Generative Models for Fast Simulation of Electromagnetic and Hadronic Showers in Highly Granular Calorimeters

ML4Jets 2022

Erik Buhmann, **Sascha Diefenbacher**, Engin Eren, Frank Gaede, Daniel Hundhausen, Gregor Kasieczka, William Korcari, Anatolii Korol, Claudius Krause, Katja Krüger, Peter McKeown, Lennart Rustige, Imahn Shekhzadeh, David Shih

sascha.daniel.diefenbacher@uni-hamburg.de

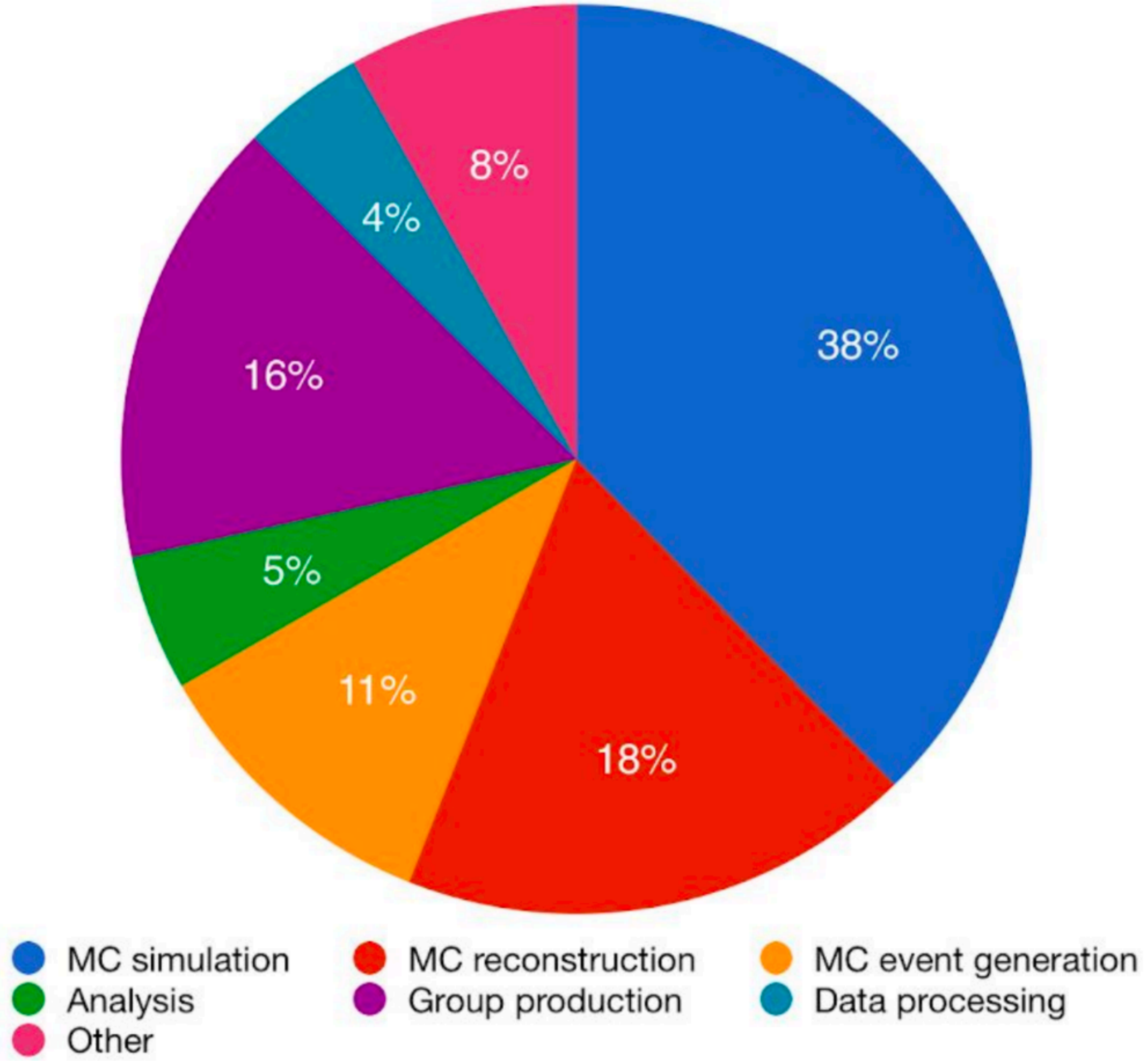
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

DESY.

Generative Simulation



- MC simulation large part of computing
- Speed up:
  - Train ML model on small dataset
  - Draw majority of samples form ML model
    - ➔Amplify original data set
    - ➔Significantly faster

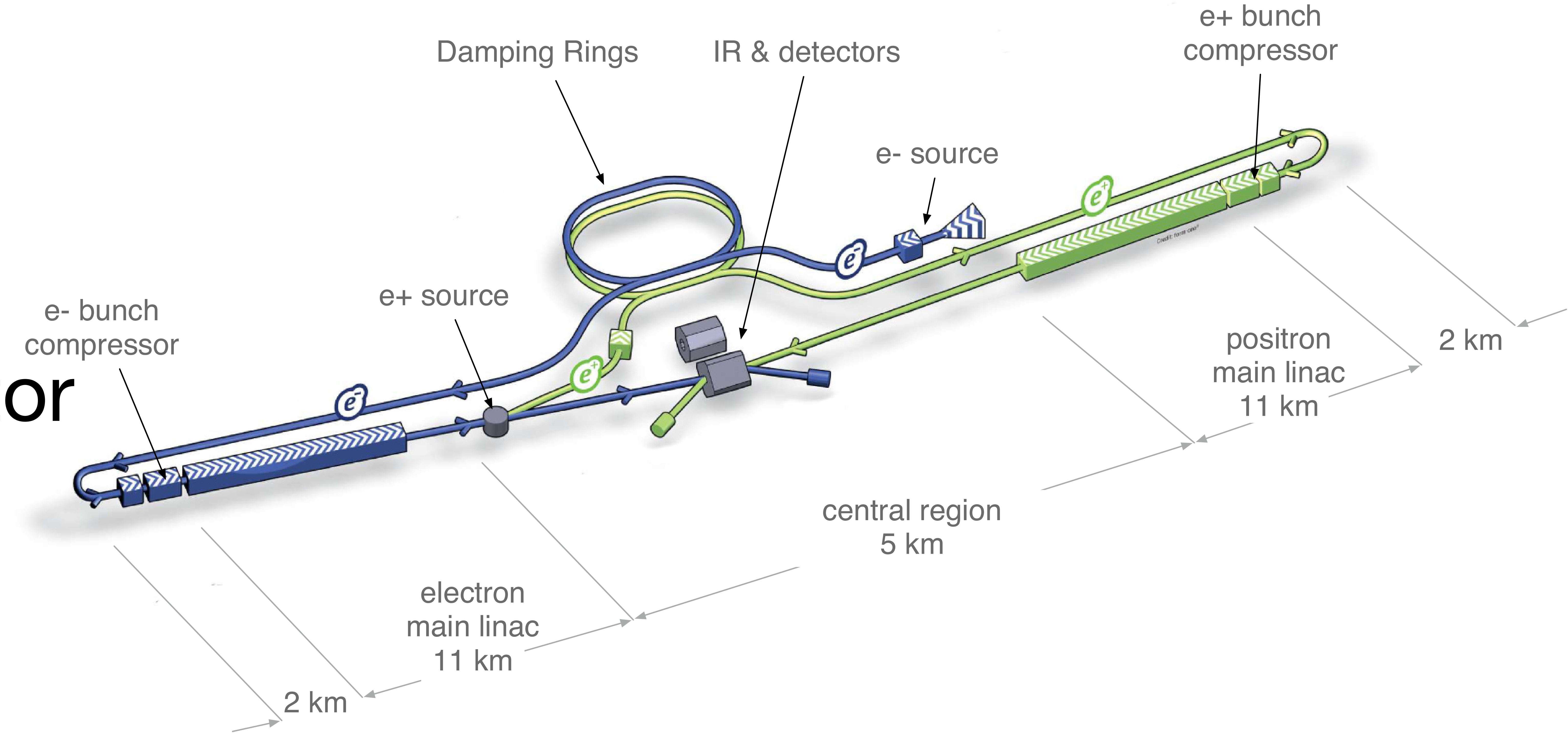
Catmore et. al. **ATLAS HL-LHC Computing Conceptual Design Report,** CERN-LHCC-2020-015 ; LHCC-G-178

Butter et al.: **Amplifying Statistics using Generative Models:** NeurIPS ML4PS 2020, 2008.06545

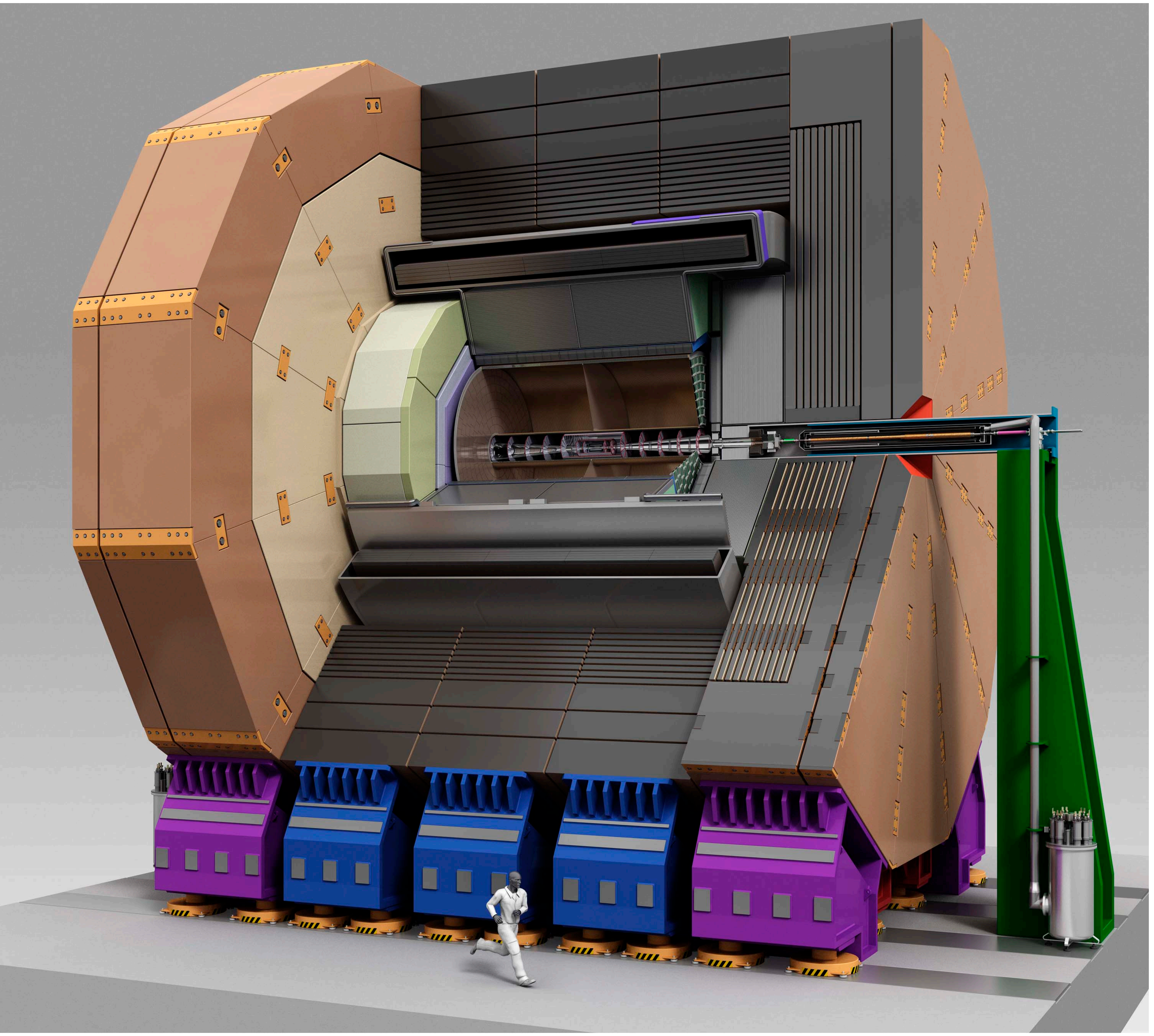
International Linear Collider

- Proposed linear lepton collider
- Designed for precision EW and Higgs measurements

- Dual detector setup
  - International Large Detector
  - Silicon Detector
- Optimised for Particle Flow reconstruction methods



