Proc. CMP-MIC, Santa Clara, CA, Feb. 1999.

A CMP MODEL COMBINING DENSITY AND TIME DEPENDENCIES

Taber H. Smith[1], Simon J. Fang[2], Duane S. Boning[1], Greg B. Shinn[2], and Jerry A. Stefani[2]

[1] MIT Microsystems Technology Laboratories, Cambridge MA 02139

[2] Silicon Technology Development, Texas Instruments, Inc., Dallas TX 75243

Abstract

This work reviews and compares density and time-dependent dielectric CMP models. Model fits to experimental CMP data demonstrate that a density model alone cannot explain the polishing of medium-low and low density features. It is shown that time-dependent models, which take into account contact with down areas, can explain this effect. These models are combined into a single model, whose parameters can be chosen to provide a 50% decrease in fitting error over that of the density model. It also shown that even the combined time-density model cannot explain the polishing of certain low density features. Finally, variations of this model are given which significantly reduce the number of parameters to be found, and show potential for use in predicting post-polish thicknesses for arbitrary layouts.

I. Introduction

Several works have proposed models for the chemical-mechanical polishing (CMP) of inter-level dielectrics (ILDs); each of which provide various benefits. The MIT density model has the ability to predict post-CMP ILD thicknesses for arbitrary layouts [1]. This is critically important if we are to utilize a model for performing pattern-dependent run-to-run process control [2], determining dummy fill [3], or estimating circuit performance [4]. However, this model provides thickness predictions only to a first order, and falls short when predicting low density features. Burke proposed in [5] that the step height decreases exponentially with time. Tseng *et al* proposed that removal rates of raised and down areas converge exponentially to the removal rate of an unpatterned dielectric sheet film (blanket removal rate) as polish time increases [6]. However, both these models lack a clear connection to density. In addition, the model in [6] assumes the pad is always in contact with both the raised and down areas, and suggests that this removal rate profile is determined by the distribution of pressure between the raised and down areas. Grillaert *et al* provided experimental data in [7] which demonstrated that these claims are true only after a certain step height is reached. The IMEC model suggests that before this “transition” step height is reached, the removal rate of the raised areas is characterized by the blanket rate divided by the density [7]. After the transition step height is reached, the removal rate profile is the exponential model outlined in [6]. The IMEC model also suggests that the transition step height is dependent on feature density, but it is unclear how these transition step heights can be determined *a priori* for arbitrary layouts. Therefore, it is not clear that this technique would work well on typical patterned wafers, where features are intermixed and their densities are not easily calculated.

In this work, we expand the MIT density model to include the IMEC model which, like the density model, works for arbitrary layouts, but improves fitting of low density features. We will not focus on the details of the algorithm here; the details will be provided in future work. Instead, we will focus on comparisons of variations of this model with the density model, and the effects these comparisons have on our understanding of the mechanisms in dielectric CMP polishing.

Section II briefly reviews the density model and time-dependent models. An analysis of the density model fit to experimental data is given in Section III. Section IV outlines a combined density and time-dependent model, and makes comparisons of this model to the density model. Section V presents two variations of this time-density model which simplify the model form and solution. Finally, Section VI presents conclusions and discusses future work.

II. Review of the Density Model and Time-Dependent Models

The MIT density model provides a first-order approximation of post-polish dielectric thicknesses for arbitrary layouts [1]. As shown in Fig. 1, this model assumes the polishing rate of a raised area is equal to the blanket rate divided by an effective density. The effective density is determined by computing a weighted average of the feature densities within a window. During this regime, the model assumes there is no removal in the down areas. Once the step height is assumed to be completely removed, the model assumes the removal rates of both the raised and down areas equal the blanket rate. This model's key strength is its ability to efficiently predict the thickness of an arbitrary layout to a first order. This benefit comes from the weighted average of the densities within a window, or interaction distance. This averaging is necessary because the pressure distribution of force on a particular feature is affected by neighboring features.

