Applications of Sets & Graphs in HEP

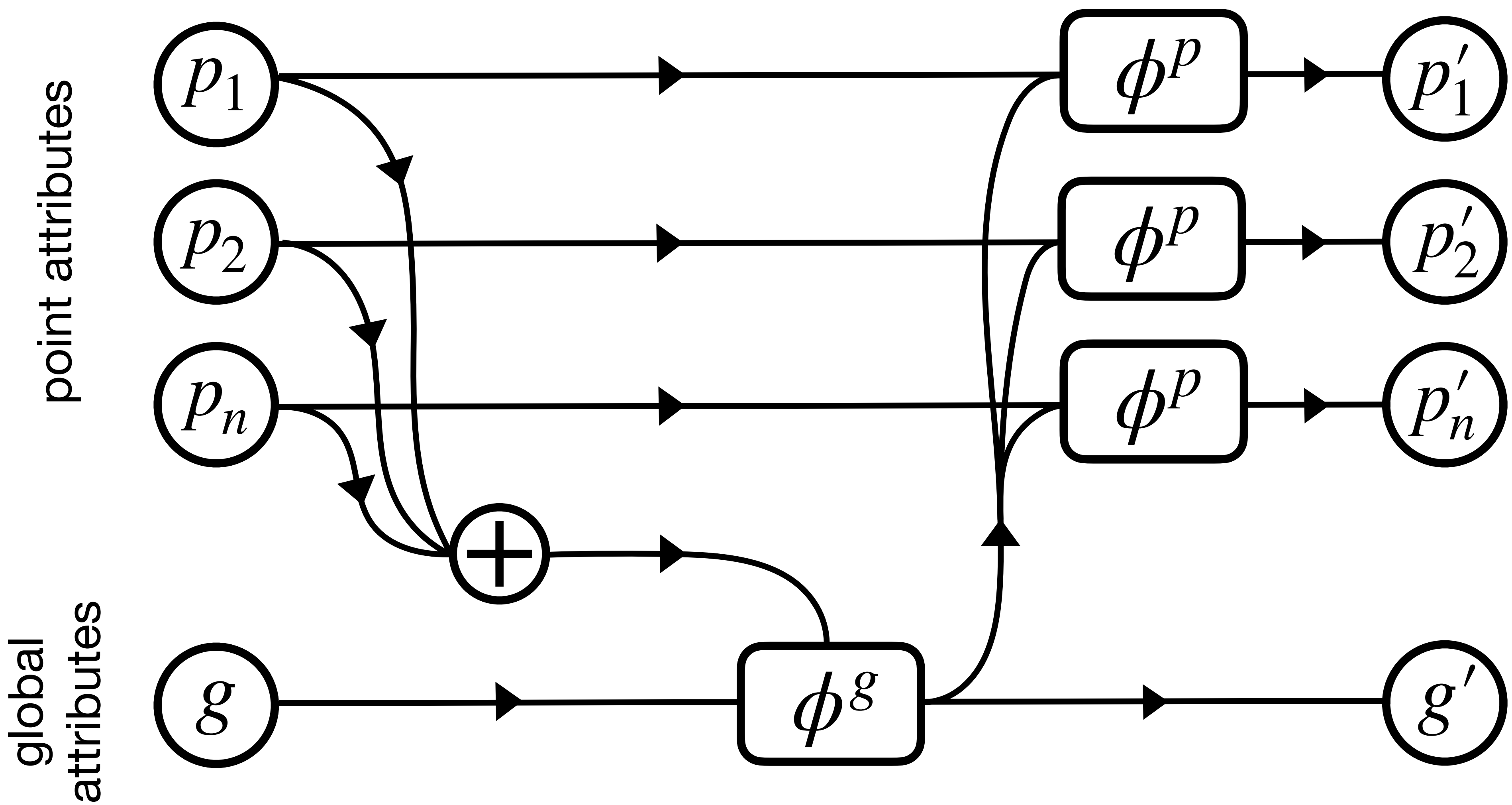
Sanmay Ganguly
(sanmay@iitk.ac.in)

