Statistical Leakage and Timing Optimization for Submicron Process Variation

Yuanlin Lu and Vishwani D. Agrawal

Auburn University
Department of Electrical and Computer Engineering
Auburn, AL 36849, USA
luyuanl@auburn.edu, vagrawal@eng.auburn.edu

Abstract

Leakage power is becoming a dominant contributor to the total power consumption and dual-V_{th} assignment is an efficient technique to decrease leakage power, for which effective design methods have been proposed. However, due to the exponential relation of subthreshold current with process parameters, such as, the effective gate length, oxide thickness and doping concentration, process variations can severely affect both power and timing yields of the designs obtained by those methods. In this paper, we propose a mixed integer linear programming method for dual-V_{th} design that minimizes the leakage power and circuit delay in a statistical sense such that the impact of process variation on the respective yields is minimized. The experimental results show that 30% more leakage power reduction can be achieved by using statistical approach when compared with the deterministic approach.

1. Introduction

With the continuing trend in the CMOS technology scaling, leakage power is becoming a dominant contributor to the total power consumption. To reduce leakage power, many techniques have been proposed, including transistor sizing, multi-V_{th}, dual-V_{th}, optimal standby input vector selection, dual-V_{DD}, transistor stacking, and body bias.

Dual-V_{th} assignment [1-6, 7-9, 11, 12] is an efficient technique for leakage reduction. Traditional deterministic approaches for dual-threshold assignment [1-6, 11, 12] to minimize the leakage power utilize the timing slack of non-critical paths to assign high V_{th} to some or all gates on those paths to decrease the leakage. These approaches can be divided into two groups: heuristic algorithms [1-4] and linear programming [5, 6, 11, 12]. Unlike a heuristic algorithm that can only guarantee a locally optimal solution, linear programming formulation ensures a global optimization.

However, the increased variation of process parameters of nanoscale devices can cause a significant increase in the leakage current because of an exponential relation between the leakage current and some key process parameters. Gate leakage is most sensitive to the variation in oxide thickness (T_{ox}). The subthreshold current is extremely sensitive to the variation in oxide thickness (T_{ox}), effective gate length (L_{eff}) and doping concentration (N_{dop}). Compared with the gate leakage, the subthreshold leakage is more sensitive to parameter variations [10]. Twenty percent variations in effective channel length and oxide thickness can cause up to 13 and 15 times differences, respectively, in the amount of subthreshold leakage current. Gate leakage can have 8 times difference due to a 20% variation in oxide thickness.

Variation of process parameters not only affects the leakage current but also changes the gate delay, degrading either one or both, power and timing yields of an optimized design. To minimize the effect of process variation, some techniques [7-9] statistically optimize the leakage power and circuit performance by dual-V_{th} assignment. Leakage current and delay are treated as random variables. A dynamic programming approach for leakage optimization by dual-V_{th} assignment has been proposed [7] using two pruning criteria that stochastically identify pareto-optimal solutions and prune the sub-optimal ones. Another approach [9] solves the statistical leakage minimization problem using a theoretically rigorous formulation for dual-V_{th} assignment and gate sizing.

Our work is motivated by the above research. A mixed integer linear programming (MILP) model is proposed to minimize leakage power with specified timing yield under process variation. This MILP method is specifically devised with a set of constraints whose size is linear in the number of gates. Although theoretical worst-case complexity of MILP is exponential, our experimental results show that actual complexity depends on the nature of the problem. To deal with the complexities of delay models and leakage calculation, two look up tables for the nominal delay and leakage current are pre-constructed for each cell. This greatly simplifies the optimization procedure. Experimental results show that 30% more reduction of leakage power can be achieved by using the statistical approach when the result is compared to a deterministic approach.

This paper is organized as follows. Section 2 presents a deterministic linear programming formulation. Section 3

discusses the statistical leakage and gate delay modeling, and proposes a statistical linear programming method for leakage minimization under process variation. In Section 4, experimental results are presented and discussed. A conclusion is given in Section 5.

