Modeling FET Variation Within a Chip as a Function of Circuit Design and Layout Choices

Josef Watts, Ning Lu, Calvin Bittner, Steven Grundon, Jeffrey Oppold

IBM Microelectronics
Essex Junction, VT, 05403, jswatts@us.ibm.com

1 ABSTRACT

In addition to the overall range of circuit characteristics expected from process variation the circuit designer needs to know how closely different circuit element will track one another. We describe a new methodology for modeling correlation between the chip mean variation different design FET and between different FET instances on a single chip. In addition it enables modeling the impact of circuit design choices on FET and circuit tracking

2 INTRODUCTION

All IC manufacturing lines are subject to random variation in process conditions which cause random variation in the chips produced. For example the average length of polysilicon gates will vary from one chip to another. We use the term chip mean variation (CMV) to describe such variations. Many chip designs use several different types of MOSFET, for example an NFET/PFET pair with a low threshold voltage (Vt) for high performance, another pair with higher Vt for low leakage and third pair with a thicker gate oxide for higher voltage I/Os. Depending on the details of process integration the CMV of the Vt between different devices types will be more or less correlated. We refer to this correlation as chip mean tracking.

Within a single chip not all MOSFETs of the same design will have exactly the same characteristics. For example the Vt of an individual MOSFET will vary randomly because the precise number of dopant atoms below the channel will vary during ion implantation. We refer to this variation as across chip variation (ACV). This particular variation is an example of **uncorrelated** ACV because each individual transistor varies in a unique way. An example of **correlated** ACV is the length variation due to lithographic effects which cause the printed length of closely spaced gates to vary in a consistent way compared to widely separated gates.

3 MODEL STRUCTURE

Our approach is to provide a process model with as much detail as possible about the distributions and their correlations and then provide the user with options as to how much of the model sophistication they use. We begin with a Monte Carlo model that includes the process variations and the correlations between. We then construct a simpler "user specified corner" model by controlling the same distributions with a smaller set of user specifiable corner parameters. Finally, we define specific values for the corner parameters to specify recommended fixed corners for static CMOS logic.

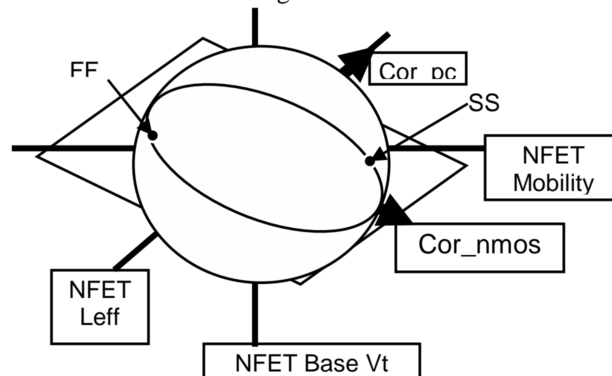

Fig. 1: The Monte Carlo model represented as a sphere provides the most complete coverage of the process space. The user specified corner models covers only the the 2-D plane and the fixed corner models each covers only a single point.

One way of thinking of this hierarchical structure is to envision the Monte Carlo model as a multi-dimensional space. Fig. 1 shows three dimension of this space, NFET channel length, NFET mobility and NFET base Vt. The Monte Carlo model is represented as a sphere in this picture because any combination of these three parameters (within the process tolerances) can occur in a Monte Carlo simulation. The complete Monte Carlo model has more dimensions but more than three is difficult to draw. The user defined corner model is a projection of the Monte Carlo model onto a smaller number of dimensions or corner parameters (see Table I). In the illustration NFET channel length is retained as an independent corner parameter (cor_pc) but mobility and Vt are collapsed to a single dimension labeled "cor_nmos". The user specified corner model is represented by a plane in this picture. Only points on this plane can be specified using the two corner parameters. The foreshortened circle is the portion of this plane containing reasonable process values. The fixed corner models appear as single points in this picture.

4 CHIP MEAN VARIATION

Chip mean variation determines the overall spread in chip performance. If more than one type of FET is used in

a chip the correlation of CMV between FET types is also important. Actually all CMOS circuits have NMOS and PMOS FETs and designers need to understand how their characteristics correlate. This is why wafer fabricators typically supply five so called corner models: FF where both NFETs and PFETs are fast, SS where both are slow, FS and SF where one is fast and the other slow, and TT for typical. The FF and SS models specify the size of the overall process variation and are used to determine the extremes of circuit performance. The FS and SF models specify how large the mistracking between NFETs and PFETs can become and are used to ensure that circuits will function correctly under the extremes of N-to-P mismatchs.

