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

Younbum Ryu, Seungsoo Han*, Sungkwun Oh and Taechon Ahn

Department of Control and Instrument Engineering, Wonkwang University
344-2 Shinyong-dong Iksan Jonbuk Korea 570-749
Email : tcahn@wonnms.wonkwang.ac.kr
* School of Electrical and Computer Engineering,
Georgia Institute of Technology
Atlanta, GA 30332, USA

Abstract

In this paper, Plasma-Enhanced Chemical Vapor Deposition (PECVD) modeling using Polynomial Neural Networks (PNN) has been introduced. The deposition of Si_3O_2 was characterized via a 2^{5-1} fractional factorial experiment, and data from this experiment was used to train PNNs using predicted squared error (PSE). The optimal neural network structure and learning parameters were determined by means of a second fractional factorial experiment. The optimized networks minimized both learning and prediction error. From these PNN process models, the effect of deposition conditions on film properties has been studied. The deposition experiments were carried out in a Plasma Therm 700 series PECVD system. The models obtained will ultimately be used for several other manufacturing applications, including recipe synthesis and process control.

I. Introduction

Silicon dioxide films deposited by PECVD are useful as interlayer dielectrics for metal-insulator structures such as Mos integrated circuits and multichip modules. The PECVD of Si_3O_2 in a Si_3H_4/N_2O gas mixture yields films with excellent physical properties. The film properties are determined by the nature and composition of the plasma, which in turn are controlled by the deposition variables involved in the PECVD of silicon oxide films. However, due to the complex nature of particle dynamics within a plasma, it is difficult to quantify the exact relationship between input factors and critical output parameters. In the past few years, several efforts have been aimed at determining such relationships using statistical response surface models [1]. Empirical process models derived from neural networks have recently been shown to exhibit superior performance in both accuracy and predictive capability.

This paper seeks to build upon this body of work, and to obtain accurate and useful manufacturing models for the PECVD process. In order to characterize the PECVD of Si_3O_2 films deposited under varying conditions, we have performed a 2^{5-1} fractional factorial experiment with three center-point replications [6]. Data from these 19 experiments was used to develop PNN process models describing seven output response. The PNNs were trained using predicted squared error.

The development of the optimal PNN model is complicated by means of simple building blocks. These include structural parameters such as the number of hidden layer neurons. In this paper, PNN process models for PECVD were first developed using default parameter set. Afterwards, the effect of these factors on PNN performance was also investigated via a second fractional factorial experiment. The results were analyzed. The optimal parameter sets that minimized the training and prediction error of the PECVD models were determined using the Si_3O_2 PECVD experimental data to examine the silicon dioxide film properties.

II. Experimental Design and Methodology

The silicon dioxide films were deposited in a Plasma-Therm 700 series batch reactor using the feed gases nitrous oxide, 2% silane in nitrogen. The deposition conditions, as shown in Table I, were varied in a 2^{5-1} fractional factorial array. Approximately 5 μm of Si_3O_2 were deposited on 4-in diameter(100) oriented silicon wafers. The PECVD system was operated at 13.56MHz, with an electrode spacing of 2.29cm. The

methodology for film characterization is described below.

A Metricon 2010 prism coupler was used to determine the thickness and index of refraction of films on a silicon wafer. A Flexus F2320 was used to measure the radius of curvature, which was used to calculate the residual stress. A Perkin-Elmer 1600-FTIR was used to obtain the infrared spectra, which were used to measure the impurity content of the films. The silanol and water concentrations of films were determined from the infrared absorbance bands at 3650 and 3330cm^{-1} using the following formulas[1]

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

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

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. The wet etch rates in a solution of 49% HF:H_2O (15 : 1 by volume) were measured using a Dektak 3030 surface profilometer. Parallel-plate capacitors were fabricated to evaluate the film permittivity. A Keithley 590 CV analyzer and a HP 4275 LCR meter were used to measure the capacitances and conductances. The permittivity was then calculated from these measurements. Based on these data, the PNN models were trained and tested. The simulated process outputs were subsequently analyzed.

