Nikita Lazarev | CV

Boston – MA, USA

Research Interests

Computer Hardware: Networking Hardware, Hardware Offload, SmartNICs, FPGAs
Computer Systems: Networking and Distributed Systems, Kernel Bypass Technologies
Application Domains: Low-Latency Systems, Datacenter Systems, 5G and vRAN

Current Research Focus: (1) Pushing further and making practical the tight integration of datacenter machines via closely-coupled with processors FPGAs; (2) Making 5G vRAN stacks as a cloud native feature by leveraging advances in low-latency networking and realtime distributed systems.

Education

Massachusetts Institute of Technology — **Boston, MA, USA**
Computer Engineering: PhD — *2024, expected*

Swiss Federal Institute of Technology (EPFL) — **Lausanne, Switzerland**
Computer Science: MS — *2018*

Bauman Moscow State Technical University — **Moscow, Russia**
Electrical Engineering & Robotics: Engineer — *2016*

Research Experience

Industry

Microsoft Azure for Operators Research (part-time researcher) — **Redmond, WA, USA**
Toward energy efficient 5G/vRAN in the cloud — *06.2022 - now*

Design of an inter-layer MPC-based control loop to enable low power consumption of cloud-native 5G systems. The technique promises up-to 30% of energy saving on average daily load on the vRAN-driven cellular networks.

Microsoft Azure for Operators Research (part-time researcher) — **Redmond, WA, USA**
Layer disaggregation and workload migration in 5G/vRAN stacks — *06.2021 - 06.2022*

Enabling MAC/PHY layer disaggregation and us-scale 5G workload migration in the future Azure vRAN infrastructure. I developed a DPDK-based component for the layer disaggregation, and implemented and tested cell migration logic in FlexRAN, both in local and distributed settings under the real 5G workload. The technique improves fault tolerance and resiliency of cloud-native 5G system.

Microsoft Research, Systems and Networking lab (internship) — **Cambridge, UK**
FPGA-based implementation of a transport network layer for the specific workload — *06.2018 - 09.2018*

Fork of the Microsoft FaRM project. The goal is to design a CPU-free datastore architecture for the applications in distributed graph databases and key-value stores. Such applications cause intense network traffic composed of small randomly addressed packets, and my task was to design the transport layer adopted for such workload. As the result, the network performance was improved by 30% in comparison with the previously used transport layer.

Microsoft Research, Systems lab (internship) — **Bangalore, India**
Hardware acceleration of real-time IoT AI algorithms — *08.2017 - 02.2018*

Two ML algorithms have been optimized at MSRI in order to fit on resource-constrained IoT devices. In this project, I was working on a design of an FPGA-based heterogeneous SoC for efficient implementation of those optimized algorithms in hardware. The designed SoC enables low-latency implementation of the algorithms, and it is generic: both algorithms can be implemented leveraging exposed SW interfaces without changing the hardware.

Samsung Research Center (part-time) **Moscow, Russia**
Junior software developer: compiler technologies and systems *2014 - 2016*

Android ART compiler and runtime optimization. Focus on compiler optimization: loop transformations, division optimization, heterogeneous back-end support. Focus on runtime optimization: hash tables for Java strings. Research focus: speed-up of class loading by static analysis of the class references.

Academia

Cornell University/MIT **Ithaca/Boston, USA**
Independent research, supervisor: Prof. Zhiru Zhang, Prof. Christina Delimitrou *02.2020 -*
FPGAs in datacenters: exploring the opportunities of closely-coupled FPGAs for improving networking performance and acceleration of datacenter tax functions.

Swiss Federal Institute of Technology (EPFL) **Lausanne, Switzerland**
Term project, supervisor: Prof. Babak Falsafi *02.2018 - 07.2018*
Development of a lightweight VLIW-like SIMD core for a neural processor

Swiss Federal Institute of Technology (EPFL) **Lausanne, Switzerland**
Research assistantship: setup of a quantum mechanics experiment *02.2017 - 07.2017*
Development and production of a fast FPGA-based PID control system for lasers

Bauman Moscow State Technical University **Moscow, Russia**
Term project *2013 - 2014*
Phasechronometric measurement device: FPGA programming and PCB design

