Ethan Berkley

3113 West 29th Street
Lawrence, KS 66047
(785) 424-5457
ethantberkley@gmail.com

EDUCATION

University of Kansas, *1450 Jayhawk Blvd., Lawrence, KS 66045*
August 2021 - May 2025. 3.94 GPA

EXPERIENCE

Undergraduate Research Assistant, Cornell University 2025-Present

Investigating specialized microarchitectures to maximize accelerator-IO interconnect bandwidth for network-intensive AI workloads.

Undergraduate Research Assistant, University of Kansas 2024-2025

Profiling large-scale services to identify the performance impact of communication inefficiencies.

Student IT Technician, University of Kansas 2022-2024

Troubleshooting software and hardware issues, configuring IT equipment, and training new members of the team.

ACTIVITIES AND AWARDS

Presenter, ACE Center for Evolvable Computing	2024
Member, International Honor Society for Computing and Information Disciplines	2023-Present
President, Stephenson Scholarship Hall, University of Kansas	2023-2025
Social Chair, Stephenson Scholarship Hall, University of Kansas	2022-2023
Eagle Scout, Boy Scouts of America	2020

PUBLICATIONS

Ethan Berkley, Guanlin Zhu, Mohammad Alian, Under review, ISCA 2026.

SKILLS

- Experienced in conducting novel research in unexplored areas
- Highly experienced in relating architectural parameters to performance implications
- Experienced in data collection, analysis, and visualization
- Highly experienced in Go, Python, C/C++, Bash, and Linux
- Research ready familiarity with Kubernetes, Docker, and Intel VTune
- Excels at group work, experienced in leadership positions

COURSE WORK

- Datacenter Architecture
- Compiler Construction
- Algorithms For HPC
- Digital Systems Design
- Embedded ML
- Embedded Systems

PROJECTS

IO Specialization | 2025

- Conducting AI workload analyses to understand access patterns and interference between IO and SM/Tensor Core traffic to HBM.
- Researching potential NUMA-like memory allocation policies to efficiently route traffic through future accelerator packages with many chiplets.
- End goal will be to maximize on-package IO bandwidth to accelerate training and inference for future multi-trillion parameter models.

Microservice Fusion | 2025

- Creating new algorithms for cloud applications to identify RPC topologies for maximizing resource efficiency while guaranteeing quality of service.
- Applying techniques for search space reduction and analytical models for end-to-end performance prediction.
- Demonstrate the potential in defining an application's RPC topology during runtime rather than application development, a technique we refer to as microservice fusion.

Benchmarking Cloud Applications in Kubernetes | 2024

- Analyzes the performance of various demo applications under a modular monolith architecture, in which process boundaries can be explicitly defined at deploy time. This allows for communication between modules to occur via both direct method calls and remote procedure calls.
- Analysis is done within a Kubernetes cluster under varying process boundaries and hardware architectures to benchmark throughput under quality of service.
- Additional profiling is done using Intel VTune to quantify microarchitectural sources of delays, such as stalls in the processor frontend due to larger code footprints, or a larger proportion of network processing cycles for processes that make frequent RPCs.

Online Multi-Agent Pathfinding | 2025

- Modified a state of the art algorithm for multi-agent pathfinding to support adding new agents after an initial timestep, simulating a realistic online environment.
- Evaluated solution cost and latency for different pathfinding granularities.

TinyML Drone vs Bird Classification | 2024

- Deployed to a XIAO ESP32-S3 with on-device camera to perform inference.
- On-chip memory constraints and real-time latency requirements required careful design considerations, including methodologies such as post-training int-8 quantization.
- Used Tensorflow to deploy a model created with transfer learning based on MobileNetv2.