www.elektrika.utm.my
ISSN 0128-4428

ELEKTRIKA
Journal of Electrical Engineering

Wireless Network On-Chips History-Based Traffic Prediction for Token Flow Control and Allocation

Ayodeji Ireti Fasiku*, Asrani Lit, Paulson Eberechukwu Numan, Shahrizal Rusli, Muhammad Nadzir Marsono

School of Electrical Engineering, Faculty of Engineering Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia.

*Corresponding author: fasiku@graduate.utm.my, Tel: 0105015754.

Abstract: Wireless network-on-chip (WiNoC) uses a wireless backbone on top of the traditional wired-based NoC which demonstrated high scalability. WiNoC introduces long-range single-hop link connecting distanced core and high bandwidth radio frequency interconnects that reduces multi-hop communication in conventional wired-based NoC. However, to ensure full benefits of WiNoC technology, there is a need for fair and efficient Medium Access Control (MAC) mechanism to enhance communication in the wireless Network-on-Chip. To adapt to the varying traffic demands from the applications running on a multicore environment, MAC mechanisms should dynamically adjust the transmission slots of the wireless interfaces (WIs), to ensure efficient utilization of the wireless medium in a WiNoC. This work presents a prediction model that improves MAC mechanism to predict the traffic demand of the WIs and respond accordingly by adjusting transmission slots of the WIs. This research aims to reduce token waiting time and inefficient decision making for radio hub-to-hub communication and congestion-aware routing in WiNoC to enhance end to end latency. Through system level simulation, we will show that the dynamic MAC using an History-based prediction mechanism can significantly improve the performance of a WiNoC in terms of latency and network throughput compared to the state-of-the-art dynamic MAC mechanisms.

Keywords: Wireless network-on-chip, Traffic prediction, Token Flow Control, Medium Access Control

Article History: received 28 August 2019; accepted 1 December 2019; published 18 December 2019.

1. INTRODUCTION

Wireless Network-on-Chip (WiNoC) interconnection emerged as a good energy efficient solution in order to enhance the performance efficiency of the conventional Networks-on-Chips (NoCs) [1, 2, 3, 4]. Nevertheless, to explore the entire advantages of this novel interconnect technology using a control algorithm, fair and efficient Medium Access Control (MAC) mechanism to enhance access of the on-chip wireless communication channel is needed [5, 6, 7]. To ensure an efficient utilization of the wireless medium in a WiNoC, in this work, a traffic aware Token Flow Control (TFC) mechanism is proposed for WiNoC to attain high link utilization, ultra-low latency, fast convergence, and low packets dropping. TFC utilizes tokens to define the concept of effective flows in Wireless Radio (WR). WiNoC is a potential solution that provides energy efficient communication while providing high bandwidth along with low latency, when the WIs are also equipped with an effective medium access control (MAC) unit [8]. The MAC requires high power transceivers to maintain the orthogonality among the coded channels, a token-based MAC mechanism was adopted by many WiNoCs. In the token passing MAC mechanism, the WIs located at various parts on the chip were organized in a virtual token ring. The permission to transmit via the wireless channel would be granted by the possession of a token at a WIs [9, 10, 11].

The movement of the token within the WIs are in round robin format, to ensure fairness in channel access. Although the token-based MAC mechanism enables simple, fair and distributed medium access, in such mechanism the period for which each WI possesses the token is independent of the volume of traffic passing through the WIs [12, 13, 14, 15]. To address such variation issue and enable an efficient wireless channel access, dynamic MAC mechanisms were proposed. In these proposed MAC mechanisms, the token possession period is adjusted dynamically depending on the volume of traffic passing through each WIs [14]. However, in order to adjust the token possession period, an accurate prediction mechanism is required to ensure maximum channel utilization. Networks-on-Chip (NoCs) are widely regarded as a promising approach for addressing the communication demands of large-scale Chip Multi-Processors (CMPs). NoC is a technology developed to address the communication deficiency experience in networking devices and demands of future many-core Systems-on-Chip (SoCs) due to its reusability, scalability and better bandwidth comparison with traditional approaches, it is fast emerging as an on-chip communication alternative for SoCs [16, 17, 18]. However, designing a high-performance low latency NoC with low area overhead has remained a challenge in this field. In academia and industry, NoCs have been more popular because the technology permits the integration of a large number of Intellectual Property (IP) cores into a single chip [19], (e.g. Tile64 [20] and Polaris [21]).

