Two Decades of Innovation in 3D-NoC and Neuromorphic Computing: From Reliable Architectures to Intelligent Systems

Abderazek Ben Abdallah
School of Computer Science and Engineering
The University of Aizu
Aizuwakamatsu, Fukushima, Japan
benab@u-aizu.ac.jp
January 13, 2025

***Abstract*—As semiconductor scaling approaches physical limits, the paradigm of computing has shifted towards many-core architectures. However, traditional 2D planar interconnects suffer from high latency and power consumption as chip sizes grow. Three-Dimensional Network-on-Chip (3D-NoC) addresses these challenges by leveraging Through-Silicon Vias (TSVs) to enable vertical communication, significantly reducing wire length. This survey presents a comprehensive review of over a decade of research conducted by our group on robust 3D-NoC design. We provide an in-depth analysis of our contributions, including: (1) High-performance look-ahead routing algorithms (LA-XYZ) that minimize pipeline stages; (2) Reliability mechanisms (LAFT, 3D-FETO, SER-3DR) that combat both hard manufacturing defects and soft errors; (3) Thermal-aware management strategies (HotCluster) that mitigate hotspot-induced aging; and (4) The application of these interconnects to Neuromorphic Computing (NASH, BTSAM), enabling scalable Spiking Neural Networks (SNNs). Furthermore, we detail the intellectual property and patents resulting from these innovations. This paper serves as a holistic reference for designing the next generation of reliable, efficient, and intelligent on-chip communication infrastructures.**

***Index Terms*—3D Network-on-Chip, TSV, Look-Ahead Routing, Fault Tolerance, Soft Error Resilience, Thermal-Aware Mapping, Neuromorphic Computing, Spiking Neural Networks, Patents.**

I. Introduction

The integration of hundreds or thousands of cores on a single chip has necessitated the move from bus-based communication to Network-on-Chip (NoC). While 2D-NoCs provided a scalable solution for a time, they are increasingly bottlenecked by long horizontal links and high power dissipation. Three-Dimensional Integrated Circuits (3D-ICs) have emerged as the solution, allowing layers of logic to be stacked vertically.

Despite the benefits, 3D-NoCs introduce severe challenges. First, vertical links (TSVs) are prone to manufacturing defects and stress-induced failures. Second, the increased power density in 3D stacks leads to thermal hotspots, which accelerate aging and reliability degradation. Third, as transistor sizes shrink, the susceptibility to soft errors (Single Event Upsets) increases.

Our research group has systematically addressed these challenges through a series of novel architectures and algorithms. This survey details our journey from optimizing basic routing logic to designing comprehensive neuromorphic systems, culminating in a portfolio of patented technologies.

II. Fundamentals of 3D-NoC Architecture

Before detailing our specific contributions, we define the baseline architecture used across our works. We utilize a grid-based 3D Mesh topology, where each router is connected to its neighbors in six directions (North, South, East, West, Up, Down) and to the local Processing Element (PE).

The vertical communication is facilitated by Through-Silicon Vias (TSVs). While TSVs reduce global wire length, they represent a critical point of failure. Our designs consistently assume a realistic fault model where TSVs can be fully or partially defective, necessitating robust recovery mechanisms.

III. High-Performance Routing Algorithms

Routing efficiency determines the latency and throughput of the entire network. Our work has focused on reducing the pipeline depth of the router.

A. The Look-Ahead Principle (LA-XYZ)

Traditional routing algorithms calculate the output port for a packet only after it arrives at the current router. This sequential process adds significant latency at every hop. To overcome this, we developed the **Look-Ahead-XYZ (LA-XYZ)** routing algorithm [1].

The core innovation of LA-XYZ is the parallelization of the routing calculation. The routing decision for the *next* router is computed at the *current* router.

1) **Mechanism:** When a packet traverses the link between Router A and Router B, the header already contains the output port information for Router B. This allows Router B to skip the Route Computation (RC) stage entirely.
2) **Pipeline Reduction:** Standard routers use a 4-stage pipeline (Buffer Write, Route Compute, Switch Allocation, Switch Traversal). LA-XYZ effectively reduces this

to a 2-stage or 3-stage pipeline for the majority of flits, resulting in a dramatic reduction in zero-load latency.

