

# Design Space Exploration for Wireless NoCs Incorporating Irregular Network Routing

Paul Wettin, *Member, IEEE*, Ryan Kim, *Student Member, IEEE*, Jacob Murray, *Member, IEEE*,  
Xinmin Yu, *Student Member, IEEE*, Partha P. Pande, *Senior Member, IEEE*,  
Amlan Ganguly, *Member, IEEE*, and Deukhyoun Heo, *Senior Member, IEEE*

**Abstract**—The millimeter-wave small-world wireless network-on-chip (mSWNoC) is an enabling interconnect architecture to design high-performance and low-power multicore chips. As the mSWNoC has an overall irregular topology, it is essential to design and optimize suitable deadlock-free routing mechanisms for it. In this paper, we quantify the latency, energy dissipation, and thermal profiles of mSWNoC architectures by incorporating irregular network routing strategies. We demonstrate that the latency, energy dissipation, and thermal profile are affected by the adopted routing methodologies. The overall system performance and thermal profile are governed by the traffic-dependent optimization of the routing methods. Our aim is to establish the energy-thermal-performance trade-offs for the mSWNoC depending on the exact routing strategy and the characteristics of the benchmarks considered.

**Index Terms**—Irregular networks, millimeter-wave wireless, network-on-chip (NoC), routing algorithms, small-world.

## I. INTRODUCTION

WIRELESS network-on-chip (WiNoC) is envisioned as an enabling technology to design low-power and high-bandwidth, massive multicore architectures [1]. The existing method of implementing a NoC with planar metal interconnects is deficient due to high latency, significant power consumption, and temperature hotspots arising out of long, multihop wireline paths used in data exchange. It is possible to design high-performance, robust, and energy-efficient multicore chips by

adopting novel architectures inspired by complex network theory in conjunction with on-chip wireless links. Networks with the small-world property have very short average path lengths, making them particularly interesting for efficient communication with minimal resources. Using the small-world approach we can build a highly efficient NoC with both wired and wireless links. Neighboring cores should be connected through traditional metal wires while widely separated cores will communicate through long-range, single-hop, wireless links. A small-world network principally has an irregular topology [2]. Routing in irregular networks is more complex, because routing methods are typically topology agnostic. Hence, it is necessary to investigate suitable routing mechanisms for small-world networks. Routing in irregular networks can be classified into two broad categories, viz., rule- and path-driven strategies [3]. Rule-driven routing is typically done by employing a spanning tree for the network. Messages are routed along this spanning tree with specific restrictions to achieve deadlock freedom. Because deadlock freedom is taken into account first for these routing strategies, minimal paths through the network for every source-destination pair cannot be guaranteed [3]. Conversely, for path-driven routing, minimal paths between all source-destination pairs are first guaranteed and then deadlock freedom is achieved by restricting portions of traffic from using specific resources such as the virtual channels [3].

We follow the above-mentioned strategies to design suitable routing mechanisms for the millimeter (mm)-wave small-world wireless NoC (mSWNoC) [4]. In the rule-based routing, a spanning tree of the network is created where data is routed along the spanning tree. An allowed route never uses a link in the up direction along the tree after it has been in the down direction once. Hence, channel dependency cycles are prohibited, and deadlock freedom is achieved [5]. However, a well-known weakness of this routing scheme is that it has a strong tendency to generate hotspots around the root of the tree structure. In the path-based routing, the network resources are divided into layers and network deadlocks are avoided by preventing portions of traffic from using specific layers [3], [6]. The achievable performance of mSWNoC depends on the efficiency of these routing algorithms. The power and thermal profiles of the system depend on how efficiently the routing mechanisms can move the traffic through the network while balancing the traffic among the network elements.

These routing algorithms have previously been studied for traditional parallel computing systems where the comparative

Manuscript received December 11, 2013; revised May 27, 2014; accepted August 1, 2014. Date of current version October 16, 2014. This work was supported in part by the U.S. National Science Foundation (NSF) under Grant CCF-0845504, Grant CNS-1059289, and Grant CCF-1162123, and in part by the Army Research Office under Grant W911NF-12-1-0373. This paper was recommended by Associate Editor L. P. Carloni.

P. Wettin is a Senior ASIC Design Engineer with Marvell Semiconductor, Boise, ID, USA. He did this work as a PhD student with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164 USA (email: pwettin@eecs.wsu.edu).

J. Murray is a Clinical Assistant Professor with the Department of Electrical Engineering and Computer Science, Washington State University, Everett, WA, USA. He did this work as a PhD student with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164 USA (email: jmurray@eecs.wsu.edu).

R. Kim, X. Yu, P. P. Pande, and D. Heo are with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164 USA (email: rkim@eecs.wsu.edu; xyu@eecs.wsu.edu; pande@eecs.wsu.edu; dheo@eecs.wsu.edu).

A. Ganguly is with the Department of Computer Engineering, Rochester Institute of Technology, Rochester, NY 14623 USA (e-mail: amlan.ganguly@rit.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCAD.2014.2351577

performance evaluation between multiple routing algorithms is predominantly done in terms of achievable saturation throughput [3], [6]. Conversely, NoCs in presence of general-purpose chip multiprocessor (CMP) benchmarks, such as SPLASH-2 [7] and PARSEC [8], operate below network saturation and their traffic density varies significantly from one benchmark to the other [9]. Hence, instead of only looking at saturation throughput, we should quantify the associated network latency, energy dissipation, and thermal profiles in presence of frequently used, nonsynthetic benchmarks.

In this paper, we first explore how the topology-agnostic irregular routing strategies should be optimized depending on the traffic patterns in an mSWNoC. We demonstrate that performance of both the rule- and path-based routings can be enhanced by taking into account the traffic characteristics compared to the traditional implementation of these routing methods. We then undertake a detailed performance evaluation for the mSWNoC architecture by incorporating these optimized irregular network routing strategies. We consider network latency, energy dissipation, and the thermal profile as the relevant metrics in this performance evaluation. We demonstrate that depending on the specific benchmarks, the energy-temperature-performance (ETP) trade-offs vary and the distribution of the on-chip traffic pattern has an important role to play. We also quantify the performance of the mSWNoC using artificially inflated traffic patterns to stress the network. We demonstrate that when the traffic injection load is increased in the mSWNoC beyond saturation then the layered routing sustains better performance compared to the tree-based counterpart.

## II. RELATED WORK

The limitations and design challenges associated with existing NoC architectures are elaborated in [9]. Conventional NoCs use multihop, packet switched communication. At each hop the data goes through a complex router/switch, which contributes to considerable power, throughput, and latency overheads. To improve performance, a methodology to automatically synthesize architectures with a few application specific long-range links inserted in a regular mesh was proposed in [10]. Subsequently, performance advantages of NoCs by insertion of long-range wireline links following principles of small-world graphs were elaborated in [11]. The concept of express virtual channels is introduced in [12]. Despite significant performance gains, in the above schemes the long-range links are designed with conventional wires. It is already shown that beyond a certain length, wireless links are more energy efficient than conventional metal wires. Hence, the performance improvements by using long-range wireless links will be more than that using wireline links [1].

A comprehensive survey regarding various WiNoC architectures and their design principles are presented in [13]. WiNoC architectures can be divided into two sub categories, viz. mesh with overlaid wireless links and hierarchical architectures with long-range wireless shortcuts. Among the first category, notable examples include design of a WiNoC based on CMOS ultra wideband (UWB) [14], 2-D concentrated mesh-based WCube architecture using sub-THz wireless links [15], and the interroutier wireless scalable express channel for NoC (iWISE)

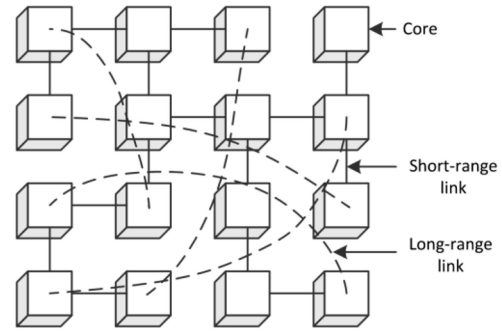


Fig. 1. SWNoC architecture with short- and long-range links.

architecture [16]. Possibilities of creating novel architectures aided by the on-chip wireless communication have been explored in [1] and [17]. These two works proposed design of hierarchical and hybrid WiNoC architectures using long-range wireless shortcuts. The whole system is partitioned into multiple small clusters of neighboring switches called subnets. In the upper level of the network, the subnets are connected via wireline and wireless links. In contrary to the mSWNoC architecture proposed in this paper, the architectures of [1] and [17] do not have perfect power-law based small-world topologies. The upper level of the hierarchy has either a mesh or ring topology with wireless shortcuts. The principal drawback of these architectures is that in presence of wireless link failures the upper level will be dominated by the multihop wireline ring or mesh and hence performance degradation is significant [18]. It is already shown that a WiNoC, where the network architecture is designed following the power-law based small-world connectivity [2], is more robust in presence of wireless link failures compared to the hierarchical counterpart [18], [19].

In this paper, our aim is to quantify the performance of the mSWNoC architecture by incorporating and optimizing different irregular network routing algorithms in terms of latency, energy dissipation, and thermal profiles.

## III. WIRELESS NoC ARCHITECTURE

Many naturally occurring complex networks, such as social networks, the internet, the brain, as well as microbial colonies exhibit the small-world property [2], [20]. Small-world graphs are characterized by many short-distance links between neighboring nodes as well as a few relatively long-distance, direct shortcuts. Small-world graphs are particularly attractive for constructing scalable WiNoCs because the long-distance shortcuts can be realized using high-bandwidth, low-energy, wireless interconnects while the local links can be designed with traditional metal wires. In this paper, we consider a small-world NoC (SWNoC) architecture, where the long-range shortcuts are implemented through mm-wave wireless links operating in 10–100 GHz range. Fig. 1 represents such a SWNoC with 16 cores, where each core is associated with a NoC switch (not shown for clarity). The SWNoC has many short-range local links, as well as, a few long-range shortcuts schematically represented by the arching, dashed, interconnects. In the following section we discuss the characteristics of the mSWNoC architecture and analyze its performance and

temperature profiles using various irregular network routing strategies with respect to a conventional wireline mesh.

