Yang, Chien-Yi

+1(858)-241-7536 chy036@ucsd.edu github.com/doctry

Research Interests

My research interest lies in multi-objective software-hardware co-optimization, computer architecture, machine learning (ML), electronic design automation (EDA), and quantum computing.

Education

University of California San Diego **September 2022 – present**
Ph.D. in Computer Science and Engineering

National Taiwan University **September 2017 – June 2021**
Bachelor of Electrical Engineering

Experience

IEEE Computer Architecture Letter **September 2023 – present**
reviewer

University of California San Diego **September 2022 – present**
PhD student

- A processing-in-memory (PIM)-specific compiler for DNN workloads. Reduced the number of cycles by 20x on a realistic PIM model (Samsung HBM-PIM) using a mixed-integer linear programming (MILP)-based optimization technique.
- Multi-objective software-hardware co-optimization with constraints for PIM that improves the constraint satisfaction rate by 3.3x, increasing the accuracy and PPA by 4.28% accuracy improvement, 35.38% power reduction,49x speedup, and 10% area reduction.
- Developed an algorithmic-neural network mixed framework for the Rectilinear Steiner Minimum Tree Problem (RSMT) to increase the scalability from 1k points of prior works to 10k points.

MediaTek **December 2021 – June 2022**
R&D Engineer

- Designed and implemented a reinforcement learning method that speeds up floorplanning by more than 3 times.
- Deployed a distributed training system of an LLM on a cluster of $>$ 100 GPUs and sped up the system by 30x by careful profiling and reducing the critical path.
- Designed an information retrieval system by augmenting transformers to reduce the error rate by 40%.

Maxeda Technology **September 2021 – December 2021**
R&D Engineer

- Designed experiments to verify a reinforcement learning-based chip design model to justify solution quality.

National Taiwan University **September 2017 – June 2021**
Research Assistant

- Designed and developed a novel qubit mapping framework and algorithm that scales up to 20,000 qubits (only 127 in prior works).

Publications

- **Chien-Yi Yang**, Minxuan Zhou, Flavio Ponzina, Suraj Sathya Prakash, Raid Ayoub, Pietro Mercati, Mahesh Subedar, Tajana Rosing. 2024. Multi-objective software-hardware co-optimization for HD-PIM via noise-aware Bayesian optimization. ICCAD'24.
- Kahng, Andrew B., Robert R. Nerem, Yusu Wang, and **Chien-Yi Yang**. "NN-Steiner: A mixed neural-algorithmic approach for the rectilinear Steiner minimum tree problem." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 12, pp. 13022-13030. 2024. (**Alphabetical Order**)
- Cheng, Chin-Yi, **Chien-Yi Yang**, Yi-Hsiang Kuo, Ren-Chu Wang, Hao-Chung Cheng, and Chung-Yang Huang. "Robust qubit mapping algorithm via double-source optimal routing on large quantum circuits." ACM Transactions on Quantum Computing (2024).
- Ennen, Philipp, Federica Freddi, Chyi-Jiunn Lin, Po-Nien Kung, RenChu Wang, **Chien-Yi Yang**, Da-shan Shiu, and Alberto Bernacchia. "Hierarchical Representations in Dense Passage Retrieval for Question-Answering." In Proceedings of the Sixth Fact Extraction and VERification Workshop (FEVER), pp. 17-28. 2023.
- Chang, F.C., Tseng, Y.W., Yu, Y.W., Lee, S.R., Cioba, A., Tseng, I.L., Shiu, D.S., Hsu, J.W., Wang, C.Y., **Yang, C.Y.** and Wang, R.C., 2022, July. Flexible chip placement via reinforcement learning: late breaking results. In Proceedings of the 59th ACM/IEEE Design Automation Conference (pp. 1392-1393).

Projects

BoCoEL

Bayesian Optimization as a Coverage Tool for Evaluating Large Language Models ***200+ stars***

- Accurately evaluate large language models with just tens of samples from your selected corpus.
- Uses the power of Bayesian optimization to select an optimal subset of samples for large language models to evaluate.

Qsyn

An open-sourced quantum circuit compilation framework ***100+ stars***

- A qubit mapping framework that scales up to 20,000 qubits.

Skills

Programming Languages: C++, Python, Rust, Tcl, JavaScript, Java, Matlab
Technologies: Git, Shell Script, Docker
Natural Languages: Chinese, English, Japanese, Spanish