

Timing Yield Enhancement Through Soft Edge Flip-Flop Based Design

Michael Wiecekowski, Young Min Park, Carlos Tokunaga, Dong Woon Kim,
Zhiyong Foo, Dennis Sylvester, David Blaauw
University of Michigan, Ann Arbor, MI

Abstract- The first evaluation of a soft-edge flip-flop is presented as an alternative to useful-skew and latch-based designs for variation compensation in a 16-bit 8-tap FIR filter in 0.13 μ m CMOS. An 11.2% performance improvement was achieved over a standard hard edge data flip-flop (9.2% when post-silicon useful-skew is applied).

I. INTRODUCTION

Increasing process variation in advanced technology nodes due to random sources, such as dopant fluctuation and line-edge roughness, along with complexity in CAD analysis for timing closure, have become limiting factors in the performance and yield of ASIC designs. Traditionally, two-phase latch-based circuits have been used to address these issues due to their relatively high tolerance to variation and their inherent ability to borrow time and pass slack between pipeline stages [1]. While these qualities are attractive from a performance standpoint, latch based systems incur significant power overhead in clock distribution along with an overall increase in design complexity. As an alternative, the combination of hard edge data flip-flops (DFF) with useful clock skew have been used to similarly enhance performance [2]. While effective at mitigating stage delay imbalance and clock tree variation, it is less effective when compensating for individual path delay variation. This is due to the fact that the useful skew can only be applied to groups of flip-flops, among which path delay may vary substantially. This limitation in achievable performance gains is shown in silicon measurements later in this work.

To address timing yield issues without the overhead and penalties seen in other techniques, we have chosen to investigate a soft edge flip-flop (SFF) that maintains synchronization at a clock edge, but also has a transparency window, or “softness”, around it. This flip-flop offers similar variation tolerance to that seen in latch based designs along with performance improvements due to time borrowing. Furthermore, the soft flip-flop provides these gains with minimal overhead in complexity, area, and power.

II. BACKGROUND

It is well known that introducing a transparency window around the clock edge provides benefits in performance by reducing sensitivity to clock skew and jitter and by allowing some degree of time borrowing between pipelined stages [3]. As a demonstration of this effect, we simulated a simple two-stage pipeline composed of balanced inverter chains with an equivalent 10 FO4 stage delay in a commercial 45 nm technology using the setup shown in Fig. 1.

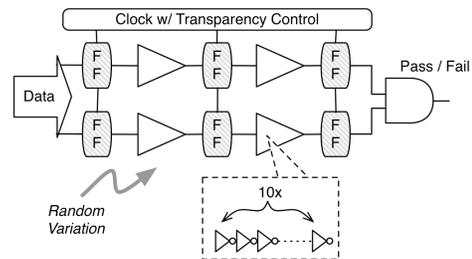


Fig. 1 - Simulation setup for testing transparency windows

Monte Carlo analysis was performed considering only random mismatch among all of the transistors in the pipeline. This mismatch induces delay imbalance among the stages, forcing one path to dictate the maximum attainable clock frequency [4]. As seen in Fig. 2, adding a small window of transparency around the clock edge directly mitigates this effect by allowing some degree of time borrowing, and in turn providing an improvement in overall performance. This improvement would be greater once global variations, clock skew, and clock jitter were considered.

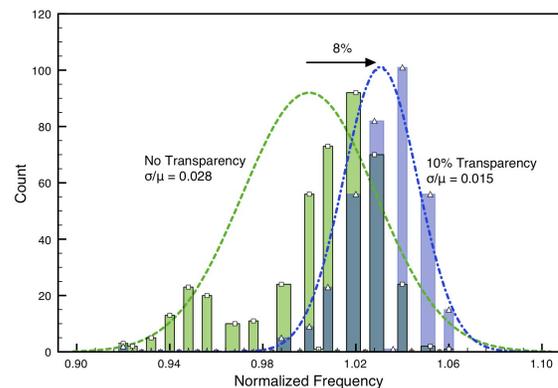


Fig. 2 – The effect of small transparency windows at nominal supply.

