Point defects and dopant diffusion in silicon

P. M. Fahey,* P. B. Griffin, and J. D. Plummer

Integrated Circuits Laboratory, Stanford University, Stanford, California 94305

Diffusion in silicon of elements from columns III and V of the Periodic Table is reviewed in theory and experiment. The emphasis is on the interactions of these substitutional dopants with point defects (vacancies and interstitials) as part of their diffusion mechanisms. The goal of this paper is to unify available experimental observations within the framework of a set of physical models that can be utilized in computer simulations to predict diffusion processes in silicon. The authors assess the present state of experimental data for basic parameters such as point-defect diffusivities and equilibrium concentrations and address a number of questions regarding the mechanisms of dopant diffusion. They offer illustrative examples of ways that diffusion may be modeled in one and two dimensions by solving continuity equations for point defects and dopants. Outstanding questions and inadequacies in existing formulations are identified by comparing computer simulations with experimental results. A summary of the progress made in this field in recent years and of directions future research may take is presented.

CONTENTS

*Present address: IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598.

LIST OF SYMBOLS

A_i	dopant atom in interstitial state
A_S	dopant atom in substitutional state
AI	dopant interstitialcy
AV	dopant-vacancy pair
AX	dopant defect, where $X=I$ or V
C_A	concentration of species A
C_A^*	equilibrium concentration of species A
$\langle C_A \rangle$	time-average concentration of species A
C_{AX}, C_X	concentration of defect AX and X not associated with A
C_O, C_{O_2}	atomic and molecular oxygen concentration
C_S	concentration of lattice sites
C_T, C_{IT}	concentration of traps, and interstitial-trap complex
ΔC_X	excess concentration of defect X above equilibrium
D, DV	arbitrary bulk defect and defect-vacancy complex
D_A, D_A^*	diffusivity of A and diffusivity of A under equilibrium conditions
$\langle D_A \rangle$	time-average diffusivity of species A
D_{AI}, D_{AV}	component of dopant diffusivity attributable to an I-type or V-type mechanism
$D_{\text{self}}, D_{\text{self}}^0$	self-diffusivity of silicon and preexponential factor of self-diffusivity
D_{SiI}, D_{SiV}	component of self-diffusivity attributable to an I-type or V-type mechanism
$\mathcal{D}$	diffusion flux of point defects
E_{AX}^b	binding energy of X to A
E_c	energy of conduction band edge
E_f	Fermi level
E_f^i	Fermi level under intrinsic conditions
E_g	band-gap energy
E_{g0}	energy constant of linear fit for temperature dependence of E_g
E_v	energy of valence band edge
E_X	energy level of defect X in band gap
ΔE_{AV}^3	energy difference between an isolated vacancy very far away from dopant atom A and a vacancy a third-neighbor site away from A
ΔE_{Coul}	Coulombic potential energy between point charges
ΔE_{X^-}	$(E_c - E_{X^-})$
$\Delta E_{X^=}$	$(E_c - E_{X^=})$
ΔE_{X^+}	$(E_{X^+} - E_v)$
$\Delta E_{X^{++}}$	$(E_{X^{++}} - E_v)$
G_X^f	Gibbs free energy of formation of defect X
$\mathcal{G}$	generation flux of point defects
H_{AX}^f, H_X^f	enthalpy of formation of AX and X
H_{AX}^m, H_X^f	enthalpy of migration of AX and X
H_g	enthalpy of electron-hole formation across the energy gap
I	silicon interstitialcy or interstitial defect
IT	interstitial-trap complex
J_A, J_{AX}, J_X	flux of A, AX, or X
L_I^b	diffusion length of I before recombination in the bulk
L_I^s	diffusion length of I before recombination at the surface
N_c, N_v	density of states in conduction and valence bands
$\mathrm{P}_s, \mathrm{P}I$	substitutional and interstitial phosphorus
Q_A	activation energy of diffusion for dopant A

Q_{AX} activation energy of diffusion for dopant A diffusing via AX defects
Q_{self} activation energy of self-diffusion
$Q_{\text{Si}X}$ activation energy for the component of self-diffusion occurring through diffusion of X-type point defects
ΔQ_A $(Q_{\text{self}}-Q_A)$
ΔQ_{AX} $(Q_{\text{Si}X}-Q_{AX})$
$\Delta Q_{AV,I}$ $(Q_{AV}-Q_{AI})$
$\Delta Q_{\text{Si}V,I}$ $(Q_{\text{Si}V}-Q_{\text{Si}I})$
$\mathcal{R},\mathcal{R}'$ recombination flux of defects at the injecting and inert surfaces
S_g entropy of electron-hole formation across the energy gap
S_X^f entropy of formation of native point defect X
S_X^m entropy of migration of native point defect X
T temperature
T bulk trap for defects
V lattice vacancy
X general symbol to denote point defect I or V
Z_{AX} net charge of the AX defect
$a_{I,V}$ capture radius for recombination between I and V
$a_{I,AV}$ capture radius for recombination between I and AV
$a_{V,AI}$ capture radius for recombination between V and AI
d_{AX},d_{A_i},d_X diffusivity of AX, A_i, and X defects
d_X^{eff} effective diffusivity of X defect
f_{AI} fraction of dopant diffusivity due to an I-type mechanism under equilibrium conditions
f_{AV} fraction of dopant diffusivity due to a V-type mechanism under equilibrium conditions
g_I^0 constant surface flux of I
g_X surface flux of defect X generated by chemical reactions
h electric field factor of dopant diffusion
h wafer thickness
k Boltzmann's constant
k_I effective bulk recombination rate
$k_{I,V}$ reaction rate constant for bulk generation and recombination of I and V in pairs
$k_{I,AV}$ reaction rate constant for reaction $I+AV\rightleftharpoons A$
$k_{V,AI}$ reaction rate constant for reaction $V+AI\rightleftharpoons A$
n electron concentration
n_i intrinsic carrier concentration
r_A tetrahedral covalent radius of dopant A
r_{Si} tetrahedral covalent radius of Si
r_X surface depletion rate of defect X due to externally imposed chemical reactions
s,s' ratio of bulk loss compared to surface loss for defect X
t time
β proportionality constant of linear fit for band-gap temperature dependence of E_g
δE_g dopant-induced change in E_g
ε scalar value of electric field
θ_{AX},θ_X multiplicity factor of AX and X in statistical calculations
λ normalized depth
μ_{AX} mobility of the AX defect species
$\nu_{I,A}$ equilibrium rate of the reaction $A+I\rightleftharpoons AI$ normalized to C_A
$\nu_{V,AI}$ equilibrium rate of the reaction $V+AI\rightleftharpoons A$ normalized to C_A
σ_X surface recombination velocity for defect X
τ time normalized to a diffusion length
τ_h surface recombination lifetime of I
τ_I,τ_V recombination lifetime of I and V in the bulk
τ_∞ surface lifetime of interstitials for a semi-infinite substrate
φ ratio of surface lifetime in finite and semi-infinite substrates
Ω volume of silicon unit cell

I. INTRODUCTION

Diffusion in solids is a classical field of study of which dopant diffusion in silicon is a subset. It is impossible to appreciate the interest in this research area that has been sustained for over 30 years without understanding the circumstances that have determined its development. The major driving force for the study of diffusion in silicon is the technological importance of dopant diffusion as an elementary process step in the fabrication of silicon-based integrated circuits (IC's). Dopant atoms are the group-V donor impurities P, As, and Sb and the group-III acceptor impurities B, Ga, In, and Al. Dopant atoms are selectively introduced into the silicon substrate during some of the many steps required to make these structures, and their subsequent redistribution by diffusion is almost always inevitable in the processing steps that follow. To be able to predict and control these diffusion steps is obviously important to the design and development of silicon IC's. If dopant diffusion exhibited only simple behavior such as that predicted by a Fick's law formulation of the problem, then we would find in the substrate erfc-type concentration profiles after constant-source depositions from the surface and approximately Gaussian-shaped profiles after ion implantation and diffusion. The prediction of how these profiles diffuse would then be a straightforward task, limited only by the experimental accuracy of dopant diffusivities. That diffusion in silicon is considerably more complicat-

ed than this became apparent during attempts to model diffusion quantitatively in the early 1960's and continues to the present day, as modern device structures require ever more stringent control of dopant distributions and as new technologies to make these IC's arise with unforeseen consequences on dopant diffusion. The unifying idea through all of this has been that since the dopant atoms dissolve substitutionally in the silicon lattice, only by interacting with native point defects, silicon interstitials or vacancies, will the dopant atoms be able to change lattice sites and thereby effect long-range diffusion. Thus, understanding dopant diffusion is fundamentally related to understanding the interactions between point defects and dopant atoms. A few examples will illustrate the nature of the problems that continue to challenge researchers in this field.

In the 1960s the *npn* bipolar transistor was the dominant type of device used in integrated circuits. The performance of this device depends critically on the width and doping level of the *p*-type base region (formed by boron diffusion) sandwiched between the *n*-type collector and emitter (then formed by phosphorus diffusion). It was quickly discovered that the diffusion of phosphorus to form the emitter greatly enhanced the diffusion of the boron base, which had already been diffused into the substrate. This made it very difficult to design devices by any method other than trial and error. It is now known that phosphorus diffusion injects excess point defects as part of its diffusion process (Sec. XIII.C). In the 1970s arsenic quickly began to replace phosphorus as an *n*-type impurity because of its lower diffusivity and better-controlled diffusion behavior. However, with the aid of more advanced tools to measure the concentration profiles, it soon become apparent that the interactions between arsenic and boron were far from negligible. A logical analysis of this system based on the assumption of diffusion under quasiequilibrium conditions (Hu and Schmidt, 1968) provided a sound understanding of this phenomenon and stands today as a landmark model (Sec. X.B and Appendix A).

About the same time as this early success came the recognition of another class of problems, oxidation-enhanced diffusion (or, as it is commonly called, OED), which even today has not been explained by a theory that adequately models experimental results. Enhanced diffusion of dopant atoms occurs when the silicon surface is oxidized, indicating that the oxidation reaction generates excess point defects (now known to be silicon interstitials; see Sec. XIV). This phenomenon, nonequilibrium by nature, is extremely important because oxidation of the silicon surface to grow SiO_2 is one of the most common steps in the fabrication of both bipolar and the now-dominant metal-oxide-semiconductor- (MOS-)type devices. Oxides serve as masks to allow selective incorporation of dopants and dielectric layers in device operation, and also act to passivate the surface. Although it is unlikely that oxidation and SiO_2 layers will be replaced by other materials and processes in the near future, recent investigations into the feasibility of using either thermally grown Si_3N_4 or SiO_2 layers nitrided by exposure to NH_3 ambient as alternative dielectric layers have brought to light the fact that these surface reactions also affect diffusion in the underlying bulk by injecting point defects from the surface.

Because of these generally undesirable and, at present, unpredictable diffusion phenomena, modern techniques for making device structures continue to use lower processing temperatures in order to minimize diffusion, and heavily utilize the process of ion implantation rather than in-diffusion to incorporate dopants into the substrate in a controlled fashion. But it is now well known that radiation damage accompanies ion implantation and that transient enhanced diffusion can result upon annealing of the damage. While lowering processing temperatures does indeed reduce diffusion, it is also the case that "anomalous" enhanced diffusion phenomena become relatively more important (and in some cases, dominant) as the temperature is decreased; presumably because the equilibrium concentrations of point defects decrease more rapidly than other thermally activated processes involved in enhanced dopant diffusion.

All of the preceding examples point out that there is a continuing need to understand diffusion processes in silicon, to improve our predictive capabilities, and to understand quickly (and hopefully control) new phenomena as they are discovered. From our perspective, the most important global requirement that new physical models and their formulations must meet in order to remain relevant to research efforts today is that they be applicable to two- and three-dimensional problems. The need to model diffusion in more than one dimension can only become more important as device geometries continue to decrease and device operation is increasingly determined by the two- and three-dimensional distributions of the doped layers that comprise the device structures. In addition, there are currently no generally available techniques that can measure concentration profiles in more than one dimension. This makes the capability of numerical simulation of diffusion utilizing sound physical models more attractive than ever to engineers trying to create new generations of devices. Unfortunately, almost all diffusion models developed to this point have been derived under the assumption of only one spatial dimension. The extension of these models to higher dimensions, while sometimes trivial, more often than not has demonstrated the inadequacy of their physical basis. This same situation historically has not been so severe for device simulation programs, since in the case of device analysis the governing equations and underlying physics are relatively well understood (for example, the governing equations are generally the current continuity and Poisson's equation). At present, no such basic framework exists for the analogous case of dopant diffusion under nonequilibrium conditions, although the last few years have seen an encouraging effort to achieve just such an end and incorporate a basic set of equations and mod-

els into two- and three-dimensional simulation programs. Our efforts at Stanford have centered on development of a computer program known by its acronym SUPREM (Stanford University PRocess and Engineering Models). The latest version of the program, SUPREM–IV, was conceived at the outset as a program capable of simulating diffusion in two dimensions. We mention this program explicitly here since we make heavy use of the capabilities of the program throughout the paper.

To cover the material mentioned above in the context of motivating the development of a general formulation of point defects and dopant diffusion in silicon, the paper has been organized in the following way. We begin the treatment of point defects in Secs. II and III by defining the meaning of interstitial- and vacancy-type point defects and present some possible configurations in the silicon lattice. In Sec. IV we qualitatively describe possible sources and sinks of point defects, and indicate in which sections of the paper these processes are dealt with in more detail. Sections V and VI present the energy requirements for formation and migration of point defects and summarize the results of attempts to measure these energies experimentally. In contrast to the case of metals, the charge states of the point defects are a very important property, especially when considering the fact that the introduction of dopants can alter the predominant charge state of a given type of point defect (i.e., interstitial or vacancy in nature). This in turn will affect the diffusion rates of the dopants themselves. Much of the remainder of the paper focuses further on the interactions between point defects and dopant atoms. Section VIII is devoted to considering the nature of these interactions and raises the important point that in a doped silicon crystal it is possible for the number of point defects associated with dopant atoms to exceed the number of isolated native point defects. We shall return to this idea several times in the paper and show that this situation cannot be easily ignored in interpreting the experimental results of diffusion studies performed under nonequilibrium conditions. Before discussing the topic of diffusion under nonequilibrium conditions, we first discuss in Secs. IX–XI self-diffusion and dopant diffusion under equilibrium conditions. The relationship between self-diffusion and dopant diffusion provides important experimental information about the mechanisms of diffusion for each. The atomistic processes underlying the vacancy, interstitial, and interstitialcy mechanisms are treated in Sec. XI and related to experimentally determined diffusivities that appear in the formulation of diffusion equations in Sec. X and Appendix A. To extend the well-developed formulation of diffusion under equilibrium conditions to cover the important case of dopant diffusion under nonequilibrium conditions, we begin by considering diffusion of point defects in an undoped silicon crystal in Sec. XII. The importance of the surface boundary conditions in affecting point-defect flow is emphasized. We also consider in detail bulk effects. A more complete examination of the coupling between interstitials and vacancies than has appeared thus far in the literature is presented. In addition, we examine two bulk processes that may also prove to be important in affecting the flow of excess point defects: the trapping of point defects on foreign impurities and defect-assisted recombination. In Sec. XIII we present the logical extension of the formulation in Sec. XII to the case in which the silicon crystal is doped. This section clarifies the assumptions that lead to the commonly accepted formulation of nonequilibrium dopant diffusion when the dopant concentrations are relatively dilute, and identifies what problems exist in the further development of a more general treatment of the problem. The last part of the paper, Secs. XIV–XVII, deals with the experimental attempts to determine some of the most basic and important properties of point defects: their diffusivities, bulk recombination constants, surface generation and loss terms, and, finally, the identification of which mechanism—interstitial-type or vacancy-type—dopant atoms diffuse by. These quantities have proved elusive, despite recent advances such as the possibility of selectively injecting either interstitials or vacancies from the silicon surface through oxidizing or nitriding reactions, the growing body of experimental data measuring the finite transit time of point defects across thin silicon wafers, and related studies that measure the decay lengths of point defects along the silicon surface when different surface coverings are used. Still, some important knowledge and insight has been gained in recent years. A summary of what we believe to be the most important developments and promising areas of future research concludes the paper.

It should be pointed out that there is still a great deal of controversy surrounding many of the basic ideas discussed in this paper. Central to much of this controversy is the question of the relative importance of vacancies and interstitials in dopant diffusion in silicon. Even today there remain strong protagonists of "vacancy-only" and "interstitial-only" points of view. We believe that the weight of evidence today favors an "interstital plus vacancy" point of view, and this is the view taken in this paper. We believe that it is the only viewpoint that is self-consistent when one examines all of the data taken together. The motivation for our work in this field and for closely following the work of others has been a very practical one. We have for many years been attempting to provide the silicon integrated circuit industry with simulation tools that correctly model the IC fabrication process (the SUPREM family of programs). Our views on basic questions such as the interstitial versus vacancy controversy have no doubt been influenced by our engineering background and orientation. However, we strongly feel that engineering programs like SUPREM are only as useful as the accuracy of their models permit, and accuracy ultimately rests on physical correctness.

We have attempted in this review paper to take as unbiased a view as possible of these and other issues. In order to develop a set of engineering tools, we have also assembled the best (in our view) physical models into a

comprehensive overall picture. We shall attempt throughout the paper to make it clear when controversy still exists about conclusions we draw. There is a great deal that remains to be done in this field; it is our hope that this paper will help in providing the overall perspective needed to guide this future work.

II. POINT DEFECTS IN SILICON

Point defects can be separated into two categories: native point defects and impurity-related defects. *Native point defects* exist in the pure silicon lattice. *Impurity-related defects* arise from the introduction of foreign impurities into the silicon lattice. Group-III elements B, Al, Ga, and In and group-V elements P, As, and Sb are a special class of impurities known as *dopants.* Their most important properties in silicon are that they are highly soluble compared to other impurities (with the exception of Ge), dissolve almost exclusively on substitutional sites, and are easily ionized.

If a native point defect is far enough away from a dopant that all of its properties are the same as if the dopant atom were not present, the point defect will be said to be *isolated* or *unassociated.* Otherwise, a point defect that exists in a state such that it is interacting with a dopant atom will be said to be *associated* with the dopant.

III. DEFINITIONS OF POINT DEFECTS

A. Native point defects

There are three native point defects of interest for silicon: the vacancy, interstitial, and interstitialcy. The *vacancy* is defined simply as an empty lattice site and will be designated V. Figure 1(a) shows three examples of how the resultant unsatisfied bonds have reconfigured themselves to accommodate the vacancy defect in the lattice. A *silicon-* or *self-interstitial* is a silicon atom that resides in one of the interstices of the silicon lattice. The dark spheres in Fig. 1(b) indicate the two possible interstitial positions with the highest symmetry. The *silicon-* or *self-interstitialcy* is distinct from the interstitial. An interstitialcy defect consists of two atoms in nonsubstitutional positions configured about a single substitutional lattice site. The term interstitialcy was first used by Seitz (1950). In analogy to the vacancy formed by removing an atom from a lattice site, an interstitialcy is formed by placing an extra atom about a lattice site. Two possibilities are shown in Fig. 1(c), where the two dark spheres represent the silicon atoms that make up the silicon-interstitialcy defect.

Because both are extra silicon atoms, a distinction between the silicon interstitialcy and silicon interstitial often is not made in the literature. Both are commonly

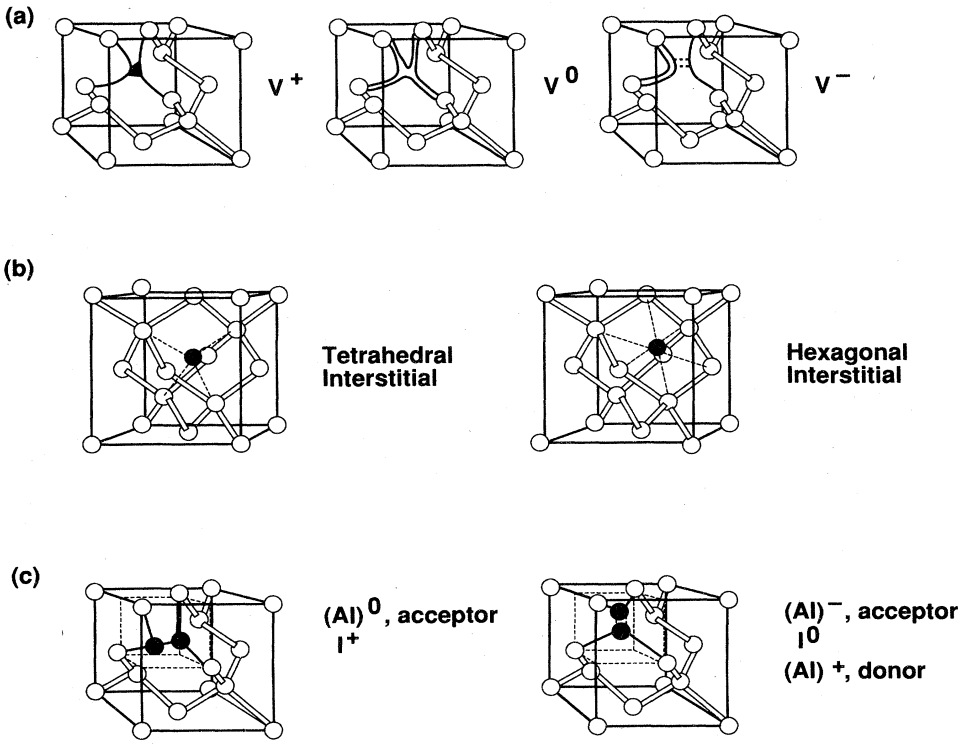

FIG. 1. Vacancy, interstitial, and interstitialcy point defects. (a) Vacancy in the +, 0, and − charge states. Darkened bonds indicate orbitals with unpaired spins, which make V^+ and V^- visible in electron paramagnetic resonance experiments. After Watkins, Troxell, and Chatterjee (1979). (b) Dark spheres indicate atoms in two different interstitial positions. (c) Interstitialcy defects. These represent silicon interstitialcy defects if both of the dark spheres are silicon atoms, and dopant interstitialcies AI if one of the spheres is a dopant atom.

referred to as self-interstitials, silicon interstitials, or, simply, interstitials. In the remainder of this paper we shall also refer to interstitial-type defects in this way, making a distinction only when it is important.

It should be mentioned that the concept of an *extended point defect* has been introduced in the literature (Nachtrieb and Handler, 1954; Seeger and Chik, 1968). An extended point defect is proposed to be similar in nature to the defects shown in Fig. 1, but not nearly as localized as these simple pictures indicate. One could define an extended vacancy as a defect involving the interactions of $N-1$ silicon atoms about N sites, and similarly an extended interstitial as a defect involving interactions of $N+1$ silicon atoms about N sites.

There is little doubt that the lattice surrounding a point defect is distorted by its presence. An extended point defect is proposed to involve changes in the lattice over many lattice sites. There is no definite transition from the situation in which a defect should be considered pointlike to that in which it should be considered extended. However, when the defect loses its pointlike character it is not amenable to any kind of simple analysis. In particular, no work has yet been done in developing even an elementary model of migration of such a defect. In all of the following discussions, point defects will be considered *not* extended in nature.

B. Dopant defects

An atom that resides on a lattice site is known as a *substitutional defect.* As mentioned earlier, dopants dissolve in the lattice almost exclusively on substitutional lattice sites. When the dopant atom occupies a substitutional site surrounded only by silicon atoms, it will be referred to as A. When a vacancy V resides next to a substitutional dopant atom, it is known as a *dopant-vacancy pair* and will be designated AV. If one of the atoms in an interstitialcy defect is a dopant atom [for example, one of the dark spheres in Fig. 1(c)], then the defect will be called a *dopant-interstitialcy pair* and written AI. If a silicon interstitial in an interstice pairs with a substitutional dopant atom, it will also be referred to as AI. If the dopant atom itself occupies an interstitial position, it will be referred to as an *interstitial dopant* and written A_i.

The dopant defects can migrate in the AV, AI, or A_i states. The dopant defects are formed by the reactions

$$A+V \rightleftharpoons AV \ , \tag{3.1}$$

$$A+I \rightleftharpoons AI \ , \tag{3.2}$$

$$A+I \rightleftharpoons A_i \ , \tag{3.3}$$

$$A \rightleftharpoons A_i+V \ . \tag{3.4}$$

These reactions and additional reactions accounting for recombination of intrinsic point defects with the defect pairs are treated in more detail in Sec. X.A. Here, it is sufficient to establish a consistent nomenclature for the diffusing species and the associated reactions. The first equation describes a *vacancy* mechanism for dopant diffusion. The next three equations [(3.2)–(3.4)] describe dopant diffusion by substitutional/interstitial(cy) interchange mechanisms. Equations (3.2) and (3.3) are known as "kick-out" reactions. The last equation [(3.4)] is known as the dissociative reaction and is also commonly referred to as the Frank-Turnbull mechanism of diffusion (Frank and Turnbull, 1956). The Frank-Turnbull mechanism is commonly believed to be unimportant for the diffusion of p- and n-type dopants in silicon. The reasons for this, the energetics of the other equations, and the implications for dopant diffusion mechanisms are discussed in detail in Sec. XI and Appendix D.

The substitutional/interstitialcy and substitutional/interstitial interchange mechanisms of Eqs. (3.2) and (3.3) are very similar. Both require generation of an interstitial or interstitialcy [referred to as an I-type or interstitial(cy)] defect when a dopant atom returns to the substitutional state, and both predict an increase in mobile dopant atoms in the presence of excess I concentrations. This makes it very difficult to differentiate between the two mechanisms experimentally, which naturally invites the question of why it is important to know which mechanism is dominant. From the point of view of predicting the redistribution of dopants under various experimental conditions, either qualitatively from physical considerations or quantitatively by numerical evaluation of the equations governing the point-defect and dopant interactions, the answer appears to be that the distinction between interstitial and interstitialcy mechanisms is not important at all.

Discussions surrounding interstitials versus interstitialcies have arisen in the past primarily for theoretical reasons. One subtle distinction between the interstitial and the interstitialcy defect was pointed out by Van Vechten (1980). The isolated silicon interstitial can only be a donor because there are no available valence band states in a perfect crystal to accommodate the wave functions of an extra unbonded silicon atom. This argument led Van Vechten to propose that the formation energy of a pure interstitial atom in a host lattice would be directly proportional to the formal valence (1–8) of the element, since this determines the number of wave functions to be accommodated by the conduction band states. Because $Z=4$ for silicon, this leads to very high estimates (9.6 eV) for the formation energy of a pure silicon interstitial defect. For many years, this analysis was used to support "vacancy-only" mechanisms of diffusion.

Yet, there is a plethora of experimental evidence that shows that the inclusion of interstitial or interstitialcy (I-type) defects is an essential feature of any theory of diffusion processes in silicon. Thus recent theoretical calculations have sought to find some configuration in the lattice for which an I-type, AI, or A_i defect has a believable energy of formation and migration.

It should be kept in mind in these debates that it is always possible for some combination of interstitial and interstitialcy mechanisms to be operative. For example, a

silicon interstitialcy may kick a substitutional dopant atom into an interstitial position, or a silicon interstitial may force a substitutional dopant atom into an interstitialcy configuration. Some experimental evidence that such combinations may take place comes from the low-temperature irradiation studies of Watkins (1964). After electron irradiation at low temperature (<200 K) of Al-doped samples, some of the Al atoms are found to be displaced into tetrahedral interstitial sites as Al_i^{++}. As discussed in Sec. VII.B, this is most likely due to the I-type defects created by the irradiation migrating to the Al atoms and kicking them into the interstitial position. However, under similar conditions carbon appears to be displaced into the split interstitialcy position shown in the right-most picture of Fig. 1(c) (Watkins and Brower, 1976) with a single bond adjoining the silicon atom, thereby resulting in the $(CI)^+$ interstitialcy defect. On the other hand, B appears as a $(BI)^0$ interstitialcy whose configuration is not known but, according to Watkins, is suggestive of a slight distortion of the bond-centered configuration in Fig. 1(c).

For the above nomenclature to be perfectly consistent, the substitutional dopant should always be designated as A_s instead of A; however, since almost all the dopants reside on substitutional sites, the subscript s will be considered understood. Frequently, it is advantageous to discuss point defects in a general way without specifying whether they are vacancy or interstitial in character. In this case we use the symbol X to refer to a point defect that can be either I or V. So, for example, the symbol AX means that the discussion is relevant to both AI and AV defects.

IV. SOURCES AND SINKS OF POINT DEFECTS: SURFACE AND BULK EFFECTS

Here we discuss qualitatively the possible sources and sinks of point defects; sections where each is covered in more detail are indicated.

A. Under equilibrium conditions

While thermodynamics predicts from fundamental principles that point defects exist in thermal equilibrium, this does not tell us how equilibrium is reached. That is, what is the physical process that creates the point defects that populate the silicon crystal? In a dislocation-free silicon crystal, there are only two possibilities. The first is that silicon atoms spontaneously leave their substitutional lattice positions, simultaneously creating interstitial and vacancy point defects,

$$I+V \rightleftharpoons 0 \ . \tag{4.1}$$

In equilibrium there is no requirement that the number of vacancy and interstitial defects be equal because one of the species may recombine at the surface, thereby lowering the total free energy of the system, and in the process giving the net free energy of formation of the companion species. The bulk process (also known as the Frenkel process) initially requires the combined energy needed to form simultaneously an interstitial and a vacancy defect and may be considered a defect creation process with a high activation energy barrier.

The second possibility is that an interstitial is created by a silicon atom at the surface moving into the bulk, and a vacancy created by a substitutional silicon atom moving to the surface. As the silicon crystal is heated, point defects flow from the surface into the bulk. Since the inverse of the surface creation process also occurs, there is a net flux into the wafer for a given point defect,

$$J_X=\sigma_X(C_X-C_X^*) \ , \tag{4.2}$$

where the superscript $*$ denotes equilibrium and σ_X has the units of velocity. The physical meaning of σ_X is discussed further in Sec. XII.B. Because the surface process (also known as the Schottky process) creates the defects independently, in equilibrium the concentration of I-type defects need not be equal to the concentration of vacancies.

An interesting demonstration of the surface creation process was recently presented by Chantre, Kechouane, and Bois (1983). These researchers used a continuous wave (cw) laser to heat and cool silicon wafers quickly. From the depth distribution of quenched-in defects they were able to deduce that vacancy defects were formed at the surface as a result of heating rather than in the bulk.

The precise details of the surface reaction are not important for either the Frenkel or the Schottky processes (Van Vechten, 1980). For example, whether a displaced atom recombines at an ideal flat surface or at a kink site on the surface does not affect the formation enthalpy of a defect, because it is defined statistically in terms of reactions occurring among ensembles of all possible configurations of the material in the limit of large systems. Since the number of surface sites varies as the square of the linear dimension and the number of bulk sites varies as the cube of the linear dimension, the atomistic process at the surface is irrelevant in determining the formation enthalpy.

For either the bulk or surface process, I and V are free to recombine independently at the silicon surfaces, so that in either case the equilibrium defect concentrations will be the same as those calculated by considering surface generation and annihilation processes alone.

Under nonequilibrium processes such as those induced by surface chemical reactions, the bulk process of Eq. (4.1) can strongly affect the behavior of the excess point defects. This is discussed in some detail in Sec. XII. In addition, in Figs. 24 and 25 of Sec. XII we demonstrate that the bulk generation of I and V can be very important in determining the kinetics of a silicon crystal approaching equilibrium, and that the bulk defect reactions cannot be ignored arbitrarily with respect to surface point-defect reactions.

B. Chemical reactions at the silicon surface

Chemical reactions at the silicon surface may alter the net generation and annihilation rates of points defects. The best known example of such a reaction is thermal oxidation of the silicon surface. It has been accepted for some time that this process injects I-type defects into the bulk, although the underlying cause of I injection is still unknown. It must also be noted that at temperatures exceeding 1150 °C, oxidation of ⟨111⟩-oriented silicon substrates appears to inject vacancies (Tan and Ginsberg, 1983).

An accepted test of which type of defect, I or V, is injected into the bulk as a result of surface reactions is to observe the growth or shrinkage behavior of *extrinsic* stacking faults. Extrinsic stacking faults are extra half planes of silicon atoms and so may grow by absorbing excess I defects or shrink by absorbing excess V. Extensive observations of stacking fault growth and shrinkage have led to a consensus among virtually all researchers in the field as to the effect of oxidation on point defects in the bulk. In Sec. XIV.F we discuss similar experiments indicating that thermal nitridation of the silicon surface by exposure to an NH_3 ambient results in the injection of vacancies, while thermal nitridation of an overlaying, pregrown SiO_2 layer causes interstitial injection.

C. Precipitation

When the solid solubility of an impurity in silicon is exceeded, precipitation of the impurity will occur. Precipitation processes can affect point-defect concentrations in two ways. First, there will in general be a volume change associated with the formation of a precipitate. For example, oxygen, which dissolves largely in the interstitial state O_i, will form SiO_2 precipitates with roughly twice the volume of a region in the silicon lattice that contains an equal number of silicon atoms. To accommodate this volume change, extra silicon atoms are expelled from the regions surrounding the precipitates as I-type defects. Second, the AX defects responsible for migration of the impurity atoms to and from the precipitates can form or break up during the precipitate growth and shrinkage processes. For example, a PI defect may diffuse to a SiP precipitate releasing an I defect as the P atom joins the precipitate.

In addition, it also possible for the native point defects themselves to coalesce into interstitial or vacancy aggregates. Defects known as *swirl defects,* which are found in as-grown silicon boules, are thought to represent the coalescence of I- and V-type defects that were introduced as the crystal was being grown. The agglomeration may occur because the energy barrier to I-V recombination is sufficiently high and the defects sufficiently far away from the crystal surfaces that, upon cooling of the crystal, condensation of the point defects occurs at a faster rate than bulk and surface annihilation processes.

D. Dislocations

Since dislocations by their nature result in regions of the crystal having more or fewer atoms per unit volume than would be found in the otherwise perfect crystal, dislocations can serve as both sources and sinks of point defects. In silicon material grown by modern techniques, the dislocation density is relatively small, so that dislocations are usually not considered to be important sources or sinks of point defects. However, it is always possible for dislocations to affect point-defect behavior, and their relative importance should always be assessed in relation to the specific experimental situation. For instance, highly dislocated silicon material was used to elucidate the mechanisms of Au diffusion by providing an efficient sink for interstitials (Stolwijk, Hölzl, Frank, Weber, and Mehrer, 1986). Dislocations may also affect point-defect concentrations indirectly by acting as preferential nucleation sites for precipitation.

E. Conversion of dopants to the substitutional state

During the redistribution of a nonhomogeneous dopant concentration profile, dopant atoms will move to undoped regions of the crystal in the AV, AI, or A_i state.[1] These species must then convert to the substitutional state in order to be consistent with the experimental observation that $C_A \gg C_{AV}$, C_{AI}, or C_{A_i}. Conversion to the substitutional state will occur through the reactions

$$AX \rightleftharpoons A + X \tag{4.3}$$

or

$$A_i \rightleftharpoons A + I \ , \tag{4.4}$$

both leading to a supersaturation of I or V.

Fahey, Dutton, and Hu (1984) have proposed that supersaturation of I results during P diffusion. The I supersaturation occurs in the absence of surface oxidation and during both the growth and shrinkage of SiP precipitates. This leaves the conversion process of either PI or P_i to the substitutional state as the most likely cause of I supersaturation. This is covered in more detail in Sec. XIII.C.

---

[1]Mechanisms, not involving point defects known as "exchange" mechanisms, have been dismissed in the past as highly unlikely processes compared to point-defect-assisted mechanisms based on energetics arguments (see, for example, the review article by Hu, 1973a). However, a new type of exchange mechanism has recently been proposed by Pandey (1986). Its possible importance for dopant diffusion is briefly discussed in Sec. XVII.

F. In-diffusion

In some deposition techniques, wafers are doped by processes in which the dopant atoms enter through the silicon surface and diffuse into the bulk. Since almost all the atoms are found in the substitutional state after the diffusion step is terminated, it is possible that the dopants enter the silicon in an interstitial-type state and eject silicon atoms into interstitial sites through one of the reactions listed above. This would lead to an I supersaturation in the bulk. It is also possible that the dopants enter the lattice in the substitutional state by occupying a vacant lattice site near the surface. This would lead to an undersaturation of V in the bulk. Whether a measurable excess or deficit of defects actually results from in-diffusion depends, of course, on how fast the dopant atoms enter and diffuse into the substrate compared to the rate of the compensating point-defect processes. At present, there is no solid experimental evidence that in-diffusion of dopants by itself significantly perturbs the concentrations of point defects from their equilibrium values.

G. Radiation damage

Equilibrium point-defect concentrations are too small to be visible to most analytical techniques (electron paramagnetic resonance, deep-level transient spectroscopy, etc.). As a result, point defects have been intentionally created by electron bombardment or irradiation with high-energy photons. These are performed at low enough temperatures that some of the defects so created are frozen in the silicon and their properties can thus be subsequently examined.

In modern integrated circuit fabrication processes, one of the most common ways to introduce dopants into the bulk is by ion implantation. Dopant ions are accelerated at energies of a few thousand to a few hundred thousand eV into the silicon substrate. This results in the displacement of some silicon atoms from their normal sites and the introduction of some dopants into nonsubstitutional positions. Upon heating to anneal the implantation damage, the defect species created by implantation will become mobile and may lead to transient enhanced diffusion. This is a research area of active interest [see, for example, the recent article by Michel *et al.* (1987)].

V. BASIC THERMODYNAMICS OF NATIVE POINT DEFECTS

A. Equilibrium concentrations

It is a well-known prediction of statistical thermodynamics that for any temperature other than 0 K, finite concentrations of point defects will exist in thermal equilibrium, reflecting the fact that this situation is the state of the crystal that minimizes its free energy. It is also a well-known result that the equilibrium concentration of a point-defect species X can be expressed as (see, for example, Swalin, 1962; Lannoo and Bourgoin, 1981)

$$\frac{C_X}{C_S}=\theta_X \exp\left[\frac{S_X^f}{k}\right]\exp\left[\frac{-H_X^f}{kT}\right], \tag{5.1}$$

where C_S is the number of available lattice sites in the crystal (i.e., number of substitutional sites, tetrahedral interstitial sites, etc.), H_X^f is the enthalpy of formation of X, the formation entropy S_X^f is the disorder entropy not associated with configuration and is usually attributed to lattice vibrations, and the term θ_X is the number of degrees of internal freedom of the defect on a lattice site (for example, spin degeneracy).

B. Determination of equilibrium concentrations

It must be stated at the outset that no experiment has definitively measured the equilibrium concentrations of vacancies or interstitials in silicon, or even the activation enthalpies of formation. Accurate determination of these quantities would be one of the most significant developments in the history of this area of research.

1. At low temperature

Native monovacancies have long been positively identified at low temperatures (see the review by Watkins, Troxell, and Chatterjee, 1979) by electron paramagnetic resonance (EPR) and deep-level transient spectroscopy (DLTS) after electron irradiation of the silicon. Interstitial-type defects have not been directly observed by EPR, but their presence has been inferred by kick-out reactions which occur during electron irradiation. The energy levels of the defects are discussed in Sec. VI.B, while the migration mechanisms of the defects are treated in Sec. VII.B.

A particular experiment that was originally thought to rule out either interstitials or vacancies as a dominant equilibrium defect was Frenkel pair production by high-energy electron irradiation. But there are experimental complications that make these determinations somewhat unreliable; they have been summarized by Flicker, Loferski, and Scott-Monck (1962). It is also possible that interactions with the electron beam enhance the migration and annihilation of the defects (Sec. VII.B). The method may tend to overestimate the threshold energy for pair production. One can write the displacement energy as

$$E_d=\Delta H_f^I+\Delta H_f^V+E_L , \tag{5.2}$$

where E_L is the energy for lattice distortion, which might be assumed to consist mostly of a bond-bending component rather than a stretching or compressing of bond lengths. This leads to estimates (Van Vechten, 1980) in the region of a few electron volts (3.2 eV for Si) for E_L. Threshold displacement energies have a wide scatter as-

sociated with practical difficulties in this conceptually simple experiment, with experimental values in the range of 12–15 eV. Because original estimates of the vacancy formation energy (Van Vechten, 1980) were often 2–3 eV, these experimental results were taken as evidence that interstitials could not be a dominant equilibrium defect.

Contrasting with these results are *ab initio* calculations, based on a more or less rigorous solution of Schrödinger's equation, for the formation energies of I and V (Baraff and Schlüter, 1984; Bar-Yam and Joannopoulos, 1984; Car, Kelly, Oshiyama, and Pantelides, 1984b). The values from Car *et al.* have been revised downwards by ≈ 1 eV since the original publication in *Physical Review Letters* (Car, Kelly, Oshiyama, and Pantelides, 1984a) and give formation energies for both I and V of 4–5 eV, with an uncertainty of ≈ 0.5 eV. The common conclusion from these studies is that I and V have comparably high formation energies, consistent in spirit with the Frenkel pair production data.

2. At high temperature

An important and controversial question about point defects at high temperatures is whether interstitials or vacancies are the dominant high-temperature defect. The question can be answered convincingly but indirectly for the most part. No one has proposed a vacancy-only or interstitial-only model that consistently explains the experimental data on dopant diffusion or metal in-diffusion at high temperatures. The critical experimental data in question relate to different in-diffusion mechanisms for different metals and to enhanced and retarded impurity diffusion, which commonly occurs during integrated circuit fabrication processes. Only an "interstitial plus vacancy" model provides a unified explanation of the phenomena.

There is no direct evidence of an equilibrium interstitial population at high temperatures, but the circumstantial evidence is very strong. In Sec. XVII the identification of the mechanisms responsible for dopant diffusion is discussed, and the conclusion is that both interstitial(cy) and vacancy mechanisms are important in equilibrium. Some recent experimental studies allow reasonable estimates of the equilibrium concentrations of I and V at diffusion temperatures. It should be kept in mind, though, that point defects can exist in multiple charge states (as discussed in the section immediately following this one) and the experimental techniques discussed here may only be sensitive to point defects in one predominant charge state.

Information on equilibrium I values has come from the work of Mantovani, Nava, Nobili, and Ottaviani (1986). These investigators have observed the in-diffusion of platinum in silicon in an effort to measure the self-diffusion of silicon. The basis of this technique is discussed in more detail in Sec. IX. They demonstrated that their results can be interpreted self-consistently if self-diffusion is dominated by an interstitial mechanism and platinum diffusion by a substitutional/interstitial interchange, or kick-out mechanism (Sec. XI). This work has been combined with a previous analysis of gold diffusion by the kick-out mechanism from Seeger (1980) to obtain an estimate of the equilibrium concentration of interstitials (Morehead, 1987). The Arrhenius expression that best fits the data is

$$C_I = 1\times 10^{27}\exp\left[-\frac{3.8\text{ eV}}{kT}\right]\text{ cm}^{-3} \tag{5.3}$$

and is plotted in Fig. 2. The preexponential factor is equal to $\theta_I C_S \exp(S_I^f/k)$. If the I defect is an interstitialcy like those shown in Fig. 1, then $C_S = 5\times 10^{22}$ cm^{-3}. This gives a value of $S_I^f = 9.9k$. It should be noted that the value of $H_I^f = 3.8$ eV is higher than a previous estimate of 2.9 eV by Seeger, Föll, and Frank (1977), who analyzed swirl defect patterns assuming that they were due to condensation of I defects upon cooling during crystal growth from the silicon melt (the interstitial nature of the A-type swirl defects observed in this study is undisputed).

Estimates of vacancy concentrations come from the positron annihilation study of Dannefaer, Mascher, and Kerr (1986). They quote a formation enthalpy of 3.22 eV (which they ascribe to a negatively charged vacancy) and an entropy factor between $6k$ and $10k$. The upper and lower limits of C_V are shown in Fig. 2. The analysis of these data has been criticized by Van Vechten (1987), who proposes that a correct interpretation of the positron data yields $H_V^f = 2.6$ eV and $S_V^f = 2.9k$; however, these values give a vacancy concentration that lies within the upper and lower bounds indicated in Fig. 2 [see the review article by Dannefaer (1986) for a detailed discussion of the model assumptions made in interpreting the positron annihilation data].

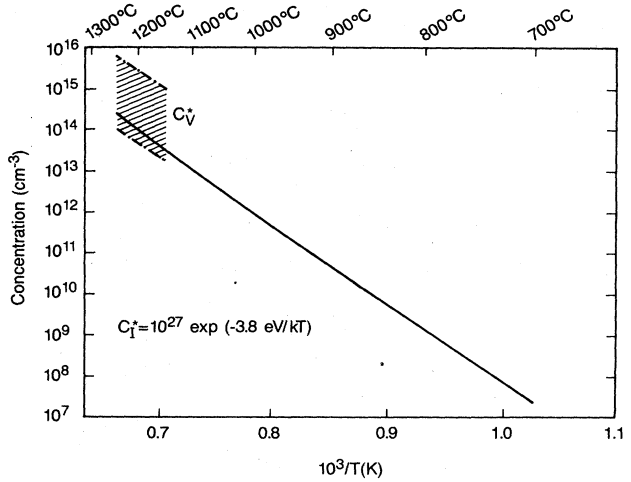

FIG. 2. Estimates, from experiment, of the equilibrium concentration of silicon interstitials (Morehead, 1987) and vacancies (Dannefaer, Mascher, and Kerr, 1986).

Putting aside possible disagreements concerning interpretation of data, the use of the positron annihilation technique for Si is a significant development. It is the only technique currently being applied to study point-defect and diffusion problems that makes measurements at the elevated temperatures where diffusion occurs and is therefore arguably the most direct experimental approach to determining equilibrium point-defect concentrations. The technique has also been applied to heavily P-doped samples (Mascher, Kerr, and Dannefaer, 1987) in an attempt to detect the increase in vacancy concentrations that is expected in heavily doped samples (Secs. VI.C and VIII.A). Signals attributed to phosphorus vacancy pairs are observed, but the interpretation of data to yield concentration values is considerably more complex than in the case of very lightly doped samples.

VI. POINT DEFECTS IN MULTIPLE CHARGE STATES

As indicated in Fig. 1, point defects can exist in different multiple charge states. It is of great interest to know how a given point-defect concentration is populated in each of its allowed charge states. There are several reasons for this.

(i) Below solubility limits, all dopant atoms are ionized at diffusion temperatures. As a result, those defects that have opposite charge sign to a dopant will have a Coulombic attraction for that dopant atom, and those of similar charge will experience a Coulombic repulsion. In addition, the charge state of the defect is expected to affect the local distortion it introduces into the silicon lattice and thereby also to affect its affinity for a dopant atom.

(ii) When the dopant concentration C_A exceeds the intrinsic carrier concentration n_i at the diffusion temperature, the silicon is said to be extrinsically doped, and the free-carrier concentration is determined by the local concentration of dopant atoms. As will be discussed in this section, the *total* number of point defects that exist in equilibrium will change under extrinsic conditions, not just the relative populations in different charge states.

(iii) Electronic charging is considered to be much faster than diffusion processes, so that even under nonequilibrium conditions the relative populations of point defects among their various charge states in steady state are expected to be the same as those calculated under equilibrium conditions

A. Calculation of equilibrium concentrations

To keep the following treatment general, we discuss a point defect X, which can be either I or V, and consider it to have stable configurations in either neutral 0, singly charged $-,+$, or doubly charged $=,++$ states with respective energy levels $E_{X^-},E_{X^+},E_{X^=},E_{X^{++}}$ in the band gap. A straightforward extension of the treatment of Shockley and Last (1957) gives the relations

$$\frac{C_{X^-}}{C_{X^0}}=\frac{\theta_{X^-}}{\theta_{X^0}}\exp\left[-\frac{E_{X^-}-E_f}{kT}\right], \tag{6.1}$$

$$\frac{C_{X^=}}{C_{X^0}}=\frac{\theta_{X^=}}{\theta_{X^0}}\exp\left[-\frac{E_{X^=}+E_{X^-}-2E_f}{kT}\right], \tag{6.2}$$

$$\frac{C_{X^+}}{C_{X^0}}=\frac{\theta_{X^+}}{\theta_{X^0}}\exp\left[-\frac{E_f-E_{X^+}}{kT}\right], \tag{6.3}$$

$$\frac{C_{X^{++}}}{C_{X^0}}=\frac{\theta_{X^{++}}}{\theta_{X^0}}\exp\left[-\frac{2E_f-E_{X^{++}}-E_{X^+}}{kT}\right], \tag{6.4}$$

where E_f is the Fermi level and C_X^0 is given by

$$\frac{C_X^0}{C_S}=\theta_X^0\exp\left[\frac{S_{X^0}^f}{k}\right]\exp\left[\frac{-H_{X^0}^f}{kT}\right]. \tag{6.5}$$

These equations are completely general in a dilute concentration approximation. Sometimes, other expressions similar to these equations appear in the literature. However, it should be recognized that alternative expressions make implicit assumptions about the relationships between the parameters in Eqs. (6.1)–(6.4). The only situation in which Eqs. (6.1)–(6.4) are no longer valid is when the doping level becomes so high that the number of lattice sites at which a defect can be considered unaffected by the local presence of dopant atoms is reduced significantly from the total number of lattice sites. The doping levels at which this breakdown occurs are discussed in more detail in Sec. VIII. In the absence of this doping effect, Eqs. (6.1)–(6.4) give the relative populations of those charged and neutral point defects that are unassociated with dopant atoms. The effect of the dopant atoms is to change the Fermi level by introducing more carriers into the silicon.

If we compare Eqs. (6.1)–(6.4) with Eq. (5.1) we can see immediately that the energy of formation of a defect depends on its charge state. For example, the energy of formation of X^- can be written

$$G_{X^-}^f=G_{X^0}^f+E_{X^-}-E_f, \tag{6.6}$$

where $G_{X^0}^f=H_{X^0}^f-TS_{X^0}^f$. In general, the term $(E_{X^-}-E_f)$ will contain both enthalpy and entropy factors. That the energy of formation of a point defect should depend on its charge states can be easily appreciated by looking at Fig. 1, where different configurations in the lattice of point defects depending on their charge states are depicted. An important consequence of the ability of point defects to change their charge states is that the concentration of charged point defects that exist in thermal equilibrium can be altered by changing the Fermi level in the silicon. This can be accomplished by doping. Furthermore, since the equilibrium concentration of charge-neutral defects X^0 is independent of the Fermi level, it follows that the *total concentration* of

point defects that exist in thermal equilibrium can be altered by doping silicon. The subject of point defects under extrinsic doping conditions is discussed below after first summarizing the experimental data of point-defect energy levels in silicon.

B. Experimental determination of energy levels

The only positive identification of energy levels that can be ascribed to isolated native point defects are those for vacancies. The levels reported by Watkins and coworkers for isolated vacancies are (Watkins, 1975; Watkins, Troxell, and Chatterjee, 1979; Newton, Chatterjee, Harris, and Watkins, 1983)

$$E_c - E_{V^-} = 0.57 \text{ eV}, \quad E_c - E_{V^=} = 0.11 \text{ eV}, \\ E_{V^+} - E_v = 0.05 \text{ eV}, \quad E_{V^{++}} - E_v = 0.13 \text{ eV}. \tag{6.7}$$

Identification of defects and their energy levels was determined using EPR and DLTS techniques on samples in which isolated vacancies had been produced by high-energy electron bombardment. A surprising feature of this charge-state assignment is that the double positive charge state $E_{V^{++}}$ is more favored than the single positive charge state E_{V^+} for any position of the Fermi level, i.e., the energy of the donor vacancy is actually lowered upon capturing a second hole, indicating that relaxation of the vacancy structure more than compensates for the Coulombic repulsion between the two carriers. This *negative-U* property of vacancies was predicted by Baraff, Kane, and Schlüter (1979) on the basis of theoretical calculations and was convincingly demonstrated (Watkins and Troxell, 1980) for the vacancy in *p*-type material by using DLTS to monitor the thermally activated hole release

$$V^{++} \rightarrow V^+ + h^+ \rightarrow V^0 + 2h^+ . \tag{6.8}$$

The limiting process is the first hole release, since for the second more weakly bound hole, carrier emission should follow immediately in a negative-*U* system. The result is that DLTS detects only the deeper level, but the amplitude of the carrier emission is twice the normal emission. Subsequent experiments confirmed this assignment and indicated that boron interstitials produced by electron irradiation also have this property with an acceptor level at $E_c - 0.45$ eV and donor level at $E_c - 0.13$ eV (Harris, Newton, and Watkins, 1982, 1983; Newton, Chatterjee, Harris, and Watkins, 1983).

Isolated silicon interstitials are not observed during electron irradiation at low temperature, presumably because they migrate quickly by an ionization-enhanced diffusion process to sinks, as discussed in Sec. VII.B. Frank (1975) has proposed that a donor level at $E_v + 0.4$ eV and an acceptor level at $E_c - 0.32$ eV correspond to the silicon self-interstitialcy (and specifically an extended interstitialcy). Seeger, Föll, and Frank (1977) have conjectured that a defect with energy level $E_c - 0.39$ eV (Mallon and Naber, 1970; Naber, Mallon, and Leadon, 1972) is a negatively charged self-interstitial, and may be the same defect giving rise to the Si-G25 EPR spectrum (Watkins, 1975). However, in a subsequent work aimed at gathering more information about the "elusive" self-interstitial, Harris and Watkins (1985) have presented experimental evidence that the Si-G25 defect is not an isolated self-interstitial. Lefèvre (1980) correlated DLTS profiles with *A*-swirl defects of an interstitial nature in float-zone wafers and proposed that acceptor levels at $E_c - 0.49$ eV and at $E_c - 0.07$ eV were related to silicon self-interstitials. These assignments for interstitial energy levels are not as generally accepted as are the levels for the vacancy defects.

Energy levels for an acceptor level corresponding to P*V*, As*V*, Sb*V*, and Bi*V* all appear to be bunched at approximately $E_c - 0.4$ eV (Elkin and Watkins, 1968).

C. Native defects under extrinsic doping conditions

Dopant concentrations are said to be extrinsic if the equilibrium carrier concentration of electrons or holes, n or p, appreciably exceeds the intrinsic carrier concentration n_i. (A plot of n_i versus temperature for silicon is shown in Fig. 3.) Since we assume that each dopant atom provides one free carrier, this condition is obviously met if $C_A > n_i$. In this case the Fermi level depends on dopant concentration. The intrinsic carrier concentration is always higher than the concentration of point defects at diffusion temperatures, so that only the ionized dopants affect the Fermi level. Using X^- as an example, we expect that when the Fermi level E_f is changed from its value under intrinsic conditions, E_f^i, C_{X^-} will change as

$$\frac{C_{X^-}}{(C_{X^-})^i} = \exp\left[\frac{E_f - E_f^i}{kT}\right] . \tag{6.9}$$

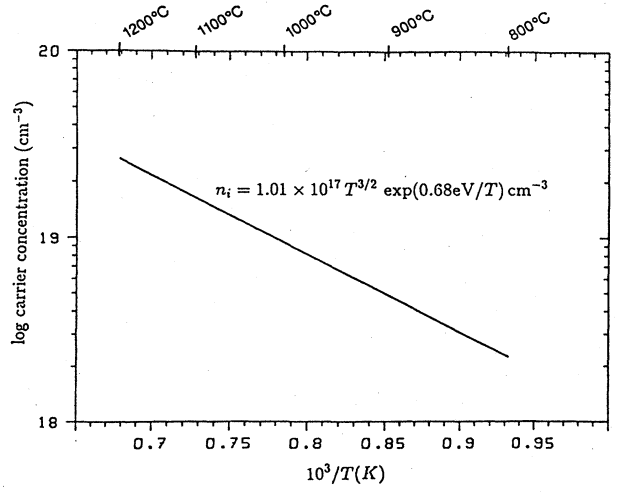

FIG. 3. Intrinsic carrier concentration in silicon vs temperature [based on the work of Morin and Maita (1954a, 1954b)].

Similar expressions follow for $X^{=}$, X^{+}, and X^{++}. Thus one can predict the changes in charged point-defect concentrations as a function of doping concentration by relating C_A to $(E_f - E_f^i)$. If we assume Boltzmann statistics are valid, we would have simply that

$$\frac{n}{n_i} = \exp\left[\frac{E_f - E_f^i}{kT}\right], \quad \frac{p}{n_i} = \exp\left[\frac{E_f^i - E_f}{kT}\right]. \tag{6.10}$$

Similar reasoning leads to the following simple relationships:

$$\frac{C_{X^-}}{(C_{X^-})^i} = \frac{n}{n_i}, \quad \frac{C_{X^=}}{(C_{X^=})^i} = \left[\frac{n}{n_i}\right]^2,$$
$$\frac{C_{X^+}}{(C_{X^+})^i} = \frac{p}{n_i}, \quad \frac{C_{X^{++}}}{(C_{X^{++}})^i} = \left[\frac{p}{n_i}\right]^2. \tag{6.11}$$

These relations show that relative to intrinsic doping conditions, X^- and $X^=$ concentrations are higher in n-type silicon and lower in p-type, while X^+ and X^{++} concentrations are higher in p-type silicon and lower in n-type.

D. Effects of band-gap narrowing

It is often necessary to try to extrapolate the experimental results found at one temperature and doping condition to other conditions. As seen in Eqs. (6.1)–(6.4), predicting the temperature dependence and concentration dependence of charged point-defect concentrations amounts to determining how the separation between their energy levels in the band gap and the Fermi level depends on these parameters. Since the band gap is known to change with both temperature and dopant concentration, some attempts have been made to incorporate "band-gap narrowing" terms into the expressions in Eq. (6.11). What is always implied in these modifications is that band-gap narrowing (or widening) has changed n_i. So, for example, in the range where Boltzmann statistics are valid we have that

$$n_i = \sqrt{N_c N_v}\exp\left[-\frac{E_g}{2kT}\right], \tag{6.12}$$

where N_c, N_v are the density of states in the conduction band and valence band edges, respectively, and E_g the energy separation between the top of the valence band and the bottom of the conduction band, is proposed to change with temperature and doping. However, when extrapolating from temperatures or doping conditions different from those at which the values of $(E_{X^-} - E_f)$, $(E_{X^+} - E_f)$, etc., were determined, it is *not* correct in general simply to modify the value of n_i in Eqs. (6.11) as is often done. We examine the two cases of interest for band-gap narrowing effects: the effect of temperature-dependent band-gap narrowing and the effect of dopant concentration on band-gap narrowing.

To take into account the temperature dependence of E_g over the limited temperature range 800–1100 °C, a linear approximation for band-gap narrowing has been used (Fair, 1981a; Antoniadis, 1983):[2]

$$E_g(T) \approx E_{g0} - \beta T . \tag{6.13}$$

Using the data of Morin and Maita (1954a, 1954b), which were measured for temperatures below 600 K, one obtains $E_{g0} = 1.46$ eV, $\beta = 2.97 \times 10^{-4}$ eV/K.

Although E_{g0} and β are determined by empirical fitting from a chosen set of data over a finite temperature range, they may be interpreted physically by defining $E_g(T)$ as the free energy corresponding to the reaction $0 \rightleftharpoons h + e$. The free-energy change can then be written (Van Vechten, 1975)

$$E_g(T) - E_g(0) = H_g - TS_g . \tag{6.14}$$

Therefore

$$S_g = -\frac{\partial E_g}{\partial T} = \beta , \tag{6.15}$$

$$H_g = E_{g0} . \tag{6.16}$$

Physically, E_g really does decrease with increasing temperature and the intrinsic carrier concentration is greater than the concentration that would be predicted assuming E_g to have the same value at diffusion temperatures as at 0 K. But as the energy separation between band edges decreases with increasing temperature, in intrinsic material the Fermi level will maintain its position midway between the two, i.e., $(E_c - E_f^i) \approx (E_f^i - E_v) \approx E_g/2$. From Eqs. (6.1)–(6.4) one can see that it is still necessary to know how the E_X's vary with temperature from one of the band edges in order to utilize information about temperature-dependent band-gap narrowing in the calculation of charged point-defect concentrations. What is usually assumed is that E_{X^-} and $E_{X^=}$ remain a fixed distance below the conduction band edge, while E_{X^+} and $E_{X^{++}}$ remain a fixed distance above the valence band edge. This is based on a proposed model for vacancies (Van Vechten, 1974; Van Vechten and Thurmond, 1976). But it should be recognized that this is a model assumption that does not follow from basic principles, and it is certainly not justified to treat interstitial defects in the same way should Van Vechten's model be proven correct.

Even more difficult problems exist when trying to include the effects of dopant-induced band-gap narrowing. In the area of silicon device physics, dopant-induced band-gap narrowing is a subject of active interest because of the effects heavy doping can have on minority carrier populations and on carrier transport (Del Alamo, Swan-

[2]The theoretical and experimental behavior of $E_g(T)$ has been discussed by Thurmond (1975).

son, and Lietoila, 1983; Lee and Fossum, 1983; Del Alamo, Swirhun, and Swanson, 1985; Dhariwal, Ohja, and Srivistava, 1985). In the area of silicon process physics, it is not clear what relevance dopant-induced band-gap narrowing has to calculations of charged point-defect concentrations at diffusion temperatures.

Fair (1979) has attempted to model the lumped effects of dopant-induced band-gap narrowing by using the relations of Eqs. (6.1)–(6.4) with a modified value of n_i for the case that $C_A \gg n_i$:

$$n_i^{\text{eff}} = n_i \exp\left[-\frac{\delta E_g}{2kT}\right] . \tag{6.17}$$

The physical interpretation provided by Fair for the nature of δE_g is that band-gap changes are induced by the lattice strain associated with dopant incorporation in the silicon (see Sec. VIII.B.3). But how does this physically affect the concentrations of charged point defects? Consider, first, that a reduced energy separation between the valence and conduction band edges will have a negligible effect on the majority carrier concentration when $C_A \gg n_i$, since almost all of the carriers come from the ionized dopant atoms, not from carriers thermally generated across the energy gap. The minority carrier concentration can be affected by dopant-induced band-gap narrowing in this way, but we do not expect this to be an important process as far as dopant diffusion is concerned. For example, in heavily n-doped material we do not expect that donor dopants A^+ will interact in any significant way with X^+ or X^{++} defects. Thus the importance of dopant-induced band-gap narrowing in changing point-defect concentrations is to affect, by doping, the separation in energy between the Fermi level and the acceptor and donor levels of the charged point defects. The δE_g term in Eq. (6.17), then, cannot freely be interpreted as the energy of band-gap narrowing. In addition, there are always questions concerning the validity of Boltzmann statistics at these high concentrations.

At present, relations like Eq. (6.17) should be viewed only as a convenient form with which to lump heavy doping effects together, but no physical meaning should be attached to δE_g.

VII. MIGRATION OF POINT DEFECTS

A. General considerations

The ideas about migration that are contained in this section apply equally well to migration of interstitials, interstitialcies, vacancies, dopant interstitials, and dopant interstitialcies. Migration of AV is a bit more complicated and so is deferred to Sec. XI.C.

