Correspondence

The Effect of Dynamic Design Processing for Yield Enhancement in the Fabrication of Deep Sub-Micron MOSFET's

Renate Sitte, Sima Dimitrijević, and H. Barry Harrison

Abstract—With the downscaling of microelectronic devices, tighter process control and more elaborate fabrication equipment need to be complemented by process correcting techniques if good quality and high yields are to be expected. Dynamic Design processing—a forward correcting technique by which some recipe values are recalculated during manufacturing—is such a technique. In this paper the effect of Dynamic Design Processing on deep sub-micron MOSFET's is presented. The results show that a parametric yield improvement in excess of 25% over conventional manufacturing can be achieved.

I. INTRODUCTION

Despite steadily tightened process control and more elaborate fabrication equipment, wafer manufacturing is still afflicted by random fluctuations of equipment and materials, which in turn has an adverse influence on parametric yield [1]. Efforts to reduce these fluctuations concentrate either on equipment, that is improvement and optimization of control settings [2], or on the product itself through forward correction. The concept of forward correction to compensate for these fluctuations has been proposed and embedded in software systems, such as statistical process controllers [3, 4] and statistical process optimization [5, 6]. Dynamic Design Processing (DDP), has been presented to increase yield [7] by counteracting lot to lot fluctuations in a statistically controlled processing environment. It can be used as a stand alone technique, and because it is compatible with any of the previously mentioned software systems, it is also suitable to be used as part of another system [8]. We present here the concept in the framework of a particular example, namely threshold voltage adjustment.

The main concept in DDP is an in-line re-calculation and modification of processing parameters so that the end result brings the product within its intended specification, although not necessarily complying with the original recipes. *Therefore the conditions for the next processing step are set, by taking into account the in-line measurements up to the current step, restoring the likelihood of obtaining the specified device parameters.* An example is the modification of the prescribed dose of an implant for threshold voltage adjustment, so as to correspond to the actual gate oxide thickness and gate oxide charge density (measured before the implantation). It has been shown for micron size devices that the effect of this forward corrective action is *equivalent* to eliminating the fluctuations (variance) of the gate oxide thickness and the gate oxide charge [7].

With current technology it is possible to manufacture 0.5 μm gate devices under tight controlled conditions. However technology will push dimensions to 0.1 μm or less in the future. At these dimensions the statistical fluctuations brought about by equipment and materials have to be virtually eliminated. For practical and financial reasons it will be necessary to complement the control of processing equipment

Manuscript received August 19, 1992; revised June 17, 1993. The authors are with the School of Microelectronic Engineering, Griffith University, Nathan Qld 4111, Australia.

IEEE Log Number 9214302.

Fig. 1. Output characteristics of 0.1 μm n-channel MOSFET. Symbols: measured data [9], lines: MINIMOS simulation.

fluctuations by correction techniques such as DDP, which can *mimic* the effect of tighter process control.

In this paper the effect of DDP on deep sub-micron devices has been studied by simulation. To achieve this a specific example was chosen, namely the process of threshold voltage adjustment. In this example conventional processing, affected by random fluctuations has been compared with its DDP processed counterpart, in which the dose of the threshold voltage adjustment implant is modified, so as to correspond to the actual gate oxide thickness and the gate oxide charge density. To present these procedures and their results, the paper has been organized in the following way: the simulation experiment is explained in the second section of this paper. The third section deals with the processing parameter recalculation. In the fourth section the results of the simulation experiment are analyzed and discussed.

II. THE SIMULATION EXPERIMENT

Few deep sub-micron MOSFET's or circuits have so far been realized [9, 10, 11]. We have chosen the data published by G. Sai-Halasz et al. [9] for our consideration. To study the effect of DDP by simulation MINIMOS [12] has been used as the simulation tool. To assert that MINIMOS is sufficiently accurate and consequently an adequate simulator for such a short channel device, the simulation of I - V characteristics have been compared to the measured I - V characteristics, prior to the DDP experiment. The comparative results are shown in figure 1. As can be seen the agreement between the simulation and the actual data suffices to be safely assumed valid for DDP analysis based on MINIMOS simulations.

