Google Research

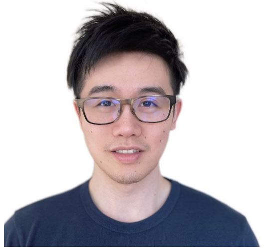
Recent Advances in Probabilistic Scientific Machine Learning for Chaotic Dynamical Systems

June 13th 2024
IMSI Workshop in Probabilistic SciML

Leonardo Zepeda-Núñez
lzepedanunez@google.com

Google Research

Who we are and what we do



Zhong Yi Wan

Leonardo Zepeda-Núñez

Fei Sha

Mission Foundational technologies that drive **efficient modeling** of large-scale, high-stake, and **computationally intensive** physical systems

Representation/Dynamics Learning: Leverage implicit representation tools for learning the dynamics of advection-dominated systems.

Probabilistic Modelling: Leverage generative AI tools for physical systems (UQ)

Google Research

Machine Learning by tasks

Machine learning can be roughly divided into 3 buckets:

ML Task	Underlying Math Problem	Applied Math Techniques
Classification	Learning a **Partition** of a domain	Meshing techniques
Regression	Learning a **Map**	Approximating functions Solving ODEs/PDEs Approximating dynamics
Generation	Learning a **Distribution**	Solving SDEs Sampling from Distributions Uncertainty quantification

Classical Problems in Numerical Analysis / Computational Maths

Difference: **Much Higher Dimension!!**

Machine Learning by tasks

Machine learning can be roughly divided into 3 buckets:

ML Task	Underlying Math Problem	Applied Math Techniques
Classification	Learning a **Partition** of a domain	Meshing techniques
Regression	Learning a **Map**	Approximating functions Solving ODEs/PDEs Approximating dynamics
Generation	Learning a **Distribution**	Solving SDEs Sampling from Distributions Uncertainty quantification

Today's talk!

Classical Problems in Numerical Analysis / Computational Maths

Difference: **Much Higher Dimension!!**

Who we are and what we do

Representation/Dynamics Learning

- A. Boral, Z. Y. Wan, **L. Zepeda-Núñez**, J. Lottes, Q. Wang, Y. Chen, J. Anderson, F Sha. Neural Ideal Large Eddy Simulation: Modeling Turbulence with Neural Stochastic Differential Equations, *NeurIPS 2023*.
- Z. Y. Wan, **L. Zepeda-Núñez**, A. Boral, F. Sha. Evolve Smoothly, Fit Consistently: Learning Smooth Latent Dynamics For Advection-Dominated Systems, *ICLR 2023*
- G. Dresdner, D. Kochkov, P. Norgaard, **L. Zepeda-Núñez**, J. A. Smith, M. Brenner, S. Hoyer. Learning to correct spectral methods for simulating turbulent flows. *TMLR 2023*.

Probabilistic Modelling

- M. A. Finzi, A. Boral, A. G. Wilson, F. Sha, **L. Zepeda-Núñez**, User-defined Event Sampling and Uncertainty Quantification in Diffusion Models for Physical Dynamical Systems, *ICML 2023*
- Y. Schiff, Z. Y. Wan, J. B. Parker, S. Hoyer, V. Kuleshov, F. Sha, **L. Zepeda-Núñez**. DySLIM: Dynamics Stable Learning by Invariant Measure for Chaotic Systems. *ICML 2024*.
- Z. Y. Wan, R. Baptista, Y. Chen, J. Anderson, A. Boral, F. Sha, **L. Zepeda-Núñez**. Debias Coarsely, Sample Conditionally: Statistical Downscaling through Optimal Transport and Probabilistic Diffusion Models, *NeurIPS 2023*.
- B. Barthel Sorensen, **L. Zepeda-Núñez**, I. Lopez-Gomez, Z. Y. Wan, R. Carver, F. Sha, and T. P. Sapsis. A probabilistic framework for learning non-intrusive corrections to long-time climate simulations from short-time training data, arXiv:2408.02688.
- B. Zhang, M. Guerra, Q. Li, and **L. Zepeda-Núñez**. Back-Projection Diffusion: Solving the Wideband Inverse Scattering Problem with Diffusion Models. CMAME 2025.
- I. Lopez-Gomez, Z. Y. Wan, **L. Zepeda-Núñez**, T. Schneider, J. Anderson, F. Sha. Dynamical-generative downscaling of climate model ensembles. PNAS 2025.
- R. Molinaro, S.Lanthaler, B. Raonić, T. Rohner, V. Armegioiu, Z. Y. Wan, F. Sha, S. Mishra, **L. Zepeda-Núñez.** Generative AI for fast and accurate Statistical Computation of Fluids, arXiv:2409.18359.

Google Research

Probabilistic Reformulation to Leverage GenAI

Focus: **High-Dimensional** problems

Probabilistic SciML ➡ two stage approach:

- Recast the problem in a “probabilistic” manner
- Leverage and tailor genAI tools to solve the new formulation

Open source code:

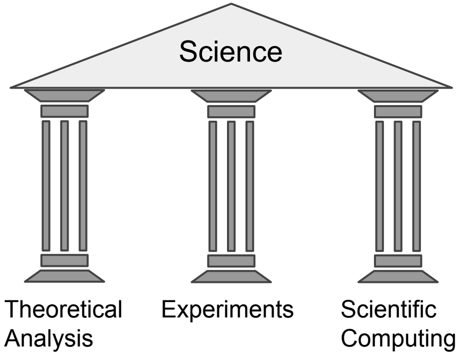
- https://github.com/google-research/swirl-dynamics

What is SciML?

Scientific ~~Machine Learning~~ Computing

What is Scientific Computing?

> *“… **Scientific Computing** is the collection of tools, techniques, and theories required to solve on a computer **mathematical models** of problems in **Science** and **Engineering**”*
>
> [Golub and Ortega]



Gene H. Golub and James M. Ortega. *Scientific Computing and Differential Equations – An Introduction to Numerical Methods*. Academic Press, 1992.

SciML: Accelerating Science and Engineering

Scientific Computing:

In silico for downstream applications but still experimental **data** is the gold standard

Issues with *in silico* workflows

- More **accurate** simulations require more **expensive** computations
- Need to run **thousands** of simulations with different parameters
- **Quantify the uncertainty** to increase robustness

SciML is the evolution of **Scientific computing** to further accelerate the development pipelines by making **computations more efficient**.

Probabilistic SciML seeks to address computational bottlenecks in **Scientific computing** dealing with probability distributions, by leveraging **generative AI** tools

Google Research

Probabilistic SciML: Two examples (or Vignettes)

Learning Stable Dynamics by Invariant Measure Matching

Y. Schiff, Z. Y. Wan, J. B. Parker, S. Hoyer, V. Kuleshov, F. Sha, L. Zepeda-Núñez. DySLIM: Dynamics Stable Learning by Invariant Measure for Chaotic Systems. ICML 2024.

Generative AI for fast and accurate Statistical Computation of Fluids

R. Molinaro, S. Lanthaler, B. Raonic, T Rohner, V. Armegioiu, Z. Y. Wan, F. Sha, S. Mishra, and L. Zepeda-Núñez, ArXiv:2409.18359

Google Research

Learning Stable Dynamics by Invariant Measure Matching

$$\partial_t u = \mathcal{F}[u(t)] \xrightarrow{\text{discretize}} u_{k+1} = \mathcal{S}_{\Delta t}(u_k),$$

$$u_k = \mathcal{S}(u_{k-1}) = \mathcal{S} \circ \mathcal{S}(u_{k-2}) = \ldots = \mathcal{S}^k(u_0)$$

known through data

$$\mathcal{D} = \{(u_j^{(i)})_{j=0}^{\ell(i)}\}_{i=1}^n$$

Invariant measure

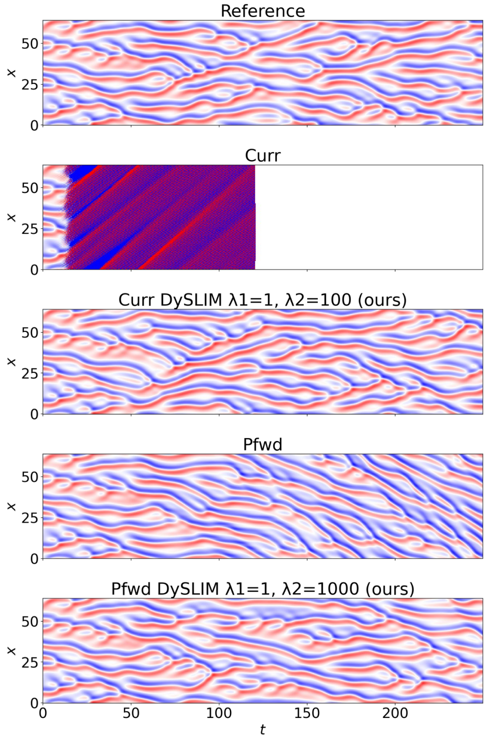
$$\{u_0^{(i)}\}_{i=1}^n \overset{iid}{\sim} \mu_0 = \mu^*$$

$$\mu_j := \mathcal{S}_\sharp^j \mu_0 = \mathcal{S}_\sharp^j \mu^* = \mu^*$$

Data-driven dynamics learning

$$\min_\theta E_j E_{u_j \sim \mu_j} \left[\|\mathcal{S}_\theta(u_j) - \mathcal{S}(u_j)\|^2 \right]$$

Unstable or goes to the wrong attractor



$$\partial_t u + u\partial_x u + \nu\partial_{xx} u - \nu\partial_{xxxx} u = 0 \qquad \text{in } [0, L] \times \mathbb{R}^+,$$

Learning Stable Dynamics by Invariant Measure Matching

Data-driven dynamics learning

$$\min_{\theta} E_j E_{u_j \sim \mu_j} \left[\|\mathcal{S}_\theta(u_j) - \mathcal{S}(u_j)\|^2 \right]$$

Constrained minimization

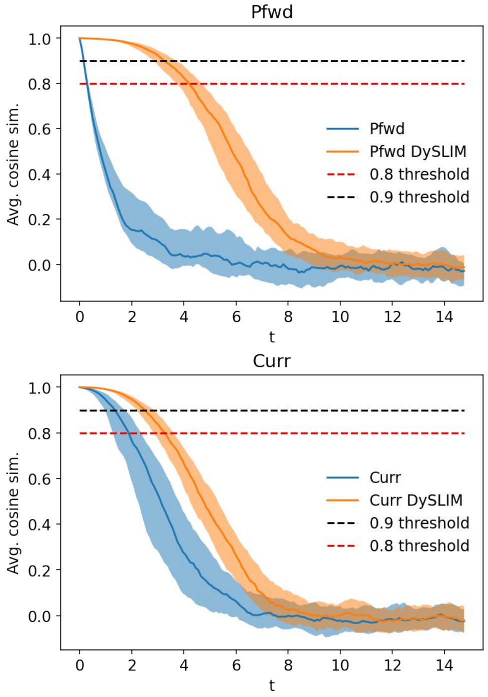
$$\min_{\theta} \mathcal{L}(\theta) \quad \text{s.t.} \quad \mu_\theta^* = \mu^*$$

Relaxed version

$$\mathcal{L}_\lambda^D(\theta) = \mathcal{L}(\theta) + \lambda D(\mu^*, \mu_\theta^*)$$

Sampling and matching

$$D(\mu^*, \mu_\theta^*) \approx D(\mu^*, (\mathcal{S}_\theta^k)_\sharp \mu^*)$$



$$D(\mu^*, \mu_\theta^*) = E_{u,u' \sim \mu^*}[\kappa(u, u')] + E_{v,v' \sim \mu_\theta^*}[\kappa(v, v')] - 2E_{u \sim \mu^*, v \sim \mu_\theta^*}[\kappa(u, v)]$$

