An IC Manufacturing Yield Model Considering Intra-Die Variations

Jianfeng Luo, Subarna Sinha, Qing Su, Jamil Kawa, and Charles Chiang
Advanced Technology Group, Synopsys Inc., Mountain View, CA, 94043 USA
{jianfeng, subarna, qsu, jamil, clc}@synopsys.com

ABSTRACT

In deep submicron feature sizes continue to shrink aggressively beyond the natural capabilities of the 193 nm lithography used to produce those features thanks to all the innovations in the field of resolution enhancement techniques (RET). With reduced feature sizes and tighter pitches die level variations become an increasingly dominant factor in determining manufacturing yield. Thus a prediction of design-specific features that impact intra-die variability and correspondingly its yield is extremely valuable as it allows for altering such features in a manner that reduces intra-die variability and improves yield. In this paper, a manufacturing yield model which takes into account both physical layout features and manufacturing fluctuations is proposed. The intra-die systematic variations are evaluated using a physics-based model as a function of a design's physical layout. The random variations and their across-die spatial correlations are obtained from data harvested from manufactured test structures. An efficient algorithm is proposed to reduce the order of the numerical integration in the yield model. The model can be used to (i) predict manufacturing yields at the design stage and (ii) enhance the layout of a design for higher manufacturing yield.

Categories and Subject Descriptors

B.7.2 [**Integrated Circuits**]: Design Aids

General Terms

Algorithms, Design, Verification

Keywords

Manufacturing Yield, Systematic Variation, Random Variation, Spatial Correlation, CMP

1. INTRODUCTION

A typical IC manufacturing process involves complex physical and chemical interactions that result in the targeted end parameters having finite variations centered around their intended values no matter how accurately controlled the manufacturing process is. Traditionally these variations manifested themselves in the form of lot to lot, wafer to wafer, and die to die variations that were design independent and that resulted in some acceptable yield loss. Yield improvement under that scenario involved tightening the process control over time. However, with feature sizes smaller than half the wavelength of the optically utilized light source and shrinking, and with corresponding tighter pitches, intra die variations are becoming dominant and are becoming most significant in determining manufacturing yield.

There are two components to die-level variations: (1) ***intra-die systematic variations*** and (2) ***intra-die random variations***. ***Intra-die systematic variations*** are strongly layout dependent. Two typical examples are: (i) intra-die critical dimension (CD) variations due to lithography and (ii) Cu & oxide thickness variations due to the Chemical-Mechanical Planarization (CMP) process. Empirical data shows that the intra-die systematic CD and thickness variations resulting from the layout pattern non-uniformity are becoming comparable to, even dominant over lot to lot, wafer to wafer, and die to die variations [1]. In recent years, a number of full-chip simulation tools have been developed to evaluate and predict intra-die systematic variations at the design step [2-3]. In addition to the intra-die systematic variation, the ***intra-die random variation*** remains an important component of the total die-level variation. This component reflects random fluctuations of process parameters within a die. The overall random variations tend to be spatially correlated in the current processes. When the lot to lot, wafer to wafer, and die to die random variations dominate the overall random variations, a perfect spatial correlation at die-level can be assumed. However, due to decreasing feature sizes and increasing die sizes, the spatial correlations of random variations between two intra-die locations do not exhibit ideal behavior any more. Instead the spatial correlation of intra-die variations decreases with increasing distances between locations. The non-perfect correlation between intra-die locations strongly impacts the total manufacturing yield.

Yield loss due to layout dependent intra-die variations cannot be rectified without some intelligent compensation at the design stage. However, the first step towards a yield-aware enhanced design is to be able to predict manufacturing yield at each design stage with some degree of accuracy. Previous work in design-stage yield prediction has focused primarily on random particle/defects related yield [4-5]. Unfortunately, random particles based models do not consider the impact of spatial correlations of intra-die variations and are therefore not sufficient to predict manufacturing yield in current and future processes.

In this paper, we propose a manufacturing yield model that can address the above-mentioned factors. To the best of our knowledge, this is the first work that addresses both intra-die systematic and all known forms of random variations. The key contributions of this paper are:

1. A yield prediction methodology that can take into account both intra-die systematic variations and all known forms of random variation. To the best of our knowledge, this is the first methodology that considers spatial correlations of intra-die variations in manufacturing yield prediction.
2. A new model to predict die yield after CMP and a novel, computationally efficient algorithm to evaluate the model.

DAC 2006, July 24–28, 2006, San Francisco, California, USA.

The remainder of the paper is organized as follows. In section 2, the variation decomposition for different time and space scales is discussed. Section 3 discusses the yield impact of CMP and a model for predicting yield loss due to this process. In Section 4, an algorithm for efficiently evaluating the yield prediction model associated with the CMP process is proposed. In section 5 simulation examples are used to demonstrate the efficiency of our yield model and algorithm. Discussion of the yield model, its applications and limitations are presented in section 6. Section 7 concludes the paper.

