Edwin G. Lim

858-776-7854 | eglim@andrew.cmu.edu | https://edwinlim0919.github.io/

SUMMARY

My work focuses on the design of efficient computer systems and hardware for modern machine learning applications, addressing energy efficiency and chip utilization hurdles that arise as larger individual applications encompass increasing portions of the datacenter capacity.

Modern large-scale machine learning (ML) applications require extraordinarily large amounts of compute and memory capacity, and the gap between what models demand and what an individual hardware platform can provide continues to grow. As silicon technology scaling slows down, individual applications must be run on multiple massively parallel hardware platforms such as Graphics Processing Units (GPUs) or Tensor Processing Units (TPUs), with some large training applications spanning more than 10,000 GPUs. Thus, datacenters are limited by their allowable electricity grid power budget and the amount of competitive ML hardware platforms they are able to acquire. In this new paradigm, every joule of energy counts, and it is imperative to efficiently utilize hardware in the face of growing infrastructural overhead. My work approaches designing ML hardware and software infrastructure from a warehouse-scale perspective, in order to enable efficient and sustainable development of modern ML applications.

My work has been recognized with the Berkeley EECS Microelectronics Undergraduate Scholarship Award in both the 2021-2022 and 2022-2023 academic calendar years, the 2023 Carnegie Institute of Technology Dean's Fellowship, and the 2023-2024 Fritsch Family Fellowship in Electrical and Computer Engineering.

EDUCATION

Ph.D. in Electrical and Computer Engineering   Fall 2023 – Present
Carnegie Mellon University   *Pittsburgh, PA*

- GPA: 4.0
- Advisors: James C. Hoe, Akshitha Sriraman

B.S. in Electrical Engineering and Computer Science   Fall 2019 – Spring 2023
University of California, Berkeley   *Berkeley, CA*

- GPA: 3.925
- Advisor: Krste Asanović

PROFESSIONAL EXPERIENCE

Ph.D. Student Researcher   Fall 2023 – Present
Carnegie Mellon University   *Pittsburgh, PA*

- Analyzing the carbon emissions associated with machine learning inference applications and developing scheduling mechanisms to reduce their environmental impact

CPU Design Verification Intern   Summer 2023
Apple   *Santa Clara, CA*

- Built tooling to generate highly configurable, pseudo-random instruction sequences for verifying new architectural features of the CPU without significant manual test-writing effort

Undergraduate Student Researcher   Fall 2021 – Spring 2023
SLICE Lab, UC Berkeley   *Berkeley, CA*

- Enabled rapid and open-source development of hyperscale systems-on-chip through high-performance scalable ASIC emulation on multiple FPGAs, and explored the benefits of hardware accelerators for hyperscale applications

Architecture Intern   Summer 2022
NVIDIA   *Santa Clara, CA*

- Developed tooling within NVIDIA's functional GPU model to measure how frequently code exercises context switch functionality, allowing the context-switch team to downscale bloated testing suites

Software Engineering Intern   Summer 2021
REX Real Estate   *Redwood City, CA*

- Developed several RESTful APIs and backends that enabled core microservices to communicate quickly and efficiently

Software Engineering Intern
Summer 2020

WhatElse.io
Berkeley, CA

- Developed web APIs and backends to scrape and show relevant information about potential clients for marketers, and migrated APIs to AWS Lambda to enable serverless cloud functionality

Teaching and Mentoring Experience

Undergraduate Student Instructor
Spring 2023

CS 152 (Computer Architecture and Engineering), UC Berkeley
Berkeley, CA

- Led discussion sections and office hours to help students with concepts such as pipelining, memory hierarchy, out-of-order processors, superscalar processors, branch prediction, multithreading, vector ISAs, cache coherence, memory consistency, and synchronization

Undergraduate Student Instructor
Fall 2021, Fall 2022

CS 61C (Great Ideas in Computer Architecture), UC Berkeley
Berkeley, CA

- Led weekly discussion/lab sections and held office hours to help students with concepts such as systems programming, assembly language programming, digital design, CPU design, caches, virtual memory, and parallel programming

