

國立臺灣大學工學院機械工程學系

博士論文

Department of Mechanical Engineering

College of Engineering

National Taiwan University

Doctoral Dissertation

時間與空間上之變異分析及其在半導體工程資料分析上的應用

Temporal and Spatial Variation Analysis and  
Its Applications to Semiconductor Engineering Data Analysis



Jakey Blue

指導教授：陳正剛 博士

Advisor: Argon Chen, Ph.D.

中華民國 99 年 1 月

January, 2010

*To my family, beloved Jolin, and dear friends*



## Abstract

Investigation of system variation is always critical to process/equipment optimization and yield enhancement in semiconductor manufacturing. Conventional variation estimate, usually the sample variance, cannot truthfully reveal the random variation if data exhibits a patterned profile or is of non-stationary distribution. The biased random variation estimate could then impact the subsequent analysis greatly. In this research, the concept of moving variance, which calculates the variance of a small number of consecutive/adjacent observations within a temporal/spatial moving window, is proposed to eliminate the impact of the pattern-induced (systematic) variation. By applying the moving variance technique to temporal profiles, such as the process states or tool signals, the tool condition can be evaluated by the proposed tool condition indicator. When dealing with spatial topography, such as the wafer metrology data, systematic variations can be identified and characterized by the proposed spatial variation spectrum (*SVS*) comprised of the spatial moving variances. Diagnosis methodologies are developed to facilitate uncovering abnormal tool conditions or systematic patterns. Properties and theories are studied as well to justify how the moving variance outperforms the conventional sample variance. With the tool condition indicator, possible tool faults can be identified and proper maintenance measures can be scheduled accordingly. With the *SVS* and its summarized indices, systematic variations can be characterized and the causal analysis for finding root causes can be further explored. The proposed methodologies are further validated through the real cases provided by local semiconductor companies.

**Keywords:** fault detection and classification (FDC), moving variance, tool condition indicator, spatial variation, systematic variation, random variation, wafer CD metrology.

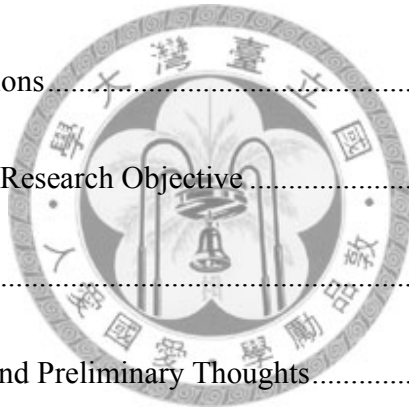
## 中文摘要

在半導體製造過程中，為了最佳化製程產能與設備利用率，以致能提升良率，系統變異的分析研究是一個相當基本且重要的課題，傳統上，工程師以統計學中的樣本變異數來估計資料的隨機變異，但該統計量在資料呈現特殊走勢或來自非平穩的分配時，往往會遭到曲解失真，以致於影響接下來進行的製程分析與最佳化。本研究提出了計算平移變異數的概念以消彌由於資料呈特殊走勢、或來自非平穩分配所造成的影響，平移變異數主要觀念在於只計算時間軸上小區間中連續(或空間中小區域內相鄰)觀測值的樣本變異數，再藉由移動該區間(或區域)來覆蓋所有觀測值，以匯整各區間(或區域)中的變異資訊。在處理時間軸變化為主的資料，例如機台在製造過程中可即時收集的參數與訊號，本研究以平移變異數為基礎發展了一機台狀態指標，以評估當前機台狀況、找出可能的機台錯誤，使後續的機台維護保養排程能夠更適當、更有彈性；而為了處理空間座標對應的觀測值，例如晶圓臨界尺度取樣量測值等，我們則利用平移變異數發展了一空間變異頻譜來描繪晶圓量測值內含的系統變異，並在空間變異頻譜之上建立了數個指標來量化整體資料的系統變異量，使後續因果分析的進行能更有效率。同時，我們亦探討了平移變異數相關的性質與理論，並試著與傳統的樣本變異數比較，證明在資料呈特殊走勢或來自不同分配時，使用平移變異數能有較精準的估計。本研究在最後並利用了國內數家半導體製造業者提供的真實數據來驗證所提出的各項理論。

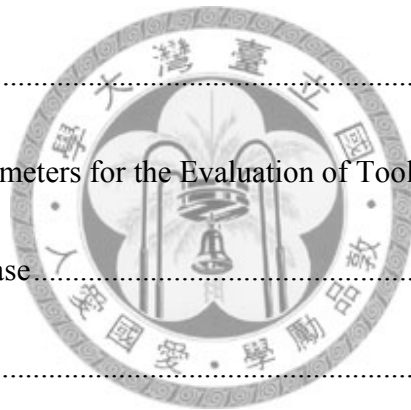
**關鍵字：**錯誤偵測與分類，平移變異數，機台狀態指標，空間變異，系統變異，隨機變異，晶圓臨界尺度量測值。

# Contents

Abstract.....	i
中文摘要.....	ii
Contents .....	iii
Contents of Figures .....	vi
Contents of Tables .....	ix
Chapter 1 - Introduction.....	1
1.1 Background and Motivations.....	1
1.2 Problem Description and Research Objective.....	2
1.3 Research Framework .....	4
Chapter 2 - Literature Review and Preliminary Thoughts.....	5
2.1 Analysis of Temporal Tool Parameters .....	5
2.2 Analysis of Spatial Wafer Topography.....	7
Chapter 3 - Variation Estimation for Temporal Tool Parameters.....	11
3.1 Distribution-based Tool Condition Evaluation.....	11
3.1.1 Moving Variance and Covariance .....	13
3.1.2 Effects of Patterns .....	16



3.2 Tool Condition Monitoring and Diagnosis.....	21
Chapter 4 - Variation Estimation for Spatial Topography Data .....	25
4.1 Characterization of Spatial Variation.....	25
4.1.1 Spatial moving variance.....	26
4.1.2 Spatial Variation Spectrum .....	28
4.1.3 Identification of Systematic Variation.....	32
4.2 Spatial Pattern Index .....	36
Chapter 5 - Case Study .....	41
5.1 Real Temporal Tool Parameters for the Evaluation of Tool Condition.....	41
5.1.1 The PECVD Tool Case.....	41
5.1.2 The PVD Tool Case.....	47
5.2 Real Wafer Topography for the Characterization of Spatial Variation .....	51
5.2.1 Hypothetical Wafer Topographies with Common Patterns.....	51
5.2.2 Wafer Metrology Data under Different Precision Levels.....	54
5.2.3 Wafer Metrology Data with Checkerboard Pattern .....	56
Chapter 6 - Conclusion .....	59
6.1 Preliminary Results.....	59



6.2 Future Study.....	61
Appendix.....	67
A. Proof for THEOREM 1 .....	67
B. Proof for THEOREM 2 .....	69
C. Proof for Corollary 1 .....	71
D. The Simulation of Hypothetical Wafer Topography .....	72
E. Proof for THEOREM 3 .....	73
F. Proof for THEOREM 4 .....	74
G. Reasoning for Conjecture 1 .....	77
Reference .....	79



## Contents of Figures

Fig. 1. The framework of this research.....	4
Fig. 2. Profile of SVID $X$ under Recipe A.....	14
Fig. 3. Profile of SVID $X$ under Recipe B.....	14
Fig. 4. Profiles of SVID $X$ and $Y$ under recipe A.....	19
Fig. 5. Profiles of SVID $X$ and $Y$ under recipe B.....	20
Fig. 6. Steps to monitor the tool condition by the generalized moving variances.....	21
Fig. 7. The visualization of a hypothetical wafer topography with dome pattern (described in Appendix D): (a) 2-D contour map; (b) 3-D response surface.....	26
Fig. 8. An illustration for calculating the spatial moving variance of window size $p = 3$ .....	28
Fig. 9. The $SVS$ of a hypothetical wafer topography with dome pattern.....	29
Fig. 10. (a) The contour map of a randomly distributed wafer topography; (b) The $SVS$ of the topography.....	31
Fig. 11. The variance of the $SVS$ and its estimated degrees of freedom.....	34
Fig. 12. The $SVS$ and its upper control limit for a randomly distributed wafer topography.....	36
Fig. 13. The concept of the calculation of spatial pattern index for the wafer topography with dome-pattern.....	37
Fig. 14. The concept of calculating the $SPI$ for the randomly distributed wafer topography.....	38



Fig. 15. PECVD condition indicator vs. recipe changes.....	42
Fig. 16. Distinct PECVD Throttle Valve profiles under two recipes. ....	42
Fig. 17. PECVD Tool condition monitoring based on generalized moving variances.....	43
Fig. 18. Abnormal throttle valve readings when tool event is issued. ....	44
Fig. 19. EWMA control chart of the PECVD tool condition ( $\lambda=0.05, l=4$ ).....	44
Fig. 20. Sampled throttle valve readings of 2500 <sup>th</sup> ~4300 <sup>th</sup> wafers before the tool event in Fig. 19. .....	45
Fig. 21. EWMA control chart for the decomposition of PECVD's generalized moving variances: SVID Variability ( $\lambda=0.05, l=4$ ).....	45
Fig. 22. EWMA control chart for the decomposition of PECVD's generalized moving variances: SVID Interrelation ( $\lambda=0.05, l=4$ ).....	46
Fig. 23. Abnormal changes of relationship between gas flow and throttle valve during ramp-up and ramp-down steps. ....	47
Fig. 24. PVD condition indicator vs. recipe changes.....	48
Fig. 25. PVD tool condition monitoring based on generalized moving variances; abnormal tool condition occurred after a Life PM was applied.....	49
Fig. 26. EWMA control chart for the PVD tool condition ( $\lambda=0.05, l=4$ ). ....	49
Fig. 27. EWMA control chart for the decomposition of PVD's generalized moving variances: SVID Variability ( $\lambda=0.05, l=4$ ).....	50

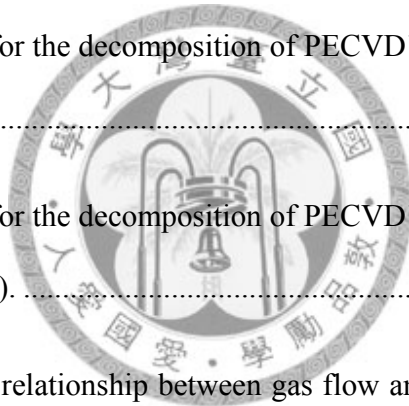


Fig. 28. EWMA control chart for the decomposition of PVD's generalized moving variances: SVID Interrelation ( $\lambda=0.05, l=4$ ). .....	50
Fig. 29. Abnormal SVID profiles of nano-torr readings vs. normal ones. ....	51
Fig. 30. Analysis of $SVS$ 's for the hypothetical wafer topographies. ....	52
Fig. 31. The upper control limit for the $SVS$ 's of the hypothetical wafer topographies. (The round shapes magnify the high-frequency parts of the spectra.).....	53
Fig. 32. Analysis of spatial variation spectra for the real wafer metrology data under different precision levels.....	55
Fig. 33. The identification of systematic variations for the real wafer metrology data under different precision levels. (The round shapes magnify the high-frequency parts of the spectra.)	56
Fig. 34. Analysis of spatial variation spectra for the real wafer topographies with checkerboard patterns.....	58
Fig. 35. The identification of systematic variations for the real wafer topographies with checkerboard patterns. (The round shapes magnify the high-frequency parts of the spectra.) ....	58
Fig. 36. Two SVID profiles with different ramp rates.....	62
Fig. 37. Moving-window variances for SVID $X$ and $X'$ . ....	63
Fig. 38. $VoVS$ for a randomly distributed topography. ....	65
Fig. 39. Histograms of the simulated spatial moving variances under spatial variation frequencies: (a) $f=0.0032$ ( $p=2$ ); (b) $f=0.5$ ( $p=308$ ). ....	78

## Contents of Tables

Table I. <i>SPI</i> and Variation Ratios of the Noise and Dome Pattern.....	39
Table II. <i>SPI</i> and variation ratios for the hypothetical wafer topographies.....	54
Table III. <i>SPI</i> and variation ratios of real wafer metrology data under different precision levels. .....	56
Table IV. <i>SPI</i> and variation ratios of real wafer topographies with checkerboard patterns. ....	58



## Chapter 1 - Introduction

### 1.1 Background and Motivations

With improvement of sensor and information technology, modern semiconductor manufacturing facilities now generate enormous amount of tool parameters, such as voltage, temperature, and wafer metrology data, such as critical dimension (CD), which all must be analyzed quickly and effectively for the process/equipment optimization and circuit design. To characterize the tool condition or process stability, the investigation of data variation within tool parameters or wafer metrology data is the key point, in particular, when the fabrication technology continues to advance.

In practice, sample variance of a dataset is usually used as the estimation of data variation. However, if the mean level of the data changes frequently or the data distribution is non-stationary, the sample variance easily gets biased. Unfortunately, shifted mean level and heterogeneous data distribution are commonly encountered during the analyses of semiconductor manufacturing data, such as the chamber temperature trends up to a steady state and then trends down according to the tool recipe. Before analyzing the variation of collected data, engineers usually need to pre-process it to get a homogeneous dataset or select the stable part of data for simplicity, both of which would possibly distort the real variation.

As the study of variation serves to be the important basis of every analysis activity for finding the root causes of yield loss, to truthfully reveal the data variation becomes an extremely critical issue for process control and circuit design and has motivated this research.

## 1.2 Problem Description and Research Objective

Nowadays, even with the highly automated and precisely monitored equipments used in a nearly dust-free clean room and operated with well-trained engineers, the occurrence of process abnormality and metrology non-uniformity still can't be avoided. With the large amount of manufacturing data collected, characterization of tool condition and wafer metrology uniformity should be carried out faithfully and efficiently. Among all of the statistical properties, sample variance of data is usually the first and the most frequently studied statistic to perform the system analyses.

The calculation of sample variance, denoted as  $S^2$ , is well-known and written in the following equation:

$$S^2 = \frac{\sum_{i=1}^n (M_i - \bar{M}_i)^2}{n-1}, \quad (1)$$

where  $M_i$  is the  $i^{\text{th}}$  random sample in the collected dataset,  $n$  denotes the number of random samples, and  $\bar{M}_i$  is the average of the  $n$  random samples. If the mean or variance within the data does not change at all, i.e. the data distribution is stationary, the sample variance is good enough to estimate the data variation statistically. However, semiconductor manufacturing data such as tool parameters and wafer metrology data always exhibit non-stationary profiles. To avoid getting a biased variation estimate, techniques for preprocessing the temporal signal data and spatial metrology data have been developed separately by engineers. Nevertheless, it would be time-consuming to identify the non-stationary properties in order to have proper methodologies developed and applied to the data, and not to mention the process data and wafer metrology are collected under distinct fundamentals.

In practice, the tool parameters are collected along the time dimension as the wafer is processed and can be demonstrated as a temporal profile while the metrology data is measured from the sampling sites on the wafer surface and is displayed as a spatial topography. To estimate the variation of a non-stationary temporal profile, engineers usually set temporal windows within a process run and calculate summarized indicators, such maximum, minimum, average, and variance of the observations within every temporal window. The tool condition is then monitored based on these summarize indicators. To characterize the variation of a non-stationary spatial topography such as wafer metrology, spatial patterns on the wafer topography are usually contrasted with a set of known systematic patterns based on engineering knowledge. The causal analysis for finding the causes to abnormalities can be done based on the identified pattern.

However, handling both temporal and spatial data requires a lot of manual efforts and domain knowledge and still might not be able to reveal the data variation truthfully. Therefore, the objective of this research is to develop a novel estimation to the non-stationary data collected no matter from processing tools or wafer metrologies. In particular, practitioners do not need to trim or transform the data in order to get a stationary distribution for further statistical analyses. The basic idea of this research is to propose a moving variance, which is calculated using a small number of consecutive (or adjacent) observations in a temporal (or spatial) moving window, such that the effect of pattern-induced variation can be removed.

With development and application of moving variance to the temporal process data, the tool condition can be evaluated based on the better estimate of tool variability. A tool condition indicator is then proposed to monitor the tool condition and identify tool faults. On the other hand, the systematic patterns on the wafer topography can be characterized by applying the moving variance technique to the spatial metrology data. Methodology for the identification of

significant patterns on wafer topography would be developed as well. In summary, a truthful estimation to data variation and its properties will be studied, and the application methodologies will be developed according in this research.

### 1.3 Research Framework

In the following chapters, literatures in relation to conventional characterization of the variation of temporal and spatial data are firstly reviewed. Then the moving variance is developed and the methodologies for applying moving variance to temporal and spatial data are described in chapter 3 and 4, respectively. Case studies are demonstrated in chapter 5 and the preliminary results and future work will be discussed in chapter 6. The organization of this research is described in Fig. 1.

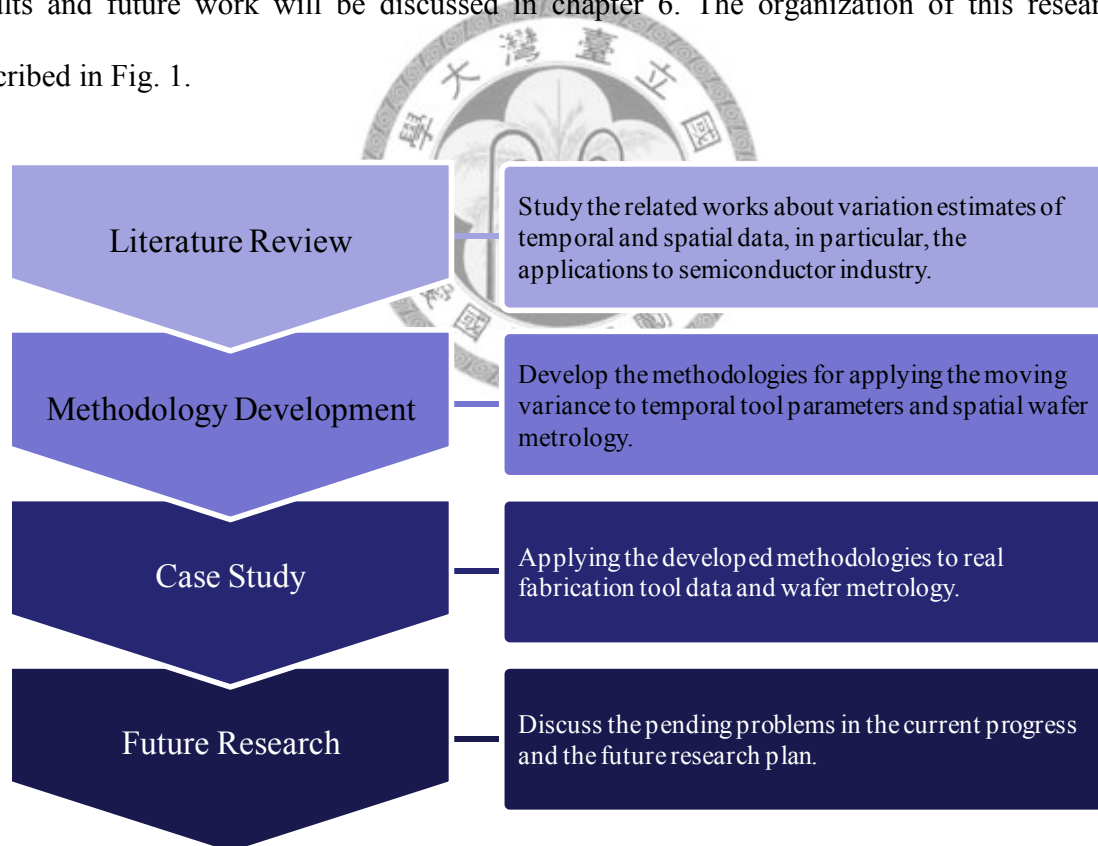


Fig. 1. The framework of this research.

## **Chapter 2 - Literature Review and Preliminary Thoughts**

In this chapter, the related works of analyzing temporal tool parameters for the evaluation of tool condition are firstly reviewed. Then, the literatures which discuss spatial variation for the pattern characterization within wafer topography are studied.

### **2.1 Analysis of Temporal Tool Parameters**

Advanced sensor and information technologies have made real-time tool data readily accessible to tool and process engineers. A significant number of tool parameters (Status Variable Identifications, SVID's) is collected during wafer processing and a large amount of tool data is acquired and available for fault detection and classification (FDC). Many IC makers have substantially improved the process capabilities by implementing FDC. With the real-time tool data, one can also evaluate the overall tool condition so that tool maintenance can be more effectively scheduled and the post-maintenance tool condition can be more easily qualified.

Conventional FDC techniques usually require data preprocessing procedures including setting temporal windows to signify steps within a process run and calculating summarized indicators, such as maximum, minimum, average and standard deviation of each SVID's observations in every temporal window. Model-based statistical process control (SPC) schemes are then used to model the temporal patterns and to monitor and detect shifts or drifts in the tool signals (Yue and Tomoyasu, 2004, Lacaille and Zagrebnoy, 2007, and He and Wang, 2007). However, the FDC indicators based on pattern modeling for specific recipes are not suitable to represent the overall tool condition, which should be independent of recipe changes or even assignable causes of local faults. The purpose of this research is not to detect specific tool faults, as the FDC technique



does, but rather to propose a recipe-independent tool condition indicator and diagnosis methodologies.

To characterize the tool condition, various types of tool data should be accounted for. With the modern sensors built in the advanced processing tool, there are usually tens or even hundreds of tool parameters (SVID's) collected. If the large amount of real-time tool data can be properly prepared and examined, one should be able to evaluate the overall tool condition for effective tool predictive maintenance (PM) schedule and for qualification of the post-maintenance tool condition. It's always a great challenge to look at so many SVIDs' real-time data all together. Gertsbakh (1977) has proposed using Discriminant Analysis to find a linear combination function of tool parameters that best distinguishes between a "good" tool and a "failed" one. Similar to Discriminant Analysis, a linear combination function of parameters with the maximum contribution to the tool condition can be found through Principal Component Analysis or Singular Value Decomposition (Stamatis *et al.*, 1992). However, the frequent change of recipes and the diversity of operations still make the overall tool condition evaluation a very difficult task.

The difficulties described above have motivated this research to focus on consolidating the large number of SVID's into a single tool condition indicator robust to the operation/recipe changes. Not only should the indicator be able to provide an easy reading for engineers to have a quick idea on the tool's overall performance, it should also serve as the basis for condition-based PM schedule. In this research, we attempt to propose a moving statistic to minimize the effect of operation/recipe changes. Then, the concept of evaluating the tool condition by observing the distribution of the tool parameters' readings (Chen *et al.*, 1998, and Chen and Wu, 2007) is employed. A tool condition evaluating method is thus proposed based on the ideas of

multivariate generalized variance (Johnson, 2002). In addition, the effect of operation/recipe changes will be minimized by using the moving statistics.

With the tool condition indicator available, the exponentially weighted moving average (EWMA) control scheme is employed to detect tool condition excursion. A two-step diagnosis method to find the causes of the abnormalities is then developed. The first step is to decompose the generalized variance into two parts: one is the variance and the other is covariance. The second step is to detect the anomalies of the SVID variability and/or relationships by observing the variance and covariance trends within the process run to discover the root cause and to predict the tool condition excursion.

## 2.2 Analysis of Spatial Wafer Topography

As the metrology technology continues to advance, more and more wafer data must be analyzed quickly and efficiently for monitoring and controlling the fabrication processes. The wafer spatial variations study is a key to yield enhancement especially when the fabrication technology enters the 32nm node. Therefore, the study of spatial variation becomes critical to both process control and circuit design (Stine *et al.*, 1997, Boning and Chung, 1996). In doing this, the systematic and random components, both of which contribute to the spatial variations, must be identified before the root cause of yield loss can be found and removed.

Kibarian *et al.* (1991) examine the spatial dependencies (referred to as spatial correlations) of the process parameters, such as polysilicon line width and film thickness, on circuit testing data. Mozumder and Lowenstein (1992), and Guo and Sachs (1993) model the within-wafer variation based on multiple response surface methods while Smith *et al.* (1998) compare it with the single response surface methods. Boning and Chung (1996) describe the concept of statistical

metrology and the decomposition of spatial variations into wafer-to-wafer, die-to-die, site-to-site variations, and residuals. Stine *et al.* (1997) characterize the wafer-level, die-level, and wafer-die spatial variations for correlation studies. The literatures extracting and characterizing the spatial variations can be mainly categorized into two groups (Vanoppen *et al.*, 2001). In one group, the variation is separated into variance components by employing methods such as analysis of variance (ANOVA) or Fourier transform. The other group identifies the distinct exemplar-based variation patterns and analyzes the impact of these factors through proper decomposition of the wafer and yield data.

To characterize the wafer-level or die-level variance components, ANOVA methods are widely applied (Stine *et al.*, 1996, and Zimmerman *et al.*, 2000, and Cain and Spanos, 2003). Significance of the components can be then ranked for further causal analysis. At the process level, Steele *et al.* (2002) assume the total critical dimension (CD) variation to be the combination of independent variance components from coating, developing, and baking processes and use design of experiments to model and understand the CD uniformity issues. Yu *et al.* (1995) employ the fast Fourier transform (FFT) to decompose the wafer CD's spatial variations into wafer spatial patterns or residual variations. Ye *et al.* (1995) and Han *et al.* (1997) analyze the pattern generator-induced mask CD errors in the spatial frequency domain and identify error contributors using the Fourier transform. Ouyang *et al.* (1998, 1999) identify the amplitude excursions in the spatial frequency domain of CD's using the spatial Fourier transform (SFT). By applying an inverse SFT, the variance components can be separated and used to explain the systematic and random errors in spatial variations. The idea of transforming observations in the space domain to a spectrum in the frequency domain is very useful for understanding the natures of the spatial variations. In particular, the low-frequency parts of the

spectra are usually considered caused by the systematic patterns while the high-frequency part is believed to be a result of the noise in the space domain (Yu *et al.*, 1995, and Ouyang *et al.*, 1998, and 1999). However, the FFT/SFT is not able to reduce the domain dimension. For example, the two-dimensional spatial wafer metrology data remains two-dimensional after transforming to the frequency domain by the SFT. Moreover, the statistical properties of the SFT spectrum in the frequency domain are not clear enough to construct a hypothesis test for identifying the existence of systematic variations.

