

2025 Activity Report

RESEARCH CENTRE: Inria Paris Centre at Sorbonne University

IN PARTNERSHIP WITH: Sorbonne Université

Project-Team

MEGAVOLT

MachinE learninG And eVOLution equaTions

In collaboration with Laboratoire Jacques-Louis Lions (LJLL), Laboratoire de Probabilités,
Statistique et Modélisation



Project-Team MEGAVOLT

Creation of the Project-Team: 2025 June 01

Each year, Inria research teams publish an Activity Report presenting their work and results over the reporting period. These reports follow a common structure, with some optional sections depending on the specific team. They typically begin by outlining the overall objectives and research programme, including the main research themes, goals, and methodological approaches. They also describe the application domains targeted by the team, highlighting the scientific or societal contexts in which their work is situated. The reports then present the highlights of the year, covering major scientific achievements, software developments, or teaching contributions. When relevant, they include sections on software, platforms, and open data, detailing the tools developed and how they are shared. A substantial part is dedicated to new results, where scientific contributions are described in detail, often with subsections specifying participants and associated keywords. Finally, the Activity Report addresses funding, contracts, partnerships, and collaborations at various levels, from industrial agreements to international cooperations. It also covers dissemination and teaching activities, such as participation in scientific events, outreach, and supervision. The document concludes with a presentation of scientific production, including major publications and those produced during the year.

Keywords

Computer sciences and digital sciences

- A6.1. – Methods in mathematical modeling
- A6.3. – Computation-data interaction
- A9.2. – Machine learning
- A9.7. – AI algorithmics

Other research topics and application domains

- B9.8. – Reproducibility

Contents

Project-Team MEGAVOLT	1
1 Team members, visitors, external collaborators	4
2 Overall objectives	4
3 Research program	4
4 Application domains	6
5 Social and environmental responsibility	6
6 Highlights of the year	6
6.1 Awards	6
7 Latest software developments, platforms, open data	6
8 New results	6
9 Partnerships and cooperations	7
9.1 International initiatives	7
9.1.1 Visits of international scientists	7
9.2 National initiatives	7
10 Dissemination	8
10.1 Invited talks	8
10.1.1 Leadership within the scientific community	9
10.2 Teaching - Supervision - Juries - Educational and pedagogical outreach	9
10.2.1 Supervision	9
10.2.2 Juries	10
10.2.3 Educational and pedagogical outreach	10
10.3 Popularization	10
11 Scientific production	10
11.1 Major publications	10
11.2 Publications of the year	10
11.3 Cited publications	11

1 Team members, visitors, external collaborators

Research Scientists

- Raphael Berthier [INRIA, Advanced Research Position]
- Borjan Geshkovski [INRIA, ISFP]

Faculty Members

- Bruno Despres [Team leader, SORBONNE UNIVERSITE, Professor Delegation, HDR]
- Gerard Biau [SORBONNE UNIVERSITE, Professor, HDR]

Post-Doctoral Fellows

- Ruiyang Dai [SORBONNE UNIVERSITE, Post-Doctoral Fellow, until Sep 2025]
- Moreno Pintore [SORBONNE UNIVERSITE, Post-Doctoral Fellow]

PhD Student

- Hugo Koubbi [DAUPHINE PSL, from Aug 2025]

Interns and Apprentices

- Thomas Giarrizzi [INRIA, Intern, from Sep 2025]

Administrative Assistants

- Derya Gok [INRIA]
- Anne Mathurin [INRIA]

2 Overall objectives

The high-level objective of the MEGAVOLT team is to bring together an expertise on evolution equations and their numerical analysis, with an expertise on machine learning (ML). Traditionally, these two communities have had limited interaction; however, some recent works demonstrate that there is a large untapped potential in crossing perspectives.

