A Compact Model Methodology for Device Design Uncertainty

R. Q. Williams, J. S. Watts, M-H. Na, and K. Bernstein

IBM Corporation, 1000 River Road, M/S 972F, Essex Junction, Vermont 05452
Email: rqw@us.ibm.com

ABSTRACT

The pace of semiconductor product development has accelerated to the point where there is significant overlap in schedules between product development and transistor design/process development. This concurrent design introduces an uncertainty in product development while the transistor design evolves to meet targets that were established early in the product development schedule. It challenges the circuit designer to anticipate possible transistor design modifications and to compensate for second-order and sometimes first-order transistor design changes as the technology matures. This work describes for the first time a compact modeling approach that allows the circuit designer to quantify the impact of this uncertainty. Simulation examples are provided for various representative scenarios.

Keywords: compact model device design uncertainty

1 Introduction

The trends toward deep technology scaling and high function chip designs such as system-on-a-chip have been well chronicled in the International Technology Roadmap for Semiconductors (ITRS) [1]. Product development groups respond to these trends while maintaining manageable time-to-market schedules (and coincidentally maintaining Moore’s Law) by initiating their designs sooner with respect to the development of next-generation semiconductor processes. Techniques such as design reuse and software design system innovations help to reduce design cycle time, however challenges in lithography, new materials, scaling, etc., remain, creating concurrency in product and technology development.

The problem of concurrency manifests itself to the circuit designer in compact models that consist of scaled versions of models from previous technology generations that have been recentered to the latest performance targets, whereas ideal models would be completely hardware-extracted from a mature fabrication process. The circuit designer must therefore have some method of assessing the possibility that changes in the details of the compact model fit will occur as device/process learning progresses.

One method for coping with concurrency is to create a compact modeling methodology that gives circuit designers the ability to explore the range of potential device design variations using circuit simulation. In this paper we introduce a new statistical modeling capability called “design distributions” for this purpose. Design distributions capture the variations seen as the transistor design evolves. Realistic responses from the model are introduced by model parameter correlations that mimic strategies used by device designers. An example of a device design tradeoff which can be modeled by design distributions is the readjustment of threshold voltage to compensate for a device capacitance target that is too high. In this case the device designer compensates using the overall goal of maintaining circuit performance.

The new modeling capability has been developed for Monte Carlo analyses, which reflect the probabilistic nature of the uncertainty, however it can also be implemented through fixed corner model parameter sets. In this work it was chosen to optimize design distributions for logic performance, but the design distribution principles can be applied to other classes of circuits with different performance goals. Examples include analog applications, in which the primary constraint could be to preserve characteristics such as $G_m$ and $G_{ds}$, or SRAM applications, where cell stability is key.

Concurrent design has also been addressed in [2]; however, that work presented a method for technology optimization rather than a tool incorporated into a compact model for the evaluation of an arbitrary circuit. The design distribution approach is distinct from conventional compact modeling techniques that simulate the uncertainty in manufacturing processes (e.g. [3]). It is possible to implement device design uncertainty into statistical circuit design optimization methods such as reference [4] however the authors are unaware of any work in this area.

2 Formulation

The general formulation for design distributions is a set of constraint equations that relate device parameter variation to the variation in other device parameters, the transistor operating conditions, and the model fit.

$$
\Delta P_j = f(\Delta P_k, q_l, m_{i}^{\text{ext}})
$$

where $P_j$ and $P_k$ are device parameters such as saturated threshold voltage $V_{T}^{\text{(sat)}}$, $q_l$ are operating conditions such as bias, device geometry, etc., $m_i$ are model parameters such as mobility $\mu_0$ (the superscript $\text{ext}$ indicates as-extracted values), and $f$ is a general transform that includes the compact
model equations as well as arbitrary relationships between the arguments to \( f \). In practice, the model equations can be isolated from \( f \) and the constraints can be implemented separately using a set of linear equations for model parameters that couple statistical variation using constant weighting factors. These linear approximations can be represented as

\[
M_i = m_i^{ext} + \sum_k w_{ik} d_{ik}
\]  

(2)

where \( M_i \) are the model parameters actually used in the compact model equations, \( w_{ik} \) are the weights, and \( d_{ik} \) are statistical (design) distributions.