The analysis consisted of the determination of threshold voltage distribution, induced by random fluctuations (distributions) of the processing parameters. These random fluctuations represent the average of processing parameter on a wafer or batch. The threshold voltage distribution is obtained by Monte Carlo simulation, one simulation cycle being as follows: random numbers are generated, such that they are normally distributed around specified parameter process values. They represent typical average fluctuations of process parameters, and are input to MINIMOS to simulate the threshold voltage value.

We present here the results of two sets of calculations. In one, representing the standard processing, all the process parameters given

in Table I are randomized¹. In the other, representing DDP, the processing value for the implant dose for threshold voltage adjustment is recalculated on the basis of the actual, measured values of the gate oxide thickness and the gate oxide charge, rather than the recipe values for these parameters. This technique provides better threshold voltage adjustment, ensuring that the gate oxide thickness and charge fluctuations do not influence the threshold voltage. The method employed for the implant dose calculation is explained in the next section. The threshold voltage is then simulated with the same fluctuations for the remaining parameters as they were in the conventional processing case, but with the newly recalculated ion implant dose value. To represent a realistic simulation, the recalculated implant dose was randomized to simulate fluctuations around the new value. The outcome of each cycle are the two threshold voltage values, which can be compared: one corresponding to DDP processing, the other to conventional processing. To obtain a sufficiently large sample size, each simulation set was repeated 1000 times. Figures 2a and 2b show the schematic flow diagram of the simulation procedures: figure 2a simulating conventional processing conditions, and figure 2b with DDP process parameter recalculation.

III. THE PARAMETER RECALCULATION PROCESS

The underlying idea of the recalculation process, is to model threshold voltage by fitting it to a second order polynomial. Because of the complexity caused by many processing parameters, the problem was reduced by selecting three suitable processing parameters to be the independent variables of the polynomial. The variables of our model were the gate oxide thickness, the gate oxide charge, and the implant dose for threshold voltage adjustment. These processing parameters strongly influence the threshold voltage, but the former two are less controllable than the implant dose, which is therefore suitable as a "recovering parameter" for the fluctuations of the previous parameters.

In our case, the fitting was based on a four dimensional model, where the threshold voltage(V_{th}) is expressed as a function of the gate oxide (tox), the implant dose (doi), and the gate oxide charge (nss):

$$V_{th} = f(tox, doi, nss) \quad (3.1)$$

The second order polynomial is then:

$$\begin{aligned} V_{th} = & a_0 + a_1(tox) + a_2(tox)^2 + a_3(doi) \\ & + a_4(doi)^2 + a_5(nss) \\ & + a_6(nss)^2 + a_7(tox)(doi) \\ & + a_8(tox)(nss) + a_9(doi)(nss) \end{aligned} \quad (3.2)$$

where $a_0, a_1, \dots, a_9$ are the polynomial coefficients. To avoid numerical problems caused by the magnitude of the quantities involved, all numbers have to be normalized with respect to their recipe value, before being used in polynomial calculations.

The data needed to determine the polynomial coefficients ($a_1, a_2, \dots, a_9$) by a fitting procedure were generated by simulating the threshold voltage at regularly spaced points in all possible combinations around 2 standard deviations of the recipe for each of the processing parameters involved. The average fitting error was $\delta V_{th}/V_{th} = 0.165\%$, which shows that the second order polynomial is a good approximation.

After fitting and finding the polynomial coefficients (which needs to be carried out only once for a process), the problem is turned

¹It should be noted that the gate oxide thickness was limited to 3 nm, which means that smaller gate oxide thickness were not included in the simulations.

Fig. 2. Schematic flow diagram of one simulation cycle, showing two paths: a) corresponds to conventional processing without recalculation; b) corresponds to Dynamic Design Processing showing process parameter recalculation for the threshold voltage adjustment implant dose. After repeated simulation cycles the distributions of the threshold voltage for both cases can be compared.

TABLE I
PROCESS PARAMETERS USED IN SIMULATION.

