Expansible Network-on-Chip Architecture

Ivan Luiz Pedroso PIRES, Marco Antônio Zanata ALVES, Luiz Carlos Pessoa ALBINI
Department of Informatics - Federal University of Paraná (UFPR) - Curitiba, Brazil,
ilppires@inf.ufpr.br, mazalves@inf.ufpr.br, albini@inf.ufpr.br

Abstract—Interconnection has a great importance to provide a high bandwidth communication among parallel systems. On multi-core context, Network-on-Chip is the default intra-chip interconnection choice, providing low contention and high bandwidth between the processing elements. However, the communication outside the chip commonly uses high performance links which have the entire communication protocol stack overhead. This paper introduces the Expansible NoC concept and architecture, which is formed by wired and wireless NoC components in order to provide a low overhead interconnection for intra-chip and inter-chip communication. ENoC couples both networks with the same simplified protocol, enabling the transmission of parallel messages directly in the NoC level. The ability of identifying new communicant on-the-fly increases its flexibility, expanding the system boundaries every time a new system is connected. The ENoC inter-chip wireless link reaches short distances working at 60 GHz with Orthogonal Frequency Division Multiplexing with Quadrature Amplitude Modulation, enabling high bandwidth communication for systems inside a single cluster rack. Experimental evaluations were performed using the Noxim simulator executing computational fluid dynamics benchmark applications. Results show that the proposed architecture improves up to 38% the performance when compared to the newest related work.

Index Terms—computer architecture, multiprocessor interconnection, system-on-chip, reconfigurable architectures, wireless networks.

I. INTRODUCTION

For intra-chip communication, Network-on-Chip (NoC) [1] represents a balanced alternative to traditional System-on-Chip (SoC) interconnection technologies such as bus and cross-bar switches. The NoC communication is performed through packets, divided into small information units called flits. Flits are transmitted from source to destination by routers, hubs and network interfaces over wired or wireless links [2].

The support of Message Passing Interface (MPI) communication in the NoC level has already been proved in an experimental Intel 48 Pentium class IA-32 cores processor [3]. This support can be very interesting to interconnect processor cores executing High Performance Computing (HPC) applications based on MPI [4]. However, such approach is bounded by the intra-chip components interconnected by the NoC.

Traditional NoC, e.g. 2D and 3D meshes, are static structures, not allowing the inclusion or removal of elements. Mangano *et al.*, [5] introduced a new concept of clustered NoC, consisting of a main NoC connected to one or several sub-NoCs. It helps to improve the network performance for clusters of SoCs by expanding resources attaching multiple sub-NoCs. However, it requires a

physical backbone in which all sub-NoCs must be connected.

Several work extended the traditional wired NoC concept to use wireless link [2], [6], [7]. The main objective is to reduce the intra-chip network diameter. Further, the wireless NoC was further extended to allow inter-chip wireless communications [8], [9]. The inter-chip communication usually is limited to a few centimeters, and all intra- and inter-chip elements must be previously known, i.e. it is neither expansible nor reconfigurable.

The present work proposes the Expansible Network-on-Chip concept (ENoC). ENoC enables the connection of multiple NoC based systems through Orthogonal Frequency-Division Multiplexing (OFDM) with 64-Quadrature Amplitude Modulation (64-QAM) wireless links. It also provides MPI communication support directly in the NoC layer. ENoC is designed for inter-chip and inter-rack communications, hence it uses high frequencies to achieve high bandwidth and throughput for low distance communications.

The main objective of the ENoC concept is the on-the-fly expansibility of clustered SoCs. It may automatically recognize and connect new devices, such as sub-NoCs, independent NoCs or even other entire SoCs. This concept takes advantage of the wireless communication paradigm, as it is possible to expand the network entities without the need of additional physical interconnections. This approach may be very useful in cluster sites such as high performance processing, high availability or load-balancing scenarios. Nevertheless, this idea can be extended for embedded systems scenario, interconnecting multiple simple devices to expand its processing power.

The main contributions of this work are:

- *Expansible NoC Architecture*: The expansible NoC architecture uses an integrated wired and wireless NoC environment in order to automatically interconnect multiple system devices into a single logical view.
- *Low Overhead Protocol*: It implements a simple NoC protocol to perform intra-chip and inter-chip communication, transferring the responsibility of the Quality-of-Service to the router hardware, simplifying the protocol stack and reducing its overhead.

The rest of this paper is organized as follows: Section II details the ENoC architecture; Section III presents an overhead evaluation between the inter-chip connections; Section IV shows the simulation results using real data; Section V contains the related work and an analysis between them and the presented architecture; and, finally, Section VI brings the final consideration and future directions.

II. Expansible Network-on-Chip

This section describes the Expansible Network-on-Chip (ENoC), an expansible interconnection architecture with hybrid wired and wireless links, for intra-chip and inter-chip communication respectively. Figure 1 illustrates the overview of two systems communicating using the Expansible Network-on-Chip proposal. ENoC extends the concepts of [5], [8] and [9] to create a computer system with the ability to on-the-fly connect to other devices using a wireless link (wireless NoC). The usage of wireless NoC eases the inclusion and removal of resources in the system, increasing its flexibility. Moreover, the communication between different systems is performed with just one low latency wireless hop, as viewed in [10]. The same low overhead protocol is used for inside and outside the chip communication, reducing thus the protocol complexity to deliver messages between systems.
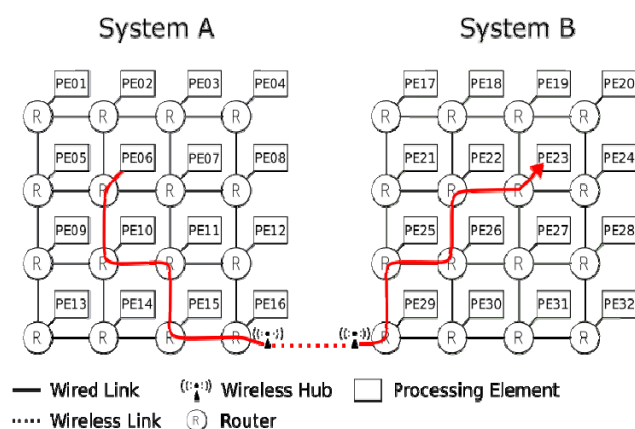

