Robust Gate Sizing by Geometric Programming*

Jaskirat Singh Vidyasagar Nookala Zhi-Quan Luo Sachin Sapatnekar
Department of Electrical and Computer Engineering
University of Minnesota
Minneapolis, MN 55455
{jsingh,vidya,luozq,sachin}@ece.umn.edu

ABSTRACT

We present an efficient optimization scheme for gate sizing in the presence of process variations. Using a posynomial delay model, the delay constraints are modified to incorporate uncertainty in the transistor widths and effective channel lengths due to the process variations. An *uncertainty ellipsoid* method is used to model the random parameter variations. Spatial correlations of intra-die width and channel length variations are incorporated in the optimization procedure. The resulting optimization problem is relaxed to be a Geometric Program and is efficiently solved using convex optimization tools. The effectiveness of our robust gate sizing scheme is demonstrated by applying the optimization on the ISCAS '85 benchmark circuits and testing the optimized circuits by performing Monte Carlo simulations to model the process variations. By varying the size of the uncertainty ellipsoids, a trade-off between area and robustness is explored. Experimental results show that the timing yield of the robustly optimized circuits improves manifold over the traditional deterministically sized circuits. As compared to the worst-case design, the robust gate sizing solution having the same area, has fewer timing violations.

Categories and Subject Descriptors

B.7.2 [**Hardware**]: Integrated Circuits—*Design Aids*

General Terms

Optimization, gate sizing

Keywords

Geometric Program, posynomial, uncertainty ellipsoid

1. INTRODUCTION

Due to shrinking of process geometries, it is becoming increasingly difficult to control the fabrication of critical device parameters. The limitations of the manufacturing process in the current technologies leads to random variations in various circuit parameters. These random perturbations from the nominal values of transistor width, channel length, oxide thickness, etc., may cause a large spread in the circuit performance measures such as the delay, power, etc. Since it is impossible to control the process-driven variations, it is essential for the design tools to account for these uncertainties and design robust circuits that are insensitive to the device parameter variations as much as possible.

The traditional gate sizing methodologies [1], [2] use Elmore delay based posynomial delay constraints to formulate the problem as a Geometric Program (GP). The use of posynomial delay models for the gate sizing problem enables the use of efficient convex optimization tools to solve the problem [3]. These conventional gate sizing tools employ a static timing analysis (STA) routine to generate the delay constraints,and then solve the GP optimization problem to determine the widths of the devices in the circuit. The minimum length is chosen for all the devices. However, due to the fact that the nominal designs are perturbed by the random process variations, a large number of chips may fail to meet the original delay specifications. This leads to a reduction in the *timing yield* of the circuit defined as the fraction of total chips whose delay does not exceed the original specified value. An obvious way to increase the timing yield of the circuit is to design for the worst-case scenario, e.g., choose a delay specification of the circuit much tighter than the required delay. This could lead to a large overhead in terms of the circuit area and the power as the optimizer may have to aggressively size the critical as well as the non-critical paths. Hence, it is necessary to develop smart worst-casing methodologies, in the presence of process uncertainties, that keep the area and the power budgets within reasonable bounds.

There have been several recent attempts to perform uncertainty-aware gate sizing to reduce the timing violations or increase the timing yield. In [4], the gate sizing problem is formulated as a non-linear optimization problem with a penalty function added to improve the distribution of timing slacks. In other works on robust gate sizing [5, 6, 7], the central idea is to capture the delay distributions by performing a statistical static timing analysis (SSTA), as opposed to the traditional STA, and then use a general non-linear programming technique to size the gates. To simplify the SSTA procedure, it is required to make assumptions such as the signal arrival time and the slope have normal distributions, and approximations such as the maximum of two or more normal distributions is also a normal distribution, which may be inaccurate. Some of these works [6],[7] ignore the significant spatial correlation component of the intra-chip parameter variations.

In this paper, we propose a novel gate sizing technique based on robust optimization theory [8]. For simplicity, we use the Elmore delay based model, but our approach is applicable to any posynomial delay model, such as the rich class of generalized posynomial delay models proposed in [3]. In our method, we first generate posynomial constraints by performing a STA. We then add *robust constraints* to the original constraints set by modeling the intra-chip random process variations in the transistor widths (W) and effec-

*This work was supported in part by the SRC under contract 2003-TJ-1092 and the NSF under award CCR-0205227.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
DAC 2005, June 13–17, 2005, Anaheim, California, USA.
Copyright 2005 ACM 1-59593-058-2/05/0006 ...$5.00.

tive channel lengths (L) as an *uncertainty ellipsoid* [9] centered at the nominal values. Under the ellipsoid uncertainty model, the resulting optimization formulation is relaxed to be a GP and is efficiently solved using the convex optimization tools.Furthermore, using a GP to perform robust gate sizing ensures that the optimizer finds a global minimum which is not guaranteed in the case of a general non-linear program. The relaxation of the robust counterpart of the conventional GP sizing solution as another GP is a major contribution of this work because, in general, it is not true that the robust versions of convex programs are also convex programs [10]. Our work is related to the conventional design centering approaches of [11] and [12]. However, unlike these methods which assume the design parameters to be normally distributed, the proposed gate sizing scheme does not require any assumptions to be made about the distributions of the parameter variations. Only the covariance matrix of the random perturbation vector is required as an input to the optimization problem. Our robust gate sizing scheme is a type of worst-case design method, but by incorporating spatial correlations in the design procedure, we reduce some pessimism in the design. Spatial intra-die correlations between the parameter variations are incorporated in the optimization scheme by using a grid-based spatial correlation model used in [13] and [14]. We focus on the intra-die variations in L and W parameters; however, the method can be easily modified to include inter-die variations. Process-driven variations in the interconnect widths and thickness can also be included in our method.

