International Journal of Production Research, 2014
Vol. 52, No. 6, 1827–1842, http://dx.doi.org/10.1080/00207543.2013.849389

Taylor & Francis
Taylor & Francis Group

A hierarchical model for characterising spatial wafer variations

Lulu Bao[a], Kaibo Wang[a]* and Ran Jin[b]

[a]*Department of Industrial Engineering, Tsinghua University, Beijing, China;* [b]*Grado Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA*

(*Received 27 September 2012; accepted 9 September 2013*)

Silicon wafers are commonly used materials in the semiconductor manufacturing industry. Their geometric quality directly affects the production cost and yield. Therefore, improvement in the quality of wafers is critical for meeting the current competitive market needs. Conventional summary metrics such as total thickness variation, bow and warp can neither fully reflect the local variability within each wafer nor provide useful insight for root cause diagnosis and quality improvement. The advancement of sensing technology enables two-dimensional (2D) data mapping to characterise the geometric shapes of wafers, which provides more information than summary metrics. The objective of this research is to develop a statistical model to characterise the thickness variation of wafers based on 2D data maps. Specifically, the thickness variation of wafers is decomposed into macro-scale and micro-scale variations, which are modelled as a cubic curve and a first-order intrinsic Gaussian Markov random field, respectively. The models can successfully capture both the macro-scale mean trend and the micro-scale local variation, with important engineering implications for process monitoring, fault diagnosis and run-to-run control. A practical case study from a wafer manufacturing process is performed to show the effectiveness of the proposed methodology.

Keywords: Gaussian Markov random field; hierarchical model; spatial data

1. Introduction

Silicon wafers are widely used as the substrate in a variety of high-tech industries and applications, including integrated circuits (ICs), sensor devices, micro-electro mechanical systems, optoelectronic components and solar cells. Because small-scale devices and structures will be fabricated on the wafer substrate, the wafer must be smooth and flat on both the macro-and micro-scales (O'Mara, Herring, and Hunt 1990; Shen, Pei, and Fisher 2006). In other words, smaller surface variations result in better wafer quality. Large surface variations are likely to result in defects such as poor thickness uniformity, cracks and breakage that may further lead to defects in the final products. Additionally, poor surface quality may limit the enlargement of the wafer diameter, which can dramatically reduce the cost of IC fabrication by placing more chips on a wafer and thereby invoking the benefits of economies of scale (Quirk and Serda 2001; Orton 2009). Therefore, the ability to improve wafer quality is a critical objective in modern semiconductor manufacturing.

In general, the geometric quality metrics of a wafer are based on the requirements of smoothness, uniformity and flatness. The metrics that are commonly adopted by the semiconductor industry include centre thickness, total thickness variation (TTV), bow, warp, total indicator reading (TIR) and maximum focal plane deviation (Lin and Wang 2011; Li, Wang, and Yeh 2013). Shen, Pei, and Fisher (2006) used relative peak and valley displacements to measure the waviness of a wafer. Fan (2000) used within-wafer non-uniformity to measure the quality of the wafers produced in a chemical-mechanical planarisation process. Essentially, all of these metrics are summary statistics; each is calculated based on multiple measurements of a single wafer. These metrics are then used to represent the quality of the wafer. For example, TTV is defined as the difference between the maximal and minimal thicknesses over several thousand measurement points on a wafer. A large TTV value indicates poor thickness uniformity. As another example, bow is defined as the distance between the centre and a fitted reference plane of wafers (see O'Mara, Herring, and Hunt (2007) for the definition of these parameters) and is a measure of wafer flatness. However, such simple summary statistics do not adequately characterise wafers because they overlook the rich spatial information of the wafers.

Figure 1 shows the thickness of a wafer in a heat map. Different colours/grayscale represent different wafer thickness values. From this example, we can see that the map contains much information about the wafer; a macro-scale thickness trend and a micro-scale thickness similarity within adjacent areas are observed. A thorough analysis of these

*Corresponding author. Email: kbwang@tsinghua.edu.cn

© 2013 Taylor & Francis

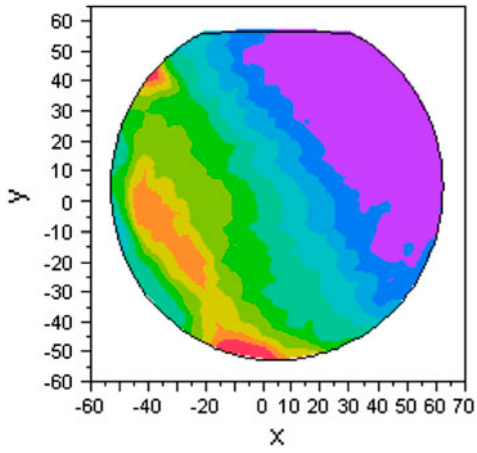


Figure 1. The heat map of a real wafer sample.

variation patterns will be given in Section 3. Based on these observations, we will develop a model that can capture the multi-scale variation patterns of wafers.

Statistical analysis based on the spatial patterns of wafers has a major advantage over analysis based on summary metrics due to the more varied information such analysis can yield. For example, the thickness of the wafer shown in Figure 1 changes sharply within a small edge area and is relatively flat throughout other large regions. Such a variation pattern cannot be explicitly reflected by TTV, as two wafers with the same TTV may exhibit totally different local thickness patterns. Instead, statistical analysis based on the rich spatial data can help link the variation pattern with the engineering process; thus, the result can provide hints for quality improvement initiatives.

The purpose of statistical analysis is construction of a quantitative model to characterise surface variation, which is important in many quality control applications. For example, unexpected process failures usually lead to changes in surface quality. Therefore, the model for wafer variation patterns can be used to characterise wafers and develop control charts for process monitoring. As another example, the manufacturing process variation can be reduced via run-to-run (R2R) process control; a model that can characterise and predict wafer quality is also needed for the design of the controller.

There are two types of quality inspection data that are available to represent the spatial patterns of wafers from a wafer manufacturing process: 2D continuous and 2D discrete data. These two types of data are usually collected after wafer preparation and fabrication processes, respectively. In the wafer manufacturing process, as shown in Figure 2, single-crystal silicon material is used to produce the wafer substrate through wafer preparation steps, such as slicing, lapping, chemical vapour deposition, polishing and testing. The wafer fabrication process then continues to create devices and structures on the top of a wafer substrate in multiple steps, such as photolithography, etching, doping, chemical mechanical planarisation and die testing. At the end of the wafer preparation processes, continuous data are used to measure the spatial variability of wafers in the data format shown in Figure 2(a). However, at the end of the wafer fabrication process, binary data are used to indicate the functionality of devices on the wafer in the data format shown in Figure 2(b).

Although geometrically similar, continuous data contain more information than binary data. Additionally, some of the defects shown on the binary map are caused by the spatial variation of wafer substrates. Therefore, we focus on the modelling and analysis of the continuous 2D wafer data map in this paper.

The remainder of the paper is organised as follows. Due to the unique spatial data structure obtained for wafer thickness analysis, a review and comparison of the existing methods for modelling surface variation is presented in Section 2. In Section 3, we analyse real wafer examples and provide insights for statistical modelling. A hierarchical model framework is then introduced for wafer variation modelling in Section 4, followed by a case study based on experimental data in Section 5. Finally, we draw conclusions and discuss the practical implications of the model in Section 6.

2. The state-of-the-art for modelling surface variation

In the literature, process monitoring, defect modelling or wafer clustering based on 2D maps have received much attention. As the two data types share similar spatial structures, the literature on the modelling of both discrete and continuous data is reviewed in this section.

