Modeling of Oxide CMP at Chip Scale

H. Wolf [1], R. Streiter [1], R. Rzehak [2], F. Meyer [2], G. Springer [2], T. Gessner [1]
[1] FhG-IZM Chemnitz, Dept. MDE
[2] Infineon Technologies, Dresden

1 Introduction

Chemical Mechanical Planarization (CMP) of oxide layers is an important process step during the formation of the Shallow Trench Isolation (STI) of present and future CMOS ICs. As for other processes, optimization by simulation is desired also for CMP, but because of the complex mutual interaction of many different concerns, physically based and experimentally verified general solutions are not sufficiently available yet.

CMP modeling takes place at various length scales. At wafer scale, FEM and BEM codes are often applied to optimize the pressure distribution across the wafer. The within-wafer non-uniformity of removal rates is reduced for very homogeneous pressure distributions that can be achieved by equipment optimization (e.g. wafer carrier, retaining ring etc.). For further improvements, the hydrodynamics of the pad-wafer contact have to be taken into account, too.

One main objective of CMP modeling at chip scale is to find out how the time dependence of material removal is affected by layout properties such as pattern density and dimensions of lines and spaces. Having determined the values of model parameters by evaluation of test patterns, predictions for arbitrary layouts should be possible with regard to the homogeneity of oxide dishing and nitride erosion across the chip. Thus, empirical CMP simulation at chip scale is applied not only for the determination of optimum process conditions, but also for layout optimization including the insertion of dummy structures (see e.g. [1]).

The evolution of feature topography during polishing is evaluated by solving the equations of linear elasticity to describe the interaction between polishing pad and wafer. Additionally, the pattern dependence of some parameters of the chip-scale models can be determined at feature scale. Moreover, elastohydrodynamic considerations at the length scales of pad asperities and slurry abrasive particles are necessary to explain how removal rates depend on process variables such as pressure and velocity.

This paper describes the representation of very extensive experimental data (about 8000 data points) at chip scale using an density-step height model, necessary extensions of the model, and the dependence of some of its parameters on process conditions.

2 Experimental

Mask set: Test structures have been prepared at Infineon Technologies using a CMP characterization mask set developed at the Massachusetts Institute of Technology [2]. A schematic representation is shown in Fig. 1. Within the density array (lower part), the layout density varies from 4 % to 72 % at a fixed pitch of 250 µm. The pitch array (upper part) consists of arrangements of equidistant lines and spaces. The pitch varies from 20 µm to 1000 µm at nearly uniform pattern density of about 50 %. The minimum feature size in both arrays is 10 µm.

Preparation of test wafers: 4 nm pad oxide and 110 nm silicon nitride were deposited on 200 mm silicon wafers followed by the patterning of 390 nm deep spaces. Afterwards, an 860 nm thick oxide layer has been deposited.

Planarization: The test wafers were polished using a Westech tool and the process conditions as follows:

Pad:	IC1000 + SUBA IV (stacked)
Slurry:	Klebosol 30 N 50
Pressure:	3, 4.5, and 6 psi
Back pressure:	1 psi
Table speed:	35, 58, and 80 rpm
Carrier speed:	110, 95, and 80 rpm
Center distance:	13.8 cm

The heights of the up and down areas have been measured on the center chip at all positions assigned to the given density and pitch values for equally incremented polishing times until complete nitride removal.

