

Comparative Analysis of Timing Yield Improvement under Process Variations of Flip-Flops Circuits

Hassan Mostafa, M. Anis, and M. Elmasry

Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada N2L3G1
 {hmostafa@uwaterloo.ca, manis@vlsi.uwaterloo.ca, elmasry@uwaterloo.ca }

Abstract—In synchronous systems, any violation of the timing constraints of the flip-flops can cause the overall system to malfunction. Moreover, the process variations create a large variability in the flip-flop delay in scaled technologies impacting the timing yield. Overtime, many gate sizing algorithms have been introduced to improve the timing yield. This paper presents an analysis of timing yield improvement of four commonly used flip-flops under process variations. These flip-flops are designed using STMicroelectronics 65-nm CMOS technology. The analyzed flip-flops are compared for power and power-delay product (PDP) overheads to achieve this timing yield improvement. The analysis shows that the sense amplifier based flip flop (SA-FF) has a power overhead and PDP overhead of 1.7X and 2.8X, respectively, much higher than that of the transmission-gate master-slave flip flop (TG-MSFF). The TG-MSFF exhibits the lowest relative power and PDP overheads of 30.87% and 9%, respectively.

I. INTRODUCTION

As CMOS technologies continue to scale towards the nanometer regime, the device parameters, such as threshold voltage, channel length, oxide thickness and mobility, will have large statistical process variations [1-5]. Consequently, these process variations will lead to delay uncertainty. Therefore, the deterministic design methodology is replaced by the statistical design methodology [6]. The process variations can be classified as die-to-die (inter-die) variations or within-die (intra-die) variations. In die-to-die variations, all devices on the same die are assumed to have the same parameters. However, devices on the same die are assumed to behave differently for within-die variations [1]. Although die-to-die variations were originally considered as the main source of process variations, within-die variations have now become the major design challenge as technology scales [3,4]. Moreover, the demand for higher performance has moved the clock frequencies up to multi-GHz in microprocessors and other advanced applications. These increased clock frequencies lead to very deep pipelining which means that hundreds of thousands of flip-flops are required to control the data flow under strict timing constraints. A violation of the timing constraints at a flip-flop can result in latching incorrect data causing the overall system to malfunction [7].

Deterministic gate sizing tools size the circuits to optimize the power-delay-product (PDP). However, due to random process variations, a large number of circuits might not meet the target delay. Consider as an intuitive example, a flip-flop that is designed for optimum PDP, which exhibits a specific target delay. Due to random process variations, the delay

can be modeled by a normal distribution with the probability density function (pdf) shown in Figure 1. Here, 50% of the total number of flip-flops will not meet the desired target delay constraint. Therefore, the flip-flops must be designed by using statistical sizing tools to improve the timing yield [8, 9]. In [10], a comparative analysis of the impact of the process variations on flip-flops power and delay is introduced. However, the analysis utilizes deterministic sizing tools to size the flip-flops, resulting in a poor timing yield (less than 50%). Therefore, the utilization of statistical sizing tools for timing yield improvement is more appropriate for an efficient and fair comparison, since the timing yield is the main concern in high performance applications. This paper provides a comparative analysis of timing yield improvement under process variations of four different flip-flops topologies, especially for the delay variability and the required power and PDP overheads for timing yield improvement. These flip-flops represent different trade-off choices between performance and power dissipation. The paper is organized as follows: Section 2 introduces the selected flip-flops designs and summarizes the flip-flops timing characteristics. Sections 3 and 4 describe the simulation setup and the simulation results, respectively. Finally, some conclusions are drawn in Section 5.

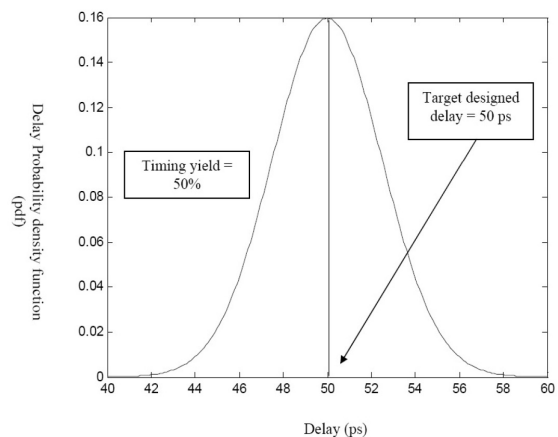


Fig. 1. The delay pdf due to process variations under deterministic gate sizing algorithms. It shows, intuitively, that up to 50% of flip flops will not meet the target delay (50 ps in this example)

