

VARUN PAL

408-748-4386 | varunpal.info | vpal2@illinois.edu

EDUCATION

University of Illinois Urbana-Champaign

Masters of Science in Computer Engineering

Bachelor of Science in Computer Engineering

– Technical GPA: **3.93**

– Relevant Coursework:

Operating Systems

Data Structures and Algs

Control Systems

Digital Systems

Applied Parallel Programming

Machine Learning

Urbana-Champaign, IL

Expected Graduation: Fall 2027

Expected Graduation: Spring 2026

Distributed Systems

Intro to Robotics

Deep Learning For Computer Vision

WORK EXPERIENCE

Cisco Systems

May 2026 – Aug 2026

Software Engineering Intern

San Jose, CA

- Re-engineered the optic-transceiver dashboard's **MongoDB analytics pipeline** over **8.4M sensor documents**, replacing regex-heavy scans with a **Docker/cron** materialization job storing **change-triggered aggregate snapshots** for **O(1) date-range lookups** — **~100ms queries vs. 16 min** raw scans with **13x storage reduction**
- Cut chart generation time **10x** via **projection-scoped queries**, **server-side aggregation**, and batched lookups; built peer-optic overlays from an **NX-OS SSH/CDP/LLDP** collector mapping **689 peer links**
- Implemented **Pydantic validation layer** enforcing typed schemas and sensor bounds at repository boundaries, catching **2 critical bugs** in upstream ingestion; fixed lane/breakout-port modeling across **3K+ optics**
- Built reproducible **scikit-learn** pipeline evaluating peer telemetry for next-day optical forecasts; **leakage-resistant** splits and a **2K-iteration block bootstrap** over **1,284** samples rejected the feature (95% CI crossed zero)

RoboVision Lab

May 2024 – Aug 2025

Summer Researcher

University of Illinois Urbana-Champaign, IL

- Researched a generalizable robotic pick and place task diffusion policy under the guidance of **Prof Saurabh Gupta**
- Trained neural net policy utilizing **self and cross attention** on DROID data set in a **transformer architecture**
- Implemented a **data-distributed parallel system**, **improving GPU utilization** from 30% to 70% avg
- Developed system for **filtering and processing large video datasets** (1.7 TB) for training robotic ML policies

ECE 391 (Operating Systems) Head TA

Aug 2026 – Present

Head Teaching Assistant

University of Illinois Urbana-Champaign, IL

- Head TA for **300-student OS** course; direct **5 TAs** and **30 CAs** across exam design, MP releases, and grading

Course Assistant (Jan 2025 – Aug 2026)

- Developed core Unix-like OS components (**Shell, Pipes**) for student projects; hold **6 weekly office hours**

Abbott

May – Sep 2023

Software Engineering Intern

Santa Clara, CA

- Designed an efficient **data parsing algorithm** using **chunking techniques** to process **large text data** (> 4GB) from glucose monitoring devices, enabling faster analysis of product performance metrics
- Created a **data visualization dashboard** using **Power BI**, providing real-time insights into product testing results

PROJECT EXPERIENCE

Out-of-Order RISC-V CPU (ECE 411) | SystemVerilog, Computer Architecture, Synopsys

Spring 2026

- Built **7-stage out-of-order RV32IM** CPU with **explicit register renaming**; synthesized at **645 MHz** / **~200K area**
- Implemented **2-way superscalar** front-end through commit and a **split LSQ** with one-hot age-tagged out-of-order load issue between older stores — **+8.3% geo-mean IPC** across 6 benchmarks
- Built **gshare branch predictor** with **64-entry BTB** achieving **81.3% branch accuracy** (vs. 47.5% baseline)

OStrich | Operating System, RISC-V, Embedded Systems

Fall 2024

- Placed **3rd out of 100** teams in OS design competition
- Implemented Unix-based OS with **paging virtual memory**, **multi-thread processes**, and **file system** on RISC-V
- Developed **UART and VGA drivers**, **interrupt handling** and **real-time scheduling**

CUDA CNN Convolution Kernel Optimization | CUDA C/C++

Spring 2025

- Optimized convolution kernel for LeNet-5 inference on Fashion MNIST dataset, implementing **matrix unrolling**, **tiled matrix multiplication**, and **kernel fusion** to achieve sub50ms run time on 10,000 images on A100 GPUs
- Optimizations include **CUDA Streams**, **Tensor Cores** with Warp Matrix Functions, and FP16 arithmetic

TECHNICAL SKILLS

Languages: Python, Go, C, C++, CUDA C/C++, SystemVerilog

Systems & Databases: MongoDB, Docker, Redis, Linux, Raft Consensus, Two-Phase Commit, ACID Transactions

Frameworks & Tools: Plotly Dash, PyTorch (DDP), scikit-learn, Pandas, Pydantic, Git, CUDA (Nsight), Performance Profiling