Scalable Neural Operator Training for High-Dimensional PDEs

Luis F. S. Ribeiro[1,2], Dibakar R. Sarkar[1], Deane Roehl[2], Somdatta Goswami[1]

[1]Department of Civil and System Engineering, Johns Hopkins University

[2]Department of Civil and Environmental Engineering, Pontifical Catholic University of Rio de Janeiro

Introduction

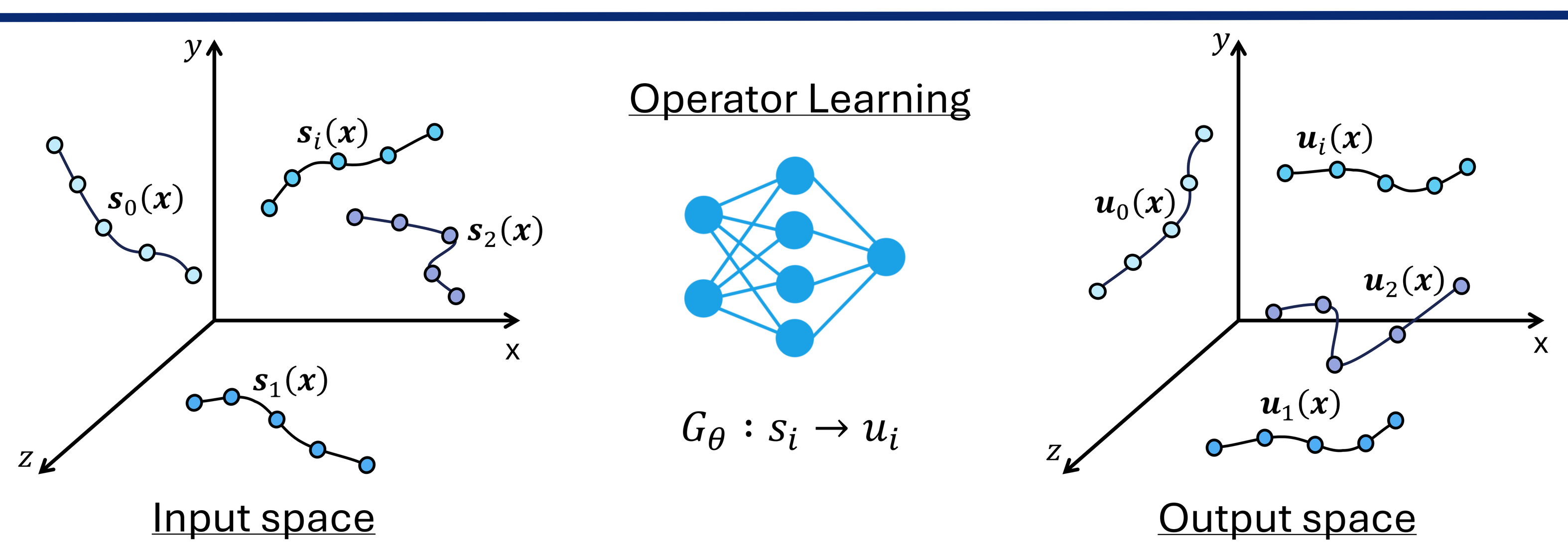

Figure 1. Graphical representation of Infinity-dimensional space mapping proposed using operator learning.

- Leverages the universal approximation capabilities of neural networks to enable efficient DeepONet learning with a flexible architecture.

