wafer exposure patterns of new products in the fab (Fig. 1 is indeed an illustrative output generated by the system).

ACKNOWLEDGMENT

The authors appreciate the generous assistance from the FAB I of the Macronix International Co., Ltd. The authors also wish to thank the referees who provided invaluable suggestions.

REFERENCES

[1] A. V. Ferris-Prabhu, "An algebraic expression to count the number of chips on a wafer," *IEEE Circuits Devices Mag.*, pp. 37–39, Jan. 1989.
[2] H. Dyckhoff, "A typology of cutting and packing problems," *Eur. J. Operation. Res.*, vol. 44, pp. 145–159, 1990.
[3] C.-F. Chien and W.-T. Wu, "A recursive computational procedure for container loading," *Comput. Ind. Eng.*, vol. 35, pp. 319–322, Oct. 1998.
[4] M. Hifi, "A DH/KD algorithm: A hybrid approach for unconstrained two-dimension cutting problems," *Eur. J. Oper. Res.*, vol. 97, pp. 41–52, 1997.

Building Neural Network Equipment Models Using Model Modifier Techniques

Manish Marwah and Roop L. Mahajan

Abstract—**In this paper, we address the problem of developing accurate neural network equipment models economically. To this end, we propose model modifier techniques in conjunction with physical-neural network models. Two model modifiers—difference method and source input method—are proposed and evaluated on a horizontal chemical vapor deposition reactor. The results show that the source input method outperforms the difference method. Further, to develop a model of comparable accuracy, the source input method reduces the number of experimental data points to approximately one fourth of those needed without this approach.**

Index Terms—**Artificial neural networks, chemical vapor deposition, modeling and simulation, semiconductor manufacturing.**

I. INTRODUCTION

One of the outstanding research problems in modeling, optimization, and control of complex manufacturing processes is the development of accurate equipment models that incorporate the underlying physics. For example, in a recent review article by Mahajan [9], it was noted that while several sophisticated physics-based computational models exist for different chemical vapor deposition systems, most of these cannot be directly applied to manufacturing. Many simplifying assumptions are made in developing these models to make them analytically and computationally tractable. As a result, although such models serve as excellent tools in describing the general trends and in delineating the different regimes of transport, they lack the level of accuracy that is generally needed to get the incremental improvement sought by an equipment engineer. A resort is therefore made to develop accurate empirical equipment models. However, for multivariable, nonlinear processes, such as those encountered in semiconductor manufacturing, extensive data points are required to build high accuracy models. This can be quite expensive. Thus, there is a need to develop techniques for building equipment models that incorporate the physics, are accurate, and yet are economical to build.

Manuscript received July 14, 1998; revised January 14, 1999.
M. Marwah was with CAMPmode, Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309-0427 USA. He is now with Qualcomm Inc., Boulder, CO 80303 USA.
R. L. Mahajan is with CAMPmode, Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309-0427 USA.
Publisher Item Identifier S 0894-6507(99)06377-0.

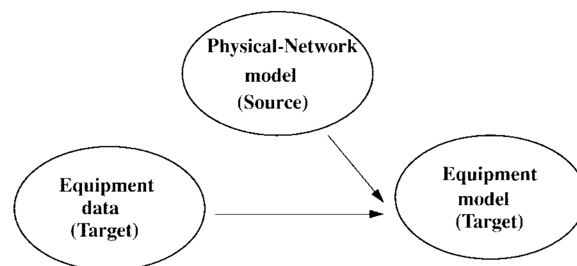


Fig. 1. Model modifier schematic.

A good starting point in realizing this objective is a physical-neural network model. The methodology to build such models is described in a number of papers by Mahajan and co-workers ([6], [10], [11], and [19]). It is shown in those papers that such models are almost as accurate as the physical models, and they additionally have the desirable characteristics of the speed and adaptability of neural networks. If these physical models can now be modified, with a few experimental points, to capture the differences between the behavior predicted by the physical–neural network model and the actual equipment, one can meet the objectives set forth above. In this paper, we explore such model modifier techniques. The underlying idea is to utilize the already existing physical–neural network model (hereafter called the source model) for developing the equipment model (or target model) instead of starting from scratch (see Fig. 1). Furthermore, the source model can be an equipment model which can be used for developing a model of another similar equipment. Recently, Nami *et al.* [8] proposed a hybrid neural network approach in which they use an approximate analytical model to build a neural network model whose output are the undetermined parameters in the analytical model.

