CMOS Compatible Many-Core NoC Architectures with Multi-Channel Millimeter-Wave Wireless Links

Sujay Deb, Kevin Chang, Miralem Cosic, Partha Pande, Deukhyoun Heo, Benjamin Belzer
School of Electrical Engineering & Computer Science, Washington State University, Pullman, WA, USA
{sdeb, jchang, mcosic, pande, dheo, belzer}@eecs.wsu.edu

Amlan Ganguly
Department of Computer Engineering
Rochester Institute of Technology
Rochester, New York, USA
amlan.ganguly@rit.edu

ABSTRACT
Traditional many-core designs based on the Network-on-Chip (NoC) paradigm suffer from high latency and power dissipation as the system size scales up due to their inherent multi-hop communication. NoC performance can be significantly enhanced by introducing long-range, low power, and high-bandwidth single-hop wireless links between far apart cores. This paper presents a design methodology and performance evaluation for a hierarchical small-world NoC with CMOS compatible on-chip millimeter (mm)-wave wireless long-range communication links. The proposed wireless NoC offers significantly higher bandwidth and lower energy dissipation compared to its conventional non-hierarchical wired counterpart in presence of both uniform and non-uniform traffic patterns. The performance improvement is achieved through efficient data routing and optimum placement of wireless hubs. Multiple wireless shortcuts operating simultaneously provide an energy efficient solution for design of many-core communication infrastructures.

Categories and Subject Descriptors
C.5.4 [Computer System Implementation]: VLSI Systems

General Terms
Performance, Design.

Keywords
Many-Core, NoC, Wireless, Small-World.

1. INTRODUCTION
The predicted evolution of many-core Systems-on-chip (SoCs) indicates a manifold increase in the number of cores on a single die over the next few years. High performance and low power are crucial for the widespread adoption of such many-core platforms. Achieving these goals cannot be attained by traditional paradigms and we are forced to re-think the basis of designing such systems, in particular the overall interconnect architecture. Networks-on-Chip (NoCs) have emerged as communication backbones to enable a high degree of integration in many-core SoCs. Despite their advantages, an important performance limitation in traditional NoCs arises from planar metal interconnect-based multi-hop links, where the data transfer between two distant blocks causes high latency and power consumption. A scalable solution can be achieved by drawing inspiration from the small-world property possessed by many natural complex networks. Such small-world networks are known for having low average inter-nodal distances with limited resources. This is achieved by adding a few long-range links in a regular lattice, resulting in a significantly lower average hop count. An attempt toward constructing small-world NoCs has been made with metal wires in the past [1]. However, that approach doesn't scale up because multi-hop wired links are required for longer distances. This paper evaluates the performance of a small-world NoC with multiple non-overlapping millimeter (mm)-wave wireless channels as long-range links. These on-chip wireless links are CMOS-compatible and do not need any new technology. But they have antenna and transceiver area and power overheads. Thus, to achieve the best performance, the wireless resources need to be optimally placed and used. To accomplish that goal, a hybrid and hierarchical network where nearby cores communicate through traditional metal wires, but long distance communications are predominantly achieved through high performance single-hop wireless links is implemented. This paper shows that network performance can be significantly improved by using a hybrid approach and by placing and using multiple simultaneously operating and non-overlapping wireless shortcuts optimally. The wireless link insertion also takes into account the target application or the traffic pattern the chip is designed for. Furthermore, we employ a distributed flow-control-based routing algorithm that ensures an optimum utilization of the wireless links. We demonstrate that the proposed mm-wave wireless NoC (mWNoC) outperforms its more traditional non-hierarchical wire line counterparts in terms of sustainable data rate and energy dissipation. It also performs better than other emerging NoC architectures like, NoC with RF-I and 3D NoC.

2. RELATED WORK
Conventional NoCs use multi-hop packet switched communication. It is shown in [2] that by using virtual express lanes to connect distant cores in the network, it is possible to avoid the router overhead at intermediate nodes, and thereby greatly improve NoC performance. Performance improvements have also resulted by inserting long range wired links following principles of small-world graphs [1].

The design principles of photonic NoC are elaborated in various recent publications [3][4]; photonic NoCs are estimated to dissipate significantly less power than electronic NoCs.

