Publications Template

#	Research Title	Field	Abstract	Year of Publication Publishing	Publishing Link "URL"
1	Semi-adaptive distributed approach for triplet-based architecture inter-core communication Network-on-Chip	**Computer Architecture**	Network-on-Chip (NoC) architectures offer significant performance improvements over traditional bus-based systems. However, as NoC designs become more complex, congestion within channels and buffers can degrade performance. Efficient routing algorithms are essential to mitigate congestion and optimize overall NoC performance. This study examines Triplet-Based Architecture (TriBA) architecture and its baseline deterministic shortest-path routing algorithm. The deterministic nature of this algorithm, combined with TriBA's inherent characteristics, may exacerbate congestion and lead to routing deadlocks, particularly in high-traffic nodes.	**2025**	https://doi.org/10.1016/j.vlsi.2025.102489

			This work introduces a novel two-stage semi-adaptive routing algorithm to address congestion within TriBA-NoC. The proposed approach leverages congestion levels within downstream buffers of TriBA sub-level apex vertices as an additional routing metric solely at the source node. The primary goal of the proposed strategy is to alleviate congestion, specifically at hot nodes, which in turn leads to reduced communication latency, shorter queuing times in downstream buffers, lower power consumption, and an enhancement in overall network throughput. Comprehensive simulations utilizing gem5-HeteroGarnet have substantiated the effectiveness of the proposed semi-adaptive routing for TriBA-NoC. The approach yields a significant decrease in latency (up to 31.71%), and a decline in buffer queuing time, with reductions reaching 43.14%. Additionally, it enhances throughput by 6.28% and lowers downstream buffer power consumption by an average of 12.15% across various traffic patterns when compared to baseline algorithms.		

			These improvements are evident in practical workloads, as demonstrated by simulations using the PARSEC benchmark suite, which reveal latency reductions between 0.91% and 45.86%. Nonetheless, these performance enhancements are accompanied by a modest power overhead, estimated at approximately 3.94%.		
2	Reactive Deadlock Avoidance Based on Focus Routing Graph Classification for Triplet-Based Architecture Network-on-Chip	**Computer Engineering**	The implementation of network-on-chip (NoC) architectures presents considerable advantages in performance relative to traditional bus-based systems. However, the sophisticated nature of NoC designs demands careful oversight of shared resources to mitigate potential performance issues. In this context, well-structured routing algorithms are essential, as they facilitate improved traffic management and minimize congestion. Furthermore, mechanisms for deadlock prevention and avoidance are integral to routing algorithms, ensuring continuous packet transmission and ultimately enhancing network performance. This article introduces a novel reactive deadlock avoidance method for Triplet-Based	**2026**	https://ieeexplore.ieee.org/document/11032167

			Architecture Intercore NoC (TriBA-cNoC) that classifies routing based on focus routing graph (FRG) size to address routing-level deadlocks caused by the combination of deterministic routing and TriBA-cNoC's inherent network characteristics. Compared to proactive techniques, it improves downstream buffer utilization and reduces power consumption. Furthermore, two shortest-path routing algorithms are introduced: DM4T-M, which incorporates a round-robin selection mechanism to alleviate congestion on critical paths and minimize hot-node formation, and TriBA-cNoC streamlined routing (TSR), a novel two-stage distributed routing algorithm that addresses the computational overhead associated with output port selection in previous algorithms. Simulation results obtained using gem5 show that the proposed approach, which integrates routing-level reactive deadlock avoidance with the proposed routing algorithms, yields improvements in latency, throughput, buffer utilization, and power consumption. TSR and DM4T-M achieve latency		

			reductions of up to 34.89% and 28.2%, respectively. Throughput increases of up to 16.94% and 7.81% are observed for TSR and DM4T-M, respectively. Moreover, the proposed approaches enhance buffer utilization by up to 13.9% and 10.44%, while reducing power consumption by up to 9.64%.		
3	A High Scalability Memory NoC with Shared-Inside Hierarchical-Groupings for Triplet-Based Many-Core Architecture	**Computer Engineering**	Innovative processor architecture designs are shifting towards Many-Core Architectures (MCAs) to meet the future demands of high-performance computing as the limits of Moore's Law have almost been reached. Many-core processors utilize shared memory hierarchies to achieve high-speed memory systems, improving memory access efficiency. However, as the number of cores multiplies, the scalability of this system is significantly constrained by the increased proportion of long-distance and Non-Uniform Memory Access (NUMA). Improving the scalability of MCAs is crucial for achieving large/super-scale general-purpose many-core processors. This work proposes a high-scalability memory Network-on-Chip (NoC) for Triplet-Based Many-Core	**2025**	https://doi.org/10.1145/3688610

