8.8 iRazor: 3-Transistor Current-Based Error Detection and Correction in an ARM Cortex-R4 Processor

Yiqun Zhang[1], Mahmood Khayatzadeh[1], Kaiyuan Yang[1], Mehdi Saligane[1], Nathaniel Pinckney[1], Massimo Alioto[2], David Blaauw[1], Dennis Sylvester[1]

[1]University of Michigan, Ann Arbor, MI,
[2]National University of Singapore, Singapore, Singapore

It is well known that technology scaling has led to increasing process/voltage/temperature/aging margins that substantially degrade performance and power in modern processors and SoCs. One approach to address these large timing margins is the use of specialized registers on critical paths that perform error detection and correction (EDAC) [1-5]. While promising, the previously proposed implementations have been limited in several ways. Most notably, they often incur large overheads beyond conventional register designs (e.g., 8-to-44 additional transistors per register). This becomes an obstacle for commercial designs and, hence, there have been no reported implementations of EDAC approaches within substantial commercial processors. Finally, the performance gain from EDAC approaches has not been thoroughly quantified in relation to competing, lower overhead approaches such as frequency binning and canary circuits/critical path monitors [6].

We propose iRazor: a new, low-overhead EDAC technique that employs a three-transistor current-sensing circuit to detect data transitions within a detection window. The proposed iRazor registers are integrated into an industrial-class ARM Cortex-R4 processor with an 8-stage pipeline. Using local detection and clock stalling, the pipeline is halted within one cycle of a detected error, allowing the EDAC technique to be integrated into the processor without requiring rollback or architectural changes. To quantify the benefits of iRazor, it is compared to a separate baseline implementation, as well as performance-binned and canary-enabled versions of the Cortex-R4 core. The iRazor Cortex-R4 operates at 843MHz in 40nm CMOS with 13.6% total area overhead compared to the baseline. This represents performance gains of 30%, 23%, and 17% compared to standard, binned, and canary-equipped R4 versions, respectively.

Similar to [1], the iRazor register uses a latch that flags any data transitions during its transparency window as an error. The data transition is performed with a 3-transistor current sensing circuit (Fig. 8.8.1). In the positive phase of CTL, the virtual rail VVSS is discharged and initialized to 0. In the negative CTL phase (error-detection phase), VVSS floats and will droop up if D changes (VVSS will be pulled high through the tri-state buffer A or feedback inverter B if D rises or falls, respectively). This switches the skewed inverter output, which is flagged as an error. Since the resulting ERR signal is a 1-0-1 pulse, it is captured by a PMOS-based dynamic OR-latch (Fig. 8.8.2). It is then propagated to the Razor timing control block after being OR'ed together with all other error signals within the processor using conventional dynamic OR gates. The Razor timing control skips the clock pulse following the occurrence of an error, providing time for the error in the pipeline to resolve (cycle 3 in Fig. 8.8.2). Following error resolution, the dynamic OR-latches are reset using the i-RESET signal and normal operation resumes. To avoid the need for a pipeline rollback mechanism, the clock pulse immediately subsequent to the error must be eliminated (Fig. 8.8.2). To accomplish this, the error signal must propagate through the OR-latch, three OR stages, the Razor control block, and clock tree to reach the clock tree leaves within the same clock cycle that the error occurs. Fig. 8.8.3 qualitatively depicts these relative delays; dynamic OR gates are employed as their speed is necessary to meet this constraint.

The relative timing of CTL and LCLK (Figs. 8.8.2 and 8.8.3) presents a number of constraints and design trade-offs. The falling edge of CTL marks the start of the detection window and must occur with sufficient delay after the rising edge of LCLK (T_{FR} in Fig. 8.8.3) to allow valid transitions of D to pass through the latch without triggering an error. Increasing T_{FR} reduces the probability of such false-positive errors and also increases time borrowing at the expense of a smaller error detection window. Monte Carlo simulation results illustrating the relationship between the time borrowing window and VVSS increase during normal operation are shown in Fig. 8.8.3. On the other hand, the rising edge of CTL must occur prior to the falling edge of LCLK to enable the corrected input data to be latched into the back-to-back inverter before the transparency window ends. This occurs during the time period T_{BK}, where the footer is re-enabled to restore VVSS. Monte Carlo simulation demonstrates the robustness of VVSS behavior for both normal operation and in case of an error (Fig. 8.8.3). The LCLK and CTL signals, and hence T_{FR} and T_{BK}, are generated by local clock generators. This avoids the need for two clock distribution networks that introduce power overhead and inter-clock mismatch. The wirelength from the skewed inverter output to the dynamic OR-latch plays a critical role in the detection window size. Hence, after initial placement of the baseline design, automated clustering is performed to assign EDAC registers to each local CLK generator / OR-latch pair (Fig. 8.8.4). After this first level of clustering, the remaining levels of the error OR-tree are determined using hierarchical iterations of clustering, followed by a new placement where the original EDAC register locations are frozen. Further iterations of placement/clustering are performed to close timing, as needed.

