早稲田大学大学院 基幹理工学研究科

博 士 論 文 概 要

論 文 題 目

Floorplan-Aware High-Level Synthesis Algorithms
and their Acceleration by Ising Computations
フロアプラン指向高位合成手法と
イジング計算機応用に関する研究

申 請 者

Kotaro	TERADA
寺田	晃太朗

情報理工・情報通信専攻　情報システム設計研究

2017 年 12 月

System LSI or system-on-a-chip (SoC) is an essential device in today's highly sophisticated information society. System LSI is already used widely in the world such as personal computers, servers, mobile phones, automobiles, and consumer electronics. In the Internet of thing (IoT) era, every "things" about to become informatization and systematized. Many SoCs with higher information processing are expected to be more and more needed. In these kinds of system designs, it is required to design a high-performance and small area of a system LSI with a lower cost and a shorter time period.

Based on the background above, high-level synthesis (HLS) has played a significant and essential role as a promising electronic design automation (EDA) technique nowadays. HLS synthesizes abstract application descriptions such as C, C++, Java, and Python into register transfer level (RTL) descriptions. Since we can begin a system design with a higher level of abstraction, it is expected to reduce a cost of design such as design time (Time-to-Market) and design errors.

The main input to HLS is a behavioral description, and HLS outputs FSMD (Finite State Machine with Datapath) which is composed of a datapath and a controller (FSM). An input behavioral description is represented as a control data flow graph (CDFG) where nodes correspond to operations and edges correspond to data-flows. An HLS algorithm typically solves (1) Scheduling problem: assigns every operation in the CDFG to a control step, (2) Allocation problem: determines the number and the type of functional units (FUs) to be used, (3) Binding problem: assigns operations and variables to FUs and registers, respectively.

Many HLS methods and tools have been proposed so far including ones for academic purposes and ones for commercial purposes. Field-programmable gate array (FPGA) comes into fashion nowadays. Not only commercial designers and experts but any users can use FPGA easily and cheaply. HLS becomes more and more essential.

However, in the deep-submicron era, interconnection delays are not negligible even in high-level synthesis. Distributed-register and -controller architectures (DR architectures) have been proposed to cope with this problem. RDR architecture is one of them.

Taking account the interconnection delays in HLS makes the optimizations difficult since we have to deal with the delays between modules (such as FUs and registers). In the first half of this dissertation (Chapter 2, 3, and 4), we propose performance-driven floorplan-aware HLS algorithms targeting RDR architectures. The objective is to minimize the latency (or maximize the performance).

Even if we optimize the system design through HLS, the floorplanning depends on the metaheuristic algorithm, typically simulated annealing (SA). Thus, the floorplanning remains the bottleneck in both speed and scalability.

Floorplanning of modules has been a significant role in VLSI design automation and it can be formulated as the "Rectangle Packing Problem." Novel physical Ising model-based computers (or annealing machines), which are the type of a non-von Neumann computer, have been recently studied and expected to solve combinatorial optimization problems efficiently. CMOS annealing machine is one of them. In those studies, a combinatorial optimization problem is mapped onto a theoretical magnetic model in statistical mechanics called Ising model.

No. 1

Ising model-based computers search the ground-state of the Ising model, which corresponds to the optimal solution of the original combinatorial optimization problem.

To implement an Ising model to physical annealing machines, we need to convert an Ising model with an arbitrary topology into an Ising model with the physical topology. This flow is called embedding. If we achieve the efficient embedding algorithm, we can increase the problem size implemented onto the annealing machines and may have better results. Embedding the Ising model to the physical models plays an important role in the programming flow of annealing machines.

In this dissertation, we propose three floorplan-aware performance-driven HLS algorithms targeting RDR architectures to cope with the increasing interconnection delays. We also propose an embedding method of fully-connected Ising model onto 20k-spin CMOS annealing machines followed by proposing a mapping of the floorplan problem for Ising model-based computers to deal with the bottleneck in the floorplanning.

Chapter 1 [Introduction] describes the research backgrounds.

Chapter 2 [A Floorplan-Aware High-Level Synthesis Algorithm with Operation Chainings Based on Inter-Island Distance] proposes a high-level synthesis algorithm using operation chainings which reduces the overall latency targeting RDR architectures. The proposed algorithm consists of three steps: The first step enumerates candidate operations for chaining. The second step introduces maximal chaining distance (MCD), which gives the maximal allowable inter-island distance on RDR architecture between chaining candidate operations. The last step performs list-scheduling and binding simultaneously based on the results of the two preceding steps. The proposed algorithm enumerates feasible chaining candidates and selects the best ones for RDR architecture. Experimental results show that our proposed algorithm reduces the latency by up to 40.0% compared to the original approach, and by up to 25.0% compared to a conventional approach. The proposed algorithm also reduces the number of registers and the number of multiplexers compared to the conventional approaches in some cases.

