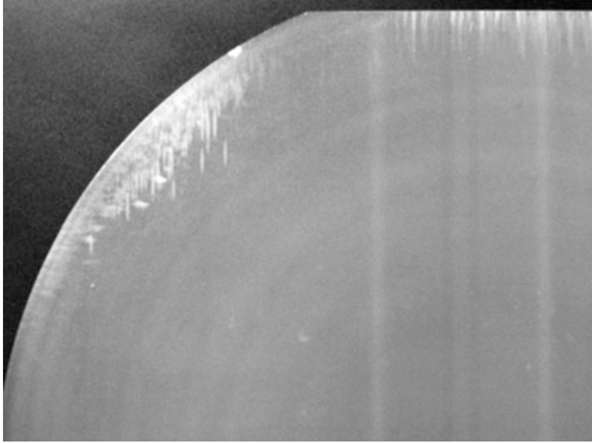

Fig. 9.31 X-ray topograph of part of a 150 mm silicon wafer after a partial regular 4 Mbit device process. The white crisscross lines are due to arrays of dislocations which have formed by excessive thermal stress in the critical <100> directions (at 45° angle with respect to the wafer flat), caused by too high temperature gradients in furnace during move-in or during ramp-up. Additional dislocation arrays on glide planes have formed near the wafer flat, due to the damage from laser marking the wafer ID number and some mechanical damage from handling the wafer in the device process

mechanical damage at the rounded edge is created which lowers the nucleation barrier for dislocation formation). In the 1970s, wafers for microelectronics did not have edge rounding yet, so the edge was very vulnerable to severe mechanical damage, with strong subsequent dislocation formation [4]. The introduction of edge rounding dramatically reduced dislocation formation in integrated circuit processing technology.

An example for dislocations generated by layer stress (rather than thermal stress) is shown in Fig. 9.32, where stress between silicon and silicon nitride (= the mask layer of local oxidation of silicon) has led to the formation of many dislocations at the oxide edges, where the layer stress had its maximum [4]. In the X-ray topograph of a 150 mm production wafer, the dislocations give rise to the white contrast which coincides with some of the structures of the 4 Mbit DRAMs devices, formed during field oxidation. Compared to the X-ray topograph of Fig. 9.31, there are no indications of the dislocations formed by excessive thermal gradients in furnaces.

As repeatedly mentioned, the task of defect engineering is to identify the factors which can prevent the defect formation. Examples for such process measures are:

- Avoiding nucleation centers (like damage at the wafer edge),
- reducing the ramp rates and the move-in temperature
- the addition of an SiO_2 layer (pad oxide) between silicon and silicon nitride as “lubrication” between the two layers with a large difference in the thermal expansion coefficients [4]

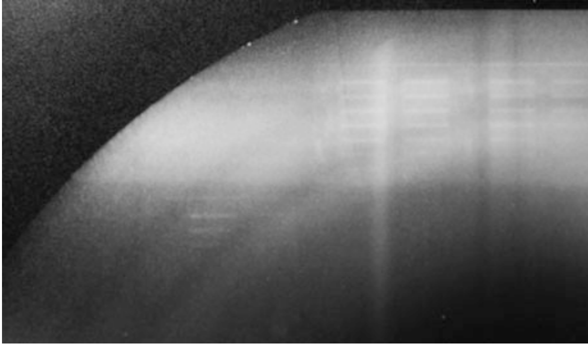

Fig. 9.32 Example for dislocations formed by layer stress, as visible from the fact that the sense amplifier structures from the 4 Mbit DRAM process wafers show up in the X-ray topograph, which implies that in those particular regions of the integrated the layer stress during one of the high temperature steps exceeded the critical yields stress

In more modern devices with design rules below quarter micron, instead of local oxidation of silicon (by the LOCOS process) the lateral insulation of devices has been implemented by etching shallow trenches into the silicon substrate and filling them with deposited oxide. Predictably, these processes also have led to problem of dislocation formation initially, but the problem could be solved e.g. by the appropriate deposition conditions and a suitable temperature time profiles to induce stress relaxation in the deposited material and to keep thermal stresses below the threshold for dislocation formation [4, 35, 36].

9.6.3 *Interaction of Defects After Initial Formation During Subsequent Processes*

It is obvious that a combination of metal contamination and stresses, as indicated in Fig. 9.11, can lead to complex phenomena via different defect mechanism paths over the course of the many consecutive process steps of microelectronics manufacturing, as already illustrated in Fig. 9.12.

An example from a real device process of the path in Fig. 9.12 where some stress, e.g. during oxidation, have led to the formation of a stacking fault, and the subsequent decoration of the stacking fault by contamination not during the actual device process but during the burn-in operation of the DRAM is represented in Fig. 9.33: The retention time fail of a particular DRAM cell in a 64 kB DRAM (after results from [4]) and the subsequent physical failure analysis showed that the failed cell had a stacking fault attached to it, as revealed by Secco defect etching. The remarkable fact is that during the initial test on a wafer at the end of the device process, that particular memory cell had shown no problems, in other words the stacking fault had formed but was not decorated by metal contamination [4]. After

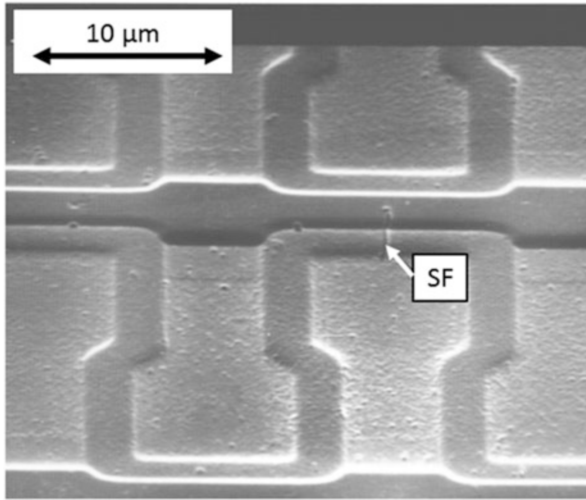

Fig. 9.33 Stacking fault in the memory cell array of a 64 kbit DRAM. The cell had been functional immediately after processing, but developed a retention time fail due to excessive leakage currents after the burn-in (i.e. operation at higher than normal operating temperature [up to 150 °C] and at an overvoltage compared to the normal operating voltage, to accelerate aging by orders of magnitude. To subject the DRAMS to burn-in is a standard procedure before they are sent to customers)

several hours of burn-in testing, however (i.e. the equivalent of months to years of operation), the cell started to degrade until the leakage current caused by the stacking fault got so large that the charge stored in the memory cell capacitor leaked before the charge was replaced in the memory refresh cycle for a DRAM. The only reasonable explanation of this finding is that during operation, impurity atoms like Fe, Ni, Co, Cu who all can diffuse even at room temperature gradually decorated the stacking fault more and more, leading to a higher and higher leakage current for the affected cell. The ability for diffusion at room temperature can be concluded from an extrapolation of the diffusion coefficient to room temperature, it is also obvious from the fact that the SPV method for the determination of Fe concentrations in Si works at all [32]: The dissociation of Fe B pairs around 150 °C implies that the interstitial Fe atoms have to move away from the Boron atoms they had formed pairs with. In fact, the time constant for re-pairing is consistent with the extrapolation of the diffusion coefficient to room temperature [37]. This example also shows directly that the complex interaction of extended defects and metal impurities is not only a yield detractor, but also a reliability and durability problem, which is the worst and often most difficult problem in defect engineering.

