Data Mining for Optimizing IC Feature Designs to Enhance Overall Wafer Effectiveness

Chen-Fu Chien, *Member, IEEE,* and Chia-Yu Hsu

***Abstract*—As global competition continues to strengthen in semiconductor industry, semiconductor companies have to continuously advance manufacturing technology and improve productivity to maintain competitive advantages. Die cost is significantly influenced by wafer productivity that is determined by yield rate and the number of gross dies per wafer. However, little research has been done on design for manufacturing and productivity enhancement through increasing the gross die number per wafer and decreasing the required shot number for exposure. This paper aims to propose a novel approach to improve overall wafer effectiveness via data mining to generate the optimal IC feature designs that can bridge the gap between integrated circuit (IC) design and wafer fabrication by providing chip designer with the optimal IC feature size in the design phase to increase gross dies and reduce the required shots. An empirical study was conducted in a leading semiconductor company for validation. The results have shown that the proposed approach can effectively enhance wafer productivity. Indeed, the developed solution has been implemented in the company to provide desired IC features to IC designers to enhance overall wafer effectiveness.**

***Index Terms*—Data mining, design for manufacturing, fab economics, manufacturing informatics, manufacturing intelligence, overall wafer effectiveness, yield optimization.**

I. Introduction

SEMICONDUCTOR industry is capital intensive in which capacity utilization and capital effectiveness significantly affect the profitability competitiveness of semiconductor manufacturing companies [1]. Analogous to the repetitive steps in making a layer cake, wafer fabrication for semiconductor devices such as microprocessors, memories, digital signal processor, and consumer electronics applications involves complex and lengthy process using similar equipment for 30 to 40 reentrant loops of cleaning, oxidation/deposition/metallization, lithography, etching, ion implantation, photo-resist strip, inspection and measurement processes. Driven by Moore's Law that the number of transistors fabricated on the same area of integrated circuit (IC) will be doubled approximately every one or two years [2], the technology has been continuously advanced to reduce IC feature sizes and the line widths. Using state of the art nano technologies, semiconductor manufacturers can fabricate an IC chip with more than a billion transistors in a chip the size of a thumbnail. However, due to diversifying product lines and shortening product life cycle in the consumer electronics era, the involved uncertainty in demand and the fluctuation of semiconductor supply chains make the virtual integration of IC design and wafer fabrication increasingly difficult [3]. On one hand, the market requires semiconductor products with increasing complexity to be developed in a shortening time-to-the-market. On the other hand, as the technology is continuously advanced and required functionality is increasing, the time left to the IC designers for the exploitation of their skills and creativity for optimizing IC designs is drastically reducing.

Data mining and manufacturing intelligence approaches have been employed for data value development, manufacturing knowledge discovery, and manufacturing informatics to maintain the competitive advantages [3]–[6]. In semiconductor manufacturing, the information and intelligence extracted from manufacturing data automatically collected about the products, WIP (work in process), metrology, productivity, yield, resources, WAT (wafer acceptance test) test, and CP (circuit probe) test can be used for cost reduction, defect diagnosis, yield enhancement, cycle time reduction, and inventory management. As the feature size keeps shrinking, most potential excursions in manufacturing should be addressed and improved during the IC design phase. The causal relationship between the layout designable variables and manufacturing response variables should be identified for IC designers in advance, but it is difficult to directly formulate the physical relationship via a mathematical model. On the other hand, empirical models may be extracted from the IC design and manufacturing data. A number of studies have addressed the modification of physical layout design to improve the manufacturing yield and quality [7]. However, little research has addressed the present problem via employing data mining to extract rules for IC designs to enhance wafer productivity in semiconductor manufacturing.

This study aims to propose a novel approach to analyze the gross die data of alternative IC feature designs to optimize the IC feature design to enhance overall wafer effectiveness (OWE) [8]. In particular, we develop the model that considers the number of gross die (# GD) and shot number (# shot) to

Manuscript received August 27, 2012; revised September 19, 2013; accepted November 4, 2013. Date of publication November 20, 2013; date of current version January 30, 2014. This study was supported by National Science Council, Taiwan (NSC100-2628-E-007-017-MY3; NSC101-2221-E-155-035, NSC 102-2622-E-007-013), and Hsinchu Science Park (102A26).

C.-F. Chien is with the Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: cfchien@mx.nthu.edu.tw).

C.-Y. Hsu is with the Department of Information Management, Yuan Ze University, Chung-Li 32003, Taiwan.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSM.2013.2291838

generate specific optimized regions of IC feature to maximize the # GD for enhancing wafer productivity, while reducing the # shot for increasing throughput and cost reduction. In particular, the throughput (i.e., wafer per hour, WPH) can be improved by reducing exposure shot per wafer. The wafer productivity can be expressed as the faction of the effective useful wafer area to the total wafer area [8]. The useful wafer area is mainly related to the number of gross dies fabricated on a wafer. To estimate the validity of the proposed approach, an empirical study was conducted in a leading semiconductor company in Taiwan. The results showed that the proposed approach can generate effective suggestions of desired rules to facilitate the layout design to adjust the IC feature to improve wafer productivity and enhance throughput.

The remainder of this paper is organized as follows. Section 2 introduced the background of data mining for semiconductor manufacturing and overall wafer effectiveness. Section 3 proposes the approach to derive alternative feature designs. Section 4 conducts an empirical study in semiconductor manufacturing. Section 5 concludes this study with discussion on contribution and future research directions.

II. BACKGROUND

Before further discussion, the notations used are shown as follows.

A_t: Total wafer area
A_a: Available area on wafer
A_e: Exposable area on wafer
A_{gross}: Gross die area on wafer
A_{good}: Good die area on wafer
l_c: length of a chip
w_c: width of a chip.
a_c: area of chip, denoting. a_c=$w_c \times l_c$
k_h: horizontal die number exposed within a field
k_v: vertical die number exposed within a field
c_h: horizontal width of scribe line
c_v: vertical width of scribe line
l_{mf}: length of a mask field
w_{mf}: width of a mask field
l_{ef}: length of an exposure field, let $l_{ef} = k_v \times (l_c + c_v)$
w_{ef}: width of an exposure field, let $w_{ef} = k_h \times (w_c + c_h) + c_h$
T: the set of data at the node
T_i: the sets of data from splitting the node by the selected attribute, $i = 1, \cdots, I$, where I is the number of branch.
S_T: the standard deviation of T
S_{T_i}: the standard deviation of T_i
x_k: IC feature, $k = 1, \cdots, K$, K is the number of IC feature
β_k: the regression coefficient
p_S: the smoothed predicted value for the internal node S
p_{S_i}: predicted value at branch S_i of the internal node S
m_S: predicted value by the linear model at internal node S
n_i: number of training data at branch S_i of internal node S
k: the smooth constant
$\hat{y}_i$: the predicted value of ith response
$\bar{y}$: the response mean value based on the model
N: the total number of sampled data.

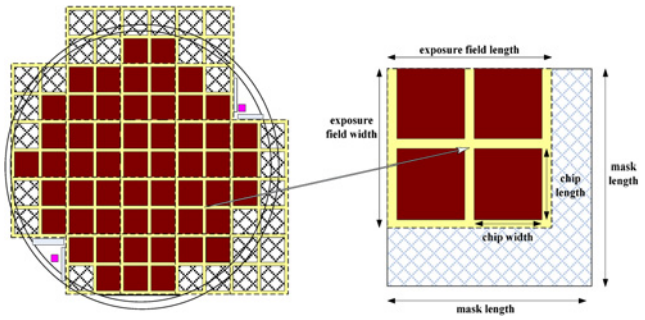

Fig. 1. The wafer exposure pattern.

