Diagnosis of Plasma Equipment using Neural Network and Impedance Match Monitoring

Byungwhan Kim

Abstract - A new methodology is presented to diagnose faults in equipment plasma. This is accomplished by using neural networks as a pattern recognizer of radio frequency (rf) impedance match data. Using a match monitor system, the match data were collected. The monitor system consisted mainly of a multifunction board and a signal flow diagram coded by Visual Designer. Plasma anomaly was effectively represented by electrical match positions. Twenty sets of fault-symptom patterns were experimentally simulated with variations in process factors, which include rf source power, pressure, Ar, and O_2 flow rates. As an input to neural networks, two means and standard deviations of positions were used as well as a reflected power. Diagnostic accuracy was measured as a function of training factors, which include the number of hidden neurons, the magnitude of initial weights, and two gradients of neuron activation functions. The accuracy was the most sensitive to the number of hidden neurons. Interaction effects on the accuracy were also examined by performing a 2^4 full factorial experiment. The experiments were performed on multipole inductively coupled plasma equipment.

1. Introduction

Processing plasmas play a crucial role in both depositing thin films and etching fine patterns. Any variability in process factors (such as radio frequency (rf) power or pressure) can cause a significant shift in plasma state. When this shift becomes large enough to change the operating conditions of the equipment beyond an acceptable level, overall product quality can be greatly jeopardized. Thus, timely and accurate diagnosis of plasma malfunction is crucial to manufacturing integrated circuits.

Algorithmic systems such as HIPPOCRATES [1] have been developed to identify process faults from statistical inference procedures and electrical measurements performed on IC wafers. Expert systems such as PIES [2] have also been designed to draw upon experiential knowledge to develop qualitative models of process behaviors. By combining the advantages of each paradigm, a hybrid diagnostic system was developed [3]. During on-line diagnosis in this system, plasma faults were diagnosed by utilizing external sensors that monitor process inputs such as power, pressure, and gas flows. Monitoring internal plasma variables, rather than power, pressure, and gas flows, is increasingly demanded due to their increased sensitivity and numerous diagnostic clues that can be extracted. Numerous in-situ sensors have been utilized for plasma diagnostics and control, which include a Langmuir probe, optical emission spectroscopy (OES), and rf sensor. An OES-based plasma monitor system [4] has been presented but lacks the ability to pinpoint root causes for abnormal plasma. Neural networks were used to detect and recognize optical emission spectra of various gas species [5] as well as to classify gas-phase infrared spectra [6]. These networks seemed promising for estimating chemical compositions of sediments analyzed with reflectance spectroscopy [7], calibrating near-infrared spectra of ampicillin trihydrate samples [8], recognizing mass spectra [9], and interpreting raman spectra of nitro-containing explosive materials [10]. In the context of diagnosis, however, neural networks have rarely been applied to in-situ plasma diagnostic data.

A new methodology is presented to diagnose a process plasma. Rather than in-situ spectroscopy-measured spectra, neural networks are used to recognize a process of plasma impedance matching. Match fault data were collected using the system designed to monitor rf match network [11]. The system consisted mainly of a multifunction board and a signal flow diagram coded with Visual Designer [12]. A total of 20 sets of faults and symptoms were obtained with variations in process factors. For this data, a backpropagation neural network (BPNN) [13] was applied to capture causal relationships between the causes and symptoms. Diagnostic accuracy, i.e., recognition rate, is quantified with the root-mean squared error metric. The accuracy is then examined as a function of training factors. Relative significances of training factors are investigated by means of an interaction model.

2. Impedance match data

The parameters involved in rf matching were collected using the monitor system previously reported [11]. Faults in the plasma were simulated by incrementally changing factor levels. The factors that were varied include source

The work was supported by a Korea Research Foundation grant (KRF-2000-003-E00160).