II. TIMING CHARACTERISTICS OF FLIP-FLOPS TOPOLOGIES

A clock signal is used in clocked registers to control the timing of the data latching process. These clocked registers can be classified into latches and flip-flops. Latches are described as level-sensitive registers, because the input data is latched when the clock signal maintains a specific voltage level. Flip-flops are called edge-triggered registers, since the input data is latched by a transition edge in the clock signal waveform. The flip-flop can sample the input data correctly if the following constraints are satisfied:

- Setup time (T_{setup}) is defined as the minimum time that the input data should be available before the clock sampling edge arrival.

- Hold time (T_{hold}) is defined as the minimum time that the input data should be available after the clock sampling edge.

The timing relations among the input data, clock signal, and output data of a flip-flop can be obtained by the following timing characteristics [11]:

- Clock-to-output delay (T_{clk-Q}) represents the delay from the sampling clock edge (Clk) to the time the latched data is valid at the output (Q).

- Data-to-output delay (T_{D-Q}) represents the delay from a transition of the input data (D) to the time the latched data is valid at the output (Q). This delay is determined as the sum of the setup time and the clock-to-output delay.

In this paper, four different flip-flops are selected to represent the various trade-off choices between performance and power dissipation. Figure 2 and 3 depict the transmission-gate master-slave flip-flop (TG-MSFF) and the modified clocked CMOS master-slave flip-flop (M-C²MOS-MSFF), respectively. Both of them are implemented by cascading two complementary latches. This master-slave implementation results in robust flip-flop with a good hold time behavior. Moreover, they are used in standard libraries [10] which makes it so important to include them in this comparison. Figure 4 shows one of the fastest flip-flops, a semi-dynamic flip-flop (SD-FF) [12]. This flip-flop can be considered as a pulsed latch, since it samples the input data to the flip-flop output during a very short transparency period around the clock sampling edge. Accordingly, the input data can arrive after the clock edge which results in a negative setup time. Therefore, this flip-flop is used in high performance VLSI applications due to its relatively short data-to-output delay (T_{D-Q}), but at the expense of a poor hold time behavior and an excessive power consumption. Figure 5 denotes a sense-amplifier based flip-flop (SA-FF) with a NAND SR-latch [13]. This flip-flop can be viewed as a compromise between the master-slave robustness and pulsed latches high performance.

III. SIMULATION PROCEDURE AND SETUP

A. Optimum PDP Design

All flip-flops are optimized for minimum PDP by using a STMicroelectronics 65-nm CMOS technology transistor model, a 1V power supply voltage, a typical process corner,

a clock frequency of 1 GHz and pseudorandom input data with a 50% data activity [11]. The measured PDP is obtained by multiplying the data-to-output delay (T_{D-Q}), and the total power consumption which includes both the internal power dissipation and the local clock/data power dissipation [11]. The optimum setup time for each flip-flop is determined to achieve minimum PDP. The optimization process is conducted by using the CFSQP (C Version Feasible Sequential Quadratic Programming) optimization technique, implemented in Spectre-RF. This algorithm is based on the finite difference perturbation (FDP) method to determine how sensitive the PDP is to each device size. Then the algorithm provides the optimal sizing and setup time to achieve the minimum PDP.

B. Impact of Process Variations on Flip-Flop Delay

Monte Carlo analysis, including the mismatch between transistors is performed on the flip-flops at the optimal PDP point. An industrial hardware-calibrated statistical STMicroelectronics 65-nm CMOS transistor model is used in this Monte Carlo analysis. In this model, the transistor parameters such as the threshold voltage and the channel length are modeled by a normal distribution within the $\pm 3\sigma$ design space. The number of the Monte Carlo analysis points used is 5000 points to provide a good accuracy. The delay, power, and PDP variability are then obtained.

