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

Computing and Data Sciences, BNL

Seungjun Lee

In collaboration with Y. Go, B. Schenke, C. Shen, Y. Ren, J. Huang, M. Chamizo-Llatas

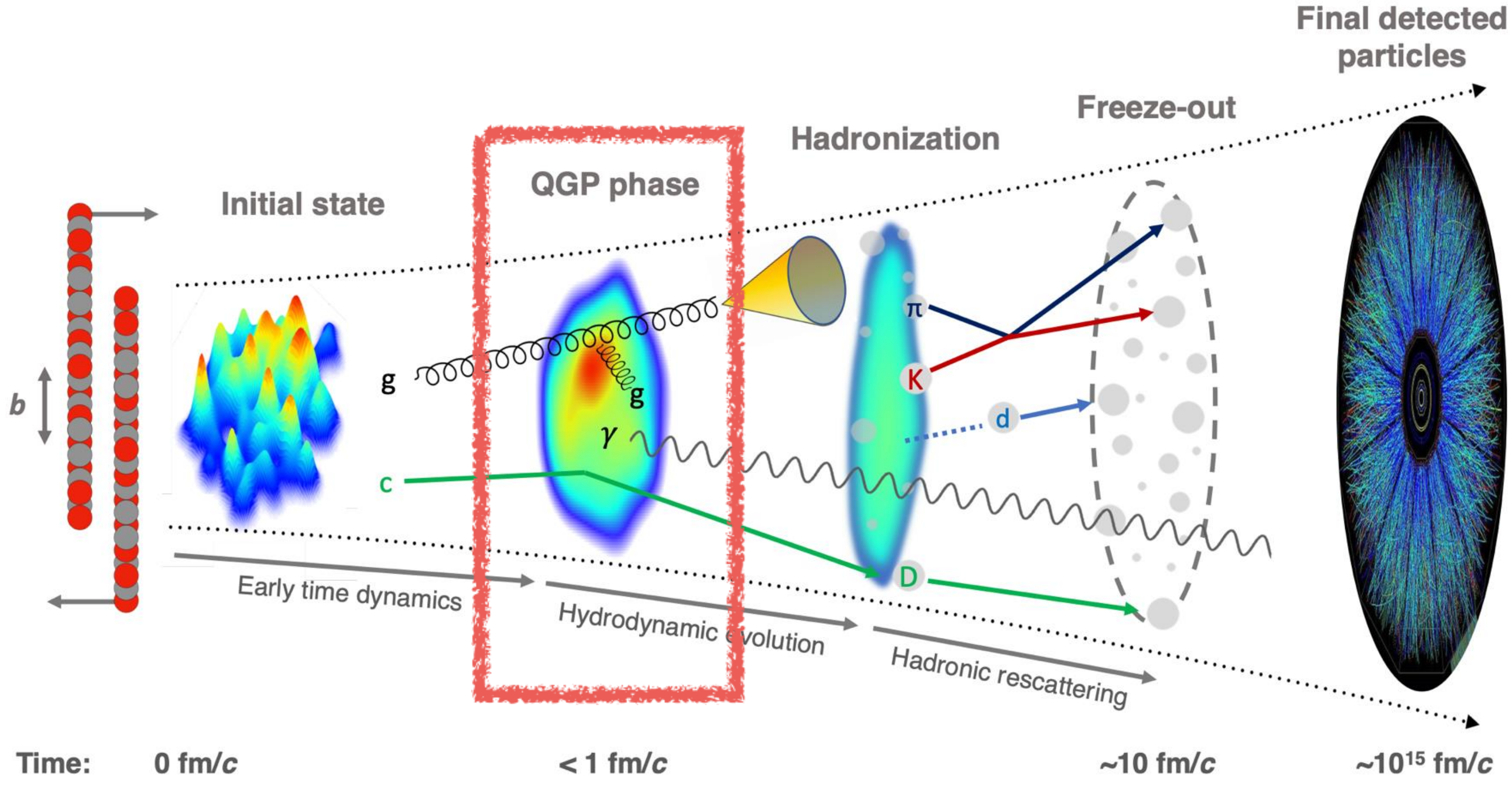
This work is funded under BNL LDRD 26-46

RHIC/AGS Annual User’s Meeting · AI/ML Workshop

05/11/2026

Heavy Ion Collision simulations

M. Arslandok, et. al., Hot QCD White Paper, arxiv:2303.17254, 2023.

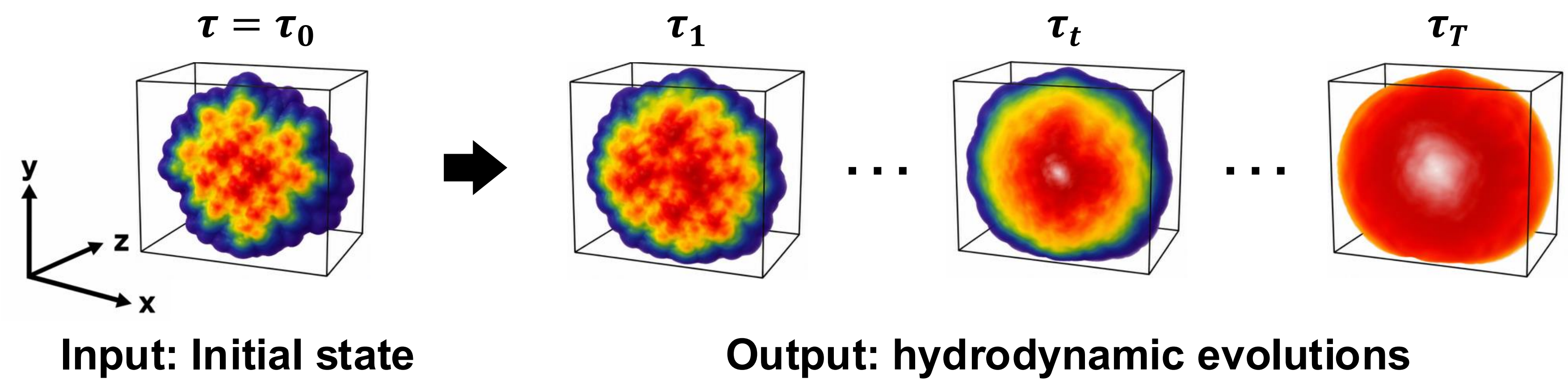


Traditional Solver (IP-Glasma+MUSIC)

	2+1D	3+1D
Time/event	~10 mins	~20 hours

- Many simulated events are required for parameter inference and uncertainty quantification.
- The hydrodynamic evolution of the QGP is a computational bottleneck, motivating fast and accurate **AI surrogate models** that can replace or accelerate the hydrodynamic stage.

