Youssef A. Elhagrasy

Vancouver, Canada
youssefelbaz32@gmail.com — LinkedIn — GitHub — Portfolio — Google Scholar

Research Interests

Algorithm-hardware co-design for efficient AI; accelerator memory systems and data movement; ASIC/RTL design and formal verification; semiconductor reliability and test; energy measurement and attribution for AI systems; photonic interconnects for computing.

Education

University of British Columbia (UBC), Vancouver, Canada — Sep 2022 – May 2027
Bachelor of Applied Science (BASc), Engineering Physics; Specialization: Electrical and Computer Engineering
Dean's Honour List; **Cumulative Average: 90.4%**
Relevant Coursework: Digital Systems Design; Digital Systems and Microcomputers; Introduction to VLSI Systems; Analog CMOS Integrated Circuit Design; Machine Learning and Data Mining; Electromagnetic Theory; Introduction to Quantum Mechanics; Systems and Control
In Progress: From Markov Decision Processes to Deep Reinforcement Learning (graduate course)
Starting January 2027: Advanced Machine Learning; Optics; Statistical Mechanics

Publications

- **Y. A. Elhagrasy**, I. Hill, and A. Ivanov, "Adaptive Sequential Test Planning for Multi-Mechanism Reliability Qualification via Bayesian Monte Carlo Tree Search," *IEEE Transactions on Reliability*, accepted, 2026. DOI. **First author.**
- **Y. A. Elhagrasy**, et al., first-author paper on where skipped multiply-accumulates pay off, from the processing element to LLM decoding (title withheld during double-blind review), submitted to *MLSys*, 2027. **First author.**
- I. Hill, R. Vázquez Guerrero, **Y. A. Elhagrasy**, R. Jalilian, P. Moheban, M. Taylor, and A. Ivanov, "Wear-out Sensor SoC Peripheral for Continuous Semiconductor Reliability Monitoring, Lifespan Prediction, and Failure Risk Management," *IEEE Transactions on Device and Materials Reliability*, under review, 2026.
- X. Xu, S. Thomas, T. Guo, L. Luo, Y. Khan, Y. Yuan, **Y. A. Elhagrasy**, M. Lanza, T. D. Anthopoulos, O. M. Bakr, O. F. Mohammed, and H. N. Alshareef, "Wafer-Scale Transfer of MXene Films with Enhanced Device Performance via 2D Liquid Intercalation," *Advanced Materials*, vol. 36, no. 51, 2024. DOI.

Industry Experience

Microsoft — Silicon Engineering Intern — May 2026 – Aug 2026
Cobalt RTL Design Team, PCIe/CXL Bridge Infrastructure
- Owned RTL features and led JasperGold formal signoff for a CHI-/AXI-compliant CXL retry/credit block, reaching **99% checker COI and 99% proof-core coverage**.
- Built agentic CDC/RDC and formal-environment workflows that reduced turnaround **5×** and cut bring-up from **2 days to 3 hours**.
- Improved three critical timing paths by up to **40 ps** with physical-design teams and root-caused a CSR parity architecture flaw that could have caused a reliability escape.

Tesla — Intern, Silicon & Autopilot Systems Validation — Sep 2025 – Jan 2026
- Developed and executed validation plans for Autopilot and MCU SoCs, PCBAs, and in-vehicle computers across functional, performance, and reliability requirements.
- Built automated test infrastructure in **C/C++ and Python** covering approximately **50 scenarios**; performed firmware flashing, bootloader configuration, and cross-layer debugging with lab measurements.

Microsoft — Silicon Engineering Intern — May 2025 – Aug 2025
Azure Hardware Systems & Infrastructure
- Characterized wafer-level test data across **10M+ datapoints** for custom silicon and gave layout- and timing-related findings to physical design teams.
- Designed an internal **RAG, knowledge graph, and LangGraph agent** for hardware code retrieval and reasoning, achieving a **3× speedup**.

Microchip Technology — Product Engineer Co-op, SERDES Characterization — Jan 2024 – Apr 2024
- Supported Python-based transmitter-jitter characterization and lab measurements for high-speed SERDES devices, including oscilloscope measurements and VCO-coupling analysis for mixed-signal design teams.

TRIUMF — Firmware & Simulation Research Intern — May 2023 – Aug 2023
- Built **C++/Geant4** simulations of particle trajectories, energy deposition, and detector response for the Scintillating Bubble Chamber dark matter experiment, with TRIUMF, SNOLAB, and Fermilab; analyzed systematic error sources.
- Selected as **1 of 3 students from 2000+ applicants**.

Research Experience

UBC — Undergraduate Researcher, Semiconductor Reliability & Test Feb 2023 – Present

Supervisor: Prof. André Ivanov

- Formulated adaptive reliability qualification across BTI, electromigration, and TDDB as a sequential decision problem; built a Bayesian MCTS-SA/EKF planner that improved yield from **20.4% to 54.2% over 5,000 iterations** (first-author paper, *IEEE Transactions on Reliability*).
- Developed BayeShield, a risk-constrained Bayesian diagnosis framework spanning **249,389 reachable belief states** and **3 binding risk ceilings**.
- Designed an **IEEE 1500-compliant DFT wrapper** around a data-acquisition IP and verified serial/parallel scan access, functional isolation, and test-mode behavior using SystemVerilog and UVM; co-authored the wear-out sensor SoC peripheral paper.

