Proceedings of the 13th
KACC, October 1998

Modeling the Properties of the PECVD Silicon Dioxide Films Using Polynomial Neural Networks

Seung-Soo Han and Kyung-Bin Song
School of Automitive Engineering
Catholic University of Taegu-Hyosung
(Office) +82-53-850-2721
(FAX) +82-53-850-2799
Email: sshan@cuth.cataegu.ac.kr

Abstract

Since the neural network was introduced, significant progress has been made on data handling and learning algorithms. Currently, the most popular learning algorithm in neural network training is feed forward error back-propagation (FFEBP) algorithm. Aside from the success of the FFEBP algorithm, polynomial neural networks (PNN) learning has been proposed as a new learning method. The PNN learning is a self-organizing process designed to determine an appropriate set of Ivakhnenko polynomials that allow the activation of many neurons to achieve a desired state of activation that mimics a given set of sampled patterns. These neurons are interconnected in such a way that the knowledge is stored in Ivakhnenko coefficients. In this paper, the PNN model has been developed using the plasma enhanced chemical vapor deposition (PECVD) experimental data. To characterize the PECVD process using PNN, SiO_2 films deposited under varying conditions were analyzed using fractional factorial experimental design with three center points. Parameters varied in these experiments included substrate temperature, pressure, RF power, silane flow rate and nitrous oxide flow rate. Approximately five microns of SiO_2 were deposited on (100) silicon wafers in a Plasma-Therm 700 series PECVD system at 13.56 MHz.

1. Introduction

In the age of industry, the process engineer must often rely on his/her own experience or on books containing recommended operating parameters for given conditions to operate the highly complex equipments. In order to facilitate more reliable results, robust process modeling methodologies are highly desirable. Accurate process models can enable prediction of process variations over a wide range of input conditions. Efficient and tighter process model often offers a competitive edge in today's complex and competitive semiconductor manufacturing.

The neural network based models have recently been shown by several researchers to exhibit superior performance in both accuracy and prediction capability [1-4]. The basic components of a neural network are processing elements (neurons) and weights (connections). The adjustable weights correspond to biological synapses. The weighted inputs to a neuron are accumulated and then passed through an activation function which determines the neuron's response. How to adjust these weights, called learning or training, are the most important in neural network modeling. Since the neural network was introduced, significant process has been made on data handling and learning algorithms. Currently, the most popular learning algorithm in neural network traning is feed forward

error back-propagation (FFEBP) algorithm. Aside from the success of the FFEBP algorithm, polynomial neural network (PNN) learning has been proposed as a new learning method. The PNN learning is a self-organizing process designed to determine an appropriate set of Ivakhnenko polynomials that allow the activation of many neurons to achieve a desired state of activation that mimics a given set of sampled patterns. These neurons are interconnected in such a way that the knowledge is stored in Ivakhnenko coefficients [5-7]. The activation level of a node is determined by a nonlinear activation function with optimal complexity called a Ivakhnenko polynomial [5].

In this paper, the PNN model has been developed using the plasma enhanced chemical vapor deposition (PECVD) experimental data. Plasma enhanced chemical vapor deposition (PECVD), which is one of the most important process in semiconductor processing, is a process where one or more gaseous species react on a solid surface (substrate) and one of the reaction products is a solid film on that substrate. The reactant gases are introduced into a reaction chamber, decomposed, and finally, reacted at a heated surface to form the thin film. A wide variety of thin films utilized in VLSI fabrication is prepared by PECVD. Based on their potential capabilities for satisfying demanding criteria, specific deposition methods have been developed to form such thin films. PECVD processes are often selected over competing deposition techniques because they offer some advantages such as high purity deposition, a variety of chemical compositions, and good economy and process control.

The film properties of SiO_2 deposited by PECVD are determined by the nature and composition of the plasma, which in turn controlled by the deposition variables involved in the PECVD process. However, due to the complex nature of particle dynamics within a plasma, it is difficult to qualify the exact causal relationship between controllable input factors (such as substrate temperature, chamber pressure, RF power, and gas flow rates) and critical output parameters.

