Modeling of Ion Energy Distribution Using Time-Series Neural Network

Suyeon Kim, Byungwhan Kim*
Department of Electronic Engineering
Sejong University
Seoul 143-747, Korea

Abstract: - The plasma is susceptible to conditions in etching and depositing thin films. In manufacturing plasma processes, ion energy distribution (IED) plays an important role in studying ion energy effects on films qualities. The IED can be used to monitor plasma conditions. For a real-time monitoring purpose, auto-correlated time-series neural network model of IED is presented. IEDs were measured by a non-invasive ion analyzer. The experiments were conducted by using a pulsed-plasma enhanced chemical vapor deposition system. Monitoring performance of neural network model was measured as a function of duty ratio and hidden neuron number. The prediction error quantified by the root-mean square error of optimized model is 0.334. The improvement of this model over the validation data is more than 57%. The proven high improvement indicates that plasma faults can be detected by means of A-NTS model of IED.

Key-Words: - Time-series model; Neural network; Ion energy distribution; Plasma; Monitoring; Diagnosis; Back-propagation neural network

1 Introduction

The plasma generally plays a crucial role in depositing or etching thin films in manufacturing integrated circuits. To maintain process quality, plasma states inside the equipment should be stringently monitored. For this purpose, a huge number of in-situ diagnostic instruments have been applied to monitor certain variation in plasma. These may include an optical emission spectroscopy [1-3], a radio-frequency (RF) impedance sensor [4-6], or a RF matching network monitor [7]. These instruments monitor plasmas by observing variations in radicals, impedance in terms of resistance and reactance, or matching variables such as a reflected power. Apart from these, ion energy distribution (IED) measured by using an ion energy analyzer can effectively be used for plasma monitoring. IED is strongly involved in determining films characteristics such as a deposition rate or etching rate. Most importantly, it may a good indicator to characterize various properties on film surfaces.

IED is expected to vary sensitively to a variation in process parameters such as a RF bias power or pressure. For a real-time monitoring purpose, a neural network time-series model constructed with diagnostic data was used to monitor plasma states [2-3, 8]. However, there were no reports on neural network model of IED.

In this study, a neural network time-series model of IED is presented. The IED data were collected by using a Non-invasive Ion Energy Analyzer [8-10]. Validation studies [9-10] show that the technique (i.e. NIEA) is well suited for use in manufacturing to monitor in real time, since it is completely noninvasive.

Anomalies in plasmas were simulated by changing a process parameter called duty ratio. The performance of IED models were optimized as a function of time variables as well as in terms of the number of hidden neurons (NHN).

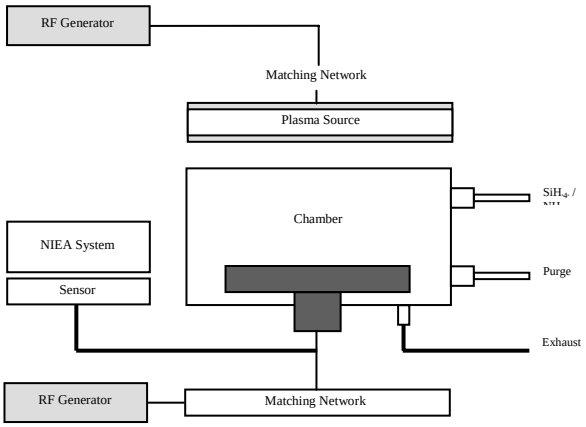

Fig. 1 Schematic of the Plasma System

2 Experimental Data

A SiN films were deposited by using a PECVD system (PLASMART^{TM}). Schematic of the equipment is shown in Fig. 1. A cylinder type of chamber has a diameter of 0.27m and a height of 0.2m. A unique more than $5\times10^{11}/cm^3$ in Ar plasma as well as a plasma double stack antenna produces a plasma density of non-uniformity of less than ±5%. Chamber vacuum was controlled by a turbo molecular pump, a rotary pump, and a vacuum gauge. Gas flow rates were precisely controlled through the mass flow controllers and the process pressure was controlled by a throttle valve. Meanwhile, a non-invasive ion energy analyzer was used to collect IED data. To simulate faults in plasma, the varied duty ratios are 80, and 100%. The flow rates of SiH_4 and NH_3 were set to 8 and 22sccm, respectively. The bias power and frequency were fixed at 50W and 250 Hz. The data acquisition time was set to 5 min. The ion energy analyzer provides various distributions such as low or high IED. Here, the low IED was used for modeling.