Course Tutor
Summer 2021, Spring 2022

CS 61C (Great Ideas in Computer Architecture), UC Berkeley
Berkeley, CA

- Led weekly tutoring sections and office hours in which I taught fundamental computer architecture and systems concepts through recitation slides and worksheets

Academic Intern
Spring 2021

CS 61C (Great Ideas in Computer Architecture), UC Berkeley
Berkeley, CA

- Helped students learn about fundamental computer architecture and systems topics by answering questions in Zoom chat and assisting in laboratory office hours

Director of Technical Operations
Summer 2020 – Spring 2021

IEEE UC Berkeley Student Branch
Berkeley, CA

- Fostered a more inclusive environment for EECS students during the COVID pandemic by creating and guiding technical projects, organizing hackathons, maintaining the club's server infrastructure, and leading weekly committee meetings

Honors and Awards

Fritsch Family Fellowship in Electrical and Computer Engineering
Spring 2024

Carnegie Mellon University
Pittsburgh, PA

- Awarded $9,500 towards tuition, stipend, and living expenses

Carnegie Institute of Technology Dean's Fellowship
Spring 2023

Carnegie Mellon University
Pittsburgh, PA

- Awarded $83,000 towards tuition, stipend, and travel expenses

Eta Kappa Nu
Spring 2022

UC Berkeley
Berkeley, CA

- Inducted into UC Berkeley's EECS honor society, with invitations extended to EECS majors in the top quarter of the junior class

EECS Microelectronics Undergraduate Scholarship Award
Spring 2021, Spring 2022

UC Berkeley
Berkeley, CA

- One of eight students to receive a $4,000 scholarship for undergraduate students planning to pursue a career in microelectronics in both the 2021–2022 and 2022–2023 academic years

Publications

Joonho Whangbo, **Edwin Lim**, Chengyi Lux Zhang, Kevin Anderson, Abraham Gonzalez, Raghav Gupta, Nivedha Krishnakumar, Sagar Karandikar, Borivoje Nikolic, Sophia Shao, Krste Asanovic. *FireAxe: Partitioned FPGA-Accelerated Simulation of Large-Scale RTL Designs.* International Symposium on Computer Architecture (**ISCA 2024**). July 2024. **Best Artifact Evaluation Award.**
Acceptance Rate: TBD
Presents the first open-source FPGA-accelerated RTL simulation platform that supports compiler-assisted RTL partitioning across multiple FPGAs for high-performance simulation of large-scale system-on-chip designs

Sagar Karandikar, Aniruddha N. Udipi, Junsun Choi, Joonho Whangbo, Jerry Zhao, Svilen Kanev, **Edwin Lim**, Jyrki Alakuijala, Vrishab Madduri, Yakun Sophia Shao, Borivoje Nikolić, Krste Asanović, Parthasarathy Ranganathan. *CDPU: Co-designing Compression and Decompression Processing Units for Hyperscale Systems.* International Symposium on Computer Architecture (**ISCA 2023**). July 2023.
Acceptance Rate: 17%
Presents the first open-source end-to-end design of hyperscale (de)compression units, a design-space-exploration of various accelerator integration strategies for datacenter SoCs, and an evaluation backed by industry hyperscale data

References

Akshitha Sriraman
Assistant Professor of ECE at Carnegie Mellon University
Pittsburgh, PA
☎ (215) 796-3381
akshitha@cmu.edu

James C. Hoe
Professor of ECE at Carnegie Mellon University
Pittsburgh, PA
☎ (412) 268-4259
jhoe@cmu.edu

Krste Asanović
Professor Emeritus of EECS at UC Berkeley
Berkeley, CA
☎ (510) 642-6505
krste@berkeley.edu

Si-En Chang
Engineering Manager at Apple
Santa Clara, CA
☎ (408) 807-8767
sien@apple.com