On the Futility of Statistical Power Optimization

Jason Cong[1,2] Puneet Gupta John Lee

[1]Computer Science Department
University of California at Los Angeles
[2]California NanoSystems Institute
e-mail: cong@cs.ucla.edu

Electrical Engineering Department
University of California at Los Angeles
e-mail: puneet@ee.ucla.edu email: lee@ee.ucla.edu

Abstract

In response to the increasing variations in integrated-circuit manufacturing, the current trend is to create designs that take these variations into account statistically. In this paper we try to quantify the difference between the statistical and deterministic optima of leakage power while making no assumptions about the delay model. We develop a framework for deriving a theoretical upper-bound on the suboptimality that is incurred by using the deterministic optimum as an approximation for the statistical optimum. On average, the bound is 2.4% for a suite of benchmark circuits in a 45nm technology. We further give an intuitive explanation and show, by using solution rank orders, that the practical suboptimality gap is much lower. Therefore, the need for statistical power modeling for the purpose of optimization is questionable.

I. Introduction

Statistical optimization via circuit sizing has been a active research topic over the last decade. The realization was that the traditional corner-based optimization (see [18]) may be too pessimistic [24], and the trend was to incorporate more and more statistical data into the optimization process.

There are many papers that explore the benefits of adding statistical delay data into the optimization process [16, 19, 21, 9, 11, 22, 10], and there are also a number of papers that also use a statistical power measure [23, 9, 15, 5, 26]. However, to the best of our knowledge, there is no publication that shows the benefits of using the statistical power measure alone.

This brings up an interesting question: How much of the improvement should be attributed to the use of a *statistical* delay model, and how much should be attributed to the use of a *statistical* power model? This question is part of a growing skepticism over the benefits of statistical optimization, and whether they outweigh its costs. Adopting statistical analyses and optimization involves considerable overhead in terms of engineering effort as well as turn-around times. It requires an almost complete overhaul of process modeling, circuit simulation and requires modifying the algorithms for statistical optimization. It is therefore important to do a thorough cost-benefit analysis of statistical optimization compared to conventional deterministic optimization methods.

The related question of statistical delay optimization vs. deterministic delay optimization has already been studied [17, 12, 6]. In [17], the claim is made that corner- or scenario-based optimization is still the most practical because:

1. Intra-die effects are still small
2. There is usually not enough information to do a full-blown statistical analysis
3. The gains of using a full-blown statistical analysis are small.

In [6] the authors quantify the difference between corner based methodologies and full statistical optimization methods. They find that with a 5% variation in stage delay, the full-blown statistical analysis and optimization gives a mere 2% improvement, and a 12% variation gives a 6% improvement over a statistical worst-case corner that employs a guardband. In [12], the tradeoff between yield and circuit delay, and the improvements in slack are examined. Significant improvements are shown for a set of benchmark circuits.

In this paper we focus on the amount of improvement that can be made by using a statistical power measure as an objective for gate sizing and l_{eff} and V_{th} assignment, when compared to the deterministic power measure. The key contributions of the paper are as follows.

- We develop a mathematical programming based framework to estimate the suboptimality gap between different power measures.
- For the common case of discrete gate sizing, we give an intuitive explanation of the suboptimality using solution rank orders.
- We show that the deterministic power measure is a good approximation for the statistical power measures, which means that the deterministic power measure can be used in place of the statistical power measures with very similar optimization results.

It is important to mention that this is independent of the model for the delay. This paper does not give judgements on the difference between statistical delay optimization and deterministic delay optimization, or the difference between static timing analysis and statistical static timing analysis. The delay is only used to generate examples for the suboptimality bounds.

The rest of the paper is organized as follows. The following section outlines the leakage power measures and mod-

TABLE I
NOTATIONS

symbol	meaning
$w_i, \vec{w}$	width of gate i / vector of gate widths
$l_i, \vec{l}$	length of gate i / vector of gate lengths
$v_{t_i}, \vec{v}_{t_i}$	threshold voltage of gate i / vector of threshold voltages
$\vec{z}$	'adjusted gate widths' $z_i = w_i e^{\alpha l_i^2+\beta l_i} e^{-\gamma v_{t_i}}$
$\mathbf{L}_{\text{wid}}$	Within-die variation random variable
$\mathbf{L}_{\text{dtd}}$	Die-to-die variation random variable

