A Customised Shell for Management and Monitoring of Dynamically Reconfigurable Accelerators

Alessandro Cilardo, Giuseppe Tiano, Nicola Mazzocca and Antonino Mazzeo

Università degli Studi di Napoli Federico II, Italy

In the last decades, Dynamic Function Exchange has become a hot topic in the field of embedded systems and High Performance Computing, since it allows FPGAs to be used in a much more flexible way. In fact, designers can divide their designs in a static part, whose functionality cannot change without reconfiguring the entire device, and one or more dynamic regions, which can be programmed at runtime, without having any impact on the static part of the design. This allows dynamic regions to host different accelerators at different times, adapting themselves to changing needs over time. However, the development of a partially reconfigurable design requires that the static part of the FPGA is equipped with some suitable support for controlling the partial reconfiguration process, which involves (among other things) using configuration interfaces and providing decoupling mechanisms for reconfigurable partitions.

This paper describes the implementation of such a support layer, called here an *FPGA shell*, whose main objectives are:

- To provide management and monitoring functions for the accelerators that are dynamically loaded in reconfigurable regions;
- To simplify the development of partially reconfigurable designs, letting designers focus only on the functionality of reconfigurable accelerators, without caring about the necessary support in the static part of the design.

In the first section, we provide a general description of the hardware and software architecture of our shell. Then, subsequent sections elaborate on the different support mechanisms provided. First of all, the shell offers the possibility to check the temperature of an accelerator and some parameters related to its utilisation. The shell also provides the possibility to access the accelerators remotely over an inter-node network, using Remote Direct Memory Access. Moreover, it offers support for freezing, saving and restoring the execution context of a hardware task. This is a really complex feature which supports suspending the computation of an accelerator in specific situations (such as a sudden increase in temperature), checkpointing of long or critical tasks, and also preemption-based management, when reconfigurable partitions are shared among multiple processes. In that respect, the last section discusses how this feature could be integrated in a virtualised environment, allowing the FPGA to be shared among different virtual machines using both preemption and a priority-based approach.