## Konstantin Burlachenko Ph.D. candidacy in Computer Science program, CEMSE division at KAUST

@ burlachenkok@gmail.com @ konstantin.burlachenko@kaust.edu.sa 

in linkedin.com/in/burlachenkok github.com/burlachenkok



i https://burlachenkok.github.io

ORCID 0000-0001-5986-0855

I have co-created systems for ML, Graphics, Vision, and Physics Simulation, exploiting hardware via DSL and using contemporary areas of Math&CS. My current focus is Federated Learning, a subfield of ML co-invented by my advisor in 2016<sup>1</sup>, which becomes a key part of the "US National AI R&D Plan" in 2023. My dissertation is "Optimization Methods and Software for Federated Learnina".

## **EDUCATION**

2020-Now	KSA: Ph.D. candidacy in CEMSE/CS Program at KAUST. Member of Prof. Peter Richtárik's Optimization	
	Machine Learning Lab in KAUST Al initiative led by Jürgen Schmidhuber. Transcript: Link-1. GPA: 3.81/4.0	
2015-2019	USA, Stanford : Graduate Non-Degree Program. <i>Transcript</i> : Link-2, <i>GPA</i> : 3.96/4.3, <i>Total Credits</i> : 28	
2015-2018	USA, Stanford : Data, Models and Optimization Graduate Certificate Link-3 (Program)	
2016 - 2019	USA, Stanford : Artificial Intelligence Graduate Certificate Link-4 (Program)	
2003-2009	Russia, Bauman Moscow State Technical University: MS in CS. GPA: Not Applicable	
Conferences and	ferences and ICLR 2024; ACM CoNEXT 2023; Rising Stars in Al Symposium 2024 and 2023, EMNLP-2022; ICML-2022; ICML	
Summer Schools	2021; NeurIPS-2021; ACM CoNEXT 2021; Regularization Methods for ML 2021; The PRAIRIE/MIAI AI 2021;	
	Oxford ML-2021; The HSE/MIPT Optimization without Border; ACM SIGGRAPH 2012.	

## **EXPERIENCE**

June 2024 September 2024

Internship in Private Federated Learning ML Team, APPLE, Cambridge, UK

Developing innovative approach for memory-efficient on-device fine-tuning of large language models Distributed Math Optimization | Al | Federated Learning

August 2021

Research Scientist Intern (AI) in AI and Systems Co-Design, META, Menlo Park, USA The internship has not happened due to 15 months process of obtaining J1 VISA to USA Distributed Optimization | Al | Federated Learning

## Now September 2020

## CS Ph.D. candidacy and a member of Prof. Peter Richtárik Optimization and ML Lab<sup>2</sup>, KAUST, KSA

- Narrow area of research: Federated Learning, Stochastic Distributed Math Optimization
- ▶ Broad area of interests: Math Optimization, AI/ML, Compute Optimization, Software Systems, GPGPU, Computer Vision, Computer Graphics, Networks, Cryptography, Control, Physical Simulation
- ➤ Sep 2022 Sep 2023 Member of SDAIA-KAUST AI (invitation letter)
- ▶ Awards: (1) Dean's Award, 2020; (2) Grant from Saudi Authority for Data and Artificial Intelligence, 2022;(3) Dean's List Award, 2023; (4) AMD MI50 from AMD Inc., 2023.; (5) Winning grant from Grand Challenge Project Proposal Shaheen III CPU, 2024; (6) Co-secured a 4-year RDIA grant for the lab, 2025. Distributed Optimization | Federated Learning | Applied and Fundamental Math | Computer Vision | Cryptography

C/C++ CUDA AVX512 Python CMake Qt/PyQt PyTorch TF Latex Git Google Tests Bash NLP

## August 2020 March 2019

## Principal Engineer Level 18 | Foundation AI Lab, HUAWEI, Moscow, Russia

- ▶ R&D in internal ML Systems middleware for HUAWEI HiSilicon and HUAWEI CBG
- ▶ Present HiSilicon solutions for engineers, scientists working with ML/AI OpenTalks.AI, HUAWEI News
- ► Awards: Grade A for a 2019-2020 Year Progress with a one-time payment bonus

Math Optimization | Al | Custome ISA | C/C++ | Python | TVM | Java | Google Protobuf | CMake | Qt | TF | SQL

## March 2019 July 2014

## Senior Developer Technology Engineer Level IC3, NVIDIA, Moscow, Russia

- ▶ Driveworks SDK SDK for self-driving cars adopted by automotive partners. Computer vision, machine learning, calibration, egomotion. Implementation and presentation of the modules internally
- ▶ PhysX/Apex SDK An industry standard for game physics simulation, graphical special effects. Internal implementation and communication with Blizzard
- cuDNN/cuBLAS libraries GPU computation libraries used by more than 1M customers in machine learning and HPC. Implementation, Documentation, and collaboration with Mathworks
- RAPIDS GPU based implementation of SkLearn, XgBoost, Pandas, I was resposible for SkLearn
- ► Awards: Funding support of visiting two NVIDIA GTC conference

