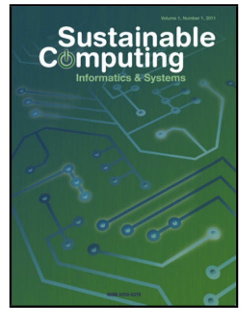
Accepted Manuscript



Title: Impact of cache voltage scaling on energy-time pareto frontier in multicore systems

Authors: Shervin Hajiamini, Behrooz Shirazi, Aaron Crandall, Hassan Ghasemzadeh, Chris Cain

PII: S2210-5379(17)30122-1
DOI: https://doi.org/10.1016/j.suscom.2018.02.011
Reference: SUSCOM 235

To appear in:

Received date: 14-4-2017
Revised date: 30-9-2017
Accepted date: 26-2-2018

Please cite this article as: Shervin Hajiamini, Behrooz Shirazi, Aaron Crandall, Hassan Ghasemzadeh, Chris Cain, Impact of cache voltage scaling on energy-time pareto frontier in multicore systems, Sustainable Computing: Informatics and Systems https://doi.org/10.1016/j.suscom.2018.02.011

Impact of Cache Voltage Scaling on Energy-Time Pareto Frontier in Multicore Systems

Shervin Hajiamini, Behrooz Shirazi, Aaron Crandall, Hassan Ghasemzadeh, Chris Cain

School of EECS, Washington State University, Pullman, WA 99164, United States

{shajiami, acrandal, hassan, ccain, shirazi}@eecs.wsu.edu

Highlights

- A frontier curve to strike a tradeoff between energy consumption and execution time of a multicore system. A set of simulations profile an execution trace of the cores with fine-grained V/F levels, based on which a multi-objective optimization problem is formulated and solved by LP.
- Investigation into the effect of scaling the frequency of L1/L2 caches with cores and the extent to which the Pareto frontier is improved over the energy-time search space.
- Comparison of the Pareto frontier with solutions provided by three heuristics in the energy-time search space. The results show that the LP-based Pareto frontier is significantly better than the performance of these heuristics as later demonstrated in Fig. 10.

Abstract

High performance computing centers need to keep up with the growing workload of varying computational characteristics. Due to their high computation rates, these computing systems consume vast amounts of energy with increasing electricity costs. As an approach to balancing computation demand with energy consumption, state-of-the-art dynamic voltage and frequency scaling (DVFS) methodologies are used for improving the energy efficiency of computing systems. However, these studies often do not explore the extent to which their solutions are close to theoretically optimal limits. This work formulates optimal boundaries for energy-time performance with a Linear Programming (LP) approach. The formulation utilizes per-core energy consumptions and execution times obtained during the profiling phase to optimize the voltage and frequency (V/F) level assignments at runtime. For each of the four benchmarks considered in this work, the optimized V/F assignments are used to bound Pareto frontiers, which trade off energy consumption and execution time. In particular, this work studies the impact of scaling the voltage and frequency of the cache subsystem in a multicore system on establishing the energy-time Pareto frontier. An unexpected result of our study is that when the frequencies of caches are not scaled with that of the cores (i.e., fixed at 2.0 GHz), the proposed LP-based technique improves the overall Energy-Delay-Product (EDP) as much as 35% compared to the traditional no-DVFS Pareto frontier. Furthermore, this work compares the performance of three heuristic-based energy-efficient DVFS algorithms to demonstrate the differences between heuristics performances and the LP-based optimal Pareto frontier.

Keywords: Multicore; DVFS; linear programming, Pareto frontier; Cache, Energy Consumption