Pablo de Oliveira Castro Herrero

I. Identification

Last name : **DE OLIVEIRA CASTRO HERRERO**
First name: **PABLO**
Date of birth: **03/12/1983**
Position : **Tenured assistant professor at Université Paris-Saclay - UVSQ**
Research unit: **Li-PaRAD Laboratoire Informatique Parallélisme, Réseaux et Algorithmique Distribuée (Computer Science lab - Parallelism, Networks, and Distributed Algorithms)**
Personal website: **www.sifflez.org**

Education

2007-2010 Computer Science Ph. D. from the University of Versailles St-Quentin-en-Yvelines, with honors. My Ph. D. was financed by a grant from CEA and carried out in the high-performance team of CEA's real-time and embedded systems laboratory (LASTRE). The thesis was supervised by Stéphane Louise, research engineer at LASTRE, and directed by Denis Barthou, professor at Bordeaux University. (https://www.theses.fr/2010VERS0048)

Title: *"Expression and optimization of data reorganizations on data flow parallelism"*

2003-2006 Télécom ParisTech engineering degree (formerly École Nationale Supérieure des Télécommunications).

2001-2003 Superior and special mathematics at Pierre de Fermat preparatory school in Toulouse (CPGE).

2001 Scientific Baccalaureate, with honors.

Career

2012-2022 Tenured assistant professor at Université Paris-Saclay - UVSQ, attached to the PRiSM (Parallelism, Networks, Systems, and Modeling) laboratory until 2015, then to the LI-PaRAD laboratory.

2011-2012 Postdoctoral fellow at Exascale Computing Research (joint laboratory with UVSQ, Intel, CEA, Genci). Characterization of HPC applications. Design of models to measure and predict the performance of parallel and sequential applications.

2006-2007 Research engineer at Télécom ParisTech. Specification and prototyping of the adaptation of a semantic wiki engine for use on an embedded platform in the ANR RNTL XWiki Concerto project framework.

2006 Engineering internship at AdaCore. Development of features to implement the Ada distributed systems annex. I upstreamed my work into the GCC/GNAT compilers and the PolyORB middleware.

II. Research activities

1. Research statement

The complete list of publications is available, at the end of the document, in section 6.

In High-Performance Computing (HPC), architectures and application codes are becoming increasingly complex to gain efficiency. In particular, new architectures are often heterogeneous and require an increased degree of parallelism. Moreover, to meet the needs of deep learning codes, new formats for representing floating-point numbers (*bfloat16*, *fp16*) have been introduced by manufacturers. This complexity presents delicate trade-offs between performance and numerical accuracy. On the one hand, it is necessary to ensure that new versions of the code retain sufficient numerical precision (e.g. when adding fine-grained parallelism or when vectorizing the code). On the other hand, it may be interesting to reduce the accuracy used by exploiting new floating-point formats to increase the computation speed and reduce memory exchanges and energy consumption.

My work over the last six years at the University of Versailles St-Quentin-en-Yvelines and the LI-PaRAD laboratory, facilitates this approach through automatic tools. In collaboration with Intel, UPVD, and ENS Paris-Saclay, I have developed the specialized compiler Verificarlo (github.com/verificarlo/verificarlo) based on LLVM, which provides an estimation of the numerical accuracy within large application codes. This compiler also allows optimizing code through mixed precision and simulating custom floating-point formats. Verificarlo relies mainly on Monte Carlo arithmetic, where the computational operators are noisy to model rounding or cancellation errors. This approach has been used to study and optimize accuracy in several industrial simulation codes such as YALES2 (Coria-CNRS), ABINIT (UCL and CEA), EuroPlexus (CEA), and more recently has been applied to the study of reproducibility in Magnetic Resonance Imaging (MRI) analysis in the Dipy software (dipy.org)

Previously, I have also been interested in auto-tuning solutions to fine-tune many factors such as the degree of parallelism, the placement of tasks, the work slicing, the choice of compilation passes, the numerical accuracy, or the execution frequency. The search for optimal parameters is a costly process that requires many measurements. The proposed solutions combine statistical learning methods to reduce the number of measurements performed to find the optimal parameters with protection/resumption approaches, deployed at the compiler and runtime interface to isolate parts of code and optimize them individually.

