A Novel Scrutiny of Different Torus Topologies: Survey

T. MadhuBala

PG Scholar, Department of Electronics and Communication Engg, PSNA College of Engineering and Technology, Tamil Nadu-624622, India.

S. Sobana

Assistant Professor, Department of Electronics and Communication Engg, PSNA College of Engineering and Technology Tamil Nadu-624622, India.

Abstract – This paper deals with the different torus topologies of the network. The topology of the network plays a very important role in Network on Chip (NoC). Topology of the network determines how the components must be placed and organized together. There are different kinds of topology used in NoC such as mesh, torus, etc. The performance of torus topology is improved by varying the structure as twisted torus, folded torus. Here a small survey of the torus based topological network and the new king torus network with its benefits is discussed. The path selection is mainly based on the routing algorithm so the shortest path between the source and destination is easily estimated. So that the overall latency gets reduced and throughput gets increased.

Index Terms – Cache coherent protocol, communication, deadlock, large scale systems.

1. INTRODUCTION

Now a day's system-on chip (SoC) are too complex to utilize an old hierarchal bus or crossbar interconnects. This is the main reason for the evolution of new technology Network on Chip (NoC). NoC improves the scalability of SoC and provides good solution to global communication within big multi-core systems. It is also used to solve the communication problem for architecture. In which topologies play very important role in network on chip. Few examples, cache-coherent SMP systems-Intel's Xeon Phi architecture in which communication is performed mainly by cache line transfers between cores and in Intel Teraflop research chip with 64 and 80 cores arranged in a 2D mesh respectively. This is based on packet-switched (PS) NoCs, which bring the advantage of high flexibility and increases the performance of the system. So that the packets can be easily transmitted from sender to receiver. Here a small survey of torus based topologies and their routing technique are discussed below.

1.1 Need of NoC

Network on Chip (NoC) is an approach for designing the communication subsystem between IP cores in a System on a Chip (SoC). Thus NoC improves the scalability of SoCs, and

the power efficiency of complex SoCs compared to other designs.

2. REVIEW

[1] Maximum multi-core and some many-core processor implement cache coherency protocols. In which this complicate the design of optimized parallel algorithms. Communication is performed completely by cache line transfers between cores. This complicates the understanding of performance properties. This paper developed a natural performance model for cache-coherent architectures and demonstrates its use with the currently most scalable cache-coherent many-core architecture. By this model, they proposed optimal and optimized algorithms for complex parallel data exchanges. All algorithms that were developed by this model pop the performance of the highly-tuned specific vendor. The model makes more global approach to satisfy the trade off between complexity and accuracy of algorithm design. So the expected model can serve as a vehicle for advanced algorithm design. The developed algorithm with the help of our analytical models leads to improve the performance over Intel's hand-tuned MPI and open MP. The performance improvement is about 4.3 times compared with the previous models. By using Direct Remote Cache Access (DRCA), leads to high performance and better degree of prediction and transparency. Optimizing the cache-coherency protocols is harder. The developed models and techniques are complex.

[2] Generally shorter and longer packets co-exist in cache-coherent NoCs. The previous existing designs for torus networks do not efficiently handle different packet size. To overcome the deadlock operations, a design consist of two virtual channels (VC), which abnormally affects the router frequency. For few optimizations it is possible to use one VC. Hence they estimate all packets as maximum-length packets, inefficiently utilizing the previous buffers. The proposed Flit Bubble Flow Control (FBFC) grants worm-hole flow control this avoids the deadlocks by using one free flit-size buffer slot inside a ring. FBFC uses one VC and does not treat short

packets as long packets. It achieves both high frequency and efficient utilization of buffer. FBFC performs 92.8% and 34.2% better than the previous models (LBS and CBS) about for synthetic traffic in a 4*4 tours. Similarly the gain of 8*8 tours is increased by 107.2% and 40.1% which is a large network. FBFC achieves an average 13.0% speedup over LBS for PARSEC workloads. Their result also show that FBFC is more power efficient than LBS and CBS, and a tours with FBFC is more power efficient than a mesh. So this technique can be used for the upcoming applications..FBFC does not treat short packets as long ones so this yields high buffer utilization. It consumes less power consumption.

[3] For every 18 months there is an increasing transistor density on chip according to Moor's law. Simultaneously, multi-core processors are also growing so that the requirement of increased performance, power efficiency. The shared memory is a predominant model, for soften programmability and increasing the motility of the system. However the memory being shared by an increasing number of cores and provide a scalable coherence mechanism which is essential for these systems. Snoopy coherence has been a good coherence scheme due to its high performance and simplicity.