The model proposed by IMEC shows that the removal rate of the raised areas, and thus the step height reduction, is not linear [7]. As shown in Fig. 2, they suggest there is an initial linear regime, where the raised area removal rate is equal to the blanket rate divided by the feature density. During this period the removal rate of the down area is zero. After the pad contacts the down area, this first regime is followed by a period where the removal rate of the raised area exponentially decreases to the blanket rate. During this regime, the removal rate of the down area exponentially increases from zero to the blanket rate. Typical plots, as well as the expressions for the removal rates in the raised and down areas as a function of polish time, are shown in Fig. 2. Here K is the blanket removal rate, h_0 is the initial step height, ρ is the feature density, τ is the exponential time constant, t_p is the polish time, t_c is the time of contact with the down area, and $h_1 = h_0 - \frac{K}{\rho}t_c$ is the transition or contact step height. The work in [7] proposes that the step height at which the pad contacts the down areas is a function of the feature density; i.e. the higher the density the smaller the contact step height.

Before we continue, consider the differences between these models, highlighted in Fig. 3. Here the density model predictions are placed over the IMEC model predictions. We see that the time at which the density model switches to the blanket removal rate is later than the time at which the IMEC model transitions to the exponential removal rate. The IMEC model suggests that removal of the down area begins before the time suggested by the density model. It suggests that the pressure distribution changes before the step height is completely removed, and the load of the force is shared with the down area. This creates a large difference in the removal rate predictions just after the exponential regime begins. We have plotted the percent differences in the amount removed determined from each model in Fig. 4. Here we see that the predictions are fairly similar at the beginning and end, but there are large differences in the middle.

III. Analysis of the MIT Density Model

We now consider experimental data in order to show that the MIT density model needs to incorporate time dependent removal rates. Wafers were patterned with an MIT CMP test mask, deposited with a 16800 Å oxide layer, and polished using an IC1000 pad with a standard process on a rotary polish tool at Texas Instruments, Inc. The wafers had 20mm by 20mm die, patterned out to the edge. Each die contained five rows and five columns of 4mm blocks with lines of varying pitch and density. The post-polish dielectric thickness data, as well as the density model predictions are shown in Figs. 5 and 6 for the raised and down areas, respectively. In Fig. 5, the low density regions correspond to the lower thickness values, i.e. points A, B, and D are low density and points B and C are medium-low. We see that the predictions of the removal in the down areas are fairly accurate for the high and medium density regions, yet fairly poor for the medium-low density regions. The predictions are accurate in the high density regions because the removal rates of these features are slow, and the step height is still too large for the pad to touch the down area. Therefore, the high and medium density features are at the beginning of the profiles shown in Figs. 3 and 4. Note that the removal in the low and medium-low density regions is over estimated. We can see in Fig. 6 that at these same locations, the down area removal in these regions is under-predicted. This is the same profile suggested by the difference of the density model and the IMEC model in Fig. 4. Therefore, the failure of the model to accurately predict the removal in locations B, C, E, and F is most likely because the pad has touched the down area before the time predicted by the density model. This causes an increase in the removal of the down area and a decrease in the removal of the raised area. On the other hand, the removal in locations A and D is over estimated in both the raised and down areas. The inaccuracy of the density model in these locations is not explained by the early contact of the IMEC model. We will return to this issue later.

IV. A Combined Density and Time-Dependent Model

In this work, we incorporate the time-dependent model into the density model to capture the benefit of modeling arbitrary layouts, while improving performance with time dependencies. We begin by integrating the expressions in Fig. 2 to obtain the amount removed in the raised areas

$$AR_u = \begin{cases} t_p K/\rho & t_p \le t_c \\ t_c K/\rho + K(t_p - t_c) + (1-\rho)\dfrac{h_1}{\tau}(1 - e^{-(t_p - t_c)/\tau}) & t_p > t_c \end{cases}$$

and the amount removed in the down areas

$$AR_d = \begin{cases} 0 & t_p \le t_c \\ K(t_p - t_c) - \rho\dfrac{h_1}{\tau}(1 - e^{-(t_p - t_c)/\tau}) & t_p > t_c \end{cases}$$

