# ERROR-RESILIENT SEQUENTIAL CELLS WITH SUCCESSIVE TIME BORROWING FOR STOCHASTIC COMPUTING

Wei-Chang Liu<sup>1</sup> Ching-Da Chan<sup>1</sup> Shuo-An Huang<sup>1</sup> Chi-Wei Lo<sup>1</sup> Chia-Hsiang Yang<sup>2</sup> Shyh-Jye Jou<sup>1</sup>

<sup>1</sup>Department of Electronics Engineering, National Chiao Tung University, Hsinchu, Taiwan <sup>2</sup>Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan

# ABSTRACT

This paper presents error-resilient sequential building blocks with time-borrowing capability without extra latches and generated clocks. The circuits are able to recover the timing errors caused by PVT variations and/or over-voltage scaling by up to half a cycle. Unlike prior works, the timing errors can be recovered dynamically through successive time borrowing without stalled cycles, retaining a constant throughput. The circuit structure can be applied to both ASICs and microprocessors. The proposed sequential cells are highly compatible with current cell-based IC design flow, for both feedforward and feedback datapaths. As a proof of concept, a design with key DSP building blocks has been verified. The results show that the performance of the DSP modules is improved by 13-15% in the worst-case operation condition, yielding a promising solution for stochastic computing under an unreliable operation condition.

*Index Terms*— Error-resilient circuit, sequential cell, time borrowing, stochastic computing.

## 1. INTRODUCTION

Stochastic computing is associated with resilience against errors. At the circuit level, errors are often caused by timing variations in a statistic sense. Advanced CMOS technology is usually adopted to meet the design specifications for high-speed signal processing at. However, advanced technology suffers from severe variations. Fig. 1 shows the simulated operating frequency versus process, voltage, and temperature (PVT) variations for a 15-stage inverter chain in a 65nm CMOS process. Three corner processes (fast-fast (FF), typical-typical (TT), and slow-slow (SS)) are considered here. In the FF corner with a 1.1V supply voltage, temperature variation (0°C to 125°C) can cause an up to 8.65% fluctuation in operating frequency. Among the process corners, the frequency fluctuation can be up to 47.82% at a 1.1V supply voltage. The frequency fluctuation becomes even more severe when taking the voltage variation into consideration. Resizing of logic gates is usually needed to compensate the speed degradation in the current chip design environment — at the expense of increased silicon area [1, 2].



**Fig. 1**. Operating frequency vs. supply voltage with respect to process, voltage, and temperature variations.

The error statistics can be leveraged to achieve robust computation [3]. An algorithmic noise-tolerance (ANT) block demonstrates the error-resilient capability by incorporating a main processing element and an estimator. With the aid of a low-complexity estimator, the estimated output with a smaller error can be used in case that the output of the main processing element largely deviates from the correct one. The *N*-modular redundancy (NMR) and soft NMR use majority vote to determine the correct output from *N* replicated processing elements. The *N*-fold hardware and power cost limits the applicability of this method. Especially, ANT and NMR do not work well when global variations exist.

Actually, the worst-case timing condition rarely occurs during the computation procedure [4]. A timing-error recovery scheme is advantageous to combat the worst operating conditions that rarely occur. The relaxed timing constraint can also be leveraged to reduce the supply voltage for further power reduction. Dynamic voltage-frequency scaling (DVFS) and body-biasing schemes [4–8] have been presented to overcome the timing drift. However, these methods need off-situ processing to determine the scaled voltage and body biasing voltage.

Regarding the in-situ approaches for stochastic computation, the Razor family [9–11] focuses on circuit-level timing error compensation. The timing errors are corrected by stalling the system or reducing the clock frequency by half.



Fig. 2. Time borrowing for successive timing errors.

A transition detector with time borrowing (TDTB) and double sampling with time borrowing (DSTB) schemes are presented in [12]. For these approaches, the hold time of the previous-stage circuits needs to be carefully tackled to prevent the short-path signal from being sampled. It needs not only a skewed clock but also extra delay buffers to fix the hold-time issues. Either reducing the clock frequency by half or stalling the system reduces the throughput and the result-ing latency is varied. Due to aforementioned issues, these approaches are only applicable to generic processors. In addition, the Razor family cannot support the logic fabric with a feedback structure.

