CCIT REPORT #724 March 2009

The Design of a Latency Constrained, Power Optimized NoC for a 4G SoC

Isask'har Walter[1], Israel Cidon[2], Avinoam Kolodny[2]
Electrical Engineering Department, Technion – Israel Institute of Technology, Haifa 32000, Israel
Rudy Beraha[3]
Qualcomm Corp. Research and Development, San Diego, California 92121, USA
[1]zigi@tx.technion.ac.il , [2]{cidon, kolodny}@ee.technion.ac.il [3]rberaha@qualcomm.com

Abstract

In this paper, we examine the process of porting a high-end commercial SoC application from a segmented bus implementation to a NoC-based one. We present several design choices and focus on the power optimization of the NoC while achieving the required performance. Our design steps include module placement optimization using simulated annealing and allocation of minimal and different capacities to links. Unlike previous studies, in which point-to-point, per-flow timing constraints were used, we introduce and demonstrate the importance of using the application's end-to-end traversal latency requirements during the optimization process. In order to quantify and evaluate the different alternatives, we report the actual throughput and timing requirements of the commercial SoC as well as the synthesis results. According to our findings, the proposed technique offers up to 40% savings in the total router area and a reduction of up to 49% in the inter-router wiring area.

1. Introduction

Application-specific systems on-chip (SoC) make extensive use of busses as the interconnect infrastructure. These busses are typically enhanced along product generations to match the increasing needs of the application. Such enhancements include increasing the bus frequency and width as well as enriching the bus semantics and transfer modes. By avoiding fundamental changes, the SoC architects can leverage their past experience in designing shared busses and successfully overcome the growing complexity of the design. However, in recent years research has shown that Network on-Chip (NoC) is likely to replace busses in future SoCs, due to superior performance, power and area tradeoffs it offers as the number of modules increases [1][2][3][4]. This is mainly attributed to the spatial parallelism and statistical multiplexing of networks, to their short, unidirectional point-to-point wires and to their scalable architecture [5].

NoCs are being adopted by companies as a means to improve design productivity. As the number of modules connected to a bus increase, the physical implementation of the bus becomes very complex, and achieving the desired throughput and latency requires time consuming custom modifications. Conversely, NoCs are designed separately from the functional units of the system to handle all foreseen inter-module communication needs. Their inherent scalable architecture facilitates the integration of the system and shortens the time-to-market of complex products.

In this work, we discuss and evaluate the design process of a NoC for a state-of-the-art SoC. More specifically, we describe our experience in porting a high-performance, power constrained 4G wireless modem application from a segmented bus based architecture to a cost optimized NoC architecture. As the design process includes many "degrees of freedom" creating a very large design space, finding the absolute optimal solution is an extremely difficult problem. Instead, we focus on some of the important choices that should be made by the system architect while selecting some well-accepted, practical solutions to other questions.

Previous work that has dealt with the design process of the NoC frequently attempted to minimize power consumption and/or maximize network performance as measured by the network's throughput and latency. When real applications are considered, simply minimizing the power consumption alone (e.g., by module placement) is impossible, as performance constraints for each given application are to be met. Similarly, maximizing performance alone is inefficient, since excessive power might be used for improving performance beyond the needs of the application. Therefore, in this paper we look for a tradeoff between the power and performance of the NoC that is characterized by a minimal power consumption that still meets the demands of all targeted applications.