Optimal design using neural network and information analysis in plasma etching

Junghui Chen[a)]
Department of Chemical Engineering, Chung-Yuan Christian University, Chung-Li, Taiwan 320, Republic of China

Paul Po-Tao Chu
Etching Process Section Module B Engineering-3 Department, Taiwan Semiconductor Manufacturing Company Ltd., Hsin-Chu, Taiwan, Republic of China

David Shan Hill Wong and Shi-Shang Jang
Department of Chemical Engineering, National Tsing-Hua University, Hsin-Chu, Taiwan 30034, Republic of China

(Received 7 May 1998; accepted 30 October 1998)

The goal of this article is to use operating data-based approaches for automating the manufacturing of submicron flash memory devices in semiconductor. This novel technique which combines the neural network and information inductive analysis has recently been proposed. It is used in this article to generate the recipe for plasma etching process design. Traditional plasma etching variables such as pressure, gas flows, temperature, rf power, etc., are used to build the neural network for predicting etching rate of polysilicon and field oxide, and the uniformity of field oxide. The information inductive analysis based on the information entropy and fuzzy clustering analysis is then utilized to look for the candidate points in each optimal region whose response surface is constructed based on the neural network model. With only a few runs, the best optimal condition getting close to the design requirements is found. Since the complexity of plasma modeling and design at the equipment level is presently ahead of theoretical method from a fundamental physical standpoint, the proposed method can effectively cope with nonlinear characteristics in the plasma etching process, giving good design directions and taking advantage of traditional statistical approaches. Using the proposed method within four runs and a total of 26 experimental points, the recipe that meets all specifications is found.
[S0734-211X(99)01901-0]

I. INTRODUCTION

Over the last decade, intensive global competition among semiconductor manufacturers has called for the need of finding a good recipe for innovative products. If a new product cannot be made just in time to meet the need in the market, it would be outdated or even no longer wanted. It is also important that accelerating product design should be done at the minimum costs without the expense of product quality. In the semiconductor field, plasma etching is regarded critical to product quality. In this article, a novel technique that is a combination of the neural network and information inductive analysis is employed to generate the recipe for plasma etching process design.

The proposed method can effectively cope with nonlinear characteristics in the plasma etching process, giving good design directions and taking advantage of traditional statistical approaches. Neural networks have demonstrated the strong capability of learning nonlinear and complex relationships between process variables without any prior knowledge of system behavior. This particularly fits the highly complex process of plasma etching. Since the complexity of plasma modeling and design at the equipment level is presently ahead of the theoretical method from a fundamental physical standpoint, many researchers have focused on the empirical approach to plasma modeling, like neural networks. In recent years, the number of applications of neural networks to etching process has increased dramatically.[1–5] Research shows that neural network models exhibit superior predictive ability over traditional statistical methods and require less experimental training data. Literature that deals with the use of neural networks to solve semiconductor fabrication problems includes: the polysilicon film growth by low-pressure chemical vapor deposition,[6] the removal of polysilicon film by plasma etching,[7] the behavior of real-time reactive ion etching process,[8] etc.

Basically, a neural-network approach can save time and money. It also learns and extracts the process behavior from the past operating information. It can be used as a model for process optimization and control design. However, a large amount of data points is usually needed for training neural network models. It is not applicable to process design since only the data located at the optimal design region are critical to building the neural network model. In order to reduce the amount of training data, a random search is utilized to look for the possible optimal regions. The information inductive analysis based on the information entropy and fuzzy clustering analysis is employed to determine the good candidate

[a)]To whom correspondence should be addressed; electronic mail: jason@wavenet.cycu.edu.tw

points at each optimal region. The information inductive analysis is regarded as a response surface management strategy for constructing and updating the neural network model. Therefore, further experimental runs should be done to verify those candidates and update the neural network model. With only a few runs, the best optimal condition which is getting close to the design requirements is found.

In this study, the proposed method is applied to the self-aligned source (SAS) process technology.[9] Etching is performed on a test structure design to facilitate the simultaneous measurement of the etching rate of polysilicon and field oxide, and the uniformity of field oxide. The first two variables are required for etching selectivity. The most important parameters during the design process are rf power, chamber pressure, bottom electrode temperature, CO flow, and Ar flow. The CO gas is considered as one of the candidates because it can help improve the oxide etching selectivity to silicon. The most challenging etching process in the SAS field oxide dry etching process is that the silicon substrate is exposed in the plasma environment during the etching process. This is different from the conventional contact via dry etching process where the silicon substrate is protected by dielectric film until the end point is approached. Through this study, the proposed method not only exhibits superior accuracy and effectiveness but also utilizes less experimental points when searching for the desired recipe.

This article is organized into five sections. In the next section, the SAS process is introduced. Section III explains the core of the proposed experimental design. Experimental study and discussion are addressed in Sec. IV. Section V gives concluding remarks.

II. SELF-ALIGNED SOURCE ETCHING PROCESS CHARACTERIZATION

In the manufacturing of submicron flash memory devices, the SAS process technology is adopted. By using the SAS design and process, the shape of the source or drain region is made so straight that there is no variation in the coupling ratio. By minimizing the variation in coupling ratio, the tight threshold voltage distribution after the program and erasing operation can thus be achieved.

