Physical Planning for the Architectural Exploration of Large-Scale Chip Multiprocessors

Javier de San Pedro, Nikita Nikitin, Jordi Cortadella and Jordi Petit
Universitat Politècnica de Catalunya

***Abstract*—This paper presents an integrated flow for architectural exploration and physical planning of large-scale hierarchical tiled CMPs. Classical floorplanning and wire planning techniques have been adapted to incorporate layout constraints that enforce regularity in the interconnect networks. Routing is performed on top of memories and components that underutilize the available metal layers for interconnectivity. The experiments demonstrate the impact of physical parameters in the selection of the most efficient architectures. Thus, the integrated flow contributes to deliver physically-viable architectures and simplify the complex design closure of large-scale CMPs.**

I. Introduction

Chip multiprocessing is rapidly evolving and increasing the power-performance efficiency of computing systems by exploiting the parallelism inherent in applications. Tile replication [1] and hierarchical CMP organizations [2] are two widely adopted methodologies for designing many-core CMPs under time-to-market pressure. Figure 1 shows an example of a hierarchical tiled CMP. The chip consists of identical tiles (*clusters*) organized in a mesh. Every cluster includes a local interconnect (bus or ring), cores (C) with private caches (L2), a shared cache (L3) and a router (R) that connects the cluster to the top-level mesh.

The problem of *architectural exploration* consists of automatically finding values for all of the system-level design parameters (number of clusters, interconnect topology, number of cores per cluster, amount of L2 and L3, etc.) to maximize chip performance under design constraints, typically area [3], [4]. A trade-off exists between the area dedicated to cores, caches and interconnect, depending on the CMP workload.

While various works exist that explore system-level parameters, significantly less research has been carried out in the field of *physical planning* for CMPs [5], [6]. Physical aspects have a significant impact on chip area and routability and may introduce important deviations in the area estimations performed during architectural exploration. Both the tendecy towards wire links and the use of *over-the-cell routing* (in which interconnect wires are routed on top of memory components) increase these deviations. In this work, we propose a architectural exploration flow that incorporates floorplanning and wire planning for more accurate area and physical viability estimation. Our proposal takes into account the use of over-the-cell routing and hierarchical tiled CMPs.

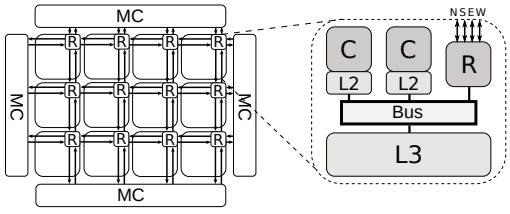

Fig. 1. Architecture of a hierarchical tiled CMP.

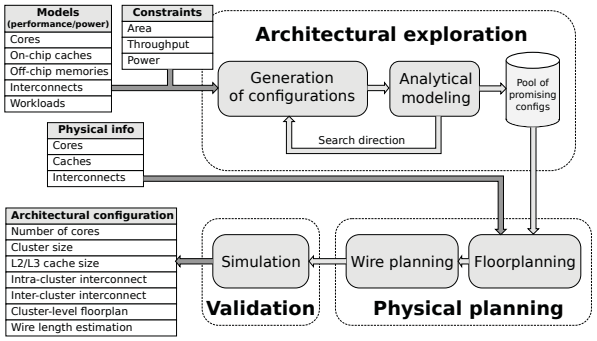

Fig. 2. The flow for architectural exploration of CMPs.

II. Exploration flow

Figure 2 presents the interface and steps of the exploration flow. The input data consists of a set of *models* and *constraints*. Examples of the models that can be used are presented in [4]. Different models of cores can be considered with various implementation aspects. The models for on-chip and off-chip memories define area and access latency. The workload is characterized by defining the memory access patterns and expected throughput for each core type. For physical planning, the aspect ratios and the number of metal layers used by each component are defined.

The output of the exploration flow is a system-level description of a configuration (or a set of configurations). This description includes all system-level parameters, a cluster-level floorplan and a interconnect wire plan.

A. Architectural exploration

The objective of architectural exploration is to efficiently reduce the design space and generate a moderate number of candidate configurations with the highest performance satisfying the area and power constraints. The efficiency in exploration is achieved by a heuristic search (both *Simulated Annealing* and *Extremal Optimization* are used) that uses an analytical model [4] for performance evaluation. This analytical model captures the contention component in the interconnect. Power consumption is also estimated using analytical models.

978-1-4673-6493-5/13/$31.00 ©2013 IEEE

B. Physical planning

This stage is the main contribution of this work. The objective of this stage is to obtain accurate estimations of chip area for all candidate configurations obtained in the previous stage, considering the physical parameters of the components, e.g. aspect ratios and number of metal layers. As it can be seen from Fig. 1, the regularity of the top-level mesh reduces the floorplanning problem from chip to cluster level.

To solve the cluster-level floorplanning problem, a heuristic search (Simulated Annealing) is also used. Only *slicing floorplans* are considered in order to reduce the search space. Bounding curves are used to efficiently model components with variable aspect ratios [7]. In the cost function, we estimate wire length using Lee's algorithm [8] instead of half-perimeter wire length to filter unroutable floorplans before trying to perform full wire planning.

