CHIP SCALE PREDICTION OF NITRIDE EROSION IN HIGH SELECTIVITY STI CMP

Jihong Choi[1], Shantanu Tripathi[1], David A.Hansen[2] and David A.Dornfeld[1]

[1]Department of Mechanical Engineering,
University of California
1115 Etcheverry hall, Hearst ave.
Berkeley, CA 94720

[2]Cypress Semiconductor Corp.
195 Champion Court, San Jose, CA 95134

CMP has been successfully used for shallow trench isolation (STI) process in front end semiconductor processing. In STI process, nitride covers active device area acting as the CMP stop layer. Amount of nitride erosion during CMP is directly related to the step height of oxide isolation structure after removing nitride, and affects device performance as well. As semiconductor technology node is decreasing beyond 90nm, variation of nitride erosion over a chip in CMP process should be controlled within tight specifications. A model based simulation tool for chip scale variation of nitride erosion would improve not only CMP compatible STI layout design but also help in optimization of high density plasma chemical vapor deposition (HDPCVD) process of oxide layer.

In this paper, a chip scale modeling scheme for better prediction of nitride erosion map integrating HDPCVD oxide topography is presented. In addition, a simulation result for a specially designed test pattern resembling production level STI layout is presented.

Keyword: Chip scale, STI CMP, HDPCVD, Nitride Erosion

INTRODUCTION

Shallow trench isolation (STI) is now a standard lateral isolation technique that facilitates deep submicron scalability of devices by allowing very small active area pitches and thus a higher packing density. Chemical mechanical planarization (CMP) has enabled replacement of previously used device isolation method, local oxidation of silicon (LOCOS), by STI. The major requirements of the CMP step are: complete removal of silicon dioxide over silicon nitride layer (nitride layer acts as a CMP stop and protects the active device area); minimal nitride erosion to prevent damage to the underlying silicon material in active areas; and minimal dishing of the trench oxide **[1]**.

The profile of the deposited silicon dioxide film using high density plasma (HDP) deposition, typically used for STI, is strongly dependent on the layout pattern. Pyramidal shapes of the overburden HDP oxide, varying with underlying active area width, results in a topography pattern density different from the layout pattern density and changing along the thickness of the overburden dielectric. CMP of this geometry causes non-uniform removal of the overburden oxide, resulting in different clearing times. Low density areas which are cleared fast experience nitride erosion and trench oxide dishing, whereas certain high density areas may have remaining silicon oxide over silicon nitride (Figure 1).

Several approaches to reduce the above mentioned within die non-uniformity (WIDNU) have been pursued. Reverse moat etch process, where the majority of the overburden dielectric over silicon nitride is etched away, assists towards a uniform CMP. Dummy features introduced in low density areas reduce pattern density variation effects. High selective slurries, with selectivity possibly greater than 100:1 (ratio of removal rate of silicon dioxide to that of silicon nitride), reduce silicon nitride erosion.

High slurry selectivity, though, could increase the formation of recesses in trench oxide, affecting the device performance. Hence, newer advanced slurry formulations use surfactant chemistry that reduces the oxide removal rate as the surface morphology becomes flat with polish time.

In spite of use of advanced high selectivity slurries there is still certain amount of nitride erosion variation across the die. Figure 2 shows an example of variation of nitride thickness over three different dies with identical pattern at different locations on a wafer polished with advanced high selectivity slurry. Within die variation of nitride thickness was less than 10nm for each die. However, within this tight variation across a die, strong pattern dependency still exists as shown in the figure. This variation is strongly dependent on initial oxide thickness variation and evolution history of oxide topography, which results in different clearing times at different locations on a die.