2. Deterministic Dual-V_{th} ILP

In the deterministic approach, the delay and subthreshold current of every gate are assumed to be fixed and without any effect of the process parameter variation. Basically, this type of methods can be divided into two groups: heuristic algorithms [1-4] and linear programming [5-6, 11, 12]. Heuristic algorithms give a locally optimal solution and linear programming formulation ensures a globally optimum solution.

An ILP that optimizes the leakage power and assigns dual-V_{th} to gates in one step [11, 12] has an advantage over an iterative procedure [5], which must assume power-delay sensitivities to be constants in a small range. Figure 1 gives the basic idea of the ILP method [11, 12] to minimize total subthreshold leakage while keeping the circuit performance by dual-V_{th} assignment.

Minimize	$\sum_i I_{subnom,i}$	$\forall i \in$ *gate number*
Subject to	$T_{POk} \le T_{\max}$	$\forall k \in PO$

Figure 1. Basic idea of using ILP to optimize leakage.

Minimize

$$\sum_i \left\{ X_i \cdot I_{subnom,L,i} + (1 - X_i) \cdot I_{subnom,H,i} \right\} \quad \forall i \quad \text{(D-O)}$$

Subject to

$$D_i = X_i \cdot D_{nom,L,i} + (1 - X_i) \cdot D_{nom,H,i} \quad \forall i \quad \text{(D-C1)}$$

$$T_i \ge T_j + D_i \quad \forall j \in \textit{fanin of gate } i \quad \text{(D-C2)}$$

$$X_i = 0 \ \ or \ \ 1 \quad \forall i \quad \text{(D-C3)}$$

$$T_{POk} \le T_{\max} \quad \forall k \in PO \quad \text{(D-C4)}$$

Figure 2. Detailed deterministic ILP formulation for leakage minimization.

A detailed version for the ILP formulation is presented in Figure 2. X_i is an integer that can only be either 0 or 1. A value 1 means that gate i is assigned low V_{th}, and 0 means that gate i is assigned high V_{th}. T_i is the latest arrival time at the output of gate i. Each gate in the design library with low and high threshold versions is characterized for its leakage in various input states and gate delay, which also depends on the fanout number, using Spice simulation.

3. Statistical Dual-V_{th} Assignment

Process variations include inter-die and intra-die variations, or global and local variations. For inter-die variations, the deterministic and statistical approaches are exactly the same. Since our objective is to have a statistical ILP formulation that enhances the deterministic approach to leakage optimization under process variations, we ignore the inter-die variation. In the remainder of this paper, process variation will only mean intra-die variation.

Leakage current is composed of reverse biased PN junction leakage, gate leakage and subthreshold leakage. In a sub-micron process, PN junction leakage is much smaller than the other two components. Gate leakage is most sensitive to the variation in T_{ox} and changes in the gate leakage due to other process parameter variations can be ignored [10]. Further, assuming T_{ox} to be a well-controlled process parameter [14, 15], we ignore the gate leakage variation in our design, focusing only on changes in the subthreshold leakage due to process variation.

Due to the exponential relation of subthreshold current with process parameters, such as, the effective gate length, oxide thickness and doping concentration, process variation can severely affect both power and timing yields of a design obtained by a deterministic method. Because fixed subthreshold leakage and gate delay do not represent the real circuit condition, statistical modeling should be used. This is discussed next.

3.1 Statistical Subthreshold Leakage Modeling

Subthreshold current has an exponential relation with the threshold voltage, which in turn is a function of oxide thickness, effective channel length, doping concentration, etc. T_{ox} is a fairly well-controlled process parameter and does not significantly influence subthreshold leakage variation [14, 15]. Therefore, we only consider variations in L_{eff} and N_{dop}.

The statistical subthreshold model can be written as [15]:

$$I_{sub} = I_{sub,nom} \cdot \exp\left(-\frac{\Delta L_{eff} + c_2 \Delta L_{Leff}^{\ 2} + c_3 \Delta V_{th,Ndop}}{c_1} \right) \quad (1)$$

Where, ΔL_{eff} is the change in the effective channel length due to the process variation and $\Delta V_{th,Ndop}$ is the change in the threshold voltage due to the random distribution of doping concentration, N_{dop}. Both are random variables with a normal distribution, N(0,1). Fitting parameters c_1, c_2 and c_3 are determined from Spice simulation.

