Sumanth Umesh

Computer Science PhD Student
University of Michigan, Ann Arbor
sumanthu@umich.edu | +1 (734) 834-6197
Google Scholar | Linkedin

Summary

My primary research focus is on optimizing memory architecture for improving the performance of memory bandwidth-bound or latency-bound workloads. My current work involves using Compute Express Link (CXL) for memory expansion and host-accelerator/accelerator-accelerator interfaces. My goal is to uncover methodologies and strategies that tackle issues such as memory shortage and bandwidth bottlenecks, ultimately contributing to more efficient and effective computing systems

Education

University of Michigan, Ann Arbor

PhD in Computer Science **[Aug 2022 - Present]**
Advisor: Prof. Reetuparna Das
GPA: 3.72/4.0

Indian Institue of Technology, Jodhpur

BTech in Electrical Engineering **[Aug 2016 - May 2020]**
Advisor: Prof. Shree Prakash Tiwari
GPA: 9.57/10.0

Work Experience

ASIC Engineer

NVIDIA, Bangalore **[Aug 2021 - Jul 2022]**

- Designed microarchitecture and RTL of forward error correction modules in the datalink layer of PCIe 6.0
- Resolved datalink layer timing and synthesis issues for PCIe - Gen 5 controller

Design Engineer

Silicon Labs, Hyderabad **[Aug 2020 - Aug 2021]**

- Designed microarchitecture and RTL of security accelerators (SHA3, Poly1305 and ChaCha20) for low-power wireless SoC
- Handled design quality, synthesis and timing checks for security sub-module
- Wrote firmware for the host ThreadArch domain-specific processor

Research Intern

Bosch Corporate Research, Bangalore **[May 2019 - Jul 2019]**

- Designed parameterized and scalable arithmetic modules (adder, subtractor and multiplier) for posit numbers as an alternative to floating point
- Optimized the design to deliver performance comparable to floating point with lower resource utilization on an FPGA

Publications

- A. Khadem, Y. Gu, X. Servot, **S. Umesh**, N. Liang, G. Oliveira, J. Gomez-Luna, O. Mutlu, R. Das. ”Cellar: CXL-enabled Larger Language Model Accelerator” *Under Submission
- **S. Umesh**, and S. Mittal. ”A survey of techniques for intermittent computing.” Journal of Systems Architecture 112 (2021): 101859.
- **S. Umesh**, and S. Mittal. ”A survey of spintronic architectures for processing-in-memory and neural networks.” Journal of Systems Architecture 97 (2019): 349-372.
- S. Mittal, and **S. Umesh**. ”A survey on hardware accelerators and optimization techniques for RNNs.” Journal of Systems Architecture 112 (2021): 101839.

Skills

Programming Languages	:	C • C++ • Python
HDL	:	Verilog • Systemverilog
Software	:	Synopsys VCS/Design Compiler • Xilinx Vivado
Simulator	:	Ramulator • ZSim
Profiling	:	VTune • PIN • Nsight Compute

Projects

CXL Based Memory Expansion for Databases

[Aug 2022 - Present]

Advisor: Prof. Reetuparna Das

- Designed a system with CXL-based memory expansion to accelerate very large main-memory databases
- Evaluated strategies to take advantage of device level parallelism between CXL devices and migrate data chunks between CXL attached memory and DRAM based on database-specific knowledge
- Implemented a CXL.mem simulator in C++ to evaluate the above strategies

CXL-enabled Large Language Model Accelerator

[July 2023 - Nov 2023]

Advisor: Prof. Reetuparna Das

- Designed processing in memory (PIM) platform for LLM inference as an alternative to GPU-based platforms
- Developed the CXL topology for inter-PIM device communication

N-Way Superscalar RISC-V Core

[Aug 2022 - Dec 2022]

Course Project | EECS 470

- Designed microarchitecture and RTL for a N-Way superscalar RISC-V core based on MIPS R10K architecture
- Implemented features such as victim cache, multi-ported I-cache, prefetching and branch prediction to achieve high IPC

Academic Achievements

- **Gold medal for the best all-round performance in the class of 2020, IIT Jodhpur**
- **Silver medal for best academic performance in Electrical Engineering, IIT Jodhpur**
- **National Talent Search Examination(NTSE) Scholarship 2014 (awarded to only 1000 high school students in the country)**

Relevant Coursework

• Computer Architecture (470) • Parallel Computer Architecture (570) • Advanced Databases (584)