Statistical Modeling for Circuit Simulation

Colin C. McAndrew
Motorola, Tempe, AZ, USA
Colin.McAndrew@motorola.com

Abstract

Robust, high yield IC design requires statistical simulation, and therefore statistical models. Simple "fast" and "slow" sets of model parameters are not sufficient to predict the manufacturing variations of all measures of circuit performance for arbitrary circuit topologies, device geometries, and biases. This paper describes an accurate and efficient approach to statistical modeling and characterization. The procedure is based on physical process parameters, and explicitly accounts for correlated and uncorrelated variations of statistical parameters. The process is generic, and so is applicable to any type of device, and emphasizes the accuracy of device electrical performance variation modeling, rather than model parameter variation modeling. This provides an accurate and simple way to model and simulate the statistical variation of circuit electrical performances.

1. Introduction

Integrated circuit (IC) manufacturing processes are statistical in nature. Besides defects, which cause catastrophic circuit failure, parametric variations are observed in electrical performances $\boldsymbol{E}$. These occur between fabs, between tool sets at the same fab, over time within the same tool set and fab, between wafers in the same lot, and between die from the same wafer. Variations also occur between devices within the same die ("mismatch"). Parametric variations can cause a die to fail to meet specs, or fail to operate functionally. Analog circuits are especially sensitive to process variations, although high-speed digital circuits in state-of-the-art processes are becoming more sensitive than in the past to statistical variations [1].

Design of ICs therefore should take proper account of the statistical nature of IC manufacturing processes. For some classes of circuits this has been done effectively and efficiently using "generic" case files. Many digital CMOS circuits are topologically similar, use minimum length devices, and have two key figures of merit, speed and power, which are highly correlated. In a simplistic sense

$$\frac{\partial V}{\partial t} = \frac{I}{C} \tag{1}$$

so models that bracket the manufacturing variations of I and C ("slow" and "fast," or "worst" and "best," for low and high current respectively, where the names relate to digital circuit speed) reasonably bracket manufacturing variations in speed and power.

These generic models could also reasonably be expected to bracket manufacturing variations in amplifier slew rate, which is a "$\partial V/\partial t$" type performance. But what about the hysteresis of a Schmidt trigger? or the phase margin or input offset voltage of an amplifier? These $\boldsymbol{E}$ are not controlled just by drive current strength.

And what about head room in an amplifier? Besides "fast" and "slow" MOSFET and BJT models, model parameter sets can include "lo" and "hi" resistance models. For "fastest" circuit operation it would seem you should mix "fast" MOSFET models with "lo" resistance models. But if transistors are connected to supply rails via resistances, then worst-case headroom comes from the combination of "fast" transistor and "hi" resistance cases.

Even the names can be misleading. For many BJTs high I_c is physically correlated with low Early voltage, and so the AC gain, $g_m r_o \approx V_A/(kT/q)$, is at its "worst" (low) value when current drive is "best" (high).