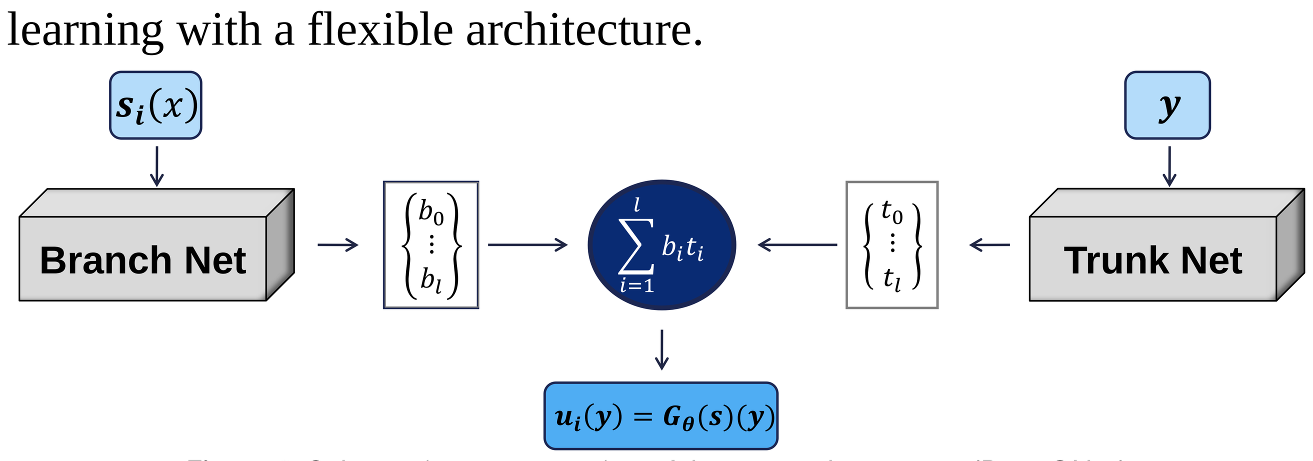

Figure 2. Schematic representation of deep neural operators (DeepONet).

Motivation

ADVANTAGES

- Real time inference
- Low generalization errors for overparametrized neural networks.

CHALLENGES

- Down-sampling input-output spaces misses fine details, reducing accuracy.
- Latent space learning is ineffective for non-dissipative systems.
- Small mini-batches limit the network's generalization ability.

SOLUTION

Developing a scalable framework for data-parallel operator learning for handle high-dimensional PDEs using JAX, optimized for multi-node and multi-GPU HPC systems.

Shallow Water Problem

$$\frac{D\boldsymbol{V}}{Dt} = -f\boldsymbol{k} \times \boldsymbol{V} - \boldsymbol{g}\nabla \boldsymbol{h} + \nu \nabla^2 \boldsymbol{V}$$

$$\frac{Dh}{Dt} = -\boldsymbol{h}\nabla \cdot \boldsymbol{V} + \nu \nabla^2 h$$

- Input space $\to \boldsymbol{h}(t=0, \boldsymbol{x}) \; \therefore (256, 256)$
- Output space $\to \boldsymbol{V}(t, \boldsymbol{x}) \; \therefore (72, 256, 256)$

$$G_\theta(h(t=0,\boldsymbol{x}))(t,\boldsymbol{x}) \to \boldsymbol{V}(t,\boldsymbol{x})$$

