

An Energy Efficient Topology Augmentation Methodology Using Hash Based Smart Shortcut Links in 2-D Mesh

Vaishali Sodani¹, Naveen Choudhary²

¹*Department Of CSE, College Of Technology & Engineering*

²*Department Of CSE, College Of Technology & Engineering*

Abstract: The Network-on-chip (NoC) paradigm was introduced as a scalable solution for on-chip communication. NoC architecture can be designed as regular or application-specific network topologies. Performance metrics based optimized designed can be achieved through application specific topologies while less design time and efforts based NoC architecture can be achieved by using regular network topologies. However, due to the lack of fast paths between remotely situated nodes, regular topology architectures for a given routing methodology may suffer from long packet latencies and higher energy dissipation. In this paper, an application specific energy efficient topology augmentation methodology is proposed to insert smart shortcut links to 2D-Mesh with practical design constraints. Exhaustive experiments are performed taking a large number of test cases to validate and test the proposed methodology. Experimental results show that the proposed methodology exhibit good performance in terms of dynamic communication energy as well as in terms of latency compared to the regular 2D-Mesh.

Keywords: Network-on-chip paradigm · System-on-Chip Communication · Dyanamic Communication Energy · Latency in Network Topology.

I. INTRODUCTION

As the number of cores on a single chip is increasing, the communication between various components on the chip is one of the greatest challenges as it directly affects the systems' performance and cost [Saman Amarasinghe. 2007]. Traditionally, the interconnection networks used for communication among the components on a chip were bus-based and point to point links [Ali, M., Welzl, M. and Zwicknagl, M. 2008]. The Network-on-chip (NoC) paradigm was introduced as a scalable solution for on-chip communication [Dally, W. J. and Towles, B. 2001]. NoC has substituted traditional bus-based and point to point networks. These on-chip networks (NoCs) can accommodate multiple flows between cores simultaneously, provides parallelism and communication concurrency. A major challenge to the efficiency of multi-core chips is the energy used for communication among cores over on-chip networks [George, B. 2012]. As the number of cores increases, this communication energy also increases which is one of the important problems in the NoC and imposing serious constraints on design and performance of both applications and architectures [Hu, J. and Marculescu, R. 2003]. Network-on-Chip (NoC) architecture can be designed as regular or application-specific (irregular) network topologies [Hu, J. and Marculescu, R. 2005]. Application specific custom network topologies are advantageous in terms of optimized design according to given performance metrics and regular network topologies are advantageous in terms of its modularity, lower design time and efforts required and thus are suitable for mass production [Teimouri, N. et al. 2011]. However, due to the lack of fast paths between remotely situated nodes and many hops between different communicating nodes regular topology architectures for a given routing methodology may suffer from long packet latencies and higher communication power dissipation [Teimouri, N. et al. 2011].

One approach to solve this problem or to reduce the energy consumption of the regular NoC is to induce the shortcut links according to the application [Ogras U.Y. and Marculescu, R. 2005]. The shortcut links will reduce the number of hops traversed by each packet, resulting in reduction of dynamic communication energy and reduction in packet latency [Ogras U.Y. and Marculescu, R. 2005].

In this paper, an energy conscious methodology is proposed to insert smart shortcut links to 2D-Mesh with practical design constraints (such as link length, port constraints and also to add minimum number of shortcuts) according to the application/traffic characteristics for a given routing function for optimizing dynamic communication energy consumption and latency in multi core chips.

Exhaustive experiments were performed taking a large number of test cases to validate and test the proposed methodology. Experimental results show that the energy efficient topology augmentation using hash based smart shortcut links 2D-Mesh (EETA-HSS 2D-Mesh) derived from the proposed methodology exhibit good performance in terms of dynamic communication energy as well as in terms of latency compared to the regular 2D-Mesh.

The remaining paper is arranged as follows: a significant review of literature is presented in section 2. In section 3, detailed description of the proposed methodology is described. Further, in Section 4, Simulations results of the proposed methodology and detailed analysis of the results are presented. At last the paper is summarized in section 5.

II. LITERATURE REVIEW

Dally, W. J. and Towles, B. 2001, introduced the on-chip interconnection network which substituted the global wiring and simplified the layout of NoC with well-controlled electrical parameters leading to significant lower power dissipation and higher bandwidth. Further, the report submitted by [George, B. 2012] proposes methods for modelling and optimizing energy consumption in multi- and many-core chips, with special focus on the energy used for communication on the NoC. Hu and Marculescu [Hu, J. and Marculescu, R. 2003] showed that mapping algorithms reduce over 60% of energy

consumption when compared to ad hoc mapping solutions. Murali and De Micheli in [Murali, S. and Micheli, G.D. 2004] implement a similar solution. The focus of their publications is to present an algorithm that maps the cores onto mesh NoC architecture under bandwidth constraints, aiming to minimize the energy consumption.

Choudhary, N. 2012, elaborated various topology designs of NoC such as regular network on chip and irregular NoCs. The popular regular topologies are mesh, tori, fat trees, spidergon, octagon, ring and k-ary n-cubes. After experimentation the author found that the Irregular NoC exhibit better performance in terms of flit latency and throughput. The irregular NoC shows that throughput is increased by 31.7% and average flit latency is decreased by 5.9 and 22.4 clock cycles in comparison to regular NoC (2D-Mesh) with XY and OE routing algorithms respectively.

Fuks, H., Lawniczak, A. 1999, investigated models of computer data networks and examined that with the insertion of additional random links impacts the performance of the entire network is enhanced. In [Ogras U.Y. and Marculescu, R. 2005], authors present a methodology to automatically synthesize an architecture where a few application-specific long-range links are introduced on a regular mesh network. Regular-mapping and irregular application specific topology-augmentation has been done by researchers. Some work has been done in this direction but either they are not application specific, or not in consensus with the chosen routing function and some do not consider practical design constraints for synchronous communication in on-chip interconnection network.

Therefore, the presented work focused on topology augmentation according to the application which is in consensus with the chosen routing function and considers practical design constraints for synchronous communication. The main focus of the research is to optimize the dynamic communication energy.

III. PROPOSED METHODOLOGY

In [Ogras U.Y. and Marculescu, R. 2005], authors present a methodology to automatically synthesize an architecture where a few application-specific long-range links are introduced on a regular mesh network. In Figure III.1, a standard 4x4 mesh network is shown which consists of processing element, router and unidirectional link with few long range links.