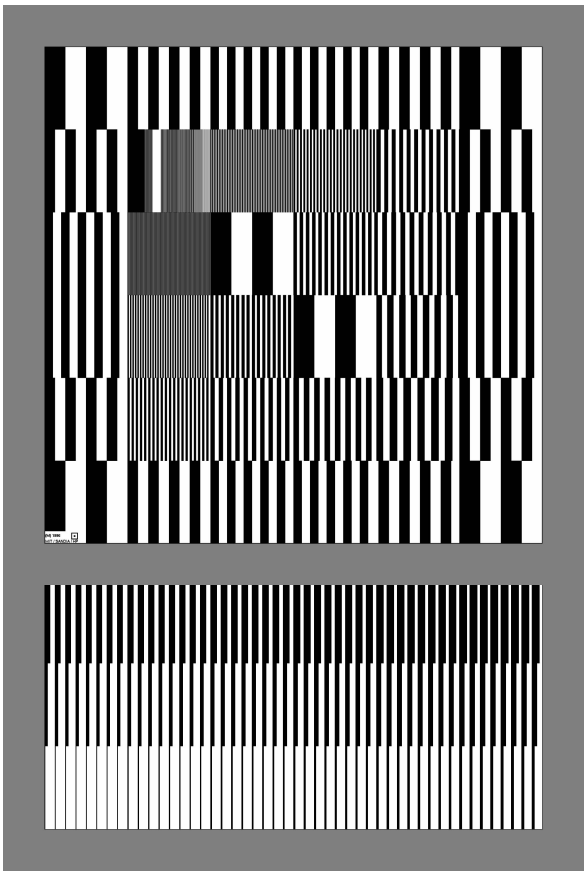

Fig. 1: Schematic representation of the test mask used. Up areas are white and down areas are black.

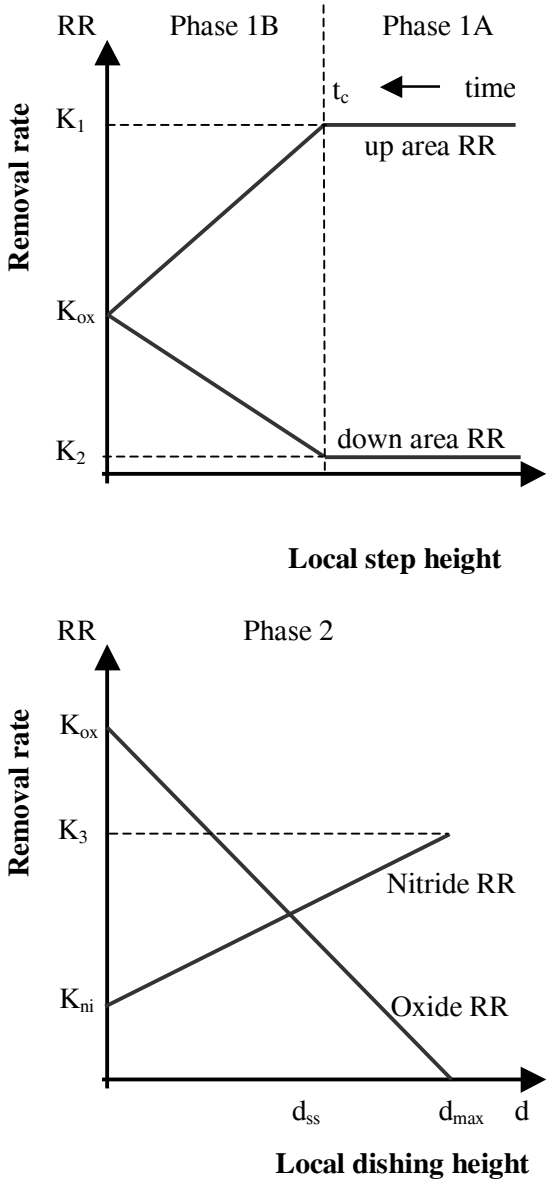

Fig. 2: Removal rate diagrams for oxide CMP [2]. Phase 1A indicates polish before the pad contacts the down area at the contact height h_c. Phase 1B indicates polish after down area has been initially contacted until nitride touch down followed by phase 2 indicating the simultaneous polish of oxide and nitride.

3 Modeling

3.1 MIT model

The density-step height model used for oxide CMP modeling at chip scale (hereafter referred to as the MIT model) is described in detail in [1] or in many other references listed at [2]. According to this model the oxide removal rates RR_u and RR_d of the up and down areas before nitride touch down depend inversely on effective density ρ and linearly on step height h as depicted in Fig 2 (phase 1A and phase1B). Similar dependences on density and dishing height d have been established for the removal rates RR_{ox} and RR_{ni} during the simultaneous polish of oxide and nitride. The rate equations are summarized in Eqs. (1) - (3).

