Statistical Leakage Minimization through Joint Selection of Gate Sizes, Gate Lengths and Threshold Voltage

Sarvesh Bhardwaj†, Yu Cao†, Sarma Vrudhula‡
†Electrical Engineering, ‡Computer Science and Engineering
Arizona State University, Tempe, AZ 85281
email:{sarvesh.bhardwaj, yu.cao, vrudhula}@asu.edu

Abstract

This paper[1] proposes a novel methodology for statistical leakage minimization of digital circuits. A function of mean and variance of the circuit leakage is minimized with constraint on α-percentile of the delay using physical delay models. Since the leakage is a strong function of the threshold voltage and gate length, considering them as design variables can provide significant amount of power savings. The leakage minimization problem is formulated as a **multivariable convex optimization** problem. We demonstrate that statistical optimization can lead to more than 37% savings in nominal leakage compared to worst-case techniques that perform only gate sizing.

I. Introduction

The leakage power has become a major cause of concern during the design of high performance nano-scale circuits. For example, it was shown in [4] that for 30% variations in the circuit delay there can be up to 20X variations in the leakage current. Scaling has also resulted in significant increase in the variations of the process and design parameters [1, 5, 3]. The most important of these variations are the variations in the effective channel length L_e, and the threshold voltage V_{th} which are due to a lack of precise control in the lithography and channel doping steps [5]. Variations in these two parameters have a significant effect on the sub-threshold leakage of a gate because of its exponential dependency on these two parameters.

The problem of leakage reduction has been addressed at the design stage by various techniques such as transistor stacking [15], sleep transistor insertion [13], body biasing [24, 16] and driving the circuit into a minimum leakage sleep state. The power savings accrued by these techniques can be further supplemented by gate sizing, dual-threshold voltage (V_{th}) and supply voltage (V_{dd}) assignment [22, 12, 9, 11] and L_e biasing [10]. However, to the best of our knowledge, L_e biasing has not been used in the past for leakage reduction in the presence of variations.

The traditional corner based design methodology treats the parameters as deterministic quantities and wastes expensive design resource in order to ensure a large guard-band on the design frequency as well as the power dissipation. Instead, a more effective methodology is to model the variations as random variables because of the stochastic nature of the underlying variations. Once this is done, the statistics (such as mean, variance etc.) of delay as well as leakage power of the circuit can be accurately estimated using the *probability density functions* (PDFs) of the parameters. A number of such statistical analysis techniques have been proposed recently [27, 20]. Although the proposed techniques predict the circuit delay or leakage accurately, their reliability with respect to what is manufactured greatly depends on the accuracy of the models used for the gate (and interconnect) delays and power [7]. While these models provide foundations for accurate statistical analysis, they are *necessary* for statistical optimization techniques.

A number of statistical optimization methods have been proposed recently [23, 19, 21, 14]. [23] uses a statistical timing analysis tool to check the satisfiability of the constraint on some percentile of the circuit delay. It then uses the *statistical sensitivities* to select the gates to be assigned high V_{th} as well as to be up-sized. In [19], a utility theoretic approach is used to identify a set of critical paths. The expected utility of the critical nodes (nodes on critical paths) is minimized subject to constraints on the expected delay and area. A robust circuit optimization technique is presented in [17] where the authors formulate the problem of maximizing the timing parametric yield as a geometric optimization problem with gate sizes as the decision variables. In [2], the problem of statistical leakage minimization using gate sizing is formulated as a geometric programming problem. The area minimization problem is solved in [21] by modeling the parametric variations using an uncertainty ellipsoid.

Introducing L_e and V_{th} as decision variables in the optimization problem instead of treating them fixed technological parameters, increases the size of the feasibility region. This can provide significant power savings that can not be achieved otherwise. Also, since V_{th} and its variance is dependent on L_e, L_e proves to be an extremely effective method of controlling the leakage variability. Instead of using a dual-V_{th} process, in which high V_{th} and low V_{th} are typically sep-

[1]Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This work was carried out at the NSF's State/Industry/University Cooperative Research Centers' (NSF-S/IUCRC) Center for Low Power Electronics (CLPE). CLPE is supported by the NSF (Grant #EEC-9523338), the State of Arizona, and an industrial consortium. This work was also supported by NSF through grant #CCR-0205227. An extended version of this paper is available at:http://veda.eas.asu.edu/papers/bhardwaj-aspdac06.pdf

0-7803-9451-8/06/$20.00 ©2006 IEEE.

arated by about $50mV$ for speed improvement and power reduction, we use a **single** V_{th} for the circuit. The use of dual-V_{th} produces extra process corners and thus, exaggerates variability. Specifically, some high V_{th} gates may be faster (and more leaky) than some low V_{th} gates due to the variation in V_{th}. This phenomenon has led to many design failures, especially in low-power applications.