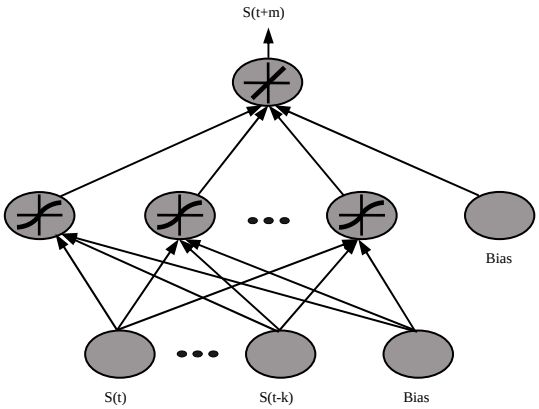

Fig. 2 Structure of the A-NTS

3. Neural Network Model

A schematic of back-propagation neural network (BPNN) [11] employed for the prediction of IED is shown in Fig. 2. As shown in Fig. 2, the output of the hidden layer is determined by a bipolar sigmoid function. The equations of bipolar sigmoid function can be written as

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

where $in_{i,k}$ and $out_{i,k}$ are the weighted input and output to the i th neuron in the k th layer, respectively. And g_b is the gradient of the bipolar sigmoid function.

The output of the output layer is determined by a linear function, which is expressed as

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

where g_l is the gradient of the linear function.

A weight update equation, commonly known as *the generalized delta rule* [12], 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. $\Delta W_{i,j,k}$ is the calculated change in the weight to minimize error (E) of all the input-output pairs. η is the learning rate. In this study, that was set to 0.01.

As shown in Fig. 2, the IED is predicted by using the past and current data of the same output. This type of time series model constructed with neural network is referred to as auto-correlated neural time series (A-NTS) model. As seen in Fig. 2, the prediction performance of A-NTS model is affected by the kinds of combinations of (k, m). In this study, either k or m was varied from 1 to 3 with an increment of 1. Therefore, A-NTS models were constructed for a total of 9 combinations. Also, the performance of A-NTS models was evaluated as a function of NHN. The NHN was increasingly varied from 2 to 9 by one. Other training factors involved in BPNN training as well as those stated earlier are shown in Table 1.

Table 1. Conditions of Modeling Parameters

Parameter	Ranges
Error Tolerance	0.1
Learning rate	0.01
Weight Distribution	±1
Gradient of Slope	1
K	1, 2, 3
M	1, 2, 3
NHN	2 ~ 9

ISBN: 978-960-6766-83-1 160 ISSN: 1790-2769

4 Results

The performance of A-NTS model was measured by the root mean square error (RMSE), defined as

$$RMSE = \sqrt{\frac{\sum_{j=1}^{q}(d_j - out_j)^2}{q}} \quad (4)$$

where q is the total number of test data, and d_j and out_j are the desired output and the calculated output of the i th neuron in the output layer, respectively. The RMSEs calculated with the training and testing data are referred to as T-RMSE and P-RMSE, respectively. The IED data collected at 100% were used to construct a neural network model. The performance of constructed model was validated with other data set at 80%. The data of 100% were separated into two training and testing data. Each data consisted of 66 patterns. For example, the performance of A-NTS model constructed at (1, 1) and NHN 2 is shown in Fig. 3(a) and (b).