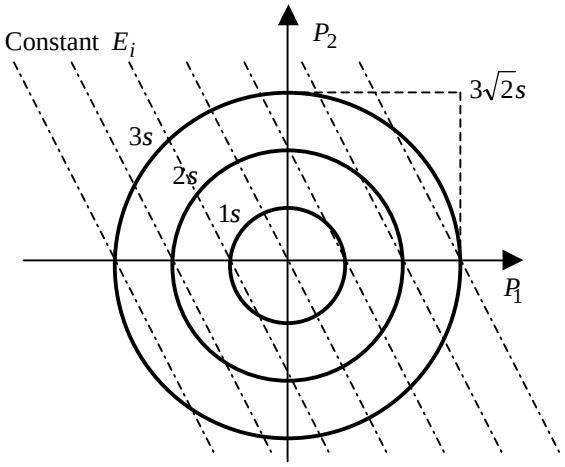

Fig. 1 Statistical Curse of Dimensionality

0-7695-1881-8/03 $17.00 © 2003 IEEE

As you include "lo/hi" models for different classes of devices available in a technology, three things happen. First, the number of possible combinations increases exponentially with the number of classes of cased device models. This makes simulating all combinations of cases impractical. Second, the "curse of dimensionality" rears its head, see Fig. 1. As you introduce $\pm ns$ variations in each of N variables the manufacturing likelihood becomes $\pm n\sqrt{N}s$. Third, because parameters and electrical performances are correlated between different categories of cased devices, unphysical combinations can occur. For example, the width variation of polysilicon resistors is highly correlated with the effective length variation of MOSFETs, and treating these as separate parameters overestimates the amount of variation predicted. A circuit designed to work with unphysical combinations of parameters will yield in practice, but this approach is uncontrolled and causes over-design.

On top of that, even if the statistical variations in the model parameters are varied to specified levels, there is no guarantee that the "case" files give any meaningful qualitative predication of the manufacturing variability in any particular E_i. This is because the sensitivities of the $\boldsymbol{E}$ to the parameters varied in the case files are not taken into account.

There is a solution to the problem. First, apply appropriate proper statistical simulation techniques, not just a "shot gun" combination of generic case files. And second, ensure that the appropriate statistical models are available, because different statistical simulation techniques require different sorts of statistical models. The keys to this process are physical process parameters as a modeling basis, sensitivity analysis, and back propagation of variance (BPV).

2. Physical modeling basis

The electrical behavior of semiconductor devices is controlled by the physical and geometric properties of materials they are made of. Geometric properties $\boldsymbol{G}$ include length and width, areas and perimeters of parasitic junctions, spacing to adjacent devices (for interconnect fringing and coupling capacitance), and spacing to adjacent trench isolation (for stress effects). These come from the physical design (layout) of the device.

Physical properties $\boldsymbol{P}$ include doping levels and profiles, their integrated effect in sheet resistance, lateral geometry variations (like poly critical dimension, and junction out-diffusions), recombination and generation time constants, and vertical dimensions like junction depths and metal, poly, and dielectric thicknesses.

Some types of devices, e.g. MOSFETs and resistors, have SPICE models that are based on physical process parameters. These $\boldsymbol{P}$ can be directly used for statistical modeling and simulation. Other devices, primarily BJTs, have SPICE models whose parameters are correlated and are not fundamental $\boldsymbol{P}$, but depend on them [2]. For these devices, the model files must be constructed with SPICE model parameters written as functions of $\boldsymbol{P}$. This can require some up-front investment of effort, but similar mappings hold for similar devices across technologies, so once this is done the effort required to update them for a new device or technology is small.

A simplified example is:

```
.par dle=0  // emitter size variation
.par rpb=1 // pinchbase sheetrho varn
.subckt npn (c b e s) le_um=5 we_um=5
.par ae=(le_um+dle)*(we_um+dle)
.par pe=2*(le_um+we_um+2*dle)
.model mymod vbic type=npn
+ is =(isa*ae+isp*pe)*rpb
+ cje=(cea*ae+cep*pe)*sqrt(rpb)
+ ver=verNom/sqrt(rpb)
+ ...
q1 (c b e s) mymod
.ends
```

Statistical variations in $\boldsymbol{P}$ fall into two classes, often termed "global" (or inter-die) and "local" (or intra-die) variations [3]. For statistical simulation purposes, it is better to consider the two types of variations to be correlated (between devices, across a die) and uncorrelated (within a die). For a particular P_j

$$P_j = P_{jo} + dP_{jc} + dP_{ju}(\boldsymbol{G}) \qquad (2)$$

where P_{jo} is the nominal value of the parameter, dP_{jc} is the variation correlated between devices within a die, and dP_{ju} is the variation that is uncorrelated between instances of a device within a die, which is a function of the geometric layout attributes of the device.

For purposes of statistical simulation, dP_{jc} is a single statistical variable for all devices sensitive to the parameter P_j in the circuit. A distinct and separate dP_{ju} is used for each separate instance of a device sensitive to P_j that is selected for mismatch analysis.