The SAS field oxide dry etching process is the most challenging etching process. As illustrated in Fig. 1, the silicon substrate has been exposed in the plasma environment since the etching process started. The SAS field oxide dry etching process is supposed to meet the criteria of both removing the field oxide effectively and sustain minimum amount of silicon loss. According to the device memory cells' data retention performance data, the silicon loss should be controlled at the level of less than 30 nm. In order to maintain reasonable oxide etching rate (throughput concern) and control the silicon etching rate to minimize silicon loss, the oxide etching recipe's selectivity of oxide to silicon on the patterned wafer should be maintained at the level of greater than 30.

Etching is performed on a test structure design to facilitate the measurement of the etching rate/uniformity of polysilicon and SiO_2. These data are required to determine the

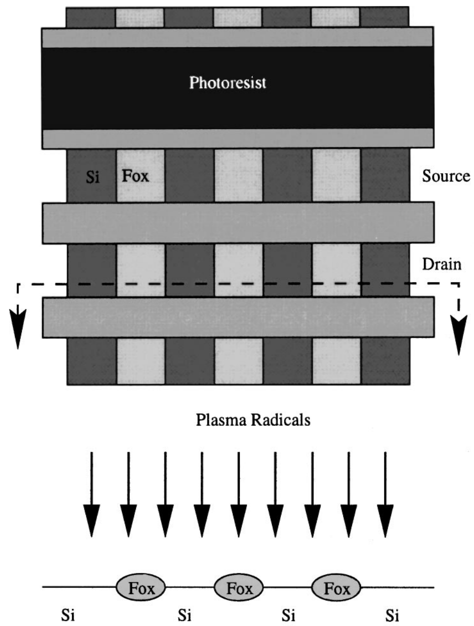


FIG. 1. In SAS etching process, Fox represents field oxide and Si is silicon.

etching selectivity of SiO_2 to silicon substrate. Test patterns are fabricated on 6-in.-diameter silicon wafers. Current oxide to silicon selectively performance of the CHF_3/CF_4 chemistry based reactive ion etch (RIE) mode oxide etcher ranges from approximately 15 to 1. According to the literature data,[10] selectivity is primarily determined by the C/F ratio in the etchant chemicals. The increase of the C/F ratio can improve the selectivity of oxide to silicon, so CO gas has been considered as one of the factors which can help improve the oxide etch selectivity to silicon. Besides CO flow, the other critical parameters in this recipe design including rf power, chamber pressure, bottom electrode temperature, and Ar flow will determine the oxide to silicon selectivity. In order to acquire the process data, an etching monitoring system transferring data from the RIE chamber onto a workstation has been designed and implemented. LAM Research RIE mode 4520 single-wafer parallel–plate system operating at 400 kHz is chosen as the plasma etching tool in this study. The patterned wafer with the deposition of SiO_2 and polysilicon is etched for 30 and 120 s, respectively. The remaining thickness of the oxide and polysilicon before and after plasma etching are measured by PROMETRIX FT-750 film thickness probe system on specific pattern opening area. The five input factors and their corresponding range of operation are shown in Table I. The etching process in this study is repre-

TABLE I. Operation ranges of input variables.

Variable	Range	Unit
Power	700–900	W
Pressure	200–300	mTorr
Bottom electrode temperature	−20–+10	°C
CO gas flow	100–300	sccm
Ar gas flow	100–300	sccm

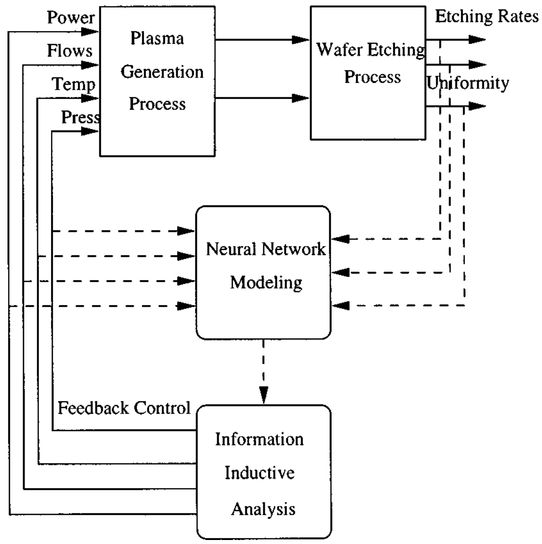


FIG. 2. Optimal experimental design architecture.

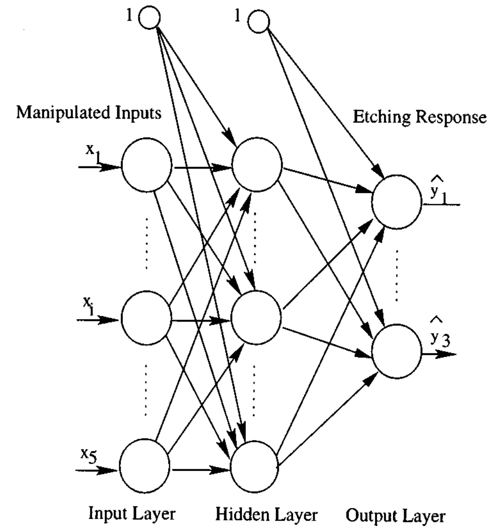


FIG. 3. Architecture of a three-layer feed forward neural network.

sentative of a typical plasma etching fabrication used for IC processes.