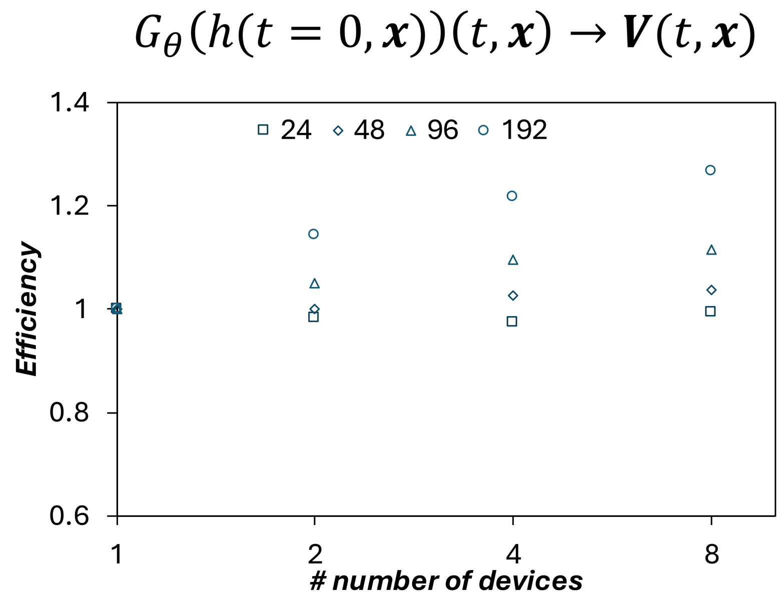

Figure 4. Efficiency values for data-driven shallow water data-parallel training.

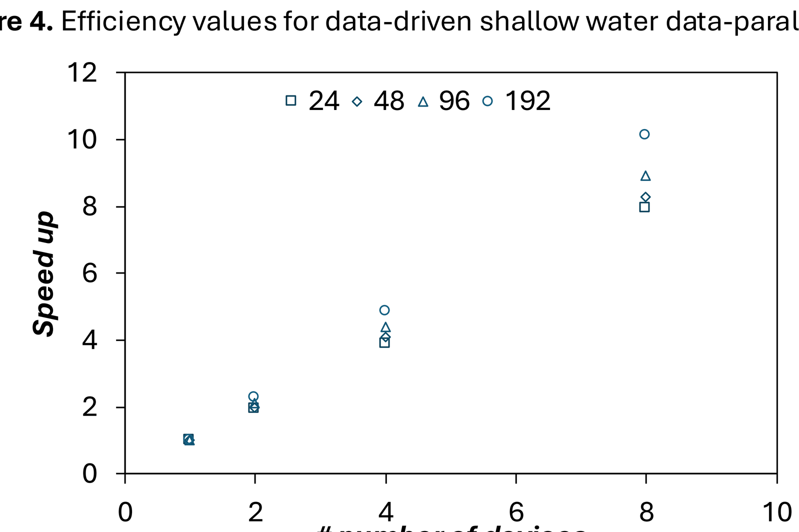

Figure 5. Speed up values for data-driven shallow water data-parallel training.

