CCIT Report #565
December 2005

Zooming in on Network-on-Chip Architectures*

Israel Cidon
Dept. of Electrical Engineering
Technion, Haifa 32000, Israel
cidon@ee.technion.ac.il

Idit Keidar
Dept. of Electrical Engineering
Technion, Haifa 32000, Israel
idish@ee.technion.ac.il

ABSTRACT

The aim of this paper is to expose the networking community to the concept of *network-on-chip (NoC)*, an emerging field of study within the VLSI realm, in which networking principles play a significant role, and new network architectures are in demand. Networking researchers will find new challenges in exploring solutions to familiar problems such as network design, routing, and quality-of-service, in unfamiliar settings under new constraints. We present a new classification of chip architectures into three categories with different requirements from their NoCs. In order to stimulate some specific research directions, we highlight research problems arising in each of these categories, focusing on routing and resource allocation (e.g., capacity assignment). We provide initial solution directions to example problems.

1. INTRODUCTION

As VLSI technology becomes smaller, and the number of modules on a chip multiplies, on-chip communication solutions are evolving in order to support the new inter-module communication demands. Traditional solutions, which were based on a combination of shared-buses and dedicated module-to-module wires, have hit their scalability limit, and are no longer adequate for sub-micron technologies [4, 12, 8, 23, 14, 15]. Current chip designs incorporate more complex multi-layered and segmented interconnection buses [1, 16, 30]. More recently, chip architects have begun employing on-chip network-like solutions [12, 13, 8, 24, 2, 17, 5]. This evolution of on-chip interconnects may evoke feelings of *déjà vu* among networking old-timers. We believe that the considerations that have driven data communication from shared buses to packet-switching networks (spatial reuse, multi-hop routing, flow and congestion control, and standard interfaces for design reuse, etc.) will inevitably drive VLSI designers to use these principles in on-chip interconnects. In other words, we can expect the future chip design to incorporate a full-fledged *network-on-a-chip (NoC)*, consisting of a collection of links and routers and a new set of protocols that govern their operation. In Section 2, we survey the reasons for the inevitable shift to NoCs in the VLSI world, while exposing the most important requirements from a NoC. We note that although some recent papers have begun to design such on-chip network architectures, the field is still in its infancy, and many challenges have yet to be tackled.

In designing a NoC, one has to address all the classical networking issues. Addressing and routing schemes need to be devised in order to allow packets traversing the same links to be routed to diverse destinations. Names meaningful to applications (such as memory and I/O addresses) need to be translated into routing efficient labels. Since the timely delivery of certain types of traffic (or signals) on the chip is crucial for performance, support for multiple quality-of-service (QoS) requirements is also essential [5, 10, 28]. Similarly, a NoC should support network level congestion control in order to accommodate excessive traffic conditions. Where congestion control is employed, fairness issues need to be considered as well. One also needs to address reliability in the face of communication soft errors that may corrupt transmitted data [28, 32]; this can be done using a combination of error-correction codes and retransmission mechanisms.

Since problems of this type have been extensively studied in the networking realm, as well as for off-chip interconnection networks [9], one may be tempted to employ well-developed networking/interconnection solutions in the NoC context. Nevertheless, a direct adaptation of network protocols to NoCs is impossible, due to the different communication requirements, cost considerations, and architectural constraints. The primarily considerations in VLSI are minimizing power dissipation and area. This has a number of implications on NoC design. First, NoC components should be extremely simple, so as to allow implementing them with a small number of logic gates and to expend as little energy as possible. In addition, power considerations render shortest-path routes highly desirable, while area considerations dictate the use of small routing tables.

Beyond the distinctive cost considerations, the requirements from a NoC also differ from their off-chip counterparts. For example, on-chip network topologies are quite restricted– they are laid onto planar layers (in silicon and/or metal), and are therefore often organized as (possibly partial) grids. Thus, elaborate layouts like high-dimension hypercubes and butterflies, which are often employed in interconnection networks [9], are not cost-effective for NoCs. Moreover, NoC topologies are fixed throughout their lifetimes. That is, they do not need to support the dynamic addition or removal of network-attached modules. Furthermore, the NoC is synthesized anew for each design [27, 5, 3], eliminating the need for standard network protocols; i.e., there is no advantage in backward compatibility of NoC protocols and architectures employed in a new chip designs with those used in previous designs, beyond the use of standard network *interfaces* to allow the reuse of modules, called *IP cores*, across

*This research is partially supported by Intel Corporation and Semiconductor Research Corporation.