From equation (1), it is obvious that I_{sub} has a lognormal distribution. The total leakage current in a circuit, which is the sum of subthreshold currents of individual gates, has an approximately lognormal distribution. Rao *et al.* [15] use the *central limit theorem* to estimate this lognormal distribution by its mean value with the assumption that there is a large number of gates in the circuit, which indeed is the case for most present day chips. Hence, the total leakage can be expressed as:

$$I_{sub,total} = \sum_i I_{sub,i} \cong E\left[\sum_i I_{sub,i}\right] = S_L \cdot S_V \cdot \sum_i I_{subnom,i} \quad (2)$$

Where,

$$S_L = \frac{1}{\sqrt{1+\frac{2\lambda_2}{\lambda_1}\sigma^2_{\Delta L_{eff}}}} \cdot \exp\left(\frac{\sigma^2_{\Delta L_{eff}}}{2\lambda_1^2 + 4\sigma^2_{\Delta L_{eff}}\lambda_1\lambda_2}\right) \quad (3)$$

$$S_V = \exp\left(\frac{\lambda_3^2\sigma^2_{\Delta V_{th,Ndop}}}{2\lambda_1^2}\right) \quad (4)$$

S_L and S_V are scale factors introduced due to local variations in L_{eff} and $V_{th,Ndop.}$ λ_1, λ_2 and λ_3 are fitting parameters. For a given process, $\sigma_{\Delta Leff}$ and $\sigma_{\Delta Vth,Ndop}$ are predetermined. Therefore, in our statistical linear programming formulation, the objective function is a sum of all nominal subthreshold leakage currents, multiplied by scale factors, S_L and S_V.

3.2 Statistic Delay Modeling

The deterministic gate delay D is given by [13]:

$$D \propto \frac{CV_{dd}}{(V_{dd} - V_{th})^{\alpha}} \quad (5)$$

where α equals 1.3 for the short channel model. Similar to the subthreshold current model, the V_{th} deviation due to the process parameter variation is also a consideration in our statistical delay model. The change of V_{th} due to the variation of process parameters can be expressed as [7]:

$$V_{th} = V_{th0} - \sum_i \beta_{X_i} \frac{X_{i0} - X_i}{X_{i0}} \quad (6)$$

where X_i is a process parameter, X_{i0} is the nominal value of X_i, and β_{Xi} is a constant for the specific technology.

To get an approximated linear relation between D and the variations of the process parameters, equation (5) is expanded into a Taylor series (7) in which only the first order term is retained because higher orders terms are relatively small and can be ignored.

$$D_{X_1,X_2,\ldots} = D(X_{10,}X_{20,}\ldots) + \sum (X_{i0} - X_i)\frac{dD}{dX_i}|_{X_{i0}} \quad (7)$$

Let $\{X_1, X_2, X_3\} = \{L_{eff}, T_{ox}, N_{dop}\}$. Combining equations (6) and (7), we get:

$$D_i = D_{nom,i}\left(1 + c_{i1}\frac{\Delta L_{eff}}{L_{eff\,0}} + c_{i2}\frac{\Delta T_{ox}}{T_{ox0}} + c_{i3}\frac{\Delta N_{dop}}{N_{dop0}}\right) \quad (8)$$

where, c_{i1}, c_{i2}, and c_{i3} are sensitivities of gate delay with respect to the variation of each process parameter and can be acquired from Spice simulation. L_{eff}, T_{ox} and N_{dop} are normal N(0,1) random variables. Therefore, in statistical analysis, D_i becomes a random variable, which also has a normal distribution.

$$\text{Let } r_i = c_{i1}\frac{\Delta L_{eff}}{L_{eff\,0}} + c_{i2}\frac{\Delta T_{ox}}{T_{ox0}} + c_{i3}\frac{\Delta N_{dop}}{N_{dop0}}, \quad (9)$$

Equation (8) becomes:

$$D_i = D_{nom,i}(1 + r_i) \quad (10)$$

Because r_i is a normal N(0,1) random variable, μ_{Di}, the mean value of D_i, is equivalent to D_{nomi}, the nominal delay of gate i.