Figure 1. Expansion scenario example

The data inside ENoC is transmitted in form of packets delivered by the routers and hubs. The packet is divided into small fractions called flits. The communication inside the systems use full-duplex wired routers, enabling parallel transmissions on every router. The communication outside the systems uses the Orthogonal Frequency-Division Multiplexing (OFDM) with 64-Quadrature Amplitude Modulation (64-QAM) wireless hub.

ENoC wired architecture consists of n tiles, in which each tile is composed of a Processing Element (PE) and a wired router, these two components are linked through a network interface. The PE can be a processor core, an accelerator, or a memory device, performing operations, offering or requesting resources to the systems. Routers are responsible for correctly deliver packets through the wired mesh.

The wireless interface of the architecture consists of one or more wireless hubs for each different system, connecting and communicating to other systems. Figure 1 presents two systems with one hub each, which are coupled to the tile 16 of system A and the tile 29 of the system B. The wireless hub can be connected anywhere in the system. Connecting it to a central tile of the wired NoC, or having multiple wireless hubs per system shall reduce the distance to all the tiles inside the system.

NoC routers generally have 5 ports: one local port to connect to the PE and the ports North, East, South and West to link with its direct neighbors. Each port has a buffer for

input and output messages. In every system, there is at least one router with one additional port to connect to the wireless hub, as shown in Figure 2. This wireless hub is responsible for listening the wireless channel to perceive new systems, and for receiving and/or transmitting packets through the wireless link. Each router uses its local routing table to discover the next hop for the packets. The local routing tables must contain routes to reach all PEs and to the wireless hub inside its own system.
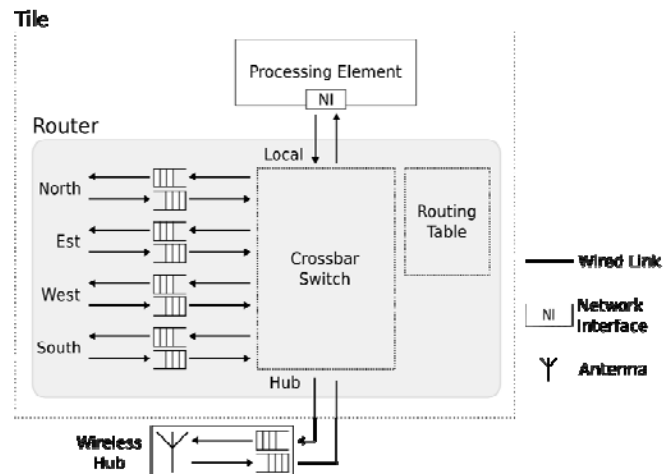

Figure 2. ENoC overview with wired routers and wireless hub

The addition of new elements (coupling connection) is performed by ENoC through the wireless link (hub), as follows:

- The wireless hub is always listening to new connections;
- In regular time intervals, it sends a beacon identifying itself;
- When a new ENoC appears in its vicinity, the wireless hub detects it;
- Both hubs exchange information about their resources, providing a new identification for the arriving communicant;
- The wireless hub works like a gateway and allows the ENoC expansion as simply as possible.

In this way, the multiple systems are linked using this wireless approach. To reach PEs in other systems, packets are always routed to the wireless hub. Note that the local routing tables do not contain entries for such PEs. Once a packet achieves the wireless hub, it is broadcasted informing the destination ID (see section II.B), in such a manner that non-target listeners can drop the packets. The wireless hub in the destination system collects the message and forwards it through the wired NoC to the destination PE. Such broadcast is clearly unsafe and can present security problems. However, at this stage ENoC is considered for well secure cluster room environments, which is supported by the fact that it uses a very small range wireless signal. Secure extensions are considered future work.

The wireless links on NoCs are well established as an alternative to the traditional wired connections. However, they are employed to connect distant sectors of the same chip, or layers of 3D-NoC, increasing the impact of this proposal.

In order to occur the expandability, it is necessary that both systems use the ENoC architecture and they must be close enough for the high bandwidth wireless signal to connect them. Furthermore, the expandability idea needs an operational system able to recognize the expansion and contraction of the resources realized by the ENoC architecture. In this way, delay and message overhead during the coupling mechanism and the energy cost are considered future work.

A. Wireless Technology

The ENoC wireless is based on the Wireless HD (WiHD) [11] working at 60 GHz frequency with omnidirectional antennas, as well as the standard OFDM for multiplexing and 64-QAM for modulation. It uses a network code rate of 13/16 and sensibility of -47 dBm achieving 6.6 Gb/s per channel, with up to 6 channels using MIMO and reaching 28 Gb/s. OFDM provides best overall performance in highly frequency selective channels as data coding and interleaving in the frequency domain captures frequency diversity more effectively [12]. The orthogonally characteristic provides power efficient signaling for a large number of users on the same channel, as each frequency is modulated with binary data to provide a number of parallel carriers each containing a portion of user data [13].

