

IE7860: Intelligent Engineering Systems

Project: Multi-layer Perceptron Networks, Radial Basis Function Networks, and Support Vector Machines

Report Due Dates: To be announced in class.

PROCESS MODELING ASSIGNMENT – REACTIVE-ION PLASMA ETCHING PROCESS:

(Primary Source: Chinnam, Ding, and May 2000)

The task here is to develop empirical models that characterize the behavior of a semiconductor fabrication process, in particular, an ion-assisted plasma etching process, using MLP, RBF, and SVM networks. Plasma modeling from a fundamental physical standpoint has had limited success. Current physics-based models attempt to derive self-consistent solutions to first-principle equations involving continuity, momentum balance, and energy balance inside a high frequency, high intensity electric field. This is normally accomplished through expensive numerical simulation methods that are subject to many simplifying assumptions, and are unacceptably slow. Since the complexity of practical plasma processes at the equipment level is presently ahead of theoretical comprehension, other efforts have focused on empirical approaches to plasma modeling.

In recent years, several groups conducting semiconductor manufacturing research are developing empirical models for predicting process quality for plasma etch processes in terms of etch rate, uniformity, and oxide and photoresist selectivities. The specific objective here is to develop these models using the experimental data collected by Himmel and May [8].

EXPERIMENTAL TECHNIQUE

The study by May, Huang, and Spanos [1991] and Himmel and May [1993] focused on the etch characteristics of n+-doped polysilicon. In that study, etching was performed on a test structure designed to facilitate the simultaneous measurement of etch rates of polysilicon, SiO₂ and photoresist. Test patterns were fabricated on 4-in diameter silicon wafers. Approximately 1.2 mm of phosphorous-doped polysilicon was deposited over 0.5 mm of thermal SiO₂ by low-pressure chemical vapor deposition (CVD). A thick layer of oxide was grown to prevent etching through the oxide by the less selective experimental recipes. Poly resistivity was measured at 86.0 W-cm. Oxide was grown in a steam ambient at 1000°C. One micron of Kodak 820 photoresist was spun-on and baked for 60 seconds at 120°C.

The etching apparatus consisted of a Lam Research Corporation Autotech 490 single-wafer parallel-plate system operating at 13.56 MHz. Film thickness measurements were performed on five points per wafer using a Nanometrics Nanospec AFT system and an Alphastep 200 Automatic Step Profiler. Vertical etch rates were calculated by dividing the difference between the pre- and post-etch thickness by the etch time. Expressions for the selectivity of etching poly with respect to oxide (S_{ox}) and selectivity of etching poly with respect to resist (S_{ph}) are percent nonuniformity (U), and are given below:

$$S_{ox} = \frac{R_p}{R_{ox}} \quad S_{ph} = \frac{R_p}{R_{ph}} \quad U = \frac{|R_{pc} - R_{pe}|}{R_{pc}} * 100$$

where R_p is the mean vertical poly etch rate over the five points, R_{ox} is the mean oxide etch rate, R_{ph} is the mean resist etch rate, R_{pc} is the poly etch rate at the center of the wafer, and R_{pe} is the mean poly etch rate of the four points located about one inch from the edge. The overall objectives are to achieve high vertical poly etch rate, high selectivities, and low nonuniformity. For a detailed discussion of the study or the process, see May, Huang, and Spanos [1991] and Himmel and May [1993].

EXPERIMENTAL DESIGN

Of the nearly dozen different factors that have been shown to influence plasma etch behavior in the literature, the study focused on the following parameters regarded as most critical: chamber pressure (P), RF power (Rf), electrode spacing (G), and the gas flow rate of CCl₄. The primary etchant gas is CCl₄, but He and O₂ are added to the mixture to enhance uniformity and reduce polymer deposition in the process chamber, respectively. The six input factors and their respective ranges of variation are shown in Table 1.

The experiments were conducted by May, Huang, and Spanos [1991] in two phases at the Berkeley Microfabrication Laboratory. In the first phase (screening experiment), a 2⁶⁻¹ fractional factorial design requiring 32 runs was performed to reduce the experimental budget. Experimental runs were performed in two blocks of 16 trials each in such a way that no main effects or first order interactions were confounded. Three center points were also added to check the model for nonlinearity. Analysis of the first stage of the experiment revealed significant nonlinearity, and showed that all six factors are significant [May, Huang, and Spanos 1991]. In order to

obtain higher order models, the original experiment is augmented with a second experiment, which employed a central composite circumscribed (CCC) Box-Wilson design [Box, Hunter, and Hunter 1978]. In this design, the 2-level factorial box was enhanced by further replicated experiments at the center as well as symmetrically located star points. In order to reduce the size of the experiment and combine it with results from the screening phase, a half replicate design was again employed. The entire second phase required 18 additional runs. In total, there were 53 data points.

