A CVD Epitaxial Deposition in a Vertical Barrel Reactor: Process Modeling Using Cluster-Based Fuzzy Logic Models

J. C. Chiou and J. Y. Yang

***Abstract*—A chemical vapor deposition (CVD) epitaxial deposition process modeling using fuzzy logic models (FLM's) has been proposed. The process modeling algorithm consists of a cluster estimation method and backpropagation algorithm to construct a number of modeling structures from the training data. A decision rule based on the multiple correlation factor is used to obtain the optimum structure of fuzzy model using the testing data. Upon the optimum structure has been reached, the gradient-descent method is used to refer the parameters of the final fuzzy model using both training and testing data. The algorithm has been applied to a nonlinear function and a vertical chemical vapor deposition process. The results demonstrate the efficiency and effectiveness of the proposed fuzzy logic model in comparison with existing fuzzy logic models and artificial neural network models.**

***Index Terms*— Chemical vapor deposition (CVD) modeling, clustering estimation method, fuzzy logic.**

I. Introduction

PROCESS modeling and control are essential to high yield, high quality, and low cost manufacturing in today's competitive semiconductor market. With the device designs shrinking to a minimum feature size of 0.15 μm by the year 2001, the process-control requirements will be more stringent and the attainment of higher yield will require the process engineer to control variability at each of the many processing steps in the microelectronics manufacturing processes. To achieve this, all the input variables controlling the desired output in a given process need to be understood and optimized for tighter control. In addition, the process controller must be quick and responsive to the variations of the input parameters. Thus, designing an effective process control technique satisfying these requirements is a very challenging job.

The first step in meeting these challenges is the development of an accurate model to describe the process. For some processes, it is possible to determine the input(s)–output(s) relationship through physical or analytical models. However, for many semiconductor manufacturing processes, physical models usually use many simplifications and assumptions which limit the model's accuracy and effectiveness. An attractive alternative is to use neural network or fuzzy logic techniques to build empirical models. These models then form the basis for process optimization and control.

Since Sugeno first introduced the FLM concept in 1985, which is known as the Sugeno-type FLM, a number of researchers have successfully applied his concept in modeling and control [1]–[3]. Various approaches have been developed to identify a FLM for a given process. However, because of the complexity of obtaining the structure of membership functions and its corresponding parameters in FLM. These approaches encountered drawbacks such as lack of efficiency during the modeling process, and fail to accurately predict their behaviors after the process has been identified. Thus, a more efficient and effective algorithm is needed in modeling the semiconductor manufacturing process.

To identify an FLM for a process, there are two major tasks: structure and parameter identifications, where each of these consists of premise and consequence parts. In principle, we cannot separate the structure identification from the parameter identification since they have a mutual relationship. For the identification of Sugeno-type FLM's, Yager and Filev [4] developed an alternative approach by transforming the problem of structure identification to estimation of distribution of input space in which the concept of sample probability distributions was introduced. This approach allows us to simplify the problem of structure identification by replacing identification of membership function of input variables with identification of the centers of cluster-like regions. However, the resulting identification algorithm still depends heavily on a nonlinear optimization procedure. Wang and Langari [5], [6] suggested a different approach in building Sugeno-type FLM's. Their approach used both the fuzzy discretization technique to determine the premise of the model and an orthogonal estimator to identify the consequence of the model. Hwang and Woo [7] utilized a fuzzy c-means clustering and GA hybrid scheme to identify the structure and the parameters of an FLM.

Most of the algorithms mentioned above have been developed for systems that can generate significant amounts of training data. An FLM trained by a large number of randomly generated data can usually model the system's behavior accurately. For the process modeling and control, however, the data generation takes time and money, thus engineers can only run designed experiments for only a few representative data. With a limited number of data points, the accuracy of an FLM is strongly dependent on its structure. As

Manuscript received March 11, 1997; revised June 26, 1998.

The authors are with the Department of Electrical and Control Engineering, College of Electrical Engineering and Computer Science, National Chiao-Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: chiou@cc.nctu.edu.tw).

Publisher Item Identifier S 0894-6507(98)08359-6.

a result, the structure identification scheme is a critical concern that demands a detailed study.

There are two important factors in developing an efficient structure identification scheme. One is the computation time, and the other is the criterion to choose an optimum structure. For a good-structure-criterion, most of the studies used the multiple correlation coefficient, R^2, to establish the correlation between the training data, the sample mean data and the output of the employed identification scheme. For a given structure, the parameter identification process will determine the right coefficients to model the given process. If $R^2 = 1$, it represents that the model fits every data point perfectly. Thus, for a case where a lot of data is available for training, a good model that fits the data well is reasonably expected. By using this model, we can predict the system behavior at nontraining parameter settings. However, for the case with a limited number of data, this expectation is questionable. Thus, a more rigorous criterion should be defined in order to obtain a good model.

In order to obtain a sufficient model from a limited number of data, an FLM for the process modeling and control is proposed in this paper, the process modeling consists of a cluster estimation method and gradient-decent method to construct a number of modeling structures from the training data. A decision rule based on the multiple correlation coefficient is used to obtain the optimum structure of the fuzzy model. Upon the optimum structure has been reached, the backpropagation algorithm is used to refine parameters of the final fuzzy model using both training and testing data. This algorithm has been applied to a nonlinear function and to a vertical chemical vapor deposition process. By comparing with other process modeling methods (both ANN and fuzzy), the proposed FLM has the advantages of simplicity, effectiveness and high predictive accuracy.