K_{ox} and K_{ni} are the blanket removal rates of oxide and nitride, respectively. Within the frame of the standard MIT model, rate constants K_i and the contact height h_c are determined as

$$\text{Phase 1A:} \quad RR_u = K_1 \qquad RR_d = K_2 \qquad \text{for} \quad h \geq h_c \tag{1}$$

$$\text{Phase 1B:} \quad RR_u = K_1 + (K_1 - K_{ox})\frac{h}{h_c} \qquad RR_d = K_1 - (K_{ox} - K_2)\frac{h}{h_c} \qquad \text{for} \quad 0 \leq h \leq h_c \tag{2}$$

$$\text{Phase 2:} \quad RR_{ox} = K_{ox}\left(1 - \frac{d}{d_{max}}\right) \qquad RR_{ni} = K_{ni} + (K_3 - K_{ni})\frac{d}{d_{max}} \tag{3}$$

$$K_1 = \frac{K_{ox}}{\rho} \qquad K_2 = 0 \qquad K_3 = \frac{K_{ni}}{\rho_{ni}} \qquad h_c = \frac{a}{\rho} \qquad (4)$$

The effective density ρ at a spatial location on the die is defined as the weighted ratio of raised to total area within a weighting region. The raised area is determined from layout after adjusting for deposition with a bias which accounts for the increase/decrease in feature size due to deposition. A 2d Gaussian weighting filter characterized by a filter length FL was used to determine the effective density by weighting the influence of nearby cell densities for each cell. FL depends on consumables and process conditions and has to be determined together with K_{ox}, a, K_{ni}, and d_{max} from experimental data by least square fitting.

3.2 Model extensions

The representation of the experimental data can be improved further, if additional effects like edge rounding, non-vanishing initial down rates, and pressure dependent contact heights and dishing are taken into account. In doing so, the number of model parameters is increased. Moreover, the formulation of the model affects the value of the filter length. Thus, filter lengths from different models cannot be compared.

Contact height: Because h_c should depend on pressure P, a linear relation

$$h_c = \frac{a_1}{\rho} P \qquad (5)$$

has been assumed in this study as follows from Hookes Law.

Edge rounding: Freestanding narrow lines are removed faster than predicted by the standard MIT model. This is called edge rounding or corner rounding and may be due to an enhanced lateral erosion. The enhanced removal is observed only in the density array. In the pitch array the narrow lines are sufficiently screened and the enhancement is suppressed. Edge rounding is empirically modeled here as an enhancement factor to the rate K_1 and depends on line width L (up) and space S (down).

$$K_1 = \Psi \frac{K_{ox}}{\rho} \qquad \Psi(L, S) = 1 + [\Psi(L, \infty) - 1]\, f(S) \qquad f(S) = 1 - \exp\left(-\frac{S}{S_{c1}}\right) \qquad (6)$$

$$\Psi(L, \infty) = \begin{cases} L_c / L & \text{for } L \le L_c \\ 1 & L \ge L_c \end{cases}$$

Parameters L_c and S_{c1} have to be determined by evaluation of experimental data.

Non-vanishing down rate: It has been observed from experimental data that the down rate between wide-spaced lines is not zero in phase 1A as predicted by the standard MIT model. For narrow trenches the removal of the down rate is suppressed by screening. The same influence should act on the contact height, too, so that the same screening function g(S) is used for both. Moreover, the down rate is also decreased at higher densities what might be related to the increased lateral strain of the polishing pad. The decrease of the down rate results in higher pressure and an increased removal rate in the up area. For empirical modeling, exponential functions have been preferred to keep the number of parameters limited. Parameters b and S_{c2} have to be determined from experimental data.

$$K_2 = K_{ox} \exp\left(-\frac{\rho}{b}\right) g(S) \qquad h_c = \frac{a_1 P}{\rho} g(S)$$

$$g(S) = 1 - \exp\left(-\frac{S}{S_{c2}}\right) \qquad \leftarrow (7)\uparrow \quad (8)\downarrow$$

$$K_1 = \Psi\left[\frac{K_{ox}}{\rho} + K_{ox} \exp\left(-\frac{\rho}{b}\right) \frac{S}{L} (1 - g(S))\right]$$

Dishing: As for the contact height, the maximum dishing d_{max} is also expected to increase with pressure and for large nitride line space

$$d_{max} = d_1 P \qquad (9)$$

Parameter d_1 depends on selectivity, local pattern dimensions, and (visco-) elastic pad properties and can be calculated using an analytical dishing model (e.g. Eq. (14) of Ref. [3]). Unfortunately, our experimental data did not allow to resolve dependences other than on pressure.

