

EDUCATION

University of Toronto

Sep. 2022 – Jun. 2024

Master of Engineering, Electrical & Computer Engineering

- **MEng Thesis (Individual): Pilot Training – Angle of Arrival and Channel Estimation in 5G Network** [Code]
- Compared Linear Minimum Mean Square Error (LMMSE) approach and Matrix-Pencil Method (MPM) for uplink channel estimation in multicell MIMO-OFDM systems with four distributed units (DUs) arranged in a square of radius 100 m, using a 3GPP-compliant MATLAB model.
- Showed that with **pilot training** in time or frequency domain, MPM yields higher angle-of-arrival accuracy than DFT-based and MMSE methods with limited antennas.
- Supervisor: Prof. Raviraj Adve

University of Toronto

Sep. 2017 – Jun. 2022

Bachelor of Applied Science, Electrical Engineering

- **Capstone Project (Teamwork): CNN NPU Overlay (MobileNetV1) on Intel Stratix 10 NX**
- Collaborated to replace the original matrix-vector unit of the Microsoft Brainwave NPU on Intel Stratix 10 NX with a **custom convolution unit (scheduler + tiles + accumulator)**, achieving 4.3× speedup and 76.6 GOP/s on MobileNetV1 inference.
- Collaborate to develop a **VLIW-based Python compiler** and **ISA simulator** to translate Keras models into NPU instructions, managing memory allocation and cycle-accurate performance verification.
- Designed a **pipelined accumulator** enabling 330 MHz operation; implemented three-stage **reduction-tree addition** and **opcode-based accumulation modes**, verified correct synchronization and 450 MHz Fmax in Quartus.
- Supervisors: Prof. Vaughn Betz, Andrew Boutros

PUBLICATIONS | PEER-REVIEWED CONFERENCES

1. Arash Ahmadian, Louis S.P. Liu, **Yue Fei**, Konstantinos N. Plataniotis, Mahdi S. Hosseini. *Pseudo-Inverted Bottleneck Convolution for DARTS Search Space*. IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), 2023.
2. Abnash Bassi, **Yue Fei**, Gilead Posluns, Mark C. Jeffrey. *Optimized Priority Scheduling for Faster Scalable Belief Propagation*. PROBABILISTIC GRAPHICAL MODELS (PGM), [In Submission], 2026.

EXPERIENCE

MediumAI · Freelance

Jan. 2026 – Present

Software Development Engineer — Hybrid, Toronto, Canada

- Build API integrations for a voice LLM-powered medical scribe agent, enabling secure multilingual clinical documentation.

iQIYI.com · Internship

Oct. 2025 – Dec. 2025

Data Analyst — Remote, Beijing, China

- Combined three-arm A/B testing with causal forests and decision-tree rules for deployable personalization, lifting renewal rates from 37% to 54.06%.

Qualcomm Canada · Permanent Full-time

Jun. 2024 – Jul. 2025

RFIC Design Verification Engineer — Onsite, Markham, Canada

- Verified UWB receiver path(analog front end); improved UWB receiver bootup by reducing LNA charging delay 90% (20 ns→2 ns); Verified WLAN CP-PLL synthesizer across 500+ channels using UVM, ensuring coverage over 2G/5G bands.
- Developed GRU-based multi-head model for analog gain line-up optimization and an MLP for VCO capacitance prediction, replacing manual tuning across 1000+ frequency targets.

University of Toronto – WIRLab · Contract Part-time

Jul. 2021 – Sep. 2021

Research Assistant — Remote, Toronto, Canada

- Explored RNN-based Transformer models and Attention mechanisms for NLP tasks.[Code]
- Applied PCA-based embedding visualization—as used in GloVe—to project high-dimensional embeddings into 2D/3D space using Python (NumPy, Matplotlib), enabling intuitive inspection of semantic clusters.

Alphawave Semi · Internship

May 2020 – Jun. 2021

SerDes Digital Verification Engineer — Remote, Toronto, Canada

- Built UVM testbenches for SerDes subsystems (clocking, datapath, SRAM), expanding coverage to 50+ scenarios; Scaled Jenkins CI/CD regression from 4 to 60+ projects, enabling 15× throughput growth.

AWARDS AND CERTIFICATES

- Edward S. Rogers Sr. Department Betz Entrance Scholarship (\$5,000) 2017
- Certificate in Engineering Business Jun. 2022

INVITED TALKS

Qualcomm Panel Discussion: *Demystifying Machine Learning* — QWomen San Diego Mar. 2025
Talk: *From Channels to States — Machine Learning in the Language of Communication and Control*

SELECTED ACADEMIC PROJECTS

COMPUTER ARCHITECTURE & SYSTEMS

Parallel Beamforming & Cache-Coherence Jan. 2024 – May. 2024

- Accelerated medical imaging by $7\times$ ($17\text{ s} \rightarrow 2.5\text{ s}$) via **16-thread data-parallel** ultrasound beamforming with Intel SIMD intrinsics and memory optimizations (`restrict`); validated correctness (RMS error $< 10^{-16}$) and scalability (1–16 threads). [\[Code\]](#)
- Designed and verified a **3-hop directory cache-coherence protocol** (MSI/MESI) in $\text{Mur}\phi$, optimizing with an **Exclusive (E)** state to eliminate unnecessary bus transactions. [\[Code\]](#)

Computer Architecture Coursework (Team of 2) Sep. 2021 – Dec. 2021

- Modeled **data hazard detection** in a modified MIPS 6-stage pipelined CPU, validating correctness via explicit stall-cycle annotation.
- Designed a **perceptron-based branch predictor**, combining local branch history with long global history to reduce MPKI under tight storage constraints.
- Implemented **Tomasulo out-of-order execution** with register renaming and dynamic scheduling, and an **MSI directory-based cache-coherence protocol**.
- Implemented a **Bouquet L1 data prefetcher** using IPCP, DPT, and a global history buffer (GHB) with confidence-based arbitration.
 - Evaluated feasibility using CACTI, showing $\sim 24\%$ area overhead and parallel access within $< 1\text{ ns}$ latency.

ERROR-CORRECTING CODES & GRAPHICAL INFERENCE

Graph-Based Error-Correcting Codes Sep. 2023 – Dec. 2023

- Implemented LDPC, fountain/LT, and Polar encoders/decoders over the binary-erasure channel, developing custom simulators in Julia and MATLAB. [\[Code\]](#)

Error Control Coding: Convolutional & BCH Codes Jan. 2023 – May. 2023

- Implemented **convolutional encoders** with Viterbi decoding and **binary BCH encoders/decoders** over $GF(2^m)$ using custom trellis structures in MATLAB.

INFORMATION THEORY & SEMANTIC COMMUNICATIONS

Sampling-Based Semantic Source Coding (One-Shot Information Theory) Jan. 2023 – May. 2023

- Implemented **Poisson functional representation**, **rejection sampling**, **importance sampling**, and hybrid Poisson + dithered quantization for 6G semantic source coding in MATLAB. [\[Code\]](#)

WIRELESS COMMUNICATIONS & SIGNAL PROCESSING

LTE Signal Processing Jan. 2022 – Apr. 2022

- Processed captured LTE signals for **time/frequency synchronization**, **OFDM demodulation**, and **channel equalization**; resampled 40 MHz RF data to the 30.72 MHz LTE rate and validated PSS/SSS detection.