Neural Operators for QGP: Promise and Limits



Why Neural Operators?

- Learns **mapping** between function spaces.
- **Fourier neural operator (FNO):** efficiently captures global structures in Fourier space.
- **Resolution-invariant:** learns operators on continuous function spaces, not limited to a fixed grid.

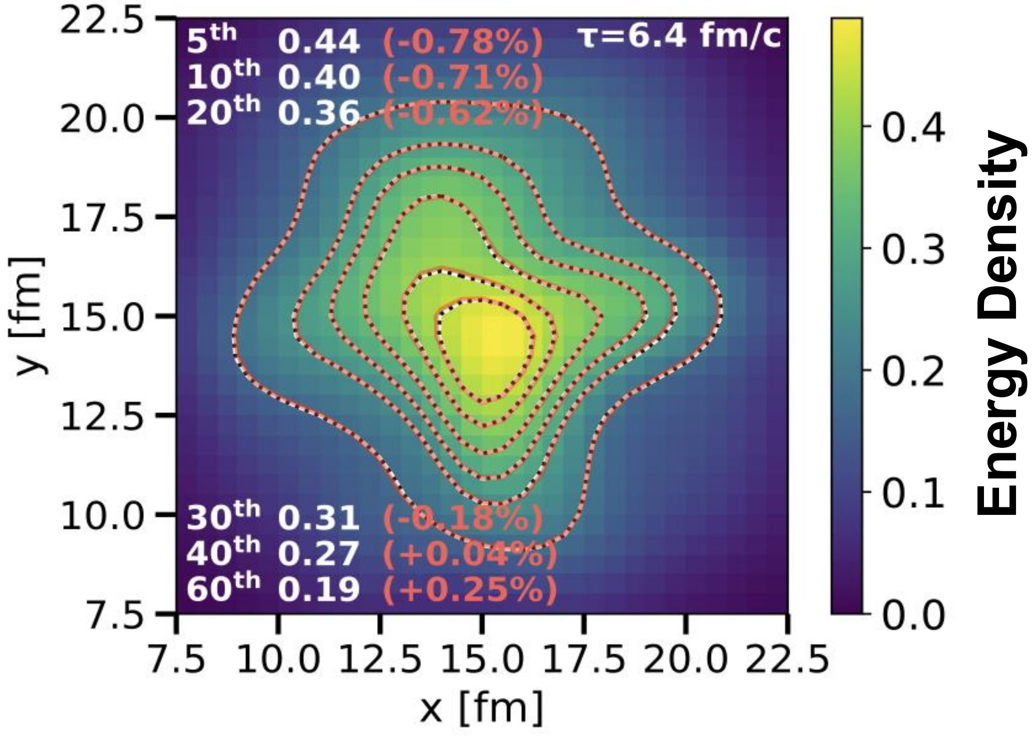
→ Strong baseline for PDE surrogate modeling but **not specific for QGP.**

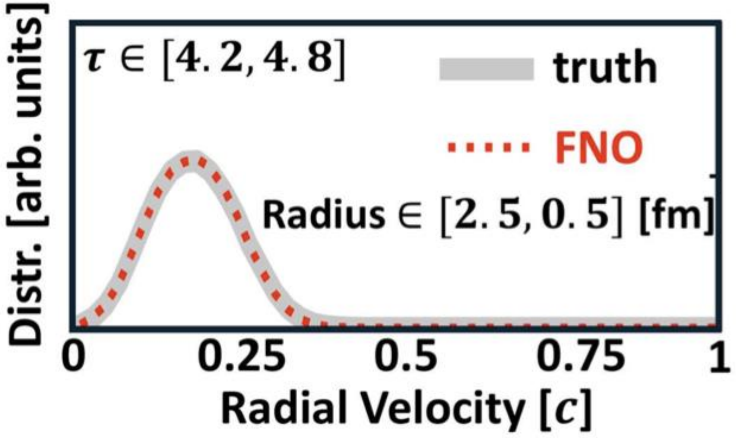
Challenges of Standard FNO for QGP

- **Dissipative dynamics over τ:** the system exhibits strong dissipation over time.
- **Evolving spatial support with sharp boundary:** the hot plasma region shrinks as it evolves, and a sharp boundary separates the fireball from the outside.
- **Long multi-step temporal rollout:** full event evolution requires accurate and stable predictions over many timesteps.

Previous Works

- Earlier ML surrogate work: stacked U-Net acceleration for 2+1D hydrodynamics (PRR 2021).
  - H. Huang, et. al., Physical Reivew Research 3, 023256, 2021.
- Stewart & Putschke (PRC 2026) demonstrated FNOs for fast prediction of 2+1D QGP evolution.
  - Predicted energy density and velocity fields, as well as flow and jet-quenching observables.
  - D. Stewart, and J. Putschke, Physical Review C 113, 014904, 2026.

