

Delay Modeling and Statistical Design of Pipelined Circuit Under Process Variation

Animesh Datta, *Student Member, IEEE*, Swarup Bhunia, *Member, IEEE*,
Saibal Mukhopadhyay, *Student Member, IEEE*, and Kaushik Roy, *Fellow, IEEE*

Abstract—Under inter-die and intra-die parameter variations, the delay of a pipelined circuit follows a statistical distribution. This paper presents analytical models to estimate yield for a pipelined design based on delay distributions of individual pipe stages. Using the proposed models, it is shown that a change in logic depth and an imbalance between stage yields can improve the design yield and the area of a pipeline a circuit. A novel statistical methodology is developed to enhance yield of a pipelined circuit under an area constraint. Based on the concept of area borrowing, the results show that incorporating a proper imbalance among stage areas in a four-stage pipeline improves design yield up to 15.4% for the same area (and reduces area up to 8.4% under a yield constraint) compared with a balanced design.

Index Terms—Gate-level sizing, pipeline design, statistical delay variation, yield enhancement.

I. INTRODUCTION

INCREASING inter-die and intra-die variations in process parameters, such as channel length (L), width (W), oxide thickness (T_{ox}), threshold voltage, etc., result in large variations in the delay of logic circuits [1]. Consequently, estimating circuit performance and designing high-performance circuits with high yield (probability that the design will meet a certain delay target) under parameter variations have emerged as serious design challenges in the sub-100-nm regime [1], [2], [8], [9]. In high-performance design, throughput is primarily improved by pipelining the data and control paths [6]. To predict and maximize the yield of a high-performance design, it is necessary to estimate the overall delay distribution of pipelined circuits by considering variation of individual stage delays. In a synchronous pipelined circuit, throughput is limited by the slowest pipe segment (segment with maximum delay) [6]. Under parameter variations, as the delays of all the stages vary considerably, the slowest stage is not readily identifiable. Thus, a statistical delay model is necessary to predict the delay distribution of a pipeline. There are multiple design

techniques to tradeoff pipeline delay for power or die area using logic synthesis, gate/wire sizing, etc. Statistical analyses of delay and techniques to enhance yield in combinational circuits have been proposed in [2], [3], [10]–[13], and [16]. Unless a worst case design is chosen for a pipeline, which guarantees that the design meets the target delay at the worst process corner, a pipeline design is bound to suffer yield loss in terms of delay failures. However, a worst case design is overly pessimistic in terms of design area/power requirement. Hence, a design methodology, which addresses yield enhancement of pipeline design under statistical delay variation with minimum impact on area/power, is mandatory. Traditionally, pipeline clock frequency has been enhanced by; 1) increasing the number of pipeline stages, which, in turn, reduces the logic depth and hence the delay of each stage [5], and 2) balancing the delay of the pipe stages [5]. However, it has been shown that if intra-die parameter variation is considered, reducing the logic depth increases the variability (defined as the ratio of standard deviation and mean of the design delay) [7]. The effect of balancing the stage delays on the overall delay under parameter variation also needs to be analyzed. Thus, traditional deterministic approaches for maximizing the pipeline throughput need to be reinvestigated to understand their effect on pipeline yield under parameter variations. In this paper, we have described a statistical delay model for a pipelined circuit and proposed a statistical design methodology for enhancing yield of a pipelined circuit considering inter-die and intra-die variations. In particular, this paper presents:

- 1) analytical models to estimate the mean and standard deviation of the overall pipeline delay from correlated individual stage delay distributions;
- 2) analysis of the effect of logic depth, imbalance in the stage delays, and correlation among different stage delays on the yield of a pipeline design;
- 3) statistical design methodology to reduce the area (or power) of a pipeline design under a yield constraint;
- 4) statistical design methodology to enhance yield of a pipelined circuit under an area budget.

Our analysis shows that, under parameter variations, a proper imbalance among stage delays can result in the improvement of the yield (or area) of a pipelined circuit.

The rest of this paper is organized as follows. In Section II, we present a statistical delay and yield model that considers the effect of parameter variations on a pipeline circuit. Section III analyzes the effect of number of pipeline stages, varying logic depth of the stages, and unbalancing stage delays on the overall

Manuscript received February 12, 2005; revised July 6, 2005 and October 10, 2005. This work was supported in part by the Marco Gigascale System Research Center (GSRC), by the Semiconductor Research Corporation (SRC), and by the National Science Foundation (NSF). This work was published in part in “Statistical modeling of pipeline delay and design of pipeline under process variation to enhance yield in sub-100 nm technology,” *ACM/IEEE Proc. of Design, Automation and Test in Europe (DATE)*, vol. 2, 2005, pp. 926–931. This paper was recommended by Associate Editor F. N. Najm.

A. Datta, S. Mukhopadhyay, and K. Roy are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: adatta@purdue.edu; sm@purdue.edu; kaushik@ecn.purdue.edu).

S. Bhunia is with the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: Swarup.Bhunia@case.edu).

Digital Object Identifier 10.1109/TCAD.2006.873886

pipeline design yield. Section IV presents an efficient statistical pipeline design methodology and the corresponding results. Section V concludes this paper.

II. YIELD ESTIMATION OF PIPELINED DESIGN UNDER PARAMETER VARIATION

In this section, we develop analytical models to estimate yield of a pipeline design with respect to a target delay (i.e., operating frequency) from the delay distribution of individual stages.

A. Process Parameter Variation

For deep submicrometer technologies, uncertainties in device and interconnect characteristics (e.g., L , W , T_{ox} , doping concentration) often lead to heterogeneous and nonmonotonic relationships among the process parameters. In general, these variations can be distinguished into two components [1].

- 1) Die-to-die physical variations (inter-die).
- 2) Within-die physical variations (intra-die).

The effect of random and systematic variations of V_{th} in the circuit delay can be taken into account by expressing V_{th} with a set of Gaussian random variables (RV). Considering the systematic components of intra-die parameter variations, the threshold voltage of a MOS device can be modeled as

$$V_{th} = V_{th0} + \Delta V_{th}, \quad \Delta V_{th} = \Delta V_{th_inter} + \Delta V_{th_intra} \quad (1)$$

where V_{th0} is the nominal threshold voltage, and ΔV_{th_inter} and ΔV_{th_intra} represent the shift in the threshold voltage due to inter-die and intra-die variations, respectively. The mean of the ΔV_{th_intra} and ΔV_{th_inter} variables is zero. Their standard deviations depend on the manufacturing process and usually lie within a few percent of the nominal value [8]. The proposed modeling framework for V_{th} variation can be extended to incorporate other process parameter variations in computing stage delay distribution parameters and correlations among them.

