

Highly-Efficient Number-Crunching-Performance SoC Macros for Singular Value Decomposition

Upasna Vishnoi*

Electrical Engineering and Computer Systems, RWTH Aachen University, Germany

Keywords: Design-space exploration; Early cost estimation; QR-Decomposition; Pareto optimization; Coordinate Rotate digital computer

Depending from the specifications a huge design space, featuring up to thousands of possible implementations is available from the QRD-architecture template [1]. In order to support design-space exploration the parameterized cost model as well as routines for pruning (e.g. according to maximum latency), Pareto optimization etc. are implemented in a MATLAB-based optimization environment [2]. The execution time for a whole design space exploration (one set of specification parameters) is in the order of a few minutes only. Cost breakdown tables and figures can be generated automatically to get detailed insights to the cost contributions of the individual building

blocks in order to identify bottlenecks and to guide optimization [3-5]. Constraints can be set on the maximum clock frequency e.g. to avoid unreasonably high clock frequencies due to o -fold Coordinate Rotate Digital Computer (CORDIC) multiplexing or on selected clock frequencies being available on a SoC [6]. Arbitrarily high throughput rates can be achieved at almost unchanged ATE complexity and latency by multiplexing parallel QRD blocks [1]. Therefore, especially the area- and energy-optimization for less challenging throughput specifications is a valuable capability of this optimization environment [1]. Finally, it can be applied in early cost estimation to support system conception and design.

Just to give an idea of the capabilities of the optimization environment exemplary results are in Figure 1a which shows an example of AT - and ATE -design spaces for complex valued integer full ($[R]$ and $[Q]$) QRD of matrix size $N=12$, iteration count $M=16$ and word length $w=16$ [7]. The whole design space with 1,440 possible implementations is pruned for $ATE \leq 10 \cdot ATE_{min}$ in Figure 1b. The execution time for this example is 1 minute 42 seconds, only [8].

Figure 2 shows the AT as well as the ATE (insets) design spaces for different QRD specifications (applying extra delay stages in the PEs). The technology used is 40 nm CMOS technology. Here, for delay and output slopes SS-technology and slow-application (temperature) corner features were used while for energy and power features derived in FF-technology and fast-application corner were applied [9,10]. Even though no fabricated silicon die would feature this combination of cost figures, it is still the adequate worst case approach to ensure meeting specified figures. The back-biasing experiments were conducted assuming a reverse back-biasing voltage of $V_{BS} = -0.5V$ in order to reduce leakage power in the FF-corner and assuming a forward back-biasing voltage of $V_{BS} = +0.5V$. Volts in order to reduce critical path delay in the SS-corner. Supply voltage is 0.8V; word length and CORDIC iteration count are specified to be $w=M=16$. Energy figures are given for the case that no clock or power gating is applied.

In Figure 2a-2c the variation with matrix sizes from $N=14$ to $N=18$ for a real-valued, integer-data format QRD is shown. The corresponding total numbers of possible implementations are 960, 1,200, and 1,440. The AT -design spaces feature hyperbola-like Pareto-optimal fronts, offering trade-offs between throughput and silicon area [11]. The ATE -design spaces are pruned for implementations featuring an ATE

