

Energy-Delay Space Analysis for Clocked Storage Elements Under Process Variations

Christophe Giacomotto¹, Nikola Nedovic², and Vojin G. Oklobdzija¹

¹ Advanced Computer Systems Engineering Laboratory,
Dept. of Electrical and Computer Engineering, University of California, CA 95616, USA
{giacomoc, vojgin}@ece.ucdavis.edu
<http://www.acsel-lab.com>

² Fujitsu Laboratories of America, Sunnyvale, CA 95616, USA
nikola.nedovic@us.fujitsu.com

Abstract. In this paper we present the effect of process variations on the design of clocked storage elements. This work proposes to use the Energy-Delay space analysis for a true representation of the design trade-offs. Consequently, this work also shows a comparison of clocked storage elements under a specific set of system constraints for typical corner design and high yield corner design. Finally, we show that designing for high yield can affect the choice of topology in order to achieve energy efficiency.

1 Introduction

The impact of process variations on Clocked Storage Elements' (CSEs) energy and delay is dependent on the sizing of the individual transistors [12]. Hence, evaluating the effect of process variations to a specific CSE topology requires a complete analysis in ED (Energy-Delay) space [15]. This analysis is then extended across a set of topologies for purpose of comparison. Several methods have been used to compare CSEs in terms of performance and/or energy [1][2][7]. The transistor tuning optimization is usually done for a given objective function or metric such as EDP (Energy-Delay Product) or Power Delay Product [1][7], and, more recently, generalized with cost function approaches [2]. However, in these cases, results are shown as a single optimum design solution and the quality of the designs is quantified using a single metric. This approach can be misleading as it fails to show all performance versus energy tradeoffs that a particular topology offers. Typically, the process of designing CSEs in mainstream high performance and low power processors starts with the choice of a topology accordingly to a rough performance and power requirements estimates. Only then, when the choice is made, transistor sizing can help meeting the energy or delay target and finally process corner variations are taken in account. In this work, the objective is to show that taking process corner variations in account can change the topology selection. This analysis reveals the impact of high yield design on an envelope of high-performance and low-power CSEs in their best energy efficient configurations.

2 Efficient Energy-Delay Approach

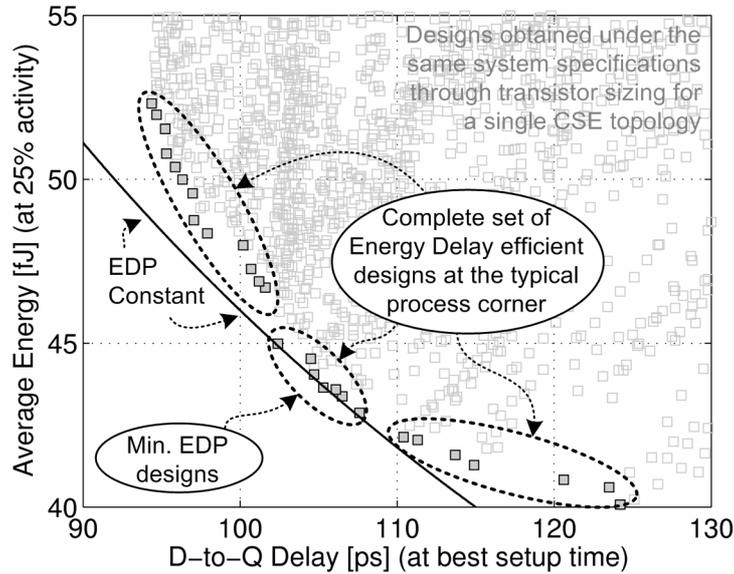


Fig. 1. Energy efficient designs for a single CSE topology through transistor sizing with fixed input/output load at the typical process corner