The ability to predict the post-CMP oxide and nitride thickness for arbitrary chip layouts is critical for CMP and deposition process optimization, layout screening or density design rule checking, pattern density equalization (e.g., dummy fill), process control, and circuit impact analysis **[2]**. Models predicting WIDNU were first developed for ILD (inter layer dielectric) CMP. Chekina et al **[3]** studied pattern dependency of CMP evolution based on fundamental contact wear mechanics. Boning et al **[4]** developed an empirical semi-physical pattern dependent model of the CMP process, integrated with a parameter extraction and process characterization methodology, to predict post-CMP oxide thickness variation across patterned chips. A finite element model **[5]** showed correlation between local contact pressure distribution and removal rate distribution on specific test patterns, which was calculated with the semi-empirical model **[4]**. Based on these models, an analytical model **[6]** for the effect of pad surface topography on chip scale pattern evolution was suggested with an experimental comparison.

Usually, the important assumption in such models is that the local removal rate in CMP is proportional to the local pressure on a feature and thus inversely proportional to effective pattern density. The effective pattern density at a position can be calculated by estimating the effect of the geometry of nearby features (within a characteristic length scale) on pad bending and thus the contact phenomena at that position. Previous models approximate the initial topography regions as either high or low features, ignoring the complexity of the initial conditions that has profound impact, particularly in case of STI, on how the features evolve during CMP. In this paper, a modeling scheme for high selectivity STI process linked with initial HDPCVD oxide topography and height variation effect is introduced.

The overall modeling scheme is described in the following section. Thereafter, a model of the pre-CMP HDPCVD oxide profile and its variation is discussed. It is followed by description of a local removal rate model to simulate CMP evolution and prediction of the final oxide and nitride profile.

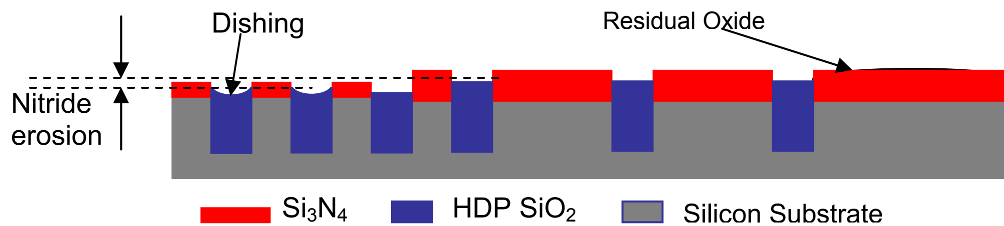
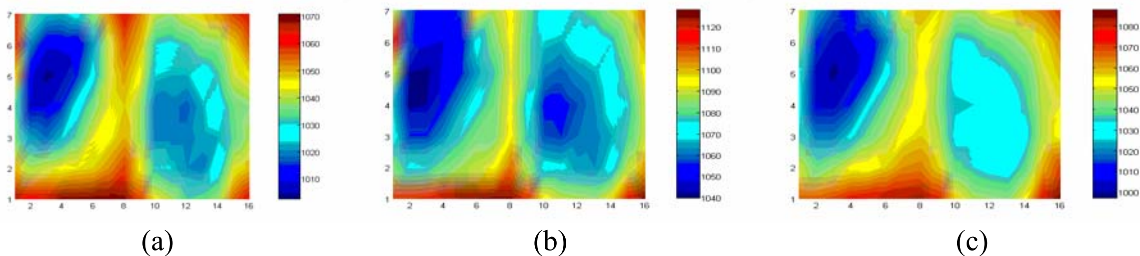


Figure 1. STI CMP problems.



(a) (b) (c)

Figure 2. Nitride thickness measured on three different dies at (a) center (b) middle (c) edge of a wafer