2. PRELIMINARIES

2.1 Conventional Gate Sizing as a GP

The conventional gate sizing problem is formulated as:

$$\begin{aligned} &\text{Minimize} && Area = \sum_{i=1}^{n} a_i W_i L_i \\ &\text{Subject to} && Delay \leq T_{spec} \\ &\text{and} && W_{min} \leq W_i, L_{min} \leq L_i \;\; \forall i = 1..n \end{aligned} \quad (1)$$

where, W_i and L_i are respectively, the width and the effective channel length of gate i, and a_i is some weight factor.Using the Elmore delay model[1], each gate i in the circuit can be replaced by an equivalent $R_{on_i}C_i$ element, where R_{on_i} represents the effective on resistance of the pull-up or the pull-down network, and the term C_i subsumes the source, drain and gate capacitances of the transistors in the gate. The expressions for R_{on_i} and C_i for a gate i are given by:

$$R_{on_i} = \frac{\alpha L_i}{W_i} \qquad C_i = \beta L_i W_i + \gamma \quad (2)$$

where, the constants α, β and γ can be derived from [2]. Both the capacitances and the on resistance of the transistors in a gate are posynomial functions of the vectors W and L. Consequently, the term $R_{on_i}C_i$ which is the equivalent delay contribution of gate i in the circuit is also a posynomial function of W and L. By breaking the circuit into a series of RC trees, and applying the Elmore delay computations at each node of the the circuit graph, the delay constraint of (1) at the primary outputs of the circuit, can be replaced by m posynomial delay constraints of the form:

$$\sum_l K_l \prod_j W_j^{a_j} L_j^{b_j} \quad \leq \quad t_i \quad (3)$$

where, m is the number of nodes in the circuit graph, K_l is a constant coefficient of the lth monomial term, which can be derived from (2), t_i is the arrival time at gate i, and a_j, b_j, the exponents of the jth components of the W and L vectors, $\in \{-1, 0, 1\}$. By substituting (3) in (1), for all gates in the circuit, the conventional transistor sizing is formulated as a GP optimization problem, having a posynomial objective function and posynomial constraints, which can be solved using the GP techniques. In Section 3.2, we show how the robust version of the standard GP formulation can be converted to another GP.

[1]The Elmore delay model is used for simplicity. Alternatively, generalized posynomial delay models [3], which have a higher accuracy, can be used for the GP formulation.

2.2 The ellipsoid set

For any vectors $\mathbf{X}$ and $X_0 \in R^n$, and a non-singular matrix $P \in R^{n \times n}$, an ellipsoid is defined as a set [9]:

$$\{\mathbf{X} : (\mathbf{X} - \mathbf{X_0})^T P^{-1} (\mathbf{X} - \mathbf{X_0}) \leq 1\} \quad (4)$$

If P is a symmetric and positive semidefinite matrix (PSD), an alternative representation of (4) is realized by making the substitution, $P^{-1/2}(\mathbf{X} - \mathbf{X_0}) = \mathbf{u}$ as:

$$\{\mathbf{X} = \mathbf{X_0} + P^{1/2}\mathbf{u} | \;\; \|\mathbf{u}\| \leq 1\} \quad (5)$$

where $\|\mathbf{u}\| = \mathbf{u}^T\mathbf{u}$ is the 2-norm of vector $\mathbf{u}$. For a symmetric and PSD matrix P, the matrix $P^{1/2}$ can be computed by the Cholesky factors of P. The ellipsoid represents a n-dimensional region, where the vector $\mathbf{X}$ varies around the center point X_0. The vector $\mathbf{u}$ characterizes the movement of $\mathbf{X}$ around X_0.

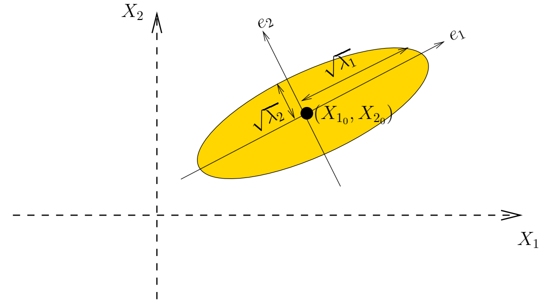

Figure 1: An uncertainty ellipsoid in 2 dimensions

Figure 1 illustrates the ellipsoid in R^2. The half-lengths of the axis of the ellipsoid are the square roots of the eigenvalues, λ_1 and λ_2, of the matrix P, and the direction of the axis is given by the eigenvectors of P, e_1 and e_2. In Section 3.1, we will introduce the concept of an uncertainty ellipsoid, based on the ellipsoid set of (5), to model the process-driven variability and formulate our robust optimization problem.