The objective is to find a recipe (or an operating condition) that satisfies the following specifications:

$$\text{etching rate of field oxide (OX E/R)} > 4000\ \text{Å/min}, \tag{1}$$

$$\text{uniformity of field oxide (OX U)} < 5\%, \tag{2}$$

$$\text{etching rate of polysilicon (Poly E/R)} < 100\ \text{Å /min}, \tag{3}$$

where uniformity is calculated by max/min.

III. INTEGRATION OF NEURAL NETWORK AND INFORMATION INDUCTION

An optimal experimental design architecture that integrates the neural network and information induction is implemented for SAS process (Fig. 2). The past operating data are collected for the feedforward neural network model to predict plasma system behavior. The information inductive tool is used to extract and suggest the new manipulated conditions. It also maintains the desired process variables at the desired region, such as etching rate or uniformity, for the next run. The main advantage of the control of the plasma variable is adjusting the manipulated parameters that are more directly related to etching characteristics. Additionally, the integrated approach provides etching modeling and process information both during and between process run-to-run. The entire methodology is described in our previous work. For a detailed coverage, see Ref. 11.

A. Modeling using neural network

Neural networks are a recent development in the field of plasma etching. They can learn the arbitrarily complex, nonlinear relationship between inputs and outputs. By training the net, the back-propagation algorithm reduces the difference between the neural network output and the actual experimental values. Once trained, each neural network represents a nonlinear or complex function for the output that it learned.[12,13] The major advantage is that the network is derived from the data presented instead of the exact form of the analytical function on which the model should be built. However, for general statistical models, a prior choice cannot be assumed for the functional form. Moreover, a small number of experiments may not be enough to develop approximation to the function. The neural network is ideally suited to semiconductor process modeling mainly because of the ability to directly learn the input–output relationships.

The process under consideration in this study is etching of the field oxide in a CO-and-Ar plasma. There are five relevant input parameters: rf power, chamber pressure, bottom electrode temperature, and the flow of the two gases. The responses of interest to be controlled are the etching rate of the polysilicon (Poly E/R), field oxide (OX E/R), and the uniformity of field oxide (OX U). They are functions of the previous five input variables. The ability of the neural network to learn the relationships between the five inputs and three outputs is attributed to the multiple parallel processing units that are interconnected to each other and the weight of the connection that stores the experimental knowledge. The structure of a network in this study composed of three layers is shown in Fig. 3. In the plasma etching application, the input layer of neurons receives external information corresponding to the five adjustable design parameters, $\mathbf{x}^p = [x_1^p, \ldots, x_5^p]$. The output layer transmits information to the outside world, corresponding to the predicted controlled or response variables (in this case, the etching rate of the polysilicon and field oxide, and the uniformity of field oxide), $\mathbf{y}^p = [y_1^p, \ldots, y_3^p]$. The input and the output components of the pth data pair are defined by $\{\mathbf{x}^p, \mathbf{y}^p\}$. The hidden layer composed of N_h neurons can be considered as representation in the fundamental physical characteristics of plasma. Its output is given by

$$h_j(\mathbf{x}^p)=\sum_{i=1}^{5} w_{ji}^h x_i^p + b_j^h, \quad j=1,2,\ldots,N_h, \tag{4}$$

and the response output is given by

$$\hat{y}_k(\mathbf{x}^p)=\sum_{j=1}^{N_h} w_{kj}^o z[h_j(\mathbf{x}^p)] + b_k^o, \quad j=1,2,3, \tag{5}$$

where w_{kj}^o and w_{ji}^h are the weights between the output and the hidden layers and the weights between the hidden and the input layers, respectively, b_k^o and b_j^h are the biases in the hidden layers and the output layers, respectively, and z is the output in the hidden layer. The hyperbolic tangent activation function is used in this study.[14] The sum square error, E, which represents the error between the predicted and targeted values is employed to evaluate the performance of the network:

$$E=\frac{1}{P}\sum_{p=1}^{P}\sum_{k=1}^{3}[y_k(\mathbf{x}^p)-\hat{y}_k(\mathbf{x}^p)]^2, \tag{6}$$

where P is the number of experimental data.

In this study, the pseudo-Gauss–Newton method algorithm[15,16] is used for training. Note that the network has no fixed topology. That is, the number of hidden neurons varies, depending on the desired network performance. Cross validation is used when training and testing the neural network. Due to the small quantity of the experimental (training) data available, a statistical technique called the leave-one-out (LOO) cross-validation scheme[17] is used to overcome the possible overfitting or bias introduced by relying on any particular division into testing and training sets. This is an attractive vehicle for generalization assessment of a neural network model. Although it is computationally expensive, time loading is not very heavy due to the small experimental data used in this study.

B. Decision making based on information induction

In this information induction step, it contains two modules: *regional optimal search* and *fuzzy information analysis*. The purpose of this step is to create some optimal regions and possible optimal operating conditions based on the neural network model generated in the previous step.

