

## **Title: Using reconfigurable computing technology to accelerate matrix decomposition and applications**

**Abstract:** Matrix decomposition plays an increasingly significant role in many scientific and engineering applications. Among numerous techniques, Singular Value Decomposition (SVD) and Eigenvalue Decomposition (EVD) are widely used as factorization tools to perform principal component analysis for dimensionality reduction and pattern recognition in image processing, text mining and wireless communications, while QR Decomposition (QRD) and sparse LU Decomposition (LUD) are employed to solve the dense or sparse linear system of equations in bioinformatics, power system and computer vision. Matrix decompositions are computationally expensive and their sequential implementations often fail to meet the requirements of many time-sensitive applications.

The emergence of reconfigurable computing has provided a flexible and low-cost opportunity to pursue high-performance parallel designs, and the use of FPGAs has shown promise in accelerating this class of computation. In this research, we have proposed and implemented several highly parallel FPGA architectures to accelerate matrix decompositions and their applications in data mining and signal processing. Specifically, we describe the following contributions:

- We propose an efficient FPGA-based double-precision floating-point architecture for EVD, which can efficiently analyze large-scale matrices.
- We implement a floating-point Hestenes-Jacobi architecture for SVD, which is capable of analyzing arbitrary sized matrices.
- We introduce a novel deeply pipelined reconfigurable architecture for QRD, which can be dynamically configured to perform either Householder transformation or Givens rotation in a manner that takes advantage of the strengths of each.
- We design a configurable architecture for sparse LUD that supports both symmetric and asymmetric sparse matrices with arbitrary sparsity patterns.
- By further extending the proposed hardware solution for SVD, we parallelize a popular text mining tool-Latent Semantic Indexing an using FPGA-based architecture.
- We present a configurable architecture to accelerate Homotopy  $\ell_1$ -minimization, in which the modification of the proposed FPGA architecture for sparse LUD is used at its core to parallelize both Cholesky decomposition and rank-1 update.

Our experimental results using an FPGA-based acceleration system indicate the efficiency of our proposed novel architectures, with application and dimension-dependent speedups over an optimized software implementation that range from  $1.5\times$  to  $43.6\times$  in terms of computation time.