The multiple access connection and collision avoidance is achieved by the Orthogonal Frequency-Division Multiple Access (OFDMA) [11] [14], OFDMA [15] is a multi-user version of OFDM which provides multiplexing operation of user data stream onto the wireless link carriers. It has a superior access technology for broadband wireless data network compared with traditional access technologies such as Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA), due to the scalability, orthogonally and ability to take advantage of the frequency selectivity of the channel [16].

B. Link layer protocol

The ENoC packet is designed to be as simple as possible, shown in Figure 3, with a header, payload and tail. The packet has two initial bits to identify its parts. The first bit indicates if the flit is the first part of a packet (header) while the second bit marks the final flit of the packet (tail). In other words:

- '10' indicates the header;
- '00' the payload;
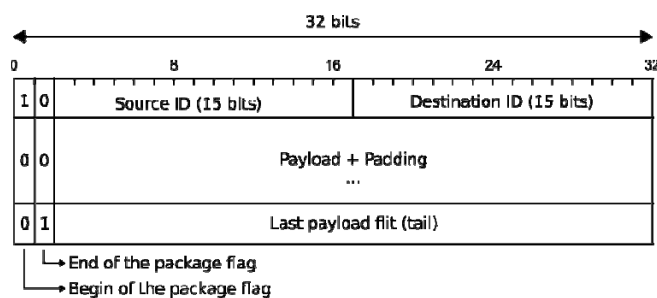- '01' the last payload flit and tail of the packet.

Figure 3. Message packet used by ENoC

The header has 15 bits to indicate the source and 15 bits to indicate the destination identification. Both fields have the same structure. Figure 4 shows the fields inside the source or destination address used in ENoC. Identification bits are divided into 7 bits to identify the PE in a local system. Due to this limit, the ENoC supports up to 128 PEs per SoC. The other 8 bits identify which wireless hub should receive the packet in case of communication with another system, hence the maximum number of systems is 256 (achieving a total of 32768 PEs). This combination allows an easy identification of each PE in the ENoC. Although these limitations are used to designate the number of PEs and total systems, different flavors of ENoC could be easily implemented. The payload must be within 1 and 1500 bytes. This range was chosen to be as small as possible to avoid padding and collision problems.
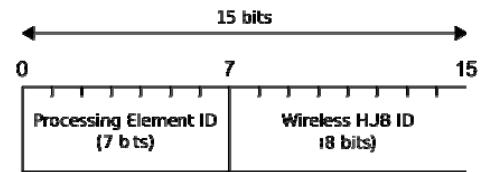

Figure 4. Fields inside source and destination address structure

III. INTER-CHIP OVERHEAD EVALUATION

This section contains an evaluation between ENoC and six other inter-chip connection technologies used on HPC clusters. It also presents the analytical analysis of the overhead (in terms of bytes) for each configuration, demonstrating the effectiveness of the ENoC low overhead network protocol. Without loss of generality, this first evaluation considers a single communication between two systems using the same intra-chip connection and different inter-chip networks.

Table I shows the parameters for each interconnection network evaluated in this paper: The 10 Gb/s Ethernet (Eth) [17]; The new Wireless Gigabit (WiGig) [18]; The InfiniBand (IB) [19]; The Wireless Interconnection with Code Division Multiple Access (WI-CDMA) [8] and The Wireless Interconnection Token-based (WI-Token) [9]. As the authors from WI-CDMA and WI-Token only reported packet size, we set it as fixed size, and for head and tail sizes we used the same parameters from ENoC. The ENoC protocol is inspired on a regular NoC [20] communication protocol (see Section II.B), while bandwidth and maximum packet size follows the WiHD parameters.

A. Theoretical analysis

To demonstrate the trade-offs between the different interconnections, a theoretical analysis of the overhead for each network was performed (Figure 5). The values were calculated based on minimal and maximum payload and also the head/tail overhead. When the data to be transmitted is smaller than the minimum payload, the packet receives padding until it reaches the minimal transmission size required. Analogously, when the data is larger than the maximum packet size, it is split in as many packets as necessary, requiring extra head/tail flits for each packet. For these situations, the extra padding or the extra head and tails are transmission overhead. The overhead of each network technology reflects their standard parameters presented on Table I.
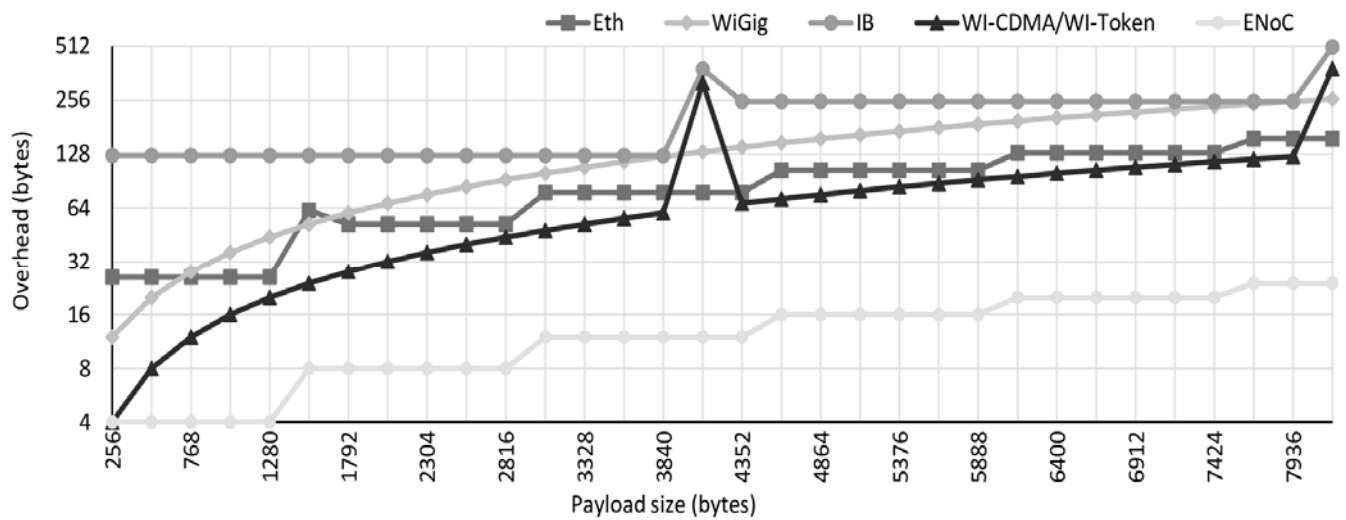
Figure 5. Overhead comparison between network links