Process Parameter	Unit	Mean Value	Std. Dev
Gate oxide thickness	nm	4.5	1.5
Channel length	μm	0.1	0.06
Gate oxide charge	cm^{-2}	5.0×10^{10}	3.0×10^{10}
Substrate concentration	cm^{-3}	7.0×10^{15}	7.0×10^{14}
<i>Threshold Voltage Adjustment Implant (Boron)</i>			
Dose	cm^{-2}	3.0×10^{12}	3.0×10^{10}
Energy	keV	15	0.15
Annealing Temperature	$^{\circ}\text{C}$	875	0.5
Annealing Time	s	2400	1
<i>Source and Drain Extension Implant (Antimony)</i>			
Dose	cm^{-2}	4.0×10^{14}	4.0×10^{12}
Energy	keV	15	0.15
Annealing Temperature	$^{\circ}\text{C}$	875	0.5
Annealing Time	s	1800	1

Fig. 3. The calculated changes in the implant dose for threshold voltage adjustment, needed to compensate variations of the gate oxide thickness and the gate oxide charge density as a second order function.

around and reformulated as a second order equation, to be solved for the implant dose for threshold voltage adjustment. The equation obtained is the model to use for recalculating the implant dose by replacing the correct (desired) and normalized threshold voltage, and the randomly generated values (which in real life are the measured values, normalized) for oxide thickness and oxide charge. In this way the "recalculated" implant dose is obtained and after reverting to its canonical form (un-normalizing) is used as replacement for the corresponding recipe value for the current batch. For the few extreme cases (less than 2%) where no real root can be found, the closest linear approximation is taken, that is, to let the root term equal zero. Figure 3 shows the implant dose for threshold voltage adjustment as a function of the gate oxide thickness and the gate oxide charge density, for the desired threshold voltage. The upper corner of the surface, which appears as folded backwards, corresponds to these less frequent fluctuation combinations where no real root can be found and linear approximation is used instead.

Fig. 4. Comparison between submicron MOS threshold voltages from DDP and conventional processing showing the histogram of simulation data, and fitted Gaussian distributions (sample size $N = 1000$ simulations each), and the upper acceptance quality level (UAQL).

IV. RESULTS AND DISCUSSION

The specified threshold value for this device is 0.213 V. From the simulation runs, the mean threshold voltage for the conventional processing was found to be $\bar{V}_{th} = 0.242$ V, with $\sigma V_{th} = 0.093$ V while for the DDP processed $\bar{V}_{th} = 0.209$ V with $\sigma V_{th} = 0.067$ V. Fig. 4 shows the histograms and fitted Gaussian distributions of both runs respectively. Two effects can be observed from the simulation data: (i) the standard deviation of the threshold voltage is reduced by 56% with DDP, and (ii) the mean value of the threshold voltage is closer, that is less than 2% within the specified value in the case of DDP, while it is shifted by 0.029 V (or 12%) in the case of standard processing.

It is not difficult to explain these effects. The standard deviation is reduced in DDP as the influence of fluctuations in the gate oxide thickness, and gate oxide charge is virtually eliminated through compensation [7]. The shift towards higher threshold voltage values in conventional processing is the result of a non symmetric distribution of the gate oxide thickness around the specified value of 4.5 nm, as can be seen from figure 5. This is caused by an increased likelihood of variations towards thicker oxides than variations towards thinner oxides, which is explained as follows: at these small dimensions the standard deviation is comparable to the mean value; variations towards thinner oxide are physically limited to the thickness of native oxide, however, this does not hold for variations towards thicker oxides, making them more likely to occur. The asymmetric gate oxide distribution is directly reflected on the asymmetric distribution of the threshold voltage, in the case of standard processing. This asymmetry is more pronounced in the case of wide fluctuations, and lessened or eliminated by tightening processing conditions. The asymmetric oxide distribution, however, does not influence the distribution in the DDP case, because the oxide thickness fluctuations are compensated by a properly calculated dose for the threshold voltage adjustment implant. DDP mitigates the demand for stringency in the process control of this otherwise crucial device parameter.

