Kinetics of Dopant Precipitation in Silicon

Scott T. Dunham
Department of Electrical, Computer and Systems Engineering
Boston University, Boston, MA 02215

Above a temperature-dependent concentration, substitutional silicon dopants are electrically inactive as well as immobile due to aggregation of the excess dopant into precipitates or clusters. Since electrical properties depend primarily on the active dopant concentrations, understanding and modeling of the activation process is critical for the prediction of device behavior. Despite the importance of aggregation processes, the kinetics of dopant activation are usually ignored and instead a simple solid-solubility cutoff or steady-state clustering is assumed. However, this approach cannot account for experimental observations such as active dopant concentrations above solid solubility or the dependence of activation on ramp-down rate. Activation kinetics have recently been addressed using clustering models, but these models are inconsistent with a number of experimental observations. In this work, we describe a kinetic precipitation model which accounts for the activation/deactivation behavior of the major dopant species in a consistent manner, including phenomena which is often associated with clustering models.

Since precipitate behavior is a strong function of size, in order to model activation/deactivation due to precipitation, it is necessary to consider the distribution of precipitate sizes as a function of time $f_n(t)$. For solute concentrations above solid solubility, there is a free energy gain upon forming a separate solute-rich phase. However, there is an energy barrier to forming the separate phase due to surface and strain energy associated with finite precipitates. The resulting free energy change upon precipitate formation can be written as

$$\Delta G_n = -nkT \ln\left(\frac{C^i}{C^{ss}}\right) + \Delta G^{\text{excess}} \qquad (1)$$

with n the number of solute atoms in the precipitate, C^i the concentration of solute near the precipitate/matrix interface, and C^{ss} the solid solubility of the solute. The excess energy is due to interface and strain energy and typically becomes proportional to the precipitate surface area for large sizes. However, for small precipitates, the excess energy can be expected to deviate strongly from this asymptotic behavior so we use an expression of the form $\Delta G^{\text{excess}} = a_1 n^{m_1} + a_2 n^{m_2}$, where m_1 m_2 and m_1 controls the asymptotic behavior and is given by the dependence of precipitate area on size (23 for spheres).

The evolution of the precipitate density can be described by a series of rate equations for each precipitate size:

$$\frac{df_n}{dt} = I_{n-1} - I_n \quad n \geq 2; \qquad \frac{df_1}{dt} = -I_1 - \sum_{n=1}^{\infty} I_n \qquad (2)$$

In order reduce the number of equations, for the larger sizes, we consider the system to be nearly continuous and rediscretize, obtaining simulation results indistinguishable from the solution to the full set of discrete equations. The rate of growth from size n to size $n+1$ is given by

$$I_n = g_n - d_{n+1} = k_n A_n (C^i f_n - C_n^* f_{n+1}) \qquad (3)$$

where k_n is the interface reaction rate and A_n is the active precipitate interface area. We assume that the barrier for hopping of a dopant atom to a site adjacent to a precipitate is the same as the bulk hopping barrier so that $k_n = D\lambda$, where D is the diffusivity and λ is the lattice spacing.

We also need to recognize that the far field concentration and the local concentration are different. This will be particularly important when the net growth or dissolution rate is high. Rather than the typical steady-state assumption, we approximate the distribution of solute atoms in the neighborhood of the precipitate by the distribution that would be present if the precipitate had always been its current size. This is always a better approximation, particularly when examining the short-time transient behavior. In this analysis, the interface concentration is given by:

$$C_n^i - C^\infty = \frac{C_n^* - C^\infty}{1 + Dkr} + [C(0) - C^\infty] \exp\left(h'^2 Dt\right) \operatorname{erfc}\left(h'\sqrt{Dt}\right)$$

where $h' = kD + 1r$ and $C^\infty = f_1$ is the bulk solute concentration.

To account for the variation of diffusivity with active doping level, we use (for donors):

$$D_{\mathrm{A}} \cong D_{\mathrm{A}}^0 + D_{\mathrm{A}}^- (nn_i) \; + D_{\mathrm{A}}^{\mathrm{H}} (C_A C_{\mathrm{ref}})^{\; n_H} \qquad (4)$$

The first two terms are the standard Fermi level dependent diffusivity, while the last term accounts for the dramatic increase in diffusivity observed above $10^{20}\mathrm{cm}^{-3}$ for group V and IV elements by Larsen *et al.* [1] ($n_H = 45$ for As). This effect is particularly significant for deactivation studies as the active doping level is above the solid solubility. For our simulations, we use activated parameters based on published work [1, 2]. Since there is insufficient data to calculate an activation energy for the last term, we keep it as a fitting parameter, while matching experiments at 1050°C. For solid solubility, we use a temperature activated expressions derived from published data [3].

