Department of Electrical and Computer Engineering

University of Rochester, Rochester, NY

Ph.D. Public Defense

Friday, April 11, 2014

10:00AM

Computer Studies Building 426

Reliable Ultra-Low-Voltage Cache Design for Many-Core Systems

Meilin Zhang

Supervised by
Professor Paul Ampadu

Energy efficiency and reliability are two main concerns in future many-core systems design. On the one hand, technology scaling makes it possible to integrate unprecedented number of transistors on a single chip, enabling many-core system design. On the other hand, this integration introduces energy and reliability issues for the whole system. Moreover, as the speed gap between processor and external memory increases, large-volume and high-density caches are required, worsening reliability and energy issues further. Supply voltage scaling is one of the most effective ways to reduce system energy consumption at the cost of performance and reliability. Low supply voltages increase the impact of process and other variations on circuit functionality and performance. Eventually, the system will fail below a minimal supply voltage V_{DDMIN}. Typically, the cache arrays in the system limit V_{DDMIN} of the whole processor. Consequentially, it is necessary to provide fault tolerance for the cache under low supply voltages to improve system energy efficiency while maintaining reliability.

We can classify cache errors as hard or soft errors. Hard errors may be caused by manufacturing defects, threshold or supply voltage variations, or device aging, and soft errors are introduced by external particle strikes or other random noise. Traditionally, most soft errors manifest as single bit upset. However, as we approach into nanometer era, the probability of multi-bit upset increases significantly because a single particle strike can cause more cache cells upset. To address both single bit upset and multi-bit upset, we propose two-layer error control codes, combining error detection capability of rectangular code and error correction capability of Hamming product code in an efficient way, to significantly improve system reliability while maintaining low area, power, and latency overhead.

To reduce supply voltage beyond normally acceptable V_{DDMIN} and maintain appropriate yield and reliability, we exploit existed double-error correcting triple- error detecting (DECTED) codes, together with cache line disabling in an efficient way to handle both hard and soft errors. The proposed method use double-error correcting triple-error detecting (DECTED) codes for each cache line—1-bit error correction for hard errors, and the other 1-bit error correction for soft errors. When there are multiple faulty cells, the cache lines will be disabled. This approach can reduce supply voltage beyond normally acceptable V_{DDMIN} and maintain appropriate yield and reliability. To further improve energy efficiency, an adaptive fault-tolerant cache architecture, which provides appropriate error control capability for each cache line based on the number of faulty cells detected, is proposed. We use SECDED codes for each cache lines to address soft errors, and extra parity bits are used when there are hard errors. Our experimental results show that the proposed method can reduce supply voltage and increase cache reliability furthermore.