While there is a vast body of literature on building neural network (NN) models and their applications, to the best of the authors' knowledge, there is no published work on the model modifier concept. Pratt [17] recently described a technique (which she refers to as neural network transfer) for classification tasks such that the training time for building the target model is reduced. However, that methodology is neither intended nor applicable for developing NN models of processes with continuous output values. Furthermore, the objective there is to reduce the time taken, while in this paper our aim is to reduce the number of data points used for training the target.

Organization of the rest of the paper is as follows. After briefly describing physical–neural network models, two model modifier techniques for building equipment models are proposed. These techniques are then tested for a horizontal CVD reactor. Finally, implications and future applications of this research are summarized.

II. PHYSICAL–NEURAL NETWORK MODELS

First proposed by Mahajan and Wang [11], these models are defined as neural network models trained on first principle physical models. To develop such a physical– neural network model (PNM), the first principles physical model for the selected process is built. Relevant

0894–6507/99$10.00 © 1999 IEEE

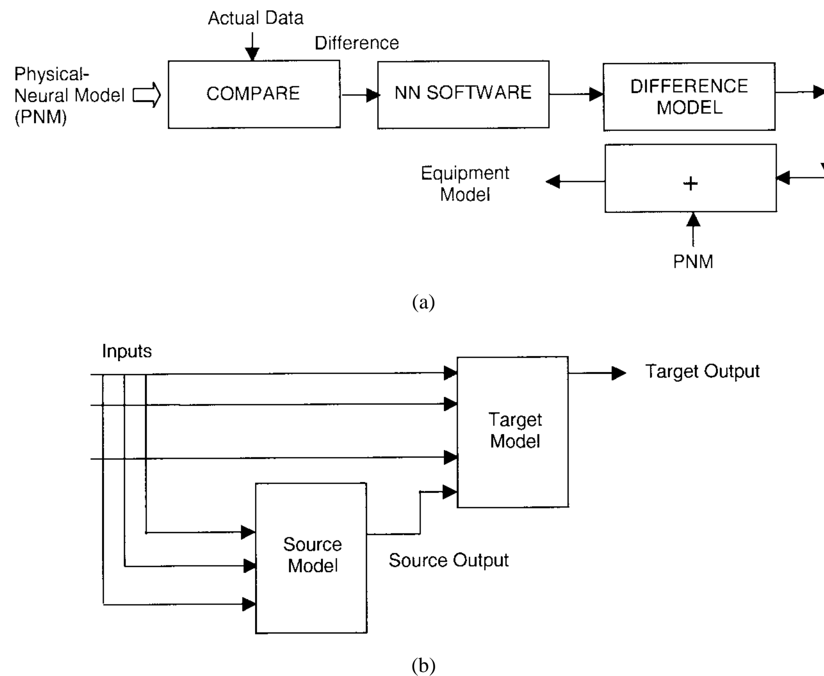


Fig. 2. Schematic of the two model modifiers: (a) difference method and (b) source input method.

input and output parameters are chosen and computer simulations are obtained for combinations of input parameters dictated by a statistical design of experiments (DOE). These numerical data are then split into training and testing data to train a neural network according to our methodology described in [13]. This methodology includes data preprocessing, "simple to complex" network structure approach, and simultaneous training and testing, where training and testing data are identified according to a statistical DOE to capture the underlying input–output relationship. For more details on neural networks and other methodologies for building NN models, the readers are referred to [1], [3], [4], [7], [15], [18], and [20].

The PNM, developed as described above, is then tested against a set of validation points and checked for the desired accuracy. If the target accuracy criterion is not met, the validation points are added to the training set and the process repeated until convergence is achieved. For an illustrated example, see [6]. This process may involve using more data points. However, as the points are obtained from a numerical model, they are not expensive.

III. From Physical–Neural to Equipment Models

The next step in building an equipment model is to update the physical–neural model to capture the specificities of the real equipment. Denoting the physical–neural model as the source model and the equipment neural model to be developed as the target model, we propose the following two modifiers to convert the source model to the target model.

