OVERLAY VARIABILITY DECOMPOSITION BY EXTRA SUM OF SQUARE AND SET OPERATIONS

Fangren Ji[1*], *Hoyoung Heo*[1]
[1]KLA-Tencor Corporation, Austin, Texas 78759, USA
Fangren.Ji@KLA-Tencor.com

ABSTRACT

Total variability in overlay can be attributed to alignment system, patterned wafer geometry, lithography process signature, and their combinations if there is any. The statistical method of extra sum of square has the advantage of extracting these three components by quantifying the amount of sum of square they explain in regression analysis. However, the order of these components coming into play matters to the final answer. By combining extra sum of square with set operations the restriction of order how multiple sources of variation make contributions to overlay is eliminated.

Keywords: Overlay, Variability Decomposition, Extra Sum of Square, Set Operations, Alignment, Patterned Wafer Geometry, Process Signature

INTRODUCTION

As feature size in semiconductor manufacturing decreases overlay control becomes extremely critical. The analysis of source of variation surrounded by the nature of complex process steps serves as a key in targeting robust overlay control. By the time metrology data is produced multiple variation contributors have come into play either solely or cooperatively, specifically by alignment system, PWG (Patterned Wafer Geometry), lithography signature from the scanner, and their combinations if there is any, within the limited research scope presented by this paper.

Extra sum of square in regression analysis [1] can be used to extract these three components. However, the order how these three components contribute to the total overlay variability measured at metrology phase after lithography process matters and therefore, the answer to how much variability in overlay can be explained by each of these three components is controversial because the answer is not unique. By combining set operations with extra sum of square, variability decomposition in this manner becomes order independent.

ALGORITHM OVERVIEW

Figure 1 summarizes the basic concept upon which this research topic is conducted and lists all the overlay variability components that will be statistically quantified in later sections.

It has become clear that sum of square will be used to represent the variability being decomposed throughout this paper. Before describing the algorithm in detail, here is a brief review regarding ANOVA in regression analysis where it states:

$$SS_{Total} = SS_{Reg} + SS_{Res} \qquad (1)$$

SS_{Total} indicates sum of square total, i.e. total variability; SS_{Reg} indicates sum of square regression, i.e. the variability that can be explained by the regression model; SS_{Res} indicates sum of square residual, i.e. the variability that cannot be explained by the regression model;

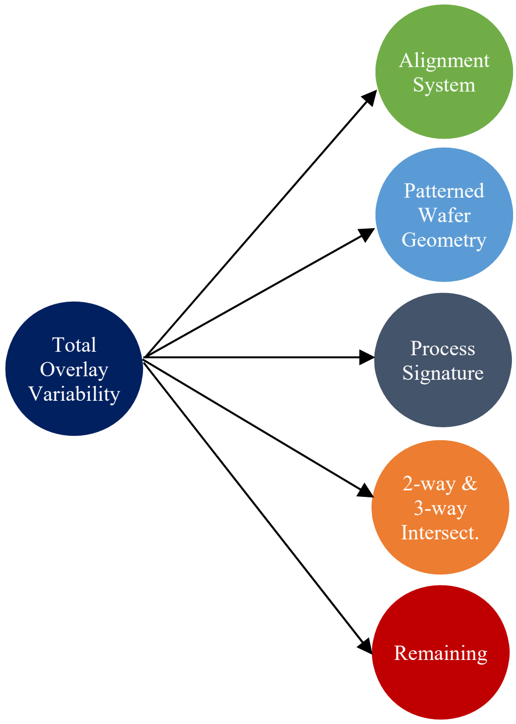

Figure 1: Variability Components to be Decomposed

ALGORITHM DETAIL

Based on what is summarized in figure 1, SS_{Reg} in equation (1) consists of all top four components in figure 1 whereas SS_{Res} is the remaining component. Combining the problem statement in figure 1 and the basic concept in equation (1) along with set operations, figure 2 visualizes the starting point of this statistical procedure in a Venn diagram.

There are three circles in the Venn diagram denoted by A, W and M. Table I provides detailed explanation about this notation.

Table I. Notation Used in Venn Diagram in Figure 2

Symbol	Meaning
A	Variability Explained by Alignment
W	Variability Explained by PWG
M	Variability Explained by Overlay Model

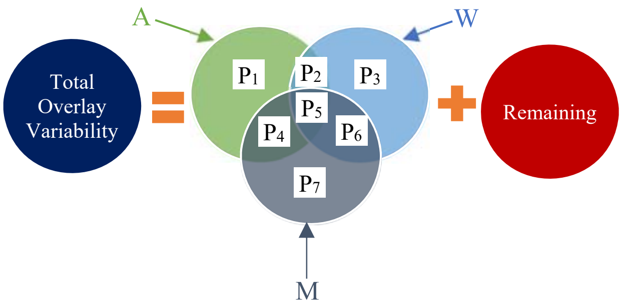

Figure 2: Venn Diagram for Variability Decomposition

P_1 to P_7 are the mutually exclusive variability subsets among A, W and M. Specifically,

$$A = P_1 + P_2 + P_4 + P_5 \tag{2}$$
$$W = P_2 + P_3 + P_5 + P_6 \tag{3}$$
$$M = P_4 + P_5 + P_6 + P_7 \tag{4}$$

It is worth noting that variability explained by process signature will be denoted by overlay model from now on for the purpose of convenience, because systematic signature from the scanner is typically quantified by overlay polynomial model.

