TOPOLOGY-AWARE QUALITY-OF-SERVICE SUPPORT IN HIGHLY INTEGRATED CHIP MULTIPROCESSORS

[1]Kalyani Bhoir
M.E, E&TC – Signal processing, Dr. D. Y. Patil School of engineering, Ambi, Pune
Email:[1]Kalyanibhoir21@gmail.com

Abstract— With increase in integration density and complexity of the system-on-Chip (SOC), the conventional interconnects are not suitable to fulfill the demands. As CMP-level quality of service(QOS) support becomes necessary to provide performance isolation, service grantees & security. The application of traditional network technologies in the form of Network-on-Chip is a potential solution. NoC design space has many variables. Selection of a better topology results in lesser complexities and better power-efficiency. In the proposed work, key research area in Network-on-chip design targeting communication infrastructure specially focusing on optimized topology design is worked upon. The simulation is modeled using a conventional network simulator Network simulator-2 (NS-2), in which by selecting proposed Topology gives reduction in traversing the longest path is observed. We evaluate several topologies for the QOS-enabled shared regions, focusing on the interaction between network-on-chip (NOC) and QOS metrics. We explore a new topology called Destination Partitioned Subnets (DPS), which uses a light-weight dedicated network for each destination node. On synthetic workloads, DPS nearly matches or outperforms other topologies with comparable bisection bandwidth in terms of performance, area overhead, energy- efficiency, fairness, and preemption resilience.

Index Terms— NoC, SoC, Routing,NS-2 simulator.

I. INTRODUCTION

The abrupt emergence of multi-core chips and their rapid proliferation have left researchers and industry scrambling for ways to exploit them. Two notable paradigms have arisen for monetizing CMPs – server consolidation and cloud computing. The former allows businesses to reduce server costs by virtualizing multiple servers on a single chip,thereby eliminating dedicated hardware boxes for each individualserver. The latter enables delivery of various clientservices from remote (i.e., “cloud”) servers. Since a single CMP can serve multiple users concurrently, hardware, infrastructure and management costs are reduced relative to a model where each user requires a dedicated CPU Unfortunately, these novel usage models create new system challenges and vulnerabilities. For instance, in a consolidated server scenario, different priorities may be assigned to different servers.

Thus, web and database servers for external customers could have a higher priority than intranet servers. But as multiple virtualized servers may be executing concurrently on a multi-core chip, traditional OS level preemptive scheduling policies can fail at properly enforcing priorities of different VMs competing for shared resources. In a cloud setting, multiple users may be virtualized on to a common physical substrate, creating a number of new concerns, including inadvertent interference among the different users, deliberate denial-of-service attacks, and side channel information leakage vulnerabilities. Researchers have recently demonstrated a number of such attacks in areal-world setting on Amazon’s EC2 cloud infrastructure, highlighting the threat posed by chip-level resource sharing on a public cloud.

Today's CMPs lack a way to enforce priorities and ensure performance-level isolation among the simultaneously executing threads. Inter-thread interference may occur in any of the shared resources present on a CMP, including caches, memory controllers, and the on-chip network. Researchers have suggested using on-chip hardware quality-of-service (QOS) mechanisms to enforce priorities, limit the extent of interference, and provide guarantees for threads sharing a substrate.

In this work, we take a network-centric, topology-aware approach to chip-level quality-of-service. To reduce performance, area, and energy overheads of network-wide QOS support, we propose to isolate shared resources, such as memory controllers and accelerator units, into dedicated regions of the chip. Hardware QOS support in the network and at the end-points is provided *only* inside these regions.

The focal point of this paper is the organization of the shared region.

We evaluate Destination Partitioned Subnets (DPS), a new topology we propose in this work. DPS uses a dedicated sub network for each destination node, enabling complexity-effective routers with low delay and energy overhead. All topologies show good fairness and experience little slowdown in the face of adversarial workloads with high preemption rates. On synthetic workloads, DPS consistently matches or outperforms mesh-based topologies in terms of performance, energy efficiency,and preemption resilience.

