Generative AI for full-detector, whole-event simulation of heavy-ion collisions

Yeonju Go[a], Dmitrii Torbunov[a], Jin Huang[a], Yihui Ren[a], Yi Huang[a],
Meifeng Lin[a], Haiwang Yu[a], Brett Viren[a], Tim Rinn[b]

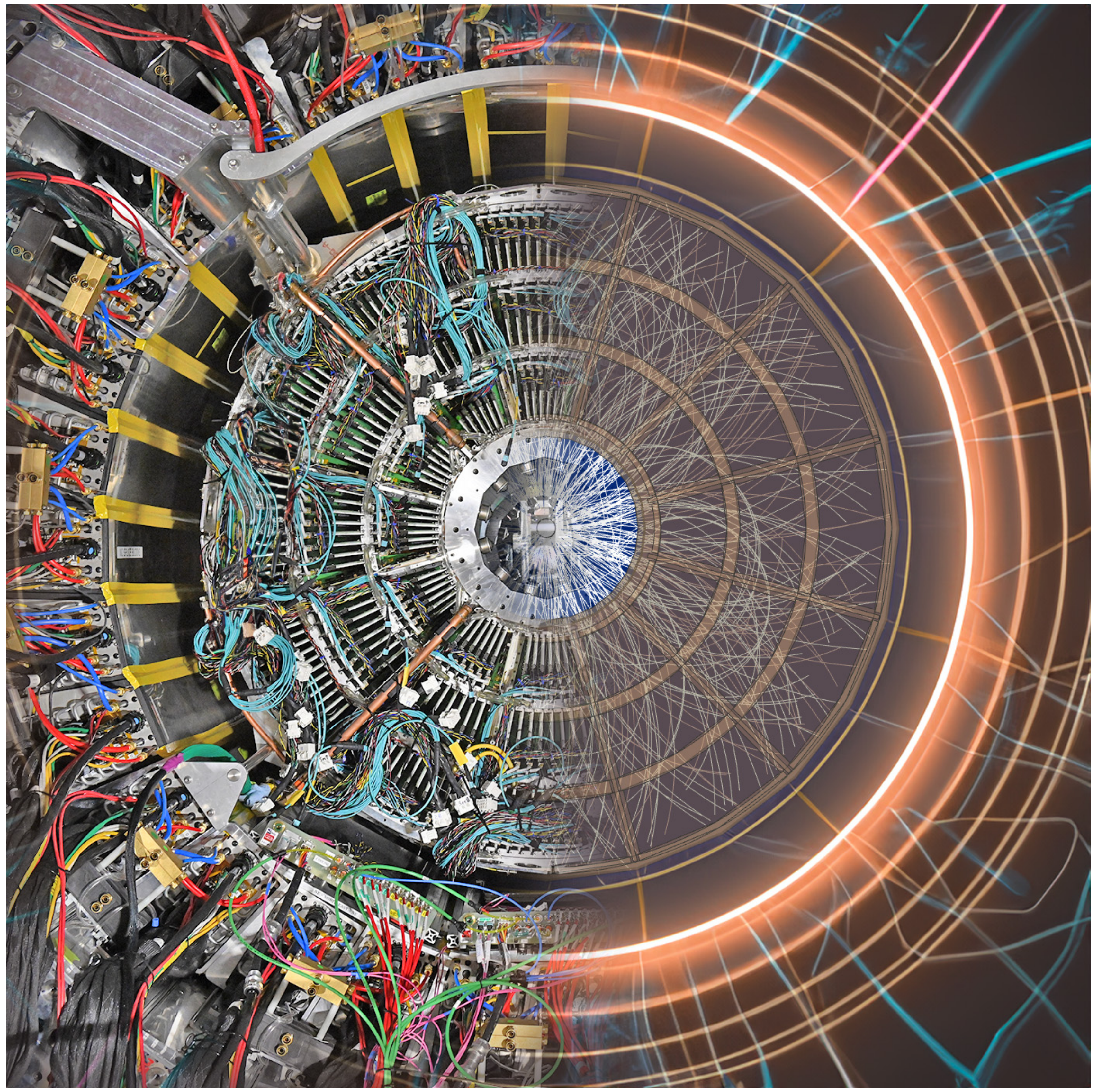
a: Brookhaven National Laboratory
b: Los Alamos National Laboratory

RHIC/AGS Annual Users' Meeting
June 11, 2024
Brookhaven National Laboratory

Brookhaven National Laboratory

Simulations of Relativistic Heavy Ion Collisions

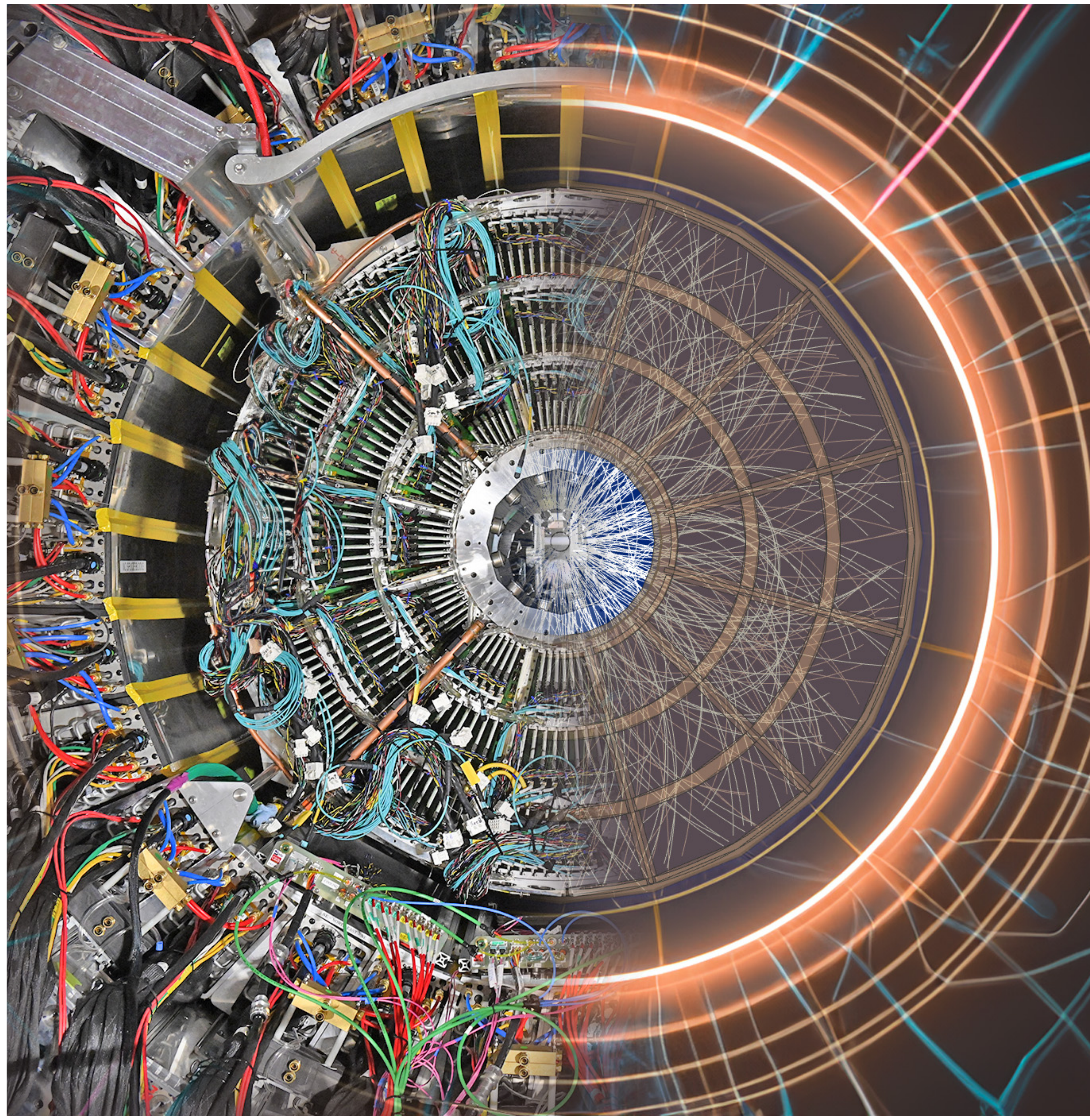
- ***O(1000) particles*** in one nuclear collision event
  + ***thousands shower steps*** per particle
  ➡ Simulation of the interaction of particles with detectors is **high complexity and computationally intensive work**



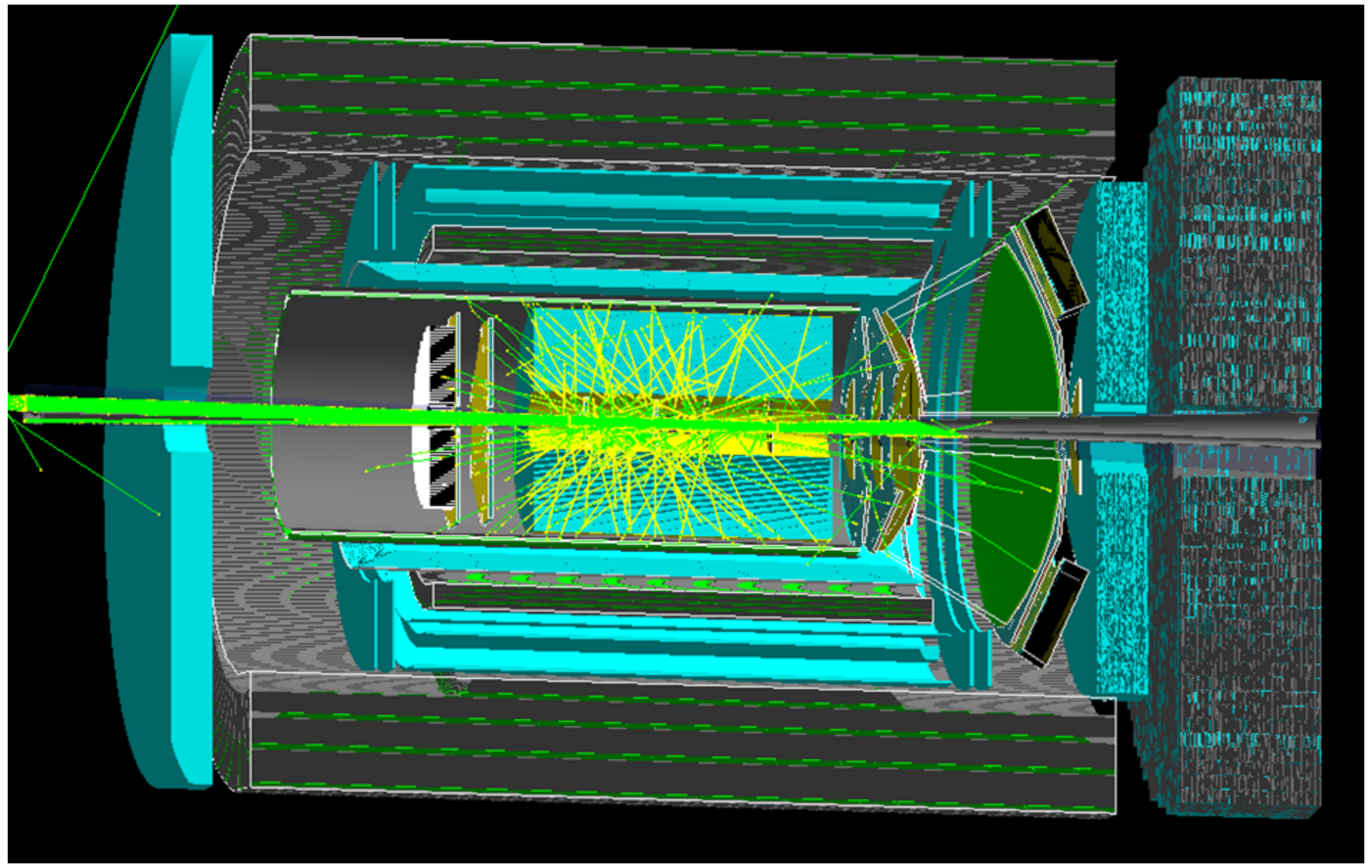
sPHENIX TPC

Simulations of Relativistic Heavy Ion Collisions

- ***O(1000) particles*** in one nuclear collision event
  + ***thousands shower steps*** per particle
  ➡ Simulation of the interaction of particles with detectors is **high complexity and computationally intensive work**
- *Electron-Ion Collider* will need a large amount of simulations of full detector with both physics and machine background



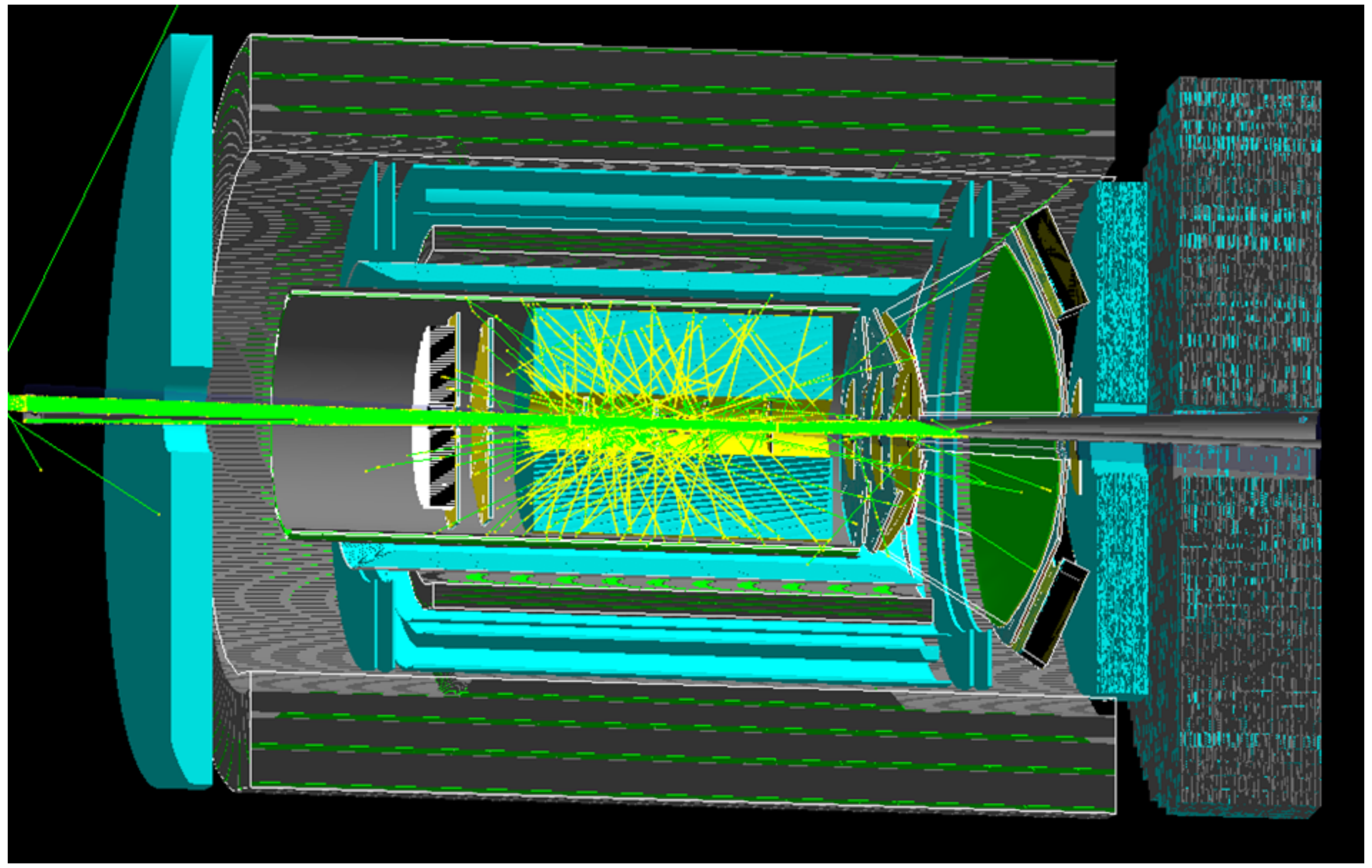
sPHENIX TPC



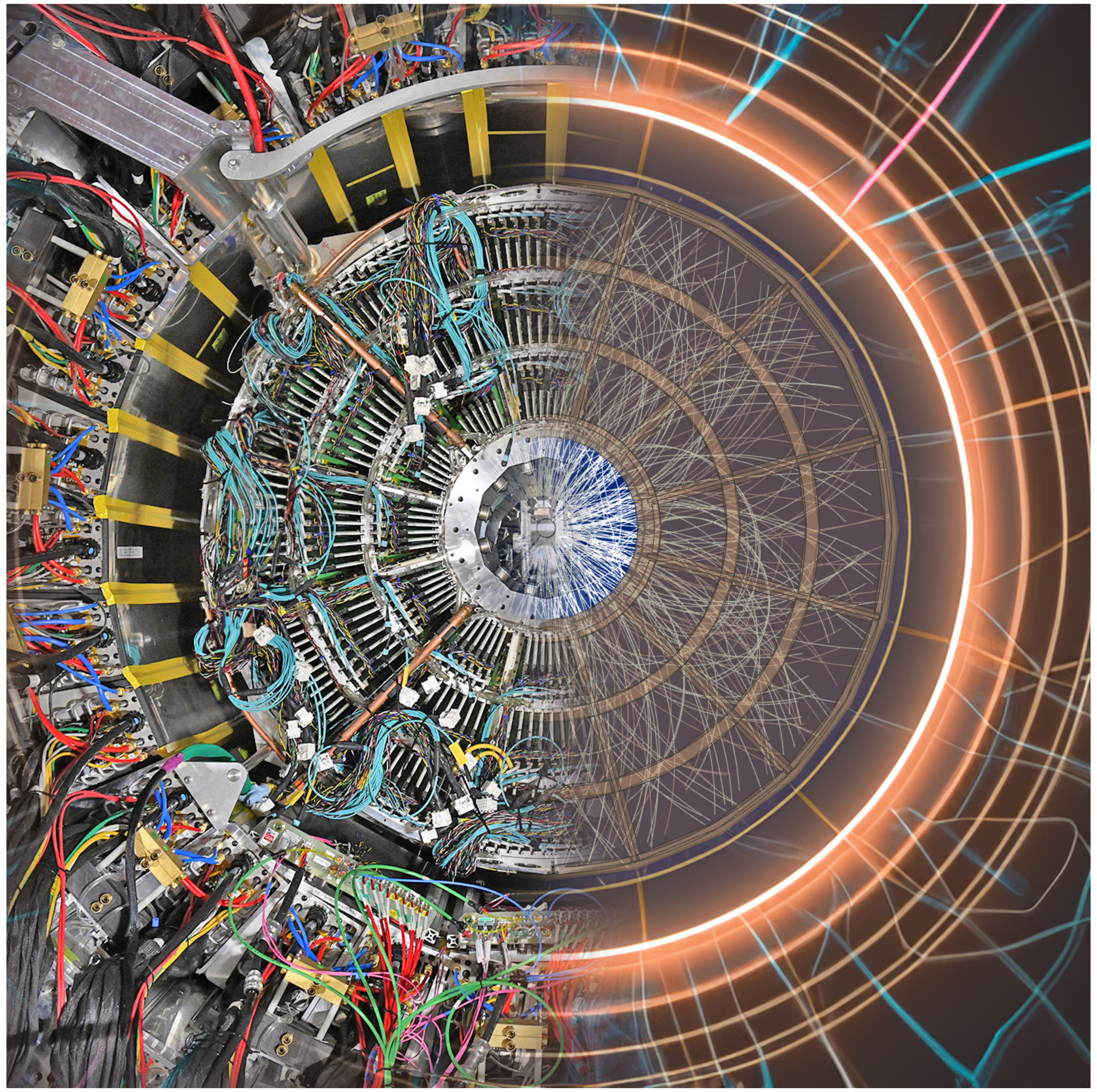
EIC CDR

Simulations of Relativistic Heavy Ion Collisions

- ***O(1000) particles*** in one nuclear collision event
  + ***thousands shower steps*** per particle
  ➡ Simulation of the interaction of particles with detectors is **high complexity and computationally intensive work**
- *Electron-Ion Collider* will need a large amount of simulations of full detector with both physics and machine background
- <u>ML can speed up and produce large amount of the heavy ion event simulations!</u>