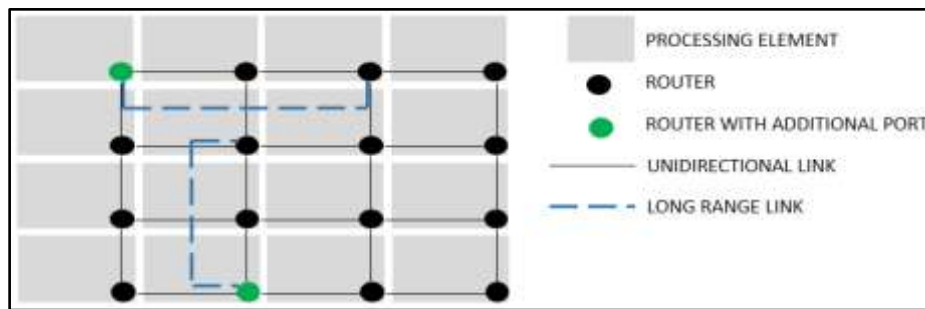


Figure III.1 Long range links are added to a standard 4X4 mesh network.

For the routers with long-range links, [Ogras U.Y. and Marculescu, R. 2005]

$$d_M(i, j) = \begin{cases} d_M(i, j) & \text{if no long range link is attached to } i \\ \min(d_M(i, j), 1 + d_M(k, j)) & \text{if } l_{(i,k)} \text{ exists} \end{cases} \quad (III.1)$$

In the above equation, $l_{(i,k)}$ means that node i is connected to node k via a long-range link [Ogras U.Y. and Marculescu, R. 2005].

An example shown in Figure III.1, demonstrates that the use of long range link decreases the number of hops traversed by a packet for reaching to the destination node. Ogras and Marculescu [Ogras U.Y. and Marculescu, R. 2005] stated that as the number of hops reduces, there is considerable reduction in the average packet latency and total energy consumed at each router is also minimized. Hence, a key improvement in the developed network obtained in the course of topology generation with minimum impact on the network topology.

Therefore, inspired from the authors [Ogras U.Y. and Marculescu, R. 2005] work, in this paper a smart topology augmentation technique which is application specific and adaptable to routing function especially table based routing is proposed. Moreover, unlike [Ogras U.Y. and Marculescu, R. 2005], the proposed methodology also take care of the practical design constraints such as link length and number of ports. Here, it should be noted that the dynamic communication energy consumption of NoC architecture depends on the amount of energy consumed in transmitting one bit of data on links between tiles (cores) as well as on the energy consumed in transporting one bit of data through a router.

A. The Dynamic Communication Energy Model in NoC:

Ye, T., Benini, L. and Micheli, G. D. 2002, proposed a NoC energy model for power consumption of network routers in which the average energy consumption of sending one bit of data from tile ti to tile tj is:

$$E_{bit}(t_i, t_j) = n_{hops} \times E_{R_{bit}} + (n_{hops} - 1) \times E_{L_{bit}} \quad (3.1)$$

Where n_{hops} is the number of routers the bit passes. $E_{R_{bit}}$ and $E_{L_{bit}}$ in Eq. (3.1) are constants for a given regular design, Their values depend on the router architecture, inter-tile link geometry (width, length, shielding, etc.) and can

usually be derived through profiling/simulation or analysis. Thus, using Eq. (3.1), the communication energy consumption can be analytically calculated, provided that the communication volume between any communicating tile pair is available. For the NoC networks with unequal link length, the $((n_{hops} - 1) \times E_{L_{bit}})$ part of Eq. (3.1) can be replaced as the summation of bit energy consumed by each link/channel in the route, the bit follows from communication source tile/core to the destination tile/cores.

In this paper, the energy model described above is used as it provides an efficient approximation for the on chip network architectures under consideration with reasonable accuracy at this level of abstraction.

Hop count: Hop count is the number of routers the bit traverses from source tile to destination tile. The communication energy model [Hu, J. and Marculescu, R. 2005] for the network on chip clearly shows that dynamic communication energy consumption linearly relates to the average number of hops the packet traverse for its journey from source tile to the destination tile. Therefore the considerable amount of dynamic communication energy saving can be achieved by reducing the hop count in the routing paths of a given on chip interconnection architecture. To reduce energy consumption, it is important to identify the energy efficient architectures. Therefore the energy-performance trade-offs need to be considered. Depending on the parameter selected an efficient methodology is proposed that is to be designed based on the selected performance metrics with help of long range interconnects for standard NoCs.

B. Distributed Table based routing and NoC:

Many deterministic as well as adaptive routing in NoC can be implemented with the help of tables in various routers of NoC. Although, our focus is on deterministic table based routing as they are most popular as they tend to reduce traffic in comparison to adaptive routing. Distributed table based routing: for the 2D-Mesh regular topologies as well as for the irregular topologies where the tables are filled offline with the respective topology supportable routing scheme.

C. HashMap and LinkedHashMap:

In the proposed methodology the smart shortcut links are selected with the help of the hashing principles by storing all the possible combinations in a LinkedHashMap datatype which offers efficient retrieval and comparison operations.

LinkedHashMap is one of the commonly used Map implementation in Java. LinkedHashMap inherits from HashMap and implements Map interface. Basic building block is same as that of HashMap.

Map is an associative array data structure used for mapping or associating value(s) with a key. So we can insert key-value pair in the map with the expectation that later on value can be looked up (or searched) given the key as the input. Insert (put) and search (get) are two most important operations on the map. In a normal array keys are indices of the array and value is the object stored the given index; whereas in associative array, keys could be of any type but still we can use them to quickly locate the value.

LinkedHashMap uses same array structure as used by HashMap with extra overhead for maintaining a linkedList(or interconnecting all elements in a particular order). This means that it stores data in the array of Entry objects.

LinkedHashMap's Entry class adds two extra pointers before and after to form a doubly linked list.

D. Proposed Methodology:

In the paper, an energy efficient topology augmentation methodology is proposed for application specific on-chip interconnection networks. In the proposed methodology a customized NoC topology is synthesized by tailoring the generic topology to achieve energy efficient communication. Eq. (3.1) clearly shows that dynamic communication energy linearly relates to the number of routers/hops the packet traverses from the source tile to the destination tile i.e. as the routers in communication path increases, the communication energy also increases. If we are able to reduce the number of hops in the communication path for a chosen routing function, the considerable amount of energy saving can be achieved in the generic NoC architecture. The reduction in the number of hops in communication path can be done by augmenting the generic topology in 2D by inducing smart shortcut links (within the same plane) according to the application characteristics and also constraining parameters (physical constraints) such as length of the shortcut link, number of added extra links (shortcut links) and the port availability per tile. The length of the shortcut link is constrained because in on-chip interconnection networks communication generally needs to be clock synchronized and to avoid the long wiring complexity for synchronous communication as well as to avoid long channel fabrication challenges, the short links are generally preferred. The total numbers of shortcut link are constrained to get maximum energy saving by adding minimum number of shortcut link and also to avoid the area overhead so as not to make complex wiring and also to avoid excessive congestion related issues. This number of added shortcut link is decided according to the application characteristics in the system for example number of shortcut links equals to some percentage of application characteristics such as 20%, 40%, 60% and so on. The port availability per tile means addition of the shortcut link to a tile will be possible if there are free ports available within the tile. Figure III.2 gives complete architecture of the proposed methodology which is mainly composed of five phases.

