

Learning Stiffness-Dependent Fluid–Structure Dynamics from Coarse Flow Representations

Chun-Jun Pu^a, Li-Wei Chen^{b,*}, Hai-Bo Huang^{a,*}

^a*University of Science and Technology of China, Hefei 230026, China*

^b*Beijing Institute of Aeronautical Systems Engineering, Beijing 100076, China*

Abstract

This paper develops a data-driven framework for long-term prediction of fluid–structure interaction (FSI) dynamics, focusing on the flow-induced vibration (FIV) of a flexible plate. A stiffness-conditioned neural evolution operator is developed by jointly representing the Eulerian flow field and the Lagrangian structural state. The flexible plate is described by 101 ordered structural tokens carrying nodal coordinates and velocities, while the nondimensional bending stiffness is introduced as a global conditioning variable. Bidirectional cross-attention couples the fluid and structural representations within a hybrid CNN–Transformer architecture. Trained with staged multi-step autoregressive rollouts and symmetry-reflected trajectories, a single operator captures three distinct stiffness-dependent response regimes, including deflected–flapping, deflected, and flapping dynamics. The predicted trajectories preserve the principal flow organization, structural oscillation characteristics, and dominant frequencies, while blind 1000-step rollouts remain bounded. The same operator also interpolates to stiffness values excluded from training.

To reduce the sensitivity of force reconstruction to under-resolved near-wall gradients, we construct a differentiable aerodynamic-force reconstruction module based on the derivative-moment transformation (DMT). By replacing conventional wall-stress surface integrals with an enclosed 2D curve integral sur-

*Corresponding author

Email addresses: liweichen.aero@gmail.com (Li-Wei Chen), huanghb@ustc.edu.cn (Hai-Bo Huang)

rounding the core vortex region, the aerodynamic lift and drag forces can be accurately reconstructed. This differentiable integration is implemented through a Signed Distance Function (SDF) and a smoothed Dirac-delta formulation, preserving gradient flow throughout the force-reconstruction procedure. Overall, the proposed framework provides an accurate and differentiable surrogate for stiffness-dependent FSI dynamics, offering a foundation for efficient parameter studies and future stiffness optimization in flow-energy-harvesting applications.

Keywords: Fluid–structure interaction; Flexible plates; Fluid-induced vibration; Neural evolution operator; Aerodynamic force reconstruction; Derivative-moment transformation

1. Introduction

Fluid–structure interaction (FSI) involving slender and flexible bodies arises in a wide range of engineering and biological systems, including aeroelastic structures, energy-harvesting devices, aquatic propulsion, and flow-responsive materials. These problems are governed by a nonlinear feedback loop: the fluid loading deforms the structure, while the resulting structural motion continuously modifies the fluid boundary and reorganizes the surrounding vortical flow. A representative configuration is the inverted flexible plate, whose downstream end is clamped while its upstream end is free. Depending on its nondimensional bending stiffness and other fluid–structural parameters, the plate can exhibit straight, deflected, periodic flapping, and more complex oscillatory states [1–3]. The transitions between these regimes are accompanied by substantial changes in the mean plate configuration, oscillation amplitude, dominant frequency, wake topology, and aerodynamic loading. These stiffness-dependent changes are also directly relevant to flow-energy-harvesting systems, whose performance depends strongly on the amplitude, frequency, and loading characteristics of the structural motion. Resolving such stiffness-dependent dynamics therefore requires a model that simultaneously represents the Eulerian flow field, the moving Lagrangian structure, and their strongly coupled temporal evolution.

High-fidelity numerical methods can reproduce these phenomena but remain computationally demanding when long trajectories or densely sampled parameter spaces are required. This cost becomes particularly restrictive in repeated-query applications such as parameter inference, uncertainty quantification, flow control, and aeroelastic design. Data-driven surrogate models provide a possible alternative by learning the evolution or solution operator directly from high-fidelity simulations. Early operator-learning frameworks, represented by DeepONet and the Fourier neural operator (FNO), demonstrated that mappings between infinite-dimensional function spaces can be approximated without solving each new PDE instance from the beginning [4, 5]. Graph-based simulators further extended learned physical modeling to irregular meshes and interacting discretized objects [6]. More recently, attention-based neural operators, including OFormer and GNOT, have provided flexible representations of nonlocal spatial interactions, irregular sampling, multiple input functions, and multiscale PDE solutions [7, 8].

Applying these ideas to FSI is more difficult than predicting a fluid field over a fixed domain. The structural geometry is itself part of the evolving physical state and determines where momentum is exchanged between the fluid and the solid. A fluid-only Eulerian surrogate must infer this moving interface implicitly from the velocity and pressure fields, so small interface errors can be reintroduced at every autoregressive step. Existing studies have explored CNN-based moving-boundary prediction, graph representations of coupled systems, and FNO-based hybrid FSI solvers [9–11]. Differentiable hybrid approaches have also embedded coarse numerical FSI models within trainable recurrent architectures, allowing gradients to propagate through the coupled rollout trajectory [12]. In parallel, codomain-attention operators have shown that tokenizing and exchanging information among different physical variables can improve the representation of multiphysics systems [13]. Nevertheless, explicit learned coupling between a dense Eulerian flow representation and an ordered Lagrangian flexible structure remains comparatively underexplored. In particular, directly concatenating a geometry mask with the fluid channels does not preserve the distinct

physical identities and discretizations of the fluid and structural states.

A second challenge concerns long-horizon prediction. Neural PDE models are often optimized using one-step or short-window errors, whereas practical FSI applications require repeated reuse of predicted states. Even a small one-step bias can accumulate during autoregressive rollout, leading to phase drift, blurred vortical structures, duplicated moving interfaces, or collapse toward an unphysical steady state. Multi-step message-passing solvers, refinement-based models, and denoising operator Transformers have been proposed to improve the stability of long PDE trajectories [14–16]. Recent systematic comparisons of one-step, differentially unrolled, and gradient-truncated unrolled training further show that temporal unrolling improves autoregressive inference by exposing the model to its own rollout distribution. Full temporal differentiability provides additional benefits within matched training configurations, while non-differentiable unrolling can still retain much of the improvement associated with reducing the rollout distribution shift [17]. Benchmark studies have further shown that autoregressive performance depends strongly on the training horizon, noise treatment, temporal resolution, and underlying PDE dynamics [18].

For FSI, this difficulty is compounded by the need to preserve not only pointwise fields but also the coupled oscillatory attractor, including the mean deformation, dominant frequency, phase-space structure, and wake organization. Moreover, most existing surrogate models are demonstrated for a single physical condition or a limited response regime. Whether one conditional evolution operator can reproduce multiple stiffness-dependent FSI regimes while remaining stable under a fully autoregressive rollout is still insufficiently established.

A third challenge is the recovery of aerodynamic forces from neural-predicted flow fields. Lift and drag are global functionals of the coupled solution and provide a more integrated physical assessment than pointwise field errors. Conventional force evaluation uses pressure and viscous tractions on the instantaneous solid boundary. On a coarse neural grid, however, the thin boundary layer and near-wall gradients are generally under-resolved, making direct surface-stress

integration sensitive to local smoothing and differentiation errors.

Control-surface formulations offer an alternative by transferring force evaluation from the body surface to a better-resolved outer-flow region. The derivative-moment transformation (DMT) expresses the unsteady force through vorticity moments and flux terms defined over a control domain and its enclosing contour [19, 20]. Control-surface force measurements and more recent weighted-integral developments have demonstrated the accuracy and diagnostic value of this class of formulations [21, 22].

For integration with neural models, force accuracy alone is not sufficient: the force-reconstruction operator should also retain derivatives with respect to the predicted velocity and pressure fields. Differentiable surrogate-based design has demonstrated the value of combining level-set geometry descriptions, automatic differentiation, and neural flow predictions [23]. Fully differentiable and GPU-accelerated CFD platforms have further enabled gradient-based inverse modeling and optimization through complete turbulence and FSI simulations [24]. Such approaches, however, generally require differentiating through the numerical solver itself. A complementary capability is to construct a differentiable force readout directly on coarse flow fields produced by a learned evolution model. This requires a continuous grid-compatible representation of the control contour and a regularized integration rule that preserves both the DMT force identity and a usable gradient path.

In this work, we develop a stiffness-conditioned neural evolution framework for long-horizon prediction and force-based assessment of a two-dimensional inverted flexible plate at $Re = 200$. The model evolves the fluid velocity and pressure on a fixed Eulerian grid while independently tracking the nodal coordinates and velocities of 101 ordered Lagrangian plate points. A convolutional encoder extracts spatial features from the flow field, which are then organized into a compact sequence of fluid tokens for Transformer attention. The position and velocity of each plate point are jointly embedded as a structural token, while the nondimensional bending stiffness is introduced through an additional global conditioning token. Bidirectional cross-attention transfers information from the

fluid to the structure and from the updated structure back to the fluid.

The training data comprise 36 original stiffness-dependent trajectories, supplemented by eight reflected trajectories. To improve recursive prediction, the model is first trained using four-step rollouts and is subsequently fine-tuned using eight-step rollouts. Starting from a single coupled state, the complete fluid–structure state is advanced recursively without intermediate ground-truth correction. The resulting model provides a unified description of three stiffness-dependent response regimes, with particular attention to the deflected and flapping dynamics.

The principal contributions of this work are summarized as follows.

(1) Coupled Eulerian–Lagrangian representation for flexible-plate FSI.

A neural evolution operator is developed to jointly represent the Eulerian flow field and the Lagrangian flexible structure, with bidirectional cross-attention coupling the two physical states. The structural state includes both nodal coordinates and velocities, while the nondimensional bending stiffness is introduced as a global conditioning variable.

(2) Unified autoregressive prediction across multiple stiffness-dependent FSI regimes.

A single stiffness-conditioned operator is trained over 36 bending stiffnesses and captures the deflected–flapping, deflected, and flapping regimes in a fully autoregressive manner. The predicted dynamics remain stable over long rollouts and interpolate to unseen stiffness conditions.

(3) Differentiable aerodynamic-force reconstruction from coarse neural flow fields.

A differentiable force-reconstruction operator is developed by combining the DMT with a fixed control contour represented by a signed distance function and a smoothed Dirac-delta formulation. The resulting formulation recovers aerodynamic forces from coarse predicted flow fields while retaining gradients with respect to the flow variables.

The remainder of this paper is organized as follows. Section 2 describes the physical FSI system, numerical solver, multi-stiffness dataset, data augmentation, and autoregressive sequence construction. Section 3 presents the coupled CNN–Transformer evolution operator, the staged rollout-training strategy, and

the differentiable DMT force-reconstruction method. Section 4 evaluates the prediction accuracy across the stiffness parameter space, representative multi-regime dynamics, aerodynamic-force reconstruction, long-horizon rollout behavior, and interpolation to unseen stiffnesses. The main conclusions and limitations are summarized in Section 5.

2. Problem Description and FSI System

We consider a two-dimensional flexible plate immersed in an incompressible viscous fluid flow. The structural boundary of the flexible plate is discretized into a set of $N_s = 101$ Lagrangian points. Each Lagrangian point $i \in \{1, \dots, N_s\}$ is characterized by its instantaneous Cartesian coordinates $\mathbf{x}_i(t) = [x_i(t), y_i(t)]^T$ and structural velocity $[u_{s,i}(t), v_{s,i}(t)]^T$. A single nondimensional bending stiffness κ is assigned uniformly to the plate in each simulation. The fluid state is defined on an Eulerian grid covering the computational domain Ω and consists of the velocity $\mathbf{u} = [u, v]^T$ and pressure p , while the vorticity ω is derived from the velocity field. The coupling at the fluid-solid interface dictates that the fluid exerts aerodynamic forces on the plate, causing large-amplitude nonlinear oscillations, which in turn modulate the vortex shedding patterns in the wake, exhibiting multiple distinct limit-cycle oscillation (LCO) modes as the nondimensional bending stiffness varies.

The physical configuration and the computational domains are illustrated in Fig. 1. The flexible plate is aligned with the incoming uniform flow in its undeformed configuration, with its downstream end clamped and its upstream end free. The full non-uniform CFD domain is used to generate the reference trajectories, whereas a near-body window containing the plate and the principal wake structures is extracted for neural prediction.

2.1. Physical configuration and governing equations

The present study focuses on the flow-induced vibration of an inverted flexible plate. The plate has a nondimensional length $L = 1$ and is initially aligned

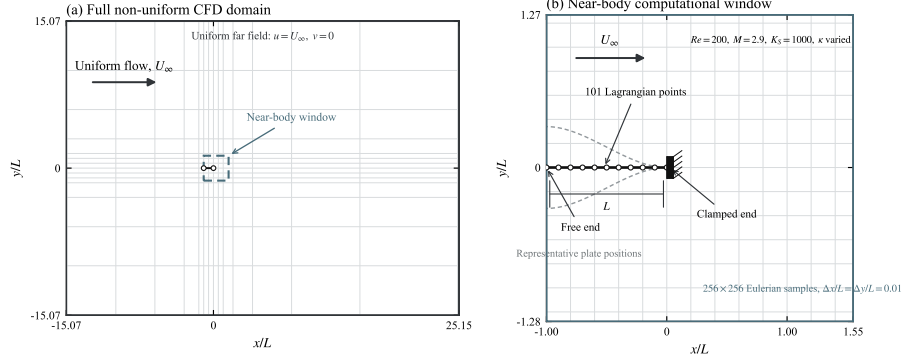


Figure 1: Physical configuration and computational domains of the flexible-plate FSI problem. (a) Full non-uniform CFD domain and the location of the near-body prediction window. (b) Near-body Eulerian window used by the neural evolution operator.

with the incoming flow. Its downstream end is clamped at $(x, y) = (0, 0)$, whereas its upstream end is free. Thus, the undeformed plate occupies $-1 \leq x \leq 0$ and $y = 0$. A uniform flow with velocity $\mathbf{u}_\infty = (U_\infty, 0)$ is imposed in the positive x direction. This inverted configuration permits large-amplitude deformation of the upstream free end and produces a strongly coupled interaction between the plate motion and the wake dynamics.

The fluid state is described by

$$\mathbf{q}_f(\mathbf{x}, t) = [u(\mathbf{x}, t), v(\mathbf{x}, t), p(\mathbf{x}, t)]^T, \quad (1)$$

where u and v are the Cartesian velocity components and p denotes the pressure channel retained in the learning problem. Its normalization relative to the pressure appearing in the governing equations is specified below.

$$\omega = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}. \quad (2)$$

The vorticity is not evolved as an independent variable by the neural model; it is derived from the velocity field and is used to characterize the wake regimes.