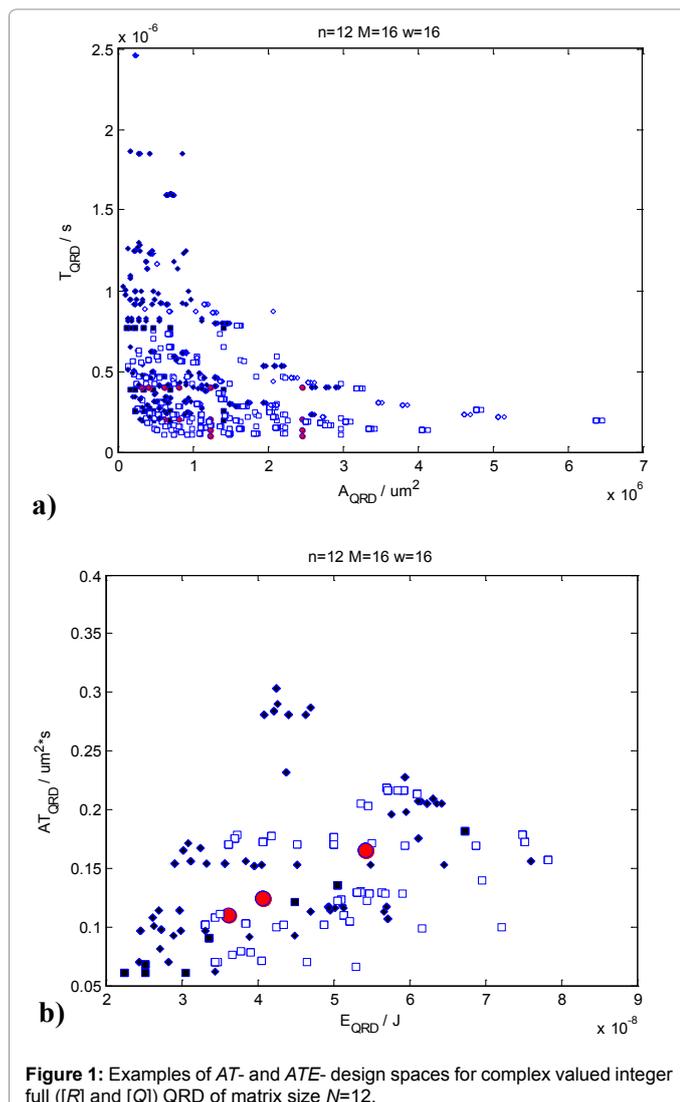


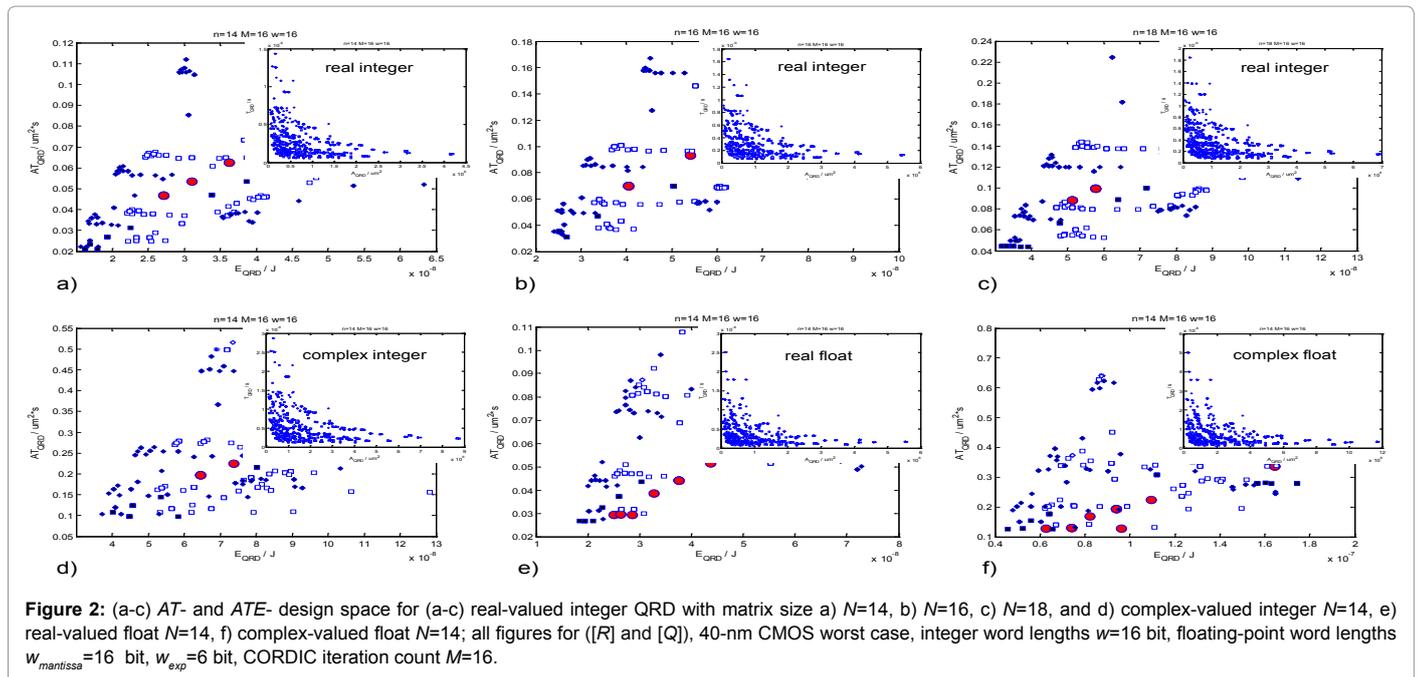
Figure 1: Examples of AT - and ATE - design spaces for complex valued integer full ($[R]$ and $[Q]$) QRD of matrix size $N=12$.