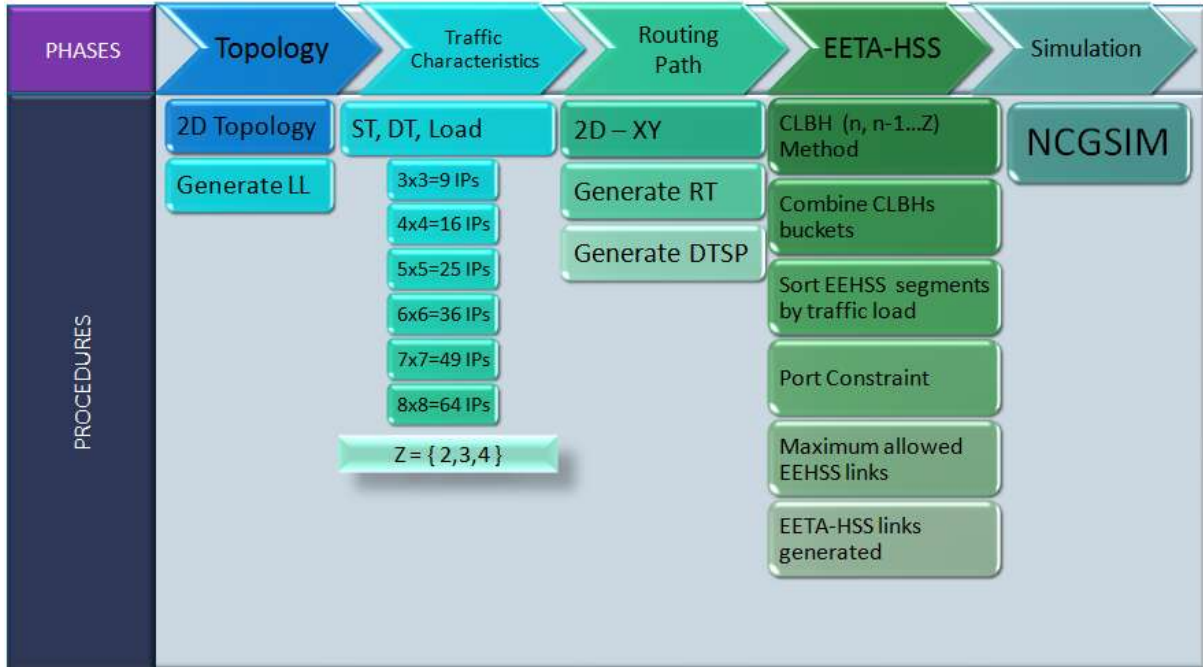


Figure III.2 Process flow diagram of the Proposed Methodology

A brief description of each phase is as follows:

Phase 1: Input Processing Phase Topology

1. 2D NoC Topology can be characterized as input for the proposed methodology.
2. The input NoC topology will describe the interconnection among the processing elements (i.e. communication infrastructure along with the available ports and link/channel lengths for example in 2D mesh, the channel length between the neighbouring tiles can be assumed as 200 mm).

Phase 2: Input Processing Phase: Traffic Characteristic Characterization

1. Application specific traffic characteristics are taken as input after converting in the appropriate format.
2. Each application traffic characteristic is represented as Source (ST) – the source tile id and destination (DT) – the destination tile id.
3. For our experimental evaluation, we have chosen the popular 2D mesh architectures. However, the proposed methodology is adaptable to any standard NoC topology. In 2D mesh, the tiles of the topology can be represented as (x, y) coordinates or it can also be represented by assigning unique tile IDs to each tile. We have assigned unique tile IDs to each tile ID of the NoC topology to keep the proposed methodology to be adaptable to any given standard NoC topology be it 2D.

Phase 3: Pre Processing Phase: Routing path generation

1. The underlying NoC architecture is assumed to support in order packet delivery. Therefore, the proposed methodology only entertains deterministic routing functions, which can give a single unique path from source tile (S_i) to destination tile (d_j). Any deterministic routing function can be given as input to this phase.
2. Given a deterministic routing function, this phase will find the corresponding routing paths (i.e. the routers and links needed to be traversed) and generate the appropriate routing table entries for each router in the path in addition to the energy required in transmitting a bit on the proposed path. For the given source tile (S_i) and destination tile (d_j) as per the traffic requirement in the application traffic characteristic for the given routing function.
3. For instance assuming 2D mesh topologies as with XY routing function as input. This phase will do the following:
 - The routing path, $r_{i,j}$, corresponds to the path to route all the packets between s_i to d_j which is determined using XY/XYZ routing algorithm for 2D respectively. Each directed edge $r_{i,j}$ has the following properties:
 - i. The routing path determines average energy per flit consumed in joules in transporting a single bit of information from s_i to d_j , i.e., E_{bit} .
 - ii. $l(r_{i,j})$ corresponds to the group of links that forms the routing path $r_{i,j}$.
 - Hence, the appropriate routing tables (RT) are generated based on 2D mesh based topology.

Phase 4: Energy Efficient Topology Augmentation using Hash based Smart Shortcut Links (EETA-HSS)

This phase comprises of eight steps for generating a topology by tailoring the generic topology to optimize energy by using hash based smart shortcut links based on application specific characteristics and chosen routing function for energy efficient communication among various tiles of NoC with practical design constraints such as number of port, channel/link length.

1. **Step I: Consecutive Links Bucket using linked Hash map (CLBH) Method** – The links that forms the routing path $r_{i,j}$ are broken into a group of consecutive (n, n-1, ..., z) links which are then stored in the respective buckets. The loads

in terms of communication energy consumption of the same links are added in the respective buckets. The parameter is design specific and depends on the channel length and clock frequency for synchronization.

2. **Step II: Combine CLBH n and CLBH (n, n-1,...z)** – These buckets holding the n consecutive links & (n, n-1,...z) consecutive links are then aggregated by combining energy load of the links having same source & destination and then merged together which represent all possible Energy Efficient Hash based Smart Shortcut: **EEHSS** links which could be inserted.
3. **Step III:** Sort EEHSS links by required Energy load (as per traffic characteristic) in the descending order and then following constraints are applied depending on the NoC topology to be augmented. The Energy load for given traffic characteristic can be estimated based on the required Bandwidth in traffic characteristics and number of links and routers and shortcuts.
4. **Step IV:** Traverse the entire EEHSS list by selecting the link with highest energy load and remove any other link which overlaps the selected link so that the list contains only those links which do not have any overlap and are sorted by energy load in descending order. This step is necessary to avoid any redundant shortcut as for deterministic routing time there can only be a fixed chosen path for given source-destination pair.
5. **Step V:** Depending on the number of ports allowed at each tile, EEHSS links with the maximum energy load are selected and rest are removed. Maximum number of EEHSS allowed is a function of number of input traffic characteristic. This constraint is basically kept to avoid excessive area overhead due to addition of excessive and in some cases redundant shortcuts.
6. **Step VI:** In this step, the augmented topology with EEHSS links is generated.

