RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030
PROGRAMME DE RECHERCHE
NUMÉRIQUE POUR L'EXASCALE

ExaDoST - Work Package 3
Exascale ML-based Analytics

WP Leaders:

Thomas Moreau (Inria Saclay)

Bruno Raffin (Inria Grenoble)

WP3 – Exascale ML-based Analytics - Objectives

Process the data **automatically** as they are produced

- Identify the algorithms which are relevant for such use cases

  ⇒ Data-integration / Anomaly detection / Unsupervised learning

- Identify/benchmark core ML building components to use these algorithms

  ⇒ Distributed / non i.i.d. learning

- Develop software bricks required to unlock / scale these use-cases

  ⇒ distributed learning paradigms and ensemble runs

  ⇒ in-situ workflows and benchmarks

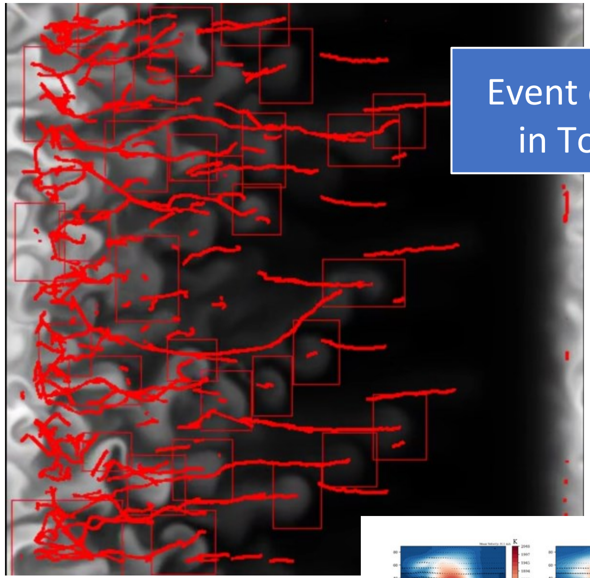
  ⇒ Distributed computing stack in python (e.g. for scikit-learn)

WP3: Participants

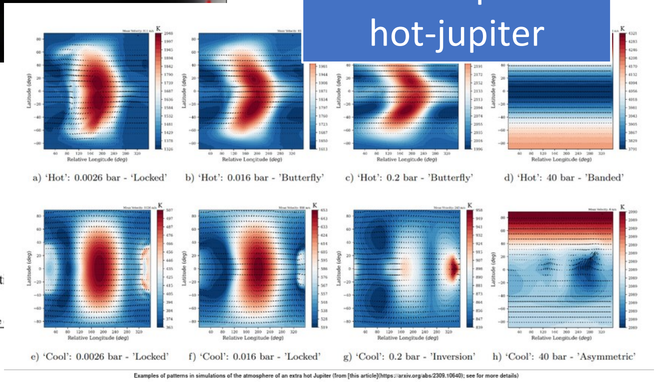
Partner	Type of position	Name of participant
Inria Saclay	Researchers	**Thomas Moreau**
	Postdocs	Mansour Benbakoura
	Engineers	Soon hired (Yoann Coudert-Osmon)
	PhD students	Jad Yehya, Hippolyte Verninas (starting 2026)
Inria Grenoble	Researchers	**Bruno Raffin**, Hadrien Hendrikx, Pedro Rodrigues
	Engineers	Abhishek Punrandare, Pierre Cesare
	PhD students	Sofia Dymchenko
CEA - Maison de la simulation	Researchers	Martial Mancip

WP3: Achievements

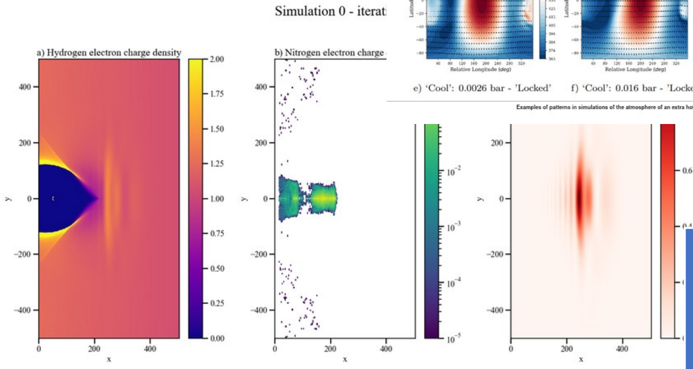
- Event detection (M. Benbakoura, T. Moreau, V. Grandgirard)
  - Literature review on (rare) event detection in computer vision and how to adapt to large physical simulations
  - Prototype with tokam2D
  - Poster in the CEA's event detection day
- Distributed optimization (H. Hendrickx, T. Moreau)
  - Realized a first benchmark on distributed PCA for large physical signals (to be open soon).
- Data Challenge (T. Moreau, M. Mancip, M. Lobet):
  - Organized three data challenges pattern classification from large physical simulations



