Equivariant Point Cloud Generation for Particle Jets

Erik Buhmann*

with Gregor Kasieczka and Jesse Thaler

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Bundesministerium
für Bildung
und Forschung

FRIEDRICH NAUMANN
STIFTUNG Für die Freiheit.

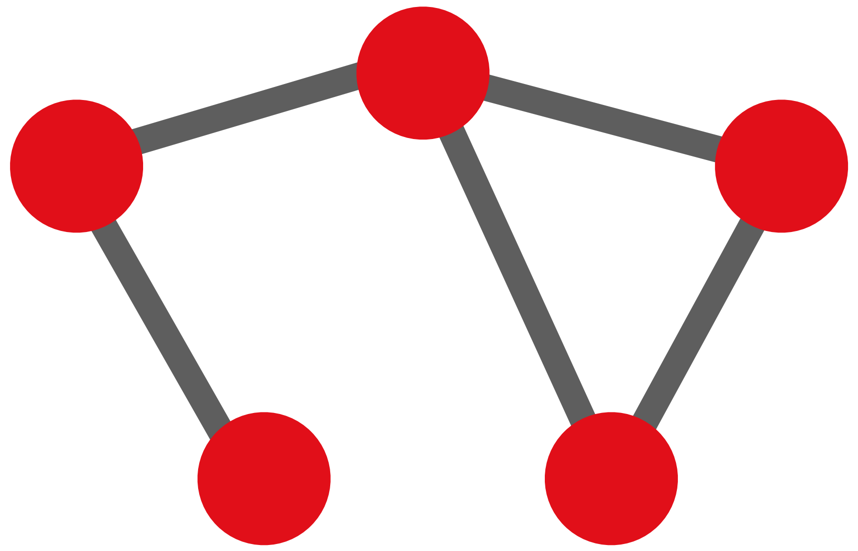
PIER
Partnership of
Universität Hamburg and DESY

* erik.buhmann@uni-hamburg.de

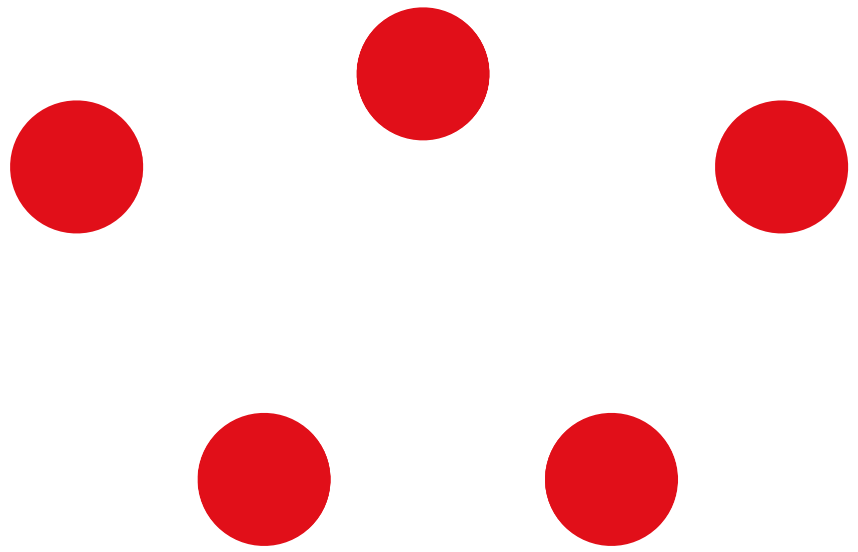
Generative Modelling of Point Clouds

- Particle jets are variable-length, unordered sets of particles
  - Natural representation as "point cloud"
- Requirements for generative model:
  A. Variable number of points N
  B. Permutation equivariance
  C. Fast generation for large multiplicities (ideally $\mathcal{O}(N)$)
- Current approaches using graph & transformer models
- Our simple Deep Sets-based generative model is equally performant, yet less complex and significantly faster

Graph



Point Cloud
(Graph without edges)