iRazor is applied to an ARM Cortex-R4 processor in 40nm CMOS. This test chip marks a significantly faster and larger implementation of an EDAC approach than previous implementations (Fig. 8.8.6). The design was completed in a fully automated fashion and required no change to the processor architecture. Fig. 8.8.4 shows the path delay histogram of the baseline and iRazor systems. All paths with <200ps slack (~16.7% of the clock cycle) had their registers replaced with EDAC registers, resulting in 1,115 iRazor registers out of 12,875 total registers in the core logic (8.7%). The total number of gates increased from 917K to 1040K when applying iRazor, mostly due to increased minimum-sized hold-time buffers (Fig. 8.8.4). Total area increased from 0.6195mm^2 to 0.7035mm^2 (13.6% overhead) for the Cortex-R4 including logic, 8KB I- and D-caches, and 12KB memory (excluding PLL).

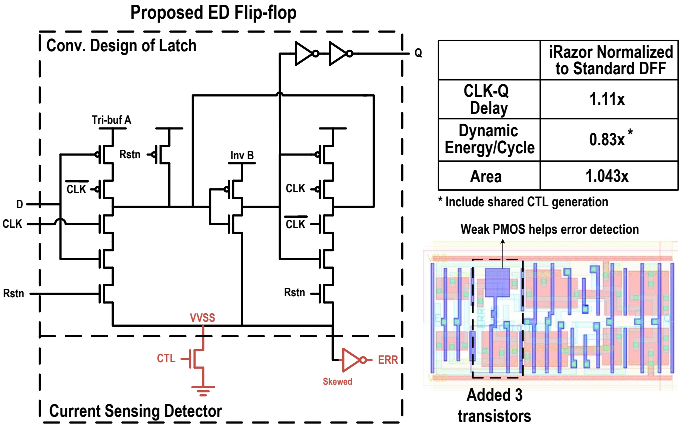
Both baseline and iRazor versions were implemented on the test chip (die photo in Fig. 8.8.7). Furthermore, the baseline processor was equipped with a RO-based canary circuit to provide a comparison between competing methods of reducing performance margins, for the first time. Fig. 8.8.5 shows measured performance across three operating voltages and five possible margining scenarios from 40 baseline dies and 40 iRazor dies, as follows:

– *Worst-case margining*: applied to all die to account for a conventional worst-case PVT condition of 85°C, 10% supply drop, and 3σ process variation.
– *Binned margining*: dies are divided into three performance bins based on process corner and each bin is margined for 85°C, 10% supply drop.
– *Simple canary*: the linear correlation between RO and processor frequencies is calculated and used to set each die's clock frequency. The linear fit is de-rated by 5% voltage margin to account for transient voltage excursions that the canary cannot capture and $2\sigma_{\text{fitting\_error}}$ to address RO-processor mistracking ($\sigma_{\text{fitting\_error}}$ is calculated across die and PVT conditions).
– *T/V specific canary*: A separate linear correlation between the RO and processor frequency is determined for each temperature/voltage condition (requiring on-die sensors). Linear fit is de-rated with 5% voltage margin and $2\sigma_{\text{fitting\_error}}$, but here $\sigma_{\text{fitting\_error}}$ is computed only across die (not V/T), reducing margins.
– *Razor-PoFF*: iRazor implementation operating at the frequency at which errors are initially observed; this provides a conservative 4.4-to-6.9% timing margin compared to the maximum possible iRazor frequency .

Figure 8.8.5 shows that a simple canary approach is about twice as effective as binning. T/V specific canary offers ~15-to-18% performance increase over the margined baseline across 0.6-to-1V, while iRazor shows 30-to-46% performance increase. Fig. 8.8.5 also compares power consumption for fixed frequency (i.e., energy). At 0.6V, iRazor provides roughly 2× better energy efficiency benefit vs. the T/V specific canary approach due to worsened canary mistracking at low voltage operation. Fig. 8.8.6 includes a comparison table of iRazor and past EDAC approaches. This work demonstrates a low-overhead sequential element and a large test chip implementation among prior EDAC works.

