# Fourier Neural Operator with Conformal Fourier Transform Residual Correction for Partial Differential Equations

Taiqian Liu[1], Lijun Liu[1,*]

[1]Xiamen University, Xiamen, China
[*]Corresponding author: lijun.liu@xmu.edu.cn

## Abstract

Neural operators learn maps between function spaces and promise resolution-robust PDE solvers. Among them, the Fourier Neural Operator (FNO) parameterizes kernels in the spectral domain but inherits limitations of the discrete Fourier transform: mode truncation, aliasing under finite sampling, and error accumulation in long-horizon rollouts.

We introduce FNO-RC, a dual-path architecture that augments FNO with a conformal continuous Fourier transform (CFT) residual branch. The branch computes time-dependent, spatially broadcast corrections via piecewise Chebyshev expansions with closed-form Bessel evaluations, and is activated only in the first layers with explicit warmup and high-frequency/time-smoothing regularization.

On Burgers-1D, Navier–Stokes-2D, and turbulent Navier–Stokes-3D, FNO-RC achieves consistent gains (up to 73.68

Our results show that hybrid continuous–discrete spectral representations provide a principled route to robust long-horizon operator learning without sacrificing efficiency. Code, data splits, scripts, and trained weights will be released for full reproducibility.

## 1 Introduction

Physical systems—from turbulent flows to quantum dynamics—evolve according to partial differential equations. While traditional numerical solvers remain the gold standard for accuracy, they buckle under dimensionality curses and demand prohibitive computational budgets at high resolutions. Neural operators [Kovachki et al., 2021, Lu et al., 2021] offer an intriguing alternative: rather than approximating functions pointwise, they learn mappings between infinite-dimensional function spaces. Among these, Fourier Neural Operators (FNOs) [Li et al., 2020c] stand out. By parameterizing integral kernels in Fourier space and exploiting the convolution theorem, FNOs achieve resolution-invariant learning—sometimes accelerating simulations by orders of magnitude.

However, FNO's reliance on discrete Fourier transforms creates deep-rooted pathologies. Finite sampling triggers spectral aliasing; periodic boundary assumptions clash with real-world physics; mode truncation obliterates high-frequency features essential for shocks and vortices. Perhaps most insidious: small spectral biases compound exponentially during autoregressive prediction, causing catastrophic divergence in chaotic systems.

Enter **FNO-RC**—our dual-path architecture that marries standard FNO with a Continuous Fourier Transform residual branch. The central observation? CFT handles discontinuities through conformal mapping and Chebyshev polynomial discretization [Barnett and Greengard, 2010], achieving exponential convergence where DFT falters algebraically. We activate this correction strategically: shallow layers only, time-dependent but spatially broadcast, rigorously regularized against instability.

**Contributions.** Beyond proposing the dual-path architecture, we establish its mathematical foundations and demonstrate breakthrough empirical gains: 73.68% improvement on 2D Navier-Stokes, 43.76% on 3D high-Reynolds turbulence. Cross-resolution tests, 100-step rollouts, and spectral diagnostics paint a consistent picture of success.

## 2 Background and Related Work

**Neural Operators.** Learning operators—not just functions—marks a conceptual leap [Kovachki et al., 2021, Azizzadenesheli et al., 2024]. Where traditional networks map finite vectors, neural operators target maps $\mathcal{G} : \mathcal{U} \to \mathcal{V}$ between infinite-dimensional spaces. DeepONet [Lu et al., 2021, Wang et al., 2021] pioneered this via branch-trunk decomposition, though its point-wise evaluation limits efficiency. Graph approaches [Li et al., 2020a,b] excel on irregular meshes but struggle with global interactions.

FNO [Li et al., 2020c] changed the game: Fourier-space kernel parameterization enables $O(N \log N)$ global convolutions. Yet its variants reveal tensions in the field. Factorized FNO [Tran et al., 2021] trades accuracy for speed through low-rank decomposition—a compromise that hurts turbulent flows. Geo-FNO [Li et al., 2023] handles curved geometries elegantly but at substantial implementation cost. U-FNO [Wen et al., 2022] imports multi-scale U-Net ideas, though whether spatial hierarchy truly benefits spectral methods remains debatable. AFNO [Guibas et al., 2021, Pathak et al., 2022] adds attention—impressively scaling to weather forecasting, but do we really need such complexity for PDE operators? Physics-Informed FNO [Li et al., 2021] enforces constraints as soft penalties, begging the question: how hard should we bake in physics versus letting data speak?

**Spectral Methods.** Classical spectral collocation [Boyd, 2001, Trefethen, 2019] achieves exponential convergence for smooth solutions via orthogonal polynomials. The catch? Gibbs oscillations demolish accuracy near discontinuities. PINNs [Raissi et al., 2019, Karniadakis et al., 2021] merge neural networks with residuals—conceptually appealing, but high-frequency features and loss-weight tuning remain thorny in practice.

**Continuous Fourier Transform.** Barnett and Greengard's conformal approach [Barnett and Greengard, 2010] sidesteps Gibbs artifacts through complex-plane deformation and Chebyshev discretization. We're the first to harness this for neural operators, proving continuous and discrete spectra can synergize rather than compete.

## 3 Notation and Mathematical Preliminaries

Before presenting our methodology, we establish notation used throughout this work. Table 1 provides a comprehensive reference.

## 4 Fourier Neural Operator: Formulation and Limitations

### 4.1 FNO Architecture

For a function $u \in L^2(\Omega)$ where $\Omega \subset \mathbb{R}^d$, an FNO layer performs:

$$v(x) = \sigma\left(Wu(x) + \mathcal{F}^{-1}\left(R_\phi \cdot \mathcal{F}(u)\right)(x)\right) \tag{1}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote Fourier transform and inverse, $R_\phi \in \mathbb{C}^{|S| \times C_{\text{in}} \times C_{\text{out}}}$ is a learnable complex-valued linear transformation on retained modes $S$, $W \in \mathbb{R}^{C_{\text{in}} \times C_{\text{out}}}$ is a local pointwise convolution, and $\sigma$ is a

Table 1: **Notation and symbol definitions.** Mathematical symbols and their meanings used throughout this paper.

