AI for Physical Dynamics: From Neural Surrogates to Foundation Models

2026-03-09

Patrick Gallinari

ISIR: Institute of Intelligent Systems and Robotics

AI4Math&Engineering@Sorbonne University

Numerical Analysis

Bruno Després Julien Salomon

Statistics

Gérard Biau

Machine Learning

Patrick Gallinari

Fluid Mechanics

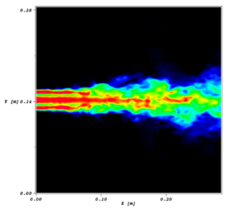
Paola Cinnella

Taraneh Sayadi

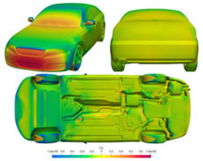
AI4Science@ISIR
Focus: Modeling Spatio-temporal dynamics with NNs

- Domains - examples

Computational Fluid Dynamics

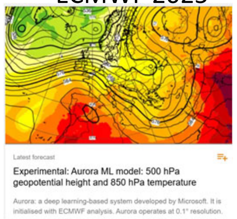


DrivaerNet 2025



Earth System Science – Weather - Climate

ECMWF 2025



Graphical design

Tompson et al. 2017

Machine learning

- Focus on methodological aspects
- Develop new ML models for problems/ challenges motivated by applications in physics

Cross-disciplinary collaborations

- Climate (LOCEAN)
  Guillaume Gastineau, Marina Levy, Sylvie Thiria
- Numerical analysis (LJLL)
  Julien Salomon
- Fluid mechanics (d'Alembert)
  Paola Cinnella, Taraneh Sayadi
- Computational cardiology (Inria Sophia)
  Maxime Sermesant

Explicit vs Implicit modeling of physical dynamics: Partial differential equations & machine learning

Spatio temporal dynamics are usually modeled through partial differential equations
Neural surrogates are **trained** on data that follow PDEs as underlying dynamics

Explicit modeling

$$\frac{\partial u}{\partial t} = g(c, f; x, t, u, \nabla u, \nabla^2 u, \ldots) \qquad \forall (x,t) \in \Omega \times \mathbb{R}_*^+$$

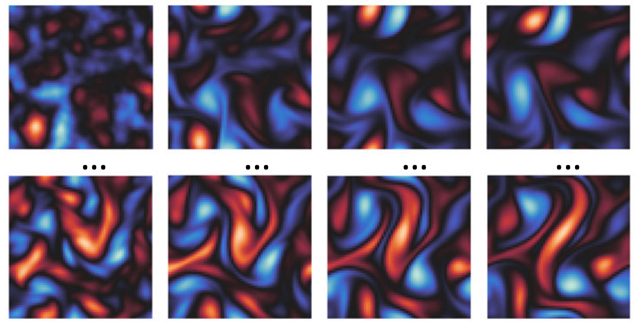
Partial Differential Equation

Initial constraint → $u(x, t = 0) = u^0(x) \qquad \forall x \in \Omega$

Boundary constraint → $B(b; u, \nabla u, x, t) = 0 \qquad \forall (x,t) \in \partial\Omega \times \mathbb{R}_*^+$

Implicit modeling

Training dataset of N trajectories



u^0 $u^{\Delta t}$ $u^{2\Delta t}$ $u^{3\Delta t}$

↓

Surrogate Model

Data-driven modeling of physical dynamics
Scientific challenges

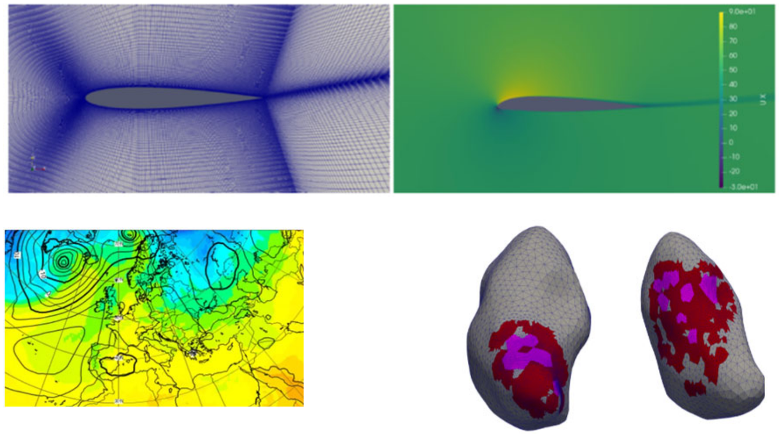
- Physical Consistency:
  - Incorporating constraints such as conservation laws, symmetries, and invariants into learning models.
- Generalization and Extrapolation:
  - Ensuring robust predictions beyond training regimes (unseen physical regimes, new domains, new sensors, etc).
- Data Scarcity & Heterogeneity:
  - Leveraging limited, noisy, or multi-fidelity data typical of scientific domains.
- Integration with Simulation:
  - Building hybrid frameworks that combine data-driven and physics-based modeling.
- Computational Efficiency:
  - Training large-scale scientific foundation models while managing cost and sustainability.
- Uncertainty Quantification
- …