Machine Learning for Particle and Astroparticle Physics, IOPB 2024

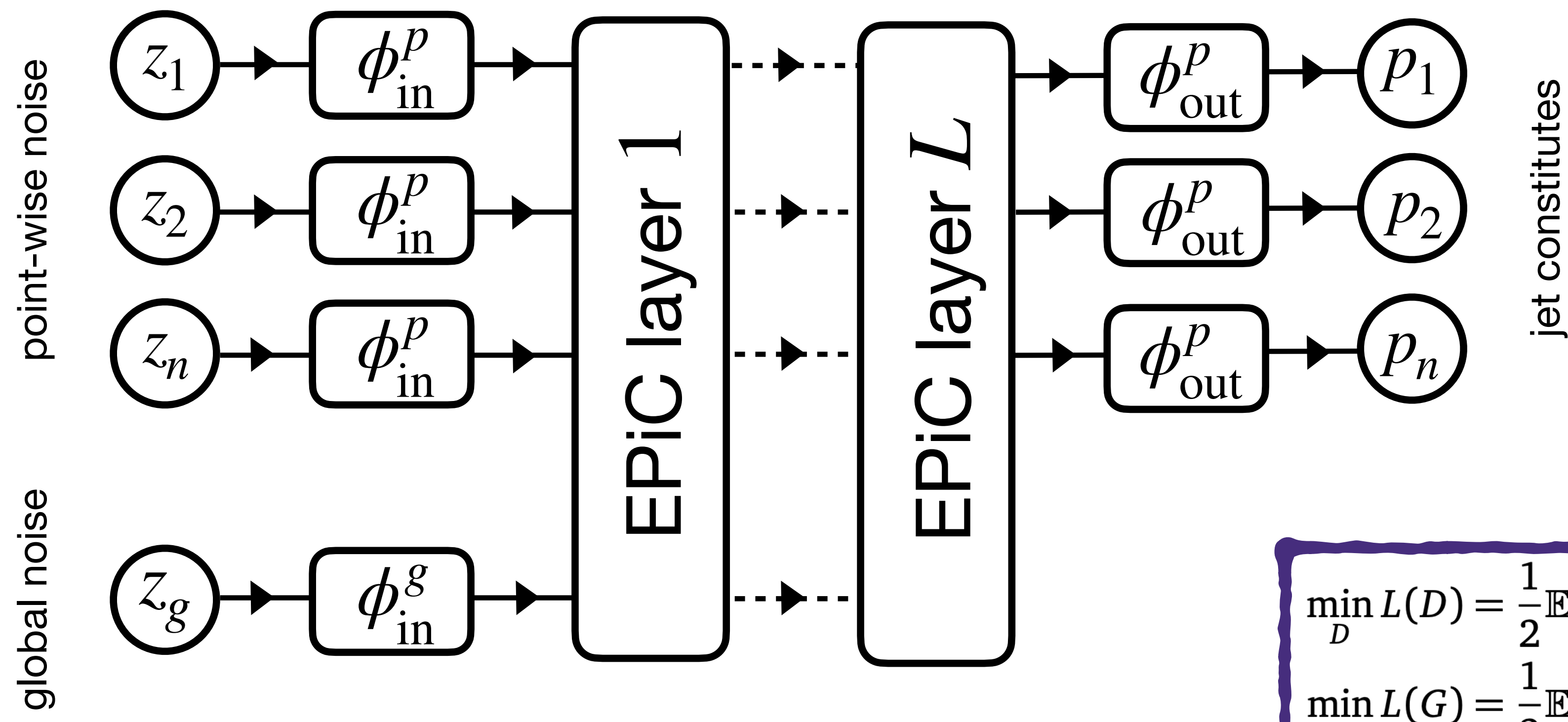
05/07/2024

Point cloud generation

EPIC paper as an example : arXiv 2301.08128 / https://github.com/uhh-pd-ml/EPiC-GAN/tree/main