The major contributions of this work are as follows:

1. Both the **mean** and the **variance** of the leakage are minimized by formulating a statistical optimization problem with **gate sizes, gate lengths** and **threshold voltage** as decision variables.
2. Experimentally verified statistical models for the gate delay and gate leakage are used that take into account the variability in various device parameters,
3. The leakage minimization problem is formulated as a **multivariable convex optimization** problem and an optimal solution is obtained.

Section II formally describes the problem of statistical leakage minimization. Section III describes the models used for the gate delay and gate leakage. The transformation of the optimization problem into a convex optimization problem is described Section IV. The Experimental results and conclusions are outlined in Section V and VI respectively.

II. Problem Formulation

Let a circuit be represented using a Directed Acyclic Graph (DAG) $G = (N, E)$, where $N = \{1, 2, .., n\}$ is the set of nodes and $E = \{(i, j) : i, j \in N\}$ is the set of edges. The nodes correspond to the gates in the original circuit. An edge $e_{ij} = (i, j)$ represents that gate i fanouts to gate j.

Let the parameter space for each gate i be defined as $\hat{u}_i = (u_1^i, u_2^i, .., u_r^i)$, where r denotes the number of parameters. In the presence of process variations, each of these parameters is a random variable. Hence, if Ω denotes the space of manufacturing outcomes, $\hat{u}_i : \Omega \rightarrow \mathbb{R}^r$ is a function that maps every outcome $\omega \in \Omega$ to a point in an r-dimensional Euclidean space. Hence, the parameters for the manufacturing outcome ω are given by $\widehat{u}_i(\omega) = (u_1^i(\omega), u_2^i(\omega), .., u_r^i(\omega))$. The **random parameters** considered in this work include the **gate length** ($L_{e,i}$) and **threshold voltage** ($V_{th,i}$). Although all the gates in the circuit are assigned the **same** V_{th}, the dependence of V_{th} on $L_{e,i}$ as well as the random variations cause the threshold voltage of each gate to be different. As channel length L_e becomes shorter, V_{th} exhibits a greater dependence on L_e and drain bias (DIBL). Larger V_{dd} and smaller L_e usually lead to sharp degradation in V_{th} (i.e., V_{th} roll-off). The $V_{th,i}$ of a gate can then be represented as

$$V_{th,i} = V_{tho} + 0.05 - V_{dd} e^{-\delta L_{e,i}}. \quad (1)$$

where V_{tho} is the long channel V_{th} and δ is the DIBL coefficient. For simplicity, V_{tho} and $L_{e,i}$ are modeled as independent normal random variables. Also, w_i and V_{dd} are modeled as a deterministic quantities. Henceforth, the explicit dependency of $\hat{u}$ on the argument ω will not be shown.

The circuit leakage and delay under this variational model are also random variables. Let I_S denote the sub-threshold leakage of the circuit and D_p denote the delay of path $p \in \mathcal{P}$, where $\mathcal{P}$ represents the set of paths in the circuit. The stochastic leakage minimization problem can now be formulated as follows

$$\min_{\omega \in \Omega} \quad I_S(\hat{u}_1, \hat{u}_2, .., \hat{u}_n, \omega) \quad (2)$$

$$\text{sub. to} \quad \mathbf{P}(D_p(\hat{u}_1, .., \hat{u}_n, \omega) \leq T_{req}) \geq \alpha \quad \forall\, p \in \mathcal{P}. \quad (3)$$

where $\mathbf{P}(X \leq x)$ denotes the probability that the random variable X is less than or equal to x. α can be considered to be a *confidence level*. As the number of manufacturing outcomes ω can be infinite, it does not make sense to solve the optimization problem for every ω. Hence, a more relevant objective would be some statistic (such as *mean* or *variance*) of the leakage current. Figure 1 shows the PDF of the leakage as a result of minimizing only the mean or the variance of the leakage. It can be seen that minimizing only the expected value of the leakage results in an increased number of chips having lower frequency (curve B). Whereas, minimizing just the variance without optimizing the mean leaves a scope of reduction in the leakage of the manufactured circuits (curve C). Hence the goal of maximizing the leakage yield can be achieved by minimizing a convex combination of the square of the mean and the variance of leakage. Thus, the new objective becomes $\lambda\, \mu^2(I_S) + (1 - \lambda)\, \sigma^2(I_S)$, where μ and σ^2 are the mean and variance of leakage. $\lambda \in [0, 1]$ controls the relative weight of the mean and the variance of the leakage in the objective.