TABLE II
MEASURES OF STATISTICAL LEAKAGE POWER

measure	symbol	expression
deterministic	$p_d(\cdot)$	$\sum k_i z_i$
mean	$p_m(\cdot)$	$\sum k_i z_i e^{\eta_i^2(\sigma^2_{\mathbf{L}_{i,\text{wid}}}+\sigma^2_{\mathbf{L}_{\text{dtd}}})/2}$
3σ-quantile[a]	$p_{3\sigma}(\cdot)$	$\sum k_i z_i e^{3\eta_i \sigma_{\mathbf{L}_{\text{dtd}}}}$
mean+3σ	$p_{m3\sigma}(\cdot)$	$p_m(\vec{z}) + \kappa\sqrt{\vec{z}^T S \vec{z}}$

[a]For inter-die variation only

els used in this paper. Section III develops the mathematical framework to estimate the suboptimality gap incurred from using the deterministic power measure in place of the statistical power measures. Section IV gives a simpler explanation of the suboptimality gap using rank orders. Section V compares the results of statistical optimization with the corresponding deterministic optimizations. Finally we conclude with ongoing work in Section VI.

II. Statistical power

In this paper uppercase bold symbols represent random-variables (e.g., $\mathbf{X}$), and uppercase non-bolded symbols represent matrices (e.g., P) or commonly used constants (e.g., V_{dd}). Vector quantities will have arrows above them (e.g., $\vec{x}$), and scalar quantities will be lowercase non-bolded (e.g., p). The principal symbols are summarized in Table I.

A. Models

In this paper, the length of the gate is assumed to be the source of power variations. The leakage random variable is modeled as a lognormal random variable, as in [20]:

$$\mathbf{P}_l = \sum k_i w_i e^{\alpha l_i^2+\beta l_i} e^{-\gamma v_{t_i}} e^{\eta_i(\Delta\mathbf{L}_{\text{dtd}}+\Delta\mathbf{L}_{i,\text{wid}})} \quad (1)$$

The random variables $\Delta\mathbf{L}_{\text{dtd}}$ and $\Delta\mathbf{L}_{i,\text{wid}}$ are assumed to be zero-mean Gaussian random variables that describe the variation of the gate lengths. In this paper, $\sigma_{\mathbf{L}_{\text{dtd}}} = 1$nm and $\sigma_{\mathbf{L}_{i,\text{wid}}} = 0.5$nm which is a good representation of the variation in 45nm.

The dynamic power random variable (ignoring short circuit current), is given by:

$$\mathbf{P}_d = \sum c_i w_i (l_i + \Delta\mathbf{L}_{\text{dtd}} + \Delta\mathbf{L}_{i,\text{wid}}) \quad (2)$$

The total power random variable is the sum of the random variables (1) and (2).

B. Measures of statistical leakage power

In this paper we will cover the statistical measures in Table II. The measures are given in terms of the "adjusted gate widths", z_i:

$$z_i = w_i e^{\alpha l_i^2+\beta l_i} e^{-\gamma v_{t_i}}$$

which incorporate the effect of l and v_t into an equivalent gate width. In the table, S denotes a covariance matrix, where the i, j^{th} entry of S contains the covariance of gates i and j, for $z_i = z_j = 1$.

Figure 1 plots the sensitivities of measures 1 to 3 for the different gates of the 45nm Nangate Open Cell Library v1.2 [3]. Note that the 3σ-quantile is used when there is only inter-die variation. When intra-die variation is also present, the mean+3σ measure is used instead.

The power measures above have useful mathematical properties. Measures 1 to 3 above are linear in z_i, and are thus concave *and* convex in z_i. Also, *all* the measures above are convex in $\vec{z}$.

C. Why do we expect the optimizations to be similar?

The statistical power and deterministic power are not similar. For example, the statistical leakage power can be larger than the deterministic leakage by 10% to 100%. It is natural to expect that the influence of these measures will also be different, and that optimizing statistical power will yield different results compared to deterministic power.

In optimization however, it is not the magnitude of the power, but the *relative* magnitude – if the statistical power is a scaled version of the deterministic power, then the optimums will be the same. To see this mathematically, we examine the optimality condition for an optimum $x^\star$: $x^\star$ is optimal if for any *feasible* $x^\star + \Delta x$,

$$f(x^\star) \leq f(x^\star + \Delta x).$$

However, this condition will also hold for any *positive* scaling of $f(x)$.

