Prediction of plasma etching using genetic-algorithm controlled backpropagation neural network

Sungmo Kim , and Byungwhan Kim *

* Department of Electronic Engineering, Sejong University, 98 Kunja-Dong. Kwangjin-Ku, Seoul, 143-747, Korea
(Tel : 82-2-3408-3729; Fax : 82-2-3408-3329 ; E-mail: kbwhan@sejong.ac.kr)

Abstract: A new technique is presented to construct a predictive model of plasma etch process. This was accomplished by combining a backpropagation neural network (BPNN) and a genetic algorithm (GA). The predictive model constructed in this way is referred to as a GA-BPNN. The GA played a role of controlling training factors simultaneously. The training factors to be optimized are the hidden neuron, training tolerance, initial weight magnitude, and two gradients of bipolar sigmoid and linear functions. Each etch response was optimized separately. The proposed scheme was evaluated with a set of experimental plasma etch data. The etch process was characterized by a 2^3 full factorial experiment. The etch responses modeled are aluminum (Al) etch rate, silica profile angle, Al selectivity, and dc bias. Additional test data were prepared to evaluate model appropriateness. The GA-BPNN was compared to a conventional BPNN. Compared to the BPNN, the GA-BPNN demonstrated an improvement of more than 20% for all etch responses. The improvement was significant in the case of Al etch rate.

Keywords: Genetic algorithm, backpropgation neural network, optimization, training factor

1. INTRODUCTION

Plasma etching is a key means to fine patterning of thin films in manufacturing integrated circuits. Predictive etch models are highly demanded not only to empirically characterize plasma processes, but to identify useful trade-offs between process responses for process optimization. Once constructed, predictive models can be effectively used to explore process parameter effects on plasma processes without conducting additional experiments. Neural networks have been promisingly used to build predictive models of various plasma processes [1-3]. Among neural networks, the backpropagation neural network (BPNN) [4] has been the most frequently applied to plasma modeling.

In most applications, predictive models are constructed by controlling only the hidden neuron variable. Apart from this factor, there exist other training factors, including the training tolerance, initial weight magnitude, or gradients of activation functions. It has been observed that each factor affects the predictive performance considerably. In the context of plasma modeling, the training factor effects have been optimized individually [1-2]. Possible interactions among them could thus not be taken into account. This has partly been circumvented by a model-based optimization technique [3]. This approach is limited in that it accompanies two types of errors. The first type of error arises from the use of a model that relates factor effects on model performance. The other is the error accompanied during the optimization. By removing the first type of error, the BPNN predictive capability is expected to improve.

For this, in this study, a generic algorithm (GA) [5] is used to optimize training factor effects simultaneously. Compared to previous works, this work is distinct in that the GA is applied directly to the training factors, not to the training factor effect model. The proposed technique was applied to experimental plasma etch data. The etching of silica films was conducted with inductively coupled plasma etch system at ETRI. To systematically characterize the etching, the experiments were conducted by means of a statistical experimental design. Additional experiments were conducted to prepare the test data. The process parameters that were varied in the experiment are radio frequency source power, bias power, and gas ratio. The etch responses modeled include aluminum (Al) etch rate, silica profile angle, Al selectivity, and dc bias.

2. EXPERIMENTAL DATA

The schematic diagram of the ICP etch system is depicted in Fig. 1. Detailed procedures to fabricate test patterns are explained in the previous work [2]. The etch process was characterized by a 2^3 full factorial experiment [6] with one center point. Resulting nine experiments are used to train the BPNN. Additional six experiments were conducted to prepare test data for model evaluation. The process parameters that were varied in the design include the radio frequency (rf) source power, bias power, and gas ratio. The total flow rate of gases, CHF_3 and CF_4, was set to 60 sccm and the flow rate of CHF_3 was varied from 10 sccm to 50 sccm. Their experimental ranges are shown in Table I.

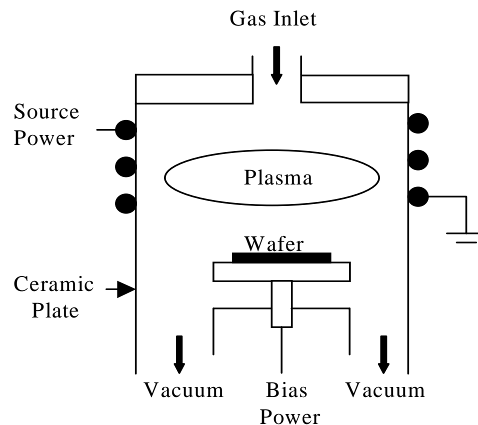

Fig. 1 Schematic of a plasma etch system

Table I: Experimental parameters and ranges

Parameter	Range	Units
Source Power	100-800	Watts
Bias Power	100-400	Watts
Gas Ratio	0.2-5.0	

The gas flow rate ratio in Table I is defined as the flow rate of CHF_3 divided by the flow rate of CF_4. In consequence, a total of 15 experiments were conducted. The etch responses modeled include aluminum (Al) etch rate, selectivity, profile angle of silica film, and DC bias. The etch rates and profile angle were estimated by using a scanning electron microscopy (SEM). The selectivity is defined as the ratio of silica etch rate to Al one. The dc bias was measured by reading a dc voltmeter embedded in the match network.