B. Variation in Pipeline Delay and Yield

Stage delays as well as individual gate delays are functions of transistor threshold voltages [9]. Using the alpha-power law as in [21], the delay of a combinational logic stage (T_{Comb}) of logic depth L can be expressed using (1) as

$$T_{Comb} = \sum_{i=1}^L \frac{C_{l_i} V_{dd}}{\beta W_i \{V_{dd} - (V_{th0} + \Delta V_{th})\}^\alpha}, \quad 1 < \alpha < 2$$

$$T_{Comb} = \sum_{i=1}^L \frac{C_{l_i} V_{dd}}{\beta W_i (V_{dd} - V_{th0})^\alpha \left(1 - \frac{\Delta V_{th}}{(V_{dd} - V_{th0})}\right)^\alpha}$$

$$T_{Comb} \approx \sum_{i=1}^L \frac{C_{l_i} V_{dd}}{\beta W_i (V_{dd} - V_{th0})^\alpha} \left(1 + \frac{\alpha \Delta V_{th}}{V_{dd} - V_{th0}}\right) \quad (2)$$

where C_{l_i} is the load capacitance of the i th gate and W_i is its width. We ignored higher order terms of ΔV_{th} since α is close to one and $\Delta V_{th} \ll (V_{dd} - V_{th0})$ (e.g., in the 70-nm

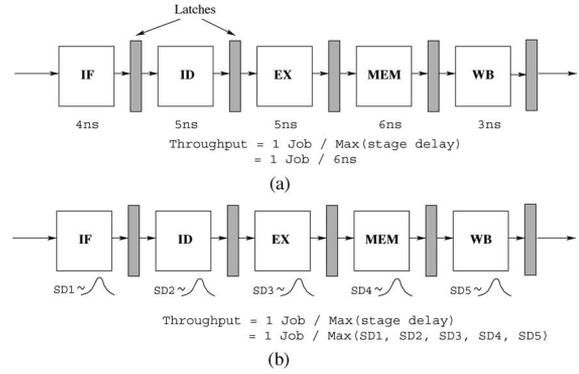


Fig. 1. Simple five-stage pipeline and throughput for (a) static delay model and (b) statistical delay model.

technology node, $\Delta V_{th} \approx 40$ mV and $V_{dd} - V_{th0} \approx 900$ mV). Equation (2) expresses combinational logic delay (T_{Comb}) as a linear combination of a number of Gaussian RVs (V_{th}). Hence, T_{Comb} can also be represented as a Gaussian RV.

It is important to note that the delay of a pipeline stage (SD) consists of the: 1) clock-to-Q delay of the flip-flop ($T_{clk-to-Q}$); 2) propagation delay through the combinational logic (T_{Comb}); and 3) setup time (T_{Setup}) [6], as shown in (3). Mahmoodi *et al.* [19] also show that the two other components of stage delay (T_{Setup} and T_{C-Q}) can also be modeled as Gaussian RVs. Let us consider a pipeline consisting of N stages (Fig. 1). If SD_i denotes the delay of the i th stage, then the overall pipeline delay (T_P) is the maximum of N individual stage delays and is given by [4]

$$T_P = \underbrace{\text{Max}}_{i=1, \dots, N} (SD_i)$$

$$= \underbrace{\text{Max}}_{i=1, \dots, N} (T_{clk-to-Q} + T_{Comb}^i + T_{Setup}^{i+1}). \quad (3)$$

Stage delays can be represented as correlated Gaussian RVs [$SD_i \sim N(\mu_i, \sigma_i)$] [1], [3] (if we neglect spatial and electrical correlations, stage delays can be considered as independent RVs). The overall pipeline delay T_P given in (3) will also be an RV [$T_P \sim N(\mu_T, \sigma_T)$]. The probability (P_D) that the pipelined circuit will meet a specific delay requirement (T_{TARGET}) is given by

$$P_D = \Pr \left\{ \underbrace{\text{Max}}_{i=1, \dots, N} (SD_i) < T_{TARGET} \right\}. \quad (4)$$

We define P_D as a direct measure of the yield of the pipelined circuit. In the next section, we present a semianalytical statistical methodology to estimate the overall pipeline delay distribution (i.e., $T_P \sim (\mu_T, \sigma_T)$).

C. Estimation of Pipeline Delay Distribution

The mean (μ_T) and standard deviation (σ_T) of the pipeline delay (T_P) depend on the mean (μ_i) and standard deviation

(σ_i) of each stage. A lower bound on μ_T is given by (using Jensen's inequality [4], [15])

$$E[T_d] = E \left[\underbrace{\max_{i=1, \dots, N} \text{SD}_i}_{\geq \underbrace{\max_{i=1, \dots, N} \{E[\text{SD}_i]\}} \right]. \quad (5)$$

The previous equation states that the mean of the overall pipeline delay will be larger than the maximum of the mean delay of all the stages. To obtain an exact estimate of μ_T and σ_T , we approximate T_P as

$$\begin{aligned} T_P &= \max\{\text{SD}_1, \text{SD}_2, \dots, \text{SD}_{N-1}, \text{SD}_N\} \\ &= \max\{\text{SD}_1, \text{SD}_2, \dots, \max\{\text{SD}_{N-1}, \text{SD}_N\}\} \\ &= \max\{\text{SD}_1, \text{SD}_2, \dots, \max\{\text{SD}_{N-2}, D_{N-1,N}\}\} \\ &= \max\{\text{SD}_1, \text{SD}_2, \dots, D_{N-2,N}\} = D_{1,N} \end{aligned} \quad (6)$$

where $D_{N-1,N}$ represents the normal approximation to $\max\{\text{SD}_{N-1}, \text{SD}_N\}$. The mean ($\mu_{N-1,N}$) and standard deviation ($\sigma_{N-1,N}$) of $D_{N-1,N}$ can be approximated as [15]

$$\begin{aligned} m_1 &= \mu_N \Phi(\alpha) + \mu_{N-1} \Phi(-\alpha) + a \varphi(\alpha) \\ m_2 &= (\mu_N^2 + \sigma_N^2) \Phi(\alpha) + (\mu_{N-1}^2 + \sigma_{N-1}^2) \Phi(-\alpha) \\ &\quad + (\sigma_N + \sigma_{N-1}) a \varphi(\alpha) \\ \alpha &= (\mu_N - \mu_{N-1}) / a \\ a^2 &= \sigma_N^2 + \sigma_{N-1}^2 - 2\sigma_N \sigma_{N-1} \rho_{N-1,N} \\ \mu_{N-1,N} &= m_1; \quad \sigma_{N-1,N} = \sqrt{m_2 - m_1^2}. \end{aligned} \quad (7)$$

Here, Φ represents the cumulative distribution function (cdf) and φ represents the probability distribution function [pdf, i.e., $\varphi(\alpha) = (2\pi)^{-1/2} \exp(-\alpha^2/2)$] of a standard normal ($\mu = 0$ and $\sigma = 1$) Gaussian RV. The correlation coefficient between SD_{N-1} and SD_N is given by $\rho_{N-1,N}$. Once $D_{N-1,N}$ is obtained, $D_{N-2,N} = \max\{\text{SD}_{N-1}, \text{SD}_N\}$ is computed by iteratively applying (6) and (7). To find $D_{N-2,N}$ using (7), we need to compute the correlation coefficient between $D_{N-1,N}$ and SD_{N-2} as [7]

$$\rho' = \frac{[\sigma_N \rho_{N-2,N} \Phi(\alpha) + \sigma_{N-1} \rho_{N-2,N} \Phi(-\alpha)]}{\sigma_{N-2}}. \quad (8)$$

In (8), ρ' is used to estimate the mean and standard deviation of $D_{N-2,N}$ from SD_{N-1} and $D_{N-1,N}$. The process is repeated $N - 1$ times by taking two variables at a time, and finally μ_T and σ_T of the overall pipeline delay $T_P (= D_{1,N})$ are estimated as in (6).

