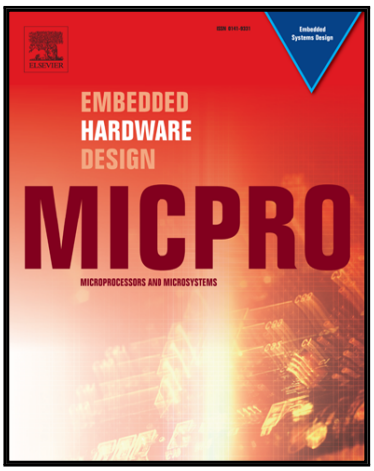

CAEMO - A Flexible and Scalable High Performance Matrix Algebra Coprocessor for Embedded Reconfigurable Computing Systems

Hendrik Woehrle, Frank Kirchner

PII: S0141-9331(16)30260-5
DOI: 10.1016/j.micpro.2017.10.005
Reference: MICPRO 2623

To appear in: *Microprocessors and Microsystems*

Received date: 21 October 2016
Revised date: 27 September 2017
Accepted date: 19 October 2017

Please cite this article as: Hendrik Woehrle, Frank Kirchner, CAEMO - A Flexible and Scalable High Performance Matrix Algebra Coprocessor for Embedded Reconfigurable Computing Systems, *Microprocessors and Microsystems* (2017), doi: 10.1016/j.micpro.2017.10.005

CAEMO - A Flexible and Scalable High Performance Matrix Algebra Coprocessor for Embedded Reconfigurable Computing Systems

Hendrik Woehrle[a], Frank Kirchner[a,b]

[a] *German Research Center for Artificial Intelligence, Robotics Innovation Center (RIC), Robert-Hooke-Str. 1, Bremen D-28359, Germany*
[b] *Robotics Group, Department of Mathematics and Computer Science, University of Bremen, Robert-Hooke-Str. 1, Bremen D-28359, Germany*

Abstract

Many applications in mobile and embedded systems like signal processing, machine learning, kinematics, dynamics, and control depend on computationally expensive matrix operations. However, such systems underlie tight constraints regarding power consumption and physical space, which prohibits the usage of powerful multicore systems. In this paper, we propose a novel scalable and power-efficient architecture for matrix algebra in FPGA-based Systems-on-Chip. The architecture is based on a linear systolic array and has been developed with a focus on flexibility in order to be adapted to different applications. We evaluate the performance, resource utilization and power consumption of different configurations and show that it provides significant speed-ups over a mobile processor and is significantly more power efficient than a standard PC.

Keywords: Matrix algebra, Hardware acceleration, Embedded systems, FPGA

1. Introduction

Dense matrix algebra is a building block in many applications that are becoming increasingly important for embedded systems. Robots, for example, utilize methods from the fields of machine learning, image processing or dynamics, control, and kinematics. Matrix operations occur frequently in the underlying algorithms. However, especially matrix multiplication is a computationally expensive operation with cubic time complexity. At the same time, we can observe an increase in the capabilities of sensors and complexity of actuated systems and, hence, a demand for powerful computing systems in order to meet the performance requirements. To keep up with this development, future computing systems have to rely on parallelism [1]. Unfortunately, it is often

Email addresses: hendrik.woehrle@dfki.de (Hendrik Woehrle), frank.kichner@dfki.de (Frank Kirchner)