

FPGA Architecture Enhancements to Support Heterogeneous Partially Reconfigurable Regions

Christophe Huriaux
University of Rennes 1
IRISA
Lannion, France
christophe.huriaux@irisa.fr

Olivier Sentieys
Inria
IRISA
Lannion, France
olivier.sentieys@inria.fr

Russell Tessier
Department of Electrical and Computer Engineering
University of Massachusetts
Amherst, MA 01003
tessier@umass.edu

I. INTRODUCTION

Partial dynamic reconfiguration has become an important feature of FPGA-based systems as the number of applications which use this capability has increased. For systems using multiple partial bitstreams, the complexity of the target reconfigurable region, which often include heterogeneous blocks such as block RAMs and DSP blocks, makes it difficult to generate a unique bitstream which can be loaded into multiple locations in an FPGA. Although the migration of homogeneous lookup-table based logic blocks across the logic fabric has been shown to be relatively straightforward for older FPGAs, the variety and organization of heterogeneous blocks in modern FPGAs now render this operation more complex. In many applications, it is desirable to place and route a partial design once, store it in memory, and then freely load it into the reconfigurable fabric with as few constraints as possible. This work addresses the need to relocate partial designs which include heterogeneous resources.

II. APPROACH

In this work we develop an FPGA architecture which allows for the placement of a partial FPGA design on the logic fabric even if the relative placement of heterogeneous blocks within the target region is not identical to the placement used to generate the bitstream for the partial design. This work has been conducted in the context of the European FP7 FlexTiles project in which a dynamically reconfigurable logic fabric is embedded in a 3-D stacked chip along with a many-core architecture. The reconfigurable logic fabric is used to load hardware-accelerated functions whose use is scheduled at run time. All communication between the fabric and many-core is made via dedicated I/O interface blocks in the fabric. This communication configuration increases the need for a flexible architecture which can handle the placement of a single application bitstream in multiple locations on the logic fabric.

Our architectural enhancements are two-fold. First, heterogeneous resources are abstracted from the logic-only array and the corresponding routing network is split. A logic-only island-style array uses a dedicated routing network (the *homogeneous routing network*) while heterogeneous resources are tied to vertical long routing lines, each dedicated to an input or output of a single resource. These *heterogeneous routing lines* are connected to multiple switch-blocks of the *homogeneous routing network*, which allows for multiple connection points to a heterogeneous resource.

Second, the generated bitstreams only include connections to heterogeneous blocks through the dedicated long lines, via the switch-boxes. This allows for a given hardware task (partial design) to be placed anywhere on the logic fabric as long as the switch-boxes of the target region include a connection to heterogeneous resources. The organization of the resources and the distribution of the heterogeneous connections among the switch-boxes are arranged to effectively abstract the heterogeneity of the fabric so that a single bitstream can be loaded at different positions.

This flexibility enhancement comes at the cost of additional area and a performance impact. The additional routing resources and enhanced switch-boxes add significant area. Moreover, due to the variable route length between the homogeneous routing network and the heterogeneous resources (i.e. since the connection to the long line can be made at multiple points), the delay of the routes going to or from a heterogeneous resource is bounded by their maximal delay.

III. RESULTS

Our target architecture was modeled using a modified version of the Versatile Place-and-Route (VPR) software which can account for two distinct routing networks. We compared a classic routing architecture, similar to the Stratix IV, with an enhanced version which includes the refinements defined in the previous section. Twelve benchmarks including between 30 and 2,977 logic blocks and between 1 and 53 heterogeneous blocks were placed and routed on these two architectures. The results show an average increase on $1.97\times$ on the channel width of the homogeneous routing network to account for the routing to dedicated switch-blocks. The delay is increased by 10% which shows the influence of the enhancements on the critical path for average positions of the task on the fabric.

ACKNOWLEDGMENT

This research is sponsored by the European Commission under the 7th Framework program within the FlexTiles project (FPT ICT-288248) and by the French Ministry of Research.