A 3D Mesh NoC Architecture with Supplementary 2D Links and Progressive Shortest-Path Routing

Abhijit Biswas*, Sourish Dhar

Department of Computer Science and Engineering, TSSOT, Assam University, Silchar, Assam 788011

abhi.021983@gmail.com, dharsourish@gmail.com

Abstract

In this paper, we study a mesh-based 3D network topology that is enhanced with a small number of additional 2D links. The 3D network is formed by vertically stacking and scaling identical 2D mesh planes along the Z-axis. Within each 2D plane, four supplementary links are added to connect the corner nodes to nodes located in the central region of the mesh. These supplementary links are activated only when the 2D routing distance exceeds N in an $N \times N \times N$ 3D mesh $N >= 4$. Under this condition, the added links consistently provide a shorter path, requiring fewer than N hops within the 2D plane, thereby reducing the overall hop count. The proposed network is evaluated using a progressive shortest-path routing algorithm. This algorithm computes the minimum 2D distance from the source node to candidate nodes equipped with supplementary links, as well as the distance from those nodes to the destination. The node that minimizes the combined 2D distance is selected, and the packet is routed through it. If the source and destination nodes are located in the same 2D plane, the packet is routed entirely within that plane. Otherwise, once the packet reaches the selected node, it is forwarded along the Z-axis, either upward or downward, to the destination plane, where it is finally routed to the target node. The results show the robustness of the proposed network using the link utilisation graph.

Keywords: *Network On Chip, System On Chip, Latency, Link Utilization, 3D NoC, Routing Algorithm.*

1. Introduction

The demand for consumer electronics is constantly pushing boundaries, requiring ever-increasing complexity and processing power within limited die space. This led to the era of the System-on-Chip (SoC), where vast numbers of components are realized onto a single piece of silicon. SoCs are the main functional blocks powering diverse applications, from communication and entertainment devices to consumer electronics [1]. However, designing and implementing these intricate systems presents significant hurdles, particularly the immense pressure to shorten time-to-market.

Traditionally, communication within these SoCs relied on rudimentary methods like shared buses or crossbar switches. Being crude, these interconnect architectures began to show their limitations as networks scaled up [2]. Shared buses, for instance, inevitably become performance bottlenecks, unable to handle the escalating bandwidth demands of modern many-core systems. Furthermore, in traditional two-dimensional (2D) chip layouts, creating dedicated point-to-point connections between numerous cores, though very efficient [3], consumes considerable power and occupies a considerable portion of valuable die area, and is also not scalable[1].

Recognising these scalability challenges, researchers in the early 2000s proposed a fundamental shift in how communication happens on a chip: the Network-on-Chip (NoC) paradigm [4]. Many popular 2D topologies for NoC have been proposed, like Mesh, Torus,

© The Author(s), under exclusive license to Digital Manuscriptpedia. 2026 Ashok Kumar et al. (eds.), Proceedings of IMPACT-2026, DMPedia Lecture Notes in Computer Science & Engineering. ISBN: 978-81-993813-9-1

hypercube, BFT (Butterfly Fat Tree) topology, etc. However, due to inherent shortcomings such as high network diameter and end-to-end delay, researchers have been compelled to explore more contemporary topologies that can mitigate these issues.

Among the various two-dimensional Network-on-Chip (NoC) topologies, mesh and torus architectures are the most widely adopted in commercial designs, owing to their structural simplicity, scalability, and ease of implementation [5].

However, the mesh topology suffers from several inherent limitations, including a comparatively large network diameter, marginal bandwidth at edge nodes, and a high hop count when communicating between diagonally opposite corner source–destination pairs. To deal with some of these issues, the torus topology was introduced. By adding wrap-around links that connect the edge nodes of a mesh, the torus maintains a uniform node degree across the network and improves path diversity. Nevertheless, torus networks also present limitations, such as wrap-around links spanning the entire die, which result in higher wire delays and, in certain communication scenarios, can still yield significant hop counts [6, 7].

Many initiatives have been taken to reduce the network diameter; however, attempts to reduce the 2D network diameter invariably introduce additional links between corner nodes or to nodes in the middle. As is seen in the case of 2D mesh. Many variants, such as 2D torus, diagonally connected mesh, C2 mesh, C mesh, diametrical mesh, XD mesh and torus, heavily rely on adding supplementary links [8]. These links do somewhat solve the issue at the cost of scalability in the 2D plane.

