Modeling MOSFET Process Variation using PSP

Josef Watts, Yoo-mi Lee, Jae-Eun Park
IBM Semiconductor Research and Development Center, Systems and Technology Group,
Essex Junction, VT 05452 USA Tel: 802-769-1874 Fax: 802-769-9659 jswatts@us.ibm.com

ABSTRACT

The PSP model has been extensively evaluated for its ability accurately match IV and CV characteristics MOSFETs measured on a single die. This enables accurate prediction of circuit behavior for circuits made of transistors which are very similar to the typical devices which were measured. However during manufacturing variations occur in the process which leads to difference in MOSFET characteristics. We examine various random sources of variation which affect all devices on a chip and attempt to model them by varying the PSP model parameters. We focus on determining a minimum set of parameter distributions for Monte Carlo modeling believing that fixed corner and user defined corners should be derived from accurate Monte Carlo models.

Keywords: PSP, Compact Models, Statistical Models.

1 INTRODUCTION

The term model has a very wide range of meanings. Even in the context of compact modeling the term is used in several ways. The set of physical and empirical equations is called a model. Since this paper deals specifically with the PSP model we will use PSP to mean the model equations. A set of values for the PSP parameters is also called a model. We will use the phrase "parameter set" for such a set of numerical values. The focus of this paper is on modeling the variation of electrical characteristics expected to result from variation in the process during wafer manufacturing. We will use the terms "process model" and simply "model" to mean the parameters and calculations for varying PSP parameters for this purpose.

A complete process model includes a Monte Carlo model; foundry defined fixed corners and parameter for the user to define their own corners. [1] In this paper we address only the Monte Carlo portion of the process model.

REQUIREMENTS

From the point of view of the user, the model is required to represent accurately the limits of electrical behavior of circuits which might be produced by the manufacturing line. If the model underestimates the possible variation the circuit design may fail to "yield", i.e. some "in spec" wafers will have few if any good die. If the model overestimates the variation the circuit may not achieve the best performance possible from the technology.

Since the model must accurately predict variation of any and all circuit characteristics the designer cares about the model builder must ensure that the range of all device characteristics are accurately modeled and that correlations between them is also accurate. Ideally this is done by accurately modeling variation of PSP parameters which have physical meaning and naturally provide correlated variation of PSP outputs. This applies to variation of several characteristic of a single FET and to variation of characteristics of devices of various sizes.

For the model builder operating in an industrial setting there are additional requirements on the model. These come from the need to be able to rapidly build the model using data from a limited set of imperfect test devices combined with device specifications which include some amount of wishful thinking. First the need to quickly build and test the model and quickly integrate it with the rest of the design system means that we need a fixed model design. Just a PSP provides a fixed set of equations and parameters to describe a single point in the process distribution; we require a fixed set of process model equations and parameters to define the process model.

The process model design consists of a fixed list of PSP parameters which will be varied, and the equations by which they will be calculated from either random variables for Monte Carlo simulation or corner parameters for user defined corners. For foundry defined fixed corners it may be necessary to vary a larger set of parameters because these corners are often required to represent several extreme condition at once, although extremes are unlikely to all occur on a single real die.

Following this methodology the modeling engineer at the stage of building the process model does not have the freedom to choose any PSP parameters desired to achieve the required variation of FET characteristics. In general she/he cannot afford time to rework the nominal model to get better response to the PSP parameters that are being varied in the process model. (Although if the response is strongly unphysical refitting may be required.) This means that the process model design must be robust in the following ways. First the electrical characteristics known to vary in the process must be sufficiently sensitive PSP parameters chosen for the process model even when the nominal parameter set is not perfectly physical. Second the process model design must produce a reasonable degree of correlation between FET characteristics of one FET and between FETs of different sizes across a range of values for the variances of individual parameters.

NSTI-Nanotech 2007, www.nsti.org, ISBN 1420061844 Vol. 3, 2007

METHODOLOGY

We begin by identifying what we believe are the main physical variations in advanced MOSFETs. We do this based on knowledge of the manufacturing process and by examining data from detailed measurements on multiple die from multiple lots of a 65nm process. These lots were manufactured at various times over approximately one year. We examine complete curves rather than a collection of points such as Ion and Vt so that we can relate electrical variations to physical process variations and select PSP parameters to vary which result in most realistic simulated devices.

Then based on the equations of PSP we search for model parameters which correspond to the physical effects which cause variation in our data. We begin with poly gate length which for short devices is the most important source of variation. It is also a desirable parameter because it is monitored inline in a physical way so the amount of variation is known without resulting to indirect methods of measurement such as back propagation of variance. [2] We add more parameters using these two criteria, strength of the parameter and availability of direct measurement.

