CURRENT UNDERSTANDING OF POINT DEFECTS AND DIFFUSION PROCESSES IN SILICON

T. Y. Tan
IBM T. J. Watson Research Center, Yorktown Heights, NY 10598

U. Gösele
Max-Planck Institut für Metallforschung, Stuttgart, FRG

ABSTRACT

In this paper we first discuss the effects of oxidation of Si which established that vacancies (V) and Si self-interstitials (I) coexist in Si at high temperatures under thermal equilibrium and oxidizing conditions. Some essential points associated with Au diffusion in Si are then discussed. Analysis of Au diffusion results allowed a determination of the I-component and an estimate of the V-component of the Si self-diffusion coefficient. A discussion of theories on high concentration P diffusion into Si is then presented. Although presently there still is no theory that is completely satisfactory, significant progresses are recently made in treating some essential aspects of this subject.

1. INTRODUCTION

Atomic diffusion in Si can proceed via both direct and indirect mechanisms. Impurity atoms having no strong bonding interactions with Si atoms, and are hence located exclusively in interstices, can jump directly between the interstices. Species such as H, He (and presumably other noble gases), and Cu⁺ and Au⁺ (presumably the Cu and Au interstitials, respectively) are supposed to diffuse in this manner. Oxygen atoms, though possess strong bonding interactions with Si atoms, are also believed to diffuse directly by jumping between the bond-centered interstitial positions. By contrast, substitutional impurity and self-diffusion in Si need intrinsic point defects as diffusion vehicles. The vacancy (V-) mechanism is known to control self-diffusion in metals as in Ge, and, as will be discussed, in Si below about 1000°C. Above 1000°C, however, the interstitial (I-) mechanism plays a prominent role in Si self-diffusion as well as in the diffusion of substitutional dopants P, B, Al and Ga.

The nature of the dominant point defect species in Si at high temperatures has been a long controversial subject, though it is clear that there can only be three possibilities: V only, I only, and I and V coexist. The controversy arises for two reasons: (i) Presumably because of the small point defect concentrations in Si, direct experimental methods have not been helpful. There is no reported absolute measurement (1) results, and there is only one very recently reported quenching result (2) which indicated the presence of V (in the opinion of the present authors, however, this does not imply that I do not exist); (ii) Theoretical calculations, which are more suitable for low temperatures, have not yielded the needed 5eV activation enthalpy and the =10k (h: Boltzmann constant)
activation entropy commonly observed for Si self-diffusion. There were several early discussions associated with some aspects of the possibility of I and V coexisting in Si: Seeger and Chik (3) considered that I and V may be contributing equally to Si self-diffusion at 800°C; Prussin (4) considered that I and V should reach a local dynamical equilibrium under oxidizing conditions; Hu (5) considered that B should be diffusing in Si via an I-mechanism while As via a V-mechanism at high temperatures. In the last few years progresses in two areas helped to clarify our understanding of the nature of point defects and diffusion processes in Si considerably. Analyses of oxidation effects on stacking fault (SF) growth kinetics (6,7) and on dopant diffusion (8-10) showed beyond reasonable doubt that I and V coexist in Si at high temperatures under thermal equilibrium as well as oxidizing conditions. Analyses of Au diffusion in Si showed that I must exist (11,12) and that the features are totally consistent with the idea that I and V coexist (13). Studies of Au diffusion into dislocation-free Si allowed to determine the I-component (14) and to estimate the V-component (15) of the Si self-diffusion coefficient. In this paper the essential points associated with these two areas are first discussed. We then examine theories on high concentration P diffusion into Si. For various reasons, we believe that there is still no satisfactory theory. However, progresses have been made in treating some essential aspects of this subject.

2. OXIDATION EFFECTS

Oxidation of Si surfaces leads to the so called oxidation-enhanced and -retarded diffusion (OED and ORD) of the substitutional dopants and to the generation of the oxidation-induced stacking faults (OSF). These phenomena need to be considered together and are due to the fact that oxidation injects I into Si (5,16). It has been shown that a consistent interpretation of the data on OED/ORD phenomena is not obtained if either I or V alone were assumed to be present under thermal equilibrium conditions (10). Therefore for our present purpose, we consider that I and V are both present in Si at high temperatures under thermal equilibrium conditions. This requires that we consider I and V have attained a local dynamical equilibrium condition (4,17), which was actually fulfilled for long time experiments (8,14,18,19). The condition is given by (17)

$$C_{Ieq}^I = C_{Veq}^V,$$

where $C_I$ and $C_V$ denote the I and V concentrations respectively and the superscript eq denotes thermal equilibrium values. Eq. (1) is arrived at via the reaction

$$I + V \rightarrow \Omega$$

(where $\Omega$ denotes the ideal lattice) under steady state conditions. Using some of the more reliable OSF size data (20-22), phenomenological but satisfactory analyses of the OSF growth/shrinkage kinetics have been carried out (6,7). Define the I and V supersaturation ratios respectively as $S_I = C_I/C_{Ieq} - 1$ and $S_V = C_V/C_{Veq} - 1$ during the oxidation, we obtain (7)

$$\frac{dR_{SF}}{dt} = -(D_{Ieq}^I + D_{Veq}^V) \frac{S_I}{kT} + D_{Ieq}^I S_I - D_{Veq}^V S_V$$