Both observed effects favor DDP processing because of: (a) yield improvement and (b) better fulfilment of the specified (optimized) performance.

To illustrate the yield improvement for our case study, a widely used guideline for the relationship between threshold voltage (V_{th})

Fig. 5. The asymmetric distribution of gate oxide thickness, caused by its physical limit (3 nm): the mean is shifted from the required 4.5 nm towards a higher value.

and the supply voltage (V_{DD}) is taken [13]–[15]:

$$V_{th} \leq 0.25V_{DD}$$

The Considered MOSFET is limited to voltage $V_{DD} = 1$ V [9], because of punch through and hot carrier problems [15]. That gives the threshold voltage limit of $V_{th} \leq 0.25$ V.

It can be seen from the graph of figure 4, that 94% DDP MOSFET's are acceptable (between 0.125 and 0.25 V), while only 63% of standard processed MOSFET's satisfy the above established criterion. The improvement in yield is 31%.

The above mentioned additional benefit from DDP is that not only more MOSFET's are in the acceptance region, but also that, as in our example, the mean threshold voltage is now better centered around 0.213 V. This means that more MOSFET's within the "accepted" class are operating at the specified threshold voltage. It is quality enhancement.

In the light of the favorable results, one might argue that an increase in the dose adversely affect other device parameters, such as drain current and transconductance. We found that the transconductance was almost unaffected by variations far beyond the required changes of the implant dose for threshold voltage adjustment, as depicted in figure 6.

With regard to the accuracy of measuring equipment, and dealing with relatively small fluctuations to measure, the question arises whether the measurement error outweighs the parameter fluctuation itself, specially in the gate oxide thickness case. In general, while systematic errors in measuring can be large for a processing parameter—say, as large as 1 nm for the oxide thickness—it is important to emphasize that the random error is typically quite small (around 0.1 nm for the oxide thickness).

V. CONCLUSION

In this paper we studied the application of Dynamic Design Processing (DDP) and presented specific results for deep sub-micron MOSFET's. Its positive impact on the fluctuations of MOSFET performance can be seen as an additional method to tight process control, which will be required for ultra small size devices. The effect of DDP has been shown to achieve better product uniformity for the threshold voltage by reducing its fluctuation to less than half the value which would be obtained with conventional processing, and to shift

Fig. 6. The transconductance is practically independent of the threshold voltage adjustment implant dose, in a range far beyond the typical fluctuations found in manufacturing.

its mean value by more than 10% closer to its desired value. This means not only better quality, but also an increase of more than 20% in yield.

The implementation, of this technique, requires the recalculation of the corrective processing parameters. For practical applications of the DDP concept, the recipe (static component) is accompanied by a dynamic component: the recalculating equation (obtained by substituting a set of process specific coefficients into a second order equation). While altering some process parameters dynamically might introduce difficulties in the process operation, this is offset by the performance and yield benefits.