3 Research program

Our research program is currently structured in three major axes, with sub-axes describing the specific objectives, as follows:

- **Axis I. Neural network architectures as dynamics** The rise of deep learning, for example through innovations like skip connections in ResNets, has led to a view of neural networks as discretized differential equations, offering a clearer temporal interpretation of layers. Our horizons include:
 - A mathematical theory of Transformers. The introduction of Transformers in marked a turning point in the artificial intelligence (AI) revolution, powering breakthroughs in natural language modeling and computer vision. With remarkable practical success, Transformers like ChatGPT have revolutionized natural language and image processing. Now, as the size of these models grows at an astonishing rate, the need to understand their inner workings has never been more urgent. The temporal perspective on Transformers leads to an interpretation as interacting particle systems:

$\dot{x}_i(t) = \mathcal{X}[(x_1(t), \dots, x_n(t))](x_i(t))$. The particles $x_i(t) \in \mathbb{R}^d$ play the role of tokens (fragments of words in language modeling, or patches of an image in image processing), and the time variable t again plays the role of a layer. In this regard, the input of a Transformer is the sequence of tokens $(x_1(0), \dots, x_n(0))$, instead of a single d -dimensional vector as in conventional neural networks. In a nutshell, the key novelty of Transformers is the introduction of the self-attention mechanism $\mathcal{X}[(x_1, \dots, x_n)](x_i) = \sum_{j=1}^n \left(\frac{e^{\beta \langle Qx_i, Kx_j \rangle}}{\sum_{\ell=1}^n e^{\beta \langle Kx_i, Qx_\ell \rangle}} \right) Vx_j$, which depends on the empirical measure of all particles $\mu(t, \cdot) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i(t)}(\cdot)$, and entails permutation-equivariance properties of the flow.

- Continuous/discrete modelling of neural ODEs and SDEs. Residual neural networks (ResNets) are state-of-the-art deep learning models. Their continuous-depth analog, neural ordinary differential equations (neural ODEs) are widely used as an idealization for which theoretical guarantees can be shown more easily. However, a rigorous mathematical framework bridging these discrete and continuous realms remains elusive. Establishing a rigorous link between these two models is not merely a technical endeavor; it promises novel insights into the workings of ResNets. Should we establish that ResNets, once trained, act as discretized neural ODEs, it would enable the application of neural ODE findings to a broad spectrum of ResNets. Theoretically, the prowess of neural ODEs in approximation and the ease of deriving their generalization bounds are well documented. Practically, neural ODEs offer benefits like training with lower memory demands and the potential for weight reduction, addressing the critical issue of memory constraints in residual network training. This inquiry marks an initial step in deciphering the implicit regularization effects of gradient descent on deep ResNets.
- **Axis II. Training dynamics of neural networks**

Neural networks accumulate two major challenges: they are non-convex and often heavily over-parameterized, with little or no explicit regularization. However it has been observed that the dynamics of neural networks converge to prediction rules that generalize well. This research axis proposes several paths to apprehend this apparent contradiction.

- Incremental/dynamical learning and implicit regularization. Incremental learning qualifies training dynamics that can be decomposed into several phases, during which different components are learned. This phenomenon occurs frequently in ML, but theoretical understanding is lacking. In this section, we propose a theoretical path of simplified models to apprehend this phenomenology. Actually we show that incremental learning builds the estimator in a specific way, and thus selects a specific solution: it induces implicit regularization. We present an analysis strategy based on *heteroclinic dynamics*.
- Structure of neural networks and function approximation. The existing theoretical approaches for the non-convex dynamics of neural networks study restricted regimes where the dynamics simplify. The “neural tangent kernel” approach linearises the dynamics around its initialization but does not explain variable selection or, more generally, the excellent practical performances of neural networks. The mean field approach only applies in the limit of a large number of neurons. The research axis that we propose proposes to use a new regime, called the *two-timescale regime*, to study non-linear dynamics (i) with a moderate number of neurons and (ii) performing variable selection.

- **Axis III. Solving PDEs with ML**

- PINNs as a global approach. Despite notable advancements, modern ML models pose challenges in interpretation and may not adhere to the fundamental mathematical laws governing physical systems. Additionally, they often struggle to extend their predictions beyond the scenarios they were trained on. Conversely, numerical or purely physical methods encounter difficulties in capturing nonlinear relationships within complex and high-dimensional systems, while lacking adaptability and being susceptible to computational issues. This situation has prompted a growing consensus that data-driven ML

methods should be integrated with prior scientific knowledge rooted in physics. Our focus will be specifically on neural networks that incorporate a physical regularization, known as PINNs.