### A. mSWNoC Topology

In the mSWNoC topology, cores are arranged in equal-area tiles over a  $20 \times 20$  mm die. Each core is connected to a switch and the switches are interconnected using both wireline and wireless links. The topology of the mSWNoC is a small-world network where the links between switches are established following a power-law model [2], [21]. In this small-world network there are still several long wireline interconnects. As these are extremely costly in terms of power and delay, we use mm-wave wireless links to connect switches that are separated by a long distance. In [1], it is demonstrated that it is possible to create three nonoverlapping channels with on-chip mm-wave wireless links. Using these three channels we overlay the wireline small-world connectivity with the wireless links such that a few switches get an additional wireless port. Each of these wireless ports will have wireless interfaces (WIs) tuned to one of the three different frequency channels. Each WI in the network is then assigned one of the three channels; more frequently communicating WIs are assigned the same channel to optimize the overall hop-count. One WI is replaced by a gateway WI that has all three channels assigned to it; this facilitates data exchange between the nonoverlapping wireless channels.

To have a detailed comparative performance evaluation of the mSWNoC we also consider a wireline-only SWNoC topology. The SWNoC topology is designed identically as the mSWNoC topology. However, the SWNoC has no wireless links. The long-range shortcuts are principally implemented through multihop metal wires.

We have assumed an average number of connections from each switch to the other switches,  $\langle k \rangle$ . The value of  $\langle k \rangle$  is chosen to be four so that the mSWNoC does not introduce any additional switch overhead with respect to a conventional mesh. Also an upper bound  $k_{\max}$ , is imposed on the number of wireline links attached to a particular switch so that no switch becomes unrealistically large in the mSWNoC. This also reduces the skew in the distribution of links among the switches. Both  $\langle k \rangle$  and  $k_{\max}$  do not include the local NoC switch port to the core.

### B. Communication and Channelization

This section describes the WI components and overall communication mechanism, which includes flow control and routing strategies for the mSWNoC.

1) *Wireless Interface (WI)*: The two principal WI components are the antenna and the transceiver, whose characteristics are outlined below.

The on-chip antenna for the mSWNoC has to provide the best power gain for the smallest area overhead. A metal zigzag antenna has been demonstrated to possess these characteristics [22]. This antenna also has negligible effect of rotation (relative angle between transmitting and receiving antennas) on received signal strength, making it most suitable for mm-wave NoC applications. Zigzag antenna characteristics depend on physical parameters like axial length, trace width, arm length,

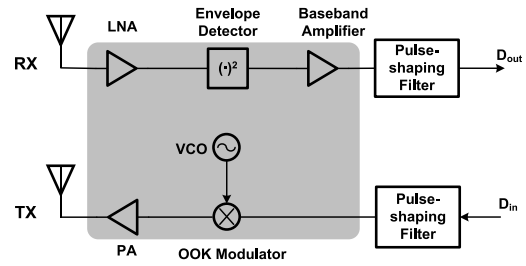


Fig. 2. Block diagram of the noncoherent OOK transceiver for mSWNoC.

bend angle, etc. By varying these parameters, the antennas are designed to operate on different frequency channels [1]. Three different channels were obtained with 3 dB bandwidths of 16 GHz and center frequencies of 31, 57.5, and 120 GHz respectively with a communication range of 20 mm. The zigzag antenna is designed with  $10 \mu\text{m}$  trace width,  $60 \mu\text{m}$  arm length, and  $30^\circ$  bend angle. For optimum power efficiency, the quarter wave antennas use axial lengths of 0.73, 0.38, and 0.18 mm, respectively. The antenna design ensures that signals outside the communication bandwidth, for each channel, are sufficiently attenuated to avoid interchannel interference.

The design of a low-power wideband wireless transceiver is the key to guarantee the desired performance of the mSWNoC. Therefore, at both the architecture and circuit levels of the transceiver, low-power design considerations need to be taken into account. At the architecture level, on-off-keying (OOK) modulation was chosen to simplify the circuit design. Noncoherent demodulation is used, therefore eliminating the power-hungry phase-lock loop (PLL) in the transceiver. Moreover, at the circuit level, body-enabled design techniques [23], including both forward body-bias (FBB) with dc voltages, as well as body-driven by ac signals, were implemented in several sub-blocks to further decrease their power consumption.

The transceiver architecture is shown in Fig. 2. The receiver (RX) includes a wideband low-noise amplifier (LNA), an envelope detector for noncoherent demodulation, and a baseband amplifier. A voltage-controlled oscillator (VCO) is not needed in the RX because noncoherent demodulation is used, which results in a power reduction by more than 30% compared to the mSWNoC transceiver of [1]. The transmitter (TX) has a simple direct up-conversion topology, consisting of a body-driven OOK modulator, a wideband power amplifier (PA), and a VCO. Using this transceiver architecture, system-level simulations were initially carried out to define the sub-blocks' design specifications [24], followed by sub-block circuit simulations and transceiver chip design.

2) *WI Flow Control*: In the mSWNoC, data is transferred via a flit-based, wormhole routing [25]. Between a source-destination pair, the wireless links, through the WIs, are only chosen if the wireless path reduces the total path length compared to the wireline path. This can potentially give rise to hotspot situations in the WIs. Many messages will try to access the wireless shortcuts simultaneously, thus overloading the WIs, which would result in higher latency and energy dissipation. Token flow control [26] is used to alleviate overloading at the WIs. Tokens are used to communicate the status



of the input buffers of a particular WI to the wireline switches, which need to use the WI for accessing the wireless shortcuts.

An arbitration mechanism is designed to grant access to the wireless medium to a particular WI, including the gateway WI, at a given instant to avoid interference and contention between the WIs that have the same frequency. To avoid the need for centralized control and synchronization, the arbitration policy adopted is a wireless token passing protocol [1]. It should be noted that the use of the word token in this case differs from the usage in the above mentioned token flow control. The wireless token passing protocol here is a simple media access control (MAC) mechanism to access the wireless channels. According to this scheme, the particular WI possessing the token can broadcast flits into the wireless medium in its respective frequency. A single flit circulates as a token in each frequency channel. All other WIs of the same channel will receive the flits, but only the WI whose address matches the destination address will accept the flit for further processing. The wireless token is released and forwarded to the next WI operating in the same frequency channel after all flits belonging to a message at a particular WI are transmitted. Packets are rerouted, through an alternate wireline path, if the WI buffers are full or if it does not have the token. As rerouting packets can potentially lead to deadlock, a rerouting strategy similar to dynamic quick reconfiguration (DQR), as presented in [27], is used to ensure deadlock freedom. In this situation, the current WI becomes the new source for the packet, which is then forced to take a wireline only path to the final destination, still following the original routing strategy restrictions, explained in the next two sub-sections below.

3) *Multiple Tree Roots (MROOTS) Routing*: The first routing strategy for mSWNoC that we consider is an up/down tree-based routing algorithm, belonging to the rule-based classification. This routing strategy utilizes a MROOTS-based mechanism [3], [5], [6]. MROOTS allows multiple routing trees to exist, where each tree routes on a dedicated virtual channel. Hence, traffic bottlenecks can be reduced in the upper tree levels that are inherent in this type of routing. In this paper, we consider three tree root selection policies, viz. a random root placement (random), a maximized intraroot tree distance placement (max distance), and a traffic-weighted minimized hop-count placement ( $f_{ij}$ ). The random root placement chooses the roots at random locations. The maximized intraroot tree distance placement attempts to find roots that are far apart in the tree [6], in order to minimize the congestion near the selected roots. Finally, the traffic-weighted minimized hop-count placement is described as follows. Selecting  $M$  tree roots will create  $M$  trees in the network, where the chosen  $M$  roots minimize the optimization metric  $\mu$  as defined in

$$\mu = \min_{\forall \text{ roots}} \sum_{\forall i} \sum_{\forall j} h_{ij} f_{ij}. \quad (1)$$

Here, the minimum path distance in hops,  $h_{ij}$ , from switch  $i$  to switch  $j$  is determined following the up/down routing restrictions. The frequency of traffic interaction between the switches is denoted by  $f_{ij}$ . As root selection only affects valid routing paths for deadlock freedom and does not alter the physical placement of links, any *a priori* knowledge of the frequency

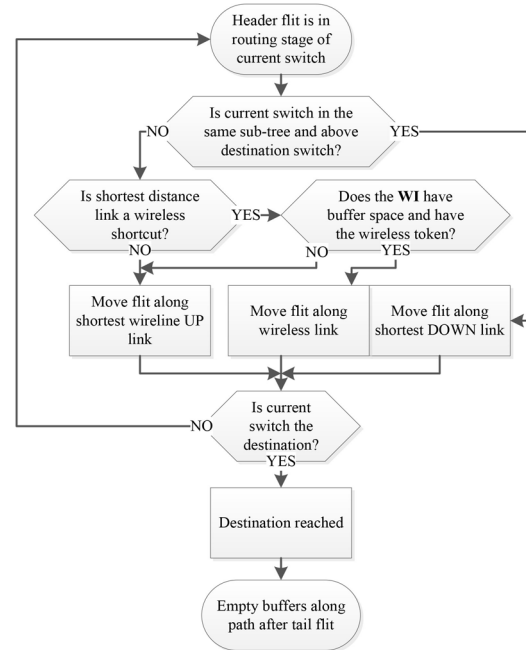


Fig. 3. MROOTS routing flow chart for mSWNoC.

of traffic interaction aids in root selection. Incorporating  $f_{ij}$  helps minimize the routed path lengths for specific workloads on the mSWNoC architecture. Breadth-first trees were used during the tree creation process to balance the traffic distribution among the sub-trees, and to minimize bottlenecks in a particular tree. All wireless and wireline links that are not part of the breadth-first tree are reintroduced as shortcuts. An allowed route never uses an up direction along the tree after it has been in the down path once. In addition, a packet traveling in the downward direction is not allowed to take a shortcut, even if that minimizes the distance to the destination. Hence, channel dependency cycles are prohibited, and deadlock freedom is achieved [5]. A flow chart of the adopted MROOTS routing strategy is shown in Fig. 3.

4) *Adaptive Layered Shortest-Path (ALASH) Routing*: The second routing strategy is an ALASH algorithm [6], which belongs to the path-based classification. ALASH is built upon the layered shortest path (LASH) algorithm, but has more flexibility by allowing each message to adaptively switch paths, letting the message choose its own route at every intermediate switch.