C. Functional Yield Improvement using Setup Time Margin

The optimum setup time determined in Subsection 3.A, is obtained by using a typical process corner to minimize the PDP. This results in a poor functional yield, since the setup time constraint of some of the flip-flop simulated Monte Carlo points will be violated. Typically, the functional yield of the flip flops with this setup time ranges from 85% to 95%. A setup time margin is added to achieve a functional yield greater than 99.9% [10]. This setup time margin is determined by sweeping the setup time and calculating the functional yield and the mean delay (T_{D-Q}). The setup time that achieves a functional yield greater than 99.9% and minimum (T_{D-Q}) is calculated.

D. Timing Yield Improvement using Gate Sizing

The delay variability is obtained from the Monte Carlo simulations by adopting the modified setup time. The timing yield of all the flip flops at the target delay (assumed to be the optimal delay achieved at minimum PDP) is less than 50%. A simplified gate sizing algorithm is employed. It is similar to that in [8] but the Lagrangian Relaxation (LR) optimization technique is replaced by the CFSQP optimization technique, implemented in Spectre RF. This algorithm utilizes the finite difference perturbation (FDP) method to determine how sensitive the delay and power are to each device size. This algorithm can be considered as one of the sensitivity based sizing algorithms. Figure 6 represents the gate sizing algorithm flow diagram. It starts with a given delay constraint (A_o) and timing yield constraint (Y_o), where (A_o) is the optimal delay obtained at minimum PDP. Then, the gate sizing values

obtained for the minimum PDP are used as an initial gate sizing values. Monte Carlo statistical analysis is then applied to obtain the delay variability. The standard deviation (σ) of the obtained delay distribution is calculated. Following that, the new delay constraint (A_o') is obtained by using the following equation:

$$(A_o') = (A_o) - n * (\sigma) \tag{1}$$

where n is dependent on the target timing yield value (Y_o) and can be obtained from the normal distribution tables. For example, in this paper, a timing yield of 99.87% ($Y_o = 99.87\%$) is required which means that " n " must equal 3.0 from the normal distribution tables. Following the calculation of (A_o'), an optimization problem is solved by employing CFSQP to determine the new gate sizing that matches the delay (A_o') and minimizes the total power consumption. These steps are repeated, until the timing yield constraint is achieved. It should be emphasized that the delay pdf changes after each iteration because the variations in the threshold voltage are a strong function of the transistor width [8]. If the delay standard deviation decreases or does not change from iteration n to iteration $n+1$, the timing yield constraint is met and the algorithm stops. However, if the delay standard deviation increases, more iterations are required to reduce the mean delay, resulting in higher power and PDP overheads. Figure 7 illustrates how this gate sizing algorithm improves the timing yield by moving the delay pdf to a shorter mean delay. Figure 8 displays the effect of the delay standard deviation on the number of iterations required.

E. Power and PDP Overhead

The last step is to repeat the Monte Carlo simulations on the improved timing yield flip-flops to obtain the delay, power, and PDP variability, as well as the power and PDP overheads required to achieve this timing yield improvement.

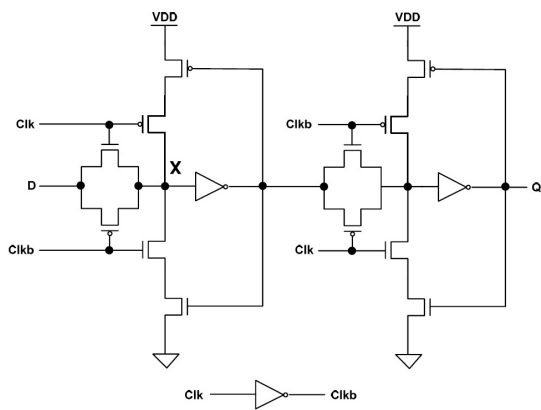


Fig. 2. The Transmission Gate based Master-Slave Flip-Flop (TG-MSFF)

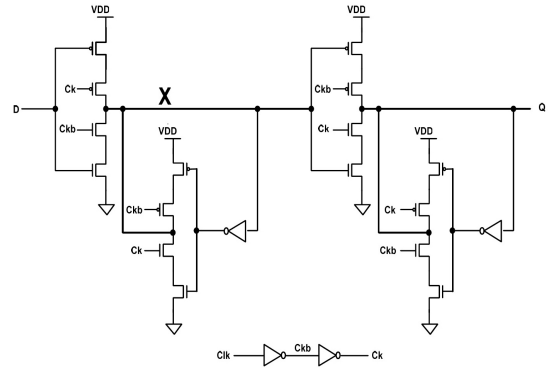


Fig. 3. The Modified Clocked CMOS Master-Slave Flip-Flop (M-C²MOS-MSFF)

