CCIT Report #677 December 2007

BENoC: A Bus-Enhanced Network on-Chip

Isask'har Walter[1], Israel Cidon[2], Avinoam Kolodny[2]
Electrical Engineering Department, Technion – Israel Institute of Technology, Haifa 32000, Israel
[1]zigi@tx.technion.ac.il , [2]{cidon, kolodny}@ee.technion.ac.il

Abstract

Recent research has shown that Network on-chip (NoC) is superior to a bus in terms of power and area for given traffic throughput requirements. Consequently, NoC is expected to be the main interconnect infrastructure in future System on Chip (SoC) and chip multi-processor (CMP). Unlike off-chip networks, VLSI modules are only a few millimeters apart, hence the cost of off-network communication among the network end-points and routers is quite low. Such off-network communication can circumvent weaknesses of the NoC, such as latency of critical signals, complexity and cost of broadcast operations, and operations requiring global knowledge or central control.

In this paper we explore the benefits of adding a low latency, customized shared bus as an integral part of the NoC architecture. While the bus is inferior to NoC in terms of data throughput, it possesses two main advantages: First, the bus is inherently capable to broadcast information. Second, the bus has lower and more predictable propagation latency. Therefore, the bus is superior to a multi-hop network for certain transactions such as broadcast of queries, fast delivery of control signals, quick exchange of small data items, network configuration and power management. Moreover, custom properties can be tailored to this particular bus in order to facilitate these specialized tasks. As a result, the Bus-enhanced NoC (BENoC) is overall more cost-effective than a traditional "busless" NoC.

We describe several applications of bus-enhanced networks, such as cache lines lookup and coherency in CMP and efficient management of SoC resources. We present an analytical comparison of the power saving in BENoC versus a network providing similar services. Finally, simulation is used to evaluate the performance of BENoC in a chip multiprocessor system which employs a distributed cache with dynamic non-uniform cache access (DNUCA).

Categories and Subject Descriptors

System-Level Design and Co-Design: Network-on-Chip (NoC)

General Terms

Performance, Design

Keywords

Network on-Chip, resource management, SoC, NoC support for CMP/MPSoC

1. Introduction

There is a large body of work advocating the use of spatial reused networks as the main on-chip interconnection infrastructure (e.g., [1]-[4]). Network architecture has been shown to be more cost effective than a system bus in terms of area, power and performance [5]. In addition, networks generally have good scalability properties, while shared busses cannot withstand the increasing bandwidth and performance requirements already seen in contemporary systems. Consequently, current state-of-the-art VLSI research often presents NoC as the practical choice for future systems. However, conventional interconnect architectures which solely rely on a network have several drawbacks when advanced services are required. In particular, the distributed nature of a network is an obstacle when global knowledge or operation is beneficial. For example, broadcast (sending information to all modules on the chip) is an inherent operation in busses and has no extra cost. However, in a typical NoC a broadcast capability either involves additional hardware mechanisms or a massive duplication of unicast messages. Broadcast is particularly expensive in NoCs that employ wormhole switching [6], as classic wormhole does not support broadcast due to the complexity of the backpressure mechanism and the requirement for small buffers. Similarly, multicast is considerably easier to implement in busses than in typical networks. Finally, multi-hop networks impose an inherent packet propagation latency for the communication between modules. This complicates the design of critical signals between remote modules. Bus properties are also valuable when global knowledge and control are useful. As current NoC implementations are strictly distributed (heavily borrowing concepts from traditional large scale networks), the system behavior and performance is often dictated by multiple local decisions. For example, arbitration for scarce resources is typically