Design Methodology for Clocked Storage Elements Robust to Process Variations

Joosik Moon , Mustafa Aktan , and Vojin G. Oklobdzija
Department of Electrical and Computer Engineering
University of Texas at Dallas, Richardson, TX, USA
{jxm055000, aktanmus, vojin}@utdallas.edu

Abstract— **In this paper we address the effects of process variations on the performance of clocked storage elements. Two types of structures are selected and evaluated by using energy-delay space analysis. The delay variations of the energy-efficient designs for each clocked storage element are measured and compared. We show how much topology selection is important in order to minimize the impact of process variations on performance and reliability.**

Keywords-Clocked storage elements; Process variations

I. INTRODUCTION

As semiconductor technology scales down, process variations start to cause increasingly detrimental impact on the performance of integrated circuit. Starting with 65-nm CMOS technology, the sensitivity to process variation became significantly large. This is mainly due to the phenomenon of velocity overshoot and the drain-induced barrier lowering (DIBL) caused by the correlations between physical parameters and increases as the technology features scale down [2]. These variations occur from wafer to wafer, die to die, and even within the single chip. Therefore, it is possible that transistors designed to have the same feature sizes in a single cell can show different performance. The challenge of designing the highly integrated circuits such as processors is that overall circuit is starting to behave in an unreliable manner due to the increased process variations [3].

The pipelining is commonly employed in high-performance processors. In pipelined designs, the output of logic operations from the previous stage need to be stored and are going to be used as the inputs for the following stages in the next clock cycle. Therefore, clocked storage elements (CSE) occupy a large portion of clock cycles and total power in the modern pipelined processor. The design of a CSE robust to process variations is important to ensure the correct functionality. In this work, the variation of delay through the critical path is explored at different delay points that are obtained from the transistor sizing. The analysis is conducted across different structures of CSEs. This paper is organized as follows. Section II presents the related work concerning the design of CSE. Section III describes how to obtain the transistor sizing with desired characteristics. In section IV and V, topologies for representative CSEs and a methodology for evaluation of process variations are presented. Section VI shows the comparative analysis of CSEs under the process variations.

II. RELATED WORKS

The design of CSEs has focused on the improvement of speed and power. The analysis and comparison of the CSEs based on the energy-efficient characteristics are performed for the design selection [4]. The effects of varying load and circuit depth within the pipeline stages are also examined [5]. This analysis provides the specific cycle time and target application for each CSE [6]. However, the impact of process variations has not been considered in CSE analysis. In addition, technology scaling motivates the selection of CSEs robust to process variations. The scaled-down technology demands the design less suffered by the process variations.

III. METHODOLOGY FOR TRANSISTOR SIZING FOR ROBUSTNESS

The transistor sizing is an important factor affecting the energy-delay tradeoff [1]. Thus, the optimal sizing can be found by the analysis of ED space. The size of each transistor is changed to evaluate the performance and energy of all possible sizing under a fixed input size and fixed output load constraint. Only one combination of transistor sizing exists that yields the minimum energy for a specific delay. The points fulfilling this condition are grouped as a subset which shows energy-efficient characteristics [7]. The subset of points with minimum energy for each delay target can be classified according to the factor affecting sensitivity. The points sensitive to the change of delay are obtained from the steep region in the subset of energy-efficient characteristics, labeled as “high energy sensitivity region”. In these points, larger transistors are used for the fast speed, which increase the energy consumption. In a similar fashion, the points with low energy consumption and high delay are obtained from the flat region, labeled as “high delay sensitivity region”, in the subset of energy-efficient characteristics. The minimum energy-delay-product (EDP) also locates on the points with energy-efficient characteristics. Therefore, the evaluation on energy-delay space is limited to the region where points have the energy-efficient characteristics.

* This work has been supported by SRC grant No.1799

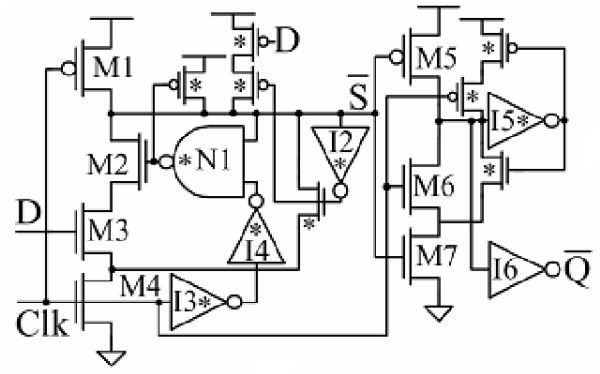

Fig. 1. UltraSPARC flip-flop (USPARC)

