NEURIPS 2026 · SPOTLIGHT

# Learned Lagrangian Models of PDEs via Euler–Lagrange Residual Minimization

Lyra Zhornyak, Eric Forgoston, M. Ani Hsieh

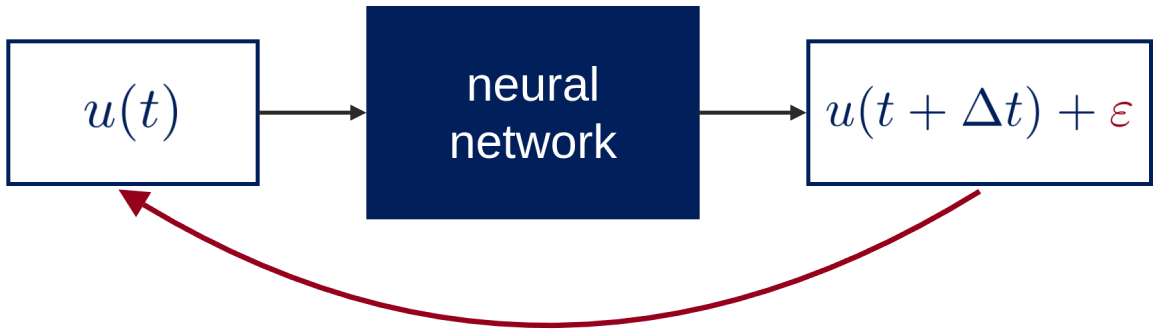
University of Pennsylvania · Montclair State University

zhornyak@engineering.upenn.edu

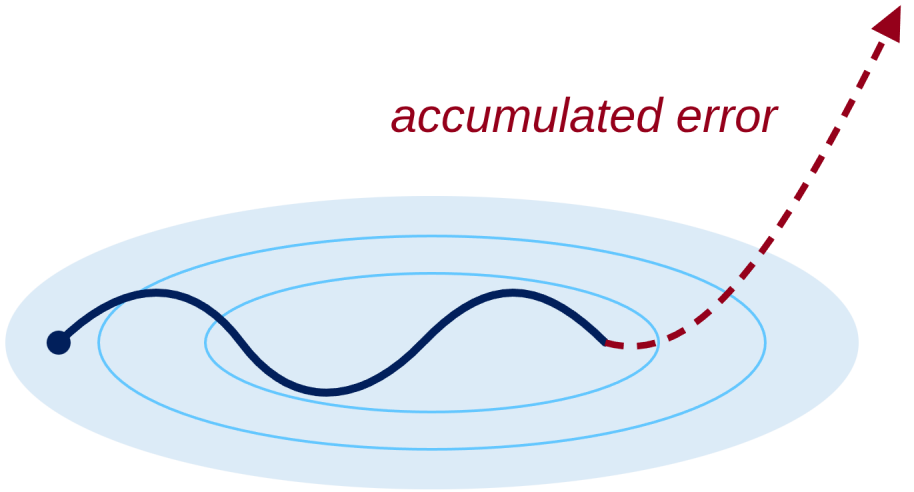
Penn Engineering

# Long-horizon forecasting requires stable iteration

*Most existing methods focus entirely on next-step prediction, allowing small errors to propagate.*




$\varepsilon$ is unconstrained and re-enters the next step



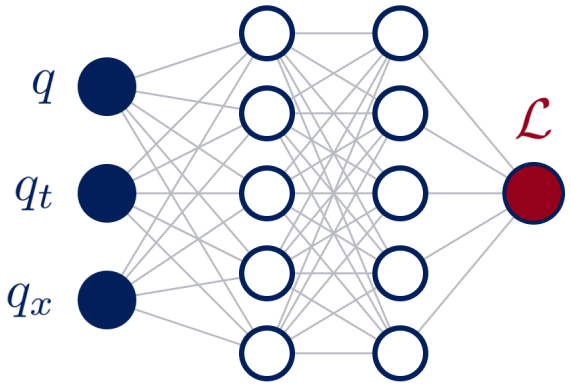

**We instead consider a variational perspective. Separate what the physics is (the Lagrangian density) from how to simulate it.**

$$S[q] = \int_X \mathcal{L}_\theta(q, q_t, q_x)\,\mathrm{d}X, \qquad \delta S = 0$$

