

AI/ML for Nuclear and Particle Physics

Yeonju Go

Brookhaven National Laboratory /
Stony Brook University

July 20-23 2026
Lehigh University

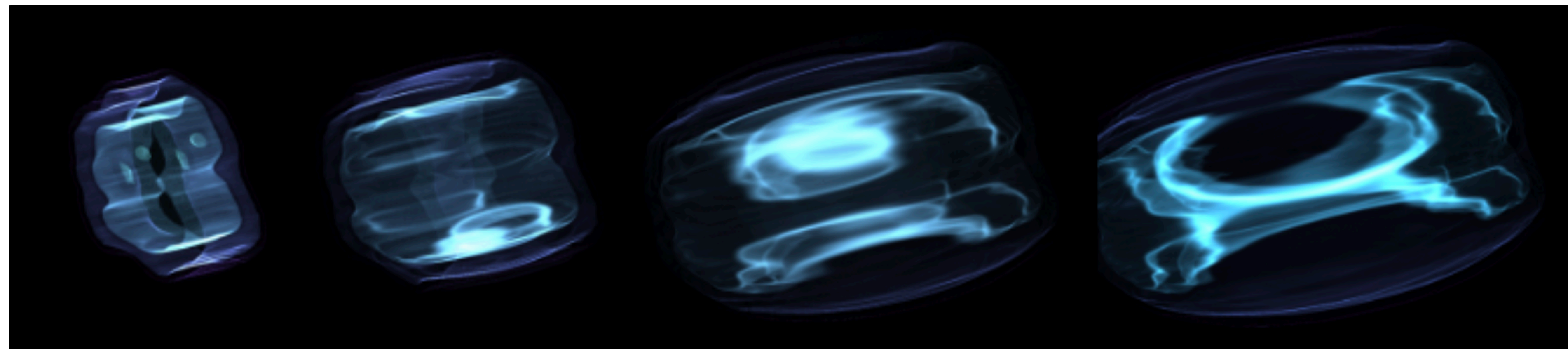


Hydrodynamic Medium Evolution

- Bayesian inference of transport coefficients (e.g. η/s and ζ/s) requires **millions of events**
- Event-by-event *hydrodynamic medium (QGP) evolution* simulations are **computationally very demanding**

	2+1D	3+1D
Simulation time per event with 1 CPU	~10 minutes	~20 hours

Quark Gluon Plasma 3+1D Energy Density Evolution *IP-Glasma+MUSIC*



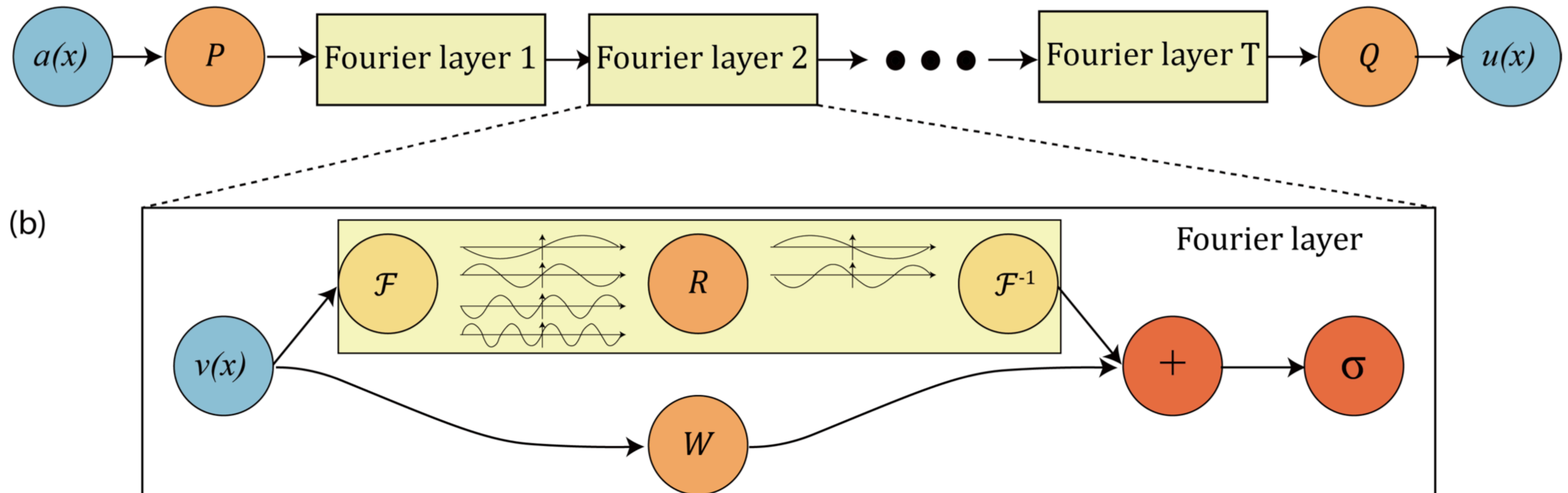
time

➡ **Use AI Surrogate Model for Fast Hydrodynamic Evolution**

Fourier Neural Operator (FNO)

arXiv:2010.08895 [cs.LG]

- **FNO** learns mappings between **functions (operators)** instead of fixed inputs/outputs
→ ideal for Partial Differential Equations (PDE)-governed evolution
- Uses **Fourier Transform** (FFT) to model global patterns efficiently in frequency space
- Well matched to **smooth but structured hydrodynamic fields across time**

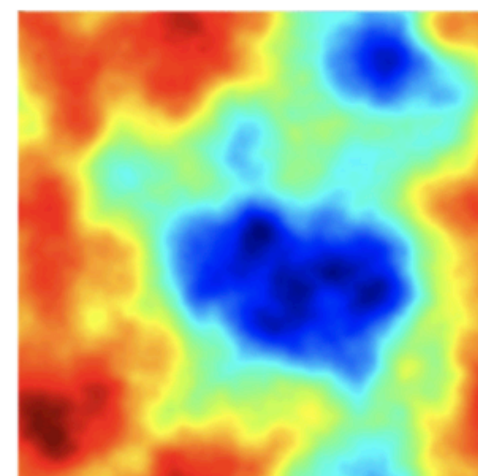


Fourier Neural Operator (FNO)

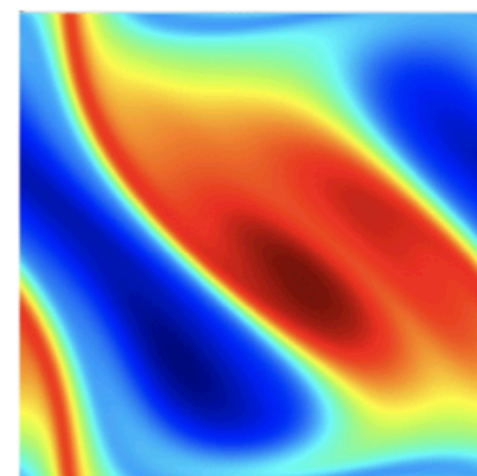
arXiv:2010.08895 [cs.LG]

- **FNO** learns mappings between **functions (operators)** instead of fixed inputs/outputs
→ ideal for Partial Differential Equations (PDE)-governed evolution
- Uses **Fourier Transform** (FFT) to model global patterns efficiently in frequency space
- Well matched to **smooth but structured hydrodynamic fields across time**
- up to **~1000× faster** than traditional PDE solvers
- **Resolution invariant**: it learns the underlying operator on the continuous function, not just a mapping between values on one fixed grid

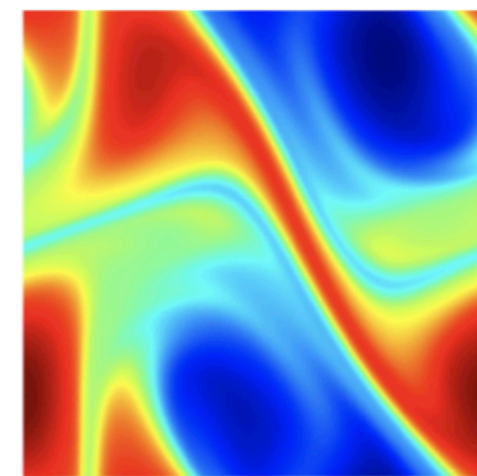
Initial Vorticity



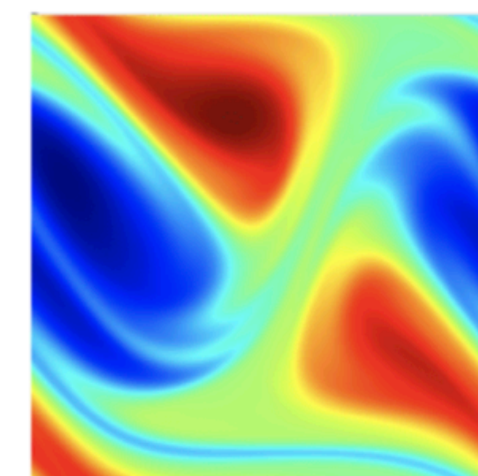
t=15



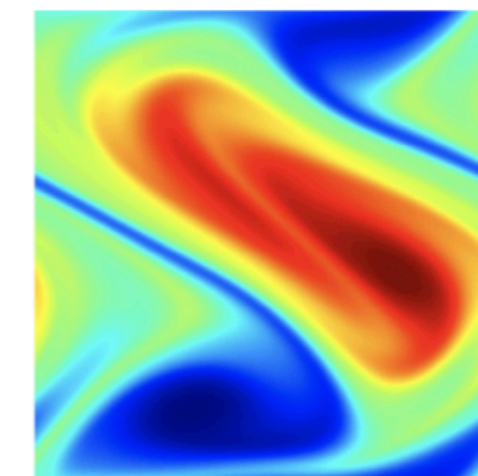
t=20



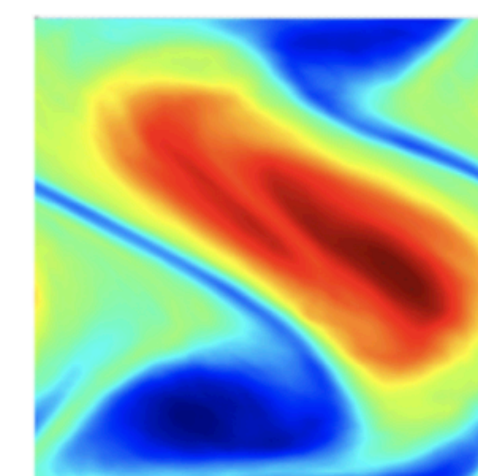
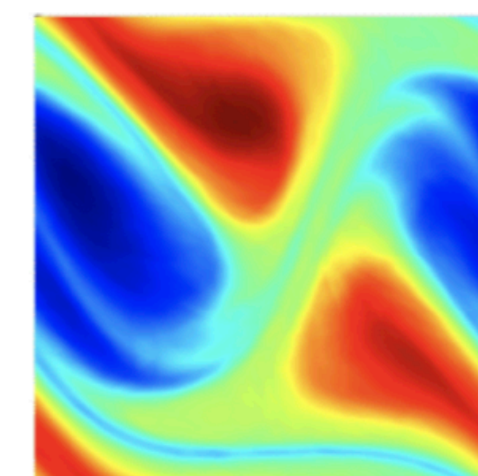
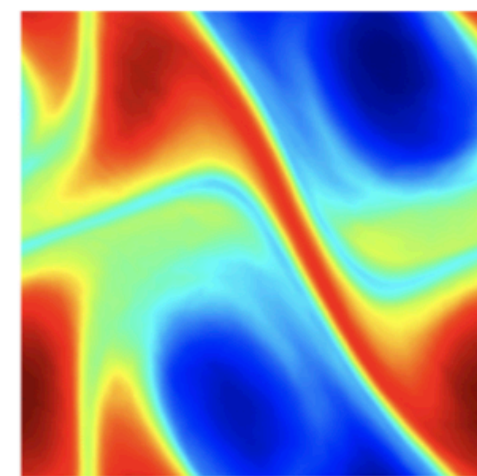
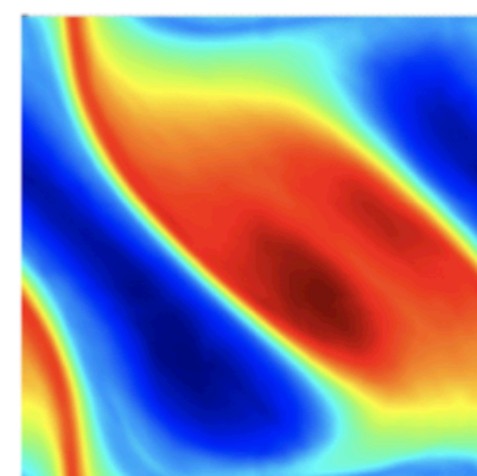
t=25



t=30



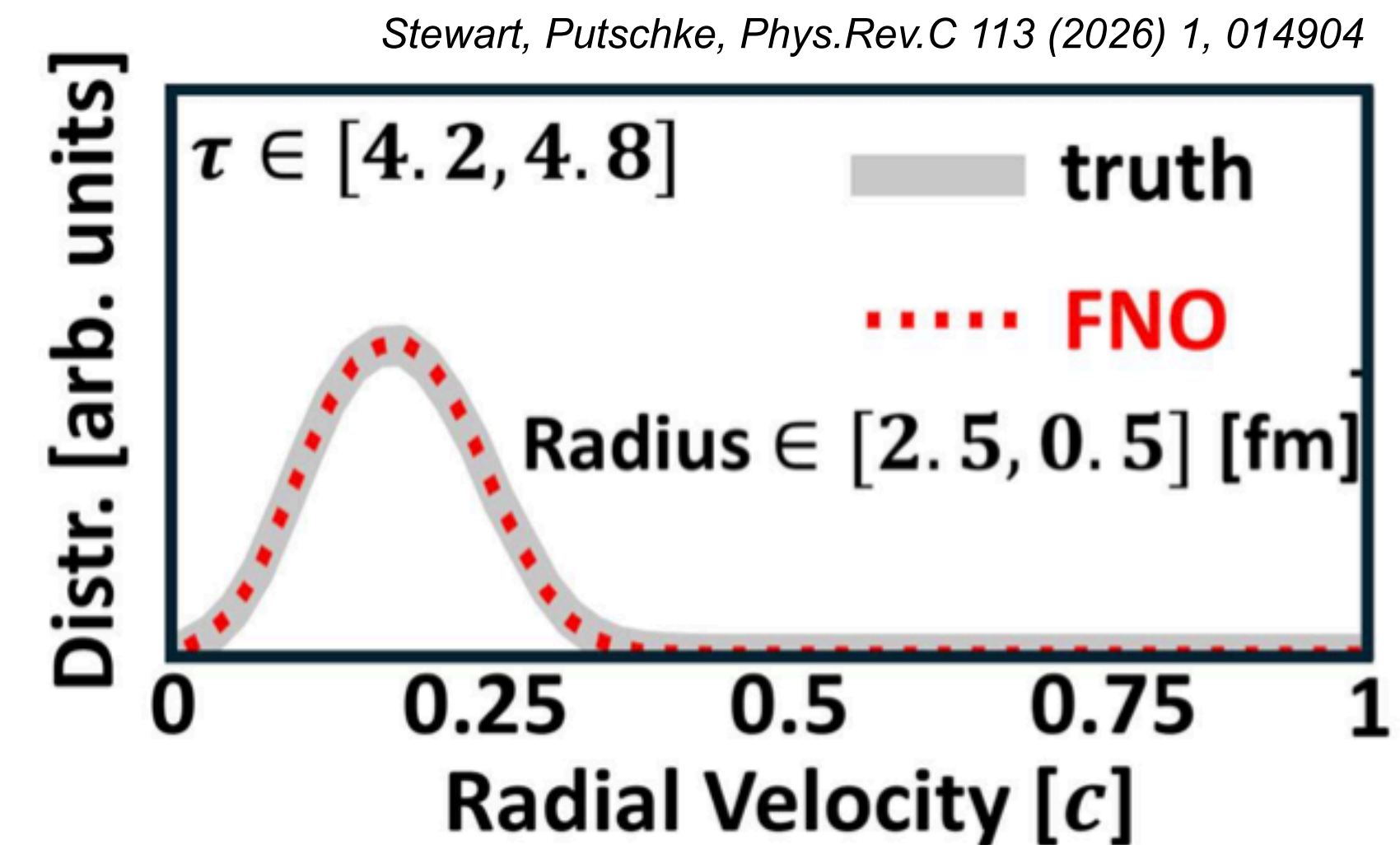
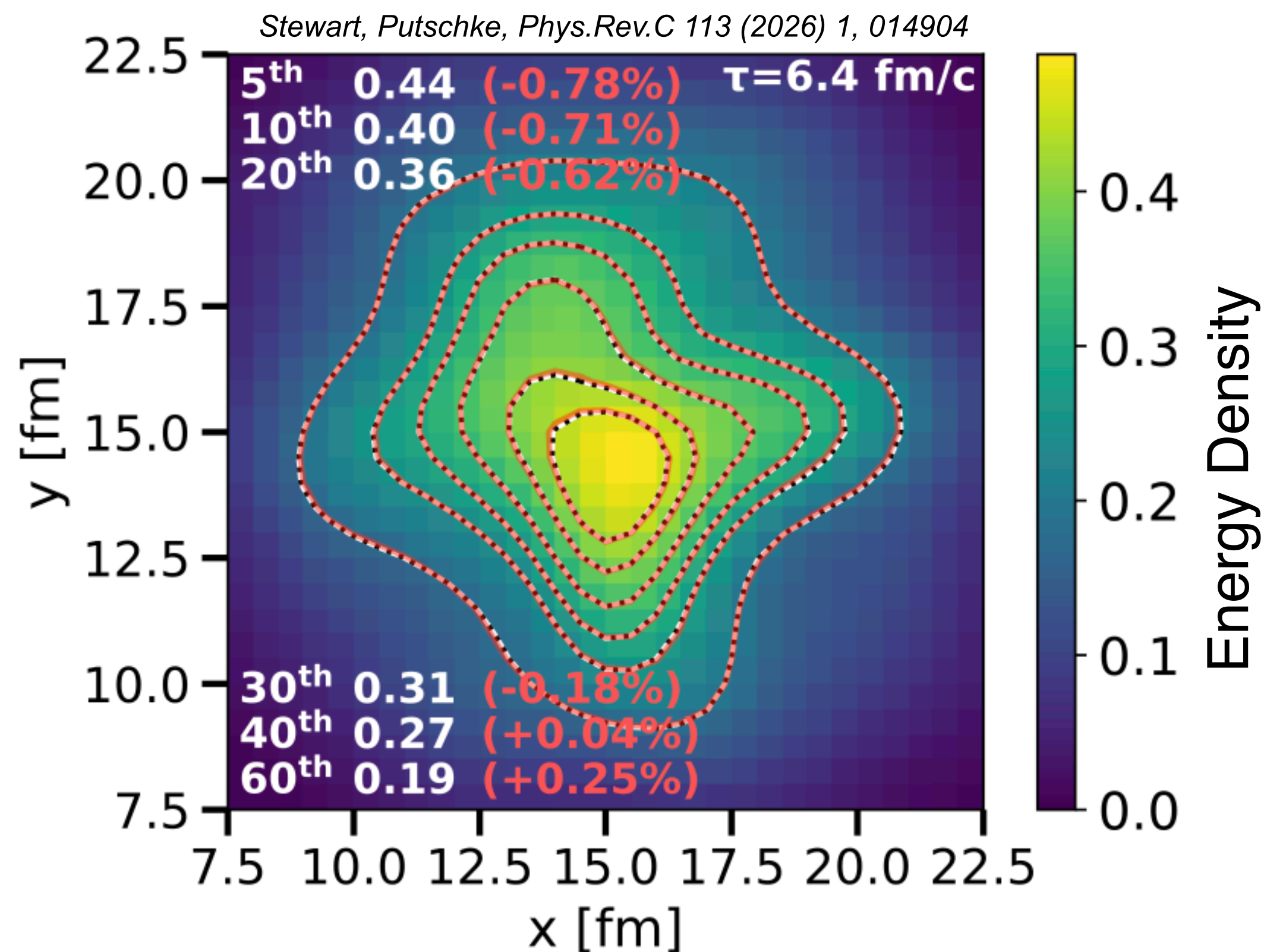
Prediction



Navier Stokes equations
(fluid dynamics)

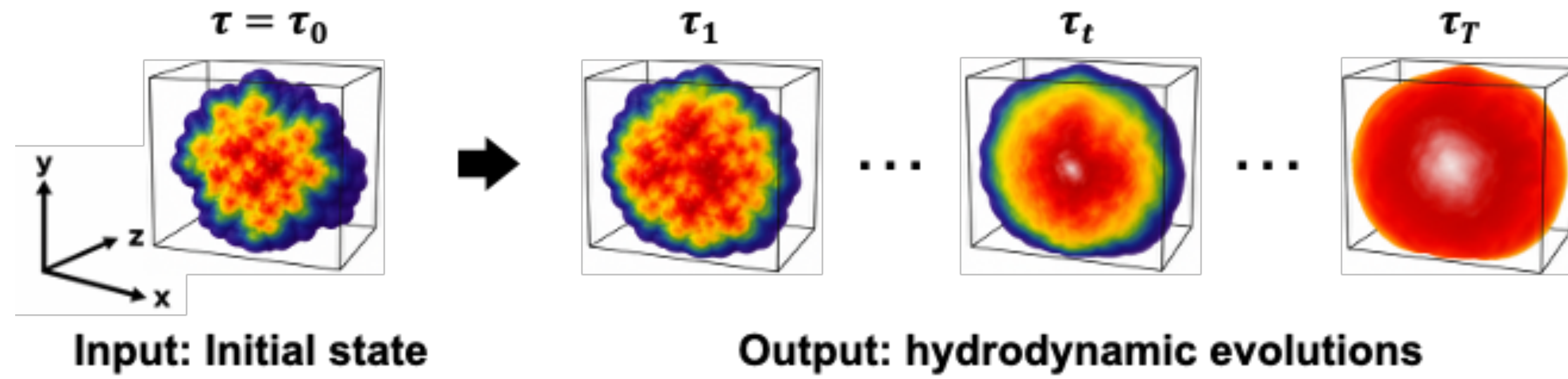
Previous Work on 2+1D Hydro Surrogate Models

- Fast developing field: stacked U-net accelerated **2+1D** simulation in 2021 ([H. Huang et. al. Phys. Rev. Research 3, 023256](#))
- Previous work ([D. Stewart, J. Putschke, Phys.Rev.C 113 \(2026\) 1, 014904](#)) has shown effectiveness of FNO in fast prediction of hydrodynamic QGP evolution
 - ➔ predicted energy density and velocity for different time step in **2+1D**, experimental observables for flow and jet quenching



- Our work extends this approach by:
 - ➔ developing modified architecture for improved fidelity
 - ➔ predicting more hydrodynamic fields (ϵ , T , P , u_x , u_y , u_z)
 - ➔ **scaling to full 3+1D evolution**

Limitation of Standard FNO



- **Challenges of Standard FNO for QGP evolution**

- ➔ **Dissipative dynamics over τ**

- ϵ , T , P decrease rapidly as the system expands and cools

- ➔ **Evolving spatial support with sharp boundaries**

- the active QGP region changes over time, with a sharp transition between the fireball and the surrounding vacuum.

- ➔ **Long temporal rollout**

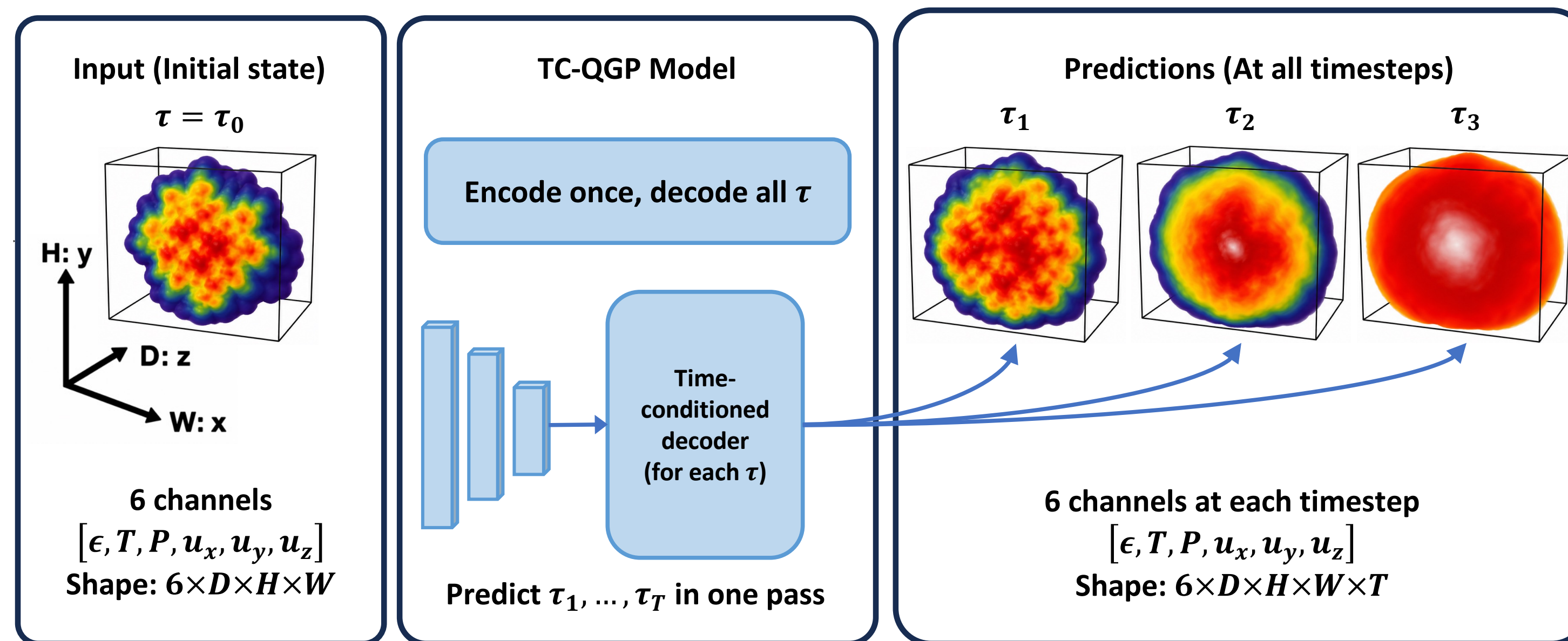
- full event evolution requires accurate and stable predictions over many time steps

- ➔ **Scaling from 2+1D to 3+1D**

- Standard FNO performs reasonably well in 2+1D, but extending to 3+1D introduces substantial memory and computational challenges

Time-Conditioned Neural Operator for QGP

- **TC-QGP Architecture**
 - ➔ **Factorized Neural Operator Encoder**: using Tucker-factorized FNO to map initial state to reduced latent space
 - ➔ **Adaptive Time-Conditioned Decoding**:
 - Embeds each τ and uses *adaptive group normalization* (AdaGN) to modulate latent features through **channel-wise scale and shift parameters**
 - This helps capture *time-dependent field evolution* and *dissipative amplitude scaling*
 - ➔ **Gated Spatial Masking**: to capture evolving field support with time-evolving sharp boundaries



TC-QGP Training Objective

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

$$\mathcal{L} = \frac{1}{T} \sum_{t=1}^T \left[\underbrace{\mathcal{L}_{L^2}(\hat{f}_t, f_t)}_{\text{pointwise field errors}} + \lambda \underbrace{\mathcal{L}_{L^2}(\nabla \hat{f}_t, \nabla f_t)}_{\text{spatial gradient discrepancies}} \right]$$

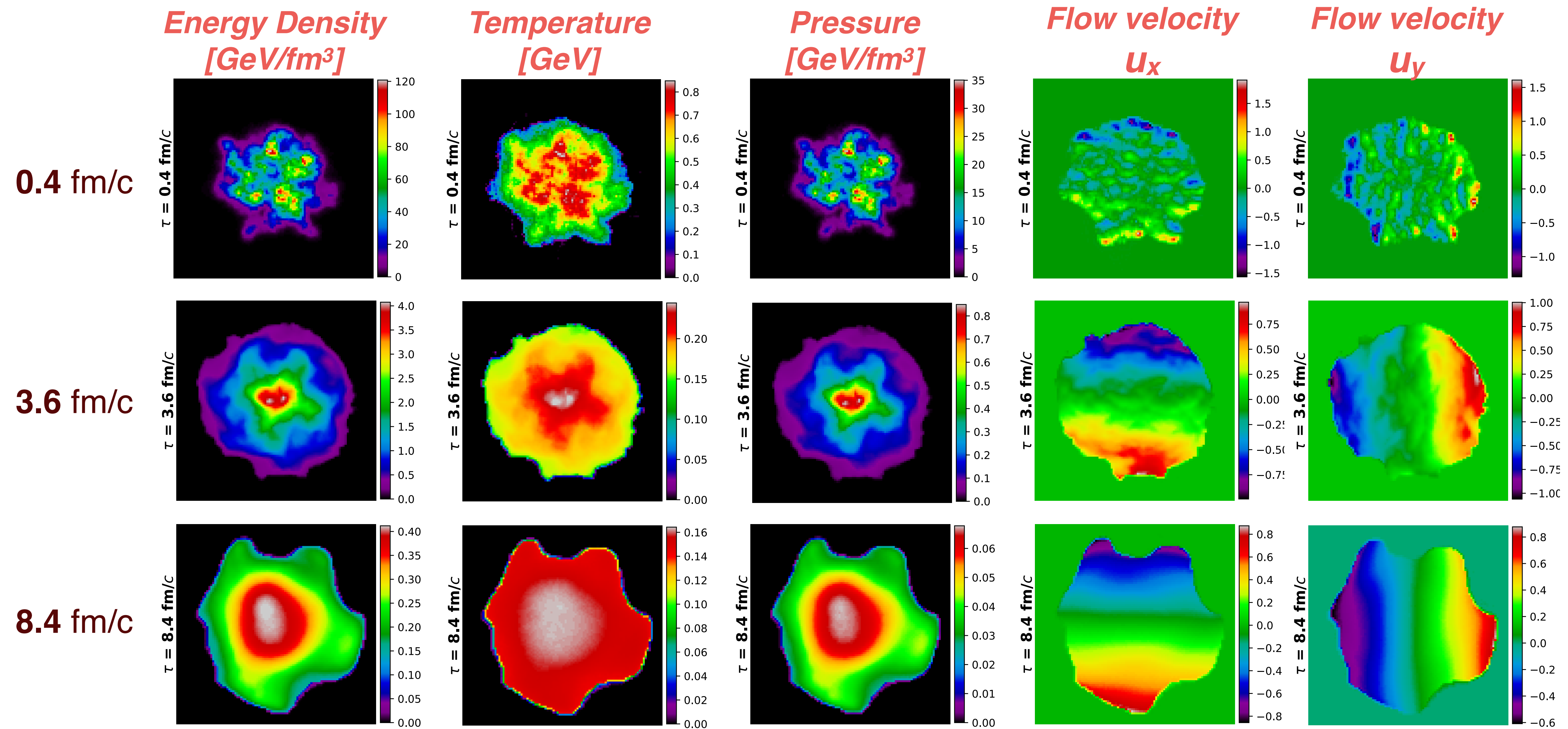
** λ is set to 0.5*

- Time rescaling:** ϵ , T and P are scaled by τ during training, serving as a physics-motivated normalization that compresses the large time-dependent dynamic range caused by expansion and dissipation

