J Intell Manuf (2006) 17:429–439
DOI 10.1007/s10845-005-0016-7

A novel method for determining machine subgroups and backups with an empirical study for semiconductor manufacturing

Chen-Fu Chien · Chia-Yu Hsu

Received: May 2005 / Accepted: December 2005

Abstract Wafer fabrication for semiconductor manufacturing consists of multiple layers, in which the displacements (i.e., overlay errors) between layers should be reduced to enhance the yield. Although it can reduce variance between layers by fixing the exposure machine (i.e. steeper or scanner), it is not practical to expose the wafer on the same machine from layer to layer for the lengthy fabrication process in real setting. Thus, there is a critical need to determine the similarity machine subgroups, in which appreciate backups for unexpected machine down can be also prioritized. This study aims to develop a novel methodology to fill this gap based on the proposed similarity measurement of systematic overlay errors and residuals. The proposed methodology was validated via empirical study in a wafer fab and the results showed practical viability of this approach.

Keywords Overlay · Similarity · Machine Subgroups · Modeling · Semiconductor manufacturing

Introduction

Semiconductor fabrication facilities (fabs) are the most capital-intensive and complex manufacturing plants today in which similar equipment and process are used to produce Integrated Circuits (IC) including microprocessors, memories, digital signal processor, and application-specific logic. The IC manufacturing process primarily consists of four phases: wafer fabrication, wafer probe, assembly and packing, and final test (Chien & Wu 2003). In particular, wafer fabrication involves the most complex and lengthy process including cleaning, oxidation/deposition/metallization, lithography, etching, ion implantation, photo-resist strip, inspection and measurement in iteration (Chien, et al. 2001).

Wafer fabrication contains multilayer wiring in which the patterned layers must overlay each other to within the tolerance to function properly. Overlay error is a displacement of the present exposure layers relative to the preceding exposure layers. Thus, microlithography that is performed on a very complicated machine, i.e., stepper or scanner, is a vital process affecting the yield as IC feature size reaching nano generation and it is also the bottleneck in fab. If the sampled overlay errors were out of specification, the photoresist should be stripped in order to rework the lithography sequence, yet the capacity loss, sampling cost, and cycle time would be increased. Ideally, the wafer should be fabricated at the same machine to reduce possible disturbance in the process and thus reduce the displacements between layers to minimize the overlay errors. In practice, the wafer was unable to be scheduled to be exposed in the same machine for all the layers for the concerns of bottleneck tool productivity and operation efficiency. There is a tradeoff between yield and productivity. In practice, engineers tend to divide the machine into subgroups rather than fixing the same machine, yet no systematic methodology was proposed.

This study aims to full the gap by developing a methodology to cluster the machine into subgroups with similar characteristics and proposing an algorithm to prioritize the appropriate backups for specific machine based on the similarity measurement of overlay error

C.-F. Chien (✉) · C.-Y. Hsu
Department of Industrial Engineering and Engineering Management National Tsing Hua University, 101 Section 2 Kuang Fu Road, Hsinchu 30013, Taiwan
e-mail: cfchien@mx.nthu.edu.tw

patterns. Indeed, little research has been done to address this issue. In particular, van den Brink, et al. (1988) proposed a matching procedure by comparing all machines to a reference machine, yet the machine should be qualified with long term stability to become a golden standard. However, the overlay error characteristics of each machine are drifting along the time and also from process to process. In practice, engineers rely on their experience and heuristic rules to determine machine groups that are often mistaken as the process being advanced and changed constantly in high-tech manufacturing.

In particular, the overlay errors can be modeled as systematic overlay errors that can thus be corrected via compensation and the residuals that are the uncorrectable portions of overlay errors (Chien, et al. 2003). Thus, this approach involves two phases for comparing machine similarity. The first phase is to cluster machine based on the systematic overlay error factors. The second phase is to consider the covariance of non-systematic overlay errors between specific machines for identify backup priority within the same group. We validated the proposed methodology for stepper subgrouping and backup prioritization with empirical data in a semiconductor fab in Taiwan. The fab has now implemented this methodology with an algorithm embedded for supporting machine grouping and backup decisions on line.

The rest of this paper is organized as follows: Section ‘Fundamental’ describes the fundamental of this approach and reviews related studies. Section ‘The approach’ presents the proposed approach and similarity measurement. Section ‘Empirical study’ compares the results of this approach with the empirical data for validation. Section ‘conclusion’ concludes with discussion of contributions and future research directions.

Fundamental

The terminology and notation used herein are as follows:

$D_{(x,y)}$	The distance between two machines or clusters
x_i	The value of parameter i in machine cluster x
y_i	The value of parameter i in machine cluster y
P	The number of overlay error factors
X, Y	Denotes $p \times 1$ matrix of machine cluster parameters
S^{-1}	The covariance matrix with p parameters
(x, y)	Intrafield coordinate system, with respect to the center of the field
(X, Y)	Interfield coordinate system, with respect to the center of the wafer
d_x, d_y	Intrafield overlay errors with respect to the intrafield coordinate system (x, y)
d_X, d_Y	Interfield overlay errors with respect to the interfield coordinate system (X, Y)
d_{x+X}	The sum of intrafield and interfield overlay errors along the x-axis
d_{y+Y}	The sum of intrafield and interfield overlay errors along the y-axis
T_x, T_y	Intrafield translation with respect to the intrafield coordinate system (x, y)
T_X, T_Y	Interfield translation with respect to the interfield coordinate system (X, Y)
T_{x+X}	Total translation along the x-axis
T_{y+Y}	Total translation along the y-axis
S_X, S_Y	Interfield scale
R_X, R_Y	Interfield rotation
B_X, B_Y	Bow coefficients
M_x, M_y	Intrafield magnification
R_x, R_y	Intrafield rotation
T_{xx}, T_{yx}	Trapezoid distortion along the x-axis,
T_{yy}, T_{xy}	Trapezoid distortion along the y-axis
W_x, W_y	Wedge distortion
D_{3x}, D_{3y}	Third-order lens distortion
D_{5x}, D_{5y}	Fifth-order lens distortion
D_{7x}, D_{7y}	Seventh-order lens distortion
ε_{x+X}	The sum of intrafield and interfield error residual along the x-axis
ε_{y+Y}	The sum of intrafield and interfield error residual along the y-axis.

