# Jingyao Zhang

jzhan502@ucr.edu | Google Scholar | Homepage | GitHub | LinkedIn | Riverside, USA

**Interest:** Enhancing the **performance** of **privacy-preserving systems** across both **software** and **hardware** domains.

## Education

**University of California, Riverside** — Riverside, USA
*Ph.D. Candidate in Computer Science, GPA: 3.7/4.0* — *Sep 2021 – Present*
*Advisor: Elaheh Sadredini*

**Xidian University** — Xi'an, China
*M.E. in Electronic and Telecommunications Engineering, GPA: 3.7/4.0; Outstanding Thesis Award* — *Sep 2018 – Jun 2021*
*B.E. in Telecommunications Engineering, GPA: 3.7/4.0; Pilot Class (Top 5% of 800+)* — *Sep 2014 – Jun 2018*

## Work Experience

**Data Center Server Platform Security, Nvidia Corporation** — *Mentor: Lohith Rangappa*
*System Software Intern* — *Jun 2025 – Sep 2025 (exp.)*

- Developed a Software Development Kit (SDK) to enable the integration of security firmware with third-party hardware.
- Designed a two-tier test suite to validate HAL (Hardware Abstraction Layer) implementations and ensure system integrity.

**Data Center Server Platform Security, Nvidia Corporation** — *Mentor: Lohith Rangappa*
*System Software Intern* — *Jun 2024 – Sep 2024*

- Developed BMC-assisted attestation to support SPDM-based remote attestation for non-PLDM devices.
- Enabled secure firmware update for non-PLDM devices by integrating Key Manifest into External Root of Trust.

**Operating System Lab, Alibaba Cloud** — *Mentor: Yue Qian*
*External Developer* — *Aug 2023 – Nov 2024*

- Developed a secure LLM inference platform with Trusted Network Gate supporting HTTP/WebSocket.
- Designed and developed a *Combined Attestation* framework for *CoCo KBS* to support GPU attestation. [repo] [demo]
- Conducted research on the *nvTrust* for Nvidia confidential computing, including verification and attestation.
- Investigated existing systems that support GPU confidential computing, such as *Azure Confidential AI*.

**Open Source Promotion Plan, Chinese Academy of Sciences** — *Mentor: Ding Ma & Jiale Zhang*
*Project Developer* — *2023 & 2024*

- Built Confidential-AI, a framework based on Attestation Agent, CDH, and Trusted Network Gate for secure AI workflows.
- Developed Tee-RefValue, a workflow that automatically generates reference measurements for user image, firmware, and kernel on the AMD SEV-SNP platform, compatible with *Confidential Containers*.
- Examined the attestation process on the AMD SEV-SNP platform, including the generation of reference measurement.
- Evaluated attestation tools across various cloud service providers, such as Google Cloud Platform and Azure.

## Publications

1. **Jingyao Zhang**, and Elaheh Sadredini. "No One-Size-Fits-All: A Workload-Driven Characterization of Bit-Parallel vs. Bit-Serial Data Layouts for Processing-using-Memory." *In Submission.*
2. **Jingyao Zhang**, Jaewoo Park, Jongeun Lee and Elaheh Sadredini. "SAIL: SRAM-Accelerated LLM Inference System with Lookup-Table-based GEMV." *In Submission.*
3. **Jingyao Zhang**, and Elaheh Sadredini. "A Near-Cache Architectural Framework for Cryptographic Computing." *In Submission.*
4. **Jingyao Zhang**, and Elaheh Sadredini. "CryptoSRAM: Enabling High-Throughput Cryptography on MCUs via In-SRAM Computing." *In Proc. of the 1st IEEE Cross-disciplinary Conference on Memory-Centric Computing (CCMCC). October 2025.*
5. Sahar Ghoflsaz, **Jingyao Zhang**, and Elaheh Sadredini. "Enabling Low-Cost Secure Computing on Untrusted In-Memory Architectures." *In Proc. of the 34th USENIX Security Symposium. August 2025.*
6. **Jingyao Zhang**, and Elaheh Sadredini. "Unlocking Energy-Efficient and High-Throughput Secure Data Communication in IoT with Memory-Centric Computing." *Poster at Proc. of the 39th International Parallel and Distributed Processing Symposium (IPDPS). June 2025.*
7. **Jingyao Zhang**, Mohsen Imani, and Elaheh Sadredini. "BP-NTT: Fast and Compact in-SRAM Number Theoretic Transform with Bit-Parallel Modular Multiplication." *In Proc. of the 60th Design Automation Conference (DAC). July 2023.*