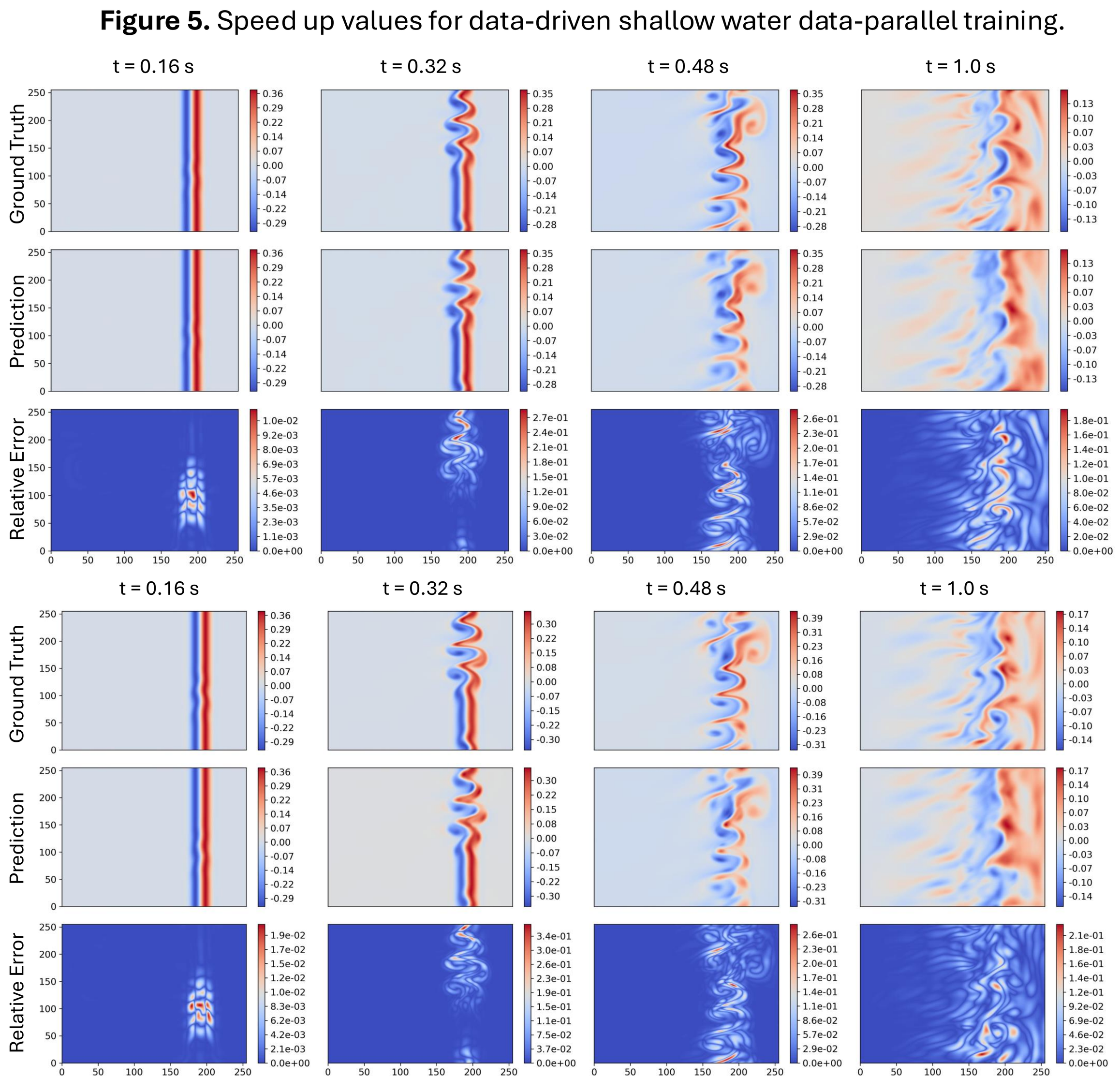

Figure 7. Representative test sample to demonstrate the accuracy of the DeepONet trained models using data-parallel schemes (8 devices).

Data Parallel Training

- Employing the single-problem multiple-data paradigm (SPMD), the training of neural operators is parallelized using the JAX shard map method.