Acknowledgements:
The authors thank TSMC University Shuttle Program for chip fabrication.

References:
[1] S. Das et al., "Razor II: In Situ Error Detection and Correction For PVT and SER Tolerance," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 32–48, Jan. 2009.
[2] K.A. Bowman et al., "Energy-Efficient and Metastability-Immune Resilient Circuits for Dynamic Variation Tolerance," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 49–63, Jan. 2009.
[3] K.A. Bowman et al., "A 45nm Resilient Microprocessor Core for Dynamic Variation Tolerance," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 194–208, Jan. 2011.
[4] D. Bull et al., "A Power-Efficient 32 Bit ARM Processor Using Timing-Error Detection and Correction for Transient Error Tolerance and Adaptation to PVT Variation," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 18–31, Jan. 2011.
[5] S. Kim et al., "Razor-lite: A Side-Channel Error-Detection Register for Timing-Margin Recovery in 45nm SOI CMOS," *ISSCC Dig. Tech. Papers*, pp. 264-265, 2013.
[6] J.L. Shin, et al., "The Next-Generation 64b SPARC Core in a T4 SoC Processor," *ISSCC Dig. Tech. Papers*, pp. 60-62, 2012.



	iRazor Normalized to Standard DFF
CLK-Q Delay	1.11x
Dynamic Energy/Cycle	0.83x *
Area	1.043x

* Include shared CTL generation

Figure 8.8.1: iRazor schematic, layout, and table of overhead. Only 3 transistors are added to the library latch design.

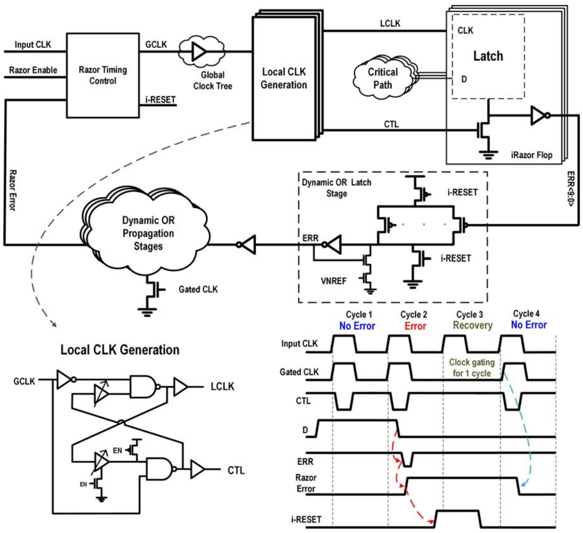


Figure 8.8.2: Diagram of overall structure of EDAC technique and local clock detection circuit. Timing and error detection waveforms and characteristics.

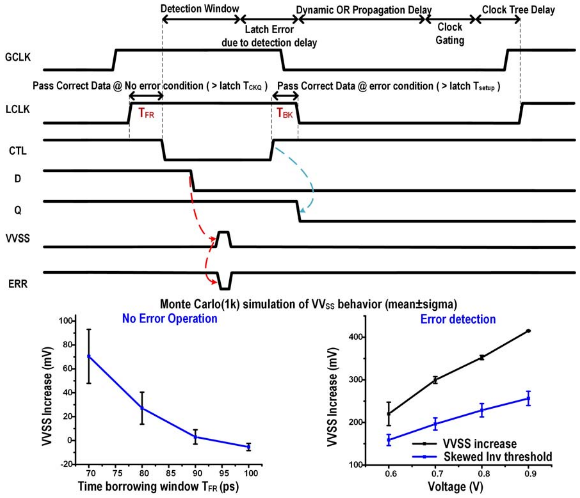


Figure 8.8.3: Conceptual timing diagram for detection scheme and timing constraints for correction scheme. Monte Carlo simulation results of VVSS at no error condition and upon error detection.