8. **Jingyao Zhang**, and Elaheh Sadredini. "Inhale: Enabling High-Performance and Energy-Efficient In-SRAM Cryptographic Hash for IoT." *In Proc. of the 41th International Conference on Computer-Aided Design (ICCAD). November 2022.*
9. **Jingyao Zhang**, Hoda Naghibijouybari, and Elaheh Sadredini. "Sealer: In-SRAM AES for High-Performance and Low-Overhead Memory Encryption." *In Proc. of the 22th International Symposium on Low Power Electronics and Design (ISLPED). August 2022.*
10. **Jingyao Zhang**, Huaxi Gu, Li Zhang, Bing Li, and Ulf Schlichtmann. "Hardware-Software Codesign of Weight Reshaping and Systolic Array Multiplexing for Efficient CNNs." *In Proc. of the 24th Design, Automation and Test in Europe (DATE). February 2021.*

## Research Experience

**AREA Lab, University of California, Riverside** — *Advisor: Elaheh Sadredini*

*Graduate Research Assistant* — *Sep 2021 – Present*

- Currently designing a TEE framework tailored for Processing-In-Memory (PIM) architectures, inspired by secure NPU designs such as sNPU; integrated lightweight tile-based permission checking, SRAM-level bank isolation without virtual memory, and NoC-aware secure data routing, enabling secure execution with minimal TCB and support for cache-line aware memory mapping.
- Developed a general-purpose compiler for domain-specific accelerators using MLIR-based Polyhedral model and E-Graph-based searching, specifically targeting workloads that involve vector and scalar kernels as well as mixed-precision workloads.
- Designed an on-chip solution for accelerating quantized language models, including dynamic data precision adaptation and efficient runtime de-/quantization.
- Developed a framework to seamlessly integrate in-SRAM computing into existing computer systems for efficient and secure on-chip processing of pre- and post-quantum cryptography.
- Developed a bit-parallel modular multiplication algorithm with implicit shifting technology for efficient and secure in-SRAM computing of the NTT, optimizing performance on a low-overhead SRAM array.
- Designed a secure in-SRAM architecture for on-chip acceleration of the AES/SHA-3 algorithm using row/lane-wise data alignment, achieving high energy and area efficiency with high throughput.

**Advanced Networking Technology Lab, Xidian University** — *Advisor: Huaxi Gu*

*Graduate Research Student* — *Sep 2018 – Jun 2021*

- Developed a hardware-software co-design framework for efficient CNNs, leveraging weight reshaping and systolic array multiplexing with genetic algorithms for optimal hardware performance.
- Built a distributed inference system for accelerating CNNs using systolic array on FPGAs, with HLS for low-level hardware description and Aurora/Ethernet protocols for inter-board communication.
- Designed a flexible and compact N $\times$ N plasmonic switch topology with a dedicated configuration algorithm that ensures re-arrangeable non-blocking, making it ideal for managing mixed traffic in data centers.
- Designed a low-loss compact plasmonic router for mesh networks in optical Network-on-Chip, exhibiting lower insertion loss and a smaller footprint compared to other structures.

## Teaching Experience

**CS 010C Introduction to Data Structures and Algorithms** — *Instructor: Patrick Miller*

*Teaching Assistant* — *Spring 2024*

- Hosted three lab sessions per week to supplement lecture, assignments.
- Held weekly office hours to answer students' questions.

**CS 161 Design and Architecture of Computer Systems** — *Instructor: Elaheh Sadredini*

*Teaching Assistant* — *Fall 2023, Winter 2024*

- Hosted three discussion sessions per week to supplement lecture, homework and lab.
- Held weekly office hours to answer students' questions.

**CS 213 Multiprocessor Architecture and Programming** — *Instructor: Elaheh Sadredini*

*Teaching Assistant* — *Fall 2022*

- Led two discussion sessions of students' presentations.
- Held weekly office hours to answer students' questions.
- Graded homework and programming assignments.

## Other Experience

**gem5 Boot Camp** Davis, USA

*Participant* *Jul 2022*

- Simulated and analyzed the performance of computer architectures, and studied the behavior of different workloads and benchmark suites on various computer architectures.
- Evaluated the impact of different design choices on system performance, such as varying cache sizes or using different interconnect topologies, and explored the effects of different microarchitectural features.