Overlay error

The overlay errors are measured with the displacements of the exposure field between present layer and previous layer. In particular, the box-in-box pattern that is exposed at the edge of the field is designed to measure overlay errors. If the inside box is patterned in the center of the outside box, there is no overlay error. Otherwise, there is overlay error as shown in Fig. 1. Indeed, the overlay errors consist of systematic and non-systematic overlay errors. Systematic errors caused by specific error factors can be corrected by compensation with corresponding machine adjustments. Non-systematic errors caused by random disturbance cannot be corrected, yet can be minimized via reducing the variance in production. Furthermore, the systematic errors can be separated into inter- and intrafield overlay errors according to their causes and effects as shown in Fig. 2a. The interfield overlay errors are related to the displacements of the fields caused by fitting problems between reticle and wafer, which can be represented with the respect to the wafer center. The intrafield overlay errors are related to the displacements of the layers caused by

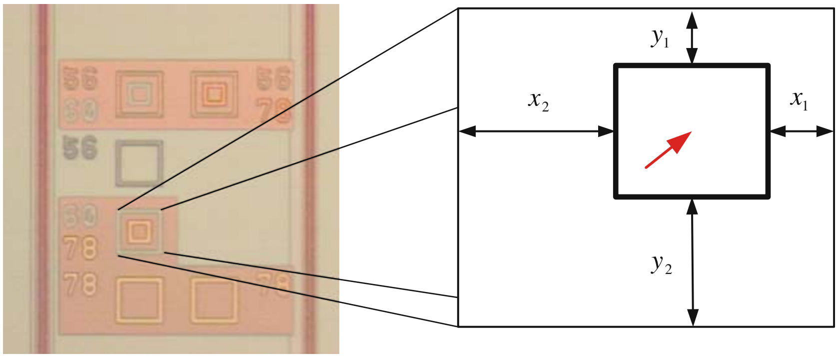

Fig. 1 Overlay measurement

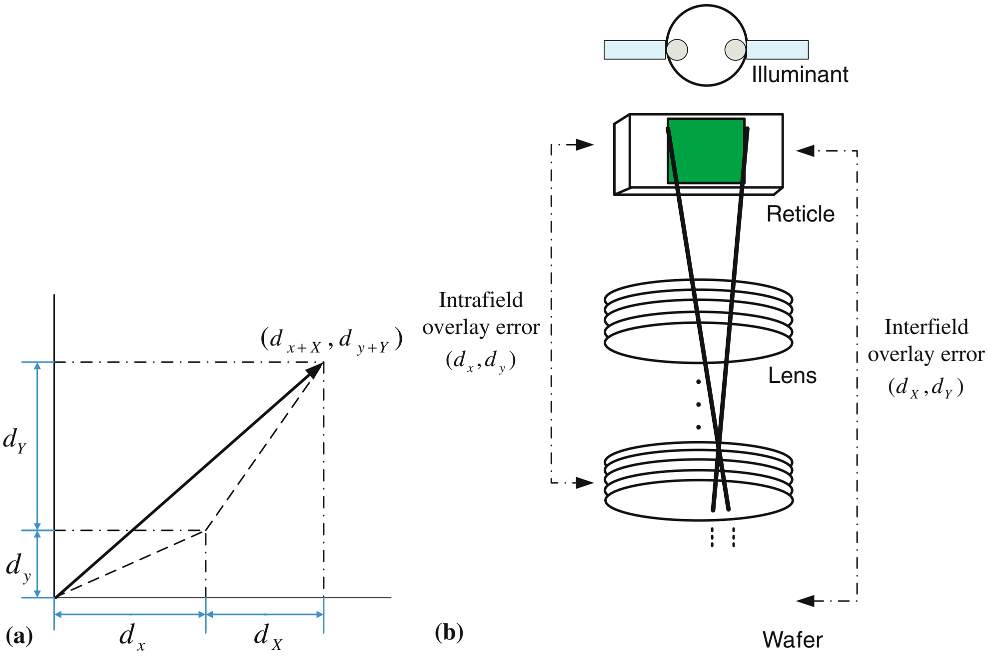

Fig. 2 Causes and effects of systematic overlay errors. **(a)** intrafield and interfield overlay error; **(b)** causes of overlay error

fitting problem between reticle and the illuminant filter lens, which can be represented with the respect to the exposure field center as shown in Fig. 2b. In practice, only total overlay errors can be measured in the *x*-and *y*-axis directions (Chien et al. 2003).

Overlay error model

A number of overlay error models for stepper were developed. In particular, Perloff (1978) first developed an overlay error model consisted of six parameters including translation, rotation, and expansion. In addition to the above factors, MacMillen and Ryden (1982) considered the trapezoid and the third-order lens distortions in their model. Furthermore, Arnold (1983) considered both grid error and lens errors simultaneously and modified Perloff model by adding the bow parameters in the x and y directions (B_X, B_Y) as follows:

$$d_X = T_X + S_X X - R_X y + B_X Y^2, \tag{1}$$

$$d_Y = T_Y + S_Y Y + R_Y X + B_Y X^2. \tag{2}$$

Then, van den Brink et al. (1988) proposed a comprehensive overlay error model including the effects of the intra- and interfield overlay errors. The intrafield model that was based on MacMillen and Ryden (1982) added the fifth-order lens distortion variables as follows:

$$\begin{aligned} d_x = {} & T_x + M_x x - R_x y - T_{xx} x^2 - T_{yx} xy + W_x y^2 \\ & + D_{3x} x(x^2 + y^2) + D_{5x} x(x^2 + y^2)^2 + \varepsilon_x, \end{aligned} \tag{3}$$

$$\begin{aligned} d_y = {} & T_y + M_y y + R_y x - T_{yy} y^2 - T_{xy} xy + W_y x^2 \\ & + D_{3y} y(x^2 + y^2) + D_{5y} y(x^2 + y^2)^2 + \varepsilon_y. \end{aligned} \tag{4}$$

The interfield model combined the model by Perloff (1978) and the bow parameters considered in Arnold