NoC are communication networks based on packet exchange, used to interconnect the cores of many-core chip

[22]. The principle components of an NoC architecture are shown in Figure 1, it is made up of the following components:

1. Routers: Its responsible for packet forwarding, according to a chosen routing protocol, until each packets' reach their destination;
2. Network Interfaces (NI): This is responsible for connecting each core to a router;
3. Links: These are the short wires between routers that consist of an interconnection medium.
4. IP core: The IP (intellectual property) core is a block of logic or data that is used in making a field programmable gate array (FPGA) or application-specific integrated circuit (ASIC) for a product.

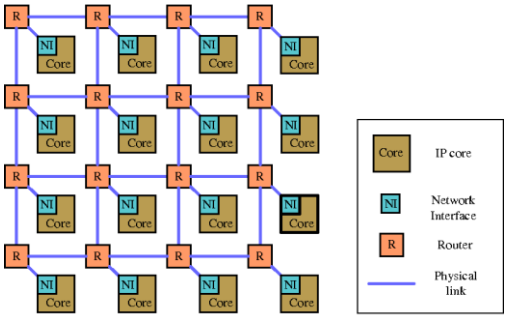


Figure 1. NoC Architecture [1]

WiNoCs are hybrid in nature and they use clustering techniques that will allow local communication through global wires while inter-cluster communication happens via wireless links. The WiNoC architecture is shown in Figure *2*. Wireless-NoC Architecture, in the WiNoC connection, three different categories of interconnections can take place within the WiNoC architecture, they are Tile-Tile, Tile-Hub and Hub- Hub as follow:

1. Tile-Tile is a wired point-to-point connection between two tile nodes.
2. Tile-Hub represents a wired point-to-point connection between a tile and a radio-hub element which can be either wired to wireless or wireless to wired connections.
3. A connection between two radio-hub elements which established a connection between two wireless routers is known as Hub-Hub interconnection [4, 23].

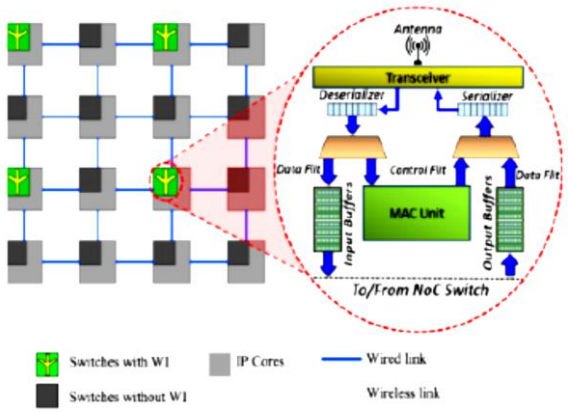


Figure 2. Wireless-NoC Architecture [24]

The rest of the paper is organized as follows: Section 2 provides a brief review of the related work in WiNoC routing algorithms. Section 3 elaborates on the methodology used in WiNoC traffic prediction. Section 4 reports the results and discussion on traffic prediction performance evaluation of the proposed methods. Finally, section 5 is the conclusion and future work of the paper.

2. RELATED WORK

Some recent works [1, 6, 11, 14, 24, 25, 26, 27] proposed the dynamic MAC mechanisms which are capable of dynamically and effectively adjust the transmission or time slots of each token based WIs on a prediction estimated by the demand of the Radio Hub (RH) (that are usually considered in daisy-chained ring topology). However, with the estimation time predicted, the token circulate via the radio hubs in round robin and enable the current radio hub that holds the token to use the radio medium in transmission mode for a period of clock cycles. Hence limits its performance benefits of using wireless link since all data transmission flows are in one circular direction and must pass through each RH on the network, result in a slower network. Therefore, to improve the efficiency in wireless channel access in WiNoC, an accurate prediction mechanism is required. In this work, we investigate the possibilities of an history-based prediction with hybrid artificial neural network (ANN) based on a prediction mechanism of dynamic MAC mechanisms for WiNoCs.