Simulation: NCGSIM (Network-on-Chip based Generic Simulator)

The performance of an augmented topology with EEHSS links is evaluated in terms of dynamic communication energy consumption as well as latency on the popular NoC simulation framework. [Vyas, K., Choudhary, N. and Singh, D. 2013]

1) Algorithm:

The architecture of the proposed methodology in Figure III.3.

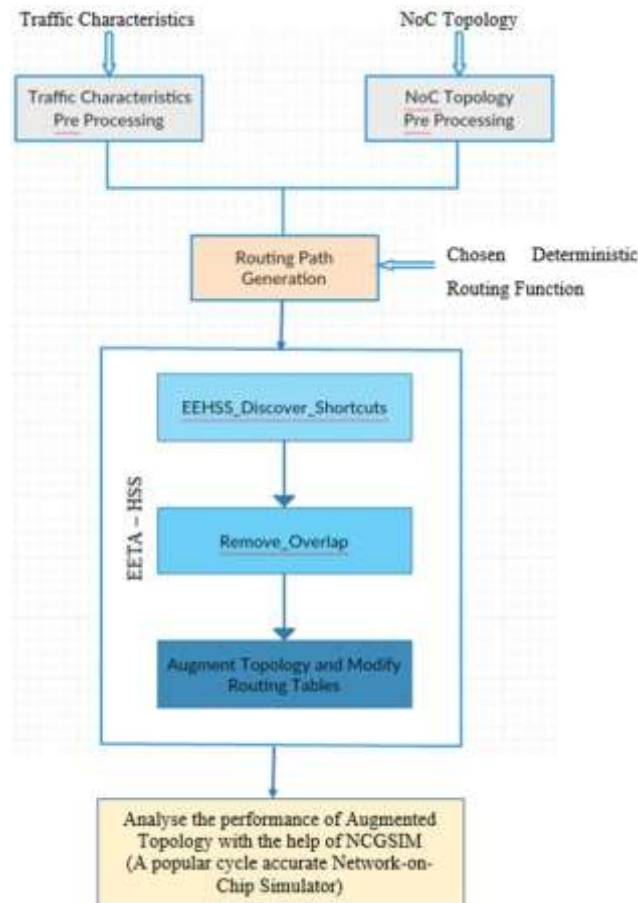


Figure III.3 Architecture of the Proposed Methodology

The description of the various modules is given below.

1. **Traffic Characteristics Pre Processing Module**

- Application specific traffic characteristics are taken as input after converting in the appropriate format.

- Each application traffic characteristic is represented as Source (ST) – the source tile id and destination (DT) – the destination tile id.
 - In 2D mesh, the tiles of the topology can be represented as (x, y) coordinated and (x, y, z) coordinates respectively or it can also be represented by assigning unique tile IDs to each tile.
- 2. NoC Topology Pre Processing Module**
2D and 3D NoC Topology can be characterized as input for the proposed methodology.
 - 3. Routing Path Generation Module**
We have chosen XY routing function for 2D regular Mesh topology.
 - 4. Input (for module EETA-HSS)**
 - Chosen standard (Generic) 2D NoC topology architecture specified (for instance 2D-Mesh).
 - Traffic Characteristics (TC) (Source tile id, Destination tile id, desired Bandwidth of communication)
 - Routes (i.e. paths) as per given Traffic Characteristics and chosen deterministic routing function.
 - Constraints: number of ports per tile (max_ports), wire length (wire_length).
 - 5. Module EETA-HSS (Traffic Characteristics (TC), Deterministic routing function, max_ports per tile, max_allowed_shortcuts)**
 - 5.1 Sub Module EEHSS_Discover shortcuts (TC , Chosen Deterministic routing function)**
 - Explore all the routing paths for each TC as per the chosen routing function. where $r = \{r_{ij} / r_{ij}\}$ is the routing path connecting *tile i & tile j*
 - For each r_{ij} links forming the routing path are broken into a group of consecutive (n, n-1,...z) links (path segments) which are then stored in the Consecutive Links Bucket using linked Hash map (CLBH n, CLBH n-1...CLBH z) respectively along with path id and Energy load per path segment.
 - The above mentioned CLBHs buckets are then aggregated by combining Energy load of the path segments having same source & destination and then merged together to form Energy Efficient Hash based Smart Shortcut (EEHSS).
 - Sort in the descending order of Energy load and call this set as S.
 - 5.2 Sub Module Remove_Overlap (EEHSS All Links)**
 - Discover all traffic load sharing segments in S (referred as Critical Energy path Segments (CEPSs)) as per the TC in the paths explored in sub module EEHSS_Discover Shortcuts and evaluate total dynamic communication energy consumed by these segments.
 - While the list EEHSS set is not empty, extract the element i with highest energy load and insert it in list EEHSS_NoOverlap Links as follows:
 - For each *path id p* stored in *path id* attribute of element i, identify all n & n-1 consecutive links in the same path id p with overlap
 - For all such overlapping segments decrease energy load proportionally and remove path id p from the respective element in list EEHSS.
 - Sort and return EEHSS NoOverlap in the descending order by Energy load.
 - 5.3 Sub Module Apply Constraints (EEHSS_NoOverlap, max_ports, max_no_allowed_links)**
 - For each link in EEHSS NoOverlap Links check if port constraint is violated or not, if it is violated remove the respective shortcut.
 - Retain the top max_no_allowed_links from EEHSS_NoOverlap and delete the rest and return the augmented topology.

IV. EXPERIMENTAL RESULTS AND DISSCUSSION

In this section the experimental setup, platform characteristics as well as the NoC simulation setup is described. Finally, after exhaustive experimentation, the obtained results were analysed.

A. Experimental setup & Platform characteristics

The platform under consideration is composed of $M \times N$ tiles which are interconnected according to the 2D-Mesh architecture as shown in Figure IV.1.