Google Research

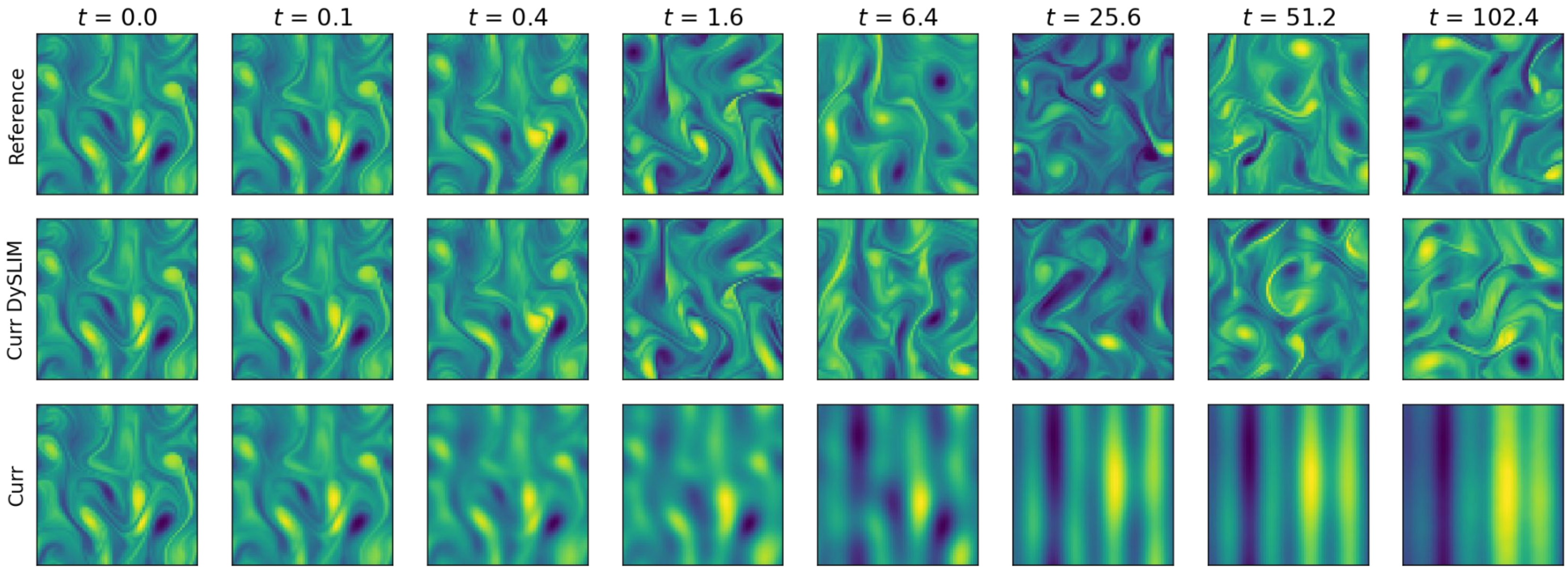
Learning Stable Dynamics by Invariant Measure Matching

$$\mathcal{L}^{\ell\text{-step}}(\theta) = \mathbb{E}_j \mathbb{E}_{u_j \sim \mu_j} \sum_{k=1}^{\ell} \omega(k) \left\| \mathcal{S}_\theta^k(u_j) - u_{j+k} \right\|^2$$

$$\mathcal{L}^{\mathrm{Pfwd},\ell}(\theta) = \mathbb{E}_j \mathbb{E}_{u_j \sim \mu_j} \sum_{k=1}^{\ell} \omega(k) \left\| \mathcal{S}_\theta(\mathbf{sg}(\mathcal{S}_\theta^{\ell-1}(u_j))) - u_{j+k} \right\|^2$$

Baseline	Batch size	Learning rate	MELR↓ ($\times 10^{-2}$) Base	MELR↓ ($\times 10^{-2}$) DySLIM	MELRw↓ ($\times 10^{-2}$) Base	MELRw↓ ($\times 10^{-2}$) DySLIM	covRMSE↓ ($\times 10^{-2}$) Base	covRMSE↓ ($\times 10^{-2}$) DySLIM	Wass1↓ ($\times 10^{-2}$) Base	Wass1↓ ($\times 10^{-2}$) DySLIM	TCM↓ ($\times 10^{-2}$) Base	TCM↓ ($\times 10^{-2}$) DySLIM
Pushforward	128	1e-4	3.19	**2.46**	0.53	0.53	6.81	**6.69**	4.64	**4.51**	3.68	**0.72**
Curriculum	64	5e-4	5.35	**1.64**	0.95	**0.45**	8.13	**6.95**	9.66	**4.76**	3.50	**2.83**
1-step	64	5e-4	2.77	**1.84**	0.44	0.85	7.93	**7.30**	16.2	**5.55**	5.39	**2.45**

Learning Stable Dynamics by Invariant Measure Matching

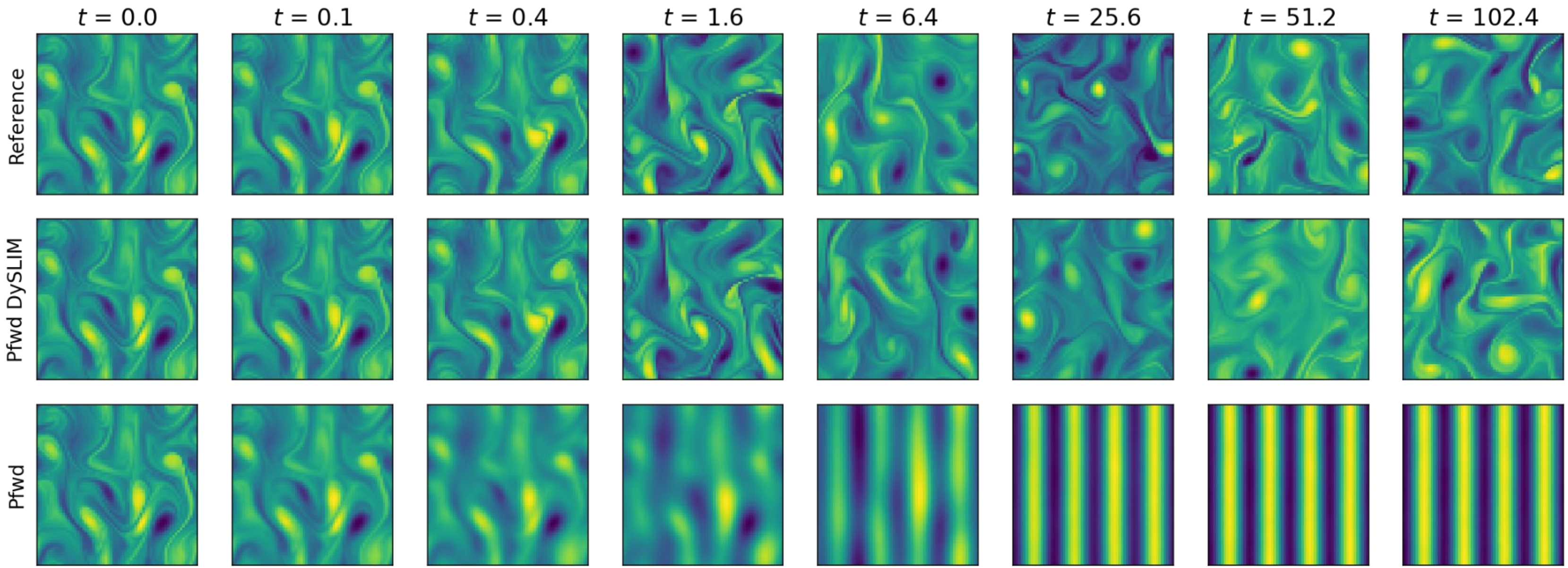


$$\frac{\partial u}{\partial t} = -\nabla \cdot (u \otimes u) + \nu \nabla^2 - \frac{1}{\rho}\nabla p + \mathbf{f} \qquad \text{in } \Omega,$$

$$\nabla \cdot u = 0 \qquad \text{in } \Omega,$$

$$\mathbf{f} = \begin{pmatrix} 0 \\ \sin(k_0 y) \end{pmatrix} + 0.1u,$$

Learning Stable Dynamics by Invariant Measure Matching



$$\frac{\partial u}{\partial t} = -\nabla \cdot (u \otimes u) + \nu \nabla^2 - \frac{1}{\rho} \nabla p + \mathbf{f} \qquad \text{in } \Omega,$$

$$\nabla \cdot u = 0 \qquad \text{in } \Omega,$$

$$\mathbf{f} = \begin{pmatrix} 0 \\ \sin(k_0 y) \end{pmatrix} + 0.1u,$$

Probabilistic SciML: Learning Statistical Solutions

Generative AI for fast and accurate Statistical Computation of Fluids

R. Molinaro, S. Lanthaler, B. Raonic, T Rohner, V. Armegioiu, Z. Y. Wan, F. Sha, S. Mishra, and L. Zepeda-Núñez, ArXiv:2409.18359

How to efficiently compute statistical solutions of fluid flows.

An Incredible Hammer: Diffusion Models

Learn a prior $p(x)$ of the high-resolution data.

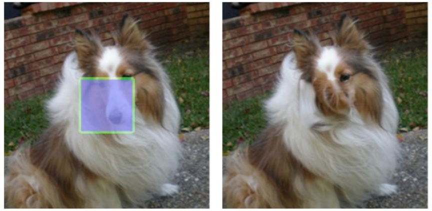
Why?

High-quality samples

High **coverage** of the distribution

Then conditional sampling

$$p(x|C'x=y)$$

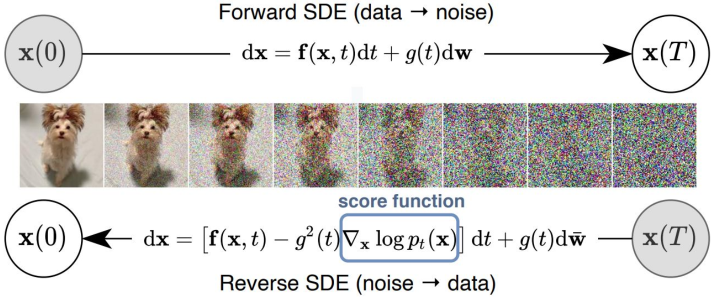


Input n = 1

[Lugmayr et al. 2021]

A cute corgi lives in a house made out of sushi.

