2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

Random generator-controlled backpropagation neural network to predicting plasma process data

Sungmo Kim, Sebum Kim, and Byungwhan Kim
Department of Electronic Engineering, Sejong University
98, Kunja-Dong, Kwangjin-Ku, 143-747, Seoul, Korea
email: kbwhan@sejong.ac.kr

***Abstract* - A new technique is presented to construct predictive models of plasma etch processes. This was accomplished by combining a backpropagation neural network (BPNN) and a random generator (RG). The RG played a critical role to control neuron gradients in the hidden layer. The predictive model constructed in this way is referred to as a randomized BPNN (RG-BPNN). The proposed scheme was evaluated with a set of experimental plasma etch process data. The etch process was characterized by a 2^3 full factorial experiment. The etch responses modeled are 4, including aluminum (Al) etch rate, profile angle, Al selectivity, and dc bias. Additional test data were prepared to evaluate model appropriateness. The performance of RG-BPNN was evaluated as a function of the number of hidden neurons and the range of gradient. For given range and hidden neurons, 100 sets of random neuron gradients were generated and among them one best set was selected for evaluation. Compared to the conventional BPNN, the proposed RG-BPNN demonstrated about 50% improvements in all comparisons. This illustrates that the RG-BPNN of multi-valued gradients is an effective way to considerably improve the predictive ability of current BPNN of single-valued gradient.**

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 response variables for process optimization. Once constructed, predictive models can be effectively used to explore process parameter effects on etching or deposition under various plasma conditions 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) has been the most frequently applied compared to other paradigms. This is mainly attributed to its high predictive ability as demonstrated by the improved predictions over statistical regression models [4-5]. In most applications, predictive models are constructed by controlling the number of hidden neurons. It has been reported that the gradient of neuron activation function affects the BPNN predictive ability considerably [1, 2, 3, 5]. The optimized gradient is typically optimized by adjusting it within certain experimental range. Meanwhile, all neurons have the same optimized gradient. It is expected that by choosing multi-valued gradients the predictive ability can be improved. Despite this expectation, there have been little studies on this concern.

In this study, a new empirical technique to overcome the current limit stated earlier is presented. This is accomplished by using a random generator (RG). The RG plays a role of generating many sets of random neuron gradients. For convenience, the RG-based BPNN is called RG-BPNN. The performance of RG-BPNN is evaluated as a function of the number of hidden neurons and the experimental range of random gradients. The performance is also compared to conventional BPNN. The data examined were collected from the etching of silica thin film, and the etch responses modeled are aluminum (Al) etch rate, profile angle, Al selectivity, and dc bias. Additional test data were prepared to evaluate model

appropriateness. The etching was conducted in an inductively coupled (ICP) plasma etch system.

2. Experimental Details

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 [3]. 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.

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 the dc voltmeter embedded in the match network.

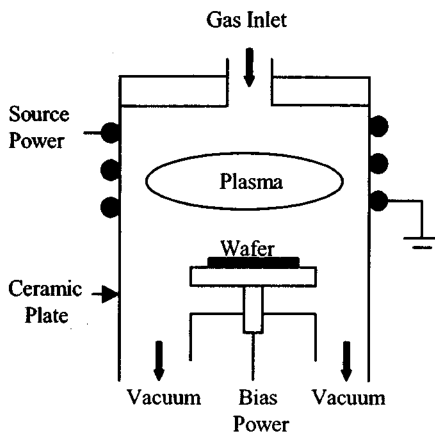

Fig .1 Schematic of a plasma etch system

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	

3. Backpropagation neural network

The BPNN consists of three layers of neurons: input layer, hidden layer, and output layer. The input layer receives external information such as adjustable process parameters contained in Table I. The output layer transmits the data and thus corresponds to the various process responses. In this study, the number of neurons in the output layer was set to unity since only etch responses were modeled individually. 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. Here, the number of the hidden layer was set to unity. The activation level (or firing strength) of a neuron in the hidden layer is determined by a bipolar sigmoid function denoted as

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

where $in_{i,k}$ and $out_{i,k}$ indicate the weighted input to the ith neuron in the kth layer and output from that neuron, respectively. The g_b represents the gradient of the bipolar sigmoid function and determines the activation level of neuron. Meanwhile, the BPNN adopts a linear function in the output layer, expressed as

$$out_{i,k} = in_{i,k} \cdot g_l \qquad (2)$$

where g_l represents the gradient of the linear function. For a given set of training factors, both gradients in (1) and (2) are to be optimized to improve the prediction accuracy. The BP algorithm by which the network is trained begins with a random set of weights (i.e. connection strengths between neurons). The Euclidean distance in the weight space the network attempts to minimize is the accumulated error (E) of all the input-output pairs, which is expressed as

$$E = \sum_{j=1}^{q} (d_j - out_j)^2 \qquad (3)$$

where q is the number of output neurons, d_j is the desired output of the jth neuron in the output layer, and out_j is the calculated output of that 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 (3). A basic weight update scheme, commonly known as the *generalized delta rule*, is expressed as

$$W_{i,j,k}(m+1) = W_{i,j,k}(m) + \eta \Delta W_{i,j,k}(m) \qquad (4)$$

where $W_{i,j,k}$ is the connection strength between the jth neuron in the layer (k-1) and the ith neuron in the layer k, and $\Delta W_{i,j,k}$ is the calculated change in the weight to minimize the E in (3) and defined as

$$\Delta W_{i,j,k} = -\frac{\partial E}{\partial W_{i,j,k}} \qquad (5)$$

Other parameters m and η indicate the iteration number and an adjustable parameter called "learning rate," respectively. The η was set to 0.01 in this study. By adjusting the weighted connections recursively using the rule (4) for all the units in the network, the accumulated E for all training vectors is minimized.