As the network communication latency depends on the characteristics of the target application, computational elements and network characteristics (e.g. network bandwidth and buffer size [2]. First of all the target applications and their associated traffic patterns and bandwidth requirements for each node in the network is determined. This application partitioning and knowledge of overall system architecture significantly impact the network traffic and helps determine the optimal network topology. Optimal network topology creates immense impact of design cost, power and performance and helps designers to choose effective and efficient routing algorithms and flow control scheme to manage incoming traffic. The design space of a NoC is very large, and includes topology choice (mesh, torus, star, etc.), circuit switched or packet switched, and other parameters (link widths, frequency, etc.). Because the traffic patterns of most SoCs can be known, a custom generated network topology and physical placement of components yields better performance and power than a regular-pattern network [4]. A NoC's buffers and links can consume near 75% of the total NoC power [5], thus there is significant benefit to optimizing buffer size, link length and bandwidth of a NoC design. Generally speaking, determining the optimal topology to implement any given application does not have a known theoretical solution. Although the synthesis of customized architectures is desirable for improved performance, power consumption and reduced area, altering the regular grid-like structure brings into the picture significant implementation issues, such as floor planning, uneven wire lengths (hence, poorly controlled electrical parameters), etc. Consequently, Exploring Alternative Topologies for Network-on-Chip Architectures ways to determine efficient topologies that trade-off high-level performance issues against detailed implementation constraints at micro- or nano-scale level need to be developed.

II. BACKGROUND

The early work and basic principles of NoC paradigm were outlined in various seminal articles, for example [7-17] and few text books [18-20]. However, the aforementioned sources do not present many implementation examples or conclusions. Networking concepts from the domains of telecommunication and parallel computer do not apply directly on chip. From a networking perspective, they require adaptation because of the unique nature of VLSI constraints and cost e.g. area and power minimization are essential; buffer space in on-chip switches are limited, latency is very important, etc. At the same time, there are new degrees of freedom available to the network designer, such as the ability to modify the placement of network endpoints. From the view point of VLSI designer, many well understood problems in the real aim of chip development methodology get a new slant when they are formulated for a NoC based system, a new trade-offs need to be comprehended. Therefore, the field offer opportunities for noble solutions in network engineering as well as system architecture, circuit technology, and design automation. [6] Current complex on-chip systems

are also modular, but most often the modules are interconnected by an on-chip bus. The bus is a communication solution inherited from the design of large board- or rack-systems in the 1990's. It has been adapted to the SoC specifics and currently several widely adopted on- chip bus specifications are available [31-34]. While the bus facilitates modularity by defining a standard interface, it has major disadvantages. Firstly, a bus does not structure the global wires and does not keep them short. Bus wires may span the entire chip area and to meet constraints like area and speed the bus layout has to be customized [35]. Long wires also make buses inefficient from an energy point of view [36]. Secondly, a bus offers poor scalability. Increasing the number of modules on-chip only increases the communication demands, but the bus bandwidth stays the same. Therefore, as the systems grow in size with the technology, the bus will become a system bottleneck because of its limited bandwidth. Recently, network-on-chip (NoC) architectures are emerging as a candidate for the highly scalable, reliable, and modular on-chip communication infrastructure platform [11]. The NoC architecture uses layered protocols and packet-switched networks which consist of on-chip routers, links, and network interfaces on a predefined topology. There have been many architectural and theoretical studies on NoCs such as design methodology [10], [11], topology exploration [21], Quality-of- Service (QoS) guarantee [22], resource management by software [23], and test and verifications [24]. In large-scale SoCs, the power consumption on the communication infrastructure should be minimized for reliable, feasible, and cost-efficient implementations. However, little research has reported on energy- and power-efficient NoCs at a circuit or implementation level, since most of previous works have taken a top-down approach and they did not touch the issues on a physical level, still staying in a high-level analysis. Although a few of them were implemented and verified on the silicon [25], [26],they were only focusing on performance and scalability issues rather than the power-efficiency, which is one of the most crucial issues for the practical application to SoC design.

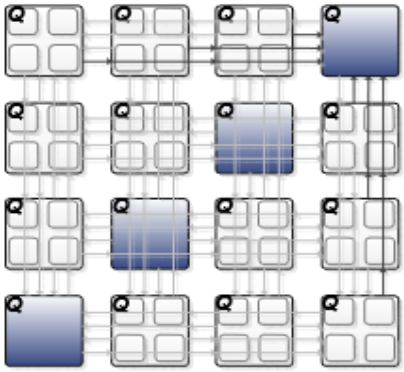
III METHODOLOGY

Network-on-Chip is a new paradigm for interconnecting today's heterogeneous IP cores based System-on-Chips (SoCs). In SoC's IP Cores are connected to network of routers using network interfaces and network is used for packet switched on-chip communication. Conventional computer design tools i.e. Network Simulator-2 utility are used for network design and simulation. It provides a versatile practice and visualization environment for the design, configuration, and troubleshooting of network environments. The work done by us uses same tool to compare two topologies. The 2-D mesh is currently the most popular regular topology used for on-chip networks in tile-based architectures, because it perfectly matches the 2-D silicon surface and is easy to implement. However, a number of limitations have been proved in the open literature, especially for long distance traffic. In this type of topology, every node has a dedicated point to point link to every other node in the network. This means each link carries traffic only between the two nodes it connects. If N is total no of nodes in network. Number of links to connect these nodes in mesh & DPS = N (N-1)/2 Each node should have (N-1) I/O ports as it require connection to every another node. The advantages are:

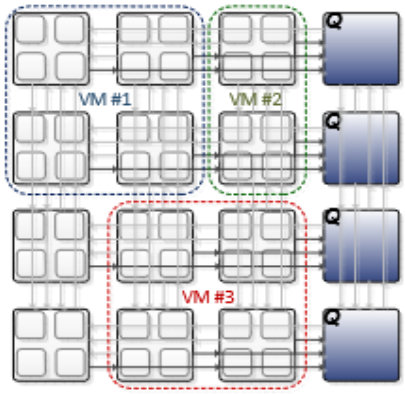
- No traffic problem as there are dedicated links. Robust as failure of one link does not affect the entire system.
- Security as data travels along a dedicated line.
-Points to point links make fault identification easy.

Disadvantages are:

-The hardware is expansive as there is dedicated link for any two nodes and each device should have (N-1) I/O ports.
- There is mesh of wiring which can be difficult to manage.
- Installation is complex as each node is connected to every node. Also in DPS topology Figure 1(b) shows a diagram of a scaled down 4x4 grid with a similar organization. One column in the middle of the grid is devoted to shared resources with one terminal per node; the rest of the network employs 4- way concentration.



(a) Baseline QOS –enabled approach



a)Topology Aware QOS Approach

As earlier studies have shown that maximum power is consumed by links and interconnect infrastructure. Reducing interconnects and links will result in lower power consumption but can also affect the performance and reliability negatively. The topology suggested by us reduces the number of links thus resulting into lower power consumption keeping same level of reliability and performance level

IV SIMULATION

Network Simulator Ns-2 The simulator, ns-2, has facilities to describe network topology, network protocols, routing algorithms and communication traffic generation. It provides basic TCP and UDP as the network transmission protocols, four routing strategies (Static, Session, Dynamic and Manual) and many mechanisms for modelling traffic generation. It is possible to generate a traffic at random, by burst or with bias towards destinations. Additionally, the simulator has the possibility of incorporating protocols, routing algorithms and traffic generation defined by the user. The simulator is written in C++ and uses OTcl (Object Tool Command Language) for building command and configuration interfaces. The source code of ns-2 is also available[5]. Ns-2 provides well documented trace format for interpreting simulation results. A graphical animator tool, nam (Network AniMator), is also built into ns-2 for user's friendly visualization of the flow of messages and the whole system simulated. In this paper, a generic NOC architecture would be modelled and simulated in ns-2 with only built-in options. Tcl is used for specifying the NOC simulation model and running the simulation.

V CONCLUSION

The results achieved in terms of time and reduction in number of links displayed here is encouraging and motivates us to take the work further. As discussed earlier the NoC technology can borrow the tools and techniques from conventional computer network technology with required customization. In our future work, we intend to test same on a standard NoC benchmark. The other design parameters on NoC will also be explored.

VI REFERENCES

[1]. M. Horowitz, R. Ho, and K. Mai, “The future of wires”, Proc. IEEE, vol. 89, pp. 490-504, Apr 2001.

[2]. Radu Marchulescu et al. outstanding research problems in NOC design: system, micro architecture and circuit perspectives IEEE transactions on computer aided design of integrated circuits and systems vol. 8 , No.1,Jan 2009.

[3]. R.Dafali, J-Ph.Diguet and M.Sevaux, “Key Research Issues for Reconfigurable Network-on-Chip”, 2008 IEEE International Conference on Reconfigurable Computing and FPGAs

[4]. Murali, S., P. Meloni, F. Angiolini, D. Atienza, S. Carta, L. Benini, G. D. Micheli and L. Raffo, “Design of application-specific networks on chips with floorplan information,” in: Proc. of the International Conf. on Computer-Aided Design, 2006.