To characterize the PECVD process, SiO_2 films deposited under varying conditions were analyzed using 2^{5-1} fractional factorial experimental design with three center point replications [8]. Parameters varied in these experiments included substrate temperature, pressure, RF power, silane flow and nitrous oxide flow. Approximately five microns of SiO_2 were deposited on (100) silicon wafers in a Plasma-Therm 700 series PECVD system at 13.56 MHz. Data from these 19 experiments was used to train polynomial neural networks to model the following critical film properties: deposition rate, refractive index, permittivity, film stress, uniformity, silanol concentration, and water concentration.

2. Experimental Design

To obtain the necessary data for modeling the PECVD process, a fractional factorial experimental design was employed. This design consisted of a 2^{5-1} fractional factorial experimental design and three center points [8]. Silicon dioxide films were deposited in a Plasma-Therm 700 series batch reactor using nitrous oxide, 2% silane in nitrogen, and nitrogen as feed gases. Approximately five microns of SiO_2 were deposited on 4″ diameter (100) oriented silicon wafers. The deposition conditions were varied according to the ranges shown in Table 1. The PECVD system was operated 13.56 MHz, with an electrode spacing of 2.29 cm.

After deposition, a Metricon 2010 prism coupler used to determine the thickness and index of refraction of the film on the wafer. These were measured at five points around the wafer to examine film uniformity. A Flexus F2320 was used to measure the change of the radius of curvature of the bare silicon substrate due to the stress in the silicon dioxide film. A Perkin-Elmer 1600 FTIR was used to obtain the infrared spectra, which were used to measure the impurity content of the films.

Table 1 Deposition parameters

Parameter	Range
Substrate Temperature	200 - 400 ℃
Pressure	0.25 - 1.8 torr
RF Power	20 - 150 watt
2% SiH_4 in N_2 Flow	200 - 400 sccm
N_2O Flow	400 - 900 sccm

The spectrum of each substrate was obtained before deposition. Afterwards, the substrate spectrum was subtracted from that of the deposited film. The silanol and water concentration of the films were determined from the infrared absorbance bands at 3650 and 3330 cm^{-1} using the following relationships [9]:

$$S=(179A_{3650}-41A_{3330}) \quad (1)$$

$$W=(-14A_{3650}-89A_{3330}) \quad (2)$$

where S is the silanol weight percent, W is the water weight percent and A_n is the optical density per micron of film at a wave number of n.

Parallel-plate capacitors were fabricated to evaluate the film permittivity. First, a thin layer (20 Å) of titanium was sputtered onto blank silicon wafers in a PAL 68000 DC sputter. A layer of gold (1500 Å) was sputtered on to the first titanium layer and followed by another 20 Å titanium layer. The titanium layers oxidize rapidly, and the titanium oxide ensures adhesion between the silicon, gold, and silicon dioxide layers. The silicon dioxide films were deposited on that layer. After deposition, another titanium-gold-titanium layer was sputtered on top of the SiO_2 film, followed by masking and etching to make the parallel capacitors. A Keithley 590 CV analyzer and a HP 4275 LCR Meter were used to measure the capacitance. Permittivity was calculated from these measurements using:

$$\varepsilon' = (Cd)/\varepsilon_0 A \quad (3)$$

where ε' is the relative permittivity, C is the capacitance, d is the film thickness, ε_0 is the permittivity of free space and A is the capacitor area. The measured capacitance was corrected for fringing using:

$$C/P=0.041k-0.077\log t+0.045 \quad (4)$$

where C is the fringing capacitance in pF, P is the perimeter of the capacitor plate, k is an approximate value of the dielectric constant, and t is the dielectric thickness in meters.