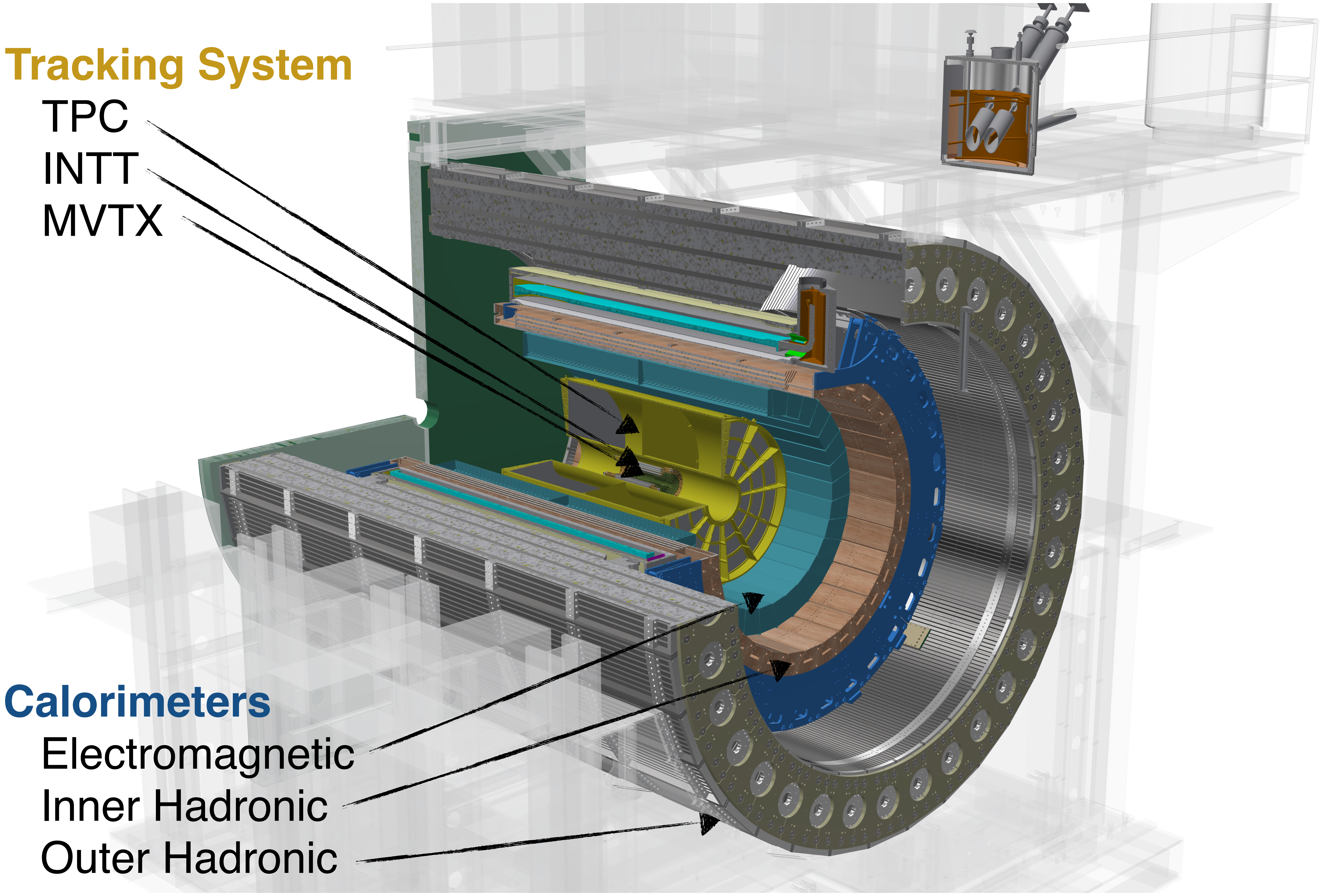
EIC CDR



sPHENIX TPC

We introduce ***full detector whole-event ML simulations*** for heavy ion collisions

sPHENIX Detector at RHIC

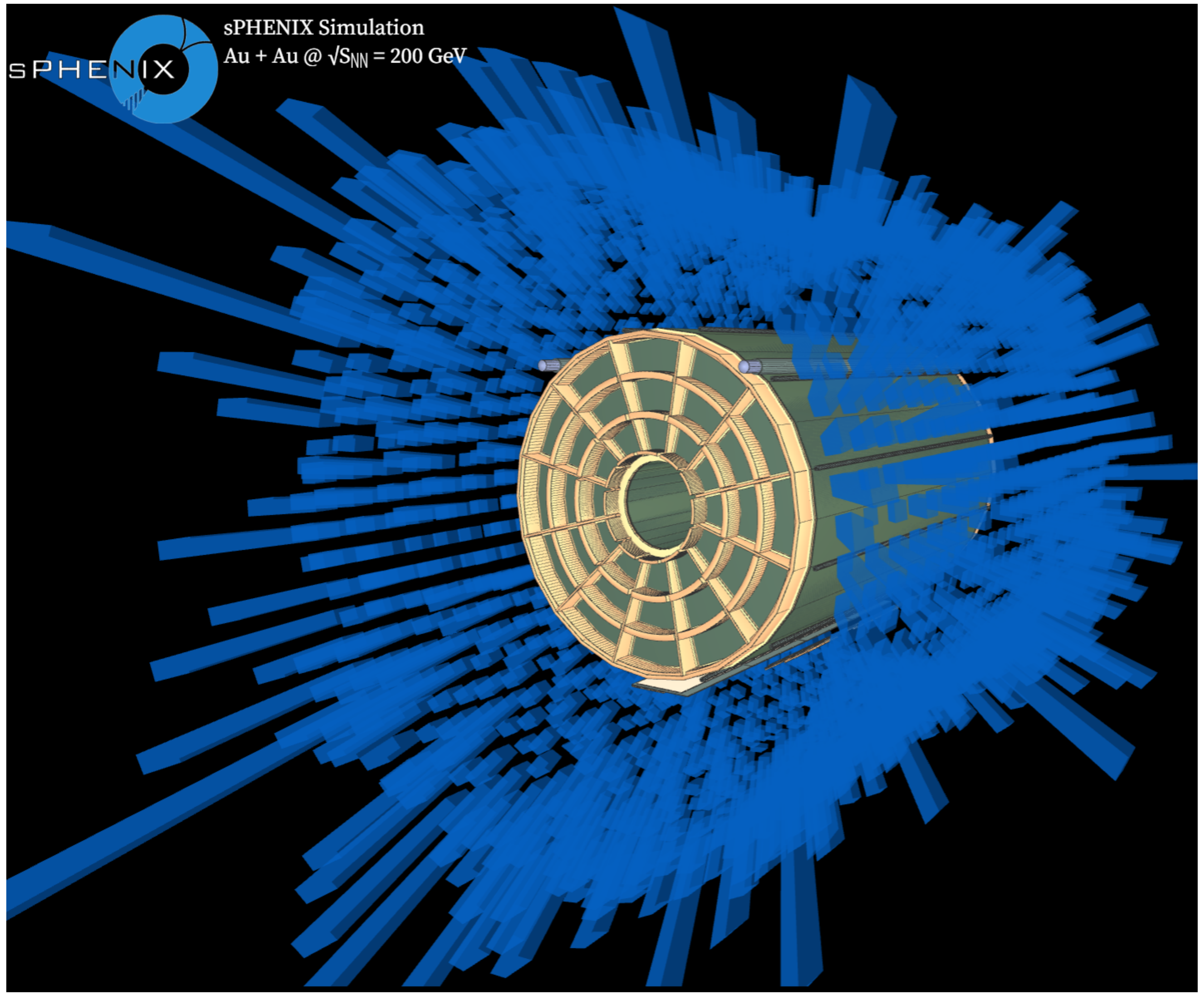


- Data taking began last year!
- High-precision tracking system + Hermetic Electromagnetic & Hadronic calorimeters

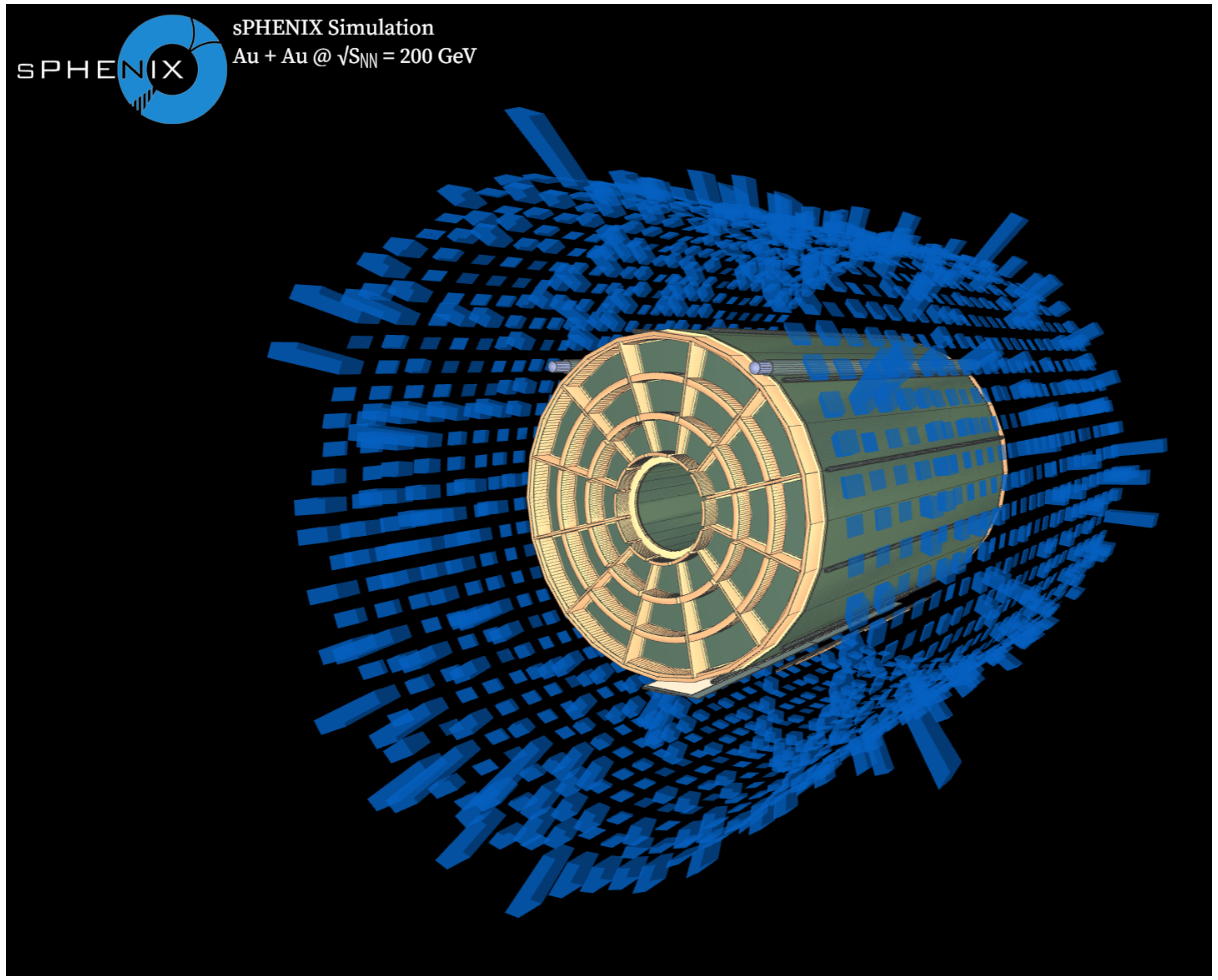
Heavy Ion Collision Event

- **HIJING** Monte Carlo event generator for Au+Au collisions at $\sqrt{s_{NN}}$=200 GeV
- **Geant4** full detector simulation with the sPHENIX geometry

Head-on collision (0-10% Centrality)

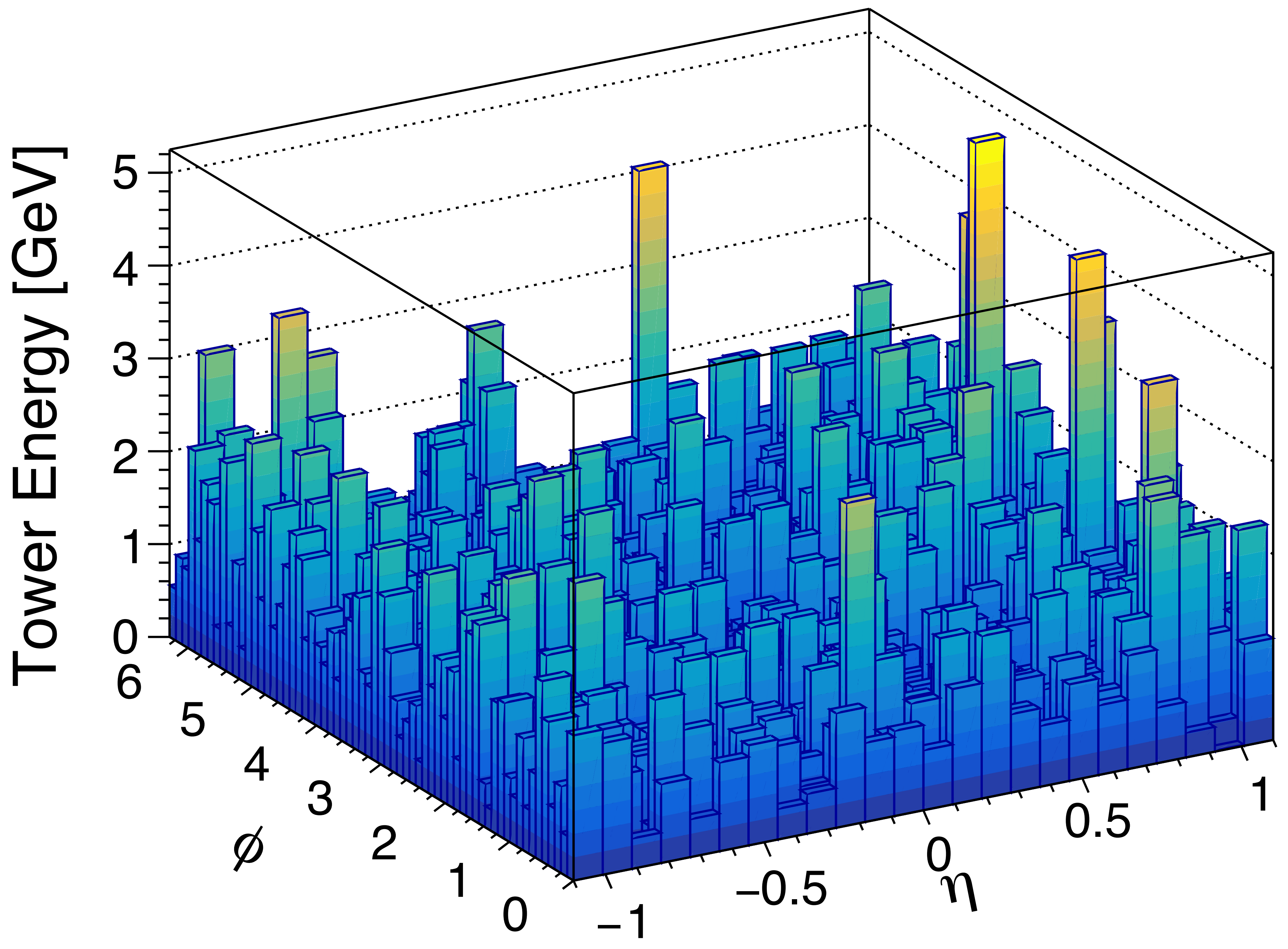


Side collision (40-50% Centrality)