AI4Science@ISIR - illustrations

- Neural operators surrogate models
- Generalization to unseen contexts
- Scaling
- Foundation models

Neural operators

- Classical numerical solvers operate on grids or meshes (finite differences, finite elements, finite volumes)



- Early neural solvers operate on tensors (grids) or on graphs (irregular meshes)

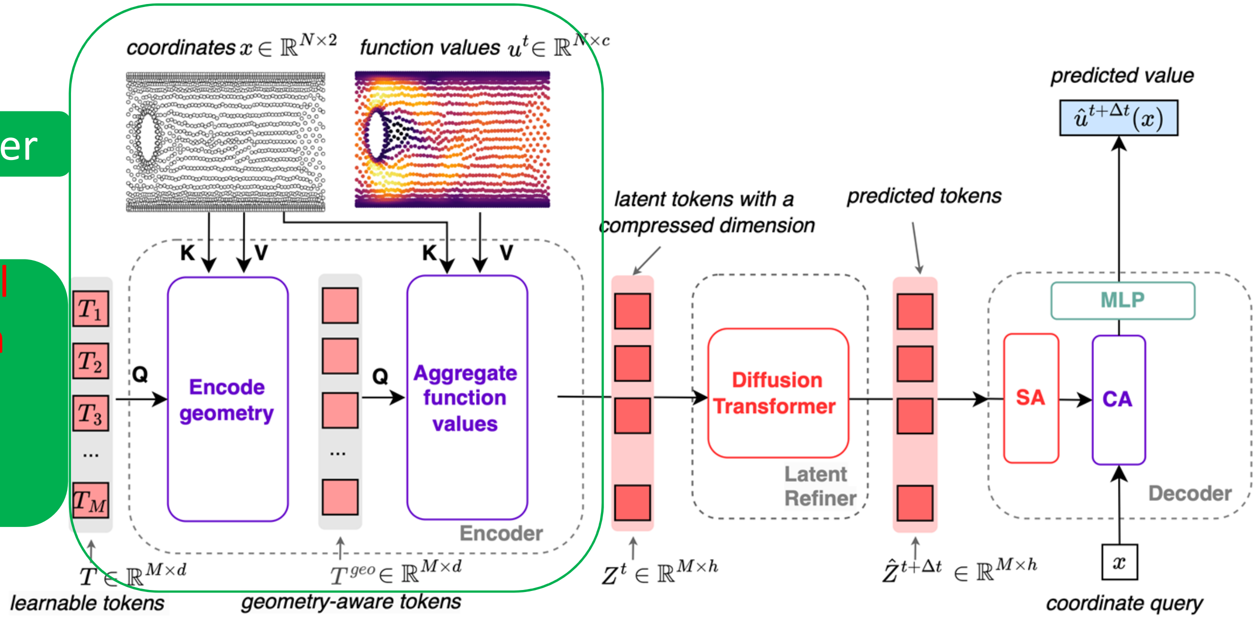
- **Neural operators** is a relatively recent topic aiming at learning maps between function spaces instead of vector spaces
  - e.g. images are considered as continuous functions
- Potential benefits
  - Functions and operators are mesh/ resolution invariant
  - Handle different geometries, multiple resolutions
  - Query at any space-time coordinate

Meshless Neural operators

AROMA: Attentive Reduced Order Model with Attention (Serrano et al. 2024)