3. Polynomial Neural Network Process Modeling

The PNN is a network transformation of $R^n \rightarrow R$, as shown in Figure 1. The PNN learning is a self-organizing process designed to determine an appropriate set of Ivakhnenko polynomials that allow the activation of many neurons to achieve a desired state of activation that mimics a given set of sampled patterns [5]. These neurons are interconnected in such a way that the knowledge is stored in Ivakhnenko coefficients. The activation level of a node is determined by a nonlinear activation function with optimal complexity called a Ivakhnenko polynomial. This function usually has a form such as;

$$y=A+\sum_{i=1}^{n} B_i x_i+\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij} x_i x_j+\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} D_{ijk} x_i x_j x_k+\cdots \quad (5)$$

where x_i, x_j and x_k are nodal input variables, and y is the output of an individual neuron. A, B_i, C_{ij} and D_{ijk} are the coefficients of the Ivakhnenko polynomial. This activation function endows the PNN with the ability to generalize with an added degree of freedom not available in statistical regression techniques.

The PNN utilizes the Group Method of Data Handling (GMDH) for data management. GMDH

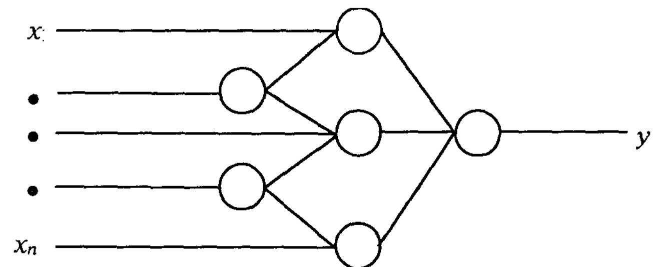

Figure 1 The basic PNN configuration

was first introduced by the Russian cyberneticist Ivakhnenko [5], and has been used to synthesize the building blocks of modeling methodology [7]. In GMDH, the data sample is divided into training and validation sets. The PNN methodology is implemented on the random input-output sets of measured training and validation data obtained from the experiments. An input matrix, consisted of measured (first layer) or calculated (second layer and more) input data, is presented to the each layer, and the output at each node is calculated by combining the inputs of each layer as pairs (or singles or triplets) and filtering this, using the least square method. At each layer, new generations of complex equations are constructed from simple forms. Survival of the fittest principle (appropriate thresholds) determines the equations that are passed on to the next layer. Only the best combination of input properties (new variables) are allowed to pass through to the next layer.

The calculated output vector is then compared to the measured output data vector, and the mean squared difference between these two vectors determines the error of this model. The number of layers is increased until the newer models begin to have poorer powers of predictability (PSE_{min}) than their predecessors. This provides the self-organizing feature of the algorithm, leading to models of optimal complexity.

From experimental work, it has been verified that the curve of the smallest values saved has the general shape shown in Figure 2. For the specific curve shown in Figure 2, one would stop the process after m iterations (generations), according to the stopping criterion.

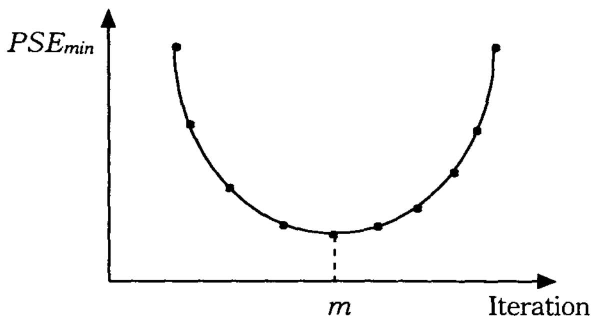

Figure 2 Stopping criterion

4. SiO_2 PNN Modeling and Results

A PNN structure may be effectively employed to model the input-output behavior of the SiO_2 PECVD experiment. Since there were five controllable input parameters and seven measured output characteristics in this SiO_2 PECVD experiment, seven different models were constructed. Output of each model represents one characteristic of the SiO_2 film. The number of inputs of each model were set to five; one controllable PECVD parameter per one input node. The Predicted Squared Error (*PSE*), an estimate of the Mean Squared Error (*MSE*) of unseen data, is used as the performance metric. Initially, PNN PECVD models were obtained using a set of input pairs and default structures. Since these rough models include all the polynomial type regression functions of input variables generated on the PNN structure, they were refined by varying the number of hidden nodes (neurons) and layers by analyzing all the *PSE*s at each layer, and select the polynomial coefficients to minimize the error. In evaluating learning (training) error, 19 PECVD experimental data was used to build the model. The prediction error of this model was determined using the trained PNN to predict the SiO_2 film properties using eight other experiment runs that occurred after the training experiments.