The proposed method is realized by using a two-dimensional mesh interconnection network that is referred to as OMNI (Ordered Mesh Network Interconnect). OMNI is an permissive solution for the Snoopy Coherent Research Processor with Interconnect Ordering (SCORPIO) processor prototype developed by MIT- a 36-core chip multi-processor supporting snoopy coherence, and fabricated by using a 45nm technology. It highly reduces the runtime by 36% in comparison to directory and Hammer coherence protocol implementations. The OMNI network achieves an operating frequency of 833MHz post-layout, occupies 10% of the chip area, and consume less than 100mW of power. The network OMNI increases the interaction between coherence mechanism and communication substrates because it's an important consideration for future multi-core designs. Chip multiprocessors (CMP) lead to reliable mechanisms for in-network coherence. This method leads to fast latency bound. The addition of filtering leads to a decrease in the operating frequency.

[4] Gaussian and Eisenstein-Jacobi integers are mainly used to construct interconnection networks with good topological properties. In this article, a unified study of the interconnection networks is presented. In addition, the results include disintegrate the edges into disjoint Hamiltonian cycles which simplifies the calculation of the Eisenstein-Jacobi distance, a distribution of distance present between Eisenstein-Jacobi router nodes and helps to find shortest path routing algorithms.

 Here a new toroidal network is introduced which is a two dimension network and the nodes are initialized using a subset of the complex numbers. In which 2D toroidal network is a subclass of the Gaussian network. Quotient rings of Gaussian and Eisenstein-Jacobi (EJ) integers give a new specific approach to the study of the topology of networks. The properties of the networks helped to simplify routing algorithms and find the integers. EJ distribution and edge-disjoint Hamiltonian cycle is found in a step by step process so it is easy to understand. There may be some problems such as fault-tolerant routing and multicasting with edge and node failures. This becomes a long process based on the number of nodes get increases in the network..

[5] This paper is designed to carry out performance-related studies of interconnection networks by using the Interconnection Network Simulator and Evaluation Environment (INSEE). It is consist of two modules in it. That is a Functional Simulator of Interconnection Networks (FSIN) and Traffic Generation module (TrGen). These two modules combined together with some other modules that extended as INSEE's functionality. The router models and topologies implemented in FSIN are included in this description. The available methods to generate traffic are described. Finally, they discussed about the resource consumption of INSEE when it manages with different systems and workloads is also evaluated. INSEE has been progressed with many processes, it also includes the evaluation of new topologies and it leads to network congestion. INSEE provides an execution time and based on the memory consumption the scale is roughly drawn and it is linear with the number of nodes. Due to the main memory consumption of the room in the queues, and the number of queues depends on the number of nodes. The main use of INSEE is because of its flexibility and its low resource requirement. INSEE is not allowed for parallel or distributed simulation. It is not able to perform large-scale simulations because of its complexity.

[6] This paper introduces a new closed-form oblivious routing algorithm called W2TURN which is the worst-case throughput optimal in 2D torus networks. W2TURN is based on a weighted random selection of path that contain at most two turns. By providing a restriction the maximum number of turns in routing path is to simply provide a deadlock-free implementation of W2TURN. When the network radix is odd, W2TURN achieves the minimum average hop counts between the source and destination that can be achieved with 2-turn paths while remaining worst-case throughput optimal, within 0.72% on a 12*12 torus. Here they describe another routing algorithm based on weighted random selection of path with two turns or ore number of turns called I2TURN and show that I2TURN is equivalent to IVAL (Improved Valiant Randomized Routing) algorithm. However, I2TURN eliminates the need for loop removal at runtime and provides a closed-form analytical expression to evaluate the average hop count.

This helps to demonstrate analytically that W2TURN strictly outperforms IVAL in average hop count. Finally they presented a algorithm for rings as Weighted Random Direction (WRD). WRD forms a closed-form expression for the optimal distribution of traffic along the minimal and non-minimal direction of a ring topology to achieve minimum average hop count while guaranteeing optimal worst-case throughput. W2TURN and WRD also perform well in the average-case, and outperform the best previously known worst-case throughput routing algorithms along the closed-form descriptions for latency and throughput over a wide range of traffic patterns.
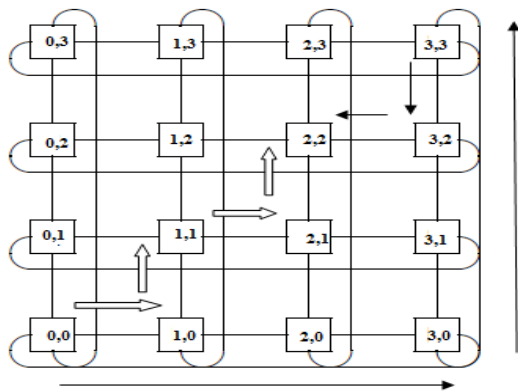

Figure.1. Routing with 2-turn paths. Both the solid and dotted lines represent different 2-turn path between (0,0) and (2,2).