With all the information having been provided so far, the solution to this problem comes down to how to solve P_1 to P_7 statistically. Intuitively, A, W and M have more direct physical meaning than P_1 to P_7. However, solving P_1 to P_7 is the key to interpreting in common language how overlay variability is being decomposed during lithography process.

Traditional overlay modeling involves only the target locations that forms the design matrix in linear regression. With another two major components namely alignment and PWG, the design matrix in this analysis will be expanded by adding two more columns that represent these two components respectively. Alignment data is usually one measurement per field. Therefore, each alignment measurement will be artificially repeated at all overlay target locations within the same field. Note that, the alignment data used here is the residual after alignment modeling because systematic signature or modeled part in raw alignment measurement is assumed to be corrected by the alignment system. Figure 3 shows how this artificial transformation step is performed.

Once repeating step is done a column vector containing corresponding alignment residuals to overlay measurements at each target location will be added to the design matrix. As far as the column vector representing PWG is concerned, predicted overlay at each target location will be used. There are several PWG prediction models currently being used in the industry such as IPD [2] which stands for In-Plane Distortion. The PWG model used in this analysis is an alternative predictive approach.

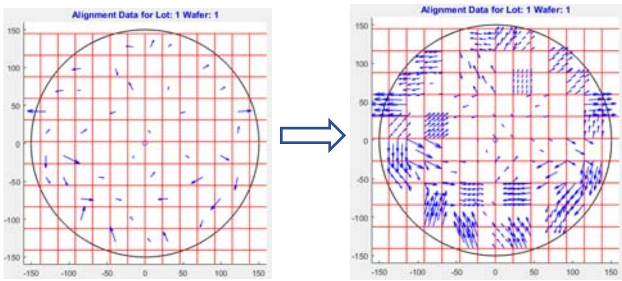

Figure 3: Repeated Alignment Measurements

For convenience purpose WX_i and WY_i will be used to denote each predicted overlay by the PWG model in this paper. Assuming there are three overlay measurements and first-order grid level overlay model with six terms is applied to our data, figure 4 shows the matrix multiplication that demonstrates the expansion of design matrix in our regression analysis.

$$\begin{pmatrix} WX_1 & ALNX_1 & 1 & 0 & Xw_1 & -Yw_1 & Xw_1 & -Yw_1 \\ WX_2 & ALNX_2 & 1 & 0 & Xw_2 & -Yw_2 & Xw_2 & -Yw_3 \\ WX_3 & ALNX_3 & 1 & 0 & Xw_3 & -Yw_3 & Xw_3 & -Yw_{31} \\ WY_1 & ALNY_1 & 0 & 1 & Yw_1 & Xw_1 & -Yw_1 & -Xw_1 \\ WY_2 & ALNY_2 & 0 & 1 & Yw_2 & Xw_2 & -Yw_2 & -Xw_2 \\ WY_3 & ALNY_3 & 0 & 1 & Yw_3 & Xw_3 & -Yw_3 & -Xw_3 \end{pmatrix} \begin{pmatrix} a \\ b \\ XTran \\ YTran \\ SymSca \\ SymRot \\ AsymSca \\ AsymRot \end{pmatrix} = \begin{pmatrix} OVLX_1 \\ OVLX_2 \\ OVLX_3 \\ OVLY_1 \\ OVLY_2 \\ OVLY_3 \end{pmatrix}$$

Figure 4: Design Matrix of Three Measurements

As shown above, $ALNX_i$ and $ALNY_i$ are the alignment measurements in X and Y direction whereas $OVLX_i$ and $OVLY_i$ are the decorrected overlay measurements in X and Y direction respectively. In the other column vector "a" and "b" are the coefficients to be estimated for PWG and alignment contribution respectively, while the other six model terms represent the typical parameters of 6-par linear grid model W1.

