

# Yuchen Tang

📍 Stony Brook, NY, USA    ✉️ [yuchen.tang.1@stonybrook.edu](mailto:yuchen.tang.1@stonybrook.edu)    📞 (631) 948-5894    🌐 [twicy.github.io](https://twicy.github.io)  
LinkedIn [yuchen-tang-b49a37190](https://www.linkedin.com/in/yuchen-tang-b49a37190)    Github [twicy](https://github.com/twicy)

## Skills

**Programming:** Proficient with C, Python, and Git; Good understanding of **ML Algorithms** and **Linux Kernel Internals**

**Parallel Programming:** ISPC, Pthreads, OpenMP, **CUDA**

**Areas of Interest:** CXL, RDMA, MLSys, HPC

**Languages:** English (fluent), Mandarin (native)

## Education

|           |   |                       |
|-----------|---|-----------------------|
| <b>MS</b> | <b>Stony Brook University</b> , Computer Engineering  | Stony Brook, NY, USA  |
|           | <ul style="list-style-type: none"> <li>Selected Courses: Advanced VLSI Systems Design, Deep Learning Algorithms and Software, Operating Systems, Parallel Processing Architectures</li> <li>GPA: 4.0/4.0</li> </ul> | Sept 2024 – Dec 2025  |
| <b>BS</b> | <b>University of Michigan</b> , Computer Science  | Ann Arbor, MI, USA    |
|           | Minor in Mathematics  | Sept 2019 – June 2022 |
| <b>BS</b> | <b>Shanghai Jiao Tong University</b> , Electrical and Computer Engineering  | Shanghai, China       |
|           |   | Sept 2017 – Sept 2019 |

## Publications

|   |           |
|---|-----------|
| <b>Practical and Scalable RDMA Connection Sharing for HPC Workload</b>  | Sept 2025 |
| Yuejie Wang, Tuo Fang, Biyu Peng, Yang Cheng, Xin Sun, Chengchao Xu, <i>Yuchen Tang*</i> , Yuxin Ren, Ning Jia, Xinwei Hu, Yunfei Du, Guyue Liu<br>(EuroSys'26) |           |
| <b>Something about CXL Live VM Migration</b>  | Dec 2025  |
| Tyler Estro, Daniel Berger, Michael Ferdman, Geoff Kuenning, <i>Yuchen Tang*</i> , Mohit Verm, Erez Zadok<br>(Coming Soon)                                      |           |

## Experience

|  |                           |
|--|---------------------------|
| <b>Stony Brook University</b> , Research Assistant and Teaching Assistant  | Stony Brook, NY, USA      |
| <ul style="list-style-type: none"> <li><b>Compas Lab</b>: Developed from scratch a device driver for an Ethernet NIC on FPGA with <b>Prof. Mike Ferdman</b>, achieving <b>50% performance boost</b>, supporting <b>NAPI-based polling</b></li> <li><b>FSL Lab</b>: Developed a working prototype with <b>Prof. Erez Zadok</b> to enable QEMU VM live migration with CXL type-3 memory devices, achieving a <b>2x faster</b> migration speed over traditional TCP-based methods in internal benchmarks</li> <li>Assisted <b>Prof. Jorge Mendez-Mendez</b> in teaching ESE 577 <b>Deep Learning Algorithms and Software</b>. Set up Gradescope and developed auto-grader public test cases, enhancing grading efficiency. Conducted Q&amp;A and demonstration sessions to clarify complex course material for students.</li> </ul> | Sept 2024 – Dec 2025      |
| <b>Huawei Technologies Co., Ltd.</b> , Operating System Engineer   | Hangzhou, Zhejiang, China |
| Owner of <b>etmem</b> , a tiered memory system manager   | Sept 2022 – Sept 2024     |
| <ul style="list-style-type: none"> <li>Enlarged memory nodes for Huawei's Taishan servers with FPGA-based CXL devices</li> <li>Optimized OpenUCX, an RDMA communication middleware for HPC workloads</li> </ul>  |                           |