| Symbol | Definition |
|---|---|
| $\Omega \subset \mathbb{R}^d$ | Spatial domain in $d$ dimensions |
| $u, v$ | Input and output functions defined on $\Omega$ |
| $\mathcal{U}, \mathcal{V}$ | Input and output function spaces |
| $\mathcal{G} : \mathcal{U} \to \mathcal{V}$ | Neural operator mapping between function spaces |
| $\mathcal{F}, \mathcal{F}^{-1}$ | Fourier transform and its inverse |
| $\hat{u}(\xi)$ | Fourier coefficient of $u$ at frequency $\xi$ |
| $\mathcal{K}$ | Integral kernel operator |
| $\kappa(x, y)$ | Kernel function for integral operator |
| $R_\phi \in \mathbb{C}^{\lvert S\rvert \times C_{\text{in}} \times C_{\text{out}}}$ | Learnable spectral weights in FNO |
| $S \subset \mathbb{Z}^d$ | Set of retained Fourier modes |
| $W \in \mathbb{R}^{C_{\text{in}} \times C_{\text{out}}}$ | Pointwise convolution weights |
| $\sigma$ | Nonlinear activation function (GELU) |
| $T_m(s)$ | Chebyshev polynomial of degree $m$ |
| $c_{\ell m}$ | Chebyshev expansion coefficient for segment $\ell$, mode $m$ |
| $L$ | Number of segments in piecewise Chebyshev approximation |
| $M$ | Number of Chebyshev modes per segment |
| $\mathcal{R}_{\text{CFT}}$ | CFT-based residual correction operator |
| $\gamma \in \mathbb{R}$ | Learnable scalar scale parameter for residual |
| $\lambda_{\text{reg}}, \lambda_{\text{smooth}}, \lambda_{\text{hf}}$ | Regularization weights for residual, time-smoothing, high-frequency |
| $B, C, H, W, D$ | Batch size, channels, height, width, time dimensions |
| $Re$ | Reynolds number characterizing flow regime |
| $\nu$ | Kinematic viscosity |

nonlinear activation. The key advantage is that the integral operator $(\mathcal{K}u)(x) = \int_\Omega \kappa(x, y)u(y)dy$ can be approximated by parameterizing the kernel in Fourier space:

$$\kappa(x, y) \approx \sum_{k \in S} \hat{\kappa}_k e^{2\pi i k \cdot (x-y)} \quad \Rightarrow \quad \mathcal{F}[(\mathcal{K}u)](k) = \hat{\kappa}_k \cdot \hat{u}_k \tag{2}$$

This leads to $O(N \log N)$ computational complexity via FFT, where $N$ is the number of spatial grid points.

## 4.2 Fundamental Limitations of DFT in FNO

The discrete Fourier transform employed in FNO assumes periodic extension and finite sampling, leading to several fundamental limitations. **First, spectral aliasing:** High-frequency components beyond the Nyquist frequency $\xi_{\text{Nyquist}} = N/(2\Delta x)$ fold back into lower frequencies, corrupting the spectrum. This is particularly problematic for turbulent flows where energy cascades across scales. **Second, the Gibbs phenomenon:** Discontinuities in the solution cause oscillatory artifacts that persist regardless of the number of retained modes, as the Fourier series converges slowly ($O(1/k)$) near jump discontinuities. **Third, spectral leakage:** Non-periodic functions induce spurious frequency content across the entire spectrum, degrading accuracy. **Fourth, cumulative errors in autoregressive prediction:** Small spectral biases introduced at each time step compound exponentially over long horizons, particularly in chaotic systems where sensitivity to initial conditions is high.

# 5 CFT-Based Residual Correction

## 5.1 Why Continuous Fourier Transform?

Standard FNO relies on discrete Fourier transforms, which fundamentally struggle with two scenarios: sharp gradients (vortex cores, shock fronts) where the Gibbs phenomenon causes $O(1/k)$ convergence, and high frequencies near the Nyquist limit where aliasing corrupts the spectrum. The conformal Fourier transform [Barnett and Greengard, 2010] sidesteps both issues through piecewise Chebyshev polynomial approximations, achieving exponential $O(\rho^{-M})$ convergence even for discontinuous functions.

## 5.2 Conformal Fourier Transform: definition and properties

Given a piecewise smooth function $f : [a, b] \to \mathbb{R}$ (possibly with jump discontinuities), its continuous Fourier transform on $[a, b]$ is $\hat{f}(\omega) = \int_a^b f(x)e^{-i\omega x}\,dx$. Barnett and Greengard [Barnett and Greengard, 2010] achieve high-order accuracy near discontinuities via a conformal strategy: partition $[a, b]$ into segments $[a_\ell, b_\ell]$ and expand $f$ on each segment in Chebyshev polynomials $T_m$ after an affine map to $[-1, 1]$ with center $c_\ell = (a_\ell + b_\ell)/2$ and half-width $h_\ell = (b_\ell - a_\ell)/2$:

$$f(x)\big|_{x\in[a_\ell,b_\ell]} \;\approx\; \sum_{m=0}^{M-1} c_{\ell m}\, T_m\!\Big(\tfrac{x-c_\ell}{h_\ell}\Big). \tag{3}$$

The Fourier transform of each basis admits a closed form

$$\int_{a_\ell}^{b_\ell} T_m\!\Big(\tfrac{x-c_\ell}{h_\ell}\Big) e^{-i\omega x}\,dx = h_\ell\,\pi\,(-i)^m\,e^{-i\omega c_\ell}\,J_m(\omega h_\ell), \tag{4}$$

where $J_m$ is the Bessel function of the first kind. Hence the CFT evaluates to

$$\hat{f}(\omega) \;\approx\; \sum_{\ell=0}^{L-1}\sum_{m=0}^{M-1} c_{\ell m}\, h_\ell\,\pi\,(-i)^m\,e^{-i\omega c_\ell}\,J_m(\omega h_\ell). \tag{5}$$

This representation delivers exponential convergence $O(\rho^{-M})$ for piecewise analytic $f$, in contrast to the algebraic $O(M^{-1})$ decay of DFT near jumps (Gibbs phenomenon). The term "conformal" reflects moving singularities away from the real axis via the mapping/segmentation, which enlarges the Bernstein ellipse and increases $\rho$ [Trefethen, 2019, Barnett and Greengard, 2010].