A. Difference Method

In the difference method, a neural network model is trained on the difference between the source and the target data. The expectation is that if this difference is a simpler function of the inputs as compared to the target, a smaller number of target data points might be required to build a good NN model. In some cases, a regression model of the difference may be good enough. The source model plus the difference model then constitutes the equipment model. This approach is shown schematically in Fig. 2(a).

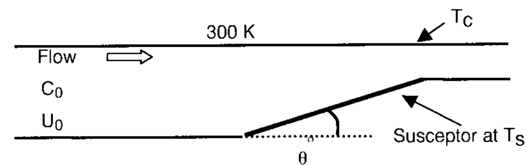


Fig. 3. Horizontal CVD reactor.

B. Source Input Method

In this method, the output of the source model is used as an input to the target network in addition to the inputs used for the source model. The architecture of the target model is shown in Fig. 2(b). Here, our rationale is as follows. Since the source model is close to the target model, the source output is some internal representation of the input data which would be useful to the target network. For the degenerate case, when the underlying target function is identical to the source function, the output of the target network should be the same as that of the source. The learning problem is then reduced to auto- association. Here again, the expectation is that the additional input would make the learning task simpler for the neural network, thereby reducing the number of target data points needed for building the equipment model.

IV. Application to Chemical Vapor Deposition (CVD)

Our test vehicle to test the two model modifier concepts discussed above is a horizontal CVD-epitaxial reactor, see Fig. 3. The graphite susceptor which serves as a support for the silicon wafers is tilted at an angle $\theta = 2.9^\circ$ to the horizontal. A carrier gas, typically H_2, with a small amount of silicon bearing species such as $SiCl_4$, SiH_4, $SiHCl_3$, or SiH_2Cl_2, and trace quantities of the desired dopant are passed over the hot wafers (at $\approx$ 1350 K). The silicon-bearing species diffuses from the bulk flow to the substrate where silicon is deposited by a chemical reaction.

For building our physical–neural network model, we used the study of Mahajan and Wei [12] who, using a finite element formulation, numerically solved the governing equations of transport and presented

TABLE I
TARGET NN MODELS

	Method of Model Development	Training $\overline{\mathcal{E}}_{\mathrm{Rel}}(\%)$	Testing $\overline{\mathcal{E}}_{\mathrm{Rel}}(\%)$	Validation $\overline{\mathcal{E}}_{\mathrm{Rel}}(\%)$	NN Structure
1	Without using source model (from scratch)	7.6	16.39	14.62	4-4-1
2	NN model of difference	1.86	1.59	12.85	4-6-1
3	Linear regression model of difference	—	—	16.90	—
4	Quadratic regression model of difference	—	—	12.96	—
5	Source as additional input	0.3	7.16	2.58	5-4-1

results for temperature, flow, and concentration fields for both the horizontal and the tilted surfaces. The growth rate (output) $\dot{m}$ was calculated as a function of various input parameters. The input parameters are the inlet silane concentration (represented by partial pressure) (C_0), inlet velocity (U_0), susceptor temperature (T_h), and downstream position (x).

Here we use 98 data points computed for a tilted susceptor ($\theta = 2.9^\circ$) with the following input parameter ranges: C_0 from 318.5 to 957.5 Pa, U_0 from 17 to 51 cm/s, T_h from 1250 to 1450 K, and x from 0.675 to 27.5 cm. For the CVD process, keeping uniform growth along x is more important; therefore, the data are given at 14 positions of x, but only at three levels of C_0, U_0, and T_h.

A. Developing Physical–Neural (Source) Model

The source NN model was built using 98 data points obtained from the numerical simulation [12]. Seventy-three of these data points were used for training while 25 points were used for testing. Our NN modeling methodology, explained in detail in [14], was used to build the neural network model. A network with the input layer containing four neurons, one hidden layer with five neurons, and one output neuron (4–5–1) produced the best results. The average relative error ($\overline{\mathcal{E}}_{\mathrm{Rel}}$) for training and testing data was 1.55 and 1.65%, respectively, where

$$\overline{\mathcal{E}}_{\mathrm{Rel}} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\hat{y}_i - y_i}{y_i} \right)^2 \tag{1}$$

where $\hat{y}_i$ is the predicted output, y_i is the actual output, and N is the number of data points.

