ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 292 - 296 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp56.pdf

THE EFFECTIVE PROGRAMMABLE RAZOR BASED TRUNCATED MAC FOR DIGITAL SIGNAL PROCESSING

K.Jyothi prakash[1],M.Hemalatha[2]

PG scholar, Dept of vlsi, sv college of engineering, tirupati, india, prakashkalikiri16@gmail.com

Asst.Professor, Dept of ece, sv college of engineering, tirupati, india,maddihemalatha@gmail.com

Abstract— Independently, programmable truncated multipliers also achieve power benefits at the expense of degradation of the output signal-to-noise ratio Fault tolerant techniques can extend the power savings achieved by dynamic voltage scaling by trading accuracy and/or timing performance against power. Such energy improvements have a strong dependent on the delay distribution of the circuit and the statistical characteristics of the input signal.. In this brief, a combination of programmable truncated multiplication is used within a fault tolerant digital signal processing (DSP) structure in which the supply voltage is reduced beyond the critical timing level. Timing modulation properties of truncated multiplication are analyzed and demonstrated to improve the performance of fault tolerant designs, reducing error correction burdens, and extending the system operating voltage range. Combining both power strategies results in lower energy consumption levels, which improve the energy savings beyond that expected when applied with a combination of both techniques with the original DSP.

Index Terms—Digital signal processing (DSP), fault tolerant, low power , Razor, reconfigurable multiplier, truncated multiplication.

I.INTRODUCTION

Truncated multiplication has been widely studied as a means of achieving both power and area improvements in the field of arithmetic circuit design, at the expense of signal degradation . As the truncated multipliers are smaller than full-precision ones, they not only achieve improvements in power consumption and area, but result in different timing distributions. The existence of synergic benefits derived from the combination of truncated multiplication and VOS using a fault tolerance strategy is presented in this brief where both techniques are applied to a custom-designed fixed point multiply and accumulate (MAC) structure.

Voltage scaling provides an effective means to lower power consumption in VLSI circuits, because scaling the supply voltage by a factor of K results in reductions in the dominating dynamic power consumption by a factor of $K2$ and yields static power benefits.However, advances in CMOS technology scaling contributed to an exponential growth of design issues derived from process voltage temperature (PVT) variations, often resulting in conservative designs that lead to a high power consumption.Some of the classic design timing constraints can be relaxed in digital signal processing (DSP) systems by applying unconventional voltage overscaling (VOS) levels to further improve energy consumption levels while maintaining signal processing performance. Two of the main streams for providing error-resiliency against timing violations are: 1) techniques that introduce an estimation or prediction sub system that monitors the system output and provides anapproximation if a fault is detected and 2) techniques that modify the data capture by augmenting the latches or flip-flops on the critical path and allotting extra execution time for operations that need a long execution time.Such techniques allow implementation of low power systems with acceptable circuit performance at the expense of either signal degradation or execution time penalties. Power savings obtained by fault tolerant techniques are dependent on both PVT variations and the circuit physical design, but are also influenced by the data input to the circuit, as the statistical timing distribution defines the percentage of samples estimated and/or corrected, thus conditioning the maximum power savings obtainable using such techniques.

This brief is organized as follows. Section II reviews some of the latest relevant VOS, fault tolerance and truncated multiplication advances. Section III details the programmable truncated multiplyand-accumulate (PTMAC) architecture used in this brief. The proposed implementation, where fault tolerance and programmable truncation are combined, is analyzed in Section IV. Finally, results for power and energy reductions are studied from postsynthesis simulations in Section V and conclusions of the technology presented are given in Section VI.

II.BACKGROUND

A. Razor and Fault Tolerance for Timing: The Razor technique was originally presented in as an approach to apply dynamic voltage scaling by dynamic detection and correction of circuit timing errors. By measuring the error rate in the circuit, the supply voltage can be tuned while the circuit is

In operation, easing the requirements imposed by conservative timing analysis. Implementation issues of Razor along with its required hardware overhead were addressed in [4] and [15], where Razor II and Bubble Razor were introduced and tested within a full system with reduced area and timing overheads, and in [5] where Razor is applied to a high-speed real-time finite-impulse response (FIR) filter. The efficiency of Razor, and the limits regarding Vdd scaling depend on the circuit timing distribution. Therefore, for any circuit implementing Razor, reducing the amount of time required to perform the average and slowest operation will significantly improve Razormerits. This is the motivation for considering the truncated multiplier which exhibits a timing profile different from the standard multiplier.

