Design and evaluation of scalable 2D, 3D and hybrid interconnects for Network-on-Chip

Submitted in partial fulfilment of the requirements

for the award of the degree of

Doctor of Philosophy

by

V. Lakshmi Kiranmai

(Roll No: 718140)

Under the supervision

of

Dr. B.K.N. Srinivasarao

Department of Electronics & Communication Engineering

National Institute of Technology Warangal

Telangana, India - 506004

2024

Approval Sheet

This thesis entitled **Design and evaluation of scalable 2D, 3D and hybrid interconnects for Network-on-Chip** by **V. Lakshmi Kiranmai** is approved for the degree of **Doctor of Philosophy**.

Examiners

Research Supervisor

Dr. B.K.N Srinivasarao
Department of ECE,
NIT Warangal, India-506004

Chairman

Prof. D. Vakula
Head, Department of ECE,
NIT Warangal, India-506004

Place:

Date:

Declaration

This is to certify that the work presented in this thesis entitled **Design and evaluation of scalable 2D, 3D and hybrid interconnects for Network-on-Chip** is a bonafied work done by me under the supervision of **Dr. B.K.N Srinivasarao** and was not submitted elsewhere for the award of any degree.

I declare that this written submission represents my own ideas and even considered others ideas which are adequately cited and further referenced the original sources. I understand that any violation of the above will cause disciplinary action by the institute and can also evoke panel action from the sources or from whom proper permission has not been taken when needed. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea or data or fact or source in my submission.

Place:

Date:

V.Lakshmi Kiranmai
Research Scholar
Roll No.: 718140

NATIONAL INSTITUTE OF TECHNOLOGY

WARANGAL, INDIA-506004

Department of Electronics and Communication Engineering

CERTIFICATE

This is to certify that the thesis work entitled **Design and evaluation of scalable 2D, 3D and hybrid interconnects for Network-on-Chip** is a bonafide record of work carried out by **V. Lakshmi Kiranmai** submitted to the faculty of **Electronics and Communication Engineering** department, in partial fulfilment of the requirements for the award of the degree of **Doctor of Philosophy** in **Electronics and Communication Engineering, National Institute of Technology Warangal, India-506004**. The contributions embodied in this thesis have not been submitted to any other university or institute for the award of any degree.

Dr.B.K.N Srinivasarao
Research Supervisor
Department of ECE,
NIT Warangal, India-506004.

Place:

Date:

Acknowledgements

I would like to express my profound gratitude to my supervisor, Dr.B.K.N Srinivasarao for his valuable guidance, constructive suggestions, unwavering support and constant encouragement throughout this research. I am thankful to him for his critical suggestions during the preparation of manuscripts. The accomplishment of this work would have been challenging without his consistent supervision and continuous feedback.

I would like to express my sincere thanks to my DSC members, to Prof.Vishwanathan N, Dr.P.Muralidhar and Dr.Mohammad Farukh Hashmi, for their insightful suggestions and encouragement during the progress of this research work. I am grateful to Head of the ECE department, Prof. D. Vakula and the former Heads of the ECE department, Prof. N. Bheema Rao, Prof. L. Anjaneyulu and Prof. P. Sreehari Rao for their continuous support and encouragement. I thank all the faculty and staff of Department of Electronics and communication Engineering for their encouragement and constant moral support.

I take this oppurtunity to extend my gratitude and convey my regards to all co-scholars, friends and my well wishers for consistently being by my side. Thanks to Dr.Ch.Balaram, Dr.Sagar, Dr.Bharat, P.Hari, Mahesh, J.Nandini, Ragini, P.Raveendra, Ravi shankar, Venkat for their motivation and support throughout my work.

I express my gratitude to all my teachers, colleagues, and relatives across various locations for their support, which played a crucial role in the successful completion of this work. Special thanks to my close friends Smt.Sudha, Smt.Swapna and Smt.Bhargavi for their motivation and moral support. I would also extend my gratitude and thank my close friend Smt.Madhavi for introducing me the intricacies of programming. I am grateful and I appreciate Mr.Teja for helping me comprehend the programming concepts.

Special thanks to Prof.B.R.Phanikumar, for the motivation and encouragement,

which helped me a lot for the successful completion of my research work. Along the way, I have met many bright and enthusiastic people who have influenced me and regardless of whether or not they know they did, I sincerely appreciate their efforts.

I extend my sincere and immense sense of gratitude to my beloved family members, my parents Sri.V.V.Suryanarayana Murty and Smt.Prabhavati for their motivation, support and blessings without which it would not have been possible for me to reach so far. I would like to convey my deepest thanks to my husband Dr.K.V.R Ravi shankar and my beloved son K.Srikrishna for their unwavering support, constant encouragement, patience, understanding and remarkable co-operation. I would like to express my sincere gratitude to my brother and sister-in-law -Sri.V.V.J.S.Subrahmanyam and Smt.Sirisha and their children Pranavi and Abhiram for their support and encouragement during my research. I express my profound gratitute to my elder brother-in-law and sister - Sri K.Ch.V.S. Kameswara Rao and Smt. Durga Satyavathi and their children, Praneeth and Lakshmi, Pratyush and Monica, second elder brother-in-law and sister - Sri K.V.S. Ramakrishna Rao and Smt. Prasanna, and their children, Shriya, Sneha for their support and encouragement throughout this endeavour. Last, but by no means least, I express my gratitude to my in-laws, Sri K.V.M. Sharma and Smt. K. Ratnam for their blessings.

Above all, I am grateful to GOD almighty for showering the grace and endowing me with the wisdom, health, and strength necessary to undertake and successfully complete this research endeavor.

V. Lakshmi Kiranmai

Abstract

Network-on-Chip (NoC) is an emerging and efficient on-chip interconnect technology. NoC is a viable option to design modular, scalable, robust communication interconnect architectures. Topology is one among many crucial design aspects of NoC, as it affects the performance of the interconnection network. Mesh is the most extensively used and favoured architecture for implementing less sophisticated SoCs due to its simple, scalable, regular structure, low-radix routers and short-range links. However, as the network scales, Mesh suffers from degraded performance because of large diameter. The present work aims at developing efficient and scalable novel topologies – 2D, 3D topology, hybrid wired topology and hybrid wired-wireless topology for on-chip interconnect architectures that outperform Mesh topology.

The objectives of the research are threefold – First, to design a hybrid wired topology i.e., combining two topologies. Second, to design a diagonal Mesh based topology by inserting diagonal links into the conventional Mesh topology retaining the simple, scalable, regular structure of Mesh simultaneously improving the performance. Third, to design a three-level hierarchical hybrid wired-wireless interconnection architecture for large networks.

To begin, the present work proposes a novel, scalable, hybrid Hexagonal Star (HS) topology for on-chip connectivity networks. The proposed topology's properties have been investigated and compared to those of the Mesh, Torus, and Honeycomb Mesh topologies. The performance of the Hexagonal Star topology has been studied and compared to that of the Mesh topology in different scenarios. The comparative studies of topological properties have indicated that the proposed topology can be a potential choice for on-chip interconnection networks. For different traffic patterns and traffic loads, HS topology has registered a reduction of packet latency ranging from 15% to 50% and from 9% to 23% for

18 nodes and 32 nodes, respectively, compared to Mesh topology. Further, the synthesis results indicate a significant reduction of area consumed by HS topology compared to the area consumed by an identically configured Mesh topology.

Second, the present work proposes DiamondMesh, an area and energy efficient diagonal mesh based topology. By incorporating diagonal links into the basic mesh topology, the proposed DiamondMesh increases network performance while keeping the Mesh topology's regular, simple, and scalable features. Topological properties of DiamondMesh have been explored and compared with that of other competitive diagonal mesh topologies. The proposed topology and other state-of-the-art diagonal Mesh topologies have been simulated and synthesised. The evaluation results indicate that there has been a significant reduction of latency compared to Mesh and other diagonal mesh topologies except DMesh and a considerable reduction of area and power compared to the DMesh topology.

Finally, in order to address the limitations of electrical interconnects, the current work investigated hybrid wired-wireless topologies. Three-level hierarchical hybrid wired-wireless Network-on-Chip (NoC) designs have been proposed and evaluated under different traffic patterns at low, medium, and high traffic loads. In the proposed three-level hierarchy, the bottom and top levels are concerned with subnet topology and wireless hub topology, respectively. The present research introduces a middle level that investigates the number of nodes that will be connected to a subnet's wireless hub. Mesh and fully connected wireless topologies have been chosen for the bottom and top levels of the hierarchy, respectively. Different hybrid wired-wireless configurations have been developed and studied by varying the number of subnets and the number of nodes to be attached to the subnet at the middle level. The objective of investigating the middle level is to reduce the number of subnets and, consequently, the number of wireless nodes in large network architectures without compromising performance. The proposed hybrid architectures outperform baseline wired Mesh architecture in terms of latency and throughput characteristics.

Index Terms : Network on Chip (NoC), Mesh, wireless NoC, hybrid NoC, Hexagonal star, DiamondMesh, Diagonal Mesh, hierarchical WiNoC, hybrid NoC topology.

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two Dimensional
3D	Three Dimensional
AAFR	Average Aaccepted Flit Rate
APL	Average Packet Latency
DOR	Dimensional Order Routing
DSM	Deep Submicron
EDP	Energy-Delay-Product
EPP	Energy per packet
FF	FlipFlop
GAD	Global average delay
HCM	HoneyComb Mesh
HN	Hub node
HS	Hexagonal Star
IC	Integrated Circuit
IP	Intellectual Property
IR	Injection Rate
LUT	Look Up Table
MPSoC	Multiprocessor System-on-Chip
NoC	Network-on-Chip
ns	nano second
NT	Network Throughput
PE	Processing Element
PIR	packet injection rate
PLP	Power-Latency-Product
SN	Subnet node

SoC	System-on-Chip
TE	Total Energy
TNP	Total Network Power
TSV	Through Silicon Via
VC	Virtual Channel
VLSI	Very Large Scale Integration
WiNoC	Wireless Network-on-chip
WI	Wireless Interface
WLs	Wireless Links
WR	Wireless Radiohub

Chapter 1

Introduction

Network-on-Chip (NoC) is an interconnect architecture that is scalable, flexible, robust, and communication-centric. It is a game-changing solution for connecting a large number of cores in the complex Multiprocessor System-on-Chip (MPSoC) domain [6]. With rapid technological advancements, the number of components integrated on a standard sized IC (a chip), has been increasing in accordance with Moore's law. The definition of a chip changes with the integration scale. In LSI (Large Scale Integration) systems, a chip represents a component of a system module. In VLSI (Very Large Scale Integration) systems, a chip represents a module of a system. In ULSI (Ultra Large Scale Integration), a chip represents an entire system which is termed as a System-on-chip (SoC). As the technology scales, the complexity of SoC increases thereby posing serious challenges to researchers. In the Deep Submicron (DSM) era, to cope with increasing performance requirements of Multi-processor Systems-on-Chip (MPSoCs), on-chip interconnect architectures have evolved from shared buses to hierarchical buses and then to bus matrix architectures. Bus architectures have limited scalability. As the complexity of the SoC architecture increases, communication becomes a bottleneck to performance. Communication as well as coordination becomes a major concern. Growing chip density and demand for efficient and scalable communication infrastructure has led to the evolution of NoCs which address the limitations of Bus infrastructure. An efficient interconnection architecture should possess the potential to exchange data with low latency, low power consumption while occupying limited space. Ideally, communication infrastructure should be designed in a way that it brings trade-off among power, area and delay metrics [6–8].

1.1 Fundamentals of Network-on-Chip (NoC)

1.1.1 Evolution of NoC

Ever since the introduction of the SoC concept, the SoC communication infrastructure has generally been facilitated by custom buses and point-to-point links. With increasing number of cores, the communication infrastructure has undergone an evolution from custom buses to NoC as shown in Figure 1.1.

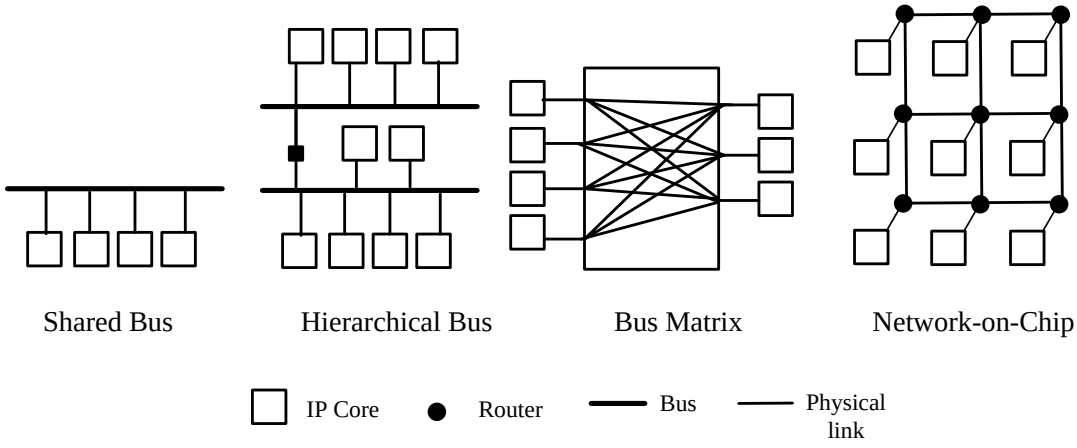


Figure 1.1 Evolution of NoC [1]

Shared buses such as AMBA bus and CoreConnect are popularly used communication infrastructure in SoCs. Even though they support a modular design approach, they become a performance bottleneck as the system bandwidth requirements scale up. Hierarchical Bus using multiple buses or bus segments alleviate the load on the main bus. This hierarchical structure allows communication between modules on the same bus segment without causing congestion on other bus segments. This structure is complex, offers less flexibility and scalability. Bus Matrix, a crossbar bus architecture, consists of a matrix switch fabric. To facilitate parallel multiple communications, this matrix connects all the inputs with all the outputs. However, the design of crossbar-based architecture is complex for multicore SoCs. With multi-core SoCs, bus architectures become a communication bottleneck. The success and scalability observed in switch-based networks and packet-based communication in parallel computing have motivated researchers to introduce the Network-on-Chip (NoC) architecture as a promising solution to meet intricate

on-chip communication challenges [1; 9].

1.1.2 NoC: A layered architecture

Network-on-Chip (NoC) has evolved as a viable solution to meet the performance, bandwidth and power demands of complex MpSoC architectures. NoC has moved from computation-centric designs to communication-centric designs. NoC architecture can be realized as a layered architecture consisting of application, transport, network, data link, and physical layers as shown in Figure 1.2. A NoC router should contain both hardware and software to support the layers [2; 3].

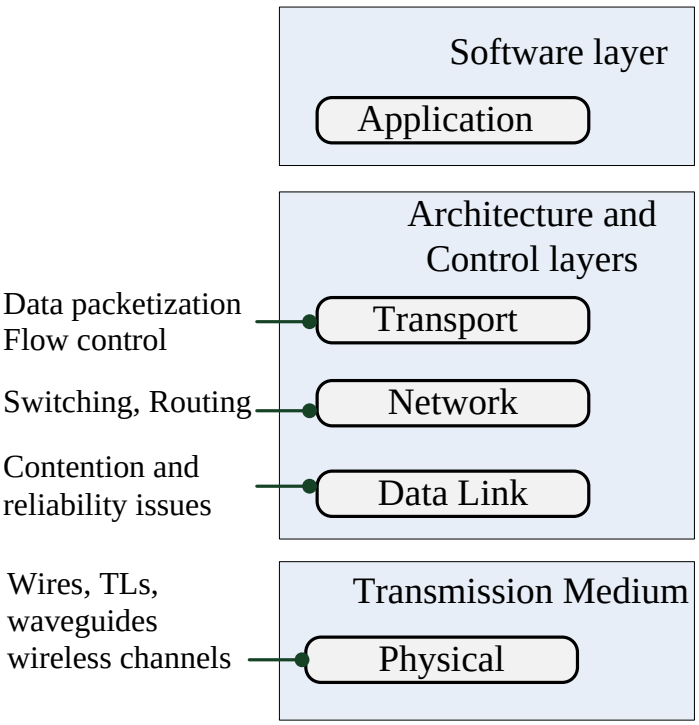


Figure 1.2 Layered architecture of NoCs [2]

1.1.2.1 Application layer

An application can be represented as a set of computational units that are interconnected by communication infrastructure to exchange data. The computational units are termed processing elements (PEs). Network on chip forms the communication infrastructure of the application. The software layer is made up of system software which provides an abstraction of the underlying hardware components. The application is thus

implemented as a set of computational and communication tasks by the software. This software is tightly coupled to the network interface (NIs) which connects the PEs to the interconnect fabric. To get the best trade-off between power and performance, application mapping and scheduling should be considered with several kinds of architecture parameters.

1.1.2.2 Transport layer

The transport layer deals with packetizing algorithms, which decompose the message into packets at the source and assemble them at the destination to reproduce the message. The granularity of packetization i.e., the size of the packet, is a significant design parameter. It has an impact on the behaviour of network control algorithms which affect the performance, power and area footprint of NoC interconnect. Also, the transport layer addresses the congestion and flow control issues. It has to implement certain flow control mechanisms to assure congestion free routing of the packets, correct delivery of data with efficient utilisation of resources.

1.1.2.3 Network layer

Network layer is responsible for switching and routing the packets from source to destination. Switching techniques are used to establish the type of connection and routing algorithms determine the path strategies to transmit the packet from source to destination. Selection and implementation of the routing algorithms depend on two important aspects – determining the packet routes and resolving the conflicts if two packets request for the same route at the same time. Also, Network layer deals with the network topology or interconnect architecture which determines the physical layout of the processing elements (PEs) and the interconnections among the PEs. A suitable topology for the application is selected first, based on which routing algorithm is chosen.

1.1.2.4 Data Link layer

The Data Link Layer is divided into two sub-layers: bottom sub-layer which is closest to physical layer is Medium Access Control (MAC) sub-Layer and the upper sub-layer which is closer to the network layer is the Data Link Control (DLC) sub-layer. MAC sub-layer is responsible for regulating access to the physical layer and resolving contention issues. DLC sub-layer is responsible for error detection and correction schemes, error-

tolerant schemes to ensure reliable operation of the system.

1.1.2.5 Physical layer

The physical layer is concerned with physical links or any other alternatives to wires which can act as the physical medium of data transmission. Bus architectures performance is limited by the long interconnect lengths, particularly in DSM technologies where wire delay prevails over gate delay. Pipelining or inserting buffers/registers to divide the longer interconnects into shorter segments can be employed to alleviate long delays. However, this approach often results in inefficient resource utilization, as the entire bus remains busy until a data transfer is completed, preventing the use of vacant wire segments by other masters. Network-on-Chip (NoC) fabrics offer a solution to the challenges posed to lengthy interconnects and inefficient utilisation of Bus resources that result in excessive wire delay. NoCs employ shorter, regularly sized wires (links) separated by switches (routers). Data packets traverse these links, also known as switch-to-switch links, within a single cycle, after which they are buffered in the routers before being routed to another link in the subsequent cycle. Unlike buses, NoCs are inherently designed to address wire delay from the outset, providing a more structured approach that simplifies the designer's efforts and reduces overall design time. In the giga billion transistor era or kilo-core MPSoC era, the wires (interconnects) cannot scale to the extent technology scales. So, in order to cope with the performance requirements of the complex MPSoCs, which even traditional NoCs are not able to meet, other promising technologies have evolved. The promising technologies which can replace the traditional metallic interconnects include Photonic Interconnects, Wireless Interconnects, Carbon Nanotubes (CNT).

1.1.3 Basic building blocks of NoC

A typical NoC architecture as shown in Figure 1.3 comprises a set of processing elements (PEs) or IP core interconnected by means of a micro network of routers. The routers of the network are linked by multiple wire segments known as links. A network interface (NI) of a PE makes a logical connection between the PE and the network. NI packetizes the data generated by PE. The NI is connected to a router of the micro network. The input buffers of the router accepts data from NI or from other routers of the network. In Figure 1.3 (a), Mesh topology is considered [3] .

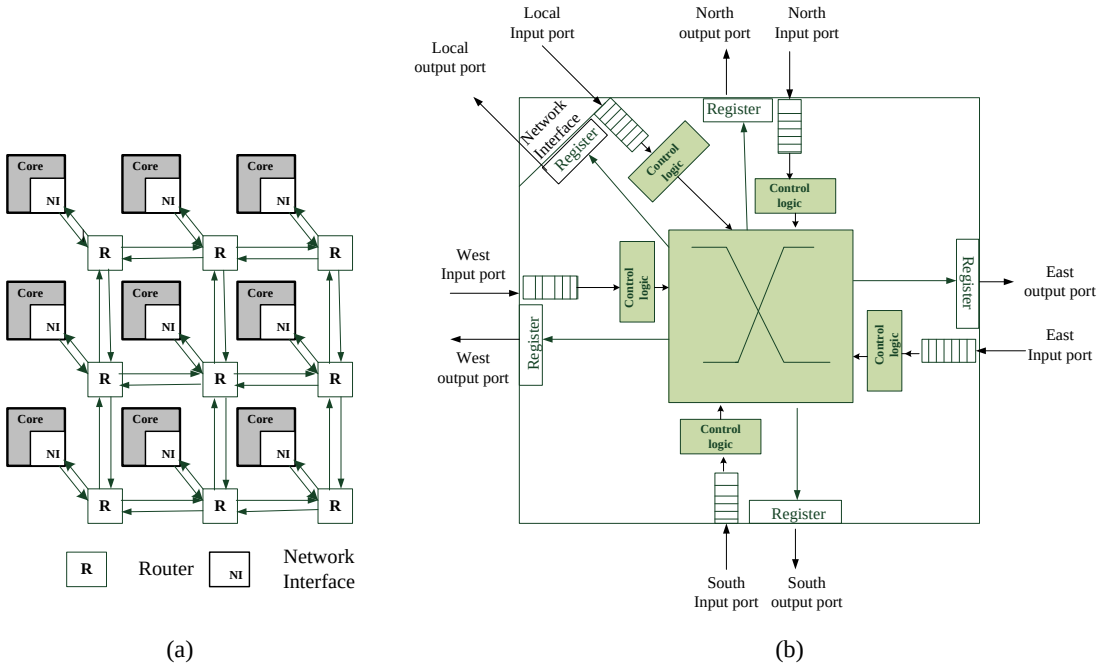


Figure 1.3 (a) A typical NoC architecture; (b) A generic NoC router [3]

The basic building blocks of the NoC fabric are the Links, Router and Network interface.

A Link is composed of a set of wires and connects two routers in the network to transmit data among the cores of the architectures. Link consists of one or more physical or logical channels, each channel comprises a set of wires.

NoC router comprises a number of input and output ports, buffers, control logic, crossbar switch. Input and output ports connect the router to its neighbouring routers in the network or the PE. The control logic block implements the routing and flow control policies such as routing computation, arbitration. Crossbar switch connects the input ports to the output ports. The primary design aspects of the router includes implementing routing and flow control strategies, arbitration and switching logic, buffering.

Network Interface also called Network adapter makes the logical connection between the PEs and the network. It decouples computation from communication. It implements two interfaces, a front end and a backend. Frontend manages the core requests while backend handles the network protocol such as assembling and disassembling of packets,

implementing synchronization protocols and thus facilitating a layered design approach.

Topology, routing and flow control are critical design aspects of NoC architectures. The subsequent sections provide an overview on these aspects.

1.1.4 Topology

Topology is one among many key design aspects of the interconnection architecture. Topology refers to the physical arrangement of nodes and the connecting links among the nodes of the interconnection architecture. It is the most critical design aspect of NoC, as it deals with wire lengths, node degree and routing strategies that affect the performance of the architecture. Topology determines the number of hops required for a packet to traverse between a pair of source and destination nodes thereby influencing the network latency and throughput. Topology determines the wired link lengths thereby influencing the area and power consumption of the architecture [10]. Further, topology determines the total number of alternate paths between a pair of nodes. With a greater number of alternate paths, higher throughput can be achieved and faster exchange of messages can be achieved by employing adaptive routing strategies. During the past few years, many NoC topologies like Mesh, CMesh, Torus, Spidergon, Octagon, Tree, Mesh of Tree, Honeycomb, Ring, Hexagonal, etc. have been proposed [10–14]. Star, Ring, Spidergon, Binary Tree, Mesh, and Torus are among the commonly utilized topologies in Network-on-Chip (NoC).Each topology has its own merits and demerits. Ring Topology is the most simple and one of the most widely employed topologies. In the Ring topology, all the nodes are connected in a circular fashion forming a closed loop as shown in Figure 1.4 (a). In this, each core is connected to exactly two neighboring cores. Design, deployment, and troubleshooting are easier. However, it has very poor scalability and poor path diversity.

Star topology connects all the nodes to a single centrally located hub node as shown in Figure 1.4 (b). Irrespective of the number of nodes, the network diameter excluding the hub is 2. The topology is simple and offers a minimum hop count. However, failure of the central hub results in the failure of the entire system. It is suitable for small network sizes.

Spidergon is similar to Ring topology enriched with across links between opposite nodes as shown in Figure 1.4 (c). For a node, clockwise, counterwise and across links

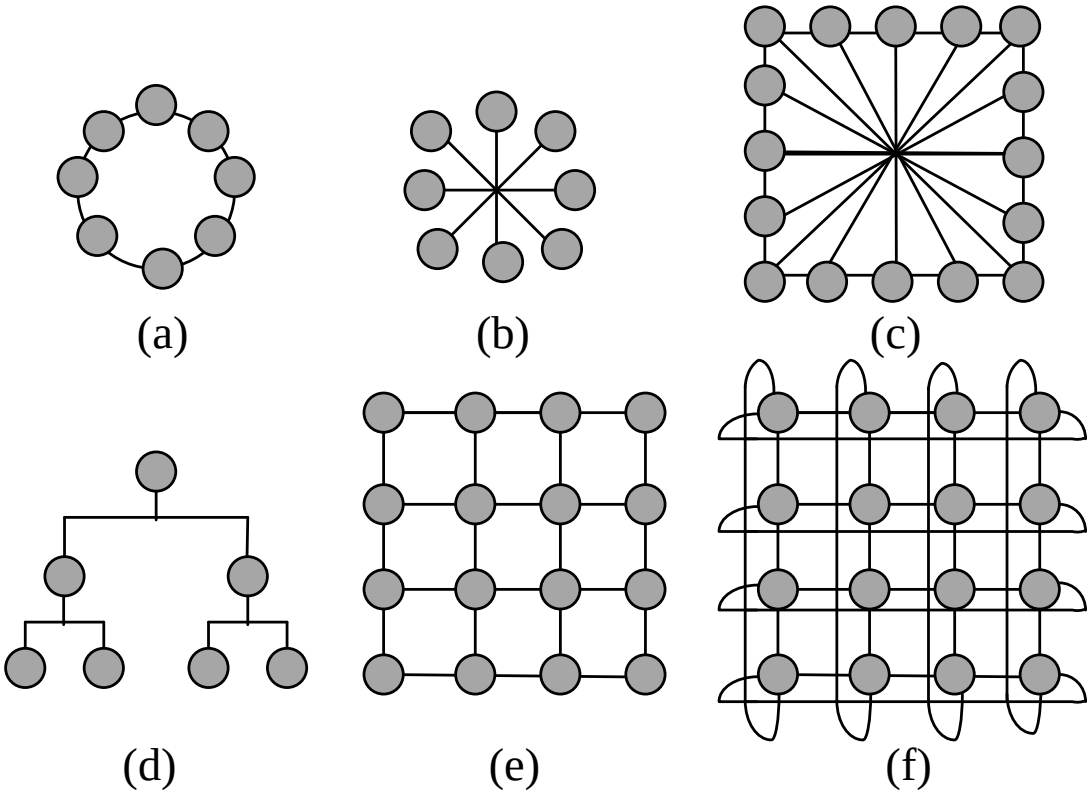


Figure 1.4 NoC Topologies [1]

are present. This topology is simple, regular with vertex symmetry, and offers a constant node degree of 3 translating to simple and cost-effective routers. The topology is only for an even number of nodes. However, the topology has poor path diversity and fault tolerance.

Mesh topology comprises cores or nodes connected as a grid as shown in Figure 1.4 (d). Each single core is connected to a router. Each router except those at the corners and boundaries is connected to four neighboring routers. It has a simple, regular, and scalable structure. Mesh provides good fault tolerance because multiple paths exist between a pair of nodes. It has a drawback of a larger diameter and it increases with an increase in system size. This is due to the nonuniform node degree. The degree of the corner, edge, and inner nodes is 2,3,4 respectively. Because of this the associated bandwidth of corner and edge nodes is lower than that of the inner nodes.

Torus topology, a topology similar to the Mesh topology, is shown in Figure 1.4 (e). The limitation of the large network diameter of Mesh is addressed by Torus topology. Torus is formed by directly connecting the end nodes that are in the same row or same

column resulting in uniform network diameter of 4, a better bisection width. However, as the system size grows, the long wrap-around links leads to increased latency and increased complexity.

Considering the benefits of scalability, path diversity, fault tolerance, link complexity, short range links, simple structure of Mesh topology, it has become a preferred choice for less sophisticated networks. However, as the system size grows, the network diameter of the Mesh increases resulting in diminished performance.

1.1.4.1 Significant topological parameters considered

Any interconnect network's performance is determined by its topological characteristics such as network diameter, bisection width, node degree, number of links, network cost, and packing density.

1. **Network diameter (Diameter):** It is the shortest distance between the two most distant nodes in the network. In other words, once the shortest path length from every node to all other nodes is calculated, the Network diameter is the longest of all the calculated path lengths.

2. **Node degree:** The node degree is the number of edges connecting a node to its neighbouring nodes. (Node refers to router or processing element (PE). If the node is a processing element, the node degree is '1' i.e., only the one edge that connects the PE to its router. If the node is a router, then the node degree includes the edges that connect the router to its neighboring routers and the edge that connects the router to the PE).

3. **Bisection width:** The bisection width of a network is the minimum number of links required to be removed for a network or graph to be bisected.

4. **Average Distance:** It is the average of the distances (hop count) between all pairs of nodes in a network.

5. **Network cost:** The Network cost is the scalar product of the Node degree and Network diameter.

$$Network\ cost = Network\ diameter \times Degree$$

6. **Scalability:** Scalability is defined as the property of expanding the network size, getting consistent performance.

7. **Link:** The physical channel that connects a node to another node.

8. **Packing density:** The packing density of a network is defined as the total number of nodes per total network cost. It indicates the size of the chip area of the VLSI layout.

   $$Packing\ density = (Total\ number\ of\ nodes)/(Network\ diameter \times Node\ degree)$$

The interconnection network with smaller diameter, larger bisection width, smaller node degree, a greater number of links, low network cost and high packing density achieves superior latency and throughput characteristics. A smaller diameter indicates fewer hops to reach the most remote nodes. Large bisection width enables a faster exchange of information. More number of links support higher bandwidth. Smaller node degree means an easier to build network. Larger packing density indicates smaller chip area of VLSI layout [10].

Mesh is the most extensively used and favoured architecture for implementing less sophisticated SoCs due to its simple, scalable, regular structure, low-radix routers and short-range links. However, as the network scales, Mesh suffers from degraded performance because of large diameter [15]. There are numerous approaches to optimize the performance of the network. Some of the approaches are bypassing some intermediate nodes by inserting express channels in k-ary n-cube networks to reduce latency [16], inserting long range links between distant nodes in the standard mesh [17–19], employing larger buffers, increasing number of virtual channels (VCs) [20; 21], improving router architecture [22; 23], inserting diagonal links in the regular Mesh [24–27], combining two topologies [28; 29] or two interconnect technologies i.e., hybrid topologies [30; 31].

1.1.5 Routing

Routing determines the traversal path of a packet from the source node to the destination node. Routing algorithms can be classified based on several categories. Routing

algorithms can be classified as source or distributed based on the node at which routing decision takes place. In source routing, pre-computed routing tables are stored at the node's network interface (NI). In distributed routing, each packet carries the source address and destination address. The routing decision is implemented in each routing either by a routing table or by executing a function. Further, depending on the adaptability, both source and distributed routing can be classified as deterministic (or static), oblivious and adaptive (or dynamic). In deterministic routing, packets consistently follow a predefined path from source to destination, ensuring in-order delivery. In contrast, oblivious routing selects paths randomly or cyclically. In both deterministic and oblivious routing, current state of the network is not taken into account. In contrast, adaptive routing makes decisions based on the current state of the network such as congestion, links availability e.t.c., dynamically choosing alternative paths to bypass congested or faulty links. Consequently, in-order delivery of the packets is not assured, necessitating packet reordering at the destination network interface (NI). Adaptive routing can be further classified as progressive routing and backtracking routing. In progressive routing, the header moves forward, reserving a new channel at each router. Backtracking routing permits the header to backtrack as well, releasing previously reserved channels. Backtracking algorithms are mainly used for fault-tolerant routing. Deterministic and adaptive routing can be categorized as minimal or non-minimal, depending on the number of hops taken from source to destination. In minimal routing, routing path from the source to the destination is the shortest possible between two nodes. Non minimal routing can use longer paths when a minimal path is not available, which can be advantageous for circumventing network congestion. Delay and power consumption are higher in non-minimal routing than in minimal routing as it requires more number of hops. Adaptive routing that follows a minimal path from source to destination can be classified as fully adaptive and partially adaptive [10].

The key challenge for any routing algorithm is to ensure that the routing remains free from livelock, deadlock and starvation conditions. Livelock occurs when a packet does not reach its destination, because it enters a cyclic path. Deadlock occurs when a packet does not reach its destination, because it is blocked at some intermediate resource. Starvation occurs when a packet does not reach its destination, because some resource does not grant access while it grants access to other packets.

1.1.6 Flow control

Flow control determines how network resources such as buffers, links e.t.c., are optimally allocated to the packets traversing in the network. The objective of flow control techniques is to avoid or manage congestion, provide error or fault recovery mechanisms for re-transmission of the packets in the network. Flow control protocols can be implemented at an end-to-end level or at a switch-to-switch level of the network. An end-to-end flow control results in higher average packet latency compared to a switch-to-switch flow control [10].

Credit-based flow control protocol is a standard protocol implemented for end-to-end scenario. In this protocol, an upstream node keeps count of data transfers, and the available free slots are referred as credits. When a transmitted data packet is consumed or forwarded, a credit is sent back. Switch-to-switch flow control can be further classified as flit level and packet level. In a switch-to-switch flow control, packet level shows higher latency than that of flit level. In NoC, flit level flow control technique is widely used. There are three types of switch-to-switch flit-level flow control protocols, namely, STALL/GO, T-Error, and ACK/NACK. STALL/GO is a simple implementation of an ON/OFF flow control protocol. It uses only two control wires: one that signals data availability going forward, and another that signals buffer status going backward, indicating either filled buffers (”STALL”) or free buffers (”GO”).” The T-Error flow control protocol is very complex as compared to other flow control protocols. It aims to improve either link performance or system reliability by catching timing errors. The T-Error protocol handles communication over physical links by either extending the distance between repeaters or increasing the operating frequency compared to a conventional design. T-Error lacks a really thorough fault handling in a real-time system operating in a noisy environment. T-Error’s fault handling in a real-time system operating in a noisy environment is not adequately robust. In the ACK/NACK protocol, a copy of a data flit is stored in a buffer until an ACK signal is received. Upon receiving an ACK signal, the flit is removed from the buffer. However, if a NACK signal is received, the flit is scheduled for retransmission [10].

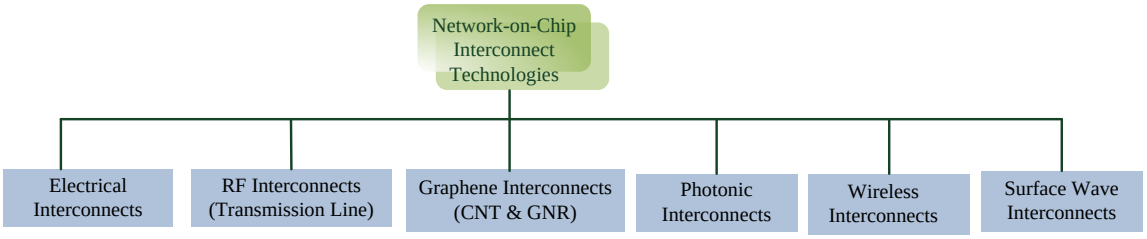


Figure 1.5 On-Chip interconnect technologies

1.1.7 Interconnect technologies

As the trend to integrate more number of components on a single chip and demand for MPSoCs continues to grow, even the communication-centric NoC fabric becomes a limiting factor to the performance and power budgets of the architecture. This is because of the technological limitations of the electrical interconnects to scale power and delay at the same rate as that of transistors. The traditional electrical interconnects are not able to meet the requirements of complex SoCs, because with an increase in system size, the parasitic resistances and capacitances of the electrical interconnects increase leading to high power dissipations. The electrical interconnects may not be able to meet the power and performance demands of the future MPSoCs. Either they have to be augmented or even replaced with other emerging interconnect technologies like photonics and wireless.

Based on the interconnect technology, on-chip interconnects can be categorised as shown in Figure 1.5. Chapter 2 provides a review of the basic concepts, limitations, advancements of state-of-the-art and emerging on-chip interconnect technologies.

1.2 Motivation

Hybrid topologies have gained focus in NoC research. Hexagonal-shaped tiles pack efficiently without gaps between them. Non-Manhattan layouts reduce the wire lengths thereby reducing the area and power of the architectures. To fully utilize the benefits of non-Manhattan layouts, the architectures are to be implemented on hexagonal or octagonal ICs with maximal chip area usage. From previous studies, it is observed that hexagonal topologies have been less explored. Considering the benefits of hexagonal-

shaped structures and non-manhattan layouts, the authors were motivated to explore hexagonal-based hybrid topologies.

Further, with a demand for simple and scalable architectures to meet the performance requirements of complex SoC, there is a need to explore the topologies that can surpass Mesh topology in performance while retaining the simple, scalable and regular structure of Mesh. One among many approaches to address the limitation of Mesh topology is to incorporate diagonal links in the baseline Mesh in such a way that it outperforms Mesh in performance simultaneously bringing a trade-off among area-performance-power metrics in comparison with state-of-the-art diagonal mesh-based topologies.

Traditionally, Electrical wired interconnects have been utilised as communication infrastructure for SoCs. Electrical interconnects offer benefits such as lower power consumption, higher bandwidth, and smaller footprint for short-range communication. However, technology advancements enable the integration of more components on a single chip. With technology scaling, resistance of the wires increases, leading to higher interconnect delays. Network performance is significantly influenced by the interconnect wire delays, thus making it increasingly important. With growing chip density and demand for MPSoC, electrical interconnects may no longer meet the stringent performance and power requirements. This necessitates the exploration of alternative technologies, such as photonics and wireless communication.

1.3 Research Objectives

The objectives of the research are threefold – Following are the objectives composed for the present work:

- To design an area efficient high performance hybrid topology for on-chip interconnection networks.
- To design a simple, scalable energy efficient high-performance topology for on-chip interconnection networks suitable for many core SoCs.
- To design a hybrid wired-wireless on-chip interconnect architecture for kilo-core

architectures

Precisely, the work aims at developing efficient and scalable novel topologies – 2D-3D topology, hybrid wired topology and hybrid wired-wireless topology for on-chip interconnect architectures that outperform traditional Mesh topology.

1.4 Organization of the Thesis

The thesis presents a detailed study of and investigation on different 2D, 3D and hybrid topologies of on-chip interconnect architectures with a motive to address the limitations of widely used state-of-the-art interconnect topologies and to explore topologies that can bring a trade-off among the latency, power and area metrics enabling their application in complex SoCs.

The thesis is organized into six chapters. The following provides the outline of the chapters of the thesis.

Chapter 1 presents a brief introduction to the basic concepts, motivation and objectives of the present research.

Chapter 2 provides a literature review of the state-of-the-art on-chip and emerging interconnect technologies, on-chip interconnect topologies, and the numerous approaches that improve the efficiency of the architectures.

Chapter 3 proposes a novel hexagonal-based hybrid wired topology named Hybrid Hexagonal Star. The topological parameters of the proposed topology have been explored. Design methodology, performance evaluation, simulation and synthesis findings have been discussed.

Chapter 4 proposes a diagonal Mesh based topology named DiamondMesh. Design methodology, performance evaluation, simulation and synthesis findings have been discussed. DiamondMesh has been compared with state-of-the-art diagonal Mesh-based topologies. Further, the 2D Diagonal Mesh-based topologies are extended to 3D and have been evaluated. Also, different 3D-heterogeneous Diagonal Mesh-based topologies have been suggested and evaluated.

Chapter 5 proposes three-level hierarchical hybrid wired-wireless Network-on-Chip (NoC) architectures. Various hybrid wired-wireless configurations have been proposed and examined under various traffic patterns at low, medium, and high traffic loads. 256-node and 1024-node hybrid wired-wireless architectures have been analysed and compared with that of identically configured 256-node and 1024-node baseline wired Mesh architectures respectively in terms of the metrics Global Average Delay (GAD), Network Throughput (NT), Total Energy (TE), Energy-Delay-Product (EDP) and Energy per packet (EPP).

Chapter 6 summarizes the findings of the thesis and proposes scope for further research.

Chapter 2

Literature

2.1 On-chip interconnect topologies

This section provides a survey of on-chip topologies, specifically focusing on hexagonal-based, diagonal mesh-based, and hybrid wired-wireless topologies.

The mesh topology, which is illustrated in Figure 2.1(a), is widely employed in on-chip interconnect architectures owing to its simple, scalable, and regular structure. Nonetheless, its performance is constrained by its narrow bisection and large diameter. As a result, it is necessary to investigate potential strategies for enhancing the performance of the Mesh or alternative topologies that can deliver exceptional performance for large networks. There exist multiple methodologies to augment the performance of NoC architectures when applied to larger nodes. These encompass various strategies such as integrating express channels between distant nodes, utilizing sizable buffers, incorporating diagonal connections into the baseline mesh, merging two topologies, investigating alternative mesh structures, and contemplating alternative technologies.

2.1.1 Numerous approaches to enhance the performance of NoCs

Kim *et al.* [17] have described the usage of high radix routers and mapping of flattened butterfly topology (FBT) for on-chip networks. FBT has been developed by combining or flattening the routers in each row of the conventional butterfly network into a single router. In comparison to conventional topologies such as Mesh and concentrated Mesh (CMesh), the FBT topology with high-radix routers and long-range links has demonstrated a decrease in latency and energy usage. Umit Y. Ogras *et al.* [18]

have introduced a novel design methodology for inserting application-specific long range links into a baseline Mesh, thereby reducing average packet delay and increasing Mesh throughput. The insertion of long range links depends on the traffic provided to that specific application. Grot *et al.* [23] suggested Multidrop Express Channels (MECS) networks built of unidirectional point-to-multipoint links based on express cubes. MECS has outperformed other similar topologies in terms of low-load latency and energy efficiency. With the increase in network size, inserting long range links appears to be an inefficient way as the complexity, area overhead and power consumption increases. Alternatively, network performance can be enhanced by increasing the number of virtual channels and buffer resources. However, this process consumes more space and energy. Feiyang Liu *et al.* [20], A. V. Bhaskar *et al.* [21] have investigated the performance of virtual-channel router at different buffer schemes. The results have indicated that upon increasing the number of virtual channels, the throughput has improved and latency has reduced. However, beyond a particular limit, increasing the number of virtual channels has not significantly improve throughput or latency characteristics. Nicopoulos *et al.* [22] have developed a novel router architecture called ViChaR, a dynamic Virtual Channel Regulator that dynamically assigns VCs based on network traffic. Instead of the static resource allocation employed by conventional routers, ViChaR permits the deployment of a flexible and dynamically varying virtual channel management method. At each input port, ViChaR allows a customizable number of VCs. Compared to conventional router with identical buffer size, ViChaR has showed performance and power improvement under various traffic patterns with more complicated logic. Sayed MS *et al.* [23] have presented a novel Flexible router architecture that makes efficient use of available buffers to alleviate the contention problem. In contrast to ViChar, this router architecture can be built with or without virtual channels. However, in a Flexible router design, packets may be received out of sequence. Wang *et al.* [28] have proposed a new interconnect architecture called TM that includes some advantages of both Mesh and Torus topologies. TM can be derived from Torus by removing one link in each row (column) properly to break the cycles. Following certain rules, the links have to be removed such that the average distance and diameter of the network reduces. The authors have also proposed deterministic and fully-adaptive deadlock-free routing schemes for the 2D TM network. TM has been shown to perform between Mesh and Torus. Moudi *et al.* [29] have proposed a new topol-

ogy called x-Folded TM. x-Folded TM is a TM topology that is folded according to the x-axis. x-Folded TM has reduced average distance, diameter and cost when compared to that of Mesh, Torus and TM topologies. The authors have discussed and analysed the topological properties of the x-Folded TM and also presented a routing and deadlock avoidance scheme for x-Folded TM topology.