Systematic wafer spatial variations usually form patterns on the two-dimensional wafer map. The advancement of information technology has helped enhance the 2-D/3-D visualization of the wafer metrology data and thus facilitate the causal analysis when studying spatial variations. Wong *et al.* (2002) propose a three-step methodology to characterize the line-width variation. Spatial analysis first decomposes the CD's metrology data into several variance components. Causes with similar spatial signatures, defined based on engineering knowledge, are then classified by contributor-specific measurements. Unanticipated components are finally classified as residuals. Vanoppen *et al.* (2001) apply the methodology for breaking down and ranking of the systematic sources of line-width variations. Evaluation of the exposure tool performance in relation to the contributors of line-width variation is also presented. Burch *et al.* (2008) and Inani *et al.* (2006) recently propose a failure signal detection algorithm (FSDA) which serves as a yield fault detection and diagnosis solution integrating several practical data mining techniques and engineering data analysis methods. FSDA first identifies the known failure metrics and spatial/reticle zones in preparation of wafer data. Wafers with similar patterns are then clustered together. Characteristics of the patterned clusters are used for the drilldown yield analysis to identify the root causes of yield loss.

Most existing methods discussed above, no matter pre-assuming the variance components or pre-defining the exemplar variation patterns, require knowing related issues/faults of the process/tool in advance and categorize the unexplained parts as residuals. Engineering knowledge is definitely helpful in analyzing the variation components for specific failure types. However, engineering knowledge is usually acquired through a high-price learning process where faults or yield losses are found in the later stages of fabrication with corresponding engineering causes learned to locate in the much earlier stages. In fact, any systematic pattern must be associated with certain physical issues. The engineering knowledge can thus be established through relating the data excursion to the out-of-control processes. It is our belief that we should have the data reveals itself as much as possible so that the corresponding engineering knowledge can be learned as early as possible. Therefore, a model-free methodology without priori knowledge to reveal systematic patterns is proposed in this research so that possible problems and respective knowledge can be learned immediately after they occur.

Since the pattern-dependent uniformity is thought to reveal critical information about the spatial variation, a novel spatial variation spectrum (*SVS*) generated based on the calculation of spatial moving variances to characterize the spatial variations is proposed. The *SVS*, which manifests the significance of the topographical systematic patterns, will be further summarized as an overall systematic pattern index based on the formation of the spectrum. The objective of this research is to provide the methodology for faithfully summarizing spatial topography into reasonable indices which thus makes the further causal analyses more efficiently.

## Chapter 3 - Variation Estimation for Temporal Tool Parameters

The idea of tool condition evaluation based on data distribution is firstly introduced in this chapter. Moving variance and covariance statistics are then proposed and consolidated to evaluate the tool condition. The EWMA control scheme and a two-step procedure are employed and developed to identify the root causes to the abnormalities.

### 3.1 Distribution-based Tool Condition Evaluation

To evaluate a process's performance by comparing the distribution of the multiple product quality characteristics against their specifications, a measure known as multivariate process capability indices (*PCP*'s) is usually employed (Taam *et al.*, 1993, and Kotz and Johnson, 1993). The following equation shows the idea of the multivariate process capability indices:

$$C_p = \frac{\text{volume of specification region}}{\text{volume of region containing 99.73\% of distribution}}$$

The same idea is used to a machine capability index (*MCI*) for evaluating the tool condition by Chen and Wu (2000 and 2007).

In reality, to estimate and obtain the specification region requires prior engineering knowledge about the equipment behavior. For example, the chamber pressure is controlled by the throttle valve and flow rates of various gases. The pressure level and the flow rates are usually set by the recipe. The throttle valve is then manipulated by an internal control loop to attain the target pressure level. Even for a given recipe, it's still quite a difficult task to set the specifications for the throttle valve. It's even more difficult to find its relationship with the pressure to form an elliptical spec region. Some works (see, for example, Taam *et al.*, 1993, and Chen and Tsai, 2004)

have tried to establish the specification regions. However, in the high-mix production environment, it's just not practical to set specification regions for hundreds or even thousands of recipes and modify them each time certain tool maintenance is performed.

To borrow the idea of using the multivariate distribution as the tool performance and to keep away from the difficulties of setting the specification regions for all SVID's, recipes, and tools, a statistical measure known as generalized variance (Johnson, 2002) is proposed. Generalized variance, the determinant of the covariance matrix, is in effect proportional to the volume of data distributed in the multi-dimensional variable space and can be used to measure the dispersion of a tool's variability during a process run. The basic assumption here is that there should be a regular size of data distribution under a normal process run. When a tool condition becomes unstable, the distribution of SVID's grows larger. By taking the determinant, the generalized variance is also able to consolidate the large amount of tool data into a single indicator.

Variances of individual SVID's and the covariance among SVID's, i.e., elements of the generalized variance, are all greatly affected by the temporal profiles of SVID's. Different recipes directly result in different temporal profiles. Recipes, however, are changed frequently during production and should not be accounted an impact factor on tool condition. This research proposes to use the moving variance and covariance, of which properties will be studied to minimize the effect of temporal profiles and, thus, to avoid the influence of recipe changes. The tool condition indicator is then calculated by taking the determinant of the moving covariance matrix and is called a generalized moving variance.

### 3.1.1 Moving Variance and Covariance

Tool condition depends greatly on the variability of tool parameters during wafer processing. Each processing step, no matter it's a temperature ramp-up step or a deposition step, the smaller the variability the better the tool condition. However, conventional calculation of sample variance and covariance is quite sensitive to the non-stationary pattern displayed in each SVID's temporal profile. The pattern often incurs a bias in estimation of the tool variability. The non-stationary patterns are distinct not only among different SVID's but also from one recipe to another. Fig. 2 and Fig. 3 show two distinct profile patterns of a tool SVID  $X$  under two recipes (A and B). A variance/covariance estimate independent of recipe changes is thus critically needed to more faithfully reflect the tool's real variability. In this research, we propose to use the pooled moving variance/covariance, which are calculated using a small number of consecutive observations in a moving time window, to be the estimates of the tool variability. The proposed estimate will be shown model-free and minimizing the effect of the patterns.

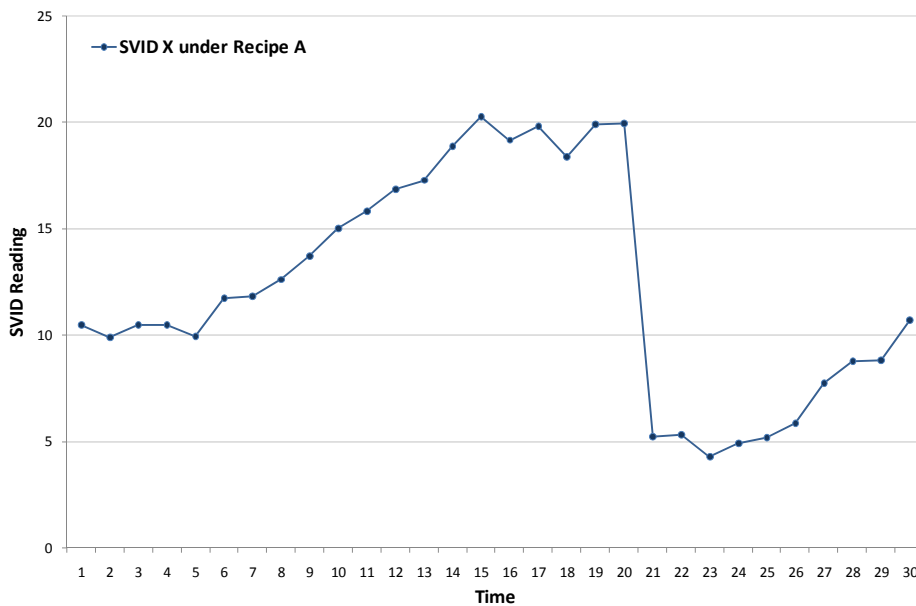




Fig. 2. Profile of SVID  $X$  under Recipe A.

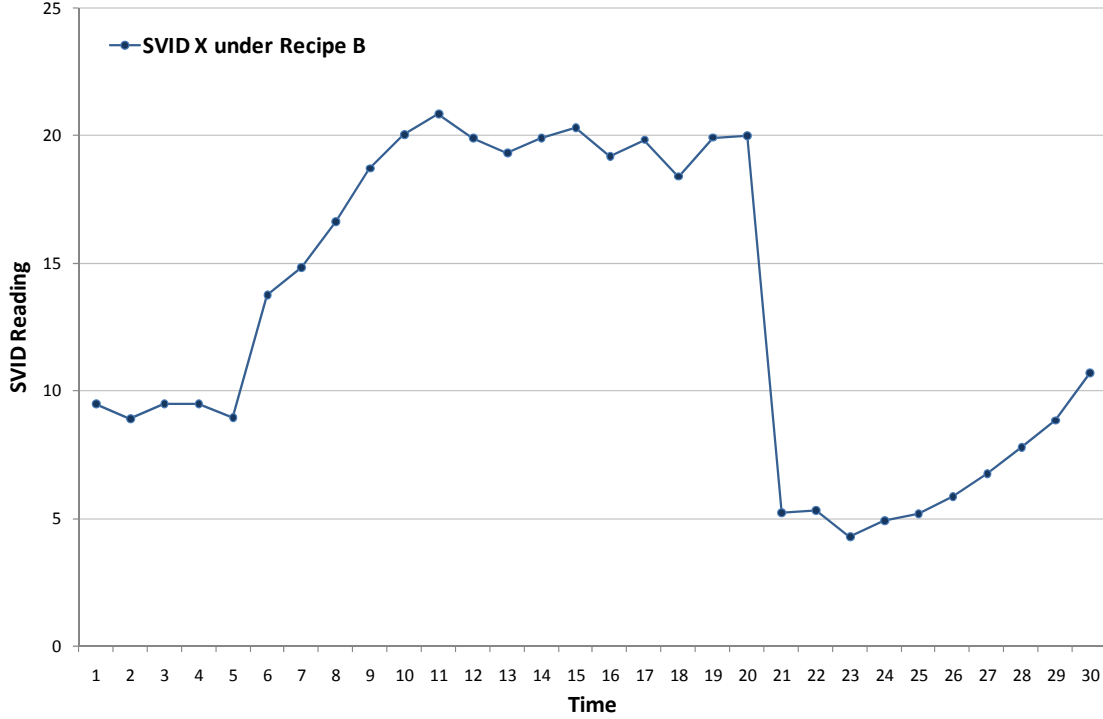


Fig. 3. Profile of SVID  $X$  under Recipe B.

Assume that there are two SVID's,  $X$  and  $Y$ , with true variability represented by variances  $\sigma_X^2$  and  $\sigma_Y^2$ , respectively, and covariance,  $\sigma_{XY}$ . Each SVID collects  $n$  sample data from a process run under a specific recipe. That is, two corresponding temporal series of random samples,  $X_1, X_2, \dots, X_n$  and  $Y_1, Y_2, \dots, Y_n$ , are collected for SVID  $X$  and  $Y$ . The sampling rate of SVID observations is assumed constant; i.e., time intervals between any two successive observations are assumed equal and fixed. Our attempt now is to estimate the true tool variability,  $\sigma_X^2$ ,  $\sigma_Y^2$  and  $\sigma_{XY}$ , using the two temporal series of observations. The conventional estimates are the sample variance, which is similar to (1), and covariance:

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}; S_Y^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}; \text{ and} \quad (2)$$

$$S_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}. \quad (3)$$

where  $\bar{X}$  and  $\bar{Y}$  represents the averages of the two series, respectively.

With patterns inherent in the temporal profiles, the above estimation could be quite inaccurate. In the literature, models are usually built first to describe the patterns. Variability is then estimated using the model residuals. This sort of estimates is, however, model-based and ineffective because the models themselves are sensitive to operation/recipe changes. To minimize the effects of patterns and to find the true variability without modeling the patterns, we propose using moving variance and covariance based on the following moving windows of size  $p$  ( $< n$ ):

$$W_1^X = \{X_1, \dots, X_p\}, W_2^X = \{X_2, \dots, X_{p+1}\}, \dots, W_{n-p+1}^X = \{X_{n-p+1}, \dots, X_n\}, \text{ and}$$

$$W_1^Y = \{Y_1, \dots, Y_p\}, W_2^Y = \{Y_2, \dots, Y_{p+1}\}, \dots, W_{n-p+1}^Y = \{Y_{n-p+1}, \dots, Y_n\}$$

for the temporal series of SVID  $X$  and  $Y$ , respectively.

The sample variance is then calculated for each of the moving windows. A total of  $n-p+1$  sample variances are obtained for each SVID temporal series. These sample variances can be now pooled together to obtain a moving sample variance for each SVID, that is,

$$\hat{\sigma}_X^2 = \frac{1}{n-p+1} \sum_{j=1}^{n-p+1} S_{W_j^X}^2 \text{ for SVID } X \text{ and}$$

$$\hat{\sigma}_Y^2 = \frac{1}{n-p+1} \sum_{j=1}^{n-p+1} S_{W_j^Y}^2 \text{ for SVID } Y. \quad (4)$$

where  $S_{W_j^X}^2$  and  $S_{W_j^Y}^2$  are sample variances calculated for the  $j^{\text{th}}$  moving windows  $W_j^X$  and  $W_j^Y$ , respectively.

Likewise, a moving covariance can be obtained by pooling together the sample covariance calculated from the  $n-p+1$  moving windows:

$$\hat{\sigma}_{XY} = \frac{1}{n-p+1} \sum_{j=1}^{n-p+1} S_{W_j^X W_j^Y}^2 \quad (5)$$

where  $S_{W_j^X W_j^Y}^2$  is the sample covariance calculated using the random samples in the  $j^{\text{th}}$  pair of moving windows  $W_j^X$  and  $W_j^Y$ .

### 3.1.2 Effects of Patterns

To show that the proposed moving statistics are relatively robust against the patterns of SVID temporal profiles, we first define two pattern elements: *linear drift* and *step change*. The pattern of a SVID temporal profile can be seen as formed by segments of linear drifts with step changes in between any two successive segments.

We first look at the linear drifts. Linear-drift patterns of SVID  $X$  and SVID  $Y$  can be characterized by a linear function of the observation count  $i$  during the process run:

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} a\sigma_X \\ b\sigma_Y \end{bmatrix} i + \boldsymbol{\varepsilon}_i \quad \text{for } i=1, \dots, n, \quad (6)$$

where  $a \neq 0$  and  $b \neq 0$  are the amounts of drifts, expressed by proportions of  $\sigma_X$  and  $\sigma_Y$ , between two successive observations of SVID  $X$  and SVID  $Y$ , respectively and

$$\boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{X,i} \\ \varepsilon_{Y,i} \end{bmatrix} \text{ for } i=1, \dots, n \quad (7)$$

are independent and follow an identical bivariate normal distribution with the mean vector equal to 0 and covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}.$$

$\boldsymbol{\Sigma}$  is the true process variability that can be used to evaluate the tool condition and that we intend to estimate.

**THEOREM 1.** By setting the size ( $p$ ) of the moving window smaller than the number ( $n$ ) of observations in one process run, the moving variance and covariance in (4) and (5) as the estimates of the true variability  $\boldsymbol{\Sigma}$  under the linear-drift pattern defined in (6) have smaller bias than the conventional sample variance and covariance have. That is,

$$E(\hat{\sigma}_X^2) - \sigma_X^2 < E(S_X^2) - \sigma_X^2; \quad E(\hat{\sigma}_Y^2) - \sigma_Y^2 < E(S_Y^2) - \sigma_Y^2; \quad \text{and} \quad |E(\hat{\sigma}_{XY}) - \sigma_{XY}| < |E(S_{XY}) - \sigma_{XY}|.$$

*Proof:* see Appendix A.

**Corollary 1.** With  $p = 2$ , the estimate bias of the true variability  $\boldsymbol{\Sigma}$  under the linear-drift pattern defined in (6) is minimized by the moving variance and covariance in (4) and (5).

*Proof:* This result can be easily obtained from (29) and (31) in Appendix A. □

Now, we define the step-change patterns of SVID  $X$  and  $Y$  by a step function of the observation count  $i$ :

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{cases} \boldsymbol{\varepsilon}_i & \text{for } i = 1, \dots, n_1 \\ \begin{bmatrix} c\sigma_X \\ d\sigma_Y \end{bmatrix} + \boldsymbol{\varepsilon}_i & \text{for } i = n_1 + 1, \dots, n \end{cases} \quad (8)$$

where  $0 < n_1 < n$ ;  $c \neq 0$  and  $d \neq 0$  are the step-change sizes, expressed by proportions of  $\sigma_X$  and  $\sigma_Y$ , at the  $(n_1+1)^{\text{th}}$  observations in the temporal series of SVID  $X$  and SVID  $Y$ , respectively; that is, there are  $n_1$  observations taken before the step change while the following  $n_2 (= n - n_1)$  observations have a constant mean difference from the previous  $n_1$  observations. The i.i.d.  $\boldsymbol{\varepsilon}_i$  for  $i = 1, \dots, n$  are as defined as in (7) with a variance/covariance matrix  $\boldsymbol{\Sigma}$  representing the tool variability to be estimated.

**THEOREM 2.** By setting the size ( $p$ ) of the moving window smaller than the number of observations before and after the step change during one process run, i.e.,  $p \leq \min(n_1, n_2)$ , and such that

$$\frac{p+1}{6(n-p+1)} < \frac{n_1(n-n_1)}{n(n-1)}, \quad (9)$$

the moving variance and covariance in (4) and (5) as the estimates of the true variability  $\boldsymbol{\Sigma}$  under the step-change pattern (8) have smaller biases than the conventional sample variance and covariance have. That is,

$$E(\hat{\sigma}_X^2) - \sigma_X^2 < E(S_X^2) - \sigma_X^2; \quad E(\hat{\sigma}_Y^2) - \sigma_Y^2 < E(S_Y^2) - \sigma_Y^2; \quad \text{and} \quad |E(\hat{\sigma}_{XY}) - \sigma_{XY}| < |E(S_{XY}) - \sigma_{XY}|.$$

*Proof:* see Appendix B.

**Corollary 2.** With  $p = 2$ , the estimate bias of the true variability  $\boldsymbol{\Sigma}$  under the step-change pattern (8) is minimized by the moving variance and covariance in (4) and (5).

*Proof:* see Appendix C.

Now, back to the examples in Fig. 2 and Fig. 3, though the SVID exhibits distinct temporal profiles under two different recipes, the estimate biases, 1.87 and 2.05, by the sample moving variance are small while estimate the biases, 27.73 and 37.32, by the conventional sample variances are much larger. Fig. 4 and Fig. 5 illustrate the temporal profiles of two SVID's,  $X$  and  $Y$ , under recipe A and B, respectively. Again, the estimate biases, 0.19 and 0.27, by the sample moving covariances of SVID  $X$  and  $Y$  appear quite small while the estimate biases, 20.62 and 30.77, by the conventional sample variances are much larger.

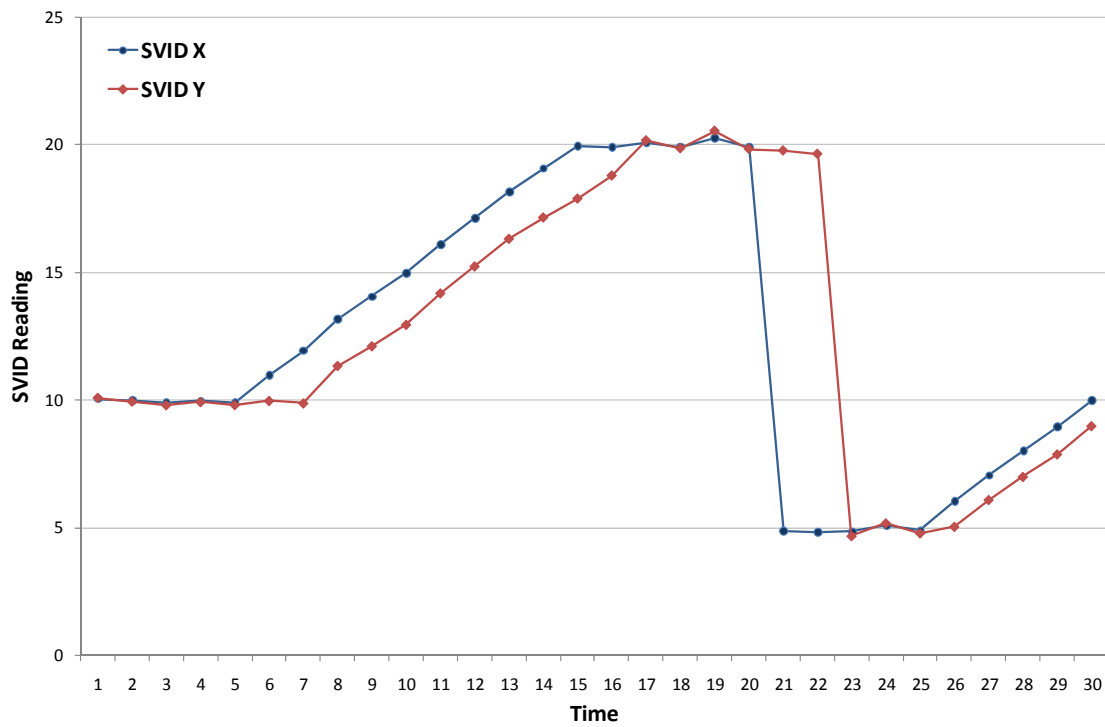


Fig. 4. Profiles of SVID  $X$  and  $Y$  under recipe A.

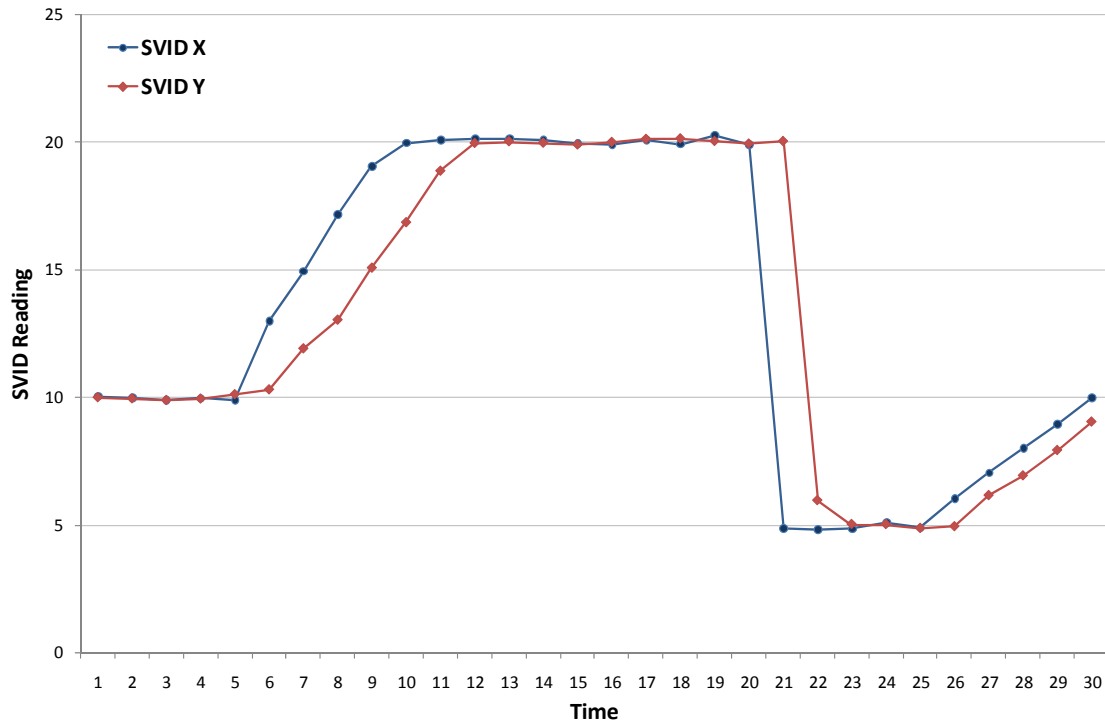


Fig. 5. Profiles of SVID  $X$  and  $Y$  under recipe B.

With the above moving variances and moving covariances, a moving variance/covariance matrix  $\mathbf{S}$  for a tool with  $v$  SVID's can be then calculated:

$$\mathbf{S} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1v} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \hat{\sigma}_{v1} & \cdots & \cdots & \hat{\sigma}_{vv} \end{bmatrix}.$$

The generalized moving variance is then the determinant of  $\mathbf{S}$  and is used as the tool condition indicator. The steps to generate a sequence of indicator values to monitor the tool condition are summarized in Fig. 6.

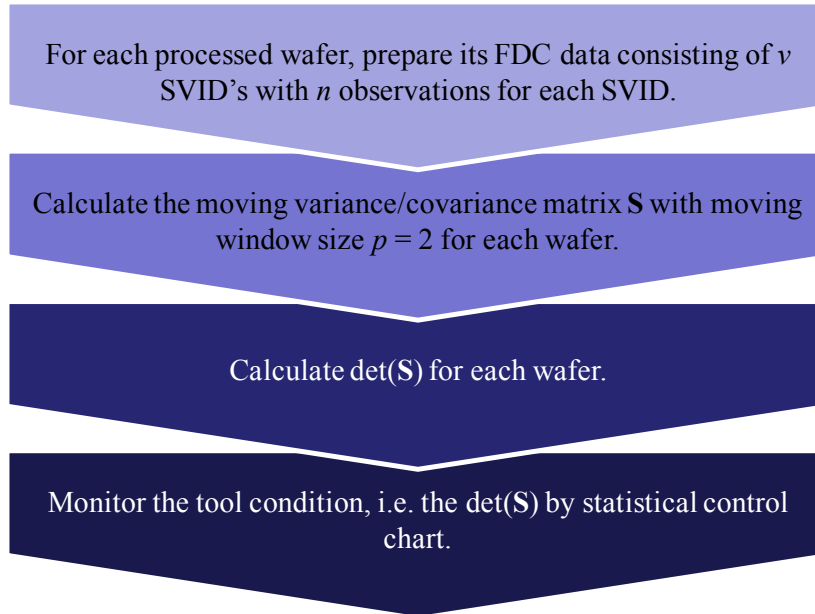


Fig. 6. Steps to monitor the tool condition by the generalized moving variances.

