Endpoint Detection in Low Open Area TSV Fabrication Using Optical Emission Spectroscopy

Ja Myung Gu, Paragkumar A. Thadesar, Ashish Dembla, Muhannad S. Bakir, Gary S. May, and Sang Jeen Hong

***Abstract*— Through-silicon via (TSV) technology is a key enabler for 3-D and 2.5-D integration, which provides low-power and high-bandwidth chip-to-chip communication. During TSV fabrication, over-etching may cause notching at the base of the TSVs, resulting in TSV diameter variations. Endpoint detection (EPD) techniques are critical for controlling TSV diameter, and detecting the endpoint for low open areas presents a serious challenge to process engineers. In this paper, a hybrid partial least squares-support vector machine model for optical emission spectroscopy data are successfully demonstrated for an EPD of low open area TSVs. Accurate EPD results are shown for 120, 80, and 25 μm diameter TSVs.**

***Index Terms*— 3-D integration, endpoint detection (EPD), optical emission spectroscopy (OES), through-silicon via (TSV).**

I. Introduction

HISTORICALLY, device scaling has driven low-cost and high-performance computing systems, but its advantages in system performance gains are gradually limited due to the slower rate of development of system interconnection and packaging technologies. Therefore, the demand for disruptive interconnection and packaging technologies has increased in recent years, with particular interest in 3-D and 2.5-D integration of integrated circuits (ICs) [1]. Through-silicon vias (TSVs) play a critical role in enabling 3-D and 2.5-D integration. A schematic representation of an electronic system featuring TSV-based 3-D ICs assembled on an interposer is shown in Fig. 1. The dimensions of TSVs need to be accurately controlled to meet the desired electrical and thermomechanical specifications of a 3-D or 2.5-D system.

To fabricate TSVs, either blind vias are etched first in silicon followed by back side silicon polishing [2], or vias are etched through the wafer thickness with a thin suspended SiO_2 layer at their base [3]. The latter approach alleviates the requirement for silicon polishing after via etching, and it is the focus of this paper. TSV etching is performed using the Bosch process, which consists of alternating etch and passivation cycles [4]. SF_6 is used as the reactive etch gas, and C_4F_8 is used for fluorocarbon sidewall passivation. For given TSV dimensions, the number of etch cycles needs to be accurately determined during silicon via etching to avoid under- or over-etch at the base of the vias. In the case of overetching, positive ions accumulate over the SiO_2 etch stop layer at the base of the vias, causing repulsion to the incoming positive ion flux. The repelled incoming positive ion flux can attack the silicon sidewall at the base of the vias, resulting in notching, as shown in [5, Fig. 2]. The instant during an etching process when a material (silicon) is completely etched without under or overetch is called the etch endpoint and the process of identifying the endpoint and terminating etching is called endpoint detection (EPD). Correct detection of the etch endpoint in TSV improves the control of the critical dimension.

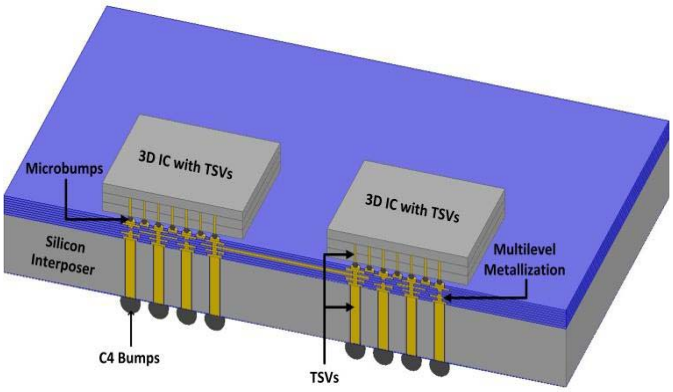

Fig. 1. System including 3-D ICs with TSVs on a silicon interposer with TSVs and multilevel metallization.

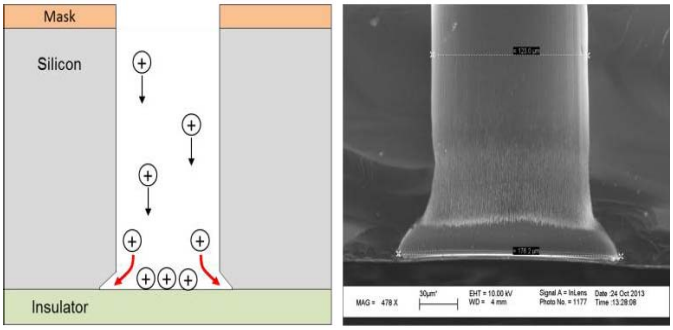

Fig. 2. (a) Notching mechanism in TSV fabrication using Bosch process. (b) Cross section of a 120 μm diameter via showing notching at the base.

Manuscript received February 10, 2014; revised April 26, 2014; accepted April 27, 2014. Date of publication June 3, 2014; date of current version June 30, 2014. Recommended for publication by Associate Editor B. Dang upon evaluation of reviewers' comments.

J. M. Gu and S. J. Hong are with the Department of Electronic Engineering, Myongji University, Yongin 449-728, Korea (e-mail: goojm76@gmail.com).

P. A. Thadesar, A. Dembla, M. S. Bakir, and G. S. May are with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: pthadesar3@gatech.ed; adembla@gatech.edu; muhannad.bakir@mirc.gatech.edu; gary.may@coe.gatech.edu).

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/TCPMT.2014.2323070

EPD has been demonstrated in the literature for the etching of polymer films using a Langmuir probe [6], blanket polysilicon using laser interferometry [7], and chrome dry-etching using RF sensors [8]. However, it is difficult to predict the endpoint during TSV fabrication for two key reasons: 1) the Bosch process used for silicon etching involves time division multiplexing of etch and passivation cycles, requiring incorporation of multiplexing in an EPD methodology and 2) the aggregated TSV area typically consumes <1% of the die area, requiring an EPD methodology capable of detecting a very low quantity of etch byproducts.

EPD for etching a silicon-on-insulator wafer has been demonstrated in [9] by implementing an envelope follower algorithm using optical emission spectroscopy (OES). The EPD method using OES focuses on identifying the wavelength corresponding to a chemical species that shows a pronounced transition at the endpoint. However, the signals for demonstrating the envelope follower algorithm using OES used both the passivation and etch steps. Therefore, the envelope follower algorithm was influenced by the passivation step, where the SiF_x (etch byproduct) intensity includes carbon noise from C_4F_8. Additionally, with respect to EPD for TSV etching using OES, low open areas and high TSV aspect ratios can result in extremely low signal-to-noise ratios and weak intensity of the selected wavelength to be detected.

To perform EPD during the etching of TSVs with low open areas and high aspect ratios, this paper proposes and experimentally validates a hybrid partial least squares-support vector machine (PLS-SVM) model using OES that solely focuses on the silicon etch step during the Bosch process. The proposed hybrid model is successfully validated for the EPD of TSVs with low open areas in their diameters of 120, 80 and 25 μm. This paper is organized as follows. Section II describes the modeling techniques used. Section III describes the sensor system setup and sample preparation. The hybrid PLS-SVM EPD model is described in Section IV. Section V presents the results and verification of the proposed EPD model. While we emphasize TSVs in this paper, the results of the research under consideration can be used in various applications, including MEMS, microfluidics, and electronic device fabrication.

