Efficient Surrogate Modeling for Fast Hydrodynamic Evolution of the Quark-Gluon Plasma

Seungjun Lee [1] **Yeonju Go** [1] **Bjoern Schenke** [1] **Chun Shen** [2] **Yihui Ren** [1] **Jin Huang** [1] **Maria Chamizo-Llatas** [1]

Abstract

We propose a neural operator-based surrogate model for fast simulation of 2+1D and 3+1D quark-gluon plasma (QGP) hydrodynamic evolution. Accurate event-by-event simulations require solving computationally expensive PDEs, often taking hours per event. While neural operators enable efficient operator learning, existing approaches assume static domains and do not explicitly capture dissipative dynamics, evolving spatial support, or scale efficiently to spatiotemporal settings. We introduce a time-conditioned neural operator that jointly models field evolution, energy dissipation, and dynamic spatial support. Our approach combines a factorized spatial neural operator with a time-conditioned decoder that adaptively modulates latent features over proper time, along with a gated spatial masking module to capture the evolving support of the plasma fireball. Experiments on simulation data generated by the IP-Glasma+MUSIC framework show accurate predictions of key physical observables while achieving orders-of-magnitude speedup over traditional solvers. These results highlight the potential of our proposed model as a scalable and high-fidelity surrogate for QGP hydrodynamics.

1. Introduction

The quark-gluon plasma (QGP) is a deconfined state of strongly interacting matter produced in ultra-relativistic heavy-ion collisions at facilities such as the Relativistic Heavy Ion Collider (RHIC) and the Large Hadron Collider (LHC) (Shuryak, 1980; Gyulassy & McLerran, 2005). Connecting the properties of this hot, dense medium to experimentally measured final-state observables requires a multi-stage simulation pipeline: an initial-state model (e.g., IP-Glasma (Schenke et al., 2012)), viscous relativistic hydrodynamics (e.g., MUSIC (Schenke et al., 2011; Paquet et al., 2016)), particlization, and a hadronic afterburner. The hydrodynamic stage, which governs the bulk collective evolution of the plasma, is the most computationally demanding component, typically consuming several minutes per event for 2+1D (boost-invariant) evolution and up to 20 hours for a full 3+1D simulation on a single CPU core.

Extracting transport coefficients such as the shear viscosity-to-entropy ratio η/s and bulk viscosity ζ/s through Bayesian inference (Bernhard et al., 2019; Nijs et al., 2021) or performing uncertainty quantification requires millions of independent event-by-event simulations, making the computational cost prohibitive. This motivates the development of fast and accurate surrogate models that can replace or accelerate the hydrodynamic stage.

Data-driven approaches based on deep learning have recently shown great promise for scientific simulation acceleration. In particular, Fourier Neural Operators (FNOs) (Li et al., 2021) learn mappings between infinite-dimensional function spaces and are well-suited to PDE-governed dynamics. Unlike standard convolutional networks, FNOs operate in Fourier space and are therefore resolution-invariant, making them directly applicable to spatially structured hydrodynamic fields. However, the standard FNO was developed for stationary or single-step operator learning and does not naturally handle (i) the monotonically dissipating energy and shrinking spatial support of the fireball, (ii) the sharp, moving boundary separating the hot plasma interior from the vacuum, or (iii) the multi-step temporal rollout required for full event evolution.

A concurrent independent study by (Stewart & Putschke, 2026) applies standard FNOs to 2+1D QGP evolution within the JETSCAPE Monte Carlo framework and evaluates flow and jet-quenching observables. While that work establishes FNO as a viable surrogate for this problem, it uses a single-pass, time-agnostic FNO with standard per-channel normalization and does not address the temporal dissipation structure, evolving spatial support, or the extension to 3+1D. Our work directly addresses these open challenges.

In this work, we make the following contributions:

[1]Brookhaven National Laboratory [2]Wayne State University. Correspondence to: Seungjun Lee <slee19@bnl.gov>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

- We propose a **time-conditioned neural operator for QGP (TC-QGP)** that encodes the initial condition with a Tucker-factorized FNO (TFNO) and decodes all future timesteps in a single forward pass via proper time-conditioned adaptive group normalization, capturing the dissipative temporal dynamics of the fireball.
- We introduce a **gated spatial masking (GSM)** module that learns to suppress activations in vacuum regimes, preserving the sharp evolving boundary of the plasma.
- We demonstrate **sub-percent predictive accuracy** across all five hydrodynamic fields $(\epsilon, T, P, u_x, u_y)$ for 2+1D evolution with a 10^2–$10^3\times$ speedup over MUSIC, and extend the framework to full **3+1D** evolution ($\epsilon, T, P, u_x, u_y, u_z$ on a $28 \times 108 \times 108 \times 40$ domain), where the target speedup exceeds $5000\times$.

2. Related Works

Neural Operators. Neural operators (Kovachki et al., 2023) learn mappings between function spaces, providing resolution-invariant predictions for complex physical systems. Fourier neural operator (FNO) (Li et al., 2021) has achieved state-of-the-art results in fluid dynamics and weather forecasting (Pathak et al., 2022) by parameterizing integral kernels in the spectral domain. Subsequent work has expanded the FNO framework through tensorized formulations (Tran et al., 2021; Kossaifi et al., 2023), geometry-aware variants (Li et al., 2023a;b), and operator transformers (Cao, 2021; Li et al., 2022; Lee, 2022; Hao et al., 2023; Lee & Oh, 2024; Wu et al., 2024). The most directly related work is (Stewart & Putschke, 2026), which applies a standard FNO as a surrogate for the MUSIC solver within the JETSCAPE Monte Carlo framework (Putschke et al., 2019). In this work, we employ the Tucker-factorized FNO (TFNO) (Kossaifi et al., 2023) as the spatial backbone for its memory efficiency in 3+1D settings.

Time Conditioning. Conditioning network activations on an auxiliary variable is a broadly used technique for incorporating external context into learned representations. Feature-wise linear modulation (FiLM) (Perez et al., 2018) introduced channel-wise scale and shift from an auxiliary embedding, and was later applied to visual reasoning and multi-task learning. Recently, adaptive group normalization (AdaGN) (Dhariwal & Nichol, 2021) conditions each residual block on both the noise level and class label, enabling fine-grained control over the denoising trajectory; this idea was further scaled in (Peebles & Xie, 2023). For time-dependent PDEs, temporal conditioning has been explored in autoregressive neural operators (Li et al., 2021) and time-continuous frameworks such as neural ODEs (Chen et al., 2018), but these either accumulate rollout error or require expensive sequential inference. We adapt AdaGN to the non-autoregressive neural operator setting, conditioning the decoder on proper time τ so that the model simultaneously captures temporal field structure and the dissipative energy scaling of the QGP fireball.

Learning on Compact Support and Sharp Interfaces. Many scientific systems exhibit compactly supported structure, where nonzero values are confined to a physically active region surrounded by a quiescent background. Global spectral operators such as FNO (Li et al., 2021) exhibit a spectral bias toward low-frequency, smooth modes, making them prone to boundary smearing in the presence of sharp interfaces or discontinuities. This limitation has been addressed in multiphase flow modeling. (Wen et al., 2022) augments spectral operators with a parallel U-Net branch to preserve high-frequency boundary features, while (Wang et al., 2024) explicitly encodes interface geometry to mitigate spectral bias. In QGP hydrodynamics, however, the evolving support of the plasma is irregular, highly dynamic, and varies event-by-event, making explicit interface supervision impractical. We instead introduce a gated sparsity module that implicitly learns the evolving support from data. Independently, the H^1 Sobolev loss component in our training objective (Son et al., 2021), which penalizes gradient discrepancies, provides a complementary regularization that improves boundary fidelity without requiring explicit support knowledge.

3. Preliminaries

3.1. Relativistic Viscous Hydrodynamics for QGP

In the Milne coordinate system (τ, x, y, η_s), where $\tau = \sqrt{t^2 - z^2}$ is the longitudinal proper time and η_s is the space-time rapidity, the bulk evolution of the QGP is governed by local conservation of the energy-momentum tensor:

$$D_\mu T^{\mu\nu} = 0, \tag{1}$$

where $T^{\mu\nu} = (\epsilon + P)u^\mu u^\nu - P g^{\mu\nu} + \pi^{\mu\nu}$ includes the local energy density ϵ, pressure P, four-velocity u^μ, and shear stress tensor $\pi^{\mu\nu}$. The equation of state $P = P(\epsilon)$ closes the system. In the boost-invariant (2+1D) limit, η_s-dependence is dropped, reducing the spatial domain to (x, y). For the full 3+1D case, all four coordinates are retained.

Given an initial condition at τ_0 specifying the field $\mathbf{f}_0(x, y, \eta_s) = [\epsilon, T, P, u_x, u_y, u_z]_{\tau_0}$, the MUSIC solver propagates the system forward in proper time, producing a sequence $\{\mathbf{f}_\tau\}_{\tau=\tau_0}^{\tau_{\max}}$. Our goal is to learn a surrogate operator $\mathcal{G} : \mathbf{f}_0 \mapsto \{\hat{\mathbf{f}}_\tau\}_{\tau>\tau_0}$ that approximates this mapping without solving Eq. (1) explicitly.