(1983) as follows:

$$d_X = T_X + S_X X - R_X Y + B_X Y^2 + \varepsilon_X, \tag{5}$$

$$d_Y = T_Y + S_Y Y + R_Y X + B_Y X^2 + \varepsilon_Y. \tag{6}$$

Based on the existing models, Lin and Wu (1999) added additional seventh-order lens distortion parameters, in which a total of 26 parameters have to be estimated as follows:

$$\begin{aligned} d_{x+X} = {} & T_{x+X} + S_X X - R_X Y + B_X Y^2 + M_x x - R_{xy} \\ & - T_{xx} x^2 - T_{yx} xy + W_x y^2 + D_{3x} x(x^2 + y^2) \\ & + D_{5x} x(x^2 + y^2)^2 + D_{7x} x(x^2 + y^2)^3 + \varepsilon_{x+X}, \end{aligned} \tag{7}$$

$$\begin{aligned} d_{y+Y} = {} & T_{y+Y} + S_Y Y + R_Y X + B_Y X^2 + M_y y + R_y x \\ & - T_{yy} y^2 - T_{xy} xy + W_y x^2 + D_{3y} y(x^2 + y^2) \\ & + D_{5y} y(x^2 + y^2)^2 + D_{7y} y(x^2 + y^2)^3 + \varepsilon_{y+Y}. \end{aligned} \tag{8}$$

As listed in Table 1, the factors considered in the existing theoretical models are increasingly complicated. Theoretically, sufficient data makes it feasible to solve a multiple regression analysis model that incorporates high-order factors. However, to correctly estimate, meaningfully interpret, and effectively compensate the effect of every parameter in the model is difficult. Furthermore, the number of samples required for the model increases exponentially along with the increasing number of factors considered (Chien et al. 2003).

Indeed, only 20 or 25 overlays are sampled in practice when measuring and controlling the overlay errors. Insufficient data makes the effects of certain factors indistinguishable from (or easily confounded with) other factors. Focusing on real setting, Chien et al. (2003) proposed an overlay model based on the data collected in the stepper directly to bridge the gap between theoretical model and manufacturing practice to estimate systematic overlay errors effectively with the limited number of samples. The proposed overlay model that considered the intra- and interfield overlay errors are as follows:

$$\begin{aligned} d_{x+X} = {} & T_{x+X} + S_X X - (N + \theta)\, Y \\ & + M'_x x - R_x y + \varepsilon_{x+X}, \end{aligned} \tag{9}$$

$$\begin{aligned} d_{y+y} = {} & T_{y+y} + S_Y Y - (\theta - N)\, X \\ & + M'_y y - R_y x + \varepsilon_{y+Y}, \end{aligned} \tag{10}$$

where the variable M_x is defined by $M'_x = M \times M_x$ given the design lens magnification parameter M Similarly, the variable M'_x is defined by $M'_x = M \times M_x$ Furthermore, the variable N denotes non-orthogonality and θ denotes interfield rotation, which is linked by the X-and Y- interfield rotation, R_X and R_Y, are as follows:

$$N = \frac{R_X - R_Y}{2}, \tag{11}$$

$$\theta = \frac{R_X + R_Y}{2}. \tag{12}$$

The proposed model has the advantage to employ the empirically assessable data to effectively estimate the effects of the correctable causes. The coefficients of the overlay error factors considered in the model can be empirically derived through regression analysis (e.g., Johnson and Wichern 1992) of the given number of sampling data (i.e., X, Y, x, y, d_{x+X}, and d_{y+Y}).

Similarity measurement

To measure a similarity level between two steppers or machine groups, three measurement scales including distance, correlation coefficient, and association coefficient (Subhash 1996) can be employed in this study.

First, distance is usually used to measure similarity with one or many parameters. In particular, Minkowski distance is the general form for measuring distance between machine x and y as follows:

$$D^{\text{Min}}_{(x,y)} = \left[\sum_{i=1}^{p} (|x_i - y_i|)^n \right]^{1/n} \tag{13}$$

Euclidean distance, as $n = 2$ in Eq. 13, is the mostly used measurement as follows:

$$D^{\text{E}}_{(x,y)} = \left[\sum_{i=1}^{p} (x_i - y_i)^2 \right]^{1/2} \tag{14}$$

Because Euclidean distance is not scale invariant, the meaningful application requires the variables in the same scale. Alternatively, city-block distance (Manhattan distance) is another form of Minkowski distance as $n = 1$.

$$D^{\text{C}}_{(x,y)} = \sum_{i=1}^{p} |x_i - y_i| \tag{15}$$

Other n values in Minkowski distance will result in different types of distance measurements, yet are seldom used. If the variables are correlated, the Mahalanobis distance is designed to consider correlation among the variables.

$$D^{\text{Maha}}_{(x,y)} = (X - Y)' S^{-1} (X - Y) \tag{16}$$

Second, correlation coefficient can be used to represent the similarity. For example, Pearson correlation

Table 1 Considered factors in stepper overlay error models

Existing models	Considered overlay error factors	
	Intrafield factors	Interfield factors
Perloff (1978)		Translation Interfield rotation Scale
MacMillen and Ryden (1982)	Translation Magnification Intrafield rotation Trapezoid Third-order lens distortion	
Arnold (1983)	Translation Magnification Intrafield rotation Trapezoid Third-order lens distortion	Translation Scale Interfield rotation Bow coefficient
van den Brink et al. (1988)	Translation Magnification Intrafield rotation Trapezoid Wedge distortion Third-order and fifth-order lens distortions	Translation Scale Interfield rotation Bow coefficient
Lin and Wu (1999)	Translation Magnification Intrafield rotation Trapezoid Wedge distortion Third-order, fifth-order, and seventh-order lens distortions	Translation Scale Interfield rotation Bow coefficient
Chien et al. (2003)	Translation Magnification Intrafield rotation	Translation Scale Interfield rotation Orthogonality

Table 2 A 2 × 2 contingency table

Machine i	Machine j		
	1	0	Totals
1	a	b	$a+b$
0	c	d	$c+d$
Totals	$a+c$	$b+d$	$n=a+b+c+d$

coefficient can be employed to measure the linear association between two variables.

