

Title: "Towards Exploring CXL-SSD"

Speaker: Speaker: Lokesh N. Jaliminche (ljalimin@ucsc.edu)

Abstract:

In recent years, the memory demands of data center applications have grown exponentially. Applications such as Deep Learning Recommendation Systems (DLRMs) rely on embeddings to deliver high-quality recommendations, often spanning from multi-gigabytes to terabytes in size. Additionally, these applications demand small granular access (typically at the 64/128-byte). Unfortunately, as memory requirements increase, DRAM scaling faces limitations, forcing investment in costly infrastructure or resorting to FAR memory solutions.

On the other hand, NAND flash memory has exhibited remarkable density scaling (50x over the past decade). Coupled with the performance capabilities of PCIe-based CXL interconnects, major device vendors are exploring the prospect of utilizing NAND Flash memory as working memory through CXL. However, these devices are still in their nascent stages of development, lacking user-friendly platforms for further research.

To address this issue, we are working on developing CXL-SSD emulation with QEMU, which can be a free-to-use and easy-to-distribute experimentation platform. Such a platform can help accelerate the experimentation and exploration of CXL-SSD. We believe that this platform will facilitate in-depth investigations into CXL-SSD, addressing critical research questions, such as application changes required to utilize these devices and conducting comprehensive design space exploration.