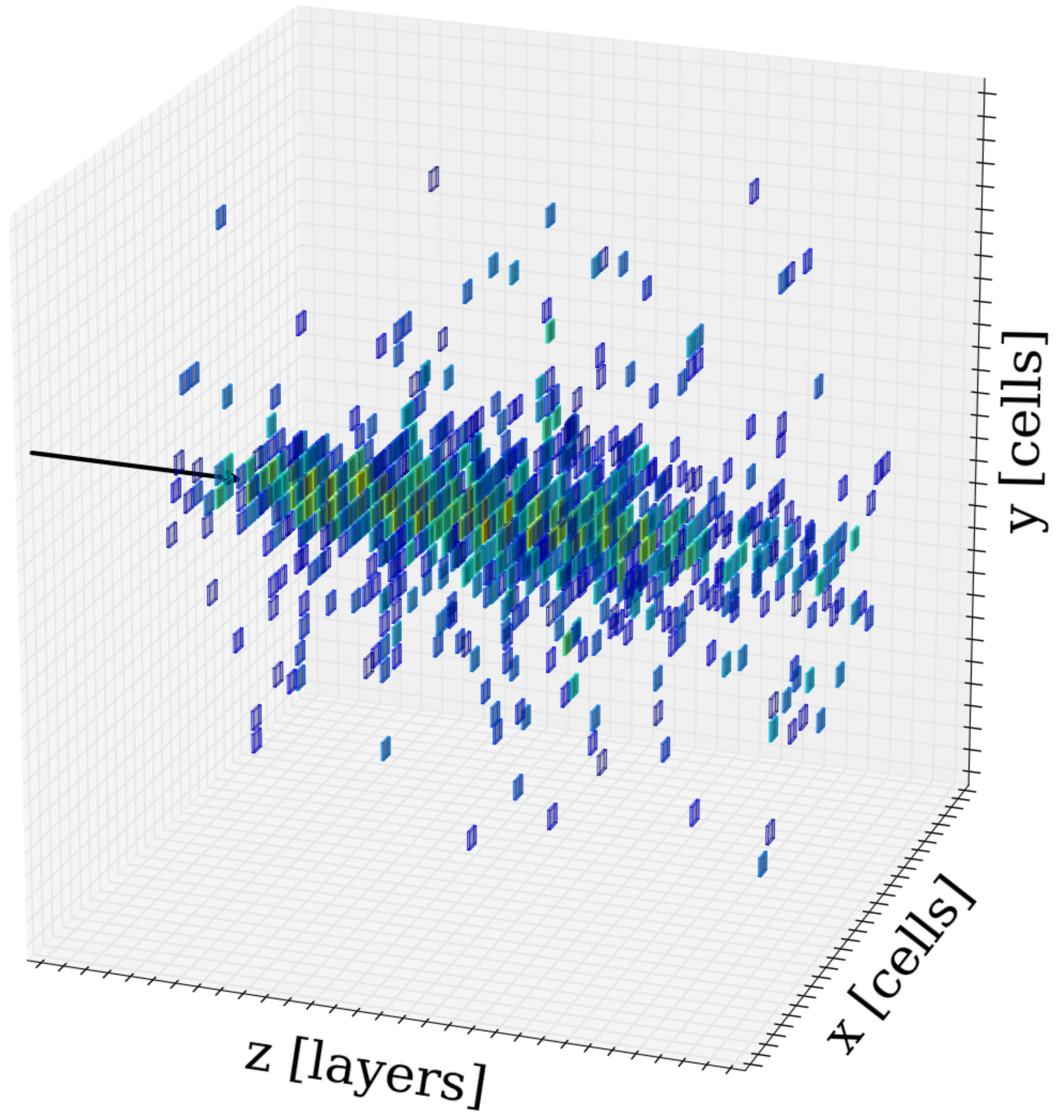
ILD Calorimeters



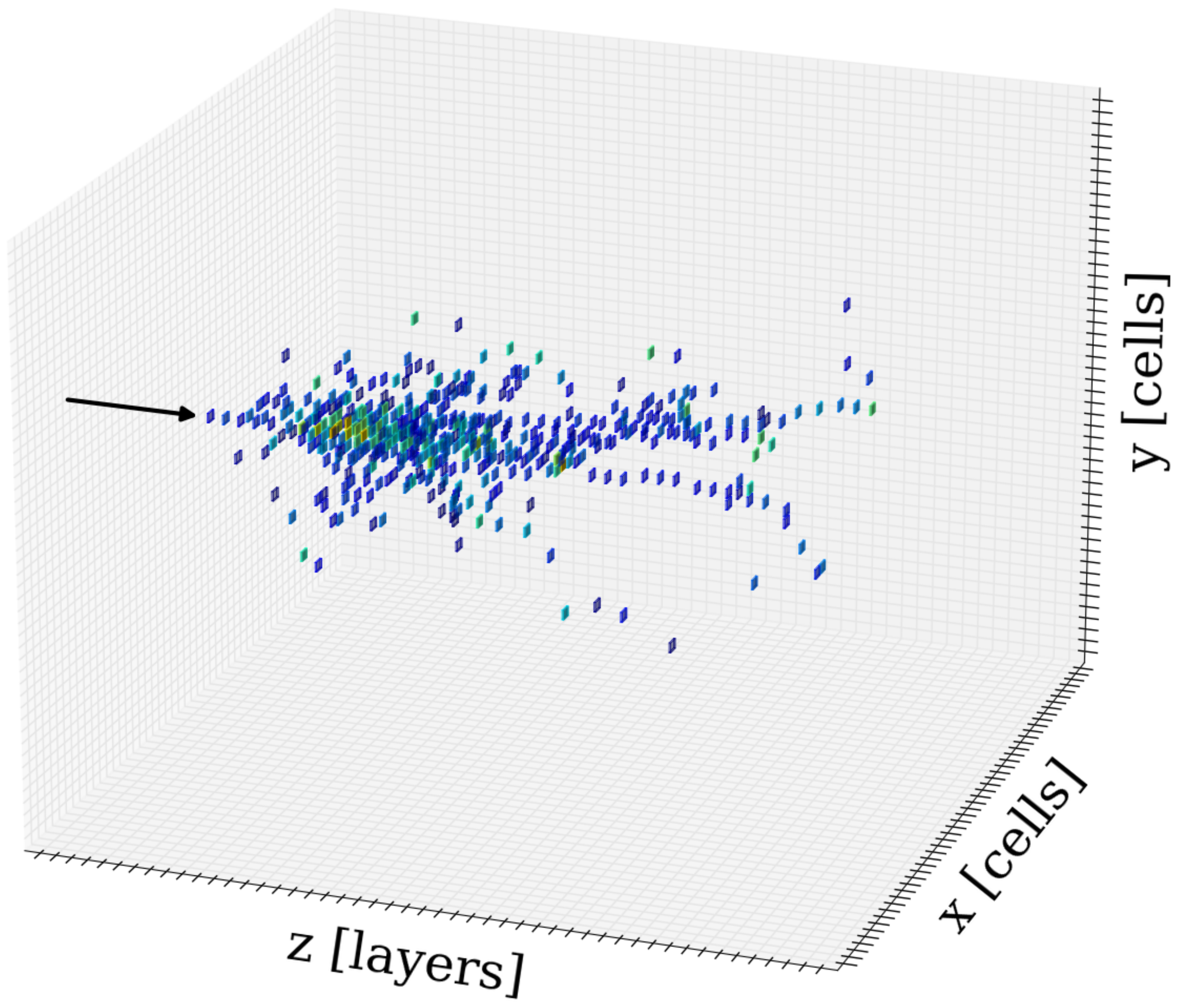
- Particle Flow optimised
  - Requires highly granular calorimeters
- ILD electromagnetic calorimeter
  - Active silicon, passive tungsten
  - 30 layers, 5mm x 5mm cells
- ILD hadronic calorimeter
  - Active scintillator, passive iron
  - 48 layers, 30mm x 30mm cells

Shower Dataset

Photon shower



Charged pion shower



Training data:

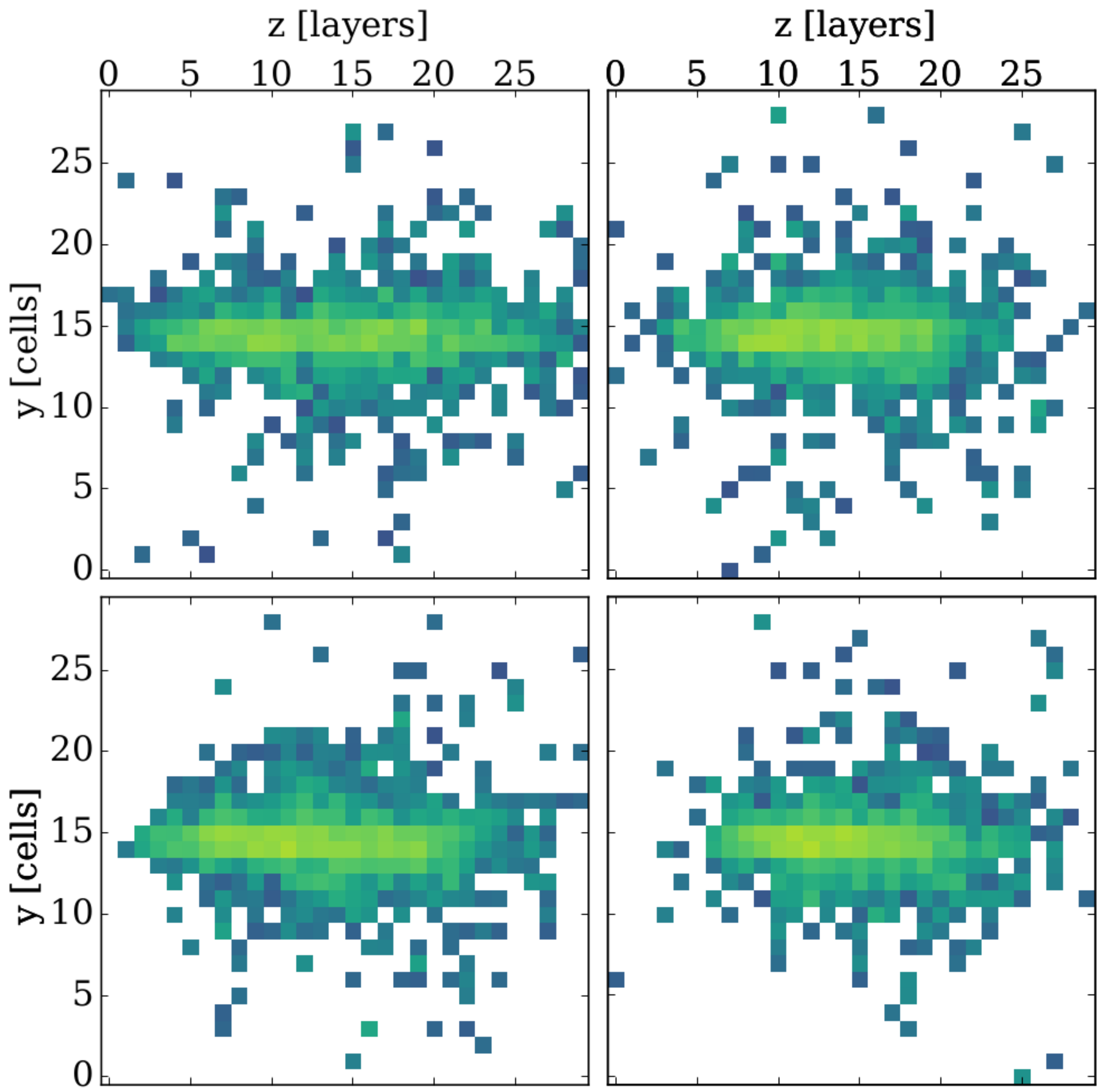
- Photons / charged Pions
- 1 million / 500k showers
- 10 to 100 GeV
- Fixed incident point & angle
- Project to grid
- 30×30×30 / 25×25×48

Buhmann et al.: **Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed** (2020) 2005.05334

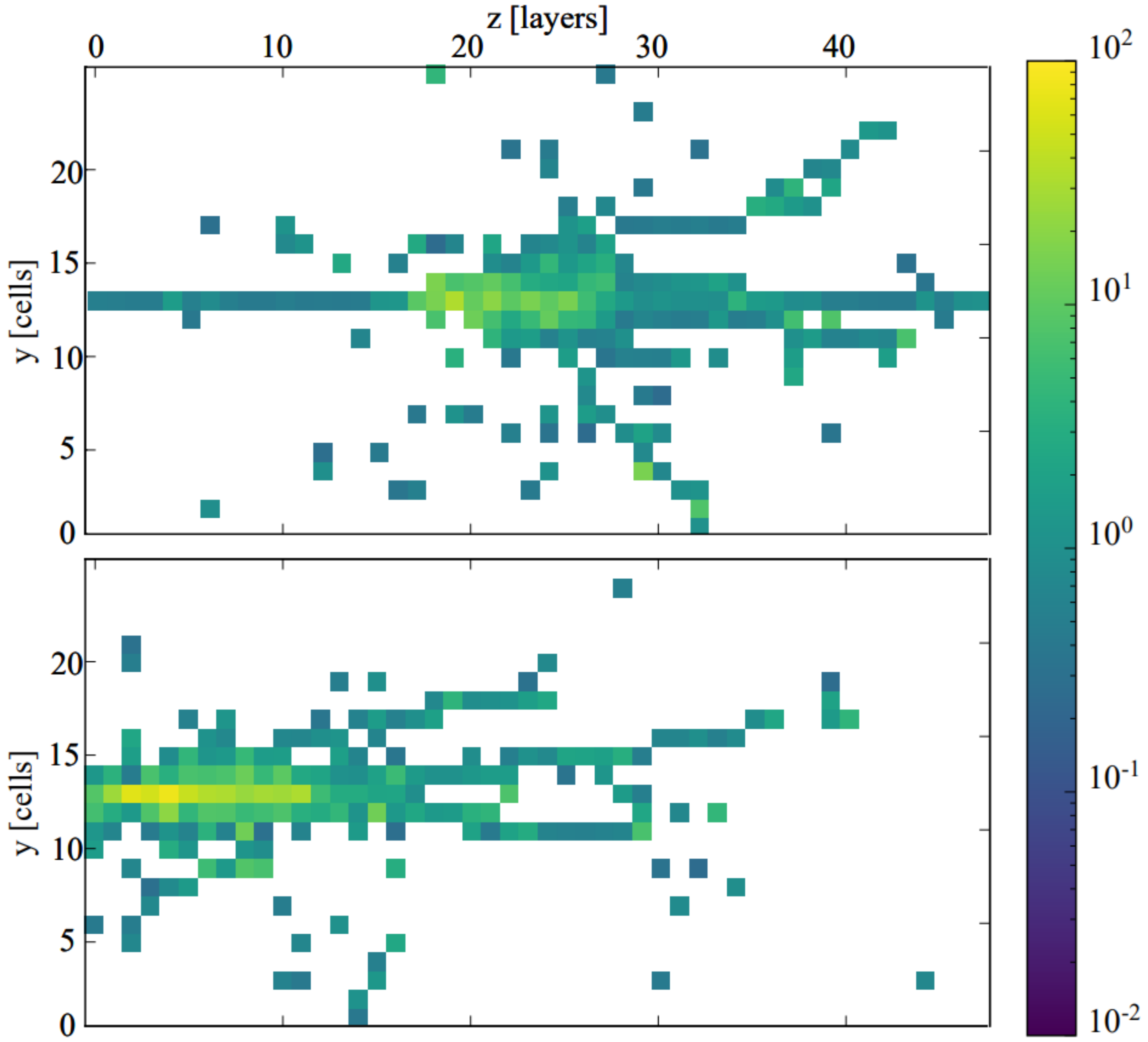
Buhmann et. al. **Hadrons, Better, Faster, Stronger:** (2021) 2112.09709

Shower Dataset

Photon showers



Charged pion showers



Pion showers significantly more complex

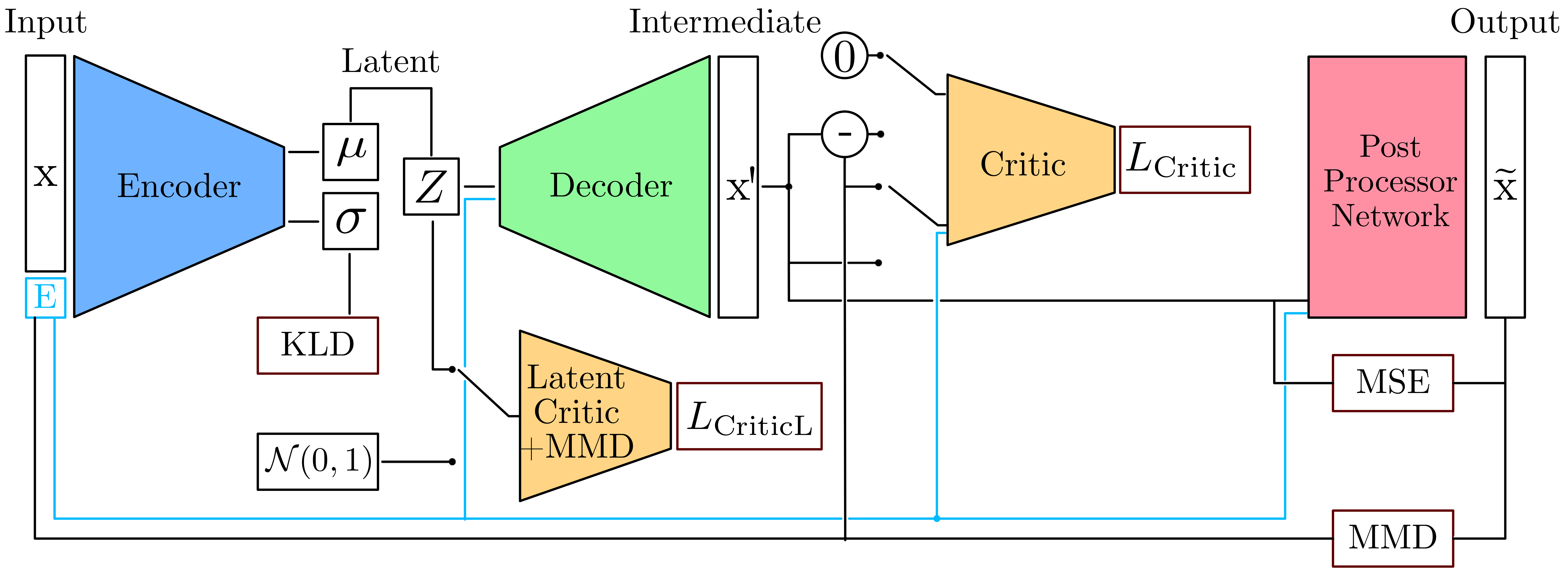
Training data:

- Photons / charged Pions
- 1 million / 500k showers
- 10 to 100 GeV
- Fixed incident point & angle
- Project to grid
- 30×30×30 / 25×25×48

Buhmann et al.: **Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed** (2020) 2005.05334

Buhmann et. al. **Hadrons, Better, Faster, Stronger:** (2021) 2112.09709

Bounded Information Bottleneck Autoencoder

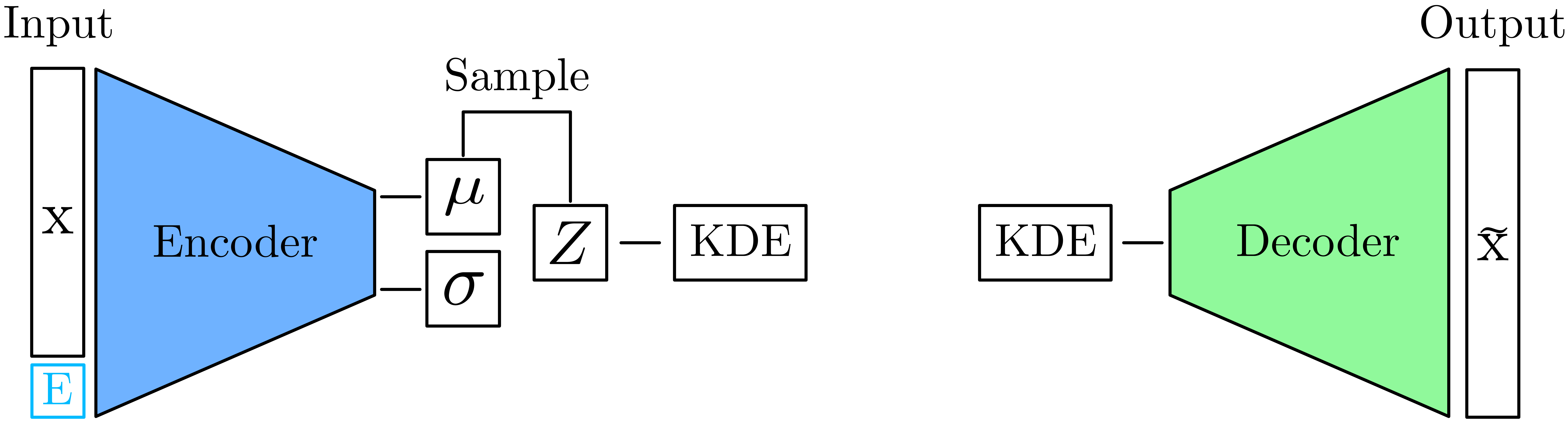


- Combines VAE and GAN approaches
- Final Post Processor network for fine tuning

Slava Voloshynovskiy et al.:
Information bottleneck through variational glasses: 1912.00830

