BIAS AND VARIANCE IN SINGLE AND MULTIPLE RESPONSE SURFACE MODELING

MIT Microsystems Technology Laboratories, Cambridge MA 02139

Abstract — This work examines the single response surface (SRS) and multiple response surface (MRS) modeling of wafer uniformity using experimental data from a chemical-mechanical polishing process. It is shown that the SRS estimate of uniformity measures total variation whereas the MRS estimate measures only spatial variation. The experimental results demonstrate that the MRS method is a biased estimator, whereas the SRS method is not. Finally, it is shown that the bias in the MRS estimate can lead one to choose a non-optimal process recipe.

Introduction

The modeling of wafer-level uniformity in semiconductor manufacturing is becoming increasingly important as wafer sizes increase from 200mm to 300mm diameters. Original work by Guo and Sachs demonstrated the advantages of the multiple response surface (MRS) method over that of conventional single response surface (SRS) methods in modeling spatial uniformity [1]. However, that work did not directly consider the effects of noise in its comparison of these techniques. The SRS methodology has been used extensively and is well understood, both in terms of its ability to model non-random functions as well as the effect noise may have on the model parameters. The MRS (or “site modeling” [1,2]) technique, on the other hand, is fairly young and work is necessary to understand its potential as well as its limitations in order to use it properly. This paper continues the exploration of the MRS technique begun in [1]. In particular, we answer key questions regarding these estimation techniques; namely, is each method unbiased, and what is the variance of each method?

SRS and MRS Estimators in Estimating Uniformity

In single response surface (SRS) modeling, a parameter of interest (e.g. uniformity) is computed as a function of measured values (e.g. thicknesses), and then regressed over one or more repetitions, possibly as a function of the process conditions. Multiple response surface (MRS) modeling, on the other hand, first regresses a model for each measured value over one or more repetitions, and then computes the parameter of interest based on the multiple individual models.

Before we consider our experimental data, we would like to briefly examine the nature of these two modeling techniques. We consider the class of functions of the form

\[ y_{m,n} = f_n + w_{m,n}, \]  

where \( f_n \) is some deterministic spatial function dependent on the location of the site \( n \), and \( w_{m,n} \) is a noise term which is independent and identically distributed (i.i.d.) over \( N \) sites on each of \( M \) wafers. This class of functions assumes that the spatial function is not dependent on the wafer number (repetition). We believe that this class of functions represents a large number of processes in the semiconductor industry.

To use the SRS method to model the uniformity of this process, we first take each of the \( M \) wafers and calculate a sample variance. Then we find the least-squares estimate of the variance given the \( M \) sample variances as a function of the process settings. For now, we consider only a single process setting, and in this case, we need only estimate a constant (i.e. a fixed variance at a particular process setting) rather than a function (i.e. the variance as a function of process settings). Therefore, the least-squares estimate is equivalent to the mean of the \( M \) sample variances. The resulting SRS estimator calculates the sample variance from the \( N \) sites on each of the \( M \) wafers, and takes the mean over the \( M \) wafers to obtain an estimate of the process variance. The variance of a single wafer with our assumed structure is

\[ V(y_{m,n}) = E_n\{(f_n + w_{m,n} - E_n\{(f_n + w_{m,n})\})^2\}. \]  

(2)

where \( E_n\{\} \) represents the expectation over the set of sites, \( n \), on the wafer. Note that

\[ E_n\{(f_n + w_{m,n})\} = E_n\{f_n\} + E_n\{w_{m,n}\} = \bar{f} + \bar{w}. \]  

(3)

Combining (2) with (3) and simplifying we obtain

\[ V(y_{m,n}) = E_n\{(f_n - \bar{f})^2\} + E_n\{(w_{m,n} - \bar{w})^2\} - 2E_n\{(f_n - \bar{f})(w_{m,n} - \bar{w})\}. \]  

(4)

Since \( (f_n - \bar{f}) \) and \( (w_{m,n} - \bar{w}) \) are both zero mean functions over the surface of the wafer, the last term in (4) is zero. Thus, we see that the variance of one wafer is

\[ V(y_{m,n}) = V(f_n) + V(w_{m,n}). \]  

(5)

This statement says that if, at a single process setting, the output is composed of a fixed spatial function, and the noise across the surface of the wafer is i.i.d., then the variance across the wafer is the sum of the variance of the fixed function (which we will refer to as the spatial variation) and the variance of the noise (which
we will refer to as the site variation). Since the SRS estimator averages samples of this wafer level variation over several wafers, it is an estimate (average) of the combined spatial and site variation.