The LASH algorithm takes advantage of the multiple virtual channels in each switch port of the NoC switches in order to route messages along the shortest physical paths. In order to achieve deadlock freedom, the network is divided into a set of virtual layers, which are created by dedicating the virtual channels from each switch port into these layers. The shortest physical path between each source-destination pair is then assigned to a layer such that the layer's channel dependency graph remains free from cycles. A channel dependency is created between two links in the source-destination path when a link from switch  $i$  to switch  $j$  and a link from switch  $j$  to switch  $k$  satisfies the following condition,  $pathlength(i) < pathlength(j) < pathlength(k)$ , where  $pathlength(X)$  is the length of the minimal path between switch  $X$  and the original

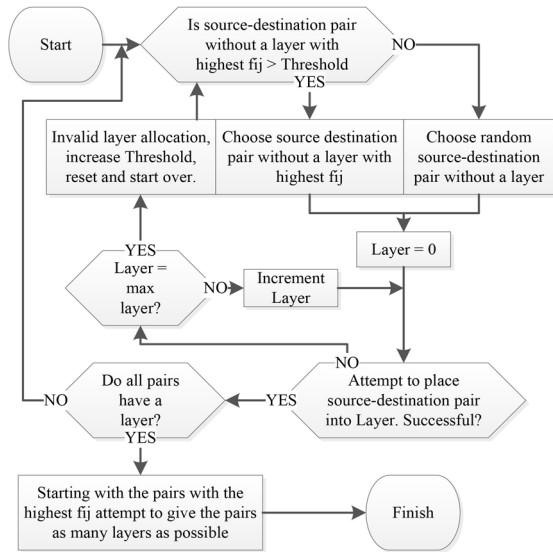


Fig. 4. Priority layering function flow chart used for ALASH to allocate the layers.

source switch. When a layer's channel dependency graph has no cycles, it is free from deadlocks as elaborated in [6].

For ALASH, the decision to switch paths is based on the current network conditions. We use virtual channel availability and current communication density of the network as the two relevant parameters for this purpose. The communication density is defined as the number of flits traversing the given switch or link over a certain time interval. In order to increase the adaptability of the routing, multiple shortest paths between all source-destination pairs are found and then included into as many layers as possible. The message route through the network depends on the layers each source-destination pair use. Therefore, the layering function that controls how the layers are allocated for each source-destination pair has an impact on the latency, energy, and thermal profile of the mSWNoC.

In this paper, we consider three different types of layering functions, a randomized uniform layering function (uniform), a layer balancing function (virtual), and a priority-based layering function (priority). The uniform layering function selects source-destination pairs at random while allocating layers to them, giving each source-destination pair an equal opportunity for each layer unless including the path results in a cyclic dependency. The virtual layering function uses the *a priori* knowledge of the frequency of traffic interactions,  $f_{ij}$ , in order to evenly distribute source-destination pairs with large  $f_{ij}$  values across the different layers. In contrary, the priority layering function allocates as many layers as possible to source-destination pairs with high  $f_{ij}$ . This improves the adaptability of messages with higher  $f_{ij}$  by providing them with greater routing flexibility. As an example, Fig. 4 shows the priority layering function flow chart.

The main goal of these different layering functions is to help distribute the messages across the network in such a way so as not to induce a load imbalance in any layer and hence, any particular network switch. The virtual layering function does this by reducing the opportunity that source-destination pairs with high  $f_{ij}$  values will use the same layer. The priority layering function does the same by allowing the source-destination

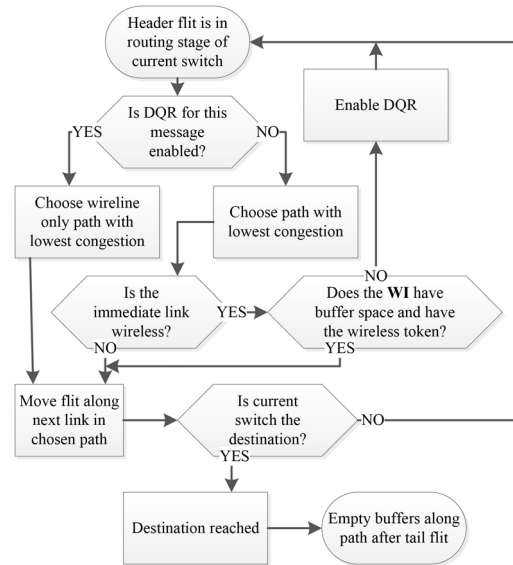


Fig. 5. ALASH routing flow chart for mSWNoC.

pairs with high  $f_{ij}$  values to get more layers to use adaptively. This avoids use of the same layer by another source-destination pair with high  $f_{ij}$  values. For every layer balancing technique, it is possible to induce deadlocks if a message is allowed to switch back and forth between two or more layers. Hence, a message is not allowed to revisit a layer that it has previously traveled in to maintain deadlock freedom. A flow chart of the adopted ALASH routing strategy is shown in Fig. 5.

#### IV. EXPERIMENTAL RESULTS

In this section, we evaluate the performance and temperature profiles of the mSWNoC, SWNoC, and conventional wireline mesh-based NoC architectures. We further evaluate the influence of root selection for the MROOTS routing strategy, as well as evaluate the influence of the layering function for the ALASH routing strategy.

We use GEM5 [28], a full system simulator, to obtain detailed processor- and network-level information. We consider a system of 64 alpha cores running Linux within the GEM5 platform for all experiments. Three SPLASH-2 benchmarks, FFT, RADIX, LU [7], and seven PARSEC benchmarks, CANNEAL, BODYTRACK, VIPS, DEDUP, SWAPTION, FLUIDANIMATE, and FREQMINE [8] are considered. These benchmarks vary in characteristics from computation intensive to communication intensive in nature and thus are of particular interest in this paper. The interswitch traffic patterns in terms of normalized switch interaction rates for the above-mentioned benchmarks are shown in Fig. 6. The benchmarks that are computation intensive, FFT, RADIX, LU, SWAPTION, DEDUP, FLUIDANIMATE, VIPS, and FREQMINE, have low median switch interaction rates. Conversely, the benchmarks that are communication intensive, CANNEAL and BODYTRACK, have higher median switch interaction rates than the others. It can be seen in Fig. 6 that the medians of the interaction rate for these two benchmarks are higher than the medians of the other benchmarks. The median switch interaction rates of the other benchmarks are low, but are not

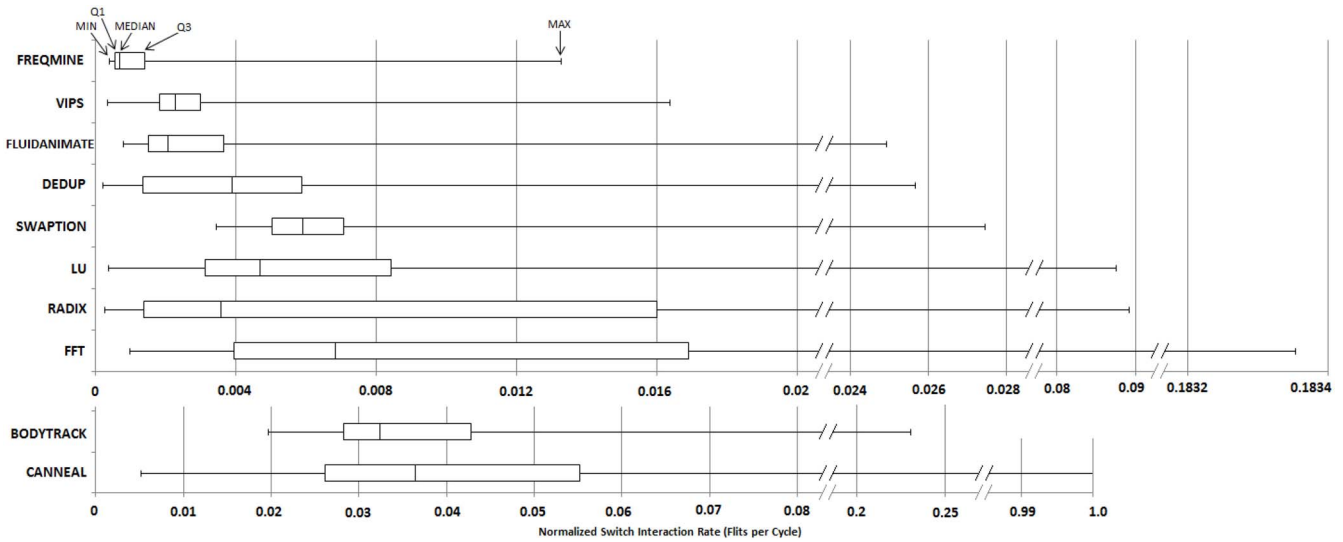


Fig. 6. Normalized switch interaction rate for the SPLASH-2 benchmarks: FFT, RADIX, and LU and the PARSEC benchmarks: FREQMINE, VIPS, FLUIDANIMATE, DEDUP, SWAPTION, BODYTRACK, and CANNEAL.

exactly the same. As an example, FFT has a relatively high median switch interaction rate when compared to the other computation intensive benchmarks, but when compared to the communication intensive benchmarks, it is an order of magnitude lower. The switch interaction rate of these benchmarks plays an important role in the overall latency, energy dissipation, and thermal profiles of the mSWNoC, as explained later.

The width of all wired links is considered to be the same as the flit width, which is 32 bits in this paper. Each packet consists of 64 flits. The NoC simulator uses switches synthesized from an register transfer level (RTL) level design using Taiwan Semiconductor Manufacturing Company (TSMC) 65-nm CMOS process in synopsys design vision. All ports except those associated with the WIs have a buffer depth of two flits and each switch port has four virtual channels. Hence, four trees and four layers are created in MROOTS and ALASH, respectively. The ports associated with the WIs have an increased buffer depth of eight flits to avoid excessive latency penalties while waiting for the token. Increasing the buffer depth beyond this limit does not produce any further performance improvement for this particular packet size, but will give rise to additional area overhead [1]. Energy dissipation of the network switches, inclusive of the routing strategies, were obtained from the synthesized netlist by running synopsys prime power, while the energy dissipated by wireline links was obtained through HSPICE simulations, taking into consideration the length of the wireline links. The processor-level statistics generated by the GEM5 simulations are incorporated into multicore power, area, and timing (McPAT) to determine the processor-level power values [29].

After obtaining the processor and network power values, these elements are arranged on a  $20 \times 20$  mm die. The floorplans, along with the power values, are used in HotSpot [30] to obtain steady state thermal profiles. The processor power and the architecture-dependent network power values in presence of the specific benchmarks are fed to the HotSpot simulator to obtain their temperature profiles.

