Prediction of Etch Profile Uniformity Using Wavelet and Neural Network

Won Sun Choi, Myo Taeg Lim, and Byungwhan Kim*

Abstract: Conventionally, profile non-uniformity has been characterized by relying on approximated profile with angle or anisotropy. In this study, a new non-uniformity model for etch profile is presented by applying a discrete wavelet to the image obtained from a scanning electron microscopy (SEM). Prediction models for wavelet-transformed data are then constructed using a back-propagation neural network. The proposed method was applied to the data collected from the etching of tungsten material. Additionally, 7 experiments were conducted to obtain test data. Model performance was evaluated in terms of the average prediction accuracy (APA) and the best prediction accuracy (BPA). To take into account randomness in initial weights, two hundred models were generated for a given set of training factors. Behaviors of the APA and BPA were investigated as a function of training factors, including training tolerance, hidden neuron, initial weight distribution, and two slopes for bipolar sigmoid and linear function. For all variations in training factors, the APA was not consistent with the BPA. The prediction accuracy was optimized using three approaches, the best model based approach, the average model based approach and the combined model based approach. Despite the largest APA of the first approach, its BPA was smallest compared to the other two approaches.

Keywords: Back-propagation neural network, plasma etching, model, optimization, uniformity, wavelet.

1. INTRODUCTION

Since the early 1990s, intelligent systems such as neural networks have been widely used to model plasma etching and deposition processes [1-3]. Despite the complexity of plasmas, neural networks have greatly contributed to capture and visualize nonlinear relationships between process inputs and etch responses. There exist many etch responses to model, which may include a typical etch rate, selectivity, profile, and non-uniformity. Among them, the profile was approximated with an angle or anisotropy. Due to this approximation, detailed variations along the profile surface were not sensitively characterized. Previous prediction models are thus limited in their ability to reveal variations in profile completely. To circumvent this deficiency inherent in the conventional metric, a wavelet technique was recently applied to capture micro variations along a profile surface [4]. Compared to the conventional metric, the wavelet-based one demonstrated improved sensitivity to variations in process inputs. In the context of plasma etching, non-uniformity was defined solely for etch rate. Another non-uniformity for etch profile has never been presented in literatures mainly due to the approximation nature of the conventional metric.

In order to gain insight into a profile in progress, a new method to measure profile non-uniformity must be defined. In this study, this is accomplished by a discrete wavelet transformation in conjunction with a simple formula traditionally employed for non-uniformity characterization. Moreover, a novel prediction model of profile non-uniformity is constructed using a back-propagation neural network. As a function of training factors, the performance is evaluated in terms of the average and best prediction accuracy. For this modeling, a 2^{4-1} fractional factorial

Manuscript received January 7, 2003; revised February 4, 2004; accepted February 9, 2004. Recommended by Editorial Board member Sun Kook Yoo under the direction of Editor Jin Bae Park. This work was supported by the Information and Display R & D Center, 21 C Frontier R & D Program, Korean Ministry of Science and Technology, under grant number M102KR01001-03K1801-01111.

Won Sun Choi is with the Division of Large Scale Integration, Samsung Electronics, Kiheung (e-mail: wonsun.choi @samsung.com).

Myo Taeg Lim is with the School of Electrical Engineering, Korea University, 5-1, Anam-Dong, Seoul 136-701, Korea (e-mail: mlim@elec.korea.ac.kr).

Byungwhan Kim is with the Department of Electronic Engineering, Sejong University, 98, Kunja-Dong, Kwangjin-Ku, Seoul 143-747, Korea (e-mail: kbwhan@sejong.ac.kr).

* Corresponding author.

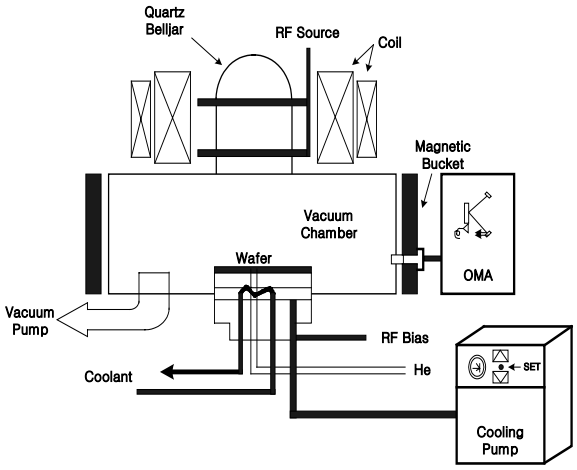

Fig. 1. A schematic of a Pinnacle 8000 helicon plasma etch system.

experiment [5] was conducted while collecting additional experiments. The data were collected from the etching of a tungsten material in a SF_6 plasma. This process was empirically modeled and related etch mechanisms were investigated [2].

