Mihail Popov

Tenured Research Scientist

Homepage
✉ *mihail.popov@inria.fr*

I am interested in high performance computing and machine learning.

▬▬ Experience

2021 - current   **Research Scientist ISFP**, *Inria STORM/TADaaM*, Bordeaux, France.
- Using AI/LLM methods for verification (**IPDPS24**).
- Exploring ML methods for performance/energy optimization (**IPDPS22, MPLR, JPDC**).
- Studying and classfying I/O behaviors.

2020   **Senior Software Researcher**, *Huawei 2012 Research Labs*, Edinburgh, UK.
- Developed compiler transformations for AI.

2018 - 2019   **Postdoc Researcher**, *UART Architecture Group Uppsala University*, Uppsala, Sweden.
- Designed ML models to optimize/explore hardware parameters and NUMA effects (**ICS20**).
- Optimized performance on NUMA systems by exploring a larger search space of thread and page mappings. The additional exploration cost is amortized with sampled execution (**ICS19**).

2017   **Postdoc Researcher**, *University of Versailles Saint-Quentin (UVSQ)*, Versailles, France.
- Characterized HPC applications performance over the last decade (**Proceedings of the IEEE**).
- Improved autotuning of compiler optimization with clustering (**CCPE**).

2013 - 2016   **Ph.D. Student**, *UVSQ/Paris-Saclay*, Versailles, France.
Title: Automatic Decomposition of Parallel Programs for Optimization and Performance Prediction
Ph.D. advisors: Assistant Prof. Pablo de Oliveira Castro and Full Prof. William Jalby
- Decomposed applications into codelets with **CERE**. Codelets map loops or OpenMP regions and are replayed as standalone programs. (**IPDPS15, TACO**).
- Proposed a piecewise holistic approach to tune compiler optimizations and thread configurations through codelets. Codelet autotuning achieves better speedups at lower tuning cost (**Euro-Par**).

2016   **Intern**, *UVSQ/Intel Software Tools*, Champaign, USA.
- Worked on profiling methods and simulation with codelets.

2013   **Intern**, *Exascale Computing Research (ECR)*, Versailles, France.
- Participated in the elaboration of a method to reduce the benchmarking cost. I clustered codes with similar computation patterns and which are sensitive to the same architectural changes (**CGO**).

2012   **Intern**, *ECR*, Versailles, France.
- Developed a method to quantify a distance between applications using compression tools.

▬▬ Ph.D. Students

2022 - ongoing   **Lana Scravaglieri**, Vectorization with numerical accuracy control for multi-precision simulation codes, `https://theses.fr/s351408`.

2024 - ongoing   **Asia Auville**, *Co-advised with Emmanuelle Saillard*, Large Language Models for Detection and Correction of Errors in HPC Applications, `https://theses.fr/s401487`.

▬▬ Awards and Grants

2024   **Inria exploratory action**, 129k€, Large Language Models for Detection and Correction of Errors, `https://www.inria.fr/en/llm4dice`.

2022   **Atos/Eviden Plan de relance**, 47k€, Statistical methods for system error detection.

2016   **Euro-Par publication in the best paper issue**, *3% acceptance rate*, 5 our of 176.

Academic Supervision

2024   **Asia Auville**, *Enserb master student*, worked on models to detect MPI errors.

2024   **Arvid Lorén**, *Uppsala University bachelor thesis*, worked methods to optimize Intel hybrid E/P cores. Bachelor thesis accessible at `https://www.diva-portal.org/smash/get/diva2:1888603/FULLTEXT01.pdf`.

2024   **Patrick Gutsche**, *ENS Lyon student*, modeled code embedding capabilities for performance and energy tuning.

2023   **Jad El Karchi**, *Enserb master student*, worked on models to detect MPI errors. Work published at **IPDPS24**.

2023   **Angel Hippolyte**, *Bordeaux University master student*, worked on on tuning different compiler optimizations. `https://inria.hal.science/hal-04315144`.

2023   **Frederic Becerril**, *ENS Lyon student*, explored and modeled different MPI parameters resulting in performance and energy gains.

2021   **Lana Scravaglieri**, *Enserb master student*, Worked on NUMA/prefetch optimizations across datasets. Work published at **JPDC**.

2019   **Anastasia Stupnikova**, *Uppsala University master thesis*, explored machine learning methods applied to optimize NUMA effects. Work published at **ICS20**.

2019   **Isaac Sánchez Barrera**, *Ph.D. student at Barcelona Supercomputing*, explored machine learning methods applied to optimize NUMA/prefetch effects. Work published at **ICS20**.

2015   **Florent Conti**, *University of Versailles bachelor student*, worked on lock OpenMP methods in checkpoint restart strategies. Work published at **IPDPS15**.

Teaching

2025   **Head teacher**, *16 hours*, Enseirb master course, Parallel programming.
Introduction to parallel programming with OpenMP and CUDA.

2022 - 2025   **Head teacher**, *12 hours*, Enseirb master course, Introduction to cryptography.
Presented the basics of symmetric and asymmetric cryptography.

2021 - 2024   **Teaching Assistant**, *20 hours*, Enseirb, C project.
Supervised students to implement games using C.

2022 - 2023   **Teaching Assistant**, *20 hours*, Enseirb, Introduction to cryptography.
Supervised exercises on the basics of symmetric and asymmetric cryptography.

