PROGRESS IN PREDICTING TRANSIENT DIFFUSION

Conor Rafferty
Bell Laboratories, Lucent Technologies

Abstract: Transient diffusion accounts for the majority of profile displacement in many modern processes. The last few years have seen considerable progress in understanding this anomalous diffusion, and predictive models have already been deployed in process development. Nevertheless, the evolving reduction in junction depth continues to challenge modeling capability. This work critically re-examines the theoretical basis of the present models and explores their limitations.

1. Introduction

Modeling transient diffusion has become essential to predicting modern device profiles. This work examines first the assumptions underlying a widely-used framework[1] describing defect-dopant interaction, and then reviews aspects of diffusion specific to transient effects.

2. Defect - dopant interaction

The diffusion of dopants in the presence of point defects takes place through pairing reactions of the form

$$A + I \rightarrow AI$$
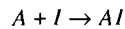
where the substitutional dopant species A is immobile but the paired species AI is mobile, as are the defects I. In addition one must consider that both defects and dopants can exchange charge with the background sea of electrons and holes generated by thermal generation across the bandgap. One might anticipate a large number of reactions and mobile species need to be tracked in order to predict dopant diffusion. In the worst case, one must solve $5mn$ partial differential equations, where m is the number of point defects (2 in silicon) and n is the number of dopants simultaneously diffusing, and 5 represents one neutral and two positive and negative charge states. However an analysis of the time constants of the reactions quickly suggests how the set of equations can be simplified without loss of accuracy.

2.1 Charging reactions

The charging and discharging of point defect species with the background sea of carriers can be represented by reactions such as

$$I + e^- \rightarrow I^-$$
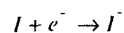
Standard reaction rate analysis gives for the rate of forward reaction where a is a capture radius of order the lattice con-

$$\partial I / \partial t = -4\pi a (D_I + D_e)[I][e^-]$$

stant and D_I, D_e are the diffusivities of interstitials and electrons, respectively. Neglecting D_I in comparison with D_e and using $[e^-] \sim n_i \sim 10^{18} cm^{-3}$ at 700C, the time constant of all such reactions is less than 1ps. For all practical purposes, defect and dopant charging can be considered instantaneous. The population of all charged species is therefore in instantaneous equilibrium with that of the corresponding uncharged species, and concentration equations need only be solved for either the uncharged species, or, equivalently, the total species (sum over all charge states).

2.2 Pairing reactions

Dopants are considered to diffuse as pairs with the native defects. By "pair" one considers both a bound combination of a defect and a substitutional dopant[1], or a substitutional atom promoted to an interstitial position by a kick-out reaction[2,3]. The reaction of pairs can be described in a similar fashion as for charging. Consider a dopant marker in the presence of a supersaturation of interstitials generated in some fashion, for instance TED. The reaction between dopant and defect can be written

$$\partial [A] / \partial t = -k_f [I][A] + k_r [AI]$$

If pairing is in steady state then at any instant the pairs can be computed from defect and dopant populations via $[AI] = (k_f / k_r)[A][I]$, and the full set of equations can be reduced from $5mn$ to just $m + n$, as is done in Suprem4[4], Prophet[5] and Floops[6].

The time constant for **all** dopant to be converted to pairs is

$$\frac{1}{k_f [I]} = \frac{1}{4\pi a D_I [I]} = \frac{1}{4\pi a D_I I^* S} = \frac{\tau_0}{S}$$

where I^* is the interstitial background concentration and S is the local supersaturation. Although separate values of the interstitial diffusivity and background are notoriously difficult to measure, values of their product agree closely even when taken from experiments of quite different type. Therefore the above time constant can be estimated with some confidence, giving a value of about an hour at 800C even if there is no supersaturation ($S = 1$). With a typical supersaturation, the time constant drops to under a second, but is still not negligible. On the basis of rate constants alone, therefore, one cannot exclude the need to take into account dynamic pairing.

