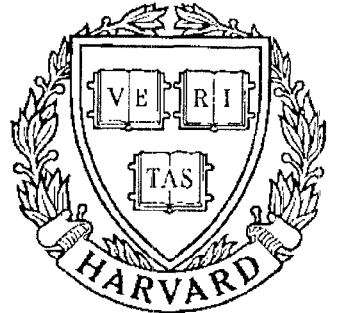


TECHNICAL RESEARCH REPORT



S Y S T E M S
R E S E A R C H
C E N T E R



*Supported by the
National Science Foundation
Engineering Research Center
Program (NSFD CD 8803012),
Industry and the University*

The Impact of CD Control on Circuit Yield in Sub-Micron Lithography

by L. Milor

The Impact of CD Control on Circuit Yield in Sub-Micron Lithography

Linda Milor

Department of Electrical Engineering and
The Systems Research Center
University of Maryland
College Park, MD 20742

ABSTRACT

As tolerance as a percent of feature size increases for sub-micron technologies with increased scaling, yield losses due to circuit performance fluctuations will increase. Therefore for sub-micron technologies a tradeoff has to be made between circuit performance yield and the purchase of more expensive processing equipment that can more tightly control critical dimensions. At the same time, the development time of a circuit that is to be manufactured on a process with higher parameter tolerances will increase, and this has to be traded off with the process development time needed to reduce tolerances. In this paper, the performance yield problem for sub-micron technologies is addressed, as it relates to tolerance in geometric feature sizes and alignment. Using a statistical model of process fluctuations, examples are presented showing that different tolerance requirements are needed for different circuits.

1. INTRODUCTION

No two circuits manufactured with the same design will perform exactly the same because of inherent randomness in the manufacturing process. One major source of yield loss in manufacturing is lithography spots, due to dust and scratches on the masks or on the wafer itself. The impact of lithography spots on yield will depend on the particle densities, their geometrical distributions, and the circuit layout. As feature sizes have decreased with scaling, the size of a dust particle that can eliminate a wire or short two wires together has decreased, resulting in more stringent requirements on the allowed levels of particulate contamination during processing.

The other source of yield loss has to do with randomness in process parameters. This randomness is due to a number of factors, including lithographic resolution, alignment, etching, and contamination in the silicon substrate. Variations in process parameters will result in variations in circuit performance. For example, the delay through a logic gate may increase or decrease, or the frequency response of a filter may change. Large variations in process parameters will result in circuits that fail their performance specifications.

This second source of yield loss has been largely ignored in the circuit design community, especially for low volume designs. This is because predicting circuit performance for industrial circuits by simulation is computationally very intensive, requiring days of CPU time for system components. Determining the impact of process variations on circuit performance requires not one, but multiple circuit simulations, where typically a Monte Carlo [1] or worst case analysis [2] is performed. Hence statistical simulation is currently only computationally feasible for very small circuit components, and good methods of statistical simulation of systems have not yet been developed. Hence from the circuit designer's perspective, it is hoped that variations in the process are sufficiently small so that yield loss due to performance variations is minimal and statistical simulation can be avoided.

However, with the scaling of technology and the escalating cost of fabrication equipment, it has become harder for process engineers to maintain low tolerances on parameters. In fact as we have moved from a 0.8 μ m to a 0.35 μ m technology, the tolerance in gate length as a fraction of gate length has gradually increased for typical processes (Fig. 1). This trend poses several questions. Will circuit designers be able to continue to ignore statistical circuit design? What tolerances are needed on param-

eters so that statistical circuit design can be avoided? What are the most critical improvements in the process needed to reduce variations in circuit performance? How does the set of critical process parameters relate to circuit design, and can we draw some conclusions about specific classes of circuits? This paper is a preliminary study that attempts to address these issues. To do this we will look at the affect of process parameter tolerances on the performance of two prototype circuits, and consequently a statistical model of process fluctuations is required.

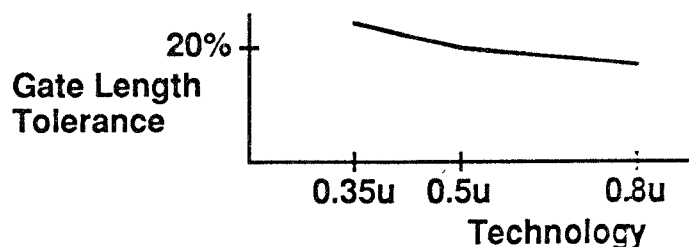


Fig. 1. Gate length tolerance as a function of technology.