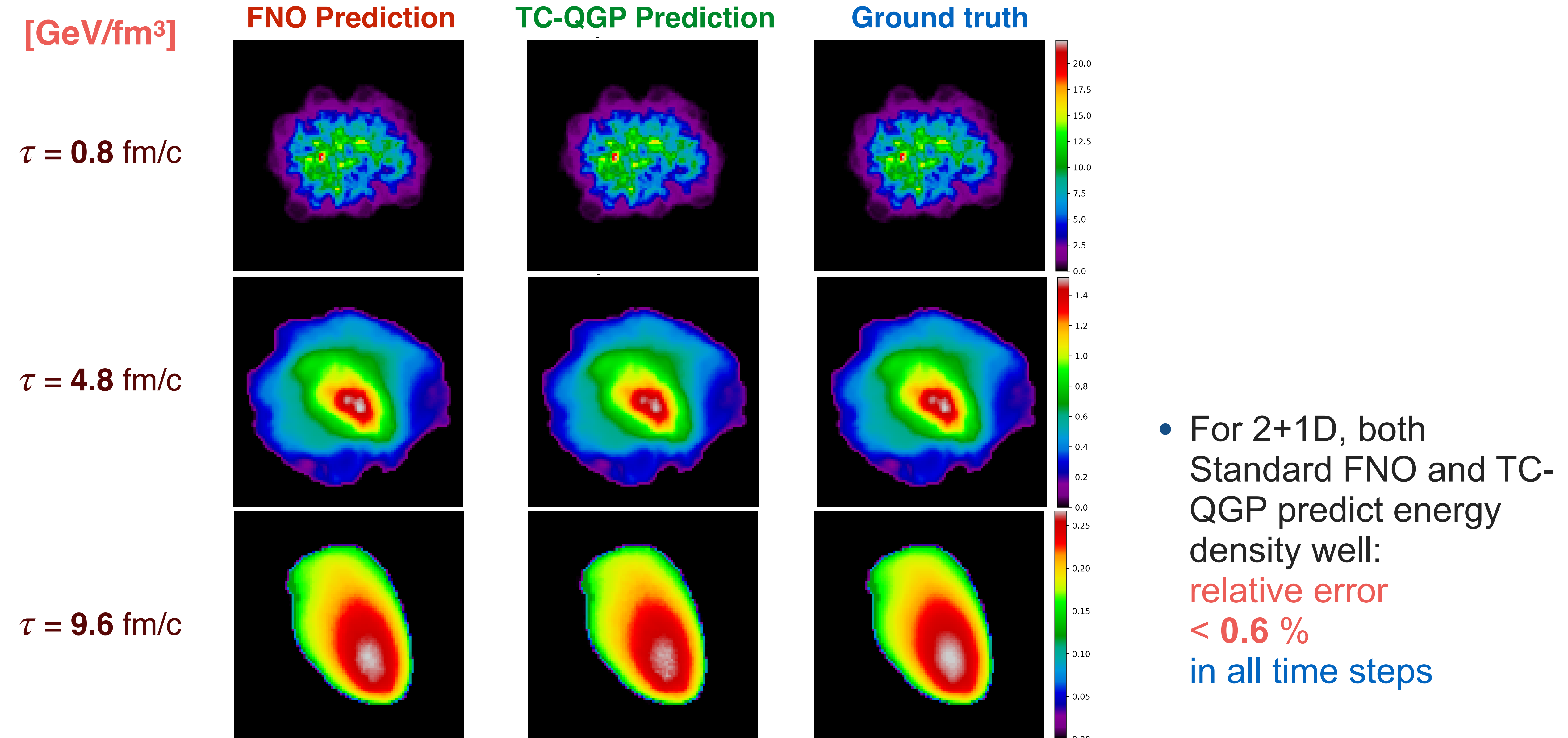
Data and Training Setup

- **Data**
 - ➔ **IP-Glasma + MUSIC** for Au+Au collisions at $\sqrt{s_{\text{NN}}}=200$ GeV
- **2+1D**
 - ➔ $(x \times y) = (100 \times 100)$ grid with $\Delta x = \Delta y = 0.2$ fm
 - ➔ $\tau \in [0.4, 10]$ fm/c with $\Delta\tau = 0.16$ fm/c
 - ➔ 5 fields (**energy density ϵ , Temperature T , Pressure P , Flow velocity u_x, u_y**)
 - ➔ 612 events used: 552 for training / 60 for testing
- **3+1D**
 - ➔ $(\eta \times x \times y) = (28 \times 128 \times 128)$ grid with $\Delta x = \Delta y = 0.2$ fm
 - ➔ $\tau \in [0.4, 8.4]$ fm/c with $\Delta\tau = 0.2$ fm/c
 - ➔ 6 fields (**energy density ϵ , Temperature T , Pressure P , Flow velocity u_x, u_y, u_z**)
 - ➔ 250 events used: 225 for training / 25 for testing

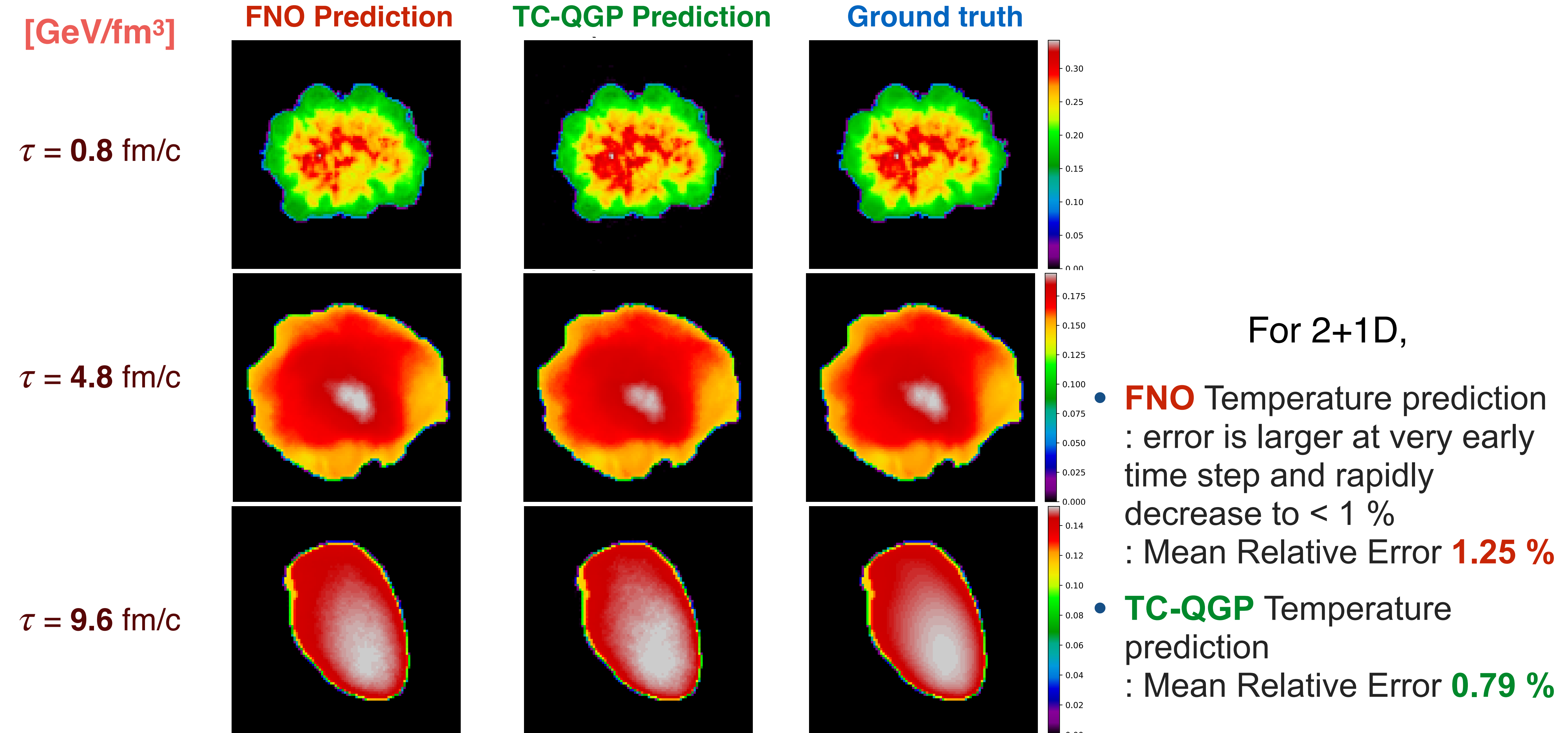
IP-Glasma+MUSIC Hydro 2+1D Evolution



Result: 2+1D Energy Density



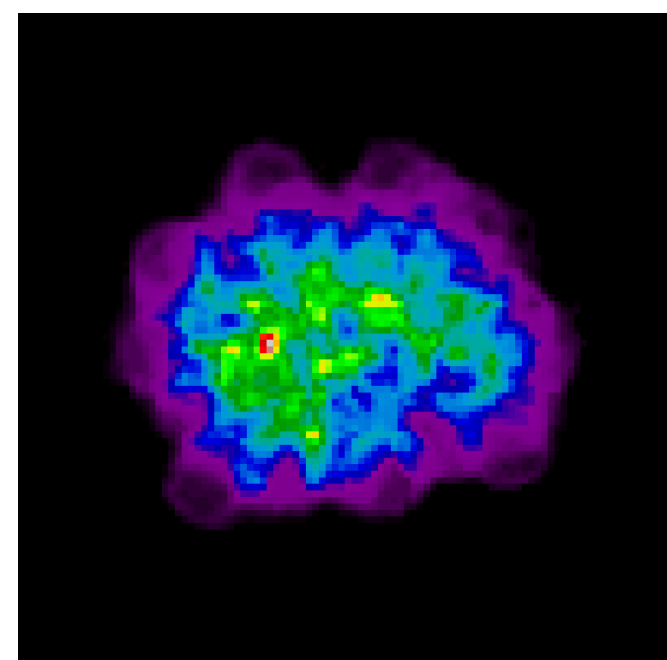
Result: 2+1D Temperature



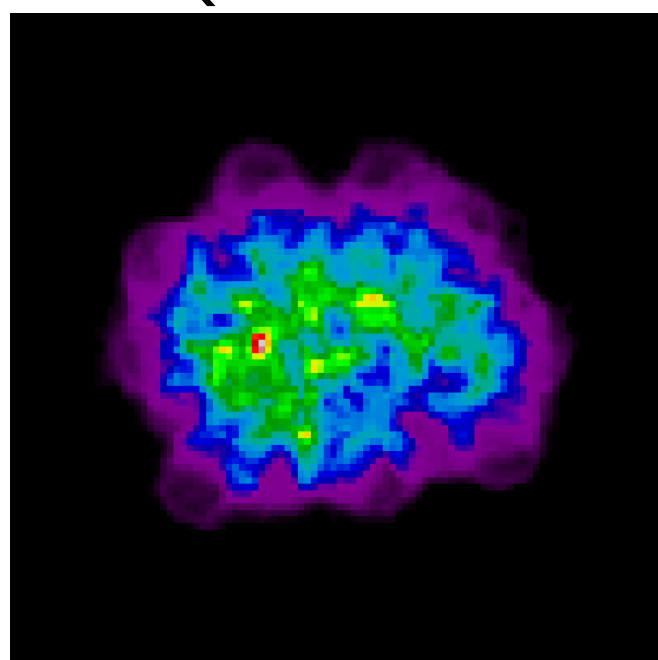
Fourier Spectrum Validation: 2+1D Energy Density

$\tau = 0.8 \text{ fm/c}$

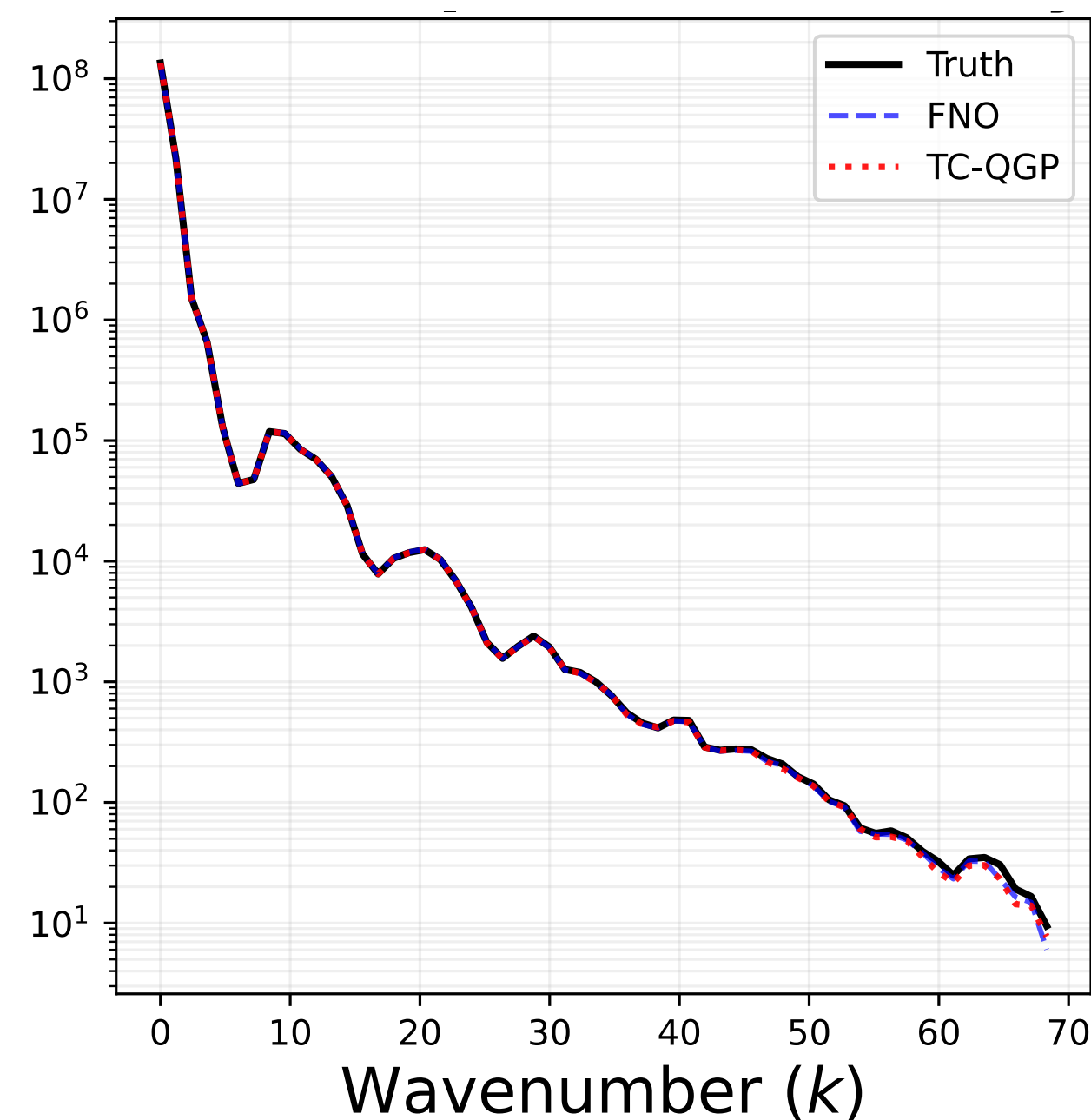
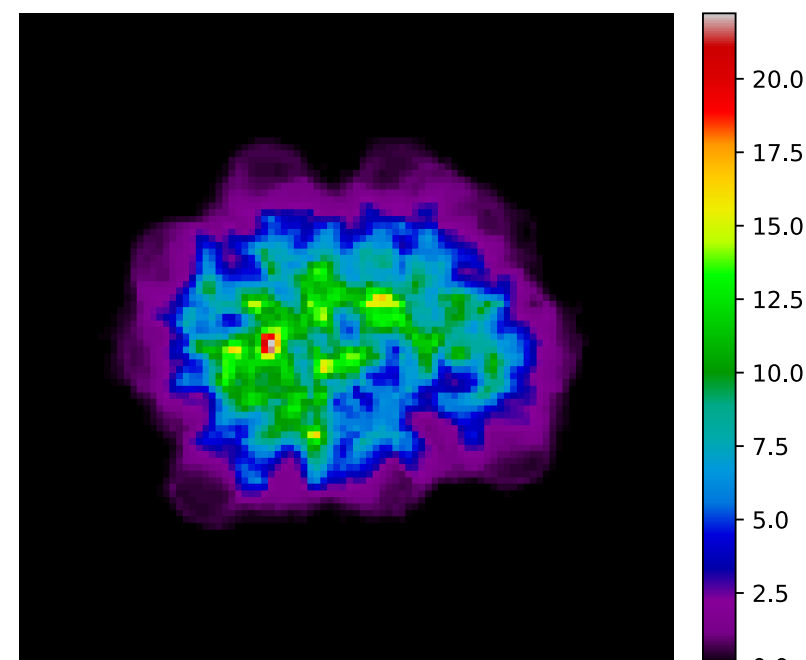
FNO Prediction



TC-QGP Prediction

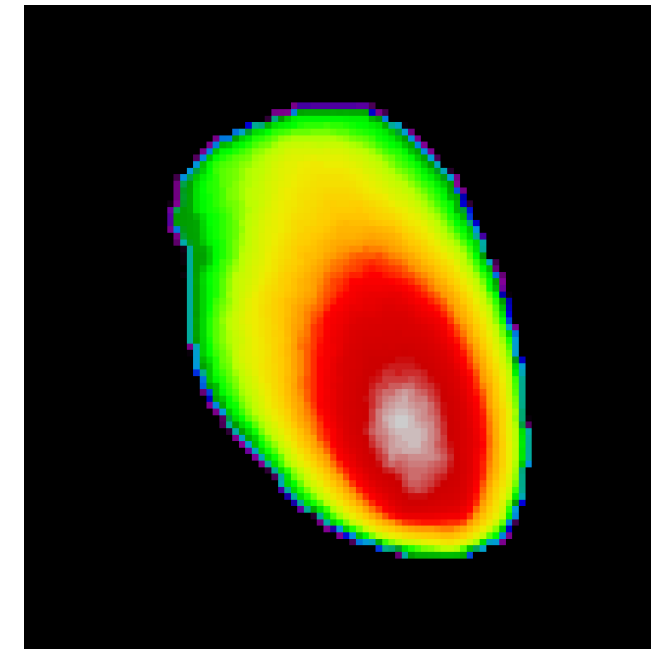


Ground Truth

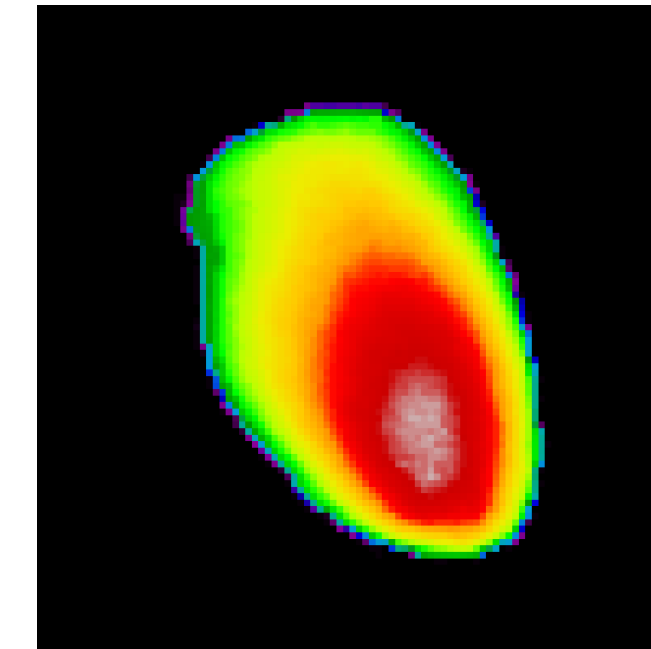


$\tau = 9.6 \text{ fm/c}$

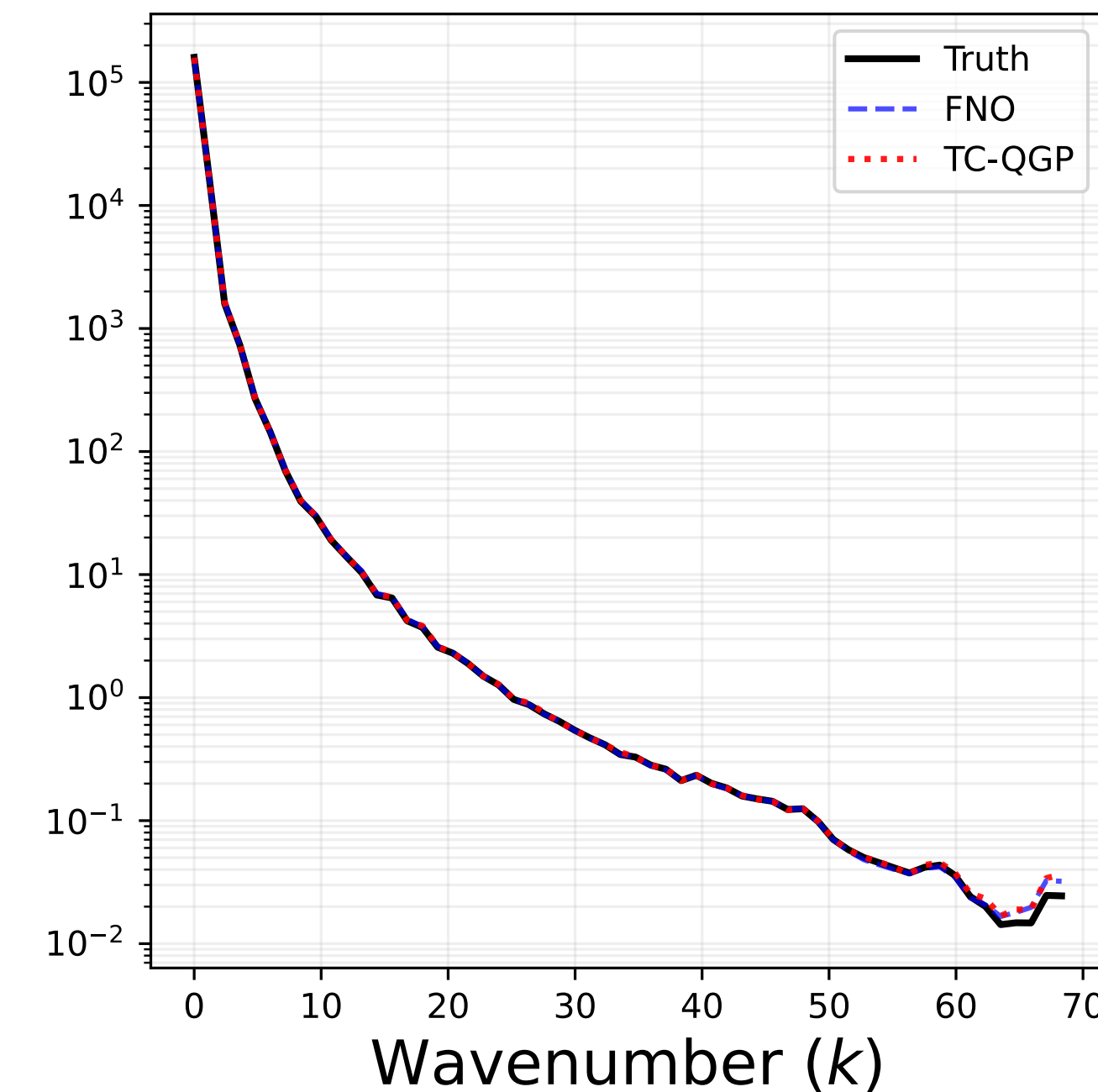
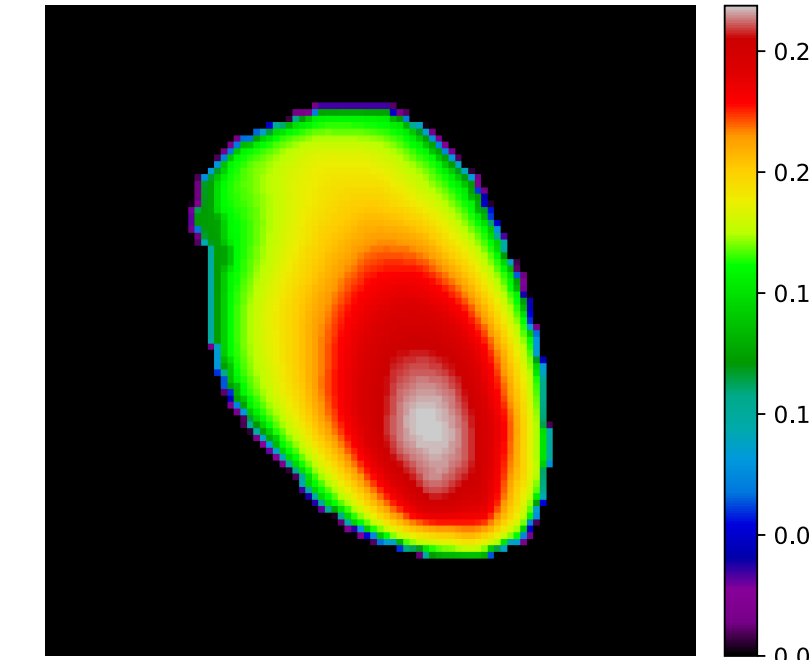
FNO Prediction



TC-QGP Prediction



Ground Truth

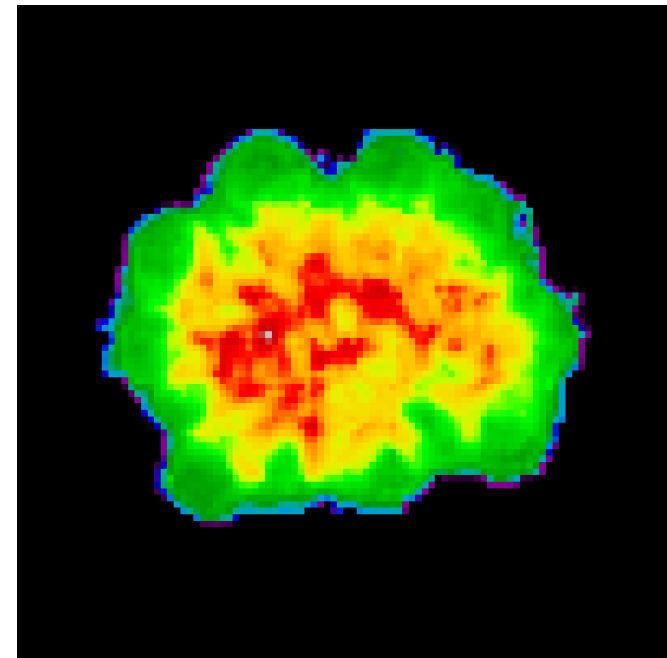


- Fourier frequency distributions for ϵ are well reproduced in both FNO and TC-QGP

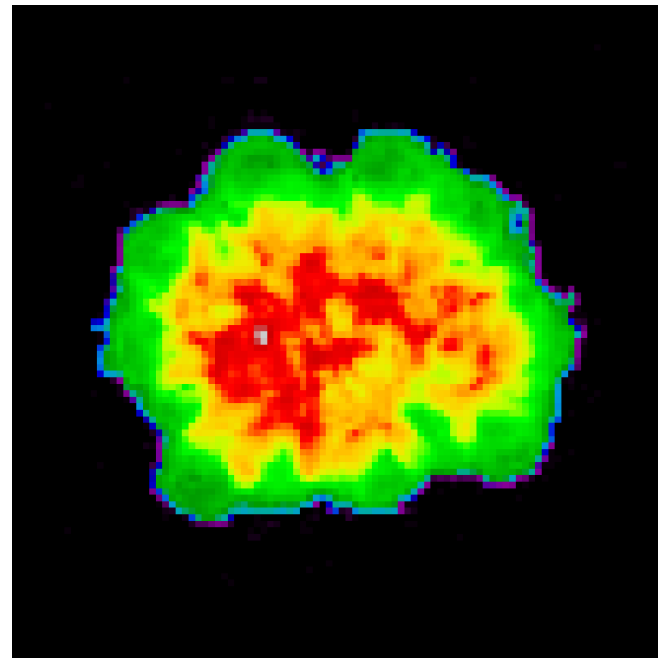
Fourier Spectrum Validation: 2+1D Temperature