A. Wafer Exposure Pattern

The IC designer will design the IC feature large enough to accommodate the microcircuitry of ICs to ensure desired functionality. Then, IC designers will issue new tape-out including the information of IC feature and pattern layout size to order wafer fabrication service. According to the designed IC feature, photolithography engineers will determine a wafer exposure pattern with needed exposure shots to gain a number of gross dies. Iterative cutting algorithms [9,10] have been developed to maximize the number of gross dies exposed on a wafer while minimizing the number of exposure shots, subject to throughput, profit, and yield such as the flat bottom line, dummy pattern, the sharing rules of the dummy pattern, the width of the scribe line, the position of the alignment mark and the mask size. Moreover, the WPH can be increased through reducing the # shot of exposure machine (i.e., stepper or scanner). Dummies are used on a wafer where a full exposure field cannot be generated due to exposure constraints [11], in which a dummy can be generated by shielding a part of mask. In addition, a technique of exposure field extension was developed to reduce the shot number around the wafer edge that cannot yield complete gross dies [12]. As shown in Fig. 1, the utilization of mask field implies the efficiency of each shot [8] that, given the same exposable wafer space, the feature of IC design with higher mask field utilization will need fewer # shot and thus result in higher WPH. However, the IC feature with higher mask field utilization not necessary achieves more gross dies than those with lower mask field utilization. Therefore, it is crucial to find the optimal design of IC feature for exposing few # shot per wafer while maximizing # GD simultaneously. Although algebraic equations were developed to estimate the numbers of gross dies [13], [14], the existing approaches cannot provide specific suggestions to help IC designers to determine the optimal designs of IC features for overall wafer effectiveness.

B. Economies of Scale and Economies of Speed

Semiconductor manufacturing companies continuously compete on the basis of achieving economies of scale and economies of speed since semiconductor industry is capital intensive. Economies of scale are required to enhance capital effectiveness via enhancing the capacity and equipment utilization. Economies of scale can be achieved by enhancing overall wafer effectiveness [8] and the implementation of lean

practices and reduction of waste [15]–[18] such as improving throughput by reducing equipment downtime and improving process yield by reducing process excursions. Moreover, the growth of optical proximity correction and resolution enhancement techniques has increased the cost of photomask generation that should be controlled for profitability [19].

Economies of speed are critical as semiconductor products in consumer electronics era compete with increasingly shortening time-to-market [3], in which unit price of the ICs is exponentially reduced in short product life [20]–[24]. Economies of speed that are associated with revenue generation depending on the unit price of products to be sold [21] can be achieved by reducing cycle time of fabrication [4], [5], [23] and accelerating the yield learning rate [6], [24], [25]. Thus, semiconductor fabs strive to maintain competitive advantages by reducing cycle time, enhancing the yield, and fast ramping up to volume production under capacity constraint to reduce the average unit cost to improve returns of capital investments.

This study addressed the issues of economies of scale and economies of speed by reducing unit cost via maximizing the number of gross dies per wafer and improving the throughput via reducing the number of shots per wafer.

C. Data Mining for Manufacturing Intelligence

Many data mining approaches have been developed to extract manufacturing intelligence from the data collected from the information systems from shopfloor-level to enterprise-level to enhance the decision quality [3]–[6], [26]. Although some analytic applications provided by the vendors can be used for cost reduction, productivity improvement, quality improvement and on-time delivery, a lot of information and intelligence are hidden behind the huge data that are automatically collected in the production system [27].

The methodologies for design for manufacturing are developed to ensure that a product can be manufactured repeatedly, consistently, reliably, and cost effectively by considering the criteria needed for the manufacturing objectives starting at the conceptual design stage and implementing these objectives across the design, manufacturing, and assembly processes [7]. With the increasingly complicated IC design and advanced semiconductor manufacturing technology, embedded intelligence in the on-line system to provide design guidelines to prevent further defect and improve production effectiveness has become increasingly important.

D. Overall Wafer Effectiveness

To capture the overall wafer exposure performance for identifying and analyzing hidden performance loss, OWE, was proposed for measuring overall wafer exposure effectiveness [8]. The total wafer area can be divided into five basic wafer area statuses, including total wafer area (A_t), available area (A_a), exposable area (A_e), gross die area (A_{gross}), and good die area (A_{good}). As shown in Fig. 2, the key blocks of the wafer area could particularly associate with the basic statuses and sub-statuses to achieve the wafer area and track resolution for the improvement of wafer exposure effectiveness. *OWE* indices can be defined with wafer area statuses and related

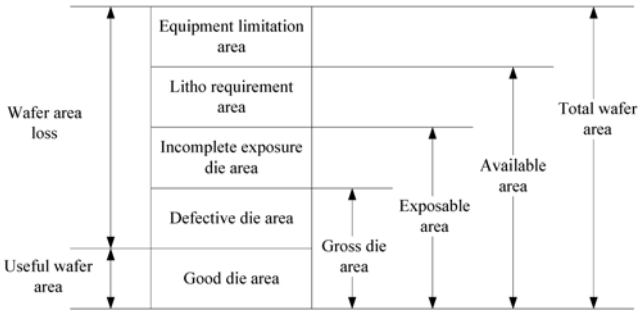

Fig. 2. Conceptual framework of OWE [8].

factors. In particular, *OWE* is the fraction of good die area to total wafer area and is expressed as follows:

$$OWE = \frac{\text{gooddiearea}}{\text{totalwaferarea}} \times 100\% \tag{1}$$

where total wafer area means that the whole area of raw wafer determined by the used wafer size and good die area is the area of produced dies that passed wafer probe test. *OWE* can identify different types of wafer area losses according to the equipment, lithography technology, exposure pattern, and production variation. By considering different wafer area loss, *OWE* can also be expressed by availability efficiency (E_A), litho efficiency (E_L), rate efficiency (E_R), and yield efficiency (E_Y) as follows:

$$OWE = E_A \times E_L \times E_R \times E_Y \tag{2}$$

where

$$E_A = \frac{\text{availablearea}}{\text{totalwaferarea}} \times 100\% \tag{3}$$

$$E_L = \frac{\text{exposablearea}}{\text{availablearea}} \times 100\% \tag{4}$$

$$E_R = \frac{\text{grossdiearea}}{\text{exposablearea}} \times 100\% \tag{5}$$

$$E_Y = \frac{\text{gooddiearea}}{\text{grossdiearea}} \times 100\% \tag{6}$$

Although maximizing *OWE* has an important industrial and commercial value [8], little research has been done to investigate various improvement directions systematically through the *OWE* indices. Considering wafer exposure in practice, the total wafer area can be divided into several kinds of area as shown in Fig. 3. The useless wafer area can be classified into equipment limitation loss, litho requirement loss, incomplete exposed die loss, and defective die loss with the corresponding area statuses. According to the information of decomposed indices, decision makers can follow specific directions to reduce the corresponding wafer area loss and enhance the effectiveness of *OWE* indices, respectively.

III. Proposed Approach

Fig. 4 shows the proposed framework including four phases: problem definition, data preparation, IC feature model construction, and result evaluation and implementation.

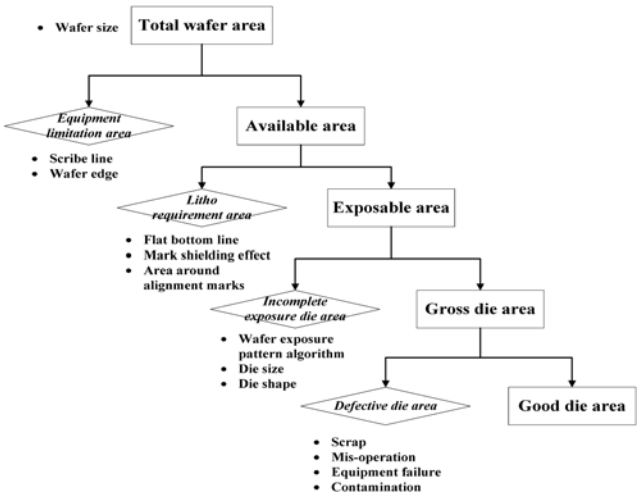

Fig. 3. Breakdown of total wafer area [8].

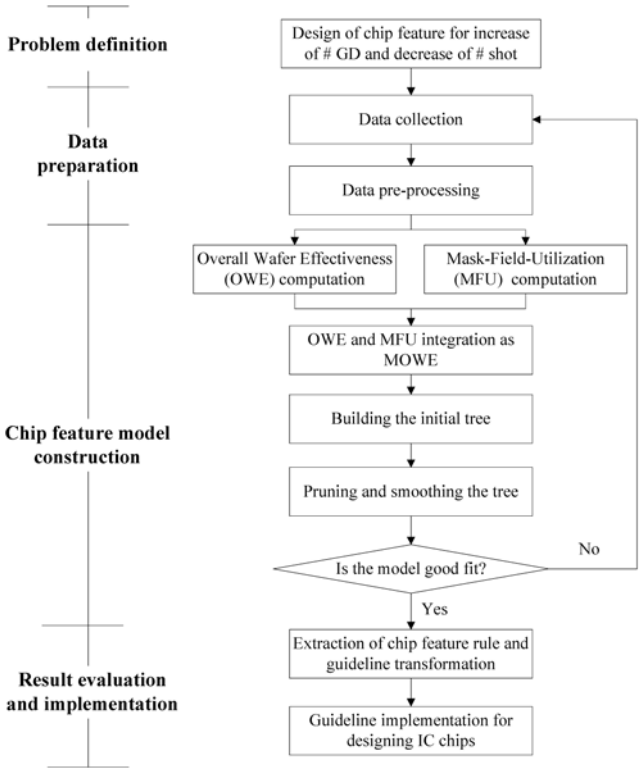

Fig. 4. Research framework for optimizing feature designs of chips.