4 Results and discussion

4.1 Blanket removal rates

For each combination of pressure and table speed the values of the model parameters have been determined by least square fitting of the corresponding up and down heights measured for every time step until complete nitride removal. For the blanket rates K_{ox} and K_{ni}, a linear dependences on pressure P and a sublinear dependence on relative velocity V could be identified as shown in Fig 3.

$$\frac{K_{ox}}{nm/s} = 0.7 \left(\frac{P}{psi}\right)\left(\frac{V}{m/s}\right)^{0.6}$$

$$K_{ni} = \frac{K_{ox}}{2.5} \qquad (10)$$

Selectivity remains constant at a value of about 2.5 for all process conditions investigated. No offset at P V = 0 has been observed.

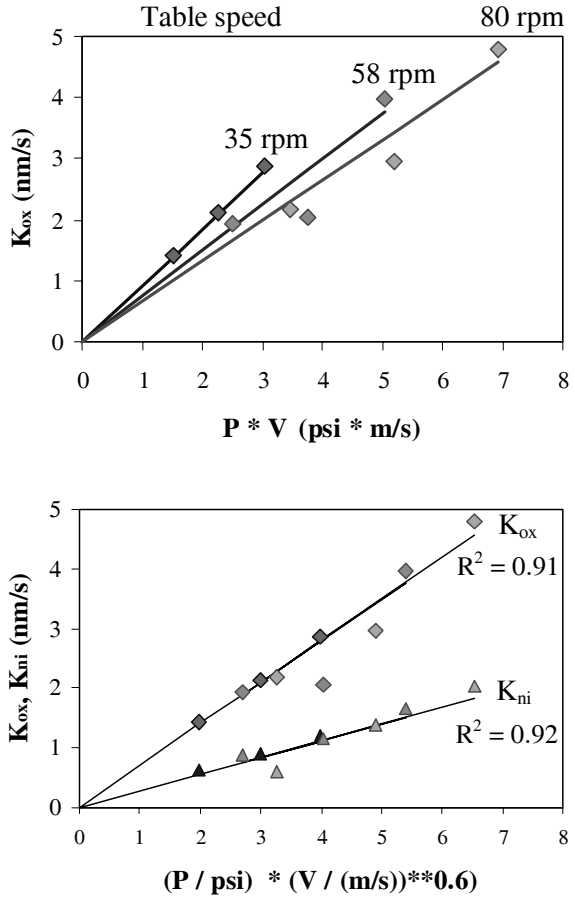

Fig. 3: Dependence of the blanket removal rates of oxide and nitride on pressure and relative velocity. The data points indicate the results of the individual fit for every set of process conditions. The lines show the simulation result based on Eq. (10).

The pressure exponent of unity is in accordance with Preston's findings for glass polishing [4], close to the value of 5/6 derived by Tseng and Wang [5], in agreement with the predictions of the comprehensive model developed by Qin et al. [6], but in contradiction to other models predicting values down to 0.5 [7]. Note that the derivation of the MIT model implies a linear dependence of blanket rates on pressure [2].

The value of the velocity exponent of 0.6 is between the limits of 1 for momentum transfer by direct solid contact between pad, abrasive particles, and wafer (boundary lubrication) and of 0.5 for hydrodynamic momentum transfer (hydrodynamic lubrication). A similar value of 0.65 was derived from experimental data by Hocheng et al. [8]. Eq. (10) can be transformed into a Prestonian representation

$$K_{bl} = K_p(V)\ P\ V \qquad K_p(V) \sim V^{-0.4} \qquad (11)$$

where K_{bl} stands for both blanket rates K_{ox} and K_{ni}. The Preston coefficient K_p is closely related to the coefficient of friction [9]. If the dependence of K_p on V/P is analyzed in terms of the Stribeck curve [10], the negative slope in Fig. 4 indicates that pad, abrasive particles, and wafer are in semi-direct contact characterized by elasto-hydrodynamic (mixed) lubrication. The decrease of K_p with increasing V/P is an expression of the growing influence of hydrodynamic effects in context with an increasing slurry film thickness [11].

