



An Energy Conscious Topology Augmentation Methodology for On-Chip Interconnection Networks

By Samta Jain, Vaishali Sodani & Naveen Choudhary

College of Technology and Engineering, India

Abstract- On-chip communication, modular, scalable packet-switched micro-network of interconnects, generally known as Network-on-Chip (NoC) architecture can be designed as regular or application-specific (irregular) network topologies. Application specific custom network topologies are advantageous in terms of optimized design according to given performance metrics and regular network topologies are advantageous in terms of its modularity, lower design time and efforts required and thus are suitable for mass production. So to offer the advantages of both the topologies this paper proposes a methodology to augment the regular topology according to the application characteristics. The experimental results demonstrate that the proposed methodology can reduce dynamic communication energy consumption by on average of 32.79% and reduction in average per flit latency by on average of 16.22% over regular 2D NoC architecture.

Keywords: *network-on-chip, application specific topology, interconnection network, dynamic communication energy.*

GJCST-E Classification : *C.1.2, C.2.0, C.2.1*



Strictly as per the compliance and regulations of:



An Energy Conscious Topology Augmentation Methodology for On-Chip Interconnection Networks

Samta Jain ^α, Vaishali Sodani ^σ & Naveen Choudhary ^ρ

Abstract- On-chip communication, modular, scalable packet-switched micro-network of interconnects, generally known as Network-on-Chip (NoC) architecture can be designed as regular or application-specific (irregular) network topologies. Application specific custom network topologies are advantageous in terms of optimized design according to given performance metrics and regular network topologies are advantageous in terms of its modularity, lower design time and efforts required and thus are suitable for mass production. So to offer the advantages of both the topologies this paper proposes a methodology to augment the regular topology according to the application characteristics. The experimental results demonstrate that the proposed methodology can reduce dynamic communication energy consumption by on average of 32.79% and reduction in average per flit latency by on average of 16.22% over regular 2D NoC architecture.

Keywords: network-on-chip, application specific topology, interconnection network, dynamic communication energy.

I. INTRODUCTION

Network-on-Chip designs consist of a number of interconnected devices (e.g. general or special purpose processors, peripherals, memories etc.) where communication between these system modules is achieved by sending packets over an interconnection network. Packets are transported from one to another module via the interconnection networks [1]. Thus, the interconnections among multiple cores on a chip have a significant impact on communication and performance of the chip design in terms of dynamic energy consumption, latency, throughput etc. So topology selection is an important choice in the design of NoC and in Network on Chip topology design is one of the significant factor that affects the net delay of the system and the dynamic communication energy of the system. Topological parameters, like hop count, wiring layout are closely related to the performance and power dissipation of a network [13]. In other words, it can be said that the NoC performance can be greatly affected by the underlying topology and the routing function followed by the communication cores.

NoC architecture can be designed as regular or application specific network topology. Application-

specific network topologies are designed to make the system more application centric [2]. As a result, the performance objectives such as power consumption, latency, area of the chip, number of routers in the system can be optimized effectively with the given design constraints [3]. However the loss of modularity in application-specific NoC generally leads to increased time-to-market in comparison to the modular generic standard NoC topologies.

Regular topologies like 2D mesh are more structured and built not considering much about the application characteristics. A regular NoC topology assumes a homogeneous distribution of routers, which leads among others to lower design time and cost. Therefore regular NoC topologies are suitable for mass production due to lower design cost and effort required. Additionally, regular topologies are highly reusable and impose the minimum re-design effort, in case they are employed to different applications.

However, due to the lack of fast paths between remotely situated nodes and many hops between different communicating nodes regular topology architectures for a given routing methodology may suffer from long packet latencies and higher communication power dissipation. Although the regular structure significantly address the implementation issues, such as floor planning, unequal lengths of wire, hence results in poorly controlled electrical parameters.

In order to meet the communication requirements, accelerate time-to-market and cut down the communication energy consumption, there is a great need to find new design alternatives of customizing the regular architectures as per the communication requirements. We therefore focused our research on exploring and finding an efficient methodology to tailor the regular 2D mesh topology so that the dynamic communication energy can be optimized. The methodology will optimize the dynamic communication energy consumption by reducing the hop count or by reducing the number of short links according to the traffic characteristics of the application for a given routing function. Tailoring the regular topology shall reduce the design time in comparison to the design of application specific NoC from stretch with reasonable performance advantages.

