While technology scaling enables increased density for memory cells, the intrinsic high leakage power of conventional CMOS technology and the demand for reduced energy consumption inspires the use of emerging technology alternatives such as eDRAM and Non-Volatile Memory (NVM) including STT-MRAM, PCM, and RRAM. The utilization of emerging technology in Last Level Cache (LLC) designs which occupies a significant fraction of total die area in Chip Multi-Processors (CMPs) introduces new dimensions of vulnerability, energy consumption, and performance delivery. To be specific, a part of this research focuses on eDRAM Bit Upset Vulnerability Factor (BUVF) to assess vulnerable portion of the eDRAM refresh cycle where the critical charge varies depending on the write voltage, storage and bit-line capacitance. This dissertation broaden the study on vulnerability assessment of LLC through investigating the impact of Process Variations (PV) on narrow resistive sensing margins in high-density NVM arrays, including on-chip cache and primary memory. Large-latency and power-hungry Sense Amplifiers (SAs) have been adapted to combat PV in the past. Herein, a novel approach is proposed to leverage the PV in NVM arrays using Self-Organized Sub-bank (SOS) design. SOS engages the preferred SA alternative based on the intrinsic as-built behavior of the resistive sensing timing margin to reduce the latency and power consumption while maintaining acceptable access time.

On the other hand, this dissertation investigates a novel technique to prioritize the service to 1) Extensive Read Reused Accessed (ERRA) blocks of the LLC that are silently dropped from higher levels of cache, and 2) the portion of the working set that may exhibit distant re-reference interval in L2. In particular, we develop a lightweight Multi-level Access History Profiler to efficiently identify ERRA blocks through aggregating the LLC block addresses tagged with identical Most Significant Bits into a single entry. Experimental results indicate that the proposed technique can reduce the L2 read miss ratio by 51.7% on average across PARSEC and SPEC2006 workloads.

In addition, this dissertation will broaden and apply advancements in theories of subspace recovery to pioneer computationally-aware in-situ operand reconstruction via the novel Logic In Interconnect (LI2) scheme. LI2 will be developed, validated, and refined both theoretically and experimentally to realize a radically different approach to post-Moore's Law computing by leveraging low-rank matrices features offering data reconstruction instead of fetching data from main memory to reduce energy/latency cost per data movement. We propose LI2 enhancements to attain high performance delivery in the post-Moore's Law era through equipping the contemporary micro-architecture design with a customized memory controller which orchestrates the memory request for fetching low-rank matrices to customized Fine Grain Reconfigurable Accelerator (FGRA) for reconstruction while the other memory requests are serviced as before. The goal of LI2 is to conquer the high latency/energy required to traverse main memory arrays in the case of LLC miss, by using in-situ construction of the requested data dealing with low-rank matrices. Thus, LI2 exchanges a high volume of data transfers with a novel lightweight reconstruction method under specific conditions using a cross-layer hardware/algorithm approach.

Major: Computer Engineering

Educational Career:
Bachelor’s of Computer Engineering, BS, 2009, Sepahan Institute of Technology
Master’s of Computer Engineering, MS, 2012, Amirkabir University of Technology
Master’s of Computer Engineering, MS, 2016, University of Central Florida

Committee in Charge:
Ronald F. DeMara, Chair, Electrical Engineering and Computer Science
Mingjie Lin, Electrical Engineering and Computer Science