NETWORK ON CHIP: AN OVERVIEW ANDREVIEW Of LITERATURE

[1]Binayakanta Senapati School of CSE
REVA UNIVERSITY
Bengaluru, India binayks.binay07@gmail.com

[1]Himashree M School of CSE
REVA UNIVERSITY
Bengaluru, India himashreem07@gmail.com

[1]Chanukya Surya School of CSE REVA UNIVERSITY
Bengaluru, India
chanukyasurya@gmail.com

[1]Dikshitha AV School of CSE
REVA UNIVERSITY
Bengaluru, India dikshithaav06@gmail.com

[2]Dr. Sanju V School of CSE
REVA UNIVERSITY
Bengaluru, India sanju.v@reva.edu.in

Abstract.

NoC (Network On Chip) is a new developing technology that is centered on a chip-based interconnection network. This approach was developed to address the traditional SoC (System On Chip)'s inefficiency as the compute capacity required to perform various tasks rose gradually over time. This article is a condensed version of what we learned from numerous research papers and other reliable sources while doing in-depth study on the intriguing idea of Network-On-Chip (NoC).

Keywords—Network-On-Chip (NoC), System On Chip (SoC), NoC Topologies, RiCoBiT, Multiprocessor System On Chip (MPSoCs)

I. INTRODUCTION

We are right now living in that specific period of a time in which new age state of the art advancements are being created pretty much consistently across the world. SoC (System on Chip) is an innovation that ignited this pattern by joining all the equipment elements of a common place PC framework into a minuscule gadget known as an Integrated Circuit (IC) or a chip. For correspondence and information move between many a circuit, SoC utilizes Common Bus Architecture. Nonetheless, as the requirement for handling abilities developed pair with the quantity of cycles to be done on a chip, SoC arrived at its scaling limit. It demonstrated wasteful on the grounds that as the quantity of cycles to be done on the gadget rose, so did the vital voltage and power dispersal. Nonetheless, on the grounds that the chip's power utilization is controlled, this can risk the capacity and working of circuit. Because of this control, the chip has acquired shortcoming as postponement. To resolve this multitude of issues, a novel thought known as NoC (Network-On-Chip) has been proposed, in which a solitary hub in the circuit can associate straightforwardly and proficiently with every hub in the circuit. NoC is a modified version of SoC that uses a concept called global wiring to replace the common bus design of SoC. Communication between the nodes can be made immediately via these cables. NoC architecture may be categorized into several topologies based on the number of connections established and the form in which wires are linked, with each topology having its own set of pros and cons. Furthermore, information about NoC is briefly discussed in the following sections.

II. NETWORK-ON-CHIP (NOC)

NoC or Network-On-Chip is an organization-based intercommunication subsystem on an incorporated circuit, for the most part between various modules/nodes in a System on chip (SoC). Or on the other hand in basic words, Network-On-Chip (NoC) is a plot for getting sorted out correspondence between working modules or hubs present on a similar chip. This innovation applies the hypothesis and techniques in viewof the ideas of PC organizations to on chip correspondence and brings regular upgrades over the transport and other existing correspondence designs. Because of its design, it groups a significant benefit over other existing structures as far as overseeing power proficiency and electrical properties on chip.

A. WHY NOC?

The network on the chip is a technology which is based on router-packet switching networks between the System On Chip (SoC) modules. For the most part, the wires involve a lot of the region of the chip, and this is where NoC comes into picture. SoCs with NoC interconnect textures can achievehigher working frequencies since NoC innovation lessens the equipment expected for exchanging and steering undertakings.

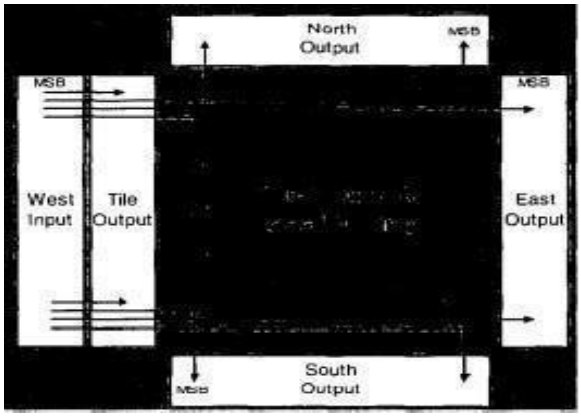
B. BENEFITS OF NOC ARCHITECTURE

- Independent layer implementation and optimization: Layers should be carried out and enhanced. The divisionof the current issues to more modest issues with a distinct normal connection is a typical way to deal with basically the difficulties of complex designing.

