

Highly-Flexible and Optimized, Area and Energy-Efficient Matrix-Decomposition Accelerators based on Givens Algorithm and CORDIC Rotations

Upasna Vishnoi* and Tobias G Noll

Electrical Engineering and Computer Systems, RWTH Aachen University, Germany

Abstract

Matrix decomposition accelerators are attractive SoC components for many applications with a wide range of specifications. In this work, a new family of highly area- and energy-efficient modular matrix decomposition architectures based on the givens-algorithms and CORDIC rotations are elaborated. Accurate algebraic cost models enable for early cost estimation as well as for cross-level optimization over architecture, micro-architecture and circuit-level using a rich set of parameters. Quantitative results for an exemplary QRD application, implemented in 40-nm CMOS technology, are presented.

Introduction

Matrix-decomposition and -factorization techniques are powerful tools of linear algebra and have applications in topics that span sciences. Specifically, QR, Eigen Value (EVD) and Singular Value (SVD) decomposition are applied in sophisticated digital signal processing for image compression, reconstruction and restoration as well as in decoding, space and noise filtering, parameter estimation and source separation in communications, to name just a few. Interesting and very successful research on systolic architectures for real-time implementations of matrix-decomposition algorithms has been conducted in the 80s and 90s of the last century [1,2]. It was found that in comparison to e.g., Householder-based approaches, Givens- and Jacobi-type algorithms feature a high degree of inherent parallelism and can be implemented by mapping the required rotations onto hardware-efficient Coordinate Rotate Digital Computer (CORDIC) operations [3]. However, during those times VLSI-CMOS technology allowed only for implementation of a single CORDIC processor per chip. Today's very-deep submicron CMOS technologies allow for the realization of high-throughput, low-energy CORDIC macros on a fraction of a square millimeter of silicon area and microwatts of power dissipation [4]. By this, the implementation of high-performance matrix-decomposition modules to be used as number crunching SoC processor sub-macros has become feasible.

In this work, new architectures and a new methodology for the quantitative optimization of matrix-decomposition-processor SoC-sub-macro implementations for the use in challenging application domains featuring a wide variety of requirements and specifications are elaborated. An attractive target architecture consists of a flexible software-controlled Application Specific Instruction Processor (ASIP) executing the matrix-decomposition algorithms and being accelerated by a large farm of dedicated CORDIC blocks. The architecture template should allow for the decomposition of real- as well as complex valued matrices. Floating-point capability can be achieved by operating the individual CORDIC blocks in block exponent arithmetic. In order to achieve maximum area and especially energy efficiency, the optimization has to be performed concurrently on all levels of CMOS design, from the algorithmic level down to the physical implementation level. Only by that, the interactions between decisions on the different design levels being imposed by the features of today's very-deep submicron CMOS technologies can be properly considered. A key element of this approach is the elaboration of quite accurate,

parameterized algebraic cost models for area, throughput, latency and energy of CORDIC macros.

QRD Decomposition Accelerators

QR decomposition today for instance is applied in multiple-input/multiple-output (MIMO) channel detection [5-7]. Suitable CORDIC implementations were investigated and optimized in a 40 nm CMOS technology [4]. Based on the costs of the elementary arithmetic CORDIC sub-functions, a MATLAB-based design-space evaluation environment has been elaborated [8]. Key elements to this approach are pre-characterization of all function slices (adders, shifters etc.) as well as proper interconnect models derived from floorplans. This approach has been successfully validated and optimized by benchmarking against the features of selected exemplary CORDIC implementations [9]. Highly interesting results were achieved especially on the usefulness of applying carry-accelerated (carry-select-) and redundant (carry-save-) adder structures.