Third, associated coefficients including Phi Coefficient (Anderberg 1973) and Yule's Q (Yule 1900) were proposed to measure similarity between binary variables. Considering a 2×2 contingency table as shown in Table 2, Phi coefficient is defined as follows:

$$\phi_{(x,y)} = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \qquad (17)$$

and the Yule's Q coefficient can be also defined as follows:

$$Q_{(x,y)} = \frac{ad-bc}{ad+bc}. \qquad (18)$$

Clustering analysis

The steppers in a fab can be divided into different subgroups based on the similarity measurements, while maximizing within-clusters homogeneity and the heterology between clusters. Clustering analysis includes hierarchical and non-hierarchical methods (Anderberg 1973; Johnson and Wichern 1992). Hierarchical clustering methods that do not require determining the cluster number in advance include agglomerative and divisive approaches. Agglomerative clustering starts with the clusters with individual steppers and merges the closest clusters according to their similarities. As the similarity threshold decreases, all the clusters are eventually merged into the same group. Divisive clustering approaches work in the opposite direction by splitting the initial cluster into two clusters. These subgroups are then further divided into dissimilar clusters as the similarity threshold increases until each of the clusters contains only an individual stepper. The employed similarity index such as distance or correlation measurement can be used to organize the clustering results into a hierarchical tree of clusters with a dendrogram. In particular, Ward's method that is one of the most extensively applied hierarchical clustering denotes each cluster with its centroid and merges two clusters into one with the increasing sum of square. Alternatively, non-hierarchical clustering approaches are design to group machines into *K* clusters, in which the number of clusters can be specified in advance or determined as part of the algorithm. In particular, *K*-means method is a well-known algorithm for non-hierarchical clustering that requires the determination of initial cluster number in advance. Then, each machine is assigned to the cluster within the closest centroid until the cluster centroid does not change. In addition, artificial neural network methods can also be employed for clustering analysis. For example, Kiang, et al. (1995) proposed an interactive clustering approach based on Self-organizing map (SOM) network for conventional group technology involved in cell formation. Chien et al.(2002) combined spatial statistics and Adaptive Resonance Theory (ART) neutral network for wafer binmap clustering for defect diagnosis and yield enhancement.

Cellular manufacturing and group technology (GT)

Most of the existing studies for machine grouping are related to solve cell formation in group technology (GT) that involves grouping the parts into families and the machine into cells. Through part-family formation, the parts with similar function and design characteristics are grouped to be produced within the same cell. Through machine-cell formation, the machines are grouped to produce one or more part-families. Use of GT can minimize material handling, cost saving, time compression, reduction of setup time, reduction of production lead time, reduction of rework rate, and improvement of human relations (e.g., McAuley 1972; Kusiak 1985, 1987).

In cellular manufacturing, the relationship between parts and machines can be formed as part-machine matrix A_{ij}, where A_{ij} is binary and represented as the part i need machine j ($A_{ij} = 1$) or not ($A_{ij} = 0$). The part-machine matrix can be represented the production route, production volume, and processing time of part i in machine j. Various approaches have been proposed to solve the part-machine grouping problem including matrix based approach (King 1980; Chan and Milner 1982; Kusiak 1985; Seifoddini and Wolfe 1986; Al-Sultan 1997), mathematical programming (Srinivasan, et al. 1990; Tam 1990; Kusiak and Cho 1992), graph theory based approach (Chandrasekharan and Rajagopalan 1986; Vannelli and Kumar 1986; Kiang 1995), fuzzy logic (Ben-Arieh and Triantaghyllou 1992), and genetic algorithm (Moon and Chi 1992) and neural networks (Kamal and Burke 1996, Al-Sultan and Fedjki 1997).

Although a number of group technologies were proposed, the present problem is different and thus the existing GT approaches can not deal with the needs for determining the stepper subgroups and backups in lithography process. First, the relationship between the steppers is not a part-machine relationship and the steppers are not routed. Second, most of the above approaches used the incidence matrix to determine the part-family and machine-cells. However, the overlay errors considered in this study are caused by several overlay error factors caused by the misalignment problem. Thus, the steppers similarities cannot be compared directly with specific factors or the similarity coefficients proposed in the existing GT studies without decomposing measured overlay errors via the proposed approach.

The approach

Problem structuring

Most machines utilized in semiconductor manufacturing are often set in parallel as a machine group. The overlay errors will be increased if the wafers are fabricated by the machines with different overlay error patterns between different layers and thus the displacements between layers are increased. Thus, Intel emphasizes importance of "copy exactly" in their machine configurations. Nevertheless, most of semiconductor fabs tend to maintain different types of machines at the same time so as to avoid being dominated by specific equipment vendors. However, utilization of bottleneck machine such as stepper will be low because the capacities are wasted during waiting. In order to increase the throughput while avoiding yield loss caused by serious displacements between layers, it is crucial to determine appropriate threshold of similarity to cluster the machines into groups with similar overlay error characteristics in lithography. Furthermore, when a machine is not available, other machines in the same group should be able to be allocated as backup to maintain the throughput without affecting yield.

Comparing overlay similarity

This approach compared machine similarity on the basis of the similarity of overlay errors including systematic and non-systematic overlay errors. Following Chien et al. (2003), this approach considers the overlay error factors including the intrafield errors of translation, magnification, and rotation and the interfield errors of translation, scale, rotation, and orthogonality as input variables for clustering. First, we applied Ward's method to investigate alternative numbers of clusters with different similarity thresholds. Then, *K*-means method is employed to further refine the clustering result. In particular, three indexes are used to evaluate the numbers of cluster (Subhash 1996) as follows. The Root Mean Square Standard Deviation (RMSSTD) that is the pooled standard deviation of all the variables composing the new cluster can be used to measure the within-cluster homogeneity. The *R*-square can be used to measure the heterology between the clusters. The Semi-partial *R*-square (SPR) can be used to measure the homogeneity of merged clusters, i.e., loss of homogeneity within the merged cluster.