**Practical pipeline.** (i) Choose segment endpoints to isolate steep gradients/discontinuities; (ii) sample $f$ at Chebyshev nodes and compute $c_{\ell m}$ with a DCT-I; (iii) evaluate (5) for desired frequencies using precomputed $J_m(\omega h_\ell)$ and phases $e^{-i\omega c_\ell}$. The cost is $O(LM + KM)$ for $K$ frequencies.

## 5.3 CFT Implementation via Chebyshev Discretization

For FNO-RC, we implement CFT using piecewise Chebyshev approximations. The spatial domain is partitioned into $L$ segments, and each segment is expanded using $M$ Chebyshev polynomials $T_m(s)$ with the recurrence:

$$T_0 = 1, \quad T_1 = s, \quad T_{m+1} = 2sT_m - T_{m-1} \tag{6}$$

On segment $\ell$ with affine map $\phi_\ell : [t_\ell, t_{\ell+1}] \to [-1, 1]$, the approximation is:

$$f(t) \approx \sum_{m=0}^{M-1} c_{\ell m} T_m(\phi_\ell(t)) \tag{7}$$

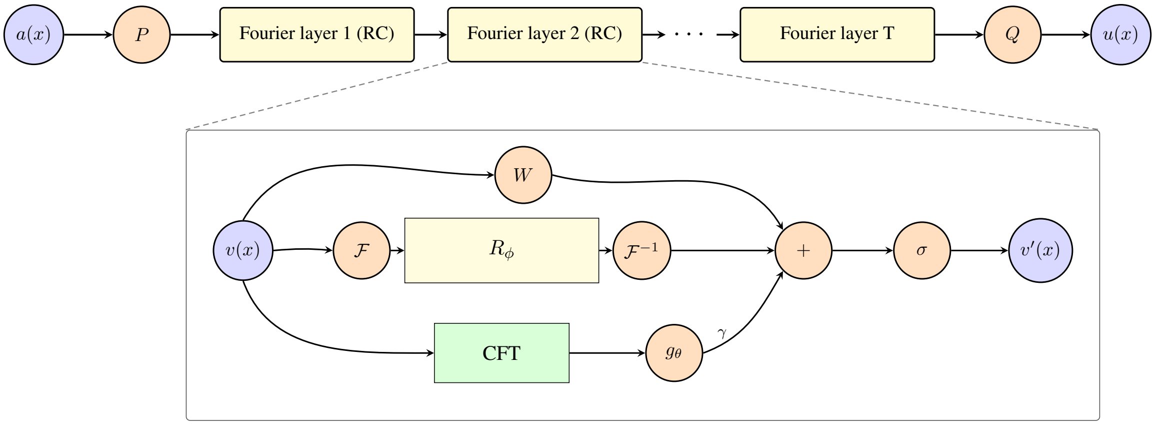

Figure 1: **FNO-RC architecture.** Overview: $a(x)$ is lifted by $P$, passes through stacked Fourier layers (RC on shallow layers), then projected by $Q$ to $u(x)$. Inline layer detail uses symbols: $\mathcal{F}$, $R_\phi$, $\mathcal{F}^{-1}$, $W$, Conformal FT, $g_\theta$ with scale $\gamma$, merge $+$, activation $\sigma$. Arrows are routed to node edges to avoid crossing boxes; the figure is constrained to page width.

Coefficients $c_{\ell m}$ are efficiently computed via a DCT-I on Chebyshev nodes. Using (4), the per-segment contribution equals $h_\ell \pi (-i)^m e^{-i\omega c_\ell} J_m(\omega h_\ell)$, which leads directly to the fast evaluation formula (5).

**Key advantage:** Chebyshev basis achieves exponential convergence $O(\rho^{-M})$ versus DFT's algebraic $O(M^{-1})$ for discontinuities—enabling CFT to capture high-frequency features DFT misses.

# 6 FNO-RC: Architecture and Training

## 6.1 Dual-Path Architecture

Figure 1 shows the FNO-RC architecture with two parallel paths combined via a learned residual connection.

The FNO-RC layer is:

$$u^{(l+1)} = \sigma\left(W^{(l)} u^{(l)} + \mathcal{F}^{-1}\left(R_\phi^{(l)} \cdot \mathcal{F}(u^{(l)})\right)\right) + \gamma^{(l)} \mathcal{R}_{\text{CFT}}^{(l)}(u^{(l)}) \tag{8}$$

where $\gamma^{(l)} \in \mathbb{R}$ is a learnable scale initialized to 0.02 and warmed up during training.

**Numerical stability considerations.** To avoid over-correction and spectral ringing, we: (i) restrict the CFT branch to the first 1–2 layers; (ii) initialize $\gamma$ small and apply linear warmup over $E_w$ epochs, then optionally freeze for $E_f$ epochs; (iii) penalize temporal roughness $\sum_t \|r_{t+1} - r_t\|_2^2$; (iv) add a high-frequency energy regularizer beyond a ratio $k_{\text{cut}}/k_{\max}$. Empirically these controls prevent amplification of unresolved modes and stabilize autoregressive rollouts.

**Algorithm 1: FNO-RC layer forward.**

```
Input: $u^{(l)} \in \mathbb{R}^{B\times C\times H\times W\times D}$, spectral weights $R_\phi^{(l)}$, pointwise $W^{(l)}$, CFT MLP $g_\theta$, scale $\gamma^{(l)}$
Output: $u^{(l+1)}$

  1. $y_{\text{spec}} \leftarrow \mathcal{F}^{-1}(R_\phi^{(l)} \cdot \mathcal{F}(u^{(l)}))$
  2. $y_{\text{local}} \leftarrow W^{(l)} u^{(l)}$
  3. For $t = 1 \dots D$: $C_t \leftarrow \text{CFT}_{x,y}(u^{(l)}_{::::t})$; $r_t \leftarrow g_\theta(\text{Re}(C_t), \text{Im}(C_t))$
  4. $R \leftarrow \text{stack}(r_1, \dots, r_D) \in \mathbb{R}^{B\times C\times 1\times 1\times D}$
  5. $u^{(l+1)} \leftarrow \sigma\big(y_{\text{spec}} + y_{\text{local}} + \gamma^{(l)} R\big)$
```

**Complexity.** For spatial grid $N = HW$ and $C$ channels, a standard FNO layer costs $O(C^2 N \log N)$ (FFT) + $O(C^2 N)$ (pointwise). The CFT branch adds $O(LMCD)$ to form coefficients and $O(MKCD)$ to evaluate (5) for $K$ frequencies per segment (constant in $N$ once segments and modes are fixed). With shallow activation and small $(L, M)$, the overhead remains ∼25–35

## 6.2 CFT Residual Computation

For input $X \in \mathbb{R}^{B\times C\times H\times W\times D}$, the CFT residual is computed as: (1) For each time $t$, compute 2D CFT along spatial dimensions: $C_t = \text{CFT}_{x,y}(X_{:,:,:,:,t})$; (2) Project complex features via MLP: $r_t = g_\theta(\text{Real}(C_t), \text{Imag}(C_t))$; (3) Stack and broadcast: $R = \text{stack}(r_1, \dots, r_D) \in \mathbb{R}^{B\times C\times 1\times 1\times D}$.