Now consider how we would use the MRS modeling technique to model the nonuniformity of this process. First, we model each site individually across the \( M \) wafers, and then calculate the variance of the resulting models. Again, because we have a process with a fixed setting, the least-squares model of the process output at each site reduces to finding the mean (across the \( M \) wafers) of the process output at each site. We then estimate the uniformity by calculating the sample variance of the \( N \) site models. In terms of our assumed model above, the average of each site across \( M \) wafers is:

\[
E_m\{f_n + w_{m,n}\} = E_m\{f_n\} + E_m\{w_{m,n}\} = f_n, \tag{6}
\]

since \( E_m\{w_{m,n}\} = 0 \) and \( f_n \) is non-random. The MRS estimator then calculates the variance across the site models, and thus, it is a calculation of the variance of the non-random spatial function alone. Therefore, when using the MRS technique for modeling uniformity, we obtain an estimate of the spatial variance alone, not the combined variance measured by the SRS method.

It should be pointed out that the MRS estimate of uniformity as we have defined it here (based on the method outlined in [1]) is not a correct estimate of the total variance of the wafer. Therefore, it is not strictly correct to compare these estimators directly (as they measure different things). However, in this paper we will test these estimators in their ability to approximate their respective limiting values. Comparing modified versions of these estimators which do measure identical quantities will be the subject of another paper.

**Mean and Variance of the SRS and MRS Techniques**

In order to examine the mean and variance of the SRS and MRS estimators, we applied these techniques to experimental data taken from a Strasbaugh 6EC chemical-mechanical polisher (CMP) at MIT. A 1 \( \mu \)m blanket thermal oxide layer on each of 15 unpatterned wafers was polished for 3 minutes with a typical process recipe (Process A). The wafers were measured at 49 sites using a KLA-Tencor UV1280. The final film thicknesses at each site on the 15 wafers from Process A are plotted in Fig. 1. Subsets of 2 wafers, 3 wafers, and so on up to 15 wafers were then randomly selected. Within each of these wafer subsets, site subsets of the 49 sites were randomly selected (namely 26 sites, 27 sites, and so on up to 49 sites). For each combination of wafers and sites, the SRS and MRS estimators were applied to the corresponding subset of data. For each number of wafers and number of sites combination, the process was repeated 500 times (each with a different particular combination). The sample mean and sample variance of the SRS and MRS estimators were calculated from these 500 samples for each number of wafers and number of sites combination.

The sample mean of the SRS and MRS estimators, as a function of the number of wafers and number of sites, are shown in Fig. 2. Since the SRS estimator is a measure of the combined spatial and site variation and the MRS estimator is a measure of the spatial variation alone, then the SRS estimator should be greater than the MRS estimator. From Fig. 2, we see that the MRS surface lies under the SRS surface, which supports this principle. Further, we can see from Fig. 2 that the sample mean of the SRS estimator is apparently not a function of the number of wafers or the number of sites used to model uniformity. Therefore, the SRS estimator appears to be unbiased (with respect to the number of wafers or number of sites). Note that this is only true if the value of the surface in Fig. 2 is equal to the sum of the true site noise and true spatial noise (which can only be verified with constructed data in which these are known). Fig. 2 also shows that the sample mean of the MRS estimator, on the other hand, changes as a function of the number of wafers used to calculate it. It does not, however, appear to be a function of the number of sites. This is a very interesting observation which we will return to later. We conclude that the MRS estimator of uniformity (spatial variance) is biased (with respect to the number of wafers). We also see in Fig. 2 that sample mean of the MRS estimator appears to be approaching a finite limit as the number of wafers used increases. As outlined above, this limit should equal the spatial variation alone.

As mentioned above, these results indicate that the MRS estimator of uniformity is biased as a function of the number of wafers. Also note that this is independent of the number of sites. If the bias were a function of the number of sites, and were decreasing as the number of sites were increased, then we might conclude that we could remove the bias in the estimator by measuring a large number of sites on the wafer. However, the results obtained indicate that, even if one has several hundred measurements per wafer, it does not drive the expected value of the estimator to the true value. Only the number of repetitions (wafers) aids in driving the expected value of the estimator to the true value. This demonstrates that even if the MRS estimator is less sensitive to noise (lower variance), it may be less sensitive around an incorrect value. We will see later how this has several implications when choosing between the SRS and MRS estimators.