For a specific CSE topology, there is only one combination of transistor sizes that yields minimum energy for a given delay. As the entire design space is explored, a subset of combinations remains that represents the configurations that yield the smallest energy for each achievable delay. This subset is referred to as the energy efficient characteristic for a CSE [2]. Fig. 1 shows such characteristic where the D-to-Q delay represents the minimum achievable delay which occurs at the optimum setup time and the average energy is calculated for 25% data activity with a 1ns clock period. As shown in Fig. 1, from this characteristic, a wide range of ED points are possible. For the low energy sizing solutions, the delay has a high sensitivity to the transistor sizing, and for the high speed sizing solutions, the energy has a high sensitivity to the transistor sizing. Fig. 1 shows that the minimum EDP, typically used as an ad-hoc energy-performance tradeoff metric, is achieved for a range of possible configurations. Restricting the design space to EDP solutions would discard all the other potential design solutions and be misleading on the energy or delay achievable by the topology. In general case, however, the optimum design point depends on the parameters of the environment of the CSE such as the energy-efficient characteristic of the logic block used in the pipeline and target clock frequency [15]. Hence, depending on the surrounding logic the CSE design chosen may be in the high energy sensitivity region or the high delay sensitivity region (Fig. 1). In our analysis, we compare entire energy efficient characteristics of the CSEs, rather than a single energy-delay metrics. In this way, the entire space of possible designs is explored and the impact of process variations onto a topology and between topologies can be fully evaluated.

3 CSE Simulation Methodology

3.1 Circuit Setup

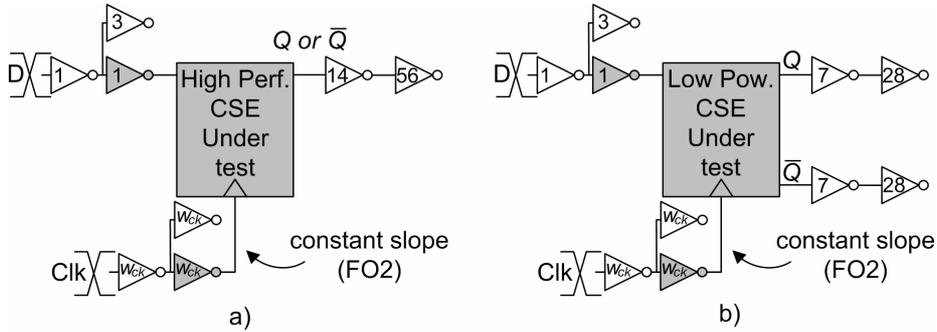


Fig. 2. Simulation setup for single ended CSEs, W_{clk} is sized to achieve an FO2 slope for the clock input, a) High Performance setup, b) Low Power setup

For this process corner evaluation and topology comparison we chose to limit our analysis to single ended flip-flops and master-slave latches. In this work we propose two distinct setups:

- High performance (Fig. 2a) where one output is loaded, either Q or Qb, whichever comes first in terms of delay and a load of 14x min. sized inverters which is considered representative of a typical moderate to high capacitive load of a CSE in a critical path [1].
- Low power (Fig. 2b) where both outputs are loaded with 7x min. sized inverters. The worst case delay (D-to-Q vs. D-to-Qb) is reported for this setup.

In both setups shown in Fig. 2, the input capacitance of the CSE under test is limited to a maximum equivalent capacitance of 4 minimum sized inverters and is driven by a minimum sized inverter. These limitations restrict the scope of this comparison since load and gain have a significant impact on the ED behavior of each CSE topology. Independently, for low power designs, the simulation setup requires further restrictions on the CSE topology itself: the input must be buffered (i.e. no passgate inputs are allowed), and the output must be buffered as well (i.e. no state element on the output). Our setup requires that the slope of the clock driving the CSE must remain constant. As the configuration under test changes, the load of the clock changes as well. In order to accommodate for this variation, the size of the clock driver (W_{clk} in Figure 2) is chosen to keep the FO2 slope characteristic.

3.2 Delay and Energy Quantification

The primary goal is to extract an accurate energy efficient characteristic of sizing configurations for each Flip-Flop and Master-Slave latch. These energy efficient

curves must include layout and wire parasitic capacitance estimates, which are re-evaluated for each combination of transistor sizes tried. The set of H-SPICE simulations are done with a nominal 130nm process and the granularity for the transistor width is set to 0.32um, which is the minimum transistor width in this technology. The FO4 delay for this technology is 45ps. In order to accurately quantify delay for each transistor size combination and for each topology, the setup time optimization must be completed as well [1]. Nedovic et al. [6] show, in the same technology, a minimum D-Q delay zone flat for at least 10ps of D to clock variation for all CSEs presented. The granularity chosen for the simulations performed in this work was set to 5 ps, which yields a negligible D-to-Q delay error vs. setup time. The energy is measured by integrating the current necessary for the operation of the CSE, the clock driver and the data driver(s) at the nominal voltage of operation as shown by the gray elements in Fig. 2. This energy is quantified for each type of state transition (0→0, 0→1, 1→0, and 1→1) over a 1ns clock period and combined to obtain the total energy for any desired activity factor [8]. For this technology node and the clock period we use, the offset in energy due to leakage is negligible.