Rather than understanding physically what causes variations in device performance, an alternative is to numerically analyze extensive model parameter sets extracted from a large statistical sample of devices [4]. While being generic, not device specific, this approach requires significant effort, encounters problems when parameters are extracted using optimization (and so are statistically "noisy"), gives no physical insight into root causes of variation, and is neither predictive nor portable. More important, it requires separate techniques to characterize uncorrelated and correlated statistical parameters. Physical modeling is preferred.

3. Statistical characterization process

There is a temptation to directly measure variations in the process parameters $\boldsymbol{P}$, but there are several problems with this. Consider L_{eff} of a MOSFET, a key statistical parameter for short channel devices. There are dozens of published methods to determine L_{eff}, and they give different values. Which should be used for statistical characterization? SPICE models are approximations, so what guarantee is there that putting in $\pm 3s$ variations in L_{eff} will give $\pm 3s$ variations in $\boldsymbol{E}$? And different SPICE models will give different values for the variations in $\boldsymbol{E}$ for the same variation in L_{eff}.

At heart, the goal of modeling is to correctly represent $\boldsymbol{E}$, it is not to model $\boldsymbol{P}$ accurately. Circuits function or not depending on the electrical performance of the devices from which they are built, not directly on parameter values. So given specified variations in $\boldsymbol{E}$, how can variations in $\boldsymbol{P}$ be derived that accurately model the desired statistics of $\boldsymbol{E}$?

If the fluctuations in $\boldsymbol{P}$ are not overly large (i.e. if a process is manufacturable), then to a good approximation

$$dE_i = \sum_j \frac{\partial E_i}{\partial P_j} dP_j \,. \tag{3}$$

The variance in E_i then follows as

$$s^2_{E_i} = \sum_j \left(\frac{\partial E_i}{\partial P_j} \right)^2 s^2_{P_j} \,. \tag{4}$$

If s_{E_i} are specified, then $\partial E_i / \partial P_j$ can be calculated from the SPICE models, and (4) gives a set of simultaneous linear equations that can be solved for s_{P_j}. The procedure is termed BPV, for backward propagation of variance.

With s_{P_j} characterized, this provides models that can be used for distributional statistical simulation, i.e. for Monte Carlo type analyses.

There are some interesting properties of this characterization procedure. First, note that it must be applied separately for each SPICE model that is being characterized. If two different MOSFET models require statistical characterization, the BPV process can give different values for s_{P_j} for each model (because the sensitivities $\partial E_i / \partial P_j$ differ). While this at first may seem strange, if you are used to a mindset where parameters are physical and sacrosanct and should be measured directly, it gives the desired result: the models will predict the same variation in $\boldsymbol{E}$, which is the goal of modeling for circuit simulation.

Second, there may be more $\boldsymbol{E}$ than $\boldsymbol{P}$. When (4) are solved this gives a least squares solution.

Third, if some P_j are characterized directly, then the associated columns of the matrix of squared sensitivities in (4) are subtracted from the left-hand side, and the variances in the remaining $\boldsymbol{P}$ calculated. Direct characterization of P_j is termed FPV, for forward propagation of variation.

Finally, analysis of (4) can provide very useful sanity checks on models, data, and the selection of $\boldsymbol{E}$ for statistical modeling. If the selected $\boldsymbol{E}$ are not sufficient to make the $\boldsymbol{P}$ mathematically observable, then the matrix of squared sensitivities in (4) becomes mathematically poorly conditioned. This then tells you that you have chosen inappropriate $\boldsymbol{E}$ (but does not tell you how to choose the $\boldsymbol{E}$ properly, see below for more discussion). If sensitivities from the SPICE models are inaccurate, or if some FPV $\boldsymbol{P}$ are inaccurate, then (4) can generate a negative variance. While this at first seems problematic, it is in fact a significant feature of BPV that it will not generate statistical models unless all the data and SPICE models are self-consistent and reasonable.

