# A Demand-Aware Predictive Dynamic Bandwidth Allocation Mechanism for Wireless Network-on-Chip

Naseef Mansoor, Md Shahriar Shamim, Amlan Ganguly
Rochester Institute of Technology
Rochester, NY, USA
{nxm4026, ms5614, axgeec}@rit.edu

## ABSTRACT

Long distance data communication over multi-hop wireline paths in conventional Networks-on-Chips (NoCs) cause high energy consumption and degradation in bandwidth. Wireless interconnects in the millimeter-wave band have emerged as an energy-efficient interconnection paradigm for multi-core chips interconnected with NoCs. However, spatial variations in traffic distribution and temporal variations in workloads can exert variable bandwidth demands on the NoC fabric. Wireless interconnects which do not require a physical layout of interconnects can be utilized to mitigate this issue. In order to dynamically allocate variable bandwidth to the wireless transceivers depending on the demand, the design of a dynamic and efficient Medium Access Control (MAC) mechanism to grant access to the on-chip wireless communication channel is needed. In this paper, a history based predictor, which can predict the bandwidth demand of the wireless nodes in the wireless NoC is designed. Based on these predicted demands we propose the design of two MAC mechanisms that are able to dynamically allocate bandwidth to the wireless transceivers. Through system level simulations, we show that the demand-aware MAC mechanisms are more energy efficient as well as capable of sustaining higher data bandwidth in wireless NoCs.

## CCS Concepts

• Hardware → **Radio frequency and wireless interconnect.**

## Keywords

Network-on-Chip, Wireless interconnect, Dynamic bandwidth allocation, medium access mechanism

## 1. INTRODUCTION

Network-on-Chip (NoC) has emerged as a communication infrastructure for the multi-core System-on-Chips (SoCs) [3]. However, due to multi-hop data communication over the metal interconnects, traditional mesh based NoC architectures are performance and energy inefficient. Long range metal wires in a mesh based NoC [19] and ultra-low-latency and low-power express channels between communicating cores [15] have been proposed as a solution to overcome these inefficiencies. However, according to International Technology Roadmap for Semiconductors (ITRS, 2012) the performance gain of these approaches is limited due to metal/dielectric based interconnection paradigm. Hence, novel interconnect technologies such as on chip photonic interconnects

[22], on-chip multi-band RF transmission line interconnects (RFI) [6], and wireless interconnects [14] have been explored in recent time as an energy efficient solution for long range on-chip data communication. Both photonic and RF interconnect based NoCs are capable of achieving low latency and low power dissipation due to single hop communication between distant cores. However, these technologies need the additional physically overlaid optical waveguides or microstrip transmission lines to enable data transmission. On the other hand, CMOS compatible long-range wireless shortcuts operating in the millimeter-wave (mm-wave) frequencies [9] do not require laying out of physical interconnects. Wireless NoCs (WiNoCs) with on-chip miniature antennas operating in the mm-wave bands are shown to be able to communicate between wireless interfaces (WIs) deployed across a die to improve the performance [9]. However, the bandwidth of the mm-wave wireless channels is limited by the state-of-the-art transceiver design. Design of multiple non-overlapping channels to enable Frequency Division Multiple Access (FDMA) is a non-trivial challenge from the perspective of transceiver design and is not easily scalable. Hence, multiple WIs share a single frequency wireless channel. Consequently, such WiNoCs require a Medium Access Control (MAC) mechanism which will enable multiple WIs to share this wireless channel without any interference and ensure optimal utilization of the available bandwidth. Code Division Multiple Access (CDMA) based MAC mechanism proposed in [25], requires overheads for maintaining synchronization for the sake of preserving orthogonality between code channels in transmitters. Hence, simple and distributed Time Division Multiple Access (TDMA) based MAC mechanism like token passing [5], [11] has been proposed for on-chip mm-wave technology. In the token passing MAC mechanism, a wireless token circulates among the WIs in a round robin fashion to ensure fairness.

Depending upon dynamic task mapping, task migration and varying workloads, modern and future multi-core chips will have dynamically varying traffic patterns. This spatial and temporal variation in traffic patterns is also expected in future heterogeneous SoCs integrating CPU, GPU, ASIC and memories on the same die [8]. As shown by the authors in [18], even for a uniform random traffic pattern all NoC routers are not utilized identically (i.e. spatial variation) due to particular routing algorithms. A heterogeneous NoC architecture with low and high bandwidth links and switches was proposed to address this issue in [18]. Such variation in traffic patterns through NoC switches is reflected in the wireless channel bandwidth demands at the WIs of the WiNoCs as well. This is depicted in Figure. 1 where the bandwidth demands measured as the data rate through the various WIs are observed. We monitor the bandwidth demand of the WIs measured as the normalized data rate through them in a 64 core WiNoC with 12 WIs. The WIs are distributed in a mesh based NoC with wormhole switching according to the optimization heuristic described in [17]. Uniform random traffic following a self-similar temporal injection pattern is used for this evaluation. From the spatial variation captured in
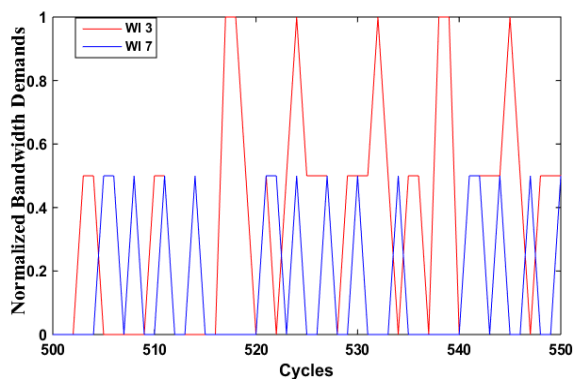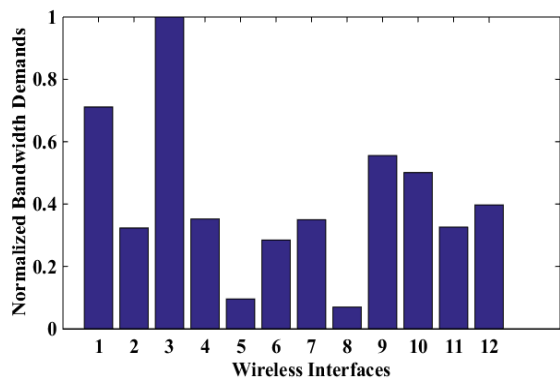
**(b)**

**Figure 1. Spatial and Temporal Variation in WI bandwidth demands. a) Spatial variation in bandwidth demand among different WIs. b) Temporal variation in normalized bandwidth demands.**

Figure. 1(a), we see all the WIs have different data rates and hence different demands on the wireless bandwidth even for uniform traffic distribution. It can be observed that the bandwidth demand varies widely between the WIs. Figure. 1(b) shows the temporal variation in normalized bandwidth demand for two selected WIs in the system to highlight the fact that not only spatially the bandwidth demand varies temporally as well even for uniform random traffic.

