Optical Switching and Networking 25 (2017) 100–108

Contents lists available at ScienceDirect

Optical Switching and Networking

journal homepage: www.elsevier.com/locate/osn

Online multi-application mapping in photonic Network-on-Chip with mesh topology

Akram Reza

Department of Computer Engineering, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran

ARTICLE INFO

Keywords:
Network-on-Chip
Photonic NoC
On-line mapping
Allocation
Preemptive
Contiguous allocation

ABSTRACT

The silicon photonic interconnection is one of the promising solutions with high bandwidth and low power consumption for overcoming electrical Network-on-Chip challenges. This communication infrastructure provides, to integrate hundreds processing cores on a chip, then this structure is suitable for running multi-application simultaneously. In this paper, a chain of algorithms have been proposed for online multi-application mapping in Photonic Network-on-Chip (PNoC). This chain contains Boundary allocation, Boundary migration and inter sub-mesh Loss Aware (LA) mapping for improving performance constrained parameter. Experimental results indicate a considerable improvement of combining the Boundary allocation and migration scheme, about 33.8% improvement in average execution time and 37.6% in energy deception.

1. Introduction

Network-on-Chip (NoC) has been suggested as scalable communications infrastructure with a proper performance and power consumption for multi-core chips. But with growing up the number of processing cores in a chip, the network delay and power consumption become a major challenge in electrical NoC design and network scalability will be stopped [1–3]. Due to the recent progress in the CMOS-compatible silicon nano-photonic technologies, the usage of photonic interconnection networks has become possible as a promising solution in chip-scaled interconnection networks in order to overcome the mentioned challenges [1].

By using Wavelength-Division Multiplexing (WDM) in photonic communications can increase the communication bandwidth considerably and reduce communication delays. On the other hand, the power consumption in photonic network components such as Micro-Ring Resonators (MRRs) and Waveguides are independent of data rate and distance. There are two major challenges for implementing photonic NoC. First, the impossibility of implementation photonic processing core and the second problem is the signal regeneration. For regenerating optical signal, Optical-Electrical-Optical (O-E-O) convertors are required. These types of convertors cause the additional power overhead and degrade performance of the system. Consequently, it forces system designers to pay special attention to insertion loss (power loss) and noise in the photonic signals, in order to ensure the integrity of After the should not be plural during the transmission. The insertion loss would disturb the signal integration and network scalability in the network [4]. Two approaches have been proposed for reducing insertion loss in photonic networks [5]: The first approach in architecture a new topology and router (switch) for decreasing insertion loss [5–9]. The second approach in application level tries to decrease insertion loss by suggesting new routing [10] and mapping algorithm [11–13]. In this paper, the application level approach is selected (mapping) to reduce system delay and power by reducing insertion loss, wait and response time of online multi-application mapping.

By using photonic Network-on-Chip as chip interconnection, we can integrate a large number of cores in a single chip. Therefore, multi-application can be run simultaneously on a chip, for multi-application mapping algorithm is needed. Mapping multi-application on a NoC might be accomplished by design time (offline mapping) or at run time (online mapping). In online mapping, the assignment and ordering of applications are performed while the other applications are executed. However, the computational overhead of mapping algorithm may intensify delay and energy consumption of the application at runtime. The multi-application mapping on NoC could be employed in two situations: when the number of cores in a chip is more than sum of cores required to run all applications (unbounded mapping); or when the number of cores in a chip is less than sum of cores involved to run all applications (bounded mapping), and the core allocation could be happened preemptively or non-preemptively.

To reduce complexity of online multi-application mapping, mapping has broken in two major functions application allocation and application mapping. In allocation function, determined a region for an application and mapping function determined task location in a selected reign. There are two schemes for processor allocations: contiguous (a selected region is accretion) and non-contiguous

E-mail addresses: a.ak.reza@gmail.com, a.reza@qodsiau.ac.ir.

http://dx.doi.org/10.1016/j.osn.2017.04.002
Received 22 October 2016; Received in revised form 15 April 2017; Accepted 16 April 2017
Available online 18 April 2017
1573-4277/

(a selected region is not accretion). The non-contiguous allocation, which can lead to global traffic and increases message delay. It might also increase insertion loss in photonic layer dramatically. Therefore, in this paper contiguous allocation is selected for PNoC. The main problem in contiguous allocation is the external fragmentation (when the number of free nodes is more than nodes required for an application but free nodes not adjacent, the allocation is not performed). One of the solution, is application migration in order to integrate free nodes, but this solution can be done in the preemptive allocation. In this paper for solving external fragmentation, preemptive allocation strategy with migration are selected. The online multi-application mapping in PNoC with mesh topology with BNPC allocation strategy (Bounded, preemptive and Contiguous allocation), Boundary migration and mapping algorithm have been suggested to improve system performance. To accomplish the desired mapping, some steps must be taken: firstly, finding appropriate sub-mesh size for an arrival application; secondly, finding sub-mesh location on mesh topology for online allocation application; and finally, mapping application cores in determined sub-mesh. To reduce online mapping time overhead, the second and third steps are run simultaneously.