Kernel Density Estimation BIB-AE

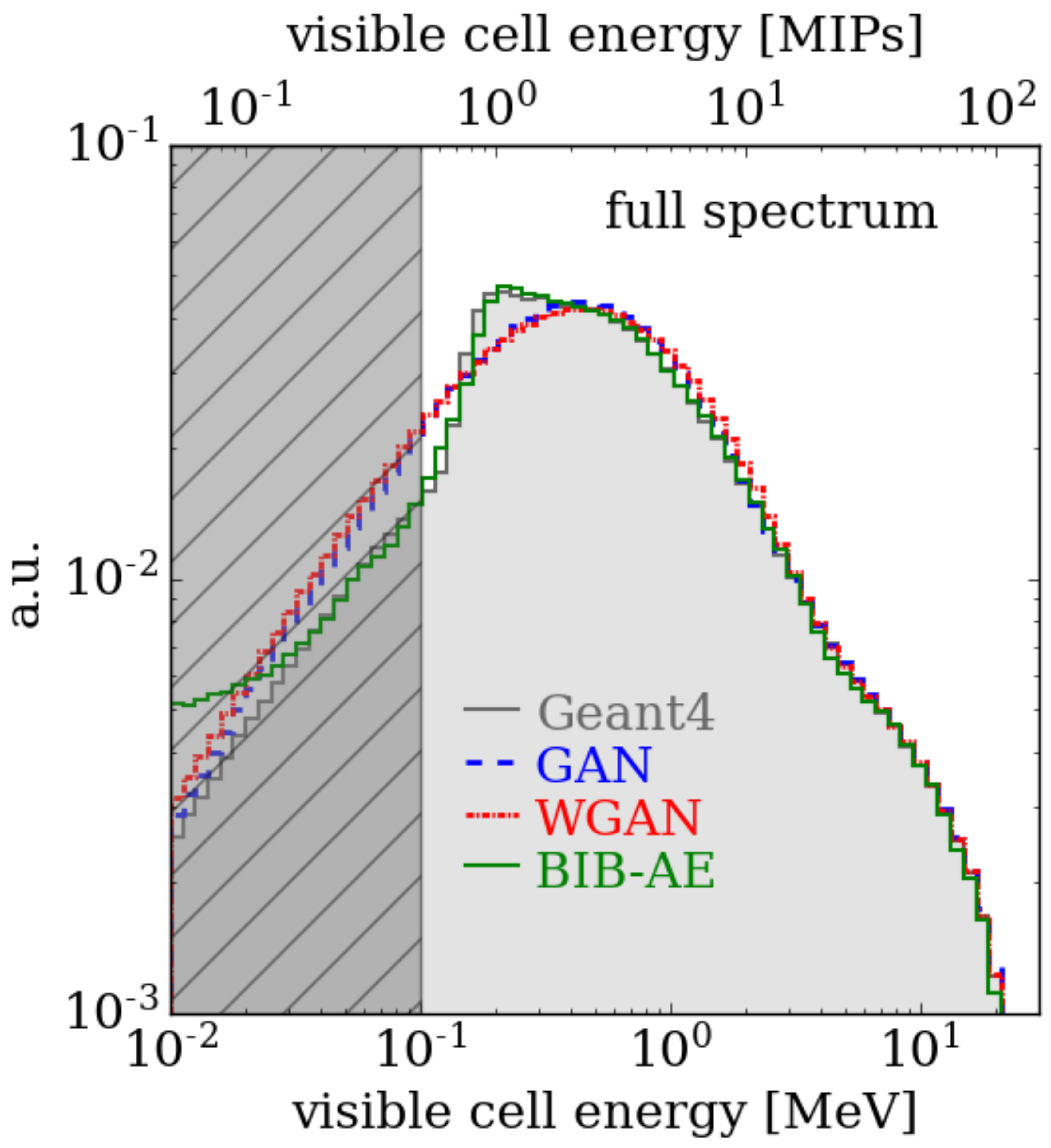
- Encode real shower images
- Fit KDE to latent space
- Sample latent space from KDE for generation



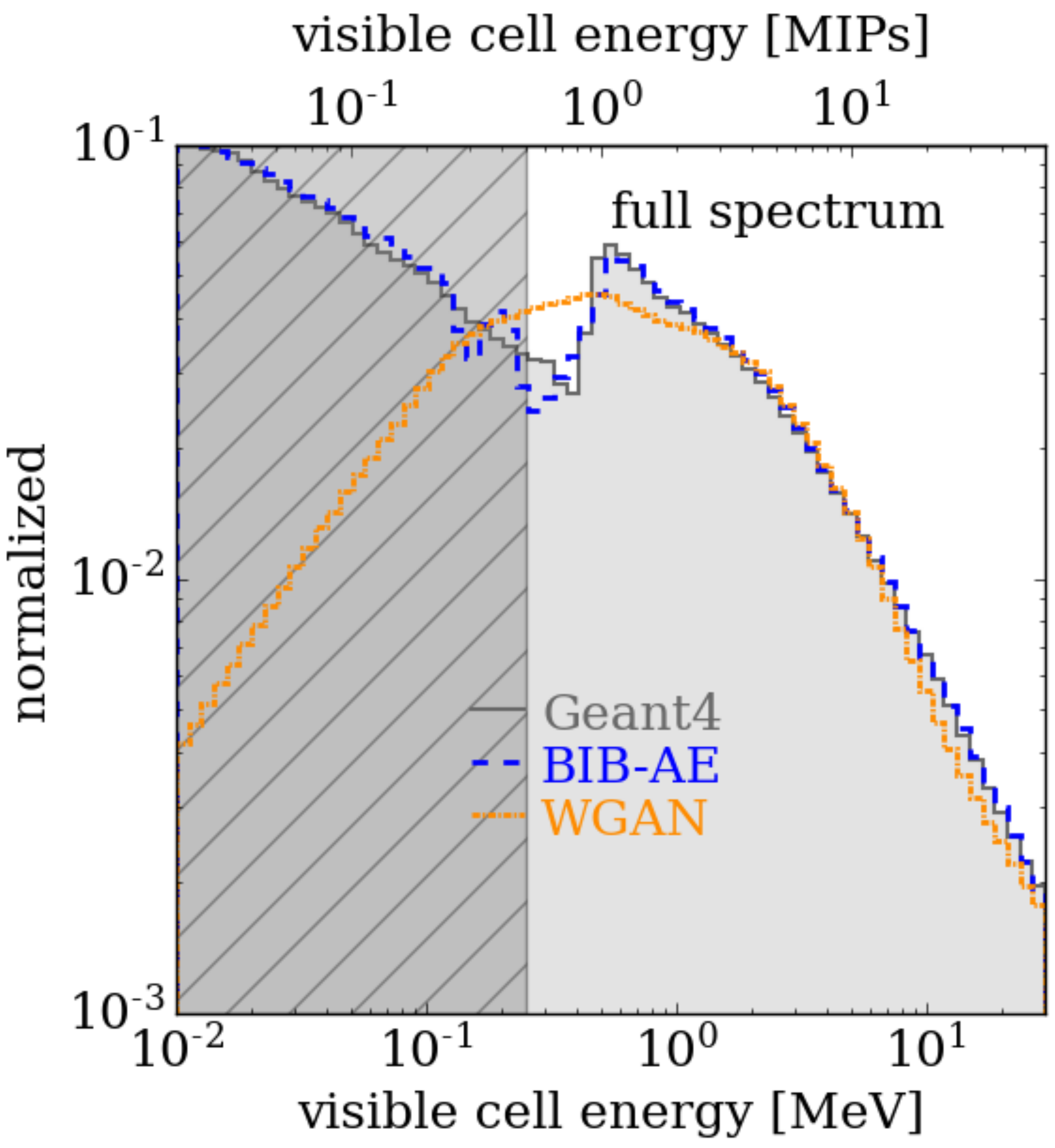
Buhmann et. al. **Decoding Photons: Physics in the Latent Space of a BIB-AE Generative Network:** (2021) [2102.12491](2102.12491)