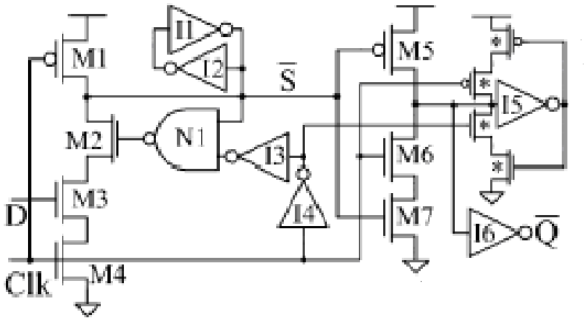

Fig. 2. Semi-Dynamic Flip-Flop (SDFF)

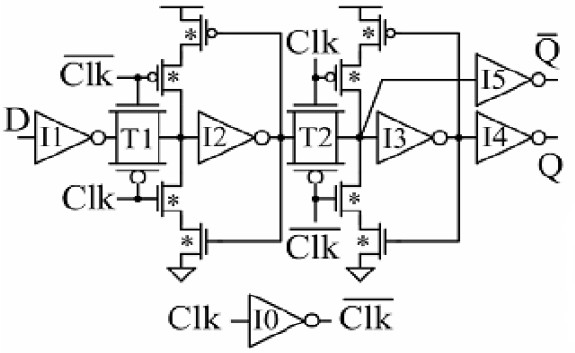

Fig 3. Transmission-Gate Master-Slave Latch (TGMS)

IV. TYPES OF CSES USED IN THIS EXPERIMENT

In this work, single-ended topologies are considered which represent the major group of structure among CSEs. Dynamic structure and static latch are the major groups for single-ended CSEs. We have examined two representative structures that are already being used in industry. The more robust will prevail and continued to be used as technology scales down. The first structure is the UltraSPARC flip-flop (USPARC, Fig. 1, [8]) which is redesigned from the Semi-Dynamic Flip- Flop (SDFF, Fig. 2, [9]) to reduce the soft error rates. The second one is the transmission-gate master-slave latch (TGMS, Fig. 3, [10]) implementing a typical master-slave latch topology

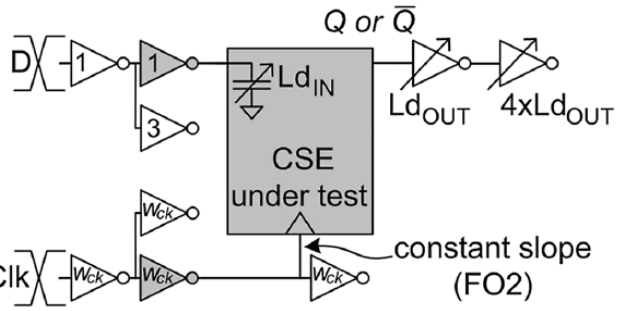

Fig. 4. Simulation set-up for single-ended CSEs

with transmission gates. The USPARC is used in Sun UltraSPARC-III and the TGMS is used in IBM PowerPC 603 processor.

In the dynamic structures such as SDFF and USPARC, the high speed is achieved by the pre-charged node that is evaluated during the active clock. However, relatively high energy is consumed because of the switching activity associated with pre-charging operation. TGMS, on the contrary, operates with the low power consumption because the pulse generation is not needed in this structure.

V. METHODOLOGY FOR EVALUATING PROCESS VARIATIONS EFFECTS

In order to determine the sizing combinations in the energy-efficient region, all the data of energy-delay characteristics are extracted through H-Spice simulations. The 45nm technology node from the latest version of the Predictive Technology Models (PTM) is used for the transistor models. The minimum width of a transistor is taken to be 0.1μm and the transistor size is incremented to the minimum width. The delay and energy of CSEs are extracted from each combination of transistor sizing. The simulation set-up for a single-ended CSE is shown in Fig. 4. As shown in the schematic of set-up, the data input and clock input are connected to the inverters which drive the pulse with the appropriate slope. The transistor sizing determines the varying load of clock to provide the clock pulse with the constant slope of FO2. The size of the clock driver is also determined by the transistor sizing. Two inverters that buffer CSE's output provide the fixed output load.

The gate length (L_g), oxide thickness (t_{ox}), and threshold voltage (V_t) are the physical parameters selected to evaluate the effects of process variations on the performance of CSEs. The gate length is an important physical parameter affecting the performance since it is directly related to the effective gate length (L_{eff}). In the aggressively scaled-down technology, DIBL effects increases the sensitivity to variation of L_{eff} because the DIBL affects the difference of threshold voltages [11].