TABLE I. NETWORK PARAMETERS

	Eth	WiGig (per channel)	IB	WI-CDMA	WI-Token	ENoC
Head + Tail. (B)	26	4	126	-	-	4
Min. Payload (B)	46	4	256	-	-	4
Max. Payload (B)	1500	144	4096	-	-	1500
Min. Packet Size (B)	72	8	382	256	256	8
Max. Packet Size (B)	1526	148	4222	256	256	1504
Data Rate (Gb/s)	10	8	50	6	16	25

ENoC has the lowest overhead for all considered network interconnections due to the lightweight and flexible communication protocol present inside the NoC. Once NoC routers provide Quality-of-Service (QoS) on the hardware level, the protocol can adopt a very small head and tail also accepting tiny payload requests, reducing thus, the overall overhead. Hence, this protocol is the correct choice for the proposed Expansible Network-on-Chip architecture, providing low overhead communication inside the chip and also among multiple systems. Moreover, the gains can be leveraged by the support of MPI messages in the NoC level [3], reducing thus the complexity to send and receive application messages.

B. Correlation to the real benchmark data

The second analysis evaluates data from the benchmark applications from the NAS Parallel Benchmark (NPB) to demonstrate the advantages of using ENoC when executing HPC applications. To perform this evaluation, the communication traces of all NPB applications [21] with the input size "A" were used.

Figure 6 shows that, most of the messages (55%) are smaller than 4096 bytes, with a considerable number of messages (19%) being smaller than 512 bytes. Depending on the protocol, the communication overhead might have a huge impact on the final throughput for these applications.

This problem can even affect the performance of state-of-the-art high speed networks such as the InfiniBand due to its requirements for minimum and maximum packet size, respectively 256 Bytes and 4096 Bytes. In fact, more than 16% of the InfiniBand packets would require extra padding. A similar problem occurs with WI-CDMA and WI-Token as they use fixed packet size of 256 bytes. In these cases, only 16% of the workload messages fits in a single packet, with

less than 4% not requiring extra padding. For ENoC, near 40% of all messages fit inside a single packet, and very few packets (less than 5%) requires extra padding.

IV. SYSTEM EVALUATION

This section contains the evaluation of the Expansible NoC proposal. The experimental setup consists of two separate systems connected by a NoCs inter-chip mesh network. This configuration maximizes the wireless bandwidth and minimizes possible collusion on the wireless link. In this evaluation the network technologies considered are: the 10 Gb/s Ethernet (Eth) standard [17]; The new Wireless Gigabit (WiGig) standard [14] which works on the 60 GHz frequency with four channels and each channel transmission rate varies between 7 Gb/s and 10 Gb/s; The InfiniBand (IB) considers 50 Gb/s as predicted by authors for the year of 2017 [19]; The Wireless Interconnection with Code Division Multiple Access (WI-CDMA) [8] and the Wireless Interconnection Token-based (WI-Token) [9] are considered with parameters reported by the authors.

Shamin *et al*. [8] present an approach (WI-CDMA) of wireless interconnects systems at 60 Ghz with a seamless communication backbone, which enable data exchange between cores from multiple chips. It uses CDMA as MAC protocol and fixed size packet achieving a data rate of 6 Gb/s covering few centimeters. Later, authors improve this work in [9] (WI-Token) using Token-based MAC instead of CDMA reaching a data rate of 16 Gb/s, with the same distance coverage.