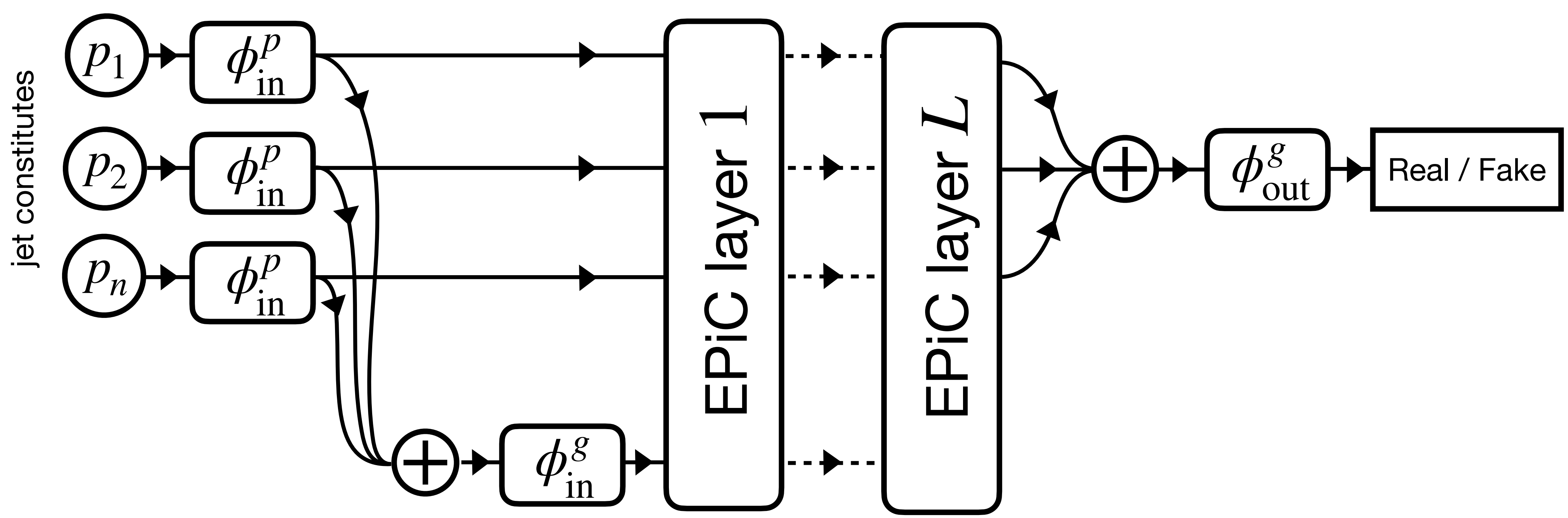


The generator and discriminator

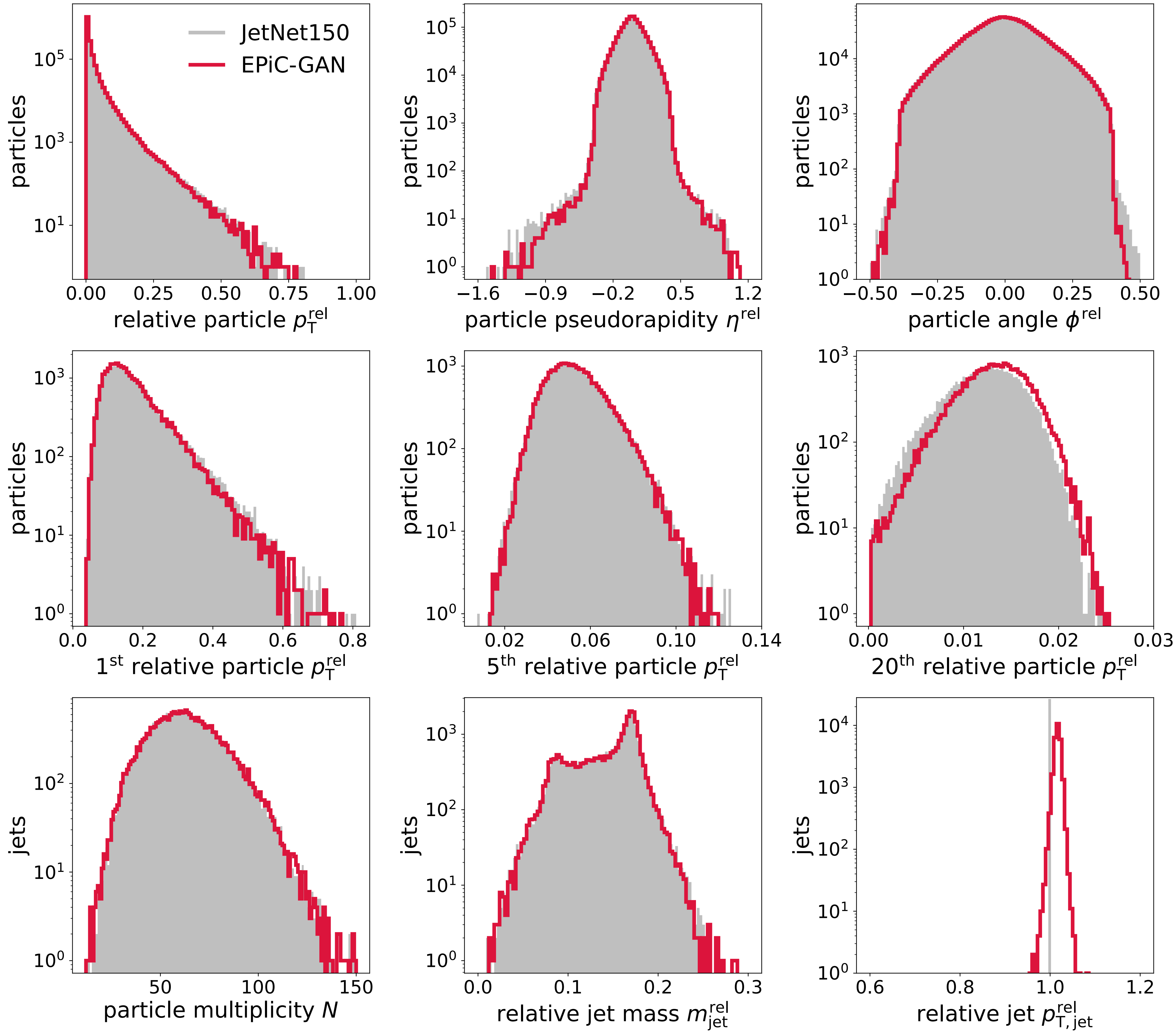


$$\min_D L(D) = \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[(D(\boldsymbol{x})-1)^2] + \