2.1.2 Hexagonal based topologies

In this section, the study has focussed on exploring topology alternative to Mesh. Hexagonal and geometrically similar to hexagonal topology like Honeycomb have been reviewed. Hexagonal and HoneyComb networks find applications in many scientific fields like digital image processing [32], computer graphics [33] , cellular networks [34]. Chen *et al.* [11] developed an addressing scheme, routing and broadcasting algorithms for hexagonal mesh topology. The addressing scheme is cumbersome and routing algorithms are complex. Carle *et al.* [35] proposed a simplified addressing scheme and routing algorithms for hexagonal interconnections. The proposed broadcasting algorithm is efficient but the addressing scheme exhibits asymmetry resulting in complicated routing. Fabian Garcia *et al.* [34] proposed a new addressing scheme which provides an efficient routing protocol. Xiao *et al.* [36] developed a 8-neighbor mesh topology and 6-neighbour hex grid topology which provide a much more effective inter-processor interconnect to reduce the application area, power consumption and total communication link lengths. Both topologies increase local connectivity while keeping much of the simplicity of a mesh-based topology. Decayeux *et al.* [37] developed a 3D hexagonal network that can be built with 2D hexagonal meshes. The authors also presented a new addressing scheme and an optimal routing algorithm for a 2D hexagonal network based on the distance formula. This 3D architecture is suitable for cellular mobile computing indoor environment. Saini *et al.* [38] explored the 2D hexagonal mesh and compared with 3D Mesh. The results showed that with the proposed routing algorithm, for small PIR (Packet Injection Rate), the overall performance of 2D hexagonal mesh is better than 3D Mesh. For the increased loads, 3D mesh performs better than the 2D hexagonal meshes. Garcia *et al.* [39] studied higher-dimensional hexagonal networks. Addressing scheme and a simple routing algorithm are developed for the proposed hexagonal network but topological parameters are not explored and compared with the existing baseline topologies. For higher dimensions also, the degree remained

three. However, this is obtained by adopting complicated 3D structures. Stojmenovic *et al.* proposed 2D-Honeycomb Mesh topology [12] and explored topological properties and communication algorithms. Further, honeycomb networks with rhombus and rectangle as the bounding polygons were presented. However, the presented routing algorithm was not deadlock-free and thus was infeasible for the on-chip networks. A.W.Yin *et al.* [40] presented a 3D honeycomb structure by extending the existing 2D topology, stacking them vertically layer by layer. A.W.Yin *et al.* also presented the deadlock-free routing algorithm for 3D honeycomb topology. The 3D structure is simple as the implementation involves straightforward stacking of layers. However, the node degree is increased from three to four. A.Yin *et al.* [41], presented the comparison between Honeycomb Mesh and rectangular Mesh topology and the simulation results showed that honeycomb outperformed mesh in terms of area, power and communication delay. Jean Carle *et al.* [42], described higher dimensional Honeycomb networks, 3-D and k-D where k>3. Further, the topological properties, routing and broadcasting algorithms were presented. The results showed that the cost of k-D honeycomb mesh was slightly better than that of k-D mesh. The above studies suggest that the hexagonal-like structure can be considered as an alternative NoC topology for mesh.

2.1.3 Diagonal Mesh topologies

With an objective to propose an alternative topology that can outperform the widely used Mesh topology, this work has focussed on Diagonal Mesh-based topologies known for their simplicity, scalability, and regularity, similar to Mesh. Moreover, incorporating diagonal links has been found to be an efficient way to enhance the performance of Mesh with a moderate level of complexity. In this section, Diagonal Mesh topologies including DMesh, XDMesh, and ZMesh shown in Figure 2.1 have been reviewed.

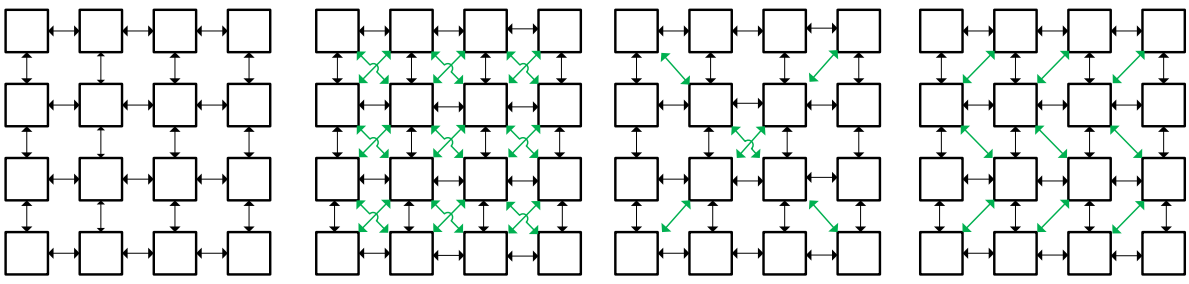


Figure 2.1 (a) Mesh; (b) DMesh; (c) XDMesh; (d) ZMesh

In recent times, several diagonally linked mesh-based topologies have been proposed. Chifeng Wang *et al.* [24] have presented a diagonal mesh topology dubbed DMesh, which, as illustrated in Figure 2.1(b), contains diagonal connections across the network. DMesh has adopted an X-architecture routing approach to minimize latency with moderate implementation cost overhead. The evaluation results have indicated that employing diagonal links has been a more area- and power-efficient strategy for NoCs for improving network performance than using larger buffers.

Md.Hasan Furhad *et al.* [25] have suggested an extended diagonal mesh topology termed XDMesh, as illustrated in Figure 2.1(c), to improve network performance and lower the Mesh NoC's energy consumption. In terms of latency, throughput, energy consumption and area overhead, XDMesh has surpassed other leading-edge topologies such as mesh, extended-butterfly fat tree and diametrical mesh.

N. Prasad *et al.* [27] have proposed an energy-efficient diagonal mesh topology, named ZMesh, shown in Figure 2.1(d). The topological properties of ZMesh such as network diameter, bisection width, and the number of edges have been analysed. Performance comparison of ZMesh with other topologies including Mesh, DMesh, CMesh, PDNoC in terms of latency and power has been performed under different synthetic and real-time traffic patterns but only for injection rates varying from 0.05 to 0.3. The results have shown that ZMesh has performed better than Mesh, PDNoC, CMesh. It has been observed that the ZMesh has the lowest power-latency product (PLP) for injection rates up to 0.3. However, beyond the injection rate of 0.3, PLP of ZMesh has increased drastically when compared with that of DMesh.

2.2 On-chip interconnect technologies

2.2.1 Electrical interconnect technology

Traditionally, electrical also called metallic interconnects have been used as the communication infrastructure for SoCs. Metallic interconnects have the advantages of lower power consumption, higher bandwidth and lesser area for short range communication. The basic block diagram for electrical interconnects is shown in Figure 2.2. Signalling in

this approach is based on charging and discharging of the wire segments to send information.

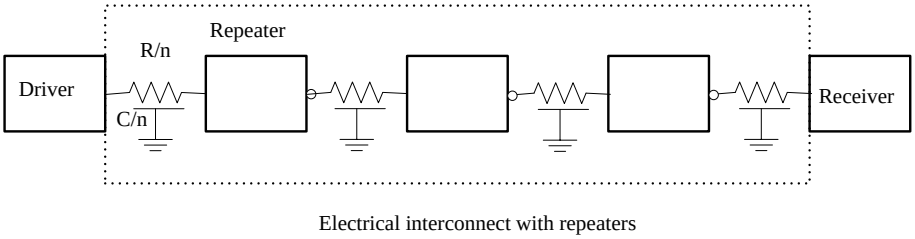


Figure 2.2 Electrical communication system

Today's technological advancements in ICs enable economically feasible production of larger die sizes resulting in the increased average length of the interconnect wires and the associated parasitic resistance, capacitance and inductance effects. Therefore, the role of interconnects has become predominant in today's complex SoCs and a thorough analysis of the interconnect wires has been essential. This is because they introduce parasitic elements - resistance, capacitance and inductance that significantly influence the energy dissipation, propagation delay and reliability of the SoCs. Electrical circuit modeling of the interconnects as a function of its parameters estimates and approximates the real behaviour of the interconnect wires. The behaviour of the circuit at a given node is determined by a few dominant parameters. Based on the length and geometry of the wire, different models exist in the literature. Some of these electrical circuit models of the interconnects include the lumped RC model, distributed RC model, transmission line model [43; 44].

Limitations and advancements

Ever since the development of integrated circuits, Aluminium, Al has been the most prevalent material used for interconnects because of its low resistivity and good processing capability. Later, aluminium-copper alloy has been used for better reliability. Gradually, copper, Cu has emerged as a suitable material for on-chip interconnects because of its lower resistance, higher reliability and conductivity compared to those of Al [45]. Though Cu has the benefit of high conductivity, its high diffusivity leads to the diffusion of copper atoms into the surrounding dielectric that isolates individual Cu interconnect lines

[46; 47]. Copper interconnects require a barrier and liner to prevent the diffusion of Cu through the dielectric. TaN and Ta has been introduced as barrier and liner respectively [48; 49]. Continuous technology scaling leads to steep rise of the resistivity of copper due to increased grain boundary and surface scattering. Additionally, the barrier and liner occupy a large cross-section of the scaled cu interconnect. Thus, with wire scaling, the effective conductive part of cu decreases. The decreasing dimension leads to higher resistivity due to electron scattering and reduced electromigration resistance [50; 51]. Scaling cu interconnect dimensions in line with technological advancements is challenging due to increased resistivity and barrier and liner concerns. This necessitated either investigating ways to extend the usability of cu for reduced wire dimensions or exploring alternate metals with superior conductivity and reliability that replace cu. Various strategies have been investigated to adapt cu to reduced dimensions without compromising the performance and reliability of interconnects. Some of these strategies include thinning barrier or liner or using advanced liner materials [52–55], hybrid metallization schemes [56], alternate deposition strategies [57]. Continued downscaling of cu interconnects or extending cu with the aforestated strategies becomes challenging at nano dimensions due to manufacturing and reliability concerns [58]. This has demanded the exploration of viable alternate materials to replace cu [59–62]. Properties and prospects of several alternate materials that replace cu have been investigated. Promising refractory metals such as Mo, Os, Ir, Ru, and Rh can be considered as possible candidates to replace cu as they have high melting temperatures, relatively low bulk resistivity and smaller inelastic mean free path than Cu. Additionally, they are stable in contact with SiO2 allowing for barrierless metallization [59]. Studies have inferred that though Ruthenium (Ru) and Cobalt (Co) have higher resistivity than that of cu, they can be suggested as a replacement for cu because the interconnects developed with Ru and Co can conduct without barrier layers and thus have better conductance than that of cu interconnects [61; 63; 64]. Sub 10nm technology, possibilities of employing alternate materials like Ru, co to cu as interconnects have been explored. C.-K. Hu *et al.* [60; 65] have investigated the resistivity and electromigration reliability of Cu, Ru, and Co on-chip interconnections. A similar resistivity size effect has been observed in Cu, Ru and Co. However, Non-linered or by using a little liner Co and Ru interconnects can have better interconnect resistance than Cu in future interconnect technologies below 5nm, if Cu liner cannot be scaled down below 2 nm. Below 3nm,

to address the limitations of cu and other metals, GNRs and CNTs have emerged as a potential choice. GNRs and CNTs are detailed in the subsequent sections.

Furthermore, technology scaling enables integration of larger number of components on the chip. The increased number of cores requires multi-hop complex routing which leads to increased network latency. In order to reduce the number of hops and lower the network latency, one or more express links or long range links can be inserted, but this increases the energy consumption as long-range links require repeaters. Consequently, electrical wired interconnects are not capable of meeting high-performance requirements of the complex MPSoC architectures. To address the limitations of electrical wired interconnects and to improve the performance and energy efficiency of complex SoCs, significant research has been going in the field of interconnect technologies.

Richard *et al.* [66] has studied the feasibility of all-electrical near speed-of-light on-chip communication. By exploiting the wave nature of the interconnects, high speed systems can be built. This can be achieved by modulating the digital data with a high frequency carrier. The modulated signal is transmitted through a low loss transmission line like a microstrip transmission line. When compared to the electrical interconnects augmented with repeaters, the modulated signalling approach (Baseband to RF conversion) shows a five times reduction in delay. Consequently on-chip RF interconnects (RF-I) have emerged. Subsequent section elaborates RF-Is. RF-I technology is one among various emerging interconnect technologies, with others including photonic/optical, Plasmonic, wireless, surface wave. These emerging technologies are elaborated in the following sections.

2.2.2 RF interconnect (RF-I) technology

RF-I approach is based on transmission of waves rather than current or voltage signalling. In electrical wired interconnects, the entire length of the wire has to be charged or discharged to signify '1' or '0'. In RF-I approach, the transmission can be achieved by modulating and transmitting the EM wave either through a guided medium or free space. The guided medium is a transmission line (TL) on chip. Miniatured on-chip antennas are used for transmitting/receiving the modulated data in RF-wireless interconnects. Data is modulated onto a high frequency carrier either by amplitude/phase modulation. As the

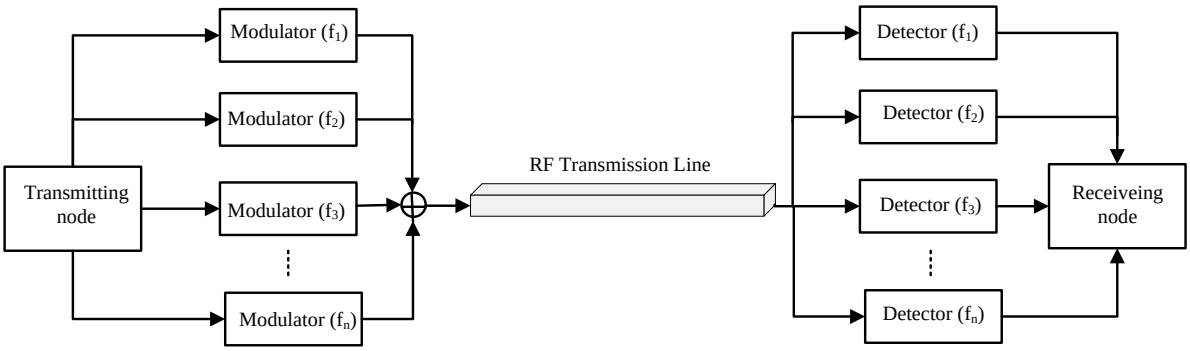


Figure 2.3 Schematic of TL based RF-I [4]

propagation of EM wave is at the effective speed of light in the dielectric layers of the chip, minimum latency can be achieved. Various transmission line structures like microstrip transmission line (MTL), co-planar waveguide (CPW), differential TL are feasible structures for implementing RF-Is. Further, employing Frequency Division Multiple Access (FDMA) for RF-I approach results in higher data rates and also enables concurrent communication using different frequency bands, transmitting through the shared transmission line [4; 67–70]. The basic schematic of transmission line based RF-I is shown in the Figure 2.3. Frequency multiplexing is considered in this schematic.

With the introduction of RF-Is, the physical transmission medium is no longer confined to a direct-coupled metallic wire. Instead, the transmission can be established either wired (transmission lines) or wireless (using on-chip integrated antennas) medium with the support of capacitor couplers. These couplers reduce baseband noise and direct current (dc) power consumption. Further, conventional wired on-chip interconnects rely solely on time division multiplexing (TDMA). This TDMA-interconnect (TDMA-I) system is constrained by fixed and nonreconfigurable architecture that results in high propagation delay and an inability to support simultaneous transmissions on the same physical medium. In contrast, RF interconnects enable the use of combination of different multiple access schemes like code division multiple access (CDMA) & frequency division multiple access (FDMA) schemes. This facilitates high data rates and simultaneous transmissions as well as reduction in latency and power consumption [68].

Limitations and advancements

Scaling beyond 1000 cores on-chip architectures is challenging for transmission line based RF-I. This is because of the large hop counts, long interconnects, complex layout requirements of complex SoCs. On-chip TLs are not effective at very high frequencies. Also, for long TLs, inter-channel interference between adjacent TLs may pose a serious problem. On-chip wireless interconnects have been investigated to be a promising solution to enable long-distance communication in larger NoC architectures. Wireless interconnects facilitate smaller hop counts between distant nodes and also they avoid long wires and high wiring complexity. RF/microwave wireless interconnects can substitute TL based RF-Is for long distance communication in complex kilo-core SoC architectures [67; 71].

2.2.3 Graphene-based interconnect technology

Copper is the most widely used material for realising traditional on-chip electrical interconnects. With the technology scaling, copper wires resistivity is rising steeply, thereby impacting the performance and reliability in terms of circuit delay, temperature and current-carrying capacity. Carbon nanomaterials with their attractive properties have been found to potentially replace copper for interconnects and passive devices in ICs, including vias and through-silicon vias (TSVs). Hong Li *et al.* [72] have discussed the prospects of various carbon nano materials as interconnects and passive devices. The authors have also provided a comparative analysis of these materials as interconnects with that of optical and RF interconnects.

Graphene, a carbon allotrope, is a flat monolayer of carbon atoms tightly packed into a 2D honeycomb lattice. It is the basic building block of the two potential structures - Carbon Nano Tubes (CNT) and Graphene Nano Ribbons (GNR) that have gained focus in the research of on-chip interconnects. CNTs and GNRs have extraordinary physical properties because of their unique structure. Also, the excellent electrical, thermal and mechanical properties of GNRs and CNTs have made them promising candidates for the next generation interconnects. A CNT can be thought as a rolled-up graphene sheet in which the edges of the sheet are joined together to form a seamless tube while a GNR is obtained by patterning graphene to a strip [72].

Carbon Nano Tubes (CNTs)

Carbon nano tubes (CNTs), because of their extraordinary electrical, thermal and mechanical properties find application as VLSI interconnects [73; 74]. CNTs can exist in different configurations like single-walled CNT (SWCNT), multi-walled CNT (MWCNT) and mixed CNT bundle (MCB) i.e., a combination of SWCNT & MWCNT. SWCNT comprises of single rolled graphene sheet where as MWCNT comprises concentric shells of graphene. SWCNTs and MWCNTs can be realised as isolated structures or bundled structures to enhance performance. SWCNTs and MWCNTs have nearly same current carrying capabilities. However, fabrication of MWCNTs is easier compared to that of SWCNTs. Significant research has been done in the study of CNTs including the RLC modelling of CNTs, performance analysis of CNTs over Cu interconnects at circuit level and system level, feasibility study of CNTs as global interconnects and comparison with cu global interconnects. Circuit models have been developed for metallic SWCNTs [75], MWCNTs [76], mixed MCBs [77] and analysed the impact of SWCNTs on the performance of VLSI designs. Pasricha *et al.* [78] have investigated the system level performance of SWCNT bundles and MCBs. Incorporated the RLC models into a system-level environment and analysed the overall performance of various CMP applications. The results indicated 1.3x and 1.5x performance gains with SWCNT bundles and MCBs over conventional cu interconnects respectively. Suraj Subash *et al.* [77] have explored the electrical parameters including resistance, capacitance, inductance of SWCNT, SWCNT bundles, mixed bundles of SWCNTs & MWCNTs. The findings have indicated that for interconnect applications minimising the lengths is advantageous. Additionally, placing only MWCNTs on the outer periphery and a metallic SWCNT in the centre of the bundle is recommended. Moreover, to have higher proportion of metallic CNTs in the bundle is always favourable.

Graphene Nano Ribbons (GNRs)

GNRs are planar, one-dimensional narrow strips of graphene with a width less than 50nm. Significant research has been going in exploring the utilisation of GNRs as interconnects [79–85]. Based on the shape and orientation, GNRs are classified as armchair GNRs (ac-GNRs) or zigzag GNRs (zz-GNRs). Armchair GNRs are either metallic or semicon-

ducting whereas zigzag GNRs are always metallic in nature. GNRs can be developed as single layered GNR (SLGNR) or multi-layered GNR (MLGNRs). The resistivity of MLGNRs is lower than that of SLGNRs. MLGNRs conductivity is limited by edge scattering for sub 20nm. To address these limitations and improve the conductivity of MLGNRS, intercalation-doped ML-GNR interconnects have been proposed [86]. Naeemi *et al.* [87] have studied the conductance of SLGNR, and proposed a compact physics based circuit model for GNRs [88]. Electrical characteristics of SLGNR at radio frequency have been analysed in [89]. An accurate impedance modeling methodology to analyse the high frequency behaviour of GNR interconnects is presented in [90]. In [91], complex phenomena occurring at high frequencies in GNRs, such as anomalous skin effect (ASE), highfrequency resistance and inductance saturation, intercoupled relation between edge specularity and ASE, and the impact of the linear dimensions on impedance have been investigated. It has been observed that SLGNRs have lower resistivity than cu and CNT bundles. However, the resistance of SLGNR is too high to be viable. Henceforth, MLGNRs have been proposed. Even though, MLGNRs have fabrication benefits over CNTs, for MLGNRs to match their performance with CNTs, proper intercalation doping is required [92]. Intercalation is the process of inserting metal atoms or compounds into pristine MLGNR layers. Furthermore, MLGNRs are classified into top contact MLGNRs (TC-MLGNR) and side contact MLGNR (SC-MLGNR) depending on their connection with surrounding devices and interconnects. In TC-MLGNRs, only the top layer is connected to the surrounding contacts where as in SC-MLGNRs, all the layers are connected to the surrounding contacts resulting in reduction of the resistance. Compared to TC-MLGNRs, SC-MLGNRs offer reduced resistance. Conversely, fabrication of SC-MLGNRs is challenging while that of TC-MLGNRs is comparatively easier [93]. Xu *et al.* [86] have presented a comprehensive conductance and delay analysis of SLGNR, MLGNR and intercalation doped MLGNR interconnects. Cui *et al.* [94] have examined the signal transmission characteristics of MLGNR interconnects, considering both capacitive and inductive couplings between adjacent GNR layers. Nasiri *et al.* [95] have conducted an analysis on the stability of GNR interconnects and have inferred that increasing the length and width enhances the stability of MLGNRs. The analysis of MLGNR interconnects in [86; 94; 95] has been carried out by neglecting interlayer resistance based on the consideration that the effective scattering resistance decreases with the increase in the number of layers. Conversely, it is

noted that the increase in the number of layers in MLGNR does not necessarily result in the decrease of resistance of MLGNR interconnects. Instead, the increase in the effective resistance saturates as the number of layers increases [96]. Hence, in an alternative study [96], interlayer resistance has also been considered along with the capacitive and inductive couplings between adjacent MLGNR layers for modelling MLGNR interconnects. It has been noted that the impact of interlayer resistance is minimal on the performance of the MLGNR interconnects as the interconnect length increases. Furthermore, MLGNR with interlayer resistance exhibits superior performance compared to cu interconnects for longer interconnects.

In the recent years, numerous studies on GNR interconnects have emerged. Subhajit Das *et al.* [97] performed thermal stability analysis of TC-MLGNR, SC-MLGNR and cu interconnects for different interconnect lengths under different chip operating temperatures. It has been inferred that SC-MLGNRs exhibits superior performance in terms of delay and power compared with that of TC-MLGNR and cu interconnects in terahertz frequency range. SC-MLGNR could be a possible material for high frequency applications such as nano-antenna, wireless communication. An alternate study [93], indicates that, nearly specular global-length intercalation doped MLGNR interconnects are more promising in terms of reduced propagation delay, power dissipation and Power-Delay-Product when compared with MCB and Cu interconnects.

As mentioned previously, intercalation doping enhances the performance of MLGNRs. FeCl3 intercalation doped MLGNRs with an optimum number of layers demonstrated lower resistivity than that of cu interconnects of same dimensions for sub 10nm widths [98]. FeCl3 intercalation doped MLGNRs have exhibited a superior performance and reliability with respect to cu interconnects for sub 20nm widths [99]. The temperature-dependent performance analysis of coupled TC-MLGNR interconnects has been carried out with different intercalation doping materials including AsF5, FeCl3, and Li for 13.4 nm technology nodes at the global level. This investigation has considered edge roughness and crosstalk effect. The results have inferred that Li-doped TC-MLGNR when compared with SC-MLGNR, MWCNT, and copper (Cu) interconnects could be a suitable material for global interconnect applications in a thermally variable environment [100]. A comprehensive analysis of p (Phosphorus) doped, N (Nitrogen), B (Boron) and Al (Aluminium)

doped zigzag GNRs has been conducted. Electronic, transport properties and stability of the interconnects have been analysed wherein N doped zigzag GNRs have lower delays w.r.t other considered p doped GNR, Al doped, B doped zz-GNRs [101]. Another approach to enhance the stability and performance of MLGNRs is dielectric insertion. SC-MLGNRs can exhibit superior performance because of their higher conduction channels. However, due to interlayer electron hoping that reduces MFP, SC-MLGNRs transform into graphite. This limits the number of vertically stacked layers in SC-MLGNRs. To address the limitations of SC-MLGNRs and improve the MFP and higher conductivity per layer, graphene double layer (GDLS) structure has been proposed. In GDLS, a thin layer of dielectric is inserted between two GNR layers. This structure improves carrier mobility, MFP, lowers the scattering rate of electrons in each layer [102]. Analytical models for analysing scattering rate , MFP and carrier mobility in dielectric inserted SC-MLGNRs have been proposed [103–105]. In an alternate study, electronic, transport, crosstalk and stability properties of MLGNRs have been investigated and inferred that the insertion of dielectric (hafnium oxide (HfO2)) between GNR layers, in conjunction with silicon dioxide (SiO2) yields best performance characteristics when compared with cu, undoped, intercalation-doped MLGNR interconnects [106].

2.2.3.1 Limitations and advancements

In spite of the swift advancements in the utilization of Carbon Nanotubes (CNTs) and Graphene Nanoribbons (GNRs) as on-chip interconnects, the integration of these graphene-based materials remains a challenge for current nanoscale technology. By Diameter Controlling the alignment of CNTs within a bundle, as well as the edge irregularity of GNRs, significantly complicates the fabrication and growth processes of graphene. Cu–graphene hybrid interconnects address the limitations of both copper and graphene-based interconnects while exploiting the superior performance of each, are regarded as a promising technology for future interconnect solutions [107–111].

2.2.4 Photonic (Optical) Interconnect Technology

Fibre optics has replaced electrical wires for long distance communication since long time. With the advancements in nano photonics, now optics has expanded its horizon on

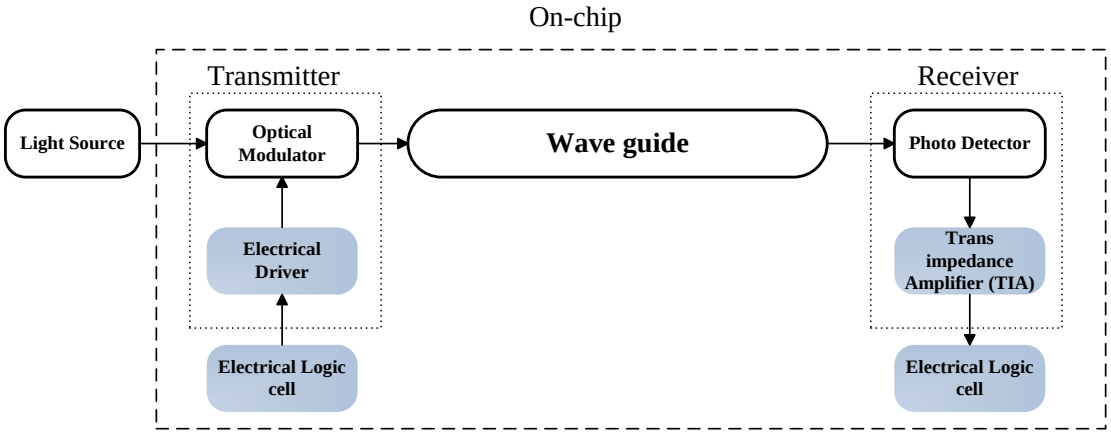


Figure 2.4 Schematic of Photonic/Optical interconnect technology [5]

to chip level short distance communication. Recent studies have found that on-chip optical interconnects outperform traditional electrical interconnects in terms of Bandwidth, Latency and Power efficiency. However, optical interconnects have certain challenges and limitations.

2.2.4.1 Basic concepts

Photonic i.e., Optical communication system has four key components – Light (optical) Source, Optical Modulator, Photo detector, Wave guides. Basic communication system of photonic interconnect technology is shown in Figure 2.4. To facilitate the data transmission through optical interconnects, the electrical data needs to be converted to optical signals. A light source generates optical signal which is directed to an optical modulator. The optical modulator converts the electrical signals from an electrical driver circuit into a modulated optical signal. Subsequently, the modulated optical signals are transmitted to the receiver node via waveguide. A photo detector at the receiving node detects the optical signal, amplifies it and converts it back to electrical signal which replicates the transmitted data. The light source can be located off-chip or on-chip. In case of off-chip light source, the signal needs to be guided onto the chip by means of couplers.

Significant research has been undertaken to explore the CMOS compatible optical components such as light sources [112–116], optical modulators [117–119], waveguides

[120–122], photo detectors [123–125].

To overcome the data rate and bandwidth limitations of electrical network-on-chip architectures (ENoC), optical interconnect technology has been emerged. Optical NoC (ONoC) offers various advantages over ENoCs. Optical interconnects offer high bandwidths in the THz range, high speed of propagation, reduced electromagnetic interference. Additionally, optical signalling results in low power consumption, since it is independent of the distance traversed by the optical signal. This proves beneficial in long distance communication at chip level. All-optical NoCs or fully optical NoCs (ONoC) use only optical links to transmit data among all nodes, with no electrical wires involved in data transmission. However, ONoCs also face challenges like integration of optical components and temperature susceptibility of optical interconnects. Each technology has its unique merits and demerits. To optimize the advantages of these technologies, hybrid architectures incorporating electrical interconnects for short distance communication and optical interconnects for long distance communication have been explored. Based on the communication requirements, availability of technology and workload patterns, either fully optical architectures or hybrid optical architectures have to be chosen.

The key challenge in designing optical interconnects is to deliver higher performance than electrical interconnects while keeping static power overheads and optical resources requirements low. The researchers have proposed various ONoC designs, broadly classified as Crossbar based designs, Multistage designs and opto-electrical (hybrid interconnect) architectures.

Crossbar based designs: A crossbar is a shared bus that nodes can use to read and write messages. ONoCs can achieve Collision-free, all-to-all communication by employing an optical crossbar. Fully connected crossbars architectures have poor scalability. Channel sharing crossbar architectures - Single-Write-Multiple-Read (SWMR) or Multiple-Write-Single-Read (MWSR), Multiple-Write-Single-Read (MWSR), Multiple-Write-Multiple-Read (MWMR) with various arbitration mechanisms for coordinating shared channels have been explored. Vantrease *et al.* [126] proposed Corona, a MRSW optical buses crossbar architecture with token-ring arbitration mechanism. Pan *et al.* [127] proposed FlexiShare, a multiple-writer-multiple-reader (MWMR) optical crossbar, channel sharing architecture to reduce over-provisioning of channels. The authors introduced a novel pho-

tonic token-stream mechanism for channel arbitration and credit distribution to improve link utilisation.

Multistage designs: Multistage designs for on-chip optical networks are more scalable and efficient. They bring tradeoffs between power, latency, and complexity by dividing larger network into multiple smaller subnetworks. Joshi *et al.* [128] proposed photonic clos network, a multi stage architecture where all the Clos routers are implemented electrically and the inter-router channels utilise photonics. The photonic clos architecture consumes less power compared to global photonic crossbar architecture. Morris *et al.* [129] proposed a two-hop multi level (multi stage) architecture termed ET-PROPEL (Extended-Token based Photonic Reconfigurable On-Chip Power and Area-Efficient Links), where in the first level, grid rows (x-dir) are subnets fully connected with a photonic crossbar and in the second level, different rows (ydir) are connected by a token-ring arbitrated shared photonic link. The first level and second level are combinely called as T-PROPEL. In the third level, four T-PROPELs are combined using fatree topology to form ET-PROPEL by using arbitration free photonic crossbars. ET-PROPEL significantly reduce power and improve performance over electrical networks. Cheng Li *et al.* [130] proposed LumiNoC, a photonic multistage NoC in which utilises an electrical router to interconnect multiple optical subnets. LumiNoC employs purely photonic, dynamic channel scheduling, distributed arbitration mechanism to achieve low latency without degrading throughput.

Opto-electrical (hybrid) designs: Opto-electrical hybrid networks can be categorised into three different categories. In the first category, the control and data planes are separated. As implementing standard features such as buffering, routing, and processing headers is very difficult in optical networks, ENoC is employed as a control network overseeing the data transmission process where as the payload data is transmitted through optical network. For example, Shacham *et al.* [131] proposed an optical, circuit-switched, mesh-based network controlled by electrical control packets. In this, every photonic message transmission is preceded by an electronic control packet termed as a path-setup packet. The path-setup packet is routed on the electronic network, to acquire and set up a photonic path for the message. Once the path is established, photonic messages are transmitted without buffering. The second category of hybrid designs is clustered based designs in which the network is partitioned into two parts - electrical and optical, both are

used to transmit data. ENoC is utilised for short distance communication whereas ONoC is used for long distance communication. The entire network is divided into multiple smaller subnetworks called clusters. Intra-cluster communication is dealt by ENoC where as Inter-cluster communication is handled by ONoC. Pan *et al.* [132] proposed Firefly architecture which employs a hierarchical crossbar NoC topology with clusters of nodes connected through local electrical networks, and optical links overlaid for inter-cluster communication. Meteor, proposed by Shirish *et al.* [133], is a cluster based hybrid architecture that combines configurable concentric photonic ring waveguides with a traditional 2D electrical Mesh. HOME, proposed by Kwai *et al.* [134] is a similar opto-electrical hybrid architecture that employs optical waveguides for long distance communication and electrical interconnects for short distance communication. To improve network throughput and reduce latency, HOME proposes a new set of protocols. The third category of hybrid designs is layered based architectures - 3D multi layered architectures employing both optical and electrical interconnects. In 3D opto-electrical hybrid architectures, multiple active layers exist and single or multiple photonic layers can be utilised. OPAL, proposed by Pasricha *et al.* [135], employs multiple active layers and multiple photonic layers. Local communication is handled by electrical Mesh whereas inter-layer and intra layer global communication is handled by photonic links. Similarly, PHENIC, proposed by Ahmed *et al.* [136], is a 3D layered opto-electrical hybrid architecture which utilises optical layer for high bandwidth transfer and an electrical control layer for path control.

2.2.4.2 Limitations and advancements

Despite the advantages of optical links such as distance-independent power consumption, low-latency, and high bandwidth compared to electrical links, they have certain limitations, namely, the size and sensitivity of Si-based optical devices. Optical components are constrained by diffraction, setting a lower bound on the dimensions of the devices with which light can be confined. Optical devices with dimensions on the scale of the signal wavelength have large device footprints compared to CMOS electronic devices. The size mismatch between micrometer-scale photonic and nanometer-scale electronic components hinders large-scale electro-optic integration on a single chip. Plasmonic interconnects address the limitations of photonic interconnects. Surface plasmon polari-

tons (SPP) have interesting properties that addresses these limitations. A plasmon is a quasi-particle formed by hybrid electron–photon oscillations occurring at the interface between a conductor and a dielectric. Plasmons resemble light waves confined to the surface of a metal maintaining the energy and bandwidth benefits of photonic interconnects. Plasmonic devices have smaller footprints. However, spp waves are more susceptible to damping due to high ohmic losses, lack wave length division multiplexing (WDM) support and are suitable only for short distances [137]. Therefore, the authors Wassel *et al.* [137] proposed a hybrid photonic/plasmonic channel where in plasmonics is utilised to realise optical components and photonic for waveguiding. This hybrid channel provides moderate bandwidth while maintaining the distance independent power consumption. This channel exhibits a high degree of temperature and process variation tolerance. Several studies have been conducted on hybrid opto-electric NoCs incorporating hybrid photonic-plasmonic interconnects and on plasmonic devices to investigate the applicability of the emerging plasmonics in NoCs [137–143].

Graphene with its high carrier mobility and broadband optical absorption, emerges to be a promising material to realise high speed, broadband optoelectronic devices compatible with Si [144–147].

2.2.5 Wireless Interconnect Technology

Wireless interconnect technology is the most promising technology for scalable, complex on-chip interconnect architectures. Thanks to advancements in nanotechnology, on-chip wireless communication has become a viable option. In Wireless NoCs (WiNoC), the wired links of ENoCs are replaced with wireless links. In wireless communication, electrical signals are converted to EM waves (RF/microwave) and transmitted through free space using miniatured antennas integrated on the chip. The block diagram of wireless interconnect is shown in the Figure 2.5 . Wireless interconnects offer superior data rates and bandwidths, enable single-hop communication between distant nodes resulting in reduced latency and power consumption, contribute to less complex layouts as compared to conventional wired electrical interconnects.

On-chip wireless communication can be either purely wireless or a combination of wired-wireless links i.e., hybrid wired-wireless communication. Although wireless links

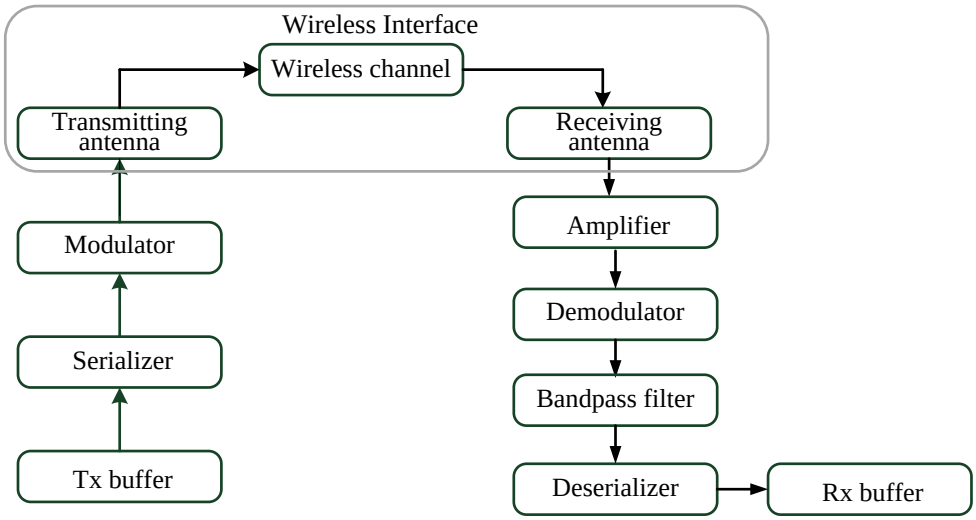


Figure 2.5 On-Chip wireless communication system

offer the benefits of low latency over long distances, they have the drawbacks of area and power overheads attributable to the transceiver circuitry. In contrast, a wired NoC has the advantage of smaller area compared to WiNoC. Considering the merits and demerits of each of these technologies, the strategic approach of combining both wired and wireless to create a hybrid wired-wireless architecture proves effective in optimising the power, area and delay. Research on WiNoCs is twofold – one is the wireless medium and the other is the architecture. With regard to medium, on-chip wireless transmission can be possible with millimeter wave (MMW) antennas [148], Radio-frequency interconnects (RF-I) [149], Carbon Nano Tubes (CNTs) [150], Graphene [151–155]. With regard to architecture, numerous studies on purely wireless [154; 156; 157] and hybrid wired-wireless NoC [158–162] architectures are available in the existing literature.

Considerable research on on-chip antennas [157; 163–167], channel modelling [168–170], MAC mechanisms [171–176] has been going on to make realisation of WiNoCs feasible in practical many-core chip architectures.

2.2.5.1 Limitations and advancements

Wireless communication technology is now advancing into new realm. Until now, MM-wave [148], UWB [156] on-chip wireless communication or Graphene enabled wire-

less network-on-chip (GWNoC) [177] have been investigated. Recent advancements in nanophotonics has led to the emergence of optical antennas facilitating Free space optics (FSO) i.e., optical wireless (OW) communication [178–182]. Light waves propagating in free space transmits data by means of nano optical antennas. Optical communication provides high data rate, high bandwidth, reduces propagation loss and delay. However, increasing the number of cores on the chip increases the complexity of switching, routing, layout, power losses due to wave guide crossings. Thus, implementation of optical NoCs for large size sophisticated NoCs becomes challenging. On the other hand, wireless communication provides an efficient solution to replace long distance wired communication, decreasing the layout complexity and simplifying routing. However, RF wireless communication fall short in providing high bandwidths comparable to optical NoCs. Therefore, the synergistic integration of both optical and wireless technologies leverages the advantages of each, emerging as a viable solution for the future exascale complex NoCs.

2.2.6 Surface wave Interconnect (SWI) Technology

Surface wave interconnects, a nascent type of interconnections comparable to RFIs and OIs, provide a communication medium and serve as a substitute to metal interconnects. The Zenneck surface wave is a 2D electromagnetic (EM) wave with non uniform characteristics that depends on a supporting surface. This surface, designed as a waveguide, confines the EM wave within a two-dimensional medium rather than enabling it to propagate in three-dimensional free space [183]. The supporting surface medium can be either a dielectric coated conductor layer or a corrugated conductor surface. RFIs can be implemented as either waveguide transmission line RFIs or free space wireless RFIs. RF wireless eliminate the need for a physical transmission medium and possesses inherent broadcasting capabilities. RF waveguide transmission lines exhibit low power dissipation compared to RF wireless. The Zenneck surface wave amalgamates the benefits of these two types enabling reduced power dissipation for significantly larger coverage areas than wireless RF technology [184]. A few studies exist on SWIs for on-chip communication, particularly on multicast patterns [183; 185–187].

2.2.7 Comparison of on-chip interconnect technologies

In Table 2.1, a brief summary of the comparison of key features of on-chip interconnect technologies is presented [188].

Table 2.1 Summary of comparison of key features of on-chip interconnect technologies

Feature	Technology	Summary
Transmission method	Electrical	Electron transport through metal conductor
	RFI(TL)	EM waves
	OI	Light waves
	Wireless	EM waves
	SWI	Surface waves
Transmission medium	Electrical	Metal wire (or) CNT (or) GNR
	RFI(TL)	Transmission line (or) Waveguide
	OI	Optical waveguide
	Wireless	wireless
	SWI	supporting surface designed as a waveguide
Design requirements	Electrical	Metal conductors, Repeaters for long range links
	RFI(TL)	Transceiver circuitry, transmission lines or waveguides
	OI	On-chip photonic components (optical source, modulators, photodetectors, waveguides)
	Wireless	On-chip antennas (metal or graphene based), transceiver circuitry
	SWI	Guiding surfaces
Bit-rate transparency	Electrical	Electrical routers switch with every bit of the transmitted data, leading to a dynamic power dissipation that scales with the bit rate
	RFI(TL)	Energy dissipation is essentially independent from the bit rate.
	OI	Energy dissipation is essentially independent from the bit rate.
	Wireless	Energy dissipation is essentially independent from the bit rate.
	SWI	Energy dissipation is essentially independent from the bit rate.
Signal Decay	Electrical	High (increases exponentially without repeaters).
	RFI(TL)	Low signal decay and dissipation.
	OI	Very low signal decay and dissipation
	Wireless	High (inversely proportional to distance).
	SWI	Low signal decay and dissipation(inversely proportional to square root of distance).
Reliability	Electrical	Possible cross-talk exists.
	RFI(TL)	Cross-talk exist (capacitor and inductor coupling).
	OI	High signal integrity.
	Wireless	Noise coupling to the antenna and possibility of multi-path interference.
	SWI	Less subject to noise coupling.
Fanout	Electrical	Needs extra power for multidrop bus (stubs) and lowers propagation velocity.

	RFI(TL)	Stubs cause impedance discontinuity, which will lead to signal reflection.
	OI	Require optical splitters and combiners that decay the optical signal (3dB per splitter).
	Wireless	Limited by transmission signal propagation cover area only
	SWI	Limited by transmission signal propagation cover area only.
Bandwidth	Electrical	Limited
	RFI(TL)	100 to 200 Gbps.
	OI	a few Tbps
	Wireless	tens to hundreds of Gbps
	SWI	100 to 200 Gbps.
Power consumption	Electrical	High. Dynamic power that is proportional to the wire capacitance and voltage.
	RFI(TL)	Relatively tolerable.
	OI	High static power dissipation.
	Wireless	High free space power dissipation.
	SWI	Relatively tolerable.

This section has provided an overview of state-of-the-art and emerging on-chip interconnect technologies, highlighting their limitations and recent advancements. Each of these technologies possesses its own merits and demerits. Consequently, hybrid NoC architectures integrating two or more technologies to leverage the strengths of each, could be a viable alternative to traditional copper based on-chip interconnects.

Chapter 3

Hexagonal Star: Hybrid topology for NoC

3.1 Introduction

Over the past few years, many NoC topologies like Mesh, CMesh, Torus, Spidergon, Octagon, Tree, Mesh of Tree, Honeycomb, Ring, Hexagonal etc., have been proposed [10–14]. Each topology has its own merits and demerits. Mesh and Torus topologies are the widely used topologies because of their simple, regular, symmetrical and scalable structure. However, as the network scales, these topologies suffer from degraded performance. Interconnect wire delays play a predominant role in deciding network performance. The possible ways to reduce the network latency is to minimize the number of wires i.e., links in the architecture, and to reduce the wire length of the interconnects. Many of the existing topologies are based on rectilinear structure i.e., square or rectangular structures, some of the topologies are based on circle or tree structure and very few on hexagonal structures. Moreover, traditionally, topologies follow the Manhattan layout for realizing the interconnects i.e., the interconnect wires are laid in vertical or horizontal directions only. With the lithographic advancements and introduction of X-architecture layout, it is possible to layout the interconnect wires in any arbitrary orientation like 45°, 135°. Realizing the wires with 45° or 135° orientation, the effective wire length can be reduced up to $1/\sqrt{2}$ times the length of wire with vertical or horizontal direction. So, topologies with wires oriented at 45°, 135° and combined with horizontal or vertical wires show considerable reduction in the interconnect wire length over the topologies with only vertical and horizontal wires [189; 190].