### 3.2 Tool Condition Monitoring and Diagnosis

Tool condition diagnosis can help determine the type of tool maintenance to be performed when the tool condition is found unstable or abnormal. To identify the abnormal patterns shown by the tool condition indicator, statistical control charts can be employed. Here, the well-known Exponentially Weighted Moving Average (EWMA) control scheme (Lucas and Saccucci, 1990, and Borror *et al.*, 1999) is adopted to detect the tool condition excursion. Let  $\Delta_k$  be the tool condition indicator, i.e., the generalized moving variance, for wafer  $k$ . The EWMA statistic,  $Z_k$ , of the tool condition indicator for wafer  $k$  is calculated as follows:

$$Z_k = \lambda \Delta_k + (1 - \lambda) Z_{k-1}, \quad (10)$$

where  $\lambda$  ( $0 < \lambda \leq 1$ ) is the smoothing constant.

The steady-state EWMA control limits are:



$$UCL = \mu_{\Delta} + l\sigma_{\Delta}\sqrt{\frac{\lambda}{2-\lambda}} \text{ and } LCL = \mu_{\Delta} - l\sigma_{\Delta}\sqrt{\frac{\lambda}{2-\lambda}}, \quad (11)$$

where  $\mu_{\Delta}$  and  $\sigma_{\Delta}$  are the mean and standard deviation of in-control  $\Delta_k$ 's, and  $l$  determines the width of the control window. For a standard Shewhart control chart (Montgomery, 2005) and data following the normal distribution, the average run length (ARL) with  $3\sigma$  is known to be 370.4. Since  $\Delta_k$  is not likely to be the normal distribution, we may follow suggestions by Borror *et al.* (1999) to use  $l=2.492$ , 2.703, and 2.86 with corresponding  $\lambda=0.05$ , 0.1, and 0.2, respectively, for the EWMA control scheme to obtain approximately the same ARL=370.4. In this research, after discussions with process engineers and experiments from the actual production data, we set  $\lambda=0.05$  and  $l=4$  for a lower false alarm rate acceptable to the semiconductor manufacturing practice. The setting of the EWMA control scheme should not be fixed across all tool types and situations and should be adjusted according to the natures of the tool processes, which in turn affect the false alarm rate, and by intensive discussions among engineers.

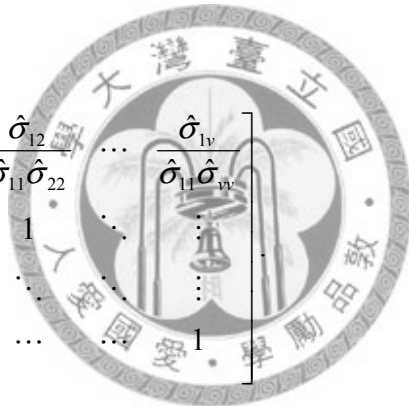
To investigate the unstable or abnormal tool condition, a two-step diagnosis method is developed. The first step is to decompose the generalized moving variance into two parts: one is the moving variance and the other is moving covariance. Given a tool with  $v$  SVID's, a moving variance/covariance matrix  $\mathbf{S}$ ,

$$\mathbf{S} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1v} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \hat{\sigma}_{v1} & \cdots & \cdots & \hat{\sigma}_{vv} \end{bmatrix}$$

can be calculated for each processed wafer. We can set all the moving covariances on the off-diagonal to 0's and then take the determinant of the new matrix  $\mathbf{D}$ , which is in effect the multiplication of all moving variances, to be an indicator which only considers **SVID variability**.

$$\mathbf{D} = \begin{bmatrix} \hat{\sigma}_{11} & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_{22} & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \hat{\sigma}_{vv} \end{bmatrix}.$$

To consider the **SVID interrelation**, the influence of moving variance within  $\mathbf{S}$ , i.e. the effect of  $\mathbf{D}$ , should be removed. Thus, another matrix  $\mathbf{R}$ , which depends solely on the relationships among SVID's, is proposed:

$$\mathbf{R} = (\mathbf{D})^{-\frac{1}{2}} \mathbf{S} (\mathbf{D})^{-\frac{1}{2}} = \begin{bmatrix} 1 & \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{11}\hat{\sigma}_{22}} & \cdots & \frac{\hat{\sigma}_{1v}}{\hat{\sigma}_{11}\hat{\sigma}_{vv}} \\ \frac{\hat{\sigma}_{21}}{\hat{\sigma}_{11}\hat{\sigma}_{22}} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\hat{\sigma}_{v1}}{\hat{\sigma}_{11}\hat{\sigma}_{vv}} & \cdots & \cdots & 1 \end{bmatrix}$$


With  $\mathbf{D}$  and  $\mathbf{R}$ , the property:  $\det(\mathbf{S}) = \det(\mathbf{D})\det(\mathbf{R})$  can be directly obtained from the relationship between the determinants of variance/covariance matrix and correlation matrix, which is:

$$\det(\mathbf{R}) = \det[(\mathbf{D})^{-\frac{1}{2}} \mathbf{S} (\mathbf{D})^{-\frac{1}{2}}] = \det(\mathbf{D}^{-\frac{1}{2}}) \det(\mathbf{S}) \det(\mathbf{D}^{-\frac{1}{2}}) = \frac{\det(\mathbf{S})}{\det(\mathbf{D})}.$$

The determinant of  $\mathbf{R}$ , which is actually the ratio of determinant of  $\mathbf{S}$  to that of  $\mathbf{D}$ , is then used as the indicator that considers SVID interrelation only. It can be seen that the product of two decomposed indicators will be exactly the same as the proposed tool condition indicator.

With the decomposition of the generalized moving variance in the first step, the cause of the tool condition excursion can be identified to be the SVID variability or the SVID interrelation. If the tool condition excursion results from the SVID variability, the second step is to one by one investigate all the sample variances which constitute a moving variance for one SVID as shown in (4), for example,

$$S_{W_1^X}^2, S_{W_2^X}^2, \dots, S_{W_{n-p+1}^X}^2, \text{ for SVID } X,$$

Similarly, we can one by one investigate the sample covariances within a moving covariance for a pair of SVID as shown in (5), for example,

$$S_{W_1^X W_1^Y}, S_{W_2^X W_2^Y}, \dots, S_{W_{n-p+1}^X W_{n-p+1}^Y}, \text{ for SVID } X \text{ and } Y,$$

if the tool condition excursion results from the SVID interrelation. The main purpose of second step is to discover the abnormal SVID variability and/or interrelations among SVID's.

By using the proposed two-step diagnosis method, one can analyze the abnormal tool condition to identify the root causes as unstable variability of SVID's or abnormal relations among SVID's are observed. Furthermore, the profiles of SVID's can be checked to get the full explanations for the abnormal tool condition.

## Chapter 4 - Variation Estimation for Spatial Topography Data

The variation within spatial topography data usually manifests systematic patterns visually on the topography. Since the pattern-dependent uniformities are thought to manifest critical information about the spatial variation, a novel spatial variation spectrum (*SVS*) generated based on the calculation of spatial moving variance is proposed to characterize the spatial variations. The *SVS*, which reveals the significance of the topographical systematic patterns, is further summarized as an overall systematic pattern index based on the formation of the spectrum. The index provides a quick understanding of the systematic patterns displayed in the wafer topography and can be explained by three levels of variation frequency ratios: high, middle, and low.

### 4.1 Characterization of Spatial Variation

Assume that  $n$  random measurements, such as the thickness or linewidth, are taken from sampled metrology sites.  $M_i$  denotes the  $i^{\text{th}}$  random measurements at the metrology site with an Euclidian coordinate  $(m_{x,i}, m_{y,i})$  on the wafer (the origin is referred to the center of the wafer). As can be seen in Fig. 7, a basic understanding of this kind of dataset could be done by drawing a 2-D contour map (Fig. 7a) or a 3-D response surface (Fig. 7b).

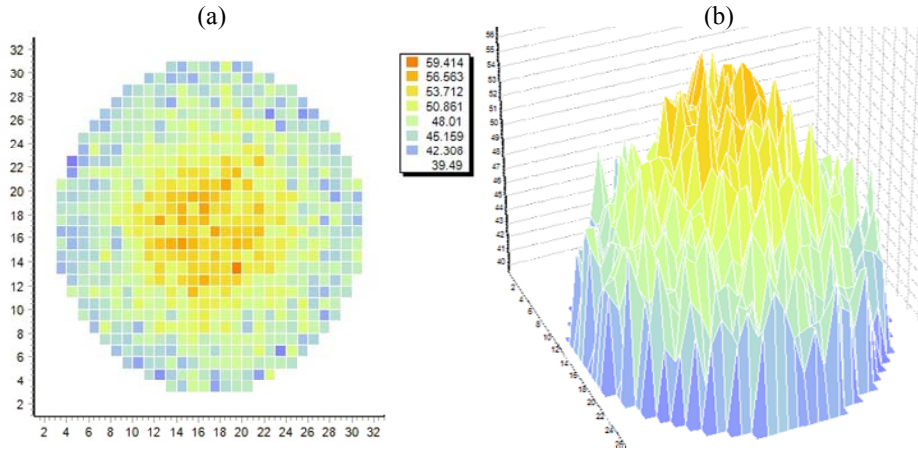


Fig. 7. The visualization of a hypothetical wafer topography with dome pattern (described in Appendix D):

(a) 2-D contour map; (b) 3-D response surface.

To analyze the spatial variation of wafer topography, the sample variance described in (1) is mostly calculated to characterize the spatial variation. However, the sample variance is significantly biased if the observations comprise systematic components, which usually result in systematic patterns on the wafer surface as shown in Fig. 7. In this section, we further extend the concept of moving variance for a temporal series to the spatial moving variance for spatial area of observations. The spatial moving variances of different sizes of spatial moving window then form a *SVS* which characterizes the spatial variations of the topography.

#### 4.1.1 Spatial moving variance

The spatial moving variance utilizes the spatial information, i.e. the Euclidian coordinates, to help decide the size and constituents of a spatial moving window. Given the random measurements  $M_i$ ,  $i=1, \dots, n$ , we first calculate the Euclidian distances for all pairs of observations and get a symmetric distance matrix  $\mathbf{E}$  with all 0's on its diagonal:

$$\mathbf{E} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ d_{n1} & \cdots & \cdots & d_{nn} \end{bmatrix} = \begin{bmatrix} 0 & d_{12} & \cdots & d_{1n} \\ d_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ d_{n1} & \cdots & \cdots & 0 \end{bmatrix}, \quad (12)$$

where  $d_{ij}$  is the distance between  $i^{\text{th}}$  and  $j^{\text{th}}$  observations.

For the  $i^{\text{th}}$  row in  $\mathbf{E}$ , we can sort the distances,  $d_{i1}, \dots, d_{in}$ , in an ascending order and find out the  $k^{\text{th}}$ -nearest measurement for measurement  $i$ , which is denoted as  $M_{i(k)}$ . One special case is that  $M_{i(1)}$  is actually  $M_i$  itself because the nearest observation for  $i$  would be itself based on the distances in (12). A spatial moving window with size  $p$  (where  $2 \leq p \leq n$ ) for observation  $i$  can be then defined as:

$$W_i^p = \{M_{i(1)}, M_{i(2)}, \dots, M_{i(p)}\}, i=1, 2, \dots, n \text{ and } 2 \leq p \leq n. \quad (13)$$

The sample variance for the observations within the spatial moving window  $W_i^p$  is then calculated and denoted as  $S_{W_i^p}^2$ . Given a size  $p$ , a total of  $n$  sample variances will be obtained from the  $n$  spatial moving windows ( $W_i^p, i=1, 2, \dots, n$ ). These sample variances are then pooled together to be the spatial moving variance for window size  $p$ , that is,

$$S_p^2 = \frac{1}{n} \sum_{i=1}^n S_{W_i^p}^2, \quad (14)$$

where  $S_{W_i^p}^2$  is the sample variance calculated for the  $i^{\text{th}}$  spatial moving window  $W_i^p$ , and  $2 \leq p \leq n$ . Furthermore, we can define the expected spatial variance for window size  $p$  as the expected value of  $S_p^2$ . That is

$$\sigma_p^2 = E(S_p^2) \text{ for } 2 \leq p \leq n.$$

When  $p = n$ , the moving window becomes the whole dataset, i.e. all  $n$  observations are used to calculate the sample variance. Thus,  $S_n^2$  is exactly equal to the sample variance in (1).

Fig. 8 shows an example for calculating the spatial moving variance given window size  $p=3$ . Assume the sampling locations are equal-spaced vertically or horizontally, and the vertical space between two adjacent observations is slightly larger than the horizontal one. For the first observation  $m_1$ , the two nearest observations, i.e.  $m_{1(2)}$  and  $m_{1(3)}$ , are  $m_2$  and  $m_6$ , respectively. For the observation at each metrology site, there is a spatial moving window consisting of three observations accordingly and thus generates a sample variance. Then, the  $n$  sample variances for  $p=3$  can be used to calculate  $s_3^2$  in (14).

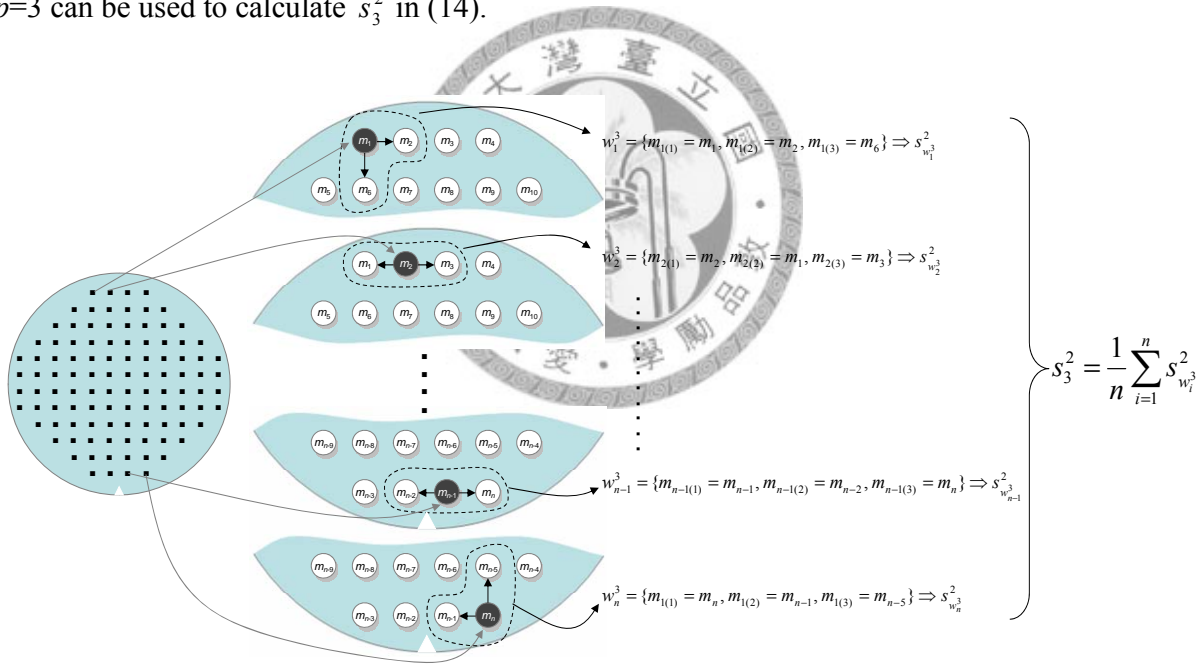


Fig. 8. An illustration for calculating the spatial moving variance of window size  $p = 3$ .

#### 4.1.2 Spatial Variation Spectrum

By varying the size of spatial moving window from 2 to  $n$ , we will have  $n-1$  spatial moving variances:

$$S_2^2, S_3^2, \dots, S_{n-1}^2, S_n^2, \quad (15)$$

which are defined as the components of the *SVS*. Since the total number of metrology sites  $n$  may differ from wafer to wafer, using  $p$ , whose range depends on  $n$ , as the domain for the variance spectrum is not appropriate. We define the spatial variation frequency  $f = p / n$  ( $0 < f \leq 1$ ) for the spectrum domain. A smaller (larger)  $f$  represents a higher (lower) frequency because it covers a smaller (larger) area for calculation of spatial moving variances. The spatial variation frequency not only defines the frequency domain of the spectrum, but also allows comparing spectra calculated from wafers with different numbers of observations as long as the measurements are rather symmetrically and uniformly distributed over the entire wafer. The spatial moving variances can be plotted against the spatial variation frequency to illustrate the structure of spatial variations. Fig. 9 shows a spectrum calculated based on the simulated data in Fig. 7.

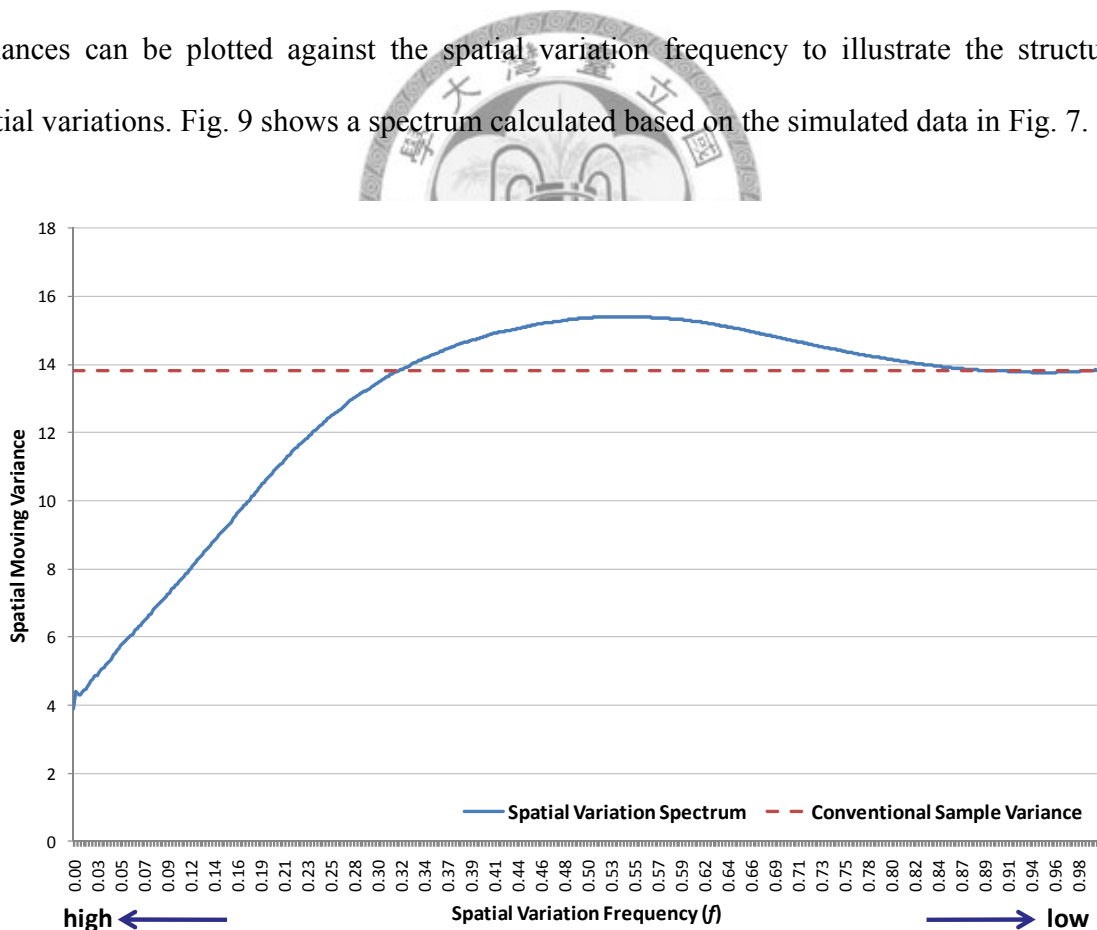
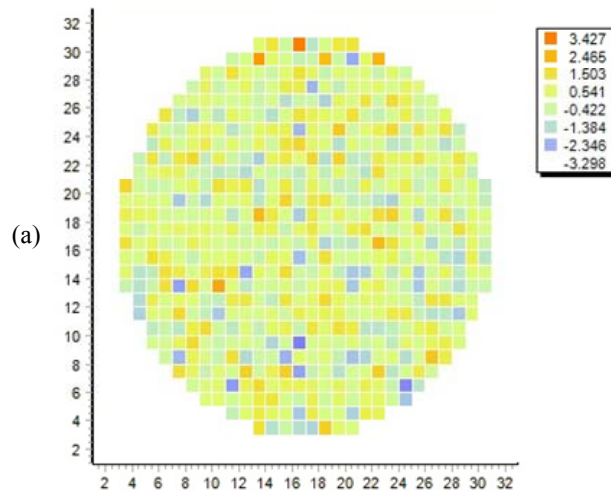
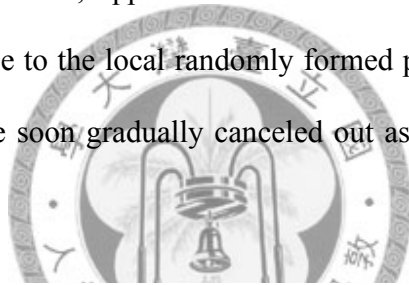


Fig. 9. The *SVS* of a hypothetical wafer topography with dome pattern.



The horizontal dashed line in Fig. 9 denotes the sample variance calculated by (1). As can be seen, the spatial sample variances over the range of middle to low frequencies ( $f > 0.3$ ) are larger than those over the range of higher frequency ( $f < 0.3$ ). The larger sample variances with spatial windows covering at least 30% portion ( $p/n > 0.3$ ) of the wafer surface reflect the pattern.

Fig. 10 shows a metrology data with observed variation randomly distributed and its corresponding  $SVS$ . Because there is no dominant pattern, the  $SVS$  appears to be random around the overall sample variance. The entire spectrum is rather stationary as compared to that in Fig. 9 despite there are small fluctuations locally. In particular, the spatial moving variances of high frequencies, i.e. small window sizes, appear to be more unstable. This is because the small windows are extremely sensitive to the local randomly formed patterns. However, the impact of the local, small patterns will be soon gradually canceled out as the window size increases, and the  $SVS$  becomes flat.



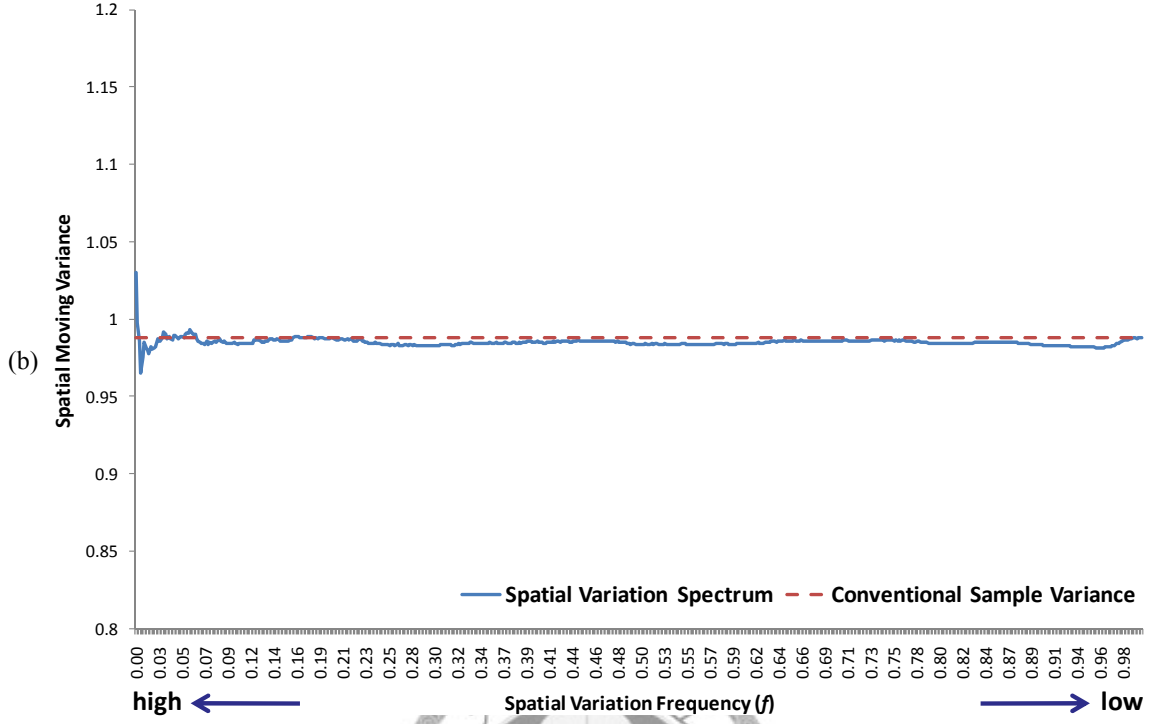


Fig. 10. (a) The contour map of a randomly distributed wafer topography; (b) The *SVS* of the topography.

To further understand the properties of *SVS*, we assume the wafer spatial variations are purely the random variations. Let the random measurement  $M_i$  be defined as:

$$M_i = a + \varepsilon_i \text{ for } i=1, \dots, n, \quad (16)$$

where  $a \neq 0$  is the mean level of all metrology sites, and

$$\varepsilon_i \text{ for } i=1, \dots, n, \quad (17)$$

are independent and follow an identical normal distribution with zero mean and variance  $\sigma^2$ . And let the spatial systematic variations be variations other than random noises satisfying (16) and (17). Then we have the following theorem.