Figure 9.34 provides further direct evidence of the detrimental electrical effects of stacking faults by causing leakage current via the joint electrical effect of metal contamination and stacking faults is illustrated:

By a spatial correlation of stacking fault densities and the measured leakage current in the power MOS devices in one whole wafer it could be established that the

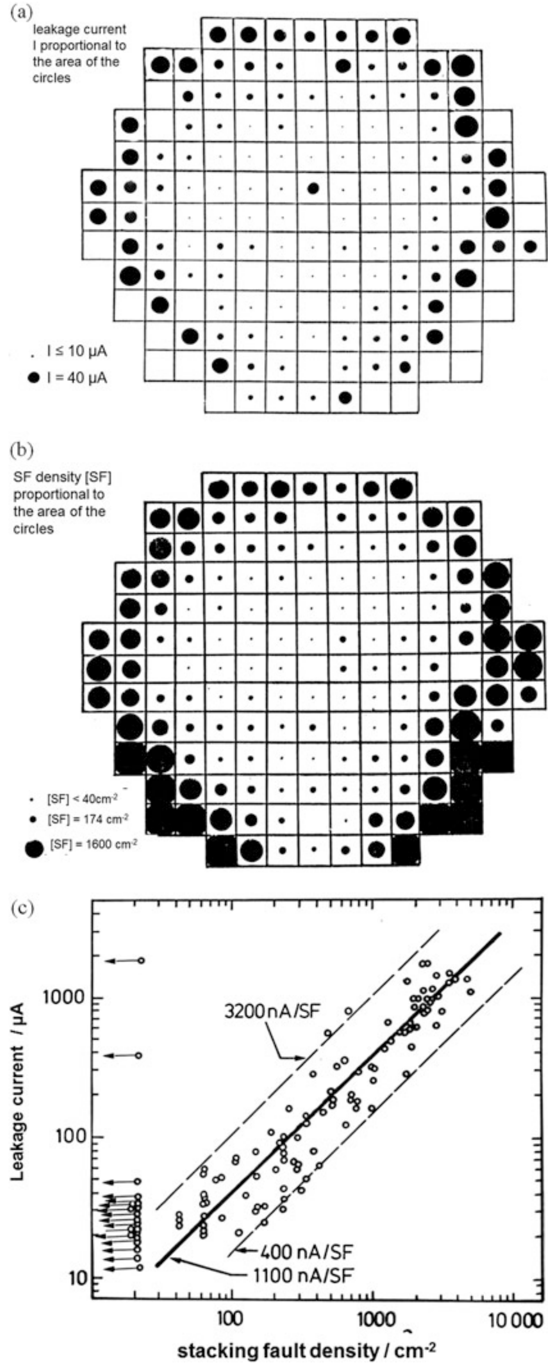

Fig. 9.34 Comparison of the spatial distribution of leakage currents of MOS power devices (**a**), stacking fault densities of (**b**), and a correlation between them number of stacking faults in one devices and its leakage current (**c**)

leakage current caused by one stacking fault was approximately 1 μA. This is orders of magnitude more than for uncontaminated stacking faults: General experience shows that clean stacking faults will cause leakage currents of the order of picoamps or a few nanoamps. The observed comparatively high values and first and foremost, the scatter are both a strong indicator for decoration of the stacking faults by metal contamination.

Another detrimental interaction of metal contamination and extended defects has already been shown in Fig. 9.18, where the formation of stacking faults during oxidation is dramatically enhanced if before the oxidation step, the metal precipitates are formed in a preceding high temperature step. The precipitates subsequently lower the nucleation barrier for stacking fault nucleation by a significant amount for all the following process steps.

It is thus clear that metal contamination has a two-fold risk potential for microelectronics production:

- the enhancement of nucleation of extended lattice defects (stacking faults, dislocations) by lowering the nucleation barrier
- The subsequent decoration by metal impurities to render these defects highly electrically active

In other words, this particular yield detraction path proceeds in three steps:

Metal contamination and precipitation → formation of extended defects → decoration of those defects by metals

To further illustrate this important mechanism, Fig. 9.35 demonstrates that the leakage current from stacking faults increases with the Fe impurity concentration.

The interactions during a complete device process which can consist of up to 1000 individual process steps can be quite complex. Another fairly common and typical example is the interaction between a super saturation of self-interstitials from oxidation with residual implantation damage, as shown in Fig. 9.36. The implantation of boron enhances the formation of stacking-fault-like defects in a subsequent well oxidation step, in this case the residual damage from boron implantation lowers the nucleation barrier for stacking fault formation.

9.6.4 Monitoring and Process Control During Mass Production on the Product Level

As mentioned in the previous section, metal contamination is a considerable and ubiquitous risk for yield and reliability issues in microelectronics production. Therefore, additional monitoring not only on the equipment level, but on the product level is appears certainly advisable, since this monitors all process steps and all potential defect scenarios. It will be demonstrated in this section, which introduce two additional aspects not previously mentioned, that this is not only advisable, but clearly necessary:

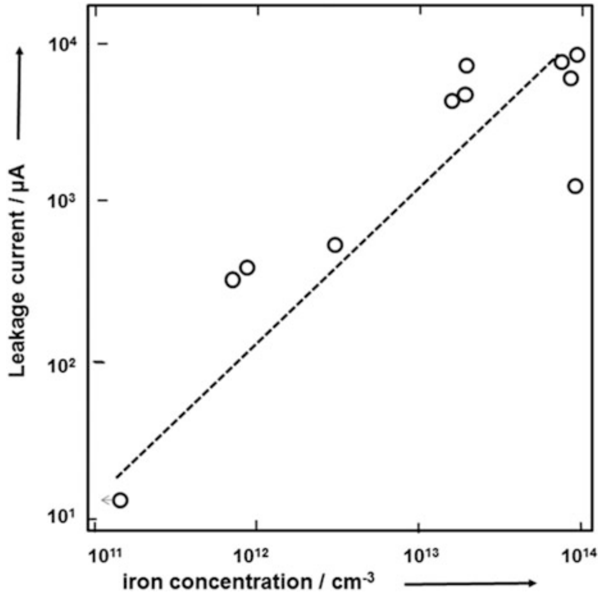

Fig. 9.35 Leakage current caused by one stacking fault as a function of the average Fe concentration in the silicon material

