Rahul Bera

✉ write2bera@gmail.com
https://rbera.com
Last updated: Aug, 2026

Research Interest

My research interest lies in computer architecture, especially on high-performance microarchitecture and memory system design, with a focus on practical data-driven policies in processor. I am also interested in novel, fundamentally-efficient computing systems, for emerging applications (e.g., large-scale drug discovery, personalized medicine).

Education

Sept 2019 – Dec 2025 **ETH Zürich**, *Doctor of Science*,
Advisor: Prof. Dr. Onur Mutlu.
Thesis: *"Mitigating the Memory Bottleneck with Machine-Learning-Driven and Data-Aware Microarchitectural Techniques"*, defended on October 1, 2025

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

Jul 2010 – Jun 2014 **Jadavpur University**, *Bachelor of Engineering*.
Thesis: *"Design and Analysis of Logarithmic Multipliers and Dividers using VHDL"*

Professional Experience

March 2026 – Present **Huawei Switzerland**, *Senior Researcher*.
- Working on disaggregated memory system and next-generation CPU/NPU core design

March 2023 – Aug 2023 **Processor Architecture Research Lab, Intel Labs**, *Graduate Research Intern*.
- Mentors: Anant V. Nori and Sreenivas Subramoney

Feb 2017 – Aug 2019 **Processor Architecture Research Lab, Intel Labs**, *Architecture Researcher*.
- Mentors: Anant V. Nori and Sreenivas Subramoney

May 2015 – Dec 2015 **Advanced Micro Devices**, *Co-op Engineer*.
- Mentor: Dr. Kanishka Lahiri

Honors and Recognitions

2026 **Distinguished artifact award at HPCA 2026**, *for "Athena: Synergizing Data Prefetching and Off-Chip Load Prediction via Online Reinforcement Learning"*.

2025 **MLCommons ML and Systems Rising Stars**, *awarded in recognition of the research at the intersection of machine learning and systems*.

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

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

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

Key Publications

Please visit `https://dblp.org/pid/250/2580.html` *for a complete list of publications.*

MICRO 2026 **Trail: Effective, Low-Cost and Scalable Temporal TLB Prefetching via Page-Table-Embedded Deltas**,
Konstantinos Kanellopoulos, Konstantinos Sgouras, Harsh Songara, Rahul Bera, Onur Mutlu.

IEEE Micro 2026 **Machine Learning-Driven Intelligent Memory System Design: From the On-Chip Caches to the Storage**,
Rahul Bera, Rakesh Nadig, Onur Mutlu.

ICS 2026 **Harmonia: Enhancing Data Placement and Migration in Hybrid Storage Systems via Multi-Agent Reinforcement Learning**,
Rakesh Nadig, Vamanan Arulchelvan, Rahul Bera, Taha Shahroodi, Gagandeep Singh, Andreas Kosmas Kakolyris, Ismail Emir Yüksel, Mohammad Sadrosadati, Jisung Park, Onur Mutlu.

ISCA 2026 **Revelator: Rapid Data Fetching via System-Software-Guided Hash-based Speculative Address Translation**,
Konstantinos Kanellopoulos, Konstantinos Sgouras, Harsh Songara, Andreas Kosmas Kakolyris, Vlad-Petru Nitu, Spiros Galanopoulos, Rahul Bera, Konstantina Koliogeorgi, Rakesh Kumar, Onur Mutlu.

HPCA 2026 **Athena: Synergizing Data Prefetching and Off-Chip Load Prediction via Online Reinforcement Learning**,
Rahul Bera, Zhenrong Lang, Caroline Hengartner, Konstantinos Kanellopoulos, Rakesh Kumar, Mohammad Sadrosadati, Onur Mutlu.
Distinguished artifact award at HPCA 2026.

HPCA 2026 **Conduit: Programmer-Transparent Near-Data Processing Using Multiple Compute-Capable Resources in SSDs**,
Rakesh Nadig, Vamanan Arulchelvan, Mayank Kabra, Harshita Gupta, Rahul Bera, Nika Mansouri Ghiasi, Nanditha Rao, Qingcai Jiang, Andreas Kosmas Kakolyris, Yu Liang, Mohammad Sadrosadati, Onur Mutlu.

ASPLOS 2025 **Virtuoso: Enabling Fast and Accurate Virtual Memory Research via an Imitation-based Operating System Simulation Methodology**,
Konstantinos Kanellopoulos, Konstantinos Sgouras, F. Nisa Bostanci, Andreas Kosmas Kakolyris, Berkin Kerim Konar, Rahul Bera, Mohammad Sadrosadati, Rakesh Kumar, Nandita Vijaykumar, and Onur Mutlu.

ASPLOS 2025 **CIPHERMATCH: Accelerating Homomorphic Encryption-Based String Matching via Memory-Efficient Data Packing and In-Flash Processing**,
Mayank Kabra, Rakesh Nadig, Harshita Gupta, Rahul Bera, Manos Frouzakis, Vamanan Arulchelvan, Yu Liang, Haiyu Mao, Mohammad Sadrosadati, and Onur Mutlu.

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, and 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, and 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, and 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.

Patents

2024 **Methods & Apparatus for Efficiently Processing Load Instruction Sequences**,
Adithya Ranganathan, Rahul Bera, Joydeep Rakshit, Sujit Mahto, Anant V. Nori, Jayesh Gaur, and Sreenivas Subramoney.
US Patent US20260003626A1

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

Invited Talks

May 2025 **Mitigating the Memory Bottleneck with Data-Driven and Data-Aware Microarchitectures**,
Stanford University, Meta, and Tenstorrent.

Nov 2024 **Data-Driven and Data-Aware Microarchitectures for Mitigating Memory Bottleneck in High-Performance Computing System**,
Apple Lonestar Design Center, Austin, US.

Nov 2024 **Data-Driven and Data-Aware Microarchitectures for Mitigating Memory Bottleneck in High-Performance Computing System**,
AMD Research and Advance Development, Austin, US.

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

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.