3) **Throughput:** Simulation results demonstrate that LA-XYZ achieves significantly higher saturation throughput compared to standard Dimension Order Routing (DOR) in 3D meshes.

B. Extensions for Deadlock Freedom

Ensuring deadlock freedom in 3D-NoCs is complex due to the additional vertical dimension. Our algorithms utilize virtual channels (VCs) and careful restriction of turns (e.g., prohibiting certain adaptive turns) to break cycles in the channel dependency graph, ensuring that packets are never permanently blocked.

IV. Reliability and Fault Tolerance

Reliability is the cornerstone of our research. We categorize our contributions into Hard Fault Tolerance (permanent physical defects) and Soft Error Resilience (transient upsets).

A. Hard Fault Tolerance: LAFT and 3D-FETO

TSVs are fragile. A single broken TSV can sever communication between layers. To address this, we proposed the **Look-Ahead Fault-Tolerant (LAFT)** algorithm [2] and the **3D-FETO** architecture [5].

1) Dynamic Bypass Mechanism: LAFT extends the look-ahead concept to include fault detection. If the preferred output port (calculated one hop ahead) is known to be faulty:

- The router dynamically recalculates an alternative path.
- The algorithm prioritizes non-faulty links that still advance the packet toward the destination.
- If all minimal paths are blocked, the algorithm switches to a non-minimal adaptive mode to circumnavigate the faulty region.

2) 3D-FETO Architecture: 3D-FETO integrates this logic into hardware. It employs a "Fault Table" within each router that stores the status of neighboring links. This table is updated via a lightweight link-testing protocol. The architecture ensures graceful degradation: as the number of faults increases, latency increases marginally, but connectivity is maintained.

B. Soft Error Resilience: SER-3DR

As technology scales below 22nm, radiation-induced soft errors can flip bits in the router's buffers or control logic. We developed the **Soft-Error Resilient 3D Router (SER-3DR)** [6].

1) Datapath Protection: We utilize Error Correcting Codes (ECC), specifically Single-Error Correction Double-Error Detection (SECDED) Hamming codes, to protect data flits in the buffers.

2) Control Logic Protection: Protecting the control logic is harder because it changes every cycle. SER-3DR employs *Temporal Redundancy*: critical control signals are computed twice in slightly shifted clock phases. A voting logic compares the results. If a discrepancy (SEU) is detected, the pipeline is stalled for one cycle to recompute, preventing the error from propagating to the crossbar.

C. TSV Defect Recovery

In [7], we introduced a non-blocking testing method for TSVs. Traditional testing requires pausing the system. Our method utilizes spare time slots in the transmission to inject test vectors, allowing for runtime health monitoring of TSVs without interrupting active applications.

V. Thermal-Aware Design and Management

Heat is a major killer of 3D-ICs. The inner layers of a 3D stack cannot easily dissipate heat, leading to thermal gradients.

A. HotCluster: Thermal-Aware Defect Recovery

We proposed **HotCluster** [9], a novel approach to defect recovery.

- **Problem:** Using spare TSVs to repair faults is standard. However, activating a spare TSV in a hot region may cause it to fail quickly due to electromigration.
- **Solution:** HotCluster aggregates thermal data from on-chip sensors. When a TSV fails, the recovery algorithm selects a redundant path or spare TSV located in a *cooler* region of the cluster, even if it is slightly less optimal in terms of latency. This "thermal-aware redundancy" significantly extends the Mean Time To Failure (MTTF) of the entire chip.

VI. Neuromorphic 3D-NoC Systems

The convergence of artificial intelligence and hardware design has led to the rise of Neuromorphic Computing. Unlike traditional von Neumann architectures, neuromorphic systems mimic the biological brain's structure, consisting of massive arrays of neurons and synapses. Our research group has pioneered the integration of 3D-NoC interconnects to solve the scalability and communication bottlenecks inherent in these massive Spiking Neural Networks (SNNs).

