Application- and platform-agnostic runtime power management of heterogeneous embedded systems

PRiME Project: Power-efficient, Reliable, Many-core Embedded systems, www.prime-project.org

Cyber Physical Systems Research Group, Electronics & Computer Science, University of Southampton, UK

Abstract

Increasing energy efficiency and reliability at runtime is a key challenge of heterogeneous many-core systems. We demonstrate how contributions from the PRiME project integrate to enable application- and platform-agnostic runtime management that respects application performance targets. We consider opportunities to enable runtime management across the system stack and we enable cross-layer interactions to trade-off power and reliability with performance and accuracy. We consider a system as three distinct layers, with abstracted communication between them, which enables the direct comparison of different approaches, without requiring specific application or platform knowledge. Application-agnostic runtime management is demonstrated with a selection of runtime managers from PRiME, including linear regression modelling and predictive thermal management, operating across multiple applications. Platform-independent runtime management is demonstrated using two heterogeneous platforms.

Background

Energy consumption and reliability are two important optimization objectives for many-core embedded platforms. These two objectives motivate the need for runtime management (RTM); however, current algorithms generally focus on specific application/platform combinations. In this demonstrator, we present a generic and cross-platform approach to standardise the runtime management of software applications executing on heterogeneous platforms.

Demonstrator

Figure 1 shows the two different parts of the PRiME demonstrator. The first part demonstrates how platform-agnostic RTM can be achieved by profiling the behaviour of one application on two platforms. The second part shows application-agnostic RTM with different algorithms controlling the same set of applications. Device, RTM and application layers communicate through an API with knobs and monitors. Platforms and RTMs are summarised below.

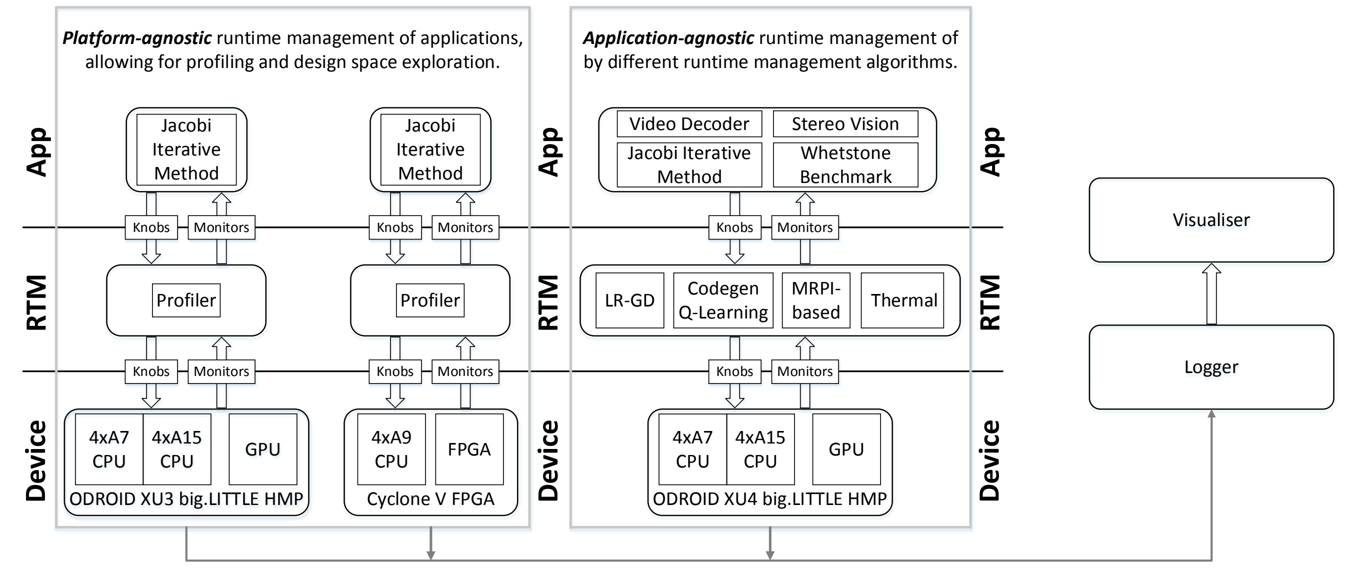

Figure 1: Block diagram of PRiME Demonstrator showing key components.

PRiME Runtime Management Approaches

Linear regression modelling: Model-based prediction with gradient descent control of application & device parameters.

Code generated Q-Learning: Uses reinforcement learning of application behaviour to predict system performance requirements. Code generated from a formal model.

Memory Reads per Instruction (MRPI): Uses processor performance counters to detect memory-bound applications.

Predictive thermal management: Extends MRPI to include regression modelling based temperature prediction.

Demonstrator Platforms

Odroid XU3 & Odroid XU4: quad-core ARM® Cortex®-A7 and quad-core ARM Cortex-A15 in a HMP big.LITTLE™ configuration with an ARM Mali™-T628 GPU.

Altera® Cyclone® V: dual-core ARM Cortex-A9 CPU with FPGA Fabric running OpenCL kernels for the Jacobi iterative method.

This work was supported by the PRiME programme grant EP/K034448/1 (http://www.prime-project.org) and EPSRC grant EP/L000563/1.