$\tau = 0.8 \text{ fm/c}$

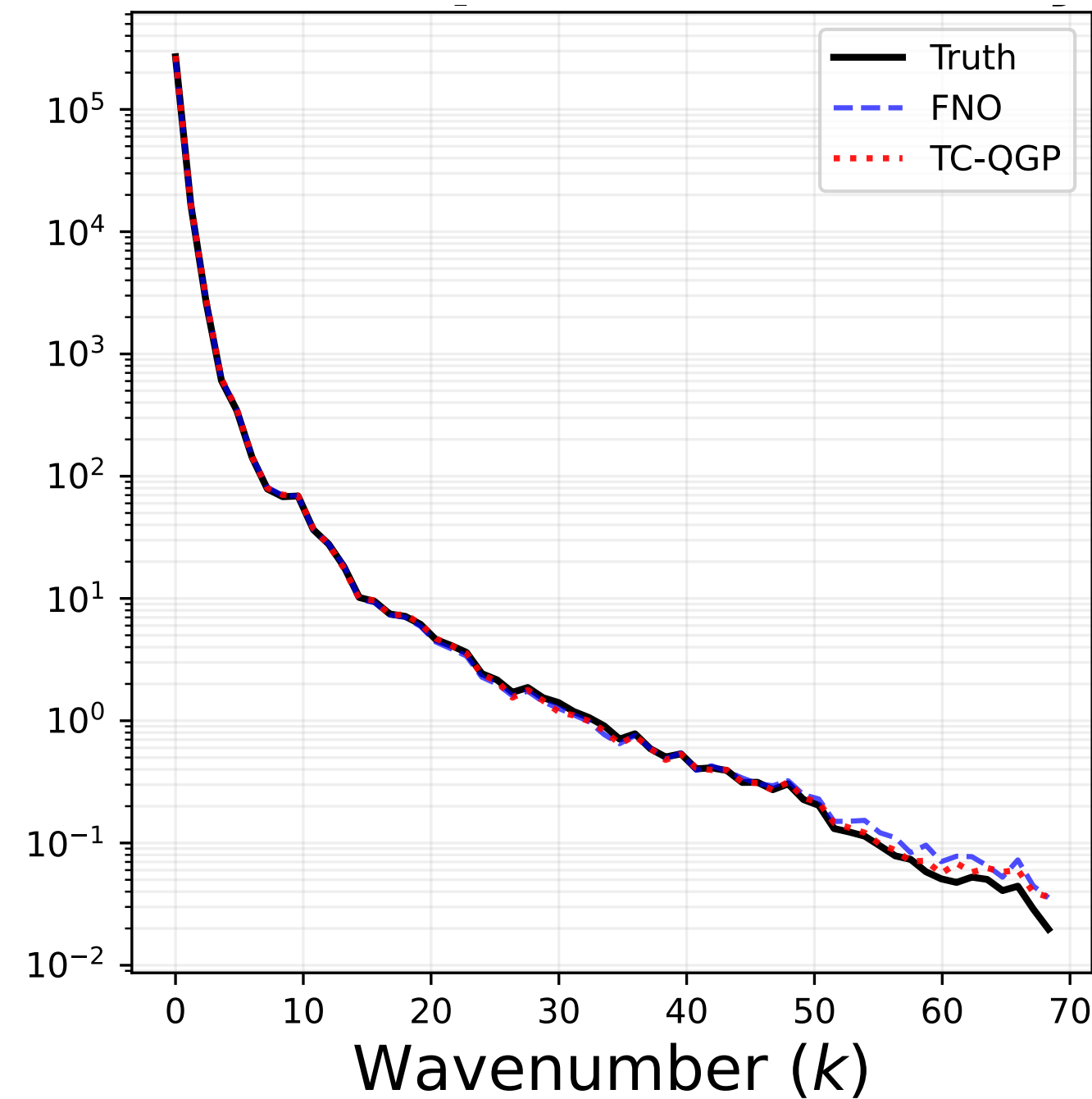
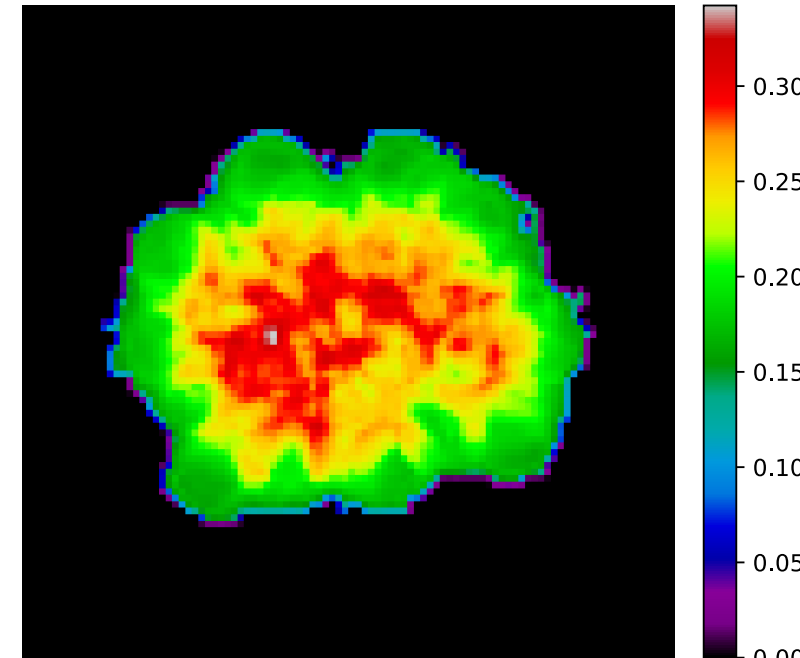
FNO Prediction



TC-QGP Prediction

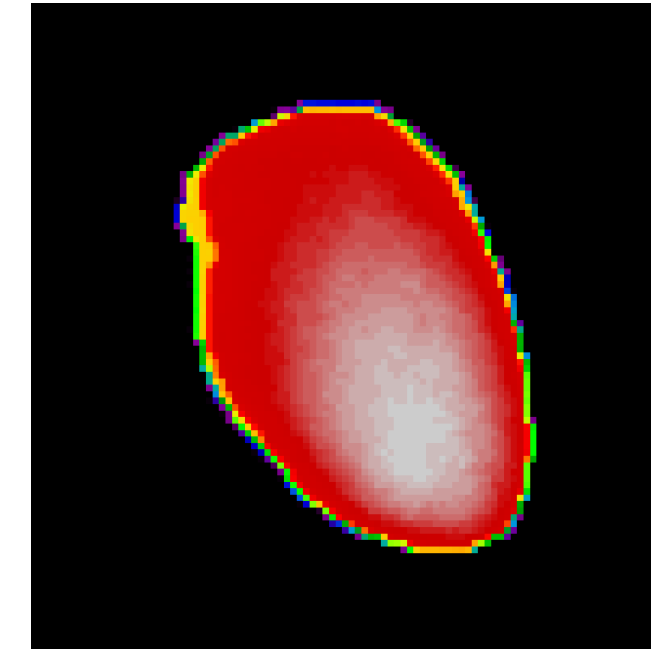


Ground Truth

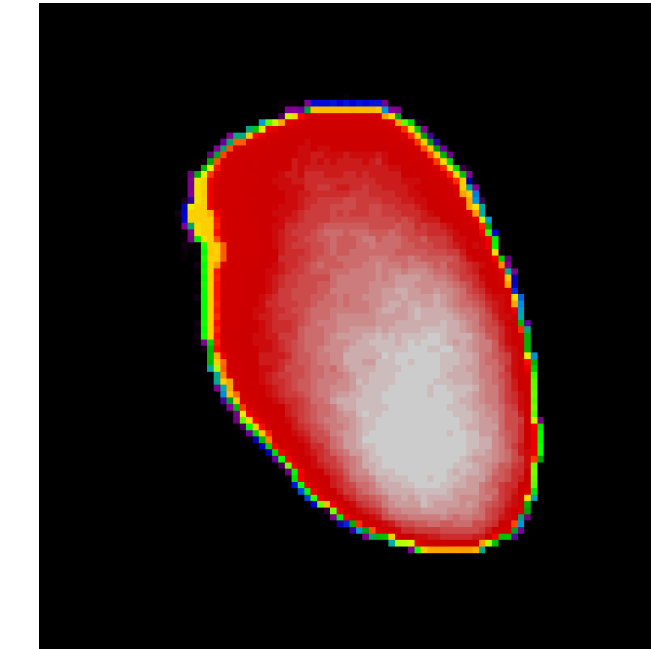


$\tau = 9.6 \text{ fm/c}$

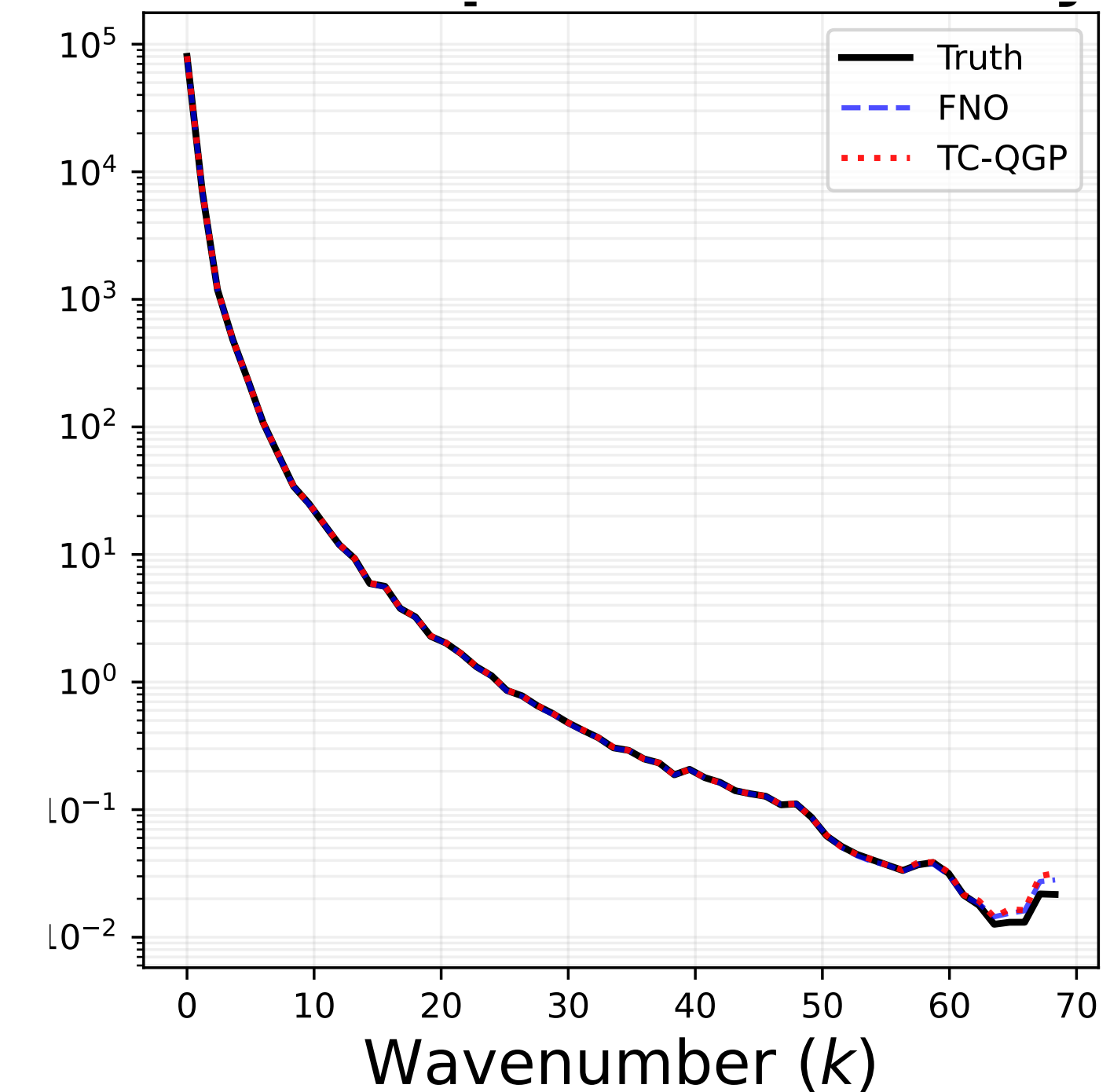
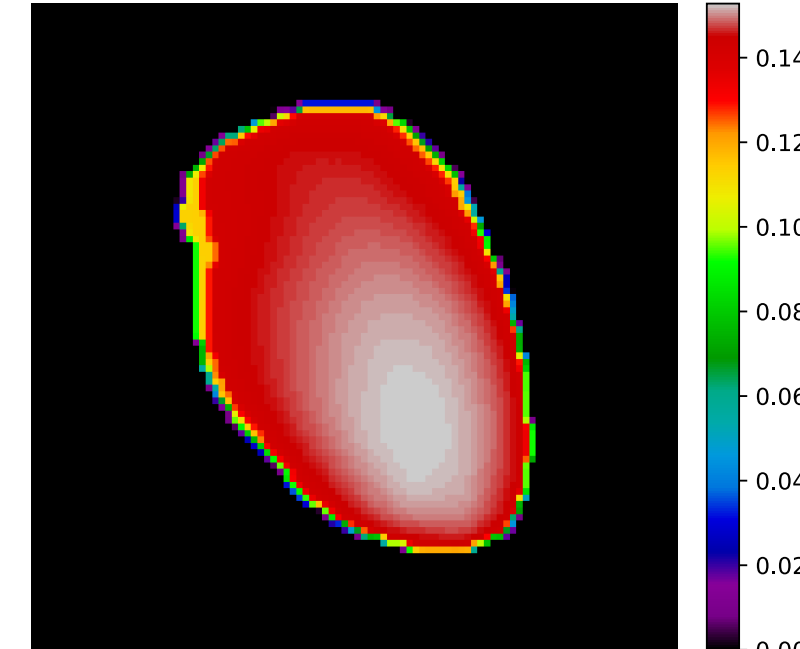
FNO Prediction



TC-QGP Prediction



Ground Truth



- For **temperature at early times**, Fourier frequency predictions are slightly worse

Out-of-distribution Handling (2+1D)

- Tested on **different initial condition** i.e. different hotspot size
➔ training: hotspot size of ~ 0.2 fm, testing: ~ 0.1 fm (Sharp) or ~ 0.4 fm (Smooth)

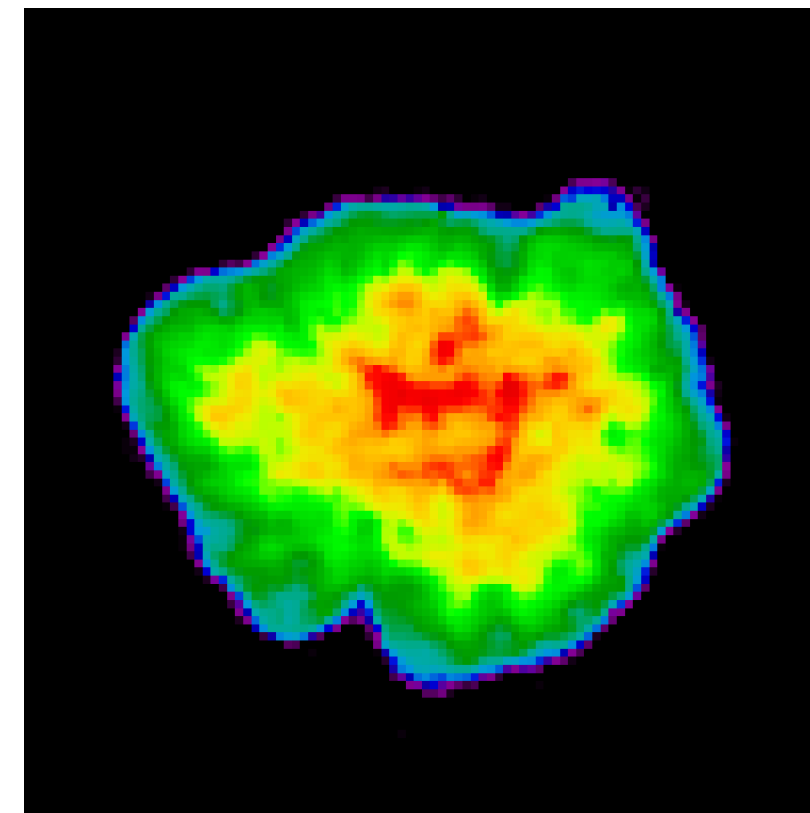
- Both models **still capture the global structure** with increased error

➔ TC-QGP slightly better

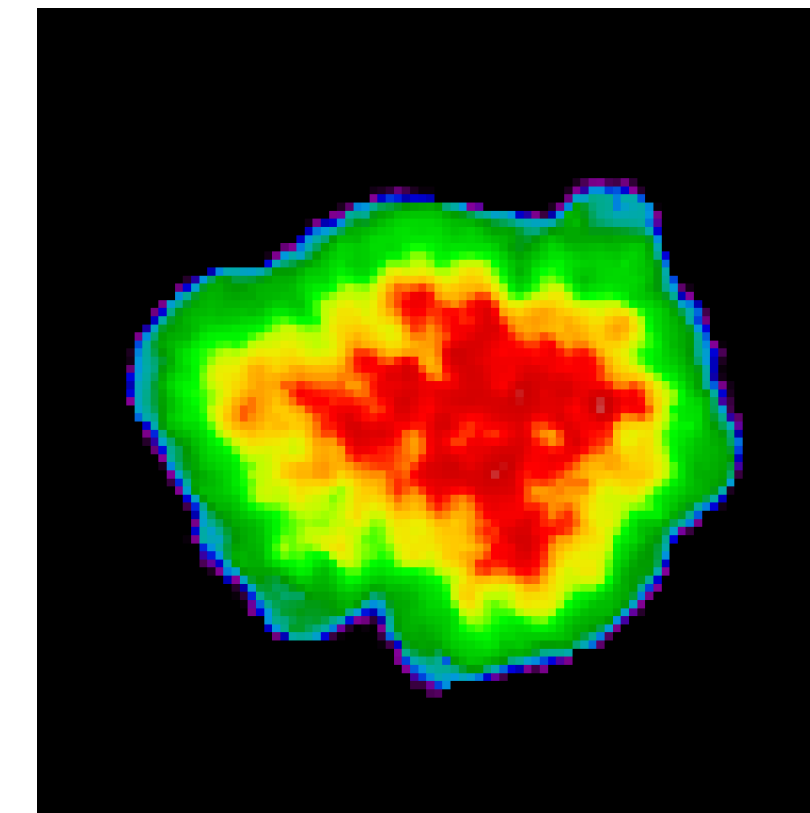
➔ **Training on mixed hotspot sizes would reduce the error**