Although several circuit-level timing-error resilient approaches have been proposed, they are not supported by the current automatic tools due to the deterministic nature of static timing analysis (STA). Especially, most of the approaches do not guarantee a fixed throughput and are not applicable to stream-based designs. In this paper, a timing-error resilient sequential cell compatible with the current cell-based design flow is proposed. The proposed sequential cell is a time borrowing master-slave flip-flop (TBMSFF) that can recover the timing errors by leveraging successive time borrowing rather than stalling/adjusting the clock period.

## 2. TIME BORROWING MASTER-SLAVE FLIP-FLOP

Logic delay is always the main concern in the sequential circuit design. Fig. 2 shows an example of timing relationship between the pipeline stages. In a synchronous design, the computation time of each stage should be less than a clock period so that the correct data can be sampled at the next rising clock edge. In the typical-case (TC) condition, the timing requirement is satisfied in most cases. In the worst-case (WC) condition, however, the computation time in some stages may become longer than that in the TC condition. Incorrect outputs occur if any of stages fails to meet the timing constraint.

The proposed TBMSFF features asynchronous and delayed sampling characteristics. Two sampling schemes are proposed for the TBMSFF. One is sampling at the rising edge and the other is sampling while the input stage completes its computation. As shown in Fig. 2, once the timing error oc-



(c) Type-III TBMSFF

Fig. 3. Architectures of TBMSFF for inter-stage registers.

curs, the next stage will wait until the current stage completes its tasks and then trigger the next stage. The advantage of this scheme is that the timing error can be successively handled instead of being recovered whenever it occurs. The processing time can be borrowed successfully until the task is completed.

Conventionally, an master-slave flip-flop can be decomposed into a master latch (ML) and a slave latch (SL), and it has two operation modes: opaque-transparent mode and transparent-opaque mode. In this work, three types of TBMSFF are proposed. The type-I TBMSFF is designed for the datapaths between registers in which time borrowing is allowed. The type-II and type-III TBMSFFs are for primary inputs and primary outputs, respectively. The circuit details are depicted in Fig. 3. For the type-I TBMSFF, there are two more operation modes: opaque-opaque mode and transparent-transparent mode. The two additional operation modes are realized by adding an input control signal ctrl\_i with an AND gate in the clock path of the SL and an output control signal ctrl\_o with an NAND gate in the clock path of the ML, respectively. Time borrowing is executed by closing the ML and SL at the same time or by opening the ML and SL at the same time in the positive-level clock period. For the type-II TBMSFF, it has only one additional opaque-opaque mode by adding an input control signal *ctrl\_i* with an AND gate in the clock path of the SL. In contrast, the type-III TBMSFF has an additional output control signal *ctrl\_o* with a NAND gate in the clock path of the ML to realize the additional transparent-transparent mode.

Fig. 4 illustrates a pipeline stage with a type-II TBMSFF and a type-III TBMSFF. Assume the relative outputs of the previous stage have no timing errors. If the previous stage



Fig. 4. Timing diagram of the proposed TBMSFF.

outputs D have timing errors, then the type-II TBMSFF should be replaced by type-I TBMSFF. If the next stage inputs Q are in the critical path and timing errors occur, the type-III TBMSFF should be replaced with a type-I TBMSFF to latch the inputs.

The input control signal  $ctrl_{-i}$  and the output control signal  $ctrl_{-o}$  are generated by the TD circuit to detect the computation status. When the computation is done, there are no transitions and the outputs of TD,  $ctrl_{-i}$  and  $ctrl_{-o}$ , rise to 1. In the situation without timing errors, the SL and ML are controlled by clock as a normal SL and ML in the MSFF.

When timing errors occur, the inputs in critical paths need to be held until the computation is done and the outputs need to use time borrowing to pass the correct data to next stage. Therefore, the SL is controlled by  $ctrl_i$  and keeps closed until the transition is done. The latched inputs prevent the incorrect outputs due to the short-path problem so that the hold time violations can be prevented. The output port Q holds the previous value of input D in the opaque-opaque mode and keeps the data to next stage when the timing errors occur. When the transition of next stage is done, the operation mode changes to opaque-transparent as a normal MSFF. Because D is updated in the internal node n at the clock rising edge, Q keeps the latest value of D while the SL is opened in the opaque-transparent mode. Therefore, the next stage has the latest data of D after time borrowing. Meanwhile, the ML will be controlled by ctrl\_o and keeps opened until the transition is done. Thus, the output port Q varies with the input signal  $\underline{D}$  in the transparent-transparent mode and stores the latest data from previous stage while the timing errors occur. When the transition of previous stage is done, the operation mode will change to opaque-transparent as a normal MSFF.