Although the values of the statistical power and the deterministic powers may be quite different, the trends are similar – the mean power is larger for all of the gates, as is the quantile power, etc. For example consider the top plot in figure 1, which shows the power vs. size sensitivities for the combinatorial cells in the Nangate Library [3]. The different sensitivities all follow the same trend. This suggests that the optimizations will be similar as well.

To see why this happens for statistical power, we take as an example the 3σ-quantile power expression:

$$p_{3\sigma}(\vec{z}_i) = \sum e^{3\eta_i \Delta\mathbf{L}_{\text{dtd}}} k_i z_i.$$

If the η_i are all equal, then the effect of variation will be seen equally for each gate and the objective will be a scaled version of the deterministic power:

$$\begin{aligned} p_{3\sigma}(\vec{z}_i) &= e^{3\eta\Delta\mathbf{L}_{\text{dtd}}} \sum k_i z_i \\ &= e^{3\eta\Delta\mathbf{L}_{\text{dtd}}} p_d(\vec{z}_i). \end{aligned}$$

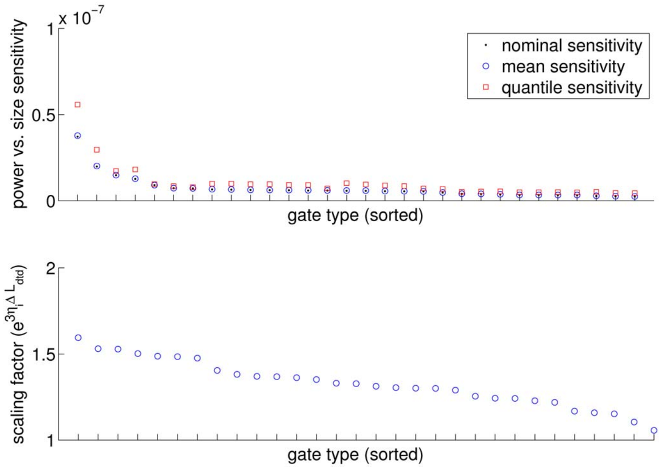

Fig. 1. The top plot shows the power vs. size sensitivities of the combinatorial logic cells in Nangate. The general pattern of each of the sensitivities vary similarly, suggesting that the optimizations will be similar as well. The bottom plot shows the modeled scaling factors $e^{3\eta_i \Delta \mathbf{L}_{\mathrm{dtd}}}$ for the combinatorial logic cells in the Nangate cell library. Although there is a significant difference in scaling values, the trend is not large enough to make a significant difference on the optima.

The actual case is a middle ground – the values of $e^{3\eta_i \Delta \mathbf{L}_{\mathrm{dtd}}}$ are different for different gates (see the bottom plot in Fig 1). However, in section III we will see that the effects are not large enough to make a large difference in the optimized powers.

III. Suboptimality bounds

The central question in this paper is whether the deterministic power solution $x_d^\star$ is a good approximation for the statistical power optimum $x_s^\star$. This generally requires information about the space of timing-feasible solutions ($\mathcal{T}$), which is difficult to describe, and is highly problem dependent, making it hard to give an exact answer. However, there is a way to solve a simpler problem with very little assumption on the structure of $\mathcal{T}$.

In this section we consider the following question: suppose we approximate the solution to the statistical power optimization problem:

$$\begin{array}{lll} \text{minimize} & p_s(\vec{w}) & \text{(statistical power)} \\ \text{subject to} & \vec{w} \in \mathcal{T} & \text{(timing constraint} \\ & & \ \text{\& discreteness)} \\ & \vec{w} \in \mathcal{B} & \text{(size bounds on } \vec{w}\text{)} \end{array} \qquad \text{(S)}$$

using the deterministic power optimum, $\vec{w}_d^\star$, which is the solution to the problem:

$$\begin{array}{lll} \text{minimize} & p_d(\vec{w}) & \text{(deterministic power)} \\ \text{subject to} & \vec{w} \in \mathcal{T} & \text{(timing constraint} \\ & & \ \text{\& discreteness)} \\ & \vec{w} \in \mathcal{B} & \text{(size bounds on } \vec{w}\text{).} \end{array} \qquad \text{(D)}$$

How good of an approximation will this be, and what is a bound for the suboptimality of this solution?