In Fig. 4 the general situation for a defect migrating between two equivalent sites, s and s', is illustrated. The potential energy barrier to migration E_X^m at its saddle point s' is shown. The magnitude of E_X^m, of course, depends on the type of defect and the available paths for its

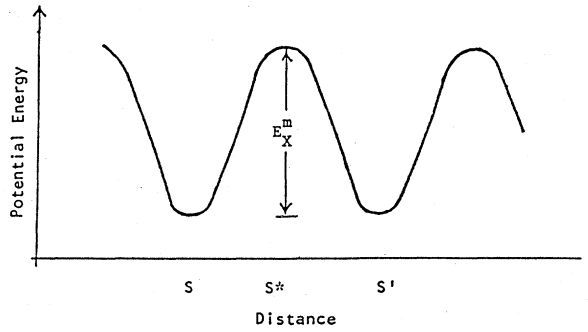

FIG. 4. Potential-energy diagram of a defect as it migrates between two equivalent lattice sites.

movement. The frequency with which the defect surmounts the energy barrier of its saddle point depends on the dynamic interactions of the defect with the surrounding host atoms. If E_X^m can be determined from experiment, an Arrhenius plot of E_X^m versus temperature can be used to extract values of H_X^m and S_X^m, although the physical meaning of these terms is not as clear as it is for the case of H_f^X and S_f^X [see, for example, the discussion by Swalin (1973)].

Two general observations can be made concerning the point-defect migration process. The first is that the equilibrium concentration of defects at the lowest energy configuration is just the concentration predicted by thermodynamics, as discussed in the previous two sections. One can see, then, that a theory capable of calculating G_X^f must be capable of predicting which configuration of X corresponds to its lowest energy state. The other observation is that for a given type of defect, I or V, the configuration in the lattice will depend on the charge state of the defect. Some examples are seen in Fig. 1. Thus the energy of the defect at its lowest energy state and at its saddle point will depend on the charge state of the defect.

The effects of ionization on defect migration rates have been discussed by Bourgoin, Corbett, and Frisch (1973). Changes in migration rates that result because the energy barrier between the lowest energy configuration and the saddle point is different when the defect changes charge state, have been termed *ionization-enhanced diffusion.* Two other types of ionization effects have also been recognized. Both are based on the realization that calculations of the relative populations of a given type of defect in its various charge states, as presented previously, reflect an ensemble averaging of the system, whereas in a kinetic picture, individual defects charge and discharge many times compared to the observation periods of defect diffusion. If we focus on a single defect X, which may exist in two different charge states X^x and X^y, the defect will alternately change between its two charge-state configurations, and it is possible that this change involves alternating between two lattice positions.

Bourgoin and Corbett (1972) first proposed that successive changes in charge state lead to motion of the in-

terstitial through the lattice. Recent *ab initio* theoretical calculations (Baraff and Schlüter, 1984; Bar-Yam and Joannopoulos, 1984; Car, Kelly, Oshiyama, and Pantelides, 1984a, 1984b) determined that the equilibrium configuration of the neutral silicon interstitial I^0 and of the charged interstitial I^{++} are at different lattice positions and that many pathways are available for migration.

Although these specific theoretical predictions have not been firmly established experimentally, they serve as specific examples of an intriguing idea: as the defect alternates between two charge states corresponding to two different lattice site locations, the defect may migrate *athermally.* This mechanism, also known as the *Bourgoin-Corbett mechanism,* has been offered as an explanation for the apparently high mobility of *I*-type defects, which, as discussed below, is observed from irradiation experiments performed at cryogenic temperatures. In addition, it is also possible that the charging and discharging process, corresponding to electron-hole recombination or carrier trapping, disturbs the local phonon modes, which may also change the rate of defect migration. This process has been termed the *energy release mechanism.* Experimental results that show that ionization effects can be important are included in the following discussion.

B. Experimental determination of defect migration energies

One general type of experiment to determine point-defect migration rates aims at creating point defects by irradiation with high-energy electrons, photons, or ions at temperatures low enough that the radiation-generated defects cannot migrate. Since the energy for creating the excess defects is provided externally, by subsequently monitoring the annealing behavior of the excess defect concentration at higher temperatures one may deduce the migration energies of the defects.

In these irradiation experiments, I and V are produced in equal pairs. However, even when the electron irradiation is performed at 4.2 K, only isolated vacancies are found still "frozen" in the material. Furthermore, with intentionally doped silicon it is found after irradiation that the number of group-III dopants found in interstitial sites is equal to the number of isolated vacancies. It is generally accepted that these findings demonstrate the I defect created by electron bombardment is mobile even at 4.2 K, and that the fast moving I reacts with substitutional group-III atoms A_s to produce interstitial group-III atoms A_i [see the review article by Watkins (1975)]. So although the value for H_I^m cannot be determined from these experiments, the results clearly show that the energy of migration must be very small indeed. The annealing behavior of the vacancies left behind in the radiation experiments leads to the following values of H_V as deduced from Arrhenius plots (Watkins, 1975; Watkins, Troxell, and Chatterjee, 1979; Newton, Chatterjee, Harris, and Watkins, 1983):

$$\begin{aligned} &H_{V^0}^m \approx 0.33 \text{ eV}, \quad T=150\text{–}180 \text{ K}, \\ &H_{V^{++}}^m \approx 0.45 \text{ eV}, \quad T=220 \text{ K}, \\ &H_{V^=}^m \leq 0.18 \text{ eV}, \quad T=70\text{–}85 \text{ K}. \end{aligned} \tag{7.1}$$

In related studies, using the same method of electron bombardment at low temperatures, the annealing behavior of AV pairs has been observed. The values found are

$$\begin{aligned} &H_{(PV)^0}^m = 0.94 \text{ eV}, \quad H_{(PV)^-}^m = 1.25 \text{ eV}, \\ &H_{(AsV)^0}^m = 1.07 \text{ eV}, \quad H_{(SbV)^0} = 1.28 \text{ eV}, \\ &H_{(BiV)^0}^m = 1.46 \text{ eV}. \end{aligned} \tag{7.2}$$

Hirata, Hirata, and Saito (1969) provided the values for PV, AsV, SbV, and BiV pairs migrating in the temperature range 150–250 °C. Elkin and Watkins (1968) presented data for PV, AsV, and SbV in the temperature range (−17)–100 °C which show excellent agreement with the data of Hirata *et al.* and identified these AV pairs as being in the neutral charge states. Kimerling, DeAngelis, and Diebold (1975) showed data for $(PV)^0$ in the temperature range 100–150 °C, confirming yet again the value of $H_{(PV)^0}^m = 0.93$ eV, and further showed that PV in the negative charge state has a 0.32 eV higher energy of migration.

Another type of experiment creates point defects by ion implantation of the dopant atoms themselves. After annealing at higher temperatures, enhanced diffusion of the dopants is observed for some transient period. If this is due to point defects created by the implantation process, then one can suppose that the activation energy of diffusion during the beginning of the transient period reflects the migration energy of the defect responsible for diffusion, since the energy of formation was supplied by the implantation process rather than by thermal energy. Pennycook, Narayan, and Holland (1984) created excess point defects by high-dose implantation of Sb, and upon recrystallization by solid-phase epitaxy observed greatly enhanced diffusion of Sb and the simultaneous formation of interstitial loops. They attributed the enhanced diffusivity of Sb to enhancement of its I component of diffusion and deduced that $H_{SbI} = 1.8 \pm 0.2$ eV. Kalish, Sedgewick, Mader, and Shatas (1983) also observed a 1.8-eV activation energy of As diffusion for ion-implanted samples that were rapidly annealed with a bank of tungsten-halogen lamps. Due to the experimental errors and unresolved theoretical questions surrounding the interpretation of these types of studies, one should not regard these experiments as definitive.

All of the above experiments were performed under conditions such that dopant diffusion was negligible at the temperature at which the radiation was produced. Two different groups have produced radiation damage by proton bombardment at temperatures high enough that dopant diffusion may be observed directly during the de-

fect creation process. Masters and Gorey (1978, 1979) studied the radiation-enhanced diffusion of B, P, and As in the temperature range 600–900 °C. They found similar enhancements for B and P, and these were larger than the enhancements for As. They have proposed that the diffusion enhancements they observe are due to the split silicon vacancy, which diffuses with a migration enthalpy of 1.5 eV. This conjecture is not easily supported by other experiments. A more reasonable interpretation of these data has been given by Gösele, Frank, and Seeger (1979), who propose that the data indicate $H_I^m \geq 1.5$ eV, if one also assumes that P and B interact more strongly with I-type defects than does As. This last assumption has strong experimental support from many sources (as discussed in Sec. XVII). In separate studies, Lucas *et al.* (1979) and Loualiche *et al.* (1982) investigated dopant diffusion during high-temperature proton irradiation. The theory used to analyze their data has many fitting parameters and relies on many assumptions. They propose that the migration enthalpy of boron by an I-type mechanism is about 1.2 eV in the temperature range of 500–800 °C.

In a completely different type of experiment, Seeger, Föll, and Frank (1977) analyzed swirl defect patterns, which are interstitial in nature, and deduced that H_I^m is between 1.6 and 1.7 eV. The analysis of Morehead (1987), alluded to in Sec. V.B, which utilizes combined data from Pt and Au diffusion studies, yields an expression for the interstitial diffusivity d_I of

$$d_I = 0.2 \exp(-1.2 \text{ eV}/kT) \text{ cm}^2\text{sec}^{-1} , \tag{7.3}$$

where d_I is defined by the Fick's law expression for interstitial flux J_I is

$$J_I = -d_I \frac{\partial C_I}{\partial x} . \tag{7.4}$$

There is yet another type of experimental approach to determining point-defect diffusivities, which has been utilized with increasing frequency in recent years. This approach is to try to deduce the values of d_I or d_V by measuring the transit time of point defects across a known thickness of silicon. Somewhat surprisingly, interpretation of the data obtained from these experiments is not as straightforward as one might initially expect. We briefly mention the origin of these difficulties in the following section and defer a detailed discussion of experimental results until Sec. XVI.B.

C. Charge-state-dependent diffusivities

Because the migration rate of a given type of point defect may depend on its charge state, it is pertinent to ask how the experimentally determined diffusivity is related to the individual diffusivities of the defects in their different charge states. Using interstitials as an example, we imagine an experimental situation (to be dealt with in some detail in Sec. XII) where I are injected from one of the surfaces into the bulk. Let us suppose for simplicity that I exist predominantly in only the I^0 and I^+ charge states. Let us further make the reasonable assumption that the electronic charging process is quick enough so that the ratio of C_{I^+}/C_{I^0} at each point in the interstitial concentration profile is given by Eq. (6.3). Then, if we assume that diffusion of both I^0 and I^+ can be described by a Fick's law expression like Eq. (7.4), it is a simple matter to show that the flux of the total I concentration, $J_I^{\text{total}} = J_{I^0} + J_{I^+}$, is given by

$$J_I^{\text{total}} = d_I^{\text{eff}} \frac{\partial C_I^{\text{total}}}{\partial x} , \tag{7.5}$$

where

$$d_I^{\text{eff}} = d_{I^0} \frac{C_{I^0}}{C_{I^0} + C_{I^+}} + d_{I^+} \frac{C_{I^+}}{C_{I^0} + C_{I^+}} . \tag{7.6}$$

Further discussion of effective diffusivities related to ionization effects can be found in the paper by Bourgoin, Corbett, and Frisch, (1973). The extension to other charge states and the case for vacancies is obvious, but there are many realistic situations where Fick's law is not an adequate formulation of the problem. For example, in Sec. XII we consider the effects of recombination between I and V, and in Sec. XII.E the effects that impurities have as either traps or recombination centers. Only under certain conditions can the flow of excess point defects be described by a simple Fick's law formulation. We also show in Sec. XIII the complexities introduced by considering point-defect diffusion across nonhomogeneous extrinsically doped layers.

VIII. FORMATION OF DOPANT DEFECTS

Although it is an experimental certainty that dopant atoms dissolve in the silicon lattice almost completely on substitutional lattice sites, at any given instant of time some finite fraction will exist in a dopant-defect state, as discussed in Sec. III. Only this fraction of dopant atoms is available to take part in diffusion, unless one supposes that dopant atoms spontaneously exchange lattice positions with neighboring silicon atoms. It is of interest to calculate the probability that at any instant of time a dopant atom is in an AX state, corresponding to the cases in which a dopant atom is next to a vacancy, next to an interstitial, or in the dopant-interstitialcy state. Derivations of the kind that led to Eq. (5.1) can be performed. The results of such a derivation (Lidiard, 1960; Hu, 1969) show that the concentration of AX defects is given by

$$C_{AX} = \theta_{AX} \frac{C_A C_X}{C_S} \exp\left[\frac{E_{AX}^b}{kT}\right] , \tag{8.1}$$

where E_{AX}^b is defined to be the binding energy of the AX defect, and the θ_{AX} factor takes into account the number of equivalent ways of forming the AX defect at a particu-

lar site (e.g., $\theta_{AV}=4$). In Eq. (8.1), C_X represents the isolated—i.e., unassociated—concentration of defect X, so that its value can be determined from the relations in Sec. V.[3] For an attractive potential (i.e., positive E^b_{AX}), Eq. (8.1) states that the formation energy of a point defect is lower in the vicinity of a dopant atom. For example, the formation energy of a vacancy or interstitial is lower in the sites next to a dopant atom. Since some interaction potential (repulsive or attractive) must exist between A and X, it necessarily follows that the formation energies of point defects are affected by the presence of the dopants in addition to the defect-charging effects discussed in Sec. VI.

Equation (8.1) will be referred to many times in the following discussions. Therefore, a brief discussion of the limits of its validity is appropriate.

A. Limitations of the dilute concentration approximation

In the dilute concentration approximation, the number of point defects in the crystal is assumed to be much smaller than the number of lattice sites. But since dopant atoms are themselves point defects and can be introduced into the silicon substrate to arbitrarily high levels (for example, by ion implantation), the dilute concentration approximation will break down for heavy doping levels. Experiments are always performed under the condition that $C_A \ll C_S$, so that reduction of available lattice sites by simple substitution is never a factor. In fact, it is observed experimentally for all of the dopants that above a certain concentration of A (which is always much less than C_S) the dopant atoms will no longer dissolve completely on substitutional sites. The classical interpretation of this phenomenon is that the solid solubility of the silicon has been exceeded, resulting in the precipitation of the dopant in a second phase. This is discussed further in Sec. VIII.A.1.

Another question concerning the dilute concentration approximation is whether the concentration of AX defects might actually exceed the isolated A concentration. Hu, Fahey, and Dutton (1983) have shown that, from basic experimental results and simple theoretical reasoning, this situation will never occur.

More important questions concerning the dilute concentration approximation arise when we consider what fraction of the point-defect concentration may be viewed as unassociated with dopant atoms. If the majority of point defects are associated with dopant atoms, then several common assumptions that are usually made in analyzing diffusion processes will no longer be valid. A detailed examination of this situation is presented in Sec. VIII.A.2

[3]We should keep in mind that, under extrinsic doping conditions, C_A and C_X are not independent for a charged defect X.

Furthermore, it is important to recognize that Eq. (8.1) gives only the number of point defects that are nearest neighbors to a dopant atom and therefore must underestimate the concentration of point defects associated with dopant atoms. If the dopant concentration is high enough, the number of available lattice sites at which a point defect can be considered unperturbed by the presence of dopant atoms will begin to diminish significantly. As the doping concentration is increased further, at some point the average spacing between dopant atoms will become small enough that all point defects in the crystal are affected by the presence of the dopant atoms. At what concentration this occurs depends of course on the spatial extent of the dopant–point-defect interaction potential. Lidiard (1960) considered the case in which the dopant-vacancy interaction potential extended only to a nearest-neighbor distance. Hu (1969) extended the treatment of Lidiard to include the case of vacancies in different charge states and the possibility of partial ionization of the dopants. The results of their analyses show that the total number of vacancies that can be considered unassociated with dopant atoms is given by just the number that one would calculate for a lattice with only $N_S-(Z+1)N_A$ sites (here N represents number and Z nearest-neighbor number). That is, the sites occupied by A and the Z nearest neighbors of A have been excluded as possible sites for occupation. If the dopant-vacancy interaction potential extends beyond the nearest-neighbor sites, then these other sites would also be excluded in the calculation. Thus the modification of C_S in this manner in Eq. (6.5) is the first-order correction for the dilute concentration approximation. An approach that takes into account the exclusion of more sites has been offered by Lannoo and Bourgoin (1981). Mathiot and Pfister (1982) have calculated that for an interaction potential between A and X that is effective at a third-neighbor site, when dopant concentrations reach about 3×10^{20} cm^{-3} ($=0.6$ at. % in silicon) all of the defects X can be considered associated with dopant atoms. This phenomenon is presented in more detail in Sec. VIII.A.3.

1. Precipitation and clustering

Experimentally, it is known that above a certain concentration the dopants are electrically inactive. Experimental values are indicated in Fig. 5 (the other meaning of the figure is explained in the section immediately following this one). There are clearly two important purposes for developing models that take into account the electrical inactivation of dopant concentrations under heavy doping conditions: (i) to predict the fraction of dopants that are electrically active, and (ii) to predict profile movements as a result of diffusion.

The classical interpretation of these heavy doping effects is that at any given temperature only a limited concentration of dopants will dissolve substitutionally in the silicon matrix before the well-known phenomenon of precipitation occurs. It is now well established that pre-

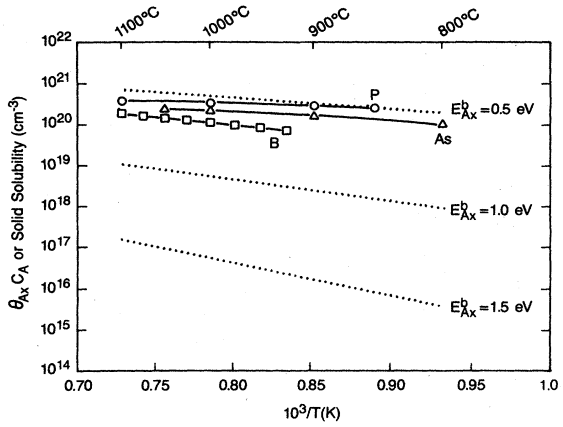

FIG. 5. Solid solubility limits (solid lines and symbols) for P (Nobili, Armigliato, Finetti, and Solmi, 1982), B (Armigliato, Nobili, Ostoja, Servidori, and Solmi, 1977), and As (Nobili, Carabelas, Celotti, and Solmi, 1983). The dashed lines are lower estimates of the dopant concentration at which the number of AX defects, as given by Eq. (8.1), is equal to the number of isolated X point defects and is valid below the solubility limits.

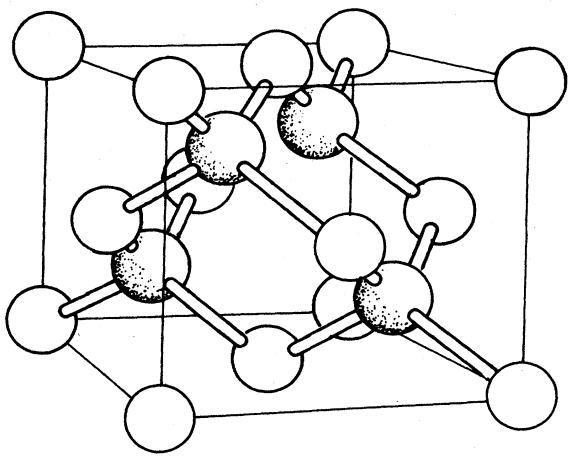

FIG. 6. Possible configuration of an As cluster involving four As atoms (shaded spheres). From Hu (1973a).

cipitation is responsible for the electrical inactivity of boron (Armigliato, Nobili, Ostoja, Servidori, and Solmi, 1977) and phosphorus (Nobili, Armigliato, Finetti, and Solmi, 1982). The inactivation of Sb under most common experimental conditions has also been attributed to precipitation (Guerrero, Pötzl, Stingeder, Grasserbauer, Piplitz, and Chu, 1985). It has been shown, however, that annealing Sb-implanted samples at temperatures too low for diffusion to occur (which is necessary for precipitates to form) can still result in the deactivation of Sb by the formation of some type of Sb-vacancy complex (Nylandsted Larsen, Pedersen, Weyer, Galloni, Rizzoli, and Armigliato, 1986).

Precipitation phenomena very similar to that found for P have been shown to occur for As (Nobili, Carabelas, Celotti, and Solmi, 1983). However, there is a competing theory to account for electrically inactive As, known as *clustering* (Hu, 1973a), in which it is proposed that multiple As atoms form some new configuration (perhaps involving point defects such as As-vacancy clusters), which is electrically inactive at room temperature. A proposed tetratomic cluster is depicted in Fig. 6. Precipitates may contain many thousands of dopant atoms; the size distribution of precipitates is a function of the initial degree of supersaturation above solid solubility and the thermal treatment given to the sample. In contrast, it is proposed that clusters are composed of a few dopant atoms in specific configurations. Clusters exist in equilibrium with isolated dopant atoms just as AX defects coexist with isolated A, whereas precipitates are regions of the crystal that have formed a second phase of the solvent and solute constituents. All researchers agree, though, that at a high enough concentration, As precipitation will eventually occur. The identification of electrically inactive As with clusters or SiAs precipitates is still debated. Most researchers who have tried to model As diffusion profiles quantitatively seem to favor the clustering explanation. Arguments in favor of precipitation have recently been summarized by Nobili (1983).

One of the most important pieces of experimental evidence concerning the two possible explanations is the fact that, for a given diffusion temperature, the carrier concentration saturates above a certain value of As concentration. In other words, the amount of electrically active As has a temperature-dependent maximum that does not depend on the total As concentration above this limit. This behavior is to be expected in a precipitation model for equilibrium between two phases (i.e., precipitates and the As dissolved in Si). On the other hand, a clustering model in general predicts that there be some mass-action relation between the isolated substitutional As and the As in a clustered state, so that there is a functional relationship between the total As concentration and the electrically active As concentration. It has been shown (Guerrero, Pötzl, Tielert, Grasserbauer, and Stingeder, 1982) that a cluster model can be reconciled with experimental data only if it has the following properties. If the cluster contains m As atoms, then at the high temperatures where the clusters form (typically this means $T \geq 800\,^{\circ}\mathrm{C}$), the cluster has a positive charge of $m-1$, and at room temperature (where the carrier concentration is measured) the cluster of m As atoms is electrically neutral. This means that at temperatures high enough for significant diffusion to occur, $(m-1)/m$ of the clustered As atoms still denote electrons. Considering the effects of electronic charging on point-defect concentrations, as presented in Sec. VI, a big difference between clustering and precipitation models for As is that in the precipitation models the electron concentration saturates at the values indicated in Fig. 5, independent of total As concentration, whereas in the clustering models the electron concentration (at the diffusion temperature) will continue to increase with total doping concentration. At present, there are still no unambiguous diffusion ex-

periments that have differentiated between clustering and precipitation models, although one would expect that the differences in predicted electron concentrations at diffusion temperatures would lead to different simulated results for profile movement. This question has not been directly addressed in the literature. In fact, no specific model of precipitation has yet been proposed other than to assume that the fraction of dopants above the electrical solubility is immobile. (In reality, of course, precipitates continuously form and dissolve by dopant atoms joining and leaving the precipitates by diffusion.)

Two more pieces of experimental data have recently been reported which have attempted to reveal the atomistic identity of electrically inactive As. One is the detection by transmission electron microscopy of very small (15–30 Å) As precipitates (Armigliato, Nobili, Solmi, Bourret, and Werner, 1986), although in concentrations too low to account for all of the electrically inactive As. The other is the use of extended x-ray-absorption fine-structure spectroscopy (EXAFS) of heavily As-doped silicon (Erbil, Cargill, and Boehme, 1985; Erbil, Weber, Cargill, and Boehme, 1986), indicating localized regions of Si and As on alternating lattice sites, similar to the arrangement of Ga and As atoms in GaAs. While proponents of clustering models would interpret the microscopy results as evidence in support of the contention that clustering accounts for most of the inactive As, proponents of the precipitation model would argue that the findings of the EXAFS study suggest coherent precipitates rather than clusters of As atoms, and that this coherency is the reason why very small precipitates cannot be detected by electron microscopy. It should not go unnoticed, however, that the tetratomic cluster pictured in Fig. 6 has just the structure of As existing on alternating lattice sites (i.e., as next nearest neighbors) that was found in the EXAFS investigation. Such experimental data are also consistent with recent total-energy calculations, which indicated that a vacancy-As_4 cluster is an energetically favored defect complex (Pandey, Erbil, Cargill, and Boehme, 1988). A coherent precipitate might just be a larger version of such a cluster (i.e., involving many more As atoms), in which case one might ask whether clusters are the embryos of SiAs precipitates and clustering and precipitation occur together with one process dominant over the other depending on the experimental conditions.

2. Concentrations of associated and unassociated point defects

Although it is seldom discussed in the literature, a very important consideration of dopant-atom–point-defect interactions is whether the majority of point defects that exist in thermal equilibrium are associated or unassociated with dopant atoms. The answer to this question is important for two major reasons. First, if the concentration of point defects associated with dopant atoms is greater than the unassociated concentration, then in a kinetic (versus ensemble) picture it necessarily follows that, when diffusing through the crystal, the point defects spend most of their time next to dopant atoms. Therefore the experimentally measured point-defect diffusivity can depend on the concentration and nature of the dopant species; this situation is analyzed in detail in Sec. XII.E.1. Second, it is possible to selectively create either I or V defects as a result of chemical reactions at the silicon surface. As an example we take the case of I injection into the silicon substrate. Once injected into the bulk, the excess I may recombine through one of the reactions

$$I+V\rightarrow 0,\quad I+AV\rightarrow A,\quad I+A\rightarrow AI\ . \tag{8.2}$$

The recombination rate of the excess I concentration can be very much affected by the presence of AV defects, depending on the relative efficacy of the above reactions and the relative concentrations of AV and V defects. (This topic is discussed in more detail in Sec. XII.E.) The trapping of I by dopant atoms will slow down the diffusive flow of I through the wafer (Sec. XII.E).

Given the possible importance of the above phenomena under realistic experimental conditions, it is desirable to estimate at what dopant concentrations one might expect these effects to manifest themselves. As mentioned at the beginning of Sec. VIII.A, this estimation, of necessity, relies on a knowledge of the spatial extent of the A-X interaction potential. However, a reasonable estimate of the doping levels at which the heavy doping phenomena outlined above may become important can be made using Eq. (8.1) and defining the critical dopant concentration as that value of C_A at which $C_{AX}/C_X=1$. This occurs for the condition

$$\theta_{AX}C_A=C_S\exp\left[-\frac{E^b_{AX}}{kT}\right]\ . \tag{8.3}$$

A plot of $\theta_{AX}C_A$ vs $1/T$ is shown in Fig. 5 for binding energies of $E^b_{AX}=0.5$, 1.0, and 1.5 eV. The dashed straight lines are for $C_S=5\times10^{22}$ cm^{-1}, which is appropriate for AV pairs and interstitialcy dopants. Also shown are the solid solubilities of B, P, and As in silicon. Since $\theta_{AX}\geq 1$, and the number of defects associated with A will in general be greater than the number predicted by Eq. (8.1), the value of C_A at which most point defects can be considered associated with A will be *less* than the values indicated in Fig. 5.

This figure demonstrates that dopant-defect association effects can manifest themselves at relatively low dopant concentrations if E^b_{AX} is large enough or kT is low enough. This important possibility and its consequences have not been recognized in the literature.

The nature and magnitude of possible A-X interaction potentials are discussed in Sec. VIII.B.

3. Percolation phenomenon

The discussion in VIII.A.2 that led to Fig. 5 concerned the heavy doping effect where *most* of the point defects

are associated with dopant atoms. Mathiot and Pfister (1982) have discussed the case in which *all* of the point defects are associated with dopant atoms. The only way in which this may happen is if the dopants are spaced so closely together that there are no lattice sites far away enough from any of the dopant atoms that the point defects may be considered unassociated. Figure 5 shows, on the other hand, that for large enough binding energy E^b_{AX} it is still possible to have most of the point defects associated with dopant atoms even if the spacing between atoms is relatively great.

The situation in which all of the point defects are associated with dopant atoms is depicted in Fig. 7. Following Mathiot and Pfister (1982), we picture in Fig. 7(a) the path a vacancy defect follows moving between two dopant atoms, labeled A and A' (similar schematic plots can be drawn for interstitialcy and interstitial diffusion mechanisms). A plot of the vacancy potential as a function of position in the lattice is shown in Fig. 7(b). It can be seen that, as the vacancy diffuses away from A, its potential is lowered by the presence of A'. If we have many dopant atoms distributed at such close spacings throughout the lattice, then the vacancy formation energy everywhere is lowered by the presence of the dopant atoms.

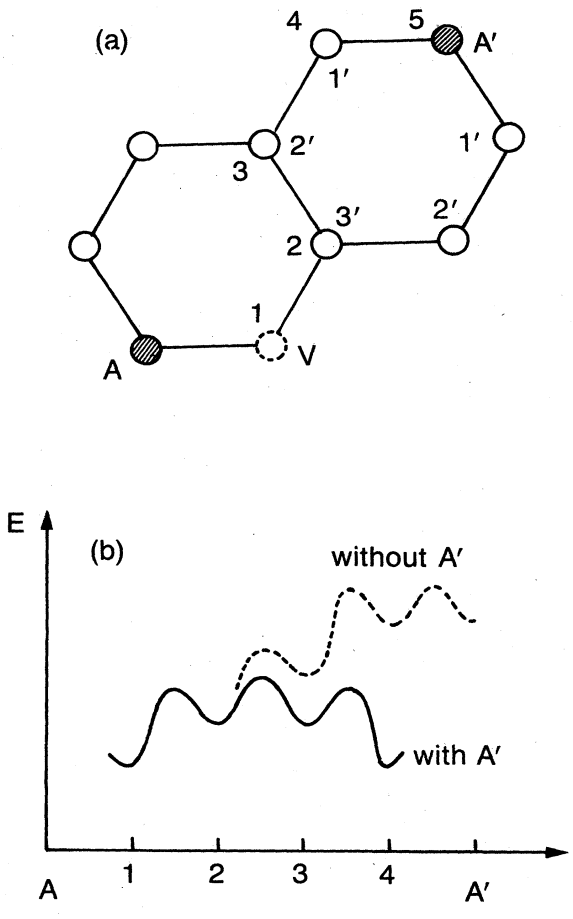

FIG. 7. Percolation phenomenon for dopant diffusion by a vacancy mechanism. After Mathiot and Pfister (1982).

Treating the situation depicted in Fig. 7 is a very difficult task. Clearly in such a situation we have a system of interacting particles and thus have to abandon one of the most fundamental assumptions in the more established treatments of diffusion processes: that the diffusing particles execute diffusion jumps independently. Thus, for example, it is certainly the case that

$$J_V \neq d_V \frac{\partial C_V}{\partial x} \tag{8.4}$$

in Fig. 7. This departure from Fick's law occurs already for the situation of $C_{AX} \geq C_X$ depicted in Fig. 5, but at least is amenable to some formal treatment (Hu, 1969).

The physical situation shown in Fig. 7 has actually been realized experimentally at Consiglio Nazionale delle Ricerche, Istituto Lamel, in Bologna (although the dominant point defects are not necessarily vacancies). Researchers implanted into silicon high doses of either P (Nobili, Armigliato, Finetti, and Solmi, 1982) or As (Angelucci, Celotti, Nobili, and Solmi, 1985) and subsequently used laser annealing to activate the dopants. The high concentrations of substitutional dopants thus achieved (approaching 10 at. %) ensure that at the onset of the thermal anneals, which followed the laser annealing, the initial condition of the silicon is something like Fig. 7. The basic phenomenon observed is a very rapid deactivation of the dopant atoms into precipitates with no anomalously high, long-range diffusion of dopants. Since the dopants must diffuse in order to come together and form precipitates, perhaps a percolation situation such as Fig. 7 allows for an enhanced rate of homogeneous nucleation of the precipitates because of the large concentrations of point defects present. But this should only be regarded as a speculative comment, since no formal treatment of the problem is yet available.

Some attempts to model the percolation situation of Fig. 7 by a traditional approach to the problem have been made by Mathiot and Pfister (1982, 1983); however, the predicted results are at variance with the experiments cited above.

B. Nature of the dopant-defect interaction potential

1. Experimental determination of E^b_{AX}

Hirata, Hirata, and Saito (1969) studied the different annealing behavior of irradiated samples that contained group-V dopants: P, As, Sb, and Bi. Comparing their results with other works for P, As, and Sb (Elkin and Watkins, 1968) and P (Kimmerling, DeAngelis, and Diebold, 1975) leaves little room for doubt that they have observed the annealing of AV pairs. Hirata, Hirata, and Saito (1969) reported the following lower-limit estimates for binding energies:

$$E^b_{PV} \geq 1.04 \text{ eV}, \quad E^b_{AsV} \geq 1.23 \text{ eV}, \\ E^b_{SbV} \geq 1.44 \text{ eV}, \quad E^b_{BiV} \geq 1.64 \text{ eV}. \tag{8.5}$$

These resulted from annealing at temperatures of 413–433 K. It is most likely that these energies correspond to the binding between A^+ and V^-. There are no equivalent experimental results for binding between A and I.

2. Coulombic interactions

If we accept for the moment that a simple Coulombic potential exists between an ionized dopant atom and a charged point defect, then we have

$$\Delta E_{\text{Coul}} = \frac{q_1 q_2}{\varepsilon_{\text{Si}} r}, \tag{8.6}$$

where q_1 and q_2 are the charges on the two interacting species, r the distance between them, and ε_{Si} the macroscopic dielectric constant of Si ($\varepsilon_{\text{Si}}=11.7$). For two unit charges of opposite sign this gives a potential that falls off at a rate of 1.22 eV/r Å, resulting in a binding energy of 0.52 eV for unit charges spaced a nearest-neighbor distance apart (2.35 Å in silicon).

There are a few problems with the concept of a Coulombic potential applied to a dopant atom and point defect interacting in proximity. First, it is clear that the macroscopic dielectric constant ε_{Si} is not appropriate for close-range dopant-defect interactions (Hu, 1973a). Second, the charge associated with point defects cannot necessarily be treated as a fixed ionic core. Consider the interstitialcy dopants, for example, in Fig. 1(c), where there are not two spatially separated charge centers; or consider the vacancy defects in Fig. 1(a), whose charges arise from bound carriers with wave functions that will extend to neighboring sites. The substitution of a dopant atom for one of the four nearest-neighbor silicon atoms in Fig. 1(a) makes evident the conceptual difficulties in trying to describe such close-range interactions with a simple Coulombic interaction potential given by Eq. (8.6).

It is much easier to believe that Eq. (8.6) might be valid when the dopant atom and the charged point defect are separated by many lattice sites. In Fig. 8 we have plotted the interaction potential energy as a function of the distance separating the oppositely charged dopant and point defect. Also included in the figure is a temperature axis corresponding to the temperature at which $\Delta E_{\text{Coul}} = kT$. Thus, for a given value of temperature, the plot gives the capture radius of the dopant atom for a point defect of equal and opposite charge. When the temperature is relatively low, the capture radius is large and Eq. (8.6) is expected to be valid. But one can also see that the magnitude of the interaction energy in Fig. 8 never approaches the data of Hirata, Hirata, and Saito (1969) listed above, and, in fact, even the *difference* between E^b_{PV} and E^b_{BiV} is greater than the highest values of ΔE_{Coul} that might be

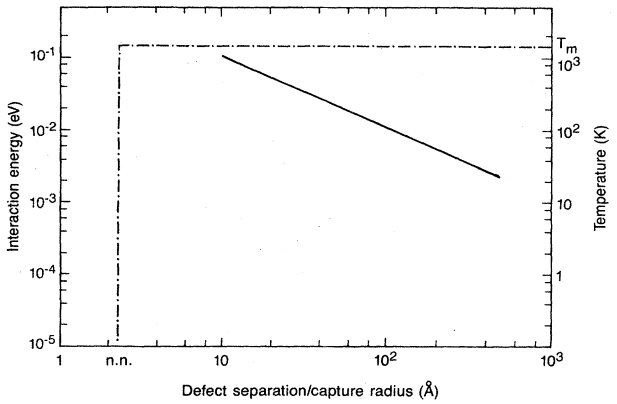

FIG. 8. Coulombic interaction energy between two point charges, assuming macroscopic dielectric constant of silicon and the corresponding temperature of equal kT, as a function of distance. Horizontal line is the melting point of silicon. Vertical line is the nearest-neighbor separation in silicon. After Lannoo and Bourgoin (1981).

expected to be accurate. Even assuming a doubly charged point defect will not provide an energy high enough to be consistent with these data. This suggests that some close-range interaction that is non-Coulombic in nature may account for the binding between dopants and point defects.

3. Non-Coulombic interactions

Some distortion of the lattice must occur in accommodating either native point defects or dopant atoms. Most previous discussions on the subject of non-Coulombic interactions between dopant and point defects have focused on the elastic interactions arising from the difference in tetrahedral bonding radii between the dopant and silicon. The experimental justification for this approach is based on the observation that the covalent bonding length between two atoms (as measured by x-ray analysis) is very nearly the same even when measured in different crystals. Values of tetrahedral bonding radii for atoms given by Pauling (1960) are shown in Fig. 9, which was reproduced from Pauling's book.

Those atoms with shorter bonding radii than silicon, $r_{\text{Si}}=1.17$ Å, are commonly said to be "smaller" than silicon. Dopants in this category are P and B ($r_{\text{P}}=1.10$ Å, $r_{\text{B}}=0.88$ Å). All other dopants (Al, Ga, In, As, Sb, and Bi) are "larger" than silicon ($r_A=1.26$, 1.26, 1.44, 1.18, 1.36, 1.45 Å, respectively). One readily observable effect of the "size" of the atom is the change in lattice parameter for silicon with increasing concentrations of dissolved impurities. Thus smaller atoms such as P and B are found to contract the silicon lattice (McQuhae and Brown, 1972; Celotti, Nobili, and Ostoja, 1974; Herzog, Csepregi, and Seidel, 1984), while the larger atoms, Ge (Dismukes, Ekstrom, and Paff, 1964; Herzog, Csepregi, and Seidel, 1984), As (Bublik, Gorelik, and Dubrovina,

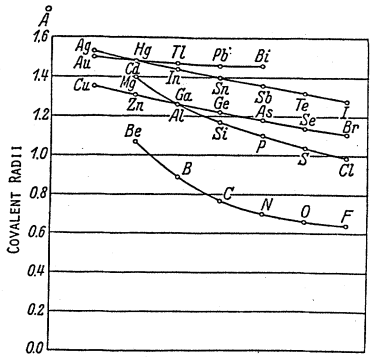

FIG. 9. Values of tetrahedral covalent radii for sequences of atoms. From Pauling (1960).

1968), and Sb (Teague, Yagnik, Long, Gerson, and Lafleur, 1971), are all found to dilate the silicon lattice. It is easy to imagine that an oversized dopant atom would find more elastic relief in the vicinity of a vacancy point defect than an interstitial point defect, and, conversely, that an undersized dopant atom would naturally prefer an interstitial point defect. If we look at the E_{AV}^{b}'s listed in Eq. (8.5), it can be seen that the ordering $E_{PV} < E_{AsV} < E_{SbV} < E_{BiV}$ is in qualitative agreement with this argument, since the ordering of the covalent radii are $r_P < r_{As} < r_{Sb} < r_{Bi}$. However, based on the experimental evidence discussed in Sec. XVII, it appears that Ga, Al, and In interact preferentially with interstitial defects over vacancies, even though their covalent radii are all larger than that of silicon.

The real problem in using size arguments to discuss point defects and diffusion may be that the tetrahedral bonding radius is simply not the correct factor to use in quantitative arguments. For example, calculations based on the elastic distortions that result because of the different covalent radii of the dopants in the lattice generally indicate that the extent of the distortion will be short range in nature, falling off as r^{-3} to r^{-5} (Hu, 1973a). However, these treatments do not take into account the distortion introduced by the associated point defect itself. The distortion introduced by the point defect will in general depend on its charge state. If the distortion extends to nearest-neighbor sites and beyond, the interaction distance will obviously be greater than that predicted by considering only the dopant atom. Furthermore, while the local distortion introduced by unassociated I or V, or isolated substitutional dopant atoms, may sometimes be treated effectively using simple models, when dopants and point defects are in proximity the resulting local state of the silicon lattice may not be correctly described by adding the separate properties of the isolated defects. Thus the effective interaction distance and magnitude of the binding energy might be greater or less than that inferred from analyses of the isolated dopant and point defect.

The most sophisticated theoretical treatment of the dopant-defect interaction potential has been presented by Car, Kelly, Oshiyama, and Pantelides (1985) using total-energy calculations. Their results for the binding energy of PV pairs are

$$E_{P^+V^0} = 1.8 \text{ eV}, \quad E_{P^+V^-} = 2.25 \text{ eV} . \tag{8.7}$$

The calculated value of P^+V^0 is greater than the lower limit in Eq. (8.5) and is thus consistent with experiment. But the theoretical treatment indicates that the non-Coulombic contribution to E_{PV} can be large and dominant over the Coulombic component. One can see from Fig. 5 and the discussion in Sec. VIII.A.2 that, if the value of E_{PV} were as large as that predicted by the total-energy calculations, almost all of the V defects would be associated with P atoms even for very low concentrations of P. If the trend of binding energies followed those indicated in Eq. (8.5), then we could conclude the same for Sb, As, and Bi at even lower concentrations.

It can be concluded that local distortions of the lattice arising from the differences in bonding lengths for dopant and silicon atoms is a real effect. However, in determining the non-Coulombic interaction between point defects and dopant atoms, this can only be considered an indicative factor, not necessarily the one upon which to build a model. Distortions introduced by the point defects themselves and their possible charge-state dependencies must be considered. When point defects and dopant atoms are in intimate contact, the resulting local state of the lattice may be quite different from what would be deduced from their separate lattice properties.

IX. SELF-DIFFUSION: EXPERIMENTAL STUDIES

Self-diffusion studies are of great interest in the investigation of point-defect interactions with dopant atoms, since self-diffusion can be viewed as the limiting case of dopant diffusion, in which dopant atoms introduce no distortion in the lattice and carry no excess charge.

One of the most common techniques for measuring self-diffusion rates is to monitor the diffusion of radioactive silicon isotopes. Allowing for both I and V components of diffusion, the diffusivity of the radioactive tracers is given by (Hu, 1985a)

$$D_{self} = D_{SiI} + D_{SiV} \tag{9.1}$$

$$= (\phi_I + 1) d_I \frac{C_I}{C_S} + (\phi_V + 1) d_V \frac{C_V}{C_S} , \tag{9.2}$$

where d_I and d_V are the diffusivities of I and V, respectively, C_S is the number of lattice sites, and ϕ_I and ϕ_V are the correlation factors of diffusion for each mechanism. It is important to note that Eq. (9.2) was only recently shown to be the correct expression for self-diffusion measured by tracer diffusion (Hu, 1985a). Previously, Eq. (9.2) with the $(\phi_X + 1)$ factors replaced by ϕ_X was the accepted expression.

Almost all self-diffusion studies show a good fit to an

Arrhenius expression

$$D_{\text{self}}=D^0_{\text{self}}\exp\left[-\frac{Q_{\text{self}}}{kT}\right]. \tag{9.3}$$

If one mechanism of diffusion is dominant over the other, then from the discussion in Secs. V and VII it is easily seen that

$$Q_{\text{self}}=H^m_X+H^f_X, \tag{9.4}$$

so that the experimentally determined value of Q_{self} is simply the sum of the migration and formation enthalpies of the point-defect species responsible for self-diffusion. Interpretation of D^0_{self} is not so straightforward. The components can be broken up as

$$D^0_{\text{self}}=\theta_X(\phi_X+1)d^0_X\exp\left[\frac{S^f_X}{k}\right]. \tag{9.5}$$

The d^0_X factor must include lattice vibration and entropy of migration terms, but the theoretical understanding of what constitutes this factor is too poor to make any definitive conclusions based on experimental results. Still, the large values of D^0_{self} compared to the case of self-diffusion in metals was the original reason that the dominance of extended point defects in self-diffusion of silicon was proposed (Seeger and Chik, 1968).

To interpret correctly the results of self-diffusion experiments, it is critical to know which mechanism or combination of mechanisms is responsible for self-diffusion. Arguments have been put forth in favor of a vacancy mechanism (Van Vechten, 1974), an interstitial mechanism (Car, Kelly, Oshiyama, and Pantelides, 1985), an interstitialcy mechanism (Seeger and Chik, 1968; Frank, Seeger, and Gösele, 1981), and, recently, a mechanism that involves no point defects at all (Pandey, 1986). At present, this issue is still unresolved. Even though self-diffusion studies have not unambiguously identified what mechanism is responsible for diffusion, at the very least it is recognized that any proposed models of point defects and dopant diffusion must be consistent with the data from the experimental investigations.

One of the most important results of self-diffusion experiments is that the activation energy of self-diffusion is approximately 1 eV greater than the activation energies of dopant diffusion. Arrhenius expressions for D_{self} from tracer experiments under intrinsic doping conditions have been reported by a number of investigators. The values of D^0_{self} and Q_{self} from these experiments do not show good agreement between different investigations, especially when compared to expressions for D_{self} in germanium, which show rather good agreement [see the review article by Frank, Gösele, Mehrer, and Seeger (1984)]. In part, this may be due to the fact that oxygen in silicon, which is now known to affect point-defect behavior in the bulk (Mizuo and Higuchi, 1982b), was an uncontrolled factor in early investigations. Carbon content in silicon may also prove to be a factor. On the other hand, a rather good fit to an Arrhenius expression for D_{self} is found for almost all investigations. In the temperature range 1050–1300 °C, values for Q_{self} range from 4.7 to 5.1 eV and values for D^0_{self} from 900 to 9000 $\text{cm}^2/\text{sec}^{-1}$.

Equation (9.1) indicates that over an extended temperature range a non-Arrhenius behavior may be observed if more than one type of self-diffusion mechanism is operative. In practice it is difficult to determine D_{self} over a wide temperature range using a radioactive tracer, because the only isotope that is readily obtainable is ^{31}Si, which has a half-life of only 2.6 h. To try to circumvent this limitation, three other measurement techniques have been developed: use of Ge as a tracer; a nuclear profiling technique using implanted ^{30}Si; and monitoring of heavy-metal diffusion with some model assumptions.

Since Ge is in the same column of the Periodic Table as Si, dissolves substitutionally with no excess charge, and has a covalent radius close to that of Si (see Fig. 9), the diffusion of Ge in silicon was investigated with the thought that it may be similar to self-diffusion (McVay and DuCharme, 1973, 1975).[4] From the analysis of Sec. XI, it can be seen that there is good reason to expect the activation energy of diffusion by a vacancy mechanism to be close for Si self-diffusion and Ge in Si, whereas the situation is less clear for the comparison of Ge and self-diffusion occurring by I-type mechanisms. Utilizing ^{71}Ge as the tracer diffusant has the advantage of a relatively long half-life (11.2 days) compared to ^{31}Si and thus allows for the possibility of diffusion measurements at lower temperatures. Hettich, Mehrer, and Maier (1979) measured Ge diffusivities over the extended temperature range 875–1300 °C. Measurements of Ge diffusivity by Ogino, Oana, and Watanabe (1982) between 1100 °C and 1300 °C agree well with the results of Hettich *et al.*, and both results show good agreement with D_{self} measurements in the same temperature range. The later Ge diffusion measurements of Dorner, Gust, Predel, and Roll (1984), however, show higher diffusivities than the previous two works. In addition, whereas the measurements of Hettich *et al.* show a non-Arrhenius behavior over an extended temperature range (875–1300 °C), the measurements of Dorner *et al.* show only a single activation energy of diffusion of 5.35 eV over the same temperature span. The most recent Ge diffusion data of Bouchetout *et al.* (1986) show a break in Arrhenius behavior at about the same temperature as that found by Hettich *et al.* (around 1050 °C); but, only one data point at the lowest temperature, 1000 °C, provided evidence of

[4] In this regard, it is of interest to note that results of tracer measurements for C, which like Ge carries no net charge, but whose bonding radius is much smaller than that of Si, are very different from results found for Si and Ge tracers. Measurements of D_C by Newman and Wakefield (1962) using ^{14}C tracer give a value of $1.9\exp(-3.1\ \text{eV}/kT)\ \text{cm}^{-2}\text{sec}^{-1}$. Kalejs, Ladd, and Gösele (1984) have shown C to have a large component of diffusion corresponding to an I-type mechanism.

non-Arrhenius behavior. It should also be noted that there is excellent agreement between the study of Bouchetout *et al.* and Dorner *et al.* for the values of D_{Ge} measured in the temperature range 1100 °C–1200 °C. While there is some discrepancy between these studies, within experimental error Ge appears to have a substantially higher activation energy of diffusion than dopant diffusion in the high-temperature region ($\geq$ 1100 °C) with values close to those found for self-diffusion. Recent results from Fahey, Iyer, and Scilla (1989) are of great importance to the question of the validity of using Ge tracer diffusion as a substitute for Si self-diffusion tracers. These investigators demonstrated that at 1050 °C, the approximate break point in the studies showing non-Arrhenius behavior, Ge diffusion is enhanced by *both* interstitial and vacancy injections. If future work can further justify the use of Ge tracers as a means of measuring Si self-diffusion, then it can be concluded from the work of Fahey *et al.* that self-diffusion has both interstitial and vacancy components of diffusion, as first proposed by Seeger and Chick (1968); see also the paper by Frank *et al.* (1984). Further work is needed to determine whether one type of diffusion mechanism, interstitial or vacancy, is dominant at temperatures above and below 1050 °C.

A second technique utilizes a nuclear reaction profiling method, (p,γ) resonance broadening, to profile the natural (stable) ^{30}Si isotope in samples ion-implanted with silicon (Hirvonen and Anttila, 1979; Demond, Kalbitzer, Mannsperger, and Damjantschitsch, 1983). These studies have covered the temperature range 830 °C–1200 °C and find activation energies about 4.0–4.2 eV using data only at or below 1100 °C. This is lower than the activation energies of diffusion from the silicon tracer studies (whose data are concentrated in the 1100–1300 °C temperature range). The values of activation energy from the Ge study of Hettich *et al.* also show a similarly low activation energy of 3.9 eV for temperatures between 850 °C and 1000 °C, but, as we mentioned before, this is in sharp disagreement with the Ge study of Dorner *et al.*, which gives a value of 5.35 eV in the same temperature range.

If we ignore the work of Dorner *et al.* for the moment, it would be possible to conclude that Ge diffusion is a valid technique for measuring Si self-diffusion and that, combined with the tracer and (p,γ) data, it reveals that Si self-diffusion has an activation energy of about 5 eV for temperatures greater than 1100 °C and about 4 eV for temperatures less than 1100 °C. But the study of Dorner *et al.* cannot be dismissed so easily, considering that care was taken to use float-zone crystals, a large body of data was taken, and the secondary-ion mass spectroscopy (SIMS) technique and its implementation in this study should be the most accurate of all the other profiling techniques. There is no obvious answer why the Ge diffusion studies of Dorner *et al.* and Hettich *et al.* show such different results. The most important question surrounding the data from the (p,γ) studies centers around the possible role of implantation damage. There is actually a fair amount of experimental evidence available on this subject; high-dose self-implantation has been studied as a means of amorphizing the surface regions of silicon substrates and thereby eliminating unwanted channeling phenomena that occur during B implantation into single-crystal substrates. These studies have shown that the implant conditions used in the (p,γ) studies will initially amorphize the surface region. Subsequent annealing will readily recrystallize this surface layer, but residual ion implantation damage remains. It has been shown that enhanced diffusion of dopants can take place during anneals of samples prepared in this way because of point defects generated from the residual damage (Angelucci, Cembali, Negrini, Servidori, and Solmi, 1987). Previous studies have also shown that the diffusion effects can be a complicated function of the implantation conditions (Solmi, Angelucci, Cembali, Servidori, and Anderle, 1987). It should not be concluded from the above discussion that the (p,γ) studies are unavoidably compromised by the existence of implantation damage; however, the possibility of implantation damages affecting diffusion must be acknowledged.

An alternative technique to tracer diffusion involves monitoring Au diffusion in silicon (Morehead, Stolwijk, Meyberg, and Gösele, 1983; Stolwijk, Schuster, and Hölzl, 1984). Four model assumptions are made: (i) Self-diffusion is dominated by an I-type mechanism. (ii) Long-range Au diffusion is accomplished by a substitutional/interstitial interchange (or "kick-out") mechanism. (iii) Interstitial gold Au_i diffuses much faster than I. (iv) There is a mass-action relation between Au_i, substitutional gold Au_s, and I. This last point has been discussed in detail by Kitagawa, Hashimoto, and Yoshida (1981), with the conclusion that it is a valid assumption.

Recently, Mantovani, Nava, Nobili, and Ottaviani (1986) have shown that diffusion of Pt in silicon can be used in the same way as Au diffusion. They combined their results with the Au diffusion results of Stolwijk, Schuster, and Hölzl (1984) and the ^{31}Si radioactive tracer measurements of Mayer, Mehrer, and Maier (1977) to obtain an expression for D_{self} over the temperature range 700 °C–1385 °C of

$$D_{self}=1400\exp(-5.01\ \text{eV}/kT)\ \text{cm}^{-2}\text{sec}^{-1}\ . \tag{9.6}$$

Morehead (1987) has attempted to fit the combined data from the Au and Pt diffusion studies by computer simulation and reports that the expressions

$$D_{SiI}=4000\exp(-5.0\,\text{eV}/kT)\ \text{cm}^{-2}\text{sec}^{-1}\ , \tag{9.7}$$

$$D_{SiV}=40\exp(-4.6\ \text{eV})/kT)\ \text{cm}^{-2}\text{sec}^{-1}\ , \tag{9.8}$$

give "reasonable" fits to the data and are consistent with the theory that was used to derive them. An Arrhenius plot of $D_{self}=D_{SiI}+D_{SiV}$ using these expressions is presented in Fig. 10.

Recently, experimental results have been reported by

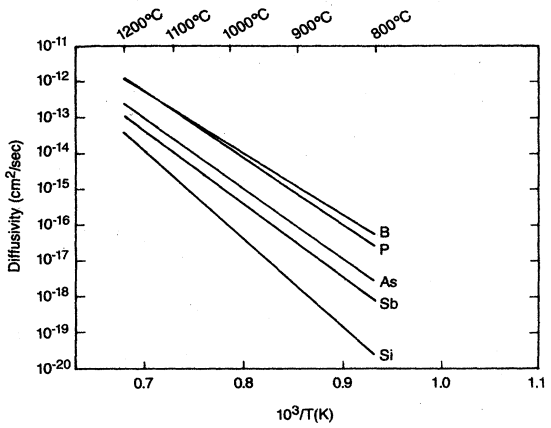

FIG. 10. Experimental values of self- and dopant diffusivity under intrinsic doping conditions. An analytic expression for self-diffusion is given in Eqs. (9.7) and (9.8). Expressions describing dopant diffusivities under intrinsic and extrinsic conditions are summarized in Appendix B.

Aziz *et al.* (1985) on the effect of hydrostatic pressure on self-diffusion. The enhanced diffusion rate observed at high pressure is supporting evidence that self-diffusion is dominated by an I-type mechanism, but more experiments are necessary in order to make definitive conclusions about the mechanism of self-diffusion, since (as pointed out by these investigators) enhanced diffusion under pressure does not in itself prove that an I-type mechanism is dominant over a V-type mechanism. In this regard, it is worth mentioning that, in a related study, Nygren *et al.* (1985) observed enhanced diffusion of As under hydrostatic pressure, and yet the results of all other studies (Secs. XIV and XVII) demonstrate the importance of the vacancy mechanism in effecting As diffusion.

X. EQUILIBRIUM FORMULATION FOR DOPANTS

Since diffusion by its nature is a nonequilibrium process, some qualification must be made to the statement that dopant diffusion occurs under equilibrium conditions. Basically, one can regard diffusion as occurring under quasiequilibrium conditions if there are no external sources of point defects (as outlined in Sec. IV) and there is a mass-action relation between A, X, and AX defects. One of the most important features of diffusion under these conditions is that the nature of the defect species participating in diffusion, I or V, need not be specified to model dopant diffusion, which of course in turn means that equilibrium diffusion experiments cannot distinguish which type of point defect is responsible for diffusion. In spite of this, many investigators incorrectly assume that concentration-dependent diffusion is evidence of a vacancy mechanism. Experiments to determine the fractional I and V components of diffusivity for different dopants are discussed in Sec. XVII.

A. Diffusion under intrinsic conditions

It is assumed that dopant atoms move only when they are in one of their defect states (defined in Sec. III.B): dopant vacancy pair AV, dopant interstitialcy AI, or dopant interstitial A_i. In the following we represent the interstitial-type diffusion mechanism by AI; there is no conceptual difference if A_i is substituted in place of AI.

Ignoring for the moment the charge-state-dependent properties of point defects, the flux of dopant A, which can diffuse by either I- or V-type mechanisms, is given by

$$-J_A = d_{AV}\frac{\partial C_{AV}}{\partial x} + d_{AI}\frac{\partial C_{AI}}{\partial x} , \tag{10.1}$$

where d_{AV} and d_{AI} are the diffusivities of AV and AI defects. The atomistic significance of these terms will be considered in Sec. XI. Here we note only that this is a simple Fick's law formulation, which should be valid for the case of diffusion of dilute dopant concentrations under intrinsic doping conditions. As we indicate below, its validity can be confirmed self-consistently by experiment. We wish to reexpress the flux of A in terms of the experimentally measurable quantity C_A; this requires finding a relationship between C_A, C_{AI}, and C_{AV}. This can be accomplished by examining the chemical reactions that convert C_A into C_{AV} or C_{AI} and vice versa. During diffusion AV and AI defects will form and break up through reactions (to be considered in more detail below) of the type

$$AX \rightleftharpoons A + X . \tag{10.2}$$

Under quasiequilibrium conditions it is assumed that local equilibrium between A, X, and AX is attained at each point in the diffusing dopant profile, so that we may write

$$\frac{C_A C_X}{C_{AX}} = K(T) , \tag{10.3}$$

where K is a constant depending only on temperature [as in Eq. (8.1)]. But under equilibrium conditions it must be the case that $C_I = C_I^*$ and $C_V = C_V^*$. Therefore C_X is a constant in Eq. (10.3), since equilibrium concentrations of isolated native point defects do not depend on dopant concentration under intrinsic doping conditions. This means that $C_{AX}/C_A = (C_{AX}/C_A)^*$ must be a constant, depending only on temperature, for each type of defect. This condition is required in order to be consistent with experimental observations that, under intrinsic doping conditions, diffusivity does not depend on dopant concentration. If C_{AX}/C_A were not a constant for each type of defect, the fraction of dopant that was in a diffusing state, i.e., AI or AV, would depend on the local dopant concentration, and nonconstant diffusivities would result. From Eq. (10.3) we can then write

$$\frac{\partial C_{AX}}{\partial x} = \frac{C_{AX}}{C_A}\frac{\partial C_A}{\partial x} . \tag{10.4}$$

Equation (10.1) can then be simplified to

$$J_A = -\left[d_{AV} \left[\frac{C_{AV}}{C_A} \right]^* + d_{AI} \left[\frac{C_{AI}}{C_A} \right]^* \right] \frac{\partial C_A}{\partial x} \tag{10.5}$$

$$= -(D^*_{AV} + D^*_{AI}) \frac{\partial C_A}{\partial x} \ . \tag{10.6}$$

The diffusivity that is measured experimentally is of course

$$D^*_A \equiv D^*_{AV} + D^*_{AI} \tag{10.7}$$

and depends only on temperature.

Let us now examine the defect reactions in more detail. There are five species involved, A, AV, AI, or A_i, V, and I, and five reaction equations:

$$I + A \rightleftharpoons AI \quad \text{or} \quad I + A \rightleftharpoons A_i \ , \tag{10.8}$$

$$V + A \rightleftharpoons AV \ , \tag{10.9}$$

$$I + AV \rightleftharpoons A \ , \tag{10.10}$$

$$V + AI \rightleftharpoons A \quad \text{or} \quad V + A_i \rightleftharpoons A \ , \tag{10.11}$$

$$I + V \rightleftharpoons 0 \ . \tag{10.12}$$

The first reaction represents a substitutional-interstitial(cy) interchange process and is known as a "kick-out" reaction. The second reaction is the dopant-vacancy formation reaction. These reactions are examined in the context of dopant diffusion mechanisms in Sec. XI. The forward reactions of the remaining three processes represent recombination between I-type and V-type defects. These recombination reactions are discussed in Sec. XII. Spontaneous formation of AI or A_i from A by the reverse reaction of Eq. (10.11) is known as a *dissociative* reaction; it is a competing process to the kick-out reaction for forming dopant atoms in an interstitial-type state. The reverse reaction of Eq. (10.12) is the Frenkel pair generation process introduced in Sec. IV.A. There is clearly an analogy between the dissociative reaction and Frenkel pair generation. In the dissociative reaction a substitutional dopant atom becomes an interstitial-type defect leaving behind a vacancy, while in the Frenkel reaction a substitutional silicon atom becomes an interstitial-type defect leaving behind a vacancy.

When applied to impurity diffusion, the dissociative mechanism of diffusion is also called the Frank-Turnbull diffusion mechanism (Frank and Turnbull, 1956). The reverse reaction of Eq. (10.10) does not have a name attached to it. The spontaneous formation of AV can be viewed as happening when Frenkel pair generation occurs at a lattice site next to a substitutional dopant atom; the silicon interstitial so generated diffuses away, leaving behind a vacancy adjacent to the dopant atom. This is a competing process to the formation of AV defects by reaction (10.9).