- Extracted NPU operators in deep learning models for swift modeling and simulations

**Wonders Information Co., Ltd.**, Data Warehouse Maintenance Intern

Shanghai, China

- Used Python to process data and update Apache Hive databases accordingly

May 2021 – Aug 2021

## Projects

---

### Parallel Programming: SGEMM + SSSP

Sept 2025 – Dec 2025

Advisor: Prof. Mikhail Dorojevets [🔗](#)

- Requirements: Perform **SGEMM** and report two minimum values and their indices first, then find and display the **Single Source Shortest Weighted Path** between them
- Implemented a series of high-performance matrix multiplication programs with *ISPC*, *Pthreads/OpenMP*, *CUDA*, *MPI* ensuring **robustness across all edge cases** (e.g., small and irregular matrices).
- Designed and optimized CUDA kernels using *shared memory tiling*, *thread tiling*, *float4 vectorization* and *double buffering*, achieving **80%–85% of cuBLAS SGEMM performance**, heavily endorsed by the Instructor.
- Used warp reduction and atomic operations to guarantee correct and swift update of both the minimum value and index. Used a two stage approach to tackle SSSP, one for neighbor exploration and relaxation, second for convergence.

### CXL-based QEMU VM live migration

Apr 2025 – June 2025

Advisor: Prof. Erez Zadok [🔗](#), Dr. Tyler Estro [🔗](#), Mr. Daniel Berger [🔗](#), Prof. Mike Ferdman [🔗](#), Prof. Geoff Kuenning [🔗](#)

- Implemented a new QEMU VM live migration strategy, utilizing a CXL2.0 memory pool as the intermediate bouncing buffer, to support fast Pre-Copy strategy
- Implemented a working proof-of-concept prototype on emulated CXL hardware, achieving a **2× faster** migration speed over traditional TCP-based methods in internal benchmarks
- Helped implement QEMU-level CXL tiering solution, where only hot pages are transferred and cold pages are remapped rather than migrated, making it **zero-copy** by definition

### Ethernet Network Driver Development for AMD AXI FPGA

Oct 2024 – Mar 2025

Advisor: Prof. Mike Ferdman [🔗](#)

- Implemented both TX and RX logic from scratch, 900+ lines of clean, well-documented kernel code, supporting multiple features such as **multi-fragment packet transmission** and **NAPI-based polling**
- Manipulated two ring buffers of hardware descriptors and registers to ensure safe DMA operations, properly handling erroneous situations and fringe cases
- Instructor's comment: "I've been giving this project to CS students for YEARS, you're the first ones to pull it off."

### RDMA communication middleware (openUCX) optimization

Oct 2023 – May 2024

Advisor: Yunfei Du [🔗](#), Tuo Fang, Yuxin Ren [🔗](#), Guyue Liu [🔗](#)

- Helped implement **XRC service type**, efficiently bringing down QP numbers thus increasing RDMA scalability
- Helped resolve the **Head-of-Line blocking** problem with a simple yet effective two-level message slicing policy
- Helped set up a finer **credit-based flow control** mechanism
- Evaluated our implementation based on both micro-benchmarks and real-world applications

### Memory node expansion using FPGA-based CXL 1.1 devices

Apr 2023 – Dec 2023

- The company's tiered memory system, etmem, identifies cold and hot pages and migrates them between tiers of memory to best **exploit fast memory for performance and slow memory for capacity**
- Observed the frequent **page ping-pong** caused by original, coarse-grained access bit checking method
- Developed a substitute kernel module that multiplexed the original etmem workflow, selectively listening to either access bit checking or fine-grained hardware access counter results

## Achievements

---

### Nina and Vinod Rathore Scholarship

October 15, 2025

This award celebrates your hard work, dedication, and exceptional academic achievements.

### Huawei Future Star

May 2023

Awarded for heavily contributing to the maintenance and optimization of etmem