Theme: New Memory

- Sub theme: New memory platform

The traditional memory hierarchy consisting of CPU Cache, DRAM, and storage is showing inefficiencies with the advent of new applications such as Big Data and Machine Learning, where we are seeing a trend that computing bottlenecks has shifted from flops to memory. This has given rise to memory-centric architectures that aims to solve memory bottlenecks in new applications.

Compute Express Link(CXL) is a new technology that allows for different memory hierarchies compared to traditional computer architectures. While the idea is promising, there is still much to learn about the capabilities and practical benefits of systems based on CXL. This theme explores memory platforms using CXL and takes a comprehensive approach, considering factors such as memory architecture, device requirements, potential applications, total cost of ownership, and more.

The most powerful computers and applications today are limited by their memory in one way or the other. The prevalence of heterogeneous computing, heterogeneous memory architectures, deeper memory hierarchies and increased adoption of disaggregated architectures are increasing the complexity of the challenges faced in overcoming the memory wall. While evolving cache-coherent interconnect technologies such as CXL can pave a path for shifting from CPU-centric computing to Memory-centric computing, there is a need to revisit and redesign the systems to address the performance, programmability, and efficiency challenges faced by these technologies in a unified manner.

The topics of interest for this proposal include, but are not limited to:

- Software, hardware and co-design approaches for efficient utilization of heterogeneous and unified memory systems, disaggregated shared/pooled memory systems, and processing-near-memory
- Memory-centric programming models, languages and optimizations for the above memory systems
- Memory Architecture for Large Language Model
- Compiler, runtime, and system techniques for optimizing the data placement (and compute, if applicable) and utilization of the above memory systems

- Software-Defined and Tiered memory systems and architectures
- Software-stack enhancements and hardware architectures to blur the boundary between memory and storage
- CPU and accelerator innovations to support heterogeneous memory technologies
- Performance scalability of large-memory systems and data movement reduction

Requirements : Proposals without detailed description of following items will NOT be reviewed for selection.

- Minimum number of Project PI/Co-PI : 2; means Project Investigator (PI) must consolidate a set of interrelated projects aimed at addressing complex and shared challenge together. PI should actively collaborate with diverse and complementary PIs, adopting a comprehensive approach that tackles the problem from various angles instead of using a compartmentalized approach
- Propose a tightly structured project which includes technical and project that demonstrates clear progress, are aggressive but achievable, and are quantitative (Objectives and Key Result(OKRs), Key performance indicator).
- Clearly define the merit of the proposed innovation compared to competing approaches and the anticipated outcome. Ideally, this will include quantitative projections for performance improvement that are tied to representative values included in authoritative publications. Proposals should incorporate an assessment of the prospective product, supported by relevant justifications and a strategic plan for possible commercialization.
- Budget must include overhead cost and specific percentage charging rate. The budget allocation is designed to provide support for a minimum of two project investigators and student researchers for a duration of one year, as outlined in the provided details. (Max $300,000/year)

Information

- Collaborative projects must be submitted as a single proposal in which a single award is being requested : The involvement of partner organizations should be supported through subawards administered by the submitting organization.
- The duration of the award is subject to variation depending on the type of resource being funded, typically ranging 2 to 3 years (or potentially up to 5

years). Nevertheless, an annual update on project milestones will be required.

- Limit on Number of Proposals per PI (or co-PI): 1 ; Each individual PI is limited to participate one proposal.

Our research interest and related publications can be found in MSL website. Topics related to our research will be highly considered.