- Simplified application customization:
  Commands which we give and the reactions which we get are essentially bundles sent by the network at the transport layer, and anything which is done at the network layer ought to help the transport of these packets. When contrasted with a customary interconnect,this works on the customization and execution of the interconnect for a specific application.

- Supports multiple topologies and options for differentparts of the network:
  NoC interconnect allows for the use of various optimizations and topologies across different modulesof the network.

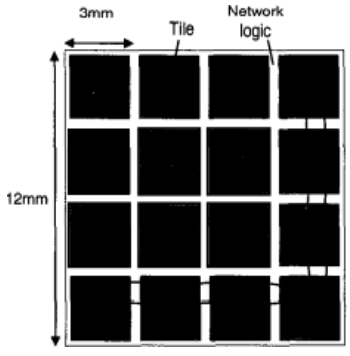
- Simplified feature development, interface interoperability, and scalability:
  Interconnects should be upgraded to support future features such as new types of transactions and burst modes. Adding new features to a connection that is organized into layers requires only minor adjustments to the layer that supports the feature. Similarly, if a network infrastructure architecture or transportation technologies outperforms the original, the old network can be replaced with no need for a network design. Traditional interconnects struggle to reach the requisite performance in systems with hundreds of masters and slaves. Interconnects built for dozens of masters and slaves are incapable of supporting the numerous numbers of parts required by today's systems. It is relatively simple to partition the network into subnetworks using NoC connection, with bridges, pipeline stages, and clock-crossing logic as needed across the whole network.



III. EXAMPLE ON CHIP INTERCONNECTION

Let us consider an example to understand what exactly an on chip-interconnection is. Let us consider a chip(12mm*12mm) which is further partitioned into16(3mm*3mm) tiles as shown in Figure1. Each tile in the above fig. holds all the required client logic. Every tile uses these client logic peripherals to connect with other tiles, and each tile has a network interface that comprises an input port for inserting a packet to be transferred to another tile and an output port for receiving packet data from another tile. A 256-bit data field is present on both the ports.

Figure1: Network logic and partitioning the chip into module tiles



IV. ROUTER ARCHITECTURE

So, in this case, the router is a device that is in charge of passing data packets in the proper direction. Each tile's router is made up of five input and output controllers (i.e., each controller for each direction as shown in the Figure2)

In Figure2, the west input controller is connected to four other output controllers, which means that the tile input and output controllers are present at the western-side of the tile. Here the network uses a simple virtual channel flow control. Each input channel contains a buffer space and an input state logic for every virtual channel to receive the incoming data packets. And each output port provides a one stage buffer to every input port connection so that the outing going data packet links with the input port of the next tile. The interesting fact is that routers, along with a few upper layer wires, take up only 6.6 percent of tile region. Because of the NoC's benefits, the network controls both pre-scheduled and dynamic load very efficiently.

Figure2: Connections from the west input to the output controllers

V. CHALLENGES IN ARCHITECTURE AND DESIGN

When compared to SoC, even though the pins and wires are more abundant in on-chip network, since all of them are organized, the on-chip interconnection network offers a massive advantage over SoC to invent new technological aspects and optimization techniques on chip. Because wires and pins are much more abundantly available in on-chip networks, each tile can be connected to every other tile present in the network and that too in more efficient and faster way when compared to SoC, which is very much pin, and wire limited. For example, a tile i has a bundle of N wires that must be properly linked to a tile j. The local logical computer supervises these wires to see if their state changes. So, depending on the shape in which the wires are connected, and the no. of connections made, NoC's connections are classified into various topologies where few topologies provide more advantage over other topologies depending upon the no. of connections made. Much of the advantage of NoC comes from the regular and structured wiring. So, the well-organized wires and electrical parameters enable us to design a very efficient and high-performance computing chip.