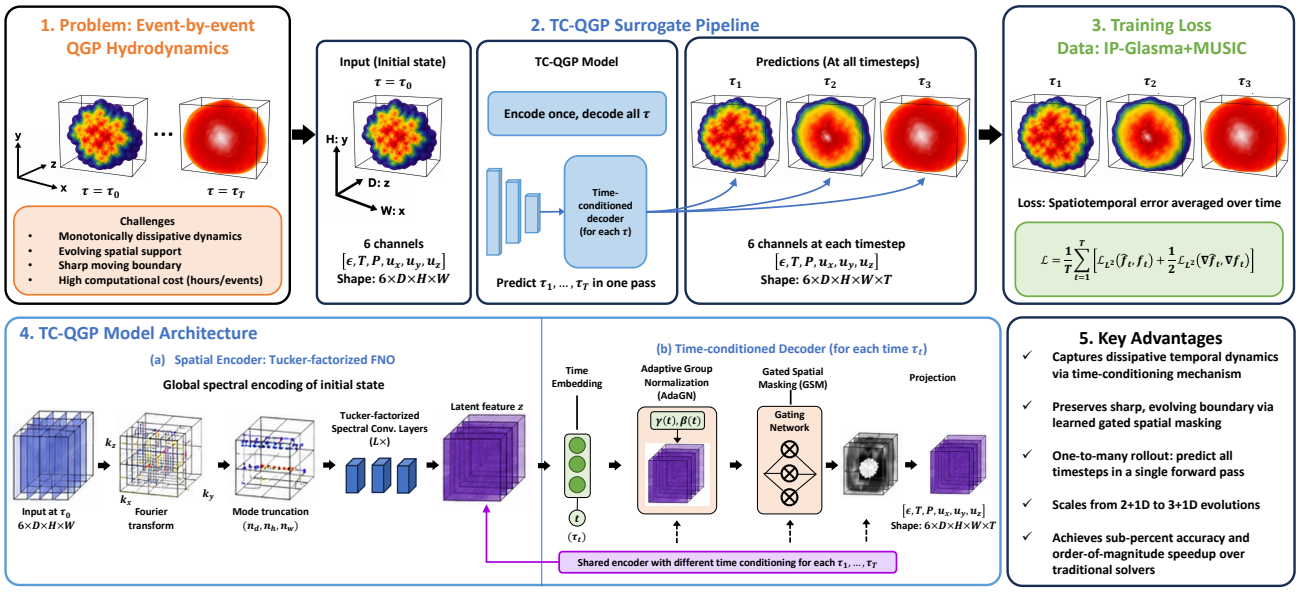

Figure 1. **Overview of the TC-QGP surrogate modeling framework.** (1) **Problem:** Event-by-event QGP hydrodynamics exhibits dissipative dynamics, evolving spatial support, moving sharp boundaries, and high computational cost. (2) **Pipeline:** From an initial state at τ_0 with six fields $[\epsilon, T, P, u_x, u_y, u_z]$, the model encodes once and predicts all T future timesteps in a single forward pass. (3) **Training:** A spatiotemporal loss combining field and gradient errors, averaged over time. (4) **Architecture:** A Tucker-factorized FNO encoder produces a latent representation, followed by a time-conditioned decoder with AdaGN and gated spatial masking (GSM). (5) **Advantages:** Captures temporal dynamics and sharp boundaries, enables one-to-many rollout, scales to 3+1D, and achieves high accuracy with significant speedup.

3.2. Neural Operators

A neural operator (Kovachki et al., 2023) approximates a mapping between two function spaces $\mathcal{G} : \mathcal{A} \to \mathcal{U}$, where $a \in \mathcal{A}$ and $u \in \mathcal{U}$ are functions defined on bounded domains $D \subset \mathbb{R}^d$. Unlike standard deep networks, neural operators are discretization-invariant: the same set of learned parameters defines the operator at any resolution.

Fourier neural operator (FNO) (Li et al., 2021) parameterizes the operator via a series of Fourier integral operators composed with pointwise nonlinearities. Each FNO layer computes:

$$v_{l+1}(x) = \sigma\big(W_l v_l(x) + \mathcal{F}^{-1}\big[R_l \cdot \mathcal{F}[v_l]\big](x)\big), \quad (2)$$

where v_l is the l-th layer representation, W_l is a local linear operator, $\mathcal{F}$ denotes the Fourier transform, R_l is a learnable complex weight tensor truncated to the $k_{\max}$ lowest Fourier modes, and σ is an elementwise activation. By operating in Fourier space, FNO efficiently captures global, long-range correlations in the input field.

The Tucker-factorized FNO (TFNO) (Kossaifi et al., 2023) decomposes the spectral weight tensor R_l using Tucker decomposition:

$$R_l \approx \mathcal{G}_l \times_1 U_l^{(1)} \times_2 U_l^{(2)} \times_3 U_l^{(3)}, \quad (3)$$

where $\mathcal{G}_l$ is a compact core tensor and $U_l^{(i)}$ are factor matrices. This reduces the parameter count and memory footprint substantially in high-dimensional (e.g., 3D spatial) settings, which is essential for 3+1D QGP evolution.

Table 1. TC-QGP configuration for each setting. The decoder (AdaGN, GSM, projection) shares the same spatial dimension as the encoder.

	2+1D	3+1D
Spatial domain $\mathbf{S}$	100×100	$28\times108\times108$
Encoder	2D TFNO	3D TFNO
Fourier modes	$(k_H, k_W)=(50, 50)$	$(k_D, k_H, k_W)=(14, 54, 54)$
Spatial convolutions	Conv2d	Conv3d
Fields C	5: $(\epsilon, T, P, u_x, u_y)$	6: $(\epsilon, T, P, u_x, u_y, u_z)$
Output steps T	60	40

4. Method

We describe our proposed architecture, the **time-conditioned QGP operator (TC-QGP)**, which predicts the full multi-step temporal evolution of the QGP hydrodynamic fields from a single initial condition. The architecture is identical in design for both the 2+1D and 3+1D settings. The two cases differ only in the spatial dimensionality of the encoder and the convolution operators used in the decoder, as summarized in Table 1. The overall pipeline is illustrated in Figure 1. We use $\mathbf{S}$ to denote the spatial domain: $\mathbf{S} = H\times W$ for 2+1D (boost-invariant) and $\mathbf{S} = D\times H\times W$ for 3+1D (rapidity-dependent). Given the initial field $\mathbf{f}_0 \in \mathbb{R}^{C\times\mathbf{S}}$ at proper time τ_0, the model outputs a predicted sequence $\hat{\mathbf{F}} \in \mathbb{R}^{C\times\mathbf{S}\times T}$ at T future timesteps.

4.1. Factorized Neural Operator Encoder

The encoder maps the initial condition $\mathbf{f}_0$ to a shared spatial latent representation that captures the global structure of the plasma fireball:

$$\mathbf{z} = \mathrm{TFNO}(\mathbf{f}_0) \in \mathbb{R}^{d_h \times \mathbf{S}}, \tag{4}$$

where TFNO applies L layers of Tucker-factorized spectral convolution as in Eq. (2), with hidden dimension d_h=64. Tucker factorization substantially reduces the memory footprint of the spectral weight tensors, which are critical in the 3+1D case. In 3+1D, the lower mode count along the rapidity direction (k_D=14) reflects the smooth η_s-dependence over the D=28 rapidity slices, while the larger transverse modes (k_H=k_W=54) resolve the finer lateral structure from nucleon fluctuations across the 108×108 transverse grid. In 2+1D, the symmetric mode count (k_H=k_W=50) covers the isotropic transverse structure of the boost-invariant fireball on the 100×100 grid. Because $\mathbf{z}$ is computed once from the initial condition and reused across all T decoding steps, the encoding cost is fully amortized. This single-pass design avoids the error accumulation inherent in autoregressive rollout and ensures globally consistent spatial representations throughout the evolution.

4.2. Adaptive Time-Conditioned Decoding

QGP hydrodynamics exhibits a broad dynamic range in time: rapid dissipation of energy density and temperature at early proper times transitions to slower collective flow build-up at later times. This temporal structure is the same physics challenge in both 2+1D and 3+1D. A static decoder cannot capture it, so we condition each decoding step on the normalized proper time $\tilde{\tau}_t = (t\,\Delta\tau)/(\tau_{\max} - \tau_0)$.

Time Embedding. A two-layer MLP with GELU activation (Hendrycks & Gimpel, 2016) maps $\tilde{\tau}_t$ to a d_h-dimensional vector:

$$\mathbf{e}_t = \mathrm{MLP}(\tilde{\tau}_t) \in \mathbb{R}^{d_h}. \tag{5}$$

Adaptive Group Normalization (AdaGN). The shared latent $\mathbf{z}$ is modulated per timestep via AdaGN (Dhariwal & Nichol, 2021):

$$\mathrm{AdaGN}(\mathbf{z}, \mathbf{e}_t) = \mathrm{GN}(\mathbf{z}) \odot (1 + \boldsymbol{\gamma}_t) + \boldsymbol{\beta}_t, \tag{6}$$

where $[\boldsymbol{\gamma}_t, \boldsymbol{\beta}_t] = W_{\mathrm{emb}}\,\mathbf{e}_t$ are broadcast over $\mathbf{S}$. This channel-wise modulation enables the decoder to track both jointly the temporal field structure and the dissipative energy scale—an effect that applies equally in 2+1D and 3+1D.

The full per-timestep prediction is:

$$\hat{\mathbf{f}}_t = \mathrm{Conv}_1\big(\mathcal{G}\big(\mathrm{AdaGN}(\mathbf{z}, \mathbf{e}_t)\big)\big), \tag{7}$$

where Conv_1 is a pointwise convolution (1×1 in 2+1D, 1×1×1 in 3+1D) projecting from d_h to C channels, and $\mathcal{G}$ is the gated spatial masking module. All T frames are produced in a single forward pass.