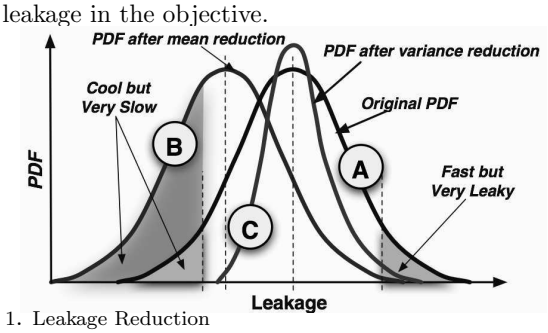

Fig. 1. Leakage Reduction

III. Leakage and Delay Models

A. Statistical Leakage Model

Let I_S represent the sub-threshold leakage of a circuit. As explained above, the I_S in the presence of the variations is random variable. In this work the sub-threshold leakage of a circuit is modeled using the model shown in (4) [25].

$$I_S = \sum_{i \in N} I_o \frac{w_i}{L_{e,i}^k} e^{\left(\frac{-V_{th,i}}{S}\right)}, \quad k > 1 \quad (4)$$

where $V_{th,i} = V_{tho} + 0.05 - V_{dd}(\delta_1 - \delta_2 L_{e,i})$ with $\delta_1 - \delta_2 L_{e,i}$ being approximation of $e^{-\delta L_{e,i}}$, $\delta_1, \delta_2 > 0$. I_o is th nominal sub-threshold leakage and k and S are positive fitting parameters. The summation in (4) is over all the nodes in the circuit. The above model captures the dependence of the sub-threshold leakage on the *all* the decision variables. Also, the dependence of V_{th} on L_e and V_{dd} has been taken into account. This model was fitted to the data from SPICE to obtain the parameters. From (4), we see that the sub-threshold leakage is inversely proportional to the gate length $L_{e,i}$. Since $L_{e,i}$ has been assumed to be a normally distributed random variable, the expectation of $L_{e,i}^{-k}$ does not

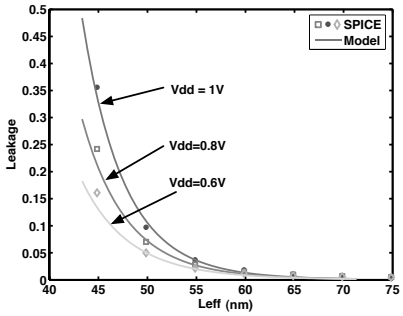

Fig. 2. Leakage Approximation

exist. Hence, we approximate the function $L_{e,i}^{-k}$ by writing it as $e^{-k \log L_{e,i}}$ and approximating $\log L_{e,i}$ by a quadratic function of $L_{e,i}$. Thus, the sub-threshold leakage is

$$I_S = \sum_{i \in N} I_o \frac{w_i}{L_{e,o}^k} e^{\left(\frac{-(V_{tho} + a_1 L_{e,i} + a_2 L_{e,i}^2)}{S}\right)}, \quad k > 1 \qquad (5)$$

where a_1 and a_2 are some constants. The accuracy of the above approximation is shown in figure 2 which shows the variation of sub-threshold leakage with $L_{e,i}$. The parameters V_{tho} and $L_{e,i}$ are modeled as $V_{tho} = V_{To} + V_\xi$ where, V_{To} is the designer specified value of the threshold voltage and V_ξ is a zero mean normal random variable $N(0, \sigma^2(V_{To}))$. Notice, that the variation in the threshold voltage is dependent on the specified value of V_{To}. The variations in V_{tho} are modeled using the Pelgrom's model [18] as

$$\sigma^2(V_{th,i}) = \frac{k}{L_{e,i} w_i}. \qquad (6)$$

Similarly, the gate length is also modeled as $L_{e,i} = L_{o,i} + L_{\xi,i}$ where $L_{\xi,i}$ is a zero mean normal random variable $N(0, \sigma_L^2)$. The variations in the gate length are assumed to be independent of the specified value of the gate length. Hence, the sub-threshold leakage of the circuit can be written as

$$I_S = \sum_{i \in N} I'_{o,i} e^{\left(\frac{-(V_\xi + (a_{11} + a_{12} V_{dd}) L_{\xi,i} + a_2 L_{\xi,i}^2)}{S}\right)}, \quad k > 1 \qquad (7)$$

where

$$I'_{o,i} = I_o \frac{w_i}{L_{e,o}^k} e^{\left(\frac{-(V_{To} + (a_{11} + a_{12} V_{dd}) L_{o,i} + a_2 L_{o,i}^2)}{S}\right)}, \quad k > 1 \qquad (8)$$