2. EXPERIMENTAL DATA

A schematic of a Pinnacle 8000 helicon plasma etch system is shown in Fig. 1 A high density plasma is produced in a source quartz by coupling 13.56 Mhz radio frequency (RF) power to an antenna via a corresponding network. The coil surrounding the quartz provides an efficient transfer of energy into the center of the plasma as well as an effective confinement of the plasma, thereby increasing plasma density. A magnetic bucket encircling the process chamber comprised of 24 permanent magnets with alternating polarity, yielding highly uniform plasma. The temperature of the chuck holder was controlled by a SC-90 TAITAC cooling pump. By cooling the coolant through the heat exchanger and subsequently by compressing it with the pump, the substrate temperature could be cooled down to a temperature as low as −50℃. The chamber pressure is maintained at the base level of 10^{-6} Torr by using a turbo pump and rotary pumps. Gas flow rates are precisely controlled through mass flow controllers.

Test patterns were fabricated on 8-inch diameter silicon wafers of (100) orientation. The pattern was structured as a 3500Å W with a 600Å TiN diffusion barrier layer and a 500Å Ti adhesive film on a 4000 Å silicon dioxide film. Deep ultra-violet photoresist of 1 ㎛ was spun and baked for 30 minutes at 120℃. To examine variations in etch profiles, a scanning electron microscopy was used to take pictures of etched patterns.

The process was characterized with a 2^{4-1} fractional factorial experiment. The factors that were varied in the design include a RF source power, bias power, chuck temperature, and SF_6 flow rate. The experimental range is contained in Table 1. To evaluate the appropriateness of the model, 7 additional experiments were conducted.

Table 1. Experimental ranges of input factors.

Factor	Range	Units
Source Power	1500 - 2500	watts
Bias Power	0 – 30	watts
SF_6	80 – 120	sccm
Temperature	-50 – 10	℃

3. WAVELET THEORY AND NEURAL NETWORK

3.1. Wavelet theory

Due to the property of efficient space-frequency localization, a wavelet can effectively be used to examine variations of a signal or an image. The DWT is mathematically detailed in previous works [6]. Rather than repeating it, the fundamentals of DWT are briefly explained. When the DWT is applied to a function f, f can be approximated as

$$f(t) = \sum_k (f, \phi_{j0,k})\phi_{j0,k} + \sum_{j>j0} \sum_k (f, \psi_{j,k})\psi_{j,k}, \quad (1)$$

where ϕ is an orthonomal basis for the scaled subspace V_j of a central subspace V_o. The other ψ forms an orthonomal basis for the subspace W_j, the complement of V_j. Both $(f, \phi_{j_0,k})$ and $(f, \psi_{j,k})$ are referred to as the approximation (or scale) and detail (or wavelet) coefficients, respectively. As represented in (1), f can be approximated as the sum of the approximation of f at level j_0 in addition to the details concerning f at $j > j_o$. For a given image, the decomposition is successively conducted by a pair of low and high pass filters, separately in two directions. The low and high pass filters provide the approximation and details of the image, respectively. The decomposition is first applied to each row of an image array. The high pass and low pass sub-images thereby obtained are each separately filtered column-wise, resulting in 4 sub-images corresponding to low-low-pass, low-high-pass, high-low-pass, and high-high pass row column filtering, respectively. The DWT of a 2-dimensional image is mathematically detailed [7].

3.2. Neural networks

Although many paradigms were available, a back-propagation neural network (BPNN) [8] was chosen.