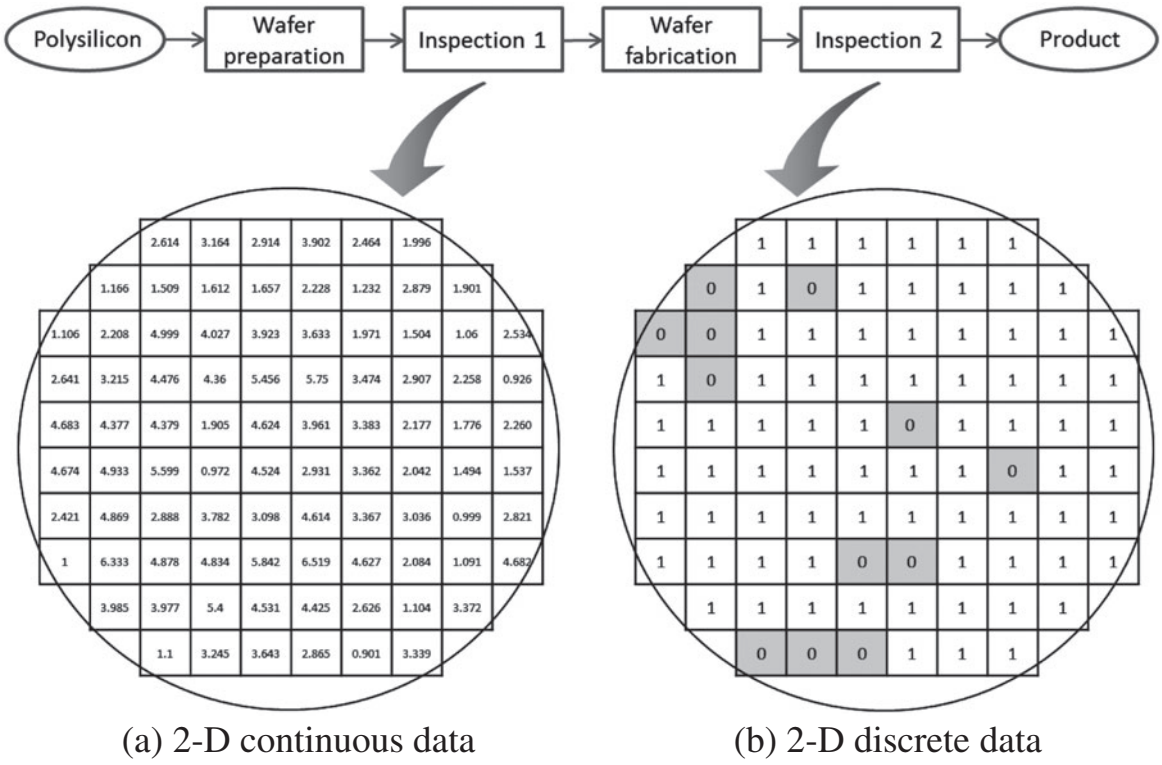


Figure 2. Two types of data collected from a wafer manufacturing process.

2.1 *Modelling 2D discrete data*

The discrete data acquired after the wafer fabrication process are mainly used for process monitoring and root cause diagnosis. Early research targeted the detection of spatial clustering defects on wafers. As Albin and Friedman (1989) noted, the Poisson distribution is generally assumed when defect data are encountered. However, defects tend to appear in clusters due to the spatial structure from the wafer manufacturing process. Albin and Friedman (1989) proposed the Neyman Type A distribution, which is more suitable for modelling the number of defects on individual wafers.

The modelling and analysis of binary 2D wafer maps are also important for diagnosing the faults in the wafer fabrication process. Hansen, Nair, and Friedman (1997) classified the defect patterns on wafers into two categories: random defects and clusters. A Markov random field is used to characterise the clustered defects. A similar approach was used by Hwang and Kuo (2007) in a model-based clustering strategy that treats the observed defects as the composition of both global defects resulting from random causes and local defects resulting from assignable causes. A spatial non-homogeneous Poisson process is used to model the global defects; a bivariate normal distribution or principal curve is then used to model the local defects. A model-based clustering algorithm is further applied to identify the characteristics of local defect clusters. Liu, Chen, and Lu (2002) proposed a neural network to extract patterns from the 2D wafer binary map. Wang, Wang and Lee (2006) also focused on the identification of spatial defect patterns on wafers. Hsu and Chien (2007) proposed a data-mining framework to identify failure patterns and classify wafers into different clusters based on the 2D binary wafer map.

Overall, the work on monitoring, modelling and analysing failure patterns based on binary 2D wafer maps is an effective approach for studying the fabrication process. The knowledge gained from these analyses is valuable for quality improvement and has yielded enhancement in practice. However, these approaches are used for discrete data rather than continuous data.

2.2 *Modelling 2D continuous data*

In wafer manufacturing, the variations or defects of the wafer substrates will be transmitted to the fabrication process from the preparation process. The continuous 2D wafer map contains rich information about wafer quality and therefore should be further studied. There are a few methods for the modelling of continuous data.

During image processing, data usually exhibit a similar 2D structure. One of the most facile ways to model 2D continuous data is to build a polynomial response surface model, in which the surface is estimated as the sum of approximated polynomial functions and random errors. Taam (1998) used a second-order polynomial model to characterise a 2D continuous surface. However, spatial correlation was not considered in this approach. In computer graphics, approaches such as spline (Lee, Wolberg, and Shin 1997) and wavelet (Valette and Prost 2004) are used as interpolation methods for image reconstruction. However, the characteristics of the wafer and the manufacturing process are ignored in these models (Jin, Chang, and Shi 2012).

There is also a rich body of literature in geostatistics for modelling continuous data with spatial structures. These models can achieve better performance in the inference, prediction and estimation of surface variability by considering spatial dependence (Haran 2011). The most commonly used modelling techniques include the Kriging and random field models.

Kriging is an interpolation method that predicts a value at an unobserved location from its neighbouring observations. This method provides more accurate global approximations for surfaces than the traditional response surface methods (Simpson et al. 2001). Although the mean part of the Kriging method could be a constant, it can perform as well as the second-order response surface model (Simpson et al. 1998). Kriging also performs better than the cubic spline methods (Voltz and Webster 1990).

Kriging has been widely used in many areas, such as hydrogeology (Tonkin and Larson 2002; Chiles and Delfiner 2012) and mining (Journel and Huijbregts 1978; Richmond 2003). Ordinary Kriging is usually used to model deterministic processes, such as the input–output relationship in computer experiments (Qian and Wu 2008). In a stochastic simulation, the nugget effect is usually added, which changes the model to a stochastic Kriging model to account for the random noises present in a heteroscedastic process (Yin, Ng, and Ng 2011). Although Kriging has been used in many cases, the use of the technique still suffers from issues, such as the selection of an appropriate model structure in the regression mean part (Martin and Simpson 2005) or the specification of the correlation matrix structure.

Besides Kriging, Gaussian Markov random field (GMRF) models can also be used for spatially correlated 2D data. GMRF models can be considered as a special case of either the Gaussian random field or Markov random field models. The Gaussian random field model assumes that the variables follow a multivariate Gaussian distribution, while the Markov random field model assumes the Markov property among the sites. The GMRF model further applies the Markov constraint to the correlation structure of a multivariate Gaussian distribution: the conditional probability of one site, given all its *immediate* neighbours, is equal to the conditional probability of the site, given *all* other sites. Because of the Markovian property, the GMRF models the surface with comparable performance to a full Gaussian model but significantly reduces computational costs (Hrafnkelsson and Cressie 2003; Hartman 2006). Moreover, the GMRF model is also flexible in handling data and can be used to investigate irregular 2D map shapes (Lindgren and Rue 2008) and spatial-temporal cases (Allcroft and Glasbey 2003). If the precision matrix is not full rank, then the model reduces to an intrinsic Gaussian Markov random field (IGMRF) model. Huang (2010) used an IGMRF model to characterise the local variability of the length of nanowires.

Because the specification of the dependency of the Markov random field is realised through a conditional probability, the Markov Chain Monte Carlo (MCMC) method is usually used to estimate the GMRF models. Based on the 2D continuous data of wafers that we obtained, the spatial correlation structure satisfies the assumptions of the IGMRF model. Therefore, we use an IGMRF model to characterise wafer surface variation in this work.

3. Analysis of variation patterns on the wafer surface

The data used in this paper are collected from the wafer preparation process. To construct an appropriate model for the wafers, we must obtain engineering knowledge for the variation patterns. Therefore, we first introduce the wafer manufacturing process from an engineering perspective because the variations are directly introduced by the process. We then analyse and interpret the macro-and micro-scale variation patterns from experimental data.

3.1 *The wafer preparation process*

In the wafer preparation process, a lapping process is a critical step that improves wafer uniformity. The primary quality metrics are mainly determined by this process. The continuous data are also collected right after the lapping process. Therefore, we first give a brief introduction of this process.