The estimate needs to be reconsidered in the following light, however. The pairing reaction time constant represents the time to convert all the dopant to pairs. In fact the pairing reaction may reach its target when only a small fraction of the dopant has been converted to pairs; the time constant for this would be smaller by the same fraction. This brings up the question of what level of pairing is in steady state with defects and dopants, a separate and very interesting question

discussed in the next section.

2.3 Pairing equilibrium

The equilibrium ratio of paired dopant to unpaired dopant can be estimated as follows. In a situation with a flat supersaturation S of defects, the flux of dopant diffusing is equal to the flux of pairs diffusing, since all dopant moves as pairs. Thus $D_{AI}\nabla[AI] = SD_A\nabla[A]$ where D_A is the macroscopic equilibrium diffusivity of impurity A and D_{AI} is the diffusivity of a dopant pair. It follows that the ratio of paired dopant to unpaired is reciprocally equal to the ratio of their diffusivities:

$$\frac{[AI]}{[A]} = S\frac{D_A}{D_{AI}}$$

What limited evidence exists regarding the diffusivity of dopant pairs favors high values. First-principles calculation of BI pair motion[3] predicts a migration energy of only 0.4eV, hence a diffusivity ~12 decades faster than dopants. Measurements of boron interstitials show rapid migration at 450C[8], and room temperature migration of BI pairs has been reported[7]. At the very least, BI and PI pairs must diffuse 4 decades faster than B or P, since diffusivity enhancements of that order have been measured. Now if, for instance, pairs diffuse "only" six decades faster than impurity diffusion, then the paired fraction of dopant at any time cannot be more than $10^{-6}S$ and the overall time to pair would be $(\tau_0/S)10^{-6}S = 10^{-6}\tau_0$ or a few milliseconds at 800C (independent of supersaturation). One is drawn to the conclusion that the pairing reaction, like the charging reaction, rapidly reaches equilibrium, on a timescale much smaller than experimental timescales, and no separate equation needs to be solved for the pairs.

A more subtle question regards the equilibrium ratio of pairs to free defects:

$$[AI]/[I] = [A]k_f/k_r$$

Since defects are much less numerous than dopants, the paired fraction of defects might very well be expected to approach unity. Applying the previous pair-to-dopant relation to equilibrium conditions, one can relate the micro and macroscopic diffusivities $D_A = D_{AI}I^*k_f/k_r$. Eliminating reaction rates, the paired defect fraction is

$$\frac{D_A[A]}{D_{AI}I^*}$$

Progress is again blocked by the lack of information regarding pair diffusivities. If one speculates that $D_{AI} \approx D_I$, the denominator can be estimated and one finds that the paired fraction of defects is small at high temperatures but approaches unity around 800C, due to the higher activation energy of self diffusion compared to dopant diffusion. Thus

the question of whether pairs are dilute with respect to defects at low processing temperatures hinges on the question of whether the pairs diffuse faster or slower than free interstitials, and remains open at present.

2.4 Pair diffusion lengths

An intriguing deduction can be made regarding the individual diffusion hops of a diffusing dopant. The lifetime of a pair is the reciprocal of the reverse reaction rate $\tau_{AI} = 1/k_r$. The distance which the pair travels in this time is on average $\lambda = \sqrt{D_{AI}\tau_{AI}}$. Substituting the forward reaction rate $k_f = 4\pi a D_I$, one obtains

$$\lambda = \sqrt{D_A/4\pi a D_I I^*}$$

Surprisingly, the pair diffusion length can be calculated in terms of macroscopic, measured quantities without any adjustable parameters. It is something of a relief then to find that substituting values for boron gives an individual hop distance of about 37A at 800C! Not only is this self-consistent with treating diffusion at a device length scale in continuum form, it agrees well with experimental estimates of the individual hop distance[8]. The predicted value at 625C is about twice that measured[8] which might be due to extrapolating both D_A and $D_I I^*$ well outside their measured range, or to experimental difficulties estimating λ from very slight amounts of diffusion.