In the following, we will describe a method for creating suboptimality bounds. First a simple set $\mathcal{T}'$ is constructed that contains $\mathcal{T}$. Optimizing the statistical power over this simpler set will return a lower bound on the statistical power optimum. This lower bound is then compared with the statistical power of the approximate solution, $p_s(\vec{w}_d^\star)$, to bound the accuracy of the approximation. This is described in detail below.

A. Relaxed constraints, enclosing sets and lower bounds

The difficult part of w, l and v_t optimization is the timing constraint and the discreteness constraint; the statistical power measures in Table II are easy to minimize. Thus, to find a quick lower bound, we must first relax the timing constraint with a looser constraint. In other words, we would like to *relax* the constraints, by enclosing the timing feasibility and discreteness condition in a simple, convex set.

Relaxing the constraints of a problem turns the resulting solution into a lower bound for the true solution. For example, consider the sets $\mathcal{T}_0 \subseteq \mathcal{T}_1 \subseteq \ldots \subseteq \mathcal{T}_k$ and the sequence of problems:

$$\begin{array}{lll} (\mathrm{P}_i) & \text{minimize} & p_s(\vec{w}) \\ & \text{subject to} & \vec{w} \in \mathcal{T}_i \\ & & \vec{w} \in \mathcal{B} \subseteq \mathbb{R}^n. \end{array}$$

If the optimal solution of problem (P_i) is $\vec{w}_i^\star$, then we have the property that:

$$p_s(\vec{w}_0^\star) \geq p_s(\vec{w}_1^\star) \geq \ldots \geq p_s(\vec{w}_k^\star).$$

In other words the optimal value for the relaxed problem is a lower bound for the original problem.

The intuition for this is the fact that the constraints in the relaxed problem enclose the constraints on the original problem. Thus, the optimal solution in the original problem is also *feasible* for the relaxed problem. In the process of solving the relaxed problem, the solver is free to choose a better point in the larger space, making the resulting optimum a *lower bound* for the original problem.

B. Linear functions, optimum solutions and enclosing sets

For certain classes of functions, it is easy to find a simple set that encloses the optimum. The following analysis will derive a set using the properties of linear functions, but the results also hold for more general functions.[1]

The key to finding an enclosing set for the constraints is to start with an optimal solution and leverage the fact that any other feasible point cannot be better. For example, if $\vec{w}_d^\star$ is optimal for problem (D), then

$$\forall \vec{w} \in (\mathcal{T} \cap \mathcal{B}): \; p_d(\vec{w}_d^\star) \leq p_d(\vec{w}). \qquad (3)$$

For linear functions, the inequality on the right side can be rewritten in a simple form. This is because any linear

[1] Specifically, equation (4) holds whenever $p_d(\vec{w})$ satisfies:

$$\{\vec{w} \mid p_d(\vec{w}^\star) \leq p_d(\vec{w})\} \subseteq \{\vec{w} \mid 0 \leq \nabla p_d(\vec{w}^\star)^T(\vec{w} - \vec{w}^\star)\}$$

This includes concave functions and functions that are convex and non-decreasing along its gradients.

function $f(x)$ can be expressed in the form $f(x) = f(x_0) + s^T(x - x_0)$ where $s = \nabla f(x)$:[2]

$$
\begin{aligned}
f(x^\star) &\le f(x^\star) + s^T(x - x^\star) \\
0 &\le s^T(x - x^\star)
\end{aligned}
$$

Applying this to (3) with $s = \nabla p_d(\vec{w}_d^\star)$ shows that

$$(\mathcal{T} \cap \mathcal{B}) \subseteq \mathcal{T}' = \{\vec{w} \mid 0 \le s^T(\vec{w} - \vec{w}_d^\star)\}. \tag{4}$$

This gives the relaxed problem:

$$
\begin{array}{ll}
\text{minimize} & p_d(\vec{w}) \\
\text{subject to} & \vec{w} \in \mathcal{T}' \\
& \vec{w} \in \mathcal{B} \subseteq \mathbb{R}^n
\end{array} \tag{5}
$$

C. Creating lower bounds for related problems

The above analysis seems a little circular – the optimum is required to create a lower bound for the optimum. However, the utility emerges when we use the same enclosing sets to find lower bounds for related problems.