A soft edge flip-flop implementation to create such a transparency window has been previously presented in [7], but its application space was focused on time borrowing in on-chip interconnects. Other implementations with similar transparency windows have been based on pulsed latches or hybrid latch-flip-flops [3,6-10]. These designs require the generation of a precisely timed pulse that is triggered from the arriving clock edge. Robust generation of this pulse under process and environmental variation is becoming increasingly difficult and has led to limited use of this technique in only the most timing critical circuits. The transparency window of the

design presented in this work makes use of a standard master-slave flip-flop with a delay element driving the clock from the slave stage to the master stage. This delay element can be implemented using a variety of circuits and is inherently more tolerant to variation than a pulse generator. We discuss our proposed implementation in the following section.

III. SOFT EDGE FLIP-FLOPS

In order to measure the effectiveness of the proposed soft edge technique, a common data flip-flop (DFF) design was used allowing us to mask out any circuit specific effects and to specifically isolate the influence of the transparency window. As shown in Fig. 3, a traditional master-slave flip-flop was modified to perform as a soft flip-flop (SFF) by separating the clocks to the master and slave latching stages.

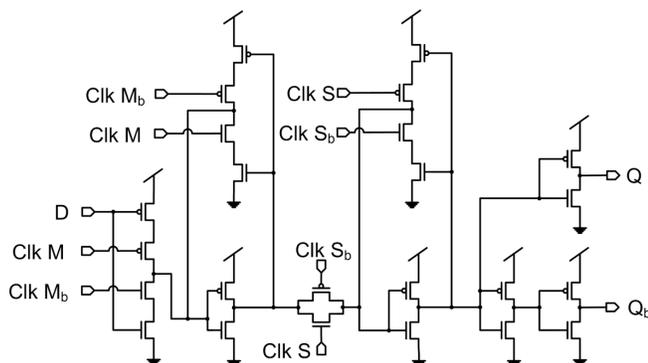


Fig. 3 – Soft edge flip-flop schematic.

To generate a transparency window and create the proposed soft clocking edge, the master stage clock is delayed with respect to the slave stage. Since the master stage is transparent low and the slave stage is transparent high, this delay results in a short window each cycle where both stages are transparent as shown in Fig. 4.

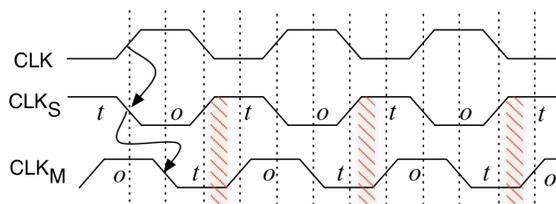


Fig. 4 - Delaying the master clock (CLK_M) relative to the slave clock (CLK_S) creates a window of transparency (t = transparent, o = opaque).

As discussed in Section II, this transparency window provides a direct improvement in overall performance. Creating this window by delaying the master stage clock is much simpler and more robust than generating a local timing pulse as in pulsed registers. This follows from the fact that even under extreme variation where the delay approaches zero, the soft flip-flop converges to the same functionality as a standard data flip-flop. This is in contrast to pulsed registers where a reduction in the pulse width can result in functional failure. In this work, we adopted a simple approach to generating the master stage delay that allowed flexibility in

choosing the size of the transparency window, for the purpose of experimentation. As shown in Fig. 5, tunability was implemented by using a series of different length inverter chains and selecting one as the master delay element using a scan controlled multiplexer.

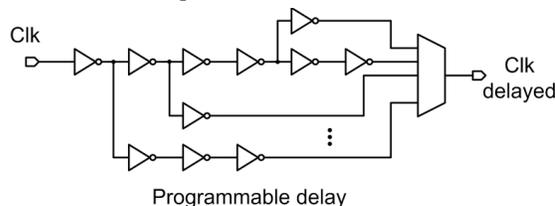


Fig. 5 - Soft flip-flop delay generator.

This delay generator allowed tuning of the transparency window in 10 ps steps from 0 to 250 ps. In a practical implementation, the flip-flop and the delay generator would be specifically designed and optimized for a particular performance target. In addition, it is important to point out that the power overhead of the delay generated can be easily amortized through sharing among groups of flip-flops. Similar techniques are used to share pulse generators in other methodologies and have proven effective [10].