B.Voltage Scaling Beyond Vdd−crit

Dynamic power consumption is the dominating component in many arithmetic unit circuits because of the high toggling profile of such structures. The switching component of the energy consumed by a digital gate is defined as Pavg = α0→1CLV2 dd $f$clk in where α0→1 is defined as the average number of times in each clock cycle (at a frequency $f$clk) that a node with capacitance CL makes a powerconsuming transition. Reducing the supply voltage by a factor of K results in a quadratic improvement in the power consumption rate of CMOS logic.

Scaling of Vdd results in timing penalties which increase as Vdd approaches the threshold voltages of the devices .The relationship between the circuit delay (τd) and the supply voltage Vdd is given by $\tau d = CLV$dd$/\beta(V$dd $- Vt\ )\alpha$, where CL is the load capacitance, β is the gate transconductance, Vt is the device threshold voltage, and α is the velocity saturation index. We refer to the critical supply voltage of a given architecture Vdd−crit, as the minimum supply voltage where timing on the critical path is met for any expected PVT variations.

Scaling the supply voltage to Vdd = K ·Vdd−crit, where $0 < K < 1$ is referred to as VOS; although this technique results in further energy reductions almost proportional to K2, scaling Vdd below the critical supply voltage results in critical timing failures for certain input combinations under certain PVT conditions. This is impractical for use with designs that do not apply fault tolerant schemes.

C. Truncated Multiplication

In systems where it is not necessary to compute the exact least significant part of the product, truncated multipliers allow power, area, and timing improvements by skipping the implementation of sections of the least significant part of the partial product matrix.

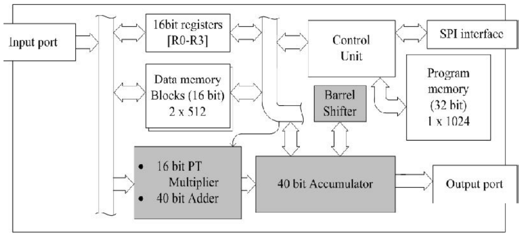
Instead of computing the full-precision output, the output is that from the sum of the first ($N + h$) columns (where $0 \le h \le N$), where N is the operand width, plus an estimation of the erased bits.

In many applications, product values generated by fixed width $N \times N$ bit multipliers are truncated or rounded back to the original bitwidth in latter stages of the algorithm flow. Truncation allows a way of reducing the complexity of the multiplier unit by replacing the lower parts of the partial product matrix by a smaller compensation circuit, and its variants range from very aggressively truncated applications to faithfully rounded truncated multipliers Programmable and configurable approaches to truncated multiplication use fixed-width structures that can be operated at reduced resolutions by disabling parts of the partial product generation. The introduction of programmable truncation in a fixed-point multiplier facilitates modifying not only he multiplier power, but also the timing of the system where the multiplier is embedded. This also alters the original critical path (OCP) of such an arithmetic block, making the architecture virtually faster where the active critical path (ACP) τACP < τOCP. This characteristic of the PTM over the overall and maximum delay is exploited with fault-tolerant schemes to achieve lower minimum energy consumption limits.

III. PTMAC—A LOW-POWER DSP WITH PTM

To extend the usage of PTM to general DSP architectures, the PTMAC was introduced and analyzed in and.PTMAC, designed as a vehicle to exercise PTM in low-power biomedical applications with a need for modest DSP such as ECG filtering or fall detection, will be utilized in this brief as a platform to combine the benefits of programmable truncation and fault tolerance.The proposed DSP, as depicted in Fig. 1, includes a control unit operating in a five-stage pipeline, program and memory blocks in a multibus Harvard configuration, some I/O connectivity and an arithmetic unit consisting of a MAC structure with a 16-bit PTM, a 40-bit accumulator, and a 40-bit barrel shifter for scaling and rotating the accumulated value. The total gate count of the original PTMAC chip is 47 k, and it is estimated (postsynthesis) to have a maximum power consumption of 78.46 μW/MHz. Timing analysis of the proposed PTMAC architecture indicates that the critical path is located within the MAC structure of the arithmetic unit; therefore, energy savings derived from the application of voltage scaling approaches will be constrained by the signal propagation time through the arithmetic unit. An experimental approach to combine the delay-modulation capabilities of programmable truncation and the benefits of fault tolerance is explored in the following sections as a way to achieve a flexible unit that trades energy for signal and performance degradation.

Fig. 1. PTMAC top level diagram.