The ENoC was based on the Wireless HD (WiHD) [11], achieving up to 28 Gb/s by merging 6 channels at 60 GHz. However, a more realistic transmission rate of 25 Gb/s was used on this evaluation.
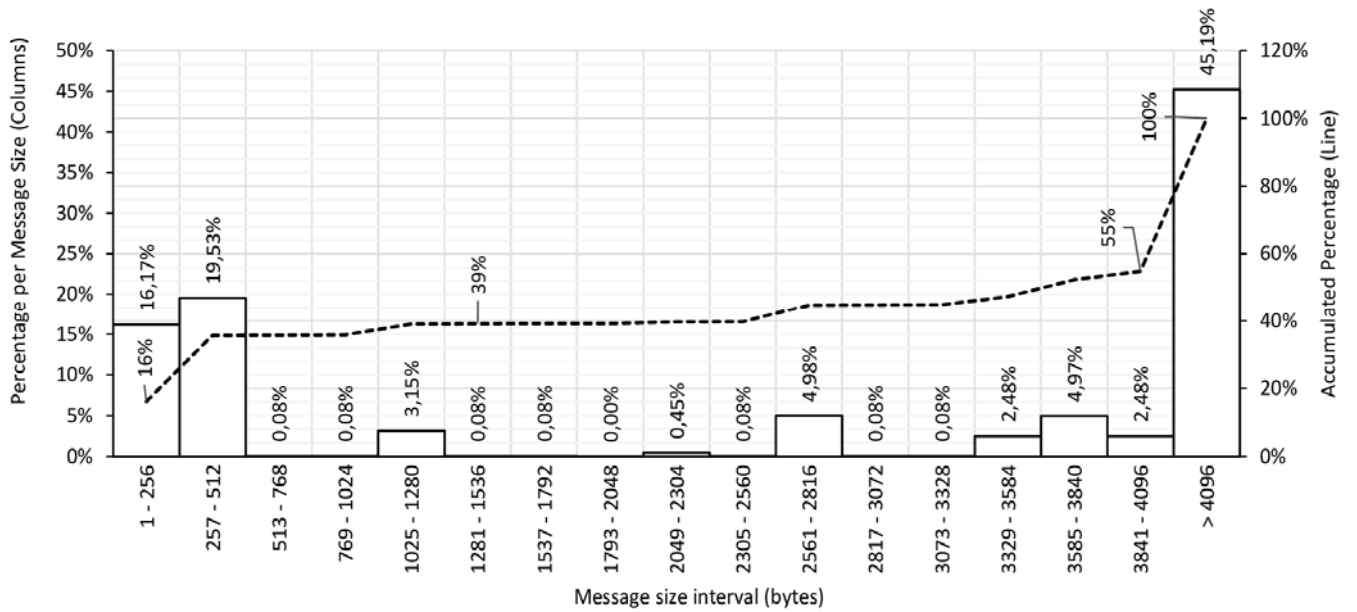
Figure 6. Histogram of messages sizes from NPB applications

Regarding the packet format, ENoC protocol is inspired on a regular NoC [20] communication protocol. Its bandwidth and maximum packet size follows the WiHD standard parameters.

Firstly, the extension made on the Noxim simulator to support the experiments is presented. Then, the parallel benchmark suite used to generate the workload for the evaluated interconnections is introduced. Finally, the parameters and metrics used in the evaluation together with the results and analysis are presented.

Our simulation focus on connecting different systems (from an HPC rack, for instance) with ENoC or other network technologies. The wired system topology is based on a 2D mesh used by many manufactures such as the Intel processors Nehalem, Sandy-Bridge, Haswell, Skylake, Xeon Phi Knights Landing, etc. Each network was modeled following standard parameters: packet head and tail size, minimum and maximum payload and packet size, and data rate, as shown in Table I (Sec. III). The flit size was set to 32 bits representing a word. Furthermore, an Expansible Network-on-Chip using the maximal theoretical data rate (ENoC-MAX) configuration based on ENoC was considered. This configuration considers the maximum bandwidth of the WiHD specification independently of multiplexing and modulation limitations, achieving approximately 40 Gb/s.

A. Simulation Environment

Noxim [22] is an open-source cycle-accurate simulator developed in System-C and C++ for heterogeneous wired and wireless NoC architectures which provides performance and power consumption analysis. The simulator works with two main elements: tile nodes and communication infrastructure. Tile nodes are the computational or storage nodes. The communication infrastructure consists of router(s) for each tile interconnected by wired links with their neighbors and possibly the wireless hub element. The hub is wired connected with one or more tiles and wireless connected with other hubs. Therefore, the simulator offers three communication patterns: tile-to-tile, tile-to-hub and hub-to-hub.

The simulator is configurable, expansible and open source. It was modified according the following requirements:

- *Input Trace:* this modification allows the user to create its own workload or to use an existing one. Moreover, it allows the usage of real data in the simulated scenario.
- *Send-wait-send*: a flow controller was implemented to control the injection of packets. It guarantees that packets will be completely transmitted, i.e. a new packet is only transmitted after the tail flit of the previous packet transmitted by the same PE reaches the destination. The application is considered finished when the last message of all traces reaches their destinations.

B. Workload Applications

This work uses the trace of all messages transmitted by eight applications of the NPB benchmark suite. The NPB version 3.3 parallelized with MPI primitives was used. It contains parallel applications implementing numerical methods for aerodynamic simulation problems. The following NPB applications uses double-precision floating-point and were written in FORTRAN: Block Tridiagonal (BT), Conjugate Gradient (CG), Fast Fourier Transform (FT), Lower and Upper triangular system (LU), Multigrid (MG) and Scalar Pentadiagonal (SP); while application Data Traffic (DT) and Integer Sort (IS) were written in C and uses mainly integer and logical operations.

