

赵怿龙

电子邮件: sjtuzyl@sjtu.edu.cn 手机号码: 15221833996 个人主页: <https://xiaoke0515.github.io>

研究方向

基于 DRAM 的存内（近存）计算，包括存内计算架构的内存管理、调度及相关应用加速。

教育经历

2022.9 – 2026.6 (预计毕业)	上海交通大学 (SJTU), 上海, 中国 计算机科学与技术, 博士	导师: 蒋力 研究员
2018.9 – 2021.3	上海交通大学 (SJTU), 上海, 中国 计算机技术, 硕士	GPA: 3.49/4.0
2014.9 – 2018.6	上海交通大学 (SJTU), 上海, 中国 电子科学与技术, 学士	GPA: 3.51/4.3

论文&专利 (完整列表[见此](#))

- **Yilong Zhao**, Fangxin Liu, Zongwu Wang, Mingjian Li, Mingxing Zhang, Chixiao Chen, and Li Jiang, "BLADE: Boosting LLM Decoding's Communication Efficiency in DRAM-based PIM", in *Proceedings of the 31st Asia and South Pacific Design Automation Conference (ASP-DAC'26)* (Accepted)
- **Yilong Zhao**, Fangxin Liu, Xiaoyao Liang, Mingyu Gao, Naifeng Jing, Chengyang Gu, Qidong Tang, Tao Yang, and Li Jiang, "STAMP: Accelerating Second-order DNN Training Via ReRAM-based Processing-in-Memory Architecture," in *Proceedings of the 16th International Symposium on Advanced Parallel Processing Technology (APPT'25)*
- **Yilong Zhao**, Mingyu Gao, Huachen Zhang, Fangxin Liu, Gongye Chen, He Xian, Haibing Guan, and Li Jiang, "PUSHtap: PIM-based In-Memory HTAP with Unified Data Storage Format," in *Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'25)*
- **Yilong Zhao**, Mingyu Gao, Fangxin Liu, Yiwei Hu, Zongwu Wang, Han Lin, Ji Li, He Xian, Hanlin Dong, Tao Yang, Naifeng Jing, Xiaoyao Liang, and Li Jiang, UM-PIM: DRAM-based PIM with Uniform & Shared Memory Space, in *51st International Symposium on Computer Architecture (ISCA'24)*
- Weidong Cao, **Yilong Zhao**(共同一作), Adith Boloor, Yinhe Han, Xuan Zhang, and Li Jiang, "Neural-PIM: Efficient Processing-In-Memory with Neural Approximation of Peripherals," in *IEEE Transactions on Computers, (TC 21)*
- **Yilong Zhao**, Zhezhi He, Naifeng Jing, Xiaoyao Liang, and Li Jiang. Re2PIM: A Reconfigurable ReRAM-Based PIM Design for Variable-Sized Vector-Matrix Multiplication. In *Proceedings of the 2021 on Great Lakes Symposium on VLSI (GLSVLSI'21)*

专利:

- 蒋力, **赵怿龙**, “可重构架构、加速器、电路部署和计算数据流方法”。发明专利, 申请号: 202010910280.5; 授权号: CN112181895B

- 蒋力, 赵怿龙, 崔晓松, 陈云, 廖健行, “神经网络电路”。发明专利, 申请号: 202010729402.0;
公开号: CN114004344A

项目经历

上海交通大学, 先进计算机体系结构实验室; 上海期智研究院

华为合作 面向光通信、无线通信的存算一体实现 研究项目

2021.03-2022.07

研究目的是基于存算一体技术实现光通信与无线通信的接收机, 负责工作如下:

- 设计基于存算一体的光通信、无线通信系统总体架构, 包括算子拆分与算法的重构。
- 实现部分算子的电路仿真。
- 针对高功耗计算模块 AI 化, 在误差允许的条件下达到比现有数值算法更低的计算量。

上海交通大学, 先进计算机体系结构实验室,

华为合作 基于 ReRAM 的高效可靠 DNN 加速器技术研究 项目

2019.04 – 2020.04

项目研究基于 ReRAM 的 DNN 加速器中, 提升计算可靠性以及利用稀疏性提升能效, 负责工作如下:

- 设计并编写了一个周期准确的 ReRAM 神经网络加速器架构的仿真器。仿真器以 GEM5 为基础编写, 可以仿真存储器与架构中各计算电路模块的交互行为, 从而更准确地仿真计算周期数与功耗等指标。
- 根据项目要求, 改写仿真器以支持稀疏性网络的计算与可靠性的评估, 仿真器的结果作为项目评估算法的重要指标。
- 设计一个基于 ReRAM 的 DNN 加速器架构, 以支持稀疏性网络的计算。