4.3. Gated Spatial Masking for Evolving Field Support

A challenge common to both 2+1D and 3+1D QGP modeling is the sharp, time-evolving boundary between the hot plasma fireball and the surrounding vacuum: scalar fields (ϵ, T, P) and velocity fields are identically zero outside the fireball. Fourier-space global operations smear activations across the full domain, corrupting this physical boundary. This problem is more severe in 3+1D due to the additional rapidity dimension, but it is present in 2+1D as well.

We introduce a gated spatial masking (GSM) module:

$$\mathcal{G}(\mathbf{h}) = \mathbf{h} \odot \sigma\big(\mathrm{Conv}_1(\mathrm{ReLU}(\mathrm{Conv}_3(\mathbf{h})))\big), \tag{8}$$

where Conv_3 (kernel size 3, channels $d_h \to d_h/4$) and Conv_1 (kernel size 1, channels $d_h/4 \to d_h$) are 2D in the 2+1D case and 3D in the 3+1D case. The sigmoid gate $\sigma(\cdot) \in (0,1)^{d_h \times \mathbf{S}}$ learns near-zero activations in vacuum regions and near-unity inside the fireball, enforcing a soft spatial mask that tracks the irregular, event-by-event fireball geometry without explicit boundary annotations.

4.4. Training Objective

We train with a composite spatiotemporal loss that penalizes both pointwise field errors and spatial gradient discrepancies, averaged over all output timesteps:

$$\mathcal{L} = \frac{1}{T}\sum_{t=1}^{T}\Big[\mathcal{L}_{L^2}(\hat{\mathbf{f}}_t, \mathbf{f}_t) + \lambda\,\mathcal{L}_{L^2}(\nabla\hat{\mathbf{f}}_t,\, \nabla\mathbf{f}_t)\Big], \tag{9}$$

where the relative L^2 loss is defined as $\mathcal{L}_{L^2}(\hat{u}, u) = \frac{\|\hat{u}-u\|_{L^2}}{\|u\|_{L^2}}$. The first term penalizes pointwise field errors, while the second penalizes discrepancies in the spatial gradients of the predicted fields. $\lambda = \frac{1}{2}$ controls the relative contribution of the gradient penalty. The losses are computed over the spatial domain $\mathbf{S}$ per channel and timestep, with $\mathbf{S} = H \times W$ for 2+1D and $\mathbf{S} = D \times H \times W$ for 3+1D. The gradient penalty term is motivated by Sobolev training (Son et al., 2021) and serves two complementary roles: it sharpens predicted field boundaries by suppressing spurious oscillations near the fireball edge, and it preserves spectral fidelity at high frequencies where the pointwise L^2 loss alone is insufficient, complementing the implicit boundary regularization provided by the GSM module.

4.5. Proper Time Scaling.

The energy density and pressure fields vary by orders of magnitude across the hydrodynamic evolution, from the hot, dense early phase at $\tau \sim \tau_0$ to the dilute late-time freeze-out. This broad dynamic range makes it difficult for the AdaGN decoder to track the dissipative energy scale from a static latent representation. To mitigate this, we apply a proper-time rescaling to the energy density and pressure channels: $\varepsilon(\tau) \leftarrow \tau \cdot \varepsilon(\tau)$, $P(\tau) \leftarrow \tau \cdot P(\tau)$, $\tau = \tau_0 + (t-1)\Delta\tau$, which compresses the dynamic range of

Table 2. Comparison of FNO and TC-QGP across all test datasets. Mean relative error (MRE, %) is averaged over test events and all timesteps. Computational cost is measured on a single NVIDIA A100 GPU (batch size 1). The 2+1D ID/OOD splits share a single trained model, hence identical computational cost.

	2+1D ID		2+1D OOD-Sharp		2+1D OOD-Smooth		3+1D	
	FNO	TC-QGP	FNO	TC-QGP	FNO	TC-QGP	FNO	TC-QGP
Mean Relative Error (%)								
ε	0.59	0.59	7.61	**7.51**	6.51	**6.47**	17.44	**4.08**
T	1.25	**0.79**	6.40	**5.58**	5.54	**4.98**	13.89	**1.66**
P	1.40	**0.58**	8.76	**8.41**	7.36	**7.20**	20.32	**4.54**
u_x	1.17	**1.12**	11.93	**11.67**	10.28	**10.10**	28.10	**8.52**
u_y	1.09	**1.21**	13.04	**13.00**	**11.30**	11.32	27.73	**7.80**
u_z	–	–	–	–	–	–	31.44	**5.44**
All	1.10	**0.86**	9.55	**9.23**	8.20	**8.01**	23.15	**5.34**
Computational Cost (single A100, batch size 1)								
Parameters	FNO: 151.0M		TC-QGP: **21.4M**				356.6M	382.4M
GPU Memory (MB)	FNO: 1,193		TC-QGP: **127.2**				6,585	**4,459**
Inference Time (ms)	FNO: 34.4		TC-QGP: **29.7**				**108.8**	230.4

the training targets and provides the AdaGN decoder with a physically meaningful reference scale at each decoding step, eliminating the need for explicit normalizations. These are applied consistently in both settings: 2+1D (τ_0=0.4 fm/c, $\Delta\tau$=0.16 fm/c) and 3+1D (τ_0=0.4 fm/c, $\Delta\tau$=0.2 fm/c).

5. Experiments

5.1. Experimental Setup

Datasets. We use hydrodynamic simulation data generated with the IP-Glasma + MUSIC framework for Au+Au collisions at $\sqrt{s_{\rm NN}} = 200$ GeV. The 2+1D dataset contains 612 events on a 100×100 grid over 60 timesteps ($\tau \in [0.4, 10.0]$ fm/c, $\Delta\tau$=0.16 fm/c), with five fields $[\epsilon, T, P, u_x, u_y]$; 552 events are used for training. The training events are produced from IP-Glasma initial conditions with a typical hotspot scale of ~ 0.2 fm. The 3+1D dataset contains 250 events on a $28\times108\times108$ grid over 40 timesteps ($\tau \in [0.4, 8.4]$ fm/c, $\Delta\tau$=0.20 fm/c), with six fields $[\epsilon, T, P, u_x, u_y, u_z]$; 225 are used for training and 25 are held out for testing.

Test sets and out-of-distribution (OOD) evaluation. For 2+1D, we evaluate on three disjoint test sets:

- **ID** (in-distribution): 60 events held out from the same hotspot-scale distribution as the training set.
- **OOD-Sharp**: 20 events generated with a finer hotspot scale, yielding initial conditions with sharper than the training distribution.
- **OOD-Smooth**: 20 events generated with a coarser hotspot scale, yielding initial conditions with smoother than the training distribution.

The two OOD sets share the same domain, time grid, and fields as the ID set, and differ only in the initial-state hotspot scale. They allow us to measure how the surrogate behaves when the initial-state spectrum is shifted away from the training distribution in either direction.

5.2. Quantitative Results

2+1D in-distribution accuracy. On the 2+1D ID test set, both models reach sub- to near-percent accuracy, but TC-QGP improves on FNO across every scalar field. The overall MRE drops from 1.10% to 0.86% ($\sim 22\%$ relative reduction), driven by large gains on temperature ($1.25 \to 0.79\%$) and pressure ($1.40 \to 0.58\%$, a 59% reduction). Energy density is already sub-percent under FNO (0.59%) and is matched by TC-QGP, while the most strongly time-dependent fields—temperature and pressure—improve the most. This pattern is consistent with the AdaGN decoder operating on the proper-time-rescaled latent representation, indicating that the time-conditioning mechanism, rather than raw spectral capacity, is the source of the improvement. The transverse flow components are nearly tied between models, as the small-amplitude flow field is dominated by smooth low-k structure that both models capture well.

2+1D out-of-distribution generalization. The OOD-Sharp/Smooth splits probe the spectral bias of the surrogate by shifting the initial-state spectrum to higher or lower wavenumbers. Both models degrade by nearly an order of magnitude under either shift, confirming that the FFT-based spectral kernel they share is the dominant OOD bottleneck rather than the temporal architecture: from $\sim 1\%$ ID error to 9.55%/9.23% MRE on OOD-Sharp, and to 8.20%/8.01% on OOD-Smooth (FNO/TC-QGP). TC-QGP nonetheless retains a consistent edge across nearly every field on both OOD splits, with the largest absolute improvement again on temperature ($6.40 \to 5.58\%$ for OOD-Sharp; $5.54 \to 4.98\%$ for OOD-Smooth), suggesting that AdaGN time-conditioning continues to correct for amplitude drift even when the spatial spectrum is shifted. Per-timestep field, error, and spectral comparisons that visualize how this degradation manifests in space and frequency are provided in Appendix B. Closing the remaining gap to ID accuracy