The applications from NPB suite have different problem sizes (benchmark classes) named in ascending order as S, W, A, B, C and D [23]. These evaluations used the problem size A, which is the most used in real machine tests, due to its medium size and reasonable execution time to be evaluated in a simulated environment. The number of threads was set to sixteen in order to simulate four ENoC systems with 2x2 mesh size. Each NPB application have a different communication pattern, already published in different papers [24], [25].
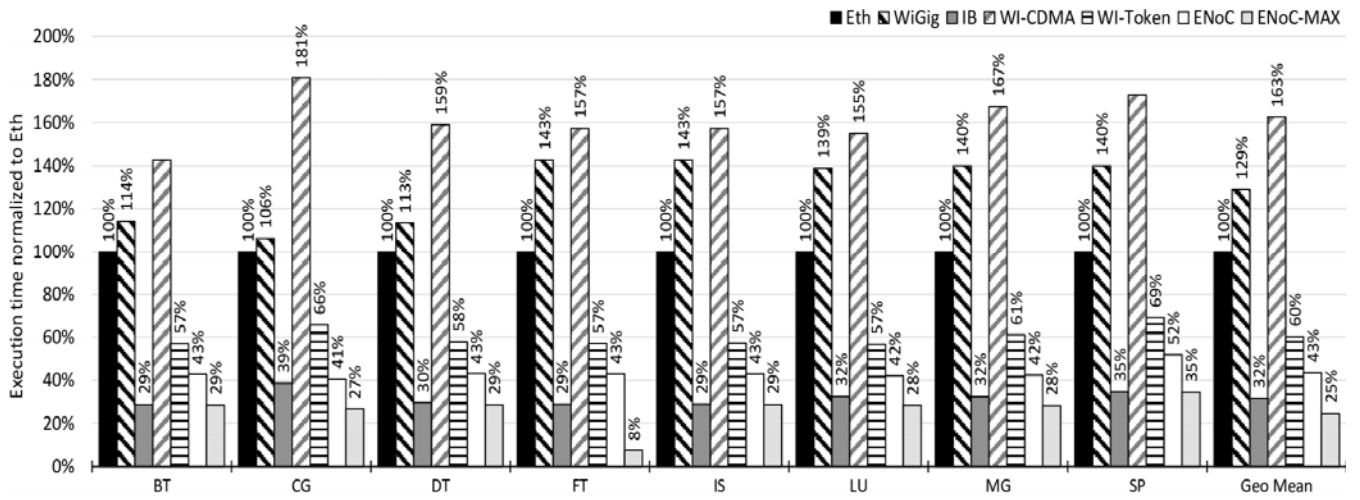
Figure 7. NPB applications execution time normalized to Ethernet results

TABLE II. OVERHEAD FOR NPB APPLICATIONS

Applications	Eth	WiGig	IB	WI-CDMA	WI-Token	ENoC	ENoC-MAX
BT	0,13%	0,02%	0,60%	0,04%	0,04%	0,02%	0,02%
CG	9,22%	0,71%	49,54%	27,09%	27,09%	0,71%	0,71%
DT	0,93%	0,07%	4,97%	2,72%	2,72%	0,07%	0,07%
FT	0,06%	0,01%	0,29%	0,13%	0,13%	0,01%	0,01%
IS	0,10%	0,01%	0,50%	0,21%	0,21%	0,01%	0,01%
LU	2,96%	0,40%	13,82%	1,06%	1,06%	0,40%	0,40%
MG	1,72%	0,13%	10,40%	6,24%	6,24%	0,13%	0,13%
SP	0,15%	0,02%	0,68%	0,03%	0,03%	0,02%	0,02%

To create a realistic workload to evaluate ENoC, an MPI wrapper was created to trace all the communication messages sent by the MPI parallel applications. During this trace generation, information regarding message size, origin and destination are stored and split per thread basis. The trace from each of the 16 threads was used to feed the different PEs.

C. Results and Discussions

This section presents the results regarding execution time and overhead obtained from the modified version of Noxim. Figure 7 shows the execution time necessary for each interconnection network to transmit all messages of the workload applications. The results for all the evaluated interconnections are normalized to the Ethernet execution time.

Regarding execution time, on average ENoC performed 2.30x, 2.96x, 3.71x and 1.38x faster than Eth, WiGig, WI-CDMA and WI-Token respectively. These results show the final impact of bandwidth and protocol overhead. However, even using a minimal protocol overhead, the ENoC performed worse than IB due to its lower bandwidth (IB was 1.37x better than ENoC on average). ENoC outperformed WI-CDMA and WI-Token, basically due to the benefits of using OFDM which enables high bandwidth with low conflict rate. Considering ENoC-MAX, it performed on average 3.45x, 4.44x, 1.09x, 5.56x and 2.07x faster than Eth, WiGig, IB, WI-CDMA and WI-Token respectively. It also overcomes the IB due to the minimal protocol overhead.

Table II shows results regarding the percentage of extra bytes transmitted due to the overhead (consisting of header, tail and eventually extra padding) for each evaluated network. In this table, the columns WI-CDMA and WI-Token are the same, a similar behavior occur between ENoC and ENoC-MAX. This is because these wireless interconnections have the same minimum and maximum packet size. It is possible to observe that ENoC has a tiny overhead compared with WI-CDMA and WI-Token, its overhead is smaller 76% on average, enabling high performance gains.

ENoC has the smallest overhead due to its simplified protocol, while the IB has the highest one on average due to the minimum packet size restriction. It is possible to correlate the overhead with the performance results. For instance, when comparing IB for the applications with less than 15% overhead (BT, DT, FT, IS, LU, MG, SP), ENoC performed 30% worse on average. However, when comparing the CG application which achieved 49% overhead on IB, it is possible to see that the performance difference is reduced to only 5%.

It is possible to notice that ENoC-MAX outperforms the IB even with a 20% smaller bandwidth. It presents better performance (9% on average) due to the simplified communication protocol used for outside chip communications. However, ENoC-MAX depicts a theoretical hypothesis based on very efficient multiplexing and modulation techniques.

Analyzing all results, ENoC specifications are easily achievable. Although it presents a performance smaller than the InfiniBand, it has a very small infrastructure cost, aggregating expansibility with zero cost. ENoC is the only proposal which present such characteristics with a small performance cutback.

V. RELATED WORK

Multiple researchers, for example [1] [20] [26], evaluated the usage of NoC to interconnect multiple processing cores with private or shared cache memory systems aiming to provide a high performance and low contention communication among on-chip resources. Nowadays, processors [27] are using NoC to interconnect cores to multiple last-level cache memory banks and the memory controller together. Embedded systems and SoCs [28] are adopting NoC to interconnect the multiple processing elements to increase performance.