**THEOREM 3.** The spatial variation must consist of variations other than random noises, i.e. systematic variations, if the expected spatial variance spectrum is uneven for  $0 < f \leq 1$ .

*Proof:* see Appendix E.

Based on THEOREM 3, we know that the expected *SVS* should display as a horizontal line if the wafer topography purely consists of random variation. Otherwise, the topography with *SVS* exhibiting a non-stationary pattern must consist of both the systematic and random variations.

#### 4.1.3 Identification of Systematic Variation

Our attempt is to detect whether the metrology data is only randomly distributed or contains systematic patterns. To check if the *SVS* is calculated from the data consisting of random variation only, we study the relation between the conventional sample variance and the proposed spatial moving variance, and develop the following theorem.

**THEOREM 4.** If one takes a sample of  $p$  measurements from a total of  $n$  measurements, there will be  $C_p^n$  combinatorial possibilities. The average of the sample variances of the  $C_p^n$  combinatorial samples is then equal to the sample variance of the  $n$  measurements (1). That is:

$$\frac{1}{C_p^n} \sum_{k=1}^{C_p^n} S_{p,k}^2 = S^2, \text{ for } p=2, \dots, n, \quad (18)$$

where  $S_{p,k}^2$  is the sample variance of the  $k^{\text{th}}$  sample among the  $C_p^n$  possible samples.

*Proof:* see Appendix F.

THEOREM 4 describes another perspective on the composition of the sample variance in terms of many sample variances of sampling subsets. Based on this concept, we have the following conjecture for the proposed spatial moving variances of  $p \leq n-1$ .

THEOREM 4 provides another perspective on the composition of the sample variance in terms of sample variances of combinatorial subsets. Based on this theorem,  $S_{n-1}^2$  is equal to  $S_n^2$  if the metrology sites are symmetrically distributed such that the  $n$  samples of  $S_{n-1}^2$ , i.e.  $W_i^{n-1}$ ,  $i=1, \dots, n$ , are the same as those in the  $C_{n-1}^n$  combinatorial samples. From THEOREM 1, to detect the existence of systematic variations, we need to detect whether the spatial variances are uneven. A hypothesis test of uneven spatial variances can be developed if the probability distribution of  $S_p^2$  is known under (16) and (17). If  $S_{n-1}^2 = S_n^2$ , it is clear that

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2 \quad \text{and} \quad \frac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi_{n-1}^2,$$

where  $\chi_{n-1}^2$  is a chi-square distribution with the degree of freedom  $n - 1$ . For  $2 < p < n-1$ ,  $C_p^n > n$ . The  $n$  samples used for calculating the proposed spatial moving variance,  $S_p^2$ , are only the subset of the  $C_p^n$  combinatorial samples. It is thus unclear what distributions  $S_p^2$  will follow. However, THEOREM 2 provides some clues that lead to the following conjecture.

**Conjecture 1.** If  $S_p^2$ ,  $2 \leq p \leq n-1$ , is calculated from a wafer metrology data defined in (16) and (17), then

$$\frac{v_p S_p^2}{E(S_p^2)} = \frac{v_p S_p^2}{\sigma_p^2} = \frac{v_p S_p^2}{\sigma^2} \sim \chi_{v_p}^2, \quad 2 \leq p \leq n-1, \quad (19)$$

where  $v_p$  denotes the degree of freedom for the  $\chi^2$  distribution and is no greater than  $n - 1$ .

*Reason:* see Appendix F.

To determine  $v_p$  for  $2 \leq p \leq n-1$ , we perform a Monte-Carlo simulation study and obtain the results in Fig. 11. As  $p$  decreases from  $n$ ,  $\lfloor n/2 \rfloor < p < n-1$  ( $1/2 < f < 1$ ), the variance of  $S_p^2$  (solid line) increases because the  $n$  samples for calculating the  $S_p^2$  is becoming a smaller subset of the  $C_p^n$  combinatorial samples. For  $p \leq \lfloor n/2 \rfloor$ , the  $n$  samples is becoming a larger subset of  $C_p^n$  combinatorial samples as  $p$  continues to decrease from  $\lfloor n/2 \rfloor$ . However, the variance of  $S_p^2$  sharply turns up when  $p$  is decreased to be less than 30 ( $f < 0.05$ ) because these particularly small window sizes are sensitive to both local random patterns and unbalanced sampling of observations for calculation of  $S_p^2$ .

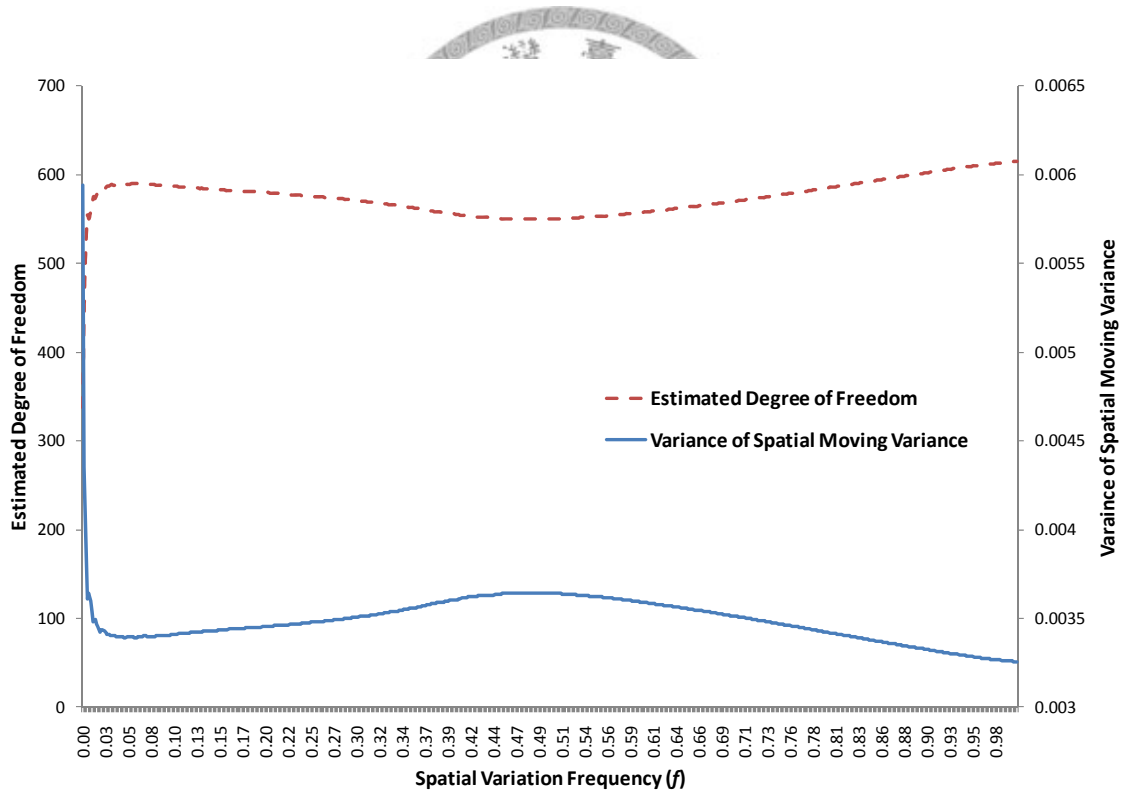


Fig. 11. The variance of the  $SVS$  and its estimated degrees of freedom.

With the Monte-Carlo estimate of the variances of  $S_p^2$  and the assumption of the  $\chi^2$  distribution (Conjecture 1),  $v_p$  in (19) can be then estimated as:

$$v_p = \frac{2(E(s_p^2))^2}{V(s_p^2)} = \frac{2\sigma^4}{V(s_p^2)}, 2 \leq p \leq n \quad (20)$$

where  $V(s_p^2)$  is the sample variance of 100,000 Monte-Carlo simulations  $s_p^2$ 's (see Appendix G).

The dashed line in Fig. 11 indicates the estimated degrees of freedom. Based on the estimated degrees of freedom, we can now construct the following hypothesis test.

$H_0$ :  $\sigma_2^2 = \sigma_3^2 = \dots = \sigma_{n-1}^2 = \sigma_n^2$  versus

$H_1$ : there exist systematic variations.

To reject  $H_0$ , we choose the smallest spatial moving variance:

$$S_{\min}^2 = \min\{S_2^2, S_3^2, \dots, S_{n-1}^2, S_n^2\}. \quad (21)$$

as the test base and test if all the rest of the spatial moving variances are equal to  $S_{\min}^2$ :

$H_0$ :  $\sigma_p^2 = S_{\min}^2$ , for  $p \neq \arg \min_i \{S_k^2, k=1, \dots, n\}$ .

Since  $S_{\min}^2$  is used as the comparison base, only the one-sided test is required. Let  $\chi_{v_p, 1-\alpha}^2$  denote the critical value of the chi-square distribution with  $v_p$  degrees of freedom and the type I error probability  $\alpha$ . An upper control limit (UCL) can be constructed as:

$$\text{UCL} = \sigma_p^2 \leq \frac{\chi_{v_p, 1-\alpha}^2 \cdot S_{\min}^2}{v_p} \text{ for } 2 \leq p \leq n - 1. \quad (22)$$

With the upper control limit, if there is at least one spatial moving variance exceeds the control limit,  $H_0$  is rejected and systematic variations are said to exist; otherwise, there is no evidence to say that the wafer spatial variations consist of any systematic variation. Fig. 12 shows an

example of using the upper control limit with  $\alpha = 0.05$  to check if there exists any systematic variation for the pure-noise metrology data in Fig. 10. As can be seen, the whole spectrum lies within the control limit and there is no sufficient statistical evidence to deny that the metrology data is only randomly distributed.

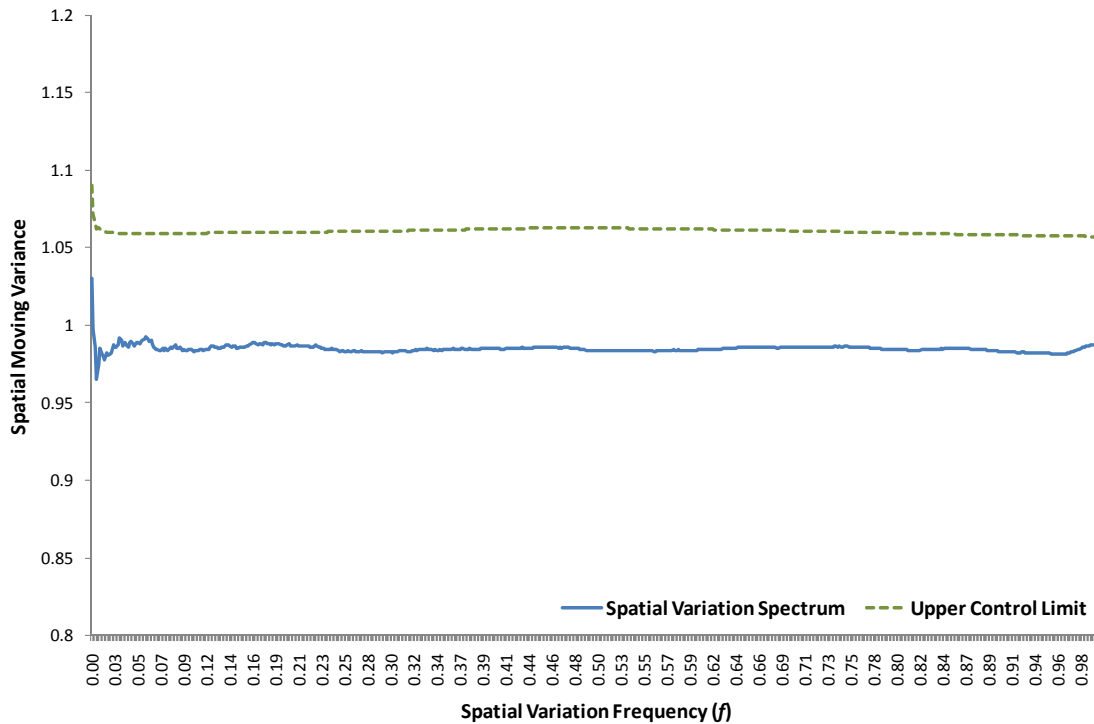


Fig. 12. The *SVS* and its upper control limit for a randomly distributed wafer topography.

## 4.2 Spatial Pattern Index

Even with the hypothesis test proposed above, it would be useful to provide simple indices summarizing the *SVS* to evaluate the significance of the systematic variations. A spatial pattern index (*SPI*) and three variation ratios of high, middle and low frequencies are proposed. The root causes associated with an identified systematic pattern can be then investigated based on the index and the three variation ratios.

Based on THEOREM 1, the spectrum is expected to be a horizontal line if the wafer spatial variations purely consist of random variations and thus the spatial pattern index for a pure noise should be near zero accordingly. For the wafers consisting of systematic patterns, we expect to see a larger *SPI* value indicating a more significant systematic variation. The spatial pattern index (*SPI*) is proposed:

$$SPI = \frac{\sum_{i=2}^n (S_i^2 - S_{\min}^2)}{\sum_{i=2}^n S_i^2}, \quad (23)$$

where  $S_{\min}^2$  defined in (21) is used as an estimation to the random variation.

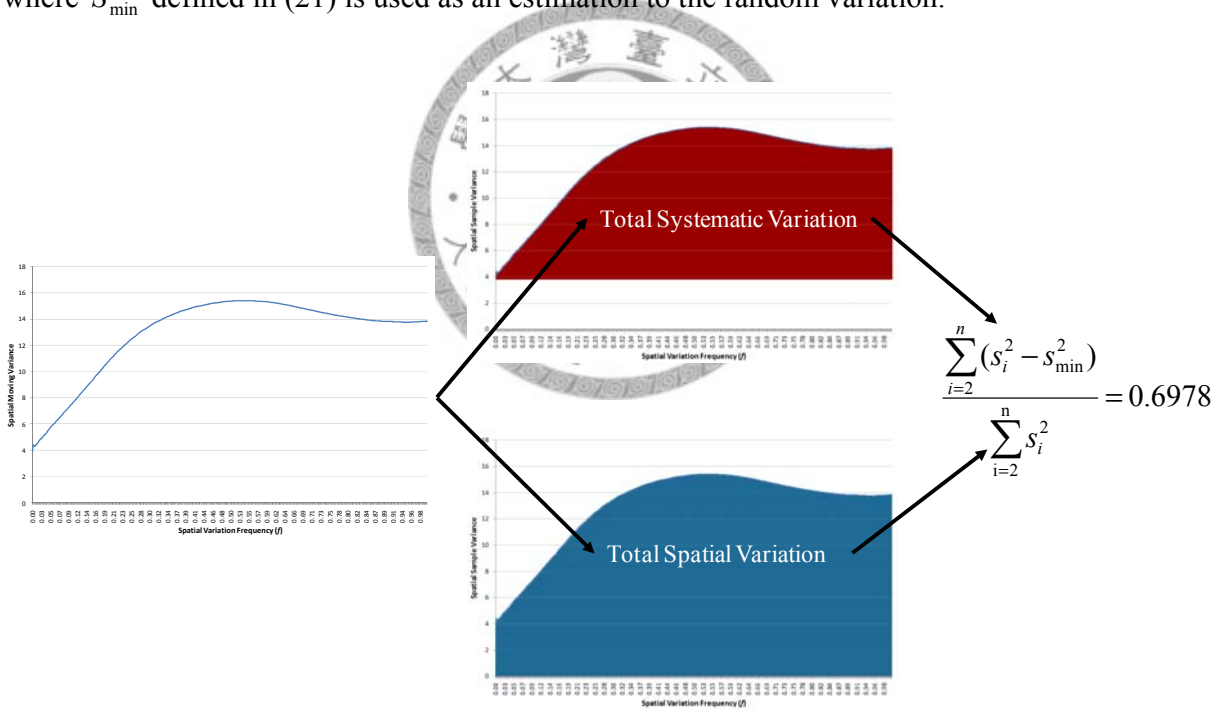


Fig. 13. The concept of the calculation of spatial pattern index for the wafer topography with dome-pattern.



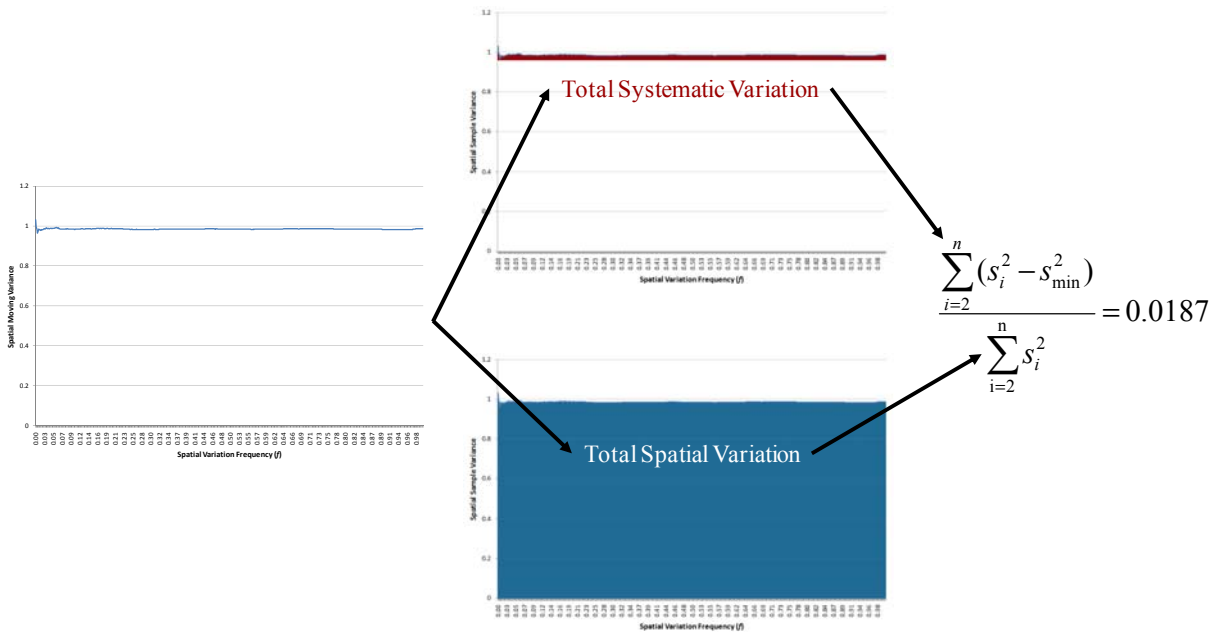


Fig. 14. The concept of calculating the *SPI* for the randomly distributed wafer topography.

The denominator in (23) represents the total spatial variation of the spectrum and the numerator calculates the total systematic variation by removing the random variation from the spatial variation. Therefore, *SPI* is ranged between 0 and 1. Fig. 13 shows the concept when calculating the *SPI* for the *SVS* of the dome-patterned topography discussed in Fig. 9, where the spatial moving variances of middle variation frequencies arise due to the domed systematic pattern in the center of the wafer. As can be seen in Fig. 13, the total spatial variation is illustrated by the middle-up figure while the systematic variation is depicted by the middle-bottom figure. The ratio of the total systematic variation to the total spatial variation is said to be the spatial pattern index (*SPI*), which is 0.6978 in this example.

For a randomly distributed wafer metrology data as in Fig. 10, its  $SPI = 0.0187$  shown in Fig. 14 is very close to zero because the systematic variation portion is relatively small compared to the total spatial variation. To further explain the *SPI*, three variation ratios are calculated by dividing the spectrum into three parts, i.e. high, middle, and low frequencies, which are:

$$\text{High-frequency variation ratio} = \left( \frac{\sum_{j=2}^{\lfloor \frac{n}{3} \rfloor} S_j^2}{\sum_{i=2}^n S_i^2} \right) \times 100\%, \quad (24)$$

$$\text{Mid-frequency variation ratio} = \left( \frac{\sum_{j=\lfloor \frac{n}{3} \rfloor + 1}^{\lfloor \frac{2n}{3} \rfloor} S_j^2}{\sum_{i=2}^n S_i^2} \right) \times 100\%, \quad (25)$$

$$\text{Low-frequency variation ratio} = \left( \frac{\sum_{j=\lfloor \frac{2n}{3} \rfloor + 1}^n S_j^2}{\sum_{i=2}^n S_i^2} \right) \times 100\%. \quad (26)$$

The *SPI*'s and variation ratios for the hypothetical examples in Fig. 13 and Fig. 14 are listed in Table I, where the three ratios for the randomly distributed topography are almost the same because the data is pattern-free. However, the middle-frequency variation ratio of the dome-patterned topography is larger than the high/low ones. It implies that the large  $SPI = 0.8849$  is majorly caused by a medium-size pattern.

With these indices summarized from the *SVS*, one can quickly grasp the significance of the systematic variations. These indices provide simple but effective information about the wafer spatial variations and can be further used to find the root causes.

Table I. *SPI* and Variation Ratios of the Noise and Dome Pattern.

	Noise	Dome
<b>Spatial Pattern Index (<i>SPI</i>)</b>	0.0187	0.6978
<b>High-freq. Variation Ratio</b>	33.19%	24.60%
<b>Mid-freq. Variation Ratio</b>	33.32%	38.82%
<b>Low-freq. Variation Ratio</b>	33.49%	36.58%



## Chapter 5 - Case Study

### 5.1 Real Temporal Tool Parameters for the Evaluation of Tool Condition

#### 5.1.1 The PECVD Tool Case

Real time data of more than 16,000 pieces of wafers with 23 SVID's are collected from a PECVD tool of a local Fab. The sampling rate is down to one observation a second. As one can imagine, the minimum number of SPC charts would be at least 23 for engineers to monitor this tool. Moreover, observations within different temporal windows over the SVID profile would be summarized as different indicators and thus result in hundreds of charts. Maintaining these charts for monitoring a tool is costly and, especially, doesn't consider the interactions among SVID's. In this section, the proposed moving variances and covariances of the collected SVID's are firstly calculated to comprise a moving variance/covariance matrix for each wafer. As described in Fig. 6, the generalized moving variances for these wafers are then calculated and used as a tool condition indicator of the PECVD tool. We can draw a trend chart of the tool condition indicator by the wafer processing sequence and monitor the tool condition. In Fig. 15, the proposed indicator of normal tool condition is quite robust against the recipe changes, indicated by the vertical gray lines despite the SVID profiles under the two of this tool perform obviously different in patterns and even in processing time-length (see Fig. 16 for an example of throttle valve profiles).

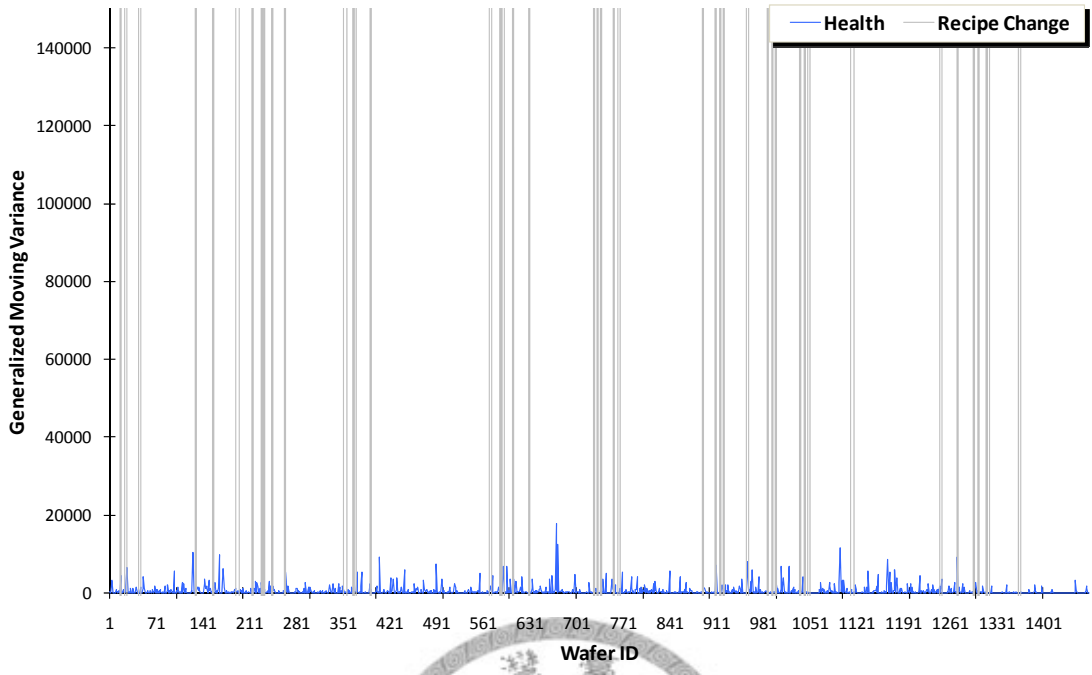


Fig. 15. PECVD condition indicator vs. recipe changes.

