LETTER   *Special Section on VLSI Design and CAD Algorithms*

Increase in Delay Uncertainty by Performance Optimization

Masanori HASHIMOTO$^{†a)}$ *and* **Hidetoshi ONODERA**†, *Regular Members*

SUMMARY   This paper discusses a statistical effect of performance optimization to uncertainty in circuit delay. Performance optimization has an effect of balancing the delay of each path in a circuit, i.e. the delay times of long paths are shortened and the delay times of short paths are lengthened. In these path-balanced circuits, the uncertainty in circuit delay, which is caused by delay calculation error, manufacturing variability, fluctuation of operating condition, etc., becomes worse by a statistical characteristic of circuit delay. Thus, a highly-optimized circuit may not satisfy delay constraints. In this paper, we demonstrate some examples that uncertainty in circuit delay is increased by path-balancing, and we then raise a problem that performance optimization increases statistically-distributed circuit delay.
key words: performance optimization, delay increase, statistical timing analysis, delay uncertainty, transistor sizing

1.   Introduction

In VLSI design, many techniques for reducing circuit delay and power dissipation are utilized at every design phase in order to satisfy given design requirements. Delay optimization methods basically detect the longest path and optimize the circuit for reducing the longest path delay. Some power reduction methods slow down blocks/cells whose timing constraints are not tight. Performance optimization therefore can be regarded as an operation that shortens long paths and lengthens short paths in a circuit. The delay times of many paths in a circuit are equalized by performance optimization, which is called path-balance.

There are several sources that cause uncertainties in circuit delay time, such as error in delay calculation, manufacturing variability, and fluctuation of operating conditions. The delay uncertainties are classified into two groups; global change and random change. The global change applies to all gates and wires similarly in a certain region, while the random change fluctuates all gates and wires independently according to a certain statistical distribution. This paper focuses on the random change. In recent VDSM technologies, intra-chip variability in transistor characteristics, which corresponds to the random change, is comparable to inter-chip variability [1], [2]. The estimation errors,

such as gate delay model and interconnect capacitance extraction, can be roughly treated as random fluctuation. Therefore the random change is not negligible and should be considered properly.

In this paper, we examine the effect of path-balancing to uncertainty in circuit delay. Due to the statistical characteristic of circuit delay, the average value of statistically-distributed circuit delay becomes large when the number of long paths increases. So far, this increase of statistically-distributed circuit delay caused by path-balancing has not been well discussed. We raise a notice that performance optimization increases statistically-distributed circuit delay due to the statistical effect of path-balancing.

2.   Statistical Effect of Performance Optimization on Circuit Delay Time

The circuit delay, which is the maximum path delay time in a circuit, $D_{circuit}$ is represented as follows.

$$D_{circuit} = \max_i D_i \quad (i = 1, 2, ..., n), \tag{1}$$

where D_i is the path delay time of the i-th path, and n is the number of the paths in the circuit.

Let us show a simple example of the statistical effect caused by the max operation. Suppose that D_i is distributed independently according to a normal distribution N(6,1). We examine the distribution of $D_{circuit}$ under several values of n. Figure 1 shows the distribution of $D_{circuit}$. When n increases, the average of $D_{circuit}$ becomes large and the standard deviation of $D_{circuit}$ becomes small.
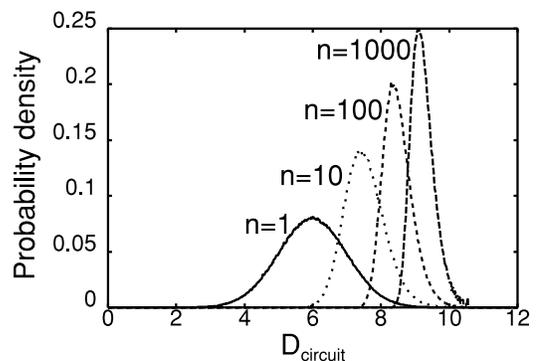
Fig. 1   Effect of max operation (n is varied).

We show another example. We fix n to 100, and vary the standard deviation σ of D_i. Figure 2 shows the distribution of $D_{circuit}$. We can see that the average and the standard deviation of $D_{circuit}$ become large, when the standard deviation of D_i increases.

Many delay optimization techniques have been proposed so far, for example, division into pipeline stages, clock scheduling, technology mapping, gate/transistor sizing, buffer insertion. As for power optimization, gate/transistor sizing, multiple supply voltage technique, multiple threshold voltage technique, for example, are used. The above performance optimization techniques basically modify circuits such that long paths are shortened and short paths are lengthened. This operation equalizes the delay times of many paths in the circuit. Figure 3 explains the concept of path-balancing.