As channel bandwidth is an important resource in all types of WiNoCs, judicious and adaptive deployment of that resource is key to maximizing the performance and energy gains in such novel NoC paradigms. Although a dynamic resource or bandwidth allocation mechanism can be designed and used for any WiNoC architecture regardless of the baseline MAC, in this work we evaluate our dynamic allocation mechanism for a token passing based WiNoC as it has a simple distributed MAC with a CMOS compatible mm-wave physical layer. In the token based WiNoC, the duration of the token at any WI governs the time for which that WI has access to the shared channel to transmit data. Hence, the basic principle of this work is to give longer access to the WI that experiences a higher volume of traffic being transmitted through it generating a higher demand for bandwidth. This will reduce underutilization of access slots in WIs that do not experience a high volume of traffic while granting longer access to those that need it resulting in an improvement in overall system performance. In order to satisfy this spatial and temporal variation in the bandwidth demand of the WIs, each WI should be able to dynamically adjust their transmission duration in the token based TDMA MAC scheme. However, such systematic dynamic MACs for WiNoCs have not been investigated extensively in literature. In this work, we propose two distributed dynamic MAC mechanisms: Proportionate TDMA (P-TDMA) and Dynamic TDMA (D-TDMA), where the WIs are able to dynamically adjust the transmission duration based on the predicted demand of the WIs. We propose a methodology to predict the bandwidth demand of a WI based on current and past bandwidth demands of the WIs. Based on these predicted demands the dynamic bandwidth allocation mechanisms in the MAC will enable WIs with higher demands access the wireless medium for longer. Through detailed system level simulations, we show that the proposed MAC mechanisms utilizing the predicted demands in wireless bandwidth perform better than the baseline token passing MAC for both synthetic and application based workloads.

## 2. RELATED WORKS

A comprehensive survey regarding various WiNoC architectures and their design principles is presented in [10]. A wireless NoC architecture augmented with directional on-chip planner log periodic antennas is explored in [23] for simultaneous multi-channel communications. However, placing multiple directional antennas without interference among them is not trivial. In order to communicate via the wireless channel, a MAC mechanism is required that allows WIs to transmit through the wireless channel without any interference. Due to energy, area and memory constraints, complex MAC mechanisms for WiNoCs are not suitable. Hence, design of low-overhead and efficient MAC scheme is one of the major challenges in designing a WiNoC as identified in [1]. A synchronous and distributed medium access mechanism (SD-MAC) is proposed in [26] for the Ultra-Wide-Band (UWB) WiNoCs where impulse based transceivers are used that limits the communication range of the WIs to a millimeter. In order to acquire access to the medium, local arbitration between the WIs using wired links is required in such MAC mechanism. Hence, such MAC mechanism cannot be adopted for WiNoCs where the WIs is more than a millimeter apart. In [14], a hybrid MAC mechanism combining both TDMA and FDMA is reported for WiNoCs based on Carbon Nanotube (CNT) antennas. However, the CNT based wireless technology is difficult to integrate in current CMOS process. On the other hand, miniature antennas operating in the mm-wave frequencies are CMOS compatible and are nearer term solution [9]. A CDMA based MAC mechanism is proposed in [25] for mm-wave WiNoCs to efficiently utilize the wireless bandwidth. Orthogonal Walsh codes are used in this MAC mechanism to enable concurrent wireless transmission through the wireless channel. However, the transceivers should be precisely synchronized to ensure the orthogonality of the code channels. Such synchronization is difficult to achieve among WIs distributed over a large multicore chip. Similar to the CDMA MAC mechanism, a distributed MAC mechanism is proposed in [13] for mm-wave WiNoCs that uses simple orthogonal request packets. These request packets are processed at each WI and permission to the wireless channel is granted by a priority based mechanism. However, maintaining orthogonality among these channels is difficult to achieve. Moreover, this method has an overhead of maintaining the state of current transmission at each transceiver. It is shown in [16] that CSMA based MAC mechanisms suffer from degradation in performance and energy efficiency with increase in traffic load due to higher probability of contention. A token based MAC mechanism is adopted for several mm-wave WiNoCs [5], [11]. In the token passing based MAC mechanism, the access to wireless medium is granted by a token circulation among the transceivers and requires no global synchronization mechanism. However, such token passing mechanism is agnostic of the

utilization of WIs. A dynamic radio access control mechanism (RACM) utilizing token passing mechanism is proposed in [20] where the unused token slots are redistributed among the WIs. However, this mechanism does not take into account the bandwidth requirement or actual utilization of the WIs. In this work, we propose a dynamic MAC mechanism where the predicted bandwidth demand of the WIs is used for time slot allocation.

# 3. TOKEN BASED MAC MECHANISM FOR WINOCS

The MAC mechanism is required to ensure an interference free communication via the wireless medium. Complex MAC mechanisms used in macro scale networks are not suitable for on chip environment due to their high implementation overhead [13]. Hence, in order to access the energy efficient wireless medium in a distributed fashion, authors in [5], [11] proposed a low-overhead token passing MAC mechanism for WiNoCs. In this section, first we discuss the baseline token passing MAC mechanism for WiNoCs followed by the required bandwidth prediction mechanism and the two proposed dynamic MAC mechanisms: P-TDMA MAC and D-TDMA MAC.