The first module, *region optimal search*, extracts features from the previously built model. In product and process development, the feature of interest is the condition that satisfies the desired optimal operation. However, multiple local optima are frequently encountered. It is often necessary to rate alternative local optima based on secondary objectives such as robustness, safety, etc. Therefore, nongradient based search procedure, like random search or genetic algorithm,[18,19] should be used. It should be pointed out that this module is what we do with this population. Therefore, the optimal regions can be gotten. Figure 4(a) shows that the representation of the performance surface $Z(x_1,x_2)$ of a neural network model is built for the experimental data

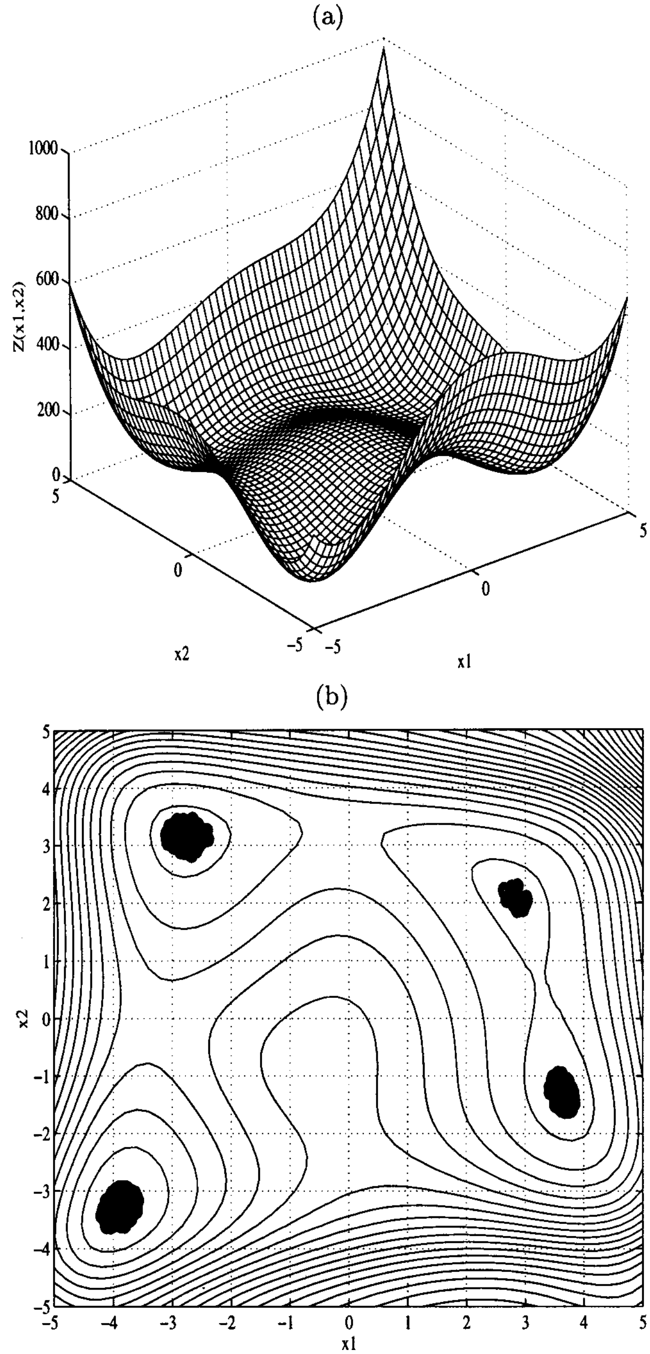


FIG. 4. Optimal-search algorithm. (a) The function of two variables has four local minima. (b) The optimal regions are found based upon the search results. The background shows the contour of the original model.

(Z,x_1,x_2). The optimal regions generated by the optimal search against the contour background of the neural network model are depicted in Fig. 4(b).

In reality, it is impossible to perform experiments at all the points in all the optimal regions. The secondary module, *fuzzy information induction*, is proposed here to select the most representative candidate points. Then the experiments will be performed only at those candidate points. The fuzzy information induction technique is composed of the fuzzy classification and information analysis for classifying the possible optimal regions.

The fuzzy c-means (FCM)[20,21] is used here for classification. The purpose of the clustering process is to distill a certain number of homogeneous clusters or classes from a large data set and to classify concise representation of the individual local optimal behavior. With the preset number of clusters performing FCM, the degree of its membership μ_{ik} can be obtained to characterize the data point as being, to a greater or less degree, a member of the appropriate cluster. The data points for each class also correspond to the cluster center $\mathbf{c}^i$. Here, μ_{ik} is a fuzzy membership which measures the degree of association of the kth data point $\mathbf{x}^k$ with the ith cluster class and $\mathbf{c}^i$ is the center of the ith cluster.

In general, more classes can help clarifying the picture of classification. However, additional classes increase our burden since we have to perform test experiments at each clustering center. The more classes we use, the more experiments we need to perform. Our goal is not to produce the crispiest classification but to locate optima as quickly as possible. Hence, an information induction is used to determine the optimal number and location of the next set of experiments. The information analysis technique is based on the concept of entropy and predicted performance for classifying the possible optimal regions. The entropy measures how well a set of cluster means classifies the data points, and energy measures how well a set of cluster means performs if it is chosen as the next set of experiments. The composite information formula is defined:

$$F=U-TS. \tag{7}$$

This equation is balanced by three terms: information energy, U, information entropy, S, and normalization factor, T.