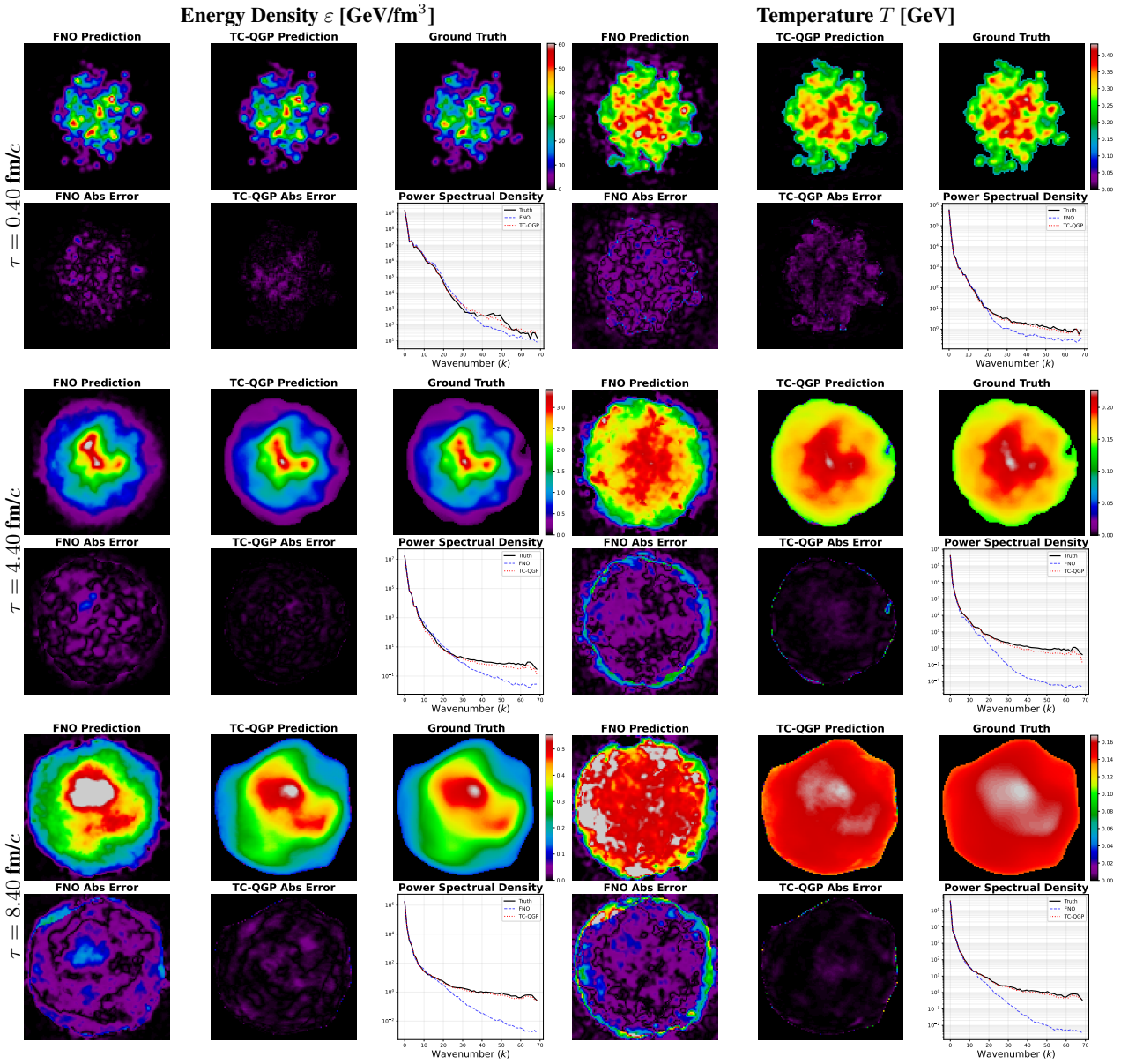

Figure 2. **3+1D QGP hydrodynamic field predictions at three proper times.** Each of the six panels contains two rows of sub-plots: the top row shows (from left to right) the FNO prediction, TC-QGP prediction, and ground truth at a fixed rapidity slice (η_s slice index 14 of 28, corresponding to the mid-rapidity plane $\eta_s \approx 0$); the bottom row shows the FNO absolute error, TC-QGP absolute error, and the power spectral density of each prediction. The two columns of the figure correspond to energy density ε [GeV/fm^3] and temperature T [GeV]; the three rows correspond to proper times $\tau = 0.40$, 4.40, and 8.40 fm/c. The FNO baseline (blue dashed) consistently underestimates high-k spectral power and develops large boundary artifacts at late times, while TC-QGP (red dotted) reproduces the ground truth spectrum and spatial structure throughout the evolution.

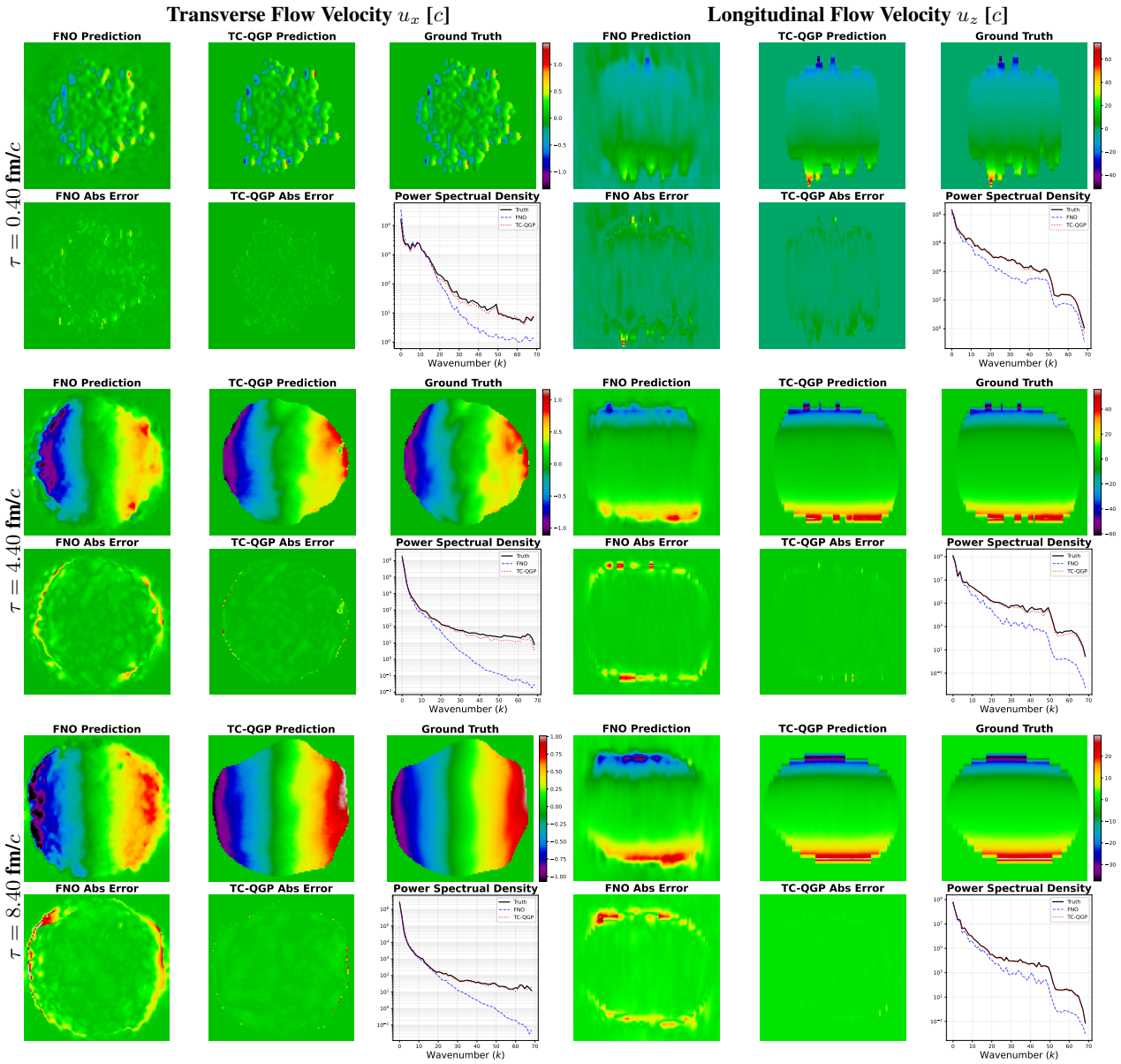

Figure 3. **3+1D QGP hydrodynamic velocity predictions at three proper times.** Layout matches Figure 2. The two columns correspond to the transverse flow velocity u_x [c], visualized on the x-y plane at rapidity slice index 14 of 28 ($\eta_s \approx 0$), and the longitudinal flow velocity u_z [c], visualized on the x-η_s plane at y index 54 of 108 (mid-y); the three rows correspond to proper times $\tau = 0.40$, 4.40, and 8.40 fm/c. The transverse cut for u_x exposes the event-by-event dipole pattern of collective flow, while the rapidity cut for u_z exposes the antisymmetric (rapidity-odd) longitudinal flow profile that is unique to the 3+1D setting and absent from the 2+1D boost-invariant case. The FNO baseline (blue dashed in the PSD panels) underestimates high-k spectral power and develops boundary artifacts that contaminate the vacuum region with spurious flow at late times, while TC-QGP (red dotted) tracks the ground truth spectrum and spatial structure throughout the evolution.

will likely require augmenting the training distribution with mixed hotspot scales, which we leave to future work.

3+1D accuracy. Scaling to the much harder 3+1D problem widens the gap dramatically. The FNO baseline degrades to 23.15% overall MRE—reflecting the combined difficulty of the additional rapidity dimension, the longer multi-step rollout, and the coupling between dissipative dynamics and sharp evolving boundaries. TC-QGP reduces the overall MRE to 5.34%, a 4.3× improvement. The gains are uniform and large on every channel: temperature drops from 13.89% to 1.66% (8.4×), energy density from 17.44% to 4.08%, pressure from 20.32% to 4.54%, and the longitudinal flow u_z—absent in 2+1D and dominated by rapidity-odd structure—from 31.44% to 5.44%. These improvements demonstrate that TC-QGP's three architectural components address complementary failure modes of the FNO baseline: AdaGN handles dissipative amplitude drift, GSM enforces the evolving compact support, and the Tucker-factorized backbone controls the parameter budget at 3D resolution.