A. NASH: Scalable Digital Neuromorphic System

In [11], we introduced **NASH (Neuromorphic Architecture with 3D-NoC)**, a comprehensive platform designed to support large-scale SNNs with on-chip learning capabilities.

1) Architecture: NASH utilizes a tile-based architecture where each tile contains a Crossbar Neuro-Core (CNC) and a router. The CNC implements digital Leaky Integrate-and-Fire (LIF) neurons. The 3D-NoC serves as the "axon" network, transporting spike packets between layers.

- **Spike Packetization:** Unlike point-to-point analog wires, NASH packetizes spikes (Address Event Representation) to travel over the digital network, allowing for virtual connectivity that exceeds physical wiring limits.
- **On-Chip Learning:** NASH implements Spike-Timing-Dependent Plasticity (STDP) in hardware. This allows the system to adapt synaptic weights in real-time based on the correlation of pre-synaptic and post-synaptic spike timing.

2) Key Contribution: The primary contribution of NASH is its scalability. By leveraging the vertical TSV connections, we demonstrated that the average hop count for spike traversal is significantly reduced compared to 2D implementations, leading to lower energy consumption per spike and faster inference times on datasets like MNIST.

B. Fault-Tolerant SNN Mapping

As neuromorphic chips scale to wafer-level sizes, manufacturing defects become unavoidable. In [12], we addressed the challenge of mapping SNNs onto 3D-NoC substrates containing faulty cores.

1) Problem Formulation: A standard mapping algorithm might place a critical neuron on a faulty tile, rendering the entire network non-functional. The goal is to find a permutation of neuron-to-tile assignments that avoids all defects while minimizing the total wirelength (communication cost).

2) Proposed Methodology: We developed a **Fault-Tolerant Mapping Algorithm** utilizing Particle Swarm Optimization (PSO).

1) **Defect Map Awareness:** The algorithm takes a "Defect Map" as input, identifying which tiles in the 3D stack are non-functional.
2) **Optimization:** The PSO algorithm iteratively explores the search space, penalizing solutions that use faulty tiles or result in high latency.
3) **Results:** Our method successfully maps complex SNN topologies onto chips with up to 10% defect rates without significant degradation in classification accuracy, ensuring the brain-like property of resilience.

C. HeterGenMap: Heterogeneous Mapping

Modern neuromorphic systems are increasingly heterogeneous, incorporating different types of neuron cores (e.g., fast-spiking vs. power-efficient) or process variations. In [13], we proposed **HeterGenMap**, an evolutionary framework for such systems.

1) Genetic Algorithm Approach: HeterGenMap employs a Genetic Algorithm (GA) to optimize the placement of neurons.

- **Chromosome Encoding:** Each chromosome represents a potential mapping of the SNN.
- **Fitness Function:** The fitness function is multi-objective, accounting for:

$$F = w_1 \cdot E_{comm} + w_2 \cdot T_{max} + w_3 \cdot L_{avg} \quad (1)$$

where E_{comm} is communication energy, T_{max} is maximum temperature, and L_{avg} is average latency.

This work is crucial for utilizing "chiplet" based neuromorphic designs where different layers may have different performance characteristics.

D. BTSAM: Thermal-State-Aware Mapping

Spiking activity in SNNs is often bursty, leading to rapid temperature fluctuations. In [14], we introduced **BTSAM (Balanced Thermal-State-Aware Mapping)**.

1) Dynamic Remapping: Unlike static mapping methods, BTSAM considers the time-varying thermal profile of the chip.

- **Hotspot Prediction:** The system monitors activity counters to predict impending hotspots.
- **Balancing:** It strategically places highly active neurons (hubs) in cooler regions of the 3D stack, often utilizing the outer layers which have better heat dissipation properties compared to the inner layers.

BTSAM ensures that the neuromorphic chip operates within a safe thermal envelope, preventing thermal throttling that would disrupt the precise timing required for SNN operation.