A. Problem definition

To enhance the competitive advantages of wafer fabrication with the trend of micro profit, cost reduction through increasing the wafer throughput including gross die number per wafer and WPH is important. The existing studies have proposed an algorithm to maximize throughput based on the fixed IC feature in fabrication phase [9], [10]. The IC feature is a crucial factor for the number of gross die and WPH. However, the effect of the optimal feature for maximum # GD per wafer and required # shot per wafer was not considered during the design phase.

The causal relationships between the IC feature, and # GD and # shot, should be identified for the IC designer in advance. However, the improved direction is difficult to find out because of the two-dimensional nonlinear relationship between # GD, # shot, and the feature. The relationships can not be formulated directly by a physical and mathematical equation. Both of the models, # GD and # shot, are varied with the IC feature. As shown in Fig. 4, each type of the chip layout has its particular model of # GD and # shot. If we want to adjust the feature to obtain more gross dies and fewer # shot than the original feature, it could be located in certain regions based on some design limitation. However, it cannot be arbitrarily adjusted step by step because there are thousands of alternative feature designs. In practice, the optimal designs with maximum # GD and minimum # shot are not always the best ones, while the modification of best chip layout design is difficult and time-consuming. As shown in Fig. 5, we can select the feature design with acceptable wafer productivity rather than selecting the best chip feature according to some design limitation, e.g., the chip area is fixed or changed.

The # GD and # shot are mainly influenced by chip feature including chip length and chip width due to the geometric characteristics and the layout pattern algorithm. Therefore, according to the same layout pattern algorithm, the chips are in the same area, yet with different shapes, i.e., the combination of chip lengths and chip widths, which could result in different # GD and # shot. In particular, given the limitations of designed products, the designer can select alternative features that can expose more # GD and fewer # shot than the original design does. As shown in Fig. 6, the origin chip layout design is to expose two dies along the x-direction and one die along the y-direction, which will produce 17 gross dies with 19 # shot. Patterns P1 and P2 adjust the chip shape given the same chip area, which can not only produce extra die but also use fewer # shot. Furthermore, while the chip area can be changed, more features could be designated. Pattern P3 scaled the original size with larger MFU and resulted in 9 shots, yet losing a die. Finally, Pattern P4 adjusted the size with different orientations and gained extra 6 dies via 9 shots. Therefore, the IC feature should consider not only the chip width and chip length but also the chip area and chip shape. However, the effect of IC feature for # shot and # GD in wafer fabrication was usually considered during the IC design process. The meaningful information is hidden behind the huge data which have been recorded during the design and manufacturing process.

B. Data preparation

The database of IC design features and wafer exposure are needed to integrate, in which generally consists of amount of data including missing, noisy, inconsistent data [4]. The designed features often have two-dimensional effect on the manufacturing response variable. For example, the chip area which is composed of chip width and chip length significantly influences the gross die per wafer.

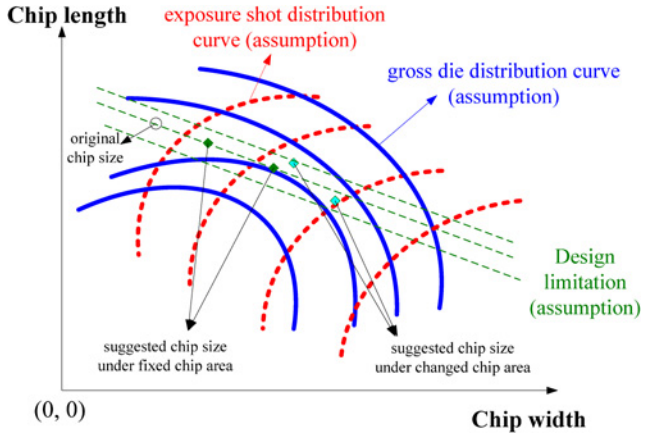

Fig. 5. # GD and # shot with specific design limitation.

This study collected the necessary data from the chip placement database including the gross die number per wafer, exposure times, and the corresponding feature design, including chip length, chip width, and chip area. The components of layout design are located with respect to the two-dimensional coordinate and could be from different data resolution such as wafer level, mask level and chip level. Furthermore, to identify the relevant variables and enhance the construct validity of the feature model, basic statistical analysis and simulation of the causal relationships were used. The collected data often include different kinds of problems and need specific data pre-processing methods for improving mining effectiveness. Data pre-processing focuses on finding data flaw and perform the corresponding treatment including data integration, data cleaning, data transformation, and data reduction [27].

C. Construct the Model to Determine IC Features

1) *Mask-field-utilization weighted OWE (MOWE):* This study employs the proposed *OWE* for measuring the exposure effectiveness of gross numbers and incorporating with utilization of mask field. In particular, the exposed # shot in lithography process is significantly influenced by the utilization of mask field. Therefore, the *MOWE* is applied to consider wafer exposure effectiveness and efficiency and is defined as follows:

$$MOWE = OWE \times MFU \tag{7}$$

In particular, the MFU is represented the utilization of mask field, shown in Fig. 7 and defined as follows:

$$MFU = \frac{l_{ef} \times w_{ef}}{l_{mf} \times w_{mf}} \times 100\% \tag{8}$$

The exposure field is determined by the IC feature (w_c, l_c) and scribe line width (c_h, c_v). For example, a mask field layout with 2 dies along the horizontal and vertical direction on the mask is shown in Fig. 7. Exposure field size along the horizontal and vertical direction can not be larger than the mask field size (w_{mf}, l_{mf}) along the horizontal and vertical direction, respectively.

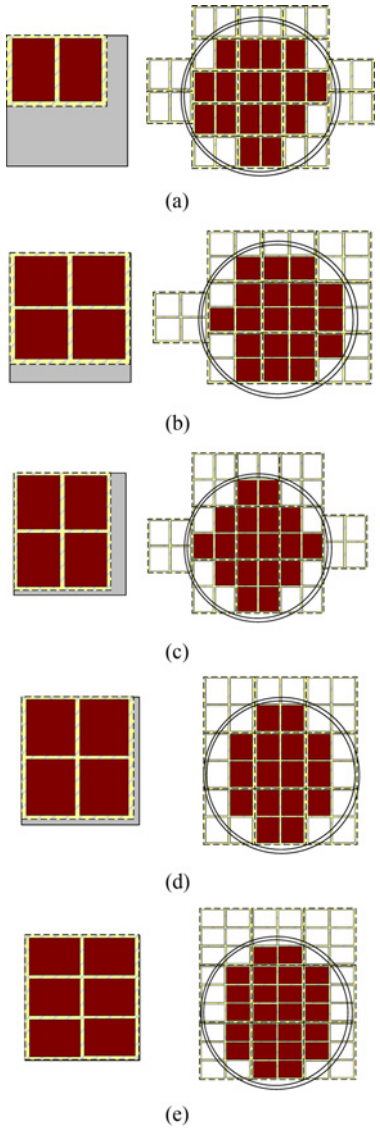

Fig. 6. Illustration of different features, (a) original Pattern (17 GD, 19 shot), (b) Pattern P1 (18 GD, 10 shot), (c) Pattern P2 (18 GD, 10 shot), (d) Pattern P3 (16 GD, 9 shot), (e) Pattern P4 (24 GD, 9 shot).

The integrated *MOWE* is the response variable and the features are independent variables that are determined at the design phase. In addition, the model effectiveness is evaluated based on other random sample data.

2) *Model tree:* The model for determining IC feature is to construct the relationship between alternative feature designs and the *MOWE*. Artificial neural network (ANN) techniques have been used for prediction and modeling the complex relationship. ANN is an technique to statistical regression, time series analysis due to its advantages including higher accuracy, less prior knowledge of model construction, and less data assumptions. However, the result of ANN structure is difficult to interpret the relationship and to provide understandable rules or and formulas. Moreover, the determination of parameters in ANN such as learning rate, momentum, number of hidden layers, and number of neurons is time-consuming via trial and error.