From the previous studies on hexagonal die and non-manhattan architecture lay-

outs [189–195], it is evident that to fully utilize the benefits of non-manhattan layouts, the architectures are to be implemented on hexagonal or octogonal ICs with maximal chip area usage. Further, earlier studies on processor tile shapes [36] show that rectangles or square shapes pack with four neighbors, whereas circle and hexagonal tile shapes pack six neighbours. Circle-shaped tiles pack together with wasted space between tiles, whereas the hexagonal-shaped tiles pack efficiently without gaps between tiles retaining the 6-nearest-neighbour property. Literature shows that hexagonal topologies have been less explored. This forms the basis of motivation to explore hexagonal-based topologies. In this attempt to explore the hexagonal-based topologies, a novel hybrid Hexagonal Star (HS) topology has been proposed. In this chapter, the static performance of the topology has been assessed by comparing its parameters with those of baseline topologies. Further, other important aspects such as scalability, routing, and dynamic performance are also administered and a preliminary analysis has been performed. This evaluation and analysis form a basis for further investigation of the proposed topology & suggested scalable architectures as well as the exploration of alternative Hexagonal topologies.

3.2 Hybrid Hexagonal Star (HS)

3.2.1 Proposed HS Topology

Interconnecting outer nodes as hexagons and the inner nodes as stars form the basic block of the proposed topology. The minimum number of nodes that can be interconnected in this fashion is 18 nodes. Figure 3.1 (a) shows the basic block of the Hexagonal Star (HS) topology. Figure 3.1(b) shows the HS topology extended to interconnect 32 nodes. This topology is scalable and can be extended in a layered manner to interconnect a large number of nodes, as shown in Figure 3.2. Each processing element (PE) i.e., a core in the system is connected to a router. The routers are interconnected in the network according to the topology through which all the PEs communicate. Figure 3.1 and Figure 3.2 shows only routers.

The size of the topology can be specified in terms of the number of nodes or the number of hexagonal stars or the dimension of the network. The dimension of the network can be defined as the number of layers of hexagonal stars. Dimension and layer can be

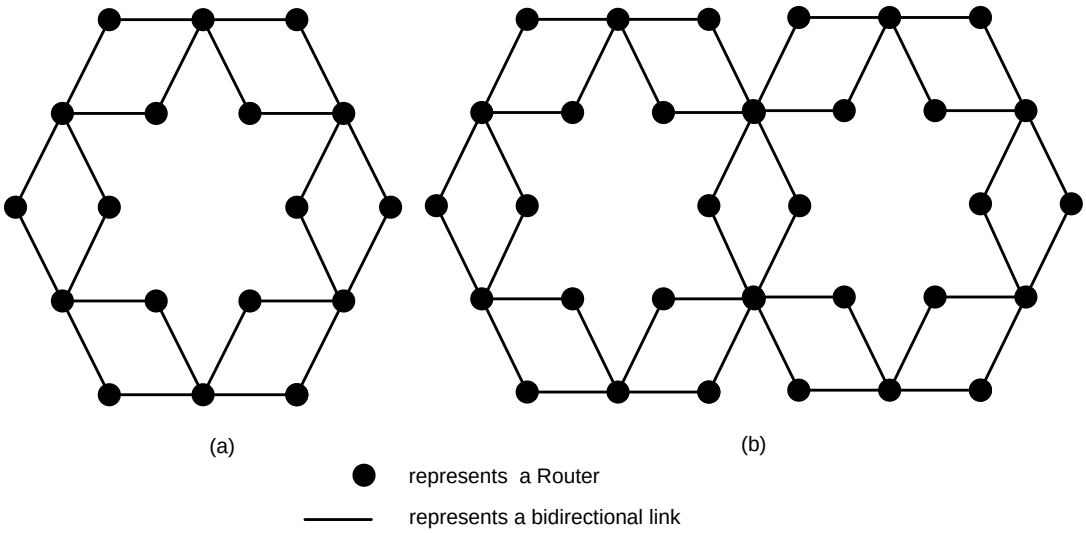


Figure 3.1 (a) Basic block of Hexagonal star(18 nodes); (b) Two hexagonal stars(32 nodes)

used interchangeably. One-dimensional topology means one hexagonal star structure, as shown in Figure 3.1 (a). It is denoted as HS_1. The second dimension, HS_2, can be obtained by adding six hexagonal stars around the six faces of the hexagonal star, HS_1, with some shared nodes as shown in Figure 3.2. The next dimension can be obtained by adding a layer of hexagonal stars to the existing system i.e., HS_d is obtained from HS_{d-1} by adding a layer of hexagonal stars.

The number of hexagonal stars to be added to the existing dimension HS_{d-1} to obtain the next higher dimension HS_d is given by Eq.(3.1)

$$h_d = 6(d-1) \quad \text{for} \quad d >= 2 \tag{3.1}$$

here d is the dimension of the network.
For example if d=2, six Hexagonal stars need to be added to HS_1 to obtain HS_2 and for d=3, twelve Hexagonal stars need to be added to HS_2 to obtain HS_3

The total number of hexagonal stars in the network of dimension 'd' is given by Eq.(3.2)

$$H_d = 1 + \sum_{t=1}^{d} 6(t-1) \tag{3.2}$$

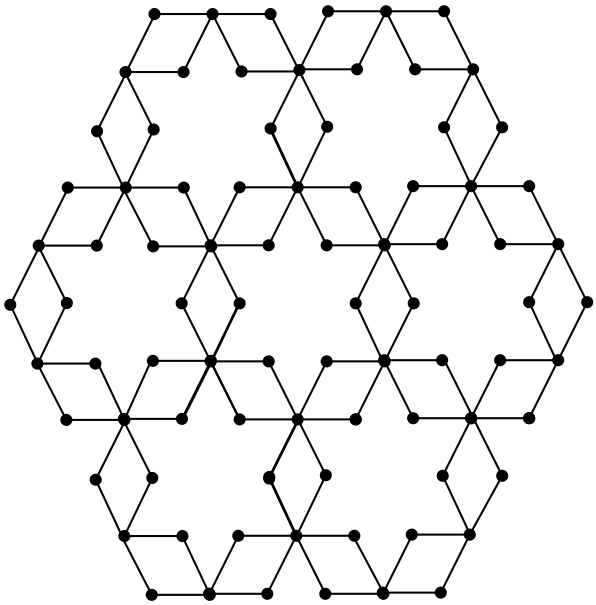


Figure 3.2 Two-dimensional Hexagonal star topology for 84 nodes

The total number of nodes in the network of dimension 'd' is given by Eq.(3.3)

$$N = 24d^2 - 6d \tag{3.3}$$

The Eq.(3.4) is used to calculate the number of nodes (N) for a given number of hexagonal stars (H)

$$N = -0.0823H^2 + 11.551H + 8.5573 \tag{3.4}$$

The number of nodes in the network for a given number of hexagonal stars in a layer is given in Table 3.1.

Table 3.1 Number of nodes for a given number of Hexagonal Stars

Dimension (d)	Number of the hexagonal stars in the layer or dimension (h)	Total number of Hexagonal Stars in the network(H)	Total number of Nodes in the network(N)
1	1	1	18
2	1	2	32
	2	3	43
	3	4	54
	4	5	65
	5	6	76
	6	7	84
3	1	8	98
	2	9	106
	3	10	114
	4	11	125
	5	12	136
	6	13	144
	7	14	152
	8	15	163
	9	16	174
	10	17	182
	11	18	190
	12	19	198

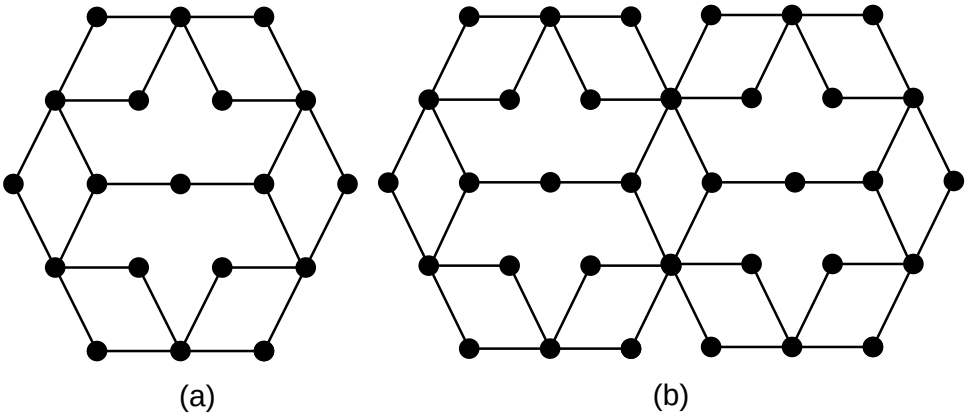


Figure 3.3 (a) Block of Hexagonal star V1 (HS-V1) (19 nodes); (b) Two Hexagonal Stars (HS-V1) (34 nodes)

For meeting the other network sizes, the proposed Hexagonal Star is slightly modified and named as Hexagonal Star-V1 (HS-V1). In the proposed HS-V1, an additional node (Processing element along with its connected router) is placed at the centre of HS to increase the size of the network by one. With this, the architecture can be implemented with 19 (minimum) nodes and can be extended to 34 nodes as shown in Figures 3.3 (a) and (b) respectively. This topology is scalable and can be extended in a layered manner

to interconnect a large number of nodes similar to HS topology.

With HS and HS-V1 topologies, odd-numbered network-sized architectures like 19, 43, and 91 can also be implemented as per the application requirements by retaining the symmetry and regularity in the structure, unlike other baseline topologies like Mesh and Torus.

The total number of nodes in HS-V1 topology with a network of dimension 'd' is given by Eq.(3.5)

$$N = 27d^2 - 9d + 1 \tag{3.5}$$

The Eq.(3.6) calculates the number of nodes (N) in HS-V1 topology for a given number of hexagonal stars (H)

$$N = -0.0823H^2 + 12.551H + 8.5573 \tag{3.6}$$

A heuristic approach has been considered to construct the topology. Following are the steps that detail the construction of the topology.

Step 1: Consider a regular hexagon (Figure 3.4 (a)) and a hexagram, also called Star of David (Figure 3.4 (b))

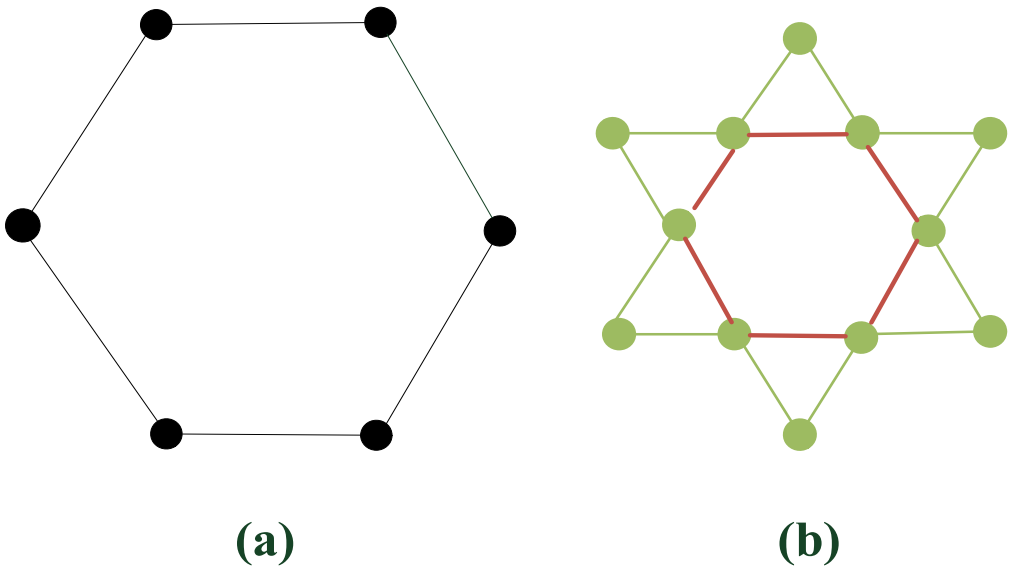


Figure 3.4 (a) Regular Hexagon; (b) Hexagram or star of David

Step 2: Remove the interior hexagon of the hexagram (represented in red colour in

Figure 3.4 (b)) to form a star like structure as shown in Figure 3.5.

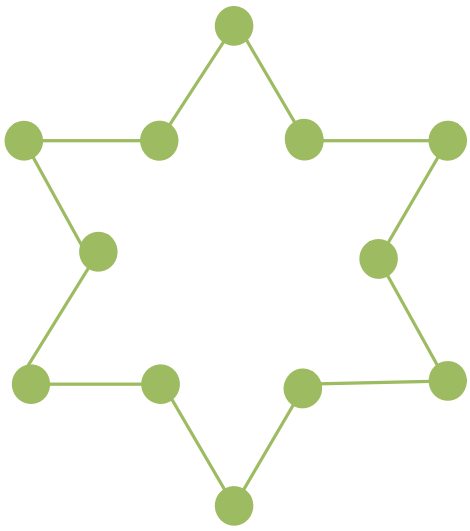


Figure 3.5 star

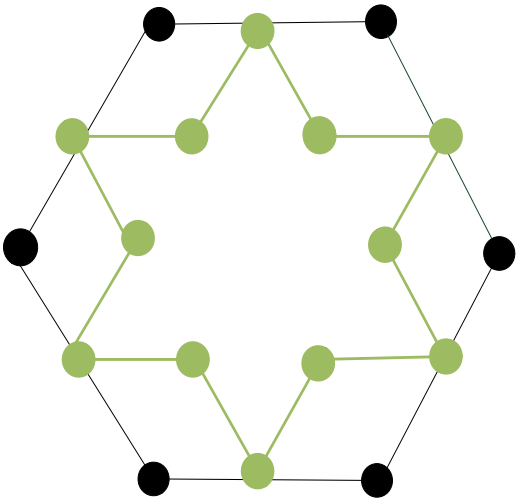


Figure 3.6 Primary block of Hexagonal Star

Step 3: Unite the hexagon(Figure 3.4 (a))and the star (Figure 3.5) to construct the hexagonal star (HS) structure. This forms the primary block of the HS topology (Figure 3.6).

Step 4: There are different approaches to extend the topology to interconnect a

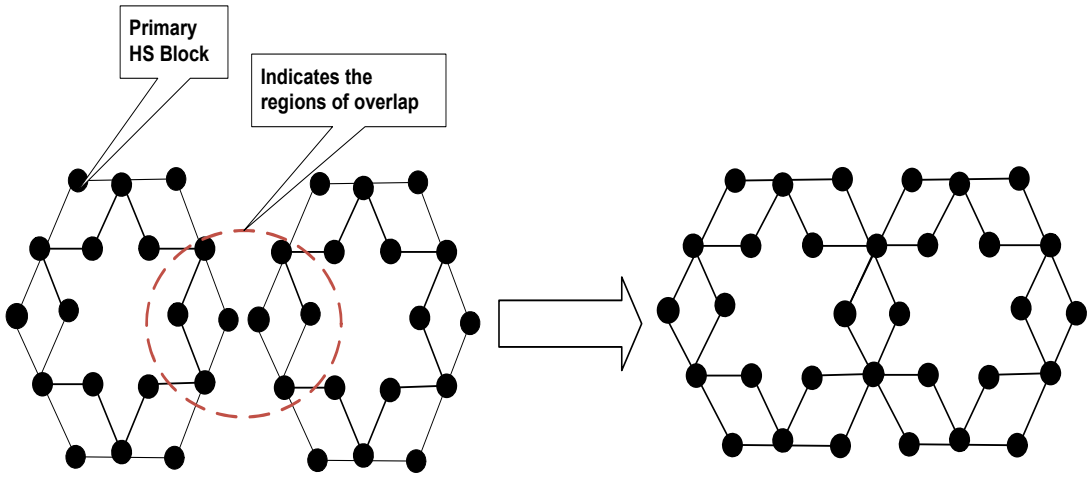


Figure 3.7 Adding 1st HS block

larger number of nodes. One of which is proposed and analyzed in this study. It is to extend the topology in a layered manner by adding the HS blocks around the six faces of the hexagonal star, overlapping some nodes. Figure 3.7 and Figure 3.8 show the way to add the HS blocks in the second layer of the topology. By this approach, the symmetry of the topology is retained.

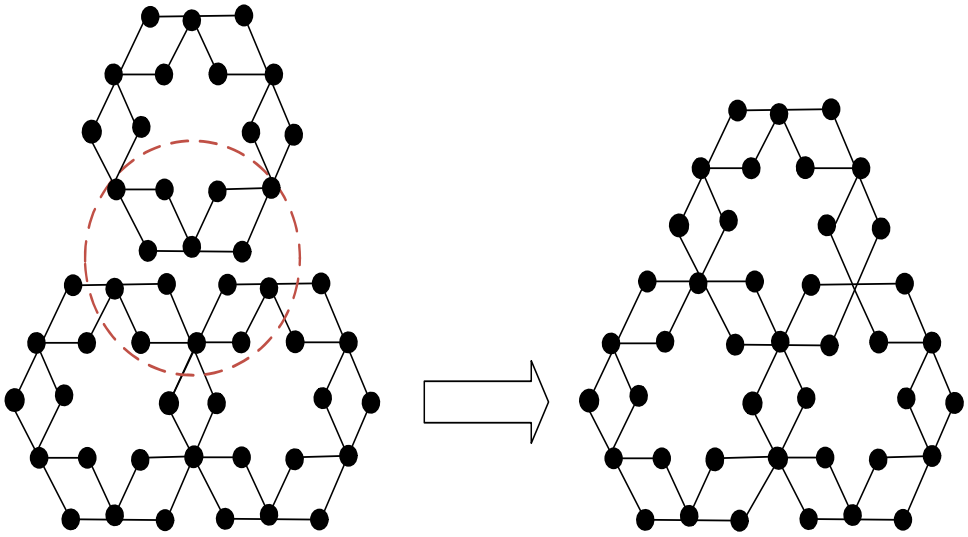


Figure 3.8 Adding 2nd HS block

Other possible approaches to extend the topology are suggested in section 3.2.5.

3.2.2 Router and organization of links

Considering the HS topology that is depicted in Figure 3.1 and Figure 3.2, there are routers with different degrees (from 3 to 7). The links that connect two neighboring routers are either horizontal (E,W) or oriented in 45° or 135° (NE,NW,SE,SW). The realization of the diagonally i.e., 45° or 135° oriented links has to be done in an efficient way to reduce the interconnect length, utilising the euclidean optimum so as to minimize the area overhead. There are two possible ways to layout the architectures with 45° or 135° oriented wires – Manhattan layout and X-architecture layout as depicted in Figure 3.9. With X-architecture for realizing layouts with 45° or 135° oriented wires, the effective link length can be reduced up to $1/\sqrt{2}$ times the length of wire laid with Manhattan layout. Thus the diagonally oriented links i.e., NE, NW, SE, SW directed links in HS topology can be efficiently realized with X-architecture chip design. The router with the maximum degree in HS topology, with its port directions, is depicted in Figure 3.9 (c).

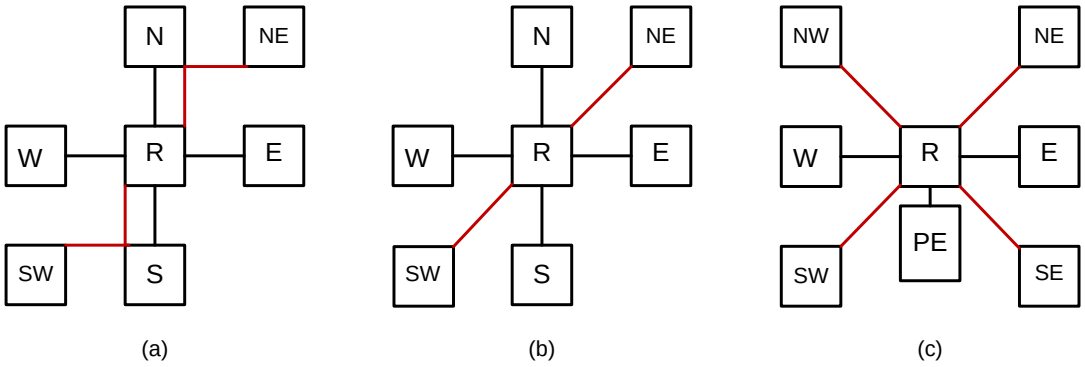


Figure 3.9 (a) Manhattan layout style; (b) X-architecture layout style; (c)Router with all port directions including the local port that connects PE in the HS topology using X-architecture layout style

- Possible layouts are marked in Red
- E- East, W-West, N-North, S- South, NE- North East, NW - North West, SE -South East, SW -South West, R - Router, PE - Processing Element

3.2.3 Buffer Resources

The router model adopted for experimentation uses wormhole packet switching, credit-based flow-control, and virtual channel input-buffered router. An identical router

model and the same parameters including the number of virtual channels (vcs) per port and number of buffers per vc have been used to analyse both Mesh and HS topologies except the number of ports per router. The number of ports per router is different in both the topologies because of the topological differences i.e., the way in which the nodes are arranged and connected for the same network size. Because of the topological differences, the number of links that connect the nodes thereby the number of ports per router thereby the number of buffer resources per router differ for the topologies.

Buffer resources contribute a significant part to the power and area overhead of the NoC architecture. The total number of buffer resources required for a topology of a given network size gives a fair comparison of the area occupied by the networks as all other factors that affect the area remain nearly same for both HS and Mesh topology.

The total number of buffer resources that are required in a network of a given size can be calculated as follows:

Step1: Number of buffer resources per router (br) = Number of input ports of the router × No. of VCs per each port × Number of buffers per each VC

Step2: Total number of buffer resources = $\sum_{i=1}^{N} br_i$, where N=total number of routers in the network and br_i is the number of buffer resources of router i.

In a much simpler way, classify the routers into different groups based on the node degree i.e., number of ports. Calculate the number of buffer resources per router based on the node degree as in step 1 and multiply it with the number of routers in that node degree group. Sum up the number of buffer resources obtained for each node degree routers group to get the total number of resources in the network of a given size.

As an illustration, consider 18 node network size HS and Mesh topologies with 4 virtual channels and 4 buffer depth configurations per port per each router.

For 32 node HS topology,

No. of routers with 7 ports = 2

Number of buffer resources = $2 \times 7 \times 4 \times 4 = 224$

No. of routers with 5 ports = 8

Number of buffer resources = $8 \times 5 \times 4 \times 4 = 640$

No. of routers with 3 ports = 22

Number of buffer resources = 22 × 3 × 4 × 4 = 1056

Total number of buffer resources in 32 node HS topology = 224+640+1056 = 1920

For 32 node Mesh topology,

No. of routers with 5 ports = 12

Number of buffer resources = 12 × 5 × 4 × 4 = 960

No. of routers with 4 ports = 16

Number of buffer resources = 16 × 4 × 4 × 4 = 1024

No. of routers with 3 ports = 4

Number of buffer resources = 4 × 3 × 4 × 4 = 192

Total number of buffer resources in 32 node Mesh topology = 960 +1024+192 = 2176

Comparison of total number of buffer resources required to be employed by HS and Mesh topologies of same size and identical virtual channel, buffer configurations is presented in Table 3.2. From Table 3.2, one can observe that HS topology requires fewer buffer resources than Mesh topology of identical network size. It indicates that the area occupied by HS topology will be lesser compared to that of Mesh topology.

Table 3.2 Comparison of total number of Buffer Resources for HS and Mesh topologies

	4 VC 4 BD (16 BR)		2 VC 4 BD (8 BR)	
Topology	18 nodes	32 nodes	18 nodes	32 nodes
HS	1056	1920	528	960
Mesh	1152	2176	576	1088
% Redu in BR	8.34	11.76	8.34	11.76

[BD : Buffer Depth; BR : Buffer Resource; % Redu: % Reduction
VC : Virtual Channel; BD : Buffer Depth]

Table 3.3 Topological Parameters

Topology	Diameter	Average Node Degree	Number of Links	Bisection Width
Hexagonal Star (HS)	$1.25N^{0.5428}$	$3.46N^{0.0228}$	$1.4677N - 3.2279$	$2.09N^{0.4621}$
Hexagonal Star V1 (HS-V1)	$1.501N^{0.51}$	$3.44N^{0.0286}$	$1.5164N - 3.9191$	$2.28N^{0.5021}$
Mesh	$P + Q - 2$	$3.35N^{0.0665}$	$P(Q-1) + Q(P-1)$	$\min(P, Q)$
Torus	$[P/2] + [Q/2]$	5	$[P(Q-1) + Q(P-1)] + P + Q$	$2 \times \min(P, Q)$
Honeycomb Mesh	$1.63N^{0.5} - 1$	$2.64N^{0.0844}$	$1.3597N - 2.6367$	$0.82N^{0.5}$

Notes: In this analysis, the average degree of the Router node (instead of the maximum degree) is considered. This is to estimate the network cost accurately. P is the number of nodes in the X-direction and Q is the number of nodes in the Y-direction for Mesh and Torus. $N = PXQ$, N refers to the total number of nodes in the network. As Bisection width is an integer value and for Hexagonal Star topology, its value is always a multiple of 4, the values obtained from the formula are rounded to the nearest multiple of 4. For honeycomb Mesh, the value is rounded to the nearest integer.

3.2.4 Topological parameters

Topological parameters of the proposed HS topology have been analysed and compared with Mesh, Torus and Honeycomb Mesh topologies. The topologies considered for evaluating HS topology are shown in Figure 3.10 . The topology parameters are presented in Table 3.3. Based on the analysis of the topological parameters, a numerical comparison of the topological parameters is performed for the proposed HS and HS-V1 topologies and compared with the state of the art network topologies and is given in Table 3.4.

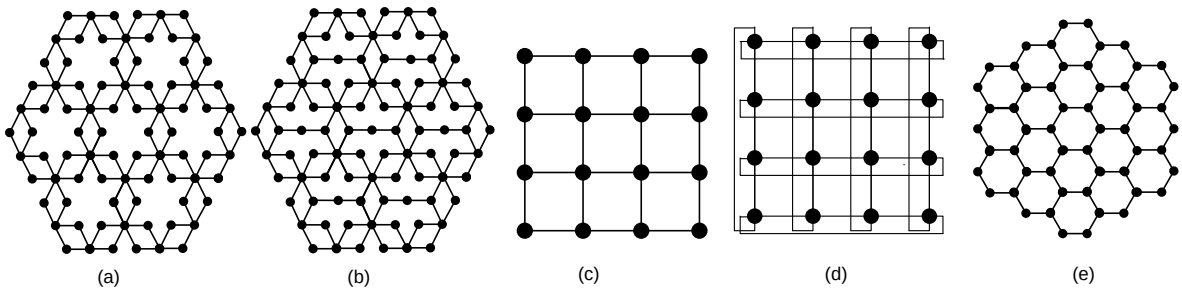


Figure 3.10 (a) 2D-Hexagonal Star; (b) 2D-Hexagonal Star-V1; (c) Mesh; (d) Torus; (e) Honeycomb Mesh

For the NoC topologies considered, topological parameters - average node degree, network diameter, bisection width, network cost and packing density for different node

Table 3.4 Numerical comparison of the Topological Parameters

Topology	Number of nodes	Average Node degree	Network diameter	Network cost	Bisection width	Packing density
Hexagonal Star (HS)	18	3.69	6.01	22.18	8	0.81
	32	3.74	9.22	34.48	12	1.04
	54	3.78	10.92	41.28	12	1.31
	76	3.81	13.15	50.10	12	1.51
	84	3.82	13.88	53.02	16	1.58
	125	3.86	17.23	66.51	20	1.88
	144	3.87	18.6	71.98	20	1.99
	198	3.89	22.11	86.01	24	2.3
Hexagonal Star V1 (HS-V1)	19	3.75	6.74	25.28	10	0.75
	34	3.81	9.07	34.56	16	0.98
	58	3.87	11.9	46.05	16	1.26
	70	3.89	13.1	50.96	16	1.37
	82	3.91	14.2	55.52	16	1.48
	124	3.96	17.54	69.46	28	1.79
	148	3.98	19.2	76.42	28	1.94
	199	4.01	22.32	89.5	28	2.22
Mesh	18	4.06	7	28.42	3	0.63
	32	4.21	10	42.1	5	0.76
	54	4.37	13	56.81	6	0.95
	76	4.47	21	93.87	4	0.81
	84	4.5	17	76.5	7	1.1
	125	4.62	28	129.36	5	0.97
	144	4.66	22	102.52	12	1.4
	198	4.76	29	138.04	9	1.43
Torus	18	5	4.5	22.5	6	0.8
	32	5	6	30	10	1.07
	54	5	7.5	37.5	12	1.44
	76	5	11.5	57.5	8	1.32
	84	5	9.5	47.5	14	1.77
	125	5	15	75	10	1.67
	144	5	12	60	24	2.4
	198	5	15.5	77.5	18	2.55
Honeycomb Mesh (HCM)	16	3.33	5.52	18.38	3	0.83
	32	3.54	8.22	29.1	5	0.98
	54	3.7	10.98	40.63	6	1.22
	76	3.8	13.21	50.1	7	1.41
	85	3.84	14.02	53.84	8	1.46
	125	3.97	17.22	68.36	9	1.73
	145	4.01	18.63	74.71	10	1.83
	198	4.12	21.94	90.4	12	2.09

configurations have been plotted as shown in Figure 3.11 (a)-(e). Based on this graphical analysis, the performance of the proposed topology can be estimated and compared with the existing topologies. These figures indicate that the average node degree, diameter, and network cost of Hexagonal Star topology are less than that of Mesh topology. The average node degree of HS topology is less than that of the Torus, but the diameter of HS is observed to be more than that of the Torus. This results in a higher network cost of HS topology when compared to that of Torus. Yet, the long wraparound links of the Torus make it less preferred. In this analysis, square, as well as rectangular Mesh and Torus are considered. As the number of nodes has increased, the network diameter has increased almost linearly for HS, HCM topologies. But for Mesh and Torus, it is not the

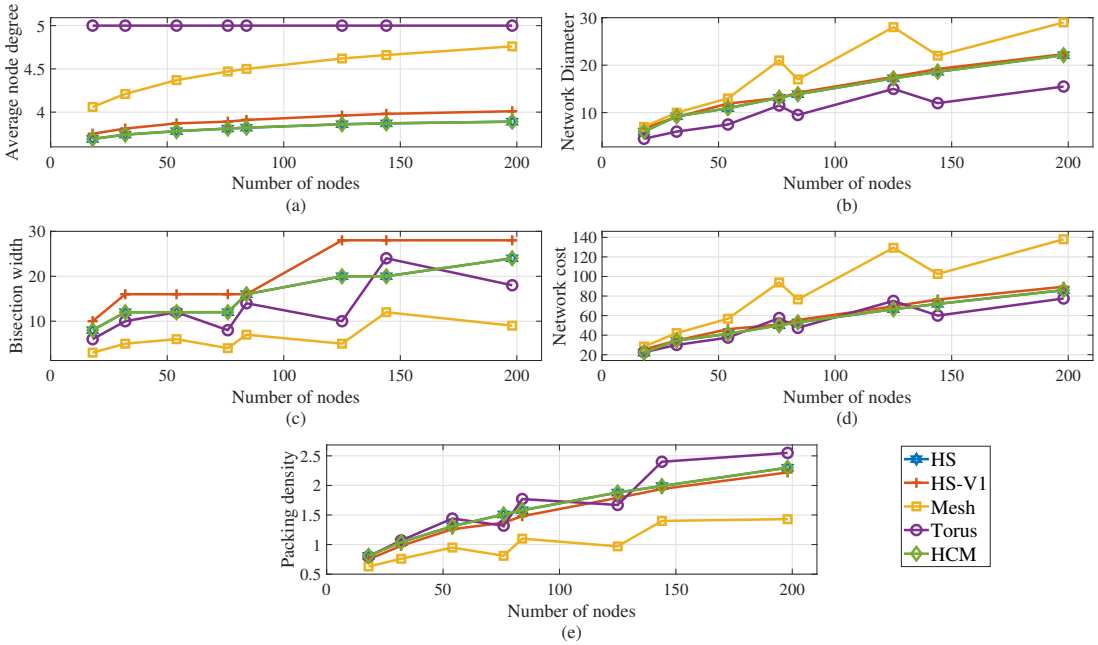


Figure 3.11 Graphical representation of the comparison of topological parameters

case. If square Mesh or square Torus are only considered for comparison, a linear increase in diameter may have been observed. But in the plots, some network sizes that can be realized only with rectangular Mesh or Torus are also considered, because of which the non-linearity has raised. The diameter and thereby the network cost of HCM topology has increased linearly with the number of nodes. These are slightly less than those for HS topology for fewer nodes, but as the number of nodes has increased, the diameter and the network cost are observed to be slightly higher for HCM than HS topology.

The bisection width of HS and HS-V1 topologies is higher than that of Mesh and HCM. It can be observed that the bisection width of HS and HS-V1 is higher than that of Torus for most of the cases except a few. This proves that more paths between any source and destination pair exist in HS topologies compared to Mesh, Torus, and HCM. The efficient design of routing algorithms for HS topology will result in higher performance of HS topology compared to all other considered topologies. The Packing Density of HS topology has been found to be more than that of Mesh and HCM and less than that of Torus topology. The larger package density indicates the smaller chip area of the VLSI design layout. So, it has become evident that the HS occupies a smaller chip area compared to Mesh and HCM.

3.2.5 Possible scalable architectures

There are other possible ways to extend the basic block of the HS topology to interconnect larger number of nodes other than the one which has been illustrated in the above section. These ways include vertical stacking, Mesh of Hexagonal stars and Honey comb of hexagonal stars as shown in Figure 3.12. These suggested scalable HS topologies have to be investigated in the future work. The present work focuses on the proposed scalable architecture i.e; the layered architecture with sharing of the nodes as depicted in Figure 3.2. It is simple and is feasible to realize it with two XY co-ordinate axes. *Ratatoskr* simulator [196] that uses XYZ coordinate system for locating the nodes has been chosen for implementing HS topology. For 2D implementation, X and Y axes are used whereas for 3D implementation Z axis can be used to represent the layers that are vertically stacked.

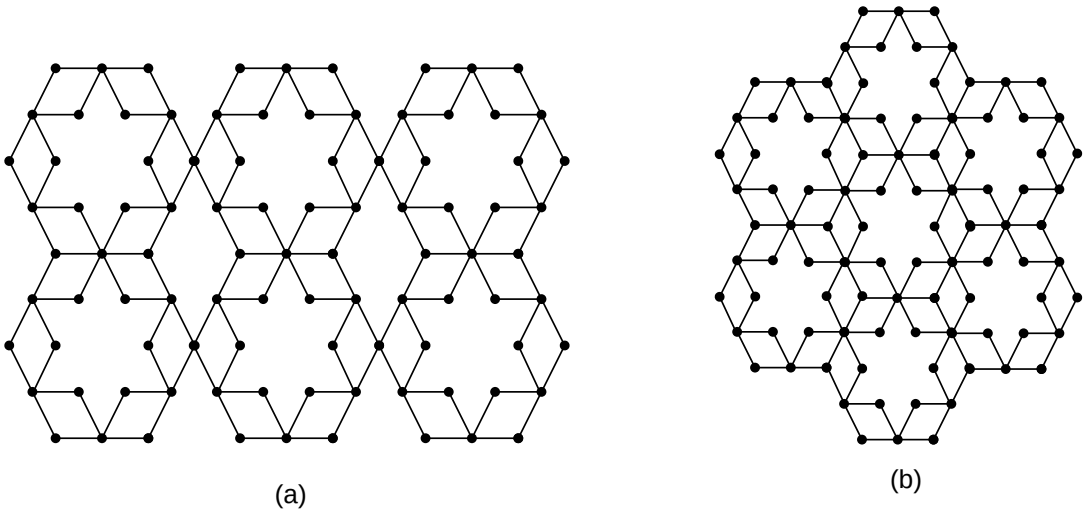


Figure 3.12 Scalable architectures of HS topology (a) Mesh of Hexagonal stars ; (b) Honeycomb Hexagonal star

3.3 Evaluation

Evaluation is two-fold. One is to examine the performance of the proposed topology, simulation has been conducted for different scenarios. The other one is to estimate the area required for implementing the topology, the topology has been synthesized on Xilinx Vivado design compiler [197]. The simulation and synthesis results of HS topology have been compared with the popularly used identically configured Mesh topology. The

simulation has been carried out using the *Ratatoskr* simulator [196].

Ratatoskr simulator, an open-source framework for power, performance, and area analysis of Networks-on-Chips (NoCs), supports cycle-accurate simulation and heterogeneous 3D integration. It is implemented in C++ (C++17) using SystemC 2.3.3 class library Parameters. The simulator is modular and flexible as the code is implemented using inheritance and polymorphism programming paradigms. It includes a synthesizable router-hardware implementation, a cycle-accurate (CA) NoC simulator and a transaction-level application model. It supports different synthetic patterns, netrace and task graphs. *Ratatoskr* provides an easy two-step user interaction: - A single point-of-entry allows to set design parameters and automatic generation of PPA reports. Also, it provides a GUI of the simulated network. With the motive to investigate the proposed HS topology in different aspects such as area, power, and 3D heterogeneous integration in the future work, *Ratatoskr* simulator has been employed for the preliminary evaluation of the topology.

The routing strategy, simulation methodology, and results discussion are presented in the following sections.

3.3.1 Routing algorithm

The routing algorithm determines the path along which the packet has to traverse from source to destination. A new static routing algorithm has been proposed for the HS topology. The routing strategy in HS topology has been based on the shortest path routing. The topology has been implemented in Ratatoskr simulator [196] in which XYZ coordinate system has been used for locating the nodes. To keep the addressing scheme and distance calculation simple, the xyz coordinate system has been used to locate the nodes. For a given pair of source and destination, the routing algorithm fetches all the connected nodes of the source node and calculates the distance from each of the connected nodes to the destination. The packet is then routed to the connected node (CN) that has the shortest distance to the destination. If two connected nodes have the same distance to the destination, the algorithm chooses any one node randomly. The same process is repeated with the connected node as the source until the packet reaches the destination. This ensures that the packet always traverses in the shortest possible path. The shortest paths in a network are nearly simple paths in that a node is visited not more than once.

This avoids looping and thereby deadlock. Thus, the algorithm is simple and deadlock-free. The routing algorithm is illustrated in Algorithm 1.

Algorithm 1: Pseudo Code for Shortest Path routing in HS

Input: Current Source, src(x_s, y_s) address and Destination, dst (x_d, y_d) address
Output: Next Router, the packet to be forwarded to
Process:
if *src==dst* **then**
    Route the packet to the local port of the PE
else
    Fetch the connected nodes list of the Source node
    if *the destination is in the connected nodes list* **then**
        Route the packet to the connected node i.e the destination node.
    else
        Compute the distance from each of the connected nodes to the destination.
        Get the connected node (CN) with the shortest distance.
        Route the packet to that connected node. (CN)
src==CN
Repeat the process until src==dst

3.3.2 Performance analysis

18-node and 32-node HS and Mesh networks have been simulated using the *Ratatoskr* simulator [196]. For Mesh topology, static XY routing has been used, whereas, for HS topology, the above proposed static routing has been used. The network has been simulated for 10000ns under different synthetic traffic patterns - uniform, transpose, tornado, and bit complement. Each node(Processing element or router) has been clocked with a clock delay of 1ns. For the 32-node network, an input buffered router with 4 virtual channels (VC), each VC with 4 buffer depths has been considered. For the 18-node network, an input buffered router with 2 virtual channels, each VC with 4 buffer depths has been considered.