Using the plate length L , reference velocity U_∞ , fluid density ρ_f , and convective time scale L/U_∞ , we define the pressure appearing in the governing equations as $p^\star = p_{\text{dim}}/(\rho_f U_\infty^2)$. The dimensionless incompressible Navier–Stokes

equations are

$$\nabla \cdot \mathbf{u} = 0, \quad (3)$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p^* + \frac{1}{Re} \nabla^2 \mathbf{u}, \quad (4)$$

where

$$Re = \frac{U_\infty L}{\nu} = 200. \quad (5)$$

and ν is the kinematic viscosity. The dimensionless fluid stress tensor is written as

$$\boldsymbol{\sigma}_f = -p^* \mathbf{I} + \frac{1}{Re} (\nabla \mathbf{u} + \nabla \mathbf{u}^T). \quad (6)$$

For the learning problem and the subsequent aerodynamic-force reconstruction, the retained pressure channel p is normalized by the dynamic pressure $q_\infty = \frac{1}{2} \rho_f U_\infty^2$. Consequently, $p = p_{\text{dim}}/q_\infty = 2p^*$. This distinction is purely a normalization convention and does not change the underlying dimensional pressure field.

The plate centreline is discretized by $N_s = 101$ Lagrangian nodes connected by 100 geometrically nonlinear beam elements. The structural state associated with node i is defined as

$$\mathbf{s}_i(t) = [x_i(t), y_i(t), u_{s,i}(t), v_{s,i}(t)]^T, \quad \mathbf{X}_s(t) = \{\mathbf{s}_i(t)\}_{i=1}^{N_s} \in \mathbb{R}^{N_s \times 4}. \quad (7)$$

Although the structural solver retains the complete finite-element degrees of freedom, only the in-plane nodal coordinates and velocities are retained in the learning problem.

The principal control parameter is the nondimensional bending stiffness

$$\kappa \equiv K_B = \frac{B}{\rho_f U_\infty^2 L^3}, \quad (8)$$

where B denotes the dimensional bending rigidity. The nondimensional axial or extensional stiffness is held fixed at

$$K_S = 1000. \quad (9)$$

Although the formulation permits a spatially varying local stiffness k_i , all Lagrangian points in each simulation are assigned the same value, such that

$k_i = \kappa$ for $i = 1, \dots, N_s$. The nondimensional mass ratio and Poisson ratio are fixed at

$$m^* = \frac{\rho_s h}{\rho_f L} = 2.9, \quad \nu_s = 0.25, \quad (10)$$

respectively, while the structural damping coefficients are set to zero. The fluid traction is transferred to the Lagrangian structure, and the no-slip condition requires the interpolated fluid velocity to match the local structural velocity on the moving plate. Consequently, the plate deformation modulates the vortex formation and shedding process, while the resulting fluid loading drives the structural motion.

2.2. Coupled numerical solver

The high-fidelity trajectories are generated using a partitioned FSI solver. The incompressible flow is advanced by a D2Q9 multiple-relaxation-time lattice Boltzmann method on a nonuniform Cartesian grid. The full fluid grid contains 1468×985 points and covers

$$-15.073 \leq x \leq 25.150, \quad -15.073 \leq y \leq 15.073. \quad (11)$$

The mesh is refined in the vicinity of the plate and gradually stretched towards the far field. A uniform far-field state is imposed at the outer boundaries.

The fluid and structure are coupled through an immersed-boundary forcing procedure. At every coupled time step, the fluid variables are first advanced, the immersed-boundary force is evaluated and transferred to the plate, and the structural state is subsequently updated. The plate dynamics are solved using a geometrically nonlinear finite-element formulation with a Newmark time-integration scheme. The fluid reference values used by the source solver are $L = 1$, $U_\infty = 0.025$, and $\rho_f = 1$, yielding a convective reference time $T_{\text{ref}} = L/U_\infty = 40$. The internal solver time step is $\Delta t = 0.01$, corresponding to the dimensionless interval

$$\frac{\Delta t}{T_{\text{ref}}} = 2.5 \times 10^{-4}. \quad (12)$$

The flow and structural states used for data construction are written every 0.1 dimensionless time units. Before storage, the nodal coordinates are scaled by L , both the fluid and structural velocity components are scaled by U_∞ , and the retained pressure channel p is scaled by the dynamic pressure $q_\infty = \frac{1}{2}\rho_f U_\infty^2$. Thus, the stored pressure satisfies $p = 2p^*$. Dimensional quantities are denoted explicitly only when needed hereafter.

2.3. Multi-stiffness FSI dataset

Independent simulations are conducted by varying only the nondimensional bending-stiffness parameter κ , while all other fluid and structural parameters remain unchanged. The reference dataset consists of 36 independently simulated flexible-plate cases, with κ uniformly sampled from 0.05 to 0.40 at an interval of 0.01. The set of independently simulated stiffnesses is therefore defined as

$$\mathcal{K}_{\text{train}} = \{0.05 + 0.01j \mid j = 0, 1, \dots, 35\}. \quad (13)$$

Collectively, these 36 stiffness conditions represent the characteristic plate vibration modes and coupled response regimes considered in this study. The reflection-based augmentation used to construct the final training set is described in Section 2.4.

2.3.1. Stiffness-dependent response regimes

Previous studies have shown that the dynamical response of an isolated inverted flexible plate can be classified into several characteristic regimes, including the straight, flapping, deflected, and deflected-flapping modes. In the present study, the selected bending-stiffness range focuses on three dynamically active regimes: the deflected-flapping, deflected, and flapping modes.

The structural response is characterized by the peak-to-peak transverse displacement A of the free tip, reported in normalized form as

$$\frac{A}{L} = \frac{\max_{t \in \mathcal{T}} y_{\text{tip}}(t) - \min_{t \in \mathcal{T}} y_{\text{tip}}(t)}{L}, \quad (14)$$

where L is the undeformed plate length and $\mathcal{T}$ denotes the statistically stationary sampling interval.

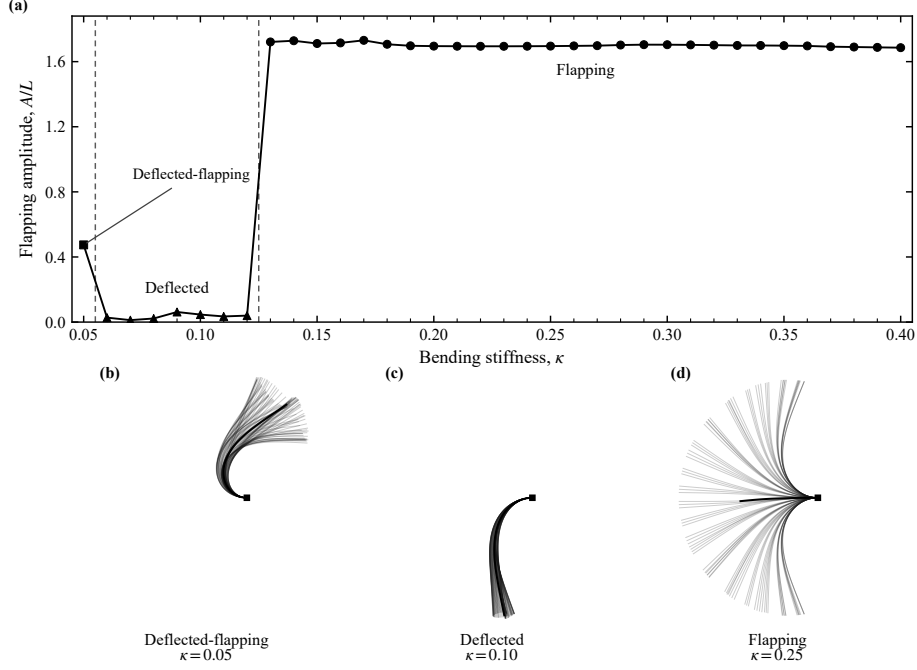


Figure 2: Stiffness-dependent response regimes of the isolated inverted flexible plate. (a) Normalized peak-to-peak transverse displacement of the free tip. (b–d) Representative full-body envelopes of the deflected–flapping, deflected, and flapping modes, respectively. The dashed lines visually separate the response regimes represented by the discrete stiffness samples.

As shown in figure 2, the densely sampled stiffness cases reveal three distinct structural-response regimes. At $\kappa = 0.05$, the plate exhibits the deflected–flapping mode, characterized by a strongly deflected mean configuration accompanied by finite-amplitude oscillations. For $0.06 \leq \kappa \leq 0.12$, the response is classified as the deflected mode, in which the plate oscillates weakly around a bent mean configuration. When $0.13 \leq \kappa \leq 0.40$, the plate undergoes sustained large-amplitude flapping across the undeformed centerline. These three regimes collectively describe the characteristic stiffness-dependent structural responses covered by the present dataset.

For each value of κ , a statistically developed segment over

$$250.1 \leq t \leq 280.0 \quad (15)$$

is extracted from the original FSI simulation. The raw output interval is $\Delta t_{\text{raw}} = 0.1$, giving 300 synchronized fluid–structure snapshots for each condition. This late-time interval is selected to focus on the developed stiffness-dependent oscillatory responses rather than the initial transient from the undeformed state. For every stiffness condition, the selected trajectory contains at least two complete oscillation cycles, thereby retaining the essential periodic evolution of both the plate motion and the associated wake dynamics.

For neural prediction, a local region surrounding the moving plate and its near wake is extracted from the full numerical domain. The learning domain is

$$\Omega_{\text{ML}} = [-1.00, 1.55] \times [-1.28, 1.27] \quad (16)$$

and is represented on a uniform 256×256 Cartesian grid. Each stored CFD snapshot contains the velocity components u and v , the pressure p , and a binary geometry mask. Since the moving plate is represented independently by its Lagrangian state, the geometry mask is not supplied to or predicted by the neural evolution operator. The fluid state used in the learning problem is therefore

$$\mathbf{Q}_f^n = [u^n, v^n, p^n] \in \mathbb{R}^{3 \times 256 \times 256}. \quad (17)$$

The corresponding structural state contains the nodal positions and velocities,

$$\mathbf{X}_s^n = [x_i^n, y_i^n, u_{s,i}^n, v_{s,i}^n]_{i=1}^{101} \in \mathbb{R}^{101 \times 4}. \quad (18)$$

To reduce temporal redundancy, one snapshot is retained from every five raw outputs. The effective time interval of the learning problem is therefore

$$\Delta t_{\text{ML}} = 5\Delta t_{\text{raw}} = 0.5. \quad (19)$$

Each stiffness condition consequently contains 60 frames at

$$t = 250.1, 250.6, \dots, 279.6. \quad (20)$$

The 36 original stiffness conditions contain 2160 synchronized fluid–structure states. Eight additional reflected trajectories, each containing 60 frames, contribute another 480 states. The augmented training set therefore contains 44 trajectories and 2640 synchronized fluid–structure states in total.

Table 1: Summary of the physical system and multi-stiffness dataset.

Quantity	Value
Reynolds number	$Re = 200$
Plate length	$L = 1$
Mass ratio	$m^* = 2.9$
Poisson ratio	$\nu_s = 0.25$
Fixed extensional stiffness	$K_S = 1000$
Training bending stiffnesses	0.05–0.40, $\Delta\kappa = 0.01$ (36 cases)
Structural discretization	101 nodes, 100 beam elements
Full fluid grid	1468×985
Learning grid	256×256
Learning-domain extent	$[-1.00, 1.55] \times [-1.28, 1.27]$
Raw snapshots per stiffness	300
Raw output interval	$\Delta t_{\text{raw}} = 0.1$
Temporal subsampling factor	5
Learning interval	$\Delta t_{\text{ML}} = 0.5$
Subsampled frames per stiffness	60
Predicted fluid variables	u, v, p
Predicted structural variables	$x_i, y_i, u_{s,i}, v_{s,i}, i = 1, \dots, 101$

2.4. Preprocessing, reflection augmentation, and autoregressive sequence construction

To reduce the directional imbalance associated with the deflected responses, reflection-based data augmentation is applied to the trajectories at

$$\mathcal{K}_{\text{mirror}} = \{0.05, 0.06, \dots, 0.12\}.$$

For a reflection about the plate centreline, the fluid variables are transformed according to

$$u^{\text{mir}}(x, y, t) = u(x, -y, t), \quad v^{\text{mir}}(x, y, t) = -v(x, -y, t), \quad p^{\text{mir}}(x, y, t) = p(x, -y, t), \quad (21)$$

while the structural state is transformed as

$$x_i^{\text{mir}} = x_i, \quad y_i^{\text{mir}} = -y_i, \quad u_{s,i}^{\text{mir}} = u_{s,i}, \quad v_{s,i}^{\text{mir}} = -v_{s,i}. \quad (22)$$

The stiffness value is unchanged under reflection. This procedure introduces eight additional trajectories without adding new stiffness values, resulting in 36 original trajectories and eight reflected trajectories for training.

The velocity and pressure channels have different numerical scales. Global minimum and maximum values are evaluated separately for $q_c \in \{u, v, p\}$ over all 44 original and reflected training trajectories. The fluid channels are then normalized according to

$$\tilde{q}_c = \frac{q_c - q_{c,\min}}{q_{c,\max} - q_{c,\min} + 10^{-8}}. \quad (23)$$

The structural coordinates and velocities are retained in their nondimensional physical scales without an additional dataset-wise min-max normalization. The stiffness condition is mapped to $[-1, 1]$ using

$$\tilde{\kappa} = 2 \frac{\kappa - \kappa_{\min}}{\kappa_{\max} - \kappa_{\min}} - 1, \quad \kappa_{\min} = 0.05, \quad \kappa_{\max} = 0.40. \quad (24)$$

Training samples are constructed using multi-step autoregressive rollout windows. Starting from an initial coupled state $(\mathbf{Q}_f^n, \mathbf{X}_s^n)$, the model recursively predicts

$$\left(\widehat{\mathbf{Q}}_f^{n+j}, \widehat{\mathbf{X}}_s^{n+j} \right), \quad j = 1, \dots, N_r, \quad (25)$$

where N_r denotes the rollout horizon adopted for a given training configuration, and the prediction at step j is used as the input to step $j + 1$. This construction allows the rollout horizon to be refined without changing the definition of the underlying fluid-structure sequence.

During training, zero-mean Gaussian perturbations with a standard deviation of 0.005 are added at each rollout step to the normalized fluid variables and all four components of the structural state. This exposes the network to off-trajectory states and improves its robustness during free autoregressive rollout.

The model adopted in this study is trained using a staged rollout strategy. It is first optimized with a four-step autoregressive horizon ($N_r = 4$). The resulting R4 checkpoint is then used to initialize an eight-step autoregressive fine-tuning stage ($N_r = 8$). The selected R4-to-R8 checkpoint is used for all neural predictions and aerodynamic-force reconstructions reported in this paper. For each 60-frame stiffness trajectory, the R4 and R8 stages contain 56 and 52 admissible rollout windows, respectively, corresponding to 2464 and 2288 training windows over the 44 original and reflected training trajectories.