2. Doctoral and scientific supervision

Doctoral co-supervision

(02/2021- ongoing) El-Mehdi El-Arar - Ph.D. co-supervision of the thesis *"Stochastic models for numerical error evaluation."*

Co-advisor: Pablo de Oliveira Castro (50%)
Supervisor: Devan Sohier (50%)
Collaboration with E. Petit (Intel)

Funded by the ANR Interflop project. This work resulted in an oral presentation in May 2021 at the 12th Arithmetic Computing Meeting and in a publication [9] at an international conference. A journal submission is in progress.

(10/2016-12/2019) Yohan Chatelain - Ph.D. co-supervision of the thesis "*Tools for debugging and optimizing floating-point computations in an HPC context*."

Co-advisor: Pablo de Oliveira Castro (50%)
Supervisor: William Jalby (50%)
Collaborations with E. Petit (Intel) and D. Defour (UPVD)

Funded by Exascale Computing Research (ECR) in partnership with Intel and CEA in the "ECR Numerical Analysis" project framework. This work has led to two publications [12,13] at international conferences.

Current position of Y. Chatelain: Postdoctoral fellow at Concordia University, Big Data Infrastructures team, Montreal, Canada.

(10/2013-10/2016) Mihail Popov - Ph.D. co-supervision of the thesis "*Automatic decomposition of parallel programs for optimization and performance prediction*."

Co-advisor: Pablo de Oliveira Castro (50%)
Supervisor: William Jalby (50%)

Funding from the research project Investissement d'Avenir of the Fonds pour la Société Numérique, Environnement Logiciel pour le Calcul Intensif (ELCI) and from the MontBlanc 3 project. This work has resulted in two publications in international journals [5, 6] and three publications [15, 16, 17] in international conferences.

Current position of M. Popov: Research scientist at INRIA Bordeaux, STORM team.

Supervision of post-doctoral students

(01/2020-) François Coppens - Post-doctoral fellow at the TREX Center of Excellence. His work focuses on optimizing Sherman-Morrison and Woodbury kernels for QMC codes. Journal submission in process.

(10/2019-12/2021) Matei Istoan - Post-doctoral fellow at Exascale Computing Research in collaboration with Intel (E. Petit) and UPVD (D. Defour). His work focuses on obtaining numerical profiles on calls to elementary functions in simulation codes to optimize them in computation time, memory, and energy. This work has been published in an international journal [3] and two international conferences [10, 12].

(07/2013-10/2013) Yuriy Kashnikov - Post-doctoral fellow at the Exascale Computing Research joint lab. His work focused on performance analysis of HPC applications using machine learning. This work contributed to two international conference publications [17, 18].

Supervision of internships in MSc and engineering programs

(05/2022-09/2022) Mathys Jam - M2 Computation High Performance Simulation (CHPS) internship at Université Paris-Saclay, *"Optimization of TREX kernels by statistical learning methods"*. 50% co-supervision within the TREX center of excellence in collaboration with Eric Petit.

(**05/2021-09/2021**) **Aurelien Delval -** Final internship at ISTY engineering school, *"Development of the Verificarlo-CI tool"*. 100% supervision within the TREX project. The work of this internship has been published as a free software (https://github.com/verificarlo/vfc_ci_tutorial), a journal submission is in progress.

(05/2020-09/2020) Alan Vaquet - M1 Computation High Performance Simulation (CHPS) internship at Université Paris-Saclay, *"Adding a function instrumentation backend in Verificarlo"*. 100% supervision. The work of this internship contributed to a journal publication [3].

(05/2018-09/2018) Damien Thénot - M1 Computation High Performance Simulation (CHPS) internship at Université Paris-Saclay, *"Development of a tool for visualizing digital traces: VeriTracer"*. 100% supervision. The work of this internship was published as a free software (github.com/verificarlo/verificarlo/tree/veritracer/postprocess/visualizer).