# Learn the Lagrangian, use its structure to predict

## Local

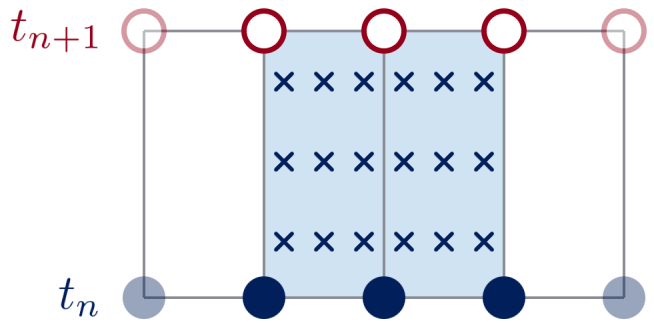
Euler–Lagrange residual




$$R = \frac{\partial \mathcal{L}}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial q_t} - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial \mathcal{L}}{\partial q_x}$$

## Patch

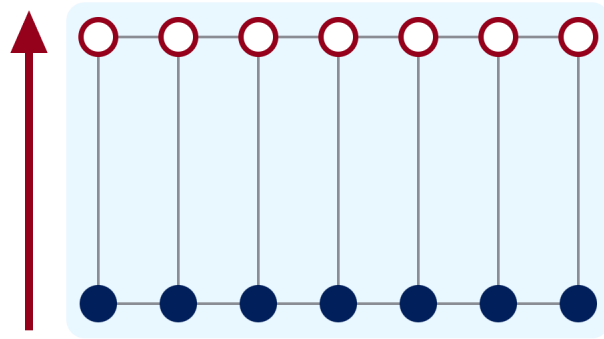
Hermite interpolation




$$J = \sum_{i=1}^{n_\omega} \omega_i \, R^2\left[\hat{q} \,\middle|\, (\tilde{t}_i, \tilde{x}_i)\right]$$

## Global

Damped Newton sweeps

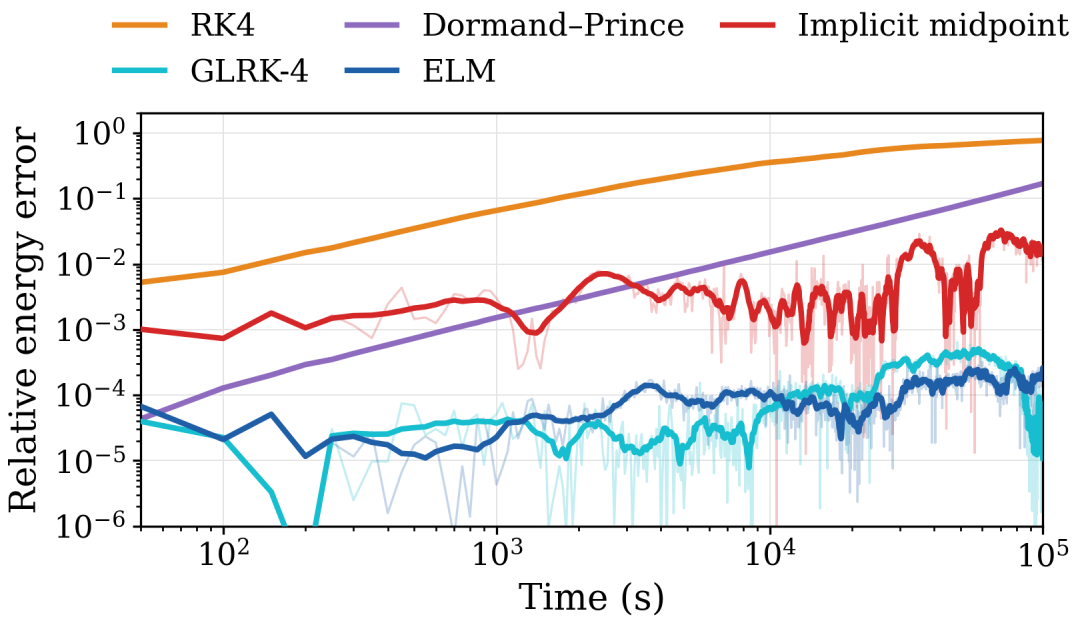


$$\gamma^{(k+1)} = \gamma^{(k)} - \lambda\left(\nabla_\gamma^2 J\right)^{-1} \nabla_\gamma J$$

Model error manifests *only* as a phase error.
Integration error is responsible for any gradual drift in energy.

