Available online at www.sciencedirect.com
SCIENCE DIRECT®
INTEGRATION, the VLSI journal 39 (2005) 12–28
INTEGRATION the VLSI journal
www.elsevier.com/locate/vlsi

Development of a large word-width high-speed asynchronous multiply and accumulate unit

S.C. Smith*

Department of Electrical and Computer Engineering, University of Missouri-Rolla, 133 Emerson Electric Co. Hall, 1870 Miner Circle, Rolla, MO 65409, USA

Received 24 September 2004; received in revised form 3 November 2004; accepted 30 November 2004

Abstract

This paper details the design of the fastest known asynchronous Multiply and Accumulate unit (MAC) architecture published to date. The MAC architecture herein is based on the MAC developed in Smith et al. (J. Syst. Archit. 47/12 (2002) 977–998). However, the MAC developed in Smith et al. (2002) contains conditional rounding, scaling, and saturation (CRSS) logic, not present in other comparable MACs (Twenty-Sixth Hawaii International Conference on System Sciences, vol. 1, 1993, pp. 379–388; Asian South-Pacific Design Automation Conference, 2000, pp. 15–16; Sixth IEEE International Conference on Proceedings of ICECS, vol. 2, 1999, pp. 629–633); thus making the comparison between the MAC developed in Smith et al. (2002) and other delay-insensitive/self-timed MACs in the literature not completely fair, in favor of the other MACs. This paper first details the removal of the CRSS logic from the MAC developed in Smith et al. (2002), and describes its subsequent optimal re-pipelining, in order to provide a more fair comparison. This yields a speedup of 1.12. Secondly, this paper details the application of the NULL Cycle Reduction technique (The 10th International Workshop on Logic and Synthesis, 2001, pp. 185–189; Gate and throughput optimizations for NULL convention self-timed digital circuits, Ph.D. Dissertation, School of Electrical Engineering and Computer Science, University of Central Florida, 2001) to the MAC's feedback loop, and subsequent re-pipelining of the feed-forward partial product generation

*Tel.: +1 573 341 4232; fax: +1 573 341 4532.
E-mail address: smithsco@umr.edu (S.C. Smith).

0167-9260/$ - see front matter
doi:10.1016/j.vlsi.2004.11.001

and summation circuitry to further increase throughput, resulting in an additional speedup of 1.31 (a speedup of 1.46 over the MAC from Smith et al. (2002). Lastly, the bit-wise completion strategy is utilized in lieu of full-word completion to decrease the area required by 6% and also increase the MAC's throughput an additional 1%.

Keywords: NULL convention logic; Gate-level pipelining; NULL cycle reduction; Computer arithmetic

1. Introduction

For the last two decades the focus of digital design has been primarily on synchronous, clocked architectures. However, as clock rates have significantly increased while feature size has decreased, clock skew has become a major problem. High performance chips must dedicate increasingly larger portions of their area for clock drivers to achieve acceptable skew, causing these chips to dissipate increasingly higher power, especially at the clock edge, when switching is most prevalent. As these trends continue, the clock is becoming more and more difficult to manage. This has caused renewed interest in asynchronous digital design.

Correct-by-construction asynchronous paradigms, such as NULL Convention Logic (NCL) [1], have been demonstrated to require less power, generate less noise, produce less EMI, and allow for easier reuse of components, compared to their synchronous counterparts, without compromising performance [2]. Furthermore, these paradigms should allow for much greater flexibility when designing complex circuits, like a System-on-Chip (SoC), since the circuits are self-timed; thus the effort required to ensure correct operation under all timing scenarios should be drastically reduced, compared to equivalent synchronous designs. Also, the self-timed nature of correct-by-construction SoCs should allow for previously designed and verified functional blocks to be reused in subsequent designs, without necessitating any significant modifications or retiming effort within a reused functional block, and may provide for simpler interfacing between the digital core and non-traditional components.

A major trend in the semiconductor design industry is to continually produce increasingly faster processors. For high-speed Digital Signal Processors (DSPs), one of the main processing bottlenecks is the Multiply and Accumulate unit (MAC). In order to produce faster DSPs, faster MACs must be designed and utilized. For large word-width MACs, the Carry-Propagate Adder (CPA) in the feedback loop becomes the main bottleneck to increasing performance. The self-timed nature of the NCL paradigm utilized for the MAC developed herein helps to mitigate the bottleneck of the feedback loop by allowing the wavefronts to move more independently of each other and by providing for average-case delay in the CPA, instead of the worse-case delay required for a clocked design. Furthermore, the use of a delay-insensitive paradigm allows for the NULL Cycle Reduction (NCR) technique [3,4] to be applied to the feedback loop to further increase its performance. Once the throughput of the feedback loop bottleneck is maximized, the feed-forward partial product generation and summation circuitry can be re-pipelined to match the throughput achieved by the redesigned feedback loop; thus maximizing throughput for the entire MAC unit.