Point Cloud: $C = (P, \boldsymbol{g})$

P: Point's attributes $\boldsymbol{p}_i$

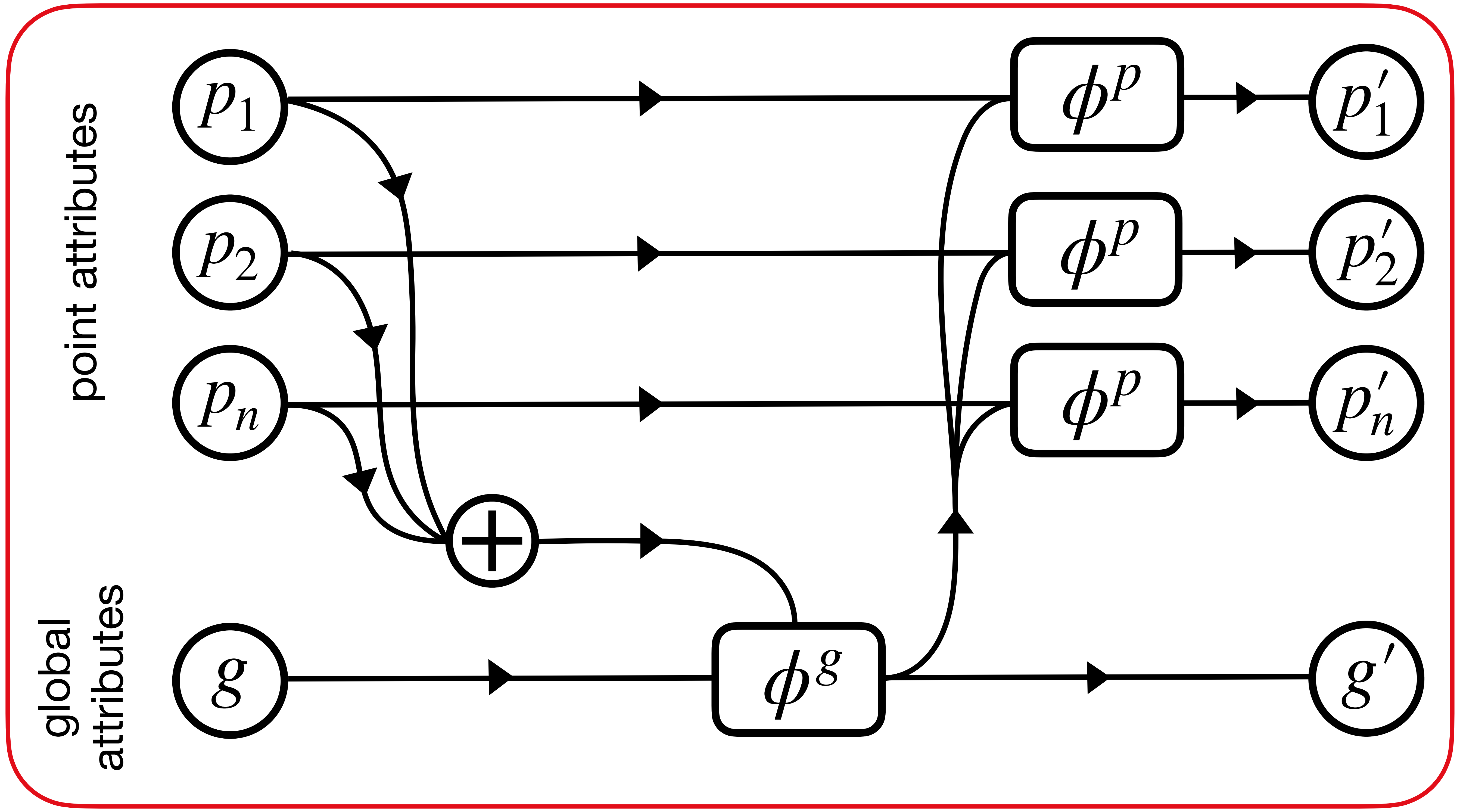
$\boldsymbol{g}$: Global attributes

Equivariant Point Cloud (EPiC) Layer

- 2 step update process per layer:
  1. Global attributes $\boldsymbol{g}$ are updated based on particle-wise attributes $\boldsymbol{p}_i$
  2. Particle attributes $\boldsymbol{p}_i$ are updated base on the updated global attributes $\boldsymbol{g}'$
- Control of communication between local vectors via:
  A. Length of global vector $[\boldsymbol{g}]$
  B. Number of stacked EPiC layers R
- Optimisation of these two hyperparameters to develop minimal generative model

$+$: sum & mean aggregation $\rho^{p->g}$

ϕ : 2-layer MLP



Computations in layer: (1) $\boldsymbol{g}'_i = \phi^g(\boldsymbol{g}, \rho^{p->g}(P))$

(2) $\boldsymbol{p}'_i = \phi^p(\boldsymbol{g}', \boldsymbol{p}_i)$

Equivariant Point Cloud (EPiC) GAN

- Generator (470k parameters):
  - Multiple R Equivariant Point Cloud (EPiC) layer stacked
- Discriminator (100k parameters):
  - Invariant Deep Sets structure
- Particle multiplicity defined when sampling (KDE estimation)
- Trained for 1000 epochs, best epoch chosen based on jet mass distribution on validation set

---

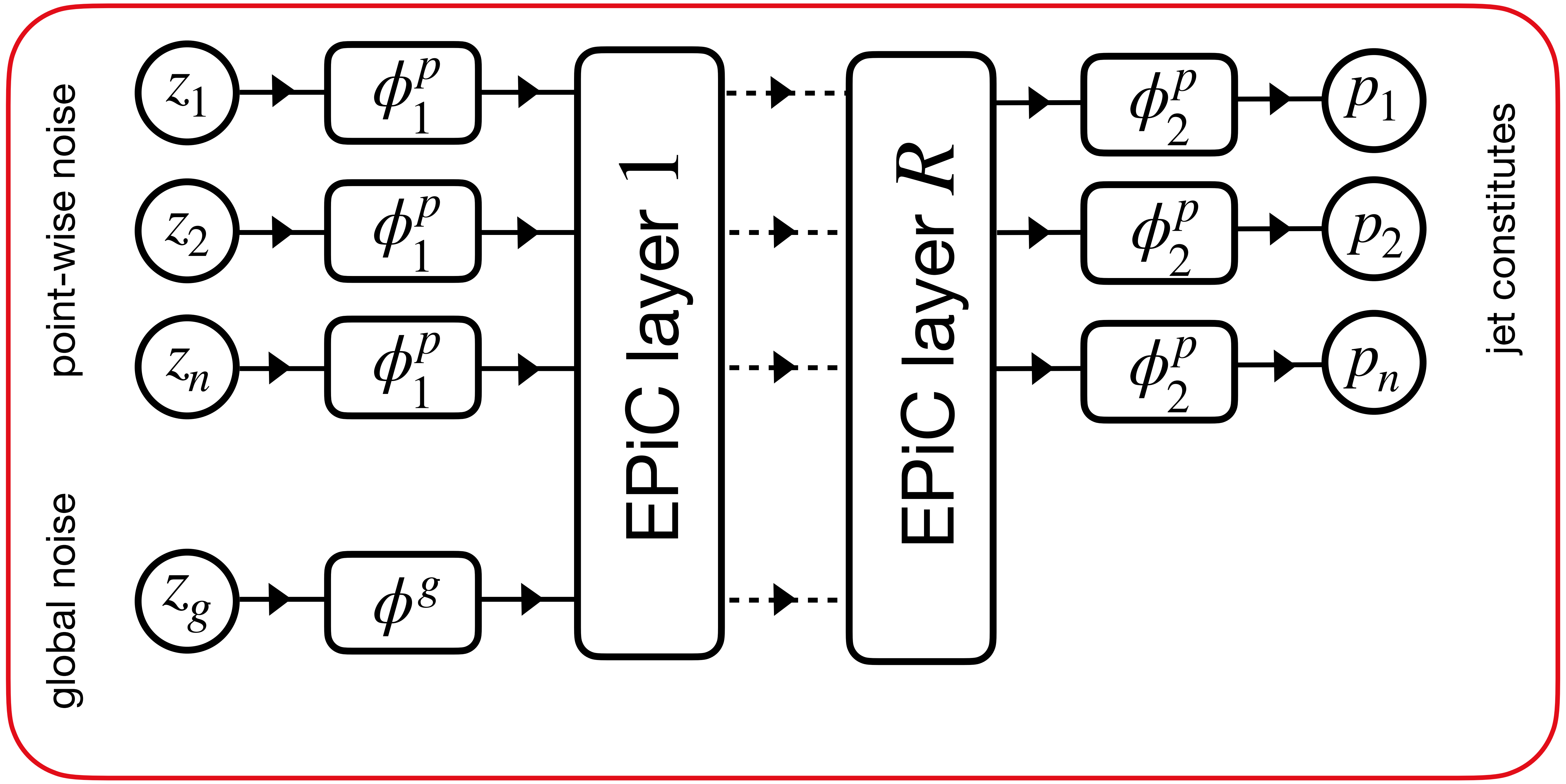
<u>LSGAN loss objective:</u>

$$\min_D L(D) = \frac{1}{2}\mathbb{E}[(D(x)-1)^2] + \frac{1}{2}\mathbb{E}[(D(G(z)))^2]$$