The rest of this paper is organized as follows: in Section 2, overview of the related works in PNoC, application allocation, application migration and application mapping are given. Section 3 is dedicated to some definitions in photonic networks, allocation definitions and notations. In Section 4, the structure of the new approach for online mapping is explained. It contains the new Boundary allocation and migration algorithms to minimize external fragmentation and wait and response time, and the LA mapping to reduce insertion loss. Section 5 reports simulation results and the evaluation of improvements in physical layer. Finally, Section 6 concludes the paper.

2. Related work

Application mapping problem is an NP-hard problem [14]. Mapping can be divided in two distinct sets, online and offline mapping [15]. In online mapping, find the performance bottleneck and load balance traffic among the network, are as main objectives. However, the computational overhead of the online mapping can increase delay and energy consumption of application.

2.1. Photonic Network-on-Chip

In recent years, most studies in photonic on-chip networks, have followed two general models: wavelength-selective routing and space routing. In the wavelength-selective routing, the transmission of photonic signals is controlled by MRRs which act as selective filters and determine the path signals taken along the network by means of their wavelength. In this method, each source node can address its destination by selecting a unique wavelength, and afterwards, the photonic signal is forwarded on the network through MRR filters. One of the advantages of such model is its low transfer latency and using wavelength-based routing, which can lead to a transfer speed close to the speed of light. However, in this method, some of the wavelengths are wasted due to signal propagation controlling, and thus, they cannot be exploited for data transfer with WDM. Therefore, the communication bandwidth performance may be decreased [5,16–20].

In space routing method, electro-optic broadband MRRs are used and WDM signals of multiple wavelength channels are forwarded in parallel along the path. In this method, all photonic paths should be established beforehand by means of circuit switching mechanism. Afterwards, the control is administrated by an electrical network with a low bandwidth and admissible power consumption. The advantage of such method is providing the possibility to exploit the whole optical spectrum by utilizing WDM mechanism, thus leading to high-bandwidth communications. However, the circuit-switching protocol can impose additional overhead and in general, it can increase network latency [3,5,21].

In this paper, space routing is employed in order to provide high-bandwidth communication links and achieve better network scalability. Photonic interconnection networks are entirely different in quiddity with electrical networks. Therefore, it is essential for designers to study both the concept and functionality of different photonic network components, such as waveguide [4], photonic switch element (PSE) [1,16], filters [16], waveguide crossing (crossing, for short) [22], waveguide bending (bending, for short) [4], modulators [23], and detectors [24]. Waveguide, an important element in photonic networks, is responsible for transferring photons.

2.2. Application allocation

The application allocation strategy could be categorized in two sets: contiguous allocation, and non-contiguous allocation. In contiguous allocation, the application is only allocated in set of adjacent nodes to reduce communication overhead [13,25,26]. External fragmentation is the major tribulation factor of this approach. It occurs when some allocation and de-allocation is executed [27]. There is no resolution in preemptive allocation although some methods suggest migration as an appropriate solution [27]. Contiguous processor allocation schemes also include a wide range of procedures such as stack-based allocation [28], adjacency allocation [29], adaptive scan allocation [9], quick allocation [25], best/first fit allocation [11], and Row allocation [13]. In non-contiguous allocation technique [17–19], applications can be allocated in non-adjacent nodes in a network.

2.3. Application migration

Application migration obstructions have been also contemplated extensively in review of literature [20–30]. Various types of migration algorithms are employed in application allocation, that include methods of managing external fragmentation, traffic distribution, load balancing [31–33] and fault tolerance [34–36] or to solving external fragmentation in application contiguous allocation [13,27].

In [27], online dynamic compaction single corner (ODC-SC) and online dynamic compaction four corner (ODC-FC) have been introduced for application migrations, in a two-dimensional mesh topology to prevent external fragmentation. The ODC-SC technique attempts to detect the destination of sub-meshes to direct a sub-mesh in such a way to obtain larger free fragment of processors. It also ensures that the transferred active application is not in interference with the other active applications, in a mesh. ODC-FC is optimized version of ODC-SC that contributes a larger region of adjacent free nodes, by moving the applications more selectively.

The Row migration has been introduced for application migrations [13], when external fragmentation occurred, the selected sub-mesh migration into lower row to integrate free node in upper row.

2.4. Mapping

The mapping algorithm application on Network-on-Chip is classified into two different categories: single mapping application [37–41] and multiple mapping application [42,43]. The Nmap algorithm is one of the well-known single application mapping approach [41].

Some researchers have recommended online multiple application mapping. In [38,44], authors have presented heuristics dynamic application mapping with an initial application mapping phase, followed by a dynamic mapping phase. The dynamic mapping phase might be applied in all procedures. It could be employed in first free (FF), nearest neighbor (NN), minimum maximum channel load (MMC), minimum average channel load (MAC), and path load (PL). In case of FF technique, the NoC selects the first free node which can execute requested application while the network is searched by column with lowest numbers. The NN mapping method is similar to FF. The only difference is that the requested application is located on the free