[Saharia et al, 2020]

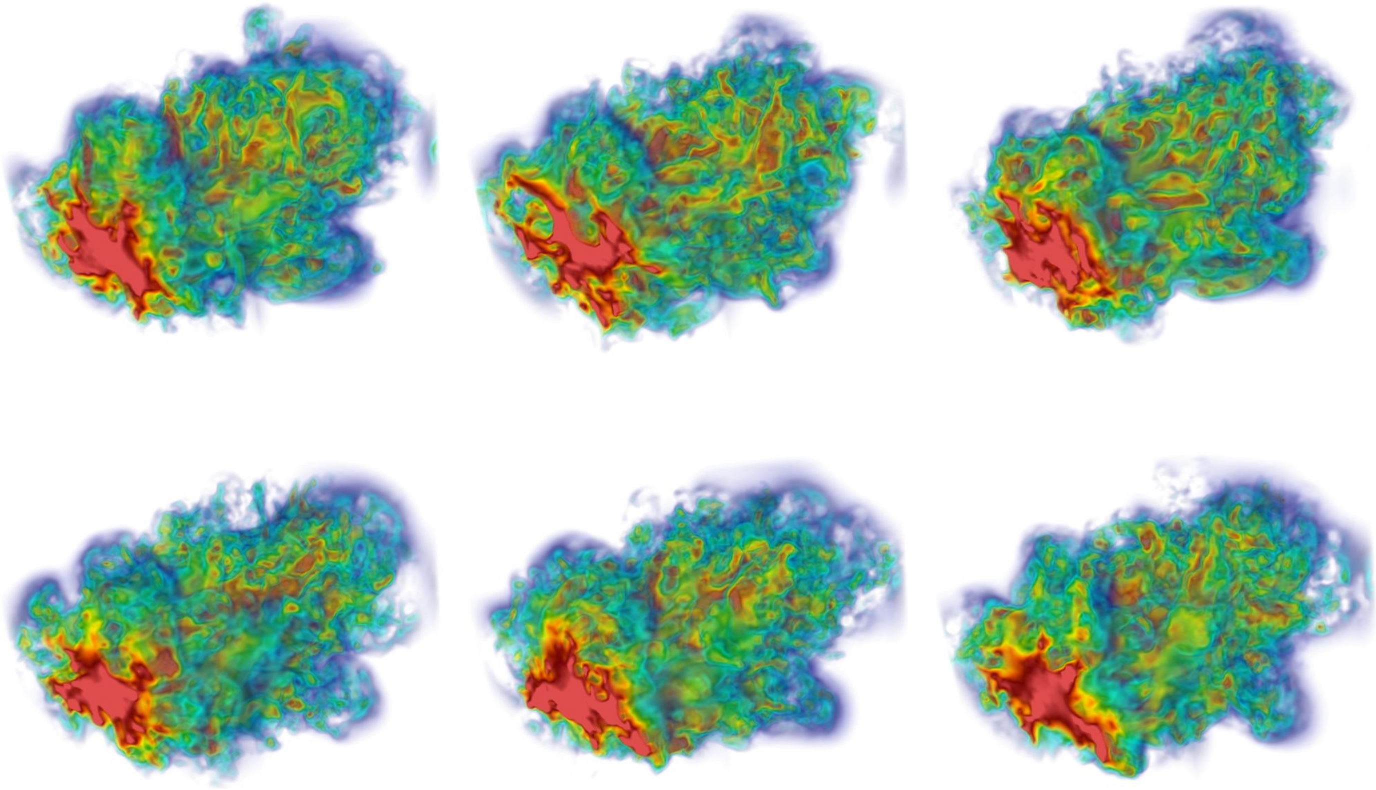


[Song et al, 2020]

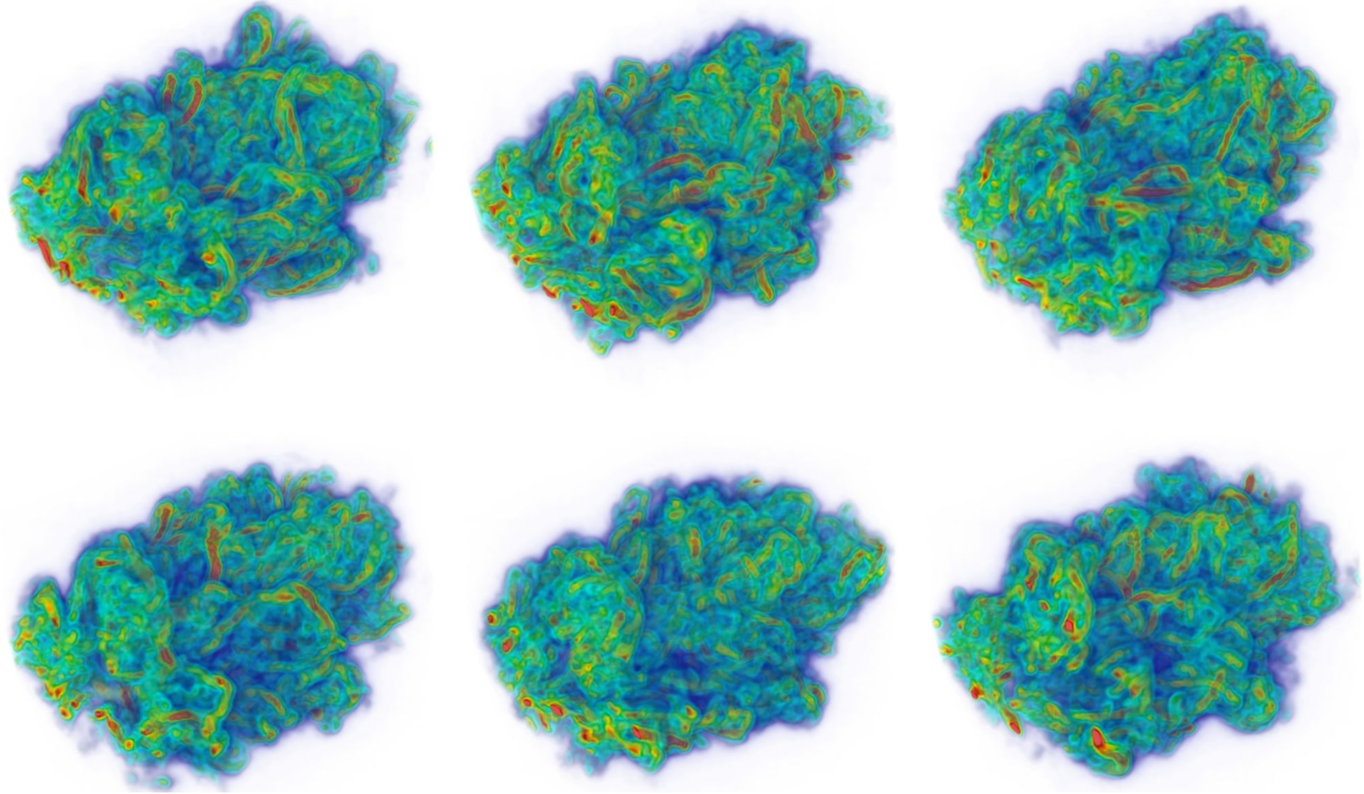
$$\mathbb{E}[x_0|x_t] = \frac{x_t + \sigma_t^2 \nabla_{x_t} \log p(x_t)}{s_t} := \hat{x}_0(x_t)$$

More details: http://smai.emath.fr/cemracs/cemracs23/summer-school.html

Quick Quiz: Real or Generated Samples?



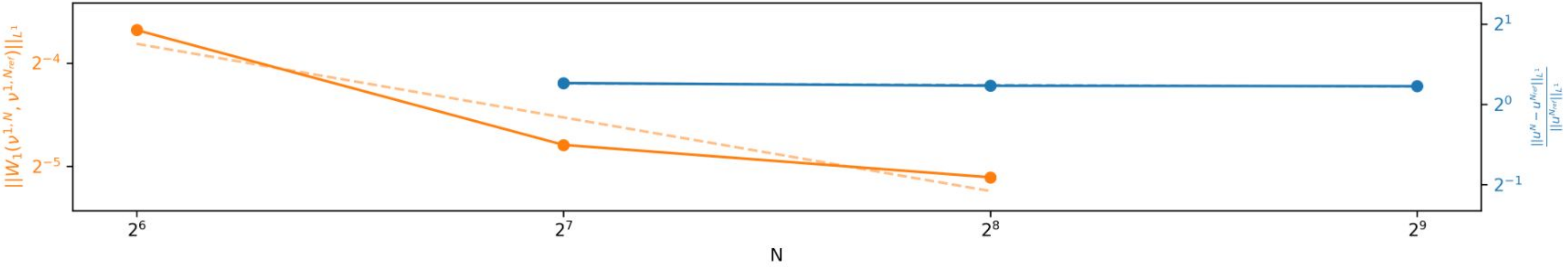
Quick Quiz: Real or Generated Samples?



Google Research

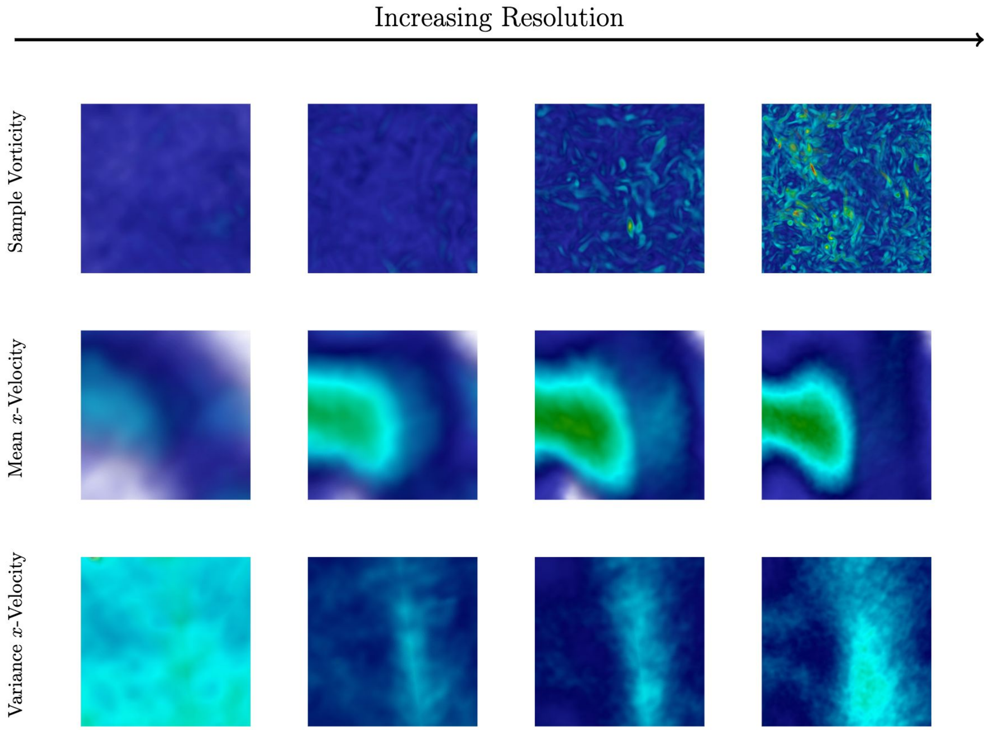
Why Statistical Solutions?

Due to the chaotic nature of system, its numerical solution **doesn't converge** under mesh refinement (3D Cylindrical Shear Flow problem)



Even though there is no convergence per trajectory, the **statistics** of the flow do converge.

Why Statistical Solutions?