II. Fuzzy Logic Model

In generally, there are two methods in developing an FLM. One is the relational matrix methodology used by Pedrycz [9], [10]. The other is in the form of fuzzy relations suggested by Takagi and Sugeno [1], which uses internal functions instead of fuzzy sets to generate the output of the model. The second approach results in higher accuracy in modeling and requires less calculation during computation [11]. An FLM consists of four major elements: membership functions, internal functions, rules and outputs. A brief description of these four elements if given as follows.

A. Membership Functions

The membership function calculates the membership grading for any given value of the input variable. The ranges of all the input variables are transformed into $[-1, 1]$ domain in order to ease the proceeding development of the proposed algorithm. Note that the membership grading are ranged from 0 to 1; where "0" means no contribution of that variable and "1" means a full contribution of that variable. For the ith rule, the membership function, $A_l^i(x_l)$, calculates the membership grading for input variables, $x_l (l = 1, 2, \cdots, k)$, where k is the number of the input variables. Generally, we choose $A_l^i(x_l)$ to be bell-shaped or triangle-shaped with a maximum value equal to 1 and minimum value equal to 0. In the present development, the Gaussian-type membership function is used and given by

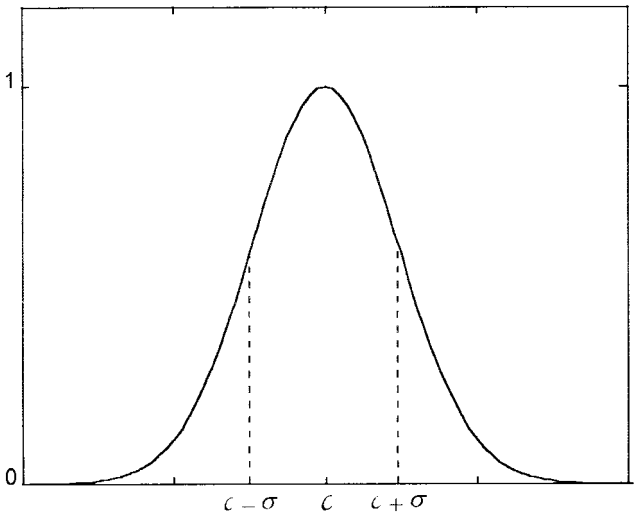

Fig. 1. Gaussian type membership function.

$$A_l^i(x_l) = \exp\left[-\left(\frac{x_l - c_l^i}{\sigma_l^i}\right)^2\right] \tag{1}$$

where c_l^i and σ_l^i are called location and shape parameters which need to be determined. As shown in Fig. 1, each of there parameters has its physical meaning, i.e., the location parameter determines the center of the corresponding membership function and the shape parameter is the width of that membership function.

B. Rules and Internal Functions

In Sugeno type's FLM, the ith rule is defined as follows:

if x_1 is F_1^i, and x_2 is F_2^i and $\cdots$ and
x_k is F_k^i, then the rule output is
$$f^i(x_1, x_2, \cdots, x_k) = p_0^i + p_1^i x_1 + p_2^i x_2 + \cdots + p_k^i x_k$$

where F_j^i is fuzzy set. The output $f^i(x_1, x_2, \cdots, x_k)$ of the ith rule is the internal function with parameters $p_0^i, p_1^i, p_2^i, \cdots, p_k^i$ needed to be determined. With respect to each rule, the membership function and the internal function are used to determine the rule's output. The use of several internal functions accounts for the "fuzziness" of the FLM. For a crisp logic analysis, like a regression model, a single function such as nth order polynomial is used to represent the system's behavior. Instead of using a single function, the FLM uses several functions to cover the input range. As a result, the representation of the system is the integration of those functions rather than a single crisp correlation. Such integration is truly fuzzy, and the fuzzy integration can actually correlate nonlinear behaviors accurately [11].

C. Output

The output of the FLM is the weighted average of the rule outputs. In the present development, the *and* operator is replaced by the product operator. The weight of each rule's output is assumed to be the product of the grading, $A_1^i(x_1)$, $A_2^i(x_2)$, $\cdots$, $A_k^i(x_k)$, which can be substituted by using a Gaussian type membership function:

$$\begin{aligned} w^i &= A_1^i(x_1) \times A_2^i(x_2) \times \cdots \times A_k^i(x_k) \\ &= \exp\left[-\tfrac{1}{2}\sum_{l=1}^{k}\left(\frac{x_l - c_l^i}{\sigma_l^i}\right)^2\right] \end{aligned} \tag{2}$$

where i is the index of fuzzy rules, $i = 1, 2, \cdots, n$, and n is the total number of rules. Thus, the weight average of the rules output yields (3), shown at the bottom of the page, which represents the model output corresponding to the jth input pair $(x_{1j}, x_{2j}, \cdots, x_{kj})$. Note that j is the index of data set. The model output given in (3) can be used to calculate the membership and internal function's parameters by comparing with the given data. When the model is established, i.e., the parameters are obtained, it can be used to predict new results by giving new inputs. The total number of parameters of the membership functions and internal functions to be solved is $n \times (3 \times k + 1)$. These, in turn, are determined by minimizing the square of instantaneous errors between the data and the calculated outputs of the FLM.