We then assume that the feature density, ρ, can be replaced by an effective density, as in the density model. We are then left with the challenge of using the effective densities and these equations to explain the experimental data (pre- and post-polish measurements for raised and down areas) from an arbitrary layout. We will outline three methods for doing this. In each of these methods, we need to find K, τ, the t_c for each measurement site, and the effective density of each measurement site. As in the density model, we assume the effective density is determined by calculating the average density within a window, and that the window size is determined by a single parameter known as the planarization length [1].

The first method for determining these parameters picks a planarization length, and calculates the effective density for each measurement site. Using the measurements and effective densities, we perform a multivariate constrained optimization to find K and τ, as well as a contact time t_c for each measurement site. This process is repeated until the parameters which provide the best fit (i.e. minimum mean squared error) of the model to the experimental data are found. We will not go into the optimization details in this paper. However, it can be shown that the following constraints are necessary to maintain positive removal rates in both the raised and down areas.

$$\tau \geq 0, t_c \geq 0, t_c \geq \frac{\rho h_0}{K}, t_c < \frac{\rho h_0}{K} - \tau, \textit{and} \quad K(t - t_c) < \rho \frac{h_1}{\tau}(1 - e^{-(t - t_c)/\tau}) \;\; \textit{for} \; (\forall t | (t_c < t < t_p))$$

Using the experimental CMP data that we used for the density model above, this time-density model was fit to the data. The results are shown in Figs. 7 and 8. The raised area fit of the time-density model is a 50% improvement over the original density model. This improvement is largely in the low density regions, A through F. The down area fit of the time-density model is also 50% better than the original density model. Here we see a significant improvement in the low density region B, in the medium-low density regions C and E, and in the medium density region F. The early removal of the down area material over that of the density model significantly improves the predictions in these regions. Unfortunately, predictions in the low density regions A and D still have significant error. As we stated in our analysis of the density model, we did not expect the time-density model to correct these locations. It is possible that these errors are caused by poor measurement data. However, it is more likely that these effects are real. Figs. 7 and 8 indicate that the time-density model is unable to predict enough removal in the raised areas of the low density regions without over-estimating the removal in the down areas. It is possible that the microstructure of the pad asperities is having an additional effect not captured by the simple asperity "pad contact" model. The discrete nature of the asperities may result in different time-averaged pressure distributions between the up and down areas in these low density regions in a way that deviates from the IMEC model assumption, causing it to fail to accurately predict the removal of the up or down areas caused by the asperities of the CMP pad. We will consider such effects in later work.

V. Variations in the Time-Density Model

The previous method has a few problems. First, the large number of parameters (three plus the number of measurement sites) causes the determination of the model to be computationally intensive. Second, having a variable contact time for every site may cause over-fitting of the data. Thus we may be able to fit the data, but not be able to predict the thicknesses on other data sets. Third, these variable contact times make it difficult to predict post-polish thicknesses for arbitrary layouts. Finally, the optimal contact times result in contact step heights that have a functional dependence on density, which conflicts with the findings of [7]. Fig. 9 shows the fitted contact times determined from the optimization of the time-density model, plotted against effective density. The contact times above 40% density are plotted at the time of polish, meaning these features have not yet contacted the down area. These fitted contact times lead to the contact step heights shown in Fig. 10. Again, the contact heights beyond 40% are determined by the maximum value of the polish time. Results in [7] indicate that the contact step height increases monotonically with decreasing density. However, the contact heights below 40% do not agree with these results.