Meshless Encoder

Encodes spatial information on problem geometry in a latent space



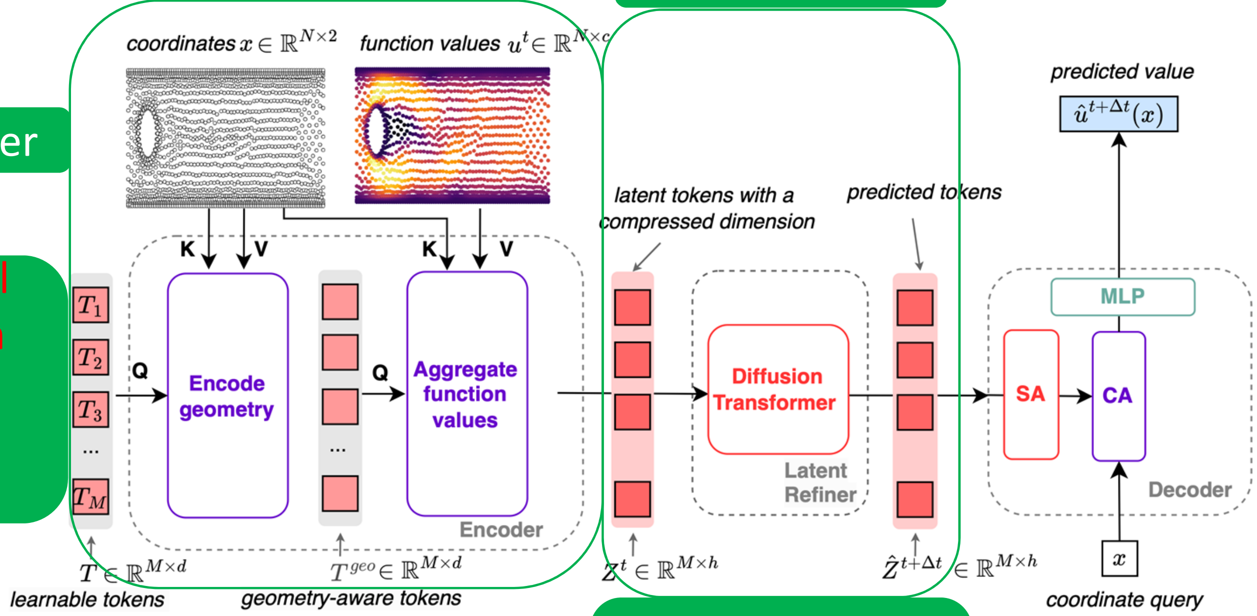
Meshless Neural operators

AROMA: Attentive Reduced Order Model with Attention (Serrano et al. 2024)