Now, $I'_{o,i}$ is a deterministic function of the assigned parameters. Also, as the underlying circuit parameters V_ξ and $L_{\xi,i}$ are assumed to be statistically independent, the mean of the leakage can be computed as shown in (9).

$$E[I_S] = \sum_{i \in N} I'_{o,i} E\left[e^{\left(\frac{-V_\xi}{S}\right)}\right] E\left[e^{\left(\frac{-((a_{11} + a_{12} V_{dd}) L_{\xi,i} + a_2 L_{\xi,i}^2)}{S}\right)}\right] \qquad (9)$$

The expectation of the two functions dependent on the V_ξ and $L_{\xi,i}$ can be obtained by computing the expectation of a random variable U that is an exponential function of a zero mean normal random variable $W \sim N(0, \sigma_W^2)$. Thus the two functions in (9) have a general form $U = \exp(-(W + aW^2)/b)$. Table I gives the values of a and b for the functions having an exponential dependence on V_ξ and $L_{\xi,i}$ in (9).

Since the leakage has an exponential dependency on a linear function of V_ξ, for the V_ξ dependent term, $a = 0$ in the general form above. The mean of U can be computed using (10).

TABLE I
VALUES OF a AND b FOR THE FUNCTIONS DEPENDENT ON V_ξ AND $L_{\xi,i}$

Parameter	a	b
V_ξ	0	S
$L_{\xi,i}$	$\frac{a_2}{(a_{11} + a_{12} V_{dd})}$	$\frac{S}{(a_{11} + a_{12} V_{dd})}$

$$E[U] = \left(1 + \frac{2a}{b}\sigma_W^2\right)^{-\frac{1}{2}} \cdot \exp\left(\frac{\sigma_W^2}{2b^2 + 4\sigma_W^2 ab}\right) \qquad (10)$$

The second moment of U can also be computed from (10) by replacing b by $b/2$. Similarly, the second moment of the leakage can be computed by computing the expectation of

$$I_S^2 = \sum_{i,j \in N} I'_{o,i} I'_{o,j} e^{\left(\frac{-(2V_\xi + (a_{11} + a_{12} V_{dd})(L_{\xi,i} + L_{\xi,j}) + a_2 (L_{\xi,i}^2 + L_{\xi,j}^2))}{S}\right)}, \qquad (11)$$

The procedure for computing the expectation of the above function $E[I_S^2]$ is analogous to the steps followed in computing the mean of the leakage. Using the second moment of the leakage, the variance can be computed using $\sigma^2(I_S) = E[I_S^2] - (E[I_S])^2$. The objective function of the leakage minimization problem can now be obtained using the mean and the variance of the leakage.

B. Statistical Delay Model

In the presence of process variations, the gate delays are random variables. For this work, we use the physical delay model proposed in [7]. Assuming that the transistors operate in the saturation mode, the proposed model can be simplified into the form shown in (12).

$$E[d_i] = \alpha\left(\frac{\beta_1}{w_i} + \beta_2\right) \frac{L_{e,i} V_{dd}}{(V_{dd} - V_{th,i})^2} \left(1 + \frac{(V_{dd} - V_{th,i})}{\gamma L_{e,i}}\right) \qquad (12)$$

The parameters for these models are obtained by performing SPICE simulation and by fitting these models to SPICE data. The accuracy of the mean of the delay is shown in Figure 3. The average error compared to the data from SPICE simulations is around 3-4% over ±25% range of $L_{e,i}$ for different values of the supply voltage and threshold voltage.

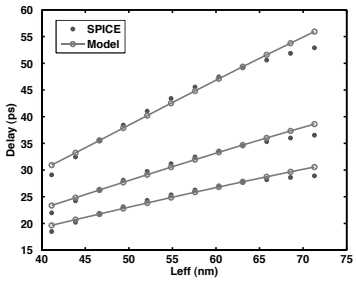

Fig. 3. Mean delay as a function of gate length for a NAND gate

Although the general form of the mean delay is not a linear function of the device parameters, for the values of the parameters in the saturation mode ($V_{dd} \in [0.8, 1.2]$) the delay can be safely assumed to have linear dependence of these parameters [7] (Figure 3 supports the linear dependence on L_e). Thus, the delay can be modeled as a normal random variable without having a significant impact on the accuracy.