We propose evaluating a mesh-based 3D topology augmented with supplementary 2D links, formed by vertically stacking and scaling the network along the Z-axis. Four supplementary links connect the four corner nodes to the four central (middle-region) nodes in each 2D mesh plane. These links are invoked when the 2D routing distance exceeds N in an NNN 3D mesh ($N \geq 4$), as they always provide a shortest path of less than N hops in the 2D plane.

The paper is structured into four main sections. In Section 2, we review the relevant literature for this work. Section 3 introduces the 3D mesh-based topology under consideration and explains the routing algorithm used. Section 4 presents and analyzes the results from our simulations, and finally, Section 5 offers the conclusions drawn from this study.

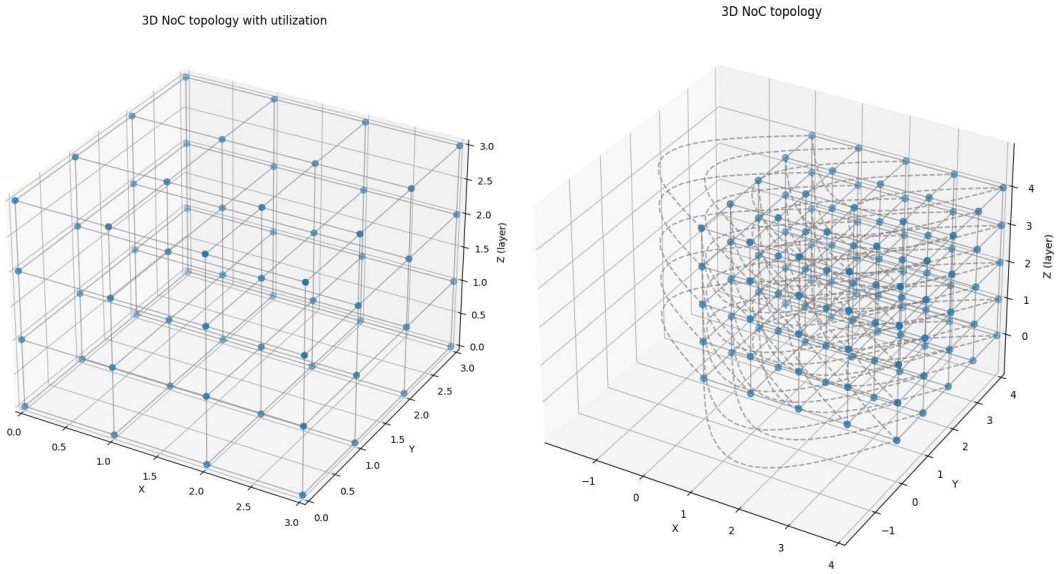
2. Literature Review

Owing to the structural advantages of the Network-on-Chip paradigm over older bus-based systems, a purely two-dimensional structure eventually exposed its own limitations, especially as chip makers continued to push the boundaries of integration density. These limitations became the bedrock of strong motivation for researchers to explore the potential of building NoCs in three dimension, some of the inherent limitations are listed as follows [1, 9, 10, 11, 12, 13, 14, 15]:

- Increasing Wire Length and Latency.
- Scalability issues on the same plane.
- Floor planning limitations.
- Power and Heat generation problem.

The prospect of overcoming the inherent limitations of 2D designs mentioned above has become the crucial motivation for the research on 3D NoC architectures. By extending networks into the third dimension, these architectures offer a more compact and flexible design space, making them an attractive direction for continued exploration and development [16, 2, 17].

Among the most popular 3D NoC is the 3D mesh which is a direct and intuitive extension of the popular 2D mesh topology into the third dimension, facilitated by TSV technology for vertical links which is characterized by arranging the nodes in a 3D grid (*XYZ*), each of the node is connected to six neighbors: East, West, North, South, Up, Down except for the edge nodes which are connected in east, South, Up and Down [16, 18, 4]. This topology, however suffers from a large network diameter and also suffers from congestion towards the middle router [19]. To alleviate some of the limitations mesh topology, 3D torus have been proposed which is again an extension of the 2D torus [20]. However, wrap around link poses the risk of cyclic dependencies which a perfect recipe for deadlock, also the long wrap-around links incur long wire delays [21]. The Figure 4 depicts the 3D mesh and tori.