Out-of-distribution (OOD)-Smooth, Temperature, $\tau=0.8$ fm/c

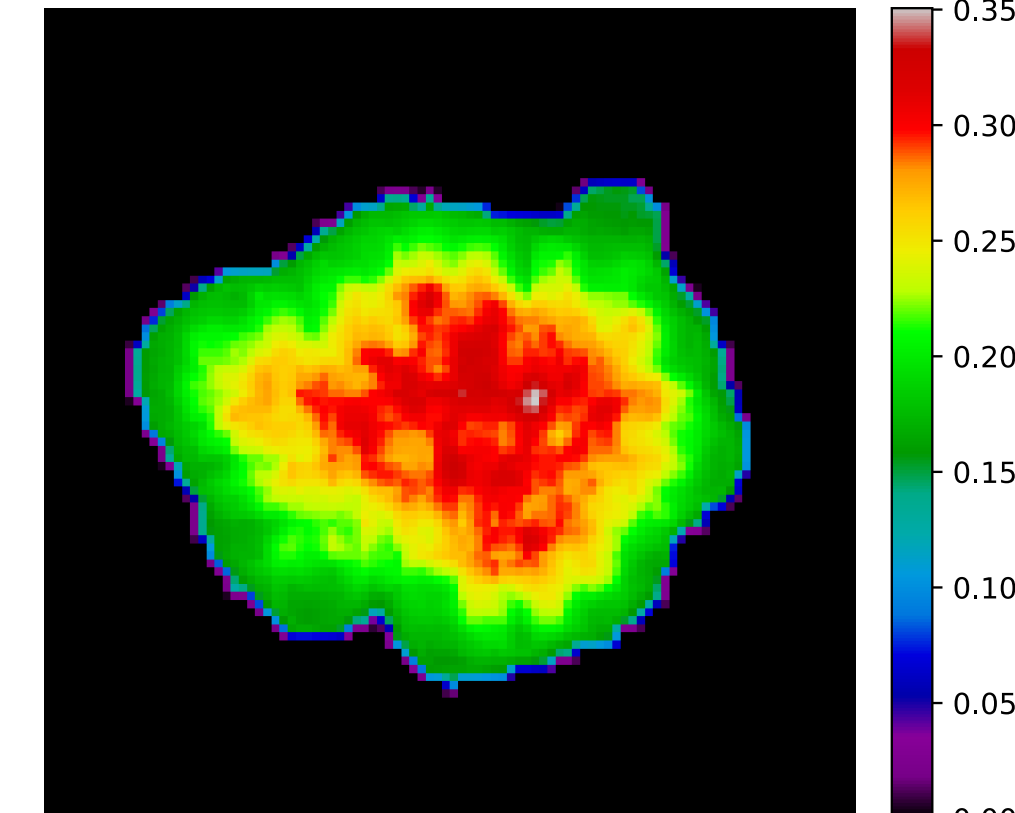
FNO Prediction



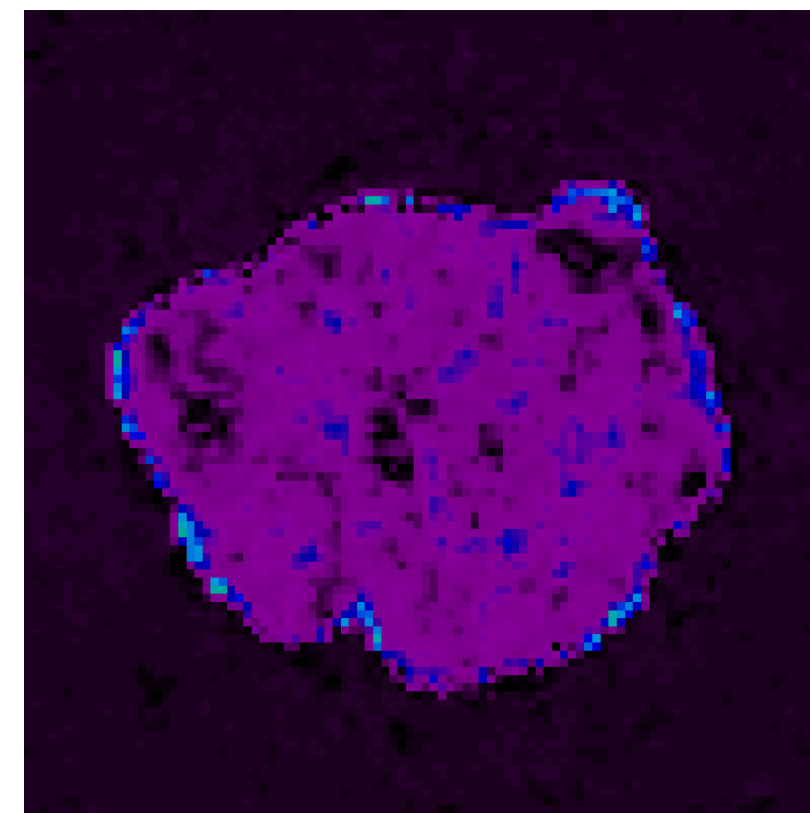
TC-QGP Prediction



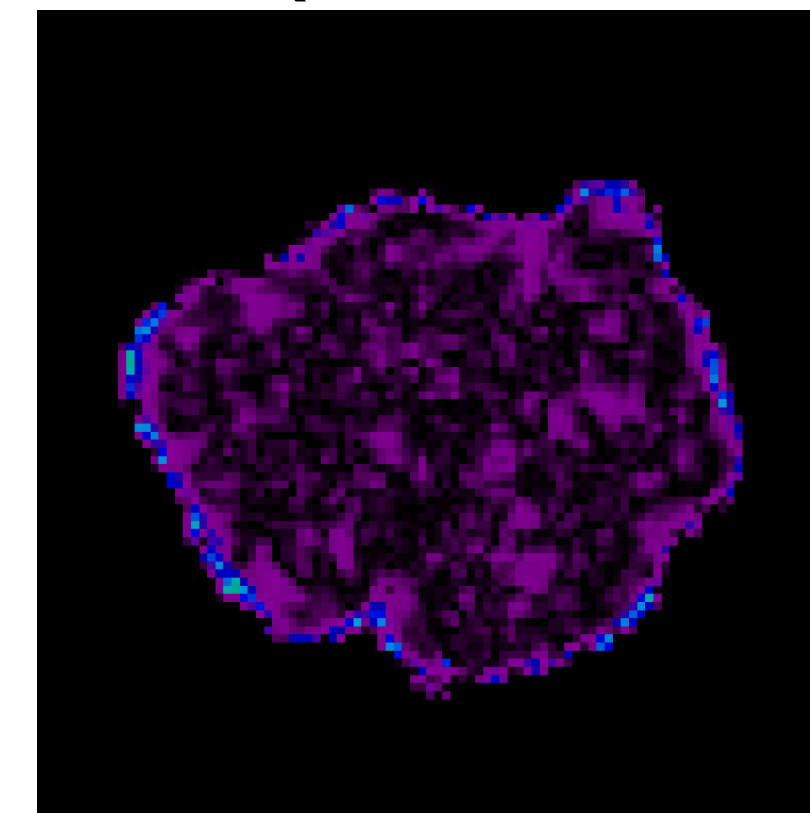
Ground Truth



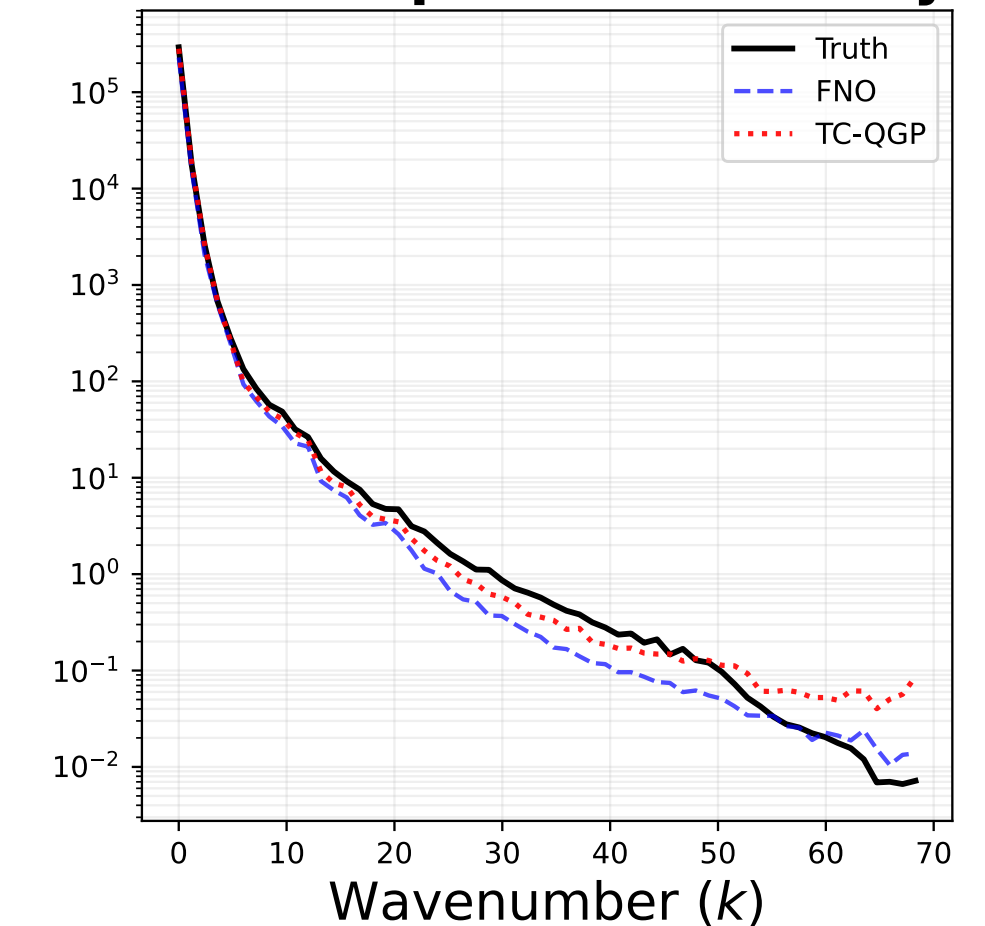
FNO Abs Error



TC-QGP Abs Error



Power Spectral Density



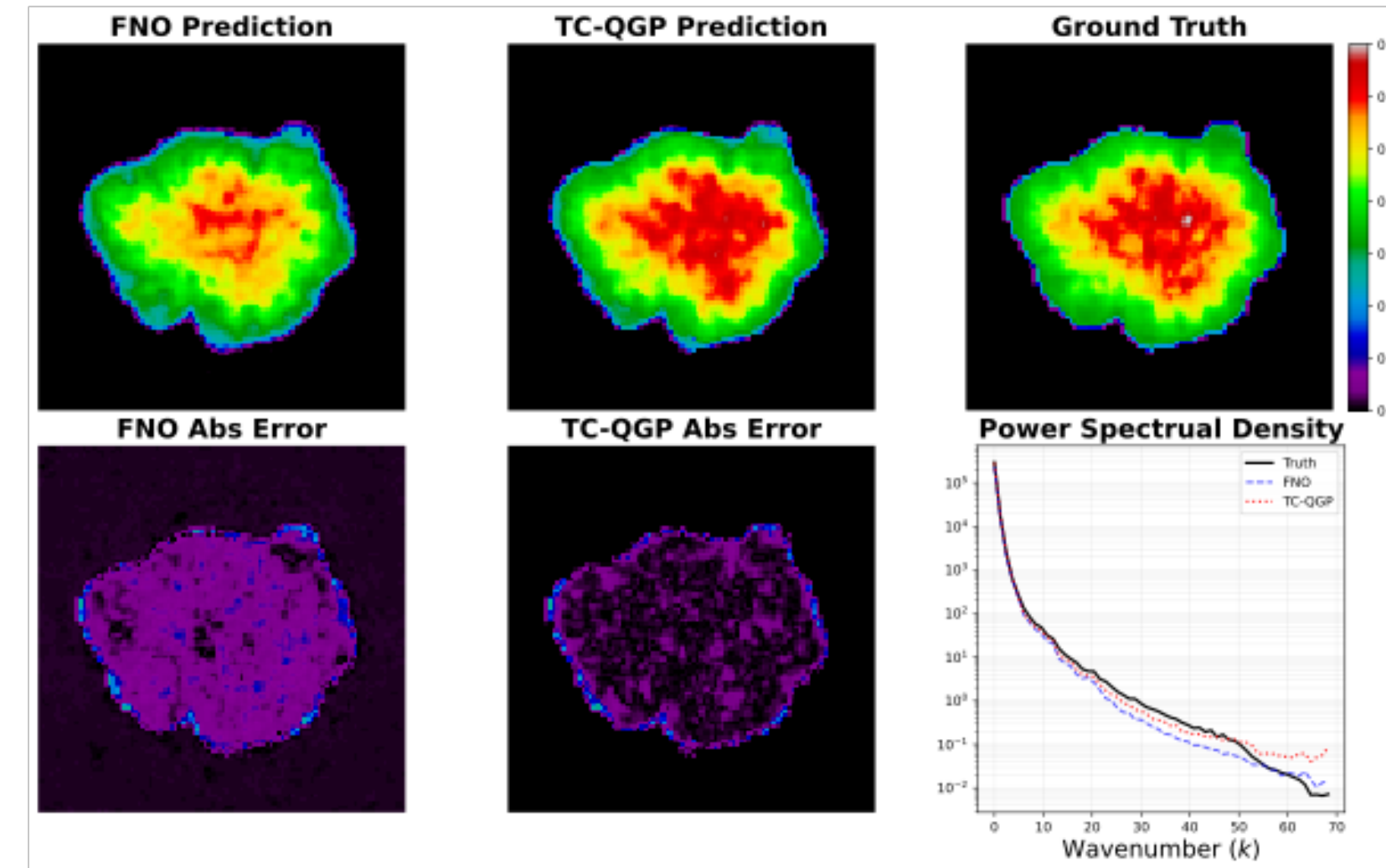
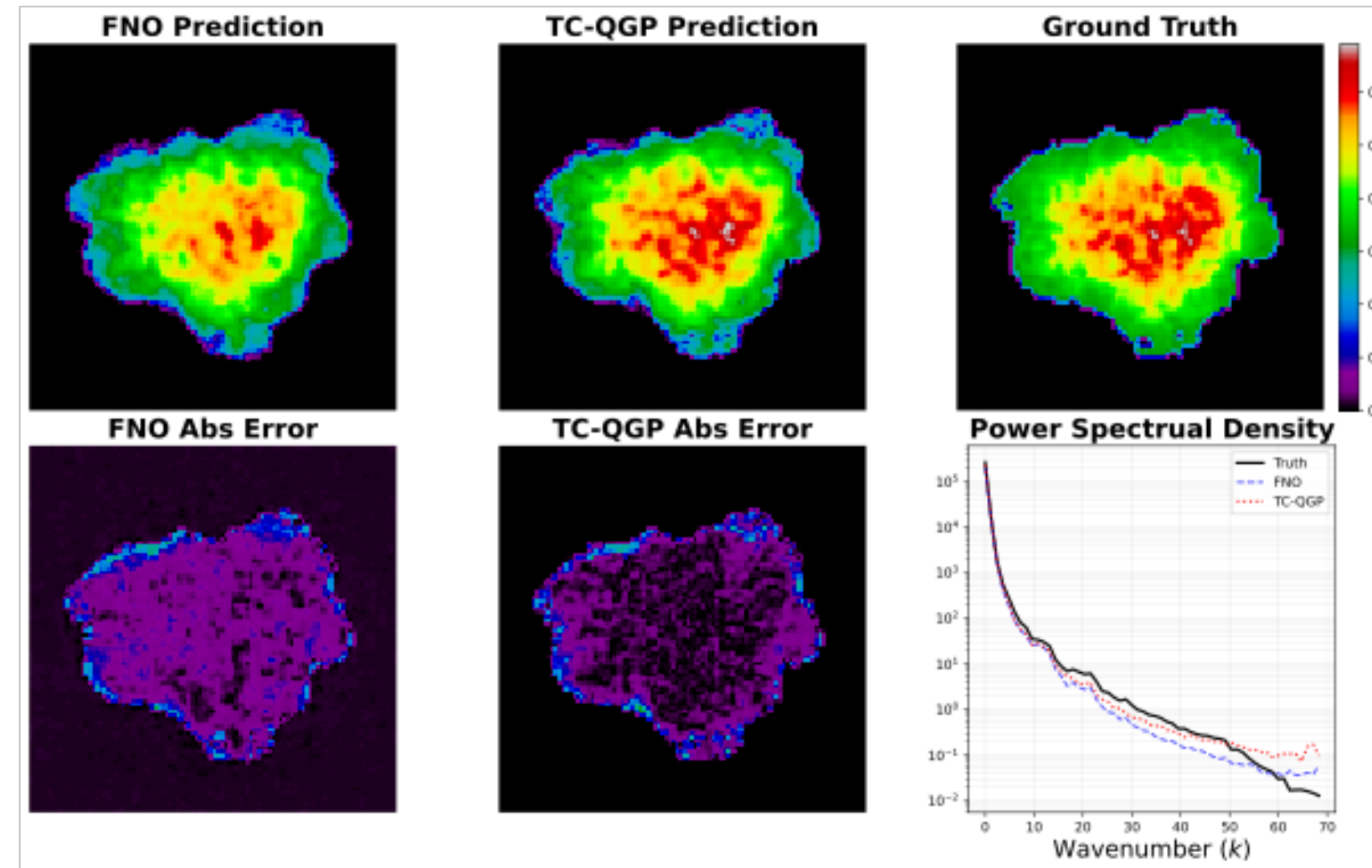
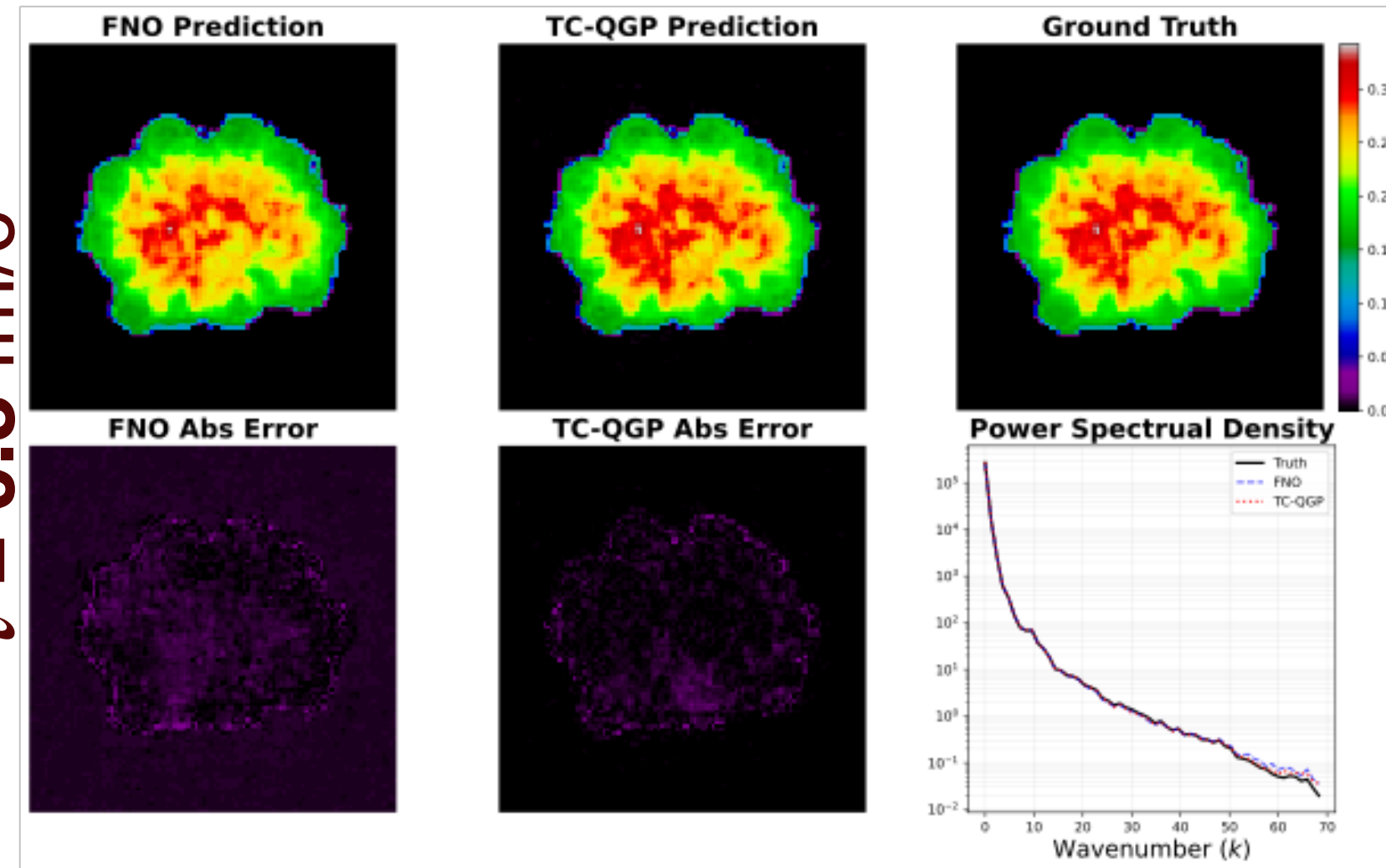
00D (2+1D) Temperature: 0.8, 9.6 fm/c

In-distribution

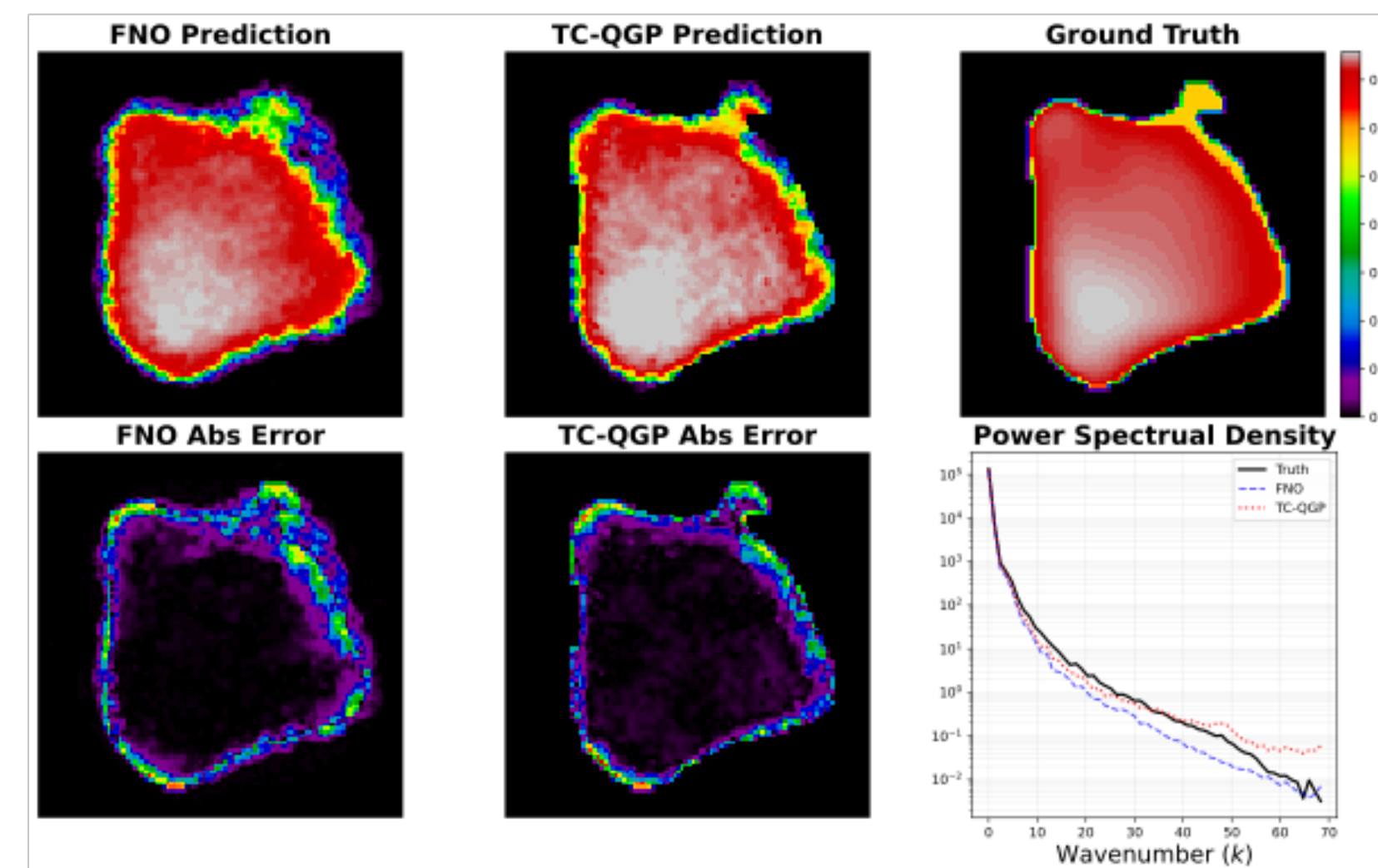
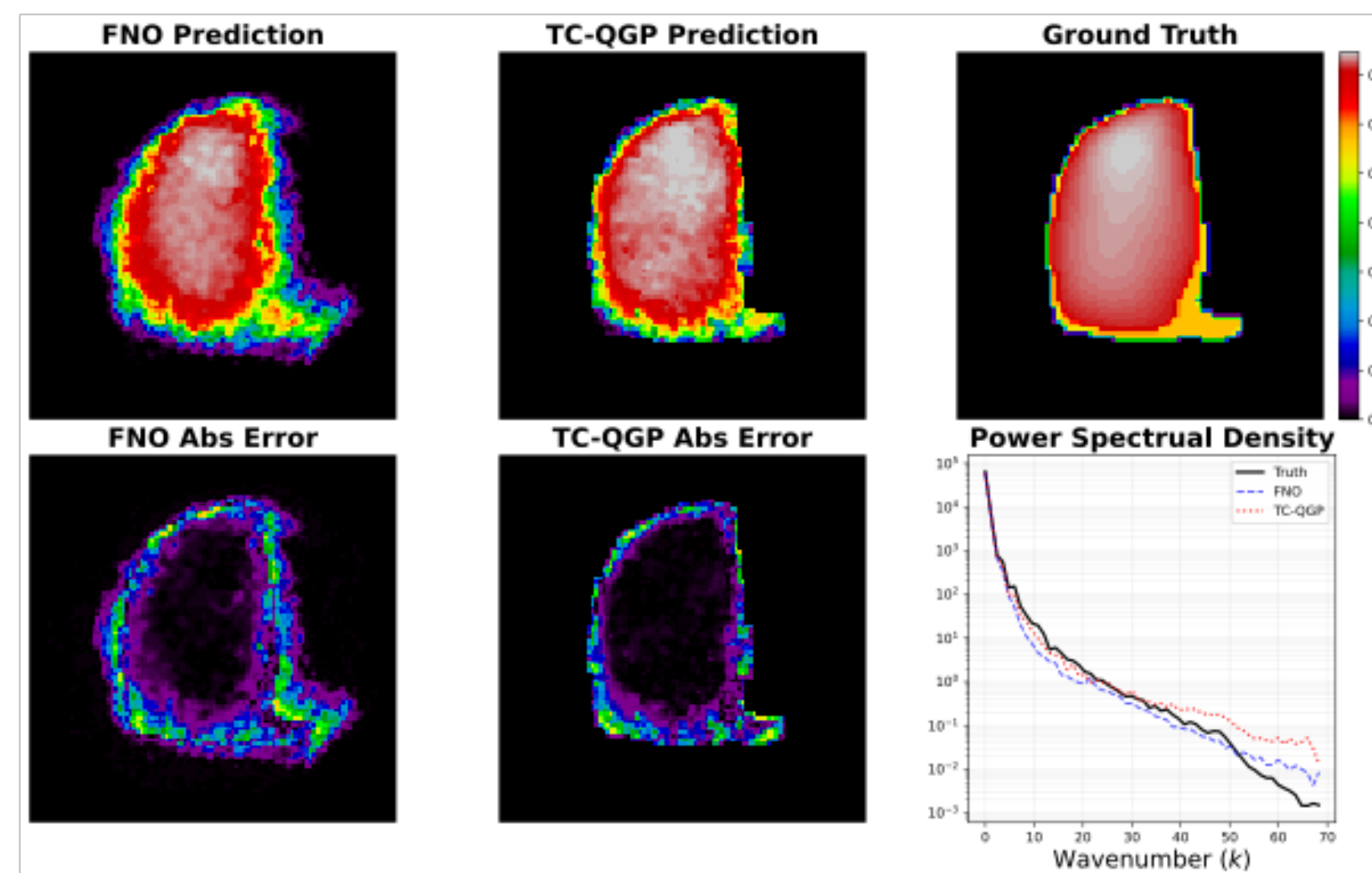
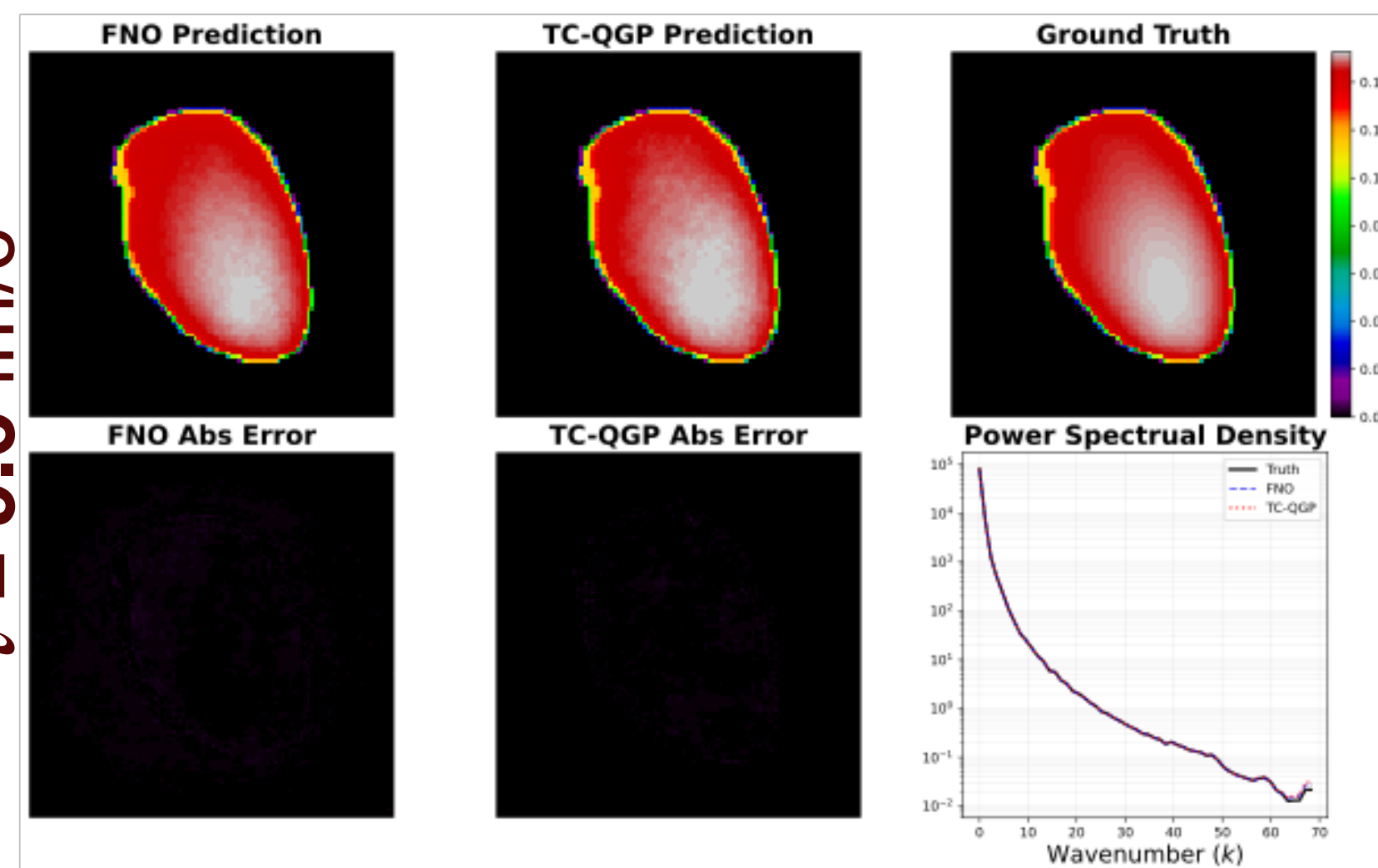
Out-of-distribution (OOD)-Sharp

Out-of-distribution (OOD)-Smooth

$\tau = 0.8 \text{ fm/c}$



$\tau = 9.6 \text{ fm/c}$



2+1D: FNO vs TC-QGP vs MUSIC

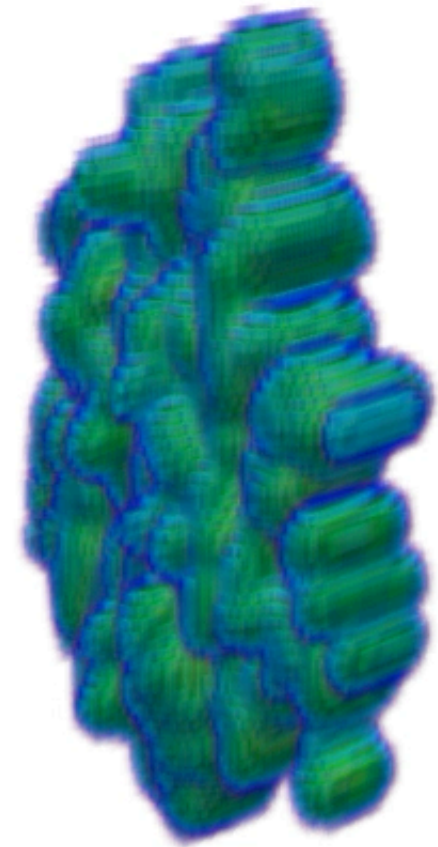
	2+1D ID		2+1D OOD-Sharp		2+1D OOD-Smooth	
	FNO	TC-QGP	FNO	TC-QGP	FNO	TC-QGP
<i>Mean Relative Error (%)</i>						
ε	0.59	0.59	7.61	7.51	6.51	6.47
T	1.25	0.79	6.40	5.58	5.54	4.98
P	1.40	0.58	8.76	8.41	7.36	7.20
u_x	1.17	1.12	11.93	11.67	10.28	10.10
u_y	1.09	1.21	13.04	13.00	11.30	11.32
u_z	–	–	–	–	–	–
All	1.10	0.86	9.55	9.23	8.20	8.01
<i>Computational Cost (single A100, batch size 1)</i>						
Parameters	FNO: 151.0M		TC-QGP: 21.4M			
GPU Memory (MB)	FNO: 1,193		TC-QGP: 127.2			
Inference Time (ms)	FNO: 34.4		TC-QGP: 29.7			

- **TC-QGP** outperforms FNO in fidelity sub-percent level accuracy
- **TC-QGP** ~7× fewer parameters, ~9× less memory with comparable inference time

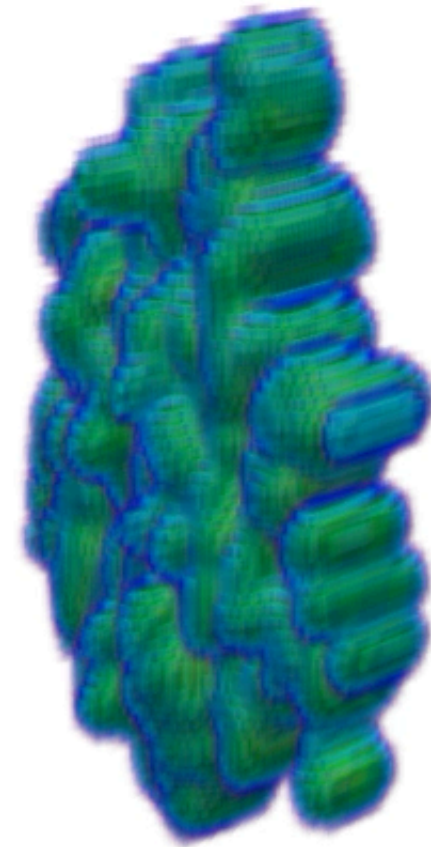
MUSIC: ~10 mins per event → TC-QGP: ~0.03 s per event: O(10,000) speedup !

3 + 1D Full Evolution: Temperature

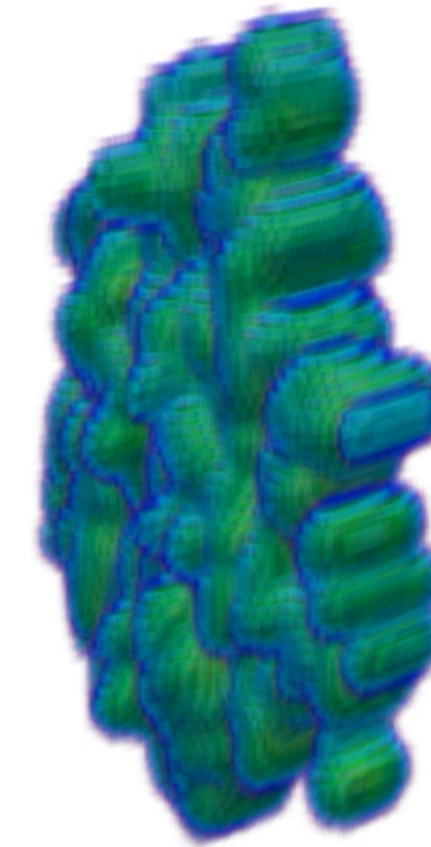
FNO



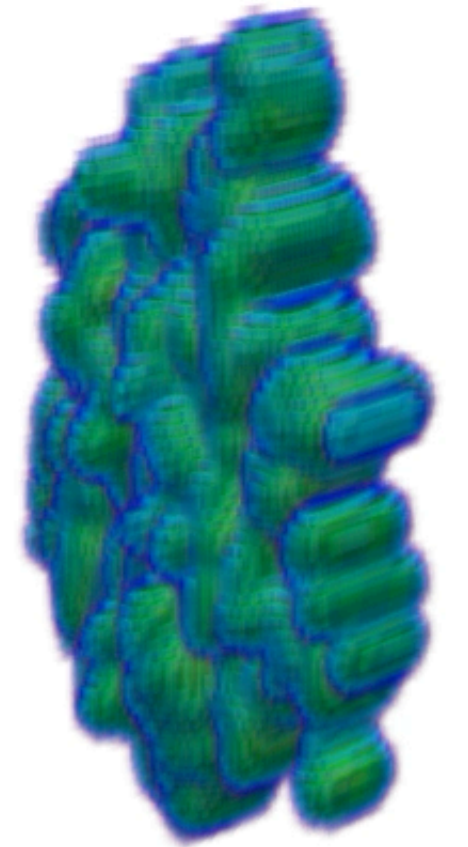
TC-QGP w/o GSM



TC-QGP



Ground Truth

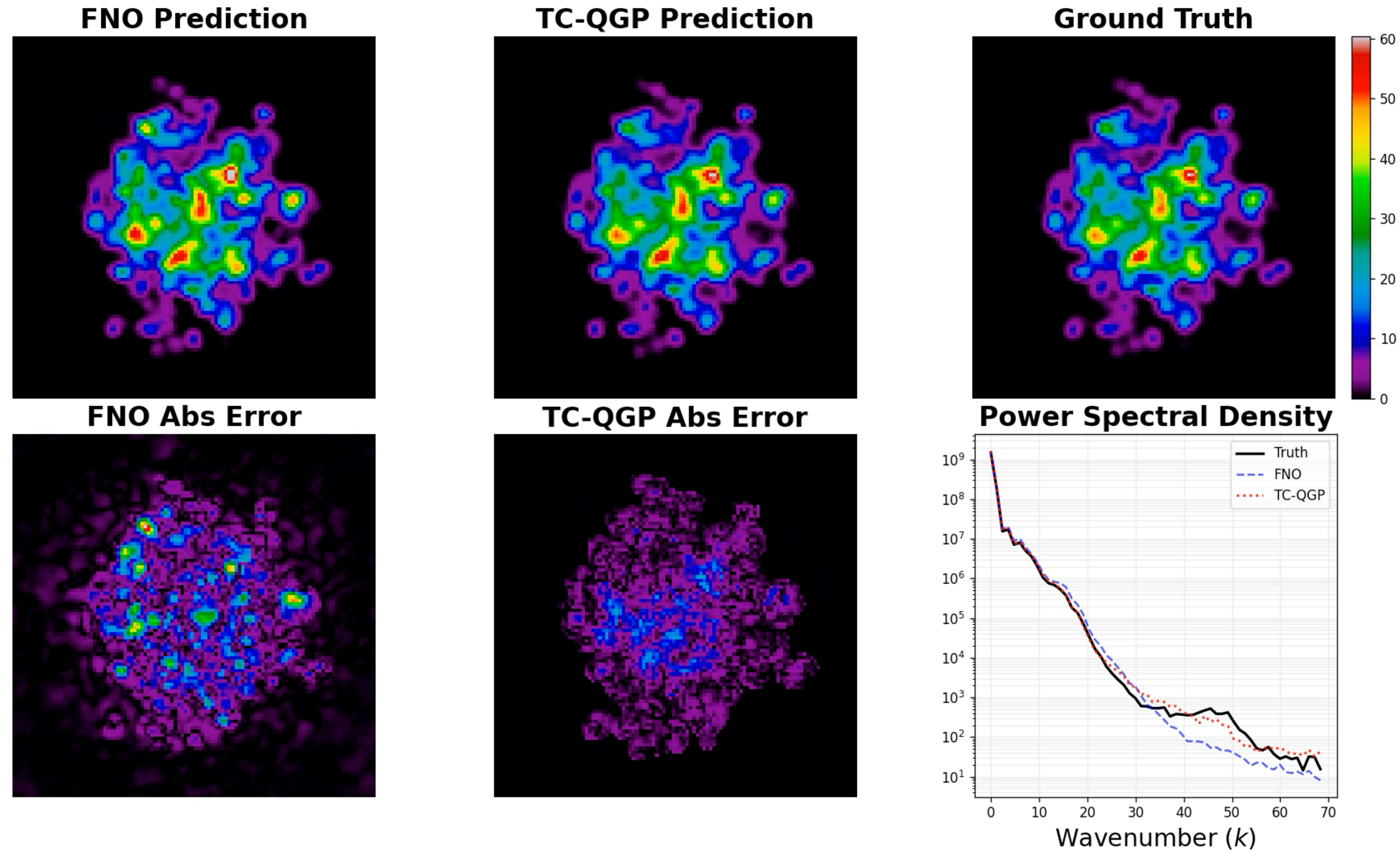


- Ablation: *TC-QGP without Gated Spatial Masking (GSM)* partially outperforms FNO, but the boundary remains blurry