The question naturally arises of what we can say about the relative importance of each of the above reactions and how they are related to the experimentally measured diffusivities. What we show here is that, for the quasiequilibrium conditions of the present discussion, it does not matter how AV and AI (or A_i) defects are formed. All that matters are the equilibrium concentrations $(C_{AI}/C_A)^*$ and $(C_{AV}/C_A)^*$. To see this, we can begin by writing down all the kinetic equations implied by the five reactions displayed above. For example, for A we can write

$$\begin{aligned} \frac{\partial C_A}{\partial t} = & (-k_{I,A} C_A C_I + k'_{I,A} C_{AI}) \\ & + (-k_{V,A} C_A C_V + k'_{V,A} C_{AV}) \\ & + (-k_{I,AV} C_{AV} C_I + k'_{I,AV} C_A) \\ & + (k_{V,AI} C_{AI} C_V - k'_{V,AI} C_A) \ . \end{aligned} \tag{10.13}$$

Each term in parentheses corresponds, respectively, to the first four reactions in Eqs. (10.8)–(10.11). We can write down similar equations for C_{AI}, C_{AV}, C_I, and C_V, including their diffusive fluxes. Adding the equations for C_A, C_{AI}, and C_{AV} gives

$$\frac{\partial C_A}{\partial t} + \frac{\partial C_{AI}}{\partial t} + \frac{\partial C_{AV}}{\partial t} = d_{AV} \frac{\partial^2 C_{AV}}{\partial x^2} + d_{AI} \frac{\partial^2 C_{AI}}{\partial x^2} \ . \tag{10.14}$$

Using Eq. (10.4), we can write

$$\left[\frac{C_A + C_{AI} + C_{AV}}{C_A} \right]^* \frac{\partial C_A}{\partial t} = \left[d_{AV} \left[\frac{C_{AV}}{C_A} \right]^* + d_{AI} \left[\frac{C_{AI}}{C_A} \right]^* \right] \frac{\partial^2 C_A}{\partial x^2} \ . \tag{10.15}$$

Since $(C_{AI}/C_A)^* + (C_{AV}/C_A)^* \ll 1$, this derivation shows that diffusion can be easily modeled by solving the Fick's law equation

$$\frac{\partial C_A}{\partial t} = D^*_A \frac{\partial^2 C_A}{\partial x^2} \ , \tag{10.16}$$

using the definition of D^*_A in Eqs. (10.5)–(10.7). For quasiequilibrium diffusion under intrinsic doping conditions, it is not necessary to know the details of the point-defect reactions involved in order to model diffusion adequately. This is also the case for quasiequilibrium diffusion under extrinsic conditions, as shown in the following section and in Appendix A. Conversely, diffusion studies performed under quasiequilibrium conditions do not provide information on the diffusion mechanisms involved. But there are many situations in which it is necessary to model dopant diffusion under nonequilibrium conditions, in which case point-defect reactions must be taken into account explicitly. This is discussed in Sec. XIII and Appendix D.

B. Diffusion under extrinsic conditions

Under extrinsic doping conditions, equilibrium carrier concentrations will change with ionized dopant concentrations, so that the Fermi level will vary over the spatial extent of the diffusing profile. This introduces two new factors, which are not present under intrinsic conditions: (i) As the Fermi level changes, the equilibrium ratio of $C_A C_X / C_{AX}$ will also change, leading to concentration-dependent diffusivities. (ii) Spatial variation of the Fermi level implies the existence of an electric field; this internal electric field is caused by the gradient of the ionized dopant concentration and results in an additional drift component of diffusion. These effects can be accounted for in a straightforward manner; however, the derivation is a bit lengthy and is deferred to Appendix A.

1. Experimental values of dopant diffusivities

A plot of dopant diffusivities, under intrinsic doping conditions, for the common dopants B, P, As, and Sb is shown in Fig. 10. The experimental accuracy of these measurements is discussed in Sec. XI; values for other dopants are also presented. A summary of expressions used to model dopant diffusion in the SUPREM III simulation program, under both intrinsic and extrinsic doping conditions, is included in Appendix B.

2. Single-species diffusion

If we consider a single species of dopant diffusing in the silicon, the results, taking all of the above factors into account, are surprisingly simple. Using a donor impurity as an example, Appendix A shows that the time evolution of the dopant profile is described by the solution of the equation

$$\frac{\partial C_A}{\partial t} = \frac{\partial}{\partial x}\left[D_A^* \frac{\partial C_A}{\partial x} \right], \tag{10.17}$$

where

$$D_A^* = h\left[D^i_{A^+X^0} + D^i_{A^+X^-}\left[\frac{n}{n_i}\right] + D^i_{A^+X^=}\left[\frac{n}{n_i}\right]^2 \right] \tag{10.18}$$

and h is defined to be

$$h \equiv 1 + \frac{C_{A^+}}{2n_i}\left[\left[\frac{C_{A^+}}{2n_i} \right]^2 + 1 \right]^{-1/2}. \tag{10.19}$$

The i superscripts denote intrinsic conditions and equilibrium is assumed. The h factor varies between 1 for $n \ll n_i$ and 2 for $n \gg n_i$. The h factor always acts to enhance the dopant diffusivity for either N- or P-type dopants. An example of simulations of a constant source diffusion of B is shown in Fig. 11, where the effects on profile shape and depth of penetration are very evident.

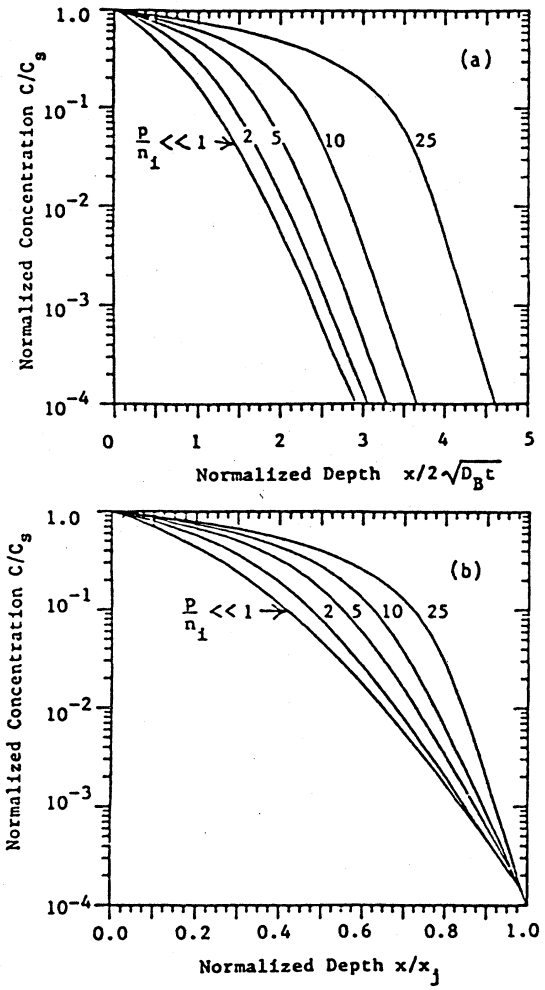

FIG. 11. Effect of concentration-dependent diffusion on B profile shapes, with p/n_i as a parameter. C_S is the surface concentration. (a) D_B is the value of B diffusivity when $p \ll n_i$. (b) The junction depth x_j is taken as $x_j = 10^{-4} C_S$.

The h term is usually referred to as the "electric field" factor, and the derivation in Appendix A shows that h represents the correction factor due to a spatially varying Fermi level. It clearly does not simply represent an additional drift component of diffusion on the AX defects because of the expected behaviors of positive, neutral, and negative AX species in the presence of an electric field. The h factor also includes the effect of the electric field on the X defects which interact with AX.

The key assumptions made in deriving the h factor are that the pairing reactions for the defects and dopants are in equilibrium and that the concentrations of the pairs and the defects are small, so that the Fermi level is determined only by the concentration of ionized dopants. Thus the charged pairs and isolated charged defects are affected by the E field but do not influence it themselves. Given these assumptions it is easy to see why the h factor acts on all AX species in the same manner. For ionized dopants A^+ in a concentration gradient, the mobile elec-

tron profile "spills over" and gives rise to an E field that enhances the migration of the positive pair A^+X^0. The charge-neutral pair A^+X^- experiences no electric field directly, but its distribution is determined by the distribution of the X^- defects. The X^- defects are determined by the Fermi level (the n/n_i term) *and* by the E field term and "pileup" where the E field is highest by just the factor h. This gives rise to the factor

$$h\left[\frac{n}{n_i}\right] \tag{10.20}$$

for the concentration-dependent distribution of the A^+X^- pair. The singly negative pair $A^+X^=$ experiences a retarding electric field, which varies with concentration as h, but the distribution of the pairs is determined by the $X^=$ defects. These have a $(n/n_i)^2$ term due to the Fermi level and are doubly affected by the E field because of their charge state. Thus the net effect on the $A^+X^=$ pair is a single h term.

The underlying origin of the h factor is discussed in Appendix A in terms of the change in electrochemical potential of A with C_A, which of course is the driving force for transport. The particular form of h in Eq. (10.19) results from the assumption of Boltzmann statistics; however, the limiting value of 2 also results if Fermi-Dirac statistics are used.

3. Multiple-species diffusion

If there is more than one dopant species present, so that changes in the Fermi level (and therefore electrochemical potential) result from diffusion of more than one chemical species, the situation is much more complex. These types of interactions are extremely important considerations in device fabrication, and yet even some very recent works have not applied the electric field factors correctly to realistic diffusion problems. A brief derivation of multiple-species diffusion is included in Appendix A along with a list of the assumptions that are made.

As an example of the complex behavior of multiple-species diffusion, Fig. 12 shows the resulting profile shapes for the donor As and acceptor Ga as a result of their interaction during diffusion (Mallam, Jones, and Willoughby, 1981). The dip in the Ga profile is very characteristic of this type of diffusion and not at all intuitively obvious. Its prediction was considered a solid verification of the initial treatment of the problem proposed by Hu and Schmidt (1968). More importantly, both the amount of Ga that lies outside the As-doped region and the depth of penetration are substantially reduced by the presence of the As. It is well known that these same effects occur for B in place of the Ga and are very important considerations for creating bipolar transistors with As-doped emitters and B-doped bases, since the width of the base and its total dose are critical parameters in the electrical performance of the device.

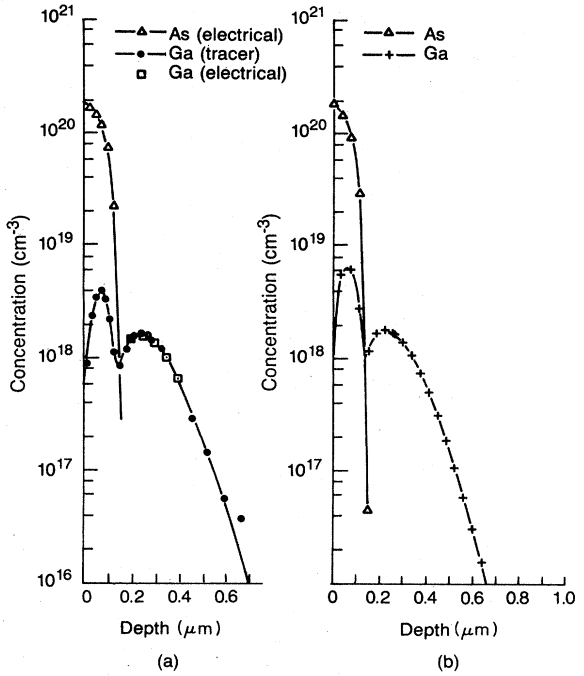

FIG. 12. Codiffusion of a donor dopant As and acceptor dopant Ga: (a) experimental result; (b) simulated result using the formalism of Appendix A. From Mallam, Jones, and Willoughby (1981).

In the As-doped region, Ga diffusion is inhibited because of three factors: (i) The change in Fermi level reduces the concentration of X^+ and X^{++} defects (Sec. VI) and thus reduces the $D_{Ga^-X^+}$ and $D_{Ga^-X^{++}}$ components of diffusion. (ii) The electric field factor is strong due to the sharp concentration gradient of the As and is in a direction opposing the diffusion of Ga out of the As-doped region. (iii) Pairing between As^+ and Ga^- immobilizes some of the Ga. The last of these processes, pairing, is often overlooked when interpreting experimental data, but its effect could be important, as discussed below.

4. Isoconcentration studies

One type of experiment for determining the charge-state components of D_A is to control the Fermi level independently of the dopant whose diffusivity is under study. This can be accomplished by diffusing a dopant species in a region that is homogeneously doped to a desired concentration with either another dopant or an isotope of the dopant under study. By performing a series of such experiments, one attempts to construct a plot of D_A as a function of n/n_i or p/n_i and thus deduce the relative values of the different charge-state components by assuming a D_A of the form of Eq. (10.19). For a substrate homogeneously doped with the dopant that controls the Fermi level, the h factor is equal to unity.

Some early attempts at isoconcentration studies were made by Masters and Fairfield (1969), who studied the diffusion of ^{76}As in heavily doped ^{75}As backgrounds; Makris and Masters (1971), who studied the diffusion of Ga in heavily doped B backgrounds; and Makris and Masters (1973), who studied the diffusion of ^{32}P in heavily doped ^{31}P backgrounds. These studies demonstrated that concentration-dependent diffusion could be modeled effectively by assuming D_A's of the form

$$D_{A^+}=D_{A^+X^0}+D^i_{A^+X^-}\frac{n}{n_i}$$

or (10.21)

$$D_{A^-}=D_{A^-X^0}+D^i_{A^-X^+}\frac{p}{n_i}\ .$$

It should be noted, however, that P diffusion cannot be modeled under nonisoconcentration diffusion conditions by such a formulation (Sec. III.C).

Recent examples of isoconcentration studies are the investigation of ^{10}B diffusing in a region heavily doped with ^{11}B (Miyake, 1985a; Willoughby *et al.*, 1986) and Sb diffusion in the presence of high concentrations of As (Fair, Manda, and Wortman, 1986) or P (Nishi, Sakamoto, and Ueda, 1986). As expected, with increasing ^{11}B doping above n_i (and thus increasing hole concentration) the diffusivity of ^{10}B increases and demonstrates what was already known from nonisoconcentration studies: that B diffusion can be modeled assuming D_B of the form given in Eq. (10.21). Enhancement of Sb diffusion occurs for both As and P doping. Fair, Manda, and Wortman (1986) concluded that the amount of Sb diffusion enhancement is the same for equal levels of As and P background doping by comparing their data for As background doping with data from P background doping from other investigations. This is expected for a model of equilibrium diffusion conditions, since the diffusivity would then be controlled only by the local Fermi level. However, there are neither enough data nor great enough experimental accuracy between these studies to support a definite conclusion. In addition, the more recent work of Nishi, Sakamoto, and Ueda (1986) shows higher enhancement of D_{Sb} in P-doped silicon than the enhancements in As-doped silicon seen by Fair, Manda, and Wortman for the same carrier concentration. Both works show that D_{Sb} is better modeled by an expression of the form

$$D_{Sb}=D_{Sb^+X^0}+D_{Sb^+X^=}\left[\frac{n}{n_i}\right]^2\ . \quad (10.22)$$

However, recent work by Nylandsted Larsen *et al.* (1988) and Andersen *et al.* (1988), using either As or P-doped samples, has shown that for even higher doping levels than those used by Nishi *et al.* and Fair *et al.*, Sb diffusivity increases dramatically. At the highest doping levels ($\sim 2-5\times 10^{20}$ cm^{-3}), an $(n/n_i)^4$ dependence on Sb diffusivity is observed. In addition, Nylandsted Larsen *et al.* observed from Mössbauer spectroscopy that, coincident with this fast diffusion, Sb atoms increasingly occupy *nonsubstitutional* sites (up to 80%). The nonsubstitutional state of the Sb is not the same as the Sb-vacancy complexes or precipitates found in previous studies using Mössbauer spectroscopy (Nylandsted Larsen *et al.*, 1986). These very recent results are quite surprising and show that isoconcentration experiments may not be as easy to interpret as previously thought.

It should be mentioned that self-diffusion has also been studied by the isoconcentration method, but the results are ambiguous [see the review article by Frank, Gösele, Mehrer, and Seeger (1984)].

5. Dopant pairing

All of the results for isoconcentration studies cited above involved donor dopants diffusing in regions heavily doped with donors, and acceptor dopants diffusing in regions heavily doped with acceptors. Studies have also been performed in which donors diffuse in acceptor-type backgrounds and acceptors in donor-type backgrounds. We have already noted above that Ga and B diffusion are retarded by high concentrations of As. Miyake (1985b) has observed reduced diffusion of B in heavily doped P layers. The purpose of some of these studies has been to determine the component of D_A that proceeds by interactions with neutral-charge defects. For example, when Sb is diffused in a heavily doped B background, the Sb^+X^- and $Sb^+X^=$ components should be effectively suppressed, leaving only the Sb^+X^0 component to contribute to diffusion. The results of such a study (Fahey, 1983; Fair, Manda, and Wortman, 1986) show, however, that Sb diffusivity is reduced below that which would be expected based on Fermi-level effects alone. P diffusion in a heavily doped background of B also exhibits this same behavior (Fahey, 1983). This most likely indicates that *dopant pairing* reactions

$$Sb^+ + B^- \rightleftharpoons (Sb^+B^-),\quad P^+ + B^- \rightleftharpoons (P^+B^-)\ , \quad (10.23)$$

are effectively reducing diffusion below that predicted by Fermi-level effects alone. This pairing reaction is the most reasonable explanation for the same large retarded diffusion observed for B diffusing in layers heavily doped with As (Willoughby *et al.*, 1986). Donor-acceptor pairing has been observed on an atomic scale using In with the perturbed angular correlation (PAC) technique. Pairing between In and P, As, and Sb have been reported (Forkel *et al.*, 1986; Wichert, Swanson, and Quenneville, 1986).

XI. MECHANISMS OF DOPANT DIFFUSION

In Sec. X.A it was shown that for a dopant A, whose diffusion is mediated by a point defect of type X (where X is either a vacancy-type or interstitial-type point defect), the diffusivity under intrinsic doping conditions is given by

$$D_{AX}^{i}=d_{AX}\left[\frac{C_{AX}}{C_A}\right]^{i}. \tag{11.1}$$

Under equilibrium conditions D_{AX}^{i} depends only on temperature. Experimentally, one always measures $D_A=D_{AV}+D_{AI}$, which also depends only on temperature (the superscript denoting intrinsic conditions is dropped for the rest of this section). This temperature dependence has been measured for dopants in many different studies. A compilation of these data can be found in the book edited by Wöhlbier (1986). By plotting D_A vs $1/T$, a fit to the Arrhenius expression

$$D_A=D_{A0}\exp\left[-\frac{Q_A}{kT}\right] \tag{11.2}$$

is usually made. The quantity Q_A is referred to as the *apparent activation energy* of diffusion.

We are most interested in examining the theoretical interpretation of Q_A for different diffusion mechanisms and relating it to measured values. Meaningful discussion of D_{A0} is much more difficult, since its atomistic interpretation is not as straightforward as that for Q_A. In addition, the experimental error in determining D_{A0} empirically is inherently much greater than the error in determining Q_A.

The remainder of this section examines the energetics of diffusion by the vacancy, interstitialcy, and interstitial mechanisms. It should be noted that the interstitialcy and interstitial mechanisms presented below should technically be called substitutional/interstitial(cy) interchange mechanisms. However, since for silicon it is always understood that the dopants of most interest (B, Ga, In, P, As, and Sb) are dissolved almost entirely on substitutional sites and the relevant discussion concerns contrasting diffusion by vacancy-type or interstitial-type mechanisms, we follow the accepted convention for Si diffusion studies of referring to the substitutional/interstitial(cy) interchange mechanism as the interstitial(cy) mechanism.

We do not analyze the dissociative (Frank-Turnbull) mechanism (briefly introduced in Sec. X.A) because it is not believed that dopants diffuse by this mechanism. This means that the kick-out mechanism is assumed to be dominant over the dissociative mechanism for the interstitial(cy) component of dopant diffusion. The validity of this assumption is supported by the analysis for nonequilibrium diffusion in Sec. XIII.A and Appendix D. Qualitatively, for the dissociative mechanism to be dominant over the kick-out mechanism generally requires that some measurable fraction of an impurity dissolve in the interstitial state and that either the vacancy component of self-diffusion D_{SiV} be orders of magnitude larger than the interstitial(cy) component D_{SiI}, or the diffusivity of A in an interstitial-type state, d_{AI}, be much larger than the diffusivity of I and V. But we know that dopants are highly substitutional in nature. We also know from present experimental evidence that although one component of self-diffusion may be dominant over the other at a given temperature, it is unlikely that D_{SiV}/D_{SiI} could be large enough to offset the small values of $(C_{AI}/C_A)^*$. There is also only experimental evidence contrary to the idea that the diffusivity of AI or A_i is much greater than I or V. All of these are in the direction opposing the dissociative mechanism.

A. Fundamental question of self-diffusion and dopant diffusion

The experimental determination of Q_A for the various dopants clearly indicates that the apparent activation energy of self-diffusion can be 1 eV or more greater than the apparent activation energy of dopant diffusion. For a given dopant, there are two basic explanations for this phenomenon: (i) dopant diffusion proceeds by the same mechanism as self-diffusion but with a lower activation energy, or (ii) dopant diffusion proceeds by a different mechanism than self-diffusion.

If (i) is true, then a model should be proposed that explains this lowering of activation energy. Explanation (ii) is actually an unsatisfactory interpretation of diffusion experiments, since there is solid evidence (Sec. XVII.A) that for $T>1050\,^{\circ}\mathrm{C}$ some dopants diffuse primarily through an I-type mechanism (P and B), while other dopants diffuse primarily through a V-type mechanism (Sb and As), and yet the activation energies of dopants that diffuse by either type of mechanism are lower than the activation energy of self-diffusion over the same temperature range. Thus a key question is what differences in activation energy between self-diffusion and dopant diffusion are predicted by I-type and V-type diffusion mechanisms.

To analyze the energetics of diffusion, the following quantities are defined:

$$Q_{\mathrm{self}}-Q_A\equiv\Delta Q_A,\quad Q_{SiX}-Q_{AX}\equiv\Delta Q_{AX}. \tag{11.3}$$

ΔQ_A is the difference in activation energy of diffusion between self-diffusion and dopant diffusion that will be measured experimentally for dopant A. The quantity ΔQ_{AX} is the difference in activation energy between the components of self- and dopant diffusion that proceed by a mechanism involving X-type point defects.

In Sec. IX it was shown that

$$Q_{SiX}=H_X^m+H_X^f. \tag{11.4}$$

Similarly, we state that

$$Q_{AX}=H_{AX}^m+H_{AX}^f, \tag{11.5}$$

where H_{AX}^m is the enthalpy of migration of the AX defect and H_{AX}^f is the enthalpy of formation of the AX defect. H_{AX}^m is associated with the d_{AX} term in Eq. (11.1), while H_{AX}^f is associated with the C_{AX}/C_A term.

Before presenting an energetics analysis of dopant diffusion, we first summarize the experimental values for Q_A.

B. Experimental determination of activation energies: Accuracy of measurements

A selective listing of Q_A values is given in Table I. A more complete list of data, including preexponential factors, can be found in the compilation of Wöhlbier (1986). In attempting to choose dependable data from the many studies reported in the literature, preference has been given to techniques that utilized some type of profile analysis over studies that monitored only penetration depths. These data are supposed to be valid for diffusions performed under intrinsic doping conditions; however, in some early studies the importance of differentiating between intrinsic and extrinsic conditions was not fully appreciated. It is relatively easy to show that under extrinsic conditions Q_A will be lowered compared to its value under intrinsic conditions if charged point defects are involved, since their enthalpies of formation are lowered under extrinsic conditions (Sec. VI.A).

A brief comment on the accuracies of the Q_A's in Table I is appropriate. Determinations of Q_P by different studies using a variety of techniques agree to within 0.2 eV and are considered accurate. A greater spread in values is found for Q_{As} than for Q_P although all studies seem to be performed with reasonable care and techniques that are reliable in principle. The Q_A values for Sb and In determined by Fuller and Ditzenberger (1956) were obtained by measuring penetration depths as revealed by staining of *pn* junctions; the accuracy of this technique is always questionable. Still, theirs are the only data for In diffusion, and their value of Q_{Sb} is in good agreement with the other works. Unfortunately the value of Q_{Sb} from the data of Adda is from unpublished work and thus not available for critiquing. The Q_{Sb} data from reference (i) of Table I are from SIMS measurements of buried marker layers and in principle are very accurate, although only temperatures of 1100 °C and 1000 °C were used in this experiment. Few investigations have examined the diffusion of Bi. The technique of Pommerrenig (1965) is reliable in principle. The three middle values for Q_B in Table I agree very well, but the lower (3.5 eV) and upper (3.87 eV) values are in very poor agreement, even though, as in the case of As, all the studies seem to have been performed with care, utilizing techniques that should give reasonable results. The determination of Q_{Al} was made some time ago by Miller and Savage (1956), which makes it difficult to assess the reliability of the measurements, although the technique of capacitance-voltage (CV) analysis is reliable in principle. It should be noted, however, that measurements of Al diffusion utilizing the CV profiling technique [e.g., the recent results of Wilson (1987)] show that Al diffusion can exhibit some anomalous effects. The only accurate study of Ga diffusion from which Q_{Ga} can be derived is that of Makris and Masters (1971).

It should also be noted that the work of Ghoshtagore (1971) was not included in the table even though diffusion measurements were performed for all of the dopants listed. This is because the data of Ghoshtagore invariably show diffusivity values that are significantly lower than any other study.

The general agreement between different measurements of Q_A is not always good. This may be due to a lack of control of oxygen in the silicon (as discussed previously for self-diffusion) or the fact that different temperature ranges are used in the various experiments. But poor agreement in Table I between Q_A values for the same dopant only serves to emphasize that the activation energy of self-diffusion is typically greater than or approximately equal to 1 eV more than the activation energies of dopant diffusion.

TABLE I. Apparent activation energies of diffusion.

Donor A	Q_A (eV)	
P	3.51, 3.61, 3.61[a], 3.66[b], 3.67	c,d,e,f,g
As	4.05, 4.08, 4.11, 4.23, 4.34[a]	c,h,d,i,j
Sb	3.89[a], 3.98, 4.05	k,l,m
Bi	4.12	n
Acceptor A	**Q_A (eV)**	
B	3.25, 3.46[b], 3.50, 3.51, 3.87	o,f,g,p,c
Al	3.36	q
Ga	3.75[a]	r
In	3.60	l

[a]Values recalculated from original data.
[b]Calculated by Fair (1981a) using data from various sources.
[c]Hill (1981).
[d]Ishikawa, Sakina, Tanaka, Matsumoto, and Niimi (1982).
[e]Makris and Masters (1973).
[f]Fair (1981a).
[g]Lin, Antoniadis, and Dutton (1981).
[h]Chiu and Ghosh (1971).
[i]Armstrong (1962).
[j]Masters and Fairfield (1969).
[k]Guerrero, Jüngling, Pötzl, Gösele, Mader, Grasserbauer, and Stingeder (1986).
[l]Fuller and Ditzenberger (1956).
[m]Adda, unpublished work; data presented by Fair, Manda, and Wortman (1986).
[n]Pommerrenig (1965).
[o]Willoughby, Evans, Champ, Yallup, Godfrey, and Dowsett (1986).
[p]Kurtz and Yee (1960).
[q]Miller and Savage (1956).
[r]Makris and Masters (1971).

C. Vacancy mechanism

The most obvious way for a substitutional dopant atom to migrate through the host lattice is by moving onto an adjacent vacant lattice site. A schematic of this process is shown in Fig. 13. Because of the apparent plausibility of this process, and the experimental observation almost 30 years ago that vacancies were the dominant point defect in metals (Simmons and Baluffi, 1960), the mechanics of the vacancy mechanism in various crystal structures has been worked out in some detail. For silicon (i.e., dia-

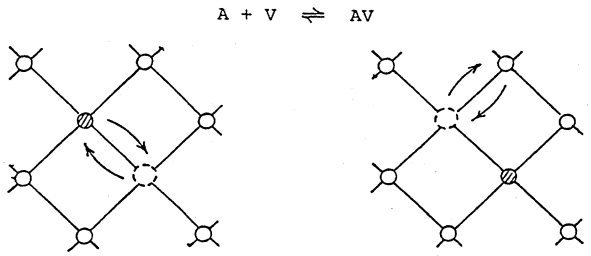

FIG. 13. Dopant diffusion by the vacancy mechanism. In the diamond lattice a vacancy must move at least a third-neighbor distance away from the dopant to return along a different path.

mond structure), the energetics of dopant diffusion by the vacancy mechanism were developed by Hu (1973b).

Perhaps the most important aspect of this mechanism is that diffusion does *not* occur by a simple exchange process, for then the dopant atom and vacancy will continue to occupy alternate lattice sites and no long-range migration of A will take place. After exchange, in order for one diffusion step to be completed, the vacancy must diffuse some distance away from the dopant atom so that it may return to a different lattice site adjacent to A along a different path. In bcc and fcc crystals the vacancy must diffuse only to a second-nearest-neighbor site in order to return along a different path. The two-dimensional lattice shown in Fig. 13 demonstrates, conceptually, this property of three-dimensional fcc and bcc lattices. In the diamond structure of the silicon lattice, the vacancy must diffuse to at least a third-nearest neighbor to complete one diffusion step. The two-dimensional hexagonal lattice of Fig. 7 schematically illustrates this property.

A potential-energy diagram of the vacancy diffusion process is indicated in Fig. 14. Here, a generalized vacancy potential as a function of distance (expressed in coordination site number) is plotted. In this figure, H_V^m is the migration enthalpy of an unperturbed vacancy diffusing through the lattice, E_{AV}^b is the binding energy of the AV pair, and ΔE_{AV}^3 is defined as the difference in potential energy between a vacancy very far away from the dopant atom and one at a third-nearest-neighbor site

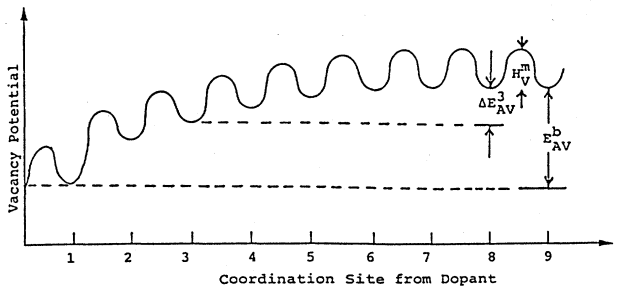

FIG. 14. Vacancy potential as a function of coordination site away from a substitutional dopant. From Hu (1973b).

from the dopant. Hu's analysis (1973b) shows that Q_{AV} can be viewed as effectively consisting of

$$Q_{AV}=\underbrace{H_V^f-E_{AV}^b}_{H_{AV}^f}+\underbrace{H_V^m+E_{AV}^b-\Delta E_{AV}^3}_{H_{AV}^m}. \tag{11.6}$$

Since the activation energy of self-diffusion is simply given by

$$Q_{SiV}=H_V^f+H_V^m, \tag{11.7}$$

it follows that

$$Q_{SiV}-Q_{AV}\equiv\Delta Q_{AV} \tag{11.8}$$

$$=\Delta E_{AV}^3. \tag{11.9}$$

That is, the activation energy of dopant diffusion by a vacancy mechanism may be an amount $\Delta Q_{AV}=\Delta E_{AV}^3$ lower than the corresponding activation energy of self-diffusion by the same mechanism.

Physically, Hu's analysis shows that although the energy of formation of AV is decreased with respect to V because of the attractive potential between A and V, this same attraction increases the migration energy of AV compared to V, since the vacancy must surmount at least some fraction of this potential barrier in order to effect diffusion of A. If the A-V interaction potential extends further than a third-nearest-neighbor site, the vacancy does not need to surmount the entire barrier (i.e., completely dissociate), and A may diffuse with an activation energy lower than that of self-diffusion. For an increasingly strong interaction potential beyond a third-nearest-neighbor distance, complete dissociation of the AV complex becomes less probable, and A and V diffuse as a pair.

The results of Hu's analysis are very important. They show that a small value of ΔQ_{AV} can result even for a very large binding energy if the interaction potential does not extend beyond a third-nearest-neighbor site. Equation (11.6) also shows that, of necessity, $E_{AV}^b>\Delta Q_{AV}$. Thus not only the strength of the interaction potential, but also its nature, i.e., short range or long range, will determine the effectiveness of a vacancy mechanism for diffusion of A.

If self-diffusion occurs through a vacancy mechanism, then it follows that $E_{AV}^b>1$ eV. This proposition is consistent with the experimental values quoted in Sec. VIII.B, but also has the consequence that the concentration of AV pairs will exceed the concentration of isolated V at concentrations below 10^{19} cm^{-3} (Sec. VIII.A.2 and Fig. 5). Under this condition, the effects of trapping and dopant-assisted recombination may become important for nonequilibrium conditions (Secs. VIII.A.2 and XII.E).

1. Charge-state dependence of activation energy

For a simple Coulombic potential, the strongest attraction would be between either A^- and V^{++} or A^+ and $V^=$. According to Eq. (8.6), a doubly charged defect located a third-nearest-neighbor distance from a dopant

atom ($r=4.5$ Å) will give ΔQ_{AV} equal to 0.5 eV. This is smaller than the 1-eV difference between dopant diffusion and self-diffusion. Thus a rather strong, non-Coulombic interaction potential extending beyond the third-nearest-neighbor distance is necessary to explain the difference between self-diffusion and dopant diffusion for $T > 1050$ °C if both are proposed to take place via a vacancy mechanism.

Before accepting this, however, we may ask whether postulating that self-diffusion occurs primarily through vacancies with a different charge state from those that assist dopant diffusion can possibly lead to an increase in ΔQ_{AV} over that found by assuming both diffuse with vacancies in the same charge state. It is a relatively easy matter to show (Fahey, 1985) that an increase of ΔQ_A arising from such possibilities can occur only if the vacancy charge state that dominates self-diffusion is *not* the charge state that leads to the lowest activation energy of self-diffusion.

A related question is whether one can deduce which vacancy charge state leads to the lowest possible value of Q_{AV}. One might intuitively suspect that the lowest energy occurs for the vacancy charge state that is most strongly attracted to the dopant, but this is not necessarily so. Using Eq. (11.6) and the example of a donor dopant, we find that[5]

$$\begin{aligned} Q_{A^+V^0} &\equiv H^m_{V^0} + H^f_{V^0} - \Delta E^3_{A^+V^0} , \\ Q_{A^+V^-} &\equiv H^m_{V^-} + H^f_{V^-} - \Delta E^3_{A^+V^-} , \\ Q_{A^+V^=} &\equiv H^m_{V^=} + H^f_{V^=} - \Delta E^3_{A^+V^=} . \end{aligned} \tag{11.10}$$

As an example, consider the difference in Q_{A^+V} for a donor dopant interacting with V^0 and V^-. The above relations yield

$$(Q_{A^+V^-} - Q_{A^+V^0}) = (H^f_{V^-} - H^f_{V^0}) + (H^m_{V^-} - H^m_{V^0}) + (\Delta E_{A^+V^0} - \Delta E_{A^+V^-}) . \tag{11.11}$$

Although this type of analysis makes clear what atomistic parameters contribute to activation energies, none of these parameters is known with enough accuracy to actually make a prediction of $(Q_{A^+V^-} - Q_{A^+V^0})$ at diffusion temperatures. In addition, dynamic charging effects, as discussed in Sec. VII, may also be an important factor, significantly complicating the analysis.

2. Dopant dependence of activation energy

It is an experimental certainty (Table I) that Q_A is not the same for all dopants. Since most investigators in the area of point defects and diffusion have favored, until recently, the vacancy mechanism of diffusion for all the dopants, one would expect that some attempt has been made to explain what set of circumstances might lead to a difference in Q_A for two different dopants both diffusing through a vacancy mechanism. Surprisingly, this question has never been discussed in the literature.

Suppose for the moment that both Sb and P diffuse by a vacancy mechanism. Then we would have that

$$Q_{SbV} - Q_{PV} = \Delta E^3_{PV} - \Delta E^3_{SbV} . \tag{11.12}$$

But since $Q_{SbV} > Q_{PV}$, one would have to conclude that $\Delta E^3_{PV} > \Delta E^3_{SbV}$: i.e., a vacancy at a third-nearest-neighbor site from a P atom is attracted more strongly to the P atom than it would be to an Sb atom. As discussed in Sec. VIII.B.3, this is contrary to the usual qualitative arguments that an oversized Sb atom has a stronger attractive potential for a vacancy than an undersized P atom. Although we believe that the difference between Q_{Sb} and Q_P is most likely a result of Sb's diffusing primarily by a vacancy mechanism and P's diffusing primarily by an interstitial-type mechanism, the above analysis is a specific example of the following general conclusion.

The activation energy for dopant diffusion by a vacancy mechanism, Q_{AV}, will be the same for all n-type dopants and the same for all p-type dopants, unless the interaction potential between A and V extends beyond a third-nearest-neighbor distance and has an appreciable non-Coulombic component.

D. Interstitialcy mechanism

Much less formal analysis has been directed at the atomistic picture of dopant diffusion by an interstitialcy mechanism than has been devoted to the vacancy mechanism, and less effort has been expended in relating it to the experimentally determined values of Q_A. In this section a rudimentary analysis of the interstitialcy mechanism is presented in much the same way as the vacancy mechanism was treated above.

Dopant diffusion by the substitutional/interstitialcy interchange mechanism takes place by the processes illustrated schematically in Fig. 15. An isolated dopant atom in the substitutional state is approached by a silicon interstitialcy. The migration of the silicon interstitialcy

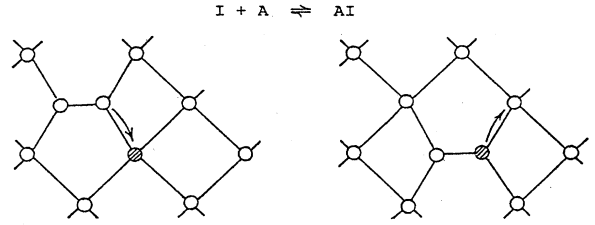

FIG. 15. Dopant diffusion by a substitutional/interstitialcy interchange mechanism via "kick-out" reactions.

[5]These equations are valid under both intrinsic and extrinsic doping conditions. Under extrinsic doping conditions the vacancy formation enthalpies are a function of doping as discussed in Sec. VI.

takes place when one of the atoms in the interstitialcy moves toward an adjacent lattice site (already occupied by a silicon atom), where it re-forms the interstitialcy defect with the new silicon atom. The silicon interstitialcy may also form a dopant interstitialcy by the same process, which then migrates in a fashion similar to the silicon interstitialcy.

An important difference exists between this mechanism and the vacancy diffusion mechanism. Whereas migration of a dopant atom by a vacancy mechanism requires that the diffusing defect (AV pair) must at least partially dissociate, the interstitialcy mechanism will operate only if the diffusing defect (AI) does *not* dissociate.

A potential-energy diagram of the migration and formation/dissociation processes is shown in Fig. 16. The binding energy in this diagram is defined by the equation[6]

$$\frac{C_{AI}}{C_A}=\theta_{AI}\frac{C_I}{C_S}\exp\left[\frac{E^b_{AI}}{kT}\right] . \tag{11.13}$$

The diagram basically shows that the attractive potential between I and A determines the probability of survival of the AI defect between its diffusion jumps. Using this diagram it is readily apparent that

$$Q_{AI}=H^m_{AI}+H^f_{AI}=H^m_{AI}+H^f_I-E^b_{AI} . \tag{11.14}$$

Since the Q for self-diffusion is given by

$$Q_{SiI}=H^m_I+H^f_I , \tag{11.15}$$

we have that

$$\Delta Q_{AI}\equiv(Q_{SiI}-Q_{AI})=(H^m_I-H^m_{AI})+E^b_{AI} , \tag{11.16}$$

so that the activation energy of dopant diffusion by an interstitialcy mechanism may be an amount of ΔQ_{AI} lower than the corresponding energy for self-diffusion by the same mechanism. One may also note that to explain a lower activation energy for dopant diffusion than for self-diffusion completely in terms of an interstitialcy mechanism, a long-range interaction potential is not necessary as it is in the case of the vacancy mechanism.

In this treatment, each successful jump of AI is independent of the previous ones. As a consequence, the dopant interstitialcy is viewed as moving as an uncorrelated particle and there is, therefore, no fundamental relationship between H^m_{AI} and H^m_I. In contrast, for the vacancy mechanism, H^m_{AV} is related to H^m_V through Eq. (11.6). It can also be seen from the potential diagram that one would expect H^m_{AI} to be less than E^b_{AI} for an interstitialcy mechanism to be effective. On the other hand, from Eq. (11.6), it is not necessarily true that $H^m_{AV}<E^b_{AV}$.

[6]This is actually a bit of an artifice, since the silicon interstitial disappears once it has formed the AI defect, but the concept of what is going on in Fig. 16 is clear, and the use of Eq. (11.13) is self-consistent in the following discussion.

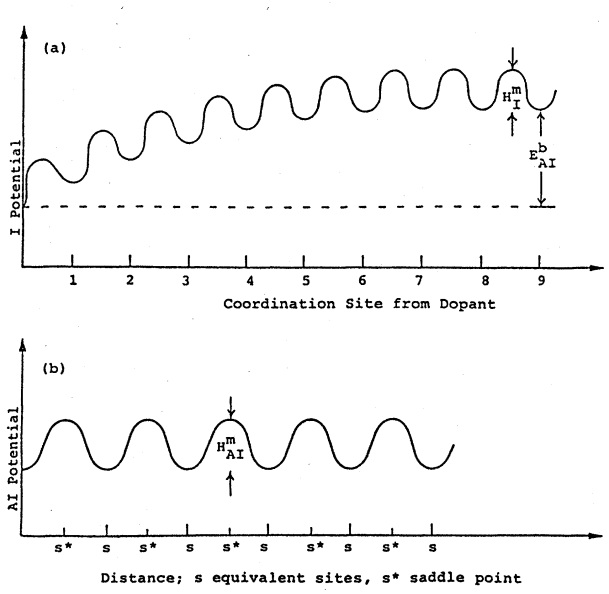

FIG. 16. Potential diagram for the substitutional/interstitialcy interchange mechanism: (a) formation and dissociation of AI; (b) migration of AI.

To explain a value of ΔQ_A approximately equal to 1 eV purely in terms of an interstitialcy mechanism requires that $\Delta Q_{AI}\approx 1$ eV. Equation (11.16) shows that E^b_{AI} may actually be lower than ΔQ_A if $H^m_{AI}<H^m_I$. But since we know that I-type defects migrate over much greater distances than is possible for AI defects, we suspect that this is not true and $E^b_{AI}>\Delta Q_A$ (but in the absence of experimental data, this remains an "educated guess"). As an example, we believe that the experimental evidence suggests that all p-type dopants and P diffuse by an I-type mechanism at temperatures ≥ 1050 °C. A brief look at Table I shows, then, that we need $E^b_{AI}>\Delta Q_{AI}\approx 1.4$ eV. While there is no *a priori* reason to believe that such a large binding energy is not possible, the discussion in Sec. VIII.A.2, and Fig. 5, indicate that, as a consequence of this strong binding, the concentration of AI defects will exceed the number of isolated I defects at dopant concentrations of about 10^{17} cm^{-3} and above. (See Secs. VIII.A.2 and XII.E for a discussion of nonequilibrium effects, which may occur when AX defects exceed isolated I defects in concentration.)

1. Charge-state dependence of activation energy

One may wonder whether ΔQ_{AI} is greatest if dopant atoms diffuse by interacting primarily with interstitialcies in a different charge state than that responsible for self-diffusion by an interstitialcy mechanism. By analogy to the case for vacancies, it is a simple matter to show (Fahey, 1985) that a larger value of ΔQ_{AI} will result by assuming self-diffusion and dopant diffusion to occur primarily with interstitialcies in different charge states rather than the same charge state, only if the charge state of I

that is responsible for self-diffusion is not the charge state that results in the lowest activation energy of self-diffusion.

For dopants interacting with silicon interstitialcies in different charge states, the charge-state dependence of Q_{AI} can be examined using Eq. (11.4). As for the case of the vacancy mechanism, a brief consideration of the charge-state dependence of Q_{AI} shows that, although this charge-state dependence could be significant, not enough quantitative information on the terms contributing to Q_{AI} is available to predict which charge state of AI leads to the lowest value of Q_{AI}.

2. Dopant dependence of activation energy

The dopant dependence of Q_{AI} follows directly from Eq. (11.14). As an example, the difference in Q_{AI} between B and In is seen to be

$$Q_{\mathrm{B}I}-Q_{\mathrm{In}I}=(H^m_{\mathrm{B}I}-E^b_{\mathrm{B}I})-(H^m_{\mathrm{In}I}-E^b_{\mathrm{In}I})\ . \quad (11.17)$$

In general, one would not expect that $Q_{\mathrm{B}I}=Q_{\mathrm{In}I}$. However, since both H^m_{AI} and E^b_{AI} are probably determined largely by the strength of the bonds holding the AI defect together, it is possible that an increase in E^b_{AI} may signify a corresponding increase in H^m_{AI}, although there is no fundamental relationship between E^b_{AI} and H^m_{AI}. We make the following conclusion.

The activation energy for dopant diffusion by an interstitialcy mechanism, Q_{AI}, will in general be different for all the dopant atoms unless there is a fixed relationship between the difference in binding energy and migration enthalpy of the AI defect.

This last conclusion may be compared with the analogous statement made at the end of Sec. XI.C.

E. Interstitial mechanism

Dopant diffusion by the substitutional/interstitial interchange mechanism is shown in Fig. 17. A silicon interstitial like those pictured in Fig. 1(b), migrating through the interstices of the lattice, approaches a substitutional dopant atom. If the interaction causes the dopant atom to be displaced into an interstitial site, the dopant atom will now migrate through the interstices as a dopant interstitial until it takes up a substitutional site again by dislodging a silicon atom from a substitutional site. These processes are described by the reaction

$$A_s+I\rightleftharpoons A_i\ , \quad (11.18)$$

where the subscript on A_s is included to make clear that it is a substitutional dopant atom that is displaced into an interstitial position.

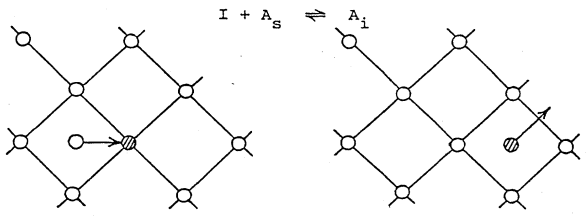

FIG. 17. Dopant diffusion by a substitutional/interstitial interchange mechanism via "kick-out" reactions.

It is readily apparent that the interstitialcy and interstitial mechanisms are similar in nature. One may attempt to draw for the interstitial mechanism analogous potential diagrams to those shown for the interstitialcy mechanism in Fig. 16, in order to analyze Q_{A_i}, which by definition is given by

$$Q_{A_i}=H^f_{A_i}+H^m_{A_i}\ . \quad (11.19)$$

As in the case of the interstitialcy mechanism, there is no fundamental relationship between $H^m_{A_i}$ and H^m_I. Further, $H^f_{A_i}$ is not related to H^f_I through a binding-energy term, although it is easy to see that the effect of a strong attraction between substitutional A and a silicon interstitial I is to enhance the probability of formation of the mobile defect A_i. This is essentially the same situation as for the interstitialcy mechanism.

XII. NONEQUILIBRIUM FORMULATION FOR POINT DEFECTS

The most important sources and sinks of point defects that are responsible for nonequilibrium conditions were listed in Sec. IV. We are primarily interested in describing point-defect behavior in the bulk that results from chemical reactions occurring at the silicon surface, and in particular the case of thermal oxidation, which injects silicon interstitials. Figure 18 shows two cases that are of practical interest to analyze. Figure 18(a) represents the case of oxidation of silicon in a limited surface region. It is of great interest to be able to predict the vertical and lateral extent of the I excess, characterized by the respective diffusion lengths before recombination L^b_I and L^s_I, since this is by far the most common situation to arise during process simulation for which solutions to point-defect equations are needed. Figure 18(b) shows a situation corresponding to a silicon-on-insulator (SOI) structure. There has been a renewed interest in fabricating electron devices in SOI structures because of the finding that good quality bipolar (Greeneich and Reuss, 1984; Münzel, Albert, and Strack, 1984) and metal oxide semiconductor (MOS) devices [see the review article by Hemment (1985)] can be made by utilizing high-dose oxygen implantation to create the buried insulator. This technique may prove economical in small-scale manufacturing. In addition, as shown in detail in Sec. XVI, thin silicon layers like those shown in Fig. 18(b) have an important use as test structures to determine model parameters such as point-defect diffusivities, surface recombination velocities, etc.

Experimental evidence (Sec. XIV.C) indicates that, accompanying the supersaturation of I during thermal oxidation, there is an undersaturation of V. The commonly proposed explanation is that this undersaturation is due to recombination of the excess I with V. There is ample evidence that this same phenomenon occurs for I injection from other reactions (Secs. XIV.F and XIII.C). The equivalent case of I undersaturation resulting from vacancy injection has also recently been shown to occur (Sec. XIV.F). Therefore, in the following section we first examine the differential equations of I and V flow, which are coupled by I-V recombination. We then discuss analytic solutions for I injection, which are decoupled from the V diffusion equations and take into account bulk recombination by assuming a time-invariant concentration of recombination centers. These solutions provide important insight into the physical significance of point-defect model parameters and also demonstrate the maximum effect that bulk recombination can have in a specific experimental situation. We then examine under what conditions I and V equations can be effectively decoupled. Finally we consider the possibility that bulk recombination of I occurs with defects other than, or in addition to, isolated vacancies.

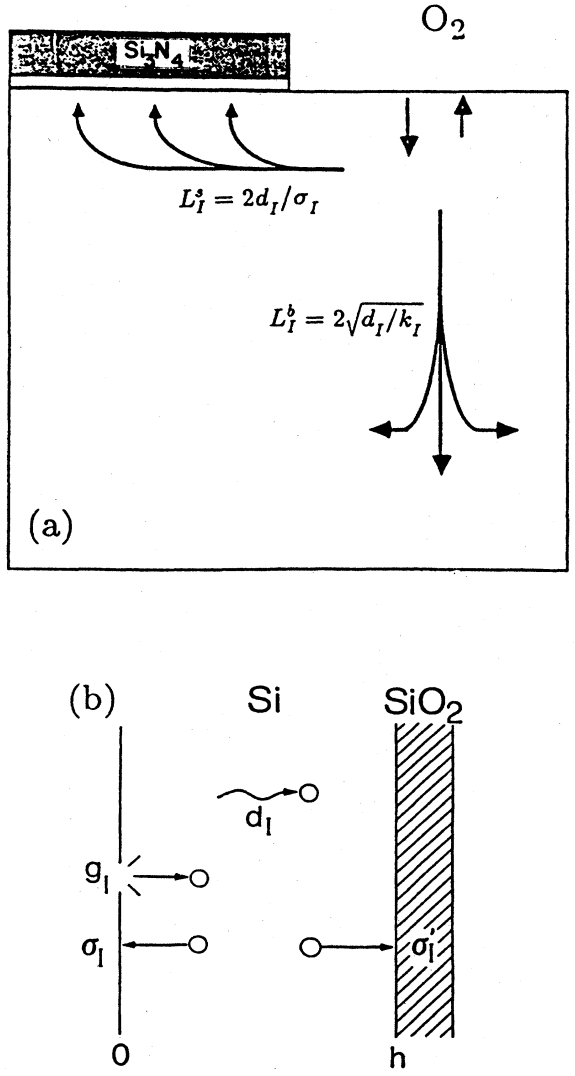

FIG. 18. I injection into the silicon substrate. (a) A masked region of the silicon is oxidized. The two-dimensional distribution of excess I can be characterized in terms of the surface and bulk decay lengths L_I^s and L_I^b, respectively. (b) Injection into a silicon-on-insulator (SOI) structure. This is a practical application of information obtained from test structures of thin silicon wafers.

A. Coupled *I* and *V* equations

If we focus only on the behavior of the native point defects and ignore for the moment any internal sources and sinks for point defects, the continuity equations require that

$$\frac{\partial C_I}{\partial t} = d_I \frac{\partial^2 C_I}{\partial x^2} - k_{I,V}(C_I C_V - C_I^* C_V^*) ,$$
$$\frac{\partial C_V}{\partial t} = d_V \frac{\partial^2 C_V}{\partial x^2} - k_{I,V}(C_I C_V - C_I^* C_V^*) , \tag{12.1}$$

where $k_{I,V}$ is the reaction constant describing the recombination process between I and V:

$$I + V \rightarrow 0 . \tag{12.2}$$

For a chemical reaction at the surface that injects a net flux g_I or g_V of either I or V, respectively, these coupled equations are to be solved subject to the boundary conditions at $x = 0$,

$$g_I + d_I \frac{\partial C_I}{\partial x} - \sigma_I (C_I - C_I^*) = 0 ,$$
$$g_V + d_V \frac{\partial C_V}{\partial x} - \sigma_V (C_V - C_V^*) = 0 , \tag{12.3}$$

where surface loss terms are modeled with surface recombination velocities σ_I and σ_V.

These equations have been analyzed in some detail by Hu (1985b, 1985c). One of the most important results of this analysis concerns the approach to steady state. It had long been assumed that during I injection, steady state would be reached when (Sirtl, 1977)

$$C_I C_V = C_I^* C_V^* . \tag{12.4}$$

Thus during I injection the vacancy undersaturation will be the inverse of the I supersaturation in steady state. But this conclusion ignores the influence of the surface on point-defect behavior. To understand this, we focus on the behavior of point defects at the Si/SiO_2 surface during oxidation.

Under nonoxidizing conditions, the Si surface generates and absorbs both I and V through normal surface generation and loss mechanisms. At equilibrium, the flux of each species into the silicon equals the flux out. When oxidation is begun, an excess of I are generated at the

surface and flow into the bulk. If we ignore any bulk recombination mechanisms, the I excess $\Delta C_I = C_I - C_I^*$ that can be supported at the Si/SiO_2 interface at steady state is simply

$$\Delta C_I = g_I / \sigma_I \ , \tag{12.5}$$

which results from setting the flux in, g_I, equal to the flux out, $\sigma_I(C_I - C_I^*)$. If we include bulk recombination, the excess will necessarily be less than this because some of the injected I recombine with V. But what determines the undersaturation level of V? Actually, there are two factors: the bulk and surface properties of the silicon. Under equilibrium conditions, the net bulk generation and recombination rates of I and V are equal. During I injection, there is an increase in recombination rate because of the excess I, resulting in a deficit of V. But because of this deficit, the balance at the surface between flux in and flux out of V is upset. Since there are fewer V available for surface loss, there is a net flux of V into the silicon (assuming $g_V = 0$) from the oxidizing surface, which tends to replenish V. This leads to the requirement that

$$C_I C_V > C_I^* C_V^* \tag{12.6}$$

in the near-surface region. Ignoring the effect of the surface, we would arrive at the relationship of Eq. (12.4).

Figure 19, from Hu's analysis (1985b), compares the behavior of C_I and C_V predicted by solving Eqs. (12.1) and (12.3) with that predicted by assuming the relation in Eq. (12.4) *a priori* [or by solving Eqs. (12.1) with $\sigma_I = 0$]. Although V is everywhere undersaturated because of recombination with I, the surface acts as a natural source of vacancies, so that only for regions in the bulk some distance away from any surfaces is the relationship of Eq. (12.4) met. A more complete numerical analysis of the coupled equations (12.1) has been presented by Yeager and Dutton (1985) and shows that much more complex behavior for V can occur than that indicated by Fig. 19. In Fig. 20 we show the results of simulations using the SUPREM IV program. Figure 20(a) shows the kinetic be-

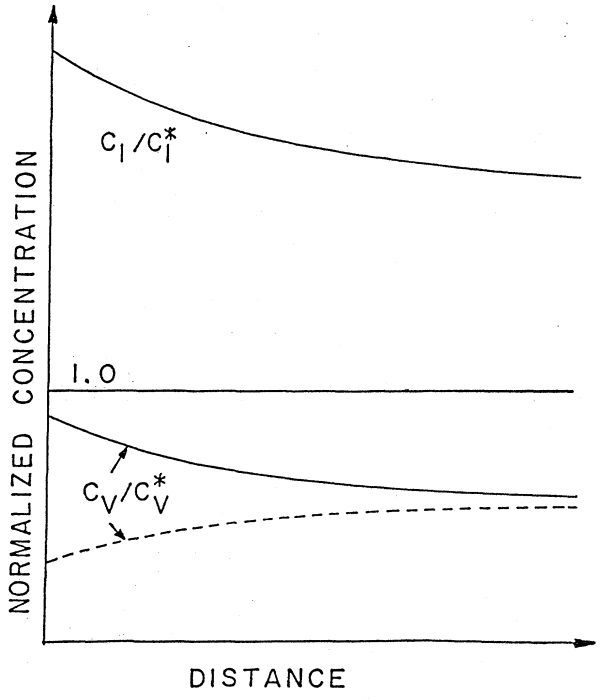

FIG. 19. Behavior of undersaturated vacancy concentration that results from I injection. Dashed line shows the behavior predicted by assuming the relationship of Eq. (12.4). Solid line shows the behavior predicted by Hu (1985b) by solving Eqs. (12.1) utilizing the boundary conditions of Eq. (12.3).

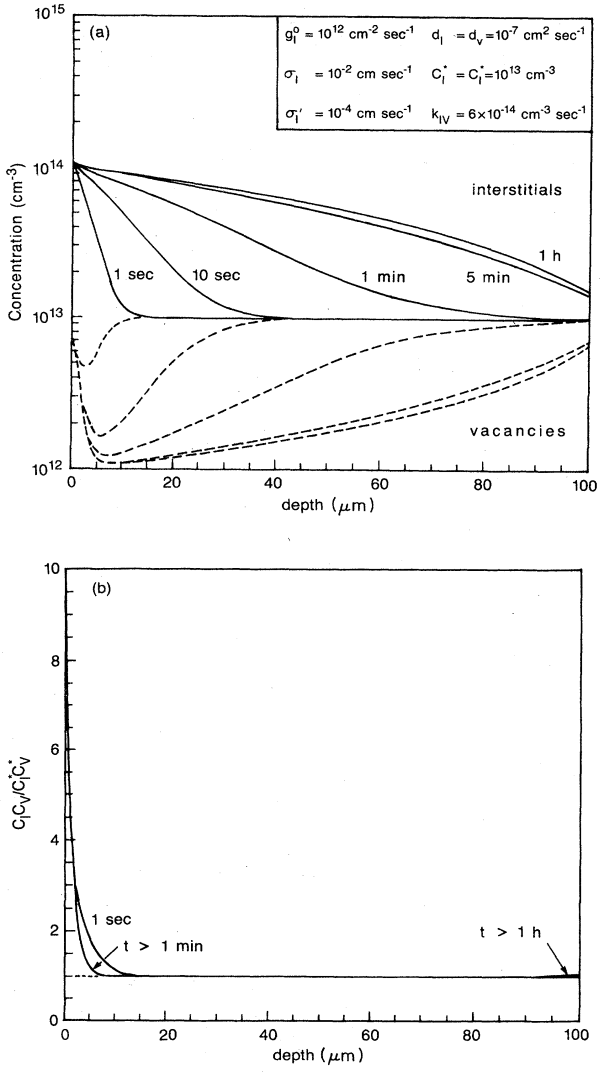

FIG. 20. Numerical simulation of the coupled point-defect equations: (a) kinetic behavior of C_I and C_V; (b) corresponding behavior of the $C_I C_V$ product. The value of $k_{I,V}$ is chosen to correspond to the case of diffusion-limited recombination between I and V; the existence of an energy barrier to recombination will lessen the coupling between the two point-defect species.

havior of V and I across a 100-μm wafer using the numerical parameters of the inserted table (these numbers were chosen based on the experiments described in Secs. XV and XVI. Figure 20(b) shows the corresponding behavior of $C_I C_V / C_I^* C_V^*$.

Now it may be argued that perhaps these analyses are invalid for a number of reasons. What if $g_V \neq 0$ and is actually negative (i.e., absorbs V during oxidation), thus reducing the effectiveness of the surface as a source of V during oxidation? This question has already been addressed by Hu (1985b). By including g_V in the analysis that leads to Eq. (12.6), it was shown that this inequality remains valid. One may also ask, what if σ_V of an oxidizing surface is so small that the surface cannot replenish the substrate with vacancies faster than they are annihilated by I-V recombination? In fact, if the oxidizing surface *could* replenish the substrate with V as fast as they are annihilated by I-V recombination, there would not *be* any V deficit. To obtain the relationship of Eq. (12.4) what one really assumes is that the rate of supply of V to the bulk from an oxidizing surface is small compared to the Frenkel pair production rate. This is the qualitative requirement of any set of boundary conditions, not just of Eqs. (12.3); this requirement need not be satisfied in general.

B. Surface and bulk processes

In this section the physical interpretation of the quantities g_I, σ_I, and $k_{I,V}$ are examined. Alternative models for surface conditions are discussed in Sec. XIV.E.

1. Surface flux

The surface fluxes g_I and g_V in the boundary conditions of Eq. (12.3) result from chemical reactions; therefore their functional form depends on the surface reaction under discussion. For the case of oxidation, Hu (1985c) has examined the consequences of assuming that $g_I(t)$ is proportional to the oxidation rate. Because oxidation of silicon follows a well-known linear/parabolic growth rate, reflecting the reaction-limited and diffusion-limited time regimes of SiO_2 growth, this leads to $g_I(t)$ of the form (Deal and Grove, 1965)

$$g_I(t) = \frac{A}{\sqrt{t_0 + t}} , \tag{12.7}$$

where t_0 is a time constant delineating the transition between linear ($t \ll t_0$) and parabolic ($t \gg t_0$) growth rates. A more general and convenient expression for g_I (or g_V) is

$$g_I(t) = A\,(t_0 + t)^{-n} . \tag{12.8}$$

This allows the empirical modeling of the surface flux as proportional to some power and also ensures, through the use of t_0, that the flux does not become infinite at time zero. For the case of *constant source injection,* we write

$$g_I(t) = g_I^0 \tag{12.9}$$

in all of the following discussions.

2. Surface loss

Writing the flux at the surface due to surface loss as

$$-\sigma_I (C_I - C_I^*) \tag{12.10}$$

is in effect stating a phenomenological model. It allows one to change boundary conditions between the extremes of a surface that is a perfect sink for point defects and one that is a perfect reflector. The basic assumption is that at any given instant of time, aside from any processes caused by chemical reactions, a certain fraction of I at the surface will leave the silicon substrate, and that a constant number of I are generated per unit time at the surface through normal thermal processes. One situation in which time-dependent σ_I is expected is the case of thermal nitridation of SiO_2. As discussed in Sec. XIV, a thermally grown SiO_2 layer, when exposed to NH_3, undergoes some chemical reaction that changes the composition of the film (and, in particular, the composition at the Si interface) and results in I injection into the bulk. Experiments also suggest (Sec. XV) that an oxidizing surface has the ability to reabsorb I at a much faster rate than the nonoxidizing Si/SiO_2 interface, implying the necessity of a time-dependent σ_I in Eq. (12.10).

These last two examples already are strong motivation for trying to understand what the physical meaning of σ_I is and whether the underlying assumptions that lead to Eq. (12.10) might break down under realistic cases. Very little work has been done on this problem. Hu (1985b) has considered a model in which surface loss is due to the capture of I at surface kinks. If there is no energy barrier to capture, σ_I is given by

$$\sigma_I = \pi \rho a_0 d_I , \tag{12.11}$$

where ρ is the density of surface kinks at which the capture is assumed to take place, and a_0 is a capture radius of atomic dimension. Scheid and Chenevier (1986) have raised the pertinent question of whether the number of capture sites might not become saturated at some point. Clearly, more thought must be given to this problem area.

