Understanding And Modeling Ramp Rate Effects On Shallow Junction Formation

Srinivasan Chakravarthi,[a] Alp H. Gencer,[b] Scott T. Dunham[c] and Daniel F. Downey[d]

[a]Department of Manufacturing Engineering, Boston University, Boston, MA 02215.

[b] Avant! Corp., TCAD Division, Wellesley, MA.

[c]Department of Electrical Engineering, University of Washington, Seattle. WA. 98195.

[d] Varian Semiconductor Equipment Associates, Gloucester, MA 01930.

The achievement of ultra-shallow source/drain extensions requires low energy implants combined with very short time anneals in order to minimize diffusion. In this paper, we demonstrate the ability of models which include accurate extended defect kinetics to predict the effect of changing ramp rates on final junction profiles.

INTRODUCTION

Transient enhanced diffusion (TED) has been the dominant effect in determining junction depths for the past decade. Because TED effects decrease with temperature, the use of rapid thermal processing (RTP) with fast ramp rates has become an important tool for achieving shallow junctions. Fabrication of sub-40nm source/drain extensions requires low energy implants along with fast-ramp spike anneals to minimize diffusion. In addition to reducing implant depths, TED effects also reduce with implant energy because the proximity of the surface provides faster annihilation of excess interstitials from implant. Ramp rates are particularly critical for lower implant energies. Sub-1keV implants show a more pronounced decrease in junction depth with an increase in ramp rates, while the reduction in junction depth for higher energy implants has been found to saturate for larger ramp rates.[1,2] The goal of this work is to understand and model the effect of RTP annealing cycle on junction behavior so that RTP processes can be optimized for shallow, low resistance junctions.

MODELING RAMPS

Fig. 1 shows a schematic illustration of the temperature versus time profile during a ramp-up. For the simulations, a linear ramp-up is assumed. After the lamps are turned off, it is assumed the wafer temperature drops chiefly by radiating heat to the furnace walls and hence the ramp-down rate decreases at lower temperatures. During the cool down, the heat loss is proportional to T^4, so:

$$T = T_f \left(1 - \frac{3tR_{\text{cool}}}{T_f}\right)^{-0.33} \tag{1}$$

where, T_f is the maximum attained temperature, R_{cool} is the cooling rate (K/s) at T_f (K) and t is time elapsed after reaching T_f.

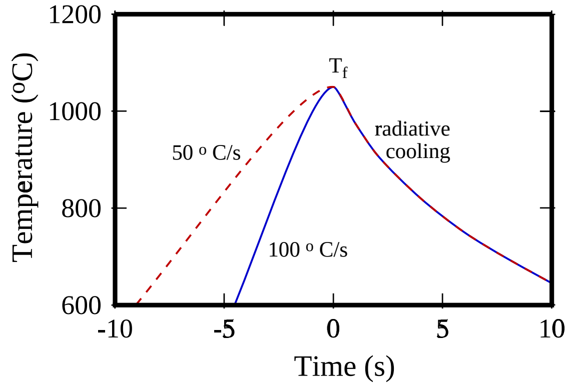

Fig. 1: Temperature versus time profiles used for the simulations. Ramp-up is modeled as linear whereas cool down is radiative.

INTERSTITIAL EXTENDED DEFECTS

For {311} defects, we use the analytic kinetic precipitation model (AKPM) which is a computationally-efficient version of the moment based model developed in our previous work.[3] The equations describing the evolution of the extended defect population in the Analytic Kinetic Precipitation Model (AKPM) are:

$$\frac{\partial m_0}{\partial t} = I_1 = D_I \lambda \left[C_{\mathrm{I}}^2 - m_0 C_{\mathrm{ss}} \gamma_0 \right] \tag{2}$$

$$\frac{\partial m_1}{\partial t} = 2I_1 + D_I \lambda m_0 \left[C_{\mathrm{I}} - C_{\mathrm{ss}} \gamma_1 \right], \tag{3}$$

where m_0 and m_1 are the zeroth (number of clusters) and first (number of clustered atoms) of the defect distribution. The parameters γ_0 and γ_1 are functions of the average size (m_1/m_0) and were derived based on experimental knowledge about the size distribution of {311} defects.[3] We also compare this model to a simple solid solubility model (SSS) with diffusion limited rates. These models have been calibrated to {311} and TED data to yield the same results for long time, constant temperature anneals. However, the AKPM model includes the changes in interstitial super-saturation as the {311} population ripens, thus giving greater TED at short times and less as the anneal progresses. Fig 2(b) shows comparison to experimental enhancement data (from Chao *et al.*[5]) of boron marker layer during TED from a Si implant using the AKPM and SSS models.

