Youwei Xiao

Chinese Name: 肖有为
+86 185 1920 4005
shallwe@pku.edu.cn | uv-xiao.github.io | GitHub

Education

Bachelor of Science degree in Computer Science — September 2018 - July 2022
School of Electronics Engineering and Computer Science, Peking University — Beijing, China

Ph.D. Candidate in EDA — September 2022 - now
School of Integrated Circuits, Peking University — Beijing, China

Research

My research focuses on software techniques for MLSys, computer architecture, and EDA, with emphasis on domain-specific languages (DSLs) and compiler techniques.

Hardware Synthesis & EDA: I develop multi-level intermediate representations and synthesis tools that bridge high-level specifications and RTL implementations. Key contributions include *Cement* (FPGA '24), a cycle-deterministic embedded HDL in Rust, and *Hector* (ICCAD '22), an MLIR-based multi-level IR for hardware synthesis. I also explore e-graph techniques for hardware optimization in the *SkyEgg* project.

Computer Architecture: I work on automated generation of domain-specific accelerators and custom instructions. *Cayman* (DAC '25) combines application profiling with design space exploration for accelerator generation. *ISAMORE* (ASPLOS '26) uses e-graph anti-unification to find reusable custom instructions.

Hardware-Software Co-Design: I lead the *APS* project (ICCAD '25 Invited), an open-source framework that generates optimized processor architectures, RTL implementations, and compiler support from target applications—towards fully automated co-design without manual intervention.

ML Compilers & Systems: I am actively researching tensor compiler optimization, distributed compilation, mega-kernel compilation, and LLM inference optimization including KV-Cache management for agentic AI infrastructure.

Publications

- **Youwei Xiao**, Chenyun Yin, Yitian Sun, and Yun Liang. 2026 (to appear). Finding Reusable Instructions via E-Graph Anti-Unification. In Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS '26).
- **Youwei Xiao**, Yuyang Zou, Yansong Xu, Yuhao Luo, Yitian Sun, Chenyun Yin, Ruifan Xu, Renze Chen, and Yun Liang. 2025. APS: Open-Source Hardware-Software Co-Design Framework for Agile Processor Specialization. In Proceedings of the 44th IEEE/ACM International Conference on Computer-Aided Design (ICCAD '25). [Invited Paper]
- Weijie Peng*, **Youwei Xiao***, Yuyang Zou, Zizhang Luo, and Yun Liang. 2025. Clay: High-level ASIP Framework for Flexible Microarchitecture-Aware Instruction Customization. In Proceedings of the 44th IEEE/ACM International Conference on Computer-Aided Design (ICCAD '25). (*Equal contribution)
- **Youwei Xiao**, Fan Cui, Zizhang Luo, Weijie Peng, and Yun Liang. 2025. Cayman: Custom Accelerator Generation with Control Flow and Data Access Optimization. In Proceedings of the 62nd ACM/IEEE Design Automation Conference (DAC '25).
- **Youwei Xiao**, Zizhang Luo, and Yun Liang. 2025. cmt2: Rule-Based Hardware Description in Rust with Temporal Semantics. In 5th Workshop on Languages, Tools, and Techniques for Accelerator Design (LATTE '25).
- Fan Cui, **Youwei Xiao**, Kexing Zhou, and Yun Liang. 2025. An Empirical Comparison of LLM-based Hardware Design and High-level Synthesis. In Proceedings of the 2025 ACM/SIGDA International Symposium on Field Programmable Gate Arrays (FPGA '25).

- Fan Cui, Chenyang Yin, Kexing Zhou, **Youwei Xiao**, Guangyu Sun, Qiang Xu, Qipeng Guo, Yun Liang, Xingcheng Zhang, Demin Song, and Dahua Lin. 2024. OriGen: Enhancing RTL Code Generation with Code-to-Code Augmentation and Self-Reflection. In Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design (ICCAD '24).
- **Youwei Xiao**, Zizhang Luo, Kexing Zhou, and Yun Liang. 2024. Cement: Streamlining FPGA Hardware Design with Cycle-Deterministic eHDL and Synthesis. In Proceedings of the 2024 ACM/SIGDA International Symposium on Field Programmable Gate Arrays (FPGA '24).
- Ruifan Xu, **Youwei Xiao**, Jin Luo, and Yun Liang. 2022. HECTOR: A Multi-Level Intermediate Representation for Hardware Synthesis Methodologies. In Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design (ICCAD '22).

Preprints

- Yuyang Zou, **Youwei Xiao**, Yansong Xu, Chenyun Yin, Yuhao Luo, Yitian Sun, Ruifan Xu, Renze Chen, and Yun Liang. Aquas: Enhancing Domain Specialization through Holistic Hardware-Software Co-Optimization based on MLIR. arXiv:2511.22267.
- **Youwei Xiao**, Yuyang Zou, and Yun Liang. SkyEgg: Joint Implementation Selection and Scheduling for Hardware Synthesis using E-graphs. arXiv:2511.15323.
- **Youwei Xiao**, Zizhang Luo, Weijie Peng, Yuyang Zou, and Yun Liang. Cement2: Temporal Hardware Transactions for High-Level and Efficient FPGA Programming. arXiv:2511.15073.

Honors and Awards

Second Place in EDAthon 2020 — August 2020
Programming Competition on Electronic Design Automation — IEEE CEDA Hong Kong Chapter

Shenzhen Stock Exchange Scholarship — 2019-2020
Awarded to 17 students from EECS, Peking University

Merit Student — 2018-2019 & 2019-2020
Awarded to top 10% students for comprehensive excellence

Yang Fuqing & Wang Yangyuan Academician Scholarship — 2018-2019
Awarded to only 5 students at EECS, Peking University, for academic excellence

Second Prize in NOI 2017 — July 2017
National Olympiad in Informatics — China Computer Federation

Second Prize in APIO 2017 — May 2017
Asia Pacific Informatics Olympiad (China District) — China Computer Federation

Tutorials

APS: An MLIR-Based Hardware-Software Co-design Framework — January 2026
ASP-DAC 2026 — Tokyo, Japan

Agile Hardware Specialization: A toolbox for Agile Chip Front-end Design — May 2025
ISEDA 2025 — Xi'an, China

Agile Hardware Specialization (AHS): A toolbox for Agile Chip Front-end Design — March 2025
ASPLOS 2025 — Rotterdam, Netherlands

Agile Hardware Specialization: A toolbox for Agile Chip Front-end Design — March 2025
DATE 2025 — Lyon, France

AHS: An EDA toolbox for Agile Chip Front-end Design — January 2025
ASP-DAC 2025 — Tokyo, Japan

Projects and Research

APS: Agile Processor Specialization 2023 – Present
Open-source framework for automated processor specialization from applications to RTL GitHub

Cement: Cycle-Deterministic eHDL 2022 – Present
Embedded HDL in Rust with temporal transactions for FPGA design GitHub

Hector: Multi-Level Hardware IR 2021 – 2022
MLIR-based multi-level intermediate representation for hardware synthesis GitHub

ISAMORE: Instruction Reuse via E-Graphs 2024 – 2025
E-graph anti-unification for finding reusable custom instructions

Cayman: Custom Accelerator Generation 2024 – 2025
Accelerator generation with control flow and data access optimization

SkyEgg: Hardware Synthesis with E-Graphs 2024 – Present
Joint implementation selection and scheduling using e-graphs

Skills

Programming Languages: Rust, C++, Python, Scala, Verilog, CUDA
Compiler Infrastructure: MLIR, LLVM
ML Frameworks: PyTorch, Tilelang