(04/2016-09/2016) Yohan Chatelain- M2 Master Calcul Haute Performance Simulation (CHPS) internship at Université Paris-Saclay, *"Construction of an energetic prediction model in the HPC context"*. 100% supervision within the MontBlanc 3 project.

(05/2016-07/2016) Nicolas Derumigny - BSc internship École Normale Supérieure de Lyon, *"Placement strategies on heterogeneous architecture"*. 100% supervision. This internship led to multiple fixes being upstreamed to the GEM5 project.

(05/2015-09/2015) Yohan Chatelain - M1 Master High Performance and Simulation (CHPS) internship, *"Value profiling and specialization in LLVM"*. 100% supervision.

(05/2015-09/2015) Rayhana Ziara - M1 Master High Performance and Simulation (CHPS) internship, *"Extension of the CERE tool for MPI region capture/resume"*. 100% supervision.

(05/2014-09/2014) Florent Conti - M1 Master SeCReTS Cryptography and Computer Security internship, *"Performance analysis by clustering in the CERE tool"*. 100% supervision.

(04/2013-09/2013) Mihail Popov - Final internship at ISTY engineering school, *"Computing a distance between programs through static analysis of performance indicators"*. 100% supervision.

3. Outreach and scientific dissemination

Participation in program committees of international conferences

Member of the program committee (2015-2020) of the International **Workshop on Programming Models for SIMD/Vector Processing (WPMVP)** attached to the ACM Principles and Practice of Parallel Programming (PPoPP) conference (ppopp20.sigplan.org/home/WPMVP-2020).

Member of the program committee (2014-2019) and organizer (2017 and 2018) of the **Software Engineering for Parallel Systems (SEPS)** international workshop attached to the ACM SIGPLAN Systems, Programming, Languages, and Applications: Software for Humanity (SPLASH) conference (conf.researchr.org/home/seps-2019).

Member of the program committee of the international **PDP** conference **in 2020 and 2021** (Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, www.pdp2020.com).

I am also a reviewer for the **ACM TACO, Elsevier JSS, Elsevier Software X, Concurrency and Computation: Practice and Experience** journals.

Online courses

Creation of an online course (MOOC) on compiler construction in collaboration with Samuel Tardieu at the Institut Mines-Télécom, https://compilation-course.github.io/.

Software productions

To ensure the reproducibility of research work and to facilitate the adoption of new techniques by the community, I pay particular attention to the development and distribution of software accompagnying my research work.

Verificarlo [1,2,12,13,14] allows the automatic instrumentation and analysis of the numerical accuracy of a computational program. For this purpose, Verificarlo, built around the LLVM compiler, replaces each floating-point operation with its stochastic equivalent in Monte Carlo arithmetic or various other backends such as VPREC (backend simulating variable precision [12]). Verificarlo was

registered with the Agency for the Protection of Programs in 2017 with IDDN FR 001 080014 000 RC 2017 000 10600.

Distributed under the LLVM license at github.com/verificarlo/verificarlo.

CERE (Codelet Extractor and REplayer) [6,15,16] automatically decomposes a program into a set of elementary computational kernels called *codelets*. CERE replays each codelet in isolation for different input data sets and with varying parameters (compiler options, target architecture, degree of parallelism, etc.). Codelet decomposition accelerates optimization and performance measurement in HPC or embedded codes.

Distributed under LGPL license at github.com/benchmark-subsetting/cere.

I am also the developer and maintainer of **IRVM,** a virtual machine for executing the TREE intermediate code described by A. W. Appel in his "*Modern Compiler Implementation in ML/C/Java*" book series. Distributed under the GPL at github.com/pablooliveira/irvm/. IRVM is used in my compilation courses at UVSQ and by Samuel Tardieu in his SE202 compilation course at Télécom ParisTech. From this collaboration sprang the project to make a joint Compilation MOOC between our two institutions.

From 2016 to 2019, I was a contributing member of Debian Science, a working group to promote the dissemination and use of free software in the Debian science community. I maintained several Debian packages allowing statistical analysis using R (https://bit.ly/3RrP3Gi).