Wireless interconnection has emerged as an energy efficient solution to overcome delay experience in wireline conventional NoC, to fully utilized the benefit of this technology, an efficient MAC is required to ensure easy access to the on-chip wireless communication channel in WiNoC. To adapt to different traffic demand from application running on a multicore system, MAC mechanism must be design to adjust the token in the WIs based on predicted traffic demand of the RH rather than following the round robin techniques. There are many researches in this area but none take into consideration adjusting the TFC based on traffic demand, Narayana [28] proposed an ANN to predict the temperature profile in a multi-core chip, the ANN thermal predictor was trained and applied to predict the temperature at any given time based on the utilization of the chip components but not considering TFC and its was limited to temperature prediction only.

Murugesan [14] designed and trained the ANN to be able to predict the traffic demands of WIs in a WiNoC but the token still circulate within the WIs in a round robin format. Moreover, [1, 14, 29] integrated this prediction mechanism using ANNs with the token-time allocation mechanism but effect of traffic data on TFC was not considered as the token are circulated among the WIs in a round robin format, with increase in the number of WIs, it will require more time before returning to the initial state. Mansoor et.al [27] proposed a dynamic MAC mechanism in which the predicted bandwidth demand of the WIs are used for time slot allocation. The token time were adjusted based on predicted bandwidth demand but each token still circulated within the WIs in round robin format.

Mansoor et al. [29] improved on their earlier work and designed a low complexity and accurate traffic tracking mechanism that predict traffic demand of a WI. They proposed two dynamic MAC mechanisms that were able to adjust the slot durations based on the predictions. Then token time were adjusted based on the predicted bandwidth

demand and Proportional Integral Differential (PID) but the wireless token still circulated within the WIs in round robin format and it will require more time when the number of radio hub increases to meet the present request of moving from hundreds of cores to thousands of cores [30]. Hence, we propose an history plus ANN based traffic prediction algorithm technique that will reduce token waiting time by enable WIs with highest traffic to transmit, this can improve the efficiency and performance of the WiNoC.

3. METHODOLOGY

In this section, we present the experimental set-up of the proposed MAC based WiNoC architectures. Although, many different WiNoC architectures has been proposed in literature, for this work we have considered a Mesh based hybrid WiNoC architecture, as shown in Figure *2*. Wireless-NoC Architecture (i.e. WiMesh) with both wired and wireless links as a test case. We adopt a Mesh based WiNoC architecture as a test-bed for evaluating the proposed MACs. In the WiMesh architecture, each core is connected to a NoC switch using a wire line link. The switches are then connected with other switches in its cardinal directions (i.e. NSEW) using wireline interconnects to form a regular Mesh. The Mesh was chosen as it is a conventional NoC topology used in several multicore based products [31, 32] and is relatively easy to design, verify, and manufacture. To provide single-hop shortcuts among the distant NoC switches to reduce the data transfer over multihop wireline paths, the wireless interconnects are overlaid on top of this Mesh topology by deploying the WIs at some of the NoC switches.

3.1 Traffic Prediction

In order to predict the traffic demand of the WIs, a simple history-based prediction mechanism was used [26]. In the history-based prediction, the bandwidth demand of a WI in a token period is predicted based on the moving average of bandwidth demand of that WI over previous token periods. Hence, the predicted traffic B^{TPj+1} demand for token period j + 1 is calculated by equation (1).

$$\hat{B}^{TP^{j+1}} = \frac{BD^{TP^{j}} + B\bar{D}^{TP}}{2} \quad (1)$$

where BD^{TPj} is the actual traffic demand of a WI over the token period j and BD^{TP} is the average traffic demand for that WI from token period 0 to j − 1. The steady state demands of the WIs are captured by the moving average of the past token periods and the most recent variation in the demand is captured by the demand of the last period. Hence this prediction mechanism captures both long term and instantaneous traffic demands of a WI. This predicted demand value is then used to adjust the token possession period for the next token period.

4. RESULTS AND DISCUSSION

In this section, we evaluate the performance of the proposed dynamic MAC mechanisms in a Mesh based WiNoC (WiMesh) architecture. In our experiment, we use 8 RHs deployed over the conventional mesh-based architecture in which allocation of token to each RH depends on the traffic demand of the RH and also putting into consideration the fear sharing, so that a particular RH will not holds the token forever.

4.1 Uniform-random Traffic Pattern

Random traffic was generated using random number generation function in Python ranging from 50 to 200, the flow chart of the random number generation procedure is shown in

Figure 3. Flow Chart for Random Traffic Generator

.