## 3.1 Token Passing MAC

In a token passing MAC mechanism, the access to wireless medium is granted by the possession of a token. Only the WI possessing the token can transmit via the wireless medium. The token circulates between the WIs as a token flit in a round robin fashion to ensure fairness of access to the wireless medium. Each WI holds the token for a fixed number of time slots where one time slot is same as the system clock cycle. After this allocated number of time slots, a WI passes the token to the next WI. We define this number of time slots used by a WI as the token possession period (i.e. *tpp*). The number of time slots required for the token to complete one circulation through all the WIs and return to an initial WI is defined as the token period (i.e. *TP*). Hence, the token period contains both the time slots for data flits as well as the token flits. In order to enable such MAC, each WI needs to be equipped with a MAC unit. The MAC unit contains three registers, ***ID_self***, ***ID_next*** and ***HasToken***. The ***ID_self*** and ***ID_next*** stores the address of that WI and the address of the next WI where token will be sent after the token possession period. The ***HasToken*** indicates the presence of token in the WI. When a token flit with a destination address set to ***ID_self*** is received, the MAC unit sets the ***HasToken*** and initiates the token possession



**Figure 2. Architecture of Predictive Dynamic MAC unit.**

period counter, $C_{tpp}$. When this counter expires, indicating the end of the token possession period, a token flit containing the fields *TokenID*, *NextWI*, and *PrevWI* is constructed and transmitted by the WI currently possessing the token. The field *TokenID* is an identifier to differentiate the token flit from data flit transmitted via the wireless medium. The ***ID_next*** and ***ID_self*** are used to set the field, *NextWI* and *PrevWI* respectively. Although the token is circulated among the WIs in a round robin fashion, these fields are necessary in the token to enable a distributed token passing mechanism without relying on synchronization between the WIs distributed in the entire WiNoC. We consider this MAC scheme as the baseline in this work. However, due to the dynamic traffic variation, the token possession period of the WIs need to be dynamically adapted.

## 3.2 Bandwidth Demand Prediction Mechanism

In the proposed MAC mechanisms, a simple history based predictor is used to predict the bandwidth demand of a WI. A history based predictor is chosen to reduce the overheads of the prediction algorithm. The predicted bandwidth demand, $\hat{B}^{TP^{j+1}}$ for token period *j+1* is calculated by,

$$\hat{B}^{TP^{j+1}} = \frac{BD^{TP^j} + \overline{BD^{TP}}}{2} \qquad (1)$$

Where $BD^{TP^j}$ is the actual bandwidth demand of a WI, measured as the total number of incoming flits in the wireless port over the token period *j* and $\overline{BD^{TP}}$ is the average predicted bandwidth demand for that WI from token period *0* to *j-1*. The moving average of the past token periods capture the steady state demand of the WIs. Whereas, the demand of the last period captures the transient or most recent variation in the demand. Hence, the predicted bandwidth demand captures both long term and instantaneous bandwidth demands of a WI. Such a moving average based prediction is a common method based on the discretization of the principles of Proportional and Integral (PI) feedback control [2]. This predicted demand value is then used to allocate the time slots for the next token possession period.

## 3.3 P-TDMA MAC

In the P-TDMA scheme, the token possession period for each WI is dynamically adapted based on the predicted proportional bandwidth demand of the WI compared to other WIs. The number of time slots in the token possession period of a WI is allocated dynamically at the start of each token period to cope with the varying bandwidth demand of the WIs. However, this allocation of time slots is constrained in such a way that the token period remains constant between allocations. The allocated time slots of a WI, *i* at token period, *j+1*, $s_i^{TP^{j+1}}$ is given by,

$$s_i^{TP^{j+1}} = \frac{\hat{B}_i^{TP^j}}{\sum_{i=1}^{N} \hat{B}_i^{TP^j}} \times S_{TP} \qquad (2)$$

where, $\hat{B}_i^{TP^j}$ is the predicted bandwidth demand for WI *i* at token period *j* calculated using (1), $S_{TP}$ is the number of time slots for data flits in the token period and *N* is the number of WIs. Due to this proportional allocation of time slots, WIs with greater predictive bandwidth demands will have more time slots in the token possession period compared to those with lower bandwidth demands.

To enable this P-TDMA mechanism the dynamic MAC unit at each WI contains three counters: demand counter, $C_{demand}$, token period counter, $C_{TP}$ and token possession period counter, $C_{tpp}$. The MAC
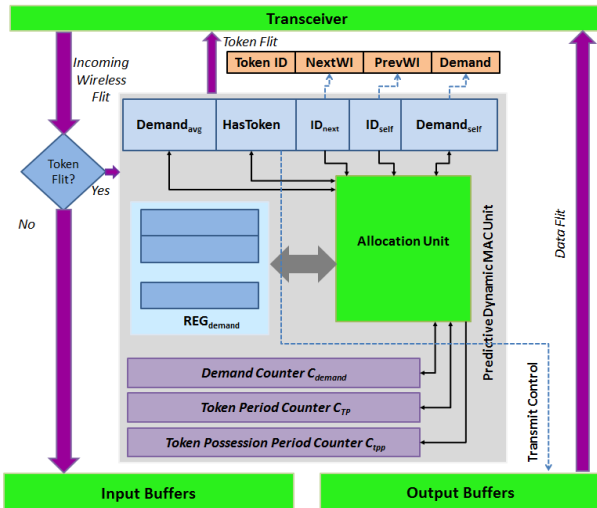
unit also contains five registers: $ID_{self}$, $ID_{next}$, HasToken, $Demand_{avg}$ and $Demand_{self}$ to store its own ID, ID of the next WI in the round robin circulation of token, indicate the possession of the token, average bandwidth demand and the predicted bandwidth demand of the WI. A register file, $REG_{demand}$ is used to store the predicted bandwidth demand of other WIs. The demand counter, $C_{demand}$ is used to count the utilization over a token period. The token period counter, $C_{TP}$ counts down to zero from the constant token period value. When the token period counter, $C_{TP}$ expires, the $C_{tpp}$ is loaded with number of time slots for the next token period, calculated using (2). Then the value of $C_{demand}$ and $Demand_{avg}$ is used to calculate the predicted bandwidth demand. This prediction is stored in the $Demand_{self}$. Then, $Demand_{avg}$ is updated using the $Demand_{avg}$ and $Demand_{self}$. After this, $C_{demand}$ is reset to zero to capture the bandwidth demand of the next token period. However, in order to determine the number of time slots in a distributed fashion, the value of register, $Demand_{self}$ of each WI is shared with other WIs. To achieve this, we propose to add a field Demand in the token flit, populated with $Demand_{self}$ to share the predicted bandwidth demand of the WI passing the token. When this token flit is broadcast by the releasing WI and received by other WIs (due to non-directional zig-zag antennas), the value in the field Demand is used to update the register file, $REG_{demand}$ at all the WIs. Even if a WI has no data packets to transmit, it receives the token from the previous WI, updates the token with its demand and passes it to the next WI.