2. BACKGROUND

In this section, we provide a brief overview of the different types of process parameter variations.

2.1 Variation Decomposition

Process parameters vary at different scales in time and space. These variations can be classified as lot-lot, wafer-wafer, die-die, and intra-die [1]. The lot-lot and wafer-wafer variations are mainly temporal while the die-die variations are primarily spatial. Due to the long range nature of those variations, they are usually very low frequency in nature. This means their impact on two locations within a die are the same. Hence, it is reasonable to assume a perfect intra-die correlation for these components of random variations.

The intra-die variation can be decomposed into two components. One is the systematic spatial variation which is mainly caused by the layout dependency of the process. A key example of this variation is the pattern and density dependency of the intra-die oxide and Cu thickness in the CMP process. The other component of the intra-die variation is the random variation that is caused by typical process fluctuations around its nominal value and is spatially correlated.

Based on the above discussion, a quality indicative value to describe a process parameter for a location (x, y) is denoted as $p(x, y)$ and can be expressed as

$$\begin{aligned} p(x,y) &= \mu + f_{l\_l} + f_{w\_w} + f_{d\_d} + f_i(x,y) \\ &\quad + \varepsilon_{l\_l} + \varepsilon_{w\_w} + \varepsilon_{d\_d} + \varepsilon_{f\_f}(x,y) \\ &= \mu + f_{l\_l} + f_{w\_w} + f_{d\_d} + f_i(x,y) + \varepsilon(x,y) \end{aligned} \quad (1)$$

where μ is the overall mean, f_{l_l} is the systematic lot-lot variation, f_{w_w} is the systematic wafer-wafer variation, f_{d_d} is the systematic die-die variation, $f_i(x, y)$ is the systematic intra-die variation, ε_{l_l} is the random lot-lot variation, ε_{w_w} is the random wafer-wafer variation, ε_{d_d} is the random die-die variation and $\varepsilon_i(x, y)$ is random intra-die variation.

In the following discussion, we assume that $f_{l\_l} + f_{w\_w} + f_{d\_d} = 0$. This simplification does not affect our analysis. Therefore, Eq. (1) can be simplified as

$$p(x, y) = \mu(x, y) + \varepsilon(x, y) \quad (2)$$

where $\mu(x, y) = \mu + f_i(x, y)$ and $\varepsilon(x, y) = \varepsilon_{l\_l} + \varepsilon_{w\_w} + \varepsilon_{d\_d} + \varepsilon_{f\_f}(x, y)$.

Note that if the random variation was dominated by lot-lot, wafer-wafer, and die-die random variation then the correlation of $\varepsilon(x, y)$ at two intra-die locations (x_1, y_1) and (x_2, y_2) would be close to 1. But, when intra-die random variation is much larger than the sum of lot-lot, wafer-wafer, and die-die random variations, the correlation is solely determined by the intra-die component $\varepsilon_{f\_f}(x, y)$. In all other cases, the correlation lies somewhere between the correlations of intra-die random variations and 1.

2.2 Variations Handled by Our Model

In this work, we make the following assumptions on the nature of the intra-die variations.

If there are n locations that we are interested in, we represent the random variables p at the n locations by an n-dimensional random variable vector $\boldsymbol{p}$. Eq. (2) can be further written as

$$\boldsymbol{p} = \boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad (3)$$

where $\boldsymbol{\mu}$ is a n-dimensional vector representing the systematic components of the variables and $\boldsymbol{\varepsilon}$ is a n-dimensional vector representing the overall random components. We assume that $\boldsymbol{\varepsilon}$ satisfies a multivariate normal distribution $N(0, \sum)$, where $\sum$ is a n×n covariance matrix. Thus, $\boldsymbol{p}$ satisfies a multivariate normal distribution $N(\boldsymbol{\mu}, \sum)$. Furthermore, it is assumed that (i) the variances σ^2 of $\varepsilon(x, y)$ at each location are equal to each other. Then there is

$$\rho_{i,j} = \sigma_{i,j}/\sigma^2 \quad (4)$$

where $\sigma_{i,j}$ is the (i, j)th entry of $\sum$, $\boldsymbol{\rho}$ is the correlation matrix and $\rho_{i,j}$ the (i,j)th entry of the correlation matrix. This assumption is usually valid for most process parameters. (ii) the correlation is solely a function of distances.

3. CMP YIELD

We use CMP as an example to demonstrate our yield modeling methodology. CMP is a key enabling processes for planarization as well as for patterning copper interconnect in the deep sub-micron IC fabrication. The Cu interconnect is patterned primarily by four sequential steps: (i) a deposition process to form a dielectric layer on the wafer (ii) a plasma etch process to generate the trenches in the dielectric layer (iii) a deposition process to fill up the trenches with copper, and (iv) a CMP process to remove the bulk copper from the top of the dielectric layer leaving copper in the trenches as interconnect.