258
for the OSF with a radius \( R_{SF} \). Here \( A \) is the area per atom in the fault (6.38\( \times 10^{-16} \) cm\(^2\)), \( \Omega \) the atomic volume (2\( \times 10^{-23} \) cm\(^3\)) and \( \gamma \) the stacking fault energy (0.026eV/atom). The dimensionless quantity \( \alpha_{\text{eff}} \) has a numerical value of \( \approx 2 \) and contains all factors related to interaction potentials between \( I \) and the Frank partial dislocation binding the OSF. Empirically, for (100) Si wafers oxidized in dry \( O_2 \), the OSF sizes are fitted satisfactorily by (6,7)

\[
R_{SF} = 1640\exp\left(-\frac{2.5\text{eV}}{kT}\right) t^{3/4} - \frac{4.8\times10^9}{kT}\exp\left(-\frac{5.02\text{eV}}{kT}\right) t \quad \text{cm.} \tag{4}
\]

We use Eq. (1) to relate \( S_I \) and \( S_V \) by \( S_V = -S_I/(1+S_I) \). A comparison of Eqs. (3) and (4) then yields (8,9),

\[
S_{I,02} = 6.6\times10^{-9} \exp(2.52\text{eV/kT}) t^{-1/4}, \tag{5}
\]

where the superscript \( 02 \) denotes the fact that this quantity is only suitable for (100) wafers oxidized in dry \( O_2 \). Eq. (5) is a most important piece of information obtained from OSF studies: it is used to obtain quantitative fittings of OED/ORD data. \( S_I \) is an instantaneous value, the time averaged value \( \bar{S}_I \) used in the actual OED/ORD data fitting is obtained by replacing the pre-exponential value in Eq. (5) by \( 8.8\times10^{-10} \) and by noting that \( t \) then denotes the oxidation duration.

In the presence of both \( I \) and \( V \), the substitutional dopant diffusivity is given by

\[
D^S = D^S_I + D^S_V, \tag{6}
\]

where \( D^S_I \) and \( D^S_V \) are respectively the \( I \)- and \( V \)-component of the dopant diffusivity. Under an oxidation which perturbs the thermal equilibrium \( I \) and \( V \) concentrations, the dopant diffusivity changes to

\[
D^S = D^S_I(C_I/C_1) + D^S_V(C_V/C_V). \tag{7}
\]

In terms of the normalized diffusivity enhancement defined as \( \Delta_{ox}^S = D_{ox}^S/D^S - 1 \), the use of Eqs. (1) and (7) yields (8,10)

\[
\Delta_{ox}^S = (2G_{ox}G_{S,ox}S_{ox}-1)/(1+S_1), \tag{8}
\]

where \( G \) is the fractional \( I \)-component of the dopant diffusivity under thermal equilibrium conditions defined as \( G = D^S_I/D^S \). Eq. (8) applies to an instantaneous diffusion enhancement. In experiments the time averaged value \( \Delta_{ox}^S \) is measured, but Eq. (8) still holds to a good approximation by having \( \bar{S}_1 \) used together with \( \Delta_{ox}^S \). A plot of Eq. (8) is shown in Fig. 1 for three \( G_1 \) values. The value of \( \Delta_{ox}^S \) is either positive or negative (OED or ORD) depending on the value of \( G_1 \) and \( S_1 \). In Fig. 2 we show a fitting of available Sb ORD data to Eq. (8). It is seen that the fitting is quite satisfactory on a quantitative basis with \( G_1 = 0.02 \). This kind of good fitting indicates that the model of \( I \) and \( V \) coexisting and attained local dynamical equilibrium is correct, particularly in light of the fact that no other reasonably thought of model can be equally satisfactory (10). The use of Eq. (8) allows to determine \( G_1 \) for a dopant at a given temperature. Some such values have been given elsewhere. To briefly summarize, it is found that the \( I \)-mechanism plays a prominent role in the diffusion of B, Ga, P and Al.
at high temperatures, i.e., $G_I > 0.5$ for these dopants at a temperature above 1100°C.

The above discussion applies for experiments with sufficiently long oxidation times. For short oxidation times, a transient exists for I and V to reach the local dynamical equilibrium. This is due to the presence of an energy barrier against I and V recombinations. The existence of such an energy barrier was suggested by Wertheim (25) to explain the fact that the rate of A-center (E = 0.16eV) production by high energy electron irradiation is strongly dependent upon the irradiation temperature (below room temperature). For high temperature cases, Hu (26) suggested that such a barrier of a few to several eV may exist, so that to a large extent I and V may behave independently. The experiments on ORD of Sb (18, 19) shown in Fig. 2 demonstrated that for oxidation times longer than a few hours local dynamical equilibrium according to Eq. (1) is reached. By performing similar experiments for shorter times (5 to 60 min.) Antoniadis and Moskowitz (9) showed that Eq. (1) is not readily satisfied when oxidation first started. At 1100°C they found a small OED for a 5 min. oxidation which gives way to ORD for an oxidation longer than 10 min., and, the diffusion retardation approaches the value expected for local dynamical equilibrium at about 60 min. Their findings may be rationalized as follows: during oxidation I are injected into Si by the SiO$_2$-Si interface but it takes about 1 hr at 1100°C for I-V recombination to proceed to such an extent that local dynamical equilibrium is reached. As a consequence, at the beginning there is no V undersaturation so that the oxidationuced I supersaturation enhances Sb diffusion via the small I-component of its diffusivity ($G_I = 0.02$).

We use Waite's theory of diffusion controlled reaction (27) to obtain from

$$\tau_{\text{diff}} \leq \frac{c}{4\pi D I V}$$

an estimate of the time $\tau_{\text{diff}}$ required to establish the local dynamical equilibrium in the Antoniadis-Moskowitz experiment if no recombination barrier were present. With $D = 10^{-5}$ cm$^2$ sec$^{-1}$ and the recombination radius $r_{\text{re}} = 5 \times 10^{-8}$ cm, Eq. (9) yields $\tau_{\text{diff}} < 0.05$ sec which is about 10$^5$ shorter than observed. We therefore conclude that I-V recombination is controlled by the overcoming of a recombination barrier that exceeds the Gibb's free energy of diffusion (of I and V) by $AG$. In this picture the factor of 10$^5$ by which the experimental $\tau$ value differs from that estimated from Eq. (9) arises from the Boltzmann factor exp($AG/kT$). Antoniadis and Moskowitz interpreted their observation in terms of an enthalpy barrier corresponding to $AH = 1.4$ eV assuming that in the expression

$$\Delta G = \Delta H - T \Delta S$$

the entropy contribution is negligibly small. By contrast, Gusele et al. (28) proposed that the main part of $\Delta G$ originates from the term $T \Delta S$ where $\Delta S$ is negative, i.e., from an entropy barrier, which is due to a consideration of the large pre-exponential factor associated with $D$.