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[(D(G(\boldsymbol{z})))^2],$$

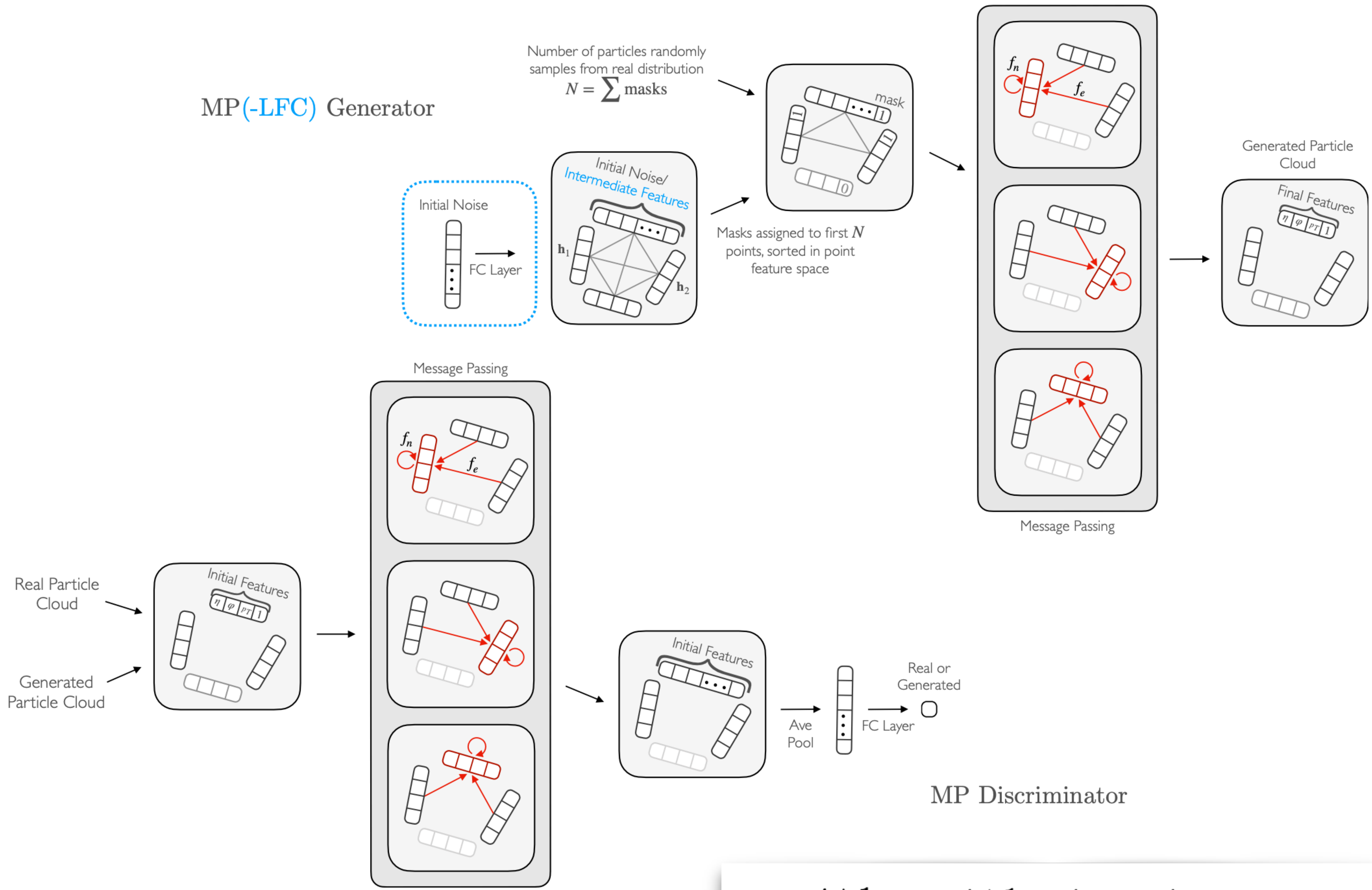
$$\min_G L(G) = \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[(D(G(\boldsymbol{z}))-1)^2].$$