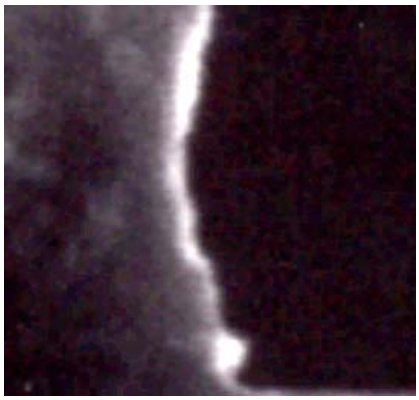

Fig. 2. Extracted profile from an original SEM photo.

The BPNN consisted of three layers, input, hidden and output. The input layer receives external information such as the four process factors (source power, bias power, SF_6 flow rate, substrate temperature). Then, the output neurons produce certain predictions. Here, the number of output neurons was set to unity since only one profile nonuniformity was modeled. 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 neurons is an import training factor. The number of hidden layers was set to unity in this study. The error the network attempts to minimize is the accumulated error (E) of all the input-output pairs, which for a given test input pattern is expressed 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 ith neuron in the layer, and out_i is the calculated output from the same neuron. In the BP algorithm, the 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* [8], 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 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 so called "learning rate," respectively. By adjusting weighted connections recursively using the rule in (3) for all units in the network, the accumulated E over all the input vectors is to be minimized.

4. RESULTS

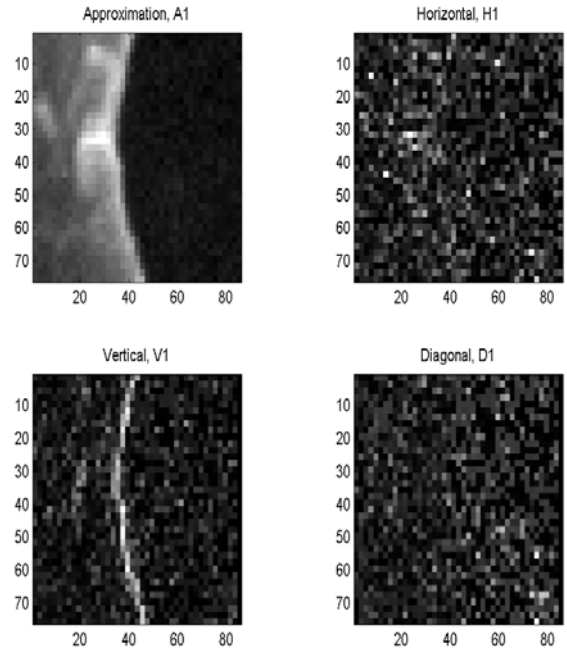

Fig. 3. Wavelet transformed SEM image.

4.1. DWT-characterized profile non-uniformity

A SEM photo is shown in Fig. 2. The etching condition is 2000 W source power, 15 W bias power, -20℃ temperature, and 100 sccm SF_6. As represented in Fig. 2, there exists some curvature in the profile, which cannot be quantified using the conventional metric. To circumvent this difficulty, wavelet technique is used to characterize micro variations on the profile surface. The SEM photo was scanned using a scanner of 150 dpi resolution in a gray scale. An image file of 650 by 500 was then obtained, from which the profile image of 96 by 96 was cut by means of iPhotoPlus 4. The profile was converted in the bit map format rather than in JPEG format so as to prevent any loss of information. When DWT is applied to the BMP file, 4 blocks appear on the screen as depicted in Fig. 3, each corresponding to an approximation of the horizontal, vertical and diagonal blocks.

A combined image (Im_c) of the vertical and horizontal images, represented as Im_v and Im_h respectively, are obtained as

$$Im_c = \sqrt{Im_v^2 + Im_h^2} . \qquad (4)$$

A profile contour was then extracted by applying the histogram method to the combined image characterized by (4). Using the gray levels of images, a rare profile image was made and is depicted in Fig. 4. The DWT was then implemented on the image in Fig. 4 four times consecutively. The original image was finally compressed down to a 3 by 3. The resulting nine coefficients are contained in Table 2.

Table 2. Wavelet coefficients at 4 scale level.

