Robust Digital Signal Processing for Wireless SoCs

P N Whatmough†‡ and I Darwazeh†

† University College London, ‡ ARM Ltd.

Abstract: Excessive process variability coupled with on-chip supply droop in wireless System-on-Chip (SoC) design has lead to the threat of transient timing errors in digital circuits. We propose a novel FIR filter architecture that mitigates sub-critical timing violations as they occur in the pipeline structure by momentarily bypassing affected coefficients. The proposed architecture is implemented in a 90nm CMOS process technology and verified using full model SPICE simulations. At the typical process and temperature corner, the proposed architecture can be scaled in voltage down to the point of first failure at 730 mV, thereby achieving a 53% power saving, with no detectable degradation in stop-band attenuation characteristics.

1. Introduction.

Finite Impulse Response (FIR) filters are a key class of digital signal processing (DSP) functions. Efficient implementation in terms of silicon area and power consumption can be challenging for high-throughput programmable FIR accelerators, since they require many parallel multiply-accumulate (MAC) operations. For DSP applications, the requirements on maximum clock frequency are often modest compared to other domains such as general purpose computing. However, power consumption is often critical, particularly for battery-operated devices. As deep sub-micron (DSM) CMOS geometries scale, process, voltage and temperature (PVT) variability is greatly exacerbated which impairs power efficiency, since a higher supply voltage is required to safely cope with such variability. This higher supply voltage ensures that the worst-case circuit delay due to PVT variation does not result in timing violations which jeopardize correct operation. In addition, it is unlikely that any ICs will ever be subject to the worst-case parametric variation for which ICs are margined conventionally. Hence, we design for a worst-case scenario that may rarely occur in practice. For low-power designs, there are significant gains to be made by providing circuit robustness using alternative methods to excessive voltage margin [1].

Adaptive design is a recent approach to IC architecture that offers an alternative to the increasingly large proportion of energy wasted due to PVT margin [2]. The essential concept is to dynamically tune the supply voltage and/or clock frequency to the native speed of each individual IC during functional use. The “in situ error detection” [1][3] approach enables the dynamic voltage scaling (DVS) to cross into the sub-critical region where timing errors may occur and in turn circuit techniques must be employed to detect and correct the errors at the point in the circuit at which they occur. The goal of this scheme is to virtually eliminate PVT safety margins since the majority of the uncertainty is handled deterministically by the error detection logic and is therefore attractive for low-power systems.

This paper is concerned with examining application of in situ error detection to the FIR filtering algorithm. Section 2 briefly describes existing approaches to DVS for DSP systems. In Section 3 we propose an FIR architecture with circuit-level error detection and algorithmic error mitigation. This is verified by simulations with results presented in section 4.

2. Dynamic Voltage Scaling for DSP Systems

There is considerable interest in DVS for low-power DSP systems. Sub-critical operation has also been studied, with the bulk of published work focused on an approach termed Algorithmic noise tolerance (ANT) [4]—[6]. The majority of these techniques seek to exploit the fact that for arithmetic operations constructed of ripple-carry adders, the propagation delay for each output bit increases monotonically with the bit order. Hence, it can be assumed that timing errors will generally occur in the most-significant bits (MSBs) first. With errors concentrated in the MSBs, error detection based on inspection of the output samples becomes feasible for a highly correlated (narrow-band) signal, because errors will be large in magnitude. Estimation theory is employed to predict the output of a DSP function in order to perform error detection by comparison and, if necessary, error correction by replacing the erroneous sample with the predicted version. Appropriate estimation methods include linear prediction [4], reduced-precision redundancy [5] and adaptive error-cancellation [6]. The design of the prediction block presents additional design complexity and circuit overhead.

In general, the ANT approaches rely on the condition that timing violations will generate errors of large magnitude in order to perform error detection. Ensuring this condition in the context of a commercial standard-

cell DSM design presents a number of concerns at the implementation level, relating to environmental variability, input signal statistics and dominant net delays.