Event detection in Tokam2D



Classification d'atmosphère hot-jupiter



Détection de bulles dans des plasmas

WP3: Achievements

- Melissa (A. Purandare, S. Dymchenko, P. Cesare, B. Raffin):
  - Adios2 as transport layer (hard to get a performance boost vs ZMQ)
  - Adaptive input parameter sampling (1 paper @ AI4Science, SC24 workshop)
  - Invited talk at WANT workshop, ICML 24
- SBI (P. Rodrigues, B. Raffin, T. Moreau)
  - First Steps into active learning for SBI (Camille Touron, M2 internship)
  - Discussions with EDF on generative flows and SBI
  - Participate in sbi package (TU Tubingen) and a tutorial paper
- AI4Science (P. Rodrigues, T. Moreau, J. Le Sommer, B. Raffin)
  - GAP workshop @ Grenoble: https://gap2024.sciencesconf.org/
  - Chair Proposal @ Grenoble AI Cluster (MIAI)
  - Sacl-AI4Science workshop @ Saclay

Tracking patterns in 2D plasma turbulence simulations

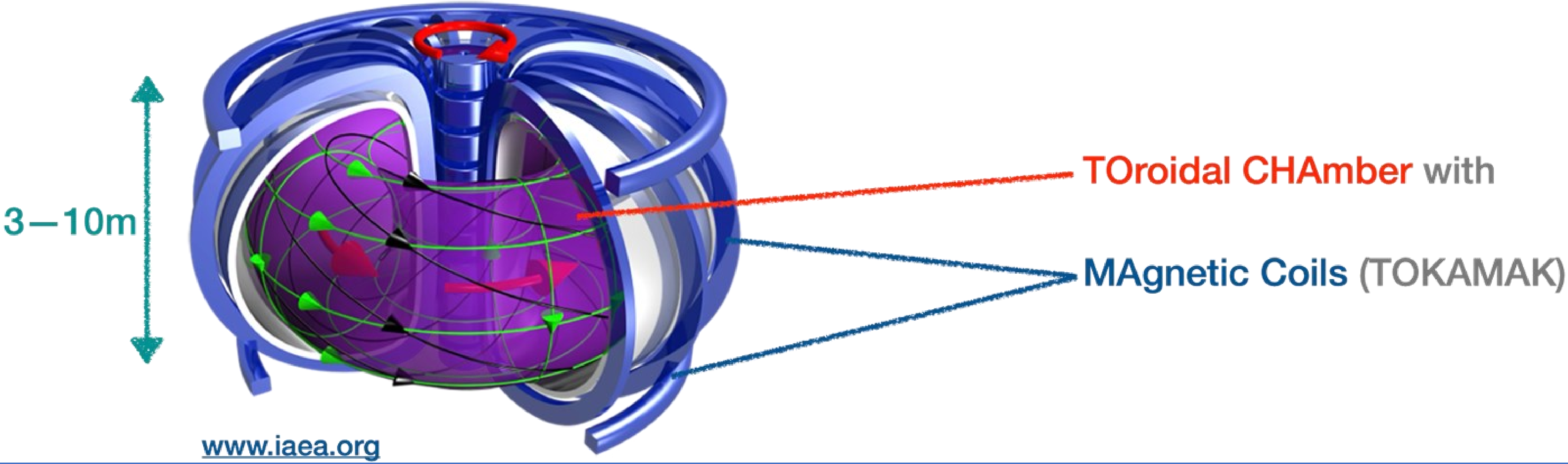
M. Benbakoura, H.Taher, R. Varennes,
G. Dif-Pradalier, V. Grandgirard, M. Lobet, M. Mancip, T. Moreau, E. Serre, D. Zarzoso

Paper in preparation.

Magnetic confinement fusion

Goal:

1. Confine a plasma (≃ gas of charged particles) with a magnetic field,
2. Heat until fusion reactions occur.



www.iaea.org

Turbulent transport hinders fusion

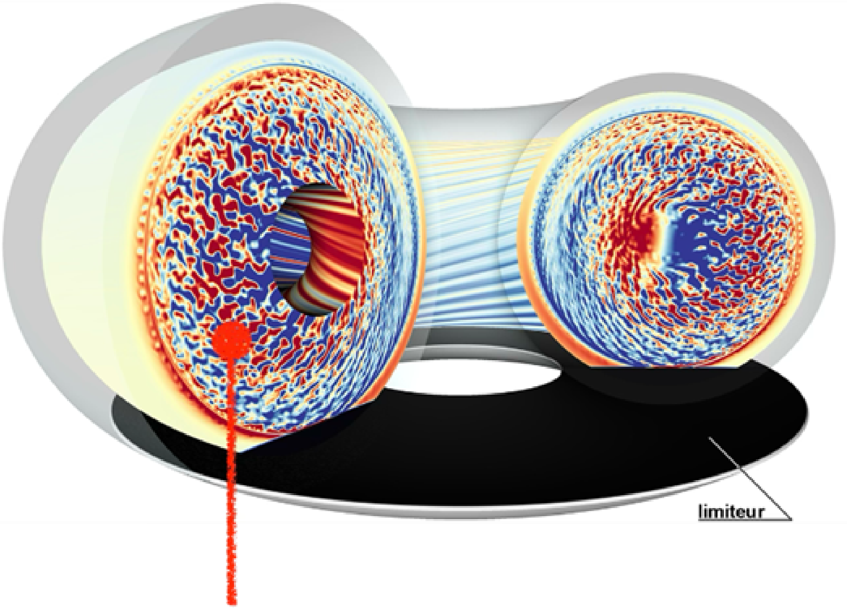
Movie from MAST tokamak
(100,000 frames per second)