Residuals in the proposed overlay error model that are composed of non-systematic overlay errors including lens distortions and random errors cannot be corrected through compensation. In particular, lens distortions include trapezoid (i.e., tilt errors), anamorphism, wedge distortion, asymmetrical errors, and high-order lens distortion. To compare the similarity of non-systematic overlay errors, Pearson correlation coefficients can be used to measure the correlation between residuals in the *x*-and *y*-axis directions, respectively.

Integrated similarity measurement

Then, we proposed the following procedure to measure the integrated similarity between steppers by considering both systematic and non-systematic overlay errors:

(a) Calculate the similarity based on systematic overlay errors as follows:

$$S_{ij} = \begin{cases} M/2, & \text{if the machine } i \text{ and } j \\ & \text{belong to the same cluster,} \\ 0, & \text{otherwise,} \end{cases} \tag{19}$$

where M is the machine number.

(b) Calculate the similarity based on non-systematic overlay errors. The similarity score X_{ij} is used to denote the similarity of residual along the *x*-axis.

$$X_{ij} = M/2 \times r^x_{ij}, \tag{20}$$

where the r^x_{ij} is the correlation coefficient of residuals along the *x*-axis between machine *i* and machine *j*. The similarity score Y_{ij} is used to denote the similarity of residual along the *y*-axis.

$$Y_{ij} = M/2 \times r^y_{ij}, \tag{21}$$

where the r^y_{ij} is the correlation coefficient of residual along the *y*-axis between machine *i* and *j*. These two similarity scores are weighting averaged with b_1 and b_2 according to the residuals along the *x*-axis and *y*-axis, respectively. That is, $b_1 + b_2 = 1$.

(c) Aggregate the total weighted similarity score W_{ij} to denote the degree of similarity between machine *i* and *j* as follows:

$$W_{ij} = w_\mathrm{s} S_{ij} + w_\mathrm{r}(b_1 X_{ij} + b_2 Y_{ij}), \tag{22}$$

where the weighting of similarity scores for systematic and non-systematic overlay errors, i.e., w_s and w_r, are determined according to the average proportion of overlay errors that can be explained by the overlay error model and the rest proportion of residuals. That is, $w_\mathrm{s} + w_\mathrm{r} = 1$. In other words, if the goodness of fit of overlay model is high, the weighting of systematic overlay errors for similarity measurement will also be high. In sum, the above scores S_{ij}, X_{ij}, Y_{ij}, and W_{ij} are defined to be less or equal to half of the total stepper number and the higher scores denote higher similarity between steppers *i* and *j*.

Empirical study

Experiment design and data preparation

An empirical study was conducted to validate the proposed methodology in a fab in Taiwan. Following the concept of experimental design, real data ten steppers was collected at the same date. This company is one of the largest DRAM producers worldwide. Owing to sampling cost and production constraints in practice, only 20 overlays on a wafer were measured in practice, i.e., four overlays per field and five exposure fields.

Overlay error modeling

To decompose the measured overlay errors into systematic overlay errors and residuals, the assessed data are used to fit the empirical model (Chien et al. 2003) by using least square method. Table 3 summarizes the *R*-squares of regression analysis results for the steppers.

Table 3 R-squares of fitted models for different steppers

stepper *R*-squares	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
R_x^2	0.886	0.856	0.828	0.751	0.927	0.933	0.961	0.574	0.887	0.835
R_y^2	0.954	0.837	0.981	0.975	0.818	0.957	0.163	0.787	0.877	0.866
Average *R*-squares	0.920	0.847	0.905	0.863	0.872	0.945	0.562	0.681	0.882	0.851

The *R*-square means the proportion of total variability in the overlay errors that can be explained through the systematic overlay error factors. In the other word, *R*-square represents the goodness of fit of the proposed model with the input data from the corresponding stepper.

Comparison of systematic and non-systematic similarity

To compare, the similarity of systematic overlay errors between the ten steppers, we used the systematic error factors, i.e. T_{x+X}, T_{y+Y}, S_X, S_Y, R_X, R_Y, M'_x, M'_y, R_x, and R_y, as the factors for clustering by following the proposed procedure.

First, the Ward's method was used to determine the appropriate number of clusters with the dendrogram for linkage distances between ten steppers as shown in Fig. 3. The cluster number can be determined according to the overlay error tolerance, namely, four. Furthermore, the three proposed indexes for comparing alternative clustering numbers are shown in Fig. 4. The RMSSTD and SPR are increasing while the *R*-square is decreasing as the cluster number is reduced from four to three. That is, the within-cluster homogeneity and the between-cluster heterology were both decreasing as the cluster number reduced from four to three as shown in Fig. 4. Therefore, the construct validity of determining the cluster number to be four was also validated.

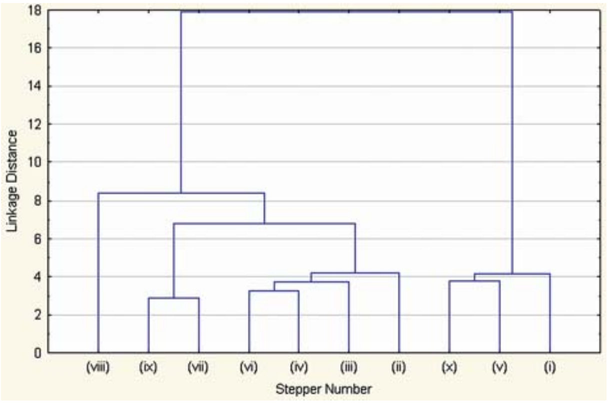

Fig. 3 Dendrogram for linkage distances between ten steppers

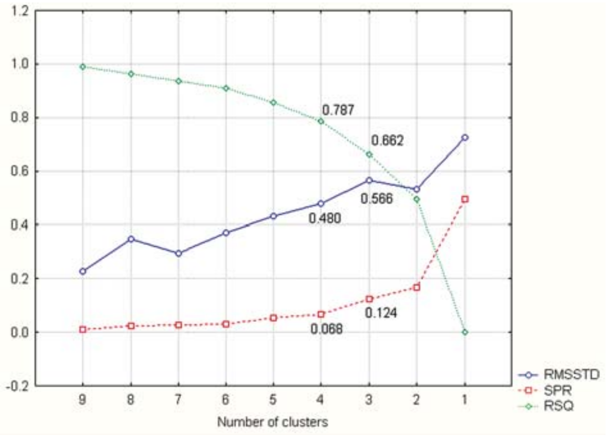

Fig. 4 Alternative cluster numbers versus three clustering indexes

Table 4 *K*-means clustering of steppers ($K = 4$)

Cluster	Steppers			
1	(8)			
2	(2)	(3)	(4)	(5)
3	(1)	(5)	(9)	
4	(7)	(10)		

Second, *K*-means algorithm with $K = 4$ was applied to further refine the clustering result and the results were summarized in Table 4.