II. PLS and SVM Methodologies

A. Partial Least Squares

PLS regression is a well-known technique in many fields of industrial process control as well as analytical, physical, and clinical chemistry [10]–[12]. The pioneering work of PLS was done by Wold in the field of econometrics in [13]. As a 1-D reduction method, PLS has special processing capability in small sample cases. PLS also adequately employs information independent variables, which is different from the properties of principal component analysis [14]. PLS obtains the most information from independent variables and dependent variables separately and makes the correlation between these two variables the greatest. The principle behind PLS is to search for a set of components (called latent vectors) by performing a simultaneous decomposition of two matrices, **X** and **Y,** with the constraint that these components explain as much as possible of the covariance between **X** and **Y**. PLS is based on the following relationships for matrix decomposition [15]:

$$\mathbf{X} = \sum_{l=1}^{L} \mathbf{t}_l \mathbf{p}_l^T + \mathbf{E} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{1}$$

$$\mathbf{Y} = \sum_{l=1}^{L} \mathbf{u}_l \mathbf{q}_l^T + \mathbf{F} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} \tag{2}$$

where l represents the number of latent variables retained in the PLS model, **t** contains information related to the samples of the input matrix **X**, **u** contains information related to the samples of the output matrix **Y**, **p** contains information related to the variables of the input matrix **X**, and **q** contains information related to the variables of the output matrix **Y** the vectors **t** and **u** are called score vectors, and the vectors **p** and **q** are called loading vectors.

Furthermore, **T**, **U**, **P** and **Q** represent the score and loading matrices describing the PLS decomposition, respectively; and **E** and **F** are the residuals of the input and the output matrices, respectively. The **X** components and **Y** components are chosen so that the relationship between the successive pairs of principal components is as strong as possible by manipulating the inner relation. **U** is determined as

$$\mathbf{U} = \mathbf{T}\mathbf{B} \tag{3}$$

where the coefficient matrix **B** is a diagonal matrix of weights optimized to maximize the covariance between the components in **U** and **T**. An adjusted version of the nonlinear iterative PLS (NIPLS) algorithm, described in [15], can be used to calculate the PLS regression model. Predictions from PLS models are obtained using the multivariate regression formula [16], [17]

$$\hat{\mathbf{Y}} = \mathbf{X} \cdot \mathbf{P}\mathbf{B}\mathbf{Q}^{\mathbf{T}}. \tag{4}$$

B. Support Vector Machines

SVMs represent a machine learning method for both classification and regression problems based on statistical learning theory [18], [19]. SVMs have been used in classification and identification in [20] and [21].

A SVM attempts to create a hyperplane between the two sets of data for classification. For example, consider Fig. 3, where two different classes are shown (class A with circles and class B with triangles). The SVM attempts to place a linear boundary (represented by a solid line) between the two classes. The nearest data points from the linear boundary are used to define a margin between the classes [called the support vectors (SVs)], as shown by the gray points in Fig. 3. When the two classes are divided linearly, data classification is accomplished by defining a hyperplane that divides the training data set so that all the training data points of the same class are on the same side of the hyperplane, while maximizing the distance between the hyperplane and the SVs.

Given a training set or training samples $S = \{\mathbf{x}_i, y_i\}$ with $i = 1, 2, \ldots, N$, where $\mathbf{x}_i$ denotes the components of the N-dimensional column vector that are the classifier inputs,

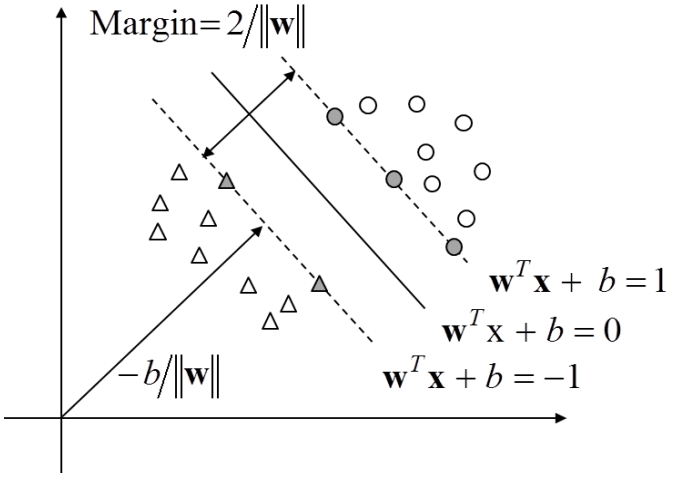

Fig. 3. Example of separation of two classes by an SVM.

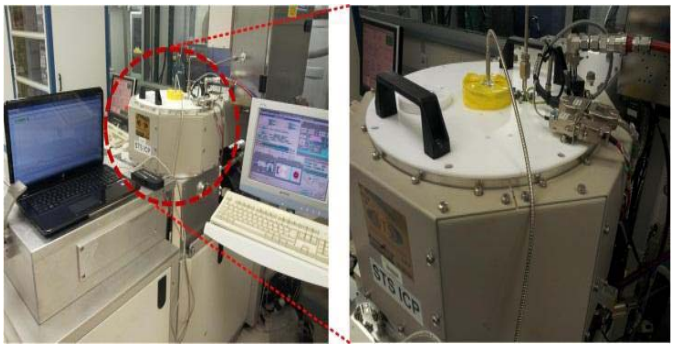

Fig. 4. OES setup with sensor mounted on top of ICP etching tool process chamber (closer image shown in the right).

TABLE I
BOSCH ETCHING PROCESS CONDITIONS

Parameter	Etching	Passivation
Time (sec)	11	8
Gas flow (sccm)	SF_6 130, O_2 13	C_4F_8 13
Pressure (mTorr)	17	17
Top Coil power (W)	800	800
Bottom Platen power (W)	17	0

$y_i$denotes the class labels that are the classifier outputs (and can be 1, −1 for class 1, class −1, respectively), and N is the size of the training set. The linear function [22] that separates both classes is

$$D(x) = \text{sgn}\,[\,f(x)\,] \tag{5}$$
$$f(x) = \mathbf{w}^T x + b \tag{6}$$

where $\mathbf{w}$ is a d-dimensional column vector and has the same dimension as the $x$column, b is a scalar, sign(•) is the sign function, the operator $(\bullet)^T$ is the transpose operator, and $D(x)$ returns the label of the class that is classified $\mathbf{x}$. The hyperplane that separates both classes is given by $f(\mathbf{x}) = 0$. Therefore, the vector $\mathbf{w}$ and scalar b determine a position of the separating hyperplane, $\mathbf{w}$ determines the orientation, and b determines

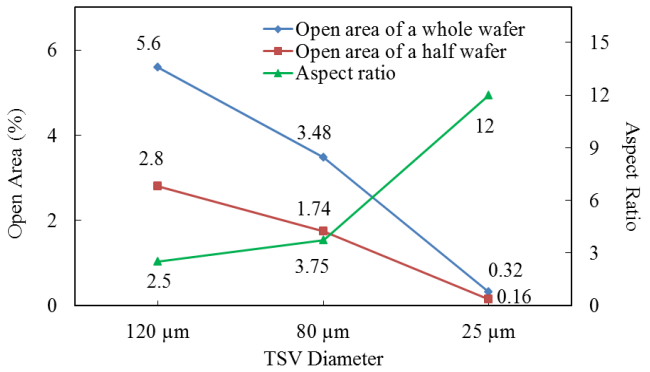

Fig. 5. Patterned open areas and aspect ratios of samples.

the hyperplane separation to the origin or reference. When the training set is linearly separable

$$y_i f(\mathbf{x}_i) \geq 1 \quad \forall i = 1,\ 2, \ldots,\ N. \tag{7}$$

and samples of the training set become correctly classified.