- Fusion reaction => $T > 10^8$ K ($\simeq 6 \times T_{solar\_core}$),
- => Strong **plasma turbulence**,
- => Turbulent heat and particle transport,
- ↳ Hinders fusion.

Blob / filaments:
Carry most of the particles and heat

Realistic simulations are expensive

- **Ab-initio turbulence simulation => (gyro)kinetic theory,**
- **6 dimensions:**
  - 3 in space,
  - 2 in velocity space,
  - 1 in time,
- **=> Demanding in computing power + storage.**



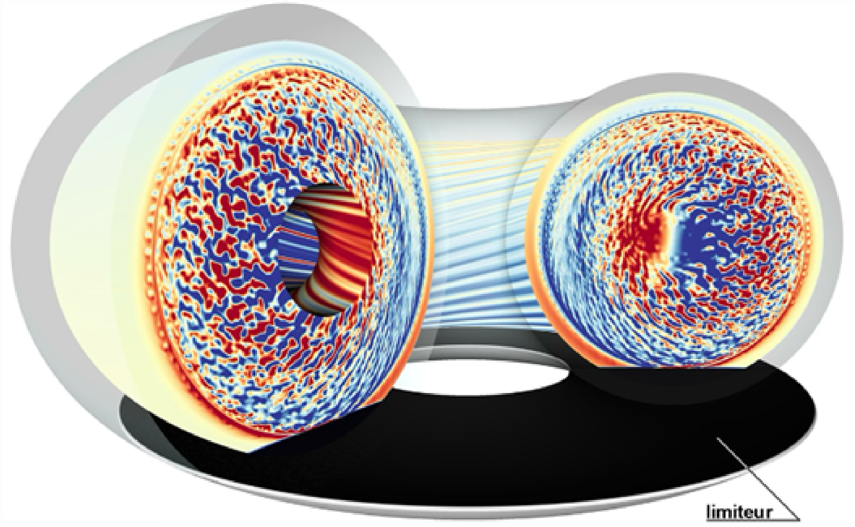
One velocity distribution per (x, y, z)

Event detection for distributed simulations

Two use cases:

- Detect rare / meaningful events
  ↳ Trigger diagnostics;

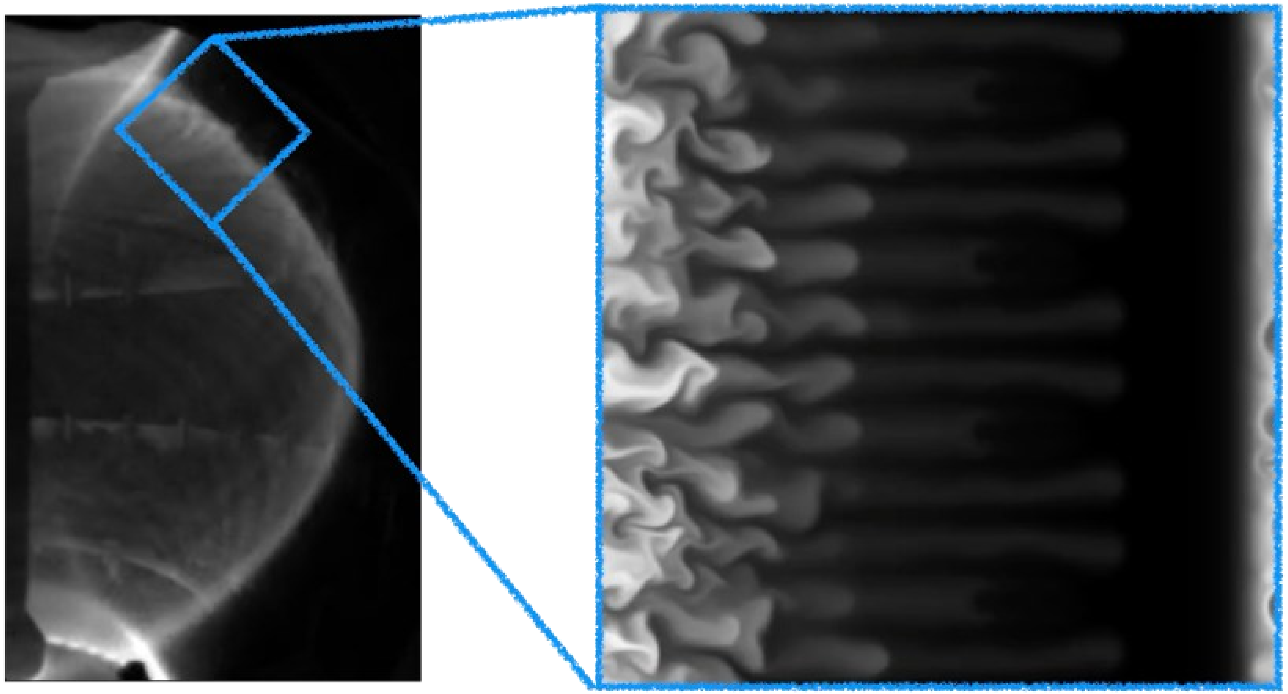
- Detect anomalies
  ↳ Detect failure on one node.