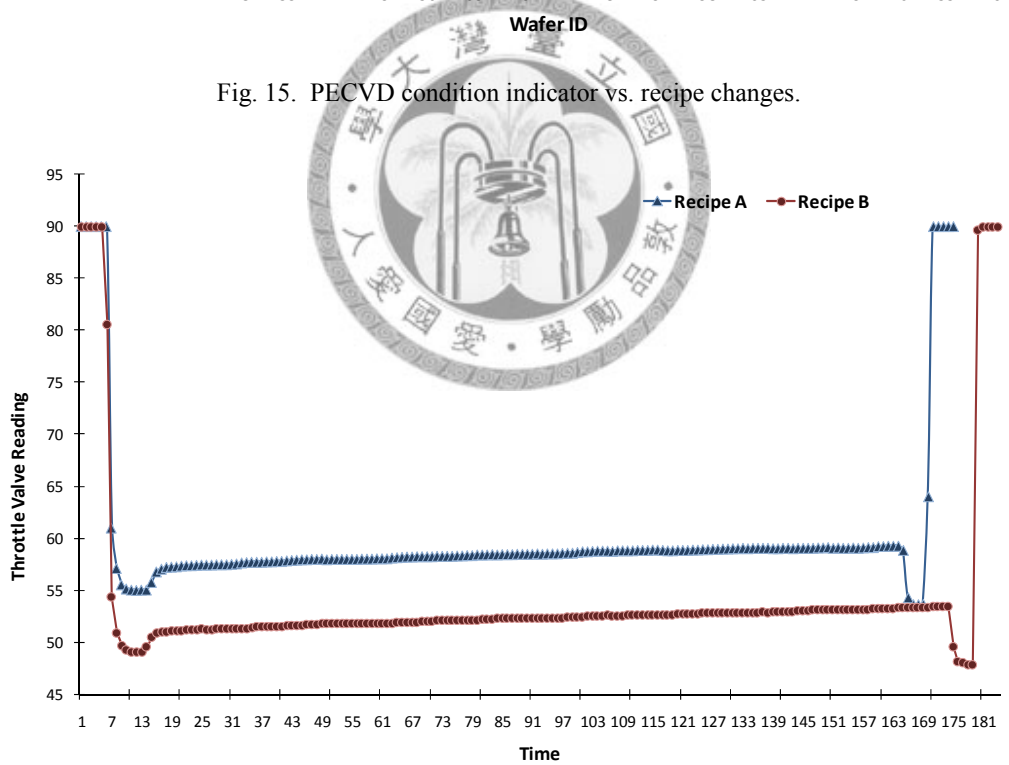


Fig. 16. Distinct PECVD Throttle Valve profiles under two recipes.

Furthermore, we can put some vertical lines to indicate the time points of events, such as tool alarms, or preventive maintenance schedules, and investigate how the trend of the tool condition

indicator responses to the events. As can be seen in Fig. 17, the dashed vertical line indicates the time point of a tool event regarding abnormal throttle valve readings in-between two Monthly PM's (MPM), and the vertical bold line indicates that a Life PM, which involves part replacement due to expiration of the expected life spans, is performed right after the event. Fig. 18 shows that the throttle valve readings appear to drift up when the tool event is issued. By looking at the EWMA control chart in Fig. 19 which is constructed based on equations (10) and (11) with  $\lambda=0.05$  and  $l=4$ , the proposed indicator displays abnormal tool condition starting the 2500<sup>th</sup> wafer long before the tool event. However, as can be seen in Fig. 20, the throttle valve readings appear quite normal during this tool alarm period (2500<sup>th</sup>~4300<sup>th</sup> wafers) and, hence, the tool FDC system is not able to detect any particular fault.

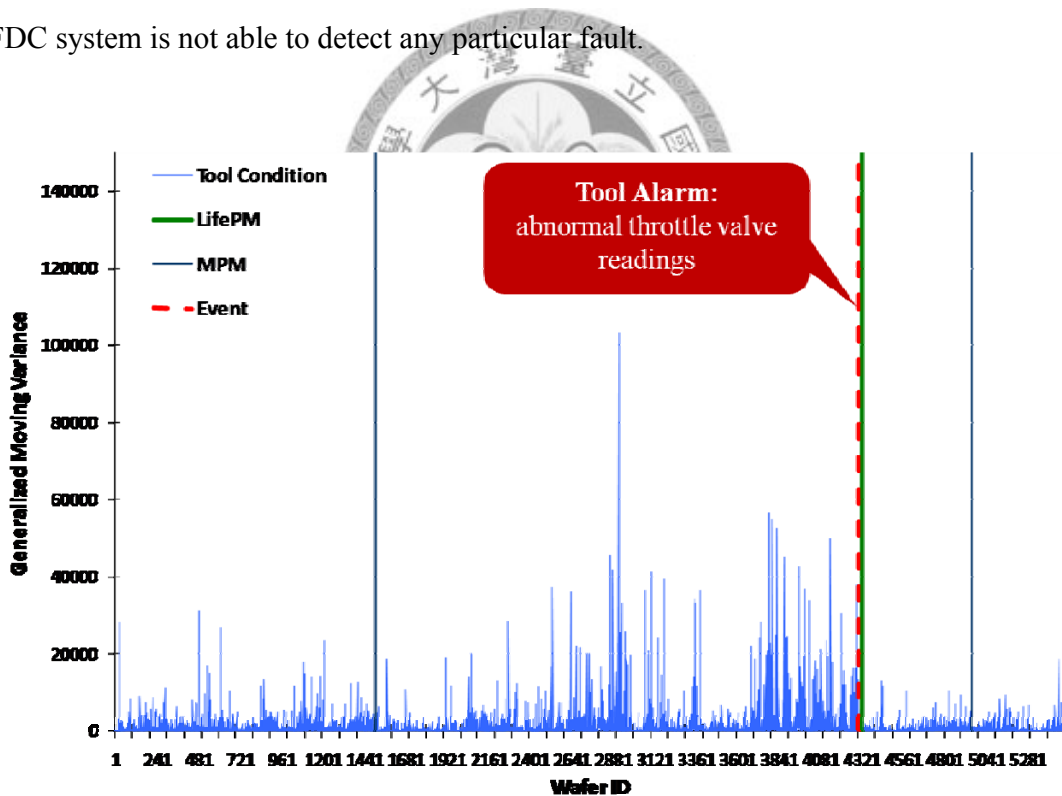


Fig. 17. PECVD Tool condition monitoring based on generalized moving variances.

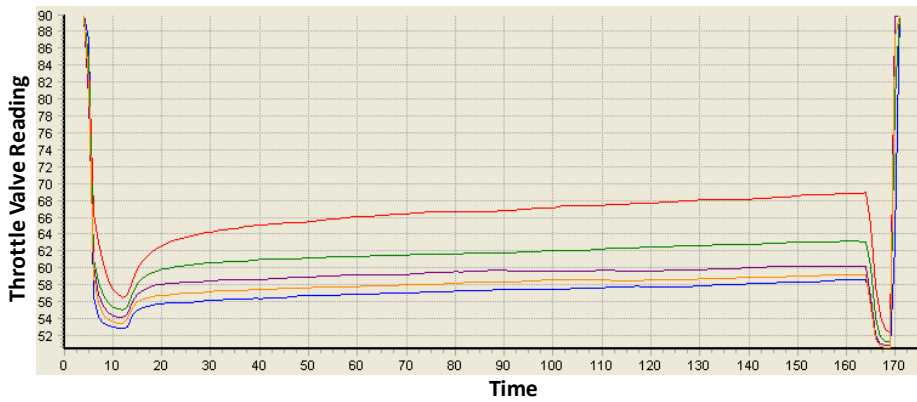


Fig. 18. Abnormal throttle valve readings when tool event is issued.

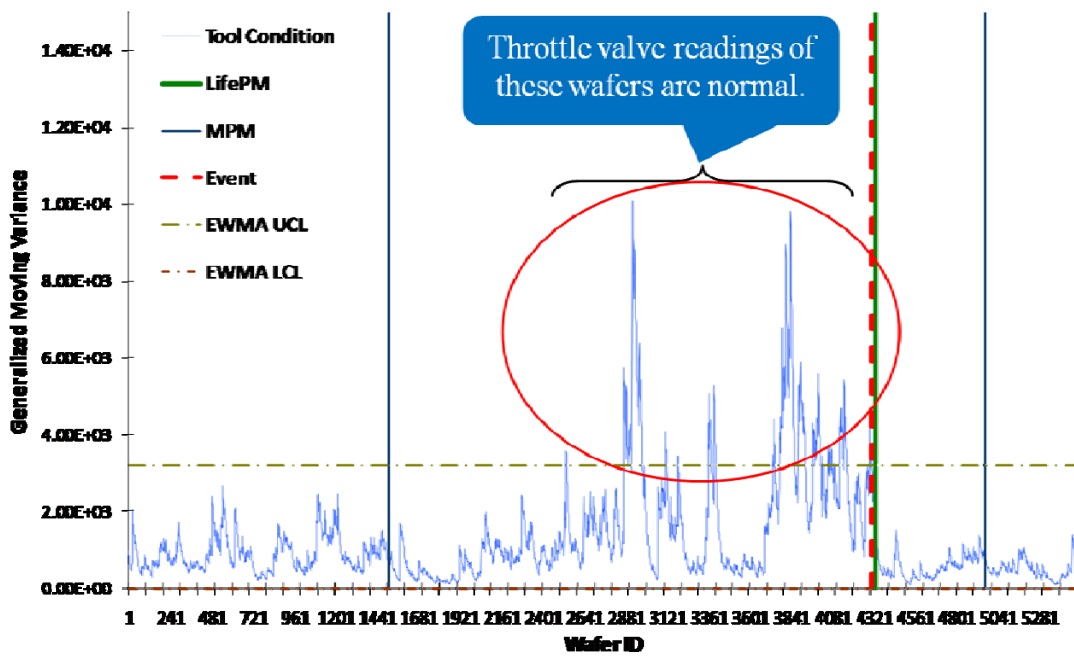


Fig. 19. EWMA control chart of the PECVD tool condition ( $\lambda=0.05$ ,  $l=4$ ).

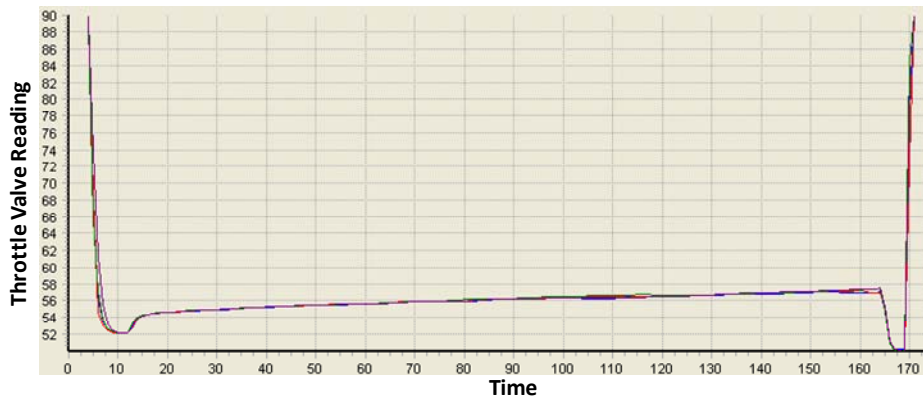


Fig. 20. Sampled throttle valve readings of 2500<sup>th</sup>~4300<sup>th</sup> wafers before the tool event in Fig. 19.

To analyze the possible tool problem behind the tool condition alarm, the proposed two-step diagnosis method is used. First, the generalized moving variances are decomposed into two parts: SVID variability and SVID interrelation, which are shown in Fig. 21 and Fig. 22, respectively. It can be seen that the tool condition excursion is mainly caused by the excursion in the SVID interrelation while the SVID variability appears to be quite normal.

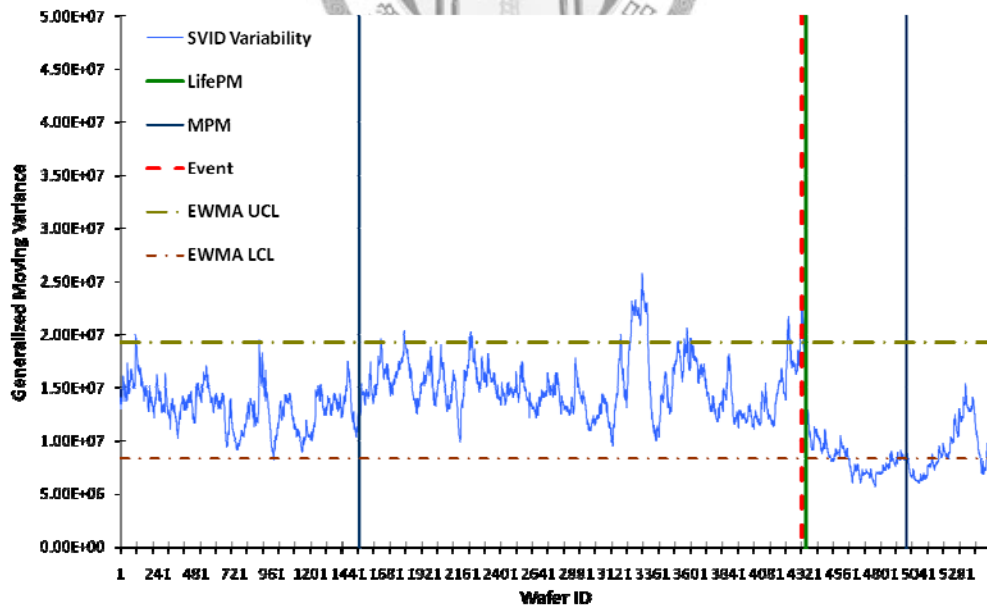


Fig. 21. EWMA control chart for the decomposition of PECVD's generalized moving variances: SVID Variability

( $\lambda=0.05, l=4$ ).



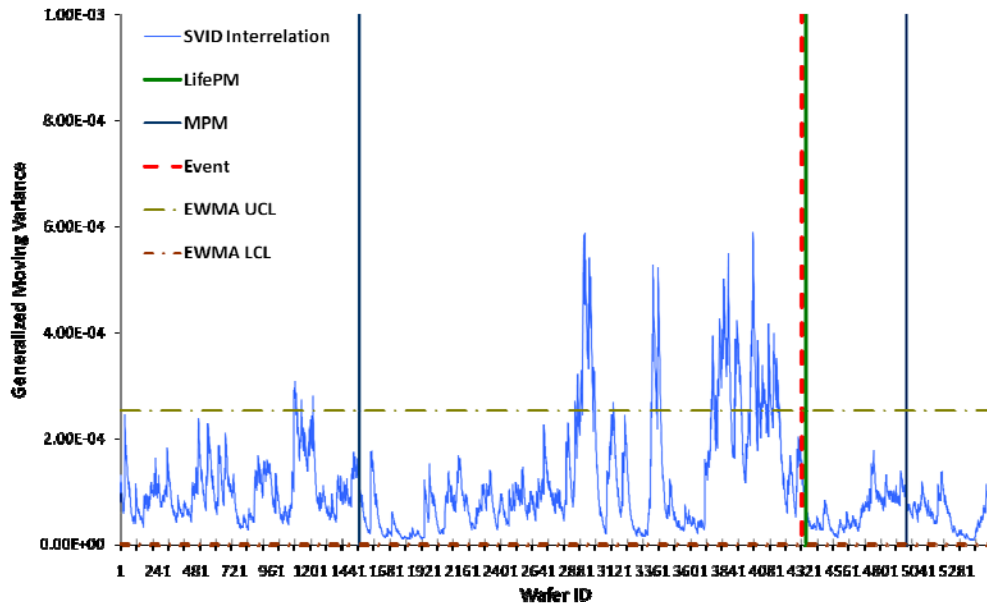


Fig. 22. EWMA control chart for the decomposition of PECVD's generalized moving variances: SVID Interrelation ( $\lambda=0.05, l=4$ ).

Since the tool condition excursion results from the SVID Interrelation, the trend of moving covariance of the small moving windows within each wafer processing run is then analyzed. In the process of the PECVD tool, there are three major processing steps: ramp-up, deposition, and ramp-down. We first check and draw the trend charts of the moving-window covariances, i.e. the  $S_{W_j^X W_j^Y}$  in equation (5), between throttle-valve and other SVID's. Some excursions of the relationship between gas flow and throttle valve are found. The upper trend chart in Fig. 23 shows the moving-window covariances between gas flow and throttle valve during the ramp-up step while the bottom one is of the ramp-down step. As shown in Fig. 23, the relationship between the gas flow rate and the throttle valve reading has been changed significantly during the ramp-up and ramp-down steps. The SVID profiles of throttle valve and gas flow are further investigated and thus the abnormal relationship change between the two SVID's is confirmed.

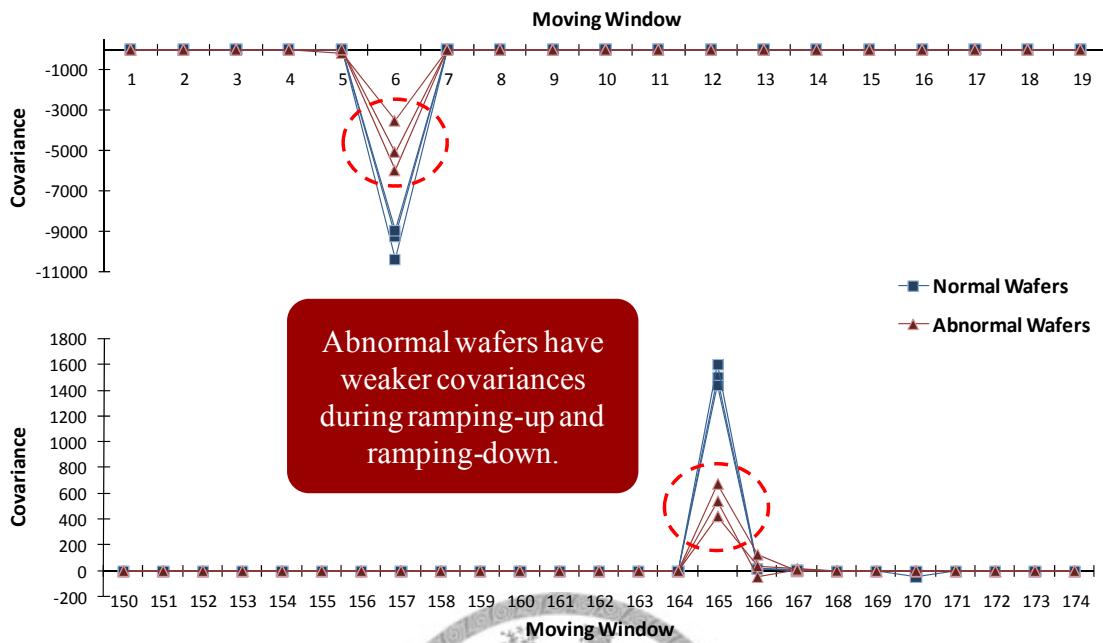


Fig. 23. Abnormal changes of relationship between gas flow and throttle valve during ramp-up and ramp-down steps.

As a result, the cause to the tool alarm is found to be an abnormal relationship between the throttle valve reading and the gas flow rate. The abnormal relationship between the two SVID's is detected by the tool condition indicator long before the tool event. Though this tool event is recovered by a regular Life PM, our study shows that the PM could have been performed earlier to minimize the undesired process output.

### 5.1.2 The PVD Tool Case

Another FDC data of more than 40,000 pieces of wafers with 9 SVID's collected from a PVD tool is analyzed as well. The sampling rate is also down to one observation a second. Similarly, the proposed indicator of normal tool condition is quite robust against the recipe changes in Fig. 24. Following the same procedure, the PVD tool condition is calculated and monitored in Fig. 25.

By looking at the EWMA control chart in Fig. 26, there were some tool alarms occurred after the Life PM, indicated by the vertical bold line. The tool condition appeared extremely unstable as compared to the tool condition before the PM. However, the tool FDC system still cannot detect any particular fault during this tool alarm period. After a regular Quarterly PM (QPM), the tool condition then became stable again. Therefore, the two-step diagnosis method is applied again. As can be seen in Fig. 27 and Fig. 28, there're much more excursions in the SVID variability as compared to that in the SVID interrelation. Then, individual profiles of the 9 SVID's are investigated. As shown in Fig. 29, the nano-torr readings of the wafers in-between the Life PM and QPM display unstable trends (dashed lines) after being processed for 13 seconds as compared to the solid lines which indicates nano-torr readings of the wafers after the QPM.

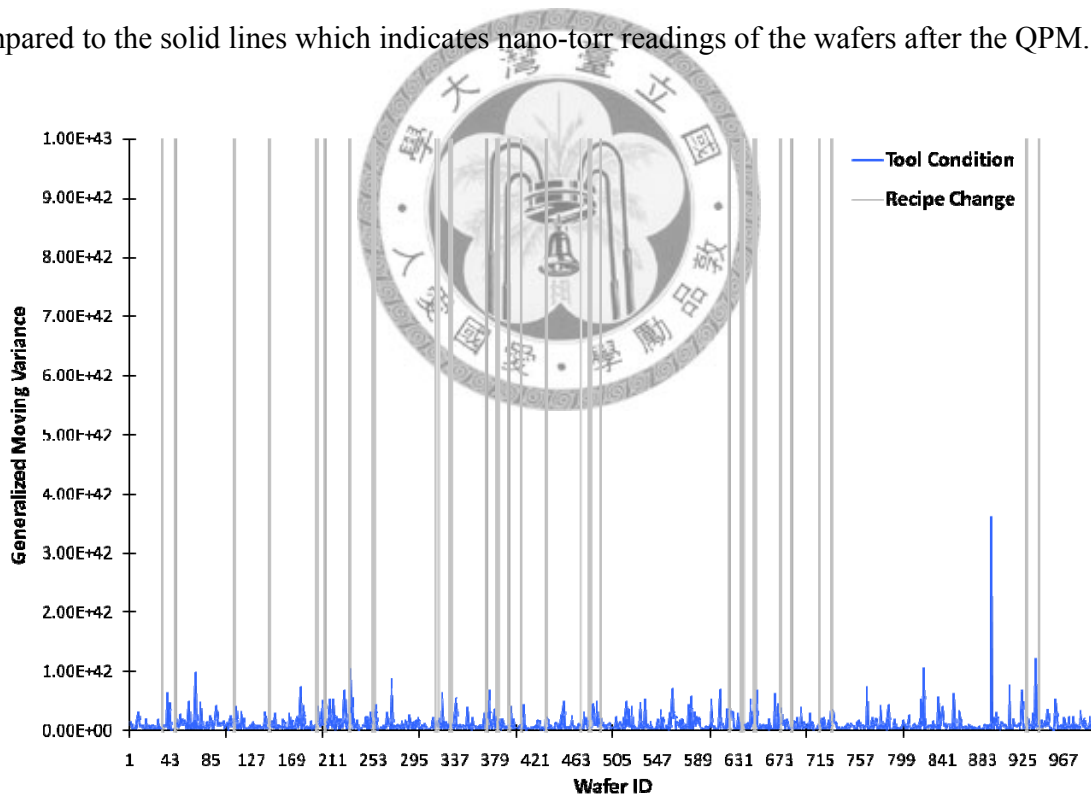


Fig. 24. PVD condition indicator vs. recipe changes.

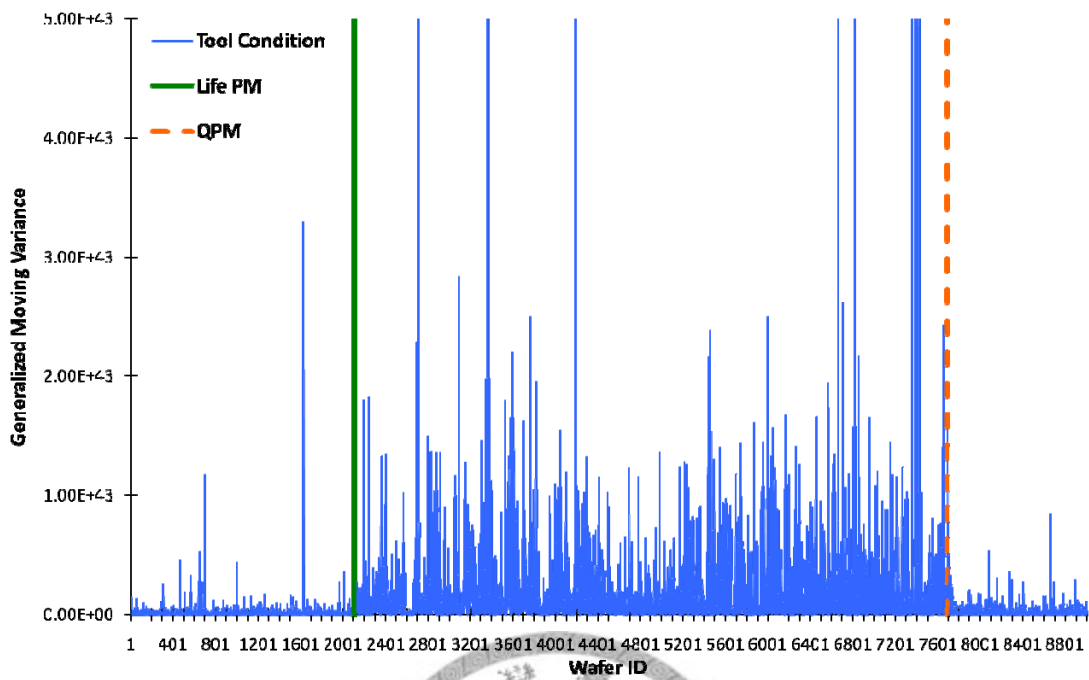


Fig. 25. PVD tool condition monitoring based on generalized moving variances; abnormal tool condition occurred after a Life PM was applied.