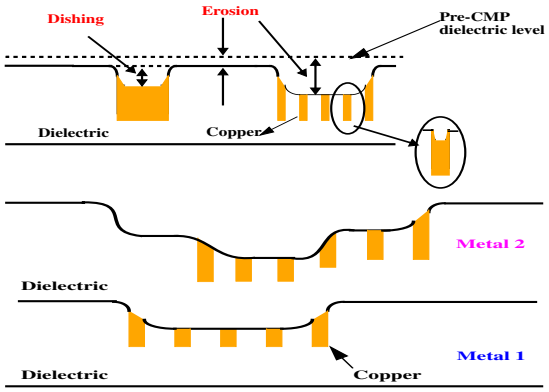

Figure 1: Typical post-CMP topography [2]

Ideally, a CMP process should produce perfectly flat Cu interconnects with uniform thickness across the wafer. However, this is not the case in reality because the Cu thickness at a location (x, y) on the wafer is affected by different layout patterns in a design, and a number of process parameters. The different layout patterns can be characterized by layout density, layout perimeters, line width, and so on [2]. The layout patterns are typically not uniform in a design. The process parameters include the incoming Cu deposition thickness, barrier deposition thickness, Cu polishing rate, barrier/oxide polishing rate, CMP down pressure, velocity, polishing time, and so on [9]. They usually fluctuate around their nominal values.

Fig. 1 shows typical Cu thickness variation after CMP. The thickness variation can be caused by underlying layout differences and/or by random variations of the process parameters. The thickness variations include copper dishing, dielectric erosion, and multi-layer cumulative topographies variation. When the thickness variation is large, it may cause a circuit open and/or a circuit short. For instance, when the copper is not totally removed from the top of the dielectric layers, a circuit short occurs. On the other hand, if the copper in the trench is over polished, a circuit open occurs. The variation can also cause defocus issues in the

lithography process following CMP. This can result in zero yield in extreme cases.

3.1 CMP Yield Prediction Model

To ensure a chip's functionality and yield after CMP it is necessary to ensure that the Cu thickness falls within a specification bounded by an upper specification limit (USL) and a lower specification limit (LSL). This specification is dictated by the design requirements on one hand and manufacturing capabilities on the other.

We define ***CMP Yield*** as the probability that all thickness values across a die fall between USL and LSL.

Mathematically, yield Y can be expressed as:

$$Y = \int_L^U \int_L^U \int_L^U ...\Phi(\boldsymbol{p})dp_1dp_2...dp_n$$
$$= \int_L^U \int_L^U \int_L^U ...\frac{e^{\left(-0.5(\boldsymbol{p}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{p}-\boldsymbol{\mu})\right)}}{\sqrt{(2\pi)^n|\Sigma|}}dp_1dp_2...dp_n \quad (5)$$

where U is used to denote USL and L is used to denote LSL, n denotes the number of locations where the thickness is monitored in the yield prediction process and Φ denotes the joint distribution of the thickness variation at these n locations. The joint distribution of the thickness at the n locations can be written as:

$$\Phi(\boldsymbol{p}) = \frac{e^{\left(-0.5(\boldsymbol{p}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{p}-\boldsymbol{\mu})\right)}}{\sqrt{(2\pi)^n|\Sigma|}} \quad (6)$$

where $\boldsymbol{p}$ is the n-dimensional random variable vector and $|\sum|$ is the determinant of the covariance matrix [7].

Typically, a large number of locations need to be tracked for good yield prediction. A typical number in the order of 10^5 to 10^6 is needed to predict the systematic thickness variation $\boldsymbol{\mu}$ accurately. For example, for a chip size of 4mm×4mm, a number of 160,000 locations are needed by meshing the chip into 10μm×10μm tiles.

A direct numerical integration of Eq. (5) with dimension n in the order of 10^5 to 10^6 is not practically feasible. It would require a large amount of computation time and a huge amount of memory without guaranteeing any numerical accuracy. Another issue with the above method is that a lot of manufacturing data would be needed to populate the covariance matrix. A chip of size 4mm×4mm with a mesh size of 10μm would contribute 160,000 locations across the chip. The covariance matrix $\sum$ as a function of locations has to be obtained from manufacturing by using test structures similar to those in [6]. For our example, to obtain the covariance / correlation matrix needed, measurements at $(400^2+400^2)^{1/2} = 576$ locations (around one location every 10μm) are ideally needed. This is an impractical and expensive proposition.

In the next section, we propose a novel and computationally efficient algorithm that can reduce the order of numerical integration and also reduce the amount of test-data needed without sacrificing accuracy.

4. YIELD PREDICTION ALGORITHM

In the following, we present yield prediction algorithms for two cases: (1) special case: perfect correlation of random variations across the die (2) general case: correlation across the die decreasing with distance.

The first case can be used to predict yield for small dies or when correlation data is unavailable from the fab. However, for large dies it is incorrect to assume a perfect correlation. Typical manufacturing data suggests that the correlation decreases gradually with distance. For such cases, we propose a novel computationally efficient algorithm in Section 4.2 for evaluating yield while taking spatial correlations into account.