B. Equipment (Target) Data

The data for the actual process was generated from an experimentally based correlation derived in [2] and is given below

$$\dot{m} = 7.23 \times 10^6 \frac{D_0 T_s P_0}{R T_0^2 \delta(x)} \exp\left[-\frac{2 D_0 T_s T_m}{49 T_0^2 \tan\theta} \cdot \left(\delta(0) - \delta(x) + 0.2 \ln \frac{\delta(0)}{\delta(x)} \right) \right] \tag{2}$$

where $\dot{m}$ is the deposition rate in micrometers per minute, D_0 is the diffusion coefficient of silane at 300 K (0.2 cm^2/s), T_m, T_0, and T_s refer to the gas temperature in the reactor, ambient temperature (300 K), and the susceptor temperature respectively, p_0 is the partial pressure of silane at the inlet of the reactor, R is the gas constant (8.31 $\times$ 10^7 erg K^{-1}) and $\delta(x)$ is the boundary layer thickness at downstream distance x and is given by

$$\delta(x) = \frac{7}{\sqrt{U_T(x)}} - 2 \tag{3}$$

where U_T is the mean velocity as a function of x.

A central composite designed experiment [16] was deployed to identify the target data. Accordingly, 25 data points (4 corner points, 2 $\times$ 4 star points + 1 center point) were used for both training and testing. For validation, we used 98 points corresponding to the input settings used for developing source model. The output for all of these points was calculated using (2). In a practical situation, it may not be possible to use such a large validation set. However, here we use a large set to achieve a high degree of confidence in our results.

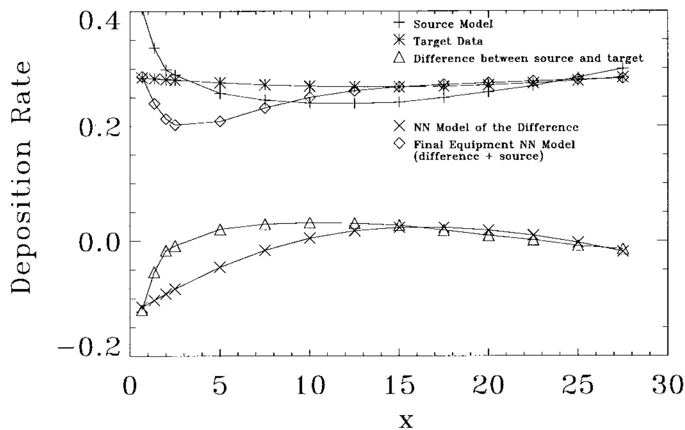
C. Building Equipment (Target) Model

Table I summarizes the results of the target model built using the two model modifier techniques discussed in Section III. In all these models, only 25 central composite data points were used for both training and testing. Row 1 shows the errors when no help is taken from the source model in building the target NN model. The relative errors for both the training and the testing data are high, suggesting the insufficiency of data points for model development. The relative error on the 98 validation points is equally high, 14.62%, which is even greater than the relative difference between the source and the target models (13.84%). The clear indication is that when no use is made of the existing source model, starting from a smaller set of target data points results in inaccurate mapping.

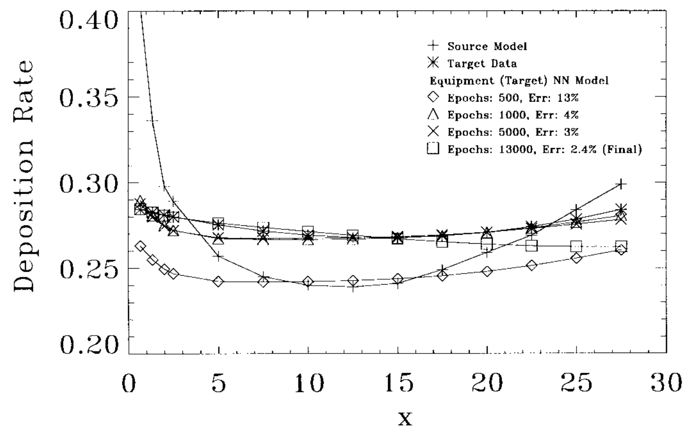
Rows two, three, and four of Table I list the results with the difference method. As mentioned in Section III-A, here the difference between the source and the target is modeled. Three cases are considered. In row two, the results are for a neural network trained on the difference, while in rows three and four, the results are for a linear and a quadratic regression model fitted to the difference, respectively. Although the performance with the difference neural network model is slightly better than with the quadratic model and significantly better than with the linear regression model, the validation error is still too large for claiming any significant advantage through the use of source model. This is not entirely unexpected. The underlying assumption in the difference method is that the difference output is a simpler function of the inputs. This may not always be true, as seen in the bottom part of Fig. 4(a), which shows a plot of the difference deposition rate $\dot{m}$ (target $\dot{m}$ − source $\dot{m}$), as a function of downstream distance x. Clearly, the "difference" deposition rate is as complex a function as the source. As a result, no simplification is offered by using the difference approach. The final model (equipment NN model), see the top part of the figure, is thus not very accurate since only 25 data points were used for building the difference NN model. Our expectation is that the difference approach would work better in situations where one of the important input parameters might have been missed in the original source model. Then, the difference data might reveal a simpler output–input relationship.