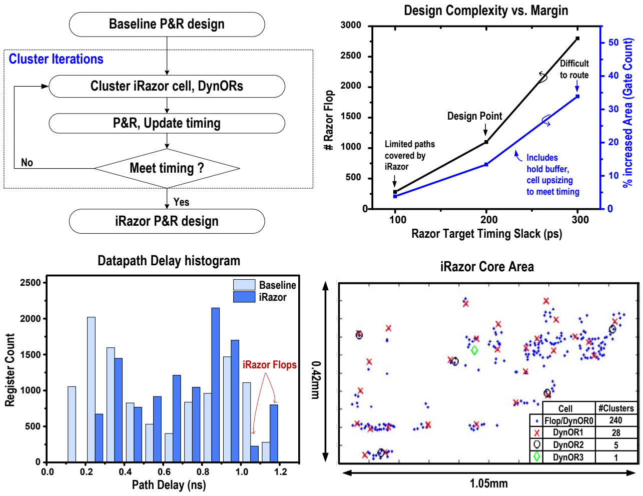


Figure 8.8.4: Architecture-independent procedure for iRazor replacement and clustering; design complexity vs. targeted timing slack of iRazor; path delay histogram; iRazor cluster positions.

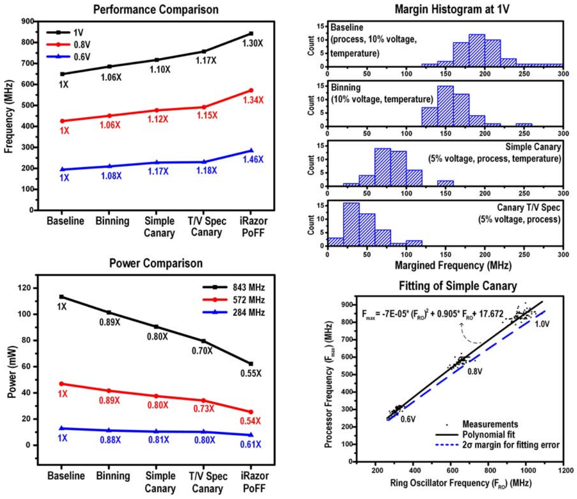


Figure 8.8.5: Performance and power comparison; margin histogram relative to the reference frequency of baseline chip at room temperature; simple canary fitting based on measured data.

		Razor II JSSC'09 [1]	TDTB JSSC'09 [2]	DSTB JSSC'11 [3]	ARM JSSC'11 [4]	Razor-lite ISSCC'13 [5]	iRazor
	Type	Latch	Latch	Latch	Flip-Flop	Flip-Flop	Latch
EDAC Level	Extra # of Transistor (I)	31 (8 Shared)	15	26	28+ delay chain	8	1.46 (II)
	Possible Datapath Metastability	No	No	No	Yes	Yes	No
	FF Power Overhead	28.5%	-9%~ -13% (III)	14%~ 34%	Not Reported	2.7%	12.5%
	FF Area Overhead	Not Reported	Not Reported	Not Reported	Not Reported	33%	**4.3%**
Test Chip	Total Area Overhead	Not Reported	Not Reported	3.8%	6.9%	4.42%	13.6%
	# Razor cell/ # Total FF	121/ 826	Not Reported	12%	503/ 2976	492/ 2482	1115/ 12875
	# Gate Count	65K (IV)	123K (IV)	Not Reported	Not Reported	Not Reported	1040K
	Technology	130 nm	65 nm	45 nm	32 nm	45 nm	40 nm

(I) Compared to standard 24T DFF
(II) Listed extra # of Transistor includes transistor count in local clock generation (30 Trs, Fig. 2), amortized over average # of latches per cluster and compared to 24T FF. 3 transistors added to each latch making -5 transistors compared to 24T FF.
(III) Only clock overhead compared to standard flip-flop (IV) Transistor counts divided by 4

Figure 8.8.6: Comparison table of previous ECAD and iRazor.

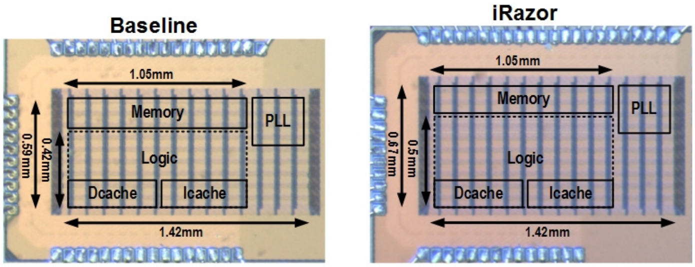


Figure 8.8.7: Die photo of baseline and iRazor Cortex-R4 processor in 40nm CMOS.