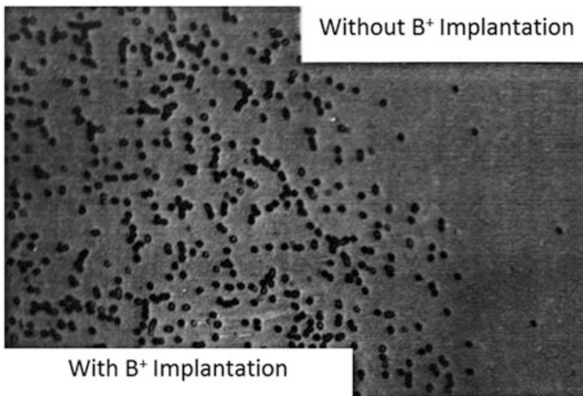

Fig. 9.36 Stacking-fault-like defects formed after boron implantation and subsequent oxidation. The area which has not been exposed to any implantation is free of defects. The residual damage from the boron implantation has caused the formation of defects during oxidation, by coalescence of excess self-interstitials

- The emergence of three-dimensional rather than planar device structures introduce additional failure paths not captured in the failure path map of Fig. 9.11, so a screening on the product levels is not only advisable but indispensable, since this defect mechanism cannot be detected by exclusive equipment monitoring.

- Control of defect formation and metal contamination on the product level can be done in a very efficient and cost effective manner via the electrical test results at the end of the production process, if it is known which of the tested performance parameters are especially sensitive to metal contamination and defects. This has the advantage that electrical tests which are performed on a routine basis for 100 % of the product can be used to detect any significant metal contamination and/or crystal defect formation that has escaped the equipment screening process. The reason that this is a "fail-safe" way to check for metal contamination and/or crystal defect formation is that such electrical testing is done for 100 % of the integrated circuits produced, so it is impossible that such contamination or crystal defect formation events can remain undetected. Also, since the electrical tests are performed after the complete production process, the test results it will represent effects of the accumulated contamination and crystal defects, including their interactions.

To elucidate these aspects, we consider the example of trench cells for 4 Mbit DRAMS: With the introduction of trench capacitors (which look more like a bore hole than an elongated trench, see e.g. Fig. 9.37a), it was noticed that a small number of cells were affected by dislocations which had developed in the immediate vicinity of the electrically defective memory cell, due to the stress concentrations in the corners of the 3D-structure [35, 36]. Figure 9.37a shows the etch figure of such a trench-induced dislocation. To unambiguously prove that the observed electrical retention time fails were due to such trench-induced dislocations, the number of dislocations revealed by defect etching in a complete 256 Kbit memory cell block were counted and compared to the number of retention time fails in that block. Figure 9.37b shows that there is indeed a very good correlation.

The production lots from the initial phases of process development had a high enough number of such defects to generate such statistical evidence that the trench-induced dislocations were indeed the cause for the retention time fail cells. In routine full-scale production, the situation is very different: There will be very few such retention time fail cells in one 4 Mbit DRAM memory, leave alone in one of the sixteen 256 k cell blocks. In routine monitoring, the occurrence of such very few single cells with retention time fails will be analyzed statistically per production lot (which comprises typically 10,000–20,000 DRAMs), and if any increase beyond the normal statistical variation of the number of fail cells per production lot is observed, physical failure analysis on that particular cell will be employed to find out whether trench-induced dislocations or other crystal defects are responsible for the retention time fail through increased leakage currents.

Another common procedure to monitor whether any abnormal contamination and/or crystal defect levels have occurred by electrical test results is to correlate the electrical test results to the "usual suspects" in terms of process steps or equipment, i.e. to see whether there e.g. a significant difference in the level of electrical failure classes between he different essentially identical pieces of equipment used for the same process. The idea behind this analysis strategy is, that if a particular process step is known to potentially cause that failure mode, then there is a high probability

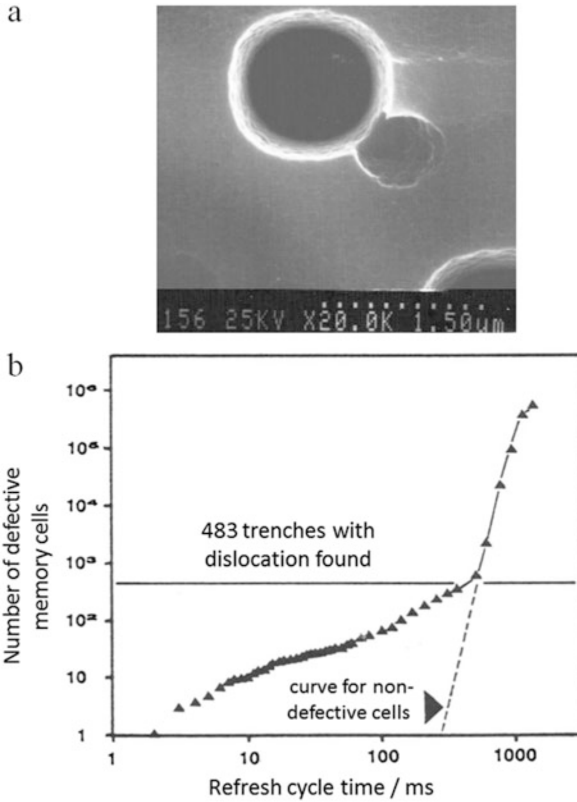

Fig. 9.37 (**a**) SEM micrograph of the etch pit of a trench-induced dislocation, as delineated by Secco defect etching. The round holes are the dry etched holes for trench capacitors, the "irregular" etch pits near the trench delineate the dislocation. (**b**) Plot of the cumulative statistics of the refresh time of the 256 K cells of a 256 K cell block. The portion of the *curve* with the *shallow slope* is representative of defective cells with too short a refresh time, the steep part of the *curve* is representative of the natural distribution of refresh times for cells without any defects, i.e. with design related leakage paths rather than leakage caused by a decorated dislocation. The number of cells in the defect branch tallies very well with the count of trench induced dislocations [35, 36]

that the extent of the failure will differ from machine to machine. The differences can be expected due to the complex nature of defect formation. In addition, there can be different equipment states that influence the defect/contamination level, which can be due to either the state of maintenance or the usage for different processes other than the suspect process. Figure 9.38 shows an example.