#### A. Wireless Transceiver Performance

The wireless transceiver circuitry was designed and laid out using TSMC 65-nm standard CMOS process and its characteristics were obtained through post-layout simulation using agilent ADS momentum and cadence specter as well as measurements using a cascade on-wafer probing station. The overall power consumption of the transceiver is 31.1 mW based on post-layout simulation, including 14.2 mW from the RX and 16.9 mW from the TX. With a data rate of 16 Gb/s, the equivalent bit energy is 1.95 pJ/bit. The chip layout photos of the proposed mm-wave transceiver are shown in Fig. 7. The total area overhead per wireless transceiver, which includes all of the components shown in the gray region in Fig. 2, turns out to be 0.17 mm<sup>2</sup>. The physical dimensions of each component is shown in the chip micrograph of Fig. 7. Both the energy and area have been reduced compared to our initial design [1], mainly due to the noncoherent RX topology which eliminates the VCO, which was discussed in Section III-B.

Fig. 8(a) and (b) illustrates the simulated time-domain waveform at the TX output and RX after demodulation. The waveform is an OOK-modulated signal with a 60 GHz carrier frequency and 16 Gb/s baseband data rate. As shown in Fig. 8(a), the peak voltage amplitude reached is 300 mV even at the shortest pulse, which is equivalent to having 0 dBm on a 50-Ohm antenna load. The modulated signal was then fed into the RX after passing through a simulated channel and antenna model. Fig. 8(b) shows the demodulated signal at the RX baseband amplifier output, and it can be seen that the signal matches the transmitted baseband data. Also, shown in Fig. 8(c), the eye in the eye diagram of the demodulated signal is wide open indicating the signal has decent quality after demodulation. With such RX and TX front-ends, the OOK target data rate of 16-Gb/s (or higher) can be achieved.

#### B. Determination of the mSWNoC Topology

We determine the exact topology of the mSWNoC based on the principles of the small-world graph as discussed in



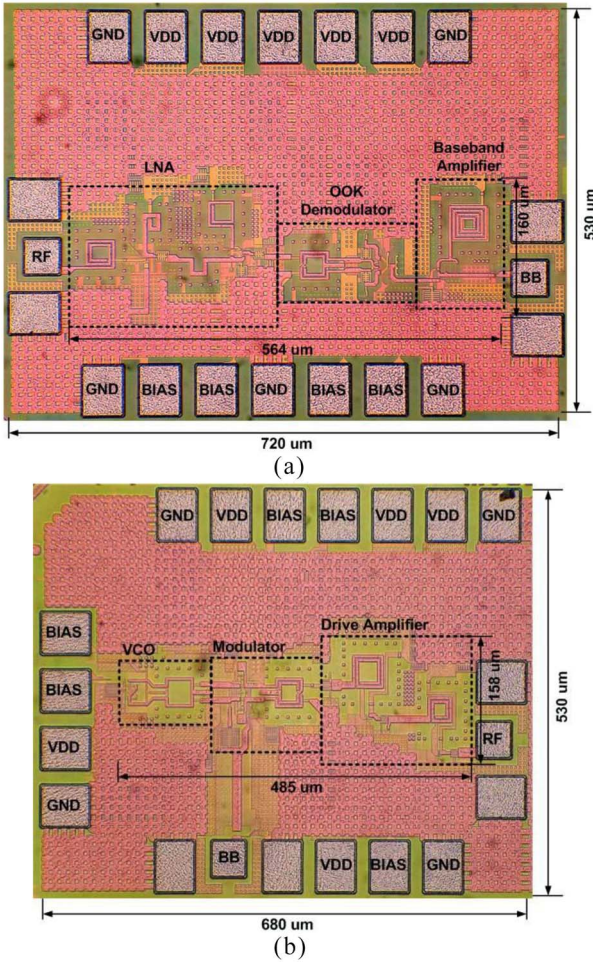


Fig. 7. Chip micrograph of the proposed (a) mm-wave RX and (b) mm-wave TX.

Section III-A. First, our aim is to determine the suitable maximum number of ports,  $k_{\max}$ , for the switches of the mSWNoC. Fig. 9 shows the variation of the normalized throughput and energy with respect to  $k_{\max}$ . From Fig. 9, it can be seen that the optimum number for  $k_{\max}$  is 7 as it optimizes the throughput and energy. Hence, we use this value for  $k_{\max}$  when creating the mSWNoC architecture. We then augment the network by adding WIs. From [1], it is shown that WI placement is most energy efficient when the distance between them is at least 7 mm, for the 65 nm technology node. Next, we determine the optimum number and the placement of WIs. For each possible number of WIs, we considered their suitable placements. It is shown in [17] that for placement of WIs, a simulated annealing (SA) based methodology converges to the optimal configuration much faster than the exhaustive search. Initially, the WIs of all frequency ranges are placed randomly, with each switch having equal probability of getting a WI. To perform SA, an optimization metric,  $\beta$ , is established which is closely related to the connectivity and performance of the network. The metric  $\beta$  is proportional to the average distance, measured in hops, between all source and destination switches. A single hop in this paper is defined as the path length between a source and destination pair that can be traversed in one clock cycle. To compute  $\beta$  the shortest distances between all pairs of switches are computed. The distances are weighted with

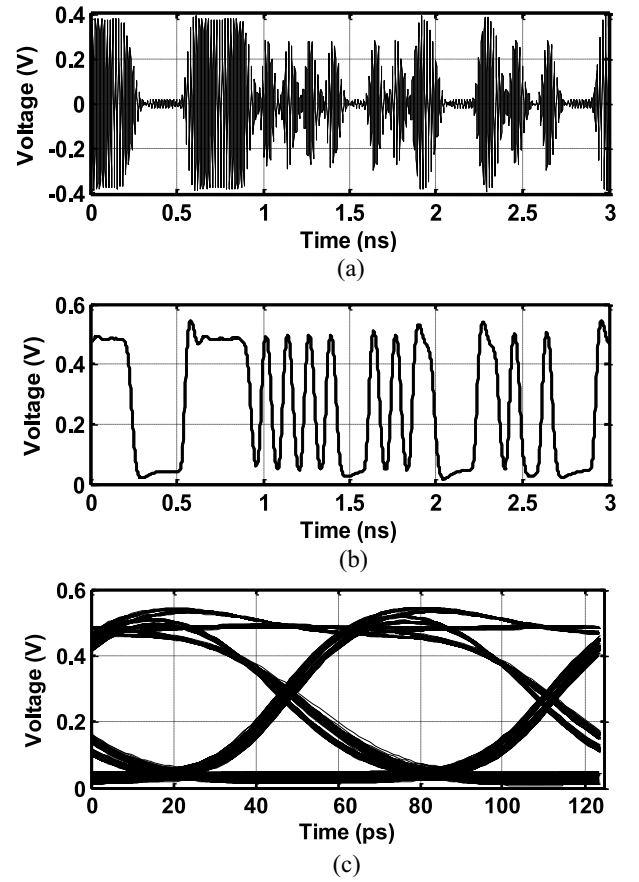


Fig. 8. (a) Simulated time-domain waveform of the transmitted radio frequency (RF) signal at TX output. (b) Demodulated baseband signal at RX output. (c) Eye-diagram of the demodulated signal.

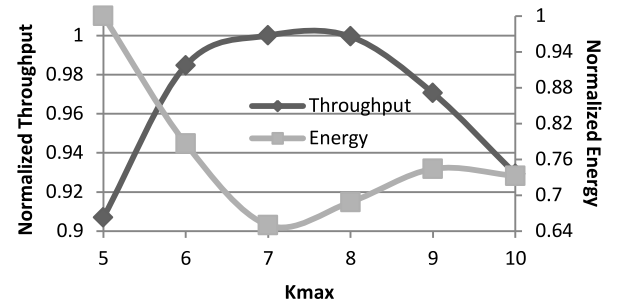


Fig. 9. Normalized throughput and energy dissipation of a 64 core mSWNoC with variation of  $k_{\max}$ .

the normalized frequencies of communication between switch pairs. The optimization metric  $\beta$  can be computed as seen in

$$\beta = \sum_{\forall i} \sum_{\forall j} h_{ij} f_{ij} \quad (2)$$

where  $h_{ij}$  is the distance in hops between the  $i$ th source and  $j$ th destination switches and  $f_{ij}$  is the frequency of communication between the  $i$ th source and  $j$ th destination. This frequency is expressed as the percentage of traffic generated from  $i$  that is destined for  $j$ . Fig. 10 shows the variation of normalized bandwidth and energy dissipation with respect to the number of WIs. It can be seen that with increasing number of WIs bandwidth increases and energy dissipation decreases.

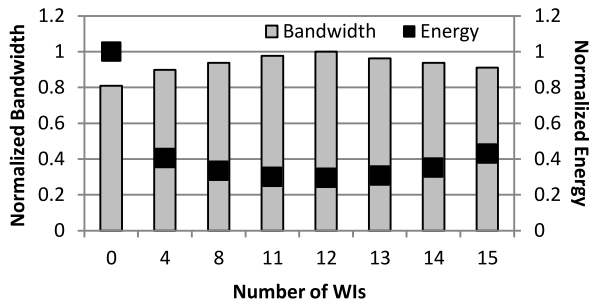


Fig. 10. Normalized bandwidth and energy dissipation tradeoff for the 64-core mSWNoC with variation of the number of WIs.

Increasing the number of WIs improves the connectivity of the network as they establish more one-hop shortcuts. However, the wireless medium is shared among all the WIs and hence, as the number of WIs increases beyond a certain limit performance starts to degrade due to the large token returning period [1]. Moreover, as the number of WIs increases, the overall energy dissipation from the WIs becomes higher, and it causes the packet energy to increase as well. Considering all these factors, we see that the optimum number of WIs for the 64-core system is 12.

### C. Performance Evaluation

In this section, we present the latency, network-level energy dissipation, and thermal characteristics of the SWNoC and mSWNoC by incorporating MROOTS- and ALASH-based routing strategies. We also compare the performance of the small-world architectures with respect to the traditional wireline mesh.

1) *Routing Strategy Optimization*: In this subsection, we evaluate the performance of the root placement optimization and layering function optimization for the MROOTS and ALASH routing strategies, respectively. As discussed in Sections III-B3 and III-B4, the placement of the roots for MROOTS-based routing and the layering function for ALASH-based routing can affect the performance of the mSWNoC architecture. Figs. 11 and 12 show the variation of the energy-delay product with respect to the location of the roots for MROOTS and the layering function for ALASH for the various benchmarks considered in this paper. It can be seen in Fig. 11 that for the MROOTS routing, traffic-weighted minimized hop-count placement ( $f_{ij}$ ) obtains the minimum energy-delay product. This is due to the fact that the highest communicating switches are placed at the roots, effectively allowing for shortest path routing to be employed for these switches. Hence, the  $f_{ij}$  root selection strategy is used for MROOTS for the following performance evaluations in Sections IV-C2 and IV-C3.

