# Apoorva Karnik

515-528-4007 • apoorvakarnik@gmail.com • www.linkedin.com/in/apoorva-karnik-a1830738/

Platform Architect & Technical Manager with 12+ years of experience in high-performance CPU performance modeling, pre & post silicon performance evaluation, HW-SW co-optimization for CPU/GPU workloads.

## WORK EXPERIENCE

**Platform Architect (Engineering Manager) | CPU Workloads** ***(Platform Architecture, Apple, Cupertino)*** **Jan 2024 – Present**

- Managed an engineering team focused on end-to-end performance characterization of CPU/GPU workloads, driving HW-SW co-design and systemic optimizations
- Partnered with cross-functional teams to resolve low-level system bottlenecks and refine control loops for optimal performance and efficiency
- Characterized real-world workloads, including GenAI (MLX, PyTorch, llama.cpp), CAD applications, and industry benchmarks (SPEC, GeekBench), using HW/SW profiling to surface latency, throughput, scaling, and memory-system "glass-jaw" limitations, and drove software, algorithmic, or architectural changes
- Developed a novel trace-driven simulation framework to accurately model multi-processing workloads and evaluate their performance on future SoC generations

**Engineering Manager | Performance Verification** ***(Silicon Engineering Group, Apple, Cupertino)*** **July 2019 – Jan 2024**

- Led pre- and post-silicon performance validation for CPU, memory subsystem & multi-core performance across the entire range of Apple SoCs
- Owned performance verification for Apple's Matrix Compute Engine (AMX); developed directed performance tests (GEMM and GEMV kernels of different size and types) to ensure HW met performance targets
- Influenced the architecture of Performance Monitoring Unit (PMU) and 'always-on' control loop counters, defining the hardware telemetry required for real-time control feedback loops
- Partnered with RTL and modeling teams to bridge the gap between architectural intent and silicon implementation
- Spearheaded several post-silicon tuning of hardware "widgets" and tunables, optimizing silicon to extract maximum performance and efficiency
- Collaborated with System Software to calibrate thermal and power control loops, maximizing sustained performance across varying operating points

**Senior Performance Verification Engineer** ***(Silicon Engineering Group, Apple, Cupertino)*** **Sept 2016 – July 2019**

- Created a lightweight kernel to run single/multi-core performance tests (C/Assembly) on bare metal
- Worked on C/C++ based performance model for last level caches, matrix-engine and several other features
- Drove significant changes to ARM atomics implementation for Apple CPUs; created multi-processor assembly tests to evaluate performance of several ARM synchronization primitives including atomics and exclusive instructions
- Helped root cause performance issues seen in iOS, Application software and suggest workarounds
- Contributed to developing a software tuning guide for application developers

**CPU Performance Verification Engineer** ***(Silicon Engineering Group, Apple, Cupertino)*** **Aug 2014 - Sept 2016**

- Worked on a C/C++ based performance model for state-of-the-art CPU, specially focusing on CPU back-end (Load store unit, Prefetchers, Caches and interface to memory subsystem)
- Modeled CPU features related to memory barriers, LD/ST exclusives, and atomics
- Wrote a cycle accurate load-store dependency shadow model in C++
- Wrote micro benchmarks in C/Assembly to analyze/evaluate the system performance and suggested improvements
- Analyzed iOS application workloads for performance improvements and suggested improvements to design

**Performance Modeling Engineer** ***(Qualcomm, Raleigh)*** **June 2013 - Aug 2014**

- Developed a C/C++ based performance model for a complex interconnect system including core complex, memory controller, DRAM and other peripherals
- Wrote micro benchmarks for analyzing system latency, BW and suggested improvements to design
- Analyzed workloads in SPEC2k, SPEC2k6 to suggest improvements to the interconnect subsystem
- Implemented various design features for QoS guarantees, LLT (low latency traffic) latency/BW improvements

**Functional Analyst Intern** ***(Deutsche Bank Global Technologies, Raleigh)*** **June 2012 - Aug 2012**

- Analyzed the flow and working of eSpear (Trade Life Cycle Monitoring System) to suggest speed improvements
- Worked with the business team to prepare a functional requirement spec and worked on implementation

## EDUCATION

**Master of Science in Computer Networks** **4.0/4.0**
North Carolina State University, Raleigh NC 2013
**Bachelor of Electronics & Telecommunication** **74%**
Nagpur University, India 2010

## RELEVANT ACADEMIC COURSES

**Graduate**: Advanced Computer Design and Technology, Computer Performance modelling, Architecture of Parallel Computers, Operating Systems, Computer Networks, Internet Protocols
**Other:** Machine Learning (Certificate course from Stanford through Coursera), Operating Systems (Stanford CS)