4. Scientific responsibilities

Research grants

ANR-InterFLOP project (2020-2024) https://interflop.github.io/interflop/

I am the **UVSQ's** project **leader for** the ANR (French national research agency) InterFLOP project. The InterFLOP project aims at providing a modular and scalable platform to analyze and optimize the floating-point computation costs of programs. The project takes into account new floating-point paradigms.

TREX European Centre of Excellence in Exascale Computing (2020-2023) https://www.trex-coe.eu/

I **lead the numerical precision task** within the TREX European Center of Excellence. The TREX Centre of Excellence aims to develop, promote and maintain high-performance open-source software solutions in the field of quantum chemistry and materials science, ready to exploit the upcoming exascale architectures. TREX brings together European scientists, High Performance Computing (HPC) experts and SMEs working on extremely accurate simulations in stochastic Quantum Monte Carlo (QMC) methods to develop software solutions that will greatly improve the diffusion of these unique computing tools, so far restricted to a specialist's community.

MontBlanc 3 Project (2015-2018) www.montblanc-project.eu

I was the **technical coordinator of the Europe Horizon 2020 MontBlanc 3 project** for UVSQ and I was responsible for several tasks in the work packages "Software Tools", "Hardware Accelerators" and "Applications and Benchmarks". The MB3 project was funded by Europe in its H2020 program (Horizon 2020).

The MB3 project was awarded the 2017 *HPCwire Editors' Choice Award for Best HPC Collaboration* and Stars of Europe 2019. The MB3 project brought together industrial partners (Bull/ATOS, ARM, AVL) and academic partners (University of Versailles Saint Quentin, CNRS, Barcelona Supercomputing Centre, Swiss Federal Institute of Technology in Zurich, University of Stuttgart, University of Cantabria, University of Graz). This project has designed a high-performance computing architecture of the Exascale class based on ARM processors, taking care to obtain the maximum efficiency of the codes. It has developed a software ecosystem allowing to exploit the proposed architecture efficiently.

Software Environment for High Performance Computing project (2014-2017)

I was the **leader of the "Analysis and Optimization Tools" task** in the ELCI project with the various partners Algotech, Bull, Cenaero, Cerfacs, Coria, Inria, Kitware, Onera, Safran. The ELCI project was financed by the "Investissement d'Avenir" program of the Fonds pour la Société Numérique.

5. Other activities and responsibilities

Pedagogical responsibilities

Since 2017, I am **co-head of the first year Master of Science in High Performance Computation and Simulation** (CHPS, chps.uvsq.fr) at Université Paris-Saclay.

I was the **pedagogical director** of the fourth-year engineering cycle at ISTY (Institut des Sciences et Techniques des Yvelines) from 2013 to 2016.

Participation in juries and selection committees

In 2021, I participated, as an examiner, in the jury of **Mathieu Stoffel's Ph.D. thesis**, entitled "Software-level improvement of energy efficiency associated with the execution of a high-performance computing application on a supercomputer".

In 2021, I participated, as an examiner, in the jury of **Nestor Demeure's Ph.D. thesis**, entitled "Managing the compromise between performance and accuracy in simulation codes".

In 2018, I participated in the selection committee for a position of Assistant Professor in the 27th section at the University of Versailles Saint Quentin.

In 2016, I participated in the selection committee for the recruitment of a PRAG position in mathematics and computer science at the Villebon Georges Charpak Institute (www.villebon-charpak.fr).

In 2016, I participated, as an examiner, in the jury for **Abdul W. Memon's Ph.D. thesis**, entitled "Crowdtuning: towards practical and reproducible auto-tuning via crowdsourcing and predictive analytics".

Participation in working groups

From 2014 to 2017 I participated in the pedagogy working group in charge of the development of the computer science curriculum of **Villebon Georges Charpak** (https://www.villebon-charpak.fr/). This group puts into practice innovative pedagogical approaches such as project-based learning or interdisciplinary teachings.