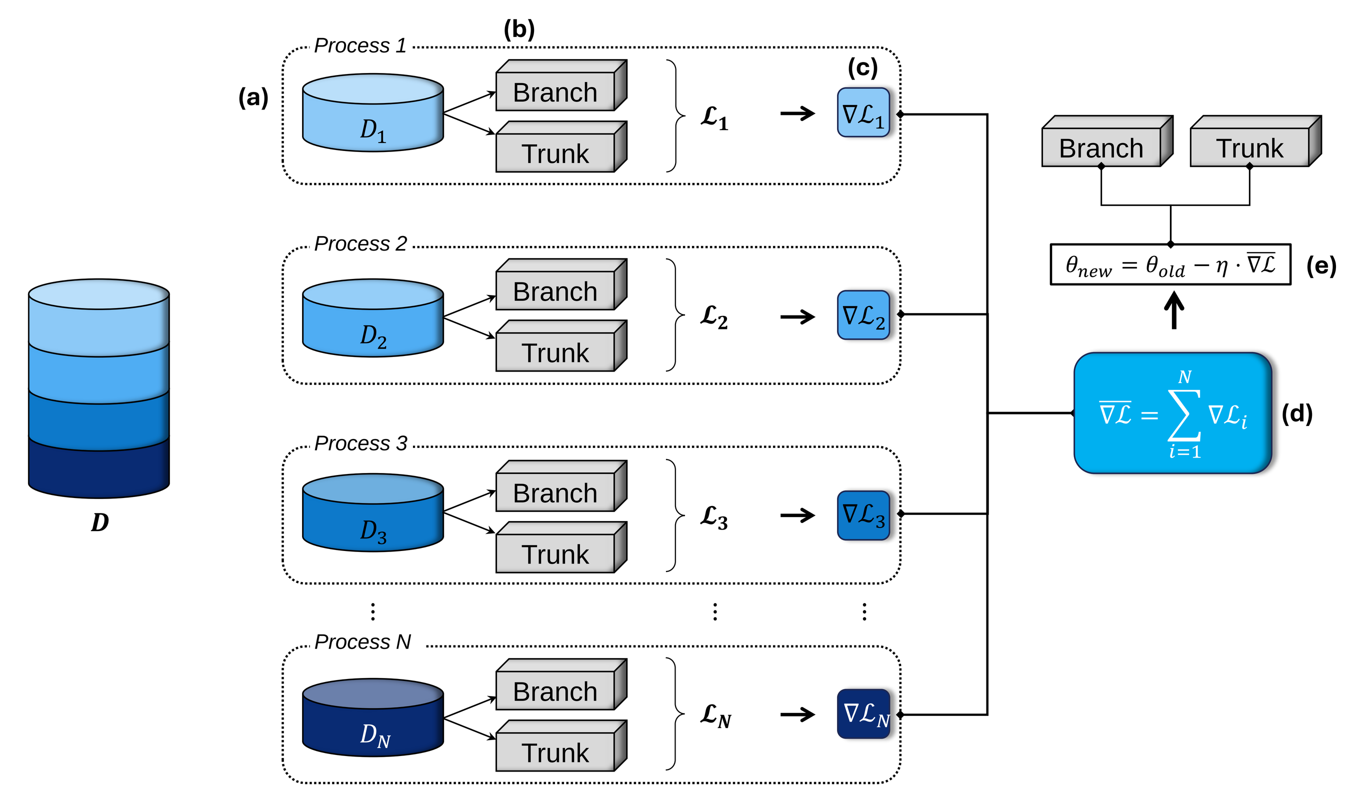

Figure 3. Graphical representation of data parallel framework for operator learning

Conclusions

- The efficiency of the proposed DP training is evident in its superior performance, which exceeds the minimum admissible (0.8) for all cases and is particularly pronounced (superior to 1) in the majority of the evaluated scenarios.
- The speed up consistently surpasses the ideal (linear) scaling and exhibits a marked enhancement as the number of devices increases, thereby demonstrating the effective management of high-dimensional problems.
- Future research directions include the integration of domain decomposition with the proposed data parallel operator learning.
- The code will be available at https://github.com/Centrum-IntelliPhysics

Rockfish Highlights

- Enables efficient training on massive datasets (90GB+), overcoming memory and computational limits.
- For large operator models, the training time has been reduced by up to 10x.

Acknowledgements

NSF, NAIRR Pilot, CAPES, CNPq

Burgers Problem

$$\frac{\partial \boldsymbol{u}}{\partial t} + u\frac{\partial \boldsymbol{u}}{\partial x} = \nu \frac{\partial^2 \boldsymbol{u}}{\partial^2 t}$$

- Input space $\to u(t=0, \boldsymbol{x}) \; \therefore (2000, 101)$
- Output space $\to u(t, \boldsymbol{x}) \; \therefore (2000, 101, 101)$