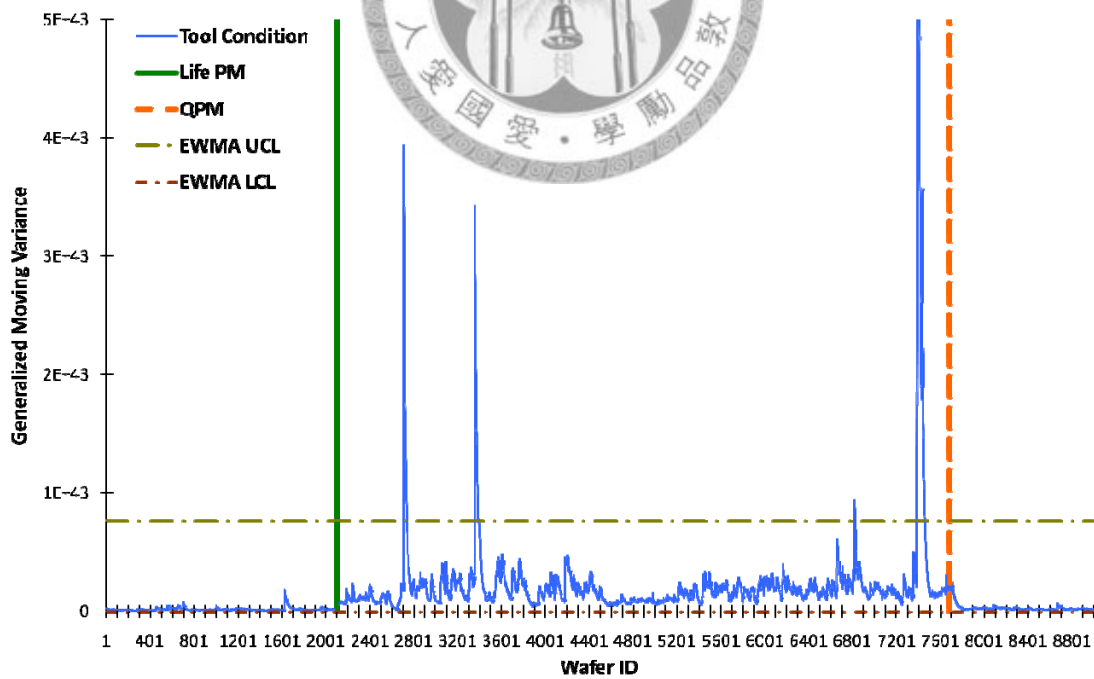


Fig. 26. EWMA control chart for the PVD tool condition ( $\lambda=0.05, l=4$ ).

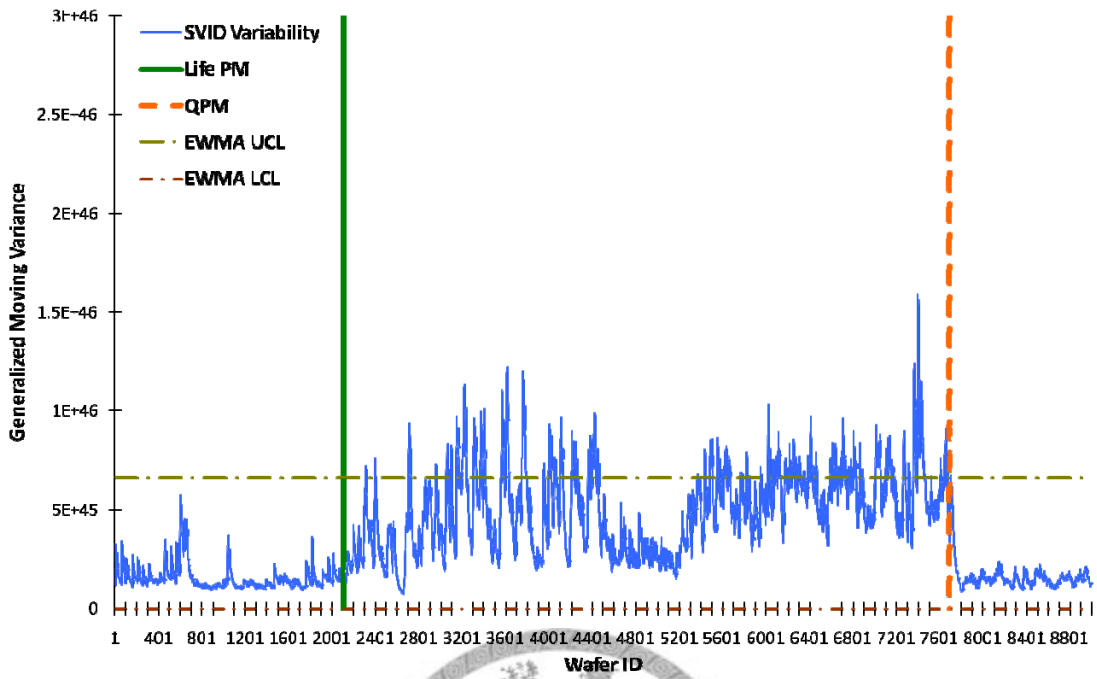


Fig. 27. EWMA control chart for the decomposition of PVD's generalized moving variances: SVID Variability

( $\lambda=0.05, l=4$ ).

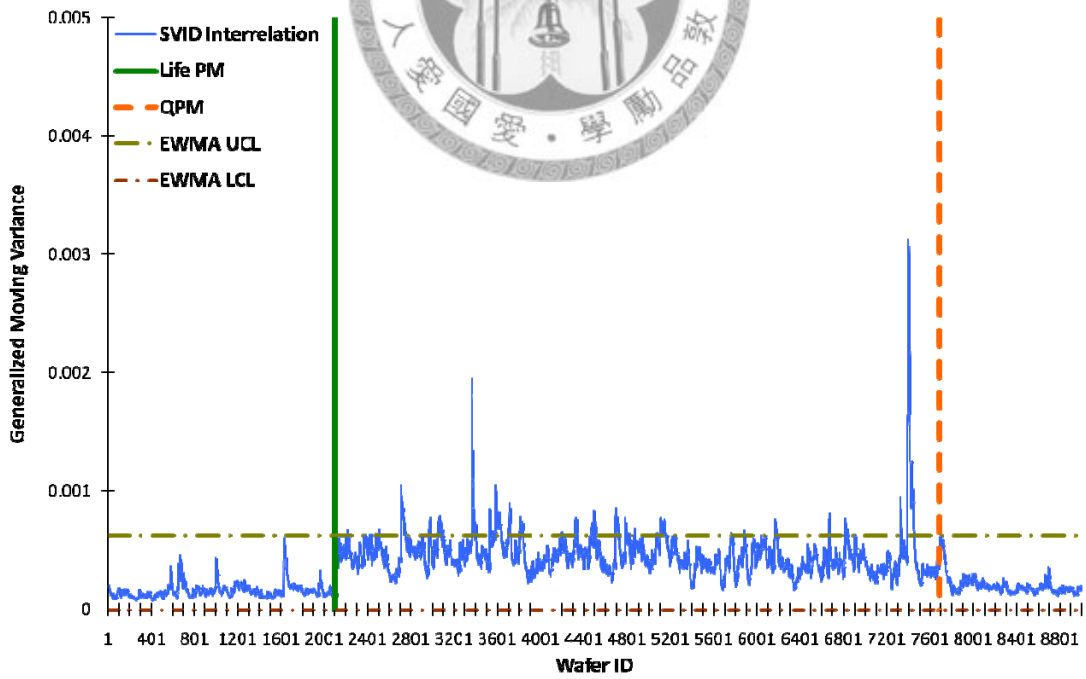


Fig. 28. EWMA control chart for the decomposition of PVD's generalized moving variances: SVID Interrelation

( $\lambda=0.05, l=4$ ).

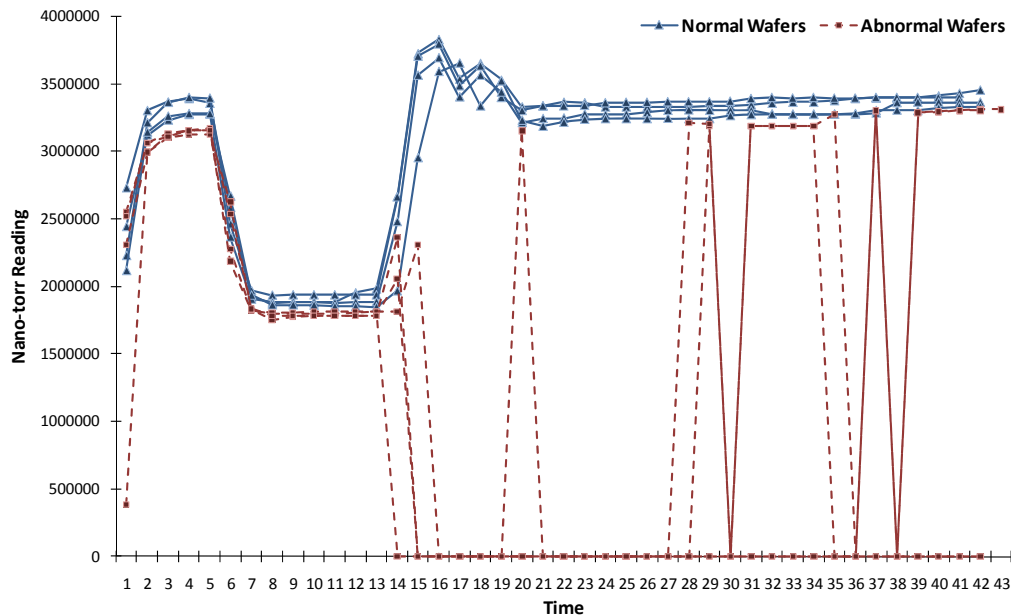


Fig. 29. Abnormal SVID profiles of nano-torr readings vs. normal ones.

In this case, the tool condition excursion resulted from an improper PM is detected by the proposed indicator. After performing the regular QPM, the nano-torr readings return to normal and the tool condition indicator behaves accordingly.

## 5.2 Real Wafer Topography for the Characterization of Spatial Variation

To validate the proposed methodology which identifies systematic patterns resulted from systematic variation within the spatial variation of wafer topography, hypothetical wafer data are generated and two sets of real wafer CD metrology data are also collected and analyzed.

### 5.2.1 Hypothetical Wafer Topographies with Common Patterns

Firstly, we simulate four common wafer topography patterns with noise disturbance: x-direction drift, y direction drift, dome, and donut (see Appendix D), to validate the proposed methodology. The contour maps for the five wafer topographies (including the random noise) are plotted in Fig. 30 and the spatial variation spectra are calculated as well. As can be seen in Fig. 30, the x- and y-

direction drifting patterns share very similar spectra where the slight difference is resulted from the random variation. Theoretically, a rotated pattern wouldn't change all the sample variances within spatial moving windows given a window size  $p$ . Therefore, the spatial moving variances, i.e. the spectra, of the two patterns are similar. The spectrum of the fourth systematic pattern, the donut, reaches a stably high level of variation of mid-to-low frequencies because the rise and fall of the donut pattern almost covers the whole wafer. Only the small-size spatial moving windows contain nearly-equal measurements and thus result in small sample variances.

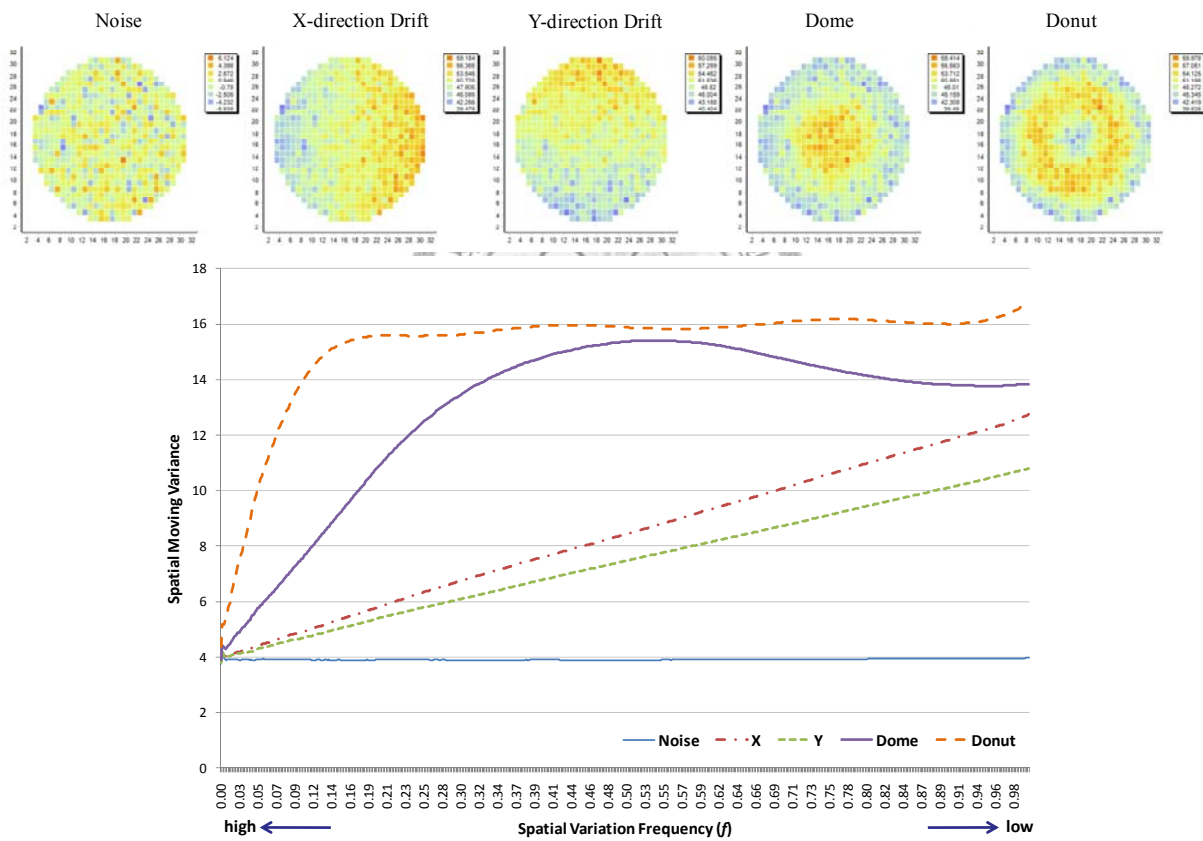


Fig. 30. Analysis of  $SVS$ 's for the hypothetical wafer topographies.

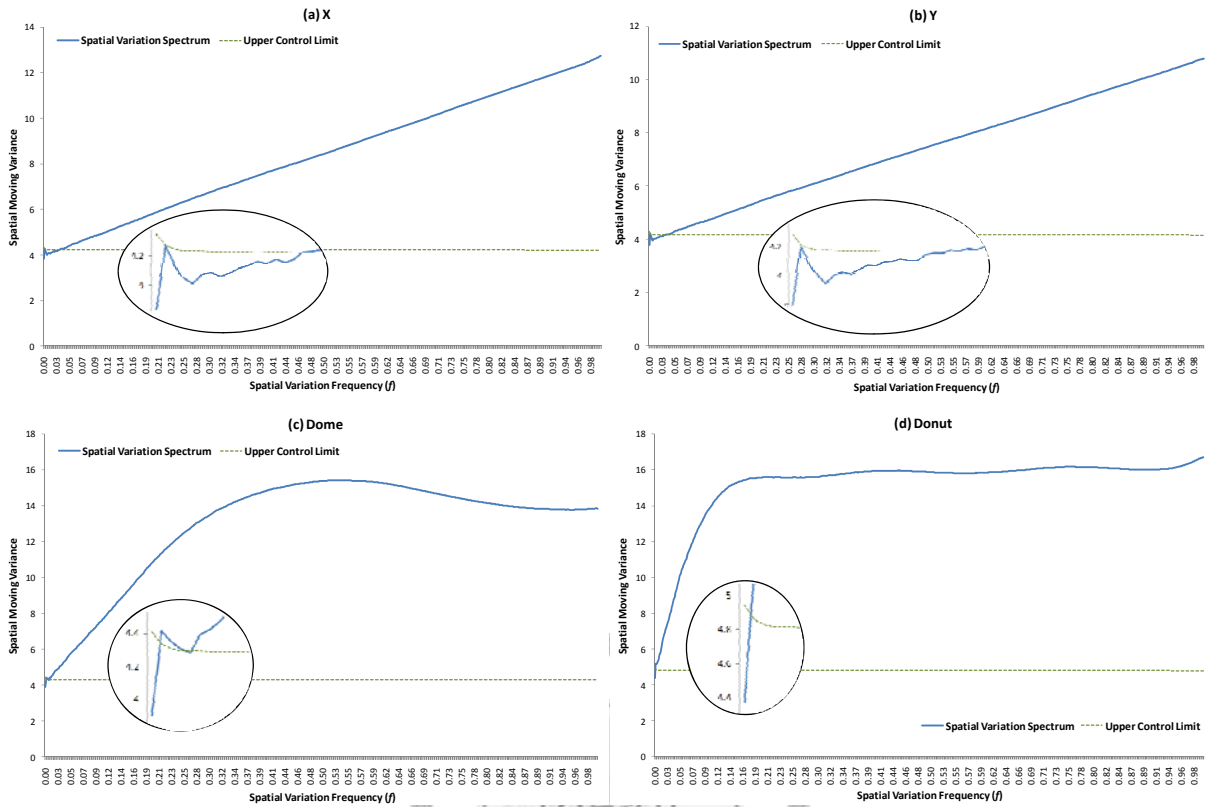


Fig. 31. The upper control limit for the *SVS*'s of the hypothetical wafer topographies. (The round shapes magnify the high-frequency parts of the spectra.)

To test if the four spectra are significantly distinct from that of a randomly distributed topography, the upper control limit described in 4.1.3 is built to identify the existence of systematic variations. As can be seen in Fig. 31, the four spectra go out of the control limits quickly as  $f$  increases. Therefore, the spatial variations of the four hypothetical topographies are said to contain not only random but systematic variations.

The spectra are further summarized into the *SPI*'s and the variation ratios in Table II. The four topographies with systematic patterns all have significantly-high indices as compared to that of pure random variation. In particular, the low-frequency variation ratios of x- and y-direction drifting patterns are very high because the drifting pattern actually covers up the whole wafer.



Table II. *SPI* and variation ratios for the hypothetical wafer topographies.

	Noise	X	Y	Dome	Donut
<b>Spatial Pattern Index (<i>SPI</i>)</b>	0.0372	0.5420	0.4882	0.6978	0.7130
<b>High-freq. Variation Ratio</b>	33.11%	21.78%	23.03%	24.60%	29.88%
<b>Mid-freq. Variation Ratio</b>	33.23%	33.39%	33.40%	38.82%	34.70%
<b>Low-freq. Variation Ratio</b>	33.67%	44.84%	43.57%	36.58%	35.42%

### 5.2.2 Wafer Metrology Data under Different Precision Levels

Real wafer metrology data from a local fab company are also collected for the methodology validation. Wafers after the post-exposure bake (PEB) process are usually measured by optical critical dimension (OCD) and scanning electron microscope (SEM), respectively. Although the measurement precision of OCD is better than SEM for the 65nm technology node, the recipe and library generation of OCD is more complicated and time consuming. Ke *et al.* (2005) proposed a concept on OCD-like CD SEM measurement which is said to be the average line width (ALW) or contact hole diameter (ACD) measurement at high resolution and low magnification CD SEM. These measurements, i.e. SEM CD, ALW/ACD, and OCD, manifest different significances of systematic variation of the same wafer due to the measuring precision. Therefore, the proposed *SVS* and *SPI* would be an appropriate way to analyze the systematic variations of wafers, and thus examine the precision of the measurements.

The spatial variation spectra of the three kinds of measurements and the control limits for the identification of the systematic variation of the spectra are calculated and plotted in Fig. 32 and Fig. 33. Then, the *SPI*'s and variation ratios are summarized in Table III. As shown in Fig. 33, the overall variation for the less-precise measurements, SEM CD, is much higher than those of ACD and OCD. This is because SEM CD consists of much noise, and thus the spatial pattern index ( $SPI = 0.3384$ ) is much lower than the other two. The OCD which provides measurements with high precision exhibits a clear pattern on the contour map and a very high *SPI* ( $= 0.9465$ ).

The phenomenon that the high-frequency spatial moving variances of OCD are smaller than that of ACD explains the ACD actually consists of more random noise than OCD. The high frequency variation ratios which decrease from SEM CD to OCD as the measurements are more precise also tell the same story. Moreover, a larger low-frequency variation ratio, as compared to that of the high or middle frequencies, indicates a clearer pattern with larger coverage of the wafer.

As a result, the *SVS* and its summarized indices can be used to examine the precision levels of different measuring methods for the wafer topography. The systematic and random variations of a wafer can be truthfully captured as the precision of measuring methods changes.

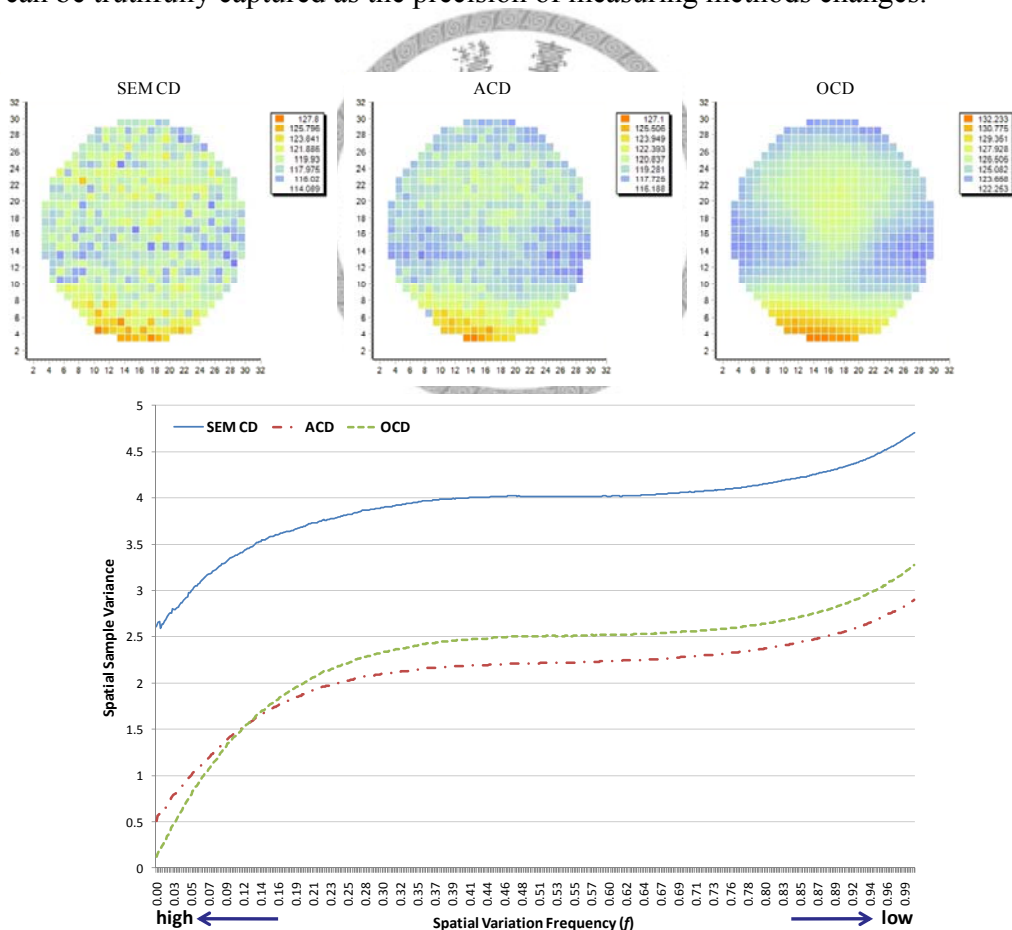


Fig. 32. Analysis of spatial variation spectra for the real wafer metrology data under different precision levels.

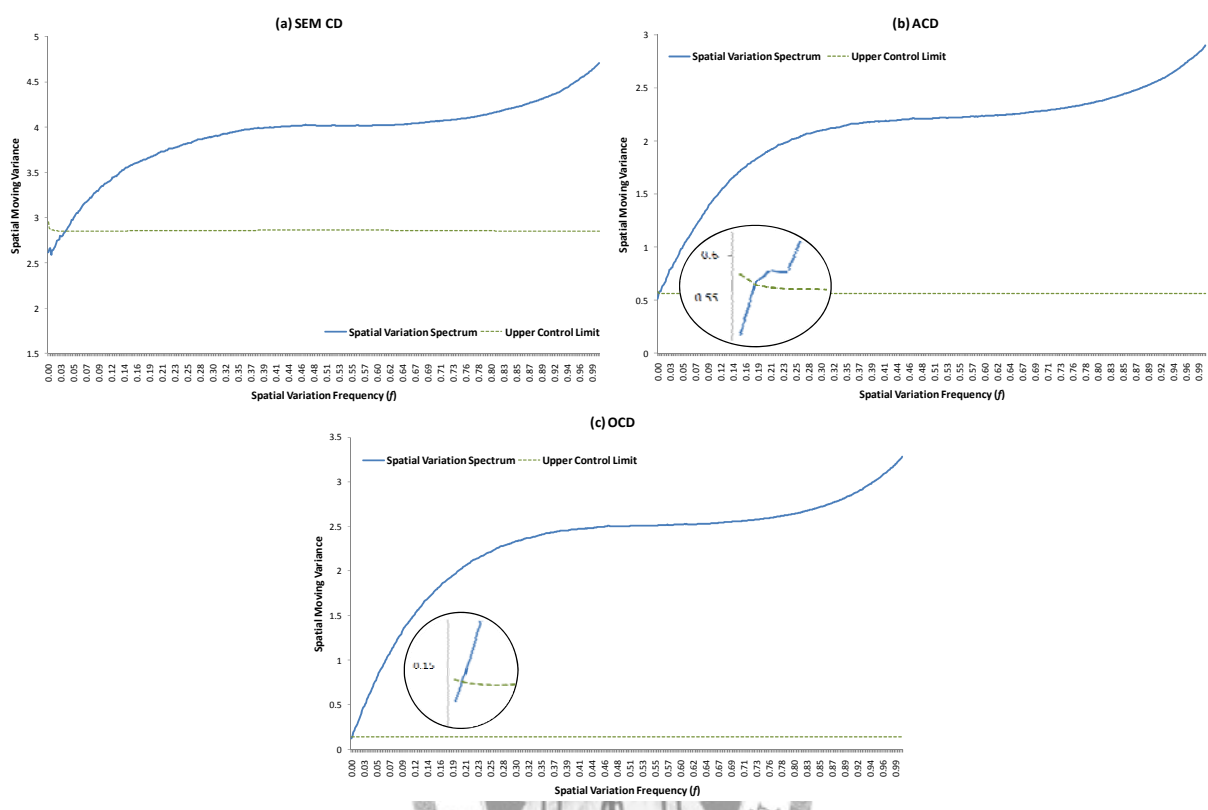


Fig. 33. The identification of systematic variations for the real wafer metrology data under different precision levels.