4.1 Perfect Correlation across the Die

When assuming a perfect correlation between the random variations at any two intra-die locations all entries in the correlation matrix are equal to 1. This causes the rank of $\sum = 1 \leq$ n (or $\sum$ to not be a full-rank matrix). Thus the inverse of $\sum$ does not exist. In this case Eqs. (5) and (6) cannot be used to calculate the joint probability of the Cu thickness at the n locations. Assuming that the maximum nominal Cu thickness of the n locations is Max and that the minimum nominal Cu thickness is Min as shown in Fig. 2 it follows that the probability for all Cu thickness to fall between the USL and LSL (therefore the yield) can be calculated as:

$$Y = \int_L^\infty PDF(Min, x)dx + \int_{-\infty}^U PDF(Max, x)dx - 1 \quad (7)$$

where PDF is the univariate normal distribution function. The first integral is the probability for all Cu thickness values to fall above LSL. The second integral is the probability for all Cu thickness values to fall below USL. Their sum minus 1 is the probability that all Cu thickness values fall between USL and LSL.

Eq. (7) implies that the probability Y of the Cu thickness value at all locations to fall within the specification is solely determined by two locations: one with the maximum nominal thickness Max and the other with the minimum nominal thickness Min. All other locations do not affect the overall probability. The perfect correlation and the equal variances of the random variations across the chip ensure that the thickness values at all locations are co-varying and the variation amounts are the same. Therefore, the location with a maximum nominal thickness always has a maximum thickness and the location with a minimum nominal thickness always have a minimum thickness. Mathematically, this can be attributed to the degeneration of the n×n covariance matrix into a scalar. The probability of the two locations to be within the specification is the same as the probability that all thickness values within the die fall within the specification.

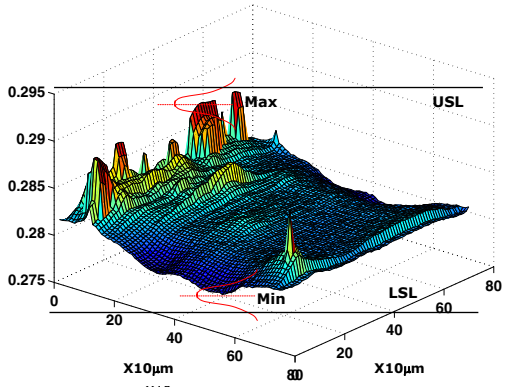

Figure 2: Yield prediction for the case of perfect across-chip correlation

4.2 Correlation decreases with distance

Manufacturing data shows that correlation decreases very gradually with increasing distance. The left sub-plot of Fig. 3 shows a high correlation (= 0.99) of the fab measured post-CMP thickness values for two locations which are 80μm away from each other. The slope of the line in the left sub-plot indicates the two locations approximately have the same variance. Another set of fab data shows that as distance of locations increases to as large as 4000μm, the correlation is still as high as 0.88, right sub-plot of Fig. 3. Hence, it is safe to assume that the correlation between locations within a certain distance, say 200μm, is equal to 1. Thus, a much coarser grid can be used for sampling and obtaining

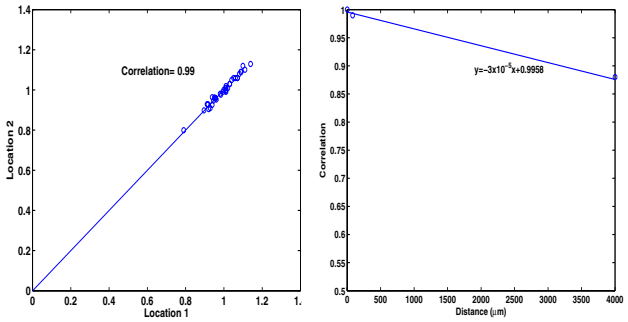

Figure 3: Spatial correlations of random variation as function of distances

manufacturing data. This fact is used in our algorithm to reduce the order of the numerical integration greatly, thereby speeding up the computation. Note that the exact relationship between correlation and distance may be linear or non-linear, depending on the process specifics. An example of a typical linear relationship between correlation and distance is shown in the right graph of Fig. 3.

To compute yield, we first compute the probability that the thickness at all locations falls below USL (denoted as Y_{max}) and the probability that the thickness at all locations falls above LSL (denoted as Y_{min}). When the across-chip correlation of the random variation is relatively strong, which is usually the case for a mature process, the probability for all tiles to fall between LSL and USL (Y in Eq. (5)) can be approximated as

$$Y \approx Y_{max} + Y_{min} - 1 \tag{8}$$

Note that Eq. (8) is an exact solution in two cases: (i) USL $= \infty$ ($Y = Y_{min}$) and (ii) LSL $= -\infty$ ($Y = Y_{max}$).