VI. SYSTEM ON CHIP (SOC)

With the intricacy of the inserted framework expanding what's more, the extent of current silicon innovation spreading, there is a development towards independent frameworks executed on one chip just, characterized as an On-Chip-System. A SoC is an incorporated circuit with a solitary chip, that houses PC peripherals like the CPU (through a chip or microcontroller), memory, input/yield ports, and auxiliary stockpiling on a solitary substrate like silicon. Since all parts are housed on a solitary surface, SoCs consume least measure of force and take up less space than multi-chip frameworks. SoCs are most generally found in the cell phone industry. Due to the utilization of SoCs, makers of such gadgets are presently ready to observe the littlest item that gives satisfactory execution. SoCs likewise affected installed frameworks by preparing for single-board PCs that are a lot more modest and compact. ARM is upheld by most of the SoCs available today. Drafting the NoC with all its boundaries for each SoC requires the utilization of computerized configuration apparatuses. Such apparatuses should be based on specific plan processes and integrate the calculations which are improved into their system. Bolotin et al. [6] propose a plan approach for SoCs that incorporates the three stages recorded beneath:

- First, determine the target SoC's QoS needs.
- Second, the network is customized through propermodule placement and a static, shortest path routing algorithm with a minimal gate-count cost.
- Finally, perform network load balancing byallocating adequate link capacity to reduce congestion and ensure that each communication flow's multi-class QoS requirements are met.

Now we'll describe about design phases, routing, and resource allocation:

A. ROUTING:

Shorter pathways in a preset topology can result in an unbalanced load distribution between links. The energy gain of routing on the shortest paths, for starters, is compelling. Second, pooling traffic across a smaller number of excessive-capacity lines reduces latencies compared to routing traffic over shorter channels. A number of routes with the same length and overall capacity Some links with a high degree of link sharing can eliminate the need for additional routers and links, thereby reducing chip space and power consumption. As a result, the shortest path routing scheme with path aggregation is very appealing. Practical NoC topologies become uneven meshes due to the variety in shape and size of modules in VLSI formats, as well as the requirement to physically segregate modules and NoC infrastructural facilities in the SoC environment. For messy meshes, there have been two simple alternate solution routing techniques:

a. source routing (SR), where a sequence of routing instructions is carried by packets
b. routing distribution (DR), where each intermediate router looks for the destination

Routing tables are typically used heavily by both SR and DR (RT). DR tables are found in each router. They contain the output port values and are labelled with the packet's destination address. Each source has its own set of SR tables.

CAPACITY ALLOCATION:

Capacity allocation is a major area of study in SoC. For broad flow distributions, variable connection capacity, and numerous virtual channels, a novel wormhole network analysis methodology is being developed. An evaluation is used as a foundational component in a recursive optimization process that gradually reduces connection capacity until latency requirements are met. For each link, the minimal number of cables as well as the minimum speed for a certain number of cable connections are chosen. Because of the speed, the designer can lower the voltage of any router, lowering the static and dynamic power dissipation.

VII. DIFFERENCE BETWEEN NOC AND SOC

SoC is a single chip that comprises a collection of unique and interconnected devices that can be used to solve a wide range of problems. The term "Network On Chip" refers to a technology that establishes connections between components within a SoC or processor to ensure maximum data transmission speeds while reducing the number of physical connections required. A network switch-based network between SoC modules could make up the network on a chip. NoC technology minimizes the required hardware for performing operations and switching, allowing SoCs with NoC communication networks to access high-performance speeds. In NOC, communication is formed via a touch of a packet. The delivery of packets by most hops is completed by switching. It generally follows the protocol stack's reinforced structure. A Network Interface and Router Processing functions are included in Network on Chip.

VIII. PACKET SWITCHING AND ARCHITECTURE

Switching describes how messages travel along their path. Packet switching is a network switching mechanism that does not require a connection. Packet switching combines the advantages of message and circuit switching. Data is delivered in packets rather than continuous streams in packet switching. When one node delivers a file to another, the file is split into packets to determine the most effective way to send the data over the network at the time. The packets are then routed to their destination via network devices, where they are rearranged for use by the receiving device. The header is on top, the payload is in the middle, and the tail is on the bottom of the packet.