Similar principles to trace electrical performance problem due to either silicon wafer issues or to processing problems/equipment issues during the photovoltaic cell manufacturing process are being used on PV mass production. A particularly instructive example of how SPC can help to stabilize a process and improve the manufacturing process has been reported in the PhD Thesis of Dinkel [16]. He could

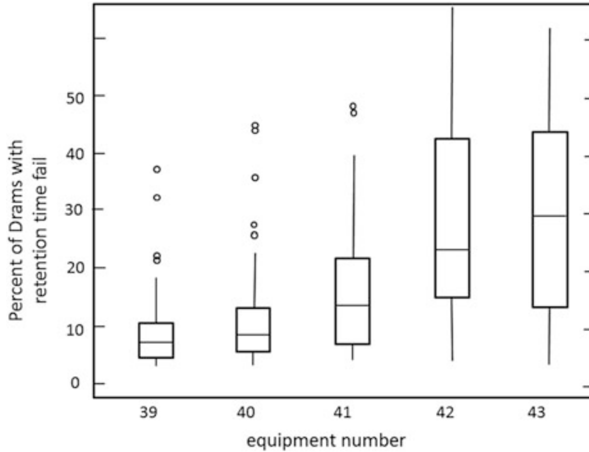

Fig. 9.38 Box plot of the probability of finding at least one retention time-affected cell in a 4 Mbit DRAM, as sorted by the oxide etching equipment (five identical machines were used for the particular process step). It is obvious that there are significant differences, and that the equipment 39 is the best, and equipments 42/43 are about a factor of four worse. It was possible, by contamination studies, to identify the root cause: The higher incidence of retention time fails was due to a Molybdenum contamination, after removal of the source for that contamination, all machines were on the level of 39, i.e. the residual occurrence of retention time fails was now mainly due to other reasons

show that by the application of SPC, the variability in the most important process parameters in cell production was reduced by a factor of 2, and that this led to a much faster improvement in the average efficiency of the cells compared to a reference production line with business as usual, compare Fig. 9.39a and b.

Although this study in the photovoltaics industry has been focused not primarily on metal contamination and defect formation, it is representative for any efforts to improve processes. A more stable process will in itself, as a side effect, also improve that situation regarding crystal defects and contamination. It will, in particular allow in a much better way to identify significant contamination or defect sources, by the "equipment benchmarking" method demonstrated in Fig. 9.38.

The recently discovered potential-induced degradation (PID) phenomenon [38] has not been understood yet, although many different hypotheses have been put forward [39]. What emerges at this stage, that it appears highly likely, that mechanisms and principles of defect formation that have been identified over decades of research in the microelectronics industry also apply here [17, 38]. It has been shown that there is a strong similarity to the mechanism describe in Sect. 9.6.2, in which dislocations are generated by the layer stress caused by the different thermal expansion coefficients of silicon and silicon nitride. Exactly that situation is found in most photovoltaic cells: The silicon nitride antireflective coating is directly on top of the silicon wafers. As mentioned in Sect. 9.6.2, the solution to this problem was the introduction of a silicon dioxide buffer layer between the silicon nitride

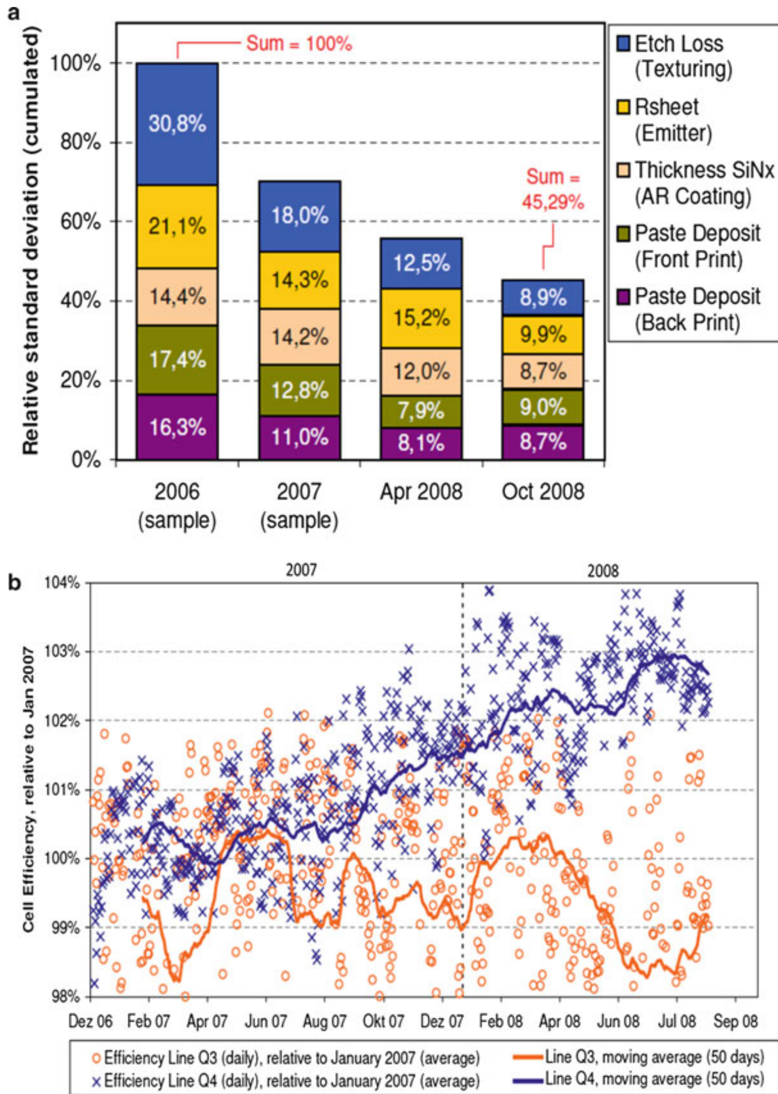

Fig. 9.39 (**a**) In photovoltaic cell production the five most important parameters which cause variability in the energy conversion efficiency of the PV cells are the etch loss during texturing the as-sawn raw wafer, the sheet resistance of the emitter, the thickness of the antireflective silicon nitride coating and the silver paste deposit steps for both the front side and the back side metallization. By the introduction of SPC, the variability in these key control characteristics could be reduced by more than a factor of two. It should be noted that another key control characteristic is the minority carrier lifetime of the silicon wafer, this was not the scope of the work [16]. (**b**) Measured cell efficiency (daily average of the complete production volume in two different production lines Q3 and Q4, which contained practically identical equipment). Line Q3: without SPC, Line Q4: with SPC. It is clear that a systematic and continuous improvement of the efficiency was only possible for reduced variability of the key control characteristics, as listed in Fig. 9.39a. It is not directly visible that among other things, the defect level was positively affected by the process improvements, hence an increase in the efficiency was achieved [16]

and the silicon, which served as a kind of lubricant to mitigate the layer stress. If for photovoltaic cells which are prone to the PID effect, such a buffer layer is introduced, the PID effect is absent. It appears therefore likely that one important root cause of the potential induced degradation is the formation of dislocations and stacking faults by the layer stress between silicon and silicon nitride.