An important assumption that the algorithm makes is that random variations are perfectly correlated and their variances are approximately equal to each other for tiles that are within a certain distance of each other. The data presented in Fig. 3 shows that these assumptions are realistic and holds for typical manufacturing processes.

DEFINITION 1. *A* **perfect correlation circle (PCC)** *of a tile is a region surrounding the tile where the random variation is perfectly correlated with the random variation in the tile. The variances of the surrounding tile are equal to the variance of the tile. All the tiles that overlap a PCC are said to be* **covered** *by the PCC.*

Typically, the radius of the PCC is a fixed value and depends on a particular process. Next, we discuss an algorithm that uses the concept of PCC to efficiently compute Y_{max} and also present an illustrative example.

Algorithm **Compute_**Y_{max} can be summarized as follows:

1. Mesh the chip into small tiles.
2. Evaluate the thickness value of each tile using a CMP simulator. This captures the systematic layout dependent intra-die variation.
3. Set $m = 0$; $S_{PCC} = \phi$.
4. While there exists at least one tile that is not covered by any of the PCCs in S_{PCC},
   (a) Identify the tile T_m with maximum nominal thickness that is not covered by any PCC in S_{PCC}.
   (b) Determine the PCC C_m of T_m.
   (c) $S_{PCC} \leftarrow S_{PCC} \cup C_m$.
   (d) $m \leftarrow m + 1$.
5. Obtain correlations between the m (maximum nominal thickness) tiles as a function of their distances. Manufacturing data as that in right graph of Fig. 3 is usually used for this step.
6. Build a $m \times m$ correlation/covariance matrix for the m locations.
7. Compute Y_{max} to be equal to the probability of these m locations to fall below USL. Genz' algorithm (detailed in Section 4.2.1) is used to efficiently compute this value.

The above procedure significantly reduces the size of the problem by clustering the tiles into groups and representing each group with a single tile. Thus, the correlation matrix has smaller dimension. This is similar to the case in the previous section, where the whole correlation matrix degenerates into a scalar.

A similar method can be used to evaluate Y_{min}, which is the probability of the thickness values at all location to be greater than the LSL. In this case, PCCs are placed around the tiles with minimal nominal thickness values. The overall yield is then calculated by Eq. (8).

Below, we present an example to illustrate in further detail the steps of the **Compute_**Y_{max} algorithm and provide a sense of the run-time benefits.

EXAMPLE 1. *Assume a chip with a small size of 120μm × 130μm (shown in Fig. 4). The chip is meshed into 10μm×10μm tiles. It is assumed that the nominal Cu thickness μ(x, y) in all tiles is obtained by a CMP simulator. The radius of the PCC is set to 3.5× the tile size, graphically illustrated using circles of radius= 35μm in Fig. 4.*

First the tile with the maximum nominal thickness across the whole chip is found and is denoted as Max1 (Fig. 4). All the tiles in the PCC around Max1 are contained in circle 1. Next, the tile with the maximum nominal thickness outside of circle 1 is located and is denoted as Max2 (Fig. 4). The tile with the maximum nominal thickness outside of the union of circle 1 and circle 2 is denoted as Max3. The above procedure is repeated until all the tiles in the die are covered by a circle. The chip in Fig. 4 is covered by 8 circles surrounding 8 maximum locations from Max1 to Max8.

The tiles in a PCC are perfectly correlated. In addition, the variance of random variations is the same for all tiles in the PCC. Then the probability of the thickness values at tiles surrounding Max1 (mint green tiles) to be smaller than USL is solely determined by Max1 similar to that shown in the case of Fig. 2. Outside the circle surrounding Max1, Max2 can represent the tiles with the highest probability to fall beyond USL (red tiles). Similarly, outside the union of circles surrounding Max1 and Max2, they are Max3 and its surrounding tiles. After the above procedure the nominal thickness values Max1∼ Max8 at the 8 maximum locations as well as the correlations between the 8 locations (functions of their relative distances) are obtained. The perfect correlation in a PCC ensures that the probability for the thickness values of the 8 locations to fall below the USL is close to the probability for the thickness values of all the locations on the die to fall below USL.

From the correlation matrix ρ_{max} of the 8 locations we can obtain their covariance matrix $\sum_{max}$ and calculate their joint distribution using Eq. (5). The only difference here is that -∞ should replace LSL L in Eq. (5) to calculate Y_{max}. Note that the size of the correlation matrix is 8×8. This is much smaller than that of 12×13×12×13= 156×156 obtained when applying Eq. (5) directly to all the tiles. For a chip with a typical size of 4mm×4mm, the size of the correlation matrix can be reduced from 160,000×160,000 to around 2,500×2,500. When the perfect-correlation distance is increased from the 35μm to 200μm the size of the correlation matrix (around 100×100) gets even smaller.

The above example illustrates clearly how the order of integration can be significantly reduced using the proposed algorithm. However, a typical integration order after the reduction could still be in the range of several hundreds. In such a situation a direct numerical integration is still not feasible. A numerical algorithm in next subsection proposed by Genz [7] is used to solve this problem.