Irrespective of the detailed origin of the barrier, with $AG = 1.4$ eV, the barrier limited reaction time, $\tau_{\text{bar}}$, for establishing the local dynamical equilibrium between I and V is given by

260
\[
\tau_{\text{bar}} = \left( \frac{n}{4\pi D^S_{IV}} \right) \exp(1.4eV/kT).
\]

Eq. (11) is plotted in Fig. 3 using \(D^S = 0.5(D^I_{\text{eq}} + D^V_{\text{eq}})\) with the quantities \(D^I_{\text{eq}}\) and \(D^V_{\text{eq}}\) taken from Eq. (21) which will be discussed. It is seen that \(\tau_{\text{bar}}\) increases rapidly with a decrease of the temperature and attained values of 2, 17 and 170 days at 1000, 950 and 900°C respectively. Even if it is assumed that these results are over-estimated by a factor of 10, \(\tau_{\text{bar}}\) values are still 5hrs, 1.7 days and 17 days at 1000, 950 and 900°C respectively. One significance of this finding is that in the expression

\[
D^S = D^S(C^I_{\text{eq}}/C^I_{I}) + D^S(C^V_{\text{eq}}/C^V_{V}),
\]

the quantities \(C^I_{I}\) and \(C^V_{V}\), arising from experimental perturbation of the respective thermal equilibrium values, are not related by Eq. (1) in an experiment performed in such low temperature range and having time durations much less than the appropriate \(\tau\) values. For device processing in the temperature range of 850 to 950°C for a few hrs, should there be any reason that the point defect thermal equilibrium values are perturbed, the change only applies to the perturbed species independently, i.e., \(I\) and \(V\) need to be considered as not having attained the local equilibrium condition.

The fact that I-V dynamical equilibrium is attained extremely slowly at lower temperatures offers the possibility of measuring the quantities \(D^I\) and \(D^V\) independent of each other and independent of the \(I\) and \(V\) concentrations. There already exist some OSF and OED/ORD experiments conducted with the intention of measuring \(D^I\) at 1100°C or above (29-31). However, because of the temperatures and oxidation times involved, I-V local dynamical equilibrium should have been reached in these experiments and we have shown that such experiments yield an effective diffusivity given as (15,24,32)

\[
D^\text{eff}_{IV} = (D^I_{\text{eq}} + D^V_{\text{eq}})/(C_{I}^\text{eq} + C_{V}^\text{eq}),
\]

instead of \(D^I\). Depending upon the relative magnitude of \(C_{I}^\text{eq}\) and \(C_{V}^\text{eq}\), which are unknown quantities themselves, Eq. (13) may yield an approximate value for either \(D^I\) or \(D^V\). We have argued that the diffusivity values obtained from these experiments (29-31) should be interpreted as \(D^V\) (24,32). However, for the lower temperature range, e.g., 850 to 950°C, the I-V local dynamical equilibrium will not be reached for an experiment involves less than a day's time. For such cases, effects associated with dopant diffusion or stacking fault growth would be solely due to the point defect species whose thermal equilibrium concentration has been perturbed by the experimental condition. This means the other species does not contribute to the experimental deviations observed if its thermal equilibrium concentration has not been also directly perturbed by the experimental condition. We can now suggest an experiment to directly determine \(D^I\): High concentration P diffusion into Si proceeds efficiently in the temperature range noted above and the diffusion causes a quite high I supersaturation in the Si interior. Thus, experiments similar to those performed for OED/ORD and OSF (29-31) can be carried out at the low temperature range (850 to 950°C) with surface P diffusion replacing oxidation to obtain \(D^I\). Similarly, \(D^V\) can be determined upon finding an interface reaction which injects V into Si. Presently we do not know with
certainty of such a reaction.

3. DIFFUSION OF GOLD INTO DISLOCATION-FREE SILICON

Gold atoms, which under thermal equilibrium conditions may be incorporated in Si either as substitutional (Au :Au\(^3^+)\) or interstitial (Au\(_i\):Au\(_s\)) atoms, may diffuse via either the kick-out mechanism (11) or the Frank-Turnbull mechanism (33). The two mechanisms have in common that long range transport of Au atoms occur via migration of Au\(_i\) atoms which may either jump from interstice to interstice to remain as Au\(_i\) atoms or from interstice to lattice positions to become Au\(_s\) atoms. This is true because though the thermal equilibrium concentration of Au\(_s\) is much larger than that of Au\(_i\), the mobility of Au\(_i\) is much much larger than that of Au\(_s\) and hence long range transport of Au\(_s\) may be ignored. However, the two mechanisms differ in the ways Au\(_i\) and Au\(_s\) atoms interchange. In the kick-out mechanism the interchange involves I according to (11)

\[ \text{Au}_i \leftrightarrow \text{Au}_s + I. \]  

(14)

This mechanism creates an I supersaturation in the crystal which is balanced by I out-diffusion. This means if the mechanism is operative than I are involved in Si self-diffusion but it does not a priori mean that V are not contributing to Si self-diffusion. In the Frank-Turnbull mechanism the interchange process involves V according to (33)

\[ \text{Au}_i + V \leftrightarrow \text{Au}_s. \]  

