Title:
Dynamic cache reconfiguration based techniques for improving cache energy efficiency

Abstract:

Modern multicore processors are employing large last-level caches, for example Intel's E7-8800 processor uses 24MB L3 cache. Further, with each CMOS technology generation, leakage energy has been dramatically increasing and hence, leakage energy is expected to become a major source of energy dissipation, especially in last-level caches (LLCs). The conventional schemes of cache energy saving either aim at saving dynamic energy or are based on properties specific to first-level caches, and thus these schemes have limited utility for last-level caches. Further, several other techniques require offline profiling or per-application tuning and hence are not suitable for product systems.

In this research, we propose novel cache leakage energy saving schemes for single-core and multicore systems; desktop, QoS, real-time and server systems. We propose software-controlled, hardware-assisted techniques which use dynamic cache reconfiguration to configure the cache to the most energy efficient configuration while keeping the performance loss bounded. To profile and test a large number of potential configurations, we utilize low-overhead, micro-architecture components, which can be easily integrated into modern processor chips. We adopt a system-wide approach to save energy to ensure that cache reconfiguration does not increase energy consumption of other components of the processor. We have compared our techniques with the state-of-art techniques and have found that our techniques outperform them in their energy efficiency. This research has important applications in improving energy-efficiency of higher-end embedded, desktop, server processors and multitasking systems. We have also proposed performance estimation approach for efficient design space exploration and have implemented time-sampling based simulation acceleration approach for full-system architectural simulators.