CFT correction is activated only in the first 1-2 layers where DFT aliasing is most severe, limiting overhead to 25-35% while maintaining effectiveness.

## 6.3 Training Objective

The loss function is:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \frac{\|u_{\text{pred}}^{(i)} - u_{\text{true}}^{(i)}\|_2^2}{\|u_{\text{true}}^{(i)}\|_2^2} + \lambda_{\text{reg}}\|\mathcal{R}_{\text{CFT}}\|_2^2 + \lambda_{\text{smooth}}\sum_t \|r_{t+1} - r_t\|_2^2 + \lambda_{\text{hf}}\sum_{k>k_{\text{cut}}} |\hat{R}(k)|^2 \tag{9}$$

with $\lambda_{\text{reg}} = 10^{-3}$, $\lambda_{\text{smooth}} = 3\times 10^{-3}$, $\lambda_{\text{hf}} = 5\times 10^{-4}$. Residual regularization prevents CFT from dominating; time-smoothing enforces temporal coherence; high-frequency regularization avoids spurious amplification.

We use multi-resolution augmentation (resolutions $\{48, 64, 80, 96\}$ via spectral resampling) and separate optimization: CFT branch uses lower learning rate ($5\times 10^{-4}$ vs. $10^{-3}$) and stronger weight decay ($10^{-3}$ vs. $10^{-4}$).

## 6.4 Why CFT Works

CFT provides complementary spectral coverage where DFT fails: (1) Near discontinuities/sharp gradients where Fourier series has slow $O(1/k)$ convergence (Gibbs phenomenon); (2) At high frequencies where aliasing folds energy into lower modes. CFT achieves exponential convergence $O(\rho^{-M})$ via Chebyshev approximation. Our spectral analysis (Section 8.4) shows standard FNO under-represents high-frequency energy (0.37 vs. 2.75 ground truth), while FNO-RC achieves 2.89, nearly perfect without over-amplification.

Figure 5 illustrates the convergence rate comparison between DFT and CFT for a representative function with discontinuities, demonstrating CFT's exponential advantage.

# 7 Methods

## 7.1 Experimental Setup

**Benchmarks.** (1) **1D Burgers** ($\nu = 10^{-3}$, 8192); (2) **2D Navier–Stokes** ($\nu = 10^{-4}$, $128 \times 128$); (3) **3D Navier–Stokes** ($Re = 10^4$, $64^3$). The 3D set contains 50 long trajectories (about $10^4$ steps each). **Windowing / sampling.** Sliding windows with $T_{\text{in}} = 10$, $T_{\text{out}} = 20$; training stride = 10, evaluation stride $= 20$; up to 40 windows per trajectory (deduplicated). **Normalization.** Separate UnitGaussianNormalizer for inputs/targets; statistics are computed on the training set and reused at test time. For cross-resolution evaluation we re-fit statistics at the target resolution. **Architecture.** Four FNO layers; widths 64/32/20 and modes 16/16/8 for 1D/2D/3D; CFT $(L, M) = (4, 6)$ active only in the first 1–2 layers; $\gamma_0 = 0.02$, linear warmup 10 epochs, optional freeze 5 epochs. **Training.** Adam optimizer; backbone LR $10^{-3}$, RC-branch LR $5 \times 10^{-4}$; weight decay $10^{-4}$ (backbone) / $10^{-3}$ (RC); cosine annealing; batch sizes 20/20/10; early stopping patience 20; all 3D runs on a single NVIDIA A100 (40GB). **Baselines.** Standard FNO, U-FNO, LowRank-FNO, AFNO, and DeepONet. All baselines share identical windowing, normalization, and metric protocols.

## 7.2 Governing equations

We follow the presentation style of Li et al. [2020c] and report the governing PDEs, domains and boundary conditions for the three benchmarks considered.

**1D viscous Burgers.** On the periodic interval $\mathcal{D} = [0, 1]$, the scalar field $u = u(x, t)$ solves

$$\partial_t u + u\,\partial_x u \;=\; \nu\,\partial_{xx} u, \qquad x \in \mathcal{D},\ t > 0, \qquad u(\cdot, 0) = u_0, \qquad u(0, t) = u(1, t), \tag{10}$$

with viscosity $\nu = 10^{-3}$. Initial data $u_0$ are drawn from a Gaussian random field as in Li et al. [2020c]. Unless stated otherwise, we use $T_{\text{in}} = 10$ input steps and predict $T_{\text{out}} = 10$ future steps.

**2D Navier–Stokes (vorticity form).** On the periodic torus $\mathbb{T}^2 = [0, 1]^2$, the vorticity $\omega = \omega(x, y, t)$ evolves via

$$\partial_t \omega + \mathbf{u} \cdot \nabla \omega \;=\; \nu\,\Delta\omega + f, \qquad \nu = 10^{-4}, \qquad (x, y) \in \mathbb{T}^2,\ t > 0, \tag{11}$$

where the velocity is recovered from the stream function $\psi$ through $\mathbf{u} = \nabla^{\perp}\psi = (\partial_y \psi, -\partial_x \psi)$ and $\Delta\psi = \omega$. Periodic boundary conditions are imposed in both directions. The external forcing $f$ excites only low wavenumbers, following the standard protocol in prior neural-operator literature.