I belong to the **Atelier d'Écologie Politique Francilien (Political Ecology Workshop** ecopolien.hypotheses.org), which brings together members of higher education and research institutions from all disciplines who wish to build a community of scientists in the Ile-de-France region gathered around ecological issues. The workshop intends to weave links across disciplines and reflect on how to involve citizens in the construction of participative science. Écopolien has been selected for a cycle of seminars funded by the Maison des Sciences de l'Homme (MSH) Paris-Saclay in 2020, 2021, and 2022.

Participation in official institutional bodies

I was a staff representative on the UVSQ **Health, Safety and Working Conditions Committee (CHSCT)** from **2015 to 2018**.

6. Full publication list (2010-2022)

International peer-reviewed journals

[1] Numerical uncertainty in analytical pipelines lead to impactful variability in brain networks. G. Kiar, Y. Chatelain, P. de Oliveira Castro, E. Petit, A. Rokem, G. Varoquaux, B. Misic, A. C. Evans, and T. Glatard. *PLOS ONE*, 16(11), 2021.

[2] Confidence Intervals for Stochastic Arithmetic. D. Sohier, P. de Oliveira Castro, F. Févotte, B. Lathuilière, E. Petit, O. Jamond. ACM Transactions Mathematical Software (**TOMS**), 47(2), 2021.

[3] A Study of the Effects and Benefits of Custom-Precision Mathematical Libraries for HPC Codes. E. Brun, D. Defour, P. de Oliveira Castro, M. Istoan, D. Mancusi, E. Petit, A. Vaquet. IEEE Transactions on Emerging Topics in Computing (**TETC**), 2021.

[4] Comparing perturbation models for evaluating stability of neuroimaging pipelines. Gregory Kiar, P. de Oliveira Castro, P. Rioux, E. Petit, S. T Brown, A. C. Evans, T. Glatard. The International Journal of High Performance Computing Applications, 34(5):491-501, 2020.

[5] Piecewise holistic autotuning of parallel programs with CERE. M. Popov, C. Akel, Y. Chatelain, W. Jalby, P. de Oliveira Castro. Concurrency and Computation: Practice and Experience **(CPE)**, doi/10.1002/cpe.4190, 2017**.** Special issue for distinguished papers.

[6] CERE: LLVM Based Codelet Extractor and REplayer for Piecewise Benchmarking and Optimization. P. de Oliveira Castro, C. Akel, E. Petit, M. Popov, W. Jalby. ACM Transactions on Architecture and Code Optimization (**TACO**), 12(1):6, 2015. Acceptance rate 30%

[7] Adaptive Sampling for Performance Characterization of Application Kernels. P. de Oliveira Castro, E. Petit, A. Farjallah, W. Jalby. Concurrency and Computation: Practice and Experience (**CPE**), 25(17):2345-2362, 2013. Special issue for distinguished papers.

Book chapters

[8] DSL Stream Programming on Multicore Architectures. P. de Oliveira Castro, S. Louise, D. Barthou. In Sabri Pllana and Fatos Xhafa, editors, Programming Multi-core and Many-core Computing Systems. John Wiley and Sons, 2012.

International peer-reviewed conferences

[9] The Positive Effects of Stochastic Rounding in Numerical Algorithms. E. El Arar, D. Sohier, P. de Oliveira Castro, and E. Petit. In *29th IEEE Symposium on Computer Arithmetic,* ARITH 2022, Virtual Conference, 2022.

[10] Shadow computation with BFloat16 to compute numerical accuracy. D. Defour, P. de Oliveira Castro, M. Istoan, E. Petit. *28th IEEE Symposium on Computer Arithmetic*, ARITH 2021, Virtual Conference, 2021.

[11] Custom-Precision Mathematical Library Explorations for Code Profiling and Optimization. D. Defour, P. de Oliveira Castro, M. Istoan, E. Petit. In *27th IEEE Symposium on Computer Arithmetic, ARITH 2020*, pages 121-124, 2020.