In response to these problems, we have tested two variations of this method. These utilize contact step heights which have a functional dependence on either the effective density or the feature line space. The first variation utilizes the functional dependence $h_1 = k_1 + k_2 e^{-\rho/\delta}$, where h_1 is the contact step height, and k_1, k_2, and δ are variable constants. This relationship restricts the contact step height to exponentially decrease with increasing density. This reduces the number of parameters in the previous method from 3+N to 6. In addition, the relationship of contact step height on density could be re-used for model prediction on other arbitrary devices.

The results of the model fit for this variation are shown in Figs. 11 and 12. Here we can see that this model also works quite well. There is a slight decrease in the quality of fit in the raised areas. This suggests that there is indeed a strong correlation of the contact height to density. The optimal contact step height dependence on density using this functional form is shown in Fig. 13. The functional dependence of the contact step height in this case is very different from that determined with the variable contact times above. However, this dependence on density agrees with that suggested in [7]. This suggests that the fit from the previous method was most likely over-fitting. Fig. 14 shows the model fit errors for both the raised and down areas. We can see from this figure that there appears to be larger errors around the 50% density region. The last 15 data points in Figs. 11 and 12 are all 50% density lines with pitch varying from 25 to 250 μm. These errors indicate that there may be a functional dependence of the contact step height on pitch or line space. Therefore, the second variation utilizes the functional dependence $h_1 = k_1 + k_2 l + k_3 l^2 + k_4 l^4$, where l is the feature line space, and k_1, k_2, k_3, and k_4 are variable constants. This reduces the number of parameters in the original method from 3+N to 7. Again, this relationship of contact step height on line space could be used for model prediction on arbitrary devices. The results of the model fit for this variation are shown in Figs. 15 and 16. Here we see that the line space dependence does improve the fit of the raised areas. Unfortunately, the dependence on line space shown in Fig. 17 suggests that the contact step height has a minimum which is not at a line space of zero. Intuitively, we would expect that the pad would be more able to reach down into larger line spaces, giving larger contact heights as line space increases. Therefore, the variation in contact height with line space shown in Fig. 17 should be treated cautiously, and may be due to confounding with density or other effects. Other test masks may be necessary to separate these effects.

VI. Conclusions and Future Work

We have demonstrated that a density model for dielectric CMP is not sufficient to completely characterize the removal in medium to low density features. We have shown that differences in the density model and the IMEC time-dependent model suggest a combined model would improve fitting. We have shown that a combined time-density model provides up to a 50% improvement in fitting errors of both raised and down area thicknesses. We have provided variations of this model which significantly reduce the number of model parameters and provide an opportunity for us to predict post-polish thicknesses for arbitrary layouts.

Future work will focus on understanding the contact height dependencies on density, line space, and pitch. In addition, we will work to demonstrate that this model accurately predicts post-polish thicknesses over time, as well as over different devices.

References

[1] B. Stine *et al.*, *CMP-MIC*, p. 266, Santa Clara, 1997.
[2] T. Smith *et al.*, "Device Independent Advanced Process Control of CMP," *in preparation.*
[3] B. Stine *et al.*, *IEEE Trans. on Electron Devices*, Vol. 45, No. 3, pp. 665-679, March 1998.
[4] V. Mehrotra *et al.*, *International Electron Devices Meeting*, San Francisco. CA, Dec. 1998.
[5] P. Burke, *VLSI Multilevel Interconnect Conf.*, pp. 379-384, Santa Clara, CA June 1991.
[6] E. Tseng *et al.*, *CMP-MIC*, p. 258, Santa Clara, 1997.
[7] J. Grillaert *et al.*, *CMP-MIC*, p. 79, 1998.

Acknowledgments

The authors would like to thank Joost Grillaert, Dr. Marc Meuris, and IMEC for valuable discussions regarding the IMEC model.