The routing and sequencing information is found in the Header section. The payload is where data is sent. The tail of the packet is a fault correcting code that is placed at the end of said packet.

A. STORE-AND-FORWARD:

Store-and-forward switching is the technique for switching data packets in which the switching device receives the data frame and then checks for faults before passing the packets. It makes it possible to transport uncorrupted frames quickly. The data packet will not be transferred from the starting address to the destination address immediately when using this method. Rather, it awaits until the complete data packet arrives at a node before moving on to the next. Complete data packet is preserved in memory using this technique. Following that, the source address, destination address, and CRC are all verified. If there are no problems, the data packets are delivered to the target address in their entirety. This method avoids truncation or damage to data packets as they travel to their destination. This method is called as store-and-forward because it requires storing, validating, and then forwarding.

B. VIRTUAL-CUT-THROUGH:

Messages are not delivered as a complete data packet in virtual-cut-through switching; instead, they are forwarded immediately once they are retrieved by the network node. The buffer property on network nodes must match the size of the data packet. It is used to temporarily store packets. If a node's buffer is full, it simply sends a chunk of the data packet to the next node as soon as it receives it. While this sort of packet switching lowers packet transfer delay, it also increases the risk of broken or unfinished data packets being sent.

C. WORMHOLE SWITCHING:

Wormhole switching is a basic flow control mechanism ina computer network that relies mostly on fixed connections.Wormhole switching is a flow control approach that is a subclass of flit-buffer flow control. Every packet of data is furthermore categorized to flits in wormhole packet switching, and nodes are allotted buffers that are the similar size as the flits. In a pipelined sequence, they are then transported from the starting node to the end node. Despite the fact that wormhole switching overcomes the problem of blocked packets, the wormhole switching delay is much like the cut-through packet switching time.

IX. ROUTING PROBLEMS

Certain problems are likely to develop during the transmission of a packet from the sender to the receiver. These issues might arise because of congestion or a delay.Here are a few of these topics:

A. DEAD-LOCK:

When there are two packets that are waiting to be routed occur at the same moment, this is known as a routing deadlock. When each packet's resources aren't available to the other, this happens. It's a phenomenon that occurs anytime there's a cyclic resource dependency. The flow control mechanism in use is in charge for network resource management. Because no additional resources are allocated, packets that fail to release resources cause the route to be blocked.

B. LIVE-LOCK:

A live-lock occurs when data packets are continuously routed around the destination without being delivered. It's a phenomenon caused using adaptive routing algorithms, in which messages are redirected in the goal of finding a new way to their destinations. This type of routing difficulty is different from deadlock in that it runs continually, whereas deadlocks are always interrupted. Setting the TTL (time to live) or prioritizing older packets for transmission first can help resolve a live-lock.
When nodes want to communicate with each other, they infuse their messages into the network. A uniform injection model follows when all the nodes send their messages at the same time, with the network which is clear of messages.
This can be compared to dynamic injection, as per to which nodes can infuse their messages at random times. Livelock can occur if dynamic injection is used. It will not occur if the fixed injection is used. Many routing policies can also be used to prevent livelock.

C. STARVATION:

When the packet with the lowest priority fails to reach its destination, it is referred to as starvation, and the packet with a higher priority is always given more resources. Because of which, low priority packets are always denied the resources required to reach their destination. Several routing strategies can be employed to prevent starvation. The most basic strategy is to provide each node its own injection queue, where it stores the messages, it wants to infuse into the network. Non-adaptive routing algorithms ignore all these factors, lowering the suggested system's practicality and efficiency.

X. NOC TOPOLOGIES

A. MESH TOPOLOGY:

This is one of the most popular interconnection designs. Every node in this architecture is linked to its neighboring node to form a mesh. The mesh consists of n rows and m columns. The X and Y coordinates of each of these nodes are used to address them in a network. It is a form of direct topology in which nodes are connected directly. It enhances scalability while simultaneously providing significant route variety. The data is shared among all nodes. It has a large number of source-to-destination paths as well as an efficient addressing mechanism that prevents network outages. Regardless, one of the major disadvantages of this architecture is that the diameter grows exponentially as the number of nodes increases. This is due to inconsistency in the degree. All nodes are linked together in a 2D lattice, and neighboring nodes are linked together. In a mesh architecture, inter-switch delays and network robustness may be avoided. Furthermore, the linked bandwidth differs significantly from node to node, with edge and corner nodes having the least capacity. The Manhattan Street network is another name of this topology.