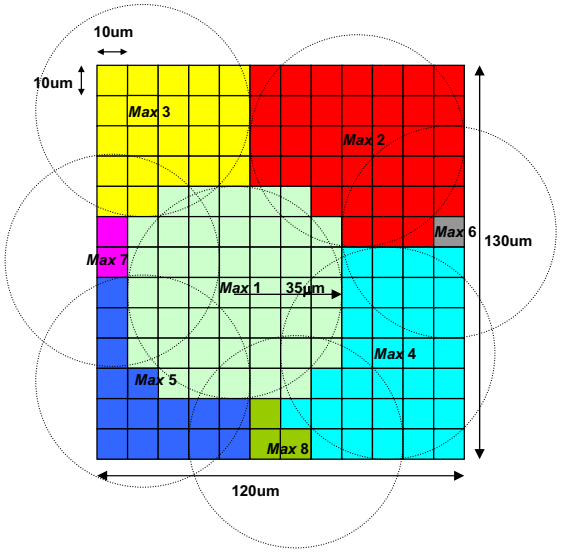

Figure 4: Perfect correlation circles surrounding maximum thickness values

4.2.1 *Genz's Algorithm for Multi-Dimensional Numerical Integration*

The key of Genz's algorithm involves three transformations of the covariance matrix $\sum_{max}$ after the order reduction. They are described briefly below. For a detailed explanation the reader is referred to [7].

***Transformation* 1:**

Because the covariance matrix is symmetric and positive definite it can be decomposed into two triangular matrices through Cholesky decomposition:

$$\sum = \boldsymbol{C}\boldsymbol{C}^T \tag{9}$$

where $\boldsymbol{C}$ is a lower triangular matrix and $\boldsymbol{C}^T$ is a upper triangular matrix. Let $\boldsymbol{p}\text{-}\boldsymbol{\mu} = \boldsymbol{C}\boldsymbol{y}$, there is $\mathrm{d}\boldsymbol{p} = \mathrm{d}p_1\mathrm{d}p_2\ldots\mathrm{d}p_n = |\boldsymbol{C}|\mathrm{d}\boldsymbol{y} = |\boldsymbol{C}|\mathrm{d}y_1\mathrm{d}y_2\ldots\mathrm{d}y_n$ and

$$\boldsymbol{L}\text{-}\boldsymbol{\mu} \le \boldsymbol{C}\boldsymbol{y} \le \boldsymbol{U}\text{-}\boldsymbol{\mu} \tag{10}$$

where $\boldsymbol{y}$ is a n-dimensional vector, y_1, y_2.. y_n are the n components of the vector $\boldsymbol{y}$, $\boldsymbol{L}$= (L, L, ... L) a n-dimensional vector representing the lower specification limit and $\boldsymbol{U}$=(U, U, ... U) a n-dimensional vector representing the upper specification limit.

For each component y_i in vector $\boldsymbol{y}$, Eq. (10) can be further written as

$$[L-\mu_i-\sum_{j=1}^{i-1}(C_{ij}y_j)]/C_{ii} \le y_i \le [U-\mu_i-\sum_{j=1}^{i-1}(C_{ij}y_j)]/C_{ii} \tag{11}$$

Note that $\boldsymbol{C}$ is an upper triangular matrix. Thus only y_m ($m < i$) determines y_i's upper limit and bottom limit. This property is critical for transformations 2 and 3.

Substitution of Eqs. (9)- (10) into Eq. (5) yields:

$$Y = \frac{1}{\sqrt{(2\pi)^n}} \int_{L_1'}^{U_1'} e^{-\frac{y_1^2}{2}} \ldots \int_{L_i'(y_1,y_2,\ldots y_{i-1})}^{U_i'(y_1,y_2,\ldots y_{i-1})} e^{-\frac{y_i^2}{2}} \ldots \int_{L_n'(y_1,y_2,\ldots y_{n-1})}^{U_n'(y_1,y_2,\ldots y_{n-1})} e^{-\frac{y_n^2}{2}} dy_n dy_{n-1}\ldots dy_1 \tag{12}$$

where $L_i'(y_1,y_2,\ldots y_{i-1}) = [L-\mu_i-\sum_{j=1}^{i-1}(C_{ij}y_j)]/C_{ii}$

and $U_i'(y_1,y_2,\ldots y_{i-1}) = [U-\mu_i-\sum_{j=1}^{i-1}(C_{ij}y_j)]/C_{ii}$.