We have conducted a series of simulations utilizing the model described above, and we find that the model is able to predict observed behavior, including effects that are inconsistent with both standard clustering as well as classical nucleation and growth models. Figure 1 shows the comparison of the model to isochronal arsenic annealing data of Nobili *et al.* [4] for two different implant doses, while Figure 2 shows the striking behavior observed for a two-step anneal. Both sets of data were fit using a single set of parameters (dependence of excess energy on size and activation energy for last term in Eq. 4). The model replicates well the observed behavior including the must faster deactivation of the heavier dose (including retrograde regime) and the reverse annealing following a temperature step. The cause of the latter effect is illustrated in Figure 3, which shows precipitate densities before and after the temperature step. Since the model considers subcritical as well as stable precipitates (in contrast to classical nucleation and growth models), it also accounts for incomplete activation for concentrations below solid solubility as observed for arsenic and boron and often used to justify the use of clustering models. Similar agreement between experiments and simulations are found for other dopants. As an example, Figure 4 shows a comparison to data from Nobili *et al.* [5] for a series of isochronal anneals. In summary, we show that precipitation can account for the kinetic activation/deactivation behavior observed for silicon dopant species in a consistent manner.

References

[1] A. N. Larsen, K. K. Larsen, P. E. Andersen and B. G. Svensson, *J. Appl. Phys.* **73**, 691 (1993).

[2] R. B. Fair, in **Impurity Doping Processes in Silicon**, ed. F. F. Y. Yang (North-Holland, New York, 1981) p. 315.

[3] G. Masetti, D. Nobili and S. Solmi, in **Semiconductor Silicon**, eds. H. R. Huff and E. Sirtl (Electrochemical Society Proc. **77-2**, Pennington, NJ, 1977) p. 648.

[4] D. Nobili, A. Carabelas, G. Celotti and S. Solmi, *J. Electrochem. Soc.* **130**, 922 (1983).

[5] D. Nobili, A. Armigliato, M. Finetti and S. Solmi, *J. Appl. Phys.* **53**, 1484 (1982).

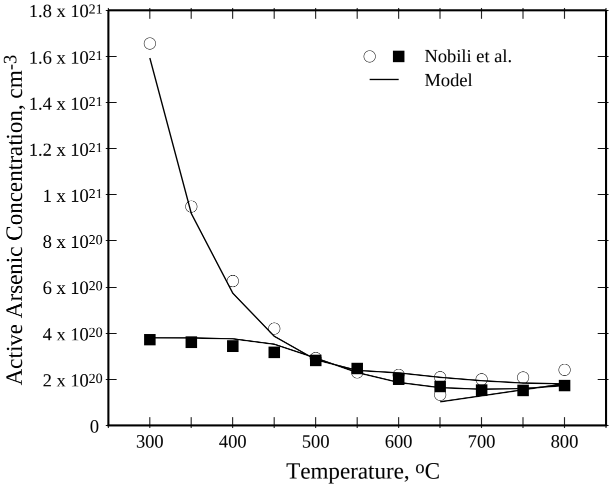

Figure 1: Peak active arsenic concentration for isochronal annealing of implants with initial peak doping concentrations of $3 \times 10^{21}\mathrm{cm}^{-3}$ and $4 \times 10^{20}\mathrm{cm}^{-3}$. Experimental results are from Nobili *et al.* [4]. Following laser anneal which activated the entire implant, the samples were annealed for 1h at each temperature from 300°C to 800°C in 50°C. Both the model and the data manifest much more rapid precipitation for the higher dose. This is because the critical size and thus the maximum activation barrier to form critical nuclei decrease with supersaturation. Following the last isochronal anneal at 800°C, a sample was subsequently annealed for 30h at 650°C, demonstrating that the system remains well above solid solubility.