Fig. 5. Architecture of (a) MAC and (b) CORDIC modules.

Because  $\underline{Q}$  varies with  $\underline{D}$  in the transparent-transparent mode, the internal node  $\underline{n}$  and output port  $\underline{Q}$  have the latest data of  $\underline{D}$  after time borrowing.

The timing diagram is also shown in Fig. 4. In the beginning of cycle 1 and cycle 2, the computation time is less than one clock cycle. In this condition, the TBMSFFs operate as a normal MSFF. In the beginning of cycle 3, the computation of cycle 2 is not completed. Therefore, the *ctrl* signal keeps in the low level. When the output for next stage is ready, the *ctrl* changes to high and the *clk\_m/clk\_s* changes to low/high. Then the TBMSFFs operate as a normal MSFF in clock positive level. In the beginning of cycle 4, the borrowed time is accumulated. However, it is returned in the end and the operation becomes the same as in cycle 3. This clearly demonstrate that the proposed TBMSFFs can bear the successive timing errors.

## 3. EXPERIMENTAL VERIFICATION

In order to verify the functionality of the proposed TBMSFF, a test chip that includes two key DSP modules is designed. One DSP module is a multiply-accumulate (MAC) unit, which is used to verify the functionality of the TBMSFF in a feedback loop. The other one is a pipelined CORDIC (coordinate rotation digital computer), which is used to verify the functionality in a feedforward datapath with different timing conditions in each stage.

The block diagram of MAC is shown in Fig. 5(a). The input register is the proposed type-III TBMSFF. Since the output register is also the input register of MAC, it may suffer from the timing violations of the next stage and current stage simultaneously. Therefore, we use the type-I TBMSFF for the output register. An extra pipeline stage is added for time borrowing of the output stage. The block diagram of the CORDIC structure is shown in Fig. 5(b). There are 20 CORDICs cascaded with four pipeline stages. In this design, the type-III TBMSFF is adopted for the input of second pipeline stage. The registers between second and third pipeline stages are the type-I TBMSFF. The type-II TBMSFF



Fig. 6. Layout and design details of the test design.

 Table 1. Performance of the proposed test design

	MAC	CORDIC			
Silicon Area (µm <sup>2</sup> )					
Combinational Logic	1557.26	8549.33			
TD	538.55	394.08			
DFF & ISDFF	2988.57	4485.40			
TBMSFF	1090.15	1846.20			
Clock Speed (MHz)					
Conventional DFF	109.89	68.03			
ISDFF	125.00	78.13			
TBMSFF	126.58	77.52			

is used at the output registers of third pipeline stage.

The TBMSFF cells are modified from the DFF cells from the standard cell library. Since latches are not fully supported in the current STA design flow, the TBMSFF is treated as a pseudo-DFF with extra control signals to facilitate the timing analysis. For implementation, only the pipeline registers in the critical path are replaced. An interleaving sampled FF (ISFF) is used for comparison. The ISFFs are triggered by different edges of the clock alternatively, creating a timing margin of half cycle for time borrowing in each stage. The layout of the test design is shown in Fig. 6. The chip is designed with the aid of automatic tools in a 40nm CMOS technology. For each DSP module, the layout is created in a bottom-up manner. The MAC and CORDIC units are created first along with their own power rings. Table 1 summarizes the implementation result. The area overhead of the proposed TBMSFFs is 20-38%, depending on the type of the TBMSFF. The area overhead can be further reduced through tailored full-custom design. The proposed TBMSFFs improve the timing performance by 13-15% when the clock duty cycle is set to 50% in the worst-case condition.

Table 2 shows the comparison of prior works and the proposed error-resilient scheme. The proposed design has the lowest hardware overhead since only one AND gate and/or

	Razor [9],	Razor	TDTB,	This
	Razor II [10]	Lite [11]	DSTB [12]	work
Hardware	High	Middle	High	Low
overhead				
Timing	Stall	FS	TB & FS	TB
recovery				
Single-cycle	No	Yes	Yes	Ves
correction	110			105
Throughput	Variable	Variable	Variable	Fixed
EDA tool	Low	Low	Low	High
compatibility	LUW			mgn
Applicability	$\mu$ -processor	$\mu$ -processor	$\mu$ -processor	ASIC
				& $\mu$ -processor

Table 2. Comparison of distinct timing resilient designs

TB: time borrowing, FS: frequency scaling

one NAND gate needs to be added without complex peripherals. The delay buffer insertion for short-path problem can be avoided since the timing error detection is performed before sampling. Since the timing error detection is derived from the asynchronous operation, no generated clock and/or multi-phase clock are required. Because stalled cycles are not required, the timing errors can be tackled within one cycle and the throughput keeps constant. Moreover, the proposed TBMSFFs are highly compatible with the current EDA tool chain since it can be modelled as conventional DFFs with two cascading latches. The effort on customization is therefore minimized. Due to the constant-throughput feature, the proposed TBMSFF cells are applicable to both stream-based applications-specific ICs (ASICs) and micro-processors.