Chapter 3 [A Floorplan-Aware High-Level Synthesis Algorithm with Multiple-Operation Chainings Based on Path Enumeration] proposes a floorplan-driven high-level synthesis algorithm using multiple-operation chainings composed of two or more operations, and reduce the overall latency targeting RDR architecture. The proposed algorithm enumerates multiple-operation-chaining path candidates before performing scheduling/binding. Based on them, we find out optimal ones taking into account RDR floorplan information. Experimental results show that our algorithm reduces the latency by up to 30.4% compared to a conventional approach, and reduces the latency by up to 24.4% compared to the algorithm proposed in Chapter 2, but is only effective to the limited benchmarks.

Chapter 4 [A Floorplan-Driven Bitwidth-Aware High-Level Synthesis Algorithm Using Operation Chainings] proposes a bitwidth-aware high-level synthesis algorithm using operation chainings targeting RDR architectures. Our proposed algorithm optimizes bitwidths of functional units and utilizes the vacant islands by adding some extra functional units to realize effective operation chainings to generate high-performance circuits without increasing the total area. Experimental results show that our proposed algorithm reduces the latency by

up to 47% compared to the algorithm proposed in Chapter 2 without area overheads by eliminating unnecessary bitwidths and adding efficient extra FUs for RDR architectures.

These algorithms above successfully reduce the latency compared to the conventional algorithms coping with the increasing interconnection delays. However, the SA-based floorplanning remains the bottleneck in both speed and scalability. To deal with the bottleneck above, we try to apply the floorplanning problem to the forthcoming Ising model-based computers.

Chapter 5 [A Fully-Connected Ising Model Embedding Method for 20k-Spin CMOS Annealing Machines] proposes a fully-connected Ising model embedding method onto 20k-spin CMOS annealing machine, and prove that the ground state of the Ising models obtained from our proposed method is equivalent to that of the original Ising model. Experimental results effectively show (a) our proposed method embeds Ising models using less physical spins compared to the de facto standard conventional method in the practical problem size (b) the probability of feasible solutions and the solution quality using our proposed method is better than those of the conventional method when solving a combinatorial optimization problem.

Chapter 6 [Rectangle Packing by Ising Computers] proposes a mapping of "Rectangle Packing Problem" for solving it by the Ising model-based computers. In our proposed mapping, a sequence-pair is represented on an Ising model, and the energy function to obtain the optimal solution of the problem is constructed. Our proposed approach maps a "Rectangle Packing Problem" with N rectangles onto a $3N^3$-spin logical Ising model. Experimental results demonstrate that by the proposed approach we can solve the problem with nine rectangles at the maximum on a fully-connected annealing machine and the problem with three rectangles at the maximum on 20k-spin CMOS annealing machine.

Chapter 7 [Conclusion] summarizes this dissertation and gives some future works. In conclusion, we can see a reasonable expectation of using Ising model-based computers for the practical problems. However, we have many tasks. Corrections of spins when interpreting them to reduce the unsatisfactions of the constraints as a post-process in the application layer proposed in Chapter 7, and improving the mapping to reduce the number of required physical spins are ones of the future works. To implement and evaluate our HLS solutions onto FPGA devices is also one of the future works.