The hyperplane that has the maximum distance between itself and the closet samples is the maximum margin hyperplane, called the optimal separating hyperplane. The classifier with the optimal separating hyperplane is the classifier with the greatest generalization capability. To find the optimal separating hyperplane, it is convenient to use three parallel hyperplanes, as shown in Fig. 1, such that

$$\text{Optimal hyperplane}: f(x) = \mathbf{w}^T x + b = 0 \tag{8}$$
$$\text{Positive class's hyperplane}: f(x) = \mathbf{w}^T x + b = +1 \tag{9}$$
$$\text{Negative class's hyperplane}: f(x) = \mathbf{w}^T x + b = -1. \tag{10}$$

The hyperplanes of the positive and negative classes must consist of some training data, which are known as SVs that contain all necessary information to create the classifier hyperplane. The Euclidean distance between the two hyperplanes (8) and (9) is called a margin and can be expressed as $2/\|\mathbf{w}\|$, as well as can be combined into (12). Thus, the objective of the SVM is to maximize the distance between the hyperplanes of the two classes. The maximum margin algorithm can be described as follows:

$$\min\ \frac{1}{2}\|\mathbf{w}\|^2 \tag{11}$$
$$\text{s.t}\quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1,\ 2, \ldots, N. \tag{12}$$

The above function can be minimized using Lagrange multipliers [23]. Lagrange multipliers are used to solve optimization problems with linear equality and inequality constraints. The function of a Lagrange multiplier is to incorporate the constraints into the expression one wants to minimize or maximize. A Lagrange equation can be defined as follows:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i (y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1) \tag{13}$$

where $\alpha = [\alpha_1, \ldots, \alpha_N]$ and $\alpha_i \geq 0$. The task is to minimize (13) with $\mathbf{w}$ and b, and to maximize it with respect

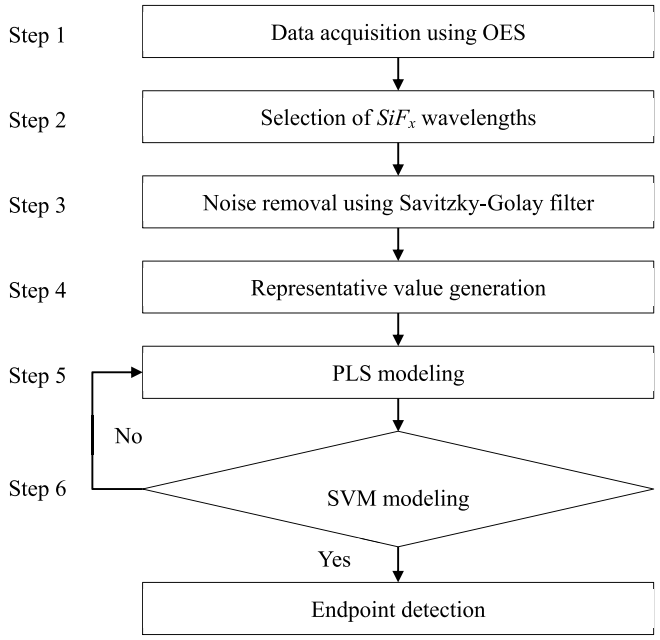

Fig. 6. Flowchart of the proposed PLS-SVM model.

to α_i. At the optimal point, we have the following saddle point equations:

$$\frac{\partial L(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i = 0 \tag{14}$$

and

$$\frac{\partial L(\mathbf{w}, b, \alpha)}{\partial b} = -\sum_{i=1}^{N} \alpha_i y_i = 0 \tag{15}$$

The other (inactive) constraints can be discarded without changing the optimal point. It is consistent therefore to set the Lagrange multipliers for the inactive constraints to zero. This condition can be summarized as

$$\alpha_i \left[y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 \right] = 0 \qquad \forall i = 1, 2, \ldots, N. \tag{16}$$

Substituting (14) and (15) into (13) gives the dual quadratic optimization problem

$$L_{\text{dual}}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \left(\mathbf{x}_i^T \mathbf{x}_j \right) \tag{17}$$

subjected to

$$\sum_{i=i}^{N} \alpha_i y_i = 0, \quad \alpha_i \geq 0 \quad \forall i = 1, 2, \ldots, N. \tag{18}$$

Thus, by solving the dual optimization problem, one obtains the coefficients α_i, which are required to express $\mathbf{w}$ to solve (11). Substituting (14) into (6) gives the dual form of the decision function of the following classifier:

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i (\mathbf{x}_i^T \mathbf{x}) + b. \tag{19}$$

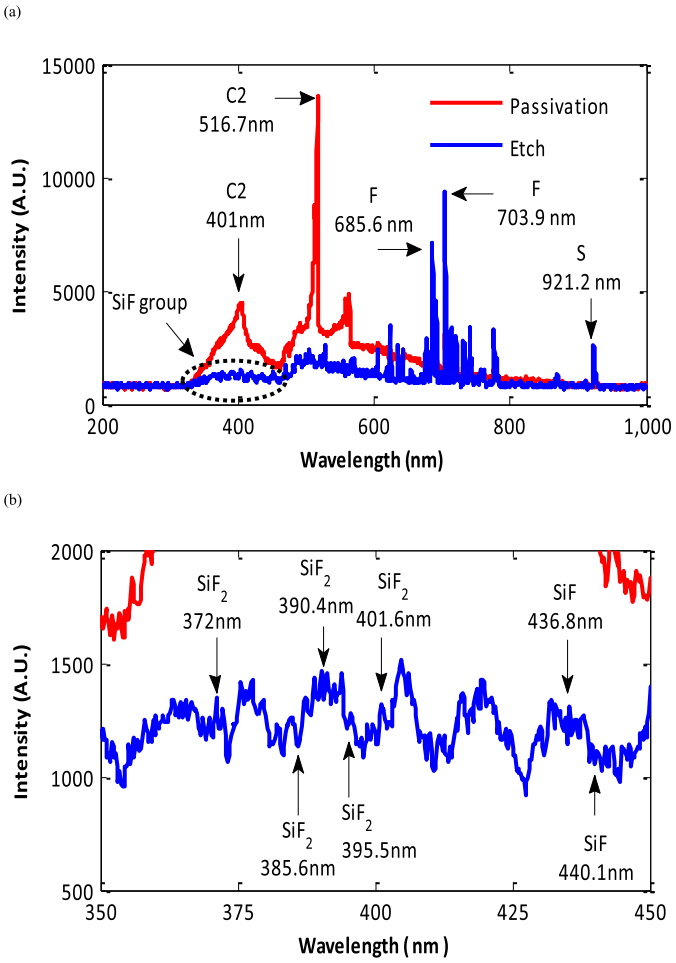

Fig. 7. (a) Passivation and etch step spectra for silicon etch process; the plasma emission spectra of both steps are dramatically different. (b) Magnified image (SiF group) of (a); selected SiF_x species in etch step during silicon etching.