D. Estimation of Yield

Using (4), the probability of delay failure (P_D) or yield of a pipeline can be estimated as [7]

$$P_D = \Pr \left\{ \bigcap_{i=1, \dots, N} (\text{SD}_i \leq T_{\text{TARGET}}) \right\}. \quad (9)$$

The exact estimation of P_D is possible by assuming the stage delays (SD_i) to be independent Gaussian RVs [7], i.e.,

$$P_D = \prod_{i=1}^N \Phi \left(\frac{T_{\text{TARGET}} - \mu_i}{\sigma_i} \right). \quad (10)$$

If the variables are correlated, such a simplification is not possible. To estimate P_D considering correlated SD_i s, we approximate the overall pipeline delay [$T_P \sim N(\mu_T, \sigma_T)$] as a Gaussian RV [with μ_T and σ_T estimated using (7) and (8)]. Using this assumption, P_D is given by [7]

$$P_D = \Pr \{T_D \leq T_{\text{TARGET}}\} = \Phi \left(\frac{T_{\text{TARGET}} - \mu_T}{\sigma_T} \right). \quad (11)$$

E. Verification of Proposed Delay and Yield Models

To determine P_D , we need to determine the correlation between SD_i and SD_j , for all i, j . There can be two kinds of correlation between SD_i and SD_j , namely: 1) electrical correlation and 2) spatial correlation. It has been demonstrated that T_{Comb} and T_{Setup} of a stage are not electrically correlated with another stage [19]. However, T_{C-Q} of a stage depends on the input data arrival time with respect to clock, i.e., T_{Comb} of the previous stage. As the data arrival time gets closer to the clock edge, T_{C-Q} of a stage experiences increasing correlation with stage delay of the previous stage [19]. Hence, the assumption that stage delays are electrically independent or uncorrelated is valid if the data arrival time is reasonably earlier than the clock edge for all the stages. Another form of correlation between the pipe stages is the spatial dependence due to inter-die and within-die device parameter fluctuations.

We have considered the spatial correlation among stage delays by using a grid approach similar to the one described in [16]. We divide the design area of the die into an array of small squares, each housing a pipe stage and its corresponding latch. The amount of correlation among the neighboring pipe stages is stronger than stages that are far apart in the die. Based on these facts, we model the systematic correlation component of $V_{\text{th_intra}}$ to diminish rapidly as we move away from a pipe stage. Let us consider that the i th pipe stage is spatially correlated to previous and next j stages. Hence, the intra-die threshold voltage variation for a transistor in the i th stage is given by

$$\begin{aligned} \Delta V_{\text{th_intra}} &= \Delta V_{\text{ind_}i} + \Delta V_{i-j,i} + \dots + \Delta V_{i-1,i} \\ &\quad + \Delta V_{i,i+1} + \dots + \Delta V_{i,i+j} \\ \sigma_{\text{th_intra_}i}^2 &= \sigma_{\text{ind_}i}^2 + \sigma_{i-j,i}^2 + \dots + \sigma_{i,j+i}^2 \end{aligned} \quad (12)$$

where $\Delta V_{\text{ind_}i}$ is the independent component of the intra-die process variation due to random dopant fluctuations and $\Delta V_{i,j}$ is the spatially correlated component of the intra-die process variation. $\sigma_{\text{th_intra_}i}^2$ is the variance of the intra-die threshold voltage component of the i th pipe stage. The spatially correlated component $\Delta V_{i,j}$ is common to both i th and j th pipe stages, and its σ value reduces rapidly as the difference between i and j increases. Stage delays as well as individual gate delays

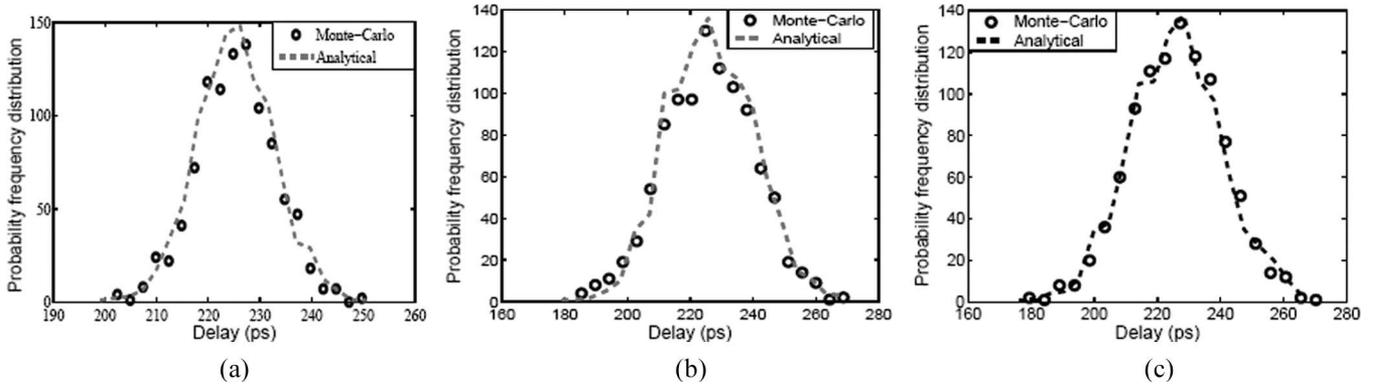


Fig. 2. Delay distribution of 12-stage inverter chain pipeline (with stage logic depth = 10) with (a) only random intra-die variation, (b) only inter-die variation, and (c) inter-die and intra-die variations.

TABLE I
PIPELINE DELAY ESTIMATION (*DENOTES VARIABLE LOGIC DEPTH)

Pipeline configuration	T_{delay} (ps)	Monte-Carlo			Analytical Model		
		μ_T	σ_T	Y (%)	μ_T	σ_T	Y (%)
20 x 6	160	155	2.82	96.4	154	2.68	98.6
15 x 8	200	198	3.27	78.2	198	2.72	77.7
15 x *	215	210	3.67	92.3	210	3.42	93.0
15 x 8 inter	240	200	29.2	88.1	199	28.9	86.7
15 x 8 inter + intra	240	201	28.6	90.3	199	28.1	91.8

are strong functions of transistor threshold voltages [14]. Since the device threshold voltages of neighboring stages are spatially correlated as described in (1) and (12), the corresponding stage delays are also correlated.

Monte Carlo simulation in HSPICE is used to determine the mean and standard deviation of the delay of each stage as well as of the overall design delay. The simulated μ_i and σ_i values for each stage are then fed into the proposed model to determine the distribution of the pipeline delay. It can be observed that the delay distribution predicted by the proposed model closely matches with the HSPICE simulation result of a 12×10 pipeline (i.e., a 12-stage pipeline, each stage having a maximum logic depth of 10) for: 1) only random intra-die variations [i.e., stage delays are independent, Fig. 2(a)]; 2) only inter-die variations [i.e., stage delays are perfectly correlated, Fig. 2(b)]; and 3) both inter-die and intra-die distributions with spatial correlation [i.e., stage delays are partially correlated, Fig. 2(c)]. Close matches have also been observed for several other pipeline configurations (Table I). Interestingly, the presence of inter-die variation significantly increases the variance of the overall pipeline delay. It can be mainly attributed to the fact that inter-die variation changes the delays of all gates in the critical path of a stage in the same direction (increases or decreases).