The path-balancing operation increases the number of the paths whose path delays are close to the maximum path delay (Fig. 3). These long paths have the possibilities of becoming the longest path in the circuit. The increase of the number of long paths corresponds to the increase of n in Fig. 1, and hence the average of $D_{circuit}$ becomes large when the number of the long paths increases. Performance optimization therefore increases statistically-distributed delay by the statistical phenomenon shown in Fig. 1.

The random variability in transistor characteristics tends to be larger when its transistor size becomes small [5]. Transistor sizing for power reduction reduces transistor widths considerably, and hence it may increase the amount of delay variability in each gate, which corresponds to the effect shown in Fig. 2.
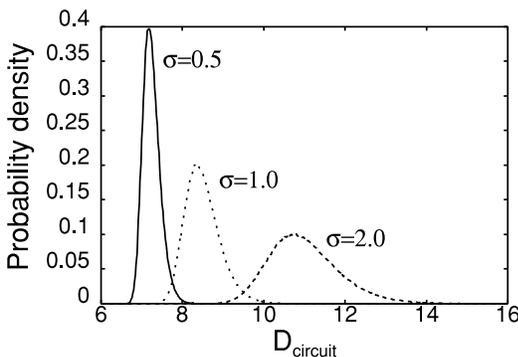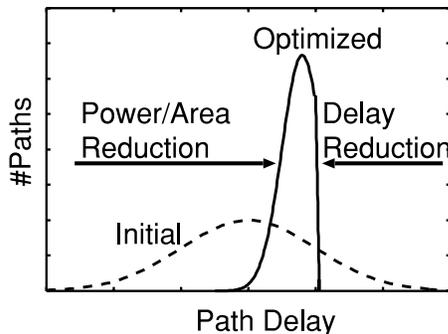
3. Experimental Analysis

This section shows some experimental results of statistical delay analysis. We reveal that statistically-distributed circuit delay increases by path-balancing operation.

We use the ALU part of a vector processor (dsp_alu) [6] and the circuit (des) included in LGSynth93 benchmark set for the experiments. These circuits are synthesized and mapped by a commercial logic synthesis tool [7] under tight delay constraints. The target library is a standard cell library used for actual fabrication in a $0.35\,\mu$m process with three metal layers. These circuits are placed and routed, and the wire capacitances are extracted from the layouts. We use these circuits as initial (not path-balanced) circuits. The number of gates used in dsp_alu and des are 14370 and 3837, respectively.

In order to obtain the path-balanced circuits, we utilize a transistor sizing method for performance optimization. We optimize the initial circuits by continuous transistor sizing for minimizing power dissipation under the delay constraint such that the delay does not increase from the initial value. Here in transistor sizing, the typical delay is evaluated, i.e. delay fluctuation is not considered. The optimization method used for the experiments is a heuristic method that reduces power dissipation greedily based on the result of sensitivity analysis [4]. Figures 4 and 5 represent the distributions of path delay in the initial and optimized circuits. The number of paths whose path delays are close to the longest path delay increases drastically, which corresponds to the increase of n in Fig. 1.

Fig. 2 Effect of max operation (σ is varied).

Fig. 3 Path-balancing effect caused by performance optimization.
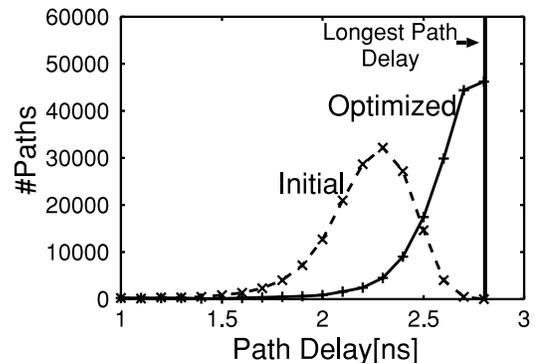
Fig. 4 Distributions of path delay (des).
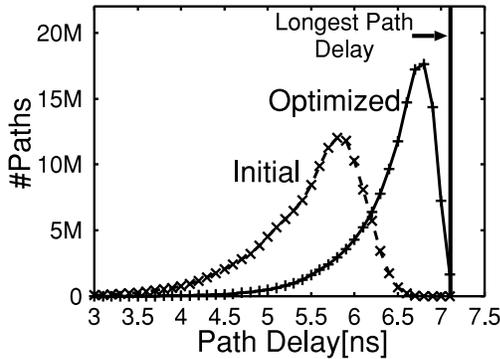
Fig. 5　Distributions of path delay (dsp_alu).

