QUANG DUONG

duongquang1@gmail.com | quangduong.me | quangmire

EDUCATION

University of Texas at Austin ..... Fall 2015 - Present
- PhD in Computer Science [GPA: 4.0] ..... Fall 2020 - Present
- BS/MS in Computer Science + BS in Mathematics [GPA: 3.9482] ..... Fall 2015 - Spring 2020

Relevant Coursework
- **Graduate**: Advanced OS, Compilers, Programming Languages, Deep Learning Seminar, NLP Seminar, Machine Learning, NLP, Computational Linguistics, Automatic Speech Recognition
- **Undergraduate**: OS, Comp Arch Seminar, Comp Arch, CV, Neural Networks, Data Mining

EXPERIENCE

Graduate Research Assistant at UT Austin ..... Spring 2021 - Present
- Reduced the size ($>1000\times$) and latency ($>100\times$) of neural temporal prefetchers while improving generalization by reformulating the temporal prefetching prediction problem.
- Streamlined on-chip temporal prefetching to improve storage efficiency by 33%.
- Extracted salient branches from profiled neural branch predictors for compression.

Architecture Research / CPU Performance Intern at Arm ..... Summer '22 / '23 / '24 / '25
- Modulated front-end speculation to exchange performance for energy savings at a 1-to-20 ratio.
- Uncovered a 3% performance opportunity by accessing academic cache management policies.
- Provided double-digit traffic reduction by employing ML-based prefetcher throttling algorithms.

Graduate Research Assistant at Applied Research Laboratories ..... Fall 2019 - Summer 2020
- Proposed and implemented data format for 4-10$\times$ speedup in processing time.
- Migrated code base from Python 2 to 3 and wrapped legacy C++ code into Python libraries with SWIG and Cython to streamline data analysis.

Honors Scholar / Student Tech at Applied Research Laboratories .. Summer 2018 - Summer 2019
- Leveraged clustering and reinforcement learning techniques to iteratively refine estimations of ionospheric model parameters that were competitive with commercial software.
- Achieved real-time speed by adaptively sampling a fast surrogate model.

FRI Research Fellowship / TIDES Research Fellowship ..... Summer 2016 / Summer 2017
- Utilized genetic algorithms to approximate 3D models parametrically for non-expert modeling.

PUBLICATIONS

Streamlined On-Chip Temporal Prefetching **[HPCA]** 2026
- Quang Duong and Calvin Lin

A New Formulation of Neural Data Prefetching **[ISCA]** 2024
- Quang Duong, Akanksha Jain, and Calvin Lin

SKILLS

- **Languages**: Python, C/C++, Rust, HTML/CSS, Go, Java, JavaScript, x86, LaTeX, SQL, MATLAB, Julia
- **Libraries**: numpy, PyTorch, TensorFlow, sklearn, scipy, matplotlib, React, Flask, OpenGL
- **Comp Arch Simulation**: ChampSim, Dynamorio, gem5, SimPoint / LoopPoint
- **Other Skills**: Unix Systems (Ubuntu, Debian, Arch), Vietnamese

TEACHING

Co-Instructor for Graduate Prediction Mechanisms for Comp Arch .... Spring '22 / '23 / '24 / '25 / '26
TA for Graduate Parallel Systems ..... Fall 2020 / Summer 2021
TA for Undergraduate Software Engineering ..... Spring 2018
Mentor for Freshman Research Initiative ..... Spring 2017

ACTIVITIES

Artifact Evaluation Committee Member [MICRO] 2023
Artifact Evaluation Committee Member [PPoPP] 2022
Co-Organizer of Machine Learning Data Prefetching Competition at [ISCA/MLArchSys] 2021