3. GATE SIZING UNDER PROCESS VARIATIONS

The posynomial constraints of (3) can be represented as:

$$f_i(\mathbf{X_0}) \quad \leq \quad t_i \quad (6)$$

where $f_i(\mathbf{X_0}) = \sum_\mathbf{l} \mathbf{K_l} \prod_\mathbf{j} \mathbf{W_{j0}^{a_j} L_{j0}^{b_j}}$ represents the ith constraint function, $\mathbf{X_0}$ is the vector representing the nominal W and L. The conventional GP optimization assigns a set of optimal W_0 and L_0 to the vector $\mathbf{X_0}$ so that each delay constraint is satisfied, i.e., $f_i \leq t_i$ for all constraints i, and the area objective is minimized.

Due to the process variations, the components of vector $\mathbf{X}$, the transistor W and L are no longer deterministic quantities, but behave as random perturbations around their nominal values. As a result, the constraint function $f_i(\mathbf{X_0})$ changes to $f_i(\mathbf{X_0} + \delta\mathbf{X})$. For the cases when the new value of the constraint function $f_i > t_i$,

the effect of the random process variations leads to the original constraints being violated and timing failure for the circuit.

Assuming that the random parameter perturbations around the nominal values are small, the new value of the constraint function f_i can be approximated by a first order Taylor series expansion as:

$$\begin{aligned} f_i(\mathbf{X_0}+\delta\mathbf{X}) &= f_i(\mathbf{X_0})+\sum_{\mathbf{j}}\frac{\delta\mathbf{f_i}(\mathbf{X})}{\delta(\mathbf{X_j})_\mathbf{0}}(\mathbf{X_j}-\mathbf{X_{j_0}}) \\ &= f_i(\mathbf{X_0})+\nabla_{\mathbf{X_0}}\mathbf{f_i}(\mathbf{X_0})\delta\mathbf{X} \\ &= \sum_l K_l\prod_j W_{j0}^{a_j}L_{j0}^{b_j}+ \\ & \quad \nabla_{\mathbf{X_0}}(\sum_l K_l\prod_j W_j^{a_j}L_j^{b_j})\delta X \end{aligned} \tag{7}$$

where $\nabla_{\mathbf{X_0}}$ represents the gradient calculated at the nominal values of W and L, and δX represents the random variation in the widths and channel lengths, around the nominal values.

In (7) the term, $\nabla_{\mathbf{X_0}}(\sum_l K_l\prod_j W_j^{a_j}L_j^{b_j})\delta X$ is the variational term representing the effect of process variations added to the nominal term $\sum_l K_l\prod_j W_{j0}^{a_j}L_{j0}^{b_j}$. To safeguard against the uncertainty of process variations, it is necessary to meet the constraint, $f_i < t_i$, for the maximum value of the variational term. In other words:

$$\begin{aligned} &\sum_l K_l\prod_j W_{j0}^{a_j}L_{j0}^{b_j}+ \\ \max_{\forall\delta X}&(\nabla_{\mathbf{X_0}}(\sum_l K_l\prod_j W_j^{a_j}L_j^{b_j})\delta X) \le t_i \end{aligned} \tag{8}$$

In the following sections, we show that by employing the concept of an uncertainty ellipsoid, the constraint of (8) can be formulated as a posynomial constraint, so that the robust optimization formulation remains a GP, and can be efficiently solved. Our robust GP formulation is applicable for all cases where the original constraints are in the form of a generalized posynomial [3].

3.1 Uncertainty Ellipsoid

For a random vector $\mathbf{X}$ of size n, with a mean X_0 and the covariance matrix P, the uncertainty ellipsoid (also called the variance ellipsoid) is defined by the ellipsoid set of (5) [9]. The uncertainty ellipsoid represents region of variation of the random vector $\mathbf{X}$ around the mean vector X_0. The variational vector $\mathbf{u}$ characterizes the uncertain movement of $\mathbf{X}$ around X_0. As seen from Figure 1 and equation (5), the maximum variation of $\mathbf{X}$ around X_0 is bounded due to the fact that the vector $\mathbf{u}$ has a norm $\|\mathbf{u}\| \le 1$.

We use the uncertainty ellipsoid to model the process variations that randomly perturb the transistor W and L around the nominal values for which they were designed. As the random vector $\mathbf{X}$ of W and L varies around the nominally designed vector X_0, the variations are considered to be bounded within the ellipsoid regions defined by (5). In other words, the variation δX from X_0 is given by $\delta X = P^{1/2}\mathbf{u}$ with $\|\mathbf{u}\| \le 1$.