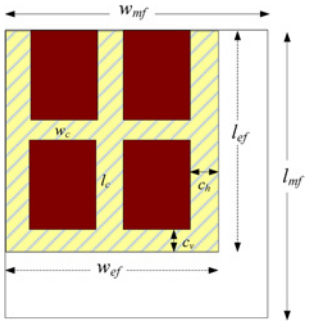

Fig. 7. Mask field layout and mask-field-utilization [8].

In this study, model tree [28] is used to construct the structural representation of the *MOWE* and the feature and transform the relationship into a series of IF-THEN rules by the recursive partitioning. Model tree is an extension of regression tree and the tree growth model is similar to CART (classification and regression tree) [29]. Both of them develop a tree structure by performing recursive partitioning for tree growth. In particular, the CART selects the splitting attributes that give the greatest expected reduction of either variance or absolute deviation. The splitting criterion of model tree is used maximum expected error reduction for determining which attribute is the best to split the training data at a particular node. Model tree predicts the continuous class through building a piecewise linear model at the leaf node instead of using a constant value for predicting continuous class.

Comparing the result representation with regression tree, model tree can generate smaller tree structures and more clear expression of relations [28]. Comparing the prediction accuracy with ANN, model tree can obtain the similar prediction performance in low root mean square error (RMSE) [30–32]. Furthermore, model tree can be trained easily and the results can be expressed as understandable rules and equations for knowledge interpretation and implementation.

In particular, M5′ model tree [33] is applied to construct the relationship between the *MOWE* values and the corresponding feature. Indeed, M5′ model tree is the reform of original M5 model tree [28] with two main construction stages. Firstly, a tree induction is recursively partitioned by minimizing the intra-subset variation of the target variable. Secondly, the tree is pruned back by replacing sub-trees with linear regression functions to minimize the estimation error. As shown in Fig. 8, chip width and chip length at different size are build the tree structure model for *MOWE* prediction. In particular, the tree structure is equivalent to the feature state space. The sub-trees are branched from the root node based on the splitting criteria. Each leaf is an independent model by a linear regression function. For example, if $l_c \leq 3$, $w_c \leq 2.5$, then the *MOWE* value is attached to Model 1.

In building the initial IC feature design model, a training dataset enters the root node of the tree, a test is applied to search all the possible splits for predictor attributes, choose the best discrimination among the *MOWE*, and determine which child node will encounter next recursively. The tree grows the branch by the splitting criterion, in which the descendent nodes are with smaller standard deviation than the sample in the parent node. The standard deviation of a set of samples is regarded as a measurement error at that node. In particular, the attribute which has the maximum expected error reduction is selected by the Standard Deviation Reduction (SDR) as follows:

$$SDR = S_T - \sum_i \frac{|T_i|}{|T|} \times S_{T_i} \tag{9}$$

where T is the set of data at the node, and $T_1, T_2, \cdots, T_k$ are the sets of data which result from splitting the node by the selected attribute. S_T is the standard deviation of the set of data T, and S_{T_i} is the standard deviation of the set of sample T_i. Based on the splitting criterion, the recursive partitioning process is not terminated until the fraction of standard deviation of T reaches the user-defined threshold, or the minimum sample size of the leaf is reached at the node. The linear multivariate regression model is built at each terminal node as $LM(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k$, where $x_1, x_2, \cdots, x_k$ are the feature attribute values and $\beta_0, \beta_1, \cdots, \beta_k$ are the regression coefficients calculated by the least square method. The branch nodes are pruned into leaf nodes if the estimated error of pruning tree is lower. The pruning process applies the estimate of expected error at each node for the testing dataset. The expected error for the unseen sample is multiplied by $(n + v)/(n - v)$ to compensate the underestimated expected error, where n is the number of training data for the node and v is the number of parameter in the linear regression model. Thus, the sub-tree will be pruned back from a leaf node and simplify the linear model in place for the internal node as long as the expected estimation error decreases.

The predicted value of unseen data is used in a smoothing process to compensate the sharp discontinuities that will inevitably occur between adjacent linear models at the leaves of the pruned tree rather than using the raw value directly [33]. In particular, the prediction accuracy could be improved by smoothing effect for the model constructed from few training sample for improving. The predicted value is smoothed along the branch back to the root by integrating the predicted values from the linear model at that node. The smoothed predictive value p_S for the internal node S is calculated as follows:

$$p_S = (n_i p_{S_i} + k m_S)/(n_i + k) \tag{10}$$

where p represents the predicted value at the branch S_i of internal node S, m_S is the predicted value by the linear model at the internal node, n_i is the number of training sample at branch S_i of internal node S, and k is the smoothing constant.

D. Result evaluation and implementation

To estimate the validity of the proposed approach, this study applies RMSE and relative error as the indices for evaluating

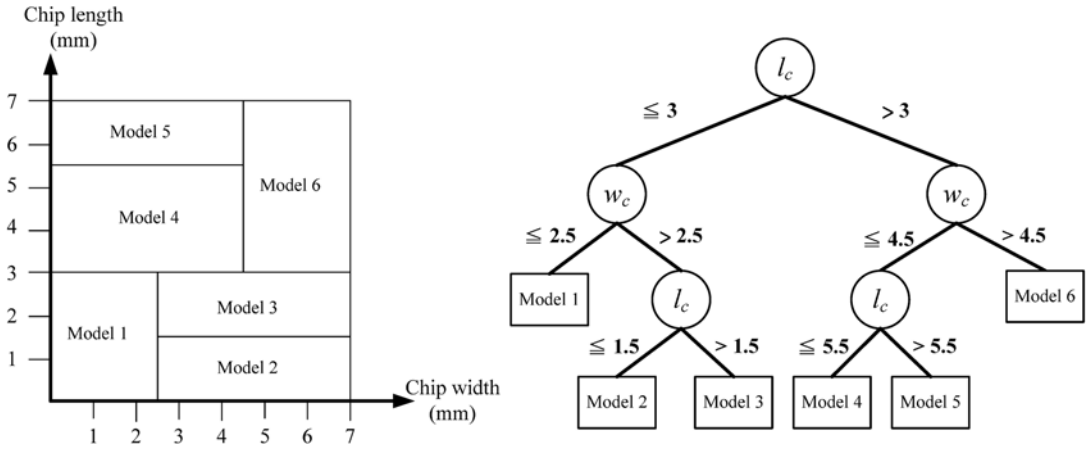

Fig. 8. Example of model tree for deriving IC feature design rules.

the IC feature design model. RMSE represented the prediction performance as follows.

$$RMSE = \sqrt{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \Big/ N} \tag{11}$$

where $\hat{y}_i$ is the predicted value of ith response and N is the total number of sampled data. Lower RMSE value represents the higher accuracy for predicting the wafer exposure effectiveness. Another useful measure is relative error, which is denoted as the ratio of the variance of the residual to the variance of the response mean value [28], which represents the explanation ability if model based on the feature model. The value of relative error should be as small as possible and used for evaluation the effectiveness of model. Relative error is defined as follows:

$$relativeerror = \frac{E\,(y_i - \hat{y}_i)^2}{E\,(y_i - \bar{y})^2} \tag{12}$$

According to the constructed IC feature model, the alternative designs of feature can be adjusted lightly to increase *OWE* and *MFU* simultaneously during the design phase based on the searching procedure as shown in Fig. 9. Firstly, the original IC feature, scribble line width, mask size are and the feature limitations of chip design such as the fixed chip area have to be considered. Next, in order to provide transparent feature area which can capture more # GD, and lower # shot, we produce a set of alternative feature designs and estimate their *MOWE*, *OWE* and *MFU* based on the constructed model. Then, the candidate features are selected based on the thresholds. Finally, all candidate feature sets are visualized through two-dimensional plot to show the alternative improvement directions.

Given the constraints of the layout design, the maximum *MOWE* can be identified by the suggested combination of feature within a range. In the real setting, IC designers do not need the optimal IC feature design for maximizing *MOWE*, and require to capture the recommendatory area where the chip feature can result in acceptable or approximate maximum

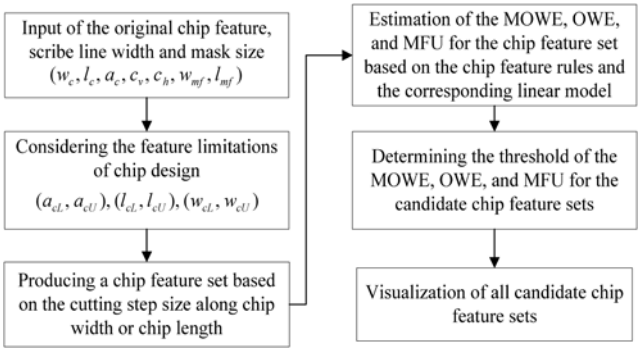

Fig. 9. Procedure of searching advised IC features.