In real-world applications, classification problems are usually linearly inseparable. When classes are not linearly separable, the training samples are transformed into a high-dimensional space where linear class separation is possible. The data transformation is accomplished via a function $\phi(\bullet)$, which takes the data from the input space to a feature space where classes are linearly separable. To get a classifier with good performance, we often solve the dual optimization problems for nonlinear soft margin algorithms and adjust the parameters using a cross-validation strategy. By introducing slack variables to relax the hard margin constraints, an optimization problem for a soft margin was proposed in [18] and [19]. Then, the problem given by (11) and (12) becomes

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \zeta_i$$

$$\text{subjected to } \zeta_i \geq 0, \quad y_i (\mathbf{w}^T \varphi(\mathbf{x}_i) + b) \geq 1 - \zeta_i \qquad \forall i = 1, 2, \ldots, N \tag{20}$$

where the constant $C \in [0, \infty)$ is known as a penalty parameter and creates a tradeoff between the number of misclassified samples in the training set and a separation of the rest samples

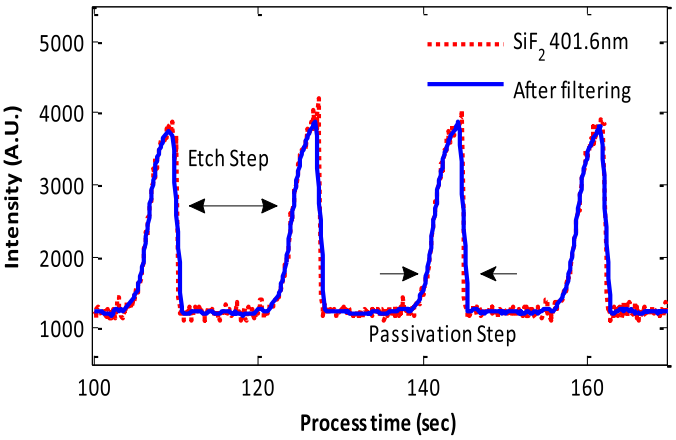

Fig. 8. Filtered intensity of SiF_2 401.6 nm using the SG filter.

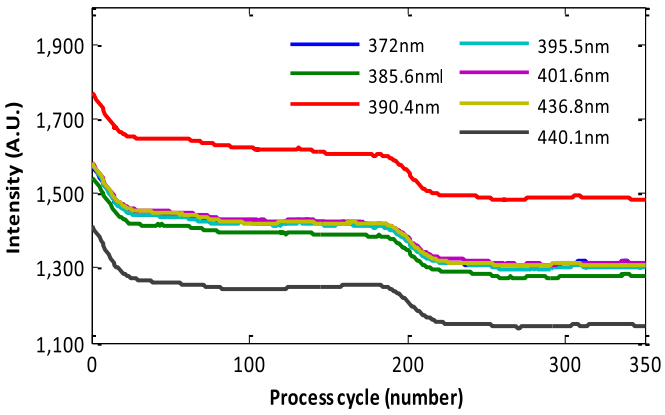

Fig. 9. Mean integrated values of SiF_x intensity during etch cycles.

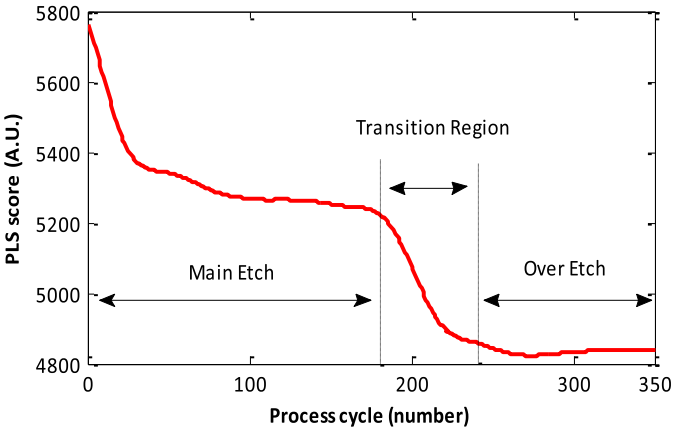

Fig. 10. Generated endpoint signal in 120-μm diameter TSVs from the first run of back-to-back etching using a PLS.

with maximum margin. The function ϕ is a feature map associated with kernel function $K(\mathbf{x}_i,\mathbf{x}_j) = \phi(\mathbf{x}_i)\phi(\mathbf{x}_j)$ and slack variable $\zeta_\mathbf{i}$ is introduced to allow some of the training patterns to be misclassified or to lie inside the classifier margin. Given the kernel function, the dual optimization problem of (20) is simplified as

$$\max \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,\,j=1}^{N} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

$$\mathbf{w} = \sum_{i=1}^{N} y_i \alpha_i \phi(\mathbf{x}_i) \qquad \sum_{i=1}^{N} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le C,$$

$$\forall i = 1,\ 2, \ldots, N. \qquad (21)$$

Equation (21) shows that the optimal hyperplane, in feature space, can be written as the linear combination of training samples with $\alpha_i \neq 0$. These informative samples, known as SVs construct the decision function of the classifier based on the kernel function [24]

$$f(\mathbf{x}) = \mathbf{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b\right). \qquad (22)$$

III. Experimental Setup

TSV etching was performed in an STS inductively coupled plasma-reactive (ICP) ion etch system using the Bosch process with an OES system installed on top of process chamber, as shown in Fig. 4. The OES system used in this experiment is a Korea Spectral Products SM440 optical sensor. The optical fiber is fixed on the topside quartz window view port of the chamber during the etching process to collect plasma glow discharge directly through a 300-μm diameter fiber with the detection range 200–050 nm.

To fabricate the TSV sample wafer, a silicon dioxide layer was first deposited on one surface of an approximately 300 μm thick and 4-in silicon wafer, followed by photolithography to form a photoresist etch mask for silicon etching on the other side of the wafer. The Bosch process conditions were kept constant at the base values, listed in Table I, with varying process cycles depending on TSVs feature size. The equipment included two independent 13.56 MHz RF power sources: one for the coil around the etching chamber to create the plasma, and the other connected to the wafer electrode to control the RF bias potential of the wafer with respect to the plasma. Back side helium pressurization was used to provide sufficient heat transfer between the wafer and the electrode to maintain a constant wafer temperature.

Three different sizes of TSV diameters (120, 80, and 25 μm) were used to verify the accuracy of the proposed hybrid PLS-SVM model to detect endpoint with varying open areas and aspect ratios, as shown in Fig. 5. To avoid wafer-to-wafer thickness variation in the samples for model establishment and verification, TSV etching was performed on one half of the wafer first while keeping the rest of the wafer was covered. Once the etching of the first half was completed, the second half was etched guided by the EPD prediction to verify the accuracy of the suggested hybrid PLS-SVM model.