Returning to the microelectronics industry, there is the severe problem that Fe, Co, Ni and Cu can even diffuse at room temperature, so that a gradual decoration of the extended lattice defects during device operation is possible, as demonstrated by the examples described earlier (see also [4]). The general conclusion is that a high additional risk for early failure after 10–1000 h of operation, i.e. reliability and durability risks, can be caused by these failure mechanisms. In other words, the "synergetic action" of metal contaminants and crystal defect will lead to the worst problem for any manufacturer, namely reliability issues, and therefore preventive countermeasures are mandatory. Incidentally, there is strong evidence [39] that the PID effect in PV, which is a reliability issue, is also associated with the decoration of the partial dislocations bounding stacking faults.

Since it is next to impossible under mass production environments to completely avoid metal contamination, in addition to monitoring an minimizing metal contamination and the driving forces for crystal defect formation, an additional strategy is normally implemented, which is similar to the ideas behind taking out an insurance policy: Extra process measures are taken, which will mitigate or suppress the detrimental effects of any residual metal contamination. Such process measures are called gettering, the different gettering types that have been invented and implemented are described in the following Sect. 9.6.5.

9.6.5 Gettering of Metals by Intentionally Introduced Defects

As mentioned in the previous section, metal contamination is a considerable risk for yield and the reliability of products in microelectronics production. In real life, such small but unavoidable risks with a large damage potential are mitigated by taking out an insurance policy. Exactly the same principle applies in microelectronics: Several gettering schemes have been developed over the years, an overview inspired by a figure in Chap. 13 [7] is shown in Fig. 9.40.

The two most frequently used gettering schemes used in the microelectronics industry involve the intentional formation of extended defects, namely

1. backside damage gettering (extended defects on the wafer back surface are provoked by slight controlled mechanical damage by proprietary, sometimes accompanied by a low temperature oxide LTO) and
2. intrinsic gettering, which involves the intentional and controlled formation of oxygen precipitates and bulk microdefects (stacking faults) in the bulk of the wafer, but not at the surface. This is illustrated in Fig. 9.41.

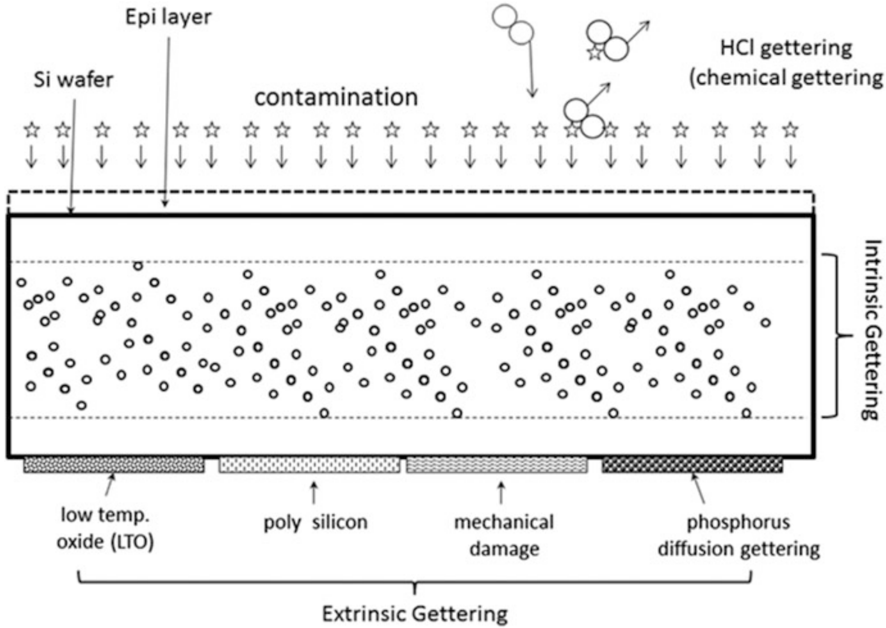

Fig. 9.40 Overview of different gettering schemes (after Shimura [7], Fig. 14, page 595) to reduce metal contamination and reduce the associated risks. Chemical gettering by chlorine containing gases in furnaces tries to remove the metal before they can even enter the silicon wafer, since chlorides of the important metals are volatile. Once the metals are in the silicon, they have to be kept away from residing in the top surface layer where the active device areas are. This is either accomplished through various intentional damage creation schemes on the wafer back surface or it is using the natural tendency in CZ-silicon for oxygen precipitates to form in the bulk, but not near the surface. In all cases, the tendency of metal contamination to aggregate at defects is used to keep them away from active device areas. Naturally this technique cannot be used in power devices if the whole wafer is active in the device operation. Not directly shown in this graph: Gettering by solubility enhancement via high doping levels

Like with any process, some control has to be established that the intended "insurance policy" is really effective. This can be either, through only the monitoring of the intentional defect density, or better by correlation of the gettering type and intensity with the electrical results on devices.

Figure 9.42 is an example of the latter approach: The density bulk micro defects (BMDs) which are generated as a result of oxygen precipitation (oxygen precipitates, stacking faults and dislocation all three defect types are visible via their respective etch artefacts in Fig. 9.41). Below a density of 10^3 cm^{-2}, there is no appreciable positive effect on the yield of gate oxide test structures (the BMD density is measured as an area density, as visible on the etched surface, from the thickness of the etched layers the volume density can be calculated, but it is not needed for such a correlation). Between 10^3 and 10^4 cm^{-2} there is a transition