IV. EXPERIMENTAL SETUP

An experimental test chip was designed containing one core with a 16-bit 8-tap FIR filter and a second core with only inverter chains. To determine the effects of the proposed soft edge technique, the pipeline of each core was segmented using both soft flip-flops and standard data flip-flops. This was accomplished using the scheme shown in Fig. 6 where data is fed to an SFF and a DFF simultaneously and their outputs are multiplexed to the succeeding stage of combinational logic. In addition, the aforementioned delay generator was included and shared among all SFFs in each stage bank. The area overhead of this arrangement was only 7.2% and would be dramatically lower given a more intelligent delay generation scheme. A second delay generator, also shown in Fig. 6, was included to add tunable useful skew to each bank.

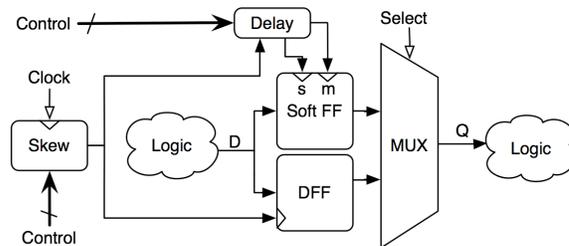


Fig. 6 - Method for selecting between SFF and DFF.

The four-stage FIR filter was designed as shown in Fig. 7 with particular attention paid to balancing all of the pipeline stages. The multipliers were divided into three stages and the final adder took only a single stage. The launching and receiving synchronization elements were chosen as hard edge DFFs in order to close any time borrowing that might occur due to transparency. In order to directly compare the

performance of the SFF to a traditional DFF under the same process and environmental variations, each synchronization element contained both flip-flops and received the same clock from the local distribution network.

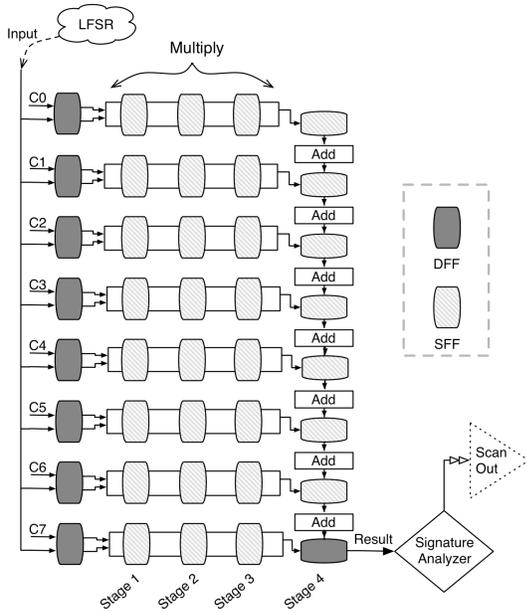


Fig. 7 – Four-stage FIR filter implementation.

The arrangement of the second test core containing four stages of only inverter chains is shown in Fig. 8. Just as in the FIR filter, SFF and DFF elements were included at each stage boundary with DFF elements on both ends to provide time borrowing closure.

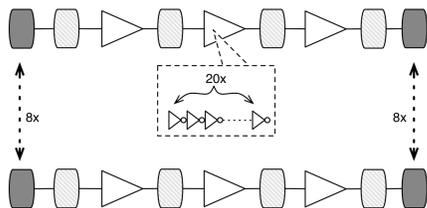


Fig. 8 - Four stages of balanced inverter chains.

Each stage was exactly balanced in the nominal case and contained a chain of inverters equaling a 20 FO4 stage delay. This structure was chosen as a test bench for the proposed soft edge technique at reduced supply voltages. Since the relative sensitivity to the threshold voltage increased at lower supplies, any local mismatch due to random sources, such as random dopant fluctuations (RDF), would express themselves as imbalance in the inverter path delays. As a result, the effect of the transparency window on random variation induced mismatch could be isolated from its systematic counterpart and quantified with respect to performance.

V. MEASURED RESULTS

We fabricated the described FIR filter and balanced inverter pipeline in a 0.13 μm CMOS technology with areas of 0.48 mm^2 and 0.1 mm^2 , respectively as shown in Fig. 9.



Fig. 9 - 0.13 μm test chip micrograph.

Performance measurements at nominal supply voltage of the FIR (1.2V) for 33 die showed an 11.7% improvement in the maximum operating frequency of the SFF circuits when compared to the DFF, and 9.2% improvement when optimal post-silicon skew was applied as seen in Fig. 10.