Towards large-scale simulations

Before going to 5D+time:

- Smaller simulation domain,
- Lower dimensionality (fluid 2D, 3D),
- => The **TOKAM2D** code.

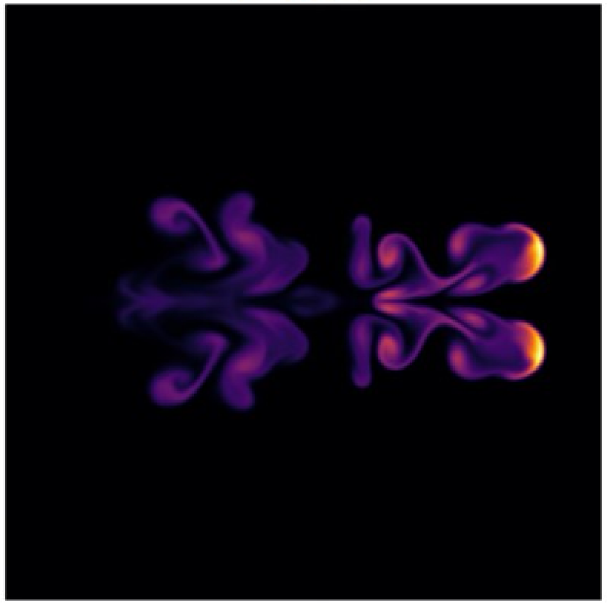


Detect and track turbulent structures?

Problem statement

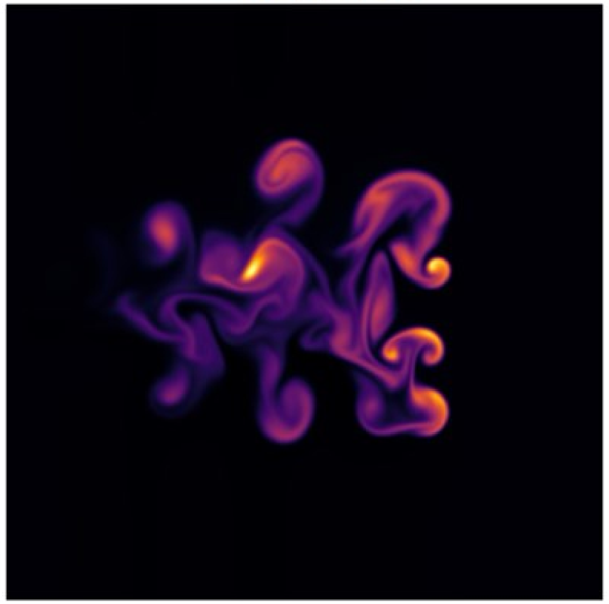
How to ensure model generalization?

blob_i



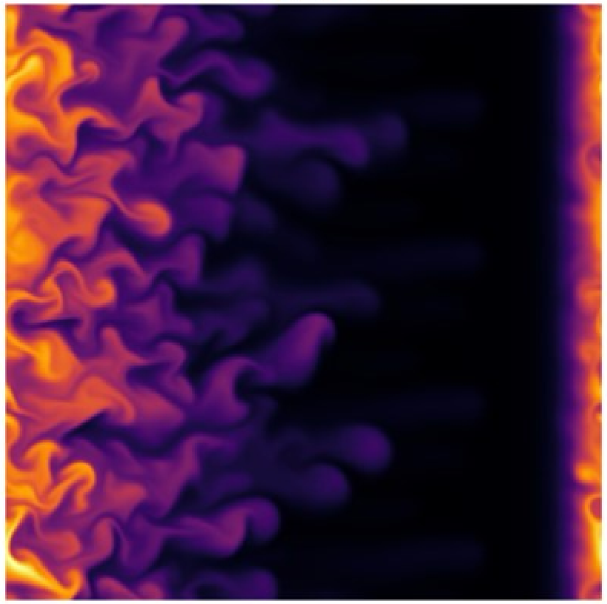
- 2000 frames,
- 20 annotated,
- 176 boxes.

blob_dwi



- 2000 frames,
- 10 annotated,
- 106 boxes.