IV. PLS-SVM EPD Model

The proposed hybrid PLS-SVM model consists of six steps, as shown in Fig. 6.

Step 1 (Data Acquisition Using OES): The OES sensor detects plasma emission through a small quartz window located on the top of the process chamber wall via a fiber-optic cable. The OES sensor points to the center of the

TABLE II
SELECTED EQUIPMENT VARIABLES FOR FAULT DETECTION AND CLASSIFICATION RESULTS OF SVM SIMULATION WITH DIFFERENT PENALTY PARAMETERS AND KERNEL PARAMETER $d = 2$

TSV Diam. (μm)	Number of Training Data	Number of Misclassified Data depending on C					
		C =0.1	C =1	C =5	C =10	C =20	C =50
120	100	10	0	0	0	0	0
80	100	16	1	1	15	0	0
25	100	13	13	0	0	0	0

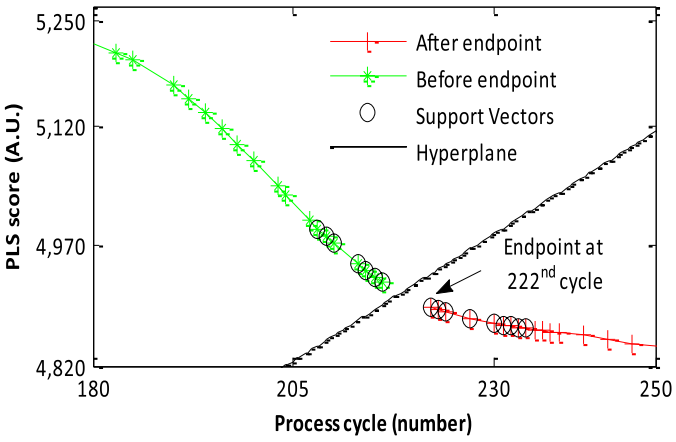

Fig. 11. Endpoint prediction by applying PLS-SVM. Endpoint was detection at 222 s cycle.

wafer. To acquire the OES signal during processing, the signal integration time is very important. The detector in the OES sensor captures light for a defined length of time (called the integration time). A longer the integration time results in a higher the intensity of the captured plasma signal. However, in TSV fabrication, if the integration time is set longer than the duration of the etch step, the OES signal will unintentionally capture information from the passivation step, thereby making the endpoint more difficult to determine. The integration time in this paper was fixed at 100 ms. To verify the endpoint, the first half of wafer was over etched to completely remove the silicon. The first half of a 120-μm diameter TSV was etched for 350 cycles, and then the acquired OES data were analyzed after TSV etching to predict the endpoint for the second half of the sample.

Step 2 (Selection of SiF_x Wavelengths): In the Bosch process, the gas used for silicon etching is SF_6, while C_4F_8 is the main gas used in the passivation step. Fig. 7(a) shows the OES data from both steps. In the plasma, SF_6 dissociates into atomic fluorine and heavy SF_5^+ ions [25], as shown in (23). The resulting plasma leads to silicon etching as shown in chemical reaction (24)

$$SF_6 + e^- \rightarrow SF_5^+ + F + 2e^- \quad (23)$$

$$Si\ (solid) + nF(gas) \rightarrow SiF_x(gas). \quad (24)$$

Most of the EPD methods using OES focus on identifying a single wavelength that corresponds to a chemical species that shows a pronounced transition at the end point [26]–[28]. However, this single wavelength method is limited in detecting an endpoint when the open area becomes small or the signal is too weak in its magnitude.

The most significant byproducts of silicon etching using an SF_6 plasma are SiF_x radicals, as shown in the Fig. 7(b), where seven SiF_x peaks have been obtained from [29] and [30]. In this paper, these major SiF_x peaks were employed to extract an endpoint signal.

Steps 3 and 4 (Noise Removal Using Savitzky–Golay (SG) Filter and Representative Valve Generation): To extract a meaningful endpoint signal during the Bosch process, the spectra of SiF_x signals were filtered using a SG filter [31] to eliminate noise from the SiF_x related peaks, as shown in Fig. 8. In this method, a smoothing window width is selected and the point located at the center of the window is replaced with the estimated value obtained by the polynomial curve fitting. The general equation of the simplified least squares convolution for the acquired OES data are as follows:

$$Y_j^* = \sum_{i=-m_L}^{m_R} C_i Y_{j+i} \quad (25)$$

where Y_j^* is the replacing value, Y_j is the respective resultant OES data, C_i is the coefficient for the ith observation of OES data, and the index j is the running index of the original ordinate data table. The parameter m_L is number of data points to the left of j, and m_R is the number of data point to the right of j.

Note that for a coefficient $C_i = 1/(m_L + m_R + 1)$, (25) becomes a filter where the center point is the average of the data in the window surrounding the point. The novelty introduced by Savitzky and Golay lies in the approximation of the underlying function by a polynomial of higher order instead of the coefficient C_i. Therefore, for each data point value Y_j, $m_L \leq j \leq m_R$, we fit a polynomial to all $m_L + m_R + 1$ points and set Y_j^* to the value of that polynomial at position j When the SG filter is applied to SiF_x wavelengths, there are two parameters that must be determined. The first parameter is the size of the smoothing window ($m_L = m_R$). Usually, a larger value of a smoothing window produces a smoother result at the expense of flattening sharp peaks. The second parameter is a coefficient C_i. In this paper, we set SG filter with a smoothing window size 31 and a polynomial degree 4. After the seven selected

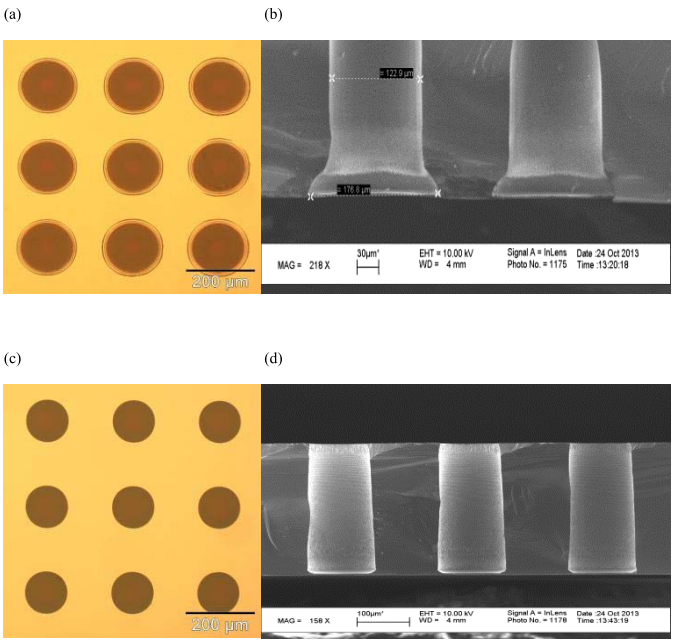

Fig. 12. (a) Bottom view and (b) cross-sectional view of 120-μm diameter TSVs after etching the first half of a wafer for 350 cycles. (c) Bottom view and (d) cross-sectional view of 120-μm diameter TSVs after etching the second half of a wafer for 222 cycles as predicted endpoint.