3.3 Statistical Dual-V_{th} Assignment ILP

In statistical approach to minimize leakage power by dual-V_{th} assignment (Figure 3), the delay and subthreshold current are both random variables, and η is the expected timing yield. The power yield is not considered because in Section 4 (Results) we will find that the statistical approach can get about 30% additional leakage power reduction for most circuits compared to the deterministic approach.

Minimize	$I_{sub,total}$	(11)
Subject to	$P(T_{POi} \le T_{\max}) \ge \eta$	(12)

Figure 3. Basic ILP for statistical dual-V_{th} assignment.

In Figure 3, T_{POi} is the path delay from primary input to the i_{th} primary output and is assumed to have a normal (Gaussian) distribution $N(\mu_{TPOi}, \sigma^2_{TPOi})$. Inequality (12) allows leakage to be optimized with timing yield η and it can be expressed into a linear format by the percent point function Φ^{-1} [16]:

$$T_{POi} + \sigma_{POi} \cdot \Phi^{-1}(\eta) \le T_{\max} \quad (13)$$

In statistical linear programming (Figure 4) all variables, except X_i, are random variables with normal distribution. Comparing the deterministic ILP (Figure 2) and statistical ILP (Figure 4), we observe the following differences:

- The deterministic gate delay in (D-C1) is extended to (S-C1) and (S-C2) to get the mean and standard deviation of the statistical delay.
- (D-C2) is extended to to (S-C3) through (S-C6) to get the mean and standard deviation of the statistical arrival time T_i at the output of gate i.

- (D-C4) is updated to (S-C8) to ensure certain timing yield under process variation.

Minimize

$$S_L \cdot S_V \cdot \sum_i I_{subnom,i}$$
$$= S_L \cdot S_V \cdot \sum_i \{X_i \cdot I_{subnom,L,i} + (1 - X_i) \cdot I_{subnom,H,i}\}$$
$$\forall i \in \text{gate number} \qquad \text{(S-O)}$$

Subject to

$$\forall i \in \text{gate number}$$
$$\mu_{Di} = X_i \cdot D_{nom,L,i} + (1 - X_i) \cdot D_{nom,H,i} \quad \text{(S-C1)}$$
$$\sigma_{Di} = r_i \cdot \mu_{Di} \quad \text{(S-C2)}$$
$$\mu_{Ti} \geq \mu_{Tj} + \mu_{Di} \quad \forall j \in \text{fan in of gate } i \quad \text{(S-C3)}$$
$$\sigma_{Tj,Di} = k(\sigma_{Tj} + \sigma_{Di}) \quad \text{(S-C4)}$$
$$temp_{Ti} \geq \mu_{Tj} + \mu_{Di} + 3\sigma_{Tj,Di} \quad \text{(S-C5)}$$
$$\sigma_{Ti} = (temp_{Ti} - \mu_{Ti})/3 \quad \text{(S-C6)}$$
$$X_i = 0 \text{ or } 1 \quad \text{(S-C7)}$$
$$\forall k \in PO$$
$$T_{PO_k} + \sigma_{PO_k} \cdot \Phi^{-1}(\eta) \leq T_{\max} \quad \text{(S-C8)}$$

Figure 4. Detailed formulation of statistical dual-V_{th} assignment ILP.

3.4 Linear Approximations

In linear programming, all the expressions and constraints should be linear functions. However, in statistical analysis, some nonlinear operations are present. We, therefore, use linear approximations.

- **ADD, *A* = *B* + *C***

If *B* and *C* are N(μ, σ²) random variables, then their sum *A* also has a normal distribution. 'Add' is a linear function, but in statistical analysis, to obtain the standard deviation σ_A, we must deal with $\sigma_A^2 = \sigma_B^2 + \sigma_C^2$, which is a nonlinear operation. Considering,

$$\frac{(\sigma_B + \sigma_C)^2}{2} \leq \sigma_B^2 + \sigma_C^2 \leq (\sigma_B + \sigma_C)^2$$

one can find a linear approximation [19, 20]:

$$\sigma_A = \sqrt{\sigma_B^2 + \sigma_C^2} = k(\sigma_B + \sigma_C) \quad \text{with} \quad k \in [\frac{\sqrt{2}}{2}, 1]$$

- **MAX, *A* = *MAX(B, C)***

If *B* and *C* are N(μ,σ^2) random variables, *A* does not necessarily have a normal distribution [17, 18]. However, a normal approximation with following mean and standard deviation has been used [19, 20]:

$$\mu_A = \max(\mu_B, \mu_C) \quad (14)$$
$$\sigma_A = \{\max(\mu_B + 3\sigma_B, \mu_C + 3\sigma_C) - \mu_A\}/3 \quad (15)$$

The error in this approximation has been shown to be small [19, 20].