REFERENCES

- [1] W. Maly, A. Strojwas and S. Director, "VLSI Yield prediction and estimation: A unified framework" *IEEE Trans. Computer-Aided Design* Vol. CAD-5, pp. 114–10, Jan. 1986.
- [2] R. S. Guo and E. Sachs, "Modeling, optimization and control of spatial uniformity in manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 6, no. 1, pp. 41–57, Feb. 1993.
- [3] E. Sachs, R. Guo, S. Ha and A. Hu, "Process control system for VLSI fabrication," *IEEE Trans. Semicond. Manuf.*, vol. 4, no. 2, pp. 134–144, May 1991.
- [4] A. J. Macdonald, A. J. Walton, J. M. Robertson and R. J. Holwill, "Integrating computer aided manufacturing and process simulation for controlling the production of ASICs," presented at the 8th Australian Conf. on Microelectronics, Brisbane, Australia, July 1989.
- [5] K. P. Mozumder, C. R. Shyamsundar, and A. Strojwas, "Statistical control of VLSI fabrication processes: A framework," *IEEE Trans. Semicond. Manuf.*, vol. 1, no. 2, pp. 62–71, May 1988.
- [6] C. R. Shyamsundar, P. K. Mozumder, and A. J. Strojwas, "Statistical control of VLSI fabrication processes: A software system," *IEEE Trans. Semicond. Manuf.*, vol. 1, no. 2, pp. 72–82, May 1988.
- [7] R. Sitte, S. Dimitrijevic and H. B. Harrison, "Dynamic design processing of integrated circuit fabrication for an On Target end product," in *Process Module Metrology, Control, Clustering, Proc. SPIE 1594*, pp. 92–97, 1991.
- [8] R. Sitte and H. B. Harrison, "Techniques to enhance yield in VLSI fabrication," in *Proc. 9th Microelectronics Conf.*, (The Institution of Radio and Electronics Engineers, Australia) 1990, pp. 141–146.
- [9] G. A. Sai-Halasz, M. R. Wordeman, D. P. Kern, E. Ganin, S. Rishton, D. S. Zichermann, H. Schmid, M. R. Polcari, H. Y. Ng, P. J. Restle, T. H. P. Chang, and R. B. Dennard, "Design and experimental technology for 0.1- μ m gate-length low temperature operation MOSFETs," *IEEE Electron Device Lett.* vol. 8, pp. 463–466, 1987.
- [10] M. Aoki, T. Ishii, T. Yoshimura, Y. Kiyota, S. Iijima, T. Yamanaka, M. T. Kure, K. Ohju, T. Nishida, S. Okazaki, K. Seki and K. Shimohigashi, "Design and performance of 0.1- μ m CMOS devices using low impurity

- channel transistors (LICTs)," *IEEE Electron Device Lett.* vol. 13, no. 1, Jan. 1992.
- [11] M. Iwase, T. Mizuno, M. Takahashi, H. Niiyama, M. Fukumoto, K. Ishida, S. Inaba, Y. Takigami, A. Sanda, A. Toriumi and M. Yoshimi, "High Performance 0.1- μ m CMOS devices operating at room temperature," *IEEE Electron Device Lett.*, vol. 14, no. 2, pp. 51-53, Feb. 1993.
 - [12] "MINIMOS," Tech. Univ. Wien, Inst. Mikroelektronik, Austria 1990.
 - [13] F. M. Klaassen, "Design and performance of micron-sized-devices," *Solid State Electron.*, vol. 21 no. 3, pp. 565, 1978.
 - [14] J. R. Pfister, J. D. Shott and J. d. Meindl, "Performance limits of CMOS ULSI," *IEEE J. Solid State Circuits*, vol. SC-20, p. 253, 1985.
 - [15] J. R. Brews, "The submicron silicon MOSFET," *Submicron Integrated Circuits* R. K. Watts, Ed., New York: Wiley, 1989, p. 20.

Integrated Processing of MOS Gate Dielectric Structures

G. W. Rubloff, M. Offenbergl, and M. Liehr

Abstract—Integrated processing of MOS gate dielectric structures has been carried out in an ultraclean, multichamber processing system as a prototype for cluster (multichamber) processing in order to understand the potential and limitations of cluster processing for this application. With the enhanced surface cleanliness and contamination control enabled in this process environment, etching of the clean Si surface can occur when very low concentrations of oxygen species impinge on the surface at elevated temperatures, an intrinsic consequence of Si-O chemistry. Such etching leads to statistical roughening of the Si surface topography and thereby degrades the electrical properties of subsequently fabricated MOS structures. Identification of the etching reaction leads directly to prescriptions for cluster process integration which prevent etching: (1) ramping up to thermal oxidation temperature in an ambient with sufficient oxygen concentrations; and/or (2) in-situ formation of a passivating oxide surface layer as part of the pre-oxidation cleaning step.

I. INTRODUCTION

The FET gate dielectric structure is one of the most critical elements in advanced CMOS technology because it requires thermal oxide layers well under 10 nm, highly perfect interface and film electrical properties, high reliability, and acceptable yield and very high levels of integration. Manufacturability of the FET gate faces profound obstacles, particularly in contamination and defect density control. Although particulate contamination normally dominates such concerns, reactive impurities (primarily molecular species) may also be a crucial factor in determining the quality and yield of FET gate dielectric structures, especially with thinner structures and new process and tool elements.