Google Research

Statistical Solutions

$$\partial_t u(x,t) + \mathcal{L}\left(u, \nabla_x u, \nabla_x^2 u, \ldots\right) = 0, \quad \forall x \in D \subset \mathbb{R}^d, t \in (0,T),$$
$$\mathcal{B}(u) = 0, \quad \forall (x,t) \in \partial D \times (0,T),$$
$$u(0,x) = \bar{u}(x), \quad x \in D,$$

Solution operator $\mathcal{S} : [0,T] \times \mathbb{X} \mapsto \mathbb{X}$

$$u(t) = \mathcal{S}^t(\bar{u}) = \mathcal{S}(t, \bar{u})$$

Discrete solution operator $\mathcal{S}^{t,\Delta} : \mathbb{X}^\Delta \mapsto \mathbb{X}^\Delta$

Statistical solution $\hat{\mu}_t = \mathcal{S}^t_{\#}\bar{\mu}$

$\uparrow \lim_{\Delta \to 0}$

Approximate statistical solution $\hat{\mu}_t^\Delta = \mathcal{S}^{t,\Delta}_{\#}\bar{\mu}$

$$\mathrm{Lip}(\mathcal{S}^{t,\Delta}) \to \infty \qquad \lim_{\Delta \to 0} W_p(\hat{\mu}_t^\Delta, \hat{\mu}_t) = 0$$

Extended Statistical Solutions

$$\left(\mathcal{S}^{t,\Delta} \times \mathtt{id}\right) : \mathbb{X} \to \mathbb{X} \times \mathbb{X}$$

$$\left(\mathcal{S}^{t,\Delta} \times \mathtt{id}\right)(\bar{u}) = \left(\mathcal{S}^{t,\Delta}(\bar{u}), \bar{u}\right)$$

Approximate extended statistical solution

$$\mu_t^\Delta := \left(\mathcal{S}^{t,\Delta} \times \mathtt{id}\right)_{\#} \bar{\mu}^\Delta$$

extended statistical solutions

$$\mu_t = \lim_{\Delta \to 0} \mu_t^\Delta = \lim_{\Delta \to 0} \left(\mathcal{S}^{t,\Delta} \times \mathtt{id}\right)_{\#} \bar{\mu}^\Delta$$

Conditional Representation

$$\mu_t(du, d\bar{u}) = \boxed{P_t(du \,|\, \bar{u})}\,\bar{\mu}(d\bar{u}),$$

We need to learn the conditional probability

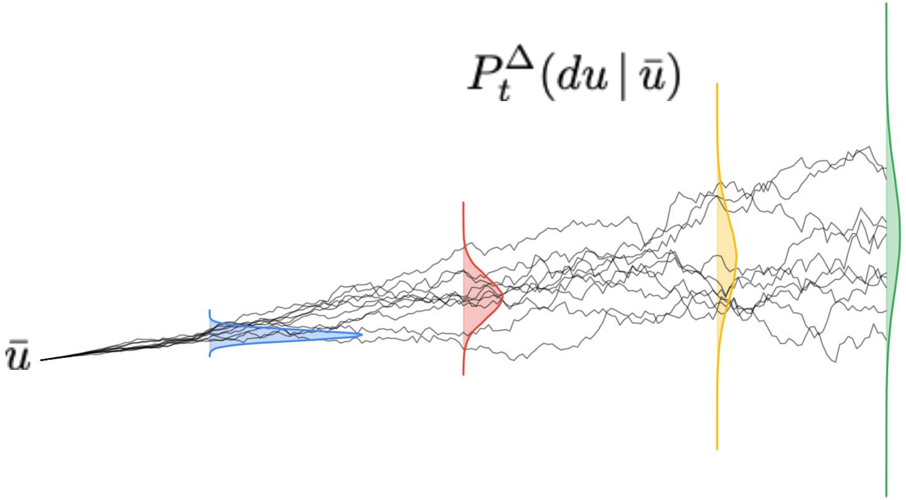
$$\mu_t^{\Delta}(du, d\bar{u}) = \boxed{P_t^{\Delta}(du \,|\, \bar{u})}\,\bar{\mu}^{\Delta}(d\bar{u})$$

Only have access to the approximate conditional probability

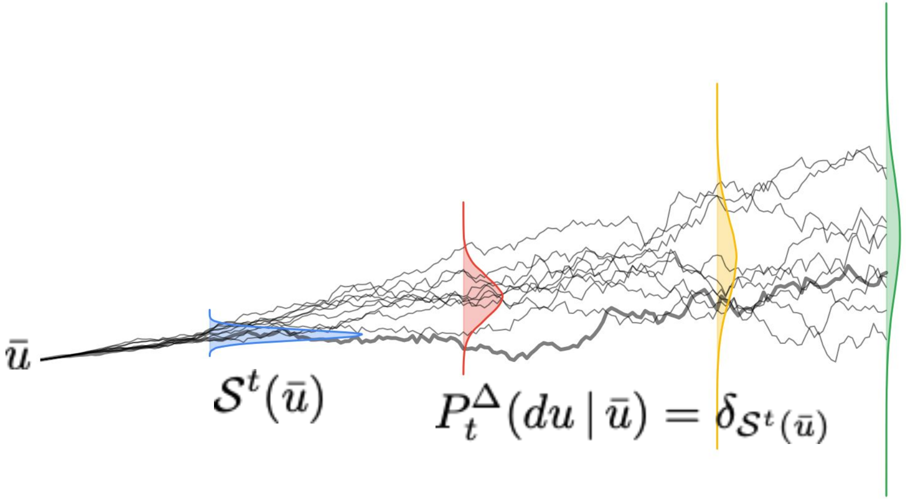
We need to **sample** from this conditional probability

If the operator is well-behaved then
it goes back to the deterministic case $P_t(du \,|\, \bar{u}) = \delta_{\mathcal{S}^t(\bar{u})}$

Conditional Representation

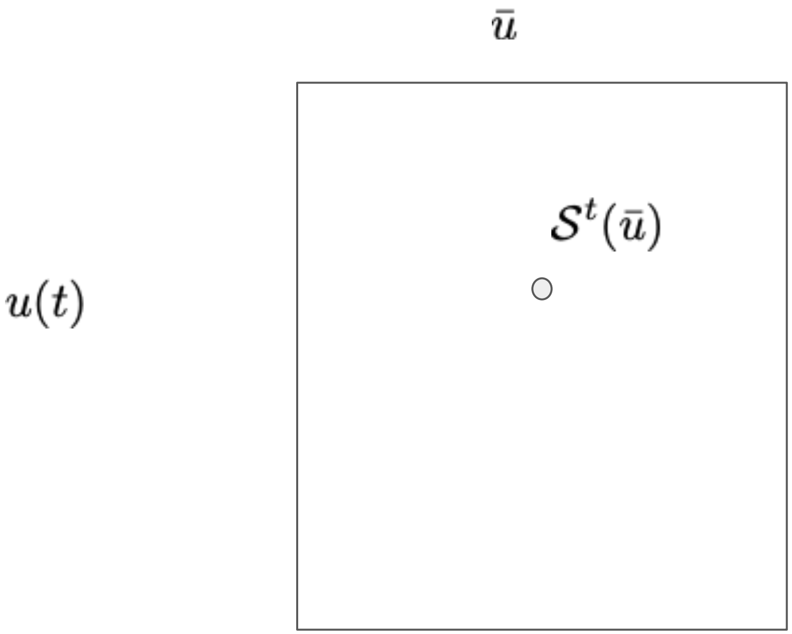


Conditional Representation



Conditional Representation

$$P_t^{\Delta}(du \mid \bar{u}) = \delta_{\mathcal{S}^t(\bar{u})}$$

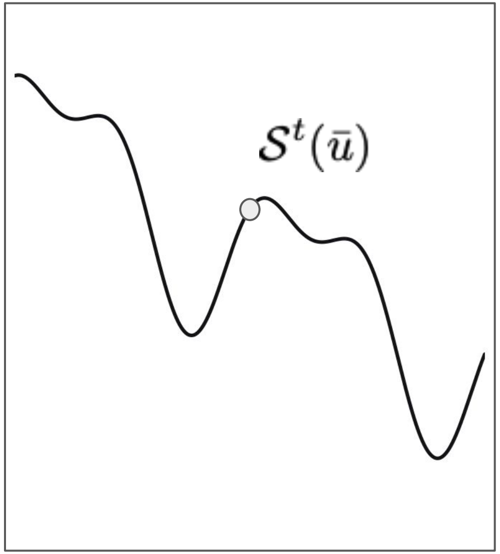


Conditional Representation

$$P_t^{\Delta}(du \mid \bar{u}) = \delta_{\mathcal{S}^t(\bar{u})}$$

$\bar{u}$

$u(t)$

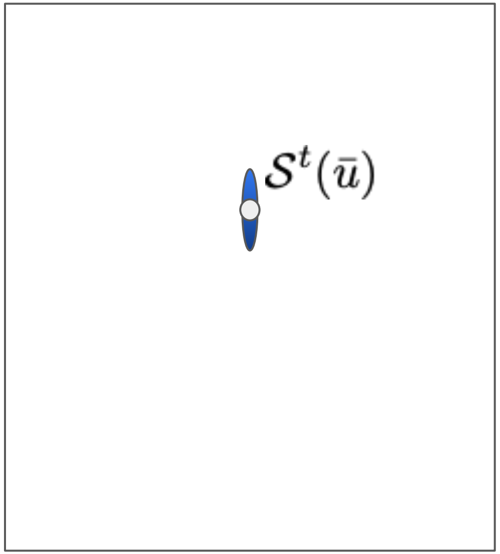


Conditional Representation

$$P_t^{\Delta}(du \mid \bar{u})$$

$\bar{u}$

$u(t)$

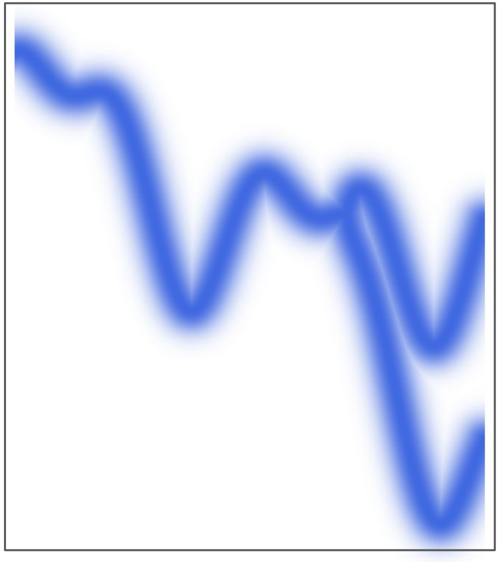


Conditional Representation

$$P_t^{\Delta}(du \mid \bar{u})$$

$\bar{u}$

$u(t)$

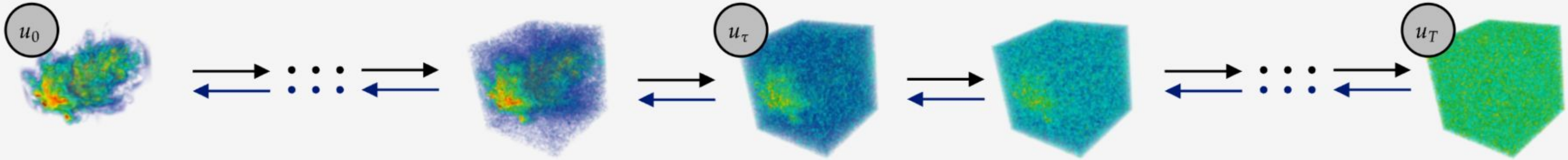


Methodology

$$u_{\tau=0} \sim P_t^{\Delta}(du|\bar{u})$$

Forward and Reverse (Sampling) Step

$$u_\tau = \frac{\dot{s}_\tau}{s_\tau} u_\tau d\tau + s_\tau \sqrt{2\dot{\sigma}_\tau \sigma_\tau} dW_\tau$$