III. Model Identification

A. Model Identification for Training Data

For a given collection of crisp training data $(x_{1j}, x_{2j}, \cdots, x_{kj}, y_j)$, the unknown structures and parameters of the membership and internal functions are determined by both the cluster estimation method and gradient-descent method. The accuracy of the FLM is given by the multiple correlation coefficient, R^2 [3], which is defined as follows:

$$R^2 = 1 - \frac{\sum_{j=1}^{m}(\hat{y}_j - y_j)^2}{\sum_{j=1}^{m}(\hat{y}_j - \overline{y})^2} \tag{4}$$

where $\hat{y}_j$ is the output of the FLM, y_j is the data point used for the model training, $\overline{y}$ is the sample mean of the data, and m is the total number of data sets. The cluster estimation method developed by Chiu [12] is used here as the coarse tuning process of a fuzzy model identification algorithm. Now we consider each data point as a potential cluster center and define a measure of the potential of data point x_i as

$$V_i = \sum_{j=1}^{m} \exp\left(-\frac{4}{r_a^2}\|x_i - x_j\|^2\right) \tag{5}$$

where $\|.\|$ denotes the Euclidean distance and r_a is a positive constant. The measure of the potential for a data point is related to the distances to all other data points. Note that the more distances between two neighboring data points that exceed the constant r_a, the less effect in measuring the potential. On the other hand, the larger of the constant r_a, the less effect on the neighboring points which yield more cluster centers. Initially, the constant r_a is used to define the radius of a neighborhood that consists of a set of training data points. Applying the cluster estimation method which consists of a set of potentials, V_i, to the training data, the number and the location of cluster centers are found. Note that, after the potential of every data point has been calculated, the data point with the highest potential is chosen to be the first cluster center. As concluded by Chiu [12]: "When we apply the cluster estimation method to a collection of input/output data, each cluster center is in essence a prototypical data point that exemplifies a characteristic behavior of the system. Hence, each cluster center can be used as the basis of a rule that describes the system behavior." Thus, the fuzzy rules are chosen to be equal to the number of cluster centers. As indicated from our examples, the result of the present coarse tuning process matches the training data well. When the coarse tuning process is completed, the fine tuning process of a fuzzy model identification algorithm that is based on the gradient-descent method is used. Here, we can then reformulate this problem by using a minimization of the square of instantaneous error between the output of the fuzzy logic model $\hat{y}_j$ and the current output reading y_j with respect to the unknown parameters p_o^i, p_l^i, c_l^i and σ_l^i of the internal functions and membership functions, i.e.,

$$E_j = \tfrac{1}{2}(\hat{y}_j - y_j)^2 = \tfrac{1}{2}e^2. \tag{6}$$

Applying the chain rule to (6), we obtain consequently the equations for updating the estimates of the unknown parameters p_o^i, p_l^i, c_l^i and σ_l^i:

$$\begin{aligned} p_0^i(t+1) &= p_0^i(t) - \alpha_0 \frac{\partial E_j}{\partial p_0^i} \\ &= p_0^i(t) - \alpha_0 v_j^i e \end{aligned} \tag{7}$$

$$\begin{aligned} p_l^i(t+1) &= p_l^i(t) - \alpha_1 \frac{\partial E_j}{\partial p_l^i} \\ &= p_l^i(t) - \alpha_1 v_j^i e_{lj} \end{aligned} \tag{8}$$

$$\hat{y}_j(x_{1j}, x_{2j}, \cdots, x_{kj}) = \frac{\sum_{i=1}^{n}\left\{\exp\left[-\frac{1}{2}\sum_{l=1}^{k}\left(\frac{x_{lj} - c_l^i}{\sigma_l^i}\right)^2\right]\right\}(p_0^i + p_1^i x_{1j} + p_2^i x_{2j} + \cdots + p_k^i x_{kj})}{\sum_{i=1}^{n}\left\{\exp\left[-\frac{1}{2}\sum_{l=1}^{k}\left(\frac{x_{lj} - c_l^i}{\sigma_l^i}\right)^2\right]\right\}} \tag{3}$$

$$\begin{aligned} c_l^i(t+1) &= c_l^i(t) - \alpha_2 \frac{\partial E_j}{\partial p_l^i} \\ &= c_l^i(t) - \alpha_2 v_j^i (p_0^i + p_1^i x_{1j} + p_2^i x_{2j} \\ &\quad + \cdots + p_k^i x_{kj} - \hat{y}_j) e \frac{x_{lj} - c_l^i(t)}{[\sigma_l^i(t)]^2} \end{aligned} \tag{9}$$

$$\begin{aligned} \sigma_l^i(t+1) &= \sigma_l^i(t) - \alpha_3 \frac{\partial E_j}{\partial \sigma_l^i} \\ &= \sigma_l^i(t) - \alpha_3 v_j^i (p_0^i + p_1^i x_{1j} + p_2^i x_{2j} \\ &\quad + \cdots + p_k^i x_{kj} - \hat{y}_j) e \frac{[x_{lj} - c_l^i(t)]^2}{[\sigma_j^i(t)]^3} \end{aligned} \tag{10}$$

and

$$v_j^i = \frac{w_j^i}{\sum_{i=1}^{n} w_j^i} = \frac{\exp\left[-\frac{1}{2}\sum_{l=1}^{k}\left(\frac{x_{lj} - c_l^i}{\sigma_l^i}\right)^2\right]}{\sum_{i=1}^{n}\left\{\exp\left[-\frac{1}{2}\sum_{l=1}^{k}\left(\frac{x_{lj} - c_l^i}{\sigma_l^i}\right)^2\right]\right\}} \tag{11}$$