A number of statistical models have been developed for the manufacturing process for various technologies. For MOS technologies, in one approach variations are modeled by approximately 50 independent normally distributed variables relating to line-width variations, diffusivity of boron, arsenic, and phosphorus, oxidation, the silicon surface, mobility, etc [3]. A simulator has been developed that maps a set of process controls (temperatures and times) and the random variables to circuit device parameters, which can later be used to simulate circuit performance [4]. The independent random variables in the model are not easily measurable directly, but for a given process they can be extracted using sets of electrical measurements made on devices [5]. A traditional approach to the problem of mapping random process variables to device parameters uses numerical models for each fabrication step to generate a doping profile (SUPREME [6]), which in turn is mapped to an I-V characteristic by a device simulator (MINIMOS [7]). Both the process simulator and the device simulator solve a system of partial differential equations. For a single set of statistical variables, finding a set of device parameters can take hours of CPU time using the traditional approach. To get around this problem, the approach taken in [4] is to use analytical models, and hence statistical process simulation is very efficient. Nevertheless the reliance on analytical models significantly reduces the flexibility and accuracy of the algorithm. And in fact, the models used are not valid for sub-micron technologies.

An alternative approach has been derived by extracting circuit parameters through measurements on fabricated devices. An independent subset of device parameters can be determined by finding the variables that explain most of the variation in the data. In one paper the major sources of variation were found to be due to gate length offset, gate width offset, gate oxide capacitance, and flat band voltage [8]. In another paper, the independent set of primary statistical variables was found to contain gate length offset, gate width offset, oxide thickness, flat band voltage, substrate doping concentration, surface mobility, and lateral junction depth [9]. For both models parameters are assumed to be constant on a chip, but to vary according to a normal distribution between chips. This makes sense for digital circuitry, but since mismatch in devices is not modeled, these models cannot be used for analog circuits. In [10] a two-level extension of the model in [9] is proposed for analog circuits. In this model, each random variable for each device is determined by three distribution parameters: the global mean, the standard deviation on a wafer modeling between chip variations, and the standard deviation within a chip (Fig. 2). The main drawback of the process models developed in [8]-[10] is that they have not been derived for sub-micron processes. Differences can be expected for submicron devices, since they are modeled by a large set of empirical curve fitting parameters not included in earlier models. Furthermore many of these curve fitting parameters are highly correlated. Consider for example the channel mobility. For sub-micron devices channel mobility is not a constant, but varies with position in the channel. At the circuit level, channel mobility is treated as a curve fitting parameter, which has been found to be highly correlated with gate length. Hence for a submicron device, channel mobility and

gate length offset cannot be treated as independent random variables. Clearly a more complex model is needed for sub-micron devices.

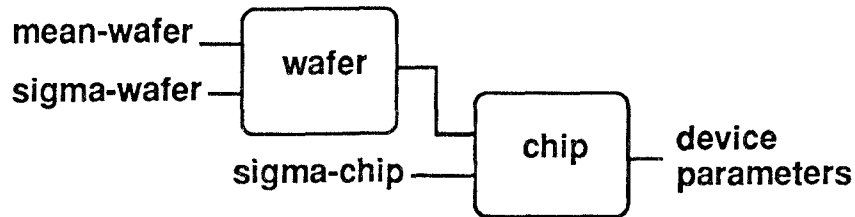


Fig. 2. Two-level model used to generate device parameters incorporating within chip parameter variations.

In this paper we need to develop a statistical process model for sub-micron devices. We will limit the scope of our model to include only those parameters relating to two-dimensional device geometries and alignment. The large set of variables relating to doping profile variations, hot electron effects, oxide properties, etc. are neglected. Therefore we aim to use circuit examples to distinguish between gate length and width offset and alignment errors.

This paper is organized as follows. In the next section we begin by describing a sub-micron device and discuss the major sources of variation in order to derive the statistical model. In Section 3, we present two circuit examples, one digital and one analog, and determine which parameters have the most significant effect on performance. We conclude in Section 4.