In evaluating the strength of a parameter we need to ensure that the parameter is effective for a variety of different device designs and for variation in the fitting of a device. To evaluate the effectiveness of parameters for a range of reasonable good parameter sets we moved the parameter in question in both directions enough to significantly effect the fit to hardware, adjusted other parameters to roughly restore the fit, then evaluated the strength of the parameter in the new model. We are in the process of fitting several devices for a 65nm technology with PSP, including different oxide thicknesses and different drain engineering. We will use all off these models to verify that our chosen parameters are consistently useful for modeling process variation.

In choosing which parameter to use to model variation of a particular physical effect we look for parameters which have a strong effect and which effect the FET behavior in a qualitatively different way than any parameters currently on the list. Qualitatively different may mean affecting a different device characteristic such as linear vs. saturated threshold voltage. Or it may mean affecting the same characteristic for different geometries such all long channel vs. short channel threshold voltage.

Because we are designing a reusable template for a process model, not a specific process model our stopping criteria is not when the magnitude of the process variation is modeled but when the parameters selected cover all of the unique modes of variation observed in the data.

The PSP model recognizes global and local parameters sets. For our process model we work exclusively with the global parameter set.

VARIATION EFFECTS

The physical length and width of the FET gate varies due to variation in the lithography and etch processes. This variation is monitored in line using electrical and physical means and so the magnitude of these distributions is well known. For short and/or narrow devices these variations have large effects on threshold voltage, drive current and current derivatives. These effects are easily modeled by measuring devices of multiple sizes on a single chip and fitting the PSP global parameter set to this data. We encode this variation as a variation in the PSP parameters **LVARO** and **WVARO**. (We will use bold font to identify PSP parameters without endlessly repeating that phrase.)

Other researchers have shown[3] and we also observe that there is additional variation in effective length due to the variation in the amount of overlap of the poly gate on the source and drain diffusions. This variation is coupled to variation in the overlap capacitance and we model it as a common variation of **LAP** and **LOV** which represent the length of the overlap in calculations of electrical length and overlap capacitance respectively. While the size of the **LVARO** variation is set based on direct measurement of physical poly length, the size of the **LAP** variation provides the modeler with a needed degree of freedom to match the total electrical variation observed in short devices. Additional variation is applied to **CFR** to match the total variation observed in the overlap capacitance.

Another variation which is measure in line is the threshold voltage of both long and short FETs. The variation is larger for short devices than for long devices and this is in part due to the effect of channel length on the threshold voltage of short devices. To determine how to model threshold variation we measured a series of FETs of varying length and width on each of our die. We observed two primary modes of variation in the length scaling data (fig. 1). There is an overall rising or lowering of the entire curve as would be expected from variation in the channel doping, affecting all channel lengths. And there is variation in the roll-off of Vt at short channel length. Surprisingly we did not observe a significant variation in the amount of roll-up (reverse short channel effect) as would be expected from a variation in halo implant dose. We also looked for but did not observe significant variation in the short channel DIBL (Drain Induced Barrier Lowering which reduces threshold voltage at high drain bias). As can be seen fig. 1b there is in variation how high the Vt rises at short length before beginning to drop, however this consistently seems to be more related to the strength of the normal short channel effect. We speculate that this relates to variation in source/drain extension junction depth and vertical profile of the halo rather than total halo dose.

To model the overall shifting of the threshold there are two likely candidate PSP parameters, **NSUBO** which represents the channel doping and **VFBO** which represents the flatband voltage. To choose between these we examined behavior of threshold with body effect for devices of various lengths across our hardware sample. We did this visually looking for individual Vt vs. Vbody curves of the

same length devices which cross each other. This would indicate an independent variation of effective doping and flatband voltage since variation of either alone would increase or decrease Vt at body biases. We observed very few crossings and they were of nearly parallel, closely spaced curves.