How do you choose the $\boldsymbol{E}$? There are two criteria. First, they need to be measures of device performance that highly correlate with circuit performance. Second, they need to make the $\boldsymbol{P}$ mathematically observable, i.e. so that as noted above the matrix of squared sensitivities in (4) is well conditioned. They do not need to be chosen so they have a one-to-one relationship with $\boldsymbol{P}$, or are at some bias condition or geometry that is far removed from typical circuit applications. Knowledge of circuit applications and device operation guides this selection.

As a specific example, one of the simplest ways to try to get direct information related to emitter resistance for a BJT is to bias it with an open collector, and with a base current significantly higher than used in normal circuit operation. This causes the device to operate so far removed from a normal operating regime that modeling tends to be poor (models are optimized for normal operation, and simple lumped resistance models are only accurate over limited bias ranges), and the device operates in a completely different fashion than in a real circuit.

Besides distributional models, for Monte Carlo like simulations, IC design does require case files. From the lines of constant E_i in Fig. 1, it is apparent that to find extreme case files for E_i, at a defined ns probability level of being manufactured, the problem to solve is

$$\max_{\boldsymbol{P}} dE_i = \sum_j \frac{\partial E_i}{\partial P_j} dP_j \text{ subject to } \sum_j \left(\frac{dP_j}{s_{P_j}} \right)^2 = n^2 \,. \tag{5}$$

Applying the method of Lagrange multipliers gives

$$dP_j = \pm \frac{n s_{P_j}^2 \left(\partial E_i / \partial P_j\right)}{\sqrt{\sum_k s_{P_k}^2 \left(\partial E_i / \partial P_k\right)^2}} . \quad (6)$$

This gives case files specific for extremes in E_i, but as these require simulation of the sensitivities $\partial E_i / \partial P_j$, which are circuit dependent, the process cannot be run as part of a normal model generation process, but needs to be run dynamically during the IC design process. And the specific case files are valid for only one circuit topology and device geometry selection, and separate specific case files are required for each E_i.

Consequently, generic case files that can be used for a shotgun type approach to simulation are also valuable (these are the generic case files like "slow" and "fast"). These can be generated by selecting extreme (generally $\pm 3s$) values for E_i for devices, rather than circuits, and using nonlinear least squares optimization to find the $\boldsymbol{P}$ that generate the desired $\boldsymbol{E}$. Care must be taken in the process, to ensure that there is physical consistency between the $\boldsymbol{E}$. For example, simultaneously specifying high I_C and high V_A for a BJT is physically unrealizable.

This same optimization procedure is invaluable for tuning models to nominal process conditions. This obviates the need to select a "golden" wafer for modeling (and no wafer is ever perfectly nominal for every parameter), all that is required is a reasonable wafer.

The characterization flow is:

1. Extract models from a "reasonable" wafer.
2. Define and add the mappings from $\boldsymbol{P}$ to model parameters.
3. Define $\boldsymbol{E}$ and obtain distributions, verify $\boldsymbol{P}$ are observable in $\boldsymbol{E}$.
4. Generate typical models by optimizing $\boldsymbol{P}$ to adjust $\boldsymbol{E}$ to nominal targets.
5. Generate distributional models for $\boldsymbol{P}$ by solving (4) for the variances of $\boldsymbol{P}$ (this requires simulating the sensitivities $\partial E_i / \partial P_j$ from the SPICE models).
6. Define targets for $\boldsymbol{E}$ for generic case files, generate $\boldsymbol{P}$ to model these $\boldsymbol{E}$ using nonlinear least squares optimization.
7. For each E_i, generate specific extreme case files using (6).

A general-purpose program does all appropriate calculations, for completely general models, to generate these statistical models in a few minutes on an engineering workstation.

Fig. 2 and Fig. 3 show manufacturing data and modeled values for drain currents for wide/long and wide/short PMOS and NMOS devices. The dotted box is the manufacturing specification. The dashed hexagon is a convex hull connecting generic case files, shown as the heavy plusses at the corners. The ellipses are 1-, 2-, and 2-σ probability contours generated from a 500 sample Monte Carlo simulation using the distributional model. The models accurately capture the spread in the data, and also the correlation structure, which differs between long and short devices.

