

NANDITA VIJAYKUMAR

Department of Computer Science
Email: nandita@cs.toronto.edu
Web: <http://www.cs.toronto.edu/~nandita/>

University of Toronto
40 St. George St.
Toronto, Canada

RESEARCH INTERESTS

Computer systems and architecture with a focus on the interaction between programming models, compilers, systems, and architecture; Systems and architectures for robotics and machine learning; Compilers; Memory systems; GPUs; Heterogeneous Systems;

PROFESSIONAL EXPERIENCE

University of Toronto, Assistant Professor [Jun 2020 - present]
Department of Computer Science

University of Toronto Scarborough, Assistant Professor [Jun 2020 - present]
Department of Computer and Mathematical Sciences

Intel Labs, Research Scientist [Dec 2019 - Sep 2020]
Memory and Accelerator Lab

ETH Zurich, Visiting Researcher in the Systems Group [Apr 2018 - Dec 2019]
Rich Cross-Layer Abstractions for Specialized Architectures

Nvidia Research, Graduate Intern with Dave Nellans and Eiman Ebrahimi [Jun 2017 - Aug 2017]
A Holistic Cross-Layer Abstraction to Express and Exploit Data Locality in GPUs

Microsoft Research, Graduate Intern with Olatunji Ruwase and Trishul Chilimbi [Jun 2016 - Aug 2016]
Compressed and Optimized Models for Deep Neural Network Training

Intel, Graduate Intern with Chris Wilkerson (Intel Labs) and Kingsum Chow (Intel SSG) [Jun 2014 - Dec 2014]
Architectural Support for Managed Languages

Advanced Micro Devices, Design Engineer [July 2011 - July 2013]
Architecture/Performance Modeling

ABB, Undergraduate Intern [Jan 2011 - May 2011]
Design and Verification of Low Power Control Products

BEML, Bangalore, Summer Intern [Jun 2009 - Aug 2009]
Embedded and Control Systems in Metro Trains.

Advanced Micro Devices, Undergraduate Intern [Aug 2009 - Dec 2009]
Architecture/Performance Modeling

EDUCATION

Carnegie Mellon University Aug 2013 – Oct 2019 Advisors: Prof. Onur Mutlu, Prof. Phillip B. Gibbons Thesis: <i>Rethinking cross-layer abstractions to enhance productivity, portability, and performance.</i>	Ph.D. in Electrical and Computer Engineering
Carnegie Mellon University Aug 2013 – Oct 2019 Advisors: Prof. Onur Mutlu, Prof. Phillip B. Gibbons	Masters in Electrical and Computer Engineering Current GPA: 3.90/4.00
PES Institute of Technology Undergraduate Thesis Advisor: Prof. B. K. Arunkumar Aug 2007 – May 2011	B.E. Electrical Engineering GPA: 9.68/10.00

Undergraduate Research: *Neural networks and fuzzy logic in designing control systems for motor drives.*

