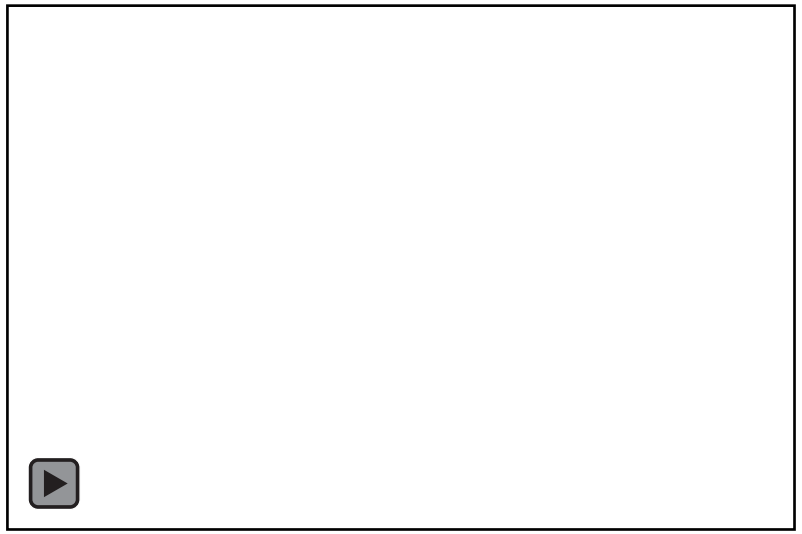
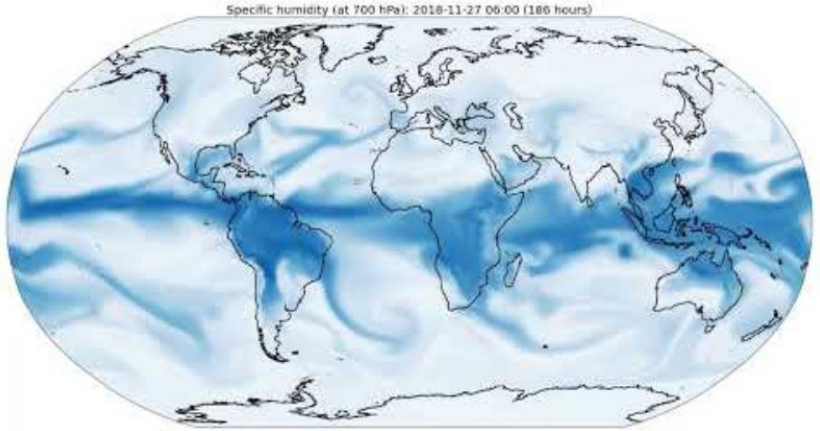
Data-driven Whitney forms for structure preserving models

Nat Trask
University of Pennsylvania

Autoregressive models in state of the art AI4Science



TheWell



GraphCast: AI model for faster and more accurate global weather forecasting

Dominant strategy: Physics-agnostic next token prediction of field data

Goal:

Construct a physics+geometry-based framework to guarantee
Physics + Uniform Stability

Objective and “Physics-by-Construction” philosophy

Model acceleration

Learn a cheap low-dimensional problem trained to give the same answer as an expensive one to get real-time models

Model discovery

Governing physics are too complex to derive by pencil and paper, so use experiment instead (multiscale, multiphysics, epidemiology)

Physics-informed

Multi-objective optimization

Numerical properties driven by optimization error

$$\mathcal{L} = ||u - u_{\text{data}}||^2 + ||\mathbf{L}[u] - b]||^2$$

vs.

Strong physics

Equality constrained optimization

Physics in the architecture

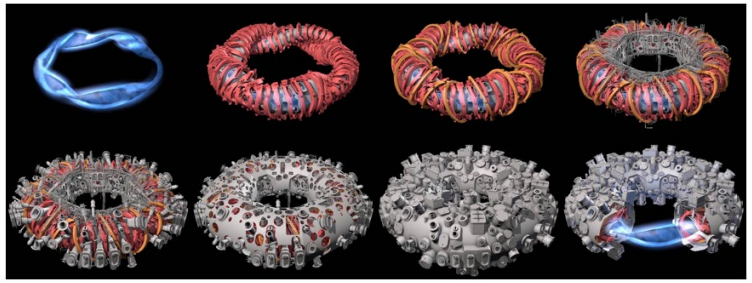
Physics independent of sample complexity or optimizer error

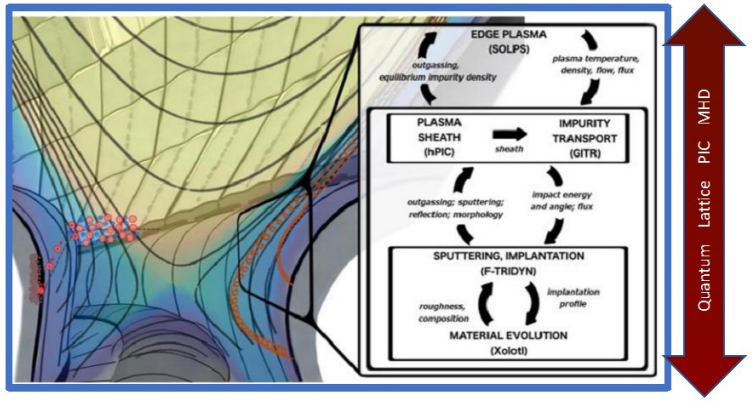
Comparably heavier software lift

$$\mathcal{L} = ||u - u_{\text{data}}||^2$$

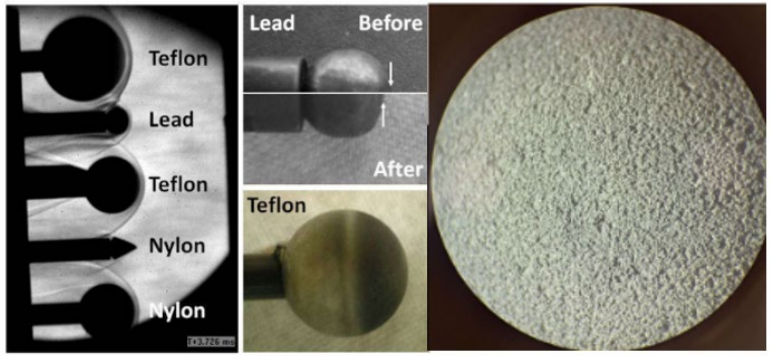
$$\textbf{such that } \mathbf{L}[u] = b$$

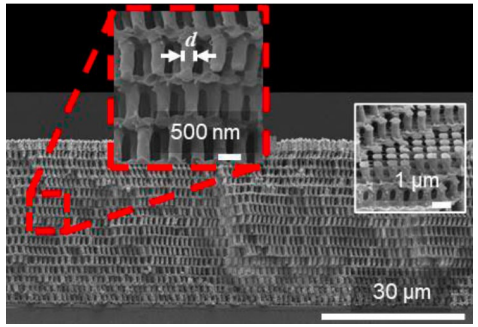
Structure-preservation requirements for data-driven models



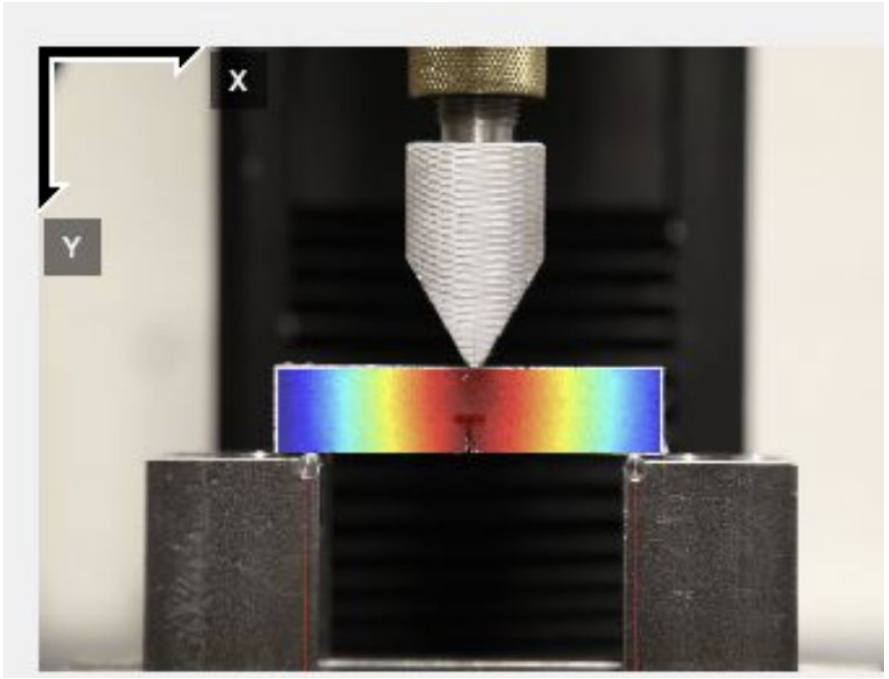


Fusion reactor optimization
Commercialization of complex geometry reactor designs
Need exact handoff of conserved fluxes, Hugoniot relations, Gauge symmetries





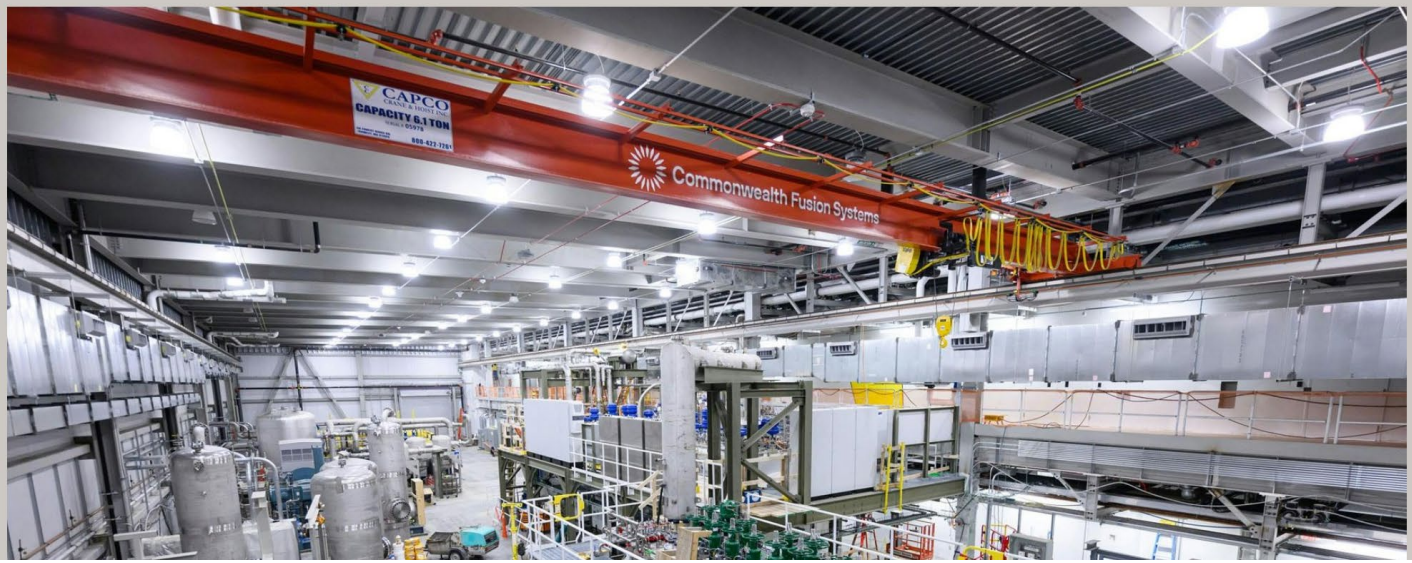
Hypersonic metamaterial design
Treat non-equilibrium physics in multiscale extreme physics processes
Need exact mass transfer, non-equilibrium chemistry, fluctuation/dissipation balance



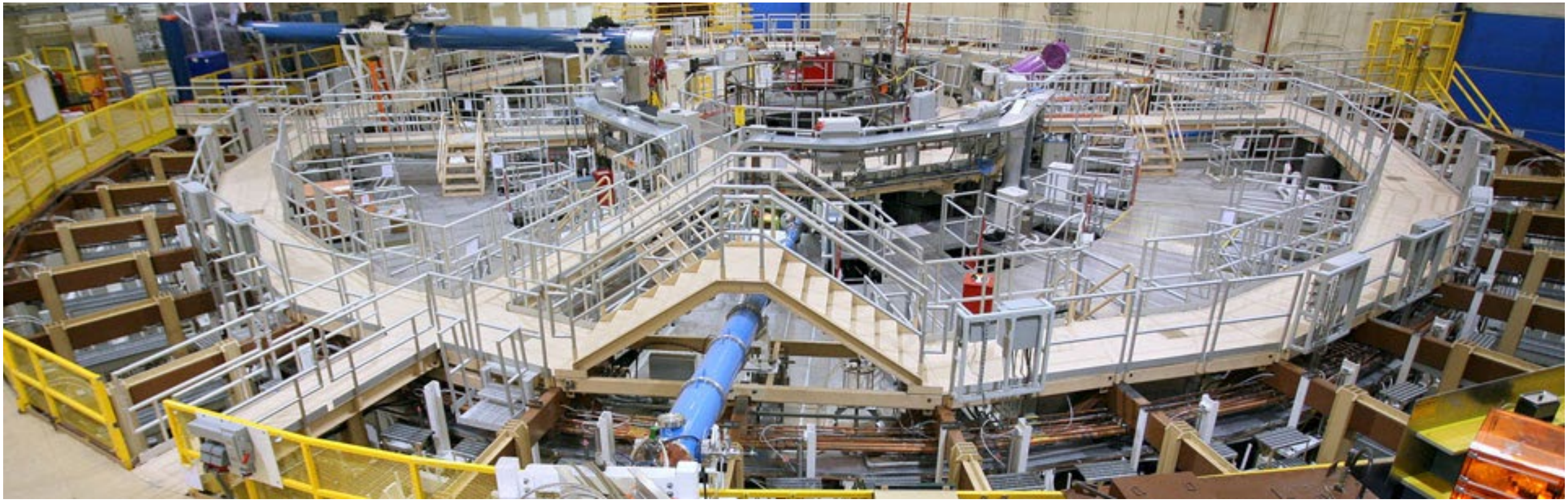
Data-driven fracture
Homogenized mechanics for multiscale materials from experimental DIC
Exactly balance strain, Griffiths fracture energy for rare events conditioned on mesoscale geometry

08.28.2025

Commonwealth Fusion Systems Raises $863 Million Series B2 Round to Accelerate the Commercialization of Fusion Energy



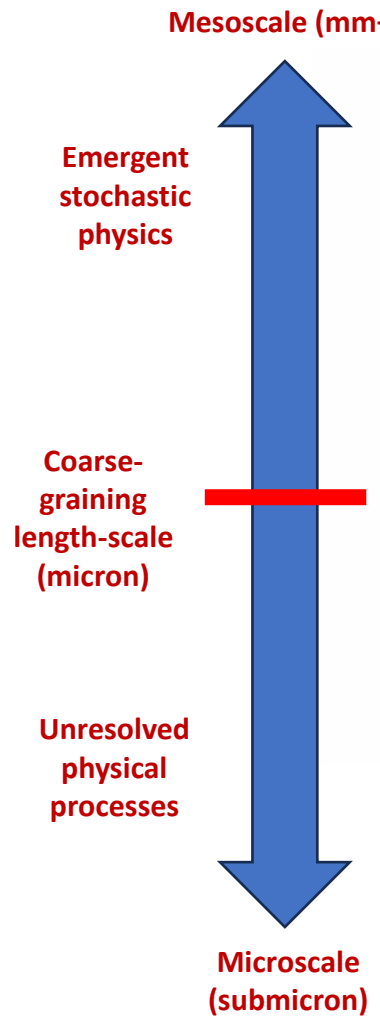
The Z-pinch pulsed power fusion facility at SNL

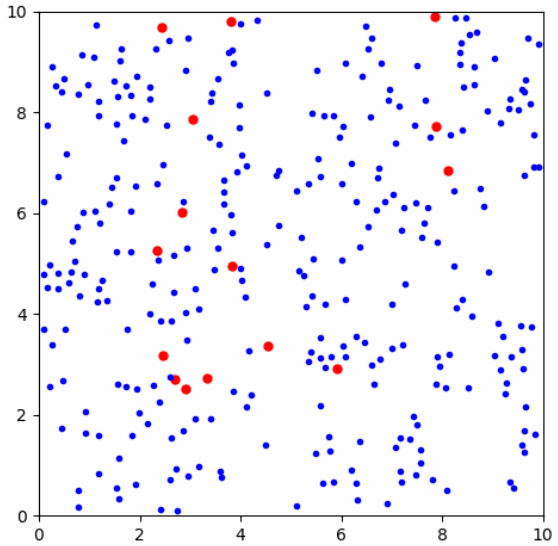


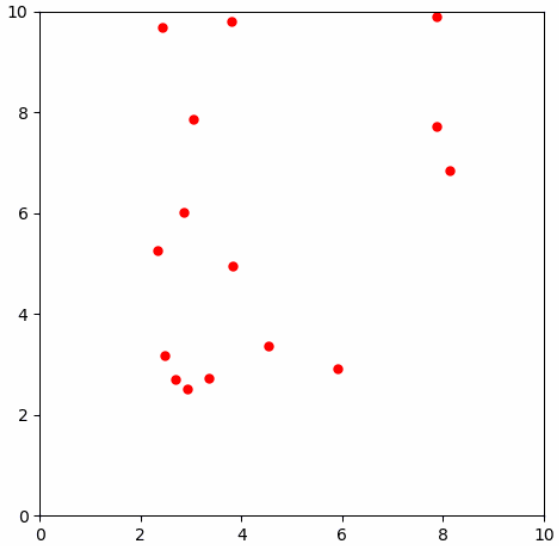
Z-Machine at SNL
Pulsed power fusion facility for generating extreme environments

A “gun” for harvesting electrons to bombard materials under test

Emergent stochastic physics from coarse-grained dynamics







$$\mathrm{d}\boldsymbol{x}_t = \left(L\frac{\partial E}{\partial \boldsymbol{x}} + M\frac{\partial S}{\partial \boldsymbol{x}} + k_B \frac{\partial}{\partial \boldsymbol{x}} \cdot M \right) \mathrm{d}t + \sqrt{2k_B M}\,\mathrm{d}W_t$$

Reversible Irreversible dissipation Thermal noise

Needs geometric dynamics beyond exterior calculus stuff I'll talk about today

QR code to preprint

Tool 1

Whitney forms

Tool 2

Hybridizable domain decomposition

Tool 3

Conditional neural operators through cross-attention transformers

Model Class 1

Nonlinear boundary value problems

Model Class 2

Uniformly stable autoregressive dynamics

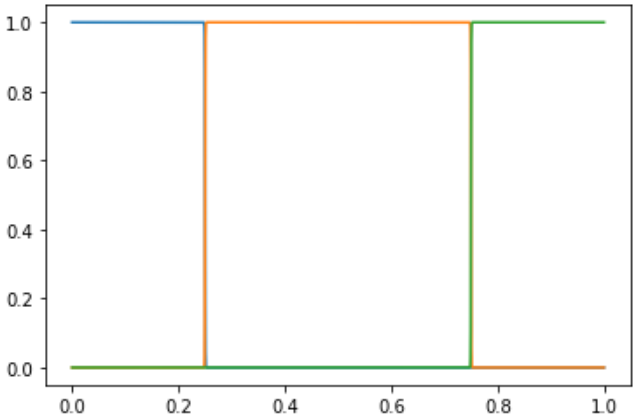
Mathematical preliminaries: partition of unity

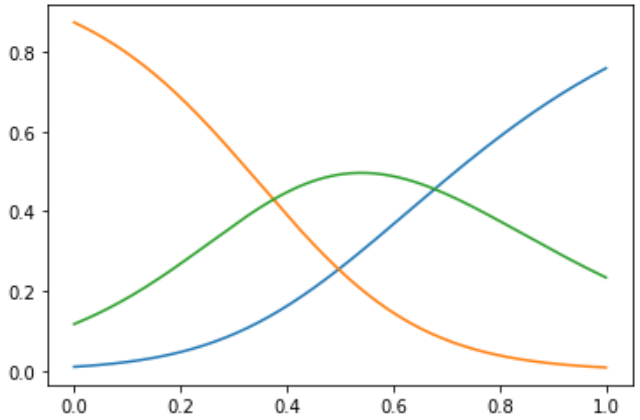
Definition: *Partition of unity (POU)*
A collection of functions $\{\phi_i\}_{i=1,\dots,N}$ satisfying

- $\phi_i > 0$
- $\sum_i \phi_i = 1$

Example:
Consider a partition of $\Omega \subset \mathbb{R}^d$ into disjoint cells $\Omega = \bigcup_i C_i$. Then the indicator functions $\phi_i(x) = \mathbb{1}_{C_i}(x)$ form a POU.

$$\phi_i(x) = \text{softmax} \circ NN(x;\theta)$$





Softmax activation function

$$\frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

POU corresponding to Cartesian mesh vs categorical embedding architecture for logistic classification

Mathematical preliminaries: partition of unity

Theorem: Partitions of unity are closed under linear convex combinations

Let $(\phi_1, \ldots, \phi_n)$ be a partition of unity:

$$\sum_{i=1}^{n} \phi_i(x) = 1, \quad \phi_i(x) \geq 0.$$

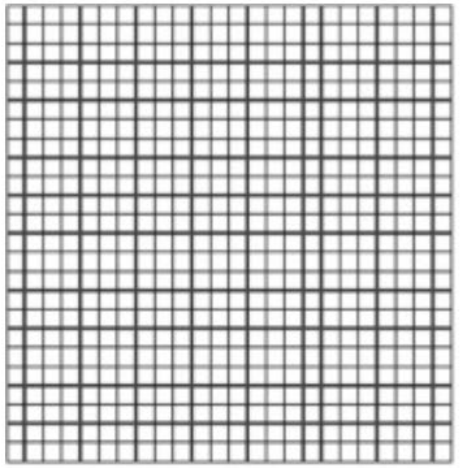
Let $W = (w_{ij})$ be an $m \times n$ matrix with $w_{ij} \geq 0, \sum_{j=1}^{n} w_{ij} = 1.$

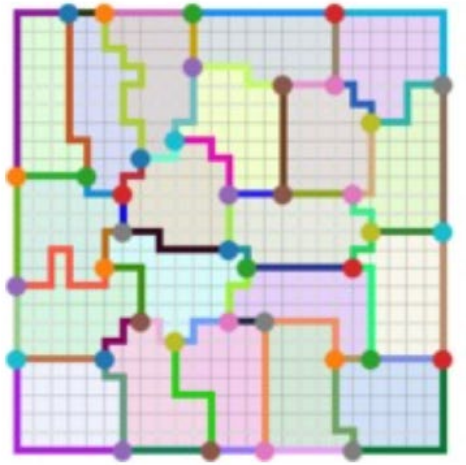
Define new functions:

$$\psi_i(x) = \sum_{j=1}^{n} w_{ij}\phi_j(x)$$

Then:

$$\sum_{i=1}^{n} \psi_i(x) = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\phi_j(x) = \sum_{j=1}^{n} \phi_j(x) \left(\sum_{i=1}^{m} w_{ij} \right) = \sum_{j=1}^{n} \phi_j(x) = 1$$





Tool one: Data-driven Whitney forms

$$\begin{array}{ccccccc} H(grad) & \xrightarrow{\nabla} & H(curl) & \xrightarrow{\nabla\times} & H(div) & \xrightarrow{\nabla\cdot} & L^2 \\ \uparrow & & \uparrow & & \uparrow & & \uparrow \\ \mathcal{W}_0 & \xrightarrow{d_0} & \mathcal{W}_1 & \xrightarrow{d_1} & \mathcal{W}_2 & \xrightarrow{d_2} & \mathcal{W}_3 \end{array} .$$

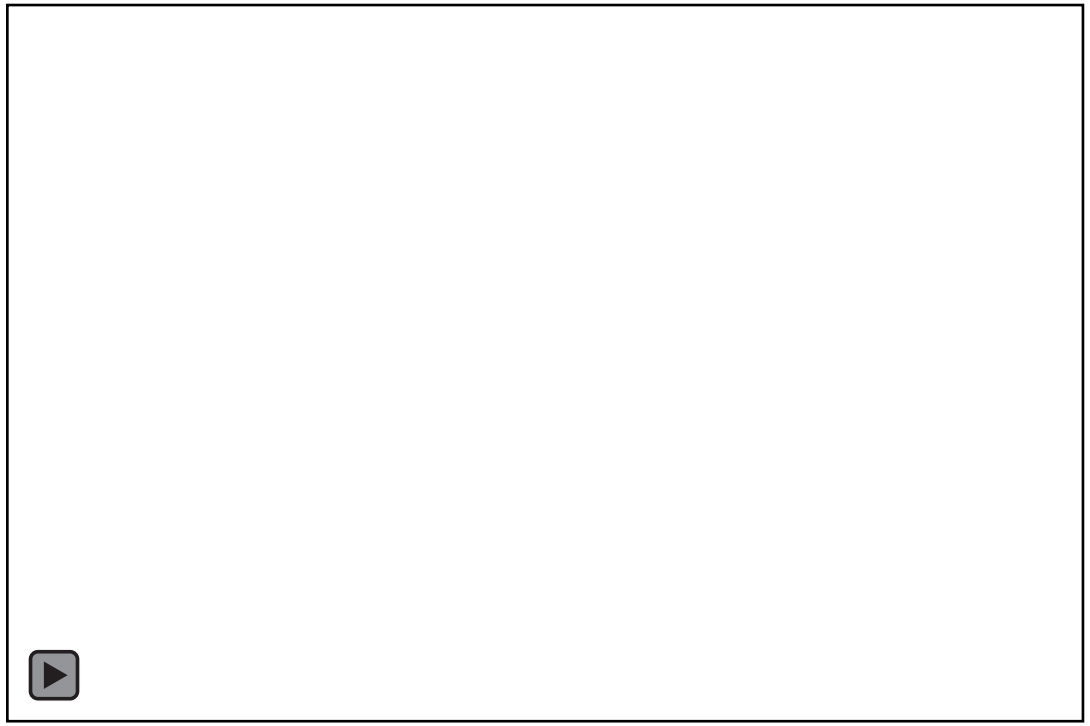
$$\mathcal{W}_{j_0\cdots j_k} = k! \sum_{i=0}^{k} (-1)^i \lambda_{j_i} \mathsf{d}\lambda_{j_0} \wedge \cdots \widehat{\mathsf{d}\lambda_{j_i}} \wedge \cdots \wedge \mathsf{d}\lambda_{j_k}$$

$$\mathcal{W}_0 = \text{span}\{\lambda_i\},$$

$$\mathcal{W}_1 = \text{span}\{\lambda_i \nabla\lambda_j - \lambda_j \nabla\lambda_i\},$$

$$\mathcal{W}_2 = \text{span}\{\lambda_i \nabla\lambda_j \times \nabla\lambda_k + \lambda_j \nabla\lambda_k \times \nabla\lambda_i + \lambda_k \nabla\lambda_i \times \nabla\lambda_j\}$$

$$\mathcal{W}_3 = \text{span}\{\lambda_i \nabla\lambda_j \cdot (\nabla\lambda_k \times \nabla\lambda_l) - \lambda_j \nabla\lambda_i \cdot (\nabla\lambda_k \times \nabla\lambda_l) + \lambda_k \nabla\lambda_i \cdot (\nabla\lambda_j \times \nabla\lambda_l) - \lambda_l \nabla\lambda_i \cdot (\nabla\lambda_j \times \nabla\lambda_k)\}$$



Red: POU on cells/0-forms
Blue: Boundary of POUS/1-forms

$$\mathcal{W}_i = \lambda_i$$

In limit of disjoint partitions, we recover indicator functions and Dirac measures

$$\mathcal{W}_{ij} = \lambda_i \nabla\lambda_j - \lambda_j \nabla\lambda_i$$

Use ML to learn convex combination W and obtain data-driven de Rham complex

Candidate model form for learning well-posed boundary value problems

Theorem 2.1 (Gustafsson [39]). *Consider the nonlinear system of equations*

$$\mathbf{A}x + \epsilon \mathbf{F}(x) = b, \tag{7}$$

where $\epsilon > 0$ *and* $\mathbf{F}$ *is a vector-valued nonlinear function with Lipschitz constant* C_L

$$||\mathbf{F}(x) - \mathbf{F}(y)||_2 \leq C_L ||x - y||_2.$$

Define $\tau = \epsilon C_L ||A^{-1}||$. *If* $\tau < 1$, *then* (7) *has a unique solution.*

Proof. See [39], Appendix A.3. □

Corollary 2.2. *Assume* $\mathbf{A}$ *is invertible and satisfies the Poincare-like inequality*

$$||x||_2^2 \leq C_p x^\intercal \mathbf{A} x \tag{8}$$

then $\tau \leq \epsilon C_p C_L$, *and following Theorem 2.1,* (7) *has a solution if* $\epsilon C_p C_L < 1$.

Proof. See Horn and Johnson [43] for a discussion of Loewner ordering and monotonicity of the matrix inverse. By Loewner ordering,

$$x^\intercal x \leq C_p x^\intercal \mathbf{A} x \implies x^\intercal \mathbf{A}^{-1} x \leq C_p x^\intercal x,$$

and we can bound $||\mathbf{A}^{-1}|| \leq C_p$. □

FEEC provides building blocks to construct our linear “anchor”

Theorem 2.3. *Denote by* $\delta_{i,jk} = \delta_{ij} - \delta_{ik}$ *the graph gradient operator mapping* $\mathbb{R}^N \to \mathbb{R}^{N\times N}$, *and* $\mathbf{M}_i$ *the mass matrix associated with* $\mathcal{W}_i$. *Then the following identities hold, for* $q, u \in \mathcal{W}_0$ *and* $v, J \in \mathcal{W}_1$.

$$(v, \nabla u) = \hat{v}^\intercal \mathbf{M}_1 \delta \hat{u}$$

$$-(J, \nabla q) = \hat{q}^\intercal \delta^\intercal \mathbf{M}_1 \hat{J}$$

Consider the Poisson equation in mixed form, namely, find $(u, J) \in \mathcal{W}_0 \times \mathcal{W}_1$ *such that for any* $(q, v) \in \mathcal{W}_0 \times \mathcal{W}_1$

$$(J, v) + (\nabla u, v) = 0$$

$$(J, \nabla q) = < J, q > + (f, q).$$

If $J \cdot \hat{n}|_{\partial\Omega} = 0$, *the resulting system of equations*

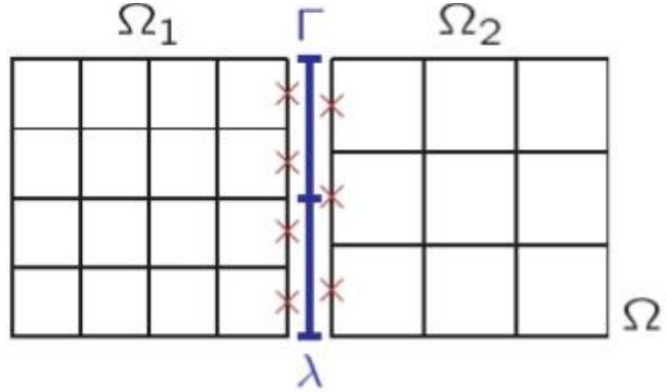
$$\mathbf{L}\hat{u} := \delta^\intercal \mathbf{M}_1 \delta \hat{u} = \mathbf{M}_0 \hat{f}$$

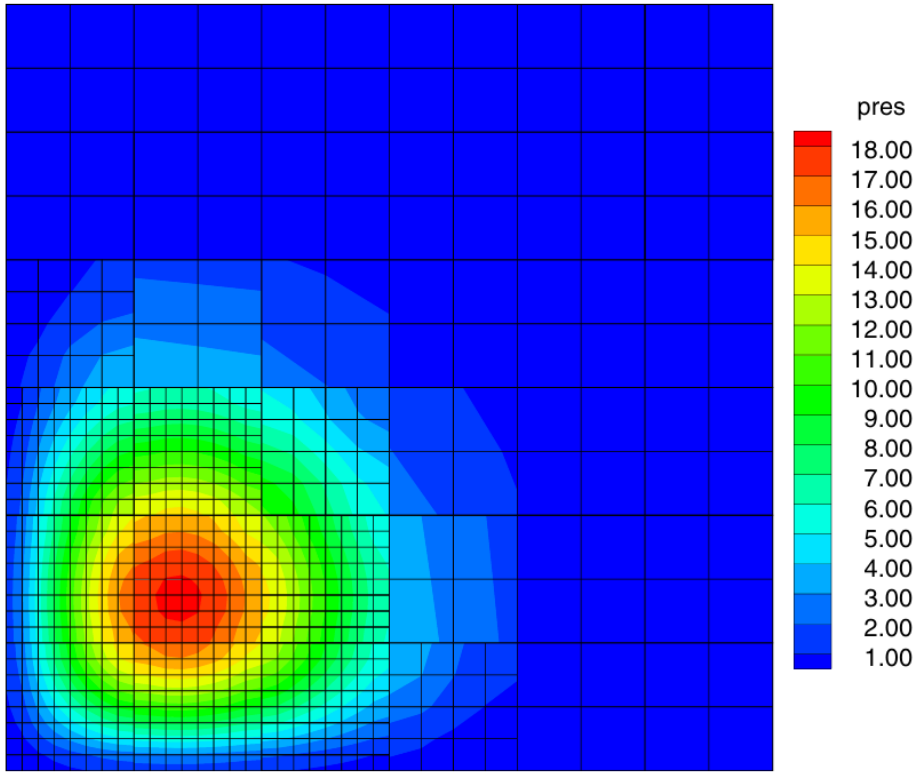
has a unique solution modulo a constant vector in the null-space, and the stiffness matrix $\mathbf{L}$ *corresponding to the discretized Hodge Laplacian is symmetric positive definite with a Poincare inequality*

$$\hat{u}^\intercal \mathbf{M}_0 \hat{u} \leq C_p(h) \hat{u}^\intercal \mathbf{L} \hat{u}.$$

Tool two: Mortar methods

$$\mathbf{u} + K\nabla p = 0$$
$$\nabla \cdot \mathbf{u} = f$$





$$(K^{-1}\mathbf{u}_h, \mathbf{v})_{\Omega_i} - (p_h, \nabla \cdot \mathbf{v})_{\Omega_i} + \langle \lambda_h, \mathbf{v} \cdot \mathbf{n}_i \rangle_{\Gamma_i} = 0 \quad \forall \mathbf{v} \in \mathbf{V}_h(\Omega_i),$$
$$(\nabla \cdot \mathbf{u}_h, q)_{\Omega_i} = (f, q)_{\Omega_i} \quad \forall q \in Q_h(\Omega_i),$$
$$\sum_i \langle \mathbf{u}_h \cdot \mathbf{n}_i, \mu \rangle_{\Gamma_i} = 0 \quad \forall \mu \in \Lambda_h,$$

- $\mathbf{V}_h(\Omega_i)$ is the Raviart–Thomas space on Ω_i,
- $Q_h(\Omega_i)$ is the space of piecewise constants on Ω_i,
- Λ_h is the mortar space on interfaces (typically piecewise polynomials),
- $\mathbf{n}_i$ is the outward unit normal to $\partial\Omega_i$.

Michel Bernadou and Pierre-Arnaud Raviart, An analysis of some mortar finite element methods, RAIRO Analyse Numérique, 1976.
Arbogast, Todd, et al. "A multiscale mortar mixed finite element method." Multiscale Modeling & Simulation 6.1 (2007): 319-346.

New idea: A fully implicit family of nonlinear dynamics with a mortar shooting method

$$\mathbf{V}_h^i = \mathcal{Q}_h^i \otimes \mathcal{V}_h^i \otimes \mathcal{M}_h^i, \text{ where } \mathcal{Q}_h^i \subseteq \mathbf{L}^2(\Omega_i),\ \mathcal{V}_h^i \subseteq \mathbf{H}^1(\Omega_i), \text{ and } \mathcal{M}_h^i \subset \mathbf{L}^2(\partial\Omega_i)$$

$$(J, v) + (u, \frac{d}{dt}v) = \lambda_{i+1}v_{i+1} - \lambda_i v_i$$

$$(\frac{d}{dt}J, q) - (\mathcal{N}[u, J], q) = 0$$

$$\lambda_i = u_i^0$$

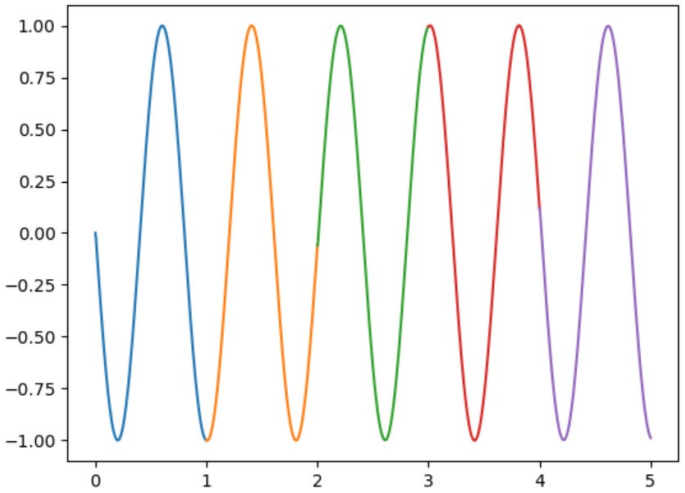
$$J(t_i) = J_i^0,$$

An example rollout

Selecting N to recover the harmonic oscillator

$$u_t = J,$$

$$J_t = -\omega^2 u$$



The mortar method naturally generates summation by parts structures

Summation by parts
A discrete counterpart to integration-by-parts generating a telescoping sum

$$\sum_{i=0}^{N} \alpha_{i+1} - \alpha_i = \alpha_N - \alpha_0$$

Lemma 2.2 (Discrete summation by parts formula). *For piecewise constant $u \in \mathcal{M}_h^i$ and continuous P1 $J, v \in \mathcal{V}_h^i$ on the interval $[t_k, t_{k+1}]$,*

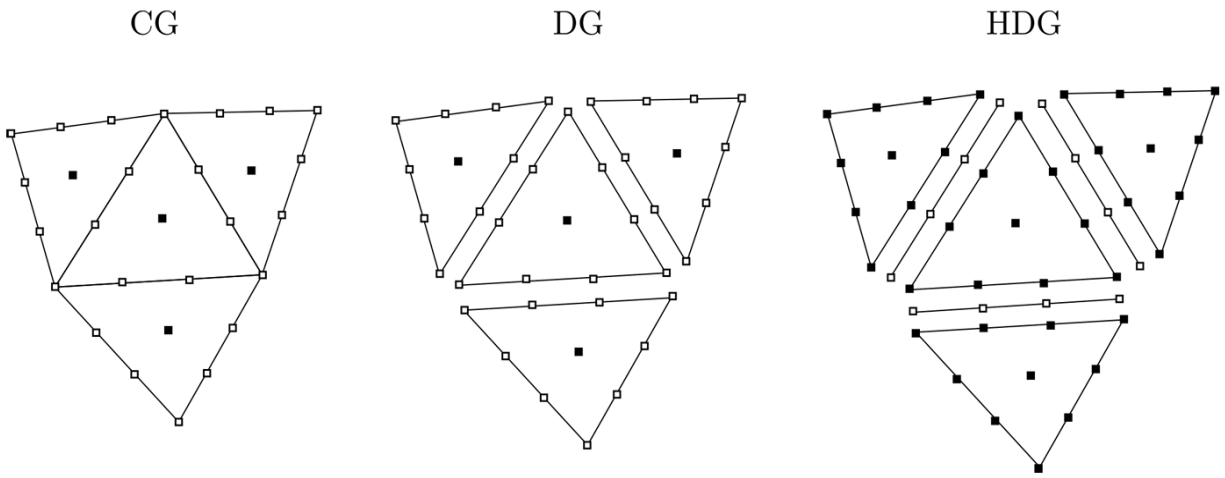
$$\int_{t_k}^{t_{k+1}} Jv\,dt + u_k\big(v(t_{k+1}) - v(t_k)\big) = u_{k+1}v(t_{k+1}) - u_k v(t_k). \tag{4}$$

Summing these identities over all fine elements $k = 0, \ldots, M-1$ in a given subdomain Ω_i yields

$$\sum_{k=0}^{M-1} \left[\int_{t_k}^{t_{k+1}} Jv\,dt + u_k\big(v(t_{k+1}) - v(t_k)\big)\right] = u(t_{i+1}^-)v(t_{i+1}^-) - u(t_i^+)v(t_i^+). \tag{5}$$

Will allow us to consistently transfer energy across domains

Mortar methods are an example of a *hybridizable scheme*



$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} u \\ \lambda \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

$$S\lambda := BA^{-1}B^T\lambda = BA^{-1}f - g$$

$$u = A^{-1}(f - B^T\lambda)$$

Unified Hybridization of Discontinuous Galerkin, Mixed, and Continuous Galerkin Methods for Second Order Elliptic Problems

Authors: Bernardo Cockburn, Jayadeep Gopalakrishnan, and Raytcho Lazarov | AUTHORS INFO & AFFILIATIONS

Traditionally useful for hardware utilization
In machine learning contexts, we don't need to propagate through the entire time series

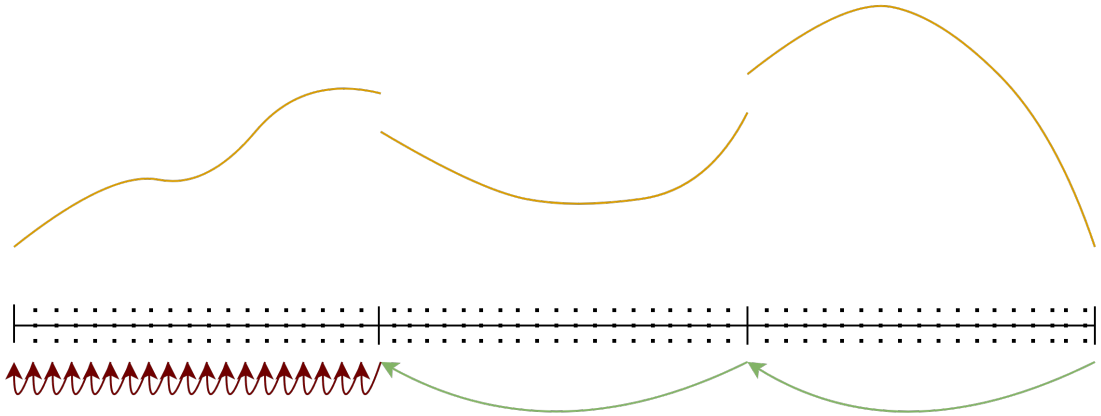


Figure taken from Guosheng Fu's slides (http://staff.ustc.edu.cn/~yxu/hdg.pdf)

What classes of dynamics can this hybridizable architecture catch

Broadly, $y'' = F(y)$

Conservation of energy independent of rollouts for linear and nonlinear systems

Theorem 2.3 (Discrete conservation of Hamiltonian for arbitrary number of rollouts). *Consider the rollout of Equation 6 over N subdomains. Then the following discrete Hamiltonian is preserved, independent of the number of subdomains.*

$$\left[\frac{1}{2}|J|^2 + \frac{\epsilon}{2}|\pi^* u|^2 - \lambda \pi^* u\right]_{t_N} = \left[\frac{1}{2}|J|^2 + \frac{\epsilon}{2}|\pi^* u|^2 - \lambda \pi^* u\right]_{t_0}. \tag{8}$$

Theorem 2.4. *For the modified nonlinear forcing $\mathcal{N} = V'(u)$, we do not obtain the same telescoping structure, but instead preserve the discrete energy*

$$\left[\frac{1}{2}|J|^2\right]_{t_0}^{t_N} + \sum_{i=0}^{N-1}\sum_{k=0}^{M-1} V'(u_k)\,(u_{k+1} - u_k) = 0 \tag{10}$$

Discrete Stieltjes integral
Recovers ∫ V’ du in
continuum limit

Discrete stability principle for introduction of abstract dissipative operators

To treat dissipation, we consider the addition of an abstract bilinear for $a : \mathcal{Q}_h^i \times \mathcal{Q}_h^i \to \mathbb{R}$, which we require to induce a norm $||\cdot||_a$ as follows

$$a(\pi J, \pi J) := ||J||_a^2. \tag{15}$$

As an example, the bilinear form associated with a linear dissipative oscillator $a(J, v) = \beta(J, v)$, $\beta > 0$, satisfies this property. Incorporating alongside our nonlinear Hamiltonian system we obtain

$$\begin{aligned} (J, v) + (u, \frac{d}{dt}v) &= \lambda_{i+1}v_{i+1} - \lambda_i v_i \\ (\frac{d}{dt}J, q) + a(\pi J, q) - (u, q) &= 0, \end{aligned} \tag{16}$$

The additional term provides the following stability results.

Theorem 2.5. *Define the discrete energy at the end of the i^{th} rollout as*

$$\mathcal{E}_i = \left[\frac{1}{2}|J|^2 + V(\pi^* u) - \lambda \pi^* u\right]_{t=t_i} \tag{17}$$

Then the energy in the system at the N^{th} rollout is bounded by the initial energy.

$$\mathcal{E}_N \leq \mathcal{E}_0 \tag{18}$$

Uniform stability of gradients independent of number of rollouts

To analyze the stability of backpropagation through multiple rollouts, consider the sequence o states $Y_k = (\lambda_k, J_k)$ at subdomain boundaries and the map

$$Y_{k+1} = \Phi(Y_k; \theta)$$

defined implicitly by solving (22) on Ω_k. Differentiating with respect to θ yields

$$\frac{\partial Y_{k+1}}{\partial \theta} = J_k \frac{\partial Y_k}{\partial \theta} + R_k,$$

where $J_k = \frac{\partial \Phi(Y_k;\theta)}{\partial Y}$ is the local Jacobian and $R_k = \frac{\partial \Phi(Y_k;\theta)}{\partial \theta}$. Iterating this recurrence,

$$\frac{\partial Y_N}{\partial \theta} = J_{N-1} \cdots J_0 \frac{\partial Y_0}{\partial \theta} + \sum_{k=0}^{N-1} \left(\prod_{j=k+1}^{N-1} J_j \right) R_k.$$

The first term vanishes because Y_0 is fixed and therefore independent of θ. If $\|J_k\| \leq L < 1$ for all k then following a geometric series argument

$$\left\| \frac{\partial Y_N}{\partial \theta} \right\| \leq \frac{\max_k \|R_k\|}{1 - L},$$

Uniform stability of gradients independent of number of rollouts

$$\mathcal{A} = \underbrace{\begin{bmatrix} \mathbf{M}_{\mathcal{V}} & \delta^{\top} & -\frac{\partial B_1}{\partial \lambda_{i+1}} \\ \delta & 0 & 0 \\ 0 & 0 & I \end{bmatrix}}_{\mathcal{A}_0} + h \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ \partial_{\hat{f}} \widehat{\mathcal{N}} & \partial_{\hat{u}} \widehat{\mathcal{N}} & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\mathcal{D}},$$

Lemma 3.1 (Bounded inverse of the local Jacobian). *Let $\mathcal{A} = \mathcal{A}_0 + \mathcal{D}$ be given by (23) with $\mathcal{Q}_h = \mathrm{dg}\mathbf{P}_0$ and $\mathcal{V}_h = \mathbf{P}_1$. Assume:*

1. *The $(\mathrm{dg}\mathbf{P}_0/\mathbf{P}_1)$-pair satisfies a discrete inf–sup condition, so that $\|\mathcal{A}_0^{-1}\| \leq C$ uniformly in h.*
2. *The partial derivatives of $\widehat{\mathcal{N}}$ with respect to $(\hat{u}, \hat{J})$ are uniformly bounded by a constant L.*

Then

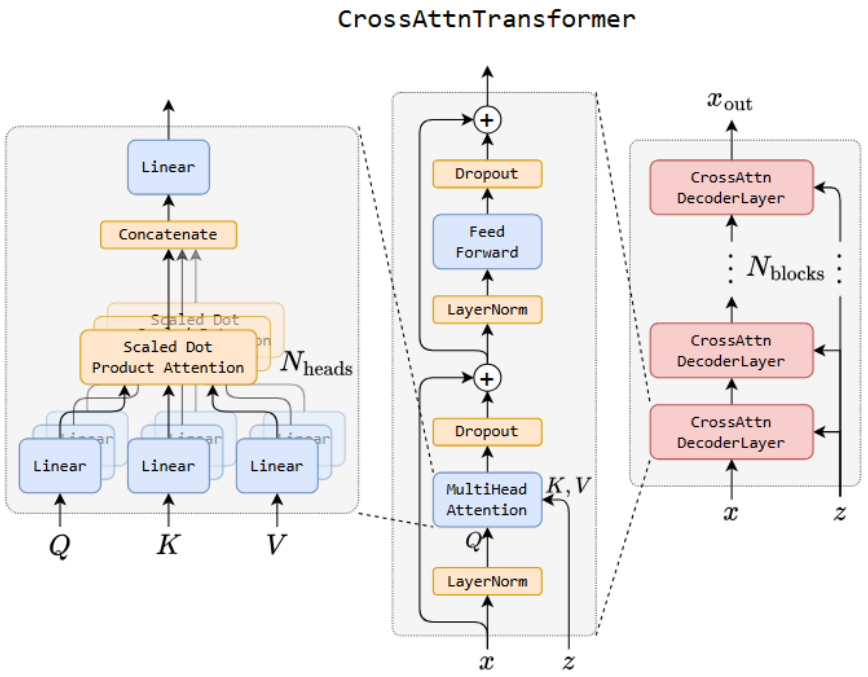
$$\|\mathcal{A}_0^{-1}\mathcal{D}\| \leq ChL,$$

and for sufficiently small h this norm is less than 1. Consequently,

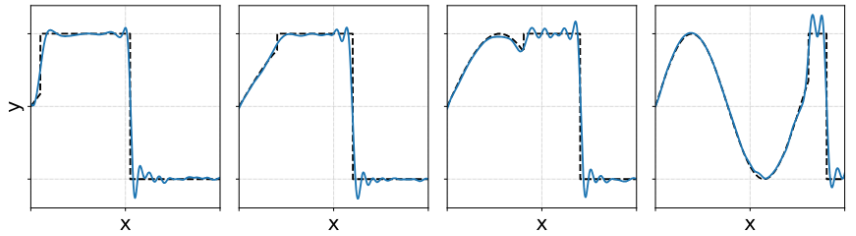
$$\mathcal{A}^{-1} = \sum_{k=0}^{\infty} (-\mathcal{A}_0^{-1}\mathcal{D})^k \mathcal{A}_0^{-1}, \qquad \|\mathcal{A}^{-1}\| \leq \frac{\|\mathcal{A}_0^{-1}\|}{1 - ChL},$$

and the bound is uniform in the number of subdomains.

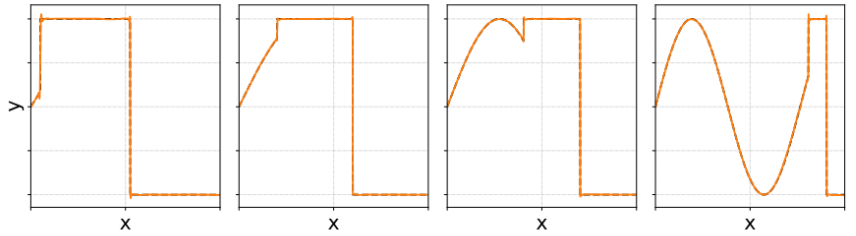
Tool three: a cross-attention transformer for operator inference conditioned on Z



(a) The multi-head attention implementation uses the built-in PyTorch module `torch.nn.MultiheadAttention`. The middle block is one layer of our `CrossAttnTransformer`: a standard decoding transformer without a self-attention component. Layer normalization is applied only over the final (per-token) dimension. The feedforward network is a two-layer MLP that expands and contracts the embedding dimension by a factor of 2, with a dropout layer after its activation. The `CrossAttnTransformer` is simply N_{blocks} of these layers stacked in sequence, each conditioned upon the same z.

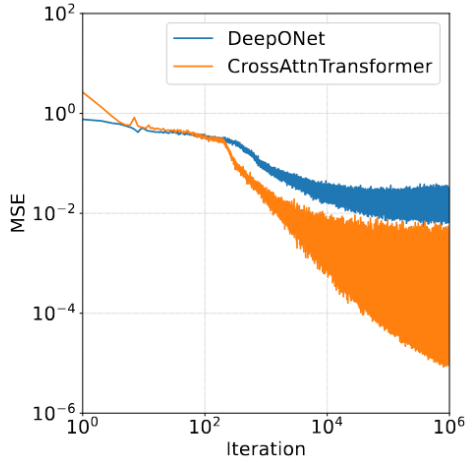


(a) Representative reconstructions for a vanilla DeepONet model (395,776 parameters).



(b) Representative reconstructions for proposed cross attention transformer (398,081 parameters).

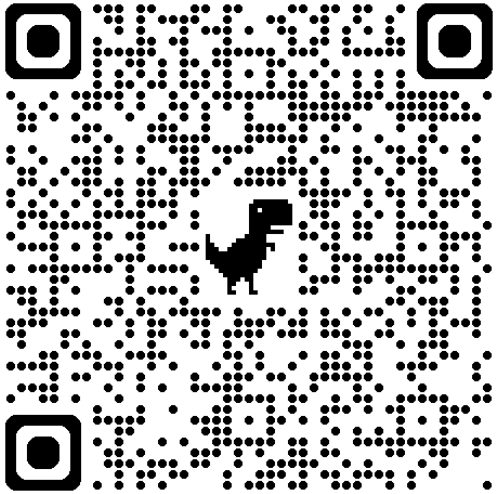
$$f(x) = \begin{cases} \sin(z_1 x) & \text{if } x < z_2 \\ 1 & \text{if } z_2 \le x < z_2 + \dfrac{1 - z_2}{2} \\ -1 & \text{if } x \ge z_2 + \dfrac{1 - z_2}{2} \end{cases}$$



(c) Convergence during training for both, demonstrating a three order-of-magnitude reduction in loss in the infinite data limit.

Conditional Neural Whitney Forms

Real time FEM conditioned on sensors to build digital twins

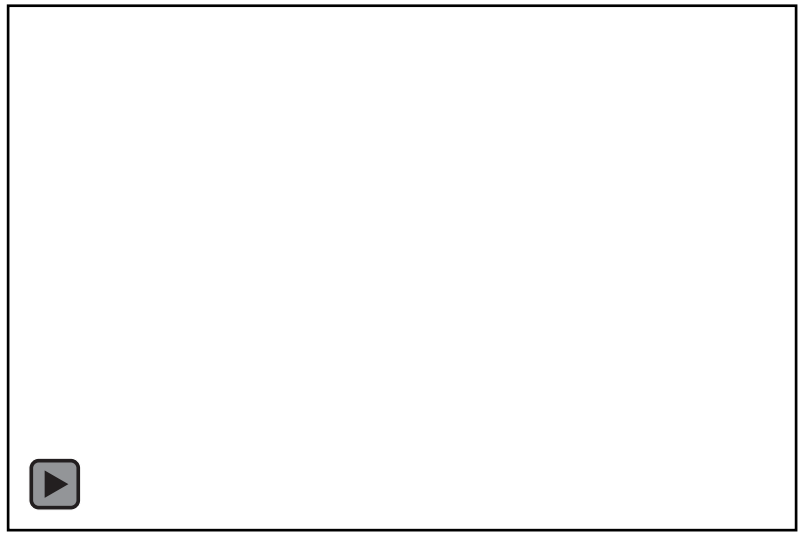


Hybridizable Neural Operators

Structure preserving autoregressive forecasting

(paper in a couple weeks)

Structure preserving neural operators via finite element exterior calculus



$$\mathcal{W}_i = \lambda_i$$

$$\mathcal{W}_{ij} = \lambda_i \nabla \lambda_j - \lambda_j \nabla \lambda_i$$

Neural Whitney forms
Differentiable architecture parameterizing control volumes and their boundaries

$$\nabla \cdot \mathbf{w} = f$$

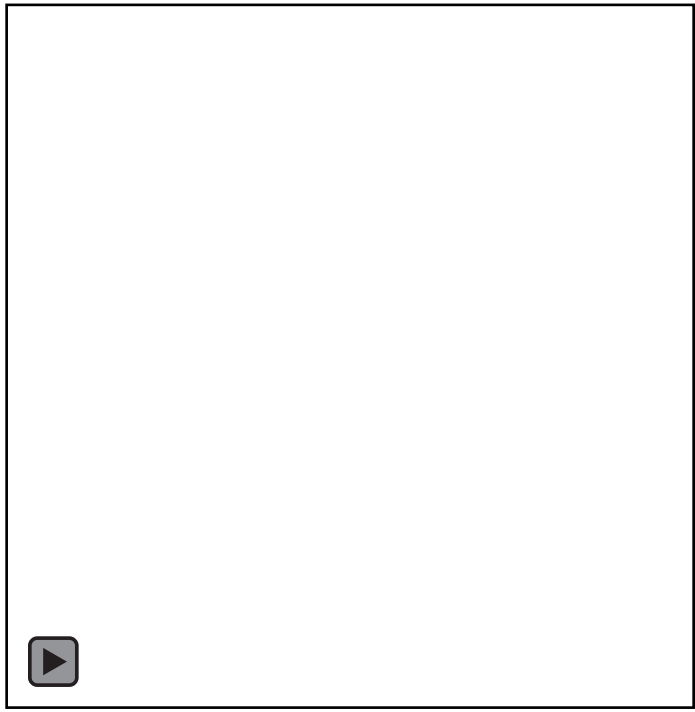
Conservation balance
Exact physics treatment

$$\mathbf{w} = -\nabla u + \mathcal{N}[u, \mathbf{w}; \theta]$$

Black-box generalized fluxes
Diffusion stabilized nonlinearity w/ uncertainty

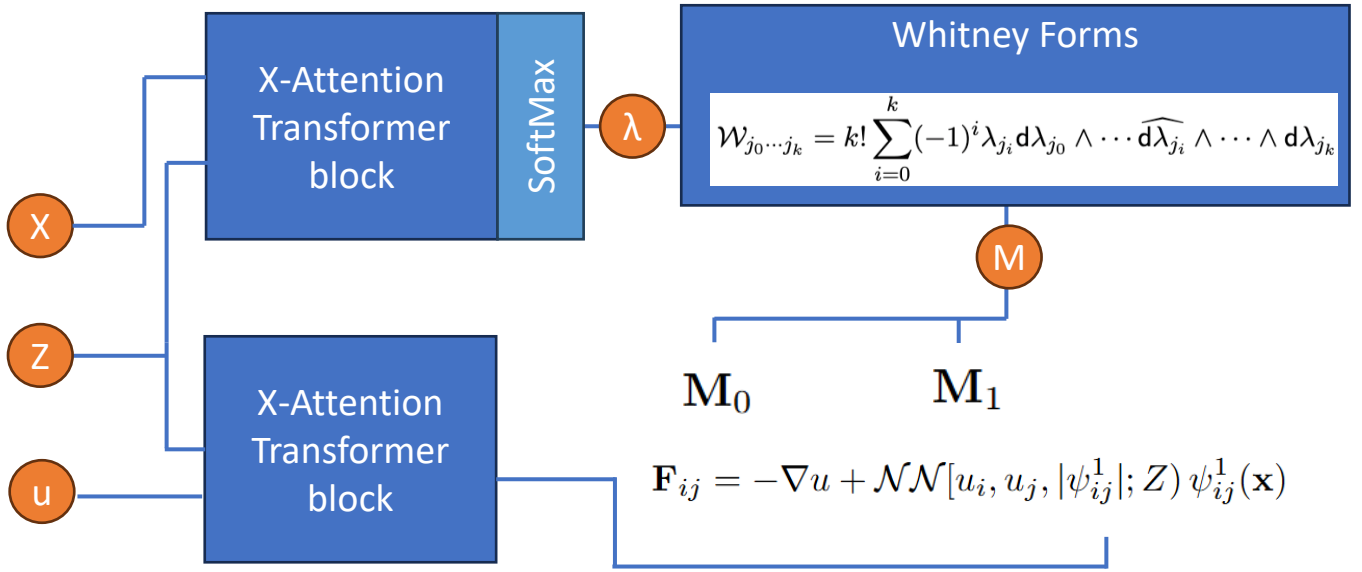
$$\underset{\mathbf{A},\theta}{\operatorname{argmin}} ||\mathbf{u} - \mathbf{u}_{data}||^2 + \epsilon^2 ||\mathbf{w} - \mathbf{w}_{data}||^2$$
$$\text{such that } a(\mathbf{u}, \mathbf{v}; \mathbf{A}) + N_{\mathbf{v}}[\mathbf{u}; \theta] = b(\mathbf{v})$$

Equality constrained optimization
Guaranteed structure preservation independent of data size



Data-driven FEM
Simultaneously identify control volumes and integral balance laws whose solution matches data

Real-time DTs via Conditional Neural Whitney Forms



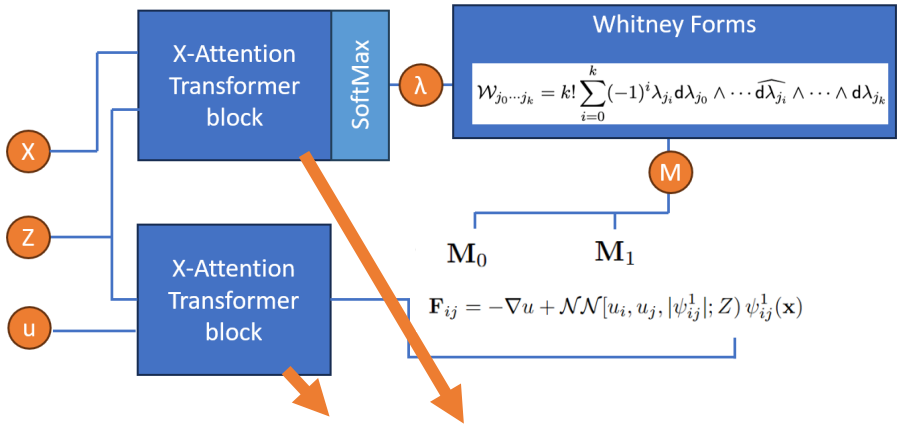
Massive strides in *conditional generative modeling*, generating images conditioned on a prompt (above)

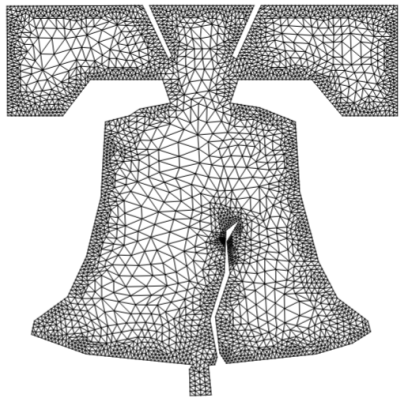
We extend the idea to sample from the space of finite element models conditioned on an input Z (sensor readings, parameterized geometry, or a latent variable) 1000x faster than standard FEM model

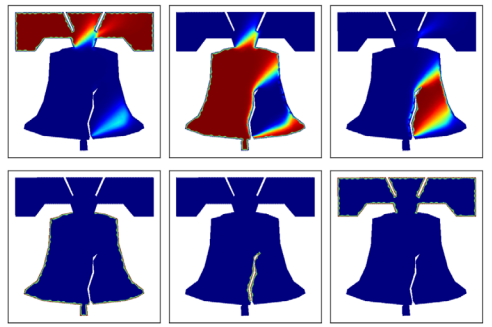
$$\underset{\mathbf{A},\theta}{\operatorname{argmin}} ||\mathbf{u} - \mathbf{u}_{data}||^2 + \epsilon^2||\mathbf{w} - \mathbf{w}_{data}||^2$$
$$\text{such that } \delta_0^\intercal \mathbf{M}_1 \delta_0 \hat{u} + \delta_0^\intercal \mathbf{M}_1 \mathcal{NN}[\hat{u}] = \mathbf{f}_\theta$$

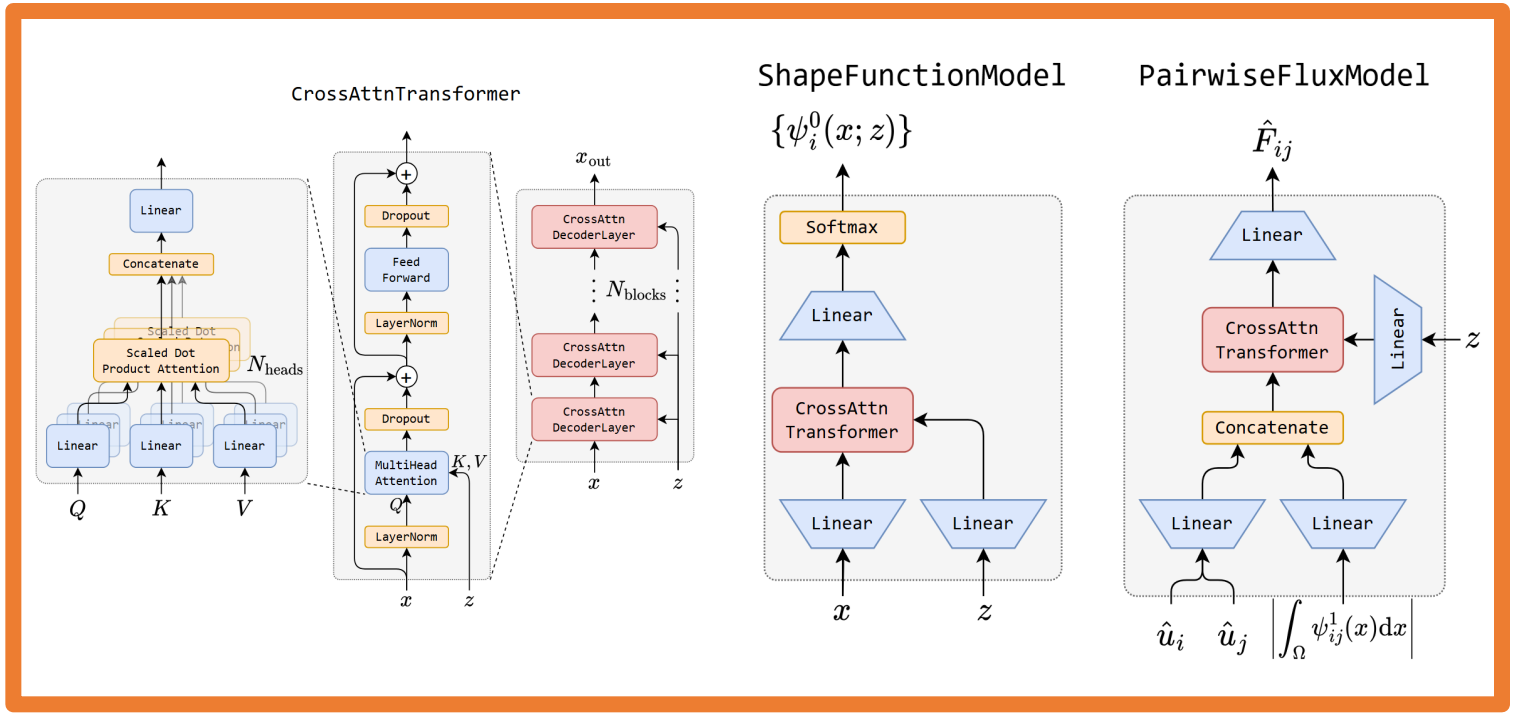
Left taken from: Synthesizing Coherent Story with Auto-Regressive Latent Diffusion Models
Xichen Pan[1], Pengda Qin[2], Yuhong Li[2], Hui Xue[2], Wenhu Chen

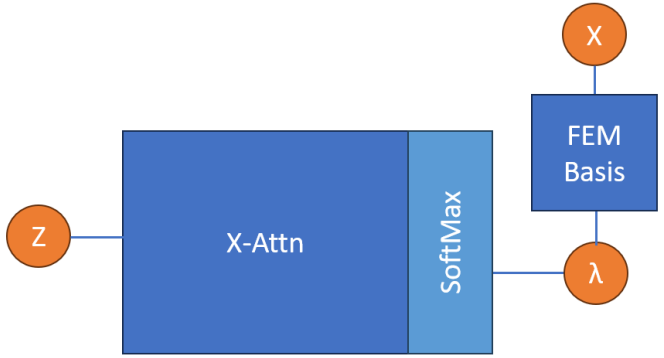
Under the hood – non-invasive cross-attention defines FEM space and physics







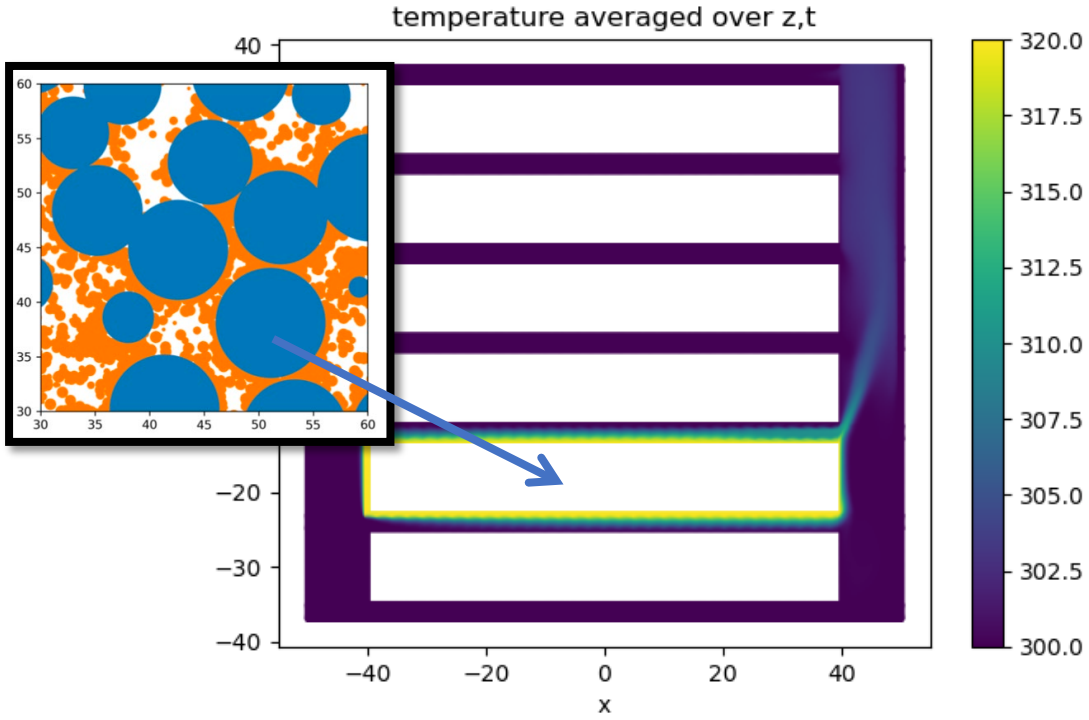


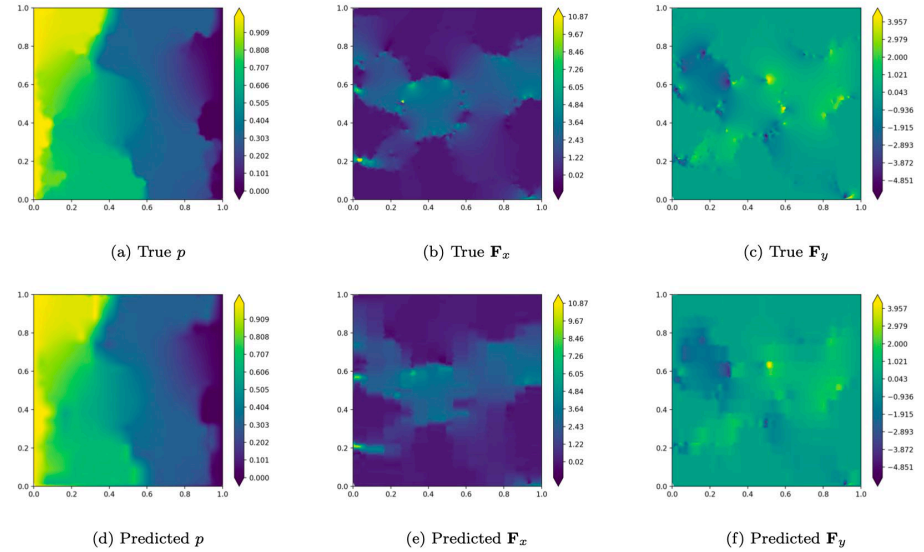


$$\mathbf{M}_0 = (\psi^0_i, \psi^0_j)$$

$$= (W_{ia}\lambda_a, W_{jb}\lambda_b) = W_{ia}M_{ab}W_{jb}$$

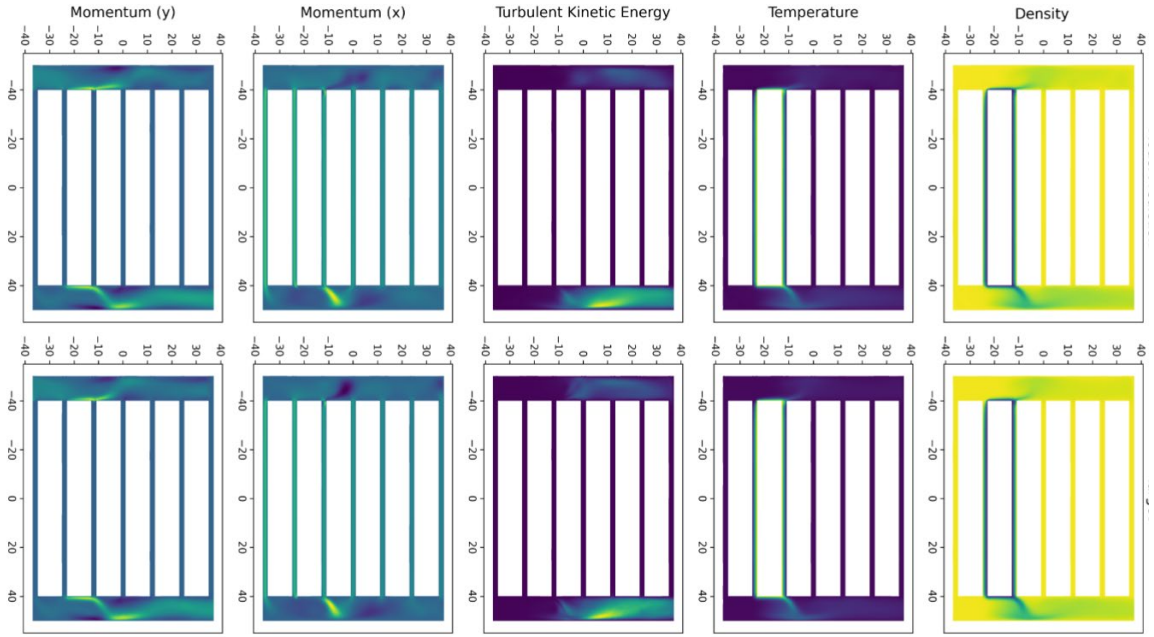
Modular digital twins bridging scales in Li-ion battery packs





LES

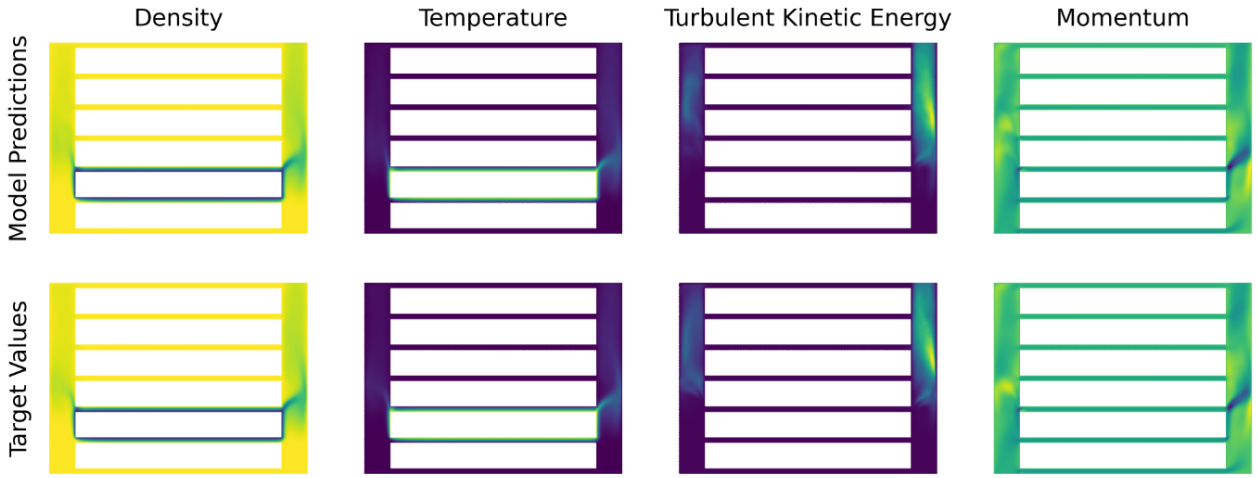
Learned DT



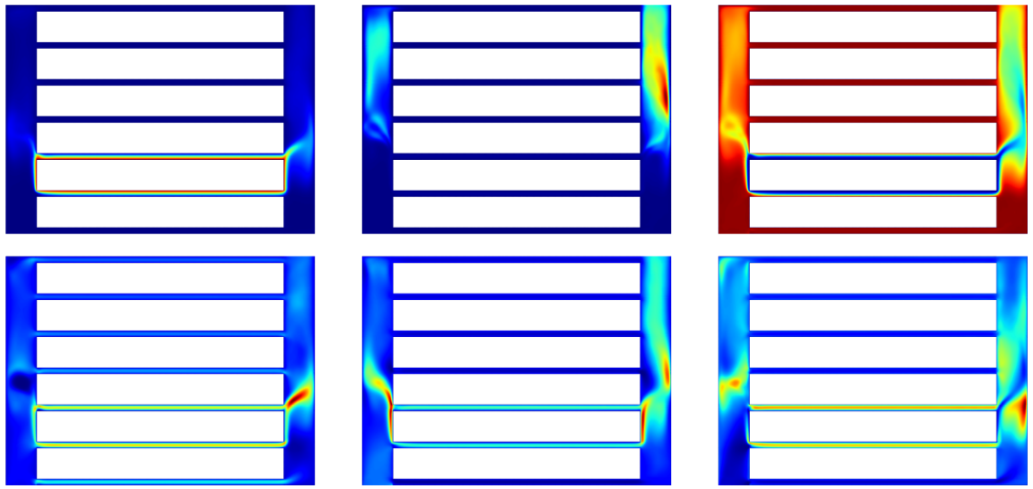
Replace a 5.89M finite element simulation of as-built battery with 8 data-driven elements w/ ~0.1% error, 10k CPU-hour LES simulation with <<1 sec data-driven RANS simulation

Structure preservation allows bidirectional and modular coupling from **material to engineering scales**

Real time RANS from LES – **a 3.11 x 10^9 speedup**

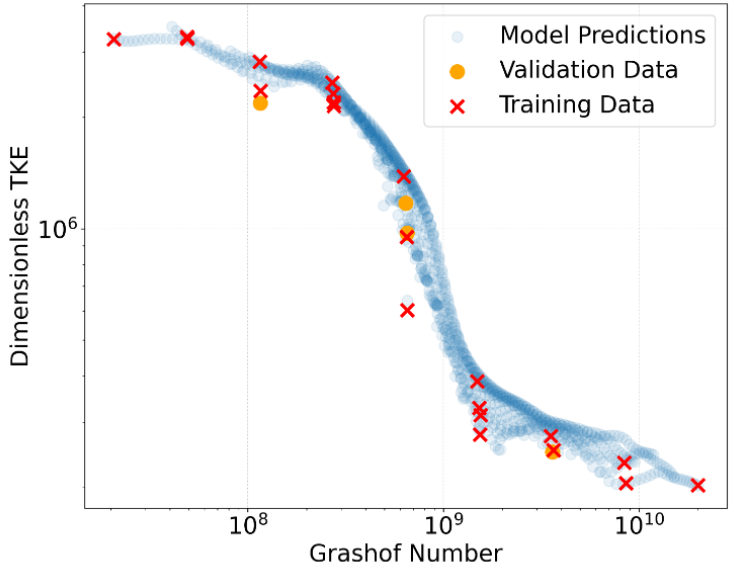


(a) The model predictions (left) and the targets (right), for all scalar fields, for the held-out validation example for which $\Delta T = 31.6$ K, $\mu = 6.7e4$ g/cm-s.



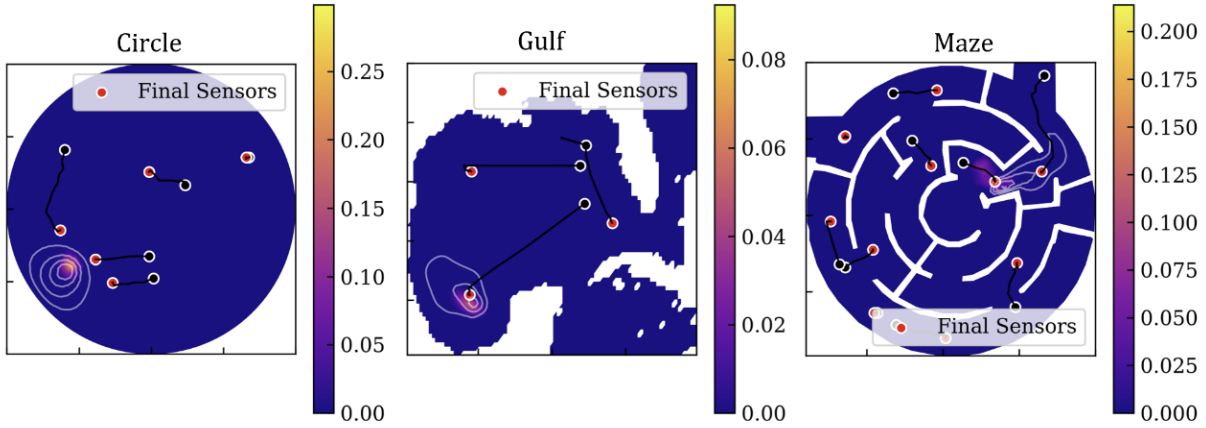
(b) The six learned shape functions for the training example shown in Figure 7a. All but one shape function are used to capture different aspects of the plume off the hot battery module (the second from the bottom of the rack). The shape functions used to enforce Dirichlet boundary conditions around each battery module, and for the surrounding wall, are not shown.

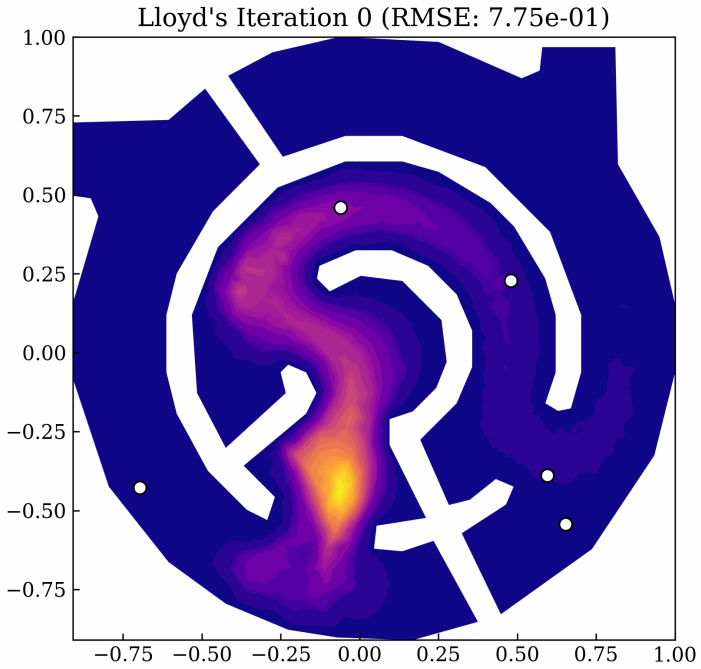
$$\partial_t \rho + \nabla \cdot F_1(\rho, T, k, M) = 0$$
$$\partial_t T + \nabla \cdot F_2(\rho, T, k, M) = 0$$
$$\partial_t k + \nabla \cdot F_3(\rho, T, k, M) = 0$$
$$\partial_t M + \nabla \cdot F_4(\rho, T, k, M) = 0$$

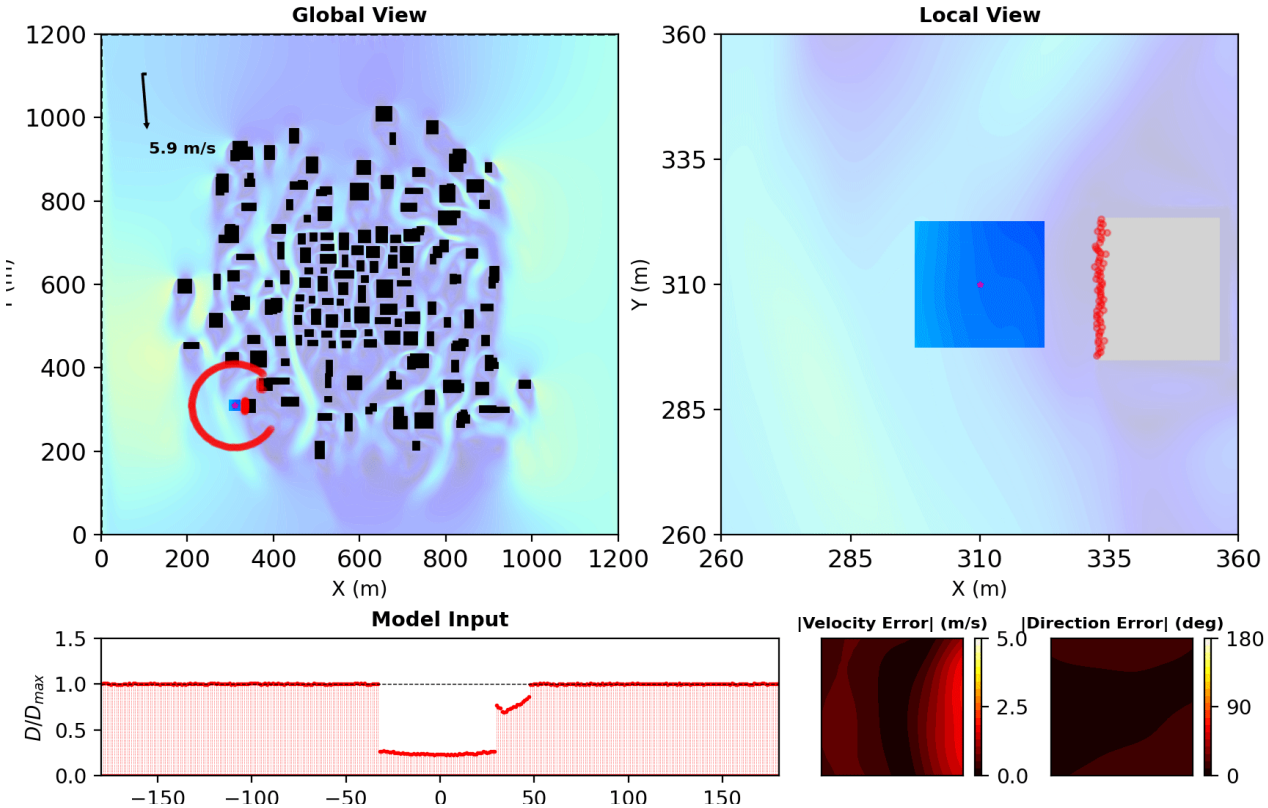


(a) Predicted dimensionless TKE integrated over domain. Despite training on only 20 LES, we capture the transition to turbulence at $Gr = 10^9$.

A platform for evaluating real-time physics on distributed sensor networks





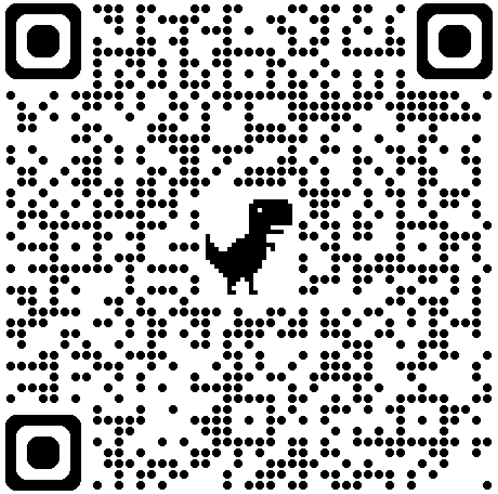


Left: Control scheme to identify physical model for source location for a stationary *(left)* and moving *(right)* source.

Right: Collaboration with robotics lab at Upenn (**Folk**, Hsieh, Kumar) to build digital twins in an urban environment

Conditional Neural Whitney Forms

Real time FEM
conditioned on sensors to build
digital twins



Hybridizable Neural Whitney Forms

Structure preserving
autoregressive forecasting

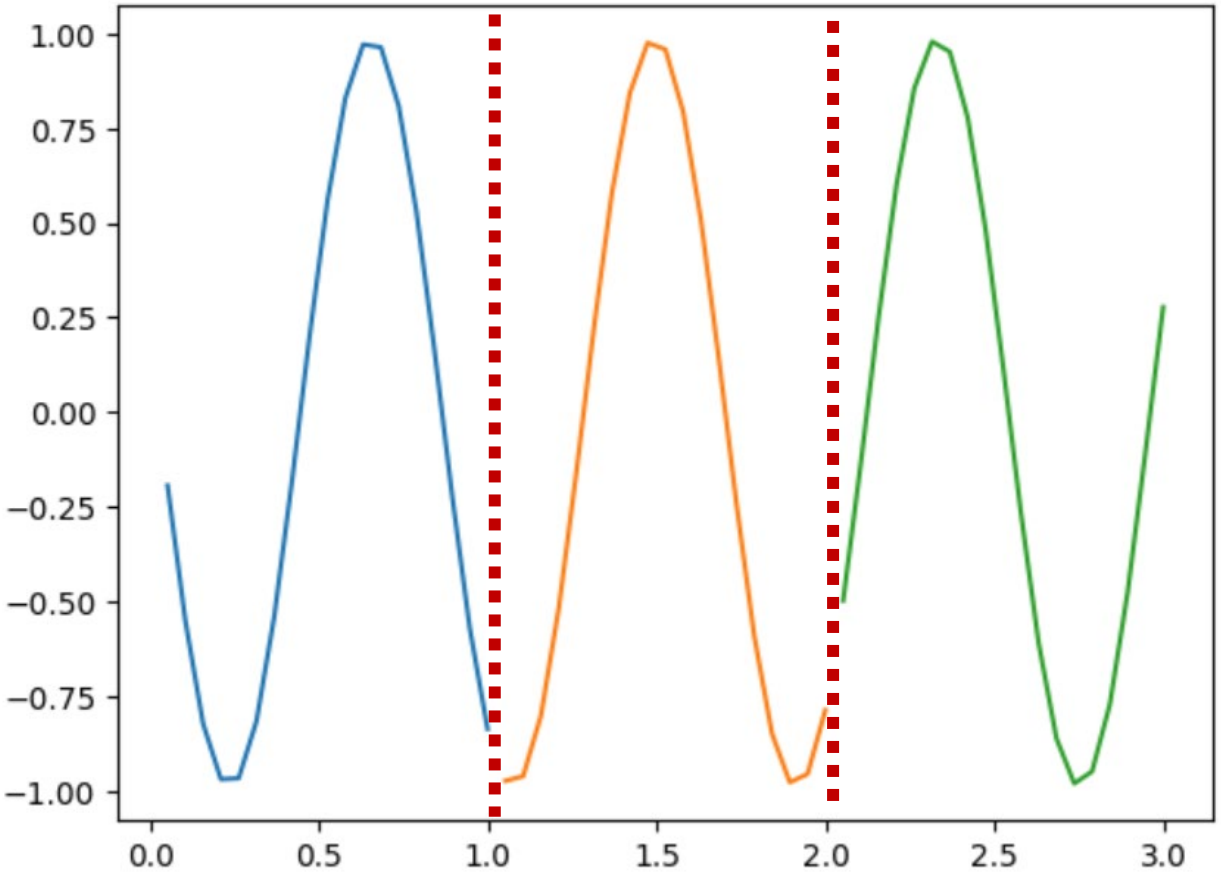
$$(J, v) + (u, \frac{d}{dt} v) = \lambda_{i+1} v_{i+1} - \lambda_i v_i$$

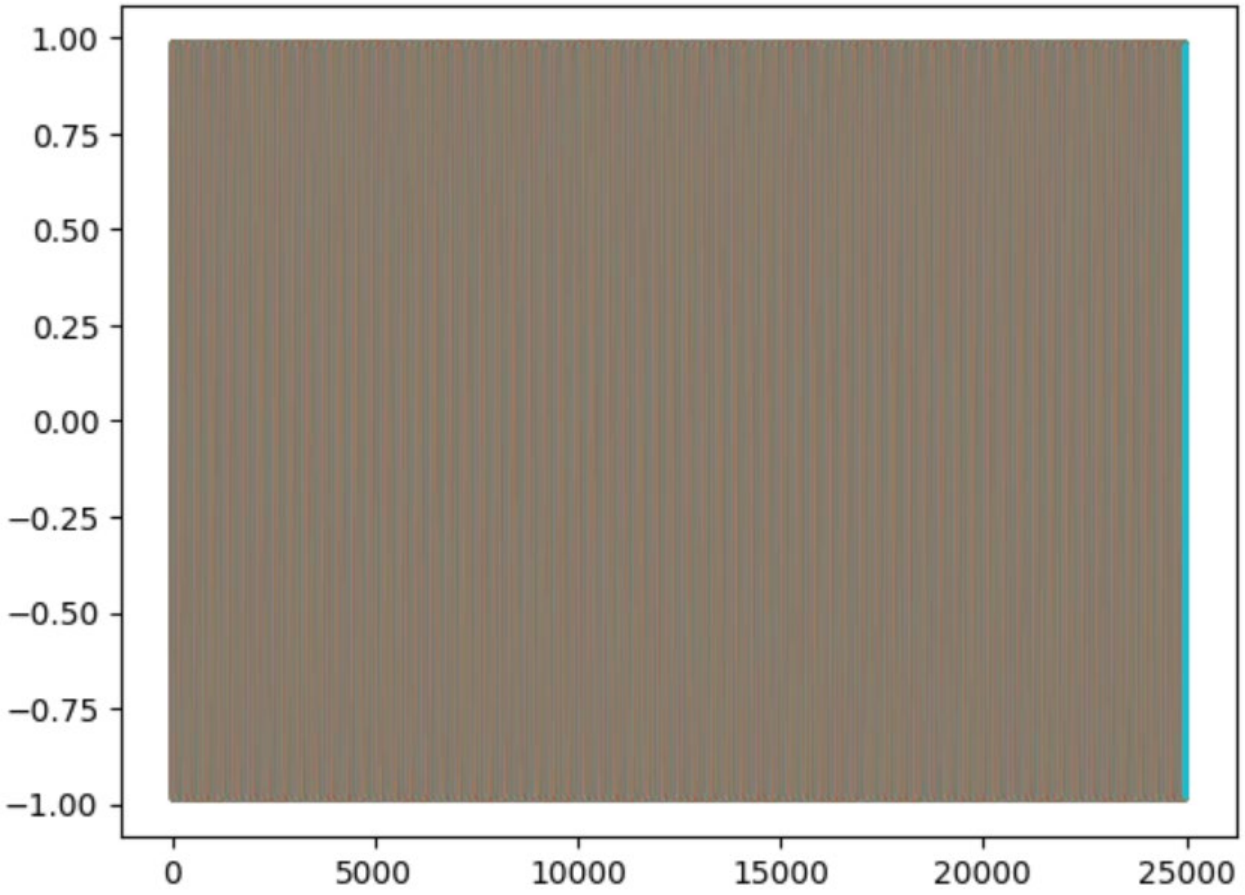
$$(\frac{d}{dt} J, q) - (\mathcal{N}[u, J], q) = 0$$

$$\lambda_i = u_i^0$$

$$J(t_i) = J_i^0,$$

Recovering theory – Hamiltonian dynamics

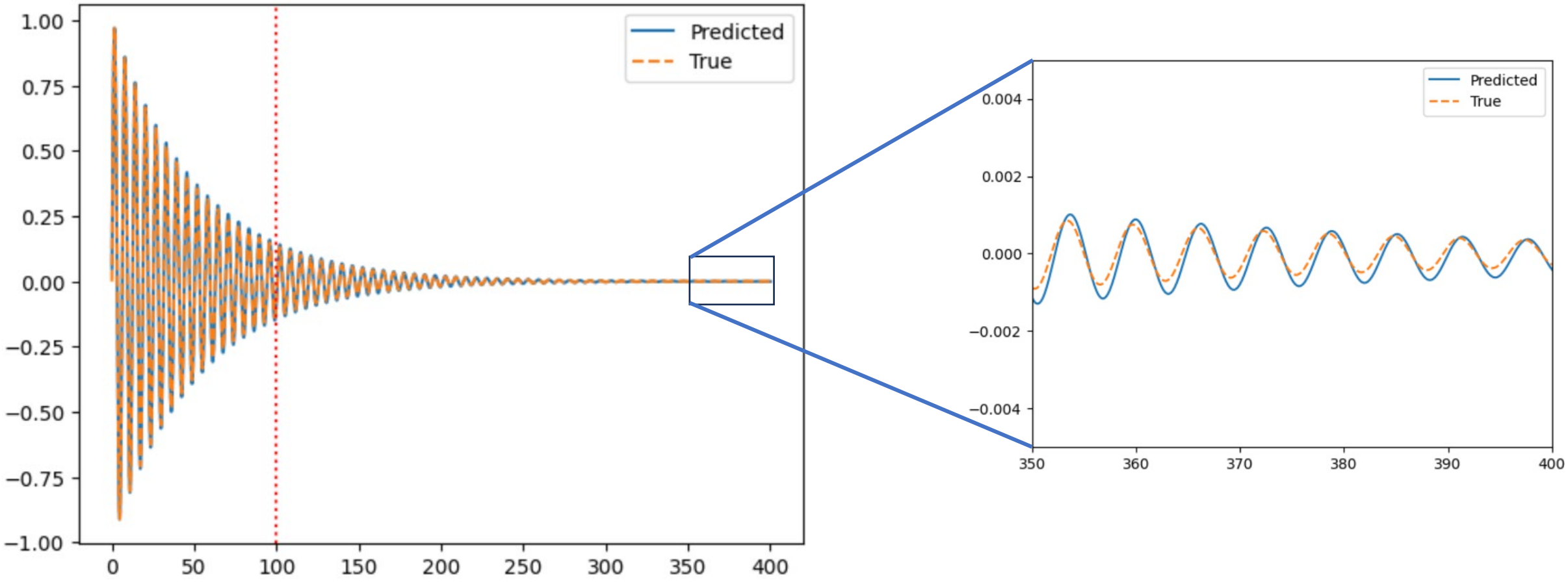




Rollout of 20K+ periods from training on 2

Theorem shows model can capture a Hamiltonian and forecast with energy independent on number of rollouts

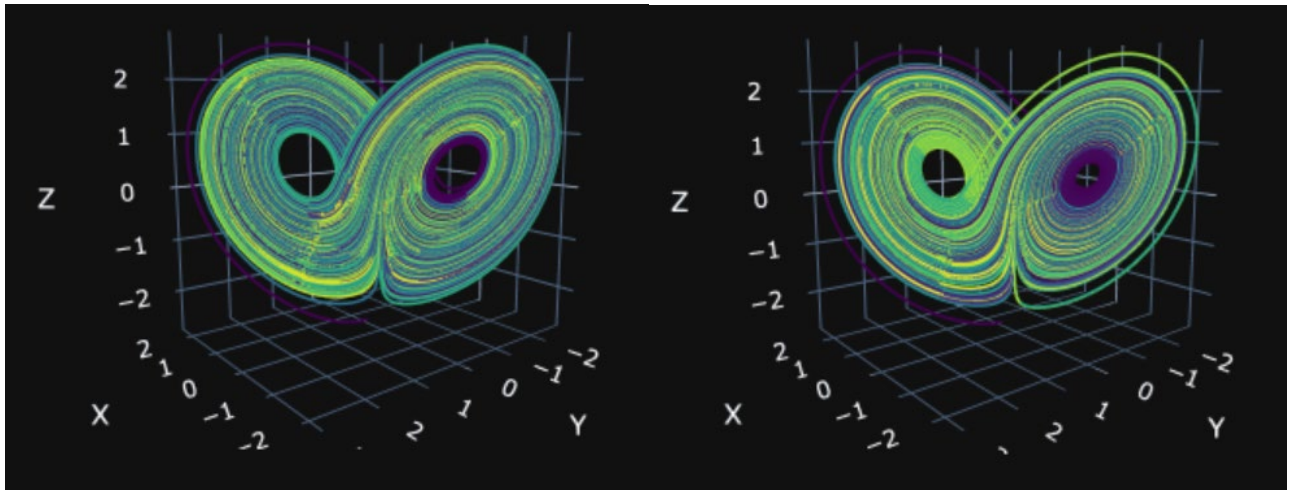
Recovering theory – dissipative dynamics

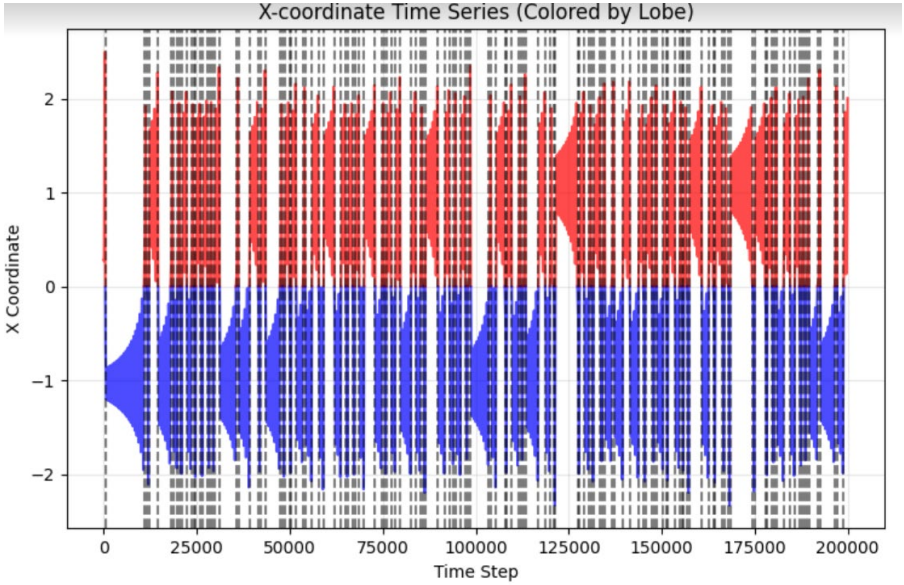


Forecasting dissipative pendulum from short term observation

Surprisingly good

Autoregressive roll-outs – chaotic dynamics



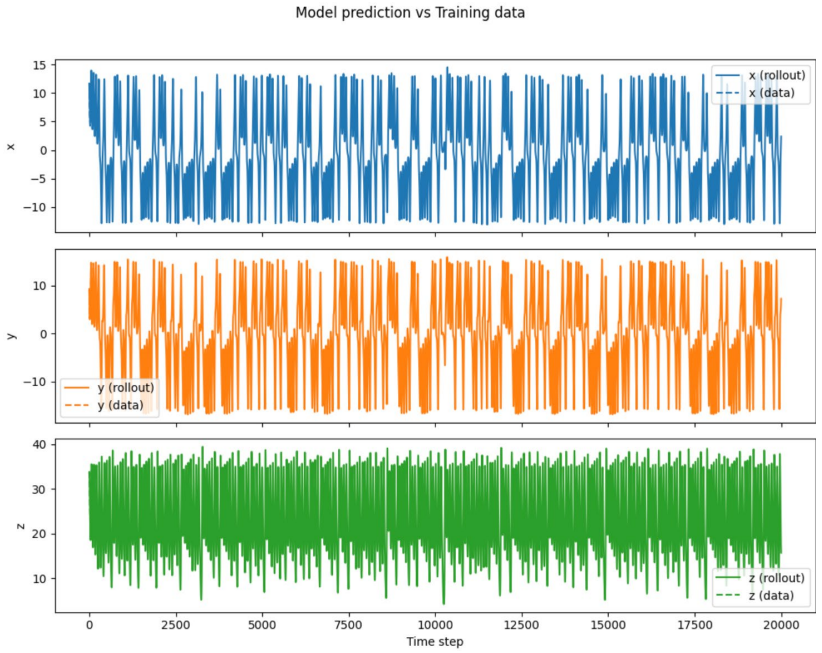


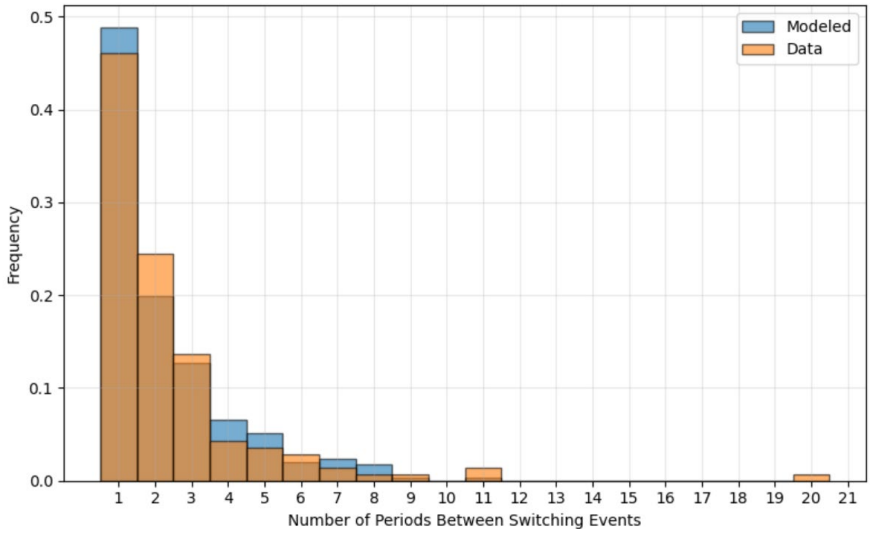
Lorenz trained on ~1 Lyaponuv period

Not possible to capture trajectories

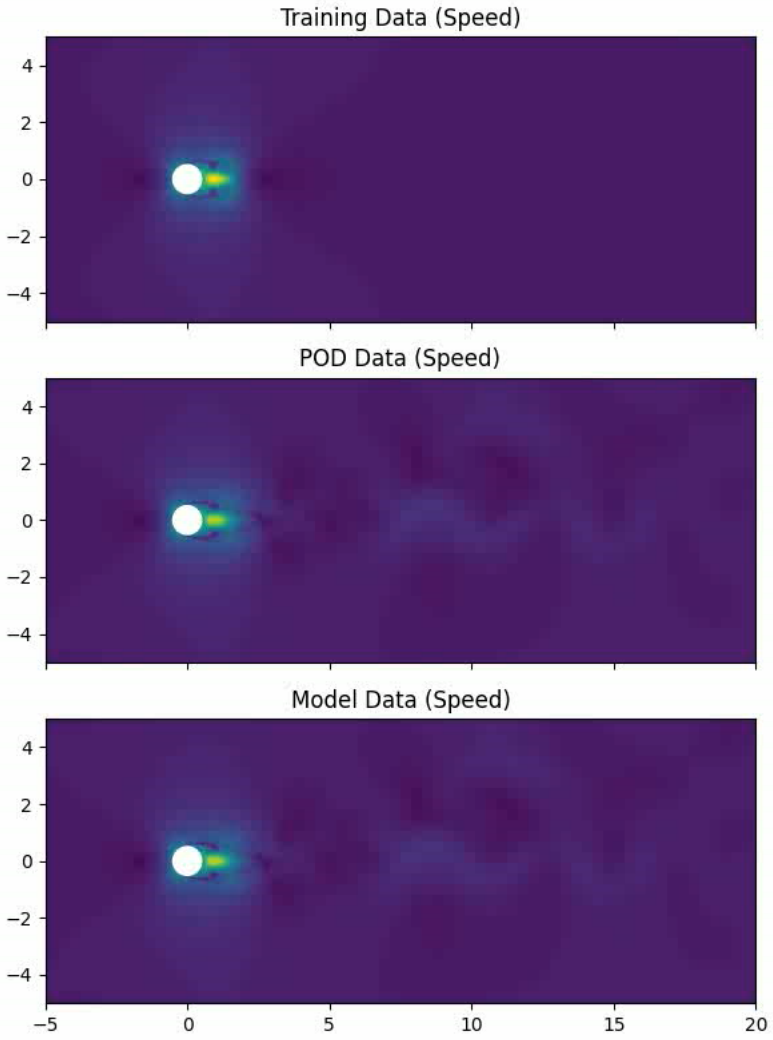
Capture attractor geometry

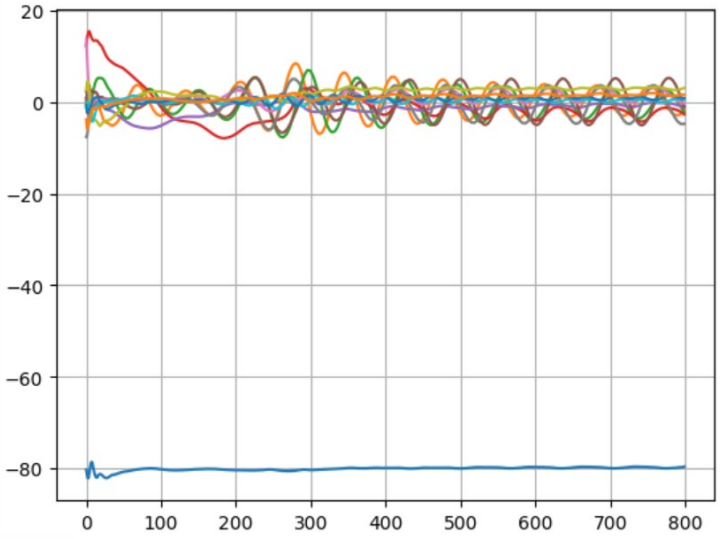
Capture statistics of switching

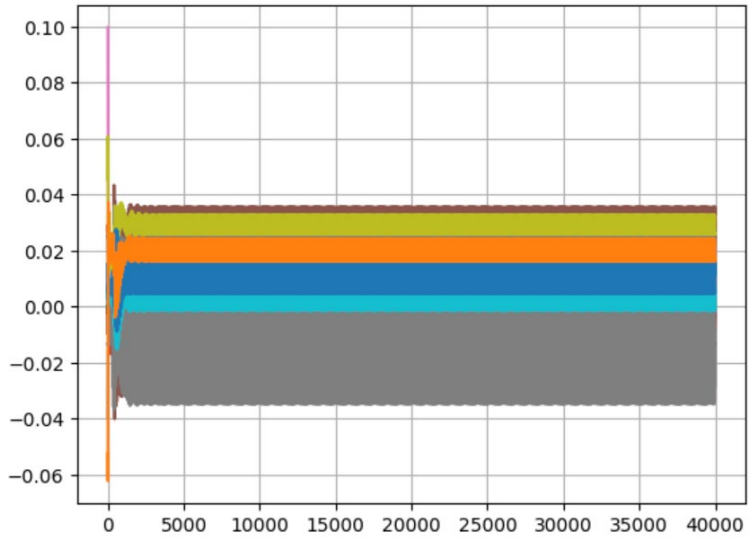




Autoregressive roll-outs – PCA modes of flow past a cylinder







Long term forecasting of von Karmann vortex

Many software engineering details here to scale up to large numbers of PCA modes…

Integrating with vision transformers – end-to-end training of latent dynamics

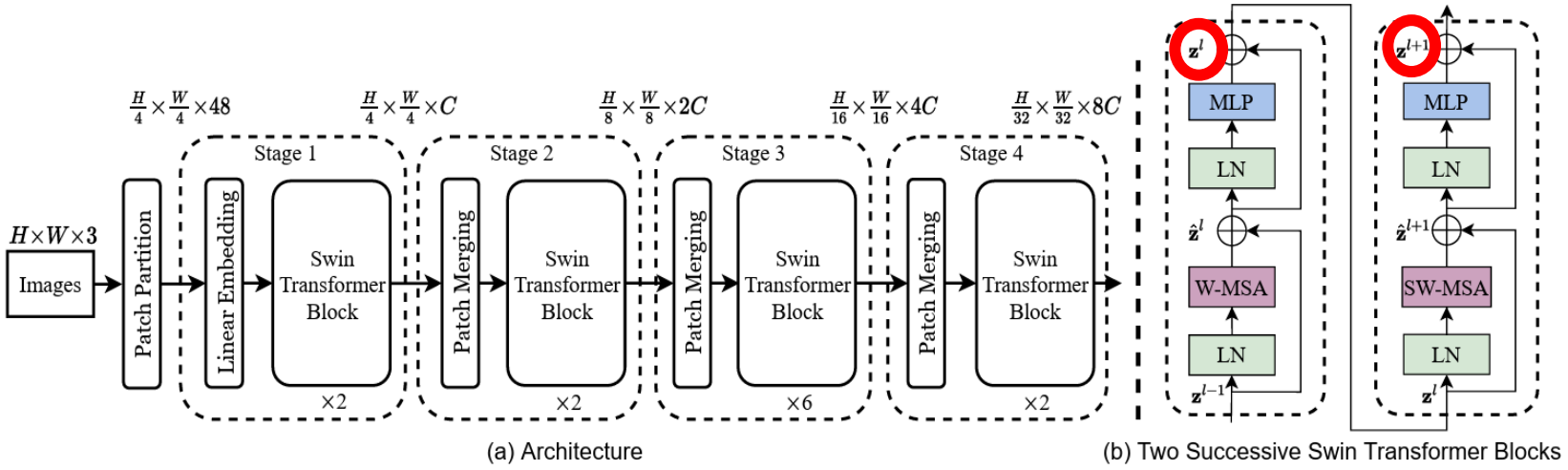
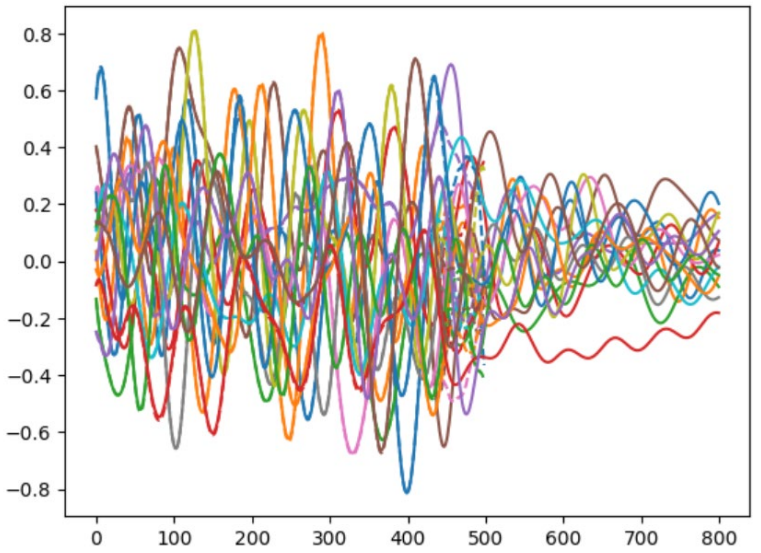


Figure 3. (a) The architecture of a Swin Transformer (Swin-T); (b) two successive Swin Transformer Blocks (notation presented with Eq. (3)). W-MSA and SW-MSA are multi-head self attention modules with regular and shifted windowing configurations, respectively.



$$(J, v) + (u, \frac{d}{dt}v) = \lambda_{i+1}v_{i+1} - \lambda_i v_i$$

$$(\frac{d}{dt}J, q) - (\mathcal{N}[u, J], q) = 0$$

$$\lambda_i = u_i^0$$

$$J(t_i) = J_i^0,$$

Swin transformer: Hierarchical vision transformer using shifted windows

Z Liu, Y Lin, Y Cao, H Hu, Y Wei, Z Zhang, S Lin, B Guo

Proceedings of the IEEE/CVF international conference on …, 2021 · openaccess.thecvf.com

Abstract

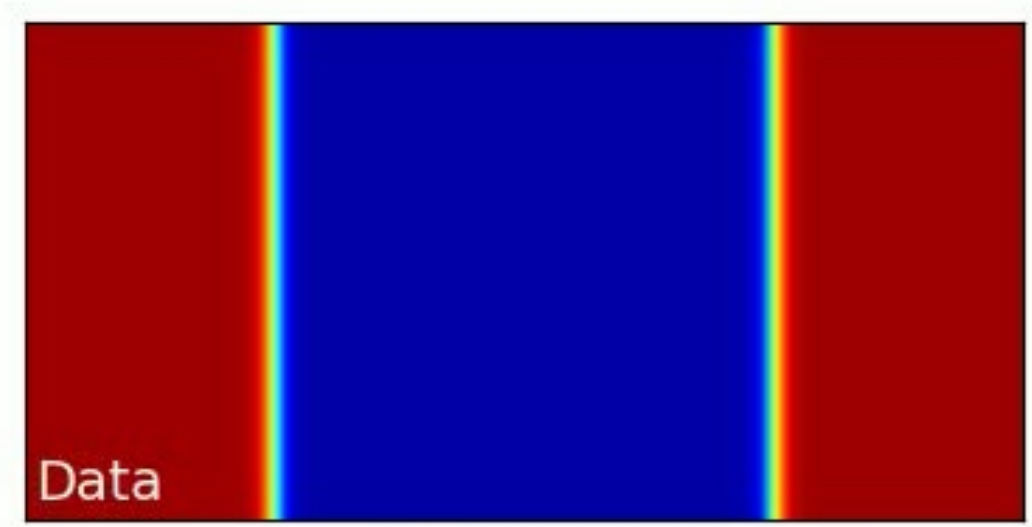
This paper presents a new vision Transformer, called Swin Transformer, that capably serves as a general-purpose backbone for computer vision. Challenges in adapting Transformer from language to vision arise from differences between the two domains, such as large variations in the scale of visual entities and the high resolution of pixels in images compared to words in text. To address these differences, we propose a hierarchical Transformer whose representation is computed with Shifted windows. The

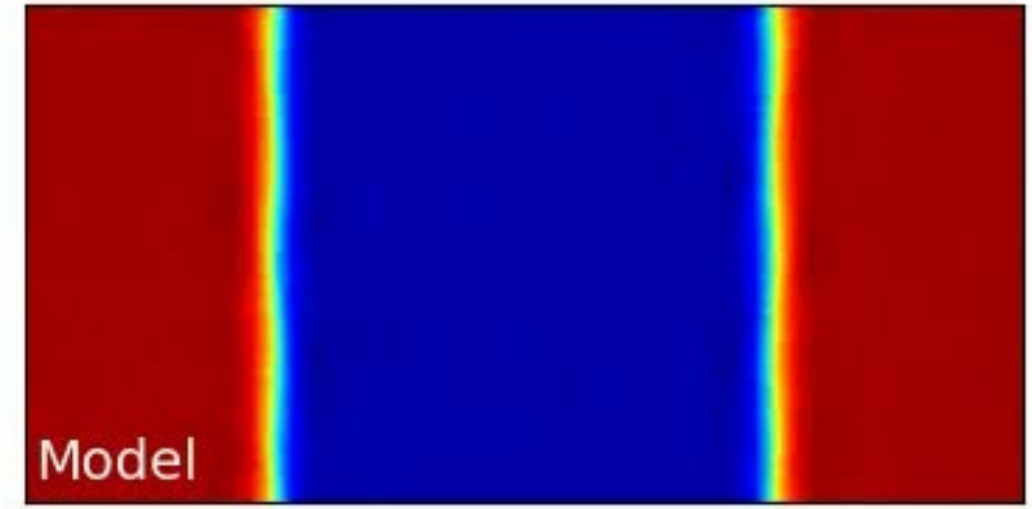
SHOW MORE

☆ Save Cite Cited by 35101 Related articles All 10 versions Import into BibTeX

Sign in

Integrating with vision transformers – end-to-end training of latent dynamics





The Well

Home Tutorial Datasets Data Format Data Visualizatior

Periodic shear flow

One line description of the data: 2D periodic incompressible shear flow.

Longer description of the data: A shear flow is a type of fluid characterized by the continuous deformation of adjacent fluid layers sliding past each other with different velocities. This phenomenon is commonly observed in various natural and engineered systems, such as rivers, atmospheric boundary layers, and industrial processes involving fluid transport. The dataset explores a 2D periodic shearflow governed by incompressible Navier-Stokes equation.

Associated paper: Paper 1, Paper 2, Paper 3.

Data generated by: Rudy Morel, CCM, Flatiron Institute.

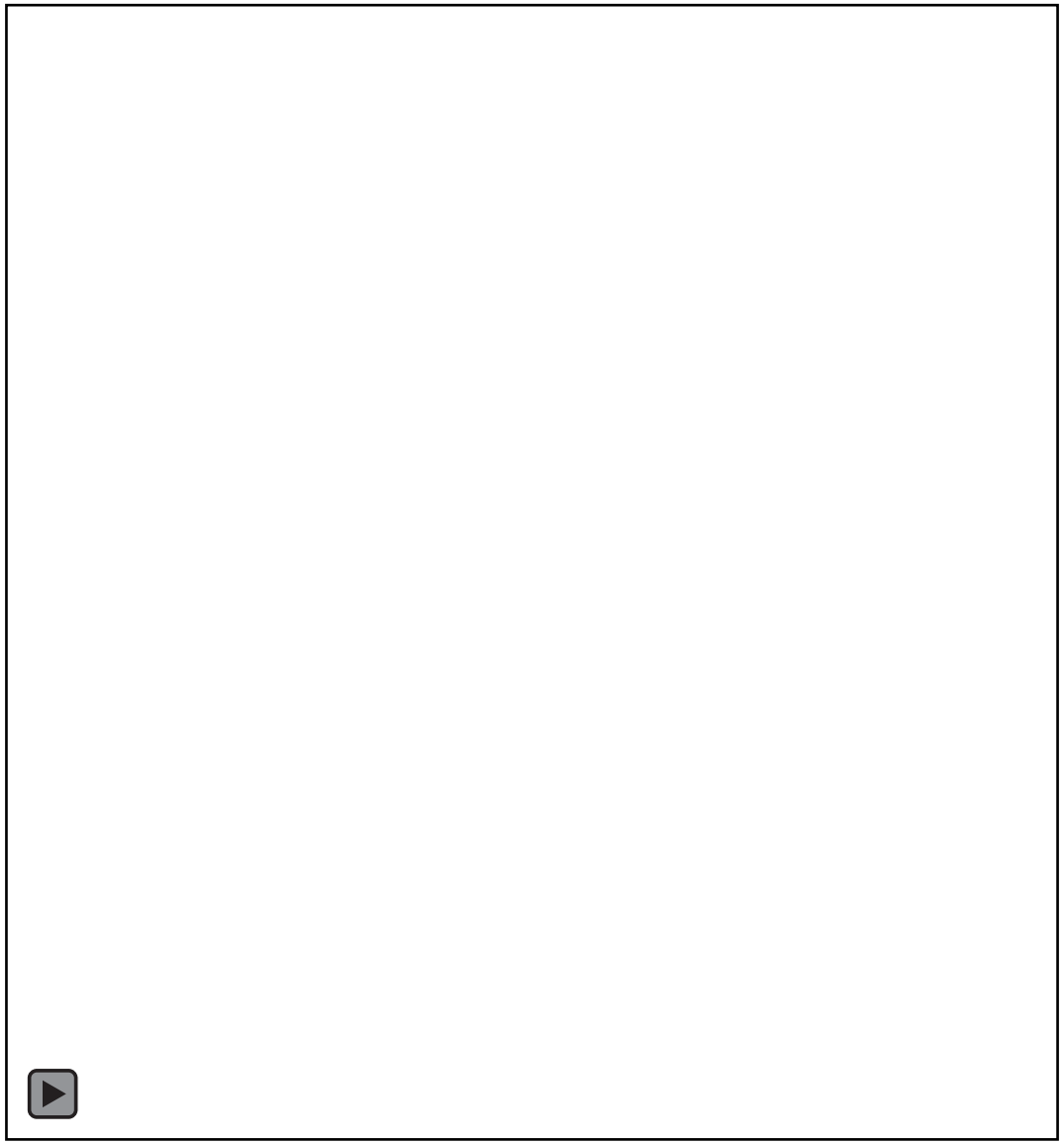
Code or software used to generate the data: Github repository, based on the software Dedalus.

Equation:

While we solve equations in the frequency domain, the original time-domain problem is

$$\frac{\partial u}{\partial t} + \nabla p - \nu \Delta u = -u \cdot \nabla u \,,$$
$$\frac{\partial s}{\partial t} - D \Delta s = -u \cdot \nabla s \,,$$

Integrating with vision transformers – end-to-end training of latent dynamics



The Well

Home Tutorial Datasets Data Format Data Visualizatior

Periodic shear flow

One line description of the data: 2D periodic incompressible shear flow.

Longer description of the data: A shear flow is a type of fluid characterized by the continuous deformation of adjacent fluid layers sliding past each other with different velocities. This phenomenon is commonly observed in various natural and engineered systems, such as rivers, atmospheric boundary layers, and industrial processes involving fluid transport. The dataset explores a 2D periodic shearflow governed by incompressible Navier-Stokes equation.

Associated paper: Paper 1, Paper 2, Paper 3.

Data generated by: Rudy Morel, CCM, Flatiron Institute.

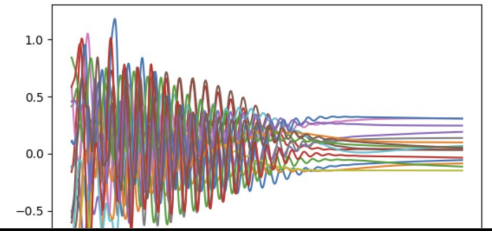
Code or software used to generate the data: Github repository, based on the software Dedalus.

Equation:

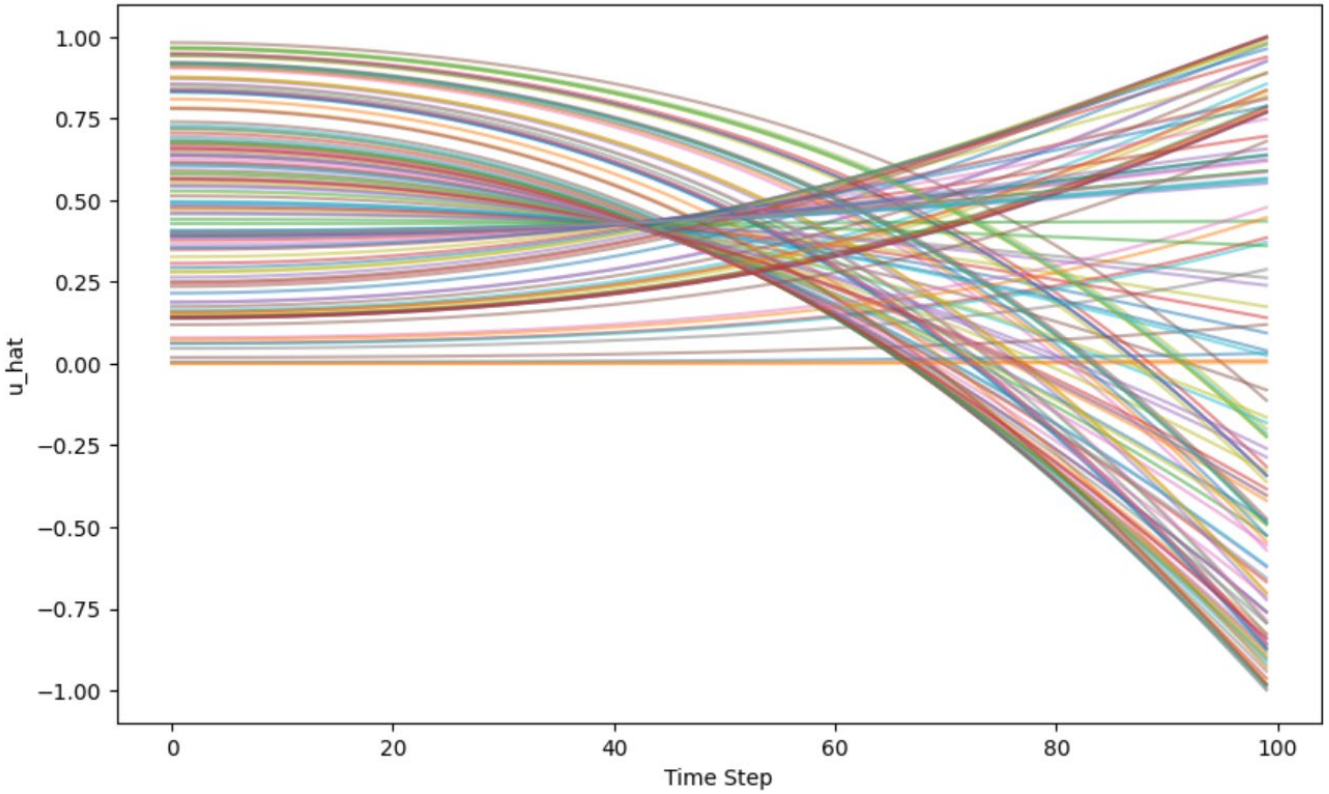
While we solve equations in the frequency domain, the original time-domain problem is

$$\frac{\partial u}{\partial t} + \nabla p - \nu \Delta u = -u \cdot \nabla u \,,$$
$$\frac{\partial s}{\partial t} - D \Delta s = -u \cdot \nabla s \,,$$

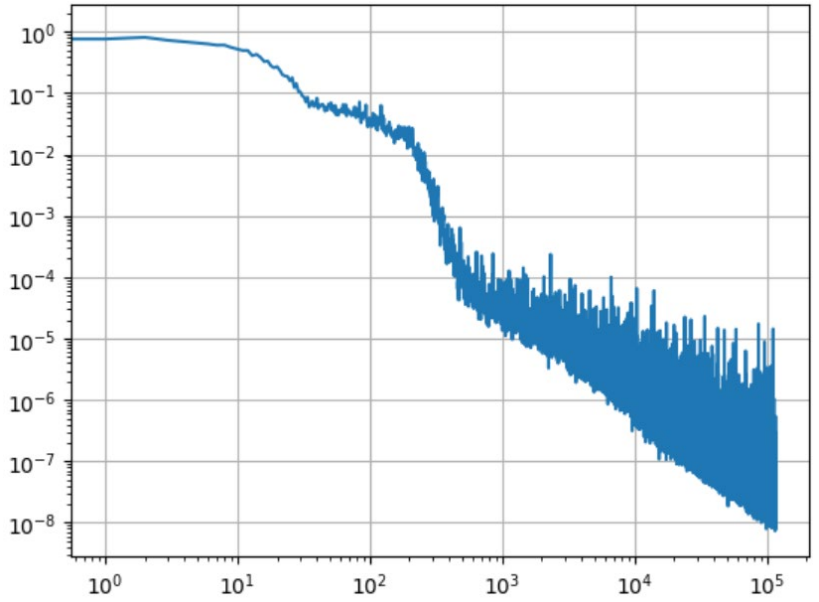
Integrating with vision transformers – end-to-end training of latent dynamics



AI4Science/Science4AI – Regression flow maps



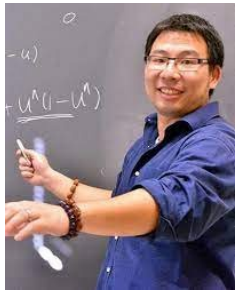
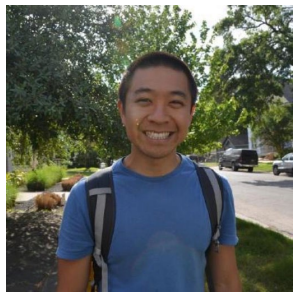
$$x \in [0,1] \rightarrow \sin 2\pi x \in [-1,1]$$



Supervised initial/final condition for conventional supervised learning

Stability allows arbitrarily deep networks which access accuracy surpassing conventional MLPs, KANs, etc

We can have it all!
Don't need to choose between black-box autoregressive methods and rigorous math/physics











Data-driven exterior calculus

Xiaozhe Hu – Tufts
Andy Huang, Jonas Actor- SNL
Brooks Kinch - UPenn

Metriplectic bracket discovery

Anthony Gruber – SNL
Kookjin Lee - ASU
Max Win, Quercus Hernandez - UPenn
Panos Stinis - PNNL

Autoregressive Rollouts

Brooks Kinch - UPenn
Martine Dyring-Hansen – NTNU
Sunniva Meltzer - SINTEF

Optimal Coverage, Colloids, Polymer melts

Ben Shaffer, Ani Hsieh, Paulo Arratia – UPenn
Thomas O'Connor– CMU
Pep Espanol – UNED Madrid