The dataset spans stiffness-dependent changes in plate deformation, oscillation amplitude, characteristic frequency, and wake topology. These qualitative response regimes are first identified from the plate envelopes and phase-aligned vorticity fields and are subsequently quantified using the free-tip trajectory, dominant frequency, mean drag, and root-mean-square lift. This organization provides the physical basis for evaluating whether a single conditional neural evolution operator can reproduce multiple FSI regimes and remain stable during long-term autoregressive prediction.

3. Methodology

The core of the proposed framework consists of two main pillars: a hybrid CNN-Transformer neural evolution operator that models the autoregressive spatiotemporal FSI coupling, and a differentiable aerodynamic force reconstruction module based on the DMT force formulation, facilitated by an SDF-based soft integration method. An overview of the complete framework is shown in Fig. 3.

3.1. CNN-Transformer Hybrid Evolution Operator

To resolve the multi-scale characteristics of the flow field and the discrete nature of the structure, we map the Eulerian fluid field and the Lagrangian solid points into a unified latent space using a hybrid tokenization strategy.

At each discrete time step n , the normalized fluid state is represented by

$$\tilde{\mathbf{Q}}_f^n = [\tilde{u}^n, \tilde{v}^n, \tilde{p}^n] \in \mathbb{R}^{3 \times H \times W}, \quad H = W = 256. \quad (26)$$

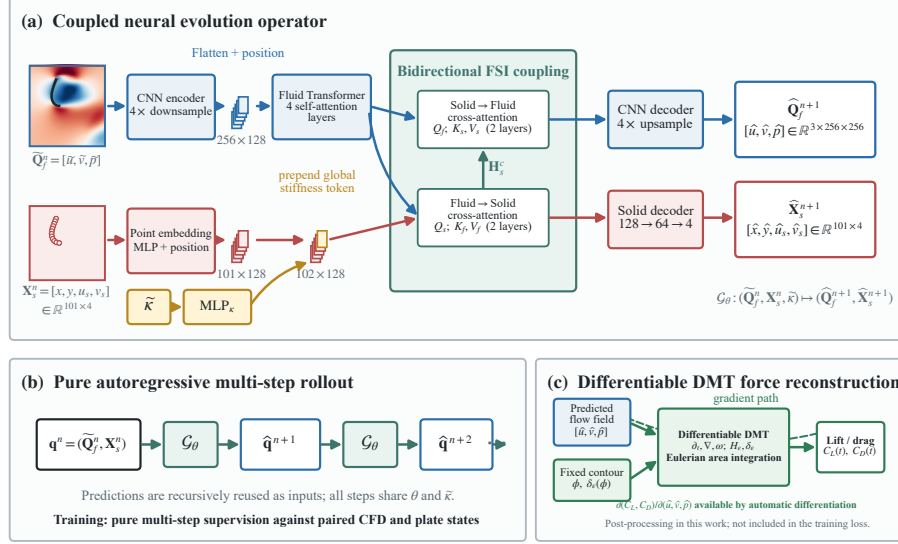


Figure 3: Overview of the proposed framework. (a) The Eulerian flow field and the Lagrangian plate states, including nodal positions and velocities, are encoded as fluid and structural tokens, respectively. The normalized bending stiffness is introduced as a global conditioning token, and bidirectional cross-attention exchanges information between the two physical representations. (b) The coupled state predicted at each time step is recursively reused as the input for subsequent autoregressive evolution. (c) Aerodynamic lift and drag are reconstructed from the predicted flow field using a differentiable DMT operator defined on a fixed level-set control contour.

The instantaneous state of the flexible plate is represented by $N_s = 101$ Lagrangian points,

$$\mathbf{X}_s^n = [\mathbf{s}_1^n, \mathbf{s}_2^n, \dots, \mathbf{s}_{N_s}^n]^T \in \mathbb{R}^{N_s \times 4}, \quad \mathbf{s}_i^n = [x_i^n, y_i^n, u_{s,i}^n, v_{s,i}^n]^T. \quad (27)$$

The dimensionless bending stiffness is supplied to the network as the normalized scalar $\tilde{\kappa}$. The one-step coupled evolution learned by the model is expressed as

$$\left(\hat{\mathbf{Q}}_f^{n+1}, \hat{\mathbf{X}}_s^{n+1} \right) = \mathcal{G}_\theta \left(\tilde{\mathbf{Q}}_f^n, \mathbf{X}_s^n, \tilde{\kappa} \right), \quad (28)$$

where $\mathcal{G}_\theta$ denotes the neural evolution operator parameterized by θ , and the hat denotes a predicted quantity. Repeated application of equation (28) produces an autoregressive rollout of the coupled fluid–structure dynamics.

3.1.1. Fluid Tokenization via CNN

At each time step n , the normalized two-dimensional fluid state $\tilde{\mathbf{Q}}_f^n \in \mathbb{R}^{3 \times H \times W}$, consisting of the velocity components $\tilde{u}$ and $\tilde{v}$ and the pressure $\tilde{p}$, is processed by a convolutional encoder. The purpose of the encoder is to extract hierarchical spatial features and convert the dense Eulerian flow field into a compact sequence of fluid tokens suitable for global attention.

The convolutional encoder consists of four successive downsampling blocks. Each block employs a 4×4 convolution with stride 2 and padding 1, followed by batch normalization and a GELU activation. The spatial resolution and channel dimension evolve as

$$\begin{aligned} 3 \times 256 \times 256 &\rightarrow 32 \times 128 \times 128 \rightarrow 64 \times 64 \times 64 \\ &\rightarrow 128 \times 32 \times 32 \rightarrow 128 \times 16 \times 16. \end{aligned} \quad (29)$$

The resulting feature map can be expressed as

$$\mathbf{F}_f^n = \mathcal{E}_{\text{CNN}}(\tilde{\mathbf{Q}}_f^n) \in \mathbb{R}^{d_m \times 16 \times 16}, \quad d_m = 128, \quad (30)$$

where $\mathcal{E}_{\text{CNN}}$ denotes the convolutional encoder and d_m is the latent embedding dimension.

The feature map is subsequently flattened along its two spatial dimensions to produce $N_f = 16 \times 16 = 256$ fluid tokens. A learnable positional embedding $\mathbf{P}_f$ is added to retain the spatial identity of each token:

$$\mathbf{H}_{f,0}^n = \text{Flatten}(\mathbf{F}_f^n) + \mathbf{P}_f \in \mathbb{R}^{N_f \times d_m}, \quad N_f = 256. \quad (31)$$

Before interacting with the structural representation, the fluid-token sequence is processed by a four-layer Transformer encoder. Each layer contains four-head self-attention, a pre-normalization structure, and a feed-forward network with hidden dimension $4d_m = 512$. The encoded fluid representation is therefore given by

$$\mathbf{H}_f^n = \mathcal{T}_f(\mathbf{H}_{f,0}^n) \in \mathbb{R}^{N_f \times d_m}, \quad (32)$$

where $\mathcal{T}_f$ denotes the fluid Transformer encoder.

This hybrid representation is particularly suitable for the present FSI problem. The convolutional encoder extracts local spatial features associated with velocity gradients and vortical structures while reducing the original 256×256 Eulerian grid to only 256 latent tokens. Applying self-attention directly to all 256^2 grid points would be computationally prohibitive because its cost increases quadratically with the sequence length. The CNN-based tokenization reduces the attention sequence length by a factor of 256, making global interaction modeling computationally tractable.

The subsequent Transformer encoder complements the local receptive fields of the convolutional layers by capturing nonlocal dependencies between the near-plate flow and the downstream wake. The resulting fluid representation therefore combines local flow features, long-range wake interactions, and computationally efficient global attention before coupling with the flexible-plate representation.

3.1.2. Solid Tokenization and Physical Embedding

The structural state is represented by $N_s = 101$ ordered Lagrangian points. Since this representation is already discrete, each point is directly embedded as a structural token without further spatial downsampling. For the i th point, the instantaneous nodal state

$$\mathbf{s}_i^n = [x_i^n, y_i^n, u_{s,i}^n, v_{s,i}^n]^T$$

is mapped into the latent space by an embedding MLP and combined with a learnable positional embedding:

$$\mathbf{h}_{s,i}^{n,0} = \text{MLP}_s(\mathbf{s}_i^n) + \mathbf{p}_{s,i}, \quad i = 1, \dots, N_s. \quad (33)$$

Here, $\mathbf{p}_{s,i}$ distinguishes the fixed ordering of the Lagrangian points, while $\mathbf{s}_i^n$ describes both the instantaneous geometry and kinematics of the plate. The resulting structural-token sequence is

$$\mathbf{H}_{s,0}^n = [\mathbf{h}_{s,1}^{n,0}, \dots, \mathbf{h}_{s,N_s}^{n,0}]^T \in \mathbb{R}^{N_s \times d_m}, \quad d_m = 128. \quad (34)$$

The normalized bending stiffness $\tilde{\kappa}$ is represented by an additional token,

$$\mathbf{h}_\kappa^0 = \text{MLP}_\kappa(\tilde{\kappa}) + \mathbf{p}_\kappa, \quad (35)$$

which is prepended to the structural-token sequence:

$$\mathbf{H}_{s\kappa,0}^n = [\mathbf{h}_\kappa^0; \mathbf{H}_{s,0}^n] \in \mathbb{R}^{(N_s+1) \times d_m} = \mathbb{R}^{102 \times 128}. \quad (36)$$

The stiffness token provides a global conditioning variable for the structural representation and participates together with the Lagrangian tokens in the subsequent cross-attention operations.

3.1.3. Bidirectional Cross-Attention Coupling Mechanism

The fundamental FSI interaction is modeled via a bidirectional cross-attention mechanism, alternating between Fluid-to-Solid ($f \rightarrow s$) and Solid-to-Fluid ($s \rightarrow f$) blocks. For the $f \rightarrow s$ coupling:

$$\hat{\mathbf{H}}_s^n = \text{softmax} \left(\frac{\mathbf{Q}_s \mathbf{K}_f^T}{\sqrt{d_k}} \right) \mathbf{V}_f + \mathbf{H}_s^n. \quad (37)$$

where $\mathbf{Q}_s$ comes from the solid tokens, while $\mathbf{K}_f$ and $\mathbf{V}_f$ come from the fluid tokens.

$$\hat{\mathbf{H}}_f^n = \text{softmax} \left(\frac{\mathbf{Q}_f \mathbf{K}_s^T}{\sqrt{d_k}} \right) \mathbf{V}_s + \mathbf{H}_f^n. \quad (38)$$

Both directional coupling modules are implemented using two-layer Transformer decoders. Each decoder layer consists of four-head attention, pre-normalization, residual connections, and a feed-forward network with hidden dimension $4d_m = 512$. For $d_m = 128$, the feature dimension of each attention head is $d_k = 32$.

The fluid-to-solid attention operates between $N_s + 1 = 102$ structural tokens, including the stiffness token, and $N_f = 256$ fluid tokens. Its attention matrix therefore has the dimension

$$\mathbf{A}_{f \rightarrow s} \in \mathbb{R}^{(N_s+1) \times N_f} = \mathbb{R}^{102 \times 256}. \quad (39)$$

The subsequent solid-to-fluid attention uses the updated structural representation as its memory, giving

$$\mathbf{A}_{s \rightarrow f} \in \mathbb{R}^{N_f \times (N_s+1)} = \mathbb{R}^{256 \times 102}. \quad (40)$$

The two operations are evaluated sequentially, such that the structural tokens are first conditioned on the fluid representation and the fluid tokens are then updated using the coupled structural representation. This formulation couples the Eulerian and Lagrangian representations in the latent space without introducing an explicit pointwise interpolation operator.

3.1.4. Autoregressive Rollout and Pure Supervision

The updated latent states are passed through the decoders to predict the next time step. The available trajectory for each stiffness contains $T_{\text{seq}} = 60$ frames. The network is trained autoregressively using pure multi-step supervision.

Following the sequence construction described in Section 2.4, a staged autoregressive training strategy is adopted. The network is first trained using a four-step rollout horizon ($N_r = 4$). The resulting R4 checkpoint is subsequently used to initialize a second training stage with an eight-step rollout horizon ($N_r = 8$). Within each training window, the predicted fluid and structural states are recursively fed back to the model without replacement by the corresponding reference states. All prediction and force-reconstruction results reported in this study are obtained using the selected R4-to-R8 checkpoint.

The training objective is defined as the sum of the fluid and structural mean-squared errors over the N_r recursively predicted steps. Thus, $N_r = 4$ in the initial training stage and $N_r = 8$ in the subsequent fine-tuning stage:

$$\mathcal{L} = \mathcal{L}_f + \mathcal{L}_s, \quad (41)$$

where

$$\begin{aligned} \mathcal{L}_f &= \frac{1}{N_r} \sum_{j=1}^{N_r} \text{MSE} \left(\widehat{\mathbf{Q}}_f^{n+j}, \tilde{\mathbf{Q}}_f^{n+j} \right), \\ \mathcal{L}_s &= \frac{1}{N_r} \sum_{j=1}^{N_r} \text{MSE} \left(\widehat{\mathbf{X}}_s^{n+j}, \mathbf{X}_s^{n+j} \right). \end{aligned} \quad (42)$$

The structural loss is evaluated jointly over the two nodal-coordinate components and the two structural-velocity components. The two loss components

are assigned equal weights in the present model.

3.2. Differentiable Aerodynamic Force Reconstruction via DMT and SDF

Standard aerodynamic-force evaluation relies on direct integration of the pressure and viscous tractions over the moving solid boundary ∂B . For neural-predicted flow fields on a relatively coarse Eulerian grid, this evaluation can be sensitive to inaccuracies in the near-wall pressure and velocity gradients. We therefore employ the DMT formulation [19] on a fixed closed control contour Σ enclosing the plate in the better-resolved outer-flow region.

In its general vorticity-moment form, the total aerodynamic force comprises the time derivative of the vorticity moment within the enclosed domain Ω_c and a boundary flux integral along Σ :

$$\mathbf{F} = -\rho \frac{d}{dt} \int_{\Omega_c} \mathbf{r} \times (\omega \mathbf{k}) d\Omega + \oint_{\Sigma} \mathbf{T}(\mathbf{u}, \omega, \mathbf{r}, \mathbf{n}) ds \quad (43)$$

where $\mathbf{r}$ is the position vector, $\mathbf{k}$ is the out-of-plane unit vector, and $\mathbf{T}$ collectively represents the pressure, viscous-vorticity, and momentum-flux contributions on the control contour.

A critical challenge in implementing Eq.43 within an end-to-end neural network is rendering the curve integral differentiable with respect to the Eulerian grid representation. Explicitly tracking the curve Σ breaks grid continuity. To achieve a fully differentiable line integral, we introduce a continuous Signed Distance Function (SDF) following the implicit boundary treatment methodology thoroughly established in deep learning fluid surrogates [23]. Here, an SDF field $\phi(\mathbf{x})$ is constructed to implicitly represent the integration contour Σ :

$$\phi(\mathbf{x}) = \begin{cases} -d(\mathbf{x}, \Sigma) & \text{if } \mathbf{x} \in \Omega_c \\ 0 & \text{if } \mathbf{x} \in \Sigma \\ d(\mathbf{x}, \Sigma) & \text{if } \mathbf{x} \in \Omega - \Omega_c \end{cases} \quad (44)$$

where $d(\mathbf{x}, \Sigma)$ denotes the Euclidean distance from a point $\mathbf{x}$ to the curve Σ . The outward unit normal vector is rigorously obtained as $\mathbf{n} = \nabla \phi / \|\nabla \phi\|$.