Meshless Encoder

Encodes spatial information on problem geometry in a latent space

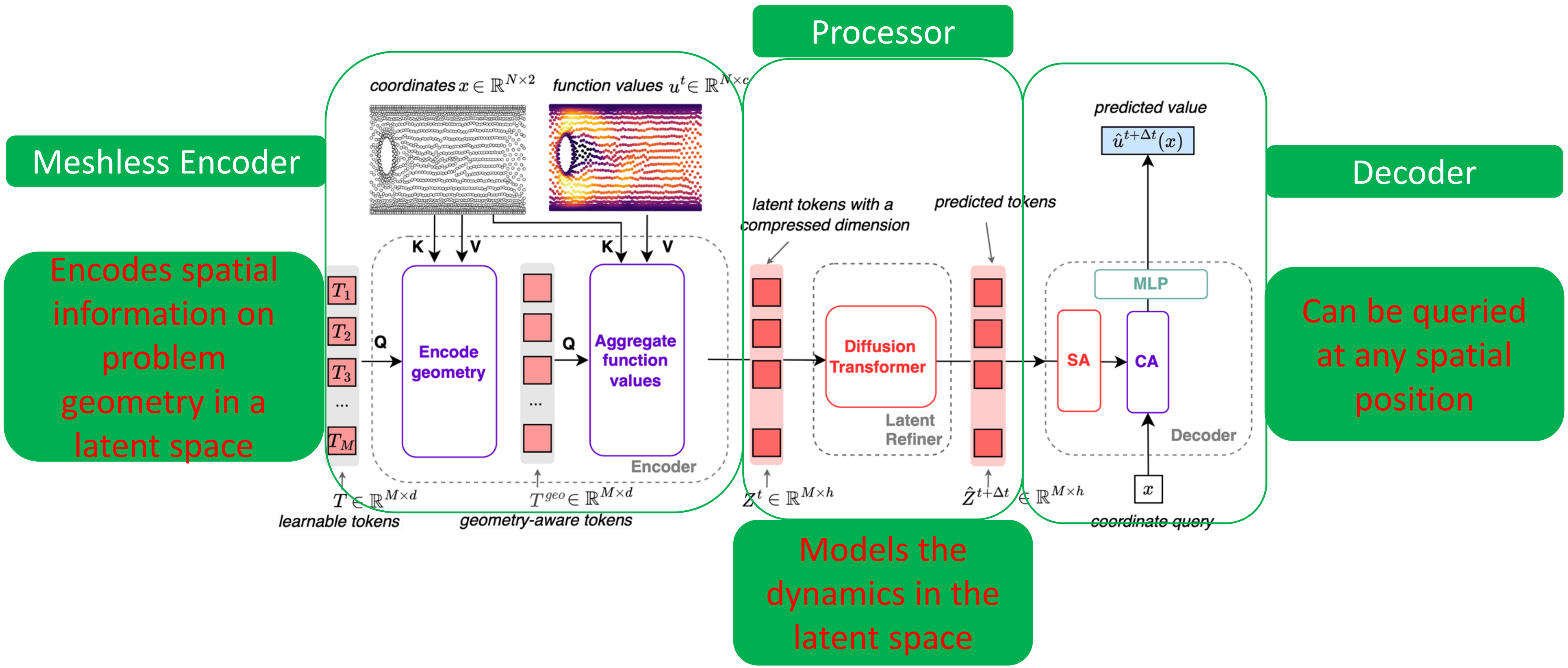
Processor



Models the dynamics in the latent space

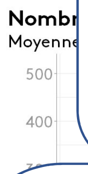
Meshless Neural operators

AROMA: Attentive Reduced Order Model with Attention (Serrano et al. 2024)

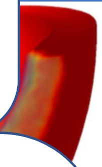


Generalization problem for physical dynamics
Solving parametric equations with data driven models

Mod



c tissue from
020



- One underlying phenomenon – different environments
  - Similar to solving parametric PDEs
  - Without knowing the parameters

Modeling : Meta-Learning

- Learn from a sample of environments, generalize on new ones

- How: few shot adaptation

→ Infer the unknown dynamics from a few samples

Generalization problem for physical dynamics

ZEBRA - In-context generative pretraining (Serrano et al. 2025)

- Inspired by In-context learning in NLP decoders (LLMs)

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

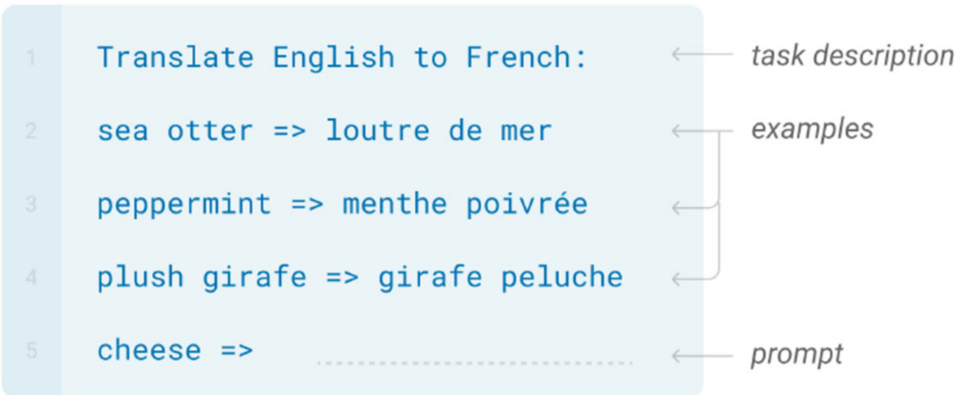


Fig. Brown et al. 2020 (GPT3)

In-Context examples ($u_{n+1} = 3u_n - 1$)

Prompt

Sequence 1: 1, 2, 5, 14, …
Sequence 2: 2, 5, 14, 41, …
Sequence 3: 3, 8, 23, 68, …
Sequence 4: 4, ?

Query new initial condition

Answer *Sequence 4: 4, 11, 32, 95, etc..*

Without optimization LLMs are somehow capable of inferring the correct pattern in the sequences, i.e the underlying parameters: $u_{n+1} = 3u_n - 1$

Fig. Serrano 2025

No gradient update – only context

Generalization problem for physical dynamics
ZEBRA - In-context generative pretraining (Serrano et al. 2025)