In order to allow the *abutment* of different clusters without any wiring overhead in the connections, a symmetry constraint is introduced during wire planning for the global interconnect wires. Additional constraints are also defined to enforce the *adjacency* between cores and their private caches, thus avoiding performance penalties caused by accessing distant caches. Wire planning is performed only at the link level, and not for the individual wires of each link. We create a 3D grid where each horizontal plane represents a metal layer of the cluster, and the cell size is equivalent to the minimum width of a link. Blockages are then inserted on this grid according to the metal layers occupied by each CMP component. A SAT solver is then used to find the required routes according to the architectural configuration and the constraints.

For every promising candidate obtained during the architectural planning, a set of the floorplans with a minimized combination of area and wire length is generated. Afterwards, wire planning is performed to verify the routability of every floorplan and generate a more accurate estimation of the wire length. Only the best routable floorplans are selected for performance validation.

C. Validation

The validation stage uses simulation to improve the performance estimations of the selected candidates. In our implementation, we used a modified version of BookSim.

III. Experimental results

In this experiment, we show the impact of physical planning on architectural exploration. An exploration was performed with constraints on total chip area (350 mm^2) and power (200 W). The area and performance models for cores were obtained by scaling publicly available Intel Core 2 Duo data. For the workload model, we characterized two SPEC2006 benchmarks (*namd* and *soplex*) as described in [4].

In Fig. 3, we plot the configurations with the highest throughput. The x-axis indicates the estimated performance of each configuration. The y-axis shows three approximations of the chip area for every configuration: *block area* is the sum of areas from the individual components; *best floorplan*

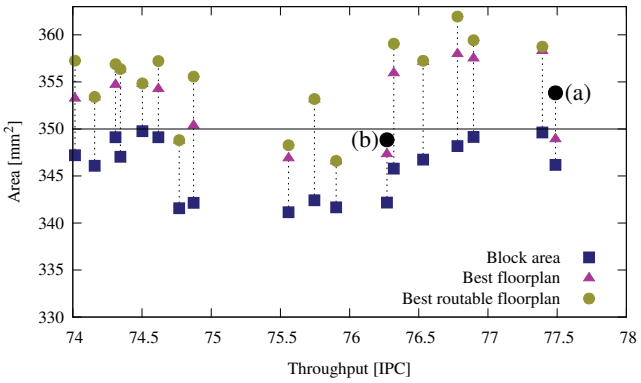

Fig. 3. Comparison of area estimations with and without physical planning

indicates the minimum possible floorplan area, even if it has routability problems; *best routable floorplan* shows the area for the smallest *routable* floorplan.

In order to satisfy the area requirement, the designer will not be able to select the best configuration *(a)* but rather *(b)*, with a 1.6% throughput loss. Even though area and throughput are similar, the architectural parameters are rather distinct. This result gives an insight for the designer to consider an alternate group of architectures with improved physical parameters. Figure 4 shows *(b)* after the full exploration flow.

Acknowledgments

This research has been funded by a gift from Intel Corp., project CICYT TIN2007-66523, and FPI grant BES-2008-004612.

References

[1] J. Balfour and W. J. Dally, "Design tradeoffs for tiled CMP on-chip networks," in *Proc. Intl. Conf. on Supercomputing*, 2006, pp. 187–198.

[2] R. Das, S. Eachempati, A. Mishra, V. Narayanan, and C. Das, "Design and evaluation of a hierarchical on-chip interconnect for next-generation CMPs," in *High Performance Comp. Arch.*, Feb. 2009, pp. 175–186.

[3] T. Oh, H. Lee, K. Lee, and S. Cho, "An analytical model to study optimal area breakdown between cores and caches in a chip multiprocessor," in *ISVLSI*, may 2009, pp. 181–186.

[4] N. Nikitin, J. de San Pedro, J. Carmona, and J. Cortadella, "Analytical performance modeling of hierarchical interconnect fabrics," in *International Symposium on Networks-on-Chip*, May 2012, pp. 107–114.

[5] T. T. Ye and G. D. Micheli, "Physical planning for on-chip multiprocessor networks and switch fabrics," in *Int. Conf. Application-Specific Systems, Architectures, and Processors*, 2003, pp. 97–107.

[6] D. Bertozzi, A. Jalabert, S. Murali, R. Tamhankar, S. Stergiou, L. Benini, and G. De Micheli, "NoC synthesis flow for customized domain specific multiprocessor systems-on-chip," *IEEE Transactions on Parallel and Distributed Systems*, vol. 16, no. 2, pp. 113–129, Feb. 2005.

[7] D. F. Wong and C. L. Liu, "A new algorithm for floorplan design," in *Proc. ACM/IEEE Design Automation Conference*, 1986, pp. 101–107.

[8] F. Rubin, "The Lee path connection algorithm," *IEEE Trans. Comput.*, vol. 23, no. 9, pp. 907–914, Sep. 1974.

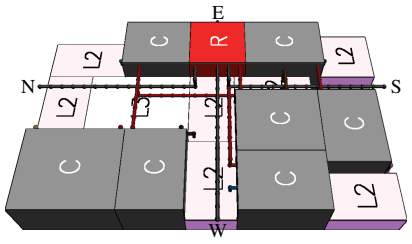

Fig. 4. Final floorplan with wire planning information