早稲田大学　博士（工学）　学位申請　研究業績書

氏 名　　　寺田　晃太朗　　　印

（2017 年　12 月　現在）

種 類 別	題名、　　発表・発行掲載誌名、　　発表・発行年月、　　連名者（申請者含む）
a. 論文 学術誌 原著論文	① ○ **K. Terada**, M. Yanagisawa, and N. Togawa, "A bitwidth-aware high-level synthesis algorithm using operation chainings for Tiled-DR architectures," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E100-A, no. 12, pp. 2911–2924 Dec. 2017. ② ○ **K. Terada**, M. Yanagisawa, and N. Togawa, "A high-level synthesis algorithm with inter-island distance based operation chainings for RDR architectures," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E98-A, no. 7, pp. 1366–1375, Jul. 2015.
c. 講演 国際会議	③ ○ **K. Terada**, M. Yanagisawa, and N. Togawa, "A floorplan-driven high-level synthesis algorithm with multiple-operation chainings based on path enumeration," in *Proceedings of 2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2129–2132, Lisbon, Portugal, May 2015. ④ ○ **K. Terada**, M. Yanagisawa, and N. Togawa, "A floorplan-driven high-level synthesis algorithm with operation chainings using chaining enumeration," in *Proceedings of 2014 IEEE Asia Pacific Conference on Circuits and Systems (APC-CAS)*, pp. 248–251, Ishigaki, Japan, Nov. 2014.
国内学会	⑤ <u>寺田晃太朗</u>，田中宗，林真人，山岡雅直，柳澤政生，戸川望，"20K スピン CMOS アニーリングマシンを対象とした完全結合イジングモデルマッピング手法," 日本物理学会 2017 年秋季大会，盛岡市，Sep. 2017. ⑥ （査読あり）<u>寺田晃太朗</u>，田中宗，林真人，山岡雅直，柳澤政生，戸川望，"20K スピン CMOS アニーリングマシンを対象とした完全結合イジングモデルマッピング手法と評価," 情報処理学会 DA シンポジウム 2017 論文集，pp. 163–168，加賀市，Sep. 2017. ⑦ （ポスター発表）長谷川健人，石川遼太，<u>寺田晃太朗</u>，川村一志，多和田雅師，戸川望，"組込みデバイスと FPGA を用いたナンバーリンクソルバの設計と実装," 情報処理学会 DA シンポジウム 2017 ポスター発表，加賀市，Aug. 2017. ⑧ <u>寺田晃太朗</u>，柳澤政生，戸川望，"演算ビット幅に基づく演算チェイニングを用いた RDR アーキテクチャ向け性能指向高位合成手法," 電子情報通信学会 2016 年ソサイエティ大会講演論文集，p. 71，札幌市，Sep. 2016.

早稲田大学　博士（工学）　学位申請　研究業績書

種 類 別	題名、　　発表・発行掲載誌名、　　発表・発行年月、　　連名者（申請者含む）
c．講演 国内学会	⑨（ポスター発表）寺田晃太朗，長谷川健人，川村一志，多和田雅師，戸川望，"機械学習と FPGA を用いたナンバーリンクソルバ，" 情報処理学会 DA シンポジウム 2016 ポスター発表，加賀市，Sep. 2016. ⑩ 寺田晃太朗，柳澤政生，戸川望，"DFG のクリティカルパス最適化に基づく演算チェイニングを用いた RDR アーキテクチャ対象高位合成手法，" 信学技報，VLD2016-05，pp. 41–46，北九州市，May 2016. ⑪（ポスター発表）寺田晃太朗，川村一志，多和田雅師，藤原晃一，戸川望，"機械学習を用いたナンバーリンクソルバ，" 情報処理学会 DA シンポジウム 2015 ポスター発表，加賀市，Aug. 2015. ⑫（査読あり）寺田晃太朗，柳澤政生，戸川望，"演算チェイニングの候補列挙・選択アルゴリズムを用いたフロアプラン指向高位合成手法，" 情報処理学会 DA シンポジウム 2015 論文集，pp. 17–22，加賀市，Aug. 2015. ⑬（査読あり）寺田晃太朗，柳澤政生，戸川望，"多段演算チェイニングを利用した配線遅延を考慮した高位合成手法，" 情報処理学会 DA シンポジウム 2014 論文集，pp. 115–120，下呂市，Aug. 2014. ⑭（査読あり）寺田晃太朗，柳澤政生，戸川望，"演算チェイニング候補列挙に基づく配線遅延を考慮した高位合成手法，" 第 27 回回路とシステムワークショップ論文集，pp. 440–445，淡路市，Aug. 2014.
e．その他 業績賞等	2017 年 8 月，情報処理学会 DA シンポジウム 2017 アルゴリズムデザインコンテスト最優秀賞． 2016 年 9 月，情報処理学会 DA シンポジウム 2016 アルゴリズムデザインコンテスト最優秀賞． 2016 年 9 月，情報処理学会 DA シンポジウム 2016 アルゴリズムデザインコンテスト特別賞． 2016 年 9 月，情報処理学会 第 176 回システムと LSI の設計技術研究発表会 優秀発表学生賞． 2015 年 8 月，情報処理学会 DA シンポジウム 2015 アルゴリズムデザインコンテスト優秀賞（学生部門）．

早稲田大学　博士（工学）　学位申請　研究業績書

種　類　別	題名、　　　発表・発行掲載誌名、　　　発表・発行年月、　　　連名者（申請者含む）
e．その他 研究費・ 助成金	早稲田大学理工学術院総合研究所，若手研究者支援事業アーリーバード，"超低消費電力を実現する革新的アニーリングマシンアプリケーション技術，" 2017 年 6 月–2018 年 3 月，総額 80 万円. 日本学術振興会特別研究員奨励費，"高性能かつ小面積を実現するフロアプランを考慮した LSI 上位設計技術，" 2016 年 4 月–2018 年 3 月，総額 190 万円（2016 年度：100 万円，2017 年度：90 万円）.