PUBLICATIONS

- Echo: Compiler-based GPU Memory Footprint Reduction for LSTM RNN Training* [ISCA 2020]
Bojian Zheng, Nandita Vijaykumar, Gennady Pekhimenko
- SysScale: Exploiting Multi-domain Dynamic Voltage and Frequency Scaling for Energy Efficient Mobile Processors* [ISCA 2020]
Jawad Haj-Yahya, Mohammed Alser, Jeremie Kim, A. Giray Yaglikci, Nandita Vijaykumar, Efraim Rotem, and Onur Mutlu
- SMASH: Co-designing Software Compression and Hardware-Accelerated Indexing for Efficient Sparse Matrix Operations* [MICRO 2019]
Konstantinos Kanellopoulos, Nandita Vijaykumar, Christina Giannoula, Roknoddin Azizi, Skanda Koppula, Nika Mansouri Ghiasi, Taha Shahroodi, Juan Gomez-Luna, and Onur Mutlu
- CROW: A Low-Cost Substrate for Improving DRAM Performance, Energy Efficiency, and Reliability* [ISCA 2019]
Hasan Hassan, Minesh Patel, Jeremie S. Kim, A. Giray Yaglikci, Nandita Vijaykumar, Nika Mansouri Ghiasi, Saugata Ghose, and Onur Mutlu
- A Case for Richer Cross-layer Abstractions: Bridging the Semantic Gap with Expressive Memory* [ISCA 2018]
Nandita Vijaykumar, Abhilasha Jain, Diptesh Majumdar, Kevin Hsieh, Gennady Pekhimenko, Eiman Ebrahimi, Nastaran Hajinazar, Phillip B. Gibbons, Onur Mutlu
- The Locality Descriptor: A Holistic Cross-Layer Abstraction to Express Data Locality in GPUs* [ISCA 2018]
Nandita Vijaykumar, Kevin Hsieh, Eiman Ebrahimi, Phillip B. Gibbons, Onur Mutlu
- Gaia: Geo-Distributed Machine Learning Approaching LAN Speeds* [NSDI 2017]
Kevin Hsieh, Aaron Harlap, Nandita Vijaykumar, Dimitris Konomis, Greg Ganger, Phillip B. Gibbons, Onur Mutlu
- SoftMC: A Flexible and Practical Infrastructure for Enabling Experimental DRAM Studies* [HPCA 2017]
Hasan Hassan, Nandita Vijaykumar, Samira Khan, Saugata Ghose, Kevin Chang, Gennady Pekhimenko, Oguz Ergin, Onur Mutlu
- Zorua: A Holistic Approach to Resource Virtualization in GPUs* [MICRO 2016]
Nandita Vijaykumar, Kevin Hsieh, Gennady Pekhimenko, Samira Khan, Ashish Shrestha, Saugata Ghose, Adwait Jog, Phillip B. Gibbons, Onur Mutlu
- Accelerating Pointer Chasing in 3D-Stacked Memory: Challenges, Mechanisms, Evaluation* [ICCD 2016]
Kevin Hsieh, Samira Khan, Nandita Vijaykumar, Kevin K. Chang, Amirali Boroumand, Saugata Ghose, Onur Mutlu
- Transparent Offloading and Mapping (TOM): Enabling Programmer-Transparent Near-Data Processing in GPU Systems* [ISCA 2016]
Kevin Hsieh, Eiman Ebrahimi, Gwangsun Kim, Niladrish Chatterjee, Mike O'Connor, Nandita Vijaykumar, Onur Mutlu, Stephen W. Keckler
- Toggle-Aware Bandwidth Compression for GPUs* [HPCA 2016]
Gennady Pekhimenko, Evgeny Bolotin, Nandita Vijaykumar, Mike O'Connor, Onur Mutlu, Todd C. Mowry, Stephen W. Keckler
- ChargeCache: Reducing DRAM Latency by Exploiting Row Access Locality* [HPCA 2016]
Hasan Hassan, Gennady Pekhimenko, Nandita Vijaykumar, Vivek Seshadri, Donghyuk Lee, Oguz Ergin, Onur Mutlu
- A Case for Core-Assisted Bottleneck Acceleration in GPUs: Enabling Flexible Data Compression with Assist Warps* [ISCA 2015]
Nandita Vijaykumar, Gennady Pekhimenko, Adwait Jog, Abhishek Bhowmick, Rachata Ausavarungnirun, Chita Das, Mahmut Kandemir, Todd C. Mowry, Onur Mutlu

BOOK CHAPTERS

- Decoupling the Programming Model from Resource Management in Throughput Processors* [Many Core Computing: Hardware and Software, IET, 2019]

Nandita Vijaykumar, Kevin Hsieh, Gennady Pekhimenko, Samira Khan, Ashish Shrestha, Saugata Ghose, Adwait Jog, Phillip B. Gibbons, Onur Mutlu

A Framework for Accelerating Bottlenecks in GPU Execution with Assist Warps [Advances in GPU Research and Practice, Morgan Kaufmann, 2016]

Nandita Vijaykumar, Gennady Pekhimenko, Adwait Jog, Saugata Ghose, Abhishek Bhowmick, Rachata Ausavarungnirun, Chita Das, Mahmut Kandemir, Todd C. Mowry, Onur Mutlu

OPEN-SOURCE TOOLS AND INFRASTRUCTURE

Expressive Memory: A Full-System Cross-Layer Interface in CPUs

A cross-layer interface implemented in RISC-V cores on an FPGA with full-stack support. It enables compiler, OS, and architecture research in hardware-software codesigns by supporting efficient communication of higher-level program information to hardware components (ISCA 2018).

(Artifact analysis work in preparation for submission.)

SoftMC: Software Memory Controller

An FPGA-based testing platform that can control and test memory modules designed for the commonly-used DDR interface with a C++-based API (HPCA 2017).

<https://github.com/CMU-SAFARI/SoftMC>

IMPICA: In-Memory Pointer Chasing Accelerator

A gem5-based simulator that models an in-memory pointer chasing accelerator, its corresponding driver, and its applications (ICCD 2016). The simulator has been used as a starting point for PIM (processing-in-memory) research.