(15)

This mechanism creates a V undersaturation in the crystal interior which is balanced by V in-diffusion. If operative, it means that V are involved in Si self-diffusion but it does not mean I are not contributing to Si self-diffusion. We now assume that local dynamical equilibrium between I, V, Au\(_i\), and Au\(_s\) is established. The necessary and sufficient condition for this assumption to be true is that the three reactions (2, 14, 15) have reached their equilibrium state. This requires that two out of the three reactions to be sufficiently fast. It is not required that the third reaction be equally fast, since if equilibrium conditions are satisfied by two of the reactions, then the third is automatically under equilibrium conditions. Under this assumption the normalized Au concentration \( C = C_0/C_{eq} \) to be measured in an experiment on Au diffusion into \( ^8\text{Si} \) may be approximately described by

\[ \frac{3C}{C} = \frac{3}{\delta x} (D_{eff} \frac{\partial C}{\partial x}) \]  

(16)

with the effective diffusion coefficient given by (13)

\[ D_{eff} = (C_{eq}^2 + D_{V,V}C_{eq} + D_{I,I}C_{eq})D_{I,I}C_{eq}/C_{eq}. \]  

(17)

In Eq. (17) contributions to the Au\(_i\) - Au\(_s\) interchange process due to both kick-out and Frank-Turnbull mechanisms are accounted for. In deriving Eq. (17) the
following approximations are used

\[ C_i = C_{eq}^i, \quad (C_i: \text{Au}_i \text{ concentration}) \]  

\[ C_{eq}^s \geq C_{eq}^v, \]  

\[ C_s \gg (C_{eq}^{eq}v^{1/2}). \]  

If only \( V \) and no \( I \) are present Eq. (17) yields a constant effective Au \( S \) diffusity

\[ D_{eff}^V = D_{eq}^V/C_{eq}^v. \]  

The same result holds if \( D_{eq}^V \) is not zero but among the reactions (2,14,15) only (15) operates sufficiently fast. If only \( I \) and no \( V \) are present Eq. (17) yields a strong concentration dependent diffusivity

\[ D_{eff}^I = C^{-2}D_{eq}^I/C_{eq}^i. \]  

The same result holds if \( D_{eq}^V \) is not zero but among the reactions (2,14,15) only (14) operates sufficiently fast. As observed by Stolwijk et al. (14), Au concentration profile after diffusion into dislocation-free Si at and above 800°C can not be describable by the constant diffusity given by Eq. (19) but are satisfactorily fitted by the use of Eq. (20). Fig. 4 shows a typical experimental profile due to Stolwijk et al. It can be seen that an erfc-function type profile as expected for \( D_{eff}^V \) failed to fit the data while the use of \( D_{eff}^I \) is quite satisfactory. This shows that \( I \) are contributing to Si self-diffusion but it does not mean that \( V \) are not contributing for the reasons that: (i) It can not be certain that this is not only due to the fact that of the reactions (2,14,15) only (14) operates sufficiently fast; and (ii) The strong dependence of \( D_{eff}^I \) on \( C^{-2} \) reduces the effect of \( D_{eq}^V \) by this factor for most part of the experiment for which \( C^{eq} \) and \( C^{-2} \) hold. Indeed Morehead et al. (13) found that while almost perfect fitting were obtained for short time diffusion profiles small but observable deviations occurred if Eq. (20) were used but a much better fitting is obtained by the use of Eq. (17). This not only shows that \( V \) are contributing to Si self-diffusion but also allowed to estimate that \( D_{eq}^V \) at 1000°C. The quantity \( D_{eq}^I \) has been determined by Stolwijk et al. (14) as

\[ D_{eq}^V = 914 \exp(-4.84 eV/kT) \text{ cm}^2 \text{ sec}^{-1}. \]  

We have already shown that at T\( > 900°C \) reaction (2) is expected to become ineffective. At still lower temperatures, as will be discussed, reaction (14) is also expected to become ineffective. In that case only reaction (15) will be operating and an erfc-function type profile should show up. Willcox et al. (34) found such a case at 700°C. With their data we obtained that \( D_{eq}^V = 8.8 \times 10^{-7} \text{ cm}^2 \text{ sec}^{-1} \) at 700°C (15). This together with the data of Morehead et al. (13) yields the rough estimate
Eq. (21) is plotted in Fig. 5 in which a data point on $D_{C_{\text{eq}}}$ obtained by analyzing the Ni precipitation behavior (15) of Kitagawa et al. (35) is also shown. An important point is, as rough estimates, the sum of the formation and migration entropies are now obtained from Eq. (21) as 3k for V and 10k for I. That is, while I are quite extended (3), V are fairly point-like. Experimentally observed $D_{c}$ values from relatively higher temperature measurements are usually of $10^{3} \exp (-5eV/kT)$ while those from relatively lower temperature experiments are of $\exp (-4eV/kT)$ (36-40). These values are quite compatible with Eq. (21).

We now make an estimate to show that at very low temperatures the kick-out mechanism, reaction (14), can become ineffective, while the Frank-Turnbull mechanism, reaction (15), still operates. It is important to note that if this can not be true then it can not be explained why Au diffuses into dislocation-free Si apparently via the kick-out mechanism at higher temperatures but the Frank-Turnbull mechanism at a very low temperature. Eqs. (19) and (20) show that Au diffusion would appear as dominated by the kick-out mechanism in $\gamma_{V}$ long as $C^{2}D_{C_{\text{eq}}}^{V}/D_{C_{\text{eq}}}^{V} > 1$ holds. Because for a large part of the experiment $\gamma_{V} > 1$ holds, the kick-out mechanism may still appear as dominating the diffusion process even for a case for which the frequency of the kick-out events is less than that of the Frank-Turnbull events. On the other hand, the frequency of the kick-out events should decrease more rapidly than that of the Frank-Turnbull events with a decrease in temperature because the former involves the creation of I which is characterized by an activation enthalpy (the I formation enthalpy) while for the latter there is not yet a good reason for assuming that a high enough energy barrier exists to prevent an Au atom drops in a V. In the following we show that under thermal equilibrium conditions a quantitative estimate can be made which indicates that the above discussed situation can be true.