A schematic illustration of the lapping equipment is shown in Figure 3. To start the lapping, wafers are first placed on the lower plate, and the upper plate presses against the lower plate when pressure is applied. During the lapping process, the upper and lower plates rotate in opposite directions. Abrasive slurry is fed into the lapping process; the

abrasive particles in the slurry are used to remove the wafer material. The summary metrics introduced in Section 1 are used to measure the wafer quality. Due to the complexity of the lapping mechanism, different points on a wafer are lapped following different moving paths. Therefore, actual lapped wafers are not ideally flat or smooth. More information about the lapping process can be found in Marinescu, Uhlmann, and Doi (2006), Lin and Wang (2012) and Wang and Han (2013).

From engineering knowledge, the wafer quality is directly affected by the wear of the lapping plates, the density and efficiency of the slurry used in production and the controllable process variables, such as time, pressure and rotation speed. Because of similar manufacturing conditions are applied across the entire wafer, a strong spatial correlation of the thickness measurements within the wafer is observed. The variation patterns will be analysed by incorporating such engineering knowledge into the statistical models.

3.2 *The macro-scale wafer variation pattern*

If a wafer exhibits large surface variation, then it may have a high defect ratio in the downstream wafer fabrication step. Figure 4 shows the thickness heat maps of two wafers. Both wafers are flat and thick on one side but rough and thin on the other side, which is the macro-scale variation pattern that we identified from the real samples. Engineering knowledge further indicates that such macro-scale variation patterns can be attributed to the uneven distribution of the slurry and forces in the slicing and lapping processes, where the removal rate is not uniform (Zhao et al. 2011). Additionally, compared to the reference plane (the horizontal flat edge on the top), the variation patterns show different changing directions. The wafer in Figure 4(a) changes in the horizontal direction, while the wafer in Figure 4(b) changes in the vertical direction. The rotation effect of the wafer is one challenge in its modelling.

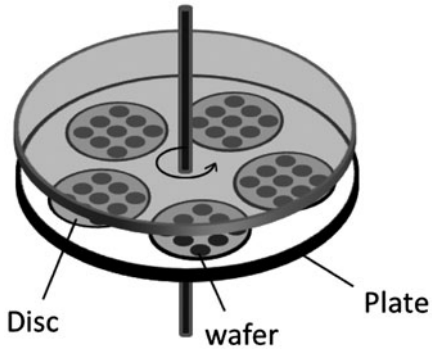


Figure 3. A schematic illustration of the lapping equipment.

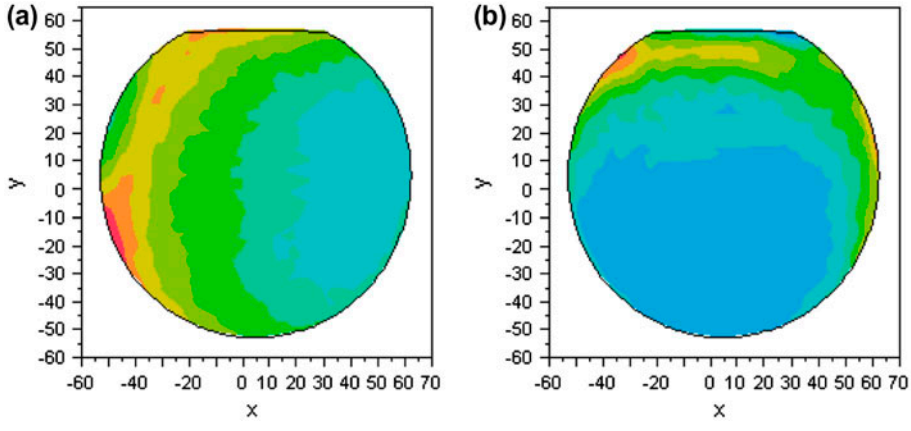


Figure 4. Wafer thickness heat maps with similar patterns but different alignment angles.

3.3 *The spatial dependency and micro-scale wafer variation*

An important and unique feature of the 2D wafer surface data is the spatial dependency among different sites. The quality measurement of one site is correlated with others, and such correlation is higher when the sites are closer to each other. In this section, we investigate the spatial dependency based on a lattice of 86 sampling points taken from each wafer. The selection of sampling sites is discussed by Jin, Chang, and Shi (2012), and we do not focus on this problem. The distribution and positional indices of these sampling points are shown in Figure 5. The centre area is missing from the measurement.

The 'neighbours' of a site are defined as the sites that surround it. Figure 5 also shows the neighbours (shaded) of sites #1, #33 and #75. Because the readings near the centre area are missing, site #33 only has neighbours to its right. In building the statistical model in the next section, the neighbourhood of relationships is designated by a set, which shows the flexibility of the model in handling irregular data structures. Some literature has studied this problem and proved that this neighbourhood structure is both simple and effective (Xu and Huang 2012). Considering that the neighbours sharing the same border (side neighbours) and diagonal neighbours have different distances from the target site, we can assign different weights to distinguish them. The selection of different weighting schemes will be addressed in the case study.

The dependency among the sites is indicated using their partial correlation, which measures the degree of association between two random variables with the effect of all other sites removed. For example, the partial correlation coefficients of site #23 on the wafer are shown in Figure 6. Figure 6(a) shows the partial correlation between site #23 and other sites, while Figure 6(b) shows the ten sites with the largest partial correlations with site #23. From these two figures, we observe that a spatial dependency does exist. Additionally, the immediate neighbours of the site account have the highest dependency, which indicates the Markovian property in spatial dependency: each site only depends on the values of its neighbours. This observation makes the GMRF model applicable for the wafer surface.

The site dependency only shows the micro-scale variation of the wafer. The primary geometric metrics, such as the TTV and TIR, are dominated by the combined effect of the macro-scale and micro-scale variations. Therefore, we propose a hierarchical IGMRF model to characterise both scales of variation.

4. Modelling wafer surface variation

As the wafer exhibits spatial dependency with the Markovian property, we model the wafer surface as a GMRF and adopt a hierarchical approach for the modelling process. In general, hierarchical modelling can be used to imply a complicated distribution without losing the briefness of a conditional distribution (Ntzoufras 2011). Additionally, the hierarchical structure makes Bayesian analysis more robust, the interpretation and calculation simpler, and the approximation easier (Robert 2007). In this section, we first introduce the general IGMRF model and then extend the model to fit the continuous 2D wafer map.

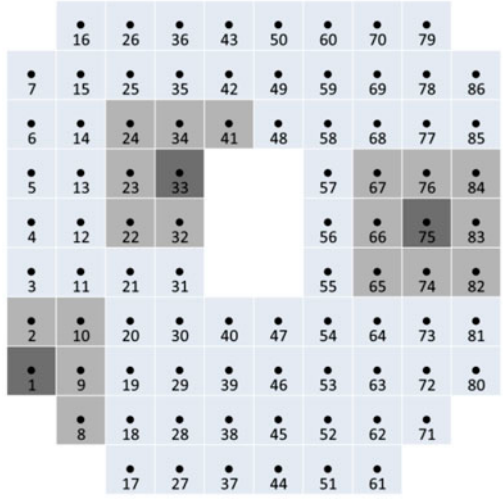
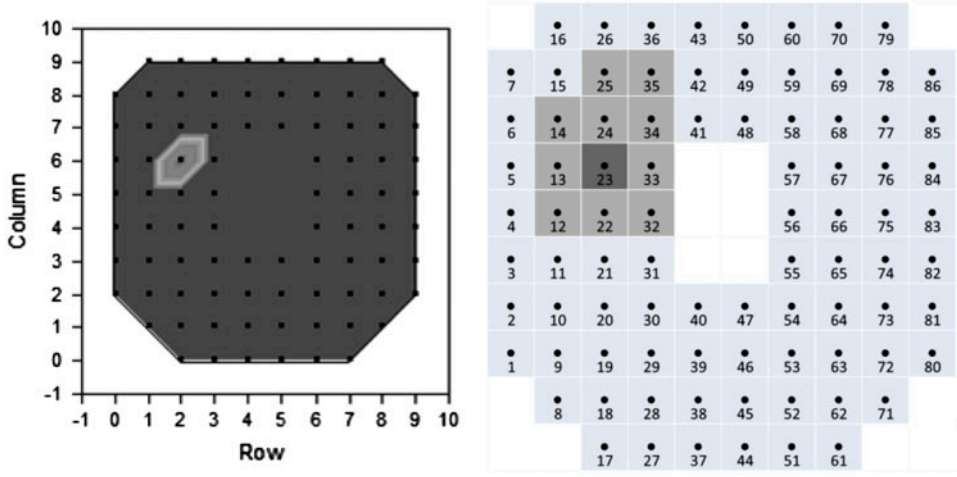


Figure 5. The definition of the neighbours of sites #1, #33 and #75.