Then, to evaluate the residual similarity between the steppers, the Pearson correlation coefficients were applied with the results summarized in Tables 5 and 6. The larger value of residual correlation coefficients represents that the corresponding machine pair has higher degree of similarity. In addition, positive correlation coefficient represents that the two steppers have the same residual direction.

Integrated similarity measurement

Following the proposed procedures, we calculate the integrated similarity measurement between the steppers via combining the above measurements based on systematic and non-systematic overlay errors. For instance, considering stepper (1), because 92% overlay errors can be explained by overlay error model, the weighting

Table 5 Correlation coefficients of residuals along the *x*-axis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	1.00	0.13	0.77	0.51	−0.24	0.55	0.82	0.64	0.17	0.36
(2)	0.13	1.00	0.24	−0.24	0.79	0.47	0.54	−0.51	−0.14	0.31
(3)	0.77	0.24	1.00	0.39	−0.10	0.69	0.70	0.53	0.30	0.62
(4)	0.51	−0.24	0.39	1.00	−0.26	0.14	0.18	0.51	0.22	0.63
(5)	−0.24	0.79	−0.10	−0.26	1.00	0.20	0.22	−0.76	−0.03	0.18
(6)	0.55	0.47	0.69	0.14	0.20	1.00	0.73	0.33	0.28	0.49
(7)	0.82	0.54	0.70	0.18	0.22	0.73	1.00	0.29	0.16	0.36
(8)	0.64	−0.51	0.53	0.51	−0.76	0.33	0.29	1.00	0.22	0.21
(9)	0.17	−0.14	0.30	0.22	−0.03	0.28	0.16	0.22	1.00	0.02
(10)	0.36	0.31	0.62	0.63	0.18	0.49	0.36	0.21	0.02	1.00

Table 6 Correlation coefficients of residuals along the *y*-axis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	1.00	0.84	−0.31	0.07	0.82	−0.48	0.51	−0.85	0.58	−0.42
(2)	0.84	1.00	−0.32	0.27	0.89	−0.53	0.74	−0.75	0.60	−0.27
(3)	−0.31	−0.32	1.00	0.44	−0.24	0.68	0.23	0.32	0.10	0.38
(4)	0.07	0.27	0.44	1.00	0.21	0.10	0.62	−0.02	0.43	0.27
(5)	0.82	0.89	−0.24	0.21	1.00	−0.58	0.71	−0.79	0.53	−0.37
(6)	−0.48	−0.53	0.68	0.10	−0.58	1.00	−0.05	0.45	0.08	0.35
(7)	0.51	0.74	0.23	0.62	0.71	−0.05	1.00	−0.37	0.69	0.10
(8)	−0.85	−0.75	0.32	−0.02	−0.79	0.45	−0.37	1.00	−0.48	0.55
(9)	0.58	0.60	0.10	0.43	0.53	0.08	0.69	−0.48	1.00	−0.15
(10)	−0.42	−0.27	0.38	0.27	−0.37	0.35	0.10	0.55	−0.15	1.00

w_s equals 0.920 and thus w_r equals 0.08, as shown in Table 7. Based on the clustering result of systematic overlay error factors, Steppers (1), (5) and (9) belong to the same subgroup and thus the corresponding similarity scores S_{ij} were recorded as $10/2 = 5$, while the others were denoted as 0. As for non-systematic overlay errors, the correlation coefficients between residuals were also derived. Thus, the integrated similarity scores W_{ij} were aggregated with the weighting and the results were summarized in Table 8.

Comparison of the results of this approach and existing rules

To validate the proposed methodology, we compared the results with the existing rules for backups in this fab. Table 8 lists the existing backup steppers in the fab based on wafer pilot run. Also, according to the integrated similarity measurement W_{ij} for each stepper i and alternative backup stepper j in Table 8, the backup steppers with the similar characteristics of overlay errors can be identified and prioritized. The comparison results showed that steppers (5), (7), and (10), the derived backup steppers are the same as the existing rules. For steppers (2), (4), and (6), the derived backups are more than those in the existing rules and thus suggested additional backup with sufficient similarity. Alternatively, for steppers (1), (3), (8), and (9), the derived backups are less than those in the existing rules. In particular, the proposed method did not suggest any backup for stepper (8) due to its poor similarity with the others. Furthermore, the proposed methodology provides an explicit logic for judging the backup decisions that used to rely on experienced engineers. Thus, the results demonstrated the practical viability of this approach. Furthermore, comparing to the existing rules, this approach can determine the machine groups effectively and also derive the appreciate backups for each stepper with various similarity thresholds in the light of tradeoffs between yield and operational efficiency.

Conclusion

This study proposed an effective approach and explicit decision process for determining critical machine subgroups and backups based on the factors caused overlay errors in wafer fabrication and thus the yield loss due to mismatch can be avoided. For each stepper, the backup steppers with the similar overlay error characteristics can be identified and prioritized according to the proposed similarity measurement and adjustable thresholds. We validated this approach with an empirical study in a wafer fab. The results derived from real data for backups were compared with the existing rules and the results demonstrated convergent validity of this