We replace the hard line integral over Σ with an area integral over the Eulerian domain Ω using a regularized Heaviside function H_ϵ and its derivative, the smoothed Dirac-delta function $\delta_\epsilon(\phi)$. To remain consistent with the discrete implementation, a compactly supported cosine regularization is employed:

$$H_\epsilon(\phi) = \begin{cases} 0, & \phi < -\epsilon, \\ \frac{1}{2} \left[1 + \frac{\phi}{\epsilon} + \frac{1}{\pi} \sin \left(\frac{\pi\phi}{\epsilon} \right) \right], & |\phi| \leq \epsilon, \\ 1, & \phi > \epsilon, \end{cases} \quad (45)$$

$$\delta_\epsilon(\phi) = \frac{\partial H_\epsilon}{\partial \phi} = \begin{cases} \frac{1}{2\epsilon} \left[1 + \cos \left(\frac{\pi\phi}{\epsilon} \right) \right], & |\phi| \leq \epsilon, \\ 0, & |\phi| > \epsilon. \end{cases}$$

where ϵ controls the half-width of the regularized narrow band. For all aerodynamic-force results reported in this study, $\epsilon = 0.51 \max(\Delta x, \Delta y) = 0.0051$.

Consequently, a line integral of a generic integrand I over the curve Σ is approximated by a differentiable two-dimensional area integral weighted by the regularized Dirac-delta function:

$$\oint_\Sigma I \, ds \approx \int_\Omega I \delta_\epsilon(\phi) \|\nabla \phi\| \, d\Omega. \quad (46)$$

Applying this regularized SDF-based transformation to the general vorticity-moment equation gives the following approximate area-integral representation:

$$\mathbf{F} \approx -\rho \frac{d}{dt} \int_\Omega \mathbf{r} \times (\omega \mathbf{k}) H_\epsilon(-\phi) \, d\Omega + \int_\Omega \mathbf{T} \delta_\epsilon(\phi) \|\nabla \phi\| \, d\Omega \quad (47)$$

The explicit-contour and SDF-based formulations retain the same physical terms in the DMT force identity, differing only in how the control-contour integrals are represented numerically.

This formulation seamlessly maps the curve integral onto the 2D Eulerian grid predicted by the CNN. Since spatial integration on a fixed grid and point-wise multiplication with $\delta_\epsilon(\phi)$ are inherently tensor operations in modern deep learning frameworks, the entire force reconstruction process becomes rigorously differentiable. Error gradients from force-based loss functions can propagate

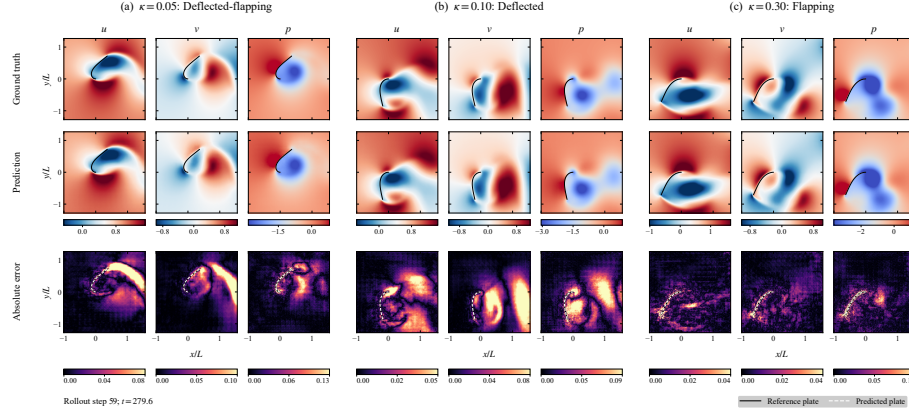


Figure 4: Flow-field reconstruction at rollout step 59 for three representative response regimes: (a) $\kappa = 0.05$; (b) $\kappa = 0.10$; and (c) $\kappa = 0.30$. Within each group, the columns show u , v , and p , and the rows show the CFD ground truth, neural prediction, and pointwise absolute error.

back through the softened Dirac-delta weighting seamlessly to both the CNN fluid tokens and the Transformer’s solid structural representation.

4. Results and Discussion

4.1. Prediction across stiffness-dependent response regimes

4.1.1. Representative regimes from one stiffness-conditioned operator

Following the physical classification introduced in Section 2, three stiffnesses are selected to represent the distinct responses contained in the data set:

$$\kappa = 0.05 \quad (\text{deflected-flapping}), \quad \kappa = 0.10 \quad (\text{deflected}), \quad \kappa = 0.30 \quad (\text{flapping}). \quad (48)$$

All three trajectories are produced by the same trained operator; only the stiffness-conditioning token and the initial coupled state are changed. The first state is supplied from the corresponding CFD trajectory, after which the fluid fields and the positions and velocities of the 101 plate nodes are advanced recursively for 59 steps.

Figure 4 shows that the predicted plate geometry and the principal high- and low-speed regions remain aligned with the CFD solution after repeated

autoregressive reuse. The distinction among the three regimes is not imposed by a discrete class label. It emerges from conditioning the common evolution map on κ and from recursively coupling the resulting structural motion back into the fluid representation. In particular, the model retains the biased mean configuration at $\kappa = 0.05$, the small-amplitude deflected response at $\kappa = 0.10$, and the large-amplitude nearly symmetric flapping at $\kappa = 0.30$.

The field and structural errors at rollout step n are measured as

$$e_q^n = \frac{\|\hat{q}^n - q^n\|_2}{\|q^n\|_2}, \quad q \in \{u, v, p\}, \quad (49)$$

and

$$e_s^n = \left[\frac{1}{2N_s} \sum_{i=1}^{N_s} \|\hat{\mathbf{x}}_{s,i}^n - \mathbf{x}_{s,i}^n\|_2^2 \right]^{1/2}. \quad (50)$$

Here, $\mathbf{x}_{s,i}^n = (x_i^n, y_i^n)^T$ and $\hat{\mathbf{x}}_{s,i}^n = (\hat{x}_i^n, \hat{y}_i^n)^T$ denote the reference and predicted positions of the i th Lagrangian point, respectively. The quantity e_q^n is the relative L_2 error over the Eulerian grid, whereas e_s^n is the coordinate RMSE over the ordered Lagrangian points.

As shown in Fig. 5, the errors remain bounded over the evaluated interval, although their magnitude depends on both the variable and the response regime. Averaged over the 59 predicted states, the relative errors are 1.89%, 5.39%, and 5.41% for u , v , and p , respectively, at $\kappa = 0.05$; 1.41%, 4.08%, and 3.36% at $\kappa = 0.10$; and 1.13%, 2.63%, and 3.94% at $\kappa = 0.30$. The corresponding mean plate-coordinate RMSE values are 2.91×10^{-3} , 9.17×10^{-4} , and 1.53×10^{-3} , respectively. The plate-coordinate errors remain of the same order across the three response regimes, indicating that the structural evolution is preserved even in the large-amplitude flapping case.

4.1.2. Accuracy across all 36 training stiffnesses

The representative cases provide a qualitative view of the three regimes, whereas the complete stiffness sweep tests whether the same operator behaves consistently throughout the parameter set. Figure 6 summarizes the mean and final-step errors for all 36 training stiffnesses.

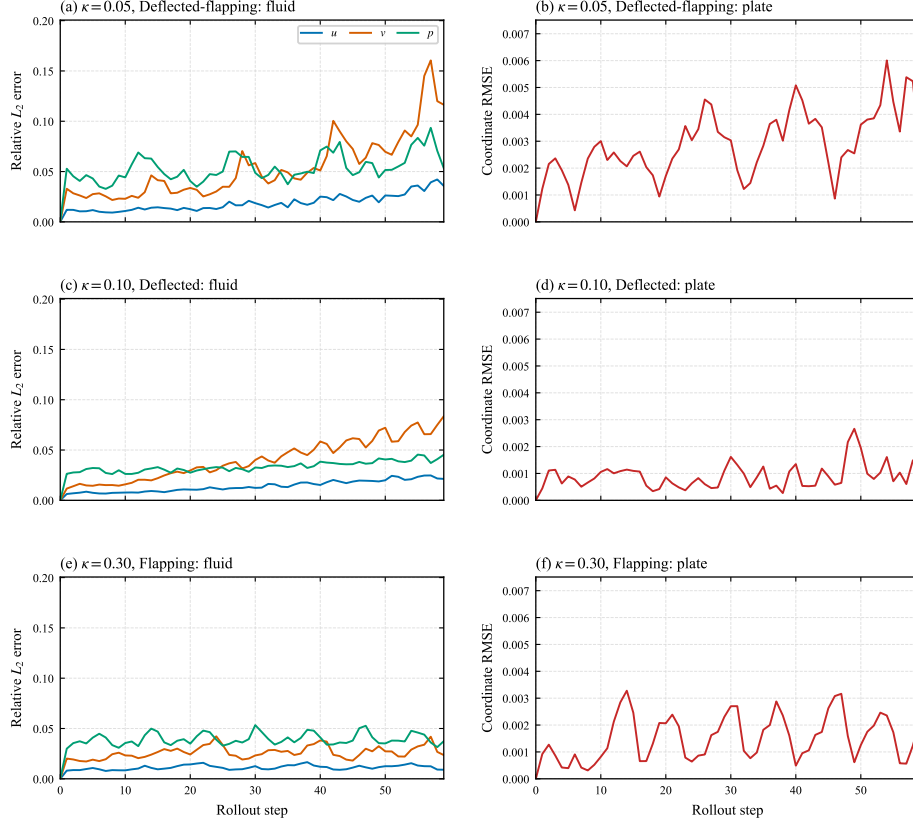


Figure 5: Stepwise rollout errors for the three representative response regimes over 59 predicted states. The left column shows the relative L_2 errors of u , v , and p , while the right column shows the coordinate RMSE of the plate. The rows correspond to $\kappa = 0.05, 0.10$, and 0.30 , respectively.

Hereafter, the plate-coordinate RMSE defined in Eq. (50) is denoted by e_{xy}^n . Since the present model also predicts the two structural velocity components, the corresponding plate-velocity RMSE is defined as

$$e_{uv}^n = \left[\frac{1}{2N_s} \sum_{i=1}^{N_s} \left\| \begin{bmatrix} \hat{u}_{s,i}^n \\ \hat{v}_{s,i}^n \end{bmatrix} - \begin{bmatrix} u_{s,i}^n \\ v_{s,i}^n \end{bmatrix} \right\|_2^2 \right]^{1/2}. \quad (51)$$

Across the 36 training conditions, the maximum mean relative errors are approximately 2.19% for u , 6.72% for v , and 5.72% for p . The maxima for u and v occur at $\kappa = 0.09$, whereas the maximum mean pressure error occurs at

Table 2: Mean errors over the 59 predicted states for eight representative stiffnesses selected from the 36 training conditions. Fluid errors are relative L_2 errors in percent; e_{xy} and e_{uv} are the nondimensional RMSEs of the plate position and velocity, respectively.

κ	Regime	e_u (%)	e_v (%)	e_p (%)	e_{xy}	e_{uv}
0.05	Deflected-flapping	1.89	5.39	5.41	2.91×10^{-3}	4.35×10^{-3}
0.10	Deflected	1.41	4.08	3.36	9.17×10^{-4}	1.68×10^{-3}
0.15	Flapping	1.04	3.26	4.21	1.19×10^{-3}	1.54×10^{-3}
0.20	Flapping	0.86	2.12	3.54	7.75×10^{-4}	1.09×10^{-3}
0.25	Flapping	0.87	2.20	3.63	6.15×10^{-4}	8.10×10^{-4}
0.30	Flapping	1.13	2.63	3.94	1.53×10^{-3}	2.23×10^{-3}
0.35	Flapping	1.51	3.58	4.28	1.51×10^{-3}	2.56×10^{-3}
0.40	Flapping	1.86	3.79	5.06	3.02×10^{-3}	4.72×10^{-3}

$\kappa = 0.13$. The maximum mean plate-position and plate-velocity RMSE values are 4.45×10^{-3} and 6.22×10^{-3} , respectively, both occurring at $\kappa = 0.31$. Because the reported errors compare time-aligned snapshots point by point, an accumulated phase difference between the predicted and reference oscillations can produce a relatively large instantaneous error without implying a qualitative failure of the predicted dynamics.

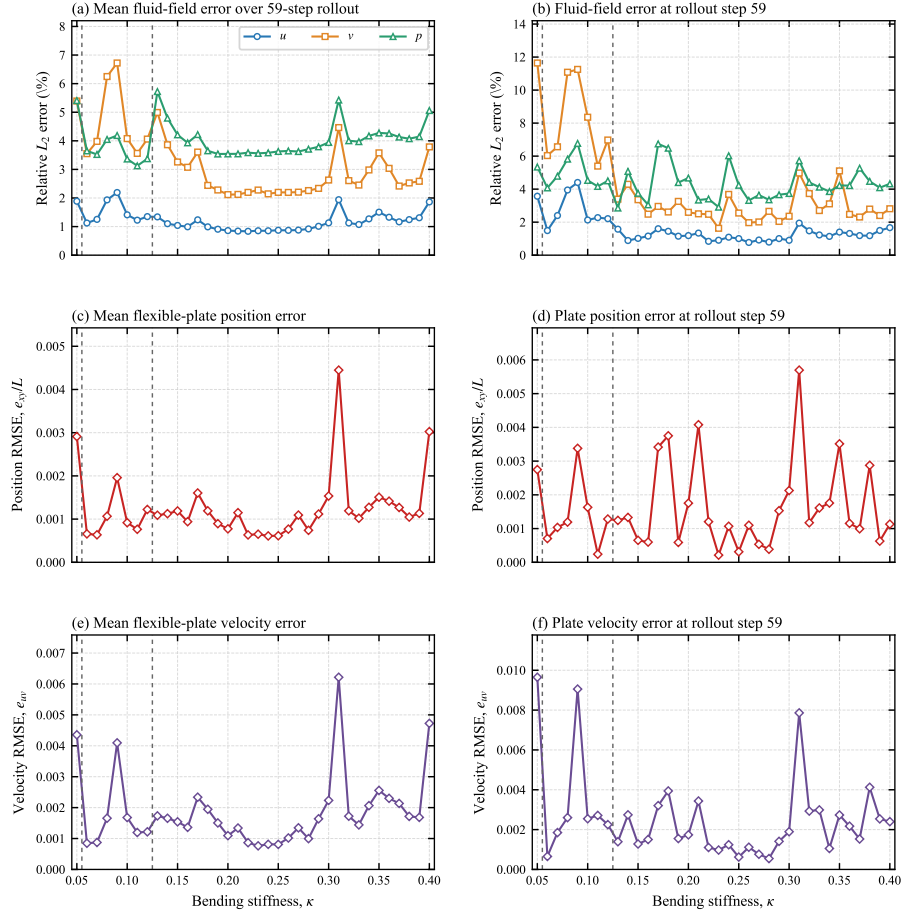


Figure 6: Prediction errors across the 36 training stiffnesses. (a) Mean fluid-field relative L_2 error over the 59 predicted states. (b) Fluid-field relative L_2 error at rollout step 59. (c) Mean flexible-plate position RMSE over the 59 predicted states. (d) Plate-position RMSE at rollout step 59. (e) Mean flexible-plate velocity RMSE over the 59 predicted states. (f) Plate-velocity RMSE at rollout step 59. The dashed separators indicate the deflected–flapping, deflected, and flapping regimes.