Table 3.5 Comparison of HS and Mesh topologies simulated under uniform traffic pattern varying injection rates

	Average Packet Latency(ns)			**Average Flit Latency(ns)**		
Injection Rates	HS-18	Mesh-18	%Reduc	HS-18	Mesh-18	%Reduc
0.01	12.13	15.19	20.14	9.04	12.10	25.29
0.04	12.90	16.19	20.32	9.58	12.81	25.21
0.07	13.87	17.67	21.51	10.30	13.84	25.58
Injection Rates	HS-32	Mesh-32	%Reduc	HS-32	Mesh-32	% Redu
0.01	14.58	18.38	20.67	11.48	15.29	24.92
0.04	17.00	20.19	15.80	13.17	16.53	20.33
0.07	20.48	22.57	9.26	15.89	18.32	13.26

Notes: ns: nanoseconds, % Redu: % Reduction , HS-18 : 18 node Hexagonal Star topology, HS-32 : 32 node Hexagonal Star topology, Mesh-18 : 18 node Mesh topology, Mesh-32 : 32 node Mesh topology

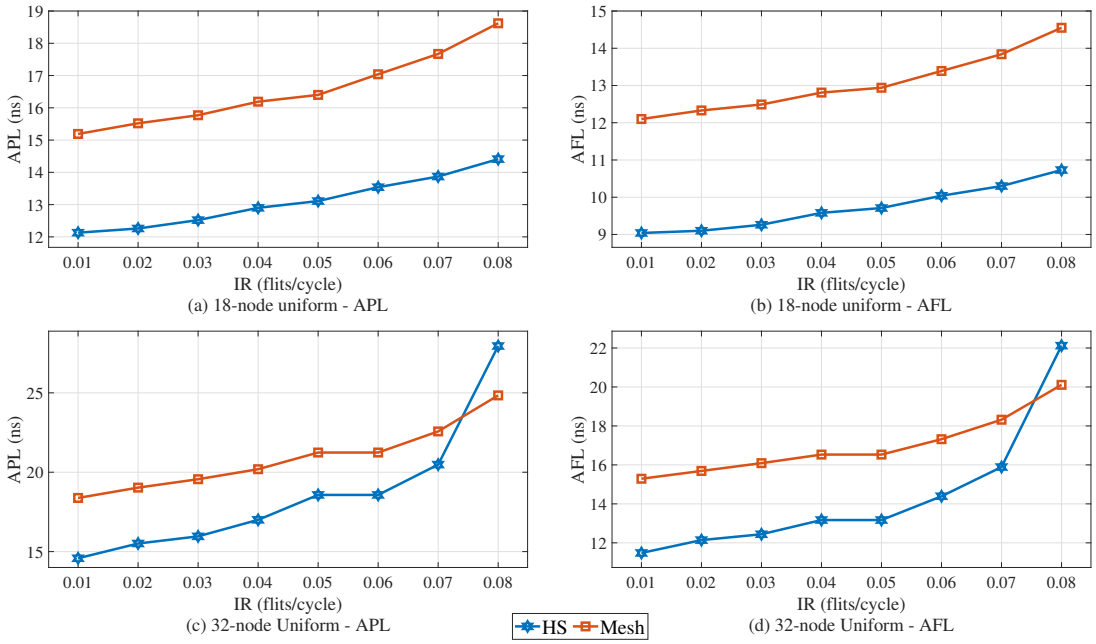


Figure 3.13 Simulation results of HS and Mesh topologies for uniform traffic pattern

Table 3.6 Comparison of HS-18 and Mesh-18 topologies simulated under transpose traffic pattern varying injection rates

	Average Packet Latency(ns)			**Average Flit Latency(ns)**		
Injection Rates	HS-18	Mesh-18	%Reduc	HS-18	Mesh-18	%Reduc
0.01	12.76	16.59	23.09	9.73	13.56	28.24
0.04	13.26	16.90	21.54	10.03	13.77	27.16
0.07	14.02	17.45	19.66	10.58	14.17	25.34

Notes: ns: nanoseconds, % Redu: % Reduction , HS-18 : 18 node Hexagonal Star topology, Mesh-18 : 18 node Mesh topology

Table 3.7 Comparison of HS-32 and Mesh-32 topologies simulated under transpose traffic pattern varying injection rates

	Average Packet Latency(ns)			**Average Flit Latency(ns)**			**%DeliveredPackets**	
InjectionRate	HS-32	Mesh-32	%Redu	HS-32	Mesh-32	%Redu	HS-32	Mesh-32
0.01	14.44	18.12	20.31	11.31	15.06	24.90	100.00	100.00
0.04	16.95	22.72	25.40	13.01	18.44	29.45	100.00	100.00
0.07	33.51	44.89	25.35	27.13	37.40	27.46	86.97	80.22

Notes: ns: nanoseconds, % Redu: % Reduction , HS-32 : 32 node Hexagonal Star topology, Mesh-32 : 32 node Mesh topology

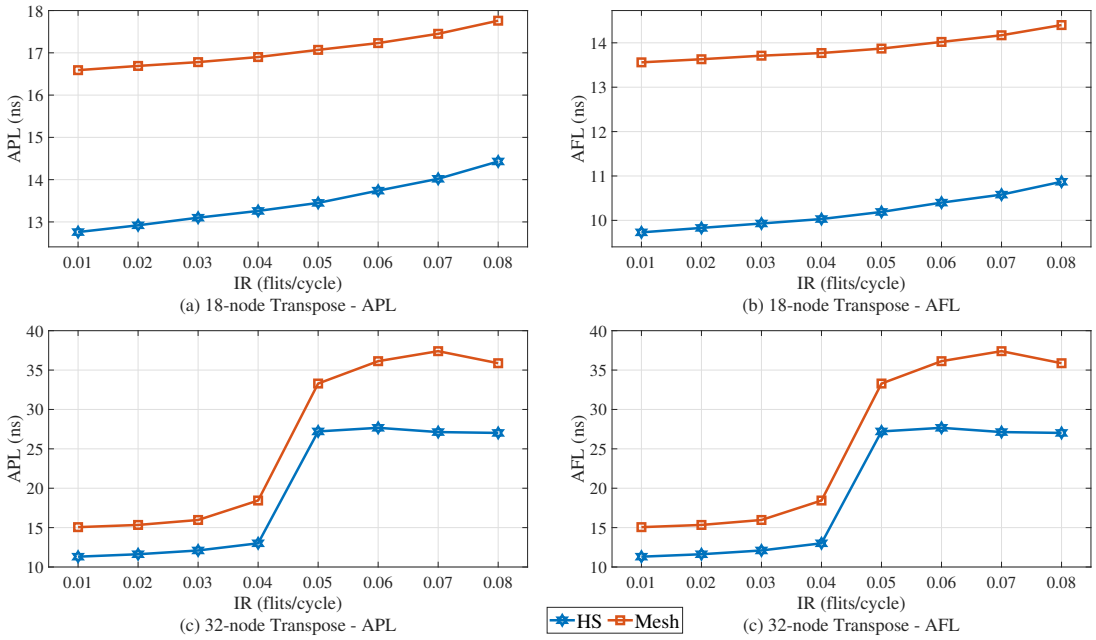


Figure 3.14 Simulation results of HS and Mesh topologies for transpose traffic pattern

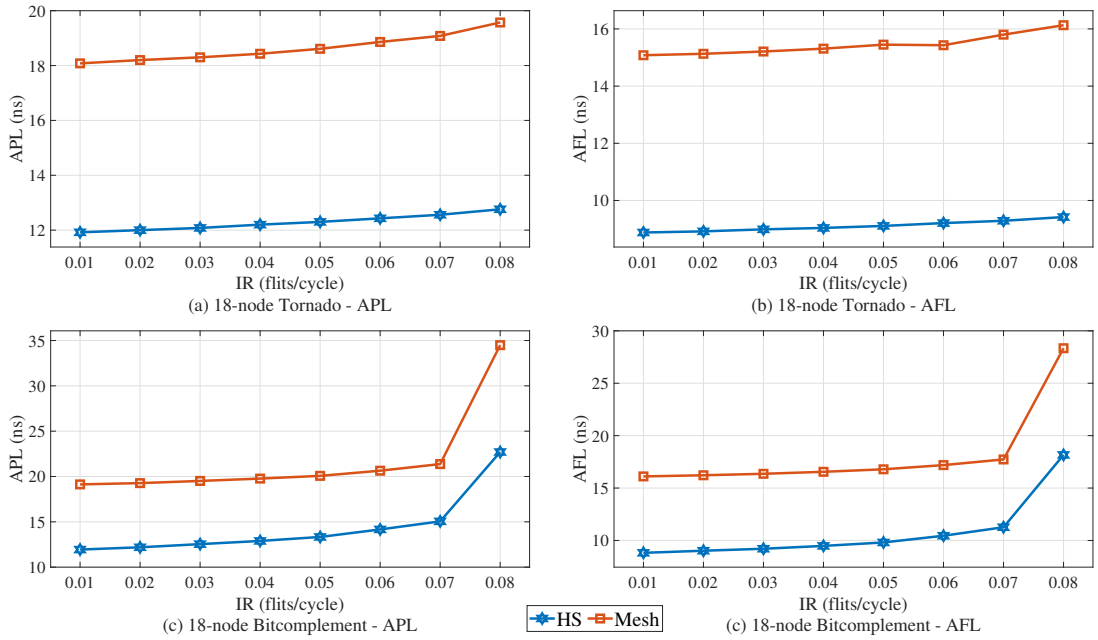


Figure 3.15 Simulation results of HS and Mesh topologies for tornado and bit complement traffic pattern

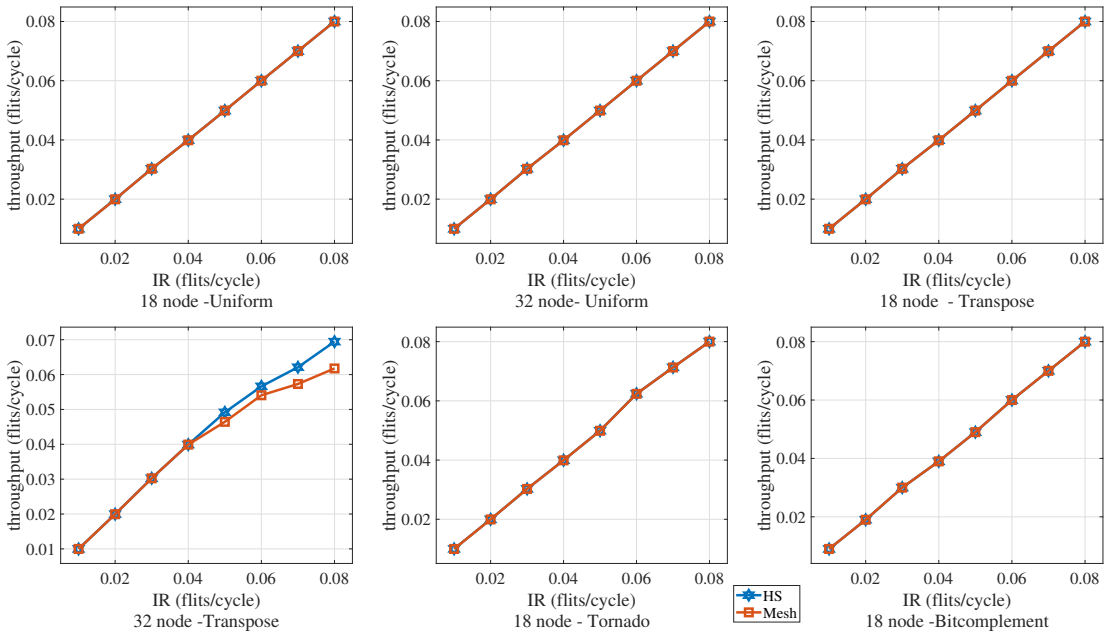


Figure 3.16 Throughput characteristics of HS and Mesh topologies for varying injection rates from 0.01 to 0.08

Table 3.8 Comparison of HS and Mesh topologies simulated under tornado and bit complement traffic patterns varying injection rates

	Average Packet Latency(ns)			**Average Flit Latency(ns)**		
Injection Rate	HS-18	Mesh-18	%Redu	HS-18	Mesh-18	%Redu
Tornado						
0.01	11.92	18.08	34.07	8.88	15.08	41.11
0.04	12.20	18.43	33.80	9.04	15.31	40.95
0.07	12.56	19.08	34.17	9.29	15.80	41.20
Bit complement						
0.01	11.94	19.13	37.58	8.82	16.11	45.25
0.04	12.89	19.77	34.80	9.48	16.55	42.72
0.07	15.05	21.37	29.57	11.26	17.73	36.49

Notes: ns: nanoseconds, % Redu: % Reduction , HS-18 : 18 node Hexagonal Star topology, Mesh-18 : 18 node Mesh topology

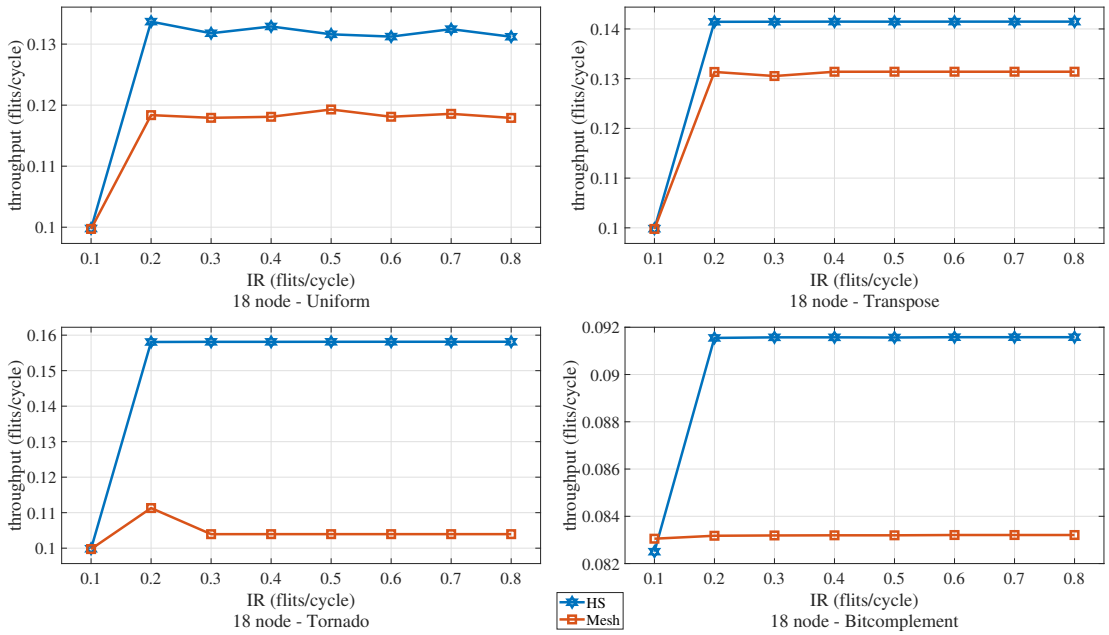


Figure 3.17 Throughput characteristics of HS and Mesh topologies for varying injection rates from 0.1 to 0.8

Mesh and HS networks with identical configuration parameters have been simulated under different traffic patterns by varying the traffic loads i.e., injection rates (IR). Simulation results have been tabulated as well as plotted. As such, for the sake of brevity, results for only three injection rates have been tabulated to observe the latency reduction and saturation throughput values. Tables 3.5, 3.6, 3.7 and 3.8 show the simulation results

Table 3.9 Comparison of HS-18 and Mesh-18 topologies simulated under different traffic patterns varying injection rates from 0.1 to 0.8

	Average Packet Latency(ns)			**Average Flit Latency(ns)**			**%Delivered Packets**	
InjectionRate	HS-18	Mesh-18	%Redu	HS-18	Mesh-18	%Redu	HS-18	Mesh-18
Uniform								
0.1	15.53	21.16	26.61	11.67	16.56	29.53	99.79	99.72
0.2	22.03	30.03	26.64	17.23	23.36	26.24	66.84	59.18
0.3	22.39	30.09	25.59	17.58	23.40	24.87	43.93	35.38
0.4	22.19	30.07	26.21	17.39	23.40	25.68	33.22	29.52
0.5	22.52	29.73	24.25	17.72	23.10	23.29	26.32	23.86
0.6	22.57	29.99	24.74	17.75	23.33	23.92	21.87	19.68
0.7	22.30	29.82	25.22	17.51	23.19	24.49	18.92	16.94
0.8	22.47	30.10	25.35	17.66	23.41	24.56	16.40	14.74
Transpose								
0.1	15.42	18.85	18.20	11.62	15.29	24.00	99.84	99.78
0.2	19.19	23.39	17.96	14.58	18.09	19.40	70.72	65.68
0.3	19.75	23.84	17.16	14.84	18.36	19.17	47.15	43.51
0.4	19.94	23.41	14.82	14.76	18.28	19.26	35.37	32.85
0.5	19.65	23.26	15.52	14.96	18.47	19.00	28.29	26.28
0.6	19.55	23.15	15.55	14.64	18.23	19.69	23.58	21.90
0.7	19.55	23.15	15.55	14.64	18.23	19.69	20.21	18.77
0.8	19.55	23.15	15.55	14.64	18.23	19.69	17.68	16.42
Tornado								
0.1	13.19	22.26	40.75	9.70	18.09	46.38	99.84	99.75
0.2	16.19	32.23	49.77	12.08	26.00	53.54	79.04	51.97
0.3	16.39	32.67	49.83	12.26	26.83	54.30	52.71	34.65
0.4	16.26	33.07	50.83	12.17	26.67	54.37	39.53	25.99
0.5	16.30	32.83	50.35	12.21	26.64	54.17	31.63	20.79
0.6	16.39	32.90	50.18	12.11	26.79	54.80	26.36	17.33
0.7	16.39	32.90	50.18	12.11	26.62	54.51	22.59	14.85
0.8	16.39	32.90	50.18	12.11	26.62	54.51	19.77	12.99
Bit complement								
0.1	24.68	38.07	35.17	19.91	30.56	34.85	82.50	83.05
0.2	17.32	38.48	54.99	12.67	30.08	57.88	45.77	41.59
0.3	17.82	38.24	53.40	12.57	29.95	58.03	30.52	27.73
0.4	17.82	38.03	53.14	12.57	29.92	57.99	22.89	20.80
0.5	17.97	38.15	52.90	12.72	29.95	57.53	18.31	16.64
0.6	17.67	37.74	53.18	12.57	29.89	57.95	15.26	13.87
0.7	17.67	37.74	53.18	12.57	29.89	57.95	13.08	11.89
0.8	17.67	37.74	53.18	12.57	29.89	57.95	11.45	10.40

Notes: ns: nanoseconds, % Redu: % Reduction , HS-18 : 18 node Hexagonal Star topology, Mesh-18 : 18 node Mesh topology

of Mesh and HS topologies simulated under different traffic patterns. It can be observed from the tabulated results, that the HS topology has shown a considerable reduction in the average packet latency (APL) and average flit latency (AFL) compared to that of Mesh topology. From Table 3.5, when the networks are simulated under uniform traffic, in comparison with Mesh topology, HS topology has noted a reduction of packet latency ranging from 20% to 22% and 9% to 20% for 18 and 32 nodes, respectively. HS (18-node and 32-node) and Mesh(18-node and 32-node) simulated under uniform traffic have achieved 100 percent throughput i.e., all the packets that are injected into the network have been delivered successfully. This has been observed for all the injection rates between 0.01 to 0.08. The same can be noticed from the throughput plots in Figure 3.16.

From Tables 3.6 and 3.7, when the networks are simulated under transpose traffic, in comparison with Mesh topology, HS topology has noted a reduction of packet latency ranging from 18% to 23% and 16% to 25% for 18 and 32 nodes, respectively. Further, 18-node HS and 18-node Mesh simulated under transpose traffic have achieved 100 percent throughput i.e., all the packets that are injected into the network have been delivered successfully. But the throughput of 32-node HS and 32-node Mesh simulated under transpose traffic has decreased with the increase in injection rate. 32-node HS topology has shown better throughput characteristics compared to 32-node Mesh topology which can be inferred from the results obtained from the comparison of the percentage of delivered packets. The same can be noticed from the throughput plots in Figure 3.16. The saturation throughput for HS is 0.07 whereas for Mesh, it is 0.06. From Table 3.8, when the networks are simulated under tornado traffic, in comparison with 18-node Mesh topology, 18-node HS topology has noted about 34% reduction of packet latency. When the networks are simulated under bit complement traffic, in comparison with 18-node Mesh topology, 18-node HS topology has noted a reduction of packet latency ranging from 29% to 38%. The throughput characteristics of 18 node HS and Mesh for tornado and bit complement traffic patterns have been plotted and can be observed from Figure 3.16 that 100 percent throughput has been achieved.

Figure 3.13, Figure 3.14, and Figure 3.15 present the graphical analysis of the simulation results to visually illustrate the trend of the latency characteristics. One can observe that the APL and AFL of the 18-node HS topology are consistently lower than those of the 18-node Mesh topology for the considered traffic patterns at all injection rates. A similar trend has been observed with 32-node HS and 32-node Mesh topologies.

Table 3.9, shows the simulation results of Mesh and HS topologies simulated for higher injection rates under different traffic patterns. It can be observed from the tabulated results, that the HS topology has shown a considerable reduction in the APL and AFL compared to those of Mesh topology. A reduction in average packet latency of about 25% for uniform traffic, about 17% for transpose, about 50% for tornado, and 50% for bit complement can be observed. As the trend of latency characteristics for higher injection rates is similar to that of the trend observed with lower injection rates, the plots for higher injection rates are omitted. Figure 3.17 shows the throughput characteristics of 18

node HS and 18 node Mesh topologies for higher injection rates. It can be observed from the plots that the saturation throughput of HS is higher than that of Mesh. 32-node HS has shown no improvement in the latency results when compared to that of the 32-node Mesh simulated for higher injection rates. However, with the benefit of larger diameter and larger bisection width of HS topology compared to those of Mesh topology, efficient dynamic routing can result in superior performance of HS topology for higher injection rates also. This has to be further investigated.

The simulation results corroborate the topological parameter-based study performed in section 3.2.4. It can be seen from the simulation results, that the HS topology has shown a considerable reduction in APL and AFL compared to that of Mesh topology. This is because the network diameter of HS topology is less than that of Mesh topology. Also, in HS topology, routers with different degrees, '3', '5', '7' are distributively located in the architecture in a way that the routers with higher degrees are located at the center of the topology where in congestion is more. In Mesh topology, the maximum degree is only '5'. With fewer links, HS topology has shown improved latency and throughput characteristics compared to that of an identically configured Mesh topology. This is because of the heterogeneous distribution of the routers with varying degrees and higher degree routers being located at the most congested central region of the topology.

3.3.3 Synthesis results

The 18 and 32-node HS and Mesh topologies have been synthesized on Xilinx vivado design compiler with Virtex7(VC709) as the target device. Virtual channel input buffered router with two virtual channels (VC), each VC with four buffers has been considered for implementing the topology. A packet size of 32 flits, each flit with 32 bits has been considered.

The synthesis results in Table 10 infer that HS topology has consumed less space when compared with that of an identically configured Mesh topology. When an 18-node network is considered, HS topology has utilized 12558 LUTs (Look Up Table), 19714 FFs(Flip-Flops) whereas Mesh has utilized 14522 LUTs, and 21566 FFs. Overall, a reduction of 13.5% and 8.58% in the utilization of LUTs and FFs respectively for 18-node HS topology compared to that of 18-node Mesh topology has been noted. A reduction of

17.5% and 11.6% in the utilization of LUTs and FFs respectively for 32-node HS topology compared to that of 32-node Mesh topology has been noted.

Table 3.10 Synthesis Results

Topology	Resource	Utilization	Available	Utilization%
Mesh-32	LUTs	28915	433200	6.67
	FFs	40790	866400	4.71
HS-32	LUTs	23839	433200	5.50
	FFs	36026	866400	4.16
Mesh-18	LUTs	14522	433200	3.35
	FFs	21566	866400	2.49
HS-18	LUTs	12558	433200	2.90
	FFs	19714	866400	2.28

Notes: HS-18: 18 node Hexagonal Star topology, Mesh-18: 18 node Mesh topology, HS-32: 32 node Hexagonal Star topology, Mesh-32: 32 node Mesh topology, LUT : Look Up Table, FF: Flip-Flop

3.4 Summary

This chapter is dedicated to exploring and evaluating a 2D hybrid topology employing electrically wired links which is capable of outperforming widely used conventional Mesh topology. To fulfill this objective, a novel, scalable, symmetrical hybrid Hexagonal Star topology has been proposed and evaluated. Following are the primary findings that are drawn from the study:

- The quantitative and graphical analysis of the topological parameters has shown that the network cost of HS is lower than that of HS, and HCM and slightly higher than that of Torus. The packing density of HS is higher than that of Mesh, and HCM and lower than that of Torus. The bisection width of HS is higher than that of Mesh, Torus, and HCM.

- Experiments were conducted for different scenarios to analyse the performance of the proposed topology and the results were compared with Mesh topology.

- The simulation results showed that HS topology indicated a reduction of packet latency ranging from 15% to 50% and 9% to 23% for 18 and 32 nodes, respectively in comparison to Mesh topology.
- For an 18-node network, the proposed HS topology outperformed Mesh topology in terms of Latency for low and high injection rates. For 32 32-node networks, HS topology showed a reduced latency compared to Mesh topology for low injection rates only.
- The results from synthesis of HS and Mesh topologies showed that HS topology consumed less area when compared to that of Mesh topology.
- The preliminary analysis of the topological parameters and simulation evaluation showed that the HS topology outperformed Mesh in terms of latency in certain scenarios.

Summing up, the proposed topology with the benefits of a hexagonal-shaped structure, lower network diameter, lower network cost, higher packing density, and higher bisection width in comparison with Mesh topology can be considered to be a potential choice for on-chip interconnection networks. Efficient routing strategies and flow control mechanisms to exploit the benefits of large bisection width of the HS topology need to be investigated in future research. Further, the proposed topology is to be evaluated for larger network sizes under different traffic patterns.

In the next chapter, the focus is on exploring NoC architectures that are capable of meeting the performance demands of complex SoC, while maintaining the simple, scalable, and regular structure of Mesh. To achieve this objective, the chapter introduces and evaluates DiamondMesh, a diagonal Mesh-based topology. This topology proves to be a well-balanced diagonal Mesh topology, bringing a trade-off among area-performance-power metrics. Further, 2D-DiamondMesh is extended to 3D-DiamondMesh and is evaluated in the context of large-scale networks.

Chapter 4

DiamondMesh : A diagonal Mesh topology for NoC

4.1 Introduction

The prime design objective of on-chip interconnect networks is to design performance-area- power optimized NoC topology. Numerous interconnect topologies have been proposed as part of the NoC paradigm. Of these topologies, Mesh is the most extensively used and favoured architecture for implementing less sophisticated SoCs due to its simple, scalable, regular structure, low-radix routers, and short-range links but with a drawback of large diameter [15]. There are numerous techniques for optimising the mesh topology's performance and enabling its application in complex SoCs. The traditional method is to increase the number of virtual channels or to increase the buffer capacity. This results in an increase in network expenses associated with area overhead and energy consumption. An alternative technique is to incorporate diagonal links in the regular mesh-based NoC and employ high radix routers to enhance the performance of the NoC. In recent times, several diagonal mesh topologies have been proposed. Each topology is distinguished by the arrangement of diagonal links included in the regular mesh topology. A few diagonal mesh topologies are DMesh [24], XDMesh [25], PDNoC [26], ZMesh [27]. Though the existing diagonal mesh topologies improve the performance of the Mesh based NoC in terms of latency, they introduce power and area overheads to the network. In this context, the present chapter proposes a diagonal mesh topology that improves the performance of the Mesh-based NoC and also brings a trade-off among latency, area, and power parameters compared to all other considered diagonal mesh-based topologies. The novelty of the proposed topology lies in the pattern in which the diagonal links are incorporated

in the baseline Mesh topology. Most importantly, the topology has been manifested to be more balanced than the other considered diagonal mesh-based topologies in terms of the area-performance-power tradeoffs. DiamondMesh has more links compared to Mesh, and XDMesh, which facilitates reduced latency. It has fewer diagonal links compared to DMesh which facilitates it to be area and power efficient. It has the same number of links compared to ZMesh, but the pattern in which the diagonal links are incorporated differs, which facilitates to reduction of diameter thereby reducing latency. DiamondMesh is suitable to realize square as well as rectangular Mesh-based NoCs, unlike XDMesh which is suitable to realize only square Mesh-based NoCs. The details of the proposed topology have been discussed in the following section.

4.2 2D-DiamondMesh

4.2.1 Proposed DiamondMesh topology

DiamondMesh topology is developed by integrating diamond-like patterned diagonal links in the baseline Mesh topology. Figure 4.1 shows the DiamondMesh topology.

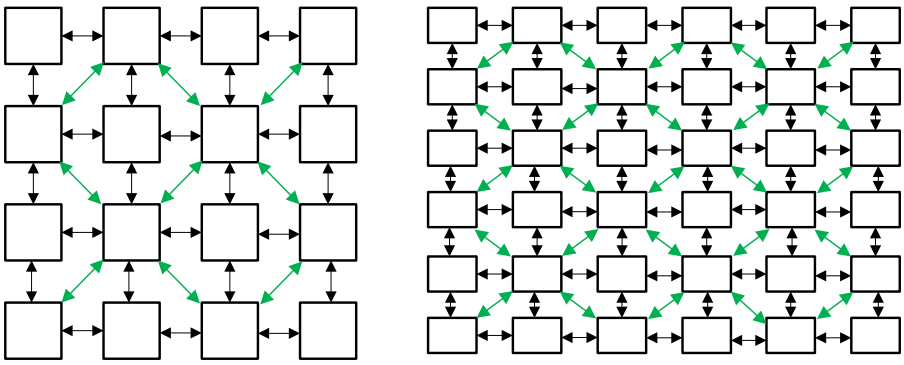


Figure 4.1 (a)16 Node DiamondMesh; (b)36 Node DiamondMesh

Theoretical definition of Diamond topology is given as follows: Consider P×Q DiamondMesh topology where P denotes the number of nodes in X direction and Q denotes the number of nodes in Y direction. It consists of a set of nodes,

$N = (x, y) \quad \| \quad 0 <= x <= (P-1), \quad 0 <= y <= (Q-1)$

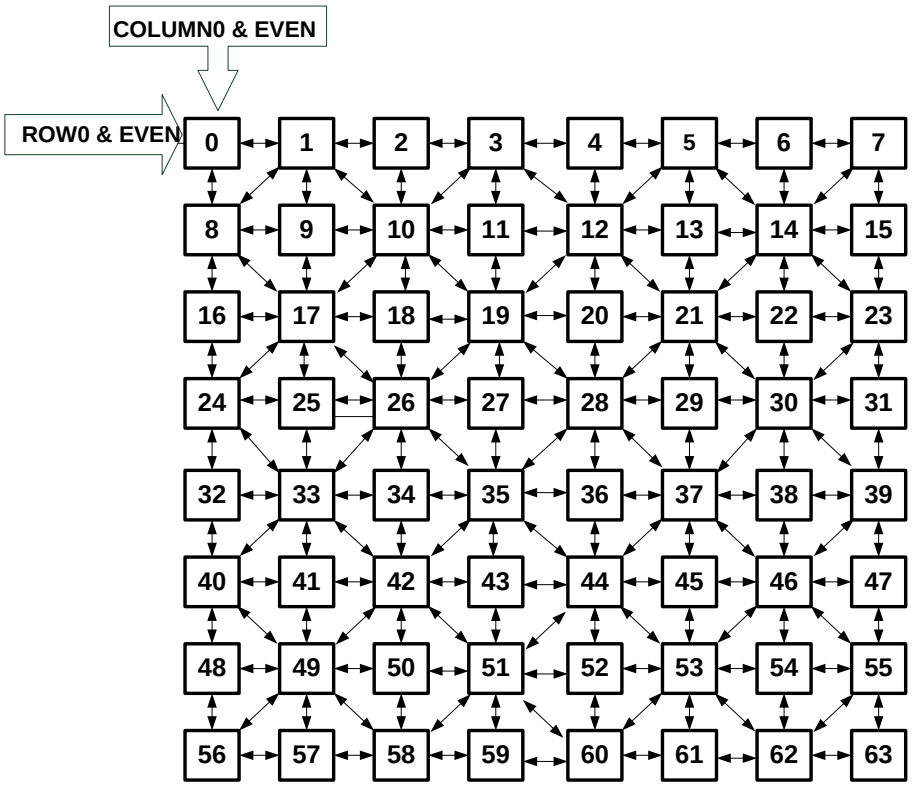


Figure 4.2 8×8 DiamondMesh

Two nodes N_1=(x_1, y_1) and N_2=(x_2, y_2) are connected to each other by horizontal i.e., East(E) or West(W) or vertical i.e., North(N) or South(S) link if their distance is equal to one, according to the euclidean distance. Integrating the diagonal links is based on the following two rules. Both the rules must be satisfied. Rule 1: For a node R, there exists a diagonal link to its diagonally located node in any of the four directions: NorthEast (NE), NorthWest (NW), SouthEast (SE), SouthWest (SW) if their distance is equal to $\sqrt{2}$, according to the euclidean distance. Rule 2: For Node R= (x,y), if x is even and y is odd or x is odd and y is even, there exists a diagonal link between R and all the other diagonally located nodes of R, else, no diagonal link exists i.e., if x,y both are even or both are odd then no diagonal link exists for the node R.

As an illustration, consider 8×8 DiamondMesh topology as shown in Figure 4.2. The leftmost column is column 0 and is considered as even. The topmost row is Row 0 and is considered as even. Consider any random node, say, node 19. The coordinates (x,y) of node 19 are '3', and '2' respectively where x represents the column number and y represents the row number. In this case, x is odd and y is even as such there exists diagonal

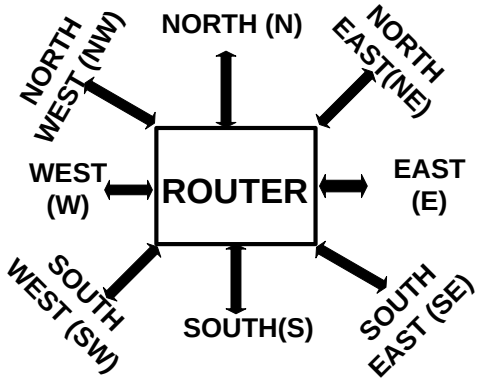


Figure 4.3 Router with Port Directions

links for node 19 to all its diagonally located neighboring nodes. Thus, it has diagonal links to node 12, node 10, node 28, and node 26 in NE, NW, SE, and SW directions respectively. Also, it has vertical and horizontal links connecting its neighboring nodes that are at distance equal to one, according to the Euclidean distance. Thus, node 19 has horizontal links to node 20, and node 18 in E, and W directions respectively and vertical links to node 11, and node 27 in N, and S directions respectively.

4.2.2 Router and Port Directions

Routers with different degrees exist in the topology. Port directions of the router with a maximum degree are depicted in Figure 4.3. Port directions of a router are based on that router's degree and its placement in the topology. As an example, Table 4.1 summarizes the Port directions of a router in 8×8 topology, shown in Figure 4.2.

4.2.3 Topological Parameters

Any interconnect network's performance is determined by its topological characteristics. For the investigated topologies shown in Figure 4.4, several critical characteristics such as network diameter, bisection width, number of links, network cost, and packing density have been evaluated and compared.

Consider M×N ($M=2^P$, $N=2^Q$ where P, Q $\in$ N) network, a network having M rows and N columns. Each processing element has a router connected to it. All the routers in

Table 4.1 Port directions of the Routers in 8×8 DiamondMesh

Router Placement	Node Degree*	Port directions of the Router
Top left corner	2	E, S
Top right corner	3	W, S, SW
Bottom left corner	3	E, N, NE
Bottom right corner	2	W, N
Left edge- Middle, odd row	5	E, N, S, NE, SE
Left edge- Middle, even row	3	E, N, S
Right edge-Middle, odd row	3	W, N, S
Right edge-even row	5	W, N, S, NW, SW
Top Edge-Middle, odd column	5	E, W, S, SE, SW
Top Edge-Middle, even column	3	E, W, S
Bottom Edge-Middle, odd column	3	E, W, N
Bottom Edge-Middle, even column	5	E, W, N, NW, NE
Middle, odd row, odd column	4	E, W, N, S
Middle, even row, even column		
Middle, odd row, even column	8	E, W, N, S, NE, NW, SE, SW
Middle, even row, odd column		

Node Degree* (number of ports of the router), here refers to the degree of the router excluding the port that connects the router to the Processing Element.
Top row is referred as Row0 (even), Left edge is referred as Column0 (even)

the network are interconnected in accordance with the topology chosen. Topological parameters of the proposed DiamondMesh topology have been explored and compared with those of the baseline Mesh, fully connected diagonal mesh topology-DMesh, an extended diagonal Mesh topology-XDMesh and ZMesh topology and is presented in Table 4.2.

Figure 4.5 shows the comparison of topological parameters. As can be seen in Figure 4.5(a), the network diameter of DMesh, DiamondMesh, and XDMesh is small and nearly identical for all network sizes whereas the network diameter of Mesh and ZMesh is large. DiamondMesh network diameter is 33.34% and 49.21% less than the network diameter of Mesh topology for 16 and 4096 node networks respectively. Figure 4.5(b) compares the bisection width of the topologies. Out of all the topologies, DMesh has the highest bisection width and Mesh has the lowest bisection width. The bisection width of XDMesh is close to that of Mesh for all the network sizes. The bisection width of DiamondMesh is in between DMesh and Mesh and is the same as that of ZMesh. Here, the average node degree is considered to estimate the network cost accurately as different nodes have different node degrees. From Figure 4.5(c), Mesh has the lowest average node degree among all the considered topologies. The average node degree of XDMesh is close to that of Mesh.

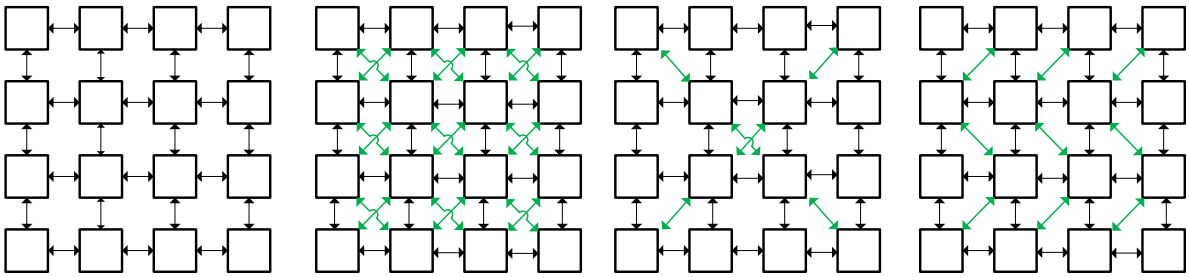


Figure 4.4 (a) Mesh; (b) DMesh; (c) XDMesh; (d) ZMesh

The average node degree of DiamondMesh and ZMesh is same and is lower than that of DMesh. DMesh has the highest average node degree because of more number of diagonal links. From Figure 4.5(d), 4.5(e), one can observe that the network cost and packing density of XDMesh are better when compared to all other topologies. This may indicate that XDMesh is area-efficient. However, considering other network parameters, XDMesh results in poorer performance in terms of latency and throughput when compared to that of DMesh and DiamondMesh. Next to XDMesh, DiamondMesh has minimum network cost and maximum packing density. So, for latency-critical applications, DMesh is the best but at the expense of area and power. DiamondMesh appears to be more area and power-efficient than DMesh. DiamondMesh is expected to outperform all other topologies except DMesh in terms of latency and throughput. Thus, DiamondMesh emerges as a suitable and optimized topology in terms of area, power, latency, and throughput. The following section summarizes the simulation findings. The results corroborate the topological parameter-based study.

4.3 Evaluation

4.3.1 Experimental Setup

BookSim2.0 simulator [198] has been employed to simulate and evaluate DiamondMesh and other potential topologies. Power and area metrics have been evaluated using the orion area-power model [199] that is included in the BookSim simulator. Since the BookSim simulator provides area and power models along with cycle-accurate simulation, and has been used to evaluate existing diagonal mesh topologies like ZMesh, it was chosen

Table 4.2 Topological Parameters

Topology	Network diameter	Bisection width	Number of edges (E)	Maximum node degree
Mesh	$M+N-2$	$\min(M,N)$	$2MN-M-N$	5
DMesh	$M+N-2-\min\{(M-1),(N-1)\}$	$3\times\min(M,N)-2$	$4MN-3M-3N+2$	9
XDMesh	$M-1 \quad or \quad N-1$	$\min(M,N)+2$	$2MN-2$	7
ZMesh	$M+N-2-\min\{(M-1)/2,(N-1)\}$	$2\times\min(M,N)-1$	$3MN-2M-2N+1$	7
DiamondMesh	$M \quad or \quad N \quad if \quad M=N$ $\max(M,N)-1$ if M≠N	$2\times\min(M,N)-1$	$3MN-2M-2N+1$	9

Number of Routers=Number of Processing elements=MN
Each edge has two unidirectional Links. In the table, the edges that interconnect the routers are only considered.
Total number of edges in the network= E+MN
Total number of Links in the network= 2 X (E+MN)
Node refers to Processing element or Router.
Average node degree = (2E+MN)/MN
For XDMesh topology, M=N

to evaluate DiamondMesh. For 16- and 64-node networks, the latency and power performance of each of the topologies under consideration has been examined using a variety of synthetic traffic patterns, including Uniform, Shuffle, and Bit Reversal as well as real application workloads. Area metrics have also been evaluated for the topologies. Table 4.3 shows the network configuration parameters.

4.3.2 Routing Algorithm

For evaluating the topologies, dimensional order routing has been considered. To evaluate Mesh topology, XY dimensional order routing has been considered. To evaluate diagonal link-based topologies including DMesh, XDMesh, ZMesh, and DiamondMesh, DXY dimensional order routing has been considered. Algorithm 1 presents pseudo code for DXY Dimensional Order Routing (DOR). Consider a source(src) and destination(dst) pair where src represents the source node number and dst represents the destination node number.

Algorithm 1 Pseudo Code for DXY DOR

Input: Current Source id (src) and Destination id (dst)
Output: Router to which packet has to be routed
Process:

```
if src==dst then
    Eject
else if src<dst then
    if src and dst are in the same column then
        Route packet towards South
    else if column of src<column of dst then
        if src and dst are in the same row then
            Route packet towards East
        else
            if SE link exists then
                Route packet towards SE
            else
                Route packet towards East
    else
        if SW link exists then
            Route packet towards SW
        else
            Route packet towards West
else
    if src and dst are in the same column then
        Route packet towards North
    else if column of src<column of dst then
        if src and dst are in the same row then
            Route packet towards West
        else
            if NE link exists then
                Route packet towards NE
            else
                Route packet towards East
    else
        if NW link exists then
            Route packet towards NW
        else
            Route packet towards West
```

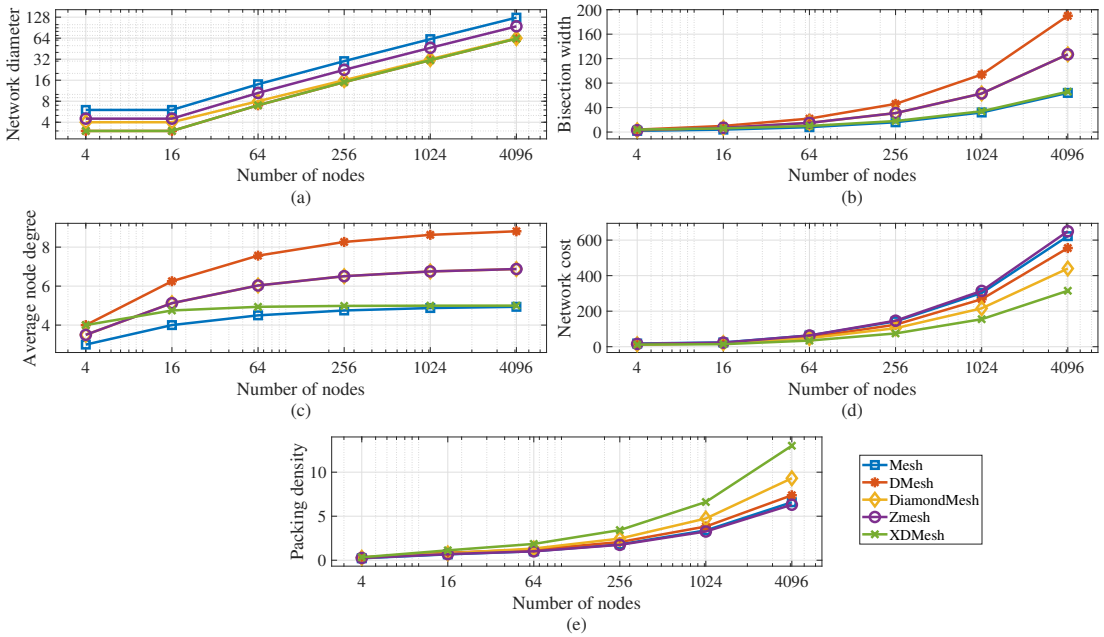


Figure 4.5 Graphical comparison of topological parameters of the considered topologies; (a)Network Diameter; (b)Bisection Width; (c)Average Node Degree; (d)Network Cost; (e) Packing Density

4.3.3 Performance Analysis for synthetic traffic

In this section, simulation results of the considered topologies have been compared and a detailed analysis of the results has been performed. The metrics considered to evaluate and compare the topologies include average packet latency (APL), total network power (TNP), power-latency product (PLP), i.e., PLP is the product of APL and TNP, average accepted flit rate (AAFR) i.e., throughput, average hop count, and area.

Figure 4.6(a) compares the average packet latency of the considered topologies for a 16-node network at various injection rates under uniform traffic patterns. From the results, it can be observed that the latency values for DiamondMesh topology are closer to that of DMesh topology for injection rates less than 0.6. The latency values for DiamondMesh have been consistently lower than that of the latency values obtained for all other topologies at all injection rates. Figure 4.6(b) compares the total network power of all the topologies at various injection rates. Out of all the considered topologies, DMesh has exhibited the highest TNP values because of more number of links and Mesh has exhibited lowest TNP values. TNP values of DiamondMesh are in between DMesh and Mesh. Also, one may observe that the TNP of DiamondMesh is almost very close

Table 4.3 Network configuration parameters

Parameter	Value
Network Size	16 Nodes (4×4) 64 Nodes (8×8)
Number of Virtual Channels	2
VC Buffer Size	4
Traffic Pattern	Uniform, Transpose, Shuffle, Bit Reversal
Routing Function	Dimension-Order Routing XY for Mesh and DXY for other topologies
Packet Size	4 Flits
Channel Width	128 bits
Router Frequency	2GHz
Process Technology	32nm

to that of ZMesh and XDMesh for injection rates lower than 0.6. Another parameter Power-Latency-product (PLP) has been calculated and compared for all the topologies. Power-Latency product can be considered as a figure of merit as it indicates a trade-off between power and latency. Figure 4.6(c) compares PLP of all the considered topologies. The results have inferred that for injection rates less than 0.5, DiamondMesh has exhibited the lowest PLP. For lower injection rates, Mesh, ZMesh, and XDMesh also have lower PLP values and are closer to that of DiamondMesh, and DMesh topologies. However with an increase in the injection rates, PLP values of these topologies have increased drastically. At higher injection rates, the PLP values of DiamondMesh have been slightly higher than that of DMesh. However, the values have been consistently lower than those of Mesh, ZMesh, and XDMesh for all injection rates. Figure 4.6(d) shows the throughput characteristics of the considered topologies. DMesh topology has the maximum throughput whereas ZMesh has the minimum throughput among all the compared topologies. The throughput characteristics of DiamondMesh are almost close to that of DMesh. The saturation throughput of DiamondMesh has been higher than that of all other topologies except DMesh. The saturation throughput of DMesh has been slightly higher than that of DiamondMesh but at the expense of more number of links. Overall, for a 16-node network with identical configuration, at 0.3 injection rate, DiamondMesh has shown a reduction of 13% in APL, a reduction of 5% in PLP compared to Mesh. DiamondMesh has shown a reduction of 9.5% in APL, a reduction of 13% in PLP compared to ZMesh. DiamondMesh has shown a reduction of 7% in APL, a reduction of 5% in PLP compared

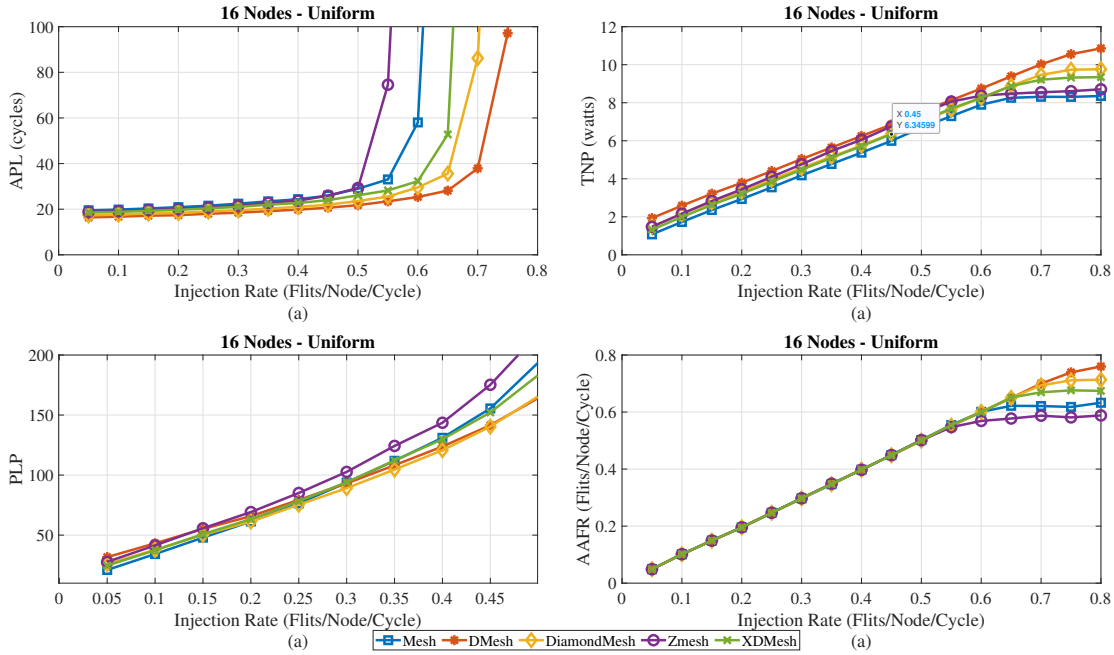


Figure 4.6 Comparison of 16 node network simulated under uniform traffic pattern.

to XDMesh. DiamondMesh has shown a reduction of 5% in PLP but an increase of 5% in APL compared to DMesh. One can observe that the saturation throughput is 0.63, 0.75, 0.7, 0.58 and 0.67 for Mesh, DMesh, DiamondMesh, ZMesh and XDMesh respectively.