Manuscript received: Aug. 22, 2001 accepted: April 9, 2002

Byungwhan Kim is with Department of Electronic Engineering, Sejong Universit 98, Goonja-Dong, Kwangjin-Gu, Seoul, 143-747, Korea

power, pressure, Ar, and O_2 flow rates. Their experimental ranges and corresponding levels of increments are contained in Table 1. A total of twenty faults, five for each factor, thus simulated. A total of six variables involved in impedance matching were measured simultaneously, including two match positions, related match errors, forward power, and reflected power. Among them, the first two positions were selected as representative of faults in the plasma. Positions in each experiment were initially set to the same 5.94 and 6.80 volts, for impedance and phase motors, respectively. From the collected data, forty data elements were sampled and subsequently quantified with statistics, statistical mean, and standard deviation. Four statistical variables, two for each position, could thus be prepared. Due to little variations in the reflected power, its raw data was used. These five variables constitute a symptom and are presented to an input layer of the network. Meanwhile, the output pattern of a fault consists of four elements, each corresponding to a process factor. Depending on the degree of variation in the factor level, the fault severity of each element was varied from 0.2 to 1.0 with an increment of 0.2.

Table 1 Experimental factors and ranges involved in simulating fault data

Process Factors	Range	Increments	Units
Source Power	500 – 900	100	Watts
Pressure	7.5 – 17.5	2.5	mTorr
O_2 Flow Rate	50 – 90	10	sccm
Ar Flow Rate	50 – 130	20	sccm

3. Backpropagation neural network

The BPNN is the most widely used in plasma data modeling. A schematic of a BPNN is exhibited in Fig. 1. Rather than a conventional bipolar (B) sigmoid function, a linear (L) function was used in this study in the output layer since this network has proven effective in improving prediction accuracy [14]. Related gradients are represented by G_B and G_L. A basic weight update scheme, commonly known as the *generalized delta rule*, is expressed as:

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

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. Other m and η indicate the iteration number and an adjustable parameter called the "learning rate," respectively. The $\Delta W_{i,j,k}$ in (1) is defined as:

$$\Delta W_{i,j,k} = -\frac{\partial E}{\partial W_{i,j,k}}. \qquad (2)$$

By adjusting weighted connections recursively using the rule of Eq. (1) for all the units in the network, the network attempts to learn the relationships between fault symptoms and roots.

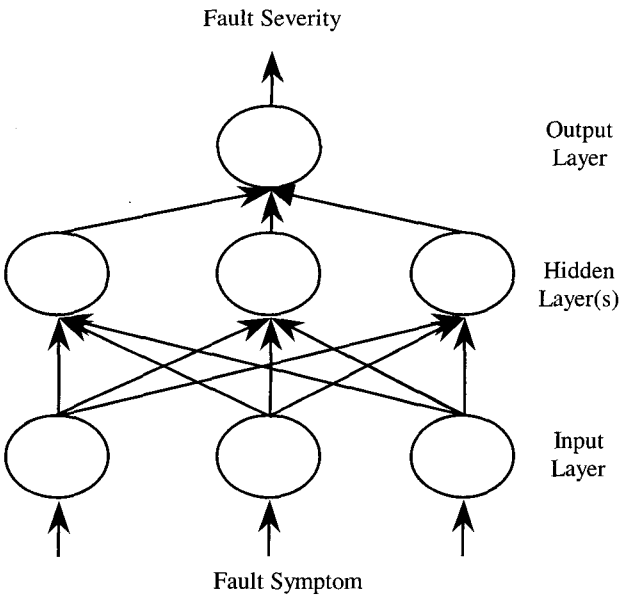

Fig. 1 A schematic of backpropagation neural network with a linear neuron in the output layer.

4. Diagnosis

4.1 Model Evaluation

Many training factors are typically involved in training a BPNN [15-16], including initial weight distribution, number of hidden units, and gradients of activation functions. Each factor was experimentally varied to examine its effect of diagnostic accuracy. Experimental factor ranges and their respective increments are contained in Table 2.

Table 2 Experimental training factors, ranges, and increments

Training Factors	Ranges	Increments
Initial weight distribution	± 0.4 – ± 1.8	0.2
Number of hidden units	2 – 7	1
G_B	0.6 – 2.0	0.2
G_L	0.4 – 1.2	0.1

Diagnostic accuracy, i.e., recognition rate, was measured with the root mean-squared error (σ), which is expressed as

$$\sigma = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij} - \hat{y}_{ij})^2}{pq}} \qquad (3)$$

where y_{ij} and $\hat{y}_{ij}$ indicate the actual and predicted fault severities for the ith output neuron and j th test pattern.