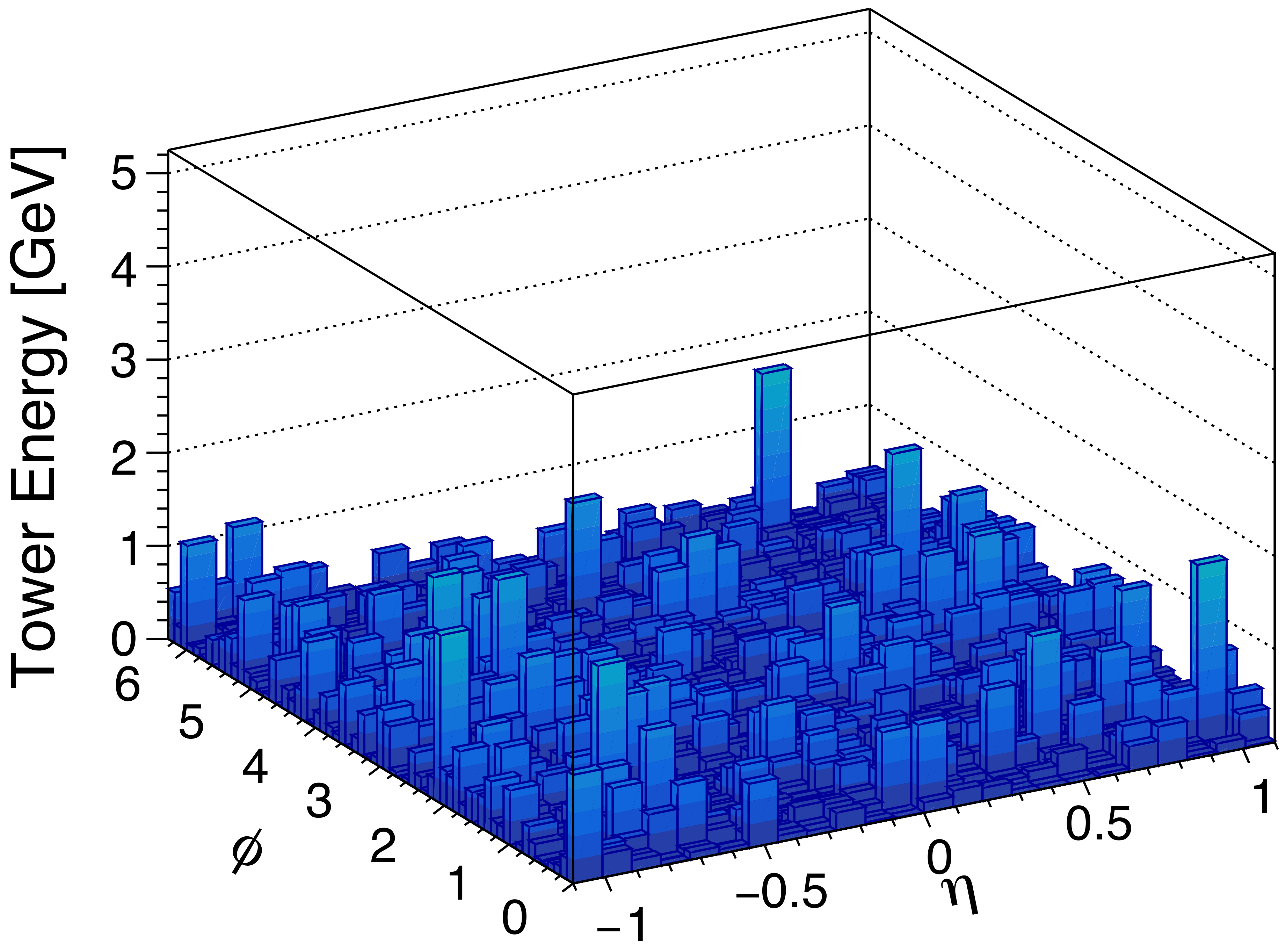


Tower Distributions

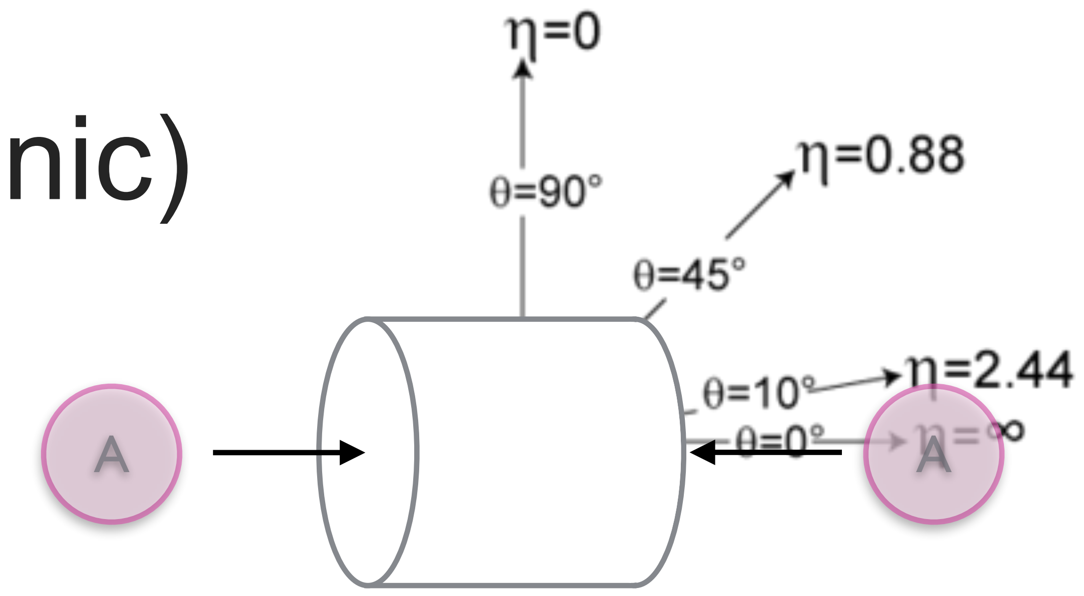
0-10% Centrality



40-50% Centrality



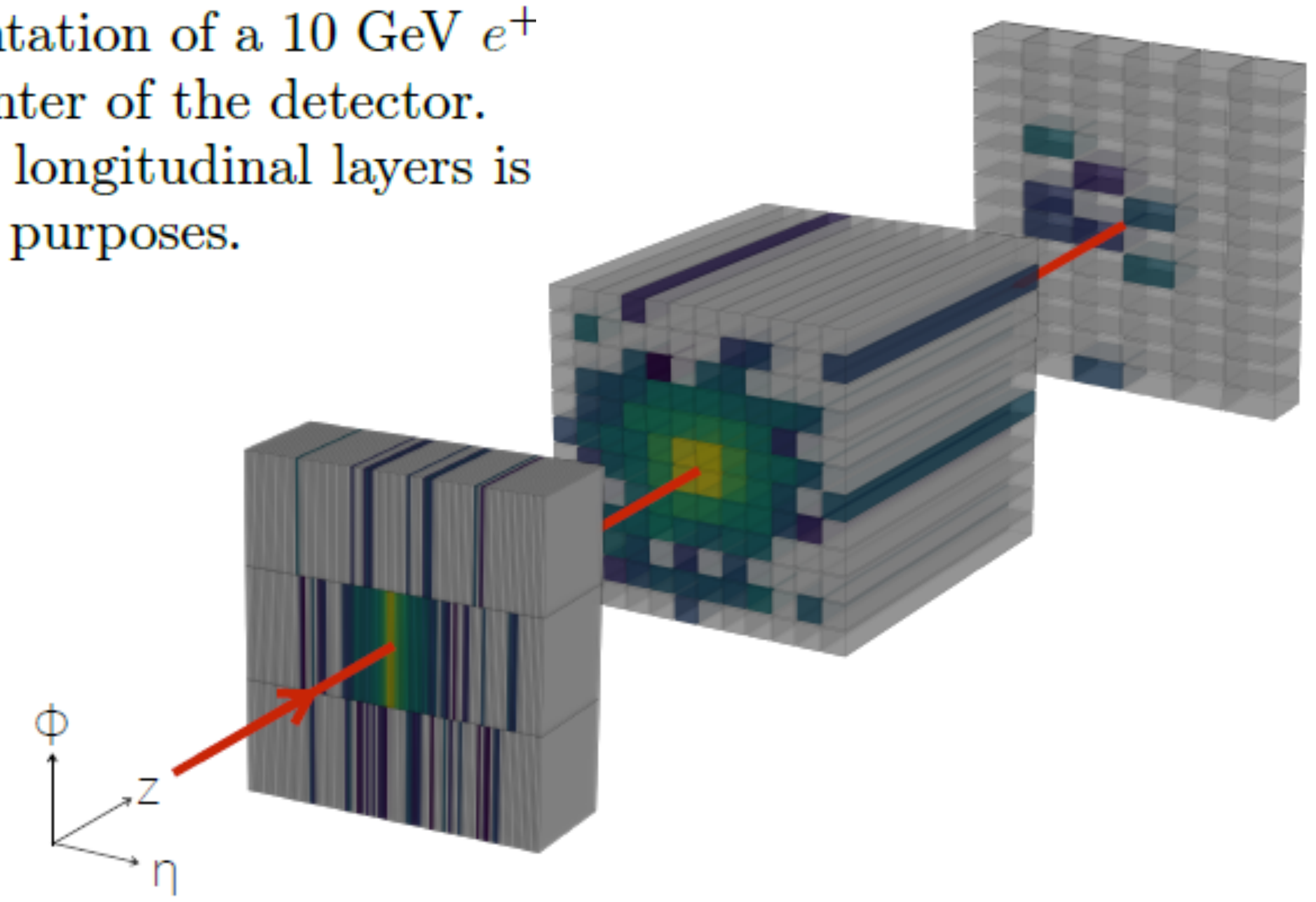
- Full calorimeter **towers** (Electromagnetic + Inner hadronic + Outer hadronic)
  - ➡ $-1.1 < \eta < 1.1, \quad 0 < \phi < 2\pi$
  - ➡ (24 x 64) bins in (η, ϕ)



Generative AI

- **Generative Adversarial Networks (GAN)**
  - ➡ actively used in high energy physics *(e.g. arXiv:1712.1032, arXiv:2209.07559, EPJC 80 (2020) 688, arXiv:2210.14245)*

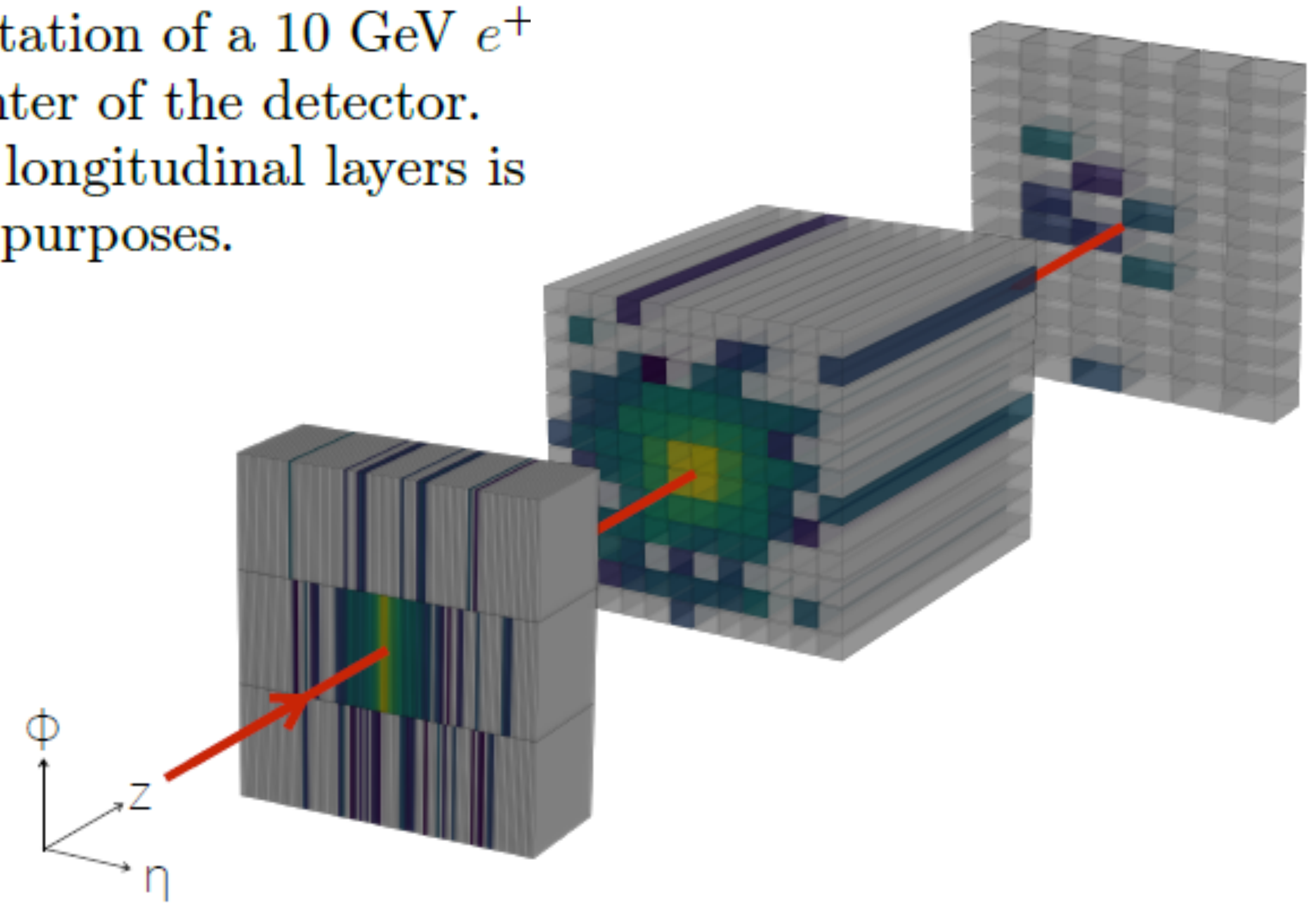
FIG. 2: Three-dimensional representation of a 10 GeV e^+ incident perpendicular to the center of the detector. Not-to-scale separation among the longitudinal layers is added for visualization purposes.

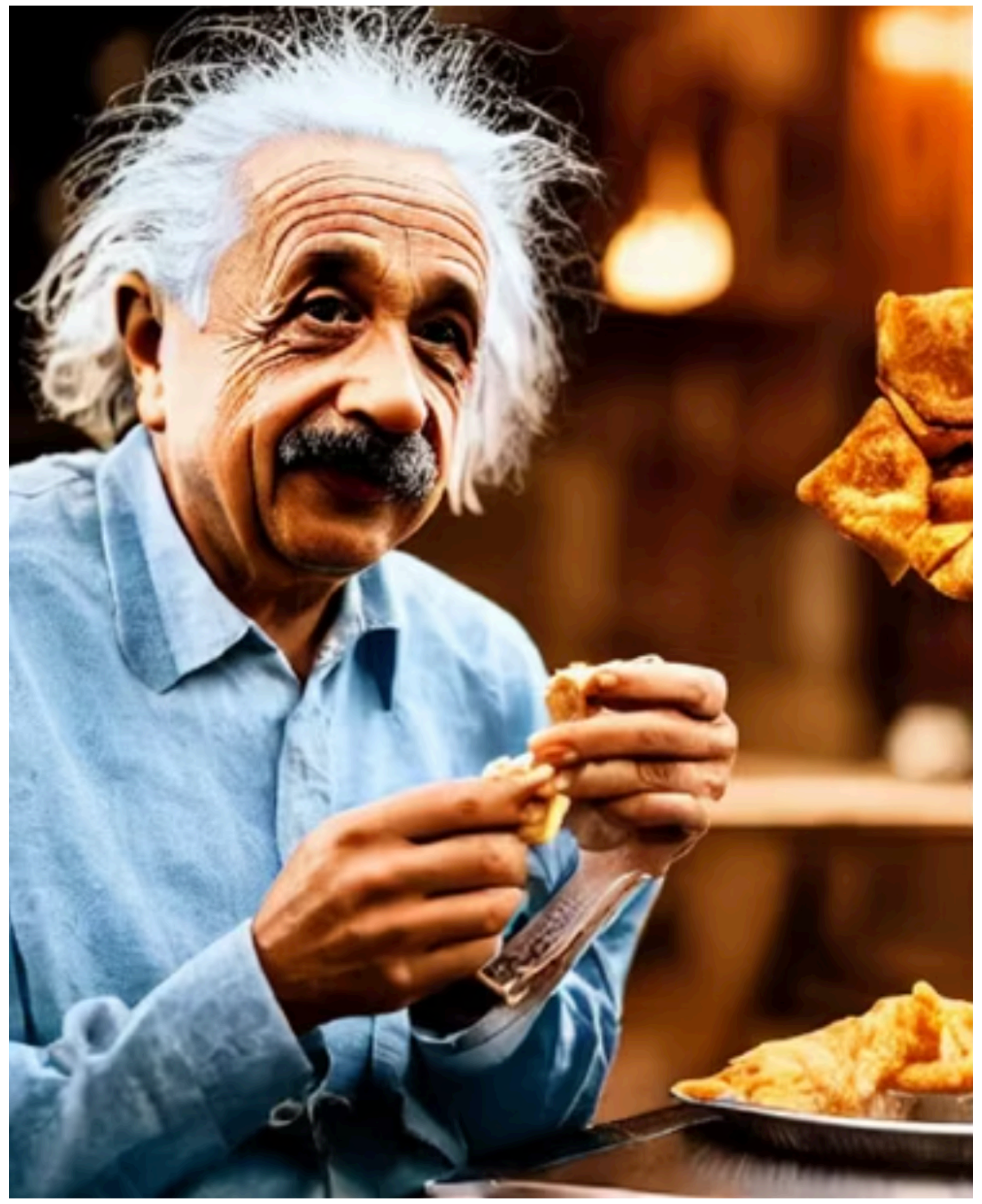


Generative AI

- **Generative Adversarial Networks (GAN)**
  ➡ actively used in high energy physics
  (e.g. arXiv:1712.1032, arXiv:2209.07559, EPJC 80 (2020) 688, arXiv:2210.14245)
- **Diffusion Models**: text-to-image generation in industry (e.g. StableDiffusion, Midjourney, Dalle-2)

FIG. 2: Three-dimensional representation of a 10 GeV e^+ incident perpendicular to the center of the detector. Not-to-scale separation among the longitudinal layers is added for visualization purposes.