Result: 3+1D Energy Density

- At mid-rapidity

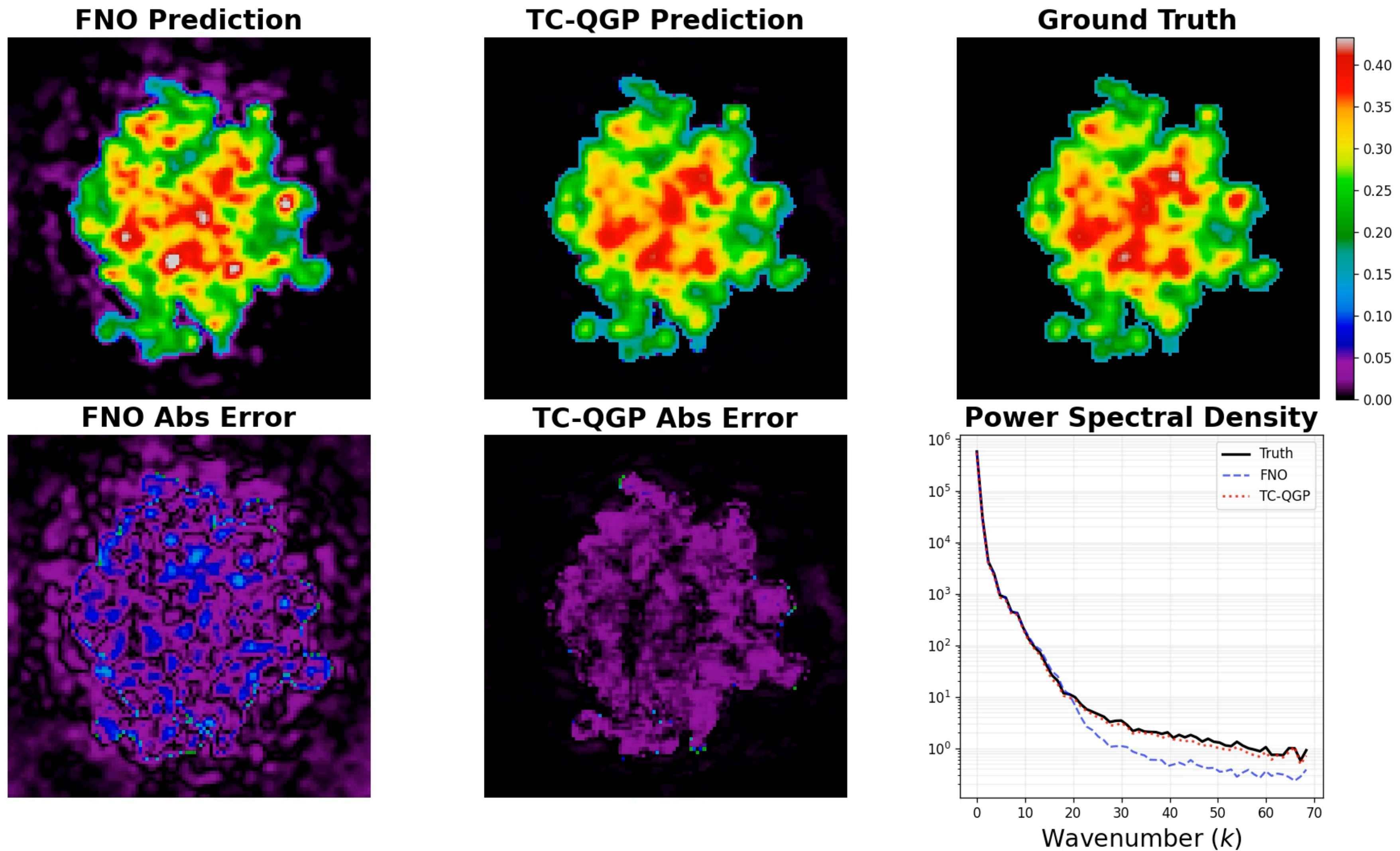
$\tau = 0.60 \text{ fm}/c$



Result: 3 + 1D Temperature

- At mid-rapidity

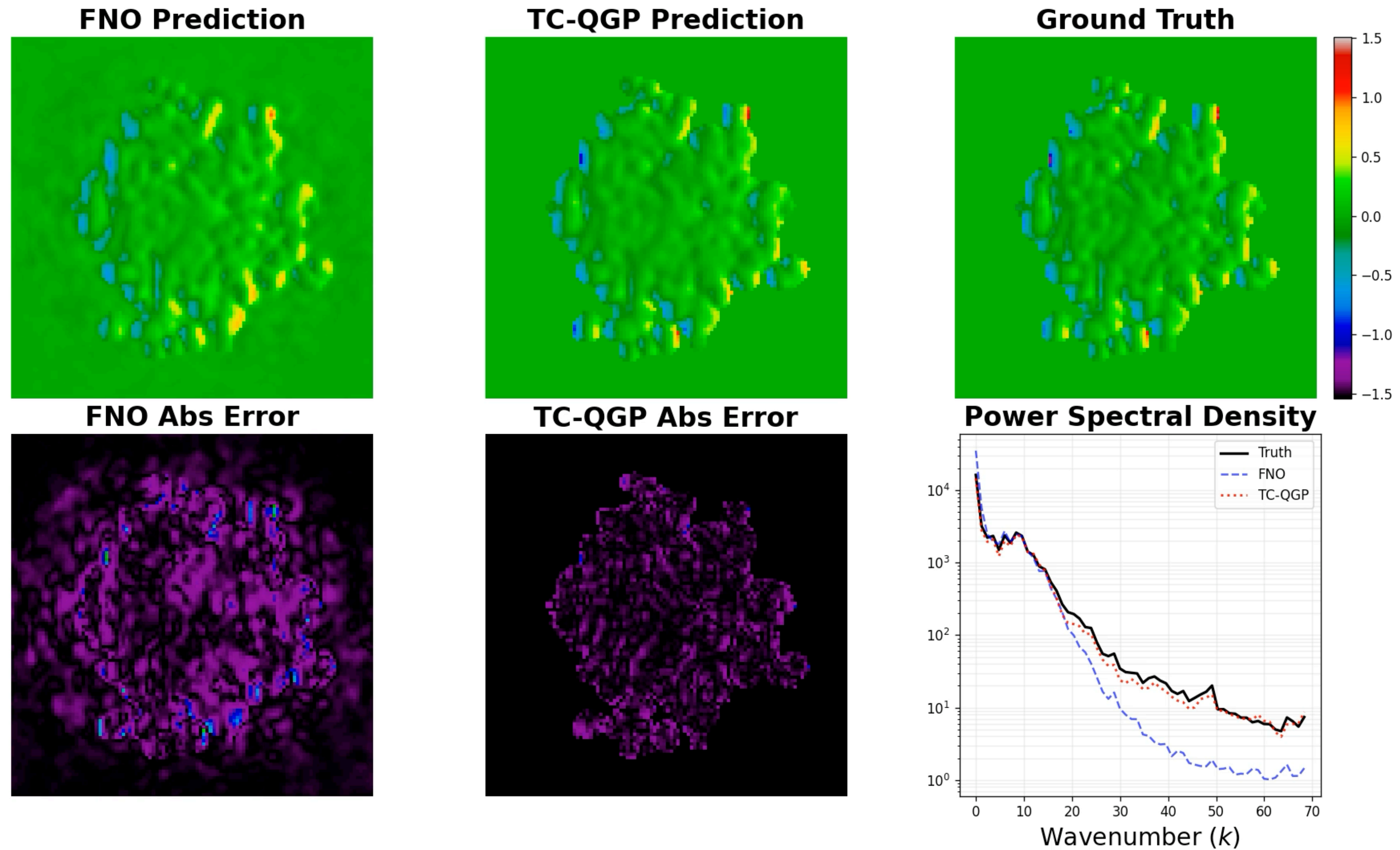
$\tau = 0.60 \text{ fm/c}$



Result: 3+1D Flow Velocity u_x

- At mid-rapidity

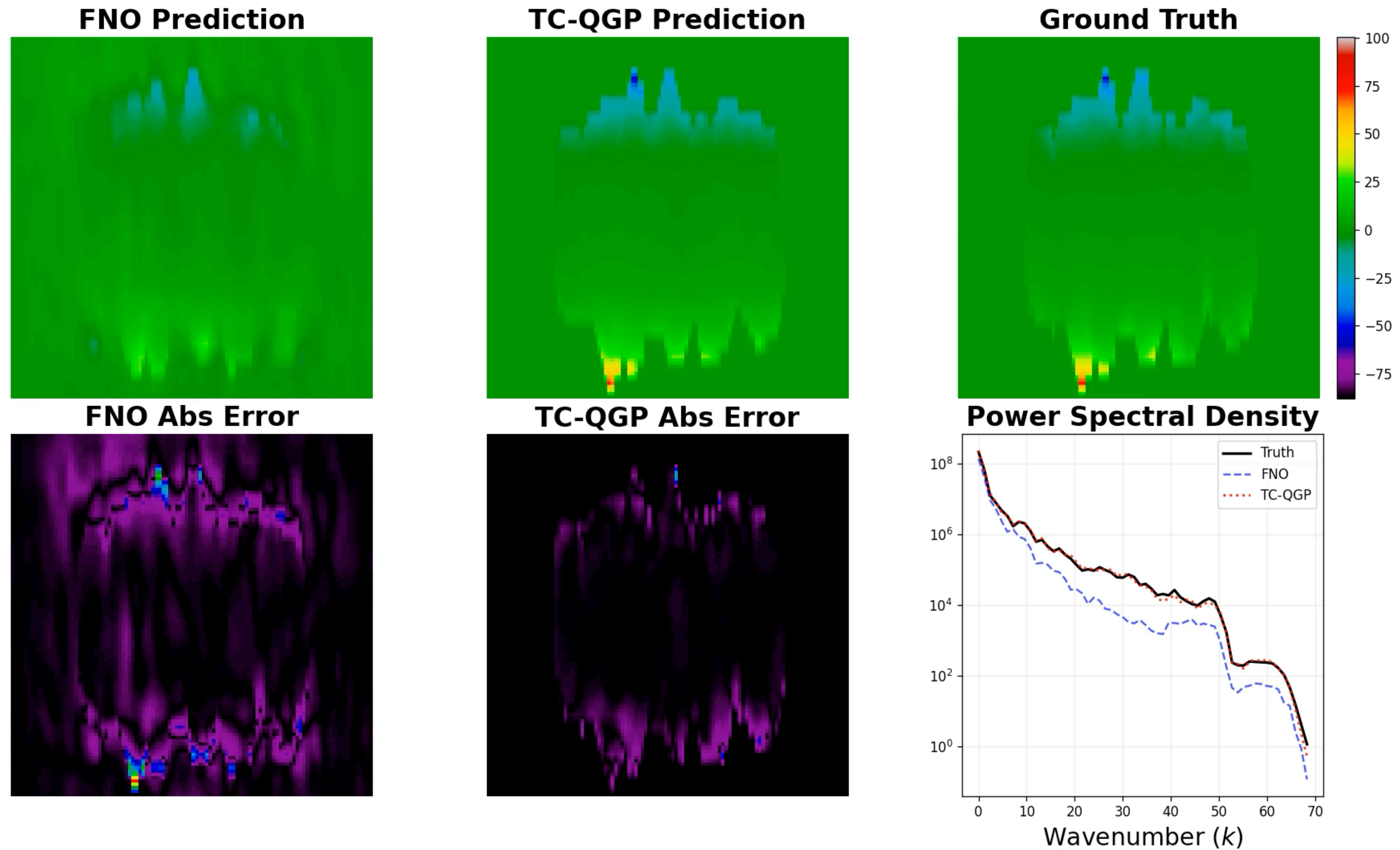
$$\tau = 0.60 \text{ fm/c}$$



Result: 3+1D Flow Velocity u_z

- At $y \sim 0$

$\tau = 0.60 \text{ fm}/c$



Result: 3+1D Summary

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 NVIDIA A100 GPU, 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

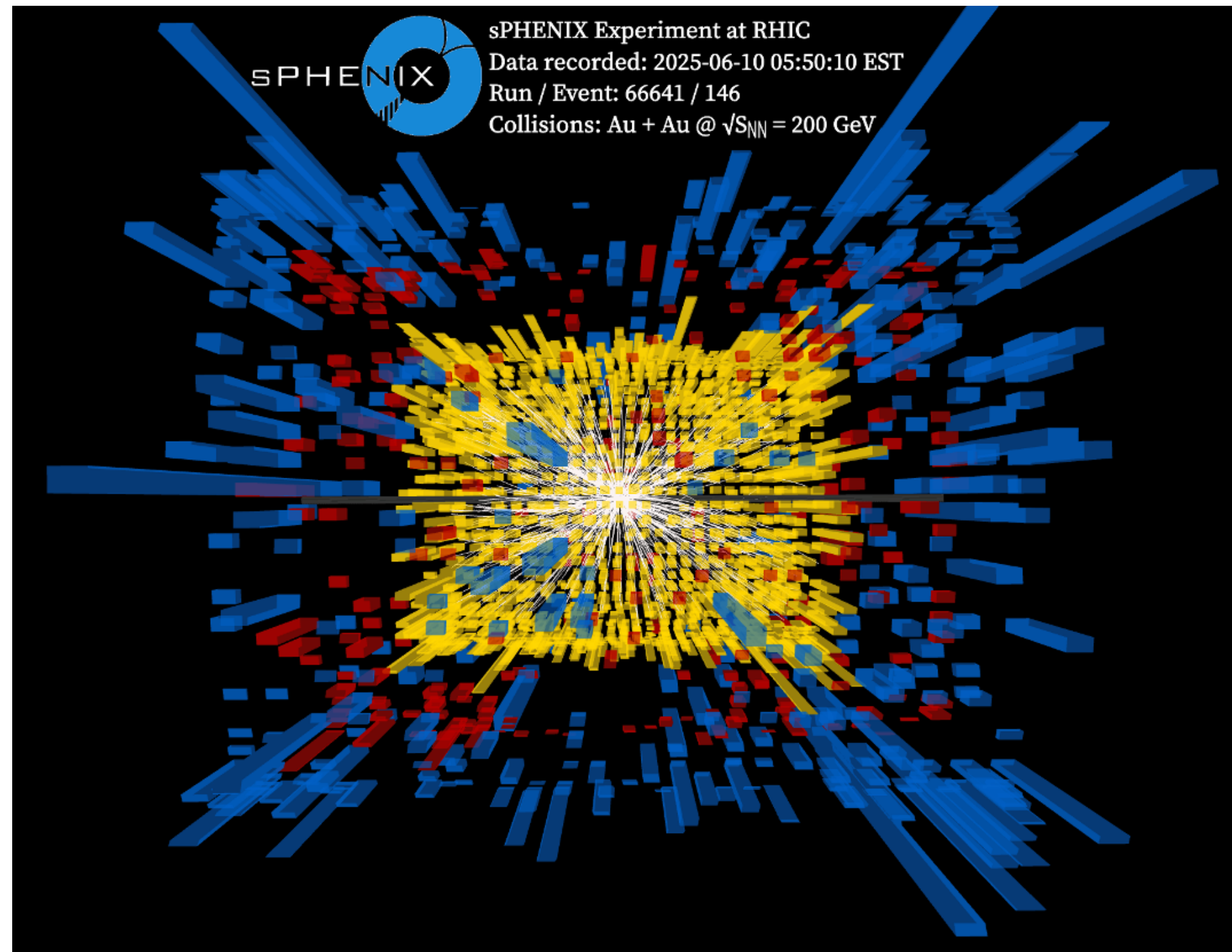
- TC-QGP substantially improves 3+1D prediction accuracy: ~5% overall MRE vs. ~23% for FNO
- TC-QGP is more memory efficient than standard FNO

MUSIC: ~20 hours per event → TC-QGP: ~0.2 s per event: O(100,000) speedup !

High Energy Collision Data Taking

Collision rates $\sim 80\text{kHz}$

Raw data: $O(\text{PB/s}) \rightarrow 10^{15}$ bytes



Triggering

*Is it interesting event? $\rightarrow$ if yes, save!
(Real-time filtering)*

Streaming readout

*Let's save everything by **compressing** data!*

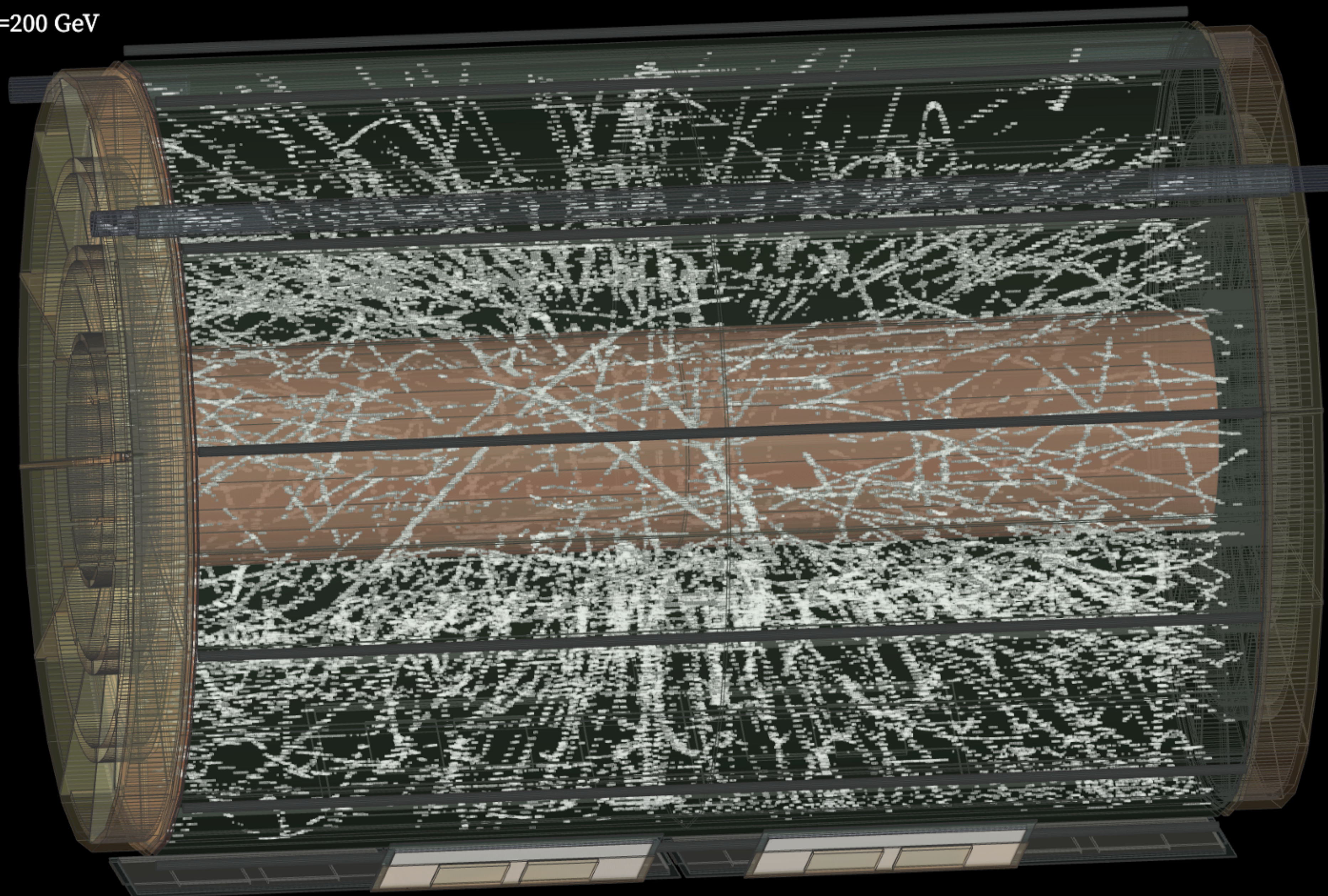


sPHENIX Time Projection Chamber

100 Hz ZDC, MBD Prescale: 2, HV: 4.45 kV GEM, 45 kV CM, X-ing Angle: 2 mrad

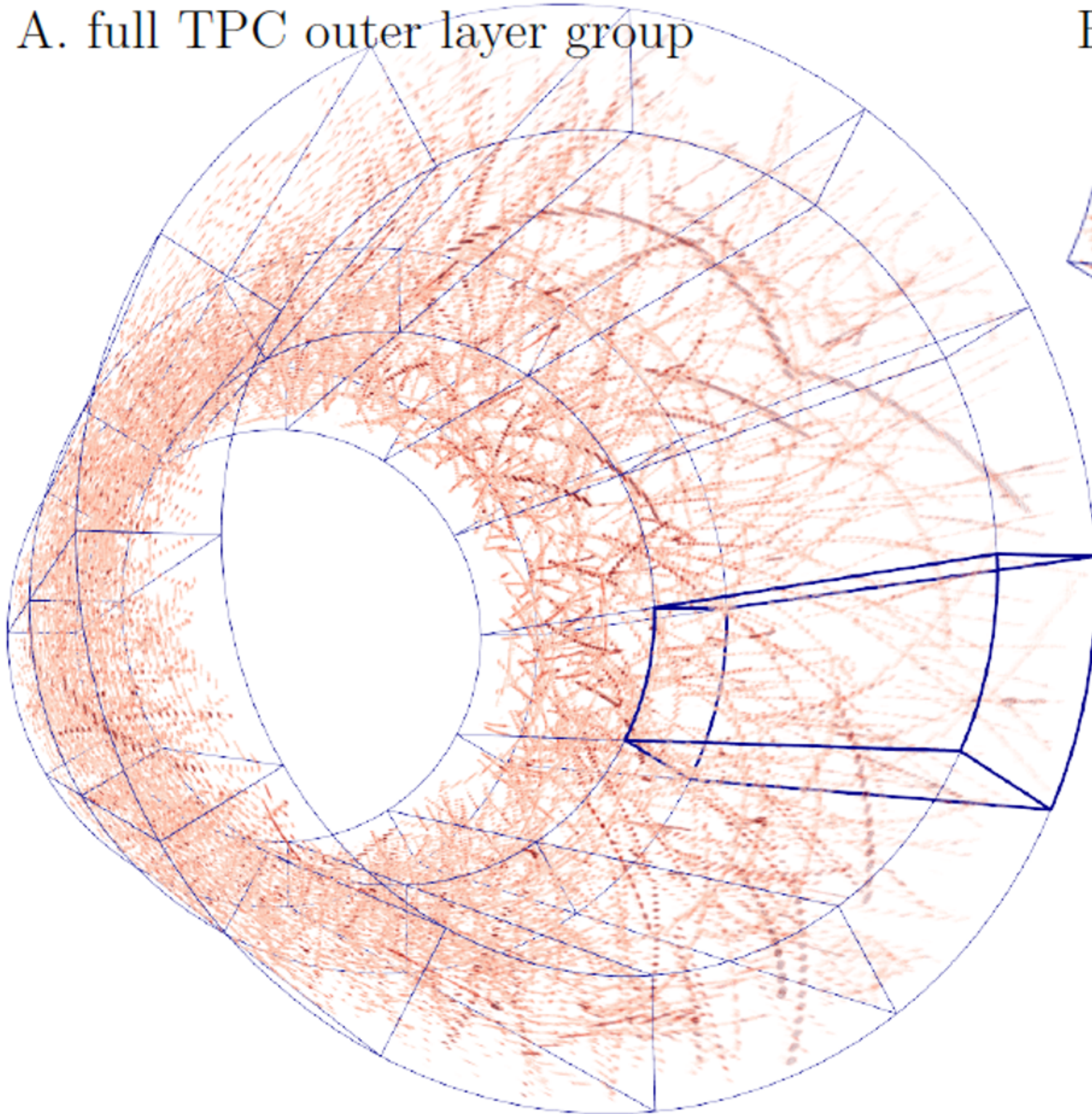
2023-06-23, Run 10931 - EBDC03 reference frame 89