*Corresponding author: Upasna Vishnoi, Electrical Engineering and Computer Systems, RWTH Aachen University, Germany, Tel: +49241801; E-mail: vishnoiupasna@gmail.com

Received January 08, 2018; Accepted January 18, 2018; Published January 25, 2018

Citation: Vishnoi U (2018) Highly-Efficient Number-Crunching-Performance SoC Macros for Singular Value Decomposition. J Electr Electron Syst 7: 251. doi: 10.4172/2332-0796.1000251

Copyright: © 2018 Vishnoi U. 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.



not being smaller than one tenth of the optimum ATE . Filled markers depict carry-ripple and unfilled markers depict carry-select adder-based implementations. Latency optimal implementations are depicted by red-filled circles. Squares mark latencies up to two, and diamonds up to five times the minimum latency. For this word length, ATE -optimal implementations in the lower left corner are solely carry-ripple adder based [1].

Figure 2d-2f show the results for matrix size of $N=16$ for complex-valued/integer - $w_{mantissa}=16$ and the exponent word length is $w_{exp}=6$. As can be seen from comparing the ATE -optimal implementations, the overhead for floating-point data format extensions is rather small (in the order of 26 % for AT and 14 % for E). In contrast to that, the overhead for the extensions for complex-valued matrix processing is quite high: Both, area and period costs are increased by a factor of more than two, resulting in a 4.9 times larger AT -complexity. Energy also is increased by more than a factor of 2.3.

Acknowledgment

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

- Vishnoi U, Meixner M, Noll TG (2016) A Family of Modular QRD-Accelerator Architectures and Circuits Cross-Layer Optimized for High Area- and Energy-Efficiency. The Journal of Signal Processing Systems 83: 329-356.

- Vishnoi U, Noll TG (2013) A Family of Modular Area- and Energy-Efficient QRD-Accelerator Architectures. IEEE International Symposium on System-on-Chip (SoC Tampere), Finland.
- Vishnoi U, Noll TG (2013) Cross-Layer Optimization of QRD Accelerators. IEEE Proceedings of 39th European Solid-State Circuits Conference (ESSCIRC), Bucharest, Romania, pp: 263-266.
- Cavallaro JR, Elster AC (1991) A CORDIC Processor Array for the SVD of a Complex Matrix. Elsevier Science Publishers, pp: 227-239.
- Vishnoi U, Noll TG (2011) Area- and energy-efficient CORDIC Accelerators in Deep Sub-micron CMOS Technologies. Advances in Radio Science.
- Kung SY (1987) VLSI Array Processors. Upper Saddle River, New Jersey, USA: Prentice-Hall.
- Ercegovic MD, Lang T (2003) Digital Arithmetic (1st edn.), Elsevier.
- Senning Ch, Staudacher A, Burg A (2010) Systolic-Array based regularized QR-Decomposition for IEEE 802.11n Compliant Soft-MMSE Detection. International Conference on Microelectronics ICM.
- Salmela P, Burian A, Sorokin H, Takala J (2008) Complex-Valued QR Decomposition Implementation for MIMO Receivers. Proceedings in IEEE ICASSP, pp: 1433-1436.
- 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.
- Vishnoi U, Meixner M, Noll TG (2012) An Approach for Quantitative Optimization of Highly Efficient Dedicated CORDIC Macros as SoC Building Blocks. SOCC International Conference, pp: 242-247.