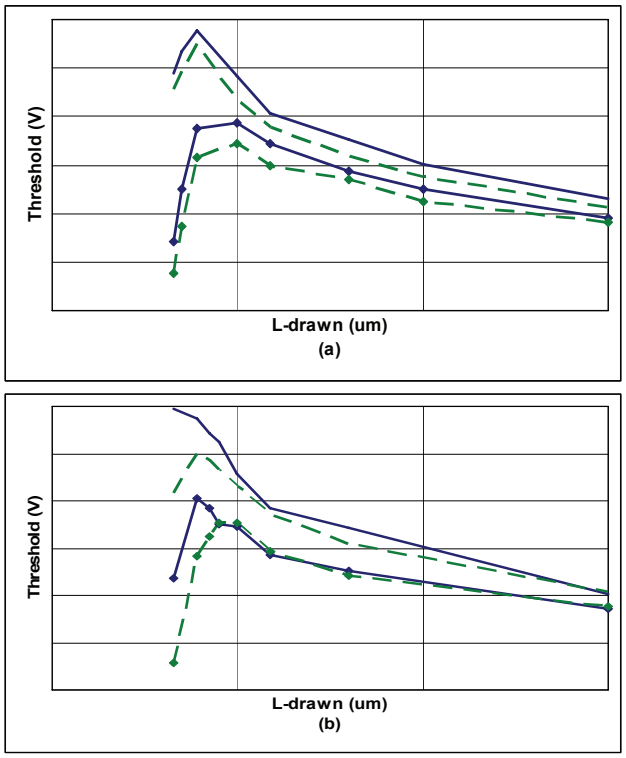

Figure 1. Two modes of threshold variation. (a) Overall variation of channel doping. (b) Variation in short channel roll-off. Lines are linear Vt, lines with markers are saturated Vt. Solid and dashed curves represent different die from different lots.

We also calculated body effect (the rise in Vt between Vbody=0V and Vbody=-1V) and looked at the correlation of that to threshold at Vbody=0. The correlation was very strong and the total variation of body effect was fairly small. We therefore chose **NSUBO** to model the length independent variation of both threshold voltage and body effect. This is a very robust parameter, having a consistent effect even when it is misfit significantly and **VFBO** and other parameters are used to compensate.

The short channel roll-off is modeled by **FOL1** and **FOL2,** which reduce the effective channel doping at short lengths by the following equation. [4]

$$NEFF = Nsub\left(1 - FOL1\frac{L_{EN}}{L_E} - FOL2\left(\frac{L_{EN}}{L_E}\right)^2\right)$$

where *Nsub* is a combination of channel and halo doping and $\frac{L_{EN}}{L_E}$ is a reference length divided by the effective length. Neither of these parameters alone is robust with respect to variation in fitting. It is possible to reasonable approximate the roll-off with many combinations of these two and in some cases one or the other has little. In order for this equation to produce the typical of variation of Vt vs. length seen in fig. 2 **FOL1** and **FOL2** should be varied by the same fraction, otherwise unphysical shape are produce in attempting to get the total variation required. This choice works well for the NFET we have been examining but must be verified with other device designs.

We also examined the behavior of Vt with W. Narrow devices show significant variation due to variation in printed channel width which we discussed above and variation in doping due to the stochastic nature of ion implantation in small devices which must be added to the chip mean variation model which this paper addresses. Examining both the shape of Vt vs. width curves and the total magnitude of Vt variation in narrow devices we did not see anything which requires an addition source of threshold variation in the model for narrow devices.

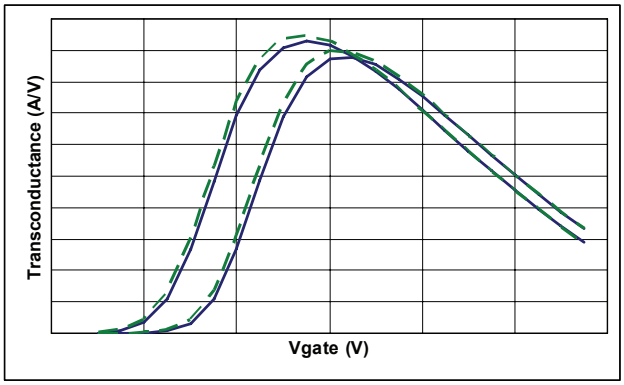

Figure 2. Variation of long channel transconductance. Two FET with nearly identical threshold have different peak gm, implying different mobility. Mobility degradation with both Vgate and Vbody are similar.

We next examined current and transconductance in long wide FETs at low drain bias. This region of operation is very sensitive to mobility as a function of gate bias but not to most other source of variation except for threshold variation. We observed a modest but significant variation in the magnitude of the peak gm. Figure 2 show gm for two FETs with nearly identical Vt but different peak gm values. Oxide thickness variation could cause this effect but Tox is monitored in line and the fractional oxide variation is quite small. It is important for gate current modeling and will need to be included in the complete process model for that purpose.

We attribute most of this difference to variation in mobility and find that **UO** is effective to model it. **UO** has a consistent effect even when the fit is adjusted over quite a wide range using **UO** in combination with other mobility

NSTI-Nanotech 2007, www.nsti.org, ISBN 1420061844 Vol. 3, 2007

parameters. We notice that behavior of gm with gate bias and body bias is quite consistent despite variations in Vt and peak gm. So we conclude that no variation in any of the mobility degradation terms is required.