When several different pairs of FET types are used in a single chip design additional circuit characterization needs arise. For example a logic designer will need to ensure that a data path composed of low leakage, high Vt transistors always evaluates before the clock signal, propagating through high speed, low Vt transistors latches the data. This must be true for both the fastest and slowest chip expected from the manufacturing line. Similarly a mixed signal designer needs to ensure correct interaction between logic blocks built with low voltage, thin oxide FETs and analog blocks built with high voltage, thick oxide FETs.

4.1 Physical Basis

The partial correlation between device types comes from the fact that any pair of device types share some common fabrications steps and each has its own process for other steps. For example one pair may share a common gate dielectric process while third device will have a separate gate dielectric. A pair of NFETs with different Vt may share an implant that sets the Vt for one and while the Vt of the other is set by that implant plus another which the first device does not receive.

One approach would be to create model distributions for physical process steps and apply them to the appropriate devices[1]. We use this approach for gate oxide thickness providing one distribution for each gate dielectric process. However the manufacturing process is too complicated to apply this approach to all of the process steps.

4.2 Empirical Method

To capture partial correlations we use the following empirical approach method.

1) Set the gate oxide thickness tolerance based on line control specifications
2) On a device by device basis set tolerance for base Vt, Leff, mobility and series resistance to match electrical line control specifications.
3) For each device pair establish tolerance for ΔVtsat and ΔIdsat. (see Appendix)
4) For each device pair determine the sensitivity of ΔVtsat and ΔIdsat to Δ(Vth0) and ΔLeff. From these solve for tolerance on Δ(Vth0) and ΔLeff to give the correct ΔVtsat and ΔIdsat.
5) For base Vt and Leff convert the device tolerances and delta tolerances into a covariance matrix.
6) Use Principal Component Analysis (PCA) to find a set of independent distributions and weights to reproduce the required device tracking.

This processes establishs the Monte Carlo model for FET DC (and some AC) characteristics. For user specified corners the corner parameters specified in Table I can be used to set each FET type to the desired point in the process space.

Table I

Corner Parameter	Distributions controlled
Cor_tox	Thin gate oxide thickness
Cor_dgtox	Thick gate oxide thickness
Cor_pc	FET Leff, W for poly resistors, etc.
Cor_rx	FET Weff, W for diffused resistors, etc.
Cor_nmos	Mobility, base Vt, series resistance, overlap and junction capacitance, Leff for all N-type MOSFETs
Cor_nmos	Mobility, base Vt, series resistance, overlap and junction capacitance, Leff for all P-type MOSFETs
Cor_xxxfet	Mobility, base Vt, series resistance, overlap and junction capacitance, Leff for the particular MOSFET with model name xxxfet. There is one cor_xxxfet parameter for every FET design in the technology.

A traditional fast N-slow P corner for all NFET-PFET pairs can be achieved by setting cor_nmos to a positive number and cor_pmos to a negative number. As another example the case of an overall slow process with the low Vt FET pair faster than the regular Vt pair could be achieved by setting cor_nfet and cor_pfet to a negative number and cor_lvtnfet and cor_lvtpfet to a negative number of a smaller magnitude. In this system some model parameters are controlled by more than one corner parameters. In these cases a fraction of the total Monte Carlo variation is allocated to each corner parameter and the effects are additive.

Having the ability to model the tracking between different types of FET allows the designer to balance the gain of trading off speed against power in non-critical timing paths with the impact on path to path tracking. The Monte Carlo model allows the user predict the impact of mistracking for typical circuits. They can then define corner parameters which give the same degree of mistracking on the typical circuits and use them to characterize every book in a library.

5 ACROSS CHIP VARIATION

To some degree anything you can measure will vary between two supposedly identical FETs on a single chip. We model three sources of across chip variation. Variation of FET electrical parameters due to random density and placement of dopants under the FET channel which we will call Vt matching because its effect is largest on threshold. Variation of the length of the FET channel due to variation in poly gate etched dimensions which we call across chip length variation (ACLV). Variation of width of the FET channel which we call across chip width variation (ACWV).