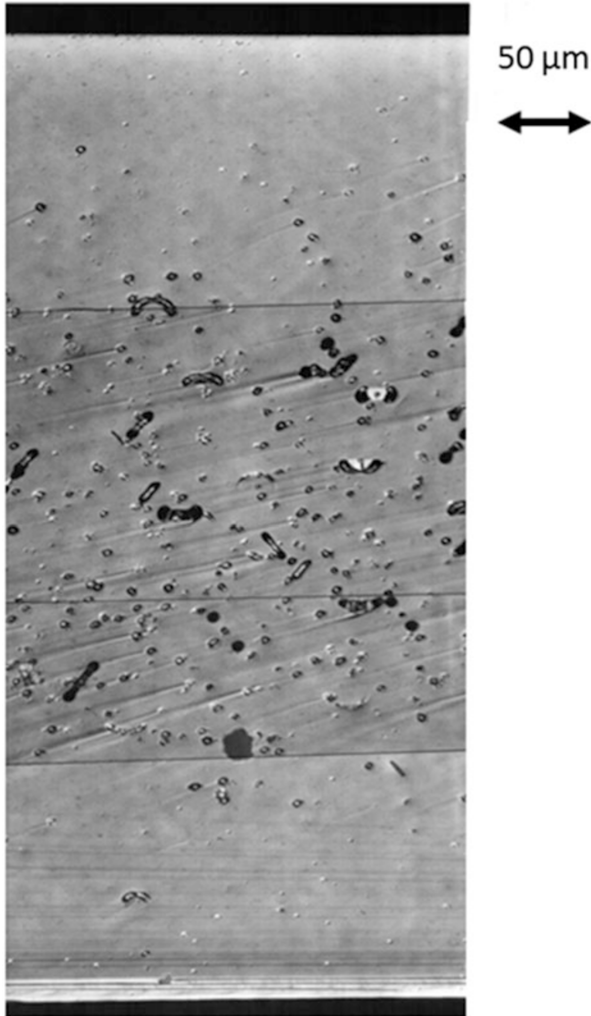

Fig. 9.41 Cross section of a wafer from the front to the back surface, which shows that the denuded zone which also forms also on the back surface. It is also visible that there can be small residual defects in the denuded zone

zone, and for higher densities, the gate-oxide-yield approaches 100 %. The residual variation is due to the variability of the impurity concentration.

A slightly different approach to testing the effectiveness of intrinsic gettering using spatial correlations has been used in an experiments by Falster and Bergholz [33], in which the most common impurities Fe, Ni, Cu, Co and in addition for Pd have shown the intrinsic gettering action of oxygen precipitates directly. To this end, wafers were prepared which were oxidized once or had received a partial or a full CMOS heat process (without the actual device process, with respect to the oxygen

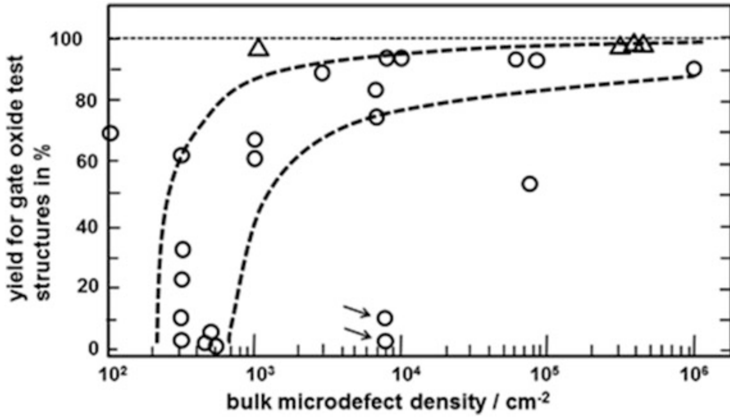

Fig. 9.42 Positive effect of intrinsic gettering. The bulk microdefects (BMDs) can prevent the detrimental effect of a significant metal contamination on the gate oxide integrity test structures, provided the density is higher than about 10^4 cm^{-2}. The *arrowed* points are from lots with more heavy metal contamination

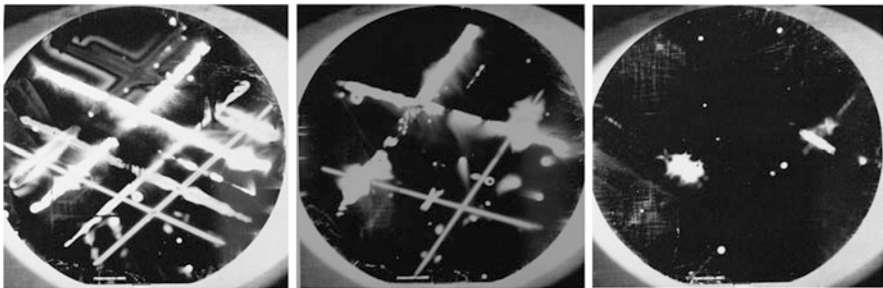

Fig. 9.43 Haze test on wafers contaminated in a similar manner on the wafer back surface as shown in Fig. 9.24. The leftmost wafer has received an initial oxidation after crystal growth and wafer manufacturing (i.e. negligible oxygen precipitation), the center wafer has received a partial CMOS heat process simulation (i.e. moderate oxygen precipitation), the wafer on the right has received a full CMOS process simulation so that gettering is strong. Also visible: significant dislocation formation in the latter wafer which indicates a softening of the material, too soft for the harsh haze test. All wafers had been intentionally contaminated by Fe, Ni, Cu, Co and Pd as indicated in Fig. 9.24

precipitation, the partial or full CMOS thermal simulation is equivalent to the actual device process). Subsequently the three types of wafers with four levels of "built-in" intrinsic gettering were contaminated in the same manner as shown in Fig. 9.24. From Fig. 9.43 it can be seen that for the wafer that had only received one oxidation, there is little gettering visible in this haze test, whereas wafers with the partial and full CMOS heat process simulation show strong gettering via the reduction of the incidence of haze on the front surface: The impurities have been gettered by the bulk microdefects before they could precipitate near the surface.

In addition to defect-induced gettering via oxygen precipitates in the bulk or intentional damage at the wafer back surface, gettering by heavily doped layers is also commonly used, as already mentioned before, this gettering type utilizes the solubility enhancement by pairing of charged impurities with dopant atoms, by the introduction of additional species for the impurity in question and last but not least by dynamic effects, such as the Si self-interstitial injection during e.g. phosphorus diffusion, which enhances the formation of SiP precipitates [40]. The increase in the solubility can be several orders of magnitude in total [22], which is certainly one reason that epitaxial wafers on a heavily doped substrate enhance significantly the robustness of CMOS processes against contamination.

In photovoltaics, gettering is also very important, although it was presumably never introduced intentionally, but came as a beneficial side effect of the cell and process design: Phosphorus diffusion gettering is highly effective during the emitter diffusion in cell which use p-type silicon as the starting wafer. Equally important, the gettering of a heavily aluminium-doped back surface layer (from the aluminium paste covered back surface, for reliable electrical contacts) is beyond any doubt contributing significantly to very effective gettering and thus results in a high resilience of the cells to metal contamination. It is well known and a consequence of the need for low cost production, that there are much higher metal contamination levels in PV production than in microelectronics. As already briefly mentioned before, the more modern cell designs (lower doping of the P-doped emitter, selective emitter doping both to improve the conversion efficiency for blue light) and the introduction of dielectric layers on the back surface can be expected to reduce the gettering efficiency. Thus predictably, contamination-related new defect types will come up in mass production of modern cell designs, and will have to be addressed along the principles explained in this chapter. Thus, this additional risk factor has to be taken into account when converting mass production from the simple cells to the more sophisticated cell concepts. It is clear that this will need additional efforts in terms of defect and contamination control.