3.3 Simulated Topologies

In this work, we examine most of the conventional single-ended topologies of the CSEs used in the industry. The CSEs are divided in two classes: High Performance and Low Power CSEs. High performance topologies consists of the Semi-Dynamic

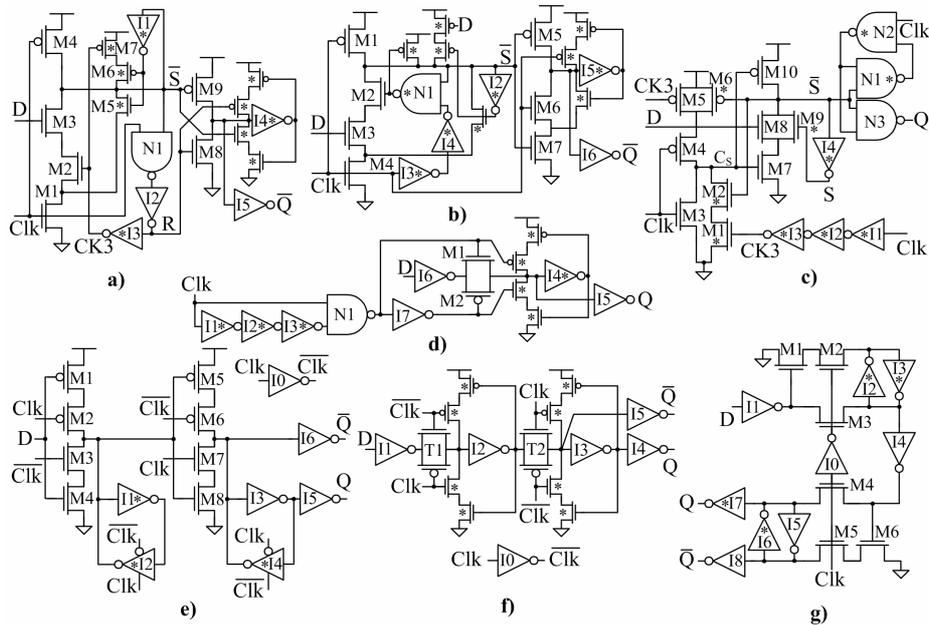


Fig. 3. Clocked Storage Elements: a) IPP: Implicitly Pulsed Flip-Flop with half-push-pull latch, b) USPARC: Sun UltraSPARC III Semi-Dynamic Flip-Flop, c) STFF-SE: Single Ended Skew Tolerant Flip Flop, d) TGPL: Transmission Gate Pulsed Latch, e) Modified C²MOS Master Slave Latch, f) TGMS: Transmission Gate Master-Slave latch, g) WPMS: Write Port Master Slave latch.

Flip-Flop [9] used in the Sun UltraSPARC-III (USPARC, Fig. 3b), The Single Ended Skew Tolerant Flip-Flop (STFF-SE, Fig. 3d)[6], the Implicitly Pulsed Flip-Flop with half-push-pull latch (IPP, Figure 3a) [8] and the Transmission Gate Pulsed-Latch (TGPL, Figure 3c) [7]. STFFSE and IPP are based on the Sdff dynamic structure, however STFFSE significantly improves the speed of the first stage and IPP improves energy by increasing driving capability of the second stage. The original TGPL had to be modified to fit in this comparison by adding the inverter from the input D to the pass gate in order to achieve sufficient input and output driving capability, otherwise impossible with our setup.