https://arxiv.org/abs/2410.03437

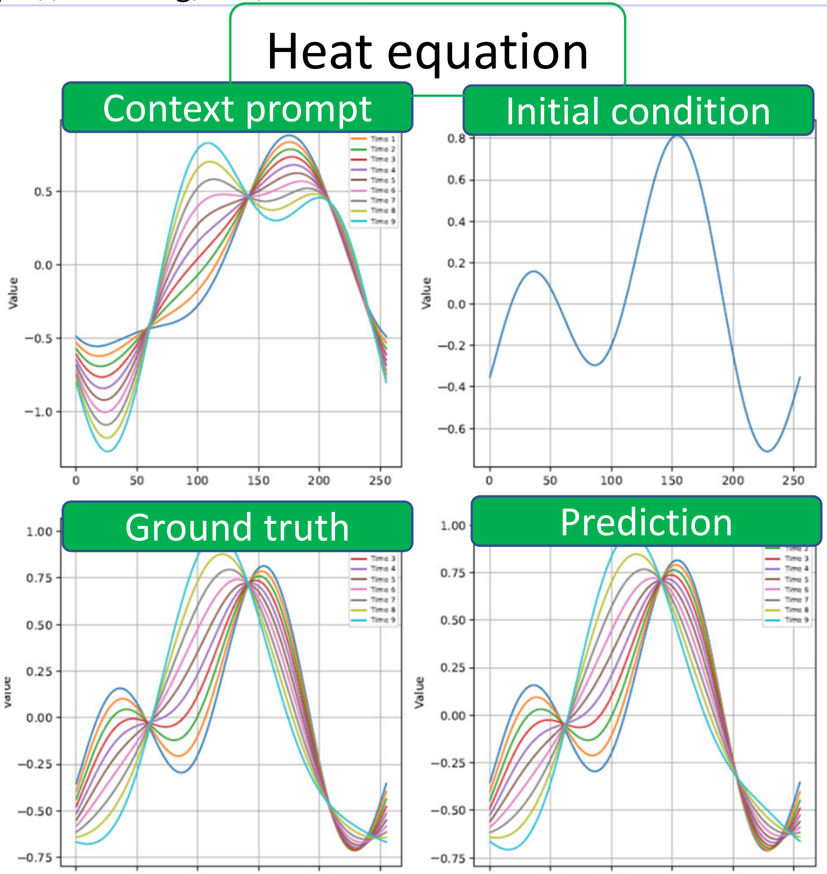


Figure 26: **One-shot adaptation on Heat**

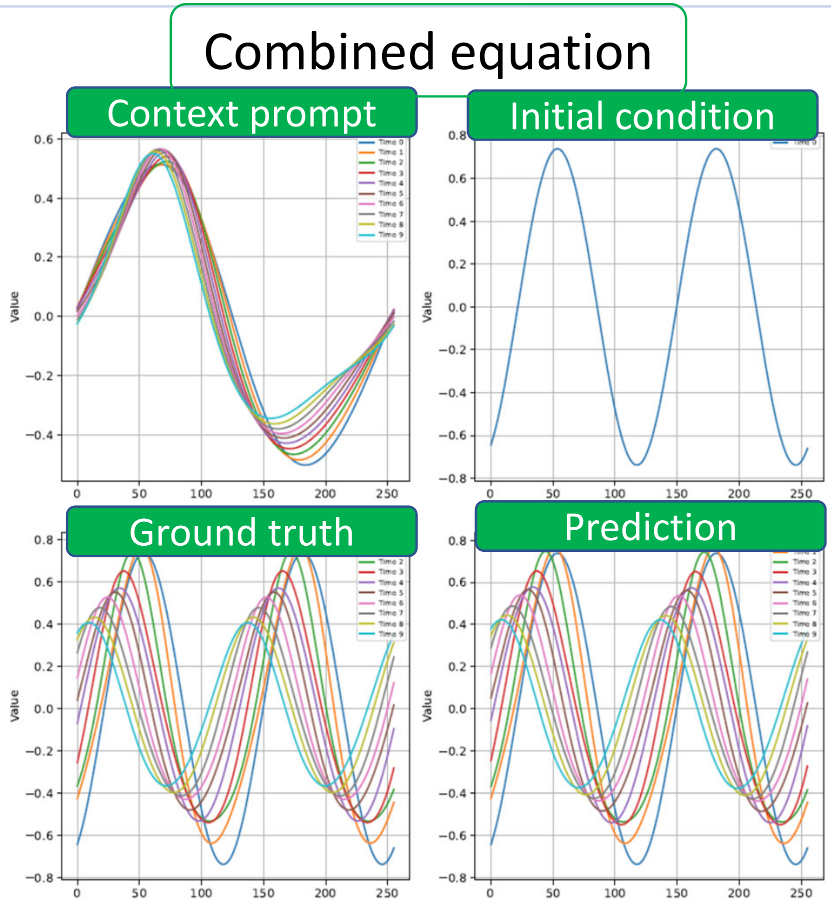


Figure 30: **One-shot adaptation on Combined**

Scaling ML solvers

ML evaluation are often performed on simple problems
Recent attempts to develop more realistic datasets and scale neural networks