PROCESS MODELING USING MLP, RBF, AND SVM NETWORKS

The task here is to design and train an "optimal" MLP, RBF, and SVM networks to recognize the inter-relationships between the process input variables and outputs, using the 53 input-output data pairs provided in Appendix A.

Your approach, analysis, and results presentation in the "typed" report, with respect to MLP, RBF, and SVM networks, will be evaluated using the following criteria:

- Were well accepted research procedures used in the analysis?
 - o Justify the chosen loss function and selection of appropriate transfer and kernel functions/parameters.
 - o What pre-processing techniques were employed (e.g., normalization of outputs and inputs) and why?
- Were structured experiments utilized for design, training, and evaluation of the networks?
 - o What is the impact of initialization method on model quality and learning?
 - o How are the hyper-parameters tuned (e.g., grid search, random search)? Report results.
 - o What is the impact of the learning algorithm? Report differences in computational efficiency and result quality.
 - o In instances with multiple outputs, is there a benefit to building a separate model for each output?
- Was generalization demonstrated in the designed networks?
 - o Is model performance effectively and completely evaluated?
 - o What strategies have been employed and why (e.g., k-fold cross-validation strategy, regularization)?
 - o What evidence do you have to claim that bias-variance dilemma has been properly addressed?
 - o If there are class-imbalances, how were they addressed?
- Were results adequately justified and explained?
- Were the results from the MLP, RBF, and SVM models adequately compared?
- The overall quality of the typed report in terms of grammar and organization.

REFERENCES:

- Box, G.E.P., Hunter, W., and Hunter, J., *Statistics for Experimenters*, Wiley Publishing Company, NY: New York, 1978.
- Chinnam, R.B., Ding, J. May, G.S., " Intelligent Quality Controllers for On-line Parameter Design," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 13, No. 4, pp. 481-491, 2000.
- Himmel, C.D. and May, G.S., "Advantages of Plasma Etch Modeling using Neural Networks over Statistical Techniques," *IEEE Trans. on Semiconductor Manufacturing*, vol. 6, pp. 103-111, 1993.
- May G.S., Huang J., and Spanos, C.J., "Statistical Experimental Design in Plasma Etch Modeling," *IEEE Trans. on Semiconductor Manufacturing*, vol.4, pp. 83-98, 1991.

NON-SEPARABLE CLASSIFICATION – PATTERN RECOGNITION IN PROCESS CONTROL CHARTS:

Evaluate the performance of MLP, RBF, and SVM networks on the problem of pattern recognition in process “location” control charts. In particular, the neural network is expected to automate the implementation of the runs rules for control charts based on X-Chart (that plots individual observations across time).

A process is declared to be out of statistical control when:

1. A point falls beyond the control limits, usually set at \pm three standard deviations
2. Two out of three consecutive points fall beyond the same \pm two standard deviation band
3. Four out of five consecutive points fall beyond the same \pm one standard deviation band
4. A run of seven consecutive points falls:
 - a) Above the center line
 - b) Below the center line
5. A run of seven consecutive points falls:
 - a) In a continuous upward pattern
 - b) In a continuous downward pattern
6. Nine out of ten points fall within the \pm one standard deviation band

Evaluation criteria would include the generalization capability of the neural network and the rationale behind the construction of an “optimal” neural network. Number of inputs per pattern is 10 (to accommodate runs rule #6) and number of outputs per pattern is 7 (to accommodate a process in control (first output) and each of the 6 runs rules violations).

Example Patterns:

-----INPUT-----										-----OUTPUT-----						
										IN CONTROL? -----TYPE OF FAULT IF OUT OF CONTROL-----						
-1.691	-2.426	1.440	-2.688	2.240	-2.577	1.541	-0.045	-1.093	-0.498	1.000	0.000	0.000	0.000	0.000	0.000	0.000 (Ho is True)
0.650	-2.725	1.492	-0.981	2.556	-1.943	1.664	-0.958	-1.558	-3.572	0.000	1.000	0.000	0.000	0.000	0.000	0.000 (Rule I Violation)
-0.174	-0.121	1.555	-0.175	1.784	0.934	-0.385	-2.906	-0.644	-2.525	0.000	0.000	1.000	0.000	0.000	0.000	0.000 (Rule II Violation)

The following data sets are available on the course website on Blackboard:

Training data set is made up of 2500 patterns (1000 patterns representing a process in control and 250 patterns representing each of the six rule violations) and the file is labeled “SPC-Training.dat”