Generative AI

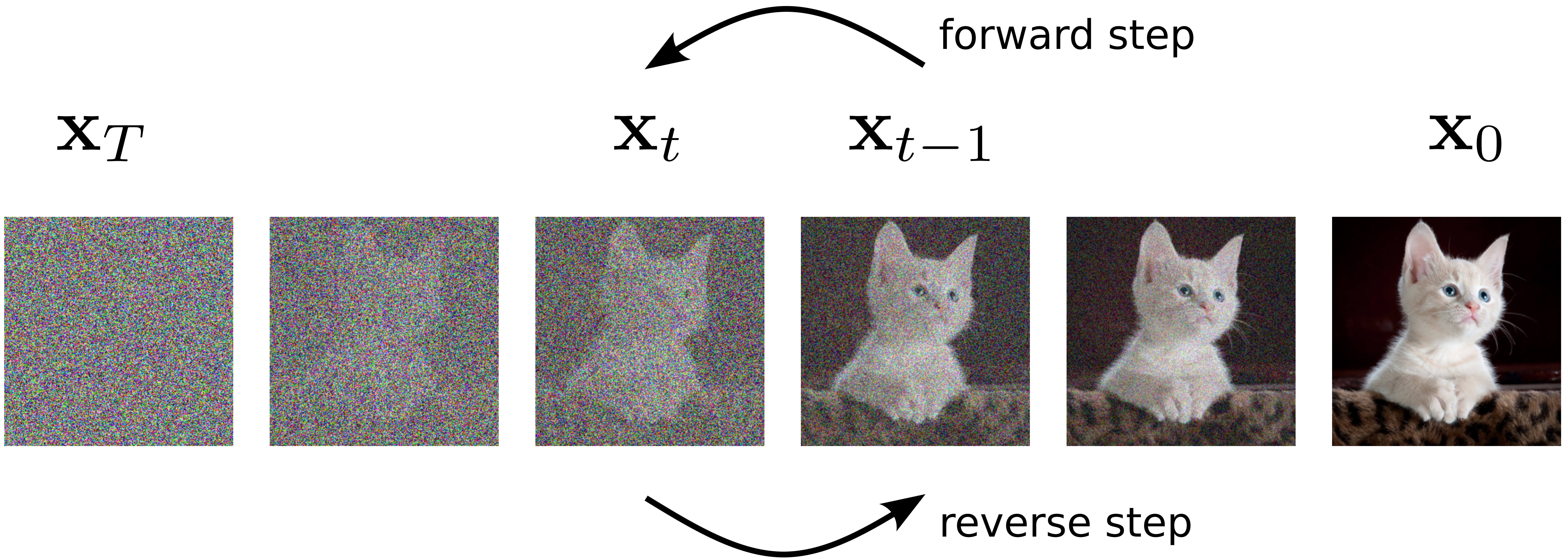
Generative AI

- **Diffusion Models**:
  text-to-image generation in industry
  (e.g. StableDiffusion, Midjourney, Dalle-2)
  ➡ known for high fidelity

Diffusion Model (DALL·E3 by OpenAI) generating a sPHENIX meeting
Note difficulty in generating features such as text

Denoising Diffusion Probabilistic Model (DDPM)

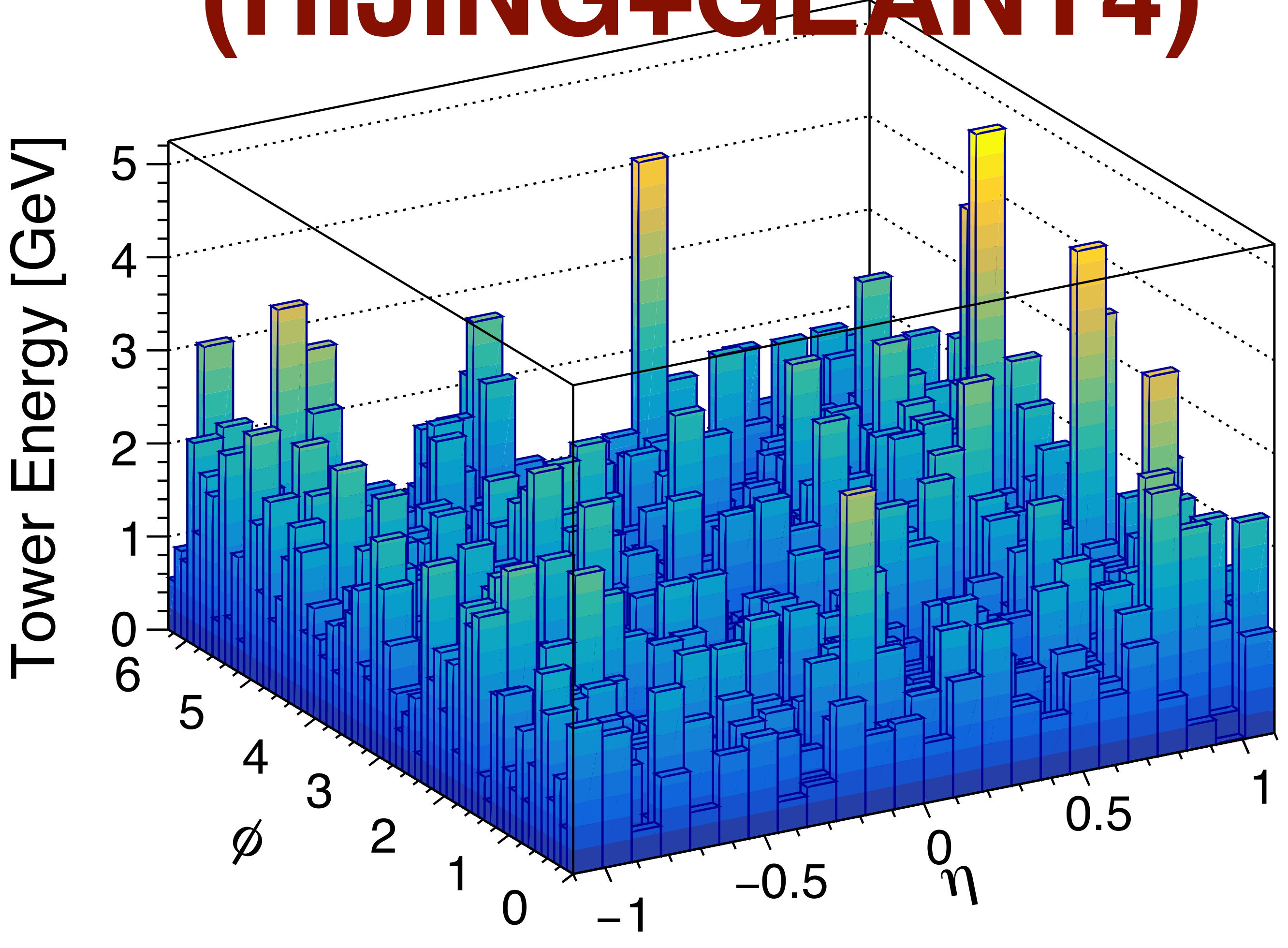
- DDPM provides ***high quality data from random noise***
- **Forward** process: add random gaussian noise
- **Reverse** process: use neural network and <u>generate data</u>
- In real application, O(1,000) steps are used



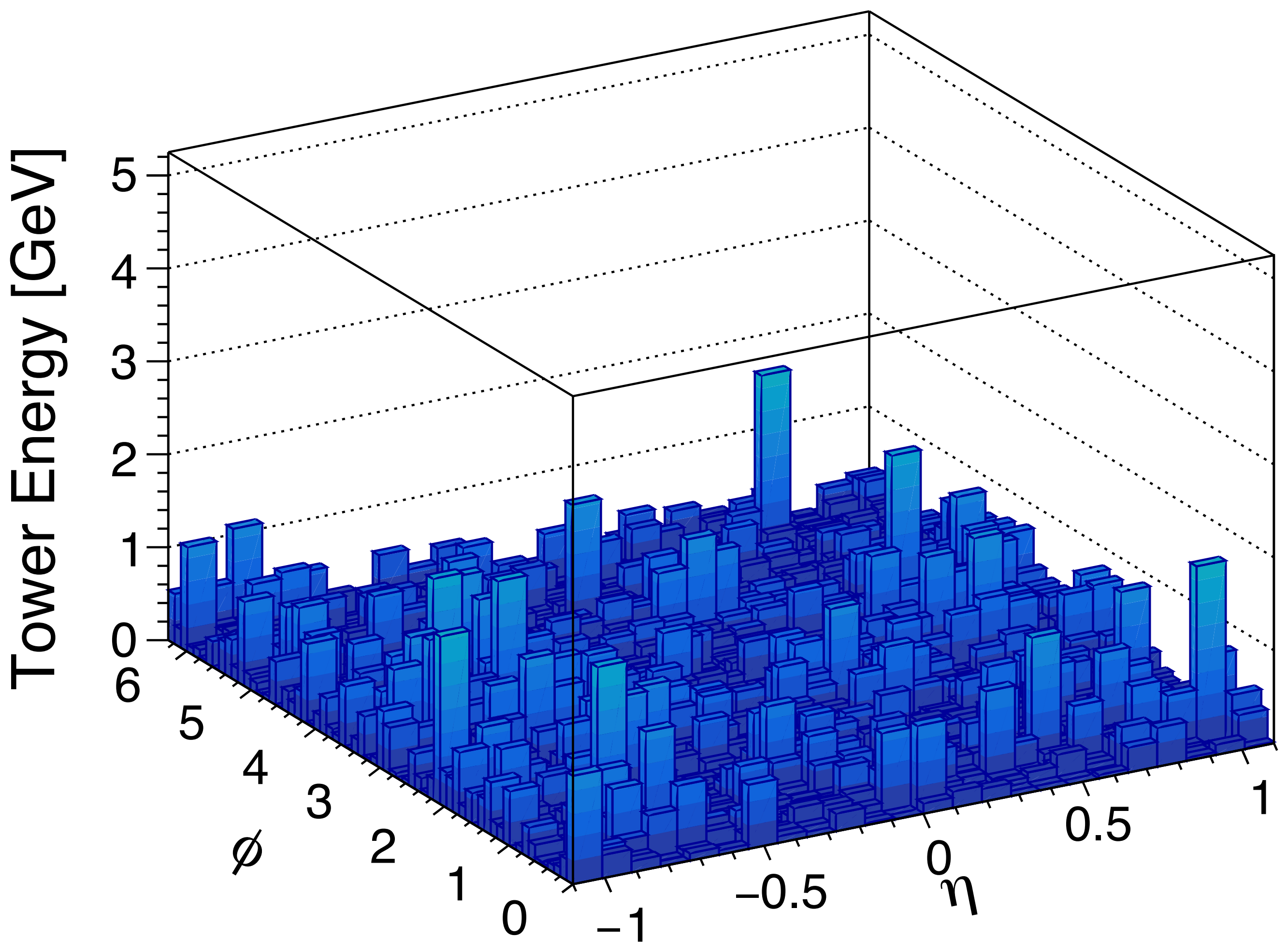
Display of Generated Events

Training sample (HIJING+GEANT4)

0-10% Centrality



40-50% Centrality

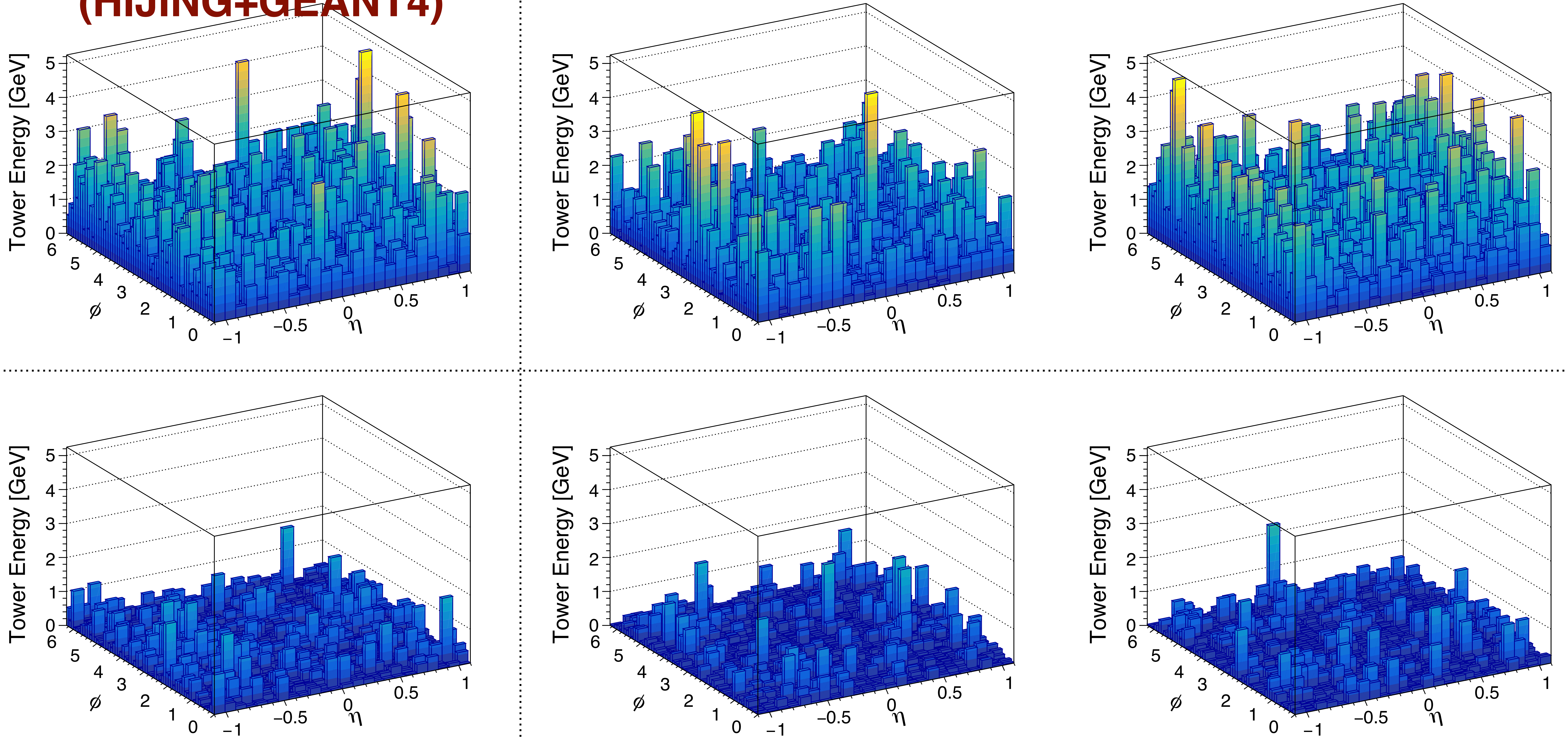


Display of Generated Events

Training sample (HIJING+GEANT4) | **Generated (DDPM)** | **Generated (GAN)**

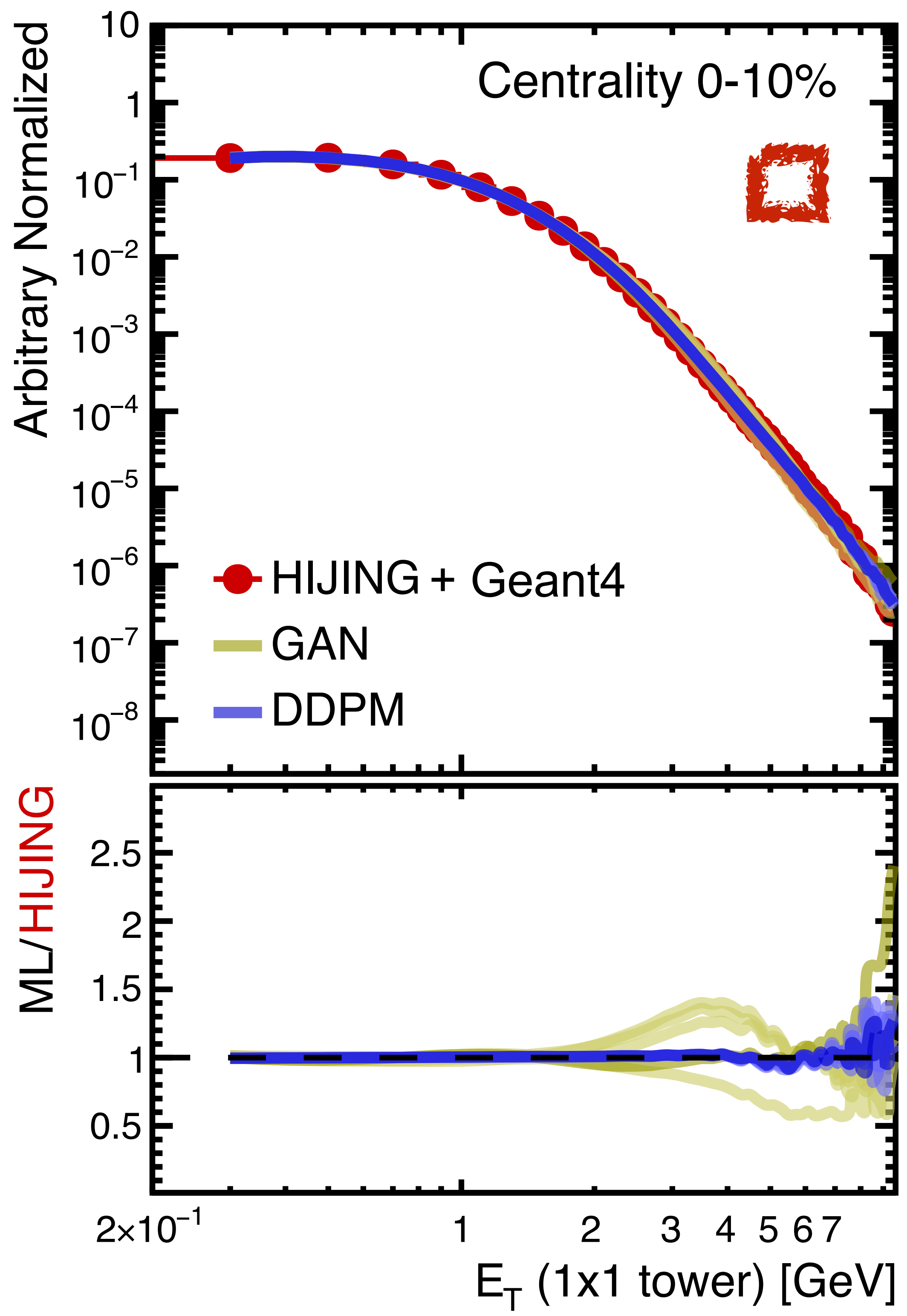
0-10% Centrality

40-50% Centrality

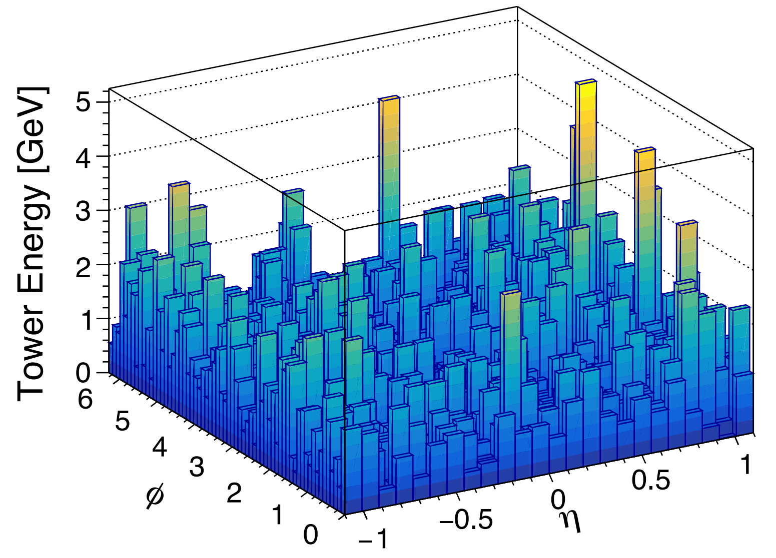


Performance: Transverse Energy (0-10%)