<Table 1 Deposition Parameters>

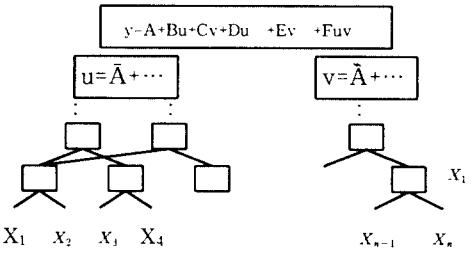
Parameter	Range
Substrate Temperature	200-400 °c
Pressure	0.25-1.8 torr
RF power	20-150 W
2% SiH_4 in N_2 flow	200-400 sccm
N_2O flow	400-900 sccm

III. PNN Process Modeling

The highly complex particle interactions with in a plasma have limited the success of process modeling from a fundamental physical standpoint[1]. Recently, PNN's, as well as neural networks[1] have emerged as a more attractive alternative to physical models and empirical statistical methods. The PNN's, however, are still in the first stage of their research on the semiconductor process modeling. The inherent property of the PNN is to model complex systems using simple building blocks. Namely, the PNN possess the capability of learning (training) arbitrary nonlinear mapping between noisy sets of input and output patterns. The PNN learning is a self- organizing process designed to determine an appropriate set of Ivakhnenko polynomials[1] that allow the activation of many simple parallel processing units to achieve a desired state of activation that mimics a given set of sampled patterns. The PNN's learning(training) capability can be attributed in part to the fact that its architecture crudely resembles that of the human brain. These nodes (called 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;

$$y = A + \sum_{i=1}^{n} B_i X_i + \sum_{n=1}^{n} \sum_{j=1}^{n} C_{ij} X_i X_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} D_{ijk} X_i X_j X_k + \cdots\cdots (3)$$

where X_i, X_j and X_k is nodal input variables, and y is the output of an individual neuron (node). 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 basic PNN configuration is like Figure 1.



<Figure 1. Basic PNN Configuration>

The PNN leads to self-organizing heuristic hierarchical models of high degree with automatic elimination of undesirable variable interactions. In contrast with the conventional regression technique, this scheme has several distinct advantages, a smaller data set is required, the computational time and resources are reduced and the final structure of

the PNN does not need to be specified. In addition, high-order regression often leads to a severely ill-conditioned system of equations. However, the PNN avoids this by constantly eliminating variables and interactions at each layer, and helps to reduce linear dependence. Therefore, complex systems can be modeled without specific knowledge of the system or massive amounts of data.

Since the PNN algorithm is based on approximation (not interpolation), it is not well proper for computer experiments, where exact values of functions are known, but it tends to minimize the average error and can eliminate outlier effects as noise.

Ⅲ. 1 Structure of PNNs

The main concept of the PNN to be utilized for data management purpose is based on the GMDH that was first introduced by the Russian Cyberneticist Ivakhnenko, and was used to synthesize the building blocks of modeling methodology. This approximation technique, i.e., PNN, based on the perceptron principle with a neural network-type architecture is used to model the input-output relationship of a complex process system.

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 and those that are discarded, that is, only the best combination of input properties (new variables) are allowed to pass through to the next layer. The model obtained after each layer is progressively more complex than the model at the preceding layers. To avoid an overfit, the data sample is divided into a) the training set, which is used for the generation of several computing alternative models and b) the testing set, which is used to test the accuracy of the models generated and for the selection of the best models at each layer. This provides the self-organizing feature of the algorithm, leading to models of optimal complexity.

The number of layers is increased until the newer models begin to have poorer powers of predictability than their predecessors. This indicates over-specialization of the system. The final model is an estimate of each performance function y in eqn.(3) as a function of two or three variables, which are themselves functions of two more variables, and so on. The network result is a very sophisticated model from a very limited data set.

