Rahul Bera

✉ rahul.bera@safari.ethz.ch
https://sites.google.com/view/rahulbera/home
Currently in Zürich, Switzerland

Research Interest

My main research interest is in computer architecture, focused around high-performance microarchitecture and memory system design, with a focus on practical data-driven policies inside processor. I am also excited about novel, fundamentally-efficient computing paradigms for machine learning algorithms, specifically targeted to healthcare, critical patient monitoring, and personal medicine applications.

Education

2019–Present **ETH Zürich**, *Doctor of Philosophy*,
Advisor: Prof. Dr. Onur Mutlu.
Expected graduation: mid 2025

2014–2017 **Indian Institute of Technology, Kanpur**, *Master of Technology*,
Advisor: Prof. Dr. Mainak Chaudhuri.
Masters Thesis: Adaptive Prefetch Filter to Mitigate Prefetcher Induced Pollution

Professional Experience

April 2023 – Aug 2023 **Processor Architecture Research Lab, Intel Labs**, *Research Intern*.
- Mentors: Anant V. Nori and Sreenivas Subramoney
- Worked on dynamic instruction elimination for high-performance and power-efficient processor design.

Feb 2017 – Aug 2019 **Processor Architecture Research Lab, Intel Labs**, *Architecture Researcher*.
- Mentors: Anant V. Nori and Sreenivas Subramoney
- Worked on high-performance prefetcher design and near-cache computation for DNN inference.

May 2015 – Dec 2015 **Advanced Micro Devices**, *Co-op Engineer*.
- Mentor: Dr. Kanishka Lahiri
- Developed a simulator to model AMD Data Fabric for performance projections of highly-threaded SoCs.

Awards and Honors

2024 **Best paper award in ISCA 2024**, *for "Constable: Improving Performance and Power Efficiency by Safely Eliminating Load Instruction Execution".*

2023 **Distinguished artifact award in MICRO 2023**, *for "Victima: Drastically Increasing Address Translation Reach by Leveraging Underutilized Cache Resources".*

2022 **Best paper award in MICRO 2022**, *for "Hermes: Accelerating Long-Latency Load Requests via Perceptron-Based Off-Chip Load Prediction".*

Key Publications

ISCA 2024 **Constable: Improving Performance and Power Efficiency by Safely Eliminating Load Instruction Execution**,
Rahul Bera, Adithya Ranganathan, Joydeep Rakshit, Sujit Mahto, Anant V. Nori, Jayesh Gaur, Ataberk Olgun, Konstantinos Kanellopoulos, Mohammad Sadrosadati, Sreenivas Subramoney, Onur Mutlu.
Best paper award at ISCA 2024.

MICRO 2023 **Victima: Drastically Increasing Address Translation Reach by Leveraging Underutilized Cache Resources**,
Konstantinos Kanellopoulos, Hong Chul Nam, F. Nisa Bostanci, Rahul Bera, Mohammad Sadrosadati, Rakesh Kumar, Davide Basilio Bartolini, and Onur Mutlu.
Distinguished artifact award at MICRO 2023.

MICRO 2023 **Utopia: Efficient Address Translation using Hybrid Virtual-to-Physical Address Mapping**,
Konstantinos Kanellopoulos, Rahul Bera, Kosta Stojiljkovic, Can Firtina, Rachata Ausavarungnirun, Nastaran Hajinazar, Jisung Park, Nandita Vijaykumar, Onur Mutlu.

MICRO 2022 **Hermes: Accelerating Long-Latency Load Requests via Perceptron-Based Off-Chip Load Prediction**,
Rahul Bera, Konstantinos Kanellopoulos, Shankar Balachandran, David Novo, Ataberk Olgun, Mohammad Sadrosadati, Onur Mutlu.
Best paper award at MICRO 2022.

ISCA 2022 **Sibyl: Adaptive and Extensible Data Placement in Hybrid Storage Systems Using Online Reinforcement Learning**,
Gagandeep Singh, Rakesh Nadig, Jisung Park, Rahul Bera, Nastaran Hajinazar, David Novo, Juan Gómez Luna, Sander Stuijk, Henk Corporaal, and Onur Mutlu.

MICRO 2021 **Pythia: A Customizable Hardware Prefetching Framework Using Online Reinforcement Learning**,
Rahul Bera, Konstantinos Kanellopoulos, Anant V. Nori, Taha Shahroodi, Sreenivas Subramoney, and Onur Mutlu.

MICRO 2021 **BurstLink: Techniques for Energy-Efficient Conventional and Virtual Reality Video Display**,
Jawad Haj-Yahya, Jisung Park, Rahul Bera, Juan Gómez Luna, Efraim Rotem, Taha Shahroodi, Jeremie Kim, and Onur Mutlu.

ISCA 2021 **REDUCT: Efficient Scaling of DNN Inference on Multi-core CPUs with Near-Cache Compute**,
Anant V. Nori, Rahul Bera, Shankar Balachandran, Joydeep Rakshit, Om J Omer, Avishaii Abuhatzera, Kuttanna Belliappa, and Sreenivas Subramoney.

MICRO 2019 **DSPatch: Dual Spatial Access Prefetcher**,
Rahul Bera, Anant V. Nori, Onur Mutlu, and Sreenivas Subramoney.

Key Patents

2023 **Method and Apparatus for Leveraging Simultaneous Multithreading for Bulk Compute Operations**,
Anant Nori, Rahul Bera, Shankar Balachandran, Joydeep Rakshit, Om Ji Omer, Sreenivas Subramoney, Avishaii Abuhatzera, Belliappa Kuttanna.
US Patent US20230205692A1

2021 **Apparatuses, Methods, and Systems for Dual Spatial Pattern Prefetcher**,
Rahul Bera, Anant V. Nori, and Sreenivas Subramoney.
US Patent US20210089456A1

2020 **Adaptive Spatial Access Prefetcher Apparatus and Method**,
Rahul Bera, Anant V. Nori, Sreenivas Subramoney, and Hong Wang.
US Patent US10713053B2

Research Talks

Nov 2022 **Taming The Memory Wall via Machine Learning Assisted Microarchitecture Design**,
Huawei Research Center, Zürich, Switzerland.

Nov 2022 **Hermes: Accelerating Long-Latency Load Requests via Perceptron-Based Off-Chip Load Prediction**,
Processor Architecture Research Lab, Intel Labs, India.

Sept 2024 **A Case for Data-Aware Microarchitecture for Alleviating Memory Bottleneck**,
Huawei Research Center, Zürich, Switzerland.