To search for the PNN architecture and coefficients of polynomial that minimized both the training error and prediction error, the following *PSE* was defined as a performance index:

$$PSE = MSE + KP \tag{6}$$

where *MSE* is the training mean squared error and *KP* is the complexity penalty (overfit penalty). The *KP* is defined by:

$$KP = PF \times \frac{2K}{N} \times S_p^2 \tag{7}$$

where K, N, and S_p^2 are determined by the database used to synthesize the network and *PF* is

the penalty factor which is set to 1 in this model. N is the number of training data, S_p^2 is a prior estimate of the true error variance, and K denotes the total number of coefficients. The penalty factor allows the designer to accommodate the computational complexity in the calculation of the error term. This term disallows addition of layers or nodes for small improvements in the predicted output.

The *PSE* takes into account the complexity of the PECVD system, and it attempts to reduce the predicted squared error. Initially, the *MSE* had a significant effect, but as the model increases in the size of training data, *KP* becomes more important. A minimum value of *PSE* will always exist, because *MSE* decreases with each additional trial but always remains nonnegative, whereas *KP* linearly increases as the number of trials increases. Therefore, *PSE* is minimized in the point of compromise of both *KP* and *MSE*. The smallest values of *PSE*s are used to determine the PNN architecture and coefficient set that simultaneously minimize both training and prediction errors. Hence, the presence of both *MSE* and *KP* ensures that *PSE* favors simple models with low error.

In performing this analysis, it has been assumed that the optimal polynomials fitted by linear regression equations were sufficient to capture the dependence of PNN performance on the optimal coefficients. The fitted final polynomials were subsequently used to generate a PNN that optimized the training and prediction errors.

Figure 3 is a synthesized optimal PNN models for the SiOH concentration. The optimal models was determined to find input nodes from the satisfied node, after iteratively calculating the *PSE* until the *PSE* was smaller than the prescribed quantity. The equation shown in the figure represents the first "Triple" in the network. The x_1, x_2, x_3 represent the first input, the second input, and the third input respectively.

The outputs obtained from each of these nodes are combined to obtain a higher-degree polynomial

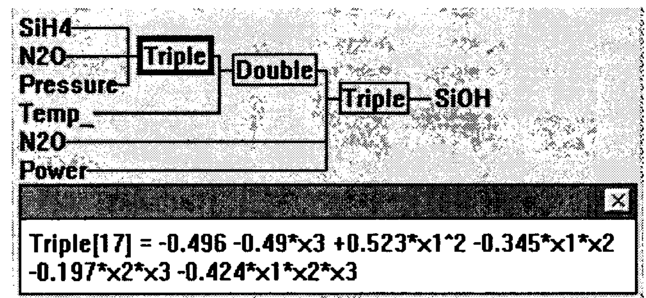

Figure 3 Synthesized PNN for SiOH

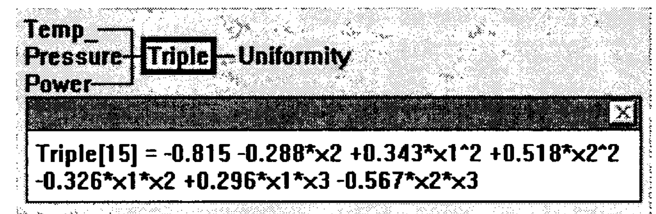

Figure 4 Synthesized PNN for uniformity

model. The PNN leads to self-organizing high degree models with automatic elimination of undesirable variable interactions. We can see the automatic elimination result of the synthesized network from the Figure 4. Although all the five inputs are used to synthesize network for the SiOH concentration (Fig. 3), only three inputs are used to synthesize network for the uniformity.