Figure 4.7(a) compares the average packet latency of the considered topologies for a 64 node network at various injection rates under uniform traffic pattern. From the results, it can be observed that the latency values for DiamondMesh topology are closer to that of DMesh topology for injection rates less than 0.3. The latency values for DiamondMesh have been consistently lower than that of the latency values obtained for all other topologies at all injection rates. Figure 4.7(b) compares total network power of all the topologies at various injection rates. Out of all the evaluated topologies, DMesh has exhibited highest TNP values because of more number of links and Mesh has exhibited lowest TNP values. TNP values of DiamondMesh are in between DMesh and Mesh. A similar trend has been observed in 16 node network. Figure 4.7(c) compares PLP of all the considered topologies. The results have inferred that for injection rates less than 0.3, DiamondMesh has exhibited lowest PLP values. For injection rates higher than 0.3, DMesh topology has the lowest PLP values. The PLP values of DiamondMesh have been consistently lower than that of Mesh, ZMesh and XDMesh for all injection rates. Figure 4.7(d) shows the throughput characteristics of the considered topologies. DMesh topology

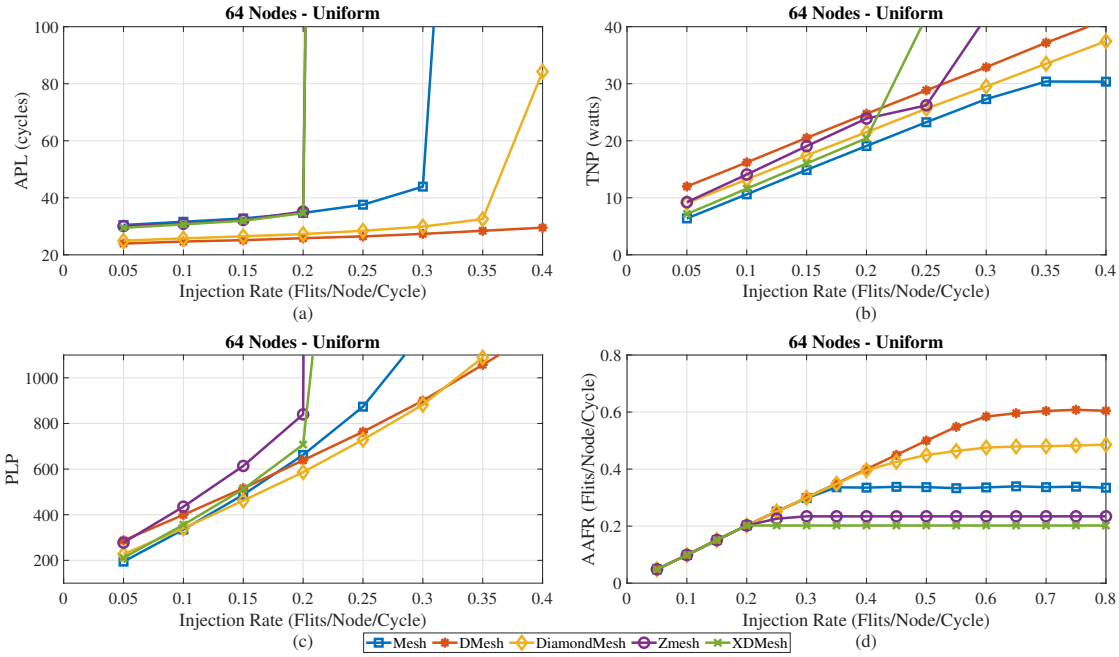


Figure 4.7 Comparison of 64 node network simulated under uniform traffic pattern

has the maximum throughput. Throughput characteristics of DiamondMesh are almost close to that of DMesh. All other topologies have exhibited very low throughput values. Overall,for 64-node network simulated under uniform traffic, the saturation throughput is 0.33, 0.6, 0.5, 0.23 and 0.20 for Mesh, DMesh, DiamondMesh, ZMesh and XDMesh respectively. Beyond, 0.2 the network has been saturated for ZMesh and XDMesh topologies, as such the metrics have been compared at 0.2 injection rates. At 0.2 injection rate, DiamondMesh has shown a reduction of 21%, 22% and 21% in APL respectively compared to that of Mesh, ZMesh and XDMesh whereas 5% increase in APL compared to DMesh. DiamondMesh has shown a reduction of 11%, 8% 30% and 17% in PLP respectively compared to that of Mesh, DMesh, ZMesh and XDMesh.

The considered topologies have been simulated under uniform traffic by varying injection rates from 0.01 to 0.08 to analyze the performance at low traffic for 64 64-node network. The results from Figure 4.8(a) have inferred that the average packet latency of DiamondMesh is considerably lower than that of all other topologies except DMesh. The latency characteristics of DiamondMesh are close to that of DMesh. From Figure 4.8(b), 4.8(c), DMesh has maximum Power and Power-Latency product values whereas Mesh has the minimum values. DiamondMesh Power and Power-Latency product characteristics are intermediate to those of DMesh and Mesh. The power-Latency product values of

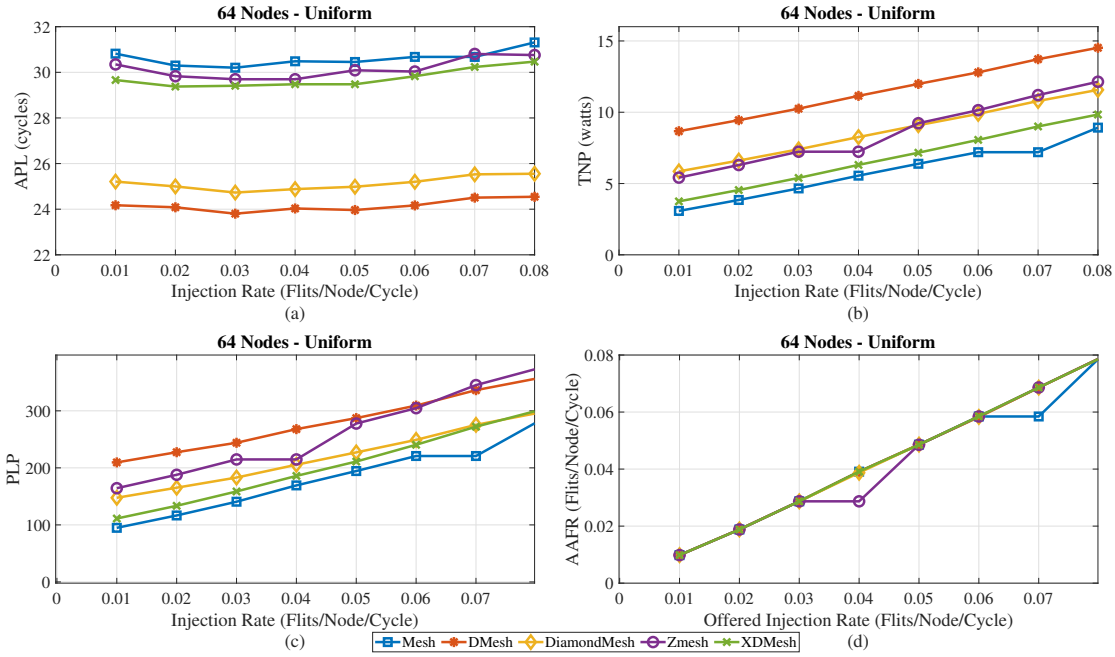


Figure 4.8 Comparison of 64 node network simulated under uniform traffic pattern for low injection rates

DiamondMesh are close to those of XDMesh and lower than those of ZMesh. From Figure 4.8(d), one can observe that all the considered topologies achieve maximum throughput at all injection rates except that Mesh and ZMesh have deviations at intermediate injection rates. DiamondMesh has shown an average reduction of 18%, 16%, and 15% in APL compared to that of Mesh, ZMesh, and XDMesh respectively. DiamondMesh has shown an average reduction of 22% and 15% in PLP compared to that of DMesh and ZMesh respectively.

Figure 4.9, Figure 4.10, Figure 4.11 and Figure 4.12 compare average packet latency, total network power, power-latency product and throughput of the considered topologies for 16, 64 nodes network under shuffle traffic pattern and bit reversal traffic pattern. The results for 16, 64 node network simulated under shuffle, bit reversal traffic patterns have exhibited a similar trend as that of 16, 64 node network simulated under uniform traffic pattern. Result values may be differing but the characteristics trend is similar in a way that the average packet latency of DiamondMesh has been consistently lower than that of Mesh, and ZMesh topologies and closer to that of DMesh topology. The total network power of DiamondMesh has been always intermediate to that of DMesh which consumes maximum power and Mesh which consumes minimum power. The total network power

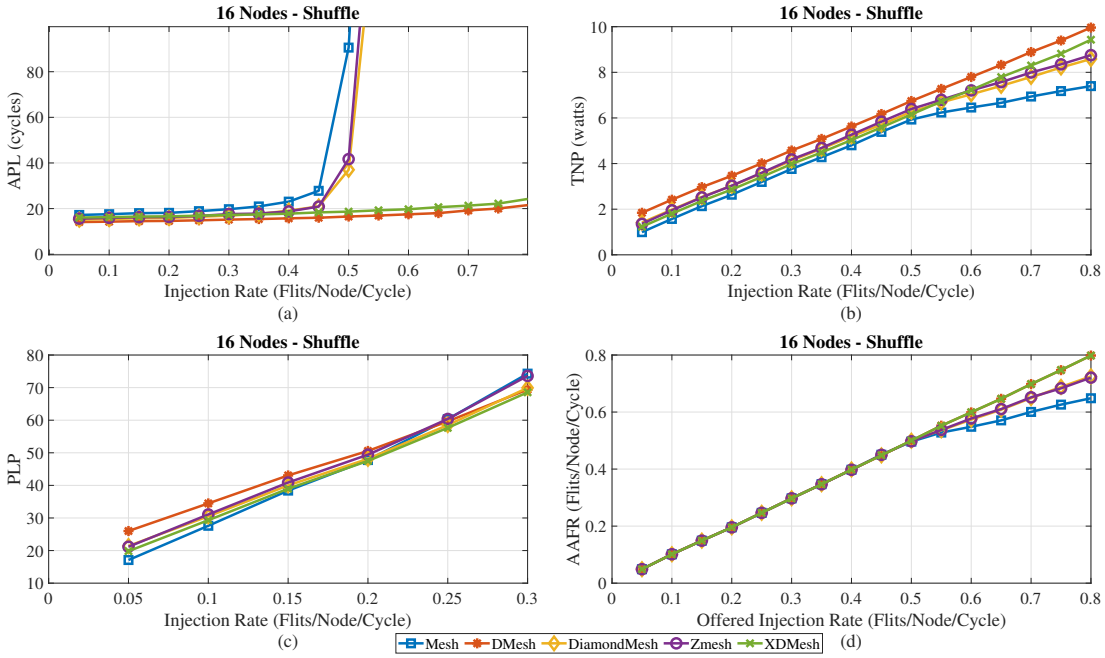


Figure 4.9 Comparison of 16 node network simulated under shuffle traffic pattern

of DiamondMesh is close to that of ZMesh and XDMesh. The Power-Latency Product of DiamondMesh has been always better than that of Mesh, ZMesh, and XDMesh topologies. For lower injection rates, the Power-Latency-Product of DiamondMesh has been lower than that of DMesh. The throughput of DiamondMesh has been always better than that of Mesh, and ZMesh topologies and it is close to that of DMesh.

Figure 4.13 shows a comparison of the average hop count of 16 and 64 node networks simulated under uniform traffic. Average hop count is calculated as a mean value of all the values of hop count obtained for different injection rates. The average hop count of DiamondMesh is very close to that of DMesh for both network sizes. The average hop count of DiamondMesh is lower than that of all other topologies.

Area metrics have also been evaluated and compared with that of the area obtained for other topologies. Table 4.4 presents the area occupied by the considered topologies for both 16-node and 64-node networks. The channel area represents the area occupied by the links. The switch area represents the area occupied by the routers. The total area represents the channel area, switch area, and other resources like buffers area. Area is directly proportional to the number of links in the topology. So, the topology with more number of links and the highest average node degree occupies more area. As can be seen in Table 4.4, DMesh has the maximum area as it incorporates more links and Mesh has

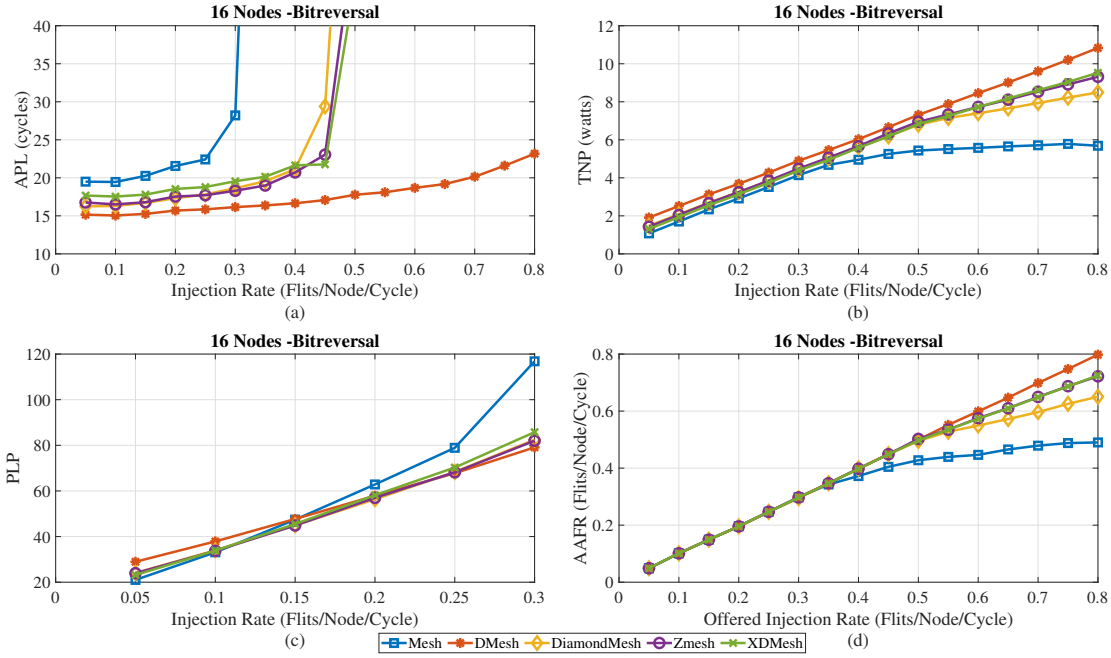


Figure 4.10 Comparison of 16 node network simulated under bit reversal traffic pattern

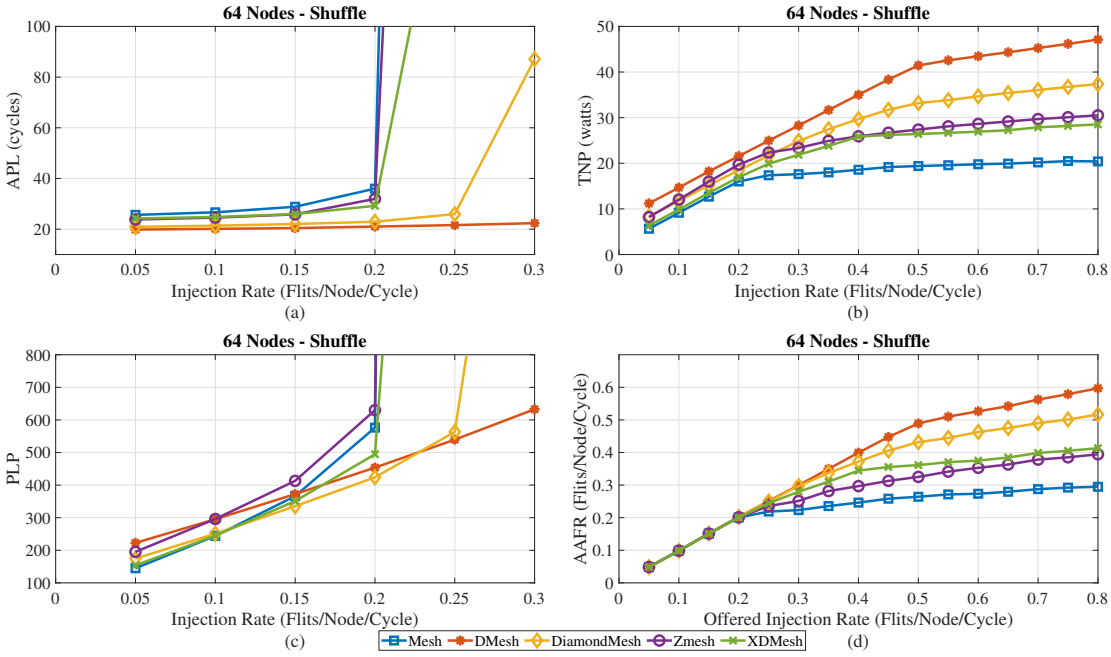


Figure 4.11 Comparison of 64 node network simulated under shuffle traffic pattern

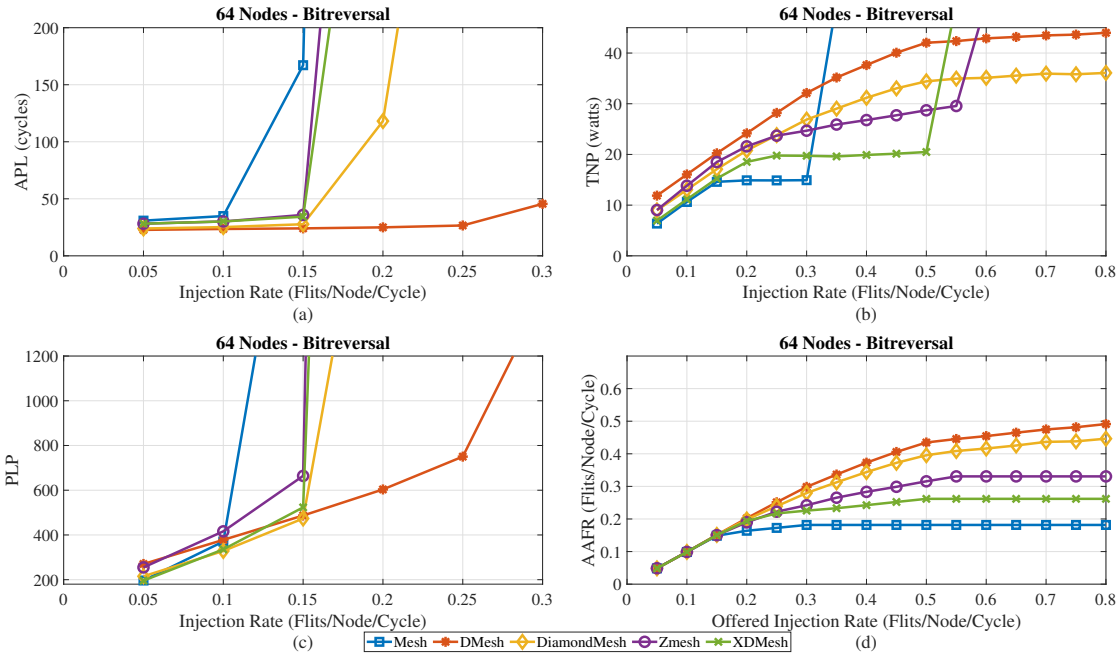


Figure 4.12 Comparison of 64 node network simulated under bit reversal traffic pattern

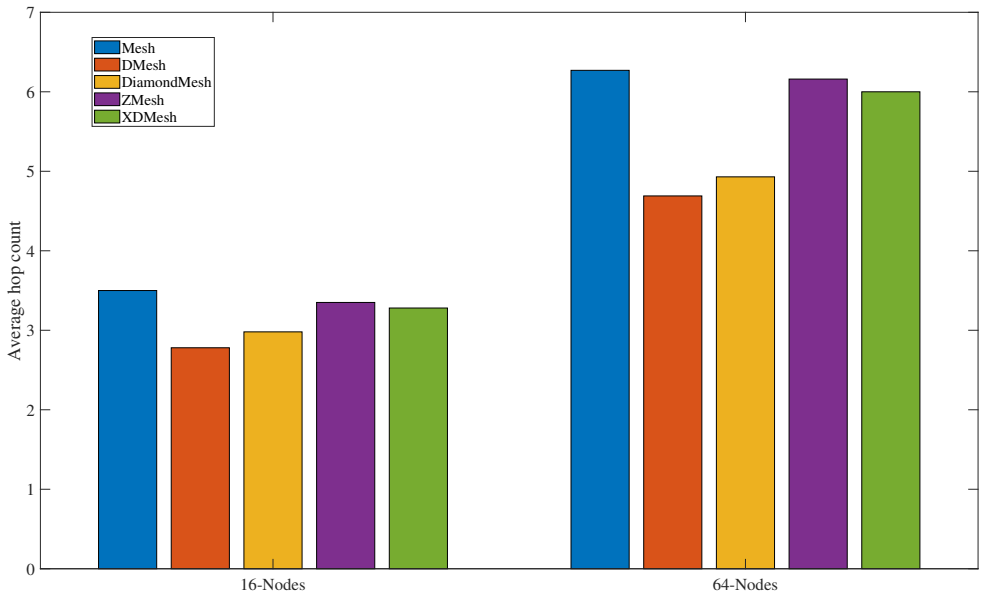


Figure 4.13 Comparison of Average Hop Count of 16 and 64 node networks simulated under uniform traffic

the minimum area. The area of DiamondMesh is intermediate to Mesh and DMesh. A reduction of 28% and 30% of the area occupied by DiamondMesh compared to that of DMesh has been noted for 16-node and 64-node networks respectively.

Table 4.4 Comparison of Area occupied by considered topologies

	Channel Area (mm^2)		Switch Area (mm^2)		Total Area (mm^2)	
	4×4	8×8	4×4	8×8	4×4	8×8
Mesh	0.0115343	0.0507511	0.11073	0.553648	0.149108	0.725195
XDMesh	0.0133117	0.0548982	0.162739	0.686188	0.207927	0.873626
DMesh	0.0168663	0.0797809	0.283535	1.61229	0.342344	1.89508
ZMesh	0.0142003	0.065266	0.188744	1.01754	0.237337	1.2447
DiamondMesh	0.0142003	0.065266	0.197132	1.08297	0.245726	1.31014

4.3.4 Performance Analysis for real application workloads

Booksim simulator integrated with netrace [200] has been employed to evaluate the proposed and considered topologies under real application workloads of PARSEC benchmark suite [200]. Netrace is a trace-based NOC evaluation methodology that captures and enforces the dependencies between packets from a full system. PARSEC benchmark suite contains many emerging applications from different areas. These applications include blackscholes, bodytrack, canneal, dedup, ferret, fluidanimate, swaptions, vips, and x264 related to financial, body tracking of a person, optimizing routing cost of a chip design, next-generation compression with data deduplication, content similarity search server, Pricing of a portfolio of swaptions, Image processing, and H.264 video encoding respectively. Region 2, the PARSEC-defined region of interest (ROI) which represents the parallel portion of the application has been considered for the evaluation. One billion cycles from each application program have been considered for evaluation. All the applications are executed on 64-node networks.

Figure 4.14, compares the normalised APL of the considered topologies for a 64-node network simulated under real application workloads. The results have inferred that the latency characteristics of DiamondMesh are close to that of DMesh and have outperformed Mesh network for all the workloads.

Figure 4.15 shows the comparison of the normalised power density of all the considered topologies simulated under real application workloads. Power density is the ratio of

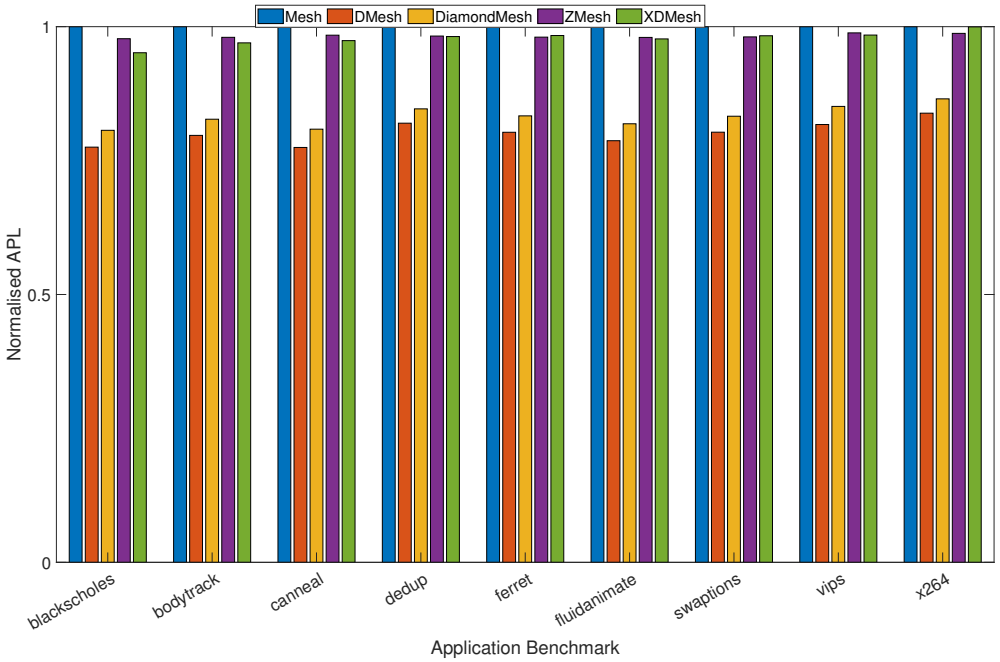


Figure 4.14 Comparison of normalised average packetlLatency (APL) 64 node network simulated under real application workloads

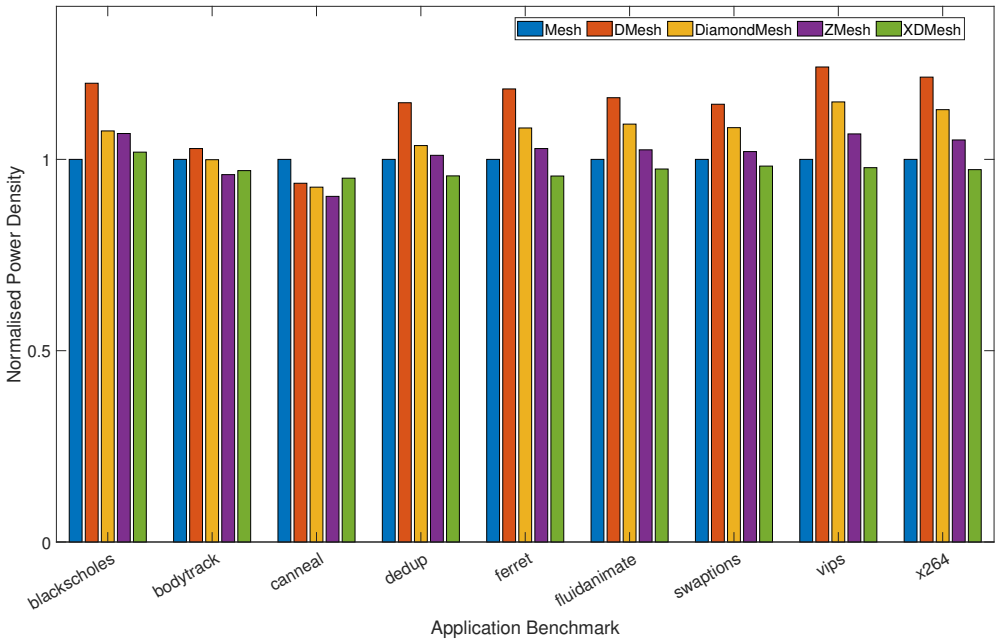


Figure 4.15 Comparison of normalised power density 64 node network simulated under real application workloads

the total network power to the total network area. It is an important metrics in assessing the thermal behaviour of the network. From the figure, it can be observed that the power density of DiamondMesh is close to that of Mesh and is lesser than that of DMesh. Higher power density may lead to higher temperatures thereby deteriorating the power integrity. DiamondMesh topology facilitates a reduction in latency over Mesh topology, simultaneously maintaining an equivalent power density as that of Mesh. The trend of TNP and PLP characteristics are similar to that of the trend observed with power density characteristics. i.e; TNP and PLP of DiamondMesh are intermediate to that of DMesh (maximum) and Mesh(minimum). This is in accordance with the results that are verified under synthetic traffic patterns for lower injection rates (Figure 3.13). As such the plots for TNP and PLP are omitted. However, with the benefits of a smaller network diameter and higher bisection width over Mesh, DiamondMesh with efficient, adaptive routing schemes may result in reduced power. This requires further investigation.

4.3.5 Synthesis results

The Mesh, DMesh, DiamondMesh, ZMesh and XDMesh topologies have been synthesized on Xilinx vivado design compiler [197] with Virtex7(VC709) as the target device. Virtual channel input buffered router with two virtual channels (VC), each VC with four buffers has been considered for implementing the topology. A packet size of 32 flits, each flit with 32 bits has been considered.

The synthesis results in Table 4.5 has inferred that DiamondMesh topology has consumed less space when compared with that of an identically configured DMesh topology. When a 16-node network is considered, DiamondMesh topology has utilized 18400 LUTs (Look Up Table), 24493 FFs(Flip-Flops) whereas DMesh has utilized 24042 LUTs, 29828 FFs. Overall, a reduction of 23.5% and 17.8% in the utilization of LUTs and FFs respectively for 16-node DiamondMesh topology compared to that of 16-node DMesh topology has been noted. The utilisation of DiamondMesh has not been much higher than that of the utilisation of ZMesh and XDMesh.

Table 4.5 Synthesis Results

Topology	Resource	Utilization	Available	Utilization%
Mesh	LUT	12916	433200	2.98
	FF	19166	866400	2.21
DMesh	LUT	24042	433200	5.54
	FF	29828	866400	3.44
DiamondMesh	LUT	18400	433200	4.24
	FF	24493	866400	2.82
ZMesh	LUT	18394	433200	4.24
	FF	24076	866400	2.77
XDMesh	LUT	16740	433200	3.86
	FF	22278	866400	2.57

Notes: LUT : Look Up Table, FF: Flip-Flop

4.4 3D-DiamondMesh

Network-on-Chips amalgamated with 3D integrated circuits (Ics) form 3D-NoCs that improve the performance, footprint, and scalability of the complex SoC architectures. Compared to 2D NoCs, 3D-NoCs have shorter links, lesser footprint, reduced latency, and power [201]. Furthermore, 3D Ics enable integration of CMOS circuits with heterogeneous technologies [202].

To evaluate the performance of larger architectures, DiamondMesh has been extended to 3D DiamondMesh and the performance under different synthetic traffic patterns for different network sizes have been analysed. Further, Different heterogeneous configurations have been proposed and evaluated. 2D-DiamondMesh layers have been stacked vertically and interconnected with TSVs (Through Silicon Vias) to form 3D DiamondMesh as depicted in Figure 4.16(a). 3D architectures of Mesh, XDMesh, ZMesh, and DMesh topologies under investigation are depicted in Figure 4.16 (b)-(e).

Heterogeneity has been introduced in a way that two alternate layers are of the same topology and the other two alternate layers are of a different topology. Five heterogeneous architectures shown in Figure 4.17 have been proposed and evaluated as part of the present work.

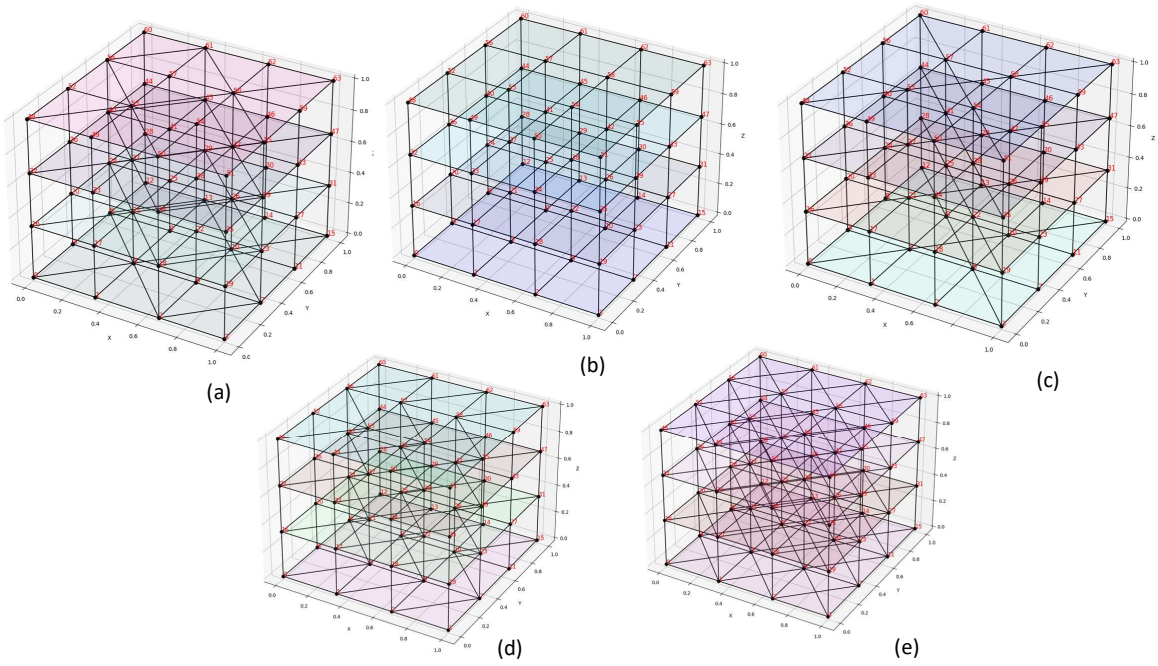


Figure 4.16 (a)3D DiamondMesh; (b)3D Mesh; (c) 3D XDMesh; (d) 3D ZMesh; (e) 3D DMesh

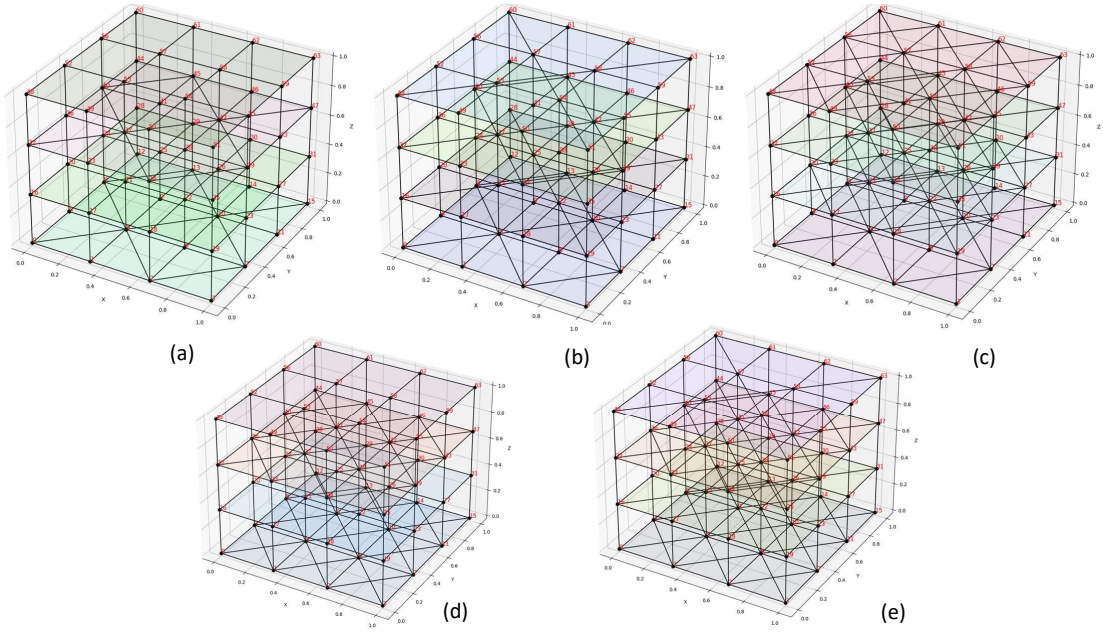


Figure 4.17 (a)DiamondMesh+Mesh; (b)DiamondMesh+XDMesh; (c) DiamondMesh+DMesh; (d) DMesh+Mesh; (e) DMesh+XDMesh

4.4.1 Evaluation

To carry out the simulation and performance evaluation of the DiamondMesh and other considered topologies, the Ratatoskr simulator [196] has been employed. The performance of the considered topologies has been evaluated under different synthetic traffic patterns that include uniform random, transpose, and bit-reversal for different network sizes. Network size is specified in terms of the total number of nodes and also as [X×Y×Z]. X denotes the number of nodes in the X-direction, Y denotes the number of nodes in the Y-direction and Z denotes the number of layers in the 3D-NoC architecture.

4.4.1.1 Methodology

Firstly, homogeneous 3D-NoC architectures have been simulated and the findings have been analysed for different network sizes. DXY routing in Algorithm 1 has been used to simulate 2D diagonal mesh-based topologies. For 3D-dagonal mesh-based topologies, DXY has been extended to the DXYZ routing algorithm to route the packets across the layers. Conventional XYZ algorithm has been used for 3D-Mesh topology. Pseudo code for DXYZ routing is shown in Algorithm 2. Secondly, the five proposed heterogeneous architectures have been simulated for 64-node and 256-node network sizes and the findings have been analysed.

Algorithm 2 Pseudo code for DXYZ routing

Input: Current Source id (src) and Destination id (dst)
Output: Router to which packet has to be routed
Process:
Step1: Follow DXY routing based on the X and Y co-ordinates of src and dst

Step2: Once the X and Y co-ordinates of src and dst are same then go to Step3 or else repeat Step1

Step3:
if *Z co-ordinate of dst is less than that of src* **then**
└ route DOWN
else
└ route UP

4.4.1.2 Performance analysis

Homogeneous architectures

Table 4.6 shows the configuration parameters to simulate the architectures. 2D, 3D-2 layer, and 3D-4 layer architectures with different network sizes from 16 nodes to a larger network size of 256 nodes have been simulated. The architectures are simulated under uniform, transpose, and bit-reversal traffic patterns.

Table 4.6 Configuration Parameters for Homogeneous configurations

Parameter	**value**
Simulation time	10000ns
No. of Virtual channels	2
Buffer depth	4
Router	Input-Buffered
Routing algorithm	DXYZ for Diagonal topologies XYZ for Mesh
Network Size ($X \times Y \times Z$)	16 nodes ($4 \times 4 \times 1$)
X – No. of nodes in X-direction	36 nodes ($6 \times 6 \times 1$)
Y - No. of nodes in Y-direction	64 nodes ($8 \times 8 \times 1$)
Z – No. of layers	32 nodes ($4 \times 4 \times 2$)
	72 nodes ($6 \times 6 \times 2$)
	128 nodes ($8 \times 8 \times 2$)
	64 nodes ($4 \times 4 \times 4$)
	144 nodes ($6 \times 6 \times 4$)
	256 nodes ($8 \times 8 \times 4$)

The simulation findings of Average Packet Latency (APL) for uniform traffic patterns have been plotted as shown in Figure 4.18 (a)-(i). From the plots, it can be observed that APL of DiamondMesh is intermediate to Mesh, ZMesh has maximum APL and DMesh has minimum APL among all other topologies. Except for DMesh, DiamondMesh surpasses Mesh, ZMesh, and XDMesh topologies in terms of APL for all network sizes. DMesh has shown reduced latency compared with DiamondMesh but at the expense of more number of links. The higher the number of links, the faster may be the performance but it leads to more area overhead and power consumption. It can be observed that the APL of ZMesh has been lower than that of Mesh up to 0.2 or 0.3 injection rates only. As the load increases, the APL of ZMesh increases and is higher than that of Mesh. APL

Table 4.7 Number of Links

	16 nodes ($4 \times 4 \times 1$)	36 nodes ($6 \times 6 \times 1$)	64 nodes ($8 \times 8 \times 1$)	32 nodes ($4 \times 4 \times 2$)	72 nodes ($6 \times 6 \times 2$)	128 nodes ($8 \times 8 \times 2$)	64 nodes ($4 \times 4 \times 4$)	144 nodes ($6 \times 6 \times 4$)	256 nodes ($8 \times 8 \times 4$)
Mesh	40	96	176	96	228	416	208	492	896
XD	46	106	190	108	248	444	232	532	952
ZMesh (or) DiamondMesh	49	121	225	114	278	514	244	592	1092
DMesh	58	146	274	132	328	612	280	692	1288

Table 4.8 Tradeoff between the number of links and APL

	16 nodes ($4 \times 4 \times 1$)	36 nodes ($6 \times 6 \times 1$)	64 nodes ($8 \times 8 \times 1$)	32 nodes ($4 \times 4 \times 2$)	72 nodes ($6 \times 6 \times 2$)	128 nodes ($8 \times 8 \times 2$)	64 nodes ($4 \times 4 \times 4$)	144 nodes ($6 \times 6 \times 4$)	256 nodes ($8 \times 8 \times 4$)
a	15.52	17.12	17.88	13.64	15.24	16.01	12.86	14.45	15.22
b	6.83	5.88	5.15	6.12	5.12	4.47	4.86	4.34	3.99
c	7.15	10.07	15.31	6.12	8.07	12.23	4.92	7.35	12.06

Notes:
a : %Reduction of No. of Links in DiamondMesh when compared to DMesh
b : % Increase of Latency for DiamondMesh compared to DMesh for injection rate=0.1
c : % Increase of Latency for DiamondMesh compared to DMesh for injection rate=0.8

of DiamondMesh has been found to be consistently lower than that of Mesh, ZMesh, and XDMesh and is closer to that of DMesh for all the loads and for all network sizes. Even though DiamondMesh and ZMesh have same number of links, DiamondMesh has performed superior to ZMesh because the network diameter of DiamondMesh has been smaller than that of ZMesh. This is because of the difference in the pattern in which the diagonal links are incorporated in DiamondMesh and ZMesh. A similar trend as that of uniform traffic patterns has been observed for transpose and bit reversal traffic patterns. As such, the plots for transpose and bit reversal traffic patterns have been omitted for brevity. Table 4.7 shows the number of links required to construct the topology. Table 4.8 shows the tradeoff between the number of links and APL for DiamondMesh and DMesh topologies.

It can be observed from the plots, at 0.8 injection rate, for 16-node and 256-node networks, there is 14.75% and 34.66% reduction of latency with DiamondMesh compared to that of an identically configured Mesh and there is 20.45% and 41.7% reduction of latency with DMesh compared to that of an identically configured Mesh. APL reduction is slightly better in DMesh compared to that of DiamondMesh but at the expense of more links. It can be observed from Table 4.8, that at 0.8 injection rate, for 16-node

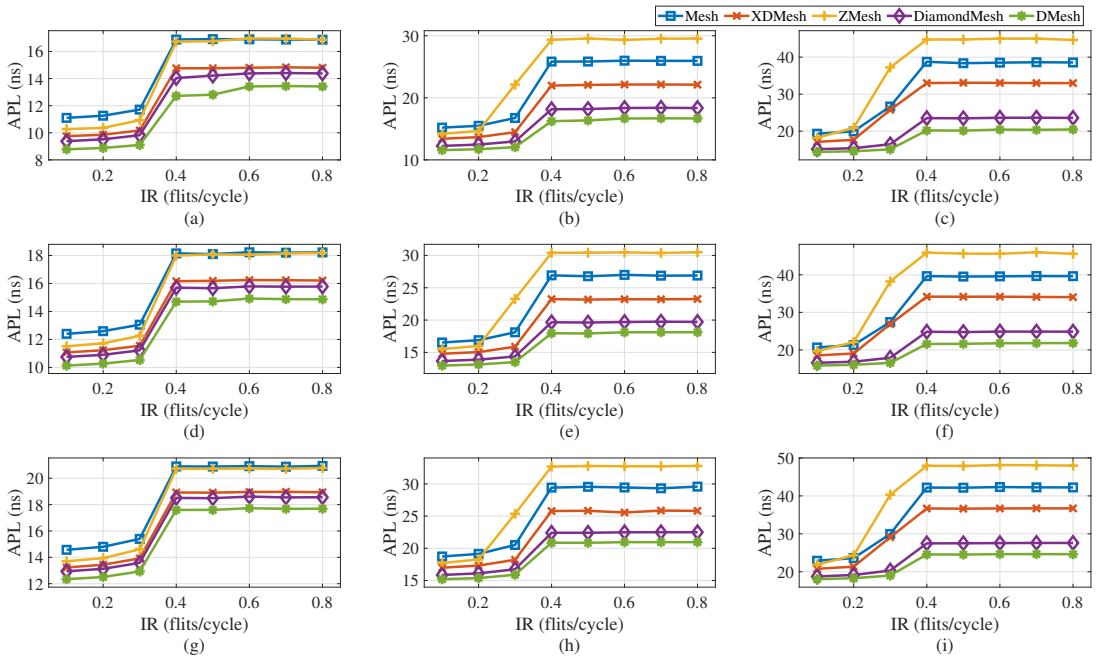


Figure 4.18 Average Packet Latency characteristics of the considered topologies (a)16 nodes(1L); (b)36 nodes(1L); (c)64 nodes(1L); (d)32 nodes(2L); (e)72 nodes(2L); (f)128 nodes(2L); (g)64 nodes (4L); (h)144 nodes (4L); (i)256 nodes (4L) 1L:1 Layer, 2L: 2 Layers, 4L: 4 Layers

and 256-node networks, the APL with DiamondMesh is 7.15% and 12.06% respectively higher than that of DMesh but with nearly 15% lesser number of links in DiamondMesh compared to that of DMesh.