The performance



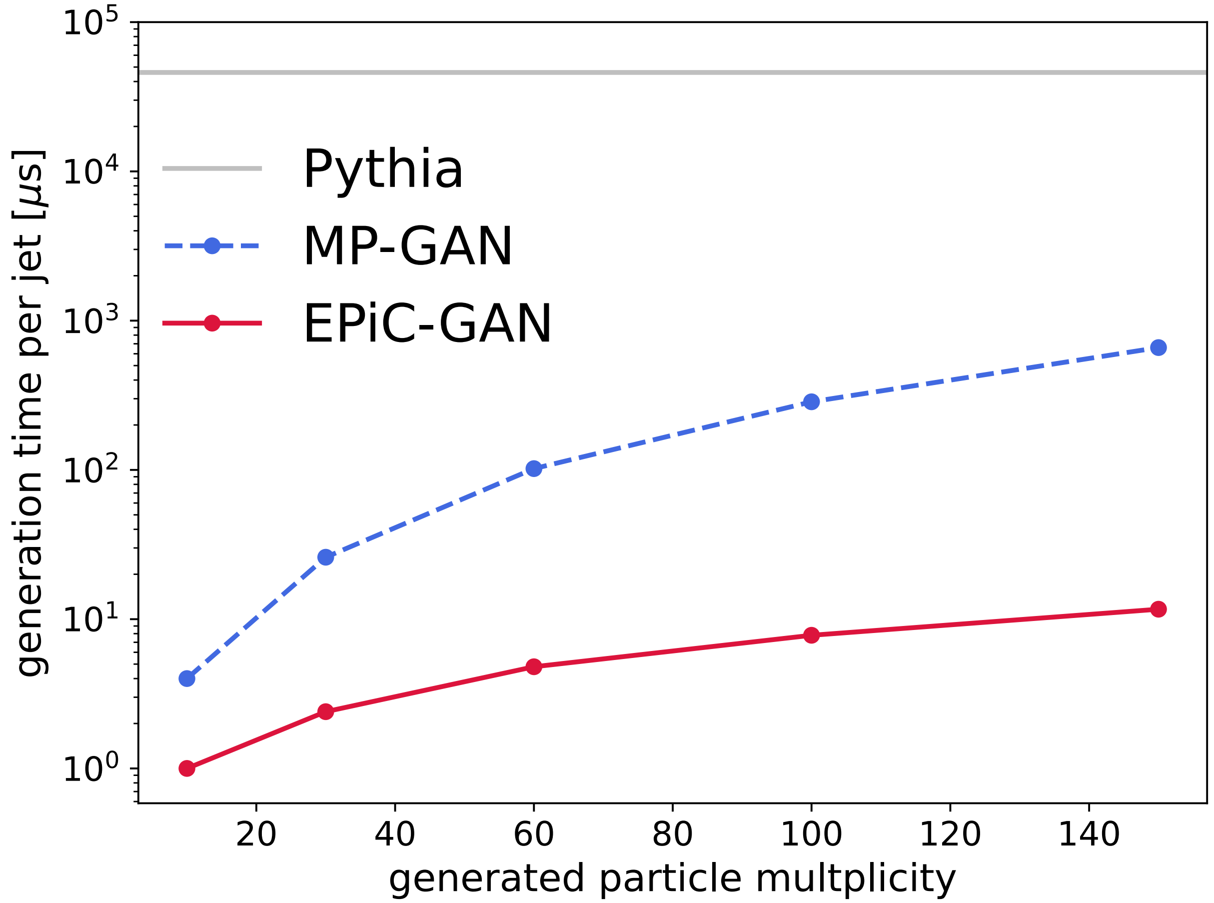
The MPGAN

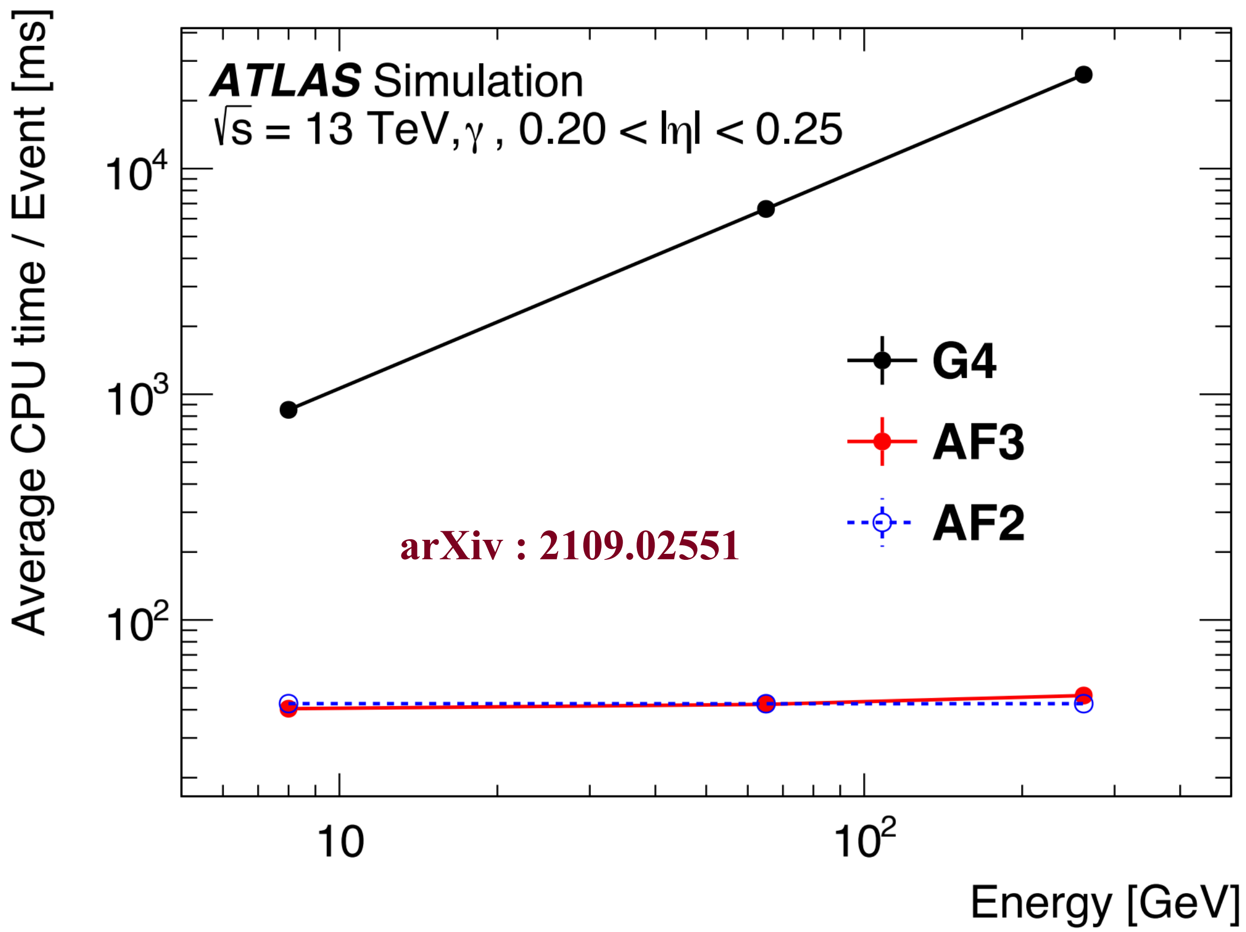


$$\mathbf{m}_{ij}^{t+1} = f_e^{t+1}(\mathbf{h}_i^t \oplus \mathbf{h}_j^t),$$

$$\mathbf{h}_i^{t+1} = f_n^{t+1}(\mathbf{h}_i^t \oplus \sum_{j \in J} \mathbf{m}_{ij}^{t+1}),$$

The major gain





A sure shot motivations to use PC generative models for HEP used cases

Generative models in HEP

arXiv > hep-ph > arXiv:2203.07460

Search...