We begin by analyzing the effect of extended defect kinetics on junctions formed via short-duration RTP. To this end, we compare the predictions of the two different models for {311} defect evolution to data from Perozziello[6] for Si implants into MBE-grown boron layers and subsequent spike anneals. As is evident from Fig. 3 and 4, the AKPM model does a better job in predicting experimental data. For fast ramp rates, the difference between the two models decreases as we go to higher final temperatures. The reason is that for higher temperatures, a spike anneal has sufficient thermal budget to anneal out all damage, completing TED. In fact, if we do a fast ramp-up and long soak at 825°C, the two models are equivalent (Fig. 5(b)). However, for slow ramp rates, even with a long soak, the difference between the two models is persistent since much of the TED occurs during the ramp up (Fig. 5(a)).

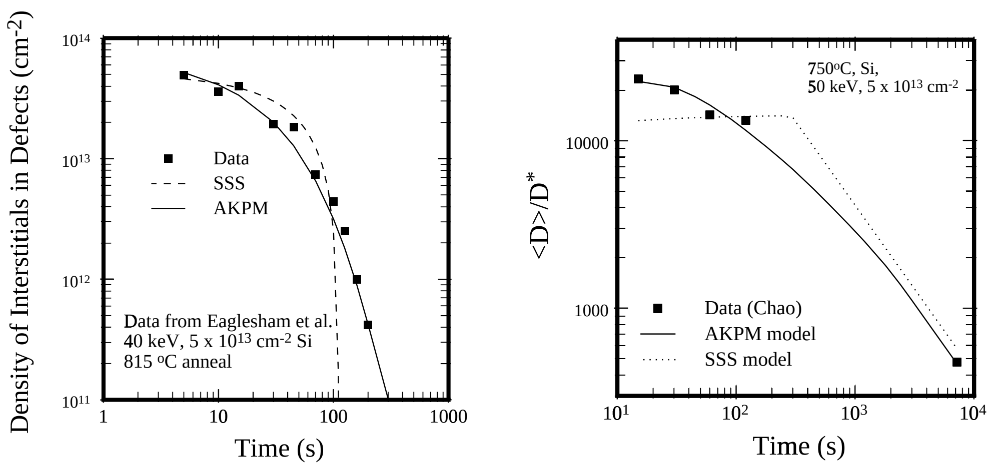

Fig. 2: Comparison to experimental data with simulation with AKPM and SSS a) {311} defects dissolution data from Eaglesham *et al.*[4] (b) Enhancement data from Chao *et al.*[5]