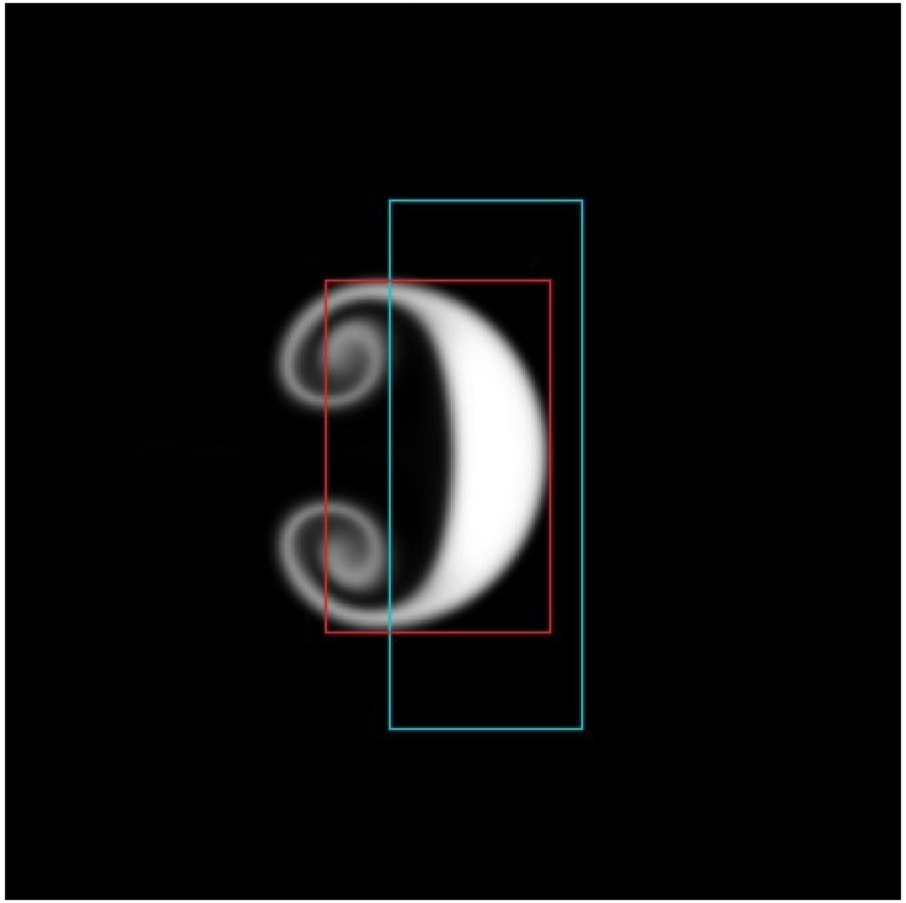
turb_dwi



- 2000 frames,
- 8 annotated,
- 373 boxes.

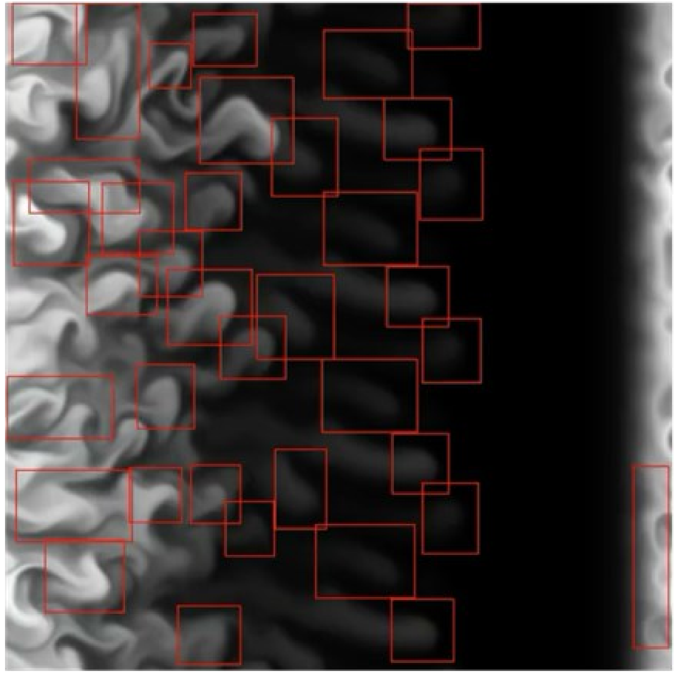
Mind the metrics!