For the Frank-Turnbull mechanism, if an Au atom and a V meet, there is a definite probability $\gamma_{I}$ that reaction (15) will go forward to produce an Au atom. Under thermal equilibrium conditions, we obtain the rate of increase in $C_{I}$ via this mechanism as

$$(dC_{I}/dt)_{PT} = \gamma_{I}(4\pi r_{V}/\eta)D_{C_{\text{eq}}}^{V}.$$  

(22)

For the kick-out mechanism, reaction (14), the Au atom has a certain frequency $\nu_{I}$ by which it produces Au and I. We do not know this frequency but if Au and I meet there is a definite probability $\theta_{S}$ for reaction to produce Au and under thermal equilibrium conditions this reaction balances that due to $\nu_{I}$. Therefore we have,

$$(dC_{I}/dt)_{ko} = \theta_{S}(4\pi r_{S}/\eta)D_{C_{\text{eq}}}^{I}.$$  

(23)

From Wilcox et al. (34) we have

$$D_{C_{\text{eq}}}^{I} = 1.8x10^{-2}\exp (-1.13eV/kT)C_{I}^{2} \text{ cm sec}^{-1}.$$  

(24)

Eqs. (21a) and (22) to (24) yield for the rate of converting Au into Au via
the kick-out mechanism to that via the Frank-Turnbull mechanism as

\[ r = \frac{(dC_s}{dt})_{FT}/(dC_s}{dt})_o = a \exp(-3.71eV/kT); \frac{C_{eq}}{V}, \]

where \( a \) is a constant of the order of \( 10^5 \). We do not know the quantity \( C_{eq} \) but it can be seen from Eq. (25) that \( r \) decreases with a decrease of temperature unless the \( V \) formation enthalpy \( \Delta H_f \) is equal to or larger than 3.71eV, which, judging from Eq. (21b), is unlikely. In Fig. 6 we show the calculated values of \( r(T) \) assuming \( \Delta H_f \) is in the reasonable range of 1.5 to 3eV. It is seen that the effectiveness of the kick-out mechanism decreases much more rapidly with a decrease in temperature than that of the Frank-Turnbull mechanism.

With this discussion, it becomes understandable why Au diffusion can appear as dominated by the kick-out mechanism (related to \( I \)) at higher temperatures, while by the Frank-Turnbull mechanism (related to \( V \)) at lower temperatures. It is expected that other metals may also exhibit a similar kinetical behavior in Si for experiments involving their diffusion: in- and out-diffusion experiments, precipitation and precipitate dissolution experiments. We suggest that by now the following rule should be kept in mind: in an experiment if it is found that \( I \) in Si appears as dominating the diffusion process of an impurity, it means that \( I \) exists as a species of Si point defect under thermal equilibrium conditions and the experimental conditions are suitable for \( I \) to appear as dominating, it does not mean that \( V \) do not also exist as a species of Si point defects under thermal equilibrium conditions; the reverse is also true. The diffusion experiments of Kitagawa et al. (41) on Ni (dominated by \( V \)) and of Mantovani et al. (42) on Pt (dominated by \( I \)) are good examples.

4. ON THEORIES OF HIGH CONCENTRATION PHOSPHORUS DIFFUSION INTO SILICON

Diffusion of -III and -V dopants shows several anomalous features that are quite thorough discussed in the literature (43-47). To summarize, Willoughby (45) and Gösele and Strunk (47) mentioned that the most prominent features are: (i) The emitter push effect, which is the extremely rapid diffusion of base dopant (B or Ga) of the doubly diffused npn transistor structure which results in an enhanced movement of the base-collector junction; (ii) The kink-tail structure of the in-diffusion \( P \) profile, which is resulted from the use of high concentration surface sources. This is shown in Fig. 7a with conveniently designated surface and tail regions. In the tail region the \( P \) diffusion rate is much higher than expected for isoconcentration studies or OED studies; (iii) Movement of buried dopant layers can be extremely rapid if the surface \( P \) concentration is high. There exist ample evidences that \( P \) rich precipitates form during \( P \) in-diffusion (48-55), we therefore add to the above list by (iv) During \( P \) in-diffusion, the \( P \) concentration in the Si surface region exceeds its normal solubility limit. We consider that the following facts have been established: (i) The dopant diffusivity enhancement are caused by supersaturation of intrinsic point defects. The enhancement are much larger than that due to oxidations; (ii) The point defect supersaturations are related to high surface concentration of \( P \), and, in some cases, of \( B \) or \( As \) (56-58); and (iii) Mechanisms involve dislocations in an essential manner, e.g., dislocation climb, are not the origin of the point defect supersaturations.
Since until recently it was widely believed that V are the dominant thermal equilibrium point defect species in Si, earlier models (59-61) put forth to explain the P anomalous diffusion have been based on calculating the V supersaturations. Among these models, that of Fair and Tsai (61) gave a most complete treatment and was apparently able to calculate quantitatively the kink–tail structure in the P diffusion profile. Even if V were the only point defect species present in Si at high temperatures, the model still cannot be regarded as satisfactory because of: (i) As pointed out by Kroger (62), the analytical development is questionable with respect to the relation between the total P concentration and the total carrier concentration on the one hand, and with respect to the derivation of the electrical field on the other; and (ii) As pointed out by Mathiot and Pfister (63), the V supersaturation was accounted for by a multiplication factor for deriving the fast diffusing tail region P diffusivity instead of solving the V diffusion equation. By correcting these problems, Mathiot and Pfister (64) found that the model of Fair and Tsai can not give a good qualitative description of the P diffusion profile. Here we mention two fundamental objections generally applicable to these models: (i) Formation of P-rich precipitates (48-55) is not accounted for; and (ii) By now there also exist ample evidences that during P indiffusion an I and not V supersaturation is present in the Si interior (55,64-67) but the V based models can only predict a V supersaturation.