(a) A partial correlation between site #23 and the other sites. The brighter color represents higher partial correlation.

(b) The ten sites with the largest partial correlations with site #23, marked with dark shading.

Figure 6. The partial correlation analysis of site #23 on a wafer.

4.1 *The hierarchical IGMRF model used to characterise wafer surface variation*

A hierarchical IGMRF model is built in multiple stages. In the first stage, a distribution assumption is specified for the observations. In the second stage, a prior model is assigned to the unknown parameters; the IGMRF is used to represent the unique spatial correlations illustrated above. In the third stage, a prior distribution is assigned to the unknown parameters of the IGMRF (Rue and Held 2005). Following this process, the form of a general hierarchical IGMRF model is constructed as follows:

$$\begin{aligned} &\textit{Stage } 1: && z_s \sim N(\mu_s, \sigma_\varepsilon^2) \\ &\textit{Stage } 2: && \mu_s = f(x_s, y_s, \theta) + w_s \\ &\textit{Stage } 3: && w_s \sim car(\kappa) \\ & && \text{other priors} \end{aligned} \tag{1}$$

We apply this three-stage model to wafer surface variation characterisation and explain the equations in greater detail.

Stage 1

Let $\mathbf{s} = (1, 2, \ldots, s, \ldots, n)^T$ be the sampling sites and z_s the thickness value at site s with coordinate (x_s, y_s). At stage 1, we assume that the wafer thickness at a site follows a normal distribution. When the mean pattern is removed, the residual varies randomly. The experimental data does not violate this assumption.

Stage 2

The mean pattern of the thickness distribution is assumed to follow the prior model $\mu_s = f(x_s, y_s, \theta) + w_s$, where $f(x_s, y_s, \theta)$ represents the macro-scale trend of site s, x_s and y_s are the coordinate locations of site s, and w_s is the micro-scale variation, modelled as an IGMRF. θ is a vector of model parameters to be explained later.

Although the thicknesses follow a generally decreasing trend from one side to the other, each wafer has a distinct direction for this trend, as shown in Figure 4. The random rotation direction is caused by both the initial placement and random self-rotation of the wafer during lapping. Without modelling the rotation effect, a data-driven modelling approach will not be able to remove this effect and characterise the micro-scale variation by the model parameters.

To model the rotation effect, we firstly define a new reference line along the centre of the wafer, as shown in Figure 7(a). A site $A(x, y)$ is then projected onto the reference line. The distance between the projected point and the centre of the wafer, O, is calculated as $d = (x \cdot \cos\theta + y \cdot \sin\theta)$, where θ is the angle of the reference line. Along the reference line, the thickness shows a trend similar to the curve in Figure 7(b). In the figure, the horizontal axis is the projected distance, and the vertical axis is the average thickness at that point. Further data analysis shows that the

relationship between the thickness and the projected distance can be well estimated by a third-order polynomial function. Therefore, we rewrite the macro-scale mean pattern $\mu(s)$ as

$$\textit{Stage 2: } \mu_s = \alpha + \beta \cdot (x_s \cdot \cos\theta + y_s \cdot \sin\theta - \delta)^3 + w_s. \tag{2}$$

The parameter vector is $\theta = (\alpha, \beta, \theta, \delta)$, where α is an intercept item, which is the baseline of the thicknesses; δ is the location shift of the curve; and β represents the magnitude of the curve. These parameters can show the flatness of the wafer and the decaying speed of the curve.

In Equation (2), w_s is modelled as a first-order IGMRF model. As the thickness of one site is mainly affected by its immediate neighbours, a first-order structure sufficiently models the spatial variability caused by the dependence of the sites. The difference in w between any two adjacent sites i and j is assumed to follow a normal distribution:

$$w_i - w_j \sim N(0, \kappa^{-1}) \tag{3}$$

where κ is the precision parameter of the random field model.

Let $\mathbf{w} = (w_1, w_2, \ldots, w_n)^T$. Based on the above assumption, it is shown in the Appendix 1 that the joint probability density of $\mathbf{w}$ is given by

$$f(\mathbf{w}|\kappa) \propto \kappa^{(n-1)/2} \exp\left[-\frac{\kappa}{2}\sum_{i\sim j}(w_i - w_j)^2\right] = \kappa^{(n-1)/2} \exp\left[-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}\right] \tag{4}$$

where $i \sim j$ means that sites i and j are neighbours; $\mathbf{Q}$ is the precision matrix with rank $n-1$ and $\mathbf{Q} \times \mathbf{1} = \mathbf{0}$. In vector $\mathbf{1}$, all elements equal 1, meaning that the sum of the row elements of $\mathbf{Q}$ equals 0. $\mathbf{Q}$ has an element in the ith row and jth column given by

$$Q_{ij} = \kappa \begin{cases} n_i, & \text{if } i = j \\ -1, & \text{if } i \sim j \\ 0, & \text{otherwise} \end{cases}.$$

The precision matrix $\mathbf{Q}$ defines the neighbourhood relationship among the sites, which is determined by the structure of the neighbourhood relationship that we specified. n_i denotes the number of neighbours of site i.

We can incorporate the difference information between diagonal and side neighbours into the statistical model by assigning different weights to them; the weight is positive and symmetric. Denoting c_{ij} as the weight of a pair of adjacent sites i and j, we have

$$c_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ share a common side} \\ c, & \text{if } i \text{ and } j \text{ are diagonal neighbours, } 0 \le c \le 1 \end{cases}.$$

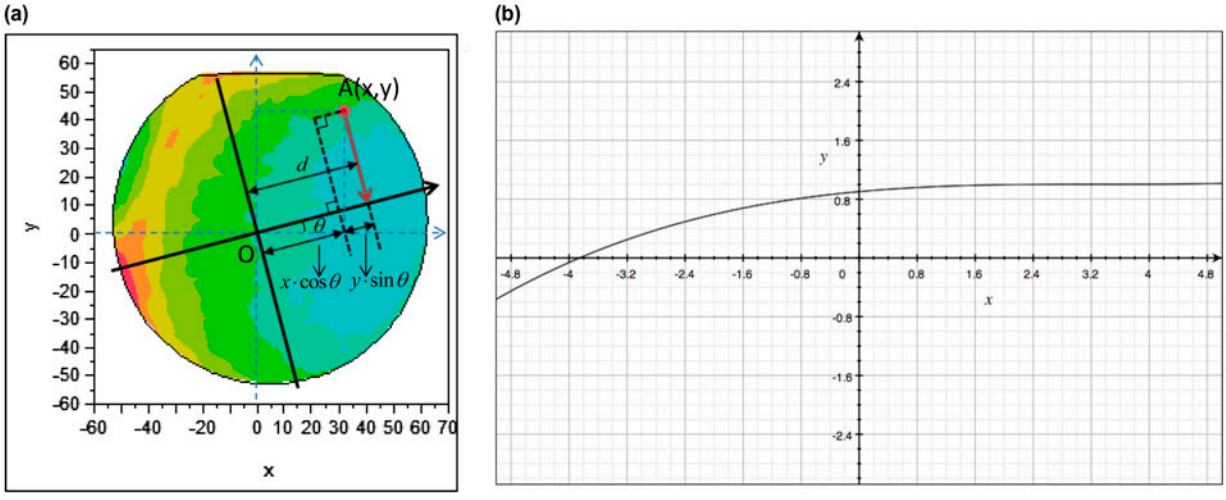


Figure 7. Modelling the rotation effect. (a) The projection of points onto the reference line. (b) The decrease pattern along the reference line.

Then the joint distribution of w_i is

$$f(\mathbf{w}|\kappa) \propto \kappa^{(n-1)/2} \exp\left[-\frac{\kappa}{2}\sum_{i\sim j} c_{ij}(w_i - w_j)^2\right]. \tag{5}$$

The precision matrix is

$$Q_{ij} = \kappa R_{ij} = \kappa \begin{cases} \sum_{k:k\sim i} c_{ik}, & \text{if } i = j \\ -c_{ij}, & \text{if } i \sim j \\ 0, & \text{otherwise} \end{cases}.$$

It can be shown that the full conditional distribution of w_i will be

$$w_i|\mathbf{w}_{-i}, \kappa \sim N\left(\frac{\sum_{j:j\sim i} c_{ij} w_{ij}}{\sum_{j:j\sim i} c_{ij}}, \frac{1}{\kappa \sum_{j:j\sim i} c_{ij}}\right), \quad i = 1, 2, \ldots, n \tag{6}$$

where $-i$ denotes all of the sites except site i. Therefore, the conditional mean of w_i is simply a weighted average of its neighbours. The details of this derivation can be found in the Appendix.