Models

Transolver 3, 2026, https://arxiv.org/abs/2602.04940

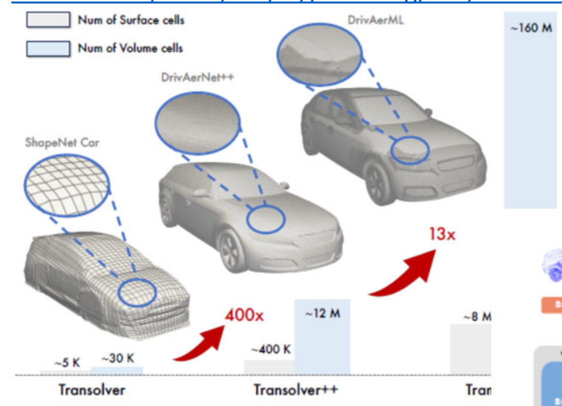
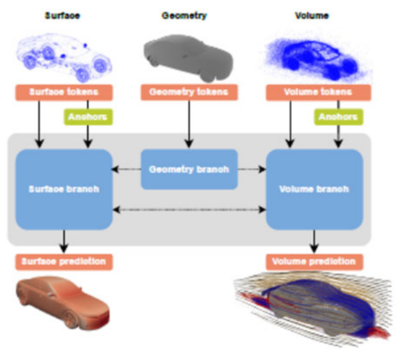


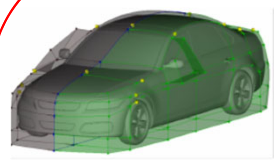
Figure 1. The maximum mesh sizes handled by Trans



AB-UPT, 2025, https://arxiv.org/abs/2502.09692

(a) **Multi-branch architecture:** AB-UPT encodes geometries into a reduced set of latent geometry tokens that are integrated into predictions via cross-attention. Interactions between surface and volume fields are modeled via cross-attention.

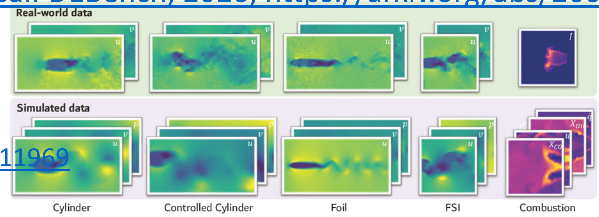
Datasets



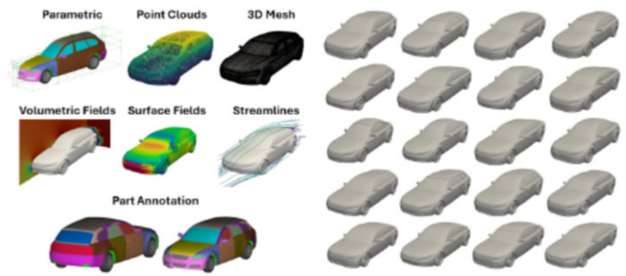
(a)

DrivaerML, 2024, https://arxiv.org/abs/2408.11969

RealPDEBench, 2026, https://arxiv.org/abs/2601.01829



DrivAerNet++, 2024, https://arxiv.org/abs/2406.09624



(a) Data modalities of DrivAerNet++. Top row: different data representations; middle row: CFD simulation results; bottom row: annotated car components.

(b) Selected samples from DrivAerNet++ showing diversity in shape with different car designs (fastback, estateback, and notchback), wheels configurations, and underbody configurations.

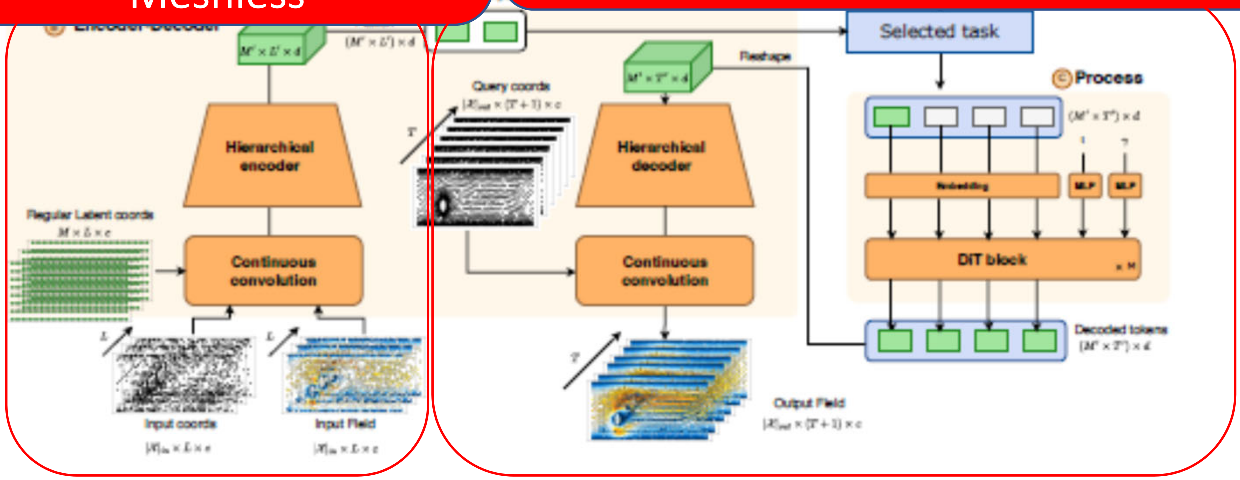
Figure 1: Data modalities and shape variations in the DrivAerNet++ dataset.