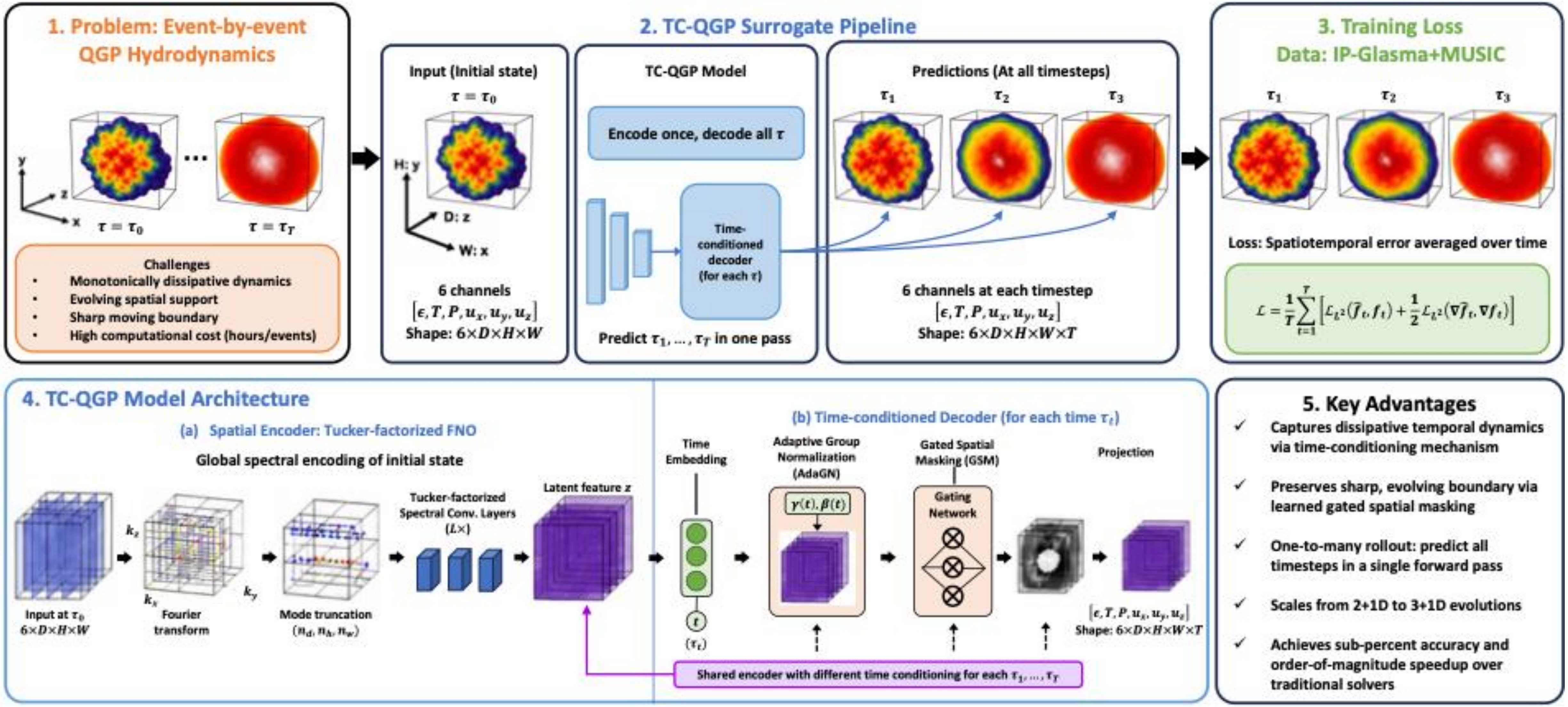




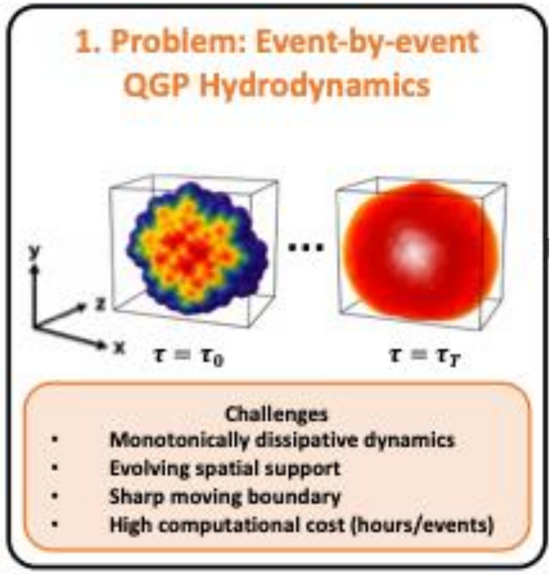
Our work extends this approach by:

- introducing **time-conditioning** and **gated spatial masking**
- predicting **all hydrodynamic fields ($\epsilon, T, P, u_x, u_y, u_z$)**
- scaling to **full 3+1D evolution**