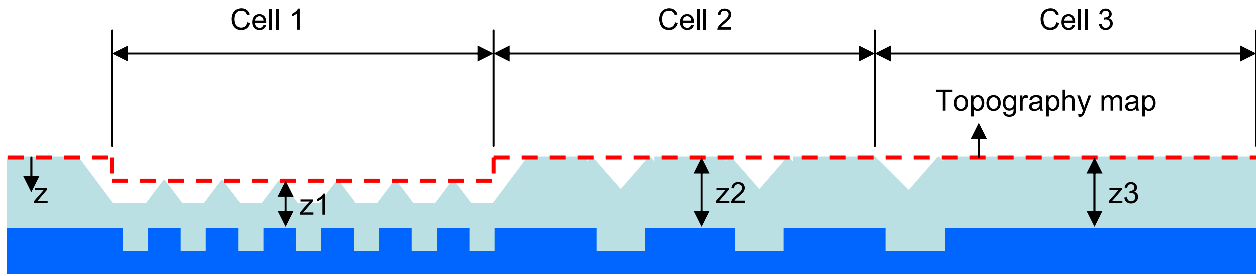


Figure 3. Discetization into basic modeling unit cells

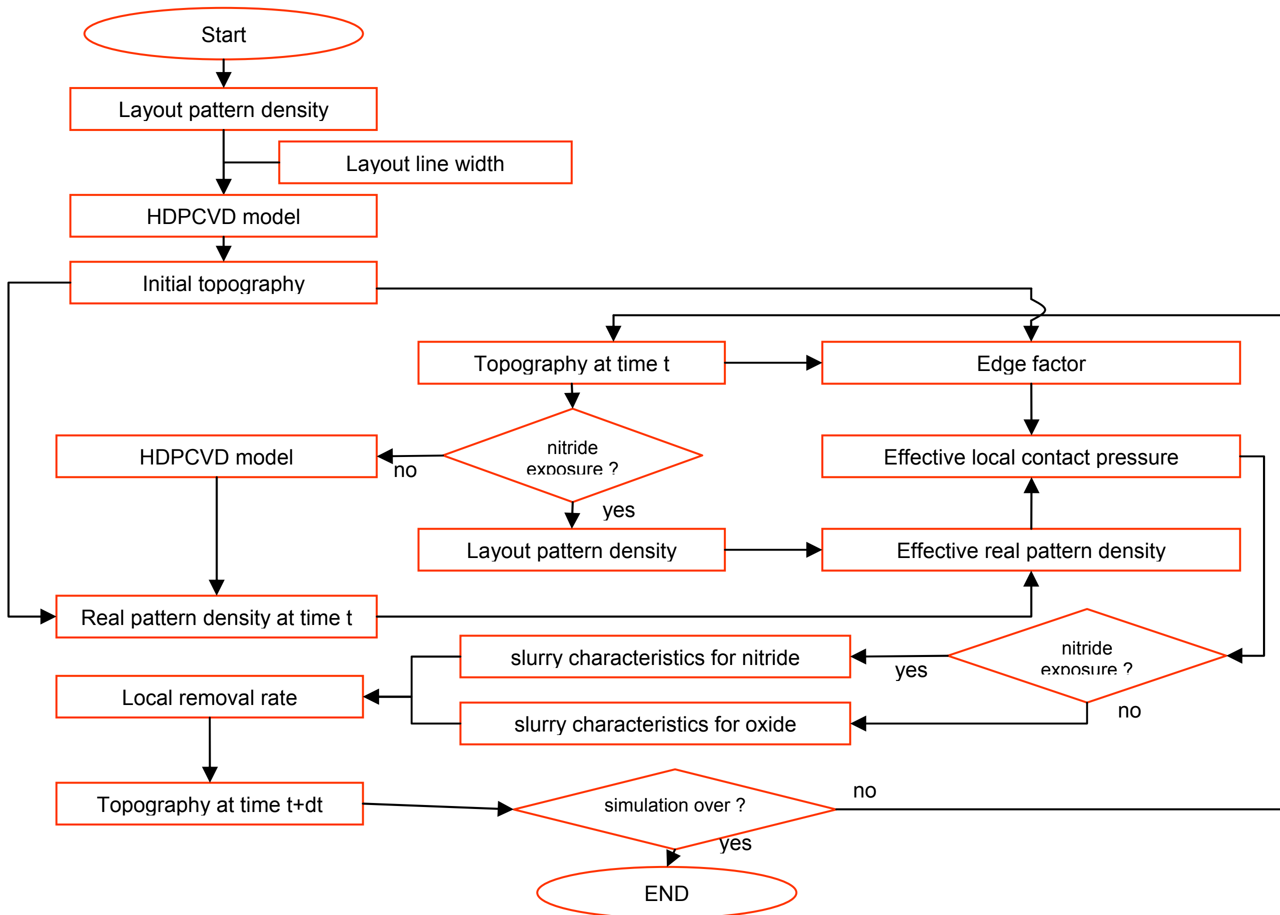


Figure 4. Simulation process flow.