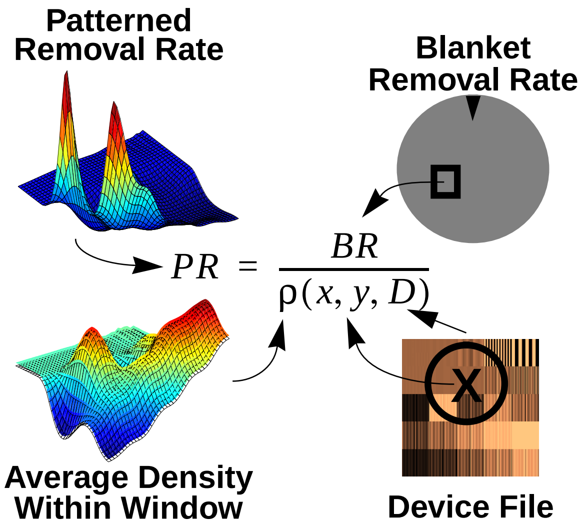

Figure 1. The MIT density model.

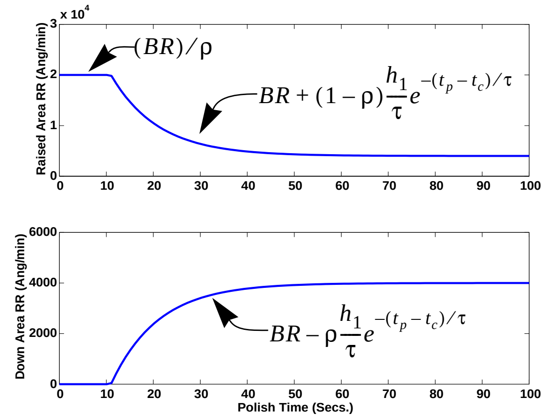

Figure 2. The removal rates of the raised and down areas using the IMEC time-dependent model.

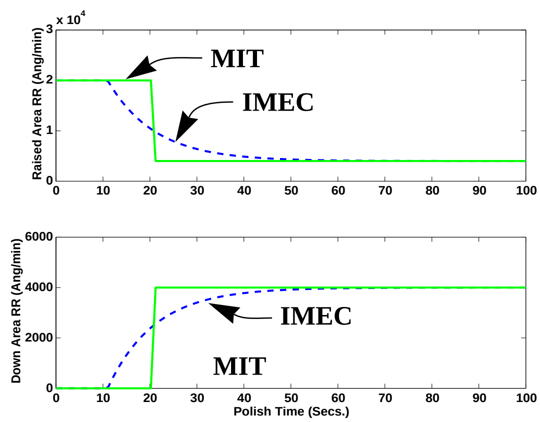

Figure 3. Removal rates of the density and time-dependent models.

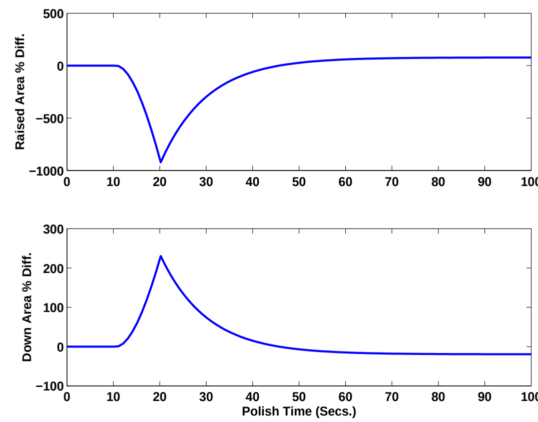

Figure 4. Percent difference in removal predictions between the density and IMEC models.

MIT Density Model Up and Down Area Predictions

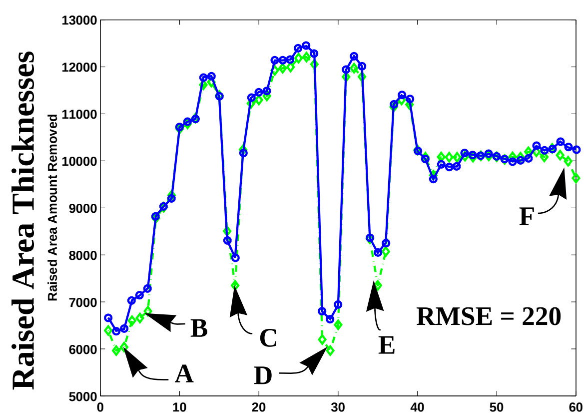

Figure 5. Actual and predicted post-polish thickness of the raised areas using the density model.