Au+Au $\sqrt{s_{NN}}=200$ GeV

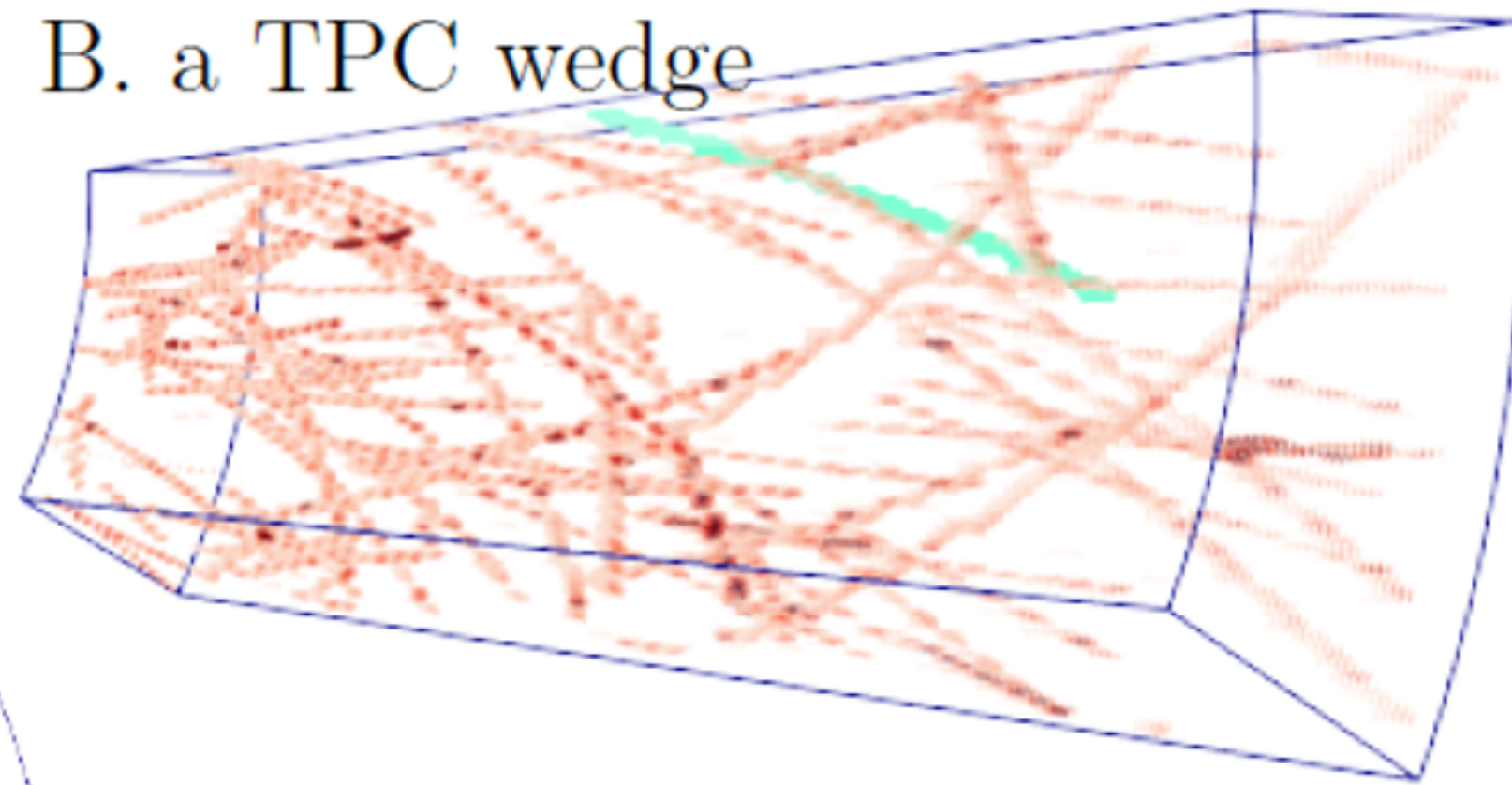


sPHENIX TPC Data for Heavy Ion Collisions

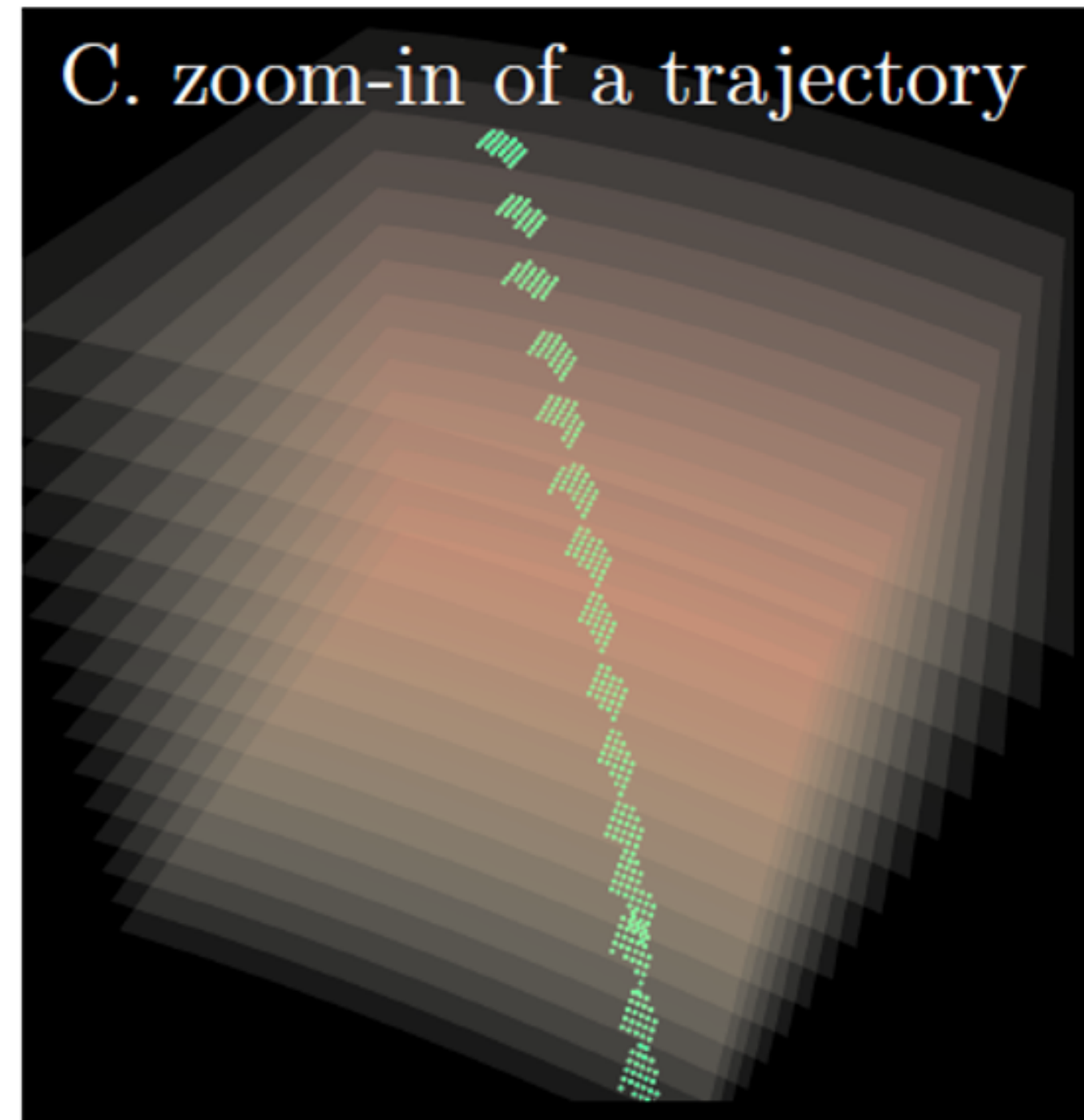
A. full TPC outer layer group



B. a TPC wedge



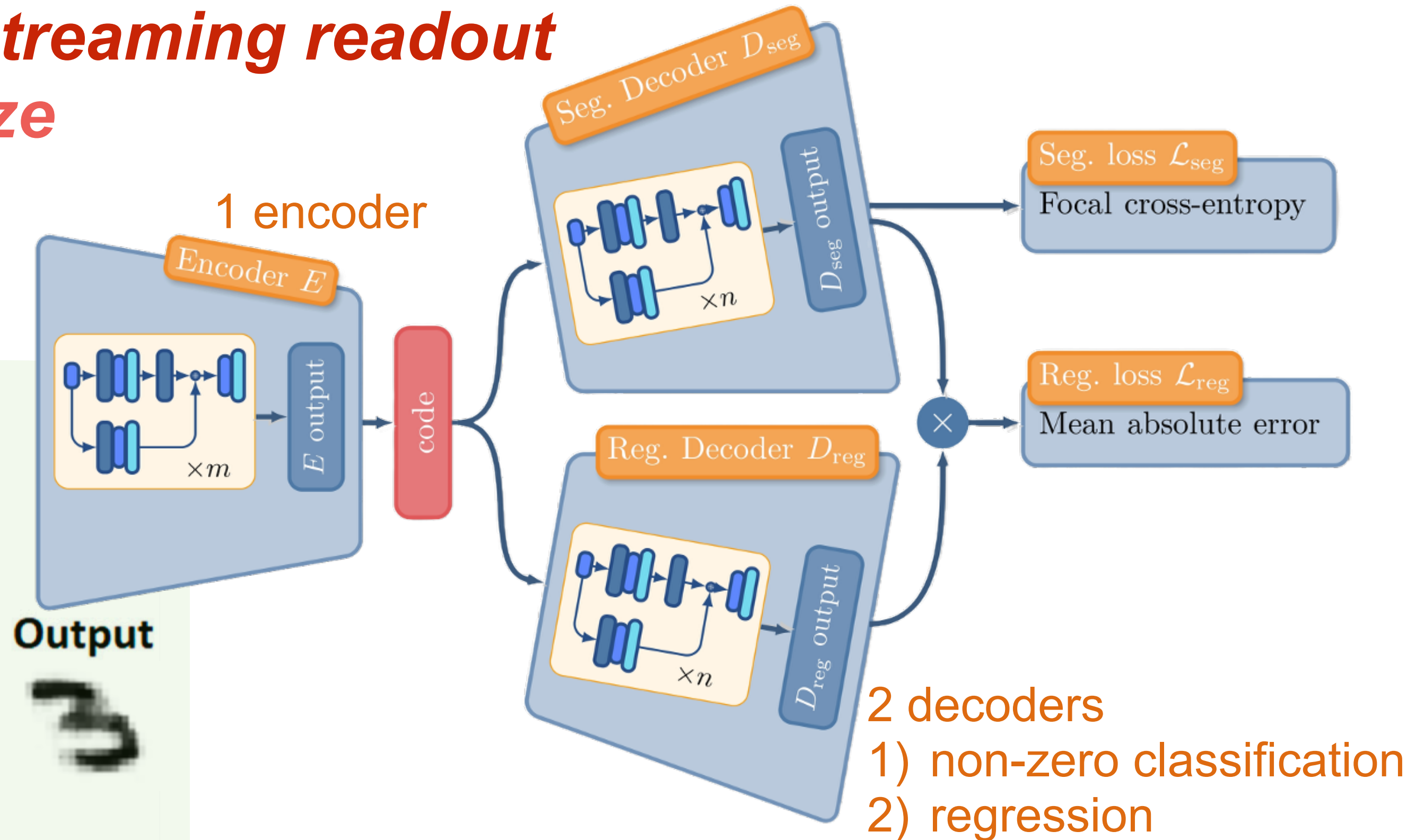
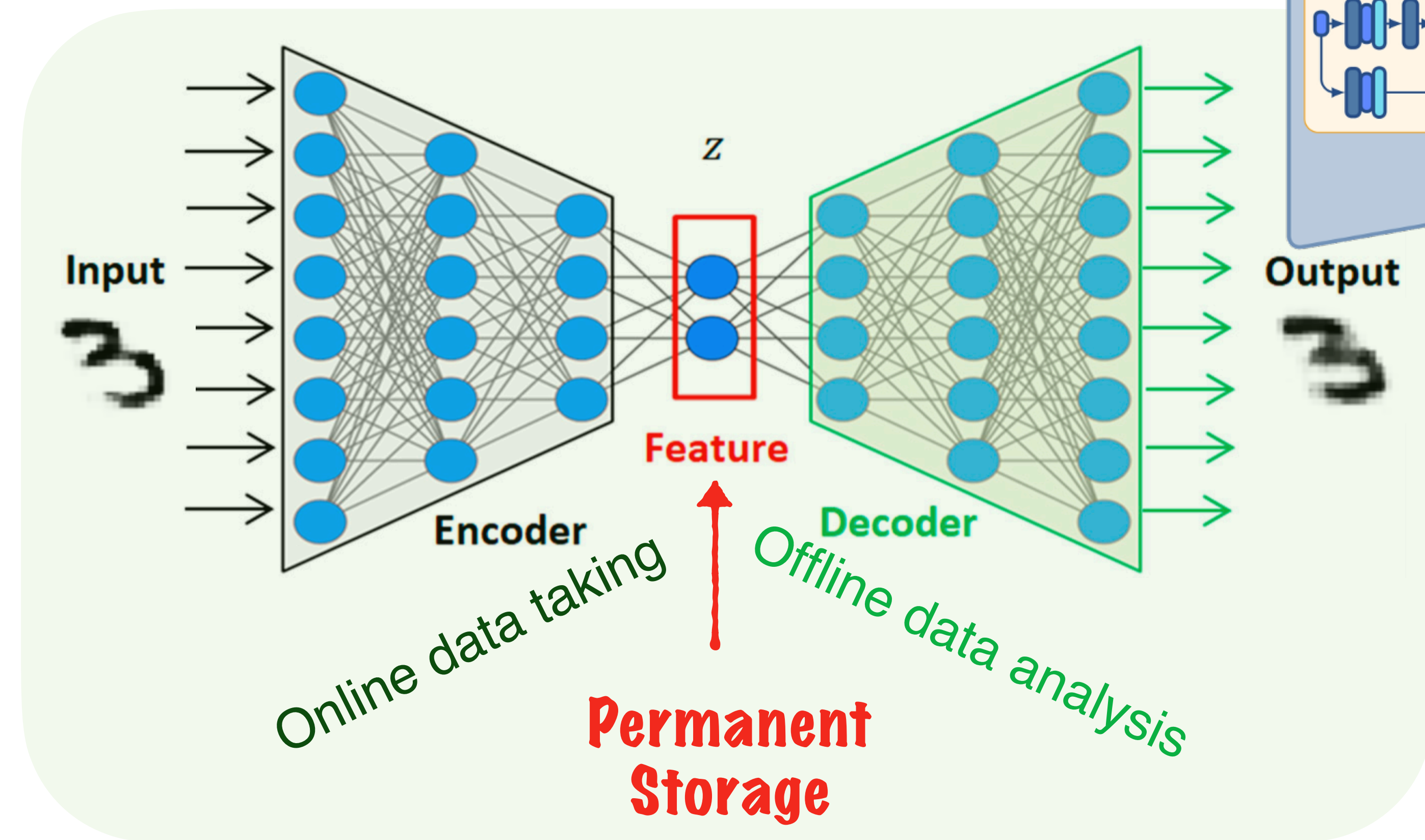
C. zoom-in of a trajectory



Data Compression using AutoEncoder

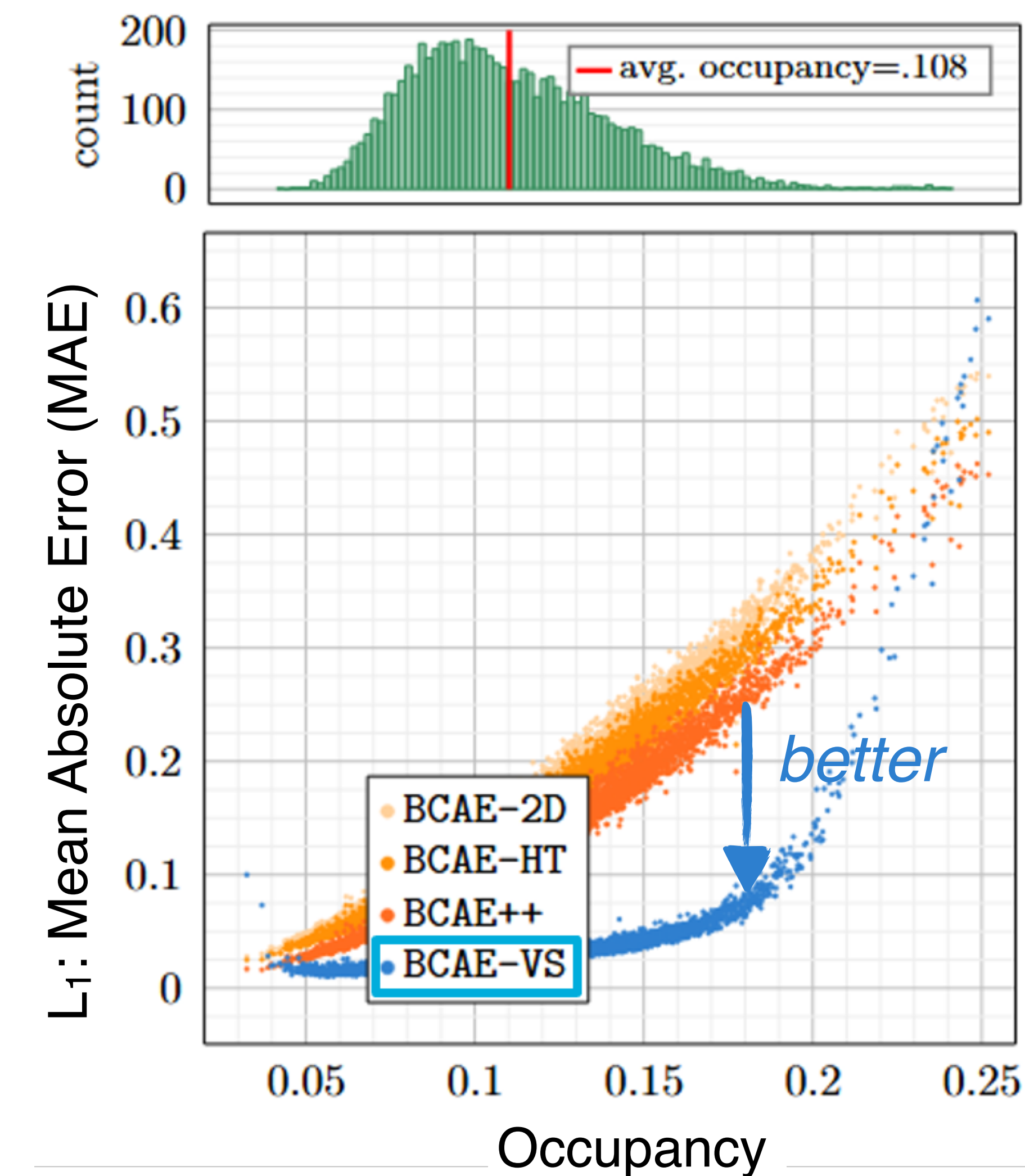
Next generation data taking strategy: **streaming readout** (no triggering) by **compressing data size**

→ record **encoded data** (small size) and **decode** for reconstruction

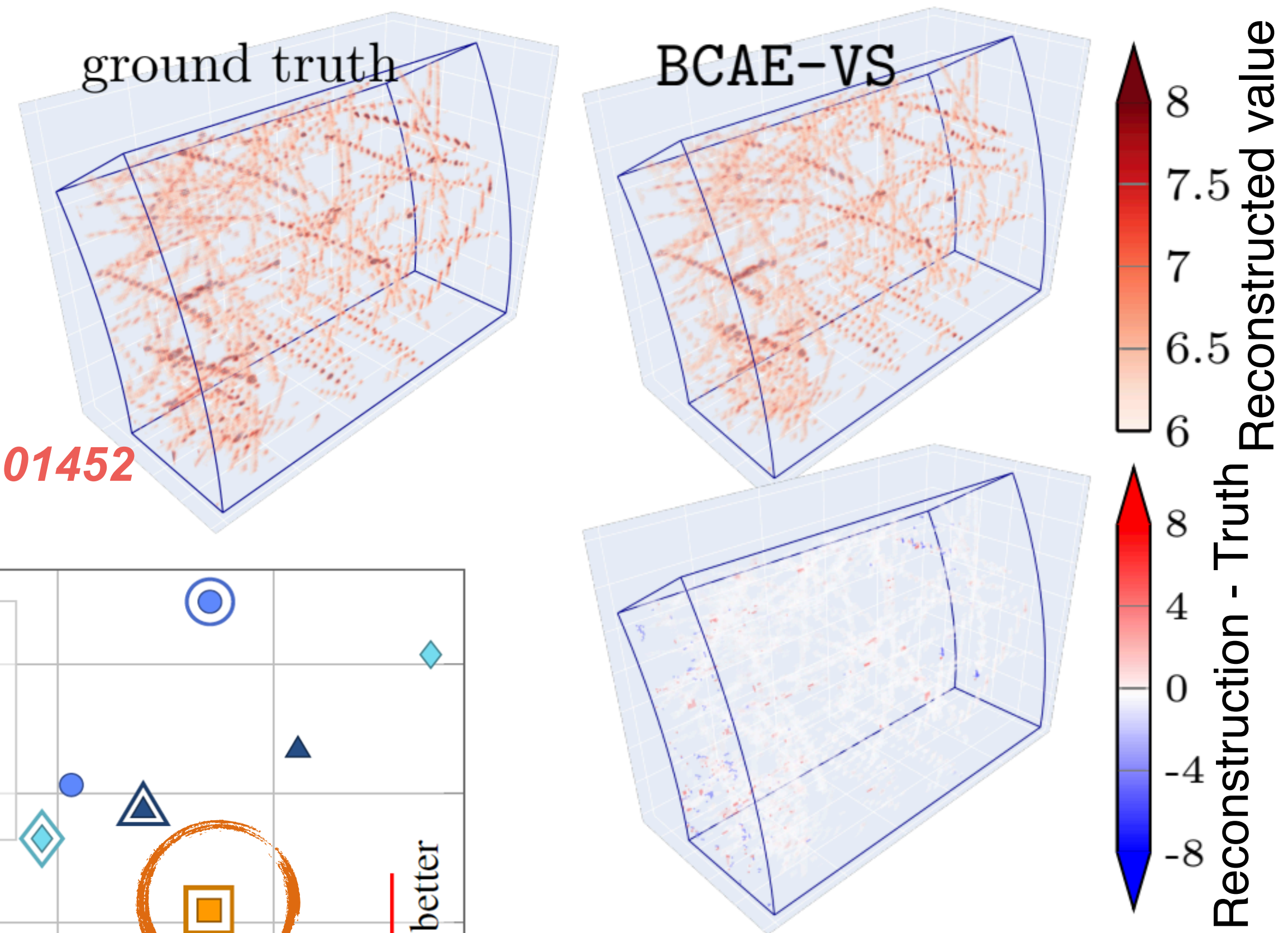
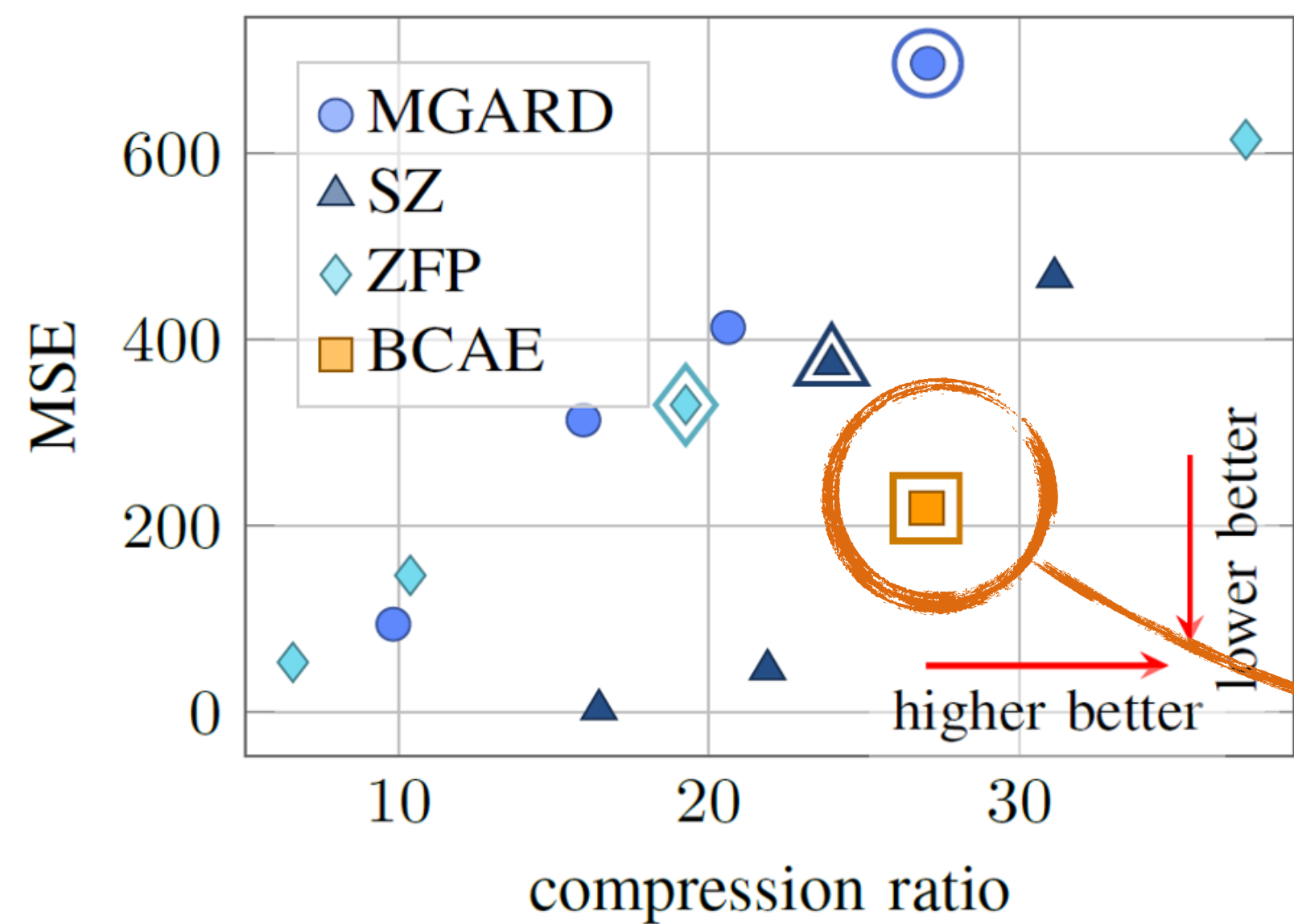


Bicephalous Convolutional Autoencoder (BCAE) model was developed to deal with zero-suppressed data (non-gaussian)

Data Compression using AutoEncoder

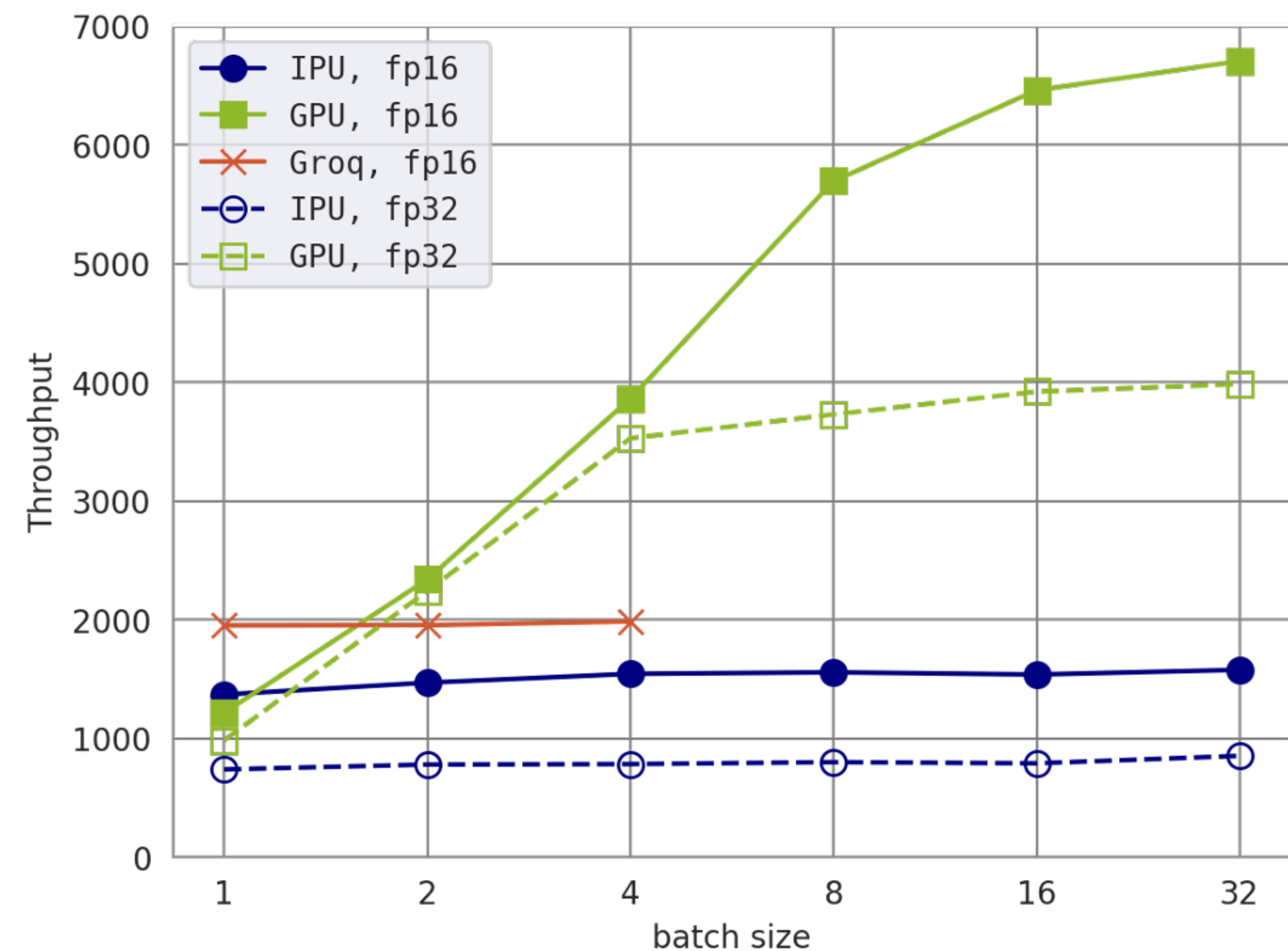
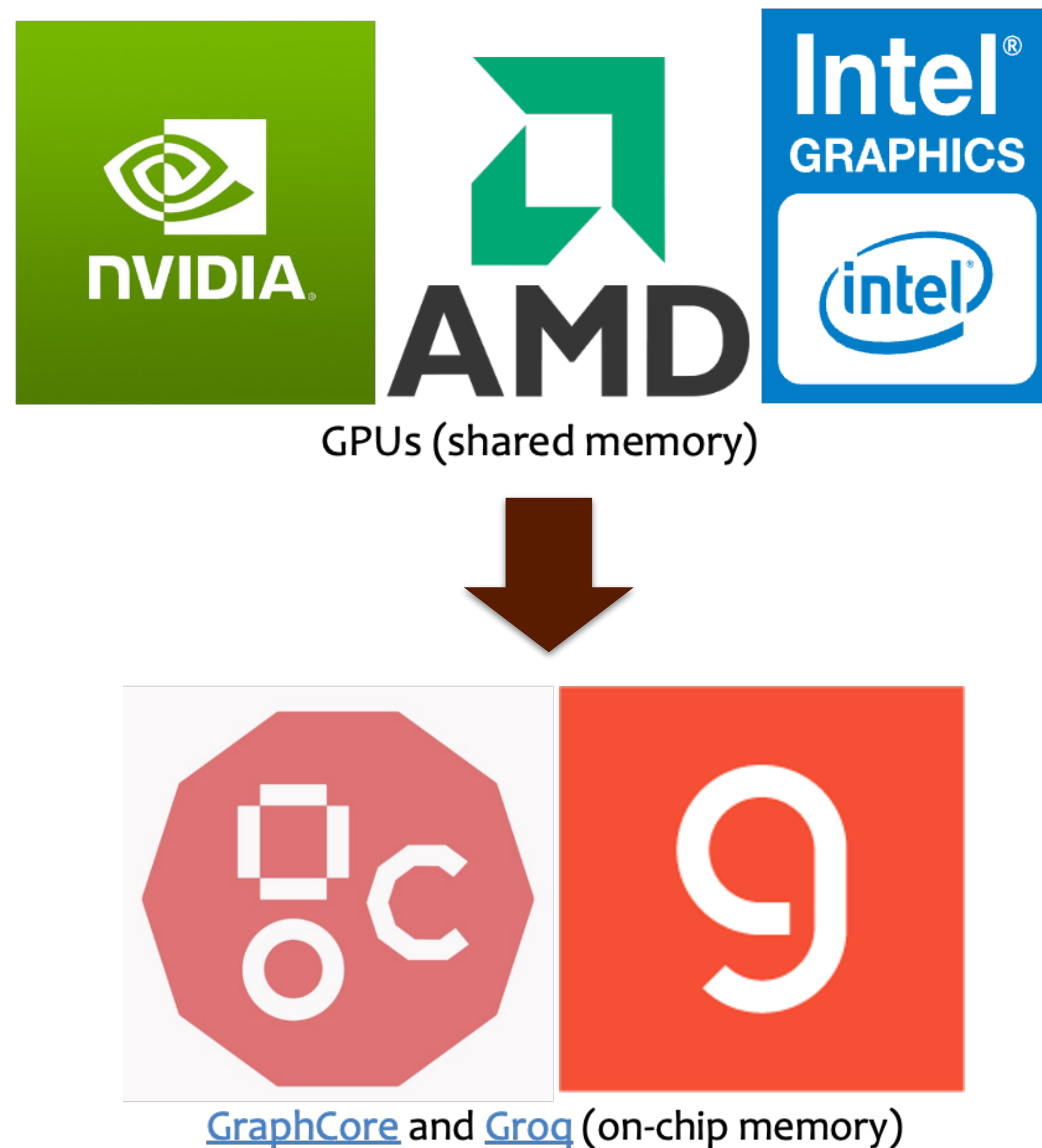


*Y.Go Patterns, 2026, 101452
arXiv:2411.11942*



BCAE-VS: Great performance for sparse data (great prospects towards EIC!)

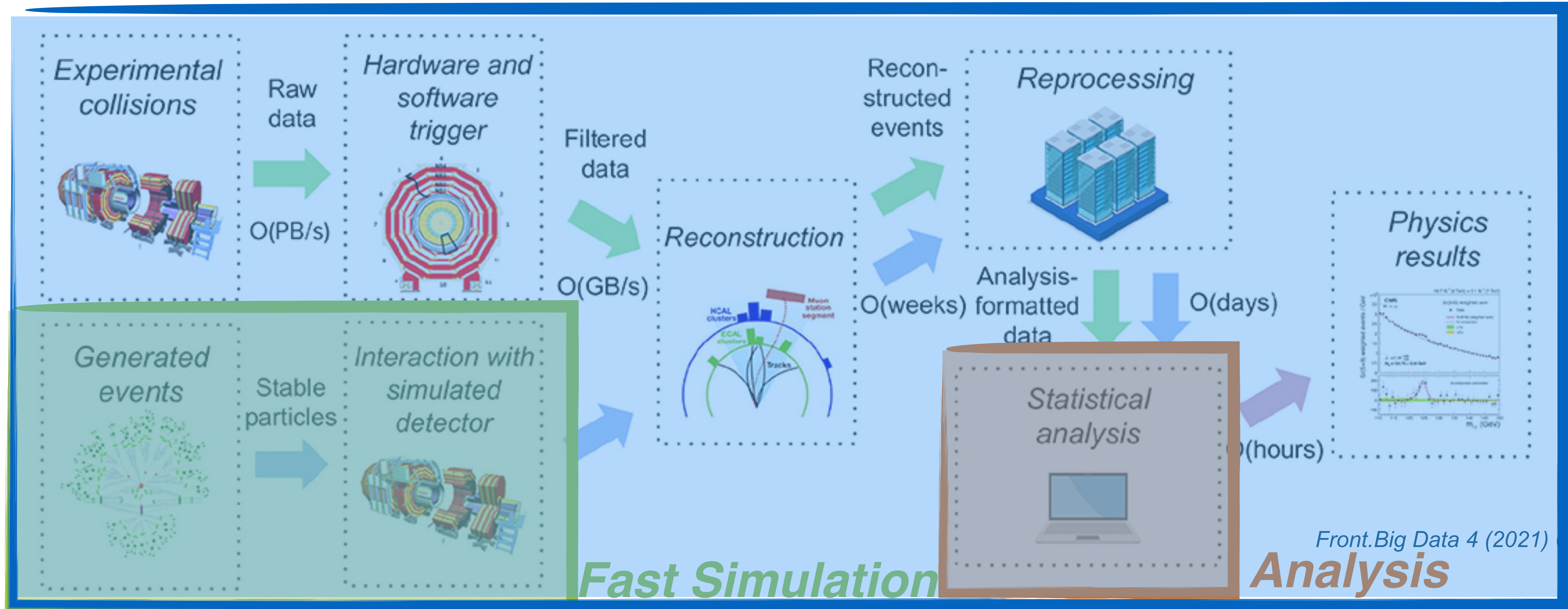
AI Accelerator Chips



- Faster compression using AI hardware instead of GPU
 - but harder to optimize AI hardwares → needs extensive development

AI-Driven Nuclear and Particle Physics

Detector design Experiment operation Data analysis Physics discovery



• Each stage of analysis requires substantial human effort, computing resources, and time

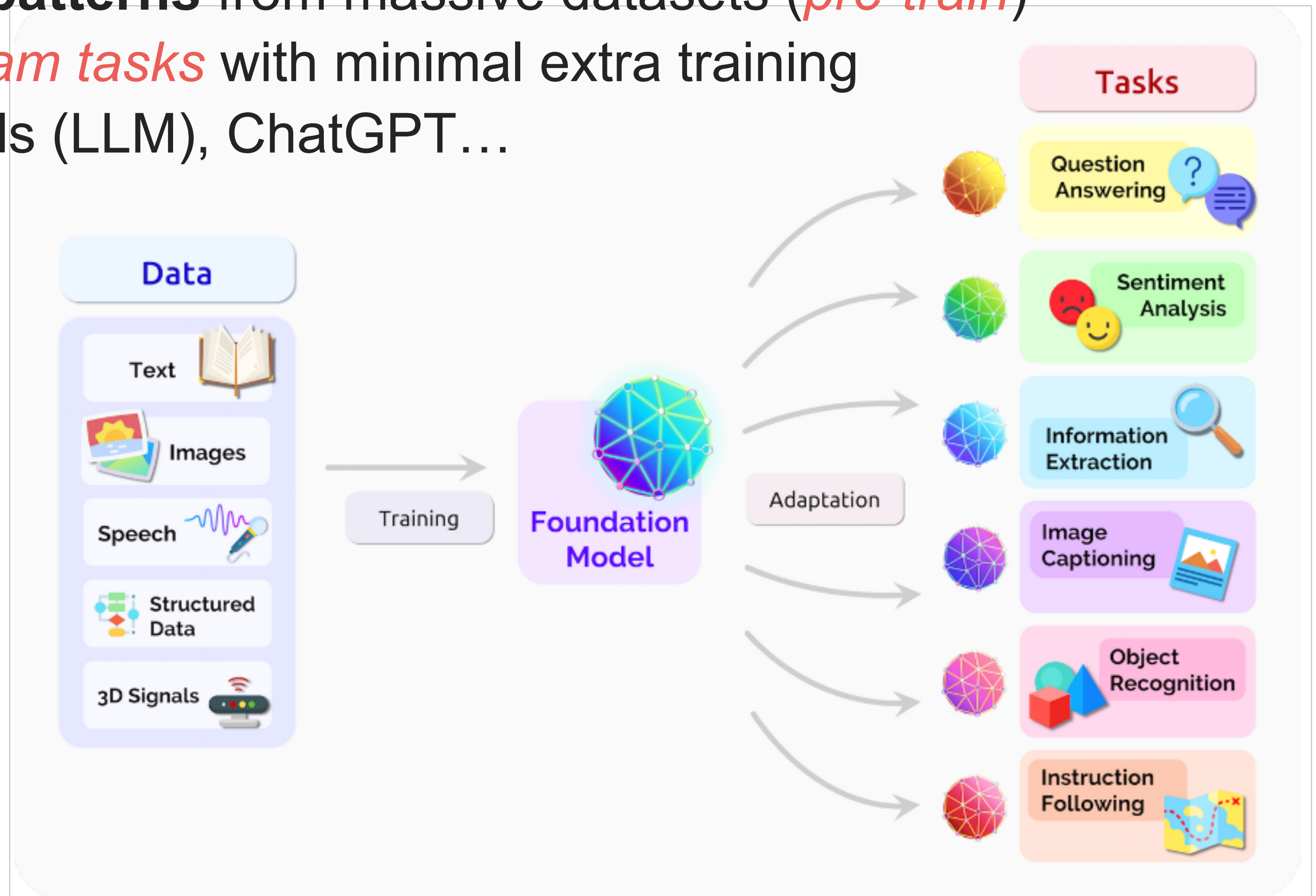
Foundation Model !