To address these issues, we propose replacing algorithmic error detection with a robust circuit-level technique that has been developed for general purpose microprocessor pipelines. The significant advantage of this approach is that we do not require that timing violations result in MSB errors and therefore we have free choice of arithmetic circuit architecture and do not need to guarantee critical paths at design time. The ability to detect timing errors allows a DVS system to tune the operating point to the critical supply voltage by maintaining a low but non-zero error rate, which is referred to as the point of first failure (PoFF). This enables us to eliminate PVT margin and track changes in environmental conditions while avoiding operation in a significantly sub-critical region which brings only a small incremental power reduction and an exponential increase in timing violations. In order to prevent intermittent timing violations from significantly degrading the filter transfer function, we propose a simple but deterministic error mitigation technique.

3. Proposed VLSI Design

In situ error detection is achieved in the proposed design using Razor flip-flops [7]. Razor flip-flops augment a conventional master-slave flip-flop with a transition detector that operates during the high phase of a pulsed clock. This arrangement can detect late arriving transitions on the input D pin to the Razor cell and flag an error event. Razor flip-flops are used to replace conventional flip-flops in the well-known transposed direct-form FIR architecture. Hence, there is a Razor flip-flop following every multiply-accumulate operation and a single-bit error flag that indicates if a timing error has effected one or more bits of the pipeline stage. An additional benefit to this approach is some protection from SEU at the Razor flip-flop as well as in the combinatorial logic that fans into it [7].

The available options for recovering from a timing error detected by the Razor pipelines are severely constrained in real-time DSP applications because variable latency is difficult to manage in the face of hard throughput deadlines. It is therefore required that the error recovery mechanism does not introduce additional latent cycles. Our proposed approach adds logic to each pipeline stage such that the individual multiply-accumulate operations, isolated by Razor flip-flop stages, prevent an erroneous tap contribution into the adder chain by bypassing the stage output using a mux and a delay register.

This approach also ensures that for a given probability of error (Pe), the worst-case algorithm performance is largely deterministic, an important consideration at the system design level. Without error-mitigation, erroneous tap results, which by nature are often large in magnitude, will propagate into the pipeline, corrupting subsequent correct results further down the pipeline.

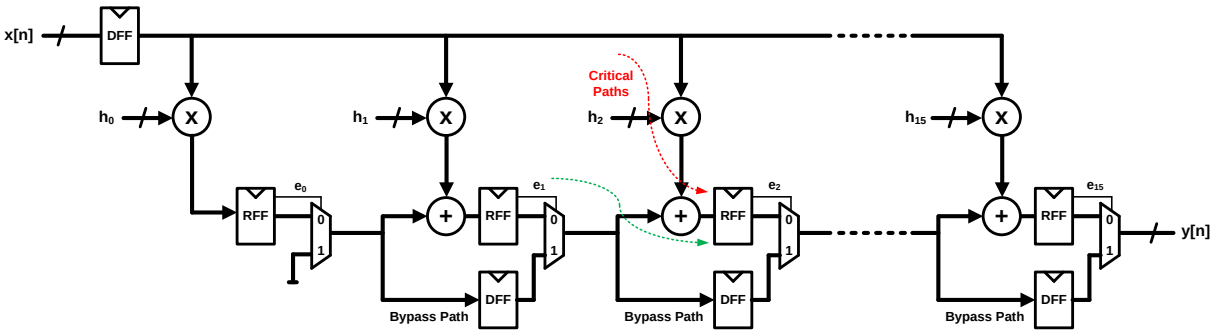

Figure 1: Proposed filter structure with error mitigation

Fig. 1 shows the proposed architecture, where x[n], y[n] and h_k represent the input signal, output signal and coefficients. This represents a modification of the FIR algorithm such that

$$y[n] = \sum_{k=0}^{N-1} \begin{cases} h_k x[n-k], e_k = 0 \\ 0, e_k = 1 \end{cases}, \quad (1)$$

where e_k is a 1-bit flag that indicates the detection of an error at the k^{th} Razor pipeline stage. Where h_k is symmetric, it is necessary to bypass both of the equal coefficients in order to maintain phase linearity while mitigating errors.