The amalgamation of two emerging paradigms, namely NoCs in a 3D IC environment, allows for the creation of new structures that enable significant performance enhancements over traditional solutions [5]. Despite these benefits, 3D architectures pose new technology challenges and the heat dissipation is a serious concern due to increased power density [6] on a smaller footprint.
Another alternative is NoCs with multi-band RF interconnects [7], wherein electromagnetic (EM) waves are guided along on-chip transmission lines created by multiple layers of metal and dielectric stack. As the EM waves travel at the effective speed of light, low latency and high bandwidth can be achieved.

Recently, design of a wireless NoC based on CMOS Ultra Wideband (UWB) technology was proposed [8]. In [9], the feasibility of on-chip wireless communication networks with miniature antennas and simple transceivers that operate at the sub-THz range of 100-500 GHz has been demonstrated. With further increases in transmission frequencies the size of antenna can be reduced, occupying much less chip real estate. One possibility is to use nanoscale antennas based on carbon nanotubes (CNTs) operating in the THz/optical frequency range [10]. The design of a small-world wireless NoC operating in the THz frequency range using CNT antennas is elaborated in [11]. Though CNT based NoCs offer orders-of-magnitude performance improvements over traditional wire line NoCs, the integration and reliability of CNT devices need more investigation. A preliminary NoC architecture with CMOS-compatible mm-wave wireless links was proposed in [12], where all the communicating nodes share a single wireless channel and hence the performance gain is limited.

This work introduces a comprehensive design methodology for a hierarchical and small-world mWNoC with multiple simultaneously operating wireless shortcuts. Through efficient distributed flow control based routing and optimal placement of wireless hubs, the mWNoC significantly outperforms traditional multi-hop NoCs.

3. PROPOSED NOC ARCHITECTURE

We propose design of a hierarchical NoC architecture with a limited number of wireless shortcuts strategically placed for optimum performance. Our goal is to use the small-world approach to build a highly efficient NoC with both wired and wireless links. The small-world topology can be incorporated in NoCs by introducing long-range, high bandwidth and low power wireless links between distant cores. The system is divided into multiple small clusters of neighboring cores called subnets. These subnets have NoC switches and links as in a standard NoC. The cores are connected to a centrally located hub through direct links and the hubs from all the subnets are connected in a 2nd level network forming a hierarchical structure. This is achieved by interconnecting adjacent hubs with wireless links and introducing a few long range mm-wave wireless links between distant hubs according to the placement scheme. The hubs connected through wireless links require wireless interfaces (WIs). As described in section 3.1, a simulated annealing (SA) algorithm is used to optimally place the WIs so as to establish optimal overall network topology under given resource constraints, i.e., a limited number of WIs. Figure 1 shows a representative hierarchical 256-core network where the subnets have a Ring-Star (a ring with a central hub connecting to every core) topology and it has 16 hubs and 7 WIs. The hubs are connected in mesh architecture with overlaid long-range wireless shortcuts on the 2nd level of the hierarchy. In this paper Mesh-RingStar architecture is used as an example since it is shown to provide the best performance-overhead tradeoff among several possible mWNoC architectures [13].

3.1 Optimum Placement of WIs

WI placement is crucial for optimum performance as it establishes high-speed, low-energy interconnects on the network. If there are \( N \) hubs in the network and \( n \) WIs to distribute, the size of the search space \( S \) is given by

\[
|S| = \binom{n}{N}
\]

Thus, with increasing \( N \), it becomes increasingly difficult to find the best solution by exhaustive search. It is shown in [11] that for placement of wireless links in a NoC, an SA based methodology converges to the optimal configuration much faster than exhaustive search. Hence, in the interest of scalability we adopt SA [14] based optimization for WI placement to get maximum benefits of using the wireless shortcuts. Initially, the WIs of all frequency ranges are placed randomly with each hub having equal probability of getting a WI.

Once the network is initialized randomly, an SA based optimization step is performed. Since only hubs contain WIs, the optimization is performed solely on the 2nd level network of hubs. The number of WIs sharing the same frequency channel is kept equal for different frequency bands along with one gateway hub (which can operate in all frequency channels). Multiple non-overlapping wireless channels are distributed among \( n \) WIs and WIs sharing the same channel form a cluster. To perform SA, an optimization metric \( \mu \) is established, which is closely related to the connectivity of the network. The metric \( \mu \) is the average distance, measured in number of hops, between all source and destination hubs. A single hop in this work is defined as the path length between a source and destination pair that can be traversed in one clock cycle. To compute \( \mu \) the shortest distances between all pairs of hubs are computed. The distances are then weighted with the normalized frequencies of communication between hub pairs. The optimization metric, \( \mu \) can be computed as

\[
\mu = \sum_{i,j} h_i f_{ij}
\]

where \( h_i \) is the distance (in hops) between the \( i^{th} \) source and \( j^{th} \) destination. The frequency \( f_{ij} \) is defined as the normalized probability of traffic interactions between subnets determined by particular traffic patterns depending upon the application mapped onto the NoC. The probability of getting access to the wireless channel for communication between any source-destination pair is designated by \( p \) which is inversely proportional to the number of WIs in a cluster \( n_i \) sharing the same frequency channel. The assumption that all the WIs are equally likely to have access to wireless channel in a cluster, \( p \) can be computed as

\[
p = 1/n_i
\]

