Combining Application Adaptivity and System-wide Resource Management: A novel approach

Eedoardo Paone, Giuseppe Massari, Patrick Bellasi, William Fornaciari, Gianluca Palermo, Cristina Silvano and Vittorio Zaccaria
DEIB, Politecnico di Milano, Italy

Overview

Next-gen many-core platforms

Application Adapativity

Adaptivity can be defined as the capability of an application to dynamically reconfigure itself according to sensitive events, i.e., changes in computational resource availability, or application-specific objectives. The latter case may introduce variability in the application resource requirements.

At design-time, we exploit a tool for Design Space Exploration (DSE), called MOST, to profile different execution of the application on the target hardware platform. Each execution is characterized by a specific configuration, consisting of a resource usage level (processing units, memory, etc...) and a set of values of some application specific parameters. The former are grouped in the set of Application Working Modes (AWM), while the latter form the set of Operating Points (OP).

Given some quality metrics (e.g., accuracy of results, QoS, performance, power consumption, etc...), MOST extracts the set of Pareto optimal configurations. Such configurations will characterize the reconfigurability of the application at run-time.

At run-time, a hierarchical and distributed resource management strategy is enforced. The Run-Time Resource Manager (RTRM) selects the Application Working Mode for each application, thus establishing the amount of resources assigned. This enables a subset of Operating Points.

The application then can set a goal with respect to its quality metrics (e.g., frames per second, accuracy of results, etc...) according to an event or certain run-time conditions. The Application-Specific Run-Time Manager (AS-RTRM) takes care of the goal set, by selecting the Operating Point that should guarantee such goal. Whether the goal is not satisfied, the application can ask for more resources to the RTRM (i.e., a change of AWM). The request will be taken into account for the next system-wide resource allocation optimization.

System-wide Resource Management

The system-wide resource management is in charge of solving the contention in the access to computational resources by multiple running applications. To do that, we are developing the BarbequeRTRM, a run-time resource management framework for multi/many-core architectures and adaptive applications.

The framework provides a support to the application adaptivity, and policies to perform optimal resource partitioning.

The system-wide resource partitioning policies are implemented as heuristics to solve multi-objective optimization problems, by taking into account:

- **Application properties**, e.g., resource requirements, Quality-of-Service and priorities;
- **Resource status**, e.g., availability, frequency, power consumption, process variation and thermal conditions;
- **Tunable optimization goals**, e.g., overall performance/QoS maximization, reconfiguration overheads minimization, resource congestion avoidance, fairness, energy efficiency.

Whether the hardware platform provides low-level control mechanisms (e.g., DVFS), further actions can be taken by the RTRM.

Run-time Library (RTLlib)

Connect the application to the RTRM and export some framework API. Define the Abstract Execution Model (AEM) that simplify the development of adaptive and reconfigurable applications.

More

Checkout DATE 2013 Friday Workshops “Platform 2012 / STHORM embedded many-core acceleration”:
“An Exploration Methodology for a Customizable OpenCL Stereo-Matching Application Targeted to P2012”, Vittorio Zaccaria et al. - Politecnico di Milano
“Run-Time Resource Management on Many-Core STHORM Platform”, Patrick Bellasi - Politecnico di Milano

http://www.2parma.eu

DATE University Booth, Grenoble, France - March 18-22, 2013