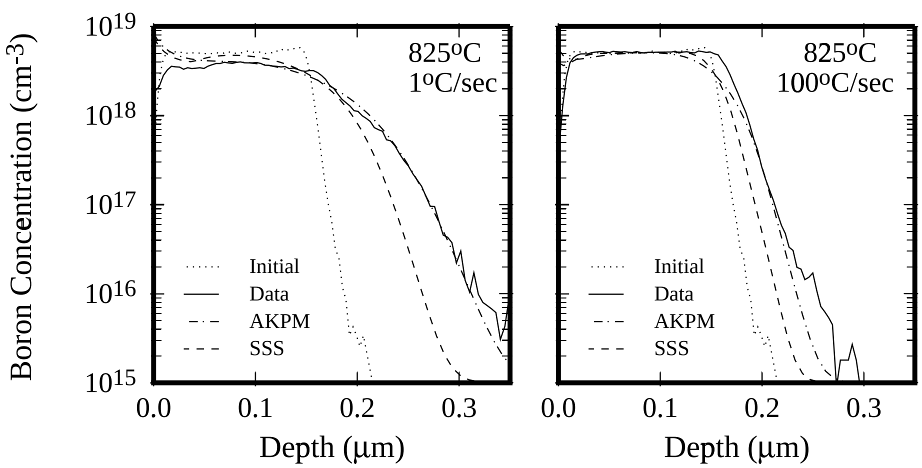

Fig. 3: Diffusion during 825°C spike anneals with ramp-up rates of (a) 1°C/s and (b) 100°C/s of MBE-grown boron marker layers following 5×10^{13} cm^{-2} Si implantation at 35 keV to create damage. Simulation with AKPM and SSS models and comparison to data from Perozziello.[6]

The AKPM model matches the experimental data because it captures the ripening of interstitial clusters. Smaller clusters maintain a larger super-saturation relative to larger, more stable clusters. Thus, the early stage of TED give the most diffusion, and if more of this stage occurs at lower temperatures (as for slower ramp-up rates), larger diffusion enhancements are seen, just as lower temperatures give greater TED for soak anneals. The effects of relatively slow ramp-up persist even for long soak anneals since during the longer time spent at low temperature, the ratio of dopant diffusion to damage elimination is larger. In short, accurate extended defect kinetics are critical whenever a large portion of TED occurs during the ramp-up or ramp-down, and thus a model (such as AKPM) which includes accurate extended defect kinetics should be deployed for such complex thermal cycles. Note that for a 825°C anneal, 100°C/s can be considered very fast, but for higher temperature spikes for which diffusion kinetics are faster, ramp rate effects persist to much higher ramp rates.

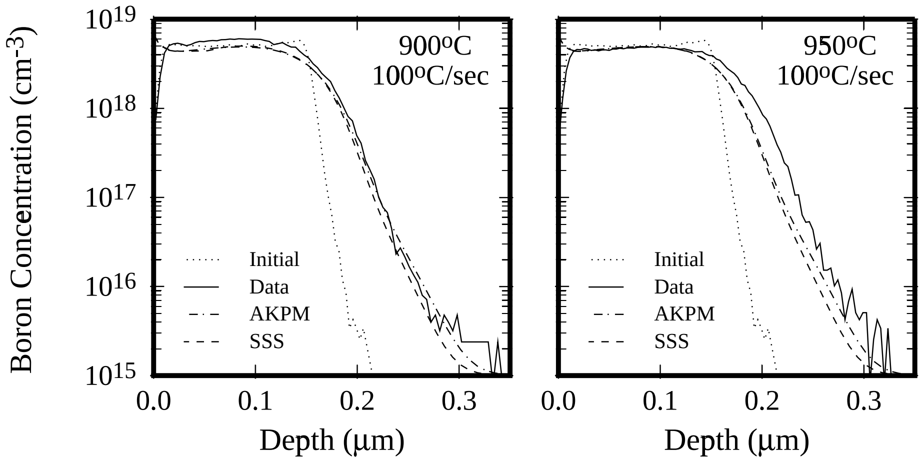

Fig. 4: Spike anneals of MBE-grown boron layer with subsequent 5×10^{13} Si implantation at 35 keV to create damage. Simulation with AKPM and SSS models and comparison to data from Perozziello.[6]

LOW ENERGY B IMPLANTS

For high concentration boron source/drain extension implants, we also need to include boron/interstitial cluster models. Small boron clusters form at lower temperatures during the ramp-up. The primary boron clusters (B_3I and BI_2) are modeled as in previous work.[7] Further boron deactivation is given by a simple solid solubility. This model has been shown to give good agreement to high temperature soak anneals where hold time dominates the thermal cycle.[8] Since the experiments considered were performed under low ambient oxygen levels, oxidation enhanced and retarded diffusion was modeled as in previous work.[8]