Testing data set is made up of 250 patterns (100 patterns representing a process in control and 25 patterns representing each of the six rule violations) and the file is labeled “SPC-Testing.dat”

Your approach, analysis, and results presentation in the “typed” report, both with respect to MLP, RBF, and SVM networks, will be evaluated using the following criteria:

- Were well accepted research procedures used in the analysis?
 - o Justify the chosen loss function and selection of appropriate transfer and kernel functions/parameters.
 - o What pre-processing techniques were employed (e.g., normalization of outputs and inputs) and why?
- Were structured experiments utilized for design, training, and evaluation of the networks?
 - o What is the impact of initialization method on model quality and learning?
 - o How are the hyper-parameters tuned (e.g., grid search, random search)? Report results.
 - o What is the impact of the learning algorithm? Report differences in computational efficiency and result quality.
 - o In instances with multiple outputs, is there a benefit to building a separate model for each output?
- Was generalization demonstrated in the designed networks?
 - o Is model performance effectively and completely evaluated?
 - o What strategies have been employed and why (e.g., k-fold cross-validation strategy, regularization)?
 - o What evidence do you have to claim that bias-variance dilemma has been properly addressed?
 - o If there are class-imbalances, how were they addressed?
- Were results adequately justified and explained?
- Were the results from the MLP, RBF, and SVM models adequately compared?
- The overall quality of the typed report in terms of grammar and organization.

NON-SEPARABLE CLASSIFICATION ASSIGNMENT – MULTI-FONT CHARACTER RECOGNITION:

Evaluate the performance of MLP, RBF, and SVM networks on the problem of multi-font character recognition. The network will be trained on upper-case English letters in selected fonts. The data that will be used for this assignment, consisting of 6 fonts (Courier, New York, Chicago, Geneva, Times, and Venice), was collected and quantized by Lee [1988]. A brief summary of the data collection method is presented here [Logar et al., 1993]:

1. The image of the letter is normalized to an 18 x 18 character matrix, where the line thickness is one and the image is represented by 0's (background) and 1's (foreground).
2. Fourteen properties similar to those proposed by Fujii and Morita (see reference) were extracted from each image. Each property is a 3 x 3 matrix, thus, for each image, a 14 x 9 matrix is generated. This is the X matrix for that image. A property recognition matrix, Y, is constructed for each image and is also a 14 x 9 matrix. It is chosen arbitrarily and is as simple as possible. W is a 9 x 9 filter matrix which maps X to Y and can be found from: $W = X^* Y$, where X^* the pseudo-inverse of X.
3. A 3 x 3 window is moved from upper left to lower right over the character image. The 9 elements in the window are multiplied by the matrix W. If the output matches a row of Y, say row k, the kth place in the count matrix is incremented by the weighting factor of that property.

Thus, the count matrix for a character contains the number of exact template matches, weighted by position. The result is 156 (26 x 6) 14-element vectors. These data sets are provided in Appendix B.

Your approach, analysis, and results presentation in the "typed" report, both with respect to MLP, RBF, and SVM networks, will be evaluated using the following criteria:

- Were well accepted research procedures used in the analysis?
 - o Justify the chosen loss function and selection of appropriate transfer and kernel functions/parameters.
 - o What pre-processing techniques were employed (e.g., normalization of outputs and inputs) and why?
- Were structured experiments utilized for design, training, and evaluation of the networks?
 - o What is the impact of initialization method on model quality and learning?
 - o How are the hyper-parameters tuned (e.g., grid search, random search)? Report results.
 - o What is the impact of the learning algorithm? Report differences in computational efficiency and result quality.
 - o In instances with multiple outputs, is there a benefit to building a separate model for each output?
- Was generalization demonstrated in the designed networks?
 - o Is model performance effectively and completely evaluated?
 - o What strategies have been employed and why (e.g., k-fold cross-validation strategy, regularization)?
 - o What evidence do you have to claim that bias-variance dilemma has been properly addressed?
 - o If there are class-imbalances, how were they addressed?
- Were results adequately justified and explained?
- Were the results from the MLP, RBF, and SVM models adequately compared?
- The overall quality of the typed report in terms of grammar and organization.

REFERENCES:

- Lee, M. and Oldham, W.J.B., "Font Recognition by a Neural Network," *International Journal of Man-Machine Studies*, Vol. 33, pp. 41-61, 1990.
- Fujii, K. and Morita, T., "Recognition Systems for Handwritten Letters Simulating Visual Nervous System," Eds. K. S. Fu, *Pattern Recognition and Machine Learning*, pp. 56-69, New York: Plenum Press.
- Logar, A.M., Corwin, E.M., Oldham, W.J.B., "Performance Comparisons of Classification Techniques for Multi-Font Character Recognition," Report, Computer Science Department, Texas Tech University, Lubbock, TX 79409, 1993.