(The round shapes magnify the high-frequency parts of the spectra.)

Table III. *SPI* and variation ratios of real wafer metrology data under different precision levels.

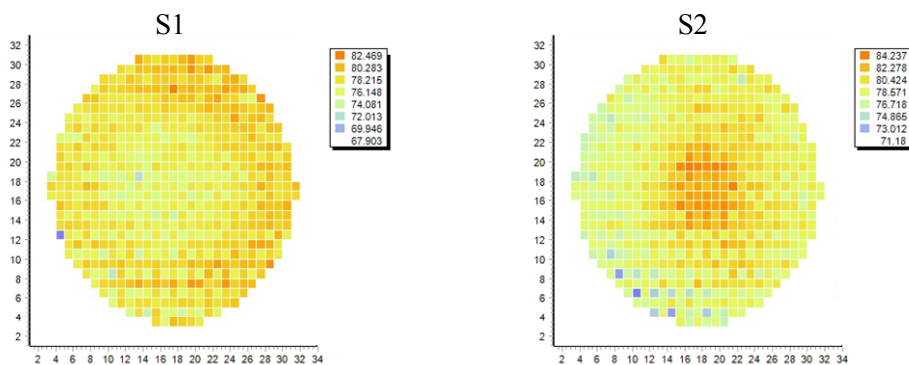
	SEM CD	ACD	OCD
<b>Spatial Pattern Index (<i>SPI</i>)</b>	0.3384	0.7581	0.9465
<b>High-freq. Variation Ratio</b>	29.55%	25.60%	23.81%
<b>Mid-freq. Variation Ratio</b>	34.20%	35.16%	36.17%
<b>Low-freq. Variation Ratio</b>	36.25%	39.24%	40.02%

### 5.2.3 Wafer Metrology Data with Checkerboard Pattern

Another set of wafer metrology data, CD's from the step-and-scan system, are also analyzed. Checkerboard patterns are sometimes formed after the scanning process because the scanning direction of the tool which scans downward when exposing the odd-numbered chips (or fields) and upward while performing exposures on even-numbered ones (A. Wong *et al.*, 2002). As can

be seen in Fig. 34, the two wafers not only consist of systematic patterns such as bowl (S1) and dome (S2), but also contain checkerboard patterns. The spectra of the two wafers are calculated and tested by the upper control limits in Fig. 35. The spatial pattern indices and variation ratios are listed in Table IV.

The alternative scanning effect of checkerboard pattern is considered to be characterized by the high-frequency spatial moving variances because the small-size spatial moving windows would cover the observations with alternative effect and show large sample variances. As can be seen in Fig. 35, the  $SVS$  starts higher than the upper control limit when the spatial variation frequency is very high ( $p=2$ ), and immediately goes lower than the control limit as the frequency becomes lower ( $f$  increases). However, it again goes out of the control limit because the bowl or dome pattern soon takes over and impacts on the spectrum. A checkerboard pattern behaves more similar to a randomly distributed topography than the bowl/dome pattern does will distort the systematic patterns and result in lower  $SPI$ 's (0.3 and 0.5244 in this case). However, it still can be identified by the high-frequency spatial moving variances through the proposed hypothesis test.



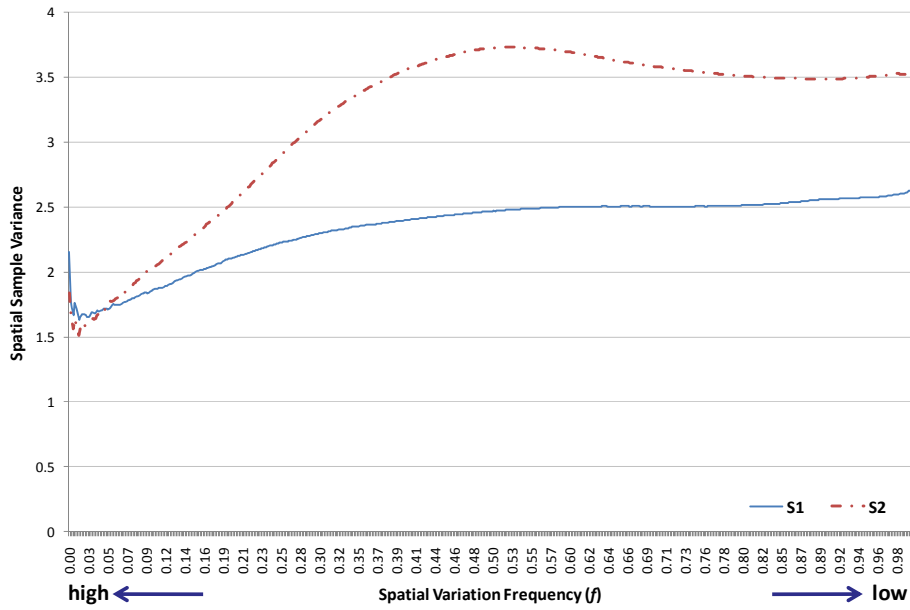


Fig. 34. Analysis of spatial variation spectra for the real wafer topographies with checkerboard patterns.

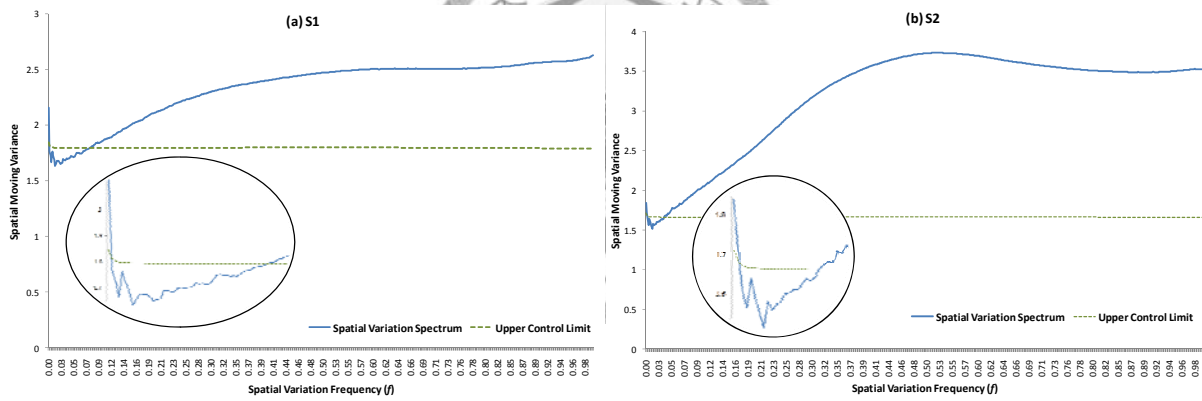


Fig. 35. The identification of systematic variations for the real wafer topographies with checkerboard patterns. (The round shapes magnify the high-frequency parts of the spectra.)

Table IV. *SPI* and variation ratios of real wafer topographies with checkerboard patterns.

	S1	S2
<b>Spatial Pattern Index (<i>SPI</i>)</b>	0.3000	0.5244
<b>High-freq. Variation Ratio</b>	28.60%	24.97%
<b>Mid-freq. Variation Ratio</b>	35.09%	38.08%
<b>Low-freq. Variation Ratio</b>	36.31%	36.95%

## Chapter 6 - Conclusion

### 6.1 Preliminary Results

A novel estimation to data variation, i.e. the moving variance calculation, for temporal profiles and spatial topographies is proposed and applied, in particular, to the semiconductor engineering data in this research. The developed moving variance technique is utilized to be the basis of the estimation methods. With truthful variation estimation, methodologies for the monitor of tool condition and the characterization of systematic variation are developed and validated.

Firstly, the calculation of moving variance and covariance to estimate the variability underlying multiple SVIDs' temporal profiles during a process run is developed. Then, the generalized moving variance is shown by both theoretical derivations and examples to be quite robust against the recipe changes. However, its insensitivity to pattern shifts and drifts of the temporal profiles also makes it incapable of detecting specific tool faults. The proposed indicator is thus a very good complementary to the existing FDC system to serve as an overall tool condition indicator. In this research, an EWMA control scheme to monitor the tool condition indicator is employed. A two-step diagnosis method is proposed to drill down the tool condition into two individual SVID variability and SVID interrelations. In 5.1.1, the change in the relationship between the gas flow and the throttle valve reading is found by the proposed tool condition diagnosis to cause the abnormal PECVD condition, which eventually leads to a tool event and a corresponding Life PM. In 5.1.2, the PVD nano-torr's abnormal reading is also correctly detected by the indicator and disappears after Quarterly PM. Both cases show all the detected abnormalities are due-to PM-related issues and the condition-based predictive maintenance could have been performed earlier.

For the analysis of topographical spatial variation, a similar idea of calculating the moving variance for a temporal series of observations to reduce the estimate bias may be used for the spatial data. Instead of the moving variance of a temporal series, the sample variance is calculated for observations from a spatial area. A model-free spatial variance spectrum (*SVS*) to analyze the spatial metrology data is proposed so that any engineering problems and issues can be learned as early as possible through identification of systematic patterns. The *SVS* is generated by the spatial moving variances with different spatial window sizes. A hypothesis test has been developed based on the statistical properties of *SVS* to detect the existence of systematic variations. Spatial pattern index (*SPI*) and three variation ratios are also developed to provide engineers with a quick look at the systematic variations. Both hypothetical and actual metrology cases are used to validate the proposed methodology. Results show that systematic patterns can be truthfully characterized by the proposed *SVS* even for metrology with low precision levels and for the unusual checkerboard patterns.

It is worth noting that the metrology sites sampled on a wafer during production may not be uniformly distributed. For the proposed spatial moving variances to work in this case, the moving window  $W_i^p$ , which consists of  $p$  random measurements including  $M_i$  and its  $p - 1$  neighboring measurements, should be now redefined as a circular area with the center at  $M_i$  and a radius  $r$ . Such an area-based moving window may cover indefinite numbers of observations. Let this area-based moving window be denoted as  $c_i^r$ . By increasing  $r$  from 0 to the diameter of the wafer,  $R$ , we can still calculate the spatial moving variances for different frequencies  $f$ , which is now calculated as  $r^2/R^2$ . Using  $c_i^r$  as the moving window is to uphold the meaning of the variance spectrum where the spatial variation for smaller  $r$ 's indicate the high-frequency

variations and are usually generated by the noises while larger  $r$ 's represent the low to mid range of frequency and are typically produced by systematic patterns. With a limited number of sampling sites (usually five or nine sites sampled per wafer) during production, the *SVS* resolution becomes lower and the spatial variance analysis may not be as effective. However, compared to conventional analysis based on just the overall sample variance, the *SVS* still reveals more information about the spatial variations and its corresponding pattern indices still provide a quick look at the significance of systematic patterns, but are, of course, less truthful due to the low sampling resolution.

## 6.2 Future Study

Despite the impact of recipe-change can be reduced and the systematic variation would be signified through applying the moving variance technique, pooling, i.e. taking the average of, the variances within moving windows, as calculated in (4), (5), and (14), will offset the high (or low) variances within minor moving windows such that the moving variance would be more insensitive to pattern changes. Sometimes, pattern-caused faults, such as a sudden ramp up/down in tool signals, signify critical tool abnormalities as well. As can be seen in the simulated example in Fig. 36, SVID  $X'$  shows an abnormal profile with irregular ramp rate as compared to the normal profile of SVID  $X$ . The temporal moving variances for SVID  $X$  and  $X'$  are 0.14 and 0.21, respectively, which do not vary a lot while the underlying variance of pure noise is 0.09.





Fig. 36. Two SVID profiles with different ramp rates.

Through looking into the moving-window variances as shown in Fig. 37, the change of ramp rate did result in a few distinguishable variances which, however, are offset by pooling the variances altogether. Therefore, the moving statistics proposed in this research perform to be quite insensitive to pattern shifts and drifts usually caused by tool recipe changes or tool faults.

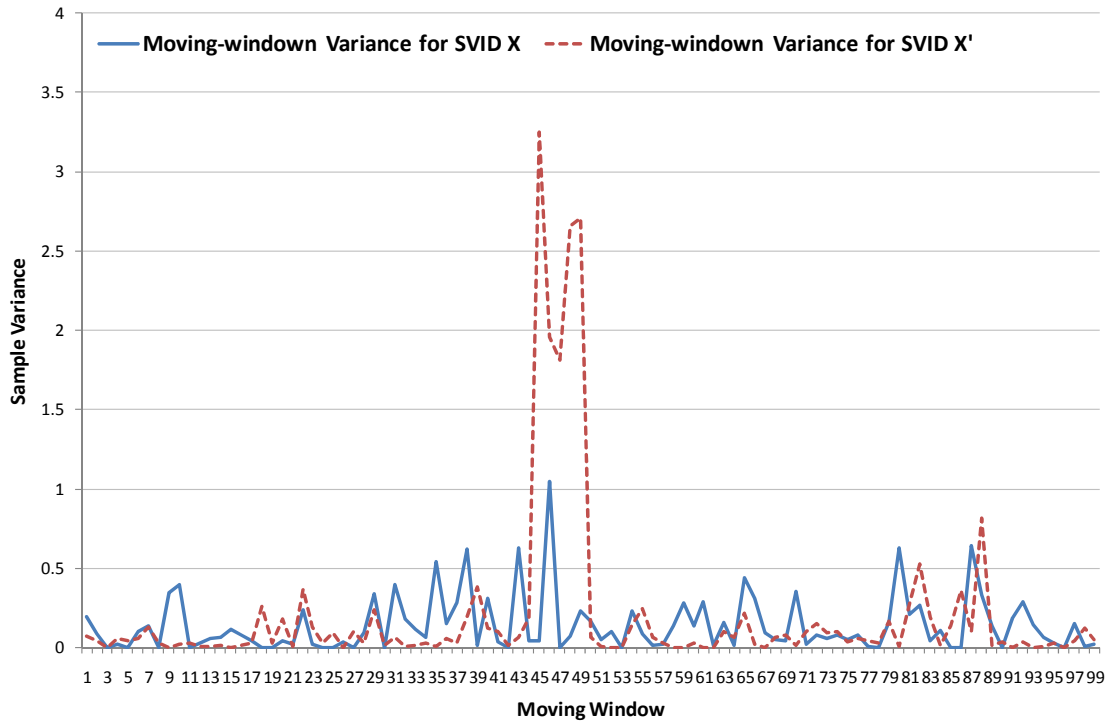


Fig. 37. Moving-window variances for SVID  $X$  and  $X'$ .

To detect abnormal patterns like DC shift or a change in ramp rate, conventional FDC practices select meaningful temporal-windows or process steps of SVID's and summarize significant statistics to be the indicators. By monitoring these indicators, the abnormal patterns of SVID profiles can be captured. However, the indicators based on the pattern modeling method are too problem-specific to represent a tool condition. Instead, the nature of variance and/or covariance is more like an indication of the process run stability.

To complement the insufficiency of moving variance, another moving statistic is proposed to justify the situation when the pattern shift or drift occurs. As compared to the calculation of moving variance described in (4) or (14), which takes the average of all moving-window variances, we calculate the variance of all moving-window variances for SVID  $X$  given the moving window size  $p$ , denoted as  $VoV_{p,X}$ .

$$VoV_{p,X} = \frac{1}{n-p} \sum_{j=1}^{n-p+1} (S_{W_j^X}^2 - \bar{S}_{W_j^X}^2)^2, \quad (27)$$

where  $\bar{S}_{W_j^X}^2$  is the average of all moving-window variances. Based on (27), the two variances of all moving-window variances for SVID  $X$  and  $X^2$ , denoted as  $VoV_{p,X}$  and  $VoV_{p,X^2}$ , are 0.03 and 0.31, which now become rather distinguishable.

For the spatial wafer topography, we can apply the same idea to extend the spatial variation spectrum as well. By calculating the  $VoV_p$ 's for  $2 \leq p \leq n$ , there will be another spectrum, defined as  $VoVS$ . Fig. 38 shows an example for the  $VoVS$  of the randomly distributed dataset used in Fig. 10 (a randomly distributed wafer topography). As can be seen, the spectrum quickly decreases to nearly 0 as the spatial variation frequency ( $f$ ) increases from 0. When the window size  $p=n$ , i.e.  $f=1$ , the  $VoV_n = 0$  because all the moving-window variances are the same and all equal to the conventional sample variance. It is also reasonable to have the largest  $VoV$  at the smallest moving window, i.e.  $p=2$ . As there are only two observations in each of the moving windows, the overlap between moving windows is very limited. Therefore, the moving-window variances of  $p=2$  would vary a lot and their variance becomes the largest.

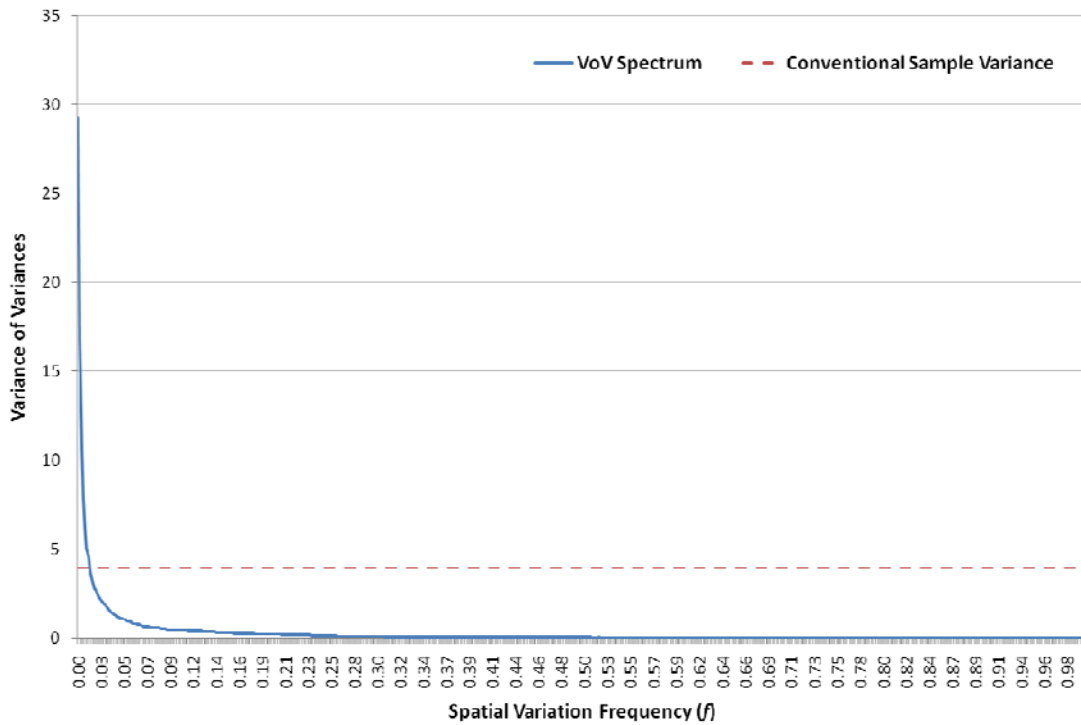


Fig. 38. *VoVS* for a randomly distributed topography.

As a result, the further study is to apply variance of moving-window variances to construct another variance/covariance matrix for the temporal tool parameters. The corresponding generalized statistic will be developed and its usefulness would be discussed. The properties of the *VoVS* calculated from wafer topography will be studied through mathematical treatments and simulated cases. Finally, we will investigate and argue if there exists any complementary property between the temporal/spatial moving variance and the variance of moving-window variances in the applications of temporal and spatial datasets.



## Appendix

### A. Proof for THEOREM 1

Since the expected value of the sample variance under the linear-drift pattern (6):

$$\begin{aligned}
 E(S_X^2) &= E\left[\frac{n\sum X_i^2 - (\sum X_i)^2}{n(n-1)}\right] = \frac{n\sum E(X_i^2) - E[(\sum X_i)^2]}{n(n-1)} \\
 &= \frac{a^2\sigma_X^2\sum i^2 + \sum E(\varepsilon_{X,i}^2)}{n-1} - \frac{a^2\sigma_X^2(\sum i)^2 + E[(\sum \varepsilon_{X,i})^2]}{n(n-1)} \\
 &= a^2\sigma_X^2\frac{n\sum i^2 - (\sum i)^2}{n(n-1)} + \sigma_X^2 = \frac{n(n+1)a^2\sigma_X^2}{12} + \sigma_X^2, \tag{28}
 \end{aligned}$$

the estimate bias size by the conventional sample variance can be obtained as:

$$E(S_X^2) - \sigma_X^2 = n(n+1)a^2\sigma_X^2 / 12.$$

Following (28), it can be also shown that:

$$E(S_{X,j}^2) = p(p+1)a^2\sigma_X^2 / 12 + \sigma_X^2 \quad \forall j = 1, \dots, n-p+1$$

and thus

$$\begin{aligned}
 E(\hat{\sigma}_X^2) - \sigma_X^2 &= \frac{1}{n-p+1} \sum_{j=1}^{n-p+1} E(S_{X,j}^2) - \sigma_X^2 \\
 &= p(p+1)a^2\sigma_X^2 / 12. \tag{29}
 \end{aligned}$$

Therefore,

$$E(\hat{\sigma}_X^2) - \sigma_X^2 < E(S_X^2) - \sigma_X^2$$

for a SVID  $X$  moving window size  $p < n$ . Following the same procedure above, it can be easily proved that

$$E(\hat{\sigma}_Y^2) - \sigma_Y^2 < E(S_Y^2) - \sigma_Y^2$$

for a SVID  $Y$  moving window size  $p < n$ .

Similarly, because

$$\begin{aligned} & E(S_{XY}) \\ &= E\left[\frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n(n-1)}\right] = \frac{n \sum E(X_i Y_i) - E[(\sum X_i \sum Y_i)]}{n(n-1)} \\ &= \frac{ab\sigma_X\sigma_Y \sum i^2 + \sum E(\varepsilon_{X,i}\varepsilon_{Y,i})}{n-1} - \frac{ab\sigma_X\sigma_Y (\sum i)^2 + E[(\sum \varepsilon_{X,i} \sum \varepsilon_{Y,i})]}{n(n-1)} \\ &= ab\sigma_X\sigma_Y \frac{n \sum i^2 - (\sum i)^2}{n(n-1)} + \sigma_{XY} = \frac{n(n+1)ab\sigma_X\sigma_Y}{12} + \sigma_{XY}, \end{aligned} \quad (30)$$

the estimate bias size by the conventional sample covariance becomes:

$$E(S_{XY}) - \sigma_{XY} = \frac{n(n+1)ab\sigma_X\sigma_Y}{12}.$$

Following (30), it can be shown that:

$$E(S_{XY,j}) = p(p+1)ab\sigma_X\sigma_Y / 12 + \sigma_{XY} \quad \forall j = 1, \dots, n-p+1$$

and thus

$$\begin{aligned} E(\hat{\sigma}_{XY}) - \sigma_{XY} &= \frac{1}{n-p+1} \sum_{j=1}^{n-p+1} E(S_{XY,j}) - \sigma_{XY} \\ &= p(p+1)ab\sigma_X\sigma_Y / 12. \end{aligned} \quad (31)$$

Since  $a \neq 0$ ;  $b \neq 0$  and  $p < n$ , we have

$$|E(\hat{\sigma}_{XY}) - \sigma_{XY}| < |E(S_{XY}) - \sigma_{XY}|.$$

□

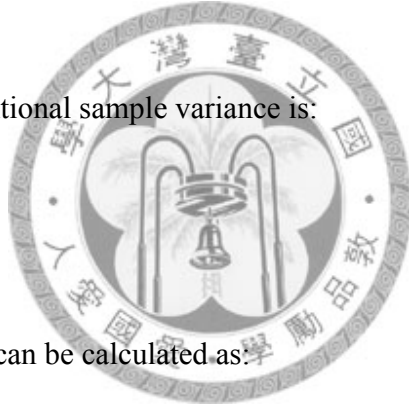
## B. Proof for THEOREM 2

Since the expected value of the sample variance under the step-change pattern (8)

$$\begin{aligned} E(S_x^2) &= E\left[\frac{n \sum X_i^2 - (\sum X_i)^2}{n(n-1)}\right] = \frac{n \sum E(X_i^2) - E[(\sum X_i)^2]}{n(n-1)} \\ &= \frac{(n-n_1)c^2\sigma_x^2 + n\sigma_x^2}{n-1} - \frac{(n-n_1)^2 c^2\sigma_x^2 + n\sigma_x^2}{n(n-1)} \\ &= \frac{n_1(n-n_1)c^2\sigma_x^2}{n(n-1)} + \sigma_x^2, \end{aligned} \quad (32)$$

the estimate bias by the conventional sample variance is:

$$E(S_x^2) - \sigma_x^2 = \frac{n_1 n_2 c^2 \sigma_x^2}{n(n-1)}. \quad (33)$$



The expected moving variance can be calculated as:

$$\begin{aligned} E(\hat{\sigma}_x^2) &= \frac{1}{n-p+1} \sum_{j=1}^{n-p+1} E(S_{w_j^x}^2) \\ &= \frac{1}{n-p+1} \left[ \sum_{j=1}^{n_1-p+1} E(S_{w_j^x}^2) + \sum_{j=n_1-p+2}^{n_1} E(S_{w_j^x}^2) + \sum_{j=n_1+1}^{n-p+1} E(S_{w_j^x}^2) \right]. \end{aligned} \quad (34)$$