We examined the output characteristics of the long channel transistors and observed significant variation in output conductance between FETs from different die and lots. Mobility and threshold effect output conductance (gds) measured at a constant bias so we looked at overlays of drain current vs. drain voltage. If output conductance variation was due to mobility or Vt we would expect to see roughly parallel curves. However we found examples where the curves crossed i.e. one FET had more conductance at low drain bias but less conductance at high drain bias. (see fig. 3)

This requires an additional variation in the model and we choose **ALPL** to model it. This parameter also varies the output conductance for short channel devices and in fact the local parameter **ALP** increases more for shorter FETs because it scales inversely with channel length. However at least for our devices at short channel length other effects such as DIBL dominate the output conductance. So we actually see a smaller relative effect of **ALPL** at shorter channel lengths which is consistent with data.

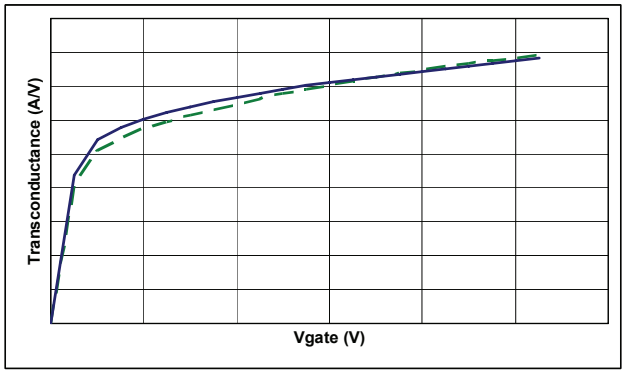

Figure 3. Long channel output conductance. Near Vds=0 the difference between these to curves is probably due to differences in threshold and mobility. These curves intersect, demonstrating the need for an independent variation of output impedance in the model.

Drain current of short channel devices is strongly influenced by series resistance in the source and drain extensions and we suspect that this will vary due to variation in spacer and hot processes. We examined this by looking at gm for short channel FETs. (see Table I) We have already included in the model a variation of mobility and threshold voltage which will affect the gm. To determine if there is additional variation due to series resistance we compare the relative variation (standard deviation divided by mean) for minimum length and long FETs. We expect the location of the peak gm to move with Vt but not the magnitude. So we conclude from the peak gm numbers that additional variation is needed for short channel transconductance beyond what is provided by long channel mobility. We note that at high Vg (110% of FET operating voltage) the long channel variation is reduced while the short channel variation is similar to that at peak gm. We believe this is because the mobility variation decreases as surface roughness and coulomb scattering decrease mobility while series resistance variation remains constant. We use **RSW1** for series resistance and find it does well on our robustness criteria.

Channel Length	Peak gm	Gm at high Vg
1.0um	1.0%	0.6%
0.05um	3.6%	3.7%

Table I: Short FET relative gm variation

We also considered the possibility that saturation velocity varied from lot to lot and would cause variation in drain current of short FETs at high drain bias. We have already discussed output conductance for short channel devices which seems adequately covered by **ALPL**. Since mobility and series resistance effect current at both high and low drain bias and velocity saturation is only important at high drain bias we compare the variation in the two currents. If there is a separate variation in saturation velocity it should cause a lack of correlation between linear and saturated current. We observed correlation coefficients of greater .85 for each of 4 FETs of different short lengths and conclude that variation of saturation velocity does not need to be modeled.

ACKNOWLEDGEMENTS

The authors would like to thank Joachim Assenmacher of Infineon Technologies for valuable input on fitting PSP and preparing corners for PSP models. We thank Prof. Gildenblat of Arizona State University for suggestions on PSP fitting and the physical meaning of various parameters. Phil Corson of IBM assisted setting up the test plans and Moe LaCroix of IBM made the electrical measurements.

CONCLUSIONS

We have presented a methodology for selecting model parameters to model FET variation. We have examined the chip mean variation in one device design in a 65nm CMOS technology and found that it can be well modeled in PSP with seven distributions and eight parameters from the PSP global set. Additional work is required to verify that this set of parameters is suitable for a wide range of FET of various designs.

REFERENCES

[1] J. Watts, et al, NANOTECH 2005,
[2] P. Drennan and C. McAndrew, IEDM 1999, p. 167-170
[3] S. J. Hong, et al, *ESSCIRC 2*004, p 141-144
[4] PSP102.1 manual,
http://pspmodel.asu.edu/downloads/PSP1021_Summary.pdf