***Transformation* 2:**

Let

$$Z_i = \Phi(y_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_i} e^{-\frac{1}{2}x^2} dx \tag{13}$$

which is the standard univariate normal distribution function. Then Eq. (12) can be further written as

$$Y = \int_{L_1''}^{U_1''} \ldots \int_{L_i''(z_1,z_2,\ldots z_{i-1})}^{U_i''(z_1,z_2,\ldots z_{i-1})} \ldots \int_{L_n''(z_1,z_2,\ldots z_{n-1})}^{U_n''(z_1,z_2,\ldots z_{n-1})} dz_n dz_{n-1}\ldots dz_1 \tag{14}$$

where

$$U_i''(z_1,z_2,\ldots z_{i-1}) = \Phi(U-\mu_i-\sum_{j=1}^{i-1}C_{ij}\Phi^{-1}(z_j))/C_{ii}$$

and $L_i''(z_1,z_2,\ldots z_{i-1}) = \Phi(L-\mu_i-\sum_{j=1}^{i-1}C_{ij}\Phi^{-1}(z_j))/C_{ii}$

***Transformation* 3:**

The final transformation is to let

$$z_i = L_i'' + w_i(U_i'' - L_i''). \tag{15}$$

Eq. (14) is written as

$$Y = (U_1''-L_1'')\int_0^1 (U_2''-L_2'')\ldots \int_0^1 (U_i''-L_i'')\ldots \int_0^1 (U_n''-L_n'')dw_n dw_{n-1}\ldots dw_1 \tag{16}$$

where

$$U_i'' = \Phi(U-\mu_i-\sum_{j=1}^{i-1}C_{ij}\Phi^{-1}(L_j''+w_j(U_j''-L_j'')))/C_{ii}$$

and $L_i'' = \Phi(L-\mu_i-\sum_{j=1}^{i-1}C_{ij}\Phi^{-1}(L_j''+w_j(U_j''-L_j'')))/C_{ii}$

Eq. (16) can be evaluated by a variety of numerical integration algorithms. The advantage of the above transformations is that it forces a priority ordering on the integration variables. Among all integration variables in Eq. (16) w_1 is the most important because all $(U_j''-L_j'')$ depend on it while w_n is the least important. This priority ordering can make numerical integration methods such as subregion adaptive algorithm work more efficiently and save computation time [7]. A simple but effective Monte-Carlo algorithm incorporating the transformations is proposed by Genz [7]. Besides Y, the algorithm can also output the numerical error estimate Err [7].

5. SIMULATION EXAMPLES

A Matlab program was written to implement the above algorithm. A chip (130nm technology) with size 4.8mm× 7.5mm is used as an example to show its efficiency. Metal layers 2-4 (M2-M4) are selected for our simulation. The USL, nominal value, and LSL of the copper thickness were set as 4580A, 3580A, and 2580A respectively, for all the three layers. The standard deviation of the Cu thickness variation was 0.012μm. The variance is 0.012×0.012μm^2. We assumed a linear reduction of the correlation with the distance and the function $y = -3\times10^{-5}x + 0.9958$ in Fig. 3 was used.

We obtained the nominal Cu thickness variation from a CMP simulation tool, which meshes the chip into 10μm×10μm tiles. The simulation tool captures the major systematic

variations including erosion, dishing, multi-layer accumulative topography variation and so on induced by layout pattern variations. The nominal Cu thickness values in matrix (size= 480×750) form, the variances and the correlation function were then input into the Matlab program for yield prediction. The yield prediction was done on a SUN UltraSPARC-II machine (CPU: 4×400MHz, RAM: 4G).

Table 1 shows the yield prediction under different radii of PCCs. It is seen that the yield prediction has converged at a radius of 1000μm for all three layers. This indicates that our algorithm works well from the convergence perspective. We tried several other chips and similar convergence at 1000μm was observed. The CPU time increased from 157 seconds at a radius of 1000μm to 1152 seconds at a radius of 350μm with no improvement in accuracy. Therefore, 1000μm can be chosen as the radius of PCCs for both efficient and accurate yield prediction. Note that the yield prediction at a PCC radius 10,000μm is the case when assuming a perfect correlation across the whole chip (chip size= 4.8mmx7.5mm). The results indicate that the yield can be greatly overestimated in this case.

PCC Radius (μm)	Yield (%)			Integration Order (Ymax /Ymin)		
	M2	M3	M4	M2	M3	M4
10000	95.1	47.0	95.1	1/1	1/1	1/1
5000	94.3	43.8	95.1	2/2	2/2	2/2
4000	94.3	41.4	95.1	4/2	4/3	3/2
3000	94.3	39.8	94.9	5/6	5/5	5/2
2000	94.3	39.2	94.4	7/8	10/9	9/11
1000	94.3	36.7	93.9	26/29	28/23	28/30
800	94.3	36.7	93.9	41/35	44/38	41/42
600	94.2	35.2	93.8	62/71	72/63	66/68
500	94.2	34.1	93.7	90/107	106/89	83/99
350	94.2	34.1	93.6	193/204	205/193	195/201

Table 1: Yield prediction under different radii of PCCs

Note that the numerical error Err increases with the order of the numerical integration (or the decrease of the radius of the PCCs and the increase of the number of PCCs covering the chip). However, no error larger than 0.8% is observed in our experiments.