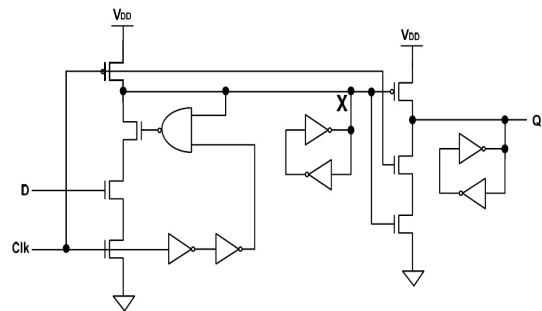


Fig. 4. The Pulsed Semi-Dynamic Flip-Flop (SD-FF)

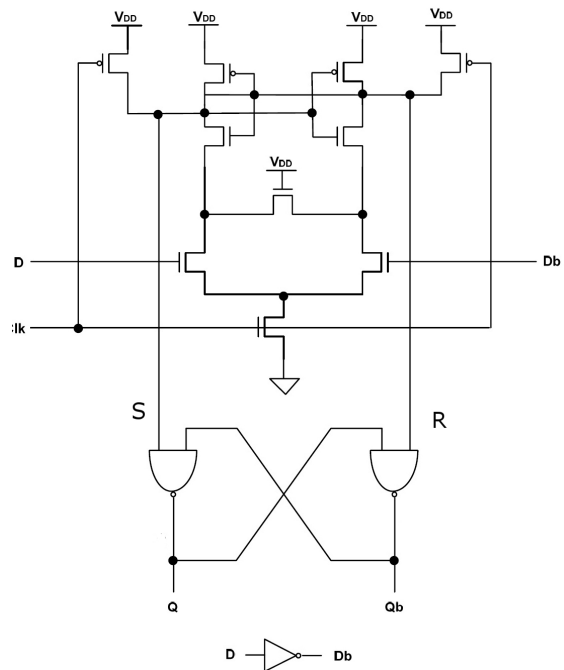


Fig. 5. The Sense-Amplifier based Flip-Flop (SA-FF)

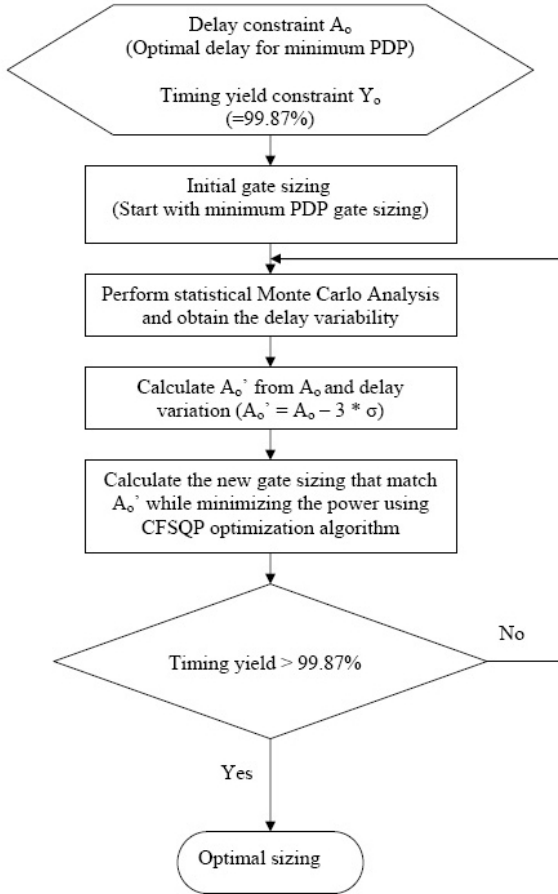


Fig. 6. The gate sizing algorithm flow diagram

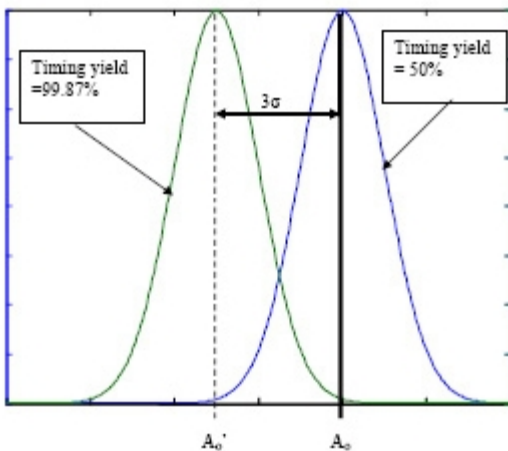


Fig. 7. The timing yield improvement under process variations employing gate sizing