Also, going by the general characteristics of the potential induced degradation in photovoltaics (PID effect, [17]), and by direct evidence [39] it appears certain that fast diffusing metal contamination is involved in the gradual deterioration of the pn-junction over months, which is typical for the PID effect. In an intentional contamination experiment, Raykov et al. [38] could show however, that Na, Li, K and Ca do not lead to the PID effect, if they are deposited on top of Si cell at an intermediate process stage. The role of ubiquitous metals such as Fe, Ni, Cu will have to be studied in order to clarify their potential role in the PID effect.

Many researchers have demonstrated the effectiveness of gettering directly in production. One particularly instructive example has been published by Jastrzebski et al. [41]. He could show that the yield of bipolar devices drops in a period of high contamination both with and without intrinsic gettering. However, the drop is much less with intrinsic gettering, so although the effect of metal contamination is not completely suppressed, it is mitigated to such a degree that it is close to the "natural" process yield variations.

Another excellent example to demonstrate the effectiveness of intrinsic gettering has been published by Hourai et al. [27]. The dramatic enhancement of OISF formation by metal contamination has already been shown in Fig. 9.18. In the same series of experiments the authors could show how intrinsic gettering can counteract the enhancement of OISF formation by metal contamination.

Traditionally, the oxygen content of silicon wafers has been regarded as the main parameter to be adjusted to a particular device manufacturing process, depending on the thermal profile of that process. From the work of Hourai [27] it is obvious that a minimum amount of oxygen has to precipitate, in "traditional" silicon wafers, a minimum oxygen content is needed to ensure this. So too low an oxygen content is not suitable to ensure the effectiveness of intrinsic gettering, but too high an oxygen content can be detrimental: Excessive oxygen precipitation will lead to warp the wafer after the production process by plastic deformation, i.e. the high large oxygen precipitates reduce the critical shear significantly.

A method pioneered by Falster et al. [42] that consists of a suitable thermal treatment with specific gas atmospheres, enables the manipulation of the oxygen precipitate nuclei in such a manner that the oxygen precipitation can be made almost independent of the oxygen content and the thermal history during crystal growth. This is achieved through erasing all existing oxygen precipitate nuclei (which are at an embryonic stage) by a rapid thermal anneal step and then using a second high proprietary temperature treatment in a suitable ambient to form nuclei in a preset density with a well-controlled denuded zone. To achieve this, a detailed understanding of defect dynamics was needed, the description of this process is beyond the scope of this chapter.

While this subsection has explained the fundamentals of intrinsic gettering without too much regard for the latest state-of-the-art microelectronic processes, in Chap. 6 Kissinger describes in detail oxygen precipitation, the delineation of oxygen-related defects by defect etches, and the impact of modern CMOS process with drastically reduced thermal budgets.

9.7 Summary and Conclusions

The dual task of defect engineering is to prevent the formation of crystal defects and to intentionally create defects for gettering in a controlled fashion.

The reduction of the minimum feature size in microelectronics and the introduction of new device designs and processes in photovoltaics have to be accompanied by vigilance regarding new contamination sources and/or new mechanisms for the formation of extended defects, such as the onset of dislocation formation at corners in 3D-device structures.

Over many past device and wafer generations, defect formation has been brought under control by weakening the "driving forces" (i.e. reduction of thermal/thermal stress and avoiding too high supersaturation of point defects) and strengthening the "impeding forces" (i.e. elimination of nucleation centers and reduction of

processing temperatures), in spite of the emergence of new defect mechanisms emerging as the technologies progressed. These principles have held over 40 years of microelectronic manufacturing. Control of defects is achieved on the one hand by a thorough scientific understanding of the very often complex defect formation and device degradation mechanisms; in addition it is necessary to employ quality engineering principles and quality management tools (such as SPC) to reliably control the defect scenarios.

The successful applicability of the general principles of defect engineering to photovoltaics and novel types of silicon wafers (perfect silicon, silicon on insulators and other innovative wafer types) can be confidently predicted. As mentioned in the last sections, there is strong evidence that the recently discovered PID effect is due to defect mechanisms in conjunction with metal impurities in the bulk of, or near the pn-junction of silicon-based solar cells, and that the formation mechanism is very similar to the formation mechanism of dislocations via layer stress in microelectronics. The analogy even includes that the solution to the problem, namely the introduction of a silicon dioxide buffer layer between silicon and silicon nitride. Moreover it is clear that with the advent of more sophisticated cell concepts (such as the PERC cell) the susceptibility to contamination and/or defect formation will increase.

It has been the overall objective of this chapter to “build a bridge” between the science behind defects in silicon and how to apply this knowledge effectively in R&D in the microelectronics, micromechanical and photovoltaic industries. In this endeavor, it was necessary to introduce the essentials of process and quality management and reliability engineering. We hope that electrical engineering and information technology students will be capable to apply materials and defect sciences effectively in their future assignments.

References

1. Falster, R., Voronkov, V.V., Quast, F.: On the properties of intrinsic defects in silicon: a perspective from crystal growth and wafer processing. Phys. Stat. Sol (B) **222**, 219 (2000)
2. SEMI M1 Standard for silicon wafers. http://ams.semi.org/ebusiness/standards/SEMIStandardDetail.aspx?ProductID=1948&DownloadID=3469
3. International Roadmap for Semiconductors ITRS. http://www.itrs.net/Links/2013ITRS/2013Chapters/2013Overview.pdf and International Technical Roadmap for Photovoltaics, http://www.itrpv.net/Reports/Downloads/
4. Kolbesen, B.O., Strunk, H.: VLSI electronics: microstructure science. In: Einspruch, N.G., Huff, H.R. (eds.) Silicon Materials, vol. 12, p. 143. Academic Press, New York (1985)
5. http://www.semi.org/Standards
6. Bergholz, W., Zoth, G., Wendt, H., Sauter, S., Asam, G.: Metal contamination control in silicon VLSI technology: fundamentals. Siemens Forsch- und Entwickl Ber **16**, 241 (1987)
7. Shimura, F.: Oxygen in Silicon. Semiconductor and Semimetal Series, vol. 42. Academic Press, New York (1994)
8. Bergholz, W.: Grown-in and process induced defects. In: Shimura, F. (ed.) Oxygen in Silicon. Academic Press, Boston (1994)