## 3.4  D-TDMA MAC

In the D-TDMA scheme, the token possession period for each WI is dynamically adapted based on the predicted bandwidth demand of the WI. However, unlike the P-TDMA scheme, the number of time slots in the token possession period is equal to the predicted bandwidth demand of a WI. Hence, the token period for D-TDMA scheme changes each time the token possession periods are calculated. The allocated time slots in the token possession period of a WI, $i$ at token period, $j+1$, $s_i^{TP^{j+1}}$ is given by,

$$s_i^{TP^{j+1}} = \min(\hat{B}_i^{TP^j}, M) \tag{3}$$

where, $\hat{B}_i^{TP^j}$ is the predicted bandwidth demand for WI $i$ at token period $j$ and $M$ is a maximum number of slots that can be allocated to any WI. This maximum number of slots ensures no WI has to wait for a large number of time slots to get access to the wireless medium. Then the number of time slot (including both data flit and token flit) in the new token period $TP^{j+1}$ calculated at the end of the current token period $TP^j$ is given by,

$$TP^{j+1} = \sum_{i=1}^{N} s_i^{TP^{j+1}} + N \tag{4}$$



**Switches with WI**     **Wired link** ——
**Switches without WI**  **Wireless link** ·········
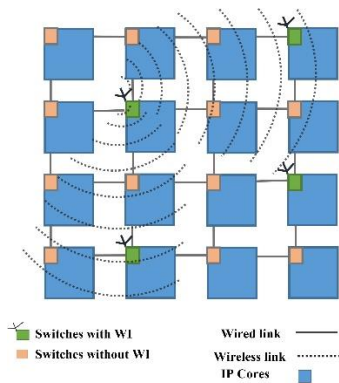                         **IP Cores** ▮

**Figure 3. WiMesh Architecture.**

where, $s_i^{TP^{j+1}}$ is the number of allocated time slots in the token possession period for WI $i$ at token period $j+1$ and $N$ is the number of WIs in the system. Hence, the maximum number of slots for countdown in the token period counter, $C_{TP}$ can be set to $N(M+1)$ thus governing its size. During operation the value of the new token period computed from (4) will be used to reset the counter unlike in the P-TDMA scheme. The MAC unit for the D-TDMA scheme contains same registers and counters as the P-TDMA scheme. The functionality of D-TDMA is same as that of the P-TDMA except for the allocation logic which follows (3).

In case of both the P-TDMA and the D-TDMA, the token carries the fields TokenID, PrevWI, NextWI and Demand to enable the regular token passing mechanism as well. The architecture of the proposed MAC mechanisms along with the token flit format is shown in Figure 2.

## 4.  EXPERIMENTAL RESULTS

In this section, we evaluate the performance and energy efficiency of the proposed dynamic MAC mechanisms in a mesh based WiNoC (WiMesh) architecture as a test case. Figure 3 shows a conceptual schematic of the WiMesh architecture. In our experiments, we use 12 WIs deployed over the conventional mesh based architecture wherein the location and number of the WIs are obtained following the heuristics designed in [17]. In this 2-step heuristic, we first iteratively optimized the placement of WIs by following a Simulated Annealing heuristic to minimize the average hopcount. Then we compare the performance of these configurations for different number of WIs and the one with the best bandwidth is chosen. We also compare the performance and energy efficiency of this WiNoC equipped with the dynamic MACs with the wired counterpart which is a conventional mesh (Mesh). Consequently, the Mesh and WiMesh have the same wireline topology with the WIs being additionally deployed in the WiMesh. We adopt wormhole switching [12] in both wired and wireless links. In the Mesh architecture, we have adopted dimension order (XY) routing which is shown to provide a deadlock free shortest path routing. In the WiMesh, the presence of the wireless links create shortcuts in the mesh and hence we adopt a shortest path routing to optimize network performance. We use a forwarding-table based routing over pre-computed shortest paths determined by Dijkstra's algorithm. This forwarding table only contains the address of the next switch in the shortest path to final destinations. Hence, each switch only has local forwarding information eliminating the need for maintaining non-scalable global routing information. Deadlock is avoided as by transferring flits along the shortest path routing tree extracted by Dijkstra's algorithm, as it is inherently free of cyclic dependencies. Using this WiMesh architecture as the baseline architecture we evaluate and compare the various demand-aware MAC mechanisms.

## 4.1  Metrics and Methodologies for Evaluation

We evaluate the proposed token based dynamic MAC mechanisms in the above mentioned WiMesh architecture platform in terms of bandwidth, energy efficiency, and packet latency. The bandwidth is measured as the data rate in bits per second successfully routed at each destination core in the NoC. The energy efficiency is measured as the packet energy, defined as the average energy (i.e. both switch and link energy) required to successfully route an entire packet from source to destination. The packet latency is measured as the number of clock cycles required to transmit one whole packet from source to destination. The average packet latency for the NoCs are estimated using a cycle accurate NoC simulator. The NoC simulator models the progress of the data flits accurately per clock
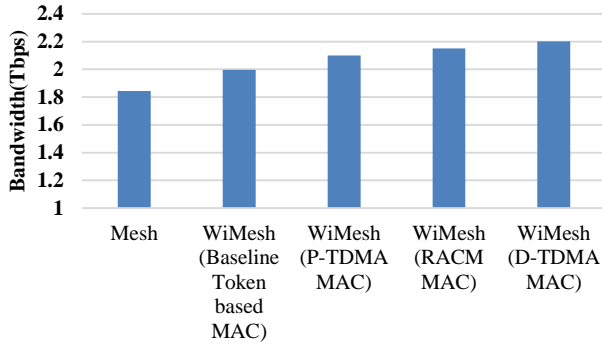
**Figure 4. Bandwidth for Mesh and WiMesh with different MACs with uniform random synthetic traffic pattern.**



**Figure 5. Packet energy for Mesh and WiMesh with different MACs with uniform random synthetic traffic pattern.**

cycle accounting for those flits that reach the destination as well as those that are stalled. We have considered a system size of 64 cores for the experiments as it is representative of current trends in multicore chip design in the industry. Ten thousand iterations were performed eliminating transients in the first thousand iterations. The width of all wired links is considered to be same as the flit size, which is considered to be 32 bits. We consider a moderate packet size of 64 flits for all our experiments. Each switch is considered to have 4 VCs with a buffer depth of 2. As the WIs handle a large volume of traffic, an increased number of VC of 8 with 16 buffer depth is used. For the baseline token passing mechanism the token possession period is considered to be 64 time slots. The bandwidth is also estimated using the NoC simulator by monitoring the number of bits arriving successfully at each core per cycle.