(a) 3D Mesh (b) 3D Torus

Figure 1: Mesh and Torus NoC topologies

Other significant 3D topologies include Hyper-mesh network QM [20], which is highly scalable but have almost the same set of disadvantages as 3D Mesh or 3D torus, XNoTs (Crossbar Connected Networks-on-Tiers) / Dimensionally-Decomposed Routers which is characterized as Each tier has PEs/routers connected by high-radix switches. Vertical links connect routers across tiers, but it did suffer from scalability issues [2]. One of the most interesting 3D topologies is 3D Small World (SWNoC), which is characterized as a regular base topology (like 3D mesh) with a few long-range “shortcut” links (typically vertical TSVs) to reduce average path length [22]. Although it reduces the network diameter but, the

irregularity becomes a bottle in designing an efficient routing path and also optimal placing is very complicated.

From the above it is therefore a well established knowledge that supplementary paths do improve the network diameter.

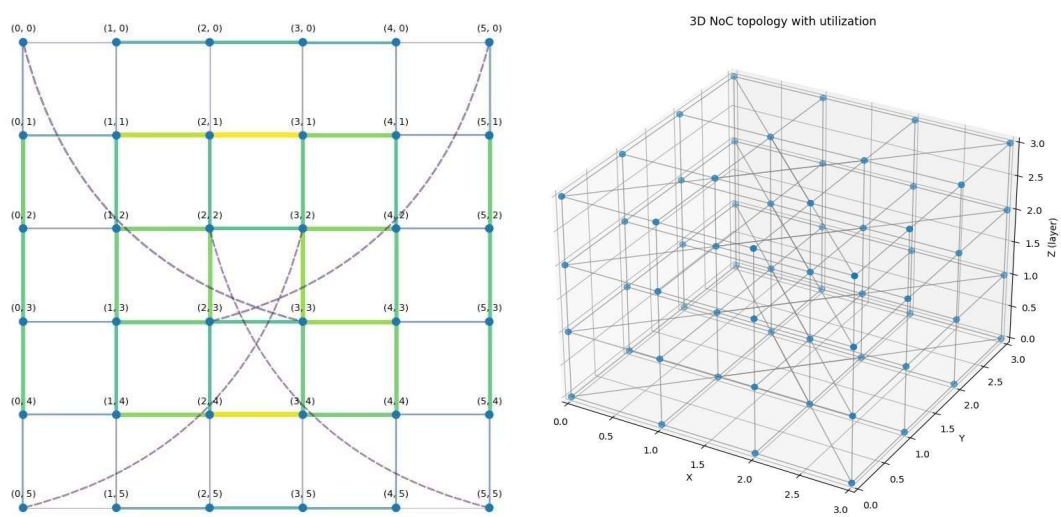
In the next section we introduce 3D Mesh NoC architecture with Supplementary 2D Links along with its progressive shortest - path routing algorithm.

3. 3D Mesh NoC Architecture with Supplementary 2D Links

Here we describe the 3D topology which is a direct and intuitive extension of the 2D mesh having supplementary connections from four corner nodes to the four nodes in the middle region of the 2D mesh topology. These nodes having supplementary nodes are called special nodes. Also, for any N **(**$N \geq 4$**)** the number of special nodes in the 2D plane remain constant i.e., 8. However, in the 3D topology, which is made from stacking these 2D layer vertically, the number of spatial nodes is given by