Recently, two new models have been proposed. One is a vacancy percolation (modified E-center) model due to Mathiot and Pfister (68), which has also been extended to include an I-component in the diffusivities of both dopant diffusion and Si self-diffusion (69). The other is a model due to Hu et al. (70) who proposed that a substantial part of P occupies interstitial positions which we shall call interstitial P model. Unfortunately, these two models seems are also not satisfactory. In the following (Sections 4.1 and 4.2) we discuss some essential points of these two models and give some detailed reasons for arriving at this conclusion. In section 4.3 we discuss a possible reason for P diffusing into Si.

4.1. The V Percolation Model

By correcting the problems of the model of Fair and Tsai and by incorporating the idea of V percolation (71), Mathiot and Pfister were able to quantitatively fit some P diffusion profiles (68). Later on they were also able to extend the model to include an I-component in the dopant and Si self-diffusion coefficients with an apparently even better success (69). Their earlier model (68) can be straightforwardly regarded as unsatisfactory since it is purely V based. The incorporation of an I-component in the dopant and Si self-diffusion coefficients (69) is a good step forward, but the model is not satisfactory because:

(i) Formation of P-rich precipitates is not accounted for;

(ii) They concluded that \( G < 0.1 \) to 0.3 for the dopants B, As and P from 900 to 1200°C for low and high concentration diffusions. For the cases of higher temperature experiments, we expect that local dynamical equilibrium between I and V to be reached, this can only result in a net V supersaturation in the Si interior. This is in contradiction with experimental results (55,64-67) that can only be explained by a net I supersaturation;

(iii) The finding that \( G < 0.5 \) for P, which requires that the observed ORD of P for (111) Si wafers oxidized for long times be explained without assuming a
net V supersaturation (as Francis and Dobson (72) and the present authors (8) have assumed), is in contradiction with the Sb OED result (10,19) which was observed under an almost identical experimental condition for observing the P ORD. A qualitatively consistent interpretation of these results is not obtained unless \( G_\text{c} > 0.5 \) for P and \( G_\text{c} < 0.5 \) for Sb and under the experimental conditions a net V supersaturation exists.

(iv) The observation of ORD of \( \beta \) at 1200°C by Hill (73) cannot be explained by their finding \( G_\text{c} < 0.5 \) for B, since in that experiment a 6% HCl was used. By analysing the OCE data of Shiraki (21) the present authors have shown that for this experimental condition a net V should have been injected (7,10). If \( G_\text{c} < 0.5 \) for B, then it should have resulted in an OED instead of the observed OPD.

The present authors believe that the V percolation idea is a very important point for which the full significance seems have not yet been realized. This subject is certainly a nice contribution of Mathiot and Pfister is well worthy of some discussions. To invoke the V-mechanism for the diffusion of a substitutional impurity species which exhibit a higher diffusivity than that of self-diffusion of the host crystal atoms, it is necessary to assume that the diffusion proceeds in the form of \( i-V \) pairs (\( i \) denotes a substitutional impurity atom), and a pair has to partially dissociate during its migration process. In Si the saddle point for the \( i-V \) pair dissociation process corresponds to having a V on the third nearest neighbour position of \( i \). The remaining binding energy between \( i \) and V determines the cohesion of the pair. For the case of having a very high impurity concentration, the impurity atoms effectively form a network or cluster in the crystal lattice in the sense that now each impurity atom would have some other impurity atoms situated at its fifth or even closer neighbour positions.

The \( i-V \) pair diffusion process changes now since upon reaching the saddle position the pair partial dissociation process, the \( V \) can also be at a third neighbour position of another \( i \). Now \( V \) can migrate to the other \( i \) atom with a potential barrier smaller than that needed for the isolated \( i-V \) pair case. This means with a sufficiently high impurity concentration, \( V \) can freely diffuse through the \( i \) network or cluster, which does not only enhance the \( V \) diffusion but also the \( i \) diffusion. This is the idea used by Mathiot and Pfister in their models (68,69).

We now mention a few points associated with the percolation idea that have not yet been explicitly considered:

(i) We believe that the percolation idea is equally applicable to an \( I \)-mechanism;

(ii) In the presence of a high impurity concentration, and hence the percolation network, we believe it is sensible to also suggest that the point defect formation energy is also lowered. This means the thermal equilibrium point defect concentrations are increased because of the presence of a high impurity concentration (not to be confused with that due to electrical charge effect);

(iii) The presence of the percolation network necessarily changes the \( i-i \) and \( i-V \) pairing potentials, and hence \( G_\text{c} \) for low and high impurity concentration cases are in principle of different values.
4.2. The Interstitial Phosphorus Model