Author $\alpha \sigma \rho$: Deptt. of CSE, CTAE, MPUAT, Udaipur.
e-mails: samtajain1991@gmail.com, vaishalisodani@gmail.com,
naveenc121@yahoo.com

In [4], the authors introduce additional random links to the mesh and torus network and investigate the performance of these networks. In [5], authors present a methodology to automatically synthesize an architecture where a few application-specific long-range links are introduced on a regular mesh network. Drawing inspiration from such work this paper proposes the methodology to insert energy conscious shortcut links with practical design constraints such as link length, port constraints. The proposed design also strives to add minimum number of shortcuts so as not to distinguish the original design in terms of area and floor-planning.

In section 2, NoC communication energy and routing functions are discussed. In section 3, the proposed methodology is given. In the next section, the experimental results are discussed and in the last section the conclusion is presented based on the results.

II. NOC COMMUNICATION ENERGY AND ROUTING FUNCTION

a) Dynamic Communication Energy Model

The dynamic communication energy model [7] for the network on chip can be defined as:

$$E_{bit}^{(t_i, t_j)} = n^{hops} \times E_{rbit} + (n^{hops}-1) \times E_{Lbit} \quad (1)$$

Where $E_{bit}^{(t_i, t_j)}$ = the average energy consumption for sending one bit of data from t_i to tile t_j , n_{hops} = the number of routers the bit traverses from tile t_i to tile t_j , E_{rbit} = the energy consumed by the router for transporting one bit of data, E_{Lbit} = the energy consumed by unit link/channel for transporting one bit of data. For the NoC networks with unequal link length, the $(n_{hops} - 1) \times E_{Lbit}$ can be replaced as the summation of bit energy consumed by each link in the path/route, the bit follows from source tile to the destination tile [21].

Hop count is the number of routers the bit traverses from source tile to tile destination tile. The communication energy model [7] for the network on chip clearly shows that the latency and power consumption linearly relates to the average number of hops the packet traverse, hence to the network topology. So we will optimize the latency and power consumption by reducing the hop count between the source tile and destination tile for a given routing function.

b) Routing Function in NoC

Routing function determines how the data is transmitted from source tile to destination tile. Routing algorithm can be grouped according to different criteria. According to the place where routing decision is made, routing algorithm can be classified as centralized, source and distributed routing algorithm. In centralized the route is chosen by a central controller, in source the route is chosen by a source router prior to a sending the packet, in distributed the route is chosen by an

intermediary routers [16, 17]. According to how a route is defined to route the packets, routing algorithm can be deterministic or adaptive. In deterministic routing, one path is calculated between source and destination and routing is done through that fixed path only.

Deterministic routing algorithm always routes the packets from source to destination along that same path [15]. Deterministic algorithm does not take into account network traffic situation or the network condition before choosing a path from source to destination. In adaptive routing algorithms multiple paths are calculated between source and destination but routing is done in one selected path depending upon the network traffic. In adaptive algorithm network traffic situation (such as network load, traffic condition) are taken into consideration. Limitations of adaptive routing algorithm are its implementation cost, complexity and more power consumption [16]. Deterministic routing algorithm is popular in NoC due to simplicity, low latency, simple routing logic, packet reach destination in correct order and reordering is not necessary [7]. It is a greedy algorithm because it always chooses a shortest path to deliver a packet from source to destination [17].

The design space of efficient deadlock-free routing is an important aspect which is more implicitly affected by the routing algorithm which determines which routes packets can take from source to destination. A deadlock is a situation where packets are involved in a circular wait for resources. Deadlock freedom is an important property for networks, since a packet deadlock can destroy possibilities of communication [18]. Deadlocks in a network may block communication between cores and may even lead to a complete failure of network. Therefore it is essential to have algorithms which have ability to handle them. The deterministic routing [7] (e.g. XY routing [14, 19] and odd-even routing [15, 19]) are the most widely used deadlock free routing algorithms for the popular 2D-mesh based on chip networks.

III. PROPOSED METHODOLOGY

In the presented work, an energy conscious topology augmentation methodology is proposed for application specific on-chip interconnection networks. In the proposed methodology a customized NoC topology is synthesized by tailoring the regular 2D mesh topology to achieve energy efficient communication.