IoU=0.45; IoMean=0.62

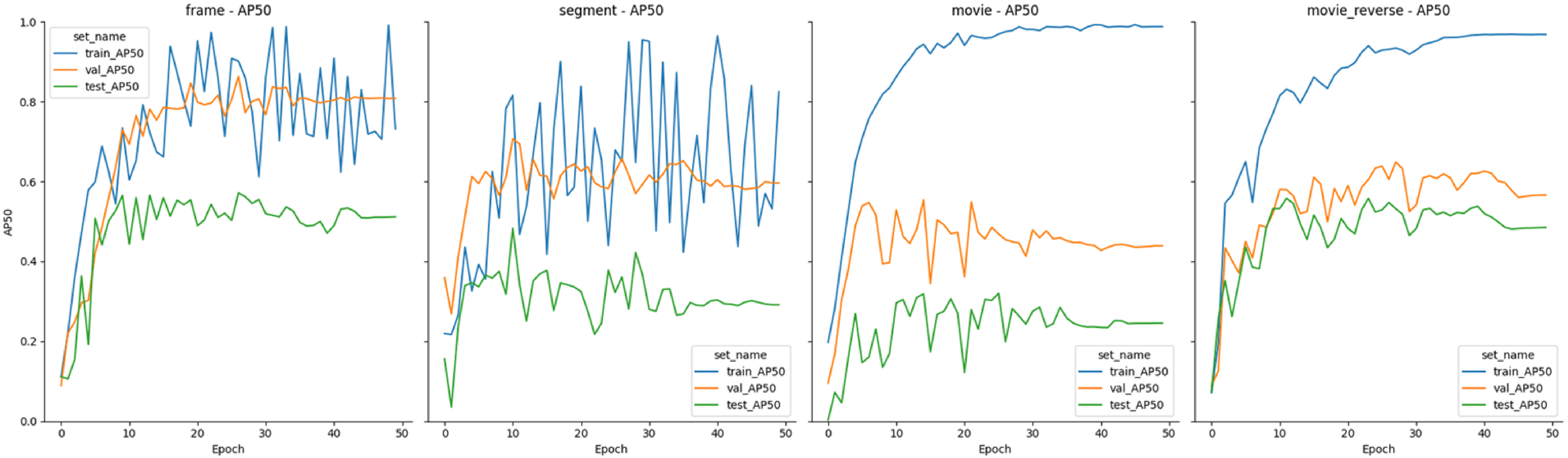


Results / prospects

- Train with `blob_i`,
- Validate on `blob_dwi`,
- Test on `turb_dwi`,
- Track with SORT (Bewley+ 16),
- Satisfactory visual evaluation.
  ↳ **TO BE QUANTIFIED.**



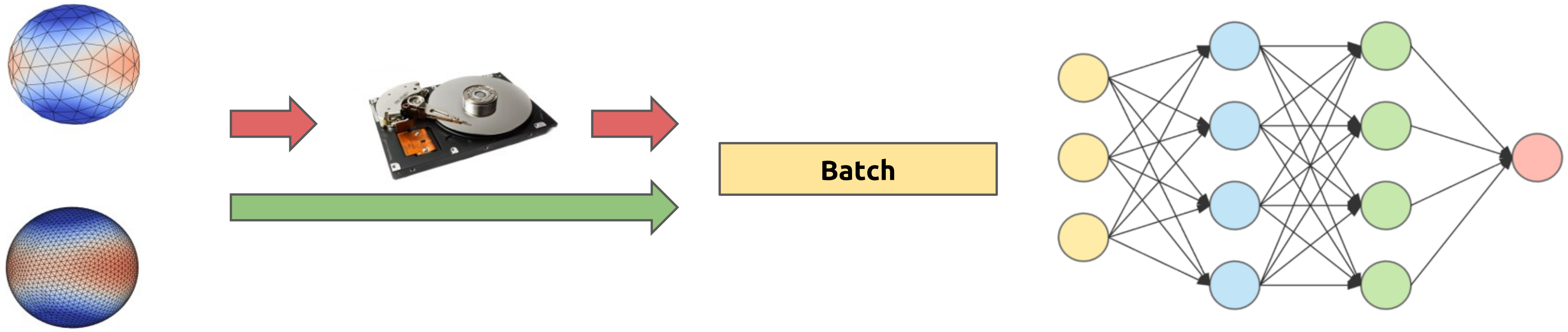
Results / prospects



Online Training With MELISSA

Abhishek Purandare
Research Engineer, Datamove, INRIA

Need for Online Training



Simulations as Data Streams: Solvers evolve PDE states, $\mathbf{x}_{t+1} = \Phi_{\Delta t}(\mathbf{x}_t; \boldsymbol{\theta}_{\text{phys}})$ generating trajectories $\tau_i = \{\mathbf{x}_{i,t}\}_{t=0}^{T}$

Learning Surrogates Online: Deep models approximate solver dynamics and update continuously as new simulation data arrive.