In the wafer example, the precision matrix is 86 × 86. This matrix is sparse due to the conditional independence between the corresponding site pairs, which can help improve the computation speed.

Stage 3

The above IGMRF model can be used to characterise any surface variation. To fit the model to the wafer examples, we need to specify all of the priors in the model. Such an assignment should be made upon the analysis of the experimental wafer sample data.

The priors of all of the parameters in this model are specified as follows:

$$\begin{aligned} \textit{Stage } 3: \quad & w_s \sim car(\kappa), \kappa \sim \Gamma(0.05, 0.00005), \frac{1}{\sigma_\varepsilon^2} \sim \Gamma(0.01, 0.01) \\ & a \sim uniform(-\infty, \infty), \beta \sim N(0.005, 1), \theta \sim uniform(0, 6.2831852), \delta \sim N(5, 2) \end{aligned}$$

In the model, *car* represents an approach to specifying an IGMRF through full conditionals, and the model is also called conditional autoregressions (CARs). The definition $w_s \sim car(\kappa)$ in Stage 3 indicates that w_s is an IGMRF with precision parameter κ. The CAR assumption makes the precision matrix sparse, as each site only depends on its immediate neighbours. Based on the data analysis in Figure 6, we confirm that such an assumption is suitable for the wafer example.

The prior of precision parameter κ is $\gamma(0.5, 0.0005)$, which is the suggested value for the precision parameter of the spatial random effects in a CAR model given by Thomas et al. (2004). For the parameter $1/\sigma_\varepsilon^2$, the conjugate prior, a Gamma distribution (or more specifically, $\gamma(0.01, 0.01)$) was suggested by Ntzoufras (2011). The rotation angle θ is assumed to be uniformly distributed, varying from 0 to 2π. A non-informative uniform distribution is assigned to α. After an initial trial based on the experimental data, we specify the normal distribution as priors for β and δ to make the algorithm converge faster. The mean of the distribution is extracted from the historical data.

4.2 *Model parameter estimation*

In this paper, we use MCMC to infer the parameters. MCMC has the advantages in simplicity, convenience, generality (Ntzoufras 2011), and is therefore widely used in Bayesian inference for highly complicated models. Based on the above-hierarchical model, the implementation of the estimation algorithm could be developed using *WinBUGS* without much difficulty.

It should be noted that the maximum likelihood estimation (MLE) could be modified and used to obtain parameter estimates in this model. One key challenge in using the MLE is that the local variation $\mathbf{w}$ follows an IGMRF model, which has an improper precision matrix and makes an explicit expression of the distribution function difficult. This is one reason that the Bayesian method is more widely used in practice (Huang 2010; Hughes and Haran 2013). To proceed with the MLE, one way is to reduce the dimension of the variables to produce a non-singular covariance matrix through the linear transformation of the data. The linear transformation will produce $n - 1$ linearly independent variables with a full-rank covariance matrix. We can then apply MLE to the new variables using the transformed data. Alternatively, one may use the Moore–Penrose inverse of the improper precision matrix and add it to the covariance matrix of

the white noise, which is a diagonal matrix. Once the desired non-singular covariance matrix is obtained, the MLE could be used in the conventional way. However, we noticed from numerical study that the likelihood function near the optimum is quite flat, which may result in a large variance in parameter estimates. Therefore, in following studies, we still rely on MCMC to estimate parameters in the models.

5. Case study

We demonstrate the proposed hierarchical Bayesian model for experimental wafer data. For each wafer, 86 thickness readings are measured at the locations shown in Figure 5. We first show the fitted model to a real wafer example, then compare the proposed model with other possible models using multiple wafer examples.

5.1 *Real sample model fitting*

Table 1 shows the estimates of the model parameters based on a sample wafer. The mean and quantiles are calculated based on 40,000 iterations in the MCMC. The trace and posterior distributions of these parameters are shown in Figure 8 for evaluation purposes, which show that the MCMC estimation procedure converges. In the example, we use a simple neighbourhood structure that does not distinguish the side neighbours and diagonal neighbours, i.e. $c = 1$.

Figure 9(a) shows the heat map of the actual thickness values, and Figure 9(b) shows that of the predicted values. The model successfully captures the macro-scale variation. The macro-scale thickness trend is characterised by the cubic expression $\alpha + \beta(x_s \cdot \cos\theta + y_s \cdot \sin\theta - \delta)^3$. To validate that this cubic curve fits the data well, we now compare two curves. One is a predictive curve that comes from our parameter estimation result,

$$\mu_s = 346.0423 + 0.0003(x_s \cdot \cos(0.5225) + y_s \cdot \sin(0.5225) - 4.7536)^3$$

the other is the experimental thickness values after removing the micro-scale variation.

Figure 10 shows the predicted and measured macro-scale thickness values. We use the projected distance, $x_s \cdot \cos\theta + y_s \cdot \sin\theta$, as the horizontal axis, so that we can see the mean pattern more clearly. The macro-scale thickness variation trend fits the cubic curve quite well after projection. Therefore, our proposed model is appropriate.

5.2 *Performance comparison*

Next, we compare the model with other candidate models fitted to the wafer data. The model performance is measured using the sum of the squared errors (SSEs), which is calculated by:

$$\text{SSE} = \sum_{s=1}^{n} (y_s - \hat{y}_s)^2 \tag{7}$$

We fit five model alternatives and compare their performance.

Model 1: $\mu + \varepsilon_s$, which uses the mean of the site thicknesses to characterise the entire surface. This model is used because the mean thickness is considered a primary quality variable in current industrial practice. A comparison with this model can show the loss of information incurred if the thickness pattern and spatial correlation are not considered.

Table 1. The estimated parameters based on experimental data.

Parameter	Mean	SD	2.50%	Median	97.50%
α	346.0423	0.0189	346.0068	346.0417	346.0812
β	0.0003	0.0001	0.0002	0.0003	0.0006
θ	0.5225	0.1837	0.1522	0.5254	0.8752
δ	4.7536	0.7300	3.3236	4.7867	6.1636
κ	74.3374	15.5785	48.2565	72.7339	109.5906
$\sigma = \sqrt{1/\kappa}$	0.1179	0.0123	0.0955	0.1173	0.1440
$1/\sigma_\varepsilon^2$	762.4495	209.3075	426.1014	736.8439	1239.3650
σ_ε	0.0372	0.0051	0.0284	0.0368	0.0484