MOWE. For example, a generic product A in a wafer fabrication is illustrated as shown in Fig. 10, in which the original IC feature is (8.922 mm × 8.981 mm) and the used mask size is 26 mm × 33 mm for 300 mm wafer. The original feature in *OWE* and *MFU* are 83.15% and 57.29%, respectively. Six different types of design with increase of *MOWE* are compared with the original feature. First, Patterns (i) and (ii) are within the same chip area, yet the chip length and chip width are varied. Both Pattern (i) and (ii) in *MOWE* and *MFU* are larger than the original feature that increase the exposure field in each exposure. The *OWE* of Pattern (i) is lower than the original. Furthermore, suppose the chip area is not fixed, the die area can be adjusted more extensively with some limitations and results in better *OWE*, and *MFU*. Patterns (iv) and (vi) can get the design with higher *OWE* and *MFU*. Patterns (iii) and (v) both capture higher *OWE* with losing a scrap of *MFU*. Although Patterns (i) and (v) do not increase both *OWE* and *MFU*, but they provide the design with maximum *MFU* and maximum *OWE*, respectively. If the wafer fabrication needs to increase WPH, Pattern (i) can be used to minimize the # shot with losing a scrap of # GD. On the other hand, Pattern (v) results in more # GD and reduction the die cost per wafer even through the increase of MFU. Indeed, the results provide a trade-off for design of IC feature between cost reduction and throughput improvement.

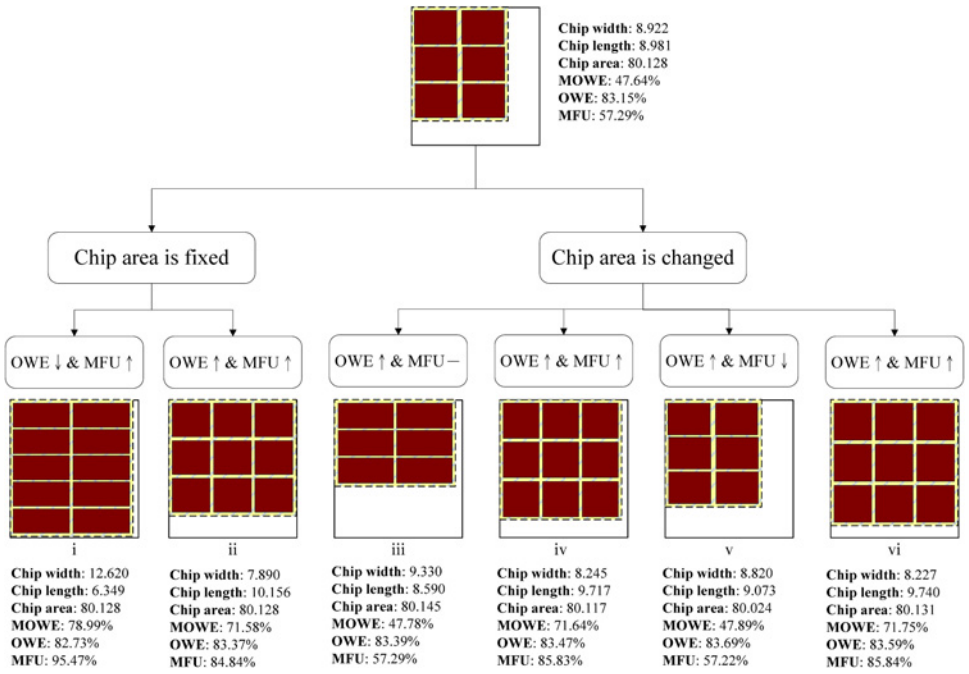

Fig. 10. Feature adjustment with different conditions.

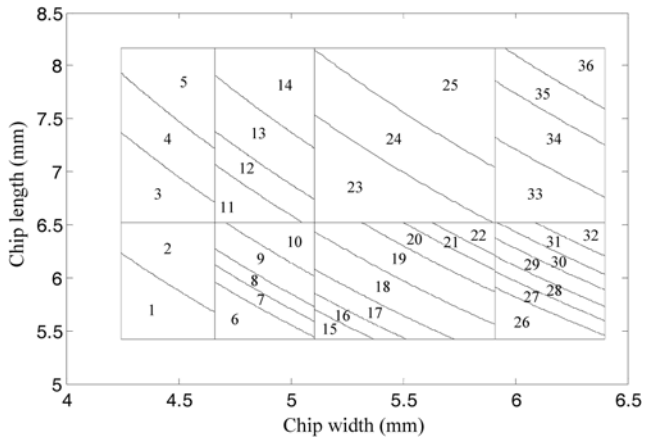

Fig. 11. Two-dimensional IC feature design rules.

TABLE II
MODEL EVALUATION BY M5′ (36 RULES)

	Training dataset	Testing dataset
RMSE (%)	0.0679	0.0796
Relative error (%)	0.0076	0.0177

IV. EMPIRICAL STUDY

A. Problem definition

An empirical study was conducted in a leading semiconductor company in Taiwan to validate the proposed approach. All of the data were collected from the 12 inch wafer fabrication, in which the width of scribe line and mask area were based on the practical die placement exposure algorithm [9,10]. Without losing generality, the data have been systematically transformed for proprietary information protection of the case company.

TABLE III
MODEL EVALUATION BY CART (742 RULES)

	Training dataset	Testing dataset
RMSE (%)	0.1172	0.1351
Relative error (%)	0.0366	0.0510

B. Data preparation

To construct the model for representing the relationship between *MOWE* and the chip design size, total 53561 data are collected from the square region where the chip width is collected from 4.24 to 6.4 and the chip length is collected from 5.42 to 8.56. The chip length, chip width and its corresponding *OWE* and *MFU* per wafer are recorded. Moreover, the related area for equipment exposure limitation including the mask size, width of scribe line, alignment mark, and wafer edge were collected. In order to increase the prediction accuracy of the model, the resolution of data intervals would be cut strictly. According to the stratified sampling strategy for covering the whole two-dimensional pattern, 33474 data are collected for constructing the model and 20087 data are randomly sampled for validation. Therefore, the *MOWE* is calculated through the above data and used as the response variable. IC features including chip length, chip width, and chip area are used for the independent variables.

C. IC Feature Model Construction

As the geometric characteristics of two-dimensional layout design and the *MOWE* are steep, the M5′ model tree is used to extract the sharp boundary and perform the model fitness well. In particular, the program of M5′ model using WEKA version 3.6.3 [34] (http://www.cs.waikato.ac.nz/ml/weka) was

TABLE IV
BASIC INFORMATION OF GENERIC PRODUCT

w_c (mm)	l_c (mm)	c_h (mm)	c_v (mm)	w_{mf} (mm)	l_{mf} (mm)	*MOWE*(%)	*OWE*(%)	*MFU*(%)
5.20	7.01	0.08	0.08	26	33	61.088	87.182	70.069

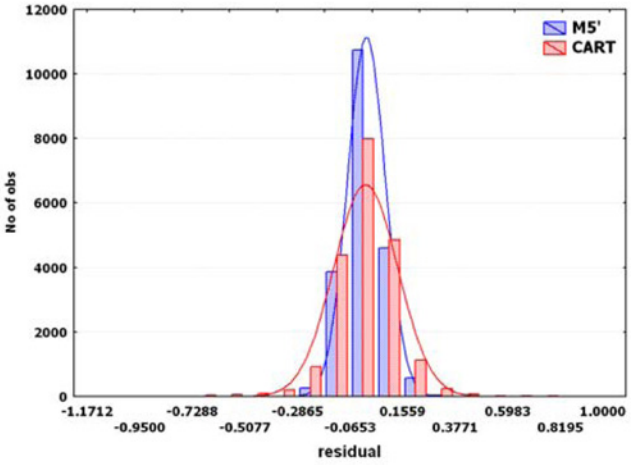

Fig. 12. Histogram of residuals of M5′ and CART.

employed in this empirical study. The minimum number at each leaf is set 30. The tree is grown and split into 36 leaves, in which each leaf is represented by a linear model. The generated IF-THEN rules and linear models which are listed in Table I, can be expressed in a specific area and provide useful information for understanding the relationships between *MOWE* and IC features. For example, rule 1 means "IF a chip length is larger than 4.24 and smaller than 4.66, and the chip width is larger than 5.42 and smaller than 6.52, and chip area is larger than 22.98 and smaller than 26.43, THEN the MOWE of the designed chip is inferred by model 1.