[5]. Srinivasan, K. and K. S. Chatha, “A low complexit heuristic for design of custom network-on-chip architectures”, in: Proceedings of DATE, 2006

[6]. Avinoam Kolodny & Li-Shiuan Peh, IEEE transactions on VLSI systems, vol 17, no. 3 , Mar 2009

[7]. D. Wiklund and D. Liu, "Switched Interconnects for System-on-a-Chip Designs," in IP2000, Oct 2000, pp. 192-198.

[8]. P. Guerrier and A. Greiner, "A Generic Architecture for On-Chip Packet-Switched Interconnections," in DATE March 2000, pp. 250-256.

[9]. A. Hemani, A Jantsch, S. Kumar, J. Oberg, M. Millberg, and D. Lindqvist , "Network-on-Chip: An Architecture for Billion Transistor Era," in Norchip, 2000

[10]. W. Dally and B. Towles, "Route Packets, Not Wires: On-Chip Interconnection Networks," in DAC, June 2001,
pp. 684-689

[11]. L. Benini and G.de Micheli, "Networks on chip: A New SoC Paradigm," IEEE computer Vol 35, No. 1 pp. 70- 78, Jan 2002

[12]. S. Kumar et al," A Network-on-Chip Architecture and Design Methodology," in VLSI, Apr. 2002, pp. 105-112

13]. P. Wielage and k. goossens, "Networks on Silicon: Blessing or Nightmare?," in DSD, Sep 2002, pp. 196- 200

[14]. J. P. Soininen , A. Jantsch, M. Forsell A. Pelkonen, J.Kreku and S. Kumar , "Extending Platform-based Design to Network –on-Chip Systems," in VLSI Design, Jan 2003, pp 401-408 S. No .Source Node Destination Node Inter m-ediate Node Time Elapsed (sec) 1 R1 R18 R10 0.003 2 RI R18 R15 0.004 3 R1 R18 R16 0.005 4 R1 R18 R17 0.007 5 R1 R18 R18 0.009 S.No. Source Node Destination Node Interm- ediate Node Time Elapsed (sec) 1 R1 R18 R17 0.001 2 RI R18 R16 0.002 3 R1 R18 R18 0.003 BIJIT - BVICAM's International Journal of Information TechnologyCopyRight©BIJIT–2011;
July–December,2011;Vol.3No.2;ISSN0973–5658
376

[15]. J. Hankel , W. Wolf and S. Chakradhar, " On-Chip Networks: A Scalable Communication-Centric Embedded System Design Paradigm," in VLSI Jan. 2004, pp 845-851.

[16]. U. Y. Ogras, J. Hu, and R. Marculescu, " Key Research Problems in NoC Design: A Hollistic Prespective," in CODES , Sep. 2005, pp. 69-75.

[17]. J. Owens et al , " Research Challenges for On-Chip Interconnection Networks," IEEE Micro, vol 27,no. 5, pp. 96-108, Sep-Oct 2007.

[18]. A. Jantsch and H. Tenhunen , Eds. , "Networks on Chip," Dordrecht , The Netherlends: Kluwer Academic Publishers, 2003.

[19]. J. Nurmi, H. Tenhunen, J. Isoaho, and A. jantsch Eds. "Interconnects – Centric design for Advanced SoC and NoC," Dodrecht , The Netherlends: Kluwer Academic Publishers, 2004.

[20]. "Networks on Chip: Technology and Tools," Morgan Kaufmann, 2006.

[21]. D. Bertozzi et al., "Xpipes: A network-on-chip architecture for gigascale system-on-chip," IEEE Circuits Syst. Mag., vol. 4, no. 2, pp. 18–31, 2004.

[22]. E. Rijpkema et al., "Trade-offs in the design of a router with both guaranteed and best-effort services for networks on chip," in Proc. Des., Autom. Test Europe Conf. Mar. 2003, pp. 350–355.

[23]. V. Nollet et al., "Operating-system controlled network on chip," in Proc. Des. Autom. Conf., Jun. 2004, pp. 256–259

[24]. J.-S. Kim et al., "On-chip network based embedded core testing," in Proc. IEEE Int. SoC Conf., Sep. 2004
,pp. 223–226

[25]. S.-J. Lee et al., "An 800 MHz star-connected on-chip network for application to systems on a chip," in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2003, pp. 468–469

[26]. M. Taylor et al., "A 16-issue multiple-program- counter microprocessor with point-to-point scalar operand network," in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2003, pp. 170

[27]. Jongman Kim Dongkook Park Nicopoulos, C. ijaykrishnan, N. Das, C.R., "Design and analysis of an NoC architecture from performance, reliability and energy perspective", symposium on architecture for etworking and communication systems, ANCS 2005, pp. 173-182 .