Neel Patel

Ph.D. candidate, Cornell University

http://www.csl.cornell.edu/~neel/ nmp83@cornell.edu

Education

Cornell University *2024–Present*
Ph.D. in Electrical and Computer Engineering
Advisor: Mohammad Alian

University of Kansas *2022–2024*
M.S. in Computer Science
GPA: 4.0/4.0

University of Kansas *2019–2022*
B.S. in Computer Science
GPA: 4.0/4.0

Research Interests

Improving performance efficiency of datacenter workloads.

Professional Experience

Research Intern, **Los Alamos National Labs** *May'23–Dec'23*, **Los Alamos, NM**
CCS-7 Applied Computer Science.

Awards and Scholarships

- ACE JUMP 2.0 SRC Researcher
- Samsung Open Innovation Contest 2nd Place Winner
- KU Locke Award Nominee (Engineering Senior of the Year)

Publications

SmartDIMM: In-Memory Acceleration of Upper Layer I/O Protocols
Patel N., *Mamandipoor A., Alian M.*
In Proceedings of HPCA 2024.

XFM: Near-Memory Acceleration of Far Memory
Patel N., *Mamandipoor A., Quinn D., Alian M.*
In Proceedings of MICRO 2023.

Accelerating Retrieval-Augmented Generation
Quinn D., Nouri M., Patel N., Salihu J., Salemi A., Lee S., Zamani H., Alian M.
In Proceedings of ASPLOS 2025.

Profiling an Architectural Simulator
Umeike J., Patel N., Manley A., Mamandipoor A., Yun H., Alian M.
In Proceedings of ISPASS 2023.

IDIO: Network-Driven, Inbound Network Data Orchestration on Server Processors
Alian M., Agarwal S., Shin J., Patel N., Yuan Y., Kim D., Wang R., Kim N.S.
In Proceedings of MICRO 2022.

Research Projects

RACER: Runtime for Accelerated Chip Multiprocessors

Designed and evaluated a data movement-aware application runtime to meet tail-latency SLOs while enabling applications to offload fixed-function kernels, like (de)compression, data movement, de/encryption, and data analytics operations to on-chip accelerators.

XFM: Accelerating Software-defined Far Memory

Designed and evaluated a system architecture which leverages a near-memory accelerator to dynamically compress/decompress Operating System pages into/out of a compressed memory pool in system memory (DRAM).

SmartDIMM: Near-Memory Accelerating Upper-Layer Network Protocols

Designed and evaluated a system architecture which offloads upper-layer network protocol processing (e.g., TLS, gzip) to a near-memory accelerator in dual-inline memory modules. Evaluated using an AES encryption accelerator on a Near-Memory Acceleration test platform (Samsung's AxDIMM) using user-facing web applications.

Accelerating Retrieval-Augmented Generation

Profiled approximate nearest neighbor and exact search schemes' performance to understand their trade-offs (search time, index generation, accuracy) in retrieval-augmented large-language model-based systems.

IDIO: Network-Driven, Inbound Network Data Orchestration on Server Processors

Profiled end-host networking performance of userspace networking software (DPDK) on server-class hardware to motivate a proposed cache optimization in server CPUs.

Profiling an Architectural Simulator

Profiled the impact of system-level and micro-architecture optimizations on the performance of an architectural simulator (gem5).

Relevant Coursework

Computer Architecture, Compiler Design, Operating Systems, Information Retrieval