Suppose the solution for problem (D) is known, and we would now like to find the a lower bound for problem (S), which has a different objective function, but identical constraints. This can be done by leveraging the solution $\vec{w}_d^\star$ for problem (D) to compute a simple, enclosing set for the constraints, as in the subsection above. The relaxed problem is then solved:

$$
\begin{array}{ll}
\text{minimize} & p_s(\vec{w}) \\
\text{subject to} & \vec{w} \in \mathcal{T}' \\
& \vec{w} \in \mathcal{B} \subseteq \mathbb{R}^n
\end{array} \tag{6}
$$

and the solution $\vec{w}'$ can be used as a lower bound on the true optimum $w_s^\star$ ($p_s(\vec{w}') \le p_s(\vec{w}_s^\star)$).

Using this lower bound, we can now find a bound for how well the deterministic solution approximates the solution for the statistical problem. The suboptimality gap between this approximation $\vec{w}_d^\star$, and the true optimum, $\vec{w}_s^\star$ is bounded by:

$$\delta_{\text{so}} = 100 \cdot \frac{p_s(\vec{w}_d^\star) - p_s(\vec{w}')}{p_s(\vec{w}_d^\star)}$$

Smaller values indicate that $\vec{w}_d^\star$ is a good approximate solution for $\vec{w}_s^\star$, and larger values indicate that it is a bad approximation. For example, if

$$\delta_{\text{so}} \le 5\%$$

then $\vec{w}_d^\star$ is a 5% approximate solution for $\vec{w}_s^\star$. In other words, using $\vec{w}_d^\star$ in place of the real optimum $\vec{w}_s^\star$, costs at most 5% (it is *suboptimal* by at most 5%).

This process also works with optimization over w, l_i and v_t. This is exactly similar to the above example, with w replaced by z, the adjusted gate widths.[3]

A surprising fact is that no properties of $p_s(\vec{w})$ are assumed. It may be non-convex and non-linear. The only assumption is that the problem (6) is solvable.

TABLE III
Suboptimality (δ_{so}) for Leakage Optimization

	p_{m}					$p_{\text{m}3\sigma}$				
	1	2	3	4	avg	1	2	3	4	avg
c432	.19%	.29%	.38%	.39%	.31%	2.6%	5.3%	7.7%	8.5%	6%
c499	.18%	.19%	.08%	.06%	.13%	2.1%	1.7%	.98%	1.2%	1.5%
c880	.2%	.23%	.24%	.19%	.22%	1.7%	2.5%	3%	2.6%	2.5%
c1355	.18%	.2%	.17%	.1%	.16%	2.1%	1.7%	1.2%	1.2%	1.5%
c1908	.23%	.24%	.23%	.24%	.23%	2%	3.9%	4.1%	4.3%	3.6%
c2670	.22%	.16%	.15%	.16%	.17%	2.8%	2.1%	2%	2%	2.2%
c3540	.2%	.23%	.23%	.23%	.22%	1.2%	1.7%	2.6%	2.6%	2%
c5315	.17%	.17%	.16%	.15%	.16%	2.7%	2.7%	2.5%	2.5%	2.6%
c6288	.2%	.19%	.23%	.22%	.21%	2%	1.5%	1.8%	1.1%	1.6%
c7552	.19%	.17%	.15%	.15%	.17%	2.2%	1.2%	1.7%	1.1%	1.5%
alu	.21%	.18%	.19%	.17%	.19%	1.9%	2.7%	2.5%	1.2%	2.1%

TABLE IV
Summary of Correction Errors

Leakage Power						
	p_{m}			$p_{\text{m}3\sigma}$		
variables	min	max	avg.	min	max	avg.
w	.004%	.03%	.01%	.07%	.65%	.19%
w, v_t	.009%	.08%	.02%	.15%	1.8%	.46%
w, v_t, l	.016%	.14%	.04%	.36%	3.5%	.86%
Total Power ($k_{\text{switch}} = .001$)						
w, v_t, l	.005%	.10%	.024%	.077%	3.3%	.98%

D. Experiment

The ISCAS '85 benchmarks and a 128-bit Arithmetic Logic Unit, ALU [1], were synthesized using the Encounter RTL compiler with the Nangate Open Cell Library v1.2, which uses the 45nm technology node. The library was fitted using the models in Section II, and the powers were averaged over the different input combinations. These designs were synthesized to four different speeds – the maximum speed, the minimum speed, and two speeds in-between. The fastest speed is labeled (1), and the slower synthesized speeds have higher cardinality (e.g., the slowest speed is (4)). The design variables in this example are the gate widths. The suboptimality for each circuit (δ_{so}) is computed and the results are presented in Table III.