Scaling ML Solvers

ECHO, Kassai 2026, Efficient Generative Transformer for Million-Point PDEs

100 x spatio-temporal hierarchical compression
Preserves fidelity
Meshless

Diffusion Transformer: learn trajectory distributions over whole trajectories avoids error accumulation



igure 2. Architecture of the ECHO framework. ECHO comprises two components: (B) a convolutional auto-encoder and (C) a DiT-based enerative process. The auto-encoder uses continuous convolutions to ingest irregular input grids of arbitrary size, map the dynamics to a egular latent grid, and hierarchically compress it; the decoder mirrors this hierarchy and applies a final continuous convolution, enabling ueries at arbitrary output locations. The DiT module is trained with a flow-matching objective to denoise latent tokens, optionally onditioned on PDE parameters. This design allows ECHO to handle irregular grids and support multiple inference tasks (A)

Solves million points problems on a single GPU

Scaling: H100 GPU with 80GB

Table 4. Super-resolution (forecast of 200% of the grid corresponding to more than **20M points**.) Forward generation (For.) on a 1024 × 1024 **Vorticity** grid. Metric: Relative MSE (lower is better). All encoder baselines are out-of-memory (OOM) except ECHO.

Task	GINO	CORAL	AROMA	CALM-PDE	Trans.++	ECHO
For.	OOM	OOM	OOM	OOM	OOM	**3.88e-1**

Foundation models for modeling dynamics
Learning from multiple physics

Models

[DPOT, 2024, https://arxiv.org/abs/2403.03542](https://arxiv.org/abs/2403.03542)

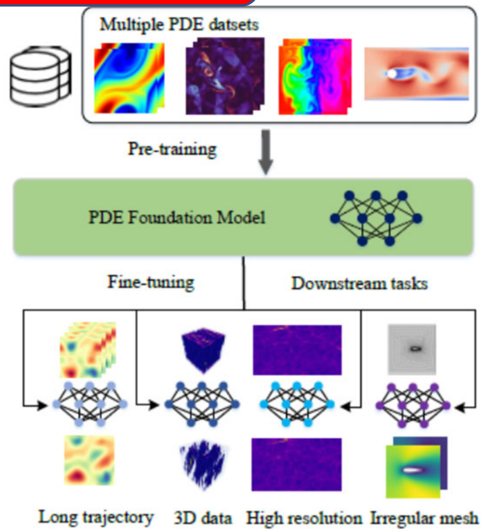


Figure 1. An illustration of pre-training a PDE foundation model using massive data from multiple PDE datasets. The pre-trained model is then used for fine-tuning different downstream operator learning tasks, which can be complex. (Best viewed in color)

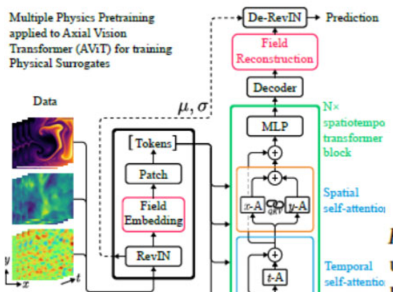
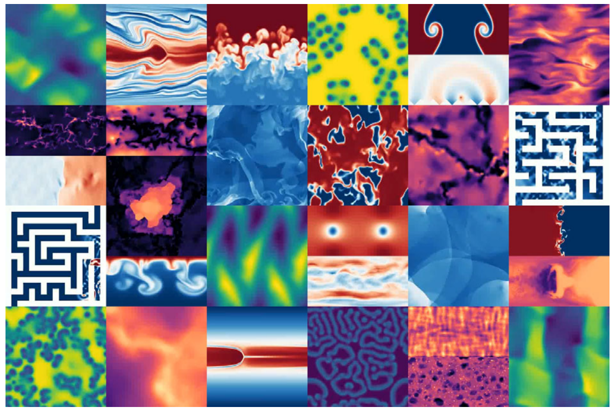


