Enabling Realistic Fine-Grain Voltage Scaling with Reconfigurable Power Distribution Networks

Waclaw Godycki, Christopher Torng, Ivan Bukreyev, Alyssa Apsel, Christopher Batten
School of Electrical and Computer Engineering, Cornell University

1 Abstract
Recent work has shown that monolithic integration of voltage regulators will be feasible in the near future, enabling reduced system cost and the potential for fine-grain voltage scaling (FGVS). In this paper, we use architecture-level modeling to explore a new dynamic voltage-frequency scaling controller called the fine-grain synchronization controller (FG-SYNC) to enable improved performance and energy efficiency at similar average power for multithreaded applications with activity imbalance. We then use circuit-level modeling to explore various approaches to implementing on-chip voltage regulation, including a new approach called reconfigurable power distribution networks (RPDNs). RPDNs allow one regulator to “borrow” energy storage from regulators associated with underutilized cores resulting in improved area and power efficiency and faster response times. We evaluate FG-SYNC and RPDNs using a vertically integrated research methodology, and our results demonstrate a 10–50% performance and 10–70% energy efficiency improvement on the majority of the applications studied compared to no FGVS, yet RPDNs use 40% less area compared to a more traditional low-power approach.

2 Motivation
Monolithic integration using a standard CMOS process provides a tremendous cost incentive for integrating closed-loop voltage regulators on-chip. Recent technology trends suggest that it is now becoming feasible to integrate switching regulators on-chip (e.g., Intel Haswell), enabling reduced system cost as well as the potential for fine-grain voltage scaling (FGVS) to exploit fine-grain activity imbalance in multithreaded applications for performance and energy efficiency benefits.

3 Target System
Our target system is an embedded processor composed of eight cores, single-issue, fine-grain, RISC, cortex, private, coherent 16 KB instruction and data L1 caches, and a shared L2/4KB unified L2 cache. We implemented the core and L1 memory system for this design in RTL and used a commercial standard-cell-based CAD toolflow targeting a TSMC 65nm process to generate layout. Each core can run at 333 MHz at 1V and the full-eight-core system is approximately 6mm2.

4 Types of Integrated Voltage Regulator
The three primary types of step-down voltage regulators are linear regulators (e.g., LDOs), inductor-based switching regulators (e.g., buck), and capacitor-based switching regulators (e.g., switched capacitor).

5 FGVS Architecture Design: FG-SYNC+
We explore a new FGVS controller called the fine-grain synchronization controller (FG-SYNC+), which exploits fine-grain scaling in level (i.e., many voltage levels, space, and time). This technique translates between levels to improve performance and energy efficiency while maintaining similar average power. FG-SYNC+ uses a thread-litigated approach with hints provided to inform the hardware about which cores are doing useful work vs. useless work (e.g., waiting for a task, waiting at a barrier).

6 FGVS Circuit Design: RPDN
We propose a new approach called reconfigurable power distribution networks (RPDNs). As shown below, RPDNs include many small "unit cells" that each contain the necessary capacitance and regulator switches required for a SC regulator. These cells can be flexibly reconfigured through a switch fabric and combined with per-core context circuits to electrically customize multiple-differentiated SC regulators on demand for cores. The inset shows how 16 unit cells can be allocated to four cores operating in four different modes.

7 Evaluation
We evaluate the performance, energy efficiency, and power of applications on our target system with each PDN. We use a 4-level, 8-domain FG-SYNC controller and account for realistic voltage-setting response times and regulator power efficiencies in each DVFS mode for varying load currents.

8 Acknowledgments
This work was supported in part by NSF CAREER Award #1149464, a Spira Fellowship, and donations from Intel Corporation and Synopsys, Inc. The authors acknowledge and thank J. Kim for his help in developing the instruction-based energy model, and Derek Lookhart, Shomea Siniraj, and Pol Rousea for their help in writing multcore application kernels.