Under this assumption, the probabilistic constraint in (3) is equivalent to

$$z_\alpha(D_p) = E[D_p] + z_\alpha \sigma(D_p) \leq T_{req} \quad (13)$$

If the gate delays are totally correlated, the above constraint can be translated in terms of the *mean* and the *standard deviation* of the gate delays as shown in (14).

$$z_\alpha(D_p) = \sum_{i \in p} E[d_i] + z_\alpha \sum_{i \in p} \sigma(d_i) \leq T_{req} \quad (14)$$

where d_i is the delay of a gate on path $p \in \mathcal{P}$. Thus the problem reduces to obtaining the expressions for the variance of the individual gate delays. For a particular gate, at a higher value of V_{th} and fixed V_{dd}, the sensitivity of the gate delay to V_{th} is very high. Instead at a lower value of the V_{th}, the gate delay is more sensitive to the V_{dd} compared to V_{th}. Hence the variation in the delay is more for larger values of the V_{th} (higher delay), although the variations in V_{th} might be small at higher values of V_{th}. Thus, in this work, the variance of the delay is computed using the model shown in (15).

$$\frac{\sigma(d_i)}{E[d_i]} = k_\sigma (E[d_i])^\zeta \quad (15)$$

where ζ and k_σ are fitting parameters. The α-percentile of the delay can now be computed as $z_\alpha(d_i) = E[d_i] + z_\alpha \sigma(d_i)$ using the models described above.

IV. Convex Optimization

A function f of variable $\mathbf{x} \in \mathbb{R}^{+n}$ is a posynomial if it has the form

$$f(\mathbf{x}) = \sum_j \beta_j \prod_{i=1}^{n} x_i^{\alpha_{ij}} \quad (16)$$

Posynomials have a useful property that they can be transformed into convex functions using the transformation $x_i = e^{y_i}$. Also, an exponential function of posynomials can be transformed into a convex function [6]. Since the objective function, which is a combination of the mean and the variance of the leakage, is a sum of exponential functions of *posynomials*, the objective function in the formulated optimization problem is convex.

In its original form, the expected delay is not a posynomial. However, it can be transformed into a posynomial by introducing the following inequality in the set of constraints

$$\frac{1}{V_{dd} - V_{th,i}} \leq t_i \quad (17)$$

and replacing $V_{dd} - V_{th,i}$ by t_i^{-1} in (12). The inequatily given in (17) is equivalent to

$$t_i^{-1} + V_{tho} + \delta_2 V_{dd} L_{e,i} \leq (1 + \delta_1) V_{dd} - 0.05. \quad (18)$$

Since $\delta_1 > 0$ and $V_{dd} \geq 0.8$, the RHS of this inequality is positive. Hence, it is a valid posynomial inequality. As a result of this inequality being a valid posynomial inequatily, the leakage minimization problem can be transformed into a convex optimization problem. Convex optimization problems are popular because efficient algorithms exist to solve them [26] as a locally optimal solution is also globally optimal. In this work, the convex optimization problem is solved using the optimization package LANCELOT [8]. The models used in the optimization correspond to the 90nm technological node. The range of the parameters, their nominal values and variances (where applicable) are given in Table II.

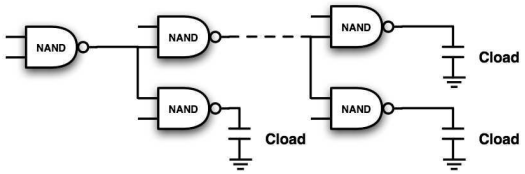

Fig. 4. Experimental circuit for the optimization problem

TABLE II
Circuit parameters for the 90nm technological node

	$L_{e,i}$ (nm)	$V_{tho}(V)$	W (size)
Mean	55.0	0.30	-
Std.Dev	5.5	0.01	-
Upper Bound	70.0	0.40	10
Lower Bound	55.0	0.15	1

TABLE III
Impact of including V_{th} as a decision variable. Results for delay = 0.349 ns

	V_{th}(V)	Area	Leakage (fA/ns)	
			Mean	Var. ($X10^{-2}$)
A (Before)	0.30	32	0.19	8.74
B (After)	0.28	12.9	0.14	4.56
Diff (%)	-	59.0	26	47.8

V. Experimental Results

A. Simultaneous V_{th} and Gate Sizing

In this section we discuss the savings that can be obtained by combining threshold voltage selection and gate sizing over simple gate sizing. For demonstration, we selected a chain of 10 NAND gates (similar to that in Figure 4) and performed gate sizing on the circuit with effective gate length and the threshold voltage fixed to their nominal values as shown in Table II. Also, V_{dd} was fixed to 1. The chain of 10 NAND gates is now optimized by minimizing the combination of mean and variance of leakage by considering V_{th} and gate sizes as decision variables. Compared to earlier works on gate sizing and threshold voltage assignment, this work performs statistical optimization and models the problem as a convex optimization and hence can guarantee the optimality of the solution while being efficient. V_{th} is treated as a continuous variable for the circuit as body biasing can provide significantly high level of granularity to achieve a given threshold voltage [24].