Cell Energy Spectrum

Photons

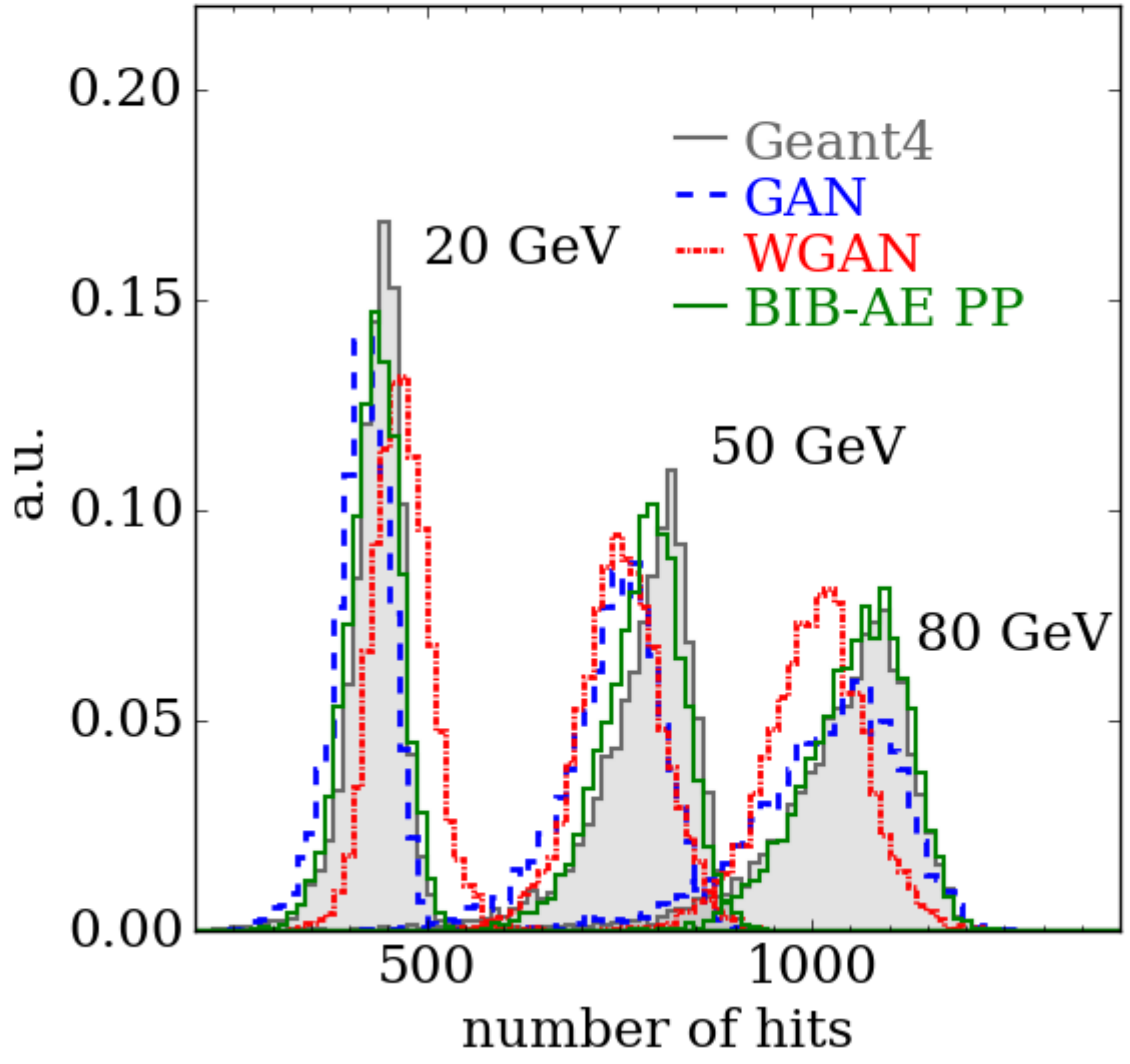


Pions

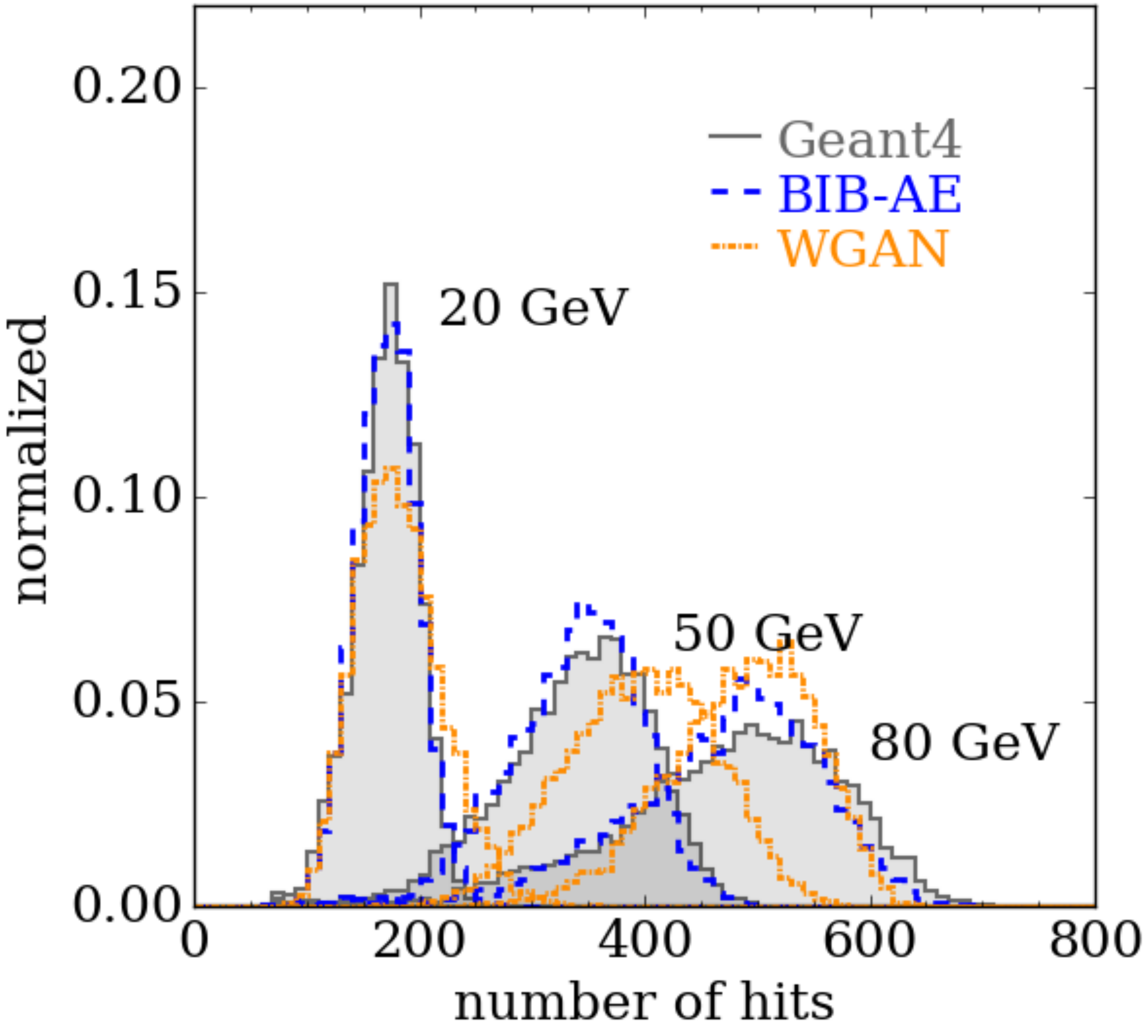


Number of Hits

Photons

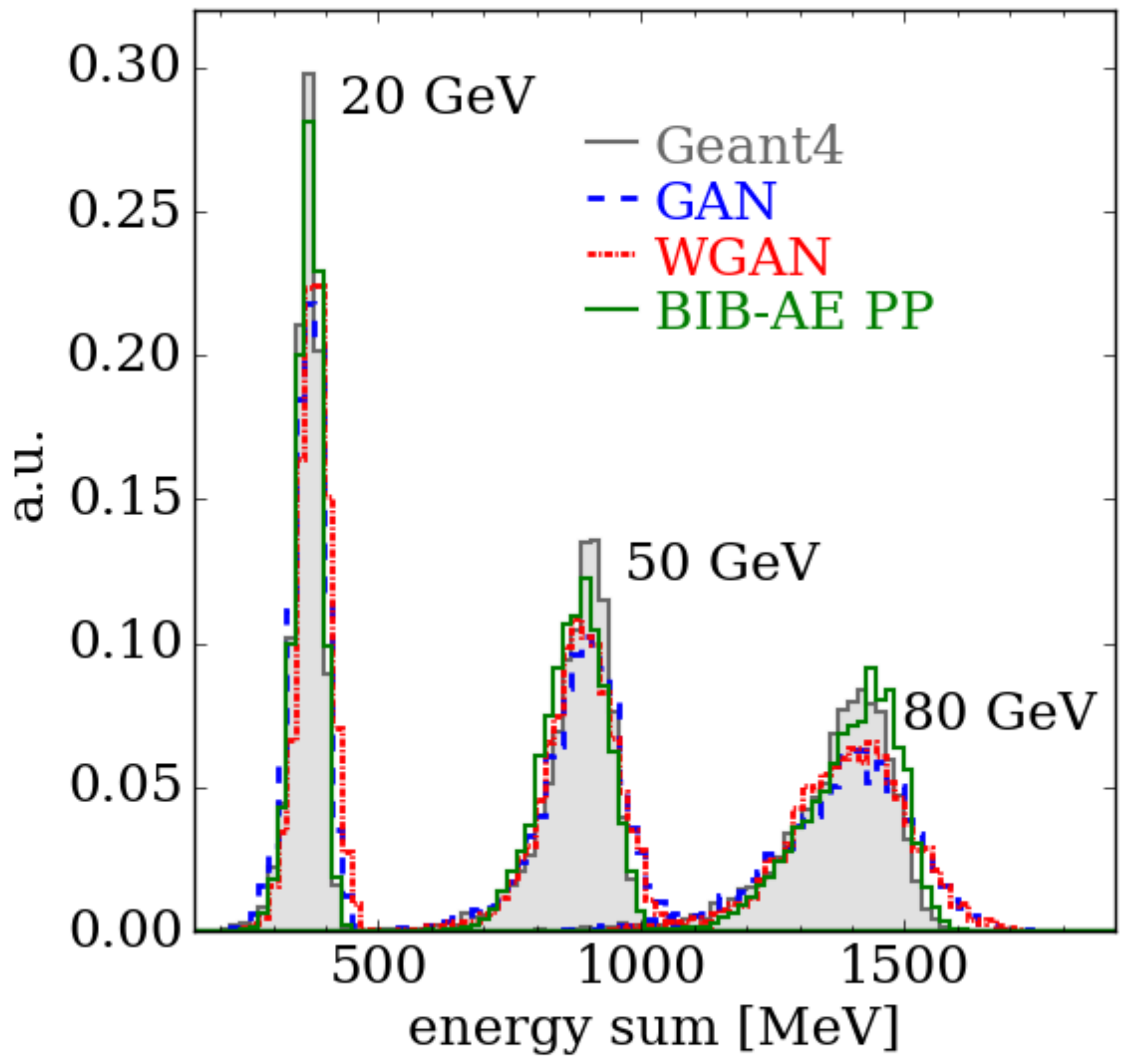


Pions

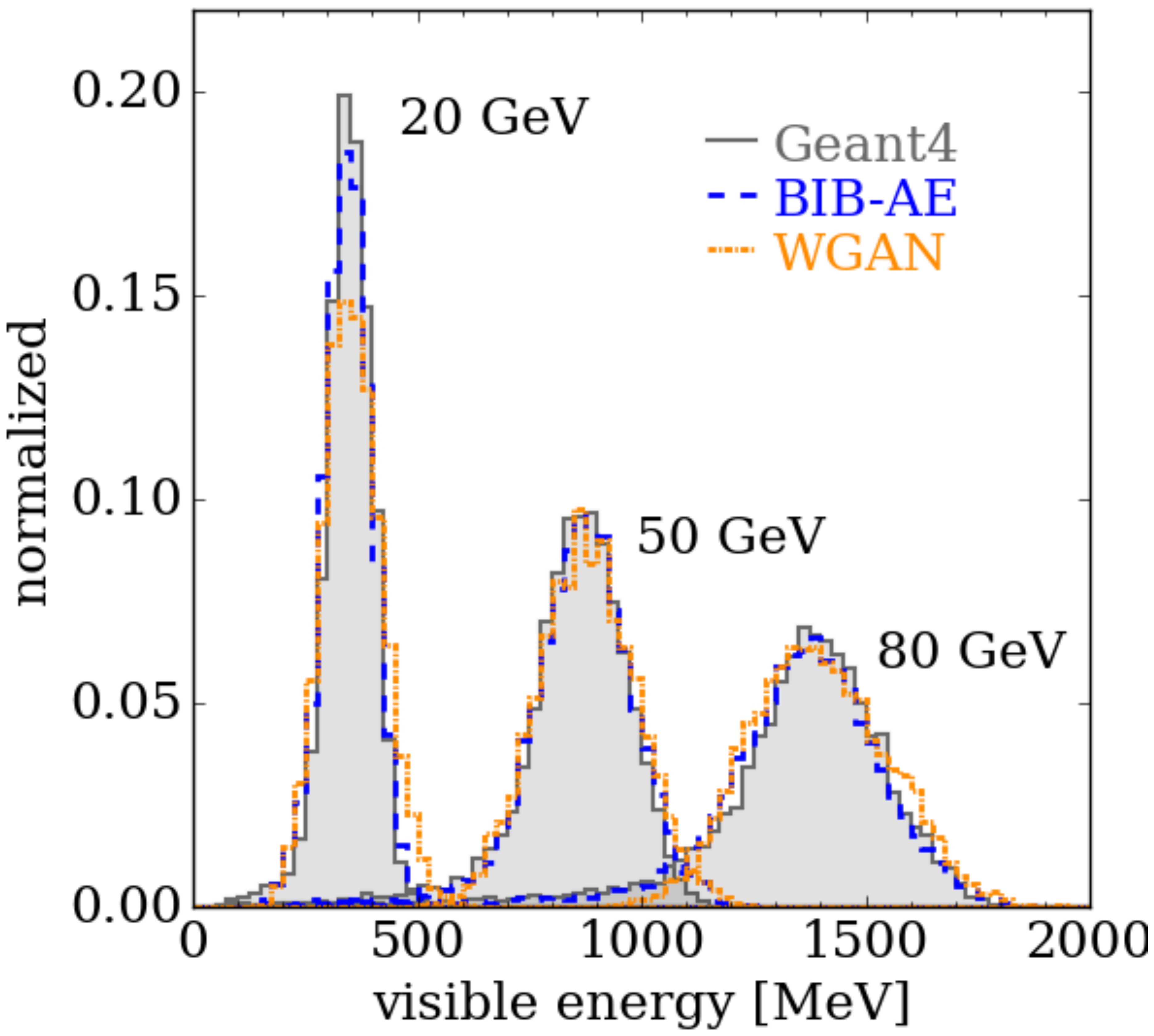


Visible Energy Sum

Photons

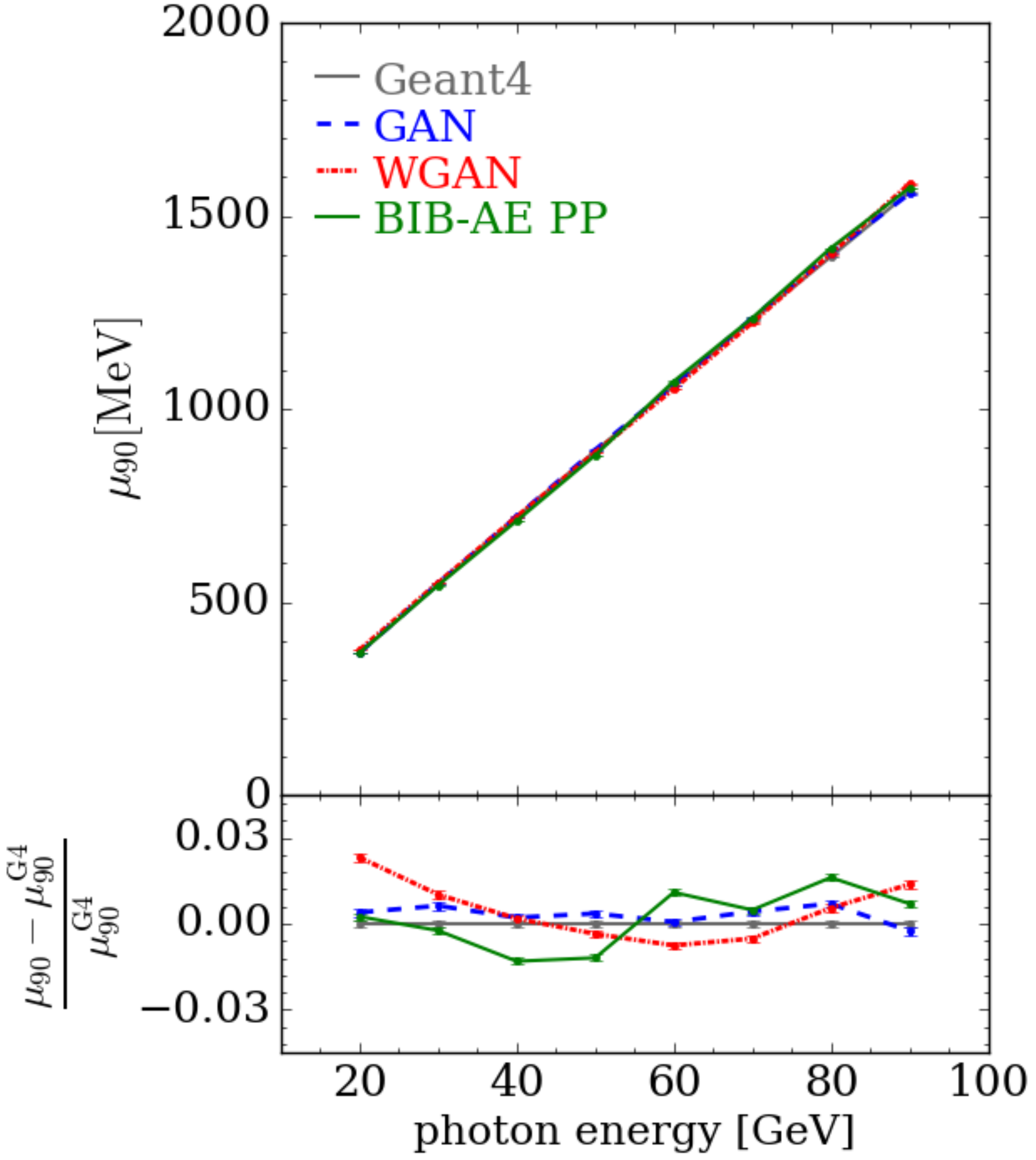


Pions

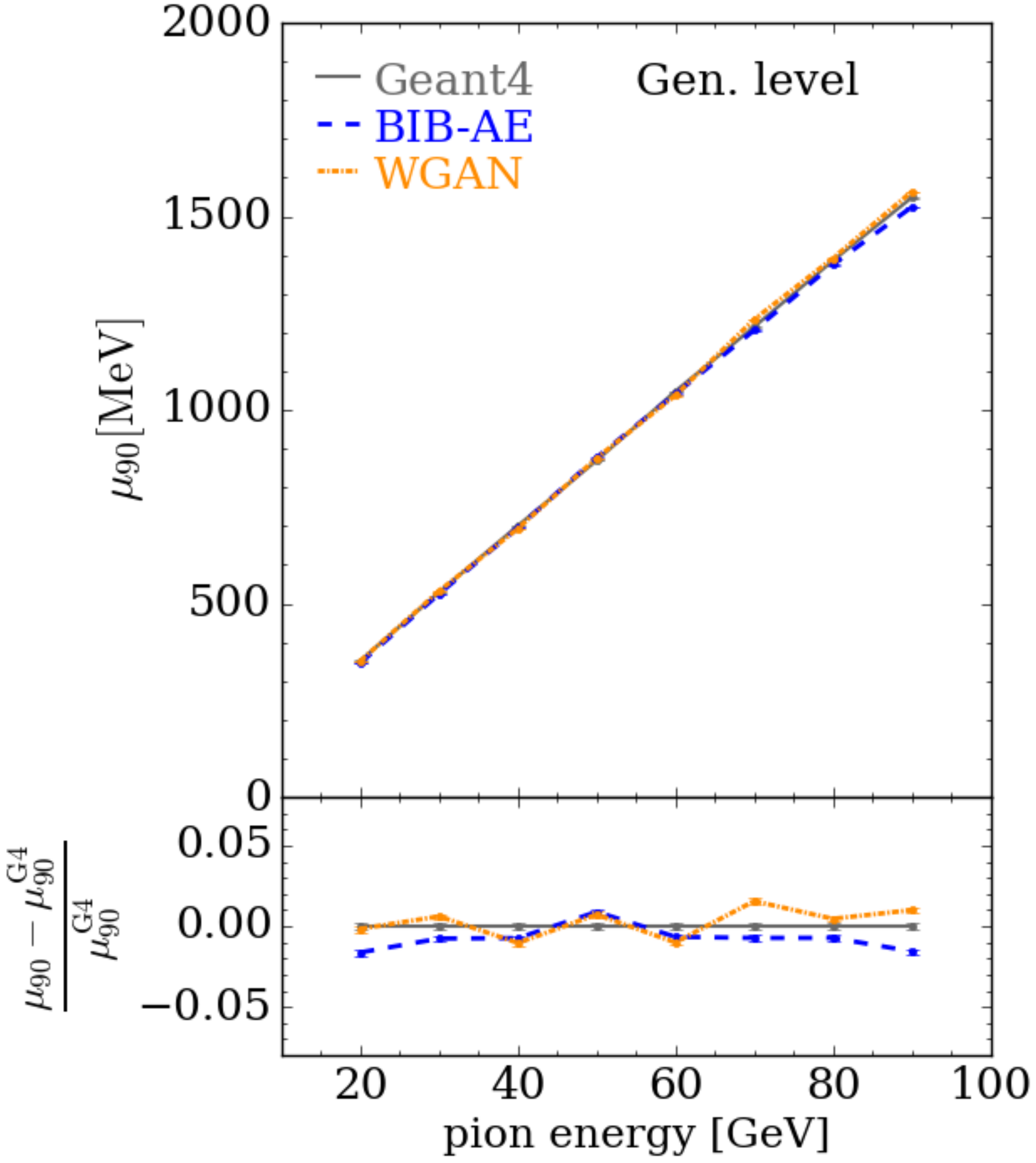


Visible Energy Sum (Means)