The distance \( (d_{ij}) \) between source and destination varies depending on whether or not wireless shortcuts are used while routing. In this case, equal importance is attached to inter-hub distance and frequency of communication.

4. COMMUNICATION SCHEME

This section describes the WI components and the adopted routing strategy.
4.1 Wireless Interface
The two principal WI components are the antenna and the transceiver, whose characteristics are outlined below.

4.1.1 On-Chip Antennas
The on-chip antenna for the proposed mWNoC has to provide the best power gain for the smallest area overhead. A metal zigzag antenna[15] has been demonstrated to possess these characteristics. This antenna also has negligible effect of rotation (relative angle between transmitting and receiving antennas) on received signal strength, making it most suitable for mWNoC applications. Zigzag antenna characteristics depend on physical parameters like axial length, trace width, arm length, bend angle, etc. By varying these parameters antennas are designed to operate on different non-overlapping frequency channels in this work.

4.1.2 Wireless Transceiver Circuits
To ensure high throughput and energy efficiency, the mWNoC transceiver circuitry has to provide a very wide bandwidth as well as low power consumption. In designing the wireless transceiver, low power design considerations are taken into account at the architecture level with a design adopted from [16]. The detail description of the transceiver circuit is out of the scope of this paper. Non-coherent on-off keying (OOK) modulation is chosen, as it allows relatively simple and low-power circuit implementation. The transmitter (TX) consists of an up-conversion mixer and a power amplifier (PA). In the receiver (RX), a direct-conversion topology is used, consisting of a low noise amplifier (LNA), a down-conversion mixer and a baseband amplifier. An injection-lock voltage-controlled oscillator (VCO) is reused for TX and RX. With both direct-conversion and injection-lock technology, a power-hungry phase-lock loop (PLL) is eliminated.

4.2 Adopted Routing Strategy
In this proposed hierarchical NoC, intra-subnet data routing depends on the Ring-Star subnet topology. In the subnet, if the destination core is within two hops on the ring from the source, then the flit is routed along the ring; otherwise the flit goes through the central hub to its destination. To avoid deadlock within the subnet, we adopt the virtual channel management scheme from Red Rover algorithm [17]. The ring is divided into two equal sets of contiguous nodes. Messages originating from each group of nodes use dedicated virtual channels. This scheme breaks cyclic dependencies and prevents deadlock.

Inter-subnet data routing requires flits to use the upper level network. By using the wireless shortcuts between the hubs with WIs of the same frequency channel, flits can be transferred in a single hop between them. If the source and destination WIs are tuned to different frequencies, flits are first routed to a gateway hub via wireless links and are then transmitted using the destination WI’s frequency channel. If the source hub has no WI, the flits are routed to the nearest hub with a WI via the wired links and are then transmitted through the wireless channel. Likewise, if the destination hub has no WI, then the nearest WI hub receives the data and routes it to the destination through wired links. Between a source and destination hub pair without WIs, the routing path with a wireless link is chosen if it reduces the total path length compared to the wired path. This can potentially give rise to a hotspot situation in the WIs because many messages try to access wireless shortcuts simultaneously, thus overloading the WIs and resulting in higher latency. Token flow control [18] and distributed routing are used to alleviate this problem. The routing adopted here is a combination of dimension order routing for the hubs without WIs and South-East routing algorithm for the hubs with wireless shortcuts. This routing algorithm is proved to be deadlock free in [1]. Consequently, the distributed routing and token flow control prevents deadlocks and effectively improves performance by distributing traffic through alternative paths.