We start with a circuit optimized using only gate sizing with the rest of the parameters assigned to their nominal values given in Table II. We investigate the effect of changing the values of various parameters on different circuit attributes such as area and leakage with the delay constraint being fixed. It should be noted that if the delay constraint T_{req} is the minimum feasible delay $T_{req,min}$ at the nominal values of threshold voltage $V_{th,nom}$ and supply voltage $V_{dd,nom}$, then the optimal threshold voltage V_{th}^* in the circuit optimized with both V_{th} and w will be lower than $V_{th,nom}$ because decreasing the V_{th} decreases the delay.

Table III summarizes the value of the parameters and the circuit attributes before and after the optimization. The value of z_α is taken to be 3, which corresponds to 99% tim-

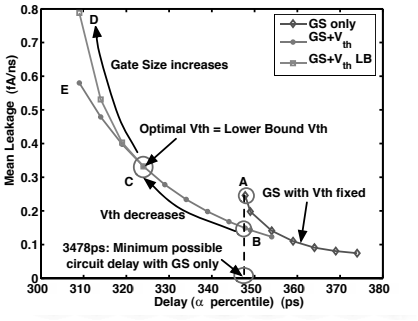

Fig. 5. Advantage of doing joint V_{th} and Gate Sizing

ing yield. The initial circuit A is obtained by performing only gate sizing with V_{th} and V_{dd} fixed to 0.3V and 1V respectively. The second circuit B is optimized by treating both the gate sizes and the threshold voltage as decision variables with V_{dd} fixed to 1V. As can be seen from the table, the circuit B has both lower area, lower mean leakage and lower variance of the leakage for the same value of the critical delay. By introducing V_{th} into the problem, we can get savings of up to 59% in the area and savings of 26% in the mean leakage. Thus we have significantly improved the leakage parametric yield without sacrificing the timing yield. Also, since the area of the circuit has reduced, it will lead to higher defect limited yield. Thus varying V_{th} is an effective method for optimizing leakage.

Figure 5 shows the trade-off of leakage and delay as a result of varying only sizes compared to simultaneous V_{th} selection and gate sizing. GS corresponds to the design methodology using only gate sizes as decision variables. As can be seen from the figure, point A corresponds to the minimum possible delay that can be achieved using GS only. However, if the V_{th} is introduced as an additional variable (with a lower bound = 0.25V) in the optimization, the feasible region is increased (because V_{th} can be changed) and we get a reduction in the objective function (point B). At point B, the threshold voltage is lower than that at point A and the area of the circuit corresponding to point B is lower than that of the circuit corresponding to point A. Also, from point B to point C, the threshold voltage proves to be the most effective method of reducing the circuit delay. For every optimal circuit between point B and point C, only the V_{th} is different, the area of all the circuits corresponding to points between B and C is the same.

At point C, the V_{th} reaches its lower bound of 0.25V. Since the threshold voltage cannot decrease any further, to achieve the critical delay, gate sizes have to be increased and thus the optimal circuits between C and D have only different area and their V_{th} is fixed to the minimum value. However, if the V_{th} were allowed to decrease further than 0.25V, we would have obtained optimal circuits having lower mean leakage as shown by the curve between the points C and E.

B. Simultaneous V_{th}, L_e *and* Gate Sizing

This section discusses the effect of introducing effective gate length and L_e as decision variables in the optimization along with V_{th} and gate sizes. Since the leakage has an exponential dependence on L_e, it is an effective factor for leakage

TABLE IV
CIRCUIT PARAMETERS FOR OPTIMIZED CIRCUIT WITH DIFFERENT METHODS FOR 99-PERCENTILE CIRCUIT DELAY OF 349 PS

	V_{th}(V)	L_e (nm)	Area	Leak. (fA/ns)	
				Mean	Var. ($X10^{-2}$)
GS	3.0	54.8	32.0	0.19	8.74
GSV	2.8	54.8	12.9	0.14	4.56
GSVL	1.71	70.0	13.2	0.12	1.49

