CCIT Report #630
July 2007

Low-Complexity Policies for Energy-Performance Tradeoff in Chip-Multi-Processors

Avshalom Elyada, Ran Ginosar, Uri Weiser

Abstract – **Chip-Multi-Processors (CMP) utilize multiple energy-efficient Processing Elements (PEs) to deliver high performance while maintaining an an efficient ratio of performance to energy-consumption. In order to utilize CMP resources, the application software is split into multiple tasks that are executed in parallel on the PEs. Dynamic frequency-Voltage Scaling (DVS) balances performance and energy consumption by dynamically varying a PE's frequency-voltage *workpoint* in order to save energy while meeting performance requirements. This work addresses DVS policies for CMP. We consider multi-task CMP applications with unknown workloads. We dynamically set frequency-voltage workpoints for each PE in the CMP, attempting to minimize a defined energy-performance criterion. Other DVS methods typically use high complexity optimization techniques, which limits the possibility of real-time implementation in performance-driven, energy-aware systems. In contrast, we investigate simple heuristic DVS policies for simplified serial/parallel task-graphs. We compare the results of our polices to a theoretical best-case solution and show that these lightweight heuristics achieve good results with low complexity. In most cases the simplest policy, named Constant, which usually keeps tasks running at a constant workpoint, is the most cost-effective one.**

Index Terms – Chip-**Multi-Processors, Dynamic Voltage Scaling, Slack Utilization, Low Power**

1. Introduction

Chip-Multi-Processors (CMP) achieve high performance while maintaining an acceptable ratio of performance to energy consumption, in comparison to traditional single-core architectures. Performance improvement techniques of single-core architectures, mainly include (1) taking advantage of shrinking gate-delays in order to increase the operating frequency, and (2) using the increased transistor density to add performance-enhancing microarchitecture features [1],. Increasing the operation frequency beyond a certain point is energy inefficient, since energy consumption is roughly quadratic in frequency. Features such as large caches, deep execution pipes and complex branch predictors yield a decreasing performance/energy return [2]. High energy consumption shortens the battery life of mobile devices, and may cause power delivery and heat dissipation problems, which consequently limit feasible frequencies and performance. CMP architectures, on the other hand, integrate multiple relatively small and simple PEs, potentially enabling linear scaling of performance [3].

Dynamic frequency-Voltage Scaling (DVS) is a widely practiced [4] and researched [5-7] technique for energy-performance tradeoff. When using DVS in CMP, a PE's frequency is altered dynamically to meet current performance requirements while consuming no more energy than is necessary. PE supply voltage is also adjusted in conjunction with frequency; usually kept at the lowest feasible value that still enables circuit operation and timing at the current frequency. Scaling the frequency-voltage workpoint (*f,V*) can result in near-quadratic energy savings [6].

In a CMP running multiple dependent tasks, DVS may save energy without degrading performance. Typically, at any given time, one or few tasks constitute(s) the performance bottleneck. Other PEs can be slowed down, saving energy without affecting total performance. We refer to the tactic of slowing down non-critical tasks as *slack-utilization*.

When all task workloads are known in advance, the *DVS policy* sets PE frequencies to utilize precisely all available time-slacks and thus save the maximum possible energy without affecting performance.

The authors are with the Electrical Engineering Department, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: avshael@tx.technion.ac.il).