Furthermore, a two-dimensional plot is used for visualization of the derived 36 feature design rules in Fig. 11. Each region is framed with the contour line. The variation of *MOWE* at different features can be visualized. The region where the chip width locates between 5.10 and 6.4 and the chip length locates between 5.42 and 6.52 is shrinking than others. The two-dimensional visualization plot can make IC designers easier to understand. The IC feature can also provide potential improvement direction for designers while the *MOWE* of chip design is low.

As shown in Table II, the RMSE of M5′ model tree for training dataset and testing dataset are 0.0679% and 0.0796%, respectively. The relative error rates are 0.0076% for the training dataset and 0.0177% for the testing dataset, which means that the use of chip length, chip width and chip area can predict *MOWE* with small tolerance.

To evaluate the performance of the M5′ model tree, CART is used based on the same training dataset and testing dataset. The IC feature design rules of CART are approximated by means of sample data at each leaf node. The criteria for stopping the tree growth include that the variance reduction

TABLE V
CHIP FEATURE ADJUSTMENT WITH DIFFERENT GROUPS

Groups	Involved rules	Average *MOWE* (%)	Average *OWE* (%)	Average *MFU* (%)
(a)	5, 13	76.164	86.940	87.605
(b)	21, 28	76.260	87.135	87.520
(c)	14	77.536	87.206	88.911
(d)	23	61.657	87.206	70.702
(e)	21	75.273	87.191	86.331
(f)	28	75.456	87.213	86.519
(g)	29	77.121	87.210	88.431

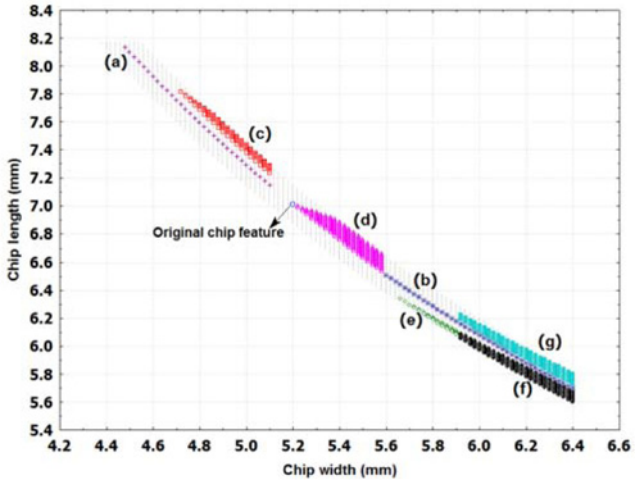

Fig. 13. The advised design of chip feature for generic product B.

is less than 0, the number of data in a leaf node is less than 30, and all the possible splits go to the same child node. As summarized in Table III, total 742 rules are derived from CART by the training dataset, and the result of RMSE and relative error rate are 0.1172% and 0.0366%, respectively. The same testing dataset is used as well as M5′ model tree. As shown in Table II and Table III, the RMSE and the relative error of M5′ model tree are lower than CART. Fig. 12 shows the histogram of prediction error of total 20087 testing chip design data through M5′ model tree and CART. The horizontal axis represents the magnitude and direction of the prediction error, and the vertical axis represents the numbers of counts. The prediction errors of M5′ model tree are symmetrical and more central toward the zero than the CART. The results demonstrate that M5′ model tree not only uses fewer rules but also have better prediction performance than CART.

1) *Result evaluation and implementation:* After constructing the rules and linear models for the modelling region, this study selects a generic chip device to evaluate the model

TABLE VI
VALIDATION RESULTS OF EMPIRICAL PRODUCT

#	Original				Advised	Advised	Improved			Increase	Decrease
	chip feature (mm×mm)	*MOWE* (%)	*OWE* (%)	*MFU* (%)	chip width (mm×mm)	chip length (mm×mm)	*MOWE* (%)	*OWE* (%)	*MFU* (%)	of # GD	of # shot
A	5.0×7.0	70.23%	83.51%	84.10%	5.54~6.35	5.51~6.32	71.20%	84.46%	84.30%	2	1
B	12.0×10.0	70.28%	82.30%	85.40%	8.53~12.84	9.35~14.07	70.76%	82.76%	85.50%	3	1
C	22.1×9.3	57.69%	79.03%	73.00%	12.63~21.92	9.41~16.36	78.88%	81.32%	97.00%	6	24
D	19.0×19.0	33.07%	78.74%	42.00%	22.75~24.32	14.84~15.87	66.56%	78.77%	84.50%	1	80

effectiveness and demonstrate the implementation. The basic information of generic product B including IC feature, scribe line width, and the mask feature are listed in Table IV. The step size along chip width is set as 0.02 mm to produce the set of alternative designs. The candidate feature designs are surpassed in the *MOWE*, *OWE*, and *MFU* for original design. However, no candidate feature is qualified based on the determined threshold. It means that the feature designs of increasing both *OWE* and *MFU* cannot be found under the original chip area. As listed in Table V, under the original chip area and different ratio of chip lengths and chip widths, the chip design in groups (a) and (b) can still get higher *MOWE* and *MFU* than original one with loss of *OWE*. In this case, the designer can select the group (b) for improving *MOWE* with losing a scrap of *OWE* or extend the limitations of IC feature for capturing overall improvement. The extension of IC feature set is produced by enlarging and narrowing chip area within the 2%. As shown in Fig. 13 and Table V, five qualified groups of IC feature designs can be found through the involved rules of IC feature designs. In particular, Groups (c) and (g) provide a superior overall performance than the other groups. Group (f) can get the largest *OWE* result. Through the constructed model for determining IC feature designs and searching procedure, the region of candidate features can be identified with the requirements. Once the limitations of IC feature are changed, the designer can also quickly find the alternative design in specific region from the two-dimensional visualization plot.

Four real products were selected in a wafer fabrication to evaluate the validity of the proposed approach. Based on the constant chip area, and the suggested designs of IC features from the decision support system were provided. As shown in Table VI, the MOWE can be improved by the advised IC feature designs. In particular, the lightly improved MOWE for product A and B can capture extra 2 dies and reduce 1 shot per wafer. The advised layout size of product C can obtain 6 dies (2% OWE) and reduce 24 shots (24% MFU) per wafer simultaneously. Furthermore, the product D with bad MFU will lead to more shot numbers for wafer exposure. The advised IC feature can enhance the MFU and decrease 80 shots at wafer exposure process. Furthermore, the constructed design rules of IC feature have been applied to effectively derive better chip layout designs in a 300 mm wafer fabrication. The extra WPH gain is estimated to be 14.57% and the benefit is about US$73 M dollars per year for the wafer fabrication. The constructed rules for IC feature designs have been implemented in the on-line system in a wafer fabrication for IC designer.

V. CONCLUSION

This study developed a novel approach to enhance overall wafer effectiveness by employing data mining to derive the optimal IC feature designs to provide useful advice to IC designers to proactively reduce average die cost by improving wafer productivity via maximizing gross dies and reducing exposure shots. In particular, this study integrates the multi-objectives of # GD and # shot by MOWE. Furthermore, the nonlinear relationship between the two-dimensional layout design and *MOWE* were extracted as understandable rules for desired IC feature regions. Indeed, the developed solution has been implemented in a leading semiconductor company to provide useful guidelines for IC designers for design for manufacturing. The empirical study has shown practical viability of this approach that can not only reduce the cost by minimizing the exposures per wafer, but also increase the yield by maximizing the distance from good dies to the wafer edge.

Little research has addressed the present problem via employing data mining to extract potentially causal relationships for constructing the model between IC design variables such as IC feature and manufacturing response variables such as yield and throughput to foster design for manufacturing. Further research should be done to develop data mining and manufacturing intelligence approaches to address other issues between design and manufacturing for effective virtual vertical integration of semiconductor supply chains [35]. Furthermore, the proposed approach can provide advice for chip designers to move blocks around to optimize the IC feature design, subject to the constraints of IC architecture that can be addressed in future studies.

