DAN WU

wudan0399.github.io dan.w@u.nus.edu GitHub Google Scholar
COM3-02-17, System and Networking Lab, 13 Computing Drive, Singapore 117417

EDUCATION

• **National University of Singapore, Singapore** *Sept. 2020 - Jul. 2025 (Expected)*
Ph.D. in Computer Science. Advisor: Tulika Mitra

• **Fudan University, China** *Jan. 2015 - Jun. 2020*
B.Sc. in Computer Science and Data Science.

INTERESTS

I have been working on software-hardware codesign, focusing on full-stack solutions for coarse-grained reconfigurable array (CGRA) and spatial accelerators, targeting neural networks and graph-related problem acceleration. I am also interested and experienced in distributed LLM training and inference.

WORK EXPERIENCE

[1] **AMD Singapore** *Jun. 2024 - Dec. 2024*
Research Intern. Manager: Haris Javaid
- Implement, profile and optimize for a new design of communication collectives in ROCm Communication Collective Library (RCCL) to reduce the communication cost in distributed LLM training. I am responsible for 1) building analysis tools to profile the communication in LLM workloads and estimate effect of different optimization techniques across various systems; 2) debugging, profiling and optimizing our new design, implemented with RCCL.

[2] **The Hong Kong Polytechnic University** *Jul. 2019 - Dec. 2019*
Research Assistant. Advisor: Jiannong Cao
- Designed an estimation model to adaptively partition neural network models by layer and deploy different parts on Raspberry Pi and laptop for object detection acceleration.

PUBLICATIONS

[1] **Dynamic Irregular Computation Offloading on Heterogeneous FPGA-GPU Systems** **In Submission**
Dan Wu, Zhenyu Bai, Pranav Dangi, Miriyala Pavan and Tulika Mitra.
- Designed a scheduler for heterogeneous systems with flexible schedules that adapt to different workload characteristics for performance and energy consumption.

[2] **Rewire: Advancing CGRA Mapping through Consolidated Routing Paradigm** **DAC'25**
Zhaoying Li, Dan Wu, Cheng Tan and Tulika Mitra.
- Changed the classic memory-free, single-edge PathFinding-based CGRA mapping with the proposed propagation-based method that allows computation reuse and multi-edge mapping in one go.

[3] **InkStream: Real-time GNN Inference on Streaming Graphs via Incremental Update** **IPDPS'25**
Dan Wu, Zhaoying Li, and Tulika Mitra.
- Designed an event-based method reducing irregular memory access and repeated computation of Graph Neural Network inference on dynamic graphs by incrementally updating node embedding.

[4] **Flip: Data-Centric Edge CGRA Accelerator** **TODAES'23**
Dan Wu, Peng Chen, Thilini Kaushalya Bandara, Zhaoying Li, and Tulika Mitra.
- Proposed a full-stack solution (compiler, simulator, and RTL implementation) for accelerating the irregular graph analysis algorithms on Coarse-Grained Reconfigurable Array originally designed for regular loop kernels.

[5] **FLEX: Introducing FLEXible Execution on CGRA with Spatio-Temporal Vector Dataflow** **ICCAD'23**
Thilini Kaushalya Bandara, Dan Wu, Rohan Juneja, Dhananjaya Wijerathne, Tulika Mitra, and Li-Shiuan Peh.

- Designed a CGRA with a flexible spatio-temporal vector dataflow execution model, reaching a balance between energy-efficient low-throughput spatial CGRAs and energy-consuming high-throughput spatial-temporal CGRAs by adjusting the reconfiguration frequency.

[6] **LISA: Graph Neural Network based Portable Mapping on Spatial Accelerators** **HPCA'22**
Zhaoying Li, Dan Wu, Dhananjaya Wijerathne, and Tulika Mitra. *Distinguished Artifact Award*
- Proposed a portable compilation framework that can be tuned automatically to generate quality mapping for varied spatial accelerators.

[7] **Mining Verb-oriented Commonsense Knowledge** **ICDE'20**
Jingping Liu, Yuanfu Zhou, Dan Wu, Chao Wang, Haiyun Jiang, Sheng Zhang, Bo Xu, and Yanghua Xiao.
- Proposed a knowledge-driven approach to mine verb-oriented commonsense knowledge from verb phrases with the help of taxonomy.

TEACHING EXPERIENCE

• **[CS3237] Introduction to IoT** *Aug. 2022 - Dec. 2022*
Teaching Assistant. Course Coordinator: Boyd Anderson
• **[CS3237] Introduction to IoT** *Aug. 2021 - Dec. 2021*
Teaching Assistant. Course Coordinator: Colin Tan and Boyd Anderson

PRACTICAL EXPERIENCE

• **Deploying AI Applications on Customized Tape-out Accelerator** *Dec. 2023 - Feb. 2024*
- Offload the computation-intensive part of an object detection neural network to our own tape-out coarse-grained reconfigurable array accelerator.

• **Dumbbell Counting** *Dec. 2020 - Jun. 2021*
- Implement an algorithm counting the dumbbell lift on the Arduino BLE 33 Sense microcontroller.

• **Integration of TVM and NVDLA** *Sep. 2020 - Nov. 2020*
- Improved the performance and compatibility of the latest industrial accelerator NVDLA by allowing the NVDLA compiler to use a highly optimized and frontend-friendly TVM model as input. Has 24 stars on GitHub.

AWARD

• **SoC Research Incentive Award** *2023*

• **Research Young Fellow Program**, Design Automation Conference *2022*

• **Research Scholarship**, National University of Singapore *2020*

• **Second Prize**, The 9th University Student Service Outsourcing Innovation Competition (top 3) *2018*

• **Meritorious Winner**, Interdisciplinary Contest in Modeling (top 9%) *2017*

• **Outstanding Undergraduate Students**, Fudan University *2017*

TEACHING ASSISTANT

• **[CS3237] Introduction to Internet of Things, School of computing, NUS** *2022 Aug-Nov*

• **[CS3237] Introduction to Internet of Things, School of computing, NUS** *2021 Aug-Nov*

SKILLS

Proficient: C++, Python, PyTorch, Docker, Shell
Intermediate: CUDA, MPI, TVM, TFLite, Embedded system development
Beginner: LLVM, Chisel, System Verilog