Alternatively, we could have chosen the variation δX in W and L to be bounded in a n-dimensional box given by $X_{min} \le \delta\mathbf{X} \le X_{max}$. However, using the box as model for bounded variation, ignores any correlation information between the random variables components of $\mathbf{X}$, as each component can move independently inside a box, assuming any values between the minimum and maximum range. Thus, optimizing for a maximum variation in such a box region would translates to an overly pessimistic design. The advantage of using the ellipsoid uncertainty model is that any correlations between the components of the random vector $\mathbf{X}$ are directly captured by appropriately populating the entries of the covariance matrix P. As will be explained in Section 4, we use this fact to incorporate the spatial correlations between the random parameter variations. To generate the uncertainty ellipsoid region, it is not required to make any assumptions about the distributions of the W and L. The only inputs needed to generate the P matrix are the standard deviations of the components of $\mathbf{X}$, which can be empirically calculated [15], and correlation factors between the components of $\mathbf{X}$, which can be derived from a spatial correlation model such as the ones used in [13] and [14].

In the next section, we show with the aid of a small example, the use of the ellipsoid uncertainty model in converting the constraint of (8) to a posynomial constraint and formulating the robust GP for gate sizing in the presence of process variations.

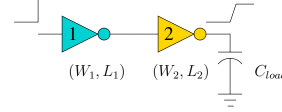

Figure 2: A simple example circuit

3.2 Robust GP formulation

We use a simple example to explain the procedure to incorporate the process variation effects in the delay constraints set. We use the toy circuit of Figure 2, comprising of just one driver gate and one load gate, for this illustration, but the idea can be generalized to arbitrarily large circuits.

Applying the Elmore delay model to the gates of circuit of Figure 2, and for simplicity neglecting the interconnect delay and the effect of drain and source capacitances of the driver gate, the delay constraint for the circuit can be written as:

$$\frac{K_1L_1L_2W_2}{W_1}+\frac{K_2L_2}{W_2} \le T_{spec} \tag{9}$$

where K_1 and K_2 are constants. As explained in Section 3, to ensure that the delay constraint of (9) is met under the effect of random process variations, the first order Taylor series expansion of the constraint function results in the following relation:

$$\begin{aligned} &\frac{K_1L_{1_0}L_{2_0}W_{2_0}}{W_{1_0}}+\frac{K_2L_{2_0}}{W_{2_0}}+ \\ \max_{\forall\delta W,\delta L}&\Bigg(\frac{K_1L_{1_0}L_{2_0}\delta W_2}{W_{1_0}}+\frac{K_1L_{2_0}W_{2_0}\delta L_1}{W_{1_0}}+ \\ &\frac{K_1L_{1_0}W_{2_0}\delta L_2}{W_{1_0}}+\frac{K_2\delta L_2}{W_{2_0}}- \\ &\frac{K_1L_{1_0}L_{2_0}W_{2_0}\delta W_1}{W_{1_0}^2}-\frac{K_2L_{2_0}\delta W_2}{W_{2_0}^2}\Bigg) \le T_{spec} \end{aligned} \tag{10}$$

where W_0 and L_0 represent, respectively, the nominal values of the transistor W and L, and δW and δL are, respectively, the random variations in W and L. Employing the ellipsoid uncertainty model of (5) for the random parameter variations, leads to:

$$\begin{bmatrix}\delta W_1\\ \delta W_2\\ \delta L_1\\ \delta L_2\end{bmatrix} = \begin{bmatrix}(P^{1/2}\mathbf{u})_1\\ (P^{1/2}\mathbf{u})_2\\ (P^{1/2}\mathbf{u})_3\\ (P^{1/2}\mathbf{u})_4\end{bmatrix} \tag{11}$$

where P is the covariance matrix of the random vector $\mathbf{X}$ comprising of the transistor W and L of the driver and the load gate of Figure 2, and $\mathbf{u}$ is the vector characterizing the variation within the 4-dimensional ellipsoid centered at the nominal values of W and L, with $\|\mathbf{u}\| \le 1$. We introduce two vectors ϕ_1 and ϕ_2 to collect the positive and negative coefficients, respectively, of the variational

parameters of (10) as:

$$\phi_1 = \begin{bmatrix} 0 \\ \frac{K_1 L_{1_0} L_{2_0}}{W_{1_0}} \\ \frac{K_1 L_{2_0} W_{2_0}}{W_{1_0}} \\ \frac{K_1 L_{1_0} W_{2_0}}{W_{1_0}} + \frac{K_2}{W_{2_0}} \end{bmatrix}, \phi_2 = \begin{bmatrix} \frac{-K_1 L_{1_0} L_{2_0} W_{2_0}}{W_{1_0}^2} \\ \frac{-K_2 L_{2_0}}{W_{2_0}^2} \\ 0 \\ 0 \end{bmatrix} \quad (12)$$

From the definitions in (11) and (12), (10) can be rewritten as:

$$\frac{K_1 L_{1_0} L_{2_0} W_{2_0}}{W_{1_0}} + \frac{K_2 L_{1_0}}{W_{2_0}} + \max_{\forall \mathbf{u}} \left(\langle P^{1/2}\phi_1, \mathbf{u} \rangle + \langle P^{1/2}\phi_2, \mathbf{u} \rangle \right) \leq T_{spec} \quad (13)$$

where $\langle a, b \rangle$ represents the inner product of vectors a and b. From the well known result of the Cauchy Schwartz inequality:[2]

$$< a, b > \quad \leq \quad \|a\| \cdot \|b\| \quad (14)$$

and the fact that in the ellipsoid uncertainty model, $\|\mathbf{u}\| \leq 1$, a sufficient condition[3] for (13) is:

$$\frac{K_1 L_{1_0} L_{2_0} W_{2_0}}{W_{1_0}} + \frac{K_2 L_{1_0}}{W_{2_0}} + \|P^{1/2}\phi_1\| + \|P^{1/2}\phi_2\| \leq T_{spec} \quad (15)$$