To estimate the average packet energy we need to estimate the energy consumption of the packets through the switches, as well as wired and wireless links. The link energy is calculated by determining the energy required to send the packet through the wired or wireless interconnects. The delay and energy dissipation on the wired link is obtained through Cadence simulations taking into account the specific lengths of each link based on the established topology in the 20mmx20mm die. For the wireless interconnects, the on-chip metal zigzag antenna and wireless transceiver designs are adopted from [5]. The wireless transceiver is shown to dissipate 2.3pJ/bit sustaining a data rate of 16Gbps with a bit-error rate (BER) of less than $10^{-15}$ while occupying an area of 0.3mm$^2$ in post-layout design using TSMC 65nm CMOS process. The NoC switches and the MAC units are synthesized from a RTL level design using 65nm standard cell libraries from CMP [7], using Synopsys. The delay and energy dissipation of these components are then incorporated in a cycle accurate NoC simulator to evaluate the packet energy. The NoC switches are driven with a 2.5GHz clock and 1V $V_{dd}$, which are the nominal frequency and voltage for the 65nm technology node. In the next sections, we present the results for synthetic and application specific traffics that demonstrate the effectiveness of the proposed bandwidth-aware dynamic MACs.

## 4.2 Comparative Performance Evaluation of the Proposed MACs with synthetic traffic

In this section, we evaluate the various MACs as discussed in section 3 on the WiMesh platform with synthetic traffic. For this experiment, we consider both uniform and non-uniform synthetic traffic patterns i.e. HotSpot and Transpose to capture the variation of the bandwidth demands on the wireless interconnects.
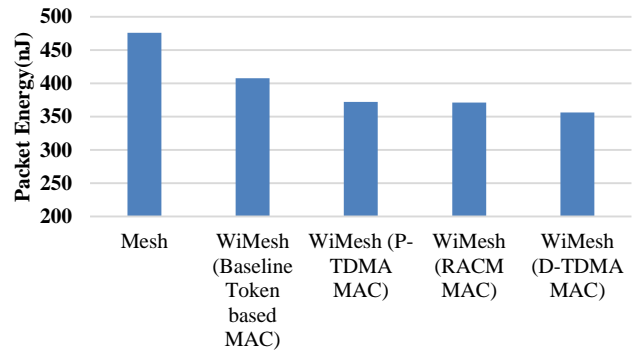
### 4.2.1 Performance evaluation for uniform random synthetic traffic pattern

In this section, we evaluate the performance of different MAC mechanisms for uniform random synthetic traffic pattern. In this traffic pattern, each core can address packets to any other core with equal probability. The peak bandwidth at network saturation for different MAC mechanisms as well as for the wired Mesh is shown in Figure 4. The wireline Mesh architecture has the lowest bandwidth compared to that of all the WiMesh architectures discussed in this paper. This is because, in the Mesh architecture, inter-core communication requires multi-hop communication over wireline paths resulting in lower bandwidth. On the other hand, in the WiMesh architectures, the wireless links help in reducing the average hop-count of the network. Due to the reduction in hop-count in the WiMesh and the transfer of data over the long distance direct wireless links, the bandwidth increase significantly compared to the wireline counterpart, Mesh. However, due to the spatial and temporal variation in the data rate of the WIs, the demanded wireless bandwidth for each WI will vary. Hence, uniformly distributing the time slots to all the WIs can result in underutilized time slots at certain WIs with no or less packets to transmit. Moreover, lower number of time slots compared to the demand at WIs which are utilized more, negatively impact the bandwidth. By periodically distributing the total time slots of a token period between the WIs based on the demand can further improve the bandwidth. Consequently, at the saturating point of the WiMesh, the bandwidth improves by 5.12% and 7.69% with P-TDMA MAC and RACM MAC compared to the baseline token based MAC in the same WiMesh architecture. The bandwidth can be improved further by dynamically allocating the time slots to each WI based on the predictive bandwidth demands. As the allocation of time slots in the WiMesh architecture with D-TDMA MAC is based on the dynamically varying demand of the WIs, the bandwidth improves significantly (i.e. 10.25%) compared to the baseline token based MAC at the saturating point of the WiMesh. Hence, the proposed D-TDMA MAC has higher bandwidth compared to the RACM and the P-TDMA MAC mechanism.

The packet energy of the Mesh and WiMesh architecture with different MAC mechanisms for uniform random traffic pattern at network saturation is shown in Figure 5. Due to the multi-hop inter-core communication, the packet energy for the wired Mesh architecture is higher than the WiMesh architecture with baseline token based MAC mechanism that uses single hop wireless links. However, due to uniform distribution of time slots among the WIs, flits in the WIs with higher bandwidth demand have to wait longer to get transmitted over the wireless channel. By allowing WIs with higher bandwidth demand to transmit more number of flits, this
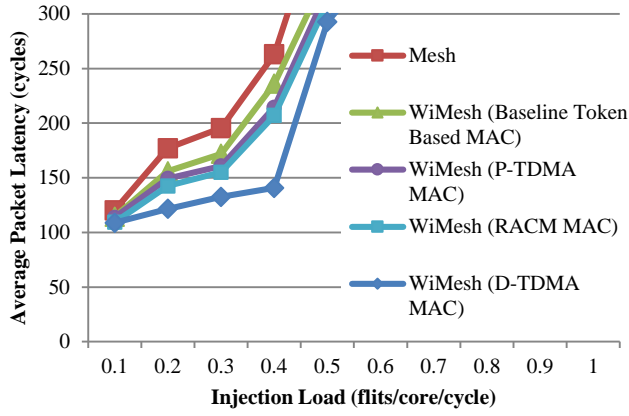
**Figure 6. Average Packet latency for Mesh and WiMesh with different MACs with uniform random synthetic traffic pattern.**