1x1 Tower | 4x4 Tower | 7x7 Tower | 11x11 Tower | All Towers

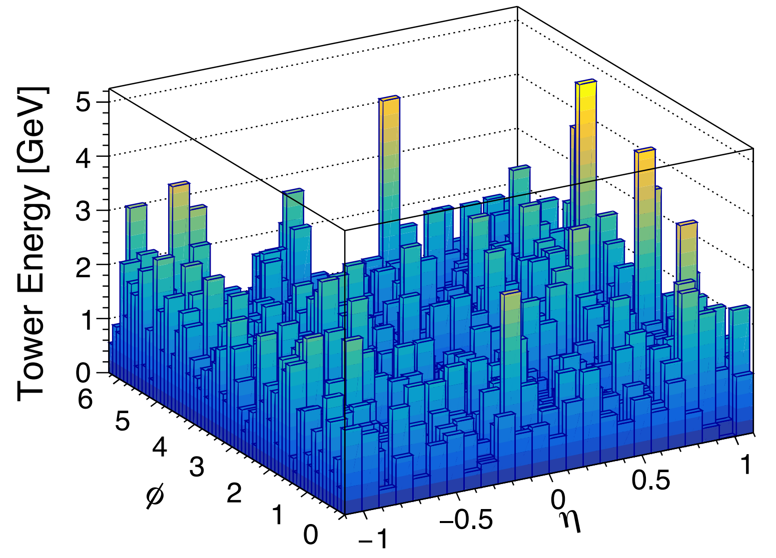


- Each model is retrained 5 times with different random seeds
- HIJING+Geant4 used as training data (600k events) and testing data (100k events)
- Both DDPM and GAN reproduce the data distribution where the data are abundant
- DDPM outperforms GAN in overall distribution w/ great stability and accuracy



Performance: Transverse Energy (0-10%)

- Each model is retrained 5 times with different random seeds
- HIJING+Geant4 used as training data (600k events) and testing data (100k events)
- Both DDPM and GAN reproduce the data distribution where the data are abundant
- DDPM outperforms GAN in overall distribution w/ great stability and accuracy



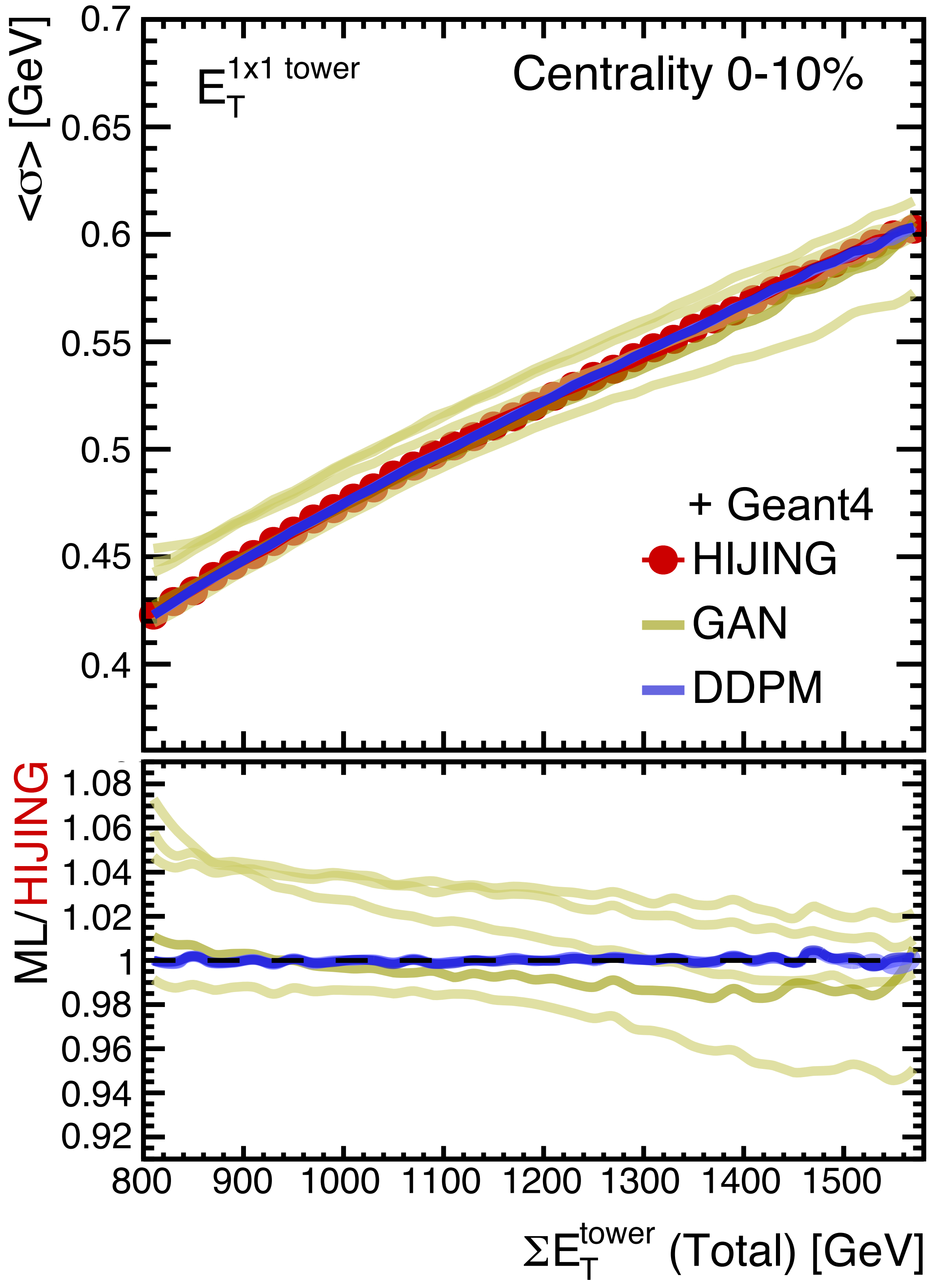
Performance: Transverse Energy Fluctuation (0-10%)

1x1 Tower

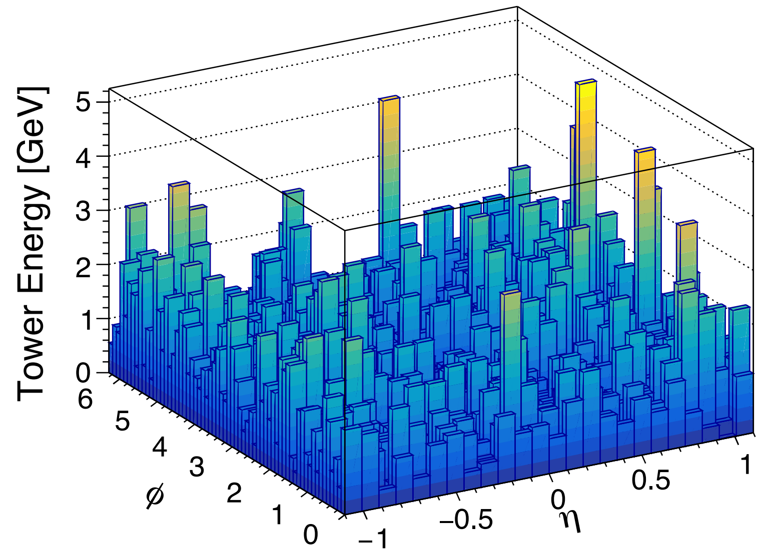
4x4 Tower

7x7 Tower

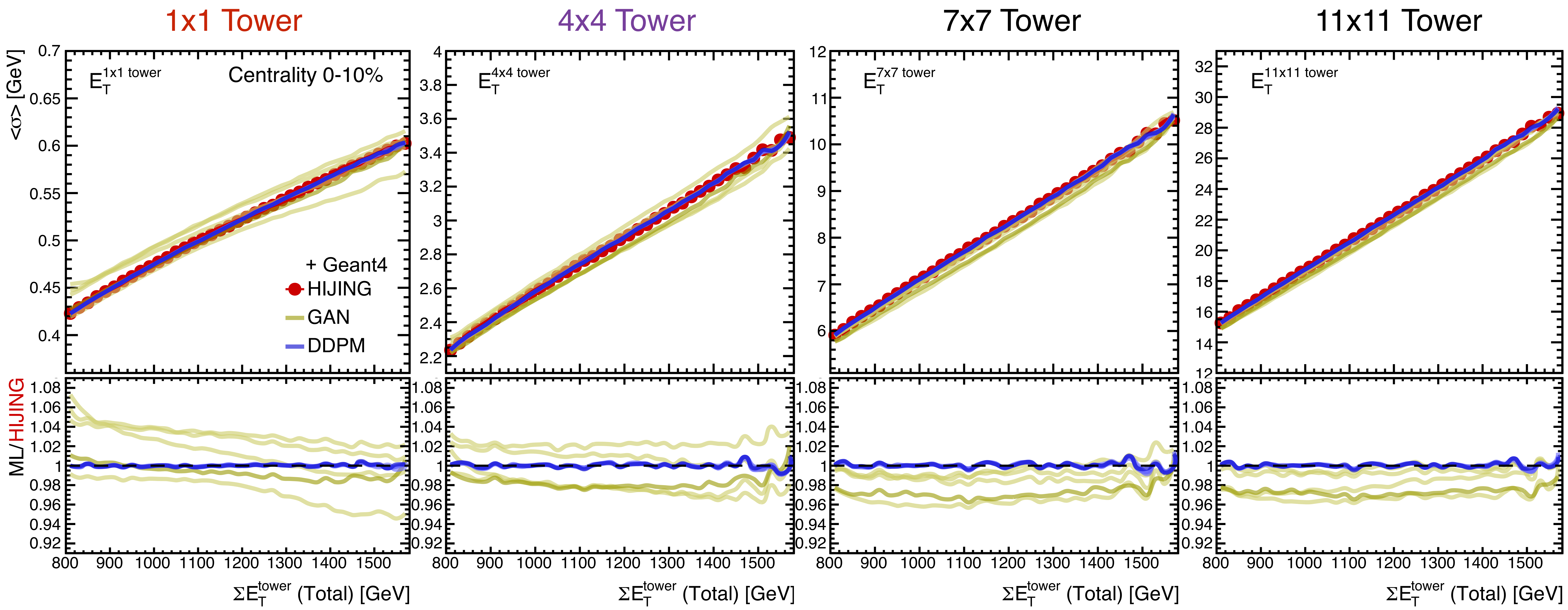
11x11 Tower



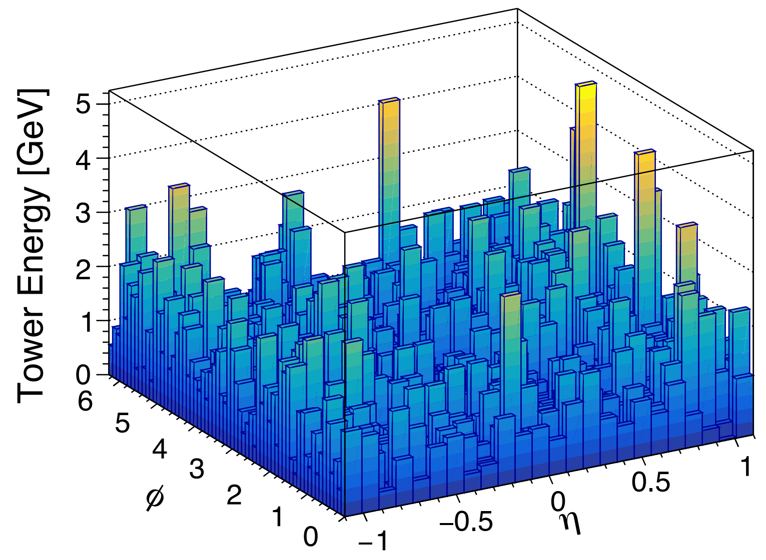
- GAN fails to describe fluctuation
- DDPM outperforms GAN w/ great stability, a few percent-level accuracy



Performance: Transverse Energy Fluctuation (0-10%)

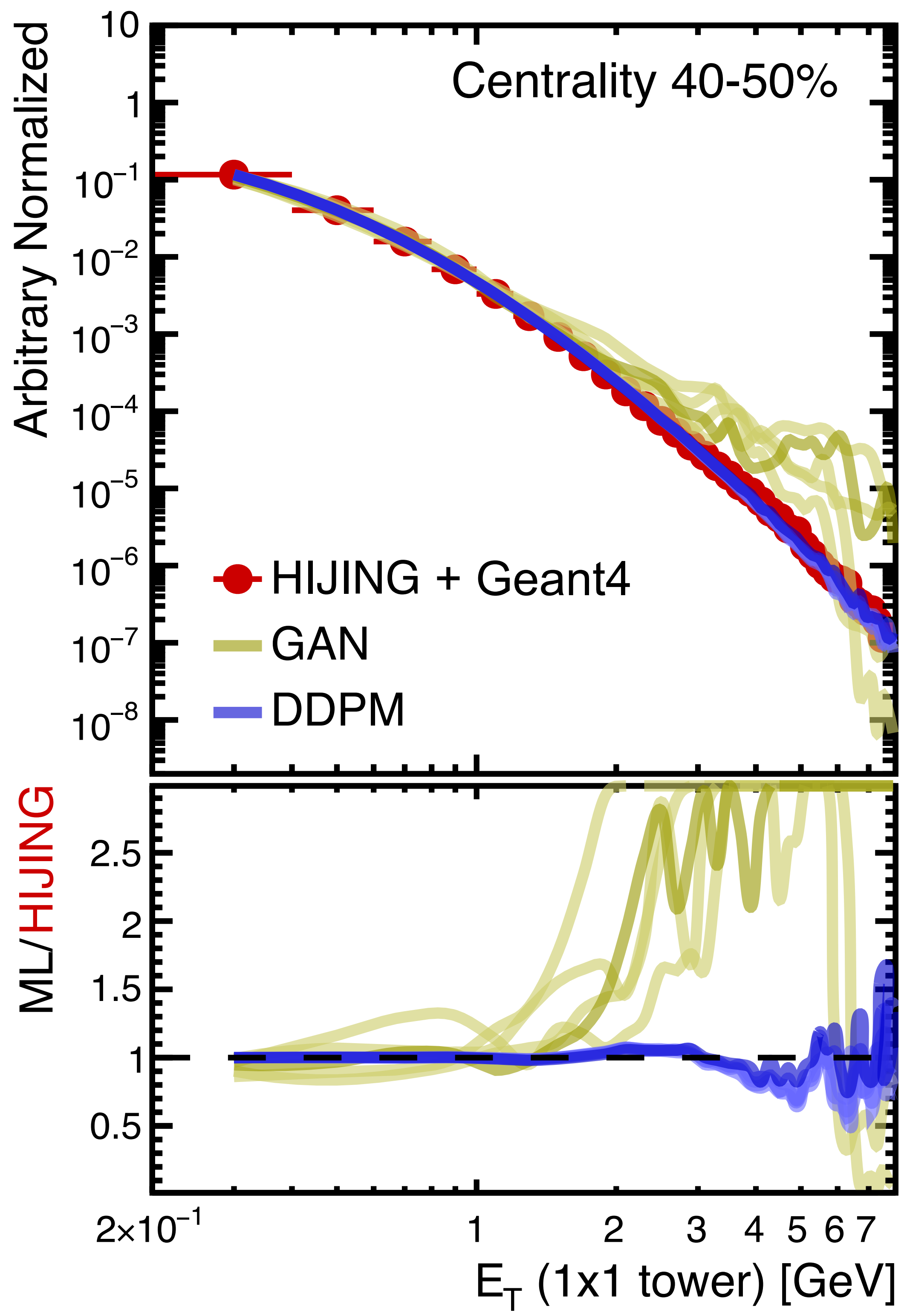


- GAN fails to describe fluctuation
- DDPM outperforms GAN w/ great stability, a few percent-level accuracy

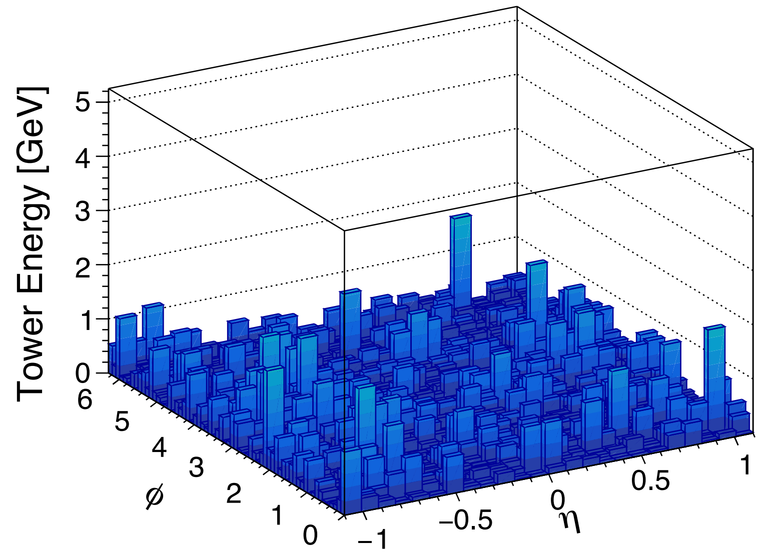


Performance: Transverse Energy (40-50%)