We then introduce two additional *robust variables* r_1 and r_2 as:

$$r_1 = \|P^{1/2}\phi_1\|, \quad \text{i.e.,} \quad r_1^2 = \phi_1^T P \phi_1$$
$$r_2 = \|P^{1/2}\phi_2\|, \quad \text{i.e.,} \quad r_2^2 = \phi_2^T P \phi_2 \quad (16)$$

The inequality of (15) is then replaced by the following constraints:

$$\frac{K_1 L_{1_0} L_{2_0} W_{2_0}}{W_{1_0}} + \frac{K_2 L_{1_0}}{W_{2_0}} + r_1 + r_2 \leq T_{spec} \quad (17)$$
$$\phi_1^T P \phi_1 r_1^{-2} \leq 1 \quad (18)$$
$$\phi_2^T P \phi_2 r_2^{-2} \leq 1 \quad (19)$$

The inequality of (17) is clearly a posynomial with the robust variables r_1 and r_2 added to the original variable list of the transistor W and L. By construction, all the elements of ϕ_1 are posynomials, and all the non-zero elements of ϕ_2 are negative of posynomials. The covariance matrix P has all non-negative elements, because a negative correlation between random variables representing the W and L variations would not have any physical meaning. Thus, the quadratic terms $\phi_1^T P \phi_1 = \sum_{i,j} P_{ij} \phi_{1_i} \phi_{1_j}$, and $\phi_2^T P \phi_2 = \sum_{i,j} P_{ij} \phi_{2_i} \phi_{2_j}$ are a summation of monomials with positive coefficients. Consequently, the constraints of (18) and (19) are also posynomials. Note that the inequality in (18) and (19) would be forced to equality of (16) by the optimizer, as the robust variables r_1 and r_2, which represent the maximum variation in the uncertainty ellipsoid, are to be minimized. Hence, by following the procedure described in the above equations, we convert the non-robust posynomial constraint of (9) to a set of robust posynomial constraints of (17-19) by introducing two additional variables.

For a general circuit, the procedure described for the example circuit of Figure 2 is repeated for each constraint. Thus, by addition of at most two additional variables for each constraint j, robustness against the process uncertainties is added to original constraint set, while still maintaining the desirable posynomial structure of the constraints. By this procedure, we convert the conventional GP formulation of the gate sizing problem to a robust gate sizing problem, which is also a GP and hence, can be efficiently solved using the convex optimization machinery.

[2] In our case, the equality in (14) also holds because there are some points in the ellipsoid set which have $\langle P^{1/2}\phi_1, \mathbf{u} \rangle = \|P^{1/2}\phi_1\| \cdot \|\mathbf{u}\|$.

[3] An equivalent condition for (13) is: $(\frac{K_1 L_{1_0} L_{2_0} W_{2_0}}{W_{1_0}} + \frac{K_2 L_{1_0}}{W_{2_0}} + \|P^{1/2}(\phi_1 + \phi_2)\|) \leq T_{spec}$. However, this does not lead to the formulation of posynomial constraints of (18) and (19).

4. INCORPORATING SPATIAL CORRELATIONS

We use the grid based spatial correlation model of [13] and [14] to incorporate the intra-die correlations between the device W and L variations.

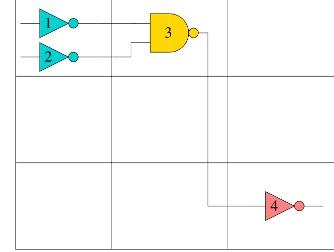

Figure 3: Grid based spatial correlation model

Figure 3 refers to such a model, where the layout area is partitioned into $m = 9$ grids. The widths (channel lengths) of the devices located in the same grid are assigned perfect correlations, device widths (channel lengths) in nearby grids are assigned high correlations and low or zero correlations in far away grids. As seen in Figure 3, gates {1,2} have perfect correlation between their widths (channel lengths), gates {1,3} and {2,3} have high correlations, where as gates {1,4} and {2,4} are uncorrelated.

For a random vector $\mathbf{X}$ representing the variations in W and L and its corresponding covariance matrix P, the entry $P_{ij} = \sigma_i \sigma_j \rho_{ij}$ denotes the covariance between components i and j of $\mathbf{X}$, where σ is the standard deviation of each random variable and ρ_{ij} is the correlation factor between the random variables i and j. By employing the spatial correlation model of Figure 3, the correlation factor between all elements of $\mathbf{X}$ is computed and stamped out in matrix P. The ellipsoid uncertainty model described in Section 3.1 then automatically incorporates the impact of correlations in the robust optimization formulation.