Due to the limited length of this paper one example of solving disjoint set P_1 will be explained as follows. All the other disjoint sets can be solved by applying similar routine with the same logic. Based on the Venn diagram in figure 2 and in terms of set operations, P_1 can be written as:

$$P_1 = A \cap (\overline{W \cup M}) \tag{5}$$

Equation (5) interprets P_1 as the sum of square that can be explained by alignment but not by either PWG or overlay model. Let us fit a model between decorrected overlay and both PWG and W1. The polynomial format of this model would be:

$$\text{OVL\_Dec} \sim \text{PWG} + \text{W1} \quad (6)$$

According to figure 2, the sum of square regression of this model would be:

$$SS_{RegWM} = P_2 + P_3 + P_4 + P_5 + P_6 + P_7 \quad (7)$$

Next let us fit the full model with all three components. The polynomial formula and model sum of square regression would be:

$$\text{OVL\_Dec} \sim \text{Alignment} + \text{PWG} + \text{W1} \quad (8)$$

and

$$SS_{RegAWM} = P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 \quad (9)$$

Combining equation (7) and (9) yields:

$$P_1 = SS_{RegAWM} - SS_{RegWM} \quad (10)$$

Equation (5) to (10) describes how to derive the quantity of P_1. All the other disjoint sets from P_2 to P_7 can be achieved similarly except 2-way and 3-way intersections can be a little more complex.

SIMULATON RESULTS

Prototyping for this algorithm was done in Matlab development environment. A dataset with multiple lots and multiple wafers per lot was used to conduct this simulation. The scale for each variability component can be a relative proportion to the total variance in decorrected overlay or an absolute value calculated based on the total σ^2. For a single wafer, pie chart is available to visualize all variability components from P_1 to P_7, plus the remaining component, while trend chart is more appropriate for multiple lots averaging the statistics among all wafers within each lot. Figure 5 and figure 6 show a pie chart for one single wafer and a trend chart for one single lot with 25 wafers respectively. A third-order composite-field model W3F3 was used in this case.

As described in algorithm detail section, displaying P_1 to P_7 from the Venn diagram may not be sufficiently intuitive. Instead, single major variability component or intersection of multiple components will be displayed in the legend. It is not difficult to map between these two formats. For instance, component "PWG&ALN" represent the sum of P_2 and P_5.

CONCLUSION

From the pie chart for a single sample wafer in figure 5, it appears that the major variability components PWG and alignment are not significant. W3F3 seems to explain a significant amount of variability in the decorrected overlay. All the 2-way and 3-way intersections are very

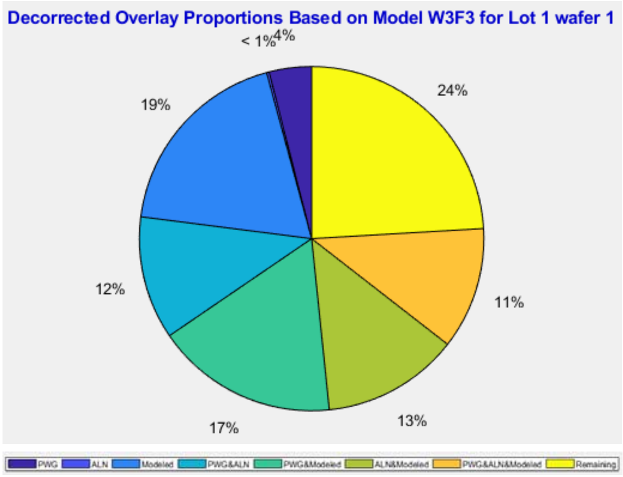

Figure 5: Single Wafer Variability Decomposition

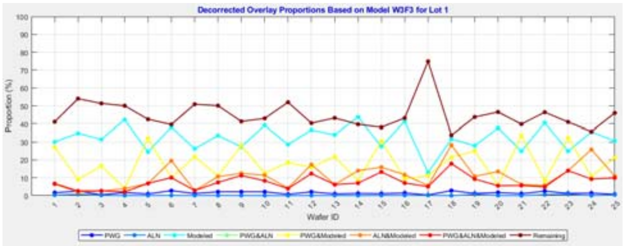

Figure 6: Single Lot Variability Decomposition

close to each other. Still there is 24% of unexplained variability which needs further investigation though it is outside of the limited scope of this study.

Trend chart for a single lot with 25 wafers in figure 6 shows significant wafer-by-wafer variation. The pattern in this variation seems to be coming from chuck variation, because odd and even numbered wafers are usually separated by stage. It is noticeable that one of the 25 wafers has higher remaining and lower modeled components than the others. This can be an indication of outlier.

The advantage of combining extra sum of square theory and set operations is that, the confounding effect between different major variability components can be quantified during modeling process while this confounding effect cannot be established by extra sum of square or set operation itself.

REFERENCES

[1] D. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis Fourth Edition*, Hoboken, NJ: Wiley, 2006.

[2] K. T. Turner, S. Veeraraghavan, and J. K. Sinha, *Relationship Between Localized Wafer Shape Changes Induced by Residual Stress and Overlay Errors*, J. Micro/Nanolithography., MEMS, MOEMS, Vol. 11(1), 2012, pp. 013001.