MODELING SCHEME

For chip scale modeling, it is not practical to use feature scale data structure because the size of data and computation time for the modeling becomes too large, whereas the possible improvement of modeling accuracy is little. Hence, the pattern layout is discretized into basic modeling cells of size much larger than small features on the chip layout (figure 3). These cells form the basic units for the evaluation of oxide thickness, nitride thickness, pattern density, pitch and line width. The size of modeling cell should be determined based on pattern layout and evaluation of the trade off between computation cost and modeling accuracy.

Within each modeling cell, nominal pattern density is calculated. In addition, based on line width and line space information, overburden oxide thickness is evaluated based on HDPCVD model (described in the following section). For every time step, topography map is updated by the local removal rate map. Local removal rate is dependent on effective local contact pressure and slurry characteristics. For the evaluation of effective local contact pressure, effective pattern density and edge impact factor is evaluated from topography map at each time step. As the topography map is updated at each time step, real pattern density changes as a function of the shape of HDPCVD high points are calculated. In addition, the removal rate vs. local contact pressure relation changes to that of nitride after nitride exposure. Figure 4 shows simulation process flow.

HDPCVD PROFILE AND PATTERN DENSITY VARIATION OVER TIME

For the prediction of post CMP profile, it is essential to have post oxide deposition profile. The dependency of post deposition profile on underlying layout can be changed by controlling deposition-sputter ratio in HDPCVD process. It is known that the lateral distance to the trench corners (λ n figure 5) changes with deposition-sputter ratio **[1]**. Also, the slope of the HDP oxide sidewalls (α in figure 5) changes with process parameters of HDPCVD. Over wide active areas, overburden oxide has trapezoidal cross section instead of the triangular cross section over small active areas. Oxide thickness over active area larger than certain value (LS_{max} in figure 5) is constant. This value can be easily measured and used as a modeling parameter. For the area with smaller trenches, the height of over burden oxide can be calculated from the slope of triangular high points (α) and trench corner bias (λ), which is function of HDPCVD process parameters and underlying active area width. Figure 7 shows an example of density map, active area width and oxide height based on this model.

At the initial stage of polishing, sharp high points of the triangular oxide overburdens polish faster than the flat oxide over wide active area if the height variation of nearby feature is in similar condition. As polishing continues, local polish rate at the small trench areas decrease as the real contact area increases and real contact pressure decreases. Real pattern density, which is defined as the real contact area divided by the area of discretized cell, changes as polishing continues until it reaches the bottom of the triangular high points. Figure 5 shows the variation of real pattern density (RPD) as a function of oxide height evolution. As an example, real pattern density evolution is shown for three different types of cells. In a cell with narrow active area (cell 1 in figure 3), real pattern density initially increases with polishing and then changes to 1 when the pyramidal regions are polished down completely. In a cell with intermediate line space (cell 2 in figure 3), pattern density starts from certain value (a in figure 6) and increases to reach 1. In a cell with large active area, real pattern density does not change from 1. After nitride exposure, oxide dishing happens much faster than nitride erosion. Hence, the real pattern density can be approximated as total nitride area divided by the cell size. The RPD after nitride exposure is used to find local pressure on nitride layer and thus the erosion rate of nitride. Parameters *a, b, c* and *d* in figure 6 are functions of modeling parameters in Figure 5. LPD1 and LPD2 in figure 6 represent layout pattern density in cell 1 and cell 2, respectively.