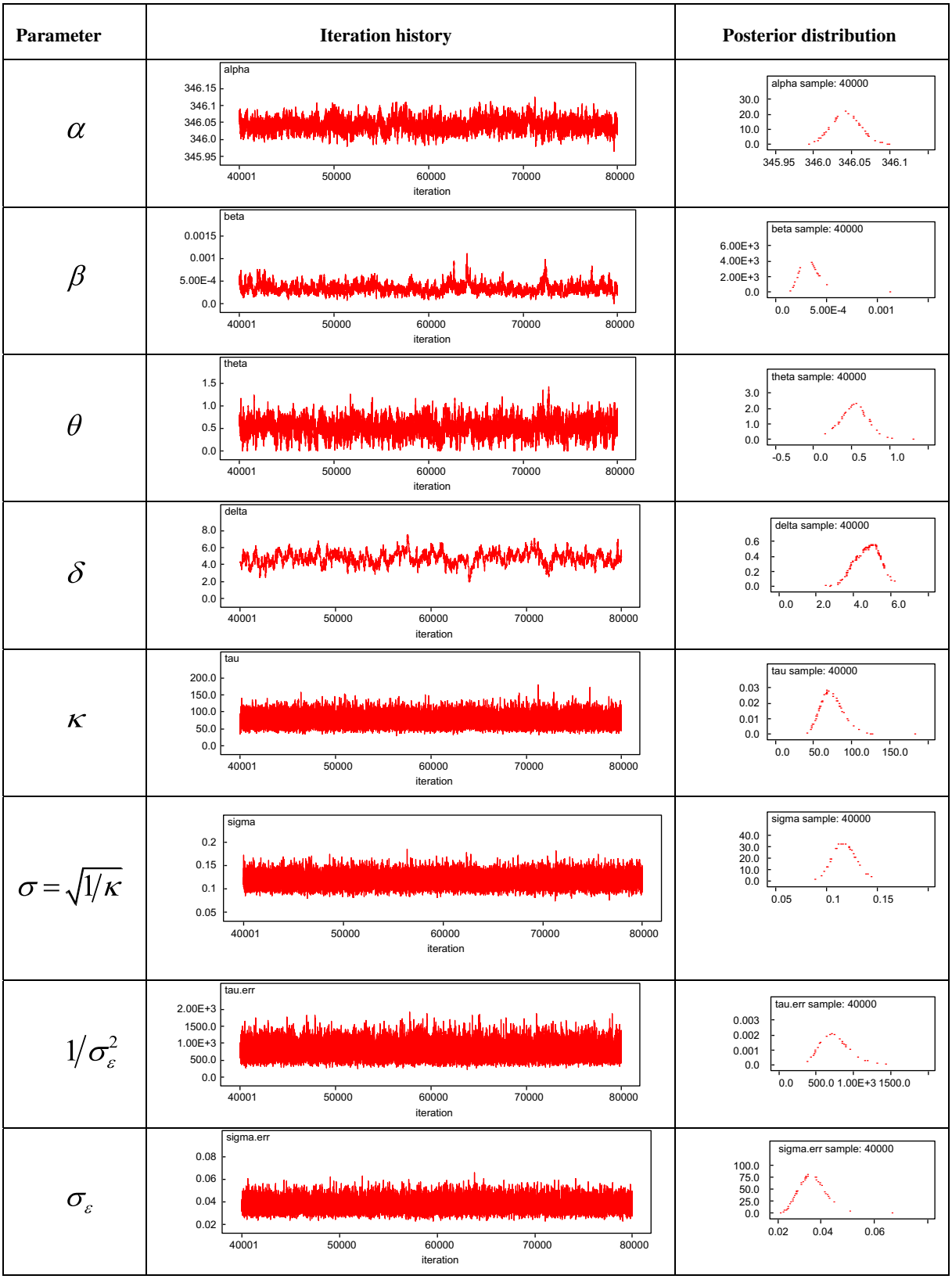
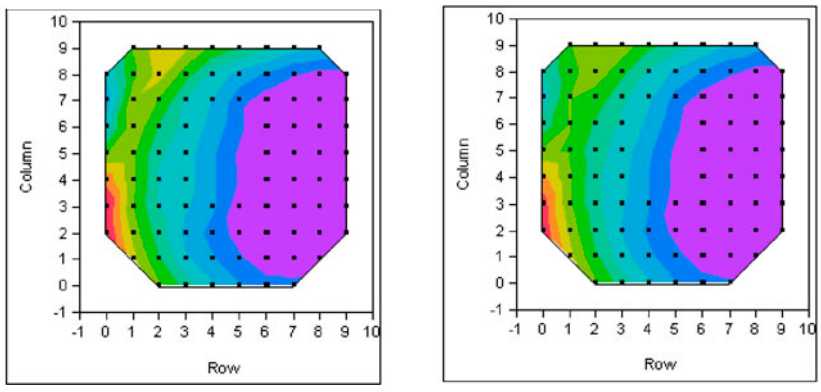


Figure 8. MCMC trajectories and distributions of the estimated parameters.



(a) Heat map of actual thickness (b) Heat map of predicted thickness

Figure 9. A comparison between the experimental and estimated thicknesses.

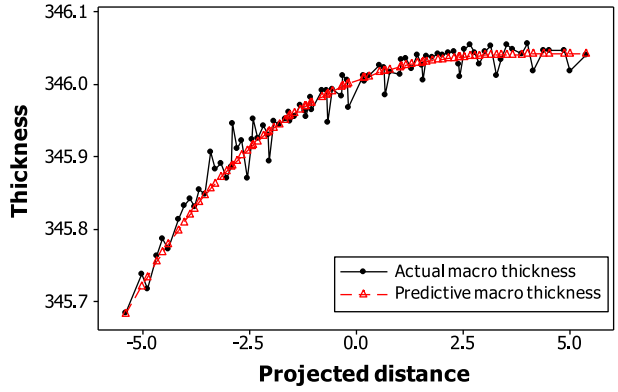


Figure 10. The predicted and measured macro-scale variation patterns.

Model 2: $\mu + a \cdot x_s + b \cdot y_s + \varepsilon_s$, which uses a linear regression model to discover the relationship of thickness and location. In Kriging analysis, the linear effect of location is commonly considered. Therefore, a comparison with this model can review the importance of considering the thickness trend pattern, which practitioners have tended to ignore.
Model 3: $\alpha + \beta(x_s \cdot \cos\theta + y_s \cdot \sin\theta - \delta)^3 + \varepsilon_s$, which uses a polynomial function to characterise the surface, without considering the spatial correlations among the sites. This model is similar to the one used by Taam (1998), which proposed a second-order polynomial model to fit the wafer surface data. However, the spatial correlations are ignored in that model.
Model 4: $\alpha + \beta(x_s \cdot \cos\theta + y_s \cdot \sin\theta - \delta)^3 + w_s + \varepsilon_s$, which uses a combination of the polynomial function and the IGMRF. This model does not distinguish between side and diagonal neighbours, which means $c = 1$.
Model 5: we use the same model as Model 4, but with a different weight for the diagonal neighbours, $c = 1/\sqrt{2}$.

The models are applied to eight real wafer samples. Table 2 shows the *SSE* of each model. It can be seen that Models 1 and 2 are generally worse than Models 3, 4 and 5 because they overlook the rotation effect. Therefore, it is important to estimate the rotation angle in macro-scale variation analysis. One exception is that Model 2 performs better than model 3 for Wafer 1. A further investigation shows that Wafer 1 has an obvious upward tilt near the near area, which is against the main trend the cubic function Model 3 tries to fit.

Comparing Models 3 and 4, the proposed Model 4 shows much better prediction performance, indicating that consideration of the spatial correlation among sites further improves the prediction accuracy. The proposed hierarchical Bayesian IGMRF, Models 4 and 5, perform the best among different model alternatives.

Between Models 4 and 5, the weighting scheme using $c = 1/\sqrt{2}$ provides more accurate results. The main effect of the weighting coefficient, c, is on the precision matrix, or equivalently, the correlation structure of different sites. In the literature, Kriging technique, in which the Gaussian Process (GP) model is widely adopted, has been used to model

Table 2. SSE comparison of models fit to experimental data.

Wafer ID	Model 1	Model 2	Model 3	Model 4 ($c = 1$)	Model 5 ($c = 1/\sqrt{2}$)
1	1.9175	0.4960	0.6332	0.0222	0.0214
2	3.3729	0.4940	0.3142	0.0408	0.0369
3	3.4492	0.5985	0.4971	0.0313	0.0293
4	0.4520	0.2556	0.1758	0.0240	0.0225
5	0.5500	0.3824	0.3655	0.0433	0.0390
6	0.3666	0.3017	0.2901	0.0272	0.0253
7	3.5348	1.5008	0.8378	0.0193	0.0178
8	0.7534	0.2777	0.2590	0.0214	0.0206

spatial correlations (see, e.g. Li and Sudjianto (2005), Qian and Wu (2008) and Zhao et al. (2011)). In Kriging modelling, different variograms may be assigned, which directly dominate the covariance matrix of the spatially correlated variables. That is, the choice of the variogram in Kriging modelling has a similar effect as the selection of the weighting scheme in GMRF modelling. In practice, an empirical variogram could be calculated from a historical data-set; a theoretical variogram is then selected from several limited alternatives to fit with the empirical one. That is, engineering data plays an important role in selecting the variogram in Kriging modelling. Similarly, in GMRF modelling, we suggest that different weighting schemes should be tested and the one with the best performance be retained.

6. Conclusions and discussion

Wafers are important substrates in semiconductor manufacturing. However, current industrial practice mainly uses summary metrics to characterise wafer quality and variation reduction. In this paper, we proposed a three-stage hierarchical model to characterise 2D continuous wafer data. In the model, the wafer thickness variation is decomposed into macro-scale and micro-scale variation components. The macro-scale variation is characterised by a third-order polynomial function that represents the rotation effects of wafers in the preparation process. The micro-scale variation is modelled as a first-order IGMRF, characterising the spatial correlation between the sites with simple neighbourhood relationships. A comparison with other potential model formulations is performed in a case study based on experimental wafer data, which indicates that the proposed model not only better the models of the wafer surface variation but also provides physically meaningful interpretations for macro- and micro-scale variations.