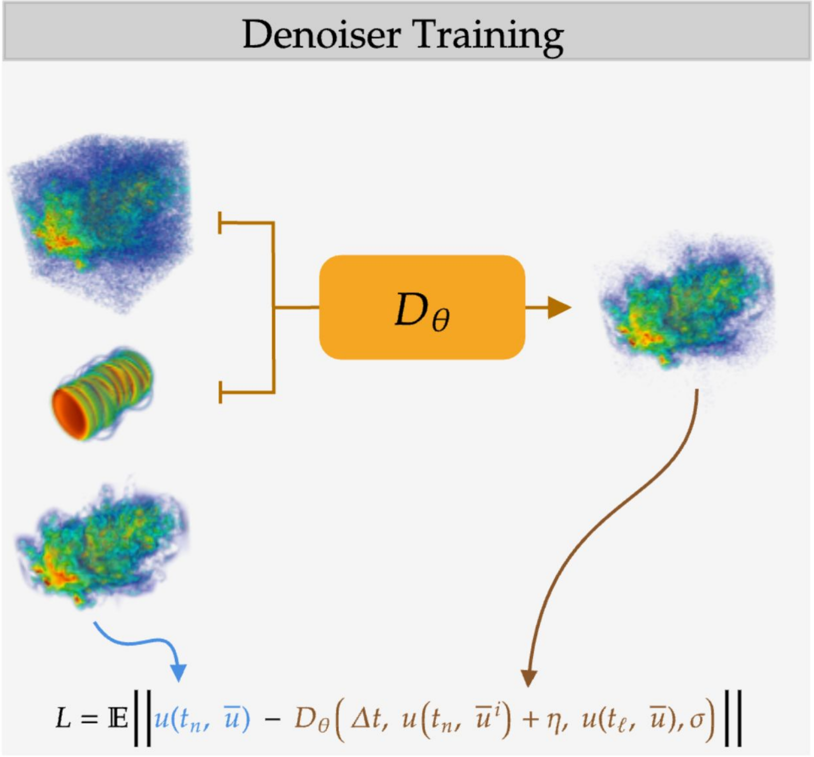
$$du_\tau = \left(\frac{\dot{s}_\tau}{s_\tau} u_\tau - 2\sigma_\tau \dot{\sigma}_\tau s_\tau^2 \nabla_{u_\tau} p(u_\tau) \right) d\tau + s\sqrt{2\dot{\sigma}\sigma_\tau}\, d\widehat{W}_\tau \longrightarrow du_\tau = 2\left(\frac{\dot{\sigma}_\tau}{\sigma_\tau} + \frac{\dot{s}_\tau}{s_\tau} \right) d\tau - 2s_\tau \frac{\dot{\sigma}_\tau}{\sigma_\tau} D_\theta(\Delta t,\, u_{\tau+1},\, \bar{u},\, \sigma_\tau) d\tau + s\sqrt{2\dot{\sigma}_\tau \sigma_\tau}\, d\widehat{W}_\tau$$

Score function

Denoiser

$$\nabla_u \log p_\tau(u_\tau) \approx \frac{D_\theta(\hat{u}_\tau, \sigma_\tau) - \hat{u}_\tau}{s_\tau \sigma_\tau^2}, \quad \text{with } \hat{u}_\tau := \frac{u_\tau}{s_\tau},$$

Methodology



Why Probabilistic Approach?

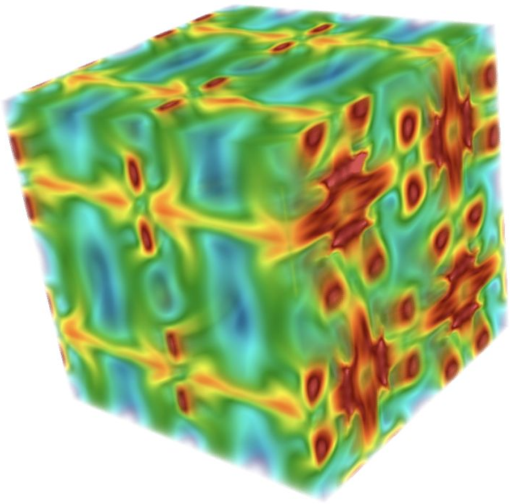
Deterministic models:

- Learn the mean
- Small variability
- Tend to force a small Lipschitz constant for stability

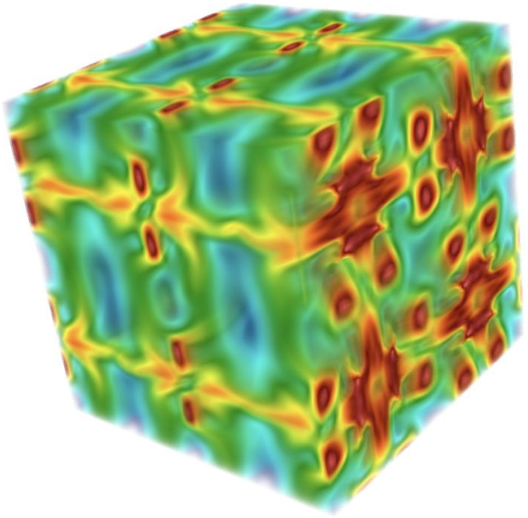
Probabilistic models:

- Sampler for the target distribution
- Moments computed by Monte-Carlo
- Distribution convergence
- Low sample complexity
- Learns transitions in the behavior of the distribution quasi-deterministic to turbulent

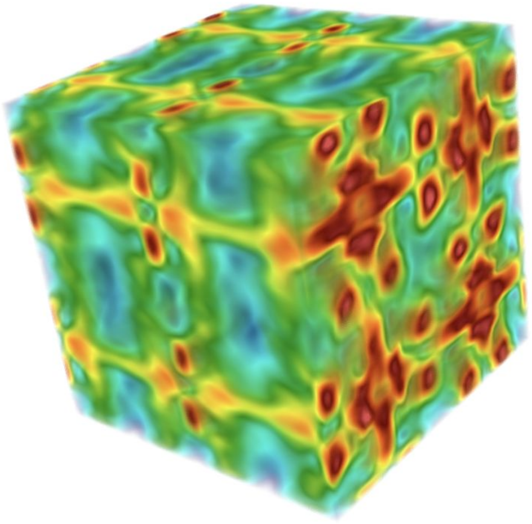
Taylor-Green Vortex (samples T=1)



DNS

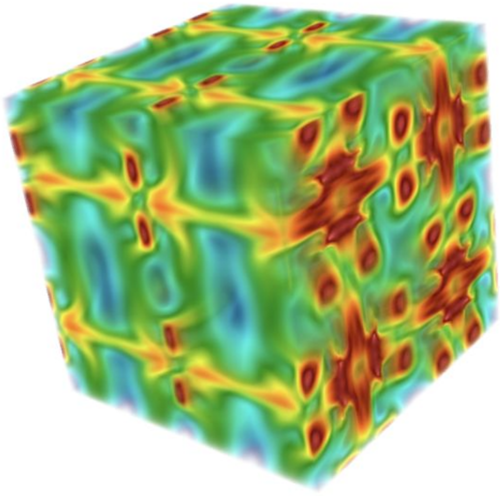


GenCFD

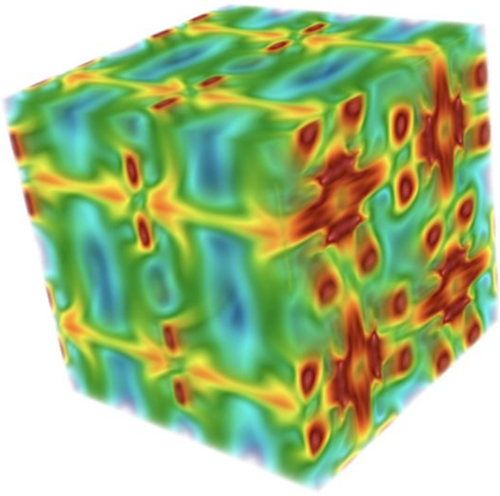


FNO

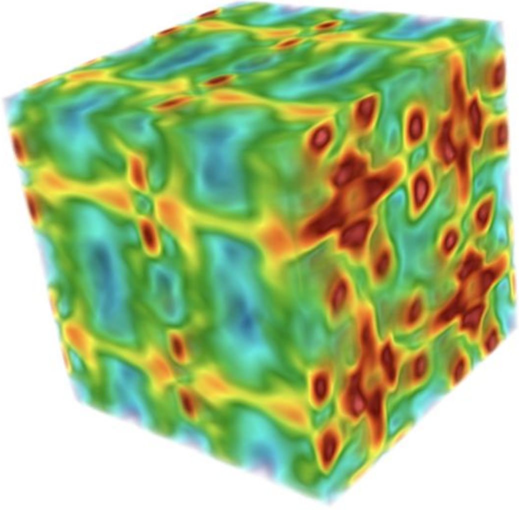
Taylor-Green Vortex (Mean T=1)



DNS



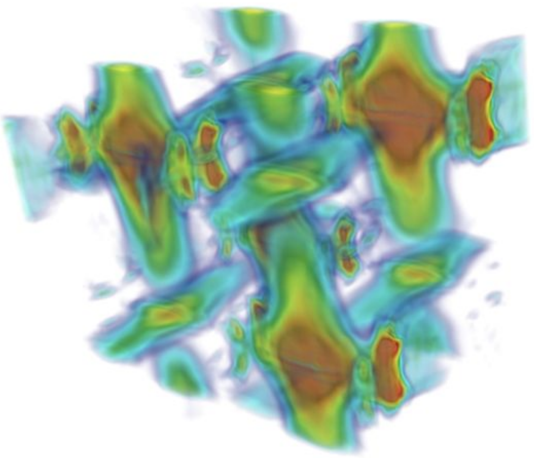
GenCFD



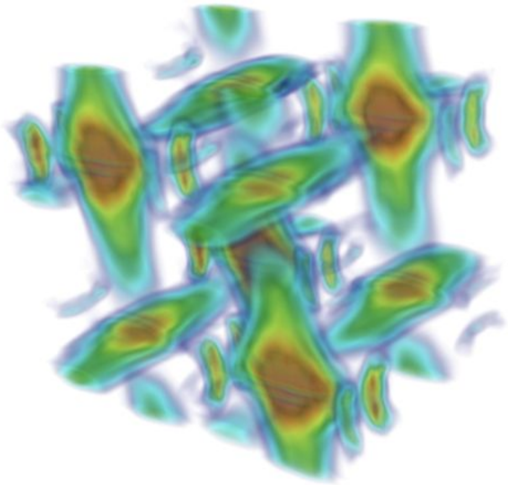
FNO

Small deviation of the mean!

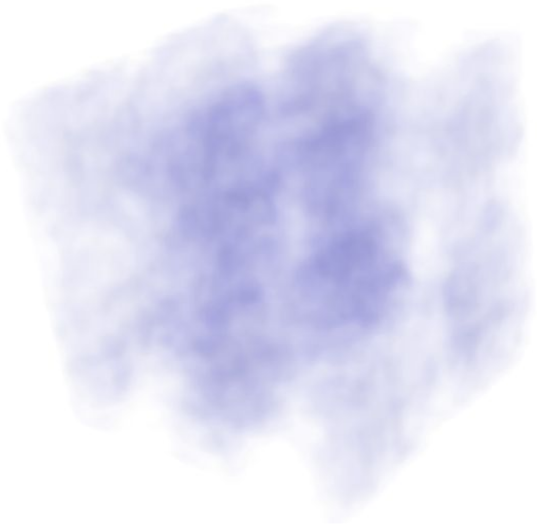
Taylor-Green Vortex (Std T=1)