In the following treatment, surface recombination velocities are assumed constant with time, although, as discussed below, for steady-state conditions the surface can be treated as having time-dependent g_I and σ_I.

3. I-V generation and recombination

Since the coupling between I and V is determined by the process of their mutual annihilation, a brief discussion of the meaning and magnitude of $k_{I,V}$ is warranted.

Bulk generation and recombination of I and V are described by the reaction equation

$$\frac{dC_I}{dt}=\frac{dC_V}{dt}=-k_{I,V}(C_IC_V-C_I^*C_V^*)\ . \tag{12.12}$$

Before writing down the atomistic constituents of $k_{I,V}$, it is worthwhile to point out that, in general, there will be some reaction barrier to I-V recombination. This can be appreciated by picturing the inverse process of Frenkel pair generation. Once I and V are created, they may diffuse away from each other until they are no longer interacting. The lattice surrounding both I and V will relax to some extent in order to minimize the energy state in the crystal. To recombine, the lattice surrounding I and V must return to the unrelaxed (higher-energy) state. This is the underlying physical basis for an energy barrier to I-V recombination. The process of I-V generation and recombination is depicted in Fig. 21. We imagine that a representative volume of silicon is allowed to be in different states specified by the presence and configuration of I and V defects. This diagram is very general. It should also be mentioned that it is possible for the Frenkel pair to exist as a stable configuration, which requires that there be a local minimum in energy when I and V are adjacent.

The recombination constant $k_{I,V}$ can be expressed as

$$k_{I,V}=\frac{4\pi a_{I,V}}{\Omega C_S}(d_I+d_V)\exp\left[-\frac{\Delta E_{I,V}}{kT}\right]\ , \tag{12.13}$$

where $a_{I,V}$ is the capture radius for I-V recombination, Ω is the volume of the silicon unit cell ($a^3/8$, with the lattice parameter $a=5.43$ Å), C_S is the density of lattice sites ($=5\times10^{22}$ cm^{-3}), and $\Delta E_{I,V}$ is the recombination energy barrier as designated in Fig. 21. Experimental evidence of an energy barrier to I-V recombination is discussed further in Sec. XIV.C.

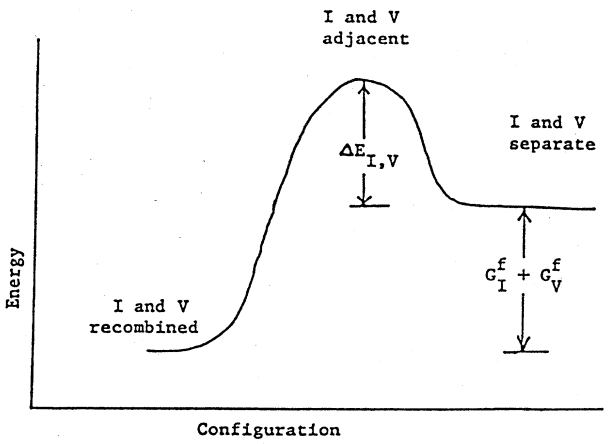

FIG. 21. Energy diagram of I-V recombination.

C. Decoupled equations

A great amount of insight into the nonequilibrium behavior of point defects can be gained by decoupling Eqs. (12.1). This can be done by rewriting the recombination term as

$$k_{I,V}(C_IC_V-C_I^*C_V^*)\Longrightarrow k_I(C_I-C_I^*)\ , \tag{12.14}$$

using I injection as an example and setting $k_I=k_{I,V}C_V^*$. This condition is fulfilled exactly if C_V is maintained at its equilibrium value even under I injection. This happens in one of two extreme cases. (i) There is a large energy barrier to recombination, so V has a negligible effect on I. (ii) Either C_V^* is very large compared to C_I, or V is replenished in the bulk in a negligible amount of time, so that, in either case, I has little effect on V, but I-V recombination has its maximum possible effect on I. It is easy to see the usefulness of this analysis, since the actual situation must lie between the two extremes of (i) and (ii). Furthermore, k_I can also be used to represent any bulk recombination process in which the recombination rate is proportional to the I excess.

In the following discussions we shall treat the one-dimensional case of I injection from one surface of a silicon wafer, and shall look primarily at the behavior of the resulting I excess

$$\Delta C_I\equiv C_I-C_I^*\ . \tag{12.15}$$

The equivalent treatment for V injection is of course the same. The analytic solutions presented in the following sections also provide sufficient conditions for the decoupling of the influence of V on the I diffusion behavior during I injection. As it turns out, these conditions are overly restrictive in some cases. This is because the analytic solutions assume a time-invariant concentration of recombination centers, while in the case of V as recombination centers, C_V and C_I are coupled through Eqs. (12.1). This is discussed in more detail after the analytic results have been presented.

We present results for three situations: (i) A wafer thin enough that the time required for point defects to diffuse across the wafer is negligible compared to surface and bulk reaction rates. (ii) A wafer thick enough that it may be considered a semi-infinite substrate. (iii) A wafer of arbitrary thickness, which demonstrates under what conditions point-defect behavior may be approximated by treating the substrate as situations (i) and (ii).

1. Uniform concentration approximation

As in the analysis of Hu (1985b, 1985c), we consider the case of a silicon wafer so thin that the value of ΔC_I can be assumed constant across the entire substrate. The criteria for such an assumption in terms of the physical parameters of the problem are discussed below, in the sections on wafers of arbitrary thickness (Secs. XII.C.3 and XII.C.4).

For a wafer of thickness h, with injection from one side, the equation for the uniform concentration approximation is

$$\frac{\partial \Delta C_I}{\partial t} = \frac{g_I}{h} - \frac{\sigma_I + \sigma_I'}{h}\Delta C_I - k_I \Delta C_I, \quad 0 \le x \le h \ , \tag{12.16}$$

where σ_I and σ_I' are the I recombination velocities on either side of the wafer. Equation (12.16) is a linear ordinary differential equation of first order and so can be integrated directly, inclusive of cases where g_I, σ_I, σ_I', and k_I are arbitrary functions of time. For the particular case of constant injection, $g_I(t) = g_I^0$, the solution is

$$\Delta C_I(t) = \frac{g_I^0}{2\bar{\sigma}_I(1+\tau_h/\tau_I)}\{1-\exp[-(1/\tau_h + 1/\tau_I)t]\} \ , \tag{12.17}$$

where

$$\bar{\sigma}_I \equiv \frac{\sigma_I + \sigma_I'}{2} \ , \tag{12.18}$$

$$\tau_h \equiv h/2\bar{\sigma}_I \ , \tag{12.19}$$

$$\tau_I \equiv 1/k_I \ . \tag{12.20}$$

So, with constant injection, ΔC_I rises like a charging capacitor and saturates at its steady-state value, $g_I^0/2\bar{\sigma}_I(1+\tau_h/\tau_I)$. The parameters τ_h and τ_I are the surface and bulk lifetimes, respectively, of I. The effect of bulk recombination is to shorten the approach to steady state and reduce the value of ΔC_I. Bulk recombination is unimportant if

$$\frac{\tau_h}{\tau_I} = \frac{hk_I}{2\bar{\sigma}_I} \ll 1 \ . \tag{12.21}$$

This is just the behavior one would suspect, since the effectiveness of the surfaces as sinks is independent of the thickness of the wafer, whereas the importance of the bulk increases with thickness since there will then be more recombination sites available per unit time. Equation (12.17) is plotted in Fig. 22 as a function of normalized time scale for the semi-infinite wafer, as described below.

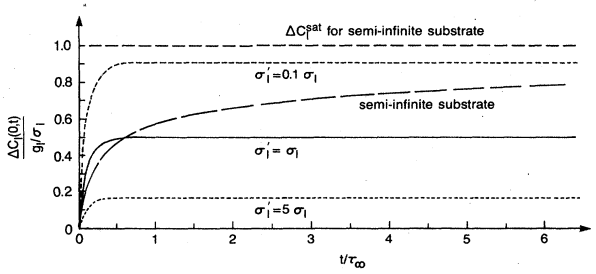

FIG. 22. Kinetics of I surface excess $\Delta C_I(0,t)$ for a semi-infinite substrate, and for a thin wafer assuming the uniform-concentration approximation and injection from one side. The solutions for the thin wafer depend on the relative value of the surface recombination velocity on the injecting side of the wafer, σ_I, and the noninjecting side, σ_I'.

For g_I not constant with time, the behavior of ΔC_I is a much more complicated function of time. The general solution of Eq. (12.16), assuming a time-dependent g_I of the form of Eq. (12.8), is included in Appendix C.

2. Semi-infinite substrate

Silicon wafers used both for making integrated circuits and in most diffusion experiments reported in the literature are typically about 300–500 μm thick. As the diameters of wafers used in the integrated circuit industry grow progressively larger, the thicknesses of the wafers continue to increase also (for mechanical stability). It is of interest, then, to analyze the predictions of point-defect behavior when it is assumed that injection takes place into a semi-infinite substrate.

For a semi-infinite substrate we have

$$\frac{\partial \Delta C_I}{\partial t} = d_I \frac{\partial^2 \Delta C_I}{\partial x^2} - k_I \Delta C_I, \quad 0 < x < \infty \tag{12.22}$$

subject to the boundary condition

$$g_I + d_I \frac{\partial C_I}{\partial x} - \sigma_I (C_I - C_I^*) = 0, \quad x = 0 \ . \tag{12.23}$$

As an example, we take the case where the influx of I is constant with time and is given by g_I^0, and bulk recombination is unimportant. The solution to this problem is (Carlsaw and Jaeger, 1959; Hu, 1985c)

$$\Delta C_I(\lambda,t) = \frac{g_I}{\sigma_I}[\mathrm{erf}(\lambda) - e^{2\sqrt{t/\tau_\infty}\lambda + t/\tau_\infty} \times \mathrm{erfc}(\lambda + \sqrt{t/\tau_\infty})] \ , \tag{12.24}$$

where

$$\lambda \equiv x/2\sqrt{d_I t} \ , \tag{12.25}$$

$$\tau_\infty \equiv d_I/\sigma_I^2 \ . \tag{12.26}$$

The parameter λ is a normalized depth in terms of diffusion lengths, while τ_∞ is the characteristic surface lifetime of I for a semi-infinite substrate, and can be contrasted to the surface lifetime of the uniform concentration approximation, $\tau_h = h/2\bar{\sigma}_I$.

Equation (12.24) is plotted in Fig. 22. It can be seen that the kinetics are very different for the semi-infinite substrate and the thin wafer, but if the surface at h is a poorer sink for excess I than the injecting surface at $x = 0$, then they both approach the same steady-state value of $\Delta C_I = g_I/\sigma_I$.

Hu (1985c) has also provided some analytic solutions for g_I with a linear/parabolic behavior and for constant source injection including bulk recombination. The effect of bulk recombination is similar to that in the uni-

form concentration approximation; at each point in space, bulk recombination will tend to shorten the time of approach to steady state and reduce the magnitude of ΔC_I.

3. Wafer of arbitrary thickness h: steady-state behavior

In some experimental situations, dopant diffusion and point-defect behavior seem to indicate that the silicon wafer is like a semi-infinite substrate, while in others, the finite thickness of the wafer is very apparent. For example, the magnitude of oxidation-enhanced diffusion on one side of a wafer does not depend on the surface condition on the other side of the wafer, which is compatible with the notion that the wafer is thick enough to be considered a semi-infinite substrate. On the other hand, for situations such as gettering of heavy metals (Sec. XVI.A) it is believed that the process is aided in part by I defects generated at one side of the wafer quickly diffusing throughout the entire substrate. Obviously, the types of behavior of ΔC_I in the uniform concentration approximation and in a semi-infinite substrate are special limiting cases of the general situation of point-defect injection into a wafer of arbitrary thickness. The analysis of this general case clarifies what behavior can be expected for a given set of experimental conditions.

For a wafer of thickness h, we are interested in obtaining solutions to the equations

$$\frac{\partial \Delta C_I}{\partial t}=d_I\frac{\partial^2 \Delta C_I}{\partial x^2}-k_I\Delta C_I, \quad 0<x<h \tag{12.27}$$

subject to the boundary conditions

$$\begin{aligned} &g_I+d_I\frac{\partial \Delta C_I}{\partial x}-\sigma_I\Delta C_I=0 \quad \text{at } x=0 , \\ &d_I\frac{\partial \Delta C_I}{\partial x}+\sigma_I'\Delta C_I=0 \quad \text{at } x=h . \end{aligned} \tag{12.28}$$

The analytic solution to this problem is a bit unwieldy compared to the solutions presented thus far and so is deferred to Appendix C.

A good amount of insight into the problem can be gained by examining the steady-state solutions. The general steady-state solution to Eq. (12.27) is given by

$$\frac{\Delta C_I}{g_I/\sigma_I}=\frac{(1+\sqrt{s'})\exp[(h-x)/L_I^b]-(1-\sqrt{s'})\exp[-(h-x)/L_I^b]}{(1+\sqrt{s'})(1+\sqrt{s})\exp(h/L_I^b)-(1-\sqrt{s'})(1-\sqrt{s})\exp(-h/L_I^b)} , \tag{12.29}$$

where

$$L_I^b\equiv 2\sqrt{d_I/k_I}=2\sqrt{d_I\tau_I} , \tag{12.30}$$

$$s\equiv k_I d_I/\sigma_I^2=\tau_\infty/\tau_I=(L_I^s/L_I^b)^2 , \tag{12.31}$$

$$s'\equiv k_I d_I/\sigma_I'^2 . \tag{12.32}$$

The parameter L_I^b is the diffusion length before recombination in the bulk. The parameter s is a measure of the importance of surface loss compared to bulk recombination. It is the ratio of I lifetime before loss at the surface to I lifetime before recombination in the bulk. Alternatively, $\sqrt{s}$ can be thought of as the ratio of the diffusion length before loss at the surface to the diffusion length before recombination in the bulk. As defined in Fig. 18(a),

$$L_I^s\equiv 2d_I/\sigma_I=2\sqrt{d_I\tau_\infty} . \tag{12.33}$$

The expression of s in terms of characteristic surface and bulk diffusion lengths is convenient, since it is possible to measure by experiment the decay length of point defects along a surface as well as the decay length in the bulk. As shown in Sec. XVI, for the case of the Si/SiO_2 surface it appears that

$$L_I^s\ll L_I^b , \tag{12.34}$$

which, by definition of the quantities involved, is the same as stating that

$$\tau_\infty\ll\tau_I \tag{12.35}$$

(i.e., surface lifetime is much less than bulk lifetimes). Thus s is always much less than unity.

Equation (12.35) is a very useful constraint in the analysis of I injection. To demonstrate this, we begin our analysis of the general case of I injection into a wafer of thickness h by noting that the behavior of ΔC_I, as developed in the previous sections, is characterized by three time constants, τ_I, τ_h, and τ_∞. There are then six possible orderings among the three time constants. However the constraint imposed by Eq. (12.35) reduces this to only three possible orderings:

$$\begin{aligned} &\text{(i)} \quad \tau_\infty<\tau_h<\tau_I , \\ &\text{(ii)} \quad \tau_h<\tau_\infty<\tau_I , \\ &\text{(iii)} \quad \tau_\infty<\tau_I<\tau_h . \end{aligned} \tag{12.36}$$

Each case can then be analyzed separately, making the problem tractable.

Cases (i) and (ii): Physically, cases (i) and (ii) correspond to the situation in which bulk recombination is not important and the behavior of ΔC_I is determined only by the thickness of the wafer.

For simplicity we first consider the case in which the recombination velocities are the same on both sides of the wafer, $\sigma_I=\sigma_I'$, and bulk recombination is negligible. Under these conditions the steady-state behavior of ΔC_I is described by the equation

$$\frac{\Delta C_I}{g_I/\sigma_I} = \frac{\varphi}{1+\varphi}\frac{h-x}{h} + \frac{\frac{1}{2}}{1+\varphi} , \tag{12.37}$$

where we have introduced a new, dimensionless variable,

$$\varphi \equiv \frac{\tau_h}{\tau_\infty} = \frac{h\sigma_I}{2d_I} = \frac{h}{2\sqrt{d_I\tau_\infty}} . \tag{12.38}$$

The factor φ is inversely proportional to how many times, on the average, an I defect may diffuse across the width of the wafer before being lost at one of the surfaces. Limiting cases for $\varphi \ll 1$ and $\varphi \gg 1$ are shown in Fig. 23(a). From this figure it is readily apparent that the condition for validity of the uniform concentration approximation is $\varphi \ll 1$. That Eq. (12.29) reduces to Eq. (12.37) in the limit of negligible bulk recombination can be easily demonstrated as follows. The condition $\tau_\infty, \tau_h \ll \tau_I$ in fact requires that $h \ll L_I^b$. By expanding the exponentials in Eq. (12.29) and keeping only linear terms, we see that Eq. (12.29) reduces to Eq. (12.37).

In the more general case where $\sigma_I \neq \sigma_I'$,

$$\frac{\Delta C_I}{g_I/\sigma_I} = \frac{\varphi^{\text{eff}}}{1+\varphi^{\text{eff}}}\frac{h-x}{h} + \frac{1-\frac{1}{2}(\varphi^{\text{eff}}/\varphi)}{1+\varphi^{\text{eff}}} , \tag{12.39}$$

where

$$\varphi^{\text{eff}} \equiv \frac{h\sigma_I\sigma_I'}{d_I(\sigma_I+\sigma_I')} . \tag{12.40}$$

If $\sigma_I = \sigma_I'$, then $\varphi^{\text{eff}} = \varphi$ and Eq. (12.39) reduces to Eq. (12.37).

Case (iii): The ordering of time constants in case (iii) corresponds to the situation in which bulk recombination cannot be ignored, and the spatial behavior of ΔC_I is determined by bulk recombination rates, even though the rate of surface loss is faster than the rate of bulk recombination.

For the condition $\tau_\infty, \tau_I \ll \tau_h$, it necessarily follows that

$$\tau_\infty \tau_I \ll (\tau_h)^2 , \tag{12.41}$$

which means that

$$h \gg L_I^b . \tag{12.42}$$

In this situation, Eq. (12.29) can be simplified to

$$\frac{\Delta C_I}{g_I/\sigma_I} \approx \frac{\exp(-x/L_I^b)}{1+\sqrt{s}} . \tag{12.43}$$

This form was used early on in some simulation programs without justification. It is obvious now that the validity of the exponential falloff of C_I with depth in the bulk requires that: (i) $s \ll 1$ for the injecting surface, (ii) the wafer thickness be much greater than L_I^b, and (iii) steady state has been reached. If these conditions are fulfilled, the time dependence of ΔC_I reflects the time dependence of $g_I/[\sigma_I(1+\sqrt{s})]$.

Figure 23(b) shows the steady-state behavior of ΔC_I for situations intermediary to the limiting cases discussed above of $h \gg L_I^b$ or $h \ll L_I^b$, maintaining only the assumption of $s \ll 1$.

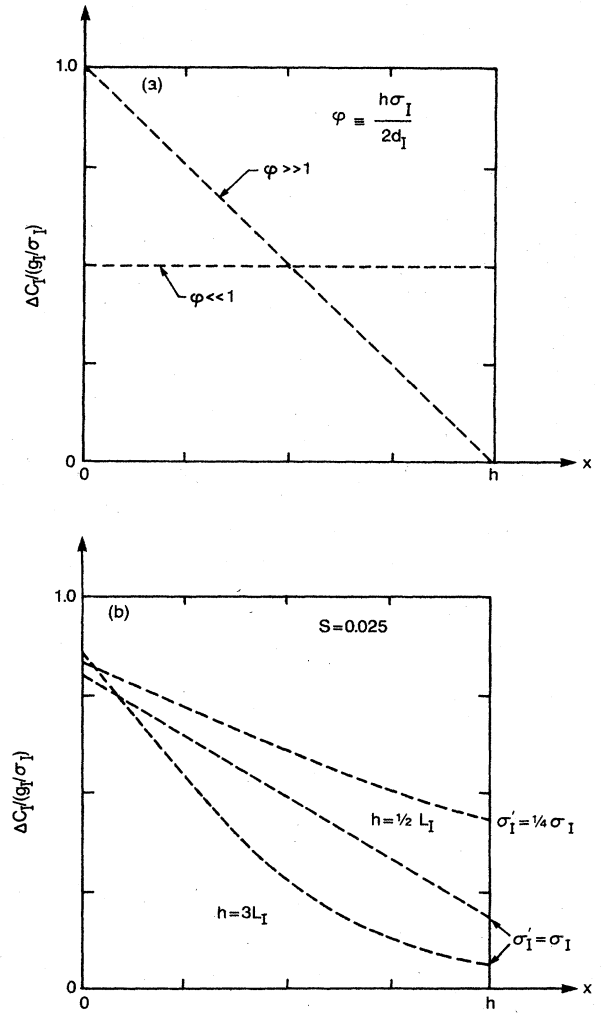

FIG. 23. Steady-state solutions for ΔC_I: (a) ignoring bulk recombination and assuming $\sigma_I = \sigma_I'$; (b) including bulk recombination.

4. Wafer of arbitrary thickness h: transient behavior

We now examine a couple of illustrative examples of the transient behavior of $C_I(x,t)$ in a wafer of thickness h during surface injection.

An interesting example is the approach to equilibrium of a wafer, initially at a temperature T_i, suddenly elevated to a higher temperature T_f, with no external chemical reactions (i.e., $g_I = 0$). A general plot of this situation cannot be made. The particular situation depicted in Fig. 24 is for a 100-μm-thick wafer. We have chosen the value of $L_I^s = 10$ μm, which is a reasonable value for a nonoxidizing Si/SiO_2 interface at temperatures around

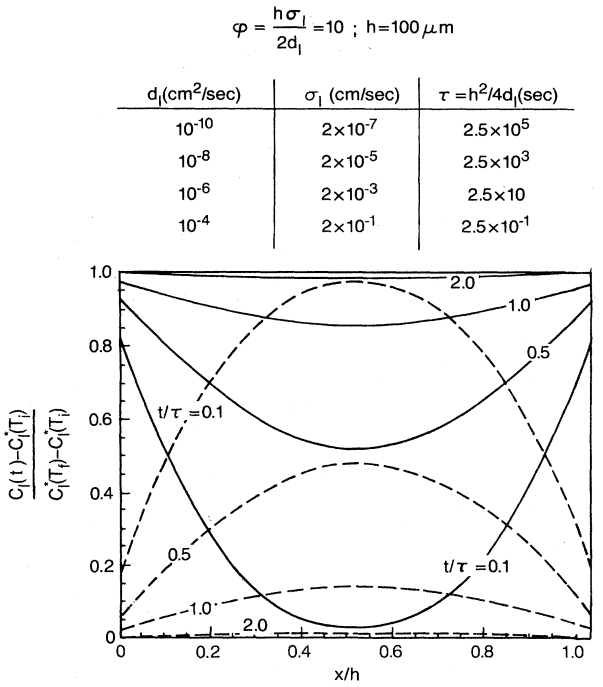

FIG. 24. Transient behavior of ΔC_I for a 100-μm-thick wafer, initially at temperature T_i, raised to T_f (solid line), and then cooled back to T_i. Bulk generation and recombination processes are neglected.

1100 °C (Sec. XVI). The solid lines show the wafer approaching its equilibrium value of $C_I^*(T_f)$ from an initial value of $C_I^*(T_i)$, assuming that the surfaces are the dominant source of point defects. The time scale has been expressed in terms of the normalized time

$$\tau = h^2/4d_I \ , \tag{12.44}$$

which is simply the time of one diffusion length across the wafer. Since the ratio of d_I/σ_I is fixed (i.e., L_I^s is fixed in this example), the actual value of τ is that given in the accompanying table. The dashed lines show the decay of C_I if the wafer is lowered again to temperature T_i after attaining equilibrium at T_f, assuming that the ratio of d_I/σ_I is constant (that is, that d_I and σ_I have the same temperature dependence). This is satisfied exactly in Hu's model for σ_I [Eq. (12.11)]. Note, however, that since the diffusivity decreases with temperature, the time scale will change as indicated in the table of Fig. 23.

What if we now include the processes of bulk generation and recombination? In Fig. 25 we use the same parameters for I as in Fig. 24 and include the process of bulk generation and recombination. In the simulation, we have assumed a value for $k_{I,V}$ that corresponds to diffusion-limited recombination (i.e., no energy barrier to I-V recombination) and have specified the same point-defect properties for V as for I, except for the specification of $C_V^* = 5C_I^*$. Note the interesting effect of the I concentration first reaching levels above C_I^* and then relaxing to its equilibrium value. Note also the

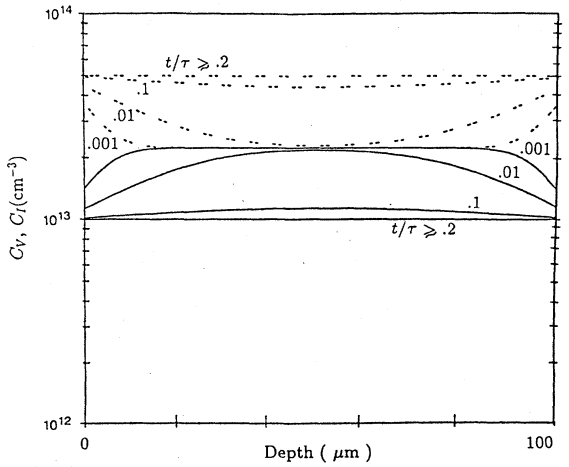

FIG. 25. Simulation similar to Fig. 24, but including bulk generation and recombination. Interstitial behavior is indicated by solid lines and vacancy behavior by the broken lines. The value of d_I is 10^{-7} cm^2/sec, $\sigma_I = 2\times10^{-4}$ cm/sec, and $k_{I,V} = 6\times10^{-14}$ cm^{-3}sec^{-1}. Parameters for V are chosen to have the same values as those for I, except that $C_V^* = 5C_I^*$.

significant reduction in time needed to achieve equilibrium (significant reduction in time occurs also if $C_V^* = C_I^*$).

With today's experimental techniques and equipment it is possible to heat and cool wafers rapidly and observe diffusive movement of dopants for times on the order of a minute by making the initial dopant profile very thin (ideally approaching a delta function in concentration). It may be possible either to determine values of d_I or d_V by observing the lag in time until equilibrium diffusion rates are achieved, or to put a lower limit on point-defect diffusivities by our inability to observe transient effects.

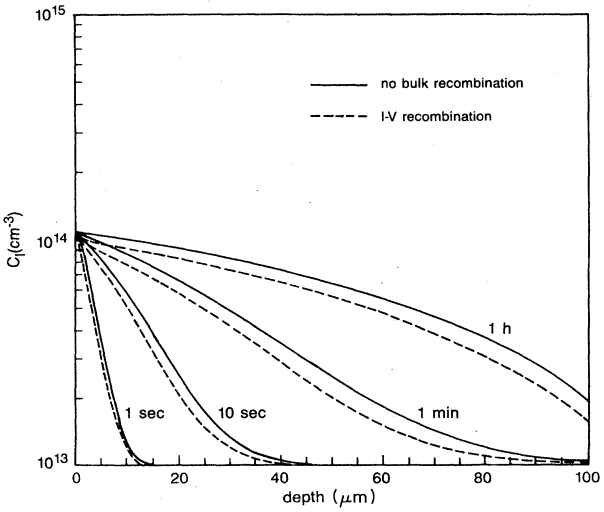

FIG. 26. The flow of I: dashed lines, using the same parameters as in Fig. 20(a); solid lines, ignoring the effect of V.

As another example, we show a comparison of C_I calculated using the solution in Appendix C under conditions identical to those used in the simulation of Fig. 20, except that $k_{I,V}$ is set equal to 0. The closeness of the two solutions in Fig. 26 indicates that V has only a minor effect on the flow of I, whereas the excess I have a major effect on C_V. But note that, with the parameters chosen, the thickness of the wafer, $h = 100$ μm, is much greater than $L_I^b \approx 8$ μm, so that from the analytic solution of Appendix C we would expect a much stronger effect of the vacancies on ΔC_I. As alluded to earlier, the analytic solutions based on a time-invariant concentration of recombination centers overestimates the effect of V on I.

D. Criteria for decoupling of *I-V* equations

Understanding the criteria under which it is valid to decouple Eqs. (12.1) is of interest for two reasons. One reason is that the problem of simulating point-defect behavior becomes much easier from a numerical point of view if I and V are not coupled. The other reason is that, in interpreting experimental results, one often needs to know whether coupling between I and V is responsible for some of the effects observed. For example, in measuring the time it takes I to traverse a silicon wafer, is the apparent diffusivity really d_I, or are we measuring an effective diffusivity that results from coupling between I

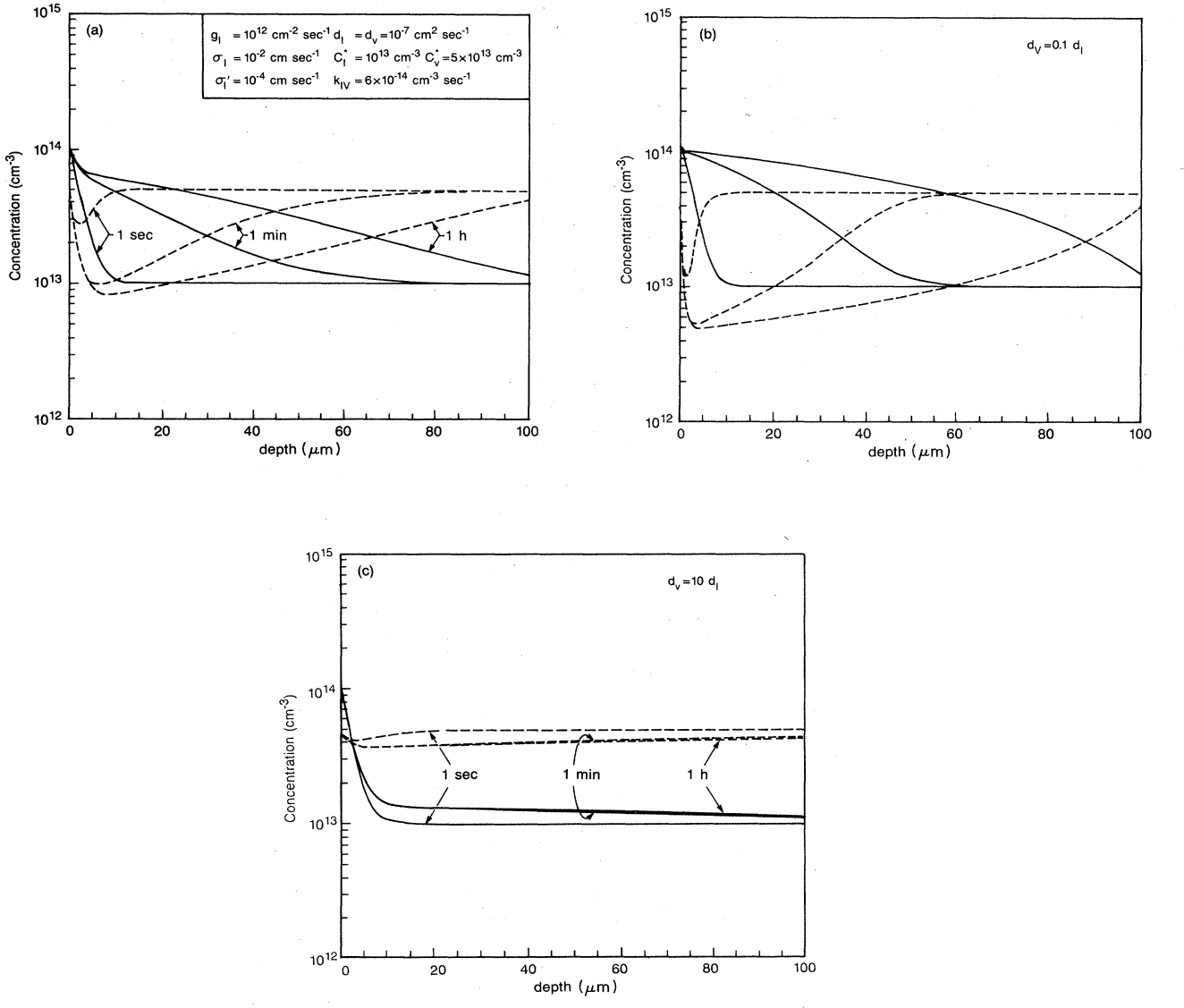

FIG. 27. The effect of *I-V* recombination on the flow of I: (a) same parameters as Fig. 20(a) except C_V^* is raised to 5 times C_I^*; (b) reducing d_V by a factor of 10; (c) increasing d_V by a factor of 10.

and V? One may think that the effect of I-V recombination can be only to slow the progression of I through the wafer. In fact, it is possible for the coupling between I and V to reduce the transit time for I across the wafer during injection [an effect already noted by Yeager and Dutton (1985)], as we shall show below.

Intuitively, it may seem very reasonable that only over a distance less than L_I^b can one expect I injection not to be affected strongly by bulk recombination. But if the recombination centers are vacancies, this condition is overly restrictive, as one can see by examining two extreme cases. If the recombination constant $k_{I,V}$ is very small, there is very little recombination, so that C_V will be maintained at approximately C_V^*, and V will have little effect on I. If, on the other hand, $k_{I,V}$ is very large, it does *not* necessarily follow that the presence of V strongly affects the flow of I, for it may be the case that V are annihilated very quickly and their effect on I is minimal. Thus it may happen that, over distances greater than L_I^b, the presence of V does little in determining the flow of I. The effect of V on I can only be determined exactly by solving the coupled point-defect equations of Eq. (12.1). However, we can make some reasonable estimates as follows.

If $C_I^* \gg C_V^*$, then I flow independently of V.

Noting that we can always state that

$$\Delta C_I(0,t) \geq \frac{g_I}{\sigma_I(1+\sqrt{s})} , \tag{12.45}$$

the case of $C_V^* > C_I^*$ can be broken down roughly into two situations.

If $C_V^* \gg C_I^*$, and

$$\frac{g_I}{\sigma_I(1+\sqrt{s})} \ll C_V^* , \tag{12.46}$$

then $C_V \approx C_V^*$ during I injection and the restrictions of the analytic solutions apply.

If $C_V^* > C_I^*$, and

$$\frac{g_I}{\sigma_I(1+\sqrt{s})} \gtrsim C_V^* , \tag{12.47}$$

the importance of V in affecting the flow of I depends on what aspect of I flow is under study. One important aspect is whether the supersaturation of I (i.e., the value of C_I/C_I^*) is affected in the near-surface region, since this is the quantity that determines the important phenomenon of oxidation-enhanced diffusion (Sec. XIV). We already see from Eq. (12.45) that for values of $\sqrt{s} = L_I^s/L_I^b$ less than one, which seems to be the case from experiment, the presence of V will have only a minor effect on the value of $\Delta C_I(0,t)$. Within a distance less than $L_I^b = 2(d_I/k_{I,V}C_V^*)^{1/2}$ away from the injecting surface, the presence of V will not cause I to decay substantially from this value. The exact criterion for a wafer of thickness h can be deduced from the analytic solutions in Appendix C. If we are concerned with the transit time of I across the wafer, then we must consider two cases: $d_I \ll d_V$ and $d_I \gg d_V$. We show in Fig. 27(a) simulation results using the same parameters as those of Fig. 20 except for an increased equilibrium concentration of C_V^*, so that $C_V^* > C_I^*$. In Fig. 27(b) we show the result of reducing d_V by a factor of 10. This has the effect of retarding the progression of the excess interstitial wave front through the wafer. Figure 27(c) shows the situation for which d_V is 10 times greater than d_I. This has the effect of shortening the time it takes for excess I to traverse the wafer, although the I profile is highly distorted compared to the other cases seen thus far. Note that this last effect requires that $d_V C_V^* \gg d_I C_I^*$. Determining the mechanism of self-diffusion would determine the magnitude of the ratio of $d_V C_V^*$ to $d_I C_I^*$ for a given temperature (Sec. IX).

E. Bulk recombination with defects other than vacancies

Diffusion experiments are not always reproducible from laboratory to laboratory. This naturally invites the question of whether there is some difference between the silicon material used in separate studies that affects point defects. We have already mentioned that oxygen in silicon is known to affect point-defect behavior (Sec. IV.C), and that it may explain the poor agreement between various self-diffusion studies (Sec. IX). Carbon is another impurity present in all silicon wafers that is known to interact with point defects. We have also mentioned the existence of a class of microdefects known as swirl defects, which are thought to be the aggregates of native point defects that coalesce during the cooling of silicon in the growth process. These few examples illustrate that there are always some defects in the bulk, and this raises the possibility that point defects may interact with them and possibly affect the outcome of otherwise well-controlled experiments.

Formulation of bulk recombination processes of excess point defects depends on the nature of the defects with which I and V recombine. Here, we briefly examine the simplest types of recombination reactions, which nevertheless demonstrate the essential effects that the presence of bulk defects will have on excess point-defect flow. We continue to use the example of I injection from one side of a silicon wafer. One important result of this analysis is that these reactions with non-native defects offer a possible explanation for the inconsistent results of some diffusion experiments; certain experiments indicate point defects have diffusivities greater than 10^{-7} cm^2/sec^{-1}, while others yield values at the same temperature on the order of only about 10^{-9} cm^{2}sec^{-1}. This problem is discussed in more detail in Sec. XVI.

1. I injection with trapping

It is known from low-temperature irradiation experiments of Watkins and co-workers that impurities in sil-

icon are very effective at trapping self-interstitials. This trapping is the reason that isolated self-interstitials have not been found in irradiated samples and is also convincing evidence that the diffusion of self-interstitials is very high, even at low temperatures (4.2 K); see Sec. VII.B and the review article by Watkins (1975). In this section we consider trapping at much higher temperatures (> 800 °C) where dopant diffusion occurs. This presentation will show that self-interstitials with very high diffusivities can appear to diffuse much more slowly because of their periodic trapping at impurities. This means that the "effective" diffusivity that is measured experimentally can have a significantly different activation energy of diffusion and preexponential factor than the "free" interstitial [see the review paper by Wert and Frank (1983) for a fuller discussion on trapping].

We consider the reaction of I with a bulk defect T, which traps the interstitial through the reaction

$$I+T\rightleftharpoons IT \ . \tag{12.48}$$

The trap can be any defect that has an affinity for I. Since the reaction is supposed to be reversible, we have at equilibrium that

$$\frac{k_{I,T}}{k'_{I,T}}=\frac{C_I^* C_T^*}{C_{IT}^*} \ . \tag{12.49}$$

This leads to the coupled set of equations

$$\frac{\partial C_I}{\partial t}=d_I\frac{\partial^2 C_I}{\partial x^2}-k_{I,T}C_T\left[C_I-C_I^*\frac{C_{IT}/C_T}{(C_{IT}/C_T)^*}\right] , \tag{12.50}$$

$$\frac{\partial C_{IT}}{\partial t}=k_{I,T}C_T\left[C_I-C_I^*\frac{C_{IT}/C_T}{(C_{IT}/C_T)^*}\right] ,$$

where for the simplicity we have assumed both T and IT to be immobile. It is also assumed that

$$C_T^{\text{total}}=C_{IT}+C_T \tag{12.51}$$

is fixed at a uniform value throughout the bulk. As in Eq. (12.13), $k_{I,T}$ is given by

$$k_{I,T}=\frac{4\pi a_{I,T}}{\Omega C_S}d_I \ , \tag{12.52}$$

and in analogy to Eq. (8.1),

$$\frac{k_{I,T}}{k'_{I,T}}=\frac{C_I^* C_T^*}{C_{IT}^*}=\frac{C_S}{\theta_{IT}}\exp(-E_{IT}^b/kT) \ , \tag{12.53}$$

where E_{IT}^b is the binding energy of the IT defect and θ_{IT} is a geometrical factor of order unity.

For a wafer initially at equilibrium with a uniform concentration of T, the effect of the traps is essentially to slow the progression of excess I through the wafer, while the value of ΔC_I at the injecting surface still approaches the steady-state value of g_I/σ_I found for a substrate without any traps. The critical variables are the ratio of C_{IT}/C_I at equilibrium and the value of g_I/σ_I compared to C_T^{total}. The behavior of I injection with trapping can be characterized as follows.

If $C_{IT}^* \ll C_I^*$, then the I flow independently of T. Note that under this condition, the presence of T will have no effect on the flow of I, regardless of the value of C_T^{total}.

If $C_{IT}^* \gtrsim C_I^*$, then there are two general situations. (a) If

$$\frac{g_I}{\sigma_I}\frac{C_{IT}^*}{C_I^*}\ll(C_T^{\text{total}}-C_{IT}^*) \ , \tag{12.54}$$

then the I flow through the bulk with an effective diffusivity[7]

$$d_I^{\text{eff}}=\frac{d_I C_I^*}{C_I^*+C_{IT}^*} \ . \tag{12.55}$$

Physically, this means that local equilibrium is obtained between the formation and dissociation reactions of Eq. (12.48), since this requires I diffusion only over atomic distances. Over much larger distances, the change in diffusivity reflects the fact that an I point defect executing a random walk through the lattice spends a fraction $C_I^*/(C_I^*+C_{IT}^*)$ of its time as a freely moving defect and $C_{IT}^*/(C_I^*+C_{IT}^*)$ of its time attached to a T defect. In this scenario, $C_{IT}/C_{IT}^* \approx C_I/C_I^*$ and $C_T \approx C_T^*$. Note also that the effective diffusivity as written above works self-consistently towards case (a) where $C_{IT}^* \ll C_I^*$ and T has little effect on I. The restriction of Eq. (12.54) simply means that the excess I generated at the surface cannot be so great as to completely fill the traps; this is a potential possibility, since the total number of defects C_T^{total} is fixed at a finite value, while no restriction on possible I injection levels has been made. This behavior is shown in Figs. 28(a) and 28(b) for $C_{IT}^* < C_I^*$ and $C_{IT}^* > C_I^*$. The restriction of Eq. (12.54) is satisfied in the case of Fig. 28(a) and is borderline in the case of Fig. 28(b). It should be mentioned that for the low equilibrium concentrations of point defects indicated by experiments, $C_X^*/C_S < 10^{-9}$, it is most likely the case that $C_{IT}^* \ll C_T^{\text{total}}$ [see Eq. (8.1)] and except for very large surface generation, as outlined below, Eq. (12.55) should hold.

(b) If $C_{IT}^* \gtrsim C_I^*$ and

$$\frac{g_I}{\sigma_I}\frac{C_{IT}^*}{C_I^*}>(C_T^{\text{total}}-C_{IT}^*) \ , \tag{12.56}$$

then an appreciable fraction of the traps become filled and Eqs. (12.50) must be solved exactly. As the traps become full, the concentration of unoccupied traps, C_T, tends towards zero, making it difficult for I to find available T. This situation is somewhat analogous to the situ-

[7]Thus the analytic solutions for the semi-infinite bulk provided by Hu (1985c) and the wafer of arbitrary thickness included in Appendix C can be used with an effective diffusivity of d_I^{eff}, and σ_I, σ'_I, and g_I all multiplied by d_I^{eff}/d_I.

ation already considered in Fig. 26, where it is demonstrated that the retarding effect of I-V recombination is reduced significantly by the annihilation of the recombination centers—in that case, V. In a similar manner, the retarding effect of the traps is lessened as the traps become filled.

Figures 28(c) and 28(d) show two representative examples of I injection that are in the regime of Eq. (12.56). Since the number of T defects is fixed at C_T^{total}, C_{IT} cannot exceed this value no matter how high the injection level. Thus, even though C_{IT}^* is greater than C_I^*, there is only a limited effect of the traps on the flow of I. This is a clear indication of the importance of including the reverse reaction to IT formation in the formulation. Figure 28(c) shows that the progression of ΔC_I through the wafer is still retarded because of the high concentration of T compared to I, but this retarded diffusion is not nearly so great as the 21-fold reduction predicted by using Eq. (12.55). Figure 28(d) shows a situation in which the surface generates excess I to a concentration above the total trap concentration. As expected, the presence of the traps has much less effect on the flow of I than in Fig. 28(c).

For a given value of C_T^{total}, the ratio C_{IT}^*/C_I^* can be estimated using Eq. (12.53). Figure 5 can be used to estimate the concentration of C_T^{total} at which C_{IT}^* will equal C_I^*, by replacing C_A with $C_T^{\text{total}} - C_{IT}$. In fact, this raises the point that dopant atoms must behave to some extent

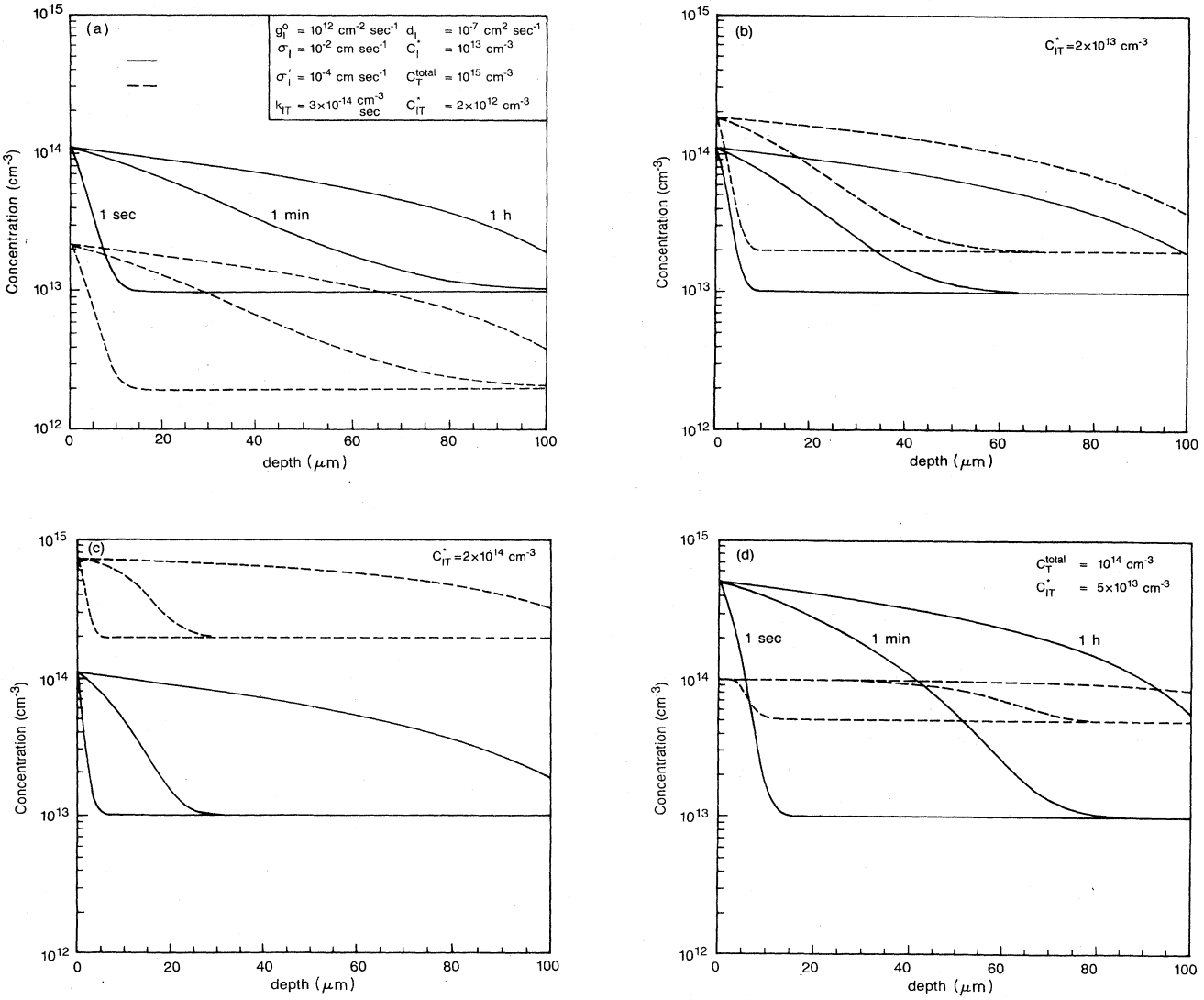

FIG. 28. Behavior of C_I in a trapping model: (a) C_{IT}^* is less than C_I^* and thus has little effect on C_I; (b) C_{IT}^* is greater than C_I^* and causes the retarded diffusion of I predicted by Eq. (12.55); (c) I are retarded less than predicted by Eq. (12.55) because the traps are being completely filled; (d) the traps have increasingly less effect on the flow of I because the traps are quickly filled by a strong surface generation flux.

like traps for excess point defects (although in this case we would now have mobile traps). For example, B atoms may act as traps for excess I by forming BI defects, or As atoms act as traps for excess V by forming AsV pairs. For the case of dopants, we expect that C^*_{AX} will always be so small compared to C_A^{total} that even under injection conditions $C_{AX} \ll C_A^{\text{total}}$. Assuming that $d_{AX} \ll d_X$, the behavior of C_X and C_{AX} will approach that shown in Figs. 28(a) and 28(b). This is discussed further in Sec. XIII.A.

2. Defect-assisted recombination

As an alternative to the direct recombination of I and V as in Eq. (12.2), it is also possible for I to recombine through reactions such as

$$I + DV \rightleftharpoons D \ , \tag{12.57}$$

where D is some defect that has an affinity for V, and we can picture the recombination process using a figure similar to that of Fig. 21. Possible defects could be oxygen-vacancy complexes or dopant-vacancy pairs. How likely is such a reaction to be dominant over the direct recombination of I with V? The full problem entails solving a system of coupled partial differential equations involving C_I, C_V, C_{DV}, and C_D, and will not be dealt with here. Still, some idea of the possible importance of reactions such as Eq. (12.57) can be obtained by making a few reasonable approximations. First, let us assume we have the case in which $C^*_{DV}/C_D^{\text{total}} \ll 1$. As mentioned previously for the case of IT defects, this is likely to be so for such DV pairs because of the low equilibrium concentrations of V. Then during injection of I it will certainly be true that $C_D \approx C^*_D$. Let us further assume that the equilibrium ratio

$$C^*_V/C^*_{DV} = C_V/C_{DV} \tag{12.58}$$

is maintained during I injection. In this picture, V concentrations quickly apportion themselves between isolated V, and V associated with D.

In this case, we can write for the interstitial part of the system of coupled equations

$$\frac{\partial C_I}{\partial t} = d_I \frac{\partial^2 C_I}{\partial x^2} - [k_{I,V} + k_{I,DV}(C_{DV}/C_V)^*](C_I C_V - C^*_I C^*_V) \ . \tag{12.59}$$

The comparative importance of I-DV recombination and I-V recombination is determined by the time constant

$$\frac{1}{\tau_I} = \frac{1}{\tau_{I,V}} + \frac{1}{\tau_{I,DV}} \ , \tag{12.60}$$

where

$$\tau_{I,V} = 1/k_{I,V}C^*_V, \quad \tau_{I,DV} = 1/k_{I,DV}C^*_{DV} \ . \tag{12.61}$$

Recombination of I with DV will be a more important factor then direct recombination with V if

$$\frac{k_{I,DV}C^*_{DV}}{k_{I,V}C^*_V} \gtrsim 1 \tag{12.62}$$

or

$$\frac{a_{I,DV}}{a_{I,V}} \frac{d_I}{d_I + d_V} \frac{\exp(-\Delta E_{I,DV}/kT)}{\exp(-\Delta E_{I,V}/kT)} \frac{C^*_{DV}}{C^*_V} > 1 \ . \tag{12.63}$$

By substituting D for A and V for X in Fig. 5, we can estimate the concentration of D at which $C^*_{DV} = C^*_V$ with different assumed binding energies of DV. Equation (12.63) shows that, even for concentrations of C^*_{DV} lower than C^*_V, the recombination of I with DV can be more important than recombination of I with V.

The coupled behavior of I and V for the case described above will be the same as a case without DV centers, but with a new k_I value of $k_{I,DV}(C_{DV}/C_V)^*$ and a reduced vacancy diffusivity of $d_V^{\text{eff}} = d_V C^*_V/(C^*_V + C^*_{DV})$, and will thus tend to have the behavior demonstrated in Fig. 27(b).

XIII. NONEQUILIBRIUM FORMULATION FOR DOPANTS

A. Intrinsic doping conditions

Consistent with the fact that intrinsic doping conditions require dilute concentrations of dopant atoms ($C_A/C_S < 10^{-3}$ for $T > 800\,^\circ$C), it is assumed that a labeled dopant atom diffuses independently of the other dopant atoms. For such a system of noninteracting particles, diffusing by either an I-type or a V-type mechanism, we can write that

$$D_A = D_{AV} + D_{AI} = d_{AV}\frac{C_{AV}}{C_A} + d_{AI}\frac{C_{AI}}{C_A} \ . \tag{13.1}$$

This gives the instantaneous value of D_A. In an actual diffusion study, the quantity that is measured experimentally is the time-averaged value of the diffusivity, $\langle D_A \rangle$. To relate the time-averaged diffusivity that results under nonequilibrium conditions to the diffusivity found under equilibrium conditions, it is convenient to introduce as a parameter the fractional interstitial component of diffusion under equilibrium conditions,

$$f_{AI} \equiv \frac{D^*_{AI}}{D^*_{AI} + D^*_{AV}} \ . \tag{13.2}$$

The diffusion could proceed equally well by using the fractional vacancy component of diffusion [$f_{AV} = (1 - f_{AI})$] as a parameter. Using the above definition, it follows from Eq. (13.1) that the enhancement or retardation in diffusivity that is observed experimentally is

$$\frac{\langle D_A \rangle}{D^*_A} = (1 - f_{AI})\frac{\langle C_{AV} \rangle}{C^*_{AV}} + f_{AI}\frac{\langle C_{AI} \rangle}{C^*_{AI}} \ . \tag{13.3}$$

The only approximations used in obtaining this relationship are that C_A is dilute (so we may use Fick's law) and that

$$C_A^{\text{total}} = C_A + C_{AI} + C_{AV} \approx C_A \ . \tag{13.4}$$

This is true under equilibrium conditions and will also be the case under nonequilibrium conditions unless the supersaturation of I or V is quite high. But if Eq. (13.4) is true, from the treatments in Secs. XII.E.1 and XII.E.2 we expect that

$$\frac{\langle C_{AX} \rangle}{C_{AX}^*} = \frac{\langle C_X \rangle}{C_X^*} \ . \tag{13.5}$$

Intuitively, one can picture this requirement's being met by considering diffusion under equilibrium conditions of a dopant concentration profile that is initially Gaussian. When the silicon crystal is heated, point defects are thermally generated at the surface and in the bulk and rapidly apportion themselves between isolated X and associated AX defects. Consequently, even for short times after heating the wafer, the dopant profile begins to broaden, as predicted by a constant diffusivity. If we generate excess defects at the surface, this same apportionment process occurs and we again see the profile broaden with a diffusivity independent of concentration, although at a different rate because of the change in point-defect concentrations. We can then rewrite Eq. (13.3) as

$$\frac{\langle D_A \rangle}{D_A^*} = (1 - f_{AI}) \frac{\langle C_V \rangle}{C_V^*} + f_{AI} \frac{\langle C_I \rangle}{C_I^*} \ . \tag{13.6}$$

Figure 29 shows the behavior of D_A as a function of I supersaturation or undersaturation using Eq. (13.6), for different assumed values of f_{AI}. The amount of enhanced or retarded diffusion is determined by the competing effects of raising the I component of diffusion while lowering the V component, or vice versa. Limiting cases are shown for strong and weak coupling of I and V, as discussed in Sec. XII.A. In discussing the meaning of Fig. 29, it is best to keep in mind that it is the value of $\Delta \bar{D}_A$ which is determined experimentally and the value of $\Delta \bar{C}_I$ that is deduced from assuming different f_{AI} values and the amount of coupling between I and V.

We focus first on the behavior of D_A in the presence of excess I, i.e., $\Delta \bar{C}_I > 0$ in Fig. 29. The figure shows that retarded diffusion during I injection can only occur for $f_{AI} < 0.5$. Even for this case, enhanced diffusion can be observed for a dopant with small f_{AI} if the I supersaturation is high enough. If f_{AI} approaches 1, then the enhanced diffusivity of the dopant directly reflects the level of I supersaturation, so that

$$\Delta \bar{D}_A \approx f_{AI} \Delta \bar{C}_I \tag{13.7}$$

regardless of any accompanying V undersaturation. Of course, since in this case f_{AI} approaches 1, measurement of the diffusivity enhancement D_A / D_A^* gives an excellent approximation for C_I / C_I^*.

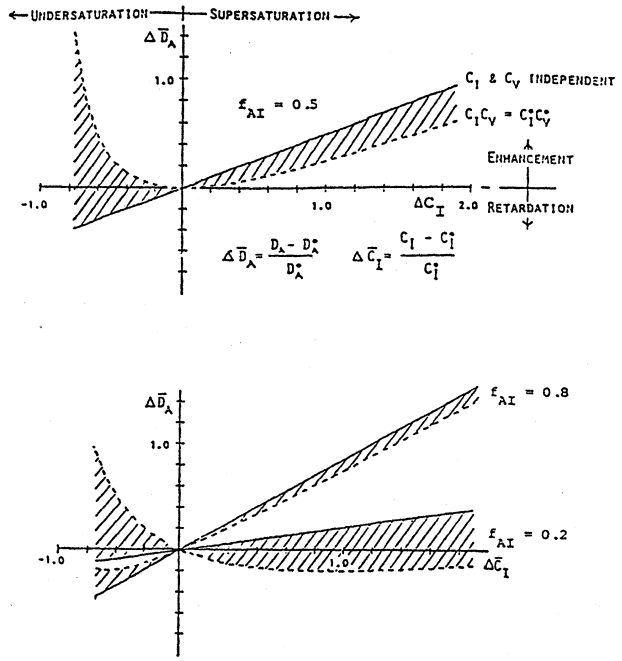

FIG. 29. Change in dopant diffusivity with change in I excess, as predicted by Eq. (13.6). Solid lines show the behavior ignoring the coupling between I and V; dashed lines show the behavior assuming the strongest possible coupling between I and V.

The behavior of D_A as a function of I undersaturation, i.e., $\Delta \bar{C}_I < 0$ in Fig. 29, may at first seem a little strange, but this is only because this type of plot is not symmetrical about the vertical axis. The solid lines simply show the degree of retarded diffusion, as predicted by Eq. (13.6), that results for an assumed I deficit. But how did this deficit come about? If it is a result of vacancy injection, then one must also take into account the enhancement to the vacancy component of D_A. The dashed line of Fig. 29 thus shows the *minimum* value of $\Delta \bar{D}_A$ that will result during vacancy injection for an assumed value of $\Delta \bar{C}_I < 0$. It can be seen from the figure that only a dopant with a value of f_{AI} close to 1 will exhibit large reductions in diffusivity during V supersaturation.

To model the behavior of D_A quantitatively as a function of point-defect concentration changes in Fig. 29, we must calculate $\langle C_I \rangle / C_I^*$ and $\langle C_V \rangle / C_V^*$ in order to utilize Eq. (13.6). The easiest situation would be one in which these point-defect concentrations could be calculated independently of the presence of the dopant atoms, and then their values simply inserted into Eq. (13.6). The treatment in Secs. XII.E.1 and XII.E.2 demonstrated, however, that we may not always be free to make this assumption, since dopants might act as either traps or recombination centers for excess point defects. However, for all practical cases, dopant layers are very thin compared to the length over which point defects diffuse. Therefore it is unlikely that direct coupling between

point defects and dopant atoms will need to be included as in the treatments in Secs. XII.E.1 and XII.E.2. This is clear in the case of trapping, since its major manifestation would be to slow the diffusivity of excess point defects through the doped layer, and a thin layer would be comparable to the diffusion length of point defects only for very short times, even with a reduced point-defect diffusivity. In addition, during this short time there would be a concentration gradient of excess point-defect concentrations across the layer, leading to a nonhomogeneous broadening of the dopant profile, which is never observed experimentally. For the case of defect-assisted recombination such as $I + AV \rightarrow A$, we would imagine this reaction to have little effect on the overall system because of the small number of additional recombination centers that result from a thin dopant layer. Unless experiments prove otherwise, it appears reasonable to use values of $\langle C_I \rangle / C_I^*$ and $\langle C_V \rangle / C_V^*$ calculated without the dopants present in Eq. (13.6) in order to predict D_A under nonequilibrium conditions. Still, in Sec. XVII.A.4 experimental results are presented that are not easily modeled by Eq. (13.6), even though the conditions of the experiment are believed to satisfy the assumptions underlying this expression.

B. Extrinsic doping conditions

We now arrive at an area that has received little experimental attention and not much in the way of theoretical analysis: the nonequilibrium behavior of point defects under extrinsic doping conditions. This situation arises quite frequently during device fabrication, since there are many common steps that require oxidation of the silicon surface when extrinsically doped layers are present in the substrate.

Perhaps the most ambitious attempt at treating this problem is the work of Mathiot and Pfister (1984). These authors essentially write down a series of point-defect reaction equations and then solve continuity equations. There are many questionable assumptions made in this work; some are inconsistent on theoretical grounds (e.g., that percolation phenomena could be described by a Fick's law formulation of diffusion; Sec. VIII.A.3) or directly at odds with experiment (e.g., that B forms B-B_i pairs at high doping levels rather than precipitating; Sec. VIII.A.1). But aside from questions on particular model assumptions, it can be appreciated that the treatment of Mathiot and Pfister is a straightforward approach to dealing with complex phenomena by including all of the complexities in the model system. Simplified test cases can then be used to examine which processes are dominant in a given situation. The major drawback of this type of approach is that, because there are so many adjustable parameters, it is not realistic to believe that good agreement between simulations and a limited set of experimental data really validates the assumptions that go into the model equations. At present, there simply are not enough experimental data available to analyze extrinsic diffusion effects in this way. More simplified treatments have been presented by Morehead and Lever (1986), Mulvaney and Richardson (1987), and Orlowski (1988). The analyses are similar in nature and essentially solve a subset of the equations presented by Mathiot and Pfister (although the nomenclatures in the papers are different enough that this is not obvious). These are all promising attempts at trying to determine nonequilibrium diffusion profiles under extrinsic doping conditions, but more work needs to be done. For example, all of the above papers skirt the issue of boundary conditions by imposing *a priori* either equilibrium or some specified time dependence on point-defect behavior at the surface. In the absence of a definitive formulation for the problem, we examine qualitatively what general effects extrinsic doping introduces into equilibrium point-defect behavior.