IV. RAZOR-BASED PTMAC

The combination of a PTM and a fault tolerant system allows such a system to modulate the average and maximum delay times in the MAC unit at run time. Therefore, the number of errors that need correction at any *V*dd level can be trimmed down by reducing the multiplier accuracy. This technique also enables lower functional *V*dd values that can be applied before nonrecoverable failures appear in the system, delivering lower optimum operation voltages which result in lower energy expenditure levels.To explore the independent benefits and interactions between fault tolerance and truncated multiplication, Razor PTMAC was designed as an evolution of PTMAC. To that end, a Razor-enabled version of the original DSP was designed and implemented using Xilinx ise 14.3

A. Razor Implementation

To achieve the fault tolerance, the accumulator unit of the PTMAC was replaced by a fault tolerant version named Razor Accumulator where the original flip-flops were substituted by a version of the Razor registers presented in and The proposed augmented cells were designed and stored as library cells for postsynthesis insertion. Such a cell follows the original implementation Razor implementation, replacing the shadow latch within the Razor registers with a shadow-flip-flop to avoid synthesis issues. The metastability detector required in Razor implementations was modeled as the delay of an inverter added as a constraint to the hold time of the Razor accumulator. In this way, all timing violations potentially causing metastability are then detected as timing errors, providing a lower bound for the performance of Razor.Static timing analysis of PTMAC demonstrated that the only registers situated at potentially critical paths within PTMAC were located in the accumulator, as the multiplication and accumulation of the input data is performed within a clock cycle. Therefore, flipflops capturing the 10 most significant bits of the accumulator were replaced by Razor flip-flops. Insertion of the Razor flip-flops and the associated control logic resulted in an increase of 18% of the area.

Since the hold constraint only limits the maximum duration of the positive clock phase and does not affect the clock frequency [4],single clock was utilized to drive both main and razor flip-flops with both transition edges providing flexibility to configure the extra time allowed by the shadow registers by configuring the duty cycle of the clock. A delay of 25% of the overall clock cycle was selected, which results in an asymmetrical clock signal.The selection of a short error detection phase, enabled a strategy whereby a barrier formed by transparent latches was situated between the compression tree of the multiplier and adder blocks. During the high phase of the clock, the partial products generation begins, but the signals provided by the multiplier are blocked at the latch barrier. With the four stages of the Razor error detection-correction cycle indicated. input, while during the low cycle of the clock, the latches become transparent and signals are free to pass to the adder.

1) Architectural Replay*:* To correct the system when a Razor error occurs, the architecture has to provide the ability to allocate an extra clock cycle for the faulty data to be replaced with the correct one. From different possible pipeline management strategies suggested in the Razor literature, architectural replay was considered the most suitable for PTMAC. Since all the operations perform write or read or arithmetic operations, a simple repeat strategy where a stall flag is issued in the presence of Razor errors, allows the flexibility to correct faulty results with a small area overhead while avoiding issues with postincrementing address pointers.The execution cycle of an instruction on the Razor-augmented PTMAC (RPTMAC) can be thus divided in four possible stages.

a) **EP:** The initial half clock cycle is the execution phase where the instruction starts its execution, but because of hold time requirements it does not reach any of the augmented registers.

b) **AP**: The second half clock cycle is the arrival phase (AP) where the instruction finishes its execution and data is allowed to reach the destination registers. Instructions failing to do so will generate either an Error or a System Failure.

c) **EDP**: The third half clock cycle is the error detection phase, where signals that failed completion can finish their execution causing a Razor error. Instructions failing to finish during the third stage will cause a System Failure, limiting the minimum supply voltage applicable to the system.

d) **ECP:** The fourth stage is the error correction phase (ECP). In the event of an Error being flagged in EDP, a multiplexer will feed the output of the value previously captured by the Shadow Latch into the input of the Main Flip-flop, and the error signal will be cleared. Stages EP and AP represent a regular execution stage of a pipeline,while the last two stages (EDP and ECP) overlap with the execution stage of the next instruction in Fig. 2. Detecting an error in the EDP phase causes either the ECP of the faulty instruction or the AP phase of the following one to update the output of the Razor registers