CUDA | GLSL | C++ | AARCH64 | SSE2/ARM NEON | Linux | Windows | PS4 | XBox | OpenGL | Google Tests | GitLab Perl Python CMake Make Qt Git TensorFlow Computer Vision Graphics Deep Learning CppCheck

<sup>1.</sup> Federated Learning: Strategies for Improving Communication Efficiency [J.Konečný, H.B.McMahan, F.X.Yu, P.Richtarik, A.T.Suresh, D.Bacon, NIPS 2016]

<sup>2.</sup> To avoid academic disputes over authorship order and focus on work, the Lab adopts the concept of "Every Author as First Author" arXiv:2304.01393

## July 2014 May 2013

### Senior Developer Engineer | Yandex Video Team, YANDEX, Moscow, Russia

- ► Text and statistical machine learning features for Yandex Video Search
- ▶ Infrastructure for storage and analysis of web documents with embedded video on the WWW
- ► Awards: Two one-time payment bonuses due to the delivery of products in production

C++ Google Protobuf JavaScript Bash Python Computer Science HTML/JS/CSS SVN MapReduce ML

## April 2013 March 2012

### Team Lead Physics Engine Developer, FITTING REALITY, Moscow, Russia

- ▶ Develop library for clothing simulation in CUDA and in OpenCL with facade interface to C++/C#
- Custom render engine for clothing visualization compatible with OpenGL 1.2
- ▶ Prepare elements of the demo for investors. Carry internal MATH/CS/PHYS training
- ► Manage a team of 4 members
- ► Awards: Funding visit of SIGGRAPH 2012, USA conference

C++ C OpenGL GLSL Qt Posix WinAPI QMake CUDA OpenCL Physics Graphics Ogre C#

## March 2012 September 2010

## Software Developer Engineer, ACRONIS, Moscow, Russia

- ► Key member of GUI team for Acronis Backup and Recovery 2011 Enterprise
- ▶ Profiling and optimization of the codebase working in user/kernel space for Windows OS

C++ C WinAPI WinDbg VmWare Specialized GUI library SVN SysInternals CppCheck ASM x86 AqTime

## September 2010 March 2009

## Senior Software Developer Engineer, CAPITAL RESEARCH, Moscow, Russia

Developed Firefox plugin to create the three-dimensional HTML view for basic HTML elements

Firefox C++ WinAPI HTML/JS/CSS Windows OpenGL GLSL SVN

## June 2009 December 2006

## C++ Programming Engineer, FLINT AND CO, Moscow, Russia

- Created several computer games with computer vision and graphics part, hardware drivers
- Spent time on factory floors to test and analyze the quality of my solutions. Carry trips to customers

  C++ SDL Posix WinApi Development Image Library Low-level programming Computer Vision OpenGL SVN

## November 2006 March 2006

## C++ Programming Engineer (Part-time work), ASTRASOFT TECHNOLOGY, Moscow, Russia

▶ Developed visual elements of management system based on Qt and OpenGL

C++ Qt Windows OpenGL SVN

## \* Presentations

MAY-2024 APPLE Inc, UK, USA, KSA remotely: Research talk for APPLE Inc.

MAY-2024 ICLR 2024, Austria: Error Feedback Reloaded

APR-2024 NVIDIA Inc, USA and KSA remotely: Research talk for NVIDIA Inc.

MAR-2024 MLSS 2024, Japan: Error Feedback Reloaded

FEB-2024 Al Symposium, KSA: Unlocking FedNL: Self-Contained Compute-Optimized Implementation

DEC-2023 ACM Distributed ML2023, France: Federated Learning is Better with Non-Homomorphic Encryption.

JUN-2023 SIAM, USA: Fl\_PyTorch: Optimization Research Simulator for FL

MAR-2023 VCC OPEN HOUSE 2023 event, KSA: FedNL. Making Newton-Type Methods Applicable to FL.

DEC-2022 EMNLP 2022, Abu Dhabi, UAE: Presenter in KAUST Al Iniative Booth.

OCT-2022 CS Ph.D. Proposal Defense, KSA: Dissertation Title: Optimization Methods and Software for Federated Learning. Committee members: Eric Feron, Marco Canini, Peter Richtarik.

JUL-2022 ACM Symposium, Italy: MARINA: Faster non-convex distributed learning with compression.

MAR-2022 Al Symposium KAUST, KSA: FL\_PyTorch: Optimization Research Simulator for FL

DEC-2021 ACM Distributed ML2021, Rome: FL\_PyTorch: Optimization Research Simulator for Federated Learning.