The critical paths through the multiply-accumulate operators are illustrated with red arrows, where "DFF" refers to a conventional library flip-flop and "RFF" indicates a Razor flip-flop. The modifications create a secondary critical path group from the late-arriving mux select through the adder to the next razor flip-flop (green arrow), which must be suitably constrained to achieve sufficient timing slack. Hold time constraints are an important consideration with Razor implementations since early arriving transitions that reach a Razor flip-flop during the high phase of the clock will trigger a false error detection event. However, in the proposed architecture this is not an issue, since there are no significant fast paths through the arithmetic logic. Since the coefficients are programmable in our design, we do not take advantage of symmetric folding or employ hardwired coefficient multipliers.

4. Simulation Results and Discussion

The proposed architecture was applied to a 16-tap FIR filter in a 90nm CMOS process. We employ 10-bit coefficients, 8-bit input data words and a 20-bit adder chain. The two's complement arithmetic implementations were chosen freely by the synthesis tool resulting in a delay-optimized Booth Wallace implementation with carry-save representation between the multiply and sum operations. There is an area and power overhead for a Razor flip-flop compared to an equivalent library flip-flop. The area overhead per bit is approximately double. Further, the proposed architecture requires an additional mux and flip-flop stage for each filter tap. Total area and power overhead as compared to a conventional transposed direct-form implementation are estimated to be 26% and 24% respectively, at the nominal supply voltage. The total area of the layout is 0.084 mm^2 (of which 60% is combinatorial logic and 40% registers).

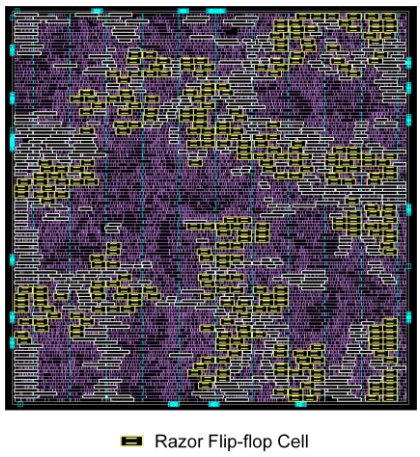

Figure 2: Layout of the Robust FIR Filter

Post-layout SPICE simulation with extracted parasitics was used to evaluate the proposed architecture over a range of supply voltage scaling conditions with an otherwise typical environment (TT process, 25°C). The coefficients used for the simulations are for a symmetric low-pass response with normalized cut-off frequency, $\omega_c = \pi/2$ radians/sample. 10,000 input vectors, consisting of uniform noise samples with maximum amplitude of 0 dBFS, were used to test the filter amplitude response up to the Nyquist frequency. The minimum clock period at the slow corner was found using conventional static timing analysis to be 2.4 ns (420 MHz). This clock period was kept constant while the supply voltage was scaled down from the process nominal of 1 V. To gauge the algorithmic performance of the proposed filter we employ a measure of stop-band rejection, η, defined as the ratio of pass-band power to stop-band power, both of which are defined as integrals of the power spectral density, Φ(ω);

$$\eta = 10.\log_{10}\left(\frac{\int_0^{\omega_c}\Phi(\omega).d\omega}{\int_{\omega_c}^{\pi}\Phi(\omega).d\omega}\right). \quad (2)$$

Results presented in Figs. 3 indicate a 53% power saving from 28.62 mW at 1 V (full PVT margin) to 13.41 mW at 730 mV, which is the PoFF (proposed zero margin operating point); the reduction in η (2) is undetectable. Beyond this point, P_e increases by almost a decade per 10 mV reduction in V_{dd} and η subsequently degrades at

5.5 dB per 10 mV. As expected, this suggests that there is little to be gained from operating beyond the PoFF with the proposed architecture.