where v_j^i are the normalized weights of the individual cells. In the present development, the gradient-descent method is used to minimize the instantaneous error. Since this method is basically a hill-climbing technique, it runs the risk of being trapped in a local minimize, where every small change in synaptic parameters p_o^i, p_l^i, c_l^i and σ_l^i increases the square error function E_j. The issue of estimating the initial parameter is crucial in this method. Another important issue is the number of fuzzy rules. In order to resolve these issues, the cluster estimation method is used to obtain an initial estimation of the membership functions and internal functions of the fuzzy rules from the training data. Note that, the initial estimations of the parameters c_l^i, p_0^i are the coordinates of the ith cluster center $(x_{1j}^*, x_{2j}^*, \cdots, x_{kj}^*, y_j^*)$ and, the parameter σ_l^i is defined as the width from the constant r_a of the cluster estimation method, i.e., $\sigma_l^i = k \times r_a$. By using the results obtained from Chiu and Yager [12], [13], the parameter k is set to be equal to $3/4\sqrt{2}$, and the parameter p_l^i is set to zero. As mentioned previously, in the present process modeling, the number of fuzzy rules is set to equal to be the number of cluster centers. The method is used extensively to update the parameters of the model until $R^2 \geq R_{\min}^2$. Note that the criterion $R^2 \geq R_{\min}^2$ is to assure that the FLM fits the training data accurately.

Given the initial r_a and training data, the present identification procedure consists of a coarse and fine tuning process that can be applied to establish a fuzzy rule. Note that, the structure obtained from this procedure is by no means an optimum structure with correct membership functions. We will deal with this issue in the next section.

B. Final Model Identification

To establish an FLM with a limited number of data, first, we need to identify an optimum structure from many intermediate FLM's that are obtained from the previous given procedure. Then, the final FLM can be established with optimum structure and all available data. The most challenging issue for the final model establishment is the quick identification of an optimum structure with proper assigned membership functions. A search algorithm similar to Tan, Xie and Lee's [14] has been proposed for the identification of the optimum structure. This algorithm partitions the available data into training and testing groups. During the search, the training data are used to the cluster estimation method and gradient-descent method for structure and the parameter identification to set up intermediate FLM's, and the testing data are used to evaluate these models' prediction accuracy. The accuracy of the model that fits the training data is represented by R_{train}^2, and the one that fits the testing data is represented by R_{test}^2. These R^2's can be calculated using (4). The proposed search algorithm provides the proper constant r_a instead of the number of membership functions for each input variable [14]. We start the algorithm with the smallest constant r_a, where it usually takes the largest number of cluster centers, fuzzy rules, and with the most complex FLM. The model's complexity is in general decreased by incrementing the value of constant r_a for the cluster estimation method. The incremental procedure is stopped when the total number of tuned parameters $n \times (3 \times k + 1)$ is smaller than the number of available data. The search algorithms can be summarized as follows.

1) Specify the input variables $x_{1j}, x_{2j}, \cdots, x_{kj}$, and the output variables y_j, $j = 1, 2, \cdots, k$.
2) Given $R_{\min}^2$ and Δr_a.
3) Provide the training and testing data.
4) Begin the search algorithm and set up the constant $r_a = \Delta r_a$.
5) Use the cluster estimation method and gradient-descent method to identify the FLM from training data until $R_{train}^2 \geq R_{\min}^2$ is satisfied. Test the corresponding FLM structure by using the testing data and record the R_{test}^2 as $R_{test}^2(r_a)$.
6) Increase the constant r_a with Δr_a and execute step e).
7) Repeat step f) until the total number of $n \times (3 \times k + 1)$ parameters less than the number of m available data.
8) Select the constant r_a that is corresponding to the maximum $R_{test}^2(r_a)$ as the optimum r_a, i.e., r_{a_optm}.
9) Set up the constant $r_a = r_{a\_optm} - 0.8\Delta r_a$ and repeat step e), f), g), and h), but in step f) the increment Δr_a is changed to $0.2\Delta r_a$.
10) Upon the r_{a_optm} has been reached, the cluster estimation method is used from training data and the gradient-descent method is used to refine the parameters of the final fuzzy logic model using both training and testing data.

To obtain the optimum structure, instead of identifying the number of fuzzy rules, we accomplished the task by changing the constant r_a of the cluster estimation method that is used to automatically identify the proper number of fuzzy rules that equal to the number of cluster centers. The advantage of this method is that we can establish the FLM efficiently and accurately. During the search, $R_{\min}^2$ is a necessary condition used to assure the trained model fits the training data well, and R_{test}^2 is used to identify the structure with the best testing

results. The incremental constant Δr_a is an important factor during computations, it determines the number of different FLM's that must be identified. From the observation of our numerical experiments, the parameter $R^2_{\min} = 0.99$ and $\Delta r_a = 0.1$ are used by considering both the computer run time and the accuracy, the value of $R^2_{\min}$ may be changed for different modeling process.

For the process modeling, as mentioned earlier, the number of data could be limited. Therefore, we have to follow the design of experiment (DOE) concept to identify critical input variables and to generate critical data. Different DOE approaches can be found in many text books. However, it is still a relatively new concept of partitioning the data into training and testing groups. For an efficient model establishment, the data should meet the following two requirements.

1) The test result of the model should reflect the model's overall accuracy in representing the process.
2) The testing data should be used for the final model training after the optimum structure is identified.

As a result, the testing data should be distanced from the training data during the group partition. An example for the partition is to generate the data using a central composite design and to generate the testing data using the Box-Behnkan design. This approach will be later illustrated in our case studies.