9. http://www.itrpv.net/Reports/Downloads/2015/
10. http://www.oecd-ilibrary.org/energy/world-energy-outlook_20725302;jsessionid=3y9ykh6ljjtq.x-oecd-live-03
11. Bergholz, W., Gilles, W.: Impact of research on defects in silicon on the microelectronics industry. Phys. Stat. Sol (B) **222**, 5 (2000)
12. Obry, M., Bergholz, W., Cerva, H., Kürner, W., Schrems, M., Sachse, J.U., Winkler, R.: The role of metal contamination and crystal defects in quarter micron technology. In: Abe, T., Bullis, W.M., Kobaysashi, S., Lin, W., Wagner, P. (eds.) Proceedings of the 3rd International Symposium Defects in Silicon, p. 133. The Electrochemical Society, Pennington (1999)
13. Kolbesen, B.O., Cerva, H., Gelsdorf, F., Zoth, G., Bergholz, W.: Process-induced defects in silicon VLSI technology. In: Proceedings of the Semicon Europe, Zürich, Mar 1992
14. Bergholz, W., Landsmann, D., Schauberger, P., Schöpperl, B.: Contamination monitoring and control in device fabrication. In: Kolbesen, B.O., Claeys, C., Stallhofer, P., Tardif, F. (eds.) Crystalline Defects and Contamination: Their Impact and Control in Device Manufacturing, p. 69. The Electrochemical Society, Pennington (1993)
15. Bergholz, W., Zoth, G., Gelsdorf, F., Kolbesen, B.: Metal contamination in ULSI technology. In: Bullis, W.M., Gösele, U., Shimura, F. (eds.) Defects in Silicon II, p. 21. The Electrochemical Society, Pennington (1991)
16. Dinkel, T.: Integrated efficiency engineering in solar cell mass production. PhD Thesis, Jacobs University Bremen (2010)
17. Raykov, A.: Potential-induced degradation – a multi-level problem. PhD Thesis, Jacobs University Bremen (2015)
18. Graff, K.: Metal Impurities in Silicon-Device Fabrication. Springer Series in Materials Science (Book 24). Springer, Berlin (1990)
19. Pyztek, T., Keller, P.: The Handbook for Quality Management. Mc GrawHill, New York (2013)
20. Graff, K.: Transition metals in silicon and their gettering behaviour. Mater. Sci. Eng. B **4**, 63–69 (1989)
21. Tuck, B.: Introduction to Diffusion in Semiconductors. IEE Monograph Series, vol. 16. Peter Peregrinus Ltd., on behalf of the Institution of Electrical Engineers, Stevenage (1974)
22. Gilles, D., Schröter, W., Bergholz, W.: Impact of the electronic structure on the solubility and diffusion of 3d transition elements in silicon. Phys. Rev. **B41**, 5770 (1990)
23. Winkler, R., Behnke, G.: Gate oxide quality related to bulk properties and its influence on DRAM device performance. Semicond. Silicon. **94**, 673 (1994)
24. Hannay, N.B.: Bell Lab reports (1958)
25. Pell, E.M.: Ion drift in an n-p junction. J. Appl. Phys. **31**, 291 (1960)
26. Bergholz, W.: Analysis of extended defects. In: Schulz, M. (ed.) Landolt Börnstein Handbook of Physics, Neue Serie 22B, p. 126. Springer, Berlin (1988)
27. Hourai, M., et al.: Behavior of defects induced by metallic impurities on Si (100) surfaces. Jap. J. Appl. Phys. **28**, 2413 (1989)
28. Nikkei Microdevices May 1990
29. Lehmann, V., Föll, H.: Minority corner diffusion length measurements in silicon wafers using a Si-electrolyte contact. J. Electrochem. Soc. **135**, 2831 (1988)
30. Hellmann, D., Rother, M., Hill, M., Bergholz, W., Riedlbauer, M.: Influence of quartzglass on silicon wafers application studies and examples from device production. In: Schidt, D.N. (ed.) Contamination Control and Defect Reduction in Semiconductor Manufacturing III, p. 285D. The Electrochemical Society, Pennington (1994)
31. Bergholz, W., Landsmann, D., Schauberger, P., Wittman, J., Hoffmann, H.: Relevance and effects of metal contamination on device processes and parameters. In: Proceedings of the Technical Conference SEMICON Europa, Geneva, 5 Apr 1995
32. Zoth, G., Bergholz, W.: A fast, preparation-free method to detect iron in silicon. J. Appl. Phys. **67**, 6764 (1990)
33. Falster, R., Bergholz, W.: The gettering of transition metals by oxygen-related defects in silicon. J. Electrochem. Soc. **137**, 1548 (1990)

34. Bergholz, W., Mohr, W., Drewes, W., Wendt, H.: Defect-related gate oxide breakdown. Mater. Sci. Eng. **B4**, 359 (1989)
35. Dellith, M., Gelsdorf, F., Bergholz, W., Booker, G.R., Kolbesen, B.O.: TEM, etching studies of fabrication-induced defects in 4 M DRAMs. Inst. Phys. Conf. Ser. **117**, 169 (1991)
36. Dellith, M., Booker, G.R., Kolbesen, B.O., Bergholz, W., Gelsdorf, F.: On the formation of trench-induced dislocations in dynamic random access memories (DRAMs). Inst. Phys. Conf. Ser. **134**, 235 (1993)
37. Zoth, G., Bergholz, W.: Metal contamination control by diffusion length measurements – principles and practice. In: Proceedings of the Technical Conference Productronica, Munich, Nov 1997
38. Raykov, A., Hahn, H., Stegemann, K.-H., Kutzer, M., Storbeck, O., Neuhaus, H., Bergholz, W.: Towards a Root Cause Model for the Potential-Induced Degradation in Crystalline Silicon Photovoltaic Cells and Modules. PV EUSEC, Paris, p. 2998 (2013); Raykov, A., Stegemann, K.-H., Hahn, H., Bitnar, B., Kutzer, M., Neuhaus, H., Bergholz, W.: On the PID inhomogeneities. Presented at the EU PVSEC, Amsterdam (2014)
39. Naumann, V., Lausch, D., Hähnel, A., Bauer, J., Breitenstein, O., Graff, A., Werner, M., Swatek, S., Großer, S., Bagdahn, J., Hagendorf, C.: Explanation of potential-induced degradation of the shunting type by Na decoration of stacking faults in Si solar cells. Sol. Energy Mater. Sol. Cells **120**(Part A), 383–389 (2014)
40. Schröter, W., Kühnapfel, R.: Model describing phosphorus diffusion gettering of transition elements in silicon. Appl. Phys. Lett. **56**, 2207 (1990)
41. Jastrzebski, L., Soydan, R., McGinn, J., Kleppinger, R., Blumenfeld, M., Gillespie, G., Armour, N., Goldsmith, B., Henry, W., Vecrumba, S.: A comparison of internal gettering during bipolar, CMOS, and CCD (high, medium, low temperature) processes. J. Electrochem. Soc. **134**, 1018–1025 (1987)
42. Falster, R., Voronkov, V.V.: The engineering of intrinsic point defects in silicon wafers and crystals. Mater. Sci. Eng. B **73**, 87 (2000)