At first glance, one would expect extrinsic doping of the silicon to have a pronounced effect on point-defect behavior. Consider the hypothetical example shown in Fig. 30 of a heavily doped arsenic layer whose peak con-

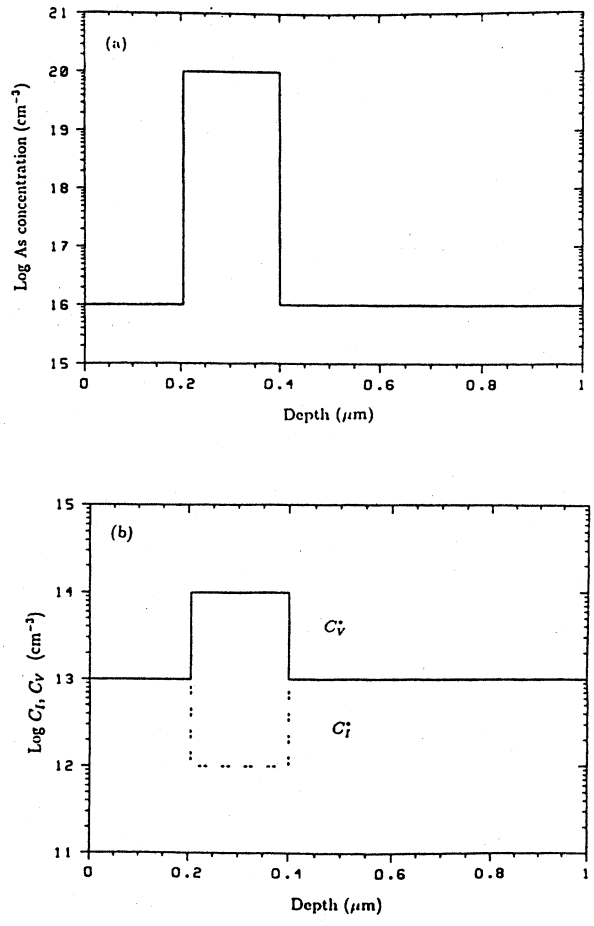

FIG. 30. Hypothetical test structure for nonequilibrium point-defect behavior under extrinsic doping conditions: (a) a thin layer of heavily doped As; (b) the corresponding equilibrium point-defect concentrations.

centration is greater than n_i. Concerning the changes in surface conditions, it is well known that the oxidation rate of such a layer is much faster than that of a layer whose surface concentration is lower than n_i. This may imply that the generation rate of interstitials is correspondingly enhanced. But if the surface condition is changed enough to alter the oxidation rate so drastically, perhaps the surface generation and loss fluxes of point defects are also affected. In the bulk, we have four obvious factors to consider.

(1) We expect the number of vacancies, both free and associated, to increase significantly due to the Fermi-level effects (Secs. VI and VIII). Thus we could find I-V recombination to be considerably more important than in the lightly doped bulk.

(2) The equilibrium concentration of interstitials will also change with doping. Figure 30 shows the case in which it is assumed that I exists primarily in the neutral and positively charged states, so that C_I^* is reduced in heavy n-type doping. If the interstitial excess, $\Delta C_I = C_I - C_I^*$, is still determined by the balance of g_I/σ_I, then this implies a different I excess under extrinsic doping conditions from that of the intrinsic doping case for the same values of g_I and σ_I.

(3) Just as there are electric field terms affecting dopant-dopant interactions (Fig. 12 and Appendix A), there should also be such terms affecting charged point defects traversing the heavily doped layer.

(4) Since the relative values of C_I^* and C_V^* will change with doping, this implies through Eqs. (13.1) and (13.2) that f_{AI} will change with doping.

The problem of point-defect fluxes across the Si/SiO_2 interface is difficult to handle because of our lack of understanding of the surface boundary conditions, as discussed in Sec. XII.B. Even in the case of boron, where there is no large enhancement in oxidation rate when $C_B > n_i$, the preferential segregation of boron into the growing oxide still brings up the question of whether the processes responsible for injecting and absorbing I at the Si/SiO_2 interface are affected by heavy boron doping. We should like briefly to address points (1)–(4) by accepting for the moment that we have eliminated the boundary condition problem by creating a heavily doped layer beneath the silicon surface (by ion implantation and epitaxial growth of a lightly doped layer, for example).

First, while point (1) is true, namely, that there could be increased I-V recombination in the As layer, in typical applications the thickness of the layer would be rather small, on the order of half a micron or less. The extent of defect recombination in such a thin layer is likely to be small, although if there were a large number of efficient recombination centers in the layer, the overall flow of point defects could be affected. The effect of excess I on V within and outside the layer is less obvious. Points (2) and (3) concerning the change in equilibrium point-defect concentrations and electric field effects are intimately related, as the discussion in Appendix A makes clear. One may be tempted to use the results obtained already in Appendix A, substituting As^+ for the high-concentration donor dopant and I^+ for the supposed lower-concentration donor dopant. However, the fact that equilibrium point-defect concentrations are a function of doping is not handled for the case of dopant-dopant diffusion interactions in the same way it should be handled for the case of charged I or V point defects injected through a heavily doped layer. For example, if we simulated the case of a constant-source deposition of Sb^+ (with $C_{Sb} \ll n_i$ at the surface) through a heavily doped As layer, this would not simulate the case of I^+ diffusing through the As layer. The equivalent case for dopants would be to consider the As concentration profile as being constructed by a sequence of materials with electronic properties identical to silicon, but with each layer having a different Sb segregation coefficient in order to mimic the effect of the dependence of C_I^* on electron concentration. In addition, such a simulation would need to include the effect of excess I on the As diffusion—not only the changes in C_{AsI}/C_{AsI}^* and C_{AsV}/C_{AsV}^*, but also, as Eq. (13.3) indicates, the doping dependence of f_{AsI}, which is point (4).

Obviously, extrinsic doping complicates the situation considerably over the intrinsic case. Because of the lack of experiments, it is impossible at this time to verify any proposed formulation of nonequilibrium diffusion under extrinsic doping conditions. Still, it is useful to consider what behavior one should expect in the simplest cases and move progressively to more complicated conditions. Experiments can then identify the conditions under which simple assumptions are inadequate.

The simplest case to consider is a dopant species that diffuses by only one mechanism, I-type or V-type, under both intrinsic and extrinsic doping conditions. Whether this is the case or not, when a wafer is heated in an inert ambient, i.e., no chemical reactions at the surface, point-defect concentrations inside and outside the heavily doped layer quickly reach their equilibrium values and diffusion proceeds as discussed in Appendix A. This is consistent with common experience. In the type of potential well illustrated in Fig. 30, this means that C_V^* is some factor higher in the As-doped layer in a short time. If we now inject vacancies from the surface, we expect that this same ratio is again quickly achieved, i.e., C_V/C_V^* is the same inside and outside the heavy doped layer. Note that this requires that the vacancy excess, $\Delta C_V = (C_V - C_V^*)$, *not* be the same inside and outside the As layer. If As diffused completely by a vacancy mechanism, we would have a very simple answer for the effect of V injection on As diffusion: the As diffusivity would be enhanced by the same C_V/C_V^* factor as calculated for a lightly doped substrate. Similarly, if B diffused completely by an I-type mechanism, we would expect B diffusion to be enhanced by the same factor C_I/C_I^* as calculated for a lightly doped substrate.

The next simplest case to consider is one in which f_{AI} is neither zero nor one, but V injection does not affect I concentrations, and I injection does not affect V concen-

trations. Then we would expect Eq. (13.6) to hold during I injection, and a similar expression to hold for V injection, but f_{AI} to depend on the doping concentration. What form this dependence might take is discussed presently in Sec. XVII.D. Qualitatively, what would result is the same supersaturation of point defects as for a lightly doped substrate, but with an enhancement factor that changes with dopant concentration. So, for example, it is believed that As has a non-negligible I component of diffusion for $C_{As} < n_i$ and diffuses almost completely by a vacancy mechanism for C_{As} much above n_i, due to the increase in negatively charged vacancy concentrations. We would expect I injection to have little effect on the diffusion rate of that part of the profile above n_i and to have some measurable effect on that part of the profile below n_i.

The most complicated situation, of course, is one in which the injected point defects enhance one component of diffusion while reducing the other, and f_{AI} depends on doping. It is difficult to see how these effects could be separated out from the small body of experimental data that are available, especially considering that it is an open question what changes in boundary conditions might result in going from lightly doped to heavily doped surfaces. Even if we were able to separate out extrinsic effects by some carefully controlled experiments with heavily doped buried marker layers, we might find it necessary to adjust the surface conditions empirically anyhow in order to simulate measured results. This situation is perhaps one of the most problematic in the further development of models of point defects and dopant diffusion, but certainly some exploratory experiments in this area are warranted.

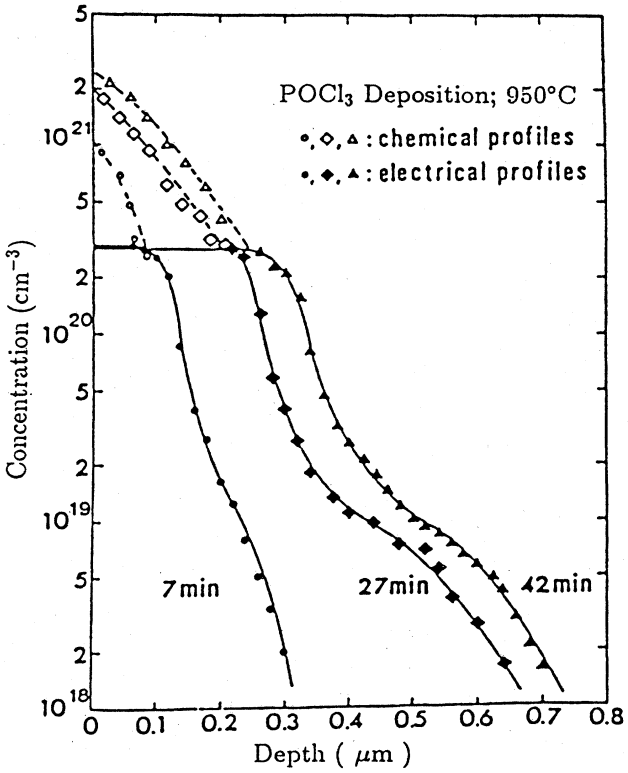

FIG. 31. Anomalous profile shapes for high-concentration P diffusion. From the work of Masetti, Nobili, and Solmi (1977).

C. The anomalous case of P diffusion

1. Experimental evidence

Although P diffuses in a well-behaved manner when the peak P concentration is below n_i at the diffusion temperature (for example, an initially Gaussian-shaped profile will maintain its Gaussian shape when broadening due to diffusion), when the P concentration is greater than n_i, the resulting diffusion profile and associated phenomena have most often been described by the adjective "anomalous." An example of P concentration profiles from surface depositions utilizing a $POCl_3$ source is shown in Fig. 31. The shape of the profile is one of the most salient features of high-concentration P diffusion and is usually described as exhibiting a "kink" and "tail" for the lower-concentration part of the profile. The experimental data also show that a large fraction of the P is electrically inactive and that the carrier profile is flat in the high-concentration region and reaches a saturated value independent of the chemical concentration. These same features also exist for ion-implanted profiles (Nobili, Armigliato, Finetti, and Solmi, 1982). In addition, it is well established that high-concentration P diffusion changes the diffusion rate of other dopants that are physically removed from the area where P diffuses. Figure 32 shows the effects of P diffusions like those shown in Fig.

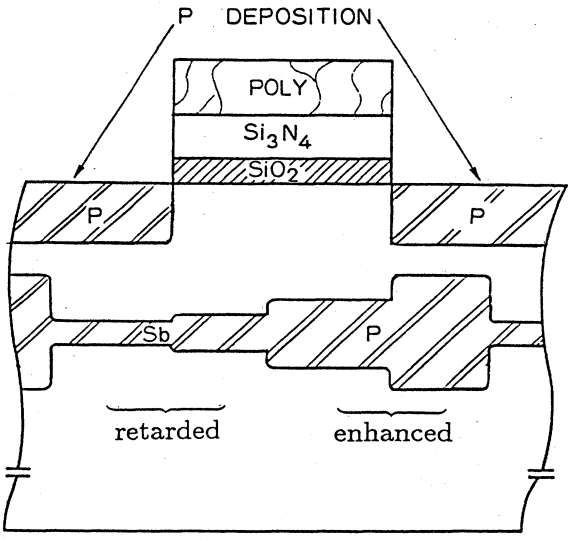

FIG. 32. Effect of high-concentration P diffusion on lightly doped buried marker layers. After Fahey, Dutton, and Hu (1984).

31 on the diffusion of buried, lightly doped marker layers of P and Sb (Fahey, Dutton, and Hu, 1984). Fahey *et al.* found that the diffusion of the P buried layer was enhanced substantially, while that of Sb was retarded. It is very well known that B layers exhibit enhanced diffusion. This became immediately apparent when bipolar transistors with P emitters and B bases were fabricated in the early 1960's. The large enhancement in the movement of the B base underneath the P emitter was termed the "emitter-dip" or "base-push" effect. Harris and Antoniadis (1983) also demonstrated that As buried marker layers are enhanced, but not nearly as much as B buried layers under identical conditions.

2. Models for P diffusion

There have been many models proposed to account for the anomalous features of P diffusion. Almost all of the early models were based on the assumption that P diffusion was dominated by a vacancy mechanism, an assumption that is now thought to be incorrect. The best known of these models is due to Fair and Tsai (1979). Their work produced a quantitative model for high-concentration phosphorus profiles that was, and still is, widely used in simulating phosphorus diffusion. The model is based on phosphorus diffusion with V^0, V^-, and $V^=$ defects and describes the kink and tail regions as the result of PV^- pair dissociation that results in a steady-state excess vacancy concentration. By fitting this model to experimental results, researchers were able to obtain quantitative parameters. Even today, this model remains the only full quantitative model for extrinsic phosphorus diffusion.

The known experimental work (next paragraph) suggests that interstitials rather than vacancies dominate phosphorus diffusion mechanisms. The use of equilibrium mass-action relationships to determine the defect and pair concentrations in the Fair-Tsai model is questionable, so that substituting interstitial charge states for vacancy charge states do not validate the model. A review of the experimental and theoretical arguments against this model was presented by Hu, Fahey, and Dutton (1983). In this same paper, these investigators attempted to develop a P model, shown in Fig. 33, based on an interstitial-substitutional mechanism of diffusion and proposed a set of coupled, simultaneous differential equations to describe P diffusing by a two-stream mechanism.

As already mentioned in Sec. VIII.A.1, after the work of the researchers at Istituto Lamel in Bologna [see the review article by Nobili (1983)], it seems reasonable to assume that the electrically inactive P is due to precipitation and that the saturated level of the carrier profile corresponds to the solid solubility of P at the diffusion temperature (see Fig. 5). The kink and tail shapes of the profile is immediately suggestive of a two-stream diffusion mechanism, in which P diffuses in two different states and the superposition of the two fluxes results in the measured profile. The identification of these two

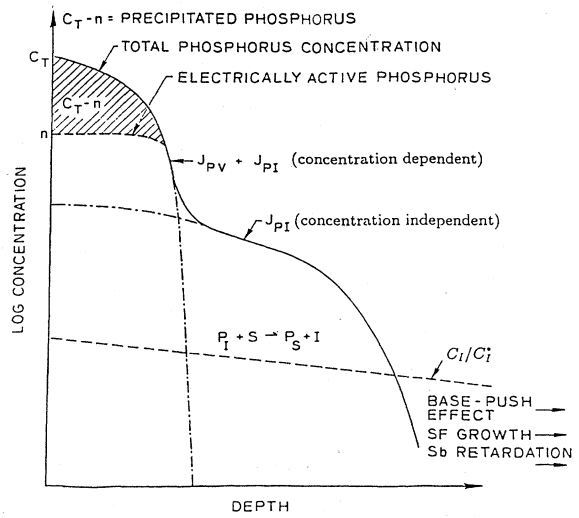

FIG. 33. Qualitative model of P diffusion consistent with experiments. After Hu, Fahey, and Dutton (1983). The identification of the point defects generated by P diffusion as *I*-type defects, and the conclusion that electrically inactive P is due to precipitation of P into a second phase, are now well established. The two-stream mechanism to account for the profile shape is still a hypothesis.

states is intimately related to the phenomena of Fig. 32. Most workers now agree that the enhanced or retarded diffusion of buried layers is due to the generation of excess point defects, caused somehow by the high-concentration P diffusion. The fact that large diffusion enhancements of P and B buried marker layers, milder enhancements for As, and retarded diffusion for Sb are observed indicates that excess I are generated by the high-concentration P (see Table II in Sec. XIV.F). Further supporting evidence is the fact that extrinsic stacking faults grow beneath the P diffusion front (Claeys, Declerck, and Van Overstraeten, 1978; Nishi and Antoniadis, 1986) and that screw dislocations climb in a direction consistent with interstitial injection during P diffusion (Strunk, Gösele, and Kolbesen, 1979).

But what, then, is the source of the I generation? Hu, Fahey, and Dutton (1983) considered four possible sources (the following processes are also discussed in a more general context in Sec. IV): (i) oxidation during deposition, (ii) in-diffusion of P at the surface ejecting substitutional silicon into interstitial sites as the P occupies substitutional positions, (iii) precipitation of P into a second phase, and (iv) P diffusing in an interstitial-type state PI and generating I by the reaction

$$S + PI \rightleftharpoons P_s + I \ . \tag{13.8}$$

This last process, (iv), is just the "kick-out" mechanism discussed in Sec. XI, where S is a substitutional silicon atom. We have included the subscript on substitutional phosphorus P_s for clarity and make no statement about whether the interstitial-type species are interstitialcies or interstitials. Fahey, Dutton, and Hu (1984) demonstrat-

ed that I supersaturation occurs in the bulk when there is no surface oxidation or in-diffusion and under conditions in which precipitates grow or shrink. Angelucci, Solmi, and Zani (1986) have also demonstrated that P precipitation does not seem to generate excess interstitials. This leaves the above reaction as a likely source of the interstitials, although possibilities (i)–(iii) could be contributory or even dominant sources of excess I depending on the experimental conditions.

While the electrical inactivation of P due to precipitation and identification of self-interstitials as the excess point defects generated during high-concentration P diffusion can be considered well established, proof that a two-stream diffusion mechanism is operative (and what species may constitute the two diffusion states) is not so straightforward. Many reports have shown that the diffusion rate in the tail region is highly correlated with the amount of I supersaturation in the bulk: the smaller the tail diffusion rate compared to equilibrium conditions, the smaller the observed I excess. A clear demonstration of this can be found in the experiment of Finetti, Masetti, Negrini, and Solmi (1980). These investigators demonstrated that diffusing P through a thin polysilicon layer, using the same $POCl_3$ reaction that gives rise to the profiles in Fig. 31, can significantly reduce the prominence of the tail and at the same time reduce the amount of enhanced diffusion of B buried marker layers.[8] Thus there is significant evidence that diffusion in the tail region proceeds by an I-type mechanism. But what of the diffusion mechanism in the high-concentration region? The nitridation experiments in Sec. XIV.F suggest that, even under extrinsic conditions, P diffusion is retarded during vacancy injection. Therefore P diffusion is likely still dominated by an I-type mechanism in the high-concentration region, although its vacancy component may be larger than under intrinsic conditions, as discussed previously for f_{AI} under extrinsic doping conditions (and also in Sec. XVII.D). A two-stream mechanism would then have to allow that P diffuse in two different interstitial charge states. Given the difficulties associated with modeling diffusion of the better behaved dopants, As and B, as outlined previously in Sec. XIII.B, one can certainly appreciate the difficulty of correctly modeling the proposed two-stream diffusion mechanism for P.

The Hu, Fahey, and Dutton (1983) model assumes that the slower component in the high-concentration region is due to PV pairs, while the tail component is attributed to an interstitial-type mechanism. The reaction of Eq. (13.8) is taken into account as well as I-V recombination and precipitation reactions. It is of interest to know the outcome of solving these equations, since all of the phenomena outlined in the experimental section above are taken into account, and the relative importance of vacancy and interstitial mechanisms can be changed in this model by the use of concentration-dependent diffusivities for vacancy and interstitial components. In a limited study, Fahey, Greenfield, and Dutton (1983) solved these equations for the case in which the interstitial diffusivity and interstitial component of P diffusivity does not depend on concentration; they found, as expected, that (a) profiles exhibit a kink and tail, (b) the electrical and chemical profiles look like experimental results, and (c) an I excess is established in the bulk and correlates with the enhanced diffusion in the tail region. *However,* these solutions predict that the concentration where the kink begins, that is, the concentration where the tail region begins, increases with time during simulations of constant-source depositions. Although the location at which the kink begins is of necessity a subjective judgment, it is still clear from multitudes of experiments that no matter what criterion is chosen, the concentration at which the kink occurs does not appear to change with time (Fig. 31 is an example). It may be that the inclusion of concentration-dependent diffusivities of I and the I component of P diffusion (i.e., Fermi-level effects) can lead to a kink concentration that is independent of time. Further development of a quantitative model along these lines was abandoned primarily for practical reasons. From a numerical point of view, solving systems of coupled differential equations with the inclusion of electric field terms is a difficult problem (aside from questions concerning their correct physical formulation) that is hard to implement in a way that would make such a model useful as an engineering tool, which is the primary goal of the model's development. From a more scientific point of view, there are now so many variables—which cannot be checked independently—in this type of model that the fitting of experimental profiles by suitable adjustment of parameters may prove nothing over the plausibility of the model based just on qualitative interpretation of experimental results. In fact, all quantitative models proposed thus far for phosphorus have claimed to be able to fit profile shapes, even though they have subsequently been found to be (i) directly at odds with other experimental results, (ii) incapable of predicting profiles under only slightly different experimental conditions, or (iii) physically untenable.

In assessing the validity of any proposed models for P diffusion based on profile fitting, it is seldom recognized that many of the data used for profile fitting are simply not suitable for model development by the approaches taken thus far. Specifically, it is most often the case that

[8]There is more than one possible explanation for the reduction in I excess by the presence of the polysilicon layer. The grain boundaries may serve as recombination planes for excess I, since it is known that oxidation-enhanced diffusion is reduced when the surface is covered with polysilicon during oxidation (Swaminathan, 1983). If precipitation processes are important in the generation of excess I, the grain boundaries may affect this contribution, since they serve as preferential nucleation sites for P precipitation (Armigliato, Servidori, Solmi, and Zani, 1980).

P concentration profiles resulting from chemical depositions have been the object of model fittings. The models presented thus far would have us believe that prediction of profile shapes after constant-source deposition and their subsequent evolution with further heat treatments can be accomplished by specifying the time and temperature of diffusion and the surface concentration of P. But it is commonly known to process engineers that the most important experimental variables, which will determine such critical parameters as total dose in the bulk after deposition, depth of penetration, etc., are not just the time and temperature of deposition, but gas flow rates and partial pressures of oxygen, the phosphorus-containing gas, and the carrier gas (either nitrogen or argon). These parameters vary, of course, with the doping system used (PH_3-doped oxides, liquid $POCl_3$ source, P_2O_5 solid sources, etc.) but are seldom known directly. In any event, the exact sequence of chemical reactions that takes place at the silicon surface is really not understood at present and is certainly not reflected in any of the parameters of the proposed models. In addition, Fig. 31 shows a feature that is a common occurrence for chemical depositions: the surface concentration does not remain constant with time during the deposition. Thus previous modeling attempts that present the result of fitting a profile for a single time must be viewed with some circumspection. Even less confidence should be placed in models based primarily on analysis of electron concentration profiles, when the existence of the flat section of these profiles clearly indicates the presence of larger amounts of electrically inactive P.

3. Prospects for developing a P diffusion model

Given the anomalous features of high-concentration P diffusion, it is little wonder that As is used almost exclusively in the fabrication of modern integrated circuits. This fact, together with the experimental and theoretical problems already encountered in trying to model this process, naturally raises the question of whether there is a sufficient driving force for any further developments that will better our understanding of P diffusion. In the authors' opinion, one area from which some new insights might be gained in the near future is the investigation of As-P interactions, such as that pictured in Fig. 34(a). This donor-donor interaction is analogous to the donor-acceptor interaction pictured in Fig. 12 of Sec. X.B. The technological driving force for understanding the donor-acceptor interaction was the desire to model the formation of emitter and base regions in bipolar transistors. The technological driving force for understanding donor-donor interactions such as that in Fig. 34(a) lies in the development of special processes to form sources and drains in MOS transistors. The goal of these processes, known as *lightly doped drain,* or LDD, processes, is to create a source/drain profile that reduces the large electric field caused by the sharp concentration gradient of the As profile. Such a profile is pictured in Fig. 34(b).

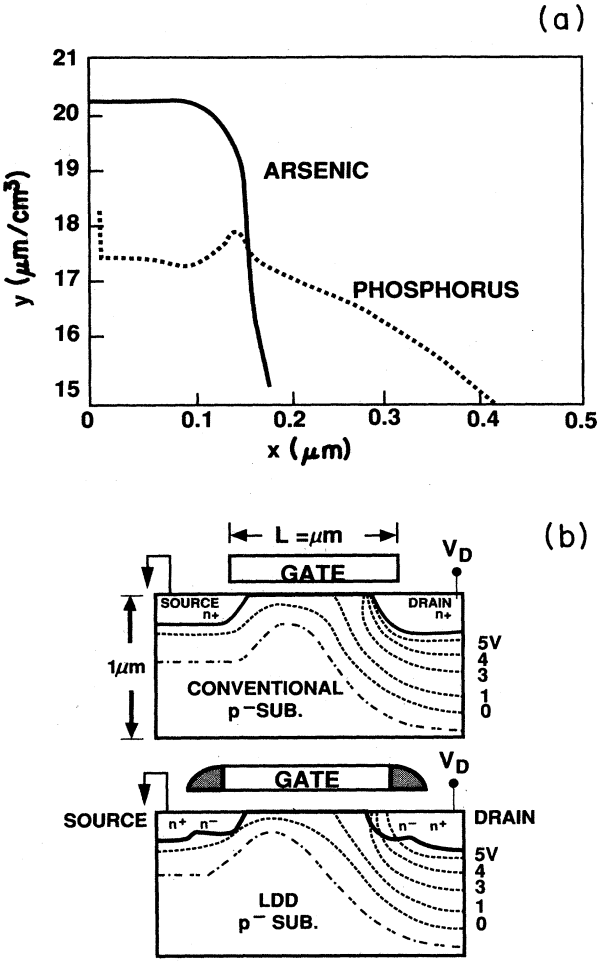

FIG. 34. Donor-donor interactions of As and P. This type of process has a technological application for the formation of lightly doped drain (LDD) structures. (a) Simulated results. (b) Reduction of the electric field by the LDD profile shape [from Bampi (1987)].

This strong electric field is responsible for injecting high-energy electrons into the gate oxide during device operation and is believed responsible for degradation of the oxide, which adversely affects the reliability of the transistor.

The basic idea in forming the LDD structure is that the high electron concentration, provided by the ionized As atoms, will result in enhanced diffusion of the low concentrations of P inside the As-doped layer and will cause the rapid diffusion of P out of the layer to form a final electron concentration profile that is much less steep than that found for As alone. The simulated results of Fig. 34(a) were obtained using the formulation of Appendix A. If the As layer is serving the same role as the extrinsically doped P region in Fig. 31, it should be possible to study the behavior of P diffusion under extrinsic doping conditions in a much more controlled way than by the technique of chemical deposition. It would be interesting to know whether there is enhanced diffusion of

B and P marker layers located below the LDD structure and whether some of these effects might be caused by ion implantation damage. A carefully designed set of experiments should be able to answer these questions.

XIV. EFFECT OF OXIDATION AND NITRIDATION ON *I* AND *V*

Although debate continues concerning the relative importance of vacancy and substitutional/interstitial(cy) interchange mechanisms for the common dopants (Sb, As, P, and B), within the last five years there has definitely been movement towards a consensus view: P and B have the most substantial interstitial components of diffusion, As exhibits both vacancy and interstitial components, and Sb appears dominated by a vacancy mechanism. Historically, evidence for these assignments came from the effects of oxidation on dopant diffusion. During oxidation of the silicon surface above an impurity layer, there is often a large increase in the impurity diffusivity above that normally observed. Figure 35 shows how this oxidation-enhanced diffusion (OED) can be monitored by beveling the silicon and staining the impurity layer after oxidation. Any other technique for determining the impurity profiles, such as spreading resistance, anodic sectioning, and sheet resistance or Hall measurements, CV measurements, or SIMS can also be used to determine the enhancement in dopant diffusivity. Hu (1974) first linked the growth of stacking faults and the OED of dopants by proposing that they had a common origin.

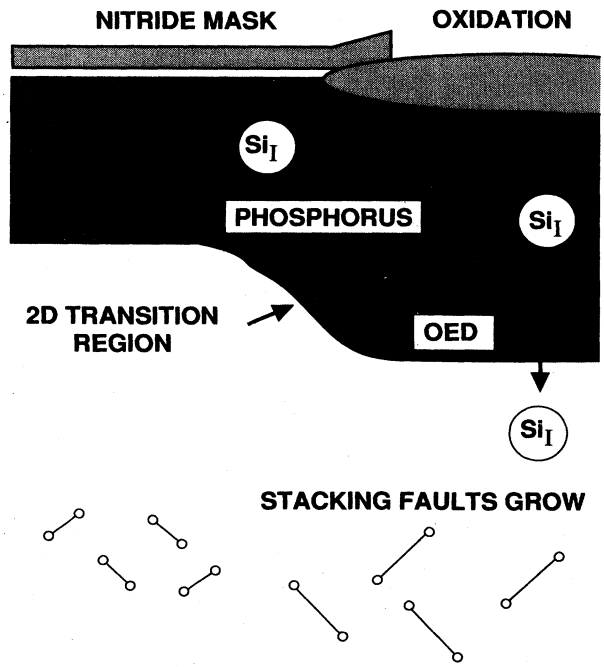

FIG. 35. A structure for monitoring the extent of oxidation-enhanced diffusion. Phosphorus or boron under the oxidizing region have higher diffusivities than the inert diffusion under the masked region.

Direct TEM observation of stacking faults showed that they were interstitial-type defects bounded by Frank partial dislocations. It was known that stacking faults grow in oxidizing ambients by the addition of silicon atoms to the dislocation ends. The common origin can then be attributed to the injection of silicon interstitials during oxidation. The enhancement of boron and phosphorus meant that a major fraction of their diffusion can take place by an interstitial-assisted mechanism. Arsenic diffusion is enhanced to a lesser extent in an oxidizing ambient. When retarded diffusion of antimony was observed under the same conditions in which B and P were enhanced, it strongly suggested that interstitial injection led to a vacancy undersaturation. The degree of this retardation indicated that antimony diffuses predominantly by a vacancy mechanism. Arsenic has both vacancy and interstitial components of diffusion and cannot be used to monitor the interstitial and vacancy concentrations during oxidation unless the kinetics of both these species are known in detail. For this reason, in the following sections, we concentrate almost exclusively on the OED of boron, phosphorus, and antimony in order to gain some insight into the underlying point-defect kinetics.

By using experiments like those outlined above, one can estimate the fraction interstitial(cy) component f_{AI} (see Sec. XVII) of the dopants. The temperature dependence of the OED for different dopants indicates how the fractional diffusion component associated with interstitials or vacancies varies with temperature. Moreover, the time dependence of OED for B, P, and Sb allows one to examine the kinetics of I-V recombination. For these reasons, much effort in many laboratories has focused on the time and temperature dependence of OED.

At first sight, it seems that the results on the time dependence of stacking fault growth are in somewhat better agreement than those for OED (Hu, 1975; Murarka and Quintana, 1977; Claeys, Laes, Declerck, and Van Overstraeten, 1978). However, experiments on stacking fault growth span a short time range before shrinkage of the stacking fault starts to occur. It seems that stacking fault growth and shrinkage are less reliable for determining the time dependence of point-defect injection than are OED experiments. Moreover, attempts to estimate the interstitial supersaturation from stacking fault kinetics are prone to model assumptions [see Nishi and Antoniadis (1986) and Sec. XVII].

Unfortunately, many of the data on the time dependence of OED appear inconsistent, with different investigators reporting different time dependencies for the OED. There has been a tendency in the literature to assume a consensus time dependence, such that the OED decreased as $\sqrt{dx/dt}$, where dx/dt is the oxidation rate. A critical review of these OED experiments will enable us to better evaluate the different models proposed for OED. These experiments and the results of similar nitridation experiments allow some conclusive statements to be made about the microscopic mechanisms of dopant diffusion.

A. Experiments on OED

Experimental data on anomalous diffusion effects during oxidation have been reported since the mid 1960s and continue to the present day. We have found that the best way to classify the data is chronologically, because later investigators, with hindsight, were able to avoid other interrelated anomalous diffusion phenomena associated with, for instance, high-concentration effects. Because the magnitude of the OED does depend on the oxidation rate, almost every investigator has at least given the time dependence of OED in the form of a power law,

$$\text{OED} \propto \left[\frac{dx}{dt}\right]^n , \tag{14.1}$$

where the exponent n is sublinear. The value n is an empirical fitting parameter, but the data do seem to fit this power law over a wide range of times. Rather than comparing both models and experiments together, we shall first look at the experimental data in terms of the power-law model, since this provides a uniform standard of comparison. Later, we shall look at the physics underlying the apparent power-law dependence and compare the different models that have been proposed.

One of the earliest reports of anomalous diffusion effects relating oxidation and diffusion was by Nicholas (1966). He found that wet oxidation caused a rapid increase in the depth of boron and phosphorus diffusions and suggested that the increase was related to the rate of oxidation. Other complicating effects present in the samples were anomalous diffusion due to crystal polishing damage and high-concentration effects. When attempts were being made to fabricate microwave transistors on ⟨100⟩ substrates rather than on the then common ⟨111⟩ substrates, Wills (1969) reported that ⟨100⟩ silicon experienced OED that was not observed in ⟨111⟩ wafers. A simple check of orientation dependence involved carrying out the diffusions in inert ambients; no difference in the inert junction depth was observed for the different orientations, as expected from the symmetry properties of the cubic lattice.

Allen and Anand (1971) reported more extensively on this effect and concluded that explanations based on differences in oxide growth rate or dopant segregation coefficient between the different orientations could not be responsible. They hinted that the defect responsible for the effect originates at the Si/SiO_2 interface and has a higher density in ⟨100⟩ silicon. In a later paper, Allen (1973) concluded that an actively growing oxide layer caused the enhanced diffusion and suggested that differences in surface bond density between the different orientations was the cause.

Masetti, Solmi, and Soncini (1973) investigated the diffusion of phosphorus in ⟨111⟩ Czochralski silicon under oxidizing and inert ambients. The authors used beveling and staining techniques to delineate the junction depths and assumed Gaussian profiles in their analysis. Some effects of concentration-dependent diffusion are present in the data, which span every 50 °C from 1000 °C to 1200 °C. A single time at each temperature was used with ambients of nitrogen, nitrogen +10% O_2, dry O_2, and steam. The conclusion was that the diffusivity increased with the oxidation rate and that the relative magnitude of the OED effect increased at lower temperatures. Attempts to determine an activation energy for the OED were somewhat premature, since the time dependence of the OED was unknown.

One of the first attempts to examine the time dependence of OED was carried out by Masetti *et al.* (1976). The same substrate orientation and experimental conditions as above were used, except that a range of times out to 16 h at 1100 °C was chosen for the work. Interpreting the results presents a problem because the diffusion under the inert ambient was not constant as expected, but instead reduced with time from a higher initial value. Surface concentrations of approximately $10^{20}/cm^3$ were used, which are a factor of 10 higher than the intrinsic concentration at 1100 °C, indicating that some concentration dependence was present for the diffusions. In addition, it is possible that oxygen precipitation in the Czochralski (CZ) wafers floods the silicon with interstitials for times on the order of hours, thereby causing changes in the inert diffusivity, as observed. Such "anomalous" effects in CZ material have been reported by Hu (1980), Mizuo and Higuchi (1982b), and Ahn *et al.* (1987). For this reason, the extraction of a time dependence of OED from this early work is questionable.

Antoniadis, Gonzalez, and Dutton (1978) examined the OED of boron in near-intrinsic conditions, after realizing that the coexistence of high-concentration, oxidation, and segregation effects made interpreting previous experimental results difficult. Spreading resistance and sheet resistance data indicated that the diffusivity enhancement during dry oxidation ranged from a factor of 2 at 1150 °C to a factor of 100 at 850 °C in ⟨100⟩ silicon, with values for ⟨111⟩ silicon being significantly lower. No retarded diffusion was observed at any temperature. Attempts to establish a time dependence based on the model of Hu (1974) for OED were, again, premature since only one time point was available at each temperature.

Another attempt to minimize anomalous diffusion behavior due to surface effects and concentration dependence and to examine only the OED dependence was carried out by Francis and Dobson (1979). An *n*-type lightly doped epilayer was used as the dopant marker layer on *p*-type CZ silicon substrates of different orientations. Spreading resistance measurements indicated that diffusion in ⟨100⟩ wafers was always enhanced in either dry or wet O_2, whereas diffusion in ⟨111⟩ wafers was enhanced in dry O_2 up to 1160 °C and was then retarded. This suggests that for long times (17 h) at high temperatures on ⟨111⟩ silicon, oxidation injects vacancies or absorbs interstitials. Later reports confirmed these qualitative observations.

Taniguchi, Kurosawa, and Kashiwagi (1980) investi-

gated boron and phosphorus OED as a function of time, temperature, and concentration in CZ silicon. We shall defer discussion of the concentration dependence to Sec. XIV.D. The temperature dependence of OED was similar to those already observed, the relative enhancement being larger at lower temperatures. Taniguchi *et al.* inferred an activation energy for OED of 2.55 eV based on measurements between 950 °C and 1150 °C. Since the activation energy for dopant diffusion is 3–4 eV, this means that large supersaturations of interstitials must exist at lower oxidation temperatures. By monitoring the OED as a function of wet oxygen partial pressure in argon, they obtained the OED dependence as a function of oxidation rate, giving a sublinear dependence with an exponent ranging from 0.27 to 0.32. Data over a small range indicated that the diffusion enhancement was the same for dry O_2 at the same growth rate. There were not sufficient data to indicate that the time dependence was the same for wet and dry oxidation.

A thorough investigation of the orientation dependence of OED was published by Hill (1981). Hill found that the diffusion coefficient increased in the order ⟨111⟩, ⟨110⟩, ⟨100⟩ during dry oxidation. In addition, retarded diffusion of boron was observed above 1150 °C in the ⟨111⟩ orientation, confirming the results of Francis and Dobson (1979). Additional confirmation came from the work of Tan and Ginsberg (1983) when they observed enhanced diffusion of antimony under the same conditions that retarded diffusion of phosphorus and boron. Enhanced antimony diffusion under these conditions suggested that there was a vacancy supersaturation at these long oxidation times. This anomalous diffusion was seen only in ⟨111⟩ wafers oxidized for long times (18 h) at 1160 °C.

Hill also determined that the diffusivity enhancement was a constant independent of oxidation rate (an exponent of zero). One may argue with the assumptions in the quantitative analysis of Hill (Gaussian approximations to initial profiles, sensitivity of analysis to peak depth, etc.), but some additional experimental evidence indicated that the time variations in OED were small. Hill monitored the OED in two regions, one of which had an initial oxide 4000 Å thick and the other bare silicon, so that differences in oxidation rate were observed on the same sample. The beveling and staining technique was used to monitor the relative changes in the junction depth and the differential OED of the dopants was observed on regions side by side on the same bevel. The absence of any difference in the final junction depths indicated that the OED was the same for both oxidation rates, at least within experimental error.

Lin, Antoniadis, and Dutton (1981) also used measurements in two regions, one with an initial oxide layer, the other bare silicon initially, to obtain information on the oxidation-rate dependence of OED. Measurements were carried out for a single time at each temperature but no attempt was made to vary the initial oxide layer thickness to obtain an optimum large difference in oxidation rates at the different temperatures. The 1200 °C result, which has the largest difference in oxidation rates (a factor of 5), indicates a sublinear dependence of OED on oxidation rate. Lin *et al.* showed that these data were consistent with previous observations based on stacking fault kinetics during oxidation, but the data cannot provide an independent measure of the oxidation-rate dependence of OED.

Ishikawa *et al.* (1982) made measurements on ⟨100⟩ and ⟨111⟩ CZ wafers for a range of times at several temperatures. Anodic sectioning and sheet resistance measurements were used to determine the profiles, and numerical solution of the diffusion equation accounting for segregation effects was used to estimate the diffusivity enhancement. The orientation dependence observed agrees with the previous data. The oxidation-rate dependence of phosphorus observed was sublinear with an exponent of 0.7–0.8 for differences of ≈ 5 in the oxidation rate.

A range of times corresponding to an oxidation-rate difference of ≈ 3 at 1100 °C was investigated by Mizuo and Higuchi (1982b), but no quantitative analysis of the results was given. Instead, junction depths and ratios of junction depths were plotted as a function of time. Yoshida, Matsumoto, and Ishikawa (1986) interpreted Mizuo and Higuchi's results in a more quantitative fashion by assuming that the diffusivity was related to the square of the junction depth. In this way, the authors determined that the time dependence for the phosphorus OED was $t^{-0.12}$, which corresponds to an oxidation-rate power dependence of 0.24 for oxidation in the parabolic growth regime.

Quantitative CV measurements of boron OED in CZ wafers were made by Miyake and Harada (1982). For differences in oxidation rate of a factor of ≈ 5 at 950 °C and 1000 °C, a sublinear dependence on oxidation rate with an exponent of 0.3 was observed, over a similar time range to that observed by Ishikawa *et al.* The authors measured in inert diffusivity to check that it was constant and corrected for segregation effects in their analysis, making the technique accurate in principle.

The CV technique was also used by Antoniadis and Moskowitz (1982a) to measure short-time enhancements in the OED and oxidation-rate dependence of dopants. The results show a relaxation from a high initial diffusion enhancement to a lower value after 60 min at 1000 °C, and if a power-law dependence is used to fit the data, the resultant value of the exponent is near 1. This finding is very different from any of the other literature results, but it is also the only experiment that monitors the OED in the linear oxide growth regime. At 1000 °C, the transition time from linear to parabolic oxide growth is about 60 min. How the value of n might vary during linear and parabolic oxide growth is addressed in Sec. XIV.E.

An elegant method of maintaining constant oxidation rates for long periods of time was described by Mizuo and Higuchi (1985). This was to vary the partial pressure of oxygen in the ambient with time. Ratios of junction

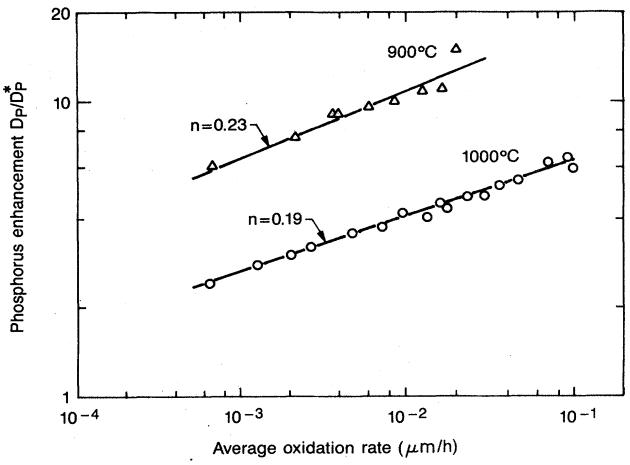

FIG. 36. Experimental data for oxidation-enhanced diffusion. These OED data from Dunham and Plummer (1986b) were obtained by spreading resistance measurements from diffusions in dry O_2 at 900 °C and 1000 °C. The value of n in the figures represents a least-squares fit of the data to the expression $\langle D_A \rangle / D_A^* = A(dx/dt)^n$.

depths in the inert and OED regions plotted against time indicated that the OED was constant if the oxidation rate was kept constant over decade time spans, for relatively thick oxides.

Dunham and Plummer (1986b) measured diffusion enhancements in CZ wafers over a two-decade span in the oxidation rates by using partial pressures of argon and dry oxygen to obtain very low oxide growth rates. Spreading resistance measurements were used to determine the diffusivity in both the inert and the oxidizing regions. The experimental results are shown in Fig. 36. The OED was found to be sublinearly dependent on oxidation rate with exponents of 0.26 and 0.19 at 900 °C and 1000 °C, respectively.

B. A consensus value for *n*

When it comes to obtaining a value for the power-law dependence from the literature values, one must, with hindsight, disregard the early results dealing with the OED time dependence. Complications from many interrelated effects influenced the OED time dependence in these experiments. Moreover, results that show that the OED time dependence was consistent with that of oxidation stacking fault (OSF) growth must be treated with caution. Most experimental investigations of OSF growth involve a decade time span, corresponding to an oxidation-rate difference of a factor of 3. Simultaneous growth and retrogrowth are also occurring, so that fits to the data involve two adjustable parameters.

It is clear, then, that only the results of Ishikawa *et al.* (1982) are in substantial disagreement with the other results in the literature. Oxidation is normally one of the most reproducible processes in the laboratory, much more so than ion implantation or diffusion, for instance. The oxide thickness values measured by Ishikawa *et al.* are in substantial disagreement with accepted values (Deal and Grove, 1965; Reisman *et al.*, 1987), suggesting some anomaly.

There is a general consensus among the other authors that the dependence of oxidation-enhanced diffusion on oxidation rate is sublinear, with a power dependence n of between 0.2 and 0.3 during parabolic oxidation. At any rate, it is less than the value of $n=0.5$ which is commonly assumed, and the difference can distinguish between some of the OED models that have been proposed. That there is any one power dependence of OED on oxidation rate is unlikely. Rather, the apparent power dependence is a reflection of a combination of nonlinear processes occurring at and close to the oxidizing interface. Measurements of the time dependence of OED can serve to distinguish between different models proposed for OED, as discussed in Sec. XIV.E. There are indications that the power dependence measured experimentally varies substantially from this consensus value for oxidation in the linear regime or for very long oxidations. However, these measurements are not trivial to carry out, as evidenced by the disagreements between some of the authors above.

A possible reason for this disagreement is that the Czochralski material used in many of these studies is much richer in terms of these complexities than has been previously realized. Oxygen and carbon are two impurities present in high concentrations in CZ material, and the effects of different time-temperature cycles on these elements do influence the point-defect processes. This is still an active area of research. What is important from a diffusion viewpoint is that the precipitation of oxygen to form SiO_2 precipitates can and does influence the point-defect populations in the silicon wafer, which may be further complicated by the fact that the type and condition of the surface covering can also play a role in the point-defect kinetics. For these reasons the material of choice in determining the fundamental OED diffusion kinetics should be float-zone material. In any case, the diffusivity in the "inert" regions must be monitored as a function of time to ensure that no other anomalous effects are occurring.

C. OED methods to probe *I-V* recombination

Most of the experiments mentioned in the preceding section dealt with OED for rather long oxidation times. It was observed that antimony diffusion was retarded while phosphorus and boron diffusion were enhanced. The kinetics leading to this steady-state situation were first investigated by Antoniadis and Moskowitz (1982a). By investigating the short-time OED kinetics, they were able to gain some insight into the process of interstitial-vacancy recombination at typical processing temperatures (900 °C–1100 °C). They measured the enhancement of phosphorus and boron at 1000 °C and the retardation of antimony at 1100 °C, using a CV technique. A surpris-

ing result of the experiment was that antimony diffusion was enhanced for very short times and that the retardation took approximately one hour to reach a steady-state value. The small enhancement implied that antimony has a small interstitial component for diffusion, which has an effect at short times before the vacancy undersaturation is complete. The time required to reach steady state provided a measure of the recombination kinetics of the interstitials and vacancies at this temperature. According to Waite's (1957) theory of diffusion-limited reactions, the recombination constant is given by

$$k_{I,V}=\frac{4\pi a_{I,V}(d_I+d_V)}{\Omega C_S} . \tag{14.2}$$

This equation can be used to estimate a value for the vacancy lifetime if some additional assumptions are made. One needs to assume that an interstitial mechanism dominates self-diffusion *or* that both interstitials and vacancies contribute equally and $C_I^* \gg C_V^*$. Then, the vacancy lifetime can be written (Gösele, Frank, and Seeger, 1983)

$$\tau_V=(k_{I,V}C_I^*)^{-1}=\frac{\Omega}{4\pi a_{I,V}D_{\text{self}}} , \tag{14.3}$$

which, for reasonable values for $a_{I,V}$ and D_{self}, gives a value for τ_V some 10^5 smaller than that determined experimentally. This led Antoniadis and Moskowitz to propose an enthalpy barrier of 1.4 eV to *I-V* recombination, so that the recombination constant is given by Eq. (12.13). Later, Gösele, Frank, and Seeger (1983) proposed an entropy barrier to recombination, in part because the large preexponential term in the self-diffusion coefficient of silicon suggests a high entropy configuration for the point defect involved.

Because of the self-diffusion term in Eq. (14.3), either of these proposed mechanisms predicts a strong temperature dependence for *I-V* recombination. Whatever the mechanism of self-diffusion, the temperature dependence of *I-V* recombination must be close to the activation energy of D_{self}. The times for *I-V* recombination should be on the order of many hours or days at lower temperatures, and the interstitial and vacancy distributions should be independent for times shorter than this; this does not seem to be the case.

Jüngling *et al.* (1987) performed a SIMS analysis of antimony profiles after oxidation at 1000 °C and 1100 °C. The data of 1000 °C show an increased amount of retardation compared to that at 1100 °C, consistent with the higher interstitial supersaturation at the lower temperatures. However, the retardation is approximately constant for times greater than 100 min, suggesting that the barrier to *I-V* recombination may not be as strongly temperature activated as expected from the above models. A summary of the short-time OED oxidation-rate dependence data obtained to date is provided in Fig. 37. We leave the discrepancy between the different authors in the time behavior of antimony at 1100 °C an open issue.

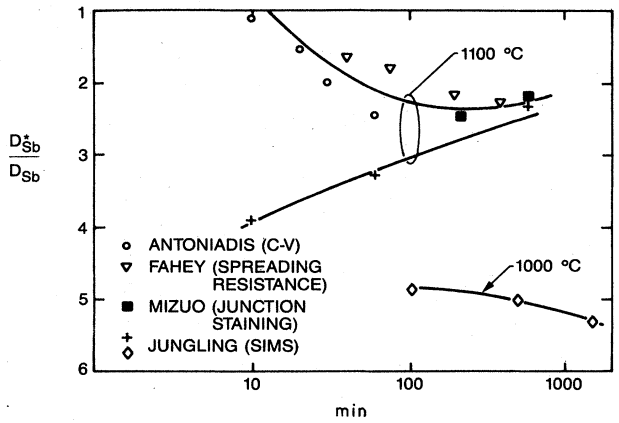

FIG. 37. Experimental results on the short-time retarded diffusion of antimony during dry oxidation.

D. OED under extrinsic conditions

For completeness, we mention the experiments that have been performed to determine OED under extrinsic conditions. Section XIII.B treats the problems from a theoretical perspective and indicates the kind of experiments that would need to be performed to decouple the effects of dopant pairing, enhanced oxide growth rates, and Fermi-level effects.

Taniguchi, Kurosawa, and Kashiwagi (1980) found that the enhancement in OED decreased sharply when the concentration of the dopant exceeded the intrinsic carrier concentration. This is open to several interpretations. The extrinsic regions might have a higher interstitial concentration due to Fermi-level effects, so the interstitial injection from oxidation has less effect. Or, there may be a change in f_{AI} with concentration, as discussed in Sec. XVII.C. Alternatively, in these experiments, where the extrinsic doped layers are adjacent to the oxidizing interface, the surface boundary conditions for interstitial injection may be changing during oxidation. Finally, the extrinsic regions may have higher vacancy concentrations due to Fermi-level effects, which could reduce the effect of the injected interstitials via recombination.

Experiments that indicate the possible utility of OED measurements under extrinsic conditions to investigate the diffusion mechanism of the dopants were performed by Ishikawa, Tomisato, Honma, Matsumoto, and Niimi (1983). The authors observed retarded diffusion of arsenic under oxidizing compared to inert conditions. This observation of retarded diffusion under interstitial injection conditions is significant because it allows more definite conclusions to be made about the arsenic diffusion mechanism at high concentrations. Arsenic seems to have both interstitial and vacancy components of diffusion, but this suggests that at high concentrations the vacancy component is more important.

Ishikawa, Matsumoto, and Niimi (1983) observed less pronounced OED of boron under extrinsic conditions

than under intrinsic conditions. In an isoconcentration experiment, Miyake (1985a) examined the OED of ^{10}B in extrinsic ^{11}B backgrounds. Interestingly enough, Miyake found that enhanced diffusion of ^{10}B tracer in the ^{11}B background was *greater* than enhanced diffusion of B under intrinsic conditions. Miyake also found that OED of B implanted into intrinsically doped substrates decreased with increasing dose; this was attributed to increased implantation damage with increased dose, creating sinks for oxidation-generated point defects. This is just the type of situation that is of practical interest to model, and it is very important to determine whether implantation damage is the dominant feature affecting diffusion. Another experiment by Miyake (1985b) involved boron OED in a uniform extrinsic phosphorus background. The magnitude of the OED decreased as the phosphorus concentration increased above the intrinsic level.

A potential problem with modeling any of the above experiments is that the oxidation takes place on a highly doped layer. This is known to affect the oxidation rate (Ho and Plummer, 1979a, 1979b) and may affect the boundary conditions for point-defect injection or absorption in an unknown manner. To circumvent this problem, a lightly doped epitaxial silicon layer was grown on an extrinsic ^{11}B substrate that contained a ^{10}B tracer profile (Kashio and Kato, 1988). The results indicated that the OED of the buried ^{10}B layer decreased as the background doping increased. Kashio and Kato attempted to determine the energy levels of the interstitial from the Fermi-level dependence of the OED. This kind of experiment on the OED of extrinsic buried layers is likely to provide important information on the point-defect kinetics in heavily doped regions (Sec. XIII.B).

E. Mathematical models for OED

A general description based on the fluxes of point defects during various physical processes is shown in Fig. 38. We chose to model the effects of surfaces on bulk point-defect populations in this way because we have found this to be a very useful conceptual framework. Specific physical evidence for this model is at present convincing, but not definitive. From an engineering perspective, however, we have been able to use this approach in quantitative numerical simulations (SUPREM-IV), which have proven useful in modeling two-dimensional diffusion profiles in advanced silicon device structures. In the following sections, we shall attempt to present the evidence favoring this approach and its current limitations.

In Fig. 38, $\mathcal{G}(t)$ is the time-dependent flux due to the generation of point defects at an active source (an oxidizing or nitriding surface). $\mathcal{D}$ is the diffusion flux into the bulk. $\mathcal{R}(t)$ and $\mathcal{R}'(t)$ are the surface recombination fluxes at the injecting and inert surfaces, respectively.

Although this simple flux description may at first sight seem to be too general to provide useful information on the OED kinetics, this is not the case. Many different de-

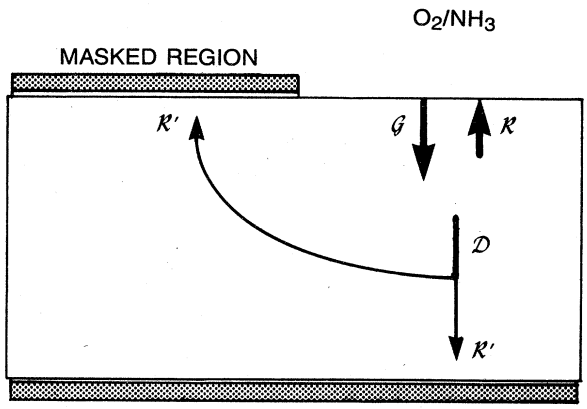

FIG. 38. A diagram of the important flux components during interstitial or vacancy injection from the surface. $\mathcal{G}$ is the generation flux of defects, $\mathcal{R}$ and $\mathcal{R}'$ are the surface loss fluxes at the injecting and the inert surface, respectively, and $\mathcal{D}$ is the diffusion flux.

tailed physical models have been proposed to explain the point-defect behavior during oxidation. In this section we present a simple method that utilizes only the fundamental assumptions of each of the proposed models, but nevertheless allows one to predict the physical behavior of the more detailed model without tediously having to solve the full differential equations. Interpreting the detailed physical models in terms of the general flux model also provides a unified basis for comparing different methods.

By examining the time dependence of each individual flux component, it is possible to determine the time dependence of the interstitial excess, which is predicted by the proposed models and determined from OED experiments. For instance, the ratio $\mathcal{G}(t)/\mathcal{R}(t)$ is a dimensionless quantity, and if we can determine the interstitial supersaturation in terms of this ratio, we immediately obtain the time dependence of ΔC_I if it depends on the balance of these two fluxes. We shall see that the interstitial supersaturation is always given by comparing the magnitude and time dependence of the $\mathcal{G}$, $\mathcal{R}$, and $\mathcal{D}$ fluxes. In many situations one of the flux components is unimportant, and the time dependence of the interstitial excess is given by the ratio of the other two. A hypothetical example would be if the $\mathcal{R}$ flux were nonexistent (i.e., no competition between generation and surface loss at the injecting interface). In this particular case, the time dependence of ΔC_I is given by the ratio $\mathcal{G}/\mathcal{D}$. Since a diffusion process has an associated flux, which varies as $t^{-1/2}$, the interstitial supersaturation remains constant if the generation flux $\mathcal{G}$ has the same order as a function of time ($t^{-1/2}$) and increases without bound if $\mathcal{G}$ decays more slowly than $t^{-1/2}$. In a physical process, there must be a limit on the interstitial supersaturation, which invokes the need for an $\mathcal{R}$ flux in the latter case.

The time dependencies of the individual flux components are not experimentally observable quantities under the usual oxidation conditions. Models with very

different starting assumptions about generation, diffusion, and recombination fluxes can then have the same overall time dependence for OED. Later, we shall see how experimental results can differentiate between some of these models. The general flux model then allows some conclusions to be drawn about the necessary behavior of the individual flux components.

The first paper that modeled the OED kinetics was by Hu (1974), and it provided a unified description of two phenomena that until then were not known to be directly related. These were the growth of stacking faults and the OED of dopants in an oxidizing ambient. Hu observed that the crystal orientation effects on both phenomena were the same and also that the effects observed were stronger in steam than dry O_2. Hu's model proposed that the oxidation was slightly incomplete and the fraction of unoxidized atoms that became interstitials supersaturated the lattice. Orientation effects were explained by proposing that surface kinks existed whose density differed on different orientations and that these acted as recombination sites for some of the interstitials. The model made the intuitively appealing assumption that the generation rate of interstitials depended on the oxidation rate, so

$$\mathcal{G}=A\frac{dx}{dt} , \tag{14.4}$$

and also assumed that the recombination of interstitials occurred in proportion to their excess concentration,

$$\mathcal{R}=\sigma_I(\Delta C_I) . \tag{14.5}$$

In addition, Hu assumed that the diffusion of the interstitials was fast, leading to a short diffusion relaxation time allowing the diffusion term to be ignored. Taking the ratio of the fluxes at the interface gives

$$\mathcal{G}/\mathcal{R}=A\frac{dx}{dt}\Big/\sigma_I(\Delta C_I) , \tag{14.6}$$

so that

$$\Delta C_I \propto \frac{dx}{dt} . \tag{14.7}$$

The basic time dependence of the OED in this case is linearly proportional to the oxidation rate. At the time the experimental evidence was not strong enough to suggest a sublinear oxidation rate dependence, and Hu returned to the fundamental ideas in this model and made suitable modifications when the experimental data were more conclusive.

Hu (1981) modified the surface regrowth condition so that two interstitials were needed to complete a recombination event, the rationale for this being shown in Fig. 39. This leads to

$$\mathcal{R}=\sigma_I(\Delta C_I)^2 . \tag{14.8}$$

The ratio now gives

$$\mathcal{G}/\mathcal{R}=A\frac{dx}{dt}\Big/\sigma_I(\Delta C_I)^2 , \tag{14.9}$$

so that

$$\Delta C_I \propto \sqrt{dx/dt} . \tag{14.10}$$

This now exhibits a sublinear dependence on the oxidation rate, which agrees better with the experimental results.

The requirement that sufficient free volume be available in the interface region for the incoming oxygen species to occupy and react with the local silicon lattice was realized by Tiller (1981). This led Tiller to propose that while some of the free volume requirement might be met by a type of SiO_2 viscous flow, it could also be satisfied by interstitial egress from the interface. The interstitials generated in this way partition into the silicon and the SiO_2 depending on the chemical potential difference between the regions for the interstitials. By assuming the existence of a transitory crystalline layer with chemical stoichiometry varying from SiO_x to SiO_2 over a 3–5 unit cell distance from the interface, Tiller obtained a qualitative explanation of the trends in OED, stacking fault, and interface charge effects, including their orientation dependence.

A particular quantitative version of this model was later proposed by Lin, Dutton, Antoniadis, and Tiller (1981). While assuming that the generation rate was directly proportional to the oxidation rate, they considered a reverse flux of interstitials into the oxide. By postulating that these interstitials were consumed in a nonstoichiometric boundary layer according to the reaction

$$\mathrm{Si}_I+\frac{x}{2}\mathrm{O}_2\rightarrow\mathrm{SiO}_x , \tag{14.11}$$

they obtained an expression for the surface loss of interstitials from the chemical kinetics of the last equation as follows:

$$\mathcal{R}=k\Delta C_I C_{\mathrm{O}_2}^{x/2} . \tag{14.12}$$

Since the oxidation rate depends directly on the oxygen concentration at the interface, we can write

$$\mathcal{R}=k_1\left[\frac{dx}{dt}\right]^{x/2}\Delta C_I , \tag{14.13}$$

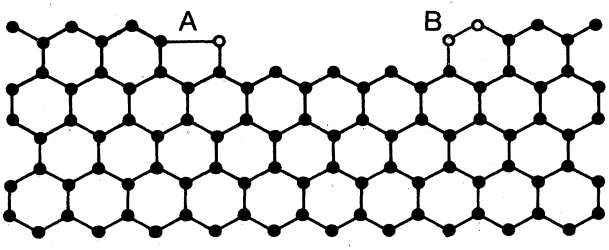

FIG. 39. Illustration of the rationale for the bimolecular recombination of interstitials at a step in the Si/SiO_2 interface. Surface regrowth at kinks or steps can occur by the capture of a pair of interstitials.

giving

$$\mathcal{G}/\mathcal{R} = A\frac{dx}{dt}\Big/ k_1\left[\frac{dx}{dt}\right]^{x/2}\Delta C_I \ , \tag{14.14}$$

so that

$$\Delta C_I \propto \left[\frac{dx}{dt}\right]^{1-x/2} . \tag{14.15}$$

Depending on the value of x, this gives rise to a sublinear dependence of the OED on the oxide growth rate. Lin *et al.* took the value of x to be 1.2 to agree with some empirical data based on stacking fault growth. Physically, it seems that x can be either 1 or 2, which allows the OED to be proportional to $\sqrt{dx/dt}$ or a constant with time, respectively. Lin *et al.* did not propose reactions with atomic and molecular oxygen but did suggest that the interface layer might somehow represent a mixture of SiO and SiO_2. If these interface reactions occur without explicit O_2 dissociation, some mechanism for stopping some of the oxidation reactions at the SiO phase must exist, since it is known that molecular oxygen is the only important transport species.