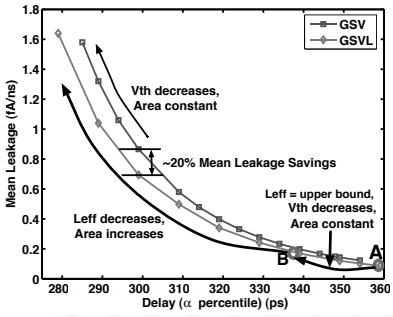

Fig. 6. Leakage-Delay tradeoffs for GSV and GSVL

reduction. We now show that L_e biasing provides significant leakage savings compared to the leakage savings obtained by simultaneous V_{th} and gate sizing. Table IV shows the parameters for the optimal circuit obtained by only gate sizing (GS), gate sizing and V_{th} assignment (GSV) and gate sizing, V_{th} and L_e biasing ($GSVL$) with a 99-percentile delay of 349ps and V_{dd} fixed to 1V. Changing the L_e from a fixed value to a variable causes the L_e to be assigned to its maximum value if the delay constraint is not very tight. The increase in L_e is shown in the third row and third column of the table. At this value of T_{req}, all the gates in the circuit have same L_e. Since, increasing the L_e increases the delay, the increase in the delay is compensated by decreasing the value of V_{th} as shown in the second column of GSVL. Thus the area of the circuit does not need to be increased. Hence, the overall effect is that the mean of the leakage reduces by around 15% with the area increasing by only 2%. Another important thing to be noticed is that the variance of the leakage decreases considerably. This is partly due to the fact that the variance of the threshold voltage reduces with increase in L_e as a result of Pelgrom's model. Thus the leakage parametric yield improves significantly without having a negative impact on the timing yield or the area of the circuit.

Figure 6 shows the leakage-delay trade-off for the two methods GSV and GSVL. In GSV, when the delay is reduced, the V_{th} has to be reduced to meet the timing constraint and thus the leakage increases. Instead, in GSVL, when the delay constraint is loose, the assignment of maximum value of L_e to all the gates in the circuit provides a leakage optimal design. Since increasing L_e increases the delay as well, V_{th} has to be reduced to satisfy the delay constraint. Hence, V_{th} for GSVL is lower than the V_{th} obtained after performing GSV for the same delay constraint. Thus, as shown in Figure 6, from point A to point B, the increase in L_e is compensated by the reduction in V_{th} and the area remains constant. At point B, the optimal design has the

TABLE V
AREA-LEAKAGE TRADE-OFF AS A RESULT OF INCLUDING L_e IN THE OPTIMIZATION

Delay (ps)	Area		Leakage (10^{-1}) (fA/ns)		Std. Dev. Leakage (10^{-1}) (fA/ns)	
	GSV	GSVL	GSV	GSVL	GSV	GSVL
289	13.54	24.83	13.2	10.3	18.73	12.00
299	13.42	24.89	8.67	6.95	12.25	7.64
309	13.42	23.94	5.81	4.98	8.21	5.25
319	13.21	18.32	3.98	3.41	5.63	3.53
329	13.12	14.78	2.78	2.41	3.93	2.45

minimum value of the $V_{th} = 0.15V$. Hence, beyond point B V_{th} cannot be decreased. Thus the decrease in circuit delay is achieved by reduction in L_e as well as increase in the gate sizes. The leakage savings in the mean leakage by including L_e in the optimization is around 20% more than that obtained using just GSV and around 37% compared to GS. The savings only increase as the delay constraint is tightened.

Table V compares the area, mean and the standard deviation of the leakage of the optimized circuit obtained by using GSV and GSVL for different values of the required time. We see from column 3 that the leakage reduction is obtained by using GSVL but at the cost of the increased area. From columns 5 and 7, we see that the mean of the leakage and the standard deviation of the leakage of the optimal circuit obtained using GSVL is much lower than that of a circuit obtained using only GSV.

VI. Conclusions

In this paper, we presented a novel methodology for simultaneously varying the threshold voltage, gate sizes and the gate length of a circuit to achieve a minimum leakage circuit. We included the effect of various process variations on different circuit parameters. A function of both mean and variance of the leakage was minimized with constraints on the α-percentile of the circuit delay. Also, to the best of our knowledge, this is the first work to include L_e as a method to reduce leakage variability. We showed that simultaneously using V_{th}, gate sizes and L_e provide significant improvement in the leakage parametric yield. We also demonstrated that we can obtain a considerably better circuit in terms of leakage and area by introducing L_e and V_{th} as decision variables in the optimization problem in addition to the gate sizes.