- *Direct:* $\hat{\mathbf{x}}_t = f_\theta(\mathbf{x}_0, t)$
- *Autoregressive:* $\hat{\mathbf{x}}_{t+1} = f_\theta(\mathbf{x}_t)$

Online Training with Melissa

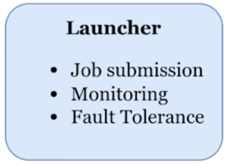
In-transit architecture with dual-server modes

- **NxM** for Sensitivity Analysis
- **Round-robin** for Deep Learning

Elastic, fault-tolerant execution

Three-component system:

- Instrumented **clients/solvers**
- **Server** that trains NN on-the-fly
- **Launcher** that orchestrates tasks



Reservoir removes online training biases:

- **Intra**-simulation
- **Inter**-simulation

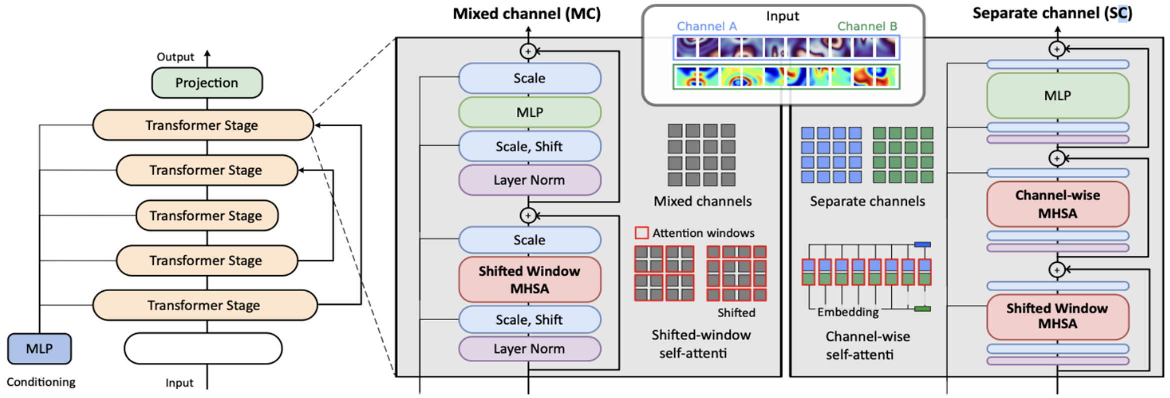
https://gitlab.inria.fr/melissa/melissa

Foundation Models for PDEs

Pretrained Physics Representations: Large neural operators trained across many PDE systems learn a *shared latent space of physical dynamics.*

Operator-Level Generalization: Instead of solving a single equation, they approximate the *mapping between function spaces* enabling generalization across PDE types, domains, and resolutions.

Reusable Scientific Backbones: Serve as foundational priors that can be *adapted, fine-tuned, or composed* for new physics regimes, drastically reducing the need for costly solver-generated data.

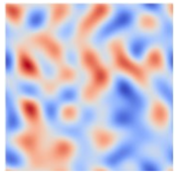


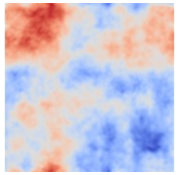
PDE-Transformer: *Efficient and Versatile Transformers for Physics Simulations* arXiv:2505.24717v1

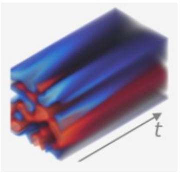
Pseudo-Offline Vs Online

Online training across 4 GPUs with different PDEs:

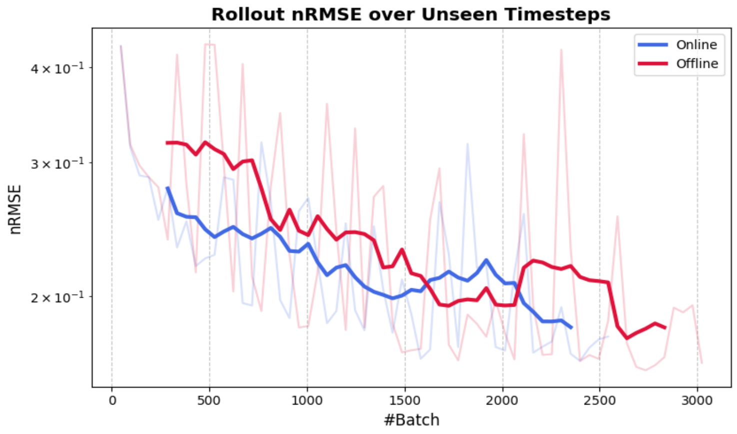
- PDE-Transformer (small) with ~33M parameters
- 2D mesh 256x256 with T=30
- Two ICs:
  - *Truncated Fourier Series*
  - *Gaussian Random Field*
- Two nonlinear PDEs:
  - *Burgers*
  - *Korteweg-De Vries*

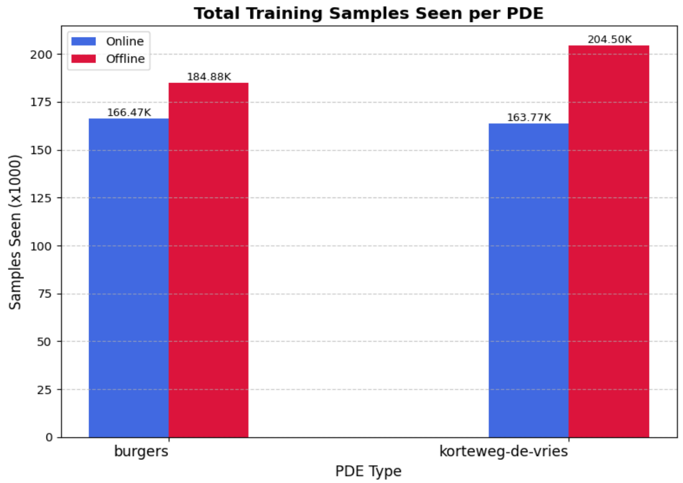






- Training around ~1.25 hours
  - *Offline runs 12 epochs across 1K Simulations*
  - *Online runs 10K Simulations*