The information energy, U, which is just the expected value of the performance index is defined:

$$U=\sum_{j=1}^{C}\frac{N_j}{N}f[\hat{\mathbf{y}}(\mathbf{c}^i)]-f_{\min}, \tag{8}$$

where $f_{\min}$ is the value of the minimum f recorded in the optimal search and $f[\hat{\mathbf{y}}(\mathbf{c}^i)]$ is the performance index evaluated at the cluster centers. The information energy is a measure of the relevance of the messages generated by the clustering analysis to the optimization procedure.

The concept of entropy, S, is used to measure the purity of a class, i.e., the distribution among the classes of the process variables within the set. The fuzzy entropy of classification which stems from the concept of Shanon's entropy[22,23] can be calculated from

$$S(\mathbf{c}^i)=\sum_{\mathbf{x}\in X}p(\mathbf{x}^k|\mathbf{c}^i)\ln[p(\mathbf{x}^k|\mathbf{c}^i)], \tag{9}$$

$$S(C)=\sum_{i=1}^{C}\frac{N_i}{N}S_i, \tag{10}$$

where $S(\mathbf{c}^i)$ is the entropy of classification for the ith cluster and $p(\mathbf{x}^k|\mathbf{c}^i)$ is the probability of finding $\mathbf{x}^k$ as a representative of the ith cluster, i.e., $p(\mathbf{x}^k|\mathbf{c}^i)=(\mu_{ik}/N_i)$ and $N_i=\Sigma_{k=1}^{N}\mu_{ik}$. Equation (10) gives weighted average of the entropies for each class. The weights are obtained by dividing the sum of the fuzzy values found in the ith cluster by their total. Note that the value of this total equals the number of points in the whole optimal regions. Therefore, the entropy of the entire classification set can be defined as

$$S(C)=\frac{1}{N}\left(\sum_{i=1}^{C}\sum_{k=1}^{N}\mu_{ik}\ln\mu_{ik}-\sum_{i=1}^{C}N_i\ln N_i\right). \tag{11}$$

In the above equation, the first term of the right-hand side, $\mu_{ik}\ln\mu_{ik}$, represents the penalty of overlapping between groups. If the data points belong to one group (i.e., $\mu_{ik}\rightarrow 1$ and $\mu_{jk}\rightarrow 0$, $i\neq j$), the contribution of these data points to the first term is negligible. On the other hand, if the data points belong to a large number of groups, μ_{ik} will be nonzero for several groups and the term, $\mu_{ik}\ln\mu_{ik}$, becomes significant. The second term of the right-hand side is a measure of the size of each cluster. The smaller the clusters are, the larger and more orderly the entropy is. Note that $S\rightarrow 0$ at $C\rightarrow N$. However, the increase in entropy with the increased number of clusters is offset by how clearly we can divide the data into groups.

In order to avoid the issue of order of magnitude difference between U and S, a proper scaling normalization factor, T, in the Eq. (7) is added to the energy term:

$$T=\frac{f_{\max}-f_{\min}}{\ln N}, \tag{12}$$

where $f_{\max}$ is the maximum f of all surviving points in the regional optimal search. During the procedure of determining the number of clusters, T remains constant. This is analogous to the thermodynamic equilibrium criterion under isothermal condition that the free energy is minimized. During the initial phase of the search, the data are relatively scattered and $f_{\max}-f_{\min}$ is relatively large. The information induction places more emphasis on obtaining the shape of the performance relation rather than finding the optimum. As the data accumulate with more new experiments, the result of regional optimal search will concentrate toward global optima, $f_{\max}-f_{\min}$ would decrease. Emphasis should be placed less on categorizing information and more on optimization. By finding out the desired class and its center, we can determine the optimal operating points.

C. Feedback testing

Based on the class center from the information induction, the final step, feedback control, can suggest changes in the input parameters and see if the output performance is improved. The process we are controlling may drift rapidly due to changes in the process. When this occurs, the optimal operating point specified by the information induction may not produce the best desired output. At this point, it may be