First we look at the qualitative behavior of junction movement due to spike anneals for ultra low energy B implants. The simulations predict a reduced effect of ramp rate on x_j (as seen experimentally)[1,2] and saturation in the junction depth with increasing ramp-up rates (Fig. 6(a)). This saturation occurs at larger ramp rates for lower energy implants. For higher energies, the faster ramp-ups retain more of the interstitial clusters and more of the TED actually occurs during the ramp-down, thus negating any reduction possible due to the faster ramp-up. In fact, for higher energies, junction depth increases slightly with increased ramp rates, as TED persists to lower temperatures during the ramp-down. For very fast ramp-up rates, the ramp-up is only a small fraction of the total thermal budget, and hence the cooling rate dominates the total junction movement. Thus, as shown in Fig. 6(b), higher ramp-down rates are effective in reducing junction depth, if the ramp-up rate is fast enough that TED is not completed. Fig. 7 shows comparison to RTP anneals from Downey et. al.[1] using two different ramp-up rate and three different implant energies. The simulations give a good match to the observed boron junction depths across the spectrum of conditions.

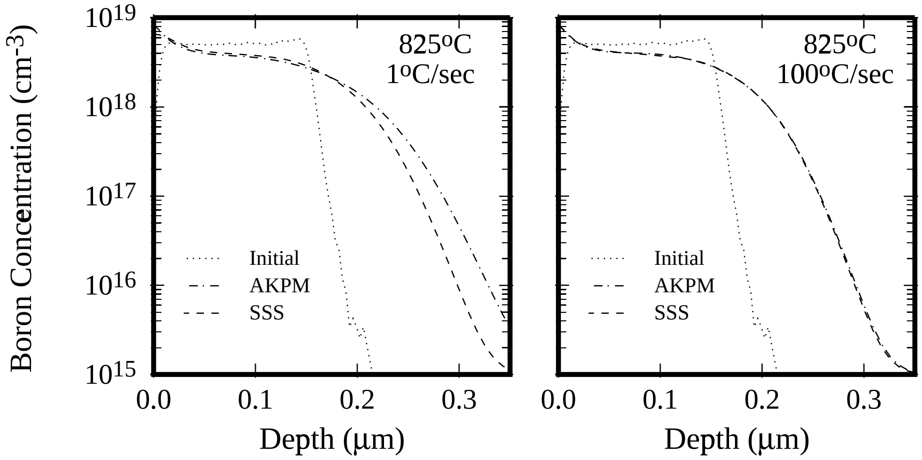

Fig. 5: Simulation of same initial conditions with a long time soak at end temperature. Since SSS and AKPM have both been calibrated to long time TED data, they produce same results in fast ramp case.

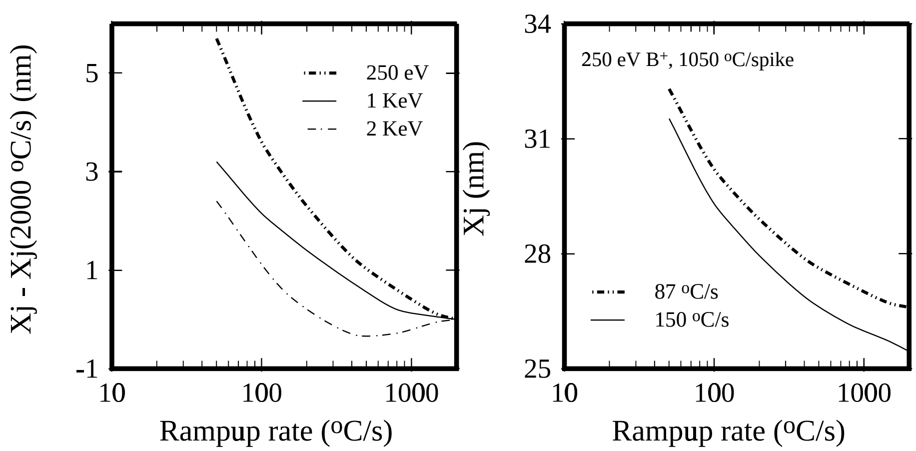

Fig. 6: The effect of ramp rate on reduction in junction depth for $1 \times 10^{15}\mathrm{cm}^{-2}$ B implants following 1050°C spike anneals. In (a), the difference in junction depth ($x_j@5{\times}10^{18}$ cm^{-3}) relative to a ramp-up rate of 2000°C/s is shown for various implant energies. In (b), the effect of changing the cooling rate is shown. Note that increasing the ramp-down rate gives a greater junction depth reduction for higher ramp-up rates, since more of the TED then occurs during ramp-down.