Table 7 Weighting for stepper *i* versus stepper *j*

Stepper				Backup steppers									
				(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	w_s	0.920	S_{ij}	–	0	0	0	5	0	0	0	5	0
	w_r	0.08	X_{ij}	–	0.653	3.849	2.533	1.192	2.731	4.120	3.183	0.840	1.788
			Y_{ij}	–	4.209	1.546	0.365	4.107	2.408	2.566	4.241	2.878	2.087
	$W_{ij} = w_s S_{ij} + w_r(b_1 X_{ij} + b_2 Y_{ij})$			–	0.194	0.215	0.116	**4.812**	0.205	0.267	0.296	**4.749**	0.155
(2)	w_s	0.847	S_{ij}	0	–	5	5	0	5	0	0	0	0
	w_r	0.153	X_{ij}	0.653	–	1.180	1.200	3.973	2.331	2.679	2.528	0.684	1.527
			Y_{ij}	4.209	–	1.603	1.370	4.468	2.642	3.724	3.752	2.979	1.332
	$W_{ij} = w_s S_{ij} + w_r(b_1 X_{ij} + b_2 Y_{ij})$			0.373	–	**4.447**	**4.430**	0.647	**4.615**	0.491	0.481	0.281	0.219
(3)	w_s	0.905	S_{ij}	0	0	–	5	0	5	0	0	0	0
	w_r	0.095	X_{ij}	3.849	1.180	–	1.927	0.475	3.465	3.524	2.671	1.475	3.117
			Y_{ij}	1.546	1.603	–	2.217	1.188	3.378	1.126	1.611	0.518	1.882
	$W_{ij} = w_s S_{ij} + w_r(b_1 X_{ij} + b_2 Y_{ij})$			0.257	**4.656**	–	**4.721**	0.079	**4.850**	0.221	0.204	0.095	0.238
(4)	w_s	0.863	S_{ij}	0	5	5	–	0	5	0	0	0	0
	w_r	0.137	X_{ij}	2.533	1.200	1.927	–	1.309	0.694	0.881	2.558	1.085	3.128
			Y_{ij}	0.365	1.370	2.217	–	1.069	0.516	3.110	0.075	2.130	1.327
	$W_{ij} = w_s S_{ij} + w_r(b_1 X_{ij} + b_2 Y_{ij})$			0.198	**4.492**	**4.599**	–	0.163	**4.399**	0.273	0.180	0.220	0.305
(5)	w_s	0.872	S_{ij}	5	0	0	0	–	0	0	0	5	0
	w_r	0.128	X_{ij}	1.192	3.973	0.475	1.309	–	1.009	1.103	3.776	0.149	0.902
			Y_{ij}	4.107	4.468	1.188	1.069	–	2.888	3.547	3.925	2.668	1.826
	$W_{ij} = w_s S_{ij} + w_r(b_1 X_{ij} + b_2 Y_{ij})$			**4.700**	0.539	0.106	0.152	–	0.249	0.297	0.491	**4.542**	0.174
(6)	w_s	0.945	S_{ij}	0	5	5	5	0	–	0	0	0	0
	w_r	0.055	X_{ij}	2.731	2.331	3.465	0.694	1.009	–	3.627	1.630	1.411	2.443
			Y_{ij}	2.408	2.642	3.378	0.516	2.888	–	0.226	2.247	0.421	1.730
	$W_{ij} = w_s S_{ij} + w_r(b_1 X_{ij} + b_2 Y_{ij})$			0.141	**4.862**	**4.914**	**4.760**	0.107	–	0.105	0.106	0.050	0.114
(7)	w_s	0.562	S_{ij}	0	0	0	0	0	0	–	0	0	5
	w_r	0.438	X_{ij}	4.120	2.679	3.524	0.881	1.103	3.627	–	1.438	0.782	1.778
			Y_{ij}	2.566	3.724	1.126	3.110	3.547	0.226	–	1.848	3.440	0.494
	$W_{ij} = w_s S_{ij} + w_r(b_1 X_{ij} + b_2 Y_{ij})$			1.464	1.402	1.018	0.874	1.018	0.844	–	0.720	0.925	**3.308**
(8)	w_s	0.681	S_{ij}	0	0	0	0	0	0	0	–	0	0
	w_r	0.319	X_{ij}	3.183	2.528	2.671	2.558	3.776	1.630	1.438	–	1.119	1.041
			Y_{ij}	4.241	3.752	1.611	0.075	3.925	2.247	1.848	–	2.408	2.726
	$W_{ij} = w_s S_{ij} + w_r(b_1 X_{ij} + b_2 Y_{ij})$			1.185	1.003	0.684	0.420	1.229	0.619	0.525	–	0.563	0.601
(9)	w_s	0.882	S_{ij}	5	0	0	0	5	0	0	0	–	0
	w_r	0.118	X_{ij}	0.840	0.684	1.475	1.085	0.149	1.411	0.782	1.119	–	0.083
			Y_{ij}	2.878	2.979	0.518	2.130	2.668	0.421	3.440	2.408	–	0.773
	$W_{ij} = w_s S_{ij} + w_r(b_1 X_{ij} + b_2 Y_{ij})$			**4.629**	0.217	0.118	0.190	**4.575**	0.108	0.250	0.208	–	0.051
(10)	w_s	0.851	S_{ij}	0	0	0	0	0	0	5	0	0	–
	w_r	0.149	X_{ij}	1.788	1.527	3.117	3.128	0.902	2.443	1.778	1.041	0.083	–
	Y_{ij}	2.087		2.087	1.332	1.882	1.327	1.826	1.730	0.494	2.726	0.773	–
	$W_{ij} = w_s S_{ij} + w_r(b_1 X_{ij} + b_2 Y_{ij})$			0.290	0.214	0.374	0.333	0.204	0.312	**4.422**	0.282	0.064	–

Table 8 Comparison of stepper backup priorities with existing rules

Stepper	Priority order of backup steppers			Existing rules based on product pilot run
	1	2	3	
(1)	(5) 4.812	(9) 4.749		(5), (9), (8), (3)
(2)	(4) 4.615	(3) 4.447	(6) 4.43	(4), (3)
(3)	(6) 4.85	(4) 4.721	(2) 4.656	(6), (4), (2), (1)
(4)	(3) 4.599	(2) 4.492	(6) 4.399	(3), (2),
(5)	(1) 4.7	(9) 4.542		(9), (1)
(6)	(3) 4.914	(2) 4.862	(4) 4.76	(3)
(7)	(10) 3.308			(10)
(8)				(1), (3)
(9)	(1) 4.629	(5) 4.575		(1), (5), (8)
(10)	(7) 4.422			(7)

approach. Considering the tradeoffs between yield loss and operational efficiency, this approach can effectively determine the machine subgroups and also derive the appreciate backups with different thresholds of similarity between the steppers. Furthermore, the results of machine grouping and backup can also be employed as the criteria for scheduling lithographic exposures to achieve the objectives of yield enhancement and tool productivity simultaneously.