2022 - 2024   **Co-organizer**, Inria, ENS student visits.
Assisted in the organization of ENS Lyon student visits.

2022 - 2024   **Teaching Assistant**, *5 hours*, Enseirb master course/Bordeaux University: HPC algo & Paper reading.
Supervised students for paper reading classes.

2019   **Head teacher**, *16 hours*, Uppsala master course, Low Level Parallel Programming.
Presented different parallelization strategies including OpenMP (threads and tasks), SIMD vectorization, C++ threads, and CUDA.

2015   **Head teacher**, *20 hours*, ISTY master course, Operating Systems Project.
Designed a project to model the Linux file system. Focused on low level structures (inodes, blocks).

2014 - 2015   **Teaching Assistant**, *36 hours*, ISTY/UVSQ, Operating Systems/Computer architecture.
- Introduction to Linux: processes and internal structures.
- Introduction to computer architecture with an emphasize on floating point arithmetics.

Volunteer Experience

My recent contributions.
- **Co-Organizer**: HPC Bugs Fest during SC 2023 & SC 2024
- **Webchair and publicty chair**: ICPP 2021.
- **PC**: COMPAS 2022, PDP 2022/2023/2024, Correctness 2023, IPDPSW 2024, ICPP 2024
- **Reviewer**: IPDPS 2022, PeerJ Computer Science 2023, SC 2023, Cluster 2023, SC 2024

Education

2010 - 2013   **Master Degree in Engineering and Computer Science**, *ISTY*, France.
European master degree of computer science.

2012 - 2013   **Master Degree in High Performance Computing**, *UVSQ/Ecole Centrale Paris.*
Master degree obtained in parallel of the last year of the master degree in engineering.

Research Interests

General   Linux, Performance Characterization, Computer Architecture
Parallelism   OpenMP (threads and tasks), MPI (Processes), Cuda, Vectorization
AI   Clustering, DNNs, Data Analysis, Automatic Differentiation, Backpropagation Optimization
Compilation   LLVM, Code Analysis, Vectorization

Publications

J. Karchi, H. Chen, A. TehraniJamsaz, A. Jannesari, **M. Popov**, and E. Saillard., "Mpi errors detection using gnn embedding and vector embedding over llvm ir," in *IEEE International Parallel and Distributed Processing Symposium*, **IPDPS**, **2024**.

L. Scravaglieri, **M. Popov**, L. Lima Pilla, A. Guermouche, A. Olivier, and E. Saillard, "Optimizing performance and energy across problem sizes through a search space exploration and machine learning," *Journal of Parallel and Distributed Computing*, **JPDC, 2023**.

A. Tehranijamsaz, **M. Popov**, A. Dutta, E. Saillard, A. Jannesari, "Learning intermediate representations using graph neural networks for numa and prefetchers optimization," in *IEEE International Parallel and Distributed Processing Symposium*, **IPDPS**, **2022**.

M. Shimchenko, **M. Popov**, and T. Wrigstad, "Analyzing and predicting energy consumption of garbage collectors in openjdk," in *Managed Programming Languages and Runtimes*, **MPLR**, **2022**.

I Sánchez, D Black-Schaffer, M Moretó, M Casas, A. Stupnikova, and **M. Popov**, "Modeling and optimizing numa effects and prefetching with machine learning," in *Proceedings of the ACM International Conference on Supercomputing*, **ICS**, **2020**. 20-minute Video.

M. Popov, A. Jimborean, and D. Black-Schaffer, "Efficient thread/page/parallelism autotuning for numa systems," in *Proceedings of the ACM International Conference on Supercomputing*, **ICS**., **2019**. 5-minute Video.

W. Jalby, D. Kuck, A. D. Malony, M. Masella, A. Mazouz, and **M. Popov**, "The long and winding road toward efficient high-performance computing," in ***Proceedings of the IEEE***, **2018**.

M. Popov, C. Akel, Y. Chatelain, W. Jalby, and P. de Oliveira Castro, "Piecewise holistic autotuning of parallel programs with cere," in *Wiley Online Library Concurrency and Computation: Practice and Experience*, **CCPE**, **2017**.

M. Popov, C. Akel, W. Jalby, and P. de Oliveira Castro, "Piecewise holistic autotuning of compiler and runtime parameters," in *European Conference on Parallel Processing*, **Euro-Par**, **2016 (included in the Best Paper Issue)**.

M. Popov, C. Akel, F. Conti, W. Jalby, and P. de Oliveira Castro, "Pcere: Fine-grained parallel benchmark decomposition for scalability prediction," in *IEEE International Parallel and Distributed Processing Symposium*, **IPDPS**, **2015**.

P. D. O. Castro, C. Akel, E. Petit, **M. Popov**, and W. Jalby, "Cere: Llvm-based codelet extractor and replayer for piecewise benchmarking and optimization," in *ACM Transactions on Architecture and Code Optimization*, **TACO**, **2015**.

P. de Oliveira Castro, Y. Kashnikov, C. Akel, **M. Popov**, and W. Jalby, "Fine-grained benchmark subsetting for system selection," in *Proceedings of Annual IEEE/ACM International Symposium on Code Generation and Optimization*, **CGO**, **2014**.