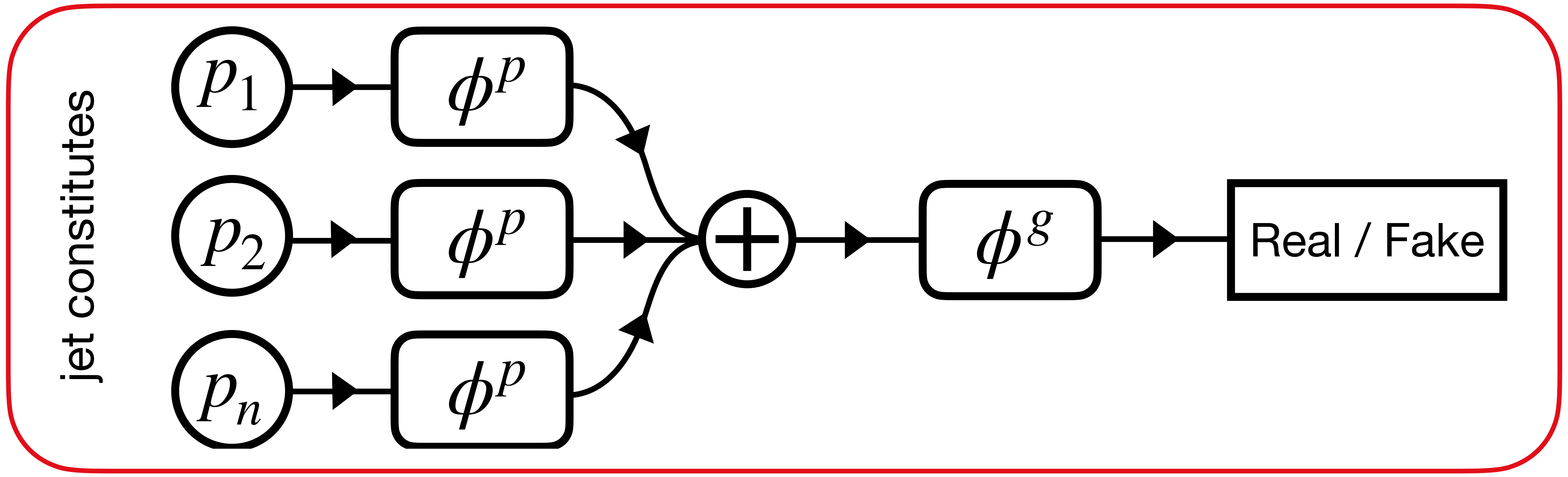
$$\min_D L(D) = \frac{1}{2}\mathbb{E}[(D(G(z))-1)^2]$$

[X Mao et al: Least Squares Generative Adversarial Networks]

Generator:

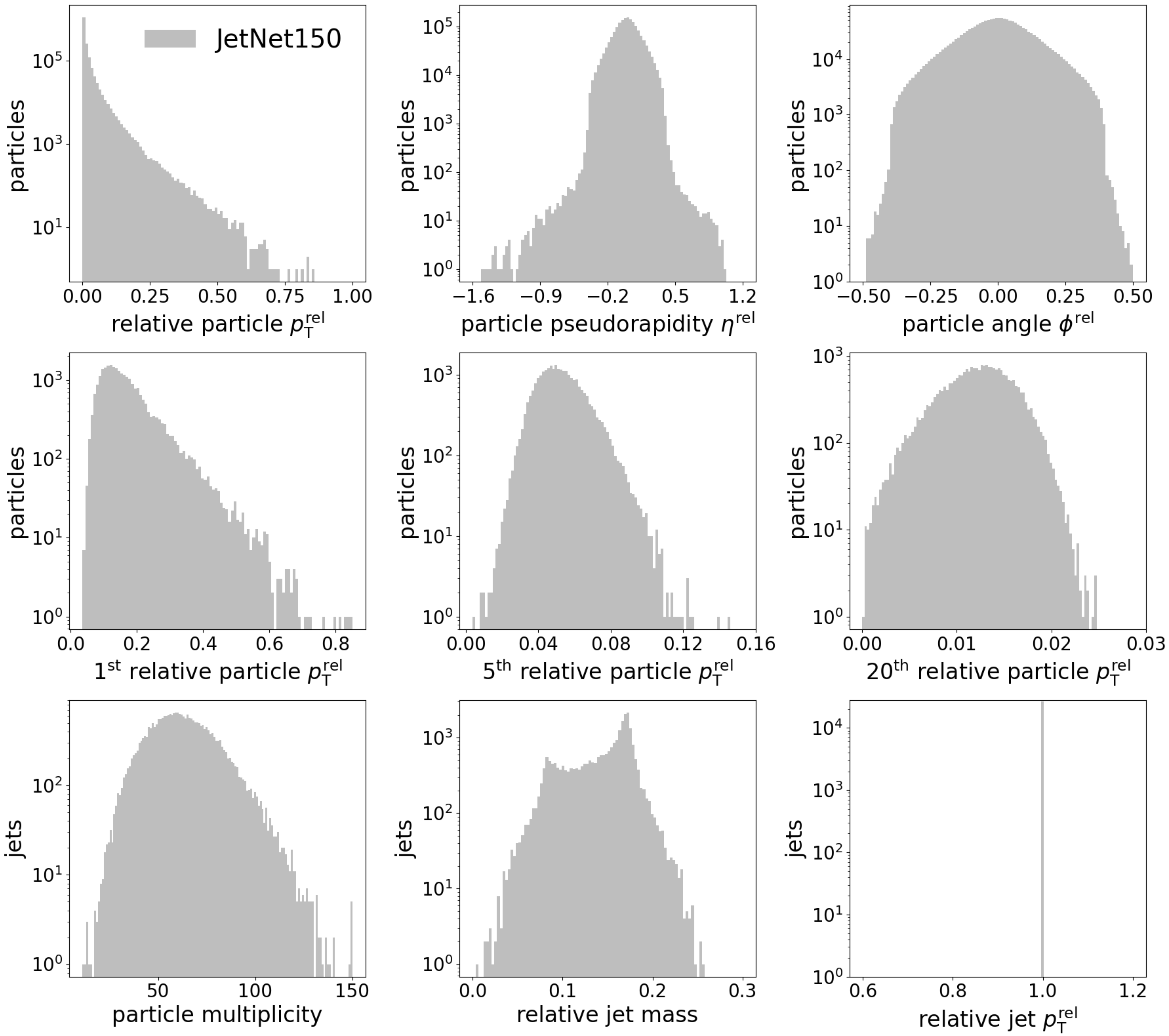


Discriminator:



Point Cloud Data

- Benchmark dataset: JetNet150 [1]
- Pythia simulated jets from proton-proton collisions
- Anti-k_{T} clustered with $R = 0.8$ and maximum particle multiplicity $N = 150$
- Particle collider coordinates normalised and centred
  - $p_{\mathrm{T}}^{\mathrm{rel}} = p_{\mathrm{T}}^{\mathrm{particle}} / p_{\mathrm{T}}^{\mathrm{jet}}$
  - $\eta^{\mathrm{rel}} = \eta^{\mathrm{particle}} - \eta^{\mathrm{jet}}$
  - $\phi^{\mathrm{rel}} = \phi^{\mathrm{particle}} - \phi^{\mathrm{jet}}$
- Here: Only top jets dataset

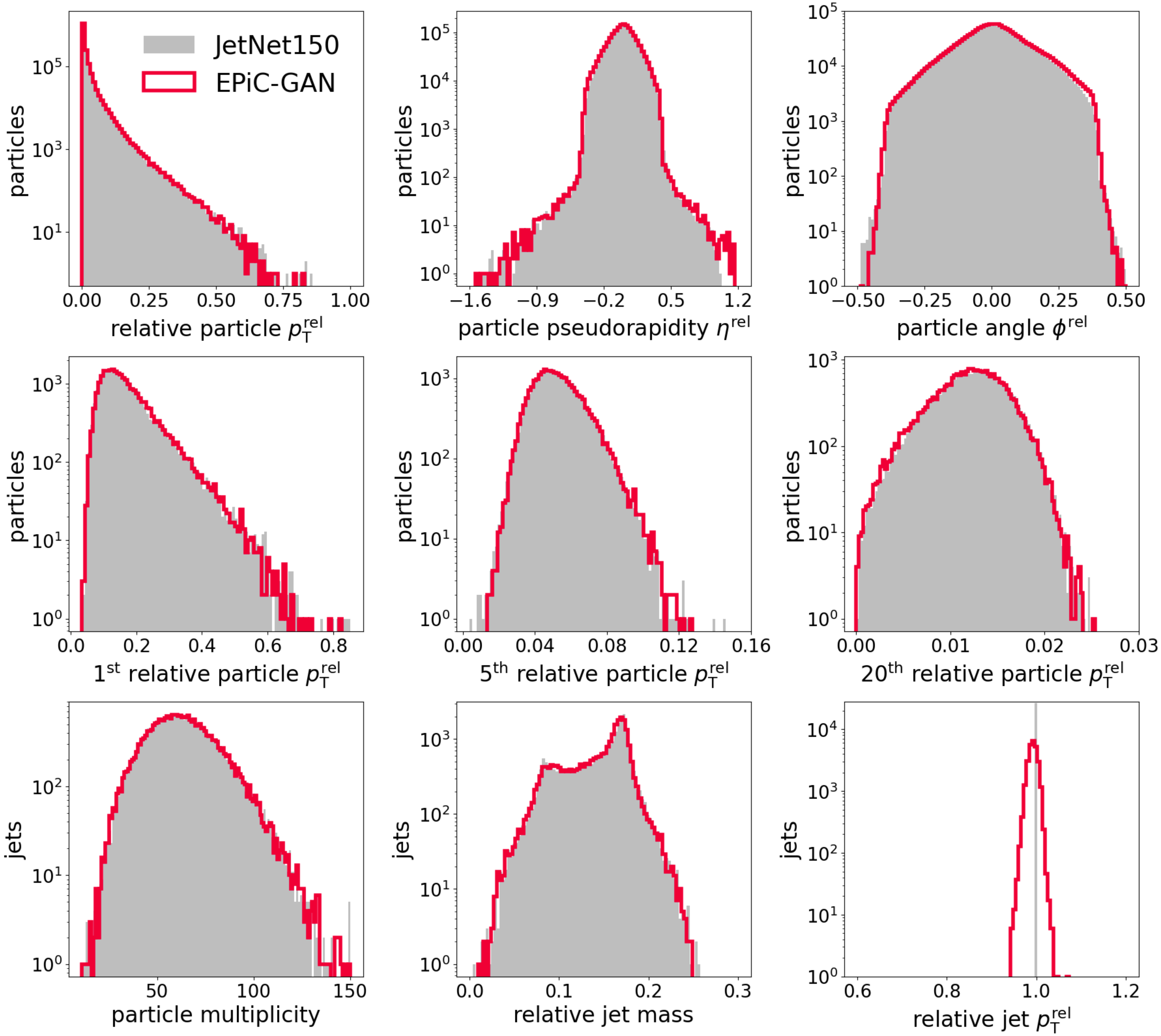


R Kansal, et al: Particle Cloud Generation with Message Passing Generative Adversarial Networks