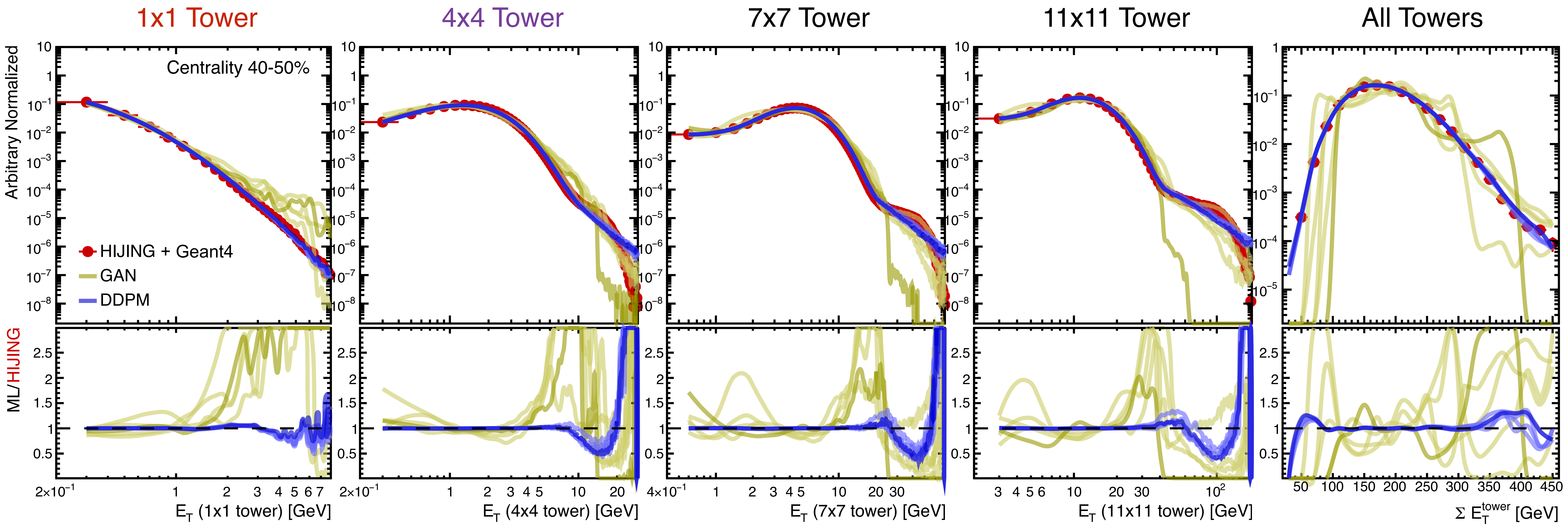
1x1 Tower　　4x4 Tower　　7x7 Tower　　11x11 Tower　　All Towers



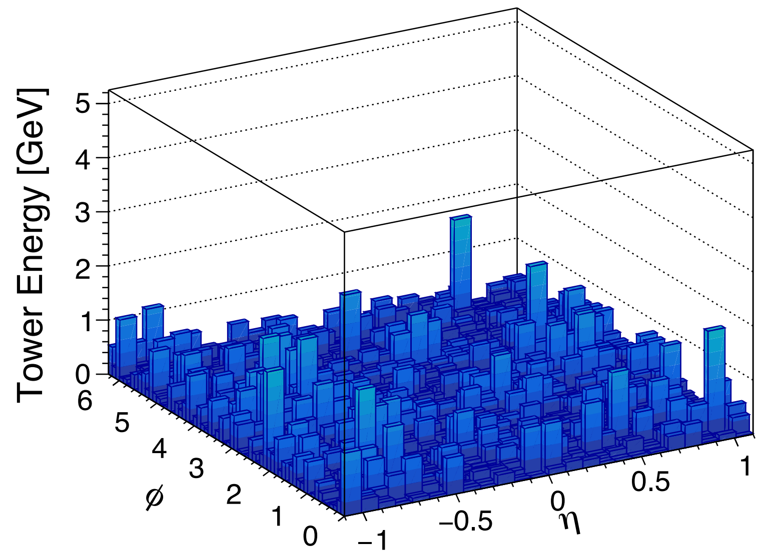
- DDPM outperforms GAN
  - ➡ great stability, good agreement with HIJING+G4 at high probability region



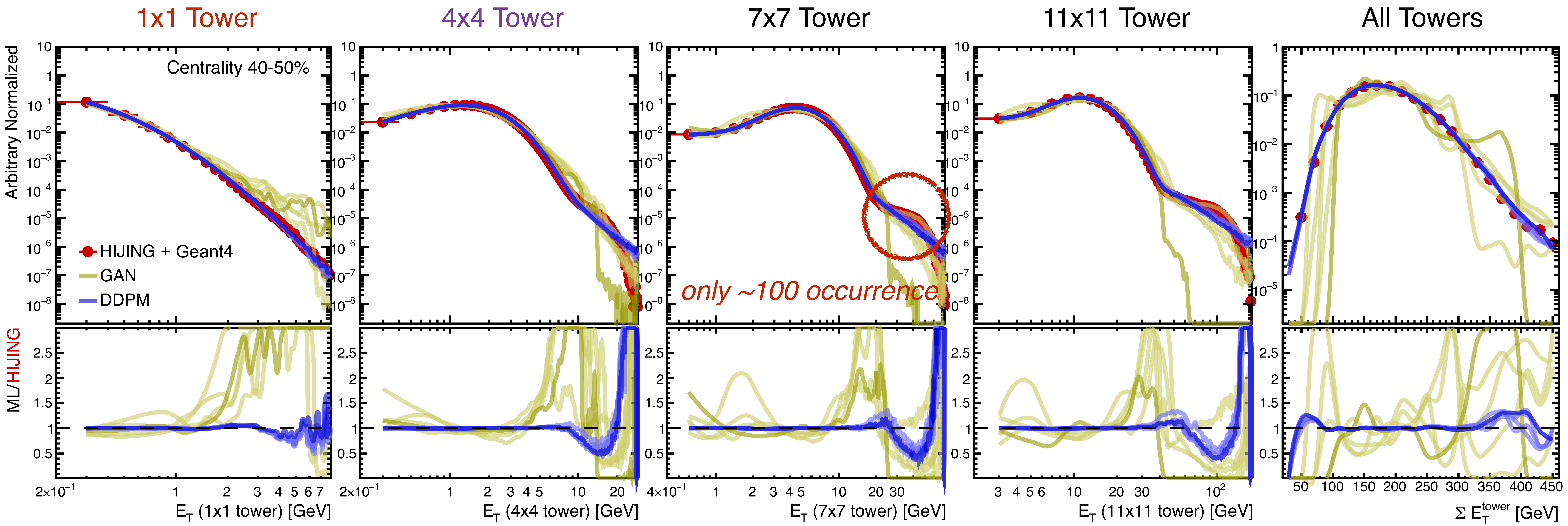
Performance: Transverse Energy (40-50%)



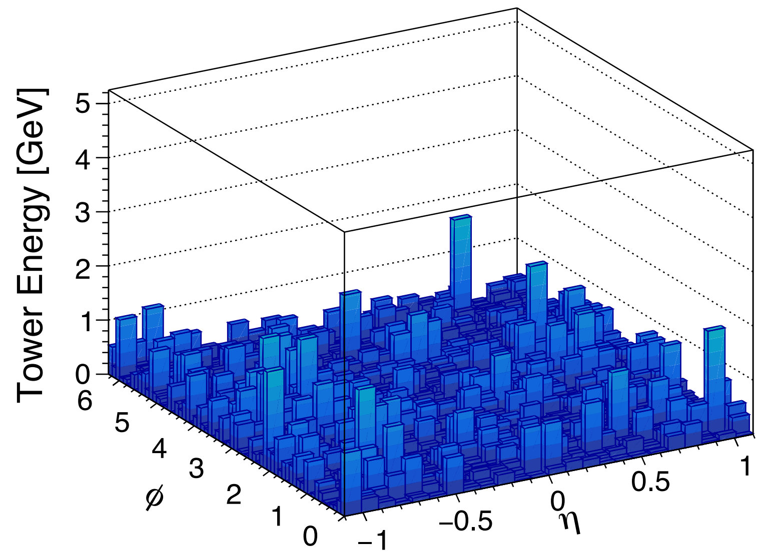
- DDPM outperforms GAN
  - ➡ great stability, good agreement with HIJING+G4 at high probability region



Performance: Transverse Energy (40-50%)

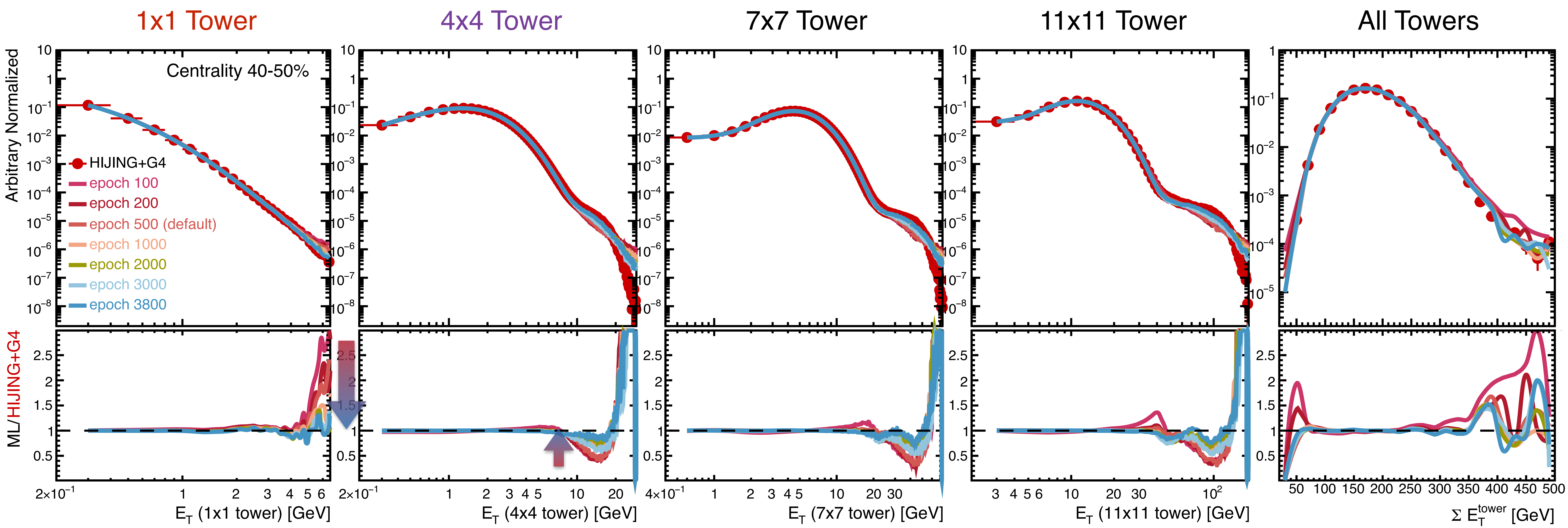


- DDPM outperforms GAN
  ➡ great stability, good agreement with HIJING+G4 at high probability region

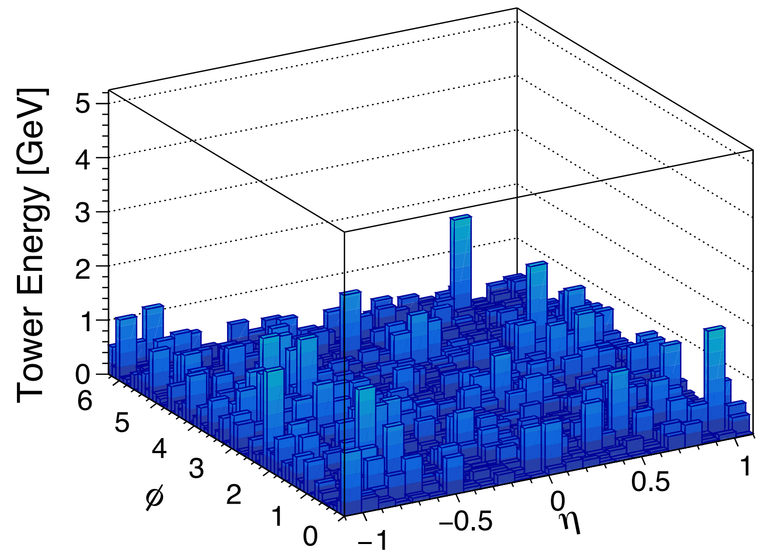


- Non-gaussian rare tail at the high energy region → challenge to reproduce

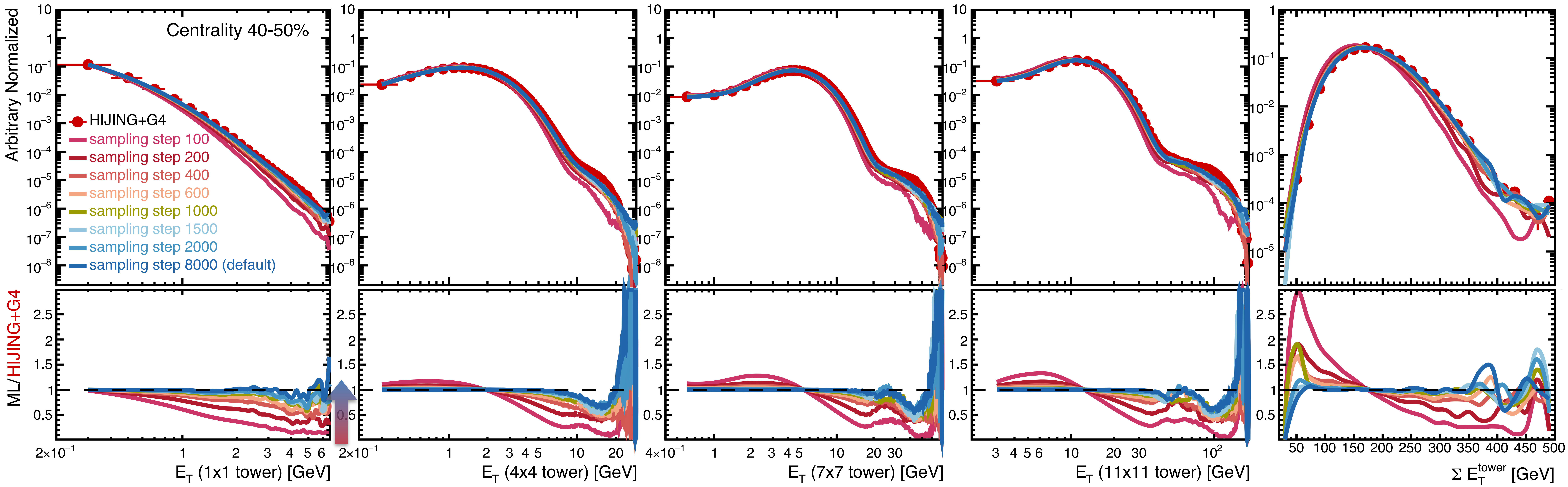
Trade-off between Training time and Fidelity



- epoch ~ training duration
- **DDPM** models with the higher epochs give **better performance**!
  - ➡ but, the higher the epochs, the *longer the training time*



Trade-off between Generation time and Fidelity



- **DDPM** models with the higher de-noising steps give **better performance**!
  ➡ but, the higher the de-noising, the *longer the generation time*

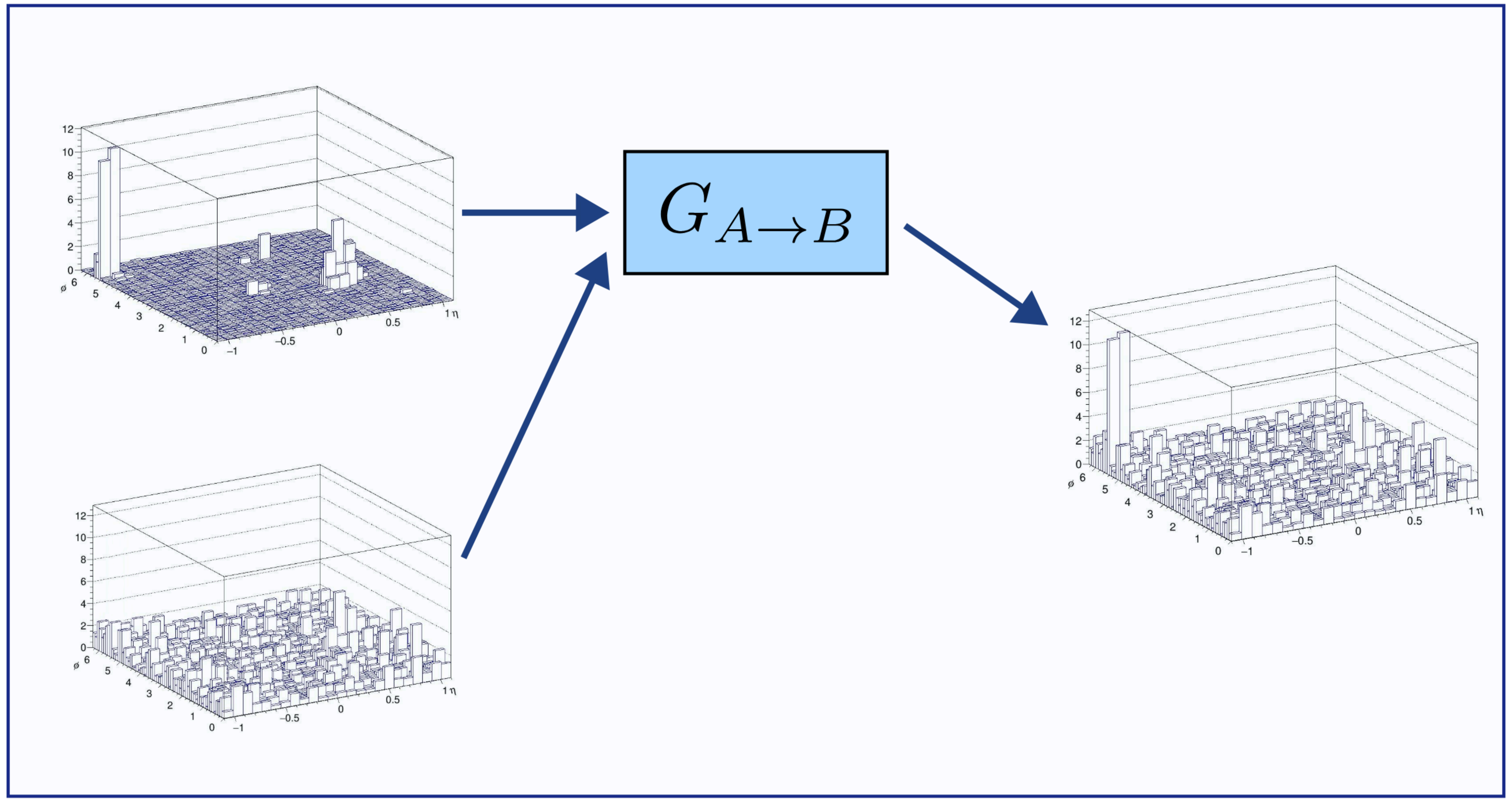
How long does it take to simulate a large sample?

