Efficient Instruction and Data Caching for High Performance Embedded Processors

A. Ferrerón Labari, D. Suárez Gracia, V. Viñals Yúfera

Grupo de Arquitectura de Computadores (gaZ)
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-976762707, Fax +34-976762043, e-mail: {ferreron, dario, victor}@unizar.es

Abstract

In the last years, embedded systems have evolved so that they offer capabilities we could only find before in high performance systems. Portable devices already have multiprocessors on-chip (such as PowerPC 476FP or ARM Cortex A9 MP), usually multi-threaded, and a powerful multi-level cache memory hierarchy on-chip. As most of these systems are battery-powered, the power consumption becomes a critical issue. Achieving high performance and low power consumption is a high complexity challenge where some proposals have been already made. Suarez *et al*. proposed a new cache hierarchy on-chip, the LP-NUCA (*Low Power NUCA*), which is able to reduce the access latency taking advantage of NUCA (*Non-Uniform Cache Architectures*) properties. The key points are decoupling the functionality, and utilizing three specialized networks on-chip. This structure has been proved to be efficient for data hierarchies, achieving a good performance and reducing the energy consumption. On the other hand, instruction caches have different requirements and characteristics than data caches, contradicting the low-power embedded systems requirements, especially in SMT (*simultaneous multi-threading*) environments. We want to study the benefits of utilizing small tiled caches for the instruction hierarchy, so we propose a new design, ID-LP-NUCAs. Thus, we need to re-evaluate completely our previous design in terms of structure design, interconnection networks (including topologies, flow control and routing), content management (with special interest in hardware/software content allocation policies), and structure sharing. In CMP environments (*chip multiprocessors*) with parallel workloads, coherence plays an important role, and must be taken into consideration.