The best model is obtained when the source output is used as an additional input to the target model, see row five. The $\overline{\mathcal{E}}_{\mathrm{Rel}}$ for validation data in this case is 2.58% which is comparable to that of the source model (1.6%). The progression of training for this method is shown in Fig. 4(b). Here the curves are plotted at 500, 1000, 5000, and 13 000 (final model) epochs. The error on the 98 validation points continually decreases with training. Further, from the plot of the $\dot{m}$ versus x, we see that after only 500 epochs, the curve looks similar to the final model. Table II lists the $\overline{\mathcal{E}}_{\mathrm{Rel}}$ versus epochs.

Finally, Fig. 5 provides a composite view of the performance of the two model modifier approaches. The deposition rate is plotted versus



(a)



(b)

Fig. 4. Deposition curves for the source model, the target data, and the NN equipment model using (a) the difference method and (b) the source input method. In case of the source input method, the equipment NN model is shown at different stages of training.

TABLE II
DYNAMICS OF TRAINING OF THE SOURCE INPUT METHOD

Epochs	$\overline{\mathcal{E}}_{\mathrm{Rel}}$
500	13%
1000	4%
5000	3%
13 000	2.4%

downstream distance x. The other inputs are at the central values ($U = 34$ cm/s, $C = 639$ Pa, and $T = 1350$ K). The reader should keep in mind that a five-dimensional plot would be required to truly represent the input–output relation. However, the two-dimensional plots are representative of the overall performance. As discussed above, the source input method clearly outperforms the other two techniques.

V. CONCLUDING REMARKS

Developing accurate equipment models for complex semiconductor manufacturing processes is generally a difficult and expensive task. In this paper, we have presented an economical way of developing an equipment neural network model that combines a physical–neural network model (trained and tested using data from analytical or numerical simulations) with a model modifier technique. Two different model modifier techniques, namely the difference method and

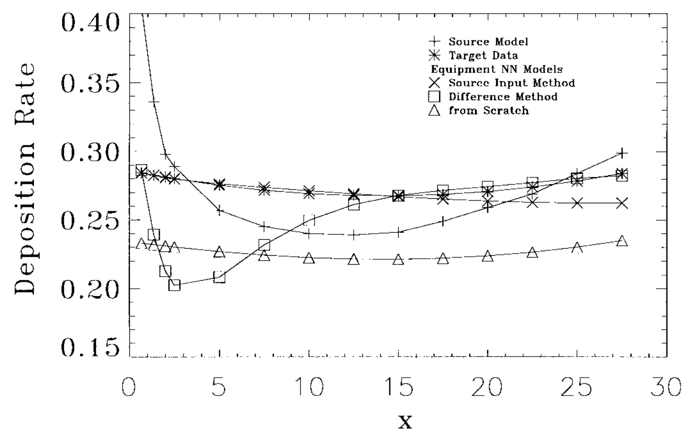


Fig. 5. Comparison of the equipment NN models developed using the source input method and the difference method.

the source input method, were proposed and tested on a chemical vapor deposition horizontal reactor. Both the methods are simple to implement. However, of the two methods, the source input method provided us the better accuracy for the example considered here.