IV. Case Studies

To demonstrate the capabilities of the proposed FLM, a two-variable nonlinear function is demonstrated first. Then, a vertical barrel chemical vapor deposition (CVD) process is studied for the purpose of comparing existing process modeling algorithms.

A. Modeling of a Two-Variable Nonlinear Function

The nonlinear equation studied by Tan, Xie, and Lee [14] is given by

$$y = \sin(2x_1 + 4x_2) \tag{12}$$

The domains of the function were ranged between $-1.0 \leq x_1, x_2 \leq 1.0$ For the given function, the training data consists of 25 evenly distributed points on 5×5 grid, and the testing data consists of 16 evenly distributed points on a 4×4 grid which lied between the 5×5 grid. Fig. 2 shows the arrangement of these points.

In order to model the test function, the final model was identified using the procedure described in Section III. The predictive capability of the final FLM is obtained by computing $R^2_{predict}$ over a large number of data. These data were obtained from the test function consisting of 441 points evenly distributed on a 21×21 grid within the domain of interest. Figs. 3 and 4 plot the curves of the test function and the corresponding curves from FLM. As shown in the figures, the FLM matched closely to the nonlinear function. If the $R^2_{predict}$ is equal to 1.0, the established FLM predicts the values of the function at these given 441 points perfectly. The $R^2_{predict}$ can be treated as the model's accuracy in predicting functional behaviors. There are three different FLM's and an ANN model were obtained from Schaible [15]. The optimum structure from each FLM was obtained using the method proposed by Tan, Xie, and Lee [14], and the optimum structure for the ANN model was obtained using the method presented by Wang and Mahajan [16]. The first type of trained FLM was

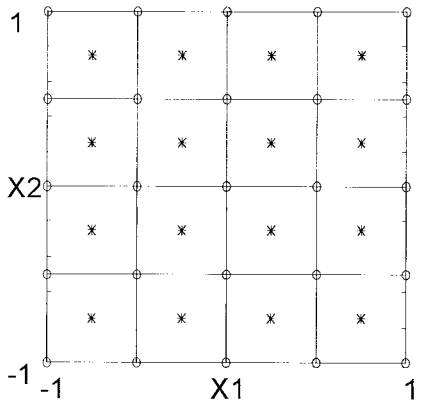

Fig. 2. Training and testing data.

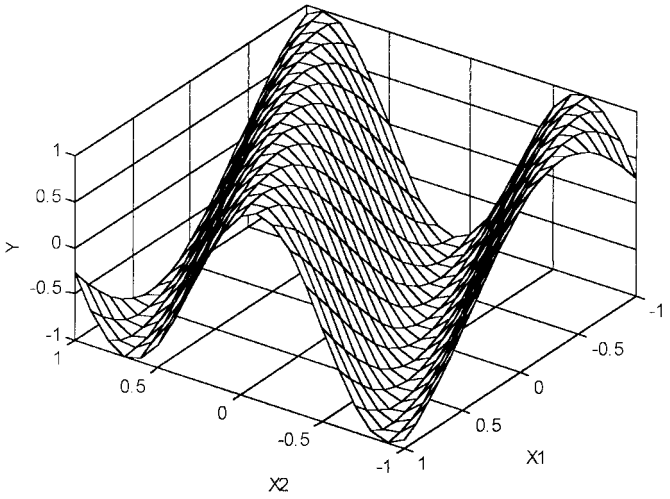

Fig. 3. Mesh from the test function.

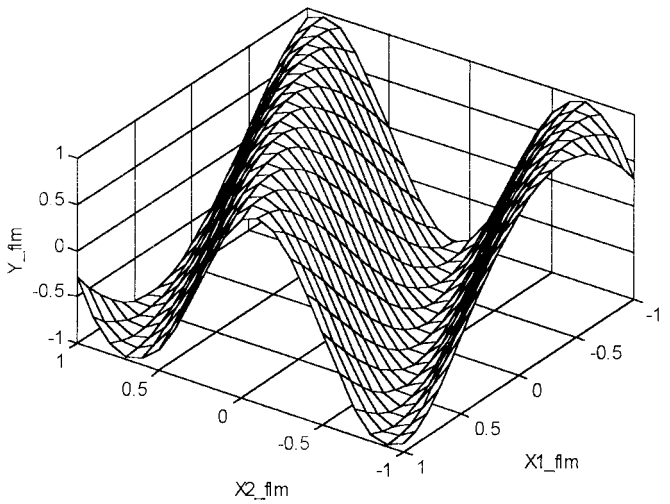

Fig. 4. Mesh of FLM from the test function.

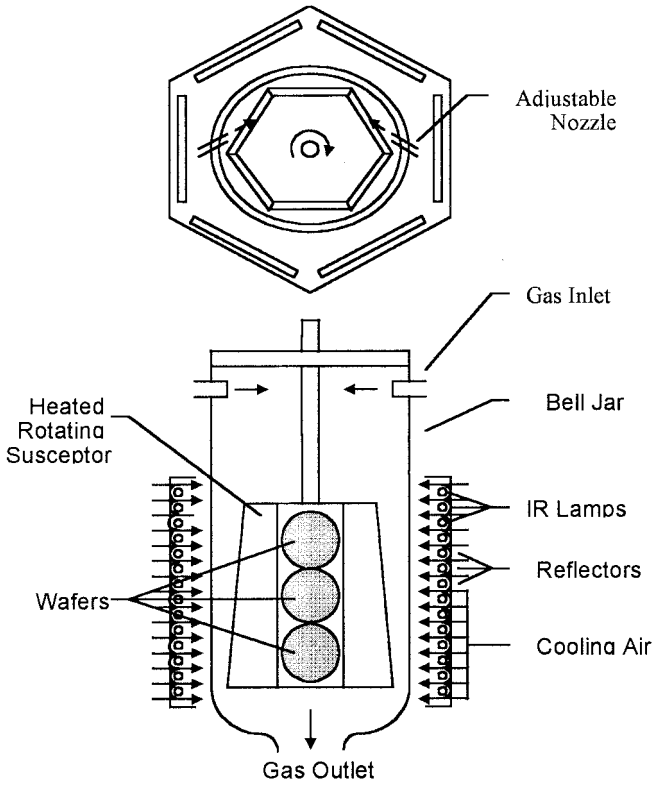

Fig. 5. Barrel CVD reactor schematic.