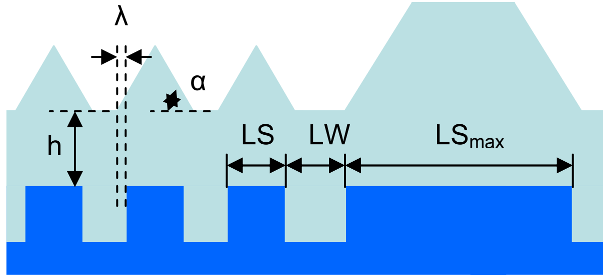


Figure 5. Modeling parameters for HDPCVD profile

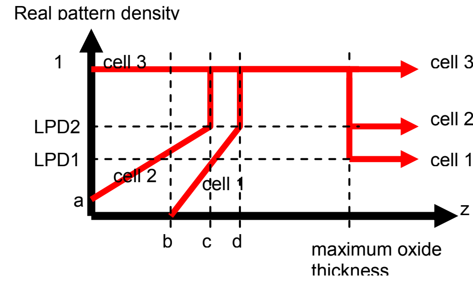
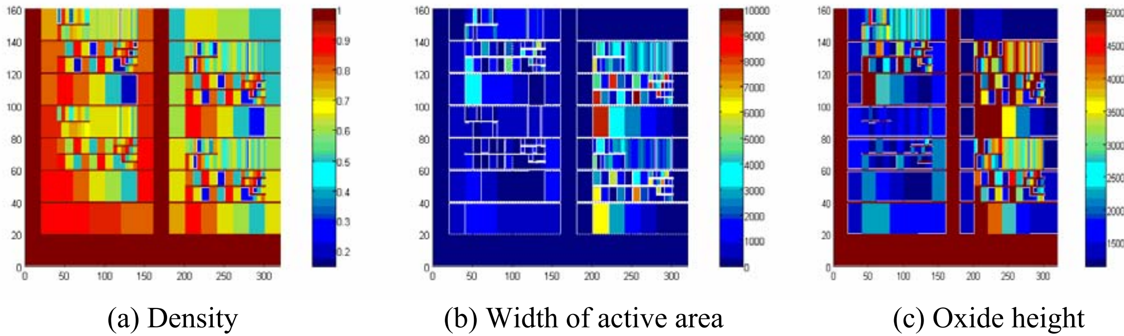


Figure 6. Real pattern density variation



(a) Density (b) Width of active area (c) Oxide height

Figure 7. Examples of nominal density map, width of active area and HDPCVD profile

LOCAL REMOVAL RATE

Pattern dependency of removal rate variation attributes to the variation of local contact pressure. Different local contact pressure at different locations can be modeled with contact mechanics **[3]**. However, this model assumes a flat smooth pad and perfect contact condition between pad surface and wafer topography. In reality, pad surface is much rougher than the wafer topography. Real contact between pad and wafer happens at pad asperity contact area, not over entire nominal contact area. Contact properties between pad and wafer change as a function of pad surface topography and pad material properties.

For a patterned wafer with uniform height and different pattern densities, local contact pressure changes as a function of pattern density. Many different sizes of real contact areas exist during CMP process. Figure 8 shows two pad asperities in contact with many wafer features. Effective value of the real contact area (R is the radius of real contact area in figure 8) is used in the model, which changes as a function of pad surface topography and pad stiffness. Within a real contact area, non uniform pressure distribution exists; higher pressure near the center and lower pressure near the edge of contact area. Pressure distribution (w(r)) can be modeled with elastic local contact model. Local contact pressure at a feature at the center of contact area changes as a function of total contact area between pad and wafer features within the real contact area and w(r). As the total contact area increases (pattern density increases), local contact pressure decreases and vice versa. In this way, effective local pressure is evaluated as a function of neighboring density information. This model can be applied to relatively flat and smooth initial topography.