#### 4. CONCLUSIONS

This paper presents timing error-resilient sequential cells with successive, multi-stage time borrowing capability. Three types of TBMSFF are proposed to replace the conventional DFFs in order to combat PVT variations. The resulting sequential circuits feature time borrowing so that the replay mechanism is not required. The proposed TBMSFFs are evaluated extensively for different application scenarios. The area overhead is minimized by adding only one AND/NAND gate in the conventional DFF. It is applicable to both feedforward and feedback datapaths. In addition, extra clocks, including multi-phase clocks, are not needed. The proposed TBMSFF is compatible with current standard cell-based design flow. This allows for integrating the TBMSFF cells into complex DSP systems. The circuit performance is improved by 13-15% in the worst-case operation condition. With the capability against timing errors, the proposed sequential cells can be deployed in applications for stochastic computing.

#### 5. REFERENCES

- J. C. Zhang, "Worst case design of digital integrated circuits," in *IEEE Int. Symp. on Circuits and Syst. (ISCAS)*, Jun. 1994, pp. 153-156.
- [2] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. Design Automation Conf. (DAC)*, Jun. 2003, pp. 338-342.
- [3] N. R. Shanbhag, R. A. Abdallah, R. Kumar and D. L. Jones, "Stochastic computation," in ACM/IEEE Design Automation Conf. (DAC), Jun. 2010, pp. 859-864.
- [4] M. Meijer and J. P. de Gyvez, "Body-Bias-Driven Design Strategy for Area- and Performance-Efficient CMOS Circuits" *IEEE Trans. Very Large Scale Integr.* (VLSI) Syst., vol. 20, pp. 42-51, Jan. 2012.
- [5] M.-E. Hwang and K. Roy, "ABRM: Adaptive -Ratio Modulation for Process-Tolerant Ultradynamic Voltage Scaling," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, pp. 281-290, Feb. 2010.
- [6] J. W. Tschanz, S. Narendra, R. Nair and V. De, "Effectiveness of adaptive supply voltage and body bias for reducing impact of parameter variations in low power and high performance microprocessors," *IEEE J. Solid-State Circuits*, vol. 38, pp. 826-829, May 2003.
- [7] H. Mostafa, M. Anis and M. Elmasry, "On-Chip Process Variations Compensation Using an Analog Adaptive Body Bias (A-ABB)," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, pp. 770-774, Apr. 2012.
- [8] H. Jeon, Y.-B. Kim and M. Choi, "Standby Leakage Power Reduction Technique for Nanoscale CMOS VLSI Systems," *IEEE Trans. Instrum. Meas.*, vol. 59, pp. 1127-1133, May 2010.
- [9] D. Ernst, N. S. Kim, S. Das, S. Pant, R. Rao, T. Pham, C. Ziesler, D. Blaauw, T. Austin, K. Flautner and T. Mudge, "Razor: a low-power pipeline based on circuit-level timing speculation," in *Proc. IEEE/ACM Int. Symp. on Microarchitecture*, Dec. 2003, pp. 7-18.
- [10] D. Blaauw, S. Kalaiselvan, K. Lai, W.-H. Ma, S. Pant, C. Tokunaga, S. Das and D. Bull, "Razor II: In Situ Error Detection and Correction for PVT and SER Tolerance," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2008, pp. 400-401.
- [11] S. Kim, I. Kwon, D. Fick, M. Kim, Y.-P. Chen and D. Sylvester, "Razor-Lite: A Side-Channel Error-Detection Register for Timing-Margin Recovery in 45nm SOI CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb.2013, pp.264-265.

[12] K. A. Bowman, J. W. Tschanz, N. S. Kim, J. C. Lee, C. B. Wilkerson, S.-L. L. Lu, T. Karnik and V. K. De, "Energy-Efficient and Metastability-Immune Resilient Circuits for Dynamic Variation Tolerance," *IEEE J. Solid-State Circuits*, vol. 44, pp. 49-63, Jan. 2009.