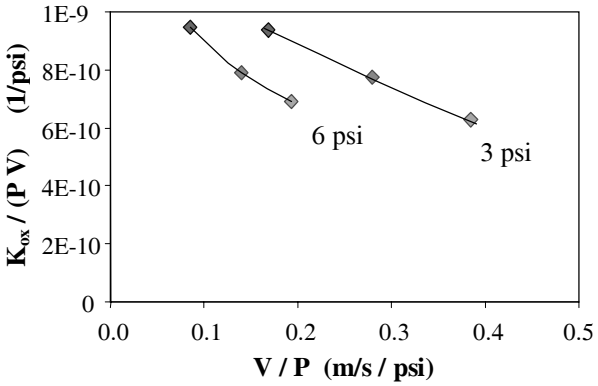

Fig. 4: Dependence of the Preston coefficient on V/P.

4.2 Filter length

The dependence of filter length FL, which characterizes the planarization behavior, on pressure and relative velocity is shown in Fig. 5. From the experimental data evaluated separately

for every set of process conditions, an upper limit of FL, a linear decrease with increasing pressure, and a sublinear increase with increasing velocity according to

$$\frac{FL}{\mu m} = 2250 - 115 \left(\frac{P}{psi}\right)\left(\frac{V}{m/s}\right)^{-0.4} \qquad (12)$$

were found. Combining Eqs. (10) and (12), a relation between FL, K_{ox}, and V is obtained

$$\frac{FL}{\mu m} = 2250 - 1.64E11\frac{K_{ox}}{V} \qquad (13)$$

which may be more general than the fitted results expressed by (10) and (12). The ratio of K_{ox} and V is equal to the material removed per sliding distance which decreases as velocity and slurry film thickness increase [11].

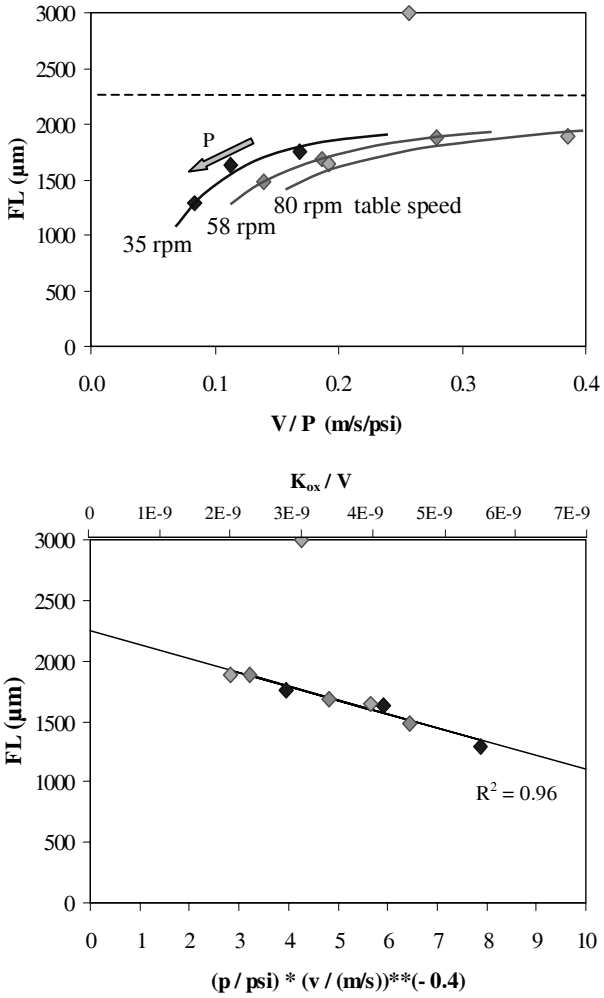

Fig. 5: *Dependence of the filter length on pressure and linear velocity. The data points indicate the results of the individual fit for every set of process conditions. The curves show the simulation result based on Eq. (12).*

