Nitin Vunnam

(469) 679-4553 | nitinvunnam06@gmail.com | linkedin.com/in/nitin-vunnam | github.com/nitinvunnam | US Citizen

Education

The University of Texas at Austin — Austin, TX
Bachelor of Science in Electrical and Computer Engineering — *May 2028*

- GPA: 3.70/4.00
- Relevant Coursework: Computer Architecture, Multithreading Programs/Architectures, VLSI, Operating Systems, Digital Logic Design, Embedded Systems, Software Design, Circuit Theory

Experience

Amazon Web Services — Seattle, WA
Software Development Engineer Intern — *Jun. 2026 – Aug. 2026*

- Shipped an artifact replication monitoring system for performing 2M+ daily replications across 250K+ pipelines
- Built the path from datastore to browser, including new APIs with targeted index lookups to prevent throttling
- Architected a cross-partition solution to show air-gapped replication progress, creating a diode-based data path
- Resolved distributed race conditions and premature-completion bugs that caused incorrect replication reporting
- Extended artifact replication APIs using atomic DynamoDB counters and persisted denominators for monitoring

CuPy — Austin, TX
Open Source Contributor — *Aug. 2025 – Oct. 2025*

- Developed fixes and regression tests for quantile and percentile operations on empty arrays and zero-length axes
- Helped prevent illegal GPU memory accesses in statistical kernels by strengthening boundary-condition validation
- Improved nanmedian edge-case handling for zero-dimensional arrays and axis validation to align with NumPy

Projects

Pipelined LC3-b Microarchitecture Simulator | *C, Assembly*

- Engineered a cycle-accurate LC-3b CPU simulator in C, spanning both single-cycle and 5-stage pipelined designs
- Modeled the full ISA datapath, control store, and microsequencer to simulate processor behavior at the cycle level
- Implemented pipeline hazards, control and data stalls, cache access timing, and per cycle state propagation logic
- Implemented interrupt and exception handling via new microcode, control sequencing, and datapath modifications
- Extended the microarch by adding virtual to physical address translation, page table walks, and protection checks

Parallel Sparse Matrix Solver Optimization | *C++, OpenMP, CUDA, HPC*

- Developed a CPU/GPU sparse matrix solver using OpenMP and CUDA to accelerate large scientific simulations
- Implemented RCM graph reordering and CSR sparse matrix operations to optimize memory locality and execution
- Benchmarked CPU/GPU implementations, analyzing scaling, memory access patterns, and reordering overhead

Concurrent Data Structures & Synchronization | *C++, Multithreading, Linux Perf, TSan*

- Built thread-safe maps and tabless using lock sharding, striping, and cache-line padding for multicore scalability
- Implemented TAS, TTAS, ticket, parking, and reader-writer locks from atomics with spin backoff and wait/notify
- Scaled sharded-map throughput from 3.0 to 51.7 Mops/s across 56 cores on TACC, achieving a 17.2× speedup
- Analyzed IPC, cache misses, and HITM coherence traffic with Linux perf to identify and fix multiple bottlenecks

Pintos Operating System | *C, x86, Kernel Development*

- Extended the PintOS kernel to support user program execution, process creation, termination, and synchronization
- Implemented system calls with user pointer validation, file descriptor management, and synchronized file I/O
- Developed parent-child synchronization to manage process lifecycles, propagate exit statuses, and reclaim resources

GPU Matrix Multiplication Optimization | *C++, CUDA, NVIDIA Nsight Compute, cuBLAS*

- Developed CUDA SGEMM kernels with coalesced memory accesses, shared-memory tiling, & 2D register blocking
- Analyzed memory bandwidth, cache behavior, register pressure, and occupancy using NVIDIA Nsight Compute
- Optimized thread-block configurations and tile sizes, benchmarking kernel throughput against cuBLAS on A100s

Technical Skills

Languages: C/C++, Python, Verilog, CUDA, Assembly, Java, JavaScript, TypeScript, HTML/CSS
Frameworks: NVIDIA NSight, OpenMP, Linux perf, Linux kernel, React.js, React Native, Node.js, Next.js, Flask
Developer Tools: FPGA, Vivado, Cadence, Synopsys, KiCad, Altium, LabVIEW, Valgrind, ARM Cortex-M0+