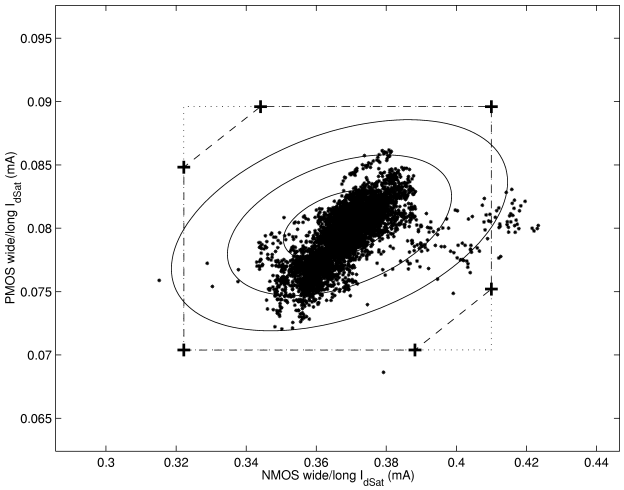

Fig. 2 Models and data from wide/long MOSFETs.

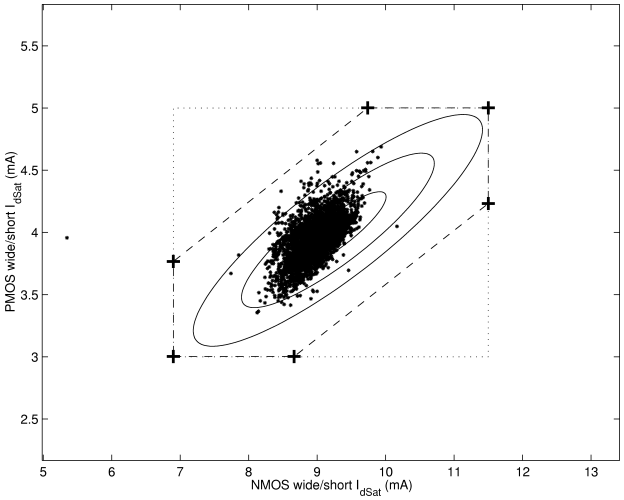

Fig. 3 Models and data from wide/short MOSFETs.

4. Specification of manufacturing variances

Given variances for $\boldsymbol{E}$, the procedures described above allow efficient generation of accurate statistical models. Where do the variances of $\boldsymbol{E}$ come from?

One source is lab measurements on a small sample of wafers. Lab measurements tend to be accurate, and especially for mismatch data can be the most reliable source. However, there is no guarantee that statistical spreads derived from a small number of wafers match the long-term variation, over multiple manufacturing sites and tools.

Manufacturing line measurements are therefore the best source of statistical data for generating statistical models. But such data needs to be interpreted carefully. The accuracy of fab measurements, especially for low currents and for capacitances, tends not to be not as good as from lab measurements. How accurate is the chuck temperature control? Especially for BJTs, whose currents vary roughly exponentially with temperature, imprecise temperature control can be manifest as incorrectly large statistical variations in some $\boldsymbol{E}$.

Further, if during a specific time period one product dominates manufacturing, and fab engineers know that this part gives better yields if the process is skewed from typical, then they will often make appropriate adjustments to maximize yield. Data measured from these lots is not statistically representative of the process, and so is not an appropriate basis for statistical design models.

Worse, most design activity occurs early in the life cycle of a manufacturing process. At this time, by definition, the long-term statistical distributions of $\boldsymbol{E}$ are not known accurately.

Therefore, although manufacturing data is the golden source of statistical variations (with the caveat that the accuracy of the data needs to be verified), more is needed for generation of statistical models. Variations defined for $\boldsymbol{E}$ should be considered as specifications (upper and lower limits), and the models are then forced to comply with these limits using the process described above. The limits are defined based on manufacturing data, expectations and extrapolations based on previous technologies and knowledge of tool sets and expected manufacturing sites, and engineering judgment and experience.