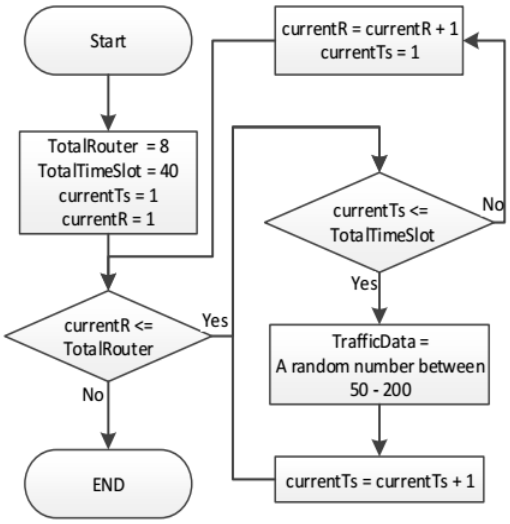


Figure 3. Flow Chart for Random Traffic Generator

4.2 History Based Prediction

History based prediction formular used in this research work is show in equation (2) with the following parameters: TS = Time Slot

$$TS^{j+1} = \frac{TS^{j} + \frac{\sum_{n=1}^{j-1}(TS)}{j-1}}{2} \quad (2)$$

Equation (2) is used for prediction based on history. To predict future time slot TS^{j+1}, we need the current data at TS^{j} and average of data from 0 to j-1, finally divide by 2. The flow chat in Figure 4 is the procedure on how to make the prediction based on past events.

The procedure shown in Figure 4: Flow Chart on History based Prediction was used to predict the RH to use token based on traffic generated history. The results of traffic generated used in figure 4 to predict which of the RH will been given access to holds token. Figure *5*:Predicted Token for Router 1 to Router 8 were graph generated for 8-Routers used with respect to the Time Slots from TS1 to TS40. The traffic generated from the first 20-TS (TS1 to TS20) were used as previous history to predict the future traffic for the next 20-TS (TS21 to TS40), then allocated the token. The blue lines in the graph generated in figure 5 is the token allocation based on traffic why the red lines are token allocation of the traffic allocation based on prediction in respect of the past history-based traffic.

4.3 Token Movement

This section compares the effect of normal token round robin with the token movement with respect to traffic requirement. Eight routers were used and subjected to the two conditions. The token movement is based on traffic demand on each router, this mean that the program decided which router will get next token in a round. The main rules used in this process is that token will be given to a router once in a round and based on RH with the highest traffic. The diagram in Figure *6*: Token Movement Comparison and Figure *7*: Token Allocation based on Maximum Traffic show the flow chart that describe the procedure of token allocation based on the traffic and the graph of the results generated comparing with the normal token in round robin with token allocation based on traffic.

5. CONCLUSION

WiNoC is an emerging communication technologies considered as a viable solution to enhance the performance of SoC. WiNoC architectures provide a scalable method to on-chip communication, however the bandwidth supplied by WiNoCs can be utilized efficiently only with the presence of an effective flow control algorithms. However, the flow control algorithms published in the literatures up-to date either rely on local information or be afflicted by large communication overhead and unpredictable traffic. Therefore, using them in the WiNoC context limit the benefit of RH. In this paper, we proposed a dynamic MAC traffic token flow control mechanism mainly on WiNoC to predict the RH token allocation but not in the daisy-chained ring topology (round robin). The MAC mechanism was able to effectively adjust the transmission slots of the RH in respect of the traffic, allocate token in respect of the highest traffic in each RHs, putting into consideration fear sharing by ensuring that no RH were given twice in a round and been able to predict future transmission based on past history. The traffic demand prediction mechanism using the history-based prediction is unable to capture the trend when there are sudden variations in the traffic which is common in on-chip environments. ANNs on the other hand, can be trained to predict and track variations in traffic with higher accuracy. Hence in the future work, ANN based prediction mechanism would be explored to predict the bandwidth demand of the RHs. Moreover, to reduce token waiting time and inefficient decision making for radio hub-to-hub communication and tile-to-hub communication, congestion-aware routing in WiNoC would be develop to enhance end to end latency.