JUL-2021 Spotlight for in ICML-2021, Virtual: MARINA Faster Non-Convex Distributed Learning with Compression.

APR-2021 NSF-TRIPODS Workshop, Virtual: MARINA: Faster Non-Convex Distributed Learning with Compression.

FEB-2020 OpenTalks.Al conference, Russia: Huawei technologies for Al developers.

JUL-2019 Educational Center Sirius, Russia: Deep Learning Course with D.Kamzolov and A.V. Gasnikov

DEC-2018 MIPT, Russia: Lectures about subtle things around Decision Trees, Gradient Boosting and Random Forest.

APR-2016 GTC 2016, USA: Presenter in Driveworks NVIDIA booth.

AUG-2012 ACM SIGGRAPH 2012, USA: Presenter in CentiLeo booth, and visitor from Fitting Reality.

## X EVALUATING THE WORK OF OTHERS

- 1. Reviewer in a peer-reviewed proceeding for ICML'22, AISTATS'23, JMLR'24, ICML'24 Workshop, ICLR'25.
- 2. Participates in annual review processes for estimating the work of my colleagues in HUAWEI, NVIDIA, and Yandex.

2

# 

PV-Tuning: Beyond Straight-Through Estimation for Extreme LLM Compression	2024
https://arxiv.org/abs/2405.14852 Presentation and proceedings to NeurIPS -2024, [Oral Paper, Top 0.4%]	
UNLOCKING FEDNL: SELF-CONTAINED COMPUTE-OPTIMIZED IMPLEMENTATION  This://arxiv.org/abs/2410.08760  Under a peer-review process	2024
ERROR FEEDBACK RELOADED: FROM QUADRATIC TO ARITHMETIC MEAN OF SMOOTHNESS CONSTANTS  Thttps://openreview.net/forum?id=Ch7WqGcGmb  Thttps://arxiv.org/abs/2402.10774  Presentation and proceedings to ICLR-2024.	2024
FEDERATED LEARNING IS BETTER WITH NON-HOMOMORPHIC ENCRYPTION  Thttps://dl.acm.org/doi/10.1145/3630048.3630182  Thttps://arxiv.org/abs/2312.02074  Presentation and proceedings to 4th ACM International Workshop on Distributed Machine Learning	2023
ERROR FEEDBACK SHINES WHEN FEATURES ARE RARE  https://arxiv.org/abs/2305.15264 Under a peer-review process	2023
FEDERATED LEARNING WITH REGULARIZED CLIENT PARTICIPATION  Thttps://icml.cc/virtual/2023/27049 https://arxiv.org/abs/2302.03662  Presentation. Workshop Federated Learning and Analytics in Practice at ICML 2023	2023
SHARPER RATES AND FLEXIBLE FRAMEWORK FOR NONCONVEX SGD WITH CLIENT AND DATA SAMPLING  https://openreview.net/forum?id=zKgJ6TWAFE https://arxiv.org/abs/2206.02275  Proceedings Transactions on Machine Learning Research (TMLR)	2022
FEDERATED OPTIMIZATION ALGORITHMS WITH RANDOM RESHUFFLING AND GRADIENT COMPRESSION  **Thttps://icml.cc/virtual/2023/27050** https://arxiv.org/abs/2206.07021*  **Presentation and proceedings to NeurIPS -2024** Presentation Workshop Federated Learning and Analytics ICML 2023**	2022
FASTER RATES FOR COMPRESSED FEDERATED LEARNING WITH CLIENT-VARIANCE REDUCTION  1. https://epubs.siam.org/doi/pdf/10.1137/23M1553820  2. https://arxiv.org/abs/2112.13097  1. Proceedings SIAM Journal on Mathematics of Data Science (SIMODS).	2021
FL_PyTorch: Optimization Research Simulator for Federated Learning  https://dl.acm.org/doi/abs/10.1145/3488659.3493775/  https://arxiv.org/abs/2202.03099  Presentation and proceedings to 2nd ACM International Workshop on Distributed Machine Learning	2021
MARINA: FASTER NON-CONVEX DISTRIBUTED LEARNING WITH COMPRESSION  **Thttps://proceedings.mlr.press/v139/gorbunov21a.html*  **Thttps://arxiv.org/abs/2102.07845  **Presentation and proceedings to ICML 2021*	2021
Personalized federated learning with communication compression  **This://openreview.net/pdf?id=dZugyhbNFY**   This://arxiv.org/abs/2209.05148  **Proceedings Transactions on Machine Learning Research (TMLR)**	21 – 2022