Efficiency. TC-QGP is also strictly more efficient than the FNO baseline at both scales. At 2+1D, it uses 21.4M parameters versus 151.0M for FNO (7× smaller), occupies 127.2 MB of peak GPU memory versus 1,193 MB ($\sim 9\times$ less), and runs slightly faster per event (29.7 vs 34.4 ms). At 3+1D, the parameter counts are comparable (382.4M vs 356.6M), but TC-QGP's peak GPU memory is reduced from 6,585 to 4,459 MB (1.5× less); its inference time of 230.4 ms is $\sim 2\times$ slower than FNO's 108.8 ms, the cost paid for T=40 time-conditioned decoder evaluations. The savings stem from the Tucker-factorized spectral weights and the single-pass time-conditioned decoder, which avoid the per-timestep duplication implicit in autoregressive rollouts. Compared to the MUSIC reference solver, which requires several minutes per event in 2+1D and up to 20 hours in 3+1D on a single CPU core, TC-QGP's inference corresponds to a 10^4–$10^5\times$ wall-clock speedup, reducing the cost of a Bayesian inference campaign over millions of events from CPU-decades to a few GPU-hours.

5.3. Qualitative Results and Ablation

For the 2+1D problem, both FNO and TC-QGP already operate at percent-level accuracy on the ID split and remain in the single-digit MRE regime under both OOD shifts; the qualitative differences are subtle but systematic and are deferred to Appendix B and Figures 5, 6, 7. For the much harder 3+1D problem, the gap is visually obvious. Figures 2 and 3 compare FNO and TC-QGP across all six fields at three proper times on a mid-rapidity slice, showing that FNO develops a prominent ring of boundary error and a high-k PSD deficit growing with τ, while TC-QGP tracks the ground truth in both space and frequency. Figure 4 shows volume renderings of the predicted temperature field at τ=2.00 fm/c. The FNO prediction is surrounded by a diffuse cloud of spurious nonzero activations in the vacuum—a direct consequence of spectral operations leaking energy across the domain boundary. Ablating GSM partially reduces the haze but still produces a slightly diffuse boundary, while full TC-QGP reproduces a compact geometry and sharp boundary in close agreement with the ground truth. Further discussions and predictions on different rapidities are provided in Appendix C and Figures 8, 9.

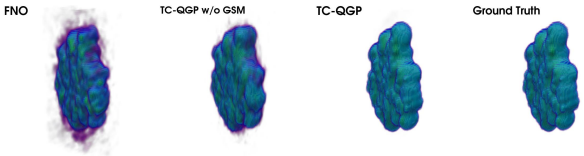

Figure 4. **3D rendering of temperature T at τ=2.00 fm/c.** From left to right: FNO, TC-QGP w/o GSM, TC-QGP, and ground truth. FNO exhibits spurious low-density activations (purple haze) in vacuum; TC-QGP w/o GSM reduces but does not eliminate the haze; full TC-QGP reproduces a compact, sharp-boundary fireball matching the ground truth.

Table 3. Ablation: effect of the GSM module on 3+1D data. MRE (%) averaged over test events and all 40 timesteps.

Field	TC-QGP w/o GSM	TC-QGP
ε	21.32	**4.08**
T	8.14	**1.66**
P	24.06	**4.54**
u_x	25.22	**8.52**
u_y	23.40	**7.80**
u_z	23.30	**5.44**
All	20.91	**5.34**

GSM ablation. Table 3 quantifies the contribution of GSM. Without GSM, the overall MRE on 3+1D rises to 20.91%, comparable to the FNO baseline, indicating that AdaGN time-conditioning alone is insufficient to model the evolving QGP profile near vacuum regions. GSM reduces MRE consistently by a factor of $\sim$ 3–5× across all fields, with the largest gains on energy density and pressure. Velocity fields also benefit substantially, suggesting that gated spatial masking also stabilizes interior flow predictions.

6. Conclusion

We have presented TC-QGP, a time-conditioned neural operator surrogate for fast simulation of QGP hydrodynamic evolution in both 2+1D and 3+1D. By combining a Tucker-factorized spatial encoder, an adaptive time-conditioned decoder, and a gated spatial masking module, TC-QGP accurately reproduces all five (2+1D) or six (3+1D) hydrodynamic field components at sub-percent to percent-level accuracy while achieving orders-of-magnitude wall-clock speedup over the MUSIC PDE solver. These results open a practical path toward large-scale event-by-event inference and multi-observable global analyses of heavy-ion collision data, tasks that were previously computationally infeasible.

References

Bernhard, J. E., Moreland, J. S., and Bass, S. A. Bayesian estimation of the specific shear and bulk viscosity of quark-gluon plasma. *Nature Physics*, 15(11):1113–1117, 2019.

Cao, S. Choose a transformer: Fourier or galerkin. *Advances in neural information processing systems*, 34: 24924–24940, 2021.

Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Dhariwal, P. and Nichol, A. Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.

Gyulassy, M. and McLerran, L. The new forms of qcd matter discovered at rhic. *Nuclear Physics A*, 750(1): 30–63, 2005.

Hao, Z., Wang, Z., Su, H., Ying, C., Dong, Y., Liu, S., Cheng, Z., Song, J., and Zhu, J. Gnot: A general neural operator transformer for operator learning. In *International conference on machine learning*, pp. 12556–12569. PMLR, 2023.

Hendrycks, D. and Gimpel, K. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

Kossaifi, J., Kovachki, N., Azizzadenesheli, K., and Anandkumar, A. Multi-grid tensorized Fourier neural operator for high-resolution PDEs. *arXiv preprint arXiv:2310.00902*, 2023.

Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A. Neural operator: Learning maps between function spaces. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Lee, S. Mesh-independent operator learning for partial differential equations. In *ICML 2022 2nd AI for Science Workshop*, 2022.

Lee, S. and Oh, T. Inducing point operator transformer: A flexible and scalable architecture for solving pdes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 153–161, 2024.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. Fourier neural operator for partial differential equations. *arXiv preprint arXiv:2010.08895*, 2021.

Li, Z., Meidani, K., and Farimani, A. B. Transformer for partial differential equations' operator learning. *arXiv preprint arXiv:2205.13671*, 2022.

Li, Z., Huang, D. Z., Liu, B., and Anandkumar, A. Fourier neural operator with learned deformations for PDEs on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023a.

Li, Z., Kovachki, N., Choy, C., Li, B., Kossaifi, J., Otta, S., Nabian, M. A., Stadler, M., Hundt, C., Azizzadenesheli, K., et al. Geometry-informed neural operator for large-scale 3d pdes. *Advances in Neural Information Processing Systems*, 36:35836–35854, 2023b.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2019.

Nijs, G., van der Schee, W., Gürsoy, U., and Snellings, R. Bayesian analysis of heavy-ion collisions with the heavy-ion computational framework trajectum. *Physical Review C*, 103(5):054909, 2021.

Paquet, J.-F., Shen, C., Denicol, G. S., Luzum, M., Schenke, B., Jeon, S., and Gale, C. Production of photons in relativistic heavy-ion collisions. *Physical Review C*, 93 (4):044906, 2016.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., Azizzadenesheli, K., et al. FourCastNet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

Peebles, W. and Xie, S. Scalable diffusion models with transformers. *arXiv preprint arXiv:2212.09748*, 2023.

Perez, E., Strub, F., De Vries, H., Dumoulin, V., and Courville, A. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Putschke, J. H. et al. The JETSCAPE framework. *arXiv preprint arXiv:1903.07706*, 2019.

Schenke, B., Jeon, S., and Gale, C. (3+1)D hydrodynamic simulation of a+a collisions. *Physical Review Letters*, 106(4):042301, 2011.

Schenke, B., Tribedy, P., and Venugopalan, R. Initial state and hydrodynamics in heavy ion collisions. *Physical Review C*, 86(3):034908, 2012.

Shuryak, E. V. Quantum chromodynamics and the theory of superdense matter. *Physics Reports*, 61(2):71–158, 1980.

Son, H., Jang, J. W., Han, W. J., and Hwang, H. J. Sobolev training for physics informed neural networks. *arXiv preprint arXiv:2101.08932*, 2021.

Stewart, D. and Putschke, J. Fast prediction of the hydrodynamic qgp evolution in ultrarelativistic heavy-ion collisions using fourier neural operators. *Physical Review C*, 113(1):014904, 2026.

Tran, A., Mathews, A., Xie, L., and Ong, C. S. Factorized fourier neural operators. *arXiv preprint arXiv:2111.13802*, 2021.

Wang, Z. et al. Cross-field interface-aware neural operators for multiphase flow simulation. *arXiv preprint arXiv:2511.08625*, 2024.

Wen, G., Li, Z., Azizzadenesheli, K., Anandkumar, A., and Benson, S. M. U-FNO: An enhanced Fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163:104180, 2022.

Wu, H., Luo, H., Wang, H., Wang, J., and Long, M. Transolver: A fast transformer solver for pdes on general geometries. In *International Conference on Machine Learning*, pp. 53681–53705. PMLR, 2024.

A. Setup Details

Training. All models are trained with AdamW (Loshchilov & Hutter, 2019) (lr=10^{-3}, weight decay 10^{-4}) for 1000 epochs on NVIDIA A100 GPUs.

Evaluation metric. We report the mean relative error (MRE),

$$\langle \epsilon_{\mathrm{rel}} \rangle = \frac{\sum_x |\hat{f}(x) - f(x)|}{\sum_x |f(x)|}, \tag{10}$$

averaged over test events and timesteps, with the spatial sum extending over the full (D, H, W) domain in 3+1D.