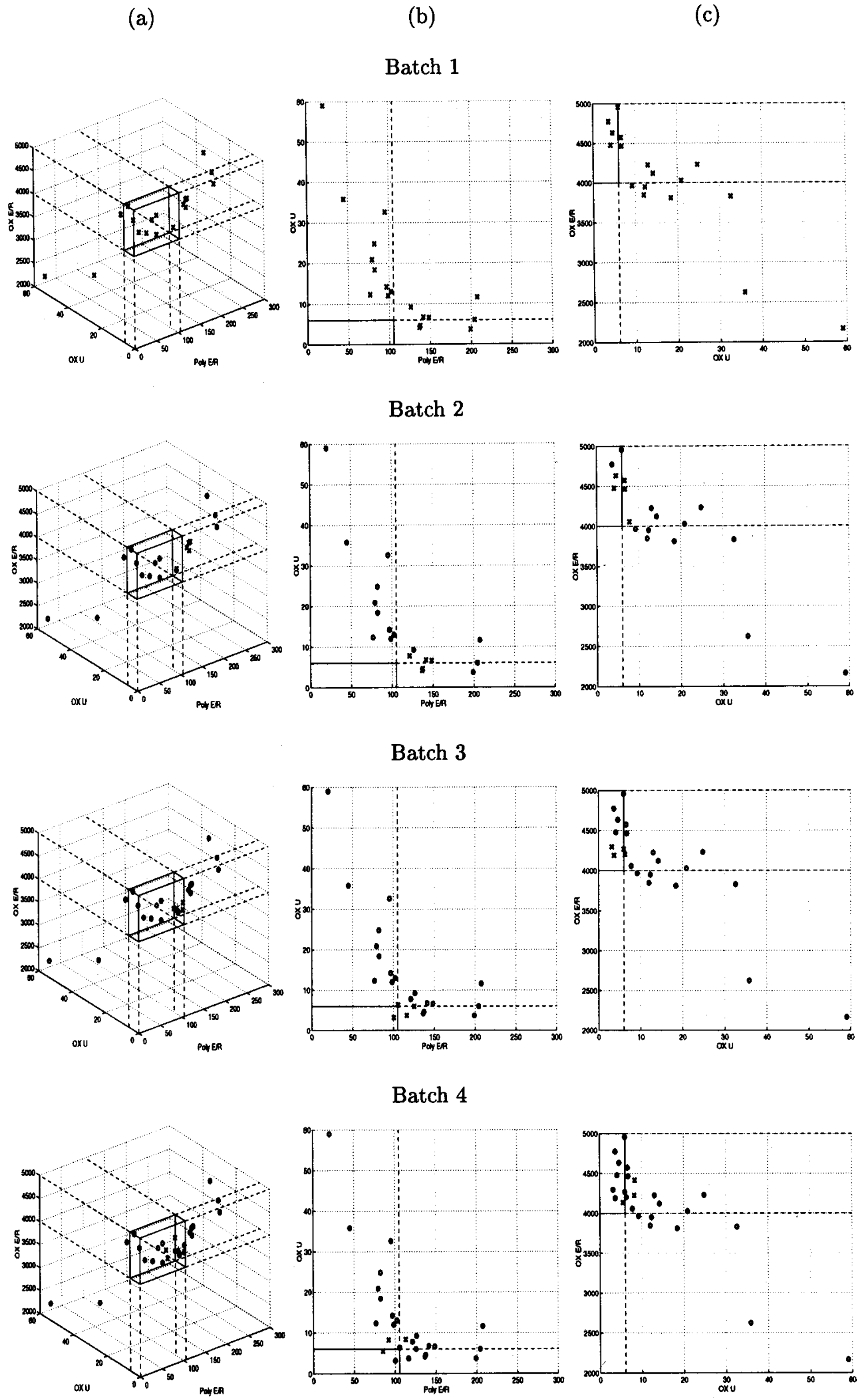


FIG. 5. Optimal etching process design at different batches. (a) Target deviation plots for the current (*x*) and the past (*o*) experiments. (b) Top view of (a) from OX U and Poly E/R directions. (c) Side view of (a) from OX E/R and OX U directions.

necessary to conduct new experiments to regenerate the response surface and classification information induction. The iterative procedures keep running until the desired performance is found. Please note that all experimental points are not wasted in each run because they provide information which is incorporated into the neural network model for the next system analysis.

The advantages of the integrated neural network and information induction in the experimental design are resummarized here. The neural network is used to build a model based on the plasma variables directly related to etching characteristics and the information induction analysis is to determine the best possible operating conditions for testing.

IV. EXPERIMENTAL RESULT AND DISCUSSION

The development of an optimal operating condition of field oxide etching consisted of two stages. In the experimental stage, some experiments are performed to collect the relevant measurement parameters which can represent the characteristics of the etching process. The next stage is to analyze the etching data gathered from these trials. Those data are used to train neural network based process models. Information induction design is subsequently employed to determine the possible optimal operating conditions. New experiments are then performed based on these conditions. The second stage should be kept running to update the neural network model and analyze the system until finding the desired performance.

According to the operator's experience, a total of 14 trials from the past historical data are initially provided. For more effective results, the etching characterization experiments generating experimental runs are statistically designed as a starting point. The performance of the first run is shown in the three-dimensional space (OX U, OX E/R, and Poly E/R) in column (a) of Fig. 5. The cubic box is the design region. The top view and side view plots are represented in columns (b) and (c) in Fig. 5, respectively. We can see there is some deviation from the desired target in the first batch of the historical data. It seems that the suggested experiments yield rather unsatisfactory results. Therefore, the proposed method will demonstrate how to improve the result and get the desired recipe.

A three-layer network is trained by the data obtained in the first run. For this application, the neurons of the input layer correspond to the five adjustable input parameters. The neurons of the output layer represent the three etching responses. Networks with one and two hidden nodes are initially tested, respectively. They are not adequate as the convergence is too slow and rms error remains high. Finally, a network with three hidden nodes solves the problem. It requires a total of 150 epochs to reduce the rms error to an acceptable level. The information induction analysis is used to analyze optimal regions based on the response surface from the neural network model. The change in U, TS, and F with the number of clusters for the first batch is shown in Fig. 6. The information energy shows four clusters that represent four major optima should be performed if optimization is our concern. However, information entropy suggests five experiments because the extra one helps us to mold the performance surface with more accuracy. Therefore, five experiments should be done in the next batch.