Engineering Expertise	
Programming Languages that I have used	C89/C99, C++20/11/03, C#, Cython, Java ,x86/AArch64, NDA ASM
Scripting Languages that I have used	Python, Bash, Perl, and Cython which is between scipting and compile languages.
DSL Languages that I have used	Gl SL, TVM, Google Protobuf, CUDA, OpenCL, Matlab, R, SQL
Frameworks	Qt, CUDA, WinApi, Posix, OpenGL, OpenCL, PyTorch, TensorFlow, CvxPy
Operating Systems	Windows, Linux based, Orbis, XBox, Android, NDA OS-es
Development Environments	QtCreator, Visual Studio, Eclipse, WinDbg, Android Studio, TexStudio, Nsight
General purpose development tools	SysInternals,AqTime,Cmake,GNU Toolchain,CppCheck,Valgrind,Git,QMake
Typing DSL Languages	Latex, HTML, XML, Markdown
Areas in which I worked	Federated Learning, Stochastic Distributed Math Optimization, Computer Vision, Statistical Learning, System Programming, HPC and GPU Programming, Math Optimization, Computer Graphics, Computational Physics, Distributed Systems
Sports achievements	The Candidate Master in chess by FIDE. (My FIDE profile). Second Place in the KAUST Chess Tournament in Oct 2022.

## TECHNICAL NOTES

Technical Note: From C++98 to C++2x

2022-2024

🖸 github.com/burlachenkok/CPP\_from\_1998\_to\_2020/blob/main/Cpp-Technical-Note.md 🛮 🗗 Short Information The technical note is dedicated for all primary C++ programming language standards: C++03/98/11/14/17/20. The note has been adopted to AMD Inc. internal education web portal. Also it has been highly assessed by: (i) Marco Foco, Head of the Delegation for Italy ISO C++ JTC1/SC22/WG21; (ii) Prof. David Keyes, Director of Extreme Compute Research Center at KAUST.

Technical Note: Exploring Python3 Language from a Computing Perspective

2023

🖸 github.com/burlachenkok/exploring-python3/blob/main/python3-note.md 🛮 🗗 Short Information The note contains a language tutorial, a description of several libraries, and a description of several low-level profiling tools for Linux/Posix OS and Windows OS Family. It has been adopted to Introduction to Data Science workshop series at KAUST.

## Selected Projects

### MATH OPTIMIZATION RESEARCH STUDIO

KAUST, 2020

☑ Project report - Math Optimizaiton Research Studio ☑ Description ☑ Bitbucket report

CS380: Math Optimization Research Studio. Self-developed command line interpreter with custom script language with self-developed backend computation in C++/CUDA.

C++ Linux Windows CUDA CMake Dot Google Test Python Bash

#### EXPERIMENTAL NEURAL NET FRAMEWORK

STANFORD, 2019

☑ Report.CS230 - 2019 ☑ Poster CS230 - 2019 ☑ bitbucket repo ☑ Presentation

CS230': Experimental Neural Net Framework. Wall-clock relative speedup compared to Google Tensor Flow x3.5.

C++ Linux Windows CUDA Python CMake

## CONVEX OPTIMIZATION SOLVERS WITH LEVERAGING INTO GPU/CPU POWER FOR AI/ML

STANFORD, 2018

Poster CS221 - 2018 Bitbucket repo

CS221: Convex optimization solvers with leveraging into GPU/CPU power for Al/ML. Lasso regression speedup with GPU is x12.6 compared to CPU standard solvers.

C++ Linux Windows CUDA Python CMake Convex Optimization

#### CONVEX OPTIMIZATION FOR MACHINE LEARNING

STANFORD, 2017

Poster CS229 - 2017. Presentation

Stanford, CS229: Convex Optimization for Machine Learning. Several solvers that work x8 faster than SkLearn.

C++ Visual Studio Numerical Linear Algebra Convex Optimization Python CMake

#### PLOTTER++. STANDALONE TOOL FOR PLOTTING IMAGES, GRAPHS, POINT CLOUDS. COMMUNICATION:TCP/IP

HOLIDAYS, 2017

An advanced scientific plotter tool suitable to use in connection with embedded systems.

C++ Linux Windows Embedded Systems Qt Python

## HTML REPORT GENERATOR FOR GOOGLE TESTS

HOLIDAYS, 2016

**☑** GitHub

Generate and compare different Google Test reports. Used by Hyundai AutoEver, and Hapsoft Consulting.

Python HTML XML

## **66** References

## **Andrew Ng**

#### **Timour Paltashev**

Assistant Professor, Stanford, Letter AMD and Core faculty, Northwestern Polytechnic University

ang@cs.stanford.edu

Timour.Paltashev@amd.com

+1 (650)725-2593

C +1 (408) 306 8508

## **Brad Osgood**

#### Jerome H. Friedman

Professor, Stanford, Under Request Professor, Stanford, Under Request

osgood@stanford.edu

@ jhf@stat.stanford.edu

+1 (650) 387-1287

+1 (650) 723-9329