The major source of error in the proposed modeling method is due to the assumption that the maximum of two Gaussian variables is also a Gaussian one (i.e., $D_{1,N}$ is Gaussian) [7], [8]. This assumption is valid only for two uncorrelated variables [7], [8]. Hence, errors in the estimation of the mean and standard deviation are expected to increase with the increase in the number of stages [Fig. 3(a)] and correlation among stage delays [Fig. 3(b)] [7]. It can be observed that the increase in error in the

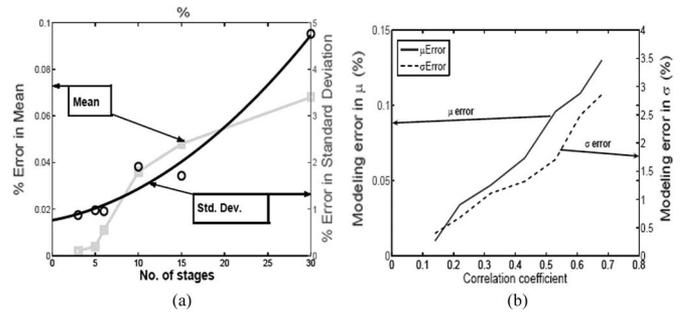


Fig. 3. Trend in pipeline delay modeling error with (a) number of stages and (b) correlation between stages.

standard deviation is more significant in both cases. However, in all cases, the errors in standard deviation and mean are less than 3% and 0.5%, respectively. The modeling error also depends on the ordering of the variables SD_i in (6). It has been shown that the error is minimum if the variables are ordered in increasing (or decreasing) sequence of their means [7]. We have used an ordering in terms of increasing mean in our estimation to minimize the error.

F. Estimation of Design Space

Using the proposed models, we can estimate the design space for the mean and standard deviation of different stages of a pipeline that will satisfy a yield constraint. Assuming the overall delay of the pipeline to be Gaussian and using (10), the upper bound of mean of each stage delay (μ_i) is given by

$$\mu_i \leq \mu_T \leq T_{TARGET} - \sigma_T \Phi^{-1}(P_D). \tag{13}$$

However, this bound does not provide any estimate of the design space in terms of σ_i of each stage. A relaxed bound on the μ_i and σ_i for the i th stage can be obtained as follows by assuming that any other j th stage meets the yield requirement with probability 1 (i.e., $\Phi(x) = 1$), i.e.,

$$\mu_i + \sigma_i \Phi^{-1}(P_D) \leq T_{TARGET}. \tag{14}$$

A relaxed bound obtained in this way is shown in Fig. 4. Equation (14) states that if the mean and/or standard deviation of any stage fall outside this bound, no pipeline design with that

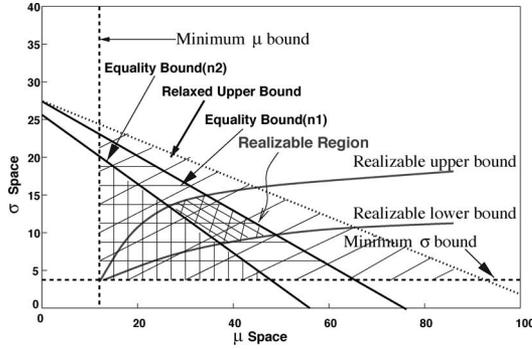


Fig. 4. Range of permissible μ and σ for each stage to meet Y_{TARGET} .

stage can ever meet the target delay and yield. A more stringent bound is obtained by assuming uncorrelated and equal stage delays and is given by

$$\left[\Phi \left(\frac{T_{\text{TARGET}} - \mu_i}{\sigma_i} \right) \right]^N \geq P_D$$

$$\Rightarrow \mu_i + \sigma_i \Phi^{-1} \left(P_D^{1/N} \right) \leq T_{\text{TARGET}} \quad (15)$$

where N is the number of pipeline stages. Depending on N , it gives a set of values for μ_i and σ_i that can meet the target yield (Fig. 4 shows two such equality bounds for $N = n1, n2$ with $n1 < n2$). It should be noted that there is a minimum bound on μ_i and σ_i (say, μ_{\min} and σ_{\min}), which depends on the minimum allowable logic depth and process specification. Moreover, the μ_i and σ_i of a combinational circuit are related parameters, and the relation determines the realizable design space for μ_i and σ_i . Considering each stage critical path as a chain of N_L inverters, a relation between μ_i and σ_i is given by

$$\mu_i = N_L \mu_{\min}; \quad \sigma_i = \sqrt{N_L} \sigma_{\min} \Rightarrow \mu_i = \left(\frac{\mu_{\min}}{\sigma_{\min}^2} \right) \sigma_i^2 \quad (16)$$

where μ_{\min} and σ_{\min} are the mean and standard deviations of a minimum-sized inverter, respectively. However, this relation (16) is independent of the number of pipeline stages. Similarly, for a maximum-sized inverter having parameters μ_{\max} and σ_{\max} , there will be another bound on realizable μ_i and σ_i (Fig. 4). Hence, for equal stage delays, a realizable region bounded by two curves is obtained by (16) with maximum- and minimum-sized gates.

III. KEY OBSERVATIONS ON PIPELINE DESIGN

A. Tradeoff Between Number of Pipe Stages and Logic Depth

In this section, we analyze the effect of logic depth and number of stages on the variability of a pipeline delay. A reduction in the logic depth, which increases the number of stages in a pipeline, improves the operating frequency [4]; however, the operation latency increases. Correlation in the delays of different gates (due to inter-die variation and spatial correlation) reduces the cancellation effect. Hence, the variability becomes a weaker function of logic depth [Fig. 5(a)]. On the other hand, increasing the number of elements in the max function reduces

its variability (Fig. 5(b), evaluated using an inverter chain pipeline with constant logic depth in each stage) [7]. It can be further observed that as the stage delays become more and more correlated, the sensitivity of the variability to the number of stages reduces [Fig. 5(b)]. In our experiments, we have used a range of correlation coefficients to observe the effect of correlation in the overall pipeline delay. To understand the effect of logic depth (N_L), number of stages (N), and relative strength of inter-die and intra-die variations, we have estimated the variability of a 120-long inverter chain pipeline (BPTM 70-nm technology [18]) with different configurations (i.e., with different N_L and N such that $N_L \times N = 120$). In each case, the delay distribution obtained by the analytical model closely matches the Monte Carlo simulation results from HSPICE. With only intra-die variation, the effect of logic depth prevails over the effect of the max function. Consequently, increasing the number of stages (i.e., reducing the logic depth of each stage) increases the variability [Fig. 5(c)]. On the other hand, as the strength of inter-die variation increases (i.e., stage delays become more correlated), the σ/μ ratio of each stage becomes a weaker function of its logic depth. Hence, the impact of max function prevails and the variability of the overall pipeline delay reduces with an increase in the number of stages [Fig. 5(c)].

B. Perfectly Balanced Versus Unbalanced Pipeline Design

Traditionally, the pipeline stages are designed for equal delay to maximize the throughput [4]. However, incorporating the imbalance among pipeline stage delays can have a positive impact on the overall yield under process variation. This is because of the fact that a balanced pipeline has a higher number of critical paths than an unbalanced design, which can adversely affect the overall design yield [5]. Imbalance can be incorporated in a balanced pipeline by using transistor sizing and/or logic restructuring. However, the impact of imbalance on the overall pipeline delay needs to be estimated.

We have performed experiments with a three-stage ALU decoder pipeline structure (Fig. 6) to understand the effects of imbalance on the pipeline delay. The combinational logic of each stage is first tuned for minimum area (using algorithm presented in [20]) for a specific overall target yield of 80% (with a pipeline delay of 179 ps). First, we keep the target yield constant for all three stages (the yield of a stage is defined as the probability that the combinational logic of that stage meets its target delay).