TABLE I
TRAINING RESULTS FOR THE TWO-VARIABLE FUNCTION $y = \sin(2x_1 + 4x_2)$

Model Type	Rules(Structure)	$R^2_{predict}$
FLM Type I	28(4,7)	0.9633
FLM Type II	12(3,4)	0.9339
FLM Type III	36(6,6)	0.9607
ANN	(2,4,1)	0.9982
Cluster based FLM	7(r_a=1.4)	0.9988

the most traditional one in which the triangular membership function along with the minimum operator was employed. These models will be hereon referred to as "Type I." The second type, hereon referred to as "Type II," also made use of triangular membership functions but employed the product operator. The third type, hereon referred to as "Type III," incorporated the quadratic membership function along with the product operator. For the purpose of comparison, the result of four different FLM's and an ANN model were listed in Table I. As a result, the cluster estimation based FLM is more accurate and simple than existing FLM's and the ANN model.

Note that, for Types I, II, and III FLM's, the number of rules is equal to the product of the two numbers denoting the structure. This is due to the fact that we cannot separate the structure identification from the parameter identification since they have a mutual relationship. Also, the structure of the three-layer ANN is based on Wang's [16] result that is constructed by giving the number of neurons at each layer. For cluster based FLM, instead of identifying the number of fuzzy rules, we try to obtain the optimum structure by varying the radius, r_a. Once optimum r_a is obtained, the optimum structure is decided by letting the final number of cluster centers equal to the number of rules. Clearly, as shown in Table I, the number of rules needed for cluster based FLM is less than Types I, II, and III FLM's. This reason is that during the process of clustering, the relationships between the data point groups and its corresponding cluster centers have been established.

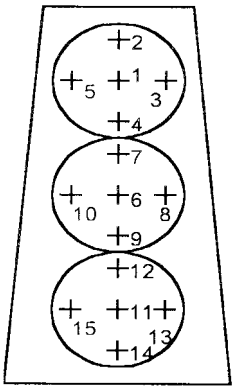

Fig. 6. Deposition thickness obtained points on the wafers.

TABLE II
SETTING OF THE CVD MODEL INPUTS

Variable Name	Center Level	Lower Oper. Level	Upper Oper. Level	Lower Test Level	Upper Test Level	Lower Start Pt. Level	Upper Start Pt. Level
LBV	66.0	58.0	74.0	62.0	70.0	46.0	86.0
RBV	76.0	68.0	84.0	72.0	80.0	56.0	96.0
JetX	5.0	3.0	7.0	4.0	6.0	0.0	10.0
JetY	5.0	3.0	7.0	4.0	6.0	0.0	10.0
H2M	45.0	34.0	56.0	39.5	50.5	17.5	72.5
H2R	55.0	44.0	66.0	49.0	60.5	27.5	82.5

B. Modeling a Simulated CVD Process

Our second test case, the CVD epitaxial growth of silicon, is a widely used process for the fabrication of thin solid films with applications in the production of microelectronics devices. This study is similar in scope to the artificial neural network modeling process proposed by Mahajan *et al.* [16]. In a reactor, as illustrated in Fig. 6 [17], silicon wafers are positioned in shallow pockets in a "susceptor," which is a multisided, slightly tapered, silicon carbide coated graphite assembly that is slowly rotated inside a cooled quartz bell jar. An array of infrared lamps backed by reflectors surrounds the bell jar. A carrier gas, typically silicon tetrachloride, dichlorosilane, or trichlorosilane, hydrogen, and a dopant such as diborotrichlorosilane, enters through two nozzles at the top of the reactor, and is exhausted through a central hole at the bottom of the reactor. Each nozzle has two degrees of freedom, one along the vertical axis and the other along the horizontal axis. Usually the two nozzles are identically oriented. The temperature is typically (1050–1200 °C) and the ambient working pressure may be either at atmospheric pressure (about 1 bar) or about one-tenth of a bar (reduced pressure mode).