The $\boldsymbol{E}$ should be chosen so that they are strongly correlated with circuit performances, and make the $\boldsymbol{P}$ mathematically observable. For example, if mobility m, channel length variation $\Delta_L = L_{drawn} - L_{eff}$, and flatband voltage V_{FB} are considered key MOSFET process parameters (along with T_{OX}, considered to be measured directly), then these parameters are observable in the threshold voltage V_{TH} of a wide/long device and the saturated drain currents I_{dsat} of both wide/long and wide/short devices. The I_{dsat} of a wide/short device is important for digital CMOS speed, and so is a good choice as an E_i. However, I_{dsat} of a wide/long device is not a relevant quantity for many circuits. Rather, long devices are generally used in analog rather than digital applications, and are operated with gate voltages near V_{TH} rather than at the supply (where I_{dsat} is measured). And conductances can be more important than currents for analog circuits. So for a wide/long device, peak g_m is a much better choice of E_i than I_{dsat} because it is more relevant to circuit performance, although both make m observable in the BPV process.

Note that nothing in the BPV characterization process specifies that the $\boldsymbol{E}$ must be measures of device performance, they could be measures of circuit performance (like ring oscillator speeds), as long as the $\boldsymbol{P}$ are observable in the $\boldsymbol{E}$. This is a significant feature of the BPV process, because it does not concentrate on directly measuring parameters, the $\boldsymbol{E}$ can be arbitrary and can include circuit, rather than device, performances. For both types of data, note that it is critical that as part of the BPV characterization process, the calculation of $\boldsymbol{E}$ from the models is done in **exactly** the same manner as it is done in measurement.

Most important, there should be an on-going process, generally termed "loop closure," in which statistical manufacturing data is regularly compared to models, and to the specifications from which the models were derived. This flags any discrepancies between the models to which circuits are being designed and actual performance of devices in manufacturing.

A final word on comparing manufacturing data to models. By definition, models are an imprecise representation of reality, models have errors. Sometimes there are observed discrepancies between data and SPICE models that cannot be fixed just by adjusting parameters. In this case an offset between statistical limits of modeled and actual performance needs to be accepted, and accounted for in design if necessary.

5. Correlated and uncorrelated components

The analyses above did not distinguish between correlated and uncorrelated components of statistical variation. This is a major benefit of the generic approach, it is valid for both types of variation, correlated and uncorrelated. In other words, BPV provides an identical modeling basis and characterization formalism for both global and mismatch variation. This is not true of other approaches, for example PCA for global variation modeling is generally coupled with direct characterization of V_{TH} and gain factor K_P mismatch for local variations. Not only does this require separate measurement and characterization procedures, it introduces modeling inconsistencies, because V_{TH} and K_P are correlated via oxide thickness, yet the characterization procedure does not account for this.

From (2), the total measured variance in some parameter P_j is

$$s^2_{P_j} = s^2_{P_{jc}} + s^2_{P_{ju}}(\boldsymbol{G}). \qquad (7)$$

Historically, the uncorrelated components of variations have been small compared to the correlated component variations. This meant that characterization of the correlated and uncorrelated components could be handled separately. Typically it was assumed that

$$s_{P_j} \approx s_{P_{jc}}, \tag{8}$$

so although data measured in manufacturing lines includes both components of variation, to a good approximation it can be directly used just to characterize dP_{jc}. Characterization of dP_{ju} was then a completely decoupled and separate step, based on mismatch measurements (although still using the BPV process). The two types of variations are then combined additively in models.

This is no longer the case. $s_{P_{ju}}$ increases as geometry decreases, and in modern technologies for the smallest devices the "uncorrelated" or local variation can be significant compared to the "correlated" or global variation. Fig. 4 shows this diagrammatically; see [5] for more specific details.