With this thought in mind, we now examine the variation of the SRS and MRS estimators. The sample variances for the SRS and MRS estimators are shown in Figs. 3 and 4, respectively. We see that the form of these two surfaces are very similar. In particular, note that
both variances decrease as the number of wafers and number of sites increase. However, in both cases the decrease with respect to the number of sites is slower than is the decrease with respect to the number of wafers. We see that the rate of decrease with respect to the number of sites and number of wafers is very similar between the two. However, the SRS variance is always larger than the corresponding MRS variance. This is further illustrated by considering the difference of the MRS variance from the SRS variance, as shown in Fig. 5. Here we see that the difference surface is always positive, indicating that the SRS variance is always larger than the MRS variance. Note that the difference is decreasing as both the number of wafers and number of sites are increasing.

Consider now, the motivation for the use of the MRS estimator proposed in [1]. The main benefits claimed were a) the modeling of uniformity with MRS provides a lower error than with SRS given the same model order, b) that the use of MRS is less sensitive to noise, and c) that the use of the MRS technique provides directionality in the model. We outlined above that the variance of the MRS estimator is always lower than that of the SRS, and thus this experimental data supports the second claim. However, the largest difference in the standard deviations is $2.8 \times 10^3$. If we take the sample mean of the MRS estimator at this point and subtract the limiting value (approximately the point with the most sites and most wafers) we obtain an estimate of the error due to the bias at this point. This value is $5 \times 10^3$. Thus the error due to the bias is larger than the benefit gained by the smaller variance. Noting that both the SRS and MRS estimators used the same model order (a zeroth order model), we conclude that the first claim is not necessarily true. We will not directly address the third claimed benefit here. Rather, we continue with the assumption that the second and third proposed benefits are true. However, we would like to further discuss the ramifications of the error due to bias with multiple process settings.

An additional 10 wafers were processed using Process A, and an additional 10 wafers at a second process recipe (Process B), with a higher table speed. The MRS modeling technique was applied to model the uniformity as a linear function of the process recipes described above. The sample variance at each process recipe was calculated from these models. The number of sites was fixed at 49; however, for each calculation the number of wafers from Process A and Process B were varied over all possible combinations (of the 10 wafers polished at each process recipe). Each possible combination was repeated 150 times and the sample mean calculated. The sample mean of the MRS estimate of uniformity at each process recipe was calculated as a function of the number of wafers from Process A and the number of wafers from Process B. Fig. 6 shows the sample mean of the MRS estimate of uniformity of Process A as one surface and the sample mean of the MRS estimate of uniformity of Process B as the second surface. Clearly, we see that if we use all 20 wafers, then Process A is found to have a lower uniformity. However, this is not typical in practice. Rather, it is often the case that more repetitions are run at the center point of an experimental design of experiments (DOE), and one or two repetitions at other settings. Suppose, in our example, that Process B is the center point, where we run 6 wafers, and we only run 2 wafers at Process A. Then, from the point $(2,6)$ in Fig. 6, we see that we would choose Process B as the lower spatial variance process. As pointed out above, Process A is the lower true variance process, and we have just chosen the wrong recipe!

The previous example demonstrates that one should be extremely careful in the use of MRS modeling; not only for process modeling, but particularly in its use for optimization and control in order to achieve the benefits possible with the approach. In particular, these results indicate that we must either correct for the bias in the MRS estimator or generate the model with the same number of wafers at each process setting.

### Conclusions and Future Work

In conclusion, we have shown that the SRS estimator measures spatial and site variation, whereas the MRS estimator measures only spatial variation. We have demonstrated that the MRS estimator is biased as a function of the number of repetitions. In addition, the error due to this bias can be larger than any benefit gained by the smaller variance of MRS over SRS. Finally, we have demonstrated that under typical DOE scenarios, this bias error could lead one to choose an non-optimal process recipe.

Future work will consider how to predict the limiting value of the MRS estimator. In addition, we will explore methods to remove the bias in the MRS estimator. Finally, we will work to analytically determine the bias and variance of the SRS and MRS estimators.

### Acknowledgments

The authors thank J. Davis (TI) for helpful comments, and the NSF/SRC ERC for Environmentally Benign Semiconductor Manuf. and DARPA under contract #DABT63-95-C-0088 for supporting this work.

### References


Figure 1. Post-polish thickness for 15 wafers plotted against the measurement site number.

Figure 2. Expected value of the estimators as a function of the number of wafers and the number of sites used to model.

Figure 3. Variance of the SRS estimator as a function of the number of wafers and the number of sites used to model.

Figure 4. Variance of the MRS estimator as a function of the number of wafers and the number of sites used to model.

Figure 5. Variance of the SRS estimator minus the variance of the MRS estimator as a function of the number of wafers and the number of sites.

Figure 6. Expected value of the MRS estimator of uniformity at two process settings as function of the number of wafers run at each setting.