[1] R Kansal, et al: JetNet150 (2.0.0) [Data set]. Zenodo

Results on JetNet150 tops

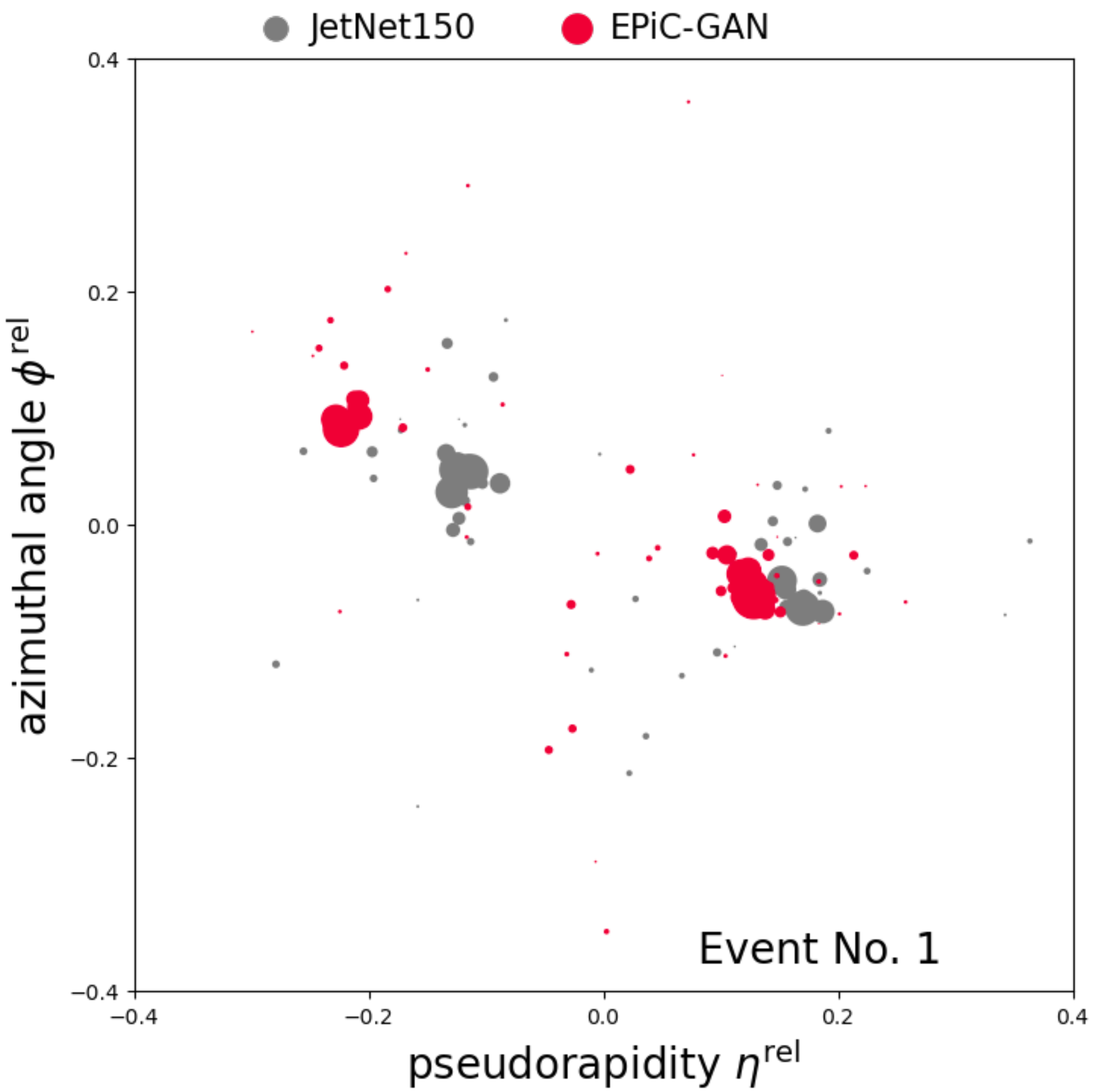
- EPiC-GAN setup:
  - $R = 6$ equivariant layers
  - $[\boldsymbol{g}] = 10$ global attributes
- High generative fidelity after six equivariant update steps
- Distributions well represented by EPiC-GAN, jet mass distributions particularly challenging
- Sharp relative jet p_{T} distribution challenging; can be resolved by calibration

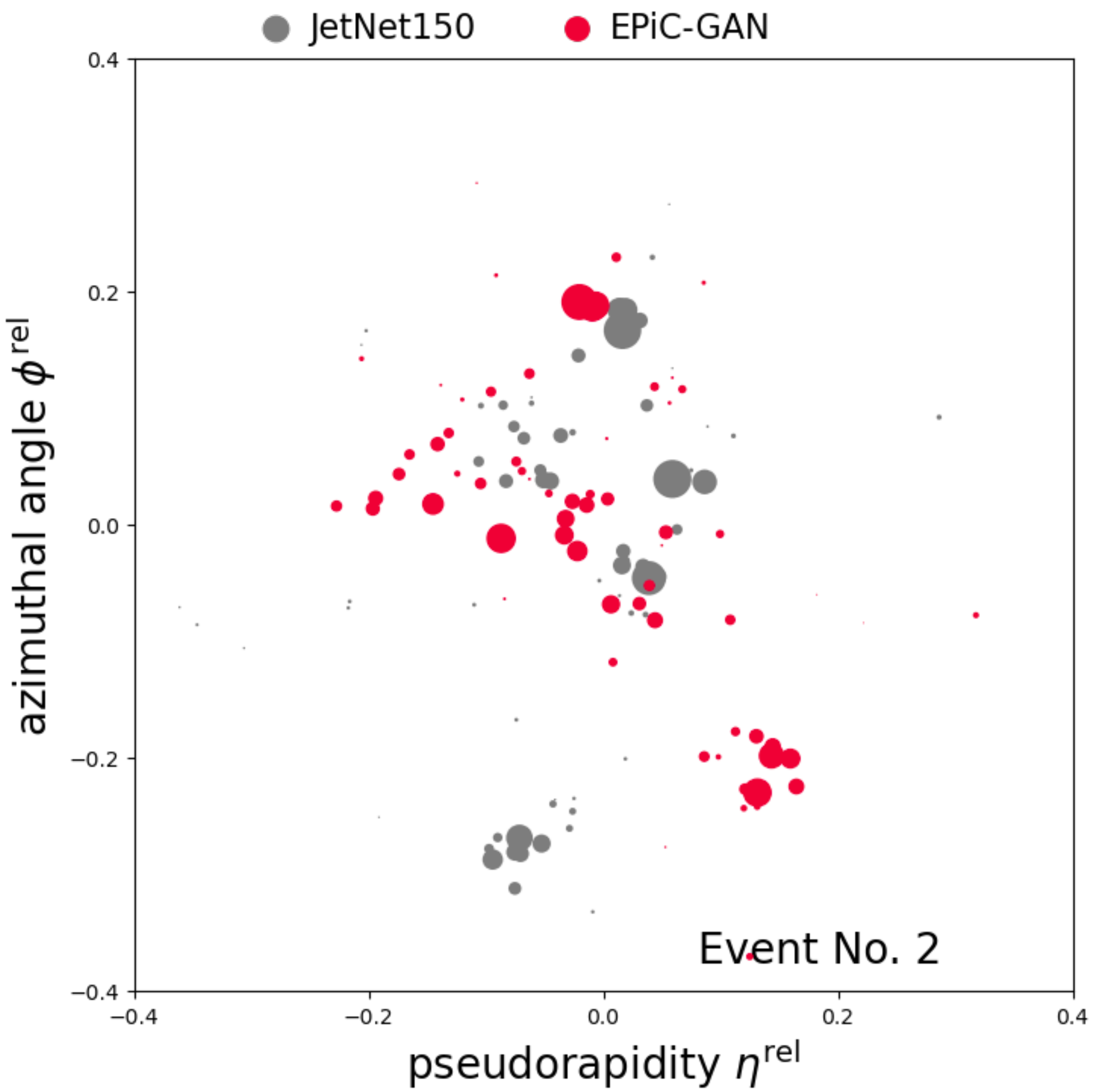


R Kansal, et al: Particle Cloud Generation with Message Passing Generative Adversarial Networks