The computed suboptimalities are small. The worst

[2] Note that s is constant over x

[3] This gives the problem:

$$
\begin{array}{ll}
\text{minimize} & p_s(\vec{z}) \\
\text{subject to} & 0 \le \nabla p_d(\vec{z}_d^\star)^T(\vec{z} - \vec{z}_d^\star) \\
& z_{min} \le \vec{z} \le z_{max}
\end{array} \tag{7}
$$

Here, $z_{\min}$ and $z_{\max}$ are the minimum and maximum values of $\vec{z}$. For example, for the i^{th} entry, they are:

$z_{i,\min} = w_{i,\min} e^{\alpha l_{i,\max}^2 + \beta l_{i,\max}} e^{-\gamma v_{t_i,\max}}$ and

$z_{i,\max} = w_{i,\max} e^{\alpha l_{i,\min}^2 + \beta l_{i,\min}} e^{-\gamma v_{t_i,\min}}$.

Interestingly, the actual values of $\vec{l}$, $\vec{w}$ and v_t do not play a direct role in the optimization above. They affect the optimization by determining a range for the values of z_i. In fact, the corresponding values of w_i, l_i and v_t may not be unique; it is only important that there is a least one combination of w_i, l_i and v_t that satisfies $z_i = w_i e^{\alpha l_i^2 + \beta l_i} e^{-\gamma v_{t_i}}$.

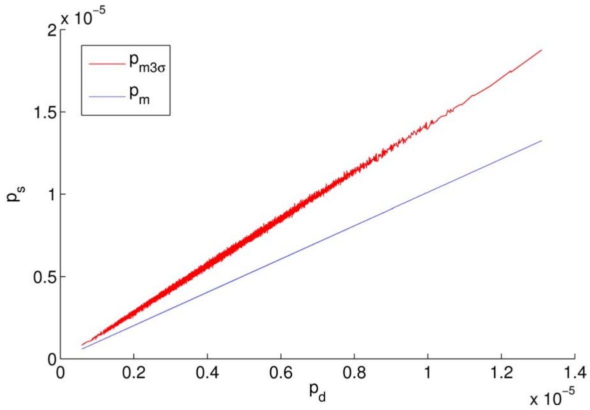

Fig. 2. Rankings of 10,000 randomly generated widths, lengths and v_t for the c432 circuit. The x-axis is the deterministic power of a point, and the y-axis is the statistical power. The power measures align well, indicating that smaller deterministic powers will correspond to small statistical powers.

case is the $p_{\mathrm{m}3\sigma}$ power measure and the circuit c432 where the the suboptimality is nearly 9%. In the majority of the other cases, however, the suboptimality is small, $< 5\%$. For the $p_{\mathrm{m}3\sigma}$ measure the average δ_{so} is 2.4%, and for the p_m measure the average δ_{so} is 0.2%. This suggests that there is little improvement to be gained by running the statistical optimization, as the improvement is on the same magnitude as modeling error.

The suboptimalities of the total power random variable were also computed in a similar way for switching probability .001. The resulting suboptimalities were *all* very small ($< 1\%$), and much smaller than in Table III. This is because the variations in dynamic power are much smaller than the variations in leakage power, and it reduces the difference between the statistical and deterministic powers.

Although examples with v_t assignment and l sizing were not included in this section, they will be covered in the following section.

IV. Solution rankings

The weakest part of the above analysis is the need for an optimum solution to compute the suboptimality. However, with discrete sizes (e.g., $\vec{w} \in \{1, 2, 4, 8\}^n$ and $\vec{l} \in \{1, 2, 3\}^n$), the problem is NP-complete [14] and the solution is not likely to be optimal. In this case, we want to compare the rankings of the deterministic power and the statistical power solutions, and show that good deterministic solutions will be good statistical solutions.