For simplicity and ease of explanation, we neglect any correlation among stage delays [correlated stage delays are considered in the final result (Tables IV and V)]. The yield target for each stage was kept at $(0.80)^{1/3} = 0.9283$ [using (15)]. In the next step, we have introduced imbalance among areas (or yields) of the three stages (by transistor sizing CASE-B, Fig. 6) in such a way that the total design yield (or area) remains constant. Resizing the transistors reduces the effective number of critical paths and improves the overall pipeline yield. To understand the reason behind this yield improvement, let us consider the area versus mean delay [normalized with respect to the target delay (T_D)] curves for each stage (Fig. 7). The stages are initially designed for equal delay, indicated by line

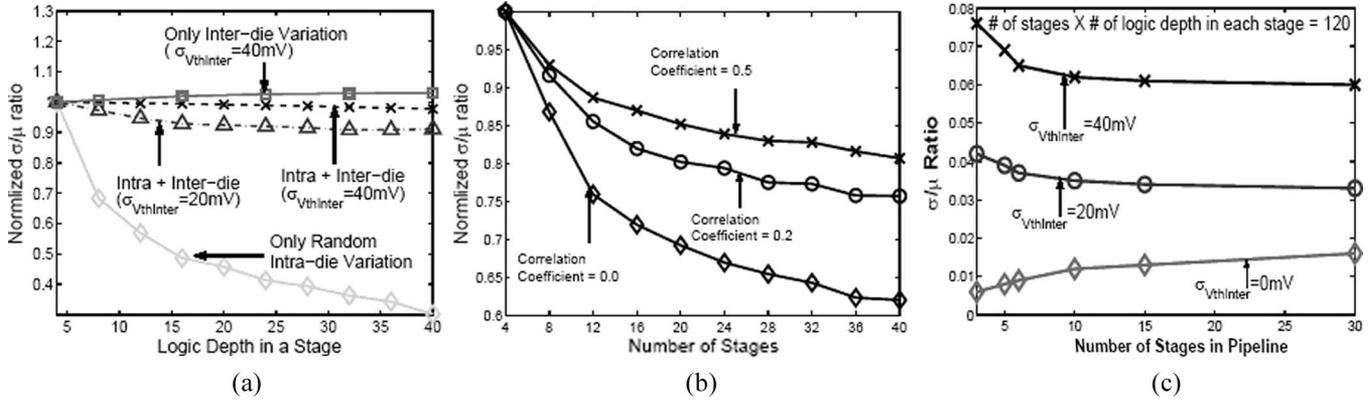


Fig. 5. Variability of (a) stage delay with logic depth, (b) pipeline design delay with number of stages, and (c) pipeline design delay with simultaneous change of logic depths and number of stages.

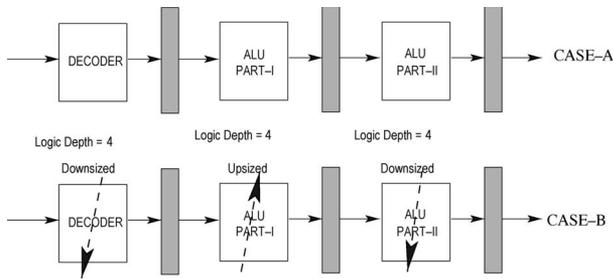


Fig. 6. Three-stage ALU decoder pipeline with different stage delays.

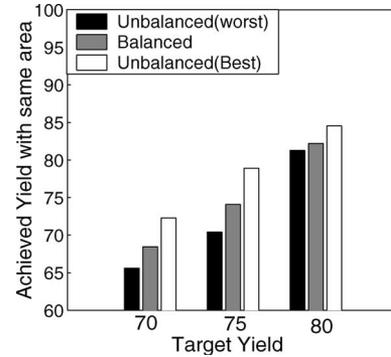


Fig. 8. Effect of unbalancing on pipeline yield.

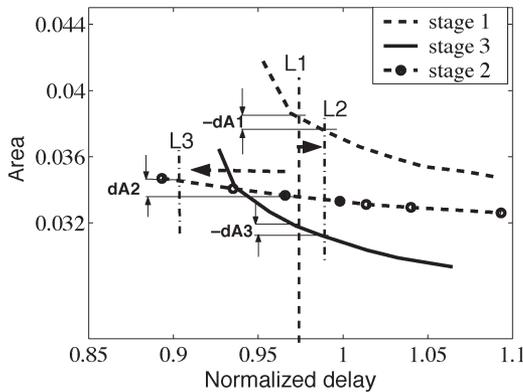


Fig. 7. Area versus delay curves of three-stage ALU decoder pipeline.

L1 in Fig. 7. This results in equal yield of $Y_0 = 0.928$ for each stage (overall pipeline yield = $Y_0^3 = 0.8$). The total area for this design is the sum of the stage areas ($A_1 + A_2 + A_3$). Now, we introduce imbalance in stage areas (or yields) by reducing the area of stages 1 and 3 (by dA_1 and dA_3), increasing their mean delays to line L2. Increasing the delays of stages 1 and 3 reduces their yield by a small amount (in this example, Y_1 becomes 91.5% and Y_3 becomes 91.8%; both are less than Y_0). However, the extra area ($dA_1 + dA_3$) can be added to stage 2, thereby reducing its delay (as indicated by line L3). Reduction of the delay of stage 2 improves its yield significantly (in this example, Y_2 increases from 92.8% to 98.8%). In this particular case as $(Y_1 \times Y_2 \times Y_3) > Y_0^3$ (i.e., here $0.988 \times 0.915 \times 0.918 = 0.83 > 0.80$), the overall pipeline

yield improves. A consistent trend of improvement for different target yields is shown for the three-stage pipeline circuit in Fig. 8. In this technique, we effectively borrow area (or yield) from stages 1 and 3 for stage 2 to improve the overall design yield (or area). We refer to this concept as area (yield) borrowing. We propose a heuristic for area allocation among the stages based on the position of the stages in their area versus delay curve. The concept of yield borrowing can be formalized as

$$\text{Calculate slope of area versus delay curve : } R_i = \left| \frac{\partial A}{\partial D} \right|$$

for each stage

if $R_i > 1$

reduction in large area

results in small increase in delay

else

reduction in small area

results in large improvement in delay

end if

end for.

$$(17)$$

It should be noted that introducing excess imbalance in stage delays may result in diminishing returns when pipeline

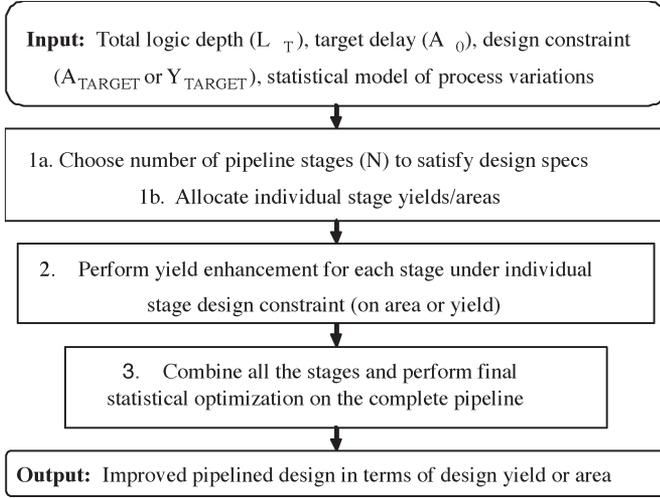


Fig. 9. Hierarchical design flow for complete pipeline design.

performance is governed by the mean delay of the slowest stage (Fig. 8). Hence, it is necessary to appropriately introduce imbalance among the stages.

IV. PIPELINE DESIGN UNDER PARAMETER VARIATION

We propose a hierarchical design flow for complete pipeline design that is efficient in terms of design effort and meets the required design specifications. It should be noted that a pipelined design consisting of complex functional blocks can be very large. This implies that global optimization of a pipeline design at the logic level under parameter variations can have large (often impractical) time and space complexity. Hence, in the proposed design flow of the pipeline, we perform design optimization of one stage at a time.