Methods - Overview



Methods – Problem Formulation



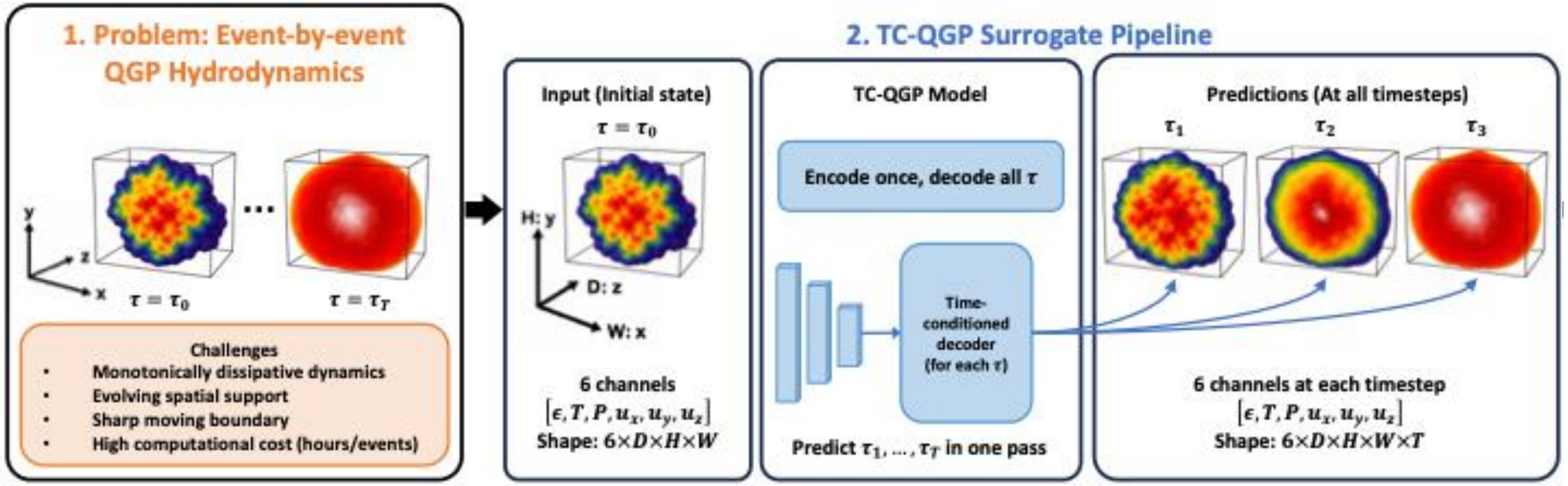
Problem Formulation

- Given an initial state at $\tau = \tau_0$ specifying the field $f_0(x, y, \eta_s) = [\epsilon, T, P, u_x, u_y, u_z]_{\tau_0}$, the MUSIC solver propagates the system forward across τ, producing a sequence $\{f_\tau\}_{\tau=\tau_0}^{\tau_T}$.
- Our goal is to learn a surrogate operator to approximate the mapping,

$$\mathcal{G} : f_0 \mapsto \{\hat{f}_\tau\}_{\tau>\tau_0}$$

- For 2+1D, $f_\tau \in \mathbb{R}^{C\times H\times W} = \mathbb{R}^{5\times100\times100}$ with 60 timesteps, for 3+1D, $f_\tau \in \mathbb{R}^{C\times D\times H\times W} = \mathbb{R}^{6\times28\times128\times128}$ with 40 timesteps.

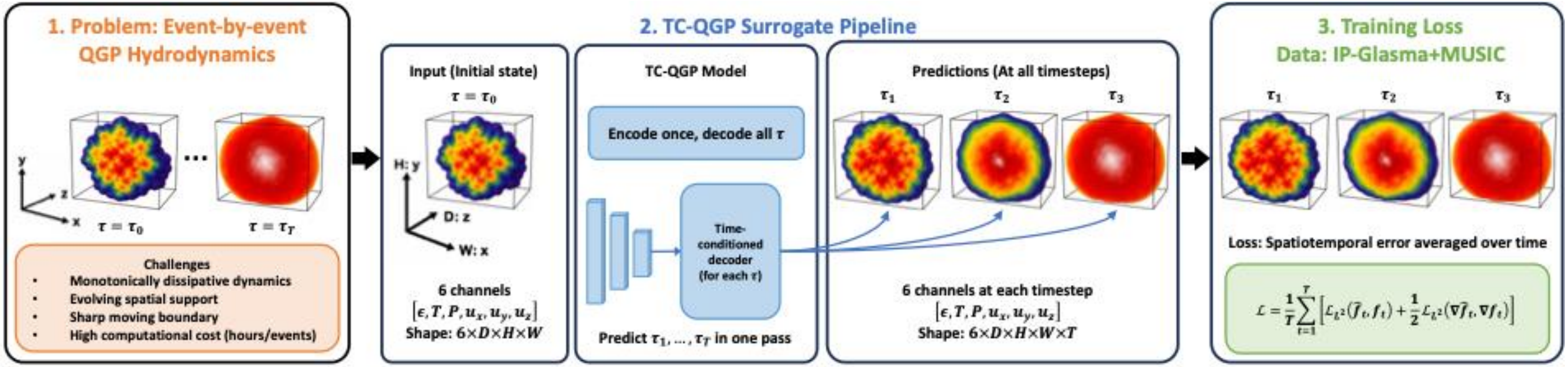
Methods – Surrogate Pipeline



TC-QGP Surrogate Pipeline

- **Factorized Neural Operator Encoder:** using Tucker tensorized FNO to map initial state to reduced latent space
$$z = TFNO(f_0) \in \mathbb{R}^{d_h \times S_z}$$
- **Adaptive Time-Conditioned Decoding:** to capture dissipative dynamics, when $[\gamma_t, \beta_t] = W_{emb} e_t$
  - **Time Embedding:** $e_t = MLP(\tilde{\tau}_t) \in \mathbb{R}^{d_h}$
  - **Adaptive modulation:** $AdaGN(z, e_t) = GN(z) \odot (1 + \gamma_t) + \beta_t,$
- Gated Spatial Masking: to capture evolving field support with time-evolving sharp boundaries
$$GSM(\hat{f}_t) = \hat{f}_t \odot \sigma\left(Conv_1\left(ReLU\left(Conv_3(\hat{f}_t)\right)\right)\right)$$