Photons

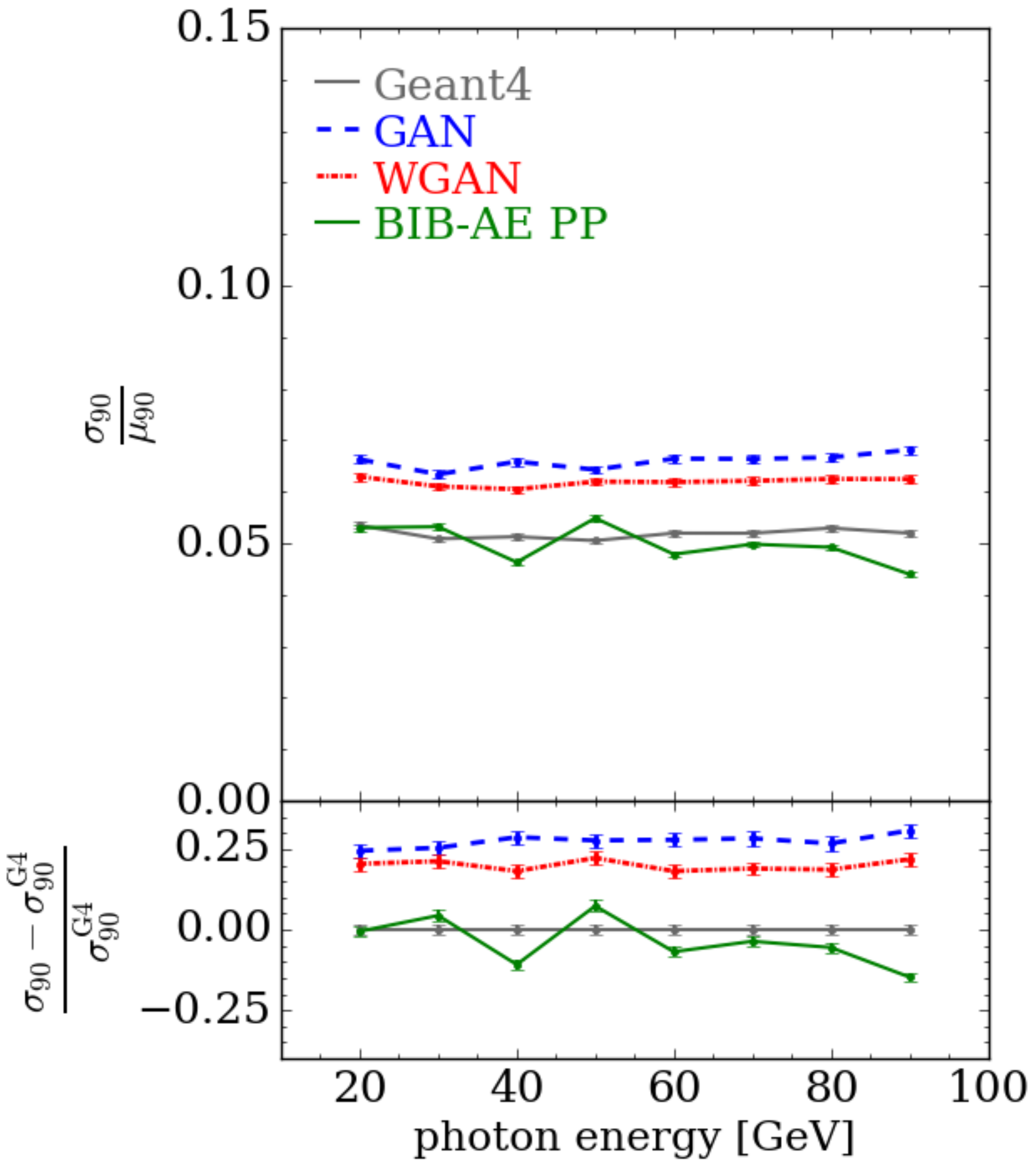


Pions

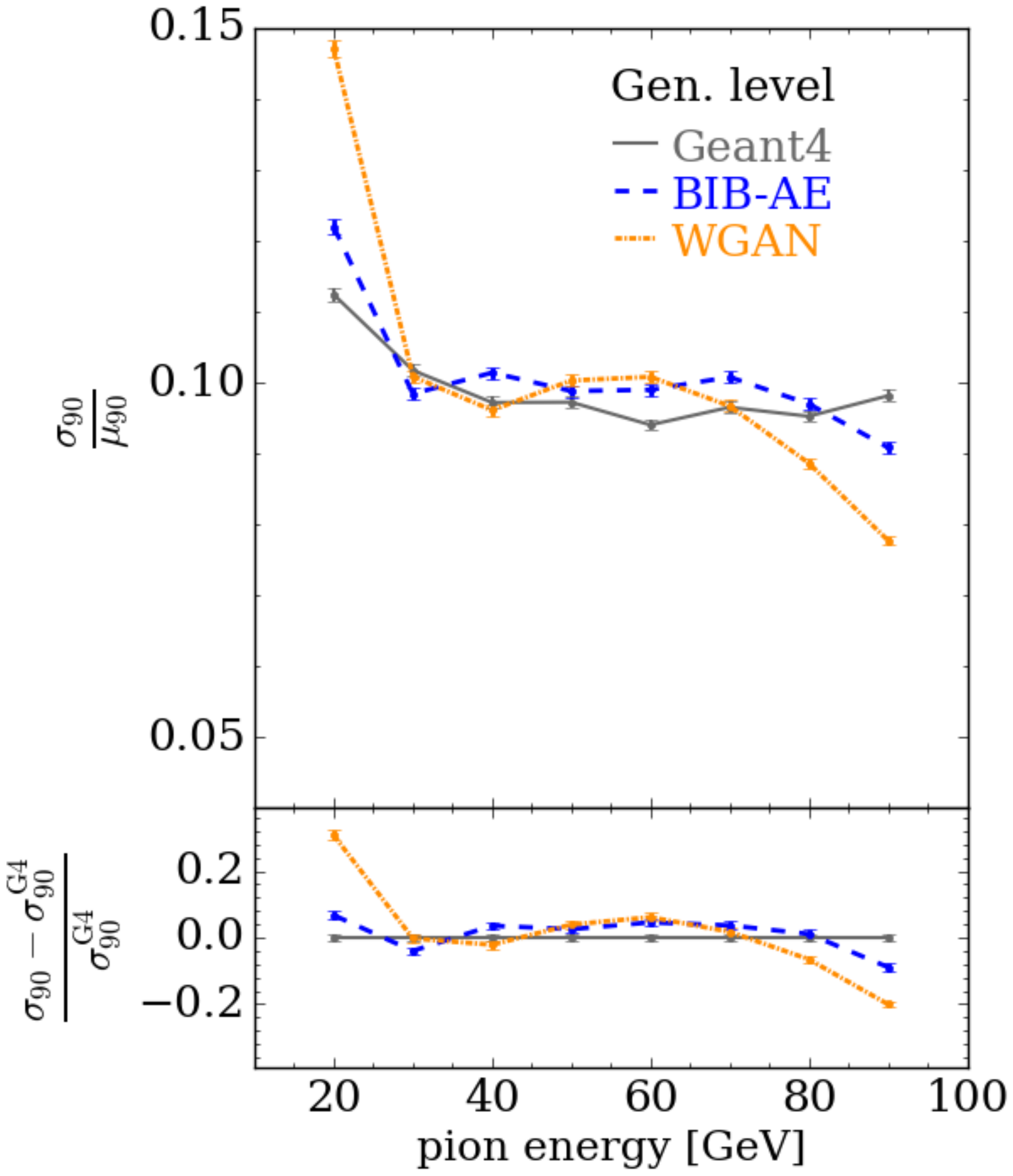


Visible Energy Sum (Widths)

Photons



Pions



Compute Times

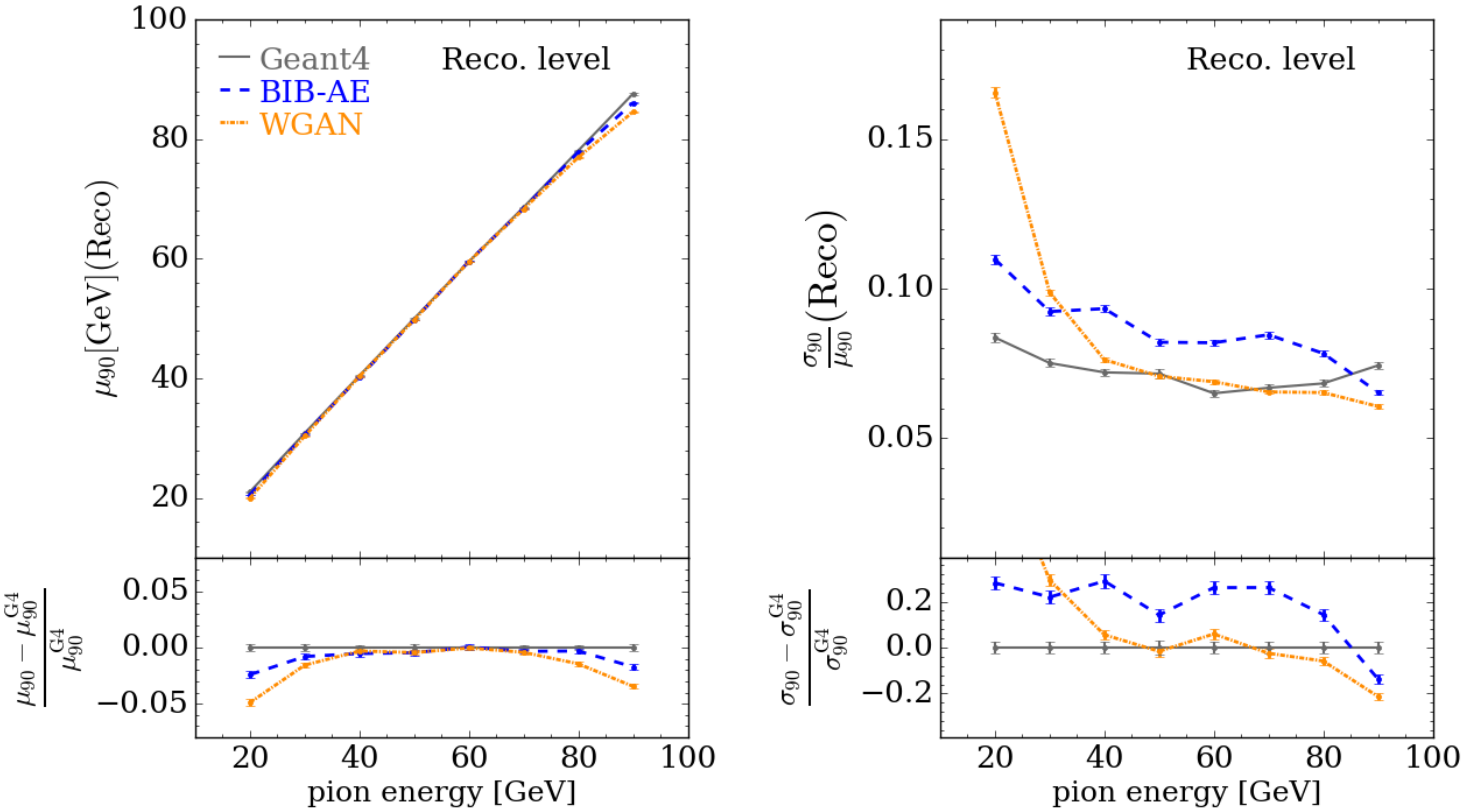
Simulator	Hardware	Photons	Speedup	Pions	Speedup
GEANT4	CPU	4082±170 ms	-	2684±125 ms	-
BIB-AE	CPU	426.3±3.6 ms	x10	350.824±0.574 ms	x8
WGAN	CPU	57.99±0.18 ms	x70	47.923±0.089 ms	×56
BIB-AE	GPU	1.42±0.01 ms	x2874	2.051±0.005	×1309
WGAN	GPU	3.34±0.01 ms	x1256	0.264±0.002	x10167

Times for fastest evaluation batch sizes, GPU times have to be normalised by cost

Buhmann et al.: **Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed** (2020) 2005.05334

Buhmann et. al. **Hadrons, Better, Faster, Stronger:** (2021) 2112.09709

Pion Reconstruction Results



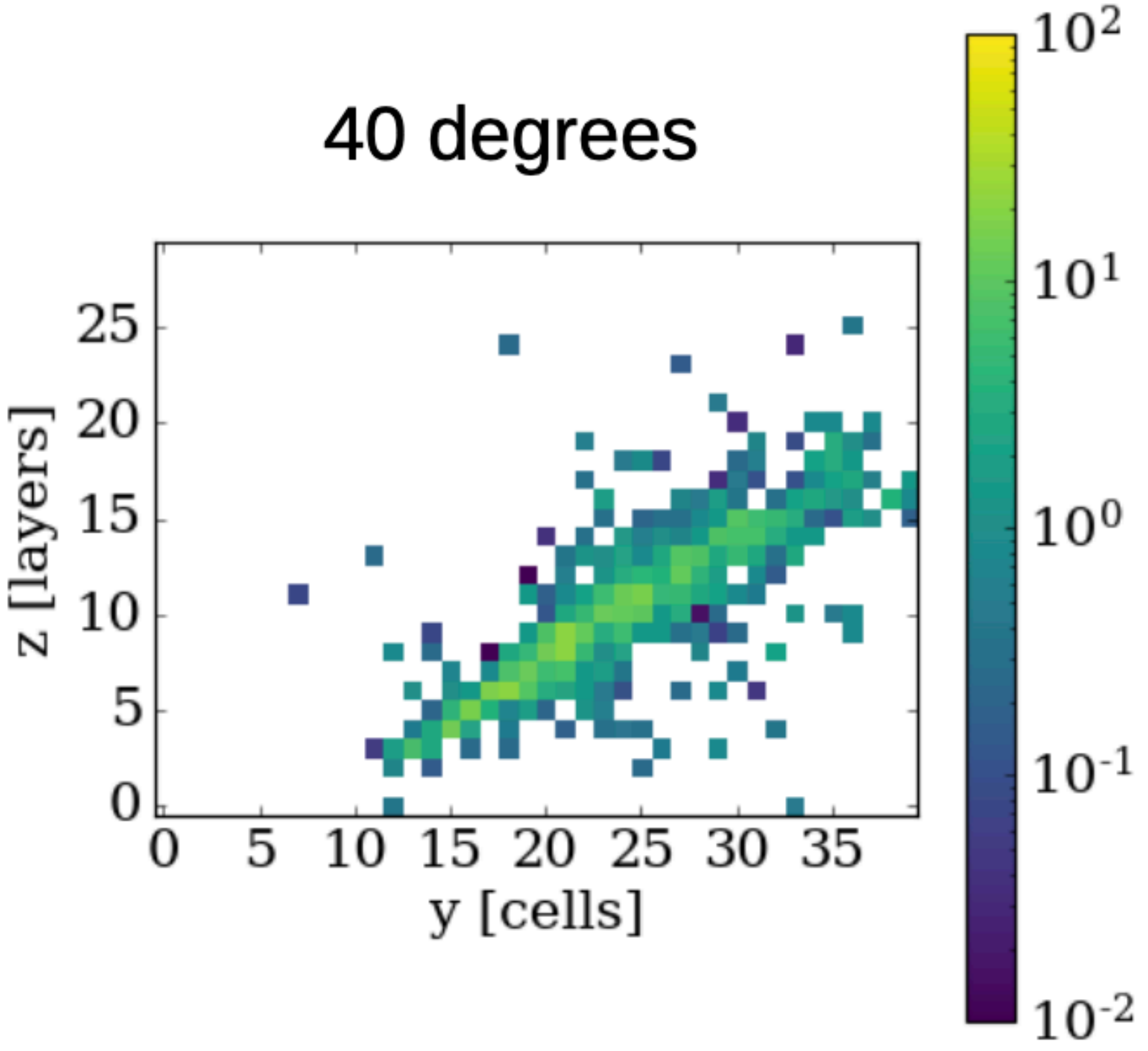
Current Status

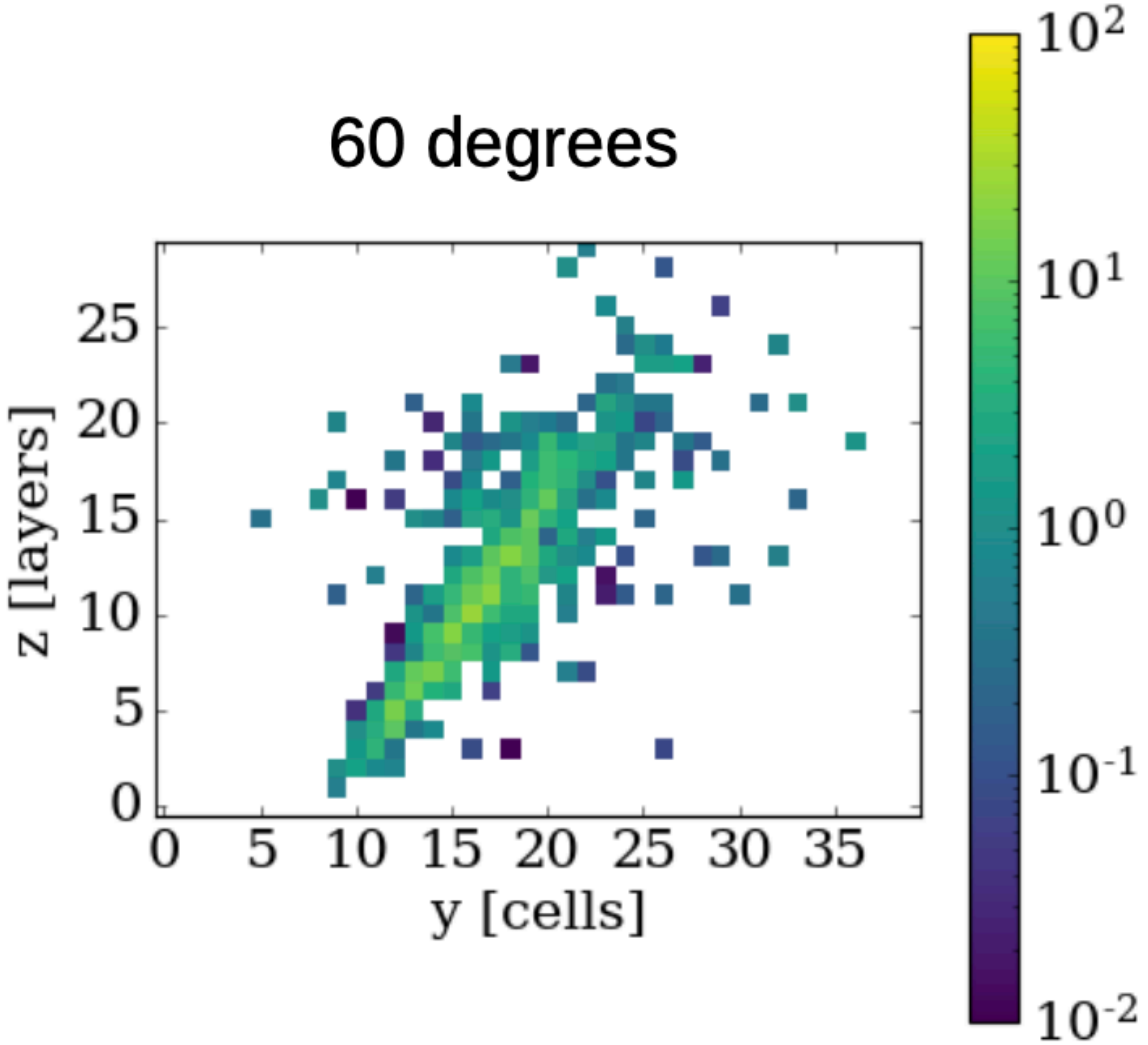
Shower Type	Generative Simulation	Energy Conditioning	Angle Conditioning	Reconstruction
EM	✓	✓	+	✗
Hadronic	✓	✓	✗	✓