5.1 Layout sensitivity

We have only characterized the layout sensitivity for ACLV. Vt matching results primarily from the fact that there is a finite number of dopant atoms randomly placed under the FET channel leading to statistical flucuations.[2] There is also a variation due to systematic variation in ion beam density. To reduce the systematic effects it is often recommended that matched transistors be layout using common centroid designs. We do not model the difference between common centroid and other layouts because our data shows the difference is small. ACWV probably has layout difference similar to ACLV but Weff variation is not generally as critical to circuit operation. The approach applied below to ACLV could also be applied to ACWV but we have not implemented that in our models.

For ACLV we model variation due to three layout variables. Distance between the devices in question, orientation of the gates (vertical or horizontal) and spacing to the next poly line (poly pitch). In addition we recognize ACLV due to line edge roughness and other uncategorized effects. Line edge roughness is included in the Vt matching because our measurement techniques is unable to remove it from that measurement. There are random distributions in Monte Carlo the other four ACLV effects.

It is valuable to model these effects separately because the first three are under the designers control. A design team could decide that all circuits will be laid out with the same orientation and pc pitch in order to reduce the ACLV. However there may be an area penalty to follow such strict layout rules. To model various choices by the design team the model provides global and instance switches for ACLV effects.

Consider orientation. Each FET can be oriented horizontally or vertically including the FET in the scribe line which is used to control channel length. If all FETs have the same orientation as the scribe line FET this effect will not contribute to total circuit variation because the scribe line device is controlled to the chip mean specification and all the others match it in this respect. To model this case the global switch for orientation ACLV is turned off. Neither the overall circuit variation or any mistracking between devices is modeled in this case.

Another case would be if all FET in a particular circuit have the same orientation but it is not known when the circuit is designed and characterized what its orientation will be when placed on the chip. For this case the parameter for orientation would be set to orientation 1 (the default) and the global switch would be enabled. The impact of this variation on overall delay would be included in the simulation but there would be no mistracking impact.

If the orientations are not all the same but are known the instance parameters can be set to indicate the orientation of each FET. During Monte Carlo simulation each orientation gets an independent ACLV contribution. The impact of ACLV on overall circuit performance is correctly modeled and only those FET which have different orientation show mistracking due to orientation. For this case the global switch can be set such orientation 1 matches the scribe line FET and these FETs see no orientation contribution to ACLV. Or the switch can be set for the scribe line FET orientation to be unknown so that each group sees a random variation.

For pre-physical design simulation when the orientation is completely unknown the instance parameter can be set to unknown and each FET is randomly assigned an orientation. In this way the advantage of controlling orientation can be accessed.

The distance and poly pitch ACLV distributions are treated in the same way. This is a simplification since each can take on many possible values. For many modeling needs it enough to break them into two groups.

5.2 Correlated and Uncorrelated ACV

ACV can be correlated or uncorrelated. Vt matching is always uncorrelated. Regardless of anything the circuit designer does each FET instance will receive a unique arrangement of dopant atoms with no predictable relationship to any other instance.

On the other hand all FETs of vertical orientation will be longer (or shorter) than all FETs of horizontal orientation by the same amount. We make a distinction between correlated and systematic in this context. On a given lot there may be a systematic poly pitch effect with all closely spaced poly lines printing smaller than isolated lines. This could be modeled by adjusting the nominal Leff depending on the poly pitch. However we do not assume that this systematic variation will be consistent from lot to lot. Instead we consider that this merely a correlated random variation, meaning that we expect on the next lot to find that closely space lines will resemble one another and differ form isolated lines. But we do not expect the magnitude or direction of the difference to be consistent from lot to lot.

Whether a particular effect should be treated as correlated or uncorrelated may depend on the situation. For example in modeling a single compact circuit distance should be treated as a correlated effect with all FETs in the

circuit receiving the same random effect from distance ACLV. However for a clock tree distributed over a wide area distance should be considered an uncorrelated effect with each instance of the clock buffer receiving a unique ACLV effect for distance.

Fig. 3 shows tracking between four identically designed ring oscillators placed at different locations on the chip. The vertical axis is delay for each ring and the horizontal axis is the chip mean delay. All of the rings have the same orientation and poly pitch so these sources of variation are turned off to calculate the tracking bounds. Since each ring is at a different location on the chip there are correlated ACLV distance effects which cause some rings to be consistently slower than others. This can be seen in the offset between the regression lines of the different rings. The scatter about the lines for each ring is a combination of uncorrelated effects and effects that are correlated within one chip but not across the entire lot.