159	685	32
207	653	80
32	877	80

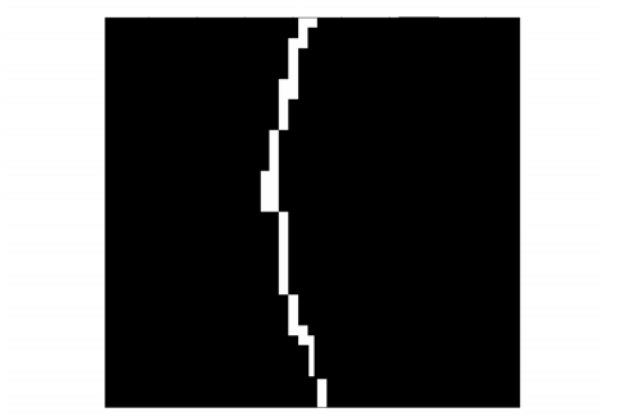

Fig. 4. Extracted profile contour using a histogram method.

Wavelet coefficients contained in Table 2 were subsequently averaged with a simple metric defined as

$$V = \frac{\sum_{i=1}^{n} v_i}{n}, \tag{5}$$

where v_i indicates each compressed wavelet coefficient. Another n represents the total number of wavelet coefficients, in this case 9. The nonuniformity is then computed as

$$\text{Non-uniformity} = \frac{|V_c - V_e|}{V_c} \times 100, \tag{6}$$

where V_c is obtained for the profile at the center. Another V_e is obtained as the mean for the four profiles at about a one inch distance from the edge.

4.2. Optimization of prediction model

The prediction accuracy of the predictive model is optimized in the following three approaches. In each optimization, the prediction accuracy is examined as a function of training factors. The effects of training factors are investigated from the standpoint of the average prediction accuracy (APA) and best prediction accuracy (BPA). The BPNN model was trained with 7 experiments and tested with 7 experiments. The prediction accuracy was quantified with the root-mean squared error (RMSE) metric defined as

$$RMSEP = \sqrt{\frac{\sum_{p=1}^{P} (T_p - O_p)^2}{P}}, \tag{7}$$

Table 3. Experimental ranges of training factors.

Factor	Range
TT	0.02-0.16
HN	2-9
IWD	± 1.6-± 3.0
GBS	0.3-3.0
GL	0.3-3.0

Table 4. Prediction accuracy versus TT.

TT	APA	BPA
0.02	14.05	11.65
0.04	14.05	11.78
0.06	14.05	11.82
0.08	14.05	11.86
0.10	14.06	11.90
0.12	14.07	11.95
0.14	14.08	12.00
0.16	14.09	12.05

where P is the size of test data, T_p is the measured etch response for the pth input, and O_p is the corresponding prediction.

4.2.1 Best model-based optimization

In the first approach, the prediction accuracy is optimized by optimizing the BPA as a function of training factors, including training tolerance (TT), hidden neuron (HN), the gradients of bipolar sigmoid function (GBS) and linear function (GL), and initial weight deviation (IWD). Particularly, a number of models were generated to take into account the randomness in IWD. In this way, an optimal plasma model was constructed [3]. The number of model generation was set to 200. The experimental ranges for the factors are included in Table 3. First, the TT effect on the prediction accuracy is examined and results are contained in Table 4. The other factors were set to their default values. In other words, the HN, IWD, GBS and GL were set to 4, ± 1.0, and 1.0, respectively. Both the average and best RMSEPs are contained in Table 4. As contained in Table 4, the APA varies little with the TT. On the other hand, the BPA increases with increasing TT. This clearly illustrates that both behaviors are completely different. This was once demonstrated in modeling Langmuir probe discharge data [3]. One BPA is determined at 0.02 and is numerically equal to 11.65 %.

The HN number was then experimentally varied. Here, the TT was set to 0.02 for the best model obtained with variations in the TT. The other factors were set to their default values stated earlier. Results are shown in Table 5. For the HN between 4 and 9, the APA varied little with HN. When decreasing the HN to the value smaller than 4, the APA increases appreciably. Meanwhile, the BPA varies randomly. The inconsistency is thus clear between the APA and

Table 5. Prediction accuracy versus HN.

HN	APA	BPA
2	15.02	11.16
3	14.34	10.77
4	14.06	11.65
5	14.06	10.22
6	13.95	11.66
7	13.94	11.05
8	13.93	10.56
9	13.88	10.73

Table 6. Prediction accuracy versus IWD.

IWD	APA	BPA
± 1.6	14.60	9.56
± 1.8	14.94	9.30
± 2.0	15.33	9.33
± 2.2	15.86	8.87
± 2.4	16.47	8.55
± 2.6	17.11	8.37
± 2.8	17.89	8.32
± 3.0	18.72	8.36

Table 7. Prediction accuracy versus GBS.