[1] R Kansal, et al: JetNet150 (2.0.0) [Data set]. Zenodo

Point Clouds: Top jet visualisation





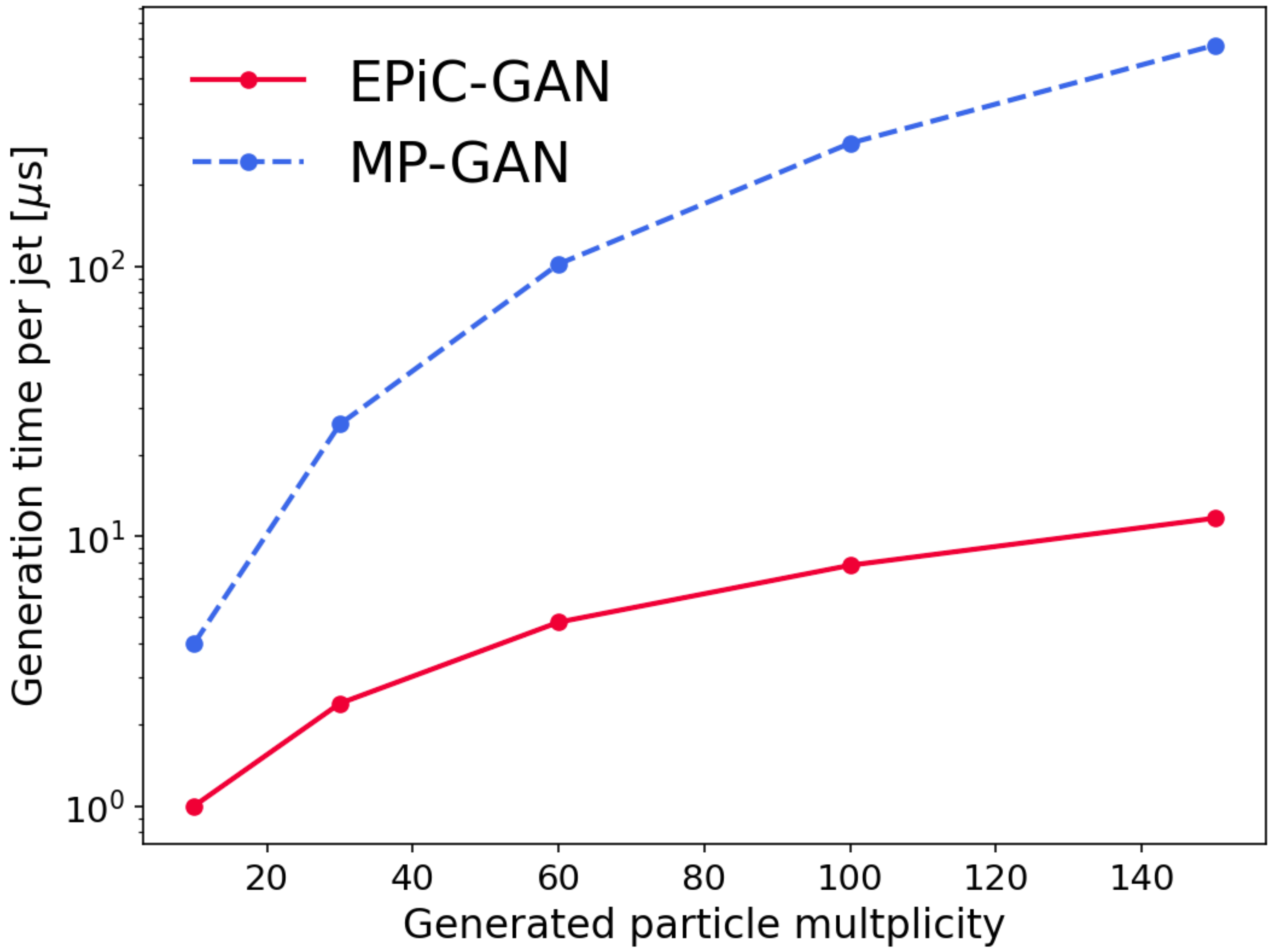
Point size corresponding to particle's transverse momentum $p_{\mathrm{T}}^{\mathrm{rel}}$

Scaling of generation time

- EPiC-GAN scales $\approx \mathcal{O}(N)$
- Message-Passing (MP)-GAN scales $\approx \mathcal{O}(N^2)$
- Batch size was adjusted for optimal generation speed

Generation time per jet:

Particles	Pythia[1]	MP-GAN*	EPiC-GAN*
$N = 30$	46 ms	26 μs	2 μs
$N = 150$	46 ms	660 μs	12 μs



* 500k jets generated on hardware:
Intel Gold-5218 (64 CPUs @ 2.3 Ghz), 384GB RAM, Nvidia A100-40GB