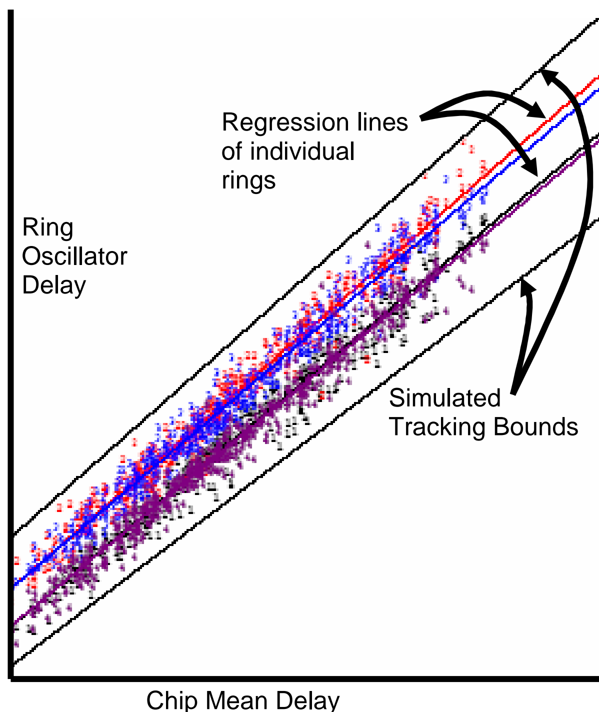

Figure 2. Individual ring delay vs. chip mean delay showing both correlated and uncorrelated ACV.

The difference between correlated and uncorrelated ACV effects can be dramatic. When timing a chain of logic gates correlated ACV effects will add much more tolerance to the delay than uncorrelated effects of the magnitude. Consider a 51 stage inverter ring oscillator with and an ACV effect that introduces a ±15 ps variation in the delay of each gate. If this effect is correlated for all the gates tolerance due to this effect for the whole chain will be ±765 ps (15x51 stages). And the percent variation of the chain will be the same as the per stage percent variation. If the effect is uncorrelated some stages will be fast and others slow. The total delay variation will be ±107 ps. ($15x\sqrt{51}$) The fractional variation of the whole ring will be only one seventh the fractional variation expected from correlated variation.

On the other hand many analog circuits depend on close matching between particular devices in the circuit. Correlated ACLV effects will have no effect on matching because both FETs receive the same variation. An uncorrelated effect will have different effects on the two FETs and so will introduce variation in the circuit function.

By default the ACLV effects for orientation, distance and poly pitch are treated as correlated while the remainder of the total ACLV distribution is treated as random. However the model also allows the designer to override the magnitude of the correlated and uncorrelated distributions, effectively making any ACLV component correlated or uncorrelated.

5.3 ACV in User Specified Corners

There are two corner parameters for ACV. Cor_acv_correlated controls all correlated components of ACLV while cor_acv_ uncorrelated controls Vt matching, ACWV and the uncorrelated portion of ACLV. Because Vt matching is larger for small geometry FETs and ACWV has a larger fractional effect on narrow devices this can be an effective tool for balance the variation of wide and narrow devices at particular corners in addition to allowing the two classes of effects to be characterized separately. The function of the global and instance control switches is preserved during user specified corners so effects that are eliminated by layout rules can be turned off.

In a user specified corner all FET normally move together when the cor_acv_... parameters are moved. In a circuit where the major ACV effect is caused by mistracking of certain pairs of FETs it will probably be impossible to find reasonable values for cor_acv_correlated and cor_acv_uncorrelated which produce enough circuit variation. We address this problem with an additional instance parameter. By defaults all devices move in the same direction with the corr_acv parameters. However this instance parameter can be set such that a particular device moves in the opposite direction compared to FETs with the default setting. This requires intelligence in setting this instance parameter but allows a fixed corner to reproduce the effect of transistor mistracking in analog circuits. This enables more rapid design experimentation and optimization.

6 APPLLICATIONS

6.1 User Specified Corners

There are many reported methods for selecting process corners either from line monitoring data or from Monte Carlo models[4,5]. In this section, we intend only to

illustrate how our process model can be used to define corners which are tailored to the users design style and design flow.