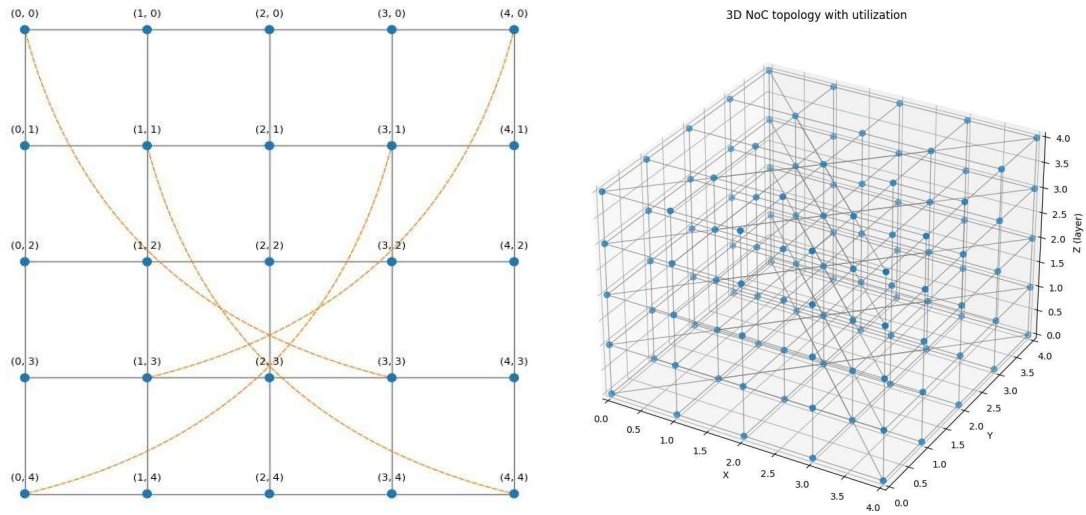
$$Number\ of\ special\ Route = 8 \times number\ of\ vertical\ layers \tag{1}$$

The special routers are placed in four corner nodes, i.e., (0,0), (0,N), (N,N), and (N,0). whereas the middle special routers are placed at $([\frac{N}{2}],[\frac{N}{2}])$, $([\frac{N}{2}]-1,[\frac{N}{2}])$, $([\frac{N}{2}],[\frac{N}{2}]-1)$, and $([\frac{N}{2}]-1,[\frac{N}{2}]-1)$. Figures 2 and Figure 3 show the connection of the 2D and 3D networks described above for $N = 4$ and $N = 5$.



(a) 2D Mesh with supplementary links (b) 3D Mesh an extension of the 2D Mesh in (a)

Figure 2: Transformation of 2D to 3D topology



(a) 2D Mesh with supplementary links N = 5 (b) Extended for N=5

Figure 3: Topology for N = 5

3.1. Routing Algorithm

Below, we describe the progressive shortest-path routing algorithm designed for this topology. Consider a 3D mesh network with dimensions $N_z \times N_y \times N_z$. Each router is identified by integer coordinates

$$c = (z, y, z),\ z \in \{0, \ldots, N_z - 1\},\ y \in \{0, \ldots, N_y - 1\},\ z \in \{0, \ldots, N_z - 1\}.$$

Routers are connected to their six nearest neighbours along the z, y, and z directions (when present). In addition, certain routers on each layer are connected by long-range special links (SP links). On layer z, these SP links form a set of undirected pairs

$$p_z = \{(a_i, b_i)\}_{i=1}^{k_z}, \qquad a_i = (x_{a_i}, y_{a_i}, z),\ b_i = (x_{b_i}, y_{b_i}, z).$$

SP links act as express connections across the *X*-*Y* plane of that layer.

Routing is decomposed into two stages:

1. **Z-alignment:** every packet first moves only along the z-dimension from its source layer z_s to the destination layer z_d.
2. **In-layer SP-assisted routing:** once $z = z_d$, the problem reduces to finding an efficient path in the 2D grid of routers on that layer, using both normal mesh links and SP links.

The vertical cost $|z_s - z_d|$ is identical for all candidate in-layer routes between the same source and destination, hence the optimization can focus on the *x*–*y* plane of the destination layer. For compactness, define the 2D projection of a 3D coordinate onto layer z_d as:

$$\pi_{z_d}(z, y, z) = (z, y, z_d),$$

and the Manhattan distance in the *x*–*y* plane by

$$dist_{2D}\big((x_1, y_1), (x_2, y_2)\big) = |x_1 - x_2| + |y_1 - y_2|$$

For each message with source $s = (x_s, y_s, z_s)$ and destination $d = (x_d, y_d, z_d)$, the router at injection time constructs a per-message plan

$$\pi^{3D} = (\text{mode, entry_sp, exit_sp, sp_used})$$

where,

- mode ∈ {XY, SP ASSIST} indicates whether the in-layer path uses only normal mesh links or a single SP hop in addition.
- entry_sp ∈ $\{0, \ldots, N_z - 1\} \times \{0, \ldots, N_y - 1\} \times \{z_d\}$ is the selected SP entry router on the destination layer (undefined if mode = XY).
- exit_SP is the corresponding SP exit router on the destination layer.
- sp_used ∈ {false, true} is a single bit that records whether the SP hop has already been taken in the in-layer phase.