2.5 Basic Diffusion Model

Having demonstrated that the charging and pairing reactions rapidly reach equilibrium, and that pairs are dilute with respect to dopants, the original Fahey-Griffin-Plummer analysis[1] can be applied with confidence to obtain the diffusion flux of dopant in a gradient of potential and defects:

$$F_A = -D_A(\psi)\left[\left(\frac{I}{I^*}\right)(\nabla A + s_A\nabla\psi) + A\nabla\frac{I}{I^*}\right]$$

where ψ is the normalized electrostatic potential, s_A is the sign of the ionized dopant, and $D_A(\psi)$ is the macroscopic dopant diffusivity as a function of potential. The origin of the defect gradient term is the product rule of calculus;

$$D_{AI}\nabla[AI] = D_{AI}(k_f/k_r)(I\nabla A + A\nabla I)$$

Efforts to dispense with this term are tantamount to disputing the product rule. An interesting feature of the flux term is that the overall flux is proportional to the ionized dopant charge *even though some defect-dopant pairs may be neutral*. The reader is referred to the appendix of [1] to see how this comes about.

2

3. Transient diffusion (TED)

It is now well established that immediately after implantation, impurities show a transient burst of diffusion which may be enhanced several thousand-fold relative to their equilibrium diffusion rate. The diffusion which occurs during this burst can exceed all other thermal contributions during processing, an occurrence which is exacerbated by the trend towards small thermal budgets. Furthermore, transient diffusion is frequently cooperative in nature; the implantation of arsenic may drive the diffusion of existing boron at some distance away. For these reasons it has become essential to model TED in order to predict device profiles.

3.1 The plus-one model

At first glance, the prediction of defect evolution after implantation might appear very difficult. Large numbers of point defects are created, and they can coalesce into various forms of extended defect. However there has been considerable success modeling TED using a very simple model of the damage, the "plus-one" model. The idea behind plus-one[9] is that implantation creates large but equal numbers of interstitials and vacancies (Frenkel pairs), as well as one extra interstitial for every implanted ion as it becomes substitutional. The equal and opposite I,V distributions recombine with one another, and the dominant contribution to TED comes from the extra ions, whose distribution is exactly that of the implanted species. This picture might sound oversimplified, but it has recently received strong experimental support. Jones[10] showed that for sub-amorphous implants, the amount of transient diffusion caused by a Si implant was independent of both beam current and substrate temperature. Since both factors affect the number of Frenkel pairs remaining after the implant, it supports the notion that the number of pairs created by implantation is not a primary variable in TED. Giles[11] implanted phosphorus at normal incidence and at high tilt angle, with the energy adjusted to make the ranges coincide. The higher energy tilted implant necessarily generates many more Frenkel pairs, yet the TED was again the same in both cases. Finally, Jones[12] also showed that even for amorphizing implants, the TED in the substrate was independent of beam current and substrate temperature. The last result is particularly surprising as the thickness of the amorphous layer is known to vary with beam conditions, yet the same number of interstitials was released to drive substrate diffusion. The conclusion from the present experimental data is that TED appears to be insensitive to the details of Frenkel pair generation during implantation, and mainly sensitive to the distribution of implanted ions.

This convenient scenario may need revision soon. Most TED experiments have been carried out using boron or phosphorus since they are interstitial diffusers and experience TED. Arsenic does not experience TED, but it certainly induces TED in other profiles. For a heavy ion such as arsenic, the distribution of lattice recoils may exceed that from the plus-one contribution. Figure 1a illustrates the relative concentrations of ions and net interstitials and vacancies for a 80keV P implant; the ion distribution dominates. In Figure 1b, however, a 50keV As implant shows a distribution of recoils

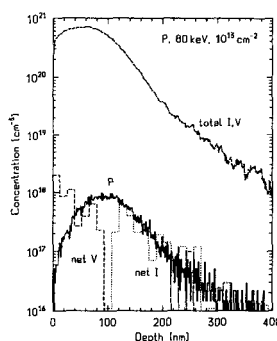which exceeds the ion distribution in both depth and dose.

