INTEGRATION the VLSI journal 60 (2018) 167–189

Contents lists available at ScienceDirect

INTEGRATION, the VLSI journal

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

# Area Constrained Performance Optimized ASNoC Synthesis with Thermal-aware White Space Allocation and Redistribution

Priyajit Mukherjee*, Sandeep D'souza, Santanu Chattopadhyay

*Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, WB 721302, India*

ARTICLE INFO

*Keywords:*
Application-Specific Network-on-Chip
Particle Swarm Optimization
Simulated Annealing
Mixed Integer Linear Programming
Communication Cost
Thermal-aware Design

ABSTRACT

Application-Specific Network-on-Chip (ASNoC) has emerged as a more efficient design alternative to the regular Network-on-Chip (NoC) topologies, which can better suit the communication requirements of an application. In this work, we have proposed a three step ASNoC synthesis procedure that targets optimization of the three major design parameters – area of the chip, communication cost (*CC*) of the resulting on-chip network and the peak temperature of the chip ($T_{peak}$). Inputs to the procedure are the core dimensions, corresponding average power consumptions and the core graph of the application. In the first step, an area-optimized floorplanning of the cores has been conducted. Considering an overhead constraint on the minimum area generated in this step, a communication cost aware floorplan has been generated in the second step. Finally, considering an overhead constraint on the minimum *CC* obtained from the second step, a thermal-aware white space allocation and redistribution (WSA & R) has been carried out. At each step, Mixed Integer Linear Programming (MILP) formulation has been made along with heuristic or meta-heuristic methods. Core placement has been performed using a Simulated Annealing (SA) based technique with its initial solution generated using a Particle Swarm Optimization (PSO) based approach. A router placement heuristic has been proposed that considers core rotation and core mirroring techniques to optimally place the primary routers. WSA & R has been performed using an iterative SA-based approach. At each stage of the design flow, we have compared our proposed method with an MILP based method and other contemporary heuristics. Significant improvements have been achieved in both *CC* and $T_{peak}$. The proposed WSR method has been able to reduce $T_{peak}$ up to 22°, considering 35% overhead on *CC*.

## 1. Introduction

With continuous technology scaling, Multi-Processor Systems-on-Chip (MPSoCs) encounter an ever-increasing inter-processor communication requirement. Thus, in the Deep Sub-micron (DSM) era, with the clock frequency in tens of GHz [1], performance of an MPSoC becomes significantly dependent on its communication infrastructure. With increase in the number of cores per chip, on-chip bus-based communication fails to support the increasing communication volume [2,3]. Network-on-Chip (NoC) has appeared as a potential scalable interconnect solution to handle this high bandwidth requirement. In NoC, the raw data to be sent between two cores is converted into a set of data packets. Such packets from different cores can be transmitted simultaneously through different routers and links present in the NoC. Thus, parallelism and pipelining techniques can be used while transmitting the packets. NoC topologies can be broadly classified into two categories - regular topologies following regular interconnection structure between routers (for example, Mesh, Torus, Ring, Tree etc.) and application-specific or custom topologies, known as Application-specific Networks-on-Chip (ASNoCs). Cores are attached to individual routers in the network. Routers are interconnected via links.

Regular NoC topologies have fixed positions of routers and links in the chip-floorplan, core sizes are also often considered to be regular. On the other-hand, Application-Specific Network-on-Chip (ASNoC) design involves performance-aware placement of the cores, routers and links. Cores may also be irregular in size. Being specific to an application, ASNoC has the potential to outperform regular NoC in terms of network performance, power consumption and area requirement [4–7]. Any ASNoC synthesis procedure includes the following two steps - placement of the cores, placement of the routers and links. While generating the core floorplan, some of the previous works [8,9] have considered the cores to be soft - the IP cores are available in the form of synthesizable RTLs or gate level net-lists. For soft cores, area of the core layout is taken to be fixed, while the width and height of the core may vary within a pre-specified range of its aspect ratio. Also, many of the previous works [5,10,7] have assumed hard cores - IP

* Corresponding author.
*E-mail addresses:* priyajit@iitkgp.ac.in (P. Mukherjee), sandeepdsouza93@gmail.com (S. D'souza), santanu@ece.iitkgp.ernet.in (S. Chattopadhyay).

http://dx.doi.org/10.1016/j.vlsi.2017.09.004
Received 27 February 2017; Received in revised form 14 May 2017; Accepted 8 September 2017
Available online 25 October 2017