[12] Automatic exploration of reduced floating-point representations in iterative methods. Y. Chatelain, E. Petit, P. de Oliveira Castro, G. Lartigue, D. Defour. 25th International Conference on Parallel Processing (**Euro-Par**), Lecture Notes in Computer Science, pages 481-494. Springer, 2019.

[13] VeriTracer: Context-enriched tracer for floating-point arithmetic analysis. Y. Chatelain, P. de Oliveira Castro, E. Petit, D. Defour, J. Bieder, M. Torrent. 25th IEEE Symposium on Computer Arithmetic, (**ARITH**), Amherst, pages 65-72. IEEE, 2018.

[14] Verificarlo: Checking Floating Point Accuracy through Monte Carlo Arithmetic. C. Denis, P. de Oliveira Castro, E. Petit. 23rd IEEE Symposium on Computer Arithmetic, (**ARITH**), Silicon Valley, pages 55-62. IEEE, 2016.

[15] Piecewise Holistic Autotuning of Compiler and Runtime Parameters. M. Popov, C. Akel, W. Jalby, P. de Oliveira Castro. 22nd International Conference on Parallel Processing (**Euro-Par**), Lecture Notes in Computer Science, pages 238-250. Springer, 2016.

[16] PCERE: Fine-grained Parallel Benchmark Decomposition for Scalability Prediction. M. Popov, C. Akel, F. Conti, W. Jalby, P. de Oliveira Castro. IEEE International Parallel and Distributed Processing Symposium (**IPDPS**), pages 1151-1160, 2015. Acceptance rate 21,8%, 108/496

[17] Fine-grained Benchmark Subsetting for System Selection. P. de Oliveira Castro, Y. Kashnikov, C. Akel, M. Popov, W. Jalby. IEEE/ACM International Symposium on Code Generation and Optimization (**CGO**), pages 132-142, 2014. Acceptance rate 28%, 29/103

[18] Evaluating Architecture and Compiler Design through Static Loop Analysis. Y. Kashnikov, P. de Oliveira Castro, E. Oseret, W. Jalby. IEEE High Performance Computing and Simulation (**HPCS**), pages 535-544, 2013. Acceptance rate 39,5%, 47/119

[19] ASK: Adaptive Sampling Kit for Performance Characterization. P. de Oliveira Castro, E. Petit, JC. Beyler, W. Jalby. International European Conference on Parallel and Distributed Computing, (**Europar**), 2012, pages 89-101. Acceptance rate 32,9%, 75/228

[20] A Multidimensional Array Slicing DSL for Stream Programming . P. de Oliveira Castro, S. Louise, D. Barthou. In Complex, Intelligent and Software Intensive Systems, International Conference, pages 913--918. IEEE Computer Society, 2010.

[21] Reducing memory requirements of stream programs by graph transformations. P. de Oliveira Castro, S. Louise, D. Barthou. In High Performance Computing and Simulation (HPCS), 2010 International Conference on, pages 171-180. IEEE Computer Society, 2010.

International peer-reviewed workshops

[22] Comparing Perturbation Models for Evaluating Stability of Post-Processing Pipelines in Neuroimaging. G. Kiar, P. de Oliveira Castro, P. Rioux, E. Petit, S. T. Brown, A. C. Evans, T. Glatard. Computational Reproducibility at Exascale 2019 (CRE2019), *in conjunction with Super Computing 2019.*

[23] Scalable Work-Stealing Load-Balancer for HPC Distributed Memory Systems. C. Fontenaille, E. Petit, P. de Oliveira Castro, S. Uemura, D. Sohier, P. Lesnicki, G. Lartigue, V. Moureau. In COLOC: 2nd Workshop on Data Locality, in conjunction with Euro-Par 2018.

[24] Computing-Kernels Performance Prediction Using DataFlow Analysis and Microbenchmarking. E. Petit, P. de Oliveira Castro, T. Menour, B. Krammer, and W. Jalby. In International Workshop on Compilers for Parallel Computers, 2012.

[25] Automatic mapping of stream programs on multicore architectures. P. de Oliveira Castro, S. Louise, and D. Barthou. In International Workshop on Compilers for Parallel Computers, 2010.