Despite these localized error variations, the neural operator reproduces the overall oscillatory evolution of the coupled system and preserves the principal spatial organization of the velocity and pressure fields over the evaluated rollout. More importantly, the same stiffness-conditioned operator successfully captures the deflected-flapping, deflected, and flapping response regimes, including their distinct mean plate configurations, oscillation amplitudes, and wake structures. Overall, these results support the interpretation of the network as a single stiffness-conditioned evolution operator rather than a collection of independently fitted response curves.

4.2. Oscillation frequency, phase-space structure, and long rollout

Pointwise field errors alone do not establish whether an autoregressive model preserves the underlying structural dynamics. A small phase drift can increase the instantaneous field error while leaving the limit-cycle topology and dominant frequency nearly unchanged. We therefore examine the free-tip displacement in both the time and frequency domains, as shown in Fig. 7.

The dominant frequencies are recovered exactly at the resolution of the sampled sequence: $f = 0.1333$ for $\kappa = 0.05$, $f = 0.4333$ for $\kappa = 0.10$, and $f = 0.1333$ for $\kappa = 0.30$. The predicted spectral peak amplitudes are also close to the CFD values. In particular, the model distinguishes the higher-frequency, small-amplitude deflected response from the two lower-frequency flapping responses, rather than reproducing a common oscillation for all stiffnesses.

The reconstructed limit cycles are further examined using the free-tip phase portraits shown in Fig. 8.

The range-normalized displacement and velocity RMSE values are 1.27% and 1.44% for $\kappa = 0.05$, 2.71% and 1.82% for $\kappa = 0.10$, and 0.20% and 0.22% for $\kappa = 0.30$. The larger normalized displacement error at $\kappa = 0.10$ should be interpreted together with its very small displacement range: normalization by this range magnifies modest absolute deviations. The predicted orbit nevertheless retains the location, orientation, and periodic structure of the CFD cycle.

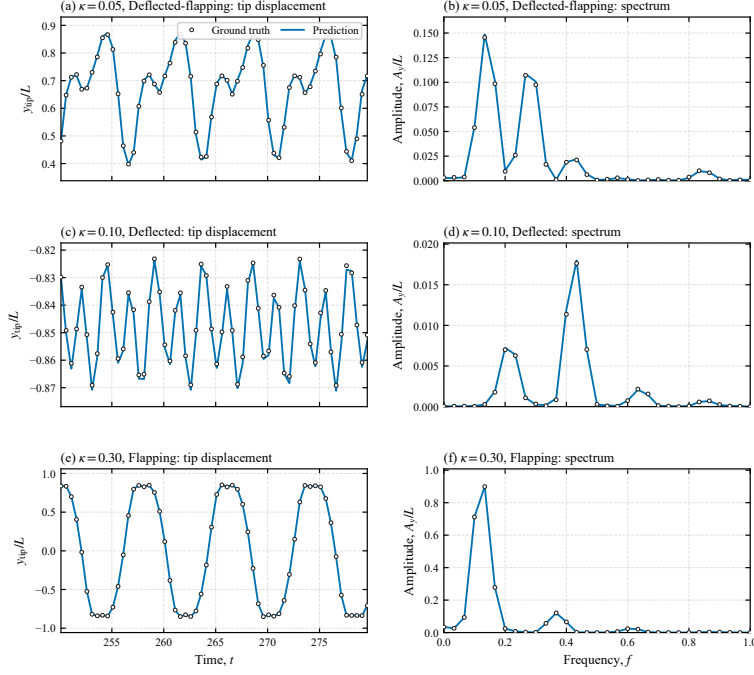


Figure 7: Free-tip displacement histories and frequency spectra for the three representative response regimes. The left column compares the predicted and CFD free-tip displacements, while the right column shows their corresponding amplitude spectra. The rows represent (a,b) $\kappa = 0.05$, (c,d) $\kappa = 0.10$, and (e,f) $\kappa = 0.30$.

Finally, a blind autoregressive rollout over 1000 steps was performed to examine long-horizon stability. The predicted coupled state remained bounded and preserved the regime-dependent periodic behavior over the full rollout. Since no synchronized CFD reference is available over this extended interval, this test is interpreted as a stability diagnostic rather than a direct measure of 1000-step prediction accuracy. Detailed stability diagnostics are provided in Appendix A.

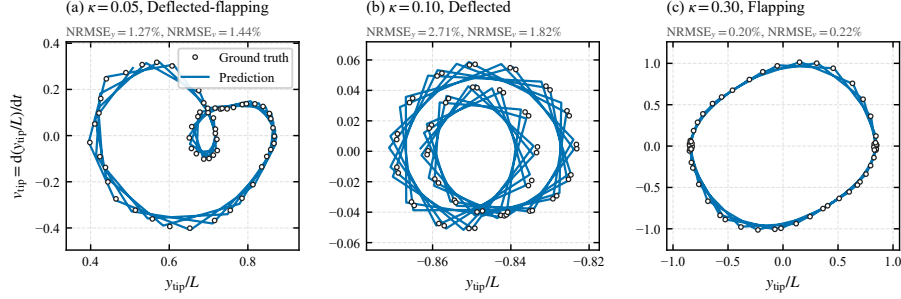


Figure 8: Free-tip phase portraits for three representative response regimes: (a) $\kappa = 0.05$, deflected-flapping; (b) $\kappa = 0.10$, deflected; and (c) $\kappa = 0.30$, flapping. The transverse tip velocity is obtained from the time derivative of the free-tip displacement. The normalized RMSE values shown above each panel quantify the displacement and velocity differences between the prediction and CFD reference.

4.3. Interpolation at two unseen stiffnesses

The 36 training stiffnesses demonstrate that a single stiffness-conditioned operator can represent the three response regimes over the sampled parameter interval. To examine interpolation between the discrete training conditions, two additional CFD trajectories are considered at $\kappa = 0.105$ and $\kappa = 0.335$. These stiffnesses were excluded from training and lie midway between the neighboring training pairs (0.10, 0.11) and (0.33, 0.34), respectively. No model parameters are updated for either test. The two cases represent different interpolation conditions: $\kappa = 0.335$ lies within the established flapping regime, whereas $\kappa = 0.105$ lies in the deflected regime, where symmetry-related plate configurations can occur.

4.3.1. Interpolation within the flapping regime at $\kappa = 0.335$.

Figure 9 compares the predicted and CFD flow fields at the final rollout step.

The model preserves the large-scale flapping pattern, the instantaneous plate configuration, and the principal spatial organization of the velocity and pressure fields. Although localized errors remain near the moving plate and the convected wake structures, the predicted coupled state remains qualitatively consistent with the CFD solution after 59 recursive steps.

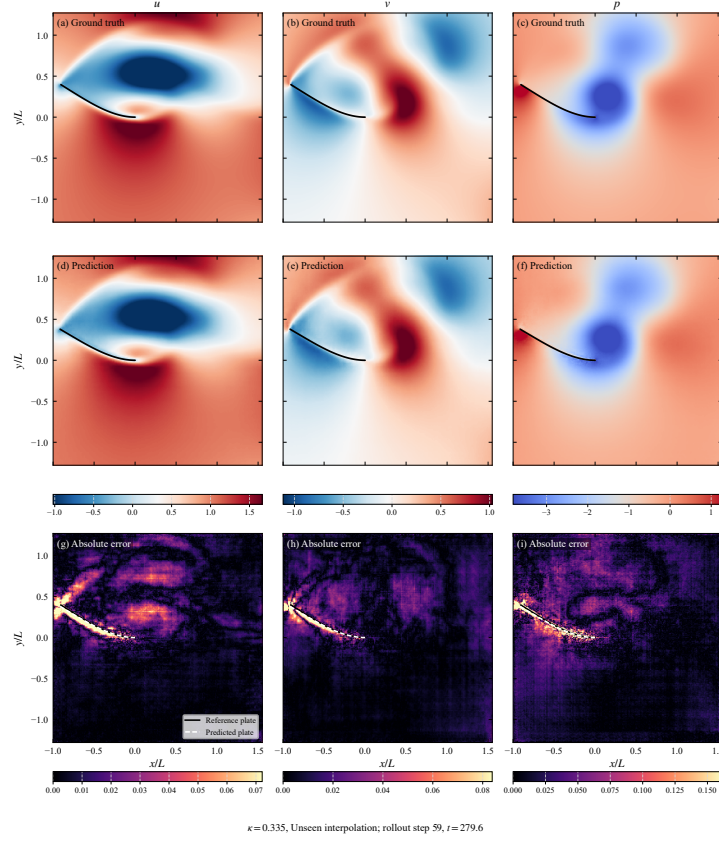


Figure 9: Flow-field reconstruction at rollout step 59 for the unseen stiffness $\kappa = 0.335$. The columns show u , v , and p , while the rows show the CFD ground truth, neural prediction, and pointwise absolute error.

The temporal evolution of the prediction errors is shown in Fig. 10.

Averaged over the 59 predicted states, the relative errors in u , v , and p are 2.20%, 5.20%, and 6.19%, respectively. The mean plate-position and plate-velocity RMSE values are 5.13×10^{-3} and 7.29×10^{-3} , respectively. At rollout step 59, the corresponding fluid-field errors are 1.97%, 3.79%, and 5.57%, while the plate-position and plate-velocity RMSE values are 7.34×10^{-3} and 8.77×10^{-3} . Both the fluid-field and structural errors remain bounded throughout the evaluated interval, indicating stable interpolation of the coupled Eulerian and Lagrangian states.

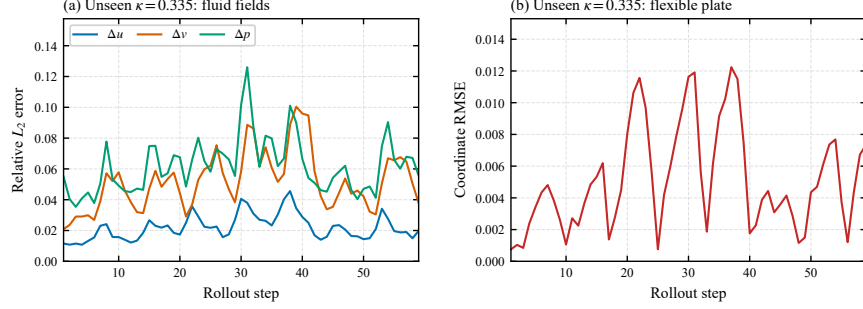


Figure 10: Evolution of the fluid-field relative L_2 errors and plate-position RMSE over the 59-step autoregressive rollout at the unseen stiffness $\kappa = 0.335$.

The corresponding low-dimensional structural dynamics are presented in Fig. 11.

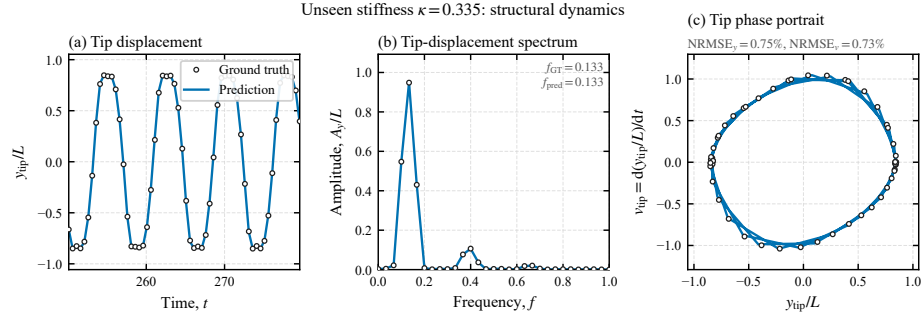


Figure 11: Structural dynamics at the unseen stiffness $\kappa = 0.335$: free-tip displacement, Fourier spectrum, and phase portrait. CFD values are shown by open circles and neural predictions by solid lines.

The predicted and CFD free-tip signals have the same dominant frequency, $f = 0.1333$. The range-normalized displacement and velocity phase-space RMSE values are 0.75% and 0.73%, respectively. The predicted orbit therefore retains the amplitude, orientation, and periodic structure of the CFD limit cycle.

4.3.2. Interpolation within the deflected regime at $\kappa = 0.105$.

Interpolation at $\kappa = 0.105$ is more sensitive because the plate response has a small oscillation amplitude and symmetry-related deflected configurations occur

in this stiffness range. Nevertheless, as shown in Fig. 12, the model recovers the upward-deflected mean configuration of the CFD trajectory and retains the principal near-plate and wake structures.

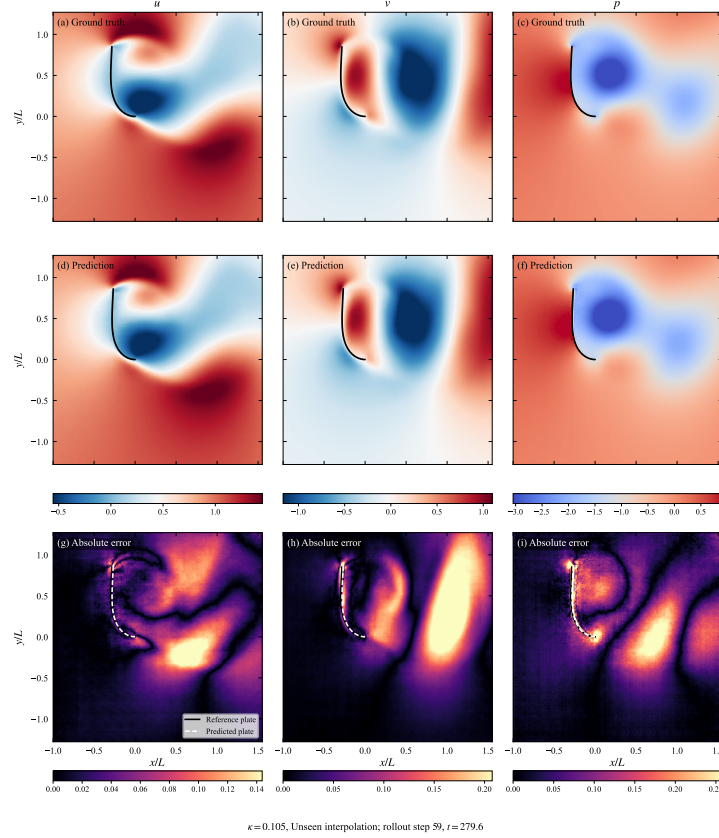


Figure 12: Flow-field reconstruction at rollout step 59 for the unseen stiffness $\kappa = 0.105$. The columns show u , v , and p , while the rows show the CFD ground truth, neural prediction, and pointwise absolute error.