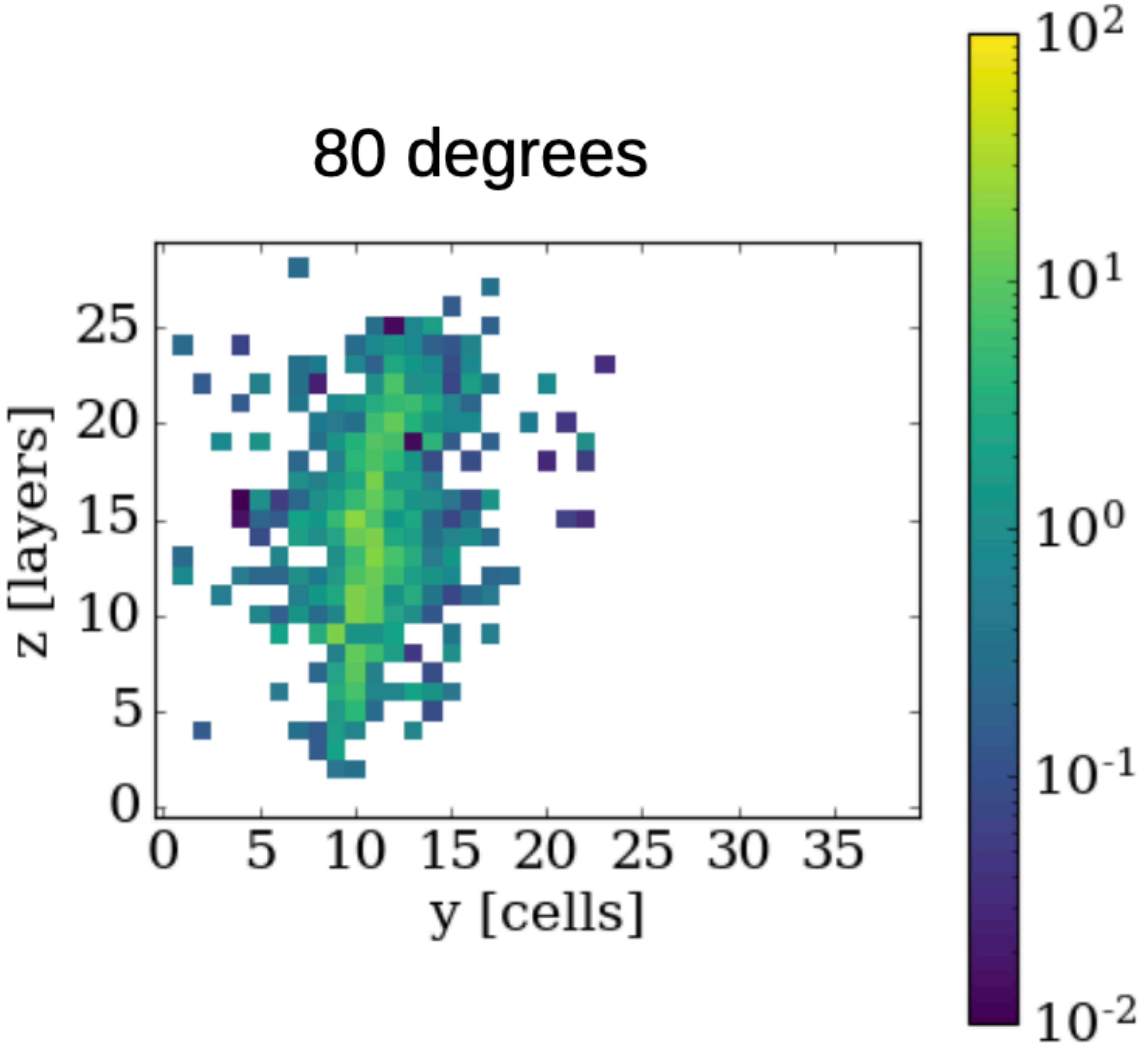
✓ **: Done** + **: In progress** ✗ **: To be done**

Angle Conditioning

- Need variable angles for practical application
- Conditioning on both photon shower angle and energy







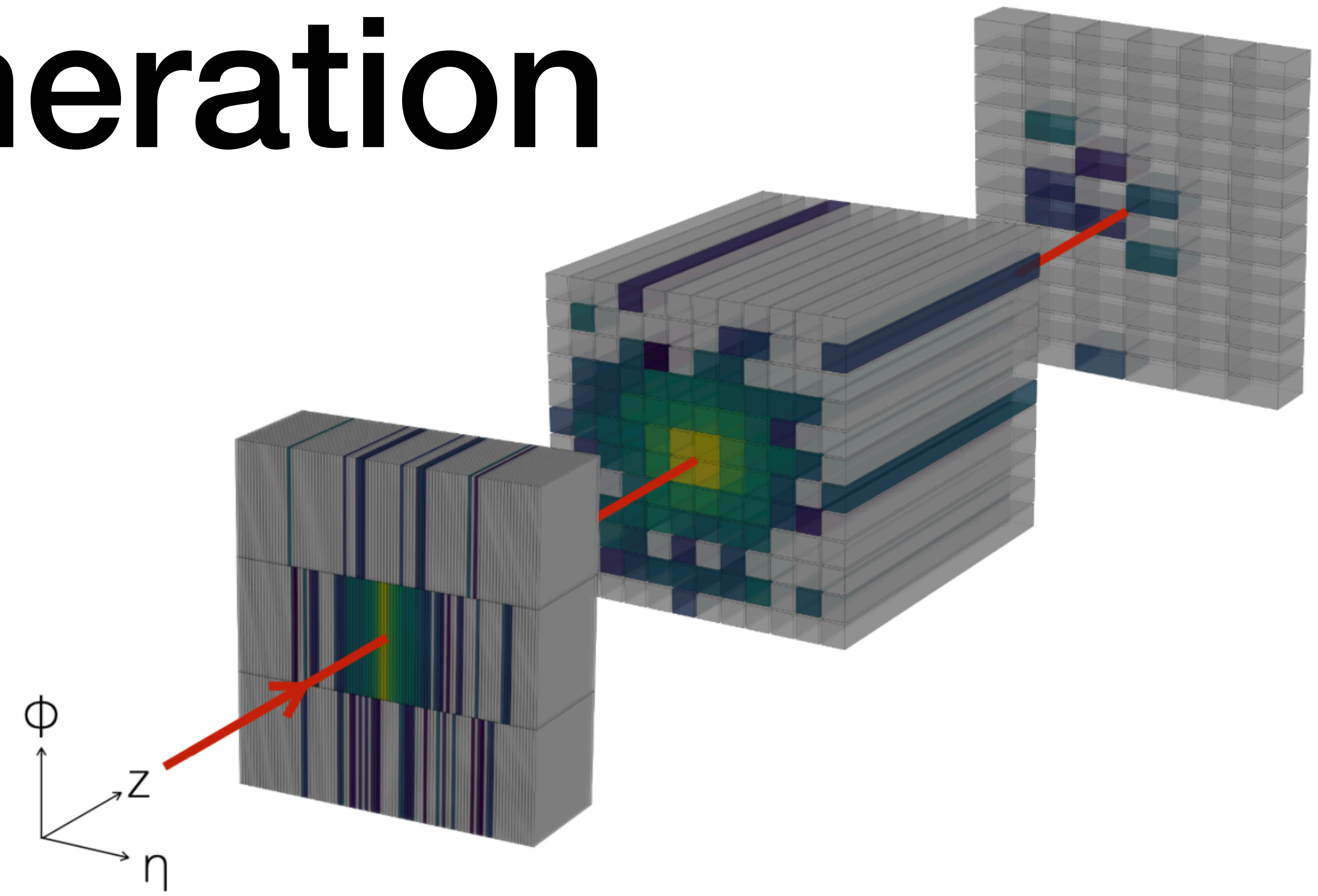
Angle Conditioning

- Additional angle input to BIB-AE and Post Processor network
- Replace latent KDE with conditional latent flow model

Angle Conditioning

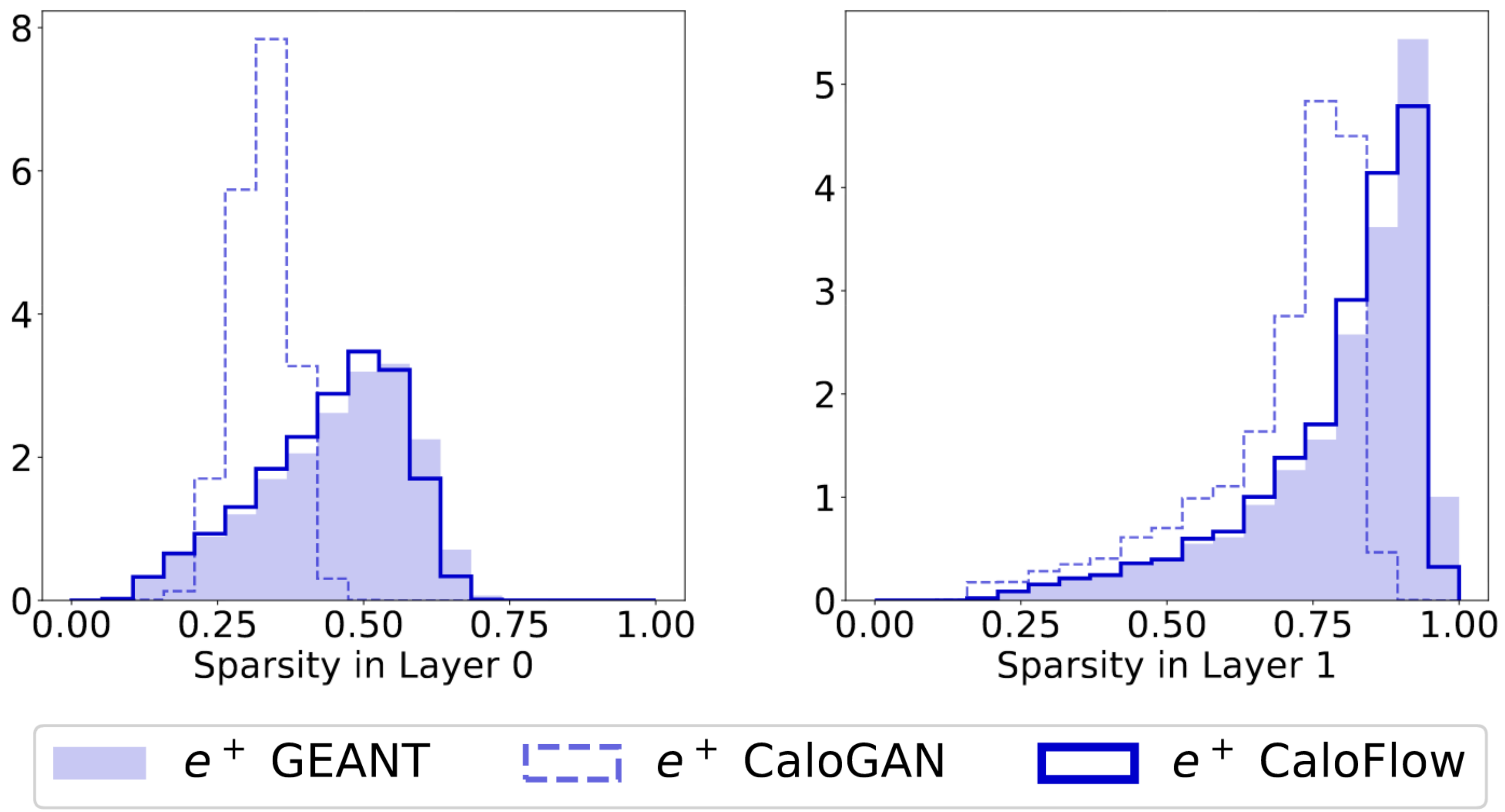
- Use PCA to reconstruct particle angle from shower
- BIB-AE accurately reproduces variety of angles and energies

Flow Based Generation



- CaloFlow:
  - Demonstration of flows on CaloGAN data set
  - 3 layers, ~500 dimensions total
  - Significant performance increase going from GAN to flow

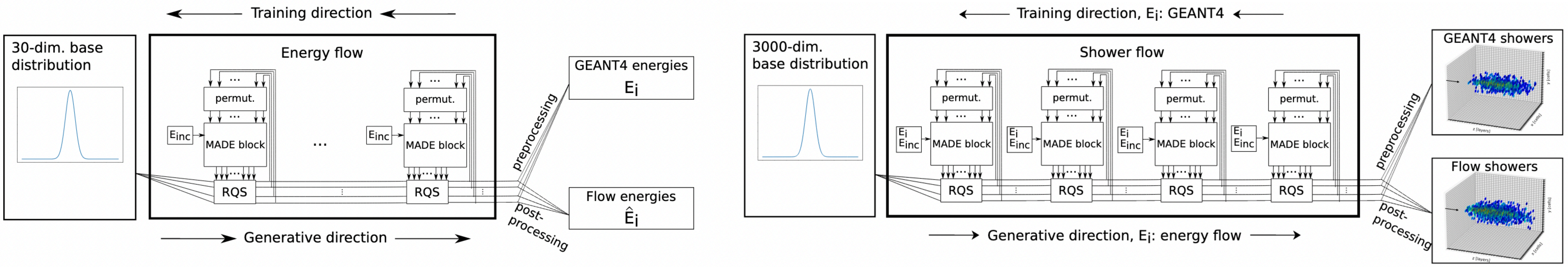
➡ Test on ILD ECAL data set



Krause et. al. **CaloFlow: Fast and Accurate Generation of Calorimeter Showers with Normalizing Flows:** (2021) [2106.05285](2106.05285)

Flow-Based Generation

- Two step generation approach:
  - "Layer flow" learns deposited energies per layer
  - "Shower flow" learns energy depositions per ECal cell



- Shower flow conditioned on layer flow output
- Additionally: Generated showers re-scaled using layer flow