Methods – Training Objective



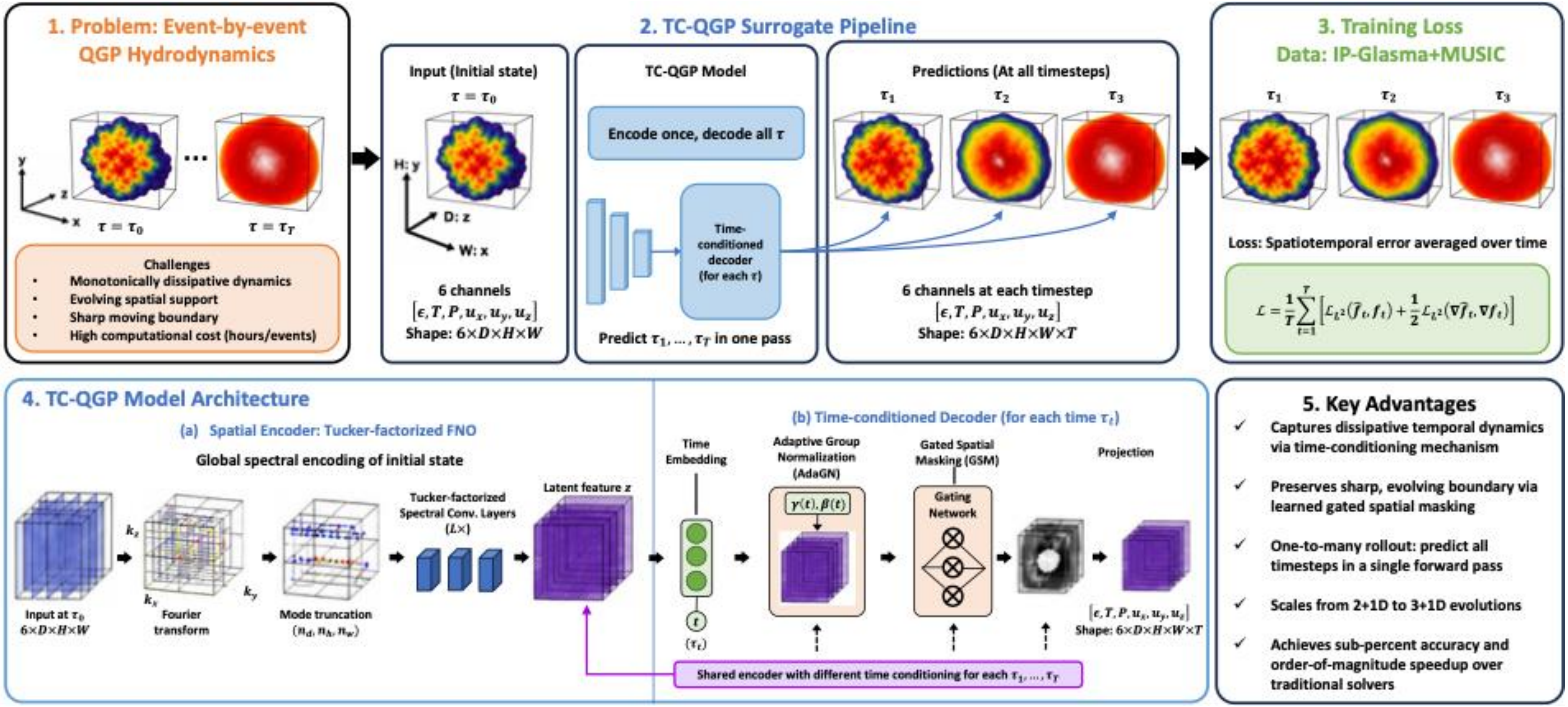
Training Objective

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

$$\mathcal{L} = \frac{1}{T}\sum_{t=1}^{T}\left[\mathcal{L}_{L^2}(\hat{f}_t, f_t) + \lambda\mathcal{L}_{L^2}(\nabla\hat{f}_t, \nabla f_t)\right],$$

- The first term penalizes pointwise field errors, while the second term penalizes discrepancies in the spatial gradients of the predicted fields.

Methods – Key Advantages



Experimental Setups

2+1D Dataset

- IP-Glasma+MUSIC for Au+Au collision at $\sqrt{s_{NN}} = 200GeV$
- 612 events: 100×100 grid $\times$ 60 timesteps
- $\tau \in [0.4, 10]\ fm/c$, $\Delta\tau = 0.16\ fm/c$
- 5 fields: ϵ, T, P, u_x, u_y
- Splits: 552 (train) / 60 (test) / 20 OOD-Sharp, 20 Smooth

3+1D Dataset

- IP-Glasma+MUSIC for Au+Au collision at $\sqrt{s_{NN}} = 200GeV$
- 250 events: $28 \times 128 \times 128$ grid $\times$ 40 timesteps
- $\tau \in [0.4, 8.4]\ fm/c$, $\Delta\tau = 0.2\ fm/c$
- 6 fields: $\epsilon, T, P, u_x, u_y, u_z$
- Splits: 225 (train) / 25 (test)

Out-of-distribution (OOD) Evaluation (2+1D)

In-distribution (ID)

- 60 events from the same hotspot-scale distribution as the training set ($\sim 0.2\ fm$)

OOD-Sharp

- 20 events with a finer hotspot scale, yielding initial conditions with sharper than the training distribution

OOD-Smooth

- 20 events with a coarser hotspot scale, yielding initial conditions with smoother than the training distribution

Evaluation Metric

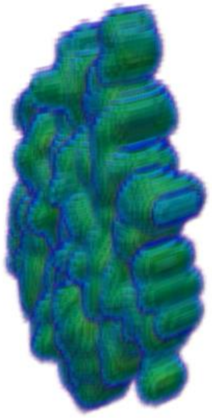
- Mean Relative Error: averaged over test events and timesteps, with the spatial sum

$$\langle \epsilon_{rel} \rangle = \frac{\sum_x \left| \hat{f}(x) - f(x) \right|}{\sum_x |f(x)|}$$

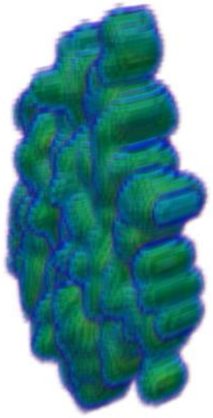
Results – 3+1D Full Evolution (including ablation)