SiF_x wavelengths were smoothed, the intensities of SiF_x were integrated over the etch cycles to generate the mean representative intensity values, as shown in Fig. 9.

The variation of SiF_x signals over the entire process cycles is straightforward In particular the accuracy of EPD can be improved using only SiF_x signals of the etch step, excluding passivation step.

Step 5 (PLS Modeling for Generating Endpoint Signal): The aforementioned PLS model was used to reduce the dimensionality the of SiF_x peaks without information loss as well as improve the transition region between the main etch and over etch to predict the endpoint through the SVM model. The **X** in (1) is comprised of the representative value from seven SiF_x peaks. The **Y** in (2) is constructed by the one of SiF_x peaks. Through the PLS decomposition, the loading vectors **P**, **Q**, and coefficient matrix B are drawn from **X** and **Y**, respectively. Therefore, the endpoint signal can be generated using (4). Fig. 10 shows the endpoint signal after etching the first half of the wafer containing 120 μm diameter TSVs. As shown, the first half was etched for 350 cycles, and the result of PLS clearly draw a distinction between the main and over etch at the transition region.

Step 6 (SVM Modeling for Detecting Endpoint): The last step of the proposed EPD method is to apply the SVM as the classification method discussed in Section IIB. To train an SVM classifier, the data points in the PLS score transition region resulting from the first half wafer etch were applied as inputs to the SVM. Using the data from the transition region instead of the entire data, the computation time and complexity of the SVM can be reduced as a result of the small training sample set size.

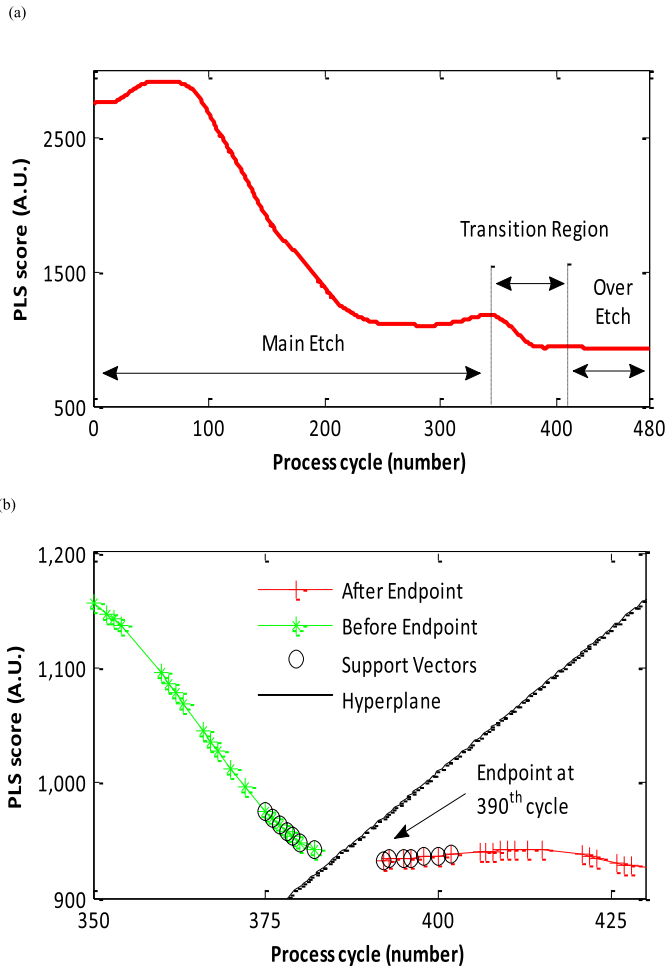

Fig. 13. (a) Generated endpoint signal in 80-μm diameter TSVs from the first run of back-to-back etching using a PLS. (b) Endpoint detection at three hundred and ninetieth cycle.

At this stage, the SVM still has a few tunable parameters that need to be determined. SVM training includes the selection of the proper kernel function parameters and the penalty parameter C shown in (20). The selection of the kernel function parameters is very important because it implicitly defines the structure of the highdimensional feature space where the maximal margin hyperplane is found. For this EPD application, the following polynomial kernel [32] is chosen because it is a popular method for nonlinear modeling and it is simple to tune:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^d \tag{26}$$

where d is a parameter that controls the width of the polynomial function. The choice of the kernel parameters is a practical problem for training an SVM because the classification result with a specific parameter is not known until after simulation. Through several simulations, we chose $d = 2$ as the optimal kernel parameter empirically.

In addition, the penalty parameter C controls the complexity of the learning machine to a certain extent and influences the training speed [33]. Since successful theoretical methods are not available for parameter selection, the parameter C has been trained with various values, and the results of training are

TABLE III
RESULTS OF THE TSVs BACK-TO-BACK ETCHING WITH 120, 80, AND 25 μm

TSV Diam. (μm)	Wafer Half	Process Cycle	Average Diameter for 100 vias (μm)		Variations between The Maximum and The Minimum Diameters (μm)	
			Top	Bottom	Top	Bottom
120	1st	350	128.9	175.7	6.90	8.96
	2nd	222	123.7	121.1	3.75	4.37
80	1st	420	84.72	122.85	5.96	12.21
	2nd	390	83.11	79.54	3.98	6.54
25	1st	650	29.61	48.39	1.84	5.43
	2nd	624	28.87	28.20	1.5	3.81

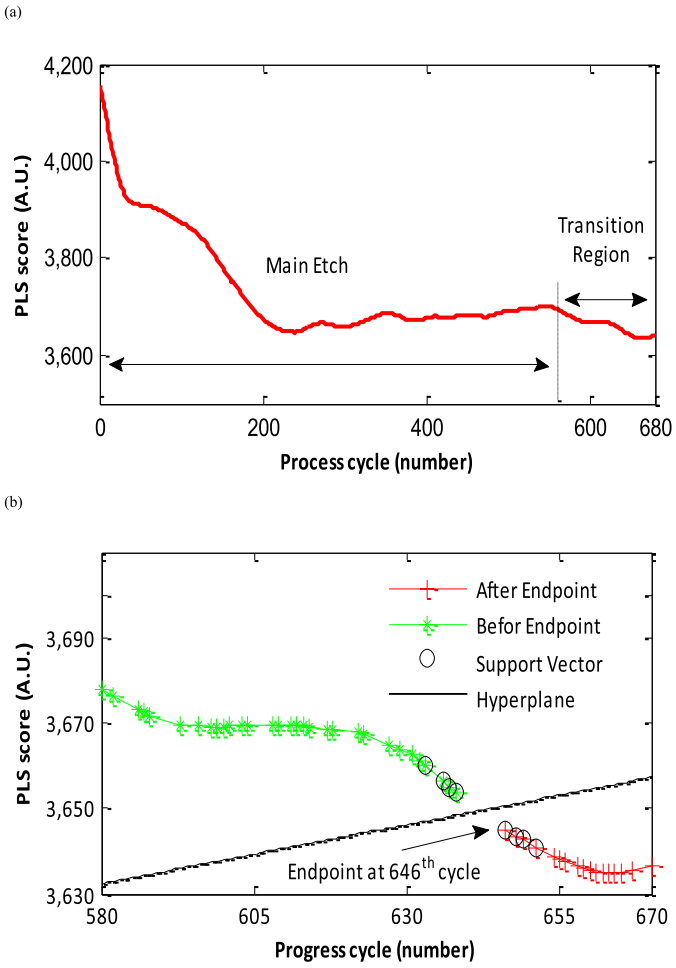

Fig. 14. (a) Generated endpoint signal in 25-μm diameter TSVs from the first run of back-to-back etching using a PLS. (b) Detected endpoint at six hundred and forty sixth cycle.