Lomonosov Moscow State University **Moscow, Russia**
Independent project *2012 - 2013*
Brain - computer interface: RF-PCB design, microcontroller firmware development, deep learning

Teaching Experience

Cornell University **Ithaca, NY, USA**
Digital logic and computer organization (ECE-2300): teaching assistant *2021*
Keywords: computer architecture, FPGA, Verilog

EPFL **Lausanne, Switzerland**
Real-time embedded systems (CS-476): teaching assistant *2018*
Keywords: FPGA, hardware accelerator, VHDL

Microsoft Research **Bangalore, India**
Tutorial: introduction to FPGAs *2017*
Keywords: FPGA, hardware accelerator

Bauman Moscow State Technical University **Moscow, Russia**
Implementation of control systems: guest lecturer *2015*
Keywords: control systems, PID, microcontrollers

Key implementation skills

Software skills

Top programming languages: C++, Python, C, Scala, Java

Domains: Kernel bypass networking, distributed systems, vRAN and 5G, realtime and embedded systems, machine learning, compiler optimization and runtime systems

Hardware skills

RTL languages: System Verilog, VHDL, Chisel

HLS languages: SystemC

Domains: SmartNICs and networking devices, AI accelerators

Key research/theoretical skills

Theory of algorithms, distributed systems, computer systems, machine learning, probability and statistics, complexity theory, basics of signal processing and control, basics of game theory

Languages

Russian: Mother tongue

English: Fluent

French: Elementary

Selected Publications

- **HiveMind: a hardware-software system stack for serverless edge swarms**
  L. Patterson, D. Pigorovsky, B. Dempsey, N. Lazarev, A. Shah, C. Steinhoff, A. Bruno, J. Hu, C. Delimitrou
  Proceedings of the 49th Annual International Symposium on Computer Architecture, June, 2022
- **A Roadmap for Enabling a Future-Proof In-Network Computing Data Plane Ecosystem**
  D. Kim, N. Lazarev, T. Tracy, F. Siddique, H. Namkung, J. Hoe, V. Sekar, K. Skadron, Z. Zhang, S. Seshan
  Arxiv pre-print, October, 2021
- **Efficient and Fast RPCs in Cloud Services with Near-Memory Reconfigurable NICs**
  Nikita Lazarev, Neil Adit, Shaojie Xiang, Zhiru Zhang, Christina Delimitrou
  Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, April, 2021
- **Dagger: Towards Efficient RPCs in Cloud Microservices with Near-Memory Reconfigurable NICs**
  Nikita Lazarev, Neil Adit, Shaojie Xiang, Zhiru Zhang, Christina Delimitrou
  IEEE Computer Architecture Letters, August, 2020
- **Phase-Chronometric System for Monitoring Turning Processes**
  D. Boldasov, N. Lazarev, A. Syritsky
  The Devices (in Russian), May, 2015

Patents

- **Project Orion: Transparently converting a shared memory channel between L1 and L2 into an over-the-network channel**
  A. Kalia, D. Kim, I. Marinos, T. Jiang, N. Lazarev, V. Bahl, Pending (Approved) US patent, 2021
- **Precise FPGA-based time measurement system for real-time turning process monitoring**
  A. Syritsky, D. Boldasov, N. Lazarev, A. Komshin
  Registration number: RU2019610688

Journal/Conference Peer Review Service

CAL'22: Computer Architecture Letters; **SC'22**: The International Conference for High Performance Computing, Networking, Storage, and Analysis; **FCCM'21**: 2021 IEEE 29th Annual International Symposium on Field Programmable Custom Computing Machines; **FPGA'21**: The ACM/SIGDA International Symposium on Field-Programmable Gate Arrays; **FCCM'20**: 2020 IEEE 28th Annual International Symposium on Field Programmable Custom Computing Machines.

Conference Organizing Committee Service

FCCM'22: 2022 IEEE 30th Annual International Symposium on Field Programmable Custom Computing Machines, New York, USA.

Achievements

- IEEE Micro "Top Picks from the Computer Architecture Conferences": Honorable Mention, 2022
- Cornell University Graduate Research Fellowship (Jacobs Scholarship), 2019
- Russian Presidential Scholarship, 2016
- Medal for Excellent Graduation, 2016
- Samsung R&D "Above and Beyond" Award, 2015