DNS

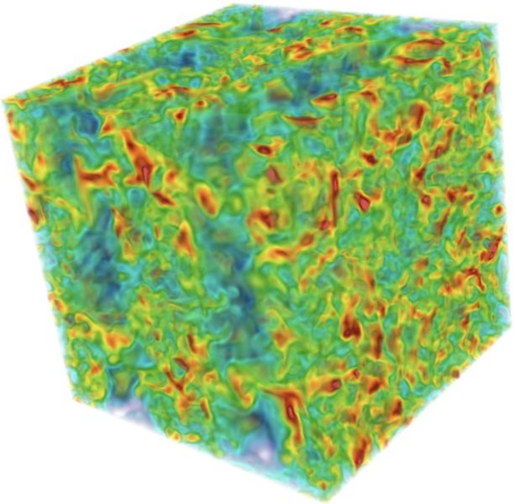


GenCFD

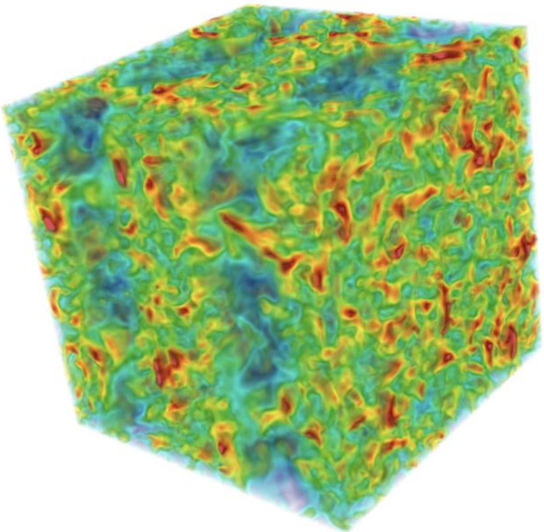


FNO

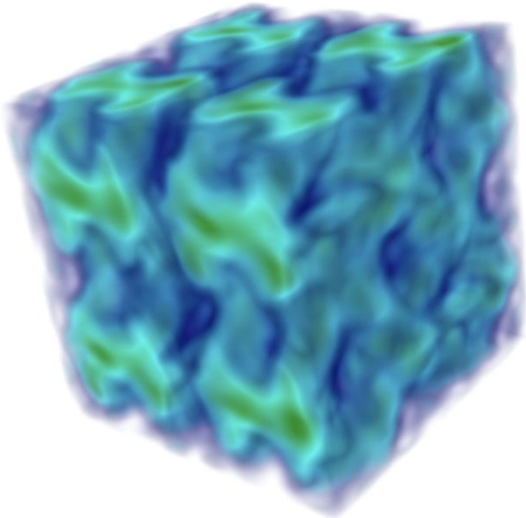
Taylor-Green Vortex (samples T=2)



DNS

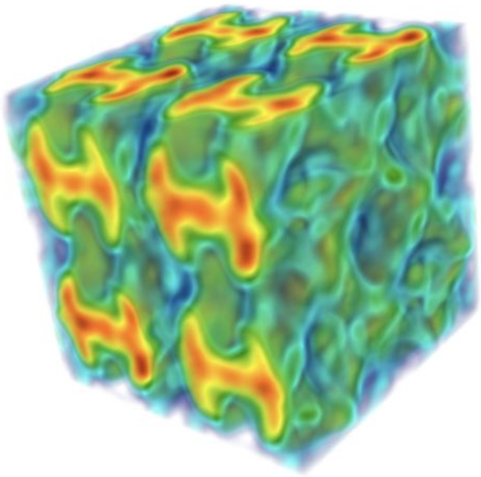


GenCFD

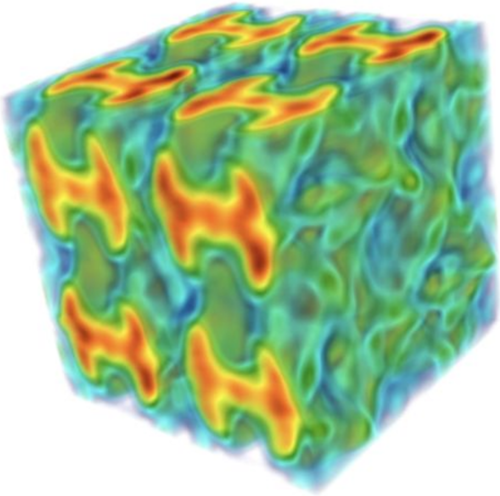


FNO

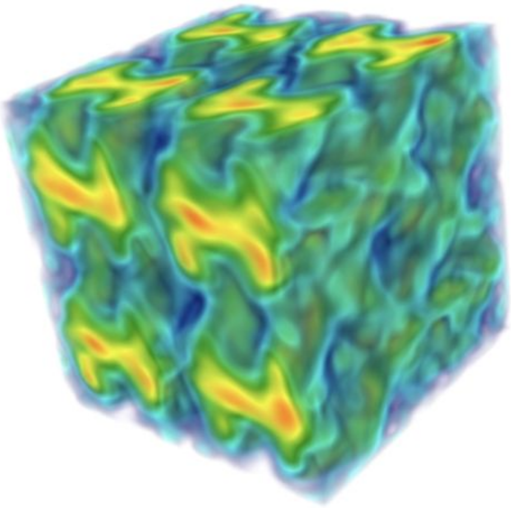
Taylor-Green Vortex (Mean)



DNS

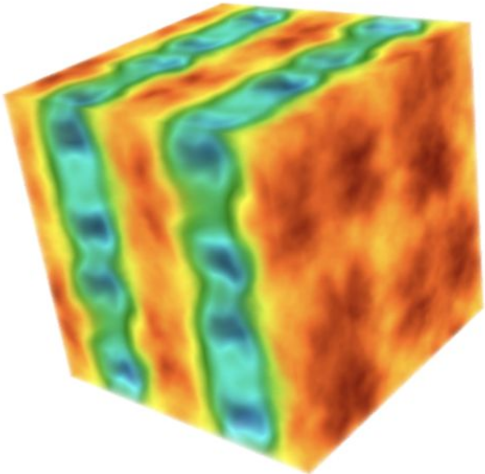


GenCFD

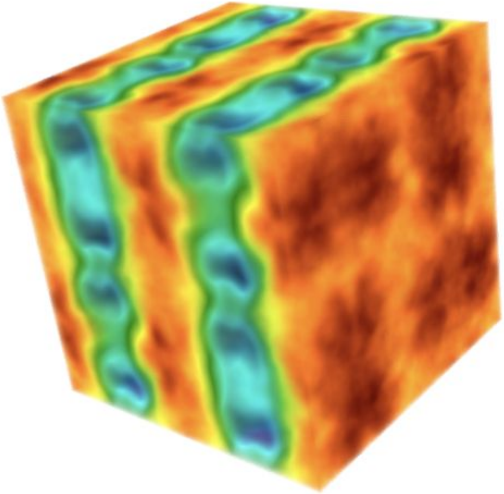


FNO

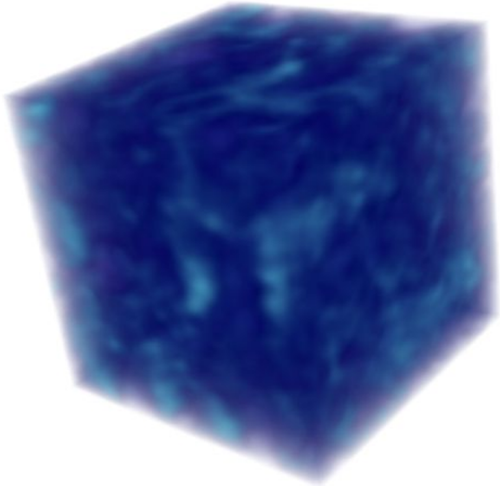
Taylor-Green Vortex (std)



DNS

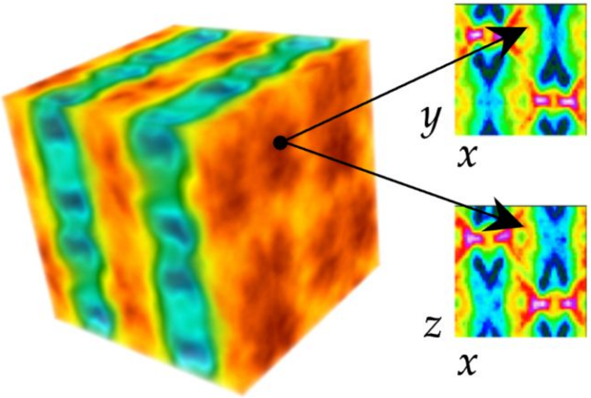


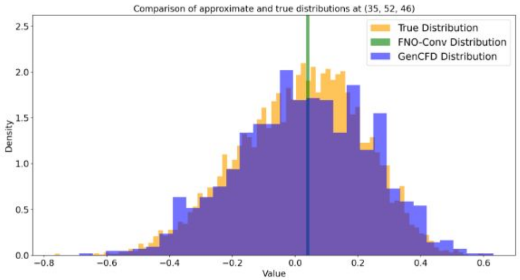
GenCFD

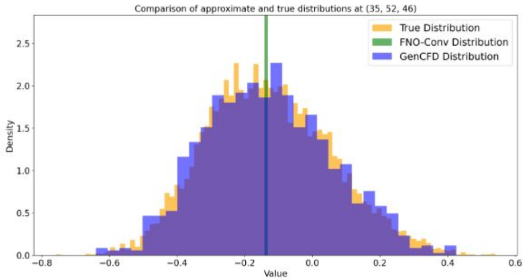


FNO

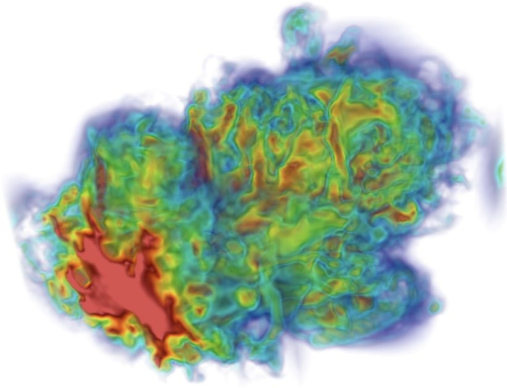
Taylor-Green Vortex (PDF)



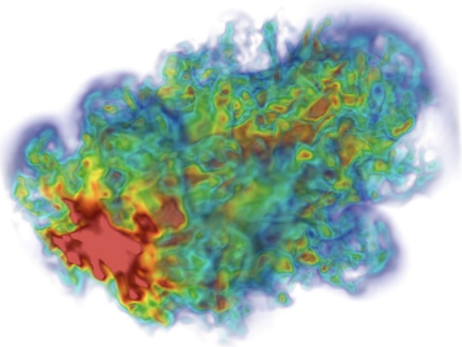




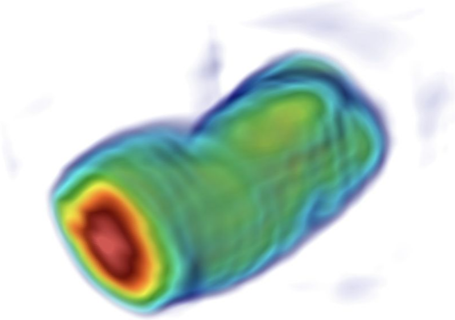
Cylinder Shear Layer (Samples)



DNS

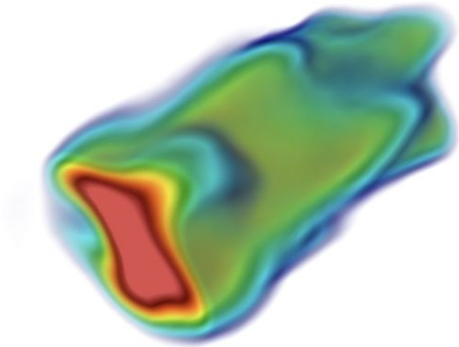


GenCFD

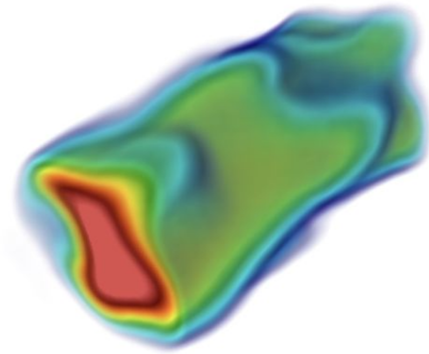


FNO

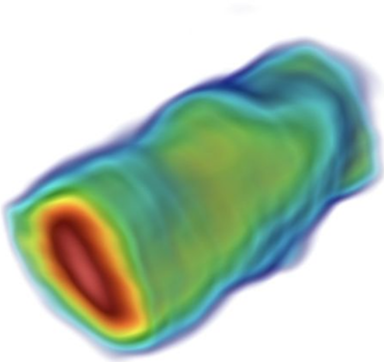
Cylinder Shear Layer (Mean)



DNS

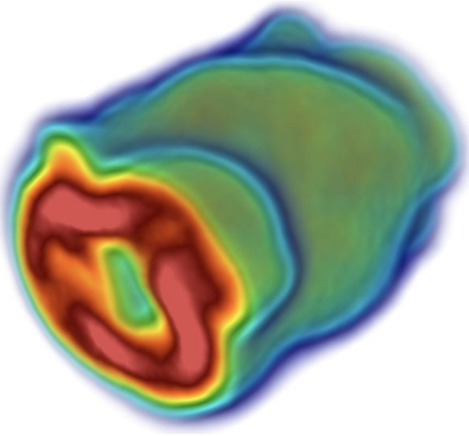


GenCFD

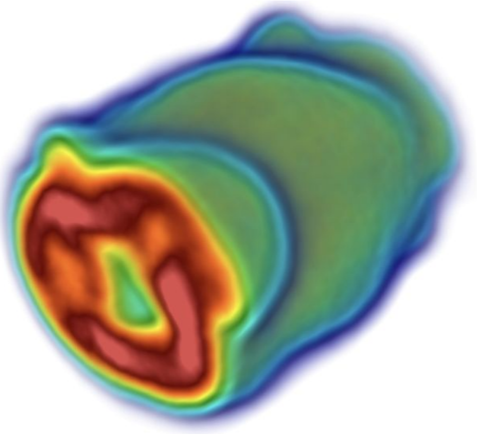


FNO

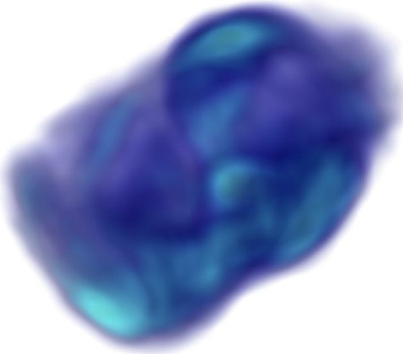
Cylinder Shear Layer (Std)



DNS

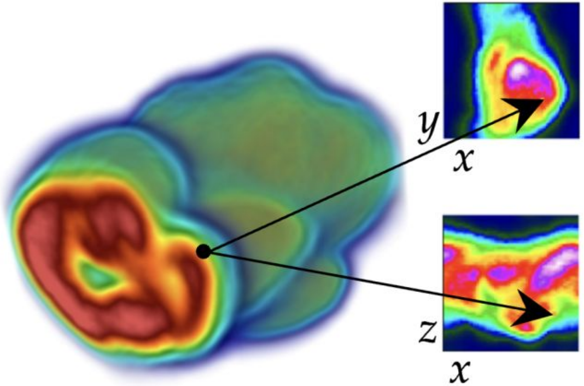


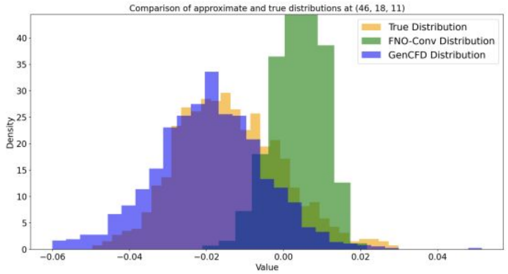
GenCFD

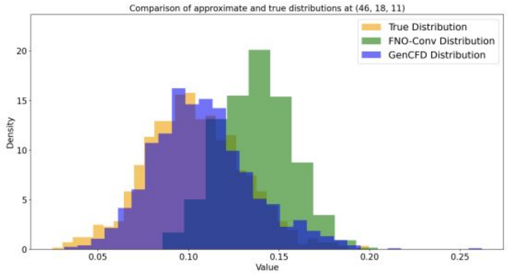


FNO

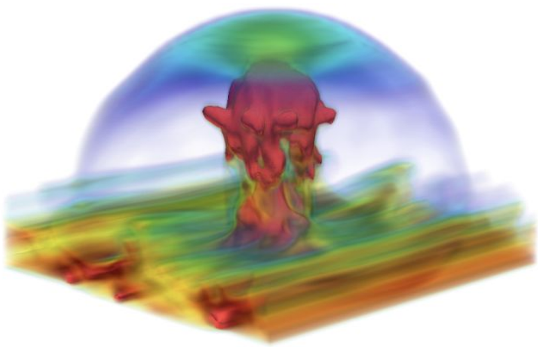
Cylinder Shear Layer (pdf)



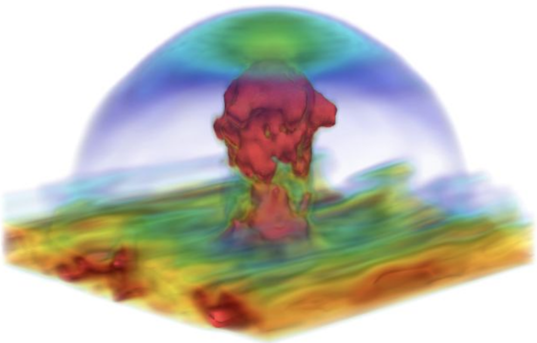




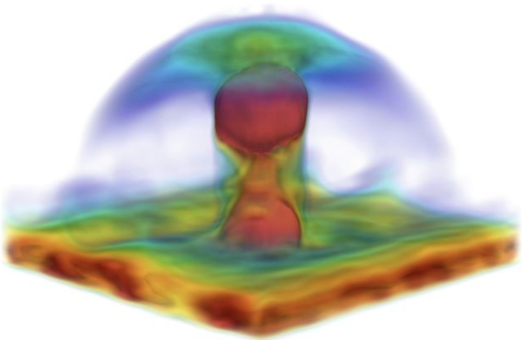
Cloud Shock Interaction Problem (samples)



DNS

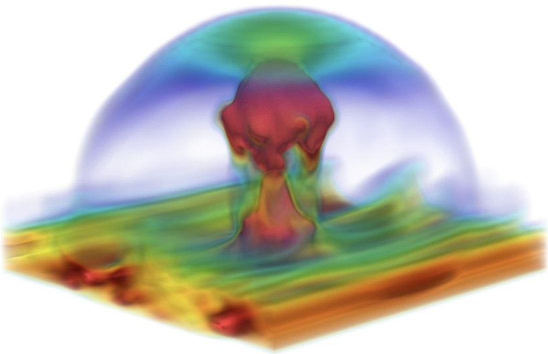


GenCFD

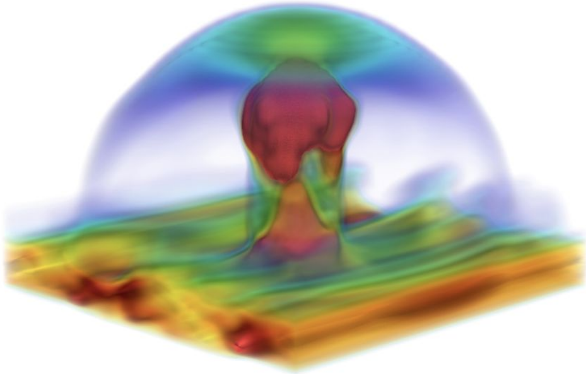


FNO

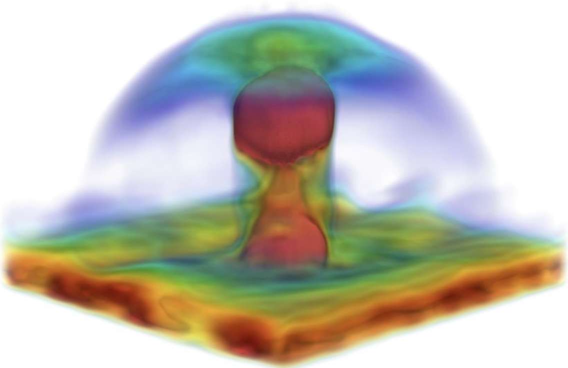
Cloud Shock Interaction Problem (mean)



DNS

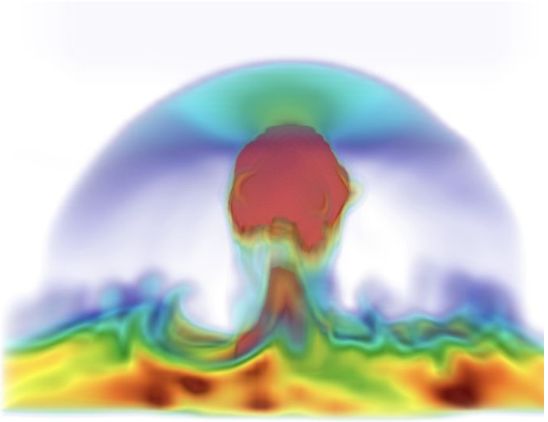


GenCFD

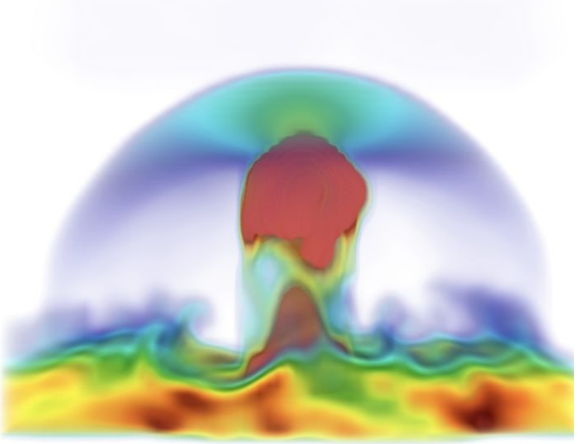


FNO

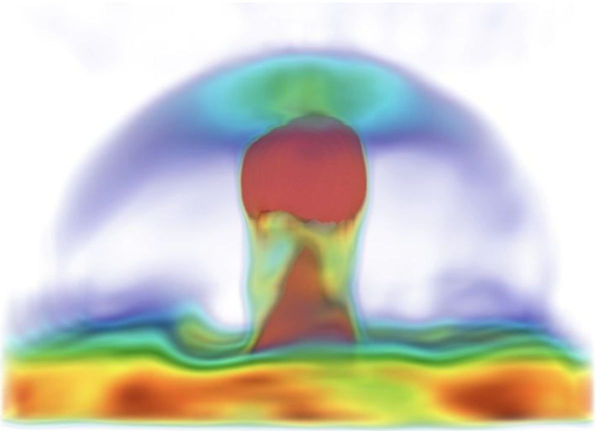
Cloud Shock Interaction Problem (Mean)



DNS

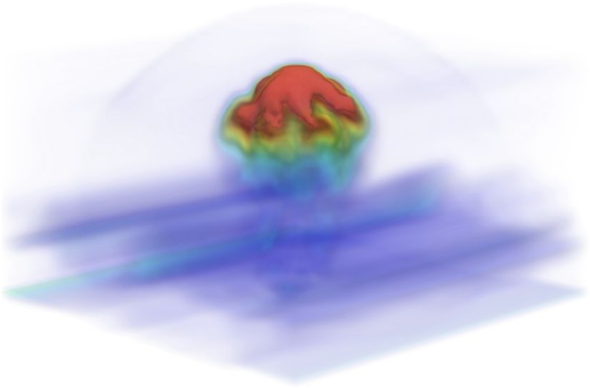


GenCFD

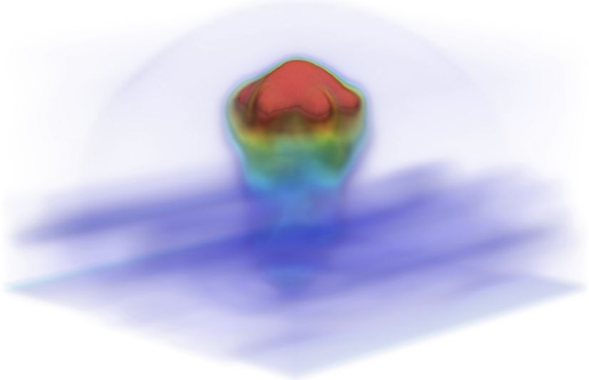


FNO

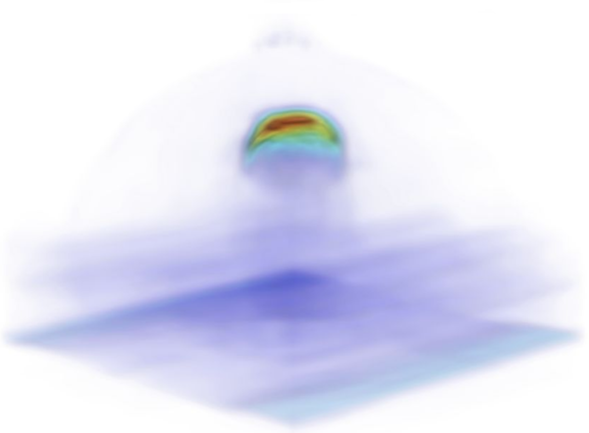
Cloud Shock Interaction Problem (Std)



DNS

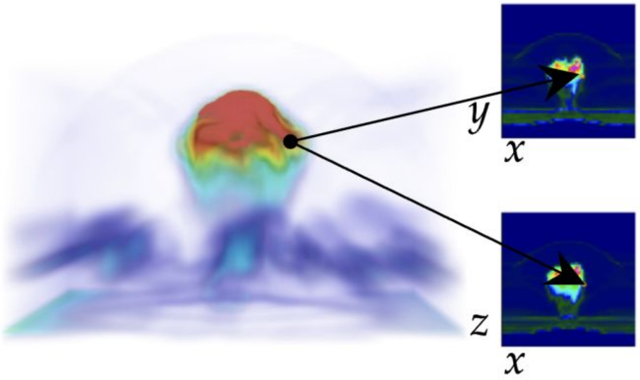


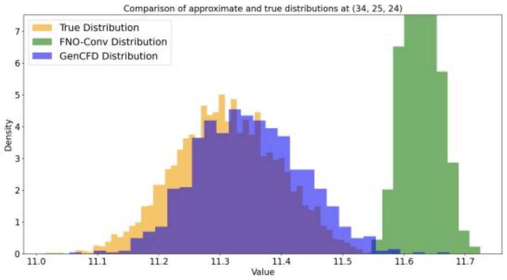
GenCFD

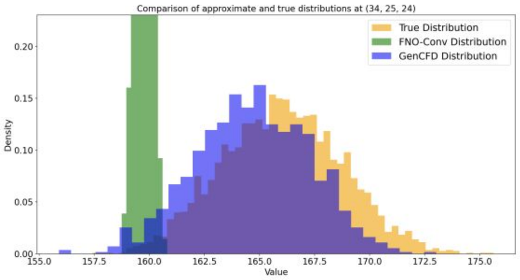


FNO

Cloud Shock Interaction Problem (std)

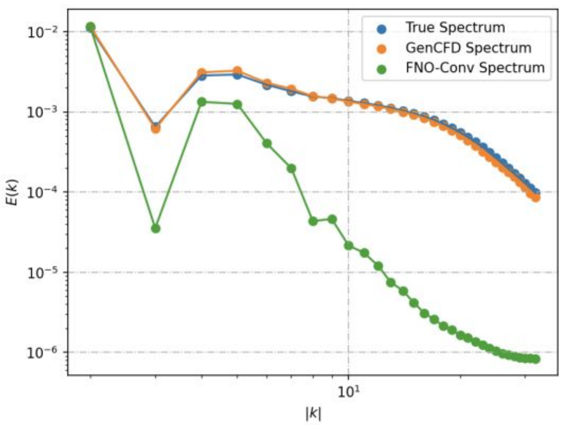




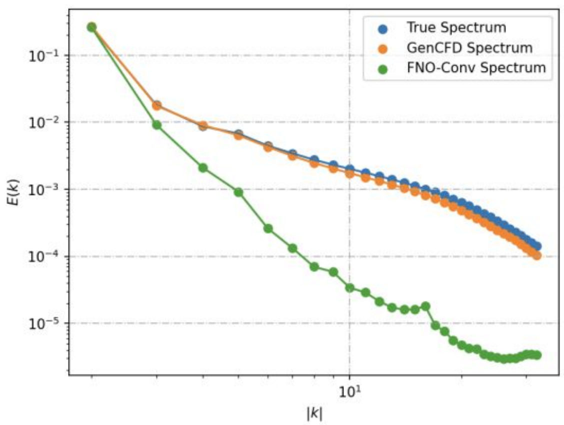


Google Research

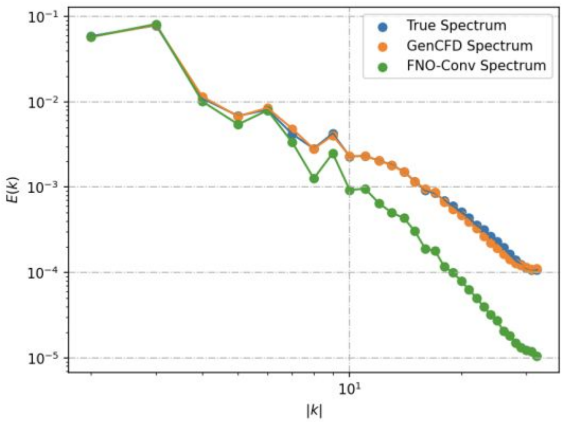
Spectrum



(a) Taylor-Green



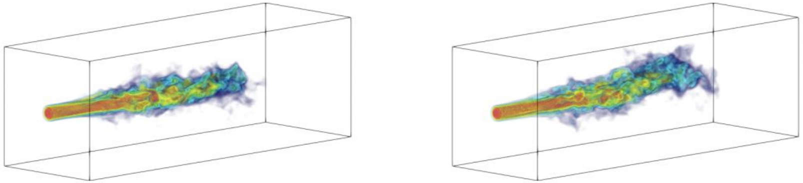
(b) Cylindrical Shear Flow



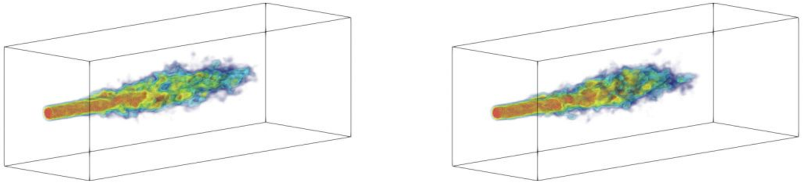
(c) 3D Cloud-Shock Interaction

Turbulent Jet

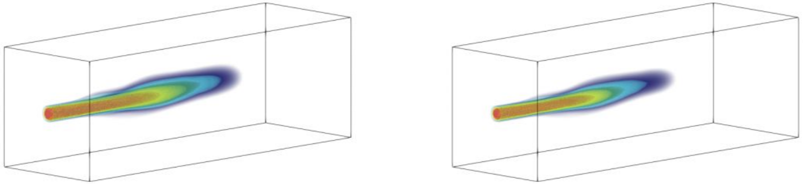
Ground Truth



GenCFD



UViT

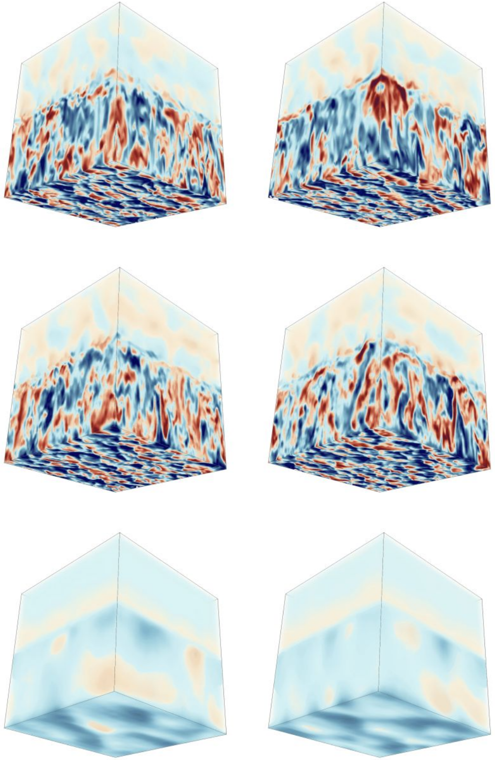


Convective Boundary Layer

Ground Truth

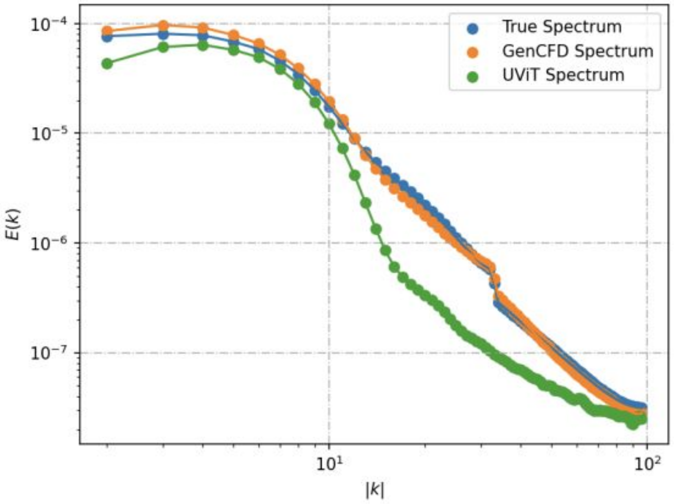
GenCFD

UViT

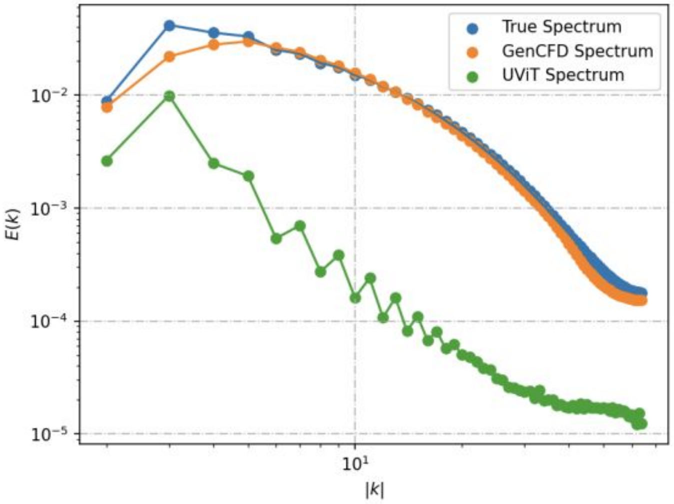


Google Research

Spectrum



(c) Nozzle flow



(d) Convective boundary layer

Conclusions

Lessons learned:

- Good software enables good research
- Stochastic description renders the learning easier
- Trading high-dimension by smoothness

Thank you!