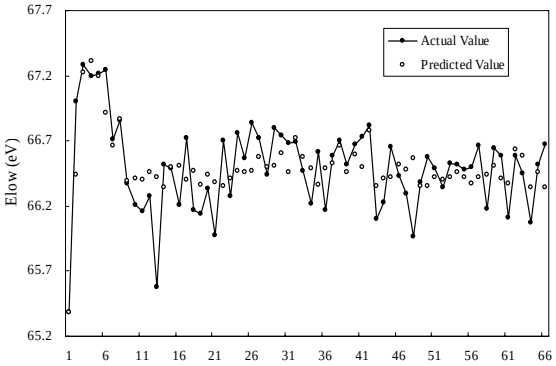

Fig. 3(a) Training values of 100% model

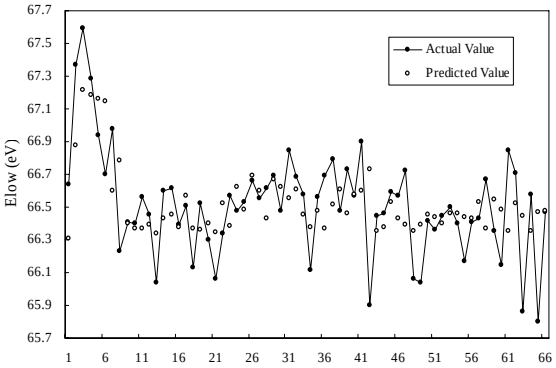

Fig. 3(b) Prediction values of 100% model

The training and testing RMSEs for the constructed A-NTS model are 0.2396, 0.2686, respectively. As shown in Fig. 3(b), the IED pattern predicted from the A-NTS model matches well the corresponding actual one.

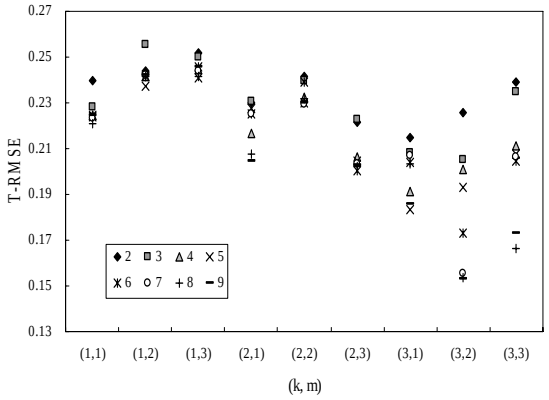

Fig. 4(a) Training errors as a function of NHN and (k, m)

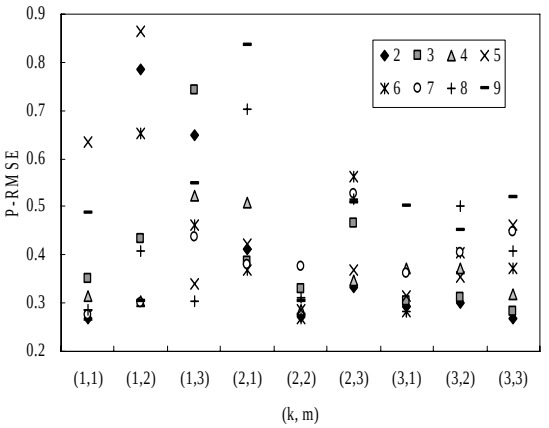

Fig. 4(b) Prediction errors as a function of NHN and (k, m)

The performance of A-NTS model is optimized as a function of NHN and (k, m) combination. The results are shown in Fig. 4(a) and (b). Figure 4(a) and (b) show the RMSE in terms of training and testing errors, respectively. As seen in Fig. 4(a), the T-RMSE seems to be affected considerably with either NHN or (k, m). Particularly, the effect of NHN is noticeable at (3, 2) and (3, 3). The smallest T-RMSE of 0.1533 is obtained at 9 NHN and (3, 2) (k, m). Compared to T-RMSE, as shown in Fig. 4(b), the variations in P-RMSE are much larger. The smallest P-RMSE of 0.2673 is obtained at 2 NHN and (3, 3) (k, m).

Figure 5 shows the variations in P-RMSE for 100% model validated with 80%, respectively.