UBC — Undergraduate Researcher, Efficient DNN Inference Jan 2026 – Present

Supervisor: Prof. Brad Quinton

- Led the project, combining learned activation thresholds with weight masks to skip multiply-accumulates, and studied where skips pay off in a Sky130 processing element through signoff, an RTL-checked CNN accelerator, and LLM decoding on an A100.
- Derived break-even rules for when to clock-gate, compress weights, skip weight fetches, or compact (first-author paper submitted to *MLSys*).

UBC — Undergraduate Researcher, Server Energy Systems Sep 2026 – Present

Supervisor: Prof. Mohammad Shahrad

- Investigating server-level energy tracking and attribution for cloud and serverless workloads, connecting application activity with OS scheduling, resource allocation, and hardware energy measurements.
- Studying how fractional-CPU enforcement, concurrency, and workload duration affect resource utilization, performance, and energy accounting in multi-tenant systems.

UBC — Undergraduate Researcher, Vision Transformers Sep 2026 – Present

Supervisor: Prof. Abdelrahman Eldesokey

- Investigating how patch-token features in Vision Transformers encode semantic and spatial structure and transfer under distribution shift; exploring self-supervised representations for visual correspondence.

KAUST — Transistor Research Intern, WISER 2023 Jul 2023 – Aug 2023

Supervisors: Prof. Husam AlShareef, Prof. Xiangming Xu

- Fabricated and characterized **MoS_2 and MXene** transistors via 2D thin-film coating for transparent electronics, using Design of Experiments, ALD, and chemical/dry etching.
- Analyzed electrical traces to study performance variability, achieving approximately **10× improvement** in device behavior; ran tests and analysis and wrote much of the methodology for the resulting *Advanced Materials* paper.

KAUST / TRYSL Tech — Hardware Intern, WISER 2022 Jul 2022 – Aug 2022

- Designed an **8-bit CPU** and microprocessor subsystems in SystemVerilog; developed RTL modules and verification testbenches in Synopsys VCS, reaching approximately **70% functional coverage**.

Selected Projects

Attention Accelerator: CUDA to RTL/FPGA 2026 – Present

- Built scaled dot-product attention across float64/Q8.8 Python models, optimized CUDA kernels, and two synthesizable SystemVerilog architectures.
- Replaced materialized attention with online softmax, reducing simulated RTL latency from **659 to 240 cycles** and intermediate storage from approximately **64 KB to 66 words** at $N = 128, D_V = 64$.
- Packaged the core as AXI4-Stream/DMA IP closing at **150 MHz**; optimized CUDA reached **3.86 TFLOP/s** and a **12.7× speedup** at $N = 4096, D = 128$. Project page.

Accelerator Memory-System Simulator 2026

- Built a CI-gated, approximately **2.6K-line Python simulator** for scratchpad banking, tensor layouts, and memory traffic; swept **6,720 configurations** and measured a **3.0× median traffic penalty** from mismatched tensor layouts.

Amazon Robotics Capstone: Optical Super-Resolution Barcode Scanner 2025 – 2026

- Co-developed a four-frame subpixel imaging pipeline with synchronized optics, motion, and camera control; calibrated shifts to **0.01–0.02 px repeatability** and below **0.05 px cross-axis error**.
- Automated PSF/MTF and ZXing-C++ testing across **2–10 mil** Code 128 targets and isolated a **3.1× vibration-induced jitter increase**.

Autonomous Driving & Sign Reading (ROS/Gazebo) 2024

- Built a robot that drives a four-zone course with 120 FPS vision and PID control and reads clue boards with a CNN (99.8% validation accuracy), scoring 57/57 points in about 40 s. With Raymond Wang.

Optimized Serial Adder 2024

- Designed a transistor-level datapath and FSM in Cadence Virtuoso (45 nm GPDK); applied logical effort for **2.4× throughput** (**58% lower critical-path delay**) and completed a DRC/LVS-clean layout.

FPGA Neural-Network Accelerator 2024

- Mapped a pretrained neural network onto a Zynq UltraScale+ FPGA using Verilog, BRAM-backed weights, parallel MAC arrays, and an AXI interface, reaching **0.5 ms/image** and **3× faster inference** than the ARM software baseline.

Autonomous Multi-Robot System 2024

- Designed H-bridge, power/current-sensing, reflectance-sensor, and actuator-control PCBs; implemented ESP32 C++ firmware, real-time control, and ESP-NOW coordination, placing **1st of 15 teams**.

Technical Skills

Languages: Python, C/C++, CUDA, SystemVerilog, Verilog, MATLAB, Tcl/Bash
Hardware: RTL design, SVA, JasperGold, UVM, DFT, CDC/RDC, FPGA, AXI, fixed-point arithmetic, Cadence Virtuoso, Vivado, DRC/LVS, probe stations, oscilloscopes
ML and Systems: PyTorch, Triton, pruning/sparsity, MCTS, Bayesian estimation, Vision Transformers, Linux, accelerator simulation, Nsight Compute, Git/CI

Honors & Awards

Faculty of Applied Science International Student of the Year (2×); Dean's Honour List (4×); Eric Vogt Scholar; International Major Entrance Scholarship; Outstanding International Student Award; Saudi RSI Top Student & Presenter