Figure 4 shows the experimental measured outputs and PNN model outputs. Black diamond represents the real experiemental measurements, and block square represents outputs of final layer obtained from the PNN model of the SiOH concentration. The graph shows that the PNN

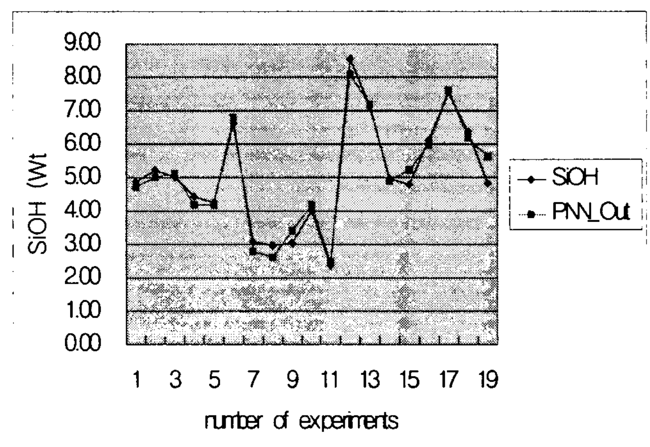

Figure 5 Experimental and PNN output for SiOH

output follows the experimental measurements very well.

5. Conclusion

A PNN modeling technique using GMDH method was investigated in this paper. In contrast to the conventional regression technique, PNN has several distinct advantages: a smaller number of data set is required, less computational time, and the final structure of the PNN does not need to be specified. Although, high-order regression often leads to a severely ill-conditioned system equations, the PNN avoids this by constantly eliminating variables and variable interactions at each layer, and helps to reduce linear dependencies. Therefore, complex systems can be modeled without specific knowledge of the system or massive amounts of data. This optimal network synthesis capability was shown using PECVD SiO_2 film characteristic data. The calculated output of the PNN model follows the real experimental measurement data well. This accurate PNN model will be used not only in analyzing the impact of each input conditions on the output characteristics, but also in synthesizing the recipes for the films with special characteristics.

6. References

[1] C. Himmel, and G. May, "Advantages of Plasma Etch Modeling Using Neural Networks Over Statistical Techniques," *IEEE Trans. Semi. Manufac.*, vol. 6, pp. 103-111, May, 1993.

[2] M. Mocella, J. Bondur, and T. Turner, "Etch Process Characterization Using Neural Network Methodology: A Case Study," *SPIE Proceedings on Module Metrology, Control and Clustering*, vol. 1594, 1991, pp. 232-242.

[3] C. Bose and H. Lord, "Neural Network Models in Wafer Fabrication," *SPIE Proceedings on Applications of Artificial Neural Networks*, vol. 1965, 1993, pp. 521-530.

[4] S. Han, S. Bidstrup, P. Kohl, and G. May, "Modeling the Properties of PECVD Silicon Dioxide Films Using Optimized Back-Propagation Neural Networks," *IEEE Trans. Comp. Pack. and Manufac.*, Part A, vol. 17, no. 2, pp. 174-182, June, 1994.

[5] A. Ivakhnenko, "The Group Method of Data Handling - A Rival to Stochastic Approximation," *Soviet Automat. Contr.*, vol. 13, no. 3, 1968.

[6] J. Farlow, *Self-Organizing Methods in Modeling GMDH Type Algorithms*, New York, Marcel Dekker, 1984.

[7] H. Madala, and A. Ivakhnenko, *Inductive Learning Algorithms for Complex Systems Modeling*, Roca Raton, CRC Press, pp. 1-42, 1944.

[8] G. Box, W. Hunter, and J. Hunter, *Statistics for Experiments*, New York, Wiley, 1978.

[9] W. A. Pliskin in "Semiconductor Silicon 1973," H. R. Huff and R. R. Burgess, Eds., pp. 506-529, *The Electrochemical Society Softbound Proceedings Series*, Preiceton, NJ.