waiting time can be reduced. This results in more data packets being routed over the energy-efficient wireless interconnects ultimately resulting in lower packet energy as can be seen for P-TDMA, RACM and D-TDMA MAC mechanisms. Among these three MAC mechanisms, D-TDMA has the lowest packet energy dissipation. This is because in D-TDMA MAC mechanism, the bandwidth allocation is done dynamically by considering the future predictions of the demands. The impact of implementing demand-aware predictive dynamic MAC is more evident in Figure 6 where the latency characteristics is shown for the architectures considered in this paper. From the figure, it can be seen that the WiMesh with D-TDMA MAC mechanism yields lowest average packet latency among all WiMesh architectures with different MAC mechanism. As the flit injection rates increase, the spatial variation in bandwidth demands among WIs also escalate. In such cases, employing demand-aware bandwidth allocation can result in significant improvement in packet latency due to less waiting time for flits in WIs with higher demands. D-TDMA results in 40.3% lower average packet latency compared to the baseline token based MAC at the saturation point.

The power, area and delay characteristics of the proposed MAC units are listed in Table 1. The total area required for the P-TDMA and D-TDMA MAC mechanism is 0.06% and 0.03% of the wireless transceiver area which itself is 0.9% of the overall chip area for the WiMesh architecture with 12 WIs. The power consumed by the P-TDMA and D-TDMA MAC is 0.483% and 0.322% of the wireless transceiver power. This power overhead is considered to evaluate the packet energy consumption. The delay of both proposed MAC mechanisms are less than a clock cycle of 400ps (2.5GHz clock frequency). However, as these MAC units operate in parallel with data transmission between the WIs, as shown in Figure 2, they have no effect on the data transfer rate.

### 4.2.2 Performance evaluation for non-uniform synthetic Traffic Pattern

In this section, we evaluate the performance of different MAC mechanisms for two non-uniform synthetic traffic patterns, HotSpot and Transpose. Due to more variable distribution of

**Table 1. Characteristics of P-TDMA and D-TDMA MAC**

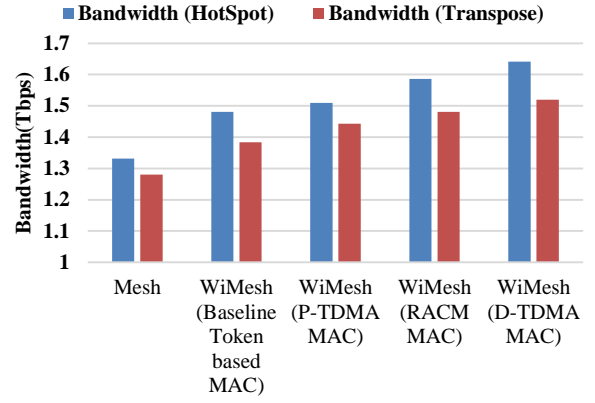| Property | P-TDMA MAC | D-TDMA MAC |
|----------|-----------|-----------|
| Power | 177.59 µW | 118.5 µW |
| Area | 208.52 µm$^2$ | 99.32 µm$^2$ |
| Delay | 370 ps | 250 ps |



**Figure 7. Bandwidth for Mesh and WiMesh with different MACs with non-uniform synthetic traffic pattern.**

bandwidth demands, these non-uniform traffic patterns can be good test cases to demonstrate the advantage of the dynamic MAC mechanisms compared to the uniform traffic pattern. In the HotSpot traffic pattern, a certain volume of traffic generated from all cores is destined towards a hotspot core. All other packets are destined to other cores following a uniform random distribution. This type of traffic pattern is fairly common for directory-based cache-coherent shared memory multiprocessor system where communication among the on-chip core and memory subsystem is more frequent [24]. In our experiment, 5% of the total traffic is destined to the hotspot core which is chosen randomly. In Transpose traffic pattern, each core generates packet only destined to cores that is diametrically opposite to it. For example, the $i^{th}$ core will only send data packets to the $(n$-$i$+1$)th$ core, where, $n$ is the total number of cores.

Figure 7 shows the bandwidth for Mesh and WiMesh architecture with different MAC mechanisms for HotSpot and Transpose traffic. It can be seen that for both non-uniform synthetic traffics, due to multi-hop nature, the bandwidth for wireline mesh is lower compared to the WiMesh with baseline token based MAC mechanism. However, due to the nature of these non-uniform synthetic traffics, few WIs end up handling significant volume of traffic compared to other WIs. Hence, providing more number of time slots to such WIs to transmit more packets further improves the bandwidth because of the reduced waiting time for accessing the wireless medium as can be seen for P-TDMA, RACM and D-
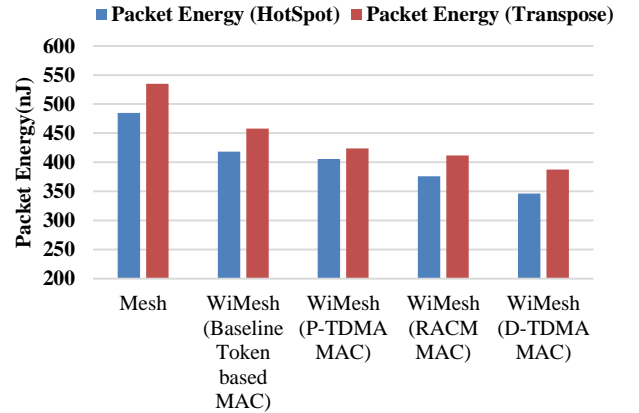


