

MODELING OF COMPETITIVE GETTERING BETWEEN DEVICES AND GETTERING SITES

A.A.Istratov^(a), W.Huber^(b), and E.R.Weber^(c)

^(a) University of California, LBNL, Berkeley, CA, USA

^(b) Sumitomo Sitix Silicon, Fremont, CA, USA

^(c) Department of Materials Science and Engineering,
University of California, Berkeley, CA, USA

Computer algorithms for modeling of gettering, reported in the past, usually used simplified models where trapping of metals by the devices was neglected. It was assumed that the denuded zone / the epi-layer had much lower trap density than the substrate. The efficiency of gettering was determined by the distance which the impurities have to diffuse to get trapped at the gettering sites, by the properties of these gettering sites, and by the cooling rate.

A realistic predictive simulation of gettering during a CMOS process should take into account that devices can be efficient metal trapping sites. This is because doping levels in the device area may reach 10^{20} cm^{-3} , which is higher than the typical doping levels in the p^+ substrate (2×10^{18} to $2 \times 10^{19} \text{ cm}^{-3}$), if epi-wafers are used. Residual implantation damage and lattice strains in the device area may provide additional mechanisms of metal trapping.

In this paper, the concept of competitive gettering by the gettering sites and devices is introduced in order to take into account the trapping of metals by the devices. Modeling of competitive gettering in p/p^+ epi-wafers and wafers with internal gettering sites was performed using the gettering simulator developed in our group for SiWEDS (Silicon Wafers Engineering and Defect Science, a university – industry cooperative research center, co-sponsored by the National Science Foundation). Iron was used as a model metal impurity. It was assumed that 10^{11} cm^{-3} of iron was homogeneously dissolved in the wafer at 1000°C . Gettering was modeled during the cooling of the wafer from 1000°C to 200°C at a constant cooling rate which varied from 0.01 to 800 degrees per second. The device area was approximated by a 100 nm thick p^+ doped layer with a doping level of 10^{18} , 10^{19} , or 10^{20} cm^{-3} .

The significance of taking into account gettering by the devices is illustrated on the example of a wafer with devices but without gettering sites. Our modeling showed that in this case, the metal concentration in the device area can substantially (by 2-3 orders of magnitude) exceed its average concentration in the wafer even after a rapid quenching of the wafer. This is because a ten-fold increase of metal concentration in a 100 nm thick device layer can be achieved by collecting metals from only a 1 micron-thick layer of the substrate adjacent to the devices. This process takes less than a millisecond at 1000°C , if the diffusing metal is iron. Therefore, the gettering of metals by the devices may substantially degrade the device yield, unless gettering sites are introduced in the bulk of the wafer and their properties are optimized to make their efficiency comparable to, or better than that of the devices. In our presentation we show that although fast cooling rates inherent in rapid thermal processing represent a challenge for gettering, optimized gettering can perform well even if the wafer is cooled in a rapid thermal processing system at a rate between 5 and 100 K/s.

Analysis of p/p^+ gettering revealed that the doping

level of the substrate is the most important parameter in determining the efficiency of p/p^+ gettering. It was found that the residual iron concentration in the device layer is inversely proportional to the square of the substrate doping level. In other words, an increase in the substrate doping level by 10 times decreases the metal concentration in the device area by 100 times.

The epi-layer width was found to be important primarily when the average distance which iron can diffuse during cooling is limited. Therefore, this dependence is likely to be observed at epi-layer widths below 10 microns primarily during fast cooling, inherent in rapid thermal processing.

The internal gettering efficiency is more sensitive to the denuded zone width and cooling rate than p/p^+ gettering because the onset of the internal gettering occurs only when iron becomes supersaturated, i.e., when its equilibrium solubility drops below its actual dissolved concentration. For 10^{11} cm^{-3} of dissolved iron, this happens at approximately 700°C . For comparison, segregation p/p^+ gettering, which does not require an impurity supersaturation, is efficient in a wider temperature range, including the temperatures where iron diffusivity is not a limiting factor. Therefore, internal gettering requires either slower cooling rates, or shorter distances between the device layer and the gettering sites. The actual distance which interstitial iron has to diffuse from the devices to the sinks consists of two components, the denuded zone width and the distance between the individual oxide precipitates. Hence, to optimize internal gettering through minimization of the diffusion distance for metals, one has to reduce the denuded zone width and increase the oxide precipitate density.

The modeling results presented in this study reveal that both internal and p/p^+ gettering techniques are compatible with the new technologies which substitute furnace anneals with rapid thermal processing which utilize faster ramp-up and cool-down rates, and deal with lower metal contamination levels than ever before. However, optimization of gettering becomes paramount since it can no longer be assumed that metals will automatically be gettering by any type of gettering sites. A high density of the gettering sites and their proximity to the devices is crucial for the gettering efficiency when low metal concentrations need to be gettering at RTP cooling rates.

The conclusion that metal concentration at devices may substantially exceed their average bulk concentration implies that one may have to reconsider the definition of the critical metal contamination level detrimental for device performance. The metal concentration that is important is not the surface contamination level of the wafer, and not the average dissolved metal concentration, but the actual metal concentration in the device area. This concentration depends greatly on the type and doping level of the devices, availability, location, and properties of the gettering sites, as well as the cooling rate. The heavily doped areas of the devices, which accumulate the largest fraction of the metals trapped by the device layer, have nearly metallic properties and as a rule cannot be significantly affected by 10^{11} - 10^{13} cm^{-3} of metals. Metals become detrimental only when they agglomerate in weakly doped areas, or form precipitates which protrude through the boundaries between the areas with different doping levels.