Figure 4: Flow Chart on History based Prediction

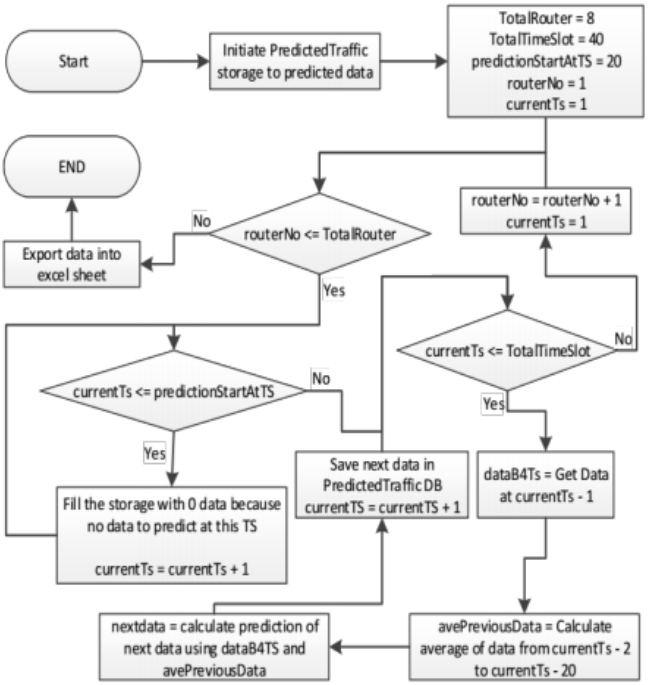
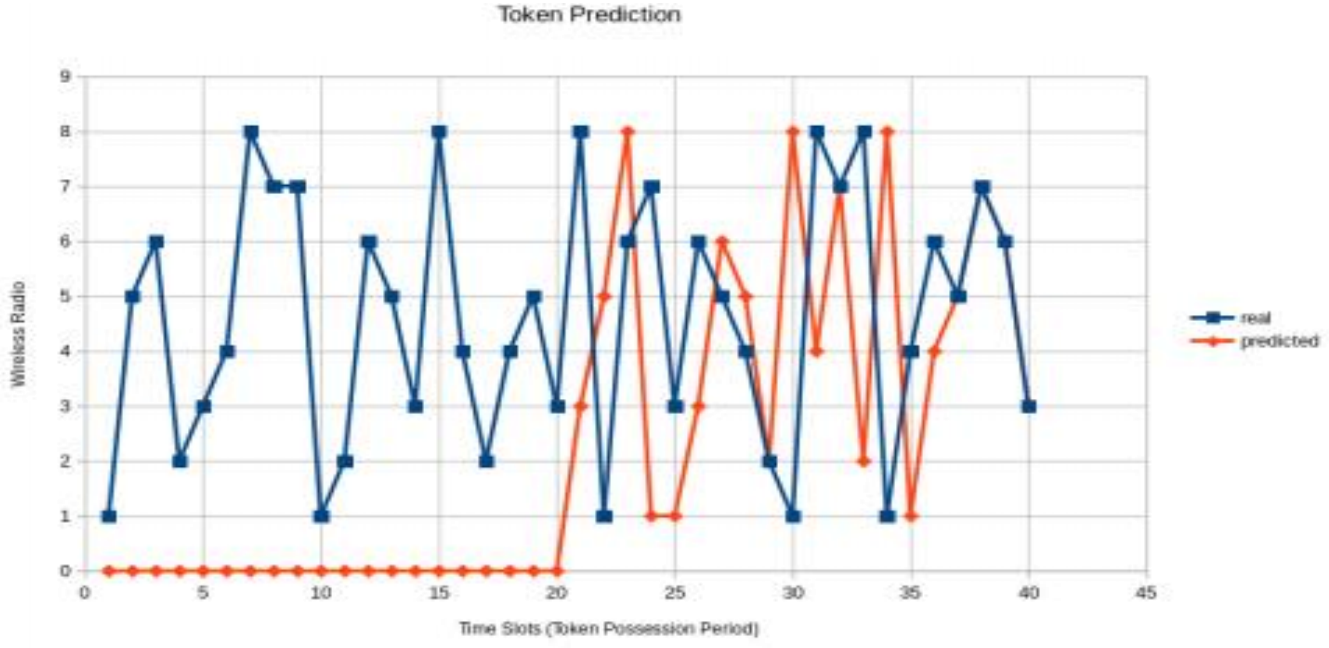