- **MAX, *A* = *MAX(B+D, C+D)***

Similarly, for function *A=Max(B+D,C+D)*, we use (16) and (17) to estimate the mean value and standard deviation of random variable *A*.

$$\mu_A = \max(\mu_B, \mu_C) + \mu_D \quad (16)$$
$$\sigma_A = \{\max(\mu_B + \mu_D + 3\sigma_{BD}, \mu_C + \mu_D + 3\sigma_{CD}) - \mu_A\}/3 \quad (17)$$

The above linear approximations are used in our statistical analysis to model the leakage optimization problem under process variation by a linear programming formulation.

4. Results

We use the BPTM 70nm CMOS technology [22]. Low V_{th} for NMOS and PMOS are 0.20V and -0.22V, respectively. High V_{th} for NMOS and PMOS are 0.32V and -0.34V, respectively. We regenerated the netlists of ISCAS'85 benchmark circuits using a cell library in which the maximum gate fanin is five. Two look-up tables for nominal gate delays and nominal leakage currents, respectively, for each type of cell were constructed using Spice simulation. A *C* program parsed the netlist and generated the constraint set for the CPLEX ILP solver in the AMPL software package [21]. CPLEX then gave the optimal V_{th} assignment as well as the minimized leakage current for the circuit.

4.1 Comparison of Leakage Power Reduction by Deterministic and Statistical Methods

To compare the power optimization results of the statistical ILP with those from the deterministic approach, we assume that all the gates have the same c_{i1}, c_{i2} and c_{i3} (sensitivities of gate delay to the variation of different process parameters) in equation (8). Therefore, each gate has the same r_i. (=σ_{Di}/u_{Di}). We assume it to be 10%. This assumption is only for the simplicity and does not change the efficacy of the statistical approach.

In Table 1, columns 4, 6 and 9 give the optimized leakage power by deterministic ILP, by statistical ILP with 99% timing yield and by statistical ILP with 95% timing yield.

Table 1. Comparison of leakage power saving due to statistical modeling with two different timing yields (η).

Circuit			*Deterministic Optimization (η =100%)*		*Statistical Optimization (η =99%)*			*Statistical Optimization (η =95%)*		
Name	**# gates**	**Unoptimized Leakage Power (µW)**	**Optimized Leakage Power (µW)**	**Run Time (s)**	**Optimized Leakage Power (µW)**	**Extra Power Saving**	**Run Time (s)**	**Optimized Leakage Power (µW)**	**Extra Power Saving**	**Run Time (s)**
C432	160	2.620	1.003	0.00	0.662	33.9%	0.44	0.589	41.3%	0.32
C499	182	4.293	3.396	0.02	3.396	0.0%	0.22	2.323	31.6%	1.47
C880	328	4.406	0.526	0.02	0.367	30.2%	0.18	0.340	35.4%	0.18
C1355	214	4.388	3.153	0.00	3.044	3.5%	0.17	2.158	31.6%	0.48
C1908	319	6.023	1.179	0.03	1.392	21.7%	11.21	1.169	34.3%	17.45
C2670	362	5.925	0.565	0.03	0.298	47.2%	0.35	0.283	49.8%	0.43
C3540	1097	15.622	0.957	0.13	0.475	50.4%	0.24	0.435	54.5%	1.17
C5315	1165	19.332	2.716	1.88	1.194	56.0%	67.63	0.956	64.8%	19.7
C7552	1045	22.043	0.938	0.44	0.751	20.0%	0.88	0.677	27.9%	0.58
Average of ISCAS'85 benchmarks				**0.24**		**29.2%**	**9.04**		**41.3%**	**4.64**
ARM7	15.5k	686.56	495.12	15.69	425.44	14.07%	36.79	425.44	14.07%	36.44