$$G_\theta(u(t=0,\boldsymbol{x}))(t,\boldsymbol{x}) \to u(t,\boldsymbol{x})$$

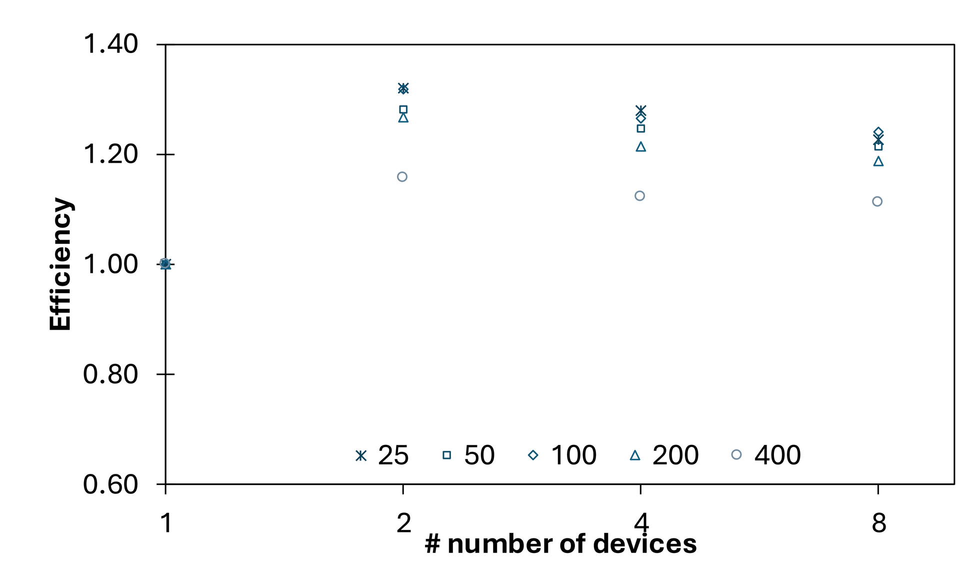

Figure 8. Efficiency values for physics-informed Burgers problem.

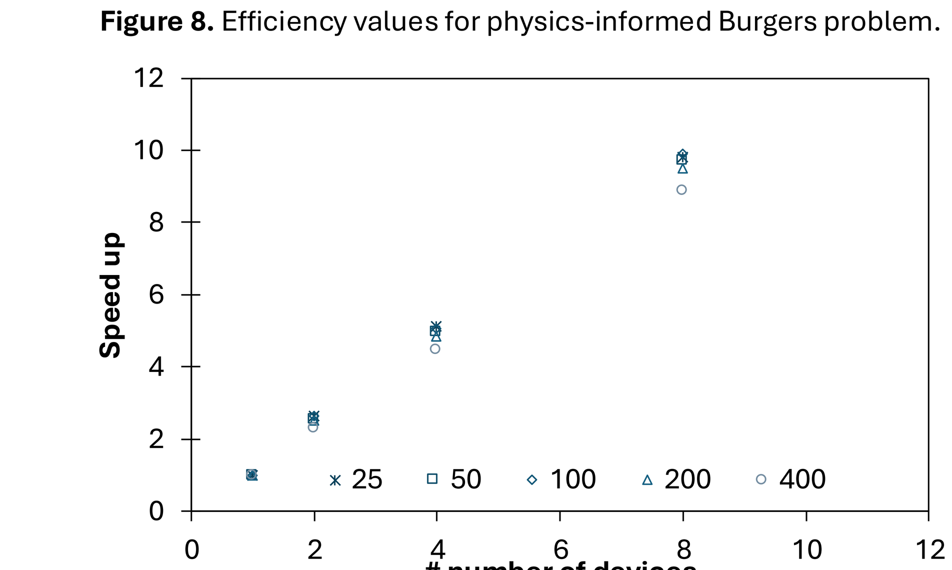

Figure 9. Speed up values for physics-informed Burgers problem.

