 ISBN: 978-81-927230-0-9

Quality of Service for Improving the Performance of Network-On-Chip : A Review

Jaya R. Surywanshi

Department of Electronics Engineering
G. H. Raisoni College of Engineering, Nagpur, India
abhijaya19@gmail.com

Dr. Dinesh V. Padole

Department of Electronics Engineering
G. H. Raisoni College of Engineering, Nagpur, India
dinesh.padole@raisoni.net

***Abstract* — SoCs typically consist of several intellectual IP blocks like accelerators imaging, video decoders, graphics, etc. including with general purpose cores. In this large heterogeneous architecture, the requirements placed by the processing on the core versus the processing on the IP block. Traditionally, bus mechanism is preferred for communication to all devices presents on it. Existing on-chip interconnection networks are usually implemented using buses. Bus-based mechanism is preferred for interconnection in SoC such as Avalon, AMBA, CoreConnect, STBus, Wishbone etc. buses are used in SoC architectures to share resources . Bus based mechanism have several interconnect issues that have direct impact on the performance of SoC. Thats why new mechanism are required for better performance as well improving the QoS of SoC . The enactment of NoC as the communication system for complex integrated systems and has been promoted by the increasing number of processing elements integrated in current MPSoCs. Quality-of-Service becomes a vital requirement in SoCs with NoCs. The Quality of Service NoC presents a preferment solution that provides high throughput and low latency transfers. Researchers have introduced different methods and techniques to support QoS. This paper presents a review on the existing work done by different authors on QoS parameters which tries to improve the performance of the NoC.**

***Keywords*— NoC,SoC,QoS**

I. INTRODUCTION

System on-Chip (SoC) is designed of heterogeneous cores on a single chip which contains the hardware as well as software and controlling various devices like microprocessor, microcontroller, DSP cores, plus different peripherals and interfaces devices. As the microprocessor and micro controller industry is moving from single-core to multi-core and eventually single-core to many-core architectures, containing tens to hundreds or thousands of identical cores arranged as chip multiprocessors, which is also required efficient and effective communications among processors. Both microprocessor and SoC needs a high-performance, scalable, flexible, user and design-friendly interconnection [1]. To provide effective and efficient communication poses a big challenge to researchers. Before the appearance of network-on-chip system, interconnection mechanism are totally based on shared buses or dedicated wires in SoC. Dedicated wires mechanism provide point-to-point connection between all IP cores and every pair of nodes. This connection is effective for small systems of a few cores. But day-by-day as the number of cores increases in SoC, the number of wires in the point-to-point architecture grows greatly and making it difficult to scale.

By comparing between the dedicated wires and the IP core and node, a shared bus should be more scalable and reusable. Share bus which is a set of wires shared by multiple cores . However, due to the inherent limitation of buses, at a time only one communication transaction is allowed and blocking communication for all other cores. The disadvantages of shared bus architectures are included high energy consumption, long data delay, widely varying latencies, widely varying data transfer rates increasing complexity in decoding/arbitration and low bandwidth [2]. It also creates a communication problem. It is very difficult to manage if hundreds of nodes are connected by shared buses. So the preferable use of shared buses is only when the limited or a few dozens of IP cores are connected. To mange with the problems in shared buses, a different hierarchical architecture is introduced which segments bus into shorter ones. This implemented bus architectures may provide the solution at some certain levels to the problems that faced by dedicated wires and shared buses. Since different buses may chronicle for different protocols, bandwidth needs, and also increase communication parallelism. Still scalability remains a major problem for hierarchical bus architectures. In order to meet the present communication requirements, on chip traffic transport and management challenges, accelerate time-to-market and minimize the communication energy consumption of huge scale SoCs. There is a great demand to find a new design alternative to the conventional bus based and point-to-point communication based architectures [1]. Network-on-Chip (NoC) cover the limitation of traditional bus-based architectures. It has been proposed as a highly scalable and structured solution to address communication problems for SoC. It has several advantages over dedicated wiring and buses mechanism system i.e. low-power consumption , high-bandwidth, low-latency, less space requirement and scalability. NoC architectures is a assured communication solution mechanism with a pre-specified clock rate regardless of the network size, which is not feasible for bus-based architectures. To overcome the limitation of conventional bus-based architectures, a number of research work and studies have demonstrated the feasibility and advantages of Network-on-Chip (NoC). Since advantages and because of the new concept of NoC, it has drawn great attention from researchers from all over the world. But to fully and proper explore to the benefits of NoC, numerous challenges and open problems are to be consign.