The wireless hubs are grouped into clusters, each tuned to a particular frequency. As the wireless hubs in a particular cluster use the same frequency and can send or receive data from any other wireless hub in that cluster, an arbitration mechanism must be designed to grant access to the wireless medium to a particular hub at a given instance to avoid interference and contention. To avoid centralized control and synchronization, the arbitration policy adopted is a wireless token passing protocol. (Note that the use of the word token in this case differs from the usage in the above mentioned token flow control.) In this scheme a dedicated token circulates in each cluster. The particular WIs possessing the wireless tokens can broadcast flits into the wireless medium in their respective clusters. The wireless token is forwarded to the next wireless hub in the same cluster after all flits belonging to a packet at the current hub are transmitted.

5. EXPERIMENTAL RESULTS
This section characterizes mWNoC performance through simulation and analysis in presence of various traffic patterns. Characteristics of the on-chip wireless communication channel and selection of the optimum number of WIs for different system sizes are presented, followed by detailed network simulations with various system sizes. We also present performance benchmarking with respect to two other emerging NoC architectures, viz., NoC with RF interconnects (RF-I) and 3D NoC.

5.1 Wireless Channel Characteristics
The metal zigzag antennas described earlier are used to establish the on-chip wireless links. Antenna characteristics are simulated using the ADS momentum tool. High resistivity silicon substrate (ρ=5kΩ-cm) is used for the simulation. To represent the worst case inter-subnet communication range between WIs, the transmitter and receiver were separated by 20 mm. The antenna’s forward transmission gains (S21) obtained via simulations are shown in Figure 2. We are able to obtain three different channels with 3 dB bandwidths of 16 GHz and center frequencies of 31, 57.5 and 120 GHz respectively. For optimum power efficiency, the quarter wave antennas use axial lengths of 0.73, 0.38 and 0.18 mm respectively in the silicon substrate. The antenna design ensures that signals outside the communication bandwidth for each channel are

![Figure 2. Antenna transmission gain (S21) for three non-overlapping channels.](image)
sufficiently attenuated to avoid inter-channel interference. The wireless transceiver circuitry is designed and simulated using TSMC 65-nm CMOS process. The OOK transceiver can sustain a data rate of 16 Gbps with a power consumption of 43.6 mW.

5.2 Optimal Number of WIs

To reduce hardware overhead, we aim to limit the number of WIs on the chip without significantly compromising the overall performance. We assume round-robin token circulation among WIs. The token is considered to be a single flit transmitted from the WI currently holding it to the next one. The smaller the token return time to a particular WI is, the better the network performance is since wireless medium acquisition delay is minimized. On the other hand, hop-count decreases with more WIs due to higher connectivity as a result of introduction of additional WIs in the network’s upper level. Since these are two opposing trends, a tradeoff needs to be established. Hence, we study achievable network bandwidth and packet energy as a function of the number of WIs. The upper level of the network is considered a mesh with three simultaneously operating wireless shortcuts and the subnet architecture is Ring-Star as shown in Figure 1. The WI clusters are equal in size and a single WI with transceivers of all frequencies acts as gateway between different clusters. Figure 3 shows that for a 512-core Mesh-RingStar system (32 subnets with 16 cores per subnet) bandwidth increases with number of WIs until reaching a maximum at 13 WIs (3 clusters of 4 WIs each and a gateway) and then it decreases. 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 determine the optimum number of WIs for 512-core mWNoC as 13. Similarly, for 8 and 16 subnet systems optimum performance is achieved with 5 and 7 WIs respectively.

5.3 Performance Evaluation

In this section we analyze mWNoC characteristics and study performance trends as the system size scales up. We consider three different system sizes, namely 128, 256, and 512 cores divided into 8, 16 and 32 subnets respectively. The die area is kept fixed at 20 mm x 20 mm for all system sizes. The NoC switch architecture is adopted from [19]. The hubs and NoC switches in the subnets have 4 virtual channels per port and have a buffer depth of 2 flits. Each packet consists of 64 flits. The WI ports have an increased buffer depth of 8 flits, which ensures that all messages trying to access wireless links are efficiently handled without compromising performance. Increasing the buffer depth beyond 8 gives no further performance improvement for this packet size, but gives rise to additional area overhead [13]. The WI wireless ports are assumed to have antennas and wireless transceivers. A self-similar traffic injection process is assumed.

Figure 3. Performance variation with different number of WIs for a 512-core system with 32 subnets.

The Mesh-RingStar architecture introduced earlier is simulated using a cycle accurate simulator. The subnet switches and the hub digital components are synthesized using 65 nm standard cell library from TSMC at a clock frequency of 2.5 GHz. The delays in flit traversals along the wired interconnects of the hybrid NoC architecture are considered when quantifying the performance. These include the intra-subnet core-to-hub wired links and the inter-hub links in the network’s upper level. The delays through the switches and inter-switch wires of the subnets and the hubs are taken into account as well.