It can be seen in Fig. 12 that the priority layering function obtains the minimum energy-delay product and improves the energy-delay product over our initial design (uniform layering) in [4] for the mSWNoC architecture employing ALASH routing. This is due to the fact that the highest communicating source-destination pairs are given the most resources, allowing the adaptability in ALASH to work at its best. Without loss of generality, as an example, Fig. 13 shows the normalized

flits per link distribution in the network for the CANNEAL benchmark. It can be seen from Fig. 13 that for both layering functions, the minimum, the first quartile, and median are the same values; while the values for the third quartile are very similar. This indicates that the traffic distribution for both layering functions is similar. We should then focus on the flits per link for the highly utilized links. Fig. 13 shows that the flit traversal for these links (indicated by the maximum) is lowered for the priority layering function. This shows that the priority layering function routes flits away from heavily utilized links in the network better than the virtual layering function. Hence, the priority layering function is used for ALASH for the following performance evaluations in Sections IV-C2 and IV-C3.

2) *Latency and Energy Characteristics*: Fig. 14 shows the average network packet latency for the various architectures using the two different routing strategies and considering the above-mentioned benchmarks. It can be observed from Fig. 14 that for all the benchmarks considered here, the latency of mSWNoC is lower than that of the mesh and SWNoC architectures. This is due to the small-world, network-based interconnect infrastructure of the mSWNoC with direct long-range wireless links that enables a smaller average hop-count than that of mesh and SWNoC [18], [19].

Both the MROOTS and ALASH routing strategies are implemented on the same mSWNoC architecture. The difference in latency arises due to the routing-dependent traffic distribution of the benchmarks. However, it should be noted that the difference in latency among the routing algorithms on the same architecture is small due to the fact that the traffic injection load for all these benchmarks is low and the network operates much below saturation [31]. However, the saturation characteristics of mSWNoC in presence of these routings will be further discussed in Section IV-C4.

It can be seen in Fig. 14 that ALASH has lower latency compared to MROOTS for the benchmarks considered in this paper. The weakness of MROOTS is that there is a strong tendency to generate traffic hotspots near the roots of the spanning trees. These traffic hotspots cause messages to be delayed in the network due to root congestion. Since ALASH does not have a tree-based routing strategy, ALASH does not have this root congestion problem. ALASH also guarantees the shortest physical path between any source and destination. On the other hand, MROOTS makes no guarantees about the message path length. Moreover, the priority-based layering function mentioned above helps ALASH also. Due to all these reasons, ALASH routing in mSWNoC outperforms MROOTS.

Fig. 15 shows the total normalized network energy dissipation for the mSWNoC, SWNoC, and mesh architectures. We consider the total network energy dissipation to compare the characteristics of the NoC architectures and their associated routing strategies under consideration here. It can be observed from Fig. 15 that for each benchmark the network energy is lower for the SWNoC and mSWNoC compared to the mesh architecture. Though the gain in latency for SWNoC/mSWNoC compared to the mesh is low due to the relatively lower injection loads, the improvement in energy dissipation brings forward the benefit of small-world architectures



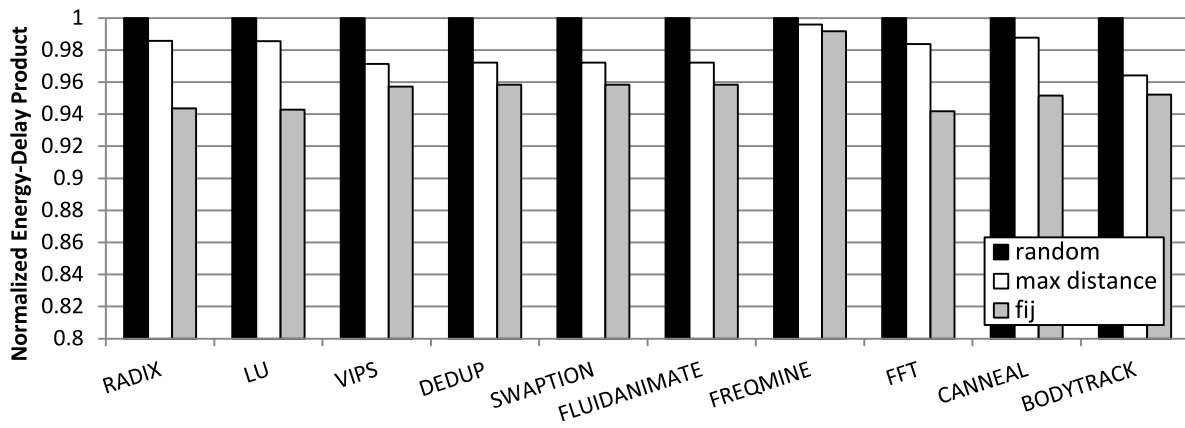


Fig. 11. Normalized energy-delay product for mSWNoC with different root placement strategies for MROOTS.

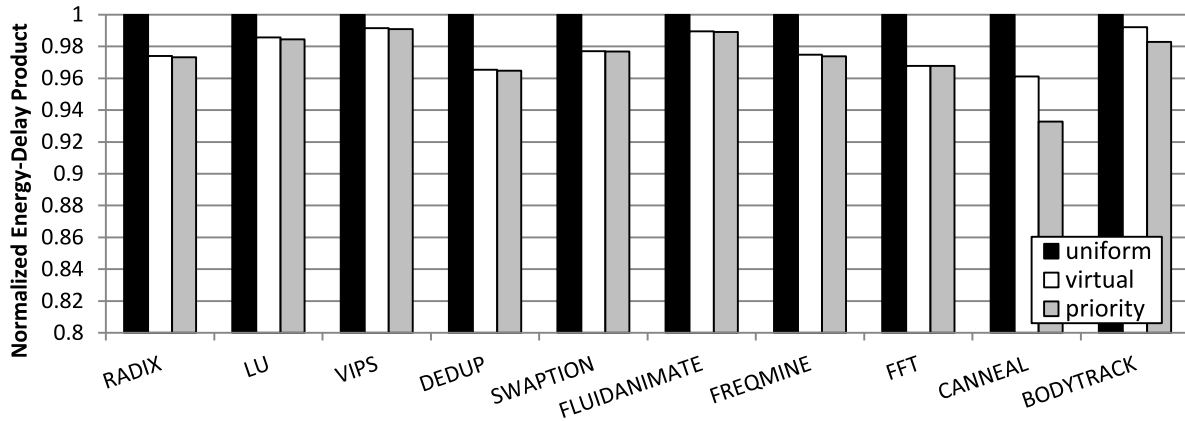


Fig. 12. Normalized energy-delay product for mSWNoC with different layering strategies for ALASH.

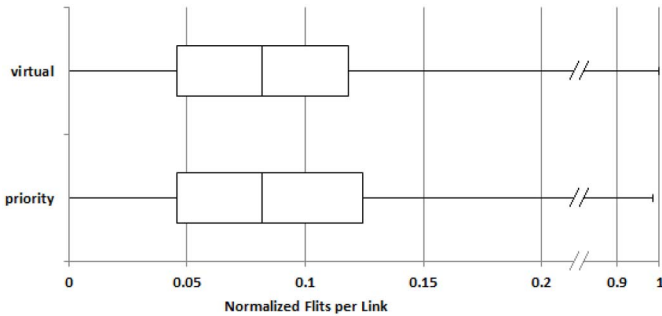


Fig. 13. Normalized flits per link distribution for the two different layering functions of ALASH for the CANNEAL benchmark.

more clearly. The two main contributors of the energy dissipation are from the switches and the interconnect infrastructure. In the SWNoC/mSWNoC, the overall switch energy decreases significantly compared to a mesh as a result of the better connectivity of the architecture. In this case, the hop-count decreases significantly, and hence, on the average, packets have to traverse through less number of switches and links. In addition, a significant amount of traffic traverses through energy efficient wireless channels in mSWNoC; consequently allowing the interconnect energy dissipation to be further decreased compared to the SWNoC architecture. It can also be observed from Fig. 15 that the energy dissipation for the two different routing strategies follows the same trend as that of

the latency. When messages are in the network longer (higher latency) they dissipate more energy. The difference in energy dissipation arising out of the logic circuits of each individual routing is very small and the overall energy dissipation is principally governed by the network characteristics.

Without loss of generality, Fig. 16 highlights the contributions of the energy dissipation for the different components of the mSWNoC and mesh architecture for the FFT benchmark. The contributors to the mesh energy are the network switches and the wireline links. The contributors to the mSWNoC energy arise from the network switches, the wireline links, and the wireless links which also include the energy of the antennas and the transceivers.

We also considered the average packet latency and network energy dissipation with different sized packets and flit lengths. We considered two different scenarios: one packet size has 64 flits per packet with 32 bits per flit and the other packet size has 32 flits per packet with 64 bits per flit. We compared the latency and normalized energy dissipation profile of the mSWNoC with respect to the mesh by varying the message characteristics as mentioned above. For brevity, we show these characteristics for two benchmarks, one with a low switch interaction rate, like RADIX, and another with relatively higher switch interaction rate, like BODYTRACK. Fig. 17 shows the average packet latency of the different packet length and flit size messages. It can be

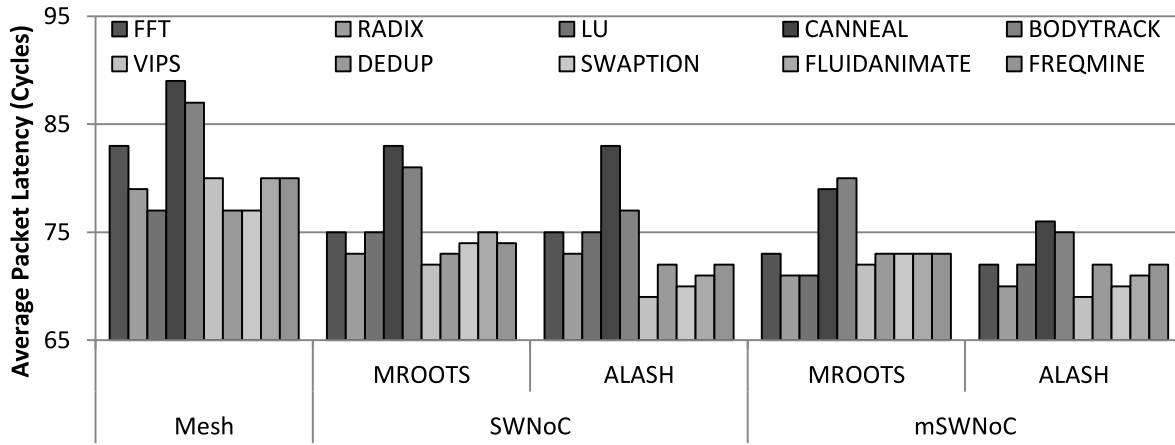


Fig. 14. Average network latency with various traffic patterns for the mesh, SWNoC, and mSWNoC architectures.