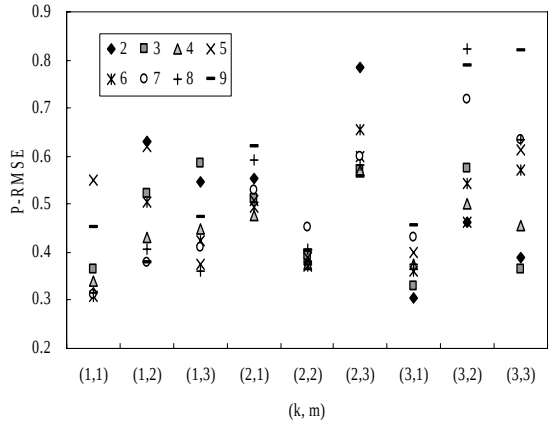

Fig. 5 Validation errors of 100% model for 80% data

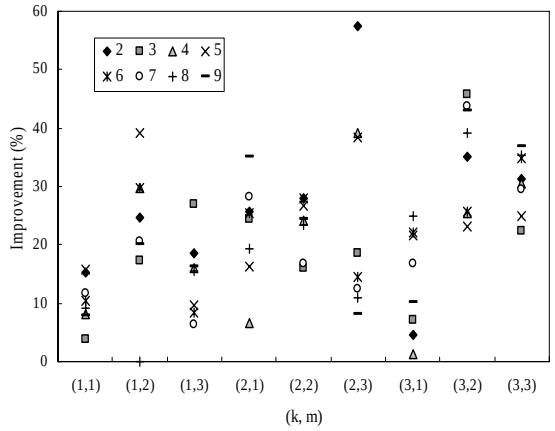

Fig. 6 Monitoring sensitivity of 100% model for 80% IED

To evaluate the monitoring performance of A-NTS model as a function of NHN and combinations, the improvements of optimized A-NTS model as to those predictions shown in Figs. 5 were calculated as

Improvement =

$$\left| \frac{(P-RMSE_{100\%} - P-RMSE_{80\%})}{P-RMSE_{80\%}} \times 100 \right| (\%) \quad (6)$$

The improvement defined above can be used as an indicative of monitoring sensitivity for the variations in either NHN or combination. The results for 80% IED are shown in Fig. 6. As shown in Fig. 6, the monitoring sensitivity varies quite differently depending on NHN or combinations. All the improvements shown in Fig. 6 are in the range 0.03-57.44%. The largest improvement occurred at 2 NHN and (2, 3) (k, m). The corresponding P-RMSE is about 0.334. In several cases, the improvement is less than 10%. The corresponding combinations and NHNs should be avoided in building A-NTS model for plasma monitoring. For the optimized model determined in Fig. 4(b), the corresponding improvement is 33.1%. This value is much smaller than the one just determined. Therefore, the model found in Fig. 6 is determined as the most accurate model.

Table 2. Evaluation of overall performance of A-NTS model as a function of combination at fixed NHN.

NHN	Combination	Improvement (%)	Sum (%)
2	(2, 3)	57.44	240.00
3	(3, 2)	45.79	182.30
4	(2, 3)	39.11	180.68
5	(1, 2)	39.24	215.70
6	(3, 3)	34.80	199.49
7	(3, 2)	43.66	185.90
8	(3, 2)	39.09	177.70
9	(3, 2)	43.00	202.19

Table 3. Evaluation of overall performance of A-NTS model as a function of NHN at fixed combination.

Combination	NHN	Improvement (%)	Sum (%)
(1, 1)	5	15.75	82.02
(1, 2)	5	39.24	181.35
(1, 3)	3	26.95	117.90
(2, 1)	9	35.16	181.01
(2, 2)	6	28.01	187.52
(2, 3)	2	57.44	199.88
(3, 1)	8	24.88	108.41
(3, 2)	3	45.79	280.80
(3, 3)	9	36.87	245.08