Figure 4 shows achievable bandwidth of the proposed mWNoC for three different system sizes considered under a uniform random spatial traffic distribution. We considered mWNoC with one, two and three simultaneously operating wireless channels. For comparison, we also present the bandwidth of three alternative architectures of the same size: (i) a flat mesh; (ii) the same hierarchical architecture as the mWNoC, but without any long-range links; and (iii) hierarchical architecture as the mWNoC, but with shortcuts implemented using buffered metal wires instead of wireless links (BWNoC). The number of wired shortcuts is kept equal to the number of WIs for different system sizes and they are optimally placed using the same SA-based optimization used for the placement of WIs. Each wired shortcut is considered to be 32 bits, which is equal to the width of a flit considered here. The wires are designed with an optimum number of uniformly placed and sized repeaters. The mWNoC with three simultaneously operating channels outperforms all the other alternatives for the three system sizes, except for the system with buffered wired shortcuts. The flat mesh architecture performs the worst due to its high average hop count. The hierarchical architecture improves the performance by reducing hop count, but the best performance is obtained from the hierarchical architecture with multiple shortcuts due to the small-world nature of the network. The hierarchical NoC with buffered wires as shortcuts results in a higher bandwidth as multiple parallel wires can operate together. But it suffers from significant energy dissipation, which is quantified in section 5.4. It can be observed that the bandwidth of the mWNoC with three non-overlapping channels improves compared to the initially proposed mWNoC with single wireless channel [12] for all the system sizes considered. Specifically, for higher system size the performance gain is more.

Figure 4. Achievable bandwidth for different system sizes.

Figure 5. Packet energy for different NoC architectures.
5.4 Energy Dissipation

We determine the mWNoC’s packet energy dissipation. The packet energy is the energy dissipated on average by a packet from its injection at the source to delivery at the destination. The energy dissipations of the switches and hubs are obtained through synthesis using Synopsys tools with 65 nm standard cell libraries from TSMC. The energy dissipated by the wireless transceiver is determined through Cadence simulations. The energy dissipation of the wired links is obtained from the Cadence layout, assuming a 20 mm x 20 mm die area.

Figure 5 shows the packet energy dissipation of the considered architectures for uniform random traffic. The energy dissipation of the hierarchical wired NoCs with or without wireline shortcuts is significantly less than that of the flat mesh architecture. This is because a hierarchical network reduces the average hop count, and hence the latency between the cores. Packets get routed faster and hence occupy resources for less time and dissipate less energy in the process. The mWNoC further improves performance by employing multiple energy efficient long range shortcuts in the hierarchical network. In Figure 6 we show the variation of per bit energy dissipation with distance for a wired and a mm-wave wireless link. From this plot it can be observed that wireless shortcuts are always energy efficient whenever the link length is 7 mm or more. In our implementation, the minimum and maximum distances between the WIs communicating using the wireless channel are 7.07 mm and 18 mm respectively. Therefore, in this design, using the wireless channel is always more energy efficient. The mWNoC with multiple non-overlapping channels has reduced packet energy for all system sizes compared to the single channel mWNoC of [12]. Also, the mWNoC significantly reduces energy dissipation compared to the other two possible wired hierarchical architectures and can reduce the packet energy dissipation by at least an order of magnitude compared to the flat mesh.

5.5 Comparative Evaluation of mWNoC

In this section we perform a comparative analysis between the mWNoC and two other emerging NoCs. The on-chip RF transmission line (RF-I) proposed in [7] is a new interconnect technology that can improve NoC performance. Hence, we designed a small-world NoC (RFNoC) by replacing the wireless communication links of the mWNoC by RF-IIs, maintaining the same hierarchical topology. Like the wireless links, these RF links can be used as long-range shortcuts. These shortcuts are optimally placed using the same SA-based optimization used for placing WIs in the mWNoC. As mentioned in [7], in 65 nm technology it is possible to have 8 different frequency channels, each operating with a data rate of 6 Gbps and used for long-range inter-subnet communications. We also considered a 3D mesh-based NoC with four layers as in [5]. Due to the high energy dissipation hierarchical NoC without shortcuts and BWNoC are not considered for this analysis.

For this comparative evaluation, we first present the normalized bandwidth with respect to flat mesh for different system sizes in Figure 7 for uniform random spatial traffic distribution. From this result it is evident that performance benefits are more prominent for bigger systems and is highest for 512-core. Consequently, a 512-core system is considered for subsequent analysis.