Fair (1981b) proposed that the sublinear OED kinetics could be explained by reactions involving atomic oxygen. The detailed mechanism involved molecular oxygen reacting at the gas-oxide surface to produce atomic oxygen, which then became the diffusing and reacting species. The concentration of molecular oxygen available for this reaction is proportional to the partial pressure of O_2 in the gas phase, by Henry's law, making the oxidation rate proportional to P_{O_2}. From the dissociative reaction, the concentration of atomic oxygen can be obtained:

$$C_{O_2} = k' C_O^2 \ . \tag{14.16}$$

Assuming atomic oxygen to be the reacting species at the interface, Fair proposed that the generation rate of interstitials was some fraction of the interface reactions, giving

$$\mathcal{G} = AC_O = A/k'' C_{O_2}^{1/2} = A\Big/ k_2\left[\frac{dx}{dt}\right]^{1/2} . \tag{14.17}$$

Allowing for an interface regrowth mechanism and ignoring the diffusion flux, as Hu did originally, gives

$$\mathcal{R} = \sigma_1 \Delta C_I \ . \tag{14.18}$$

To find the time dependence of the interstitial concentration we take the ratio

$$\mathcal{G}/\mathcal{R} = \frac{A}{(k_2\sigma_I)}\left[\frac{dx}{dt}\right]^{1/2}\Big/\Delta C_I \ , \tag{14.19}$$

so that

$$\Delta C_I \propto \sqrt{dx/dt} \ . \tag{14.20}$$

While this does give a sublinear dependence on the oxide growth rate, it relies on the questionable assumption that atomic species are the dominant transport species. In the diffusion-limited regime of growth, the partial-pressure dependence of the oxidation reaction is linearly proportional to P_{O_2}, indicating that the reaction is of first order. This precludes a dissociation reaction giving rise to atomic transport species, which would cause a higher-order reaction kinetics.

A macroscopic rather than an atomistic approach to the problem was taken by Tan and Gösele (1981). The authors proposed that the viscoelastic flow of the SiO_2 determined the supersaturation of interstitials in the silicon. The rate of strain was assumed to be directly proportional to the oxide growth rate and also nonlinearly related to the stress. By assuming that the stress was directly responsible for the interstitial supersaturation the authors proposed that

$$\Delta C_I \propto \text{stress} \propto \left[\frac{d\gamma}{dt}\right]^{1/m} \propto \left[\frac{dx}{dt}\right]^{1/m} , \tag{14.21}$$

where γ is the strain and m accounts for the nonlinear relationship between the stress and the strain. This model certainly has the desired sublinear dependence on oxidation rate, but the parameters in the model have not been determined independently.

Evidence that the interstitial diffusion coefficient might not be as fast as previously thought led Hu (1983) to include the effects of the diffusion flux in his original model, which had the generation rate proportional to the oxidation rate, and surface regrowth proportional to the excess interstitial concentration. Because the time dependence of the generation and diffusion fluxes are of the same order, we need to consider all three flux components in this model. In the other models we have examined, the investigators either explicitly or implicitly ignored the diffusion flux of interstitials into the silicon.

The $\mathcal{G}/\mathcal{R}$ flux balance leads to $\Delta C_I \propto dx/dt$ as before [Eq. (14.7)], while the $\mathcal{G}/\mathcal{D}$ flux balance leads to the excess concentration of interstitials being a constant. We can see this by making the following approximation for the diffusion flux:

$$\mathcal{D} \approx d_I \frac{\Delta C_I}{\Delta x} \approx d_I \frac{\Delta C_I}{\sqrt{d_I t}} \approx \sqrt{d_I}\,\Delta C_I t^{-1/2} . \tag{14.22}$$

During the parabolic regime of oxidation, when the oxide thickness increases as $\sqrt{t}$, the generation rate has a $t^{-1/2}$ dependence in Hu's model. This gives

$$\mathcal{G}/\mathcal{D} = At^{-1/2}/\sqrt{d_I}\,\Delta C_I t^{-1/2} , \tag{14.23}$$

so that

$$\Delta C_I \propto A/\sqrt{d_I}, \ \text{a constant} . \tag{14.24}$$

In this way, the interstitial supersaturation varies depending on the dominant flux balance. To find the transition time from one regime to the other, we can determine at what point in time the *magnitude* of the $\mathcal{D}$ flux equals that of the $\mathcal{R}$ flux. Equating gives

$$\sqrt{d_I}\Delta C_I t^{-1/2}=\sigma_I \Delta C_I \ , \tag{14.25}$$

so that

$$t_0=d_I/\sigma_I^2 \ . \tag{14.26}$$

(This is the characteristic surface lifetime of I, τ_∞, for a semi-infinite substrate discussed in Sec. XII.A.2.) For times much shorter than the crossover time t_0, the $\mathcal{G}/\mathcal{D}$ flux term is dominant and the interstitial supersaturation is better approximated by a constant, while for times much greater than t_0, $\Delta C_I(t)$ falls linearly with the oxidation rate. Figure 40 shows that the predicted power-law dependence of OED on oxidation rate varies smoothly from 0 to 1, so that over some of the time range Hu's model predicts a power-law dependence that matches the experimental results. This simple analysis shows that Hu's model can be described using the same generalized approach by considering all the important flux terms.

Hu (1985b) later went on to consider the expected time dependence of OED during the linear phase of oxidation. Here the oxide thickness varies linearly with time, so the generation flux is a constant:

$$\mathcal{G}=A\frac{dx}{dt}=g_I^0 \ . \tag{14.27}$$

The interstitial supersaturation generated by this level of injection continues to build up and eventually becomes limited by the balance of the generation and surface recombination fluxes. During the early stages of the process the flux balance of interest is

$$\mathcal{G}/\mathcal{D}=g_I^0/\sqrt{d_I}\Delta C_I t^{-1/2} \ , \tag{14.28}$$

so that

$$\Delta C_I \propto t^{1/2} \ . \tag{14.29}$$

Initially, the interstitial excess follows this power-law dependence and rises to a final value determined by the other flux balance,

$$\mathcal{G}/\mathcal{R}=g_I^0/\sigma_I \Delta C_I \ , \tag{14.30}$$

so that

$$\Delta C_I \propto g_I^0/\sigma_I \ , \tag{14.31}$$

which gives a final constant value for the interstitial supersaturation if the oxidation rate remains constant. Figure 40(a) shows the regions where the different flux terms are important for linear oxide growth. For comparison, Hu's (1985a) complete analytical solution for combined linear-parabolic oxidation is shown in Fig. 40(b).

Matsumoto, Ishikawa, and Niimi (1983) proposed a rather arbitrary variation to Hu's model by introducing a generation rate proportional to $\sqrt{dx/dt}$. Keeping the recombination rate proportional to the excess interstitial concentration and ignoring the diffusion flux gives

$$\mathcal{G}/\mathcal{R}=A\sqrt{dx/dt}\ /\sigma_I \Delta C_I \ , \tag{14.32}$$

so that

$$\Delta C_I \propto \sqrt{dx/dt} \ . \tag{14.33}$$

The authors then analyzed the data of Ishikawa *et al.* (1982) using this model and determined the variation in the fraction interstitialcy component of diffusion with temperature.

A model that linked atomic reactions at the interface, thought to be involved in the detailed oxide growth kinetics, with the point-defect reactions was proposed by Dunham and Plummer (1986a). The original Deal-Grove model for silicon oxidation supposes that molecular oxygen is the only species involved in the growth of oxide films. The linear rate constant, which models the interface-reaction-limited stages of early growth, was later found to be sublinearly dependent on the oxygen partial pressure (van der Meulen, 1972), which indicates a higher-order reaction. Ghez and van der Meulen (1972) explained this result by considering the dissociation of molecular oxygen at the interface with subsequent reactions of both the atomic and the molecular oxygen. At low temperatures, the reaction with atomic oxygen is dominant and the linear rate constant approaches a 0.5 power dependence on the partial pressure of O_2, and at high temperatures the reaction with molecular oxygen is dominant, leading to a linear dependence on O_2 partial pressure. The interface reaction that promotes the association-dissociation reaction is

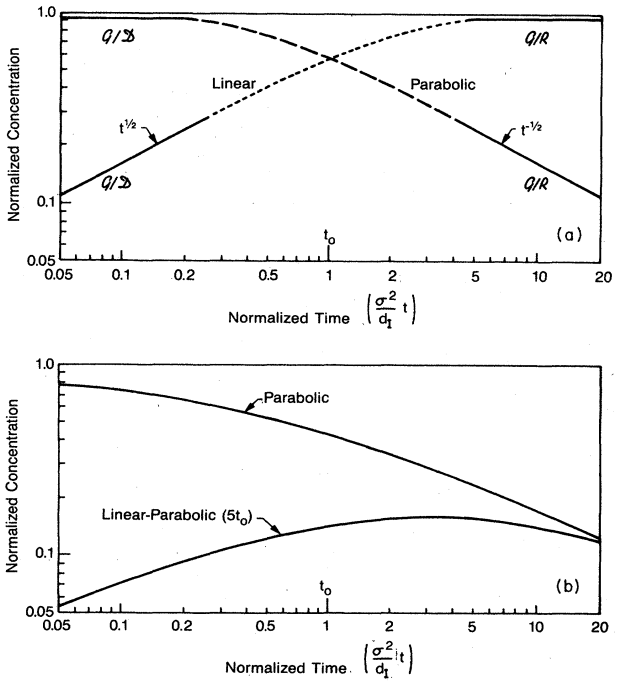

FIG. 40. Interstitial excess ΔC_I at the silicon surface. (a) Qualitative diagram of the behavior of ΔC_I for linear and parabolic oxidation obtained from the analysis of flux components. (b) Exact solution for the case of parabolic oxidataion and combined linear parabolic oxidation from Hu (1985a). The transition time from linear to parabolic oxidation is $5t_0$.

$$O_2 \rightleftharpoons 2O \ , \tag{14.34}$$

giving a quasi-steady-state O_2 concentration in terms of the O concentration as

$$C_{O_2} = k' C_O^2 \ . \tag{14.35}$$

This reaction plays a key part in Dunham and Plummer's point-defect model; they suggested that the volume expansion at the interface contributes to the interstitial generation and proposed that a certain percentage of the interface reactions resulted in the creation of an interstitial. They assumed that most of these interstitials diffuse back into the oxide, where they are further oxidized to SiO_2, with the remainder giving rise to the point-defect excess in the silicon. The diffusion flux of interstitials into the bulk was estimated to be negligible by comparison. The nonlinearity in the model is introduced by considering the diffusion of interstitials into the oxide, where they can react with the incoming molecular oxygen, which is known to be the major transport species (Fig. 41). The growth rate is determined by the molecular oxygen concentration at the gas/oxide interface, since all of the incoming oxygen is eventually transformed to SiO_2. At high temperatures the molecular oxygen is also dominant in the interface reaction. The generation flux was assumed to be some percentage of the interface reactions, so that, at high temperature,

$$\mathcal{G} = AC_{O_2} = A' \frac{dx}{dt} \ . \tag{14.36}$$

The surface loss term was determined by the solution of the diffusion equation for the interstitials, accounting for the reaction with the transport species as follows:

$$d_I \frac{\partial^2 \Delta C_I}{\partial x^2} = k C_{O_2} \Delta C_I \ , \tag{14.37}$$

which for quasiconstant oxidant concentrations gives the surface loss flux as

$$\mathcal{R} = d_I \frac{\partial \Delta C_I}{\partial x} = (d_I k C_{O_2})^{1/2} \Delta C_I \ . \tag{14.38}$$

The time dependence of the interstitial supersaturation is given by the balance of the two fluxes,

$$\mathcal{G}/\mathcal{R} = \frac{AC_{O_2}}{(kd_I C_{O_2})^{1/2} \Delta C_I} \ , \tag{14.39}$$

so that

$$\Delta C_I \propto (C_{O_2})^{1/2} = \sqrt{dx/dt} \ . \tag{14.40}$$

When atomic species at the interface are the major reactant species, the molecular oxygen concentration near the interface is given by the reaction

$$C_{O_2} = k' C_O^2 \ . \tag{14.41}$$

Dunham's assumption was that the generation rate of interstitials was some fraction of the interface reactions, giving

$$\mathcal{G} = AC_O \ . \tag{14.42}$$

The diffusion equation with the interstitials still involved reactions with bulk molecular oxygen, whose concentration near the interface was given by Eq. (14.41), so that

$$d_I \frac{\partial^2 \Delta C_I}{\partial x^2} = k C_{O_2} \Delta C_I = k_1 C_O^2 \Delta C_I \ . \tag{14.43}$$

Again, assuming quasiconstant oxidant concentrations near the interface allowed the recombination flux to be calculated,

$$\mathcal{R} = d_I \frac{\partial C_I}{\partial x} = \sqrt{d_I k_1} C_O \Delta C_I \ , \tag{14.44}$$

and the flux balance to be determined,

$$\mathcal{G}/\mathcal{R} = \frac{AC_O}{\sqrt{d_I k_1} C_O \Delta C_I} \ , \tag{14.45}$$

so that

$$\Delta C_I \propto A/\sqrt{d_I k_1}, \ \text{a constant} \ . \tag{14.46}$$

These two cases give a summary of the reactions that were considered more rigorously by Dunham and Plummer. It can be seen that the reaction with atomic oxygen at the interface produced by the dissociation reaction gives rise to a constant interstitial supersaturation, and that molecular oxygen reacting at the interface introduces a 0.5 power dependence on the oxidation rate. According to van der Meulen, both atomic and molecular oxygen contribute strongly to the oxidation at 1000 °C in ⟨100⟩ silicon, which is at least in qualitative agreement with the power dependence of 0.19 on the oxidation rate for interstitials observed by Dunham and Plummer at this temperature. If the trend shown in Fig. 36 is significant and not within experimental error, then one wonders why the exponent n is larger at 900 °C than at 1000 °C, where the interface reaction presumably favors atomic oxygen, according to Ghez and van der Meulen's (1972) model.

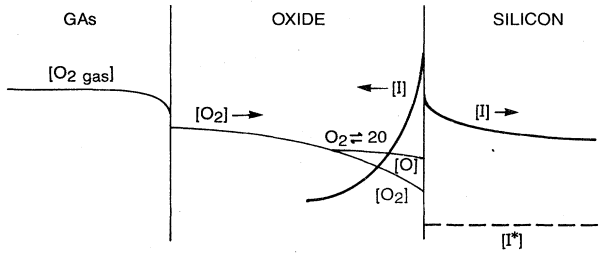

FIG. 41. A model for the Si/SiO_2 system during oxidation in a dry O_2 ambient. From Dunham and Plummer (1986a). Interstitials diffuse and react with incoming molecular and atomic oxygen, relieving the strain at the interface.

F. Information from nitridation experiments

With regard to effects on impurities and stacking faults, thermal nitridation is the direct complement of oxidation. During the nitridation of a bare silicon surface in an ammonia ambient, stacking faults shrink faster than usual, antimony diffusion is enhanced, and boron and phosphorus diffusion are retarded. It is clear that there is a supersaturation of vacancies and a corresponding undersaturation of interstitials.

An interesting subset of the nitridation experiments concerns the exposure of an existing oxide to ammonia gas, creating an oxynitride. This process has similar effects to oxidation in every way, namely, stacking fault growth, boron and phosphorus diffusion enhancements, and retardation of antimony diffusion. Table II provides a concise summary of the effects of oxidation, nitridation, and oxynitridation. Oxidation data pertain to oxidations of ⟨100⟩-oriented samples [oxidation of ⟨111⟩ surfaces at temperatures at or about 1150 °C appears to inject vacancies rather than self-interstitials into the bulk (Tan and Ginsberg, 1983)]. Nitridation results are for temperatures $\geq$ 1000 °C (nitridation studies have not been performed for lower temperatures). The nitridation results for Ga are from Fahey, Iyer, and Scilla (1989).

There are two different issues related to the nitridation reactions: the underlying causes of point-defect creation at the surface and the information provided by nitridation reactions on the mechanisms of dopant diffusion.

Turning first to the question of what causes point-defect excesses, we begin by examining what is known about direct nitridation of the silicon surface. It is worthwhile to examine the kinetics of the Si_3N_4 growth process, since it is significantly different from the more familiar oxide growth process. As seen in Fig. 42, the nitride film thickness follows a logarithmic growth law, with a final film thickness in the region of 50 Å at 1100 °C. Superimposed on the same figure is the

TABLE II. Summary of interface processes on diffusion.

Stacking faults	Oxidation $I\uparrow V\downarrow$ Grow	Oxynitridation $I\uparrow V\downarrow$ Grow	Nitridation $I\downarrow V\uparrow$ Shrink
P,B diffusion			
intrinsic	Enhanced	Enhanced	Retarded
extrinsic	Enhanced	Enhanced	Retarded
Sb diffusion			
intrinsic	Enhancement precedes retardation	Enhancement precedes retardation	Enhanced
As diffusion			
intrinsic	Enhanced	Enhanced	Enhanced
extrinsic	Retarded or no effect	Enhanced	Enhanced
Ga diffusion			
intrinsic	Enhanced	Enhanced	Retarded

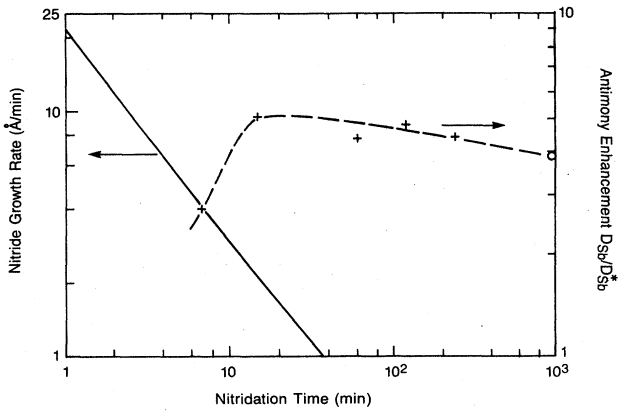

FIG. 42. A plot of the nitride film growth rate on silicon as a function of time in an ammonia ambient at 1100 °C. From Moslehi and Saraswat (1985). The enhancement in antimony diffusivity is also shown for comparison.

enhancement in antimony diffusivity, corresponding to the vacancy supersaturation during nitridation. It is clear that nitridation injects point defects even when the film growth has essentially stopped. This is one of the most interesting results to consider when trying to assess the nature of the point-defect injections that result from oxidation and nitridation processes. But even though the nitride film growth almost stops, the composition of the film continues to change (Moslehi and Saraswat, 1985), indicating that a static situation is not reached when film growth stops. In fact, if an existing thermal nitride film is annealed in an argon ambient, there continues to be an enhancement in the Sb diffusivity that persists for many hours. The enhancement is less than that in a pure NH_3 ambient and decays slowly with time. When the thermal nitride film is removed, the antimony diffusivity returns to the inert value (Ahn, Kennel, Plummer, and Tiller, 1988a). It is also known (Mizuo and Higuchi, 1982b; Ahn *et al.*, 1988b) that deposited nitride films cause enhanced Sb diffusivity, the degree of which is closely related to the stress level in the silicon nitride films. The stress level can be varied by varying the stoichiometric composition of the films. The phenomenon qualitatively resembles that under thermal nitride films, but the effect is smaller under deposited films. Ahn *et al.* (1988b) have suggested that a plausible explanation for the vacancy supersaturation under the thermal nitride film is the release of stress in the film by the generation of Frenkel defects at the Si/Si_3N_4 interface. The Si atoms could then be absorbed by the film while the companion vacancies are injected into the bulk, causing a supersaturation near the interface.

Because of the importance of the complementary process to oxidation in providing insight into the point-defect behavior, we examine the nitridation kinetics more closely. Hayafuji, Kajiwara, and Usui (1982) first showed the increased shrinkage rate of stacking faults in an am-

monia ambient. This effect occurs over the temperature range from 1050 °C to 1200 °C, indicating that there exists an undersaturation of interstitials and/or a supersaturation of vacancies in the bulk silicon. The magnitude of the stacking fault shrinkage rate depended on the partial pressure of the ammonia in the ambient, suggesting that the actual chemical reaction at the surface is important in determining the point-defect perturbations. Hayafuji *et al.* proposed that cation migration of silicon to the surface of the nitride film might cause an interstitial undersaturation, but the growth mechanism of thermal nitride films is not well understood.

Mizuo *et al.* (1983) were the first to report on the effects of nitridation on dopant diffusion. They showed that boron and phosphorus diffusion was retarded, confirming an interstitial undersaturation, for temperatures between 1000 °C and 1150 °C in ⟨100⟩ and ⟨111⟩ silicon. In addition, antimony diffusion was enhanced at these temperatures for times up to 4 h. This implies both a vacancy supersaturation and an interstitial undersaturation.

While there can be no argument as to the qualitative results presented above, one can ask if the controlling mechanism during nitridation is interstitial depletion or vacancy injection. We can argue that a vacancy injection mechanism during direct nitridation is more plausible than an interstitial depletion, primarily because a vacancy excess will not result solely from interstitial depletion. Even in the case in which a significant deficit of interstitials result from a surface depletion mechanism, it does not immediately follow that a supersaturation of vacancies will result. Since we are discussing the hypothetical situation in which there is no externally induced generation of vacancies at the surface, it follows that the supersaturation of vacancies must result from *bulk generation.* The mechanism by which this would be achieved may be pictured as follows. As interstitials are depleted at the surface by chemical reaction, C_V is initially constant throughout the bulk and equal to C_V^*. Now, as interstitials and vacancies are created through normal generation processes we imagine the bulk-generated interstitials to be quickly swept away and consumed at the surface; they are thus unavailable for recombination with the vacancies that remain behind. This introduces an imbalance in the *I-V* generation reaction, resulting in a net generation of vacancies. However, a sizable supersaturation of vacancies will result only if the bulk generation rate of vacancies is greater than those processes that serve to reduce the vacancy excess. Specifically, excess vacancies are free to diffuse away into the bulk with a flux $-d_V(\partial C_V/\partial x)$, or may recombine at the surface with a flux $-\sigma_V(C_V - C_V^*)$. Based on available data and current theoretical expectations, the diffusion of vacancies and surface recombination should be sufficiently fast that an appreciable vacancy excess at the surface will not be sustained, even for an arbitrarily large depletion of interstitials.

Additional evidence that vacancy injection is the controlling mechanism during thermal nitridation comes from the short-time kinetics data of Fahey, Barbuscia, Moslehi, and Dutton (1985) shown in Fig. 43. It seems that phosphorus retardation occurs after the antimony enhancement, so that steady-state bimolecular annihilation takes some time to occur. This closely parallels the complementary case for short-time OED kinetics (see Sec. XIV.C), with a much shorter time constant. Since the cause must occur before the effect, it implies that vacancy injection is the controlling mechanism during nitridation, but the question still remains as to what process causes the vacancies to be injected.

We next consider the case of nitridation of an SiO_2 layer. Generation of self-interstitials by nitridation of an

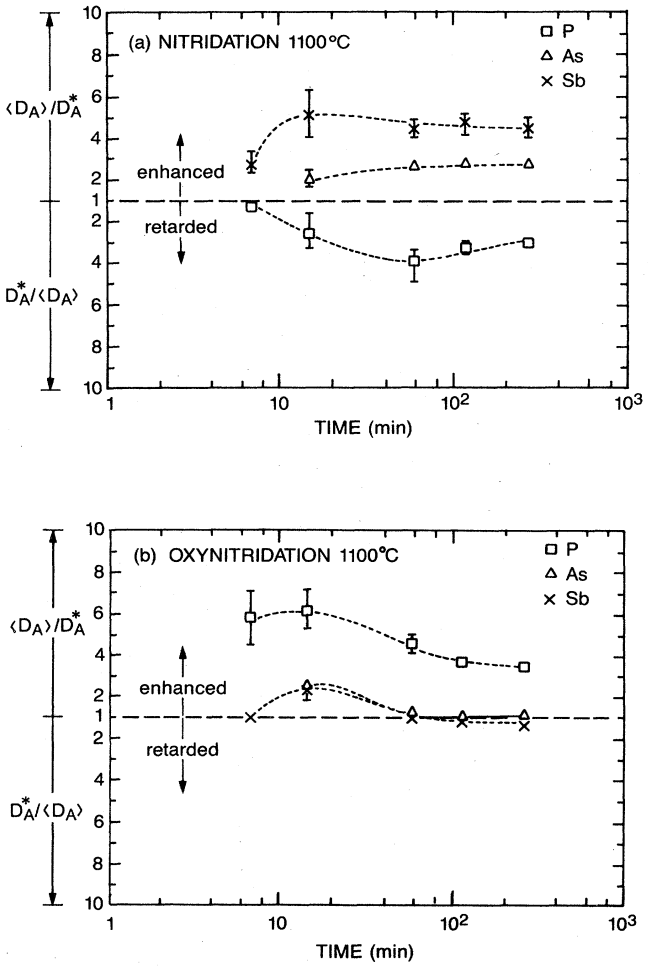

FIG. 43. The effects of nitridation of Sb, P, and As diffusion at 1100 °C: (a) direct nitridation of the silicon surface produces complementary effects to oxidation; (b) nitridation of a 350-Å oxide layer has similar effects to oxidation. From Fahey, Barbuscia, Moslehi, and Dutton (1985).

SiO_2 layer is an interesting result and contrasts sharply with self-interstitial generation during oxidation. In the case of oxidation, most thinking has been along the lines that as the silicon surface is consumed by the oxidation reaction, some silicon atoms are injected into the substrate as self-interstitials (the various proposed processes by which this occurs are discussed in Sec. XIV.E). But for nitridation of an SiO_2 layer, if there is any consumption of the silicon surface by nitriding reactions it is apparently too small to be easily observable, and yet the interstitial supersaturations that result are at least as great as those that occur during oxidation of a silicon surface in a 100% O_2 ambient. Nitrogen incorporation into the films has been studied [see, for example, the article by Han *et al.* (1985) and references therein], but it is not clear how this would lead to self-interstitial injection into the bulk. Vasquez *et al.* (1984) observed an oxygen-rich region at the nitrided-oxide/silicon interface and suggested as a possible explanation that some oxidation might be taking place at the silicon surface as a result of reaction with by-products (possibly hydroxyl species) formed from reactions of NH_3 with SiO_2. An oxygen-rich layer was also observed in the study of Han *et al.* (1985), who concluded that the oxygen present at the nitrided-oxide/silicon interface originated as a by-product of exchange reactions in the bulk of the oxide during nitrogen incorporation. But even if oxidation does take place during nitridation of SiO_2, it is still necessary to come up with an explanation of why such a relatively small oxidation rate should lead to such a large interstitial supersaturation. One possibility is that oxidation is occurring in a completely different environment from that of a freely growing oxide and that for some reason, perhaps involving film stress, a much higher number of self-interstitials are generated per amount of silicon consumed than under normal oxidation conditions. From another viewpoint, it might also be proposed that it is not the generation rate of self-interstitials that is high, but rather the surface loss that is low for a nitrided oxide. In Sec. XIV.E it is shown that the excess self-interstitial concentration $\Delta C_I = C_I - C_I^*$ is largely controlled by the ratio of the generation flux at the surface to the surface loss flux $\mathcal{G}/\mathcal{R}$. Thus a small value of $\mathcal{R}$ is just as effective as a large value of $\mathcal{G}$ in increasing an interstitial supersaturation. This explanation is at the heart of the model proposed by Dunham (1987) to account for self-interstitial supersaturations during nitridation of SiO_2. Experiments such as those described in Sec. XV might be able to test the validity of this model.

From the preceding discussions it is obvious that there is no real understanding of why nitridation reactions generate point defects. But regardless of how point defects are generated, the nitridation results taken together with the oxidation results (summarized in Table II) allow us to make more definitive statements about the mechanisms of dopant diffusion than would be possible based on the results of oxidation studies alone. The different responses of dopants to the same point-defect injection conditions at the surface (i.e., enhanced versus retarded diffusion) demonstrate that all dopants cannot diffuse by the same mechanism. The experimental results in Table II can be explained consistently by assuming Sb diffusion to be dominated by a vacancy mechanism, P, B, and Ga to be dominated by a substitutional-interstitial(cy) interchange mechanism, and As to have comparable components of both types of mechanisms. For the qualitative results of Table II, "dominated" means that more than 50% of the diffusivity measured under equilibrium conditions can be attributed to a particular mechanism. Quantitative analysis of these data is necessary before more precise estimates of the relative importance of vacancy and interstitial(cy)-assisted mechanisms can be made. A detailed analysis of diffusion data from a nitridation study performed at 1100 °C is presented in Sec. XVII.

XV. TESTING THE MODELS

Since many models have been proposed for OED in particular, the question arises of how to test the various models. It is obvious that a straightforward comparison of the models with experimental evidence (where it exists) in one dimension is inadequate, since each model explains the sublinear power dependence of OED on oxide growth rate. Some of the models can be tested by extending the experimental conditions to the regime where the oxide growth is linear, or by considering extremes of low partial pressure, high growth rates, and low or high temperatures. Another very useful test of these models involves their extension from one dimension to two dimensions. This also gives some insight into the relationship between the generation, recombination, and diffusion fluxes, which is not easily accessible otherwise. These tests also give the same information about the nitridation process and could provide a basis for a model of point-defect injection during nitridation. Recently, Griffin and Plummer (1987) used a two-dimensional test structure to determine the interface kinetics in oxidizing and nitriding ambients. The structure consisted of wide, inert, masked stripes separated by progressively narrower oxidizing (or nitriding) stripes (Fig. 44). A uniform blanket implant layer of boron (or antimony) was used to monitor the enhancement in diffusivity due to oxidation (or nitridation).

For the case of oxidation, to first order, the full OED was observed under the oxidizing regions independent of the width of the oxidizing stripe, as seen in Fig. 45. This is somewhat surprising, because the masked regions adjacent to the oxidizing stripes are thought to be efficient sinks for the injected interstitials. That this is so can be inferred from the following observations. (i) There is a different lateral extent of OED when different materials are used as masks (Sec. XVI). (ii) The lateral extent of OED falls off faster than the vertical extent of OED. Since perturbing the diffusion flux by the adjacent sinks has no effect on the OED, we infer that the interface kinetics are independent of this flux component. The inter-

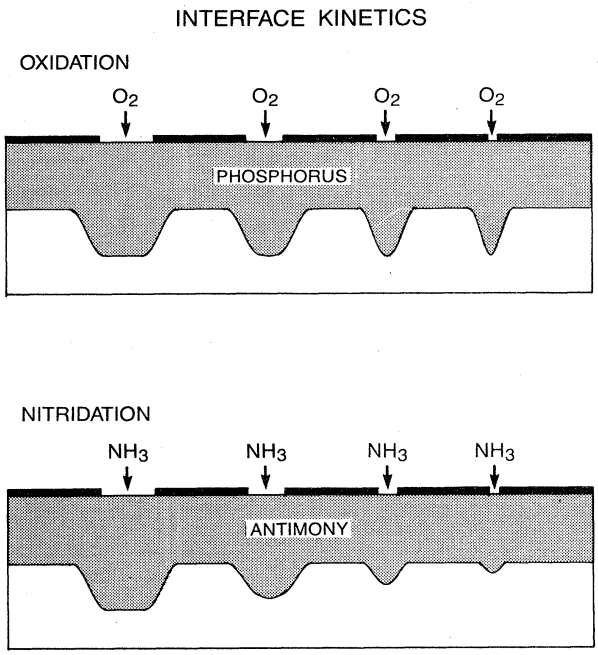

FIG. 44. Illustration of a two-dimensional test structure to probe the interface kinetics during point-defect injection. The wide inert (masked) regions act as sinks for the point defects and can influence the supersaturation under the injecting regions.

stitial excess is determined only by the balance of the generation and recombination terms at the oxidizing interface.

Simulations of the two-dimensional OED test structure using the general flux model of Fig. 38 indicate that the recombination velocity at the oxidizing interface must be much larger than at an inert pad-oxide. This means that the surface loss at an oxidizing surface must be much greater than at a nonoxidizing surface. The parameter set used for these simulations was $(g_I,\sigma_I,\sigma_I',d_I)$, where g_I is the generation rate of interstitials, σ_I and σ_I' are the recombination velocities at the oxidizing and inert surfaces, respectively, and d_I is the self-interstitial diffusivity. Effective values for two of these parameters, σ_I' and d_I, can be determined independently (see Sec. XVI). Representative values for these parameters at 1100 °C are

$$d_I = 1\times10^{-9}\ \text{cm}^2/\text{sec} \tag{15.1}$$

and

$$\sigma_I' = 4\times10^{-7}\ \text{cm/sec}\ . \tag{15.2}$$

It remains to choose values for the remaining parameters, g_I and σ_I. The generation rate g_I was chosen so that $g_I \propto \sqrt{dx/dt}$, without prejudice to any of the models mentioned in Sec. XIV.C. This choice for the time dependence means that the diffusion flux into the silicon bulk (which is of order $t^{-1/2}$) does not directly determine the interstitial supersaturation. The absolute value of the generation rate was scaled for each recombination velocity σ_I to get the same interstitial supersaturation and hence the same dopant enhancement. For comparison, the two cases considered were (i) $\sigma_I=\sigma_I'$ and (ii) $\sigma_I \gg \sigma_I'$.

As seen in Fig. 46, both models do an adequate job of modeling the OED in one dimension. When these models are used to predict the stripe-width dependence of OED, the results are very different, as shown in Fig. 47. When $\sigma_I=\sigma_I'$, the interstitials are "shared" between the oxidizing and masked interfaces, leading to less OED

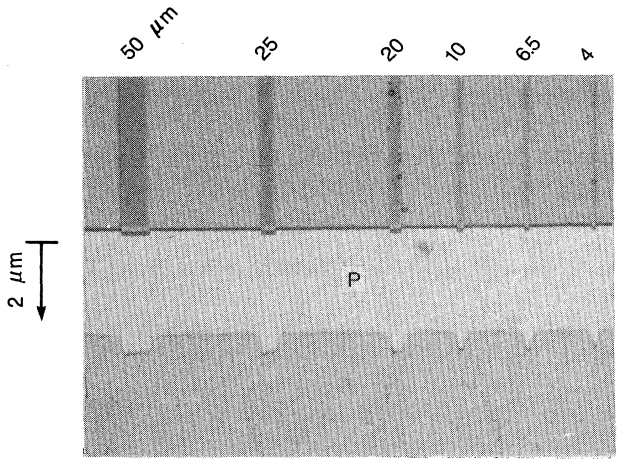

FIG. 45. Two-dimensional OED results. A photograph of a beveled-and-stained phosphorus junction after stripes of different widths at the surface have been oxidized for 4 h at 1100 °C.

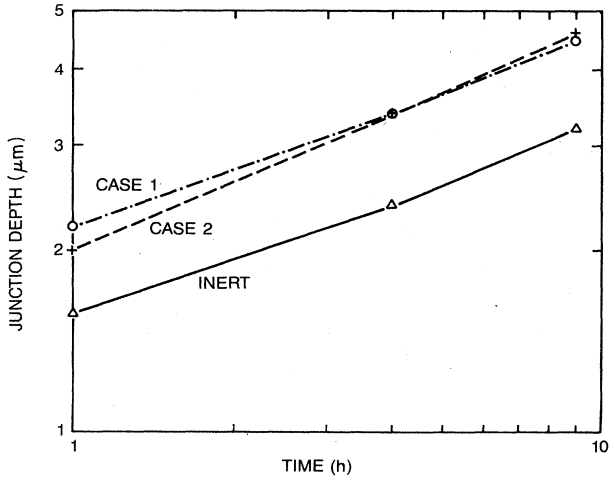

FIG. 46. Comparison of two different models for the point-defect generation at an oxidizing surface. Little variation is seen in the predictions of junction depths for a diffusion in one dimension.

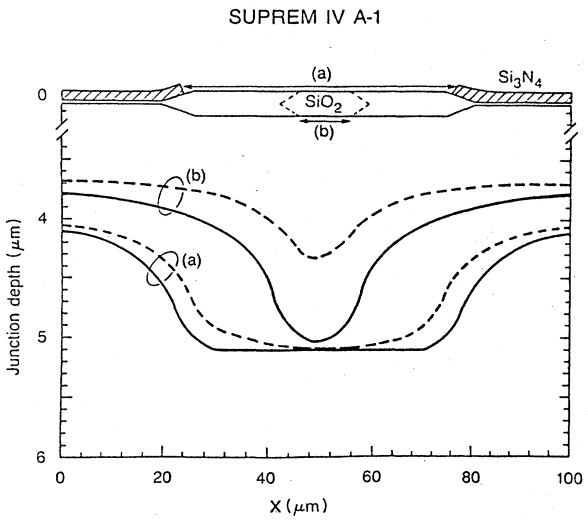

FIG. 47. Diffusion models that predict similar results in one dimension (see Fig. 46) give completely different results when applied to a two-dimensional diffusion. Dashed lines show the case where the surface recombination velocity of I at the oxidizing interface, σ_I, is the same as the recombination velocity at the nonoxidizing interface, σ_I'. Solid lines, which correspond more closely to experimental results, are calculated assuming $\sigma_I \gg \sigma_I'$.

when the dimension of the oxidizing stripe is reduced. Only the model that has the surface loss at the oxidizing interface greater than at the inert interface is consistent with the OED experiments.

This comparison enables us to draw some important conclusions about point-defect generation during silicon oxidation. It is clear that the interstitial concentration during oxidation is "pinned" at a particular value independent of the diffusion flux. A natural description of this process is that the interstitial excess is determined by the balance of two large fluxes, a generation flux and a recombination flux at the same interface. In addition, the recombination velocity at the oxidizing surface must be much greater than that at an inert surface. This means that a substantial fraction of the generated interstitials are reabsorbed at the same interface. The physical model proposed by Hu (1983, 1985b) does not account for different surface recombination velocities under oxidizing and nonoxidizing surfaces. Dunham and Plummer (1986a) neglect the surface loss flux when the film is not oxidizing. Although we used constant recombination velocities to model the dopant OED, the physical recombination process must be time dependent because there cannot be a quantum change in the recombination velocity between an oxidizing and a nonoxidizing surface. The time dependence observed in experiments where dopant OED is monitored is the overall time dependence of the ratio $\mathcal{G}(t)/\mathcal{R}(t)$, and it is difficult to separate out the individual time dependencies from straightforward OED experiments.

Experiments in which the striped pattern described above has the oxidizing regions replaced by nitriding regions show distinctly different results. The thermal nitridation of the silicon surface in an ammonia ambient injects vacancies, and this enhances the diffusion of an antimony marker layer. Now, however, the magnitude of the enhanced diffusion depends strongly on the width of the nitriding stripe (Fig. 48). This indicates that the perturbation in the diffusion flux by the adjacent inert regions (sinks) affects the vacancy excess close to the nitriding interface. We can infer from the previous simulations that the balance of the generation/recombination terms for vacancies at the nitriding interface is much less significant than the analogous process during oxidation. Intuitively, this makes some sense. There are experiments showing that a nitride interface passivates the silicon surface so that it acts as a poor sink for interstitials (Mizuo and Higuchi, 1982b; Griffin and Plummer, 1986); by extension, such a surface is likely to be a poor sink for vacancies too. Thus a small generation rate of vacancies, with little or no recombination at the injecting interface, could build up a large supersaturation. The provision of an additional surface sink for vacancies could then influence the supersaturation at the center of a narrow stripe. This gives rise to the geometry dependence observed in the nitridation experiments.

When one compares the cases of oxidation- and nitridation-enhanced diffusion, the problem of proposing a general point-defect model for injecting surfaces becomes apparent. A common feature of the different oxidation models is that the interstitial generation rate is some function of the oxidation rate. During oxynitridation, any oxidation of the silicon surface that takes place is orders of magnitude smaller than for thermal oxidation, yet the interstitial supersaturation in both is comparable. Moreover, there are different degrees of vacancy undersaturation in both cases, with less vacancy undersa-

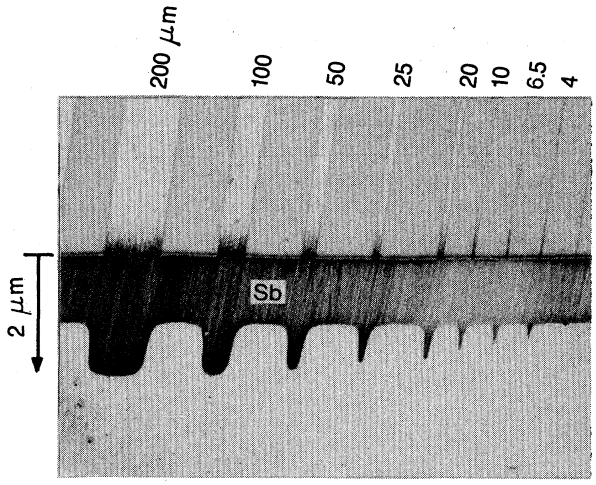

FIG. 48. Antimony diffusion after 1 h at 1100°C in an ammonia ambient. In contrast to oxidation (Fig. 47), nitridation of the silicon surface gives rise to junction depths that depend strongly on the width of the nitriding stripe.

turation during oxynitridation [compare Figs. 37 and 43(b)]. This suggests that the surface is resupplying vacancies faster during oxynitridation than during oxidation. For the case of thermal nitridation, it is clear that any vacancy injection is not correlated with the nitride growth rate. The point-defect injection and resulting supersaturation that is observed is the net result of competing processes, and these processes differ significantly between oxidation and nitridation.

Experimentally measured quantities involve the balance of different flux components whose individual time dependencies are often not experimentally observable. The advantage of interpreting the proposed models in terms of $\mathcal{G}$, $\mathcal{R}$, and $\mathcal{D}$ fluxes is now clear. Some of the fundamental relationships between the individual flux components can be determined from experiments. Beyond that, one's choice of a particular model is a matter of preference, unless specific predictions of the model can be verified. A general model for point-defect behavior, which adequately accounts for the different surface effects, is lacking. Such a model would clarify each individual point-defect injection process and would be a significant contribution to the field.

XVI. DETERMINING THE POINT-DEFECT PARAMETERS

Other experiments have been performed to determine values for the diffusivity d_I and the recombination velocity σ'_I at a nonoxidizing interface. We state at the outset of this discussion that the values obtained are *effective* values and lump the effects of bulk recombination and time-dependent surface recombination into d_I and σ'_I. The values for the diffusivity and surface recombination so obtained seem to be valid over an extensive range of experimental times, but there is some discrepancy between these values and those obtained from gettering experiments.

An important question is how important are the surface effects compared with the bulk effects. The effective values for d_I and σ'_I are useful in answering this question. There is also other evidence that suggests that surface processes are dominant in determining the concentration of point defects in typical silicon wafers. Evidence that surfaces are good sinks for interstitials comes from observations of the vertical versus the lateral extent of OED. Mizuo and Higuchi (1982b) and Griffin and Plummer (1986) determined that interstitials under pad-oxide films, were annihilated faster than under nitride films. Lin, Dutton, and Antoniadis (1979) and Taniguchi and Antoniadis (1985) found short lateral diffusion lengths under pad-oxides. These all suggest that a pad-oxide film is a good sink for interstitials.

Additional support for the dominance of surface processes over bulk processes comes from observations by Hu (1980) on the growth and shrinkage of stacking faults in silicon under different surface coverings. Stacking faults grew faster under nitride films than under pad-oxide films, which would not happen if bulk recombination of interstitials were faster than the parallel process of surface recombination of interstitials. Both nitride and oxide surface coverings do passivate the silicon surface somewhat—it is possible that a bare silicon surface is the best recombination site of all for point defects. Swirl defects, often introduced during crystal growth (Plaskett, 1965; Abe *et al.*, 1966; Föll, Gösele, and Kolbesen, 1981), are not found within 1–2 mm of the crystal surface and cannot be quenched into ordinary silicon wafers. Since swirl defects have been identified as agglomerates of point defects, this suggests that point defects prefer to annihilate at the bare silicon surface rather than agglomerate in the bulk.

A. Gettering and point defects

During the fabrication of silicon integrated circuits, the process of gettering consists of a step or a series of steps to reduce the concentration of metallic contaminants to levels of less than 1 part in 1×10^{11}. Some devices, such as solar cells, are seriously degraded by transition-metal contamination even in these small quantities. Here we shall focus only on the relationship between gettering and point defects, and how the gettering process can be used to obtain estimates of the point-defect parameters.

Gettering treatments often consist of mechanical backside damage, argon implant damage, or extrinsic phosphorus diffusions, among others. Evidence suggests that high concentrations of phosphorus inject interstitials into the bulk (Sec. XIII.C.1). This can be confirmed by observing the effect of an extrinsic phosphorus diffusion on dopant marker layers or on buried stacking faults. Bronner and Plummer (1985) showed that layers damaged by high-dose argon implantation inject point defects for long periods of time. Later experiments by Bronner and Plummer (1987) confirmed that interstitial injection was the dominant process when they observed no antimony enhancement, which would occur during vacancy injection. Both of these surface treatments are also effective gettering techniques. In addition, Lecrosnier *et al.* (1981) showed that the efficiency of gold removal during surface phosphorus diffusions or argon implant damage increased as the temperature decreased. This agrees well with the temperature dependence of the interstitial supersaturation observed under argon implant damage (Bronner and Plummer, 1985) and under phosphorus diffusions (Fahey, Dutton, and Hu, 1984).

Based on the above similarities and the observation that direct metal diffusion to an infinite sink could not account for the time dependence of gettering profiles, Bronner and Plummer (1987) proposed a causal link between interstitial injection and gold gettering. Under equilibrium conditions, the fraction of gold on substitutional sites is much greater than that on interstitial sites. The diffusivity of the substitutional gold is much smaller than that of interstitial gold. Any process that pushes

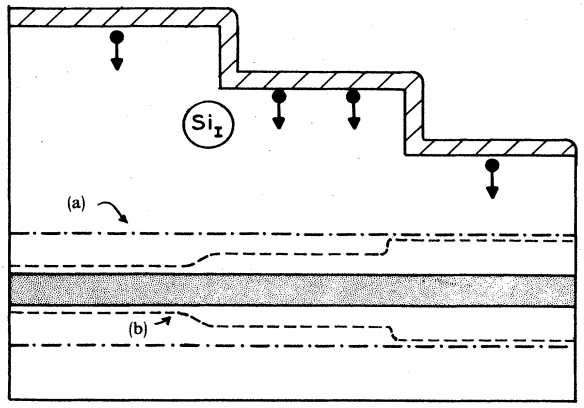

FIG. 49. Conceptual diagram of test structure to measure the silicon interstitial diffusivity. A buried dopant layer under 40 μm of epitaxial silicon was used to monitor the progress of the interstitials. Case (a) shows the expected result for a fast interstitial diffusivity; case (b) shows a depth dependence of the interstitial diffusivity.

gold from a substitutional site to an interstitial site increases the average gold diffusivity and facilitates its removal to a suitable trapping site. This gettering process depends on the kick-out mechanism, which is also thought to operate during gold in-diffusion (Sec. IX).

By analyzing the time dependence of gold gettering profiles during interstitial injection, Bronnner and Plummer estimated the diffusivity of the silicon interstitial as

$$d_I = 600 \exp(-2.44/kT) \text{ cm}^2/\text{sec} , \quad (16.1)$$

giving a value for $d_I \approx 10^{-7}$ cm^{-2}/sec at 1100 °C. Values for the product of $d_I C_I^*$ have been derived from the analysis of gold and platinum in-diffusion in silicon and the solution of the kick-out equations (Sec. IX). There is no sound experimental method to separate these components in a consistent manner, and attempts to do so have relied on many assumptions. Still, Tan and Gösele (1985) determined that d_I should be of order 10^{-7} cm^{-2}/sec at 1100 °C. Morehead (1987) has proposed even higher values for d_I. All the estimates of the interstitial diffusivity from gettering studies are consistent and give relatively high values for d_I.

These values also agree with a value for d_I derived by Griffin, Fahey, Plummer, and Dutton (1985) from the structure shown in Fig. 49. By observing the difference in OED at different depths in an epitaxial layer, they established a lower bound for d_I in agreement with the values derived from the gettering studies.

B. Other experiments to determine d_I

Many experiments to determine a value for the interstitial diffusivity involve using thin membranes of silicon and observing the transport time of the interstitials across the membrane (Mizuo and Higuchi, 1982a, 1983; Taniguchi, Antoniadis, and Matsushita, 1983; Scheid and Chenevier, 1986; Ahn *et al.*, 1987). This transport is not directly observed, but the interstitial behavior can be inferred from the enhanced diffusivity of a dopant marker layer. Ahn *et al.* (1987) confirmed that the interstitial supersaturation at the oxidizing interface was independent of the thickness of the membrane, a point that had not previously been checked. Varying the thickness of the membrane changed only the time required for the point defects to reach the opposite side of the membrane. By monitoring the OED or the growth of stacking faults at different distances from the oxidizing interface, Ahn *et al.* inferred a value for the interstitial diffusivity.

Values for the surface recombination velocity (assumed constant) of the interstitial have also been obtained from the membrane structures mentioned above, by monitoring the magnitude of the enhancement at the backside compared to that at the oxidizing interface. The agreement between Taniguchi *et al.* (1983) and Ahn *et al.* (1987) is within a factor of 3 for both d_I and σ_I', which is remarkable, considering that one used stacking faults to monitor the interstitial transport and the other used dopant marker layers.

The agreement obtained above extends to other independent measures of the point-defect parameters. Other experiments to determine d_I and σ_I' involve two-dimensional test structures (Lin, Dutton, and Antoniadis, 1979; Taniguchi and Antoniadis, 1985; Griffin and Plummer, 1986). In these cases, wide oxidizing regions are separated by progressively narrower inert regions, as in Fig. 50. The extent of the "spillover" of the interstitials

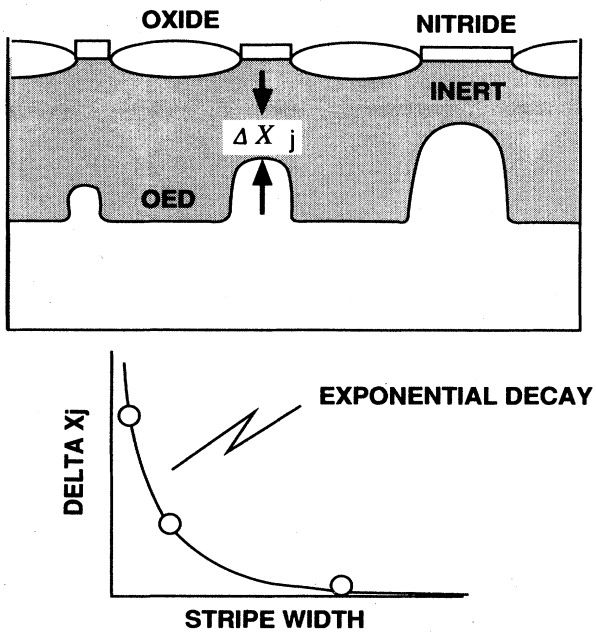

FIG. 50. A two-dimensional test structure to determine the diffusivity and surface recombination velocity of the point defects. The "spillover" of point defects from the injecting regions causes an enhanced diffusion under the inert regions, which depends on the inert stripe width.

from the oxidizing regions to the inert regions as the stripes get narrower can be monitored by the OED at the center of a normally inert stripe. Shin and Kim (1984) modeled the results of Lin *et al.* (1979) and extracted a value for the surface recombination velocity of the interstitials. Their method relied on the assumption that the interstitial diffusivity was so fast that a steady-state distribution of the point defects was present. Although the interstitial diffusion is fast compared to the dopant diffusion, the assumption of a steady-state distribution is not correct. Taniguchi and Antoniadis (1985) applied their results from the thin membrane experiments to the two-dimensional test structures and showed that they were consistent with the two-dimensional profiles. Griffin and Plummer (1986) investigated a longer time range and obtained values for the diffusivity and surface recombination velocity directly from the two-dimensional data. The transient in the lateral OED depended on the ratio of $d_I/\sigma_I'^2$ and the final steady-state value depended on the ratio d_I/σ_I'. By monitoring both the initial transient and the final steady-state behavior, Griffin and Plummer were able to obtain unique values for the two parameters.

These different investigations all give good agreement for the parameters, which is significant considering the differences in the experimental techniques used. Stacking fault growth, diffusion enhancements in thin membranes, and two-dimensional effects were also used to monitor the interstitial excess and to extract the relevant parameters. The one assumption common to all these investigations was that the recombination velocity at the inert interface was independent of time. If it were a function of time, one could imagine that the interstitial diffusivity might be very fast, but that any buildup of interstitials at the inert surface would be determined by the surface recombination kinetics. This possibility has not been carefully checked experimentally. Scheid and Chenevier (1986) have commented on the possibility that the integral of the surface loss might, over time, cause the interface reaction to saturate. Calculating the number of silicon interstitials that might annihilate at the interface requires a knowledge of the equilibrium concentration of interstitials, which is not well known.

While time-dependent surface recombination seems like a plausible resolution of the fast interstitial diffusivity observed in some experiments, we must point out that there is a significant bulk effect on interstitial diffusion. For example, Ahn *et al.* (1987) observed a factor-of-two difference in d_I between CZ and floating zone (FZ) silicon material. Mizuo and Higuchi (1982a) observed enhancements in dopant diffusivity far from an oxidizing surface in FZ silicon but saw no measurable enhancement in CZ material, even for times as long as 1000 min. A model that accounts for bulk effects on the interstitial diffusivity and reconciles the different measurements of d_I in the literature has been proposed (Griffin *et al.*, 1987). It assumes that interstitials recombine with bulk traps via the reaction

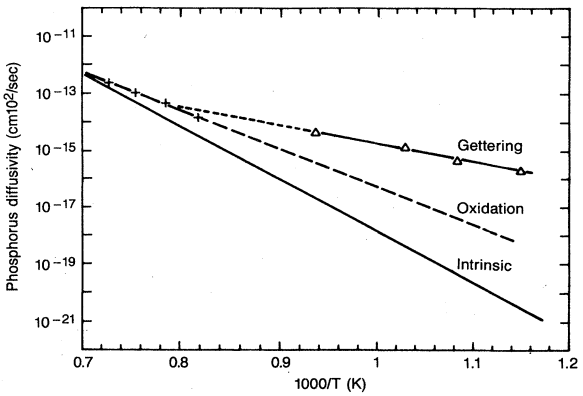

FIG. 51. Interstitial supersaturation during different surface treatments. The solid line is the intrinsic phosphorus diffusivity. The enhancement in P diffusivity from gettering (Bronner and Plummer, 1987) and from oxidation [extrapolated from Taniguchi *et al.* (1980)] is taken as a measure of the interstitial excess.

$$I + T \rightleftharpoons IT \ . \tag{16.2}$$

The presence of the traps slows down the diffusion front of the interstitials (Sec. XII.E). By assuming that silicon wafers grown by different methods contain different numbers of traps, one can explain the differences in diffusivity between FZ, CZ, and epitaxial silicon. The high apparent diffusivity observed during gettering in all materials occurs because the high interstitial excess in wafers during gettering (Fig. 51) quickly annihilates the traps, and the interstitials move freely through the bulk, as shown in Fig. 28.

Until further experimental results are available, we cannot say whether a time-dependent surface recombination in an inert region is necessary to model the interstitial kinetics. Bulk effects on the interstitial diffusivity must be included in any model, and the diffusion in Sec. XII.E shows that such a model is sufficient to account for the present experimental results.

Overall, the conceptual picture in Fig. 38 seems to provide a very useful way of understanding dopant diffusion processes and point-defect behavior in two dimensions. Enough quantitative values for the parameters in the model are now available that meaningful two-dimensional simulations can be performed in programs like SUPREM-IV. Substantial future work is required to further define and modify this picture, but the model presented here provides a background for this work.

XVII. FRACTIONAL INTERSTITIAL COMPONENT OF DIFFUSION, f_{AI}

Perhaps the most controversial single issue related to diffusion in silicon has been the identification of which mechanism is responsible for dopant diffusion: interstitial-type or vacancy-type. In recent years there

has been a growing consensus that both I and V mechanisms contribute to dopant diffusion (see Table II). There seems to be agreement for the common dopants Sb, As, P, and B that

$$f_{SbI} < f_{AsI} < f_{PI} \simeq f_{BI} , \tag{17.1}$$

where f_{AI} is the fractional interstitial component of diffusion defined in Sec. XIII. This ordering is presumed to hold under intrinsic doping conditions for temperatures where diffusion of profiles can be observed (typically, $T \geq 900\,^{\circ}\mathrm{C}$). In addition, although not as well studied, the p-type dopants from column III (Al, Ga, and In) appear to have f_{AI} values comparable to B and P (Mizuo and Higuchi, 1982c, 1982d; Antoniadis and Moskowitz, 1982b).

Most researchers also seem to agree that Sb diffusion is dominated by a vacancy mechanism. This general agreement followed very quickly after experimental determinations that Sb diffusion is retarded in the presence of excess I [Figs. 37 and 43(b)]. There is still sharp disagreement, however, over the relative importance of the interstitial mechanism in effecting P and B diffusion.

The results of the nitridation studies presented in Sec. XIV.F indicate that the diffusivities of P and B are retarded in the presence of excess V; this strongly suggests that P and B diffusion are dominated by an interstitial mechanism. Although the magnitude of retarded diffusion for P and B during V injection is much more pronounced than the retarded diffusion of Sb during I injection, it seems to be readily accepted that at least 98% of the diffusivity of Sb can be attributed to a vacancy mechanism, whereas some researchers are resistant to the idea that even 50% of the diffusivity of P and B could be attributed to an interstitial mechanism.

A. Methods for determining f_{AI}

Three methods have generally been used to estimate the fractional interstitial component of diffusion, all of which assume a dual mechanism of dopant diffusion: (1) self-consistently estimating f_{AI} by observing different dopant diffusion behavior under identical diffusion conditions; (2) assuming that $C_I C_V = C_I^* C_V^*$ and estimating f_{AI} from observations of retarded diffusion; (3) estimating the interstitial supersaturation from stacking fault growth kinetics and comparing with enhancements in dopant diffusion to find f_{AI}.

To determine the fractional I component of diffusion from nonequilibrium diffusion data, we note that Eq. (13.6) can be rearranged to yield the relationship between f_{AI} and the other diffusion parameters:

$$f_{AI} = \frac{\langle D_A \rangle / D_A^* - \langle C_{AV} \rangle / C_{AV}^*}{\langle C_{AI} \rangle / C_{AI}^* - \langle C_{AV} \rangle / C_{AV}^*} . \tag{17.2}$$

This rearrangement simply indicates that if we could measure all of the quantities on the right-hand side of the equation experimentally, then we could determine f_{AI} directly. In reality we are only able to measure $\langle D_A \rangle / D_A^*$ experimentally.

If we define the extent of retardation as

$$\Delta_R = \langle D_A \rangle / D_A^* < 1 , \tag{17.3}$$

then a dopant retarded by vacancy injection ($\langle C_{AV} \rangle / C_{AV}^* > 1, \langle C_{AI} \rangle / C_{AI}^* \leq 1$) allows an *absolute lower bound* to be placed on f_{AI} by putting $\langle C_{AI} \rangle / C_{AI}^* = 0$ (it must, in fact, be > 0) and $\langle C_{AV} \rangle / C_{AV}^* = 1$ (it must, in fact, be > 1), which gives

$$f_{AI} \gg 1 - \Delta_R . \tag{17.4}$$

A better approximation to f_{AI} can be made if one observes some other dopant B enhanced under the same conditions under which dopant A is retarded. In that case the vacancy supersaturation must be at least as great as the enhancement in the dopant diffusivity observed, and we can define the enhancement factor to be

$$\Delta_E = \langle D_B \rangle / D_B^* < \langle C_{BV} \rangle / C_{BV}^* . \tag{17.5}$$

Substituting into Eq. (17.2) gives a more realistic bound,

$$f_{AI} \gtrsim 1 - \frac{\Delta_R}{\Delta_E} . \tag{17.6}$$

This kind of analysis was used (Fahey 1985; Fahey, Barbuscia, Moslehi, and Dutton, 1985) to impose upper and lower limits on f_{AI} for P, As, and Sb (Sec. XVII.A).

Gösele and Tan (1983) determined f_{AI} from observations of retarded diffusion during high-temperature oxidation of ⟨111⟩ silicon (see Sec. XIV.A), from the retarded diffusion under deposited Si_3N_4 films (Mizuo and Higuchi, 1982b), and from the restrained diffusion under HCl-added ambients (Nabeta, Uno, Kubo, and Tsukamoto, 1976). The method of analysis relied on Eq. (17.2), the extent of retardation, and the assumption that the steady-state relationship implied by Eq. (12.4) between vacancies and interstitials holds. Using Eq. (12.4), we find that

$$\Delta_R = f_{AI} \langle C_{AI} \rangle / C_{AI}^* + f_{AV} C_{AI}^* / \langle C_{AI} \rangle . \tag{17.7}$$

This has a minimum value when

$$\langle C_{AI} \rangle / C_{AI}^* = \sqrt{f_{AV} / f_{AI}} , \tag{17.8}$$

which gives a minimum possible value of Δ_R for a given f_{AI} of

$$\Delta_R^{\min} = 2\sqrt{f_{AI} f_{AV}} = 2\sqrt{f_{AI}(1 - f_{AI})} . \tag{17.9}$$

The experimental value of the extent of retardation, which is not necessarily the minimum value, allows a bound to be placed on f_{AI},

$$f_{AI} \geq 0.5 + 0.5(1 - \Delta_R^2)^{1/2} . \tag{17.10}$$

The results obtained using this equation indicated a dominant interstitial diffusion mechanism for P, Ga, Al, and B, with the values of f_{AI}^P close to unity. This technique does not apply to As and Ge in experiments performed

thus far, since these dopants are not retarded.

In general, the literature gives very different values for f_{AI} depending on the method of measurement used (Tan and Gösele, 1985). When stacking fault growth is used to monitor the excess interstitial concentration, inferences about the value of f_{AI} have been made by comparing the enhancement in dopant diffusivity with the calculated enhancement in the interstitial concentration.

Hu (1974) first proposed that the interstitial excess was related to the stacking fault (SF) growth rate by a simple equation of the form

$$\frac{dR_{\rm SF}}{dt}=\pi a_0^2 d_I(C_I-C_I^*)\ , \tag{17.11}$$

where $a_0\approx 3.85$ Å is an estimate of the capture radius of bounding dislocation. In a later paper (Hu, 1981), a more rigorous derivation of the factors involved in stacking fault growth gave the equation

$$\frac{dR_{\rm SF}}{dt}=(\pi r_c^2/r_a^2 bC_s)\exp\left[\frac{-\Delta H}{kT}\right]\left\{d_I(C_I-C_I^*)-d_I C_I^*\left[\exp\left[\frac{\gamma}{C_s bkT}\right]-1\right]\right\}, \tag{17.12}$$

where γ is the energy per cm^2 in the stacking fault, r_c is the core radius of the dislocation, r_a is the interatomic spacing in the lattice, b is Burger's vector, and ΔH is the difference between the reaction energy barrier and the activation energy of interstitial migration, assumed to be $\geq 4kT$ for reaction-controlled growth of the stacking fault. In equilibrium ($C_I=C_I^*$), values for the shrinkage rate of stacking faults based on this equation are 60 nm/sec compared with experimental values of 98 nm/sec [from Hu (1974)]. This assumes a value for ΔH of 0.4 eV and assumes that self-diffusion occurs completely by an interstitial mechanism.