Next, the overall performance of A-NTS model was evaluated as a function of NHN or combination. For this, those improvements were summed as a function of NHN or combination. The results are shown in Table 2 and 3. In either table, the optimized NHN or combinations, their corresponding improvements, and sums are provided. For example, the information just stated is shown as 9 combinations were varied at 2 NHN. A sum of 240 is simply obtained by adding 9 improvements obtained for all combinations at 2 NHN. As shown in Table 2,

the sum is in the range 180-240%. The largest sum is achieved at 2 NHN. This indicates that the effect of combinations on improvement of A-NTS model is the most significant particularly at 2 NHN. And the largest improvement in Table 2 was obtained at the largest sum. As the NHN was varied at one fixed combination, the results are shown in Table 3. In this case, the sum is in the range 82-281%. The largest sum is obtained at (3, 2). This indicates that the effect of varying NHN is the most significant particularly at (3, 2) combination. In consequence, the largest sum in Table 3 is larger than that in Table 2. This means that varying NHN at a fixed combination yielded better performance in overall fashion. As revealed already, however, it should be noted that this does not guarantee the construction of the most accurate prediction model.

4 Conclusion

In this study, A-NTS models of IED data were constructed by using the BPNN. The performance of A-NTS model was evaluated individually as well in overall fashion. The performance of A-NTS model was detailed in terms of NHN and various combinations. The optimized model yielded a high sensitivity to the faults simulated experimentally. The demonstrated accuracy seems to be enough to detect faults concerned here, but its applicability to detect smaller variations in IED still remains a further study.

Acknowledgements

This research was supported by the Ministry of Knowledge Economy, Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA(Institute of Information Technology Advancement) (IITA-2008-C109008010030).

References:

[1] J. O. Srevenson, P. P. Ward, M. L. Smith, R. J. Markle, A plasma process monitor/control system, *Surf. Interf. Anal.* Vol. 26, 1998, pp. 124-133.

[2] S. J. Hong and G.S. May, Neural network based time series modeling of optical emission spectroscopy data for fault detection in reactive ion etching, *Proc. SPIE.*, Vol. 5041, 2003, pp. 1-8.

[3] S. J. Hong and G.S. May, Neural network-based real-time malfunction diagnosis of reactive ion etching using in situ metrology data, *IEEE Trans. Semicond. Manufact.*, Vol. 17, 2004, pp. 408-421.

[4] S. Bushman, T. F. Edgar, I. Trachtenberg, Radio Frequency Diagnostics for Plasma Etch Systems, *J. Electronchem Soc.*, Vol. 144, No. 2, 1997, pp. 721-732

[5] F. A. Bose, R. Patrick, H. Baltes, Measurement of discharge impedance for dry etch process control, *Proc. SPIE.*, Vol. 2336, 1994, pp. 101-110

[6] W. C. Roth, R. N. Carlile, J. F. O'Hanlon, Electrical Characterization of a processing of a processing plasma chamber, *J. Vac. Sci. Technol. A*, Vol. 15, No. 6, Nov/Dec 1997, pp. 2930-2937.

[7] B. Kim, 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, Jan/Feb 2000, pp. 58-62.

[8] Non-invasive Ion Energy Analyzer Plasma Monitoring System, NIEA™ System, PASMART. Inc.

[9] M. A. Sobolewski, Monitoring sheath voltages and ion energies in high-density plasmas using noninvasive radio-frequency current and voltage measurements, *J. Appl. Phys.*, Vol. 95, No. 9, 2004, pp. 4593-4604.

[10] M. A. Sobolewski, Real-time, noninvasive monitoring of ion energy and ion current at a wafer surface during plasma etching, *J. Vac. Sci. Technol. A.*, Vol. 24, No. 5, 2006, pp. 1892-1904.

[11] M. D. Baker, C. D. Himmel, G. S. May, Time Series Modeling of Reactive Ion Etching Using Neural Networks, *IEEE Trans. Semicond. Manufact.*, Vol. 8, No. 1, 1995, pp. 62-71.

[12] C. D. Himmel, G. S. May, Advantages of plasma etch modeling using neural networks over statistical techniques, *IEEE Trans. Semicond. Manuf.*, Vol. 6, 1993, pp. 103-111.