Arash Bakhtiari

arash@bakhtiari.org • bakhtiari.org • arashbakhtiari • arashb

Education

2013 – 2017 **Ph.D. in Computational Science**, *Technical University of Munich*, Germany.
- Research Topics: High Performance Computing, Parallel Algorithms, Computational Fluid Dynamics

2011 – 2013 **M.Sc. with Honors in Computational Science**, *Technical University of Munich*, Germany.
- GPA: 1.5 (scale: 1.0 - 4.0, 1.0 is the highest possible grade), Passed with high Distinction

2007 – 2011 **B.Sc. in Physics**, *Ludwig-Maximilian University of Munich*, Germany.
- GPA: 2.0 (scale: 1.0 - 4.0, 1.0 is the highest possible grade)

Professional Experiences

2020 – Now (September) **Senior HPC/DL Software Engineer**, *Blaize*, Cambridge, UK.
- Optimizing the performance of computer vision deep learning models (ResNet, OpenPose, Mask R-CNN, ...) for the Blaize Graph Streaming Processor (GSP) architecture to run inference at the edge.

2019 – 2020 (June – September) **Deep Learning Software Engineer**, *Plumerai Limited*, London, UK.
- Technically leading, designing, and developing *Larq Compute Engine*: a highly optimized deep learning inference engine for *Binarized Neural Networks* on mobile and embedded devices [MLSys'21 paper, code].
- Larq Compute Engine outperformed the current state of the art in terms of performance by a factor of 2 by using advanced optimization techniques such as CPU cache optimization, ARM NEON SIMD vectorization, multi-threading and hand-optimized kernels for ARM architecture developed in assembly programming language.
- Supervised a team of three software developers and led weekly discussions to brainstorm ideas.

2018 – 2019 (March – June) **Software Engineer**, *Intel Corporation*, Munich, Germany.
- Investigated, tested and benchmarked support of *Intel Deep Learning Boost* instructions (*AVX-512 Vector Neural Network Instructions* (VNNI)) on Intel Cascade Lake microarchitecture.
- Developed support of *Control-Flow Enforcement Technology* as a new x86 architecture extension to GNU Debugger (GDB) which is successfully shipped with Intel Parallel and System Studio 2020.
- Developed fixes and increased the coverage of testing the overall functionality of the software.

2009 – 2014 (November – October) **Software Engineer (Part-time)**, *ReliaTec GmbH*, Garching, Germany.
- Designed and Developed C++ and python parser and API for FIBEX and AUTOSAR (XML-based standardized formats for automotive electronic control units).
- Full-stack development responsibilities: system design, user interface design, source code development, unit test, system test, deployment.

2009 – 2009 (January – October) **Software Developer (Part-time)**, *Lifecycle Engineering Solutions Center*, Karlsruhe, Germany.
- Developed a simulation software for virtual wind tunnels in C++ using OpenSG framework.

Research Experiences

2014 – 2017 **PhD Student at Institute for Advanced Study**, *HPC Focus Group*, Germany.
- Developed a novel, parallel, unconditionally stable numerical algorithm to solve scalar Advection-Diffusion and incompressible Navier-Stokes partial differential equations [thesis].
- Designed and developed TbSLAS: a distributed/shared-memory parallelized fast solver which was capable of solving numerical problems with one billion unknowns on 16,384 CPU cores on the STAMPEDE system at the Texas Advanced Computing Center [SC'16 paper, code].

2014 – 2015 **Visiting Researcher at The University of Texas at Austin**, *Lab of Prof. George Biros*, USA.
- Developed a novel communication scheme in distributed-memory systems which resulted in 95% reduction of the communication overhead in semi-Lagrangian schemes (presented in SIAM Conference on Parallel Processing 2016, Paris, France).

2011 – 2013 **Master Student at Technical University of Munich**, *Lab of Prof. Hans-Joachim Bungartz*, Germany.

- Developed a novel non-blocking algorithm to overlap the computation and communication for MPI parallelized, Multi-GPU Lattice-Boltzmann solvers which resulted in parallel efficiencies of more than 90% on 24,576 CPU cores and 2048 GPUs [MDPI paper].
- Designed and implemented a distributed-memory parallelized and GPU accelerated Lattice-Boltzmann solver [thesis, code].

Technical Proficiencies

Languages	C++, Python, inline assembly, Matlab
HPC	OpenMP, MPI, CUDA, OpenCL
DL	TensorFlow, PyTorch, ONNX
Productivity	Git, LaTeX, Emacs

Publications

- T. Bannink*[1], **A. Bakhtiari***, A. Hillier*, L. Geiger*, Tim de Bruin, Leon Overweel, Jelmer Neeven, Koen Helwegen, *Larq Compute Engine: Design, Benchmark and Deploy State-of-the-Art Binarized Neural Networks*, Proceedings of the 4th MLSys Conference, San Jose, CA, USA, 2021.
- C. Riesinger, **A. Bakhtiari**, M. Schreiber, P. Neumann and H.-J. Bungartz: *A holistic scalable implementation approach of the lattice Boltzmann method for CPU/GPU heterogeneous clusters*, MDPI, Basel, 2017.
- **A. Bakhtiari**, D. Malhotra, A. Raoofy, M. Mehl, H.-J. Bungartz and G. Biros: *A Parallel Arbitrary-Order Accurate AMR Algorithm for the Scalar Advection-Diffusion* , In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis SC'16. IEEE, Salt Lake City, UT, USA, November 2016.
- **A. Bakhtiari** : *A Parallel AMR Algorithm for the Scalar Advection-Diffusion Equation*, SIAM Conference on Parallel Processing for Scientific Computing, Paris, France, April 2016.

Published Open Source Software

- **Larq Compute Engine**: A Highly optimized inference engine for Binarized Neural Networks.
- **TbSLAS**: A parallel Semi-Lagrangian/Fast Multipole Method advection-diffusion and Navier-Stokes solver.
- **Multi-GPU Turbulent LBM**: A distributed-memory parallelized Multi-GPU Lattice-Boltzmann solver.

Leadership

- As a member of Bavarian Graduate School of Computational Engineering, supervised, coordinated and led teams of seven master students to successfully finish research projects in cooperation with industry partners:

2016 – 2017 (May – March) **Carl Zeiss Microscopy**, *Data Intensive Distributed Computing Workflows in Light Microscopy (link).*

2015 – 2016 (May – March) **Siemens AG**, *CADO - Computer Aided Design Optimizer: A Topology Optimization Tool (link).*

2014 – 2015 (May – March) **GE Global Research Europe**, *SAPIENS - A New Generation MR Spectroscopy Processing, Analysis and Visualization Software (link).*

Legal Status

Work permit	EU citizenship & work permit

[1] * indicates equal contribution.