GBS	APA	BPA
0.3	24.25	8.83
0.5	21.83	8.07
0.7	19.97	8.85
1.0	17.89	8.32
1.5	15.71	9.60
2.0	14.84	9.72
2.5	14.37	9.90
3.0	14.19	10.26

BPA. One optimized BPA is obtained at five hidden neurons and its RMSEP is 10.22. Approximately 12.2% improvement is thus achieved with respect to the preceding model optimized as a function of TT.

The magnitude of IWD was then adjusted and results are shown in Table 6. Here, the TT and HN were set to 0.02 and 5 with the other factors set at their default values. The APA decreases with a reduction in IWD. In contrast, the BPA increases for the same variations in IWD. One optimized BPA is achieved at ±2.8 and is numerically equal to 8.32%. With the control of IWD, the BPA is thus improved by about 18.5% with respect to the preceding model obtained with variations in HN. This large amount of improvement indicates that the IWD plays an important role in affecting the prediction accuracy. As in the case of TT, the behavior of APA was inconsistent with that for the BPA. The GBS was then tuned and results appear in Table 7. The other TT, HN, and IWD were set to those optimized previously. The

Table 8. Prediction accuracy versus GL.

GL	APA	BPA
0.3	16.29	8.38
0.5	17.69	7.86
0.7	19.23	7.55
1.0	17.89	8.32
1.5	26.47	8.80
2.0	31.18	8.37
2.5	35.66	9.07
3.0	39.78	10.71

Table 9. Average model-based prediction accuracy.

IWD		GBS		GL	
APA	BPA	APA	BPA	APA	BPA
13.78	12.86	14.18	10.66	13.93	12.78
13.78	12.49	13.89	11.42	13.80	12.58
13.81	11.59	13.81	11.72	13.77	12.60
13.88	10.73	13.78	12.49	13.78	12.49
14.02	10.22	13.80	12.81	13.81	12.10
14.22	10.01	13.84	13.06	13.84	12.09
14.54	10.06	13.86	13.23	13.86	12.08
14.99	9.76	13.98	13.37	13.88	12.09

APA increases considerably with a decrease in the gradient. Meanwhile, the BPA continues to decrease until the GBS decreases to 1.0. For the GBSs less than 1.0, the BPA varies randomly. One optimized BPA is obtained at 0.5 and is equal to 8.07%. Compared to the preceding model, the improvement is tiny. The alternate GL was adjusted by setting the other factors set to their optimized values. Results are shown in Table 8. In general, the APA decreases with a decrease in the GL. This is contrary to what was observed for variations in the GBS. Random behavior in the BPA with the GL is depicted in Table 8. This is somewhat similar to that shown in Table 7. From this observation, it is thus noticed that both gradient affects are similar in view of the BPA. In contrast, they are opposite with respect to the APA. One optimized BPA is obtained at 0.7 and is equal to about 7.55%. Thus, the BPA was improved by about 6.4% over the preceding model.

4.2.2 Average model-based optimization

In the second approach, the prediction accuracy is optimized on the basis of APA. As seen in Table 4, the APA varies little with a variation in TT. The APA has the same value at four TTs, 0.02, 0.04, 0.06, and 0.08. The average model at 0.02 was selected. The corresponding BPA is 11.78%. Next, the HN was varied while setting the TT at 0.04. The other factors were set to their default vales. As contained in Table 5, the APA is optimized at 9 HN and the corresponding BPA is 10.73%. Table 9 contains the effects of the other training factors. In Table 9, the IWD was varied

Table 10. Combined model-based prediction accuracy.