Help | Adva

High Energy Physics – Phenomenology

[Submitted on 14 Mar 2022 (v1), last revised 28 Dec 2022 (this version, v2)]

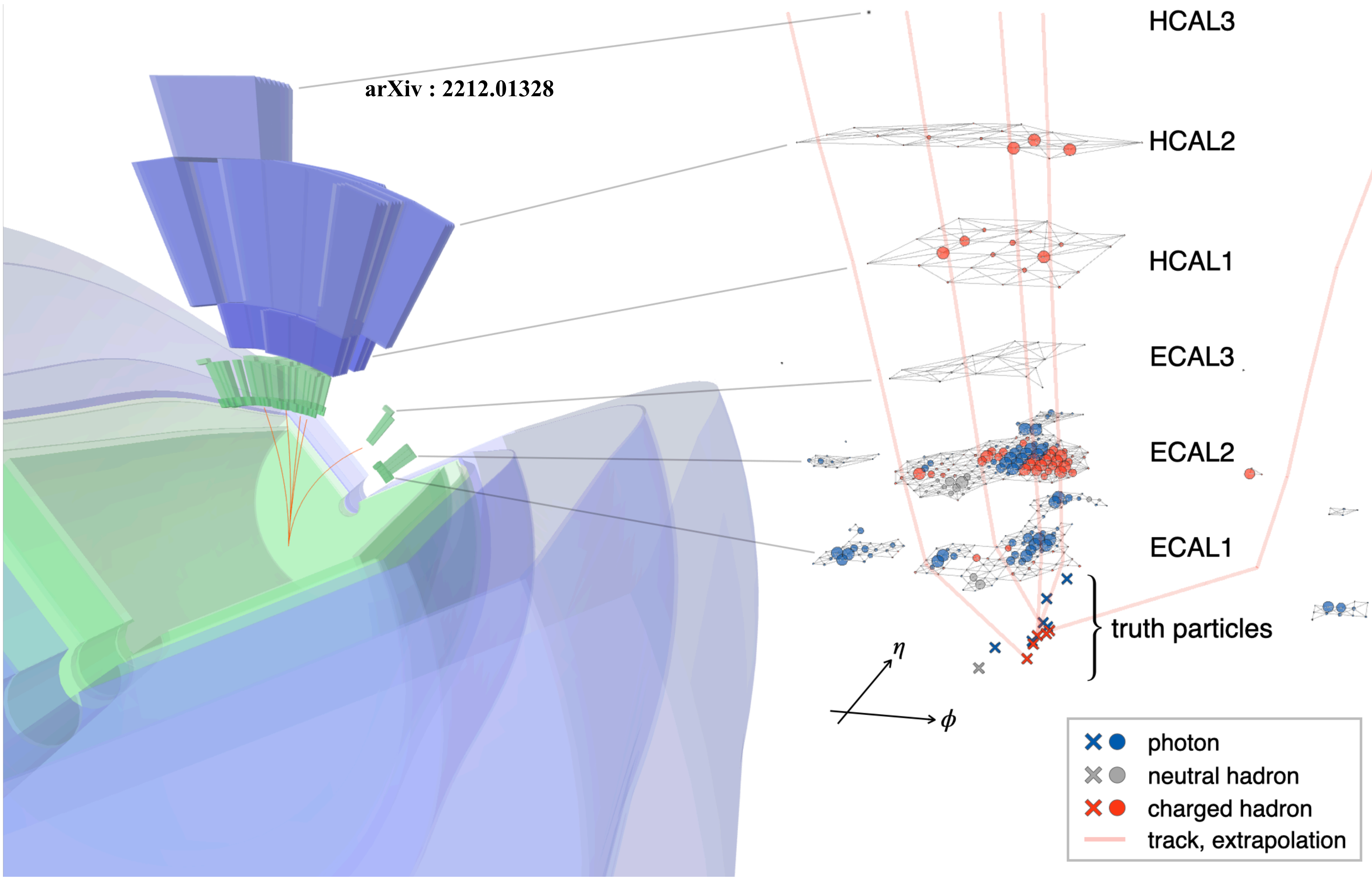
Machine Learning and LHC Event Generation

Anja Butter (ed), Tilman Plehn (ed), Steffen Schumann (ed), Simon Badger, Sascha Caron, Kyle Cranmer, Francesco Armando Di Bello, Etienne Dreyer, Stefano Forte, Sanmay Ganguly, Dorival Gonçalves, Eilam Gross, Theo Heimel, Gudrun Heinrich, Lukas Heinrich, Alexander Held, Stefan Höche, Jessica N. Howard, Philip Ilten, Joshua Isaacson, Timo Janßen, Stephen Jones, Marumi Kado, Michael Kagan, Gregor Kasieczka, Felix Kling, Sabine Kraml, Claudius Krause, Frank Krauss, Kevin Kröninger, Rahool Kumar Barman, Michel Luchmann, Vitaly Magerya, Daniel Maitre, Bogdan Malaescu, Fabio Maltoni, Till Martini, Olivier Mattelaer, Benjamin Nachman, Sebastian Pitz, Juan Rojo, Matthew Schwartz, David Shih, Frank Siegert, Roy Stegeman, Bob Stienen, Jesse Thaler, Rob Verheyen, Daniel Whiteson, Ramon Winterhalder, Jure Zupan