In 2010, Intel released a 48-core [3] processor connected on a 6x4 2D-mesh NoC with the shared memory coherency maintained through the application parallel library (such as OpenMP or MPI). This processor was released for academic research within a special MPI library which implements message passing interface on NoC level, eliminating the communication overhead due to the protocol stack. In our work, we used the same approach to provide low overhead communication inside ENoC.

Several work extended the traditional wired NoC to use wireless signal [2] [6] [7] to communicate two elements inside the same chip. Basically, wireless NoC has been proposed to reduce the number of hops during intra-chip communications. Extensions of these work were proposed [10] to perform communication with wired and wireless link depending on the distance between sender and receiver mitigating the latency of multi-hop communications. ENoC pushes the idea of wireless NoC further, providing expansibility. It allows the connection of multiple chips together as a single system, using wired NoC inside the chip and wireless NoC outside the chip.

The work of Shamin *et al*. [9] and [8] present a design of a seamless hybrid wired and wireless interconnection network for multi-chip systems. [8] proposes the use of Code Division Multiple Access achieving only 6 Gb/s in the wireless link, while [9] uses a token based collision avoidance method reaching 16 Gb/s. Even though they propose a similar interconnection structure as the ENoC one, only ENoC is expansible and reconfigurable. Nevertheless, ENoC is based on OFDM and 64-QAM and obtains 28 Gb/s data rate. Moreover, OFDM and 64-QAM are the standard multiplexing and modulation techniques for the 60 GHz WiHD channels [11].

Regarding the ability of coupling multiple chips into a single system view, it is possible to mention the SGI Rackable computer [29]. This machinery can connect multiple systems into a single view by the Operating System (OS). It uses an InfiniBand backbone to connect the multiple systems together. ENoCs propose a similar mechanism to provide the view of multiple components as a single system, enabling message communication between processing elements from multiple ENoCs. The major difference is that ENoC expansibility is transparent to the user, while [29] coupling requires the physical connection and logical configuration of the network by the system administrator. On the other hand, [29] introduces an SO which is able to deal with dynamic resources, thus we consider that most of the OS issues regarding ENoC are already solved by nowadays operating systems.

Also, related to this proposal, there is a patent [5] which designs a system composed by multiple sub-NoCs which can be attached to a centralized component. ENoC extends this idea expanding the system through wireless link, withdrawing the need of physical modification or adaptation. Furthermore, in our proposal there is no centralized component, being fully distributed and making it simpler and faster to connect or remove components.

VI. CONCLUSIONS AND FUTURE WORK

This paper introduced the Expansible Network-on-Chip (ENoC) architecture which is the first on-the-fly expansible NoC. ENoC can identify and connect to new systems without any physical modifications or adaptations, making multiple systems work as a single team on-the-fly. Moreover, it uses a low overhead communication protocol for intra-chip (wired) and inter-chip (wireless) interconnections. It is based on 60 GHz wireless communication standard, using OFDM digital modulation scheme and 64-QAM modulation, improving the performance with low conflict communication scheme.

ENoC was evaluated using eight benchmark applications from NPB suite, comparing multiple state-of-the-art interconnection infrastructures, such as 10 Gb/s Ethernet, WiGig, InfiniBand, WI-CDMA and WI-Token

Compared to the Ethernet and WiGig, ENoC performed on average 2.30x and 2.96x better respectively. It outperformed WI-CDMA and WI-Token by 3.71x and 1.38x on average. Such improvements come from the higher data rate and low overhead protocol. When comparing ENoC to the IB, which provides the twice the bandwidth, it performed 27% slower.

Exploring an extension with the maximum data rate (ENoC-MAX), it is possible to achieve on average 2.86x better than all the compared networks (10 Gb/s Ethernet, WiGig, InfiniBand, WI-CDMA and WI-Token), outperforming the InfiniBand by 9%. These results might be achieved in the near future by altering the multiplexing and modulation schemes of the wireless link.

Future work includes the evaluation of the energy consumed by the different systems. It also includes a scalability test and the examination of ENoC security issues for inside and outside system communications, considering potential malicious elements inside the expansible system. Moreover, the operational system should be researched in order to understand all the modifications necessary to support the ENoC architecture.

REFERENCES

[1] L. Benini and G. D. Micheli, "Networks on Chips: A New SoC Paradigm," Computer, vol. 35, no. 1, pp. 70-78, 2002. doi:10.1109/2.976921.

[2] S. Deb, A. Ganguly, P. P. Pande, B. Belzer and D. Heo, "Wireless NoC as Interconnection Backbone for Multicore Chips: Promises and Challenges," IEEE Journal on Emerging and Selected Topics in Circuits and Systems - (JETCAS), vol. 2, no. 2, pp. 228-239, 2012. doi:10.1109/JETCAS.2012.2193835.

[3] J. Howard *et al.*, "A 48-Core IA-32 message-passing processor with DVFS in 45nm CMOS," in 2010 IEEE International Solid-State Circuits Conference - (ISSCC), 2010, pp. 108-109. doi:10.1109/ISSCC.2010.5434077.

[4] H. C. de Freitas, L. M. Schnorr, M. A. Z. Alves, and P. O. A. Navaux, "Impact of Parallel Workloads on NoC Architecture Design," in 18th Euromicro Conference on Parallel, Distributed and Network-based Processing - (PDP), 2010, pp. 551-555. doi:10.1109/PDP.2010.53.

[5] D. Mangano and I. A. Urzi, "System for Designing Network-on-Chip Interconnect Arrangements," US Patent App. 14/940,026, 2016.

[6] 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 Journal of Solid-State Circuits, vol. 37, no. 5, pp. 543-552, 2002. doi:10.1109/4.997846.