2. A STATISTICAL PROCESS MODEL FOR SUB-MICRON DEVICES

The submicron device that is used in our examples is a lightly-doped source/drain (LDD) device with a 0.5μ channel length (Fig. 3). It is designed to be used with a 3.3V power supply and has a threshold voltage of approximately 0.9V. The I-V characteristic for the device has been generated by simulation using UDMFET [11]. UDMFET is a 2-dimensional device simulator, like MINIMOS. It uses an energy transport model to solve the energy balance equations, which is more accurate for devices with channel lengths that are less than 0.8μ , rather than the Drift-Diffusion transport model which is implemented by most device simulators. The I-V characteristic for our device is shown in Figure 4 below.

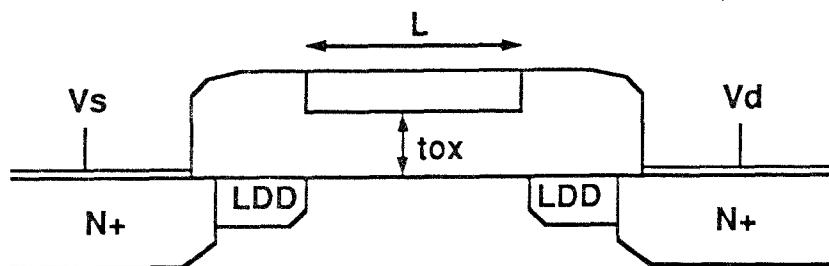


Fig. 3. 0.5μ LDD transistor.

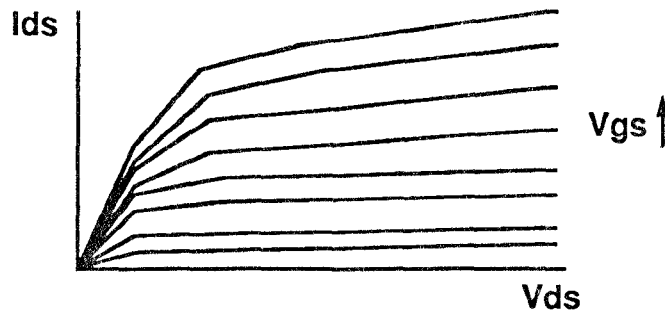


Fig. 4. I-V characteristic for the 0.5 μ LDD transistor.

Most circuit simulators cannot use I-V curves directly for circuit simulation. Instead current equations are fit to the I-V curves. We have used the empirical SPICE level 3 model [12] to simulate circuits containing sub-micron devices. We have obtained the device parameters by specifying some known parameters of the device, like oxide thickness and substrate doping, and then we have used an optimizer to find six fitting parameters: V_{T0} (zero-bias threshold voltage), μ_0 (surface mobility), v_{max} (maximum drift velocity of carriers), η (static feedback on threshold voltage), θ (mobility modulation), and κ (saturation field factor). The fit is almost perfect. Two of these parameters (κ and η) were found to correlate strongly with channel length, and μ_0 is also affected by channel length.

The device is fabricated with a self-aligning process. The transistor gate is formed by depositing polysilicon on a thin layer of oxide and etching it away where it is not needed (Fig. 5). A nitride is then deposited to cover areas that will form the drain, gate and source of the transistor. After this step, an oxide is grown and etched away. However, because the oxide has to step over the gate, it will be thicker on the sides of the gate, and some will be left over after etching (Fig. 6). An implant is then made to form the source and drain. The LDD region is formed by implanting through the oxide, which blocks some of the particles, when the source and drain regions are being implanted, followed by an annealing step which extends the source and drain regions under the oxide. The source and drain diffusions extend under the gate by a small amount (approximately 0.06 μ), reducing the channel length. Variations in the distance that the source and drain regions extend under the gate is a minor source of variation in the channel length. The process is essentially self aligned because the gate boundary is used to form the boundaries of the source and drain, defining the channel length. The major source of variation in gate length comes from the formation of the gate itself. Since the gate is defined by a masking step, the main sources of variation in its length are lithographic resolution and etching. For an optical process the tolerance we have used in our gate length is 0.1 μ . It is not known if the tolerance is due primarily to etching or lithographic resolution, but it is thought that they may have a roughly equal impact. X-ray technology because of its improved resolution is likely to improve the tolerance in gate length.

