

Zebin Yang

13731109063 2401111912@stu.pku.edu.cn

ABOUT ME

I am a PH.D. student at SEC Lab at Peking University, advised by Prof. Meng Li. My current research interest lies in **Efficient Embodied AI System, Large Language Model, Neural Network Model Compression**, and other emerging technologies. I received my bachelor's degree in applied physics from Peking University under the supervision of Prof. Meng Li.

RESEARCH INTERESTS

- Automation Engineering
- Computer Aided Design
- Application of MCUs

EDUCATION

Peking University, Beijing, China

Sept. 2024 - Now

Ph.D., School of Electronic Engineering and Computer Science

Peking University, Beijing, China

Sept. 2020 - Jun. 2024

B.S., School of Electronic Engineering and Computer Science

SELECTED HONORS & AWARDS

- UBIQUANT scholarship Dec. 2023
- Lee Wai Wing scholarship Dec. 2022
- Merit Student Sep. 2022, 2023
- Outstanding Graduates at Peking University Jul. 2024

PUBLICATIONS

- **Zebin Yang***, Renze Chen*, Taiqiang Wu, Ngai Wong, Yun (Eric) Liang, Runsheng Wang, Ru Huang, Meng Li, "MCU-BERT: Memory-Efficient BERT Inference on Commodity Microcontrollers.", ICCAD'24
- Shuzhang Zhong, **Zebin Yang**, Ruihao Gong, Runsheng Wang, Ru Huang, Meng Li, "ProPD: Dynamic Token Tree Pruning and Generation for LLM Parallel Decoding.", ICCAD'24
- Chenqi Lin*, Tianshi Xu*, **Zebin Yang***, Runsheng Wang, Ru Huang, Meng Li, "FastQuery: Communication-efficient Embedding Table Query for Private LLM Inference.", DAC'24
- Renjie Wei*, Linfeng Zhong*, Songqiang Xu*, **Zebin Yang**, Qingyu Guo, Yuan Wang, Runsheng Wang, Meng Li, "Light-Mamba: Efficient Mamba Acceleration on FPGA with Quantization and Hardware Co-design.", DATE'25
- **Zebin Yang**, Sunjian Zheng, Tong Xie, Tianshi Xu, Bo Yu, Fan Wang, Jie Tang, Shaoshan Liu, Meng Li, "EfficientNav: Towards On-Device Object-Goal Navigation with Navigation Map Caching and Retrieval.", NeurIPS'25