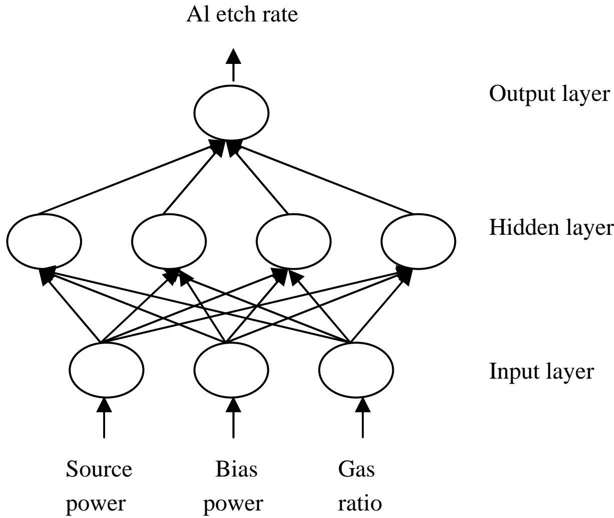

Fig. 2 Architecture of a backpropagation neural network.

3. BACKPROPAGATION NEURAL NETWORK

A typical architecture of BPNN is depicted in Fig. 2. The BPNN consists of one or more layers of neurons: input layer, hidden layer, and output layer. The input layer receives external information such as process parameters. From the output layer, predictions to a given input pattern are obtained.

The BPNN also incorporates "hidden" layers of neurons that do not interact with the outside world, but assists in performing nonlinear feature extraction on the data provided by the input and output layers. The number of hidden layer was set to one in this application. Activation level (or firing strength) of a neuron is determined by a bipolar sigmoid function denoted as:

$$out_{i,k} = \frac{1 - e^{-int_{i,k}}}{1 + e^{-int_{i,k}}} \qquad (1)$$

where $int_{i,k}$ and $out_{i,k}$ indicate the weighted input to the i th neuron in the k th layer and output from that neuron, respectively. The error (E) the network attempts to minimize is defined as:

$$E = 0.5 \sum_{i=1}^{p} (d_i - out_i)^2 \qquad (2)$$

where p is the number of output neurons, d_i is the desired output of the i th neuron in the layer, and out_i is the calculated output from the same neuron. In the BP algorithm, this error is to be minimized via the gradient descent optimization, in which the weights are adjusted in the direction of decreasing the E in (2). A basic weight update scheme, commonly known as the generalized delta rule [4], is expressed as:

$$W_{i,j,k}(m+1) = W_{i,j,k}(m) + \eta \Delta W_{i,j,k}(m) \qquad (3)$$

where $W_{i,j,k}$ is the connection strength between the j th neuron in the layer (k-1) and the i th neuron in the layer k. Other m and η indicate the iteration number and an adjustable parameter called "learning rate," respectively. The $\Delta W_{i,j,k}$ in (3) is the calculated change in the weight to minimize the E in (2) and defined as:

$$\Delta W_{i,j,k} = \frac{\partial E}{\partial W_{i,j,k}} \qquad (4)$$

By adjusting weighted connections recursively using the rule in (3) for all the units in the network, the accumulated E over all the input vectors is to be minimized.

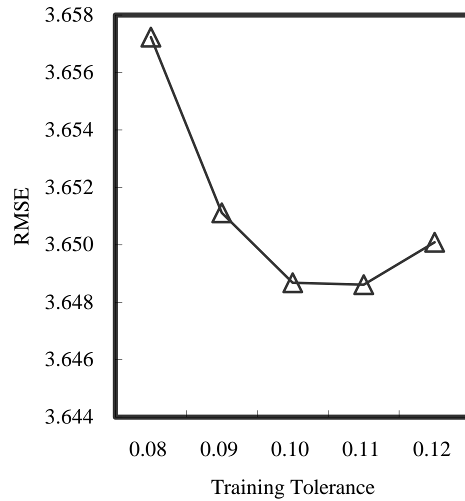

Fig. 3 Prediction accuracy as a function of Training tolerance.