TABLE III
THE OPTIMIZATION PROCESS: DATA SET 1

Exp. #	LBV	RBV	Jetx	Jety	H2m	H2r	Tvar
1	66.00	76.00	5.00	5.00	45.00	55.00	7.738
2	86.00	76.00	5.00	5.00	45.00	55.00	63.472
3	46.00	76.00	5.00	5.00	45.00	55.00	135.8
	.....	.....	.....	.....	.....	.....	.....
22	58.00	84.00	3.00	3.00	34.00	66.00	92.114
23	58.00	84.00	3.00	3.00	56.00	44.00	102.026
24	58.00	84.00	3.00	7.00	34.00	44.00	105.456
25	58.00	84.00	3.00	7.00	56.00	66.00	102.464
	.....	.....	.....	.....	.....	.....	.....
46	70.00	80.00	6.00	6.00	45.00	55.00	20.947
47	70.00	80.00	6.00	4.00	45.00	55.00	8.864
48	70.00	80.00	4.00	6.00	45.00	55.00	4.888
	.....	.....	.....	.....	.....	.....	.....
60	62.00	72.00	4.00	6.00	45.00	55.00	4.888
61	62.00	72.00	4.00	4.00	45.00	55.00	4.133

An important consideration in epitaxial growth is that the deposited layer on the wafers be uniform in thickness. The available data was generated using a model for the process [17] developed by Dr. Herb Lord. The model simulates deposition of silicon on silicon wafers, and the model output represents silicon deposit thickness. This model is a fourth order polynomial of two variables, x and y, which represent (x, y) coordinates of a point on a silicon wafer. The coefficients of the polynomial equation are functions of the six process inputs. The six input variables are given as follows: LBV is the left gas valve reading; RBV the right gas valve reading; $JetX$ the angular deviations of the gas jets from the horizontal plane; $JetY$ the angular deviations of the gas jets from the vertical plane; $H2M$ the main hydrogen gas flow rate reading; $H2R$ the rotational hydrogen gas flow rate reading. For a given recipe, or combination of the six input variables, silicon thickness is obtained at 15 different (x, y) points, at five points for each of top, middle, and bottom wafers, as shown in Fig. 5. The output of the process was the thickness variance T_{var} of these 15 points is calculated as

$$T_{var} = \frac{1}{15}\sum_{j=1}^{15}\left(\frac{\overline{T} - T_i}{\overline{T}}\right)^2 \tag{13}$$

where T_i represents each of the 15 individual thickness, and $\overline{T}$ represents the average value of 15 individual thickness.

Table II gives the operating range levels for the six input variables along with input levels used to generate testing points and central composite "star" points which are schematically shown in Fig. 7.

The case studied here is the simulated CVD process using all six input variables with T_{var} as the output variable. Three training/testing data sets were generated for the six input cases. These data sets are described as follows.

- Data set 1 (Table III) contained 45 training points and 16 testing points. The training set include 12 central composite start points, one center point, and 32 points from a one-half fractional factorial two level experimental design. The 16 test points from a full factorial two level design conducted at the test levels of the four variables that have the most effect on the CVD process: LBV, RBV, $JetX$, and $JetY$. The levels of the two remaining variables, $H2M$ and $H2R$, were held constant at their value for these 16 test points.
- Data set 2 contained 77 training points and 64 testing points in the six dimensional input space. The training set included 12 central composite star points, one center point, and 64 points from a one-half fractional factorial two levels of experimental design. The testing set consisted of 64 points from a full factorial two level design at the upper and lower test levels.
- Data set 3 table also contained 141 points and was identical to data set 2 except for the fact that 12 star points in the training set were replaced by analogous points that lie at the center of a cube face.

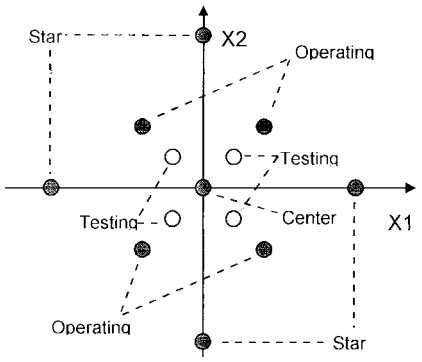

Fig. 7. Training and testing data settings.

A single prediction data set that included 200 random points uniformly distributed throughout the six dimensional

TABLE IV
TRAINING RESULTS FOR CVD MODEL USING DATA SET 1

Model Type	Rules(Structure)	$R^2_{predict}$
FLM Type I	270(3,3,5,3,2,1)	-0.4282
FLM Type II	24(2,2,3,2,1,1)	0.8766
FLM Type III	24(2,2,3,2,1,1)	0.9277
ANN	(6,6,1)	0.7298
Cluster based FLM	7(r_a=1.46)	0.9651

TABLE V
TRAINING RESULTS FOR CVD MODEL USING DATA SET 2

Model Type	Rules(Structure)	$R^2_{predict}$
FLM Type I	288(3,3,2,2,2,5)	-1.9583
FLM Type II	24(2,2,3,2,1,1)	0.7469
FLM Type III	72(3,3,4,1,2,1)	0.9300
ANN	(6,7,1)	0.7711
Cluster based FLM	8(r_a=1.48)	0.9731

TABLE VI
TRAINING RESULTS FOR CVD MODEL USING DATA SET 3

Model Type	Rules(Structure)	$R^2_{predict}$
FLM Type I	256(2,4,2,2,2,4)	-1.3444
FLM Type II	16(2,2,1,2,2,1)	0.9152
FLM Type III	16(2,2,1,2,2,1)	0.9152
ANN	(6,6,1)	0.7611
Cluster based FLM	8(r_a=2.18)	0.9750

input space was also generated for the purpose of comparing predictive capabilities of which were also obtained.

The following tables show the training results for the existing and proposed model types by using three different data sets. The data sets, FLM Types I, II, III, and ANN models are provided by Schaible [15].

Based on the data given in Tables IV–VI, Type I FLM's show poor results in predicting the CVD process. In addition, there are many derivative discontinuities in the model. As shown from the tables, Type II FLM's with the product operator improved the predictive capabilities, but the triangular function still has a derivative discontinuity at the median value of the variable. This may make the Type II FLM's undesirable in certain applications. The Type III FLM's use the quadratic membership function along with the product operator, they have the advantage of being continuously differentiable. From Tables IV, V, and VI, the Type III FLM's show more predictive capability than Type I FLM's and Type II FLM's. Note that, during the process of the present study, ANN model did not yield good results as shown from the tables.