IV. SIMULATION RESULTS AND DISCUSSION

Table 1 summarizes the simulation results for all the flip-flops. The comparison is performed for the improved timing

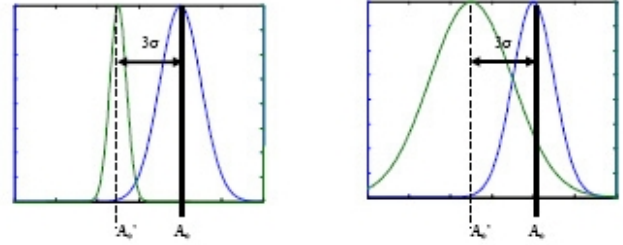


Fig. 8. The delay standard deviation effect on the algorithm number of iterations. If the standard deviation decreases, the timing yield constraint is met in a single iteration. However, if it increases, more iterations are required to achieve the target timing yield

yield flip-flops. The optimal T_{D-Q} delay is adopted as the target delay constraint for the timing yield improvement for each flip-flop. The SD-FF has 2.35X higher performance compared to the M-C²MOS-MSFF at the expense of power dissipation that is 1.4X greater than that of the M-C²MOS-MSFF. It should be noted that the delay standard deviation, reported in Table 1, differs from that introduced in [10] due to the following two facts. First, the work in [10] models the threshold voltage variations as a uniform distribution, while it is more accurate to model it as a normal distribution [1]. Secondly, the delay standard deviation is a strong function of the gate sizing. This gate sizing targets the PDP optimization in [10] while the timing yield improvement is the primary objective in this paper.

Figure 9 shows the relative power and the relative PDP overheads of the improved timing yield flip-flops. According to the results in Figure 9, the SA-FF has a power overhead of 58.2% which is 1.7X higher than that of the TG-MSFF. Moreover, the SA-FF exhibits a 25.26% PDP overhead which is 2.8X higher than that of the TG-MSFF. The reason for this is that the SA-FF implementation utilizes a symmetric cross-coupled architecture which suffers from device mismatch more than all other flip-flops. The M-C²MOS-MSFF delay standard deviation increases from one iteration to the next. Consequently, this flip-flop requires the highest number of gate sizing algorithm iterations, which increases the required power overhead to minimize the mean delay. The SD-FF has the same PDP overhead as the M-C²MOS-MSFF, while having 1.2X less power overhead. The TG-MSFF exhibits the lowest power and PDP overhead of 30.87% and 9%. This advantage is due to the fact that its delay standard deviation decreases with iterations. This flip flop requires the lowest number of gate sizing algorithm iterations. Correspondingly, its power overhead is smaller than that of each of the other flip flops.

Figures 10, 11, 12, and 13 show the average power consumption versus the T_{D-Q} delay space for the improved timing yield flip-flops. It is evident that all but one of the flip-flops have a timing yield > 99.87%, where the optimal delay is the timing constraint. The TG-MSFF exception, achieves a timing yield of 99.6%, at most. This emphasizes the need for a more efficient algorithm. However, we do not attempt

to automate the process since this is not the main purpose of this research. Moreover, a timing yield of 99.6% is close to the main timing yield objective of 99.87%. These figures further demonstrate the power $\pm 3\sigma$ variations. The TG-MSFF and SD-FF exhibits the highest power variations of 12% and 11.5%, which is 1.7X and 1.6X, respectively, higher than that of the M-C²MOS-MSFF.

Finally, a trade-off between the power overhead, required to achieve the timing yield improvement and the corresponding power variability is indicated by these simulation results; that is, the higher the required power overhead for the timing yield improvement is, the lower the power variations are. For instance, the M-C²MOS-MSFF flip-flop has the highest power overhead (53.92%) and the lowest power variability (1.18%), whereas the TG-MSFF flip-flop has the lowest power overhead (30.87%) and the highest power variability (2.0%). The SA-FF does not follow this observation due to its increased variations from the transistor mismatches.

