Shuan Yang

College of Integrated Circuits & Micro-Nano Electronics
Fudan University

sayang25@m.fudan.edu.cn
+135 8560 8985
satreeby.github.io
orcid:0009-0006-8450-6720

EDUCATION

Sep 2025 – present **Fudan University**
Ph.D. in Integrated Circuit Science and Engineering
Advisor: Prof. Chixiao Chen

Sep 2021 – Jun 2025 **Fudan University**
B.E. in Microelectronic Science and Engineering
GPA: 3.72/4.00 (Top 15%)

RESEARCH AREAS

HW/SW Co-Design — Domain-specific accelerator design for sparse and irregular workloads
GPU/CUDA optimization and computer architecture

Efficient AI — Large Language Model inference optimization (KV cache, speculative decoding, quantization)

PUBLICATIONS

(* indicates equal contribution.)

Conference

ISCA 2026 **NS-FPS: Accelerating Farthest Point Sampling via Neighbor Search in Large-Scale Point Clouds**
Jiapei Zheng*, **Shuan Yang***, Siqi He, Chixiao Chen and Qi Liu

RESEARCH EXPERIMENTS

Sep 2025 – Mar 2026 **Research Topic**: Acceleration of Core Operators in Point Cloud Neural Networks

- Reformulated Farthest Point Sampling (FPS) as an iterative neighbor-search problem, reducing complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$.
- Proposed a hardware-software co-designed acceleration framework and and conducted detailed performance analysis (simulation + workload modeling).
- Developed high-performance implementations on CPU (C++) and GPU (CUDA).

Mar 2026 – Now **Research Topic**: Audio LLM Inference Optimization (Ongoing)

- Currently exploring inference acceleration (token pruning & quantization) for Audio LLMs, focusing on bottleneck analysis and algorithm-system co-optimization.

AWARDS

Awards and Honors

2024 Second Prize, Shanghai Division, 8th National College Student Integrated Circuit Innovation and Entrepreneurship Competition

2022 Second Prize, Shanghai Division, 14th National College Student Mathematics Competition (Non-Mathematics Category)

2022 Second-Class Award, Fudan University Outstanding Undergraduate Scholarship (Top 15%)

SKILLS

Programming	Python, C/C++, LaTex, Verilog, CUDA
Deep Learning	PyTorch
Languages	Chinese (native), Engilish (CET6)
Tools & Framework	Gem5, Ramulator, Dram-sim