Our proposed pipeline design method comprises three principal steps (Fig. 9). First, an appropriate number of stages for the pipeline (N) is estimated for a given design target of yield or area. The number of pipeline stages (N) is typically specified by architectural constraints and performance (throughput) requirements of the design from system-level analysis. Once N is determined, we perform an initial yield allocation and area budgeting of individual stages. In the next step, we independently optimize each functional module for a design objective of area (yield) under constraint on the allocated yield (area). To perform statistical timing analysis of individual stages, we have used a statistical static timing analysis (SSTA) method as proposed in [16]. Given the gate-level netlist of a stage, gate sizes, and information on process parameter variations, the SSTA tool efficiently computes the mean and standard deviation of the stage delay within a small error margin (less than 1%) considering both inter-die and intra-die variations [16]. Individual stage yields can be obtained using (11) for a given target delay from the delay distribution parameters. Finally, in the last step, individually optimized stages are glued together to form a pipeline and a global optimization is performed using the concept of “unbalancing” stage yields (area), as described in Section III-B.

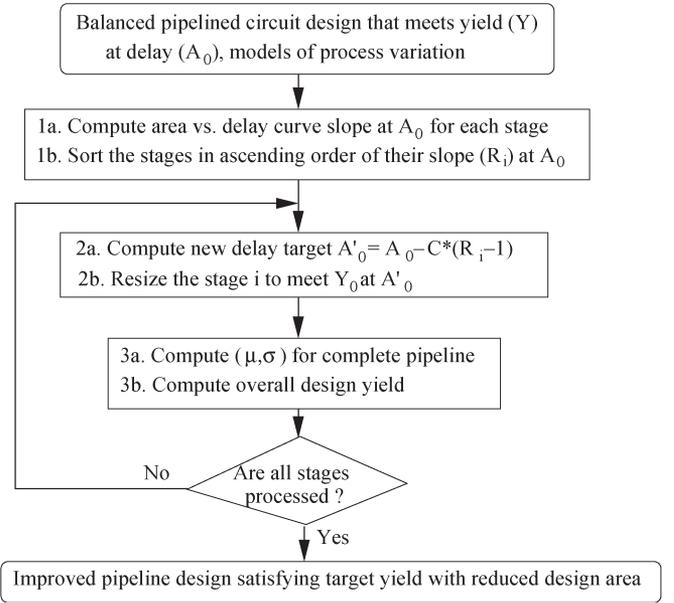


Fig. 10. Statistical design of pipeline to reduce area under yield constraint.

A. Statistical Design of Individual Stages

Individual stages of the pipeline can now be designed for a given objective (i.e., minimize area for a Y_{TARGET} or maximize yield for an A_{TARGET}). Based on the optimization objectives, we can have the following two design problems.

The problem of minimizing stage area under a yield constraint can be stated as

$$\begin{aligned} \text{Minimize Area} &= \sum_{i=1}^n \alpha_i x_i / * n = \text{No. of gates} \\ x_i &= \text{size factor of } i\text{th gate}^* / \\ \text{Subject to } \Phi \left(\frac{A_0 - \mu}{\sigma} \right) &\geq Y_0 / * Y_0 = \text{target stage yield}^* / \\ L_i \leq x_i \leq U_i, / * L_i, U_i &\text{ are min and max size}^* / . \end{aligned} \quad (18)$$

To solve the above problem, we make use of a gate-level sizing algorithm that minimizes the total area under a delay constraint, as given in [20]. This gate-sizing method is an iterative low-complexity algorithm based on Lagrangian relaxation (LR). In [20], a solution for convex gate-level sizing problem is proposed to: 1) minimize the active design area under a target delay constraint and 2) minimize circuit delay under an area constraint. We adopt method (1) here to improve the design area of each stage logic under a yield constraint with the help of an SSTA tool based on [16].

On the other hand, the problem of maximizing the pipeline stage yield under an area budget can be expressed as

$$\begin{aligned} \text{Maximize} \quad & Y_i \text{ for the } i\text{th stage at } A_0 \\ \text{Subject to} \quad & A_{\text{Comb}_i} + A_{\text{Seq}_i} \leq A_{\text{TARGET}_i}. \end{aligned} \quad (19)$$

We adopt method (2) as proposed in [20] to solve the problem proposed in (19). This sizing algorithm for minimizing the

TABLE II
AREA REDUCTION FOR FOUR-STAGE PIPELINE WITH $Y_T = 80\%$

Stage-logic	Individually Optimized		Proposed Algorithm	
	Area (%)	Yield (%)	Area (%)	Yield (%)
c3540	50	94	45	90.5
c2670	20.6	95	21.2	99.1
c1908	23.3	95	19.1	90.5
c432	6.1	94.5	6.3	99.2
Complete Design	100	80.3	91.6	80.6

TABLE III
SIX-STAGE PIPELINE WITH CORRELATED STAGE DELAYS AT $Y_T = 87\%$

Stage-logic	Individually Optimized		Proposed Algorithm	
	Area (%)	Yield (%)	Area (%)	Yield (%)
c3540	36.19	96.57	31.75	91.01
c2670	17.72	96.13	18.13	99.2
c1908	16.53	96.68	14.08	91.68
c432	4.95	96.76	5.31	99.51
c499	13.95	97.15	14.44	99.21
c880	10.73	96.45	10.73	99.57
Complete Design	100	87.79	94.12	87.63

maximum mean delay of a circuit can be used to maximize yield under statistical delay distribution.

B. Optimization of Complete Pipelined Design

In this section, we propose two statistical design methods to improve a balanced pipeline design.

1) *Statistical Design Procedure to Reduce Area of Pipelined Circuit for Target Yield:* The optimization problem of minimizing the area of a pipeline design under a given yield constraint is presented in (18). We present a design procedure (Fig. 10) to solve the above problem by efficiently employing the principle of divide-and-conquer, sizing one stage at a time such that the target yield for the complete pipeline is satisfied while the total area is minimized. From the initial balanced design of each stage (with minimum area as described in Section IV-A), we first compute the individual area versus delay curve slope at the target delay (A_0) for each stage. We then sort the stages in ascending order of their R_i values (step 1, Fig. 10). According to their R_i values and current delay distribution parameters μ and σ , each stage is resized for a modified target delay (steps 2a)–b), Fig. 10). After sizing all the stages, an SSTA on the complete pipeline is performed, and employing the proposed pipeline delay model, the overall design delay (μ, σ) is determined [step 3a)]. Since only one stage is being sized at a time, we can perform SSTA of one stage at a time to estimate the overall mean and standard deviation. This incremental timing analysis improves the computational efficiency of our method.