The pointwise discrepancies are larger than those at $\kappa = 0.335$, particularly in the transverse velocity and pressure fields. The predicted plate and flow fields nevertheless remain on the same deflected branch as the CFD reference.

The error histories in Fig. 13 remain bounded, although their magnitudes are higher than those obtained at $\kappa = 0.335$.

Averaged over the 59 predicted states, the relative errors in u , v , and p are

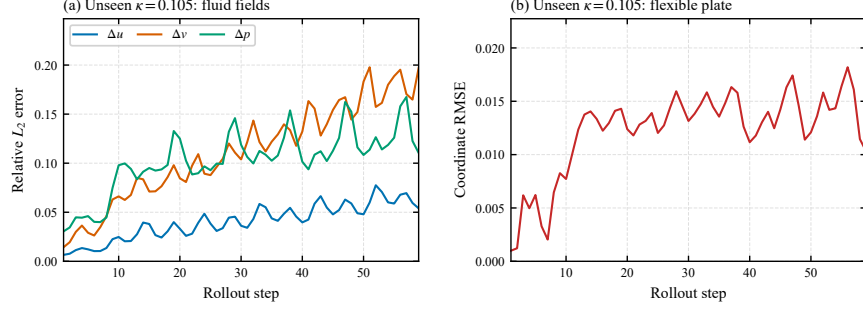


Figure 13: Evolution of the fluid-field relative L_2 errors and plate-position RMSE over the 59-step autoregressive rollout at the unseen stiffness $\kappa = 0.105$. The errors remain bounded, although their magnitudes are higher than those obtained within the flapping regime.

4.05%, 11.17%, and 10.28%, respectively. The mean plate-position and plate-velocity RMSE values are 1.22×10^{-2} and 7.98×10^{-3} , respectively. At rollout step 59, the corresponding fluid-field errors are 5.40%, 19.70%, and 11.05%, while the plate-position and plate-velocity RMSE values are 1.03×10^{-2} and 5.08×10^{-3} . The absence of monotonic error growth indicates that the increased errors do not arise from numerical divergence of the autoregressive rollout.

The structural response is examined in Fig. 14.

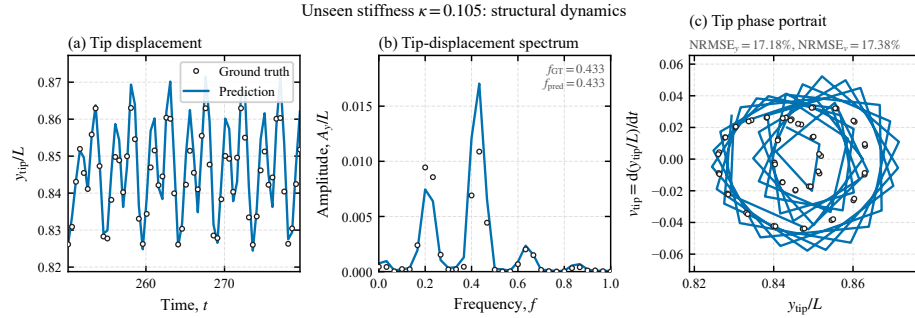


Figure 14: Structural dynamics at the unseen stiffness $\kappa = 0.105$: free-tip displacement, Fourier spectrum, and phase portrait. Although the dominant frequency and mean deflected branch are recovered, differences remain in the oscillation amplitude and phase-space trajectory.

The model recovers the CFD dominant frequency of $f = 0.4333$ and retains

a repeatable oscillatory motion. The range-normalized displacement and velocity phase-space RMSE values are 17.18% and 17.38%, respectively. Thus, the model captures the mean deflected branch and the characteristic structural frequency, although differences remain in the oscillation amplitude and phase. The interpolation at $\kappa = 0.105$ is therefore less accurate than that at $\kappa = 0.335$, but it does not exhibit an incorrect branch orientation or an unstable rollout.

Taken together, the two unseen-stiffness cases show that the same stiffness-conditioned operator can interpolate both within the flapping regime and within the branch-sensitive deflected regime. The interpolation at $\kappa = 0.335$ accurately preserves the flow-field organization and structural limit cycle, while the case at $\kappa = 0.105$ retains the correct mean branch and dominant frequency with larger instantaneous errors. These results support the use of a single conditional evolution operator across multiple stiffness-dependent FSI responses, while also identifying reduced instantaneous accuracy in the small-amplitude deflected regime.

4.4. Differentiable aerodynamic-force reconstruction from predicted flow fields

The preceding results show that the coupled neural operator can maintain a coherent moving interface while evolving the surrounding velocity and pressure fields. Reconstructing the flow and structural states, however, does not by itself provide a complete description of the fluid–structure dynamics. The aerodynamic loading is the quantity through which the fluid drives the structural motion, and the time-dependent drag and lift coefficients therefore provide a stricter physical assessment of the predicted coupled state.

Unlike pointwise flow-field errors, aerodynamic forces are global functionals of the solution. They depend simultaneously on pressure, velocity gradients, vorticity, momentum transport, and their temporal variation. Small local errors that are barely visible in an instantaneous flow-field plot may accumulate coherently during force integration. A predicted field may therefore reproduce the main wake structures while still yielding inaccurate aerodynamic coefficients.

The conventional evaluation of the aerodynamic force is based on the pressure and viscous tractions acting directly on the instantaneous body surface,

$$\mathbf{F} = \int_{\partial B} (-p\mathbf{n} + \boldsymbol{\tau} \cdot \mathbf{n}) \, ds, \quad (52)$$

where ∂B denotes the deforming plate boundary. This formulation requires accurate near-wall pressure and velocity gradients. These quantities are particularly sensitive to the spatial resolution of the neural prediction, because the thin boundary layer is only partially resolved on the coarse Eulerian grid. Direct traction integration consequently amplifies near-interface smoothing and local differentiation errors.

To reduce this dependence on the under-resolved near-wall region, the DMT is employed to express the aerodynamic force using quantities evaluated over a control domain and its enclosing contour [19]. In the present two-dimensional setting, the control surface reduces to a closed control line surrounding the plate. The force evaluation is thereby transferred from the body surface to the better-resolved outer-flow region.

4.4.1. Fixed control contours for the flexible-plate motion

The control contour must enclose the plate throughout the complete autoregressive rollout. A contour that follows the plate at every time step would introduce a time-dependent integration geometry and complicate both the numerical evaluation and subsequent differentiation. A single fixed contour is therefore constructed separately for each stiffness.

Denoting the i -th predicted Lagrangian point at rollout step n by (x_i^n, y_i^n) , the contour is obtained from the envelope of all predicted plate configurations,

$$\begin{aligned} x_0 &= \min_{n,i}(x_i^n) - m, & x_1 &= \max_{n,i}(x_i^n) + m, \\ y_0 &= \min_{n,i}(y_i^n) - m, & y_1 &= \max_{n,i}(y_i^n) + m, \end{aligned} \quad m = 0.3, \quad (53)$$

after which the four sides are aligned with the Eulerian grid. The margin separates the contour from the immediate vicinity of the plate, while the fixed

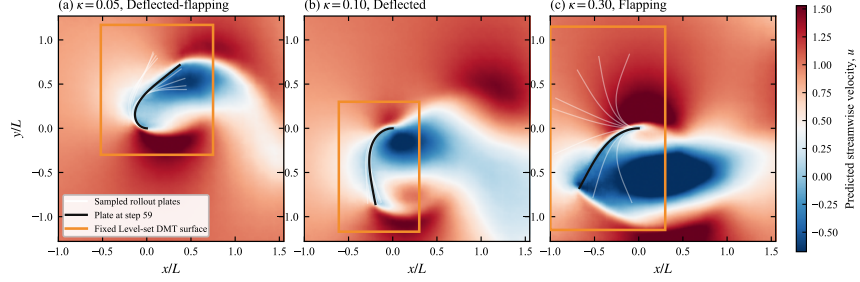


Figure 15: Fixed DMT control contours for the three representative flexible-plate regimes. The background shows the predicted streamwise velocity at rollout step 59. The white curves show sampled plate configurations during the rollout, the black curve denotes the plate at step 59, and the orange rectangle is the fixed control contour used for force reconstruction.

geometry ensures that the same integration operator is used at every rollout step.

Figure 15 illustrates the control contours for the deflected-flapping, deflected, and flapping regimes. The plate occupies substantially different regions of the domain in the three cases, so a separate contour is required for each stiffness. Within each case, however, the contour remains fixed and encloses the full structural trajectory. This construction separates the force-integration geometry from the instantaneous plate motion and provides a common basis for the explicit and Level-set DMT realizations.

4.4.2. Isolation of the DMT reconstruction error

Before applying DMT to the neural-predicted fields, the force operator is evaluated using the original CFD flow fields. This intermediate validation is necessary because a discrepancy in the final force history may originate either from the DMT reconstruction or from the neural flow prediction. Supplying resolved CFD fields removes the second source and isolates the numerical error of the control-contour formulation.

The validation is performed for the flapping case at $\kappa = 0.30$ and $Re = 200$, using 501 consecutive CFD fields at the original temporal resolution. The pressure-containing explicit Line DMT operator is evaluated on the fixed control

contour, and a cubic spline is used to differentiate the moment term in time. The resulting coefficients are compared with the aerodynamic forces reported directly by the CFD solver. The error metrics are computed using the 499 interior samples after excluding the two endpoint samples.

The range-normalized root-mean-square error is defined as

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (\hat{C}_n - C_n)^2}}{\max(C) - \min(C)} \times 100\%, \quad (54)$$

where C denotes the direct CFD coefficient and $\hat{C}$ its DMT reconstruction. This range-based normalization avoids the singular behavior of pointwise relative errors when the lift coefficient crosses zero.

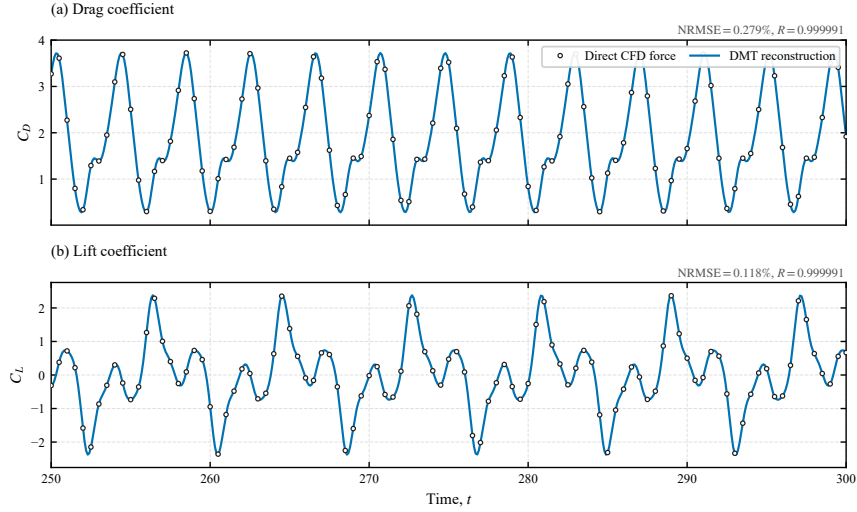


Figure 16: Validation of the pressure-containing explicit Line DMT reconstruction using the original CFD flow fields at $\kappa = 0.30$. Open circles denote the direct CFD force and solid lines denote the DMT reconstruction: (a) drag coefficient and (b) lift coefficient.

As shown in Fig. 16, the DMT reconstruction is nearly indistinguishable from the direct CFD force. The drag coefficient gives an NRMSE of 0.279% and a Pearson correlation coefficient of 0.999991, while the lift coefficient gives an NRMSE of 0.118% with the same correlation coefficient. The agreement

extends to both the principal extrema and the smaller secondary variations of the force histories.

This result establishes that the explicit Line DMT formulation can recover the aerodynamic loading without direct access to the near-wall traction. More importantly, it provides a reference level for interpreting the neural-rollout results: force errors substantially larger than the sub-percent CFD-field reconstruction error can be attributed primarily to inaccuracies in the predicted flow fields rather than to the underlying DMT formulation.

4.4.3. Aerodynamic loading recovered from neural rollouts

Having isolated the DMT reconstruction error, the same force operator is applied to the autoregressively predicted u , v , and p fields. Three stiffnesses are selected to represent the distinct structural responses contained in the data set,

$$\kappa = 0.05 \quad (\text{deflected-flapping}), \quad \kappa = 0.10 \quad (\text{deflected}), \quad \kappa = 0.30 \quad (\text{flapping}). \quad (55)$$

For each case, only the initial coupled state is supplied by CFD; all subsequent fluid and structural states are generated recursively by the neural operator. The force reconstruction therefore evaluates whether the predicted sequence retains the global pressure and momentum organization required to reproduce the aerodynamic loading.

Two realizations are considered. The explicit Line DMT samples and integrates the required quantities directly along the grid-aligned control contour. The Level-set DMT represents the same contour implicitly and replaces its line contribution with a regularized Eulerian-band integral. Both realizations receive the same neural-predicted flow sequence.

Figure 17 shows that the force histories remain qualitatively different across the three structural regimes, while their principal temporal features are consistently recovered from the neural-predicted fields. The reconstructed curves preserve the dominant peak locations, oscillation periods, mean loading levels, and phase relationships of the direct CFD forces.