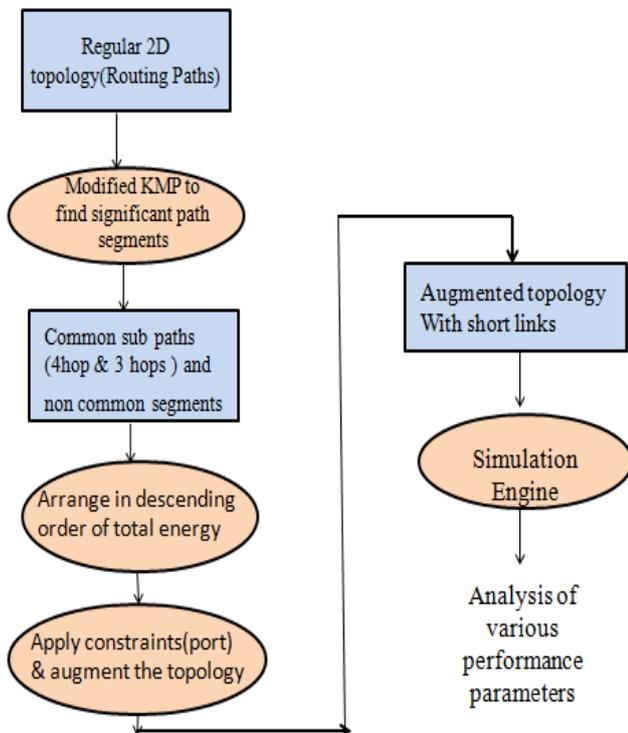


Figure 1 : Architecture of Proposed Methodology

Equation 1 clearly shows that dynamic communication energy linearly relates to the number of routers/hops the packet traverses from the source tile to the destination tile i.e. as the routers in communication path increases, the communication energy also increases. In the regular grid like NoC architecture, there is lack of fast path between two distant nodes so the packets have to traverse many hops between any two remotely situated nodes (source and destination), leading to poor dynamic communication energy dissipation. If we can able to reduce the number of hops in the communication path for a chosen routing function, the considerable amount of energy saving can be achieved in the regular NoC architecture. The reduction in the number of hops in communication path can be done by augmenting the regular mesh by inducing significant shortcut links according to the application characteristics and also constraining parameters such as length of the shortcut link and the port availability per tile. The length of the shortcut link is constrained because in NoC communication generally needs to be clock synchronized and to avoid the long wiring complexity and area overhead short links are generally preferred. The port availability per tile means add the shortcut link to a tile only if there free port is available. The augmentation of regular topology can be done by adding the significant shortcut links to the communication path segments which are carrying maximum traffic load for all the communication/routing paths for a routing function while taking the above

mentioned constraints into consideration. In the proposed methodology these significant shortcut links are find out with the help of modified KMP string matching algorithm [12].

Figure 1 gives architecture of the proposed methodology. The proposed methodology augments the regular 2D mesh topology which is customized using any of the deterministic routing function. In the proposed methodology the routing paths (decided according to chosen deterministic routing function) are input to the modified KMP [12] string matching algorithm. Here this algorithm actually dealing with the routing paths as a string. This algorithm find out the path segments of 3 consecutive and 2 consecutive links (in other words of 4 hops and 3 hops respectively) which are common in other routing paths and also gathers non common path segments. In the next step the proposed methodology arranges all these segments (common sub paths of 4 & 3 hops and non-common segments) in the descending order of total communication energy consumed by these segments. This ordering has been done so that the segments with higher dynamic communication energy consumption will be considered earlier or in other words the segments which carry the maximum traffic load can be addressed earlier, so that maximum benefits can be achieved. After that, it will add the shortcut link to the routing paths which contains that segment in the regular 2D mesh topology according to the order by keeping in check the input constraints (port availability per tile). Here adding the shortcut link to segment of 3 consecutive links or to segment of 2 consecutive links means skipping 2 hops or 1 hop in the corresponding routing path respectively. Hence by doing this we are constraining the length of additional link (shortcut link) in the augmented topology. This augmented topology is referred as shortcut inclusive 2D mesh (SI-2D mesh). For the analysis of various performance parameters SI-2D mesh is validated on NoC simulator NC-G-SIM [20].