Appendix A: Dataset for Modeling Reactive-Ion Plasma Etching Process

Inputs							Outputs			
Run	Pressure	RF Power	Electrode Gap	CCl ₄ Flow	He Flow	O ₂ Flow	Etch Rate - R_p Å/min	Etch Uniformity - U (in %)	Oxide Selectivity - S_{ox}	Photoresist Selectivity - S_{ph}
1	300	300	1.8	100	200	20	3491	14.2	6.48	2.01
2	200	400	1.8	100	50	10	3884	3.9	5.98	1.91
3	200	400	1.2	150	200	20	4931	24.8	5.39	1.85
4	300	400	1.8	150	200	20	4726	6.6	5.97	2.11
5	200	400	1.2	150	50	10	5089	12.4	5.61	2.16
6	300	300	1.8	150	200	10	3452	6.5	6.55	2.28
7	300	400	1.8	100	50	20	5164	1	8.51	2.06
8	250	350	1.5	125	125	15	4108	8.9	5.74	1.89
9	200	300	1.8	150	200	20	3494	8.3	6.24	1.32
10	300	400	1.2	150	50	20	5300	0.5	9.64	3.09
11	300	300	1.2	100	200	10	2800	12.7	4.38	1.42
12	200	300	1.2	150	200	10	3396	15.1	4.99	2.09
13	200	400	1.2	100	200	10	4176	41.2	5.01	2.58
14	300	400	1.8	150	50	10	4992	3.8	9.28	2.37
15	200	300	1.8	100	50	20	3684	0.7	6.9	2.08
16	200	400	1.8	100	200	20	2704	27.3	4.41	1.58
17	200	300	1.2	100	200	20	3515	11.6	5.84	1.99
18	300	300	1.2	150	50	10	3895	1.9	13.25	4.17
19	200	300	1.2	100	50	10	3318	22.5	5.47	2.06
20	200	300	1.8	150	50	10	3245	11.6	5.87	2.1
21	300	400	1.2	150	200	10	4112	18	4.58	1.76
22	200	400	1.2	100	50	20	4777	14.6	6.11	1.69
23	200	400	1.8	150	200	10	4229	26.7	5.91	1.78
24	300	400	2.2	100	200	20	4486	55.2	5.85	1.72
25	250	350	1.5	125	125	15	4433	14.2	6.02	1.97
26	300	300	1.2	100	50	20	4453	11.3	11.11	2.89
27	300	300	1.8	100	50	10	3839	2.7	10.55	2.98
28	300	300	1.2	150	200	20	3643	15.1	5.85	2.1
29	200	300	1.2	150	50	20	3723	24.1	6.08	2.17
30	200	300	1.8	100	200	10	2810	15.4	5.52	1.93
31	200	400	1.8	150	50	20	4703	1.8	6.35	1.85
32	300	400	1.8	100	200	10	3141	26.6	2.65	1.42
33	300	300	1.8	125	50	20	4439	2.9	15.15	3.71
34	300	400	1.2	100	50	10	4975	1.2	7.95	2.42
35	250	350	1.5	150	125	15	4500	22.8	5.41	1.99
36	250	350	1.5	125	125	3	4069	6.6	5.77	2.31
37	250	231	1.5	125	125	15	3254	2.9	9.14	3.43
38	250	350	1.5	125	200	15	4106	21.3	6.73	2.12
39	250	350	0.8	125	125	15	5139	4.7	6.61	2.99
40	369	350	1.5	125	125	15	4758	5.3	9.67	2.71
41	250	350	1.5	125	0	15	4700	1.8	9.81	2.82
42	250	350	1.5	125	125	15	4390	13.5	7.72	2.05
43	250	350	1.5	64	125	15	4007	11.4	7.35	2.05
44	250	350	1.5	184	125	15	4643	7.3	6.93	2.28
45	250	350	1.5	125	125	15	4408	5	6.42	2.06
46	250	350	1.5	125	125	15	4579	10.4	7.41	2.28
47	250	350	2.2	125	125	15	4628	8	8.43	2.4
48	250	350	1.5	125	125	15	4703	2.4	7.18	2.27
49	250	469	1.5	125	125	15	5515	6.8	6.41	1.97
50	131	350	1.5	125	125	15	3970	6	6.22	2.07
51	250	350	1.5	125	125	27	4871	7.9	7.82	2.18
52	250	350	1.5	125	125	15	4692	5.4	7.08	2.27
53	250	350	1.5	125	125	15	4768	1.9	7.08	2.27