	Generating time	Speedup	CPU/GPU
HIJING + GEANT4 (Conventional)	40 minutes / event	1	Single CPU
DDPM	1.34 s / event	~1,800X	NVIDIA RTX A6000
GAN	0.42 ms / event	~5,700,000X	NVIDIA RTX A6000

- GAN is faster, but the DDPM exhibits high fidelity in describing the truth ground (HIJING+GEANT4)
- DDPM provide a speedup of O(100), considering a 32-core CPU equivalent to a GPU

Applications

- There is a need for a large number of full-detector simulation events, e.g.
  - ➡ heavy ion collisions
  - ➡ beam background events for EIC detectors
- Train the model using a relatively modest number (at the level of millions) and then **accelerate the production** of much larger samples (at the level of billions)
- Rare signal events, e.g. jets, can be embedded in these simulation events



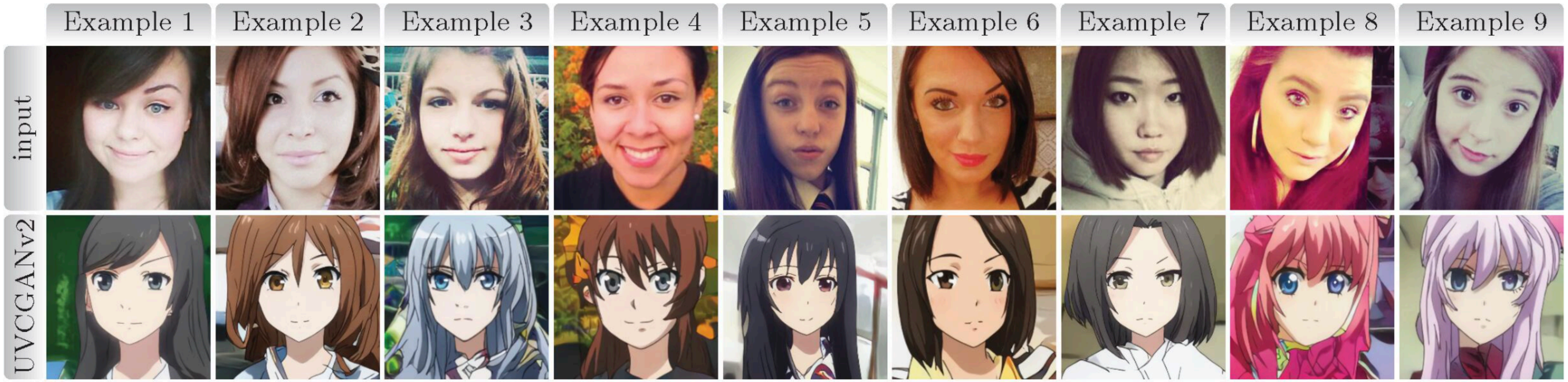
Applications - Jet Background Subtraction (1)

- **UVCGAN** (UNet Vision Transformer cycle-consistent Generative Adversarial Network)
  - ➡ **unpaired** image-to-image translation; bridging gap between simulation and data reference
  - ➡ arXiv:2203.02557 [cs.CV]

Selfie2Anime and Anime2Selfie (pdf)

A

B

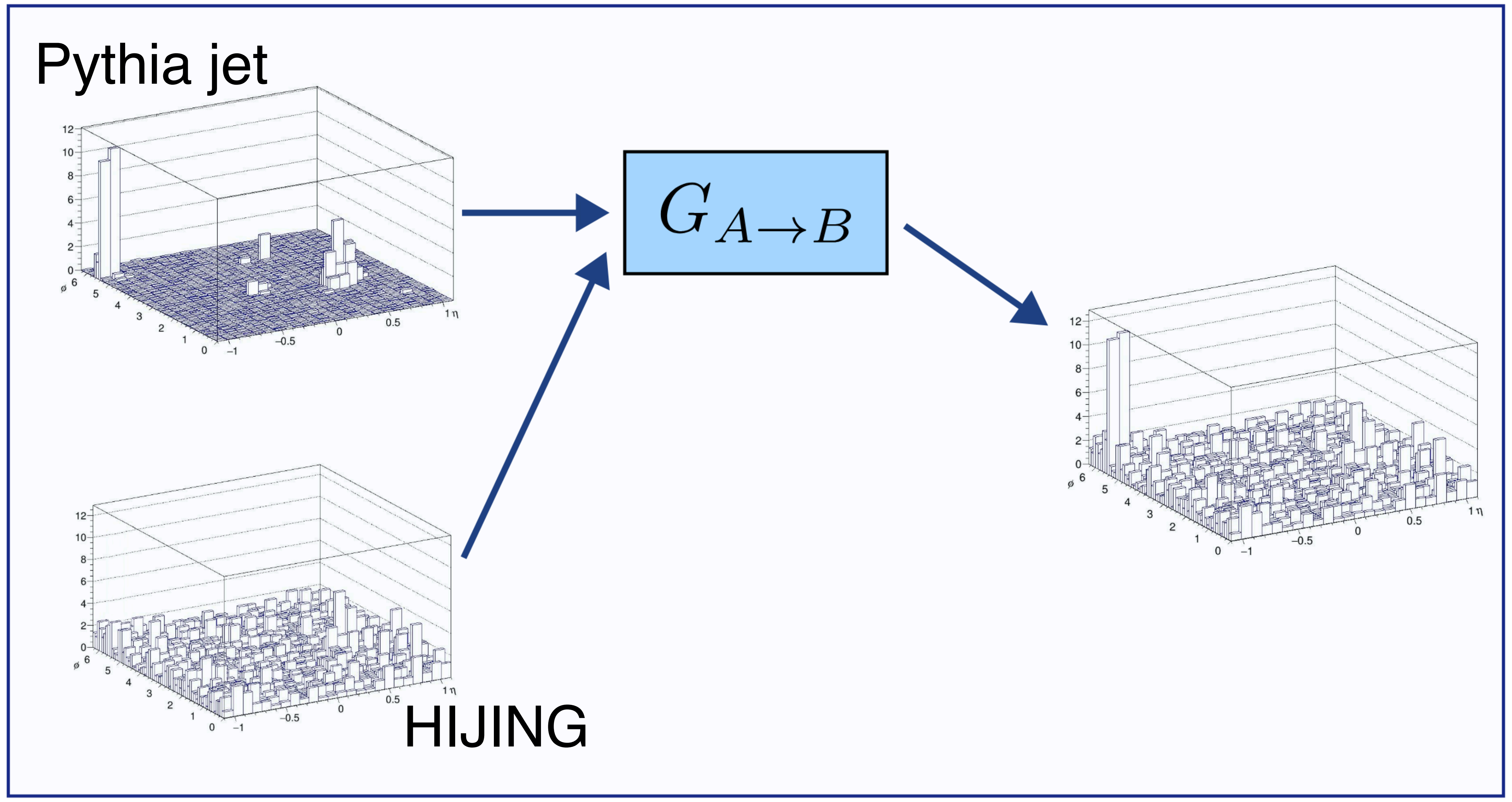


Applications - Jet Background Subtraction (2)

- Calorimeter η vs ϕ images

B

A



Input

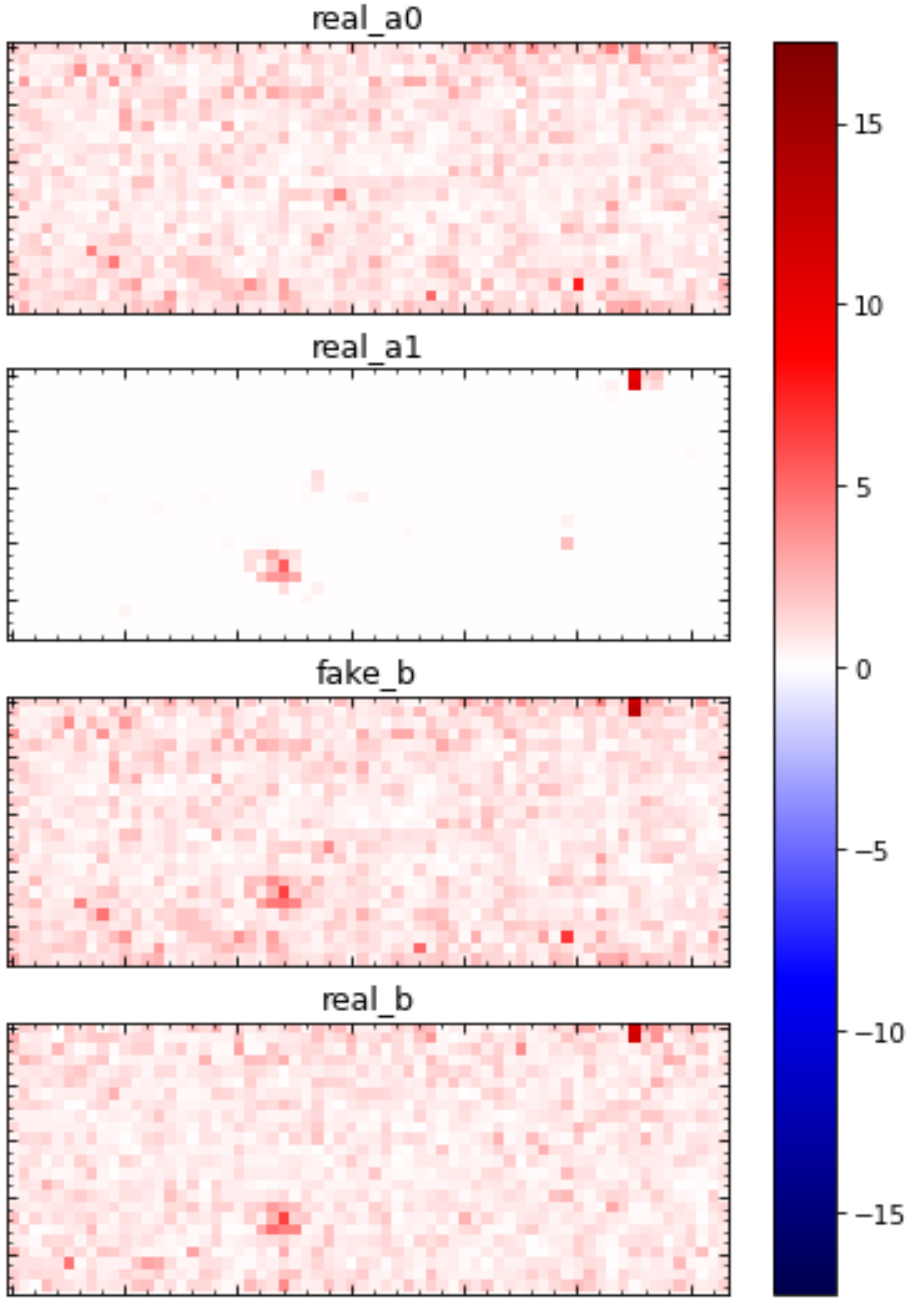
HIJING

Pythia jet

A→B Generated by UVCGAN

Reference Pythia + HIJING

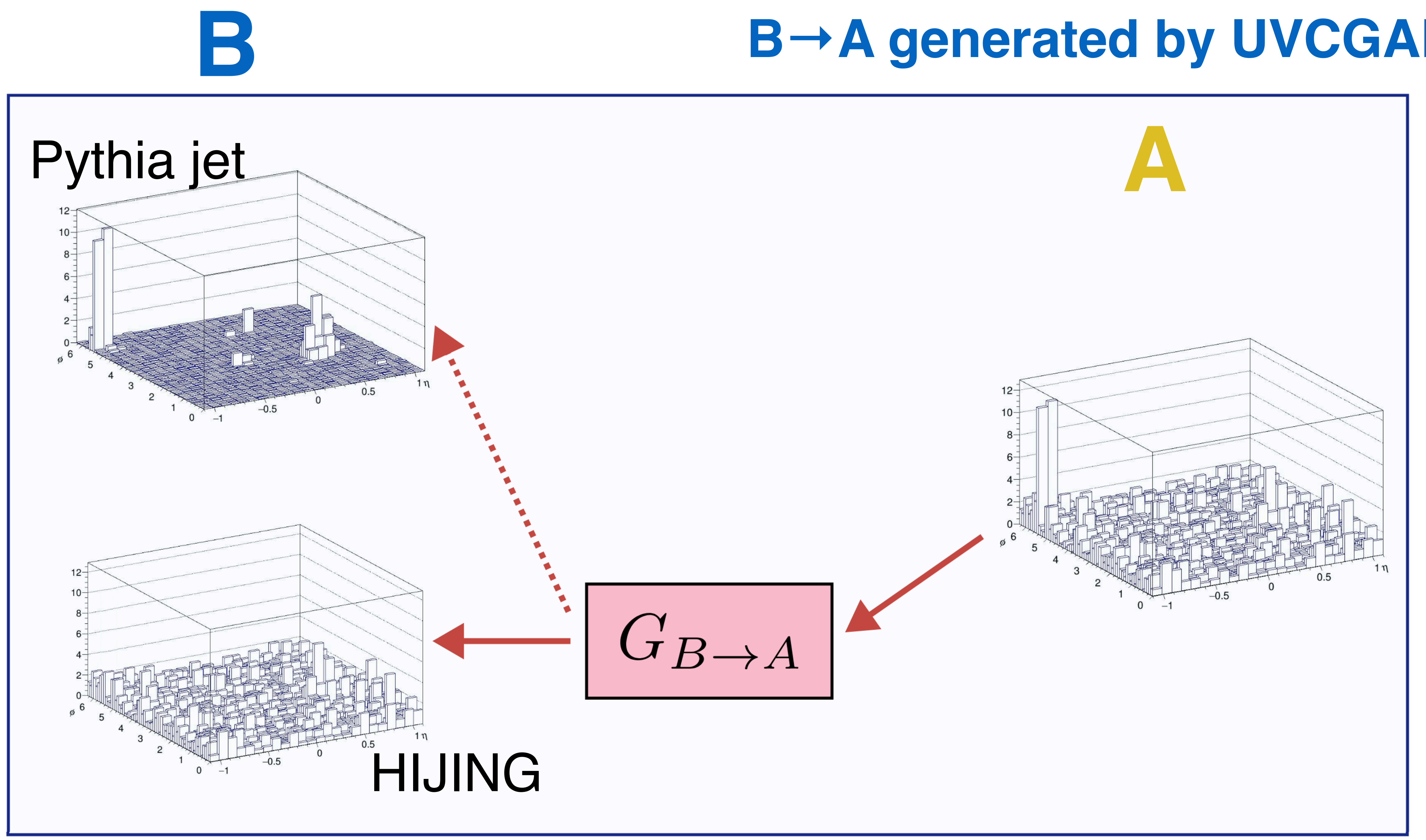
Work-in-progress

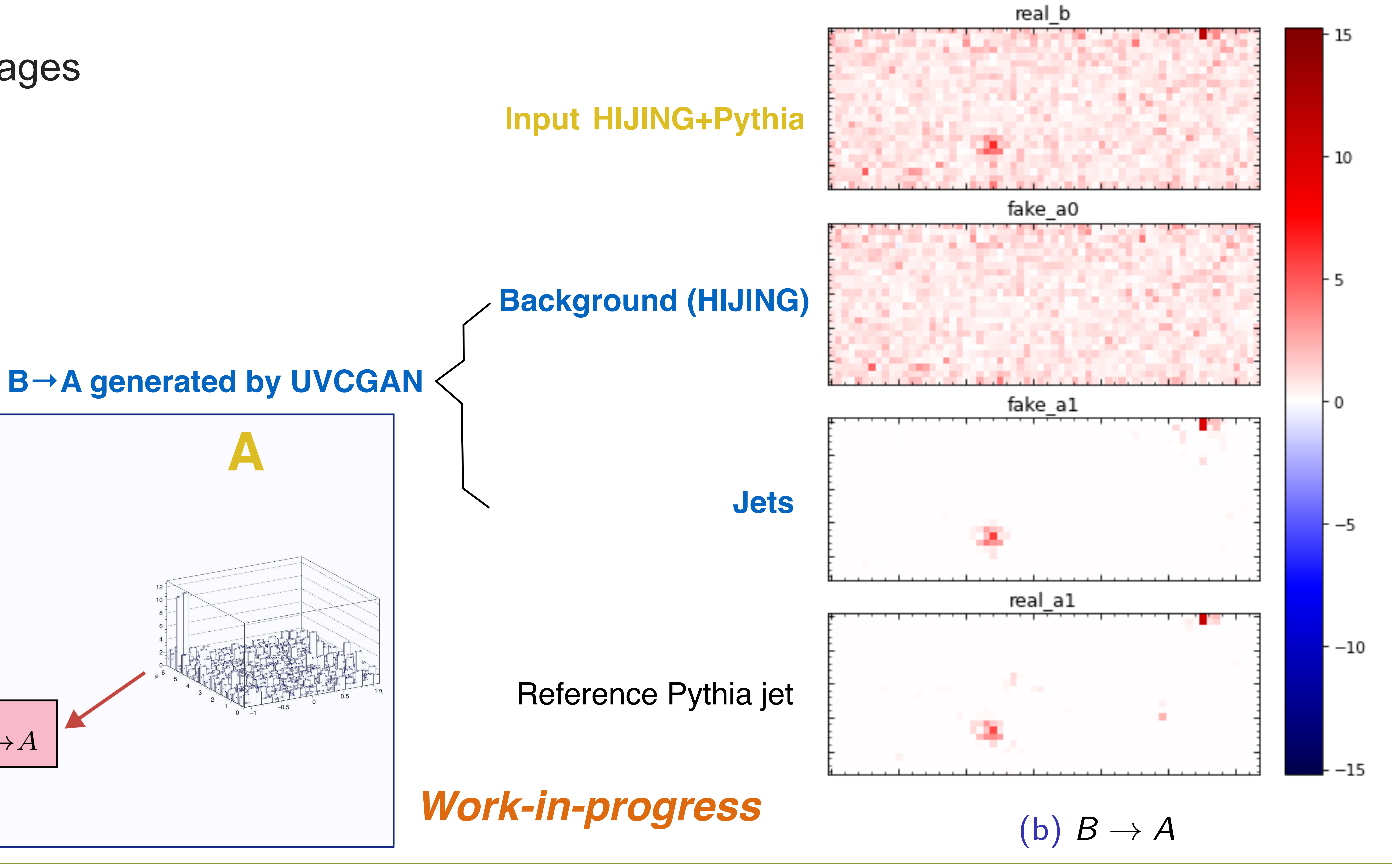


(a) $A \to B$

Applications - Jet Background Subtraction (3)