- **Temperature Field**

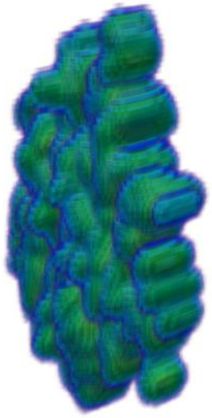
FNO



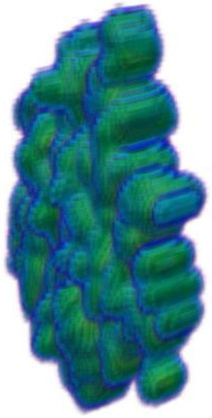
TC-QGP w/o GSM



TC-QGP

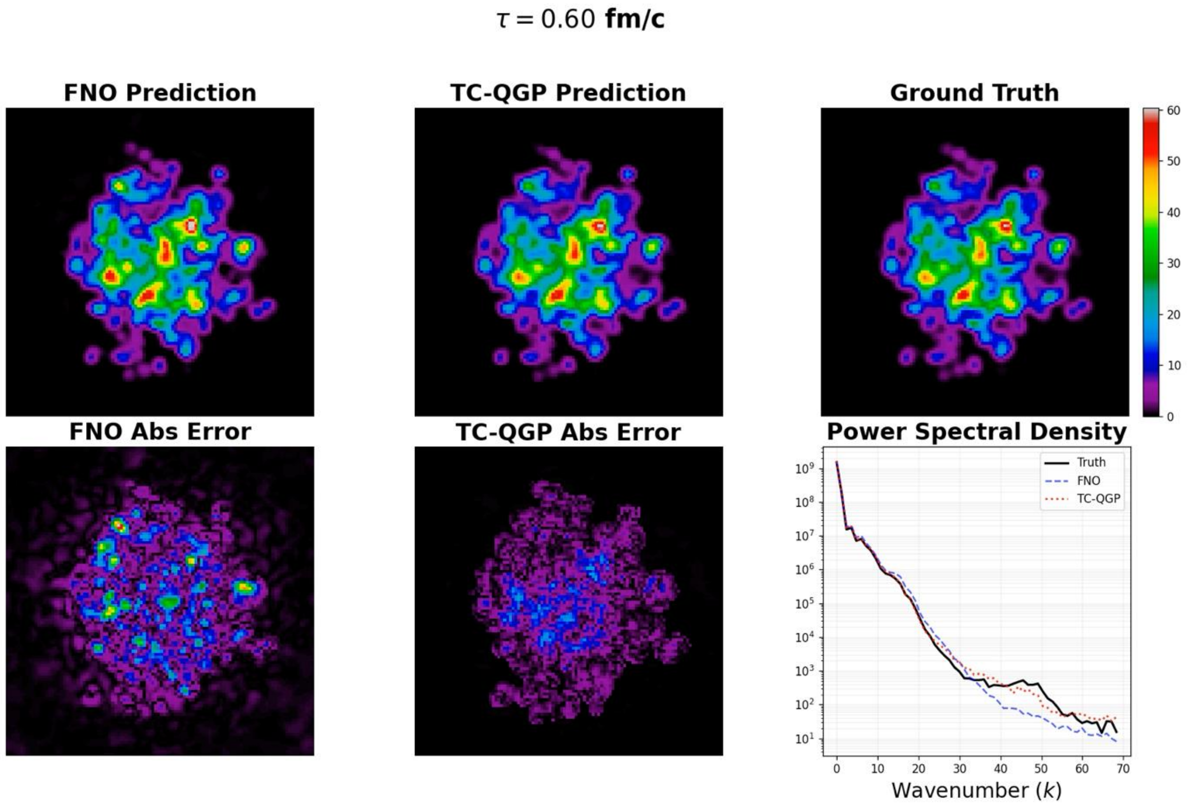


Ground Truth

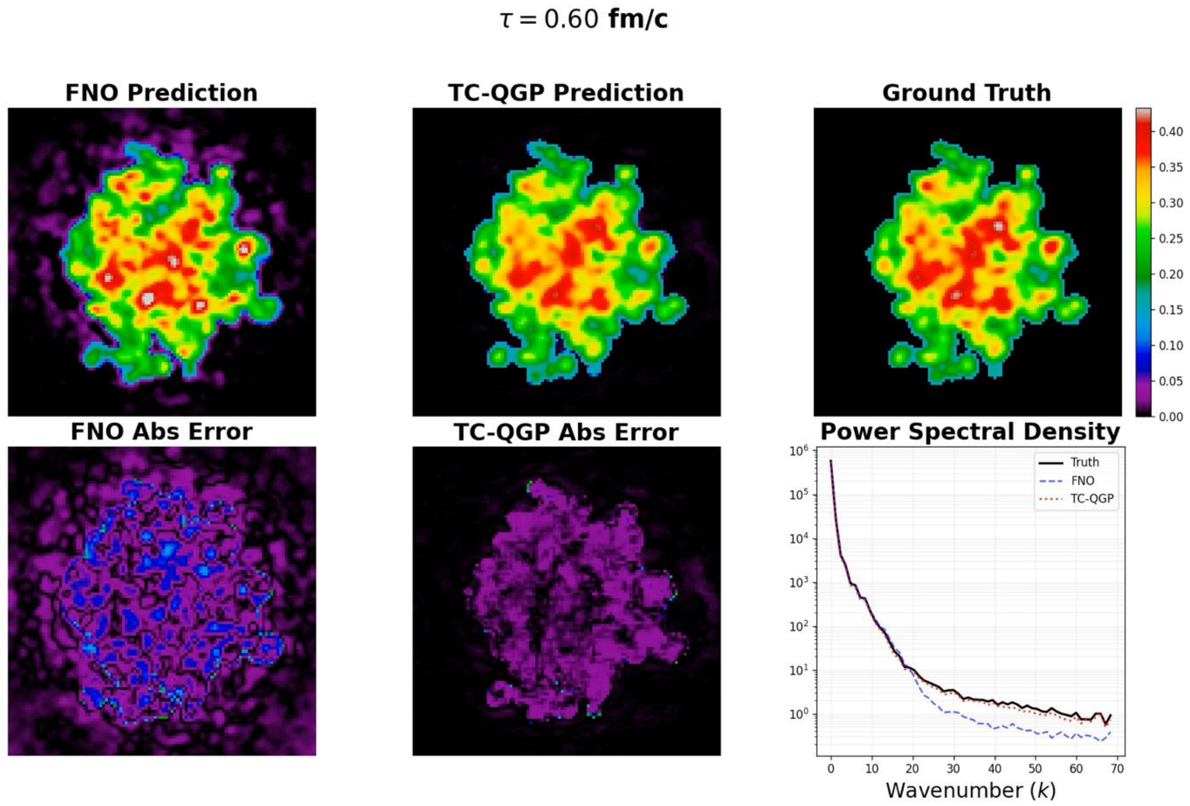


Results – 3+1D Energy Density, Temperature

- **At mid-rapidity plane $\eta_s \approx 0$**



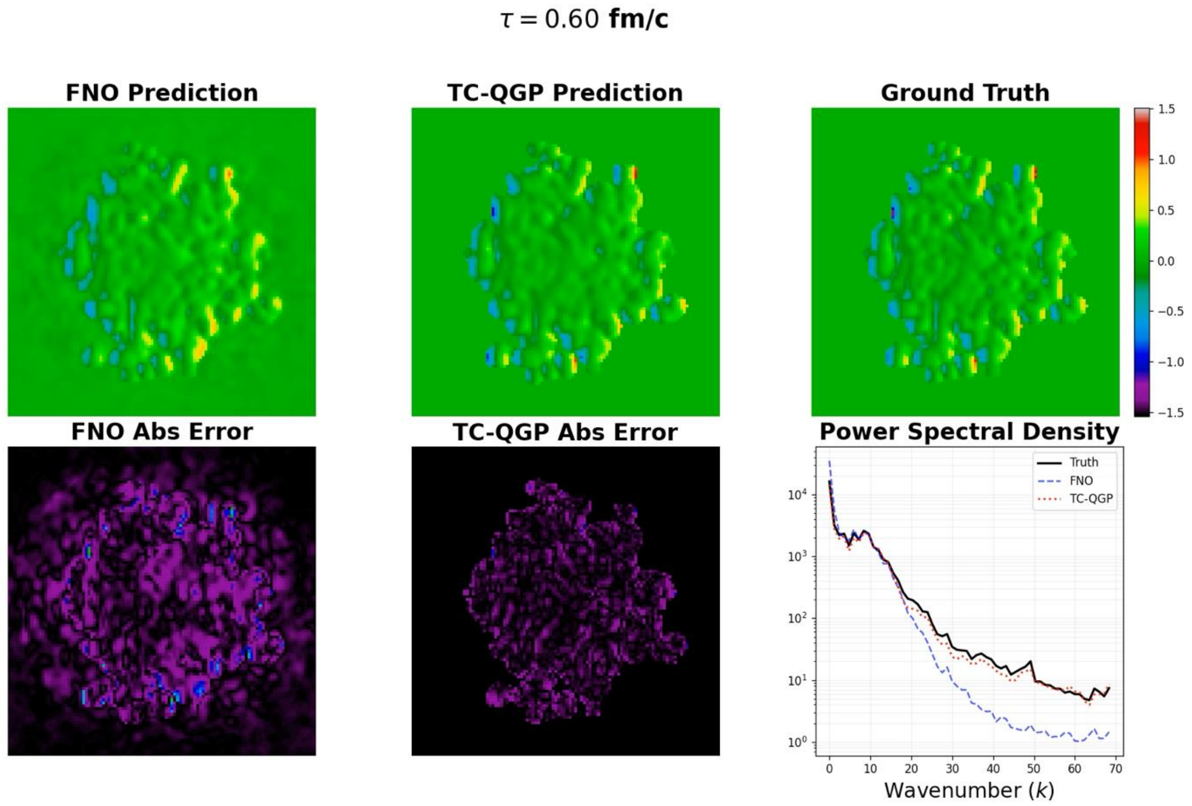
Energy Density



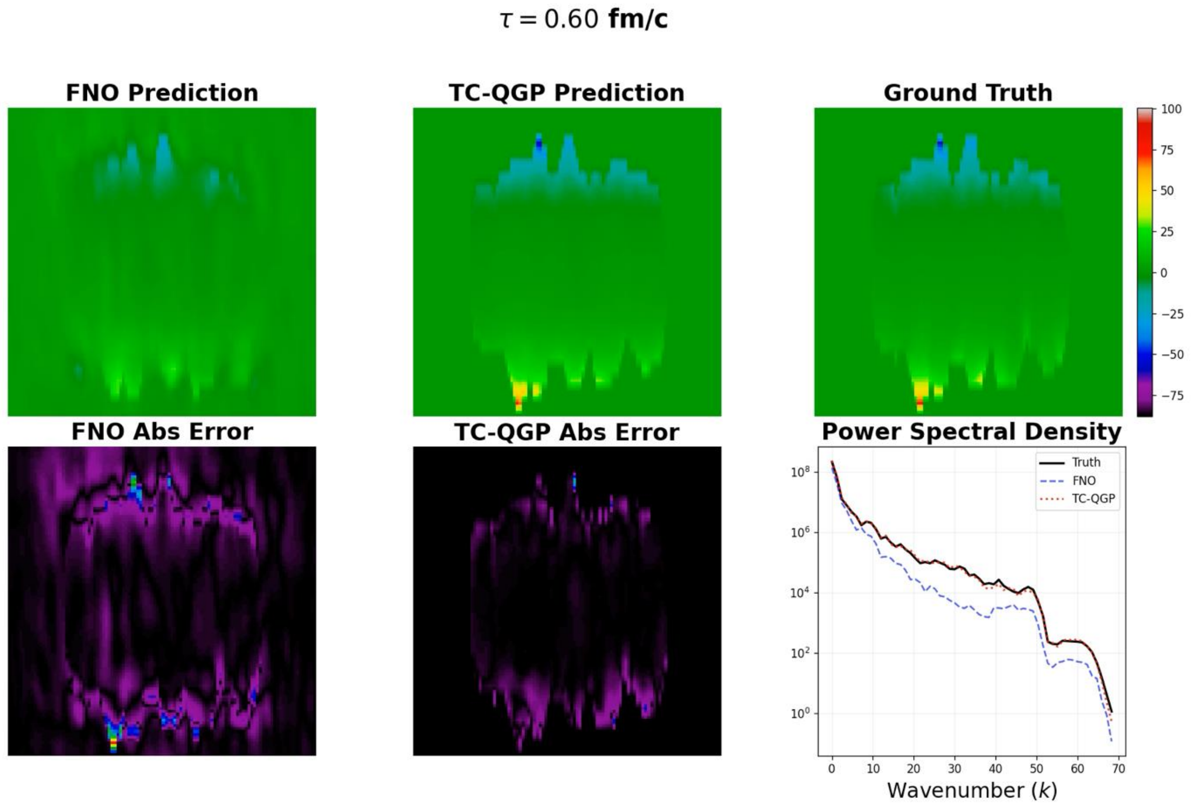
Temperature

Results – 3+1D Velocity Fields u_x, u_z

- **Velocity Fields**



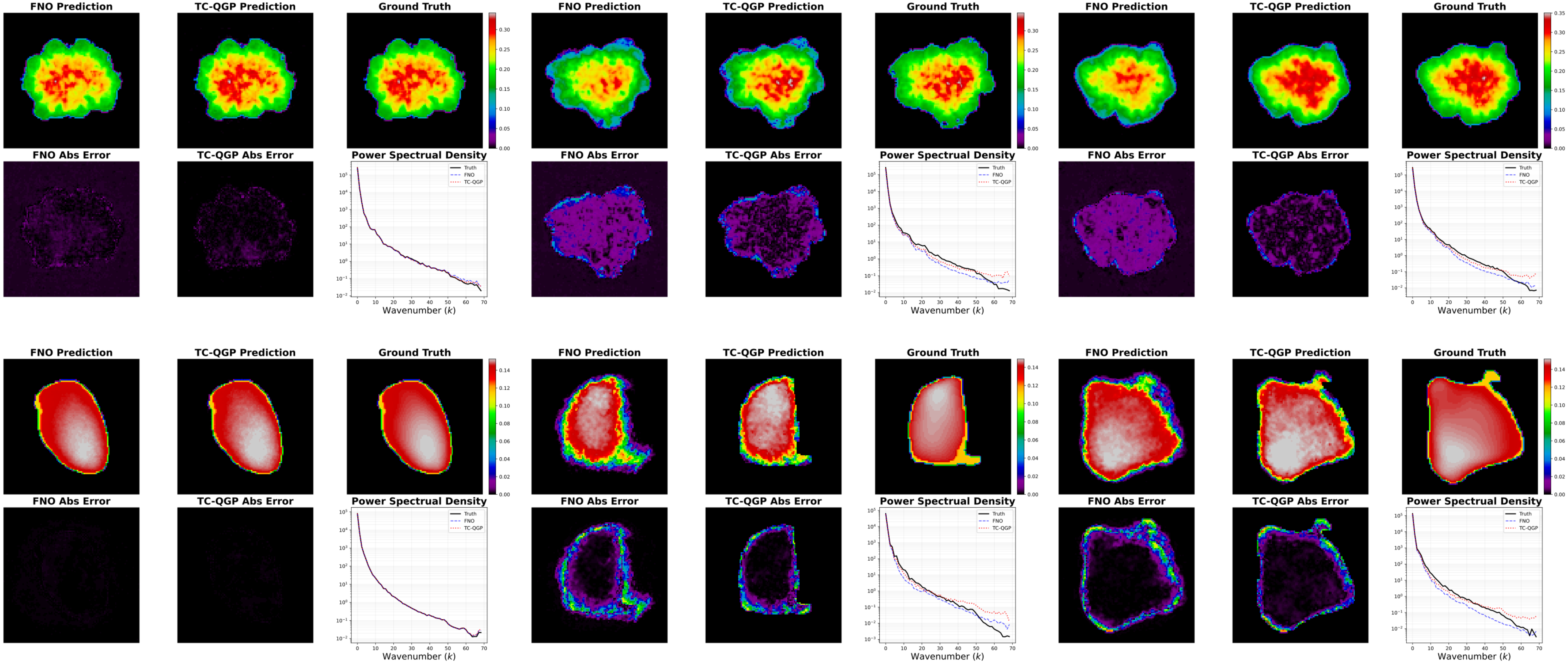
u_x on the $x-y$ plane at mid-rapidity $\eta_s \approx 0$



u_z on the $x-\eta_s$ plane at mid-y

Results – 2+1D Generalization: (0.8, 9.6 fm/c)

In-distribution (ID) | **OOD-Sharp** | **OOD-Smooth**



Results – Quantitative Results

2+1D Dataset