As a simple example, consider setting the fast and slow process corners for timing logic delays. This can be done by first timing a few representative circuits using Monte Carlo analysis. Then the same circuits are simulated in a series of runs with the corner parameters `cor_pc`, `cor_rx`, `cor_tox`, `cor_acv`, `cor_nmos` and `cor_pmos` all set to the same value and then varying that value from run to run. The fast and slow corners are specified by setting the corner parameters to the values which reproduced the fast and slow delays from Monte Carlo. These timing corners could then be used to generate fast and slow static timing rules for the entire library.

An analog designer working in the same technology could use a similar procedure to determine a set of corners which produced extreme cases for circuit characteristics important for his or her designs. The existence of multiple corner parameters rather than one fast-slow parameter allows for creation of multiple process corners which stress the circuit in different ways.

6.2 Foundry specified corners

For simple logic design, some designers prefer to use corners specified by the foundry. Since these corners must cover a variety of typical design styles, they may be more conservative than corners defined by the user for a particular library. To meet this need, we supply seven corners. Fig. 3 shows these corners and their meanings. Each of these corners is specified as a vector of values for the various corner parameters. From the model point of view these corners are exactly like user specified corners, they are defined by a set of values for the corner parameters. The only difference is that they are already defined for the user and are available on a pull-down menu within the design kit environment.

6.3 Statistical and Parameterized Timing

For statistical timing analysis of logic circuits, the timing rules need to specify not just fast or slow delay times but also sensitivities to different sources of process variation[6] This requires multiple process corners and multiple simulations of each circuit to establish their delays and their sensitivities. The process model we have described with its sophisticated Monte Carlo model and its rich set of corner parameters provides the tools needed to find and set these corners for library characterization.

6.4 Selecting User Defined Corners

There are many reported methods for selecting process corners either from line monitoring data or from Monte Carlo models[4.,5]. In this section, we intend only to illustrate how our process model can be used to define corners which are tailored to the users design style and design flow.

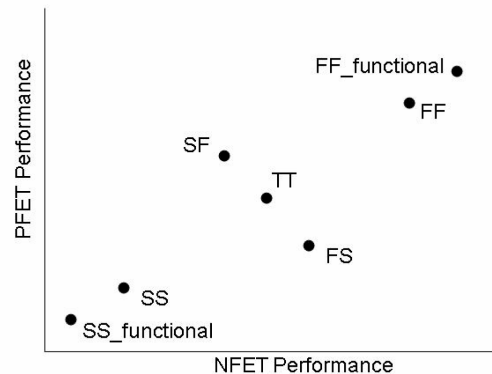

Fig. 3 Seven corners. TT-typical. FF-fast corner for logic timing. SS-slow corner for logic timing. FS-fast N slow P mismatch. SF-slow N fast P mismatch. FF_functiona and SS_functionall-All FET thresholds and on currents at 3-sigma fast or slow limits for all geometries.

As a simple example, consider setting the fast and slow process corners for timing logic delays. This can be done by first timing a few representative circuits using Monte Carlo analysis. Then the same circuits are simulated in a series of runs with the corner parameters `cor_pc`, `cor_rx`, `cor_tox`, `cor_acv`, `cor_nmos` and `cor_pmos` all set to the same value and then varying that value from run to run. The fast and slow corners are specified by setting the corner parameters to the values which reproduced the fast and slow delays from Monte Carlo. These timing corners could then be used to generate fast and slow static timing rules for the entire library.

An analog designer working in the same technology could use a similar procedure to determine a set of corners which produced extreme cases for circuit characteristics important for his or her designs. The existence of multiple corner parameters rather than one fast-slow parameter allows for creation of multiple process corners which stress the circuit in different ways.

7 CONCLUSION

We have described a unified spice process model which supports sophisticated Monte Carlo modeling as well as user defined and foundry defined process corners. The model includes correlations between different devices designs and different device instances within a single device. It describes both chip to chip and within chip variation. Multiple corner parameters allow the user to easily create simulation corners to simulate process extremes including extremes for device mistracking.

8 REFERENCES

[1] Slezak, Litschmann, Banas, Mlcousek and Kejhar, "On the correlations between process parameters in statistical modeling", NSTI-Nanoteck 2004, Vol 2, pp144-146

[2] J. Jackson, "A user's guide to principal components", John Wiley & Sons, New York, 1991

[3] Stadlober, Kocher, Rappitsch, "Simulation models for robust design using location depth methods", Quality and reliability engineering international, V19, N4, 2003, P317-326

[4] Singhal and Visvanathan, "Statistical device models from worst case files and electrical test data", IEEE trans. on Semiconductor Manufacturing, V 12, N 4, pp. 470-484, Nov. 1999

[5] Jess, Kalafala, Naidu, Otten and Visweswariah, "Statistical timing of parametric yield prediction of digital integrated circuits", DAC 2003, p932-937.