Although in this paper we presented model modifier techniques for economical building of equipment models from physical neural network models, it is noted that the technique is equally applicable for transferring an experimental-data-based neural network model from one equipment to another similar equipment. A common experience in manufacturing is that even though two machines may be from the same equipment vendor, they may not be identical in their performance. Each machine has its own unique "personality," and this thwarts the task of using a universal recipe for all the machines. The methodology proposed here can be utilized to develop unique equipment models for all the machines with relatively few experimental points. We also note that the same approach can be used to transfer models from one to another similar design.

REFERENCES

[1] N. V. Bhat and T. J. McAvoy, "Determining model structure for neural models by network stripping," *Comput. Chem. Eng.*, vol. 16, no. 4, p. 271, 1992.
[2] F. C. Eversteyn, P. J. W. Severin, C. H. J. V. D. Brekel, and H. L. Peck, "A stagnation layer model for the epitaxial growth of silicon from silane in a horizontal reactor," *J. Electrochem. Soc.*, vol. 117, no. 7, pp. 925–931, 1970.
[3] S. S Han, M. Ceiler, S. A. Bidstrup, P. Kohl, and G. S. May, "Modeling the properties of PECVD silicon dioxide films using optimized back-propagation neural networks," *IEEE Trans. Comp., Packag., Manufact. Technol. A*, vol. 17, no. 2, p. 174, 1994.
[4] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan, 1993.
[5] M. Hornik, M. Stinchcombe, and H. White, "Multilayer feed forward networks are universal approximations," *Neural Networks*, vol. 2, p. 359, 1989.
[6] A. S. Kelkar, R. L. Mahajan, and R. L. Sani, "Real-time physico-neural solutions for MOCVD," *ASME-J. Heat Transfer*, vol. 118, p. 814, 1996.
[7] B. Kim and G. S. May, "An optimal neural network process model for plasma etching," *IEEE Trans. Semiconduct. Manufact.*, vol. 7, pp. 12–21, Feb. 1994.
[8] Z. Nami, O. Misman, A. Erbil, and G. S. May, "Semi-emperical neural network modeling of metal-organic chemical vapor deposition" *IEEE Trans. Semiconduct. Manufact.*, vol. 10, pp. 288–294, May 1997.
[9] R. L. Mahajan, "Transport phenomena in chemical vapor deposition systems," *Advances in Heat Transfer*. San Diego, CA: Academic, 1996, vol. 28, pp. 339–415.
[10] R. L. Mahajan and K. C. Gupta, "Physical-neural network modeling for electronic packaging applications," in *Proc. Wireless Communications Conf.,* Boulder, CO, Aug. 19–21, 1996, pp. 157–162.

[11] R. L. Mahajan and X. A. Wang, "Neural network models for thermally based microelectronic manufacturing processes," *J. Electrochem. Soc.*, vol. 140, no. 8, p. 2287, 1993.

[12] R. L. Mahajan and C. Wei, "Buoyancy, Soret, Dufour and variable property effects in silicon epitaxy," *ASME J. Heat Transf.*, vol. 113, p. 688, 1991.

[13] M. Marwah, R. L. Mahajan, and Y. Li, "Integrated neural network modeling for electronics manufacturing," *J. Electron. Manufact.*, vol. 118, p. 148, 1996.

[14] M. Marwah, "Neural network modeling techniques for selected electronic manufacturing processes," M.S. thesis, Univ. Colorado, Boulder, 1996.

[15] J. L. McClelland and D. E. Rumelhart, *Parallel Distribution Processing*. Cambridge, MA: MIT Press, 1986.

[16] D. C. Montgomery, *Design and Analysis of Experiments*. New York: Wiley, 1991.

[17] L. Y. Pratt, "Transferring previously learned back propagation neural networks to new learning tasks," Ph.D. dissertation, Rutgers Univ., Piscataway, NJ, 1993.

[18] M. Smith, *Neural Networks for Statistical Modeling*. New York: Van Nostrand Reinhold, 1989.

[19] G. Subbarayan, Y. Li, and R. L. Mahajan, "Reliability simulations for solder joints using stochastic finite element and artificial neural network models," *ASME J. Electron. Packag.*, vol. 118, no. 3, p. 148, 1996.

[20] P. D. Wasserman, *Neural Computing: Theory and Practice*. New York: Van Nostrand Reinhold, 1989.