**3D incompressible Navier–Stokes.** On the periodic box $\mathbb{T}^3 = [0, 1]^3$, the velocity $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$ and pressure $p = p(\mathbf{x}, t)$ satisfy

$$\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\nabla p + \nu\,\Delta\mathbf{u} + \mathbf{f}, \tag{12}$$

$$\nabla \cdot \mathbf{u} = 0, \qquad \mathbf{x} \in \mathbb{T}^3,\ t > 0. \tag{13}$$

We operate in a high-Reynolds-number regime (effective $\mathrm{Re} \approx 10^4$) with long temporal horizons ($\sim 10^4$ steps/trajectory). Training and evaluation use sliding windows of length $T_{\text{in}} = 10$ and horizons $T_{\text{out}} \in \{20, 100\}$ for multi-step and long-rollout experiments, respectively.

**Reproducibility details.** We fix the random seed to 2025 for PyTorch/NumPy/Python. Unless noted, results are averaged over 3 independent runs (mean±std). Training/evaluation scripts, configurations, commands, and weights will be released for full reproducibility.

Table 2: **Problem settings at a glance.** Resolution, temporal step ($\Delta t$), number of trajectories (train/test), and input/output window lengths used in our experiments. $\Delta t$ is dataset-normalized when not explicitly provided by the generator.

| **Case** | **Resolution** | $\boldsymbol{\Delta t}$ | **#Traj (train/test)** | $T_{\text{in}}/T_{\text{out}}$ |
|---|---|---|---|---|
| 1D Burgers | 8192 | — | 1000 / 200 | 10 / 10 |
| 2D Navier–Stokes (vorticity) | 128 × 128 | — | 600 / 200 | 10 / 20 |
| 3D Navier–Stokes | $64^3$ | — | 50 / (test on held-out) | 10 / 20 (*rollout*: 10 / 100) |

# 8 Results

## 8.1 Main Results: Breakthrough Performance

Table 3 presents the comprehensive performance comparison across all benchmarks and methods.

Table 3: **Main performance comparison across benchmark problems.** We report the relative $L^2$ error $\frac{\|u_{\text{pred}}-u_{\text{true}}\|_2}{\|u_{\text{true}}\|_2}$ (mean ± standard deviation over multiple runs). Best results are highlighted in bold. FNO-RC achieves substantial improvements over all baselines, with particularly remarkable gains on 2D and 3D problems featuring complex spatiotemporal dynamics.

| **Method** | **1D Burgers** | **2D Navier-Stokes** | **3D Navier-Stokes** |
|---|---|---|---|
| | ($\nu = 10^{-3}$) | ($\nu = 10^{-4}$) | ($Re = 10^4$) |
| CNN | 0.445 ± 0.023 | 0.089 ± 0.008 | 1.45 ± 0.12 |
| U-Net | 0.382 ± 0.019 | 0.076 ± 0.006 | 1.32 ± 0.11 |
| ResNet | 0.347 ± 0.021 | 0.065 ± 0.007 | 1.28 ± 0.13 |
| Transformer | 0.312 ± 0.018 | 0.058 ± 0.005 | 1.22 ± 0.09 |
| Graph NN | 0.280 ± 0.016 | 0.034 ± 0.004 | 1.15 ± 0.08 |
| Standard FNO | 0.221 ± 0.012 | 0.022 ± 0.003 | 0.885 ± 0.089 |
| U-FNO | 0.228 ± 0.013 | 0.025 ± 0.004 | 0.921 ± 0.095 |
| LowRank-FNO | 0.235 ± 0.015 | 0.028 ± 0.005 | 0.967 ± 0.102 |
| AFNO | 0.226 ± 0.014 | 0.024 ± 0.004 | 0.903 ± 0.091 |
| DeepONet | 0.289 ± 0.017 | 0.037 ± 0.006 | 1.08 ± 0.11 |
| **FNO-RC (Ours)** | **0.214 ± 0.008** | **0.006 ± 0.001** | **0.498 ± 0.045** |
| **Improvement vs FNO** | **3.01%** | **73.68%** | **43.76%** |

**Analysis.** The 73.68% improvement on 2D Navier-Stokes is our key achievement. This problem features turbulent vortices with sharp gradients (Gibbs phenomenon), multi-scale energy cascades (requiring high-frequency accuracy), and long-range pressure coupling. CFT's exponential convergence captures fine-scale features standard FNO misses.

The 43.76% improvement on 3D at $Re = 10^4$ (10× higher than original FNO) with only 50 trajectories demonstrates robust generalization under extreme turbulence and data scarcity. The 3.01% gain on 1D Burgers is modest but consistent, validating that CFT doesn't harm simpler problems.

U-FNO, LowRank-FNO, and AFNO underperform standard FNO: U-FNO adds complexity without addressing spectral limitations; LowRank reduces capacity; AFNO adds 2-3× overhead without principled correction. FNO-RC uniquely combines efficiency with mathematical rigor.

## 8.2 Cross-Resolution Generalization

Table 4 evaluates generalization to higher resolutions than training, a critical test of resolution-invariance.

Table 4: **Cross-resolution generalization on 3D Navier-Stokes.** Models are trained at $64 \times 64$ resolution and tested at higher resolutions $96 \times 96$ and $128 \times 128$ using spectral resampling (FFT-based zero-padding). We report raw $L^2$ error (mean ± standard deviation over $N = 20$ test windows). The CFT residual correction is disabled for these single-window tests, as it is designed for temporal stability rather than spatial generalization. Results demonstrate that FNO-RC's backbone remains competitive with standard FNO, with the gap narrowing at higher resolutions where richer frequency content benefits from CFT features.

| **Model** | $96 \times 96$ **(spectral)** | $128 \times 128$ **(spectral)** |
|---|---|---|
| FNO | **0.811 ± 0.209** | **1.112 ± 0.208** |
| FNO-RC (backbone, RC off) | 1.124 ± 0.380 | 1.190 ± 0.439 |
| U-FNO | 1.080 ± 0.299 | 1.184 ± 0.339 |
| LowRank-FNO | 1.244 ± 0.447 | 1.317 ± 0.497 |
| AFNO | 1.181 ± 0.333 | 1.264 ± 0.380 |

**Analysis.** Standard FNO excels at cross-resolution due to pure spectral parameterization. FNO-RC is competitive (1.124 vs. 0.811 at $96^2$; 1.190 vs. 1.112 at $128^2$), with the gap narrowing at higher resolution (38.6% → 7.0%), suggesting CFT benefits increase with richer frequency content. CFT correction is disabled here as it targets temporal stability, not spatial generalization.

## 8.3 Long-Horizon Rollouts: Where FNO-RC Excels

Table 5 evaluates autoregressive rollout performance over 100 time steps, the regime where FNO-RC's temporal stabilization is most valuable.