<https://github.com/CMU-SAFARI/IMPICA>

AWARDS AND HONORS

Invited to Rising Stars in Computer Architecture [2018]
Georgia Tech

Invited to Rising Stars in EECS [2017]
Stanford University

Qualcomm Innovation Fellowship Finalist [2015 – 2016]
Qualcomm, USA

Benjamin Garver Lamme/Westinghouse Fellowship [2013 – 2014]
Carnegie Mellon University

Prof. MRD Merit Scholarship [2007 – 2011]
PES Institute of Technology

Spotlight Award for Outstanding achievement in deploying clustering algorithms for workload organization [2012]
Advanced Micro Devices, India

Spotlight Award for Specialized and customer specific workload analysis [2012]
Advanced Micro Devices, India

Distinction Awards for Academic Excellence [2007 - 2011]
PES Institute of Technology

INVITED TALKS AND POSTERS

Rethinking the Hardware-Software Contract: Enabling practical and general cross-layer optimizations

- ◇ VMware Research, Palo Alto, CA [July 2019]
- ◇ AMD Research, Santa Clara, CA [May 2019]
- ◇ Penn State University, State College, PA [April 2019]
- ◇ Simon Fraser University, Vancouver, BC [April 2019]
- ◇ University of Chicago, Chicago, IL [April 2019]

◇ University of Southern California, Los Angeles, CA	[April 2019]
◇ University of Waterloo, Waterloo, ON	[March 2019]
◇ Rutgers University, New Brunswick, NJ	[March 2019]
◇ Duke University, Durham, NC	[March 2019]
◇ University of Toronto, Toronto, ON	[March 2019]
◇ University of British Columbia, Vancouver, BC	[March 2019]
◇ Boston University, Boston, MA	[March 2019]
◇ University of Pennsylvania, Philadelphia, PA	[February 2019]
◇ University of California, Santa Barbara, CA	[February 2019]
◇ University of Texas at Austin, TX	[February 2019]
◇ PDL, Carnegie Mellon University, PA	[January 2019]
◇ MSR India, Bangalore, India	[January 2019]
 <i>Expressive Memory: Rethinking the Hardware-Software Contract with Rich Cross-Layer Abstractions</i>	
◇ Intel, Portland, OR	[September 2020]
◇ Penn State University, State College, PA	[November 2018]
◇ University of Illinois at Urbana-Champaign, Urbana-Champaign, IL	[November 2018]
◇ PDL Retreat, Carnegie Mellon University, Bedford Springs, PA	[October 2018]
◇ Massachusetts Institute of Technology, Cambridge, MA	[October 2018]
◇ CALCM Seminar, Carnegie Mellon University, Pittsburgh, PA	[October 2018]
◇ Rising Stars in Computer Architecture, Georgia Tech, Atlanta, GA	[October 2018]
◇ EPFL, Lausanne, Switzerland	[September 2018]
◇ ETH Zurich, Switzerland	[September 2018]
 <i>Towards Practical and Powerful Hardware-Software Interfaces to Bridge the Semantic Gap</i>	
◇ Poster at CWWMCA Workshop at MICRO-51, Fukuoka, Japan	[October 2018]
 <i>A Case for Richer Cross-layer Abstractions: Bridging the Semantic Gap with Expressive Memory</i>	
◇ ISCA-45, Los Angeles, CA	[June 2018]
 <i>The Locality Descriptor: A Holistic Cross-Layer Abstraction to Express Data Locality in GPUs</i>	
◇ ISCA-45, Los Angeles, CA	[June 2018]
 <i>A Rich Cross-Layer Interface to Enhance Application Expressivity</i>	
◇ Poster at Intel Science and Technology Center (ISTC) Retreat, Santa Clara, CA	[October 2017]
 <i>Cross-Layer Compute and Memory Abstractions for Enhanced Programmability, Portability, and Performance</i>	
◇ Poster at Rising Stars in EECS, Stanford University	[November 2017]
 <i>Zorua: A Holistic Approach to Resource Virtualization in GPUs</i>	
◇ MICRO-49, Taipei, Taiwan	[October 2016]
◇ CALCM Seminar, Carnegie Mellon University, Pittsburgh, PA	[October 2016]
 <i>A Framework for Accelerating Bottlenecks in GPU Execution with Assist Warps</i>	
◇ ETH Zurich, Switzerland	[January 2016]
 <i>A Case for Core-Assisted Bottleneck Acceleration in GPUs: Enabling Flexible Data Compression with Assist Warps</i>	
◇ ISCA-42, Portland, OR	[June 2015]
◇ Penn State University, State College, PA	[June 2015]
 <i>Energy-Efficient Data Compression for Modern Memory Systems</i>	
◇ Qualcomm Innovative Fellowship Finals, San Diego, CA	[Mar 2015]