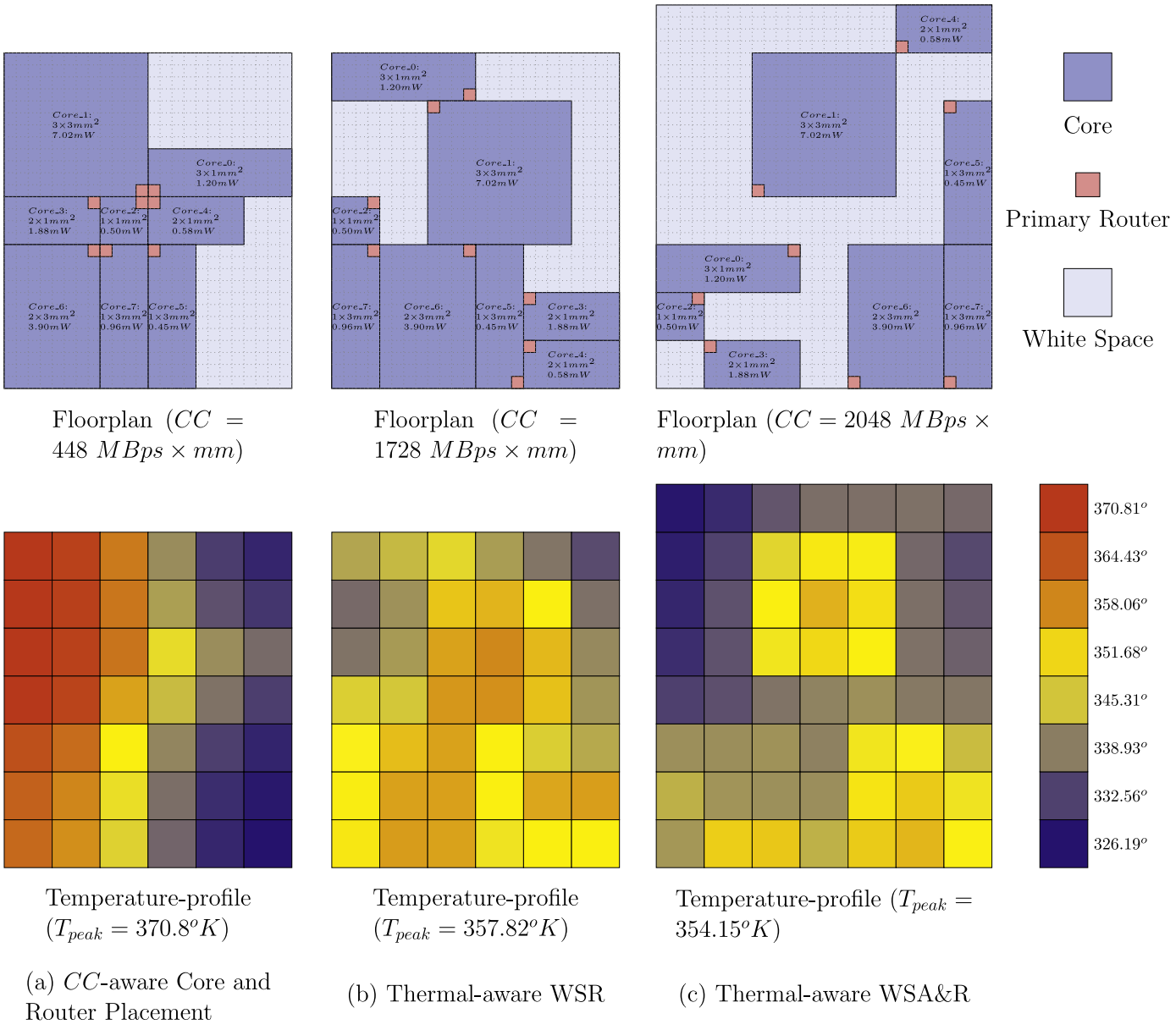


**Fig. 1.** Peak temperature reduction with thermal-aware WSR and WSA & R for the benchmark application PIP.

cores being available in GDSII file format only. In contrast to soft cores, a hard core has fixed width and height, as its transistor layout has already been generated. Due to the availability of the transistor layout of hard cores, it can be assumed that the network interface (NI) circuit of the core is also present inside the core layout (if the core is designed to be used in an NoC). As the *Primary Router* (routers which are directly connected to the cores) is required to be placed near the network interface unit of the core, it can also be assumed that the core layout contains the positions of the *Primary Routers* in the floorplan. This is feasible due to the fact that the area of a router is reasonably small compared to the area of a core [8,5,7]. In this work, we have considered the cores to be hard cores, also containing positions of the *Primary Routers* at one of the corner positions of the core layout. While placing the cores, rotation of the core layout has been supported to reach closer to the design objective. For the same reason, while placing the routers, positions of the primary routers have been altered by performing rotation and mirroring [11] of the core floorplan. After fixing the positions of the *Primary Routers*, link placement has been performed. In the process, two new types of routers may get introduced. *Junction Routers* have been placed at the intersection points of two links, while *Secondary Routers* (or repeaters) have been placed to maintain the maximum link length constraint. As the sizes of the network components are very less compared to the size of a core [8,5,7], the boundary of each core layout is considered to be extended to accommodate the positions of the junction and secondary routers and the links. Before describing the motivation and objectives of the proposed ASNoC synthesis method, some important terminologies related to this work are described next.

**Core Graph** ($CG$) [12]: Core Graph $CG(C, E)$ of an application is a directed acyclic graph where each vertex $c_i \in C$, represents a core of the application and each edge $e_i \in E$, signifies a communicating edge between two cores in it. Each edge $e_i$ has the following attributes - $src_i$, denoting the source core, $dst_i$, the destination core and $BW_i$, the communication bandwidth requirement of the edge.

**Communication Cost** ($CC$) [12]: Communication Cost is a metric to predict the network performance (latency and power consumption). It is the summation of the products of the bandwidth requirement of each edge in the Core Graph and the corresponding rectilinear distance between the