Figure 1a - distribution of ions and recoils after P implantation
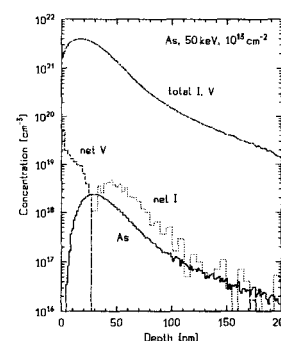
Figure 1b - distribution of ions and recoils after As implantation

Contributions of more than plus-one also arise for low energy P and BF2 implants[14].

In summary, the plus-one model has had substantial success in predicting the amount of TED, but future work will require refinement for heavy ions and low energy.

3.2 Interstitial clustering

The discovery of interstitial clusters was the jigsaw piece that brought all the pieces of the TED puzzle together. Based on Michel's cluster hypothesis[15], a solubility model of clustering was adopted as long ago as 1992 for modeling TED in Prophet[16]. However experimental support was lacking until the discovery [17] that the lifetime of {311} defects closely correlated with the duration of TED, and furthermore that the dose of interstitials contained in {311} defects approximated the implant dose (plus-one).The existence of clusters emitting interstitials at a steady rate explains the quasi-constant enhancement during TED, its independence of dose and energy, and its increasing *duration* with increasing dose and energy[18]. Given the centrality of clusters to transient diffusion, it is surprising to note that the final transient displacement is independent of the clustering model used. Figure 2 shows the local time-averaged interstitial supersaturation after a (2e13,50keV) silicon implant, either ignoring clusters, with a 1-moment cluster model[18], or with a 2-moment cluster model[19]. The identical curves result from the fact that each clustered interstitial eventually escapes to make its contribution to TED. Of course, if a transient does not run to completion, the details of cluster evolution will influence the results. Nevertheless, one can see that for the narrow purpose of computing diffusion, the details of clustering have a second order influence.

The figure also demonstrates another unusual feature of TED. The shape of the enhancement curve is identical for different doses and energies, apart from a scale factor; it is a function of temperature only. To first order, all implants cause the same enhancement curve and differ only in magnitude; an exception is at high energy when the shoulder of the enhancement curve shifts deeper into the bulk, to approximately the range of the implant.
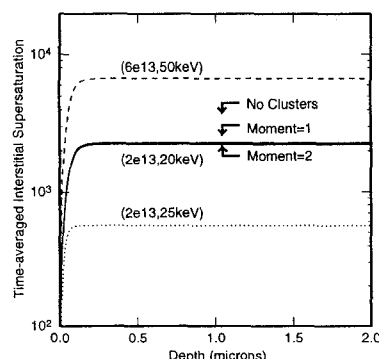
Figure 2: time average supersaturation depends primarily on implant parameters and secondarily on details of cluster model.

3.3 Dopant clustering

A striking feature of profiles following implantation is that previously substitutional dopant may become non-substitutional, as evidenced by loss of electrical activity and inability to diffuse[19]. Early models of this phenomenon focussed on 2-3 atom clusters involving interstitials and dopants. Recent data shows that a small dose of interstitials can immobilize several times its own dose of boron, suggesting that while interstitials are necessary to initiate the clustering reaction, and may be consumed by the reaction, the clustering can then proceed without being limited by interstitial supply. (Figure 3). An atomistic model capturing this behavior has

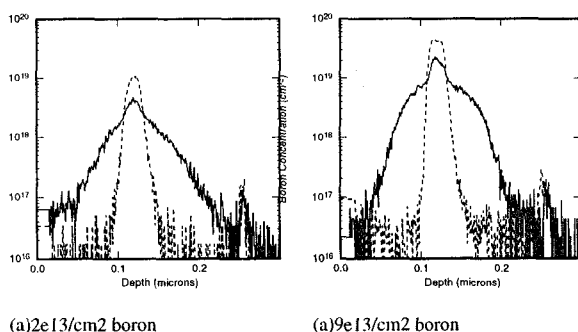(a)2e13/cm2 boron   (a)9e13/cm2 boron

Figure 3: A high dose marker shows more, not less clustering, when driven by TED from a low dose implant; clustering is not limited by the interstitial supply. TED conditions are (2e13,40keV) Si annealed for 35min at 800C.

been presented[21], based on the concept that cluster precursors are formed during implantation, which subsequently capture boron during anneal.