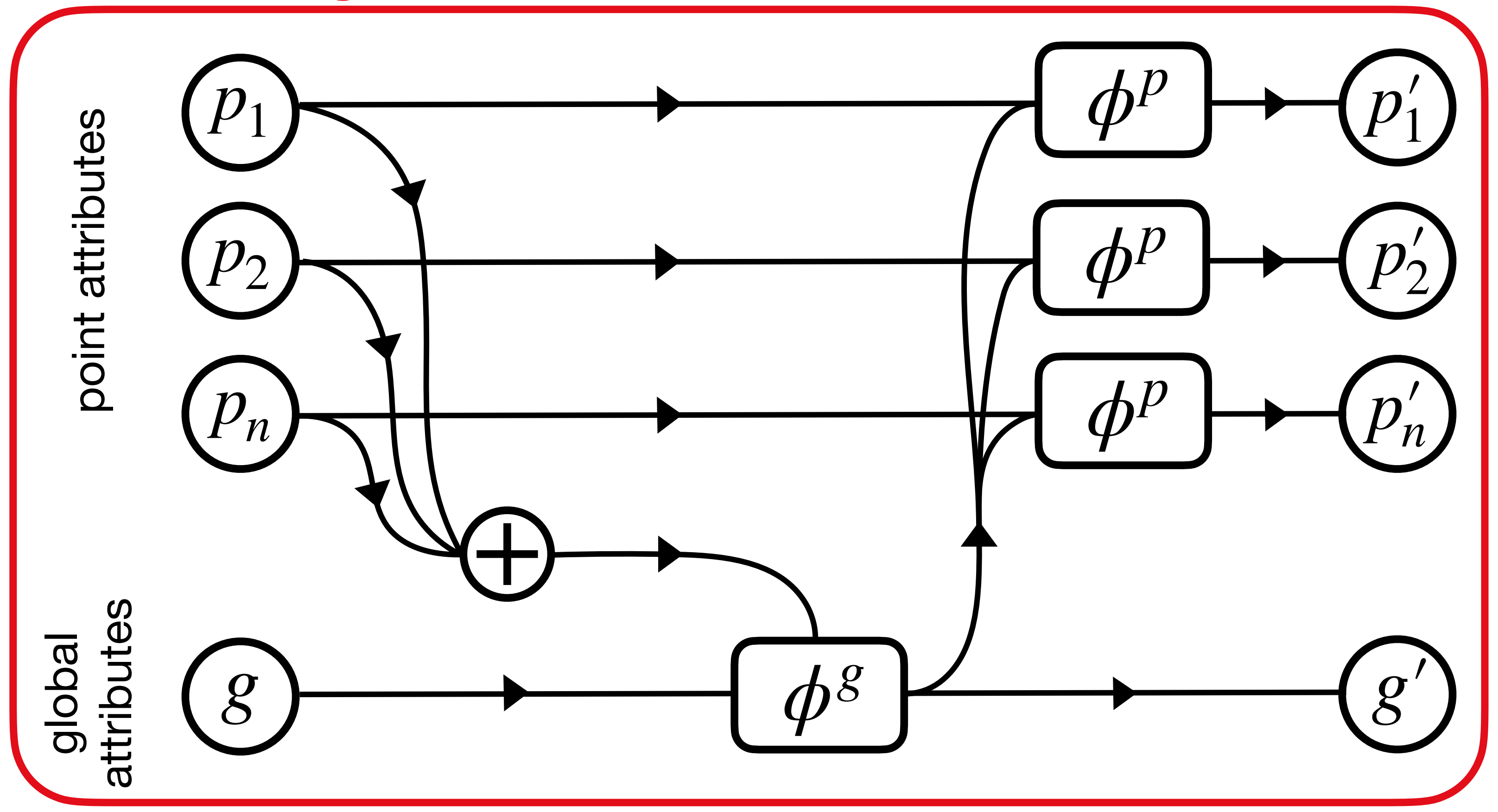
[1] R Kansal, et al: Particle Cloud Generation with Message Passing Generative Adversarial Networks

Summary

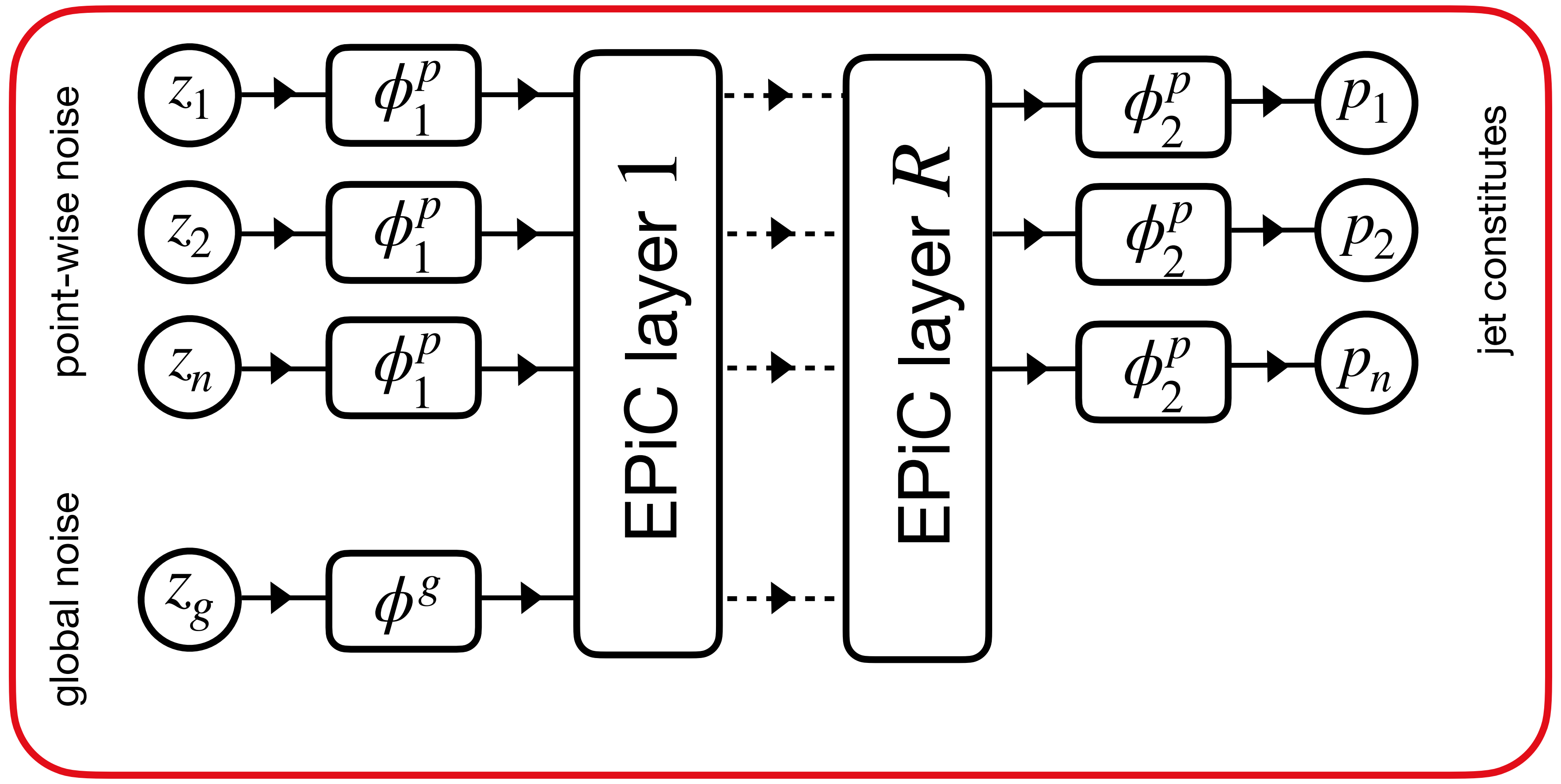
- Generative modelling of point clouds usually done with graph and transformer-based models
- Equivariant Point Cloud (EPiC)-GAN offers a simple and fast alternative
- Model generates realistic particle jets with variable particle multiplicity
- Fine-tuning possible for minimal global information & model complexity

Stay tuned for the paper!