APPENDIX

Depending on the maturity of the manufacturing process tolerance on Vtsat and Idsat tracking can be established by a combination of in line measurement, historical precedent and engineering judgement. Becasuse these quantities are defined in terms of Idsat and Vtsat of individual FET which are already measured in line no additional test time is required. Specifically we define:

$$\Delta Vt_{ij} = \left(Vt_i - Vt_j\right) - \left(Vt_i - Vt_j\right)_{nom} \quad \text{for 2 NFETs or 2 PFETs}$$

$$\Delta Vt_{ij} = \left(Vt_i + Vt_j\right) - \left(Vt_i + Vt_j\right)_{nom} \quad \text{for 1 NFET and 1 PFET}$$

$$\Delta Id_{ij} = \frac{Id_i}{Id_{i,nom}} - \frac{Id_j}{Id_{j,nom}} \qquad (1)$$

Where i and j represent two different design FETs, Id is Idsat and Vt is Vtsat for those FETs.

We calculate the sensitivities of these composite quantities to the deltas between the model parameters of the different FET types as:

$$\begin{aligned}
\frac{\partial(\Delta Vt_{ij})}{\partial(\Delta Vth0_{ij})} &= \frac{\partial Vt_i}{\partial vth0_i} + \frac{\partial Vt_j}{\partial vth0_j} \\
\frac{\partial(\Delta Id_{ij})}{\partial(\Delta Vth0_{ij})} &= \frac{\partial Id_i}{\partial vth0_i} + \frac{\partial Id_j}{\partial vth0_j} \\
\frac{\partial(\Delta Vt_{ij})}{\partial(\Delta Leff_{ij})} &= \frac{\partial Vt_i}{\partial Leff_i} + \frac{\partial Vt_j}{\partial Leff_j} \\
\frac{\partial(\Delta Id_{ij})}{\partial(\Delta Leff_{ij})} &= \frac{\partial Id_i}{\partial Leff_i} + \frac{\partial Id_j}{\partial Leff_j}
\end{aligned} \qquad (2)$$

Using the model we can evaluate the derivatives in (2) and form a matrix equation for each for the tolerances for each pair of devices.

$$\begin{bmatrix} \Delta Vt_{ij} \\ \Delta Id_{ij} \end{bmatrix} = \begin{bmatrix} \frac{\partial(\Delta Vt_{ij})}{\partial(\Delta Vth0_{ij})} & \frac{\partial(\Delta Vt_{ij})}{\partial(\Delta Leff_{ij})} \\ \frac{\partial(\Delta Id_{ij})}{\partial(\Delta Vth0_{ij})} & \frac{\partial(\Delta Id_{ij})}{\partial(\Delta Leff_{ij})} \end{bmatrix} \begin{bmatrix} \Delta Vth0_{ij} \\ \Delta Leff_{ij} \end{bmatrix} \qquad (3)$$

Using the measured or estimated tolerances on deltas of line measurements we solve (3) for the required tolerances on model parameters. We also have the toleranaces on the individual model parameters set to give the correct tolerance on individual FET Vtsat and Idsata. We can then calculate the variances and covariances from the tolerances as:

$$\begin{aligned}
\sigma_i^2 &= (\Delta Vth0_i)^2 \\
\sigma_{ij}^2 &= \frac{\Delta Vth0_i^2 + \Delta Vth0_j^2 + \Delta Vth0_{ij}^2}{2}
\end{aligned} \qquad (4)$$

for base Vt and

$$\begin{aligned}
\sigma_i^2 &= (\Delta Leff_i)^2 \\
\sigma_{ij}^2 &= \frac{\Delta Leff_i^2 + \Delta Leff_j^2 + \Delta Leff_{ij}^2}{2}
\end{aligned} \qquad (5)$$

for Leff.

We perform principal component analysis on the covariance matrices for Leff and Vth0 seperately and construct correlated distribution for each parameter across all FET types offered in the technology.

NSTI-Nanotech 2005, www.nsti.org, ISBN 0-9767985-3-0 WCM, 2005