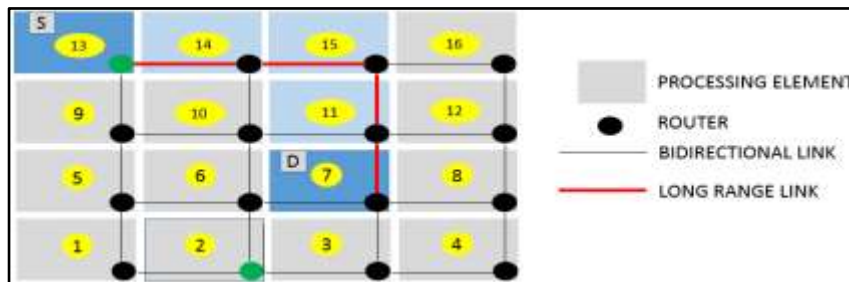


Figure IV.1 A
Mesh architecture

4x4 2D-

For 2D-Mesh interconnection architecture XY (dimension order routing) [Duato, J., Yalamanchili, S., Ni L. 2003; Choudhary, N. 2012] routing algorithm (which is a deterministic routing algorithm) is used to route the packets across the

network in the NoC. XY routing routes packets first along the X dimension then packets are routed along the Y dimension to reach its destination and it offers deadlock free paths. To compute the dynamic communication energy consumption of the 2D-Mesh interconnection architecture, the energy model of Section III is used. The proposed methodology is implemented using java programming language on Intel Core i5 based processor system with 6 GB memory (RAM) and open source operating system platform Ubuntu with various sizes of 2D-Mesh topologies. The derived augmented topology (EETA-HSS) from the proposed methodology is validated on on-chip interconnection network simulator NC-G-SIM which is discussed in the following subsection B.

B. On-Chip interconnection network simulator: NC-G-SIM setup

The performance of EETA-HSS derived from the regular 2D-Mesh using energy conscious topology augmentation methodology is evaluated and analysed by setting up a network on chip simulator NC-G-SIM. The communication is assumed to be of constant bit rate. Number of application (traffic) characteristics data sets were randomly generated using TGFF [Dick, R. P., Rhodes, D. L., & Wolf, W. 1998], with diverse bandwidth requirement of the IP Cores and randomly generated communication bandwidth requirement. For performance comparison NC-G-SIM was run for 10000 clock cycles to evaluate the network performance. The average total communication energy per flit received at destination, total dynamic communication energy and average per flit latency are taken as performance metrics. The flit latency performance parameter is defined as how much time it takes for a flit to reach the destination node from the source node.

The proposed methodology is applied to 6 set of different 2D-Mesh topology, for each topology 10 different test cases exhibiting diverse traffic patterns (generated using TGFF [Dick, R. P., Rhodes, D. L., & Wolf, W. 1998]) were generated and evaluated. For each and every test case, using the proposed methodology the table files (configuration files) were generated by taking into consideration the traffic characteristics that include specifically the source-destination pairs. Also, for the same size of topology and same test cases (same traffic scenarios) the regular topology (2D-Mesh) is generated so that the performance comparison can be made fairly with the EETA-HSS generated from the proposed energy conscious methodology. The configuration files of NC-G-SIM for a given traffic pattern are generated using the proposed methodology and supplied to the simulator as an input for performance evaluation. After having experimented on number of topology sets, the average of obtained results is represented graphically to demonstrate the effectiveness of the proposed methodology. The description and analysis of the experimental results obtained from the proposed methodology is presented in the following sub-section.

C. Experimental Results

To obtain the large range of traffic scenarios for comprehensive comparison, multiple traffic characteristic were randomly generated using TGFF [Dick, R. P., Rhodes, D. L., & Wolf, W. 1998] with diverse bandwidth requirement of the IP cores and randomly generated communication volumes.

Figure IV.2 shows comparison in terms of average total dynamic communication energy per flit between 2D- and EETA-HSS 2D-Mesh with varying number of shortcut links added with same traffic scenarios at reduced and at heavy load. Firstly, the number of hash based smart shortcut links added is equal to 20% of application characteristics then 40% of application characteristics then 60% of application characteristics, after that the 80% of application characteristics and lastly the number of shortcut links added is equal to the 100% of application characteristic which are referred in

Figure IV.2 as EETA-HSS(20)-2DMesh, EETA-HSS(40)-2DMesh, EETA-HSS(60)-2DMesh, EETA-HSS(80)-2DMesh, EETA-HSS(100)-2DMesh respectively, and this same representation is used throughout the discussion on results.

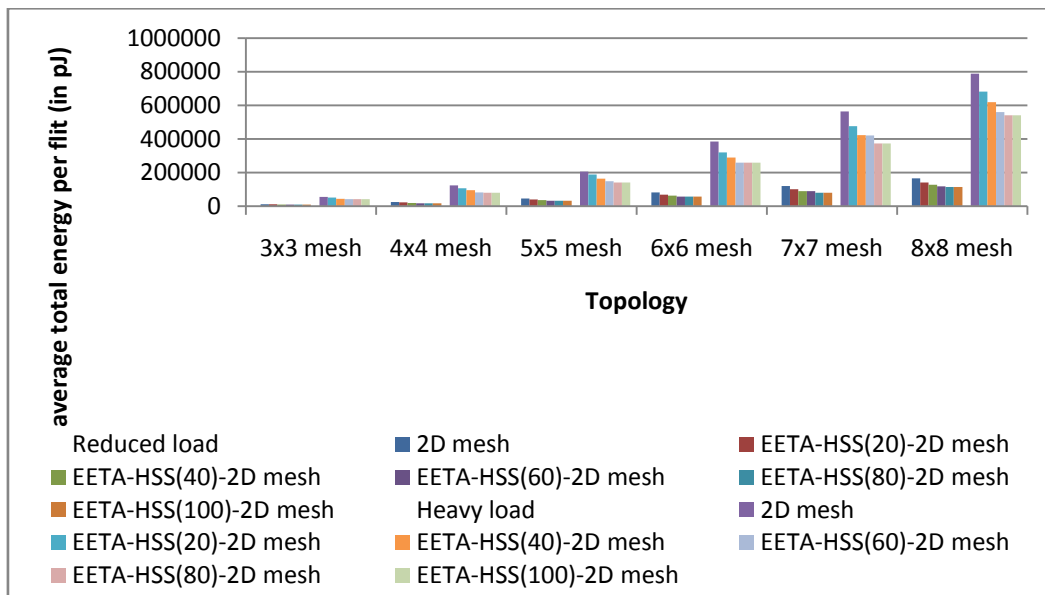


Figure IV.2 Comparison of Average Total Energy per flit of 2D-mesh and EETA-HSS 2D-mesh at reduced and heavy load for various topologies.

Figure IV.3 and Figure IV.4 gives comparison on the basis of average total dynamic communication energy per flit (in pJ) at reduced load and at heavy load between the EETA-HSS (OL)-2D mesh topology generated using proposed energy conscious methodology and 2D-Mesh for same traffic scenarios. For EETA-HSS, the number of shortcut links added is equals to the 80% of the application/traffic characteristics, the reason behind selecting this parameter (no. of shortcut link=80% of application characteristics) is explained experimentally later in this chapter.