From Table 1, we see that compared to the deterministic method, which uses the fixed values, when we use statistical models for gate delay and subthreshold leakage current, ISCAS85 benchmarks can achieve on average 29% greater leakage power saving with 99% timing yield and 41% greater power saving with 95% timing yield. The reason is that statistical model has a more flexible optimization space, while the deterministic approach assumes the worst case. For c499 and c1355, which have many critical paths due to their extremely symmetrical circuit structures, the optimization space is limited and therefore the additional power saving contributed by optimization is much smaller, especially with the higher timing yield (99%).

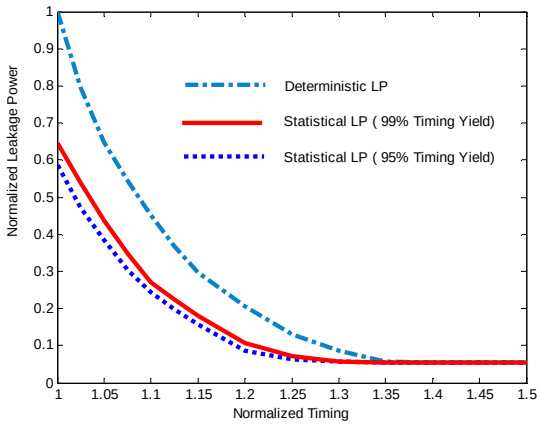

Figure 5. Power-delay curves of deterministic and statistical approaches for C432.

It is also obvious that with a decreased timing yield, higher power saving can be achieved due to the relaxed timing constraints, resulting in larger optimization space.

Figure 5 shows the power-delay curves for C432 for deterministic and statistical approaches. The starting points of the three curves, (1,1), (1,0.65) and (1,0.59), indicate that if we can reduce the leakage power to some 1 unit by deterministic approach, 0.65 unit and 0.59 unit leakage power can be achieved by using statistic approach with 99% and 95% timing yields, respectively. Lower the timing yield, higher is power saving. With a further relaxed T_{max}, all three curves will give more reduction in leakage power because more gates will be assigned high V_{th}.

4.2 Run Time of MILP Algorithm

The run time of ILP is always a big concern since its complexity is exponential in the number of variables and constraints of the problem in the worst case. However, our experimental results show that the real computing time may depend on the circuit structure, logic depth, etc., and may not be exponential. Running on a 2.4GHz AMD Opteron 150 processor with 3GB memory, many CPU run times for solving the ILP problem were less than one second (columns 5, 8 and 11 in Table 1). This is an advantage over other techniques [9] because we achieve 30% more leakage reduction with 99% timing yield but in much less CPU time.

Besides ISCAS'85 benchmark circuits, we also optimized the leakage for an ARM7 IP core, which has 15.5k combinational cells and 2.4k sequential cells implemented in TSMC 90nm CMOS process. The experimental results in the last row of Table 1 show 14% more leakage reduction achieved with 37 seconds run time and partly demonstrate the feasibility of applying our MILP approach to real circuits.

Although today's SOC may have over one million gates, it always has a hierarchical structure. ILP constraints can be generated for submodules at a lower level and the run times will be determined by the number of gates in the individual submodules. Such a technique may not guarantee a global optimization, but still would get a reasonable result within acceptable run time.

5. Conclusion

A mixed integer linear programming formulation to statistically minimize the leakage power in a dual-V_{th} process under process variations is proposed in this paper. The experimental results show that 30% more leakage power reduction can be achieved by using this statistical approach compared with the deterministic approach. In the statistical approach, the impact of process variation on leakage power and circuit performance is simultaneously minimized when a small yield loss is permitted.

Acknowledgment – Authors are grateful for assistance from Analog Devices.

6. References

[1] M. Ketkar and S. S. Sapatnekar, "Standby Power Optimization via Transistor Sizing and Dual Threshold Voltage Assignment," *Proc. ICCAD*, 2002, pp. 375-378.

[2] L. Wei, Z. Chen, M. Johnson and K. Roy, "Design and Optimization of Low Voltage High Performance Dual Threshold CMOS Circuits," *Proc. DAC*, 1998, pp. 489-494.

