

Preface

The main intention of this book is to propose innovative scalable methodologies, which support the near-optimal designing and mapping process of embedded systems and to describe the general principles and the processes which lead to the development of methodologies with near-optimal and scalable characteristics. The material of the book is based on research performed in collaboration with University of Patras and IMEC.

As the requirements of the embedded systems have increased, increased complexity is imposed to the systems. Systematic design space exploration (DSE) methodologies are needed to support a near-optimal design of the system under development, which satisfies the requirements in performance, area and energy consumption, etc., to keep the overall cost of the final design low and to derive the final design within the available short time to market. The existing DSE approaches cannot concurrently guarantee scalable exploration time and near-optimal designs, as either they require too much exploration time to find near-optimal designs due to the high number of system parameters and the correlations between them or they have to end up with a less efficient trade-off result in order to find a design within acceptable time to market.

This book presents an alternative direction in DSE methodologies which is based on systematic creation of scalable and near-optimal DSE frameworks. It describes a set of principles and processes followed by the proposed reusable DSE methodology to develop scalable and near-optimal DSE frameworks and to efficiently use them to derive scalable and near-optimal design solutions within a Pareto trade-off space. The main principles of the proposed reusable DSE methodology are as follows: (1) the top-down split, which creates into two complementary and nonoverlapping steps with unidirectional constraint propagation using asymmetries in the characteristics of the problem under study; (2) the pruning of the incompatible steps using additional constraints deriving from restricting the problem under study; and (3) the exploration of the trade-offs through parametric templates that describe the available options. The remainder of the book describes how this set of principles and processes has been applied to create near-optimal and scalable DSE frameworks, which describe the corresponding methodologies. The created methodologies consist of a

sequence of unidirectional steps: (1) the costly design iterations are avoided leading to scalable approaches and (2) the constraints between steps are not overlooked leading to near-optimal final designs.

The target domain, where the proposed reusable DSE methodology is applied, consists of real-time cost-sensitive embedded systems with loop- and data dominated applications, such as video, image and speech processing. The reusable DSE methodology presented in this book is applied to several design steps of the processor level in the design flow of the embedded systems in order to create scalable and near-optimal DSE frameworks for each design step. The design steps under study belong to both the data storage and the processing part of the system. The first part of this book is dedicated to the development of DSE frameworks which describe mapping methodologies for storing large embedded system data arrays in the lower layers of the on-chip background data memory hierarchy. The DSE methodologies for the data storage mapping are illustrated for the intra-signal in-place optimization step, i.e. to find the maximum number of concurrent alive elements during the complete execution of the application (array window). The second part of this book is dedicated to the DSE frameworks for methodologies for the processing part of SW/HW architectures in embedded systems including the foreground memory systems. The DSE methodologies for the processing part are illustrated for the scheduling and assignment techniques, which are applicable in the foreground memory and the data path, and the instruction set processor mapping step, especially for the mapping into an FPGA platform with a microprocessor core and several custom hardware accelerators.

The existing mapping methodologies for the data storage are either enumerative, symbolic/polyhedral or worst-case (heuristics) approximations. The enumerative methodologies are optimal, but their exploration time is highly increased with the increase in the number of memory accesses. The symbolic/polytope methodologies are scalable and near-optimal for solid and regular memory accesses. When they are applied in irregular memory accesses, they either have to approximate the space by applying a convex hull or have to split the iteration space into unions of polytopes, where the exploration time is increased with the increase in the number of polytopes created. The worst-case approximations lead to overestimation of the storage requirements. In this book, we applied the reusable DSE methodology to develop DSE frameworks that describe innovative intra-signal in-place optimization methodologies, which compute the storage size requirements with linear exploration time and in a near-optimal way for both regular and highly irregular memory accesses. The DSE frameworks consist of sets with the different cases that may exist for the nonoverlapping store and load memory accesses and for the overlapping store and load memory accesses. Then, scalable and near-optimal solutions are proposed per intra-signal in-place case. To our intra-signal in-place methodology, a new representation of the array memory accesses is proposed, which is appropriate to express both regular and irregular accesses in a scalable and near-optimal way. The representation consists of a general pattern formulation which describes the memory accesses in a compact and repetitive way. Pattern operations have been

developed to combine the patterns in a scalable and near-optimal way for each of the different pattern combinations which may exist in the application under study.

In the processing part of the embedded systems, a DSE framework is developed to describe the potential options that may exist for design-time near-optimal scheduling and assignment techniques. A methodology is developed which uses the DSE framework of scheduling and assignment techniques to create scalable and near-optimal parametric templates for scheduling and assignment techniques, which satisfy the target domain constraints. The proposed DSE methodology is illustrated for the target domain of complex and large application graphs. For the instruction set processor mapping design step, either the existing DSE methodologies are scalable, but they cannot guarantee near-optimality, as they overlook constraints between design parameters, or they propose a near-optimal design, but they cannot guarantee that their exploration time remains scalable, as they require several iterations of the DSE steps. In this book, a DSE framework is developed to describe a scalable and near-optimal instruction set processor mapping methodology especially for mapping instances of a loop-dominated application domain onto a partially fixed architecture platform template, which consists of one processor core and several custom hardware accelerators. The DSE methodology consists of a DSE framework with unidirectional steps. Each step is described by parametric template, i.e. the available step options presented in a parametric way. As the steps are unidirectionally connected, the steps are applied without costly design step iterations. The proposed DSE methodology explores the space by instantiating the steps, i.e. by providing valid values to the parameters of the template and propagating design constraints following the unidirectional step ordering to prune suboptimal design options of the next DSE step, leading to a scalable solution. The result is a final Pareto trade-off curve with the most relevant near-optimal designs.

We, therefore, expect this book to be of interest in academia, both for the reusable DSE methodology and for the developed near-optimal and scalable methodology per design step. In the former case, the readers obtain the principles and processes required in order to develop scalable and near-optimal DSE methodologies. As the book demonstrates how the principles and the processes are applied in the problems under study, i.e. the data storage mapping step, the instruction set processor mapping step and the scheduling and assignment problem, the reader, by following the examples in this book, can apply the principles and processes in other steps of the design flow to develop new scalable and near-optimal DSE methodologies. In addition, the reader can directly apply the proposed scalable and near-optimal DSE methodologies to solve the problems of the memory optimization, the instruction set processor mapping of loop-dominated and data-dominated applications and the development of scalable and near-optimal scheduling and assignment techniques for the target domain of large and complex application graphs.

It has been a pleasure for us to work in this research domain and to cooperate between IMEC and University of Patras. We would also like to thank many people who helped us in realizing these results and achieving these contributions. In particular we wish to mention the following: Giorgos Athanasiou, Vasileios Kelefouras, Praveen Raghavan, Dimitris Tsitsipis, Christos Panagiotou, Sofia-Maria

Dima and others. We finally hope that the reader will find the book useful and enjoyable and that the presented results will contribute to the continued progress of the field of embedded system design.

Piraeus, Greece
Leuven, Belgium
Patras, Greece
January 2014

Angeliki Kritikakou
Francky Catthoor
Costas Goutis

Scalable and Near-Optimal Design Space Exploration
for Embedded Systems

Kritikakou, A.; Catthoor, F.; Goutis, C.

2014, XVII, 277 p. 80 illus., 2 illus. in color., Hardcover

ISBN: 978-3-319-04941-0