Modelling wafer surface variation is an essential step in analysing and improving wafer quality during the preparation process. The developed model not only helps to characterise quality variation but also provides insights to the variation patterns. Such variation patterns contain valuable information for the engineering discipline regarding root cause diagnosis and process improvement. Therefore, the proposed statistical model has potential for many quality control applications. By linking the model with the engineering process, we can better understand the variation source and develop advanced methods to improve product quality. Control charts and R2R control strategy based on the spatial data and model will be developed for future research.

Specifically, the two parts of the proposed model in Equation (1) have important engineering implications. First, the rotated cubic curve in the mean part shows the uneven surface variation pattern. From the pattern analysis above, we identified this trend in the thickness, which has not received sufficient attention in the past. Based on the fitted model and the significance of the parameters, we characterise such rotation effects. This non-uniformity is possibly attributed to the flatness of the lapping plate. In the lapping process, the upper lapping plate is physically held by a corner joint. Therefore, the upper plate has the freedom to tilt in rotation. Such a tilt could introduce an uneven lapping force to the wafer; thus, this process could produce uneven thickness. Additionally, monitoring the fitted parameters with a control chart is helpful for the identification of sudden or gradual failures in the lapping process, which is important to the prevention of serious quality losses. Second, the random field and noise terms in the model show the spatial correlation and micro-scale variation. In the lapping process, the roughness of the surface is mainly dominated by the lapping slurry and controllable factors such as pressure and rotation speed. In practice, changes in the lapping slurry, including its density and particle diameter distributions, are difficult to measure. However, such changes would lead to deterioration of the wafer surface quality. Therefore, we can identify such process changes by checking the fitted parameters and further develop condition-based replacement strategies for the slurry with the hierarchical model. The use of the proposed model, which is a combination of regression and IGMRF models, is not limited to the wafer example. This model could also be extended to similar data formats with spatial dependency.

In this work, a cubic curve along the reference line is fitted to explain the macro-scale variation pattern. Such parametric models have good engineering interpretability and can be improved by incorporating engineering knowledge. They can also be easily used for statistical monitoring or process control. However, parametric models are less flexible in characterising irregular patterns. Comparatively, a non-parametric method (Raja, Muralikrishnan, and Fu 2002) may have better performance. Some popular models, such as spline (Gardner et al. 1997; Wang 1998; Tait et al. 2006), and kernel method (Qiu, Zou, and Wang 2010) have been seen used for fitting 1-dimensional profiles. Such non-parametric models are usually less constrained and have strong robustness and adaptability. Therefore, fitting the wafer surface using non-parametric models is also a topic that deserves further research efforts.

Acknowledgements

We greatly thank the editor and the anonymous referees for their valuable comments, which have helped improve this work greatly. Dr Wang's work was supported by the National Natural Science Foundation of China under Grant No. 71072012 and Tsinghua University Initiative Scientific Research Program.

References

Albin, S., and D. Friedman. 1989. "The Impact of Clustered Defect Distributions in IC Fabrication." *Management Science* 35 (9): 1066–1078.

Allcroft, D. J., and C. A. Glasbey. 2003. "A Latent Gaussian Markov Random Field Model for Spatiotemporal Rainfall Disaggregation." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52 (4): 487–498.

Besag, J., and C. Kooperberg. 1995. "On Conditional and Intrinsic Autoregressions." *Biometrika* 82 (4): 733–746.

Chiles, J. P., and P. Delfiner. 2012. *Geostatistics: Modeling Spatial Uncertainty*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

Fan, S. K. S. 2000. "Quality Improvement of Chemical–mechanical Wafer Planarization Process in Semiconductor Manufacturing Using a Combined Generalized Linear Modelling-non-linear Programming Approach." *International Journal of Production Research* 38 (13): 3011–3029.

Gardner, M. M., J.-C. Lu, R. S. Gyurcsik, J. J. Wortman, B. E. Hornung, H. H. Heinisch, E. A. Rying, S. Rao, J. C. Davis, and P. K. Mozumder. 1997. "Equipment Fault Detection Using Spatial Signatures." *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C* 20 (4): 295–304.

Hansen, M. H., V. N. Nair, and D. J. Friedman. 1997. "Monitoring Wafer Map Data from Integrated Circuit Fabrication Processes for Spatially Clustered Defects." *Technometrics* 39 (3): 241–253.

Haran, M. 2011. "Gaussian Random Field Models for Spatial Data." In *Handbook of Markov Chain Monte Carlo*, edited by S. P. Brooks, A. Gelman, G. L. Jones and X. L. Meng, 449–478. Boca Raton, FL: Chapman & Hall/CRC Press.

Hartman, L. W. 2006. "Bayesian Modelling of Spatial Data Using Markov Random Fields, with Application to Elemental Composition of Forest Soil." *Mathematical Geology* 38 (2): 113–133.

Hrafnkelsson, B., and N. Cressie. 2003. "Hierarchical Modeling of Count Data with Application to Nuclear Fall-out." *Environmental and Ecological Statistics* 10 (2): 179–200.

Hsu, S. C., and C. F. Chien. 2007. "Hybrid Data Mining Approach for Pattern Extraction from Wafer Bin Map to Improve Yield in Semiconductor Manufacturing." *International Journal of Production Economics* 107 (1): 88–103.

Huang, Q. 2010. "Physics-driven Bayesian Hierarchical Modeling of the Nanowire Growth Process at Each Scale." *IIE Transactions* 43 (1): 1–11.

Hughes, J., and M. Haran. 2013. "Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75 (1): 139–159.

Hwang, J. Y., and W. Kuo. 2007. "Model-based Clustering for Integrated Circuit Yield Enhancement." *European Journal of Operational Research* 178 (1): 143–153.

Jin, R., C. J. Chang, and J. Shi. 2012. "Sequential Measurement Strategy for Wafer Geometric Profile Estimation." *IIE Transactions* 44 (1): 1–12.

Journel, A. G., and C. J. Huijbregts. 1978. *Mining Geostatistics*. London: Academic Press.

Lee, S., G. Wolberg, and S. Y. Shin. 1997. "Scattered Data Interpolation with Multilevel B-splines." *IEEE Transactions on Visualization and Computer Graphics* 3 (3): 228–244.

Li, R. Z., and A. Sudjianto. 2005. "Analysis of Computer Experiments Using Penalized Likelihood in Gaussian Kriging Models." *Technometrics* 47 (2): 111–120.

Li, B., K. Wang, and A. B. Yeh. 2013. "Monitoring the Covariance Matrix via Penalized Likelihood Estimation." *IIE Transactions* 45 (2): 132–146.

Lin, J., and K. Wang. 2011. "Online Parameter Estimation and Run-to-run Process Adjustment Using Categorical Observations." *International Journal of Production Research* 49 (13): 4103–4117.

Lin, J., and K. Wang. 2012. "A Bayesian Framework for Online Parameter Estimation and Process Adjustment Using Categorical Observations." *IIE Transactions* 44 (4): 291–300.

Lindgren, F., and H. Rue. 2008. "On the Second-order Random Walk Model for Irregular Locations." *Scandinavian journal of statistics* 35 (4): 691–700.

Liu, S., F. Chen, and W. Lu. 2002. "Wafer Bin Map Recognition Using a Neural Network Approach." *International Journal of Production Research* 40 (10): 2207–2223.

Marinescu, I. D., E. Uhlmann, and T. Doi. 2006. *Handbook of Lapping and Polishing*. Boca Raton, FL: Chapman & Hall/CRC Press.

Martin, J. D., and T. W. Simpson. 2005. "Use of Kriging Models to Approximate Deterministic Computer Models." *AIAA Journal* 43 (4): 853–863.

Ntzoufras, I. 2011. *Bayesian Modeling Using WinBUGS*. Hoboken, NJ: John Wiley & Sons.

O'Mara, W. C., R. B. Herring, and L. P. Hunt. 1990. *Handbook of Semiconductor Silicon Technology*. Park Ridge, IL: Noyes Publications.

O'Mara, W., R. Herring, and L. Hunt. 2007. *Handbook of Semiconductor Silicon Technology*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC Press.

Orton, J. W. 2009. *Semiconductors and the Information Revolution: Magic Crystals that made IT Happen*. Amsterdam: Academic Press.