- VOFML as a local approach.

Another line of research concerns the extension of learning methods for local non-linear modules, which are set to boost the accuracy and stability of complex physics simulation codes. This theme concerns the interaction between ML and PDE resolution, which is a subject of very active international research. There is a very wide variety of points of view. The aim is to get closer to the numerical simulation of hyperbolic problems. The aim is to focus on well-targeted computational fluid dynamics (CFD) problems, in this case interface reconstruction for volume-finite numerical flow reconstruction. More generally, complex physics codes are tricky to handle, and they present many numerical or physical sub-mesh problems for which supervised learning can provide. This topic is becoming increasingly active at Sorbonne Université and some interactions are natural.

- Numerical stability of neural networks and time problems.

Stability is a central concept in time evolution PDE and numerical analysis, but one encounters fundamental difficulties when it comes to evaluating it for methods or functions created by ML. The aim of the research to be carried out is first to evaluate the Lipschitz constant of functions created in neural networks with the ReLU activation function (which has interesting properties), and then to incorporate this constant into the learning phase in order to enhance the stability of neural networks.

4 Application domains

Since the project is oriented mostly towards methodological questions, there is no specific applicative domain linked to our research.

However we aim at interacting with scientific community involved in ML for fluids with the objective of developing a basis of common research, and also with the engineering community.

5 Social and environmental responsibility

Machine learning and artificial intelligence may contribute positively to the environment for example by measuring climate change effect or reducing the carbon footprint of other sciences and activities. But it may also contribute negatively, notably by the ever-increasing sizes of machine learning models. Within the team, we work on these two aspects through our work on climate science and on frugal algorithms.

6 Highlights of the year

6.1 Awards

Gérard Biau has been elected at the French Academy of Sciences in 2025.

Borjan Geshkovski received a Google gift for his work on the mathematics of transformers. -

7 Latest software developments, platforms, open data

No software has been developed so far.

8 New results

Participants: Raphael Berthier, Bruno Després, Borjan Geshkovski.

Raphaël Berthier has worked on the theory of neural networks by drawing a new connection with sparse regression. His work focused on diagonal linear networks that are neural networks with linear activation and diagonal weight matrices. The theoretical interest of these neural networks is that their implicit regularization can be rigorously analyzed: from a small initialization, the training of diagonal linear networks converges to the linear predictor with minimal 1-norm among minimizers of the training loss. In the paper [2], RB deepened this analysis showing that the full training trajectory of diagonal linear networks is closely related to the lasso regularization path. In this connection, the training time plays the role of an inverse regularization parameter.

Bruno Després has shown that the Murat-Trombetti Theorem is a simple and efficient mathematical framework for nonsmooth automatic differentiation of maxpooling functions. In particular it gives a chain rule formula which correctly defines the composition of Lipschitz-continuous functions which are piecewise C^1 . The formalism is applied to four basic examples, with some tests in PyTorch. A self contained proof of an important Stampacchia formula is in the appendix of the article published at TMLR.

Borjan Geshkovski gives a precise optimization-theoretic and dynamical-systems interpretation of a Transformer self-attention layer’s forward pass in the “hardmax” (zero-temperature regime: in that limit, the token update can be rewritten as a Frank–Wolfe (conditional gradient) step for a quadratic objective over the convex hull of the current token embeddings (with the value matrix playing a preconditioning role), which lets him sharply characterize the geometry and long-time behavior of token motion depending on the sign of the (symmetric) key–query matrix—showing linear contraction to a single cluster at the origin in the negative semidefinite case, and (after extending the rule to the whole convex hull) a Voronoi-cell structure in the positive semidefinite case where vertices are stationary, points remain in their initial cells, and tokens move straight toward the cell’s vertex with (super-)exponential convergence; they additionally prove well-posedness of the associated singular ODE limit and then connect back to finite-temperature (softmax) attention by modeling it as a Markov chain and proving a dynamic metastability result: with high probability the finite-temperature dynamics rapidly reaches and then stays near the hardmax “near-vertex” configurations for times exponential in the inverse temperature so the hardmax analysis accurately predicts behavior over very long horizons before eventual collapse.