References

[1] International technology roadmap for semiconductors. 2003.

[2] S. Bhardwaj and S. Vrudhula. Leakage minimizatio of nano-scale circuits in the presence of systematic and random variations. In *Proceedings Design Automation Conference (DAC)*, 2005.

[3] D. Boning and S. Nassif. *Models of process variations in device and interconnect, Design of High-Performance Microprocessor Circuits*, chapter 6. IEEE Press, 2000.

[4] S. Borkar et al. Parameteric variations and impact on circuits and microarchitecture. In *Proc. DAC*, 2003.

[5] K. A. Bowman, S. G. Duvall, and J. D. Meindl. Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration. *JSSC*, 37(2):183–190, Feb 2002.

[6] S. Boyd, S. J. Kim, L. Vandenberghe, and A. Hassibi. A tutorial on geometric programming. Technical report, www.stanford.edu/∼boyd/gp_tutorial.html, 2004.

[7] Y. Cao and L. T. Clark. Mapping statistical process variations toward circuit performance variability: An analytical modeling approach. In *Proc. of DAC*, 2005.

[8] A. R. Conn, N. I. M. Gould, and P. L. Toint. *LANCELOT*. Springer-Verlag, 1992.

[9] W. H. et. al. Total power optimization through simultaneously multiple-v_{DD} multiple-v_{TH} assignment and device sizing with stack forcing. In *Proc. of ISLPED*, 2004.

[10] P. Gupta, A. B. Kahng, P. Sharma, and D. Sylvester. Selective gate-length biasing for cost-effective runtime leakage control. In *Proc. of DAC*, pages 327–330, 2004.

[11] M. Ketkar and S. S. Sapatnekar. Standby power optimization via transistor sizing and dual threshold voltage assignment. In *Proc. of ICCAD*, pages 375 – 378, 2002.

[12] D. Lee, H. Deogun, D. Blaauw, and D. Sylvester. Simultaneous state, Vt and Tox assignment for total standby power minimization. In *Proc. of DATE*, 2004.

[13] C. Long and L. He. Distributed sleep transistors network for power reduction. In *Proc of DAC*, pages 181 – 186, 2003.

[14] M. Mani, A. Devgan, and M. Orshansky. An efficient algorithm for statistical minimization of total power under timing yield constraints. In *ACM/IEEE Design Automation Conference*, 2005.

[15] S. Narendra et al. Full-chip subthreshold leakage power prediction and reduction techniques for sub-0.18-μ/m CMOS. *Journal of Solid-State Circuits*, 39(2):501–510, Feb 2004.

[16] C. Neau and K. Roy. Optimal body bias selection for leakage Improvement and Process Compensation over different technology generations. In *ISLPED*, pages 116–121, 2003.

[17] D. Patil et al. A new method for design of robust digital circuits. In *Proc. of ISQED*, 2005.

[18] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers. Matching properties of mos transistors. *IEEE Journal of Solid-State Circuits*, 24(5):1433–1439, Oct 1989.

[19] S. Raj, S. Vrudhula, and J. M. Wang. A methodology to improve timing yield in the presence of process variations. In *Proc. of DAC*, pages 448–453, 2004.

[20] R. Rao, A. Devgan, D. Blaauw, and D. Sylvester. Parametric yield estimation considering leakage variability. In *Proc. of DAC*, pages 442–447, 2004.

[21] J. Singh, V. Nookala, Z.-Q. Luo, and S. Sapatnekar. Robust gate sizing using geometric programming. In *ACM/IEEE Design Automation Conference*, 2005.

[22] S. Sirichotiyakul et al. Stand-by power minimization through simultaneous threshold voltage selection and circuit sizing. In *Proc. of DAC*, pages 436 – 441, 1999.

[23] A. Srivastava et al. Statistical Optimization of Leakage Power Considering Process Variations using Dual-Vth and Sizing. In *Proc. of DAC*, pages 773–778, 2004.

[24] J. Tschanz et al. Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage. *IEEE Journal of Solid-State Circuits*, 37(11):1396–1402, 2002.

[25] UC Berkeley Device Group. *BSIM 4.2.1 MOSFET Model - User's Manual*, 2004.

[26] P. M. Vaidya. A new algorithm for minimizing convex functions over convex sets. In *30th Annual Symposium on Foundations of Computer Science*, 1989.

[27] C. Visweswariah, K. Ravindran, K. Kalafala, S. G. Walker, and S. Narayan. First-order incremental block-based statistical timing analysis. In *Proc. of DAC*, pages 331–336, 2004.