The p indicates the number of output neurons, here 4. Another q represents the number of fault patterns to be recognized by neural networks, here 20. As a termination criterion for network training, the number of epochs was set to 50000. Here the epoch equals the total number of the input patterns presented to the input layer through the whole process of network training. Results are shown in Figs. 2–5. Fig. 2 shows the accuracy as a function of the magnitude of initial weight distribution. As depicted in Fig. 2, the accuracy drastically increases as the magnitude of the initial weight matrix increases from ± 0.4 to ± 0.6. In contrast, the accuracy significantly decreases as the magnitude increases from ± 0.6 to ± 0.8. The accuracy no longer decreases with further increasing the magnitude. The lowest accuracy is obtained as 0.349 at a fairly smaller magnitude of ± 0.4. As exhibited in Fig. 3, diagnostic accuracy decreases consistently as the number of hidden neurons increases. The lowest accuracy is thus obtained as 0.200 at a larger number of 7. In Fig. 4, the accuracy is plotted as a function of the gradient of the bipolar sigmoid function. From Fig. 4, no consistency between the accuracy and the gradient is observed. The lowest accuracy, 0.307, is obtained at 1.4. For the gradients exceeding 1.4, the accuracy seems to increase with an increase in the gradient.

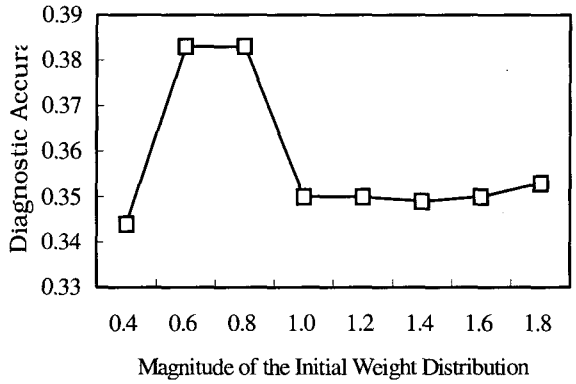

Fig. 2 Diagnostic accuracy as a function of the magnitude of the initial weight distribution.

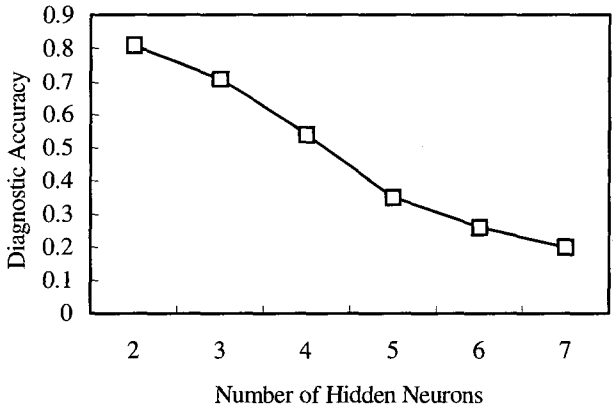

Fig. 3 Diagnostic accuracy as a function of the number of hidden neurons.

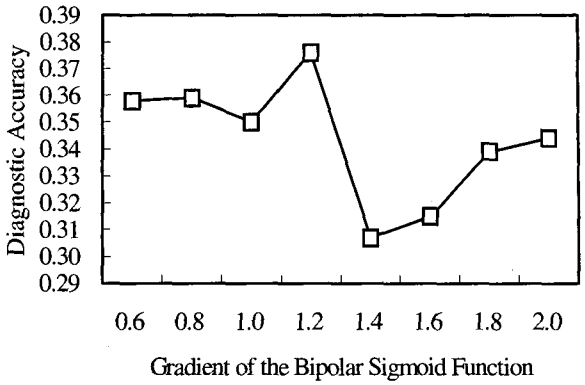

Fig. 4 Diagnostic accuracy as a function of the gradient of bipolar sigmoid function.

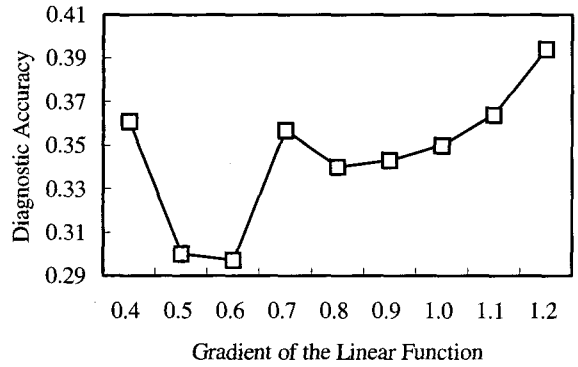

Fig. 5 Diagnostic accuracy as a function of the gradient of the linear function.