TEACHING

Carnegie Mellon University , Teaching Assistant with Prof. Phil Gibbons <i>Optimizing Compilers, Graduate</i>	[Spring 2017]
Carnegie Mellon University , Teaching Assistant with Prof. Onur Mutlu <i>Computer Architecture, Graduate</i>	[Fall 2015]
PES Institute of Technology , Teaching Assistant with Prof. Abha Tripathi <i>Power Systems Analysis, Undergraduate</i>	[Fall 2010]
PES Institute of Technology , Teaching Assistant with Prof. S. Venkatesh <i>Digital Signal Processing, Undergraduate</i>	[Spring 2010]
PES Institute of Technology , Teaching Assistant with Prof. Gayathri Devi <i>Linear Integrated Circuits, Undergraduate</i>	[Spring 2009]

STUDENTS SUPERVISED

Ruofan Liang PhD Research, University of Toronto	[2020-present]
Sankeerth Durvasula PhD Research, University of Toronto	[2020-present]
Jimmy Lin MSc Research, University of Toronto	[2020-present]

STUDENTS UNOFFICIALLY SUPERVISED

Christina Giannula PhD Research, National University of Athens. <i>Efficient synchronization of processing-in-memory architectures.</i>	[2019-present]
Hasan Hassan PhD Research, ETH Zurich. <i>DRAM testing infrastructures (HPCA 2017) and efficient DRAM substrates (HPCA 2016, ISCA 2019).</i>	[2015-present]
Konstantinos Kanellopoulos Research Internship, ETH Zurich. <i>Hardware-Software codesign for sparse linear algebra (MICRO 2019).</i>	[2018-present]
Nika Mansouri Masters Research, ETH Zurich, <i>Automatic code offload for PIM architectures.</i>	[2018-present]
Bojian Zheng Masters Research, University of Toronto. <i>EcoRNN: Fused LSTM RNN Implementation with Data Layout Optimization</i>	[2018-present]
Mehrshad Lotfi Research Internship, ETH Zurich. <i>Towards Practical and Realizable Interfaces to Enhance Application Expressivity</i>	[2018]
Abhilasha Jain Masters Research, CMU. <i>Cross-layer Interfaces for Efficient Caching (ISCA 2018).</i>	[2017]
Diptesh Majumdar Masters Research, CMU. <i>Cross-layer Interfaces for Data Placement in Heterogeneous Memories (ISCA 2018).</i>	[2017]
Ashish Shrestha Masters Research, CMU. <i>Zorua: A Holistic Approach to Resource Virtualization in GPUs (MICRO 2016).</i>	[2016]
Mahmoud Khairy Research Internship, CMU. <i>Efficient DRAM Refreshes for GPUs.</i>	[2015]
Madhav Iyengar Graduate Research Project, CMU. <i>Introducing Heterogeneity in GPU Architectures.</i>	[2015]

Jonathan Leung [2015]
Graduate Research Project, CMU. *Introducing Heterogeneity in GPU Architectures.*

Gaurav Srivastava [2015]
Graduate Research Project, CMU. *Improving Warp Scheduling in GPUs*

Elliot Rosen [2015]
Graduate Research Project, CMU. *Improving Warp Scheduling in GPUs*

Abhishek Bhowmick [2013]
Undergraduate Internship, CMU. *Enabling Flexible Data Compression in GPUs (ISCA 2015)*

GRANTS

NSF Award, *CSR-Core: Effective Data Compression for Modern Memory Systems*, National Science Foundation (Award #1423172). *Contributed to writing and ideas.* [2014–2017]

NSF Expeditions Collaborative Proposal, *Prescriptive Memory: Razing the Semantic Wall between Applications and Computer Systems.* *Contributed to writing and ideas.* [2018]

SERVICE

Program Committee member: ISCA 2021, USENIX ATC 2021, HPCA 2021, GPGPU 2020, PACT 2020, SRC PACT 2020, IISWC 2019, GPGPU 2019

Reviewer: ASPLOS 2021, CAL 2020, TACO 2020, TRETTS 2020, ICS 2018, MICRO 2017, ICS 2017, PLDI 2017, ISCA 2014-2017, MICRO 2014-2015, HPCA 2014-2017, PACT 2014, DAC 2014-2015, IISWC 2014, ICCD 2014, MICRO Top Picks 2015.

Panelist: Diversity Workshop in SOSP 2019.

Judge: SRC student research competition in SOSP 2019.