Figure 5:Predicted Token for Router 1 to Router 8



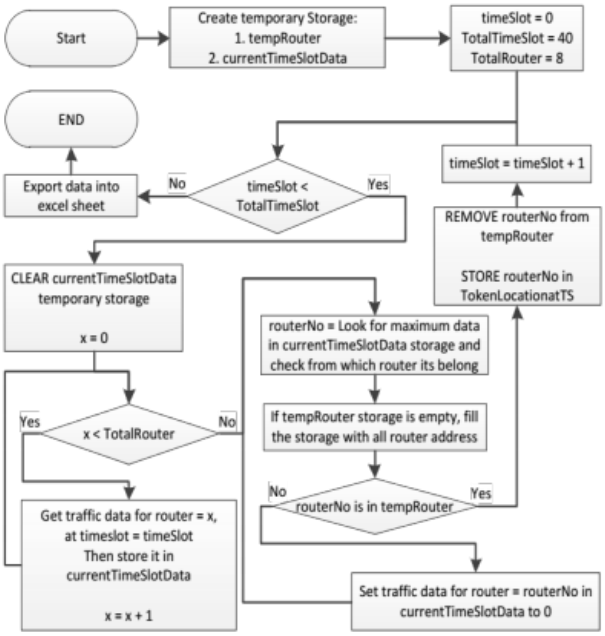


Figure 6: Token Movement Comparison

REFERENCES

[1] N. Mansoor, "Robust and Traffic Aware Medium Access Control Mechanisms for Energy-Efficient mm-Wave Wireless Network-on-Chip Architectures," scholarworks.rit.edu, NY, 2017.

[2] M. O. Agyeman, W. Zong, J. Wan, A. Yakovlev, K. Tong, and. T. Mak, "Novel hybrid wired-wireless network-on-chip architectures: Transducer and communication fabric design," in 9th International Symposium on Networks-on-Chip, UK, 2015.

[3] C. Wang,.W. H. Hu and N. Bagherzadeh, "A Wireless Network-on-Chip designfor multicore platforms," in 19th International Euromicro Conference on Parallel, Distributed and Network-based Processing , 2011.

[4] S. Wang and T. Jin, "Wireless network-on-chip: A survey," The Journal of Engineering, vol. 3, pp. 98-104, 2014.

[5] 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 21st IEEE International Conference on Application-specific Systems, Architectures and Processors, 2010.

[6] K. Chang, S. Deb, A. Ganguly, X. Yu, S. P. Sah, P. P. Pande, B. Belzer, and D. Heo, "Performance evaluation and design trade-offs for wireless network-on-chip architectures," ACM Journal on Emerging Technologies in Computing Systems (JETC), vol. 8, no. 3, p. 23, 2012.

[7] S. Deb, S. P. Sah, M. Cosic, X. Yu, A. Ganguly, K. Chang, P. Pande, B. Beizer, 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.

[8] S. Abadal, A. Mestres, J. Torrellas, E. Alarcón, 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.

[9] Z. Qian, S. M. Abbas, and C. Y. Tsui, "Fsnoc: A flit-level speedup scheme for network on-chips using self-reconfigurable bidirectional channels," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 23, no. 9, pp. 1854-1867, 2014.

[10] A. Kannan, N. E. Jerger, and G. H. Loh, "Enabling interposer-based disintegration of multi-core processors," in 48th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO) IEEE., 2015.

[11] Ganguly, M. Ahmed, R. Singh Narde, A. Vashist, M. Shamim, N. Mansoor, T. Shinde, S. Subramaniam, S. Saxena, J. Venkataraman, and M. Indovina, "The Advances, Challenges and Future Possibilities of Millimeter-Wave Chip-to-Chip Interconnections for Multi-Chip Systems," Journal of Low Power Electronics and Applications, vol. 8, no. 1, p. 5, 2018.

[12] S. Deb, A. Ganguly, P. P. Pande, B. Belzer, and D. Heo, "Wireless NoC as interconnection backbone for multicore chips: Promises and challenges," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 2, no. 2, pp. 228-239, 2012.

