

# AMEL FATIMA

+1 540-314-0569 • af3szz@virginia.edu

## EDUCATION

<b>Ph.D., Computer Engineering</b> University of Virginia, Charlottesville, VA	August' 19 - Present GPA: 4.0/4.0
<b>M.S., Computer Science &amp; Engineering (Tuition Award by the University and Fulbright Scholar)</b> State University of New York At Buffalo, Buffalo, NY	August' 17 - June' 19
<b>B.Sc., Computer Science (Gilgit Baltistan Student Scholarship for Tuition Waiver)</b> University of Engineering and Technology, Taxila, Pakistan	August' 12 - July' 16

## EXPERIENCE

<b>Research Associate, Advanced Micro Devices Research (AMD Research)</b> • Manager: Bradford Beckmann	January' 25 - August' 25
<b>Graduate Research Assistant, University of Virginia</b> • Advisor: Dr. Adwait Jog (During and after 2023)	August' 19 - present
<b>Graduate Teaching Assistant, Department of CS, University of Virginia</b> • Courses Assisted: Undergraduate Computer Architecture (CS 3330) • Responsibilities: Lectured review classes, designed and conducted labs, graded assignments, and held TA office hours	Spring' 20, Spring' 21

## PUBLICATIONS

### Characterizing and Analyzing Reverse Address Translation Overheads in Scale-Up GPU Pods

Bradford Beckmann, Tuan Ta, Amel Fatima (Under Submission)

### NetCrafter: Tailoring Network Traffic for Non-Uniform Bandwidth Multi-GPU Systems

Amel Fatima, Yang Yang, Yifan Sun, Rachata Ausavarungrun, Adwait Jog (ISCA' 2025) (22% acceptance rate)

[\[Paper\]](#) [\[Slides\]](#)

### vPIM: Efficient Virtual Address Translation for Scalable Processing in-Memory Architectures

Amel Fatima, Sihang Liu, Korakit Seemakhupt, Rachata Ausavarungrun, Samira Khan

Proceedings of the 60th Design Automation Conference (DAC' 23) (23% acceptance rate)

[\[Paper\]](#) [\[Slides\]](#) [\[Recording of the Talk\]](#)

## PATENTS

### Patent (Filed): NetCrafter: Tailoring Network Traffic in Multi-GPUs

Amel Fatima, Yang Yang, Adwait Jog

U.S. Patent No. 63/826,281, filed Jun 18<sup>th</sup>, 2025

[\[Link\]](#)

## RESEARCH PROJECTS

### Characterizing and Analyzing Reverse Translation Overheads in Scale-Up GPU Pods

Large-scale machine learning workloads often require hundreds or even thousands of GPUs to efficiently train and perform inference for large-scale language models. In such environments, interconnect performance becomes a critical bottleneck. Emerging scale-up fabrics, such as UALink, address this challenge by enabling GPUs to be organized into pods across multiple nodes, allowing direct load/store accesses between them. While this approach improves performance through inter-server load/store operations, it introduces an additional layer of translation at the destination GPU node, which we term reverse translation. In this work, we investigate the performance overheads associated with reverse translation and present a comprehensive characterization study analyzing its impact on large-scale communication collectives.

### NetCrafter: Tailoring Network Traffic for Non-Uniform Bandwidth Multi-GPU Systems.

Network bandwidth across multi-GPU clusters is a critical performance bottleneck. In this work, we characterized the network traffic and discovered that flits traversing the network are often underutilized, with many flits partially empty or entirely unnecessary. Additionally, certain flits exhibit higher latency sensitivity than others. Leveraging these insights, we developed

NETCRAFTER, a novel framework employing three advanced techniques designed to significantly reduce and optimize network traffic. Consequently, NETCRAFTER delivers substantial performance improvements across a wide range of GPGPU workloads.

#### **vPIM: Efficient Virtual Address Translation for Scalable Processing in-Memory Architectures**

3D-stacked memory technology integrates computation logic in the memory stack to reduce data movement between the CPU and memory, enabling processing-in memory (PIM). PIM systems scale in capacity and bandwidth by connecting multiple PIM stacks over a memory network. This work analyzes the virtual memory overhead in such multi-stack PIM systems and proposes vPIM, which optimizes contention in the memory network due to page table walks and reduces virtual address translation time with pre-translation.

#### **Evaluating Shared Page Table Walker Performance in on-chip TPU Style Systolic-Array Based Accelerator using FireSim.**

This project utilized FireSim, a cycle-accurate RTL simulator with FPGA acceleration on AWS F1 instances. The evaluation focused on assessing the performance of shared page table walkers within an on-chip accelerator, Gemmini. We ran diverse deep neural network workloads in a multi-tenancy environment on Gemmini, providing insights into the efficiency of shared resources specifically page table walkers in concurrent task execution.

### **COURSE PROJECTS**

---

#### **Characterizing the Performance of different RISC-V Cores Using LiteX (ECE-6501)**

Technology: Xilinx UltraScale+VCU118 FPGA Board, LiteX Framework

1. Configured SOC configurations with different RISC-V cores (rocket, blackparrot, naxriscv) in Litex.
2. Generated bitstreams, implemented them on FPGA and booted linux on all RISC-V cores.
3. Compared the fmax and resource utilization for different configurations to perform a preliminary analysis and help researchers identify relevant cores for their research.

#### **Design of 8X8 Vedic Multiplier using Kogge Stone Adder (CSE-593)**

Technology: Cadence (TSMC 180nm Technology)

1. Implemented a very fast response-time 8-bit multiplier with low propagation delay, using Radix-2 Algorithm
2. Designed and Verified the Circuit layout with DRC LVS Check.

#### **Design of a 16-Bit RISC Processor Adder (CS-590)**

Technology: VHDL, Vivado

1. Implemented a RISC Processor with 512Bytes Instruction and 512Bytes Data Memory With 16 16-bit Registers.
2. Designed my own 16-bit ISA of 24 instructions. Generated RTL and Timing Simulation using 100kHz clk.

### **AWARDS AND SCHOLARSHIPS**

---

- Travel Grant by Computing Research Association (CRA) to attend Grad Cohort, New Orleans, LA, 2022
- Fulbright Scholarship for master's degree in computer science & engineering at the State University of New York at Buffalo.
- Tuition Award by the State University of New York at Buffalo.
- Gilgit Baltistan Student Scholarship for Full tuition waiver at the University of Engineering & Technology Taxila, Pakistan.

### **PROFESSIONAL SERVICES**

---

- Artifact evaluation reviewer, MICRO 2025
- Artifact evaluation reviewer, IISWC 2025
- Artifact evaluation reviewer, SIGCOMM 2025

### **TECHNICAL SKILLS**

---

Programming: C, C++, Python, Go

HDL: VHDL, Verilog

EDA and Simulation Tools: Cadence Virtuoso-XL, Xilinx Vivado, ModelSim, Firesim, LiteX, Gem5, Zsim, Ramulator, MGPUSim, AstraSim, Omnetpp

Database: MySQL

Version Control: Git