The following simple example explains how the correlations are captured by the uncertainty ellipsoid. Consider a simple constraint involving the transistor widths of two gates:

$$\frac{K_1 W_1}{W_2} \leq t_i \quad (20)$$

For simplicity, we assume that the channel length is not varying and is included in the constant K_1. Furthermore let's assume that the gates are placed in the same grid of the spatial correlation model, hence, the variations in the two gate widths are same, i.e., $\delta W_1 = \delta W_2$. Also if the nominal gate sizes are the identical, i.e., $W_{1_0} = W_{2_0}$, the effect of process variation cancels out in the numerator and denominator of (20) and no guard-banding is required. To verify that the ellipsoid uncertainty correctly incorporates this perfect correlation affect, we apply our robust optimization procedure to the constraint function of (20). A first order Taylor series expansion of the constraint around the nominal values (W_{1_0}, W_{2_0}) and applying the ellipsoid uncertainty yields:

$$\frac{K_1 W_{1_0}}{W_{2_0}} + \frac{K_1 (P^{1/2}u)_1}{W_{2_0}} - \frac{K_1 W_{1_0} (P^{1/2}u)_2}{W_{2_0}^2} \leq t_i \quad (21)$$

However, since we have perfect correlation between W_1 and W_2, the correlation factor, $\rho_{12} = \rho_{21} = 1$. Furthermore, since the variations in W_1 and W_2 and the mean values are same, we must have $\sigma_1 = \sigma_2$. It then follows that for all vectors $\mathbf{u} = [u_1, u_2]$, which characterize the uncertainty ellipsoid, we have $(P^{1/2}u)_1 = (P^{1/2}u)_2$ and the variational term in (20):

$$\frac{K_1(P^{1/2}u)_1}{W_{2_0}} - \frac{K_1 W_{1_0}(P^{1/2}u)_2}{W_{2_0}^2} = 0$$

Thus, the ellipsoid uncertainty model easily captures the effects of correlations between random variables and incorporates the same in the optimization procedure. Incorporating the correlations in gate sizing optimization procedure ensures that the circuit is not over-designed to achieve robustness against the process variations.

5. THE GATE SIZING PROCEDURE

Figure 4 summarizes the overall flow for the robust gate sizing procedure. We start by generating the non-robust delay constraints

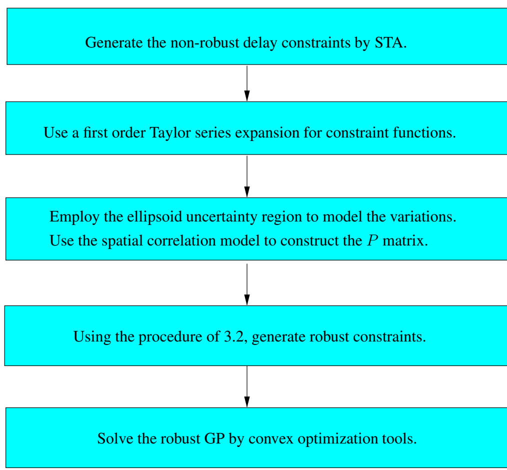

Figure 4: Overall flow of the robust gate sizing procedure.

by performing a STA. The effect of process variations on the constraints are considered by using a first order Taylor series of the posynomial constraints around their nominal values. We employ the uncertainty ellipsoid to model the random process variations. The P matrix characterizing the variance ellipsoid is constructed by using the spatial correlation model. Following the procedure described in Section 3.2, a set of robust constraints are generated. Finally, the resulting GP is solved to obtain an optimal assignment of the transistor W and L, which minimizes the area, and meets the delay constraints in the presence of random process variations.

6. EXPERIMENTAL RESULTS

The proposed robust gate sizing procedure was implemented in a C program, and an optimization software [16] was used to solve the final GP. All experiments were performed on a 2.4 GHz Linux machines having 2 GB of RAM. The robust gate sizing technique was applied to the ISCAS 85 benchmark circuits. The cell library selected comprised of inverters and two and three input NAND and NOR gates. We use a TSMC 180nm technology parameter [17] to estimate the constants for the on resistance and the source, drain and gate capacitances. We assume capacitive loading for the gates. The objective function chosen for the optimization is to minimize $Area = \sum_i m_i W_i L_i$, where m is the number of transistors in gate i.

In our method, the amount of guard-banding required in the face of process variations, is controlled by the size of the uncertainty ellipsoid, as determined by the entries of the covariance matrix, P. Each element of the P matrix is given by $P_{ij} = \rho_{ij}\sigma_i\sigma_j$, where the correlation component ρ_{ij} is obtained from the spatial correlation model described in Section 4. To use the spatial correlation model, we first place the circuits using the placement tool Capo [18], and then divide the chip area into different number of grids, depending on the circuit size, so that each grid size is of the order of $60\mu \times 60\mu$. We define $\sigma_{\mathbf{ref}}$ as the vector of maximum percentage deviations from the nominal values of W and L. The elements of $\sigma_{\mathbf{ref}}$ predicted from [15], are 25% of nominal width value and 20% of nominal channel length value. To calculate the elements of P matrix, we choose the value of $\sigma_i = K\sigma_{\mathbf{ref_i}}$, where K is a constant ≤ 1.