M3 shows significantly lower yield than M2 and M4. A detailed study of the nominal thickness values of the three layers shows that it is because the minimal nominal thickness values on M3 are closer to LSL.

To further study how the yield changes with the correlation function, we assumed a correlation-distance function $y = -7 \times 10^{-5}x + 1$ and re-ran the yield prediction. Data in Table 2 shows that the yield prediction converges at a smaller radius of PCC. In addition, the yield under the new correlation function is lower. The lower correlations of the random variations between intra-die locations contribute to the lower yields.

6. DISCUSSION

The above simulation shows that both the intra-die systematic variation and the spatial correlation of random variations can significantly affect yield. The intra-die systematic variation is primarily determined by the design layout. The random variation and its spatial correlation are determined by the manufacturing variability (fluctuation). Traditionally reducing peak-peak intra-die systematic variations, for example, the nominal thickness variation in CMP [8], is the sole objective of various layout design optimization algorithms. This methodology, without considering the random variations and their spatial correlation, may not be sufficient. As an example, consider two chips with the same peak-peak range of systematic variations but different distributions of the peaks: one has one single peak, the other have more than one peak and the peaks are far away from each other. The yield for the chip with more peaks should be lower than the yield of the chip with one peak only considering the low correlation between the random variations at the peaks. Thus a dummy filling algorithm which trades off the peak-peak range and the number of peaks may be needed.

PCC Radius (μm)	Yield (%)			Integration Order (Ymax /Ymin)		
	M2	M3	M4	M2	M3	M4
10000	95.1	47.0	95.1	1/1	1/1	1/1
5000	93.5	39.4	95.0	2/2	2/2	2/2
4000	93.5	33.4	94.9	4/2	4/3	3/2
3000	93.5	30.3	94.5	5/6	5/5	5/2
2000	93.4	28.1	93.4	7/8	10/9	9/11
1000	93.4	24.0	92.3	26/29	28/23	28/30
800	93.4	21.9	92.2	41/35	44/38	41/42
600	93.4	19.7	92.1	62/71	72/63	66/68
500	93.4	18.9	92.1	90/107	106/89	83/99
350	93.4	18.6	91.2	193/204	205/193	195/201

Table 2: Yield prediction for a quicker decrease of the spatial correlation

The above model does not consider the lot-lot, wafer-wafer, and die-die systematic variations. These variations can be incorporated into the model easily once they are available. In addition the yield model and associated algorithm proposed may be extended to other processes such as the lithography process, and plasma etch process.

7. CONCLUSION

In this paper a manufacturing yield model which takes into account both physical layout features and manufacturing fluctuations is proposed. The intra-die systematic variations are evaluated by a physics-based model as a function of a design's physical layout. The random variations and their across-die spatial correlations are obtained from data harvested from manufactured test structures. An efficient algorithm is proposed to reduce the order of the numerical integration in the yield model. The model can be used to (i) predict manufacturing yields at a design stage and to (ii) enhance the layout of a design for higher manufacturing yield.

8. REFERENCES

[1] BE Stine, DS Boning and JE Chung, "Analysis and decomposition of spatial variation in integrated circuit processes and devices," *IEEE Trans. Semi. Manu.*, Vol. 10, pp. 24-41, 1997.

[2] TE Gbondo-Tugbawa, "Chip-scale modeling of pattern dependencies in copper chemical mechanical polishing processes," *Ph.D. Dissertation, Dept. of EECS*, MIT, MA, USA, 2002.

[3] LM Liebmann, SM Mansfield, et al., "TCAD development for lithography resolution enhancement," *IBM Journal of Research and Development*, Vol. 45, pp. 651-665, 2001.

[4] AR Dalal, PD Franzon and MJ Lorenzetti, "A layout-driven yield predictor and fault generator for VLSI," *IEEE Trans. Semi. Manu.*, Vol. 6, pp. 77-82, 1993.

[5] IS Wagner and I Koren, "An interactive VLSI CAD tool for yield estimation," *IEEE Trans. Semi. Manu.*, Vol. 8, pp. 130-138, 1995.

[6] JP Cain and CJ Spanos, "Electrical linewidth metrology for systematic CD variation characterization and causal analysis," *Proceedings of SPIE*, Vol. 5038, pp. 350-361, 2003.

[7] A Genz, "Numerical computation of multivariate normal probabilities," *Journal of Computational and Graphical Statistics*, Vol. 1, pp. 141-149, 1992.

[8] R Tian, X Tang and DF Wong, "Dummy feature placement for chemical-mechanical polishing uniformity in a shallow trench isolation process," *IEEE Trans. CAD*, Vol. 21, pp. 63-71, 2002.

[9] J Luo and D Dornfeld, *Integrated Modeling of Chemical Mechanical Planarization for Sub-Micron IC Fabrication: from Particle Scale to Feature, Die and Wafer Scales*, Springer-Verlag, Berlin, Germany, 2004.