Fig. 6　Circuit delay distributions under a delay error model of $3\sigma=10\%$ (des).

Fig. 7　Circuit delay distributions under a delay error model of $3\sigma=10\%$ (dsp_alu).

Fig. 8　Circuit delay distributions under three delay error model of $3\sigma = 5, 10, 15\%$ (des).

3.1　Analysis of Delay Uncertainty

We first evaluate the impact of delay calculation error to the circuit delay uncertainty in the initial and optimized circuits. We assume an error model of gate delay such that the error of each gate is independently distributed according to a normal distribution with $3\sigma=10\%$ of its typical (no error) delay. The distribution of circuit delay is obtained by a Monte Carlo analysis as follows. We assign delay fluctuation to each gate in the circuit randomly according to the given normal distribution, and evaluate the circuit delay using a static timing analysis technique. The number of delay evaluation is 10,000. The results are shown in Figs. 6 and 7. The bar labeled "Typical" represents the delay time calculated using the typical (no error) delay time for each gate. The statistically-distributed delay of the optimized circuit increases as we expected. In des circuit (Fig. 6), the average delay of the optimized circuit is 2.98 ns, whereas the average of the initial circuit is 2.90 ns. The average delay increases by 3% by path-balancing although the circuit delay calculated from the typical delay does not change after the optimization. Also, the delay distribution of the path-balanced circuit moves far to the right of the typical delay. Therefore, in the case that the circuit is optimized considering only the typical delay, the statistically-distributed delay of the optimized circuit hardly satisfy the delay constraints.

Next, we examine the relationships between the uncertainty of gate delay and the distribution of circuit delay. We assume three models of gate delay uncertainties such that each gate delay fluctuates independently according to the normal distribution with $3\sigma=5$, 10 and 15% of its typical delay. In the case of a convex gate delay model for continuous transistor sizing, it is reported that 3σ of the estimation error in simple gates is 5 to 23% [8]. In this gate delay model, the error model of $3\sigma=15\%$ might be a reasonable assumption. We guess that the model of $3\sigma=5\%$ corresponds to the delay calculation using well-designed
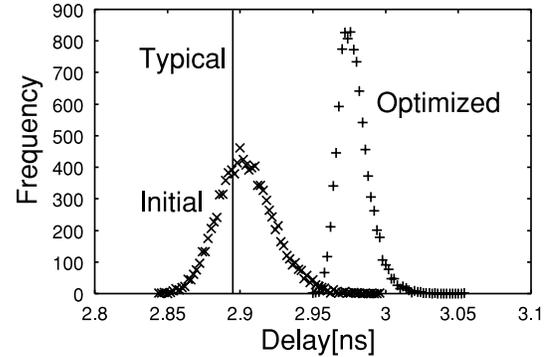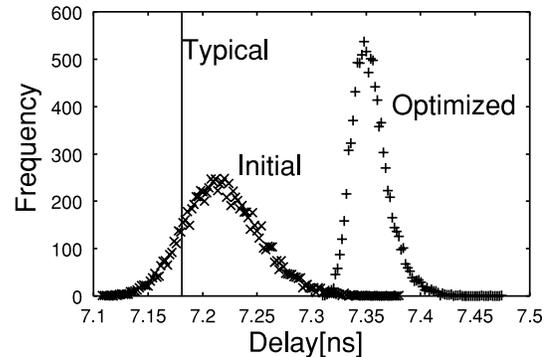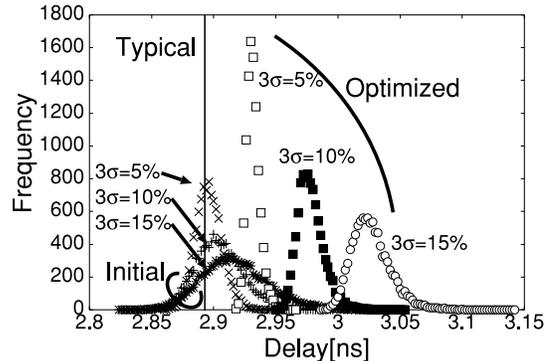
look-up tables characterized at many points (capacitive load, input transition time, transistor sizes). Figure 8 expresses the distributions of circuit delay under three error models. As the value of 3σ increases, the average and standard deviation of the circuit delay distribution becomes large, which is the same phenomenon shown in Fig. 2. Compared with the initial circuits, the increase of the statistically-distributed delay in the optimized circuit is large. Even when the accurate delay model with $3\sigma=5\%$ is used in performance optimization, there is a distinct delay difference between the statistically-

Table 1 Accuracy of statistical static timing analysis in worst-case delay calculation.