Hu et al. (70) have recently criticized most of the V models (59-61) and proposed a new model to remedy the situation. An important contribution of their work is that formation of P-rich precipitate during P in-diffusion is very much emphasized and therefore in the model this is accounted for and hence removing one major shortcoming common to all other models. They have written down the basic equations but not yet performed the needed calculations and it is therefore not possible to see whether a net I supersaturation in the Si interior can result from the model or not. An examination of the basic assumptions involved in this model has led the present authors to believe that the model in its present form is probably also not satisfactory. Their major assumptions are: (i) A portion of the P atoms in Si occupies the substitutional sites (P_s) while the rest occupy interstitial sites (P_i); (ii) Mutual conversion between P_s and P_i is very slow; and (iii) Both P_s and P_i are shallow donors. Assumption (i) is needed since they invoked a two-stream diffusion mechanism as the a priori reason for obtaining the kink-tail diffusion profile. The best way to see what is a two-stream diffusion is to imaging that there are two distinct chemical species A and B diffusing into Si simultaneously but otherwise independently and we are looking for \( C_A + C_B = C_A + C_A \), see Fig. 7b. Now there is only one P species chemically but in Si they can occupy different sites, e.g., P_s and P_i. In so far as diffusion is concerned, P_s and P_i can behave independently. Assumption (ii) is needed since if P_s and P_i are converting into each other rapidly enough then their individual identity will be lost and a kink-tail structure in the P concentration profile can not be obtained. Assumption (iii) is needed since atomically dissolved P atoms in Si is known to be electrically active. The assumption that a substantial portion of the P atoms occupies interstitial sites is very different from that invoked conventionally: P atoms occupy substitutional sites only. By definition, P_diff use the I-mechanism. Thus, there will always be a substantial number of P atoms on interstitial sites whether they are migrating or not. In the conventional picture, P atoms can diffuse via an I-mechanism and hence those migrating P atoms are on interstitial sites temporarily but there will only be a negligibly small fraction of P atoms that are migrating at any instant in time and hence for practical purposes all P atoms are occupying substitutional sites.

In the following we present four arguments which jointly constitute as a strong indication that the above mentioned assumptions involved in the model of Hu et al. do not form a realistic enough basis for regarding the model as satisfactory (admittedly, though, each argument individually may be only constituting as a weak indication):

(i) The fact that there is but one shallow donor level associated with P doping is generally accepted. While it is legitimate to postulate that P_s is also a donor, it is difficult to believe that the donor is also a shallow one, and, in particular, its energy level needs to be almost identical to that due to P_i.

(ii) High concentration P diffusion into Si can induce the formation of a dislocation network to accommodate the Si lattice parameter change in the diffused layer due to the incorporation of P atoms to a large concentration. Analyses of such dislocation networks indicated that the P-diffused layer has a smaller lattice parameter (74) which is consistent with the knowledge that P atoms are all on the substitutional sites (the ratio of the covalent radi
of P atoms to Si atoms is 0.94). In order to explain the P diffusion profiles by the two stream mechanism, it needs to be assumed in some cases $C_p$ can be almost equal to $C_s$. If this is so, it seems that by now there should be some indication that the opposite, i.e., the P diffused layer has a lattice parameter larger than that of the undiffused Si, is true.

(iii) The requirement that $P_i$ and $P$ are converting into each other very slowly is an extremely stringent condition. Considering the diffusion times involved in a typical experiment, a simple estimate would show that the energy barrier against the $P_i$-$P$ conversion needs to be higher than 3eV and quite likely in the neighbourhood of 4 to 5eV. This is the same energy barrier prevents the interstitial pair from dissociation, and hence the pair needs to be regarded as very stable. If so, it is not easy to understand why $P$ can diffuse so fast, which is also required in the model.

(iv) In an Au gettering experiment, Lecrosnier et al. (75) found that the Au concentration in the P diffused layer developed a kink-tail profile similar to the P profile. This phenomenon is consistent with the interpretation that all P atoms occupy substitutional sites but not equally consistent with the interpretation that a substantial portion of P atoms occupies interstitial sites. The interstitial P model requires that the surface region P atoms be mostly P, (diffusion slow) while the tail region P atoms be mostly P, (diffusion fast). We first mention that in the Au gettering experiment all the gettered Au atoms should be Au, since normally $C_{\text{Si}} > C_{\text{Au}}$ holds and since if they were Au, then we expect the P and Au concentration profiles follow a complementary rather than a similar distribution pattern since the charge states of the P and Au atoms should result in a repelling force (P is a donor in spite of whether in substitutional or interstitial positions, and, Au is also suppose to be a donor, i.e., Au$^+$. Assuming that Au atoms are indeed Au atoms, then the best interpretation is that P atoms are all P atoms in $C_{\text{Si}}$ now charge compensation occurs between $P^+$ and Au$^{2+}$ (presumably Au$^+$. P atoms and atomic size compensation also occurs (covalent radius of Si atom is larger than that of P but smaller than that of Au$^+$). If the P atoms are P, then we expect a large compressive stress to develop in the Si which tends to result in a complementary distribution pattern between the P$^+$ and Au$^{2+}$ concentrations. The worst case is that both P and Au atoms occupy interstitial sites since then it will result in a compressive strain as well as a charge repelling situation.

4.3. Physical Reasons for $P$ Diffusing into Si in Supersaturation

While the kink-tail diffusion profile of P has been recognized as an outstanding feature for a long time, it is only recently that the equally outstanding (and undoubtedly related) feature of P diffusion into Si in supersaturation is also recognized. Why is it so? Hu et al. (70) provided one possible reason which is dynamical in nature: P diffusion (using a P-rich oxide source) is usually carried out in a partial oxygen ambient and hence oxidation, i.e., a chemical reaction, occurs. They suggested that the excess chemical potential responsible for P entering Si exceeding its solubility limit derives from this chemical reaction (oxidation). However, this may not be the primary reason because

---

*This problem was brought to the attention of one of the present authors (TYT) by Dr. T. E. Seidel.*

269
of: (i) The I supersaturation associated with high concentration P diffusion far exceeds that due to an oxidation of Si, and, in particular, this seems to be true at very low temperatures, e.g., 750°C, for which an oxidation of Si ceases to function; (ii) There are some indications that P can diffuse into Si in a supersaturated amount for cases that the annealing ambient is inert. Nibili et al. (54) concluded that the electrically active P concentration limit (presumably the solubility limit) is about $4 \times 10^{-3}$ cm$^{-3}$ at 1100°C, but at the same temperature P can enter Si to a concentration of about $1 \times 10^{-2}$ cm$^{-3}$ in an N$_2$ driven annealing using a P-rich oxide source (76).