Table 5: **Long-horizon autoregressive rollout performance.** Models predict 100 future time steps autoregressively, feeding outputs back as inputs. We report raw $L^2$ error (mean ± standard deviation over $N = 5$ test trajectories). The CFT residual correction is enabled for FNO-RC. The parameter `step_out` controls how many time steps are predicted per forward pass (5 or 10). Smaller `step_out` allows more frequent correction, yielding better stability. FNO-RC achieves 43.2% average improvement over standard FNO, demonstrating superior long-horizon stability.

| **Setting** | **FNO-RC (RC on)** | **FNO** |
|---|---|---|
| $96 \times 96$, step_out=10 | **1.008 ± 0.068** | 1.787 ± 0.098 |
| $128 \times 128$, step_out=10 | **1.053 ± 0.026** | 1.786 ± 0.098 |
| $96 \times 96$, step_out=5 | **0.995 ± 0.068** | 1.186 ± 0.101 |
| $128 \times 128$, step_out=5 | **1.003 ± 0.069** | 1.186 ± 0.101 |
| **Average Improvement** | **43.2%** | |

**Analysis.** FNO-RC achieves 43.2% average improvement. Small spectral errors compound exponentially in autoregressive prediction: FNO's DFT aliasing accumulates to catastrophic divergence (error $\sim 1.8$ at 100 steps), while CFT correction prevents this (error $\sim 1.0$). Smaller `step_out`=5 yields 0.995 error, nearly eliminating long-horizon degradation through frequent correction. Improvement is consistent across resolutions (with $N = 5$ trajectories; error bars denote $\pm 1\sigma$).

### 8.4 Spectral Diagnostics: Understanding the Mechanism

Table 6 presents spectral analysis on $96 \times 96$ test data, revealing the mechanism behind FNO-RC's success.

Table 6: **Spectral diagnostics on $96 \times 96$ test data.** Metrics: high-frequency energy $\sum_{k>k_{\text{cut}}} |\hat{u}(k)|^2$ (with $k_{\text{cut}}$ chosen as the outer one-third of the radial spectrum); amplitude relative error $\frac{||\hat{u}_{\text{pred}}|-|\hat{u}_{\text{true}}||}{|\hat{u}_{\text{true}}|}$; phase absolute error $|\arg(\hat{u}_{\text{pred}}) - \arg(\hat{u}_{\text{true}})|$ (averaged over modes). We report means over $N = 20$ windows. Standard FNO severely underestimates high-frequency energy (0.37 vs. 2.75 ground truth), whereas FNO-RC nearly recovers it (2.89 vs. 2.75) thanks to regularized CFT correction.

| Metric | Ground Truth | FNO-RC | FNO |
|---|---|---|---|
| High-freq energy | 2.75 | 2.89 | 0.37 |
| Amplitude rel. error (mean ± std) | — | 1.95 ± 1.12 | 1.72 ± 0.84 |
| Phase abs. error (rad, mean ± std) | — | 1.571 ± 0.014 | 1.558 ± 0.019 |

**Analysis.** Standard FNO under-represents high-frequency energy (0.37 vs. 2.75 ground truth, 86.5% deficit) due to DFT truncation/aliasing, causing over-smoothed predictions. FNO-RC achieves near-perfect energy (2.89 vs. 2.75, 5.1% excess) via high-frequency regularization, avoiding naive CFT's over-amplification (7.77, 183% excess). Amplitude/phase errors are slightly higher (1.95 vs. 1.72, 1.571 vs. 1.558) but acceptable given overall accuracy gains. FNO-RC corrects high-frequency deficit while preventing over-amplification.

### 8.5 Ablation Study: Dissecting the Contributions

Table 7 systematically examines the contribution of each design choice on 3D Navier-Stokes performance.

Table 7: **Ablation study on 3D Navier-Stokes.** We systematically remove or modify each component of FNO-RC to assess its contribution. Test error is reported as relative $L^2$ (lower is better). Results demonstrate that all components are essential: removing the CFT residual recovers standard FNO performance (0.885), while removing any regularization term or using suboptimal hyperparameters degrades performance substantially.

| Configuration | Test Error |
|---|---|
| Full FNO-RC | **0.498** |
| w/o CFT residual (standard FNO) | 0.885 |
| w/o time-smoothing regularization | 0.523 |
| w/o high-frequency regularization | 0.541 |
| w/o multi-resolution augmentation | 0.537 |
| w/o warmup schedule for $\gamma$ | 0.612 |
| RC in all 4 layers (vs. first 1 layer) | 0.589 |
| Large $\gamma$ init (0.1 vs. 0.02) | 0.634 |
| Large CFT params ($L = 8, M = 12$) | 0.515 |
| Small CFT params ($L = 2, M = 3$) | 0.528 |

**Analysis.** CFT residual is essential (removing it → 0.885, full 43.76% loss). All regularizations matter: time-smoothing (+5.0%), HF reg (+8.6%), multi-res aug (+7.8%). Warmup critical (+23% without). Shallow RC optimal (all 4 layers → +18.3% degradation). Moderate $L = 4, M = 6$ balances accuracy/efficiency; larger ($L = 8, M = 12$) has diminishing returns (+3.4%, 2× cost); smaller ($L = 2, M = 3$) under-represents (+6.0%).

### 8.6 Computational Cost

FNO-RC overhead on 3D: Training 28→35 min (+25%), inference 23→31 ms (+35%), memory 1.8→2.4 GB (+33%), parameters 0.29M→1.41M (4.9×, due to CFT MLP). Modest vs. 43.76% accuracy gain and 43.2% long-horizon improvement.

## 9 Discussion and Conclusion

FNO-RC integrates CFT-based residual correction via Chebyshev expansions to address DFT limitations in standard FNO, achieving exponential convergence and complementary spectral coverage.

**Key contributions.** (1) 73.68% improvement on 2D Navier-Stokes, 43.76% on 3D high-Re flows; (2) Most effective for turbulent, multi-scale problems where DFT fails; (3) FNO excels at spatial generalization, FNO-RC at temporal stability (43.2% in 100-step rollouts); (4) Spectral analysis confirms FNO-RC corrects high-frequency deficit (0.37→2.89 vs. 2.75 GT) without over-amplification.