Leroy (1979, 1982) has proposed that the growth of stacking faults is diffusion controlled rather than reaction controlled. To explain the independence of stacking fault size from the spacing between stacking faults, even for closely spaced faults, Leroy determined that the "capture radius" of interstitials is small compared with the distance between faults, giving an equation of the form

$$\frac{dR_{\rm SF}}{dt}=\frac{2\pi d_I}{N_0}\frac{1}{\ln(r_0/b)}(C_I-C_I^*)-K_n\ , \tag{17.13}$$

where N_0 is the surface density of atoms in the stacking fault ($N_0=1.57\times 10^{15}$ cm^2), and K_n accounts for the shrinkage rate of the stacking faults in an inert ambience. The assumptions in Leroy's model have been disputed by Hu (1981) and Antoniadis (1982). In any case, both models give the same basic formulation for the shrinkage and growth of the stacking faults.

Tan and Gösele (1981, 1982b) presented both an interstitial-only form for the stacking fault growth equation,

$$\frac{\Omega}{\alpha A}\frac{dR_{\rm SF}}{dt}=-D_{\rm self}\frac{-\gamma}{kT}+D_{\rm self}S_I\ , \tag{17.14}$$

and a general form of the equation,

$$\frac{\Omega}{\alpha A}\frac{dR_{\rm SF}}{dt}=-D_{\rm self}\frac{-\gamma}{kT}+d_I C_I^* S_I-d_V C_V^* S_V\ , \tag{17.15}$$

where

$$D_{\rm self}=d_I\frac{C_I^*}{C_s}+d_V\frac{C_V^*}{C_s}\ , \tag{17.16}$$

$$S_I=\frac{C_I-C_I^*}{C_I^*}\ , \tag{17.17}$$

$$S_V=\frac{C_V-C_V^*}{C_V^*}\ , \tag{17.18}$$

$$\alpha\approx 2\ . \tag{17.19}$$

In these equations, S_I and S_V are the normalized interstitial and vacancy supersaturation, $D_{\rm self}$ is the self-diffusion coefficient of silicon, A is the area per atom in the stacking fault (6.38×10^{-16} cm^2), Ω is the atomic volume (2×10^{-23} cm^3), γ is the stacking fault energy (0.026 eV/atom), $R_{\rm SF}$ is the stacking fault radius, and α is a constant that accounts for geometrical effects and a possible reaction energy barrier.

In their analysis, as well as in most other analyses of stacking fault growth, an equation of the form of Eq. (17.14) has been used to estimate the I concentration (Fair, 1981b; Hu, 1981; Lin *et al.*, 1981; Hayafuji, Kajiwara, and Usui, 1982; Tan and Gösele, 1982a, 1982b; Antoniadis, 1983; Matsumoto, Ishikawa, and Niimi, 1983; Leroy, 1987; Chichibu, Harada, and Matsumoto, 1988). Sources of error include the arbitrary assignment of a value for α and the assumption that $D_{\rm self}$ in Eq. (17.14) can replace $d_I C_I^*$ in the general equation. If one assumes that self-diffusion occurs predominantly by an interstitial mechanism, the measured value for $D_{\rm self}$ can replace $d_I C_I^*$, but this is certainly a debatable point. Neglecting vacancies also introduces some error in the calculation of S_I, since based on nonequilibrium studies S_I may be of order unity at the high temperatures where stacking fault growth is measured. Experiments based on stacking fault growth consistently lead to values for f_{AI} for dopants that are lower by approximately a factor of 2 than those derived from nonequilibrium diffusion experiments [see Antoniadis (1982); Tan and Gösele (1985)].

An attempt to include both interstitials and vacancies in the problem using Eq. (17.15) and to account self-consistently for dopant diffusion and stacking fault growth was carried out by Yoshida, Matsumoto, and Ishikawa (1986) and Yoshida (1988). A simultaneous solution of the equations involved concluded that self-

diffusion occurred primarily with vacancies, but the results on f_{AI} for dopants were consistent with those from nonequilibrium experiments (Sec. XVII.B). However, a value of $\alpha=4$ was used, compared with a value of $\alpha=2$ suggested by Tan and Gösele (1985).

Our conclusion is that the model for stacking fault growth is simply not good enough to form an independent estimate of the I supersaturation. Quantities of order unity are being calculated (S_I at 1100 °C) while the constants in the model are simply not known to sufficient accuracy. The self-consistent approach mentioned above seems to be best in this case, and the value of S_I should be included as an adjustable parameter in the equations. The trend in the determination of f_{AI} with temperature does indicate that the vacancy component of diffusion becomes more important at lower temperatures (Matsumoto, Ishikawa, and Niimi, 1983). This agrees in spirit with recent results from Hu (1987), which have provided evidence that enhanced diffusion of B during V injection can occur under experimental conditions different than those in the nitridation studies; see also the article by Fahey and Dutton (1988).

Other estimates of f_{AI} have been presented by Mathiot and Pfister (1985) based in part on the detailed solution of a set of defect-dopant coupled differential equations for extrinsic phosphorus diffusion (Mathiot and Pfister, 1984). The calculations suggested that $f_{AI}^{P}\approx 3f_I^{SD}$. By assuming that self-diffusion is governed primarily by an interstitial-type mechanism, this suggests that phosphorus diffuses predominantly with vacancies. However, there is no general agreement on the mechanism of self-diffusion at 1000 °C, as discussed in Sec. IX. In addition, Sec. XIII. C indicates that extrinsic phosphorus diffusion is still poorly understood, and small differences in model assumptions can seriously affect these calculations.

Because of the importance of resolving this issue, we examine in detail an attempt to determine f_{AI} from nonequilibrium studies, based on the data from the nitridation experiments presented in Sec. XIV. F. We also examine the expected behavior of f_{AI} with temperature, the physical significance of different f_{AI}'s among the different dopants, and how f_{AI} may change under extrinsic doping conditions.

B. Experimental determination of f_{AI}

It is important to recognize in the following estimation process that this analysis does not rely on any model assumptions about how the point-defect excesses come about, such as whether the excess is due to I depletion or V injection. Not does it rely on any assumptions concerning relationships between I and V. In addition, for the data analyzed, great care was taken in sample preparation to ensure that results were reproducible and that dopant diffusivities under inert surface conditions were constant with time. The following method of estimating fractional components of vacancy-type and interstitial-type diffusions is believed to be more reliable than any previous estimations, specifically those made based on oxidation studies.

1. P diffusion

Before showing in detail how limits on f_{PI} are obtained from nonequilibrium studies, we make the following simple observations. Since P diffusivity is shown to be reduced to 30% of its equilibrium value in the presence of a V excess at 1100 °C, one can conclude that at least 70% of P diffusivity under equilibrium can be attributed to a nonvacancy mechanism. The treatment that follows demonstrates that f_{PI} must be approximately equal to unity at 1100 °C.

Using the direct nitridation data presented in Fig. 43(a), we first estimate a lower limit for f_{PI}. Using Eq. (17.2) we have

$$f_{PI}=\frac{\langle D_P\rangle/D_P^*-\langle C_{PV}\rangle/C_{PV}^*}{\langle C_{PI}\rangle/C_{PI}^*-\langle C_{PV}\rangle/C_{PV}^*}\ . \tag{17.20}$$

To determine an absolute lower bound for f_{PI} we assume $\langle C_{PI}/C_{PI}^*\rangle\approx 0$ (thus minimizing the I contribution to P diffusion during direct nitridation) and $\langle C_{PV}\rangle/C_{PV}^*\approx 1$ (i.e., no enhancement of the V component of P diffusion during direct nitridation). With these approximations for $\langle C_{PV}\rangle/C_{PV}^*$ and $\langle C_{PI}\rangle/C_{PI}^*$ and using the values of $\langle D_P\rangle/D_P^*$ in Table III, one finds that $f_{PI}>0.7$ at 1100 °C. But of course, if it were really the case that $\langle C_{PV}\rangle/C_{PV}^*$ remained at unity after several hours of vacancy injection, one would suspect that this apparent inability of the P to readily form PV pairs in the presence of a vacancy excess was striking evidence that P must have an extremely small V component of diffusion under equilibrium conditions, so that $f_{PI}\approx 1$. Furthermore, it is extremely unlikely that $\langle C_{PI}\rangle/C_{PI}^*$ actually attains a value of 0. If one allows this ratio to grow from 0 to something approaching the values of $\langle D_P\rangle/D_P^*$ (e.g., $\langle D_P\rangle/D_P^*\approx 0.25$ after 1 h of direct nitridation), even maintaining that $\langle C_{PV}\rangle/C_{PV}^*\approx 1$, Eq. (17.20)

TABLE III. Diffusion effects during nitridation at 1100 °C.

Time (min)	$\langle D_p\rangle/D_p^*$	$\langle D_{As}\rangle/D_{As}^*$	$\langle D_{Sb}\rangle/D_{Sb}^*$
	Oxynitridation		
7	5.70		1.00
15	6.00	2.5	2.33
60	4.5	1.2	1.00
120	3.60	1.02	0.83
275	3.30	1.02	0.72
	Nitridation		
7	0.79		2.63
15	0.39	1.90	5.00
60	0.25	2.50	4.24
120	0.30	2.60	4.66
275	0.32	2.56	4.29

will yield a value of f_{PI} close to one.

A more reasonable estimate of $\langle C_{PV}\rangle/C^*_{PV}$ can be made by using the data for Sb diffusion, which show large enhancements during V injection. We can certainly state that

$$\frac{\langle C_{SbV}\rangle}{C^*_{SbV}} \geq \frac{\langle D_{Sb}\rangle}{D^*_{Sb}} . \tag{17.21}$$

If Sb diffused entirely by a vacancy mechanism, the equality would be satisfied. However, the fact that Sb diffusivity does show some small enhancement during I injection before retarded diffusion is observed (Antoniadis and Moskowitz, 1982a; Fahey, 1984) demonstrates that there is some I component of diffusion for Sb. Therefore we are sure that

$$\frac{\langle C_{SbV}\rangle}{C^*_{SbV}} > \frac{\langle D_{Sb}\rangle}{D^*_{Sb}} \tag{17.22}$$

is satisfied. The question then is how $\langle C_{SbV}\rangle/C^*_{SbV}$ is related to $\langle C_{PV}\rangle/C^*_{PV}$. This question can be addressed by utilizing the results in Sec. XII for the case of I injection with trapping. Results in that section are applicable to the present discussion of V injection, where now the dopant atoms act as traps for the excess vacancies. In the case of dopant atoms, the number of AV pairs is sure to be much less than the number of isolated A atoms. Under this condition, it was shown in Sec. XII.E.1 that

$$\frac{\langle C_{AV}\rangle}{C^*_{AV}} = \frac{\langle C_V\rangle}{C^*_V} . \tag{17.23}$$

Therefore we expect that for the direct nitridation data in Table III we can assume

$$\frac{\langle C_V\rangle}{C^*_V} \approx \frac{\langle C_{SbV}\rangle}{C^*_{SbV}} \approx \frac{\langle C_{PV}\rangle}{C^*_{PV}} \approx \frac{\langle C_{AsV}\rangle}{C^*_{AsV}} . \tag{17.24}$$

We shall scrutinize this assumption further in the examination of f_{AsI} below.

As an example, we use these approximations with the data in Fig. 43(a) for 1 h of direct nitridation:

$$\begin{aligned} &\langle D_P\rangle/D^*_P = 0.25 , \\ &\langle D_{Sb}\rangle/D^*_{Sb} = 4.24 < \langle D_{PV}\rangle/D^*_{PV} , \\ &\langle D_{PI}\rangle/D^*_{PI} \approx 0.0 . \end{aligned} \tag{17.25}$$

Inserting these values into Eq. (17.20) gives the result

$$f_{PI} > 0.94 . \tag{17.26}$$

Using the data for 120 and 240 min with the same approximations gives $f_{PI} > 0.94$ and 0.93, respectively (the data for times shorter than 60 min are considered less accurate and only give less restrictive estimates for f_{PI}). These lower estimates are very strong evidence that $f_{PI} \approx 1$ at 1100 °C.

It is very important to note the following. Mathematically, using Eq. (17.20), the only way in which one may try to lower f_{PI} closer to the absolute lower bound of 0.7 is to propose some model whereby during vacancy injection from a nitriding silicon surface, $\langle C_{PV}\rangle/C^*_{PV} < \langle C_{SbV}\rangle/C^*_{SbV}$ in the bulk results. But as we mentioned above, if this is due to the inability of P to readily form PV pairs after some hours in the presence of a vacancy excess, then this is only further evidence that $f_{PI} = 1$.

It can be concluded that, based on the experimental evidence provided by nitridation studies, $f_{PI} \approx 1$ at 1100 °C. This conclusion follows basically from the experimental fact that P diffusion shows a large reduction of diffusivity in the presence of a vacancy excess. It is not possible to explain this phenomenon without a fractional I component of diffusion for P close to one.

2. Sb diffusion

The same estimation process as presented in detail for the case of P can be applied to Sb. In this case one can use the results for nitridation of SiO_2 shown in Fig. 43(b). For Sb we write

$$f_{SbI} = \frac{\langle D_{Sb}\rangle/D^*_{Sb} - \langle C_{SbV}\rangle/C^*_{SbV}}{\langle C_{SbI}\rangle/C^*_{SbI} - \langle C_{SbV}\rangle/C^*_{SbV}} . \tag{17.27}$$

Proceeding in a manner similar to that for P discussed above, we let $\langle C_{SbV}\rangle/C^*_{SbV} = 0$ and assume $\langle C_{SbI}\rangle/C^*_{SbI} = \langle C_{PI}\rangle/C^*_{PI} \geq \langle D_P\rangle/D^*_P$. We then find that

$$f_{SbI} < 0.22 . \tag{17.28}$$

This is consistent with published estimates based on oxidation work. The results using the nitridation data, however, are arrived at using far fewer model assumptions than previous works and thus can be considered a very reliable upper estimate for f_{SbI}.

3. As diffusion

The behavior of As diffusion in Fig. 43 is very interesting. Much less enhancement of As diffusion occurs in the presence of an I supersaturation than occurs for P in Fig. 43(b), so we can conclude that f_{AsI} must be much less than f_{PI}. On the other hand, because the diffusion enhancement of As is definitely less than that seen for Sb diffusion in the presence of a V supersaturation, we can conclude that As must have a smaller V component of diffusion than Sb. However, there is an inconsistency between the two sets of data. To see this, one can try to obtain an upper limit for f_{AsI} from the oxynitridation data and a lower limit for f_{AsI} from the data of direct nitridation.

To determine an upper limit for f_{AsI} from the I supersaturation data of oxynitridation, we write

$$f_{AsI} = \frac{\langle D_{As}\rangle/D^*_{As} - \langle C_{AsV}\rangle/C^*_{AsV}}{\langle C_{AsI}\rangle/C^*_{AsI} - \langle C_{AsV}\rangle/C^*_{AsV}} . \tag{17.29}$$

We first assume that the V component of As diffusion is reduced to 0 because of significant I-V recombination (i.e., $\langle C_{AsV}\rangle/C^*_{AsV}\approx 0$), thus minimizing the amount of enhancement seen in the presence of excess I. Then, assuming

$$\langle C_{AsI}\rangle/C^*_{AsI}=\langle C_{PI}\rangle/C^*_{PI}\geq\langle D_P\rangle/D^*_P\ , \quad (17.30)$$

we obtain the maximum possible value of f_{AsI} consistent with the fact that a much larger increase in diffusivity is found for P than for As during I supersaturation. For example, at 120 min we have $\langle D_{As}\rangle/D^*_{As}=1.02$ and $\langle D_P\rangle/D^*_P=3.60$, which means that

$$f_{AsI}<0.28\ . \quad (17.31)$$

To determine a lower bound for f_{AsI} from the direct nitridation data, we first assume that the I component of As diffusion is reduced to 0 (i.e., $\langle C_{AsI}\rangle/C^*_{AsI}\approx 0$) due to significant I-V recombination. This allows for the minimum amount of diffusion enhancement expected in the presence of a V excess. Then assuming that

$$\langle D_{AsV}\rangle/D^*_{AsV}=\langle D_{SbV}\rangle/D^*_{SbV}\geq\langle D_{Sb}\rangle/D^*_{Sb} \quad (17.32)$$

we arrive at the minimum value of f_{AsI} (or, alternatively, the maximum value of f_{AsV}) consistent with the fact that a larger diffusion enhancement is seen for Sb than for As during V supersaturation. For example, at 120 min $\langle D_{As}\rangle/D^*_{As}=2.60$ and $\langle D_{Sb}\rangle/D^*_{Sb}=4.66$, which means that

$$f_{AsI}>0.44\ . \quad (17.33)$$

But this is where we note an inconsistency. The oxynitridation data indicate that f_{AsI} must be less than about 0.3, while the direct nitridation data indicate that f_{AsI} must be greater than about 0.4. In addition, these conclusions were reached with the assumptions that $\langle C_{AsV}\rangle/C^*_{AsV}=0$ during oxynitridation and $\langle C_{AsI}\rangle/C^*_{AsI}=0$ during direct nitridation. If either of these (unlikely) restrictions is lifted, then, when the above estimations are repeated, the discrepancy between the two sets of data in Fig. 43 will only grow. This is a rather puzzling result, since all of the assumptions made in estimating f_{AsI} appear quite reasonable.

4. Concerning the inconsistencies found in calculating f_{AsI}

The fact that the diffusion behavior of As under vacancy injection and interstitial injection cannot be reconciled using Eq. (13.6) is a significant observation. It is also somewhat disappointing, considering the reasonableness of Eq. (13.6). In fact, Eq. (13.6) merely states that the changes in dopant diffusivities from equilibrium conditions reflect the fact that the number of dopant atoms in a diffusing state at any given instant of time (i.e., number of A in AI or AV states) has changed. The use of the quantity f_{AI} is only a convenient shorthand.

Mathematically, if we accept that Eq. (13.6) is correct, the fact that the data in Fig. 43 cannot be modeled by Eq. (13.6) using a unique value of f_{AsI} requires that when comparing the different diffusion behaviors of the various dopants during nitridation, for the same temperature and time, we cannot assume that the relationships

$$\frac{\langle C_V\rangle}{C^*_V}=\frac{\langle C_{SbV}\rangle}{C^*_{SbV}}=\frac{\langle C_{AsV}\rangle}{C^*_{AsV}}=\frac{\langle C_{PV}\rangle}{C^*_{PV}}\ ,$$
$$\frac{\langle C_I\rangle}{C^*_I}=\frac{\langle C_{SbI}\rangle}{C^*_{SbI}}=\frac{\langle C_{AsI}\rangle}{C^*_{AsI}}=\frac{\langle C_{PI}\rangle}{C^*_{PI}}\ , \quad (17.34)$$

are satisfied (see Sec. XIII.A). That is, the supersaturation or undersaturation of mobile dopant-defect species, which results from point defects' being injected from the surface, depends on the nature of the dopant atom. If this is true, we need both a physical explanation for why this might be so and an explanation that removes the discrepancy between the As diffusion data under nitriding and oxynitriding conditions.

One possible condition under which the relations in Eq. (17.34) may not be satisfied is the situation in which the recombination rates between injected point defects and dopant defects is not the same as that between injected point defects and native point defects. For example, one could imagine that the recombination rate between I and V is not the same as that between I and AV pairs. Thus, for a given injection level of I, it is possible that the resulting undersaturation of AV pairs could depend on the chemical identity of A. The exact conditions under which this situation might occur are discussed in more detail in Sec. XIII.E.2. Here we note only that this phenomena cannot resolve the problem of the apparent discrepancies found for the As data. In the analysis performed in Sec. XVII.A.3, we made a trial assumption that recombination between injected point defects and dopant defects was so complete that $C_{AV}\approx 0$ when I are injected and $C_{AI}\approx 0$ when V are injected. Even under this assumption we found an inconsistency when obtaining upper and lower limits for f_{AsI}. This inconsistency only grows if we accept the fact that the undersaturation of dopant defects cannot be so complete, so that there is nothing to be gained by considering potential differences in undersaturation levels of dopant defects for the different dopants.

What could explain the inconsistencies of As diffusion in the nitridation study is a plausible physical model for a situation in which the supersaturation of AV defects during V injection, or the supersaturation of AI defects during I injection, may depend on the chemical identity of A. We already rejected the idea of a trapping model in Sec. XIII.A. In Appendix D we show that it is possible $\langle C_{AsI}\rangle/C^*_{AsI}\neq\langle C_I\rangle/C^*_I$ if the nonvacancy component of As diffusion can be attributed to a dissociative mechanism, but that present experimental evidence is against this explanation. It is always possible that some residual damage from ion implantation acts to reduce point-defect supersaturation levels. For the extended anneals at

1100 °C of the nitridation studies, implantation damage is usually observed to be removed, but the injection of point defects may counteract the natural tendency of extended defects to shrink upon high-temperature anneals. This possibility is of course testable by observing whether changes in implantation conditions affect results, or by introducing dopants into the lattice by methods other than ion implantation.

One could always invoke Pandey's concerted exchange mechanism (Pandey, 1986), which does not involve point defects at all. In this way, one could explain the As results by postulating that the I diffusion component of As is small and that exchange mechanism and vacancy mechanism components are comparable. This is an almost too easy way in which to remove the As inconsistencies, because we simply introduce a new (independently untested) variable into the analysis. To be fair, however, if one were trying to devise a test of Pandey's mechanism, the nitridation experiments would be one logical proposal.

The problems encountered in trying self-consistently to explain As diffusion behavior in the nitridation study are nagging ones. It should be noted, however, that the results for P still remain clear. If one performs an analysis on the diffusion data from the nitridation study, allowing for a component of diffusion attributable to the concerted exchange mechanism, it will still result that $f_{PI} \approx 1$. This is the only logical conclusion for a dopant whose diffusivity is enhanced to such a large degree during I injection and retarded so much during V injection.

C. Temperature dependence of f_{AI}

The experiments discussed above attempted to determine f_{AI} at 1100 °C primarily because the diffusivities of the common dopants B, P, As, and Sb can be measured reliably at this temperature. A natural question then is whether we can extrapolate the results found at 1100 °C to other temperatures (this is of technological interest, since the temperatures used during the fabrication sequence of modern silicon integrated circuits are typically about 900 °C–1000 °C.) All published reports thus far have stated that f_{AI} decreases with decreasing temperature. However, as we indicate below, there is actually no experimental evidence for this conclusion from diffusion studies, and even the form of the temperature dependence has not been presented correctly.

We begin by using the definition of f_{AI} in Eq. (13.2), from which it follows that

$$\left[1-\frac{1}{f_{AI}}\right]=\frac{D_{AV}}{D_{AI}}=K_A \exp\left[\frac{-\Delta Q_{AV,I}}{kT}\right], \tag{17.35}$$

where

$$\begin{aligned}\Delta Q_{AV,I} &\equiv Q_{AV}-Q_{AI} \\ &=(H^f_{AV}-H^f_{AI})+(H^m_{AV}-H^m_{AI}) .\end{aligned} \tag{17.36}$$

It can be seen from Eq. (17.35) that f_{AI} should vary with temperature such that the quantity in parentheses on the left-hand side should exhibit an Arrhenius dependence, *not* f_{AI} itself. Although this observation may seem elementary, all works published to date that use the same definition of f_{AI} as Eq. (13.2) have attempted to fit an Arrhenius expression to f_{AI} in order to describe its temperature dependence. Equation (13.2) shows that the activation energy obtained in this way is not a physically meaningful quantity, while for a given dopant the temperature dependence of f_{AI} is determined by the difference in formation plus migration enthalpies between the dopant's I and V components of diffusion.

One way to try to measure the individual activation energies of diffusion for each component is to measure D^*_A over a large temperature range like that shown in the hypothetical example of Fig. 52. However, we can see in this example that even though the activation energy for diffusion by an I-type mechanism is 0.75 eV higher than the activation energy for diffusion by a V-type mechanism, a single activation energy intermediate between the two can fit the data very well over the temperature range 800 °C–1200 °C. This shows that an Arrhenius plot may not easily pick up the two separate components of diffusion unless a wide temperature range is chosen and a very accurate method of measuring D^*_A is employed. This has not been accomplished as yet, and there is no other experimental evidence that indicates whether f_{AI} decreases with decreasing temperature, as has been assumed in the literature. Unfortunately, we cannot answer the question of how to extrapolate the results for f_{AI} at 1100 °C to other temperatures. We can only note

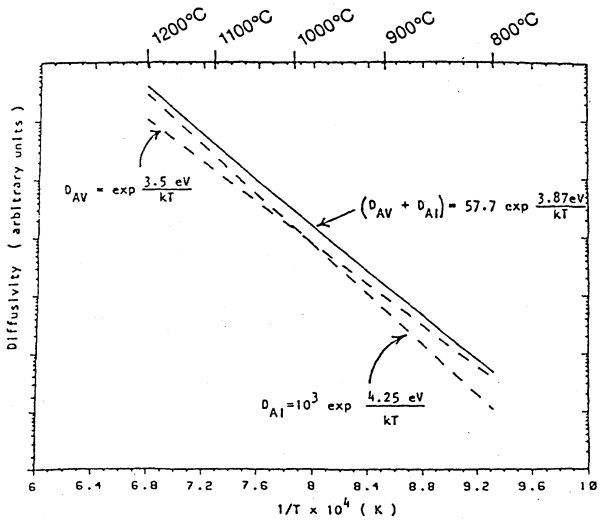

FIG. 52. Hypothetical example of dopant diffusivity. Even though the activation energy of diffusion for the I-type component is 0.75 eV higher than the V-type, an approximate Arrhenius dependence results for the measured diffusivity over a wide temperature range.

that the reduction of P and B diffusivities in the presence of excess V between temperatures of 1010 °C and 1250 °C demonstrates that the fractional V component of diffusion must be less than 0.5 in this temperature range.

D. Dopant dependence of f_{AI}

Why do some dopants diffuse preferentially by a vacancy mechanism and some by an interstitial mechanism? Experimental evidence indicates that all of the p-type dopants from column III of the Periodic Table are dominated by an I-type diffusion mechanism, while n-type dopants of column V are dominated by a vacancy mechanism, with the glaring exception of P. Thus there is no unambiguous trend of f_{AI} with tetrahedral bonding radius or charge state of the dopant. We have already examined in Sec. XI the atomistic processes that determine the apparent diffusivity for each type of mechanism. Combining this analysis with the experimental results, we arrive at the (admittedly inelegant) conclusion that there is no simple answer to our introductory question, and no simple answer is required by first-principle arguments or from incontrovertible experimental evidence.

E. Concentration dependence of f_{AI}

For a given dopant A, it is of interest to determine whether the value of f_{AI} depends on the concentration of A itself. This question was alluded to earlier, in the discussion of nonequilibrium diffusion in Sec. XIII.B. Physically, we can imagine how such a dependence can come about by comparing two uniformly doped silicon crystals. One is doped low enough that the free-carrier concentration is simply the intrinsic carrier concentration n_i. The other is doped high enough that the free-carrier concentration is much greater than n_i. Since we know that the total number of point defects that exist in equilibrium will be different in the two crystals, we can easily accept the fact that the proportion of vacancy defects to interstitial defects will in general be different in the extrinsically doped crystal from that in the intrinsically doped one. In fact, the only way in which the proportions would remain the same would be if both types of point defects were charge neutral (contrary to low-temperature experiments) or if both types of defects had exactly the same donor and acceptor levels with reference to the band edges (an unlikely proposition considering the fundamentally different nature of the vacancy and interstitial point defects). Therefore, since the ratio C_I^*/C_V^* will change with doping level, it is inevitable that under extrinsic doping conditions f_{AI} will be a function of C_A.

Since we always define f_{AI} to be the ratio of $D_{AI}^*/(D_{AI}^*+D_{AV}^*)$, the predicted behavior of f_{AI} on C_A depends on the concentration dependence of D_{AI} and D_{AV} we assume. To show a specific example, suppose we choose a simple model for a donor atom,

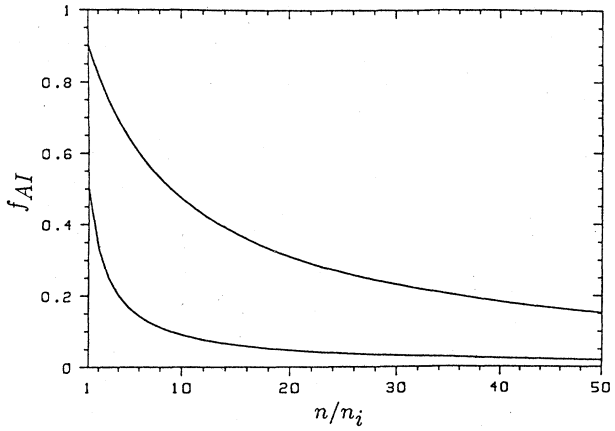

FIG. 53. Hypothetical example of the behavior of f_{AI} under extrinsic conditions. The diffusion model of Eq. (17.37) is chosen.

$$D_A^* = (D_{AI}^*)^i + (D_{AV}^*)^i \frac{n}{n_i} , \tag{17.37}$$

where the superscript i means values found under intrinsic doping conditions. This specific model assumes that dopant diffusion is completely dominated by interactions of the donor dopant A^+ only with I^0 and V^- point defects. In Fig. 53, D_A^* is plotted as a function of n/n_i assuming two different values of f_{AI} under intrinsic conditions: $f_{AI}=0.5$ and 0.9. This plot shows that even if 90% of the diffusive flux of a dopant can be attributed to an I-type mechanism under intrinsic doping conditions, it is possible under extrinsic doping conditions for the dopant diffusion to be dominated by a vacancy mechanism. This of course assumes, as implied by Eq. (17.37), that diffusion of A^+ does not occur by interactions with I^- or V^0 point defects. If interactions of A^+ with point defects in these charge states make a significant contribution to diffusion under intrinsic doping conditions, then the decrease of f_{AI} with increasing values of n/n_i will be less than that pictured in Fig. 53.

XVIII. SUMMARY AND CONCLUSIONS

The subject of point defects and dopant diffusion in silicon covers a wide range of topics. This paper has tried, as much as possible, to present a unified view of diffusion phenomena by pulling together experimental results from a wide variety of sources and discussing them in terms of basic physical models. Rather than present the historical development of this field of study, we have tried to give an idea of what we know now about each related topic. This approach necessarily brings out the limitations of our knowledge as a way of identifying what problems remain to be solved. Although the discussions in the main text may sometimes give the impression that our knowledge in many areas is indeed quite limited, we believe that real progress has been made in the last few

years. If we had attempted to write this same paper five years ago it is clear to us that a much less coherent picture of point defects and diffusion processes in silicon would have resulted.

In the remaining, we summarize our appraisal of the present understanding of point defects and dopant diffusion and show under what experimental circumstances deficiencies become apparent. The need for more information on the basic properties of point defects is emphasized, as well as the recognition of the important role that developing integrated circuit technologies will continue to play in defining new research challenges.

A. The present understanding of diffusion phenomena in silicon

One way in which to judge the present understanding of silicon diffusion processes is by assessing how satisfactory is our present ability to construct physical models of diffusion phenomena that are consistent with experiment. As mentioned in our introductory remarks, we believe that the most important general requirement that future models of point-defect and dopant interactions must meet is that these models be applicable to two- and three-dimensional diffusion problems. The practical importance of this requirement lies in the emerging need to model dopant diffusion in two and three dimensions because of the decreasing device dimensions in silicon integrated circuits. We would like to be able to write a set of governing equations that describes the basic physical processes involved in dopant diffusion and to have experimental values of all necessary coefficients specified by the system of equations. How close are we to achieving this goal, and what work remains to be done? This question is best addressed by discussing diffusion under quasiequilibrium and nonequilibrium conditions separately.

1. Quasiequilibrium conditions

For diffusion under quasiequilibrium conditions (Sec. X and Appendix A), the situation is fairly good. This is basically because under quasiequilibrium conditions the nature of the point defects participating in dopant diffusion, or any of their fundamental properties, do not need to be known explicitly in order to describe the time evolution of diffusing dopant profiles. Thus while the thermodynamics of point defects, their physical and electronic properties, and the details of the atomistic processes that underlie macroscopic diffusion (as outlined in Secs. V–XI) offer a physical basis for diffusion and lead to a self-consistent formulation of the problem, knowledge of the actual experimental values of point-defect parameters is not an essential ingredient in quantitatively modeling diffusion itself. The only areas that are less than satisfactory are the phenomena of dopant pairing (Sec. X.B.5) and heavy doping effects, including clustering, precipitation, and percolation (Sec. VIII.A). The subject of dopant pairing is not of critical importance at present. Pairing basically slows down diffusion more than is predicted by Fermi-level effects alone, but in realistic cases Fermi-level changes already predict greatly reduced diffusivities, so that pairing is most often a second-order effect. Thus although ignoring the effects of pairing can lead to large errors in deriving empirical expressions for dopant diffusivities from isoconcentration experiments (as discussed in Sec. X.B.5), in most other cases pairing is not an important factor. Clustering and precipitation, on the other hand, have a strong influence on diffusion. That precipitation processes dominate over clustering for B, Sb, and P now seems clear (although P diffusion cannot be described by quasiequilibrium conditions in this regime; see Sec. XIII.C). But the system of equations that describes dopants simultaneously diffusing and precipitating, has not been written down, nor has there been an attempt to determine whether the assumption of quasiequilibrium conditions is justified. The uncertain state of affairs for As, whether it forms clusters or precipitates or both, is a bit surprising considering the amount of work that has gone into investigating this problem. But as pointed out in Sec. VIII.A.1, the difference between coherent precipitates of small dimensions and multiatom clusters may not be a clear one when trying to interpret the physical state of the As atoms. From the discussion in Sec. VIII.A.1 it appears that an important difference between clustering and precipitation is that precipitation predicts that the electron concentration at the temperature of diffusion saturates at a value corresponding to the solid solubility of the As, while clustering predicts that the electron concentration will continue to increase beyond this value with increasing As concentration. Thus clustering and precipitation predict different behavior for the diffusion of P in a heavily doped As layer such as that shown in Fig. 34. When the As concentration is further increased, clustering theory predicts that the enhanced diffusion (compared to intrinsic doping conditions) of P will continue to increase with increasing As concentration, while precipitation predicts a leveling off of enhancement. At still higher concentrations, percolation (Secs. VIII.A.3 and Fig. 7) may become a factor, but the experimental results thus far favor rapid deactivation of dopants into clusters or precipitates rather than the sustaining of percolation regions in the crystal.

2. Nonequilibrium conditions

The situation for diffusion under nonequilibrium conditions is much less satisfactory than that for quasiequilibrium conditions. In this case, not only do we not have a basic set of equations that describes diffusion processes under many important conditions, but there is also a dearth of information on the basic properties of point defects, which will be needed in *any* formulation of nonequilibrium diffusion processes. Let us briefly review where the deficiencies appear to be, based on the at-

tempts to model these processes that were outlined in this paper.

We began the treatment of nonequilibrium diffusion in Sec. XII by treating the simplest case: diffusion of native point defects in an undoped silicon wafer. Writing the continuity equation for I and V point defects is straightforward [Eq. (12.1)], and the treatment in Secs. XII.A–XII.D shows the general types of behavior that should be expected with different relative values of point-defect parameters. But experiments that were originally designed to determine point-defect parameters such as the equilibrium concentrations C_I^* and C_V^*, point-defect diffusivities d_I and d_V, and the recombination rate between I and V have shown that nonequilibrium diffusion of point defects is not modeled well by Eq. (12.1).

Consider first the seemingly straightforward attempt to measure d_I by observing the transit time of I across silicon wafers of different thicknesses (Sec. XVI.B). The fact that for wafers of 20–100 μm thickness it takes a time of minutes to hours for I injected from one side of a wafer to cause enhanced dopant diffusion on the opposite (nonoxidizing) side of the wafer implies that d_I must be on the order of 10^{-9} cm^2sec^{-1} (at 1100 °C). But this is inconsistent with the result of the experiment indicated in Fig. 49 (also performed at 1100 °C), which unambiguously shows d_I to be orders of magnitude faster. Results from gettering (Sec. XVI.A) and low-temperature irradiation experiments (Sec. VII.B) also indicate a much faster interstitial diffusivity than the transit-time experiments. Possible explanations that could reconcile the various experimental results were put forth in Sec. XII.E and involve the interactions of native point defects with nonnative point defects. But what are these defects? All one can say at present is that if interactions with non-native defects are responsible for the apparent slowing down of the diffusing interstitial wave front, then these defects are present in the as-grown crystal and do not propagate into the epitaxially grown layer used in the experiment of Fig. 49. Unless further experiments are performed to check this and identify the nature of the defects, no definite conclusions can be made. It is also possible that surface effects may play a role in interpreting these experiments. For example, suppose that the interstitials that reach the opposite side of the wafer are at first quickly absorbed into surface traps, but as these traps become filled the interstitial excess concentration begins to rise. This too could explain the apparently low interstitial diffusivity based on present experimental results.

Questions about the boundary conditions are at least as important as questions of bulk interactions and may also prove harder to answer. Although the boundary conditions of Eq. (12.3) seem a reasonable starting point, experiments show that something much more complex is going on at the silicon surface. Figure 44 illustrates the basic importance of the boundary conditions. What is indicated in the figure is the two-dimensional effect of having alternating regions of point-defect injecting and noninjecting surfaces; the injecting regions become progressively smaller, while the noninjecting surfaces (in this case, thin layers of SiO_2 covered with Si_3N_4) remain the same width. By looking at the boundary conditions supposed by Eq. (12.3) it easy to appreciate that if the injecting and noninjecting ($g_I=0$, $g_V=0$) surfaces had comparable recombination velocities σ_I, σ_V, then as the width of the point-defect source regions became progressively smaller compared to the diffusion length of the point defects, one should expect the point-defect supersaturation under the injecting area to decrease. For the case of the oxidizing sample this is definitely not true. This is a very serious deficiency of the boundary conditions of Eq. (12.3), since the oxidation example is precisely the case of the greatest technological importance. The result indicates that the rate at which excess interstitials leave the silicon through the same surface that injects them is much greater than the rate at which they leave the noninjecting surface. But since both surfaces are oxides, one can only conclude that the surface loss term depends on time and cannot be modeled as a physical constant, as supposed in Eq. (12.3). On the other hand, the nitriding case shown in Fig. 44 shows that the ability of the noninjecting oxide surfaces to absorb excess vacancies is comparable to or greater than that of the nitriding surface that injects them. These types of effects could also be important in one-dimensional experiments such as the silicon-on-insulator (SOI) experiment depicted in Fig. 18, where the resulting supersaturation of point defects depends on the relative efficiency of the buried oxide surface in absorbing point defects compared to the injecting surface. We have reviewed alternative boundary conditions in Sec. XIV.E. Our conclusion is that we do not have a boundary condition to replace Eq. (12.3) and at present are using "effective" recombination velocities along with "effective" point-defect diffusivities and "arbitrary" time dependencies of point-defect injection rates. While this may serve the purposes of modeling over a finite range of conditions, it is hardly a satisfying solution.

Given the problems of describing and modeling point-defect diffusion when the crystal is lightly doped, problems formulating the correct sets of expressions and performing experiments to verify them are expected to be even more difficult for diffusion processes when the crystal is heavily doped. We have mentioned several times in this paper, beginning in Sec. VIII.A, the very real possibility that at the diffusion temperatures of interest (800 °C–1100 °C) the number of point defects associated with dopant atoms may exceed the number of isolated point defects even for relatively low concentrations of dopants (see Fig. 5). The point was raised in Secs. XIII.A and XIII.B that dopant atoms could play the same role as the non-native traps and recombination centers discussed in Sec. XII.E. In addition to these types of effects, the behavior of point defects diffusing under extrinsic doping conditions (such as the hypothetical potential well of Fig. 30) has not received much attention

from either a theoretical or an experimental viewpoint. Moreover, as mentioned in Sec. XIII.B, the effect of extrinsic doping on boundary conditions is completely unclear and yet could be a dominant factor.

B. Recent developments and future directions

Reviewing the successes and failures of models for the diffusion processes in silicon gives us a clearer idea of the type of work needed in the future. But coupled with this awareness should be the realization that changes in integrated circuit technology will undoubtedly play a role in redefining new goals and raising new questions (as they have in the past). We must therefore consider what basic work remains to be done that is pertinent to a wide variety of physical situations, as well as anticipate the needs of future silicon device technologies.

Examples of basic work that remains to be done are all too evident. The relatively poor understanding of nonequilibrium diffusion processes compared to those for quasiequilibrium conditions is due in large part to a lack of information on the fundamental properties of point defects: their equilibrium concentrations, diffusivities, recombination rates, etc. Any experiments that can help to determine these quantities will certainly be major contributions. Whether they come from self-diffusion studies, transit-time experiments, or techniques that attempt to measure point-defect concentrations directly, more definitive determinations of point-defect properties at elevated temperatures would be a real boon to the field. Work must continue to determine empirically the fractional I and V components of dopant diffusivities at temperatures below 1100 °C in order not only to advance modern integrated circuit technology but also to allow us to utilize dopant diffusion as a probe of point-defect behavior. Work must also continue on two- and three-dimensional diffusion effects because of their physical significance and technological importance. All of the experiments on point-defect diffusion and its effect on dopant diffusion presented in this paper utilize one-dimensional measurement techniques to obtain information on the two-dimensional diffusion behavior. This is necessary in the absence of a two-dimensional profiling technique. Several works have been presented recently reporting on progress towards developing techniques to measure two-dimensional dopant concentration profiles (Hill *et al.*, 1987; Subrahmanyan, Massoud, and Fair, 1987; Ahn and Tiller, 1988).

We mentioned in our Introduction that new effects involving point defects and dopant diffusion always seem to arise as new technologies in creating integrated circuit structures emerge. We can briefly point to two examples that have emerged only in the past year. One example is the recent finding by Hu (1987), using $TaSi_2$ layers, and Wen *et al.* (1987), using $TiSi_2$, that the formation of some silicides on the silicon surface injects vacancies into the bulk. This is an interesting development both because of the growing technological importance of silicides in integrated circuit technology (as low-resistance metals for ohmic contacts, gates, and interconnect lines) and because it gives us another means of injecting vacancies besides the thermal nitridation reactions described in Sec. XIV.F. One of the first interesting results to come out of the silicidation studies is that B diffusion is enhanced under the same conditions as Sb (Hu, 1987). This is the first solid experimental evidence of a vacancy component for B diffusion, unless one supposes that the $TaSi_2$ injects both vacancies and interstitials simultaneously. Of course the question of *why* a silicidation process injects point defects from the surface is of great interest. This question has not been satisfactorily answered for any other system, although it has already been proposed in both of the aforementioned studies that a silicon efflux during silicidation induces a vacancy influx. The experiment shown in Fig. 42 is evidence against this kind of idea in the case of nitridation of the silicon surface. A similar experiment could answer the obvious question: is the rate of vacancy injection correlated with the rate of silicidation? Further, can we therefore predict that other systems and other silicides will also inject vacancies? In a limited study (Fahey and Dutton, 1988) it was shown that oxidation of a WSi_2 layer, which also causes a large efflux of silicon atoms from the silicide/silicon interface, has no effect on the diffusion of B and P buried marker layers. But does silicidation of W inject vacancies? Does oxidation of $TaSi_2$ and $TiSi_2$? More experiments along these lines will provide essential information needed to understand what is going on and may also shed more light on the other surface processes that inject point defects.

Another example of technological changes fueling new areas of research is the growing interest in the properties of Ge in Si. Implantation of Ge has been investigated as a means of amorphizing silicon prior to dopant implantation in order to eliminate unwanted channeling effects that occur for implantation of dopants into crystalline silicon (Sadana *et al.*, 1984). There is also an interest in creating heterostructure devices utilizing strained Si_xGe_{1-x} layers [see the review article by People (1986)]. This may be an opportune time to investigate the mechanism(s) of Ge diffusion in Si. We have already mentioned in Sec. IX the results of Ge tracer experiments. Although the experimental results do not always agree between different studies, there is still an indication that Ge tracers are an effective probe for Si self-diffusion. It would be very interesting to have more information about the behavior of Ge diffusion under nonequilibrium conditions, i.e., its response to interstitial and vacancy excesses. This may provide important information concerning the mechanisms of self-diffusion in addition to providing information on Ge diffusion itself. The recent study of Fahey, Iyer, and Scilla (1989), which shows Ge diffusion to be enhanced under both I and V supersaturation conditions, is a good first step in this direction. However, other recent results indicate that some care must be taken in designing and interpreting experiments

involving Ge diffusion in Si. For example, Holland, White, and Fathy (1987) and Fathy, Holland, and White (1987) have shown that there is an enhanced oxidation rate of silicon heavily implanted with Ge and a large pileup of Ge at the SiO_2/Si interface. This result had previously been observed for bulk crystals of SiGe alloys (Margalit, Bar-Lev, Kuper, Aharoni, and Neugroschel, 1972; Neugroschel, Margalit, and Bar-Lev, 1973) and recently demonstrated to occur in strained SiGe alloy layers grown by molecular beam epitaxy (Patton, Iyer, Delage, Ganin, and McIntosh, 1987). In a related work, Pfiester and Alvis (1987) showed that oxidation-enhanced diffusion of B and the anomalous diffusion of high-concentration P were eliminated when codiffused with Ge that had been ion implanted. Is the Ge pileup at the surface altering the surface boundary conditions for point-defect injection and loss? Or is there some type of residual implantation damage that acts as a sink for point defects? Perhaps there is some other unknown bulk effect that is responsible. Experiments that incorporate Ge by methods other than ion implantation and can remove the Ge away from the silicon interface will be able to answer these questions.

Silicidation effects and questions concerning germanium in silicon are just two very recent examples of interesting subjects that have arisen because of integrated circuit technology issues. Other cases are sure to follow, indicating that there will be a continuing need to understand point-defect and dopant diffusion phenomena in silicon. It can also be said that more experimental information on the basic properties of point defects and dopant diffusion mechanisms, as outlined above, would surely speed up the process of understanding new diffusion phenomena as they arise.

C. Conclusion

While our understanding of diffusion processes in silicon is far from complete, we believe that today research in this field is on the right track and people are asking the right questions. The rapid improvement in our understanding of nonequilibrium diffusion processes in the past three years and attempts at modeling and measuring diffusion in two dimensions are encouraging examples of this assessment. This is still an interesting and exciting field, with both basic research that remains to be done and new issues of technological importance that will continue to develop and challenge us.

ACKNOWLEDGMENTS

The authors would like to express their thanks to Dr. Mark Law and Dr. Conor Rafferty for providing us with the capability to simulate the physical models discussed in this paper, and to Professor Robert Dutton, who supported and encouraged this work. We are grateful to our colleague Dr. Sung Tae Ahn for valuable discussions. The support of the Defense Advanced Research Projects Agency and the Semiconductor Research Corporation is also acknowledged.

APPENDIX A: DIFFUSION EQUATION UNDER EXTRINSIC CONDITIONS

In Sec. X.B it was stated that, under extrinsic conditions, the concentration gradient of ionized dopants introduces two effects that are not present for diffusion under intrinsic conditions: (i) an internal electric field is produced as evidenced by the spatial variation of the Fermi level E_f, and (ii) the concentrations of charged defects will also vary spatially because of the changes in E_f with local carrier concentrations. These two factors can be incorporated in a straightforward manner by extending the derivation of D_A presented in Sec. X.A.

1. Single-species diffusion

Consider first a single dopant species diffusing under extrinsic conditions. As an example, it is assumed that the extrinsic condition is brought about by n-type doping, so that we are interested in calculating the flux J_{A^+}. In this case the concentrations of A^+X^+ and A^+X^{++} are considered too small to contribute appreciably to the flux of A^+, so that only the A^+X^0, A^+X^-, and $A^+X^=$ defects are considered.

For AX diffusing in the presence of an electric field $\mathcal{E}$, the flux will be given by

$$J_{AX} = -d_{AX}\frac{\partial C_{AX}}{\partial x} + Z_{AX}\mu_{AX}C_{AX}\mathcal{E} , \tag{A1}$$

where Z_{AX} is the charge state of AX, and μ_{AX} is the mobility of the AX defect and is defined by Eq. (A1). Assuming that Boltzmann statistics are valid, the internal electric field is

$$\mathcal{E} = +\frac{\partial}{\partial x}[(E_f^i - E_f)/q] \tag{A2}$$

$$= -\frac{kT}{q}\frac{\partial}{\partial x}[\ln(n/n_i)] . \tag{A3}$$

In the case where Boltzmann statistics apply, the mobility can be related to the diffusivity through the Einstein relation

$$\mu_{AX} = \frac{q}{kT}d_{AX} . \tag{A4}$$

Equation (A1) can then be rewritten

$$J_{AX} = -d_{AX}\frac{\partial C_{AX}}{\partial x} - Z_{AX}d_{AX}C_{AX}\frac{\partial}{\partial x}(\ln n/n_i) . \tag{A5}$$

Assuming that $C_A \approx C_{A^+}$ (i.e., no precipitation or clustering), we have from the conditions of charge neutrality, $C_{A^+} + p = n$, and Boltzmann statistics, $pn = n_i^2$, that

$$\frac{n}{n_i}=\frac{C_{A^+}}{2n_i}+\left[\left[\frac{C_{A^+}}{2n_i}\right]^2+1\right]^{-1/2}, \tag{A6}$$

from which it is a simple matter to show that

$$\frac{\partial}{\partial x}[\ln(n/n_1)]=\frac{1}{2n_i}\left[\left[\frac{C_{A^+}}{2n_i}\right]^2+1\right]^{-1/2}\frac{\partial C_{A^+}}{\partial x}. \tag{A7}$$

In the general case Eq. (A1) then becomes

$$J_{AX}=-d_{AX}\frac{\partial C_{AX}}{\partial x}-Z_{AX}\left[d_{AX}\frac{C_{AX}}{C_{A^+}}\right]\frac{C_{A^+}}{2n_i}\left[\left[\frac{C_{A^+}}{2n_i}\right]^2+1\right]^{-1/2}\times\frac{\partial C_{A^+}}{\partial x}, \tag{A8}$$

where the quantity in the leftmost parentheses will be identified with D_A presently. Because the assumption of quasiequilibrium is made, we still maintain that the mass-action relation

$$\frac{C_A C_X}{C_{AX}}=K(T) \tag{A9}$$

introduced in Sec. X.A holds at each point in the diffusing profile, where K is a constant that depends only on temperature. In contrast to the intrinsic case, however, one cannot assume that C_X is constant across the spatial extent of the profile, since the concentration of charged X defects will change with the local carrier concentration. Equation (A9) implies that

$$\frac{\partial C_{AX}}{\partial x}=\frac{C_{AX}}{C_{A^+}}\frac{\partial C_{A^+}}{\partial x}+\frac{C_{AX}}{C_X}\frac{\partial C_X}{\partial x}. \tag{A10}$$

To derive a general expression for J_A, including the A^+X^0, A^+X^-, and $A^+X^=$ components, it is easiest to discuss their respective fluxes separately.

a. Flux of A^+X^0

In this case $Z_{A^+X^0}=+1$. Since $C_{A^+X^0}$ is not a function of carrier concentration, it can be assumed that C_{X^0} is constant across the profile, in which case

$$\frac{\partial C_{A^+X^0}}{\partial x}=\frac{C_{A^+X^0}}{C_{A^+}}\frac{\partial C_{A^+}}{\partial x}. \tag{A11}$$

Then we have from Eq. (A8) that $J_{A^+X^0}$ can be expressed simply as

$$J_{A^+X^0}=-\left\{1+\frac{C_{A^+}}{2n_i}\left[\left[\frac{C_{A^+}}{2n_i}\right]^2+1\right]^{-1/2}\right\}\times d_{A^+X^0}\frac{C_{A^+X^0}}{C_{A^+}}\frac{\partial C_{A^+}}{\partial x} \tag{A12}$$

$$=-hD^i_{A^+X^0}\frac{\partial C_{A^+}}{\partial x}, \tag{A13}$$

where the superscript i denotes the value found under intrinsic conditions and the quantity h has been defined as

$$h\equiv 1+\frac{C_{A^+}}{2n_i}\left[\left[\frac{C_{A^+}}{2n_i}\right]^2+1\right]^{-1/2}. \tag{A14}$$

Thus the diffusive component of J_A attributable to A^+X^0 is enhanced by a factor of h over the value found under intrinsic conditions. As C_{A^+} becomes much larger than n_i so that $C_{A^+}\approx n$, it is seen that h has a limiting value of 2.

b. Flux of A^+X^-

For the A^+X^- defects $Z_{A^+X^-}=0$, so that

$$J_{A^+X^-}=-d_{A^+X^-}\frac{\partial C_{A^+X^-}}{\partial x}, \tag{A15}$$

and no drift term is included. In contrast to the situation for A^+X^0 where C_{X^0} is assumed constant across the profile, C_{X^-} will vary with the local carrier concentration according to Eq. (6.10). Using Eq. (A10) we have

$$\frac{\partial C_{A^+X^-}}{\partial x}=\frac{C_{A^+X^-}}{C_{A^+}}\frac{\partial C_{A^+}}{\partial x}+\frac{C_{A^+X^-}}{C_{X^-}}\frac{\partial C_{X^-}}{\partial x} \tag{A16}$$

$$=\frac{C_{A^+X^-}}{C_{A^+}}\frac{\partial C_{A^+}}{\partial x}+\left[C^i_{A^+X^-}\frac{n}{n_i}\right]\frac{\partial}{\partial x}[\ln(n/n_i)]. \tag{A17}$$

Inserting this relation into Eq. (A15) yields the simple result

$$J_{A^+X^-}=-hD^i_{A^+X^-}\left[\frac{n}{n_i}\right]\frac{\partial C_{A^+}}{\partial x}, \tag{A18}$$

which says that the diffusive component of J_A attributable to A^+X^- is increased by a factor of $h(n/n_i)$ over that found under intrinsic conditions.

c. Flux of $A^+X^=$

The value of $Z_{A^+X^=}$ is -1. Using the results of Eqs. (A10) and (6.11) we can write

$$\frac{\partial C_{A^+X^=}}{\partial x}=\frac{C_{A^+X^=}}{C_{A^+}}\frac{\partial C_{A^+}}{\partial x}+\frac{C_{A^+X^=}}{C_{X^=}}\frac{\partial C_{X^=}}{\partial x} \tag{A19}$$

$$=\frac{C^i_{A^+X^=}}{C_{A^+}}\left[\frac{n}{n_i}\right]^2\frac{\partial C_{A^+}}{\partial x}+2C^i_{A^+X^=}\left[\frac{n}{n_i}\right]^2\frac{\partial}{\partial x}[\ln(n/n_i)]. \tag{A20}$$

Inserting this relation into Eq. (A9) yields

$$J_{A^+X^=} = -hD_{A^+X^=}\left[\frac{n}{n_i}\right]^2 \frac{\partial C_{A^+}}{\partial x}, \tag{A21}$$

which states that the diffusive component of J_A attributable to $A^+X^=$ defects is enhanced by an amount $h(n/n_i)^2$ over the value found under intrinsic conditions.

Combining Eqs. (A13), (A18), and (A21), we find that the general flux equation under extrinsic conditions for n-type doping is

$$J_A = J_{A^+X^0} + J_{A^+X^-} + J_{A^+X^=} \tag{A22}$$

$$= -h\left[D^i_{A^+X^0} + D^i_{A^+X^-}\left[\frac{n}{n_i}\right] + D^i_{A^+X^=}\left[\frac{n}{n_i}\right]^2\right]\frac{\partial C_{A^+}}{\partial x}. \tag{A23}$$

This is a relatively simple result. The only real assumptions made were that Boltzmann statistics apply and that charge neutrality is satisfied. Using Fermi-Dirac statistics does not alter the calculated flux significantly (Shrivastava and Marshak, 1980). In addition, under conditions in which Boltzmann statistics are not valid, there is always the question of whether effects from dopant-induced band-gap narrowing (Sec. VI.D) would leave the inclusion of Fermi-Dirac statistics an incomplete formulation of the problem. Therefore the inclusion of Fermi-Dirac statistics is not expected to improve process modeling capabilities.

The validity of the charge-neutrality assumption can be investigated directly by solving the diffusion equation subject to Poisson's equation and comparing it to the more approximate formulation presented above. Jüngling, Pichler, Selberherr, Guerrero, and Pötzl (1985) have shown that there is little difference between resultant simulated profile shapes using the two approaches. The disadvantage of solving Poisson's equation is of course the added difficulties associated with numerical solution of the system of equations. Recently, O'Brien and Srinivasan (1988) have presented a derivation of the internal electric field similar to the one above, but assuming Poisson's equation instead of charge neutrality. The results of their analysis show that the electric field calculated in this way is more accurate than the charge-neutrality approximation and can actually speed up numerical solutions (the electric field becomes smoother around pn junctions, where the assumption of charge neutrality is worst, and this relaxes the condition on how finely spaced the discretization grid needs to be).

One of the important features of the h factor is that it has a limiting value of 2 for $n \gg n_i$. This limiting behavior is also found when Fermi-Dirac statistics are used. Thus, for diffusion of a single species of dopant, the electric field factor introduces an increase in D_A by a factor of between 1 and 2.

A more thermodynamically rigorous derivation of the diffusion equation would begin with the supposition that the driving force for diffusion is the gradient in *chemical potential* rather than the gradient in chemical concentration as written in Eq. (A1) and would include the flux of X through the lattice in the derivation. Assuming the quasiequilibrium conditions implied by Eq. (A9) and the validity of Boltzmann statistics, Eq. (A23) follows directly from such an analysis (Hu, 1969).

The term "electric field" factor is appropriate terminology in that the h factor will arise only if the Fermi level varies spatially, which means that an electric field is present. That the h factor does not arise by simply including an additional drift component of diffusion on the AX defects can be deduced immediately from Eq. (A23), where it is seen that the h factor is the same for the positively charged A^+X^0 defect, the charge-neutral defect A^+X^-, and the negatively charged $A^+X^=$ defect.

2. Multiple-series diffusion

We now examine the simultaneous diffusion of multiple dopant species. For simplicity, the case in which only two different dopant species, A and B, are diffusing is examined. Extension to the situation in which many species of dopants are diffusing will be obvious from the derivation. For definiteness assume as in the previous section that extrinsic conditions are brought about by n-type doping and let A be an ionized donor A^+. Simultaneously B diffusion occurs in the same region, and we let B be either a donor or an acceptor dopant, $B^\pm$. Two situations in which such simultaneous diffusions are important are the process of emitter/base formation in bipolar transistors, in which high concentrations of As are simultaneously diffusing with lower concentrations of B, and the process of forming lightly doped drains where high concentrations of As are diffusing simultaneously with lower concentrations of P.

Derivation of flux expressions when more than one species of dopant atom is present follows straightforwardly from the treatment in the last section. Equation (A6), which reflects electroneutrality conditions under the assumption of Boltzmann statistics, is modified to

$$\frac{n}{n_i} = \frac{C_{A^+} \pm C_{B^\pm}}{2n_i} + \left[\left[\frac{C_{A^+} \pm C_{B^\pm}}{2n_i}\right] + 1\right]^{1/2} \tag{A24}$$

We then have

$$\frac{\partial}{\partial x}[\ln(n/n_i)] = \frac{1}{2n_i}\left[\left[\frac{C_{A^+} \pm C_{B^\pm}}{2n_i}\right]^2 + 1\right]^{-1/2} \times \left[\frac{\partial C_{A^+}}{\partial x} \pm \frac{\partial C_{B^\pm}}{\partial x}\right]. \tag{A25}$$

This leads to

$$J_{A^+} = -h_{A^+}D_{A^+}\frac{\partial C_{A^+}}{\partial x} \mp (h_{A^+} - 1)D_{A^+}\frac{\partial C_{B^\pm}}{\partial x} \tag{A26}$$

and

$$J_{B^\pm} - h_{B^\pm} D_{B^\pm} \frac{\partial C_{B^\pm}}{\partial x} \mp (h_{B^\pm} - 1) D_{B^\pm} \frac{\partial C_{A^+}}{\partial x} , \quad \text{(A27)}$$

where

$$h_{A^+} = 1 + \frac{C_{A^+}}{2n_i} \left[\left[\frac{C_{A^+} \pm C_{B^\pm}}{2n_i} \right]^2 + 1 \right]^{-1/2} , \quad \text{(A28)}$$

$$h_{B^\pm} = 1 + \frac{C_{B^\pm}}{2n_i} \left[\left[\frac{C_{A^+} \pm C_{B^\pm}}{2n_i} \right]^2 + 1 \right]^{-1/2} , \quad \text{(A29)}$$

and

$$D_{A^+} = D_{A^+X^0} + D_{A^+X^-} \left[\frac{n}{n_i} \right] + D_{A^+X^=} \left[\frac{n}{n_i} \right]^2 , \quad \text{(A30)}$$

$$D_{B^+} = D_{B^+X^0} + D_{B^+X^-} \left[\frac{n}{n_i} \right] + D_{B^+X^=} \left[\frac{n}{n_i} \right]^2 , \quad \text{(A31)}$$

$$D_{B^-} = D_{B^-X^0} + D_{B^-X^+} \left[\frac{n}{n_i} \right]^{-1} + D_{B^-X^{++}} \left[\frac{n}{n_i} \right]^{-2} . \quad \text{(A32)}$$

These flux equations are incorporated into the continuity equations

$$\frac{\partial C_{A^+}}{\partial t} + \frac{\partial J_{A^+}}{\partial x} = 0 , \quad \text{(A33)}$$

$$\frac{\partial C_{B^\pm}}{\partial t} + \frac{\partial J_{B^\pm}}{\partial x} = 0 , \quad \text{(A34)}$$

with separate boundary conditions for A^+ and $B^\pm$.

A discussion of dopant-dopant interactions when donor and acceptors diffuse simultaneously has been given by Hu (1969). A general derivation of dopant-dopant interactions using a thermodynamic formulation of transport has been presented by Rupprecht (1977).

APPENDIX B: NUMERICAL VALUES OF DIFFUSION CONSTANTS

Table IV contains a list of numerical values for diffusion constants used in the SUPREM simulation programs. The general form of the diffusivity for a dopant A is

$$D_A = D_{AX^0} + D_{AX^+}(p/n_i) + D_{AX^-}(n/n_i) + D_{AX^=}(n/n_i)^2 , \quad \text{(B1)}$$

where each diffusion component has a preexponential factor and activation energy of diffusion such as

$$D_{AX^0} = D^0 \exp(-E^0/kT) . \quad \text{(B2)}$$

Values for P and Sb come from Fair (1981a); values for B and As come from Chin and Barbuscia (1984). The expressions for B and As agree reasonably well with experiment, while the values for Sb and P are not as reliable. In addition, the expression for P as presented above is expected to be applicable only under intrinsic conditions or in an isoconcentration study because of the anomalies discussed in Sec. XIII.C.