Qian, P. Z. G., and C. F. J. Wu. 2008. "Bayesian Hierarchica Modeling for Integrating Low-accuracy and High-accuracy Experiments." *Technometrics* 50 (2): 192–204.

Qiu, P., C. Zou, and Z. Wang. 2010. "Nonparametric Profile Monitoring by Mixed Effects Modeling." *Technometrics* 52 (3): 265–277.

Quirk, M., and J. Serda. 2001. *Semiconductor Manufacturing Technology*. Upper Saddle River, NJ: Prentice Hall.

Raja, J., B. Muralikrishnan, and S. Fu. 2002. "Recent Advances in Separation of Roughness, Waviness and Form." *Precision Engineering* 26 (2): 222–235.

Richmond, A. 2003. "Financially Efficient Ore Selections Incorporating Grade Uncertainty." *Mathematical Geology* 35 (2): 195–215.

Robert, C. P. 2007. *The Bayesian Choice: From Decision-theoretic Foundations to Computational Implementation*. 2nd ed. New York: Springer Science+Business Media.

Rue, H., and L. Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL: Chapman & Hall/CRC Press.

Shen, J., Z. Pei, G. Fisher, and E. Lee. 2006. "Modelling and Analysis of Waviness Reduction in Soft-pad Grinding of Wire-sawn Silicon Wafers by Support Vector Regression." *International Journal of Production Research* 44 (13): 2605–2623.

Simpson, T. W., T. M. Mauery, J. J. Korte, and F. Mistree. 1998. "Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization." *AIAA Paper 98* 4758 (7): 381–391.

Simpson, T. W., T. M. Mauery, J. J. Korte, and F. Mistree. 2001. "Kriging Models for Global Approximation in Simulation-based Multidisciplinary Design Optimization." *AIAA Journal* 39 (12): 2233–2241.

Taam, W. 1998. "A case Study on Process Monitoring for Surface Features." *Quality and Reliability Engineering International* 14 (4): 219–226.

Tait, A., R. Henderson, R. Turner, and X. Zheng. 2006. "Thin Plate Smoothing Spline Interpolation of Daily Rainfall for New Zealand Using a Climatological Rainfall Surface." *International Journal of Climatology* 26 (14): 2097–2115.

Thomas, A., N. Best, D. Lunn, R. Arnold, and D. Spiegelhalter (2004). "GeoBUGS User Manual." Accessed 2012. http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/geobugs12manual.pdf

Tonkin, M. J., and S. P. Larson. 2002. "Kriging Water Levels with a Regional-linear and Point-logarithmic Drift." *Ground Water* 40 (2): 185–193.

Valette, S., and P. Prost. 2004. "Wavelet-based Multiresolution Analysis of Irregular Surface Meshes." *IEEE Transactions on Visualization and Computer Graphics* 10 (2): 113–122.

Voltz, M., and R. Webster. 1990. "A Comparison of Kriging, Cubic Splines and Classification for Predicting Soil Properties from Sample Information." *Journal of Soil Science* 41 (3): 473–490.

Wang, Y. 1998. "Smoothing Spline Models with Correlated Random Errors." *Journal of the American Statistical Association* 93 (441): 341–348.

Wang, K., and K. Han. 2013. "A Batch-based Run-to-run Process Control Scheme for Semiconductor Manufacturing." *IIE Transactions* 45: 658–669.

Wang, C. H., S. J. Wang, and W. D. Lee. 2006. "Automatic Identification of Spatial Defect Patterns for Semiconductor Manufacturing." *International Journal of Production Research* 44 (23): 5169–5185.

Xu, L., and Q. Huang. 2012. "Modeling the Interactions Among Neighboring Nanostructures for Local Feature Characterization and Defect Detection." *IEEE Transactions on Automation Science and Engineering* 9 (4): 745–754.

Yin, J., S. Ng, and K. Ng. 2011. "Kriging Metamodel with Modified Nugget-effect: The Heteroscedastic Variance Case." *Computers and Industrial Engineering* 61 (3): 760–777.

Zhao, H., R. Jin, S. Wu, and J. Shi. 2011. "PDE-constrained Gaussian Process Model on Material Removal Rate of Wire Saw Slicing Process." *Journal of Manufacturing Science and Engineering* 133 (2): 021012-1–021012-9.

Appendix 1

In this appendix, the construction of the IGMRF and its precision matrix will be introduced in detail.

The thicknesses at different locations on the wafer are correlated, which was validated by the experimental data that we collected. There are two assumptions needed for constructing a GMRF model (Rue and Held 2005): the random vector follows a multivariate normal distribution, and non-adjacent sites are conditionally independent, which means two non-adjacent sites are independent when the values of other sites are given. Based on these two assumptions, the fields we constructed must be proper, which implies that a normal distribution with a symmetrical positive definite covariance matrix is equivalent to a GMRF.

The GMRF model has two disadvantages (Besag and Kooperberg 1995). One is that the marginal variances of the variables in the vector often vary in scale, which is not desirable for modelling a smooth surface. The other is that appreciable correlations between the neighbouring sites may require the parameters to be very close to the boundary of the parameter space. One solution to these disadvantages is to use an improper version, the Intrinsic GMRF (IGMRF). The precision matrix of IGMRF is not full rank. Therefore, the above assumptions cannot be used to construct the IGMRF. Besag and Kooperberg (1995) found that a Gaussian vector with a symmetrical positive semi-definite precision matrix satisfies $\mathbf{Q}\cdot\mathbf{1}=\mathbf{0}$, where $\mathbf{Q}$ is the precision matrix and $\mathbf{1}$ is a vector with all elements equal to 1. Any linearly independent differences among the variables will have a proper Gaussian distribution. We can also construct the IGMRF by using the independent increments, which follow a proper Gaussian distribution.

For two neighbouring sites i and j, we define a normal increment as

$$w_i - w_j \sim N(0, \kappa^{-1})$$

$$f(w_i - w_j) = \frac{1}{\sqrt{2\pi}}\kappa^{1/2}\exp\left(-\frac{\kappa}{2}(w_i - w_j)^2\right)$$

For all of the unordered neighbour pairs, the joint probability density is

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{n-1}\kappa^{(n-1)/2}\exp\left(\sum_{i\sim j}-\frac{\kappa}{2}(w_i - w_j)^2\right)$$

The power of κ equals $n-1$, due to the presence of hidden linear constraints that exist among the variables. In the example of a three-element vector $\mathbf{w}=(w_1,w_2,w_3)'$, every two sites are neighbours. Hence, there are three unordered adjacent differences, $w_2-w_1, w_3-w_1, w_3-w_2$, and we can see that $w_3-w_2=(w_3-w_1)-(w_2-w_1)$.

From the above joint probability density function, we therefore have

$$f(\mathbf{w}) \propto \kappa^{(n-1)/2}\exp\left(\sum_{i\sim j}-\frac{\kappa}{2}(w_i - w_j)^2\right),$$

$$\sum_{i\sim j}-\frac{\kappa}{2}(w_i - w_j)^2 = -\frac{\kappa}{2}\sum_{i\sim j}(w_i - w_j)^2 = -\frac{\kappa}{2}\sum_{i\sim j}(w_i^2 - 2w_iw_j + w_j^2).$$

Because the number of neighbours for site i is n_i, we can write the sum in matrix form:

$$-\frac{\kappa}{2}\sum_{i\sim j}(w_i^2 - 2w_jw_j + w_j^2) = -\frac{\kappa}{2}\mathbf{w}^T\mathbf{R}\mathbf{w}$$

where $\mathbf{R}$ is called the structure matrix, which is symmetric, and

$$R_{ij} = \begin{cases} n_i, & \text{if } i=j \\ -1, & \text{if } i\sim j \\ 0, & \text{otherwise} \end{cases}.$$

Let $\mathbf{Q}=\kappa\mathbf{R}$, and then we have

$$f(\mathbf{w}) \propto \kappa^{(n-1)/2}\exp\left(-\frac{1}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}\right),$$

$$Q_{ij} = \kappa\begin{cases} n_i, & \text{if } i=j \\ -1, & \text{if } i\sim j \\ 0, & \text{otherwise} \end{cases}.$$

The conditional distribution of w_i can be obtained as

$$w_i|\mathbf{w}_{-i},\kappa \sim N\left(\frac{1}{n_i}\left(\sum_{j:j\sim i}w_j\right),\frac{1}{n_i\kappa}\right), \quad i=1,2,\ldots,n.$$