We consider both uniform and non-uniform traffic patterns in this evaluation. For non-uniform traffic patterns we use synthetic and application-based traffic distributions. We considered two types of synthetic traffic. First, a transpose traffic pattern [1] is considered where cores in a certain number of subnet pairs are considered to communicate more frequently with each other. We consider three such pairs and 50% of packets originating from one of these subnets are targeted towards the other in the pair. The other synthetic traffic pattern considered is hotspot traffic [1], where each core communicates with a certain number of subnets more frequently than with the others. We consider three such hotspot subnets to which all other cores send 50% of the packets that originate from them. Transpose and hotspot traffics are mapped in 3D NoC by selecting sets of adjacent cores to form groups (equivalent to subnets of mWNoC). In the transpose traffic three of these groups communicate with each other and also we consider three hotspot groups. We consider two application-based traffics. A 1024-point FFT is considered where each core performs a 2-point radix 2 FFT computation. Multiplication of two 512x512 matrices is used to generate another application-based traffic pattern.

Figure 8 shows the achievable overall network bandwidth for the different NoC architectures in uniform and non-uniform traffic scenarios for a 512-core system. It can be observed that in case of non-uniform traffic, due to the skewed communication pattern, certain interconnects on the path are overloaded and become bottleneck affecting the overall performance of the NoC. This is most prominent in case of hotspot traffic. The 3D mesh based NoC suffers from the fact that number of layers is limited to four, which results in poor performance for 512-core system size. Though mWNoC and RFNoC have the same hierarchical architecture, mWNoC performs better. In RFNoC once the 8 shortcuts are placed they are fixed for that traffic pattern. But, in mWNoC, 13 WIs are placed depending on the traffic and wireless communication channel can be established between any of those pairs. Moreover, the total long-range link area overhead and the layout challenges of the RFNoC are more significant compared to mWNoC. For example, for a 20 mm x 20
mm die, an RF interconnect of approximately 100 mm length has to be allocated for RFNoC following the layout of [7]. This is significantly higher than the combined length of all the antennas used in the mWNoC, which is 6.45 mm for the 512-core system.

6. AREA OVERHEAD

This section, quantifies the area overhead due to the wireless deployment in mWNoCs. The antenna lengths for the three different channels range from 0.18 mm to 0.73 mm with a trace width of 10 µm. The area of each WI transceiver is 0.31 mm², 0.32 mm² and 0.34 mm² for the three selected frequency ranges. The digital part for each WI, which is very similar to a traditional wireline NoC switch, has area overhead of 0.40 mm². Therefore, the total area overhead per hub with a WI (inclusive of transceiver and antenna) is determined to be in the range of 0.70 mm² to 0.74 mm². Gateway hubs have transceivers at all three frequencies so their total circuit area requirement is 1.3 mm². Since the number of WIs is kept limited, the overall silicon area overhead is dominated by the wireline NoC switches. For example, in case of a 512-core mWNoC integrated in a 20 mm x 20 mm die, wireless transceivers consume only 5.9 % of total silicon area. The transceiver area overhead for RFNoC is obtained from [7]. Total silicon area overheads for flat mesh, mWNoC (3 Channel), RFNoC, BWNoC and 3D mesh for a 512-core system are shown in Figure 9. The extra ports in z-dimension for 3D mesh cause the total switch area requirements to increase. The area overheads of the hubs along with the required transceivers (mWNoC, RFNoC) and buffers (BWNoC) are shown separately. The transceiver area overhead for mWNoC is marginally higher than RFNoC and BWNoC. Though the overall silicon area requirements for mWNoC, RFNoC, BWNoC and 3D mesh are higher than flat mesh, the performance benefit of these emerging NoCs clearly outweighs the associated overhead.

7. CONCLUSIONS

This paper demonstrates that by an optimal utilization of long range, high bandwidth, low power, mm-wave wireless channels, significant performance improvements can be achieved in a NoC. By incorporating a hierarchical small-world topology and multiple non-overlapping wireless channels, the proposed mm-wave NoC architecture gives considerable performance gains in presence of various traffic patterns without significant area overhead.

8. ACKNOWLEDGMENTS

This work was supported in part by the US National Science Foundation (NSF) CAREER grant (CCF-0845504) and CRI grant (CNS-1059289) and NSF CAREER Grant (ECCS-0845849).

9. REFERENCES


Figure 9. Silicon area overhead for different NoCs of size 512.