The slurry film thickness h can be estimated from the case of full hydrodynamic lubrication, as

$$h \sim \sqrt{\frac{\mu V}{A P}} \qquad (14)$$

[8,12], where µ is the dynamic viscosity of the slurry and A is the area of the wafer. A similar dependence on V/P can be seen in the upper frame of Fig. 5. At low values of V/P there is direct solid contact all over the wafer surface resulting in nearly homogeneous removal and bad planarization behavior. If the average slurry thickness increases with increasing V/P, there is still direct solid contact with strong mechanical erosion at exposed surface positions, but already hydroplane sliding at positions with greater pad to wafer separation. Thus, planarization is improved. If the slurry film thickness is increased further, the number of positions with direct solid contact decreases and planarization cannot be improved further. Changing contact mode and limited stiffness of the pad cause FL to level off at increased values of V/P.

4.3 Parameter extraction

Because the parameter extraction for every individual set of process conditions resulted also in values having both less significance than K_{ox}, K_{ni}, and FL and no clear correlation to P and V, a common fit of the whole data set has been performed based on Eqs. (10) and (12) and constant values for the remaining parameters of the extended model. The results are collected in Table 1. Using this parameter set, the rms error is 20 nm.

Parameter	Unit	Value
K_{ox}	nm/s	$0.7\ (P/psi)\left(\frac{V}{m/s}\right)^{0.6}$
L_c	µm	83
S_{c1}	µm	1160
b		0.52
S_{c2}	µm	65
d_1	nm/psi	51
FL	µm	$2470 - 148\left(\frac{P}{psi}\right)\left(\frac{V}{m/s}\right)^{-0.4}$
a_1	nm/psi	14
K_{ni}	nm/s	$K_{ox}/2.5$

Table 1: Parameter set for the extended model.

4.4 Total indicated range

The total indicated range (TIR) is a quantitative measure for the planarization achieved by the polishing process. The TIR is defined as the maximum possible vertical topography difference within a given area (chip or part of a chip). Using the standard MIT model, the TIR can be calculated [13]. For sufficiently long polish times when all step heights have become small against the contact height, the TIR depends only on the initial step height h(0) and on the variation of the effective density $\Delta\rho = \rho_{max} - \rho_{min}$ according to

$$\mathrm{TIR}(t) = h(0)\,\Delta\rho \quad \text{for} \quad t >> h(0)\frac{\rho_{max}}{K_{ox}} \qquad (15)$$

For shorter polish times the TIR is above this long-time limit. But even for long polish times, Eq. (15) is only a crude approximation because local pattern effects are ignored.

If the extended model is used, the issue is more complicated. Both edge rounding and non-vanishing down rates promote the early erosion of down areas at low density and, therefore, tend to increase the TIR in the density array. Figs. 6 and 7 show that, although polish times are sufficient, the experimental TIR is more or less

