Digital Design at a Crossroads
How to Make Statistical Design Methodologies Industrially Relevant

Ulf Schlichtmann, Manuel Schmidt
Technische Universität München
Munich, Germany
{ulf.schlichtmann | manuel.schmidt}@tum.de

Harald Kinzelbach
Infineon Technologies AG
Munich, Germany
Harald.Kinzelbach@infineon.com

Michael Pronath, Volker Glöckel
MunEDA GmbH
Munich, Germany
{michael.pronath | volker.gloeckel}@muneda.de

Manfred Dietrich, Uwe Eichler, Joachim Haase
Fraunhofer IIS/EAS Dresden
Dresden, Germany
{manfred.dietrich | uwe.eichler} @eas.iis.fraunhofer.de

Abstract—Statistical analysis is generally seen as the next EDA technology for timing and power sign-off. Research into this field has seen significant activity started about five years ago. Recently, interest appears to have fallen off somewhat. Also, while a lot of focus has been put on research fundamentals, extremely few applications in industry have been reported so far. Therefore, a group including Infineon Technologies as a leading semiconductor IDM and various universities and research institutes, as well as an EDA provider has tackled key challenges to enable statistical design in industry in a publicly funded project called “Sigma65”. Sigma65 strives to provide key foundations to allow a change from traditional deterministic design methods to future design methods driven by statistical considerations. The project starts with statistical modeling and optimization of library components and ranges to statistical techniques for designing ICs on gate level and higher levels. In this paper, we present some results of this project, demonstrating how the interaction between industrial perspective, research institutions and EDA provider enables solutions which are applicable already in the near future. After an overview of the industrial perspective of the current situation in dealing with variations recent results on both statistical timing and power analysis will be given. In addition, recent research advances on fast yield estimation concerning parametric timing yield will be given.

Keywords: Simulation, digital IC design, statistical timing analysis, statistical power analysis

I. INTRODUCTION

Circuit performances such as delay and leakage depend on the process parameters, supply voltage and temperature (PVT). For every new process generation, increasing process variations (certainly relative to their nominal values and partly also absolutely) and also an increasing sensitivity of performances to these variations can be observed [4]. This calls for new methodologies of modeling and analyzing the impact of PVT variations on circuit performances. On this foundation, advanced optimization methods can then be developed.

These variations are classified into global and local variations. Global variations affect each device on a die equally and only vary from die to die. Traditionally, these variations used to be well captured by corner cases. Different sets of process parameters are used to define slow and fast corners. A circuit is then designed such that the timing constraints are met in all corners. However, it becomes increasingly difficult to find a small number of adequate corners. Some effects that appear in recent process generations, such as the inverted temperature effect, demonstrate that the extreme values of circuit performance may not appear at the standard slow and fast corners. The increasing number of varying parameters results in a higher complexity of finding the correct corners – experimental data show e.g. that each dimension of different metal layers vary independently; therefore a thorough analysis needs to include a number of corners that is exponential in the number of metal layers. Different cells of the same design can have a different slow/fast corner possibly leading to a different corner for each cell. Additionally, it is already near to impossible for current designs to fulfill all requirements for these extreme corners as they are overly pessimistic.

The second class is local variations which affect each device on a die independently. The classical corner-based design methodology cannot handle the effects of local variations. Such variations are currently being considered in industrial design flows by the introduction of on-chip variation (OCV) factors. The basic concept is to increase the computed delay of a data path by an empirically determined OCV factor and decrease the delay of the corresponding clock path for a setup-time check, and vice versa for a hold-time check. However, this approach results in a proportional increase of the delay with the length of the path even though the local variations cancel out for longer paths. Thus, the OCV approach introduces even more pessimism into the design. Furthermore, the empirical determination of OCV factors is very tricky. If data and clock paths share some components, then a common path pessimism removal technique is applied. Meanwhile, some tool vendors offer the option to define OCV factors specifically depending on the length of paths and also their geometric extension on a die. Various IDMs started to evaluate or use such concepts. However, it appears that such advanced OCV concepts tend to end up as one complex band-aid attached on top of another insufficient band-aid. As shown in Fig. 1, an industrial
comparison of timing-path arrival times reported by classical STA with OCV shows considerable deviations from highly accurate SSTA reference results.