WRD is used in 2D torus network to achieve the minimum average hop count on rings. Reduced latency and good throughput is achieved.
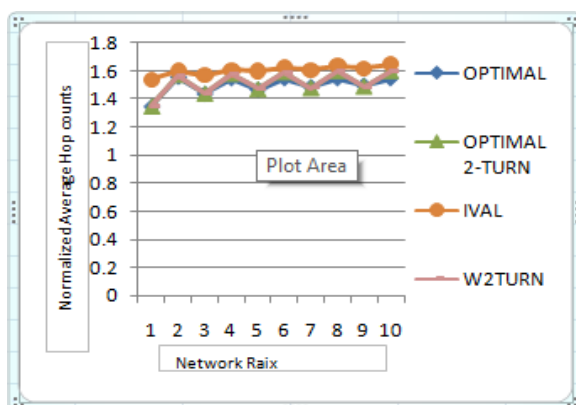

Figure.2. Comparison of average hop counts for several routing methods

[7] Tours network inter-connection is mainly used in the many current parallel computers. This is due to the main advantages of the topologies, square (2D) or cubic (3D) torus networks. 2D and 3D tori with different number of nodes per dimension have been used. In this paper they deal with the analysis of twisted 2D and 3D mixed-radix tori which solve the network bottlenecks present in non twisted network. Such topologies recover edge symmetry, and simultaneously help to balance the utilization of the links. The distance-related property of twisted tori together with a full characterization of bisection bandwidth is described. The consistent performance gain is up to 74% in accepted load. The bottleneck in the twisted tori is removed by equalizing on each dimension the length of the path traversed by packets. The networks have been tested managing both synthetic traffic and workloads from real application traces.

[8] Due to increasing performance of topological networks. A new two topologies is proposed for On Chip Networks and that is denoted as king mesh and king torus. These topologies are mainly designed for larger degree evolution of the classical mesh and torus topologies. In a king network, packets can be send by using orthogonal and diagonal path like the king on a chess board (i.e) in chess board a king can move in its all eight directions. This is mainly proposed for addressing the study of distance properties of topologies, bisectional bandwidth and path diversity and folding scheme. That is ranging from minimal distance routing technique to miss-routing technique and they concluded by analysing the topological importance of these networks.
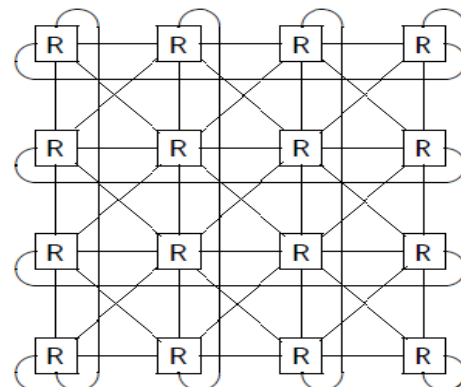

Figure.3. King Torus (4*4) network

The evolution of these new topologies serves as a better alternative to classical topologies. King topological properties offer efficient properties and they can be used in future network-on-chip systems. A better routing technique specified for king networks that are both adaptive and deadlock free, which helps to expose the topological richness. Performance evaluation of these algorithms based on synthetic traffic has been discussed. King network is suitable for large scale systems.

[9] In NoC one of the urge to reduce the choke point in communication and the clock problem from architecture. Each route in NoC adds some routers and takes a few clock periods by passing a router. When the network is in traffic congestion,

the package transmission produces more time delay. So adopting an appropriate routing algorithm provides the balance between the delay and throughput become a main problem. So they use the XY routing algorithm for 2D tori of NoC architecture for constant bit rate (CBR). This algorithm reduces the average latency per packet and increase the average throughput.

[10] In NoC, if the network is in congestion, the package transmission will produce much more delay so that the time taken to reach from master to slave will be increased. So adopting an appropriate routing algorithm to get the balance between the delay and throughput becomes the major problem. In this paper they tried to solve that problem using torus topology by using modified XY routing algorithm. 3*3 network size & traffic used is Constant Bit Rate (CBR) random traffic.