**Figure 8. Packet energy for Mesh and WiMesh with different MACs with non-uniform synthetic traffic pattern.**
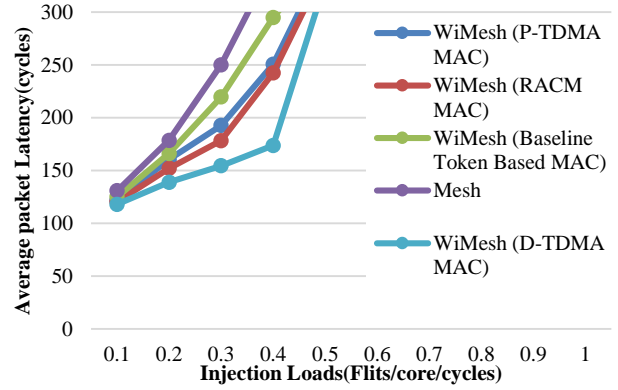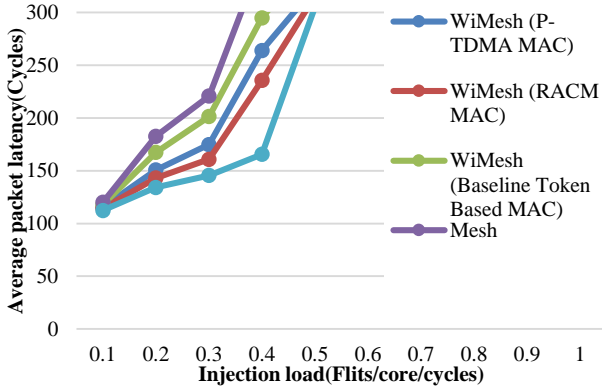
**Figure 9. Average packet latency for Mesh and WiMesh with different MACs with (a) HotSpot (b) Transpose traffic pattern.**

TDMA MAC mechanisms. Among these wireless architectures, WiMesh with D-TDMA MAC mechanism yields the highest bandwidth as the time slots are allocated dynamically according to the future bandwidth demands of the WIs. For the same reasons, D-TDMA shows 17.19% and 15.41% lower packet energy compared to baseline token based MAC mechanism for HotSpot and Transpose traffics respectively as can be seen from Figure 8.

Figure 9(a) and Figure 9(b) show the latency characteristics for HotSpot and Transpose traffic pattern for wireline Mesh and the WiMesh architecture with different MAC mechanisms considered in this paper. Similar to uniform traffic pattern, for both non-uniform synthetic traffic patterns, WiMesh with D-TDMA MAC mechanism has lowest average packet latency compared to all the architectures considered here. As explained in earlier sections, in D-TDMA MAC mechanism, time slots are allocated dynamically according to the predictive bandwidth demands of the WIs reducing the waiting time for the flits at the switch buffers. For this reason, D-TDMA shows 24.87% and 27.59% reduction in average packet latency at network saturation compared to baseline token based MAC mechanism for HotSpot and Transpose traffics respectively.

### 4.2.3 Performance Comparison for Application Specific Traffic Pattern

Based on our discussion in the previous subsection we find that the predictive D-TDMA MAC outperforms all the other MACs studied here. Hence, we compare the bandwidth and energy dissipation of the D-TDMA MAC mechanism and the baseline token passing
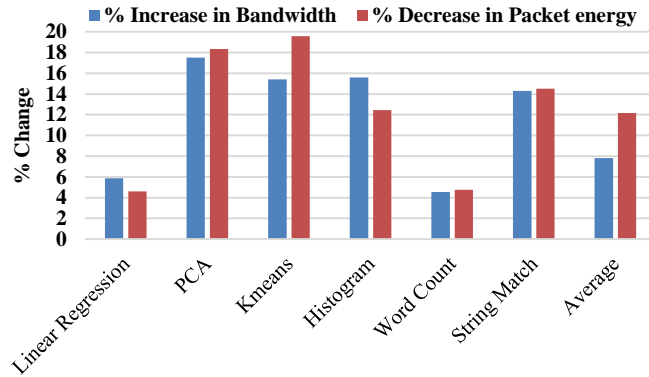
MAC in presence of real application specific traffic patterns in this section. The percentage change in bandwidth and packet energy for the WiMesh with these MACs are shown in Figure 10 for different application specific traffic patterns. It is difficult to find a single benchmark suite that captures the potential variations in traffic patterns in future multicore environments so in this paper application-specific traffic patterns are obtained from several MapReduce [21] benchmarks. We use GEM5 [4] to obtain detailed processor and network-level information. For full system simulations we consider a system of 64 alpha cores running linux within the GEM5 platform. The memory system is MOESI_CMP_directory, setup with private 64KB L1 instruction and data caches and a shared 64MB (1MB distributed per core) L2 cache. The original trace of traffic interaction between the cores, obtained from GEM5 is used to generate the benchmark traffic patterns in the NoC simulator.

The application specific traffic patterns have different spatial and temporal variation and put variable bandwidth demands on the various WIs. Hence, the bandwidth and energy efficiency of the WiMesh architecture with D-TDMA MAC compared to that with the baseline MAC varies between traffic patterns. For traffic patterns where the amount of inter-core communication is low (e.g. Linear Regression and Word Count) due to the pattern of communication, the increase in bandwidth and the decrease in packet energy for the WiMesh architecture with D-TDMA MAC mechanism is relatively lower compared to traffic patterns where the inter-core communication is high (e.g. Histogram, PCA, Kmeans). However in all the cases, the WiMesh with D-TDMA MAC mechanism has 7.81% higher bandwidth and 12.16% lower packet energy on average, compared to the WiMesh with the baseline token based MAC.

## 5. CONCLUSION AND FUTURE WORK

Wireless interconnections are one of the emerging interconnect paradigms that can emerge as a solution to the scalability and energy efficiency problems of the large NoCs. However due to task mapping, task migration, varying workloads and integration of heterogeneous components, the bandwidth demand on the WIs vary spatially and temporally. In this work, we propose two dynamic MAC mechanisms that are able to dynamically allocate time slots to each token based WI based on a predicted estimate of the demand. We show that the proposed dynamic TDMA based MAC mechanism with predictive time slot allocation outperforms the baseline token based MAC mechanism for both synthetic and application specific traffic patterns in a WiNoC.



**Figure 10. Percentage change in bandwidth and packet Energy for application-specific traffic.**

Since, dynamic resource allocation in the form of channel bandwidth allocation is an important technique that can be useful to maximize the benefits of such shared medium communication fabrics we intend to investigate its applicability in other WiNoC architectures using other types of MACs besides the token passing based access mechanism. In addition we want to explore their advantages and applicability to WiNoCs designed with different wireless interconnect technologies like CNT based antennas as well.