Specific device design constraints can be further understood in terms of the device parameters and the model parameters that are adjusted. Table 1 lists the device parameters used in this work to specify the transistor design window. The quantities in Table 1 were chosen for their impact on circuit design, ease of measurement, and ease of adjustment through transistor design changes. Note that the tolerance on circuit performance (\( d_{pref} \)) is asymmetric, meaning that the average circuit performance is explicitly constrained to be at least as high as the nominal model regardless of any device change occurring during development. Also included in Table 1 are some sample values for tolerances that might be specified by the device design and the principle model parameters used to fit the device parametric uncertainty. The tolerances are an input to the methodology and are set by the transistor designer who estimates the maximum likely device parametric variation over the transistor development cycle. As the ULSI process matures, the transistor design uncertainty decreases and the design distribution tolerances are reduced.

Table 2 shows model parameters that are modified by the design distribution algorithm. The parameter names reflect the model type that is used to demonstrate the design distribution concept. The model used in this work is the recently-approved Compact Model Council (CMC) SOI model standard, BSIMPD[5], [6]; however, the method is general and can be applied to other models as well.

### Table 1: Device parameters used to define the transistor design space

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Definition</th>
<th>Tolerance</th>
<th>Model parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_{pref} )</td>
<td>Circuit performance as measured by stage delay</td>
<td>0%0-10%</td>
<td>multiple</td>
</tr>
<tr>
<td>( I_{DS(sat)} )</td>
<td>Drain-to-source current at ( V_{DS}=V_{GS}=V_{DD} )</td>
<td>+/-10%</td>
<td>u0, vsat</td>
</tr>
<tr>
<td>( I_{off} )</td>
<td>Drain-to-source current at ( V_{DS}=V_{DD}, V_{GS}=0 )</td>
<td>0/-1.5X</td>
<td>cdsdc</td>
</tr>
<tr>
<td>( \beta_{NP} )</td>
<td>Ratio of NMOS ( I_{DS(sat)} ) to PMOS ( I_{DS(sat)} )</td>
<td>+/-10%</td>
<td>u0, vsat</td>
</tr>
<tr>
<td>( C_{gate} )</td>
<td>Total gate capacitance with ( V_{GS}=V_{DD}, V_{DS}=0 )</td>
<td>+/-5%</td>
<td>cvt terms, dlc</td>
</tr>
<tr>
<td>( C_{ov} )</td>
<td>Gate-to-diffusion capacitance at ( V_{GS}=V_{DS}=0 )</td>
<td>+/-10%</td>
<td>cgso, cgdo, cgsl, cgdl</td>
</tr>
<tr>
<td>( C_j )</td>
<td>Diffusion-to-body capacitance at ( V_{BS}=0 )</td>
<td>+/-20%</td>
<td>cjswg</td>
</tr>
<tr>
<td>( V_T (L rollup) )</td>
<td>( V_T (sat) ) rollup from long L to nominal L</td>
<td>+/-10%</td>
<td>nlx</td>
</tr>
<tr>
<td>( V_T (L rolloff) )</td>
<td>( V_T (sat) ) rolloff from nominal L to minimum L</td>
<td>+/-10%</td>
<td>dvt0</td>
</tr>
<tr>
<td>( V_T (W rolloff) )</td>
<td>( V_T (sat) ) voltage from wide W to minimum W</td>
<td>+/-10%</td>
<td>k3</td>
</tr>
</tbody>
</table>

### Table 2: Model Parameters adjusted by design distributions

<table>
<thead>
<tr>
<th>BSIMPD parameter</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>vsat</td>
<td>Saturation velocity</td>
</tr>
<tr>
<td>cdscd</td>
<td>Subthreshold slope ( V_{DS} ) sensitivity</td>
</tr>
<tr>
<td>nlx</td>
<td>( V_T ) (L rollup) adjustment</td>
</tr>
<tr>
<td>dvt0</td>
<td>Short channel effect adjustment</td>
</tr>
<tr>
<td>k3</td>
<td>Narrow width effect adjustment</td>
</tr>
<tr>
<td>cjswg</td>
<td>Gate-edge junction capacitance</td>
</tr>
<tr>
<td>ladj</td>
<td>Non-BSIMPD; equals dlc - lint</td>
</tr>
<tr>
<td>cgso, cgdo, cgsl, cgdl</td>
<td>Overlap capacitance</td>
</tr>
<tr>
<td>vth0</td>
<td>long-channel ( V_T )</td>
</tr>
<tr>
<td>u0</td>
<td>long-channel mobility</td>
</tr>
</tbody>
</table>