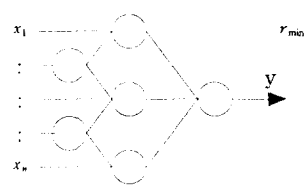
In a PNN technique, a simple form function is usually combined at each node of a polynomial neural network to obtain a more complex form. This function as an approximation represents the current model for the given training and testing sets of input-output data. This approximation is written as a second degree regression equation like eqn.(4) in a case of combining two inputs at each node.

$$y = A + BX_i + CX_j + DX_i^2 + EX_j^2 + FX_iX_j \quad (4)$$

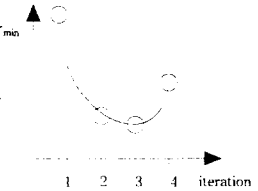
where y is the output and X_i and X_j are the two inputs.

Ⅲ. 2 Process of PNNs

The design of a variable control strategy requires the availability of a reasonable accurate model of the process system. Such models had not been available for the semiconductor process model due to the process dimensionality and the complexity of the interacting physical phenomena. Till now, principle models were almost impossible to develop in process system. In this paper, the input-output data for PECVD processes like the semiconductor system are becoming available and open-up the possibility for robust modeling tools that adopt a modeling paradigm that is based on a PNN. The PNN is a network transformation of R^n -> R functions, as shown in Figure 2.



<Figure 2. PNN Function in Process Model>



<Figure 3. Stopping Criterion>

The nodes (neurons) are low-order polynomials of a few inputs. The PNN synthesis problem may be stated as follows. From a database of I/O pairs, find a structure (site and topology)and its parameters (weight), so as to represent the database but also to

preserve generalization. The Predicted Squared Error (PSE) is used as the performance metric. The PSE is estimated from the mean squared error augmented with a complexity penalty proportional to the error variance of the output in the database. The PSE is calculated for each node of the polynomial neural network and only those nodes which satisfy a threshold criterion are used as inputs to the next layer generating new generations, thus pruning the search and assisting to terminate the network development in a finite number of layers. The best output model, in terms of the PSE, was selected as the final model for the semiconductor process system.

The PNN synthesis activities have focused over the past years on the development of self-organizing, minimal polynomial networks with good generation capabilities. We considered nine types of model equations as shown in Table 2.

<Table 2. The node equations considered for polynomial neural network synthesis>

		Number of Inputs		
		1	2	3
Degree	1	linear	bilinear	trilinear
	2	quadratic	biquadratic	triquadratic
	3	cubic	bicubic	tricubic

trilinear=$w_0+w_1*x_1+w_2*x_2+w_3*x_3$

triquadratic=trilinear+$w_4*x_1*x_2+w_5*x_1*x_3+w_6*x_2*x_3+w_7*x_1^2+w_8*x_2^2+w_9*x_3^2$

tricubic=triquadratic+$w_{10}*x_1*x_2*x_3+w_{11}*x_1^3+w_{12}*x_2^3+w_{13}*x_3^3$

From experimental work, it has been verified that the curve of the smallest values saved has the general shape shown in Figure 3. For the specific curve shown in Figure 3, one would stop the process after n iterations (generations), according to the stopping criterion.

Ⅲ. 3 PNN Optimization

A PNN structure may be effectively employed to model the input-output behavior of the Si_3O_2 PECVD experiment, i.e. to represent the relationship between the input variables of the process and the yields as the output quantities. Since there were five controllable input parameters and seven measured output characteristics in this Si_3O_2 PECVD experiment, the number of nodes (neurons) in the input and output layers were set to ten and seven (one per each output), respectively. In optimizing the PNN models of PECVD process, the number of hidden nodes (neurons) and layer are considered, respectively. Initially, PNN PECVD models were obtained using a set of input pairs (or single or triplets) and default structures. Since these rough models include all the polynomial-type regression functions (of first, second or third orders) of input variables generated on the PNN structure, they were then refined by varying the number of hidden nodes (neurons) and layers according to a 2^4-1 fractional factorial design, analyzing all the PSE's at each layer, and choosing the polynomial values as close as possible to the yields.