Figure 1. Relative deviation (in %) of timing-path arrival times reported by STA + OCV compared to highly accurate SSTA.

As process variations are forecast to get even worse for future process generations, it is vital to move on towards more realistic methods for their consideration in order to avoid the waste of area, power and engineering time. The aim should not be to gain smaller margins but to derive margin-less methodologies that are capable of capturing variation effects in a more realistic manner. An improvement in the corner-based method together with OCV might help in the near future but at the cost of more corners and more margins. But even with this extreme effort the risk of missing the real worst case can not be excluded completely. Therefore, the industry increasingly looks towards statistical design methods to provide a real breakthrough in dealing with process variations.

The publicly funded project Sigma65 strives to provide key foundations to allow a change from traditional deterministic design methods to future design methods driven by statistical considerations. The project starts with statistical modeling and optimization of library components and will range to statistical techniques for designing ICs on gate level and higher levels.

Following this brief description of the industrial perspective regarding the handling of variations, we now describe recent advances in statistical timing analysis considering waveforms. This is followed by an approach for fast yield estimation concerning parametric timing yield using Monte Carlo methods enhanced by Importance Sampling. We then give a brief overview of the industrial perspective considering leakage power variation, followed by the description of a new approach to statistical power estimation.

II. TIMING ANALYSIS

The consideration of the statistical nature of PVT variation for the delay computation leads to statistical static timing analysis (SSTA). The challenges mentioned above are clearly obvious to the vendors of EDA tools and first engines for SSTA are commercially available. However, such tools will always employ simplifications in order to analyze circuits in acceptable time. Thus, the results of these tools have to be evaluated before using them in a production design flow. Therefore, a reliable reference is needed for timing analysis. In the case of deterministic STA, this need can be addressed by a method based on stage-wise analog simulation. For SSTA, Monte Carlo analysis is usually proposed, but the size of most circuits inhibits this option due to runtime constraints. In the following, we propose a different solution, following a similar concept as for deterministic STA. In addition to the nominal simulation, the sensitivities to the process parameters are propagated through the circuit. Such a methodology is presented in the following.

A. Significance of Exact Waveforms

The miniaturization of integrated circuits results in a larger influence of effects that could be safely neglected for the timing analysis in previous process technology generations. Therefore, the models used for timing analysis are gradually advancing towards more complex implementations. Three major simplifications are used in traditional timing analysis, which will be addressed in this work: 1) modeling the driving gates by look-up tables dependent on slope and load; 2) abstracting the waveforms by arrival time and slope of a linear ramp; 3) reducing the load comprised of the gates in the fanout as well as the interconnect to a single capacitance in order to use the look-up tables. These three simplifications have yielded sufficient accuracy for previous process technologies.

However, the influence of currently neglected effects has increased in recent process technologies and is forecast to increase further. Thus, these simplifications lead to unacceptable timing results. One major influence is the actual shape of the waveforms. Due to the increasing resistive behavior of interconnects the waveforms differ significantly from standard shapes leading to large errors in the delay of the gates [1]. To address this problem in static timing analysis, current source models of the driving cells have been proposed, e.g. [2]. These models are capable of considering the shape of the waveform but more complex effects like non-linear capacitances at the connected gate due to the Miller effect can only be approximated. The most accurate approach was published in [3], which used analog simulations of stages consisting of the driving gate and the interconnect and receiving gates at the output. Thus, the entire netlist including parasitics extracted from the layout can be considered by this approach. This approach will be used in the following as a nominal analysis and basis for the statistical analysis.