Fig. 5 displays the accuracy as a function of the linear function. The accuracy drastically decreases as the gradient increases from 0.4 to 0.5. At 0.6, the lowest accuracy, 0.297, is obtained. For the gradients that exceed 0.8, the accuracy increases with an increase in the gradient. These observations reveal that the lowest accuracy can be achieved by varying the number of hidden neurons. This implies that hidden neuron variable is the most critical factor in enhancing diagnostic accuracy.

4.2 Interaction Model

Additionally, the effect of interaction between the training factors on diagnostic accuracy are investigated. For this, a 2^4 full factorial experiment [17] was conducted on the four training factors. Factor ranges employed in the design were carefully selected from Figs. 2–5 and are contained in Table 3. For each of the experimental trials, diagnostic accuracy was measured with Eq. (3). The BPNN was then trained on this data and from the trained model several plots were prepared to examine trade-offs between training factors. Fig. 6 shows diagnostic accuracy as a function of the magnitude of the initial weight distribution and the number of hidden neurons. With an increase in the number of hidden neurons, the accuracy is improved. The improvement is more appreciable at larger magnitudes of the

initial weight distribution. Meanwhile, the accuracy seems slightly improved with a decrease in the initial weight magnitude at 5, but this is unclear at 7. A sharper variation in neural response surfaces with the initial weight distribution clearly indicates that the initial weight distribution plays more dominant role than the number of hidden neurons in determining the accuracy. Fig. 7 depicts the accuracy as a function of G_L and G_B. As G_L increases, the accuracy decreases, and this behavior is observed over an entire range of G_B. Meanwhile, the accuracy seems to decrease with a decrease in G_B. The lowest accuracy is thus achieved at higher G_L and lower G_B.

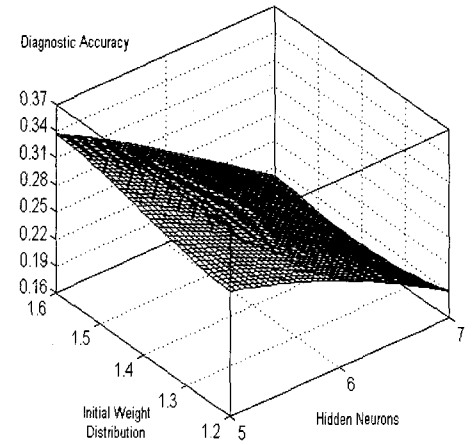

Fig. 6 Diagnostic accuracy as a function of the magnitude of the initial weight distribution and the number of hidden neurons.

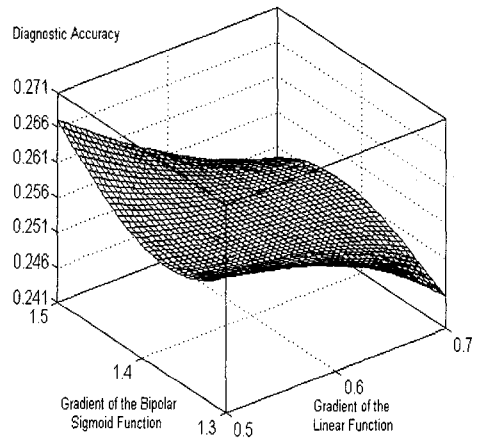

Fig. 7 Diagnostic accuracy as a function of the gradients of the bipolar sigmoid function and the linear function.

Table 3 Experimental ranges of training factors

Training Factors	Ranges
Number of Hidden Neurons	5 – 7
Magnitude of Initial Weights	±1.2 – ±1.6
G_L	0.5 – 0.7
G_B	1.3 – 1.5

5. Conclusion

A new strategy for plasma diagnosis was presented and used a neural network to recognize faults in the plasma. A total of twenty fault-symptoms were simulated and the symptoms were effectively represented by two static match positions. Plasma faults were simulated with a variation in source power, pressure, Ar, and O_2 flow rates. Diagnostic accuracy was examined as a function of training factors, which include the number of hidden neurons, the magnitude of the initial weight distribution, and the gradient of neuron activation functions. The accuracy was most sensitive to variations in the number of hidden neurons. Factor interaction effects on the accuracy were investigated via a statistically designed experiment. The method proposed can further be applied to other in-situ diagnostic data.