9 Partnerships and cooperations

9.1 International initiatives

9.1.1 Visits of international scientists

Other international visits to the team

Andrea Agazzi

Status Professor

Institution of origin: University of Bern

Country: Switzerland

Dates: 22/03 to 28/03

Context of the visit: Seminar and collaboration with Berthier and Geshkovski

Mobility program/type of mobility: Research stay, lecture

9.2 National initiatives

The activity of the project is supported by the PEPR-IA with a from Agence Nationale de la Recherche, program France 2030, reference ANR-23- PEIA-0004.

Participants: Bruno Despres, Moreno Pintore.

10 Dissemination

10.1 Invited talks

Gerard Biau

- *Foundations and Advances in Generative AI: Theory and Methods*, Paris, France (février 2025, **invité**).
- *Mathematics of Machine Learning, Physics Informed Machine Learning*, Abou Dhabi, Emirats arabes unis (février 2025, **invité**).
- *17th German Probability and Statistics Days (GPSD) 2025*, Dresde, Allemagne (mars 2025, **invité, conférencier plénier**).
- *Grand Séminaire MACS 2025*, Paris, France (octobre 2025, **invité**).

Raphael Berthier

- Dec. 2025 Institut Henri Poincaré, séminaire d'optimisation parisien (SPO)
- Oct. 2025 Université Paris 1 Panthéon Sorbonne, séminaire SAMM (statistiques, analyse et modélisation multidisciplinaire)
- Oct. 2025 ENSAE, séminaire de statistiques
- Juin 2025 University of Bern, Institute for Mathematical Statistics and Actuarial Sciences

Borjan Geshkovski

- *Physics of AI Algorithms, Les Houches*, Les Houches, France (janvier 2025).
- *Applied Mathematics Colloquium, RWTH Aachen*, Aachen, Allemagne (janvier 2025).
- *Séminaire Parisien d'Optimisation, Institut Henri Poincaré*, Paris, France (février 2025).
- *OT PDE ML Seminar, Laboratoire de mathématiques d'Orsay*, Orsay, France (mars 2025).
- *Erlangen workshop ML PDE*, Erlangen, Allemagne (avril 2025).
- *Mathematic park, Institut Henri Poincaré*, Paris, France (avril 2025).
- *NYU Paris workshop*, Paris, France (juin 2025).
- *Despres 60*, lieu à préciser (juin 2025).
- *EPFL Optimization Unplugged*, Lausanne, Suisse (août 2025).
- *Hamburg workshop on Transformers*, Hambourg, Allemagne (septembre 2025).
- *Round meanfield Venice*, Venise, Italie (octobre 2025).
- *Séminaire francilien de géométrie algorithmique et combinatoire, Institut Henri Poincaré*, Paris, France (octobre 2025).
- *Barcelona workshop on Mathematical Foundations on ML*, Barcelone, Espagne (janvier 2026).

Bruno Despres

- 6-10 janvier 2025: cours à l'Ecole CEFIPRA <https://www.iitr.ac.in/indofrench/index.html>
- 13 mars: exposé aux journées PEPR-IA (organisateur Antonin Chambolle)
- 24-26 mars: comité scientifique des premières journées de l'EMS-Tag <https://ems-tag-sciml.github.io> sur le SciML à Milan <https://www.mate.polimi.it/events/EMS-TAG-SciML-25/>

