Bodun Hu

CONTACT INFORMATION

E-mail: bodunhu@utexas.edu
Website: https://www.bodunhu.com

2317 Speedway
The University of Texas at Austin
Austin, TX 78712 USA

RESEARCH INTERESTS

Systems for ML, Operating System, heterogeneity, ML SW-HW Co-design, Distributed System

EDUCATION

The University of Texas at Austin

Ph.D. in Computer Science
Advisor: Aditya Akella

The University of Texas at Austin

M.S. in Computer Science, May 2021
Advisor: Christopher J. Rossbach

The University of Texas at Austin

B.S. in Computer Science, May 2020 (Research Distinction)

PUBLICATIONS

Liyan Tang, Grace Kim, Xinyu Zhao, Thom Lake, Wenxuan Ding, Fangcong Yin, Prasann Singhal, Manya Wadhwa, Zeyu Leo Liu, Zayne Sprague, Ramya Namuduri, **Bodun Hu**, Juan Diego Rodriguez, Puyuan Peng, Greg Durrett. ChartMuseum: Testing Visual Reasoning Capabilities of Large Vision-Language Models. *Preprint.*

Bodun Hu*, Luis Pabon*, Saurabh Agarwawl, Aditya Akella. Patchwork: A Unified Framework for RAG Serving. *Preprint.*

Bodun Hu, Shuozhe Li, Saurabh Agarwal, Myungjin Lee, Akshay Jajoo, Jiamin Li, Le Xu, Geon-Woo Kim, Donghyun Kim, Hong Xu, Amy Zhang, Aditya Akella. StitchLLM: Serving LLMs, One Block at a Time. *ACL 25.*

Ajay Jaiswal, **Bodun Hu**, Lu Yin, Yeonju Ro, Shiwei Liu, Tianlong Chen, Aditya Aeklla. FFN-SkipLLM: A Hidden Gem for Autoregressive Decoding with Adaptive Feed Forward Skipping. *EMNLP 24.*

Bodun Hu, Le Xu, Jeongyoon Moon, Neeraja J. Yadwadkar, Aditya Akella. MOSEL: Inference Serving Using Dynamic Modality Selection. *EMNLP 24.*

Henrique Fingler, Isha Tarte, Hangchen Yu, Ariel Szekely, **Bodun Hu**, Aditya Akella, Christopher J. Rossbach. Towards a Machine Learning-Assisted Kernel with LAKE. *ASPLOS 23.*

Bodun Hu and Christopher J. Rossbach. 2020. Altis: Modernizing GPGPU Benchmarks. *ISPASS 20.*

RESEARCH EXPERIENCE

Meta, Menlo Park, CA, USA.

Student Researcher **2025-Current**

Designed and optimized high-performance network stack to reduce latency in large language model (LLM) training.

Meta, Menlo Park, CA, USA.

PhD SWE Intern **2025**
Implement efficient communication collectives for LLM and recommendation models.

The University of Texas at Austin (UT Austin), Austin, TX, USA.

Research Assistant **2021 - Current**
Implement efficient inference serving systems for LLMs and GenAI.

Intel, San Jose, CA, USA.

P4 Dataplane Intern **2022**
TCP-INT: Improved Network Telemetry in TCP Transport for better e2e visibility and improved closed-loop control of TCP workloads.

The University of Texas at Austin (UT Austin), Austin, TX, USA.

Research Assistant **2017 - 2021**
LAKE: Built a generic API remoting system to expose accelerator APIs to OS kernel with close-to-native performances.
ALTIS: Designed a benchmark with improved diversity over existing GPU benchmarks by extending application domains with modern CUDA features.

The University of Texas at Austin (UT Austin), Austin, TX, USA.

Rearch Assistant **2020**
TAS: Ported TAS into P4 to facilitate TCP fast-path migration to programmable NICs.

The University of Texas at Austin (UT Austin), Austin, TX, USA.

Rearch Assistant **2016 - 2017**
G-Code-gen: Designed an automated detection system utilizing readily available hardware, which detects and terminates 3D printing processes upon identification of object defects.

Industry Experience

H3C, Chengdu, China.

Software Engineering Intern **2018**
Devised and implemented a highly effective caching strategy, resulting in a significant reduction of video streaming processing latency on Kubernetes cluster by a factor of 3x.

Wisesoft, Chengdu, China.

Software Engineering Intern **2017**
Developed a data preprocessing pipeline for improved audio classification in an air traffic control system.

Honors and Awards

Wesley W. Calhoun Jr. Endowed Scholarship

ISPASS Student Travel Award, 2020

Research Distinction by the College of Natural Sciences (UT Austin), 2020.

Teaching

CS395T: Advanced Topics in Systems and GenAI (graduate)
Teaching Assistant, UT Austin, Fall 2025

CS378: System For Machine Learning and Big Data (undergrad)
Teaching Assistant, UT Austin, Fall 2024

CS378: Multicore Operating System Implementation (undergrad)
Teaching Assistant, UT Austin, Spring 2020

TALKS

- *Altis: Modernizing GPGPU Benchmarking*, ISPASS'20 (August 2020)
- *Accelerating Kernel Access to Hardware Acceleration*, Texas Systems Symposium (November 2020)

SERVICE

- Reviewer for ACL 2025, ACL-SRW 2025
- Junior Graduate Admissions Committee, UT Austin (January 2021)