Number of Parameters

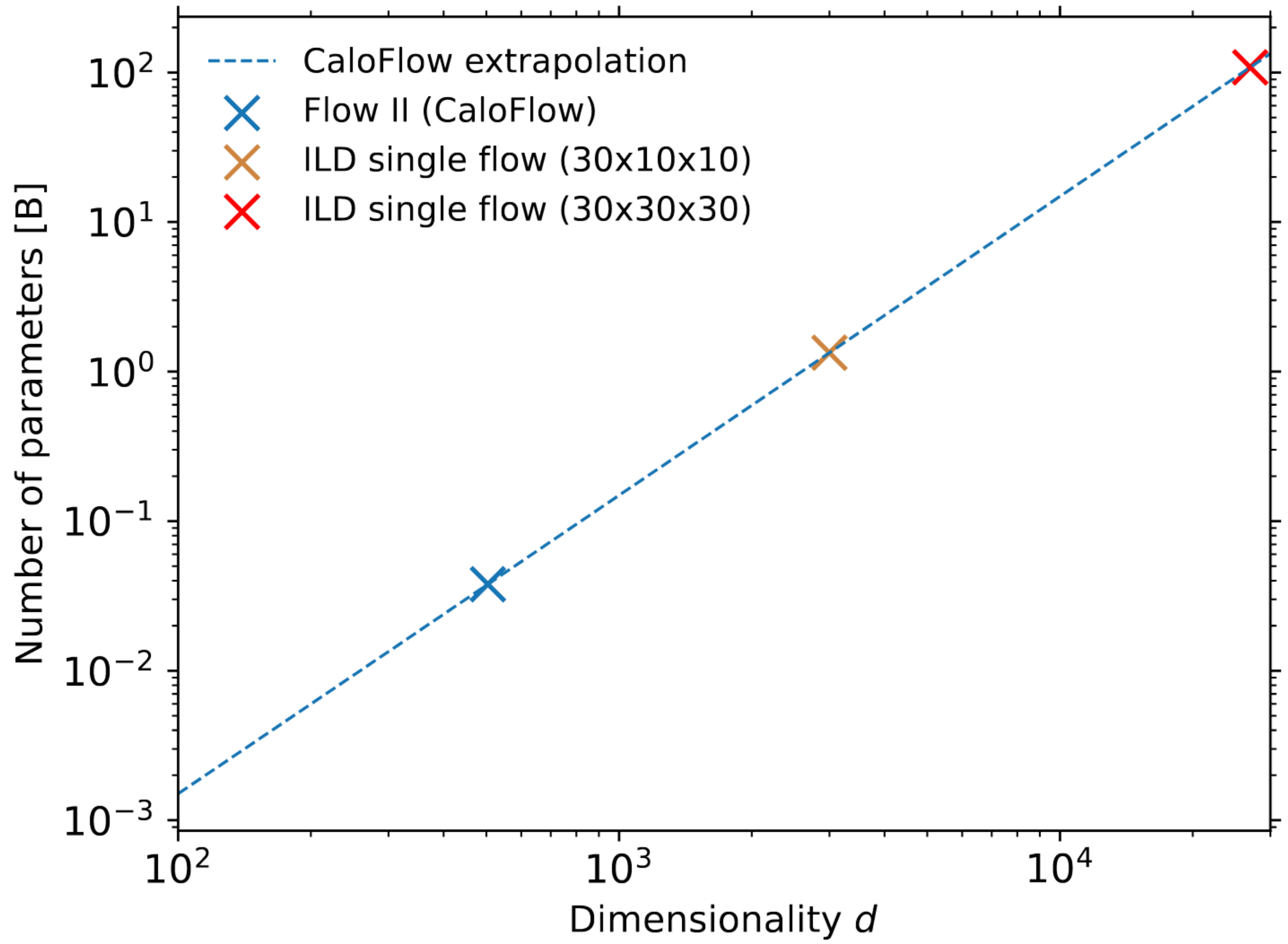
- Shower flow on 30x30x30 too large

➡ Focus in central 10x10x30 region

- O(1,000,000,000) parameters
- Still too large for single flow

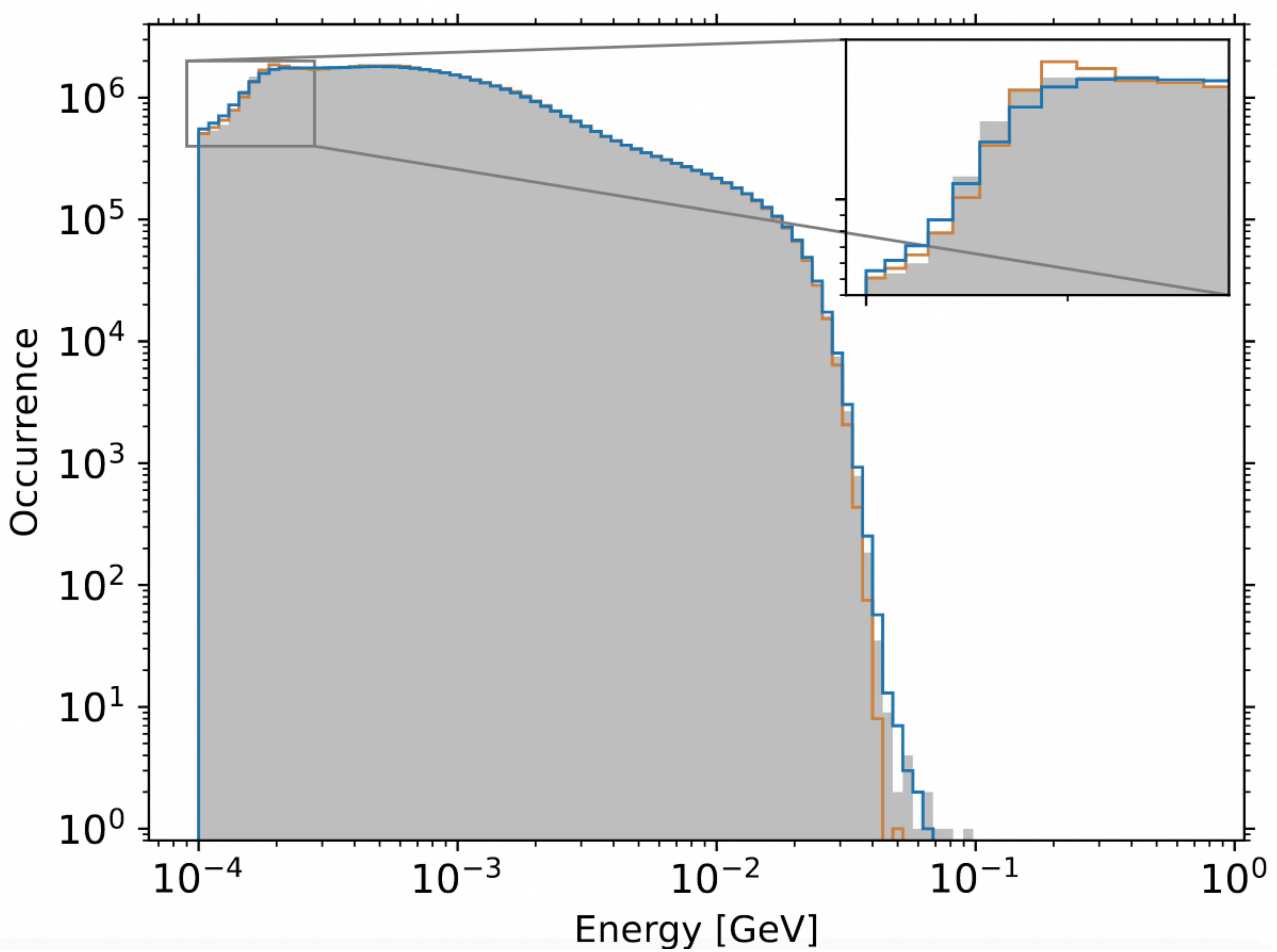
➡ Multiple flows:

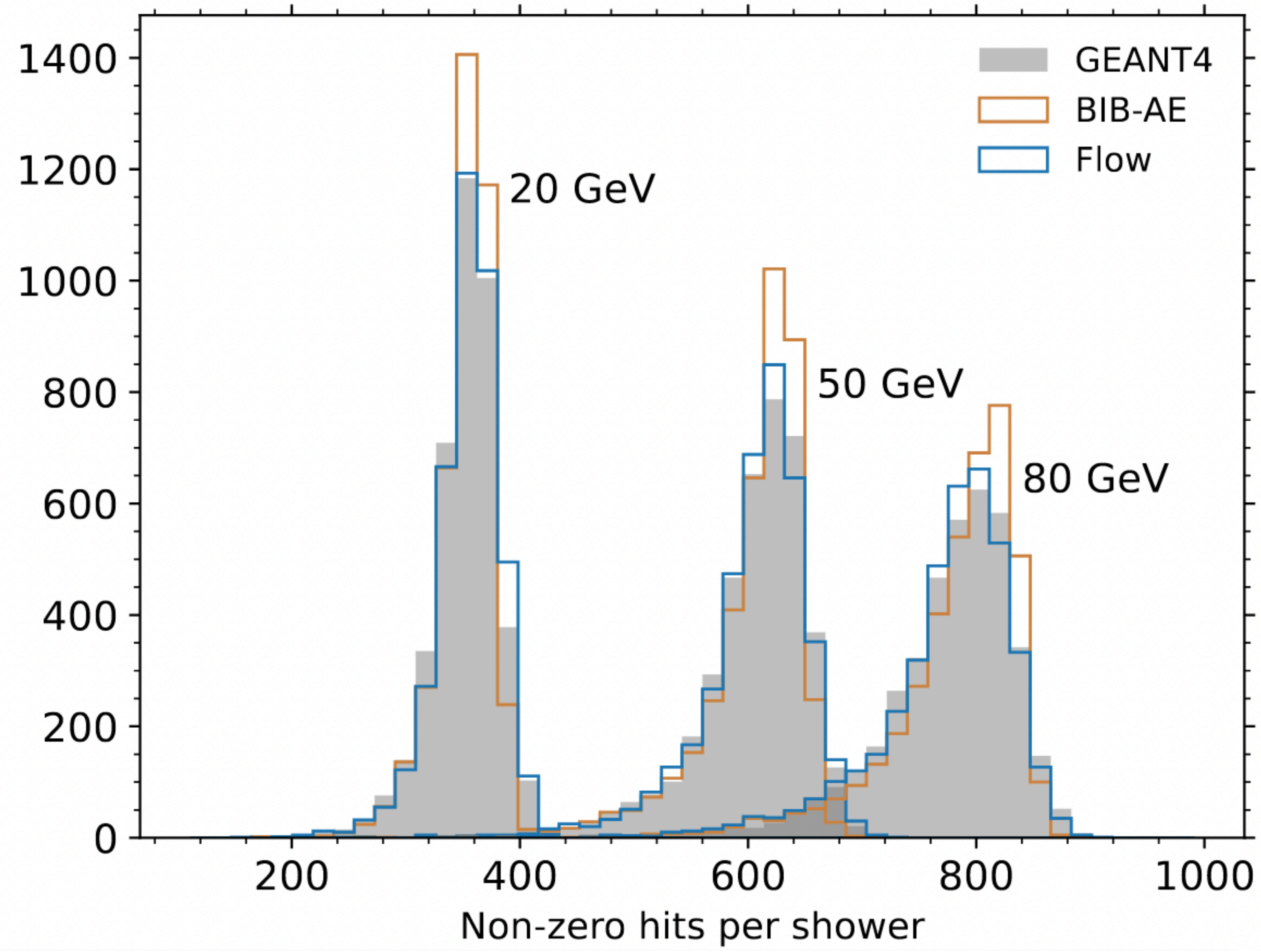
  - Each flow conditioned on previous 5 layers
  - Reduced overall model size



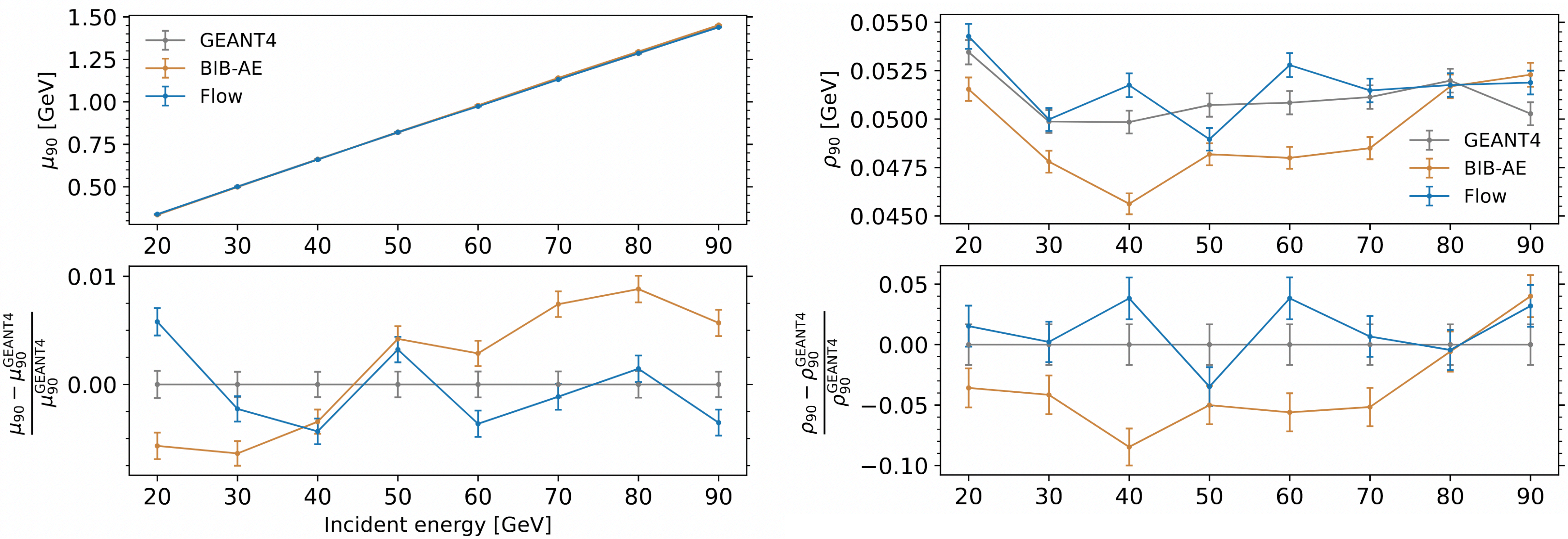
Flow i	Context features	Context shape
0	E_0, E_{inc}	$[N, 2]$
1	I_0, E_1, E_{inc}	$[N, 102]$
2	I_1, I_0, E_2, E_{inc}	$[N, 202]$
3	I_2, I_1, I_0, E_3, E_{inc}	$[N, 302]$
4	I_3, I_2, I_1, I_0, E_4, E_{inc}	$[N, 402]$
≥ 5	I_{i-1}, I_{i-2}, I_{i-3}, I_{i-4}, I_{i-5}, E_i, E_{inc}	$[N, 502]$

Cell Spectrum and Number of Hits





Visible Energy Sum (Mean and Width)



Classifier Performance

/	BIB-AE	Multiple flows
Classifier AUC	0.9912 ± 0.0018	0.8399 ± 0.0036

- Classifier trained to distinguish GEANT4 from generated showers
- 5 trainings used to estimate uncertainty
- Both methods imperfect
- BIB-AE may have tells/artefacts resulting in perfect classification
- Flow performs better