Figure 2: (Left) MPP works by individually normalizing each example using Reversible Instance Normalization (RevIN) then embedding each field individually into a shared, normalized space. A single transformer backbone can then predict the next step for multiple sets of physics. We use an AViT backbone which attends over space and time axis sequentially. Spatial attention is further split by axis, though these share linear projection weights. (Right) The embedding and reconstruction matrices are formed by subsampling a larger 1×1 convolutional filter based on input fields.

[MPP, 2024, https://arxiv.org/abs/2310.02994](https://arxiv.org/abs/2310.02994)

Datasets

[The Well, 2025, 15 TB, 16 datasets](https://polymathic-ai.org/the_well/), https://polymathic-ai.org/the_well/



Foundation models for modeling dynamics
DAVY, Castano Segade et al. 2026, Multiphysics Foundation Encoders

- Unified multiphysics architecture
- Handle heterogeneous sources/ variables
- Decoupling dynamic representation learning from downstream tasks
- Self Supervised Learning (JEPA)

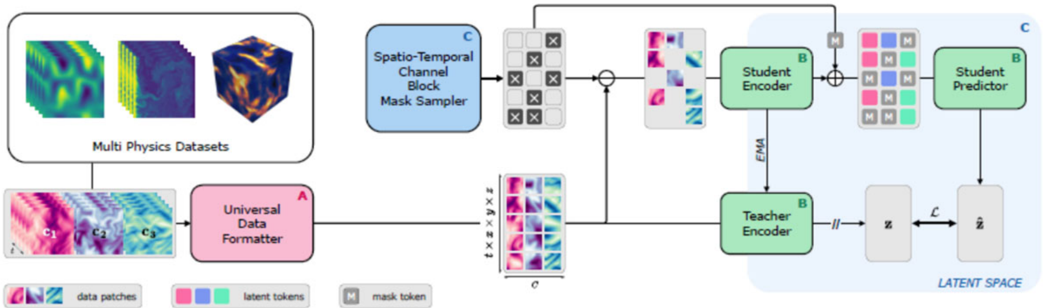
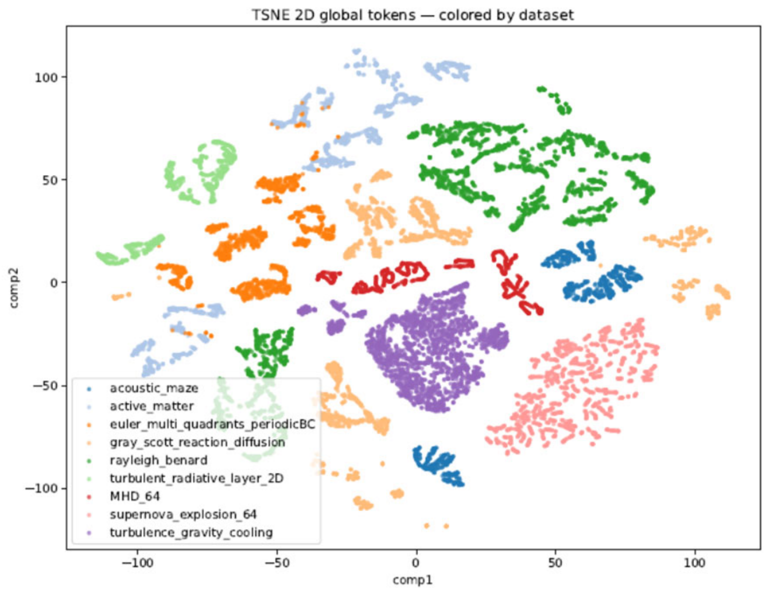


Figure 1. Overview of DAVY. DAVY is pretrained on heterogeneous multi-physics spatiotemporal data using partially observed inputs. A universal data formatter (A) maps inputs with varying spatial dimensions and physical variables to a common token representation. During pretraining, a teacher–student strategy (B) is used for representation space learning (C): the student encoder receives masked spatiotemporal and channel-wise observations and predicts latent representations produced by a teacher encoder from the corresponding unmasked inputs. The teacher parameters are updated as an exponential moving average of the student. After pretraining, the student encoder is reused as a frozen context extractor for downstream tasks.

Multiple equations
Colors = physical phenomena



(a) Multiphysics.

Thanks