- **Accuracy – Mean Relative Error (%)**

	FNO	TC-QGP
ID	1.10	**0.86**
OOD-Sharp	9.55	**9.23**
OOD-Smooth	8.20	**8.01**

→ ***Outperforms FNO across ID and OOD settings***

- **Cost (single A100, batch 1)**

	FNO	TC-QGP
Parameters	151 M	**21.4 M**
GPU Memory	1,193 MB	**127 MB**
Inference Time	34.4 ms	29.7 ms

→ ***~7× fewer parameters, ~9× less memory*** *with comparable wall time*

vs MUSIC: ~10 mins per event → TC-QGP: ~0.03 s per event

3+1D Dataset

- **Accuracy – Mean Relative Error (%)**

	FNO	TC-QGP	↓
ε	17.44	**4.08**	4.3×
T	13.89	**1.66**	**8.4×**
P	20.32	**4.54**	4.5×
u_x	28.10	**8.52**	3.3×
u_y	27.73	**7.80**	3.6×
u_z	31.44	**5.44**	**5.8×**
All	**23.15**	**5.34**	**4.3×**

- **Cost (single A100, batch 1)**

	FNO	TC-QGP
Parameters	382.4 M	356.6 M
GPU Memory	6,585 MB	**4,459 MB**
Inference Time	108.8 ms	230.4 ms

vs MUSIC: ~20 hours per event → TC-QGP: ~0.2 s per event

Conclusion

- **First surrogate modeling for 3+1D hydrodynamics of the QGP**
  - **Challenges:** dissipative dynamics, evolving boundaries, and 3+1D scale.
  - Standard FNO architectures aren't naturally designed to handle these.

- Our model enable **fast ($> 10^5 \times$ speedup)** and **accurate (within a few percent)** surrogate modeling of QGP hydrodynamic evolution
  - Opening the door to **large-scale event-by-event parameter inference studies.**

- Understanding both the underlying **physics** and the **limitations of standard ML architectures** allowed us to design a specialized model.
  - To address **distribution shift**, we need **deeper understanding** of the underlying physics.