However, in the case of HDPCVD, post deposition profile is not smooth. Hence, step height effect between different pattern density areas should be considered. This includes die edge effect and large height variations within patterned area. For a region between a high area and a low area, edge effect

is considered in the model. Lower pressure is applied near the edge of low area and higher pressure is applied near the edge of high area (figure 9). Physically, this model assumes that the local effective pressure is not only a function of local pattern density, but also height changes over certain distance. For example, sharp features with abrupt height change are polished faster than features with a smooth shape. To address this effect, edge impact factor was used in the model. For a certain location on a die, edge impact factor is defined as sum of height differences between neighboring features and the location for which the edge impact factor is evaluated. At the initial stage of polishing, sharp variations of height exist on the wafer, and those areas have high edge impact factor. Variation of edge impact factor over a die decreases during polishing as sharp height variation is smoothed as polishing continues. Figure 10 shows an example of variation of edge impact factor over time with different evaluation window sizes. The size of evaluation window changes with pad stiffness and pad surface topography. It will increase as the stiffness of the pad increases or the size of real contact area between pad asperity and wafer increases. Physical model to express the evaluation window size as a function of these parameters needs to be developed for the future.

Pressure dependency of local removal rate in high selectivity STI process can be characterized as shown in figure 11. Below a certain threshold pressure, the increase rate of material removal rate with respect to pressure is very small. Almost no material removal occurs for oxide or nitride films in this pressure range. For a certain range of pressure, removal rate increases almost linearly with pressure. Above a certain pressure range, the increase rate of removal rate drops. Hence the overall pressure dependency of the removal rate is not Prestonian This characteristic enables high step height reduction efficiency since the low area polishing happens only after the pressure at the low area reaches the threshold pressure. During simulation, pressure dependency of oxide removal is applied to the effective local removal rate map before nitride exposure. After nitride exposure, corresponding pressure dependency of the nitride removal rate (which has a much smaller slope) is used in the model. Figure 12 shows an example of topography evolution based on this model.