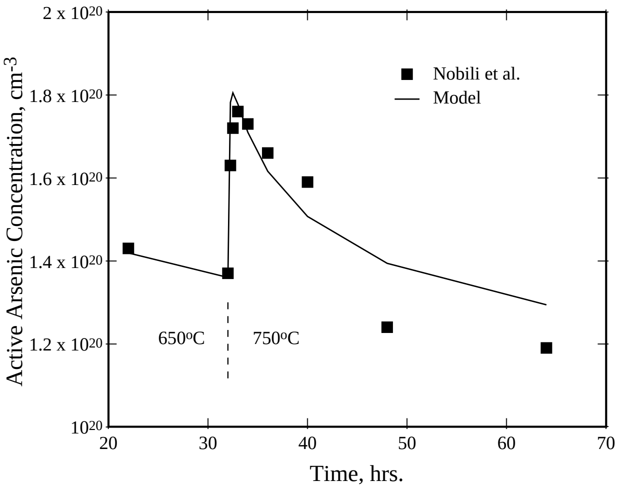

Figure 2: Peak active arsenic concentration versus time for two-step anneal first at 650°C for 32h and then at 750°C. Initial peak doping is $9 \times 10^{20}\mathrm{cm}^{-3}$. Experimental data from Nobili *et al.* [4]. Note reverse annealing following temperature change at $t = 32$h. Despite the fact that the active doping remained well above solid solubility (11×10^{20} at 750°C) the precipitation process initially reversed itself following the temperature step before eventually continuing at a faster rate characteristic of the higher temperature.

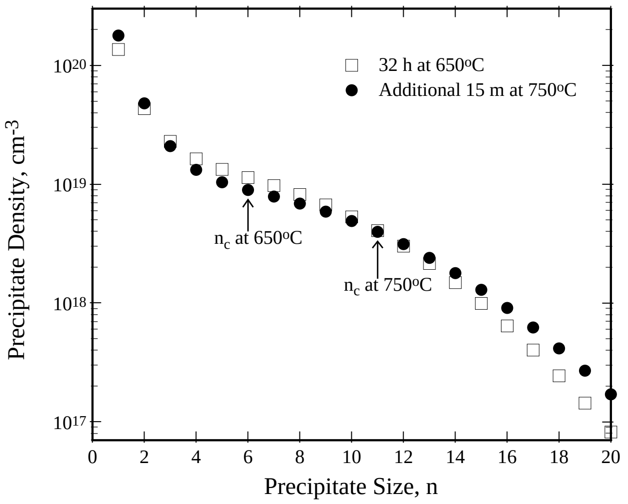

Figure 3: Simulated precipitate size distribution at the end of a 32 h anneal at 650°C and after a subsequent 15 min anneal at 750°C. Upon increasing the temperature from 650 to 750°C, the solid solubility increases, reducing the supersaturation and thus the critical precipitate size. Therefore, precipitates which previously were growing, now shrink and disappear. The result is an initial increase in the active doping concentration. In contrast, the precipitates which remain stable continue to grow and eventually dominate leading to dopant deactivation at longer times. It may be noted that this type of behavior (reverse and then forward annealing) is only possible when there is a distribution of precipitate sizes such that some shrink while others grow, rather than a single dominant precipitate size as assumed in clustering models.

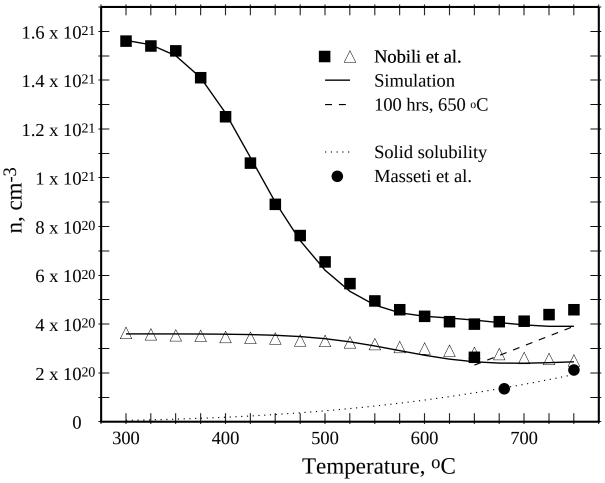

Figure 4: Peak active phosphorus concentration for isochronal annealing of implants with two different initial doping levels compared to kinetic precipitation model. Experimental results are from Nobili *et al.* [5]. Also shown for comparison are solid solubility data from Masetti *et al.* [3].