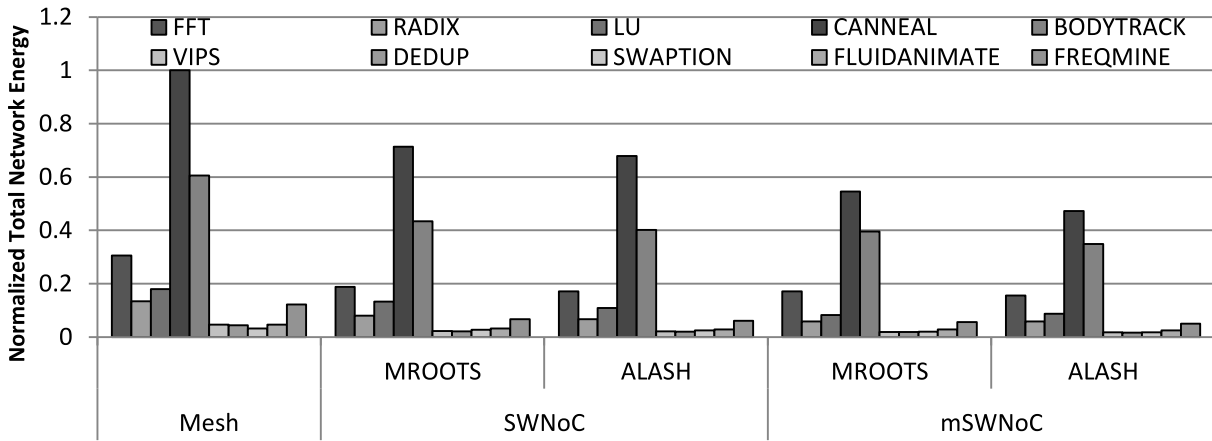


Fig. 15. Normalized total network energy with various traffic patterns for the mesh, SWNoC, and mSWNoC architectures.

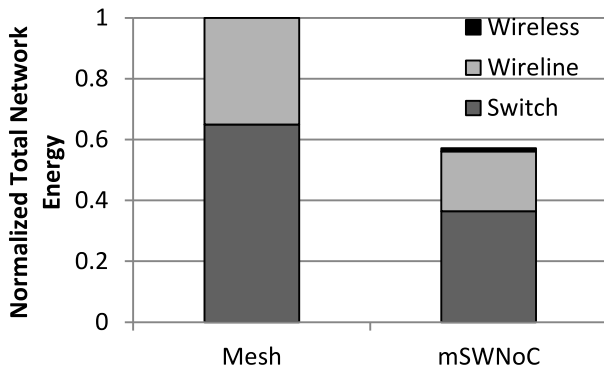


Fig. 16. Components of network energy dissipation for (a) mesh and (b) mSWNoC for the FFT benchmark.

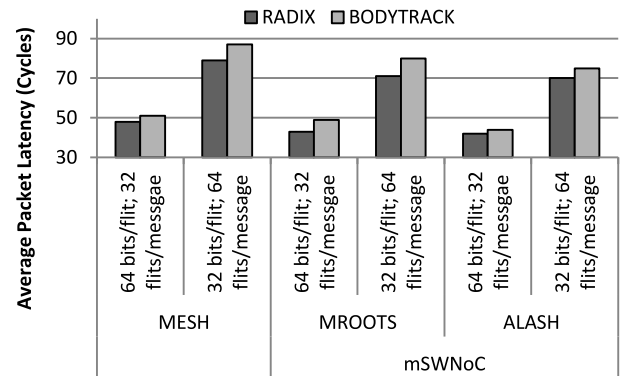


Fig. 17. Average packet latency of mesh and mSWNoC for different sized packets.

seen in Fig. 17 that although the absolute value varies for average packet latency, both ALASH and MROOTS maintain the latency improvement over mesh irrespective of the packet length or flit size. Fig. 18 shows the energy dissipation characteristics for these two packet and flit sizes. It is clear that similar to the latency characteristics, mSWNoC is always more energy efficient than the mesh.

3) *Thermal Characteristics*: In this subsection, we evaluate the thermal profile of the mSWNoC, SWNoC and mesh architectures. To quantify the thermal profile of the

SWNoC/mSWNoC in presence of the two routing strategies, we consider the temperatures of the network switches and links. We also ensured that each network element carried sufficient traffic to contribute to the overall thermal profile. The focus of this paper is to analyze the network characteristics. However, we consider the effects of the processing cores in the HotSpot simulation to accurately portray the temperature-coupling effects that the processors have on their nearby network elements.

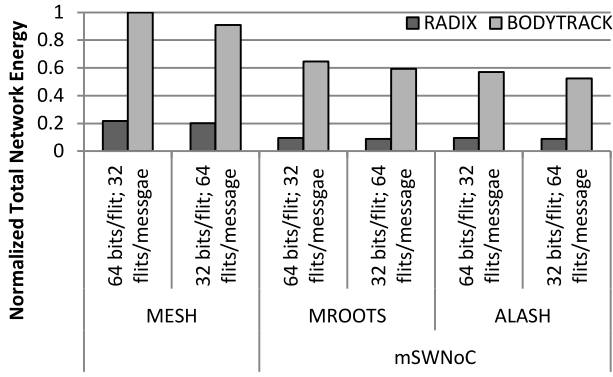


Fig. 18. Normalized total network energy of mesh and mSWNoC for different sized packets.

We consider the maximum and average switch and link temperature changes between a mesh and SWNoC/mSWNoC,  $\Delta T_{\text{hotspot}}$  and  $\Delta T_{\text{avg}}$ , respectively, as the two relevant parameters. As explained at the beginning of Section IV, the benchmarks can be put into two different categories, viz., communication and computation intensive. We consider BODYTRACK and RADIX as two representative examples for the communication and computation intensive benchmarks, respectively. However, for the other benchmarks we have observed the same trend.

Without loss of generality, we first show the variation of normalized transient temperature of the hotspot switches for the RADIX and BODYTRACK benchmarks in Fig. 19. It can be seen from this figure that the temperature initially rises with time and then ultimately saturates asymptotically at a steady-state value. We consider this steady-state temperature as the relevant parameter for further analysis.

Figs. 20 and 21 show  $\Delta T_{\text{hotspot}}$  and  $\Delta T_{\text{avg}}$  for the links and switches of the two routing strategies. It can be seen that the SWNoC/mSWNoC network architectures are inherently much cooler than the mesh counterpart. From Fig. 15, we can see that the difference in energy dissipation between the small world architectures and mesh is significant and hence, it is natural that SWNoC/mSWNoC switches and links are cooler. Fig. 20 helps depict how well each routing strategy performs in distributing the power density, and hence heat, among the network switches and links. This is due to the fact that variations in  $\Delta T_{\text{hotspot}}$  correspond to how well the routing mechanism balances the traffic within the network. The more interesting observation while analyzing the temperature profile lies in characterizing the differences among the routing strategies for the small-world architectures. ALASH performs well in distributing the traffic among the network elements. Because of this, ALASH has the lowest maximum network temperature, which can be seen in Fig. 20, where ALASH has the largest  $\Delta T_{\text{hotspot}}$ .

By observing Fig. 21, it can also be seen that the average temperature reduction in switches and links among the routing strategies is relatively unaffected. We can conclude that, reduction of the maximum temperature using ALASH has not come at the cost of increasing the average network temperature due to the inherent rerouting efforts of this strategy. Overall, it can be seen that for the routing strategies implemented,

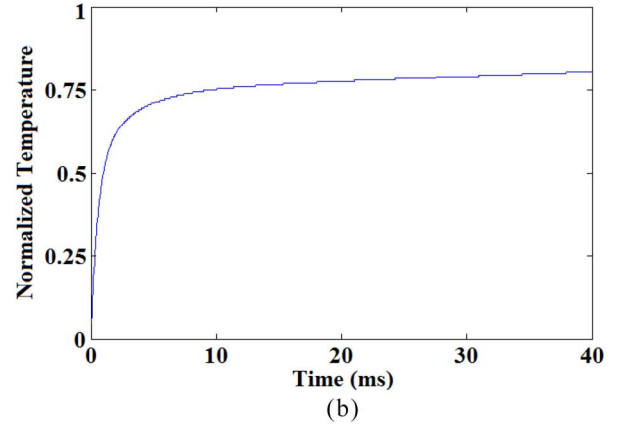
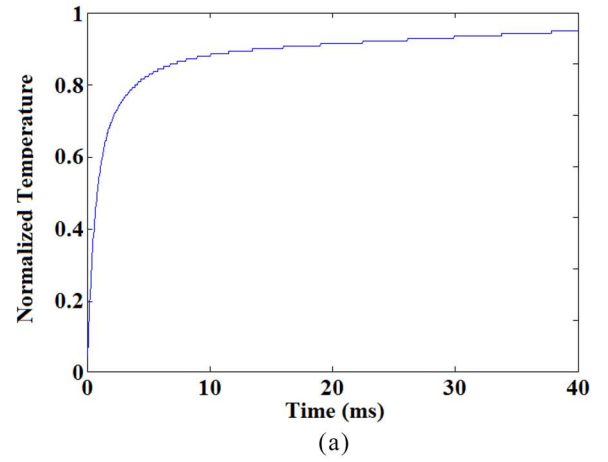


Fig. 19. Normalized transient temperature plot of the hotspot switch for (a) RADIX and (b) BODYTRACK.

we can obtain very similar latency and network energy profiles while reducing the temperature of the hotspot switches and links.

Fig. 22(a) displays the temperature distribution of the switches in the routing schemes of the RADIX and BODYTRACK benchmarks. Here, it can be seen that the MROOTS routing strategy has a larger temperature spread compared to ALASH for both the small-world architectures (the difference between the first and third quartiles is larger). The MROOTS routing strategy will form bottlenecks in the upper levels of its trees. In this case, the heat distribution will be spread further as the leaves among the trees see lighter traffic, while the near-root nodes see heavier traffic. ALASH attempts to avoid creating hotspots by having multiple shortest paths. By choosing a path that avoids local network hotspots, we can reduce the maximum network temperature quite well using the ALASH routing strategy. For RADIX and BODYTRACK, ALASH reduces the hotspot switch temperature further compared to MROOTS by 1.86 °C and 3.32 °C on mSWNoC, respectively.