Table 1: Simulation results for different flip flops designs

		(TG-MSFF)	(M-C ² MOS-MSFF)	(SD-FF)	(SA-FF)
T _{D-Q} delay	Optimal (ps)	48.61	76.36	32.55	49.6
	Mean (ps)	40.71	57.67	26.1	39.35
	σ (%)	6.71	4.62	5.89	6.51
Power	Optimal (μ W)	8.86	11.48	16.28	12.74
	Mean (μ W)	11.65	17.74	23.69	20.21
	σ (%)	2.0	1.18	1.9	1.56
	Relative overhead (%)	30.87	53.92	44.8	58.2
PDP	Optimal (fJ)	0.43	0.88	0.53	0.632
	Mean (fJ)	0.47	1.02	0.62	0.8
	σ (%)	5.54	4.5	5.6	6.0
	Relative overhead (%)	9.0	15.77	15.7	25.26

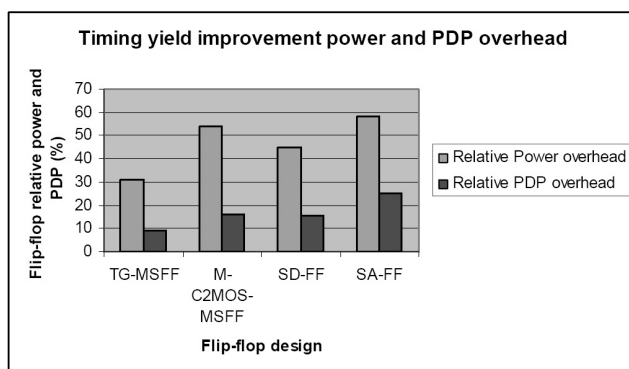


Fig. 9. The relative power and PDP overheads due to timing yield improvement

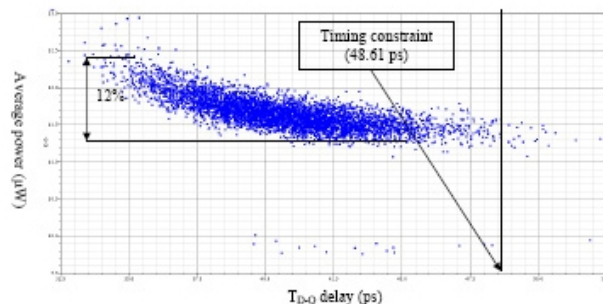


Fig. 10. The power-delay spread of the TG-MSFF

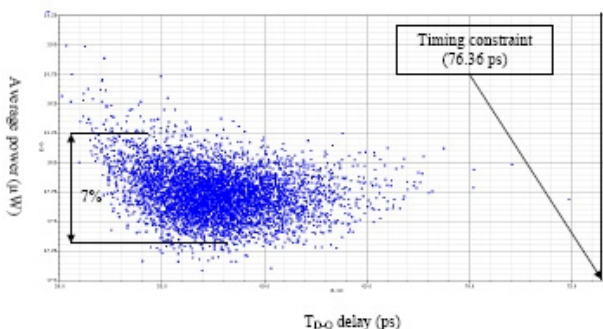


Fig. 11. The power-delay spread of the M-C²MOS-MSFF

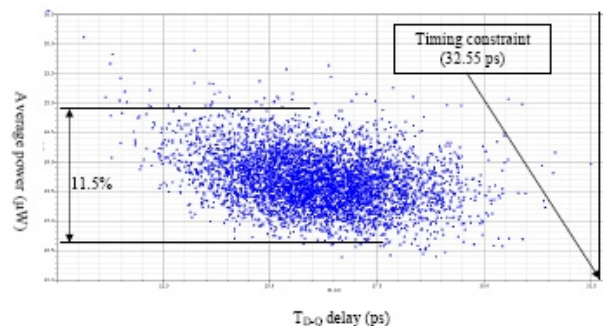


Fig. 12. The power-delay spread of the SD-FF

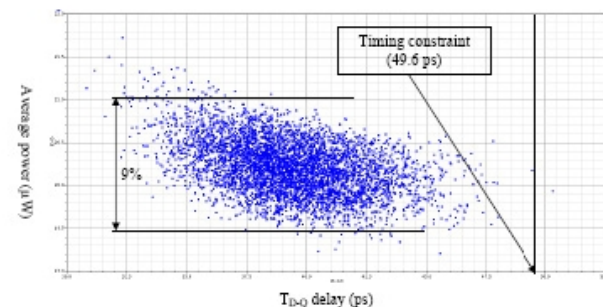


Fig. 13. The power-delay spread of the SA-FF

V. CONCLUSION

A comparative analysis of improved timing yield four commonly used flip-flop topologies is introduced. The SA-FF suffers from device mismatch which results in a power overhead of 1.7X and a PDP overhead of 2.8X higher than that of the TG-MSFF, respectively. The M-C²MOS-MSFF has relatively, the same power overhead as the SA-FF. The TG-MSFF exhibits the lowest power and PDP overheads of 30.87% and 9%, respectively, due to its decreased delay standard deviation with the gate sizing algorithm iterations. Moreover, it is observed that there is a trade-off between the required power overhead to achieve the timing yield improvement and the corresponding power variability.

REFERENCES

- [1] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter Variations and Impact on Circuits and Microarchitecture," *Proceedings of the 40th Conference on Design Automation (DAC '03)*, pp. 338-342, 2003.
- [2] S. Borkar, T. Karnik, and V. De, "Design and Reliability Challenges in Nanometer Technologies," *Proceedings of the 41st Conference on Design Automation (DAC '04)*, pp. 75-75, 2004.
- [3] K. Bowman, S. Duvall, and J. Meindl, "Impact of Die-to-die and Within-die Parameter Fluctuations on the Maximum Clock Frequency Distribution for Gigascale Integration," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 2, pp. 183-190, 2002.