More studies are needed to validate this approach in various settings. Also, a decision analysis framework should be developed to assist in machine backup decisions with specified risk levels and the involved tradeoffs between possible yield loss and potential throughput loss. Future, research should be done to develop an evolution mechanism for adjusting the results of machine subgroups and backups in light of constant process advancement and change.

Acknowledgements This research is supported by National Science Council (NSC 93-2213-E-007-008) and Industry Technology Research Institute (ITRI). The preliminary version of this research was presented at IML2005 at Waseda University sponsored by Interchange Association, Japan and NSC, Taiwan. Special thanks go to Professor Mitsuo Gen and Professor Hark Hwang for their invaluable inputs.

References

Al-Sultan, K. S. (1997). Hard clustering approach to the part family formation problem. *Production Planning and Control, 8*(3), 231–236.

Al-Sultan, K. S., & Fedjki, C. A. (1997). Genetic algorithm for the part family formation problem. *Production Planning and Control, 8*(8), 788–796.

Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.

Arnold, W. H. (1983). Image placement differences between 1:1 projection aligners and 10:1 reduction wafer steppers. *Proceedings of SPIE: Optical Microlithography, 394*, 87–98.

Ben-Arieh, D., & Traintaphyllou, E. (1992). Quantifying data for group technology with weighted fuzzy features. *International Journal of Production Research, 30*(6), 1285–1299.

Chan, H.M., & Milner, D.A. (1982). Direct clustering algorithm for group formation in cellular manufacturing. *Journal of Manufacturing systems*, *1*(1), 65–74.

Chandrasekharan, M. P., RajaGopalan, R. (1986). An ideal seed non-hierarchical clustering algorithm for cellular manufacturing. *International Journal of Production Research, 24*(2), 451–464.

Chien, C., Hsu, S., & Deng, J. (2001). A cutting algorithm for optimizing the wafer exposure pattern. *IEEE Transactions on Semiconductor Manufacturing, 14*(2), 157–162.

Chien, C., Lin, D., Liu, Q., Peng, C., Hsu, C., & Huang, C. (2002). Developing a data mining method for wafer binmap clustering and an empirical study in a semiconductor manufacturing fab. *Journal of the Chinese Institute of Industrial Engineers, 19*(2), 23–38.

Chien, C., Chang, K., & Chen, C. (2003). Design of a sampling strategy for measuring and compensating for overlay errors in semiconductor manufacturing. *International Journal of Production Research, 41*(11), 2547–2561.

Chien, C., & Wu, J. (2003). Analyzing repair decisions in the site imbalance problem of semiconductor test machines. *IEEE Transactions on Semiconductor Manufacturing, 16*(4), 704–711.

Johnson, R. A., & Wichern, D. W. (1992). *applied multivariate statistical analysis,* 3rd edn. Englewood Cliffs, New Jersey: Prentice Hall.

Kamal, S., & Burke, L. I. (1996). FACT: a new neural network-based clustering algorithm for group technology. *International Journal of Production Research, 34*(4), 919–946.

Kiang, M. Y. Kulkarni, U. R., & Tam, K. Y., (1995). Self-organizing map network as an interactive clustering tool: an application to group technology. *Decision Support Systems, 15*, 351–374.

King, J. R. (1980). Machine-component group formation in production flow analysis: an approach using a rank order clustering algorithm. *International Journal of Production Research, 18*(2), 213–232.

Kusiak, A. (1985). The part families problem in flexible manufacturing systems. *Annals of Operational Research, 25*, 561–569.

Kusiak, A., (1987). The generalized group technology concept. *International Journal of Production Research, 25*(4), 561–569.

Kusiak, A., & Cho, M. (1992). Similarity coefficient algorithms for solving the group technology problem. *International Journal of Production Research, 30*(11), 2633–2646.

Lin, Z., & Wu, W. (1999). Multiple linear regression analysis of the overlay accuracy model. *IEEE Transaction on Semiconductor Manufacturing, 12*, 229–237.

MacMillen, D., & Ryden, W. D. (1982). Analysis of image field placement deviations of a 5× microlithographic reduction lens. *Proceedings of SPIE: Optical Microlithography-Technology, 334*, 78–89.

McAuley, J. (1972). Machine grouping for efficient production. *The Production Engineering, 52*, 53–57.

Moon, Y. B., & Chi, S. C. (1992). Generalized part family formation using neural network techniques. *Journal of Manufacturing Systems, 11*(3), 149–159.

Perloff, D. S. (1978). A four-point electrical measurement technique for characterizing mask superposition errors on semiconductor wafers. *IEEE Journal of Solid State Circuits, 13*(4), 436–444.

Seiffodoni, H., & Wolfe, P. M., (1986). Application of similarity coefficient method in group technology. *IIE Transactions, 18*(13), 271–277.

Srinivasan, G., Narendran, T. T., & Mahadevan, B. (1990). An assignment model for the part-families problem in group technology. *International Journal of Production Research, 28*(1), 145–152.

Subhash, S. (1996). *Applied multivariate techniques*. New York: Wiley.

Tam, K. Y. (1990). An operation sequence based similarity coefficient for part families formations. *Journal of Manufacturing Systems, 9*(1), 55–68.

van den Brink, M. A., DeMol, C. G. M., & George, R. A. (1988). Matching performance for multiple wafer steppers using an advanced metrology procedure. *Proceedings SPIE: Integrated Circuit Metrology, Inspection, and Process Control II, 921*, 180–197.

Vannelli, A., & Kumar, K. (1986). A method for finding minimal bottle-neck for grouping part-machine families. *International Journal of Production Research, 24*(2), 387–400.

Yule, G. U. (1900). On the association of attribute in statistics: with illustration from the material of the childhood society. *Philosophical Transactions of the Royal Society of London, Series A, 194*, 257–319.