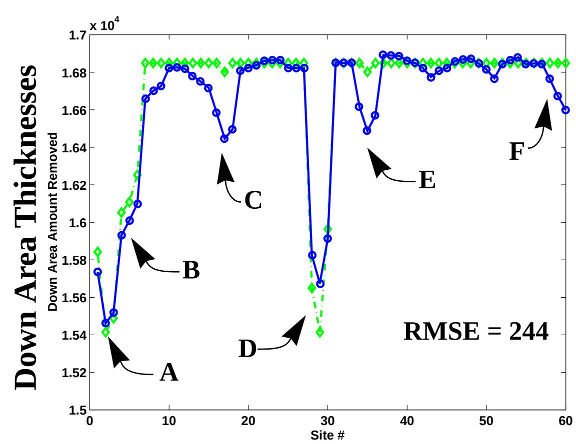

Figure 6. Actual and predicted post-polish thickness of the down areas using the density model.

Time-Density Model Up and Down Area Predictions (Floating Contact Times)

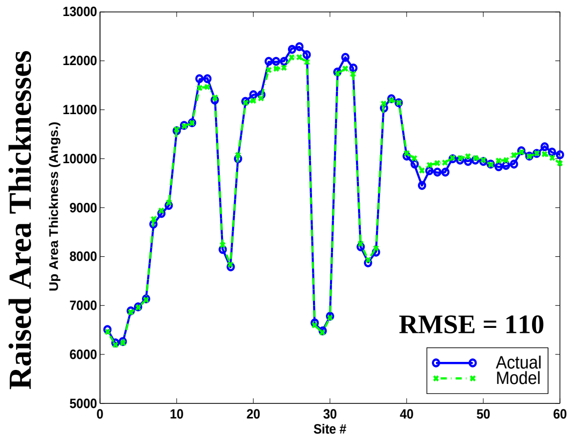

Figure 7. Actual and predicted post-polish thickness of raised areas using the time-density model.

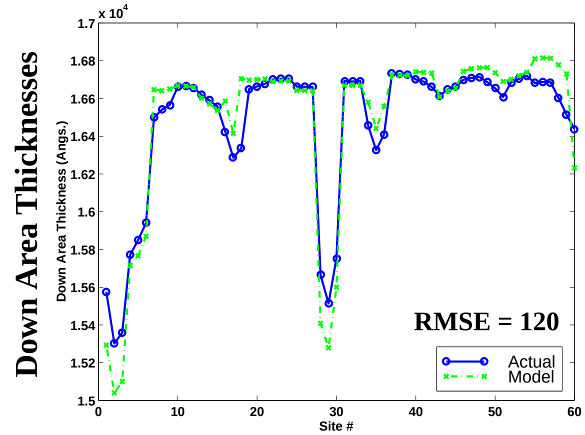

Figure 8. Actual and predicted post-polish thickness of down areas using the time-density model.

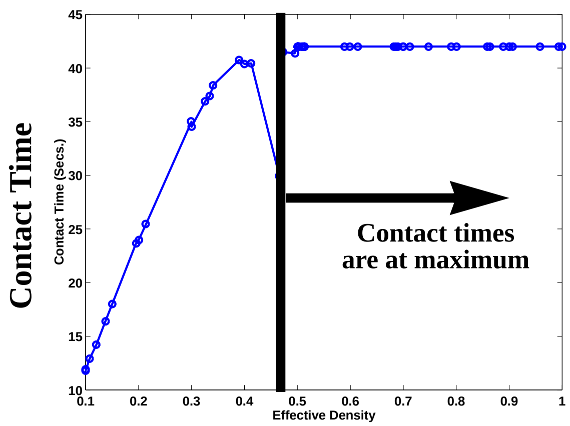

Figure 9. Extracted contact time as a function of the effective feature density.