In Fig. 5, it is obvious that most of the new experimental points are getting close to the target although they still have not yet fallen into the desired region where the circles are old experimental data and the cross points represent the results of the suggested new experiments. Repeating the same previous procedures, the suggested experiments and previous experiments done in the first batch are used to retrain a new neural network model. After information induction in the second batch (Fig. 6), four experiments are suggested to be

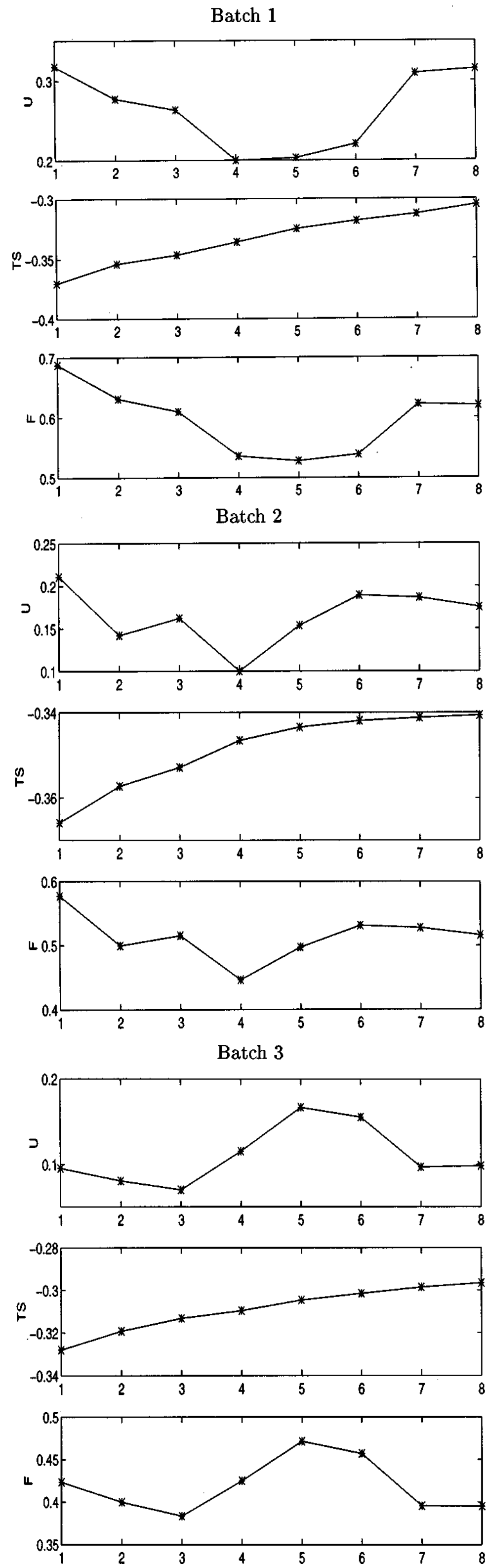
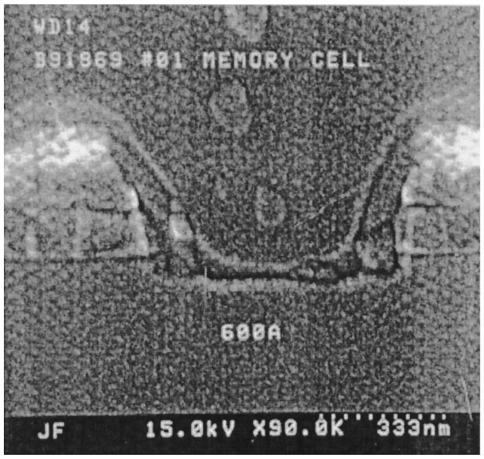


FIG. 6. Information analysis plots at different batches.

(a)



(b)

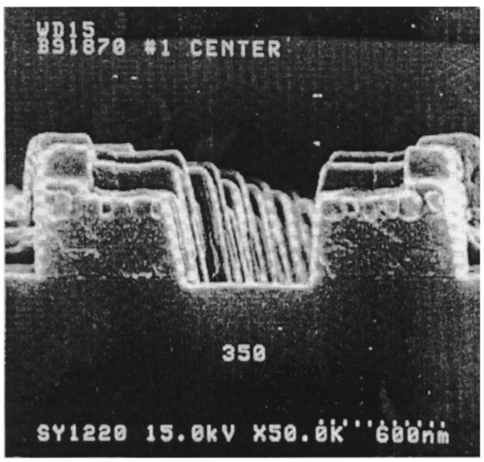


FIG. 7. Cross-section SEM of SAS etching. (a) High pressure recipe without CO gas at Si loss >600 Å. (b) Recipe with CO gas at Si loss ~300 Å.

performed. In this new run, data are getting even closer to the target. The recipe that meets both criteria is found (Fig. 5). We keep another trial to find any chance for improvement. Although the performance of the experimental data this time is still within the targeted area, there is no significant improvement. That means the procedures of experimental design are finished. When making a comparison among the information induction plots (Fig. 6), we can see that the information energy is high for all classes in the first batch. After several runs, both terms, TS and U, are decreased to the smaller values for all classes. It indicates that the neural network model is getting close to the system little by little with the new experiments added and the optimal region is getting explicit. Furthermore, the initial batches which do not reach the required specifications are not wasted, because they provide information of the response surface that can be incorporated into the neural network model.