Open problems can be classified into four main categories, including application modeling and optimization, NoC communication architecture analysis and optimization, Architecture Evaluation, NoC Communication, and NoC Design Validation and Synthesis.

NoC support the different application domain. Few applications are Multimedia application, Control application, Aerospace application, Telecommunication application, Real time application, Military, Healthcare, Networking, High performance computing application, signal processing etc. Because of that requirement it is needed a predictable and desirable quality of service (QOS) to all application point of view , even for future application demand basis. System-on-chip designers use networks on chip architecture to solve wide submicron problems, and to divide global problems into local, decoupled problems. NoCs provide IP re-use and platform based design, services through protocol stacks, and introducing guaranteed services. It also provides globally predictable behavior, as required by the user, when combining local, decoupled solutions. There are many levels of QoS commitment like completion, correctness, completion bounds etc. but it will increase the cost of NoC. There are number of different QoS improvement methods like QoS can be improved with traffic shaping techniques such as Packet prioritization and packet classification, Application classification ,Queuing at congestion points, Priority based resources allocation, Connection establishment support, Congestion control mechanism, Application Specific Optimization, Architecture Optimization etc .This work evaluate different presents methods which provide QoS for NoCs. NoC's still does not present efficient solution to provide QoS to any application. Through this work we will study the limitation of all the present methods and try to overcome this gap through our work. Outline for this paper is as follows: After the introduction, we review some existing work done by different authors in section 2. Then we discuss the our research problem in section 3, In section 4 objective and scope of the research work and finally conclusion.

II. LITERATURE SURVEY

NoC design use different methods to provide QoS. Major considerable methods are providing support to circuit switching for all IPs or for selected IPs and second one is making available priority scheduling for packet transmission.

This Section we review recent works that provide some level of QoS in NoC systems. Through this literature survey we explore different work done by different researchers by considering different methods for NoC design.

'Design and Evaluation of a High Throughput QoS-Aware and Congestion-Aware Router Architecture for Network-on-Chip', [5] propose a novel QoS-aware and congestion-aware Network-on-Chip architecture that not only implement quality-oriented network transmission and maintains a reasonable implementation cost but it also well balance packet traffic load inside the network to enhance overall throughput. Cost evaluation results also show that the propose router architecture requires negligible cost overhead but provides exceeding performance for both advanced mesh NoC platforms.

'A Quality of Service Network on Chip based on a New Priority Arbitration mechanism', [6] propose a new router architecture that use dynamic arbitration mechanism with a priority-based scheduler system to differentiate between multiple packets and entertain with various QoS requirements. The results of latency show that their router outperforms in terms of minimal router latency is less than 2 cycles as compare to other architectures.

Authors Abdul Quaiyum Ansari et al. [3] works on different Topologies like fat-tree, mesh, tours and c- mesh .While designing a NoC ,topologies are one of the most important part , which considering the performance parameters as a constraint. They considered these most popular topologies and evaluated the performance based on latency, throughput and injection rate of hope counts.

'Local Congestion Avoidance in Network-on-Chip' by Minghua Tang et al.[4] contribute their work on congestion control basis. They work on, to solve local congestion by addressing different local region size. Divide-Conquer approach and routing pressure approach based work they done in their work. Through this work it avoids congestion in every local region by maintaining routing pressure of every local region minimum. Their work shows that the local region size is closely related with the routing performance.

'Signalling approach for NOC quality of service requirements', [7] propose a new communication protocol based on new signaling mechanisms on mesh architecture for a network on chip in order to enhance mapping for QoS requirements between a communication processes and available hardware resources in NoC routers. It appears significant to study the behavior router in a context of an IP specific application in order to adjust advantage and its behavior according to that of the application.