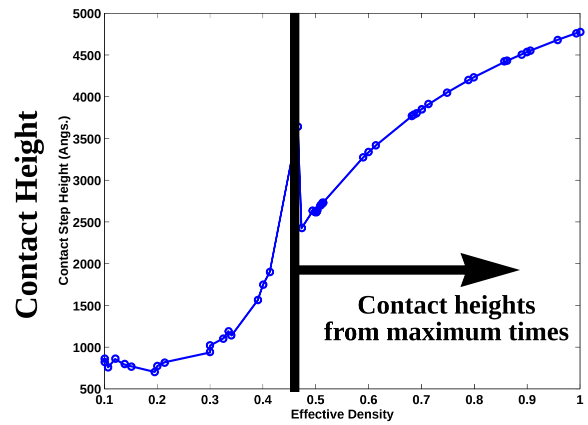

Figure 10. Step height at fitted contact time as a function of the effective feature density.

Time-Density Model Up and Down Area Predictions (Contact Step Height as a Function of Density)

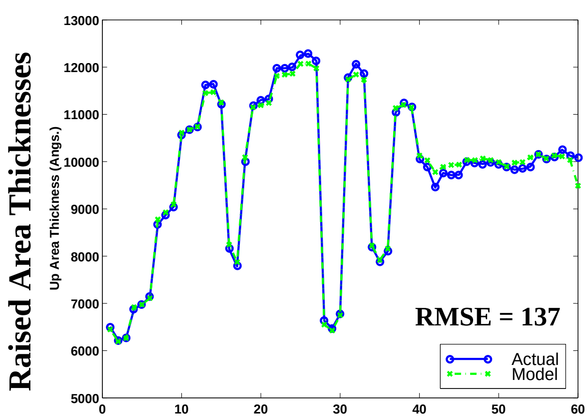

Figure 11. Actual and predicted thickness of the raised areas using time-density model with contact step height as a function of density.

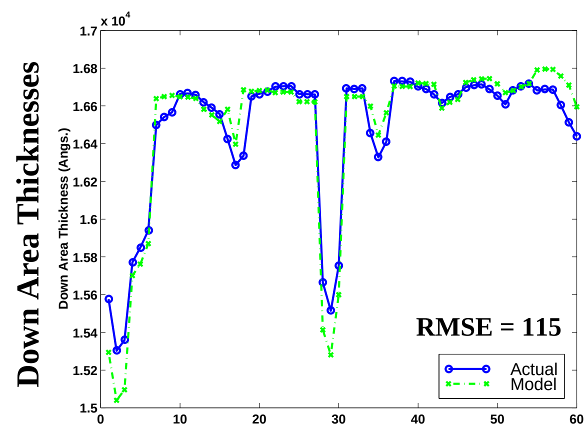

Figure 12. Actual and predicted thickness of down areas using time-density model with contact step height as a function of density.

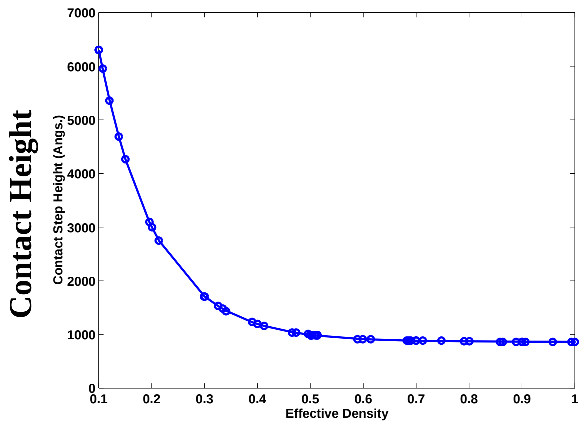

Figure 13. Step height at contact time as a function of the effective feature density.