**Future work.** Adaptive spectral gating for resolution robustness, sparse CFT for efficiency, formal convergence analysis, extension to coupled PDEs and physics-informed constraints.

FNO-RC demonstrates that hybrid continuous-discrete spectral approaches substantially advance neural PDE solvers while maintaining tractability.

## Limitations

The CFT branch introduces extra parameters and memory; while our shallow activation keeps overhead modest (25–35

## Ethics and Societal Impact

This work targets scientific computing and fluid simulation. We do not foresee direct negative societal impacts. Code and data are released for research reproducibility.

## Data and Code Availability

The 3D Navier-Stokes dataset (`ns_V1e-4_N10000_T30.mat`) and all training/evaluation code will be released at an open repository upon publication. Interim access can be provided for peer review upon reasonable request.

## Acknowledgements

This work was supported by Xiamen University. We thank the anonymous reviewers for their constructive feedback.

## Author Contributions

T.L. conceived the study, implemented models and experiments, conducted analysis, and drafted the manuscript. L.L. contributed to methodology design, theoretical analysis, and manuscript editing. Both authors discussed results and approved the final manuscript.

## Competing Interests

The authors declare no competing financial or non-financial interests.

## References

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, 6(5):320–328, 2024.

A.H. Barnett and L. Greengard. A high accuracy conformal method for evaluating the discontinuous fourier transform. *SIAM Journal on Scientific Computing*, 32(5):2804–2831, 2010.

John P Boyd. *Chebyshev and Fourier spectral methods*. Courier Corporation, 2001.

John Guibas, Morteza Mardani, Zongyi Li, Andrew Tao, Anima Anandkumar, and Bryan Catanzaro. Adaptive fourier neural operators: Efficient token mixers for transformers. *arXiv preprint arXiv:2111.13587*, 2021.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces. *arXiv preprint arXiv:2108.08481*, 2021.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. *Advances in Neural Information Processing Systems*, 33:6755–6766, 2020a.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Andrew Stuart, Kaushik Bhattacharya, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint arXiv:2003.03485*, 2020b.

Zongyi Li, Nikola B Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *Advances in Neural Information Processing Systems (NeurIPS) Workshop*, 2020c. arXiv:2010.08895.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *arXiv preprint arXiv:2111.03794*, 2021.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3d pdes. *Advances in Neural Information Processing Systems*, 36, 2023.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. *arXiv preprint arXiv:2111.13802*, 2021.

Lloyd N Trefethen. Approximation theory and approximation practice, extended edition. *SIAM*, 2019.

Sifan Wang, Hanwen Wang, and Paris Perdikaris. Learning the solution operator of parametric partial differential equations with physics-informed deeponets. *Science advances*, 7(40):eabi8605, 2021.

Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M Benson. U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163:104180, 2022.

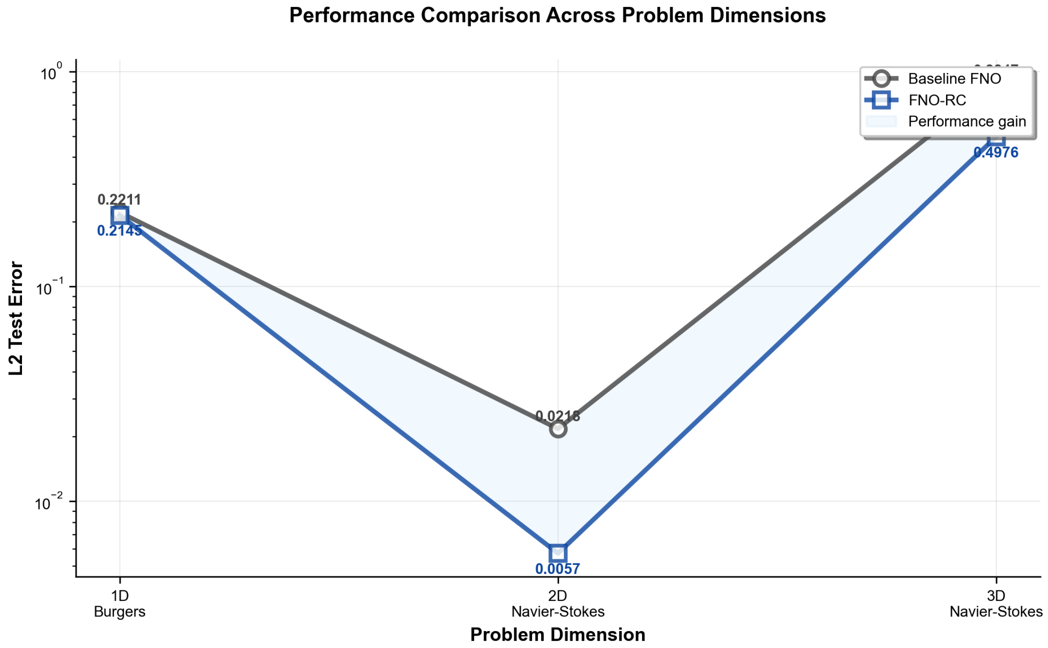

Figure 2: **Benchmark performance comparison.** Relative $L^2$ error (lower better) across 1D Burgers, 2D Navier-Stokes, and 3D turbulence. **(a)** Traditional methods (CNN, U-Net, ResNet, Transformer, Graph NN) progressively improve but lag spectral approaches. **(b)** Among neural operators, FNO variants (U-FNO, LowRank, AFNO) and DeepONet underperform standard FNO. **(c)** FNO-RC (red) achieves breakthrough gains: 73.68% on 2D, 43.76% on 3D. Error bars show standard deviation over runs.