To verify the results of our method, we generate 10,000 random samples from a multivariate normal distribution $N(\mathbf{X_0}, Q)$, where, the mean vector of the distribution $\mathbf{X_0}$ is the vector containing the optimal set of transistor W and L as determined by the the robust GP solution, and the elements of covariance matrix Q are given by $Q_{ij} = \rho_{ij}\sigma_{ref_i}\sigma_{ref_j}$. We then perform Monte Carlo simulations using these 10,000 random samples, to determine the frequency of timing violations of the chip, i.e, the number of times the delay of the circuit exceeds T_{spec}. The timing yield of the robust design is compared to that of the conventional gate sizing solution or the non-robust design. The Monte Carlo samples for the non-robust design are generated using a mean vector $\mathbf{X_0}$, containing W and L, as determined by solving a standard GP optimization. For each circuit, the value of T_{spec} is chosen to be the point of 15% slack, i.e., $T_{spec} = D_{min} + 0.15(D_{max} - D_{min})$, where D_{min} and D_{max} are, respectively, the minimum and the maximum possible delays of the circuit. We found that for all the circuits, the optimizer assigns a value of $L = L_{min}$ to all channel length variables. This can be ascribed to the fact that increasing the channel length not only has an area penalty, it also increases the delay of the circuit and does not improve the robustness of the circuit in the face of variations. However, the robust GP determines a width assignment such that the random variations for both the width and the length are accounted for.

Ckt	# of Gates	Non Robust Design			Robust Design		
		Area	%Delay Violations	Run Time (sec)	Area	%Delay Violations	Run Time (sec)
C432	616	1.00	78.76%	3.45	1.17	0.00%	18.08
C499	1262	1.00	70.52%	7.49	1.23	0.01%	29.26
C880	854	1.00	72.36%	5.30	1.14	0.00%	22.17
C1355	1202	1.00	68.76%	7.33	1.20	0.00%	36.42
C1908	1636	1.00	65.43%	14.40	1.19	0.00%	307.88
C2670	2072	1.00	60.09%	20.52	1.21	0.03%	310.34
C3540	2882	1.00	67.12%	31.70	1.11	0.02%	342.14
C5315	4514	1.00	62.25%	65.12	1.18	0.01%	817.89
C6288	5548	1.00	63.36%	98.27	1.18	0.02%	1042.44
C7552	6524	1.00	65.12%	120.35	1.22	0.03%	1245.34

Table 1: A comparison of robust and non-robust gate sizing solutions

Table 1 shows a comparison of the robust designs (R) obtained by the proposed optimization scheme and non-robust designs (NR) obtained by the conventional GP solution. For the robust circuits shown in Table 1, the size of the variance ellipsoid is chosen such that the factor $K = \sigma/\sigma_{ref} = 1$. In this case, the P matrix characterizing the uncertainty ellipsoid around nominal values vector X_0, is same as the Q matrix used to draw Monte Carlo samples from

the multivariate normal distribution $N(\mathbf{X_0}, Q)$. This size of uncertainty ellipsoid corresponds to the case of maximum uncertainty-awareness. The third column in Table 1 shows the normalized area for the non-robust design, and the fourth column indicates the percentage of Monte Carlo chips that failed to meet the T_{spec} constraint for the conventional gate sizing solution. For the robust circuits, the area overhead to incorporate robustness is listed in column six of Table 1, and the percentage of timing violations in column seven. The run times for the NR and R designs are listed in columns five and eight, respectively. As seen from Table 1, the circuits obtained by the robust gate sizing scheme are able to eliminate the effect of process variations and increase the timing yield by about 3-4 times, e.g., the timing yield for C6288 increases from 33.88% to 99.98%. However, the area overhead for some cases could be significant, e.g., 23% for C499.

The sensitivity of the delay of a circuit with respect to the process variations depends on the amount of timing slack the circuits have. Circuits with smaller slacks or tighter delay specifications would be more sensitive to the random process variations. Hence, the area overhead to guard-band against the uncertainties would be greater for the robust designs. Figure 5(a) shows the effect of non-robust optimizing of the C499 circuit for different values of T_{spec} as given by the % slack points. As seen from the figure, the number of violations increase with tighter specs or smaller slacks. Figure 5(b) shows the corresponding robust designs for the C499 circuit for each of the slack points. As seen from the figure, the area overhead for R designs to eliminate the timing violations, increases for tighter specs.

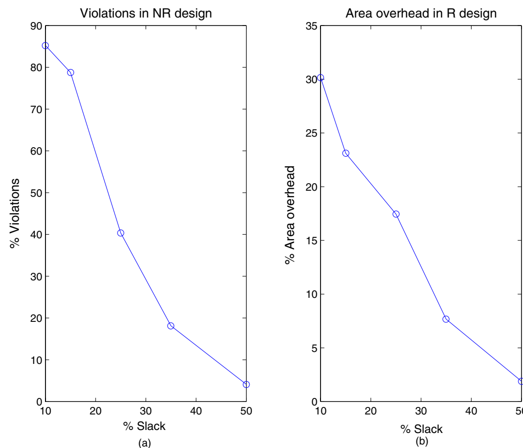

Figure 5: The non-robust and robust designs for C499 circuit for different values of T_{spec}. (a) Timing violations for non-robust designs. (b) Area overhead for robust designs.