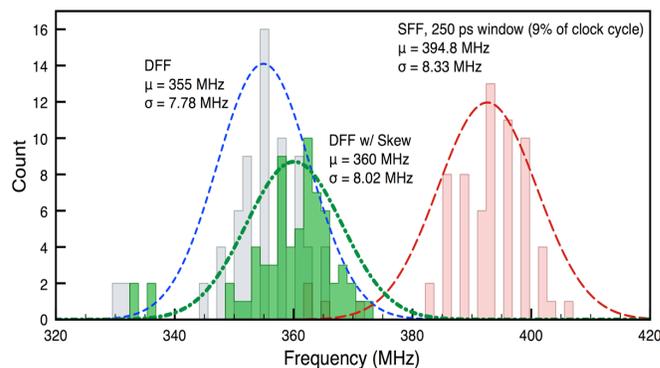


Fig. 10 - Test chip FIR filter frequency distributions for DFF, DFF with useful skew, and SFF.

In addition, all of the FIR filters were measured as a function of transparency window size using the tunable delay generators in each bank. Fig. 11 shows that an increasing window size results in a proportional increase in performance and saturates when the window size approaches 10% of the cycle time.

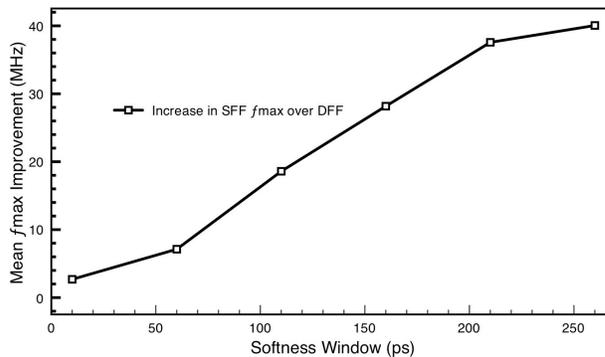


Fig. 11 - Frequency improvement of SFF over DFF FIR filter as a function of softness.

As the window sized is increased, a tradeoff occurs due to hold time constraints becoming more stringent. In addition,

the larger delay introduced to the master clock will lead to higher power consumption. It is therefore desirable to use the minimum amount of softness necessary to achieve a particular performance target. Since separate delay generators were used for each bank of the FIR filter, we could determine the effects of softness to each stage of the pipeline independently. As seen in Fig. 12, the performance improvement of each stage saturates at different values of softness. Therefore, choosing the smallest window size that results in performance saturation will minimize hold time constraints and reduce overall system power. This window size assignment can be performed at design time, and a CAD algorithm for such softness assignment was recently proposed in [13] and showed a total power overhead of 2% on average with comparable performance gains as experimentally shown in this work. Alternatively, window sizes could be dynamically programmed during post-silicon testing.

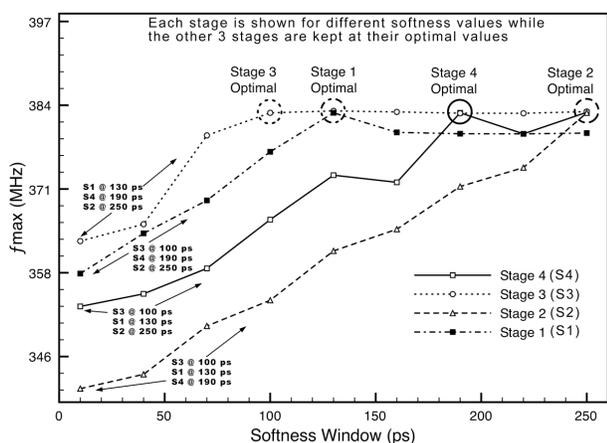


Fig. 12 - Optimization of softness windows for each FIR stage

As explained in Section IV, low voltage operation of pipelined circuits incurs an additional performance penalty due to the heightened sensitivity to threshold voltage variation. The end result is greater imbalance in the pipeline stages, translating to larger timing margins and design area [12]. The proposed soft edge technique helps to reclaim these losses by reducing the sensitivity of the system frequency to critical path delay variation. This improvement was measured for the balanced inverter core over a typical supply voltage scaling range, and is demonstrated in Fig. 13. At higher voltages near the nominal supply of 1.2V, softness has negligible effect on performance since the stages are perfectly balanced. As the voltage is lowered toward the threshold voltage, soft edge clocking reclaims over 10% of the performance lost to variation-induced stage delay imbalance.