It can be observed from the plots of 2D 64-node network and an identically configured 3D-4 layer 64-node network, 16 nodes per layer, that there is an average reduction of 40.18%, 39.07%, 38.21%, 48.44%, 19.14%, 13.49% in APL with 3D- Mesh, XDMesh, ZMesh, DiamondMesh, DMesh architectures respectively compared with their corresponding 2D architectures. The reduction in APL with 3D-Mesh, 3D-XDMesh, and 3D-ZMesh over their corresponding 2D architectures is higher compared to that of DiamondMesh and DMesh topologies. The reduction of APL with DiamondMesh compared to that of Mesh for a 2D-64node network is 22.81% whereas for a 3D-64 node network, it is only 11.28%. Similarly, the reduction of APL with DMesh compared to that of Mesh for a 2D-64 node network is 26.98% whereas for a 3D-64 node network, it is only 1.475%. Thus, with 3D-NoCs technology, the 2D architectures with larger network diameters can benefit more compared with that of the architectures with smaller network diameters. Precisely, it can be inferred that the 3D architectures outperform their corresponding identically

configured 2D architectures.

Heterogeneous architectures

Five heterogeneous architectures shown in Figure 4.17, formed by interconnecting the nodes in the alternate layers with two different topologies have been simulated under uniform traffic patterns. Table 4.9 shows the configuration parameters of the simulation. The simulation findings have been plotted as shown in Figure 4.19. An average reduction of 7.83%, 9.59%, 13.18%, 4.80%, 8.85%, 10.50%, 6.56% and 10.63% in APL can be observed with 4Layer-64node XDMesh, DiamondMesh, DMesh, DiamondMesh+Mesh, DiamondMesh+XDMesh, DiamondMesh+DMesh, DMesh+Mesh, DMesh+XDMesh architectures respectively over the conventional 4Layer-64node Mesh architecture. Similarly, an average reduction of 9.26%, 21.21%, 25.00%, 10.47%, 15.21%, 23.00%, 12.50% and 16.90% can be observed with 4Layer-256node XDMesh, DiamondMesh, DMesh, DiamondMesh+Mesh, DiamondMesh+XDMesh, DiamondMesh+DMesh, DMesh+Mesh, DMesh+XDMesh architectures respectively over the conventional 4Layer-256node Mesh architecture.

Table 4.9 Configuration Parameters for Heterogeneous configurations

Parameter	**value**
Simulation time	10000ns
No. of Virtual channels	2
Buffer depth	4
Router	Input-Buffered
Routing algorithm	DXYZ for Diagonal topologies XYZ for Mesh
Network Size ($X \times Y \times Z$) X – No. of nodes in X-direction Y - No. of nodes in Y-direction Z – No. of layers	64 nodes ($4 \times 4 \times 4$) 256 nodes ($8 \times 8 \times 4$)
Heterogeneous architectures (Topology1+Topology2) Topology1 for Layers '0' & '2' Topolog2 for Layers '1'& '3'	DiamondMesh+Mesh DiamondMesh+XDMesh DiamondMesh+DMesh DMesh+Mesh DMesh+XDMesh

It can be observed that DiamondMesh+XDMesh has a higher reduction in APL over Mesh when compared to homogeneous XDMesh over Mesh and also the reduction

in APL with DiamondMesh+XDMesh architecture is close to that of DiamondMesh with the benefit of a lesser number of links. Thus, heterogeneity in topology across the layers of 3D-NoC architecture could be a possible approach to bring a tradeoff among the key metrics like area, latency and power of the architecture. There could be several possible variations in the heterogeneity across the layers, out of which five architectures have been suggested and analysed in the present work.

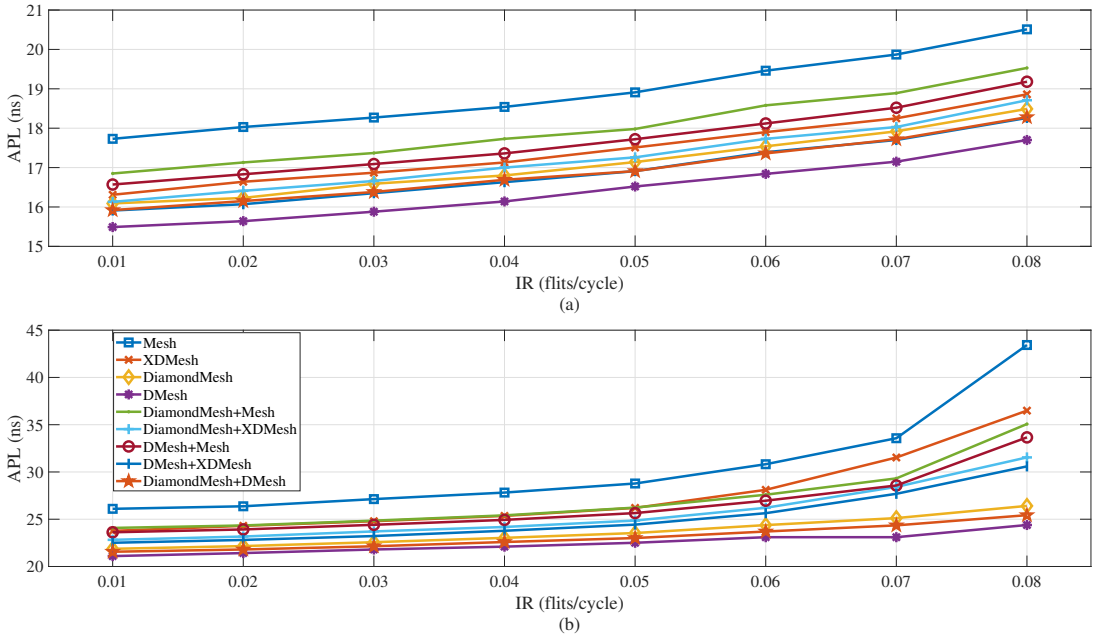


Figure 4.19 Average Packet Latency characteristics of heterogeneous architectures (a) 64 nodes(4L); (b) 256 nodes(4L)

4.5 Summary

In this chapter, a unique, energy-efficient diagonal mesh topology named DiamondMesh has been proposed. Topological parameters of the DiamondMesh topology have been evaluated and compared with other state-of-the-art diagonal mesh-based topologies. For 16 and 64-node networks, the considered topologies have been simulated under various synthetic traffic patterns and real-time benchmarks. The proposed topology and other considered topologies have been synthesised using xilinx vivado design compiler and the results have been analysed.

DiamondMesh outperforms Mesh, XDMesh, and ZMesh in latency, throughput, and

power-latency products. DiamondMesh has its latency and throughput characteristics close to that of DMesh accompanied by a significant reduction in area and power consumption. Notably, DiamondMesh occupies 28% less area for 16-node networks and 30% less for 64-node networks compared to DMesh. The power consumption of DiamondMesh exhibits an average reduction of 14% compared to DMesh. Additionally, 3D-DiamondMesh has also been evaluated and compared with the 3D architectures of Mesh, DMesh, ZMesh and XDMesh. The findings have inferred that the 3D architectures outperform their corresponding identically configured 2D architectures. Among all, 3D-DiamondMesh has demonstrated to be a more balanced topology providing a trade-off among power, latency and area. Further, five heterogeneous 3D architectures have been suggested and evaluated. The analysis suggests that introducing heterogeneity in topology across layers can effectively reduce latency with minimal area overhead or decrease area with a marginal impact on performance.

Precisely, DiamondMesh, 2D as well as 3D, proves to be a viable option for NoCs due to its optimal performance in latency, throughput, area, and power. The advantages of DiamondMesh, particularly in terms of the tradeoff between area, power, and latency, become more evident under heavy traffic loads as opposed to lighter traffic conditions.

In the following chapter, diverging from the content of chapters 3 and 4, the exploration will center on hybrid wireless interconnect topologies. While Chapters 3 and 4 concentrated on examining electrical wired Network-on-Chip (NoC) topologies capable of surpassing the conventional Mesh topology, Chapter 5 aims to delve into hybrid wired-wireless topologies capable of addressing the limitations of wired NoCs in complex architectures.

Chapter 5

Three level hierarchical hybrid wired-wireless NoC architecture

5.1 Introduction

Hybrid wireless NoC is one of the emerging NoC paradigms and is a promising solution to alleviate the scalability issues of traditional wired NoC which affect the performance of complex MPSoC architectures. Over the past few years, several hybrid hierarchical WiNoC architectures have been proposed. A hierarchical architecture with wired and wireless links (WLs) in which wired links are adapted for short range and WLs adapted as shortcuts between distant nodes for fast data transfer, will alleviate the challenges caused by the wired links.

In the existing literature on hybrid hierarchical wired-wireless NoC architectures, most of the hybrid hierarchical wired-wireless NoC architectures have been implemented as a two-level hierarchical architecture. Most of the studies have focused only on the subnet topology or wireless hubs placement and optimisation. The number of nodes to be connected to the central hub of each subnet is also an important design aspect of WiNoC architectures as it affects the performance as well as the area of the architecture. Very few studies exist in this aspect. Mostly, 256 node hybrid wired-wireless architectures have been studied and the studies have inferred that logical subnet with 16 nodes could be a viable option (i.e., for 256 nodes – 16 subnets with 16 subnet size is a viable option). However, with 16 nodes as the fixed subnet size, for large sized networks like 1024-node networks, the number of subnets would be 256, which is very high. Owing to these considerations,

in this present chapter, rather than adopting a fixed subnet size such as 16 nodes, the number of nodes in the subnets is varied. Further, the placement of nodes is also varied to form different subnet configurations thereby expanding the design space for exploration. The novelty of the proposed configurations resides in the manner in which the placement of hub nodes is varied within a subnet. Different hierarchical hybrid wired-wireless NoC architectures with varying number of subnets and varying number of nodes that are to be connected to the wireless radio hub of the respective subnet, have been proposed and analysed.

5.2 Design methodology

The proposed wired-wireless hierarchical NoC architectures have been implemented as three-level hierarchical architectures - Bottom, Middle, and Top levels. Firstly, all the nodes of the network are connected as a traditional wired mesh. Then, the wired mesh network is divided into smaller logical homogeneous regions called subnets. Each subnet has a central wireless radio hub (WR). Each WR is equipped with a wireless interface to facilitate wireless communication among the nodes. This forms the bottom level of the hierarchy of the architecture. Secondly, a fixed number of nodes within each subnet has been chosen and connected to the WR of the respective subnet. These nodes are referred to as hub nodes (HN). This forms the middle level of the hierarchy. Thirdly, all the WRs are fully connected i.e., each WR can communicate with every other WR via wireless channel(s). This forms the top level of the hierarchy. This allows for single-hop communication among distant nodes which eventually improves the performance of the network in terms of latency and throughput. The methodology to design the hybrid wired-wireless NoC architecture has been shown in Figure 5.1

Predominant research on the topology of the bottom-level and top-level hierarchy compared to the topology of the middle-level hierarchy has been noticed in previous studies. The number of nodes to be connected to the central WR of each subnet is also an important design aspect of WiNoC architectures as it affects the performance as well as the area of the architecture. As per the findings in [159], the highest efficiency of the subnets is obtained in 4×4 size. In some studies, keeping the subnet size fixed to

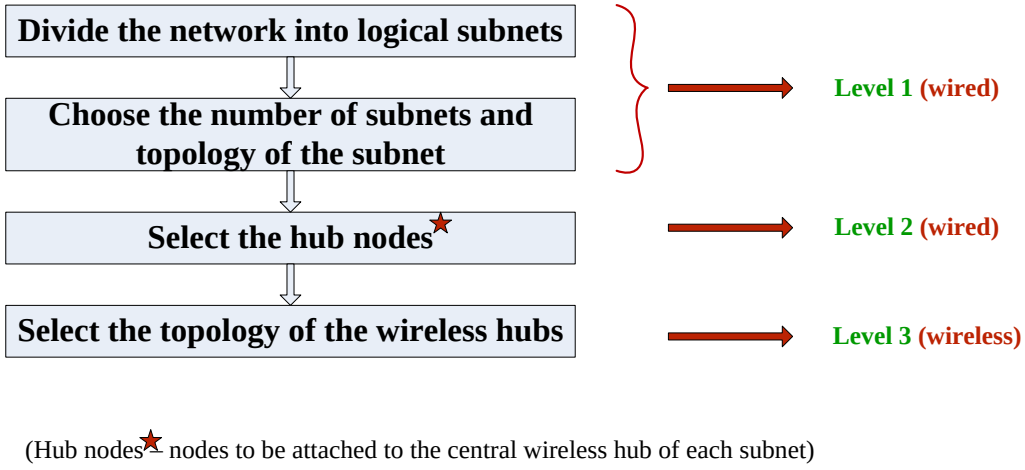


Figure 5.1 Methodology to construct hierarchical hybrid wired-wireless NoC

16 nodes, all the available nodes within a subnet are directly connected to the hub of that subnet. In that case, 16 wired links will be required. Consequently, the area of the links will increase. So, in some studies, to reduce the number of links, only four nodes are selected among all the available nodes within each subnet, to communicate with the related hub directly. The size and number of subnets have to be chosen such that neither the subnets nor the number of hubs of the top-level hierarchy becomes too large. If the subnet level or the hubs level of the hierarchy becomes too large, it causes a performance bottleneck by limiting the data throughput in that level [159]. Thus, for a 1024-node network, if the subnet size is fixed to 16 nodes, the number of subnets will be 64, making the top level too large. In this study, the number of subnets and the number of nodes sharing each WR of the subnet have been varied to form different architectures. These different architectures have been simulated and analysed for different network sizes under various traffic patterns and traffic loads.

5.3 Proposed architectures

In the present work, the mesh topology and fully connected wireless topology have been considered for the bottom level and top level of the hierarchical structure. In the middle level, the number of subnets, and the number and position of hub nodes connected to the wireless radio-hub within the subnet have been varied to form different hybrid

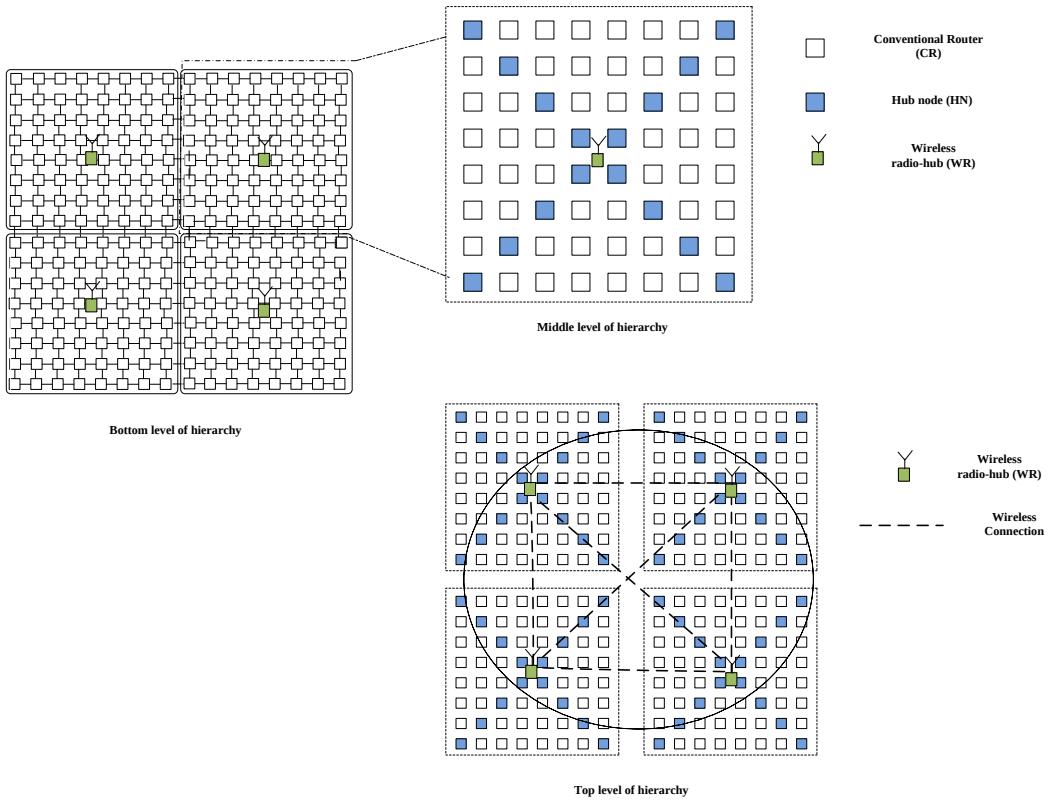


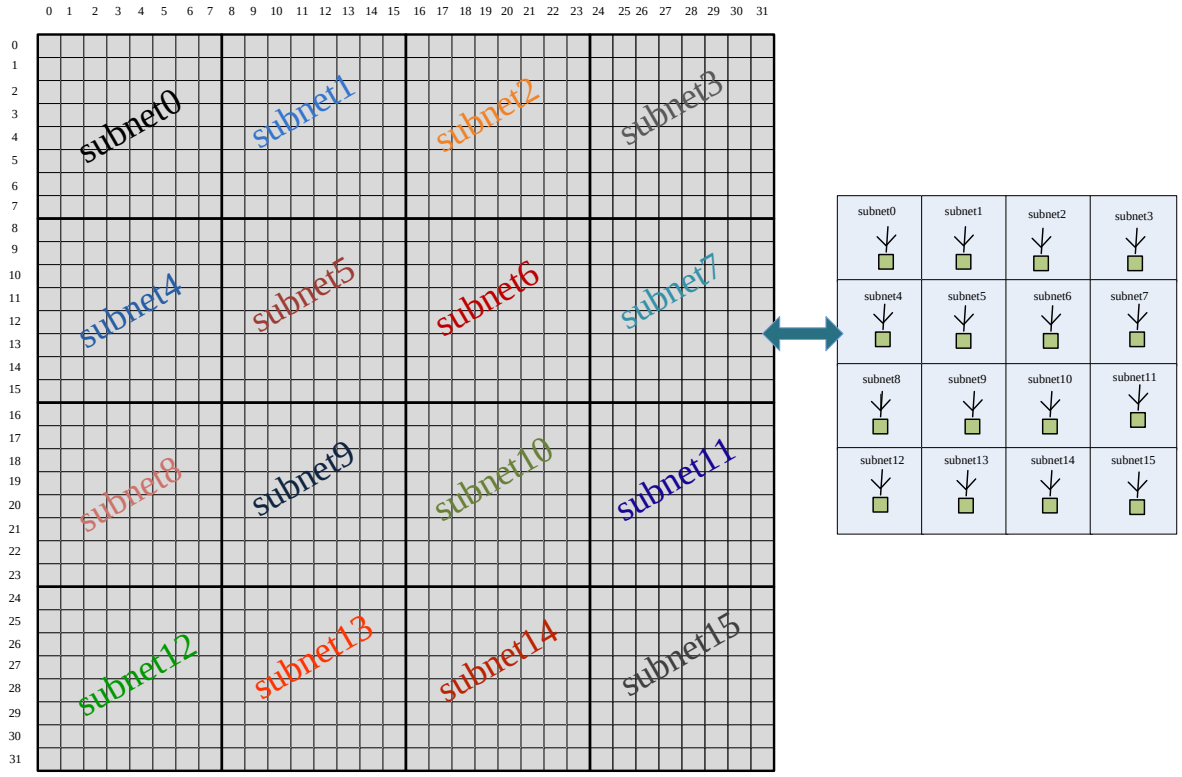
Figure 5.2 256_node hierarchical hybrid wired-wireless NoC

wired-wireless architectures. A heuristic approach has been followed to select the hub nodes of a subnet.

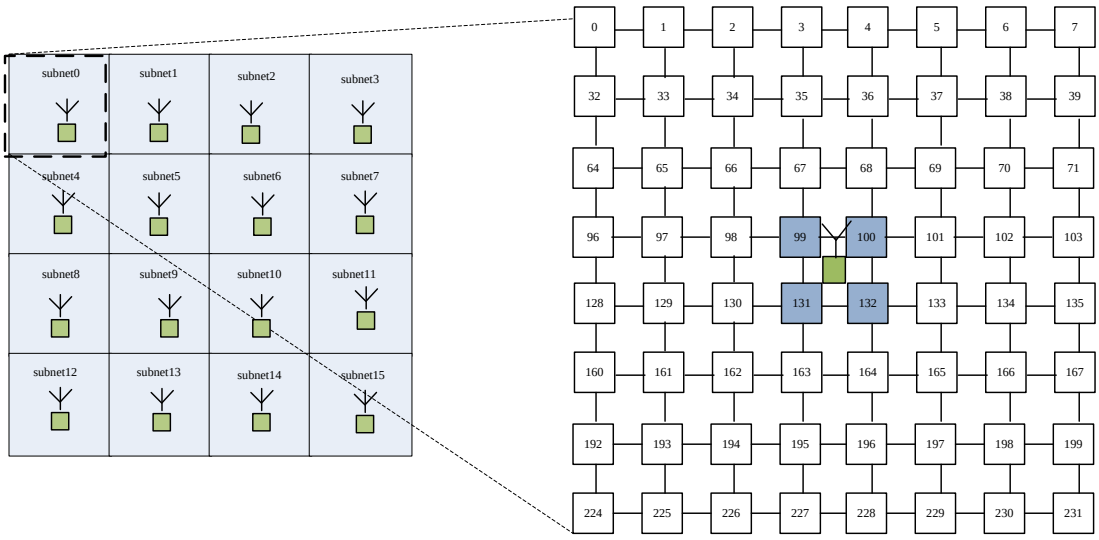
Figure 5.2 illustrates the three-level hierarchical hybrid wired-wireless NoC for a 256-node network and provides the details of the hierarchical architecture and the terminology used in the study (this figure is to illustrate the three-level hierarchy only, it is not meant to show the proposed architectures). For the middle-level and top-level hierarchy, the wired connections are omitted for the sake of clarity of the figure. The HN's of each subnet has to be wire connected (not shown in the figure) to the WR of the respective subnet.

Figure 5.3(a) and 5.3(b) show 1024-node bottom and middle-level hierarchy respectively. Top-level hierarchy i.e., the connection of WRs has been chosen as fully connected WRs in the present work. The top-level hierarchy for a 256-node network with 4 subnets is shown in Figure 5.2. Similarly, for a different network size, with the number of subnets varying, the topology of the top-level hierarchy is fixed as fully connected i.e., the WR of each subnet is connected to every other WR of the subnets. As the bottom and top levels of the hierarchy are fixed throughout the study, only the middle level of the hierarchy is

depicted in Figure 5.4. The middle level of the proposed hierarchical hybrid wired-wireless architectures is shown in Figure 5.4 and is produced by varying the number of subnets and the number of nodes that are connected to WR of each subnet.



(a) 1024-node-Bottom Level



(b) 1024-node-Middle Level

Figure 5.3 1024-node with 16 subnets

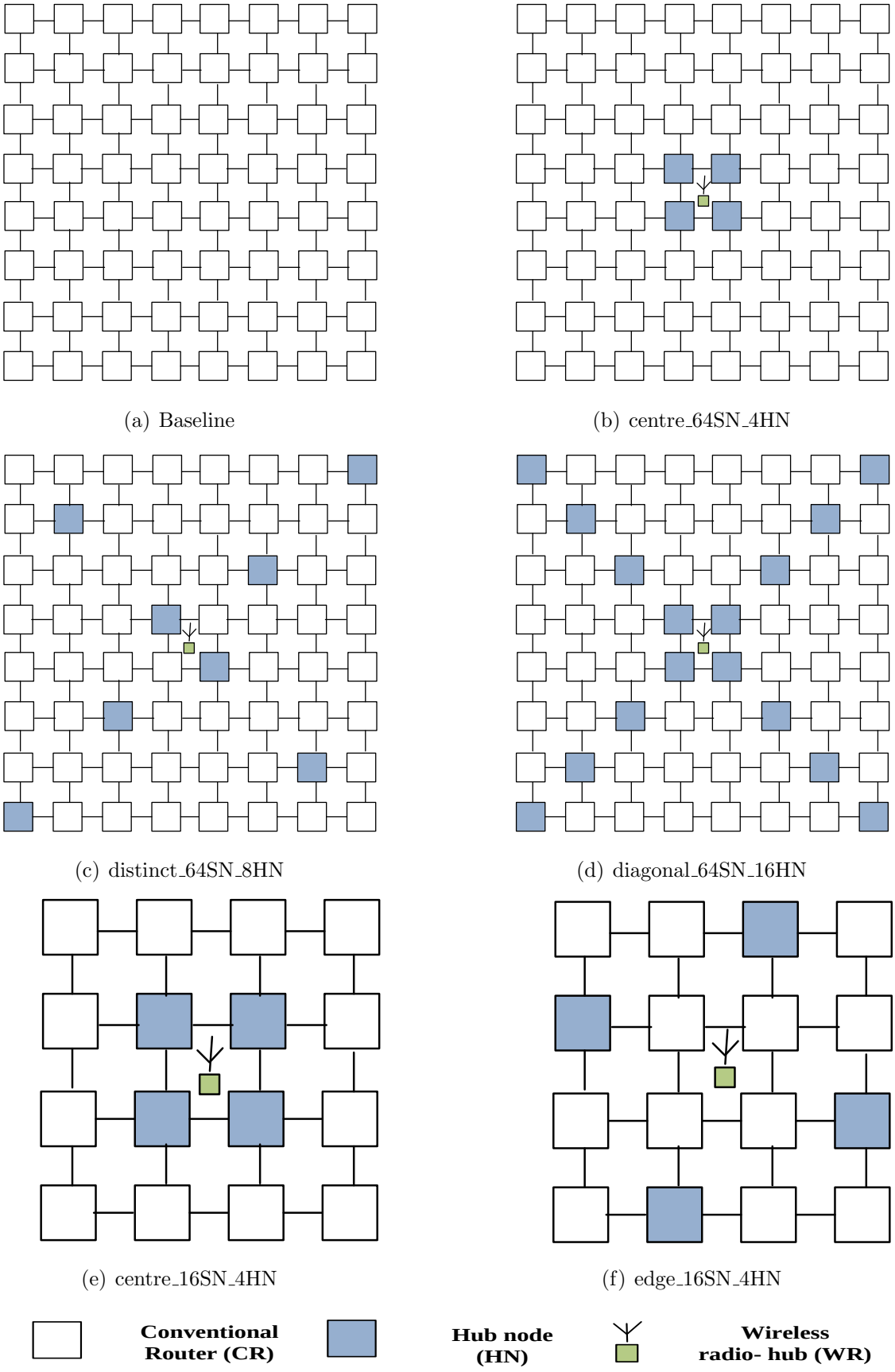


Figure 5.4 Proposed architectures

This study examined five hierarchical hybrid wired-wireless topologies, along with a baseline wired Mesh topology, as depicted in Figure 5.4. Only one subnet and the hub nodes of that subnet are shown in the figure. All other subnets of the network follow the same configuration. Each centrally located WR of a subnet is wirelessly connected to every WR of other subnets. In Figure 5.4, only the middle level of the hierarchy has been depicted. The bottom level and top level of the hierarchy are the logically divided wired Mesh and fully connected WRs, respectively. Both levels are the same in all the five proposed hierarchical hybrid wired-wireless architectures. The middle level of the hierarchy for five topologies differs and has been constructed by varying the number of nodes in the subnets (SN), the number of nodes attached to WR of the subnet i.e., the number of hub nodes (HN), and the placement of hub nodes (placement). The different architectures, also referred to as configurations or topologies are named in the format placement_#SN_#HN. Following are the four placements considered for forming different topologies.

- *centre*: the nodes located at the centre of the subnet are attached to the wireless hub.
- *edge*: The nodes located at the edge of the subnet following a pattern are attached to the wireless hub.
- *diagonal*: the nodes located on the diagonals of the subnet are attached to the wireless hub
- *distinct*: the nodes from each row and column of the subnet are selected in such a way that no two nodes are located in the same row or column

The specifications of the topologies are presented in Table 5.1. The selection of hub nodes for the topologies with center, diagonal and edge placements are evident from the naming convention. The placement i.e., the location of the hub nodes for a specific placement is always the same, it does not change for that particular configuration. This is not the case with distinct placement. For the topology with distinct placement, the hub nodes are selected such that no two nodes are located in the same row or column. There can be different possible combinations of hub nodes for a specific configuration. The number of such possible combinations depends on the subnet size and the number of hub nodes. If the subnet size is n×n and the number of hub nodes to be selected is n, then

the number of combinations of the distinct placement such that no two selected hub nodes are located in the same row or column is n! (factorial of n). So, for the configuration considered, distinct_64SN_8HN with 64 subnet size (8×8 subnet) and 8 hub nodes, the number of possible combinations is 8!=40320 combinations. One such combination of hub nodes for subnet0 of the distinct-64SN-8HN configuration is shown in Figure 5.4(c).

Table 5.1 Proposed architecture specifications

Topology	#S	#SN	#HN	Placement
Baseline	NA	NA	NA	NA
centre_64SN_4HN	16	64	4	centre
distinct_64SN_8HN	16	64	8	distinct
diagonal_64SN_16HN	16	64	16	diagonal
centre_16SN_4HN	64	16	4	centre
edge_16SN_4HN	64	16	4	edge

- #S : Number of subnets
  #SN : Number of nodes of a subnet
  #HN : Number of nodes attached to the wireless radio-hub of the subnet
  Placement : Placement of the hub nodes in the subnet
- Baseline : Wired Mesh topology
- centre_64SN_4HN : Placement of hub nodes is centre, no. of subnets is 16, no. of subnet nodes is 64, no. of hub nodes is 4
- distinct_64SN_8HN: Placement of hub nodes is distinct, no. of subnets is 16, no. of subnet nodes is 64, no. of hub nodes is 8
- diagonal_64SN_16HN : Placement of hub nodes is diagonal, no. of subnets is 16, no. of subnet nodes is 64, no. of hub nodes is 16
- centre_16SN_4HN : Placement of hub nodes is centre, no. of subnets is 64, no. of subnet nodes is 16, no. of hub nodes is 4
- edge_16SN_4HN : Placement of hub nodes is edge, no. of subnets is 64, no. of subnet nodes is 16, no. of hub nodes is 4

5.4 Evaluation

The Noxim simulator [203], is a configurable and extendible cycle accurate simulator. It enables performance and power analysis of both conventional wired and wireless NoC architectures. Noxim that supports wireless NoC simulation unlike Ratatoskr and

Algorithm 3 Pseudo Code for XYW routing

Input: Current Source id (src) and Destination id (dst)
Output: Router to which packet has to be routed
Process:

```
if src==dst then
    Eject
else if src and dst are in the same subnet then
    Follow XY routing
else
    if src is a hub node then
        if dst is a hub node then
            wireless routing from src to dst
        else
            Follow XY routing
    else
        Follow XY routing
```

Booksim simulators, used in the previous studies, was utilised to conduct simulation of the proposed five distinct hierarchical hybrid wired-wireless Network-on-Chip (NoC) architectures, as well as a baseline wired Mesh architecture. The configuration parameters are displayed in Table 5.2. The simulations were conducted for various configurations, encompassing low (0.005), medium (0.01), and high (0.05) packet injection rates (PIRs). These configurations were evaluated under uniform random traffic, transpose, bit reversal, and shuffle traffic patterns. Metrics – Global average delay (GAD), Network Throughput (NT), Total Energy (TE), Energy-Delay-Product (EDP), and Energy per packet (EPP) were measured and analysed. The routing defined for wired-wireless transmission in noxim simulator was considered. It is named as XYW for our reference and is illustrated in Algorithm 3. If the source node and destination node are in the same subnet, packets are transmitted from source to destination following the XY dimensional order routing algorithm. Wired routing has been considered for intra-subnet transmission. If the source node and destination node are in different subnets, according to XYW routing algorithm, wireless transmission takes place if and only if both the source and destination are connected to the Wireless radio hubs. If the source node has no WR and the destination node has WR, the source node routes the packets following XY routing through wired links. In the routing path, if the packet traverses through a node with WR, then between this intermediate node with WR and the destination node with WR, wireless transmission takes place. If the destination node has no WR, even if the source node or any intermediate

node in the traversal path has a WR, wired transmission only takes place following XY routing. This routing algorithm ensures less congestion at WRs while simultaneously reducing the average hop count by accommodating the wireless transmission between pairs of source nodes (or intermediate nodes) and destination nodes with WRs.

Table 5.2 Network configuration parameters

Configuration parameter	Value
Network size	256, 1024 nodes
Placement of Hub nodes	centre, edge, diagonal, distinct
Number of subnets	16, 64
Number of hub nodes	4, 8, 16
Topology	Baseline centre_64SN_4HN distinct_64SN_8HN diagonal_64SN_16HN centre_16SN_4HN edge_16SN_4HN
Simulation cycles	10000
Clock period	1000 ps
Traffic pattern	Uniform Random Transpose Bitreversal Shuffle
Packet Injection Rate (PIR)	0.005 (low), 0.01 (medium), 0.05 (high)

For evaluating distinct_64SN_8HN configuration, as it is not possible to simulate all the 40320 combinations, only a few have been chosen using a heuristic approach. Following are the steps in the heuristic approach.

- *Step 1:* All the possible combinations (40320 combinations) of hub nodes with distinct placement have been computed.
- *Step 2:* Now, out of 40320 combinations, considering one combination, the average of the sum of the distances of each node to every other node following XYW routing protocol has been calculated. Here, the distance between a pair of source node and destination node is calculated in terms of number of hops required to travel from the source node to the destination node following XYW routing.
- *Step 3:* Repeat Step2 for all combinations

- *Step 4:* Out of all the average distances calculated in the above steps, the combinations resulting in the minimum, average, and maximum values have been simulated and tested.

- *Step 5:* The simulation results for these combinations do not have much variations. So, one of such combinations has been chosen for evaluation of distinct_64SN_8HN configuration.

5.4.1 Performance analysis

All the simulation results have been plotted and the results have been discussed in detail. The result values have been normalised concerning Baseline configuration. The simulation findings for 256-node architectures simulated under uniform random traffic are shown in Figure 5.5. From Figure 5.5(a), it can be observed that diagonal_64SN_16HN has the lowest global average delay (GAD). Next to diagonal_64SN_16HN, distinct_64SN_8HN has the lowest GAD. From Figure 5.5(b), it can be observed that for low and medium PIRs, the NT is almost same for all configurations. For high PIR, all the hybrid configurations have improved NT compared to that of Baseline configuration. From Figure 5.5(c), it can be observed that for all PIRs, Baseline has the lowest TE. Compared to Baseline, centre_16SN_4HN, edge_16SN_4HN configurations have higher TE values whereas centre_64SN_4HN, distinct_64SN_8HN and diagonal_64SN_16HN configurations have TE values not much higher than Baseline. Along with the metrics, GAD, NT, and TE, two other metrics, Energy-Delay-Product (EDP) and Energy per packet (EPP) have also been considered as figures of merit. EDP indicates the tradeoff between Energy and Delay. EPP indicates the tradeoff between Energy and throughput. From Figure 5.5(d) and Figure 5.5(e), it can be observed that for low, medium PIRs, compared to Baseline, centre_16SN_4HN, edge_16SN_4HN configurations have higher EDP & EPP values whereas centre_64SN_4HN, distinct_64SN_8HN and diagonal_64SN_16HN configurations have EDP & EPP values not much higher than Baseline. For high PIR, all the hybrid configurations have EDP & EPP values not much higher than Baseline. Overall, among all the hybrid configurations, distinct_64SN_8HN and diagonal_64SN_16HN configurations have reasonably better EDP & EPP values simultaneously providing the benefits of reduced GAD and improved NT over the Baseline.

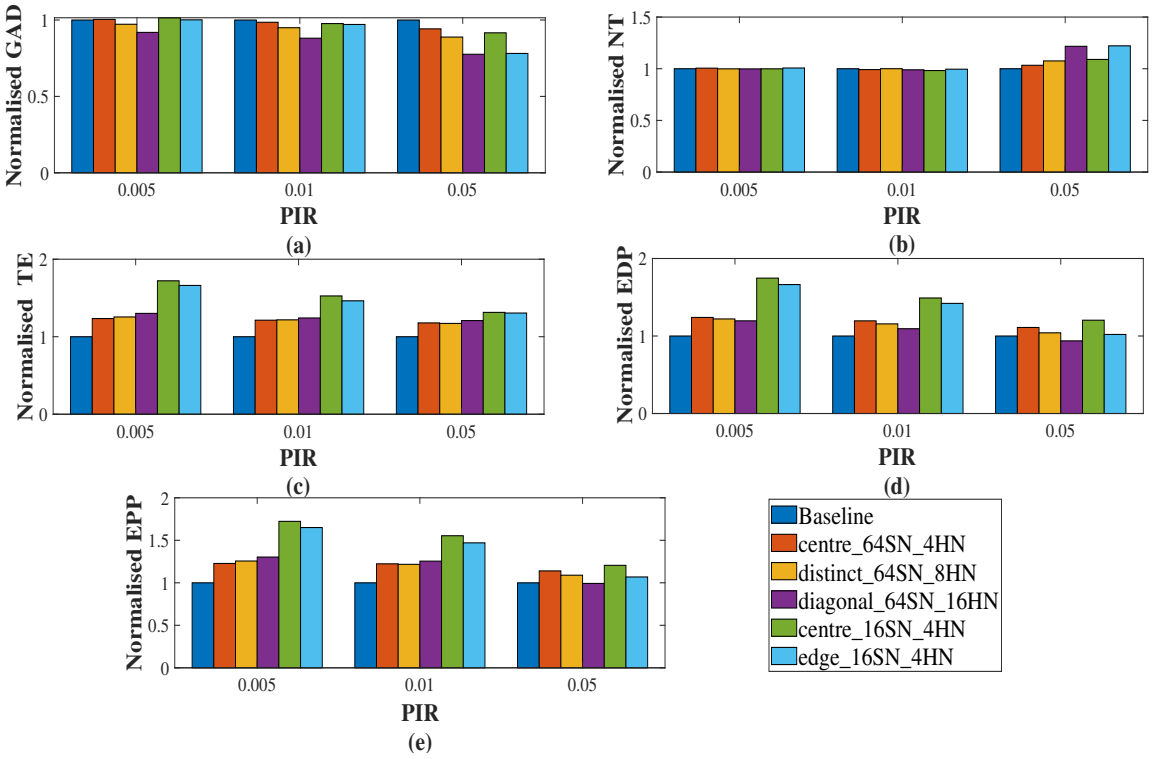


Figure 5.5 Comparison of 256-node configurations simulated under uniform random traffic pattern

The simulation findings for 256-node architectures simulated under transpose traffic have been shown in Figure 5.6. From Figure 5.6(a), it can be observed that for low, and medium PIRS, diagonal_64SN_16HN has the lowest GAD. Next to diagonal_64SN_16HN, distinct_64SN_8HN has the lowest GAD. For high PIR, the configuration, distinct_64SN_8HN has the lowest GAD. From Figure 5.6(b), it can be observed that for low and medium PIRs, the NT is almost same for all configurations. For high PIR, all the hybrid configurations have improved NT compared with that of Baseline configuration. From Figure 5.6(c), it can be observed that for all PIRs, the Baseline has the lowest TE. Compared to Baseline, centre_16SN_4HN, edge_16SN_4HN configurations have higher TE values whereas centre_64SN_4HN, distinct_64SN_8HN and diagonal_64SN_16HN configurations have TE values not much higher than Baseline. From Figure 5.6(d) and Figure 5.6(e), it can be observed that for low PIR, compared to Baseline, centre_16SN_4HN, edge_16SN_4HN configurations have higher EDP & EPP values whereas centre_64SN_4HN, distinct_64SN_8HN and diagonal_64SN_16HN configurations have EDP & EPP values not much higher than Baseline. For medium PIR, distinct_64SN_8HN, diagonal_64SN_16HN, edge_16SN_4HN configurations have EDP lower than that of Baseline. For medium PIR, compared

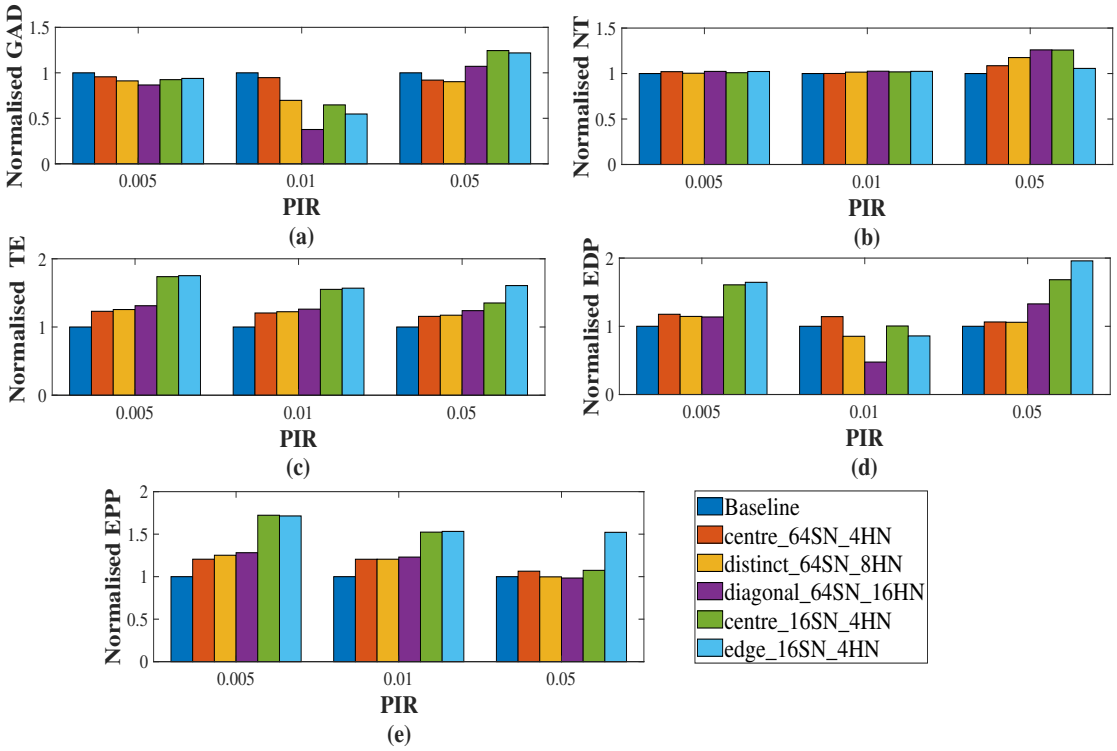


Figure 5.6 Comparison of 256-node configurations simulated under transpose traffic pattern

to Baseline, centre_16SN_4HN, edge_16SN_4HN configurations have higher EDP values whereas centre_64SN_4HN, distinct_64SN_8HN and diagonal_64SN_16HN configurations have EPP values not much higher than Baseline. For high PIR, compared to Baseline, centre_16SN_4HN, edge_16SN_4HN configurations have higher EDP & EPP values whereas centre_64SN_4HN, distinct_64SN_8HN and diagonal_64SN_16HN configurations have EDP and EPP values not much higher than Baseline. Overall, among all hybrid configurations, distinct_64SN_8HN and diagonal_64SN_16HN configurations have reasonably better EDP & EPP values simultaneously providing the benefits of reduced GAD & improved NT over the Baseline, particularly for medium PIR.

The simulation findings for 256-node architectures simulated under bit reversal traffic and shuffle traffic patterns are shown in Figure 5.7 and Figure 5.8 respectively. As discussed, for uniform and transpose traffic patterns the results for bit reversal and shuffle traffic patterns can be analysed in a similar way. Overall, for bit reversal & shuffle patterns, among all the hybrid configurations, distinct_64SN_8HN and diagonal_64SN_16HN configurations have reasonably better EDP & EPP values while simultaneously providing the benefits of reduced GAD and improved NT over the Baseline, particularly for high PIR.

