/
Article / Book Information

( )	
Title(English)	Hardware-Accelerated Modeling of Large-Scale Networks-on-Chip
( )	Chu Van Thiem
Author(English)	Thiem Van Chu
( )	: ( ), : 10994 , :2018 9 20 , : , : , , , ,
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number: 10994 , Conferred date:2018/9/20, Degree Type:Course doctor, Examiner:,,,,
( )	
Category(English)	Doctoral Thesis
( )	
Type(English)	Summary

論 文 要 旨

THESIS SUMMARY

専攻： Department of 　計算工学　専攻

学生氏名： Student's Name 　Chu Van Thiem

申請学位（専攻分野）： Academic Degree Requested 　博士 Doctor of （ 工学 ）

指導教員（主）： Academic Supervisor(main) 　吉瀬 謙二

指導教員（副）： Academic Supervisor(sub)

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

　　Networks-on-Chip (NoCs) have become integral parts of different types of computing hardware platforms including many-core processors, multiprocessor systems-on-chip, and many specialized accelerators for critically essential applications such as deep neural networks and databases. In such a hardware platform, the NoC is responsible for connecting the other components together. As the number of components that need to be interconnected increases, the overall performance becomes highly sensitive to the NoC performance. Therefore, research and development of large-scale NoCs play a key role in designing future large-scale architectures with hundreds to thousands of components.

　　A major obstacle to research and development of large-scale NoCs is the lack of fast modeling methodologies that can provide a high degree of accuracy. Analytical models are extremely fast but may incur significant inaccuracy in many cases. Thus, NoC designers often rely on simulation to test their ideas and make design decisions. Unfortunately, while being much more accurate than analytical modeling, conventional software simulators are too slow to simulate large-scale NoCs in a reasonable time. To accelerate the simulation speed, there have been some attempts to build NoC emulators using Field-Programmable Gate Arrays (FPGAs), reconfigurable hardware devices that can be reprogrammed. However, these FPGA-based NoC emulators suffer from the scalability problem. They cannot scale to large NoCs due to the FPGA logic and memory constraints. This dissertation proposes methods to address this problem, thereby enabling fast and accurate modeling of large-scale NoCs with up to thousands of nodes.

　　To overcome the FPGA logic constraints, the dissertation proposes a novel use of time-division multiplexing (TDM) where the emulation cycle is decoupled from the FPGA cycle and a network is emulated by time-multiplexing a small number of nodes. In this way, large-scale NoCs with up to thousands of nodes can be emulated using a single FPGA. The dissertation focuses on applying the TDM technique to two commonly used network topologies, two-dimensional (2D) mesh and fat-tree (k-ary n-tree), which are the bases of almost all actually constructed network topologies. It thus can be expected that the proposed methods can be extended for a wide range of networks.

　　While the TDM methods enable the emulation of large-scale NoCs, they alone are not sufficient. To achieve a high emulation speed, it is essential to address the memory constraints caused by modeling traffic workloads. There are two types of workloads used in NoC emulation: synthetic workloads and trace-driven workloads. Synthetic workloads are based on mathematical modeling of common traffic patterns in real applications. On the other hand, trace-driven workloads are based on trace data captured from either a working system or an execution-driven simulation/emulation. Currently, due to the lack of trace data of large-scale NoC-based systems, using synthetic workloads is practically the only feasible approach for emulating large-scale NoCs with thousands of nodes.

　　To overcome the memory constraints caused by modeling synthetic workloads, the dissertation proposes a method to reduce the amount of required memory so that it is not necessary to use off-chip memory even when emulating NoCs with thousands of nodes. This method not only makes the overall design much simpler but also significantly contributes to the improvement of emulation speed since accessing on-chip memory is much faster than off-chip memory. It and the proposed TDM methods enable a NoC emulator which can be used to model a mesh-based NoC with 16,384 nodes (128×128 NoC) and a fat-tree-based NoC with 6,144 switch nodes and 4,096 terminal nodes (4-ary 6-tree NoC) and is up to three orders of magnitude faster than BookSim, a widely used software-based NoC simulator, while providing the same results.

　　The dissertation shows the usability of the proposed NoC emulator by designing and modeling an effective routing algorithm for 2D mesh NoCs and evaluating it for various network sizes, from 8×8 to 128×64.

　　While synthetic workloads can provide a relatively thorough coverage of the characteristics of the emulated NoCs, evaluation on trace-driven workloads is still required in some cases such as assessing some application-specific optimizations. The dissertation takes this into account and extends the proposed NoC emulator to support trace-driven emulation which will be useful for research and development of large-scale NoCs in the future when trace data of large-scale NoC-based systems are available. Since trace data are large, they must be stored in off-chip memory. The dissertation proposes an effective trace data loading architecture and some methods to hide the off-chip memory access time and improve the scalability of the emulation architecture in terms of operating frequency and logic requirements. The evaluation results show that the extended NoC emulator is two orders of magnitude faster than BookSim when emulating an 8×8 NoC with the widely used PARSEC traces while also providing the same results, and the speedup is increased to three orders of magnitude when emulating a 64×64 NoC with trace data created based on a synthetic workload.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ（T2R2）にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).