'A Dynamic and Distributed TDM Slot-scheduling protocol for QoS-Oriented Network on Chip', [8] proposes a new scheduling protocol. This run-time distributed and dynamic protocol use for Quality of Service oriented NOC implementation and opens to work on optimization of dynamic allocation of frames.

'QNoC: QoS architecture and design process for Network on Chip', [9] defines Quality of Service (QoS) and cost model for communications in Systems on Chip (SoCs), and developed related Network on Chip (NoC) architecture and design process. It is also concluded that the QNoC architecture requires a much shorter total wire length as compare to the two other options, and while being on par with a point to point architecture in terms of power, it performance is clearly outperforms shared buses.

'Evaluation of Current QoS Mechanism in Network on Chip' [10] propose two methods. First is priority based resource allocation and connection establishment support. Both methods present limitations, especially when flows with QoS requirement. The priority mechanism does not provide rigid guarantees to flow. QoS requirement then this method will be not able to guarantee

this requirement. NoC still does not present efficient solution to provide QoS to application.

'Xpipes: A latency insensitive parameterized network-on-chip architecture for multi-processor SoC'[11] in which a designer size of Xippes changed according to application requirements. It adjusts each channel bandwidth to fulfill the requirements. Applying this method alone does not guarantee avoidance of local congestion, even if bandwidth is largely increased.

'Ethereal Network on chip: Concept, Architecture, and Implementation' [12] used circuit switching method, provides a connection oriented distinction between flows .This scheme has advantage to guarantee tight temporal bound for individual flows. However this method has multiple disadvantages like poor scalability, inefficient bandwidth usages and the setting up a circuit at runtime may long time and unpredictable latency.

Bin Li et al. [13] explore the Contention problems where shared resources are not managed efficiently and can cause resource contention problem . This problem have a big influence on the performance of the SoC and quality-of-service (QoS) of the applications running on that platform in unpredictable ways. They propose CoQoS that is a class-of-service based QoS architecture. In the SoC efficient resources sharing and management plus a guarantee of a certain level of QoS are very important. Research is going on from last decade in this area and proposed different techniques to support QoS. Researchers have proposed and still continuously working for different QoS frameworks to provide predefine levels of performance for different types of applications. However, future SoC architectures are expecting a guarantee QoS to support the overall performance. Their CoQoS architecture framework enables coordinated management of three critical shared resources .They concluded that propose architecture has low cost, and is suitable for SoC architectures.

Authors Yue Qian et al. [14] have applied network calculus theorem and develop analytical models to figure out per flow worst-case delay bound value for two scheduling algorithms first is Strict Priority Queueing (SPQ) and second one is Weighted Round Robin(WRR).They performed comparative analysis and showed that WRR is more flexible and fair than SPQ for guaranteed service. They develop their own algorithm to automate for the delay bound calculation. This allow them to allocate proper weights to individual flows to satisfy the delay requirement.

Jan Heißwolf etc al.[15] work on hardware mechanism propose a hardware supported decentralized unit in NoC for resource management strategy. Through decentralized reconfigurable resource management mechanism they offers to improve the performance and communication resource allocation within the NoC regions. The propose hardware mechanism supports communication resource for management strategy in a decentralized manner in NoCs for supporting QoS. This hardware allows supporting reconfiguration of resource allocation policies in different NoC regions.

Chouchene Wissem et al.[16] present the design of a new on chip network with support Quality-of Service (QoS) . They propose a new routers architecture which use new dynamic arbitration mechanism. This architecture works with a priority-based scheduler which differentiates between multiple packets with various QoS requirements. For verify their router results they created a wormhole input queued 2-D mesh router. They analysis the performance in terms of average throughput and latency . 4x4 mesh 2-D network was used to get the benefit of using the QoS packets and finding the saturation point. They have presented a new QoS NoC architecture work on a new dynamic arbitration that allows packets having a real time requirement to be routed with low latency.

Radu Stefan etc al.[17] propose a Circuit Switching network that supports multicast and it offers hard guarantee service in terms of bandwidth and latency per connection. Their network uses a time-division-multiplexing (TDM) and contention-free scheme. These networks possibly offer both Best-Effort (BE) as well as Guaranteed Services (GS). It is a contention-free scheme and a distributed routing model. This work supports only guaranteed-services (GS) because, GS offers a better performance-cost ratio and in fact the more likely to be required by applications in the embedded domain.