→ *AI/ML to accelerate, automate, and improve these processes*

Foundation Model

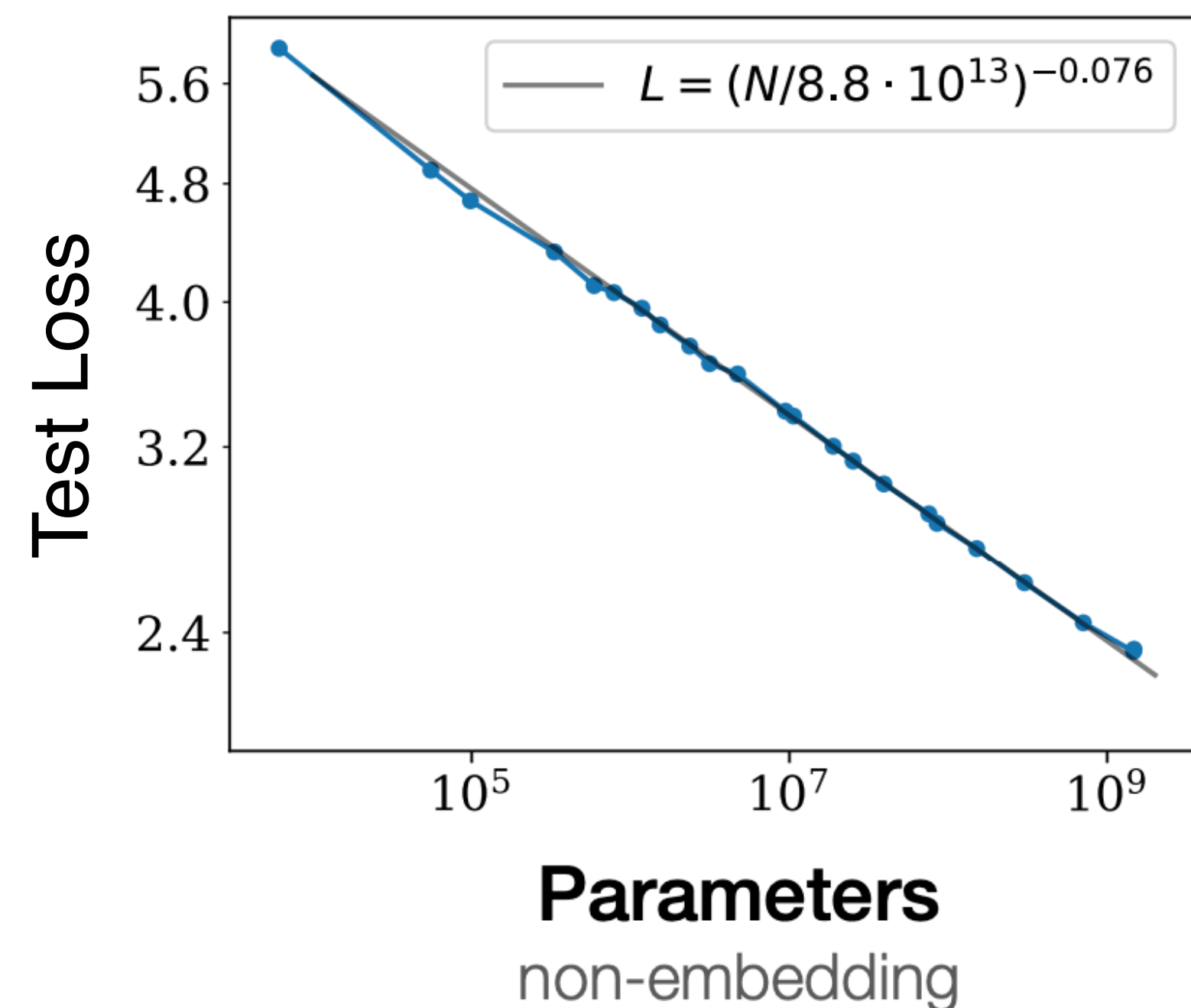
- **Foundation models**

- ➔ learn **broad features and patterns** from massive datasets (*pre-train*)
- ➔ adapted to many *downstream tasks* with minimal extra training
- ➔ e.g. Large Language Models (LLM), ChatGPT...

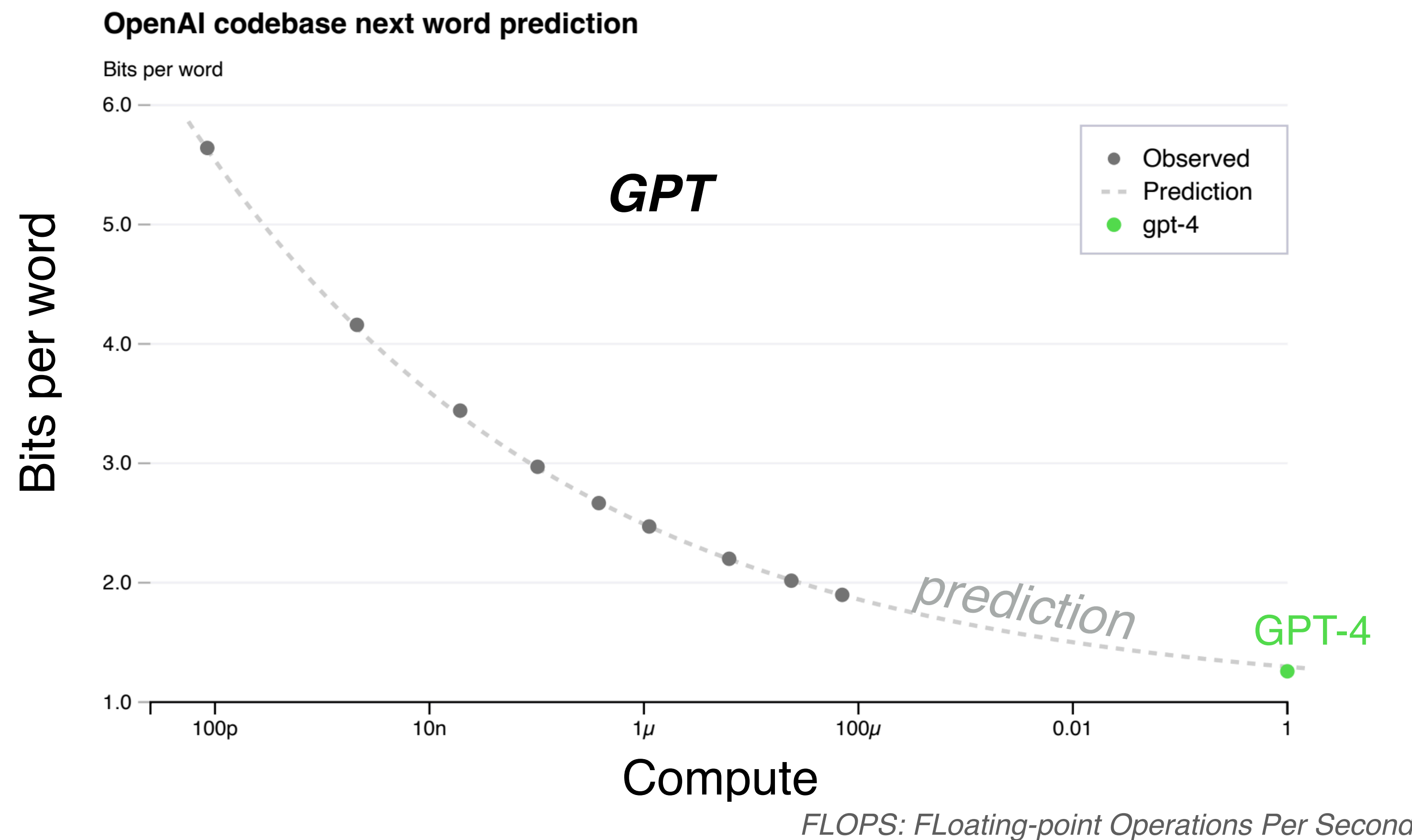


Neural Scaling Laws

- Large language model studies established that performance improves predictably with model size, dataset size, and compute, often following **power law scaling**



“Scaling Laws for Neural Language Models”
arXiv:2001.08361

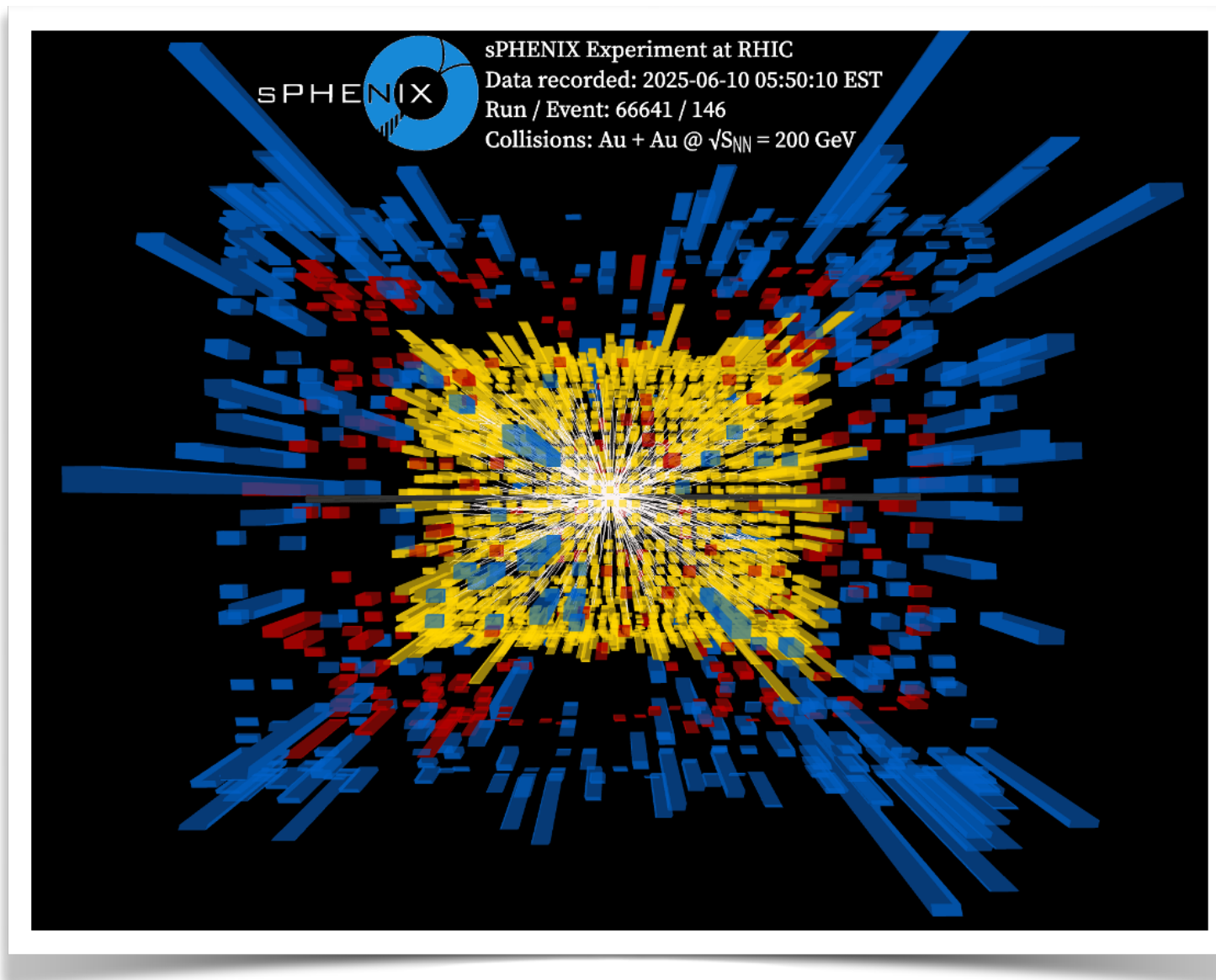


Achiam, Josh, et al. "Gpt-4 technical report." arXiv:2303.08774 (2023)

Foundation Model for Nuclear and Particle Physics

Various Downstream Tasks

*Large amount of unlabeled data available
in high energy nuclear and particle physics
e.g. sPHENIX data 300 Petabytes*



Pre-training



FM4
NPP

**minimal
training**



track finding
particle identification
noise removal
global reconstruction
jet tagging
event classification
anomaly detection

⋮

AI-driven discovery

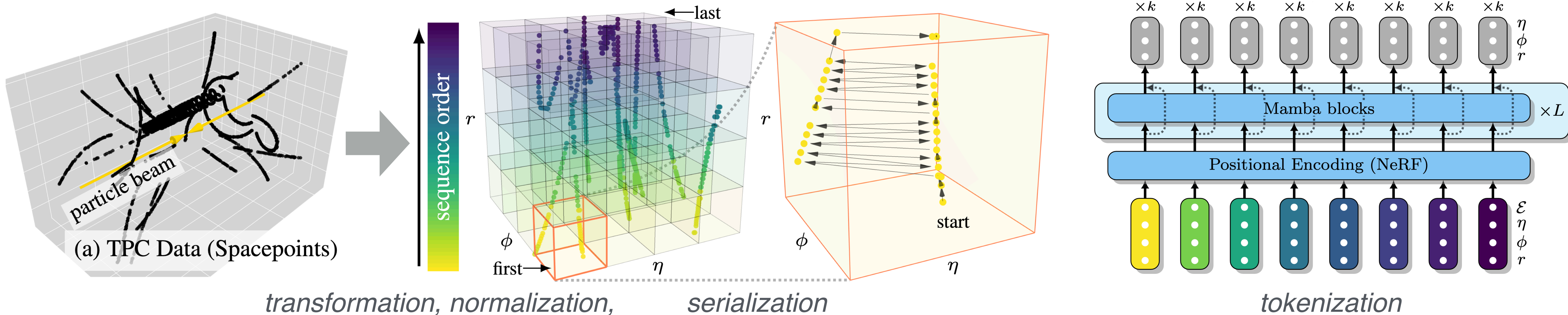
- **FM4NPP**: a **self-supervised** foundation model with inputs on unlabeled sparse detector-level data
- lightweight **Adapter** for multiple downstream tasks

Challenges and Goals for FM4NPP

- Low-level physics data are sparse **3D points**, not language, not image.
How to formulate a self-supervised learning task?
- What neural network **architecture** to use?
- Does performance scale with **model size** and **dataset size**?
- Do the learned **representations transfer to downstream tasks**?
- What is the **computational cost**? (c.f. GPT-3 costs ~ \$4.6M)
- The list goes on and on ...

FM4NPP: Training

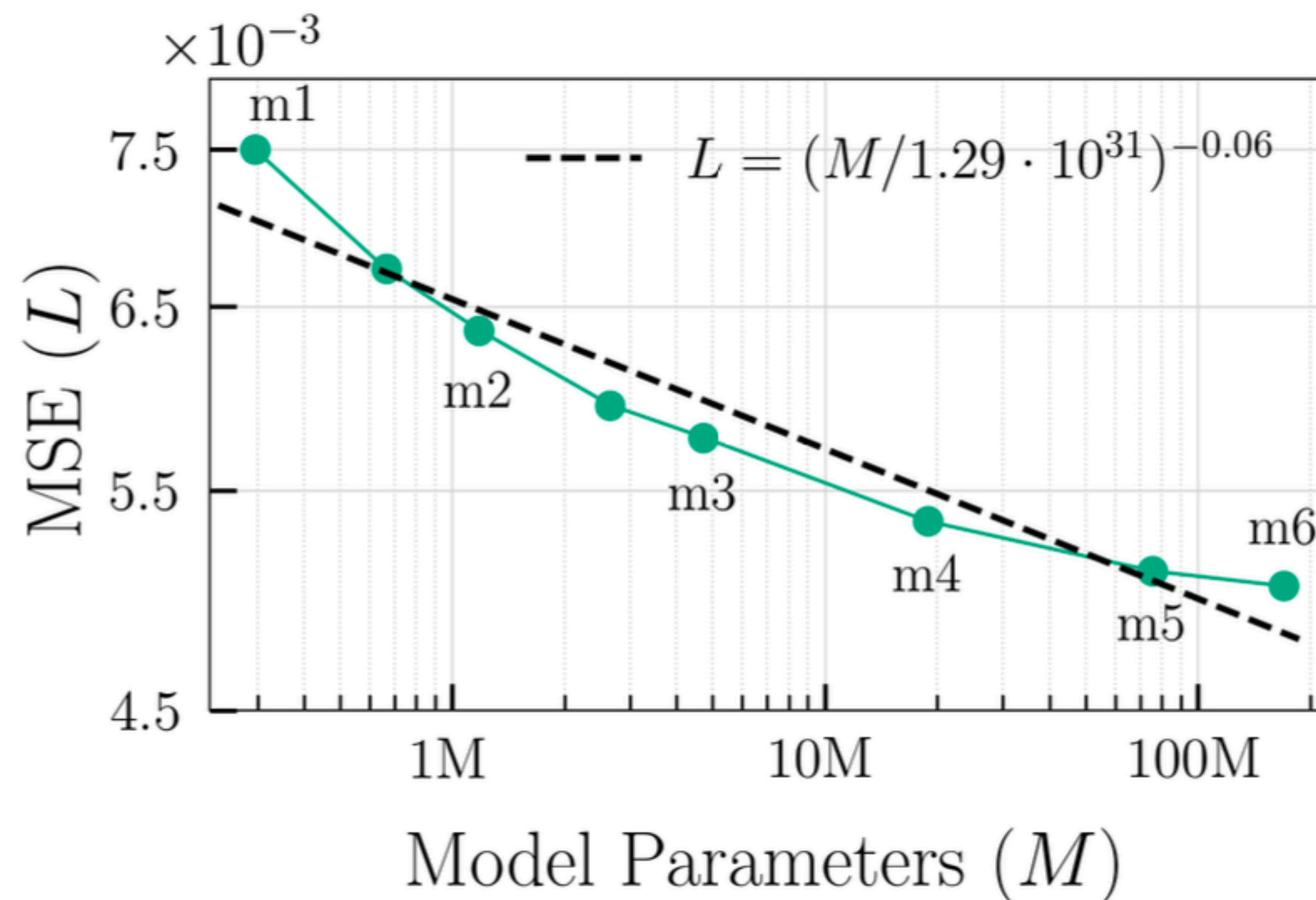
- Initial study by training with simulated Time Projection Chamber spacepoints (clusters) in 10 million $p+p$ collision events
- **Hierarchical Raster Scan (HRS)** serializes spacepoints to 1D sequence
- **Mamba** model: recent State Space Model (SSM) that improves computational efficiency while maintaining long-range sequence modeling capabilities
- **Self-supervised Next k-nearest-neighbor prediction**: instead of ordinary autoregression or “next token” prediction



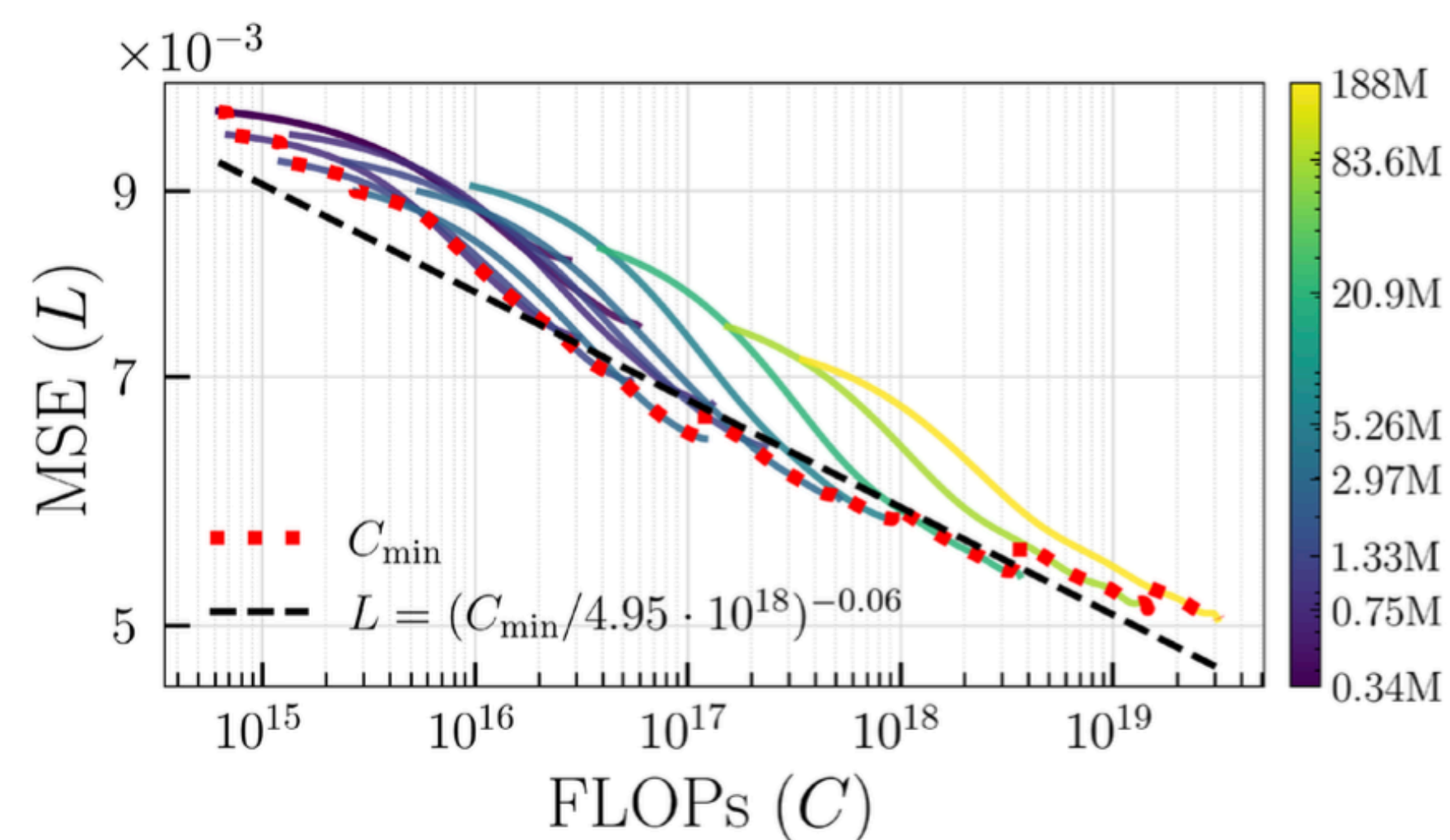
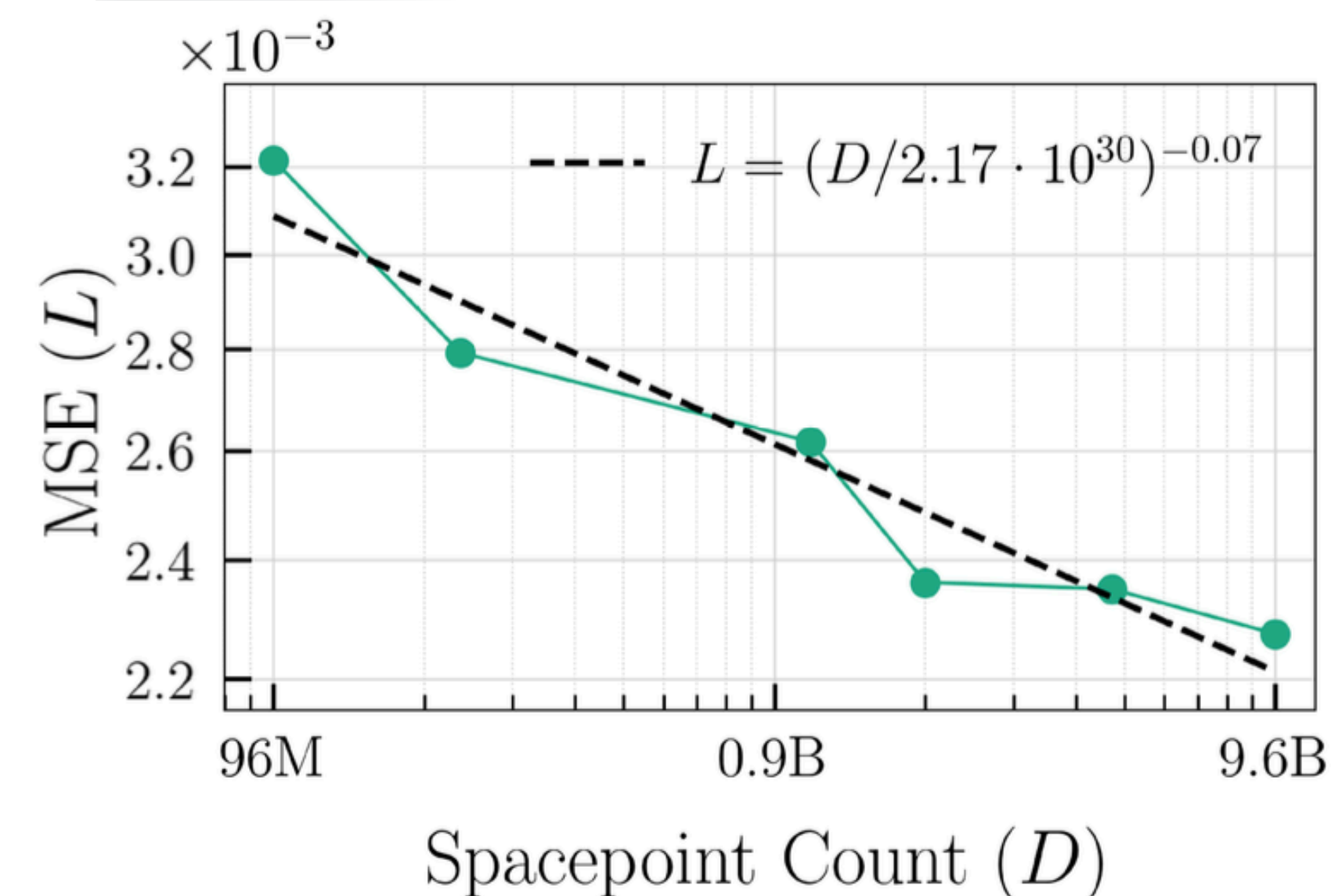
Y. Go, [arXiv:2508.14087](https://arxiv.org/abs/2508.14087) - Accepted to ICLR26

FM4NPP: Neural Scaling Behavior

Y.Go, arXiv:2508.14087 - Accepted to ICLR26



- Established **self-supervised loss scaling behavior** of **FM4NPP** versus **model parameters**, **data size** (Spacepoint Count), and **compute** (FLOPs)
 - ➔ Consistent behavior with scaling observed in LLMs

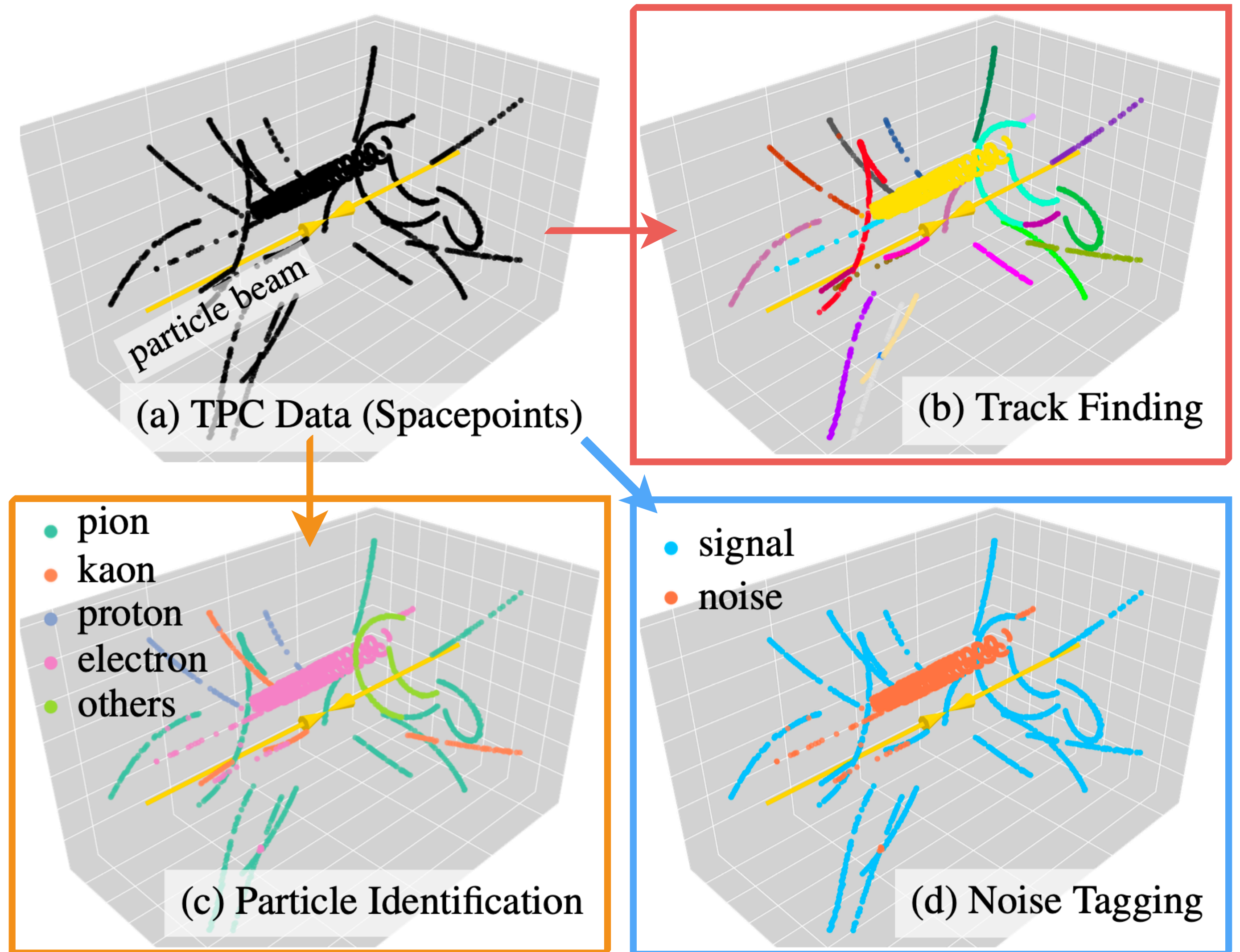


First Foundation Model that shows the neural scaling behavior with low-level detector data!

