Sathvik Swaminathan

sathvikswaminathan@gmail.com | linkedIn/sathvik-swaminathan-50b424178 | sathvikswaminathan's website

EDUCATION

B.E. Electronics and Communication Engineering; M.Sc. Physics   Hyderabad, India | Aug. 2019 – July 2024
Birla Institute of Technology and Science, Pilani

TECHNICAL BACKGROUND

Programming Languages: C, C++, Python, Verilog, RISC-V/x86 Assembly
Systems/AI Tools: PyTorch, QEMU, perf, Intel VTune (CPU analysis), Synopsys (VCS/DVE), Xilinx Vivado, Git
Technical Interests: AI/Systems performance, Memory Management, Computer Systems for ML, ML for Computer Systems
Relevant Courses: Computer Architecture, OS, Networks, Microprocessors, Digital VLSI Design, Device Physics

PUBLICATIONS

- TierTrain: Proactive Memory Tiering for CPU-based DNN Training

  Proceedings of the ACM SIGPLAN International Symposium on Memory Management (ISMM), 2025.

  Sathvik Swaminathan, Sandeep Kumar, Aravinda Prasad, Sreenivas Subramoney (Intel Labs)

PATENTS

- Bandwidth Aware Proactive Memory Tiering for LLM Inference

- Context-Aware Memory Tiering for Machine Learning Training

EXPERIENCE

PARL, INTEL LABS | Research Scientist   Bangalore | July 2024 – Present

- Profiled and characterized memory access patterns of DNN Training workloads on Intel Xeon CPUs.
- Designed software to manage memory across in a tiered memory (DRAM, Optane / CXL attached memory) architecture for DNN training workloads.
- Delivered performance gains of 35–84% in constrained memory scenarios over SOTA memory tiering techniques.
- Profiled LLM Inference workloads on Intel Xeon CPUs and identified performance bottlenecks.
- Designed an algorithm to manage KV cache and Weights across memory tiers to maximize system memory bandwidth.

PARL, INTEL LABS | Research Intern   Bangalore | Aug 2023 – June 2024

- Profiled memory access behavior, compute metrics of GNN Training workloads across the PyG and DGL frameworks using Linux *perf*, Intel Vtune, etc.
- Analyzed the impact of data migration of various tensors across different memory tiers (DRAM and Intel Optane) in NUMA systems.

NVIDIA | ASIC Design Intern   Bangalore | May 2023 – July 2023

- Interned in the SoC Verification team.
- Modified a verification framework to allow easier integration during SoC/cluster verification.

PARL, INTEL LABS | Research Intern   Bangalore | Feb 2023 – Apr 2023

- Characterized the memory access pattern of the BFS algorithm in Ligra framework.
- Optimized data placement between DRAM and Intel Optane, reducing DRAM cost by 20% while maintaining performance.

ABCR LABS | Digital Design Engineer Intern   Remote | Feb 2022 – July 2022

- Developed a **Python** script to automate the functional verification of the Shakti E-Class Processor.
- Designed and simulated the architecture of a SPI programmer in **System Verilog**.

SELECTED PROJECTS

RISC-V BASED NOC SYSTEM ⬀   Computer Architecture, Verilog

- Performed functional verification of a Network-on-Chip between RISC-V cores.

- Extended the NoC with multi-packet support between RISC-V cores.

COVERT CHANNELS ⬀   Computer Architecture, OS, C

- Designed a **cache-based covert channel** not requiring threshold calibration, implementation available here.

- Designed a **context-switch-based covert channel** that leverages page faults, achieving a bit error rate below 4%, preprint available here.

RISC-V PROCESSORS ⬀   Computer Architecture, Verilog

- Built single-cycle and pipelined RV32I compatible processors

- Supports hazard resolution (forwarding, stalls)

VIRTUAL MEMORY SIMULATOR ⬀   Computer Architecture, OS, C

- Developed a C-based simulator for virtual memory/page replacement

- Supports the clock-sweep algorithm and provides real time stats on TLB hits, page faults, and access time.