

I'm a 2nd year Ph.D. student at the University of Michigan (advised by Prof. Satish Narayanasamy), working on confidential computing and trusted hardware, with a strong background in GPU architecture, memory systems and Out-of-Order CPU architecture. I'm looking for a research or engineering internship position during the summer of 2024.

My research seeks to advance confidential computing solutions for enabling privacy-preserving data analytics solutions ranging from population-scale genomic analysis to generative AI. My approach is to develop trustworthy hardware, and use it to efficiently guarantee privacy from the rest of system components, including the operating system and system administrators.

I have made three specific contributions. One, I have helped build SECRET-GWAS, a privacy-preserving genome-wide association study platform on Microsoft Azure's confidential computing platform. SECRET-GWAS scales to over 1000 cores. We showed for the first time that we can perform regression analysis on large genomic datasets from multiple institutions in less than a few seconds, without revealing data to even the cloud service provider (under submission to Nature Methods).

Two, I invented the Toleo. Today, trusted processors (Intel SGX) support only a few hundred MBs of secure memory space. The Toleo is an innovative solution that expands trust to intelligent memory and scales secure memory space to tens of TeraBytes, which is a million times larger than what is feasible today (submitted to ASPLOS 2024).


In addition to these thesis work, in collaboration with AMD, I have also built a GPU accelerated long-read genome sequencing software artifact (minimap2). It will soon be released as part of AMD Research Open-source Project.

My internship with NVIDIA gpu architecture team on deep learning performance analysis and ubenchmarking on the Hopper generation equips me well with GPU architecture knowledge, especially with the GPU memory system. I worked on utilizing Hopper features such as TMA and cluster xbar to improve the NCCL library.

Going forward, I plan to advance confidential computing to address privacy and safety concerns of generative AI. I'm also looking into providing safe storage against ransomware attacks through trusted hardware.

Please see next page for my CV and detailed background.

JUECHU DONG

✉ joydong@umich.edu  [joydddd.github.io](https://github.com/joydddd)

SUMMARY

Juechu (Joy) Dong is a 2nd year Ph.D. student with the Computer Engineering Lab at the University of Michigan, advised by Prof. Satish Narayanasamy. Her research seeks to advance confidential computing solutions for enabling privacy-preserving data analytics solutions ranging from population scale genomic analysis to generative AI. Her current work focuses on scaling trusted memory capacity from hundreds of MB to tens of TB and developing privacy-preserving genome-wide association study platform on Azure's confidential computing platform.

EDUCATION

University of Michigan - Ann Arbor (exp.) 2027
Computer Science and Engineering, PhD
Topics: Computer System & Architecture, Trusted Hardware / Confidential Computing
Advisor: Prof. Satish Narayanasamy

University of Michigan-Shanghai Jiao Tong University Joint Institute Aug 2022
Computer Engineering, Bachelor of Science

University of Michigan - Ann Arbor Apr 2022
Computer Engineering, Bachelor of Science in Engineering, Summa Cum Laude
Selected Coursework: Comp. Architecture A, Compiler A+, Operating System A
GPA: 3.99/4.00

PUBLICATION (under submission)

Toleo: Scaling Freshness to Tera-scale Memory Using CXL and PIM ASPLOS'24
J. Dong, J. Rosenblum, S. Narayanasamy under submission

- Scale trusted memory size from hundreds of MB to tens of TB by expanding the span of trusted from a single trusted processor to an entire platform including intelligent memories.
- Design a new scheme of freshness protection that reduces the space requirement by 50x.
- Reduce deployment cost by spacing sharing one intelligent memory device among multiple CPUs.

mm2-long: Accelerating Accurate Ultra Long Genome Sequence Mapping on AMD GPU

J. Dong, X. Liu, H. Sadasivan, G. Sitaraman, S. Narayanasamy on-going

- Accelerate computational intensive chaining step in the state-of-art long sequence mapping tool minimap2 using AMD GPU by 5x.
- Optimize towards ultra long reads of 100k+ to accommodate genome sequencing technology trend.
- Develop adaptive GPU scheduling algorithm to balance highly heterogeneous workload.

SECRET-GWAS: Confidential Computing for Population-Scale GWAS RECOMB'24
J. Rosenblum, J. Dong, S. Narayanasamy under submission

- Develop a thousand-core platform on Azure Confidential Computing to conduct multi-institutional GWAS on millions of patients in less than a minute.
- Adapt Spark-based Hail genomic analysis framework to run on TEE under obliviousness requirement.
- Parallelize GWAS computation on 1k cores to achieve near linear speedup.

INDUSTRY EXPERIENCE

NVIDIA May 2022 - Aug. 2022

GPU Deep Learning Architect Intern

- Model and analyze new memory features on next-gen GPUs such as distributed shared memory, asynchronous transaction barrier, etc.
- Analyze and optimize multi-GPU data movement for deep learning workloads using Tensor Memory Accelerator (TMA).

TEACHING

Instructional Aide

2021FA, 2022WN

EECS470 Computer Architecture

Graduate Student Instructor

2023FA

EECS471 Applied Parallel Programming with GPUs

Graduate Student Instructor

2024WN

EECS570 Parallel Computer Architecture

SERVICE

Computer Engineering Lab Reading Group

2022 - present

Coordinator

- Organize weekly paper reading presentations and discussions.
- Host talks from visiting researchers and professors.

UM-SJTU Joint Institute Alumni Association

2022 - present

Founder & Vice President

- **Alumni Engagement:** Organize alumni and student gatherings.
- **Relationship Building:** Involve in expanding SJTU - UM collaborations, connecting to JI sponsors, and building industry relationships.
- **Career Advising:** Organize students career development workshops.
- **Welcoming:** Host new student orientation events, organize airport pickups, and offer settle down help.
- **Student Support:** Support students during the stressful transition to start in a new university in a new country, and during urgent crisis.

SKILLS

Programming Languages: C/C++, CUDA, HIP, SIMD, (system) verilog, bash, Makefile

Technologies/Frameworks:

GPU Tuning: nsight-compute/nsight-sys, omniperf/omnitrace/rocprof

Formal Verification: Murphi,

SIMD: avx512, avx2 on Xeon Phi

Simulation: SniperSim, DRAMSim, pinplay

Confidential Computing: Open Enclave SDK, Intel SGX

Architectures: AMD CDNA2 Instinct GPU, NVIDIA Hopper GPU, Intel Xeon Phi, Out-of-order CPU

SELECTED HONORS

Rackham International Student Fellowship

2023-2024

Awarded to 25 outstanding students among all international PhD and MS students in the university

James B. Angell Scholar

2020-2023

Achieving "A Record" for 6 consecutive terms

John Wu & Jane Sun Outstanding Scholarship

2018-2022

Awarded to 13 outstanding students among ~ 300 freshmen in SJTU-UM Joint Institute for half tuition waiver

Outstanding Academic Performance Scholarship

2018-2020

Awarded to undergraduate students at SJTU with outstanding academic performance