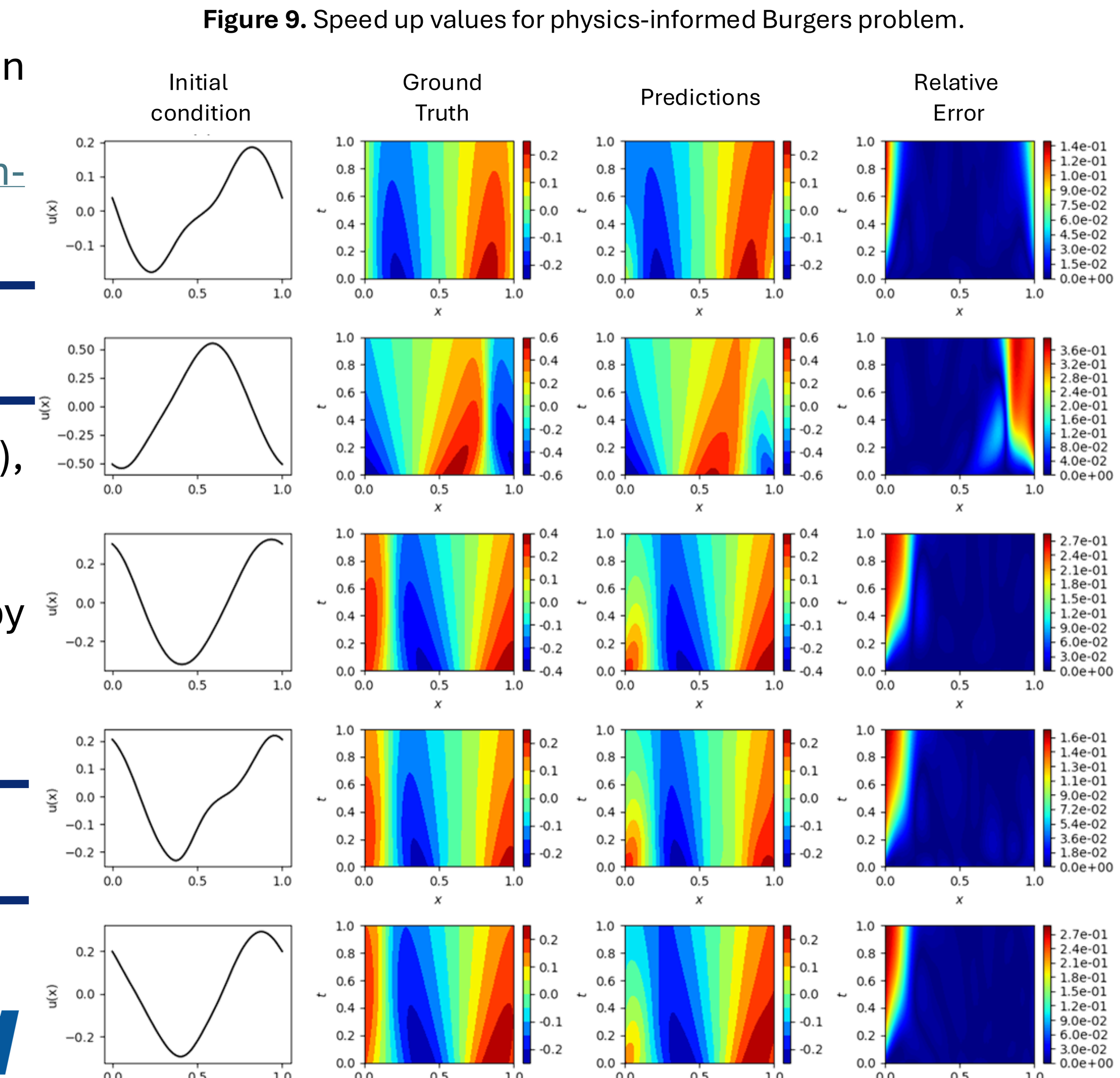

Figure 10. Representative test sample to demonstrate the accuracy of the DeepONet trained models using data-parallel schemes (4 devices).

Contact: Luis F. S. Ribeiro (lsanto13@jh.edu); and Dibakar R. Sarkar (droysar1@jh.edu)