CSEs targeted for low power operation designs are typically static structures since they require robustness of operation under all process and system variations. The common static structures are: the Master-Slave (MS) latch used in the PowerPC 603 (TGMS, Fig. 3f)[10] and commonly referred as a low power CSE [1, 3, 8]. We also included the Modified C²MOS Master-Slave latch (C²MOS, Fig. 3e) [1] and the Write Port Master-Slave latch (WPMS, Fig. 3g) [13].

3.4 Design Space Assumptions

As can be seen in Fig. 3, the number of transistors of a single topology varies from 18 to 32 transistors. However, a good part of these transistors are non-critical for the delay and must remain minimum size (shown as * in Fig. 3) for minimum energy consumption. Hence, the number of transistors that actually matter for the purpose of the extraction of the ED curve as shown in Fig. 1 is limited, often in the order of 5 to 10 transistors. Furthermore, transistor width variations are discrete and increments of the minimum size grid, which is sufficient in term of accuracy for our purpose. On top of this limitation, the lower bound for some transistors is not the technology minimum width for functionality reasons and the upper bounds are limited by the size of the output load of the CSE. Consequently, the number of possible transistor sizing combinations is in the order of a few thousands depending on the topology. Modern desktop computers and scripting languages combined with Hspice can easily handle such task in a few hours. By keeping the design solutions that achieve the lowest energy for a given delay, the extraction of a complete set of ED efficient curves per topology is possible as shown in Fig. 1.

4 Energy-Delay Curves Under Process Variations

From a practical stand point, the Energy-Delay results given by the ED curves simulated in the typical corner as shown in Fig. 1 can be misleading since they do not account for process variations. Dao et al. [14] show process corner variations and the corresponding worst cases for a single sizing solution per topology. This work extends the analysis in [14] to each design point of the ED curve, as shown in Fig. 1. The worst case delay and the worst case energy are necessary for high yield CSE design. Fast paths hazards should also be considered during implementations and we assume padding tools guarantee to cover hold times and clocking uncertainties at the same yield. In this work we assume no variability between the transistors of a single design. If transistor-to-transistor variations are taken in account, the optimization method as proposed by Patil et al. [12] has to be included as well. Effectively,

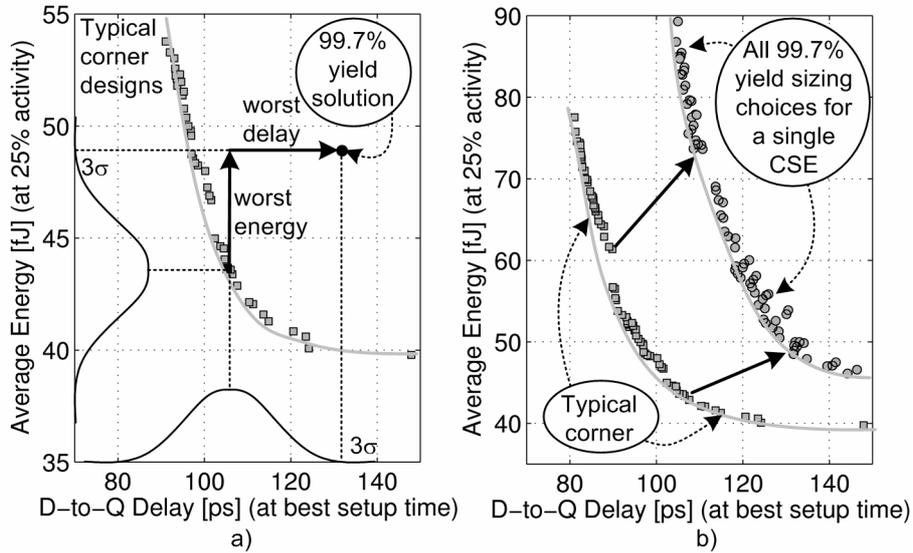


Fig. 4. Energy-Delay curves under process variations: a) Behavior of a single point for a 99.7% yield limit, b) Behavior of the energy efficient characteristic for a 99.7% yield limit

process variations shift the ED curves to higher energy and worse delay than the typical corner accordingly to the desired yield level in both energy and delay. This concept is shown for a single design in Fig. 4a: All of the designs at the typical corner are at the top of the distribution in the typical corner. If the process varies towards a faster corner or higher leakage corner, the energy increases. Similarly, if the process varies towards a slow corner the delay increases. Eventually, as we hit the desired yield (99.7% as example in Fig. 4a) in both energy and delay, the worst ED performance for that yield level is (48fJ ; 132ps) rather than (44fJ ; 105ps) at typical corner. To achieve the desired yield, the design must satisfy the new constraints based on the worst case delay and energy. This concept can be applied to all points of the energy efficient characteristic, thus obtaining the 99.7% yield ED-curves, shown in Fig. 4b.