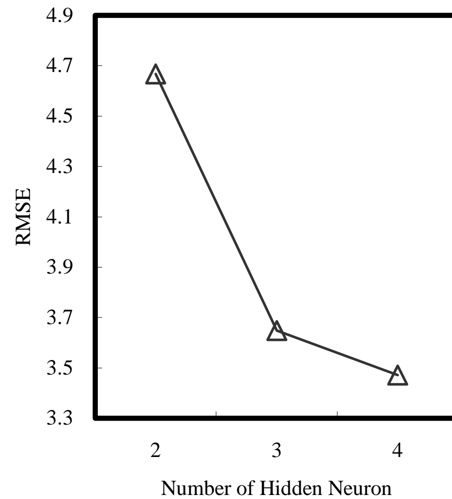

Fig. 4 Prediction accuracy as a function of number of hidden neuron.

4. RESULT

4.1 BPNN Model

For comparison, predictive models are constructed in conventional way. As an illustration, the profile of silica is exampled. For this, the training factors are optimized in two stages. In the first stage, 3 training factors but the gradients are optimized. First, the training tolerance (TT) varied from 0.08 to 0.12 by 0.01 and its effects on the prediction accuracy are depicted in Fig. 3. The prediction accuracy was measured by

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{r}(d_j - out_j)^2}{r}} \quad (5)$$

where the r indicates the total number of test vectors. In this study, the r is equivalent to 6. As depicted in Fig. 3, the RMSE varies inconsistently with the TT. The smallest RMSE is obtained at 0.11, which is equal to about 3.648 With the TT set at 0.11, the number of hidden neuron (NHN) then varied from 2 to 4. Results are shown in Fig. 4. The RMSE consistently decreases with the NHN and is minimized at 4. The corresponding RMSE is 3.472. Next, the magnitude of initial weight distribution (MIWD) varied from ± 0.2 to ± 1.6 by 0.2. Results are shown in Fig. 5. As displayed in Fig. 5, the RMSE variation with the MIWD is quite complex. At ± 0.4, the smallest RMSE of 3.436 is obtained. Compared to the preceding model determined, the RMSE varied little. This indicates that the MIWD played a little role in improving the RMSE.

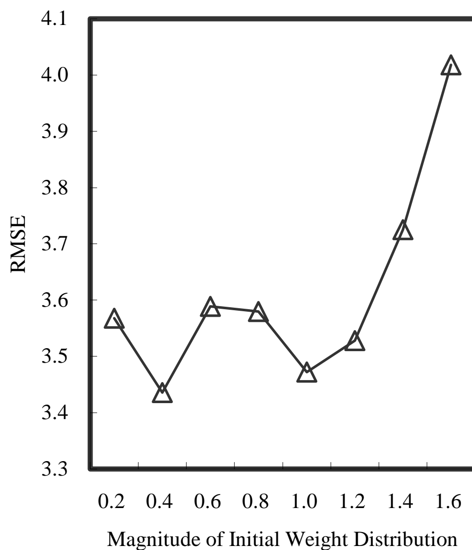

Fig. 5 Prediction accuracy as a function of magnitude of initial weight distribution.

In the second stage, the gradient effects are optimized. For this, the gradient of the bipolar sigmoid function (GBSF) varied from 0.4 to 2.0 with an increment of 0.2. The other gradient of the linear function (GLF) varied in the same range. As a result, a total of 81 combinations of the gradients were examined. Results are shown in Fig. 6. As exhibited in Fig. 6, the smallest RMSE is obtained at the 19th combination consisted of 0.8 GBSF and 0.4 GLF. The corresponding RMSE is 2.858. Compared to the optimized model in the first stage, this demonstrates an improvement of about 16.8%. This indicates that the gradient combination is an important training factor in improving the RMSE. In this way, the other 3 etch responses were modeled and results are contain in Table II. The optimized training factors and corresponding RMSEs are included.

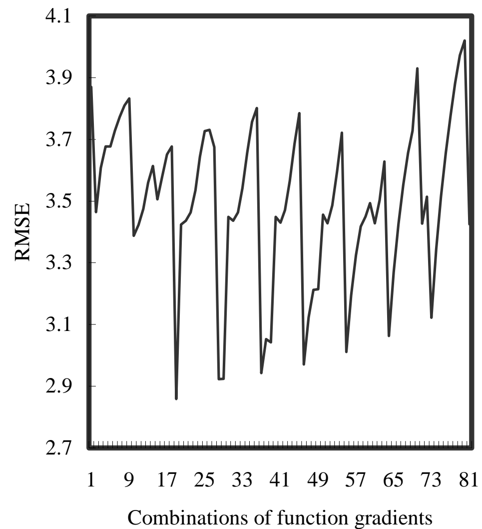

Fig. 6 Prediction accuracy as a function of gradient combinations.