IWD		GBS		GL	
APA	BPA	APA	BPA	APA	BPA
13.86	12.97	16.37	8.90	13.85	11.15
13.86	12.60	15.02	9.14	13.88	10.66
13.88	11.96	14.40	10.01	13.94	10.47
13.93	10.56	14.02	10.22	14.02	10.22
14.02	10.22	13.91	11.75	14.27	10.40
14.26	10.18	13.87	12.05	14.52	10.37
14.56	9.94	13.88	12.36	14.80	10.01
14.99	9.76	13.89	12.78	15.10	9.56

from ± 0.4 to ± 1.8 by 0.2. The experimental ranges for the other two factors, GBS and GL, are identical to those in Table 7 and Table 8, respectively. For variations in IWD, as seen from Table 9, the APA is optimized at ±0.4 and ±0.6. Here, both TT and HN were fixed at 0.04 and 9, respectively, while setting both GBS and GL to 1.0. The model obtained at ±0.6 was selected and the corresponding APA is 13.78%. For variations in GBS, the model has the smallest APA at 1.0, which is equal to about 13.78%. This is identical to that obtained in the preceding case. For the variations in the other GL, the smallest APA is almost identical to the preceding one. In this sense, both gradient effects on the APA of the model were insignificant. From this analysis, the IWD played the most significant role in improving the APA. Compared to the APA (19.23%) of the optimized model with the first approach, that optimized with the second one yields about 28.3% of improvements. In contrast, the BPA of the corresponding model for the second approach is considerably larger than that (7.55%) for the first approach. This means that the first approach is better than the second one from the standpoint of the BPA.

4.2.3 Combined model-based optimization

The third approach is based on the minimization of the sum of the APA and BPA. From Table 4, the sum is minimized at 0.02 TT and is equal to 25.71%. For variations in HN, the smallest sum is obtained at 8 HN and is equal to about 24.50%. For the other three training factors, the APA and BPA are contained in Table 10. For variations in IWD, as calculated from Table 10, the sum is minimized at ±1.2, and it is about 24.24%. Compared to the preceding model, the sum is slightly decreased. Here, the HN was fixed at 8. As represented in Table 10, the minimized sum for variations in either gradient is the same as that obtained with the variation in IWD. Both APA and BPA of the finally determined model based on the sum are 14.02% and 10.22%, respectively. From the standpoint of the BPA, the optimized model with the third approach is superior to that for the second approach, but inferior to that for the first approach. It is therefore concluded that the first approach is the most appropriate when trying to achieve a predictive model with the highest BPA.

5. CONCLUSIONS

A new metric for profile uniformity was defined by using a discrete wavelet transform. Compared to the conventional metric, the DWT-based metric is advantageous in that it can characterize complex variations along the etch profile surface in detail. In conjunction with the neural network, a predictive model of profile nonuniformity was constructed. The effects of training factors were optimized in three ways. More performance was examined with respect to the average and best prediction accuracy. The first best model-based optimization method was identified to be the most appropriate way in achieving a predictive model with the highest predictive ability. The constructed model can facilitate physical interpretation for profile nonuniformity as well as its optimization.

REFERENCES

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

[2] B. Kim, J. Sun, C. Choi, D. Lee, and Y. Seol, "Use of neural networks to model low temperature tungsten etch characteristics in high density SF_6 plasma," *J. Vac. Sci. Technol. A*, vol. 18, no. 2, pp. 417-422, 2000.

[3] 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.

[4] W. S. Choi, M. T. Lim, and B. Kim, "Wavelet detection of etching profiling surface," *Internat'l conf. Elect. Eng. 2002*, pp. 1905-1908, July, 2002.

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

[6] S. G. Mallat, "A theory for multi-resolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674-693, 1989.

[7] L. Prasad and S. S. Iyengar, *Wavelet Analysis with Applications to Image Processing*, CRC Press, 1997.

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

Won Sun Choi received the B.S. and M.S. degrees in Electrical Engineering from Korea University in 2001 and 2003, respectively. He is now part of the technical staff in the Division of Large Scale Integration at Samsung Electronics. His research interests include neural network applied semiconductor manufacturing.

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 Member at Hyundai Electronics in 1996 and a Faculty Member at Chonnam National University until 2000. He is now an Associate Professor in Electronic Engineering at Sejong University, Korea. His research interests include neurofuzzy modeling, optimization, diagnosis, and control of semiconductor processes.

Myo Taeg Lim received the B.S. and M.S. degrees in Electrical Engineering from Korea University, Seoul, in 1985 and 1987, respectively. He received the M.S. and Ph.D. degrees in Electrical Engineering from Rutgers University, U.S.A., in 1990 and 1994, respectively. Since 1996, he has served as an Associate Professor at the School of Electrical Engineering at Korea University. His research interests include modeling and control of complex systems.