Statistical Variations -- On top of the issues of deterministic timing analysis, one of the most prominent topics is the statistical variation of process parameters [4]. With increasing influence of random, local variations the corner-based methods become increasingly pessimistic and chip area and power are wasted due to lack of better analysis tools [5]. Therefore, methods for statistical static timing analysis (SSTA) were proposed, considering the delays and arrival times as random variables [6]. However, the consideration of waveform variation in SSTA remains unclear. The aim of this work is to provide a methodology to determine the sensitivities of voltages at arbitrary points of the output waveform on process parameters. The methodology is presented in a path-based version as the influence of other approximations like the
maximum computation in block-based analysis should be excluded for higher accuracy. The sensitivities allow the analysis of the waveform variation depending on the distributions of the process parameters.

These sensitivities have to incorporate the correlation between process parameters. Process parameters of devices can be correlated if the devices are located in close vicinity. The arrival time at the input of a gate is then correlated to the process parameters of the gate itself. This correlation is considered in the proposed method using the representation of a weighted sum of process parameters for signal variation.

### B. Stage-wise analog Simulation

The proposed methodology relies on the partitioning of a path into stages comprised of the driving gate, the entire interconnect and all connected receiving gates in the fan-out. Beginning from the input, each stage is simulated by a SPICE-like simulator, which yields the nominal waveform for each stage. The next step is to determine the waveform variation around this nominal waveform resulting from a variation of the process parameters.

The waveform variation is modeled as the variation of voltages of the waveform. These variations of the voltages are represented by a weighted sum of process parameters usually referred to as a canonical sum:

\[
u(t, p_0 + \delta p) \approx u(t, p_0) + \sum \chi_{\nu}(t) \delta p_i
\]

Wherein \(u(t, p_0)\) is the nominal output waveform for the nominal case of process parameters \(p\), \(\chi_{\nu}(t)\) are the sensitivities of the voltage at time \(t\) to the process parameter \(p_i\), and \(\delta p_i\) are the variations of the process parameter \(p_i\) around the nominal value.

The variation of the output waveform of one stage is divided into two contributions: 1) The variation caused by the variation of the process parameters of the stage itself and 2) the variation of the input waveform. In order to consider these two contributions, the sensitivities of the output voltage to the process parameters of the driving cell as well as the sensitivities of the output voltage to the input voltage have to be determined. The employed simulator can compute the required sensitivities during nominal simulation using adjoint network analysis. The two contributions are added to obtain the canonical sum as in (1) at the output of the stage. The path is traversed and the nominal output waveform of one stage is used as an input for the next stage. The canonical sums for all voltages of the waveform are also propagated towards the output of the path.

One major obstacle is the fact that the main influence of parameter variation is a shift of the output waveform resulting in a large variation of the voltage and the voltage variation can exceed the linear region. Therefore, an expanded representation was developed, which separates the time shift from the voltage variation and propagates the canonical sum of the time shift separately. At the end of the path, the two contributions are added in order to obtain meaningful quantities like the variation of the arrival time.

The validity of this approach relies on the linear assumption of the relationship between voltage/arrival time and process parameters. This linearity can be sufficiently assumed when separating the arrival time from the voltage variation as described above.

### C. Timing Results

The proposed method was tested on extracted paths of an industrial design with all post-layout parasitics. In order to evaluate the accuracy of the sensitivities, a SPICE Monte Carlo simulation was performed and the standard deviation of the delay \(\sigma_d\) was compared to the \(\sigma_d\) resulting from the linear combination of process parameters according to the computed sensitivities.

The variation of 4 global and 2 local parameters was obtained from the productive 90nm framework of a major IDM. Gaussian random variables are assumed for the process parameters, but the proposed method is not restricted to any particular distribution.

Fig. 2 shows how the histogram of the Monte Carlo simulation matches the probability density function obtained by adding the Gaussian random variables according to the canonical sum at the output of a path.

![Figure 2: Comparison of Monte Carlo histogram and Gaussian probability density function obtained by the proposed method](image)

Table I shows the number of gates, the total size of the interconnect, the error in the delay, the runtime of the proposed method, and the runtime of the Monte Carlo analysis for different paths.