[13] K. Kim, B. Floyd, J. Mehta, H. Yoon, C. M. Hung, D., Bravo, T. Dickson, X, Guo, R Li, N. Trichy, and J. Caserta, "The feasibility of on-chip interconnection using antennas," in IEEE/ACM International conference on Computer-aided design, IEEE Computer Society., 2005.

[14] R. Murugesan, "Artificial Neural Network Based Prediction Mechanism for Wireless Network on Chips Medium Access Control," 2017.

[15] M. Gholipour, A. T. Haghighat, and M. R. Meybodi "Hop-by-hop traffic-aware routing to congestion control in wireless sensor networks," EURASIP Journal on Wireless Communications and Networking, vol. 1, no. 1, p. 15, 2015.

[16] M. Ebrahimi, "Adaptive Routing Approaches for Networked Many-Core Systems.," 2013.

[17] A. Monemi, C. Y. Ooi, M. N. Marsono "Low latency network-on-chip router microarchitecture using request masking technique," International Journal of Reconfigurable Computing, p. 2, 2015.

[18] D. Wu, B. M. Al-Hashimi, and M. T. Schmitz "Improving routing efficiency for network-on-chip through contention-aware input selection," in 2006 Asia and South Pacific Design Automation Conference, 2006.

[19] W. J. Dally and B. Towel "Route packets, not wires: on-chip inteconnection networks," in In Proceedings of the 38th annual Design Automation Conference, 2001.

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

[21] A. M. Amorim, P. A. Oliveira, and H. C. Freitas, "Performance evaluation of single-and multi-hop wireless networks-on-chip with NAS Parallel Benchmarks," Journal of the Brazilian Computer Society, vol. 21, no. 1, p. 6, 2015.
[22] P. Guerrier and A. Greiner, "A generic architecture for on-chip packet-switched interconnections.," Design, Automation, and Test in Europe, Springer,, Dordrecht, 2008.
[23] M. A. I. Sikder, "Emerging technologies in on-chip and off-chip interconnection network," Doctoral dissertation, Ohio University, Ohio, 2016.
[24] D. DiTomaso, A. Kodi, S. Kaya, and D. Matolak, "iWISE: Inter-router wireless scalable express channels for network-on-chips (NoCs) architecture," in IEEE 19th Annual Symposium on High Performance Interconnects, 2011.
[25] N. Mansoor, P. J.S. Iruthayaraj and A. Ganguly, "Design methodology for a robust and energy-efficient millimeter-wave wireless network-on-chip," IEEE Transactions on Multi-Scale Computing Systems, vol. 1, no. 1, pp. 33-45, 2015.
[26] A. K. Mishra, N. Vijaykrishnan, and C. R. Das, "A case for heterogeneous on-chip interconnects for CMPs," ACM SIGARCH Computer Architecture News, vol. 39, no. 3, pp. 389-400, 2011.
[27] N. Mansoor, M. S. Shamim, and A. Ganguly,, "A demand-aware predictive dynamic bandwidth allocation mechanism for wireless network-on-chip," in 18th System Level Interconnect Prediction Workshop, ACM, 2016.
[28] S. Narayana, "An artificial neural networks based temperature prediction framework for network-on-chip based multicore platform," arXiv preprint arXiv:1612.04197, 2016.
[29] N. Mansoor, A. Vashist, M. M. Ahmed, M. S. Shamim, S. A. Mamun, and A. Ganguly, "A Traffic-Aware Medium Access Control Mechanism for Energy-Efficient Wireless Network-on-Chip Architectures," arXiv preprint arXiv:1809.07862., 2018.
[30] K. Astrom and T. Hagglund, "PID controllers: theory, design, and tuning," Research Triangle Park, NC: Instrument society of America, US, 1995.
[31] P. Salihundam, S. Jain, T. Jacob, S. Kumar, V. Erraguntla, Y. Hoskote, S. Vangal, G. Ruhl, and N. Borkar, "A 2 Tb/s 6X4 Mesh Network for a Single-Chip Cloud Computer With DVFS in 45 nm CMOS," IEEE Journal of Solid-State Circuits, vol. 46, no. 4, pp. 757-766, 2011.
[32] S. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, P. Iyer, A. Singh, T. Jacob, and S. Jain, "An 80-tile 1.28 TFLOPS network-on-chip in 65nm CMOS," in 2007 IEEE International Solid-State Circuits Conference, 2007.

Figure 7: Token Allocation based on Maximum Traffic