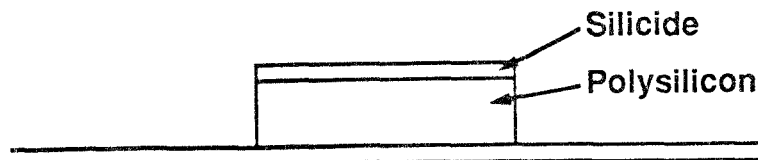


Fig. 5. The transistor gate and gate oxide.

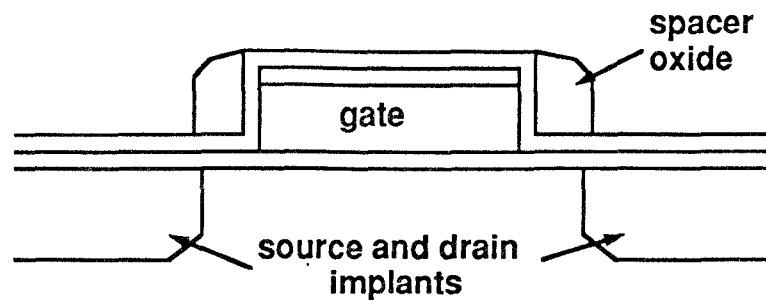


Fig. 6. The transistor channel after implantation and the spacer oxide.

The resistance associated with the source, drain, and LDD spacers degrade device performance. An angular misalignment during etching of the spacer oxide could increase the size of one LDD region, while decreasing the other, increasing resistance on one side of the channel, while decreasing it on the other side. Similarly, an angular misalignment of the implanter could also add resistance to either the source or the drain, while reducing the resistance on the other side. In addition, the gate mask could be misaligned with respect to the source and drain contacts due to overlay misalignment (Fig. 7). This would also increase the resistance on one side of the channel, while reducing the resistance on the other side, keeping the sum of resistances constant. Note however that the LDD region, where the resistance is highest remains self-aligned with the gate. The tolerance assumed for lateral movement of the gate is 0.2μ in either direction, adding 65Ω of resistance to one side, while subtracting it on the other. The variation in resistance is assumed to be independent of variation in gate length. Lithography does not play a role, since the primary source of variation in resistance is misalignment.

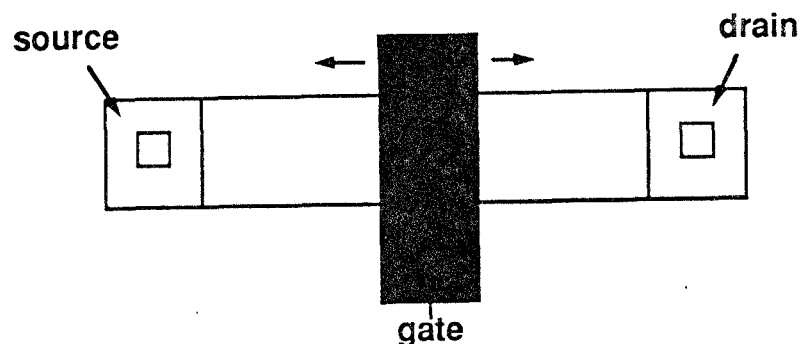


Fig. 7. Gate-contact overlay misalignment.

A final source of process variation relates to the gate width. Some variation in gate width comes from implantation and thermal annealing. This is small compared to the variation in gate width caused by lateral oxidation. A nitride covers the source and drain areas when the field oxide is grown, but some of the oxide grows underneath the nitride, reducing the channel width by as much as 0.4μ per side (Fig. 8). This variation is assumed to be independent of variations in gate length and resistance. As with resistance, gate width variation is primarily not a result of lithography.