As an experiment, the deterministic and statistical powers were computed for 10,000 randomly generated solutions widths, lengths and v_ts, for each design in the ISCAS '85 benchmark suite. The powers are computed using a lookup table generated from a Monte-Carlo simulation of the Nangate Open Cell Library, using models from the BPTM 45 [2]. Three different v_t values were used (low, high and normal), and four different gate lengths were used ($\{+0\text{nm}, +1\text{nm}, +2\text{nm}, +3\text{nm}\}$). The variations are assumed to be $\sigma_{\mathbf{L}_{\mathrm{dtd}}} = 1\text{nm}$ and $\sigma_{\mathbf{L}_{i,\mathrm{wid}}} = 0.5\text{nm}$.

In Figure 2, the deterministic vs. the statistical power is plotted for the c432 circuit (the plots for the other circuits are very similar). Visually, the relation is linear with a little error. In effect, this indicates that solutions that have good deterministic powers will have good statistical powers. In other words, a solution in the top 1% of all deterministic solutions is very likely to be in the top 1% of the statistical solutions.

To quantify the errors, the linear correlation between the two measures is be removed, and the remaining errors are normalized to a percentage of the statistical power. Intuitively, these rank errors approximate the minimum amount of correction that is required to align the deterministic power rankings and the statistical power rankings, and to make the i^{th} best deterministic solution the i^{th} best statistical solution. The average value tells us what the expected correction is, and can be interpreted as the expected suboptimality.

The minimum, maximum and average of these errors is shown in Table IV. The table reflects the data for the entire ISCAS '85 benchmark suite, and for different combinations of design variables (width only, width and v_t, etc.). The magnitude of the errors is very low ($< 1\%$ for $p_{m3\sigma}$ and $< .2\%$ for p_m), and indicates that on average, only a little correction is needed to align the results of the two different power optimizations.

V. Statistical vs. deterministic optimization

To support the results above, we optimized the circuits that were synthesized to speed 3 with mean, nominal and inter-die quantile $p_{3\sigma}$ objectives using a commercial gate-width sizing engine. Surprisingly, the optimizations yield near-identical results. Both the largest difference, and the average difference are very small (.03% and .003%, respectively) and in most of the circuits, the optimized powers are identical. This indicates that the worst-case bounds may be very loose.

We also ran the circuits through a continuous width sizing program. This used a linear delay model that is fitted from the Nangate Open Cell Library. Slew effects were ignored, and the sizing was run over three different delay values for each of the different synthesized speeds of the ISCAS '85 benchmarks. The differences here are also very small – the maximum is .07%, and the average is 0.06%. This is a further indication that the optimizations are similar, and that there is little benefit that is gained by optimizing the statistical power.

VI. Summary

In this paper we compared deterministic solutions of sizing problems with statistical solutions of sizing problems. The interesting conclusion is that the worst-case bounds on the suboptimality gap are small (usually $< 5\%$). The rankings of the solutions coincide very well, indicating that good deterministic solutions will be equally good statistical solutions. Another implication of high rank correlation is that any sizing engine is very likely to yield the same or similar solution whether it is driven by statistical or deterministic objectives. Given the results and the arguments presented in the paper, we believe that there is little value in statistical modeling of power for optimization purposes.

Our ongoing work is to extend similar arguments to statistical delay objective functions (especially the ones which are nicely expressible, such as total delay), and account for variations in v_t. Furthermore, we are investigating methods of quantifying the impact of relaxing statistical constraints to deterministic ones in optimization problems.

VII. Acknowledgments

We are very thankful to Dr. Saumil Shah for his insights on statistical power optimization. We are also grateful for support from SRC contract 2006-TJ-1460 and NSF award 811832. This research was also partially supported by the UC Discovery Grant ele07-10283 under the IMPACT program.

References

[1] Available from http://www.opencores.org.

[2] Available from http://www.eas.asu.edu/∼ptm.

[3] Nangate Open Cell Library v1.2. Available from http://www.si2.org/openeda.si2.org/projects/nangatelib.

[4] Agarwal, A., Blaauw, D., and Zolotov, V. Statistical timing analysis for intra-die process variations with spatial correlations. In *Proc. Int. Conf. Computer-Aided Design* (2003), p. 900.

[5] Bhardwaj, S., and Vrudhula, S. B. K. Leakage minimization of nano-scale circuits in the presence of systematic and random variations. In *Proc. Design Automation Conference* (2005), pp. 541–546.