[7] M. F. Chang, J. Cong, A. Kaplan, M. Naik, G. Reinman, E. Socher and S. Tam, "CMP Network-on-Chip Overlaid with Multi-band RF-interconnect," in IEEE 14th International Symposium on High Performance Computer Architecture - (HPCA), 2008, pp. 191-202. doi:10.1109/HPCA.2008.4658639.

[8] M. S. Shamim, J. Muralidharan, and A. Ganguly, "An Interconnection Architecture for Seamless Inter and Intra-Chip Communication Using Wireless Links," in Proceedings of the 9th International Symposium on Networks-on-Chip - (NOCS), New York, NY, USA, 2015, p. 2:1-2:8. doi:10.1145/2786572.2786581.

[9] M. S. Shamim, N. Mansoor, R. S. Narde, V. Kothandapani, A. Ganguly, and J. Venkataraman, "A Wireless Interconnection Framework for Seamless Inter and Intra-Chip Communication in Multichip Systems," IEEE Transactions on Computers - (TC), vol. 66, no. 3, pp. 389-402, 2017. doi:10.1109/TC.2016.2605093.

[10] D. DiTomaso, A. Kodi, D. Matolak, S. Kaya, S. Laha, and W. Rayess, "A-WiNoC: Adaptive Wireless Network-on-Chip Architecture for Chip Multiprocessors," IEEE Transactions on Parallel and Distributed Systems - (TPDS), vol. 26, no. 12, pp. 3289-3302, 2015. doi:10.1109/TPDS.2014.2383384.

[11] WirelessHD Consortium, "WirelessHD Specification Version 1.1 Overview," Specification. California, USA, 2010.

[12] R. C. Daniels, J. N. Murdock, T. S. Rappaport, and R. W. Heath, "60 GHz Wireless: Up Close and Personal," IEEE Microwave Magazine - (MMM), vol. 11, no. 7, pp. 44-50, 2010. doi:10.1109/MMM.2010.938581.

[13] T. S. Rappaport, Wireless Communications: Principles and Practice. Prentice Hall PTR, 2002.

[14] C. J. Hansen, "WiGiG: Multi-gigabit wireless communications in the 60 GHz band," IEEE Wireless Communications - (MWC), vol. 18, no. 6, pp. 6-7, 2011. doi:10.1109/MWC.2011.6108325.

[15] M. Rohling, T. May, K. Bruninghaus, and R. Grunheid, "Broad-band OFDM radio transmission for multimedia applications," Proceedings of the IEEE, vol. 87, no. 10, pp. 1778-1789, 1999. doi:10.1109/5.790637.

[16] H. Yin and S. Alamouti, "OFDMA: A Broadband Wireless Access Technology," in IEEE Sarnoff Symposium - (SARNOF), 2006, pp. 1-4. doi:10.1109/SARNOF.2006.4534773.

[17] D. J. Law, A. Healey, P. Anslow, S. B. Carlson, V. Maguire, and M. Hajduczenia, "IEEE Standard for Ethernet," IEEE Computer Society, Section One, 2015.

[18] WIFI Alliance, "60 GHz Technical Specification," WiFi Alliance, Version 1.0, 2016.

[19] Infiniband TA, "InfiniBand Architecture Specification Volume 1," InfiniBand Trade Association, Release 1.1, 2002.

[20] C. A. Zeferino and A. A. Susin, "SoCIN: a parametric and scalable network-on-chip," in 16th Symposium on Integrated Circuits and Systems Design - (SBCCI), 2003, pp. 169-174. doi:10.1109/SBCCI.2003.1232824.

[21] M. Frumkin, H. Jin, and J. Yan, "Implementation of NAS Parallel Benchmarks in High Performance Fortran," NASA, 1998.

[22] V. Catania, A. Mineo, S. Monteleone, M. Palesi, and D. Patti, "Noxim: An open, extensible and cycle-accurate network on chip simulator," in IEEE 26th International Conference on Application-specific Systems, Architectures and Processors - (ASAP), 2015, pp. 162-163. doi:10.1109/ASAP.2015.7245728.

[23] R. F. V. der Wijngaart and H. Jin, "NAS Parallel Benchmarks, Multi-Zone Versions," NASA, 2003.

[24] D. Bailey, T. Harris, W. Saphir, R. F. V. der Wijngaart, A. Woo, and M. Yarrow, "The NAS parallel benchmarks 2.0," NASA, 1995.

[25] M. Diener, E. H. M. Cruz, L. L. Pilla, F. Dupros, and P. O. A. Navaux, "Characterizing communication and page usage of parallel applications for thread and data mapping," Journal of Performance Evaluation - (JPEVA), vol. 88, pp. 18-36, 2015. doi:10.1016/j.peva.2015.03.001.

[26] H. Krichene, M. Baklouti, P. Marque, J. L. Dekeyser, and M. Abid, "SCAC-Net: Reconfigurable Interconnection Network in SCAC Massively Parallel SoC," in 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing - (PDP), 2016, pp. 759-762. doi:10.1109/PDP.2016.94.

[27] M. Yuffe, E. Knoll, M. Mehalel, J. Shor, and T. Kurts, "A fully integrated multi-CPU, GPU and memory controller 32nm processor," in 2011 IEEE International Solid-State Circuits Conference - (ISSCC), 2011, pp. 264-266. doi:10.1109/ISSCC.2011.5746311.

[28] R. Rajsuman, System-on-a-Chip: Design and Test, 1st ed. Norwood, MA, USA: Artech House, Inc., 2000.

[29] S. Saini et al., "An early performance evaluation of many integrated core architecture based sgi rackable computing system," in International Conference for High Performance Computing, Networking, Storage and Analysis - (SC), 2013, pp. 1-12. doi:10.1145/2503210.2503272.