SUMMARY

In summary, we find it is possible to model the extent of diffusion during spike anneals with varying ramp rates by considering the full thermal cycle. These models allow the optimization of RTP annealing cycles considering the trade-offs between junction depth and sheet resistance. For example, with 1050°C spike anneals, the active dose (and thus sheet conductivity) varies approximately linearly with junction depth. However, faster ramp rates allow the use of higher spike temperatures, with associated higher activation and reduced sheet resistance for the same junction depth.

Work at Boston University and the University of Washington was supported by the Semiconductor Research Corporation. We would like to thank Eric Perozziello for details and discussion regarding their experimental results.

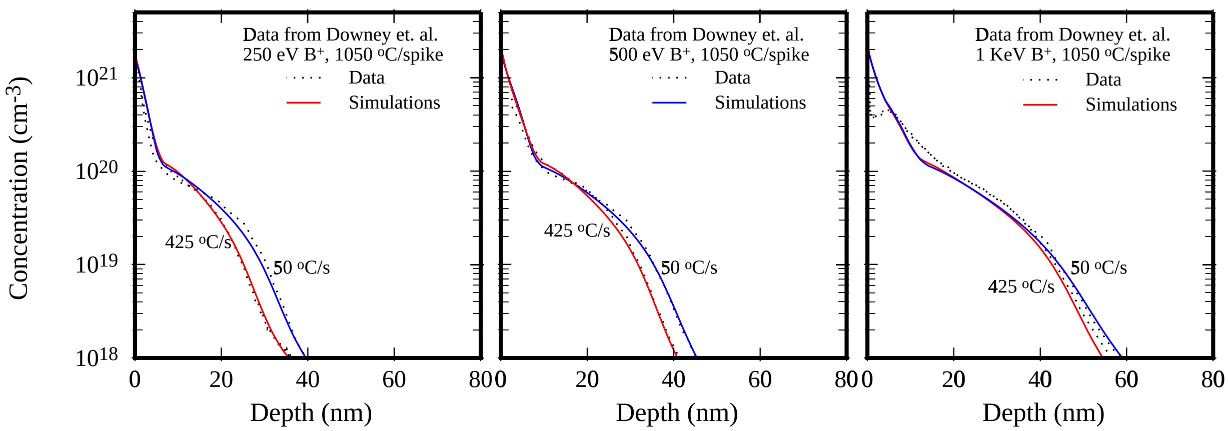

Fig. 7: Comparison of simulation and experiment for $1 \times 10^{15}\mathrm{cm}^{-2}$ (a) 250 eV B^+, (b) 500 eV B^+ (c) 1 keV and implants for a 1050°C spike anneal in 33 ppm O_2 in N_2 ambient. Ramp-up rates are 50°C/s and 425°C/s, and initial ramp-down rate is 87°C/s. Data from Downey *et al.*[1]

[1] D. F. Downey, S. F. Felch, and S. W. Falk. In Advances in Rapid Thermal Processing, Proc. of Elec. Chem. Soc. Meet. (1999).

[2] A. Agarwal, H.-J Gossmann and A. T. Fiory. In Si Front-End Processing.–Phys. and Tech. of Dopant-Defect Interactions, **568**, MRS Symp. Proc., April (1999).

[3] A.H. Gencer and S.T. Dunham. In Si Front-End Processing-Physics and Technology of Dopant-Defect Interactions, Materials Research Soc., (1999).

[4] D.J. Eaglesham, P.A. Stolk, H.J. Gossmann, and J.M. Poate. *Appl. Phys. Lett.* **65**(18), 2305 (1994).

[5] H.S. Chao. " Physics and modeling of ion implantation induced transient enhanced diffusion in silicon" Ph.D. thesis, Stanford University (1997).

[6] E. Perozziello. "Non-Equilibrium Dopant Diffusion in Silicon", Ph.D. thesis, Stanford University (1998).

[7] S. Chakravarthi and S. T. Dunham. In SISPAD Technical Digest, 1998.

[8] S. T. Dunham, S. Chakravarthi, and Alp H. Gencer. In IEDM, Technical Digest, 1998.