- [4] H. Masuda, S. Ohkawa, A. Kurokawa, and M. Aoki, "Challenge: Variability Characterization and Modeling for 65-nm to 90-nm Processes," *Proceedings of the IEEE 2005 Custom Integrated Circuits Conference (CICC)*, pp. 593-599, 2005.
- [5] A. Keshavarzi, G. Schrom, S. Tang, S. Ma, K. Bowman, S. Tyagi, K. Zhang, T. Linton, N. Hakim, S. Duvall, J. Brews, and V. De, "Measurements and Modeling of Intrinsic Fluctuations in MOSFET Threshold Voltage," *Proceedings of the 2005 International Symposiums on Low Power Electronics and Design (ISLPED '05)*, pp. 26-29, 2005.
- [6] F. N. Najm, "On the Need for Statistical Timing Analysis," *Proceedings of the 42nd Conference on Design Automation (DAC '05)*, pp. 764-765, 2005.
- [7] P. R. Gada, W. R. Roberts, and D. Velenis, "Effects of Parameter Variations on Timing Characteristics of Clocked Registers," *International Conference on Electro Information Technology*, pp. 1-4, 2005.
- [8] S. H. Choi, B. C. Paul, and K. Roy, "Novel Sizing Algorithm for Yield Improvement Under Process Variation in Nanometer Technology," *Proceedings of the 41st Conference on Design Automation (DAC '04)*, pp. 454-459, 2004.
- [9] A. Agarwal, K. Chopra, and D. Blaauw, "Statistical Timing Based Optimization Using Gate Sizing," *Proceedings of the Conference on Design, Automation and Test in Europe (DATE '05)*, pp. 400-405, 2005.
- [10] M. Hansson, and A. Alvandpour, "Comparative Analysis of Process Variation Impact on Flip-Flop Power-Performance," *Proceedings of the 2007 IEEE Symposiums on Circuits and Systems (ISCAS 2007)*, pp. 3744-3747, 2007.
- [11] V. Stojanovic, and V. G. Oklobdzija, "Comparative Analysis of Master-Slave Latches and Flip-Flops for High Performance and Low Power Systems," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 4, pp. 536-548, 1999.
- [12] F. Klass, "Semi-Dynamic and Dynamic Flip-Flops with Embedded Logic for High Performance Processors," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 712-716, 1999.
- [13] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, "Digital Integrated Circuits A design Perspective , " *second edition, Prentice Hall*, 2002.