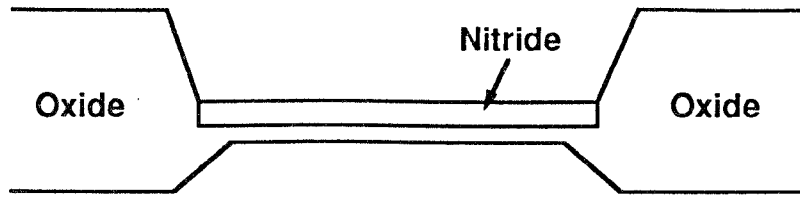


Fig. 8. Channel width reduction by lateral oxidation.

The model that we propose incorporates three independent variables: length offset, width offset, and drain resistance, together with the constraint that the sum of source and drain resistances is constant. We have included variations in parameters between chips and within a chip in our model, by using the two level parameter generator in Figure 2. The within chip standard deviation is assumed to be 15% of the between chip standard deviation. These variations are caused by problems in wafer planarity and stress. Overall, for a circuit containing N devices, the distribution of parameters is defined by a mean of each of the three parameters, the chip-to-chip tolerance for each parameter (which is equivalent to 2σ chip-to-chip variation), and N within chip standard deviations per parameter.

3 CIRCUIT EXAMPLES

Our submicron process model has been applied to two circuits, an off-chip driver and an emitter coupled pair. The off-chip driver is a string of nine inverters, where the first one is minimum sized and each consecutive stage is wider than the last by a factor of three (Fig. 9). The output capacitance that has to be driven is 10pF. The performance that we have studied is the fall time of the output by 99%. The circuit has 18 devices, and 57 parameters were needed to describe the statistical distribution. The parameters with the largest impact on fall time are listed below in Table 1. They were determined by increasing and decreasing each parameter, i , one at a time, computing fall time, and taking the difference, Δy_i . The differences are summed up to find the total change in the performance,

$$\Delta Y = \sum_{i=1}^{57} |\Delta y_i|$$

where $||$ denotes absolute value. The percentage change in Y explained by each variable is computed as follows.

$$\frac{|\Delta y_i|}{\Delta Y}$$

For this circuit the parameters that affect fall time the most are between chip variation in channel length and between chip variation in source and drain resistance. A minor impact can be contributed to the last inverter in the chip, which pulls down the output capacitance. Even a small variation in its length and width had an impact on fall time. Overall the variation in fall time was 20%. From this example it seems that impact of alignment errors and lithography/etching is roughly equal. Improvement in lithography can potentially help, but etching and alignment are as important.

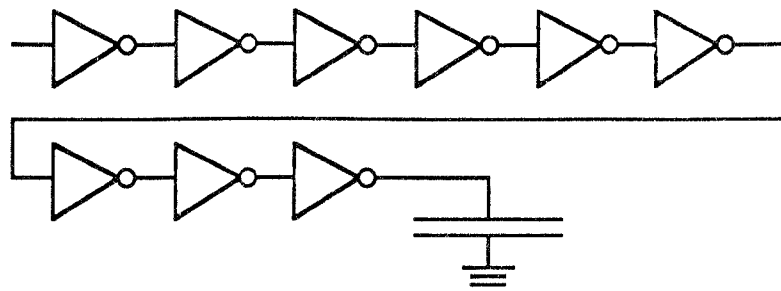


Fig. 9. The Off-Chip Driver.

Table 1

Critical Parameters	Contribution to Total Variation
source-drain resistance - wafer	41%
gate length - wafer	41%
source-drain resistance - T9 (last pull-down transistor)	6%
gate length - T9 (last pull-down transistor)	4%

We also considered an emitter coupled pair (Fig. 10). The circuit contains 6 devices, and therefore the model of the process contains 21 parameters. The performances we considered were gain and CMRR (common-mode rejection ratio). The critical parameters for both performances are shown in Table 2 below. The gain was found to be extremely sensitive to very small amounts of mismatch between the devices. For both performances the parameters that were found to be most sensitive are the local variations in channel lengths of the two input transistors. CMRR was also found to be sensitive to mismatch between the two n-channel transistors in both length and width. Clearly within chip variations in lithography have the strongest impact on performance, and an improvement in tolerance in lithography and in etching can make a difference.