APEBench: *A Benchmark for Autoregressive Neural Emulators of PDEs* arXiv:2411.00180v1

All2All Training with Melissa

Efficiently stores and samples pairs of states $(\mathbf{x}_{i,t_1}, \mathbf{x}_{i,t_2})$ where $t_1 < t_2$

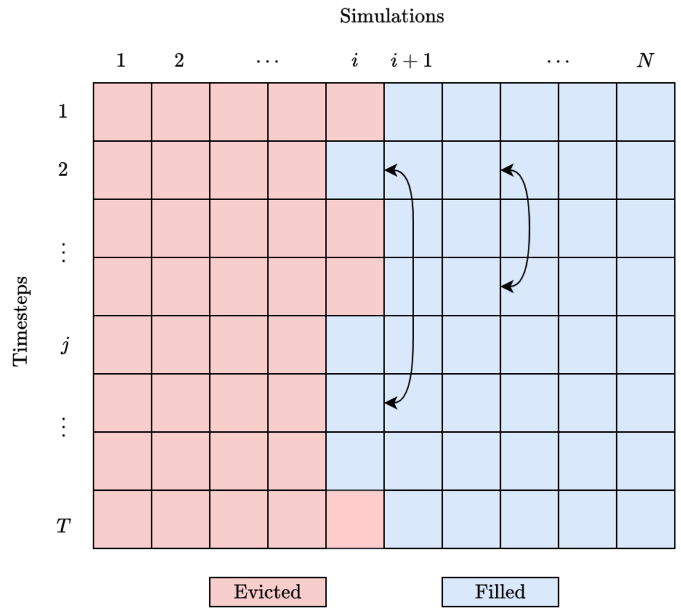
- **temporal coherence**
- **statistical diversity**

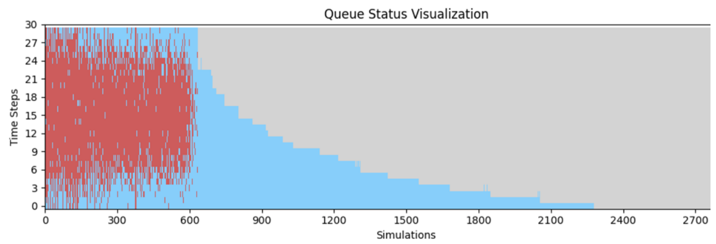
Uses a two-level selection scheme:

- select a trajectory **based on usage**
- sample a pair by giving **higher probability to rarely** used time separations Δt

Applies an eviction strategy when full:

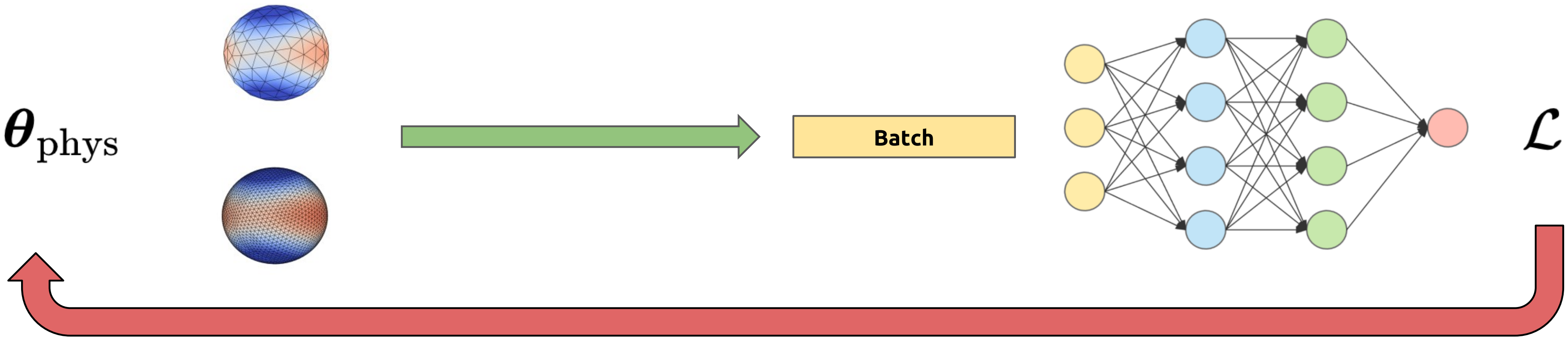
- trajectories used more often
- time-steps near the middle of the trajectory are evicted, preserving boundary information for long-range Δt sampling





Poseidon: *Efficient Foundation Models for PDEs* arXiv:2405.19101v2

Active Learning



Global-steering