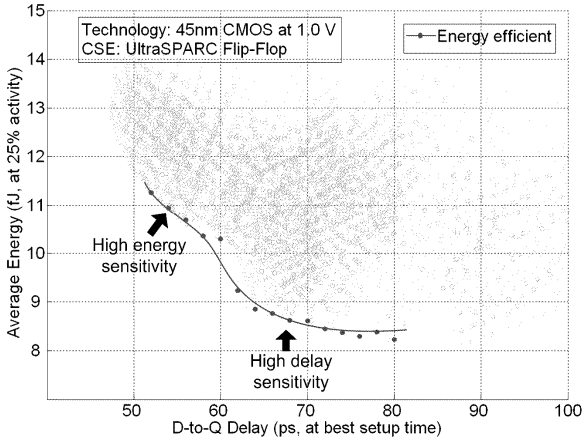

Fig. 5. Energy-efficient characteristics of the UltraSPARC Flip-flop

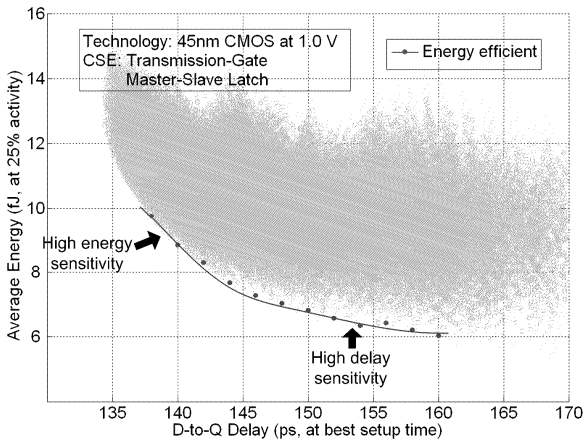

Fig. 6. Energy-efficient characteristics of the Transmission -Gate Master-Slave Latch

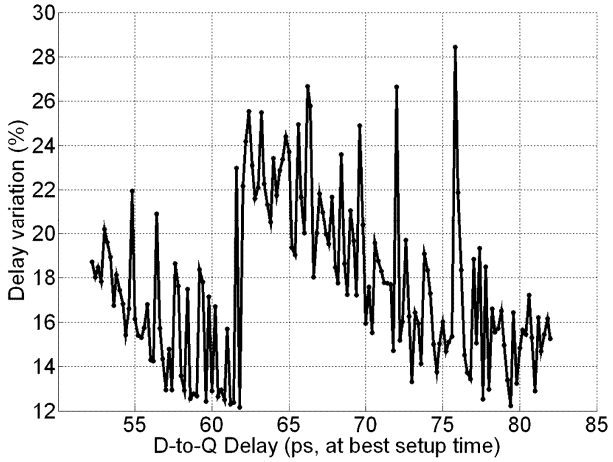

Fig. 7. The delay variations (%) of USPARC at the points in the energy-efficient region

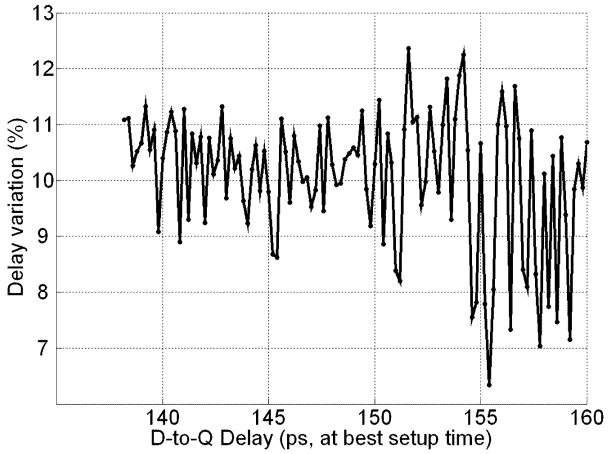

Fig. 8. The delay variation (%) of TGMS at the points in the energy-efficient region

Another important factor for the process variation is threshold voltage which defines the channel doping concentration. The effective carrier mobility and saturation velocity, important parameters affecting the performance of a device, are determined by the concentration of channel doping. Furthermore, the correlation between these parameters increases the effects of process variations. We assume that all variation parameters used in this experiment follow Gaussian distribution.

The nominal value and $\pm 3\sigma$ variations of the normal distributions are presented in the Table 1. The 3σ values are obtained from the International Roadmap for Semiconductors (ITRS). The real values of physical parameters with process variations are produced by the Monte-Carlo technique which is a statistical way of generating random numbers [12]. This technique considers not only inter-die, but also intra-die variations which becomes an important part of process variations due to the scaled-down feature size of the device. The delay simulations are iterated for 30 times at each sizing combination found in the energy-efficient region. The values of parameters are deviated from the expected feature size for each iterative simulation. These deviations are regarded as process variations and they do affect the delay of CSEs. Therefore, the different values of delay are extracted from one combination of sizing.