Acknowledgements

The work was supported by a Korea Research Foundation grant (KRF-2000-003-E00160).

References

[1] C. Spanos, "HIPPOCRATES: A methodology for IC process diagnosis," *Proc. ICCAD*, 1986.

[2] J. Pan and J. Tenenbaum, "PIES: An engineer's do-it yourself knowledge system for interpretation of parametric test data," *Proc. 5th Nat. Conf. on AI*, 1986.

[3] B. Kim and Gary S. May, "Real time diagnosis of semiconductor manufacturing equipment using a hybrid neural network expert system," *IEEE Trans. Comp. Pack. Manufac. Tech.*, vol. 20, no. 1, pp. 39–47, 1997.

[4] J. O. Stevenson, P. P. Ward, M. L. Smith, and R. J. Markle, "A plasma process monitor/control system," *Surf. Interf. Anal.*, vol. 26, pp. 124–133, 1998.

[5] M. Hyland, D. Mariotti, W. Dubitzky, J. A. McLaughlin, and P. Maguire, "Gas recognition using a neural network approach to plasma optical emission spectroscopy," *Proc. SPIE Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation III*, pp. 246–252, 2000.

[6] H. T. Mayfield, D. Eastwood, and L. W. Burggraf, "Infrared spectral classification with artificial neural networks and classical pattern recognition," *Chem. Biol. Sen.*, pp. 54–65, 2000.

[7] T. Udelhoven and B. Schutt, "Capability of feedforward neural networks for a chemical evaluation of sediments with diffuse reflectance spectroscopy," *Chemom. Intell. Lab. Syst.*, vol. 51, no. 1, pp. 9–22, 2000.

[8] E. Bertran, M. Blanco, S. Maspoch, M. C. Ortiz, and

L. A. Sarabia, "Handling intrinsic non-linearity in near-infrared reflectance spectroscopy," *Chemom. Intell. Lab. Syst.*, vol. 49, no. 2, pp. 215–224, 1999.

[9] I. Belic and L. Gyergyek, "Neural network methodologies for mass spectra recognition," *Vacuum*, vol. 48, no. 7–9, 1997.

[10] N. W. Daniel, I. R. Lewis, and P. R. Griffiths, "Interpretation of raman spectra of nitro-containing explosive materials-Part II- The implementation of neural, fuzzy, and statistical models for unsupervised pattern recognition," *Appl. Spectrosc.*, vol. 51, no. 12, 1997.

[11] B. Kim and C. Lee, "Monitoring plasma impedance match characteristics in a multipole inductively coupled plasma for process control," *J. Vac. Sci. Technol. A*, vol. 18, no. 1, pp. 58–62, 2000.

[12] Intelligent Instrumentation, Visual Designer Reference Manual.

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

[14] B. Kim, W. Choi, and H. Kim, "Using neural networks with a linear output neuron to model plasma etch process," *Sixth IEEE International Symposium on Industrial Electronics*, vol. II, pp. 441–445, 2001.

[15] B. Kim and G. S. May, "An optimal neural network process model for plasma etching," *IEEE Trans. Semi. Manufac.*, vol. 7, no. 1, pp. 12–21, 1994.

[16] B. Kim and S. Park, "An optimal neural network plasma model: A case study," *Chemom. Intell. Lab. Syst.*, vol, 56, no. 1, pp. 39–50, 2001.

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

Byungwhan Kim received the B. S. and M. S. degrees in Electrical Engineering from Korea University in 1985 and 1987, respectively, and the PH. D. degree in the School of Electrical and Computer Engineering at the Georgia Institute of Technology, Atlanta, in 1995. He was a principal technical staff at Hyundai Electronics in 1996, a post-doctoral research fellow and IPS assignee at Korea University in 1998, and a faculty member at Chonnam National University until 2000. He is now an assistant professor in Electronic Engineering at Sejong University, Korea. His research interests include neuro-fuzzy modeling, optimization, diagnosis, and control of semiconductor processes.

Tel: +82-2-3408-3729, Fax: +82-2-3408-3329

E-mail: kbwhan@sejong.ac.kr