B. TORUS TOPOLOGY:

This network is a more advanced Mesh-topology. However, Mesh-topology has a smaller radius. Consequently, the torus topology solves the problem of mesh topology diameter increasing in parallel with network size. This is accomplished through the use of direct links among end nodes in the same row or column. The main advantage of this topology is that it may enhance path variation due to its symmetric edge geometry and low hop count. This network type offers a broader range of routing options, including more basic paths. The fundamental disadvantage of this network is that as the length of wire required to connect nodes in rows and columns becomes longer, the latency increases.

C. TREE TOPOLOGY:

Tree topology is made up of parent nodes and child nodes, with the parent nodes leading to the child nodes. The nodes above the present nodes are known as their ancestors, while the nodes below them are known as their children. There are duplicated ancestors connected to each node in a fat-tree architecture, providing numerous alternate paths between the nodes. This topology gives nodes on the network access to more system resources, and it is assisted by a number of retailers. The root node, on the other hand, is its bottleneck, and its failure might bring the entire network down. Furthermore, as the tree length increases, the network configuration becomes more complicated.

D. POLYGON TOPOLOGY:

This topology is made up of many nodes, each of which is connected to every other node in the network. A circular topology in which data packets travel in a loop, from one node to the next, is the most basic example of this topology. Concatenating the ring with chords is one way to achieve a varied routing.

E. BUTTERFLY TOPOLOGY:

The nodes in this topology are organized in the shape of a butterfly, with routers between the wings. The packet's path will be pre-defined or pre-established, and it may be unidirectional or multidirectional. Take, for example, a multi-directional butterfly network. The inputs and outputs of this network are on the same side of the topology. Before being returned to the output side, the incoming packets are routed to the other side. This architecture links many computers to high-speed networks. Processor nodes, routers, and connections are the major components of this architecture. Butterfly topology design and implementation involve network characteristics such as bisection bandwidth, degree, and diameter.

F. RICOBIT TOPOLOGY:

The RiCoBiT is composed of nodes connected in a ring, each of which is linked to two different nodes in the next ring, resulting in a ring-connected binary tree. RiCoBiT is composed of concentric rings of nodes connected by binary trees. The rings in the topology are assigned a number to K, with nodes ranging from 0 to 2L-1, where L is the present ring number. The nodes are addressed according to the ring they belong to and where they are in that ring. The level numbering system starts at one, while the node address numbering system starts at zero. Ring one is the name given to the first ring. Each ring contains a total of two L nodes. The integration formula for calculating number of nodes in the entire architecture is as follows:

$$N_r = \int_{L=1}^{K} 2^L \ldots\ldots\ldots (1)$$

RiCoBiT design has several advantages. The architecture is simple, regular, and symmetrical. It's also well-organized, scalable, and modular. Performance facilitates scalability without necessitating a significant rise in area. An optimal routing algorithm provides strong support for the architecture. The architecture is particularly good for adaptive routing in various traffic situations since there are several shortest paths. The architecture has been discovered to be superior to the presently used mesh and torus architectures.