- Calorimeter η vs ϕ images





Work-in-progress

Conclusion

- Simulations of high energy nuclear experiments
  ➡ **highly complex** and **computationally intensive**
  ➡ both fidelity and speed is important
  ➡ *Generative AI can speed up and produce large amount of the heavy ion event simulations!*

Conclusion

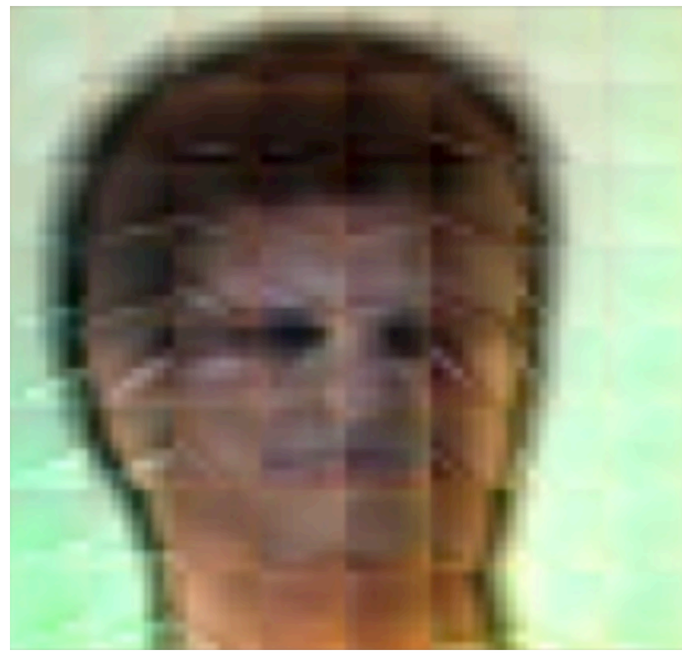
- Simulations of high energy nuclear experiments
  ➡ **highly complex** and **computationally intensive**
  ➡ both fidelity and speed is important
  ➡ *Generative AI can speed up and produce large amount of the heavy ion event simulations!*

- **Diffusion model (DDPM)** was used to generate the whole-event full-detector simulated calorimeter data in high fidelity for the first time in heavy ion collisions
  ➡ GAN used as a reference
  ➡ DDPM outperforms GAN for scientific fidelity
  ➡ trade-off found between training / generation duration and the quality of reproducing the rare feature

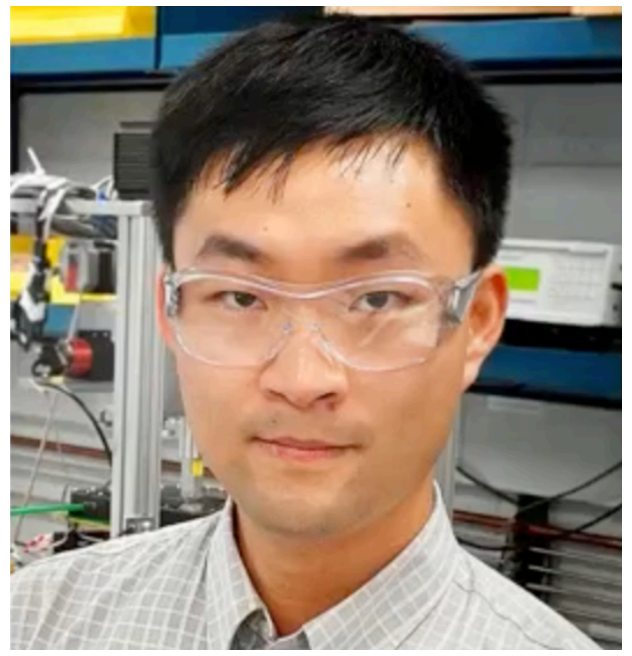
- Paper has been submitted to PRC and available: arXiv:2406.01602

Our Team

Yeonju Go



Dmitrii Torbunov



Jin Huang

Yihui Ren

- Contacts: {ygo, dtorbunov, jhuang, yren}@bnl.gov

Tim Rinn

Yi Huang

Haiwang Yu

Shinjae Yoo

Meifeng Lin

Brett Viren

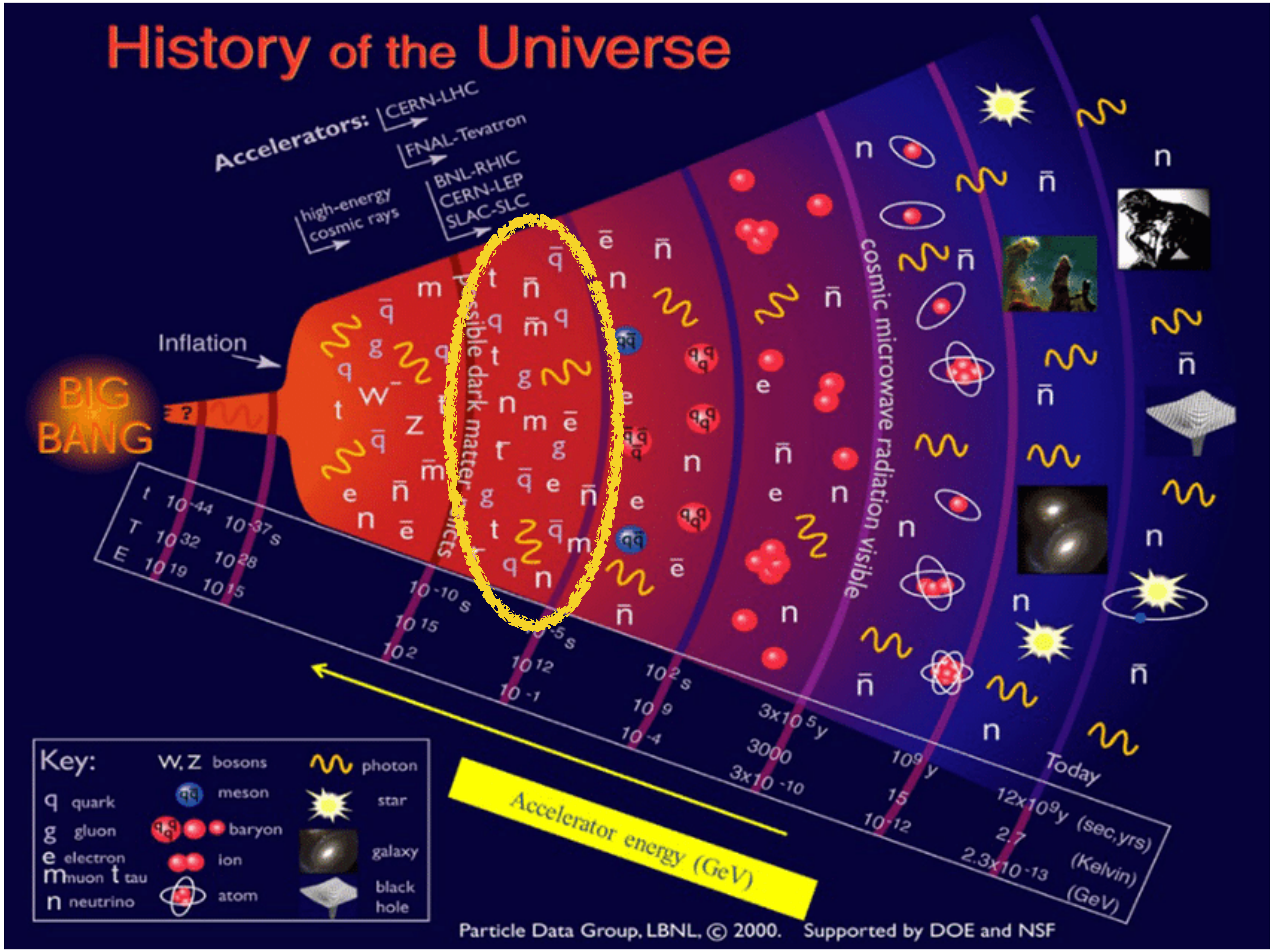
Acknowledgement

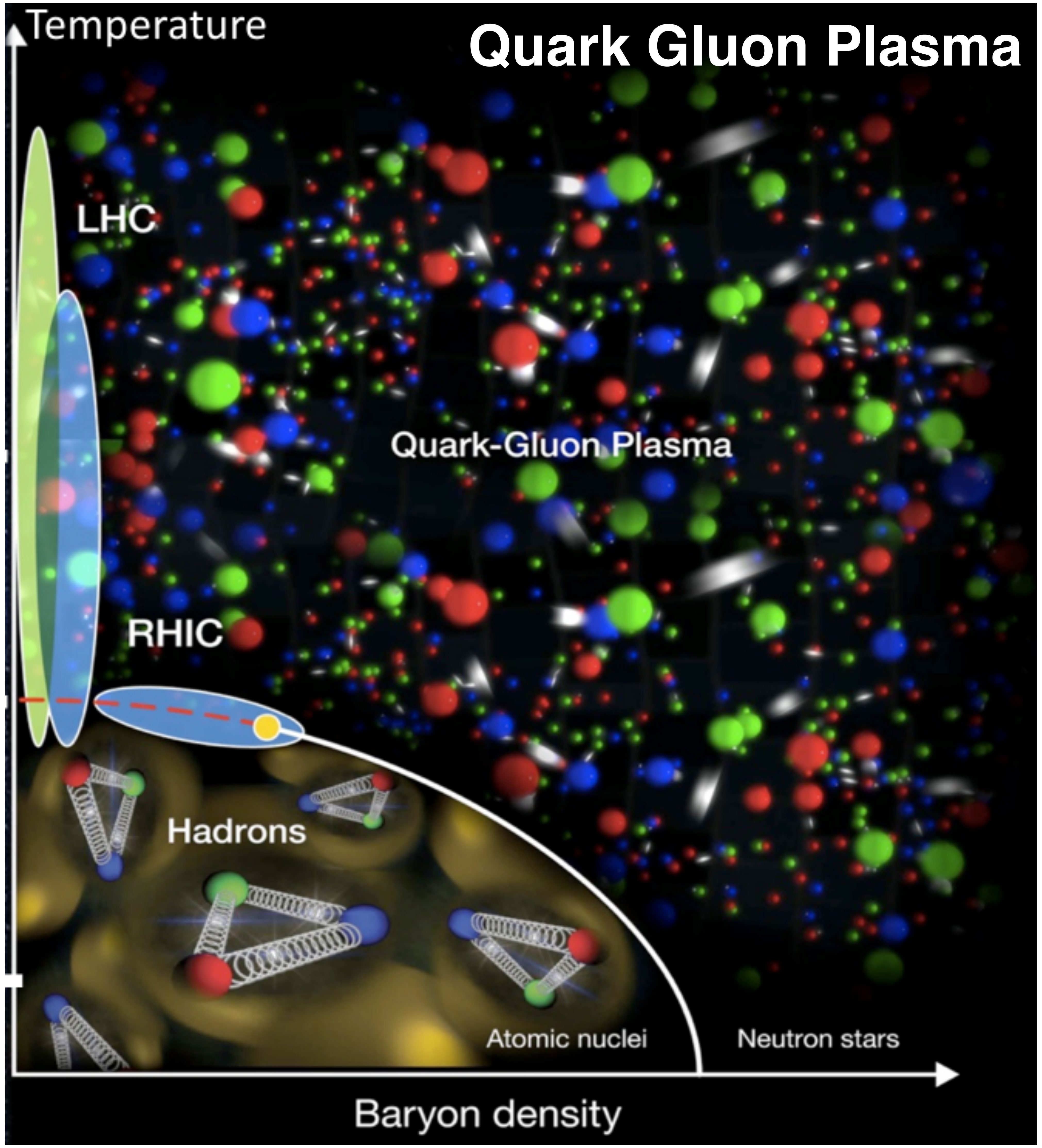
- The LDRD Program at Brookhaven National Laboratory, sponsored by DOE's Office of Science under Contract DE-SC0012704, supported this work.

- We thank the sPHENIX collaboration for access to the simulated dataset, which was used in the training and validation of our algorithm.

BACKUP

Early Universe and Quark Gluon Plasma

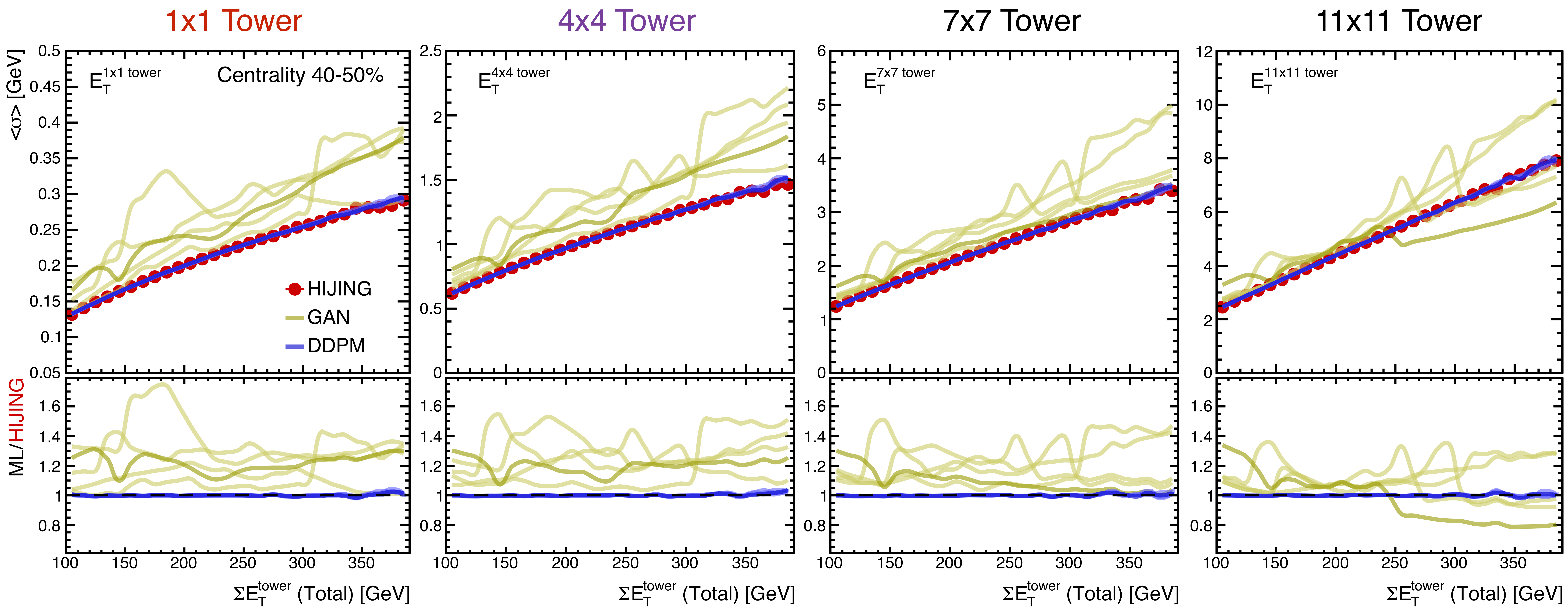




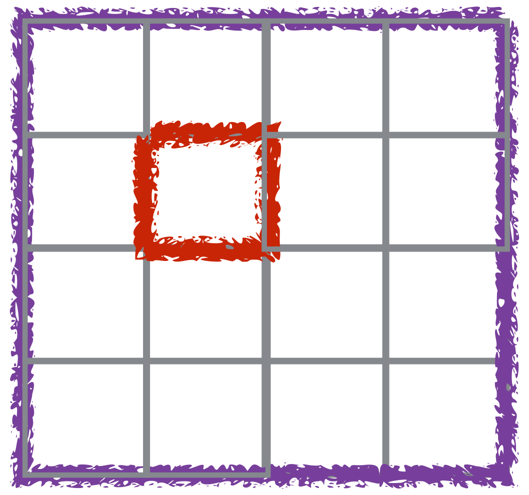
DDPM Configuration

- number of diffusion steps T: default 8000 / variation [1000, 16000]
- variance schedule β_t: default 0.1 / variation [0.02, 0.2]
- training batch size: default 128 / variation [100, 12500]
- training steps per epoch: default 2000
- epoch: default 4000 / variation [100, 4000]
- training with the Adam optimizer with learning rate 10^{-4}
- trained with 600,000 events per each centrality bin
- tested with 100,000 events per each centrality bin
- neural network architecture (U-ResNet + Attention)
- depth/width of the model
  ➡ U-Net encoder-decoder stage, channels per stage: 32, 64, 128
  each of which comprised of two ResNet blocks

Performance: Transverse Energy Fluctuation (40-50%)



- DDPM outperforms GAN
  - ➡ great stability, a few percent-level accuracy



Batch Size Dependence

- Batch size not only introduces different random seeds and but also changes variance schedule (β_t)