3D SP-Assisted Route Planning (Per Message)

Let $s_{2D} = (z_s, y_s)$ and $d_{2D} = (z_d, y_d)$ denote the 2D projections of the source and destination onto the destination layer $z = z_d$. The pure in-layer distance using only normal links is

$$D_{XY} = dist_{2D}(s_{2D}, d_{2D}).$$

For each SP link pair $(a, b) \in P_{z_d}$, with $a = (x_a, y_a, z_d)$ and $b = (x_b, y_b, z_d)$, we consider two candidate SP-assisted paths:

- Entry at a and exit at b:

$$L_1(a, b) = dist_{2D}\big(s_{2D}, (x_a, y_a)\big) + 1 + dist_{2D}\big((x_b, y_b), d_{2D}\big).$$

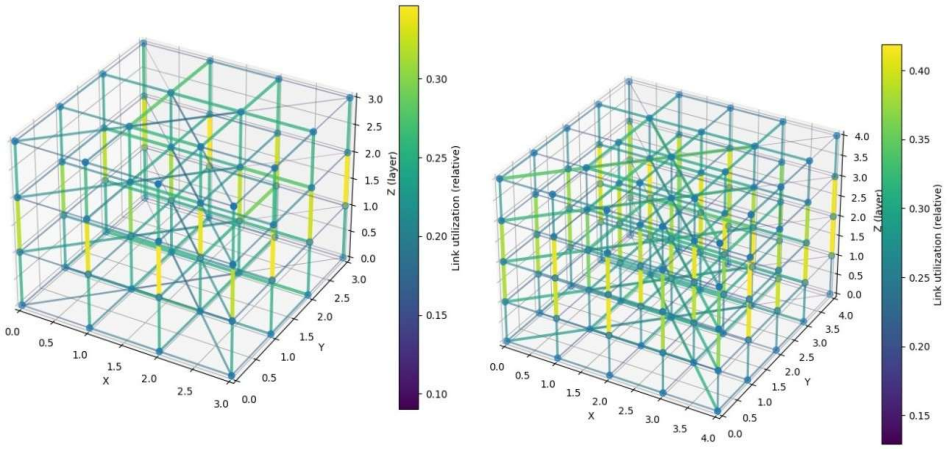
- Entry at b and exit at a:

$$L_2(a, b) = dist_{2D}\big(s_{2D}, (x_b, y_b)\big) + 1 + dist_{2D}\big((x_a, y_a), d_{2D}\big).$$

The +1 term accounts for the SP hop itself. The planner compares the best SP-assisted candidate against D_{XY} and chooses SP-assisted routing only when it strictly reduces the in-layer hop count.

4. Experimental Results

We implemented the proposed topology and evaluated its performance by computing the maximum hop count (worst case) and the corresponding worst-case latency using the proposed algorithm. The primary objective of this work is to see the rout ability and to record the link utilization, hop count and maximum latency. The figure below shows the link utilization of the proposed 3D topology for $N = 4$ and $N = 5$.



(a) N=4 (b) N=5

Figure 4: Link Utilization Graph

The table 1 below compares the maximum hop, maximum latency cycles and average latency cycles of 3D mesh and the proposed 3D topology with the progressive shortest path routing algorithm.