The procedure is applied on two example pipelined circuits (designed with several ISCAS85 benchmark circuits as the stage logic and edge-triggered D flip-flops as the sequential elements), and the results are compared with the case when the stages were individually sized for equal yield. It is worth noting that improvement in the total area of the pipeline strongly depends on the ordering in which the stages are chosen for sizing. In our design procedure, the ordering is based on the position of

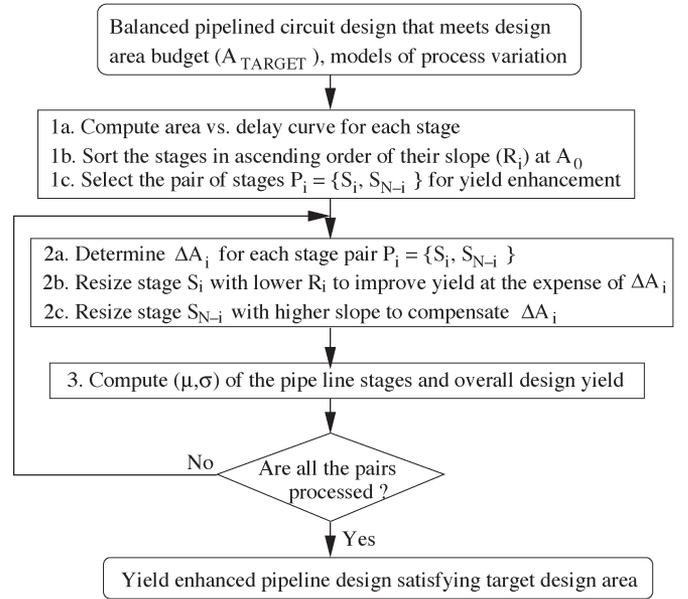


Fig. 11. Statistical design of pipeline to enhance yield under an area budget.

each stage in their area/delay curve as explained in Section III (17). This minimizes the total area to meet the target pipeline yield in the final statistical optimization step since stages with lower R_i are sized before stages with higher R_i . When applied our statistical pipeline design optimization, 8.4% of the area saving is obtained over a balanced four-stage pipelined design having the same yield (Table II) considering independent stage delays. In a six-stage pipeline design, 5.9% of the area saving for a target yield of 87% is obtained (Table III). These results show that a yield-aware statistical design step for the complete pipeline can significantly improve the design area.

2) *Yield Enhancement of Pipeline Design Under Area Budget:* In this section, we solve the yield optimization problem under an area constraint as presented in (19). The principal idea in this design solution is to judiciously allocate yield among the pipeline stages while maintaining the overall design area. We use the concept of yield borrowing (Section III-B) to improve the overall yield of the pipeline by selectively increasing area of some stages, borrowed from other stages, keeping the total design area unchanged. As in the previous method, we first obtain their individual area versus delay curve (step 1a), Fig. 11). Now, based on the position of each stage in their area versus delay curve, we rank the stages in ascending order of their slopes (e.g., for i th stage) (step 1b), Fig. 11). Next, we form $\lfloor N/2 \rfloor$ stage pairs by grouping each i th stage with the $(N - i)$ th stage from the sorted list of stages [step 1c)]. For example, if five stages of a pipeline have their area versus delay curve slopes at A_0 as $R_3 > R_5 > R_2 > R_1 > R_4$, then according to the above rule we choose (R_3, R_4) and (R_5, R_1) as the two pairs of stages and we do not change the design of second stage. Now for each pair, we enhance the yield of the stage with smaller slope (R_i) at the expense of a certain area increase (ΔA_i) using the transistor sizing algorithm.

The area overhead (ΔA_i) thus incurred is compensated by the other stage in the pair (with higher slope). After resizing each pair, we compute the overall design delay and yield

TABLE IV
 YIELD ENHANCEMENT OF A FOUR-STAGE PIPELINE

Delay		$Y_{balanced}(\%)$		$Y_{improvement}(\%)$	
T_d (ps)	FO_4	Uncorrel.	Correl.	Uncorrel.	Correl.
430	10.97	74.61	81.06	15.43	11.05
440	11.22	80.54	84.63	12.66	8.54
450	11.47	83.70	87.52	10.62	6.84
460	11.73	84.98	89.43	9.96	5.92
470	11.99	86.72	90.45	7.15	4.51
480	12.24	88.28	92.36	6.6	3.42
490	12.50	89.89	92.89	5.01	2.02
500	12.75	93.22	95.12	2.65	1.50
510	13.00	95.14	96.53	2.05	0.82

 TABLE V
 YIELD ENHANCEMENT OF A SIX-STAGE PIPELINE

Delay		$Y_{balanced}(\%)$		$Y_{improvement}(\%)$	
T_d (ps)	FO_4	Uncorrel.	Correl.	Uncorrel.	Correl.
430	10.97	58.13	70.39	24.21	8.46
440	11.22	66.61	77.25	16.07	7.11
450	11.47	75.36	83.83	14.13	6.67
460	11.73	78.06	85.32	10.68	5.38
470	11.99	81.38	87.79	9.33	4.52
480	12.24	84.52	90.27	8.08	3.47
490	12.50	85.68	90.42	5.75	2.44
500	12.75	89.55	92.78	3.54	1.77
510	13.00	92.69	95.27	2.79	1.12

parameters using the models proposed in Section II (step 3). It is worth noting that for a particular stage pair ($\{S_i, S_{N-i}\}$), the maximum yield that can be obtained by the area borrowing concept depends on the selection of the exact area ΔA_i to be traded between them [step 2(a)]. We use an iterative solution to obtain the best choice of ΔA_i by incrementing it small steps.

We present yield improvement results with the proposed design methodology for the same four-stage and six-stage pipeline designs described in Section IV-B. However, in this experiment, we have used the delay of an inverter with a fanout of four (as proposed in [17]) as a technology-independent unit of delay for our example pipeline designs. An inverter with a fanout of four similar inverters is denoted by a unit FO_4 delay in Tables IV and V. We varied the operating frequency (FO_4 stage delay also varies) to evaluate the effectiveness of the proposed unbalancing techniques over a range of pipelined circuit speeds. We have presented yield improvement results considering both independent and correlated stage delays. The yield enhancement results of a four-stage pipeline are shown in Table IV, where the third and fourth columns represent uncorrelated and correlated yields [corresponding to the yield model presented in (15) and (11)] for a four-stage balanced pipeline, where individual stages are optimized for equal stage yield under the given area budget. The fifth and sixth columns show the obtained yield improvements. The results show up to 15% (11% with correlated delays) yield improvement. The improvement gradually reduces with higher yield target. We obtain about 5% (2% with correlated delays) improvement even when the initial yield is 90% (Table IV). Table V presents a similar set of results for the six-stage pipelined design.

In both cases, we observe a large yield improvement (above 10%) in the high frequency (at a smaller delay) design (for

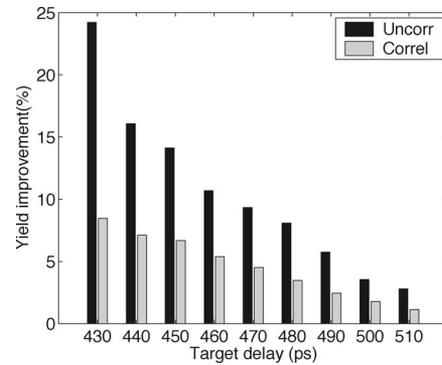


Fig. 12. Yield improvement under area constraint in six-stage pipeline.

$T_d < 450$ ps). Fig. 12 shows the trend of yield improvement at different delay targets for a six-stage pipeline design.

V. CONCLUSION

In this paper, we have investigated pipeline delay distribution under inter-die and intra-die V_{th} variations. Analytical models for estimating yield of a pipelined circuit are presented. We have observed the impact of logic depth and imbalance among stage yields on the variability of the pipeline delay and yield of the pipelined circuit for a target delay. A statistical design framework to improve the overall pipeline design, exploiting the nature of area versus delay curves of individual stages, is developed. Our results show that the traditional method of optimizing individual stages does not necessarily produce the best pipeline design in terms of yield and area. A global design enhancement step that considers the area versus delay curves of the individual stages of a pipeline is effective in improving the design yield/area in the presence of parameter variations.