Compute Times

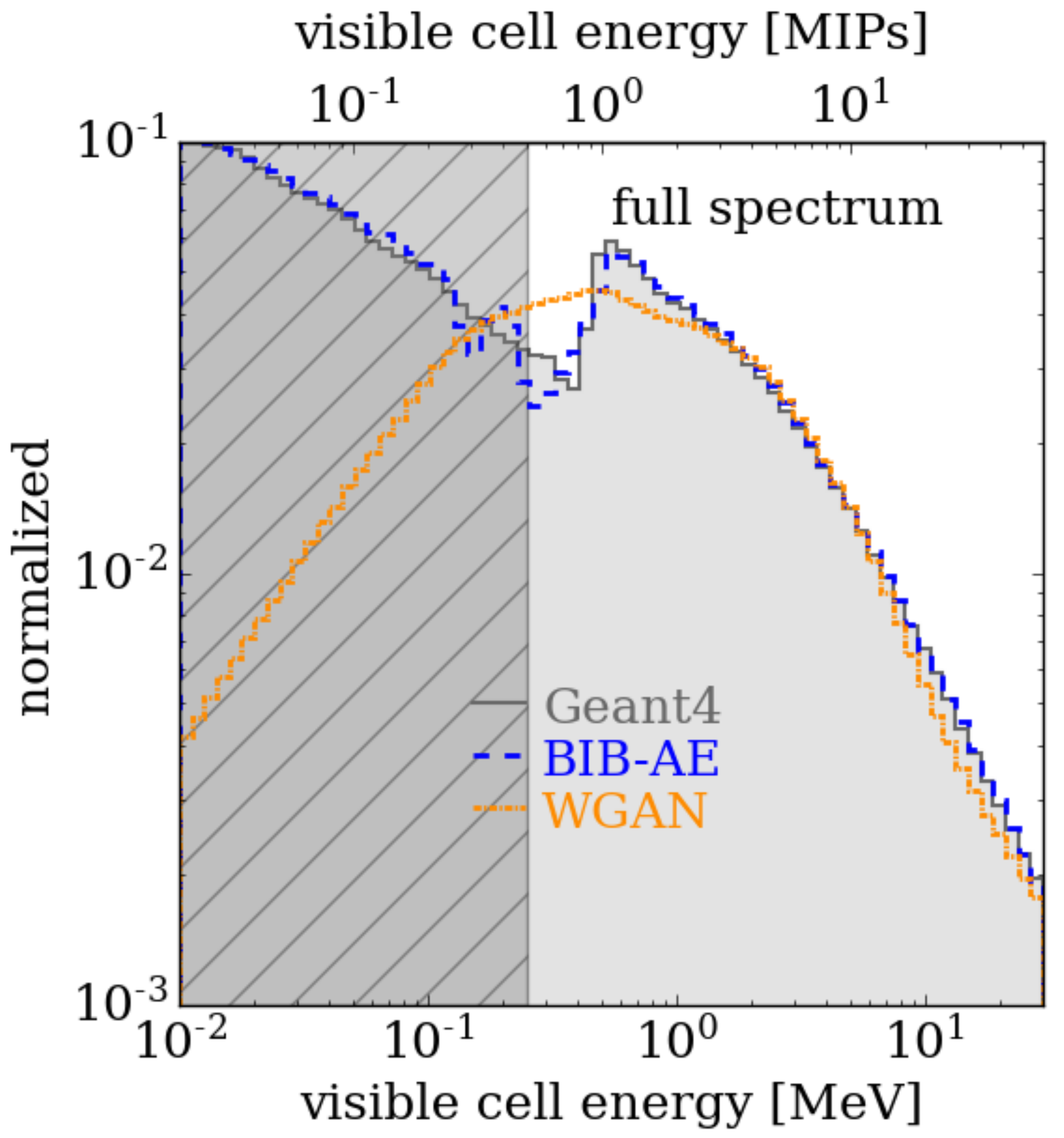
Simulator	Hardware	Photons 10x10	Speedup
GEANT4	CPU	4082±170 ms	-
BIB-AE	CPU	37.81±0.13 ms	x108
Flow	CPU	2066±31 ms	x2
BIB-AE	GPU	0.23±0.01 ms	x17745
Flow	GPU	12.51±0.01 ms	x326

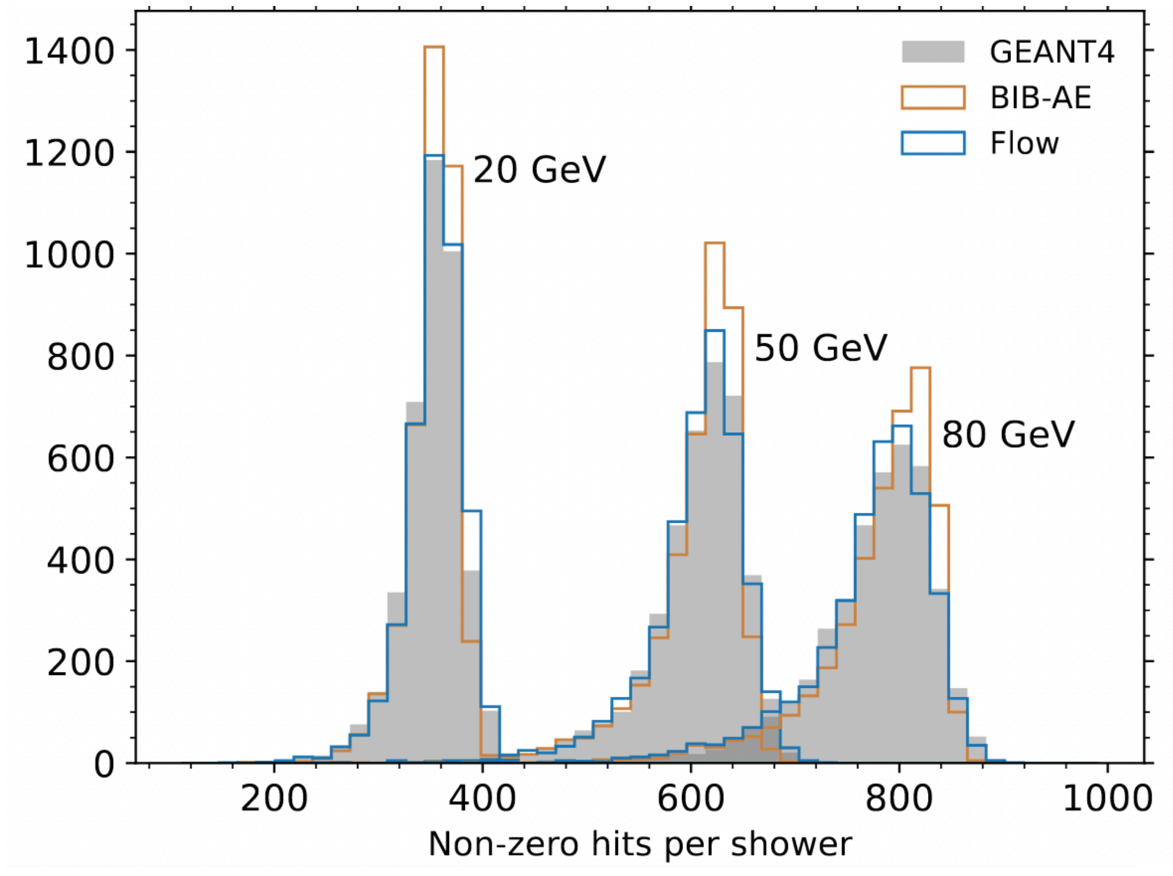
Times for fastest evaluation batch sizes, GPU times have to be normalised by cost

- MAF setup used in flow slow for generation
- Teacher-student IAF setup can lead to speedup, work in progress

Conclusion

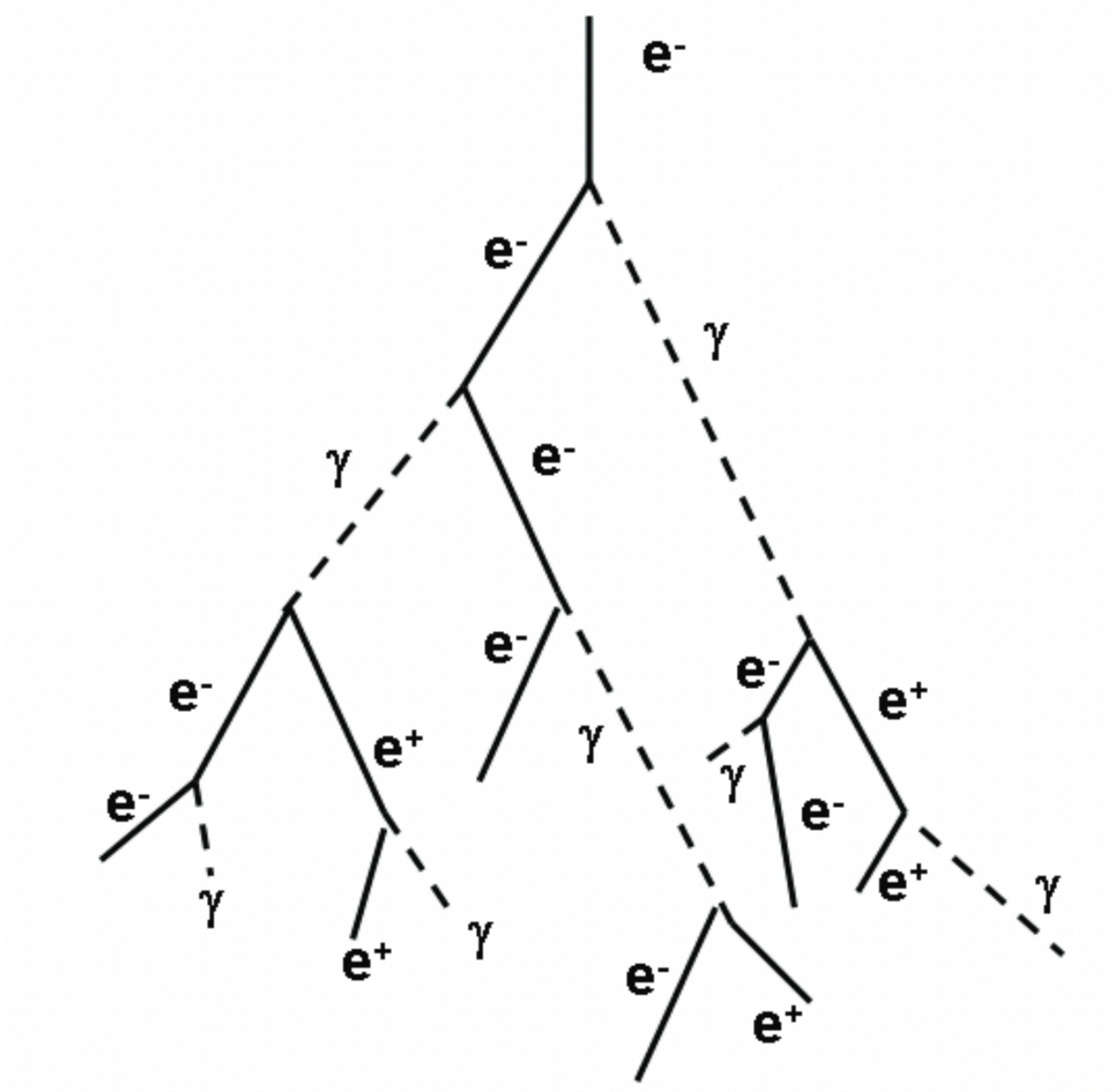
- Generative models growing field in particle physics
- Calorimeter simulation:
  - BIB-AE model accurately produces HG showers
  - Angle conditioning work in progress
  - Flow based models provide increased accuracy on smaller scale data

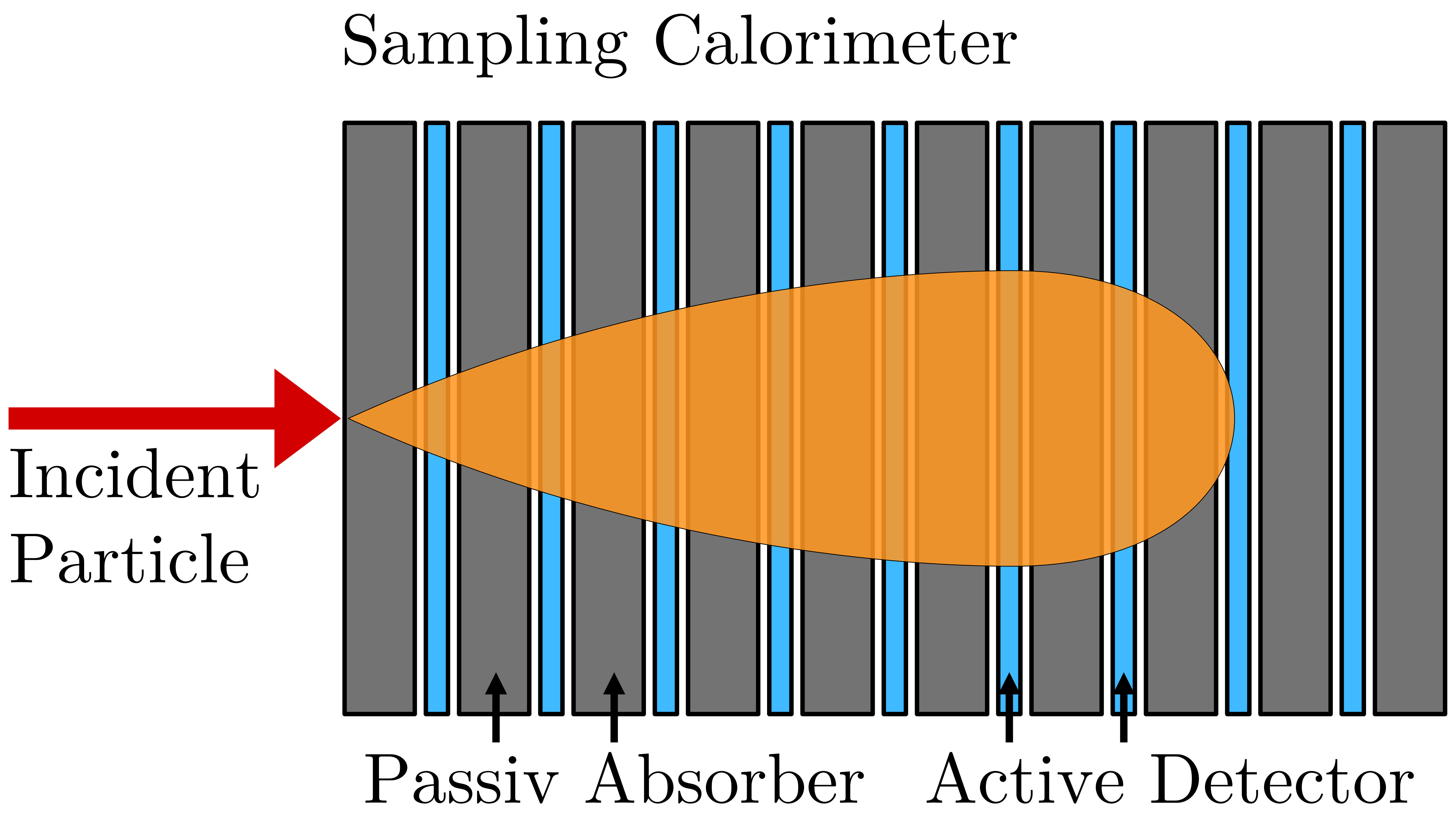




Thank you

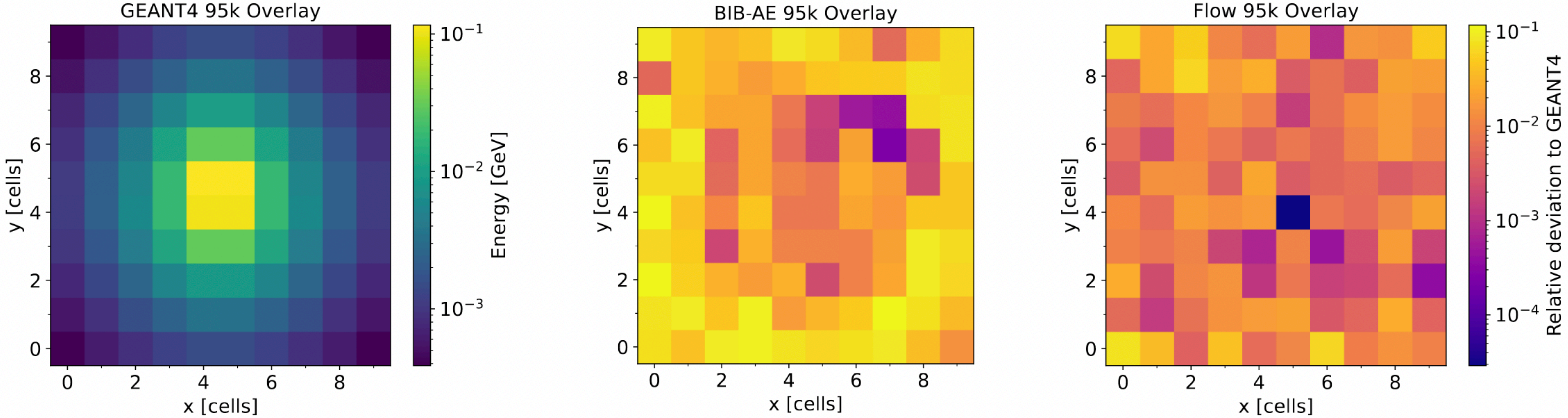
Calorimeter Simulation





- Calorimeter shower:
  - Particle interacts with material
  - Cascade of secondary particles
- Sampling Calorimeter
  - Interspersed active and passive parts
  - Only active segments are recorded

Shower Overlay



- Compare multiple flow approach to BIB-AE re-trained on 10x10x30 data set
- Overlay of flow showers has better agreement with GEANT4