Table II: Prediction performance of BPNN as a function of the number of hidden neurons and gradient.

Gradient	Number of Hidden Neurons				
	2	3	4	5	6
0.2	697.9	529.1	610.5	450.0	690.5
0.4	627.8	**395.7**	528.7	500.1	521.1
0.6	611.5	694.9	501.9	535.6	532.6
0.8	611.2	681.4	516.3	577.0	597.5
1.0	613.5	718.0	524.2	607.3	598.0
1.2	614.3	641.3	529.1	621.0	595.2
1.4	615.9	635.0	532.2	626.9	594.9
1.6	617.2	627.0	533.8	629.3	596.0
1.8	617.5	617.1	534.0	627.5	596.7
2.0	617.4	608.4	533.3	608.9	598.3

4. Results

4.1. BPNN of single valued-neuron gradient

First, the performance of conventional BPNN is evaluated. As stated earlier, the performance is investigated as a function of neuron numbers and gradient. The number varied from 2 to 6 by 1. The gradient varied within the range 0.2-2.0 with an increment of 0.2. It should be noted that given the gradient all hidden neurons are equipped with the same one. Also, another gradient (g_l) was set to unity. The other training factors were set to their default values, 0.10 training tolerance. ± 1.0 initial weight distribution. As an illustration, the Al etch rate was modeled and results are shown in Table II. The predictive accuracy was measured by the root-mean squared error (RMSE). As contained in Table II, the best model is obtained at 3 hidden neurons and 0.4 gradient. The corresponding RMSE is 395.7 Å/min. In this way, the other etch responses were evaluated and their performances are summarized in Table III

Table III: Performance of optimized BPNN with optimized gradient and hidden neurons

Etch Responses	Hidden Neurons	Gradient	RMSE
Al Etch Rate (Å/min)	3	0.4	395.7
Profile Angle (degree)	6	1.0	2.45
Selectivity	6	0.8	2.28
DC Bias (V)	5	0.4	59.3

Table IV: Prediction performance of BPNN as a function of the number of hidden neurons and range of random gradient.

Gradient	Number of Hidden Neurons				
	2	3	4	5	6
± 0.2	131.7	107.4	93.5	63.3	61.8
± 0.4	156.9	207.8	120.0	154.2	120.3
± 0.6	144.9	147.3	162.0	120.6	142.4
± 0.8	131.8	132.1	129.0	104.0	141.2
± 1.0	131.8	77.1	164.7	112.9	150.8
± 1.2	131.5	135.2	164.6	134.0	116.8
± 1.4	131.5	131.8	156.4	155.7	148.6
± 1.6	132.1	**53.7**	141.3	127.9	183.4
± 1.8	132.2	146.0	78.9	133.1	197.4
± 2.0	131.5	61.2	122.0	170.1	139.5

4.2. BPNN of multi-valued neuron gradient

Using the RG, 100 predictive BPNN models were generated for a given neuron number and random gradient. Among them, only one model of the smallest RMSE was selected. The corresponding RMSEs for the Al etch rate models determined are contained in Table IV. As seen in Table IV, one optimal model is obtained at 3 hidden neurons and in the range [0, ± 1.6]. The optimized gradients are 0.1432, 0.3336, and 0.8507. The corresponding RMSE is Å/min. Compared to conventional BPNN, the RG-BPNN demonstrates a significant improvement of 86.4%. In the same way, the other etch responses were evaluated and results are contained in Table V. The improvements over the conventional BPNN are contained in the last column of

Table V. As seen in Table V, the improvements are more than about 50% in all comparisons. This clearly indicates the advantage of multi-valued gradients.

Table V: Performance of optimized BPNN with optimized range of random gradient and hidden neurons

Etch Responses	Hidden Neurons	Range of Gradient	RMSE	Improvement (%)
Al Etch Rate (Å/min)	3	±1.6	53.7	86.4
Profile Angle (degree)	4	±0.4	1.15	53.0
Selectivity	3	±0.2	1.02	55.2
DC Bias (V)	4	±2.0	25.3	57.3

5. Conclusions

In this study, we presented a means to optimize gradients of BPNN hidden neurons. The RG played a role of generating multiple sets of neuron gradients. The performance of optimized etch models were compared to those obtained in conventional way. More than 50% improvements were demonstrated for all etch responses. A drastic improvement of about 86% was achieved for the Al etch rate. Due to high computational burden, the complexity of random initial weight was not considered. This can empirically be taken into account by generating multiple models for each of generated models for a given range of random gradient. The comparison results clearly indicate that the RG-controlled neuron gradient is an effective way to considerably improve the BPNN predictive ability.

Acknowledgements

This work was supported by Korea Health Industry Development Institute from 2000 IMT Fund.

REFFERENCES

[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. Vacuum Science and Technology A, vol. 20, no. 1, pp.146-152, 2002

[2] B. Kim and S. Park, "Characterization of inductively coupled plasma using neural network," IEEE Trans. Plasma Sci., vol. 30 no. 2, pp. 698-705, 2002.

[3] B. Kim and K. Kwon, Qualitative modeling of silica plasma etching using neural network, J. Appl. Phys., vol. 93, no. 1, pp. 76-82, 2003.

[4] B. Kim and G. Park, "Modeling plasma equipment using neural networks", IEEE Plasma Sci., vol. 29, no. 2, pp. 8-12. 2001.

[5] B.Kim and S. Park, "An optimal neural network plasma model: a case study," Chemometrics and Intelligent Laboratory System, vol..56 (1), pp. 39-50, 2001.

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