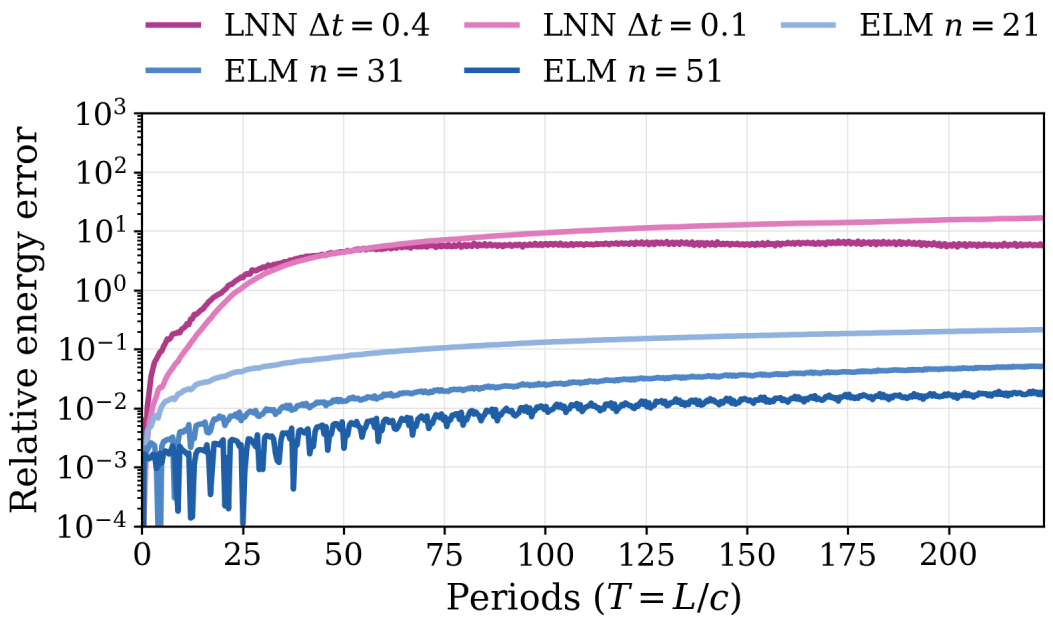
# We achieve low long-horizon energy drift