As shown in Figure IV.3 and Figure IV.4, EETA-HSS topology exhibit better performance than the regular 2D-Mesh. The EETA-HSS (OL)-2D mesh topology saves about an average of 29.59% at reduced load and 29.66% at heavy load of average total energy per flit consumption as compared to the regular 2D-Mesh.

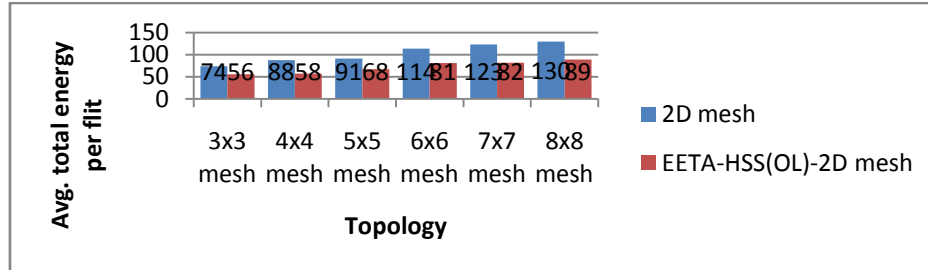


Figure IV.3 Comparison of Average Total Energy per flit between 2D-Mesh and EETA-HSS-2D Mesh (no. of short cuts= 80% of application characteristics) at reduced load for various topologies.

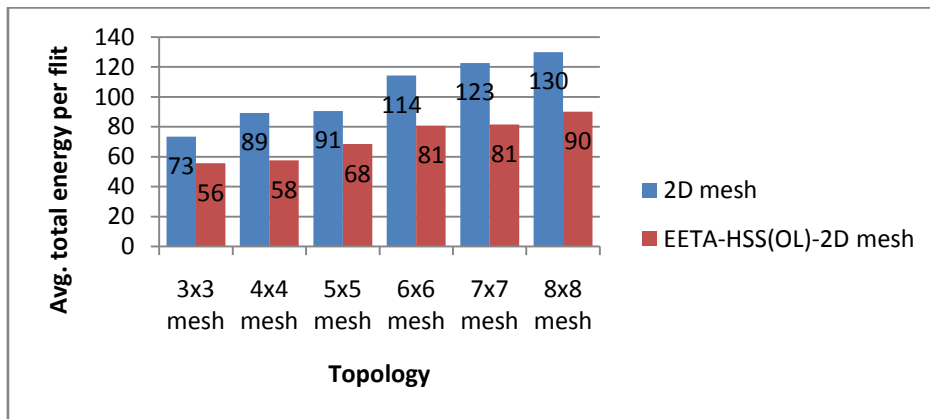


Figure IV.4 Comparison of Average Total Energy per flit between 2D-Mesh and EETA-HSS-2D Mesh (no. of short cuts= 80% of application characteristics) at heavy load for various topologies.

In the proposed methodology packets needs to traverse less number of hops from the source tile to the destination tile due to the inclusion of shortcut links, leading to reduction in average total energy per flit. Figure IV.5 gives comparison average per flit latency (in clock cycles) between 2D- Mesh and EETA-HSS(OL) 2D-Mesh generated using proposed energy conscious methodology for same traffic scenarios at reduced as well as at heavy load.

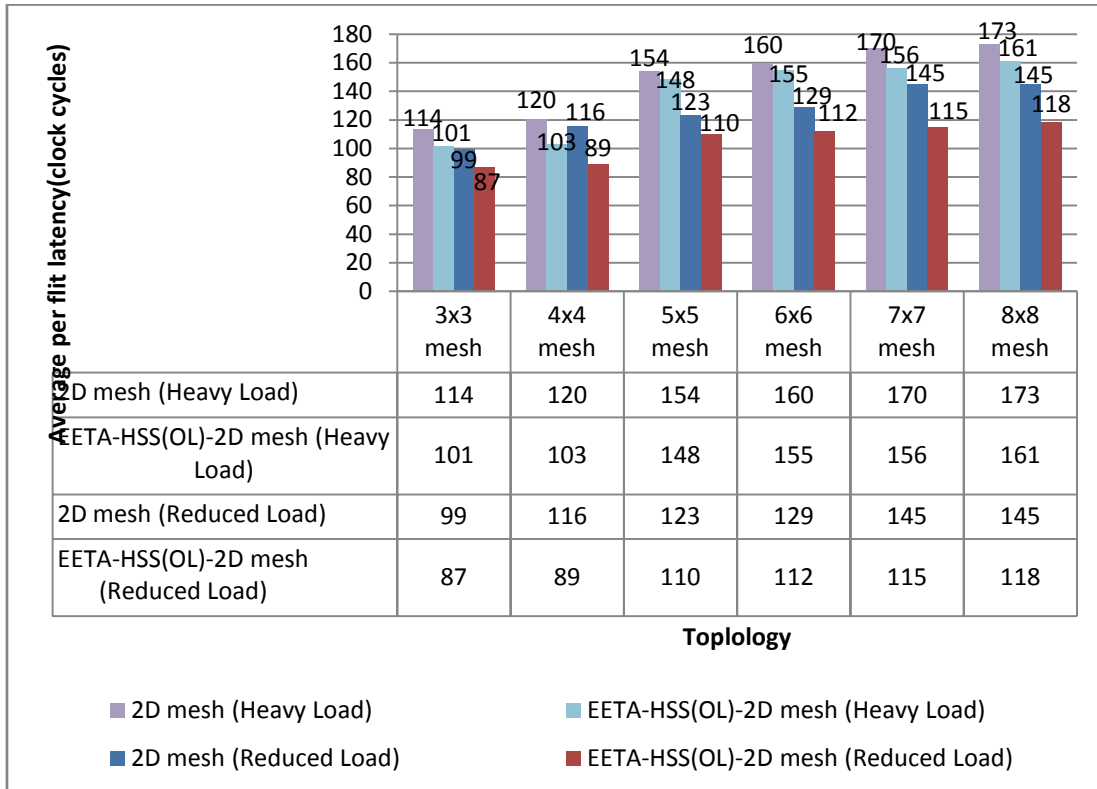


Figure IV.5 Comparison of Average per flit latency between 2D-Mesh and EETA-HSS (OL)-2D mesh (no. of Optimized Links = 80% of application characteristics) at reduced and heavy load for various topologies.

The average latency per flit get reduced on an average of 16.35% and 7.88% at reduced and heavy load respectively in EETA-HSS (OL) topology as compared to the regular 2D-Mesh.

The same scenario can be seen for the average latency per flit in Figure IV.6 and Figure IV.8. The reduction in the average per flit latency at reduced load and at heavy load is continuous till EETA-HSS 3D-Mesh (80%), after that it remains unchanged. Therefore, the maximum number of shortcut links added in 2D-Mesh is kept equals to the 80% of the application/traffic characteristics to get the maximum benefits in average total energy per flit and average per flit latency.