			Architecture (TriBA), named TriBA-mNoC. TriBA-mNoC maintains a consistent core-to-core spacing as the network scale increases, effectively preventing increased long-distance memory access latency. Moreover, it leverages an inherent advantage of shared-inside hierarchical-groupings, alleviating common NUMA issues in the NoC design. Evaluations of static network characteristics show that TriBA-mNoC outperforms most classical NoCs in network diameter, average distance, and cost. TriBA-mNoC can be integrated with TriBA in the same silicon die with a tile-like floorplan, forming a novel NoC called TriBA-NoC, which can combine the strengths of both networks to maximize the architecture performance. We evaluated the memory access performance and scalability of TriBA-NoC using mathematical evaluation models and actual simulations with real traffic (PARSEC 3.0 and SPLASH-2) at different network scales. The mathematical evaluation results indicate that TriBA-NoC achieves an aggregate		

			speedup of approximately 3x compared with 2D-Mesh for a similar number of cores. Furthermore, TriBA-NoC's single-core speedup efficiency remains stable as the number of cores increases under the same cache hit ratio, whereas 2D-Mesh experiences a rapid decline, highlighting TriBA-NoC's exceptional scalability. Finally, the actual traffic simulation results show that TriBA-NoC achieves an average memory access latency and time reduction of 25.90% to 40.50% and 5.61% to 31.69%, respectively, compared with 2D-Mesh.		
4	Deadlock-Free Strategies Based on Synchronized Hamiltonian Ring in Triplet-Based Many-Core Architecture	**Computer Engineering**	Ensuring deadlock-free data transmission in the network-on-chip (NoC) is a prerequisite for providing reliable communication services for multi-processor system-on-chip (MPSoC), directly determining the availability of NoC and even MPSoC. Existing general-purpose deadlock-free strategies are oriented to arbitrary topologies, making it challenging to utilize the features and advantages of a specific topology. Moreover, these strategies may even	**2025**	https://crad.ict.ac.cn/cn/article/pdf/preview/10.7544/issn1000-1239.202331042.pdf

			increase network latency, power consumption, and hardware complexity. In addition, due to significant differences in the regular network between routing-level and protocol-level deadlocks, existing solutions struggle to simultaneously address both types of deadlock issues, affecting MPSoC reliability. We propose a deadlock-free strategy with synchronous Hamiltonian rings based on the inherent Hamiltonian characteristics of the triplet-based many-core architecture (TriBA). This method uses the topology's symmetric axes and Hamiltonian edge to allocate independent store-and-forward buffers for data transmission, preventing protocol-level deadlocks and improving data transfer speed. Additionally, we design a directional determination method for data transmission within the same buffer using cyclic linked-list technology. This method ensures data independence and synchronous forward transmission,		

			eliminates routing-level deadlocks, and reduces data transfer latency. Based on optimizing redundant calculations in look-ahead routing algorithms, we propose a deadlock-free routing mechanism called Hamiltonian shortest path routing (HamSPR) based on a synchronous Hamiltonian ring. GEM5 simulation results show that, compared with existing solutions in the TriBA, HamSPR reduces average packet latency and power consumption in synthetic traffic patterns by 18.78%−65.40% and 6.94%−34.15%, respectively, while improving throughput by 8.00%−59.17%. In the PARSEC benchmark, HamSPR achieves maximum reductions of 16.51% in application runtime and 42.75% in average packet latency, respectively. Moreover, compared with 2D-Mesh, TriBA demonstrates an application performance improvement of 1%−10% in PARSEC benchmark.		