# 6. REFERENCES

[1] Abadal, S., Nemirovsky, M., Alarcon, E., and Cabellos-Aparicio, A., *"Networking Challenges and Prospective Impact of Broadcast-Oriented Wireless Network-on-Chip"*. In Proc. of the *ACM/IEEE International Symposium on Networks-on-Chip. NOCS '15.* ACM, Vancouver, CA.

[2] Áström, K.J., Hägglund, T., *"PID controllers: Theory, design, and tuning"*. Instrument Society of America, Research Triangle Park, NC. 1995 Jan;10.

[3] Benini, L. and Micheli, G. De, *"Networks on chips: a new SoC paradigm," in* Computer*, vol. 35, no. 1, pp. 70-78, Jan 2002.*

[4] Binkert, N. et. al. 2011. *"The gem5 simulator"*. ACM SIGARCH Comput. Archit. News, 39 (2), 1-7, 2011.

[5] Chang, K., et. al. *"Performance evaluation and design trade-offs for wireless network-on-chip architectures"*. ACM Journal of Emerg. Tech. in Comp. System, 8 (3), Article 23, 2012.

[6] Chang, M. *et al.*, "CMP network-on-chip overlaid with multi-band RF-interconnect," *High Performance Computer Architecture, 2008. HPCA 2008. IEEE 14th International Symp. on*, Salt Lake City, UT, 2008.

[7] *Chip MultiProjects. Retrieved November, 2015, from: http://cmp.imag.fr*

[8] Chung, E. S., Milder, P. A., Hoe, J. C., and Mai, K., *"Single-Chip Heterogeneous Computing: Does the Future Include Custom Logic, FPGAs, and GPGPUs?,"* Microarchitecture (MICRO), 2010 43rd Annual IEEE/ACM International Symposium on*, Atlanta, GA, 2010, pp. 225-236.*

[9] Deb, S. et al.*, "Design of an Energy-Efficient CMOS-Compatible NoC Architecture with Millimeter-Wave Wireless Interconnects," in* IEEE Transactions on Computers*, vol. 62, no. 12, pp. 2382-2396, Dec. 2013.*

[10] Deb, S., Ganguly, A., Pande, P.P., Belzer, B., and Heo, D., "*Wireless NoC as Interconnection Backbone for Multicore Chips: Promises and Challenges*," in IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 2, no. 2, pp. 228-239, June 2012.

[11] DiTomaso, D., Kodi, A., Kaya, S., and Matolak, D., *"iWISE: Inter-router Wireless Scalable Express Channels for Network-on-Chips (NoCs) Architecture,"* High Performance Interconnects (HOTI), 2011 IEEE 19th Annual Symposium on*, Santa Clara, CA, 2011, pp. 11-18.*

[12] Duato, J., Yalamanchili, S., and Lionel, N., *"Interconnection Networks: An Engineering Approach"*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2002.

[13] Duraisamy, K., Kim, R. G. and Pande,P.P., *"Enhancing performance of wireless NoCs with distributed MAC protocols,"* Quality Electronic Design (ISQED), 16th International Symp. on*, Santa Clara, CA, 2015.*

[14] Ganguly, A. Chang, K., Deb, S., Pande, P. P., Belzer, B., and Teuscher, C., "*Scalable Hybrid Wireless Network-on-Chip Architectures for Multicore Systems*," in *IEEE Trans. on Comp.*, vol. 60, no. 10, pp. 1485-1502, 2011.

[15] Kumar, A., Peh, L.S., Kundu, P., and Jha, N., *"Express virtual channels: towards the ideal interconnection fabric"*. In Proc. of the International Symp. on Computer architecture. ISCA '07. ACM, 150-161, 2007.

[16] Mansoor, N., and Ganguly, A., *"Reconfigurable Wireless Network-on-Chip with a Dynamic Medium Access Mechanism"*. In Proc of the *International Symp. on Networks-on-Chip*. NOCS '15. ACM, NY, USA.

[17] Mansoor, N., Ganguly, A. and Yuvaraj, M.P., "An energy-efficient and robust millimeter-wave Wireless Network-on-Chip architecture," *Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT), 2013 IEEE International Symposium on*, New York City, NY, 2013, pp. 19-24.

[18] Mishra, A. K. , Narayan, V. and Das, C.R., *"A case for heterogeneous on-chip interconnects for CMPs"*. In Proc. of the International Symposium on Computer architecture. ISCA '11. ACM, NY USA, 389-400.

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

[20] Palesi, M., Collotta, M., Mineo, A., and Catania, V., 2015. *"An Efficient Radio Access Control Mechanism for Wireless Network-On-Chip Architectures"*. JLPEA, 5 (2). 38-56, 2015.

[21] Ranger, C., Raghuraman, R., Penmetsa, A., Bradski, G., and Kozyrakis, C., "*Evaluating MapReduce for Multi-core and Multiprocessor Systems*," High Performance Computer Architecture, 2007. HPCA 2007. IEEE 13th International Symposium on, Scottsdale, AZ, 2007, pp. 13-24.

[22] Shacham, A., Bergman, K. and Carloni, L. P., "*Photonic Networks-on-Chip for Future Generations of Chip Multiprocessors*," in IEEE Transactions on Computers, vol. 57, no. 9, pp. 1246-1260, Sept. 2008.

[23] Shamim, M. S., et al. "Energy-efficient wireless network-on-chip architecture with log-periodic on-chip antennas." *Proceedings of the 24th edition of the great lakes symposium on VLSI*. ACM, 2014.

[24] Soteriou, V., Wang, H., and Peh, L., "A Statistical Traffic Model for On-Chip Interconnection Networks," *Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, 2006. MASCOTS 2006. 14th IEEE International Symposium on*, 2006, pp. 104-116.

[25] Vijayakumaran, V. et. al. "*CDMA Enabled Wireless Network-on-Chip*". ACM Journal on Emerging Tech and Comp Sys., 10(4). Article 28, 2014.

[26] Zhao, D., and Wang, Y., "*SD-MAC: Design and Synthesis of a Hardware-Efficient Collision-Free QoS-Aware MAC Protocol for Wireless Network-on-Chip,*" in IEEE Trans. on Comp., vol. 57, no. 9, pp. 1230-1245, 2008.