As an alternative possibility, we suggest a mechanism due to phase-equilibrium requirements to explain the phenomenon of P diffusing into Si in supersaturation. This mechanism is readily understandable if the materials involved constitute as a binary alloy system. Therefore, for our present purpose we discuss the case that the source of diffusion is gaseous P (P$_g$) at the standard pressure instead of the more commonly used P rich oxide. In Fig. 8 we show the phase diagram of the Si-P system and an appropriate free energy diagram in the temperature range of our interest. The significance of Fig. 8 is that, under phase-equilibrium conditions: (i) the two primary phases, a (Si containing atomically dissolved P) and y (P$_1$, and the binary intermediate phase SIP, can each exist alone; (ii) The a phase and SIP can coexist; (iii) SIP and P$_g$ can coexist. This means for a diffusion experiment using P$_g$ as source, to start with, the materials (Si and P$_g$) do not exist in a phase-equilibrium state. Hence, the materials will proceed to reach the appropriate phase-equilibrium states by creating coexisting phases permitted by the phase-equilibrium conditions, i.e., the above discussed possibilities (ii) and (iii). We coined the term that, before the phase-equilibrium state is reached, the materials are in a diffusion-transient state. A primary phenomenon associated with this transient state is that P will diffuse into Si in excess of its normal solubility limit which is defined per coexistence of the a and SIP phases. Refer to Fig. 8b, the normal P solubility in Si is represented by concentration A. The material near the Si surface is now, however, not coexisting with SIP but with P$_g$ which gives a P solubility in Si represented by concentration B. Since B=A has to hold, we see P enters Si at a concentration exceeding its solubility limit A. That is, by the very fact that now P$_g$ and Si coexist, the material P$_g$ directly exerts a chemical potential on Si to maintain a P concentration in excess of the P solubility derived per coexistence of Si and SIP. If kinetical conditions permit, SIP precipitates can now form inside the Si crystal or alternatively a continuous layer of SIP can form at the Si surface. Before a sufficiently thick continuous SIP layer forms, the supersaturation phenomenon will persist.

Per this discussion, we see that there is a thermodynamic reason for P to diffuse into Si in supersaturation. For experiments for which a P rich oxide layer is used as source, the argument is also applicable. The analysis leading to this conclusion is, however, considerably more complicated than the binary example just discussed. This complexity arises because a Si-O-P ternary situation is involved. We have carried out the analysis for this ternary situation and the details will be given elsewhere (77). In that analysis we have also shown that the O solubility limit is also changed. There exists one experimental result (76) indicating that this seems to be the case.
5. CONCLUDING REMARK

In this paper some essential points associated with oxidation effects and Au diffusion in Si are discussed. Progresses in these areas in the last few years seems to have established beyond reasonable doubts that I and V coexist in Si under thermal equilibrium conditions. Within the framework of coexisting I and V species, some new understanding is now emerging. We look forward for more progresses in our understanding of the nature of point defects and their roles in the materials transport properties in Si in the near future.

REFERENCES

66. R. M. Harris and D. A. Antoniadis, J. Appl. Phys. to be published.

Fig. 1. Prediction of dopant diffusivity changes due to an oxidation. The model assumes that I and V coexist in Si in thermal equilibrium and in oxidation at high temperatures and that during the oxidation I and V attained local equilibrium.
Fig. 2. Quantitative fitting of available Sb ORD data from (100) wafers to the calculated $G_T = 0.02$ curve per Eq. (8). Unfilled circle: Mizuo and Higuchi [18]; filled circle: Tan and Ginsberg [19].

Fig. 3. The recombination barrier limited time constant, $\tau_{bar}$, for I and $\nu$ to reach dynamical equilibrium as a function of reciprocal temperature calculated per Eq. (11).

Fig. 4. Gold concentration profile after a hr diffusion into dislocation-free Si at 900°C [14]. Dash line: an erfc-function fitting per Eq. (19); solid line: fitting per Eq. (20).
Fig. 5. $D_{\text{C}}^\text{eq}$ and $D_{\text{V}}^\text{eq}$ vs. reciprocal temperature.

Fig. 6. The function $r(T)/r(1000^\circ C)$ vs. reciprocal temperature, $r$ is the ratio of the efficiency of converting Au into Au due to the kick-out mechanism to that due to the Frank-Turnbull mechanism. Numbers above each curve are the possible $\Delta E_V$ values.

Fig. 7. (a) Schematic drawing of the kink-tail structure of P profile, which, hypothetically, may be obtained by adding (b) profiles of physically distinguishable A and B atoms diffusing independently.
Fig. 5. The Si-P binary system phase diagram (a) and an appropriate free energy diagram (b) showing the thermodynamic origin for P to diffuse into Si in excess of its normal solubility limit. The solubility limits A and B are obtained from the usual common tangent constructions per coexistence of the α phase with the SiP and Pg phases respectively.
DISCUSSION

ROSE: On one of your viewgraphs you showed during the phosphorus diffusion, something like 40% of it is electrically active. I wonder if you might comment further on what physical form does it take?

TAM: It takes precipitated form. A small part of it, about 15%, is identified as SiP structure. About 85% of them are only observed in terms of no more than 15-A-sized precipitates. They are too small to be identified by transmission electron microscope. Therefore they are only inferred to be SiP precipitate, but not proven.