Table II : Performance of BPNN with training factors

	TT	NHN	MIWD	GBSF	GLF	Optimized RMSE
Profile	0.11	4	0.4	0.8	0.4	2.858
Selectivity	0.12	2	0.2	1.2	1.2	2.258
DC Bias	0.08	3	1.4	0.4	0.4	53.57
Etch Rate	0.12	4	0.8	0.6	0.4	434.35

4.2 GA-optimized BPNN

Genetic algorithm was utilized to search for a particular factor setting that minimizes the prediction accuracy. In GA optimization, each training factor was coded in a real value and this resulted in a total chromosome length of 5 bits. During each computational cycle, an initial population of 100 potential solutions was created with each manipulated by the genetic operators. Next, the performance of each individual of the population is evaluated and a selection mechanism is subsequently activated to choose the best string with the highest fitness for the genetic manipulation process. The *crossover* operator takes two chromosomes and parts of their genetic information are swapped to produce two new chromosomes based on a specified crossover probability. Another mutation probability is given to the mutation operator, which randomly changes a fixed number of bits every generation. Here, those numerical probabilities of crossover and mutation used in this optimization are 0.9 and 0.1, respectively.

A particular input setting generated by GA meets a given fitness function expressed as:

$$F = \frac{1}{1 + \sum_r \mathrm{RMSE}_r} \quad (6)$$

where, r is the number of etch response to optimize. Since each etch attribute was optimized individually, the r is equal to unity. Training factors optimized by means of GA are contained in Table III along with corresponding RMSEs. The optimized models in Table II and III are compared and the results are represented in the bar graph as shown in Fig. 7. The percent improvements of GA-BPNN calculated over the conventional BPNN are indicated on the top of each bar graph. More than 20% of improvements are achieved for all etch responses. The improvement was the most drastic in the case of Al etch rate. These improvements reveal that the GA-BPNN is an effective means to improve the BPNN predictive ability.

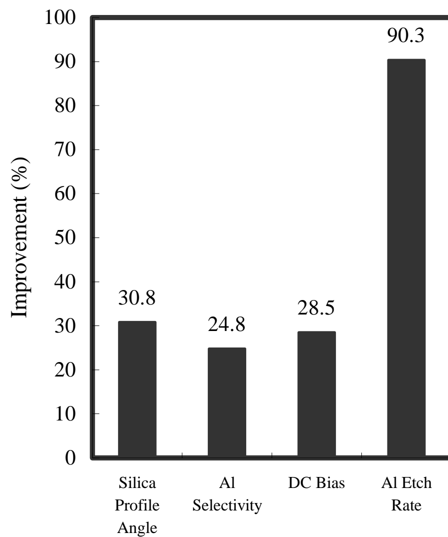

Fig. 7 The percent improvements of GA-BPNN over BPNN.

Table III : Performance of GA-BPNN

	TT	NHN	MIWD	GBSF	GLF	Optimized RMSE
Profile	0.1	7	2.11	1.4	1.27	1.972
Selectivity	0.1	4	1.79	0.48	0.52	1.694
DC Bias	1.34	3	3.15	1.69	0.83	38.216
Etch Rate	3.2	7	1.97	3.26	1.91	41.99

5. CONCLUSIONS

In this study, a means to optimize training factor effects on BPNN model was presented. This was accomplished by applying a GA directly to the training factors. The experimental data were collected by a statistical experimental design, and four etch responses were modeled. Compared to conventional model, more than 20% improvements were demonstrated for all etch responses. A drastic improvement of about 90% was achieved for the Al etch rate. The comparison results clearly indicate that the by controlling training factors using the GA the BPNN predictive ability could considerably be improved.

ACKNOWLEDGMENTS

This work was supported by Korea Research Foundation Grant (KRF-2002-041-D00117)

REFERENCES

[1] B. Kim, S. M. Kong, and B.-T. Lee, "Modeling SiC etching in C_2F_6/O_2 inductively coupled plasma using neural networks, *J. Vac. Sci. Technol. A*, vol. 20, no. 1, pp. 146-152, 2002.

[2] B. Kim, K. H. Kwon, S.-H. Park, "Characterizing metal-masked silica etch process in a CHF_3/CF_4 inductively coupled plasma," *J. Vac. Sci. Technol. A*, vol. 17, no. 5, pp. 2593-2597, 1999.

[3] B. Kim and S. Park, "An optimal neural network plasma model: a case study", *Chemom. Intell. Lab. Syst.*, vol. 56, pp. 39-50, 2001.

[4] D. E. Rummelhart and J. L. McClelland, *Parallel Distributed Processing*, Cambridge, M.I.T. Press, 1986.

[5] D. E. Goldberg, *Genetic Algorithms in Search, Optimization & Machine Learning*, Addison Wesley, 1989.

[6] D. C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, Singapore, 1991.