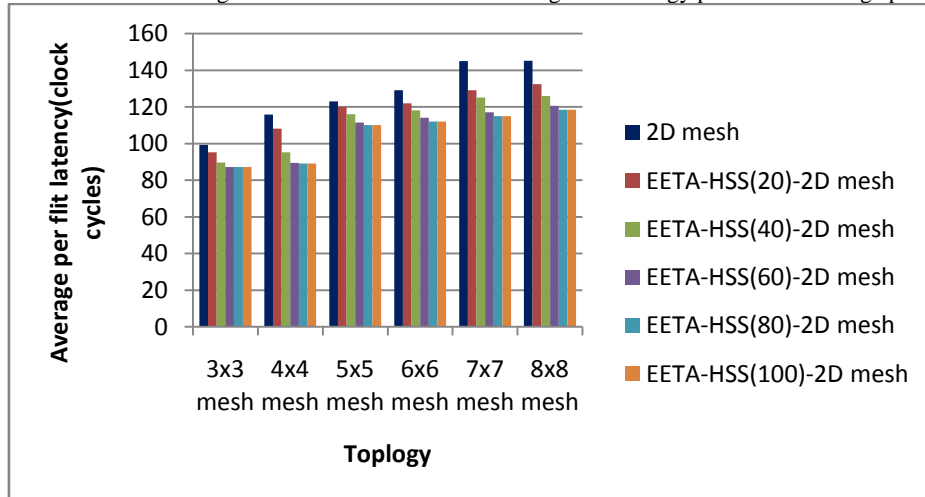


Figure IV.6 Comparison of Average per flit latency of 2D-mesh and EETA-HSS 2D-mesh at reduced load for various topologies.

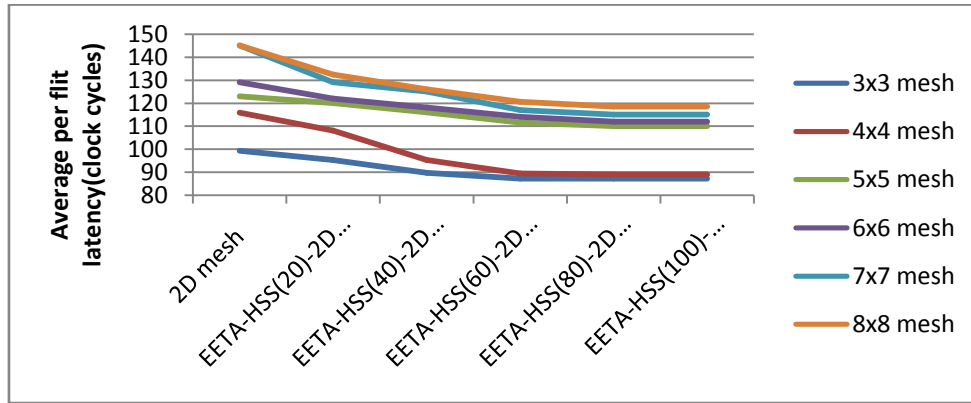


Figure IV.7 Comparison of Average per flit latency of 2D-mesh and EETA-HSS 2D-mesh at reduced load for various topologies (line graph).

This is also depicted by the line graph as shown in Figure IV.7 and Figure IV.9, which demonstrate comparison in Average per flit latency of 2D-mesh and EETA-HSS 2D-mesh at reduced load and heavy load for various topologies using the line graph. It can be seen that the lines for all the 2D topologies are getting straight after EETA-HSS (80%)-3D mesh, after that the average per flit latency remains unchanged.

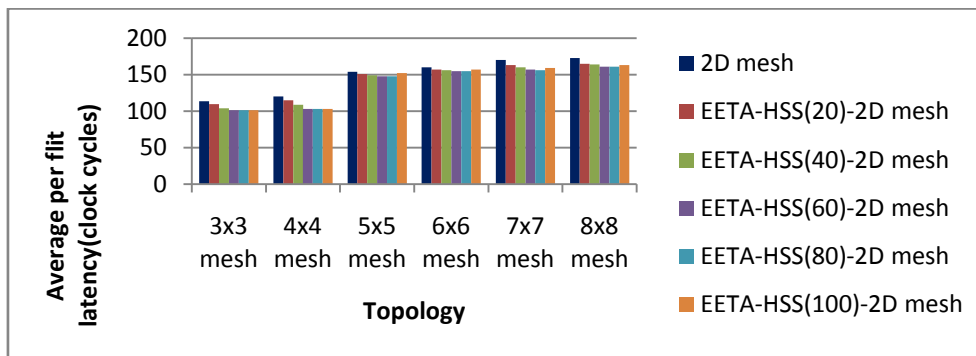


Figure IV.8 Comparison of Average per flit latency of 2D-mesh and EETA-HSS 2D-mesh at heavy load for various topologies.

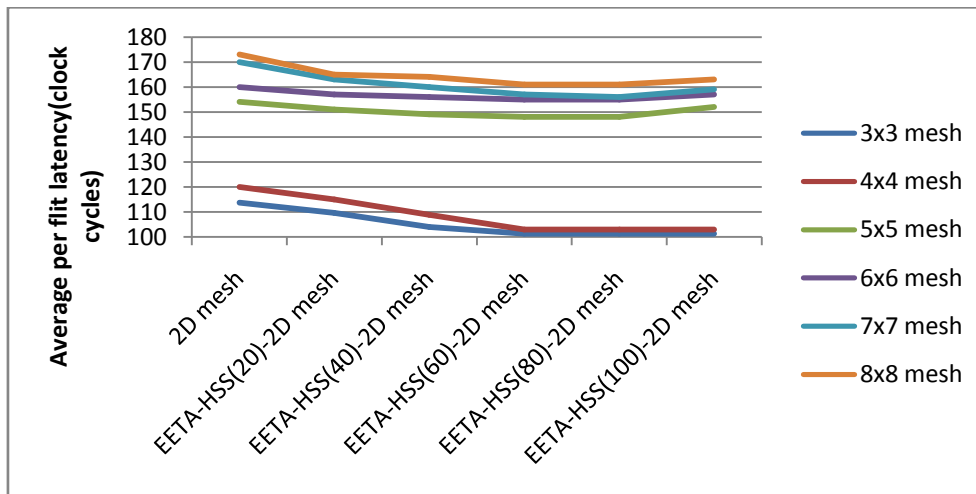


Figure IV.9 Comparison of Average per flit latency of 2D-mesh and EETA-HSS 2D-mesh at heavy load for various topologies (line graph).

Due to the insertion of shortcut links for same source destination pair the data packet has to traverse less number of hops to reach its destination leading to less switch arbitration in EETA-HSS 2D-Mesh and also as the packets needs to traverse less number of routers between source and destination leading to reduced switching delay, which therefore results in reduced dynamic communication energy and average latency as compared to the regular 2D-Mesh NoC topology architecture.