VII. Intellectual Property and Patents

Our academic research has led to the development of several protected intellectual properties, bridging the gap between theoretical architecture and practical implementation.

A. 3D-NoC and TSV Reliability Patents

- **TSV Grouping for 3D SoC [16]:** This patent (JP7488989B2, 2024) details a method for organizing Through-Silicon Vias (TSVs) into logical groups to optimize vertical bandwidth and manufacturing yield between stacked layers.
- **TSV Error Tolerant Router [17]:** Granted in 2023, this invention describes a router architecture capable of detecting specific TSV failures and dynamically rerouting packets through backup vertical links without halting the system.
- **Defect Tolerance Router [18]:** This patent (2021) covers a generalized router micro-architecture that includes built-in self-test (BIST) logic to identify and isolate internal logic defects.
- **Optical NoC System [19]:** Addressing the speed of light limits, this 2020 patent proposes a hybrid system using non-blocking photo-switches to enable ultra-low latency optical communication paths alongside electrical NoCs.
- **Error Resilience Router [20]:** This earlier patent (2018) lays the foundation for our soft-error resilience work, detailing control methods for correcting transient errors in router pipelines.

B. Neuromorphic and AI Patents

- **AI Processor [21]:** This recently granted patent (2025) describes a dedicated hardware accelerator for artificial intelligence tasks, likely integrating the NASH architecture principles for efficient edge-AI processing.
- **Spiking Neural Network by 3D NoC [22]:** Granted in 2023, this core patent protects the specific method of mapping and routing SNN spikes over a 3D-Mesh Network-on-Chip, which is fundamental to our scalable neuromorphic designs.
- **Neural Network Processor [23]:** A provisional patent filed in 2024 focusing on the next-generation architecture for neural network processing, emphasizing heterogeneity and energy efficiency.

VIII. Conclusion

This survey has consolidated over a decade of research from the Adaptive Systems Laboratory. We have demonstrated that 3D-NoC is not just a theoretical concept but a viable, high-performance backbone for future systems.

Our work on LA-XYZ established the baseline for low-latency routing. Our reliability research (LAFT, 3D-FETO, SER-3DR) proved that 3D systems can be made robust against the inherent fragility of deep sub-micron fabrication. Finally, our recent foray into Neuromorphic Computing (NASH, BT-SAM) highlights the versatility of 3D-NoC in enabling the next generation of AI hardware.

Acknowledgment

The author acknowledges the hard work of laboratory members including Akram Ben Ahmed, Khanh N. Dang, Mohamed Maatar, Mark Ikechukwu Ogbodo, and others whose theses and papers form the basis of this survey.

References

[1] A. B. Ahmed and A. B. Abdallah, "LA-XYZ: Low latency, high throughput look-ahead routing algorithm for 3D network-on-chip (3D-NoC) architecture," in *IEEE 6th International Symposium on Embedded Multicore SoCs (MCSoC)*, 2012, pp. 167–174.

[2] A. B. Ahmed and A. B. Abdallah, "Architecture and design of high-throughput, low-latency, and fault-tolerant routing algorithm for 3D-network-on-chip (3D-NoC)," *The Journal of Supercomputing*, vol. 66, no. 3, pp. 1507–1532, 2013.

[3] A. B. Ahmed and A. B. Abdallah, "Graceful deadlock-free fault-tolerant routing algorithm for 3D Network-on-Chip architectures," *Journal of Parallel and Distributed Computing*, vol. 74, no. 4, pp. 2229–2240, 2014.

[4] A. B. Ahmed, A. B. Abdallah, and K. Kuroda, "Architecture and Design of Efficient 3D Network-on-Chip (3D NoC) for Custom Multicore SoC," in *International Conference on Broadband, Wireless Computing, Communication and Applications*, 2010.

[5] A. B. Ahmed and A. B. Abdallah, "Adaptive fault-tolerant architecture and routing algorithm for reliable many-core 3D-NoC systems," *Journal of Parallel and Distributed Computing*, vol. 93, pp. 30–43, 2016.