### Chaotic double pendulum · 100,000 s




Max energy error: **ELM 0.039%** · GLRK-4 0.027% · RK4 72%

### One-dimensional wave · 10,000 s




Max energy error: **ELM 1.9–21%** · LNN 680–1,680%

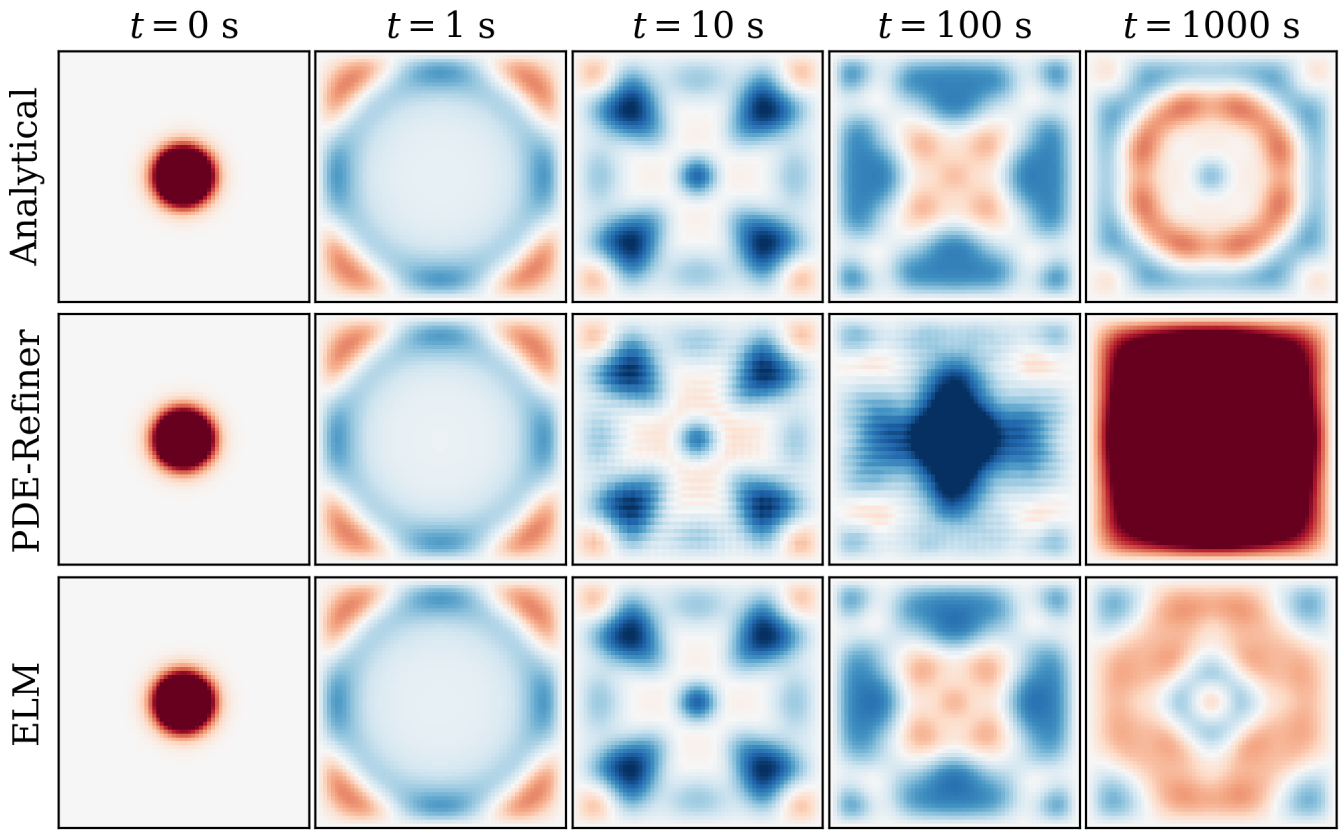
$$J = 0: \quad \mathrm{d}S_d = p(t_1)\,\mathrm{d}q(t_1) - p(t_0)\,\mathrm{d}q(t_0) + \mathcal{O}\left(\Delta t^{2n_\omega+1}\right)$$

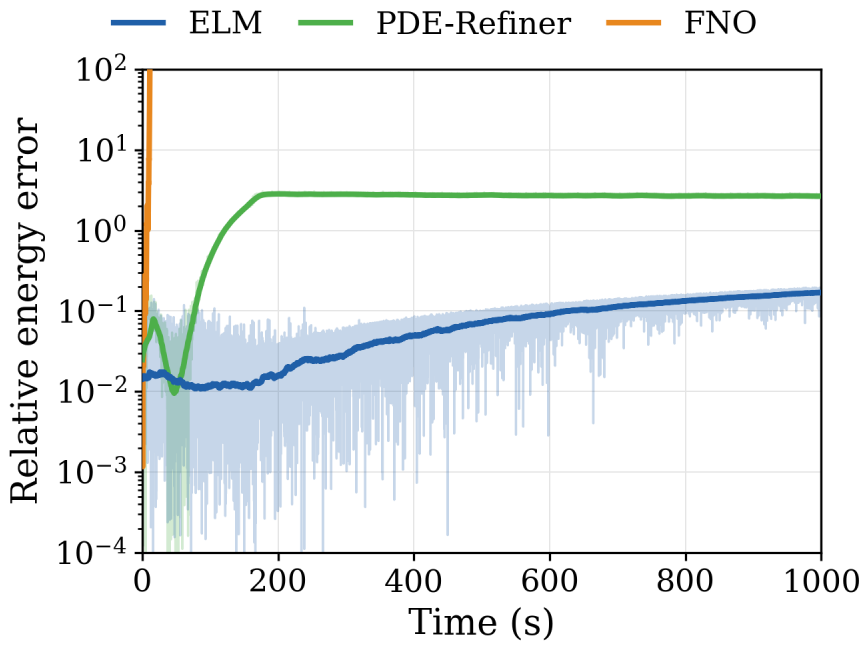
Symplectic to quadrature error; exactly symplectic for $\mathcal{L} = \frac{1}{2}\, q_t^\top M\, q_t - V(q), \quad M$ constant