V. EVALUATION OF THE RAZOR PTMAC

To evaluate the performance of the Razor PTMAC, estimations for timing and power were performed under different truncated levels using TSMC 90 nm libraries. Execution times were recorded for a total of 20 000 MAC instructions over random data, and power was measured for a demonstrative algorithm where the MAC is processing data on 50% of the instructions executed Fig. 3 displays timing histograms of (a) the original and (b) Razor PTMAC units, where different colors indicate different truncation levels. In terms of timing, the insertion of Razor on the origina PTMAC architecture was negligible, as timing in the arithmetic unit was not impacted by the fault tolerant insertion. Average and maximum execution times were 7.1 and 12 ns for the original PTMAC, and 7.4 and 11 ns for the slightly more power consuming (8% increase) Razor version. The effect of voltage scaling over the system delay was approximated by using the alpha-power model equations presented in [14] and Section II-A, with the delays obtained from TSMC 90 nm at two known voltages with the same process and temperature characteristics. As the architecture is active through the full simulation, the overall energy consumption is dominated by the dynamic power sources. Therefore, energy consumption at different values of *V*dd was calculated by applying a scaling factor *K*2 to the energy expenditure generated by the circuit logic when running the algorithm at a voltage *V*dd = *K* · *V*DD−nominal, where *V*DD−nominal = 1.21V is the nominal operating voltage provided by TSMC. Power dissipated by the memory blocks, account for approximately half of the original consumption, was not scaled with the voltage supply. Table I displays the energy performance of the Razor PTMAC architecture when executing the proposed test routine at a clock frequency of clk = 62.5 MHz and a clock duty cycle of 25%. It can be appreciated how the introduction of Razor allows pushing the architecture supply voltage beyond the critical supply of the original implementation, *V*dd−crit = 0.986 V. In such a way, energy savings can be achieved despite the initial overheads introduced by Razor (≈8%). The minimum functional *V*dd levels were in the range of [0.86–0.870] volts, while the minimum energy consumption levels were obtained at supply levels ranging 0.898–0.909 V. All power values are normalized to the power estimated for the case of nonrazor PTMAC with zero truncated columns (63.8 *μ*W/MHz), whereas maximum power consumption on the arithmetic unit is estimated to be 28.1 *μ*W/MHz when operating on random data, at full resolution, and nominal voltage (*V*DD−nominal). Reference publications in the field of the truncated multiplication estimate power consumptions of 36.6–12.2 *μ*W/MHz (16-bit multiplier, 0.13 m) in and70.3–37.6 *μ*W/MHz (16-bit multiplier and 20-bit accumulator,0.18 m) in The energy consumption levels estimated on Razor-PTMAC were normalized to the case of a PTMAC unit without fault tolerance and 0-bit truncation *E*original. Applying truncation without fault tolerance achieved minimum energy consumption levels of *E*truncation = *E*original □ *α*trunc = 90.86%, whereas the minimum consumption for fault tolerance without truncation was *EFT* =*E*original □ *α*FT = 81.25%. Assuming they are independent, the most optimistic energy levels expected after applying both techniques would *E*RPTMAC_expected = *E*original □ *α*FT □

TABLE I

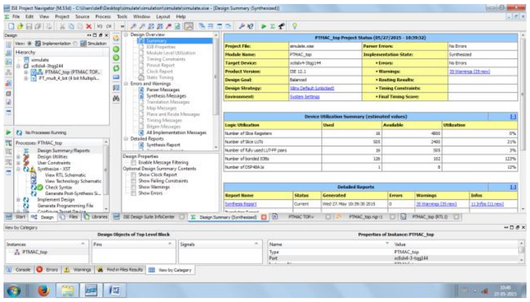
ENERGY MEASUREMENTS FOR THE RAZOR-ENABLED AND PTMAC ARCHITECTURES. VALUES NORMALIZED TO THE CONSUMPTION OF THE PTMAC IMPLEMENTATION WITH NO TRUNCATION APPLIED

Truncation	Original ptmac				Razor ptmac			
	0	16	20	24	0	16	20	24
Vdd=1.18 Vdd=1.0 Vdd=1 Vdd=0.96 Vdd=0.9 Vdd=0.88 Vdd=0.86	1 - - - - - -	0.9 8 - - - - - -	0.9 - - - - - -	0.9 - - - - - -	1 1 1 1 0 0 -	1.2 0.9 0.8 0.7 0.7 0.9 -	1 .9 .8 .7 .7 .9 -	0.9 .88 .79 .74 .74 .86 -
SNR(dB)	∞	65	41	19	∞	65	41	19

*α*razor = 74.6% while postsynthesis simulations indicated the existence of a new lower energy bound of *E*RPTMAC_simulation = 74.2%, in regions where the application of truncation allowed to reduce the number of timing errors corrected by the Fault tolerant technique.