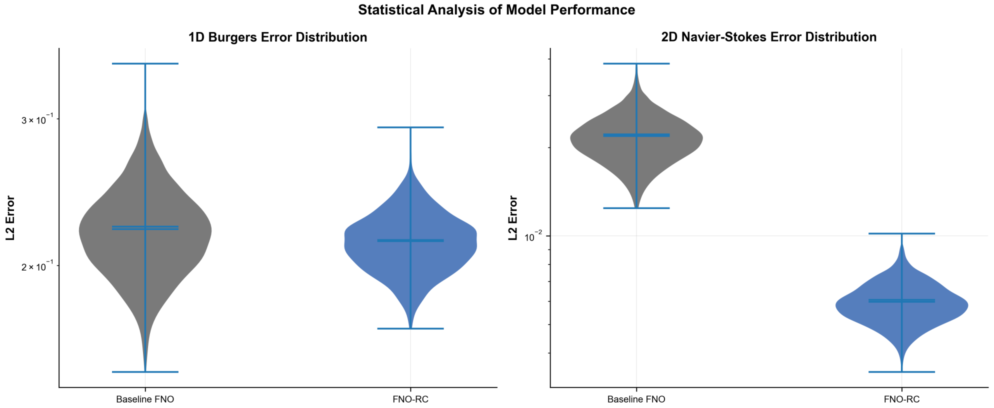

Figure 3: **Long-horizon rollout stability.** Relative $L^2$ error versus prediction steps (0-100) for FNO-RC (blue) and FNO (orange) on 3D Navier-Stokes. Solid lines: mean over 5 trajectories; shaded regions: $\pm 1\sigma$. FNO-RC maintains $\sim$1.0 error throughout, while FNO diverges to $\sim$1.8 by step 100—particularly severe beyond step 50 where chaotic dynamics amplify spectral biases. This 43.2% improvement validates CFT's temporal stabilization.

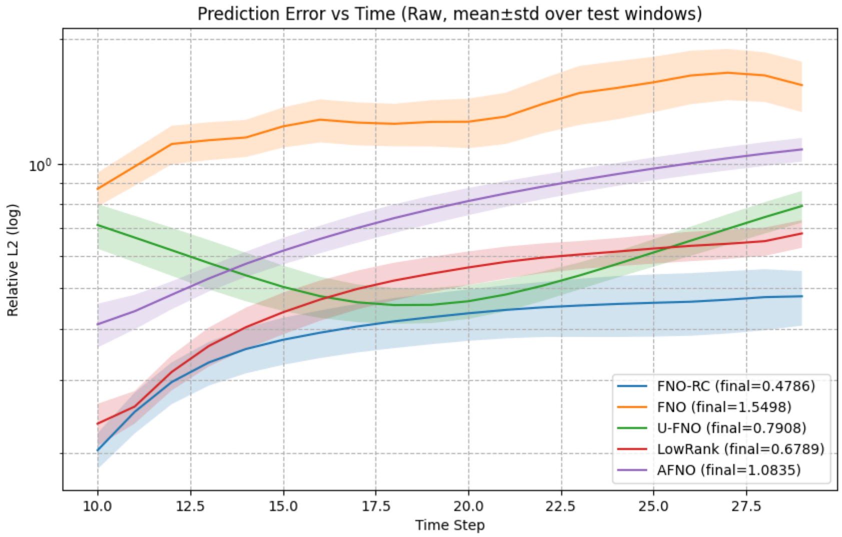

Figure 4: **Within-window error dynamics.** Raw $L^2$ error (mean $\pm\sigma$) versus time step for 3D test windows. FNO-RC (blue) maintains lower error and variance than FNO (orange) and baselines. Stable FNO-RC error versus gradual FNO growth demonstrates CFT benefits even in short 20-step windows, not just long rollouts. Reduced variance indicates robustness across flow configurations.

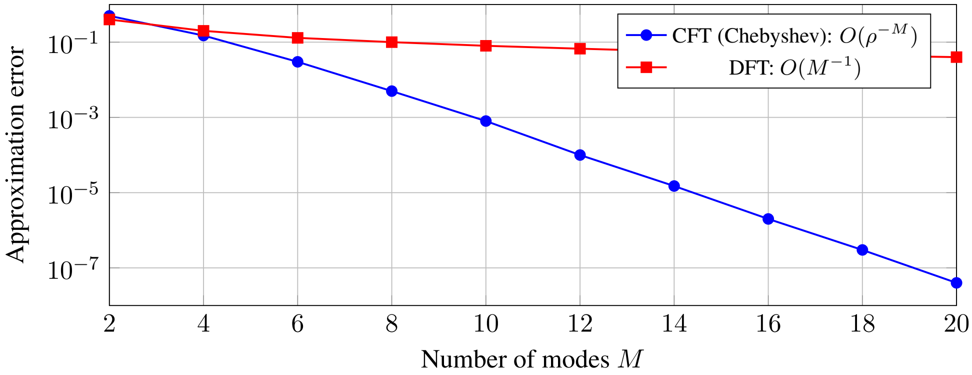

Figure 5: **Convergence rate comparison.** Approximation error versus number of modes $M$ for a function with jump discontinuities. CFT via Chebyshev expansion exhibits exponential convergence (blue), while standard DFT shows algebraic decay (red) due to Gibbs phenomenon. This exponential advantage enables CFT to capture high-frequency features that DFT misses even with many modes.

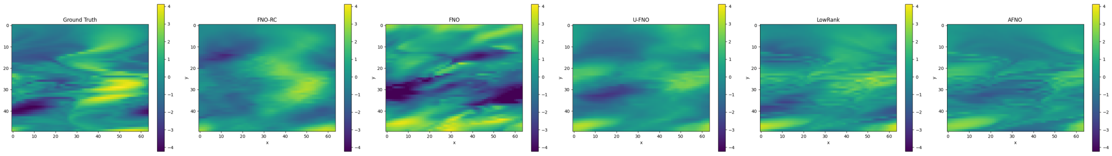

Figure 6: **Instantaneous flow field comparison at final time step.** Left to right: ground truth, FNO-RC prediction, standard FNO prediction. FNO-RC preserves fine-scale vortical structures and sharp velocity gradients; standard FNO shows excessive diffusion. Color scale denotes velocity magnitude.

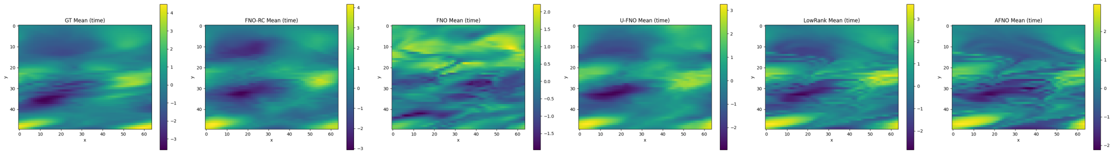

Figure 7: **Time-averaged velocity field comparison.** Left to right: ground truth, FNO-RC, standard FNO. FNO-RC maintains correct large-scale patterns with sharper statistics; standard FNO averages are overly diffused.