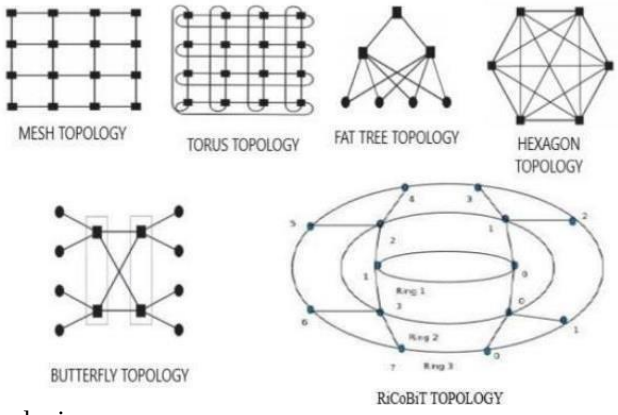


Figure 3: NOC Topologies

XI. NOC ON MULTIPROCESSOR SYSTEM

Emerging user applications in the next generation of embedded devices demand a high degree of performance. As a result, for the efficient construction of these complex future embedded structures, new techniques and connection mechanisms are required. MPSoCs (Multi- Processor System-On-Chips) are made up of complex integrated components that communicate at extremely high speeds. It would be impossible to meet the interconnectivity needs of MPSoC's with hundreds of cores by using a single bus or a hierarchy to eliminate several of the expenses of buses and MPSoC's connected via basic communication architectures. NoCs have been offered as a feasible solution to the MPSoC scaling challenge. NoCs bring packet-based communication principles to the on-chip realm, and they address many of the impending issues of interconnecting intricate designs more effectively than buses in terms of sophisticated protocols and topological design. In early NoC topology design attempts, the use of common topologies, such as meshes, such as those used in macro- networks, was predicted to lead through predictable and regular layouts. Although this is valid for architectures with homogeneous processor cores and memory, it is not valid for most MPSoC's because they are made up of heterogeneous cores, and common topologies lead to bad throughput as well as significant power and space overhead.

This is owing to the MPSoC's very non-uniform core sizes and the design's floorplan, which differs from typical topologies' regular, tile-based floorplan. An application-specific NoC with a customized architecture that meets the design objectives and limitations is required to possess an efficient on-chip connection for MPSoC's.

XII. CONCLUSION

In the paper, we explored various Network-On-Chip ideas and a few key challenges associated with packet switching network connectivity, and furthermore discussed the design of an SoC system. We additionally looked at the characters in various architectures to see how easy it would be to construct a new network. In addition, we learned how to apply network principles to the NoC system and how to construct topologies in NoC. We have also mentioned how NoC may be used to create Multiprocessor Systems. However, more research is yet to be conducted to design an NoC system that would meet the current and future needs.

REFERENCES

[1]. 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

[2]. Route Packets, Not Wires: On-Chip Interconnection Networks William J. Dally and Brian Towles Computer Systems Laboratory Stanford University Stanford, CA 94305

{billd,btowles}@cva.stanford.edu

[3]. RiCoBiT - A topology for the future multi-core processor: A concept analysis and review of literature 1 Jude Abishek Satish, 1 Hussam Taqhi, 1 Harsh Mishra, 1 Chetana Reddy P, 1 Sanju V. 1 School of Computer Science and Information Technology, REVA University,

Bangalore

[4]. Ville Rantala, Teijo Lehtonen and Juha Plosila, "Network-On-Chip Routing Algorithms," TUCS Technical Report No 779, August 2006

[5]. Sanju V, Niranjan Chiplunkar, M Khalid and P Venkata Krishna, "Design & Simulation Of SMITHA: A Structured and Scalable Architecture For 3D Network-On-Chip Systems," Advances in Systems Science and Application (2014) Vol.14 No.2

[6]. E. Bolotin, I. Cidon, R. Ginosar, and A. Kolodny. QNoC: QoS architecture and design process for network on chip. J. Systems Architecture, 50(2–3):105–128, Feb. 2004.

[7]. Network-on-Chip design and synthesis outlook: David Atienzaa, Federico Angiolini, Srinivasan Murali, Antonio Pullini, Luca Benini, Giovanni De Micheli.

[8]. Performance Analysis of various Parameters of Network-on-chip (NoC) for different Topologies: Ashish Mulajkar and Govind Singh Patel.

[9]. A. V. Bhaskar, "A Detailed Power Analysis of Network-on-Chip," 2022 IEEE Delhi Section Conference (DELCON), 2022, pp. 1-7, doi: 10.1109/DELCON54057.2022.9752850.

[10]. S. Kumar et al., "A network on chip architecture and design methodology," Proceedings IEEE Computer Society Annual Symposium on VLSI. New Paradigms for VLSI Systems Design. ISVLSI 2002, 2002, pp. 117-124, doi: 10.1109/ISVLSI.2002.1016885

[11]. U. R, S. H. and S. V., "Network-On-Chip (NoC) - Routing Techniques: A Study and Analysis," 2019 Global Conference for Advancement in Technology (GCAT), 2019, pp. 1-6, doi: 10.1109/GCAT47503.2019.8978403.