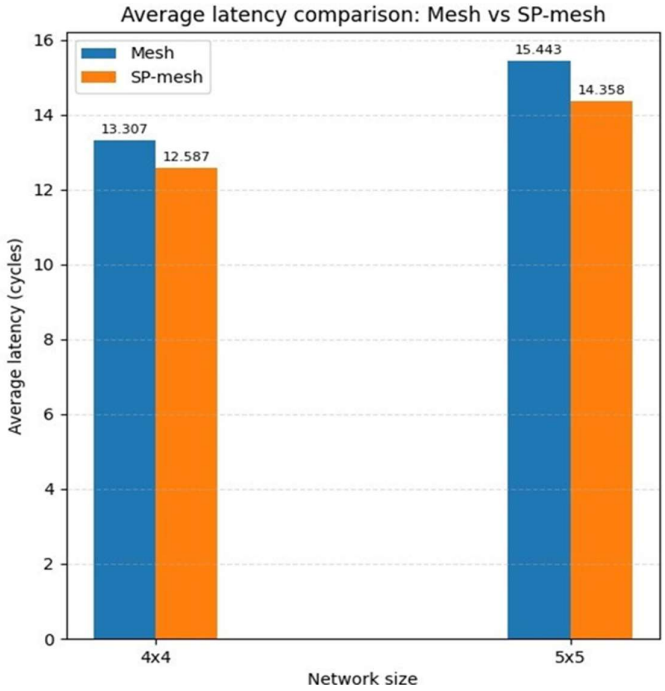


Figure 5: Comparative graph

Table 1: Comparison between maximum hop, maximum latency cycles, and average latency cycles of 3D mesh and the proposed 3D topology

Topology	Network Size	Maximum Hop Count	Maximum Latency	Average Latency
Proposed Mesh Based Topology	4 × 4	3	69	10.662
	5 × 5	4	52	11.869
Mesh Topology	4 × 4	6	72	11.710
	5 × 5	10	92	13.070

It is evident from the above table that, the proposed algorithm improves the maximum latency considerably in the 3D Mesh NoC Architecture with Supplementary 2D links over the 3D mesh topology.

5. Conclusions

This work presents a 3D mesh-based Network-on-Chip (NoC) topology augmented with four additional links that connect eight special nodes, of which four are located at the corners of the mesh and the remaining four are positioned at the center within the same 2D plane. The proposed topology is simulated using a Dijkstra-based Progressive Shortest-Path Routing algorithm. Simulation results show average latencies of 12.587 and 14.358 latency cycles for network sizes $N = 4$ and $N = 5$, respectively, which are slightly lower than those observed in a conventional 3D mesh topology. Furthermore, for $N = 5$, the proposed topology achieves an approximate 63% reduction in maximum latency compared to the standard 3D mesh of same size. However, link utilization analysis indicates that two of the additional links are heavily utilized, while the remaining two experience moderate utilization. Future work will focus on evaluating the proposed network under different traffic patterns to further examine its performance and load distribution.

References

[1] V. Dumitriu and G. Khan. "Throughput-Oriented NoC Topology Generation and Analysis for High Performance SoCs". In: IEEE Transactions on Very Large Scale Integration (VLSI) Systems 17 (2009), pp. 1433–1446. doi: 10.1109/TVLSI.2008.2004592.

[2] Dongkook Park et al. "MIRA: A Multi-layered On-Chip Interconnect Router Architecture". In: ACM SIGARCH Computer Architecture News 36 (June 2008), pp. 251–261. doi: 10.1145/ 1394608.1382143.

[3] Christian Neeb and Norbert Wehn. "Designing efficient irregular networks for heterogeneous systems-on-chip". In: Journal of Systems Architecture 54.3 (2008). System and Network on Chip, pp. 384–396. ISSN: 1383-7621. doi: https://doi.org/10.1016/j.sysarc.2007.07.006. URL: https://www.sciencedirect.com/science/article/pii/S1383762107001191.

[4] Xinyu Li and Omar Hammami. "NOCDEX: Network on Chip Design Space Exploration Through Direct Execution and Options Selection Through Principal Component Analysis". In: Industrial Embedded Systems - IES'2006. Nov. 2006, pp. 1–4. doi: 10.1109/IES.2006.357469.

[5] Inc. Arteris. Network-on-Chip Mesh Tile Topology Speeds Time-to-Market for System-on-Chip Designs Used for Generative AI. Accessed: 2025-10-31. Feb. 2024. URL: https://www.designnews. com/semiconductors-chips/revising-chip-design-to-gear-up-for-ai.

[6] William J Dally and Brian Towles. Principles and practices of interconnection networks. Morgan Kaufmann Publishers Inc., 2003.

[7] M. Mirza-Aghatabar et al. "An Empirical Investigation of Mesh and Torus NoC Topologies Under Different Routing Algorithms and Traffic Models". In: 10th Euromicro Conference on Digital System Design Architectures, Methods and Tools (DSD 2007). 2007, pp. 19–26. doi: 10.1109/DSD.2007.4341445.