TABLE IV. Diffusion parameters used in SUPREM.

Dopant	D^0 E^0	D^+ E^+	D^- E^-	$D^=$ $E^=$	(cm^2/sec) (eV)
B	0.278	0.23			
	3.40	3.40			
As	8.0		12.8		
	4.05		4.05		
P	3.84		4.44	44.2	
	3.66		4.00	4.37	
Sb	0.214		15.0		
	3.65		4.08		

APPENDIX C: ANALYTIC SOLUTIONS TO POINT-DEFECT INJECTION EQUATIONS

Analytic solutions for point-defect injection from one side of a wafer are presented below. Solutions similar to some of these can be found in the book by Carlsaw and Jaeger (1959) on heat flow. However, heat-flow equations do not account for the equivalent of the bulk recombination term, which we have included for the point-defect solutions.

1. Uniform concentration approximation

As mentioned in Sec. XII.C.1, the point-defect equation for the uniform concentration approximation,

$$\frac{\partial \Delta C_I}{\partial t} = \frac{g_I}{h} - \frac{\sigma_I + \sigma_I'}{h} \Delta C_I - k_I \Delta C_I , \quad 0 \le x \le h , \quad \text{(C1)}$$

is a linear ordinary differential of first order and therefore can be integrated to yield solutions of ΔC_I for time-varying surface conditions. In contrast to the simple monotonic behavior of ΔC_I pictured in Fig. 22 for the case of constant surface injection, much more complicated behavior results if the surface injection varies with time. A set of solutions of particular interest are the expressions for $\Delta C_I(t)$ when the surface flux $g_I(t)$ varies as

$$g_I(t) = A(t_0 + t)^{-b} , \quad \text{(C2)}$$

since this provides a convenient power-law dependence for empirical modeling. In addition, with b set equal to $\frac{1}{2}$, this expression is appropriate for a model that supposes $g_I(t)$ to be proportional to the Si oxidation rate. For simplicity, we ignore the effects of bulk recombination (the inclusion of the k_I term in the following solution is straightforward). In this case, the solution to Eq. (C1) is given by

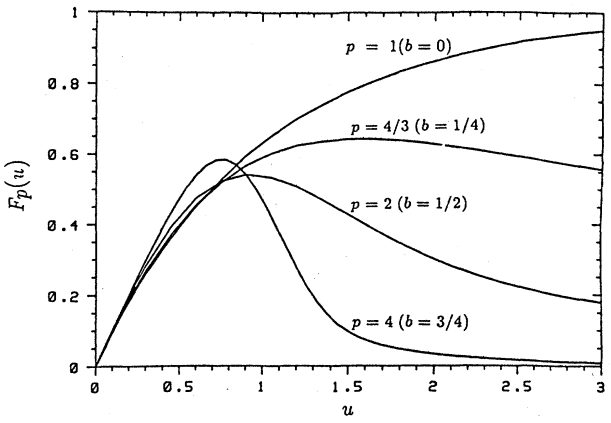

FIG. 54. Behavior of the integral function $F_p(u)$ defined by Eq. (C4).

$$\Delta C_I(t)=\frac{A}{2\bar{\sigma}_I}\frac{\tau_h^{1-b}}{1-b}(F_p\{[(t+t_0)/\tau_h]^{1-b}\} - e^{-t/\tau_h}F_p[(t_0/\tau_h)^{1-b}]) \ , \quad \text{(C3)}$$

where $\bar{\sigma}_I$ is the average value of the surface recombination velocities on the two sides of the wafer, $\tau_h \equiv h/2\bar{\sigma}_I$ is the characteristic surface lifetime of I discussed in Sec. XII.C.1, and we have defined the integral function

$$F_p(u)\equiv e^{-u^p}\int_0^u e^{v^p}dv, \quad p=\frac{1}{1-b} \ . \quad \text{(C4)}$$

The behavior of $F_p(u)$ as a function of u is shown in Fig. 54 for different values of p (or b). As an example of the complex behavior that can result for a time-varying injection rate, we compare in Fig. 55 the behavior of $\Delta C_I(t)$ for a thin wafer whose surface injection rate varies as

$$g_I(t)=At^{-1/2} \quad \text{(C5)}$$

to the behavior for the surface value of ΔC_I in a semi-infinite solid under identical injection conditions. Hu (1985c) has shown for the semi-infinite solid that $\Delta C_I(0,t)$ is given by

$$\Delta C_I(0,t)=A\sqrt{\pi/d_I}\,e^{t/\tau_\infty}\mathrm{erfc}(\sqrt{t/\tau_\infty}) \ , \quad \text{(C6)}$$

where τ_∞ is the characteristic surface lifetime of I in a semi-infinite solid. This figure should be compared with Fig. 22 for constant-source injection. It is seen that

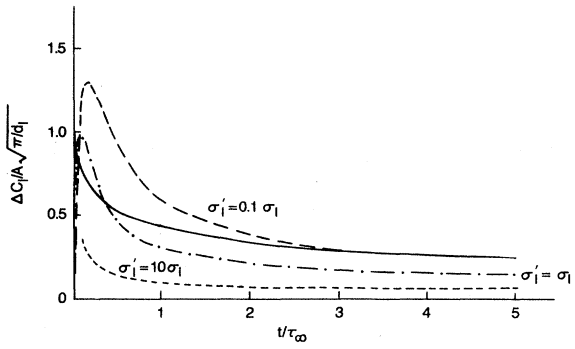

FIG. 55. Behavior of ΔC_I for a parabolic injection source [Eq. (C5)]. The solid line shows the behavior of $\Delta C_I(0,t)$ for a semi-infinite substrate as predicted by Eq. (C6); dashed lines show the behavior predicted by the uniform concentration approximation when the recombination velocity at the noninjecting side, σ_I', is allowed to vary with respect to the recombination velocity on the injecting side, σ_I.

whereas, for the semi-infinite solid, $\Delta C_I(0,t)$ monotonically decreases with time [as does $g_I(t)$], for the thin wafer, $\Delta C_I(t)$ initially increases sharply with time and then relaxes towards its steady-state behavior. This very different behavior between very thin and very thick wafers during their initial transient periods demonstrates how one may use these types of predictions either to determine point-defect parameters or to impose constraints on their magnitudes.

2. Wafer of arbitrary thickness h

The solution to the equation

$$\frac{\partial \Delta C_I}{\partial t}=d_I\frac{\partial^2 \Delta C_I}{\partial x^2}-k_I\Delta C_I, \quad 0<x<h \quad \text{(C7)}$$

subject to the boundary conditions

$$\begin{aligned} &g_I^0+d_I\frac{\partial \Delta C_I}{\partial x}-\sigma_I\Delta C_I=0 \ \text{ at } x=0 \ , \\ &d_I\frac{\partial \Delta C_I}{\partial x}+\sigma_I'\Delta C_I=0 \ \text{ at } x=h \ , \end{aligned} \quad \text{(C8)}$$

where g_I^0, σ_I, and σ_I' are constants, is given by

$$\frac{\Delta C_I}{g_I^0/\sigma_I}=\frac{(1+\sqrt{s'})\exp[(h-x)/L_I]-(1-\sqrt{s'})\exp[-(h-x)/L_I]}{(1+\sqrt{s'})(1+\sqrt{s}\,)\exp(h/L_I)-(1-\sqrt{s'})(1-\sqrt{s}\,)\exp(-h/L_I)}$$
$$-2\sum_{n=1}^{\infty}\frac{\left[4\varphi\varphi'\sin\left[\alpha_n\frac{h-x}{h}\right]+2\alpha_n\varphi\cos\left[\alpha_n\frac{h-x}{h}\right]\right]\exp(-\xi_n t)}{\alpha_n[1+(h/\alpha_n L_I)^2]\{[\alpha_n^2-4\varphi\varphi'-2(\varphi+\varphi')]\cos\alpha_n+2\alpha_n[1+\alpha_n(\varphi+\varphi')]\sin\alpha_n\}} \ , \quad \text{(C9)}$$

where

$$\xi_n \equiv [1+(h/\alpha_n L_I)^2]\frac{\alpha_n^2 d_I}{h^2} , \tag{C10}$$

$$L_I \equiv 2\sqrt{d_I/k_I} , \tag{C11}$$

$$\varphi \equiv h\sigma_I/2d_I, \quad \varphi' \equiv h\sigma_I'/2d_I , \tag{C12}$$
$$s \equiv k_I d_I/\sigma_I^2, \quad s' \equiv k_I d_I/\sigma_I'^2 ,$$

and α_n is the nth positive root of the transcendental equation

$$(\alpha_n^2 - 4\varphi\varphi')\tan\alpha_n = 2\alpha_n(\varphi+\varphi') . \tag{C13}$$

For wafer thicknesses large enough that the uniform concentration approximation is not valid, α_1 is of order unity or larger and all other α's increase with increasing n.

Note that the importance of bulk recombination is determined by the ξ_n terms through the factor $(h/\alpha_n L_I)^2$. Since α_1 yields the largest term, then

$$(h/\alpha_1 L_I)^2 \ll 1 \tag{C14}$$

is a sufficient condition to conclude that bulk recombination is unimportant. Equation (C9) shows explicitly that the effect of bulk recombination is to shorten the approach to steady state and to reduce the value of ΔC_I everywhere in the bulk.

For the case in which g_I is a function of time, a solution can be constructed from the constant-injection solution above by use of Duhamel's theorem. In this particular application, it is convenient first to note that the solution to the constant-injection case can be written in terms of steady-state and transient parts via

$$\frac{\Delta C_I(x,t)}{g_I^0/\sigma_I} = S(x) - \sum_{n=1}^{\infty} T_n(x)\exp(-\xi_n t) . \tag{C15}$$

Applying Duhamel's theorem, we have for a time-varying injection source at the surface, $g_I(t)$,

$$\Delta C_I(x,t) = \frac{g_I(t)}{\sigma_I}S(x) - \sum_{n=1}^{\infty} T_n(x)\left[\frac{g_I(0)}{\sigma_I}e^{(-\xi_n t)} + \frac{1}{\sigma_I}\int_0^t \frac{g_I(\lambda)}{\partial t}e^{-\xi_n(t-\lambda)}d\lambda\right] . \tag{C16}$$

For a linear/parabolic behavior such as Eq. (C2), the integral parts of the series can be written in terms of $F_p(u)$ as in Eq. (C3), but there is no numerical advantage in evaluating the series in this way. Solutions (C9) and (C16) are rapidly converging series except for $\xi_1 t \ll 1$. But physically this corresponds to the situation in which the wave front of excess defects is still far away from the other (noninjecting) side of the wafer, so that solutions for a semi-infinite substrate can be used.

3. Fixed boundary conditions

For completeness, we include analytic solutions for $\Delta C_I(x,t)$ in a wafer of arbitrary thickness h, in the case in which the value of ΔC_I at the injecting side is a specific function of time. In general, this type of solution is valid if the recombination velocity of the injecting surface ($x=0$) is much greater than the recombination velocity of the noninjecting surface on the other side of the wafer ($x=h$). From the results in Sec. XVI it can be seen that this type of solution should be applicable to the case of oxidation from one side of a wafer when the other side is covered by either a nonoxidizing SiO_2 layer, or Si_3N_4.

Using the same diffusion equation as Eq. (C7), subject to the boundary conditions that $\Delta C_I(0,t)$ is some specified constant ΔC_I^0 and

$$d_I\frac{\partial \Delta C_I}{\partial x} + \sigma_I'\Delta C_I = 0 \quad \text{at } x=h , \tag{C17}$$

we find that the solution for $\Delta C_I(x,t)$ is given by

$$\frac{\Delta C_I}{C_I^0} = \frac{(1+\sqrt{s'})\exp[(h-x)/L_I] - (1-\sqrt{s'})\exp[-(h-x)/L_I]}{(1+\sqrt{s'})\exp(h/L_I) - (1-\sqrt{s'})\exp(-h/L_I)} - 2\sum_{n=1}^{\infty}\frac{(\gamma_n^2+4\varphi^2)\sin(\gamma_n x/h)\exp\left[-\left[1+(h/\gamma_n L_I)^2\right]\frac{\gamma_n^2 d_I}{h^2}t\right]}{\gamma_n[1+(h/\gamma_n L_I)^2](4\varphi^2+2\varphi+\gamma_n^2)} , \tag{C18}$$

where γ_n are the roots to the transcendental equation

$$\gamma_n \cot\gamma_n = -2\varphi . \tag{C19}$$

The case of time-varying $\Delta C_I(0,t)$ can be constructed from this solution by use of Duhamel's theorem, as was done for Eq. (C9).

APPENDIX D: THE IMPORTANCE OF THE DISSOCIATIVE MECHANISM FOR DOPANT DIFFUSION

In the analysis below, we represent the nonsubstitutional fraction of dopant atoms that can diffuse by a substitutional-interstitial(cy) interchange mechanism

with the symbol AI. The identity of AI and I as either interstitialcies or interstitials does not affect the derivation. There are two point-defect reactions involving the formation of AI defects:

$$I + A \underset{k'_{I,A}}{\overset{k_{I,A}}{\rightleftharpoons}} AI \tag{D1}$$

and

$$V + AI \underset{k'_{V,AI}}{\overset{k_{V,AI}}{\rightleftharpoons}} A \ . \tag{D2}$$

Equation (D1) is the kick-out reaction discussed in Secs. X and XI and is commonly believed to be the dominant reaction by which dopant atoms interchange between substitutional and interstitial-type states. The forward reaction of Eq. (D2) accounts for recombination between V and AI, and the reverse reaction of Eq. (D2) is the dissociative reaction alluded to in Sec. X.A and in the introductory remarks of Sec. XI. This is an alternative process to the kick-out mechanism for the formation of AI and is not believed to be an important process for dopant diffusion; it is this assumption we wish to investigate further. One of the motivating factors for reexamining this assumption is that it is relevant to the question of whether the condition

$$\frac{\langle C_{AI} \rangle}{C^*_{AI}} = \frac{\langle C_I \rangle}{C^*_I} \tag{D3}$$

is a valid approximation under nonequilibrium conditions. If only the kick-out reaction of Eq. (D1) is important, then the arguments presented in Sec. XIII.A support the contention that the condition of Eq. (D3) is met under common experimental situations. But if the dissociative reaction of Eq. (D2) proceeds at a rate comparable to the kick-out reaction of Eq. (D1), then we cannot freely make this assumption. Consider, for example, the behavior of As diffusion under V and I injection conditions in Fig. 43. As discussed in Sec. XVII.A, there is a problem that arises in explaining this behavior assuming that $\langle C_{AsI} \rangle / C^*_{AsI} = \langle C_I \rangle / C^*_I$ and $\langle C_{AsV} \rangle / C^*_{AsV} = \langle C_V \rangle / C^*_V$. Because As diffusion is enhanced less than Sb diffusion during vacancy injection [Fig. 43(a)], one could reasonably conclude that the V component of As diffusion is less than that of Sb. But if the nonvacancy component of As diffusion is a substitutional-interstitial(cy) interchange mechanism, it is hard to understand why As diffusion is enhanced so little during I injection in Fig. 43(b). However, suppose that under equilibrium conditions the nonvacancy component of As diffusion comes about predominantly by dissociative rather than kick-out reactions, perhaps because of a reaction barrier to the kick-out process. Then if we inject excess I, this might have little effect in the formation of AsI defects. In fact, an enhanced formation rate of AsI may actually occur in this case by an undersaturation of V enhancing the reverse (i.e., dissociative) reaction of Eq. (D2). Although we do not show it here, it is a simple matter to demonstrate that in steady state the change in C_{AI} would tend towards the condition

$$\frac{\langle C_{AI} \rangle}{C^*_{AI}} = \frac{C^*_V}{\langle C_V \rangle} \tag{D4}$$

if the dissociative mechanism were dominant, and only in the special case of

$$\langle C_I \rangle \langle C_V \rangle = C^*_I C^*_V \tag{D5}$$

would Eq. (D3) be satisfied. If dissociative reactions are not negligible compared to kick-out reactions, it is possible for two different dopants to have exactly the same fractional interstitial and vacancy components of diffusion under equilibrium conditions and yet show completely different diffusion behavior under nonequilibrium conditions.

To assess the relative importance of dissociative and kick-out reactions, we first note that under equilibrium conditions we can write for Eq. (D1) that

$$k_{I,A} C^*_I = k'_{I,A} \left[\frac{C_{AI}}{C_A} \right]^* \equiv \nu_{I,A} \tag{D6}$$

and similarly for Eq. (D2) that

$$k_{V,AI} \left[\frac{C_{AI}}{C_A} \right]^* C^*_V = k'_{V,AI} \equiv \nu_{V,AI} \ . \tag{D7}$$

The quantity $\nu_{I,A}$ is the equilibrium formation rate, normalized to C_A, of AI from A and I, which is equal to the normalized equilibrium rate at which substantial AI converts into A and I. Likewise, the quantity $\nu_{V,AI}$ is the recombination rate under equilibrium conditions, normalized to C_A, between V and AI, and is equal to the normalized spontaneous formation rate of AI and V from A. Obviously, which process is dominant depends on the ratio of $\nu_{I,A}$ to $\nu_{V,AI}$. Using the definitions of $k_{I,A}$ and $k_{V,AI}$ in the same way as $k_{I,V}$ in Eq. (12.13) yields

$$\begin{aligned} \frac{\nu_{V,AI}}{\nu_{I,A}} &= \frac{a_{V,AI}}{a_{I,A}} \frac{(d_{AI} + d_V) C^*_V}{d_I C^*_I} \left[\frac{C_{AI}}{C_A} \right]^* \\ &\quad \times \exp\left[- \frac{\Delta E_{V,AI} - \Delta E_{I,A}}{kT} \right] \\ &= \frac{a_{V,AI}}{a_{I,A}} \left[\frac{D^*_{AI}}{D^*_{SiI}} \frac{C^*_V}{C_{Si}} + \frac{D^*_{SiV}}{D^*_{SiI}} \left[\frac{C_{AI}}{C_A} \right]^* \right] \\ &\quad \times \exp\left[- \frac{\Delta E_{V,AI} - \Delta E_{I,A}}{kT} \right] , \end{aligned} \tag{D8}$$

where $a_{V,AI}$ and $a_{I,A}$ are the capture radii for the processes. $\Delta E_{V,AI}$ is the energy barrier to recombination between V and AI; this energy barrier is conceptually similar to the energy barrier to recombination between I and V (Sec. XII.B.3 and Fig. 21, and Sec. XII.E.2). $\Delta E_{I,A}$ is introduced here as a reaction barrier to the kick-out process. Although we do not have quantitative information

for all the terms in Eq. (D8), the grouping of terms in brackets allows for some reasonable conjectures about the relative importance of the dissociative mechanism. Basically, we need the ratio of $\nu_{V,AI}/\nu_{I,A}$ to be on the order of, or much greater than, unity for dissociative reactions to be important. $C_V^*/C_{\rm Si}$ and $(C_{AI}/C_A)^*$ are estimated to be much less than one. Current estimates of the other values in brackets (see Figs. 2 and 10) predict that the bracketed term is always much less than unity. To make $\nu_{V,AI}/\nu_{I,A}$ approach unity requires that there be a large barrier to a dopant atom's being kicked into an interstitial-type state. This is hard to imagine for a dopant that diffuses by an interstitialcy-type mechanism, since it would then be hard to form AI defects at all, but it is not impossible to picture this situation for dopant atoms that diffuse as interstitials A_i, since the rate at which substitutional A atoms defects spontaneously jump into interstitial sites need not be related to how easy it is for approaching I defects to displace A into A_i sites. Although an open mind should be kept with respect to this question, based on present experimental knowledge it appears unlikely that the interstitial(cy) components of dopant diffusion can be attributed to the dissociative mechanism.

We should also mention that a similar analysis can be made to compare the reactions that involve formation of AV:

$$V + A \rightleftharpoons AV \tag{D9}$$

and

$$I + AV \rightleftharpoons A \ . \tag{D10}$$

An expression similar to Eq. (D8) results, but in this case there is no reaction barrier to the formation of AV from A and V. This makes it very likely that the forward reaction of Eq. (D9) is dominant over the reverse reaction of Eq. (D10) in forming AV.

REFERENCES

Abe, T., T. Samizo, and S. Maruyama, 1966, Jpn. J. Appl. Phys. **5**, 458.

Ahn, S. T., 1988, Ph.D. thesis (Stanford University).

Ahn, S. T., P. B. Griffin, J. D. Schott, J. D. Plummer, and W. A. Tiller, 1987, J. Appl. Phys. **62**, 4745.

Ahn, S. T., H. W. Kennel, J. D. Plummer, and W. A. Tiller, 1988a, Appl. Phys. Lett. **53**, 1593.

Ahn, S. T., H. W. Kennel, J. D. Plummer, and W. A. Tiller, 1988b, J. Appl. Phys. **64**, 4914.

Ahn, S. T., and W. A. Tiller, 1988, J. Electrochem. Soc. **135**, 2370.

Allen, W. G., 1973, Solid-State Electron. **16**, 709.

Allen, W. G., and K. V. Anand, 1971, Solid-State Electron. **14**, 397.

Andersen, P. E., A. Nylandsted Larsen, P. Tidemand-Petersson, and G. Weyer, 1988, Appl. Phys. Lett. **53**, 755.

Angelucci, R., G. Celotti, D. Nobili, and S. Solmi, 1985, J. Electrochem. Soc. **132**, 2726.

Angelucci, R., F. Cembali, P. Negrini, M. Servidori, and S. Solmi, 1987, J. Electrochem. Soc. **134**, 3130.

Angelucci, R., S. Solmi, and Q. Zani, 1986, Phys. Status Solidi A **96**, 541.

Antoniadis, D. A., 1982, J. Electrochem. Soc. **129**, 1093.

Antoniadis, D. A., 1983, in *Process and Device Simulation for MOS-VLSI Circuits,* edited by P. Antognetti, D. Antoniadis, R. W. Dutton, and W. G. Oldham, NATO ASI Series E, No. 62 (Martinus Nijhoff, Boston), p. 1.

Antoniadis, D. A., A. G. Gonzalez, and R. W. Dutton, 1978, J. Electrochem. Soc. **125**, 813.

Antoniadis, D. A., and I. Moskowitz, 1982a, J. Appl. Phys. **53**, 6788.

Antoniadis, D. A., and I. Moskowitz, 1982b, J. Appl. Phys. **53**, 9214.

Armigliato, R., D. Nobili, P. Ostoja, M. Servidori, and S. Solmi, 1977, in *Semiconductor Silicon 1977,* edited by H. R. Huff and E. Sirtl (Electrochemical Society, Pennington, New Jersey), p. 638.

Armigliato, R., D. Nobili, S. Solmi, A. Bourret, and P. Werner, 1986, J. Electrochem. Soc. **133**, 2560.

Armigliato, A., M. Servidori, S. Solmi, and A. Zani, 1980, J. Mater. Sci. **15**, 1124.

Armstrong, W. J., 1962, J. Electrochem. Soc. **109**, 1065.

Aziz, M. J., E. Nygren, W. H. Christie, C. W. White, and D. Turnbull, 1985, in *Impurity Diffusion and Gettering in Silicon,* Materials Research Society Symposia Proceedings No. 36, edited by R. B. Fair, C. W. Pearce, and J. Washburn (Materials Research Society, Pittsburgh), p. 101.

Bampi, S., 1987, Ph.D. thesis (Stanford University).

Baraff, G. A., E. O. Kane, and M. Schlüter, 1979, Phys. Rev. Lett. **43**, 956.

Baraff, G. A., and M. Schlüter, 1984, Phys. Rev. B **30**, 3460.

Bar-Yam, Y., and J. D. Joannopoulos, 1984, in *Proceedings of the Seventeenth International Conference on the Physics of Semiconductors, 1984,* edited by J. D. Chadi and W. A. Harrison (Springer, New York), p. 721.

Bouchetout, A. L., N. Tabet, and C. Monty, 1986, in *Proceedings of the Fourteenth International Conference on Defects in Semiconductors,* edited by H. J. von Bardeleben (Trans Tech, Aedermannsdorf, Switzerland), p. 127.

Bourgoin, J. C., and J. W. Corbett, 1972, Phys. Lett. A **38**, 135.

Bourgoin, J. C., J. W. Corbett, and H. L. Frisch, 1973, J. Chem. Phys. **59**, 4042.

Bronner, G. B., and J. D. Plummer, 1985, Appl. Phys. Lett. **46**, 510.

Bronner, G. B., and J. D. Plummer, 1987, J. Appl. Phys. **61**, 5286.

Bublik, V. T., S. S. Gorelik, and A. N. Dubrovina, 1968, Fiz. Tverd. Tela (Leningrad) **10**, 2846 [Sov. Phys. Solid State **10**, 2247 (1969)].

Car, R., P. J. Kelly, A. Oshiyama, and S. Pantelides, 1984a, Phys. Rev. Lett. **52**, 1814.

Car, R., P. J. Kelly, A. Oshiyama, and S. Pantelides, 1984b, in *Proceedings of the Seventeenth International Conference on the Physics of Semiconductors, 1984,* edited by J. D. Chadi and W. A. Harrison (Springer, New York), p. 713.

Car, R., P. J. Kelly, A. Oshiyama, and S. Pantelides, 1985, Phys. Rev. Lett. **54**, 360.

Carlsaw, H. S., and J. C. Jaeger, 1959, *Conduction of Heat in Solids,* 2nd ed. (Clarendon, Oxford).

Celotti, G., D. Nobili, and P. Ostoja, 1974, J. Mat. Sci. **9**, 821.

Chantre, A., M. Kechouane, and D. Bois, 1983, Physica (Utrecht) **116B**, 547.

Chichibu, S., T. Harada, and S. Matsumoto, 1988, Jpn. J. Appl. Phys. **27**, 1543.
Chin, G., and G. Barbuscia, 1984, unpublished.
Chiu, T. L., and H. N. Ghosh, 1971, IBM J. Res. Dev. **15**, 472.
Claeys, C. L., G. L. Declerck, and R. J. Van Overstraeten, 1978, in *Semiconductor Characterization Techniques,* edited by P. A. Barnes and G. A. Rozgonyi (Electrochemical Society, Princeton, New Jersey), p. 366.
Claeys, C. L., E. E. Laes, G. L. Declerck, and R. J. Van Overstraeten, 1978, in *Semiconductor Silicon 1977,* edited by H. R. Huff and E. Sirtl (Electrochemical Society, Pennington, New Jersey), p. 773.
Dannefaer, S., 1986, in *Proceedings of the Fourteenth International Conference on Defects in Semiconductors,* edited by H. J. von Bardeleben (Trans Tech, Aedermannsdorf, Switzerland), p. 103.
Dannefaer, S., P. Mascher, and D. Kerr, 1986, Phys. Rev. Lett. **56**, 2195.
Deal, B. E., and A. S. Grove, 1965, J. Appl. Phys. **36**, 3770.
Del Alamo, J., R. M. Swanson, and A. Lietoila, 1983, Solid-State Electron. **26**, 483.
Del Alamo, J., S. Swirhun, and R. M. Swanson, 1985, Solid-State Electron. **28**, 47.
Demond, F. J., S. Kalbitzer, H. Mannsperger, and H. Damjantschitsch, 1983, Phys. Lett. A **93**, 503.
Dhariwal, S. R., V. N. Ohja, and G. P. Srivistata, 1985, IEEE Trans. Electron Devices **32**, 44.
Dismukes, J. P., L. Ekstrom, and R. J. Paff, 1964, J. Phys. Chem. **68**, 3021.
Dorner, P., W. Gust, P. Predel, and U. Roll, 1984, Philos. Mag. **49**, 557.
Dunham, S. T., 1987, J. Appl. Phys. **62**, 1195.
Dunham, S. T., and J. D. Plummer, 1986a, J. Appl. Phys. **59**, 2551.
Dunham, S. T., and J. D. Plummer, 1986b, J. Appl. Phys. **59**, 2541.
Elkin, E. L., and G. D. Watkins, 1968, Phys. Rev. **174**, 881.
Erbil, A., G. S. Cargill III, and R. F. Boehme, 1985, in *Advanced Photon and Particle Techniques for the Characterization of Defects in Solids,* Materials Research Society Symposium Proceedings No. 41, edited by J. B. Roberto, R. W. Carpenter, and M. C. Wittels (Materials Research Society, Pittsburgh), p. 275.
Erbil, A., W. Weber, G. S. Cargill III, and R. F. Boehme, 1986, Phys. Rev. B **34**, 1392.
Fahey, P., 1983, unpublished.
Fahey, P., 1984, in *Computer-Aided Design of Integrated Circuit Fabrication Processes for VLSI Devices,* Stanford Electronics Laboratory, Stanford University Technical Report No. DXG501-84, July 1984, Chap. 14.
Fahey, P., 1985, in *Point Defects and Dopant Diffusion in Silicon* (Stanford Electronics Laboratory, Stanford University, Stanford, CA).
Fahey, P., G. Barbuscia, M. Moslehi, and R. W. Dutton, 1985, Appl. Phys. Lett. **46**, 784.
Fahey, P., and R. W. Dutton, 1988, Appl. Phys. Lett. **52**, 1092.
Fahey, P., R. W. Dutton, and S. M. Hu, 1984, Appl. Phys. Lett. **44**, 777.
Fahey, P., R. W. Dutton, and M. Moslehi, 1985,
Fahey, P., J. Greenfield, and R. W. Dutton, 1983, unpublished.
Fahey, P., S. Iyer, and G. Scilla, 1989, Appl. Phys. Lett. (in press).
Fair, R. B., 1979, J. Appl. Phys. **50**, 860.
Fair, R. B., 1981a, in *Processing Technologies,* edited by D. Khang, Applied Solid State Science, Supplement 2B (Academic, New York), p. 1.
Fair, R. B., 1981b, J. Electrochem. Soc. **128**, 1360.
Fair, R. B., M. L. Manda, and J. J. Wortman, 1986, J. Mater. Res. **1**, 705.
Fair, R. B., and J. C. C. Tsai, 1979, J. Electrochem. Soc. **124**, 1107.
Fathy, D., O. W. Holland, and C. W. White, 1987, Appl. Phys. Lett. **51**, 1337.
Finetti, M., G. Masetti, P. Negrini, and S. Solmi, 1980, IEE Proceedings-I **127**, 37.
Flicker, H., J. J. Loferski, and J. Scott-Monck, 1962, Phys. Rev. **128**, 2557.
Föll, H., U. Gösele, and B. O. Kolbesen, 1981, J. Cryst. Growth **52**, 907.
Forkel, D., H. Foettinger, M. Iwatschenko-Borho, S. Malzer, F. Meyer, W. Witthuhn, and H. Wolf, 1986, in *Proceedings of the Fourteenth International Conference on Defects in Semiconductors,* edited by H. J. von Bardeleben (Trans Tech, Aedermannsdorf, Switzerland), p. 557.
Francis, R., and P. S. Dobson, 1979, J. Appl. Phys. **50**, 280.
Frank, F. C., and D. Turnbull, 1956, Phys. Rev. **104**, 617.
Frank, W., 1975, in *Lattice Defects in Semiconductors, 1974,* Institute of Physics Conference Series No. 23, edited by F. A. Huntley (Institute of Physics, London), p. 23.
Frank, W., U. Gösele, H. Mehrer, and A. Seeger, 1984, in *Diffusion in Crystalline Solids,* edited by G. E. Murch and A. S. Nowick (Academic, New York).
Frank, W., A. Seeger, and U. Gösele, 1981, in *Defects in Semiconductors,* edited by J. Narayan and T. Y. Tan (North-Holland, New York), p. 31.
Fuller, C. S., and J. A. Ditzenberger, 1956, J. Appl. Phys. **27**, 544.
Ghez, R., and Y. J. van der Meulen, 1972, J. Electrochem. Soc. **119**, 1100.
Ghoshtagore, R. N., 1971, Phys. Rev. B **3**, 2507.
Gösele, U., W. Frank, and A. Seeger, 1979, in *Defects and Radiation Effects in Semiconductors, 1978,* Institute of Physics Conference Series No. 46, edited by J. H. Albany (Institute of Physics, London), p. 538.
Gösele, U., W. Frank, and A. Seeger, 1983, Solid State Commun. **45**, 31.
Gösele, U., and T. Y. Tan, 1983, in *Defects in Semiconductors II-Materials Research Society Symposium Proceedings,* Volume 14, edited by S. Mahajan and J. W. Corbett (North-Holland, New York), p. 45.
Greeneich, E. W., and R. H. Reuss, 1984, IEEE Electron Dev. Lett. **5**, 91.
Griffin, P. B., S. T. Ahn, W. A. Tiller, and J. D. Plummer, 1987, Appl. Phys. Lett. **51**, 115.
Griffin, P. B., P. M. Fahey, J. D. Plummer, and R. W. Dutton, 1985, Appl. Phys. Lett. **47**, 319.
Griffin, P. B., and J. D. Plummer, 1986, in *Materials Issues in Silicon Integrated Circuit Processing,* Materials Research Society Symposium Proceedings No. 71, edited by M. Wittmer, J. Stimmell, and M. Strathman (Materials Research Society, Pittsburgh), p. 75.
Griffin, P. B., and J. D. Plummer, 1987, in *Computer-Aided Design of Integrated Circuit Fabrication Processes for VLSI Devices,* Stanford Electron. Lab., Stanford University.
Guerrero, E., W. Jüngling, H. Pötzl, U. Gösele, L. Mader, M. Grasserbauer, and G. Stingeder, 1986, J. Electrochem. Soc. **133**, 2181.
Guerrero, E., H. Pötzl, G. Stingeder, M. Grasserbauer, K. Pi-

plitz, and W. K. Chu, 1985, J. Electrochem. Soc. **132**, 3048.
Guerrero, E., H. Pötzl, R. Tielert, M. Grasserbauer, and G. Stingeder, 1982, J. Electrochem. Soc. **129**, 1826.
Han, C. J., M. M. Moslehi, C. R. Helms, and K. C. Saraswat, 1985, Appl. Phys. Lett. **46**, 641.
Harris, R. D., J. L. Newton, and G. D. Watkins, 1982, Phys. Rev. Lett. **48**, 1271.
Harris, R. D., J. L. Newton, and G. D. Watkins, 1983, Phys. Rev. Lett. **51**, 1722.
Harris, R. D., and G. D. Watkins, 1985, *Proceedings of the Thirteenth International Conference on Defects in Semiconductors,* edited by L. C. Kimmerling and J. M. Parsey (The Metallurgical Society, Warrendale, Pennsylvania), p. 799.
Harris, R. M., and D. A. Antoniadis, 1983, Appl. Phys. Lett. **43**, 937.
Hayafuji, Y., K. Kajiwara, and S. Usui, 1982, J. Appl. Phys. **53**, 8639.
Hemment, P. L. F., 1985, in *Energy Beam-Solid Interactions and Transient Thermal Processing IV, 1985,* Materials Research Society Symposia Proceedings No. 35, edited by V. T. Nguyen and A. G. Cullis (Materials Research Society, Pittsburgh), p. 475.
Herzog, H. J., L. Csepregi, and H. Seidel, 1984, J. Electrochem. Soc. **131**, 2969.
Hettich, G., H. Mehrer, and K. Maier, 1979, in *Defects and Radiation Effects in Semiconductors, 1978,* Institute of Physics Conference Series No. 46, edited by J. H. Albany (Institute of Physics, London), p. 500.
Hill, C., 1981, in *Semiconductor Silicon, 1981,* edited by H. R. Huff and R. J. Kriegler (Electrochemical Society, Pennington, New Jersey), p. 988.
Hill, C., P. J. Pearson, B. Lewis, A. J. Holden, and R. W. Allen, 1987, in *Proceedings of the 17th European Solid State Research Conference (ESSDERC '87),* edited by P. G. Calzolari and G. Soneini (Tecnoprint, Bologna), p. 923.
Hirata, M., M. Hirata, and H. Saito, 1969, J. Phys. Soc. Jpn. **27**, 405.
Hirvonen, J., and A. Anttila, 1979, Appl. Phys. Lett. **35**, 703.
Ho, C. P., and J. D. Plummer, 1979a, J. Electrochem. Soc. **126**, 1516.
Ho, C. P., and J. D. Plummer, 1979b, J. Electrochem. Soc. **126**, 1523.
Holland, O. W., C. W. White, and D. Fathy, 1987, Appl. Phys. Lett. **51**, 520.
Hu, S. M., 1969, Phys. Rev. **180**, 773.
Hu, S. M., 1973a, in *Atomic Diffusion in Semiconductors,* edited by D. Shaw (Plenum, London), p. 217.
Hu, S. M., 1973b, Phys. Status Solidi B **60**, 595.
Hu, S. M., 1974, J. Appl. Phys. **45**, 1567.
Hu, S. M., 1975, Appl. Phys. Lett. **27**, 165.
Hu, S. M., 1980, J. Appl. Phys. **51**, 3666.
Hu, S. M., 1981, in *Defects in Semiconductors,* edited by J. Narayan and T. Y. Tan (North-Holland, New York), p. 333.
Hu, S. M., 1983, Appl. Phys. Lett. **43**, 449.
Hu, S. M., 1985a, in *VLSI Science and Technology, 1985,* edited by W. M. Bullis and S. Draydo (Electrochemical Society, Pennington, New Jersey), p. 465.
Hu, S. M., 1985b, J. Appl. Phys. **57**, 1069.
Hu, S. M., 1985c, J. Appl. Phys. **57**, 4527.
Hu, S. M., 1987, Appl. Phys. Lett. **51**, 308.
Hu, S. M., P. Fahey, and R. W. Dutton, 1983, J. Appl. Phys. **54**, 6912.
Hu, S. M., and S. Schmidt, 1968, J. Appl. Phys. **39**, 4272.
Ishikawa, Y., S. Matsumoto, and T. Niimi, 1983, in *Extended Abstracts of the Electrochemical Society,* Abstract No. 124.
Ishikawa, Y., Y. Sakina, H. Tanaka, S. Matsumoto, and T. Niimi, 1982, J. Electrochem. Soc. **129**, 644.
Ishikawa, Y., M. Tomisato, H. Honma, S. Matsumoto, and T. Niimi, 1983, J. Electrochem. Soc. **130**, 2109.
Jüngling, W., E. Guerrero, H. W. Pötzl, U. Gösele, L. Mader, M. Grasserbauer, and G. Stingeder, 1987, J. Electrochem. Soc. **133**, 2183.
Jüngling, W., P. Pichler, S. Selberherr, E. Guerrero, and H. W. Pötzl, 1985, IEEE Trans. Electron Devices **32**, 156.
Kalejs, J. P., L. A. Ladd, and U. Gösele, 1984, Appl. Phys. Lett. **45**, 268.
Kalish, R., T. O. Sedgewick, S. Mader, and S. Shatas, 1983, Appl. Phys. Lett. **44**, 107.
Kashio, T., and K. Kato, 1988, in *Technical Report of the Institute of Electronics, Information and Communication Engineers,* Vol. 88 (IEICE, Tokyo, Japan), p. 63.
Kimmerling, L. C., H. M. DeAngelis, and J. W. Diebold, 1975, Solid State Commun. **16**, 171.
Kitagawa, H., K. Hashimoto, and M. Yoshida, 1981, Jpn. J. Appl. Phys. **20**, 2033.
Kurtz, A. D., and R. Yee, 1960, J. Appl. Phys. **31**, 303.
Lannoo, M., and J. Bourgoin, 1981, in *Point Defects in Semiconductors I: Theoretical Aspects,* Springer Series in Solid-State Sciences No. 22 (Springer, New York).
Lecrosnier, D., J. Paugam, G. Pelous, F. Richou, and M. Salvi, 1981, J. Appl. Phys. **52**, 5090.
Lee, D. S., and J. G. Fossum, 1983, IEEE Trans. Electron Devices **30**, 626.
Lefèvre, H., 1980, Appl. Phys. **22**, 15.
Leroy, B., 1979, J. Appl. Phys. **50**, 7996.
Leroy, B., 1982, J. Appl. Phys. **53**, 4779.
Leroy, B., 1987, Philos. Mag. B **55**, 159.
Lidiard, A. B., 1960, Philos. Mag. **5**, 1171.
Lin, A., D. A. Antoniadis, and R. W. Dutton, 1981, J. Electrochem. Soc. **128**, 1131.
Lin, A., R. W. Dutton, and D. A. Antoniadis, 1979, Appl. Phys. Lett. **35**, 799.
Lin, A., R. W. Dutton, D. A. Antoniadis, and W. A. Tiller, 1981, J. Electrochem. Soc. **128**, 1121.
Loualiche, S., C. Lucas, P. Baruch, J. P. Gailliard, and J. Pfister, 1982, Phys. Status Solidi B **69**, 663.
Lucas, C., J. P. Gailliard, S. Loualiche, P. Baruch, J. Pfister, and R. Truche, 1979, in *Defects and Radiation Effects in Semiconductors, 1978,* Institute of Physics Conference Series No. 46, edited by J. H. Albany (Institute of Physics, London), p. 551.
Makris, J. S., and B. J. Masters, 1971, J. Appl. Phys. **42**, 3750.
Makris, J. S., and B. J. Masters, 1973, J. Electrochem. Soc. **120**, 1253.
Mallam, N., C. L. Jones, and A. F. W. Willoughby, 1981, in *Semiconductor Silicon, 1981,* edited by H. R. Huff, R. Kriegler, and Y. Takeishi (Electrochemical Society, Pennington, New Jersey), p. 979.
Mallon, C. E., and J. A. Naber, 1970, IEEE Trans. Nucl. Sci. **NS-17**, 123.
Mantovani, S., F. Nava, C. Nobili, and G. Ottaviani, 1986, Phys. Rev. B **33**, 5536.
Margalit, S., A. Bar-Lev, A. B. Kuper, H. Aharoni, and A. Neugroschel, 1972, J. Cryst. Growth **17**, 288.
Mascher, P., D. Kerr, and S. Dannefaer, 1987, Phys. Rev. B **35**, 3043.
Masetti, G., D. Nobili, and S. Solmi, 1977, in *Semiconductor Silicon, 1977,* edited by H. Huff and E. Sirtl (Electrochemical

Society, Pennington, New Jersey), p. 648.
Masetti, G., S. Solmi, and G. Soncini, 1973, Solid-State Electron. **16**, 1419.
Masetti, G., S. Solmi, and G. Soncini, 1976, Philos. Mag. 8th Series, **33**, 613.
Masters, B. J., and J. M. Fairfield, 1969, J. Appl. Phys. **40**, 2390.
Masters, B. J., and E. F. Gorey, 1978, J. Appl. Phys. **49**, 2717.
Masters, B. J., and E. F. Gorey, 1979, in *Defects and Radiation Effects in Semiconductors, 1978,* Institute of Physics Conference Series No. 46, edited by J. H. Albany (Institute of Physics, London), p. 545.
Mathiot, D., and J. C. Pfister, 1982, J. Phys. Lett. (Paris) **43**, 453.
Mathiot, D., and J. C. Pfister, 1983, Appl. Phys. Lett. **42**, 1043.
Mathiot, D., and J. C. Pfister, 1984, J. Appl. Phys. **55**, 3518.
Mathiot, D., and J. C. Pfister, 1985, Appl. Phys. Lett. **47**, 962.
Matsumoto, S., Y. Ishikawa, and T. Niimi, 1983, J. Appl. Phys. **54**, 5049.
Mayer, H. J., H. Mehrer, and K. Maier, 1977, in *Radiation Effects in Semiconductors, 1976,* Institute of Physics Conference Series No. 31, edited by N. B. Urli and J. W. Corbett (Institute of Physics, London), p. 186.
McQuhae, K. G., and A. S. Brown, 1972, Solid-State Electron. **15**, 259.
McVay, G. L., and A. R. DuCharme, 1973, J. Appl. Phys. **44**, 1409.
McVay, G. L., and A. R. DuCharme, 1975, in *Lattice Defects in Semiconductors, 1974,* Institute of Physics Conference Series No. 23, edited by F. A. Huntley (Institute of Physics, London), p. 91.
Michel, A. E., W. Rausch, P. A. Ronsheim, and R. H. Kastl, 1987, Appl. Phys. Lett. **50**, 416.
Miller, R. C., and S. Savage, 1956, J. Appl. Phys. **27**, 1430.
Miyake, M., 1985a, J. Appl. Phys. **57**, 1861.
Miyake, M., 1985b, J. Appl. Phys. **58**, 711.
Miyake, M., and H. Harada, 1982, J. Electrochem. Soc. **129**, 1097.
Mizuo, S., and H. Higuchi, 1982a, J. Electrochem. Soc. **129**, 2292.
Mizuo, S., and H. Higuchi, 1982b, Jpn. J. Appl. Phys. **21**, 281.
Mizuo, S., and H. Higuchi, 1982c, Denki Kagaku **50**, 338.
Mizuo, S., and H. Higuchi, 1982d, Jpn. J. Appl. Phys. **21**, 56.
Mizuo, S., and H. Higuchi, 1983, J. Electrochem. Soc. **130**, 1942.
Mizuo, S., and H. Higuchi, 1985, Appl. Phys. Lett. **46**, 587.
Mizuo, S., T. Kusaka, A. Shintani, M. Nanba, and H. Higuchi, 1983, J. Appl. Phys. **54**, 3860.
Morehead, F., 1987, in *Defects in Electronic Materials,* Materials Research Society Symposium Proceedings Vol 104, edited by M. Stavola, S. J. Pearton, and G. Davies (Materials Research Society, Pittsburgh), p. 99.
Morehead, F., and R. F. Lerer, 1986, Appl. Phys. Lett. **48**, 151.
Morehead, F., N. A. Stolwijk, W. Meyerberg, and U. Gösele, 1983, Appl. Phys. Lett. **42**, 690.
Morin, F. J., and J. P. Maita, 1954a, Phys. Rev. **94**, 1525.
Morin, F. J., and J. P. Maita, 1954b, Phys. Rev. **96**, 28.
Moslehi, M. M., and K. C. Saraswat, 1985, IEEE Trans. Electron Devices **32**, 106.
Münzel, H., G. Albert, and H. Strack, 1984, IEEE Trans. Electron Devices **5**, 283.
Mulvaney, B. J., and W. B. Richardson, 1987, Appl. Phys. Lett. **51**, 1439.
Murarka, S. P., and G. Quintana, 1977, J. Appl. Phys. **48**, 46.
Naber, J. A., C. E. Mallon, and R. E. Leadon, 1972, in *Radiation Damage and Defects in Semiconductors,* Institute of Physics Conference Series No. 16, edited by J. E. Whitehouse (Institute of Physics, London), p. 26.
Nabeta, Y., T. Uno, S. Kubo, and H. Tsukamoto, 1976, J. Electrochem. Soc. **123**, 1416.
Nachtrieb, N. H., and G. S. Handler, 1954, Acta Metall. **2**, 797.
Neugroschel, A., S. Margalit, and A. Bar-Lev, 1973, J. Phys. D **6**, 1606.
Newman, R. C., and J. Wakefield, 1962, in *Metallurgy of Semiconductor Materials,* Vol. 15, edited by J. B. Schroeder (Interscience, New York), p. 201.
Newton, J. L., A. P. Chatterjee, R. D. Harris, and G. D. Watkins, 1983, Physica (Utrecht) **116B**, 219.
Nicholas, K. H., 1966, Solid-State Electron. **9**, 35.
Nishi, K., and D. A. Antoniadis, 1986, J. Appl. Phys. **59**, 1117.
Nishi, K., K. Sakamoto, and J. Ueda, 1986, J. Appl. Phys. **59**, 4177.
Nobili, D., 1983, in *Aggregation Phenomena of Point Defects in Silicon,* proceedings of the Satellite Symposium to ESSDERC '82, Munich, edited by E. Sirtl, J. Goorissen, and P. Wagner (Electrochemical Society, Pennington, New Jersey), p. 189.
Nobili, D., A. Armigliato, M. Finetti, and S. Solmi, 1982, J. Appl. Phys. **53**, 1484.
Nobili, D., A. Carabelas, G. Celotti, and S. Solmi, 1983, J. Electrochem. Soc. **130**, 922.
Nygren, E., M. Aziz, D. Turnbull, J. Poate, D. Jacobson, and R. Hull, 1985, Appl. Phys. Lett. **47**, 105.
Nylandsted Larsen, A., F. T. Pedersen, G. Weyer, R. Galloni, R. Rizzoli, and A. Armigliato, 1986, J. Appl. Phys. **59**, 1908.
Nylandsted Larsen, A., P. Tidemand-Petersson, and P. E. Andersen, 1988, in *Proceedings of the Third International Conference on Shallow Impurities in Semiconductors* (IOP, Bristol), in press.
O'Brien, R. R., and G. R. Srinivasan, 1988, in *The Symposium on Process Modeling in Semiconductor Technology,* edited by G. R. Srinivasan, J. D. Plummer, and D. A. Antoniadis (Electrochemical Society, Pennington, New Jersey), p. 1.
Ogino, M., Y. Oana, and M. Watanabe, 1982, Phys. Status Solidi A **72**, 535.
Orlowski, M., 1988, Appl. Phys. Lett. **53**, 1323.
Pandey, K. C., 1986, Phys. Rev. Lett. **57**, 2287.
Pandey, K. C., A. Erbil, G. S. Cargill III, and R. F. Boehme, 1988, Phys. Rev. Lett. **61**, 1282.
Patton, G. L., S. S. Iyer, S. L. Delage, E. Ganin, and R. C. McIntosh, 1987, presented at Materials Research Society meeting, Boston, Mass., 1987.
Pauling, L., 1960, in *The Nature of the Chemical Bond,* 3rd ed. (Cornell University, Ithaca), p. 247.
Pennycook, S. J., J. Narayan, and O. W. Holland, 1984, J. Appl. Phys. **55**, 837.
People, R., 1986, IEEE J. Quantum Electron. **22**, 1696.
Pfiester, J. R., and J. R. Alvis, 1987, *Transactions of the International Electron Device Meeting,* Washington, D.C., 1987 (IEEE, New York).
Pfiester, J. R., and P. B. Griffin, 1988, Appl. Phys. Lett. **52**, 471.
Plaskett, T. S., 1965, Trans. Metall. Soc. AIME **233**, 809.
Pommerrenig, D., 1965, Acta Phys. Austriaca **20**, 338.
Reisman, A., E. H. Nicollian, C. K. Williams, and C. J. Merz, 1987, J. Electron. Mater. **16**, 45.
Rupprecht, H. S., 1977, in *Process and Device Modeling for Integrated Circuit Design,* edited by F. Van De Wiele, W. L. Engl, and P. G. Jespers (Noordhoff, Leyden), p. 33.
Sadana, D. K., W. Maszara, J. J. Wortman, G. A. Rozgonyi, and W. K. Chu, 1984, J. Electrochem. Soc. **131**, 943.

Scheid, E., and P. Chenevier, 1986, Phys. Status Solidi A **93**, 523.
Seeger, A., 1980, Phys. Status Solidi A **61**, 21.
Seeger, A., and P. Chik, 1968, Phys. Status Solidi **29**, 455.
Seeger, A., H. Föll, and W. Frank, 1977, in *Radiation Effects in Semiconductors, 1976,* Institute of Physics Conference Series No. 31, edited by N. B. Urli and J. W. Corbett (Institute of Physics, London), p. 12.
Seitz, F., 1950, Acta Crystallogr. **3**, 346.
Shin, Y. S., and C. K. Kim, 1984, IEEE Trans. Electron Devices, **ED-31**, 797.
Shockley, W., and J. Last, 1957, Phys. Rev. **107**, 392.
Shrivastava, R., and A. Marshak, 1980, J. Appl. Phys. **51**, 3222.
Simmons, R. O., and R. W. Baluffi, 1960, Phys. Rev. **117**, 52.
Sirtl, E., 1977, in *Semiconductor Silicon 1977,* edited by H. R. Huff and E. Sirtl (Electrochemical Society, Pennington, New Jersey), p. 4.
Solmi, S., R. Angelucci, F. Cembali, M. Servidori, and M. Anderle, 1987, Appl. Phys. Lett. **51**, 331.
Stolwijk, N. A., J. Hölzl, W. Frank, E. R. Weber, and H. Mehrer, 1986, Appl. Phys. A **39**, 37.
Stolwijk, N. A., D. Schuster, and J. Hölzl, 1984, Appl. Phys. A **33**, 133.
Strunk, H., U. Gösele, and B. O. Kolbesen, 1979, Appl. Phys. Lett. **34**, 530.
Subrahmanyan, R., H. Z. Massoud, and R. B. Fair, 1987, presented at the October 1987 meeting of the Electrochemical Society, Honolulu, Hawaii, Abstract No. 172.
Swalin, R. A., 1962, *Thermodynamics of Solids* (Wiley, New York), second edition, 1972.
Swalin, R. A., 1973, in *Atomic Diffusion in Semiconductors,* edited by D. Shaw (Plenum, London), p. 65.
Swaminathan, B., 1983, Ph.D. thesis (Stanford University).
Tan, T. Y., and B. J. Ginsberg, 1983, Appl. Phys. Lett. **42**, 448.
Tan, T. Y., and U. Gösele, 1981, Appl. Phys. Lett. **39**, 86.
Tan, T. Y., and U. Gösele, 1982a, Appl. Phys. Lett. **40**, 616.
Tan, T. Y., and U. Gösele, 1982b, J. Appl. Phys. **53**, 4767.
Tan, T. Y., and U. Gösele, 1985, Appl. Phys. A **37**, 1.
Taniguchi, K., and D. A. Antoniadis, 1985, Appl. Phys. Lett. **46**, 944.
Taniguchi, K., D. A. Antoniadis, and Y. Matsushita, 1983, Appl. Phys. Lett. **42**, 961.
Taniguchi, K., K. Kurosawa, and M. Kashiwagi, 1980, J. Electrochem. Soc. **127**, 2244.
Teague, J. R., C. M. Yagnik, G. J. Long, R. Gerson, and L. D. Lafleur, 1971, Solid State Commun. **9**, 1695.
Thurmond, C. D., 1975, J. Electrochem. Soc. **122**, 1133.
Tiller, W. A., 1981, J. Electrochem. Soc. **128**, 689.
Van der Meulen, Y. J., 1972, J. Electrochem. Soc. **119**, 531.
Van Vechten, J. A., 1974, Phys. Rev. B **10**, 1482.
Van Vechten, J. A., 1975, in *Lattice Defects in Semiconductors, 1974,* Institute of Physics Conference Series No. 23, edited by F. A. Huntley (Institute of Physics, London), p. 212.
Van Vechten, J. A., 1980, in *Handbook on Semiconductors,* Vol. 3, edited by T. S. Moss and S. P. Keller (North-Holland, New York), p. 1.
Van Vechten, J. A., 1987, presented at the March Meeting of the American Physical Society, New York City, paper BI 1.
Van Vechten, J. A., and C. D. Thurmond, 1976, Phys. Rev. B **14**, 3539.
Vasquez, R., M. Hecht, F. Grunthaner, and M. Naiman, 1984, Appl. Phys. Lett. **44**, 969.
Waite, T. R., 1957, Phys. Rev. **107**, 463.
Watkins, G. D., 1964, in *Radiation Damage in Semiconductors* (Dunod, Paris), p. 97.
Watkins, G. D., 1975, in *Lattice Defects in Semiconductors, 1974,* Institute of Physics Conference Series No. 23, edited by F. A. Huntley (Institute of Physics, London), p. 1.
Watkins, G. D., and K. L. Brower, 1976, Phys. Rev. Lett. **36**, 1329.
Watkins, G. D., and J. R. Troxell, 1980, Phys. Rev. Lett. **44**, 593.
Watkins, G. D., J. R. Troxell, and A. P. Chatterjee, 1979, in *Defects and Radiation Effects in Semiconductors, 1978,* Institute of Physics Conference Series No. 46, edited by J. H. Albany (Institute of Physics, London), p. 16.
Wen, D. S., P. L. Smith, C. M. Osburn, and G. A. Rozgonyi, 1987, Appl. Phys. Lett. **51**, 1182.
Wert, C. A., and R. C. Frank, 1983, Ann. Rev. Mater. Sci. **13**, 139.
Wichert, Th., M. L. Swanson, and A. F. Quennevelli, 1986, Phys. Rev. Lett. **57**, 1757.
Willoughby, A. F. W., A. G. R. Evans, P. Champ, K. J. Yallup, D. J. Godfrey, and M. G. Dowsett, 1986, J. Appl. Phys. **59**, 2392.
Wills, G. N., 1969, Solid-State Electron. **12**, 133.
Wilson, R. G., 1987, J. Appl. Phys. **61**, 933.
Wöhlbier, F. H., 1986, Ed., *Diffusion and Defect Data* (Trans Tech, Aedermannsdorf, Switzerland).
Yeager, H. R., and R. W. Dutton, 1985, IEEE Trans. Electron Devices **32**, 1964.
Yoshida, Y. M., 1988, Jpn. J. Appl. Phys. **27**, 967.
Yoshida, Y. M., S. Matsumoto, and Y. Ishikawa, 1986, Jpn. J. Appl. Phys. **25**, 1031.

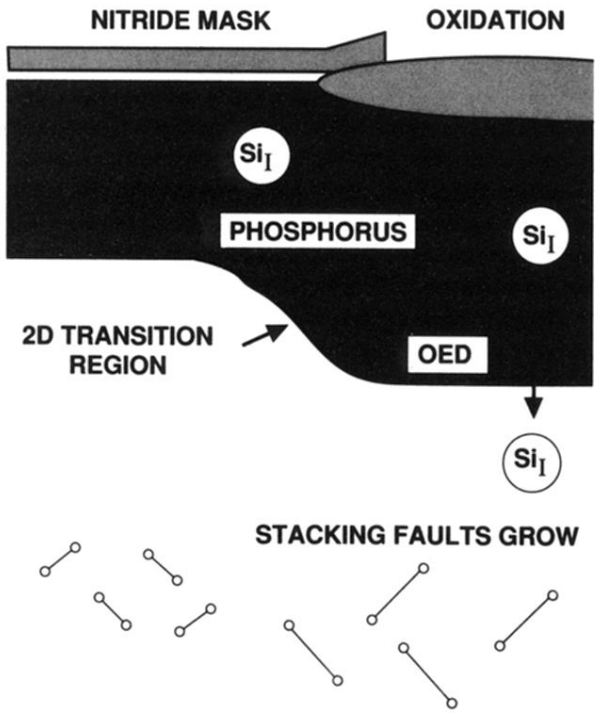

FIG. 35. A structure for monitoring the extent of oxidation-enhanced diffusion. Phosphorus or boron under the oxidizing region have higher diffusivities than the inert diffusion under the masked region.

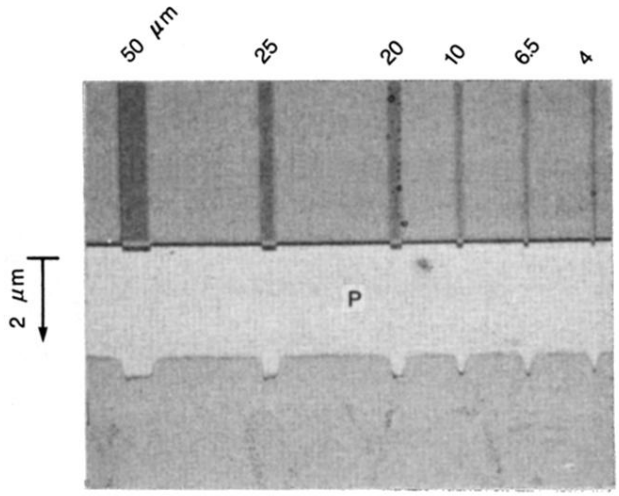

FIG. 45. Two-dimensional OED results. A photograph of a beveled-and-stained phosphorus junction after stripes of different widths at the surface have been oxidized for 4 h at 1100°C.

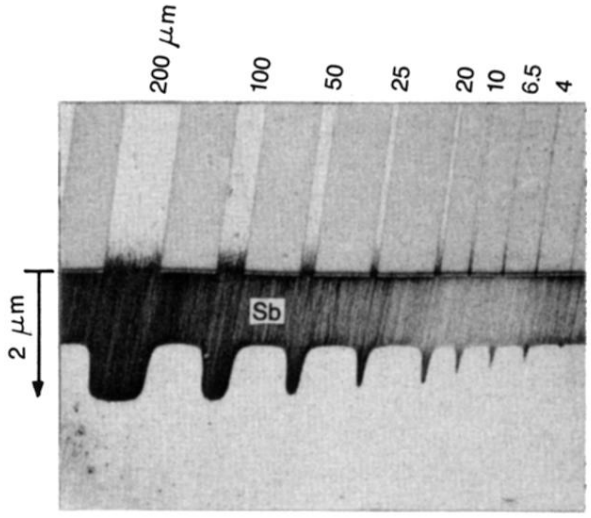

FIG. 48. Antimony diffusion after 1 h at 1100 °C in an ammonia ambient. In contrast to oxidation (Fig. 47), nitridation of the silicon surface gives rise to junction depths that depend strongly on the width of the nitriding stripe.

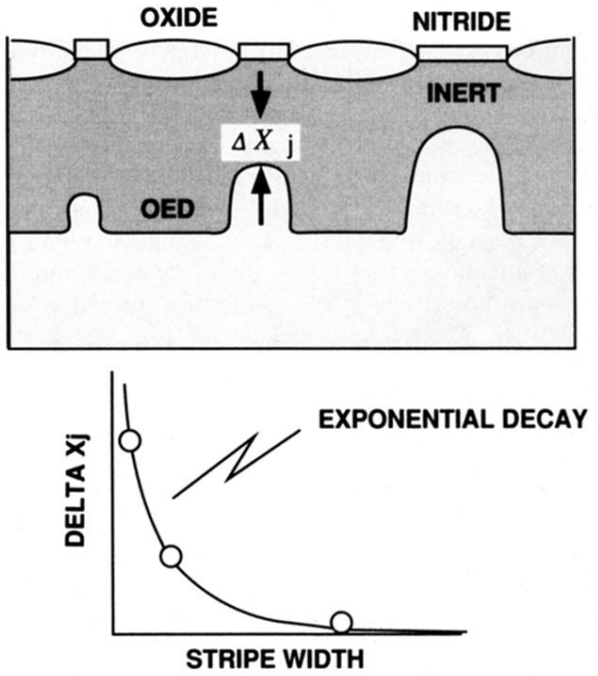

FIG. 50. A two-dimensional test structure to determine the diffusivity and surface recombination velocity of the point defects. The "spillover" of point defects from the injecting regions causes an enhanced diffusion under the inert regions, which depends on the inert stripe width.