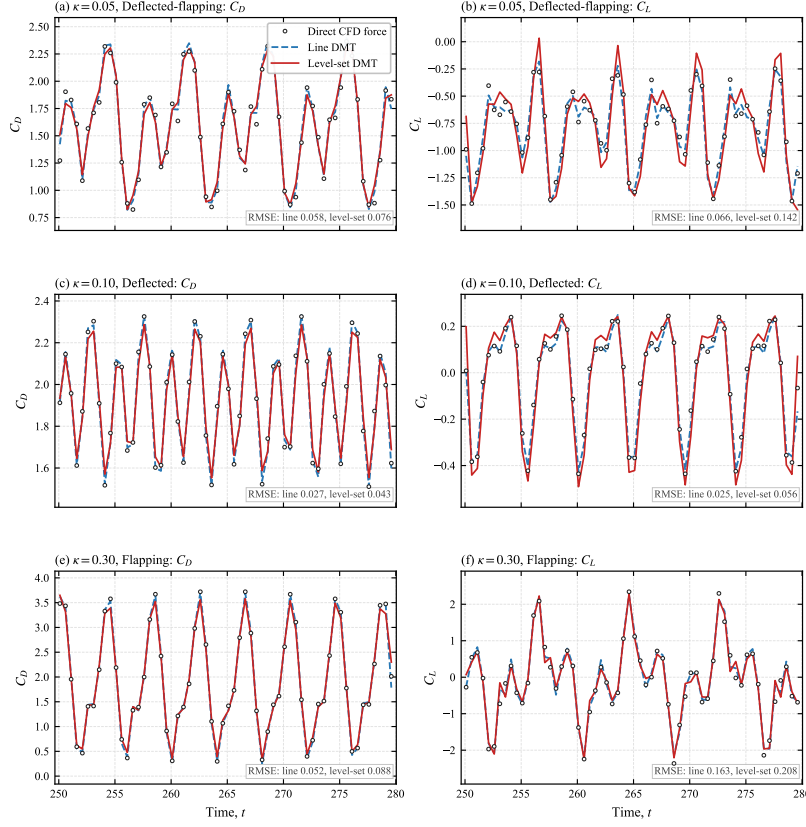


Figure 17: Aerodynamic-force reconstruction from the autoregressively predicted flow fields for three representative structural regimes. The rows correspond to $\kappa = 0.05$ (deflected-flapping), $\kappa = 0.10$ (deflected), and $\kappa = 0.30$ (flapping); the columns show C_D and C_L . Open circles denote the direct CFD force, dashed blue lines denote the explicit Line DMT reconstruction, and solid red lines denote the Level-set DMT reconstruction.

For the explicit Line DMT, the NRMSE values range from 1.52% to 5.34% across the six aerodynamic-coefficient histories. This agreement indicates that the flow fields predicted by the stiffness-conditioned operator retain sufficient outer-flow information to recover the principal aerodynamic response.

The deflected-flapping case produces a comparatively complex and asymmetric loading history. Although both DMT realizations recover its major force variations, the Level-set lift error reaches 11.49%, which is the largest error

Table 3: RMSE and range-normalized RMSE (NRMSE) of the aerodynamic coefficients reconstructed from the neural-predicted flow fields.

κ	Regime	Coefficient	Line DMT		Level-set DMT	
			RMSE	NRMSE (%)	RMSE	NRMSE (%)
0.05	Deflected-flapping	C_D	0.0576	3.74	0.0765	4.96
0.05	Deflected-flapping	C_L	0.0663	5.34	0.1424	11.49
0.10	Deflected	C_D	0.0266	3.27	0.0427	5.25
0.10	Deflected	C_L	0.0249	3.65	0.0558	8.19
0.30	Flapping	C_D	0.0521	1.52	0.0884	2.58
0.30	Flapping	C_L	0.1626	3.45	0.2075	4.41

among the representative cases. The lift signal in this regime contains several closely spaced extrema and relatively rapid variations, making it sensitive to accumulated flow-field errors and to the finite-width spatial averaging introduced by the Level-set band.

For the deflected and flapping cases, the dominant loading periods and phases remain well preserved. In particular, the two DMT realizations give similar results for the large-amplitude flapping case at $\kappa = 0.30$, despite the substantial plate excursion. This agreement indicates that the coherent periodic loading is encoded robustly in the predicted outer-flow dynamics rather than depending exclusively on the under-resolved near-wall region.

The CFD-field validation in Fig. 16 establishes a sub-percent reference error for the explicit Line DMT at $\kappa = 0.30$. The larger Line DMT errors obtained from the neural rollouts therefore primarily reflect accumulated inaccuracies in the predicted pressure, velocity, vorticity, and temporal momentum evolution. The additional difference between the Line and Level-set results also reflects the finite-width regularization used in the Level-set integration. The force comparison consequently provides a more integrated physical assessment than instantaneous field errors because it tests whether the predicted coupled trajectory remains consistent with the global aerodynamic loading.

These results demonstrate that the neural operator does not merely reproduce visually similar flow structures. Across three distinct plate-response regimes, the predicted fields preserve sufficient global information to reconstruct the dominant drag and lift dynamics. The explicit Line DMT establishes the feasibility of force recovery from the predicted coarse-grid fields, while the Level-set realization provides a differentiable approximation of the same control-contour formulation.

Aerodynamic loading at the unseen stiffnesses.

The same force-reconstruction procedure is applied to the two unseen stiffnesses examined in Section 4.3. No model parameters are updated, and both DMT realizations operate on the recursively predicted flow fields.

At $\kappa = 0.335$, the reconstructed forces closely follow the direct CFD results throughout the evaluated interval.

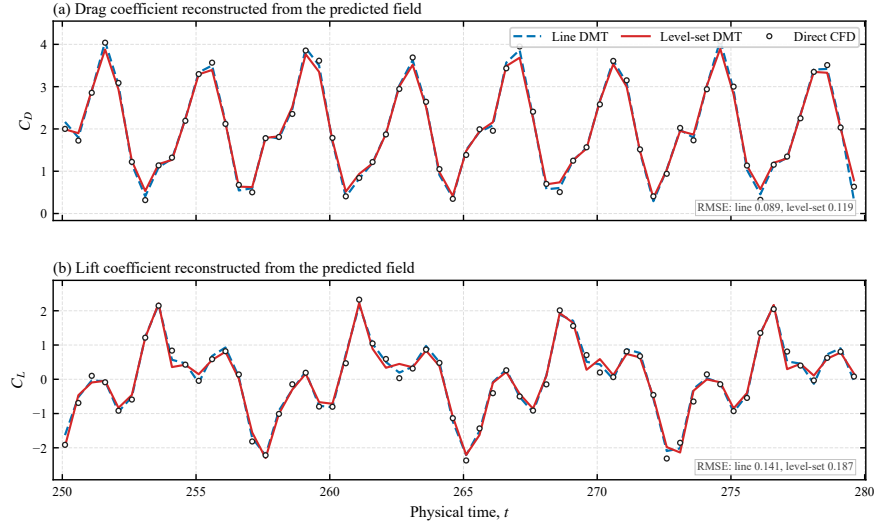


Figure 18: Drag and lift coefficients reconstructed from the predicted flow fields at the unseen stiffness $\kappa = 0.335$, compared with the forces reported directly by the CFD solver. Open circles denote the direct CFD force, dashed blue lines denote the explicit Line DMT reconstruction, and solid red lines denote the Level-set DMT reconstruction.

For the explicit Line DMT, the drag and lift NRMSE values are 2.38% and

3.00%, respectively. The corresponding Level-set DMT errors are 3.17% and 3.99%. Both realizations retain the principal extrema, oscillation period, and phase of the direct CFD forces. This agreement confirms that the interpolated flow fields preserve the global pressure and momentum organization required to recover the aerodynamic loading within the flapping regime.

The force reconstruction at $\kappa = 0.105$ is more challenging, consistent with the larger flow-field and structural errors reported in Section 4.3.

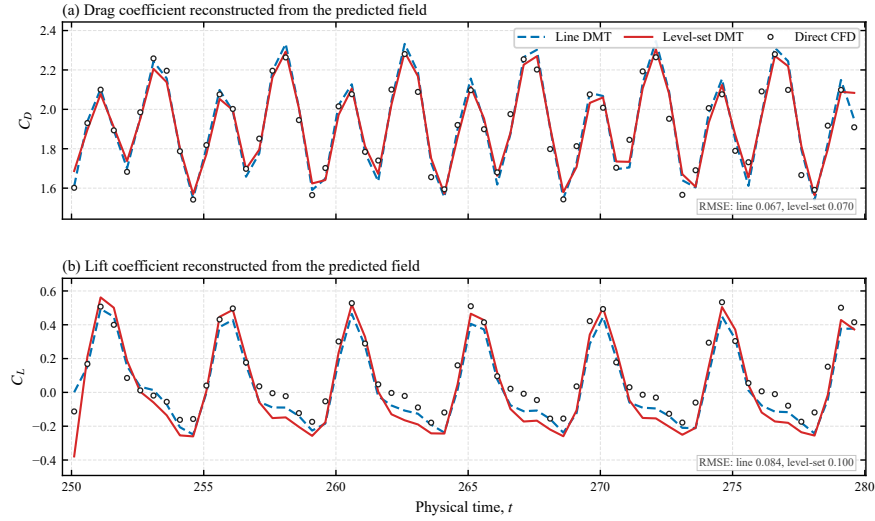


Figure 19: Drag and lift coefficients reconstructed from the predicted flow fields at the unseen stiffness $\kappa = 0.105$, compared with the forces reported directly by the CFD solver. Open circles denote the direct CFD force, dashed blue lines denote the explicit Line DMT reconstruction, and solid red lines denote the Level-set DMT reconstruction.

For the explicit Line DMT, the drag and lift NRMSE values are 9.00% and 11.73%, respectively, while the corresponding Level-set DMT values are 9.50% and 14.07%. Although these errors are larger than those at $\kappa = 0.335$, the reconstructed histories retain the principal periodic variations of the direct CFD forces. The larger discrepancies are consistent with the higher flow-field and structural errors observed in this branch-sensitive regime, while the bounded rollout errors provide no evidence of autoregressive divergence.

4.4.4. Level-set realization and differentiability

Explicit control-line integration requires the extraction and ordered traversal of discrete contour points. Although suitable for post-processing, such indexing operations are inconvenient within a tensor-based automatic-differentiation graph. The fixed contour is therefore represented implicitly as the zero level set of a signed-distance function,

$$\Sigma = \{\mathbf{x} \in \Omega : \phi(\mathbf{x}) = 0\}. \quad (56)$$

A generic contour integral is converted to an Eulerian area integral using a smoothed Dirac-delta function,

$$\oint_{\Sigma} \mathcal{I} \, ds \approx \int_{\Omega} \mathcal{I} \delta_{\varepsilon}(\phi) |\nabla \phi| \, d\Omega, \quad (57)$$

where $\delta_{\varepsilon}(\phi)$ distributes the contour contribution over a narrow band of finite width. Because the contour remains fixed, the signed-distance field, the smoothed delta function, and the normal-vector weights can be precomputed. The force reconstruction is then composed of spatial and temporal differences, pointwise tensor products, and weighted summations, all of which remain in the automatic-differentiation graph. The resulting discrete operator admits

$$\frac{\partial(C_D, C_L)}{\partial(u, v, p)} \quad (58)$$

for the complete predicted flow sequence.

The differentiable formulation makes the aerodynamic-force reconstruction available as an active functional of the predicted flow field. A conventional non-differentiable post-processing procedure can evaluate lift and drag, but it cannot transmit information on how the upstream flow prediction or a prescribed design variable should change in order to modify these quantities. In contrast, the Level-set formulation retains the dependence of the reconstructed force on every velocity and pressure value participating in the Eulerian integration.

When the force operator is composed with a differentiable neural rollout, a force-based objective $\mathcal{J}$ can, in principle, transmit sensitivity information

through the complete computational chain,

$$\frac{d\mathcal{J}}{d\alpha} = \frac{\partial \mathcal{J}}{\partial \mathbf{C}} \frac{\partial \mathbf{C}}{\partial \mathbf{q}} \frac{\partial \mathbf{q}}{\partial \alpha}, \quad \mathbf{C} = (C_D, C_L), \quad \mathbf{q} = (u, v, p), \quad (59)$$

where α denotes a differentiable model parameter, conditioning variable, control variable, or design variable. The fixed Level-set contour is advantageous in this setting because its signed-distance field and quadrature weights remain unchanged during the rollout. Consequently, no time-dependent contour extraction or discrete boundary-indexing operation interrupts the gradient path.

The regularized Level-set integration requires a balance between smoothing and grid resolution. Analytical checks of the contour integral, geometric gradient, Taylor remainder, and translating-point-vortex force are presented in Appendix B. These tests verify the accuracy of the Level-set approximation and the consistency of its automatically differentiated gradients.

In the present study, the force operator is applied only as a post-processing diagnostic, and no force-based objective is included in the training loss. The differentiable formulation nevertheless provides the computational connection required for future force-aware learning and gradient-based aeroelastic design.

Taken together, the results establish a hierarchy of predictive capability. Across the 36 training stiffnesses, a single stiffness-conditioned operator reconstructs the coupled flow fields and plate motion over three physically distinct response regimes, with bounded 59-step prediction errors and stable long-term autoregressive rollouts. At the unseen stiffness $\kappa = 0.335$, the model accurately interpolates the flow-field organization, structural limit cycle, and aerodynamic loading within the flapping regime. The more challenging case at $\kappa = 0.105$ exhibits larger instantaneous field, structural, and force errors, but the prediction recovers the correct deflected branch and the dominant structural frequency. Together with the aerodynamic-force results, these findings show that the proposed model provides a unified representation of stiffness-dependent FSI without requiring a separate surrogate for each response regime.

5. Conclusions

We have developed and systematically evaluated a stiffness-conditioned neural evolution operator for fluid–structure interaction involving an inverted flexible plate. The framework jointly evolves Eulerian flow fields and an explicit Lagrangian structural state, using 101 ordered structural tokens carrying nodal coordinates and velocities together with a global stiffness-conditioning token. Bidirectional fluid–structure coupling enables a single operator to represent the coupled evolution of the moving plate and surrounding flow.

Across 36 sampled stiffness values, the model reproduced three distinct response regimes, namely deflected–flapping, deflected, and flapping dynamics, without requiring a separate surrogate for each condition. The predicted trajectories preserved the principal flow organization, structural deformation, oscillation frequency, and phase-space structure. Blind 1000-step autoregressive rollouts remained bounded and retained the characteristic dynamical behavior of the three regimes, although these tests assess long-term dynamical stability rather than pointwise prediction accuracy over the full horizon.

The stiffness-conditioned representation also interpolated to two unseen stiffness values. The case at $\kappa = 0.335$ preserved the flow-field organization, structural limit cycle, and aerodynamic response within the flapping regime. At the branch-sensitive $\kappa = 0.105$, the model recovered the same symmetry-related deflected branch as the CFD reference and retained the dominant oscillatory behavior, although with larger instantaneous errors.

In addition, a differentiable aerodynamic-force readout was constructed by combining the derivative-moment transformation with a fixed control contour represented through a signed-distance function and a smoothed Dirac-delta formulation. The approach transfers force evaluation from the under-resolved moving boundary to the better resolved outer-flow region while retaining gradients with respect to the predicted flow variables. Validation against resolved CFD fields confirmed the accuracy of the control-contour formulation, and application to neural rollouts demonstrated that the principal aerodynamic loading can

be recovered from coarse predicted flow fields.

The differentiable force reconstruction developed here establishes a direct computational pathway from the learned fluid–structure state to aerodynamic loading and its gradients. This pathway can be extended naturally to inverse parameter identification from sparse flow, structural, or force measurements and to gradient-based optimization of stiffness and other design variables. The conditional neural operator also provides a reusable forward model for many-query parameter studies and design iterations, potentially avoiding repeated high-fidelity solutions for every candidate condition. Quantifying this computational advantage and demonstrating end-to-end inverse and design tasks are important next steps, together with broader validation under unseen parameters and incomplete observations.