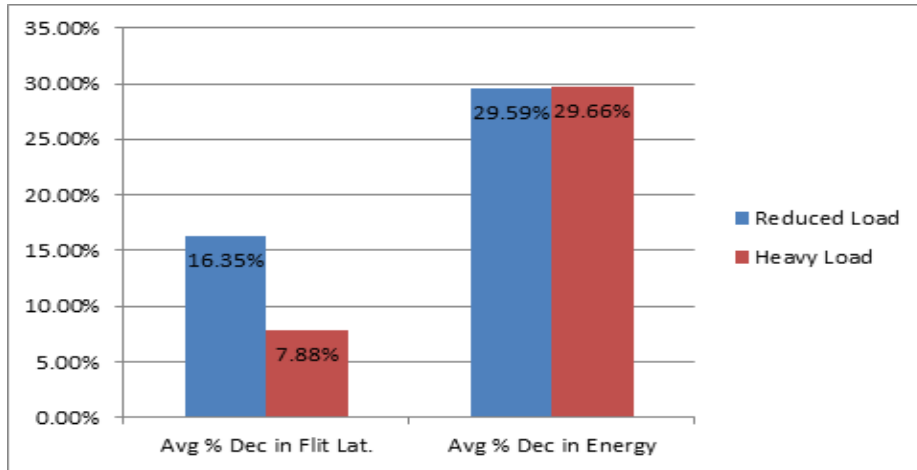


Figure IV.10 Comparison in terms of percentage reduction of Average latency per flit and Average total energy per flit at reduced and at heavy load in 2D-mesh and EETA-HSS 2D-mesh.

As shown in Figure IV.10, the reduction in average latency per flit is about an average of 16.35% and 7.88% in EETA-HSS 2D-Mesh for low (reduced) traffic load and heavy traffic load respectively as compared to the regular 2D-Mesh. The reduction in average total energy per flit consumption is about an average of 29.59% and 29.66% in EETA-HSS 2D-Mesh for reduced traffic load and heavy traffic load respectively as compared to the regular 2D-Mesh. For both the reduced and heavy load cases we are getting marginal reduction in latency as compared to the saving in total dynamic communication energy. Moreover, the gains for latency in case of heavy load decreases substantially in comparison to the reduced load.

Due to the insertion of shortcut links for same source destination pair the data packet has to traverse less number of hops to reach its destination leading to less switch arbitration in EETA-HSS 2D-Mesh and also as the packets needs to traverse less number of routers between source and destination leading to reduced switching delay, which therefore results in reduced dynamic communication energy and average latency as compared to the regular 2D-Mesh NoC topology architecture.

V. CONCLUSION

The paper proposed an efficient energy conscious methodology to tailor the regular 2D-Mesh topology by inserting shortcut links so that the dynamic communication energy and latency can be optimized. In this paper, a greedy energy conscious topology augmentation methodology for on-chip interconnection networks is proposed. The proposed methodology uses hash based smart shortcut links to augment an energy efficient 2D-mesh and 3D-mesh topology (EETA-HSS 2D and 3D-mesh) according to the traffic characteristics. The proposed methodology is greedy but fast and uses LinkedHashMap for exploring the effective and efficient energy critical shortcuts to be augmented in the given regular architecture and to use the same routing function (deterministic routing function) even after the augmentation. Further, the proposed methodology works with practical design constraints such as link length, port constraints (port availability per tile) and the proposed design also strives to add minimum number of shortcuts so as not to compromise the original design in terms of area and synchronous communication constraints.

The experimental results demonstrate that the EETA-HSS 2D-mesh has reduced the average total energy per flit consumption on an average of 29.59% (reduced load) and 29.66% (heavy load). The proposed methodology is also able to reduce the average per flit latency by on average of 16.35% (reduced load) and 7.88% (heavy load) as compared to the regular 2D NoC architecture.

In future, the possible extension to the proposed methodology is to use some optimization whereas the methodology presented in this paper is greedy. The optimization heuristics such as genetic algorithm, simulated annealing, honey bee approach etc., which may further improve the performance in terms of energy saving at the cost of increased computational complexity and computational time.

REFERENCES

- [1]. Ali, M., Welzl, M. and Zwicknagl, M. 2008. Networks on chips: scalable interconnects for future systems on chips. 4th European Conference on Circuits and Systems for Communications, ECCSC 2008- IEEE: pp. 240-245.
- [2]. Dally, W. J. and Towles, B. 2001. Route packets, not wires: On-chip interconnection networks. In: Proceedings of the 38th Design Automation Conference (DAC), IEEE held at Las Vegas during June 18-22, 2001, pp. 684- 689.
- [3]. Dick, R. P., Rhodes, D. L., & Wolf, W. 1998. TGFF: task graphs for free. In: Proceedings of the 6th international workshop on Hardware/software co-design IEEE Computer Society, pp. 97-101.
- [4]. Fuks, H., Lawniczak, A. 1999. Performance of data networks with random links. Mathematics and Computers in Simulation 51.
- [5]. George, B. 2012. Energy consumption in networks on chip: efficiency and scaling. IEEE 15th International Symposium, pp. 250-261.

- [6]. Hu, J. and Marculescu, R. 2003. Energy-aware mapping for tile-based NoC architectures under performance constraints. In: proceeding of Conference on 8th Asia and South Pacific Design Automation, IEEE held at Kitakyushu during January 21-24, 2003, pp. 233 – 239.
- [7]. Hu, J. and Marculescu, R. 2005. Energy and performance-aware mapping for regular NoC architectures. Computer-Aided Design of Integrated Circuits and Systems, Vol. 24, pp. 551-562.
- [8]. Murali, S. and Micheli, G.D. 2004. Bandwidth-Constrained Mapping of Cores onto NoC Architectures. DATE, pp.896-901.
- [9]. Ogras, U.Y. and Marculescu, R. 2005. Application Specific Network-on-Chip Architecture Customization via Long Range Link Insertion.
- [10]. Saman Amarasinghe (2007). The future of multicore programming primer. Technical report, Massachusetts Institute of Technology.
- [11]. Vyas, K., Choudhary, N. and Singh, D. 2013. NC-G-SIM: A Parameterized Generic Simulator for 2D-Mesh, 3D-Mesh and Irregular On-chip Networks with Table-based Routing. Published in Global Journal of Computer Science and Technology Network, Web and Security Volume 13, 2013
- [12]. Ye, T., Benini, L. and Micheli, G.D. 2002. Analysis of power consumption on switch fabrics in network routers. In: Proceedings of 39th Conference on Design Automation, IEEE held at New Orleans during June 10-14, 2002, pp. 524-529.