**Xilinx Summer Camp** Online

*Participant & Team Leader* *Jul 2020 – Aug 2020*

- Developed an FPGA-based distributed platform for acceleration over Ethernet, with the mother board sending a file to a watched folder on the child board for immediate program execution. [repo]

**Microsoft Innovation Center** Xi'an, China

*Intern* *Jul 2017 – Aug 2017*

- Explored the advancements and challenges in the evolution of cellular networks across generations, starting from the early analog systems to the 5G technology.

## Talks

1. **Jingyao Zhang**. "In-SRAM Computing for Cryptography." *In Xia Peisu Forum hosted by ICT.* [slides] *Dec 2023*
2. **Jingyao Zhang**. "Combined Attestation." *In Confidential Containers Community Meeting.* [slides] [video] [demo] *Nov 2023*
3. **Jingyao Zhang**. "Safeguard Your Cloud Workloads: An In-depth Look at Confidential Computing." *In Proc. of IEEE Inland Empire Data Science Workshop (IEDSW).* [link] *Nov 2023*
4. **Jingyao Zhang**. "Safeguard Your Cloud Workloads and Then Accelerate: An In-depth Look at CPU and GPU Confidential Computing." *Remotely at ByteDance in Mandarin.* ***Live Attendance:*** *∼**300**.* [slides] [video] *Oct 2023*
5. **Jingyao Zhang**. "BP-NTT: Fast and Compact in-SRAM Number Theoretic Transform with Bit-Parallel Modular Multiplication." *In Proc. of the 60th Design Automation Conference (DAC). San Francisco, CA.* [slides] [video] *Jul 2023*
6. **Jingyao Zhang**. "Inhale: Enabling High-Performance and Energy-Efficient In-SRAM Cryptographic Hash for IoT." *In Proc. of the 41th International Conference on Computer-Aided Design (ICCAD). San Diego, CA.* [slides] [video] *Nov 2022*
7. **Jingyao Zhang**. "Sealer: In-SRAM AES for High-Performance and Low-Overhead Memory Encryption." *In Proc. of the 22th International Symposium on Low Power Electronics and Design (ISLPED). Online.* [slides] [video] *Aug 2022*
8. **Jingyao Zhang**. "Hardware-Software Codesign of Weight Reshaping and Systolic Array Multiplexing for Efficient CNNs." *In Proc. of the 24th Design, Automation and Test in Europe (DATE). Online.* [slides] [video] *Feb 2021*

## Awards

**Laxmi N. Bhuyan Endowed Fellowship**, University of California, Riverside *2025*

**DAC Young Fellowship**, Design Automation Conference *2023*

**Dean's Distinguished Fellowship Award**, University of California, Riverside *2021*

**Outstanding Thesis Award**, Xidian University *2021*

**First-class Scholarship**, Xidian University (Top 14% of 560+) *2018, 2019*

**Outstanding Student Award**, Xidian University *2018, 2019*

## Grants

**Conference Travel Grant**, University of California, Riverside *2023*

**Student Travel Grant**, gem5 Boot Camp *2022*

## Services

### Reviewed Papers: 16

**Journal Paper Review:**

Elsevier Integration (1) *2025*

IEEE Computer Architecture Letters (CAL) - (13) *2023, 2024,2025*

IEEE Transactions on Dependable and Secure Computing (TDSC) - (1) *2024*

**Conference Paper Review:**

IEEE International Conference on Communication (ICC) - (1) *2024*

**Evaluated Artifacts:** 17

**Artifact Evaluation Board:**

Journal of Systems Research (JSys) - (2) *2023*

**Artifact Evaluation Committee:**

IEEE International Symposium on High-Performance Computer Architecture (HPCA) - (1) *2024*
USENIX Security Symposium - (2) *2024*
ACM European Conference on Computer Systems (EuroSys) - (3) *2024*
Annual Network and Distributed System Security Symposium (NDSS) - (2) *2024*
ACM Symposium on Operating Systems Principles (SOSP) - (2) *2023*
ACM Conference on Computer and Communications Security (CCS) - (1) *2023*
ACM International Conference On Mobile Computing And Networking (MobiCom) - (1) *2023*
USENIX Annual Technical Conference (ATC) - (1) *2023*
USENIX Symposium on Operating Systems Design and Implementation (OSDI) - (2) *2023, 2024*

## Skills

**Programming:** C, C++, Python, Verilog, Rust, MLIR

**Technologies:** Gem5, Sniper, HSpice, PyTorch, LLVM

**Languages:** Chinese (Native), English (Professional)