[6] K. N. Dang, M. Meyer, Y. Okuyama, and A. B. Abdallah, "A Soft-Error Resilient 3D Network-on-Chip Router for Highly-reliable Multi-core Systems," in *IEEE 7th International Conference on Awareness Science and Technology (iCAST)*, 2015.

[7] K. N. Dang, M. C. Meyer, A. B. Ahmed, A. B. Abdallah, and X.-T. Tran, "A Non-Blocking Non-Degrading Multiple Defects Link Testing Method for 3D-Networks-on-Chip," *IEEE Access*, vol. 8, pp. 59571–59589, 2020.

[8] K. N. Dang, A. B. Ahmed, A. B. Abdallah, and X.-T. Tran, "A Thermal-Aware On-Line Fault Tolerance Method for TSV Lifetime Reliability in 3D-NoC Systems," *IEEE Access*, vol. 8, pp. 166642–166657, 2020.

[9] K. N. Dang, A. B. Ahmed, A. B. Abdallah, and X.-T. Tran, "HotCluster: A Thermal-Aware Defect Recovery Method for Through-Silicon-Vias Toward Reliable 3-D ICs Systems," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 4, pp. 799–812, 2021.

[10] M. Maatar, K. N. Dang, and A. B. Abdallah, "Thermal-Aware Task-Mapping Method for 3D-NoC-Based Neuromorphic Systems," in *IEEE International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)*, 2023.

[11] M. I. Ogbodo, K. N. Dang, and A. B. Abdallah, "On the Design of a Fault-Tolerant Scalable Three Dimensional NoC-Based Digital Neuromorphic System With On-Chip Learning," *IEEE Access*, vol. 9, pp. 64331–64344, 2021.

[12] W. Y. Yerima, M. I. Ogbodo, K. N. Dang, and A. B. Abdallah, "Fault-Tolerant Spiking Neural Network Mapping Algorithm and Architecture to 3D-NoC-Based Neuromorphic Systems," *IEEE Access*, vol. 11, pp. 52429–52443, 2023.

[13] K. N. Dang, N. A. V. Doan, N. Nguyen, and A. B. Abdallah, "HeterGenMap: An Evolutionary Mapping Framework for Heterogeneous NoC-Based Neuromorphic Systems," *IEEE Access*, vol. 11, pp. 144095–144112, 2023.

[14] M. Maatar, Z. Wang, K. N. Dang, and A. B. Abdallah, "BTSAM: Balanced Thermal-State-Aware Mapping Algorithms and Architecture for 3D-NoC-Based Neuromorphic Systems," *IEEE Access*, vol. 12, pp. 126679–126692, 2024.

[15] M. Maatar, Z. Wang, K. N. Dang, and A. B. Abdallah, "3D-NoC-Based Neuromorphic Systems," *Book Chapter/Article*, Jan 2024.

[16] A. B. Abdallah and K. N. Dang, "A three-dimensional system on chip in which a TSV group connects between layers," Japan Patent JP7488989B2, 2024.

[17] A. B. Abdallah, K. N. Dang, and M. Hisada, "TSV Error Tolerant Router Device for 3D Network On Chip," Japan Patent JP7239099B2, 2023.

[18] A. B. Abdallah, "Defect tolerance router for network on-chip," Japan Patent JP6846027B2, 2021.

[19] A. B. Abdallah, "Optical network-on-chip system using non-block photo-switches," Japan Patent JP6747660B2, 2020.

[20] A. B. Abdallah, "Error resilience router and control method," Japan Patent JP6284177B2, 2018.

[21] A. B. Abdallah, H. H. Kun, D. N. Khanh, and S. Janning, "AI Processor," Japan Patent JP7699791B2, June 20, 2025.

[22] A. B. Abdallah, T. H. Vu, and M. Hisada, "Spiking Neural Network by 3D NoC," Japan Patent JP7277682B2, May 11, 2023.

[23] K. N. Dang, A. B. Abdallah, and N. N. Doanh, "Neural Network Processor," Japan Provisional Patent 2024-047372, 2024.