Indeed, as the feature size shrunk to nano-meter, new challenges and issues that have not been discovered in previous technology nodes need novel approaches. However, previous domain knowledge and existing models have limitations to interpret unseen phenomenon and deal with new challenges. Future studies for developing novel approaches are needed to address emerging issues for advanced semiconductor manufacturing. Furthermore, maximizing the number of gross dies and reducing the number of shots per wafer via the proposed approach can effectively reduce cost of ownership and enhance throughput of photolithography process. Since profit metrics tend to be represented by the difference between revenue metrics and cost metrics, whereas productivity can be expressed as the ratio of output (revenue) to input (cost). Further research is needed to address fab economics [17–22, 36] involved in advanced semiconductor manufacturing.

VI. Appendix

TABLE I

CHIP FEATURE RULES FOR *MOWE*

Rule	IF						THEN
	$< w_c$	$w_c \leqq$	$< l_c$	$l_c \leqq$	$< a_c$	$a_c \leqq$	
1	4.24	4.66	5.42	6.52	22.98	26.43	$MOWE = 2.125 + 0.026w_c - 0.029l_c + 2.555a_c$
2	4.24	4.66	5.42	6.52	26.43	30.38	$MOWE = 2.417 + 0.026w_c - 0.029l_c + 2.545a_c$
3	4.24	4.66	6.52	8.17	27.64	31.26	$MOWE = 1.811 + 0.026w_c - 0.040l_c + 2.048a_c$
4	4.24	4.66	6.52	8.17	31.26	33.63	$MOWE = 2.596 + 0.026w_c - 0.040l_c + 2.022a_c$
5	4.24	4.66	6.52	8.17	33.63	38.07	$MOWE = 2.806 + 0.026w_c + 0.022l_c + 2.003a_c$
6	4.66	5.10	5.42	6.52	22.98	27.76	$MOWE = 1.788 + 0.025w_c - 0.032l_c + 2.570a_c$
7	4.66	5.10	5.42	6.52	27.76	28.52	$MOWE = 2.547 + 0.025w_c - 0.032l_c + 2.543a_c$
8	4.66	5.10	5.42	6.52	28.52	29.23	$MOWE = 4.812 + 0.025w_c - 0.032l_c + 2.464a_c$
9	4.66	5.10	5.42	6.52	29.23	30.70	$MOWE = 2.025 + 0.025w_c - 0.032l_c + 2.560a_c$
10	4.66	5.10	5.42	6.52	30.70	33.25	$MOWE = 2.743 + 0.025w_c - 0.032l_c + 2.538a_c$
11	4.66	5.10	6.52	8.17	30.38	32.94	$MOWE = 1.989 + 0.025w_c - 0.047l_c + 2.045a_c$
12	4.66	5.10	6.52	8.17	32.94	34.40	$MOWE = 3.403 + 0.025w_c - 0.047l_c + 2.002a_c$
13	4.66	5.10	6.52	8.17	34.40	36.86	$MOWE = 2.268 + 0.025w_c - 0.047l_c + 2.035a_c$
14	4.66	5.10	6.52	8.17	36.86	41.67	$MOWE = 3.024 + 0.025w_c + 0.022l_c + 2.001a_c$
15	5.10	5.91	5.42	6.52	27.64	29.10	$MOWE = 1.623 + 0.001w_c - 0.019l_c + 2.054a_c$
16	5.10	5.91	5.42	6.52	29.10	29.88	$MOWE = 3.994 + 0.001w_c - 0.019l_c + 1.973a_c$
17	5.10	5.91	5.42	6.52	29.88	31.03	$MOWE = 2.437 + 0.001w_c - 0.019l_c + 2.026a_c$
18	5.10	5.91	5.42	6.52	31.03	32.84	$MOWE = 1.605 + 0.001w_c - 0.019l_c + 2.053a_c$
19	5.10	5.91	5.42	6.52	32.84	34.65	$MOWE = 2.717 + 0.001w_c - 0.019l_c + 2.019a_c$
20	5.10	5.91	5.42	6.52	34.65	35.87	$MOWE = 2.190 + 0.001w_c - 0.019l_c + 2.035a_c$
21	5.10	5.91	5.42	6.52	35.87	36.71	$MOWE = 5.652 + 0.001w_c - 0.019l_c + 1.940a_c$
22	5.10	5.91	5.42	6.52	36.71	38.53	$MOWE = 2.455 + 0.001w_c - 0.019l_c + 2.029a_c$
23	5.10	5.91	6.52	8.17	33.25	38.47	$MOWE = 1.779 + 0.001w_c - 0.026l_c + 1.632a_c$
24	5.10	5.91	6.52	8.17	38.47	41.60	$MOWE = 2.363 + 0.001w_c - 0.026l_c + 1.617a_c$
25	5.10	5.91	6.52	8.17	41.60	48.28	$MOWE = 2.646 + 0.001w_c + 0.016l_c + 1.603a_c$
26	5.91	6.4	5.42	6.52	32.03	34.91	$MOWE = 3.376 + 0.006w_c - 0.037l_c + 2.003a_c$
27	5.91	6.4	5.42	6.52	34.91	35.81	$MOWE = 2.716 + 0.006w_c - 0.037l_c + 2.022a_c$
28	5.91	6.4	5.42	6.52	35.81	36.66	$MOWE = 5.482 + 0.006w_c - 0.037l_c + 1.947a_c$
29	5.91	6.4	5.42	6.52	36.66	37.67	$MOWE = 5.901 + 0.006w_c - 0.037l_c + 1.938a_c$
30	5.91	6.4	5.42	6.52	37.67	38.60	$MOWE = 6.729 + 0.006w_c - 0.037l_c + 1.919a_c$
31	5.91	6.4	5.42	6.52	38.60	39.71	$MOWE = 5.409 + 0.006w_c - 0.037l_c + 1.955a_c$
32	5.91	6.4	5.42	6.52	39.71	41.73	$MOWE = 4.593 + 0.006w_c - 0.037l_c + 1.977a_c$
33	5.91	6.4	6.52	8.17	38.53	43.27	$MOWE = 2.220 + 0.006w_c - 0.037l_c + 1.625a_c$
34	5.91	6.4	6.52	8.17	43.27	46.44	$MOWE = 3.371 + 0.006w_c - 0.053l_c + 1.599a_c$
35	5.91	6.4	6.52	8.17	46.44	48.60	$MOWE = 3.842 + 0.006w_c + 0.026l_c + 1.576a_c$
36	5.91	6.4	6.52	8.17	48.60	52.29	$MOWE = 1.965 + 0.006w_c + 0.026l_c + 1.615a_c$

REFERENCES

[1] P. Silverman, "Capital productivity: Major challenge for the semiconductor industry," *Solid State Technol.*, vol. 37, no. 3, p. 104, 1994.

[2] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, pp. 114–117, 1965.

[3] C.-F. Chien, Y.-J. Chen, and J.-T. Peng, "Manufacturing intelligence for semiconductor demand forecast based on technology diffusion and product life cycle," *Int. J. Prod. Econ.*, vol. 128, no. 2, pp. 496–509, 2010.

[4] C.-J. Kuo, C.-F. Chien, and J. Chen, "Manufacturing intelligence to exploit the value of production and tool data to reduce cycle time," *IEEE Trans. Autom. Sci. Eng.*, vol. 8, no. 1, pp. 103–111, Jan. 2011.

[5] C.-F. Chien, C.-Y. Hsu, and C.-W. Hsiao, "Manufacturing intelligence to forecast and reduce semiconductor cycle time," *J. Intell. Manuf.*, vol. 23, no. 6, pp. 2281–2294, 2012.

[6] C.-F. Chien, S.-C. Hsu, and Y.-J. Chen, "A system for online detection and classification of wafer bin map defect patterns for manufacturing intelligence," *Int. J. Prod. Res.*, vol. 51, no. 8, pp. 2324–2338, 2013.

[7] C. Chiang and J. Kawa, *Design for Manufacturability and Yield for Nano-Scale CMOS*. Dordrecht, The Netherlands: Springer, 2007.

[8] C.-F. Chien, C.-Y. Hsu, and K.-H. Chang, "Overall wafer effectiveness (OWE): A novel industry standard for semiconductor ecosystem as a whole," *Comput. Ind. Eng.*, vol. 65, no. 1, pp. 117–127, 2013.