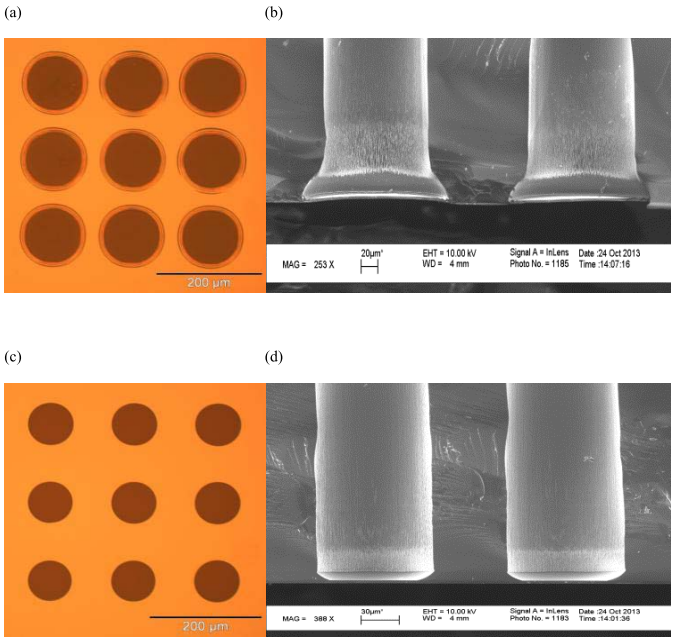

Fig. 15. (a) Bottom view and (b) cross-sectional view of 80-μm diameter TSVs after etching the first half of a wafer for 480 cycles. (c) Bottom view and (d) cross-sectional view of 80-μm diameter TSVs after etching the second half of a wafer for 390 cycles as predicted endpoint.

shown in Table II. As shown in Table II, the optimal penalty parameter C depends on the diameter of the TSV. Therefore, the SVM model used $C = 1$ for the 120 μm diameter, $C = 20$ for the 80 μm diameter, and $C = 5$ for 25 μm diameter TSVs.

As shown in Fig. 11, using the hybrid PLS-SVM model, an accurate endpoint (at 222 s cycle) computed from the transition region data of the first half wafer etch. The computed endpoint was applied during the etching of the second half of the wafer. Note that no matter which wafer half is etched, the other half was always covered. Accurate EPD was obtained for the etched 120 μm diameter TSVs using this method. Fig. 12 shows the top-view and cross-sectional images of the etched TSVs in both halves of the wafer. The images of etched TSVs in the second half of the wafer clearly illustrate accurate end point prediction.

V. VERIFICATION OF THE PROPOSED EPD MODEL

To verify the robustness of the hybrid PLS-SVM EPD algorithm, we conducted back-to-back etching of wafers containing 80 and 25 μm diameter TSVs. The endpoint obtained from the first half of the wafer was applied to accurately predict the endpoint of the second half of the

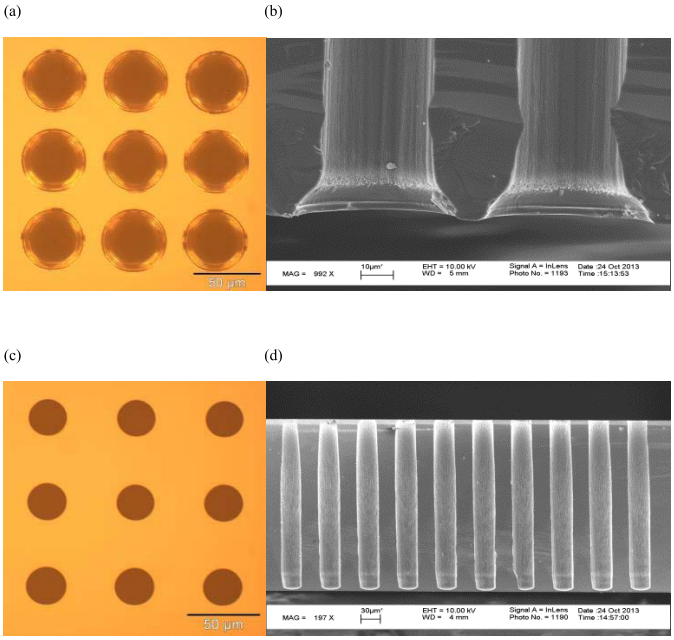

Fig. 16. (a) Bottom view and (b) cross-sectional view of 25-μm diameter TSVs after etching the first half of a wafer for 680 cycles. (c) Bottom view and (d) cross-sectional view of 25-μm diameter TSVs after etching the second half of a wafer for 646 cycles as predicted endpoint.

wafer as described previously. The first half of the wafer with 80 μm was etched for 480 cycles. PLS was applied to the representative value of SiF_x group, and the results of PLS are shown in Fig. 13(a), where the transition region between 350 and 420 cycles has been predicted. Based on PLS, the endpoint was predicted at the three hundred and ninetieth cycle using the SVM, as shown in Fig. 13(b). Using the predicted endpoint of the first half of the wafer, the second half was etched for the 390 cycles To verify these results, SEM analysis was conducted, and the results are presented in Fig. 15. The results confirm the accuracy of EDP. The back-to-back etching of wafer containing 25 μm diameter TSVs was also conducted in the same way as the wafer with 80 μm diameter TSVs, and the results confirming accuracy are shown in Figs. 14 and 16. Table III shows the average via diameters after the etching of each half of the wafer with 120, 80 and 25 μm diameter, respectively.

VI. Conclusion

EPD is an important method to control critical dimensions in TSV etching. However, EPD is difficult for features like TSVs that have low open area and high aspect ratio. To address these challenges, a hybrid PLS-SVM model using OES has been proposed for EPD, and the suggested method has been successfully demonstrated for Bosch process etching of 120, 80 and 25 μm diameter TSVs. A minor difference (2.6 μm in 120 μm TSVs, 1.61 μm in 80 μm TSVs, and 0.74 μm in 25 μm TSVs) between the top and bottom diameters of the TSVs was observed after etching, but overall the PLS-SVM based endpoint predictive model demonstrates good accuracy in TSV fabrication.

The proposed EPD methodology is not fixed for a particular TSV geometry or diameter, and it is expected that it can be adapted for smaller diameter TSVs. The EPD method with OES is also suitable for the recently suggested vertical NAND flash memory process, which employs multiple pairs of oxide/nitride layer deposition [34]. We are currently investigating plasma processing monitoring of the multiple pairs of oxide/nitride layer deposition process and monitoring it's etch process using OES.