Circuit	3σ of Gate Delay Error (%)	Monte Carlo Worst-Case Delay (ns)	SSTA Worst-Case Delay (ns)	Error (%)
Initial	5	2.93	2.93	0.0
	10	2.97	2.97	0.0
	15	3.01	3.02	0.3
Optimized	5	2.96	2.96	0.0
	10	3.02	3.02	0.0
	15	3.09	3.10	0.3
Average	-	-	-	0.1

Table 2 CPU time of worst-case delay analysis.

Monte Carlo		Statistical Static Timing Analysis
#evaluation: 10 k	#evaluation: 1	
6044 s	0.6 s	1.9 s

distributed delay and the typical delay in the optimized circuit.

3.2 Worst-Case Delay Calculation

The increase of statistically-distributed circuit delay is different between the initial and the path-balanced circuits (Figs. 6–8). So, setting a design margin to avoid the delay violation is difficult and seems not to be a good way. To avoid this problem, statistical delay calculation [9] and the performance optimization based on statistical delay model [3], [10] are desired. We then apply the statistical static timing analysis (SSTA) method [3] to the initial and optimized circuits. The circuits and the error models of gate delay are the same with those used in the previous experiment. We evaluate the worst-case delay D_{worst}. The worst-case delay D_{worst} is defined such that the probability of $D_{circuit} \leq D_{worst}$ becomes 99.87%, which corresponds to the value of $m + 3\sigma$ in a normal distribution.

Table 1 shows the accuracy of the statistical static timing analysis (SSTA) method [3]. The column "3σ of Gate Delay Error" represents the value 3σ of gate delay uncertainties. SSTA method computes the worst-case delay D_{worst} within 0.3% error, and the average error is 0.1%. SSTA method can calculate the worst-case delay accurately irrespective of the initial and the optimized circuits. Table 2 represents the comparison of CPU time needed to derive the worst-case delay. The column "Monte Carlo" corresponds to the Monte Carlo simulation whose number of delay evaluation is 10,000. Each CPU time is the average CPU time of six calculations shown in Table 1. SSTA method calculates the worst-case delay as more than three thousand times as fast as the Monte Carlo simulation with 10,000 delay evalua-

tions. SSTA method requires only threefold CPU time of the Monte Carlo simulation whose evaluation number is one. In other words, SSTA needs threefold CPU time of the usual static timing analysis, although the average error of SSTA is 0.1%.

4. Conclusion

This paper examines the statistical effect of path-balancing operation to uncertainty in circuit delay. We demonstrate some examples that uncertainty in circuit delay is increased by path-balancing. We raise a notice that path-balancing increases uncertainty in circuit delay, and demonstrate a problem that a highly-optimized circuit may not satisfy delay constraints.

References

[1] S. Nassif, "Within-chip variability analysis," Proc. IEDM, pp.283–286, 1998.

[2] M. Berkelaar and E. Jacobs, "Sources and quantification of delay variations in a 250 nm CMOS digital cell library," Proc. IWLS, 2000.

[3] M. Hashimoto and H. Onodera, "A performance optimization method by gate resizing based on statistical static timing analysis," IEICE Trans. Fundamentals, vol.E83-A, no.12, pp.2558–2568, Dec. 2000.

[4] M. Hashimoto and H. Onodera, "Post-layout transistor sizing for power reduction in cell-base design," IEICE Trans. Fundamentals, vol.E84-A, no.11, pp.2769–2777, Nov. 2001.

[5] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," IEEE J. Solid-State Circuits, pp.1433–1439, Oct. 1989.

[6] T. Iwahashi, T. Shibayama, M. Hashimoto, K. Kobayashi, and H. Onodera, "Vector quantization processor for mobile video communication," Proc. ASIC/SOC Conf., pp.75–79, 2000.

[7] Synopsys Inc., Design Compiler Reference Manual, 1999.

[8] M. Ketkar, K. Kasamsetty, and S. Sapatnekar, "Convex delay models for transistor sizing," Proc. DAC, pp.655–660, 2000.

[9] M. Berkelaar, "Statistical delay calculation, a linear time method," Proc. TAU, pp.15–24, 1997.

[10] E.T.A.F. Jacobs and M.R.C.M. Berkelaar, "Gate sizing using a statistical delay model," Proc. DATE, pp.283–290, 2000.