Salah and Tourki [18] propose a router architecture for real-time applications. They use priority based scheduling for QoS support. They differentiate the flow of packets between two categories BE and GS flows. The scheduler examines the deadlines of the incoming flows and then selecting VCs according to the flow class .Priority always goes to GS flows.

Authors results show that router achieves an optimal packets scheduling, increasing channel utilization and increase throughput ratio, reduce network latency, and avoidance of resources conflicts.

Winter and Fettweis [19] develop a hardware unit. This unit is NoCManager. This unit allocates at run-time guaranteed service VCs providing QoS in packet-switched NoCs. NoCManager works as a central NoCManage and authors argue that it is superior to the distributed technique. Hardware mechanism is not suitable when space requirement issue is presented.

III. RESEARCH PROBLEM

NoC based system design is expected to gain its popularity in futur chip design methodology .Modeling provides hints for early design exploration in order to elaborate and explore the detail system specification before actual design begins. It has also to analyse its functionality in details and to improve the sytem performance by finding the bottleneck in system. The system modeling could decide on several NoC architecture issues – topology, router type , IP placement, size of queue,and switching technique to achieve an optimized trade-off dsign based on the user requirment or the specified NoC application.With modeling ,system testing and performance diagnostic are easier and could be made faster.Without modeling mechanisum ,a dircet implementation of a system is more erroneous and several redesigneds upgrades may be required in order to achieve a workable system.

IV. OBJECTIVE AND SCOPE OF THE RESEARCH WORK

The aim of our research is to develop an analytical model of a 2D mesh NoC. This model allows prediction of the NoC performance and provide hints for early NoC based SoC design exploration. To achieve this aim,it is necessary to know the architecture of an NoC. Routers play a pivotal role in NoC based design performance.

A. NoC architecture to be support the end-to- end communication between the modules at the specific quality of services (QoS) using resources sharing of the interconnected resources.
B. Optimization of network parameter such as link bandwidth , switch configuration ,buffering ,channel interface as well as network interface for effective NoC perforamnce.
C. Implementation of contentation free communicaton for QoS.
D. The shortest latency implementation to be achieved which defeat the flexibility required by on chip networks.
E. To decided on the appropriate network architecture that support the QoS requirement i.e. Minimum latency and Maximum throughput the listed network characteristic must be consider: Different topologies, Routing scheme, Arbitration mechanism, Switching technique , End-to-End congestion, Router architecture and Flow control schemes.

V. CONCLUSION

In this paper we have presented the communication problems in SoC system based on traditional bus-based mechanism and needs of QoS in NoC. This work is focused on the quality of service management in network on chip by implementation of complete new NoC architecture with signaling approach based router architecture. We are going to implement a new architecture that supports more data processing granularity with support good latency and throughput results. This architecture maybe one of the best solution to improve the performance problem of traditional NoC. The outcome of this review can aid new researchers in developing a technique to decrease latency and to increase throughput in NoC. For this we need to know current existing techniques and work done in this area. We hope through this work we provide the good solution to the communication problems in SoC.

REFERENCES

1 J. D. Owens, W. J. Dally, R. Ho, D. N. Jayasimha, S. W. Keckler, and L.-S. Peh, 'Research challenges for on-chip interconnection networks,' Micro, IEEE, vol. 27, no. 5, pp. 96 –108, September, 2007.

2 C. Grecu, P. Pande, A. Ivanov, and R. Saleh, 'Timing analysis of network on chip architectures for mp-soc platforms,' Microelectronics Journal, vol. 36, no. 9, pp. 833–845, 2005.