A summary of the tradeoffs achievable at a system level by the use of the Razor PTMAC is shown in Fig. 4. Reductions in the supply voltage beyond the optimum point generate a steep rise in the number of Razor errors, which results in a fast increase in the pipeline correction overheads and the inability to obtain any more energy savings by applying further voltage scaling. Increasing the number of truncated columns in the proposed system provides energy savings via three mechanisms: 1) switching reductions expected from truncated multiplication; 2) a dynamic reduction of the average computation time required at any given voltage, what implies a reduction of the number of errors per operation, lowering the energy burdens derived from error correction; and 3) the introduction of truncation results in the effective critical path being shorter than the original critical path, extending the lowest feasible supply voltage range.

VI. RESULTS ANAIYSIS



PTMAC_top Project Status (05/27/2015 - 10:39:32)			
Project File:	simulate.xise	**Parser Errors:**	No Errors
Module Name:	PTMAC_top	**Implementation State:**	Synthesized
Target Device:	xc6slx4-3tqg144	**• Errors:**	No Errors
Product Version:	ISE 12.1	**• Warnings:**	35 Warnings (35 new)
Design Goal:	Balanced	**• Routing Results:**	
Design Strategy:	Xilinx Default (unlocked)	**• Timing Constraints:**	
Environment:	System Settings	**• Final Timing Score:**	

Device Utilization Summary (estimated values)			
Logic Utilization	**Used**	**Available**	**Utilization**
Number of Slice Registers	16	4800	0%
Number of Slice LUTs	505	2400	21%
Number of fully used LUT-FF pairs	16	505	3%
Number of bonded IOBs	126	102	123%
Number of DSP48A1s	1	8	12%

Detailed Reports					
Report Name	**Status**	**Generated**	**Errors**	**Warnings**	**Infos**
Synthesis Report	Current	Wed 27. May 10:39:30 2015	0	35 Warnings (35 new)	11 Infos (11 new)

Properties of Instance: PTMAC_top	
Name	Value
Type	PTMAC_top
Part	xc6slx4-3-tqg144

Power can be reduced upto 25% as compared to existing ptmac and having combinational delay of 17.63ns.

VII. CONCLUSION

The use of Razor on a PTMAC structure has been tested at a postsynthesis simulation level to study the effect and interactions of both energy reducing techniques on a previously tested DSP design. The timing and power effects of VOS with error correction and

the application of programmable truncated multiplication resulted in significant power reductions.Fault tolerance was provided by implementing a conservative approach to the Razor I technique, and achieved energy reductions of 16.7% over the original DSP implementation by enabling the reduction of *V*dd beyond the original critical supply level. Truncated multiplication was achieved by implementing a PTM, and resulted in energy savings of 7.1% of the full design. Energy reductions achieved by fault tolerant techniques are limited by the overheads required to provide error resilience and the amount of operations that need correction, therefore, they are highly influenced by the delay distribution and maximum value of the system critical paths. The introduction of the truncated multipliers achieve two goals in this scenario: 1) it reduces power on the multiplier by cancelling the switching activity within its least significant sections and 2) disables the multiplier critical path, thus reducing the error recovery overheads of Razor, and extending the applicable *V*dd range.

Results show that the application of both techniques to the proposed DSP unit allow maximum energy savings of 24.8%, improving the results obtained by independently implementing programmable truncation, fault tolerance via Razor, and the most optimistic prediction for the combination of both techniques (24.4%). This indicates that the delay-modulation properties of truncated multiplication can be exploited to improve the energy consumption of fault tolerant DSP architectures where multipliers are involved in the critical path of the circuit.

REFERENCES

[1] B. Shim, S. Sridhara, and N. R. Shanbhag, “Reliable low-power digital signal processing via reduced precision redundancy,” *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 5, pp. 497–510, May 2004.

[2] R. Hegde and N. R. Shanbhag, “Soft digital signal processing,” *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 9, no. 6, pp. 813–823, Dec. 2001.

[3] S. Das, C. Tokunaga, S. Pant, W.-H. Ma, S. Kalaiselvan, K. Lai, *et al.*, “RazorII: 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

[4] D. Ernst, N. S. Kim, S. Das, S. Pant, R. Rao, T. Pham, *et al.*, “Razor: A low-power pipeline based on circuit-level timing speculation,” in *Proc. 36th Annu. IEEE/ACM Int. Symp. Microarch.*, 2003, pp. 7–18.

[5] P. Whatmough, S. Das, and D. Bull, “A low-power 1 GHz razor FIR accelerator with time-borrow tracking pipeline and approximate error correction in 65 nm CMOS,” in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2013, pp. 428–429.