- 9-13 juin: cours "Neural Networks from the viewpoint of Numerical Analysis" à l'ENS Rennes <https://indico.math.cnrs.fr/event/13569/>
- fin juillet: Workshop Wien <https://www-thphys.physics.ox.ac.uk/research/plasma/wpi/workshop2025.html>
16th Plasma Kinetics Working Meeting, session "What can AI do for plasma physics and what can plasma physics do for AI?".
- 8-12 septembre: "NUMERICAL METHODS FOR HIGH-DIMENSIONAL DATA" La Sapienza Italie, <https://sites.google.com/uniroma1.it/numerics-high-dimensional-data/mini-courses>
- 14 octobre: séminaire invité à la TU Eindhoven <https://casa.win.tue.nl/home/event/colloquium-bruno-despres-sorbonne-university/>, redonné à Numkin 2025 "Thoughts on Mathematics, Plasma Physics and Machine Learning" <https://www.ipp.mpg.de/5518866/program-numkin-2025>
- 19-25 novembre Erice Italie: The Mathematics of Scientific Machine Learning and Digital Twins <https://mod.fau.eu/scimldt/>

Moreno Pintore

- 8th ECCOMAS Young Investigators Conference YIC 2025. Pescara, Italy. Contributed talk, Minisymposium organizer. September 17-19, 2025
- Summer school: "Numerical methods for high-dimensional data". Rome, Italy. Lesson. September 15-19, 2025
- SIMAI Conference 2025. Trieste, Italy. Invited speaker. September 1-5, 2025
- Eccomas Math2Product. Valencia, Spain. Contributed talk. June 4-6, 2025
- Joint event Euromech Colloquium on Data-Driven Fluid Dynamics/2nd ERCOFTAC Workshop on Machine Learning for Fluid Dynamics. London, UK. Contributed talk. April 2-4, 2025
- PEPR IA Days. Saclay, France. Contributed talk. March 18-20, 2025
- DTE AICOMAS Congress 2025. Paris, France. Contributed talk. February 17-21, 2025
- 3rd Workshop of UMI Group - Mathematics for Artificial Intelligence and Machine Learning. Bari, Italy. Contributed talk. January 29-31, 2025
- Indo-French Workshop on Innovative - Numerical Methods for Modern Engineering Problems. Roorkee, India. Lesson. January 6-10, 2025

10.1.1 Leadership within the scientific community

Raphael Berthier and Borjan Geshkovski (along with Francis Bach, Bruno Despres and Gerard Biau) jointly organize the monthly seminar on Analysis, Algorithms and Learning at Sorbonne Université <https://www.ljll.fr/gdt-analyse-algorithmique-apprentissage/>.

Borjan Geshkovski is part of the organizing committee of the Automath seminar at ENS Paris .

10.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

10.2.1 Supervision

Hugo Koubbi began his PhD thesis in September 2025 under the supervision of Borjan Geshkovski and Antonin Chambolle.

Thomas Giarizzi (ENS Paris) began a Master 1 internship in September 2025 under the supervision of Borjan Geshkovski.

Fabien Richard (CEA/LJLL) starts his PHD in October 2025 with Bruno Despres.

The PhD thesis of Nicola Galante (Inria-Alpines) is co-supervised with Bruno Despres and Emile Parolin (Inria-Alpines).

10.2.2 Juries

Raphael Berthier was an examiner in the PhD defense of Maksim Velikanov (Ecole Polytechnique).

Borjan Geshkovski was an examiner in the PhD defense of Raphaël Barboni (ENS Paris).

Bruno Despres was member of the PHD jury of Davide Oberto in MArch 2025, au Politecnico di Torino.

10.2.3 Educational and pedagogical outreach

Raphael Berthier teaches a course on “optimization for machine learning” in the **M2 Apprentissage et Algorithmes** of Sorbonne Université. The lecture notes are [available on github](#). RB also taught a course on “inferential statistics” in the “**mineure IA et sciences des données**” of Sorbonne Université.

Borjan Geshkovski teaches a course on particle systems and machine learning in M2 Mathématiques de la Modélisation at Sorbonne Université.

Bruno Despres teaches "Neural Networks and Numerical Analysis" in **M2 "Mathématiques de la Modélisation"** and **M2 "Apprentissage et Algorithmes"** at Sorbonne Université. He is also head of the track "Sciences des données et EDP (SDEDP)" of the M2 "Mathématiques de la Modélisation". The notes are published in [17]

10.3 Popularization

Borjan Geshkovski gave a talk at the Mathematic Park seminar at Institut Henri Poincaré .