REFERENCES

- [1] 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," *IEEE J. Solid-State Circuits*, vol. 37, no. 2, pp. 183–190, Feb. 2002.
- [2] E. T. A. F. Jacobs and M. R. C. M. Berkelaar, "Gate sizing using a statistical delay model," in *Proc. DATE*, 2000, pp. 283–290.
- [3] M. C.-T. Chao, L.-C. Wang, K.-T. Cheng, and S. Kundu, "Static statistical timing analysis for latch-based pipelined designs," in *Proc. ICCAD*, 2004, pp. 468–472.
- [4] A. M. Ross, "Useful bounds on the expected maximum of correlated normal variables," *ISE Working Paper 03W-004*, Aug. 2003.
- [5] D. D. Gajski, N. Dutt, A. Wu, and S. Lin, *High-Level Synthesis: Introduction to Chip and System Design*. Norwell, MA: Kluwer, 1992.
- [6] J. L. Hennessy, D. A. Patterson, and D. Goldberg, *Computer Architecture: A Quantitative Approach*, 3rd ed., M. Kaufmann, Ed. San Mateo, CA: Morgan Kaufmann, May 2002.
- [7] A. Papoulis and S. U. Pillai, "Probability," in *Random Variables and Stochastic Processes*. New York: McGraw-Hill, Dec. 2001.
- [8] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. DAC*, 2003, pp. 338–342.
- [9] S. R. Nassif, "Within-chip variability analysis," in *IEDM Tech. Dig.*, 1998, pp. 283–286.
- [10] S. G. Duvall, "Statistical circuit modeling and optimization," in *Proc. 5th Int. Workshop Stat. Metrol.*, Jun. 2000, pp. 56–63.
- [11] M. R. C. M. Berkelaar, "Statistical delay calculation, a linear time method," in *Proc. TAU*, 1997, pp. 15–24.
- [12] X. Bai, C. Visweswariah, P. N. Strenski, and D. J. Hathaway, "Uncertainty-aware circuit optimization," in *Proc. DAC*, 2002, pp. 58–63.

- [13] D. Arnst, N. S. Kim, S. Das, S. Pant, R. Rao, P. Toan, C. Ziesler, D. Blaauw, T. Austin, K. Flautner, and T. Mudge, "Razor: A low-power pipeline based on circuit-level timing speculation," in *Proc. MICRO-36*, 2003, pp. 7–18.
- [14] P. Fox, F. Wernicke, and R. Narayanasamy, "Statistical analysis of propagation delay in digital integrated circuits," in *Proc. IEEE Int. Solid-State Circuits Conf.: Dig. Tech. Papers*, Feb. 1972, vol. XV, pp. 66–67.
- [15] C. E. Clark, "The greatest of a finite set of random variables," *Oper. Res.*, vol. 9, no. 2, pp. 145–162, Mar./Apr. 1961.
- [16] H. Chang and S. S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single pert-like traversal," in *Proc. ICCAD*, 2003, pp. 621–625.
- [17] M. Harowitz. (2000). *VLSI Scaling for Architects*. [Online]. Available: <http://mos.satnford.edu/papers/VLSIScaling.pdf>
- [18] University of California, *Predictive Technology Model*, (2001). [Online]. Available: <http://www-device.eecs.berkeley.edu/~ptm>
- [19] H. Mahmoodi, S. Mukhopadhyay, and K. Roy, "Estimation of delay variations due to random-dopant fluctuations in nano-scaled CMOS circuits," in *Proc. CICC*, 2004, pp. 17–20.
- [20] C. Chen, C. N. Chu, and D. F. Wong, "Fast and exact simultaneous gate and wire sizing by Lagrangian relaxation," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 18, no. 7, pp. 1014–1025, Jul. 1999.
- [21] T. Sakurai and R. Newton, "Alpha-power law MOSFET model and its application to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584–593, Apr. 1990.



Animesh Datta (S'03) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology (IIT), Kharagpur, India, in 2001, and is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN.

In 2001, he joined the Advanced VLSI Design Laboratory at IIT, working on analog and mixed-signal circuit design. His research interests include yield-aware system design in scaled technologies, speed-binning-aware design optimization, and power-aware system architecture. He has re-

ceived the 2006 IEEE Circuits and Systems Society VLSI Transaction Best Paper Award.



Swarup Bhunia (S'00–M'05) received the B.S. degree (Hons.) in computer science and engineering from Jadavpur University, Calcutta, India, in 1995, the M.S. degree in computer science from the Indian Institute of Technology (IIT), Kharagpur, India, in 1997, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in August 2005.

He has worked in the electronic design automation industry on register transfer level synthesis and verification for about three years. He is currently an

Assistant Professor of electrical engineering and computer science at Case Western Reserve University, Cleveland, OH. He has more than 40 publications in refereed journals and conferences. His research interests include design methodologies for high-performance low-power testable very large scale integration systems, defect-based testing, noise analysis, and noise-aware design.

Dr. Bhunia received the Best Paper Awards at the 2003 Latin American Test Workshop and the 2004 International Conference on Computer Design.



Saibal Mukhopadhyay (S'99) received the B.E. degree in electronics and telecommunication electrical engineering from Jadavpur University, Calcutta, India, in 2000, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in 2006.

He has worked as an Intern at the High Performance Circuit Design Department, IBM T. J. Watson Research Lab, Yorktown Heights, NY, in the summers of 2003 and 2004. His research interests include the analysis and design of low-power and robust circuits using nanoscaled CMOS and circuit design using double gate transistors.

Mr. Mukhopadhyay received the IBM Ph.D. Fellowship Award, for 2004 and 2005, and the Semiconductor Research Corporation Technical Excellence Award, in 2005. He received the Best Paper Award at the 2003 IEEE Nano and 2004 International Conference on Computer Design.



Kaushik Roy (S'83–M'90–SM'95–F'02) received the B.Tech. degree in electronics and electrical communications engineering from the Indian Institute of Technology (IIT), Kharagpur, India, in 1983, and Ph.D. degree from the Electrical and Computer Engineering Department, University of Illinois, Urbana-Champaign, in 1990.

He was with the Semiconductor Process and Design Center, Texas Instruments, Dallas, TX, where he worked on field-programmable gate array architecture development and low-power circuit design.

In 1993, he joined the Electrical and Computer Engineering Faculty, Purdue University, West Lafayette, IN, where he is currently a Professor and the Roscoe H. George Professor of electrical and computer engineering. His research interests include very large scale integration (VLSI) design/computer-aided design for nanoscale silicon and nonsilicon technologies, low-power electronics for portable computing and wireless communications, VLSI testing and verification, and reconfigurable computing. He has published more than 300 papers in refereed journals and conferences, holds eight patents, and is the coauthor of two books on low-power CMOS VLSI design. He is a Purdue University Faculty Scholar. He is the Chief Technical Advisor of Zenasis Inc., and a Research Visionary Board Member of Motorola Labs (2002).

Dr. Roy has been on the Editorial Board of *IEEE Design and Test*, *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS*, and *IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS*. He was a Guest Editor for Special Issue on Low-Power VLSI in the *IEEE Design and Test* (1994) and *IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS* (June 2000), and *IEE Proceedings Computers and Digital Techniques* (July 2002). He received the National Science Foundation Career Development Award, in 1995, the IBM Faculty Partnership Award, the ATT/Lucent Foundation Award, the 2005 Semiconductor Research Corporation (SRC) Technical Excellence Award, the SRC Inventors Award, and the Best Paper Awards at the 1997 International Test Conference, IEEE 2000 International Symposium on Quality of IC Design, 2003 IEEE Latin American Test Workshop, 2003 IEEE Nano, 2004 IEEE International Conference on Computer Design, as well as the 2005 IEEE Circuits and System Society Outstanding Young Author Award.