Because  $p \leq \min(n_1, n_2)$  where the moving windows from the 1<sup>st</sup> to the  $(n_1-p+1)^{\text{th}}$  and from the  $(n_1+1)^{\text{th}}$  to the  $(n-p+1)^{\text{th}}$  will not contain the step change while each of the  $(n_1-p+2)^{\text{th}}, \dots, n_1^{\text{th}}$  moving windows contains the step change. Therefore, following (34) we have



$$\begin{aligned}
& \frac{1}{n-p+1} \left[ \sum_{j=1}^{n_1-p+1} E(S_{W_j^x}^2) + \sum_{j=n_1-p+2}^{n_1} E(S_{W_j^x}^2) + \sum_{j=n_1+1}^{n-p+1} E(S_{W_j^x}^2) \right] \\
&= \frac{1}{n-p+1} \left[ \sum_{j=1}^{n_1-p+1} \sigma_x^2 + \sum_{j=n_1+1}^{n-p+1} \sigma_x^2 + \sum_{j=n_1-p+2}^{n_1} \left( \frac{(n_1-j+1)(p-n_1+j-1)c^2\sigma_x^2}{p(p-1)} + \sigma_x^2 \right) \right] \\
&= \sigma_x^2 + \frac{1}{n-p+1} \sum_{q=1}^{p-1} \frac{q(p-q)c^2\sigma_x^2}{p(p-1)} \\
&= \sigma_x^2 + \frac{p+1}{6(n-p+1)} c^2\sigma_x^2 \tag{35}
\end{aligned}$$

and thus

$$E(\hat{\sigma}_x^2) - \sigma_x^2 = \frac{p+1}{6(n-p+1)} c^2\sigma_x^2. \tag{36}$$

From the right-hand sides of (33) and (35), we obtain the following condition:

$$\frac{p+1}{6(n-p+1)} < \frac{n_1(n-n_1)}{n(n-1)}.$$



It can be easily seen that  $E(\hat{\sigma}_x^2) - \sigma_x^2 < E(S_x^2) - \sigma_x^2$ . Following the same procedure above,  $E(\hat{\sigma}_y^2) - \sigma_y^2 < E(S_y^2) - \sigma_y^2$  can be also proved. Similarly, because the expected value of the sample covariance under the step-change pattern (8)

$$\begin{aligned}
& E(S_{XY}) \\
&= E\left[ \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n(n-1)} \right] = \frac{n \sum E(X_i Y_i) - E[\sum X_i \sum Y_i]}{n(n-1)} \\
&= \frac{(n-n_1)cd\sigma_x\sigma_y + \sum E(\varepsilon_{X,i}\varepsilon_{Y,i})}{n-1} - \frac{(n-n_1)^2cd\sigma_x\sigma_y + E[(\sum \varepsilon_{X,i} \sum \varepsilon_{Y,i})]}{n(n-1)} \\
&= \frac{(nn_1 - n_1^2)}{n(n-1)} cd\sigma_x\sigma_y + \sigma_{XY} = \frac{n_1(n-n_1)cd\sigma_x\sigma_y}{n(n-1)} + \sigma_{XY}, \tag{37}
\end{aligned}$$

the estimate bias by the conventional sample covariance is:

$$E(S_{XY}) - \sigma_{XY} = \frac{n_1(n-n_1)cd\sigma_X\sigma_Y}{n(n-1)}.$$

Following (34), it can be shown that:

$$\begin{aligned}
E(\hat{\sigma}_{XY}) &= \frac{1}{n-p+1} \sum_{j=1}^{n-p+1} E(S_{W_j^X W_j^Y}) \\
&= \frac{1}{n-p+1} \left[ \sum_{j=1}^{n_1-p+1} E(S_{W_j^X W_j^Y}) + \sum_{j=n_1-p+2}^{n_1} E(S_{W_j^X W_j^Y}) + \sum_{j=n_1+1}^{n-p+1} E(S_{W_j^X W_j^Y}) \right] \\
&= \frac{1}{n-p+1} \left[ \sum_{j=1}^{n_1-p+1} \sigma_{XY} + \sum_{j=n_1+1}^{n-p+1} \sigma_{XY} + \sum_{j=n_1-p+2}^{n_1} \left( \frac{(n_1-j+1)(p-n_1+j-1)cd\sigma_X\sigma_Y}{p(p-1)} + \sigma_{XY} \right) \right] \\
&= \sigma_{XY} + \frac{1}{n-p+1} \sum_{q=1}^{p-1} \frac{q(p-q)cd\sigma_X\sigma_Y}{p(p-1)} \\
&= \sigma_{XY} + \frac{p+1}{6(n-p+1)} cd\sigma_X\sigma_Y
\end{aligned} \tag{38}$$

and thus

$$E(\hat{\sigma}_{XY}) - \sigma_{XY} = \frac{p+1}{6(n-p+1)} cd\sigma_X\sigma_Y. \tag{39}$$

Under the condition (9),  $c \neq 0$ ;  $d \neq 0$  and  $p < n$ , it can be easily proved that

$$|E(\hat{\sigma}_{XY}) - \sigma_{XY}| < |E(S_{XY}) - \sigma_{XY}|. \quad \square$$

### C. Proof for Corollary 1

Substitute  $p=2$  in (9),

$$\frac{1}{2(n-1)} < \frac{n_1(n-n_1)}{n(n-1)}.$$

Condition (9) becomes:

$$\frac{n - \sqrt{n(n-2)}}{2} < n_1 < \frac{n + \sqrt{n(n-2)}}{2}. \quad (40)$$

Because  $p=2 \leq \min(n_1, n_2)$ ,  $n$  would be at least equal to or larger than 3. It can be easily seen that (40) holds when  $n \geq 3$ . In addition, the biases in (36) and (39) can be shown monotonically increasing with  $p$  and are minimized as  $p=2$ .  $\square$

## D. The Simulation of Hypothetical Wafer Topography

To simulate the wafer topography, a set of spatial coordinate must be generated. Firstly, we define the coordinates ranged from -135 to 135mm with a 10mm interval for a 300mm wafer and get  $28 \times 28 = 784$  grid points. After screening out the points which have radii larger than 140mm, 616 points are left and settled to be the metrology sites for generating observations. The topographies contain x- and y-direction drifting patterns, denoted as  $m_i(X)$  and  $m_i(Y)$ , can be generated as:

$$m_i(X) = a + \frac{m_{x,i}}{s} + \varepsilon_i, \text{ and} \quad (41)$$

$$m_i(Y) = a + \frac{m_{y,i}}{s} + \varepsilon_i \text{ for } i=1, \dots, 616, \quad (42)$$

where  $m_{x,i}$  and  $m_{y,i}$  are the generated coordinates,  $a$  is a constant,  $s$  is a scalar, and  $\varepsilon_i$ , for  $i=1, \dots, 616$ , are independent and follow an identical normal distribution with zero mean and variance  $\sigma^2$ .

To generate the dome and donut patterns, cosine function is employed and the radii for all the 616 points are calculated:

$$r_i = \sqrt{(m_{x,i})^2 + (m_{y,i})^2} \text{ for } i=1, \dots, 616. \quad (43)$$

The topographies with dome and donut patterns, denoted as  $m_i(Dome)$  and  $m_i(Donut)$ , can be then generated as:

$$m_i(Dome) = a + k \cdot \cos\left(\left[\frac{r_i - \min(r_i)}{\max(r_i) - \min(r_i)}\right] \cdot \pi\right) + \varepsilon_i, \text{ and} \quad (44)$$

$$m_i(Donut) = a - k \cdot \cos\left(\left[\frac{r_i - \min(r_i)}{\max(r_i) - \min(r_i)}\right] \cdot 2\pi\right) + \varepsilon_i, \quad (45)$$

for  $i=1, \dots, 616$ ,

where  $k$  denotes a multiplier which magnifies/minifies the dome and donut pattern, and  $\max(r_i)$  and  $\min(r_i)$  return the maximal and minimal radii among  $r_i$ 's in (43), for  $i=1, \dots, 616$ , respectively.

Lastly, the noise data is generated from  $N(0, \sigma^2)$ . In this research,  $a$ ,  $s$ , and  $k$  are set to 50, 25, and 5, respectively.  $\varepsilon_i$ 's of examples in Section I-III are generated based on  $N(0, 1)$  while those in the first case study are generated by  $N(0, 4)$ .

### E. Proof for THEOREM 3

Since the wafer spatial variations only consist of random variation, the sample variance within a moving window, i.e.  $S_{W_i^p}^2$ , can be expressed as:

$$S_{W_i^p}^2 = \sum_{j=1}^p \frac{(M_{i,(j)} - \bar{M}_{i,p})^2}{p-1},$$

where  $\bar{M}_{i,p} = \sum_{j=1}^p \frac{M_{i,(j)}}{p}$  is the average of observations within the moving window  $W_i^p$ . It is

known that the sample variance is an unbiased estimator of the variance, i.e.

$$E(S_{W_i^p}^2) = E\left(\sum_{j=1}^p \frac{(M_{i,(j)} - \bar{M}_{i,p})^2}{p-1}\right) = \sigma^2.$$

Therefore, the expected value of spatial variance for window size  $p$  can be simply derived as:

$$E(S_p^2) = E\left(\frac{1}{n} \sum_{i=1}^n S_{W_i^p}^2\right) = \frac{1}{n} \sum_{i=1}^n E(S_{W_i^p}^2) = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \sigma^2.$$

As a result, if the spatial variations are only random noises as in (16) and (17), the expected values of the spatial moving variances for  $2 \leq p \leq n$  must be all equal. That is, if the expected spatial moving variances are not all equal, then the spatial variations must consist of variations other than random noises. □

#### F. Proof for THEOREM 4

Let  $\bar{M}_{n-1,(i)}$  and  $S_{n-1,(i)}^2$  denote the sample mean and variance based on  $n-1$  random measurements wherein the  $i^{\text{th}}$  random measurement, i.e.  $M_i$ , is excluded. It is straightforward to derive

$$(n-1)S^2 = (n-2)S_{n-1,(n)}^2 + \frac{n-1}{n}(M_n - \bar{M}_{n-1,(n)})^2 \quad (46)$$

by replacing the overall sample mean  $\bar{M}$  in (1) with

$$\bar{M} = \frac{M_n + (n-1)\bar{M}_{n-1,(n)}}{n} \quad (\text{Casella and Berger, 2001}).$$

Similarly, the excluded measurement in (46) can be replaced by the other  $n-1$  measurements in the metrology data, and we have the following equations:

$$\begin{aligned}
& (n-1)S^2 \\
&= (n-2)S_{n-1,(n)}^2 + \frac{n-1}{n}(M_n - \bar{M}_{n-1,(n)})^2 \\
&= (n-2)S_{n-1,(n-1)}^2 + \frac{n-1}{n}(M_{n-1} - \bar{M}_{n-1,(n-1)})^2 \\
&= \quad \vdots \\
&= (n-2)S_{n-1,(1)}^2 + \frac{n-1}{n}(M_1 - \bar{M}_{n-1,(1)})^2
\end{aligned} \tag{47}$$

We can then derive the following relation

$$\begin{aligned}
& \frac{n-1}{n}(M_n - \bar{M}_{n-1,(n)})^2 + \frac{n-1}{n}(M_{n-1} - \bar{M}_{n-1,(n-1)})^2 + \cdots + \frac{n-1}{n}(M_1 - \bar{M}_{n-1,(1)})^2 \\
&= S_{n-1,(n)}^2 + S_{n-1,(n-1)}^2 + \cdots + S_{n-1,(1)}^2 \\
&\Rightarrow \sum_{i=1}^n \frac{n-1}{n}(M_i - \bar{M}_{n-1,(i)})^2 = \sum_{i=1}^n S_{n-1,(i)}^2
\end{aligned} \tag{48}$$

by expanding the sample variances in (47) with the conventional calculation of sample variances, that is:

$$S_{n-1,(i)}^2 = \frac{(n-1) \sum_{j=1, j \neq i}^n M_j^2 - (\sum_{j=1, j \neq i}^n M_j)^2}{n-2}, \text{ for } i=1, \dots, n.$$

If all the RHS's in (47) are summed up, we get:

$$\begin{aligned}
& n(n-1)S^2 \\
&= (n-2)S_{n-1,(n)}^2 + \frac{n-1}{n}(M_n - \bar{M}_{n-1,(n)})^2 \\
&\quad + (n-2)S_{n-1,(n-1)}^2 + \frac{n-1}{n}(M_{n-1} - \bar{M}_{n-1,(n-1)})^2 \\
&\quad + \quad \vdots \\
&\quad + (n-2)S_{n-1,(1)}^2 + \frac{n-1}{n}(M_1 - \bar{M}_{n-1,(1)})^2 \\
&= (n-2)\sum_{i=1}^n S_{n-1,(i)}^2 + \sum_{i=1}^n \frac{n-1}{n}(M_i - \bar{M}_{n-1,(i)})^2 \\
&= (n-2)\sum_{i=1}^n S_{n-1,(i)}^2 + \sum_{i=1}^n S_{n-1,(i)}^2 \\
&= (n-1)\sum_{i=1}^n S_{n-1,(i)}^2 \\
\Rightarrow S^2 &= \frac{1}{n}\sum_{i=1}^n S_{n-1,(i)}^2
\end{aligned} \tag{49}$$



Given  $p=n-1$  in (18), we have

$$S^2 = \frac{1}{C_{n-1}^n} \sum_{k=1}^{C_{n-1}^n} S_{I_{n-1,k}}^2 = \frac{1}{n} \sum_{k=1}^n S_{I_{n-1,k}}^2,$$

which is identical to (49).

Each  $S_{n-1,(i)}^2$  can be further expressed in terms of the sample variances of the samples with  $n-2$  measurements, and (49) becomes:

$$S^2 = \frac{1}{n} \sum_{i=1}^n S_{n-1,(i)}^2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{n-1} \sum_{j=1, j \neq i}^n S_{n-2,(i,j)}^2, \tag{50}$$

where  $S_{n-2,(i,j)}^2$  denotes the sample variance based on  $n-2$  random measurements wherein the  $i^{\text{th}}$  and  $j^{\text{th}}$  measurements are both excluded. Each  $S_{n-2,(i,j)}^2$  in the RHS of (50) is repeatedly calculated, ex:  $S_{n-2,(1,2)}^2$  and  $S_{n-2,(2,1)}^2$  are identical and counted. Therefore, (50) can be revised as:

$$\begin{aligned}
 S^2 &= \frac{1}{n} \sum_{i=1}^n S_{n-1,(i)}^2 = \frac{1}{n} \sum_{i=1}^n \frac{2}{n-1} \frac{\sum_{j=1, j \neq i}^n S_{n-2,(i,j)}^2}{2} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \frac{S_{n-2,(i,j)}^2}{2} \\
 &= \frac{1}{C_{n-2}^n} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \frac{S_{n-2,(i,j)}^2}{2}, \tag{51}
 \end{aligned}$$

which is also identical to (18) with  $p=n-2$ .

By applying the result in (49) recursively, THEOREM 4 is proved. □

## G. Reasoning for Conjecture 1

To reason Conjecture 1, in addition to the inspiration by THEOREM 4, Monte-Carlo simulations are performed to study the distributions of the spatial moving variances. 100,000 randomly distributed wafers are generated from  $N(0, 1)$ , and their spectra are calculated accordingly. One can draw a histogram and perform goodness of fit test to see how the distribution of the 100,000 spatial moving variances fit to the chi-square distributions. Here, we use  $p=2$  and 308 ( $n=616$ ) as two examples to show how they resemble the chi-square distribution. Fig. 39(a) and (b) show the histograms fit to the chi-square distributions and the  $p$ -values of the chi-square goodness of fit tests. The  $p$ -values are all near zero and the histograms look perfectly fit to the chi-square distribution curves for both  $p=2$  and 308 ( $f=0.0032$  and 0.5). □



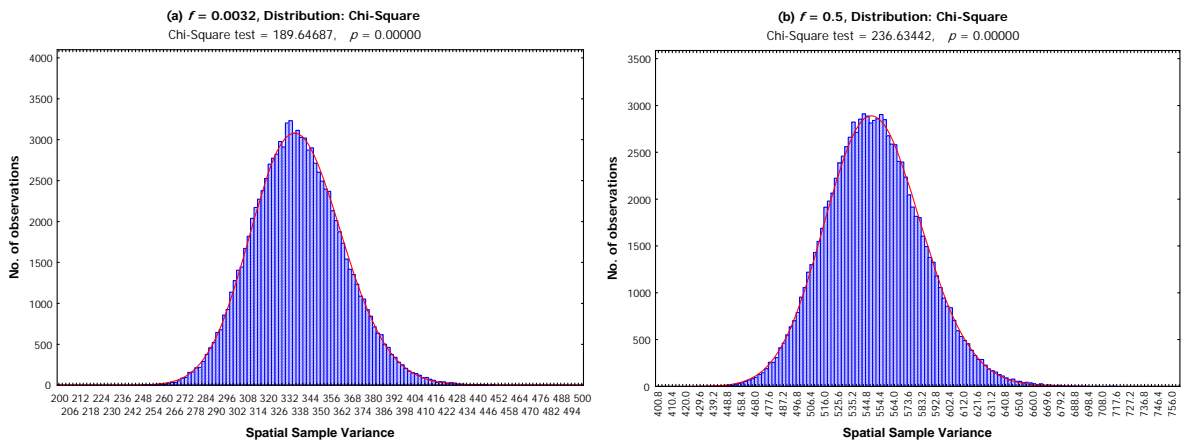


Fig. 39. Histograms of the simulated spatial moving variances under spatial variation frequencies: (a)  $f=0.0032$  ( $p = 2$ ); (b)  $f=0.5$  ( $p = 308$ ).



## Reference

- [1] D. S. Boning and J. E. Chung, "Statistical Metrology: Understanding Spatial Variation in Semiconductor Manufacturing," in *Microelectronic Manufacturing Yield, Reliability, and Failure Analysis II: SPIE 1996 Symposium on Microelectronic Manufacturing*, Austin, TX, Oct. 1996.
- [2] C. M. Borrer, D. C. Montgomery, and G. C. Runger, "Robustness of the EWMA Control Chart to Non-normality," *Journal of Quality Technology*, vol. 31, no. 3, pp 309-316, 1999.
- [3] R. Burch, P. Lin, S. Graves, and E. Antonissen, "Method and system for failure signal detection analysis," U.S. Patent 7415386 B2, Aug. 19, 2008.
- [4] J. P. Cain and C. J. Spanos, "Electrical Linewidth Metrology for Systematic CD Variation Characterization and Causal Analysis," in *Proc. SPIE*, vol. 5038, 2003, pp 350-361.
- [5] G. Casella and R. L. Berger. *Statistical Inference*. California: Duxbury Press, 2001, ch. 5, pp 218-225.
- [6] A. Chen, R. S. Guo, A. Yang and C. L. Tseng, "An Integrated Approach to Semiconductor Equipment Monitoring," *Journal of the Chinese Society of Mechanical Engineering*, vol. 19, no. 6 , pp 581-591, 1998.
- [7] A. Chen, R. S. Guo, and G. S. Wu, "Real-time Equipment Health Evaluation and Dynamic Preventive Maintenance," in *Proc. International Symposium on Semiconductor Manufacturing*, 2000, pp 375-378.
- [8] A. Chen and C. C. Tsai, 'Accommodating Engineering Knowledge in T2 Control Chart Construction for Equipment FDC,' in *Proc. International Symposium on Semiconductor Manufacturing*, 2004, pp 281-285.
- [9] A. Chen and G. S. Wu, "Real-time Health Prognosis and Dynamic Preventive Maintenance Policy for Equipment under Aging Markovian Deterioration," *International Journal of Production Research*, vol. 45, no. 15, pp 3351-3379, 2007.
- [10] A. Chen and J. Blue, "Recipe-Independent Indicator for Tool Health Diagnosis and Predictive Maintenance," *IEEE Transactions on Semiconductor Manufacturing*, vol. 22, no. 4, pp. 522-535, 2009.
- [11] I. B. Gertsbakh. *Models of Preventive Maintenance*. North-Holland, 1977, pp 229-244.

- [12] R. S. Guo and E. Sachs, "Modeling, Optimization, and Control of Spatial Uniformity in Manufacturing Processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 6, no. 1, pp 41-57, 1991.
- [13] L. Han, W. Wang, M. A. McCord, C. N. Berglund, and R. F. W. Pease, "Practical Approach to Separating the Pattern Generator-Induced Mask CD Errors from the Blank/Process-Induced Mask CD Errors Using Conventional Market Measurements," *Journal of Vacuum Science Technology B*, vol. 15, no. 6, pp 2243-2248, 1997.
- [14] Q. P. He and J. Wang, "Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 20, no. 4, pp 345-354, 2007.
- [15] A. Inani, R. Burch, J. Kim, and B. Stine, "Yield improvement using fail signature detection algorithm (FSDA)," in *Proc. ASMC*, 2006, pp. 225-227.
- [16] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall, 2002, ch. 3, pp 124-139.
- [17] C. M. Ke, C. C. Lee, Y. H. Wang, H.J. Lee, C. H. Lin, T. S. Gau, B. J. Lin, H. Kawada, K. Ueda, H. Nomura, N. Ren, and T. Iizumi, "Evaluation of Lin and Hole Measurement by High-Resolution/Low-Magnification CD SEM," in *Proc. SPIE*, vol. 5752, 2005, pp 1292-1299.
- [18] J. K. Kibarian and A. J. Strojwas, "Using Spatial Information to Analyze Correlations between Test Structure Data," *IEEE Transactions on Semiconductor Manufacturing*, vol. 4, no. 3, pp 219-225, 1991.
- [19] S. Kotz and N. L. Johnson. *Process Capability Indices*. London: Chapman & Hall, 1993, pp 37-115, 179-187.
- [20] J. Lacaille and M. Zagrebnoy, "An Unsupervised Diagnosis for Process Tool Fault Detection: The Flexible Golden Pattern," *IEEE Transactions on Semiconductor Manufacturing*, vol. 20, no. 4, pp 355-363, 2007.
- [21] R. C. Leachman. 'Closed-Loop Measurement of Equipment Efficiency and Equipment Capacity,' *IEEE Transactions on Semiconductor Manufacturing*, vol. 10, no. 1, pp 84-97, 1997.
- [22] J. M. Lucas and M. S. Saccucci, "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, vol. 32, no. 1, pp 1-29, 1990.

- [23] D. C. Montgomery. *Introduction to Statistical Quality Control*. New Jersey: John Wiley, 2005, ch. 8, pp 385-422.
- [24] P. K. Mozumder and L. M. Loewenstein, "Method for Semiconductor Process Optimization Using Functional Representations of Spatial Variations and Selectivity," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, vol. 15, no. 3, pp 311-316, 1992.
- [25] X. Ouyang, C. N. Berglund, M. A. McCord, and R. F. W. Pease, "Economical Sampling Algorithm Using Fourier Analysis for Mapping Wafer Critical Dimension Variations," *Journal of Vacuum Science Technology B*, vol. 16, no. 6, pp 3655-3660, 1998.
- [26] X. Ouyang, T. L. Deeter, C. N. Berglund, M. A. McCord, and R. F. W. Pease, "High-Throughput, High-Spatial-Frequency Measurement of Critical Dimension Variations Using Memory Circuits as Electrical Test Structures," *Journal of Vacuum Science Technology B*, vol. 17, no. 6, pp 2707-2713, 1999.
- [27] T. H. Smith, B. E. Goodlin, D. S. Boning, C. O. Oji, J. E. Chung, and H. H. Sawin, "Bias and Variance in Single and Multiple Response Surface Modeling," in *Proc. 3rd International Workshop on Statistical Metrology*, pp 60-63, Honolulu, HI, 1998.
- [28] A. Stamatis, K. Mathioudakis, and K. Papailiou, "Optimal Measurement and Health Index Selection for Gas Turbine Performance Status and Fault Diagnosis," *Journal of Engineering for Gas Turbines and Power*, vol. 114, pp 209-216, 1992.
- [29] D. Steele, A. Coniglio, C. Tang, and B. Singh, "Characterizing Post Exposure Bake Processing for Transient and Steady State Conditions, in the Context of Critical Dimension Control," in *Proc. SPIE*, vol. 4689, 2002, pp 517-530.
- [30] B. E. Stine, D. S. Boning, and J. E. Chung, "Analysis and Decomposition of Spatial Variation in Integrated Circuit Processes and Devices," *IEEE Transactions on Semiconductor Manufacturing*, vol. 10, no. 1, pp 24-41, 1997.
- [31] B. E. Stine, D. S. Boning, and J. E. Chung, "Inter- and Intra-Die Polysilicon Critical Dimension Variation," in *Proc. SPIE*, vol. 2874, 1996, pp 27-35.
- [32] W. Taam, P. Subbaiah, and J. W. Liddy, "A Note on Multivariate Capability Indices," *Journal of Applied Statistics*, vol. 20, pp 339-351, 1993.

- [33] P. Vanoppen, O. Noordman, J. Baselmans, J. V. Schoot, "Analysis of Full Wafer/Full Batch CD Uniformity Using Electrical Line Width Measurements," in *Proc. SPIE*, vol. 4404, 2001, pp 33-44.
- [34] A. K. Wong, A. F. Molless, T. A. Brunner, E. Coker, R. H. Fair, G. L. Mack, and S. M. Mansfield, "Linewidth Variation Characterization by Spatial Decomposition," *Journal of Microlithography, Microfabrication, and Microsystems*, vol. 1, no. 2, pp106-116, 2002.
- [35] J. Ye, C. N. Berglund, J. Robinson, and R. F. W. Pease, "A Review of Mask Errors on a Variety of Pattern Generators," *IEEE Transactions on Semiconductor Manufacturing*, vol. 8, no. 3, pp 319-325, 1995.
- [36] C. Yu, T. Maung, C. J. Spanos, D. S. Boning, J. E. Chung, H. Y. Liu, K. J. Chang, and D. J. Bartelink, "Use of Short-Loop Electrical Measurements for Yield Improvement," *IEEE Transactions on Semiconductor Manufacturing*, vol. 8, no. 2, pp 150-159, 1995.
- [37] H. H. Yue and M. Tomoyasu, "Weighted PCA and its Applications to Improve FDC Performance," in *Proc. 43rd IEEE Conference on Decision and Control*, 2004, pp. 4262-4267.
- [38] J. D. Zimmerman, Javed Sumra, Y. K. Leong, P. Govil, and G. Baxter, "Linewidth Uniformity Error Analysis for Step-and-Scan Systems," in *Proc. SPIE*, vol. 4000, 2000, pp 785-792.