5	Proactive deadlock prevention based on traffic classification sub-graphs for triplet-based NoC TriBA-cNoC	**Computer Engineering**	Network topology and routing algorithms stand as pivotal decision points that profoundly impact the performance of Network-on-Chip (NoC) systems. As core counts rise, so does the inherent competition for shared resources, spotlighting the critical need for meticulously designed routing algorithms that circumvent deadlocks to ensure optimal network efficiency. This research capitalizes on the Triplet-Base Architecture (TriBA) and its Distributed Minimal Routing Algorithm (DM4T) to overcome the limitations of previous approaches. While DM4T exhibits performance advantages over previous routing algorithms, its deterministic nature and potential for circular dependencies during routing can lead to deadlocks and congestion. Therefore, this work addresses these vulnerabilities while leveraging the performance benefits of TriBA and DM4T. This work introduces a novel approach that merges a proactive deadlock prevention mechanism with Intermediate Adjacent Shortest Path Routing (IASPR). This combination guarantees both	**2024**	https://doi.org/10.1016/j.micpro.2024.105091

			deadlock-free and livelock-free routing, ensuring reliable communication within the network. The key to this integration lies in a flow model-based data transfer categorization technique. This technique prevents the formation of circular dependencies. Additionally, it reduces redundant distance calculations during the routing process. By addressing these challenges, the proposed approach achieves improvements in both routing latency and throughput. To rigorously assess the performance of TriBA network topologies under varying configurations, extensive simulations were undertaken. The investigation encompassed both TriBA networks comprising 9 nodes and those with 27 nodes, employing DM4T, IASPR routing algorithms, and the proactive deadlock prevention method. The gem5 simulator, operating under the Garnet 3.0 network model using a standalone protocol for synthetic traffic patterns, was utilized for simulations at high injection rates, spanning diverse synthetic traffic		

			patterns and PARSEC benchmark suite applications. Simulations rigorously quantified the effectiveness of the proposed approach, revealing reductions in average latency of 40.17% and 34.05% compared to the lookup table and DM4T, respectively. Additionally, there were notable increases in average throughput of 7.48% and 5.66%.		
6	NxtSPR: A deadlock-free shortest path routing dedicated to relaying for Triplet-Based many-core Architecture	**Computer Engineering**	Deadlock-free routing is a significant challenge in Network-on-Chip (NoC) design as it affects the network's latency, power consumption, and load balance, impacting the performance of multi-processor systems-on-chip. However, achieving deadlock-free routing will routinely result in expensive overhead as previous solutions either sacrifice performance or power efficiency to proactively avoid deadlocks or impose high hardware complexity to resolve deadlocks when they occur reactively. Utilizing the various characteristics of NoC to implement deadlock-free routing can be significantly more cost-effective with less impact on performance. This paper proposes a	**2024**	https://doi.org/10.1016/j.parco.2024.103094

			relay routing algorithm (NxtSPR) with a shortest path property and a deadlock prevention mechanism based on a synchronized Hamiltonian ring. The proposal is based on an in-depth study of the characteristics of a Triplet-Based many-core Architecture (TriBA) NoC. We establish various important topology-related theories and perform a formal verification (proof-based) for them. By utilizing the critical subgraph and apex of TriBA, NxtSPR can pre-calculate downstream nodes forwarding ports for packets by using a concise judgment strategy. This significantly reduces the computational overhead required for data transmission while optimizing the pipeline design of routers to decrease packet transmission latency and power consumption compared to other TriBA routing algorithms. We group the data transmissions according to the levels of maximum Hamiltonian edges a packet will traverse during its transmission life cycle. Independent data transmissions between groups can avoid mutual		

			interference and resource competition, eliminating potential deadlocks. Gem5 simulation results show that, under the synthetic traffic patterns, compared to the representative (Table) and up-to-date (SPR4T) routing algorithms, NxtSPR achieves a 20.19%, 14.76%, and 5.54%, 4.66% reduction in average packet latency and per-packet power consumption, respectively. Moreover, it has an average of 18.50% and 4.34% improvement in throughput, as compared to them. PARSEC benchmark results show that NxtSPR reduces application runtime by up to a maximum of 22.30% and 12.82% compared to Table and SPR4T, and running the same applications with TriBA results in a maximum runtime reduction of 10.77% compared to 2D-Mesh.		