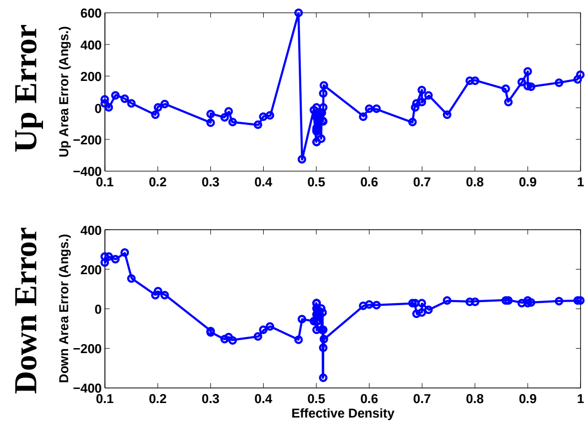

Figure 14. Model errors for raised and down areas as a function of the effective feature density.

Time-Density Model Up and Down Area Predictions (Contact Step Height as a Function of Line Space)

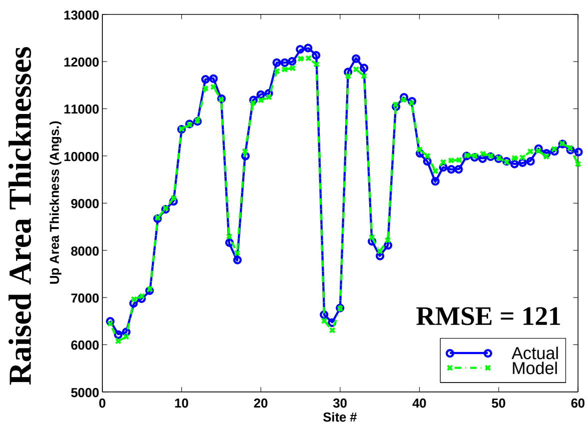

Figure 15. Actual and predicted thickness of the raised areas using the time-density model with step height as a function of line space.

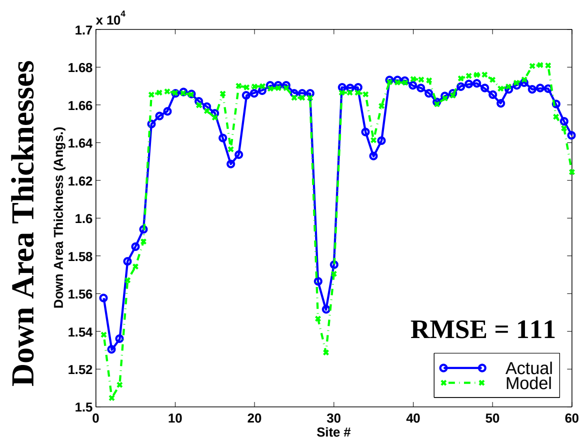

Figure 16. Actual and predicted thickness of the down areas using the time-density model with step height as a function of line space.

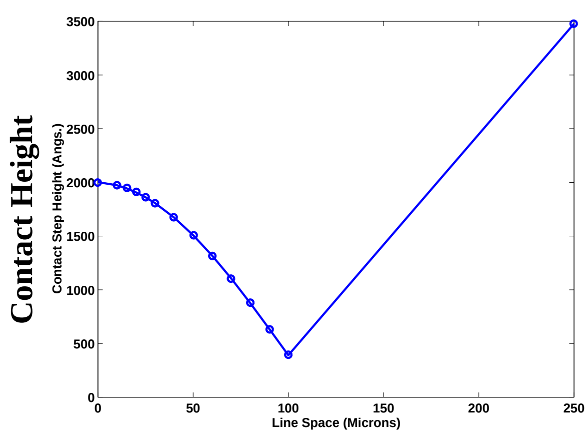

Figure 17. Fitted step height at contact time as a function of line space.