Appendix A. Long-horizon autoregressive stability diagnostics

To examine the behavior of the learned evolution operator beyond the 59-step evaluation horizon considered in the main text, blind autoregressive rollouts are performed for 1000 recursive steps at the three representative stiffnesses, $\kappa = 0.05, 0.10$, and 0.30 . Only the initial coupled state is supplied by CFD, and all subsequent fluid and structural states are generated recursively without intermediate reference-state correction.

Figure A.1 presents the corresponding long-horizon stability diagnostics. The relative differences between consecutive predicted fluid states and the coordinate RMSE between consecutive predicted plate states are used to identify unbounded growth, abrupt state changes, or collapse of the autoregressive trajectory. The free-tip spectra obtained from the final 60 predicted states are compared with those of the corresponding 60-state, statistically developed CFD reference windows.

The blind rollouts remain bounded and preserve regime-dependent periodic behavior over the complete 1000-step interval. The spectra obtained from the final 60 predicted states retain the dominant frequencies of the corresponding CFD reference trajectories. Repeated reuse of the predicted coupled state does not lead to immediate divergence, collapse to a steady solution, or unphysical growth of the plate motion.

Because a synchronized CFD reference trajectory is not available over the complete extended interval, these results should be interpreted as evidence of long-horizon numerical and dynamical stability rather than as a direct measurement of 1000-step prediction accuracy.

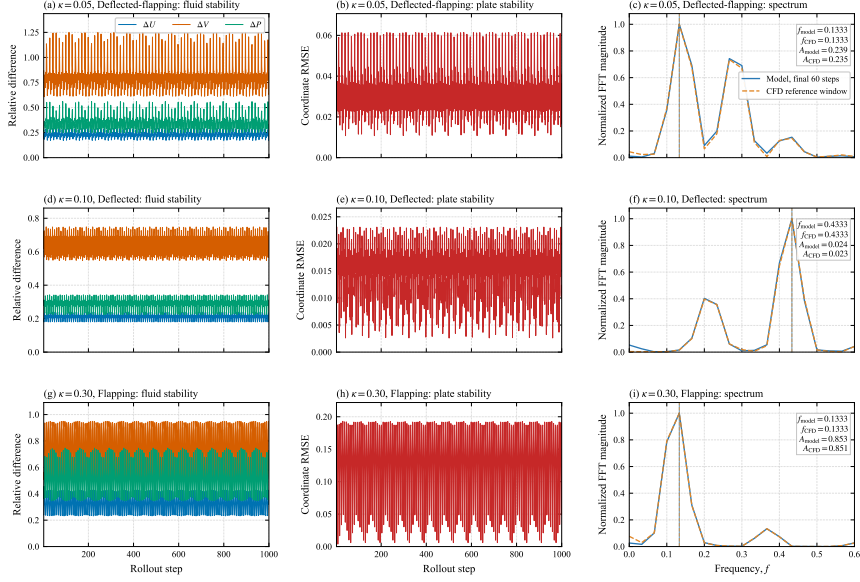


Figure A.1: Long-horizon stability diagnostics over 1000 fully autoregressive steps for $\kappa = 0.05$, 0.10, and 0.30. The left column shows the relative differences between consecutive predicted fluid states, and the middle column shows the coordinate RMSE between consecutive predicted plate states. The right column compares the normalized spectrum obtained from the final 60 predicted states with that of the corresponding 60-state CFD reference window. These quantities are used as stability diagnostics rather than errors against a synchronous 1000-step CFD trajectory.

Appendix B. Analytical verification of the differentiable force reconstruction

The Level-set quadrature and its geometric gradient are first tested on a circular contour with exact perimeter $L = 2\pi R$ and derivative $dL/dR = 2\pi$. Fig. B.1(a,b) compares the corresponding errors under grid refinement and different smoothing-band choices. Both quantities approach their analytical values when the regularized band is adequately resolved. Gradient consistency is further assessed using

$$\begin{aligned} R_0(h) &= |J(R+h) - J(R)|, \\ R_1(h) &= |J(R+h) - J(R) - hJ'(R)|. \end{aligned} \tag{B.1}$$

The fitted small-perturbation slopes are 1.01 and 2.08, consistent with the expected first- and second-order behavior, respectively (Fig. B.1(c)).

The vorticity-moment force is additionally tested using a point vortex of strength Γ translating with velocity (c_x, c_y) , for which

$$\mathbf{F} = \Gamma(-c_y, c_x), \quad \frac{\partial \mathbf{F}}{\partial \Gamma} = (-c_y, c_x). \quad (\text{B.2})$$

The computed force and its automatic-differentiation gradient agree with the analytical expressions to approximately machine precision. Differences between automatic differentiation and centered finite differences are of order 10^{-11} – 10^{-10} (Fig. B.1(d)).

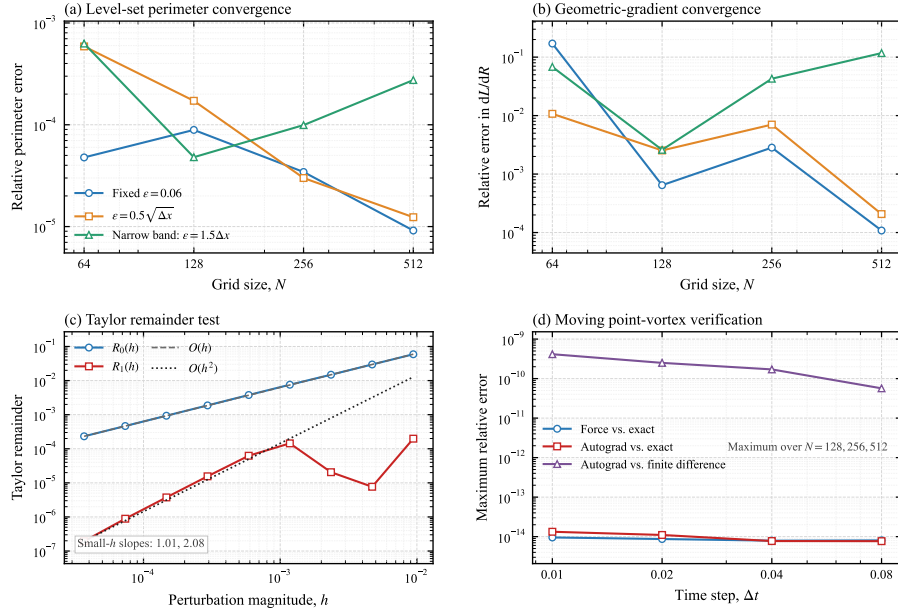


Figure B.1: Analytical verification of the Level-set integration and discrete gradients: (a) circular-perimeter convergence; (b) geometric-gradient convergence; (c) Taylor-remainder test; and (d) translating-point-vortex force and gradient verification.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant Nos. 92470120 and 12672385.

References

- [1] D. Kim, J. Cossé, C. Huertas Cerdeira, M. Gharib, Flapping dynamics of an inverted flag, *Journal of Fluid Mechanics* 736 (2013) R1. doi:10.1017/jfm.2013.555.
URL <https://doi.org/10.1017/jfm.2013.555>
- [2] J. E. Sader, J. Cossé, D. Kim, B. Fan, M. Gharib, Large-amplitude flapping of an inverted flag in a uniform steady flow—a vortex-induced vibration, *Journal of Fluid Mechanics* 793 (2016) 524–555. doi:10.1017/jfm.2016.139.
URL <https://doi.org/10.1017/jfm.2016.139>
- [3] A. Goza, T. Colonius, J. E. Sader, Global modes and nonlinear analysis of inverted-flag flapping, *Journal of Fluid Mechanics* 857 (2018) 312–344. doi:10.1017/jfm.2018.728.
URL <https://doi.org/10.1017/jfm.2018.728>
- [4] L. Lu, P. Jin, G. Pang, Z. Zhang, G. E. Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, *Nature Machine Intelligence* 3 (2021) 218–229. doi:10.1038/s42256-021-00302-5.
URL <https://doi.org/10.1038/s42256-021-00302-5>
- [5] Z. Li, N. B. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. M. Stuart, A. Anandkumar, Fourier neural operator for parametric partial differential equations, in: *International Conference on Learning Representations*, 2021.
URL <https://openreview.net/forum?id=c8P9NQVtmn0>

- [6] T. Pfaff, M. Fortunato, A. Sánchez-González, P. W. Battaglia, Learning mesh-based simulation with graph networks, in: International Conference on Learning Representations, 2021.
URL https://openreview.net/forum?id=roNqYLO_XP
- [7] Z. Li, K. Meidani, A. Barati Farimani, Transformer for partial differential equations’ operator learning, Transactions on Machine Learning Research (2023).
URL <https://openreview.net/forum?id=EPPqt3uERT>
- [8] Z. Hao, Z. Wang, H. Su, C. Ying, Y. Dong, S. Liu, Z. Cheng, J. Song, J. Zhu, GNOT: A general neural operator transformer for operator learning, in: Proceedings of the 40th International Conference on Machine Learning, Vol. 202 of Proceedings of Machine Learning Research, PMLR, 2023, pp. 12556–12569.
URL <https://proceedings.mlr.press/v202/hao23c.html>
- [9] O. Bublík, V. Heidler, A. Pecka, J. Vimmer, Neural-network-based fluid–structure interaction applied to vortex-induced vibration, Journal of Computational and Applied Mathematics 428 (2023) 115170. doi:10.1016/j.cam.2023.115170.
URL <https://doi.org/10.1016/j.cam.2023.115170>
- [10] R. Gao, R. K. Jaiman, Predicting fluid–structure interaction with graph neural networks, Physics of Fluids 36 (1) (2024) 013622. doi:10.1063/5.0182801.
URL <https://doi.org/10.1063/5.0182801>
- [11] W. Xiao, T. Gao, K. Liu, J. Duan, M. Zhao, Fourier neural operator based fluid–structure interaction for predicting the vesicle dynamics, Physica D: Nonlinear Phenomena 463 (2024) 134145. doi:10.1016/j.physd.2024.134145.
URL <https://doi.org/10.1016/j.physd.2024.134145>
- [12] X. Fan, J.-X. Wang, Differentiable hybrid neural modeling for fluid–structure interaction, Journal of Computational Physics 496 (2024) 112584.

doi:10.1016/j.jcp.2023.112584.

URL <https://doi.org/10.1016/j.jcp.2023.112584>

- [13] A. Rahman, R. J. George, M. Elleithy, D. Leibovici, Z. Li, B. Bonev, C. White, J. Berner, R. A. Yeh, J. Kossaifi, K. Azizzadenesheli, A. Anandkumar, Pretraining codomain attention neural operators for solving multiphysics PDEs, in: *Advances in Neural Information Processing Systems*, Vol. 37, Curran Associates, Inc., 2024, pp. 104035–104064. doi:10.52202/079017-3306.
URL https://proceedings.neurips.cc/paper/_files/paper/2024/file/bc75fa9843a7905bbbed9d83895a88f7f-Paper-Conference.pdf
- [14] J. Brandstetter, D. E. Worrall, M. Welling, Message passing neural PDE solvers, in: *International Conference on Learning Representations*, 2022.
URL <https://openreview.net/forum?id=vSix3HPYKSU>
- [15] P. Lippe, B. Veeling, P. Perdikaris, R. Turner, J. Brandstetter, PDE-refiner: Achieving accurate long rollouts with neural PDE solvers, in: *Advances in Neural Information Processing Systems*, Vol. 36, Curran Associates, Inc., 2023, pp. 67398–67433.
URL https://proceedings.neurips.cc/paper/_files/paper/2023/file/d529b943af3dba734f8a7d49efcb6d09-Paper-Conference.pdf
- [16] Z. Hao, C. Su, S. Liu, J. Berner, C. Ying, H. Su, A. Anandkumar, J. Song, J. Zhu, DPOT: Auto-regressive denoising operator transformer for large-scale PDE pre-training, in: *Proceedings of the 41st International Conference on Machine Learning*, Vol. 235 of *Proceedings of Machine Learning Research*, PMLR, 2024, pp. 17616–17635.
URL <https://proceedings.mlr.press/v235/hao24d.html>
- [17] B. List, L.-W. Chen, K. Bali, N. Thuerey, Differentiability in unrolled training of neural physics simulators on transient dynamics, *Computer Methods in Applied Mechanics and Engineering* 433 (2025) 117441.

doi:10.1016/j.cma.2024.117441.

URL <https://doi.org/10.1016/j.cma.2024.117441>

- [18] F. Koehler, S. Niedermayr, R. Westermann, N. Thuerey, APEBench: A benchmark for autoregressive neural emulators of PDEs, in: Advances in Neural Information Processing Systems, Vol. 37, Curran Associates, Inc., 2024, pp. 120252–120310. doi:10.52202/079017-3822.
URL https://proceedings.neurips.cc/paper/_files/paper/2024/file/d9875ebcf74bccdc5076acab0db62c-Paper-Datasets_and_Benchmarks_Track.pdf
- [19] J.-Z. Wu, Z.-L. Pan, X.-Y. Lu, Unsteady fluid-dynamic force solely in terms of control-surface integral, *Physics of Fluids* 17 (9) (2005) 098102. doi:10.1063/1.2055528.
URL <https://doi.org/10.1063/1.2055528>
- [20] J.-Z. Wu, X.-Y. Lu, L.-X. Zhuang, Integral force acting on a body due to local flow structures, *Journal of Fluid Mechanics* 576 (2007) 265–286. doi:10.1017/S0022112006004551.
URL <https://doi.org/10.1017/S0022112006004551>
- [21] D. Lentink, Accurate fluid force measurement based on control surface integration, *Experiments in Fluids* 59 (1) (2018) 22. doi:10.1007/s00348-017-2464-1.
URL <https://doi.org/10.1007/s00348-017-2464-1>
- [22] A.-K. Gao, C. Xie, X.-Y. Lu, Weighted integral methods for fluid force diagnostics in incompressible flows, *Journal of Fluid Mechanics* 1024 (2025) A57. doi:10.1017/jfm.2025.10854.
URL <https://doi.org/10.1017/jfm.2025.10854>
- [23] L.-W. Chen, B. A. Cakal, X. Hu, N. Thuerey, Numerical investigation of minimum drag profiles in laminar flow using deep learning surrogates, *Journal of Fluid Mechanics* 919 (2021) A34. doi:10.1017/jfm.2021.398.

- [24] X. Fan, X.-Y. Liu, M. Wang, J.-X. Wang, Diff-FlowFSI: A GPU-optimized differentiable CFD platform for high-fidelity turbulence and FSI simulations, *Computer Methods in Applied Mechanics and Engineering* 448 (2026) 118455. doi:10.1016/j.cma.2025.118455.
URL <https://doi.org/10.1016/j.cma.2025.118455>