

Education

2016–2020 **B.S. in Computer Science**, *The University of Texas at Austin*, Austin TX.

Related Courses

Multicore OS, Concurrency, Parallel Computing, Operating Systems, Networks, AI, Neural Network, Computer Architecture, Algorithm, NLP

Research Interests

Operating Systems, GPUs, Heterogeneous Systems, Distributed Computing & Systems, ML Systems

Experience

2018 – **UTCS System Research Lab**, *Austin TX*, Researcher, ADVISOR: CHRISTOPHER ROSSBACH.

present The use of graphics processing units (GPUs) has become increasingly popular in accelerated computing in industry in recent years. Existing benchmark suites such as Rodinia were designed to better understand heterogeneous systems. But these tools failed to keep up with the rapid evolution of GPU. We develop a new benchmark suite to better demonstrate the diversity of GPU workloads, applications, and GPU resource usage, with emphasis on neural network workload characterizations and new CUDA features.

2016 – 2017 **UTCS AI Research Lab**, *Austin TX*, Student Researcher, ADVISOR: CEM TUTUM.

3-D printing has introduced many possibilities into the manufacturing industry. However, detecting object defection in real time requires expensive equipment inaccessible to many. With a team of four, we classified the types of error presented in 3-D printing. Because there were no previous work for reference, we delivered a framework to automate the process of data collection by alternating Gcode instructions, which facilitated the training of neural networks used to classify error type present in collected data.

June – Aug **H3C**, *Chengdu, China*, Software Engineering Intern.

- 2018 ○ Reconstructed mjpg-streamer to increase jpg streams efficiency in Linux by ten percent.
- Participated in the integration of Apollo autonomous vehicle driving system into H3C cloud platform (in Go).

June – Aug **Wisesoft**, *Chengdu, China*, Junior Software Engineer.

- 2017 ○ Implemented the neural network module (in Tensorflow) to process audio files using FFT(Fourier transformation). It is used in a production level system used to classify communication in air traffic control system.
- Constructed a tool set to automatically prune neural network models and organize training data files. Deployed into production systems.

Projects

Jan 2019 **GUPS on GPU**, *Random Memory Access on GPU*.

- Constructed random memory access program to stress the dram capability of GPU with several optimizations.

April 2019 **FDTD in Cuda Graph**, *Reimplementation in Cuda Graph*.

- Rewrote part of fdtd(Finite-difference time-domain method) using Cuda graph to test new Cuda features.

April 2019 **Multicore Operating System Implementation**, *a capability-based research OS by ETH Zurich*.

- Implemented physical memory management and process spawning based on Barrelfish kernel.
- Implemented message passing in LMP and RPC for processes to communicate in a multicore system.

Spring 2018 **Kmeans in CUDA MPI**, *Kmeans in NUMA*.

- Developed kmeans with focus on MPI communication on TACC super computer to test program scalability.

Spring 2018 **Thread Pool in C**, *a thread pool implemented in C*.

- Designed a thread pool with tasks mapped upon pthreads to simulate the design of Go routines.

Skills

Languages Python, C/C++, Java, Go, Rust

Frameworks OpenMP, MPI, PyTorch, Tensorflow, Cuda, Jekyll