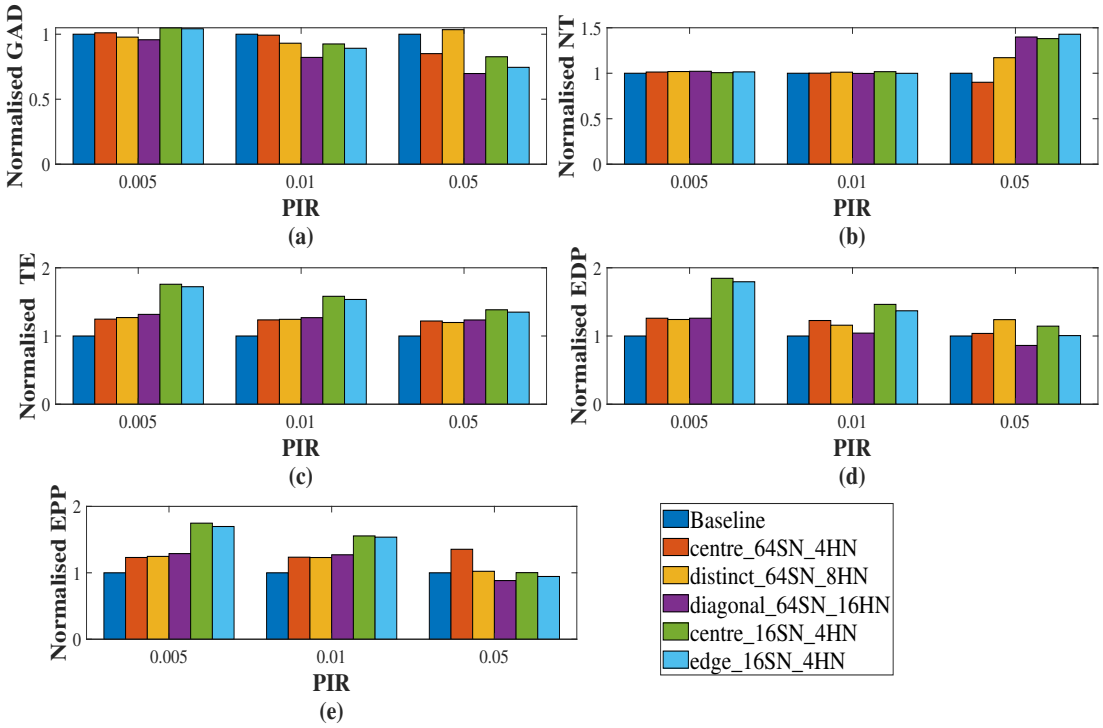


Figure 5.7 Comparison of 256-node configurations simulated under bit reversal traffic pattern

Significant reductions in GAD values or improvements in NT values obtained for different configurations when compared with that of the Baseline configuration for 256-node architectures are shown in Table 5.3.

Table 5.3 Results comparison for 256 nodes

PIR	Results analysis
0.005	**Uniform random traffic pattern** 2% and 8% reduction in GAD for distinct_64SN_8HN and diagonal_64SN_16HN respectively compared with Baseline **Transpose traffic pattern** 4%, 8%, 13%, 7% and 6% reduction in GAD for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline **Bit reversal traffic pattern** 5%,12% and 4% reduction in GAD for distinct_64SN_8HN, diagonal_64SN_16HN and edge_16SN_4HN configurations respectively compared with Baseline

	Shuffle traffic pattern 2% and 4% reduction in GAD for distinct_64SN_8HN and diagonal_64SN_16HN respectively compared with Baseline
0.01	**Uniform random traffic pattern** 5%, 11%, 2% and 3% reduction in GAD for distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline **Transpose traffic pattern** 5%, 30%, 62%, 35% and 45% reduction in GAD for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline **Bit reversal traffic pattern** 10%, 33%, 74%, 10% and 67% reduction in GAD for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline **Shuffle traffic pattern** 7%, 18%, 7% and 11% reduction in GAD for distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline
0.05	**Uniform random traffic pattern** 5%, 11%, 22%, 8% and 21% reduction in GAD for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline 3%, 7%, 21%, 9% and 22% improvement in NT for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline **Transpose traffic pattern** 8% and 9% reduction in GAD for centre_64SN_4HN and distinct_64SN_8HN respectively compared with Baseline

8%, 17%, 26%, 25% and 5% improvement in NT for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline

Bit reversal traffic pattern

4%, 23%, 6% and 17% reduction in GAD for distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline

4%, 34%, 14% and 21% improvement in NT for distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline

Shuffle traffic pattern

14%, 4%, 30%, 17% and 25% reduction in GAD for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline

17%, 40%, 38% and 43% improvement in NT for distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline

The simulation findings for 1024-node architectures simulated under uniform random traffic have been shown in Figure 5.9. From Figure 5.9(a), it can be observed that for low and medium PIRs, diagonal_64SN_16HN has the lowest GAD. Next to diagonal_64SN_16HN, distinct_64SN_8HN has the lowest GAD. For high PIR, the configuration, distinct_64SN_8HN has the lowest GAD. From Figure 5.9(b), it can be observed that for low and medium PIRs, the NT is almost same for all configurations. For high PIR, distinct_64SN_8HN has the highest throughput. From Figure 5.9(c) it can be observed that for all PIRs, Baseline has the lowest TE. For low and medium PIRs, next to Baseline, distinct_64SN_8HN and diagonal_64SN_16HN have the lowest TE. For high PIR, next to the Baseline, edge_16SN_4HN has the lowest TE at the cost of network throughput. From Figure 5.9(d), it can be observed that for low and high PIRs, next to Baseline, distinct_64SN_8HN and diagonal_64SN_16HN have the lowest EDP and they have EDP almost close to that of Baseline. For medium PIR, distinct_64SN_8HN has the lowest EDP. From Figure 5.9(e), it can be observed that for all PIRs, Baseline has the lowest EPP, next to Baseline, distinct_64SN_8HN and diagonal_64SN_16HN have the lowest Energy

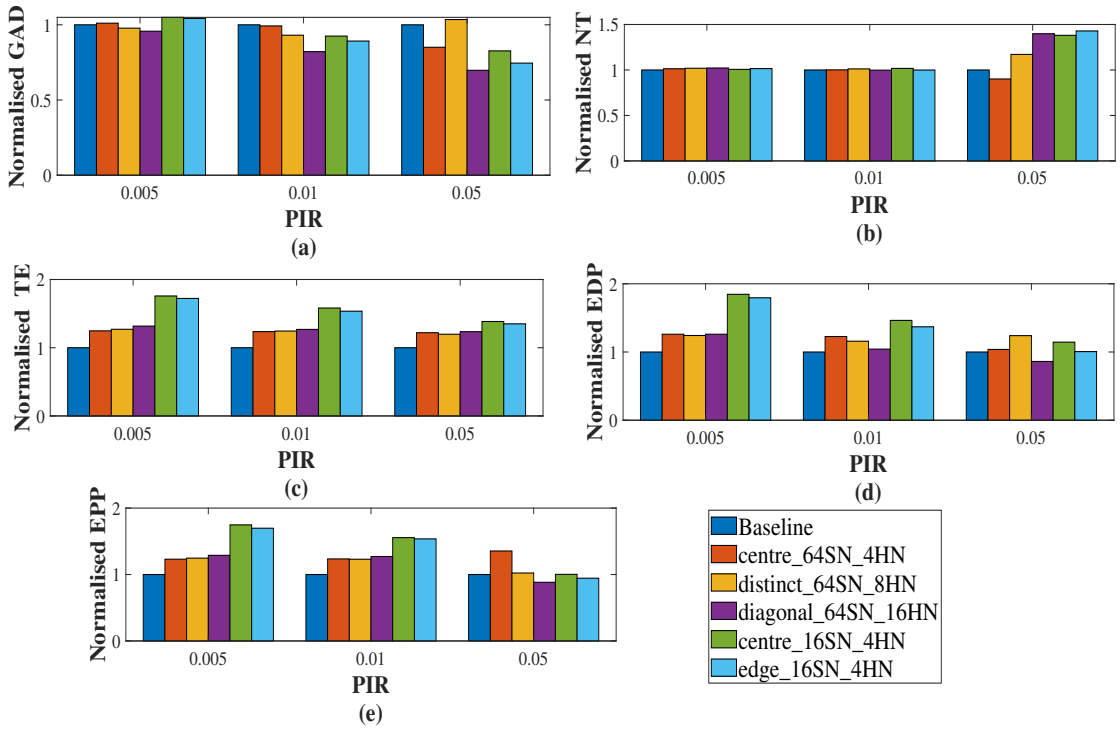


Figure 5.8 Comparison of 256-node configurations simulated under shuffle traffic pattern

per packet. For high PIR, Energy per Packet for distinct_64SN_8HN is almost the same as that of Baseline. Overall, for low PIR, distinct_64SN_8HN, and diagonal_64SN_16HN compared with Baseline have reduced latency, same network throughput, almost the same EDP and EPP. For medium PIR, distinct_64SN_8HN and diagonal_64SN_16HN have reduced latency, same network throughput, lower EDP, and almost the same EPP when compared with Baseline. For high PIR, distinct_64SN_8HN has reduced latency, higher network throughput, and almost the same EDP & EPP compared with Baseline.

The simulation findings for 1024-node architectures simulated under transpose, bit reversal traffic, and shuffle traffic patterns are shown in Figure 5.10, Figure 5.11, and Figure 5.12 respectively. As discussed for uniform random patterns, the results for transpose, bit reversal, and shuffle traffic patterns can be analysed in a similar way. Overall, for transpose, bit reversal, and shuffle traffic patterns, among all hybrid configurations, distinct_64SN_8HN, and diagonal_64SN_16HN configurations have reasonably better EDP & EPP values simultaneously providing the benefits of reduced GAD & improved NT over the baseline, particularly for low and medium PIRs.

Significant reductions in GAD values or improvements in NT values obtained for dif-

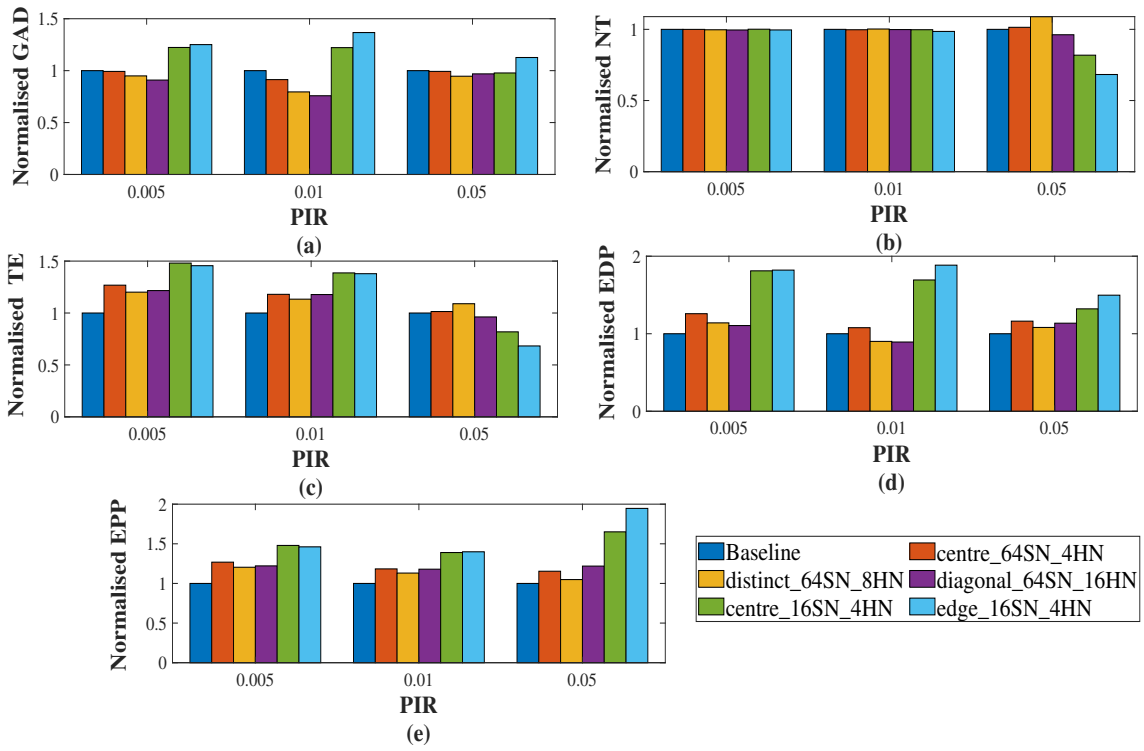


Figure 5.9 Comparison of 1024-node configurations simulated under uniform random traffic pattern

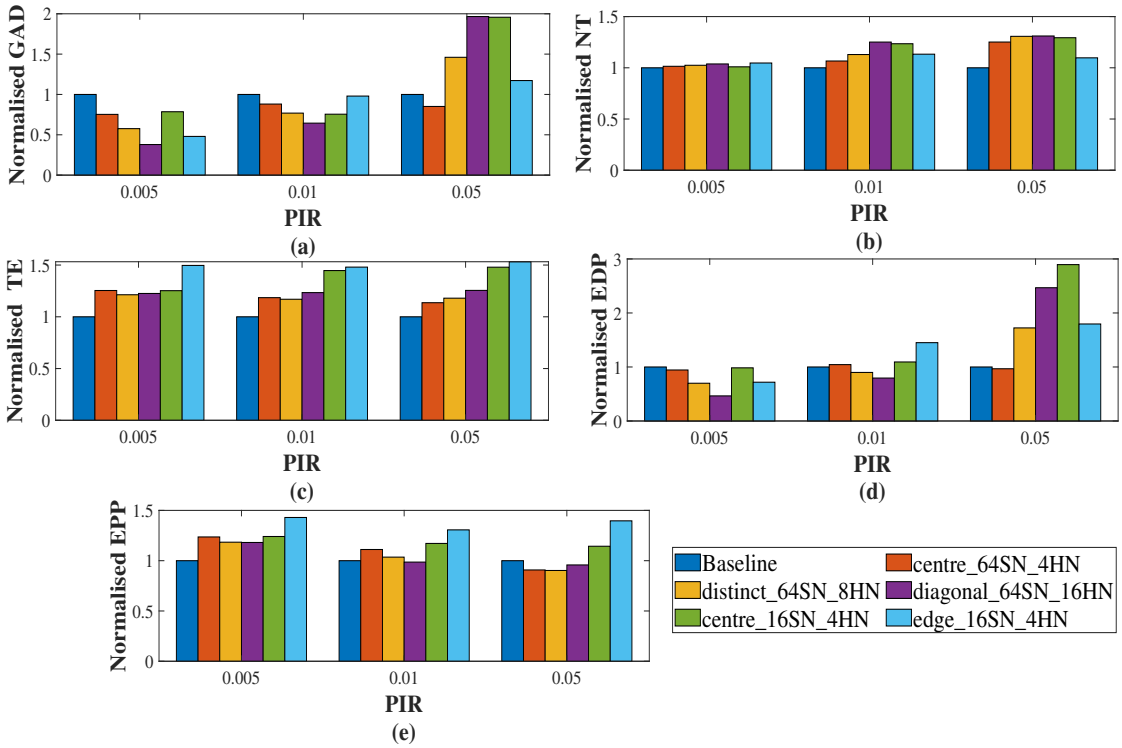


Figure 5.10 Comparison of 1024-node configurations simulated under transpose traffic pattern

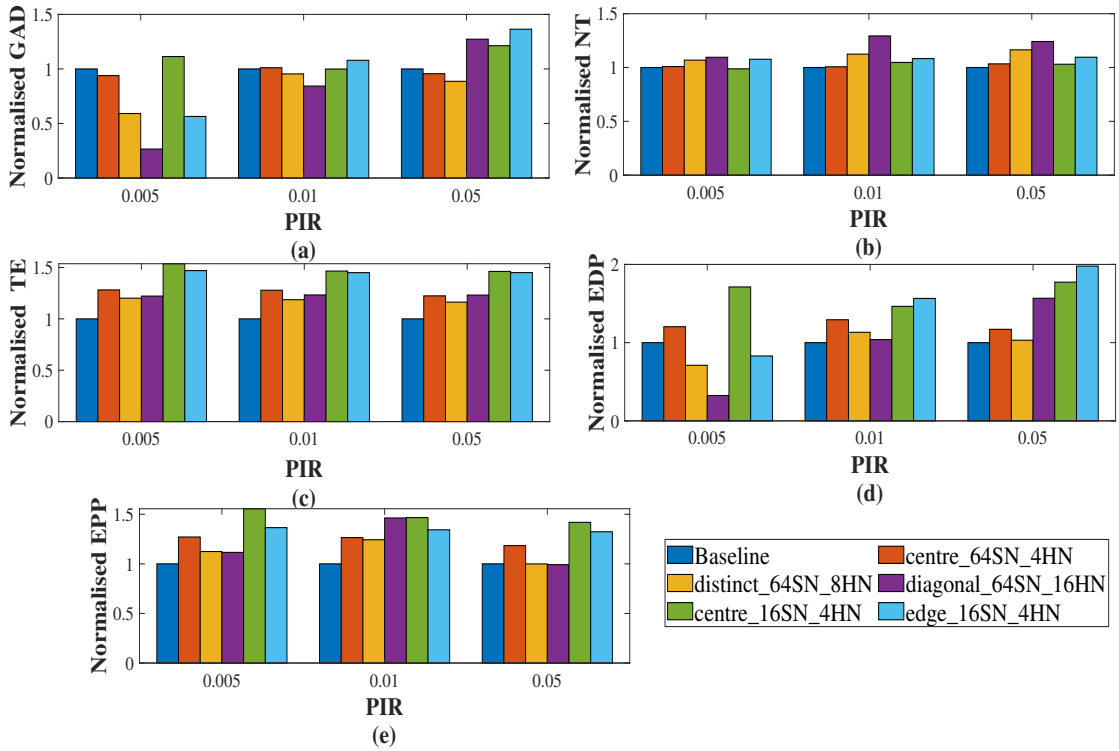


Figure 5.11 Comparison of 1024-node configurations simulated under bit reversal traffic pattern

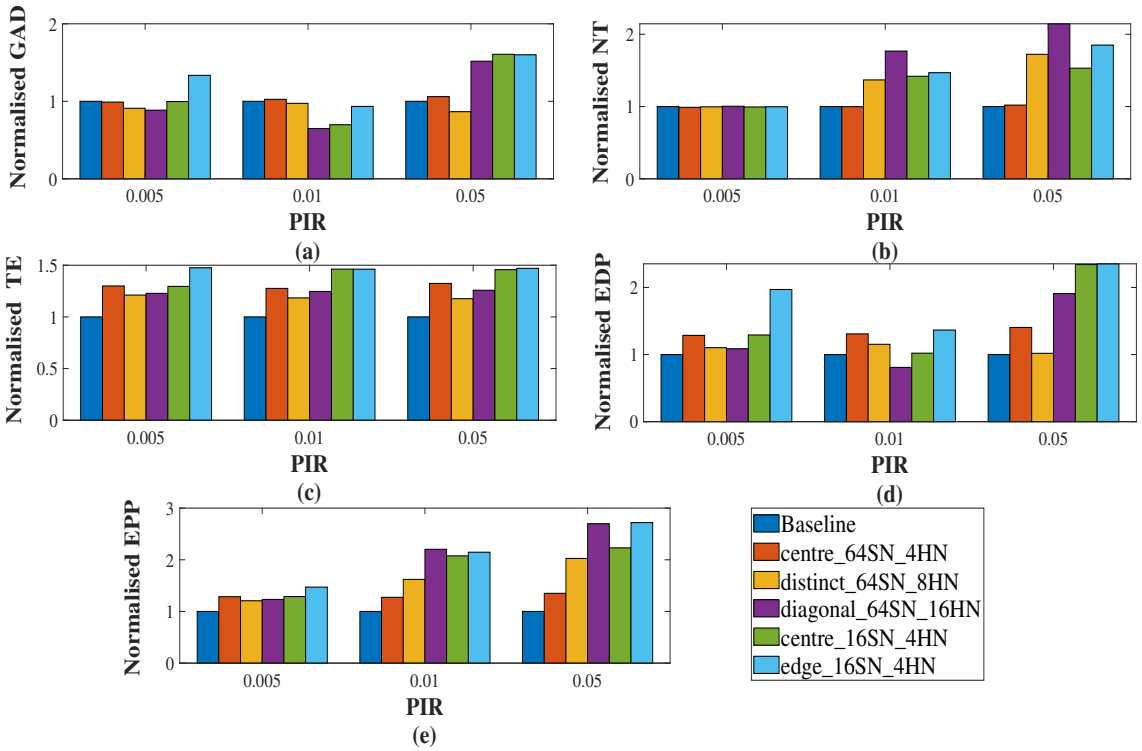


Figure 5.12 Comparison of 1024-node configurations simulated under shuffle traffic pattern

ferent configurations when compared with that of the Baseline configuration are indicated in Table 5.4.

Table 5.4 Results comparison for 1024 node networks

PIR	Results analysis
0.005	**Uniform random traffic pattern** 5% and 9% reduction in GAD for distinct_64SN_8HN and diagonal_64SN_16HN compared with Baseline **Transpose traffic pattern** 25%, 42%, 62%, 21%, 52% reduction in GAD for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN, edge_16SN_4HN configurations respectively compared with Baseline **Bit reversal traffic pattern** 41%, 73%and 43% reduction in GAD for distinct_64SN_8HN, diagonal_64SN_16HN and edge_16SN_4HN configurations respectively compared with Baseline 7%, 10%and 8% improvement in NT for distinct_64SN_8HN, diagonal_64SN_16HN and edge_16SN_4HN configurations respectively compared with Baseline **Shuffle traffic pattern** 8% and 11% reduction in GAD for distinct_64SN_8HN and diagonal_64SN_16HN compared with Baseline
0.01	**Uniform random traffic pattern** 9%, 20% and 24% reduction in GAD for centre_64SN_4HN, distinct_64SN_8HN, and diagonal_64SN_16HN respectively compared with Baseline **Transpose traffic pattern** 12%, 23%, 35% and 24% reduction in GAD for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN and centre_16SN_4HN respectively compared with Baseline **Bit reversal traffic pattern**

	5% and 16% reduction in GAD for distinct_64SN_8HN and diagonal_64SN_16HN compared with Baseline 12%, 29%, 5% and 8% improvement in NT for distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline **Shuffle traffic pattern** 3%, 35%, 30% and 7% reduction in GAD for distinct_64SN_8HN , diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline 37%, 77%, 42% and 47% improvement in NT for distinct_64SN_8HN , diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline
0.05	**Uniform random traffic pattern** 5% and 3% reduction in GAD for distinct_64SN_8HN and diagonal_64SN_16HN respectively compared with Baseline 8% improvement in NT for distinct_64SN_8HN compared with Baseline **Transpose traffic pattern** 14% reduction in GAD for centre_64SN_4HN compared with Baseline 25%, 31%, 31%, 29% and 10% improvement in NT for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline **Bit reversal traffic pattern** 4% and 11% reduction in GAD for centre_64SN_4HN and distinct_64SN_8HN respectively compared with Baseline 3%, 16%, 24%, 3% and 10% improvement in NT for centre_64SN_4HN, distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline **Shuffle traffic pattern** 13% reduction in GAD for distinct_64SN_8HN compared with Baseline

72%, 114%, 53% and 85% improvement in NT for distinct_64SN_8HN, diagonal_64SN_16HN, centre_16SN_4HN and edge_16SN_4HN configurations respectively compared with Baseline

5.4.2 Wiring and Buffer resources requirements

Links, buffer resources, and the number of wireless interfaces contribute a significant part to the power and area overheads of the NoC. The number of wired links, buffers, and wireless interfaces required for each architecture is presented in Table 5.5. For M×N Baseline wired Mesh, the total number of wired links is calculated as N(M-1)+M(N-1)[204]. Thus for 16×16 i.e. 256 node network, the total number of wired links is 480. In the architecture proposed by Ganguly *et al.* [159], the 256-node network is divided into 16 logical subnets, in which the 16 routers of the subnet are connected in Mesh topology. To connect 16 routers in Mesh topology, 24 wired links are required. Each subnet has a centrally located hub to which all the 16 routers of the subnet are wired-connected. So 16 links are required to connect the nodes of a subnet to the hub. Thus, for 16 subnets, 24×16 = 384 node-to-node wired links and 16×16=256 node-to-hub wired links are required. Also, all the hubs are connected in ring topology. So, to connect 16 hubs (hub-to-hub) in a ring topology, 16 wired links are required. To accommodate wireless transmission, 16 wireless interfaces are required as there are 16 wireless routers in the architecture. In the architecture proposed by Deb *et al.* [205], the 256-node network is divided into 16 logical subnets, in which 16 routers of the subnet are connected in Ring topology. To connect 16 routers in Ring topology, 16 wired links are required. Each subnet has a centrally located hub to which all the 16 routers of the subnet are wired-connected. So 16 links are required to connect the nodes of a subnet to the hub. Thus, for 16 subnets, 16×16 = 256 node-to-node wired links and 16×16=256 node-to-hub wired links, are required. Also, all the hubs are connected in Mesh topology. So, to connect 16 hubs (hub-to-hub) in Mesh topology, 24 wired links are required. To accommodate wireless transmission, 6 wireless interfaces are required as there are 6 wireless routers in the architecture.

The total number of buffer resources in a network of a given size can be calculated as follows:

Table 5.5 Comparison of links and buffer resources

Archi	#S	#SN	#CH	#WH	#HN	$\#L_{NN}$	$\# L_{NH}$	$\#L_{HH}$	#Links	#Bufs
1	16	16	NA	NA	NA	480	0	0	480	3840
2	16	16	16	16	16	384	256	16	656	9344
3	16	16	16	6	16	256	256	24	536	5824
4	4	64	-	4	4	480	16	0	496	4224
5	4	64	-	4	8	480	32	0	512	4608
6	4	64	-	4	16	480	64	0	544	5376
7	16	16	-	16	4	480	64	0	544	5376
8	16	16	-	16	4	480	64	0	544	5376

Archi : Referred Architetures from previous and present work
Network size of all the considered architectures in the table is 256 nodes.
1 - Baseline wired Mesh, 2 - Ganguly [159], 3 - Deb [205], 4- centre_64SN_4HN,
5- distint_64SN_8HN, 6 - diagonal_64SN_16HN, 7 - centre_16SN_4HN, 8 - edge_16SN_4HN
#S : Number of subnets
#SN : Number of nodes in each subnet
#CH : Number of conventional i.e., wired hubs in the architecture
#WH : Number of wireless hubs in the architecture
#HN : Number of hub nodes i.e., nodes attached to the hub of each subnet
$\#L_{NN}$: Number of node to node connecting short range wired links
$\# L_{NH}$: Number of node to hub connecting medium range wired links
$\#L_{HH}$: Number of hub to hub connecting long range wired links
#Links : Total number of wired links in the architecture
#Bufs : Total number of buffer resources in the architecture

The total number of buffer resources required for wired routers = number of wired links × number of virtual channels per link × number of buffers per virtual channel

The total number of buffer resources required for wireless routers i.e., wireless hubs = number of wireless hubs × number of nodes attached to the wireless hub × number of virtual channels per transmitter × number of buffers per virtual channel of a transmitter × number of virtual channels per receiver × number of buffers per virtual channel of a receiver

The total number of buffer resources in a network = Total number of buffer resources required for wired routers + Total number of buffer resources required for wireless routers i.e., wireless hubs

For a network size of 256 nodes, considering an identical configuration of 2 virtual

channels (VC) per link, 2 VCs per transmitter, 2VCs per receiver, and 4 buffers per VC for all the architectures, the total number of buffer resources required for different architectures is given in the Table 5.5. From Table 5.5, it can be observed that the total number of buffer resources required for the proposed architectures is fewer than that of the total number of buffer resources required for the architectures proposed by Ganguly *et al.* in [159] and Deb *et al.* in [205]. In both the architectures, the subnet size is small and fixed to 16 and all the nodes of the subnet are connected to the central hub of the subnet. Also, at the top level of the hierarchy, the number and placement of the wireless hubs are optimised. But for large network sizes like 1024, if the subnet size is fixed to 16, the number of subnets, also the number of hubs, will be 64. The number of wireless interfaces even after optimisation could lead to more area overhead. This issue could be addressed by the approaches proposed in the present work i.e., to increase the subnet size to 64 thereby reducing the number of wireless interfaces to 16 and varying the number of hub nodes to 8 or 16 as per the latency, area, and power requirements of the application so as not to compromise performance.

5.5 Summary

This chapter provides an in-depth exploration of the design of 256-node and 1024-node hybrid wired-wireless NoC architectures as 3-level hierarchical hybrid wired-wireless NoC architectures. Five hierarchical hybrid wired-wireless configurations have been formulated by varying the number of subnets, which correspond to the number of wireless radio hubs, number of hub nodes, and the placement of these hub nodes. All the hybrid wired-wireless and baseline wired Mesh networks have been simulated under different synthetic traffic patterns- uniform traffic pattern, transpose, bit reversal, and shuffle for low (0.005), medium (0.01), and high (0.05) PIRs. Evaluation metrics – Global average delay (GAD), Network Throughput (NT), Total Energy (TE), Energy-Delay-Product (EDP), and Energy per packet (EPP) were analysed. Among all the configurations considered, distinct_64SN_8HN configuration, characterized by 16 subnets with 64 nodes per subnet and 8 hub nodes with distinct placement such that no two hub nodes are in the same row or same column, emerges as a more efficient topology. This configuration demonstrates reduced latency, and improved throughput, with only a slight penalty in area and energy

when compared to Baseline wired Mesh.

Summing up, distinct_64SN_8HN configuration compared to all other considered wired and hybrid wired-wireless configurations has proven to be an efficient configuration as it provides a trade-off among metrics latency, throughput, area, and power. It is important to acknowledge that there could be multiple symmetrical solutions, each offering comparable or potentially improved performance. To further enhance the performance of the proposed 1024-node wired-wireless hierarchical hybrid NoC architectures, it is recommended to explore the design space of the middle level of the hierarchy, coupled with optimisation of the number and placement of wireless hubs in the top-level.

Chapter 6

Conclusions and Scope for future research

This chapter highlights the significant contributions of the extensively discussed work in the preceding chapters. Moreover, potential areas for further advancement in this research are suggested.

6.1 Conclusions

The objective of the work reported in this thesis is to explore on-chip interconnect topologies and formulate efficient and scalable novel topologies including 2D, 3D, and hybrid wired-wireless topologies to surpass the performance of traditional Mesh topology. Interconnect wire delays significantly influence network performance. Considering this, a hybrid topology based on hexagons and stars has been developed with a reduced number of wired links compared to an identically sized Mesh topology while simultaneously providing enhanced performance and reduced area benefits over Mesh topology. The proposed hybrid hexagonal star topology outperforms Mesh in latency and occupies less area, albeit with a more complex structure. Subsequently, DiamondMesh, a diagonal Mesh topology, has been developed by incorporating diagonal links in the baseline Mesh topology in further exploration of topologies. The proposed DiamondMesh topology outperforms Mesh and is established to be a well-balanced topology among state-of-the-art diagonal Mesh topologies. This topology provides enhanced performance compared to Mesh while preserving the simple, scalable, and regular structure of Mesh. However, the two contributions of the thesis have focussed on wired topologies, which may not fully meet the stringent performance and power constraints of the kilo-core architectures. Consequently, the third contribution is investigation of hybrid wired-wireless on-chip topologies,

formulating various configurations. These configurations indicated reduced latency and improved throughput with a slight penalty in area and energy when compared to Baseline wired Mesh. The key conclusions of the research documented in this thesis can be summarized as follows:

- The study proposed a novel, scalable, symmetrical hybrid Hexagonal Star (HS) topology. The quantitative and graphical analysis of the topological parameters has shown that the network cost of HS is lower than that of HS, and HCM and slightly higher than that of Torus. The packing density of HS is higher than that of Mesh, HCM and lower than that of Torus. The bisection width of HS is higher than that of Mesh, Torus, and HCM. The HS and Mesh topologies were simulated under different traffic patterns varying the injection rates for 18-node and 32-node network sizes and further these were synthesized to estimate the area occupied. The simulation results have demonstrated a reduced packet latency, ranging from 15% to 50% for 18 nodes and 9% to 23% for 32 nodes with HS in comparison to Mesh topology. The synthesis results indicated a decrease of 13.5% in LUTs utilization & 8.58% in FFs utilization for 18-node HS topology compared to 18-node Mesh topology and a reduction of 17.5% in LUTs utilization & 11.6% in FFs utilization for 32-node HS topology in comparison to 32-node Mesh topology. The simulation and synthesis results corroborate the analysis based on topological parameters. The proposed HS topology, with its hexagonal-shaped structure, lower network diameter, reduced network cost, higher packing density, and greater bisection width compared to the Mesh topology, emerges as a promising option for on-chip interconnection networks.

- The study proposed DiamondMesh topology, a diagonal mesh-based topology, is developed by incorporating diagonal links into baseline mesh topology. The study explores topological characteristics of DiamondMesh as well as other investigated topologies such as Mesh, DMesh, XDMesh, and ZMesh to assess their static performance. Notably, the Network diameter of DMesh, DiamondMesh and XDMesh is small and nearly identical for all network sizes while Mesh and ZMesh have larger network diameters. DMesh has the highest bisection width while Mesh has the lowest bisection width. The bisection width of DiamondMesh is between that of DMesh and Mesh and is the same as that of ZMesh. Next to XDMesh, DiamondMesh

has minimum network cost and maximum packing density. Consequently, DiamondMesh exhibits commendable topological characteristics consistently lying between the most favorable and least desirable values. DiamondMesh along with other investigated topologies has been simulated under different synthetic and real time traffic patterns for varying loads. Overall, the trend in characteristics infers that Average Packet Latency (APL) of DiamondMesh has been consistently lower than that of Mesh, ZMesh topologies and closer to that of DMesh topology. Total Network Power (TNP) of DiamondMesh has been always intermediate to that of DMesh which consumes maximum power and Mesh which consumes minimum power. The power-Latency-Product (PLP) of DiamondMesh has been always better than that of Mesh, ZMesh, and XDMesh topologies. For lower injection rates, PLP of DiamondMesh has been lower than that of DMesh. Throughput of DiamondMesh has been always better than that of Mesh, ZMesh topologies and it is close to that of DMesh. Furthermore, a significant reduction of 28% and 30% in the area occupied by DiamondMesh compared to that of DMesh is noted for 16-node and 64-node networks, respectively. Synthesis results align with simulation results, confirming the efficiency of DiamondMesh over DMesh in terms of area utilization.

In summary, DiamondMesh has demonstrated improved network performance while retaining the regular, simple, and scalable properties of the Mesh topology. The topology has been shown to be more balanced than the other diagonal mesh-based topologies considered in terms of the area-performance-power trade-off.

- To further explore DiamomdMesh, it is extended to 3D- DiamondMesh by stacking the 2D DiamondMesh layers vertically. 3D-DiamondMesh along with other 3D architectures, including 3D-Mesh, 3D-DMesh, 3D-XDMesh, and 3D-ZMesh have been simulated for different network sizes ranging from 16 nodes to 256 nodes and different configurations (1 Layer, 2 Layer, 4 Layer) under various traffic patterns. The results indicate that 3D architectures consistently outperform their equivalently configured 2D counterparts. Among the investigated topologies, 3D-DiamondMesh has been shown to be a balanced topology in terms of performance and area overhead. Furthermore, five heterogeneous 3D architectures have been suggested and evaluated. The analysis of the results has inferred that introducing heterogeneity

in topology across the layers can either reduce latency with a slight area overhead or reduce the area with a slight penalty in performance.

- Designed three-level hierarchical hybrid wired-wireless NoC architectures for 256-node and 1024-node networks. By varying the number of subnets i.e., the number of wireless radio-hubs, number of hub nodes, and placement of hub nodes, five hierarchical hybrid wired-wireless configurations were designed and evaluated. All the hybrid wired-wireless and Baseline wired Mesh networks were simulated under different traffic patterns for low (0.005), medium (0.01), and high (0.05) PIRs. All the hybrid wired-wireless configurations were evaluated along with the baseline wired Mesh in terms of Global Average Delay (GAD), Network Throughput (NT), Total Energy (TE), Energy-Delay-Product (EDP) and Energy per packet and observed that the hybrid configurations exhibit superior performance over baseline Mesh topology. Among all the configurations considered, distinct_64SN_8HN i.e., a configuration with 16 subnets having 64 nodes per subnet and 8 hub nodes with distinct placement such that no two hub nodes are in the same row or same column, has proven to be a more efficient topology resulting in reduced latency and improved throughput with a little penalty in area and energy compared with Baseline wired Mesh.

6.2 Scope for future research

The Work outlined in this thesis provides a foundation for potential future investigations. The present study proposed a novel hybrid hexagonal star topology and analysed the topology for small network size simulated under synthetic traffic patterns. It is imperative to evaluate the proposed topology for larger network sizes, considering diverse, real-time traffic patterns. Furthermore, future studies should explore efficient routing strategies and flow control mechanisms that can leverage the benefits of large bisection width inherent in HS topology. Also, future studies should investigate on the possible scalable architectures - vertically stacked HS, Mesh of Hexagonal Stars and Honeycomb of Hexagonal Stars that are suggested in the present study.

The study on hybrid wired-wireless configurations explores the the middle level of

the hierarchical architecture. Future research should explore the design space of this middle level of the hierarchy combined with optimisation of the number & placement of the wireless hubs in the top-level.

The current study on HS, DiamondMesh, and hybrid wired-wireless NoC configurations has focussed solely on homogeneous cores. Investigating these proposed architectures with heterogeneous cores presents a promising opportunity for further research.

Publications

List of International Journals:

1. **Lakshmi Kiranmai V.**, and B. K. N. Srinivasarao. "A Novel Hybrid Hexagonal Star Topology for On-Chip Interconnection Networks." Journal of Circuits, Systems and Computers 32.05 (2023): 2350076. – DOI: 10.1142/S0218126623500767 - (World Scientific).

2. **Varanasi Lakshmi Kiranmai**, and BK N. Srinivasarao. "Design and evaluation of an energy efficient DiamondMesh topology for on-chip interconnection networks." Design Automation for Embedded Systems 26.3-4 (2022): 161-187. – DOI: https://doi.org/10.1007/s10617-022-09266-0 - (Springer)

3. **Varanasi Lakshmi Kiranmai**, and B. K N. Srinivasarao. "Design of a three Level hierarchical hybrid wired-wireless Network-on-Chip architecture". (Under Review)

4. **Varanasi Lakshmi Kiranmai**, and B. K N. Srinivasarao, "Review of On-chip interconnect technologies" (Under Review)

List of International Conferences:

1. **V. L. Kiranmai** and B. K. N. Srinivasarao, "Impact analysis of virtual channels and Buffers on different Network-on-chip configurations," 2022 IEEE International Symposium on Smart Electronic Systems (iSES), Warangal, India, 2022, pp. 578-581, doi: 10.1109/iSES54909.2022.00126.

2. **V. Lakshmi Kiranmai**, K. N. Srinivasarao Batta, and S. Kotte,"Evaluation of 3D-Diamondmesh homogeneous and heterogeneous architectures," in TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON), 2023, pp. 1293–1298

Bibliography

[1] I. A. Alimi, R. K. Patel, O. Aboderin, A. M. Abdalla, R. A. Gbadamosi, N. J. Muga, A. N. Pinto, and A. L. Teixeira, "Network-on-chip topologies: Potentials, technical challenges, recent advances and research direction," *Network-on-Chip-Architecture, Optimization, and Design Explorations*, 2021.

[2] S. Pasricha and N. Dutt, *On-chip communication architectures: system on chip interconnect*. Morgan Kaufmann, 2010.

[3] W.-C. Tsai, Y.-C. Lan, Y.-H. Hu, and S.-J. Chen, "Networks on chips: structure and design methodologies," *Journal of Electrical and Computer Engineering*, vol. 2012, pp. 2–2, 2012.

[4] M. F. Chang, J. Cong, A. Kaplan, M. Naik, G. Reinman, E. Socher, and S.-W. Tam, "CMP network-on-chip overlaid with multi-band RF-interconnect," in *2008 IEEE 14th International Symposium on High Performance Computer Architecture*. IEEE, 2008, pp. 191–202.

[5] G. Chen, H. Chen, M. Haurylau, N. A. Nelson, D. H. Albonesi, P. M. Fauchet, and E. G. Friedman, "On-chip copper-based vs. optical interconnects: Delay uncertainty, latency, power, and bandwidth density comparative predictions," in *2006 International Interconnect Technology Conference*. IEEE, 2006, pp. 39–41.

[6] L. Benini and G. De Micheli, "Networks on chips: a new soc paradigm," *Computer*, vol. 35, no. 1, pp. 70–78, 2002.

[7] L. Benini and D. Bertozzi, "Network-on-chip architectures and design methods," *IEE Proceedings-Computers and Digital Techniques*, vol. 152, no. 2, pp. 261–272, 2005.

[8] T. Bjerregaard and S. Mahadevan, "A survey of research and practices of network-on-chip," *ACM Computing Surveys (CSUR)*, vol. 38, no. 1, pp. 1–es, 2006.

[9] N. Choudhary, "Network-on-chip: a new soc communication infrastructure paradigm," *International Journal of Soft Computing and Engineering*, vol. 1, no. 6, pp. 332–335, 2012.

[10] S. Kundu and S. Chattopadhyay., *Network-on-Chip: The Next Generation of System-on-Chip Integration.* Tayor and Francis group: CRC Press, 2014.

[11] K. G. S. M. Chen and D. D. Kandlur, "Addressing, routing, and broadcasting in hexagonal mesh multiprocessors," *IEEE Transactions on Computers*, pp. vol. 39, no. 1, pp. 10–18, January 1990.

[12] I. Stojmenovic, "Honeycomb networks: Topological properties and communication algorithms," *IEEE Transactions on Parallel and Distributed Systems*, vol. 8, no. 10, pp. 1036–1042, October 1997.

[13] S. Kundu, J. Soumya, and S. Chattopadhyay, "Design and evaluation of mesh-of-tree based network-on-chip using virtual channel router," *Microprocessors and Microsystems*, vol. 36, no. 6, pp. 471–488, 2012.

[14] A. Bouhraoua and M. E. Elrabaa, "Improved modified fat-tree topology network-on-chip," *Journal of circuits, systems, and computers*, vol. 20, no. 04, pp. 757–780, 2011.

[15] S. Kumar, A. Jantsch, J.-P. Soininen, M. Forsell, M. Millberg, J. Oberg, K. Tiensyrja, and A. Hemani, "A network on chip architecture and design methodology," in *Proceedings IEEE Computer Society Annual Symposium on VLSI. New Paradigms for VLSI Systems Design. ISVLSI 2002*, 2002, pp. 117–124.

[16] W. Dally, "Performance analysis of k-ary n-cube interconnection networks," *IEEE Transactions on Computers*, vol. 39, no. 6, pp. 775–785, 1990.

[17] J. Kim, J. Balfour, and W. Dally, "Flattened butterfly topology for on-chip networks," in *40th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO 2007)*, 2007, pp. 172–182.

[18] U. Ogras and R. Marculescu, "It's a small world after all: NoC performance optimization via long-range link insertion," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 14, no. 7, pp. 693–706, 2006.

[19] B. Grot, J. Hestness, S. W. Keckler, and O. Mutlu, "Express cube topologies for on-chip interconnects," in *2009 IEEE 15th International Symposium on High Performance Computer Architecture*, 2009, pp. 163–174.

[20] F. Liu, H. Gu, and Y. Yang, "Performance study of virtual-channel router for network-on-chip," in *2010 International Conference On Computer Design and Applications*, vol. 5, 2010, pp. V5–255–V5–259.

[21] A. V. Bhaskar and T. G. Venkatesh, "A study of the effect of virtual channels on the performance of network-on-chip," in *2015 IEEE Student Conference on Research and Development (SCOReD)*, 2015, pp. 255–260.

[22] C. A. Nicopoulos, D. Park, J. Kim, N. Vijaykrishnan, M. S. Yousif, and C. R. Das, "Vichar: A dynamic virtual channel regulator for network-on-chip routers," in *2006 39th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO'06)*, 2006, pp. 333–346.

[23] M. S. Sayed, A. Shalaby, M. El-Sayed, and V. Goulart, "Flexible router architecture for network-on-chip," *Computers & Mathematics with Applications*, vol. 64, no. 5, pp. 1301–1310, 2012, advanced Technologies in Computer, Consumer and Control.

[24] C. Wang, W.-H. Hu, S. E. Lee, and N. Bagherzadeh, "Area and power-efficient innovative congestion-aware network-on-chip architecture," *Journal of Systems Architecture*, vol. 57, no. 1, pp. 24–38, 2011, special Issue On-Chip Parallel And Network-Based Systems.

[25] M. H. Furhad and J.-M. Kim, "An extended diagonal mesh topology for network-on-chip architectures," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 10, no. 10, pp. 197–210, 2015.

[26] T. C. Xu, V. Leppänen, P. Liljeberg, J. Plosila, and H. Tenhunen, "PDNOC: Partially diagonal network-on-chip for high efficiency multicore systems," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 4, pp. 1054–1067, 2015.

[27] N. Prasad, P. Mukherjee, S. Chattopadhyay, and I. Chakrabarti, "Design and evaluation of zmesh topology for on-chip interconnection networks," *Journal of Parallel and Distributed Computing*, vol. 113, pp. 17–36, 2018.

[28] X. Wang, D. Xiang, and Z. Yu, "TM: a new and simple topology for interconnection networks," *The Journal of Supercomputing*, vol. 66, no. 1, pp. 514–538, 2013.

[29] M. Moudi, M. Othman, K. Y. Lun, and A. R. A. Rahiman, "x-folded tm: An efficient topology for interconnection networks," *Journal of Network and Computer Applications*, vol. 73, pp. 27–34, 2016.