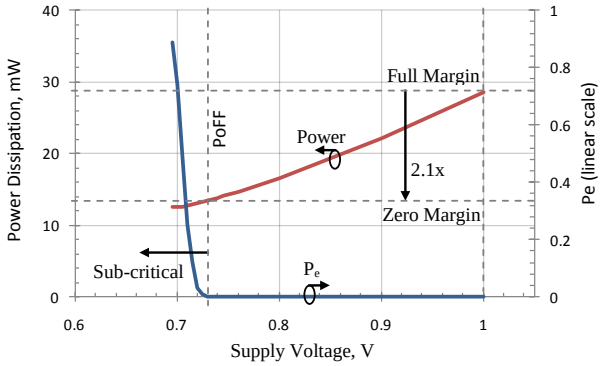

Figure 3:Simulation of Power Dissipation and Pe vs. Vdd

P_e and η are dependent on both the input data (x[n]) samples and the coefficients (h_k) and will therefore vary depending on application. We employed random stimulus over the full input dynamic range in order to analyze robustness to the worst-case error patterns, since random data contains a disproportionately high number of sign changes which exercise the critical MSB paths in arithmetic logic. Consequently, many real-world signals would generate a significantly lower P_e, in turn leading to a more graceful degradation in η than seen here.

5. Conclusion

We report and describe a novel FIR filter architecture incorporating in situ error detection and mitigation for low-voltage operation. Unlike algorithmic error detection schemes, the proposed architecture does not rely on predictable path delays that increase monotonically with bit order, but instead uses a circuit-level technique to detect timing violations. Error mitigation is achieved by bypassing erroneous taps in the adder chain, such that the filter can tolerate a small but non-zero error rate when operating with a supply voltage dynamically scaled to a marginally sub-critical operating point. The proposed filter was implemented in a 90nm CMOS process and simulated using SPICE, assuming typical conditions (TT process, 25°C). Power dissipation was found to scale with supply voltage from 28.62 mW at 1 V (nominal) to 13.405 mW at the point of first failure at 730 mV. This is a power saving of 53%, with an undetectable reduction in algorithmic performance. Beyond the PoFF, a trade-off with power and performance can be achieved although the potential gains are small compared to those achieved from PVT margin elimination. The area and power overhead at nominal supply voltage compared to the transposed direct form is conservatively estimated to be 26% and 24% respectively.

References.

[1] S. Das and D. Blaauw, “Adaptive design for nanometer technology”, in proc. IEEE Int. Symp. Circuits and Systems, pp. 77-80, May 2009

[2] J. Tschanz et al., “Adaptive frequency and biasing techniques for tolerance to dynamic temperature-voltage variations and aging,” in IEEE Intl. Solid-State Circuits Conf. Dig., pp. 292–293, Feb. 2007

[3] G. Dasika, S. Das, K. Fan, S. Mahlke, D. Bull, “DVFS in loop accelerators using BLADES”, in proc. IEEE/ACM Design Automation Conf., June 2008

[4] R. Hegde and N. Shanbhag, “A voltage overscaled low-power digital filter IC”, IEEE J. Solid-State Circuits, vol. 39, no. 2, Feb. 2004

[5] B. Shim, S. Sridhara and N. Shanbhag, “Reliable low-power digital signal processing via reduced precision redundancy”, IEEE Trans. on VLSI Systems, vol 12, no 5, pp. 497-510, May 2004

[6] L. Wang and N. Shanbhag, “Low-power filtering via adaptive error-cancellation”, IEEE Trans. on Signal Processing, vol. 51, no. 2, pp. 575-583, Feb. 2003

[7] S. Das et al., “RazorII: In situ error detection and correction for PVT and SER tolerance”, IEEE J. Solid-State Circuits, Vol. 44, No. 1, Jan. 2009