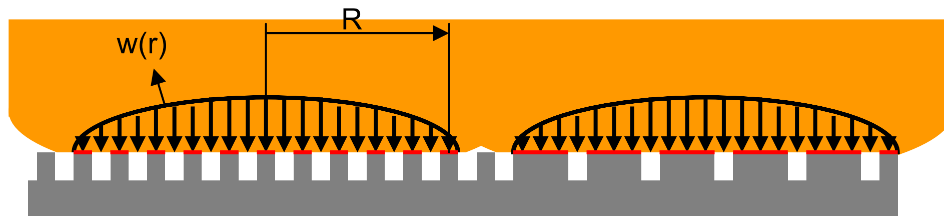


Figure 8. Contact pressure distribution in real contact area

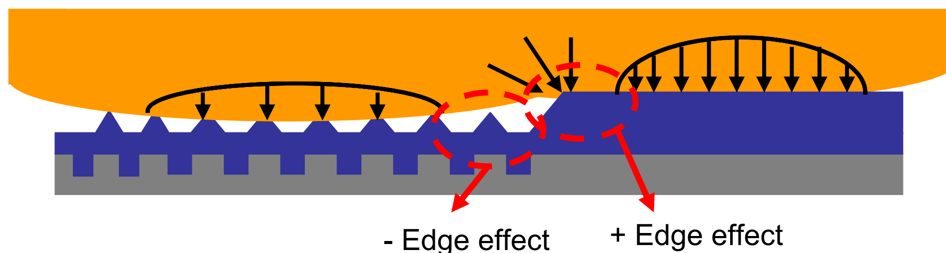


Figure 9. Edge effect

t=30 t=70 t=90

(a) Effective length scale = 50 µm

t=30 t=70 t=90

(b) Effective length scale = 1.1mm

t=30 t=70 t=90

(c) Effective length scale = 2.5 mm

Figure 10. Edge effect calculated with different effective length scales

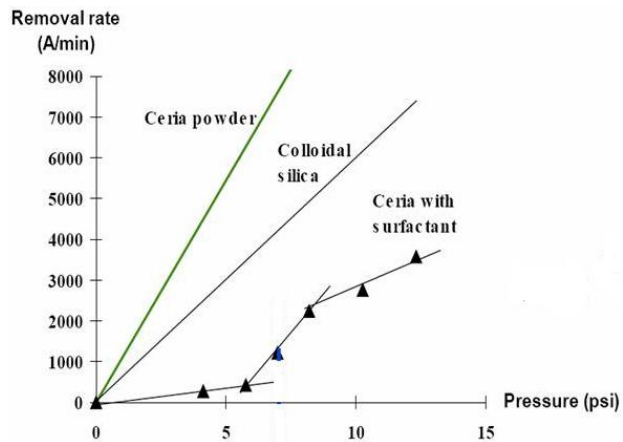


Figure 11. Removal rate vs. contact pressure of high selective slurry

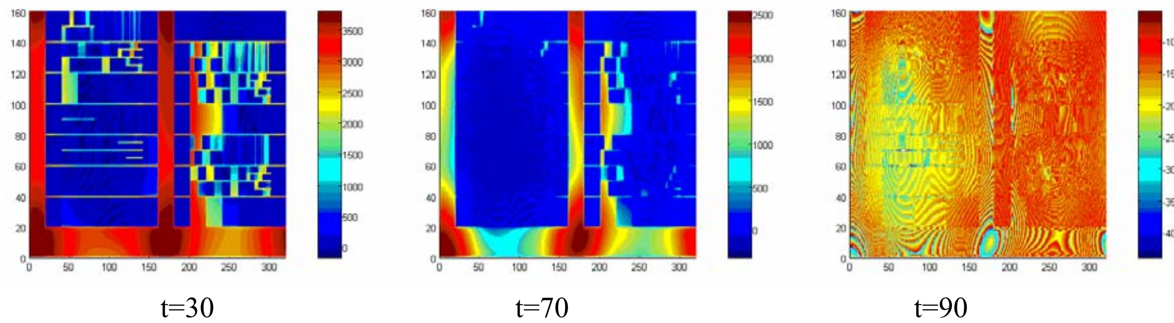


Figure 12. Topography evolution (effective length scale = 2.5mm)

CONCLUSTION

For better gap filling of high aspect ratio trenches, HDPCVD process is widely used for oxide deposition in STI process. HDPCVD process results in a strongly pattern dependent non-uniform oxide height distribution. High selectivity CMP process enables high planarization efficiency and low nitride erosion (within 10nm variation) over a chip. However, strong pattern dependency of nitride thickness variation is still observed. This pattern dependency attributes to the initial non uniformity of oxide profile. A modeling scheme to predict post-CMP oxide and nitride thickness for arbitrary chip layout was introduced with an example simulation result. This model integrates HDPCVD profile and CMP evolution model. In addition, edge impact factor was used to consider large step height effect. Experimental model calibration and verification is still underway.

REFERENCES

1. M.R. Oliver, Chemical-Mechanical Planarization of Semiconductor Materials, book, Springer-Verlag, 2004
2. X. Xie, T. Park, B. Lee, T. Tubgawa, H. Cai, and D. Boning, "Re-examining the Physical Basis of Pattern Density and Step Height CMP Models," MRS Spring Meeting, Symposium F: Chemical Mechanical Planarization, San Francisco, CA, April 2003.
3. O. G. Chekina, L. M. Keer, and H. Liang, "Wear-contact problems and modeling of chemical mechanical polishing," J. Electrochem. Soc., vol. 145, no. 6, pp. 2100–2106, June 1998. 2000.
4. D. O. Ouma, D. S. Boning, J. E. Chung, W. G. Easter, V. Saxena, S. Misra, and A. Crevasse, "Characterization and Modeling of Oxide Chemical Mechanical Polishing Using Planarization Length and Pattern Density Concepts," IEEE Trans. Semiconduct. Manufact., vol. 15, no. 2, pp. 232-244, May 2002.
5. J.Choi and D.A.Dornfeld, "Modeling of Pattern Density Dependent Pressure Non-uniformity at a Die Scale for ILD CMP", MRS Proc. Vol.816, MRS spring meeting, San Francisco, CA, April 12-16, 2004
6. J. Choi and D.A.Dornfeld, "Chip Scale Topography Evolution Model for CMP Process Optimization", Proceedings of International Symposium on Semiconductor Manufacturing (ISSM), San Jose,CA, Sept 13-15, 2005