[8] Usman Ali Gulzari et al. "Comparative analysis of 2D mesh topologies with additional com- munication links for on-chip networks". In: Computer Networks 241 (2024), pp. 110193. iSSN: 1389-1286. doi: https://doi.org/10.1016/j.comnet.2024.110193. URL: https://www. sciencedirect.com/science/article/pii/S1389128624000252.

[9] Y. Ghidini et al. "Topological impact on latency and throughput: 2D versus 3D NoC comparison". In: 2012 25th Symposium on Integrated Circuits and Systems Design (SBCCI). 2012, pp. 1–6. doi: 10.1109/SBCCI.2012.6344439.

[10] Miglani, A., Bhatia, T., Sharma, G., & Shrivastava, G. (2017). An Energy Efficient and Trust Aware Framework for Secure Routing in LEACH for Wireless Sensor Networks. Scalable Computing: Practice and Experience, 18(3), 207-218.

[11] Yu Gan, Hong Guo, and Ziheng Zhou. "3D NoC Low-Power Mapping Optimization Based on Improved Genetic Algorithm". In: Micromachines 12 (Oct. 2021), pp. 1217. doi: 10 . 3390 / mi12101217.

[12] Yigitcan Nalci et al. "ILP formulation and heuristic method for energy-aware application mapping on 3D-NoCs". In: The Journal of Supercomputing 77 (Mar. 2021). doi: 10.1007/s11227- 020-03365-0.

[13] M. O. Agyeman, A. Ahmadinia, and N. Bagherzadeh. "Performance and Energy Aware Inhomogeneous 3D Networks-on-Chip Architecture Generation". In: IEEE Transactions on Parallel and Distributed Systems 27 (2016), pp. 1756–1769. doi: 10.1109/TPDS.2015.2457444.

[14] Kun-Chih Chen, Chih-Hao Chao, and A. Wu. "Thermal-Aware 3D Network-On-Chip (3D NoC) Designs: Routing Algorithms and Thermal Managements". In: IEEE Circuits and Systems Magazine 15 (2015), pp. 45–69. doi: 10.1109/MCAS.2015.2484139.

[15] Ebadollah Taheri, A. Patooghy, and K. Mohammadi. "Cool elevator: A thermal-aware routing algorithm for partially connected 3D NoCs". In: 2016 6th International

Conference on Computer and Knowledge Engineering (ICCKE), 2016, pp. 111–116. doi: 10.1109/ICCKE.2016.7802125.

[16] Vimal, S., Khari, M., Crespo, R. G., Kalaivani, L., Dey, N., & Kaliappan, M. (2020). Energy enhancement using Multiobjective Ant colony optimization with Double Q learning algorithm for IoT based cognitive radio networks. Computer Communications, 154, 481-490.

[17] Sapna Tyagi et al. "Exploring 3D Network-on-Chip Architectures and Challenges". In: Inter- national Conference on Computer, Communications and Electronics (Comptelix). Sept. 2017, pp. 97–101. doi: 10.1109/COMAPP.2017.8079768.

[18] Gusain, Nutan, and Himanshu Sharma. "Communication-efficient federated learning in industrial IoT—a framework for real-time threat detection and secure device coordination." *International Journal on Computational Modelling Applications* 2, no. 2 (2025): 18-29.

[19] K. Gaffour et al. "A new congestion-aware routing algorithm in network-on-chip: 2D and 3D comparison". In: International Journal of Computers and Applications 45 (2019), pp. 27–35. doi: 10.1080/1206212X.2019.1679529.

[20] Shivam Tyagi and Shweta Bohare. "Review of 3-D network-on-chip topologies". In: 2011 World Congress on Information and Communication Technologies. 2011, pp. 783–788. doi: 10.1109/ WICT.2011.6141346.

[21] A. Q. Ansari, Mohammad Rashid Ansari, and Mohammad Ayoub Khan. "Modified quadrant-based routing algorithm for 3D Torus Network-on-Chip architecture". In: Perspectives on Science 8 (2016), pp. 718–721. doi: 10.1016/J.PISC.2016.06.069.

[22] Sourav Das et al. "Small-World Network Enabled Energy Efficient and Robust 3D NoC Architectures". In: Proceedings of the 25th Edition on Great Lakes Symposium on VLSI. GLSVLSI '15. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, 2015, pp. 133–138. ISBN: 9781450334747. doi: 10.1145/2742060.2742085. URL: https://doi.org/10.1145/2742060.