Within four generations and a total of 26 experimental points, a recipe that meets all specifications is found. However, the traditional two-level ($2^5=32$) factorial design requires at least 32 experiments to deal with these five input factors. Since it estimates the main effects and interactions, it can only bring the path or direction for finding the optimal performance. Being modeled as a linear plan, it cannot estimate a curvature in the response surface. If the three-level factorial design (3^5) is used to estimate the degree of curvature in the response, a full three-level five-factor factorial design requires 243 experiments. Suppose a partial factorial design[24,25] is run using the central composite design. Forty-three data points would be required to perform an initial search. This shows that the proposed method is the most effective and efficient way for the etching process.

After the simulation is completed, topography wafers have been through with the optimized process conditions and cross-section SEM analysis has been conducted. The results confirm that the CO recipe could meet the predefined process criteria. In Fig. 7(a) the wafer is processed by a high pressure regime recipe whose Si loss is still beyond satisfaction, and Fig. 7(b) is the newly developed CO recipe which exhibits the minimum Si loss of ~300 Å.

V. CONCLUSION

With the increasing use of neural networks in semiconductor process modeling, a new application is developed by the authors. This novel technique integrates two elements of artificial intelligence research for process experimental design. The neural network is used as a tool to summarize all experimental information into a mathematical model. Information induction analysis is employed to determine how many features are worth testing, i.e., finding possible optimal operating conditions. The information induction is derived from a fuzzy classification technique and information theory. The etching experimental design in this study demonstrates that run-to-run experiments can lead the process into the desired etching characteristics within specified ranges. Besides, the proposed method in the semiconductor fabrication can reduce time, cost and risk of the product and process development.

ACKNOWLEDGMENTS

This work is supported by the National Scientific Council of R.O.C. under Grant No. NSC85-2622-E007-008 and is sponsored by Taiwan Semiconductor Manufacturing Company Ltd.

[1] J. P. Card, D. L. Sniderman, and C. Klimasauskas, IEEE Trans. Neural Netw. **8**, 883 (1997).
[2] E. A. Rietman, IEEE Trans. Semicond. Manuf. **9** (1995).
[3] G. S. May, *IEEE Spectrum* (1994), pp. 47–51.
[4] E. A. Rietman and E. R. Lory, IEEE Trans. Semicond. Manuf. **6** (1993).
[5] C. D. Himmel, B. Kim, and G. S. May, in *Proceedings of the Fourth International Semi. Manufacturing Science Symposium*, 1992, p. 124.
[6] F. Nadi, A. Agogino, and D. Hoges, IEEE Trans. Semicond. Manuf. **4** (1991).
[7] C. D. Himmel and G. S. May, IEEE Trans. Semicond. Manuf. **6**, 103 (1993).
[8] D. M. Baker, C. D. Himmel, and G. S. May, IEEE Trans. Semicond. Manuf. **8**, 62 (1995).
[9] Y. Ohshima, S. Mori, Y. Kaneko, E. Sakagmi, N. Arai, N. Hosokwa, and K. Yoshikawa, in *International Electron Devices Meeting 1990*, Technical Digest, p. 95.

[10]P. S. McAnally, W. L. Krisa, L. M. Ting, and G. A. Dixit, in 1995 International Symposium on VLSI Technology, Systems and Applications; Proceedings of Technical Papers, p. 180.
[11]J. Chen, D. S. H. Wong, S.-S. Jang, and S.-L. Yang, AIChE. J. **44**, 876 (1998).
[12]K. Hornik, M. Stinchcombe, and H. White, Neural Networks **2**, 359 (1989).
[13]K. Hornik, M. Stinchcombe, and H. White, Neural Networks **3**, 551 (1990).
[14]B. L. Kalman and S. C. Kwasny, in *International Joint Conference on Neural Networks* 1992, Baltimore, MD.
[15]J. Gorodkin, L. K. Hansen, A. Krogh, C. Savrer, and O. Winther, Int. J. Neural Syst. **4**, 159 (1993).
[16]J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, New York, 1991).
[17]D. M. Allen, Technometrics **16**, 125 (1974).
[18]D. E. Goldberge, *Genetic Algorithm in Search, Optimization and Machine Language* (Addison-Wesley, Reading, MA, 1989).
[19]G. Dahlquist, and A. Bjorck, *Numerical Method* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
[20]J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Plenum, New York, 1981).
[21]J. C. Bezdek, R. Ehrlich, and W. Full, Comput. Geosci. **10**, 191 (1984).
[22]C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[23]C. E. Shannon and W. Weave, *The Mathematical Theory of Communication* (University of Illinois Press, Champaign, IL, 1949).
[24]R. H. Myers and D. C. Montgomery, *Response Surface Methodology, Process and Product Optimization using Designed Experiments* (Wiley, New York, 1995).
[25]B. Kim and G. S. May, IEEE Trans. Semicond. Manuf. **7**, 12 (1994).