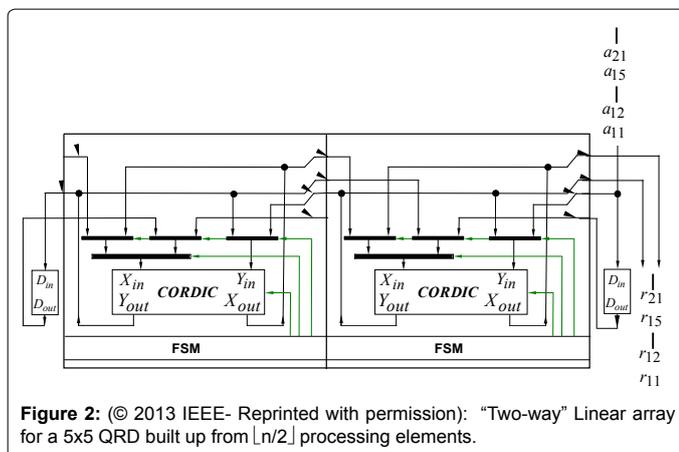
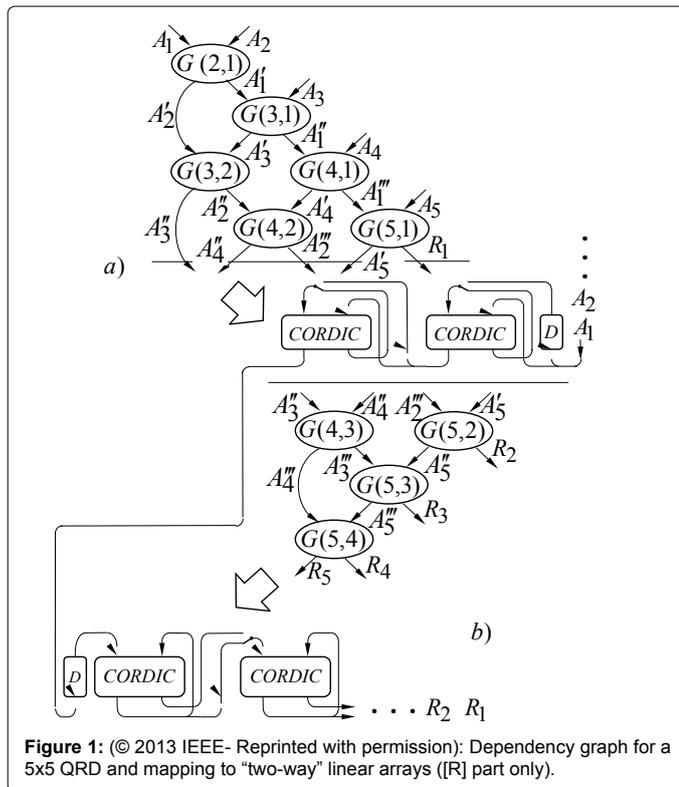
For the simple most kind of matrix decomposition technique, the QR decomposition, a new "two-way" linear array was derived [10]. This array is the result of the mapping from the dependency graph depicted in Figure 1; the resulting signal flow graph is shown in Figure 2 for the simple example of a $n \times n$ matrix with $n=5$. It is composed of $\lfloor n/2 \rfloor$ processing elements and allows for implementation of flexible, micro-code-controlled accelerators featuring lowest possible latency as well as high area- and energy-efficiency. The throughput (and utilization) of that array can be improved by a modification derived from a left-to-right flipping of the dependency graph before mapping. As was shown in ref. [10], there are several possibilities to apply multiplexing in space and/or multiplexing in time to this architecture leading to a high-dimensional design space.

*Corresponding author: Upasna Vishnoi, Electrical Engineering and Computer Systems RWTH Aachen University, Germany, Tel: +49 241 801; E-mail: vishnoiupasna@gmail.com

Received July 20, 2017; Accepted July 28, 2017; Published July 29, 2017

Citation: Vishnoi U, Noll TG (2017) Highly-Flexible and Optimized, Area and Energy-Efficient Matrix-Decomposition Accelerators based on Givens Algorithm and CORDIC Rotations. J Electr Electron Syst 6: 231. doi: 10.4172/2332-0796.1000231

Copyright: © 2017 Vishnoi U, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



Based on that as well as on the CORDIC-cost models a higher level optimization environment for whole QR decomposition processors have been elaborated. By this, one can account for the strong interactions between QR decomposition architecture-, CORDIC micro-architecture- and circuit-level, which have not been considered in any publication on QR decomposition architectures so far. The corresponding MATLAB-based cost model allows for efficient exploration of the resulting highly complex design space.