A subset of the device parameter variation from Table 1 is implemented without cross-correlations and therefore are determined by a single distribution and a single weighting factor. The model parameters from Table 2 in this category are cdscd, dvt0, cjswg, nlx, and k3. Their general constraint relation is

\[
M_i = m_i^{ext} + w_i d_i
\]

(3)

where \( i \) is one of cdscd, dvt0, cjswg, nlx, or k3, \( w_i \) is a weighting factor and \( d_i \) is a Gaussian distribution. Gaussian distributions were chosen to reflect the natural intention of the device designer to minimize change.

A second set of device parameters are controlled by multiple weighting factors because they are constrained by additional device parameters that are affected by other varying model parameters. These are long-channel threshold voltage, which is constrained by rollup and rolloff requirements

\[
v_{th0} = (v_{th0})^{ext} + w_{nlx} d_{nlx} + w_{dvt0} d_{dvt0}
\]

(4)

and the capacitance parameters \( C_{gate} \) and \( C_{ov} \). The capacitance parameters have multiple constraints because \( C_{gate} \) is defined as the sum of inversion capacitance (which is modu-
lated by \( l_{adj} \) and \( C_{ov} \).

\[
\begin{align*}
\text{l}_{adj} &= (\text{l}_{adj})^{ext} + w_{l_{adj}} d_{l_{adj}} + w_{C_{ov}}^{l_{adj}} d_{C_{ov}} \\
\text{Cov} &= (\text{Cov})^{ext} + w_{C_{ov}} d_{C_{ov}} + x_{C_{ov}} d_{C_{ov}}^{l_{adj}} + w_{l_{adj}} d_{l_{adj}} + x_{l_{adj}}^{l_{adj}} d_{C_{ov}}
\end{align*}
\]  

(5)

(6)

where here Cov refers to one of cgso, cgdo, cgsl, and cgdl since these terms contribute approximately linearly to total overlap capacitance. Weights and distributions with a superscript establish cross-correlation links. An "x" indicates a weight or distribution superscript that references the opposite device type (e.g., a PMOS term that is used in the NMOS equations). The references to the opposite type are needed to satisfy the cross-correlations inherent in the beta ratio \( \beta_{N/P} \) and stage delay \( d_{pref} \) device parameters.

The most complex constraints relate the \( I_{DS}(\text{sat}) \), \( \beta_{N/P} \), and \( d_{pref} \) device parameters. The term \( d_{pref} \) is computed from the sum of the change in N-type delay plus the change in P-type delay. The change in delay is approximated by the change in the total device capacitance (taken as a weighted sum of \( C_{gate} \), \( C_{ov} \), and \( C_{jswg} \)) divided by \( I_{DS}(\text{sat}) \). Therefore, the subspace of \( d_{pref} \) variation is bounded by limits on all of the total capacitance terms, on \( I_{DS}(\text{sat}) \), and on \( \beta_{N/P} \). The variation needed to satisfy these limits is produced by modulating the mobility \( u_{0} \) and the saturation velocity \( v_{sat} \).

\[
\begin{align*}
\text{u}_{0} &= (\text{u}_{0})^{0} + w_{u_{0}} d_{u_{0}}^{w} + x_{u_{0}} d_{u_{0}}^{x} + w_{v_{sat}} d_{v_{sat}}^{w} + x_{v_{sat}} d_{v_{sat}}^{x} \\
\text{v}_{sat} &= (\text{v}_{sat})^{0} + w_{v_{sat}} d_{v_{sat}}^{w} + x_{v_{sat}} d_{v_{sat}}^{x} + w_{u_{0}} d_{u_{0}}^{w} + x_{u_{0}} d_{u_{0}}^{x}
\end{align*}
\]  

(7)

(8)

Special distribution definitions (eqns 9-15) are used to link \( u_{0} \) and \( v_{sat} \). The subscript \( i \) refers to either \( u_{0} \) or \( v_{sat} \), and the generality of these linkage equations demonstrates that the nature of the links is contained mostly in the weight values.