We perform another series of experiments to compare our approach with a gate sizing methodology employing a worst-case design approach. The worst-case designs are obtained by iteratively solving the standard GP, but for delay specs tighter than the original required specs, until the area of the worst-case design is the same as that of the robust design. Furthermore, to explore the area-robustness trade-off we vary the size of the uncertainty ellipsoid, by choosing different values of the factor $K = \sigma/\sigma_{ref}$. We found in our experiments that the number of timing violations reduces with increase in area, for both the worst-case and the robust circuits. However, in all cases, our robust design has fewer violations than the worst-case design having the same area. On an average, the robust design has about 12% fewer violations that the worst-case design having the same area. The better performance of our robust sizing solution is not surprising because of the fact that the spatial correlation information, stored in the P matrix, is used by the optimization scheme. The worst-case circuit is expected to have a large overhead, since designing by setting tighter specs results in rendering critical some of the earlier non-critical paths. So the optimizer now has to aggressively size the gates on these paths, which results in greater chip area than actually required. Since, the runtimes for our robust gate sizing solutions are reasonably small, the user can run the optimization for different values of $K = \sigma/\sigma_{ref}$, to select the amount of robustness required against the process uncertainties, at the cost of additional chip area.

7. CONCLUSION

In this paper, we solve the gate sizing problem in the presence of process-driven variations. The procedure employs a ellipsoid set as a bounded variation model and considers the spatial correlations of the intra-die parameter variations. The original set of posynomial delay constraints are modified and converted to another set of posynomial constraints and the resulting robust GP is efficiently solved. Experimental results, on several benchmark circuits, show that the robust design significantly increases the timing yield of the chip as compared to the conventional gate sizing solutions and has a better performance than the worst-case designing approach.

8. REFERENCES

[1] J. Fishburn and A. Dunlop, "TILOS: A Posynomial Programming Approach to Transistor Sizing," in *Proc. IEEE/ACM ICCAD*, pp. 326–328, 1985.

[2] S. Sapatnekar, V. B. Rao, P. M. Vaidya, and S. M. Kang, "An Exact Solution to the Transistor Sizing Problem for CMOS Circuits Using Convex Optimization," in *IEEE Trans. on CAD*, vol. 12, pp. 1621–1634, Nov 1993.

[3] K. Kasamsetty, M. Ketkar, and S. S. Sapatnekar, "A New Class of Convex Functions for Delay Modeling and their Application to the Transistor Sizing Problem," in *IEEE Trans. on CAD*, vol. 19, pp. 779–788, Jul 1998.

[4] X. Bai, C. Visweswariah, P. N. Strenski, and D. J. Hathaway, "Uncertainty-Aware Circuit Optimization," in *Proc. ACM/IEEE DAC*, pp. 58–63, 2002.

[5] S. Raj, S. B. K. Vrudhala, and J. Wang, "A Methodology to Improve Timing Yield in the Presence of Process Variations," in *Proc. ACM/IEEE DAC*, pp. 448–453, 2004.

[6] S. H. Choi, B. C. Paul, and K. Roy, "Novel Sizing Algorithm for Yield Improvement under Process Variation in Nanometer Technology," in *Proc. ACM/IEEE DAC*, pp. 454–459, 2004.

[7] E. T. A. F. Jacobs and M. R. C. M. Berkelaar, "Gate Sizing Using a Statistical Delay Model," in *Proc. IEEE DATE*, pp. 283–291, 2000.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

[9] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, NJ, 2002.

[10] K. L. Hsiung, S. J. Kim and S. Boyd, "Robust Geometric Programming via Piecewise Linear Approximation," 2003. Submitted to Mathematical Programming, available at `http://www.stanford.edu/~boyd/rgp.html`.

[11] H. L. A-Malek and A.-K. S. O. Hassan, "The Ellipsoidal Technique for Design centering and Region Approximation," in *IEEE Trans. on CAD*, vol. 10, pp. 1006–1014, Aug 1991.

[12] S. Sapatnekar, P. M. Vaidya, and S. M. Kang, "Convexity-based Algorithms for Design Centering," in *IEEE Trans. on CAD*, vol. 13, pp. 1536–1549, Dec 1994.

[13] H. Chang and S. S. Sapatnekar, "Statistical Tming Analysis Considering Spatial Correlations Using a Single PERT-like Traversal," in *Proc. IEEE/ACM ICCAD*, pp. 621–625, 2003.

[14] A. Agarwal, D. Blaauw, and V. Zoltov, "Statistical Timing Analysis for Intra-Die Process Variations with Spatial Correlations," in *Proc. IEEE/ACM ICCAD*, pp. 900–907, 2003.

[15] S. Nassif, "Delay Variability: Sources, Impact and Trends," in *Proc. ISSCC*, pp. 368–369, 2000.

[16] Available at `http://www.mosek.com`.

[17] "TSMC: 180nm Test Data." Available at `http://www.mosis.org/Technical/Testdata/tsmc-018-prm.html`.

[18] A. Caldwell, A. B. Kahng and I. Markov, "Capo: a large-scale fixed-die placer." available at `http://vlsicad. ucsd.edu/GSRC/bookshelf/Slots/Placement.`