IV. EXPERIMENTAL RESULTS

The performance of SI-2D mesh derived from the regular 2D mesh is validated on a network on chip simulator NC-G-SIM. It is a discrete event cycle accurate simulator for Network-on-Chip performance simulation [20]. NC-G-SIM supports regular as well as irregular topology framework with source and table based routing. E_{Lbit} is calculated according to the analytical energy model presented in [8][9]; E_{rbit} for a router is calculated the help of power simulator Orion [10] for 0.18 μ m technology. **The communication is assumed to be of constant bit rate.** Number of application (traffic) characteristics data sets were randomly generated using TGFF [11], with diverse bandwidth requirement of the IP Cores and randomly generated communication bandwidth requirement. The simulation is run for 10000

clock cycles. The average total communication energy per flit received at destination and average per flit latency are taken as performance metric. The methodology is applied to 6 set of different 2D mesh topology, for each topology 10 different test cases were generated and executed. After having experimented on

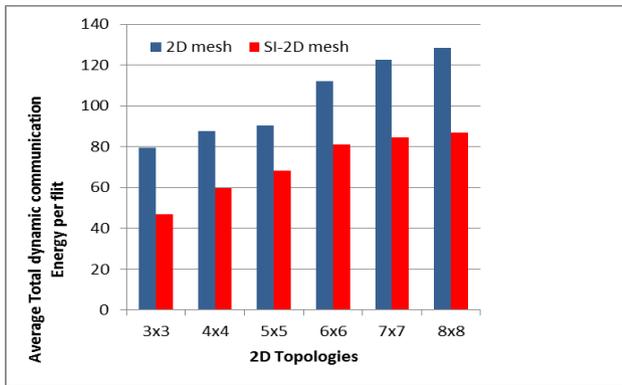


Figure 2 : Average total dynamic communication energy per flit comparative results between 2D mesh and SI-2D mesh topology

As shown in Figure 2, the short cut inclusive 2D mesh (SI-2D mesh) generated using the proposed methodology saves about on average 32.79% of dynamic communication energy consumption compared to regular 2D mesh because SI-2D mesh needs less number of hops to traverse from the source tile to the destination tile, leading to reduced dynamic communication energy and latency. Figure 3 shows comparison in terms of average per flit latency between 2D mesh and SI-2D mesh for various topologies and it depicts that the average latency per flit gets reduced by on average of 16.22% in SI-2D mesh compared to the regular 2D mesh

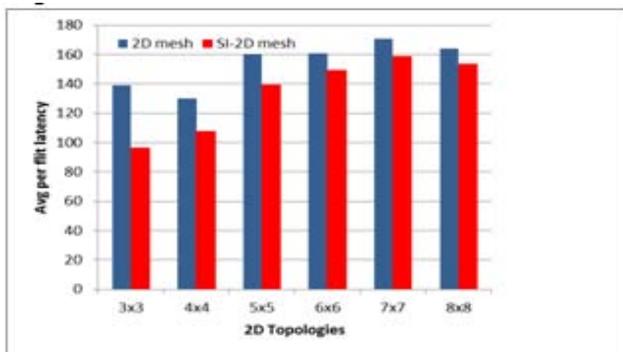


Figure 3 : Average per flit latency comparative results between 2D mesh and SI-2D mesh

It is clear that we are getting marginal reduction in latency as compared to the dynamic communication energy; the reason behind this is the occurrence of the congestion due to the generation of hotspot at the induced routers which are heavily loaded as they are bearing the traffic load which was earlier getting spread number of topology sets, the average of obtained results is represented to demonstrate the effectiveness

of the proposed methodology. The configuration files for a given traffic pattern are generated using the proposed methodology and supplied to the simulator as an input over multiple routers. However still there is gain in latency as the switching delay in path from source to destination for various application characteristics is reduced due to decrease in number of routers in the corresponding paths.

V. CONCLUSION

In this paper, a greedy energy conscious topology augmentation methodology for on-chip interconnection networks is proposed. Using the proposed methodology a shortcut inclusive topology is designed according to the traffic characteristics. The proposed methodology is greedy but fast and uses fast modified KMP string matching algorithm [12] for exploring the effective and efficient shortcuts to be augmented in the given regular architecture. The experimental result clearly shows that shortcut inclusive topology generated using proposed methodology is able to reduce dynamic communication energy leading to significant energy savings and also reduction in the average per flit latency as compared to the regular topology.

