Department of Electrical and Computer Engineering
University of Rochester, Rochester, NY
Ph.D. Public Defense

Monday April 4, 2016
2:00 PM
Computer Studies Building, Room 426

Power Delivery Aware Microarchitectures for Energy Efficient Computing

Yuxin Bai

Supervised by
Professor Engin Ipek

Abstract

Power- and energy-efficiency are significant requirements in virtually all computer systems, from mobile devices to large-scale data centers. Power delivery is a process that distributes stable supply voltages to gates within an integrated circuit (IC). The design of such a delivery network is a critical task to guarantee functionality, timing, and operation reliability, and significantly affects the power- and energy-efficiency of a high performance IC. Therefore, microarchitectural solutions that are aware of the power delivery system, should be capable of exploring a larger optimization space for energy efficient computer systems. This thesis proposes two microarchitectural techniques that leverage the design tradeoffs of the underlying power delivery networks to achieve energy-efficient computing.

First, the use of MOS current-mode logic (MCML) is explored as a fast and low- noise alternative to static CMOS logic in microprocessors, thereby improving the performance, energy-efficiency, and signal integrity of future computer systems. The power and ground noise generated by an MCML circuit is typically $10\times-100\times$ smaller than the noise generated by a static CMOS circuit, and therefore can significantly relax the typical design constraints imposed on the power delivery network. Unlike a static CMOS circuit, in which dynamic power is proportional to the clock frequency, an MCML circuit dissipates a constant power independent of the clock frequency. Although these traits make MCML highly energy-efficient when operating at high speeds, the constant static power of MCML poses a challenge for a microarchitecture that operates at a modest clock rate and with a low activity factor. To address this challenge, this thesis explores a single-core microarchitecture for MCML that takes advantage of the C-slow retiming technique, and runs at a high frequency with low complexity to save energy. This design principle diff fundamentally from the contemporary multicore design paradigm for static CMOS, which relies on a large number of gates running in parallel at modest speeds. The proposed architecture generates $10-40\times$ lower power and ground noise, and operates at a level of performance within 13% of a conventional, eight-core static CMOS system, while exhibiting $1.6\times$ lower energy and 9% les s area. Moreover, the operation of the MCML processor is robust under both systematic and random variations in transistor threshold voltage and effective channel length.

Dynamic voltage and frequency scaling (DVFS) is an effective technique used in power management. Voltage regulators are key components for power generation during the power delivery process. Emerging on-chip voltage regulators has the potential to increase the energy efficiency of computer systems by enabling the control of DVFS at a fi granularity in both space and time. A low dropout voltage regulator (LDO) is suitable for on-chip integration due to its speed, regulation quality, and area advantages. The energy conversion efficiency of an LDO, however, is dependent on the ratio of the input and output voltages, which results in energy waste when DVFS is applied over a wide voltage range. A DVFS framework that relies on a hierarchy of off-chip switching regulators and per-core on-chip LDOs is proposed. It ensures fast DVFS in nanoseconds and a more than 90% regulator efficiency over a wide voltage range. A control policy using a reinforcement learning (RL) approach is proposed to exploit the fine-granularity control of power and the high regulator efficiency enabled by the framework. Per-core RL agents learn and improve their DVFS policies independently, while retaining the ability to coordinate their actions to accomplish system level power management objectives. The proposed framework achieves 18% greater energy efficiency than a typical per-core DVFS framework using on-chip switching regulators when evaluated on a mix of 14 parallel and 13 multiprogrammed workloads. Moreover, the proposed RL policy is 21% more energy efficient as compared to an oracle policy with coarse-grained DVFS.