In varying the above factors, two important characteristics of PNN process models have been investigated, i.e. learning (or training) ability and prediction capability. In order to compare the PNN with the Neural Networks[1], these performance metrics are quantified in terms of their training (or prediction) mean squared error (MSE), which is given by;

$$\mathrm{MSE} = \frac{1}{N-1}\sum_{i=0}^{N}(yi-\hat{yi})^2 \qquad (5)$$

where N is the number of trials, y_i is the measured value of each response, and $\hat{y}i$ is the corresponding PNN process model estimate. obtained from the least square method. In evaluation learning (training) error, N ranges over the number of training trials used to build the model (in this case, the 19 PECVD experiments). As for prediction error, on the other hand, N represents some number of testing trials apart from the original training set. The prediction error in experiment was determined using the trained PNN's to predict the Si_3O_2 film properties for eight other experiment runs that occurred after the training experiment.

To search for PNN architecture and coefficients of polynomial that minimized both the predicted squared error of training set and the predicted squared error of testing set, the following PSE as a performance index was implemented at all layer of PNN structure to rank and select the better model structure for each of the seven output responses;

$$PSE = MSE + KP \quad (6)$$

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

$$KP = PF \frac{2K}{N} Sp^2 \quad (7)$$

where PF is the user defined penalty factor and is set to 1 as the default value. N is the number of training trials like eqn(5), Sp^2 is a prior estimate of the true error variance that does not depend on the particular model being considered. K denotes the total number of coefficients in the model that are estimated so as to minimize MSE. The complexity penalty term 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, while, at the same time, 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 trials, the KP term becomes more important. Therefore the PSE is optimized in the point of compromise of both the KP and the MSE. As The values of PSE are to determine the PNN architecture that simultaneously minimizes both training and prediction errors.

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. The optimal PNN models for the S_iO_2 PECVD system were determined using least square method, in order to find input nodes from the satisfied node, after iteratively calculating the PSE until the PSE was smaller than the prescribed quantity. Hence the presence of both MSE and KP ensures that PSE favors simple models with low error.

As summarized above, the training and prediction mean squared errors for each of the seven PECVD process models are shown in Table 3.

<Table 3. Error comparison between BPN and PNN>

Pecvd Response	Model Method	Train error	Train improve	Predict error	Predict improve
Ref. Index	BPN	4.69e-3	32.62%	1.99e-2	60.85%
	PNN	3.16e-3		7.79e-3	
Permitti-vity	BPN	0.613	49.76%	0.586	36.52%
	PNN	0.308		0.372	
Wet etch rate	BPN	2674	48.62%	5783	93.60%
	PNN	1374		370	
Stress	BPN	84.6	58.75%	683	83.75%
	PNN	34.9		111	
Deposi-tion rate	BPN	88.8	45.61%	1113	85.76%
	PNN	48.3		158.5	
H_2O	BPN	1.32	41.67%	3.06	46.41%
	PNN	0.77		1.64	
S_iOH	BPN	1.37	37.96%	4.72	63.77%
	PNN	0.85		1.71	
Average			45.00%		67.24%

Ⅳ. Conclusions

The properties of PECVD silicon dioxide films have been modeled using the PNNs. The PECVD process was characterized by varying five controllable parameters in a fractional factorial design. The PNNs were trained and later optimized to predict seven-key PECVD output responses. Overall, the PNN process modeling method is extremely useful and readily applicable to the empirical modeling of such complex plasma processes.

Finally, the PNN model based on a limited data set, predicted well the SiO2 PECVD process yields, Then, an enhanced database is bound to improve the predictive capabilities of the S_iO_2 PECVD model. These results were very encouraging.

Reference

[1] S.S Han, S.A. Bidstrup, and J. May, "Modeling the properties of PECVD silicon dioxide films using optimized back-propagation neural network", IEEE Trans. Comp. Pack. & Manuf., vol.17, pp.174-182, June, 1994.

[2] J. Farlow, Self-organizing methods in modeling GMDH type algorithms, New York; Marcel Dekker, pp. 1-103, 1984.