[9] C.-F. Chien, S.-C. Hsu, and C. Chen, "An iterative cutting procedure for determining the optimal wafer exposure pattern," *IEEE Trans. Semicond. Manuf.*, vol. 12, no. 3, pp. 375–377, Aug. 1999.

[10] C.-F. Chien, S. Hsu, and J. Deng, "A cutting algorithm for optimizing the wafer exposure pattern," *IEEE Trans. Semicond. Manuf.*, vol. 14, no. 2, pp. 157–162, May 2001.

[11] C.-W. Lin, H.-H. Chou, Y.-J. Wang, C.-F. Chien, J.-H. Wang, and C.-H. Hsiao, "Methods for optimizing die placement," U.S. Patent 735 307 7B2, Apr. 1, 2008.

[12] A. Varela, M. A. Maxim, D. E. Vanlare, and A. Lazar. "Field extension to reduce non-yeilding exposures of wafer," U.S. Patent Application Publication 201 201 626 22A1, Jun. 28, 2012.

[13] A. Ferris-Prabhu, "An algebraic expression to count the number of chips on a wafer," *IEEE Circuits Devices Mag.*, vol. 5, no. 1, pp. 37–39, Jan. 1989.

[14] D. K. de Vries, "Investigation of gross die per wafer formulas," *IEEE Trans. Semicond. Manuf.*, vol. 18, no. 1, pp. 136–139, Feb. 2005.

[15] R. E. Bohn and C. Terwiesch, "The economics of yield-driven processes," *J. Oper. Manag.*, vol. 18, no. 1, pp. 41–59, 1999.

[16] C. S. Tang, "Designing an optimal production system with inspection," *Eur. J. Operat. Res.*, vol. 52, no. 1, pp. 45–54, 1991.

[17] M. A. Lapré, A. S. Mukherjee, and L. N. Van Wassenhove, "Behind the learning curve: Linking learning activities to waste reduction," *Manag. Sci.*, vol. 46, no. 5, pp. 597–611, 2000.

[18] R. E. Bohn, "Noise and learning in semiconductor manufacturing," *Manag. Sci.*, vol. 41, no. 1, pp. 31–42, 1995.

[19] C. M. Weber, C. N. Berglund, and P. Gabella, "Mask cost and profitability in photomask manufacturing: An empirical analysis," *IEEE Trans. Semicond. Manuf.*, vol. 19, no. 4, pp. 465–474, Nov. 2006.

[20] R. C. Leachman, S. Ding, and C.-F. Chien, "Economic effciency analysis of wafer fabrication," *IEEE Trans. Autom. Sci. Eng.*, vol. 4, no. 4, pp. 501–512, Oct. 2007.

[21] C. M. Weber and A. Fayed, "Scale, scope, and speed-managing the challenges of multiproduct manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 23, no. 1, pp. 30–38, Feb. 2010.

[22] C. Terwiesch and R. E. Bohn, "Learning and process improvement during production ramp-up," *Int. J. Prod. Econ.*, vol. 70, no. 1, pp. 1–19, 2001.

[23] R. C. Leachman and S. Ding, "Integration of speed economics into decision-making for manufacturing management," *Int. J. Prod. Econ.*, vol. 107, no. 1, pp. 39–55, 2007.

[24] C. Weber, "Yield learning and the sources of profitability in semiconductor manufacturing and process development," *IEEE Trans. Semicond. Manuf.*, vol. 17, no. 4, pp. 590–596, Nov. 2004.

[25] S. P. Cunningham, C. J. Spanos, and K. Voros, "Semiconductor yield improvement: Results and best practices," *IEEE Trans. Semicond. Manuf.*, vol. 8, no. 2, pp. 103–109, May 1995.

[26] D. Braha and A. Shmilovici, "Data mining for improving a cleaning process in the semiconductor industry," *IEEE Trans. Semicond. Manuf.*, vol. 15, no. 1, pp. 91–101, Feb. 2002.

[27] C.-F. Chien, W.-J. Wang, and J. C. Cheng, "Data mining for yield enhancement in semiconductor manufacturing and an empirical study," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 192–198, 2007.

[28] J. R. Quinlan, "Learning with continuous classes," in *Proc. 5th Aust. Joint Conf. Artif. Intell.*, 1992, pp. 343–348.

[29] L. Breiman, J. H. Friedman, R. J. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA, USA: Wadsworth, 1984.

[30] D. P. Solomatine and K. N. Dulal, "Model trees as an alternative to neural networks in rainfall-runoff modelling," *J. Sci. Hydrol.*, vol. 48, no. 3, pp. 399–411, 2003.

[31] B. Bhattacharya and D. P. Solomatine, "Neural networks and M5 model trees in modelling water level-discharge relationship," *Neurocomputing*, vol. 63, pp. 381–396, Jan. 2005.

[32] A. Etemad-Shahidi and J. Mahjoobi, "Comparison between M5 model tree and neural networks for prediction of significant wave height in lake superior," *Ocean Eng.*, vol. 36, nos. 15–16, pp. 1175–1181, 2009.

[33] Y. Wang and I. H Written, "Inducing model trees for continuous classes," in *Proc. 9th Eur. Conf. Mach. Learn.*, 1997, pp. 128–137.

[34] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2005.

[35] W. Shih, C. Shih, C.-F. Chien, and Y. Chang, "System on a chip 2008: Global Unichip Corporation," Harvard Business School, Boston, MA, USA, Tech. Rep. HBS Case 9-608-159, 2008.

[36] C.-F. Chien, J.-K. Wang, T.-C. Chang, and W.-C. Wu, "Economic analysis of 450 mm wafer migration," in *Proc. ISSM*, 2007, pp. 283–286.

Chen-Fu Chien (M'03) received the B.S. degree (with Phi Tao Phi Honor) with double majors in industrial engineering and electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 1990, and the M.S. degree in industrial engineering and the Ph.D. degree in operations research and decision sciences from University of Wisconsin–Madison, Madison, WI, USA, in 1994 and 1996, respectively. He is a Tsing Hua Chair Professor with NTHU, the Director for the NTHU-TSMC Center for Manufacturing Excellence, and the Principal Investigator for the NSC Semiconductor Technologies Empowerment Partners Consortium, Taiwan. He was a Fulbright Scholar with University of California, Berkeley, CA, USA, from 2002 to 2003 and also received the PCMPCL Training in the Harvard Business School in 2007. From 2005 to 2008, he was on leave as the Deputy Director of the Industrial Engineering Division in Taiwan Semiconductor Manufacturing Company.

His research efforts center on decision analysis, modeling and analysis for semiconductor manufacturing, manufacturing strategy, and data mining. Dr. Chien has received seven invention patents on semiconductor manufacturing, and published three books, over 120 journal papers, and a number of case studies in Harvard Business School. He has been invited to give keynote speech in various conferences, including APIEMS, C&IE, IEEM, IML, KES, and leading universities worldwide. He received the National Quality Award from the Executive Yuan, the Distinguished Research Awards and Tier 1 Principal Investigator (Top 3%) from NSC, the Distinguished University–Industry Collaborative Research Award from the Ministry of Education, the University Industrial Contribution Awards from the Ministry of Economic Affairs, the Distinguished University–Industry Collaborative Research Award, the Distinguished Young Faculty Research Award from NTHU, the Distinguished Young Industrial Engineer Award, the Best IE Paper Award, the IE Award from the Chinese Institute of Industrial Engineering, the Best Engineering Paper Award, the Distinguished Engineering Professor by the Chinese Institute of Engineers in Taiwan, and the 2011 IEEE TASE Best Paper Award. He is an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, the Area Editor of *Flexible Services and Manufacturing Journal*, the Advisory Board Member of *OR Spectrum*, and on editorial boards of several journals.

Chia-Yu Hsu received the B.S. degree in statistics from National Cheng Kung University, Tainan, Taiwan, in 2002, and the M.S. and Ph.D. degrees in industrial engineering and engineering management from National Tsing Hua University, Hsinchu, Taiwan, in 2004 and 2009, respectively. He is an Assistant Professor with Yuan Ze University, Chung-Li, Taiwan. His research works appear in *Computers and Industrial Engineering*, *Flexible Services and Manufacturing Journal*, *Journal of the Chinese Institute of Industrial Engineers*, *Journal of Intelligent Manufacturing*, and *Journal of Quality*. He has received two USA invention patents on semiconductor manufacturing. His current research interests include manufacturing intelligence, data mining, yield enhancement, design for manufacturability, and APC/AEC.