LNN: Cranmer et al., 2020

# A learned density simulates a 2D wave for 1,000 s




**19%**
ELM energy error

**271%**
PDE-Refiner energy error

**21.9 s**
FNO diverges

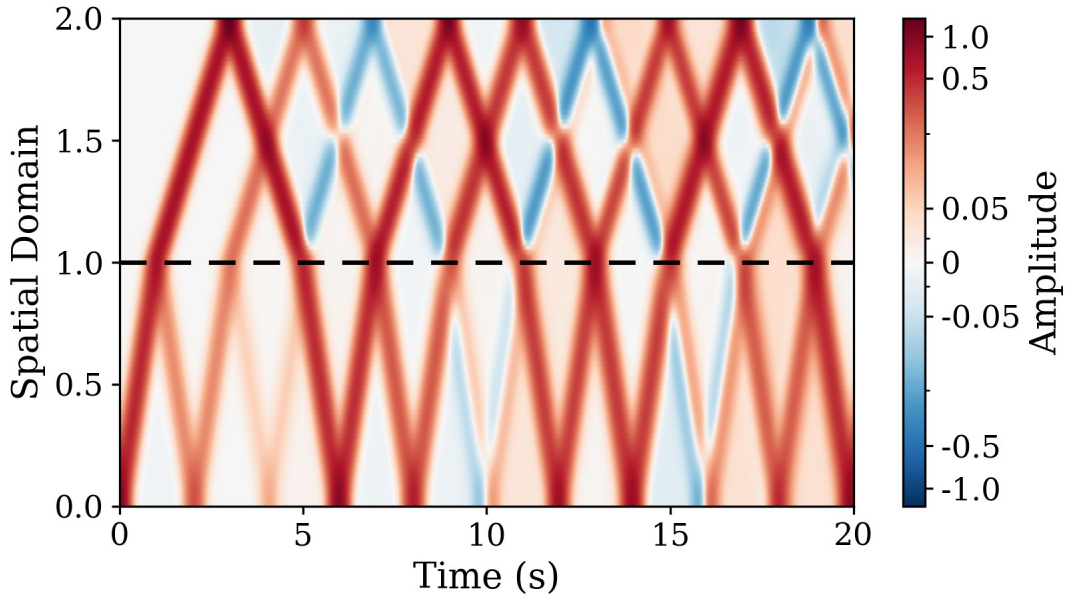
**78,000×**
Fewer parameters than PDE-Refiner



FNO: Li et al., 2021 · PDE-Refiner: Lippe et al., 2023

# One trained model transfers to new domains

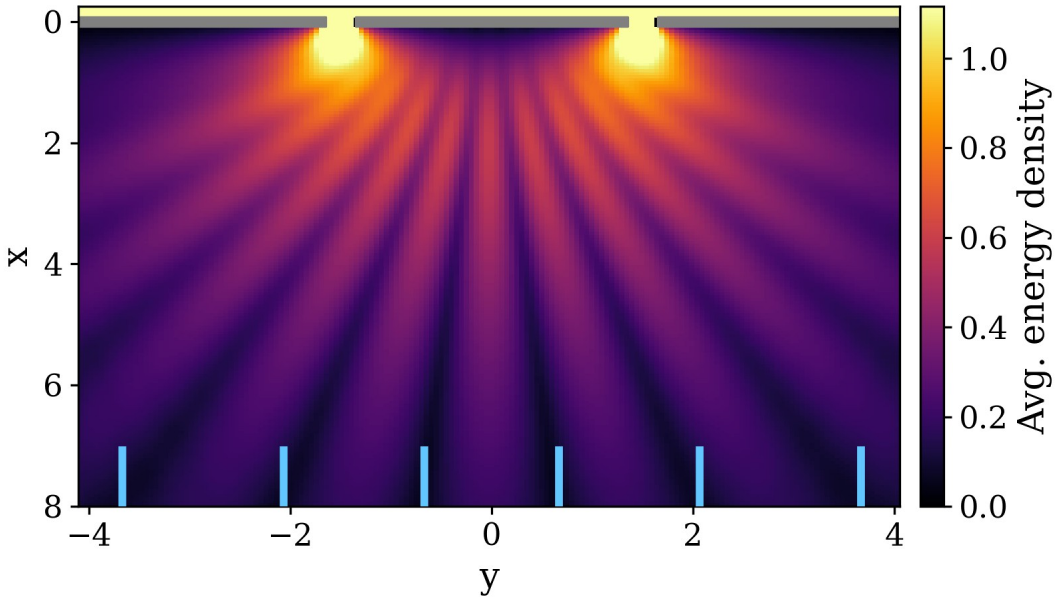
### Impedance interface (1D)




$$\mathcal{L}(x) = \big(1 - \sigma(x)\big)\,\mathcal{L}_1 + \sigma(x)\,\mathcal{L}_2$$

Two media trained separately, blended by a sharp sigmoid: partial reflection and transmission, ~1% energy drift over 20 s.