[30] P. Dai, J. Chen, Y. Zhao, and Y.-H. Lai, "A study of a wire–wireless hybrid noc architecture with an energy-proportional multicast scheme for energy efficiency," *Computers & Electrical Engineering*, vol. 45, pp. 402–416, 2015.

[31] H. Chung, C. Teuscher, and P. Pande, "Design and evaluation of technology-agnostic heterogeneous networks-on-chip," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 10, no. 3, pp. 1–27, 2014.

[32] L. Middleton and J. Sivaswamy., *Hexagonal Image Processing A Practical Approach.* ss: Springer, 2005.

[33] L. N. Lester and J. Sandor, "Computer graphics on hexagonal grid," *Computer Graphics*, vol. 8, no. 4, pp. 401–409, 1984.

[34] F. Garcia Nocetti, I. Stojmenovic, and J. Zhang, "Addressing and routing in hexagonal networks with applications for tracking mobile users and connection rerouting in cellular networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 9, pp. 963–971, 2002.

[35] J. Carle and J. Myoupo, "Topological properties and optimal routing algorithms for three dimensional hexagonal networks," in *Proceedings Fourth International Conference/Exhibition on High Performance Computing in the Asia-Pacific Region*, 2000, pp. 116–121 vol.1.

[36] Z. Xiao and B. Baas, "A hexagonal processor and interconnect topology for many-core architecture with dense on-chip networks," in *IEEE International Conference*

on Very Large Scale Integration - System on a Chip (VLSI-SoC-2012), 2012, pp. pp.125–143.

[37] C. Decayeux and D. Seme, "3D hexagonal network: modeling, topological properties, addressing scheme, and optimal routing algorithm," *IEEE Transactions on Parallel and Distributed Systems*, pp. vol. 16, no. 9, pp. 875–884, Sept. 2005.

[38] R. K. Saini and M. Ahmed, "2D hexagonal mesh Vs 3D mesh network on chip: A performance evaluation," *Int. J.Comput. digital Syst*, vol. 4, no. 01, pp. 33–41, 2015.

[39] I. S. F. Garcia, J. Solano and M. Stojmenovic, "Higher dimensional hexagonal networks," *Journal of Parallel and Distributed Computing*, vol. 63, p. 1164–1172, 2003.

[40] A. W. Yin, T. C. Xu, P. Liljeberg, and H. Tenhunen, "Explorations of honeycomb topologies for network-on-chip," in *2009 Sixth IFIP International Conference on Network and Parallel Computing*, 2009, pp. 73–79.

[41] A. Yin, N. Chen, P. Liljeberg, and H. Tenhunen, "Comparison of mesh and honeycomb network-on-chip architectures," in *2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2012, pp. 1716–1720.

[42] J.-F. M. Jean Carle and I. Stojmenovic, "Higher dimensional honeycomb networks," *Journal of Interconnection Networks*, vol. 2, no. 4, pp. 391–420, 2001.

[43] J. Rabaey, *Digital Integrated Circuits: A Design Perspective.* (Prentice-Hall, 2004.

[44] N. Patel and Y. Agrawal, "A literature review on next generation graphene interconnects," *Journal of Circuits, Systems and Computers*, vol. 28, no. 09, p. 1930008, 2019.

[45] P. C. Andricacos, "Copper on-chip interconnections: A breakthrough in electrodeposition to make better chips," *The Electrochemical Society Interface*, vol. 8, no. 1, p. 32, 1999.

[46] J. Lloyd, C. Murray, S. Ponoth, S. Cohen, and E. Liniger, "The effect of cu diffusion on the tddb behavior in a low-k interlevel dielectrics," *Microelectronics Reliability*, vol. 46, no. 9-11, pp. 1643–1647, 2006.

[47] T. K. Wong, "Time dependent dielectric breakdown in copper low-k interconnects: Mechanisms and reliability models," *Materials*, vol. 5, no. 9, pp. 1602–1625, 2012.

[48] D. Edelstein, C. Uzoh, C. Cabral, P. DeHaven, P. Buchwalter, A. Simon, E. Cooney, S. Malhotra, D. Klaus, H. Rathore *et al.*, "A high performance liner for copper damascene interconnects," in *Proceedings of the IEEE 2001 International Interconnect Technology Conference (Cat. No. 01EX461)*. IEEE, 2001, pp. 9–11.

[49] W.-L. Wang, C.-T. Wang, W.-C. Chen, K.-T. Peng, M.-H. Yeh, H.-C. Kuo, H.-J. Chien, J.-C. Chuang, and T.-H. Ying, "The reliability improvement of cu interconnection by the control of crystallized α-ta/tanx diffusion barrier," *Journal of Nanomaterials*, vol. 2015, pp. 9–9, 2015.

[50] F. Chen and D. Gardner, "Influence of line dimensions on the resistance of cu interconnections," *IEEE Electron Device Letters*, vol. 19, no. 12, pp. 508–510, 1998.

[51] R. L. Graham, G. Alers, T. Mountsier, N. Shamma, S. Dhuey, S. Cabrini, R. H. Geiss, D. T. Read, and S. Peddeti, "Resistivity dominated by surface scattering in sub-50 nm cu wires," *Applied Physics Letters*, vol. 96, no. 4, 2010.

[52] N. Jourdan, M. Van Der Veen, V. V. Gonzalez, K. Croes, A. Lesniewska, O. V. Pedreira, S. Van Elshocht, J. Bömmels, and Z. Tökei, "CVD-Mn/CVD-Ru-based barrier/liner solution for advanced BEOL Cu/Low-k interconnects," in *2016 IEEE International Interconnect Technology Conference/Advanced Metallization Conference (IITC/AMC)*. IEEE, 2016, pp. 37–39.

[53] M. H. van der Veen, N. Jourdan, V. V. Gonzalez, C. Wilson, N. Heylen, O. V. Pedreira, H. Struyf, K. Croes, J. Bömmels, and Z. Tőkei, "Barrier/liner stacks for scaling the cu interconnect metallization," in *2016 IEEE International Interconnect Technology Conference/Advanced Metallization Conference (IITC/AMC)*. IEEE, 2016, pp. 28–30.

[54] C.-L. Lo, M. Catalano, K. K. Smithe, L. Wang, S. Zhang, E. Pop, M. J. Kim, and Z. Chen, "Studies of two-dimensional h-bn and mos2 for potential diffusion barrier application in copper interconnect technology," *npj 2D Materials and Applications*, vol. 1, no. 1, p. 42, 2017.

[55] C.-L. Lo, B. A. Helfrecht, Y. He, D. M. Guzman, N. Onofrio, S. Zhang, D. Weinstein, A. Strachan, and Z. Chen, "Opportunities and challenges of 2D materials in back-end-of-line interconnect scaling," *Journal of Applied Physics*, vol. 128, no. 8, 2020.

[56] M. H. van der Veen, J. Soethoudt, A. Delabie, O. V. Pedreira, V. Vega Gonzalez, S. Larivière, L. Teugels, N. Jourdan, S. Decoster, H. Struyf, C. J. Wilson, K. Croes, and Z. Tőkei, "Hybrid metallization with cu in sub 30nm interconnects," in *2020 IEEE International Interconnect Technology Conference (IITC)*, 2020, pp. 16–18.

[57] M. van der Veen, F. Inoue, K. Vandersmissen, D. Dictus, T. Tanaka, J. Boemmels, H. Struyf, and Z. Tokei, "Conformal cu electroless seed on co and ru liners enables cu fill by plating for advanced interconnects," in *Proc. IEEE Mater. Adv. Metallization Conf.*, 2016, p. 247.

[58] K. Croes, C. Adelmann, C. Wilson, H. Zahedmanesh, O. V. Pedreira, C. Wu, A. Leśniewska, H. Oprins, S. Beyne, I. Ciofi, D. Kocaay, M. Stucchi, and Z. Tőkei, "Interconnect metals beyond copper: reliability challenges and opportunities," in *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018, pp. 5.3.1–5.3.4.

[59] C. Adelmann, L. G. Wen, A. P. Peter, Y. K. Siew, K. Croes, J. Swerts, M. Popovici, K. Sankaran, G. Pourtois, S. Van Elshocht *et al.*, "Alternative metals for advanced interconnects," in *Ieee international interconnect technology conference*. IEEE, 2014, pp. 173–176.

[60] C.-K. Hu, J. Kelly, H. Huang, K. Motoyama, H. Shobha, Y. Ostrovski, J. H. Chen, R. Patlolla, B. Peethala, P. Adusumilli *et al.*, "Future on-chip interconnect metallization and electromigration," in *2018 ieee international reliability physics symposium (irps)*. IEEE, 2018, pp. 4F–1.

[61] O. V. Pedreira, K. Croes, A. Leśniewska, C. Wu, M. H. van der Veen, J. de Messemaeker, K. Vandersmissen, N. Jourdan, L. Wen, C. Adelmann, B. Briggs, V. V.

Gonzalez, J. Bömmels, and Z. Tőkei, "Reliability study on cobalt and ruthenium as alternative metals for advanced interconnects," in *2017 IEEE International Reliability Physics Symposium (IRPS)*, 2017, pp. 6B–2.1–6B–2.8.

[62] D. Gall, "The search for the most conductive metal for narrow interconnect lines," *Journal of Applied Physics*, vol. 127, no. 5, 2020.

[63] M. H. van der Veen, N. Heyler, O. V. Pedreira, I. Ciofi, S. Decoster, V. V. Gonzalez, N. Jourdan, H. Struyf, K. Croes, C. J. Wilson, and Z. Tőkei, "Damascene benchmark of ru, co and cu in scaled dimensions," in *2018 IEEE International Interconnect Technology Conference (IITC)*, 2018, pp. 172–174.

[64] J. Kim, H. Rhee, M. W. Son, J. Park, G. Kim, J. B. Jeon, H. Song, G. Kim, B.-G. Park, J. H. Han *et al.*, "Electromigration reliability of barrierless ruthenium and molybdenum for sub-10 nm interconnection," *ACS Applied Electronic Materials*, vol. 5, no. 5, pp. 2447–2453, 2023.

[65] C.-K. Hu, J. Kelly, J. H.-C. Chen, H. Huang, Y. Ostrovski, R. Patlolla, B. Peethala, P. Adusumilli, T. Spooner, L. M. Gignac, J. Bruley, C. Breslin, S. A. Cohen, G. Lian, M. Ali, R. Long, G. Hornicek, T. Kane, V. Kamineni, X. Zhang, F. Mont, and S. Siddiqui, "Electromigration and resistivity in on-chip cu, co and ru damascene nanowires," in *2017 IEEE International Interconnect Technology Conference (IITC)*, 2017, pp. 1–3.

[66] R. T. Chang, N. Talwalkar, C. P. Yue, and S. S. Wong, "Near speed-of-light signaling over on-chip electrical interconnects," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 5, pp. 834–838, 2003.

[67] M. F. Chang, V. P. Roychowdhury, L. Zhang, H. Shin, and Y. Qian, "Rf/wireless interconnect for inter-and intra-chip communications," *Proceedings of the IEEE*, vol. 89, no. 4, pp. 456–466, 2001.

[68] M.-C. Chang, I. Verbauwhede, C. Chien, Z. Xu, J. Kim, J. Ko, Q. Gu, and B.-C. Lai, "Advanced rf/baseband interconnect schemes for inter-and intra-ulsi communications," *IEEE Transactions on Electron devices*, vol. 52, no. 7, pp. 1271–1285, 2005.

[69] S.-W. Tam, E. Socher, A. Wong, and M.-C. F. Chang, "A simultaneous tri-band on-chip rf-interconnect for future network-on-chip," in *2009 Symposium on VLSI Circuits*, 2009, pp. 90–91.

[70] C. Silvano, M. Lajolo, and G. Palermo, *Low power networks-on-chip*. Springer Science & Business Media, 2010.

[71] S.-B. Lee, S.-W. Tam, I. Pefkianakis, S. Lu, M. F. Chang, C. Guo, G. Reinman, C. Peng, M. Naik, L. Zhang *et al.*, "A scalable micro wireless interconnect structure for cmps," in *Proceedings of the 15th annual international conference on Mobile computing and networking*, 2009, pp. 217–228.

[72] H. Li, C. Xu, N. Srivastava, and K. Banerjee, "Carbon nanomaterials for next-generation interconnects and passives: Physics, status, and prospects," *IEEE Transactions on electron devices*, vol. 56, no. 9, pp. 1799–1821, 2009.

[73] K. Banerjee and N. Srivastava, "Are carbon nanotubes the future of VLSI interconnections?" in *Proceedings of the 43rd annual design automation conference*, 2006, pp. 809–814.

[74] F. Kreupl, A. P. Graham, G. Duesberg, W. Steinhögl, M. Liebau, E. Unger, and W. Hönlein, "Carbon nanotubes in interconnect applications," *Microelectronic Engineering*, vol. 64, no. 1-4, pp. 399–408, 2002.

[75] A. Naeemi and J. D. Meindl, "Design and performance modeling for single-walled carbon nanotubes as local, semiglobal, and global interconnects in gigascale integrated systems," *IEEE Transactions on Electron Devices*, vol. 54, no. 1, pp. 26–37, 2007.

[76] H. Li, W.-Y. Yin, K. Banerjee, and J.-F. Mao, "Circuit modeling and performance analysis of multi-walled carbon nanotube interconnects," *IEEE Transactions on Electron Devices*, vol. 55, no. 6, pp. 1328–1337, 2008.

[77] S. Subash and M. H. Chowdhury, "Mixed carbon nanotube bundles for interconnect applications," *International Journal of Electronics*, vol. 96, no. 6, pp. 657–671, 2009.

[78] S. Pasricha, N. Dutt, and F. J. Kurdahi, "Exploring carbon nanotube bundle global interconnects for chip multiprocessor applications," in *2009 22nd International Conference on VLSI Design*. IEEE, 2009, pp. 499–504.

[79] Soldano, Caterina, Talapatra, Saikat and Kar, Swastik, "Carbon nanotubes and graphene nanoribbons: Potentials for nanoscale electrical interconnects. vol. 2. 2013."

[80] A. Naeemi and J. D. Meindl, "Conductance modeling for graphene nanoribbon (gnr) interconnects," *IEEE electron device letters*, vol. 28, no. 5, pp. 428–431, 2007.

[81] R. Murali, K. Brenner, Y. Yang, T. Beck, and J. D. Meindl, "Resistivity of graphene nanoribbon interconnects," *IEEE Electron Device Letters*, vol. 30, no. 6, pp. 611–613, 2009.

[82] S. Agrawal, A. Srivastava, and G. Kaushal, "N-doped zigzag graphene nanoribbons for nanoscale interconnects," in *2020 IEEE 10th International Conference Nanomaterials: Applications & Properties (NAP)*. IEEE, 2020, pp. 01TPNS03–1.

[83] ——, "Understanding electron transport in halogenated graphene nanoribbons and possible application as interconnects," *Semiconductor Science and Technology*, vol. 37, no. 2, p. 025004, 2021.

[84] ——, "Electron transport in boron functionalised armchair graphene nanoribbons: Potential interconnects," *Solid State Communications*, vol. 327, p. 114209, 2021.

[85] S. Agrawal, A. Srivastava, G. Kaushal, and A. Srivastava, "Edge engineered graphene nanoribbons as nanoscale interconnect: Dft analysis," *IEEE Transactions on Nanotechnology*, vol. 21, pp. 43–51, 2022.

[86] C. Xu, H. Li, and K. Banerjee, "Graphene nano-ribbon (gnr) interconnects: A genuine contender or a delusive dream?" in *2008 IEEE International Electron Devices Meeting*. IEEE, 2008, pp. 1–4.

[87] A. Naeemi and J. D. Meindl, "Conductance modeling for graphene nanoribbon (gnr) interconnects," *IEEE Electron Device Letters*, vol. 28, no. 5, pp. 428–431, 2007.

[88] ——, "Compact physics-based circuit models for graphene nanoribbon interconnects," *IEEE Transactions on Electron Devices*, vol. 56, no. 9, pp. 1822–1833, 2009.

[89] D.-Y. Jeon, K. J. Lee, M. Kim, D. C. Kim, H.-J. Chung, Y.-S. Woo, and S. Seo, "Radio-frequency electrical characteristics of single layer graphene," *Japanese journal of applied physics*, vol. 48, no. 9R, p. 091601, 2009.

[90] D. Sarkar, C. Xu, H. Li, and K. Banerjee, "High-frequency behavior of graphene-based interconnects—part i: Impedance modeling," *IEEE Transactions on Electron Devices*, vol. 58, no. 3, pp. 843–852, 2011.

[91] ——, "High-frequency behavior of graphene-based interconnects—part ii: Impedance analysis and implications for inductor design," *IEEE Transactions on Electron Devices*, vol. 58, no. 3, pp. 853–859, 2011.

[92] C. Xu, H. Li, and K. Banerjee, "Modeling, analysis, and design of graphene nanoribbon interconnects," *IEEE Transactions on Electron Devices*, vol. 56, no. 8, pp. 1567–1578, 2009.

[93] T. Kaur, M. K. Rai, and R. Khanna, "Effect of temperature on the performance analysis of mlgnr interconnects," *Journal of Computational Electronics*, vol. 18, pp. 722–736, 2019.

[94] J.-P. Cui, W.-S. Zhao, W.-Y. Yin, and J. Hu, "Signal transmission analysis of multilayer graphene nano-ribbon (mlgnr) interconnects," *IEEE Transactions on Electromagnetic Compatibility*, vol. 54, no. 1, pp. 126–132, 2012.

[95] S. H. Nasiri, M. K. Moravvej-Farshi, and R. Faez, "Stability analysis in graphene nanoribbon interconnects," *IEEE Electron Device Letters*, vol. 31, no. 12, pp. 1458–1460, 2010.

[96] V. Kumar, S. Rakheja, and A. Naeemi, "Performance and energy-per-bit modeling of multilayer graphene nanoribbon conductors," *IEEE transactions on electron devices*, vol. 59, no. 10, pp. 2753–2761, 2012.

[97] S. Das, S. Bhattacharya, D. Das, and H. Rahaman, "Thermal stability analysis of graphene nano-ribbon interconnect and applicability for terahertz frequency," *National Academy Science Letters*, vol. 43, no. 3, pp. 253–257, 2020.

[98] D. Kondo, H. Nakano, B. Zhou, A. I, K. Hayashi, M. Takahashi, S. Sato, and N. Yokoyama, "Sub-10-nm-wide intercalated multi-layer graphene interconnects with low resistivity," in *IEEE International Interconnect Technology Conference*, 2014, pp. 189–192.

[99] J. Jiang, J. Kang, W. Cao, X. Xie, H. Zhang, J. H. Chu, W. Liu, and K. Banerjee, "Intercalation doped multilayer-graphene-nanoribbons for next-generation interconnects," *Nano letters*, vol. 17, no. 3, pp. 1482–1488, 2017.

[100] F. Zou, Z. Pan, and P. Xu, "Modeling and performance analysis of coupled multilayer graphene nanoribbon (mlgnr) interconnects with intercalation doping," *Microelectronics Journal*, vol. 141, p. 105971, 2023.

[101] S. Agrawal, A. Srivastava, and G. Kaushal, "Performance metrics of doped graphene nanoribbons basednanoscale interconnects: Dft and negf analysis," *Solid State Communications*, p. 115389, 2023.

[102] K. Hosono and K. Wakabayashi, "Dielectric environment effect on carrier mobility of graphene double-layer structure," *Applied Physics Letters*, vol. 103, no. 3, 2013.

[103] A. K. Nishad and R. Sharma, "Performance improvement in sc-mlgnrs interconnects using interlayer dielectric insertion," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 4, pp. 470–482, 2015.

[104] P. Xu and Z. Pan, "The insertion of high-k dielectric materials in multilayer graphene nanoribbon interconnects for reducing propagation delay and expanding bandwidth," *Organic Electronics*, vol. 78, p. 105607, 2020.

[105] ——, "Collaborative applying the ultra-low-k dielectric and the high-k dielectric materials for performance enhancement in coupled multilayer graphene nanoribbon interconnects," *IEEE Journal of the Electron Devices Society*, vol. 8, pp. 200–212, 2020.

[106] Sidhu, Ramneek and Rai, Mayank Kumar, "Electronic transport in doped and dielectric inserted mlgnr interconnects: Crosstalk induced delay and stability analyses at sub-threshold regime," *Microelectronics Journal*, vol. 128, p. 105524, 2022.

[107] P. Goli, H. Ning, X. Li, C. Y. Lu, K. S. Novoselov, and A. A. Balandin, "Thermal properties of graphene–copper–graphene heterogeneous films," *Nano letters*, vol. 14, no. 3, pp. 1497–1503, 2014.

[108] L. Li, Z. Zhu, A. Yoon, and H.-S. P. Wong, "In-situ grown graphene enabled copper interconnects with improved electromigration reliability," *IEEE Electron Device Letters*, vol. 40, no. 5, pp. 815–817, 2019.

[109] H. C. Lee, M. Jo, H. Lim, M. S. Yoo, E. Lee, N. N. Nguyen, S. Y. Han, and K. Cho, "Toward near-bulk resistivity of cu for next-generation nano-interconnects: Graphene-coated cu," *Carbon*, vol. 149, pp. 656–663, 2019.

[110] M. Son, J. Jang, Y. Lee, J. Nam, J. Y. Hwang, I. S. Kim, B. H. Lee, M.-H. Ham, and S.-S. Chee, "Copper-graphene heterostructure for back-end-of-line compatible high-performance interconnects," *npj 2D Materials and Applications*, vol. 5, no. 1, p. 41, 2021.

[111] K. W. Shin, Y. Cho, S.-G. Nam, A. Jung, E.-K. Lee, C.-S. Lee, M.-H. Lee, H.-J. Shin, and K.-E. Byun, "Graphene capping of cu back-end-of-line interconnects reduces resistance and improves electromigration lifetime," *ACS Applied Nano Materials*, vol. 6, no. 6, pp. 4170–4177, 2023.

[112] R. King, R. Michalzik, D. Wiedenmann, R. Jaeger, P. Schnitzer, T. Knödl, and K. J. Ebeling, "2D VCSEL arrays for chip-level optical interconnects," in *Optoelectronic Interconnects VI*, vol. 3632. SPIE, 1999, pp. 363–372.

[113] H. Rong, R. Jones, A. Liu, O. Cohen, D. Hak, A. Fang, and M. Paniccia, "A continuous-wave raman silicon laser," *Nature*, vol. 433, no. 7027, pp. 725–728, 2005.

[114] P. M. Fauchet, "Light emission from si quantum dots," *materials today*, vol. 8, no. 1, pp. 26–33, 2005.

[115] L. Pavesi, "Routes toward silicon-based lasers," *Materials Today*, vol. 8, no. 1, pp. 18–25, 2005.

[116] G. Roelkens, L. Liu, D. Liang, R. Jones, A. Fang, B. Koch, and J. Bowers, "Iii-v/silicon photonics for on-chip and intra-chip optical interconnects," *Laser & Photonics Reviews*, vol. 4, no. 6, pp. 751–779, 2010.

[117] W. M. Green, M. J. Rooks, L. Sekaric, and Y. A. Vlasov, "Ultra-compact, low rf power, 10 gb/s silicon mach-zehnder modulator," *Optics express*, vol. 15, no. 25, pp. 17 106–17 113, 2007.

[118] A. Liu, L. Liao, D. Rubin, J. Basak, Y. Chetrit, H. Nguyen, R. Cohen, N. Izhaky, and M. Paniccia, "Recent development in a high-speed silicon optical modulator based on reverse-biased pn diode in a silicon waveguide," *Semiconductor Science and technology*, vol. 23, no. 6, p. 064001, 2008.

[119] H. Jia, L. Zhang, J. Ding, L. Zheng, C. Yuan, and L. Yang, "Microring modulator matrix integrated with mode multiplexer and de-multiplexer for on-chip optical interconnect," *Optics Express*, vol. 25, no. 1, pp. 422–430, 2017.

[120] A. Sakai, H. Go, and T. Baba, "Sharply bent optical waveguide silicon-on-insulator substrate," in *Physics and Simulation of Optoelectronic Devices IX*, vol. 4283. SPIE, 2001, pp. 610–618.

[121] E. Cassan, S. Laval, S. Lardenois, and A. Koster, "On-chip optical interconnects with compact and low-loss light distribution in silicon-on-insulator rib waveguides," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 9, no. 2, pp. 460–464, 2003.

[122] G. Mashanovich, M. Milosevic, P. Matavulj, S. Stankovic, B. Timotijevic, P. Yang, E. Teo, M. Breese, A. Bettiol, and G. Reed, "Silicon photonic waveguides for different wavelength regions," *Semiconductor science and technology*, vol. 23, no. 6, p. 064002, 2008.

[123] S. Sahni, X. Luo, J. Liu, Y.-h. Xie, and E. Yablonovitch, "Junction field-effect-transistor-based germanium photodetector on silicon-on-insulator," *Optics letters*, vol. 33, no. 10, pp. 1138–1140, 2008.

[124] L. Tang, S. E. Kocabas, S. Latif, A. K. Okyay, D.-S. Ly-Gagnon, K. C. Saraswat, and D. A. Miller, "Nanometre-scale germanium photodetector enhanced by a near-infrared dipole antenna," *Nature Photonics*, vol. 2, no. 4, pp. 226–229, 2008.

[125] S. Xu, W. Wang, Y.-C. Huang, Y. Dong, S. Masudy-Panah, H. Wang, X. Gong, and Y.-C. Yeo, "High-speed photo detection at two-micron-wavelength: technology

enablement by gesn/ge multiple-quantum-well photodiode on 300 mm si substrate," *Optics express*, vol. 27, no. 4, pp. 5798–5813, 2019.

[126] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. P. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. G. Beausoleil, and J. H. Ahn, "Corona: System implications of emerging nanophotonic technology," *ACM SIGARCH Computer Architecture News*, vol. 36, no. 3, pp. 153–164, 2008.

[127] Y. Pan, J. Kim, and G. Memik, "Flexishare: Channel sharing for an energy-efficient nanophotonic crossbar," in *HPCA-16 2010 The Sixteenth International Symposium on High-Performance Computer Architecture*. IEEE, 2010, pp. 1–12.

[128] A. Joshi, C. Batten, Y.-J. Kwon, S. Beamer, I. Shamim, K. Asanovic, and V. Stojanovic, "Silicon-photonic clos networks for global on-chip communication," in *2009 3rd ACM/IEEE International Symposium on Networks-on-Chip*. IEEE, 2009, pp. 124–133.

[129] R. W. Morris Jr and A. K. Kodi, "Power-efficient and high-performance multi-level hybrid nanophotonic interconnect for multicores," in *2010 Fourth ACM/IEEE International Symposium on Networks-on-Chip*. IEEE, 2010, pp. 207–214.

[130] C. Li, M. Browning, P. V. Gratz, and S. Palermo, "Luminoc: A power-efficient, high-performance, photonic network-on-chip," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, no. 6, pp. 826–838, 2014.

[131] A. Shacham, B. G. Lee, A. Biberman, K. Bergman, and L. P. Carloni, "Photonic noc for dma communications in chip multiprocessors," in *15th Annual IEEE Symposium on High-Performance Interconnects (HOTI 2007)*, 2007, pp. 29–38.

[132] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, and A. Choudhary, "Firefly: Illuminating future network-on-chip with nanophotonics," in *Proceedings of the 36th annual international symposium on Computer architecture*, 2009, pp. 429–440.

[133] S. Bahirat and S. Pasricha, "Meteor: Hybrid photonic ring-mesh network-on-chip for multicore architectures," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 13, no. 3s, pp. 1–33, 2014.

[134] K. H. Mo, Y. Ye, X. Wu, W. Zhang, W. Liu, and J. Xu, "A hierarchical hybrid optical-electronic network-on-chip," in *2010 IEEE Computer Society Annual Symposium on VLSI*. IEEE, 2010, pp. 327–332.

[135] S. Pasricha and S. Bahirat, "Opal: A multi-layer hybrid photonic noc for 3d ics," in *16th Asia and South Pacific Design Automation Conference (ASP-DAC 2011)*. IEEE, 2011, pp. 345–350.

[136] A. B. Ahmed and A. B. Abdallah, "Phenic: silicon photonic 3d-network-on-chip architecture for high-performance heterogeneous many-core system-on-chip," in *14th International Conference on Sciences and Techniques of Automatic Control & Computer Engineering - STA'2013*, 2013, pp. 1–9.

[137] H. M. G. Wassel, D. Dai, M. Tiwari, J. K. Valamehr, L. Theogarajan, J. Dionne, F. T. Chong, and T. Sherwood, "Opportunities and challenges of using plasmonic components in nanophotonic architectures," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 2, pp. 154–168, 2012.

[138] V. K. Narayana, S. Sun, A. Mehrabian, V. J. Sorger, and T. El-Ghazawi, "Hyppi noc: Bringing hybrid plasmonics to an opto-electronic network-on-chip," in *2017 46th International Conference on Parallel Processing (ICPP)*. IEEE, 2017, pp. 131–140.

[139] J. A. Dionne, L. A. Sweatlock, M. T. Sheldon, A. P. Alivisatos, and H. A. Atwater, "Silicon-based plasmonics for on-chip photonics," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 1, pp. 295–306, 2010.

[140] K. Liu, N. Li, D. K. Sadana, and V. J. Sorger, "Integrated nanocavity plasmon light sources for on-chip optical interconnects," *ACS Photonics*, vol. 3, no. 2, pp. 233–242, 2016.

[141] Y. Liu, L. Ding, Y. Cao, D. Wan, G. Yuan, B. Huang, A. Voon-Yew Thean, T. Mei, T. Venkatesan, C. A. Nijhuis, and C. Soo-Jin, "The design of cmos-compatible plasmonic waveguides for intra-chip communication," *IEEE Photonics Journal*, vol. 12, no. 5, pp. 1–10, 2020.

[142] Z. Ge, L. Zhang, G. Wang, W. Zhang, M. Liu, S. Li, L. Wang, Q. Sun, W. Ren, J. Si, and W. Zhao, "On-chip router elements based on silicon hybrid plasmonic waveguide," *IEEE Photonics Technology Letters*, vol. 29, no. 12, pp. 952–955, 2017.

[143] M. Fukuda, Y. Tonooka, T. Inoue, and M. Ota, "Feasibility of plasmonic circuits for on-chip interconnects," *Solid-State Electronics*, vol. 156, pp. 33–40, 2019.

[144] X. Wang and B. Sensale-Rodriguez, "Graphene based optical interconnects," in *Nanoelectronics*. Elsevier, 2019, pp. 271–285.

[145] Z. Xing, C. Li, Y. Han, H. Hu, Z. Cheng, and T. Liu, "Design of on-chip polarizers based on graphene-on-silicon nanowires," *Applied Physics Express*, vol. 12, no. 7, p. 072001, 2019.

[146] B. Choubey, V. R. Kumbhare, and M. K. Majumder, "Performance analysis of graphene based optical interconnect at nanoscale technology," in *VLSI Design and Test: 22nd International Symposium, VDAT 2018, Madurai, India, June 28-30, 2018, Revised Selected Papers 22*. Springer, 2019, pp. 418–429.

[147] J. Wang, X. Zhang, Y. Chen, Y. Geng, Y. Du, and X. Li, "Design of a graphene-based silicon nitride multimode waveguide-integrated electro-optic modulator," *Optics Communications*, vol. 481, p. 126531, 2021.

[148] S. Deb, A. Ganguly, K. Chang, P. Pande, B. Beizer, and D. Heo, "Enhancing performance of network-on-chip architectures with millimeter-wave wireless interconnects," in *ASAP 2010-21st IEEE International Conference on Application-specific Systems, Architectures and Processors*. IEEE, 2010, pp. 73–80.

[149] B. A. Floyd, C.-M. Hung, and O. KK, "Intra-chip wireless interconnect for clock distribution implemented with integrated antennas, receivers, and transmitters," *IEEE Journal of solid-state circuits*, vol. 37, no. 5, pp. 543–552, 2002.

[150] N. Srivastava and K. Banerjee, "Performance analysis of carbon nanotube interconnects for VLSI applications," in *ICCAD-2005. IEEE/ACM International Conference on Computer-Aided Design, 2005.* IEEE, 2005, pp. 383–390.

[151] S. Rakheja, V. Kumar, and A. Naeemi, "Evaluation of the potential performance of graphene nanoribbons as on-chip interconnects," *Proceedings of the IEEE*, vol. 101, no. 7, pp. 1740–1765, 2013.

[152] I. Llatser, S. Abadal, A. M. Sugrañes, A. Cabellos-Aparicio, and E. Alarcón, "Graphene-enabled wireless networks-on-chip," in *2013 First International Black Sea Conference on Communications and Networking (BlackSeaCom)*, 2013, pp. 69–73.

[153] Q.-T. Vien, M. O. Agyeman, T. A. Le, T. Mak *et al.*, "On the nanocommunications at thz band in graphene-enabled wireless network-on-chip," *Mathematical Problems in Engineering*, vol. 2017, 2017.

[154] S. Saxena, D. S. Manur, M. S. Shamim, and A. Ganguly, "A folded wireless network-on-chip using graphene based thz-band antennas," in *Proceedings of the 4th acm international conference on nanoscale computing and communication*, 2017, pp. 1–6.

[155] S. Abadal, R. Guirado, H. Taghvaee, A. Jain, E. P. de Santana, P. H. Bolivar, M. Saeed, R. Negra, Z. Wang, K.-T. Wang *et al.*, "Graphene-based wireless agile interconnects for massive heterogeneous multi-chip processors," *IEEE wireless communications*, 2022.

[156] D. Zhao, Y. Wang, J. Li, and T. Kikkawa, "Design of multi-channel wireless noc to improve on-chip communication capacity," in *Proceedings of the fifth ACM/IEEE International Symposium on Networks-on-Chip*, 2011, pp. 177–184.

[157] H. I. Gaha and M. Balti, "Novel bi-uwb on-chip antenna for wireless noc," *Micromachines*, vol. 13, no. 2, p. 231, 2022.

[158] P. P. Pande, A. Ganguly, K. Chang, and C. Teuscher, "Hybrid wireless network on chip: A new paradigm in multi-core design," in *Proceedings of the 2nd international Workshop on Network on Chip Architectures*, 2009, pp. 71–76.

[159] A. Ganguly, K. Chang, S. Deb, P. P. Pande, B. Belzer, and C. Teuscher, "Scalable hybrid wireless network-on-chip architectures for multicore systems," *IEEE Transactions on Computers*, vol. 60, no. 10, pp. 1485–1502, 2010.

[160] A. Rezaei, M. Daneshtalab, F. Safaei, and D. Zhao, "Hierarchical approach for hybrid wireless network-on-chip in many-core era," *Computers & Electrical Engineering*, vol. 51, pp. 225–234, 2016.

[161] B. Bahrami, M. A. Jabraeil Jamali, and S. Saeidi, "A novel hierarchical architecture for wireless network-on-chip," *Journal of Parallel and Distributed Computing*, vol. 120, pp. 307–321, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0743731518301102

[162] A. Dehghani and K. RahimiZadeh, "Design and performance evaluation of mesh-of-tree-based hierarchical wireless network-on-chip for multicore systems," *Journal of Parallel and Distributed Computing*, vol. 123, pp. 100–117, 2019.

[163] D. Neculoiu, A. Muller, K. Tang, E. Laskin, and S. Voinigescu, "160 ghz on-chip dipole antenna structure in silicon technology," in *2007 International Semiconductor Conference*, vol. 1, 2007, pp. 245–248.

[164] A. Samaiyar, S. S. Ram, and S. Deb, "Millimeter-wave planar log periodic antenna for on-chip wireless interconnects," in *The 8th European Conference on Antennas and Propagation (EuCAP 2014)*. IEEE, 2014, pp. 1007–1009.

[165] A. Nojeh and A. Ivanov, "Wireless interconnect and the potential for carbon nanotubes," *IEEE Design & Test of Computers*, vol. 27, no. 4, pp. 44–53, 2010.

[166] H. M. Cheema and A. Shamim, "The last barrier: on-chip antennas," *IEEE Microwave Magazine*, vol. 14, no. 1, pp. 79–91, 2013.

[167] A. A. Althuwayb, "On-chip antenna design using the concepts of metamaterial and siw principles applicable to terahertz integrated circuits operating over 0.6–0.622 thz," *International Journal of Antennas and Propagation*, vol. 2020, pp. 1–9, 2020.

[168] I. E. Masri, P.-M. Martin, H. K. Mondal, R. Allanic, T. Le Gouguec, C. Quendo, C. Roland, and J.-P. Diguet, "Accurate channel models for realistic design space exploration of future wireless nocs," in *2018 Twelfth IEEE/ACM International Symposium on Networks-on-Chip (NOCS)*, 2018, pp. 1–8.

[169] Y. Chen and C. Han, "Channel modeling and characterization for wireless networks-on-chip communications in the millimeter wave and terahertz bands," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 5, no. 1, pp. 30–43, 2019.

[170] J. Wang, C.-X. Wang, J. Huang, and Y. Chen, "6g thz propagation channel characteristics and modeling: Recent developments and future challenges," *IEEE Communications Magazine*, pp. 1–8, 2022.

[171] D. Zhao and Y. Wang, "Sd-mac: Design and synthesis of a hardware-efficient collision-free qos-aware mac protocol for wireless network-on-chip," *IEEE Transactions on Computers*, vol. 57, no. 9, pp. 1230–1245, 2008.

[172] N. Mansoor and A. Ganguly, "Reconfigurable wireless network-on-chip with a dynamic medium access mechanism," in *Proceedings of the 9th International Symposium on Networks-on-Chip*, 2015, pp. 1–8.

[173] S. Abadal, A. Mestres, J. Torrellas, E. Alarcon, and A. Cabellos-Aparicio, "Medium access control in wireless network-on-chip: A context analysis," *IEEE Communications Magazine*, vol. 56, no. 6, pp. 172–178, 2018.

[174] Y. Ouyang, Y. Zhao, K. Xing, Z. Huang, H. Liang, and J. Li, "Design of wireless network on chip with priority-based mac," *Journal of Circuits, Systems and Computers*, vol. 28, no. 08, p. 1950124, 2019.

[175] Q. Gao, W. Song, Z. Lu, L. Li, and Y. Fu, "Dynamic and traffic-aware medium access control mechanisms for wireless noc architectures," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021, pp. 1–5.

[176] S. S. Rout, M. Sinha, and S. Deb, "2dmac: A sustainable and efficient medium access control mechanism for future wireless nocs," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 19, no. 3, pp. 1–25, 2023.

[177] S. Abadal, E. Alarcón, A. Cabellos-Aparicio, M. C. Lemme, and M. Nemirovsky, "Graphene-enabled wireless communication for massive multicore architectures," *IEEE Communications Magazine*, vol. 51, no. 11, pp. 137–143, 2013.

[178] P. Bharadwaj, B. Deutsch, and L. Novotny, "Optical antennas," *Advances in Optics and Photonics*, vol. 1, no. 3, pp. 438–483, 2009.

[179] A. Alù and N. Engheta, "Wireless at the nanoscale: optical interconnects using matched nanoantennas," *Physical review letters*, vol. 104, no. 21, p. 213902, 2010.

[180] M. Nafari, L. Feng, and J. M. Jornet, "On-chip wireless optical channel modeling for massive multi-core computing architectures," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2017, pp. 1–6.

[181] J. S. Dehkordi, V. Tralli, M. Barbiroli, J. Nanni, F. Fuschini, and V. Petruzzelli, "Analysis, design and performance evaluation of on-chip optical wireless links," in *Proc. ICQNM*, 2019, pp. 1–6.

[182] T. Sangeeta and J. Deny, "A study on optical interconnects to improve on-chip wireless communication using plasmonic nanoantennas and seeking dielectric nanoantenna as an alternative," in *2022 IEEE International Conference on Nanoelectronics, Nanophotonics, Nanomaterials, Nanobioscience & Nanotechnology (5NANO)*. IEEE, 2022, pp. 1–5.

[183] A. J. Karkar, J. E. Turner, K. Tong, R. AI-Dujaily, T. Mak, A. Yakovlev, and F. Xia, "Hybrid wire-surface wave interconnects for next-generation networks-on-chip," *IET Computers & Digital Techniques*, vol. 7, no. 6, pp. 294–303, 2013.

[184] A. Karkar, R. Al-Dujaily, A. Yakovlev, K. Tong, and T. Mak, "Surface wave communication system for on-chip and off-chip interconnects," in *Proceedings of the Fifth International Workshop on Network on Chip Architectures*, 2012, pp. 11–16.

[185] A. Karkar, N. Dahir, R. Al-Dujaily, K. Tong, T. Mak, and A. Yakovlev, "Hybrid wire-surface wave architecture for one-to-many communication in networks-on-chip," in *2014 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2014, pp. 1–4.

[186] M. O. Agyeman, J.-X. Wan, Q.-T. Vien, W. Zong, A. Yakovlev, K. Tong, and T. Mak, "On the design of reliable hybrid wired-wireless network-on-chip architectures," in *2015 IEEE 9th International Symposium on Embedded Multicore/Many-core Systems-on-Chip*, 2015, pp. 251–258.

[187] A. Karkar, T. Mak, N. Dahir, K.-F. Tong, A. Yakovlev *et al.*, "Network-on-chip multicast architectures using hybrid wire and surface-wave interconnects," *IEEE Transactions on Emerging Topics in Computing*, vol. 6, no. 3, pp. 357–369, 2016.

[188] A. Karkar, T. Mak, K.-F. Tong, and A. Yakovlev, "A survey of emerging interconnects for on-chip efficient multicast and broadcast in many-cores," *IEEE Circuits and Systems Magazine*, vol. 16, no. 1, pp. 58–72, 2016.

[189] C.-K. Koh and P. H. Madden, "Manhattan or non-Manhattan? A study of alternative VLSI routing architectures," in *Proceedings of the 10th Great Lakes symposium on VLSI*, 2000, pp. 47–52.

[190] S. L. Teig, "The x architecture: Not your father's diagonal wiring," in *Proceedings of the 2002 international workshop on System-level interconnect prediction*, 2002, pp. 33–37.

[191] S. Bothra, "Hexagonal semiconductor die, semiconductor substrates, and methods of forming a semiconductor die," Feb. 29 2000, uS Patent 6,030,885.

[192] M. D. Rostoker, J. S. Koford, R. Scepanovic, E. R. Jones, G. R. Padmanahben, A. K. Kapoor, V. B. Kudryavtsev, A. E. Andreev, S. V. Aleshin, A. S. Podkolzin *et al.*, "Hexagonal architecture," Jun. 18 2002, uS Patent 6,407,434.

[193] S. Teig, "Non manhattan floor plan architecture for integrated circuits," Mar. 17 2009, uS Patent 7,506,295.

[194] C. Chiang and M. Sarrafzadeh, "Wirability of knock-knee layouts with 45 degrees wires," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 6, pp. 613–624, 1991.

[195] B. Yang, J. Yu, T. Yan, and J. Li, "A non-manhattan hexagon/triangle placement paradigm," in *2004 International Conference on Communications, Circuits and Systems (IEEE Cat. No.04EX914)*, 2004, pp. 1271–1275 Vol.2.

[196] J. M. Joseph, L. Bamberg, I. Hajjar, B. R. Perjikolaei, A. G. Ortiz, and T. Pionteck, "Ratatoskr: An open-source framework for in-depth power, performance and area analysis in 3d nocs," *CoRR*, vol. abs/1912.05670, 2019.

[197] https://www.xilinx.com/support/download.html, "Xilinx vivado," June.

[198] N. Jiang, D. U. Becker, G. Michelogiannakis, J. Balfour, B. Towles, D. E. Shaw, J. Kim, and W. J. Dally, "A detailed and flexible cycle-accurate network-on-chip simulator," IEEE, pp. 86–96, 2013, iEEE international symposium on performance analysis of systems and software (ISPASS).

[199] A. B. Kahng, B. Lin, and S. Nath, "Orion3. 0: A comprehensive noc router estimation tool," *IEEE Embedded Systems Letters*, vol. 7, no. 2, pp. 41–45, 2015.

[200] https://www.cs.utexas.edu/ netrace/, "Parsec."

[201] V. F. Pavlidis and E. G. Friedman, "3-D topologies for networks-on-chip," *IEEE transactions on very large scale integration (VLSI) systems*, vol. 15, no. 10, pp. 1081–1090, 2007.

[202] W. R. Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A. M. Sule, M. Steer, and P. D. Franzon, "Demystifying 3d ics: The pros and cons of going vertical," *IEEE design & test of computers*, vol. 22, no. 6, pp. 498–510, 2005.

[203] V. Catania, A. Mineo, S. Monteleone, M. Palesi, and D. Patti, "Noxim: An open, extensible and cycle-accurate network on chip simulator," in *2015 IEEE 26th international conference on application-specific systems, architectures and processors (ASAP)*. IEEE, 2015, pp. 162–163.

[204] S. Kundu and S. Chattopadhyay, *Network-on-chip: the next generation of system-on-chip integration*. Taylor & Francis, 2014.

[205] S. Deb, K. Chang, X. Yu, S. P. Sah, M. Cosic, A. Ganguly, P. P. Pande, B. Belzer, and D. Heo, "Design of an energy-efficient cmos-compatible noc architecture with millimeter-wave wireless interconnects," *IEEE Transactions on Computers*, vol. 62, no. 12, pp. 2382–2396, 2012.