Spurious Source/Drain Underlap of Large Junction Area NFET's

Terence B. Hook

Abstract— **When a 0.35-μm CMOS technology was introduced into manufacturing, a small fraction of the tested devices exhibited symptoms of source/drain underlap, despite the fact that all other monitors were well within the design control limits. Additional measurements showed variable overlap on various monitor structures on the same chip. The specific formulation of an HF wet clean was shown to be responsible for the underlapped devices, and the problem was eliminated by altering this process step. High-volume manufacturing data are presented to show the problem and the solution.**

Index Terms—**Semiconductor device manufacture, semiconductor device reliability, semiconductor logic devices.**

I. Introduction

The technology in question is a 3.3-V, 0.35-μm CMOS logic technology utilizing nitrided gate oxide [1], [2]. As part of the basic device design, the minimum allowable overlap of the gate polysilicon to the source and the drain was carefully established by examining the NFET hot-electron shift as a function of overlap capacitance. Provided that the overlap capacitance exceeds a specific minimum value, the hot-electron behavior was well controlled, exhibiting the typical dependence on the substrate current and stress time [3]. Upon introducing the technology into manufacturing and obtaining data on a very large sample, however, a few devices were observed to violate these relationships and demonstrated electrical behavior consistent with insufficient overlap, although adjacent chips and devices on the same wafer behaved normally and the overlap monitors showed no change in capacitance. Extensive characterization of various devices on the same chips showed that only a certain topography was affected; in particular, devices with large regions of open diffused areas were underlapped, while otherwise similar devices with small diffused areas behaved normally. Process experiments established that optimizing a wet clean step eliminated this phenomenon.

Manuscript received August 14, 1998; revised March 19, 1999.

The author is with IBM Corporation, Essex Junction, VT 05452 USA (e-mail: tbhook@us.ibm.com).

Publisher Item Identifier S 0894-6507(99)06380-0.

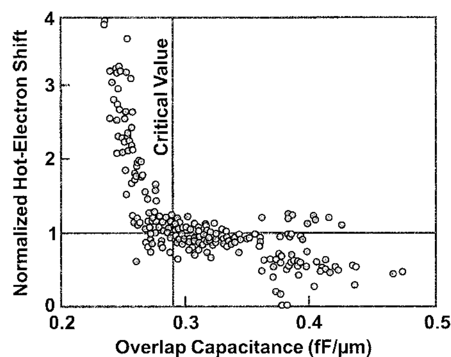


Fig. 1. NFET drain current degradation after hot-electron stress normalized for substrate current plotted against gate-source overlap capacitance. Each point is a single chip.

II. Determination of the Minimal Allowed Capacitance

For improved performance, it is desirable to minimize the overlap capacitance while maintaining adequate reliability. To determine the optimum design point, experimental wafers with different overlap capacitance were produced by modifying the spacer thickness. To assess the reliability of the devices, a large drain bias was applied for a short time, and the drain current before the stress was compared to that measured after the stress. The NFET hot-electron shift thus determined was then normalized to the measured substrate current and is plotted against the overlap capacitance (as measured on a capacitive monitor) in Fig. 1. The data indicate that if the overlap capacitance is less than 0.29 fF/μm, the hot-electron shift is anomalously large. This is a typical symptom of a phenomenon known as "prompt shift" [4], [5] and is indicative of too little doping in the source/drain regions beneath the gate and will also manifest itself in an increase in the resistance of the FET. Another aspect of inadequate overlap manifests itself in an apparent anomalously long effective channel length. Fig. 2 shows the effective channel length of NFET devices as a function of the overlap capacitance. There is a change in slope when the overlap capacitance drops below 0.29 fF/μm; at that point the effective channel length increases anomalously, rapidly exceeding the physical line width. The effective channel length is extracted from a set of devices of varying length [6] and relies on good scaling behavior across the device lengths for a physically realistic value. If the devices are substantially underlapped, then the current on the shortest devices will be unusually reduced relative to the longer devices, and the effective channel length will appear to be very long. For well-behaved device characteristics, it is necessary to avoid an underlapped condition.

III. Manufacturing Data

Having established the process variables (e.g., spacer width, source/drain drive heat cycle, etc.) so that the overlap would never be less than this critical value, and having run wafers in substantial volume, we noted that a small fraction of the scribe line devices showed exceedingly long extracted electrical effective channel length,