The new QR decomposition-template architecture has been extended for a family of modular architectures allowing also for the decomposition of complex-valued as well as floating-point matrices [11]. Just to show one exemplary result, a QR decomposition macro for 4×4 matrices and 16 bit word length implemented in 40 nm CMOS technology can be operated at a clock frequency of 910 MHz, allows for latency as small as 100 ns and a throughput as high as 500 million QR decompositions per second. Silicon area is 0.0092 mm², equivalent gate count is 36.7 k and energy is 5.15 pJ per full

QR decomposition (all features are given in worst case corners). Benchmarking with QR decomposition implementations known from literature proves the significant improvement in area-time-energy efficiency [12].

Outlook: EVD and SVD Accelerators

QR decomposition requires one-sided Givens rotations only. In future, new linear arrays for the implementation of EVD and SVD will be driven by proper mapping of the dependency graphs. This time the challenge on architecture level is that two-sided Jacobi- and Jacobi-like rotations have to be applied. Initial ideas for achieving highest possible parallelism as well as highest area and energy-efficiency are to apply composite rotations as well as customized two-port SRAM modules allowing for row- and column-matrix accesses. Again the MATLAB-based cost model will be extended for these architectures to allow for proper design space exploration and early cost estimation in 40 nm and 28 nm CMOS technology.

Finally, the realization of a matrix-decomposition-accelerator demonstrator in a 28 nm CMOS technology is envisaged, which can be used as a dongle-type high-performance SVD-accelerator to a standard computer.

Acknowledgements

This research work was done at the Chair of Electrical Engineering and Computer Systems, RWTH Aachen University, Germany under the able guidance of Professor Dr.-Ing. Tobias G. Noll.

References

1. Kung SY (1987) VLSI array processors. Englewood Cliffs, NJ, Prentice Hall, Research supported by the Semiconductor Research Corp, SDIO, NSF, and US Navy.
2. Cavallaro JR, Elster AC (1991) A CORDIC Processor Array for the SVD of a Complex Matrix. Elsevier Science Publishers, pp: 227-239.
3. Ercegovic MD, Lang T (2003) Digital Arithmetic. (1st edn.), Elsevier.
4. Vishnoi U, Noll TG (2011) Area- and energy-efficient CORDIC Accelerators in Deep Sub-micron CMOS Technologies. Advances in Radio Science 10: 207-213.
5. Senning C, Staudacher A, Burg A (2010) Systolic-Array based regularized QR-Decomposition for IEEE 802.11n Compliant Soft-MMSE Detection. International Conference on Microelectronics (ICM).
6. Salmela P, Burian A, Sorokin H, Takala J (2008) Complex-Valued QR Decomposition Implementation for MIMO Receivers. Proc. IEEE ICASSP, pp: 1433-1436.
7. Patel D, Shabany M, Gulak PG (2009) A Low-Complexity High-Speed QR Decomposition Implementation for MIMO Receivers. IEEE International Symposium on Circuits and Systems.
8. Vishnoi U, Meixner M, Noll TG (2012) An Approach for Quantitative Optimization of Highly Efficient Dedicated CORDIC Macros as SoC Building Blocks. IEEE International Conference (SOCC).
9. Vishnoi U, Noll TG (2012) Systematic Quantitative Optimization of Area- and Energy-Efficient, Number-Crunching-Performance SoC Macros for Singular Value Decomposition. IEEE International Conference on VLSI-SoC.
10. Vishnoi U, Noll TG (2013) Cross-Layer Optimization of QR Decomposition Accelerators. Proceedings of the ESSCIRC.
11. Vishnoi U, Noll TG (2013) A Family of Modular Area- and Energy-Efficient QR-Decomposition Accelerator Architectures. International Symposium on System-on-Chip.
12. Vishnoi U, Noll TG (2013) Systematically and Quantitatively Optimized Area- and Energy-Efficient Matrix-Decomposition Accelerators. International Conference on Very Large Scale Integration (VLSI-SoC).