3.4 Challenges in TED

The greatest challenge in TED modeling at present comes from matching high dose, low energy implants. At high doses, amorphization and dislocations must be taken into account, and extrinsic effects on diffusivity and background need to be characterized. Meanwhile low energy greatly increases the influence of the surface, in particular, dose loss effects[22]. Even at modest dose, low energy TED is poorly predicted by present models. Figure 4[23] shows the effects of dose loss for a (1e14,30keV)P implant, annealed at 800C

for 30 min. Although the tail diffusion is well captured, some process has "scalped" the peak region of the profile. The logarithmic scale disguises the fact that half the implanted dose is missing, even though all known effects have been taken into account including surface trapping and defect gradients.
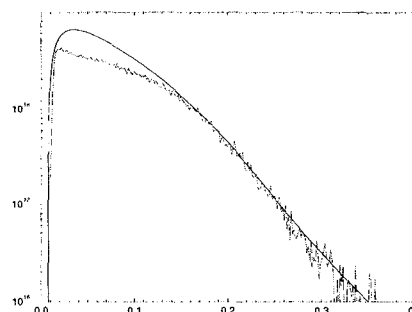

Figure 4: Measured and simulated profiles for low energy P implant annealed at low temperature.

4. Summary

Transient diffusion has come to dominate the thermal displacement of implanted dopants. Modeling the dopant-defect interaction can be reduced from a large set of equations to one for each dopant and each defect, though low temperature processing may require relaxing the dilute approximation. The simple plus-one model of damage has received surprisingly strong experimental support, and when combined with the diffusion model, satisfactorily accounts for profile displacement at medium dose and energy. High dose, low energy profiles will require improvements in models of damage, surface loss of dopant, and dopant clustering.

References

[1] P.M. Fahey, P.B. Griffin, J.D. Plummer, Rev. Mod. Phys v61 p289 (1989)
[2] C.S. Nichols et al, Phys. Rev. B v48 p.5848 (1989)
[3] J. Zhu et al, Phys. Rev B. v54 p4741 (1996)
[4] M.E. Law, C.S. Rafferty, R.W. Dutton, IEDM p518 (1996)
[5] M.R. Pinto et al, IEDM p923 (1992)
[6] H.Y. Park, K.S. Jones, M.E. Law, J. Electrochem. Soc. v141 p759 (1994)
[7] N.E.B. Cowern et al, Abstract E4.1, MRS Spring Meeting 1997
[8] N.E.B. Cowern et al, Phys. Rev. Lett v69 p116 (1992)
[9] M.D. Giles, J. Electrochem. Soc. v138 p1160 (1991)
[10] K.S. Jones et al, Abstract E3.6 MRS Spring Meeting 1997
[11] M.D. Giles et al, Abstract E1.2 MRS Spring Meeting 1997
[12] K.S. Jones et al, Abstract E1.4 MRS Spring Meeting 1997
[13] G.H. Hobler, Abstract 366, ECS Spring Meeting 1996
[14] G.H. Hobler et al, in preparation
[15] A.E. Michel, NIMB v37/38 p379 (1989)
[16] C.S. Rafferty, M.D. Giles, unpublished (1992)
[17] D.J. Eaglesham et al, Appl. Phys. Lett. v65 p2305 (1994)
[18] C.S. Rafferty et al, Appl. Phys. Lett. v68 p2395 (1996)
[19] G.H. Hobler et al, this proceedings (1997)
[20] P.Stolk et al, Appl. Phys. Lett. v66 p568 (1995)
[21] L. Pelaz et al, Abstract E6.4 MRS Spring Meeting (1997)
[22] P.B. Griffin et al, Appl. Phys. Lett. v67 p482 (1995)
[23] R. Kasnavi, private communication