### Double slit (2D)

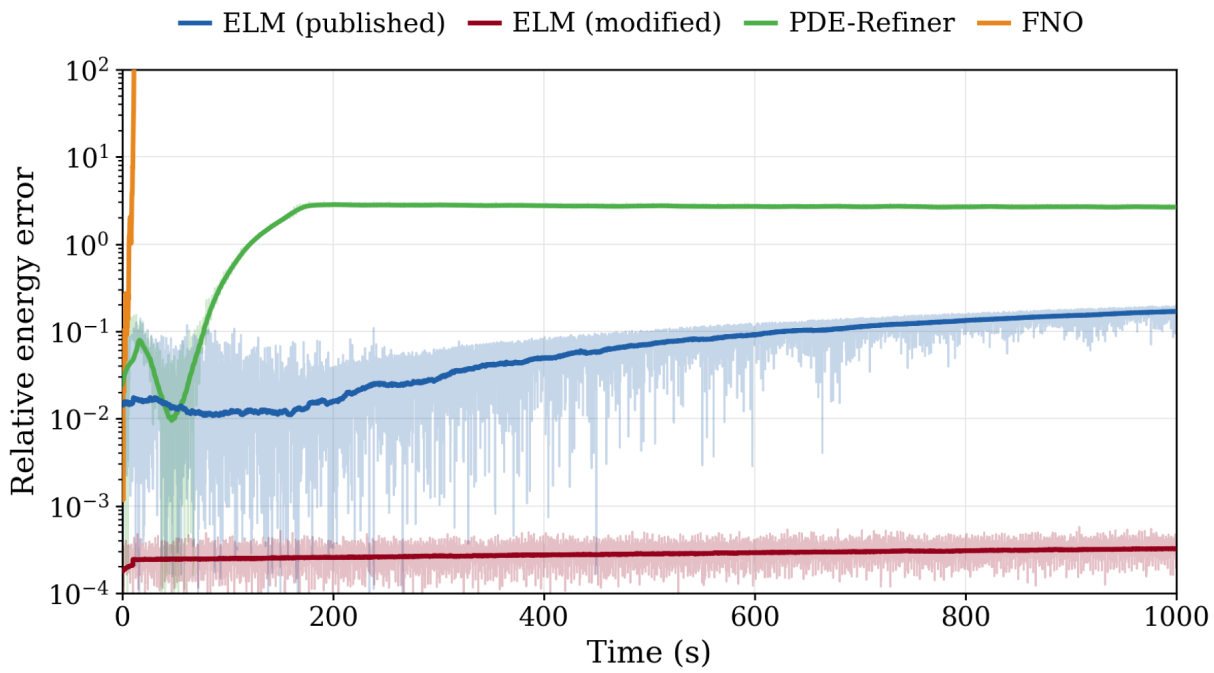



$$q\big|_{\text{barrier}} = 0, \qquad q\big|_{\text{edge}} = A\sin\omega t$$

Slits are constraints on degrees of freedom. Fringes within 6% of the fringe spacing of Fraunhofer theory.

Information about the medium and boundaries enters at inference, without retraining.

# What a symplectic formulation allows




2D wave energy error: same frozen model, same energy computation

With a small modification to the method, we can formally prove symplecticity for the class of PDEs we examine.

**20.6×**

Faster per step on the 2D wave

**5.5–6.4×**

Faster per step on the 1D wave

Results shown are with the same model, hardware and configuration.
The new step is faster and shows bounded energy on long horizons.



FNO: Li et al., 2021 · PDE-Refiner: Lippe et al., 2023

# Learn the Lagrangian, not the next timestep, and transferable dynamics follow.

**Paper & code**

lzhornyak.github.io/Euler-Lagrange-Minimization

Penn Engineering