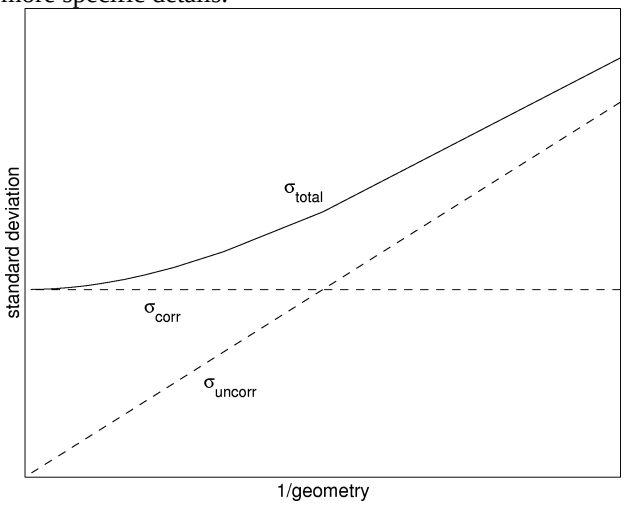

Fig. 4 Global and local variation over geometry

There are several very important ramifications of the increasing contribution of dP_{ju} to variations in P_j. First, mismatch characterization cannot be done as an independent and subsequent step to dP_{jc} characterization. It needs to be done first, and then (4) becomes

$$s^2_{E_i} - \sum_j \left(\frac{\partial E_i}{\partial P_j} \right)^2 s^2_{P_{ju}}(\boldsymbol{g}) = \sum_j \left(\frac{\partial E_i}{\partial P_j} \right)^2 s^2_{P_{jc}} \tag{9}$$

where the measured (or specified) s_{E_i} include both global and local components. If the subtraction of the FPV'd mismatch variances in (9) is not done, then the mismatch variances are double counted in modeling.

Second, for purposes of statistical simulation, a major assumption of case file simulation is that devices are correlated. If dP_{ju} dominates variations, as is starting to be the case for small geometry devices, then case file simulation can drastically overestimate the variability in circuit performance. This means circuits designed with case files will yield, however they will be over-designed.

Third, if case file simulation only involves perturbations in dP_{jc}, then from Fig. 4 it is apparent that the geometric dependence of the overall variation of P_j will not be modeled. Again, this leads to inaccuracies in circuit simulation.

These issues are just starting to become important, as the industry moves into a regime where dP_{ju} becomes comparable to or greater than dP_{jc}.

6. Conclusion

This paper has detailed an approach for defining statistical models, based on physical process parameters, and for characterizing statistical models for both case file simulation and distributional simulation. The process is efficient, taking several minutes to run on an engineering workstation, and accurate, in that it guarantees that variations in electrical performances important for circuit design are matched. The BPV process is identical for characterization of both global and local variation, and is independent of specific SPICE models.

7. References

[1] V. Mehotra, S. Nassif, D. Boning, and J. Chung, "Modeling the effects of manufacturing variation on high-speed microprocessor interconnect performance," *Proc. IEEE IEDM*, pp. 767-770, 1998.

[2] W.F. Davis and R.T. Ida, "Statistical IC simulation based on independent wafer extracted process parameters and experimental design," *Proc. IEEE BCTM*, pp. 262-265, 1989.

[3] J.B. Shyu, G.C. Temes, and K. Yao, *IEEE J. Solid-State Circuits*, vol. SC-17, no. 6, pp. 1070-1076, Dec. 1982.

[4] J.A. Power, B. Donellan, A. Mathewson, and W.A. Lane, "Relating statistical MOSFET model parameter variabilities to IC manufacturing process fluctuations enabling realist worst-case design," *IEEE Trans. Semicond. Manufact.*, vol. 7, pp. 306-318, 1994.

[5] R. Difrenza, P. Llinares, S. Taupin, R. Palla and G. Ghibaudo, "Comparison between matching parameters and fluctuations at the wafer level," *Proc. IEEE ICMTS*, 2002.