If the packet size increases then latency of different packet get increases and throughput (in Gbps) also increases but latency decreases for a normal transmission. Shortest path is found by using this modified XY routing algorithm. If the packet size gets increased, average latency per packet also get increased. If all title acts as master and all titles act as slave leads to congestion in the network so the average latency per packet increases.
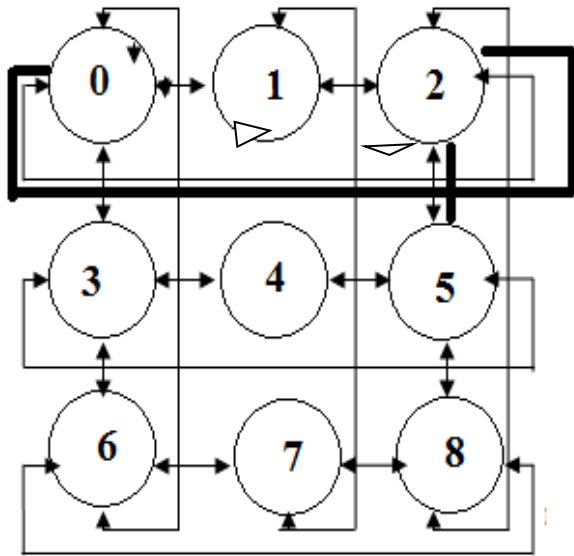

Figure.4. Example of XY Routing Algorithm

3. CONCLUSION

From the above discussion it is concluded, there are different types of torus based network and efficiency also various from one to another. We present a survey of torus based routing algorithms and their response. Different routing algorithm leads to different routing process from source to destination. Here we discuses about the DOR, 2TURN, IVAL algorithms even though they have few drawbacks. A new king topology is identified to improve the overall performance which includes low latency, high throughput, efficient routing process. This king topology can be discussed in my future.

REFERENCES

[1] S. Ramos and T. Hoefler., "Modeling communication in cache-coherent SMP system: A case study with Xeon Phi", in Proc. Int. Symp. High-perform. Parallel Distrib. Comput., 2013 ,pp. 97-108.

[2] S. Ma, Z. Wang, Z. Liu, and N.E. Jerger., "Leaving one slot empty: Flit bubble flow control for torus cache-coherent NoCs", IEEE Trans. Comput., vol. 64, no. 3, pp. 763-777, 2015.

[3] S. Subramanian., "Ordered mesh network interconnect (OMNI): design and implementation of in-network coherence", Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2013.

[4] M. Flahive and B. Bose., "The topology of Gaussian and Einsten-stein-Jacobi interconnection networks", IEEE Trans. Parallel Distrib. Syst., vol. 21, no. 8, pp. 1132-1142, Aug. 2010.

[5] J. Navaridas, J. Miguel-Alonso, J.A. Pascul, and F.J. Ridruejo., "Simulating and evaluating interconnection networks with INSEE", Simul. Modelling Practice Theory, vol. 19, no. 1, pp. 494-515, 2011.

[6] Rohit Sunkam Ramanujam., "Randomized Throughput-Optimal Oblivious Routing for Torus Networks", Student Member, IEEE, and Bill Lin, Member-2011.

[7] J.M. Camara, M. Moreto, E. Vallenjo, R. Beivide, J. Miguel-Alonso, C. Martinez, and J. Navaridas, "Twisted torus topologies for enhanced interconnection networks", IEEE Trans. Parallel Distrib. Syst., vol. 21, no. 12, pp. 1765-1778, Dec. 2010.

[8] E. Stafford, J.L. Bosque, C. Martinez, F. Vallenjo, R. Beivide, and C. Camarer., "A first approach to king topologies for on-chip networks", in Euro-Par: Parallel Processing. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 428-439.

[9] Priyanka N. Chopkar and Mahendra A. Gaikwad, Ph.D., "Review of XY Routing Algorithm for 2D Torus Topology of NoC Architecture", International Journal of Computer Application (0975-8887)-2013.

[10] Priyanka N. Chopkar and Dr. Mahendra A. Gaikwad., "Analysis of latency and Throughput of 2D Torus Topology using Modified XY Routing Algorithm", ACEEE Int J. On Communications, vol. 4, No. 2, Nov. 2013.

[11] Y. Ajima, S. Sumimoto, and T. Shimizu, "Tofu: A 6D mesh/torus interconnect for exascale computers", IEEE Comput., vol. 42, no. 11, pp. 36-40, Nov. 2009.

[12] J. Kim, W.J. Dally, S. Scott, and D. Abts, "Technology-driven, highly-scalable dragonfly topology", in Proc. Int. Symp. Comput. Archit., 2008, p. 77.

[13] S. Sobana, S.KrishnaPrabha, "An Efficient Method for Multichannel Wireless Mesh Networks With Pulse Coupled Neural Network," International Journal of Advanced Computer Science and Applications, vol. 3, No. 1, 2012.