Between the SWNoC and mSWNoC architectures, the SWNoC achieves a higher switch hotspot temperature reduction for the computation intensive benchmarks, as their traffic density is small. This can be seen in Fig. 22(a) as SWNoC has a lower maximum temperature over mSWNoC for both routing strategies for the RADIX benchmark. For these benchmarks, the benefits of the wireless shortcuts are outweighed



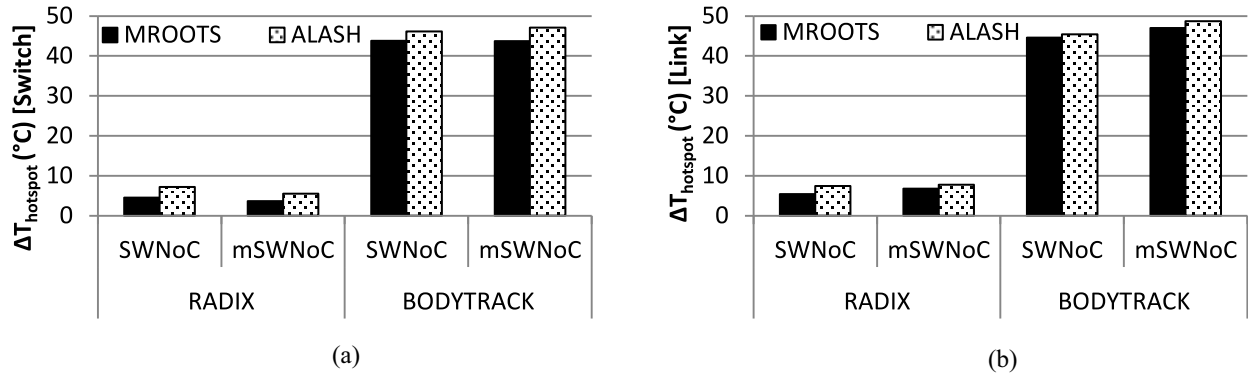


Fig. 20. Decrease in hotspot (a) switch and (b) link temperature compared to a mesh for the RADIX and BODYTRACK benchmarks.

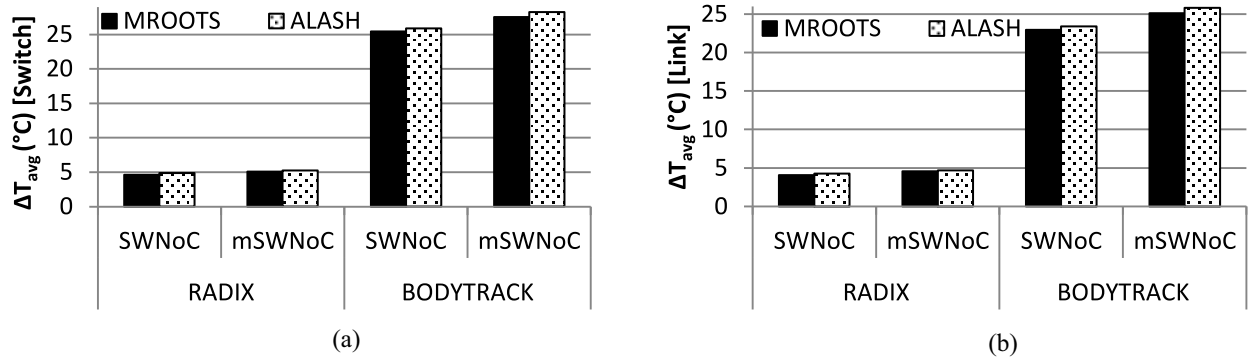


Fig. 21. Decrease in average (a) switch and (b) link temperature compared to a mesh for the RADIX and BODYTRACK benchmarks.

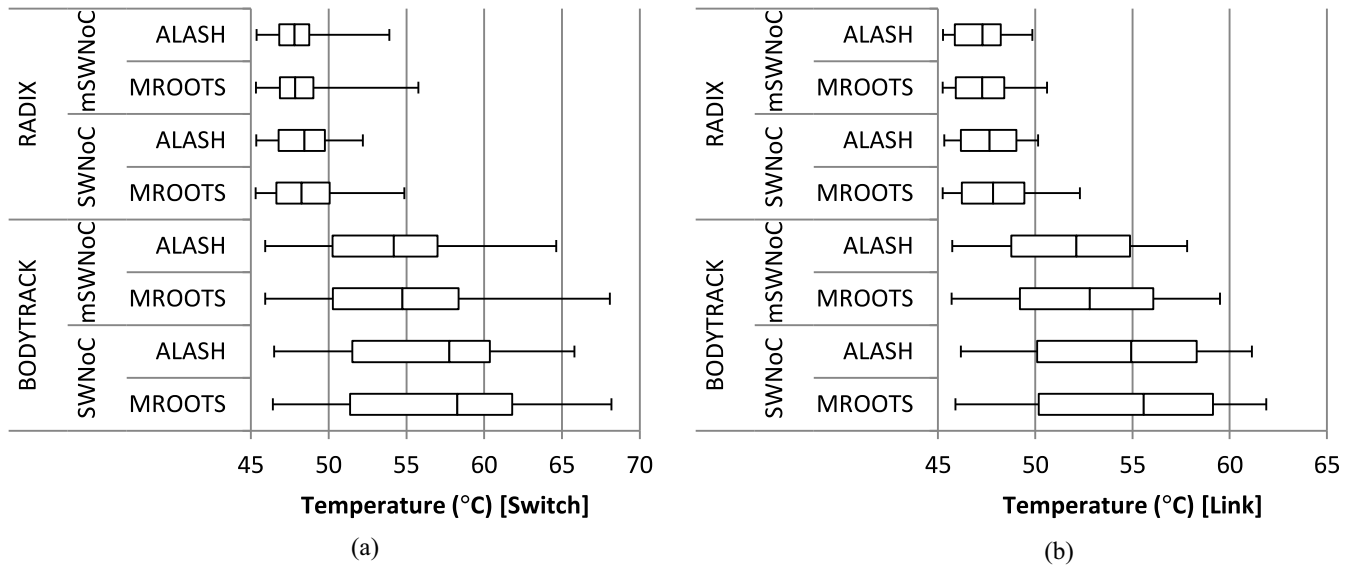


Fig. 22. Temperature distribution of (a) switches and (b) links for the RADIX and BODYTRACK benchmarks.

by the amount of traffic that the WIs attract. However, for the communication intensive benchmarks with higher traffic density, the use of high-bandwidth wireless shortcuts, in the mSWNoC, quickly relieves the higher amount of traffic that the WIs attract. In case of SWNoC, as the shortcuts are implemented through multihop wireline links, moving traffic through these wireline links takes more time and energy which correlates with less temperature reduction. This can be seen in Fig. 22(a) as SWNoC has a higher maximum temperature over

mSWNoC for both routing strategies for the BODYTRACK benchmark. Conversely, between SWNoC and mSWNoC, the mSWNoC achieves a higher link hotspot temperature reduction. This is due to the wireless links detouring significant amounts of traffic away from the wireline links.

Fig. 22(b) displays the temperature distribution of the links in the routing schemes of the RADIX and BODYTRACK benchmarks. The links follow the same temperature trend as the switches where the MROOTS routing strategy has a larger

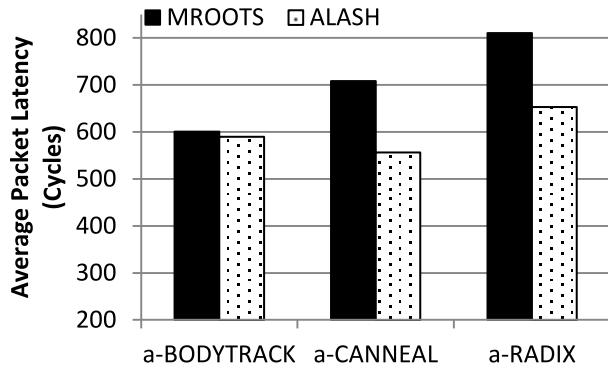


Fig. 23. Latency in network saturation of the mSWNoC for MROOTS and ALASH running a-BODYTRACK, a-CANNEAL, and a-RADIX.

temperature spread when compared to ALASH for both the small-world architectures (the difference between first and third quartiles is larger). For RADIX and BODYTRACK, ALASH reduces the hotspot link temperature further compared to MROOTS by  $0.92^{\circ}\text{C}$  and  $1.69^{\circ}\text{C}$  on mSWNoC, respectively.

4) *Performance Evaluation in Network Saturation:* As mentioned earlier, the benchmarks considered in this paper operate below network saturation. To have a detailed comparative performance evaluation of MROOTS and ALASH we also need to see the effects of these routing strategies when the network is in saturation. For this evaluation we artificially inflated the switch interaction rates for the BODYTRACK, CANNEAL, and RADIX benchmarks (a-BODYTRACK, a-CANNEAL, and a-RADIX, respectively). Fig. 23 shows the saturation latency of the MROOTS and ALASH routing strategies using these artificially inflated traffic patterns for mSWNoC. It can be seen from Fig. 23 that ALASH performs better than MROOTS. This is due to the inherent problem of any tree-based routing, where the roots of the spanning trees become traffic bottlenecks. As mentioned in Section IV-C2, when the traffic injection rates are high enough, the root switches start to become traffic hotspots. Hence, the messages get delayed in the network due to root congestion. ALASH does not have a root congestion problem and hence, outperforms MROOTS because of the adaptiveness in ALASH. It should be noted that after artificially inflating the computation-intensive benchmark loads, like RADIX, they become more communication-intensive. Hence, ALASH outperforms MROOTS in case of a-RADIX.

## V. CONCLUSION

As we demand more from our computing systems, they will be limited by power, energy, and thermal constraints. Without new energy-efficient design paradigms, producing information and communication technologies (ICT) systems capable of meeting the computing, storage, and communication demands of emerging applications will be unlikely. Millimeter-wave wireless mSWNoC is an enabling technology to design energy efficient and high bandwidth multicore architectures with improved thermal profile over conventional wireline mesh-based counterparts. In this paper, we evaluated the latency,

energy dissipation, and thermal profiles of mSWNoC in presence of rule- and path-driven irregular routing methodologies, viz., MROOTS and ALASH. In presence of these routing strategies, mSWNoC provides lower latency and energy dissipation compared to a conventional wireline mesh. Moreover, each of these routing strategies can be suitably optimized taking into account the traffic patterns generated by the benchmarks to enhance the achievable performance benefits. The difference in latency and energy dissipation profile among these routing strategies is small. However, the advantages of ALASH are clearly demonstrated specifically when the traffic load is increased. As examples, ALASH reduces the temperature of the hotspot network switch for a computation intensive benchmark, like RADIX, and for a communication-intensive benchmark, like BODYTRACK, by an additional  $1.86^{\circ}\text{C}$  and  $3.32^{\circ}\text{C}$ , respectively over MROOTS even when the system operates well below saturation. When deciding the suitable routing strategy for mSWNoC, it is clear that ALASH provides similar or better performance compared to MROOTS, while offering an improved temperature profile.