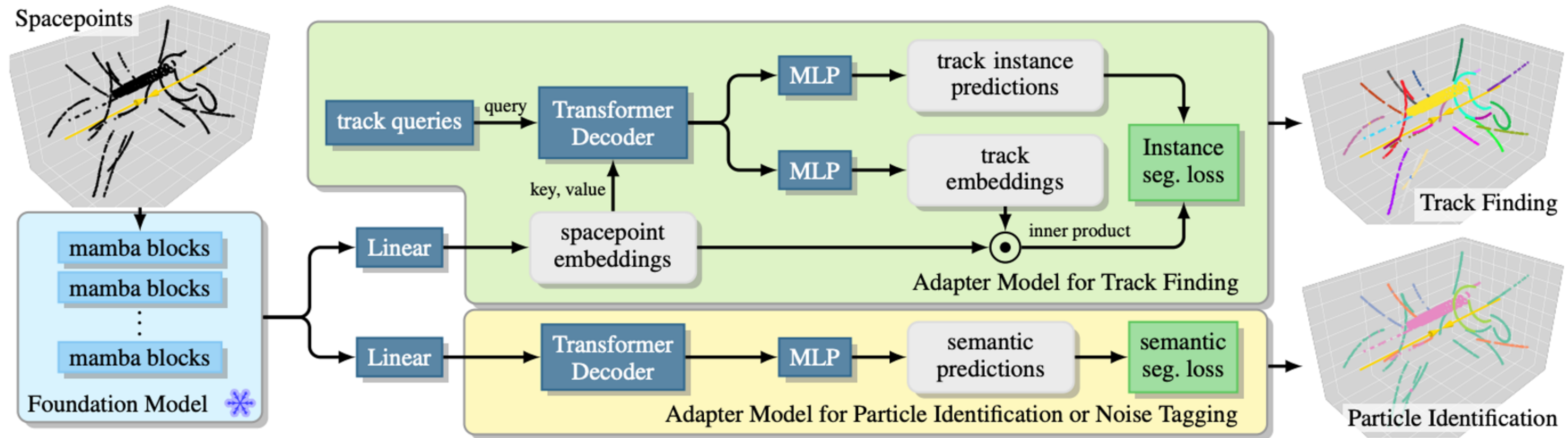
FM4NPP: Downstream Tasks

- The FM backbone is frozen, and only *lightweight decoder* heads are trained
- ➔ **Track Finding**: group spacepoints from the same particle trajectory
- ➔ **Particle Identification**: classify spacepoints based on their particle species
- ➔ **Noise Tagging**: classify spacepoints from low momentum secondary particles (“noise”)

Note: a broader range of downstream tasks will become available as additional detectors are incorporated into the foundation model



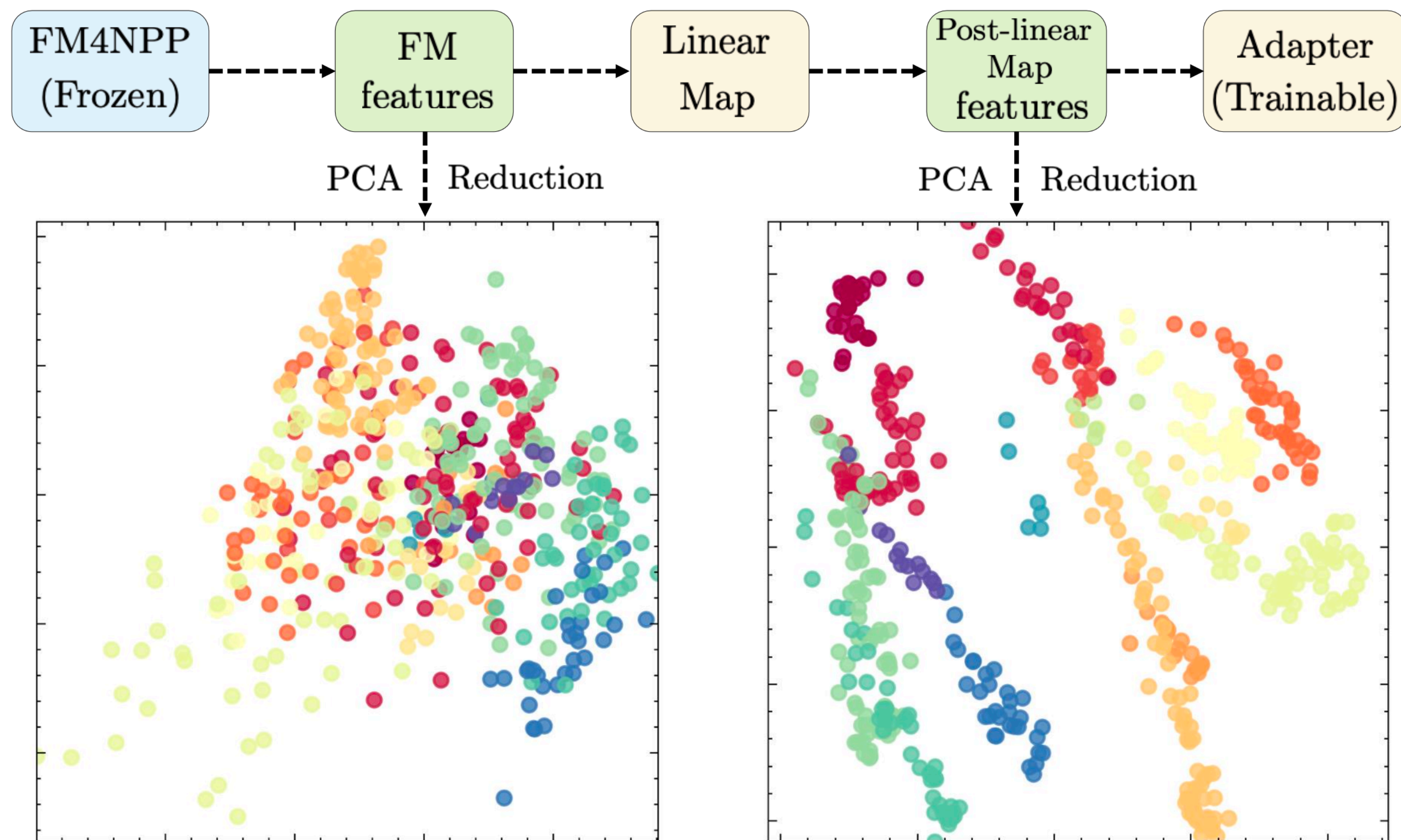
FM4NPP: Adapters for Downstream Tasks



1. FM weights are frozen
2. Lightweight Adaptor models are trainable on labeled data
3. Evaluated on three downstream tasks
: Track Finding, Particle Identification and Noise Tagging

Generalizability of FM4NPP

Visualization of learned embeddings for track finding



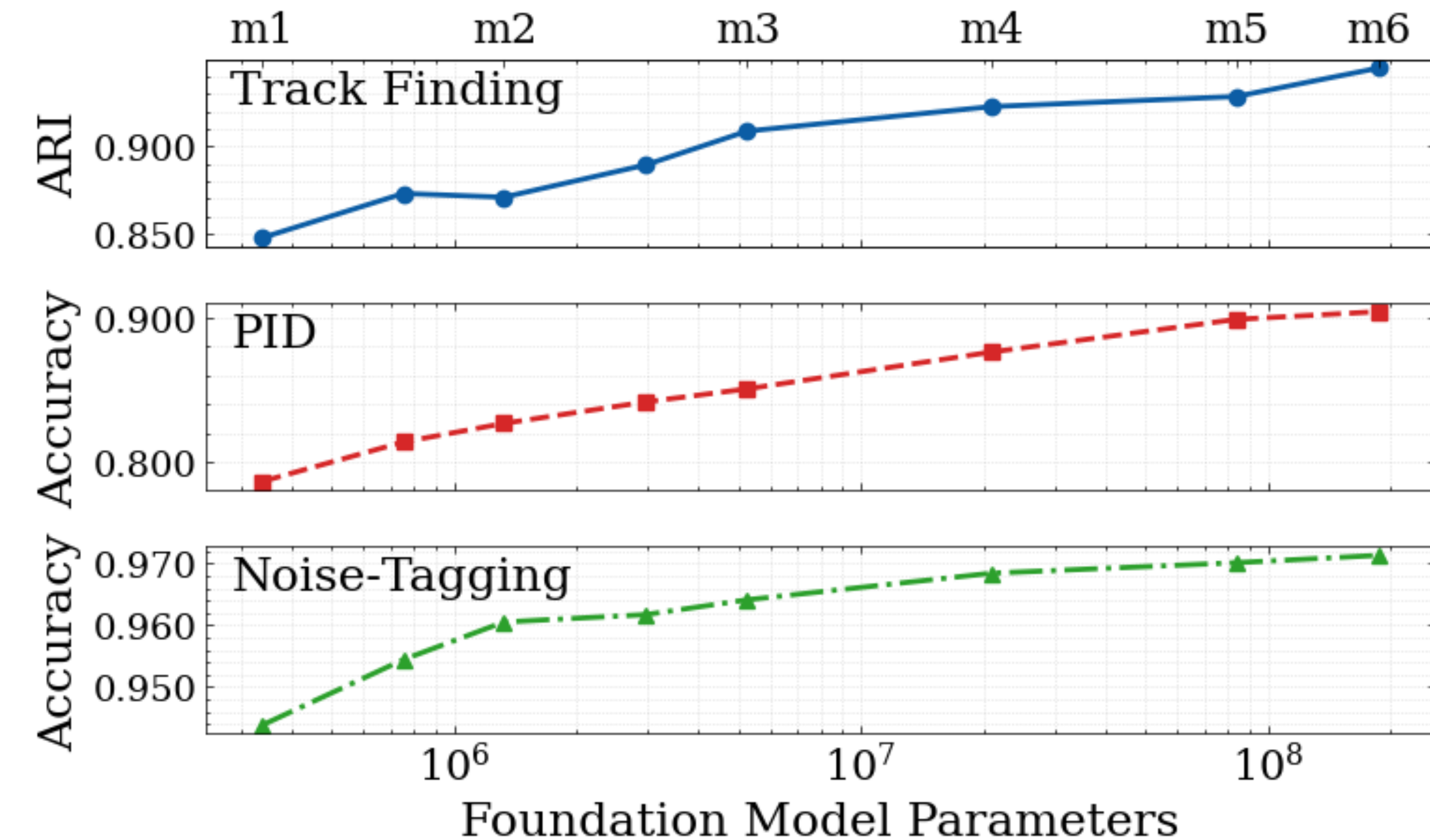
different colors are different tracks

- Features projected via dimensionality reduction (i.e. PCA)
- Raw FM embeddings: no clear separation among particle tracks
 - ➔ representations are task-agnostic
- After single linear projection
 - ➔ well-separated clusters

FM encodes rich, general-purpose representations that require minimal adaptation for task-specific use

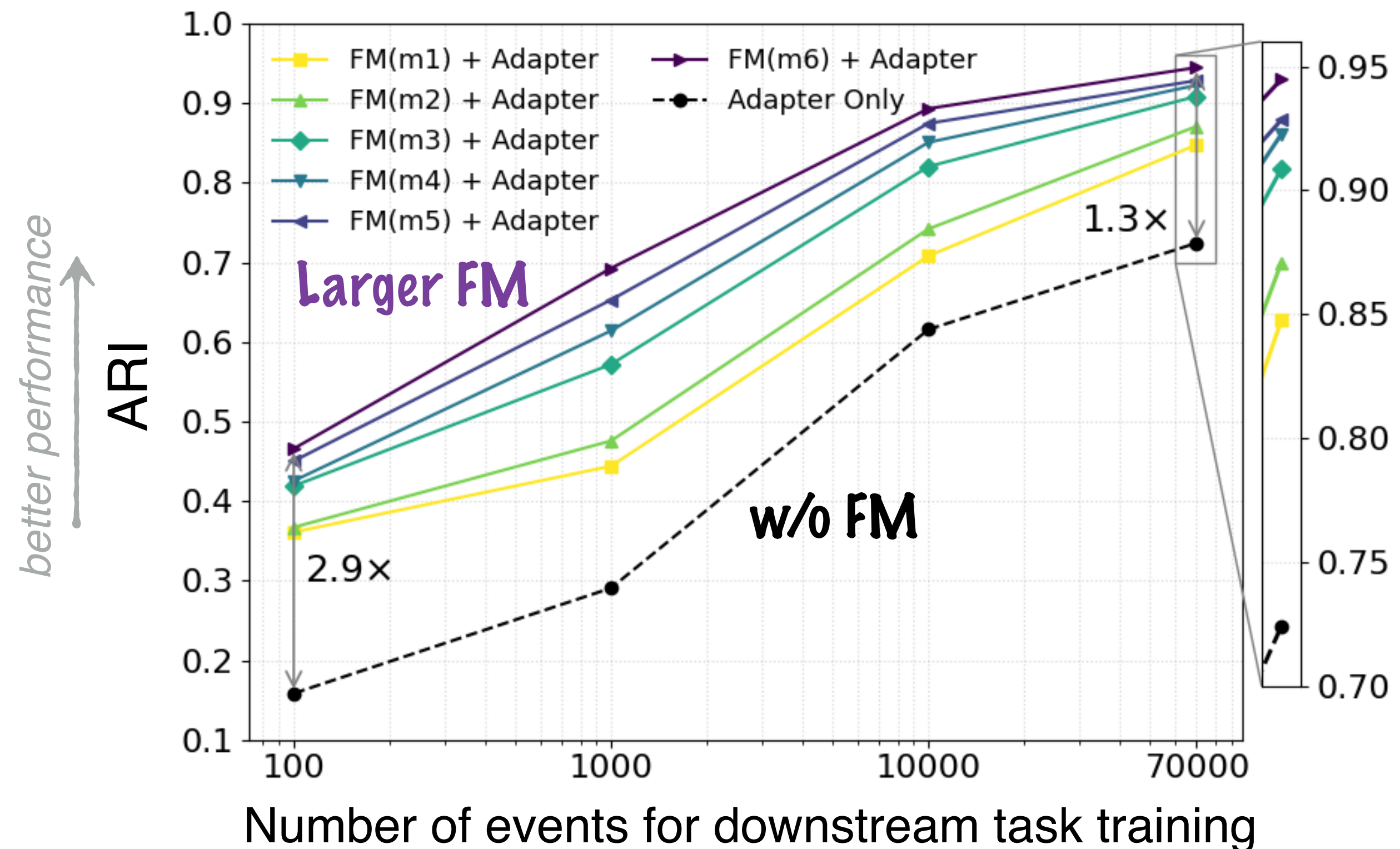
FM4NPP Neural Scaling

Y.Go, *arXiv:2508.14087* - Accepted to ICLR26



Larger FM → Better Performance

- Demonstrated **scaling behavior for downstream tasks** scaled with foundation model size and data size



Larger FM → More Data-Efficient

ARI: Adjusted Rand Index
ARI=1.0 indicates a perfect match

FM4NPP Performance on Downstream Tasks

Y.Go, arXiv:2508.14087
Accepted to ICLR26

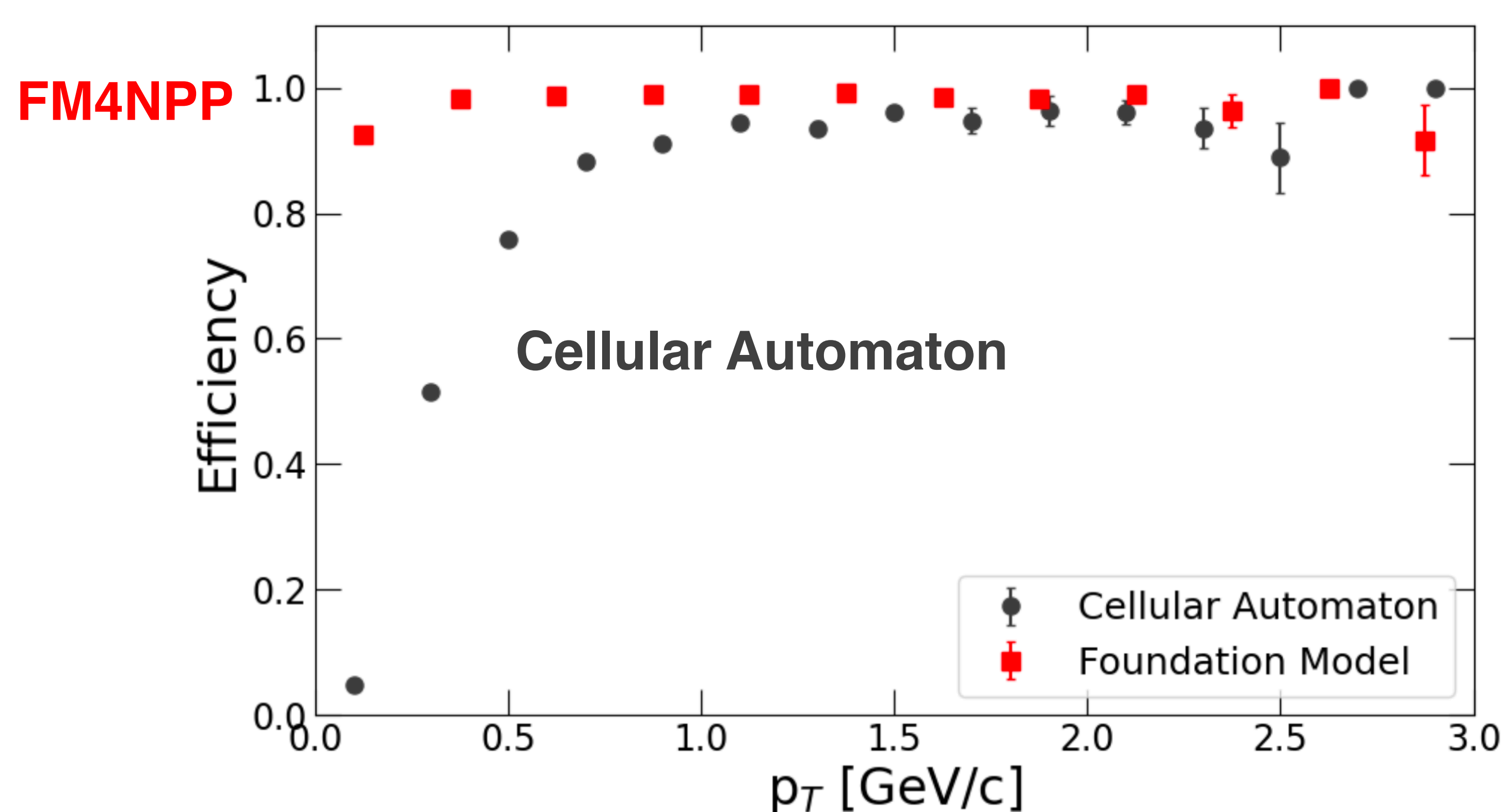
Y.Go, arXiv:2508.14087
Accepted to ICLR26

		Track Finding				
		model	#trnbl para.	ARI↑	efficiency↑	purity↑
		EggNet	0.16M	0.7256	74.19%	75.14%
		Exa.TrkX	3.86M	0.8765	91.79%	66.42%
ours (w/o FM)		AdapterOnly	2.39M	0.7243	78.01%	64.54%
ours (w/ FM)		FM4NPP (m6)	2.39M	0.9448	96.08%	93.08%

		Particle Identification			Noise Tagging				
		model	#trnbl para.	acc.↑	recall↑	pre.↑	acc.↑	recall↑	pre.↑
		SAGEConv	0.91M	0.7262	0.4563	0.6502	0.9174	0.7227	0.8165
		OneFormer3D	44.95M	0.7701	0.4897	0.5767	0.9646	0.9404	0.8948
ours (w/o FM)		AdapterOnly	0.74M	0.6631	0.3387	0.6111	0.9111	0.6215	0.8359
ours (w/ FM)		FM4NPP (m6)	0.74M	0.9039	0.7652	0.8782	0.9713	0.9367	0.9190

- **FM4NPP** shows clear performance gains compared to an “**AdapterOnly**” baseline
- **Superior performance** of tracking with FM4NPP compared to other state-of-art ML models

FM4NPP Track Finding Performance



*Detailed physics performance paper
in preparation*

- **Outperform the traditional method**

- ➔ tracking efficiency of 99.6% exceeding the sPHENIX pipeline's 94.6% (Cellular Automaton) for $p_T > 1$ GeV

Note - FM is all truth tracks with at least 5 clusters, CASeeder is only primaries with at least 20 clusters in TPC acceptance (!)

FM4NPP: Future Directions

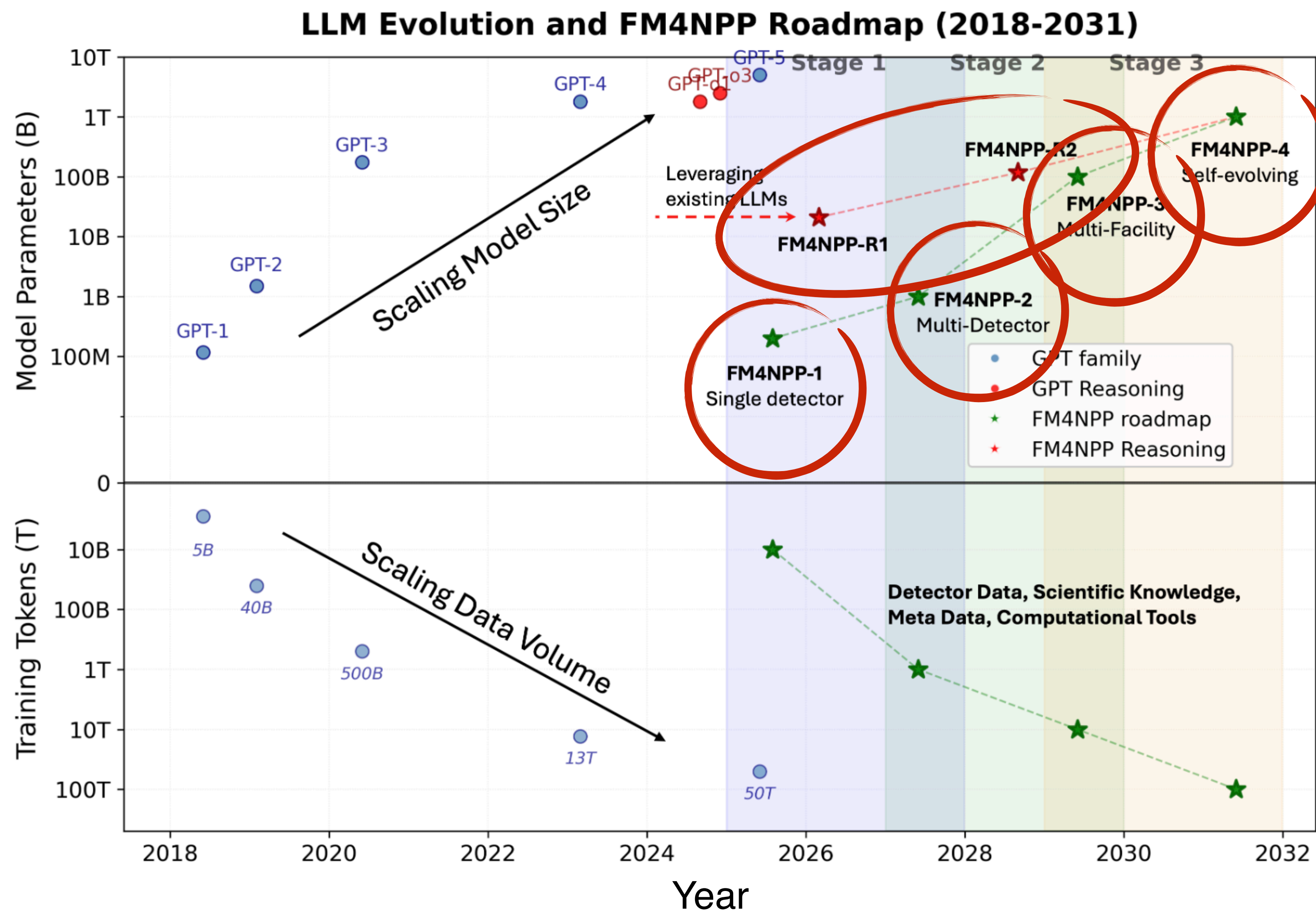
- **Current stage**

- ➔ initial stage (**FM4NPP-1**) utilized single detector (sPHENIX TPC simulation) with **180M** parameters
- ➔ Similar scaling behavior, which was first observed in ChatGPT-2

- **Next stages**

- ➔ Demonstrate on real and more complex data (e.g. heavy ion collisions), 1B parameters (**FM4NPP-1+**)
- ➔ Multi-detectors (**FM4NPP-2**)
- ➔ Demonstrate generalizability to multi-facility (**FM4NPP-3**)
- ➔ Leverage existing LLM reasoning, connect FM4NPP to theory and existing knowledge base (**FM4NPP-R1,2**)
- ➔ Self-evolving AI (**FM4NPP-4** with an integrated reasoning model)

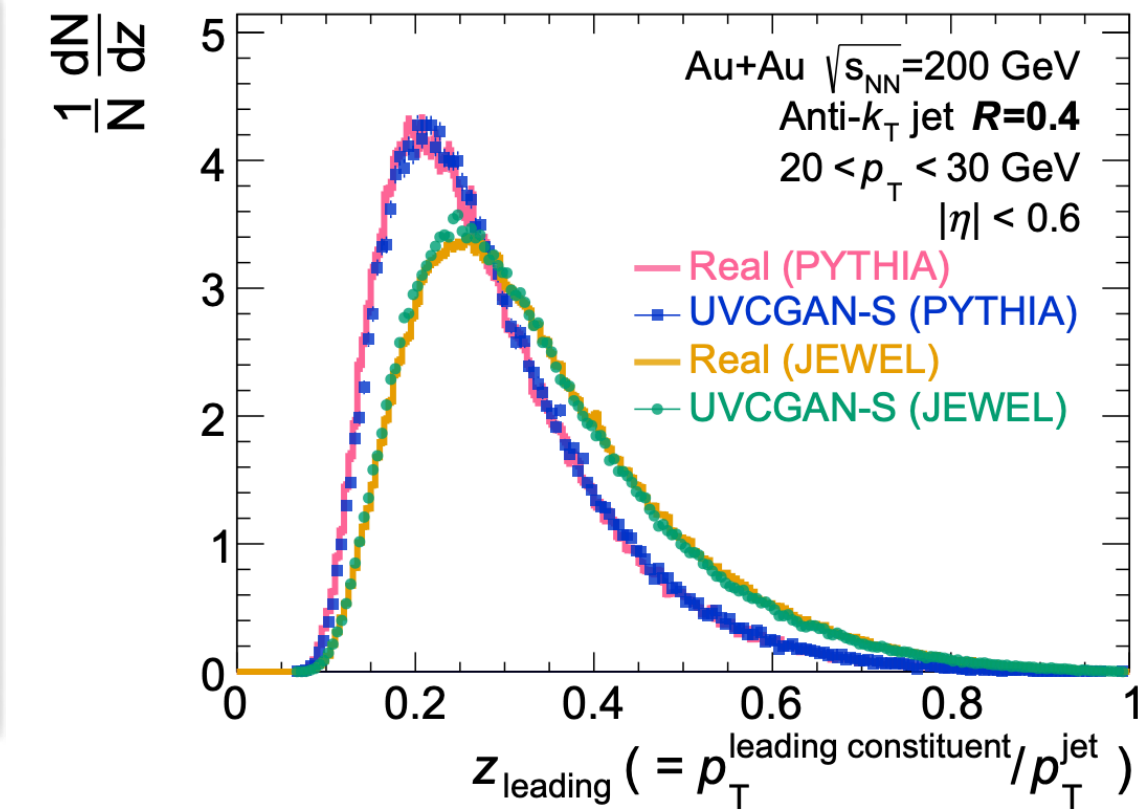
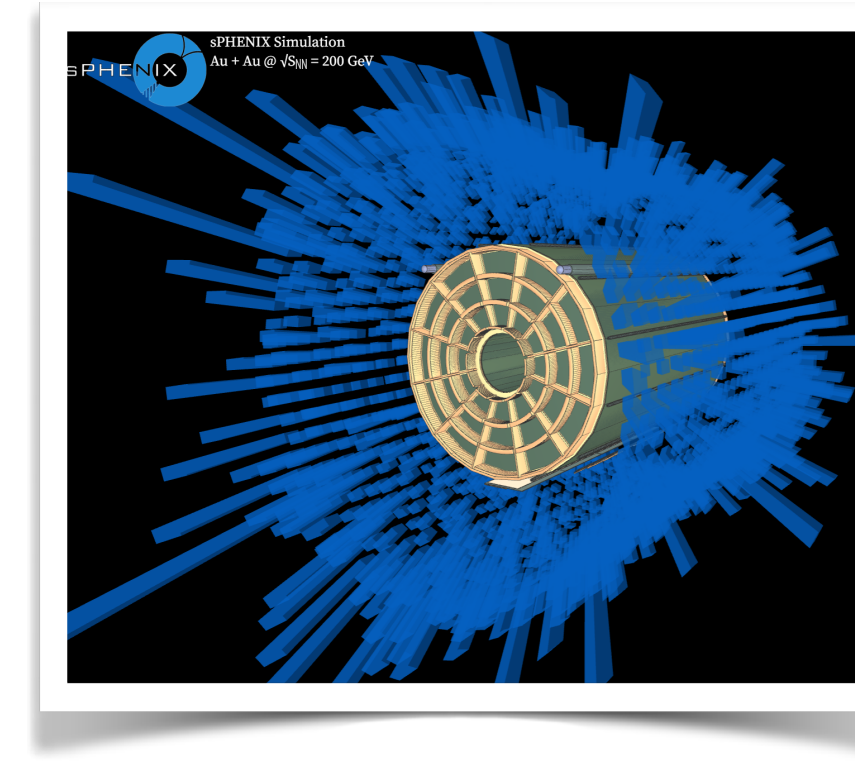
[Roadmap Whitepaper \[ssrn.5467666\]](#)



Prospects of AI in Nuclear and Particle Physics

- Task-specific AI is already delivering real value opening new opportunities

➔ e.g. **generative AI** enabling *fast simulation*, *signal extraction analysis*



- **Foundation Model for Nuclear and Particle Physics (FM4NPP)**

➔ *The ultimate goal*: autonomous, **self-evolving AI system** and **AI-driven physics discovery**



Drive a paradigm shift in nuclear and particle physics discovery !

Thank you for listening

BACK UP