\[
\begin{align*}
\Gamma_{1i} &= w_{i}^{C_{ov}} d_{C_{ov}} + w_{i}^{C_{jswg}} d_{C_{jswg}} + w_{i}^{l_{adj}} d_{l_{adj}} \\
\Gamma_{2i} &= x_{i}^{C_{ov}} d_{C_{ov}} + x_{i}^{C_{jswg}} d_{C_{jswg}} + x_{i}^{l_{adj}} d_{l_{adj}} \\
\Gamma_{3i} &= |w_{i}^{C_{ov}}| + |w_{i}^{C_{jswg}}| + |w_{i}^{l_{adj}}| \\
\Gamma_{4i} &= |x_{i}^{C_{ov}}| + |x_{i}^{C_{jswg}}| + |x_{i}^{l_{adj}}| \\
\Gamma_{i} &= \frac{\Gamma_{1i} + \Gamma_{2i}}{\Gamma_{3i} + \Gamma_{4i}} \\
d_{i}^{w} &= d_{i}^{w} (1 - \max(0, \Gamma_{i})) + \Gamma_{i} \\
d_{i}^{x} &= 2d_{i}^{x} \min(0.5, 1 - |d_{i}^{x}|)
\end{align*}
\]  

(9)

(10)

(11)

(12)

(13)

(14)

(15)

3 Results

As previously mentioned, the design distributions are demonstrated in this work through circuit simulation using the BSIMPD model. Figure 1 illustrates the influence of the constraint equations for an NMOS device in a Monte Carlo simulation (PMOS shows the same behavior). In this plot of \( C_{inv} \) versus \( C_{ov} \), the sloped edges of the data cloud reflect the constraints on \( C_{ov} \). If the terms were uncorrelated, the plot would be a rectangle. In this figure and Figure 2, the statistical distributions have been redefined with uniform distributions to emphasize the correlated model behavior. The constant vertical spacing between the diagonal edges reflects the \( C_{inv} \) constraint, while the decreasing trend of \( C_{inv} \) while \( C_{ov} \) is increasing is indicative of the constraint on \( C_{gate} \).

Figure 2 plots the mobility parameters in the model for an NMOS/PMOS pair in a Monte Carlo simulation to illustrate the cross-correlations that are required to satisfy multiple constraints. The upper horizontal limit and the right-most vertical limit are set by the current constraints on maximum N/P \( I_{DS}(\text{sat}) \). Since the current depends on the mobility, lines of constant \( \beta_{N/P} \) are related to the \( u_{0} \) (NMOS)/\( u_{0} \) (PMOS) ratio. Therefore the edges defined by positive slopes above and below the data cloud reflect the maximum and minimum \( \beta_{N/P} \) limits. Lines of constant performance (with \( \beta_{N/P} \) varying) can then be drawn approximately perpendicularly to the \( \beta_{N/P} \) edges. The minimum performance (increase) line goes through the (0,0) percentage change point on the delta \( u_{0} \) (N)-\( u_{0} \) (P) plot while the maximum performance increase is the point in the upper right-hand corner. Variations in the capacitive terms move the minimum and maximum performance limits, changing the position and shape of the allowed region.

Figure 3 shows a Monte Carlo simulation of a 15-stage NAND ring oscillator with parasitic circuit elements and illustrates how design distributions restrict the bulk of the simulation points to faster performance, and demonstrates the validity of the model formulation. For comparison, a CV/I-like metric that is similar to the model is also calculated and shows good matching with the simulation.

Finally, some circuits are not practical to simulate using statistical analysis. For these cases, a set of device distribu-
design. This uncertainty can be modeled in circuit simulation by adding design distributions to the compact model. Design distributions adjust model parameters to vary within a parametric space that is defined by the transistor designer while meeting the overall criterion of maintaining minimum logic performance. Simulation results are given that demonstrate that design distributions function as expected.

REFERENCES


4 Summary

Concurrency between ULSI chip designs and ULSI device/process technology introduces uncertainty into product
design corner files can be developed that bound statistical device distribution behavior (Figure 4). This plot shows the bounding corners for two conditions: 1) all capacitances set to their nominal value (with currents varying) and 2) all capacitances set to half of their maximum value. Corners are developed by skewing various design distributions to give specific capacitance or current limits.