From Tables IV–VI, the cluster based FLM has shown its superior predictive capability for all three of the six dimensional data sets that are studied here. Moreover, the present FLM has the advantage of being continuously differentiable. This makes the cluster based FLM more desirable to the applications that require computation of derivatives from the process model.

V. CONCLUSION

A new algorithm has been proposed for the efficient establishment of a cluster based fuzzy logic model process modeling. With a limited number of data available, the accuracy of the FLM is strongly depended on its structure. A decision rule based on the cluster estimation method and the multiple correlation coefficient is used to obtain the structure of FLM. Upon the optimum structure has been reached, the back-propagation algorithm is used to refine parameters of the FLM using both training and testing data. This algorithm has applied to a vertical CVD process. By comparing with other fuzzy and ANN modeling methods, the proposed FLM has the advantages of effectiveness and high accuracy.

ACKNOWLEDGMENT

The authors would like to thank Professor Y. C. Lee and B. Schaible of University of Colorado at Boulder for their helpful comments and suggestions during the development of this paper.

REFERENCES

[1] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, pp. 116–132, Jan./Feb. 1985.
[2] B. Schaible, H. Xie, and Y. C. Lee, "Fuzzy logic models for tanking process effects," *IEEE Trans. Fuzzy Syst.*, vol. 5, pp. 545–556, Nov. 1997.
[3] H. Xie, R. L. Mahajan, and Y. C. Lee, "Fuzzy logic models for thermally based microelectronic manufacturing processes," *IEEE Trans. Semiconduct. Manufact.*, vol. 8, pp. 219–227, Aug. 1995.
[4] R. R. Yager and D. P. Filev, "Unified structure and parameter identification of fuzzy models," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 23, no. 4, pp. 1198–1205, July/Aug. 1993.
[5] L. Wang and R. Langari, "Building Sugeno-type models using fuzzy discretization and orthogonal parameter estimations techniques," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 454–458, Nov. 1995.
[6] ——, "Complex systems modeling via fuzzy logic," *IEEE Trans. Syst., Man, Cybern. B*, vol. 26, pp. 100–106, Feb. 1996.
[7] H. S. Hwang and K. B. Woo, "Linguistic fuzzy model identification," *Inst. Elect. Eng., Proc. Control Theory Appl.*, Nov. 1995, vol. 142, no. 6.
[8] R. R. Yager and L. A. Zadeh, *An Introduction to Fuzzy Logic Applications in Intelligent Systems.* Norwell, MA: Kluwer Academic, 1992.
[9] E. Czogala and W. Pedrycz, "Some problems concerning the construction of decision making in fuzzy systems," *Int. J. Man-Mach. Stud.*, vol. 15, pp. 201–211, 1981.
[10] W. Pedrycz, "Approximate solutions of fuzzy relation equations," *Fuzzy Sets Syst.*, pp. 183–202, 1989.
[11] H. Xie, Y. C. Lee, R. L. Mahajan, and R. Su, "Process optimization using a fuzzy logic response surface method," *IEEE Trans. Comp., Packag., Manufact. Technol. A*, vol. 17, pp. 202–211, June 1994.
[12] S. L. Chiu, "A cluster estimation method with extension to fuzzy model with extension to fuzzy model identification," in *Proc. Third IEEE Conf. Fuzzy systems*, 1994, vol. 2, pp. 1240–1245.
[13] R. R. Yager and D. P. Filev, *Essentials of Fuzzy Modeling and Control*, 1st ed. New York: Wiley, 1994.
[14] J. Tan, H. Xie, and Y. C. Lee, "Efficient establishment of a fuzzy logic model for process modeling and control," *IEEE Trans. Semiconduct. Manufact.*, vol. 8, pp. 50–61, Feb. 1995.
[15] B. R. Schaible, "Modeling nonlinear processes using fuzzy logic," M.S. thesis, Univ. Colorado, Boulder, 1995.
[16] X. A. Wang and R. L. Mahajan, "CVD epitaxial deposition in a vertical barrel reactor: Process modeling and optimization using neural network models," *J. Electrochem. Soc.*, vol. 142, no. 9, pp. 3123–3132, Sept. 1995.
[17] H. A. Lord, "Convective transport in silicon epitaxial deposition in a barrel reactor," *J. Electrochem. Soc.*, vol. 134, no. 5, pp. 1227–1235, 1987.

J. C. Chiou received the Ph.D. degree in aerospace engineering science from the University of Colorado, Boulder, in 1990.

Before joining National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., in 1992, he worked at the Center for Space Structure and Control, University of Colorado, Boulder, as a Research Associate (1991–1992). His research interests include fuzzy logic modeling and control, microelectromechanical systems, servo control of CD-ROM and DVD, and modeling and control of multibody dynamic systems. He is the coauthor of an advanced text/reference book on CD-ROM system technology.

J. Y. Yang was born in Keelung, Taiwan, R.O.C., in 1970. He received the B.S. degree from the National Chiao-Tung University (NCTU), Hsinchu, Taiwan, R.O.C., in 1992. He is currently pursuing the Ph.D. degee in the Department of Electrical and Control Engineering, NCTU.

His research interests include applications of fuzzy models for nonlinear process, servo control of brushless DC motors, modeling and control of dynamic systems.