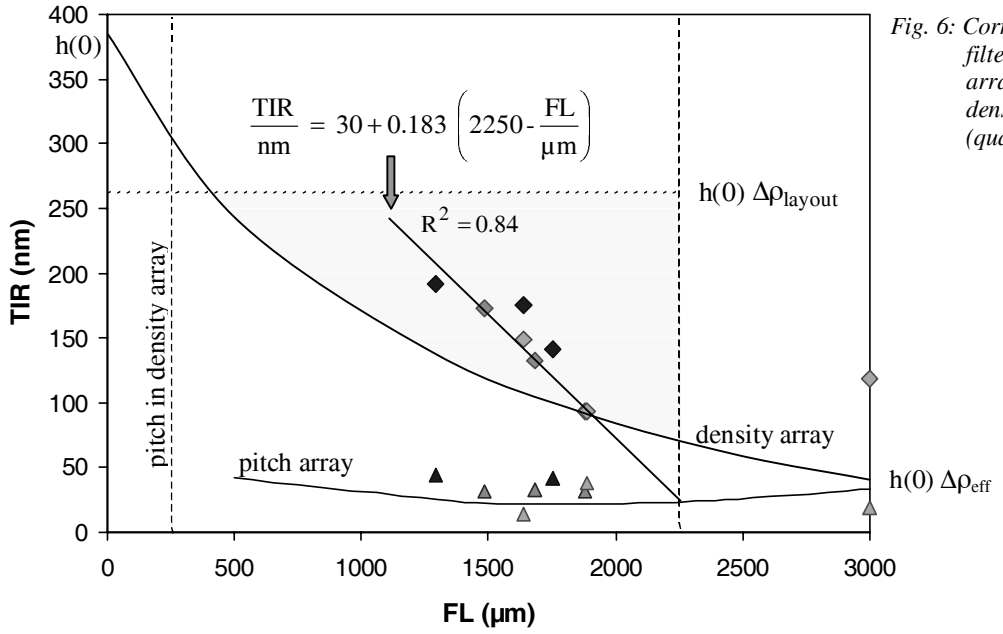

Fig. 6: Correlation of TIR and filter length for the pitch array (triangles) and the density array (quadrangles).

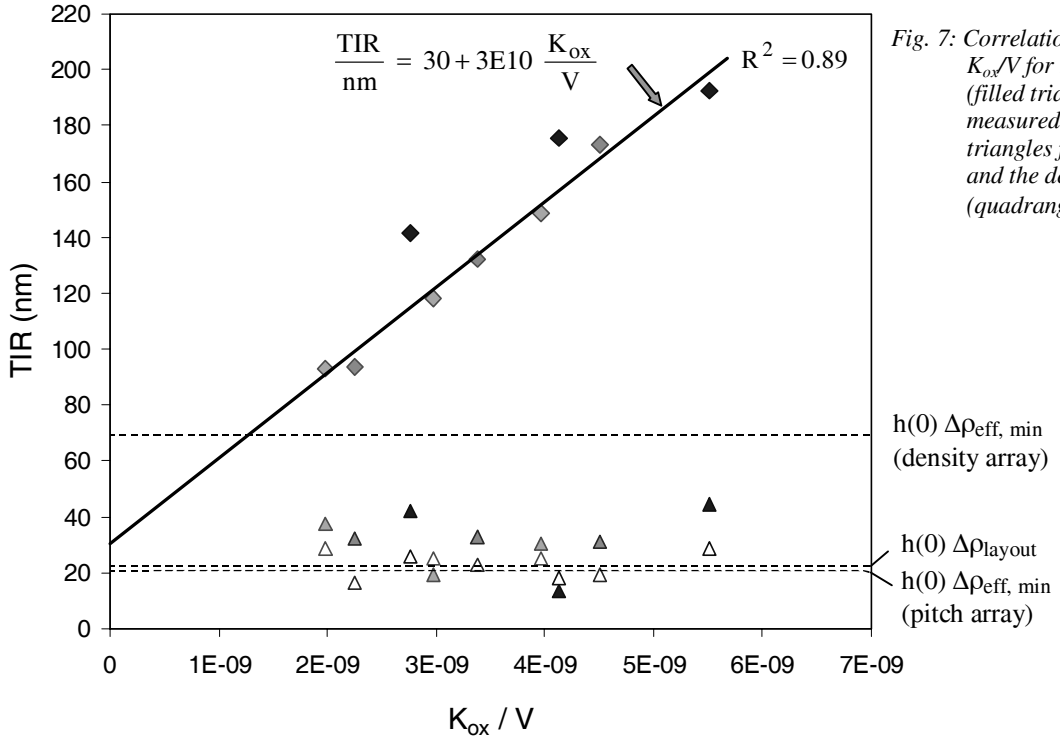

Fig. 7: Correlation of TIR and K_{ox}/V for the pitch array (filled triangles for measured TIR, open triangles for simulation) and the density array (quadrangles).

above the predictions of Eq.(15).