REFERENCES RÉFÉRENCES REFERENCIAS

- Dally, W. J., and Towles, B. (2001). Route packets, Not wires: On-chip interconnection networks. In Design Automation Conference, Proceedings, pp. 684 -689, IEEE.
- Umamaheswari, S., Rajapaul, P., J. (2011). Energy, Throughput and Area Evaluation of Regular and Irregular Network on Chip Architectures. International Journal of Distributed and Parallel Systems (IJDPS) Vol.2, No.5.
- Yu Bei, Dong Sheqin (2010). Floorplanning and Topology Generation for Application-Specific Network-on-Chip. Design Automation Conference (ASP-DAC), 2010 15th Asia and South Pacific. IEEE: 535 – 540.
- Fuks, H., Lawniczak, A. (1999). Performance of data networks with random links. Mathematics and Computers in Simulation, vol 51.
- Ogras Umit, Y. and Marculescu, R. (2005). Application Specific Network-on-Chip Architecture Customization via Long Range Link Insertion.
- Ogras Umit, Y. and Marculescu, R. (2006). It's a Small World After All: NoC Performance Optimization via Long Range Link Insertion. IEEE, VOL. 14, NO. 7.
- Hu, J., Marculescu, R. (2005). Energy and performance-aware mapping for regular NoC architectures. Computer- Aided Design of Integrated Circuits and Systems, Vol. 24: 551-562.

8. Hu, J., & Marculescu, R. (2003). Energy-aware mapping for tile-based NoC architectures under performance constraints. In Proceedings of the Asia and South Pacific Design Automation Conference, pp. 233-239.
9. Choudhary, N., Gaur, M. S., Laxmi, V., & Singh, V. (2011). GA based congestion aware topology generation for application specific NoC. In Sixth IEEE International Symposium Electronic Design, Test and Application (DELTA), pp. 93-98.
10. Kahng, A. B., Li, B., Peh, L. S., & Samadi, K. (2009). Orion 2.0: A fast and accurate NoC power and area model for early-stage design space exploration. In Proceedings of the conference on Design, Automation and Test in Europe, pp. 423-428.
11. Dick, R. P., Rhodes, D. L., & Wolf, W. (1998). TGFF: task graphs for free. In IEEE Computer Society, Proceedings of the 6th international workshop on Hardware/software co-design, pp. 97-101.
12. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein. (2009). Introduction to Algorithms. The MIT Press Cambridge, Massachusetts London, England, 3rd edition.
13. George, B. (2009). Energy consumption in networks on chip: efficiency and scaling. IEEE 15th International Symposium, pp. 250–261.
14. Dehyadgari, M., Nickray, M., kusha, A., Navabi, Z. (2005). Evaluation of Pseudo Adaptive XY Routing Using an Object Oriented Model for NOC.
15. Duato, J., Yalamanchili, S., Ni L. (2003). Interconnection Networks: An Engineering Approach. Morgan Kaufmann Publishers, San Francisco.
16. Singh J. K. Performance evaluation of different routing algorithm in network on chip master dissertation National Institute of Technology Rourkella, Odesha.
17. Adamu, G., Chejara P., Garko, A., B. (2015). Review of deterministic routing algorithm for network-on-chip. International conference on science, technology and management.
18. Holsmark, R., Palesi, M. and Kumar, S. (2006). Deadlock Free Routing Algorithms for Mesh Topology NoC Systems with Regions. Digital System Design: Architectures, Methods and Tools, 9th EUROMICRO Conference, Dubrovnik, pp. 696-703.
19. Choudhary, N. (2012). Network-on-Chip: A New SoC Communication Infrastructure Paradigm. International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, vol.1, 332-335.
20. Vyas, K., Choudhary, N. and Singh, D. (2013). NC-G-SIM: A Parameterized Generic Simulator for 2D-Mesh, 3D Mesh & Irregular On-chip Networks with Table-based Routing. Global Journal of Computer Science and Technology (GJCST-E) on Network, Web & Security 13: 7- 12.
21. Wadhvani, P., Chaudhary, N., Singh, D. (2013). Energy Efficient Mapping in 3D Mesh Communication Architecture for NoC. Global Journal of Computer Science, and Technology (GJCST-E) on Network, Web & Security 13: 1-5.

This page is intentionally left blank