3 Abdul Quaiyum Ansari ,Mohammad Rashid Ansari ,Mohammad Ayoub Khan 'Performance Evaluation of various Parameters of Network-on-chip(NoC) for Different Topologies', 2015 Annual IEEE India Conference (INDICON) ,Pages: 1 - 5, DOI: 10.1109/INDICON.2015.7443838

4 Minghua Tang, Xiaola Lin, and Maurizio Palesi,'Local Congestion Avoidance in Network-on-Chip' , IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 27, NO. 7, JULY 2016

5 Chifeng Wang, Nader Bagherzadeh, 'Design and Evaluation of a High Throughput QoS-Aware and Congestion-Aware Router Architecture for Network-on-Chip'. 2012 20th Euromicro International Conference on Parallel, Distributed and Network-based Processing.

6 Chouchene Wissem, Abid Noureddine, Abdelkrim Zitouni, 'A Quality of Service Network on Chip based on a New Priority Arbitration mechanism'. ICM 2011 Proceeding , IEEE Conference Publications ,Pages: 1 - 6, DOI: 10.1109/ICM.2011.6177349

7 Med Lassaad KADDACHI, Adel SOUDANI, Rached TOURKI. 'Signaling approach for NOC quality of service Requirements', 2008 International Conference on Signals, Circuits and Systems'.

8 Young Jin Yoon; Nicola Concer; Luca Carloni, 'A Dynamic and Distributed TDM Slot-scheduling protocol for QoS-Oriented Network on Chip'. Computer Design (ICCD), 2011 IEEE 29th International Conference Pages: 31 - 38, DOI: 10.1109/ICCD.2011.6081372

9 Evgeny Bolotin, Israel Cidon, Ran Ginosar and Avinoam Kolodny, 'QNoC: QoS architecture and design process for Network on Chip', Journal of Systems Architecture, Volume 50, Issues 2–3, February 2004, Pages 105–128

10 Mello, A. ; Tedesco, L. ; Calazans, N. ; Moraes, F.,'Evaluation of current QoS Mechanisms in Networks on Chip' System-on-Chip, Conference: System-on-Chip, 2006.

11 Dall'Osso, M. , G. ; Giovannini, L. ; Bertozzi, D. 'Xpipes: A latency insensitive parameterized network-on-chip architecture for multi-processor SoCs' 2012 IEEE Pages: 45 - 48, DOI: 10.1109/ICCD.2012.6378615

12 Goossens, K. ey al.,'Ethereal Network on chip: Concept, Architecture, and Implementation', IEEE Design and Test of Computer ,v22(5),Sep-Oct 2005 ,pp 414-421.

13 Bin Li, Li Zhao, Ravi Iyer , Li-Shiuan Peh, Michael Leddige , Michael Espig , Seung Eun Lee ,Donald Newell ,"CoQoS: Coordinating QoS-aware shared resources in NoC-based SoCs" , Elsevier Parallel Distrib. Comput (2010)

14 Yue Qian, Zhonghai Lu and Qiang Dou, 'QoS Scheduling for NoCs: Strict Priority Queueing versus Weighted Round Robin' Computer Design (ICCD), 2010 IEEE International Conference on 3-6 Oct. 2010

15 Jan Heißwolf , Aurang Zaib , Andreas Weichslgartner , Ralf König , Thomas Wild,Jürgen Teich, Andreas Herkersdorf, Jürgen Becker , 'Hardware-assisted Decentralized Resource Management for Networks on Chip with QoS' ,Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International Conference

16 Chouchene Wissem, Abid Noureddine, Abdelkrim Zitouni, ' A Quality of Service Network on Chip based on a New Priority Arbitration mechanism' , Microelectronics (ICM), IEEE 2011 International Conference on 19-22 Dec. 2011

17 Radu Stefan, Anca Molnos, Angelo Ambrose, Kees Goossens, 'A TDM NoC supporting QoS, multicast, and fast connection set-up', 2012 EDAA, Proceeding DATE 12 Proceeding of the conference on Design, Automation and Test in Europe Pages 1283-1288

18 Salah, Y.; Tourki, R. ,'Design and FPGA Implementation of a QoS Router for NoC', In: International Conference on Next Generation Networks and Services, 2011, pp. 84-

19 Winter, M.; Fettweis, G., P. 'Guaranteed Service Virtual Channel Allocation in NoCs for Run-Time Task Scheduling'. In: DATE, 20 II, pp. 1-6.