## REFERENCES

- [1] S. Deb *et al.*, "Design of an energy efficient CMOS compatible NoC architecture with millimeter-wave wireless interconnects," *IEEE Trans. Comput.*, vol. 62, no. 12, pp. 2382–2396, Dec. 2013.
- [2] T. Petermann and P. De Los Rios, "Spatial small-world networks: A wiring cost perspective," arXiv:cond-mat/0501420v2.
- [3] J. Flich *et al.*, "A survey and evaluation of topology-agnostic deterministic routing algorithms," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 3, pp. 405–425, Feb. 2012.
- [4] P. Wettin *et al.*, "Performance evaluation of wireless NoCs in presence of irregular network routing strategies," in *Proc. Design Autom. Test Eur. Conf. (DATE)*, Dresden, Germany, 2014, pp. 1–6.
- [5] H. Chi and C. Tang, "A deadlock-free routing scheme for interconnection networks with irregular topology," in *Proc. Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Seoul, Korea, 1997, pp. 88–95.
- [6] O. Lysne, T. Skeie, S.-A. Reinemo, and I. Theiss, "Layered routing in irregular networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 17, no. 1, pp. 51–65, Jan. 2006.
- [7] S. C. Woo, M. Ohara, E. Torrie, J. P. Singh, and A. Gupta, "The SPLASH-2 programs: Characterization and methodological considerations," in *Proc. Int. Symp. Comput. Archit. (ISCA)*, Santa Margherita Ligure, Italy, 1995, pp. 24–36.
- [8] B. Bienia, "Benchmarking modern multiprocessors," Ph.D. dissertation, Dept. Comput. Sci., Princeton Univ., Princeton, NJ, USA, 2011.
- [9] R. Marculescu, U. Y. Ogras, L.-S. Peh, N. E. Jerger, and Y. Hoskote, "Outstanding research problems in NoC design: System, microarchitecture, and circuit perspectives," *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.*, vol. 17, no. 1, pp. 3–21, Jan. 2009.
- [10] U. Y. Ogras and R. Marculescu, "Application-specific network-on-chip architecture customization via long-range link insertion," in *Proc. Int. Conf. Comput. Aided Design (ICCAD)*, 2005, pp. 246–253.
- [11] U. Y. Ogras and R. Marculescu, "It's a small world after all: NoC performance optimization via long-range link insertion," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 14, no. 7, pp. 693–706, Jul. 2006.
- [12] A. Kumar, L.-S. Peh, P. Kundu, and N. K. Jha, "Towards ideal on-chip communication using express virtual channels," *IEEE Micro*, vol. 28, no. 1, pp. 80–90, Jan./Feb. 2008.
- [13] S. Deb, A. Ganguly, P. P. Pande, B. Belzer, and D. Heo, "Wireless NoC as interconnection backbone for multicore chips: Promises and challenges," *IEEE J. Emerg. Sel. Topic Circuits Syst.*, vol. 2, no. 2, pp. 228–239, Jun. 2012.
- [14] D. Zhao and Y. Wang, "SD-MAC: Design and synthesis of a hardware-efficient collision-free QoS-aware MAC protocol for wireless network-on-chip," *IEEE Trans. Comput.*, vol. 57, no. 9, pp. 1230–1245, Sep. 2008.
- [15] S. B. Lee *et al.*, "A scalable micro wireless interconnect structure for CMPs," in *Proc. AMC MobiCom*, Beijing, China, 2009, pp. 20–25.

- [16] D. DiTomaso, A. Kodi, S. Kaya, and D. Matolak, "iWise: Inter-router wireless scalable express channels for network-on-chips (NoCs) architectures," in *Proc. High Perform. Interconnects (HOTI)*, Santa Clara, CA, USA, 2011, pp. 11–18.
- [17] A. Ganguly *et al.*, "Scalable hybrid wireless network-on-chip architectures for multi-core systems," *IEEE Trans. Comput.*, vol. 60, no. 10, pp. 1485–1502, Oct. 2011.
- [18] A. Ganguly, P. Wettin, K. Chang, and P. Pande, "Complex network inspired fault-tolerant NoC architectures with wireless links," in *Proc. Int. Symp. Netw. Chip (NOCS)*, Pittsburgh, PA, USA, 2011, pp. 169–176.
- [19] P. Wettin, A. Vidapalapati, A. Ganguly, and P. P. Pande, "Complex network enabled robust wireless network-on-chip architectures," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 9, no. 3, Sep. 2013. Art. ID 24.
- [20] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, Jun. 1998.
- [21] P. Wettin, J. Murray, P. P. Pande, B. Shirazi, and A. Ganguly, "Energy-efficient multicore chip design through cross-layer approach," in *Proc. Design Autom. Test Eur. Conf. (DATE)*, Grenoble, France, 2013, pp. 725–730.
- [22] B. A. Floyd, C.-M. Hung, and K. K. O, "Intra-chip wireless interconnect for clock distribution implemented with integrated antennas, receivers, and transmitters," *IEEE J. Solid-State Circuits*, vol. 37, no. 5, pp. 543–552, May 2002.
- [23] X. Yu *et al.*, "A wideband body-enabled millimeter-wave transceiver for wireless network-on-chip," in *Proc. Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Seoul, Korea, 2011, pp. 1–4.
- [24] X. Yu, S. P. Sah, B. Belzer, and D. Heo, "Performance evaluation and receiver front-end design for on-chip millimeter-wave wireless interconnect," in *Proc. Int. Green Comput. Conf. (IGCC)*, Chicago, IL, USA, 2010, pp. 555–560.
- [25] P. P. Pande, C. Grecu, M. Jones, A. Ivanov, and R. Saleh, "Performance evaluation and design trade-offs for network on chip interconnect architectures," *IEEE Trans. Comput.*, vol. 54, no. 8, pp. 1025–1040, Aug. 2005.
- [26] A. Kumar, L.-S. Peh, and N. K. Jha, "Token flow control," in *Proc. Microarchitecture (MICRO)*, Lake Como, Italy, 2008, pp. 342–353.
- [27] F. O. Sem-Jacobsen and O. Lysne, "Topology agnostic dynamic quick reconfiguration for large-scale interconnection networks," in *Proc. Int. Symp. Cluster Cloud Grid Comput. (CCGrid)*, Ottawa, ON, Canada, 2012, pp. 228–235.
- [28] N. Binkert *et al.*, "The Gem5 simulator," *ACM SIGARCH Comput. Archit. News*, vol. 39, no. 2, pp. 1–7, May 2011.
- [29] S. Li *et al.*, "McPAT: An integrated power, area, and timing modeling framework for multicore and manycore architectures," in *Proc. Microarchitecture (MICRO)*, New York, NY, USA, 2009, pp. 469–480.
- [30] K. Skadron *et al.*, "Temperature-aware microarchitecture," in *Proc. Int. Symp. Comput. Archit. (ISCA)*, San Diego, CA, USA, 2003, pp. 2–13.
- [31] P. Gratz and S. Keckler, "Realistic workload characterization and analysis for networks-on-chip design," presented at the 4th Workshop Chip Multiprocessor Memory Syst. Interconnects (CMP-MSI), Bangalore, India, Jan. 2010.



**Paul Wettin** (S'08–M'14) received the B.S. degree in computer engineering and the Ph.D. degree in electrical and computer engineering, both from the Washington State University, Pullman, WA, USA, in 2010 and 2014, respectively.

He is a Senior ASIC Design Engineer at Marvell Semiconductor, Boise, ID, USA.



**Ryan Kim** (S'13) received the B.S. degree in computer engineering from the Washington State University, Pullman, WA, USA, in 2011, where he is currently pursuing the Ph.D. degree in electrical and computer engineering.



**Jacob Murray** (S'08–M'14) received the B.S. degree in computer engineering and the Ph.D. degree in electrical and computer engineering, both from the Washington State University, Pullman, WA, USA, in 2010 and 2014, respectively.

He is a Clinical Assistant Professor and a Program Coordinator with the Department of Electrical Engineering and Computer Science, Washington State University, Everett, WA, USA.



**Xinmin Yu** (S'10) received the B.S. degree from Zhejiang University, Hangzhou, China, the M.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, and the Ph.D. degree from the Washington State University, Pullman, WA, USA, in 2002, 2006, and 2014, respectively, all in electrical engineering.



**Partha P. Pande** (M'05–SM'11) received the M.S. degree in computer science from the National University of Singapore, Singapore, and the Ph.D. degree in electrical and computer engineering from the University of British Columbia, Vancouver, BC, Canada.

He is a Professor and a Boeing Centennial Chair holder in computer engineering at the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, USA. His current research interests include novel

interconnect architectures for multicore chips, on-chip wireless communication networks, and hardware accelerators for biocomputing.

Dr. Pande is an Associate Editor-in-Chief of the IEEE DESIGN AND TEST and serves as an Editorial Board Member for the *ACM Journal of Emerging Technologies in Computing Systems* and *Sustainable Computing: Informatics and Systems*.



**Amlan Ganguly** (S'06–M'10) received the B.Tech. degree from the Indian Institute of Technology, Kharagpur, India, and the M.S. and Ph.D. degrees from the Washington State University, Pullman, WA, USA, in 2005, 2007, and 2010, respectively.

He is an Assistant Professor with the Department of Computer Engineering at Rochester Institute of Technology, New York, NY, USA. His current research interests include robust and thermally efficient multicore chips, wireless, photonic, and 3-D network-on-chip architectures.



**Deukhyoun Heo** (S'97–M'00–SM'13) received the B.S. degree from Kyungpook National University, Daegu, Korea, the M.S. degree from the Pohang University of Science and Technology, Pohang, Korea, both in electrical engineering, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1989, 1997, and 2000, respectively.

In 2000, he joined the National Semiconductor Corporation, Santa Clara, CA, USA, where he was a Senior Design Engineer involved in the development of silicon RFICs for cellular applications. Since 2003, he has been an Associate Professor with the Electrical Engineering and Computer Science Department, Washington State University, Pullman, WA, USA.

Dr. Heo has served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART II: EXPRESS BRIEFS from 2007 to 2009 and for the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES.