Beyond Closure Models: Learning Chaotic Systems via Physics-Informed Neural Operators

Chuwei Wang, Julius Berner, Zongyi Li, Di Zhou, Jiayun Wang, Jane Bae, Anima Anandkumar

Caltech

NEURAL INFORMATION PROCESSING SYSTEMS

Background

Central Task: Estimate Long-term Statistics of Chaotic Systems with Coarse-grid Simulations

$$\begin{cases} \partial_t u(x,t) = \mathcal{A}u(x,t) \\ u(x,0) = u_0(x),\ u_0 \in \mathcal{H} \end{cases}$$

$\mathcal{A}$: (Nonlinear) Operator; $\mathcal{H}$: function space of interest.

Attractor Ω: All trajectories $\{u(\cdot,t)\}$ will converge to it as $t \to \infty$.

Invariant Measure: $\mu^* := \lim_{T\to\infty} \frac{1}{T}\int_{t=0}^{T} \delta_{S(t)u} dt,\ u \in \mathcal{H},\ a.e.$ (average along traj).

Long-term Statistics: $\mathbb{E}_{u\sim\mu^*}\mathcal{O}(u)$ for measurement functionals $\mathcal{O}$.

- Key information of the physical system at dynamical equilibrium.
- Important in application: airfoil design, climate modeling, etc.
- **[*REMARK*]** Impossible to track trajectories for very-long time in chaotic system, but possible to estimate statistics (shadowing lemma).

General Methods:

- Straightforward: Fully-Resolved Simulations (FRS)
  Numerically simulate with very fine spatio-temporal grids/meshes.
  TOO EXPENSIVE! Intractable for most practical problems.
- Estimations statistics with coarse-grid simulation:
  Need to account for the large discretization error (i.e. missing information from the fine scale).
- Known as **Closure Modeling** or **Coarse-graining**.

Scheme of Closure Modeling:

- $\mathcal{F}$: filter from fine grid to coarse grid, e.g. spatial downsampling, Fourier mode truncation etc.), viewed as a mapping in function space $\mathcal{H}$.
- Filtered Dynamic $\partial_t \bar{u} = \mathcal{F}\mathcal{A}u = \mathcal{A}\bar{u} + \underbrace{(\mathcal{F}\mathcal{A} - \mathcal{A}\mathcal{F})u}_{\text{Unresolved}}$, $(\bar{u} := \mathcal{F}u)$.

$$\begin{cases} \partial_t v(x,t) = \mathcal{A}v(x,t) + clos(v;\theta),\ x \in D' \\ v(x,0) = \bar{u}_0(x),\ \bar{u}_0 \in \mathcal{F}(\mathcal{H}), \end{cases}$$

- Interpretation:
- Assign a vector field ($\mathcal{A} + clos$) in the reduced space to drive the dynamics.

How to design closure models?

- Classical Models: hand-designed. Strong physical intuition and assumptions.
- Machine Learning for Closure Models (Hopes: better expressiveness)
- **[Learning Framework]** Supervised Learning (Single-State Model)

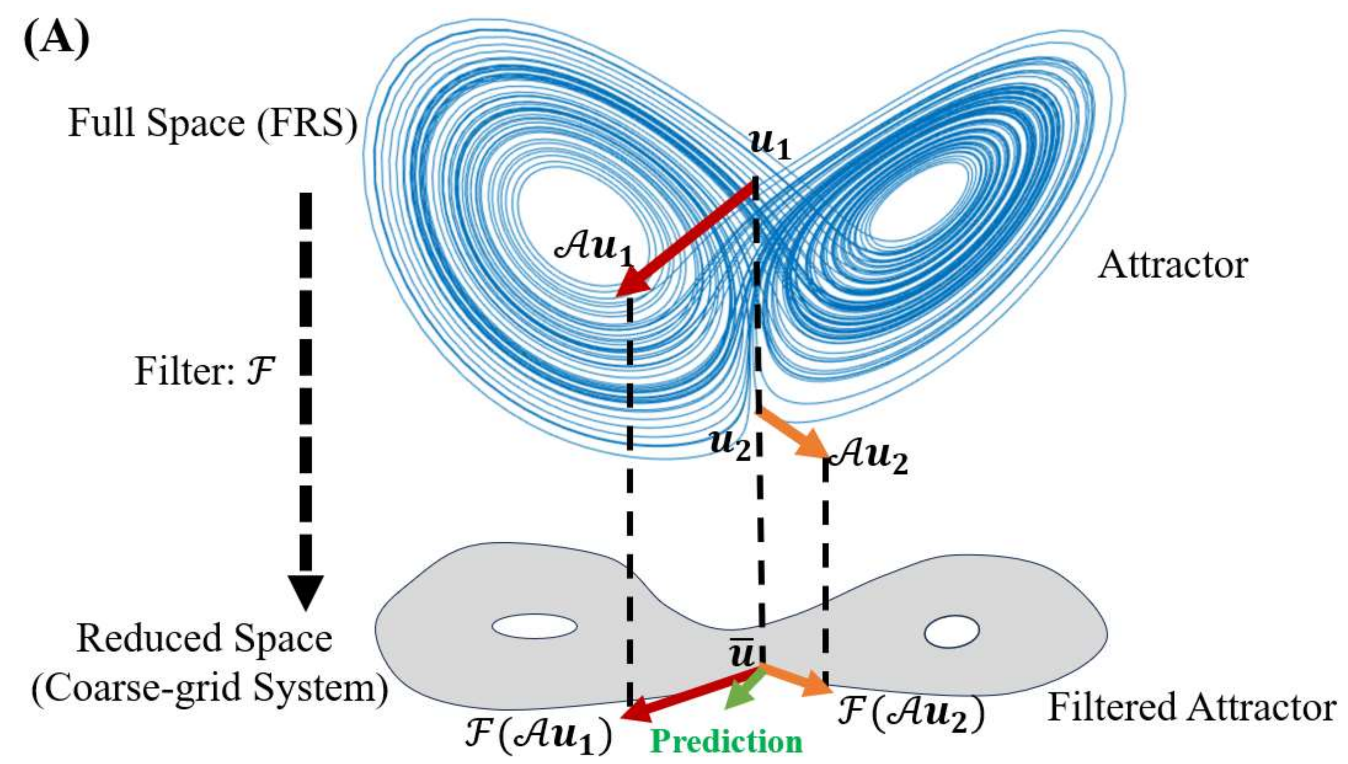
$$J_{ap}(\theta;\mathfrak{D}) = \frac{1}{|\mathfrak{D}|}\sum_{i\in\mathfrak{D}} \|clos(\bar{u}_i;\theta) - (\mathcal{F}\mathcal{A} - \mathcal{A}\mathcal{F})u_i\|^2$$

u_i: data from fully-resolved (fine-grid) simulations

- **[Advanced Variants]**
- [Posterior Training] $J_{post}(\theta;\mathfrak{D}) = J_{ap}(\theta) + \frac{1}{|\mathfrak{D}|}\sum_{i\in\mathfrak{D}} \|v_i(\cdot,\Delta t;\theta) - \mathcal{F}(S(\Delta t)u_i)\|^2$
- [History-aware Models] Model's input: $\{\bar{u}(x_i, t-s)\}_{x_i\in D',\ 0<s\le t_0}$
- [Stochastic Closure Models]

Theoretical Results

Learning-based methods should not follow previous closure modeling ansatz $\mathcal{A}\bar{u} + clos(\bar{u};\theta)$.



- ➢Learning-based closure models suffers from a large approximation error independent of model complexity, stemming from the **non-uniqueness** of the target mapping.
- ➢Leveraging history information and randomness can neither help.
- ➢A fundamental limitation for any method following the ansatz $\mathcal{A} + clos_\theta$.
- ➢To mitigate the nonunique issue, model has to use a large number of FRS data. The amount of training data is of the same order to estimate long-term statistics! – **eliminating the need for a closure model!**
- ➢One could not expect the model to generalize among different dynamics (e.g. different domain shape, different coefficient in the PDE).

Proof Idea: Functional Liouville Flow

- View functions u as particles.
- (Infinite-dimensional) Liouville eqn. for analyzing the limit distribution.

Coarse-grid Dynamics that can achieve optimal approx. of μ^*:

$\partial_t v = \mathbb{E}_{u\sim\mu_t}[\mathcal{F}\mathcal{A}u|\mathcal{F}u = v]$, μ_t: distribution of $u \in \mathcal{H}$ at time t.

For CGS and learning closure model, one can only fix a $\hat{\mu} \in \mathcal{P}(\mathcal{H})$, and evolve $\partial_t v = \mathbb{E}_{u\sim\hat{\mu}}[\mathcal{F}\mathcal{A}u|\mathcal{F}u = v]$

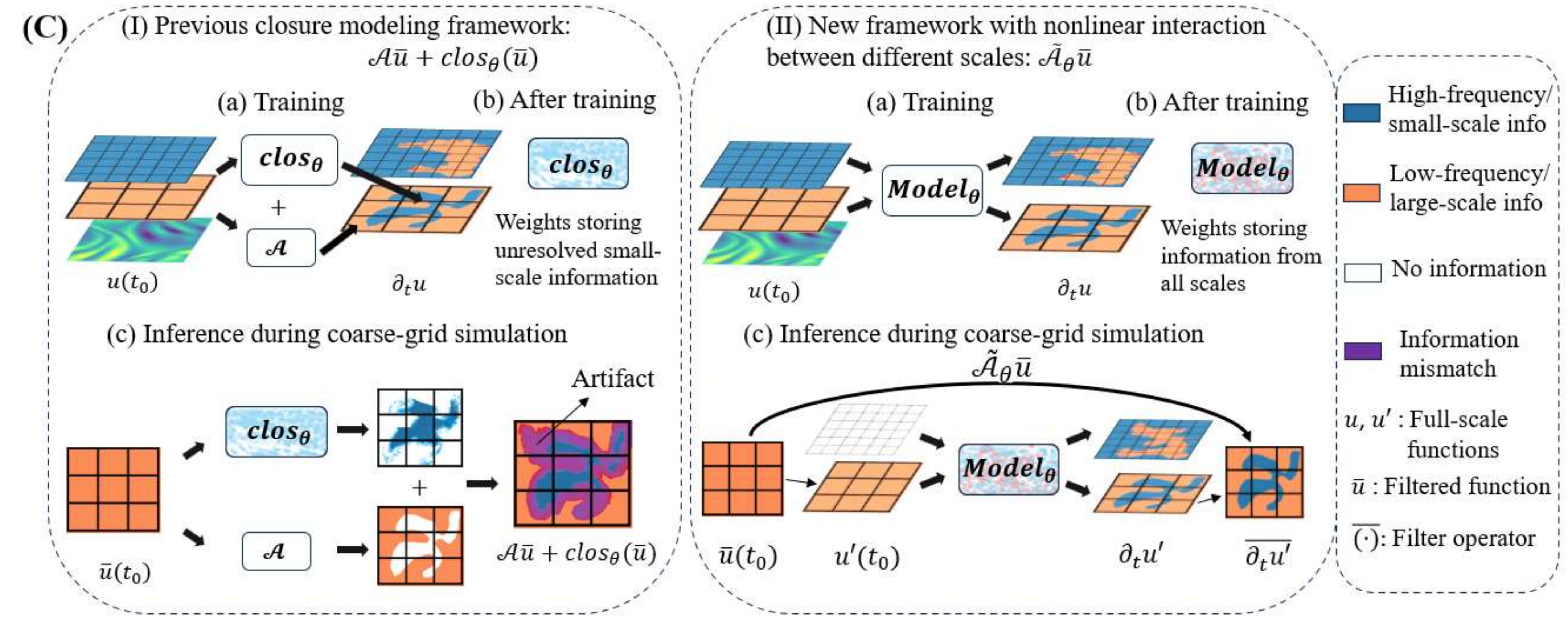
- ➢ $\hat{\mu} = \mu^*$: optimal approximation of μ^* in reduced system ($\mathcal{F}_\#\mu^*$).
- ➢In practice, the best model one can yield (assuming sufficient expressive power of NN function class) corresponds to $\hat{\mu} = \mu_{data}$.
- ➢Large gap between μ_{data} and μ^* for infinite-dim distributions!

Key Takeaway

➢We need nonlinear interaction between information from different scales (i.e. resolved part in coarse-grid system and unresolved parts)!

- ➢Previous ansatz: $\mathcal{A} + clos(\cdot,\theta)$
- ➢New Ansatz: $\tilde{A}_\theta$

New Ansatz with Physics-Informed Operator Leaning



Neural Operators

- ➢ Resolution-invariant. (Support input from both coarse-grid and fine-grid).
- ➢ $O(1)$ jump along time instead of moving with tiny time grids.
- ➢The infinitesimal generator of learned operator $\mathcal{G}_\theta$ plays the role of $\tilde{A}_\theta$
- ➢ Physics-informed learning + multi-resolution pre-training to reduce reliance on FRS data (only $\sim\mathbf{10^2}$ snapshots from **single** FRS trajectory vs $\sim\mathbf{10^5}$ in previous works).
- ➢Theoretical guarantee on optimal estimation of μ^* with coarse-grid simulations.

Theorem 3.1. *For any $h > 0$, denote $\hat{\mu}_{h,\theta} := \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\delta_{\mathcal{G}_\theta^n v_0(x)}$, any $v_0(x)$ with $x \in D'$. For any $\epsilon > 0$, there exists $\delta > 0$ s.t. as long as $\|(\mathcal{G}_\theta u)(\cdot,h) - S(h)u\|_{\mathcal{H}} < \delta, \forall u \in \mathcal{H}$, we have $\mathcal{W}_{\mathcal{H}}(\hat{\mu}_{h,\theta}, \mathcal{F}_\#\mu^*) < \epsilon$, where $\mathcal{W}_{\mathcal{H}}$ is a generalization of Wasserstein distance in function space.*

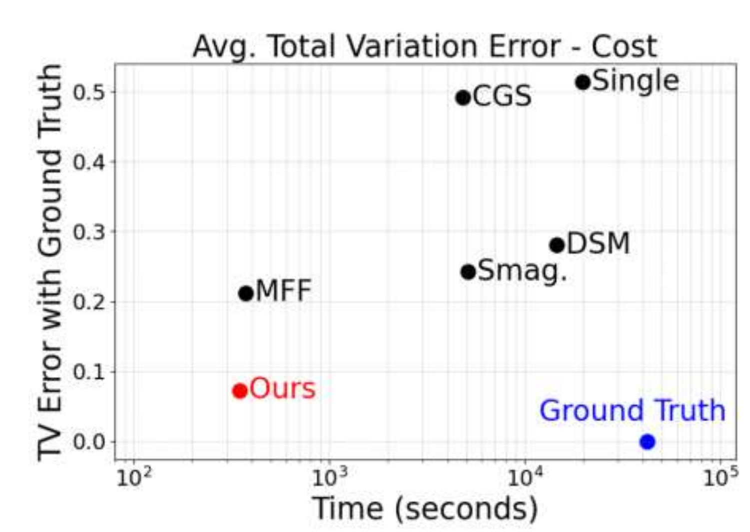
(Part of) Experiment Results

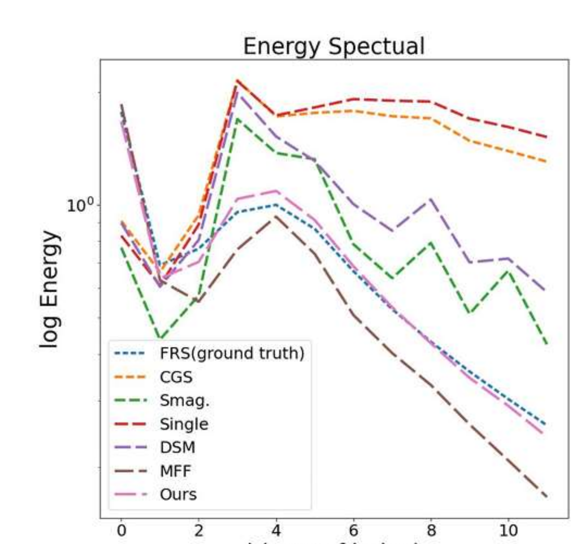
Kolmogorov Flow (2D forced Navier-Stokes eqn).

$$\partial_t \mathbf{u} = -(\mathbf{u}\cdot\nabla)\mathbf{u} - \nabla p + \nu\Delta\mathbf{u} + (\sin(4y), 0)^T, \quad \nabla\cdot\mathbf{u} = 0, \quad (x,y,t) \in [0,L]^2 \times \mathbb{R}_+$$

Method	Avg. TV	Energy	Vorticity	Variance
CGS (No closure)	0.4914	178.4651%	0.1512	253.4234%
Smagorinsky	0.2423	52.9511%	0.0483	20.1740%
Single-state	0.5137	205.3709%	0.1648	298.2027%
DSM	0.2803	74.2150%	0.0821	73.6158%
MFF	0.2123	20.7055%	0.0115	20.4410%
Our Method	**0.0726**	**5.3276%**	**0.0091**	**2.8666%**

FRS	39.70
CGS (No closure)	4.50
Smagorinsky	4.81
Single-state	18.57
DSM	13.67
MFF	**0.32**
Ours	**0.32**





The newest version: arxiv.org/abs/2408.05177