11 Scientific production

11.1 Major publications

- [1] A. Álvarez-López, B. Geshkovski and D. Ruiz-Balet. *Constructive approximate transport maps with normalizing flows*. 23rd Jan. 2025. URL: <https://hal.science/hal-04909497>.
- [2] R. Berthier. *Diagonal Linear Networks and the Lasso Regularization Path*. 22nd Sept. 2025. URL: <https://hal.science/hal-05282524> (cit. on p. 7).
- [3] B. Geshkovski, P. Rigollet and D. Ruiz-Balet. *Measure-to-measure interpolation using Transformers*. 8th Nov. 2024. URL: <https://hal.science/hal-04784260>.
- [4] B. Geshkovski, P. Rigollet and Y. Sun. *On the number of modes of Gaussian kernel density estimators*. 13th Dec. 2024. URL: <https://hal.science/hal-04842172>.
- [5] P. Marion, R. Berthier, G. Biau and C. Boyer. ‘Attention layers provably solve single-location regression’. In: *Proceedings of the Thirteenth International Conference on Learning Representations*. ICLR 2025 - Thirteenth International Conference on Learning Representations. Singapore, Singapore, 26th Feb. 2025. URL: <https://hal.science/hal-04720799>.
- [6] M. Pintore and B. Després. *A 3D Machine Learning based Volume Of Fluid scheme without explicit interface reconstruction*. 7th July 2025. URL: <https://hal.sorbonne-universite.fr/hal-05149322>.

11.2 Publications of the year

International journals

- [7] N. Doumèche, G. Biau and C. Boyer. ‘On the convergence of PINNs’. In: *Bernoulli* Vol. 31 (2025), pp. 2127–2151. URL: <https://hal.science/hal-04085519>.

Reports & preprints

- [8] A. Alcalde, B. Geshkovski and D. Ruiz-Balet. *Attention’s forward pass and Frank-Wolfe*. 2025. DOI: [10.48550/arXiv.2508.09628](https://arxiv.org/abs/10.48550/arXiv.2508.09628). URL: <https://hal.science/hal-05212729>.

- [9] A. Álvarez-López, B. Geshkovski and D. Ruiz-Balet. *Constructive approximate transport maps with normalizing flows*. 23rd Jan. 2025. URL: <https://hal.science/hal-04909497>.
- [10] S. Berrone, L. Neva, M. Pintore, G. Teora and F. Vicini. *The Zipped Finite Element Method: High-order Shape Functions for Polygons*. 2025. URL: <https://hal.science/hal-05391322>.
- [11] S. Berrone, M. Pintore and G. Teora. *The Neural Approximated Virtual Element Method for Elasticity Problems*. 8th July 2025. URL: <https://hal.sorbonne-universite.fr/hal-05150590>.
- [12] R. Berthier. *Diagonal Linear Networks and the Lasso Regularization Path*. 22nd Sept. 2025. URL: <https://hal.science/hal-05282524>.
- [13] G. Biau and C. Boyer. *A Note on k-NN Gating in RAG*. 19th Jan. 2026. URL: <https://hal.science/hal-05462307>.
- [14] B. Després. *A functional framework for nonsmooth autodiff with maxpooling functions*. 28th Apr. 2025. URL: <https://hal.sorbonne-universite.fr/hal-04846612>.
- [15] N. Galante, B. Després and E. Parolin. *A Trefftz Continuous Galerkin method for Helmholtz problems*. 1st Dec. 2025. URL: <https://hal.science/hal-05394682>.
- [16] M. Pintore and B. Després. *A 3D Machine Learning based Volume Of Fluid scheme without explicit interface reconstruction*. 7th July 2025. URL: <https://hal.sorbonne-universite.fr/hal-05149322>.

11.3 Cited publications

- [17] B. Després. *Neural Networks and Numerical Analysis*. Berlin, Boston: De Gruyter, 2022. DOI: [doi: 10.1515/9783110783186](https://doi.org/10.1515/9783110783186). URL: <https://doi.org/10.1515/9783110783186> (cit. on p. 10).