Because effective densities strongly depend on filter length, a correlation between TIR and FL is expected. Evaluation of measured heights *only for the density array* leads to

$$\frac{\mathrm{TIR}}{\mathrm{nm}} = 30 + 0.183 \left(2250 - \frac{\mathrm{FL}}{\mu\mathrm{m}} \right)$$

$$= 30 + 3\mathrm{E}10 \frac{K_{ox}}{V} \qquad (16)$$

where the second relation follows from (13).

The TIR of the density array for any individual process is found in Fig. 6 within a range determined by the maximum filter length, the initial step height, and the variations of layout density and effective density. For the process with 3 psi down pressure and 80 rpm table speed, the TIR is practically at its lower limit and cannot be reduced further with the given set of consumables. Also in the extended model, a greater value of FL corresponds to a lower TIR indicating improved planarization.

Fig. 7 shows the dependence of the TIR on the ratio of K_{ox} and V which is equal to the material removed per sliding distance. The TIR of the density array increases for more aggressive removal (e.g. higher P) and planarization gets worse. No such dependence has been found within the pitch array. The TIR varies here by some tens of nm. This is not only caused by statistical scattering of the measurements, because measured and simulated data follow the same trend.

5 Conclusions

- The MIT Model is suitable to simulate topography evolution at chip scale. Density and pitch dependent extensions such as edge rounding and non-zero down rates can reduce rms errors further.
- Oxide removal rates show a linear dependence on down force and a sub-linear dependence on relative velocity. The non-Prestonian behavior is due to changed pad-wafer contact caused by velocity dependent slurry film thickness.
- Filter length depends on details of the model and on process conditions. An upper limit and a linear dependence on material removal per sliding distance (K_{ox}/V) were found for the processes investigated.
- Within the density array the TIR depends linearly on K_{ox}/V. The TIR of the pitch array is nearly constant.

Acknowledgement

This work was financially supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (Project No 13N8406).

References

[1] B. Lee, "*Modeling of Chemical Mechanical Polishing for Shallow Trench Isolation*", Ph.D. Thesis, MIT Dept. of Electrical Engineering and Computer Science, 2002, http://www-mtl.mit.edu/Metrology/PAPERS/Lee-PHD2002-Thesis.pdf

[2] http://www-mtl.mit.edu/Metrology/

[3] Y. Guo, A. Chandra, A.-F. Bastawros, J. Electrochem. Soc. **151,9** (2004) G583-G589.

[4] F. W. Preston, J. Soc. Glass Technol. **11** (1927) 214-256.

[5] W.-T. Tseng, Y.-L. Wang, J. Electrochem. Soc. **144**,2 (1997) L15-L17.

[6] K. Quin, B Moudgil, C.-W. Park, Thin Solid Films 446 (2004) 277-286.

[7] J. Luo, D. A. Dornfeld, IEEE Trans. Semicond. Manuf. **14**,2 (2001) 112-133.

[8] H. Hocheng, H. Y. Tsai, Y. T. Su, J. Electrochem. Soc. **148**,10 (2001) G581-G586.

[9] Y. Moon, I. W. Park, and D. A. Dornfeld, Proc. Am. Soc. for Precision Engineering (ASPE), Vol. 17, pp. 83-87, ASPE Spring Topical Conference on Silicon, Monterey, April 1998.

[10] J. F. Luo, D. A. Dornfeld, Z. Mao, E. Hwang, *6th Int. Conf. on CMP for ULSI Multilevel Interconnection (CMP-MIC)*, March, 2001, Santa Clara CA, pp. 25-32.

[11] Y. Moon and D. A. Dornfeld, *Advanced Metallization Conference (AMC) in 1998*, Eds. G. Sandhu, H. Koerner, M. Murakami, Y. Yasuda, N. Kobayashi, pp. 255-260, Materials Research Society, 1999.

[12] S. Runnels, J. Electrochem. Soc. **141,7** (1994) 1900-1904.

[13] F. Meyer, J. W. Bartha, G. Springer, W. Dickenscheid, 11. CMP-Nutzertreffen, Itzehoe, Germany, 2003, http://www.isit.fhg.de/german/cmp/nutzertreffen.html