5 High Yield and Energy Efficient CSE Designs

The purpose of this section is to show the results of an Energy-Delay space analysis for a set of CSEs under specific system constraints and to see how the results translate into high yield design space. For a complete ED space analysis, other system constraints variations such as output load and supply voltage must also be included in order to provide sufficient data for a system optimization [15].

5.1 High Performance CSEs

Fig. 5 shows the results of the ED analysis for the high performance CSEs. The results consist of the composite curve of the best sizings and topologies for the fixed

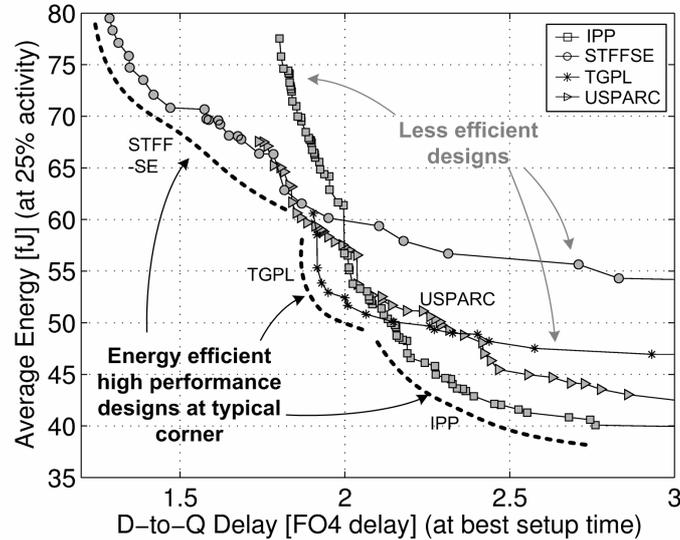


Fig. 5. Energy Efficient High Performance CSEs, Initial Comparison of the various topologies in the typical corner

input and output capacitance (Fig. 5). The results indicate that a subset of the IPP, TGPL and STFFSE ED characteristics constitute the best solutions, depending on the target delay. At $2.1FO_4$ delays and above the IPP achieve best energy efficiency and below $1.9FO_4$ the STFFSE achieve best energy efficiency. In between, there is a narrow section around $2FO_4$ in which the TGPL provides lowest energy designs. Although the USPARC flip-flop is close to the IPP and TGPL in wide range of the delay targets, in no sizing configuration it is the optimum CSE choice. It should be noted that if a smaller load is chosen in the setup (Fig. 2a), the inverter I6 (Fig. 3d) may be removed, improving the TGPL design further, and allowing TGPL to occupy wider range of the composite energy-efficient characteristic.

Fig. 6 shows the energy efficient composite characteristics extracted from Fig. 5 as well as the energy efficient characteristic for high yield, obtained as described in section 4. Designing for high yield shifts the ED curves consistently with an average of a 13% penalty in energy and a 30% penalty in delay for the STFF-SE, IPP and USPARC topologies. However, the TGPL performs worse than other studied CSE in terms of delay with a 48% penalty when process variations are taken into account. The reason for this discrepancy is the principle of operation of the TGPL. This structure relies on the explicit clock pulse to drive the pass gate (M1&M2 in Fig. 3d). Due to lower driving capability of the NAND gate N1 in Fig.3d, the pulse generator in some sizing configurations is not capable to produce full-swing clock pulse height, which further reduces the speed of the TGPL. In order to generate full-swing pulse, larger number of inverters in the pulse generator is needed. However, increasing the width of the pulse has adverse effects on the energy and on the hold time in the fast process corner.