B. 2+1D Qualitative Results

This appendix provides the per-timestep two-dimensional field maps and power-spectral-density comparisons referenced from Section 5.3, supporting the quantitative results reported in Table 2 for the 2+1D ID, OOD-Sharp, and OOD-Smooth test sets. Figure 5 shows the in-distribution case where both models perform near percent-level accuracy, while Figures 6 and 7 probe the surrogates under controlled shifts of the initial-state hotspot scale toward sharper and smoother spectra, respectively.

Early-to-mid evolution ($\tau = 0.80$ fm/c). Soon after the initial state is propagated, the multi-hotspot transverse structure of ε and T is most pronounced, and high-k accuracy matters most. On the ID split, both FNO and TC-QGP reproduce this structure with errors well under one percent and matching power spectra. On OOD-Sharp, the FNO baseline already exhibits a visible spectral deficit at high k and a faint ring of boundary error, while TC-QGP closely tracks the ground truth at all wavenumbers. On OOD-Smooth, where the initial spectrum is depleted at high k, the two models behave similarly, but FNO predictions remain slightly suppressed in amplitude across the fireball.

Mid evolution ($\tau = 4.80$ fm/c). At intermediate proper times, the fireball has expanded and developed a sharper boundary against the vacuum. ID predictions remain visually indistinguishable from the ground truth for both models. On OOD-Sharp, the FNO baseline develops a pronounced ring of boundary error in ε and T as spectral operations smear activations into the vacuum, with its power spectrum dropping clearly below the ground truth at large k; TC-QGP keeps the boundary localized and the spectrum aligned. On OOD-Smooth, the differences are smaller but follow the same trend: FNO underestimates fireball amplitude while TC-QGP preserves it.

Late evolution ($\tau = 9.60$ fm/c). By the final timestep, the fireball has substantially cooled, and most spatial support is concentrated in a narrow region, making the boundary-handling capability of the surrogate critical. On the ID split, both models continue to produce essentially exact predictions. On the two OOD splits, TC-QGP retains a clean fireball boundary with errors confined to a thin band where GSM is most active, whereas the FNO predictions show diffuse residuals and a power spectrum suppressed by up to an order of magnitude at the highest wavenumbers. These observations are consistent with the MRE gaps reported in Table 2 and reinforce the picture that the AdaGN time-conditioning and the gated spatial masking remain beneficial under spectral distribution shifts in 2+1D.

C. 3+1D Qualitative Results

This appendix expands the per-timestep analysis in Section 5.3 into a finer-grained, per-time-window discussion of the 3+1D scalar (Figure 2) and velocity (Figure 3) predictions, including channel-by-channel observations.

Early time ($\tau = 0.40$ fm/c). At the earliest timestep, both models face the most challenging regime: the IP-Glasma initial state contains fine-grained nucleon-scale fluctuations that require accurate reproduction of high-frequency spectral features. TC-QGP captures the multi-hotspot structure of ε and T with errors concentrated inside the fireball interior, and its power spectral density follows the ground truth closely across all wavenumbers. FNO exhibits a spectral deficit of roughly an order of magnitude, yielding visually smoother but physically incorrect predictions. For u_x and u_z, both models reproduce the small-amplitude early-time flow pattern, but TC-QGP achieves noticeably lower absolute errors.

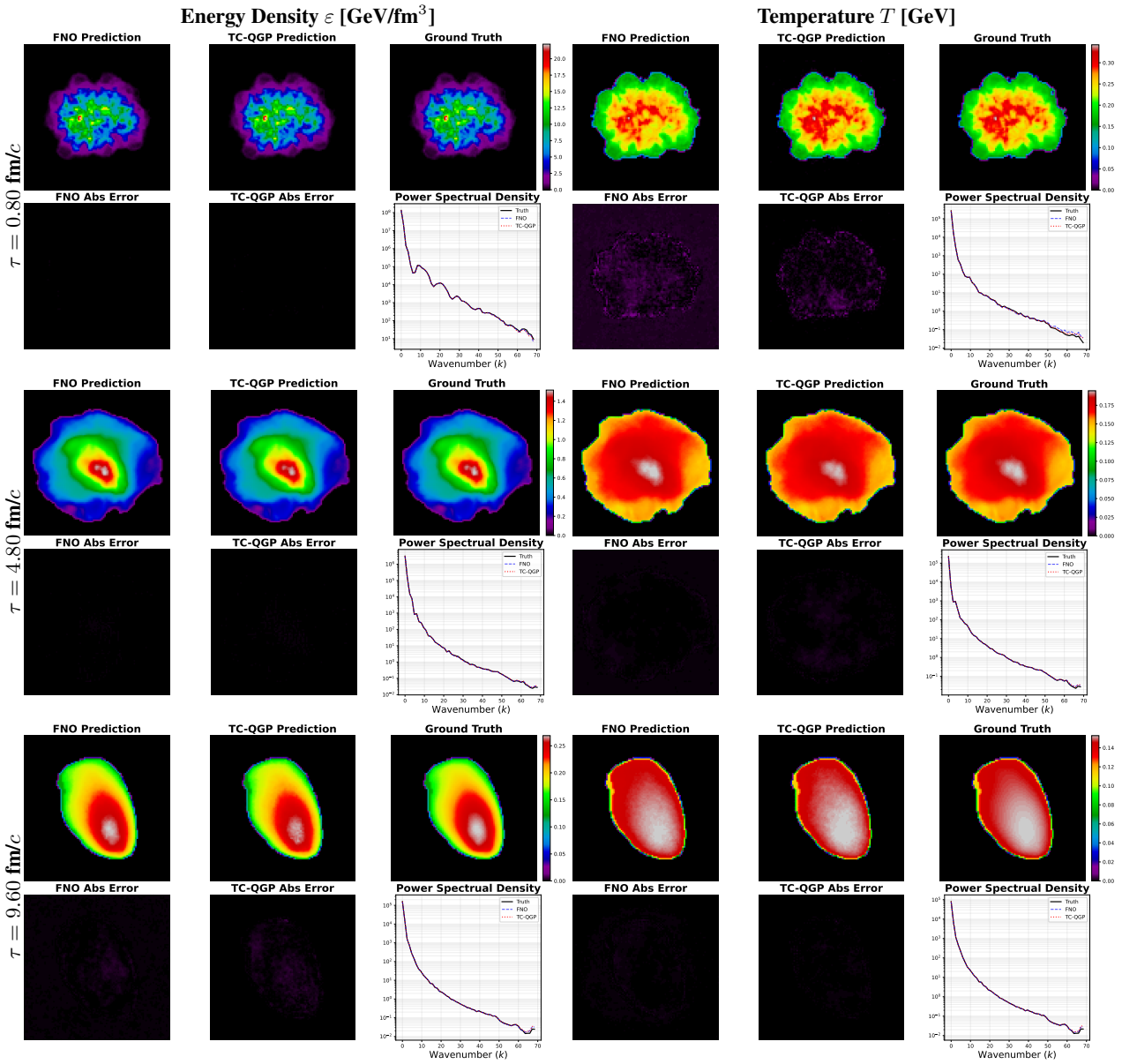

Figure 5. **2+1D QGP hydrodynamic field predictions on the in-distribution (ID) test set at three proper times.** Each of the six panels contains two rows of sub-plots: the top row shows (from left to right) the FNO prediction, TC-QGP prediction, and ground truth; the bottom row shows the FNO absolute error, TC-QGP absolute error, and the power spectral density of each prediction. Columns correspond to energy density ε [GeV/fm^3] and temperature T [GeV]; rows to proper times $\tau = 0.80$, 4.80, and 9.60 fm/c. On the ID split, both models are visually nearly indistinguishable from ground truth, with absolute errors at the percent level across the full evolution and matching power spectra over all wavenumbers, consistent with the sub-percent MRE in Table 2.

