Pasan Sanjula Perera

Phone: +94 77 590 8445 — Email: rps.perera@outlook.com
Web: https://pasxn.github.io — Github: https://github.com/pasxn
LinkedIn: https://www.linkedin.com/in/pasansperera/

EXPERTISE

Hardware Performance Modeling, Computer Architecture, Machine Learning, Parallel Computing, Hardware Acceleration, C++ Development.

TECHNICAL STRENGTHS

Programming Languages	*C, C++, Python, Java*
Build Systems	*make, CMake*
Debugging	*GDB, Valgrind*
Unit Testing	*GoogleTest, pytest*
Continuous Integration	*Github Workflows*
Version Control	*git, Github*
Tools & Frameworks	*TVM, PyTorch, OpenCV*
Operating Systems	*Unix/Linux*

EDUCATION

Sri Lanka Institute of Information Technology *2020 - 2023*
B.S. (Hons) Electrical and Electronic Engineering
Grade: Second Class, Upper Division

Ananda College, Colombo 10 *2011 - 2019*
G.C.E Advanced Level Examination
Physical Science

EXPERIENCE

AxPro Semi

Engineering Consultant *Feb. 2024 - Present*

- Designing & developing a cycle-accurate microarchitectural simulator for AxPro SoC to evaluate architectural decisions.
- Developed a C++ microarchitectural simulator supporting configurable, in-order & out-of-order, super-scalar pipelines, cache hierarchy & DRAM simulation, AxPro propitiatory SoC components, & programmable logic, capable of booting Linux & running custom kernels in programmable logic.
- Developed matmul & softmax kernels for AxPro, involving microarchitectural design for programmable logic, & host programs for software use; later integrated with PyTorch to run LLMs.
- Developed & integrated a C++ library for generating & visualizing SoC state transitions as .vcd files, with configurable signals & timeframes; extended the library for compressing waveforms to .fst using shared memory & multiprocessing .
- Developed & integrated a visualization tool for simulator pipeline & memory hierarchy in Python.
- Built a tool to visualize memory access patterns in Python, integrated with the simulator.
- Implemented a regression suite which runs in two environments for automated execution (host os & simulator os), including a visualization tool, & integrated it into CI; established CI in the company, wrote GitHub workflows.
- Set coding standards, refactored open-source code used within, enforced formatting with .clang-format & pre-commit hooks; reviewed code & guided the application team on simulator use.

ACCELR, *www.accelr.lk*

Software Engineer *Nov. 2023 - Present*

- Augmented to AxPro Semi working on a RISC-V CPU design as a CPU Performance Architect.
- Built a Python data pipeline for web crawling & text collection, with multi-stage cleaning & human feedback for quality datasets; used MongoDB for storage, was used in a RAG application.
- Created a C++ library to decode proprietary stock market data from ethernet frames for verifying HFT FPGA kernels.
- Developed a TVM-based ML inference pipeline with benchmarking for RISCV64GC, verified on QEMU (https://github.com/accelr-net/tvm-riscv-demo).

ACCELR, *www.accelr.lk*

Software Engineering Intern *Nov. 2022- Jan. 2023*

- Worked with a team of FPGA developers tasked with building a Solr/Lucene hardware accelerator for a stealth mode US startup.
- Developed & maintained the CocoTB based verification framework used in the project.
- Developed Verilog RTL modules based on C++ hardware reference model built by senior engineers.
- Involved in testing, debugging & fixing the C++ reference model & RTL design throughout the period.
- Involved in developing several Python scripts to dump & analyze intermediate data generated by Apache Solr.
- Developed a test tool to perform bench-marking tests to compare vanilla Solr & the FPGA accelerated Solr versions.

ACCELR, *www.accelr.lk*

Software Engineering Intern *Nov. 2021- Jan. 2022*

- Worked with a team from USA, India & Sri Lanka, building a Hardware Accelerated Computing solution for an open source search engine software library.
- Developed a hardware model for a hash function using C/C++ & integrated with the existing Java code base using JNI.
- Created an OpenCL kernel for the same function to execute on an FPGA using Xilinx Vitis SDK.
- Wrote Python scripts to automate extracting textual data from CSV & PDF files, reformatting & writing to JSON files.

Bot Labs

Research & Development Intern *Jun. 2021- Jul. 2021*

- Contributed to research & development of the core architecture of an upcoming product.
- Developed various proof-of-concepts using C/C++, JavaScript & WebAssembly with OpenGL.

SELECTED TECHNICAL PROJECTS

A GPU accelerated ML inference framework for the RPi *February 2023 - October 2023*

Led a team of three to design & develop a GPU-accelerated ML inference framework for Raspberry Pi, including a tensor computation library, neural network blocks, & a GPU kernel pipeline, using Python & C/C++. Designed a new algorithm to traverse ONNX graphs in runtime efficiently by pre-computing parts in compile time. Guided team technically & reviewed contributions.

Presenter Tracking Camera *February 2022 - October 2022*

Led a team of three to design & develop a low-cost presenter tracking camera for hybrid teaching, using classical CV & Kalman filters for real-time detection & tracking. Developed alpha software in Python & firmware in C/C++, peer reviewed code.