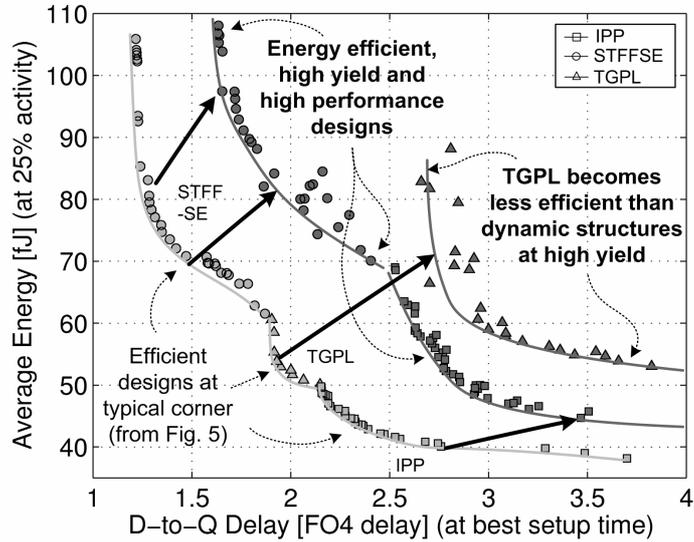


Fig. 6. Impact of high yield design (99.7%) on the energy efficient high performance CSEs

5.2 Low-Power CSEs

Static master-slave latches typically used in low power systems behave much differently than high performance topologies in term of ED performance versus sizing.

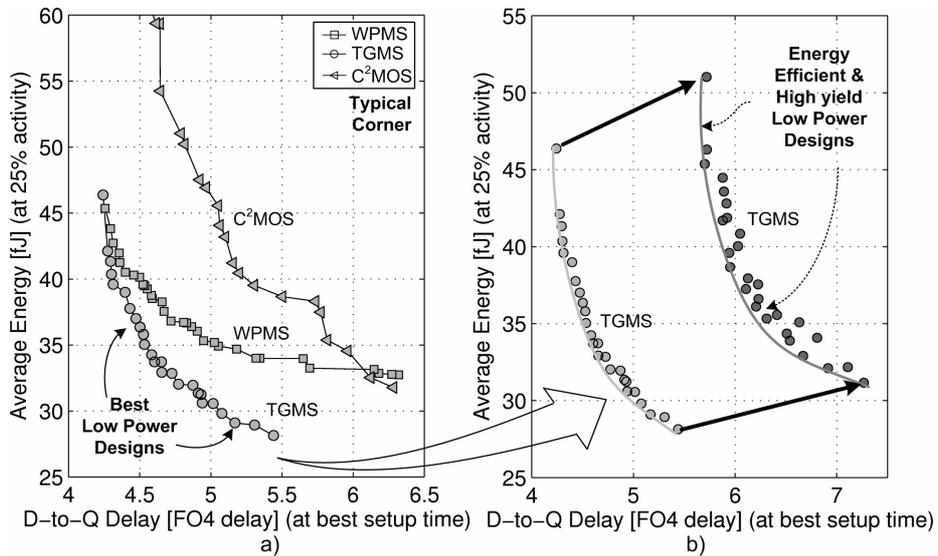


Fig. 7. Energy Efficient Low Power CSEs: a) Comparison of the various topologies in the typical corner, b) Impact of high yield design (99.7%) on the energy efficient low power CSEs (TGMS only)

Because the critical path from D to Q (or Qb) is similar to a chain of inverters, the ED performance is dependent on the gain specification. However, the slope of the energy efficient characteristic is dependent on the topology. For example, as shown in Fig. 7a, the energy of C²MOS MS latch increases rapidly as we move towards faster designs. This is due to the clocked transistors (M2-M3-M6-M7 in Fig. 3e), which must be large to maintain drive strength because they are stacked with the data transistors (M1-M4-M5-M8 in Fig. 3e). In the TGMS and the WPMS the “inverter – pass transistor” combination decouple the datapath inverters from the clock, hence allowing a more efficient distribution of the gain and yielding lower energy for faster designs than the C²MOS MS Latch. Fig. 7a reveals that the TGMS provides best ED results versus the WPMS and the C²MOS in all cases for the setup shown in Fig. 2b. The impact of the process variations is shown in Fig. 7b and represents a consistent 30% overhead in delay and 10% overhead in energy for all three master-slave designs.