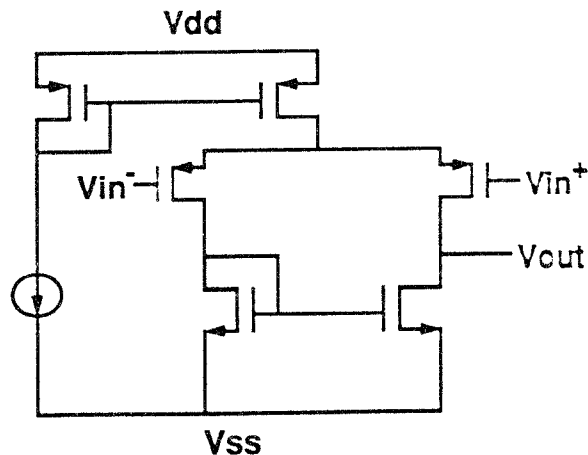


Fig. 10. Emitter Coupled Pair.

Table 2

Critical Parameters	Contribution to Variation in Gain	Contribution Variation in CMRR
gate length - T1 (input transistor)	50%	47%
gate length - T2 (input transistor)	50%	47%
gate length - T3 (n-channel pull-down)		2%
gate length - T4 (n-channel pull-down)		2%
gate width - T3 (n-channel pull-down)		1%
gate width - T4 (n-channel pull-down)		1%

4. CONCLUSION

Two examples have been presented to analyze the impact of gate length, gate width, and source/drain resistance on circuit performance. Using our result for the digital example, it appears that a statistical model neglecting variations within a chip will be sufficient for digital circuits. Furthermore variations in gate length due to lithography and etching seem to be of equal importance. Because this circuit uses very wide transistors, no impact was seen for gate width variations. For a circuit with smaller sized devices, gate width could also have an impact.

The results for the analog component are very different. It was found to be primarily sensitive to within chip variations in gate length. This indicates that improvement in lithographic techniques can have the strongest impact on the yields of high performance analog circuits and components.

5. ACKNOWLEDGEMENTS

This paper would not have been possible without the helpful and many fruitful discussions with Osamu Otani, Greg Wolokin, and Jeff Frey at the University of Maryland, College Park, Eric Boskin at the University of California, Berkeley, David Ma and Marty Peckerar at the Naval Research Laboratory, and Steve Schnur at IBM. The author would also like to thank Osamu Otani for providing the sub-micron device design.

6. REFERENCES

1. J.M. Hammersley and D.C. Handscomb, *Monte Carlo Methods*, Fletcher and Son Ltd., Norwich, 1964.
2. V. Visvanathan, "Variational analysis of integrated circuits," *Proc. ICCAD*, pp. 228-231, 1986.
3. W. Maly, A.J. Strojwas, and S.W. Director, "VLSI yield prediction and estimation: A unified framework," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, pp. 114-130, Jan. 1986.
4. S.R. Nassif, A.J. Strojwas, and S.W. Director, "FABRICS II: A statistically based IC fabrication process simulator," *IEEE Trans. Computer-Aided Design*, vol. CAD-3, pp. 40-46, Jan. 1984.
5. C.J.B. Spanos, "Parameter extraction for statistical IC process characterization," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, pp. 66-78, Jan. 1986.
6. D.A. Antoniadis, S.E. Hansen, and R.W. Dutton, "SUPREM II -- A program for IC process modeling and simulation," Tech. Rep. 5019-2, Stanford Electronics Laboratories, June 1978.
7. S. Selberherr, A. Schutz, and H.W. Potzl, "MINIMOS -- A two-dimensional MOS transistor analyzer," *IEEE Trans. Electron Devices*, vol. ED-27, no. 8, Aug. 1980.
8. P. Yang *et al.*, "An integrated and efficient approach for MOS VLSI statistical circuit design," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, pp. 5-14, Jan. 1986.
9. S. Liu and K. Singhal, "A statistical model for MOSFETS," *Proc. ICCAD*, pp. 78-80, 1985.
10. L. Milor and A. Sangiovanni-Vincentelli, "Optimal test set design for analog circuits," *Proc. ICCAD*, pp. 294-297, 1990.

11. Z. Peng, *UMDFET -- A Two-Dimensional General Device Simulator and its Application in EPROM Analysis*, M.S. Thesis, University of Maryland, 1990.
12. P. Antognetti and G. Massobrio, eds., *Semiconductor Device Modeling with SPICE*, McGraw-Hill, New York, 1988.