Integrated processing, also known as cluster or multichamber processing, has become an important trend in processing tools for

Manuscript received February 8, 1994; revised September 2, 1993. This work has been supported in part by the Office of Naval Research.

G. W. Rubloff was with IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598. He is now with the Center for Advanced Electronic Materials Processing, North Carolina State University, Box 7920 NCSU, Raleigh, NC 27695.

M. Offenbergl was with IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598. He is now with Robert Bosch GmbH, Postfach 1342, 7410 Reutlingen, Federal Republic of Germany.

M. Liehr is with IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, NY 10598.

IEEE Log Number 9214301

manufacturing, driven in large part by their potential advantage in defect density control. The variety of commercially available integrated vacuum processing tools has now embraced processes normally carried out in hot wall batch reactors, either at low pressure (e.g., LPCVD polySi, nitride, oxide) or at atmospheric pressure (e.g., thermal oxidation). This opens the door to integrated processing of FET gate dielectric structures, which requires pre-oxidation cleaning, thermal oxidation, annealing, and polySi deposition.

Process, tool, and process integration issues abound for this cluster sequences, such as wet vs. vapor cleaning, or single-wafer rapid thermal vs. hot wall batch processing. Even before these can be examined, however, it seems prudent to assess the potential and limitations of the conventional process combination when integrated into a cluster tool and exercised under ultraclean conditions. The conclusion from the present exercise is that new chemical reaction pathways may become accessible in advanced cluster processing, and these must be identified and taken properly into account in defining a manufacturable process integration sequence for the cluster.

We have employed an advanced experimental research system, [1], [2] to address the underlying issues in integrated processing of FET gate dielectric structures. This tool includes a variety of process reactors as well as *in situ* analysis chambers, all ultrahigh-vacuum (UHV) so that reactive impurities can be highly controlled to allow a thorough assessment of their role in affecting advanced chemical processes, including integrated surface cleaning, thermal oxidation, annealing, and polySi deposition. The results demonstrate that the etching reaction—in which low oxygen concentrations etch an exposed Si surface and cause roughness—must be prevented by a suitable choice of ramp-up procedure before oxidation (in sufficient oxygen) and/or by formation of an oxide passivation layer before ramp-up [3]–[5].

Following a brief summary of our previous results, in this paper we present direct chemical evidence that the etching reaction is responsible for the systematic degradation of electrical breakdown (for Al gate MOS capacitors), show that the mechanism is also operative when pre-oxidation cleaning, thermal oxidation, and polySi gate deposition are all integrated into a cluster process, and provide prescriptions—with their rationale—for cluster process integration in an ultraclean tool set.

II. EXPERIMENT

In the present studies, a hot wall UHV-based quartz reactor was employed for thermal oxidation and annealing at atmospheric pressure as well as for *in situ* deposition of polySi from SiH₄ at low pressure. The reactor is a vertical batch reactor for a small cassette of four 3.25 in diameter wafers. It is pumped by a turbomolecular pump system to reach $< 1 \times 10^{-9}$ torr base pressure after bakeout, consisting mainly of H₂, CO, and CO₂; the total base pressure with the furnace at oxidation temperature is $\sim 1 \times 10^{-8}$ torr. Purified Ar at 1 atm. was used in some cases during ramp-up to oxidation temperature as well as in annealing processes prior to oxidation. By introducing small concentrations ($\sim 0.1 - 1000$ ppm) of O₂ during preannealing in 1 atm. purified Ar, it was possible to ascertain the role of reactive impurities within the integrated gate dielectric process sequence. The hot wall reactor is connected to UHV transfer chambers ($< 1 \times 10^{-9}$ torr) which permit wafer transfer without air exposure to other tools, including an inert-ambient glove box for pre-oxidation surface cleaning and various surface analysis instruments.