<table>
<thead>
<tr>
<th>#gates</th>
<th>#res/#cap</th>
<th>err σ_d</th>
<th>rt</th>
<th>Rt(MC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>237/215</td>
<td>3%</td>
<td>923 s</td>
<td>614 · 10^4 s</td>
</tr>
<tr>
<td>35</td>
<td>241/219</td>
<td>5%</td>
<td>927 s</td>
<td>608 · 10^3 s</td>
</tr>
<tr>
<td>35</td>
<td>241/219</td>
<td>5%</td>
<td>927 s</td>
<td>608 · 10^3 s</td>
</tr>
<tr>
<td>35</td>
<td>241/219</td>
<td>5%</td>
<td>927 s</td>
<td>608 · 10^3 s</td>
</tr>
<tr>
<td>50</td>
<td>108/44</td>
<td>5%</td>
<td>1400 s</td>
<td>1.2 · 10^4 s</td>
</tr>
</tbody>
</table>

It can be seen that the results of the proposed method are very close to analog simulation but with a speedup of 600 to 800. Therefore, the proposed method is well suited as a highly accurate reference tool to analyze the accuracy of commercial tools as well as to perform a detailed analysis on a limited number of paths which have been identified as being critical by
III. RESEARCH ADVANCES ON FAST YIELD ESTIMATION

For analog circuits yield estimation methods as well as yield optimization methods [7] are well established [8]. They cannot be applied directly to digital functional blocks, as the characteristics of the path delay needs to be taken into account. Due to the maximum calculation of the delay of convergent paths, the performance function is non-linear. This effect can cause non-Gaussian distributions. Additionally, the statistical variance of the delays is often negatively affected by operating conditions – especially the load. To estimate the parametric yield of a design in the relevant region above 95%, the estimation of mean value and variance is therefore not sufficient. Instead, the distribution of the delay has to be analyzed exactly at the boundary. In this region deterministic yield estimation methods have clear advantages compared to robust stochastic estimation methods like Monte Carlo. However, regarding digital functional blocks the consideration of the special shape of the performance functions is missing. This complicates the application of standard methods.

Known deterministic and statistic yield estimation methods are adjusted to the challenge of digital design. Yield estimation methods are developed which can be applied first-time to digital circuit design handling the non-linear effects. The advantage of Monte Carlo methods over deterministic methods is that their convergence rate is independent of the number of statistical parameters. Please note that considering local variations, statistical parameters are introduced per device. But its simulation effort is increasing with accuracy. To enable fast yield estimation, Variation Reduction Techniques are used to develop new methods estimating the yield with same accuracy and less simulation effort compared to standard Monte Carlo method.

One promising technique is using Importance Sampling Analysis with a Defensive Mixture Distribution [9]. It is used to enrich the numbers of samples around the specification boundary which is located at the boundary of the process distribution for high yield designs. This can be achieved by using an artificial distribution taking just half of the samples from the original process distribution and the other half from the process distribution shifted towards the specification boundary. Applying this technique to an inverter shows that the method is most efficient if the process distribution is shifted to a point on the specification boundary, the worst-case point. This splits the samples of the shifted distribution half and half in good and bad. The worst-case point is the one point on the specification boundary with closest distance to the mean value of the process distribution.

To estimate the yield with a standard Monte Carlo method the percentage of the good samples fulfilling the specification is determined. Using Importance Sampling Techniques, the samples need to be weighted to correct the difference between the artificial distribution and the process distribution. The weighting increases the variance of the yield estimator as outliers contribute with high weights to the yield estimate. Estimating the percentage of the bad samples (100% - yield) contributing with small weights instead, showed an increase of estimation accuracy by up to 350%.

To apply the Importance Sampling yield estimation technique to a path of a digital circuit, the worst-case point needs to be calculated. This can be done with deterministic methods using WiCkeD [8]. The simulation effort of the deterministic worst-case point calculation is increasing with the number of statistical parameters. This is quickly consuming the benefit of the Importance Sampling method or even taking more simulation effort as the standard Monte Carlo yield estimation. Therefore, a method to approximate the worst-case point of the path delay has been developed. In a preparation step the worst-case points are predetermined for each inverter, NAND gate, NOR gate etc. In general, rising and falling edges have to be distinguished as their delays and therefore their worst-case points are different. Lacking any specification for the gate delays, worst-case points are calculated with a 3 sigma distance. The worst-case points of the gate delays are calculated once for a gate library. Based on this, the worst-case point of a path delay can be approximated by a liner combination summing up the worst-case points of the gate delays on the path. While simulations are necessary in the preparation step to calculate the worst-case points of the gate delays, the worst-case point of a path delay can be approximated without any simulation just by calculation.