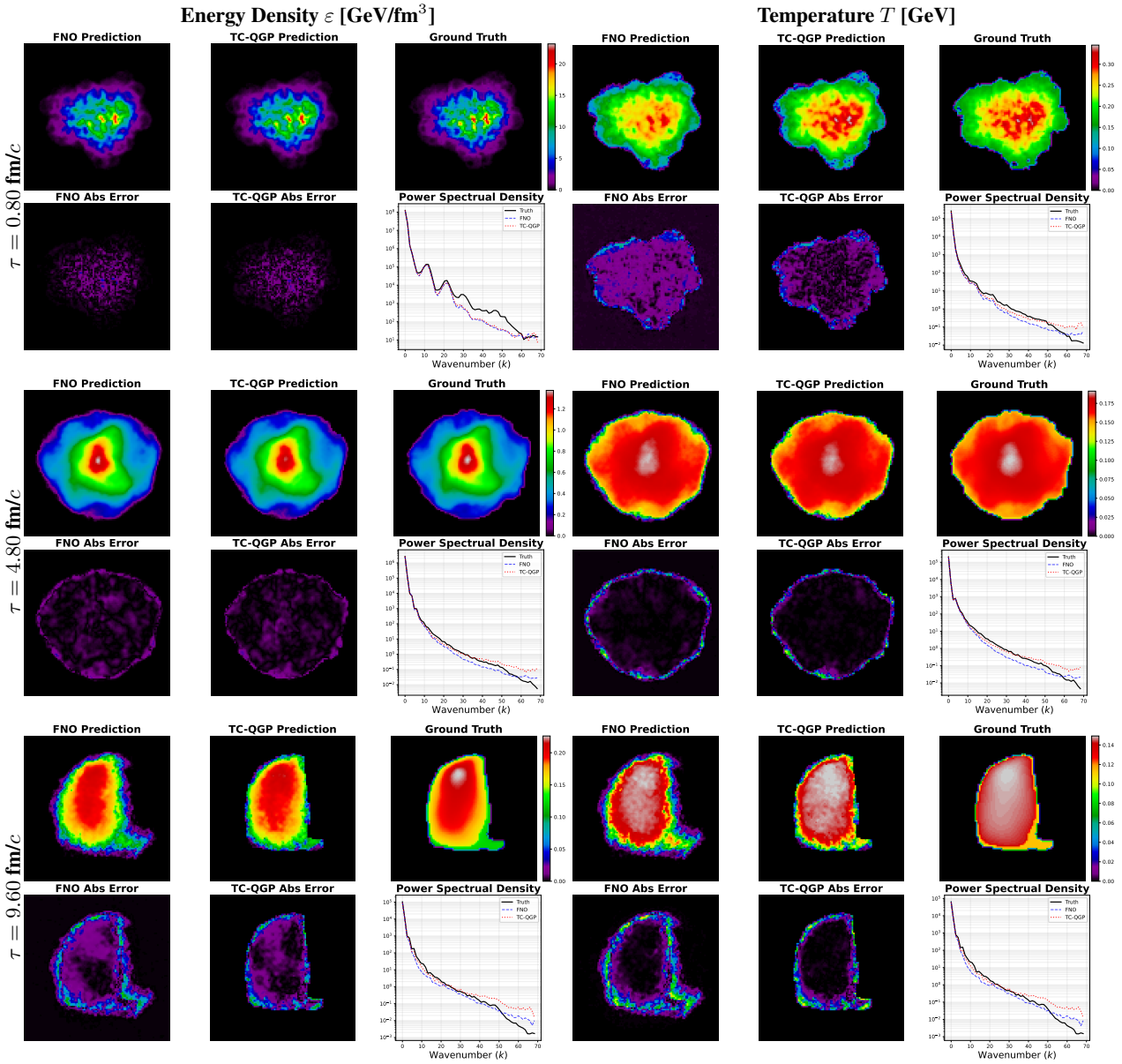

Figure 6. **2+1D QGP hydrodynamic field predictions on the OOD-Sharp test set at three proper times.** Layout is identical to Figure 5: per-panel, the top row shows FNO prediction, TC-QGP prediction, and ground truth; the bottom row shows the two absolute-error maps and the per-prediction power spectral density. Columns correspond to ε and T; rows to $\tau = 0.80$, 4.80, and 9.60 fm/c. Under the spectrally sharper initial condition, the FNO baseline visibly underestimates high-k structure and develops a ring of large boundary errors that grows with τ. TC-QGP retains a sharper fireball boundary and a power spectrum closer to ground truth at large k, in line with its lower MRE on this split (Table 2).

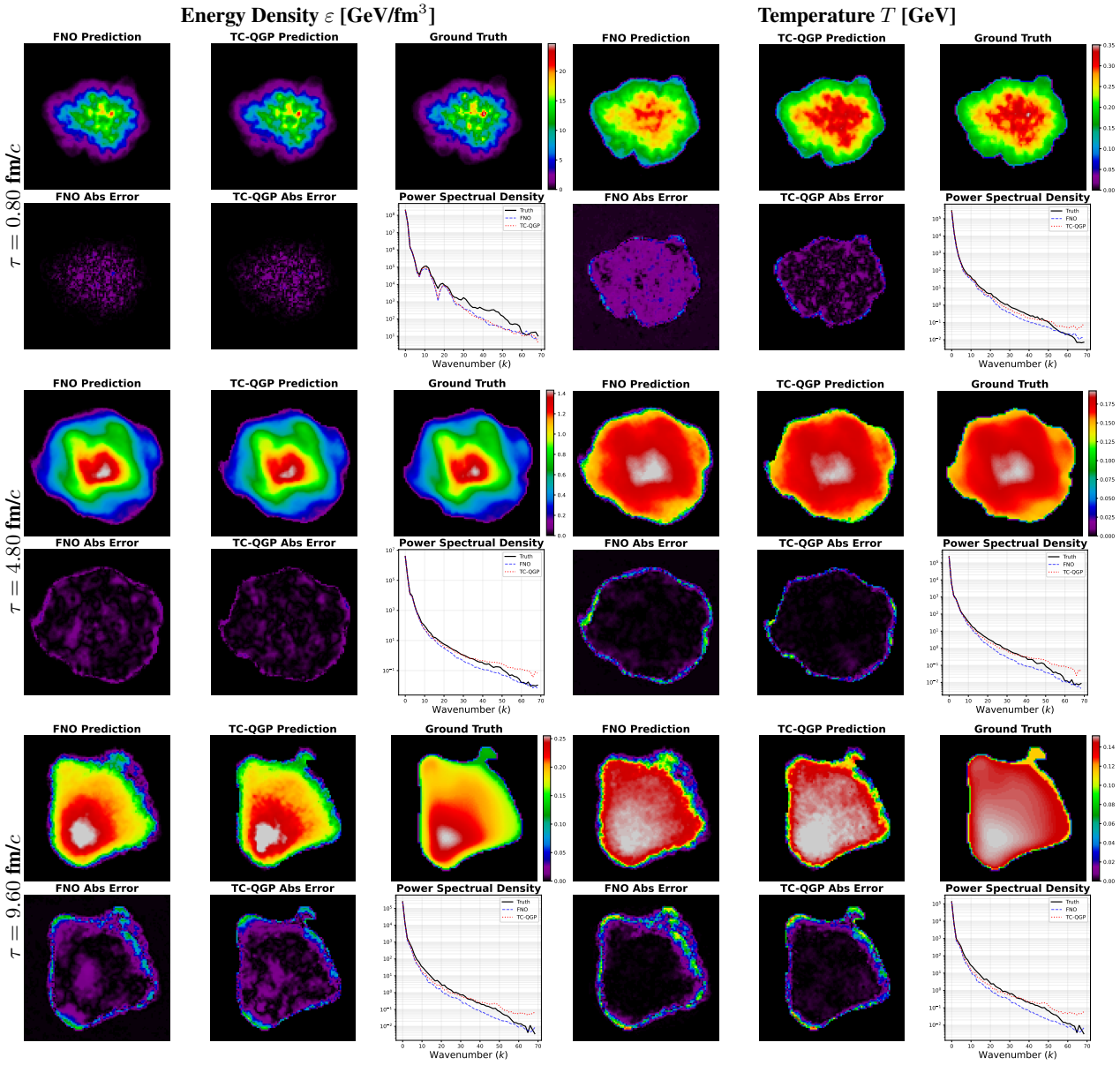

Figure 7. **2+1D QGP hydrodynamic field predictions on the OOD-Smooth test set at three proper times.** Layout is identical to Figure 5: per-panel, the top row shows FNO prediction, TC-QGP prediction, and ground truth; the bottom row shows the two absolute-error maps and the per-prediction power spectral density. Columns correspond to ε and T; rows to $\tau = 0.80$, 4.80, and 9.60 fm/c. Compared with the in-distribution case, the OOD-Smooth initial condition has lower high-k spectral content than the training distribution. Both models retain the smooth fireball geometry but exhibit a uniform amplitude bias inside the fireball; the FNO power spectrum is suppressed across all wavenumbers, while TC-QGP tracks the ground truth spectrum within a small factor at all k, mirroring the modest MRE gap reported in Table 2.

Mid evolution ($\tau = 4.40$ fm/c). As the fireball expands and develops collective transverse flow, the differences become stark. FNO predictions of ε and T display a prominent ring of large errors at the fireball boundary—a direct consequence of spectral operations smearing activations into the vacuum—while the interior loses spatial structure visible in the ground truth. The u_x field reveals an asymmetric dipole pattern from event-by-event geometry; FNO captures only the large-scale gradient and fails at the boundary, whereas TC-QGP matches both the interior field and the sharp vacuum transition. For u_z, the rapidity-odd flow profile develops a clear antisymmetric gradient along η_s; TC-QGP tracks this accurately, while FNO's spectral leakage contaminates the vacuum region with spurious longitudinal flow. In the power spectral density, FNO falls 2–3 orders of magnitude below ground truth at $k > 10$, while TC-QGP remains within a factor of a few across all modes.

Late time ($\tau = 8.40$ fm/c). The contrast is most dramatic at the final timestep. FNO predictions of ε and T exhibit widespread saturation artifacts—gray (overflowed) regions and unphysical high-amplitude patches throughout the domain—indicating an inability to track the dissipative energy scale across the full evolution. TC-QGP accurately reproduces the smooth fireball profile with errors confined to a narrow band at the boundary where GSM is most active. Its power spectral density is essentially indistinguishable from the ground truth, while FNO is suppressed by 3–4 orders of magnitude. The late-time large-scale velocity gradients with sharp boundaries are well captured by TC-QGP, whereas FNO's boundary artifacts corrupt the velocity field across the entire peripheral region.

Rapidity dependence (Figures 8, 9). The mid-rapidity view in Figure 2 captures the peak transverse fireball structure but understates the difficulty of the 3+1D setting at off-central rapidities. Figures 8 and 9 show the same scalar fields at a backward (η_s slice 8 of 28) and forward (η_s slice 20 of 28) rapidity slice, respectively. Two effects emerge that are absent at $\eta_s \approx 0$. First, the transverse footprint of the fireball is markedly smaller and its boundary sharper at large $|\eta_s|$, amplifying the spectral-leakage failure mode of FNO: the boundary-error ring visible at mid-rapidity becomes a thicker halo of spurious vacuum activations at off-central slices, and the high-k PSD deficit grows. Second, a forward/backward asymmetry is visible event-by-event due to the rapidity-odd longitudinal flow u_z discussed earlier; FNO produces nearly symmetric predictions that wash out this asymmetry, while TC-QGP preserves it across both rapidity directions. Together with the mid-rapidity view in Figure 2, these slices establish that TC-QGP's accuracy gain is uniform along the η_s axis rather than localized to the central plane.