References

[1] J. U. Knickerbocker *et al.*, "2.5D and 3D technology challenges and test vehicle demonstrations," in *Proc. 62nd IEEE ECTC*, San Diego, CA, USA, Jun. 2012, pp. 1068–1076.
[2] V. S. Rao *et al.*, "TSV interposer fabrication for 3D IC packaging," in *Proc. 11th Electron. Packag. Technol. Conf.*, Dec. 2009, pp. 431–437.
[3] P. A. Thadesar and M. S. Bakir, "Novel photo-defined polymer-enhanced through-silicon vias for silicon interposers," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 3, no. 7, pp. 1330–1337, Jul. 2013.
[4] B. Wu, A. Kumar, and S. Pamarthy, "High aspect ratio silicon etch: A review," *J. Appl. Phys.*, vol. 108, no. 5, pp. 051101-1–051101-20, Sep. 2010.
[5] C.-H. Kim and Y.-K. Kim, "Prevention method of a notching caused by surface charging in silicon reactive ion etching," *J. Micromech. Microeng.*, vol. 15, no. 2, pp. 358–361, Feb. 2005.
[6] R. M. de Castro *et al.*, "End-point detection of polymer etching using Langmuir probes," *IEEE Trans. Plasma Sci.*, vol. 28, no. 3, pp. 1043–1049, Jun. 2000.
[7] W. Wang, Z. Lan, W. Wu, and Y. Gong, "Optical interferometry endpoint detection for plasma etching," in *Proc. 8th Int. Conf. Electron. Meas. Instrum.*, Xi'an, China, Aug. 2007, pp. 4-252–4-255.
[8] H. Handa, S. Yamauchi, H. Maruyama, S. Ishimoto, M. Kosugi, and Y. Miyahara, "Process monitoring of chrome dry-etching with RF sensors for reticle production beyond 90-nm node," *Proc. SPIE*, vol. 5256, pp. 85–92, Sep. 2003.
[9] R. Westerman, D. Johnson, S. Lai, and M. Teixeira "Endpoint detection method for time division multiplex etch processes," *Proc. SPIE, Micromach. Microfabricat. Process Technol. XI*, vol. 6109, pp. 61090I1–61090I11, Jan. 2006.
[10] S. Wold, M. Sjostroma, and L. Erikssonb, "PLS-regression: A basic tool of chemometrics," *Chemometrics Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, 2001.
[11] A. McIntosh, W. Chau, and A. Protzner, "Spatiotemporal analysis of event-related fMRI data using partial least squares," *J. NeuroImage*, vol. 23, no. 2, pp. 764–775, Oct. 2004.
[12] W. Lindberg, J. Persson, and S. Wold, "Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and lignin sulfonate," *J. Anal. Chem.*, vol. 55, no. 4, pp. 643–648, Apr. 1983.
[13] H. Wold, "Estimation of Principal Components and Related Models by Iterative Least Squares," Multivariate Analysis, P.R. Krishnaiaah, ed. New York: Academic Press, 1966, pp. 391–420.
[14] R. Shadmehr, D. Angell, P. B. Chou, G. S. Oehrlein, and R. Jaffe, "Principal component analysis of optical emission spectroscopy and mass spectrometry: Application to reactive ion etch process parameter estimation using neural networks," *J. Electrochem. Soc.*, vol. 139, no. 3, pp. 907–915, 1992.
[15] P. Geladi and B. R. Kowalski, "Partial least-squares regression: A tutorial," *Anal. Chimica Acta*, vol. 185, no. 1, pp. 1–17, 1986.
[16] H. Abdi. (2003). Partial least squares regression. *SAGE* [Online]. Available: http://www.utdallas.edu/~herve/Abdi-PLS-pretty.pdf
[17] S. J. Qin, "Partial least squares regression for recursive system identification," in *Proc. 32nd Conf. Decision Control*, vol. 3. San Antonio, TX, USA, Dec. 1993, pp. 2617–2622.
[18] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.
[19] V. N. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.
[20] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, May 2010.
[21] N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 11, pp. 3592–3607, Nov. 2011.

[22] E. Vazquez-Sanchez, J. Gomez-Gil, J. C. Gamazo-Real, and J. F. Diez-Higuera, "A new method for sensorless estimation of the speed and position in brushed DC motors using support vector machines," *IEEE Trans. Ind. Electron.*, vol. 59, no. 3, pp. 1397–1408, Mar. 2012.
[23] B.-S. Yang, W.-W. Hwang, M.-H. Ko, and S.-J. Lee, "Cavitation detection of butterfly valve using support vector machines," *J. Sound Vibration*, vol. 287, nos. 1–2, pp. 25–43, Oct. 2005.
[24] M. A. Oskoei and H. Hu, "Support vector machine-based classification scheme for myoelectric control applied to upper limb," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1956–1965, Aug. 2008.
[25] J. Bhardwaj, H. Ashraf, and A. McQuarrie, "Dry silicon etching for MEMS," in *Proc. 191st Meeting Electrochem. Soc., Symp. III Microstruct. Microfabricat. Syst.*, vol. 97-5. Montreal, QC, Canada, May 1997, pp. 118–130.
[26] S. J. Pearton, F. Ren, and C. R. Abernathy, "Optical emission end point detection for via hole etching in InP and GaAs power device structures," *Mater. Sci. Eng. B*, vol. 21, no. 1, pp. 36–40, Feb. 1994.
[27] P. Bioisi, L. Drachnik, S. Dllinger, and D. Morvay, "An advanced end-point solution for <1% open area applications; contact and via," in *Proc. IEEE/SEMI ASMC*, Cambridge, MA, USA, Nov. 1996, pp. 391–396.
[28] L. Dreeskornfeld *et al.*, "Reactive ion etching with end point detection of microstructured Mo/Si multilayers by optical emission spectroscopy," *Microelectron. Eng.*, vol. 54, nos. 3–4, pp. 303–314, Dec. 2000.
[29] J. S. Horwitz, C. S. Dulcey, and M. C. Lin, "REMPI/MS detection of SiF_2 radicals by $(3+1)$ and $(1+3)$ photoionization," *Chem. Phys. Lett.*, vol. 150, nos. 1–2, pp. 165–170, Sep. 1988.
[30] C. S. Bulcey and J. W. Hudgens, "Detection of SiF radicals with multiphoton ionization spectroscopy," *Chem. Phys. Lett.*, vol. 118, no. 4, pp. 165–170, Aug. 1985.
[31] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964.
[32] S. Gunn, "Support vector machines for classification and regression," Image Speech and Intelligent Systems Group, Dept. Electron. Comput. Sci., Univ. Southhampton, Southampton, U.K., Tech. Rep. MP-TR-98-05, 1998.
[33] S. Hua and Z. Sun, "Support vector machine approach for protein subcellular localization prediction," *Bioinformatics*, vol. 17, no. 8, pp. 721–728, Oct. 2001.
[34] J. Jang *et al.*, "Vertical cell array using TCAT (terabit cell array transistor) technology for ultra high density NAND flash memory," in *Proc. Symp. VLSI Technol.*, Jun. 2009, pp. 192–193.

Ja Myung Gu, photograph and biography not available at the time of publication.

Paragkumar A. Thadesar, photograph and biography not available at the time of publication.

Ashish Dembla, photograph and biography not available at the time of publication.

Muhannad S. Bakir, photograph and biography not available at the time of publication.

Gary S. May, photograph and biography not available at the time of publication.

Sang Jeen Hong, photograph and biography not available at the time of publication.