VI. COMPARISONS OF CSES IN TERMS OF PROCESS VARIATIONS

Fig. 5 shows the ED characteristics of USPARC examined in this analysis. The ED space for TGMS is given in Fig. 6. Both figures show the energy-efficient points evaluated for the impact of process variations on the D-to-Q delay. The delay variations measured at these points are

calculated from the ratio of variations to the nominal delay. The average values are given in Table 2. The delay variation obtained from the point on the curve of energy-efficient characteristics is quite lower than the one from other region of ED space. As shown in Fig. 7 and 8, there are oscillations in delay variations at the points on the energy-efficient curve. For the USPARC, the variations for D-to-Q delay in the range 52 ps to 62 ps correspond to "high energy sensitivity region" and are usually lower than the variations in "high delay sensitivity region" which corresponds to the range starting from 65 ps to 75 ps. In case of TGMS, the pattern of delay variations shows more oscillations in the "high delay sensitivity region" than in the "high energy sensitivity region". Especially, there are big peaks of oscillation from 151 ps to 156 ps.

The mean value of delay variations is calculated from the points of energy-efficient characteristics of each CSE and compared to determine which CSE suffers less from the impact of process variations on performance. As shown in Table 2, the delay variation resulting from the process variations in the USPARC is larger than the one of TGMS. The average variation of USPARC is almost two times larger than TGMS. Even the smallest variation in USPARC is close to the largest variation in TGMS. This suggests that TGMS is a more robust structure to the impact of process variations on delay.

TABLE 1. Nominal and $\pm 3\sigma$ variation values of process variations simulated

	Gate Length (*Lg*)	**Oxide Thickness (*tox*)**	**Threshold Voltage (*Vth*)**
Nominal Value	45 nm	0.9 nm	200 mv
$\pm 3\sigma$ variation (%)	$\pm$ 12	$\pm$ 5	$\pm$ 12

TABLE 2. D-to-Q delay variation of each CSE

	Mean value of delay variations
USPARC	18%
TGMS	10%

VII. CONCLUSIONS

In this paper, we compared two CSEs representing the dynamic flip-flop and static latch in terms of the delay variations due to the process variations. The combinations of transistor sizing in the energy-efficient region are obtained by the analysis of entire ED space. The average variation of delay obtained from the points with energy-efficient characteristics is smaller in TGMS than USPARC. Results have shown that static latch is a more robust structure to process variations. However, to strengthen the result and give a sound explanation of the cause, the same analysis should be conducted for other static and dynamic structures as well. Otherwise, the number of transistors can be thought as an important factor for immunity to process variations, since the number of transistors consisting TGMS is less than that of USPARC.

REFERENCES

[1] V. G. Oklobdzija, V. Stojanovic, D. Markovic, and N. Nedovic, *Digital system clocking, high-performance and low-power aspects*, John Wiley, January 2003.

[2] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45 nm early design exploration," *IEEE Trans. on Electron Devices*, vol. 53, no. 11, Nov. 2006.

[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," *J. of Solid-State Circuits*, 37(2), Feb. 2002.

[4] V. Zyuban, "Optimization of scannable latches for low energy", *IEEE Trans. Very Large Scale Integrat. (VLSI) Syst*., vol. 11, no. 10, pp. 778-788, Oct. 2003.

[5] S.Heo and K. Asanovic, "Load-sensitive flip-flop characterizations", in *Proc. IEEE Computer Society Workshop on VLSI*, 2001, pp. 87-92.

[6] C. Giacomotto, N. Nedovic, and V. G. Oklobdzija, "The effect of the system specification on the optimal selection of clocked storage elements," *IEEE J. Solid State Circuits*, vol. 42, no. 6, June 2007.

[7] C. Giacomotto, N. Nedovic, and V. G. Oklobdzija, "Energy-delay space analysis for clocked storage elements under process variations," in *PATMOS 2006*, LNCS 4148, pp. 360 – 369, 2006.

[8] R. Heald, *et al* "A third-generation SPARC V9 64-b microprocessor," *IEEE J. Solid-State Circuits*, vol. 35, no. 11, pp. 1526-1538, Nov. 2000.

[9] F. Klass, "Semi-dynamic and dynamic flip-flops with embedded logic," in *Proc. Symp. VLSI Circuits*, 1998, pp. 108–109.

[10] G. Gerosa, *et al* "A 2.2 W, 80MHz superscalar RISC microprocessor," *IEEE J. Solid-State Circuits*, vol. 29, no. 12, pp. 1440-1454, Dec. 1994.

[11] E. Sangiorgi, N. Barin, M. Braccioli, and C. Fiegna, "32nm technology node double-gate SOI MOSFET using SiO2 gate stacks", *Nano CMOS, 2006 International Workshop on*, Jan. 2006.

[12] A. Srivastava, D. Sylvester, and D. Blaauw, *Statistical Analysis and Optimization for VLSI: Timing and Power*, Springer, 2005.