Applying the Importance Sampling method with an approximated worst-case point to a path with 3 inverters, a NOR and a NAND gate of an industrial design showed similar accuracy compared to the same experiment using calculated worst-case points. So the efficiency improvement of the Importance Sampling technique is maintained completely by using the worst-case point approximation.

Applying the overall Importance Sampling approach to that path shows a 400% efficiency improvement over standard Monte Carlo yield estimation for a specification 3σ away from the mean of the process distribution. The efficiency is increasing exponentially with higher yields.

IV. STATISTICAL LEAKAGE ANALYSIS

![Figure 3. Comparison Monte Carlo simulation vs. log-linear leakage model for a 90nm inverter at 135°C](Image)

As described above for timing analysis, leakage power also varies due to PVT variation. But there are significant differences between leakage power and timing. Firstly, the leakage power shows an extremely wide range of variation, as
it depends exponentially on some physical process parameters. Thus, it is less appealing to linearize around the nominal point. Experiments showed that the distribution of the leakage power can be approximated at first order by a log-normal distribution. However, for some cases, this assumption leads to inaccurate results. Fig. 3 shows a comparison of direct Monte Carlo simulation data versus a linearized logarithmic model for an inverter in 90nm technology at 135°C. For the log-normal leakage distribution used in the literature a simplified model to be valid, all points would need to be on a straight line. In the given case (and many other cases) however, obviously there is a significant deviation. A second order model shows a better approximation behavior, but for a correct modeling even higher order nonlinearities have to be taken into account.

After the determination of the modeling parameters for single cells, the next challenging task is the analysis of the entire circuit. The leakage analysis in the nominal case is available meanwhile, what remains to be done is the extension of this analysis to include the variation effects, leading to a full Statistical Leakage Power Analysis (SLPA). First ideas along these lines are available, based on a combination of response surface models and Monte Carlo simulations.

V. STATISTICAL POWER ESTIMATION

Power dissipation of a digital system is composed of static and dynamic power dissipation. The static power dissipation corresponds to the leakage currents which flow through the system’s components in steady-state. The dynamic power dissipation depends on short-time cross currents and the power which is necessary to charge capacitances of devices and wires during signal changes. Both power contributions depend on process variations that influence the transistor parameters. Static power distributions can be estimated taking into consideration an in general nonlinear dependency between process parameters and leakage currents. If this dependency can be described by special polynomial forms, an analytical evaluation can be applied only in special cases [10, 11, 12]. An alternative is a Monte Carlo simulation applied on the response surface models which characterize the leakage current dependency on the parameter variations.

Monte Carlo simulation could also be applied to carry out a statistical dynamic power analysis based on transistor level netlist descriptions using SPICE-like simulators. However, the required effort is in the case of larger systems beyond the accepted limits. Acceleration can be achieved replacing transistor level descriptions by behavioral models. Parameter variations influence both – transient behavior that affects charging and de-charging of capacitances and power which is necessary to load and de-load these capacitances. Thus, a method was developed to run Monte Carlo simulations using behavioral models and estimate dynamic (and also static) power in each simulation run [13]. VHDL was used to establish the behavioral models. The SAE J2748 standard [14] was applied to describe the parameter variations.