[6] Burns, S. M., Ketkar, M., Menezes, N., Bowman, K. A., Tschanz, J. W., and De, V. Comparative analysis of conventional and statistical design techniques. In *Proc. Design Automation Conference* (2007), pp. 238–243.

[7] Chang, H., and Sapatnekar, S. S. Statistical timing analysis considering spatial correlations using a single pert-like traversal. In *Proc. Int. Conf. Computer-Aided Design* (2003), p. 621.

[8] Chang, H., Zolotov, V., Narayan, S., and Visweswariah, C. Parameterized block-based statistical timing analysis with non-gaussian parameters, nonlinear delay functions. In *Proc. Design Automation Conference* (2005), pp. 71–76.

[9] Chopra, K., Shah, S., Srivastava, A., Blaauw, D., and Sylvester, D. Parametric yield maximization using gate sizing based on efficient statistical power and delay gradient computation. In *Proc. Int. Conf. Computer-Aided Design* (2005), pp. 1023–1028.

[10] Davoodi, A., and Srivastava, A. Variability driven gate sizing for binning yield optimization. In *Proc. Design Automation Conference* (2006), pp. 959–964.

[11] Guthaus, M., Venkateswaran, N., Visweswariah, C., and Zolotov, V. Gate sizing using incremental parameterized statistical timing analysis. *Proc. Int. Conf. Computer-Aided Design* (2005), 1029–1036.

[12] Guthaus, M., Venkateswaran, N., Zolotov, V., Sylvester, D., and Brown, R. Optimization objectives and models of variation for statistical gate sizing. In *Proc. Great Lakes symposium on VLSI* (2005), pp. 313–316.

[13] Jess, J. A. G., Kalafala, K., Naidu, S. R., Otten, R. H. J. M., and Visweswariah, C. Statistical timing for parametric yield prediction of digital integrated circuits. In *Proc. Design Automation Conference* (2003), pp. 932–937.

[14] Li, W. Strongly np-hard discrete gate sizing problems. *IEEE International Conference on Computer Design: VLSI in Computers and Processors* (Oct 1993), 468–471.

[15] Mani, M., Devgan, A., and Orshansky, M. An efficient algorithm for statistical minimization of total power under timing yield constraints. In *Proc. Design Automation Conference* (2005), pp. 309–314.

[16] Mani, M., and Orshansky, M. A new statistical optimization algorithm for gate sizing. In *Proc. IEEE Int. Conf. on Computer Design* (2004), pp. 272–277.

[17] Najm, F. N. On the need for statistical timing analysis. In *Proc. Design Automation Conference* (2005), pp. 764–765.

[18] Nassif, S., Strojwas, A., and Director, S. A methodology for worst-case analysis of integrated circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD) 5*, 1 (January 1986), 104–113.

[19] Patil, D., Yun, S., Kim, S., Cheung, A., Horowitz, M., and Boyd, S. A new method for design of robust digital circuits. In *Proc. of ISQED* (2005), pp. 676–681.

[20] Rao, R., Srivastava, A., Blaauw, D., and Sylvester, D. Statistical analysis of subthreshold leakage current for VLSI circuits. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on 12*, 2 (2004), 131–139.

[21] Singh, J., Nookala, V., Luo, Z., and Sapatnekar, S. Robust gate sizing by geometric programming. In *Proc. Design Automation Conference* (2005), pp. 315–320.

[22] Sinha, D., Shenoy, N., and Zhou, H. Statistical gate sizing for timing yield optimization. In *Proc. Int. Conf. Computer-Aided Design* (2005), pp. 1037–1041.

[23] Srivastava, A., Sylvester, D., and Blaauw, D. Statistical optimization of leakage power considering process variations using dual-Vth and sizing. In *Proc. Design Automation Conference* (2004), pp. 773–778.

[24] Visweswariah, C. Death, taxes and failing chips. In *Proc. Design Automation Conference* (2003), pp. 343–347.

[25] Visweswariah, C., Ravindran, K., Kalafala, K., Walker, S., and Narayan, S. First-order incremental block-based statistical timing analysis. In *Proc. Design Automation Conference* (2004), pp. 331–336.

[26] Ye, X., Zhan, Y., and Li, P. Statistical leakage power minimization using fast equi-slack shell based optimization. In *Proc. Design Automation Conference* (2007), pp. 853–858.