6 Conclusions

This work presents the impact of process variations on the choice and design of the CSEs. We show how the boundaries in which various CSEs are the most energy efficient topologies change when the yield is taken into account. For single-ended high performance CSEs, the STFFSE, the TGPL and the IPP perform best at typical corner and only STFFSE and IPP remain efficient for high yield design. For low power designs the transmission gate master-slave latch performs best in typical corner, and it remains best for high yield design. This work reveals the impact of the process corner to the Energy-Delay characteristics for each energy efficient CSE.

Acknowledgments

The authors would like to thank B. Zeydel for his suggestions on system design. They are thankful for the support provided by the Semiconductor Research Corporation grants and Fujitsu Ltd.

References

1. V. Stojanovic and V. Oklobdzija, “Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems”, IEEE JSSC, vol. 34, (no. 4), April 1999, p. 536-48.
2. V. Zyuban, “Optimization of scannable latches for low energy”, IEEE Transactions on VLSI, Vol.11, Issue 5, Oct. 2003 Page(s):778-788
3. V. G. Oklobdzija, V. M. Stojanovic, D. M. Markovic, N. M. Nedovic, “Digital System Clocking”, January 2003, Wiley-IEEE Press
4. V. Stojanovic, V. G. Oklobdzija, "FLIP-FLOP" US Patent No. 6,232,810, Issued: 05/15/2001
5. B. Nikolic, V. Stojanovic, V.G. Oklobdzija, W. Jia, J. Chiu, M. Leung, "Sense Amplifier-Based Flip-Flop", 1999 IEEE ISSCC, San Francisco, February 1999.
6. N. Nedovic, V. G. Oklobdzija, W. W. Walker, “A Clock Skew Absorbing Flip-Flop”, 2003 IEEE ISSCC, San Francisco, Feb. 2003.

7. J. Tschanz, S. Narendra, Z. Chen, S. Borkar, M. Sachdev, V. De, "Comparative delay and energy of single edge-triggered and dual edge-triggered pulsed flip-flops for high-performance micro-processors", ISLPED, 2001. 6-7 Aug. 2001 Page(s):147-152
8. N. Nedovic, "Clocked Storage Elements for High-Performance Applications", PhD dissertation, University of California Davis 2003.
9. F. Klass, "Semi-Dynamic and Dynamic Flip-Flops with Embedded Logic," Symposium on VLSI Circuits, p.108-109, 1998
10. G. Gerosa, S. Gary, C. Dietz, P. Dac, K. Hoover, J. Alvarez, "A 2.2W, 80MHz Superscalar RISC Microprocessor," IEEE JSSC, vol. 29, pp. 1440-1452, Dec. 1994.
11. M. Matsui, H. Hara, Y. Uetani, K. Lee-Sup, T. Nagamatsu, Y. Watanabe, "A 200 MHz 13 mm² 2-D DCT macrocell using sense-amplifier pipeline flip-flop scheme," IEEE JSSC, vol. 29, pp. 1482-1491, Dec. 1994. Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. Int. J. Digit. Libr. 1 (1997) 108-121.
12. D. Patil, S. Yun, S.-J. Kim, A. Cheung, M. Horowitz, S. Boyd, "A new method for design of robust digital circuits", Sixth International Symposium on Quality of Electronic Design, 2005, ISQED 2005. 21-23 March 2005 Page(s):676 - 681.
13. D. Markovic, J. Tschanz, V. De, "Transmission-gate based flip-flop" US Patent 6,642,765, Nov. 2003.
14. H. Dao, K. Nowka, V. Oklobdzija, "Analysis of Clocked Timing Elements for DVS Effects over Process Parameter Variation", Proceedings of the International Symposium on Low Power Electronics and Design, Huntington Beach, California, August 6-7, 2001.
15. H. Dao, B. Zeydel, V. Oklobdzija, "Energy Optimization of Pipelined Digital Systems Using Circuit Sizing and Supply Scaling" IEEE Transactions on VLSI, Volume 14, Issue 2, Feb. 2006 Page(s):122 - 134.