A. Modeling Approach

Non-Linear Delay and Power Models (NLDM, NLPM) describe the dependency of the delay, slope and power of each output signal of a digital cell with respect to the slope $S$ of the switching input signals and the associated output load capacitance $C$. These dependencies are provided in the digital design flow by tables that characterize $y_i=f(C,S,p_0)$ where $y_i$ is either a delay, slope or power characterizing an input-to-output transfer and $p_0$ is a vector of $n$ nominal real-valued parameters that influence delays, slopes and power. These nominal values are fixed for a given technology. Thus, the functions $f_i$ can be characterized with values at discrete data points $(C_i,S_i)$. In the case of a statistical analysis the parameters can vary and are characterized by random variables. The dependency of delay, slope and power values on the parameters can be expressed in the first approximation using first order sensitivities:

$$y_i = f_i(C, S, p_0) + \sum_{j=1}^{n} \frac{\partial f_j}{\partial p_j} \Delta p_j$$

At the beginning of each Monte Carlo run new $\Delta p_j$ are tossed up which results in a varied set of tables for slope, load and power characterizing $f_i(C, S, p_0+\Delta p)$.

![VHDL Cell model structure for statistical power simulation](image)

For each input signal change the extended cell models (see Fig. 4) determine delay and slope of the affected output signal based on equation (2) using the pre-calculated, varied lookup tables. The required arguments $S$ and $C$ are provided in different ways: the load capacitance $C$ is assumed to be constant in the current implementation and is therefore passed as generic parameter and the input slope $S$ is carried together with the logic value in an overloaded signal data type. At the same time, a corresponding power value is determined from the tables and either applied to a global variable or passed to a post-processing step allowing to create a simplified waveform of the total current.

B. Design Flow Aspects

Based on the standard NLDMLNLPMLib file information, extended by parameter sensitivities, VHDL cell models and packages which store the delay, slope time and power information are uniquely created. For a system under investigation, load capacitances are determined based on the associated SPEF (Standard Parasitic Exchange Format) file information. Thus, cell models and the netlist to simulate a system are available. In the instantiation phase of each
simulation run the parameter values can be determined based on their probability density functions. In this way, the Monte Carlo simulation runs only require to toss up parameter values at the beginning of a simulation (full arrow in Fig. 5). This is a time-saving procedure compared to a modification of the .lib files for each Monte Carlo simulation run (dashed arrow in Fig. 5).

The difference of the number of evaluated toggles of SPICE-like and digital simulations were carried out in order to compare simulation effort and accuracy.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Time (abs)</th>
<th>Time (rel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPICE-like simulation</td>
<td>10 min</td>
<td>30000</td>
</tr>
<tr>
<td>VHDL simulation (proposed model)</td>
<td>0.02 s</td>
<td>1</td>
</tr>
<tr>
<td>Standard Verilog simulation</td>
<td>0.009 s</td>
<td>0.45</td>
</tr>
</tbody>
</table>

The difference of the number of evaluated toggles of SPICE-like simulation and VHDL simulation with the extended model was about 10 % for the example.

VI. CONCLUSION

The increasing relative variation of process parameters and increasing sensitivities to these parameter variations call for a statistical analysis of performance parameters. The two statistical parameters delay and leakage have been addressed in this paper. For SSTA, a highly accurate reference was proposed. The statistical power analysis is still in early stages but some first ideas seem to give an impression of the development directions to come. In this paper, we have presented statistical methods for timing and power analysis and simulation of digital blocks with thousands of gates. Innovative approaches for exact statistical modelling of library elements and connections are the basis of these methods.

The developed methods have to extend regarding additional requirements of new process technologies. Furthermore, the methods will be improved for the application to larger digital blocks as well as to analog and mixed-signal circuits. The goal is the analysis of the whole integrated circuit and a sufficient estimation of the yield. For future development three major questions shall be stated here. Firstly, the interaction between timing and power is not properly addressed. This yields much optimization potential and should be used. Secondly, the variation aware implementation should be addressed. On one hand the process itself could be optimized but on the other hand the implementation could take the variations into account and thus reduce the overall impact of the variations on the circuit performance. Thirdly, the statistical analysis should be extended to higher levels of the design flow. A variation aware high level synthesis can further optimize the statistical behavior of the final implementation. With such tools the rising impact of variation in future process generation can be addressed - but without them it will be impossible to develop complex systems in the future.

ACKNOWLEDGEMENT

These activities are supported by the German Ministry of Education and Research (BMBF) within the project Sigma65 (ID 01M3080). The content is the sole responsibility of the authors.

REFERENCES