[3] L. Wei, Z. Chen, K. Roy, Y. Ye and V. De, "Mixed-V_{th} (MVT) CMOS Circuit Design Methodology for Low Power Applications," *Proc. DAC*, 1999, pp. 430-435.

[4] Q. Wang, and S. B. K. Vrudhula, "Static Power Optimization of Deep Submicron CMOS Circuits for Dual V_T Technology," *Proc, ICCAD*, 1998, pp. 490-496.

[5] D. Nguyen, A. Davare, M. Orshansky, D. Chinney, B. Thompson, and K. Keutzer, "Minimization of Dynamic and Static Power Through Joint Assignment of Threshold Voltages and Sizing Optimization," *Proc. ISLPED*, 2003, pp. 158-163.

[6] F. Gao and J. P. Hayes, "Gate sizing and V_t Assignment for Active-Mode Leakage Power Reduction," *Proc. ICCD*, 2004, pp. 258-264.

[7] A. Davoodi and A. Srivastava, "Probabilistic Dual-V_{th} Optimization Under Variability," *Proc. ISLPED*, 2005, pp. 143-147.

[8] A. Srivastava, D. Sylvester, D. Blaauw, "Statistical Optimization of Leakage Power Considering Process Variations Using Dual-V_{th} and Sizing," *Proc. DAC*, 2004, pp. 773-778.

[9] M. Mani, A. Devgan and M. Orshansky, "An Efficient Algorithm for Statistical Minimization of Total Power Under Timing Yield Constraints," *Proc. DAC*, 2005, pp. 309-314.

[10] S. Mukhopadhyay and K. Roy, **"**Modeling and Estimation of Total Leakage Current in Nano-Scaled CMOS Devices Considering the Effect of Parameter Variation," *Proc. ISLPED*, 2003, pp. 172-175

[11] Y. Lu and V. D. Agrawal, "Leakage and Dynamic Glitch Power Minimization Using Integer Linear Programming for V_{th} Assignment and Path Balancing," *PATMOS*, 2005, pp. 217-226.

[12] Y. Lu and V. D. Agrawal, "CMOS Leakage and Glitch Power Minimization for Power-Performance Tradeoff," *Journal of Low Power Electronics*, vol. 2, no. 3, December 2006.

[13] T. Sakurai and A. R. Newton, "Alpha-Power Law MOSFET Model and its Applications to CMOS Inverter Delay and Other Formulas," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, pp. 584-594, February 1990.

[14] R. Rao, A. Srivastava, D. Blaauw and D. Sylvester, "Statistical Analysis of Subthreshold Leakage Current for VLSI Circuits," *IEEE Transactions on VLSI Systems*, vol. 12, no. 2, Feb. 2004, pp. 131-139.

[15] R. Rao, A. Devgan, D. Blaauw and D. Sylvester, "Parametric Yield Estimation Considering Leakage Variability," *Proc. DAC*, 2004, pp. 442-447

[16] M. Mani and M. Orshansky, "A New Statistical Optimization Algorithm for Gate Sizing," *Proc. ICCD*, 2004, pp. 272-277.

[17] E. T. A. F. Jacobs and M. R. C. M. Berkelaar, "Gate Sizing Using a Statistical Delay Model," *Proc. DATE*, 2000, pp. 283-290.

[18] C. E. Clark, "The Greatest of a Finite Set of Random Variables," *J. Operations Research Soc. America*, vol. 9, pp. 145-162, 1961.

[19] F. Hu, "Process-Variation-Resistant Dynamic Power Optimization for VLSI Circuits," PhD Dissertation, Dept. of ECE, Auburn University, Auburn, Alabama, May 2006.

[20] F. Hu and V. D. Agrawal, "Input-Specific Dynamic Power Optimization for VLSI Circuits, *Proc. ISLPED*, 2006, pp. 232-237.

[21] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL: A Modeling Language for Mathematical Programming*. South San Francisco, California: The Scientific Press, 1993.

[22] BPTM: Berkeley Predictive Technology Model. See http://www-device.eecs.berkeley.edu/~ptm/.