VI. CONCLUSION

Soft edge clocking based on latch transparency windows is a well-understood technique that has been shown in this work to provide significant gains in performance for modern process technologies plagued by random sources of variation. Since high performance pulsed techniques are intolerant to variation, and hence unsuitable for variation compensation, we

have demonstrated a simple soft edge flip-flop technique as a viable alternative. We have demonstrated in silicon that this

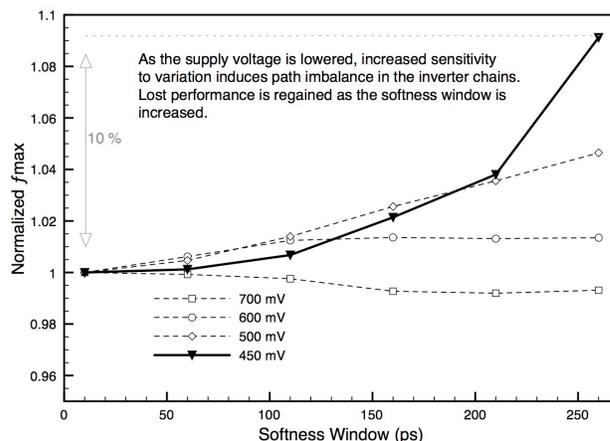


Fig. 13 - Frequency improvement of SFF over DFF inverter chains as a function of softness at low voltage

SFF can achieve gains in performance over 10% compared to traditional designs even under harsh low-voltage conditions. These gains come at minimal area and power penalties and can readily be incorporated into a standard design flow.

REFERENCES

- [1] D. Harris, *Skew-Tolerant Circuit Design*, Stanford University, 1999.
- [2] V. Nawale and T.W. Chen, "Optimal useful clock skew scheduling in the presence of variations using robust ILP formulations," *International conference on Computer-aided design*, 2006, pp. 27-32.
- [3] H. Partovi et al., "Flow-through latch and edge-triggered flip-flop hybrid elements," *IEEE International Solid-State Circuits Conference*, 1996, pp. 138-139.
- [4] 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 Journal of Solid-State Circuits*, vol. 37, 2002, pp. 183-190.
- [5] K. Bowman et al., "Time-borrowing multi-cycle on-chip interconnects for delay variation tolerance," *International symposium on Low power electronics and design*, 2006, pp. 79-84.
- [6] A. Scherer et al., "An out-of-order three-way superscalar multimedia floating-point unit," *IEEE International Solid-State Circuits Conference*, 1999, pp. 94-95.
- [7] J. Tschanz et al., "Comparative delay and energy of single edge-triggered & dual edge-triggered pulsed flip-flops for high-performance microprocessors," *Proceedings of the 2001 international symposium on Low power electronics and design*, 2001, pp. 147-152.
- [8] F. Klass et al., "A new family of semidynamic and dynamic flip-flops with embedded logic for high-performance processors," *IEEE Journal of Solid-State Circuits*, vol. 34, 1999, pp. 712-716.
- [9] H. Partovi, "Clocked Storage Elements," *Design of High-Performance Microprocessor Circuits*, A.P. Chandrakasan, W.J. Bowhill, and F. Fox, eds., Wiley-IEEE Press, 2000.
- [10] D. Krueger, E. Francom, and J. Langsdorf, "Circuit Design for Voltage Scaling and SER Immunity on a Quad-Core Itanium Processor," *IEEE International Solid-State Circuits Conference*, 2008, pp. 94-95.
- [11] V. Joshi, D. Blaauw, and D. Sylvester, "Soft-Edge Flip-Flops for Improved Timing Yield: Design and Optimization," *International Conference on Computer-Aided Design*, 2007, pp. 667-673.
- [12] A. Datta et al., "A Statistical Approach to Area-Constrained Yield Enhancement for Pipelined Circuits under Parameter Variations," *Asian Test Symposium*, 2005, pp. 170-175.