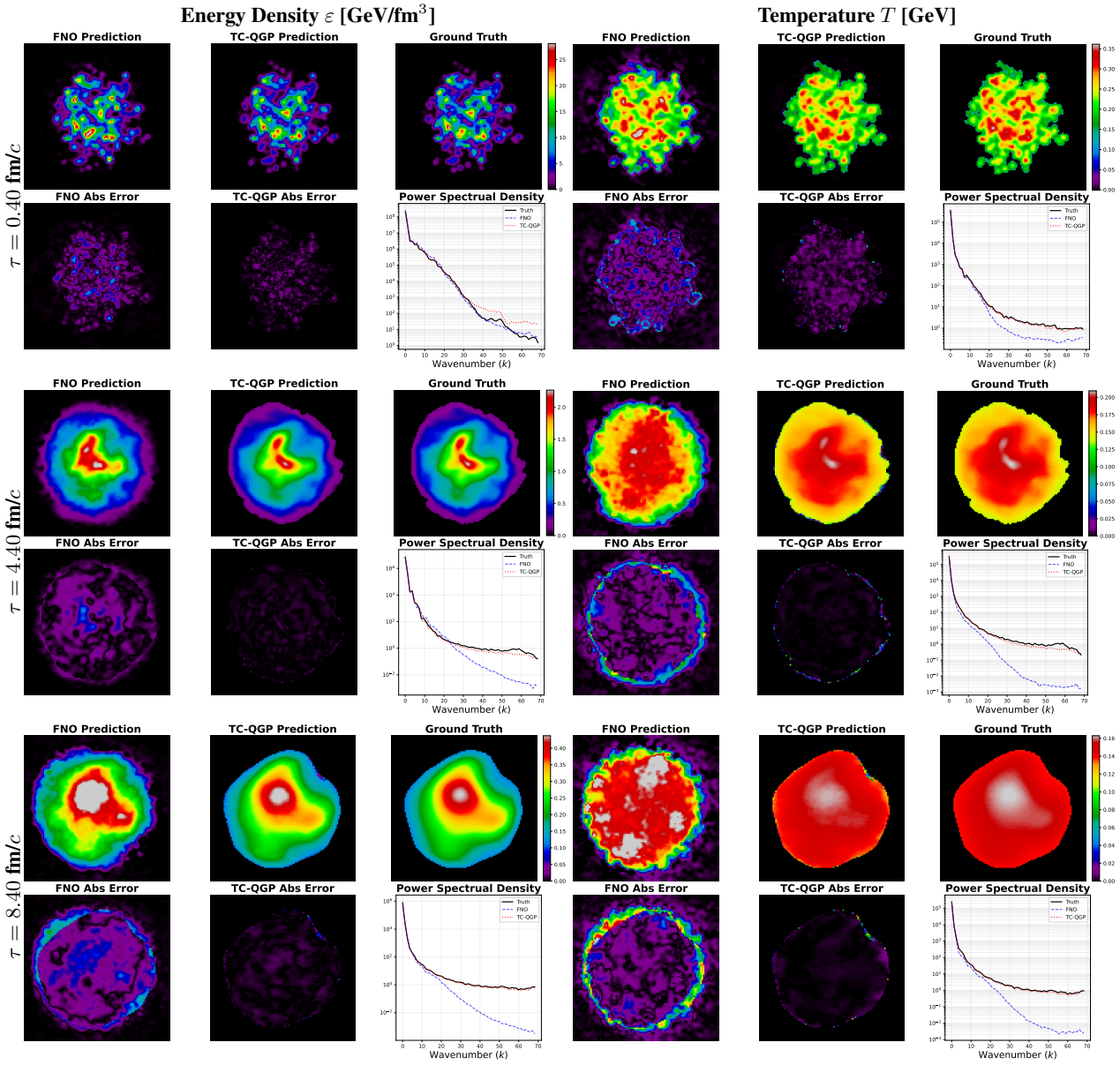

Figure 8. **3+1D QGP hydrodynamic field predictions in a backward-rapidity slice** ($\eta_s < 0$) **at three proper times.** Each of the six panels contains two rows of sub-plots: the top row shows (from left to right) the FNO prediction, TC-QGP prediction, and ground truth at a fixed backward-rapidity slice (η_s slice index 8 of 28); the bottom row shows the FNO absolute error, TC-QGP absolute error, and the power spectral density of each prediction. The two columns of the figure correspond to energy density ε [GeV/fm^3] and temperature T [GeV]; the three rows correspond to proper times $\tau = 0.40$, 4.40, and 8.40 fm/c. The FNO baseline (blue dashed) consistently underestimates high-k spectral power and develops large boundary artifacts at late times, while TC-QGP (red dotted) reproduces the ground truth spectrum and spatial structure throughout the evolution.

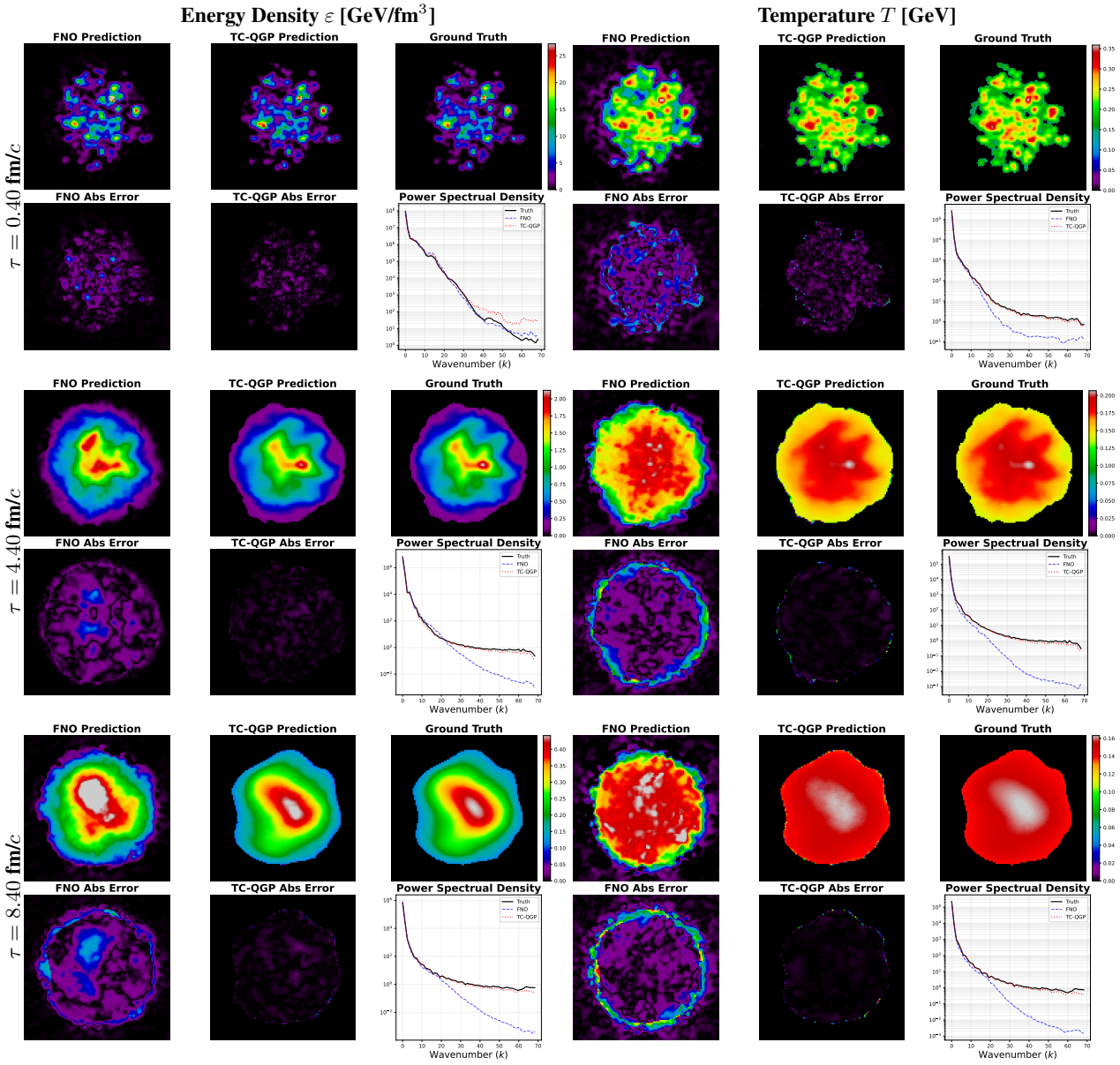

Figure 9. **3+1D QGP hydrodynamic field predictions in a forward-rapidity slice ($\eta_s > 0$) at three proper times.** Each of the six panels contains two rows of sub-plots: the top row shows (from left to right) the FNO prediction, TC-QGP prediction, and ground truth at a fixed forward-rapidity slice (η_s slice index 20 of 28); the bottom row shows the FNO absolute error, TC-QGP absolute error, and the power spectral density of each prediction. The two columns of the figure correspond to energy density ε [GeV/fm^3] and temperature T [GeV]; the three rows correspond to proper times $\tau = 0.40$, 4.40, and 8.40 fm/c. The FNO baseline (blue dashed) consistently underestimates high-k spectral power and develops large boundary artifacts at late times, while TC-QGP (red dotted) reproduces the ground truth spectrum and spatial structure throughout the evolution.