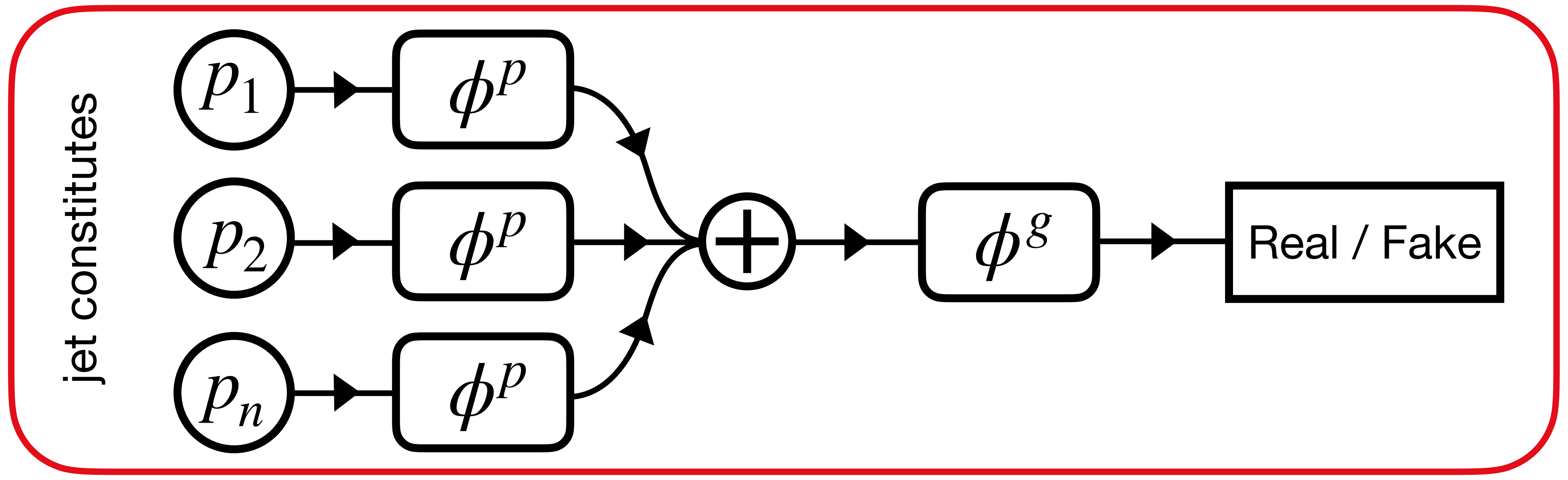
EPiC layer:



Generator:



Discriminator:



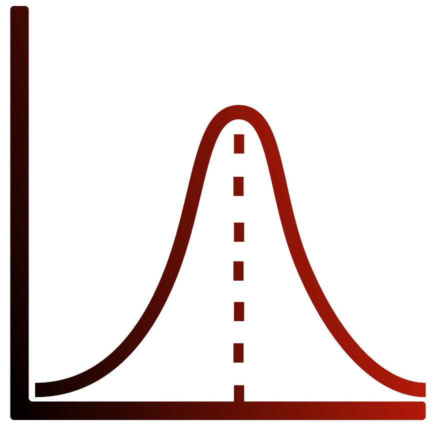
Bonus slides

EPiC-GAN Hyperparameter

- Trained for 500 / 1000 epochs (quarks / tops)
- Learning rate of Adam optimiser: 1e-4
- Maximum batch size = 128
- Pre-processing: Each feature standardised to $\mathcal{N}(0,5\sigma)$
- Generator: 470 parameters (with $R = 6$ & $[\boldsymbol{g}] = 10$)
- Discriminator: 100k parameters
- All fully connected layer size = 128
- $[\boldsymbol{p}_i]$: For each particle one local noise vector of length 10
- Training / validation / test set: 127k / 26k / 26k
- Shown plots from test set with 26k events

Generative Models: Use cases in HEP research

Amplification of statistics



- Strong inductive bias of architectures help models to learn underlying distribution
- Powerful data augmentation technique

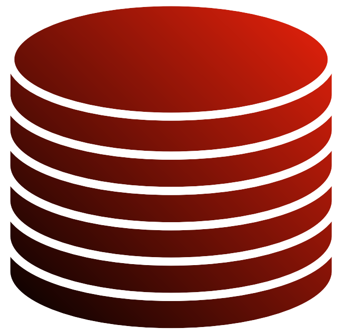
S Bieringer et al: Calomplification – The Power of Generative Calorimeter Models

A Butter et al: GANplifying Event Samples

J Kummer et al: Radio Galaxy Classification with wGAN-Supported Augmentation

[…]

Amortised computation



- Minimisation of local computing resources by upfront central model training
- Storing model weights instead of data

M Paganini et al: CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks

EB, Sascha Diefenbacher et al: Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speeds

A Butter et al: Machine Learning and LHC Event Generation

[…]

Generation from detector data

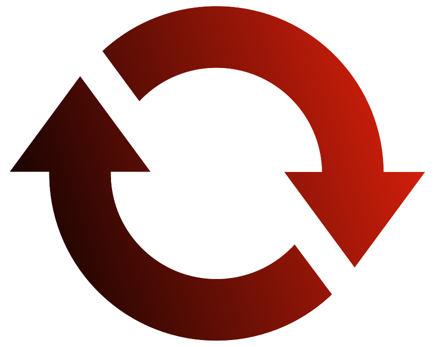
- Unsupervised training on real events instead of tuning Monte Carlo simulations
- I.e. for estimation of background densities

JN Howad et al: Learning to Simulate High Energy Particle Collisions from Unlabeled Data

A Hallin et al: Classifying Anomalies THrough Outer Density Estimation (CATHODE)

[…]

Differentiable models



- Optimisation of experimental setup based on explicit data likelihood
- Backpropagation through analysis chain

T Dorigo et al: Toward the End-to-End Optimization of Particle Physics Instruments with Differentiable Programming: a White Paper

A Adelmann et al: New directions for surrogate models and differentiable programming for High Energy Physics detector simulation

[…]