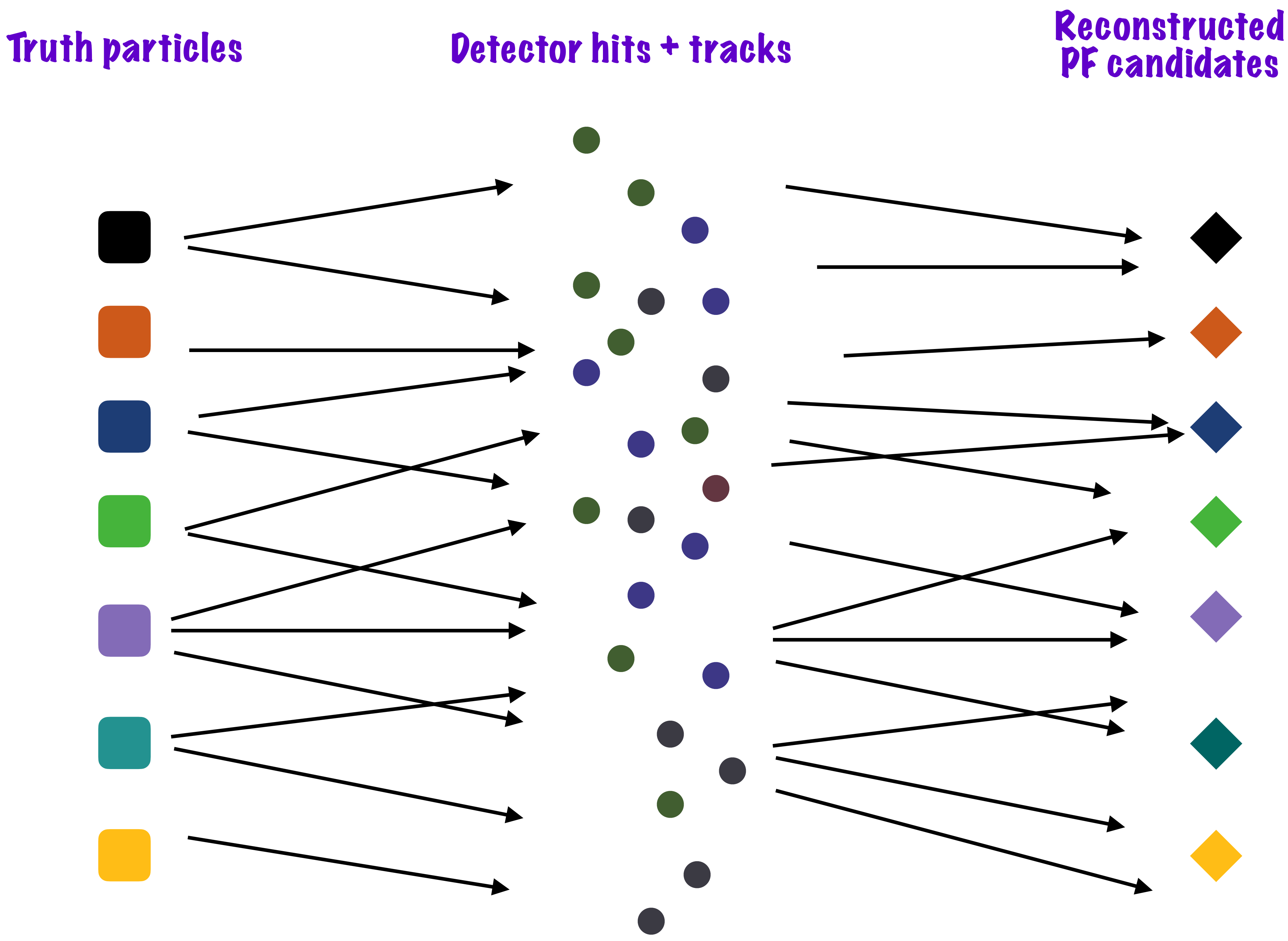
First-principle simulations are at the heart of the high-energy physics research program. They link the vast data output of multi-purpose detectors with fundamental theory predictions and interpretation. This review illustrates a wide range of applications of modern machine learning to event generation and simulation-based inference, including conceptional developments driven by the specific requirements of particle physics. New ideas and tools developed at the interface of particle physics and machine learning will improve the speed and precision of forward simulations, handle the complexity of collision data, and enhance inference as an inverse simulation problem.

Give a look if you want to see the extent of generative models in HEP

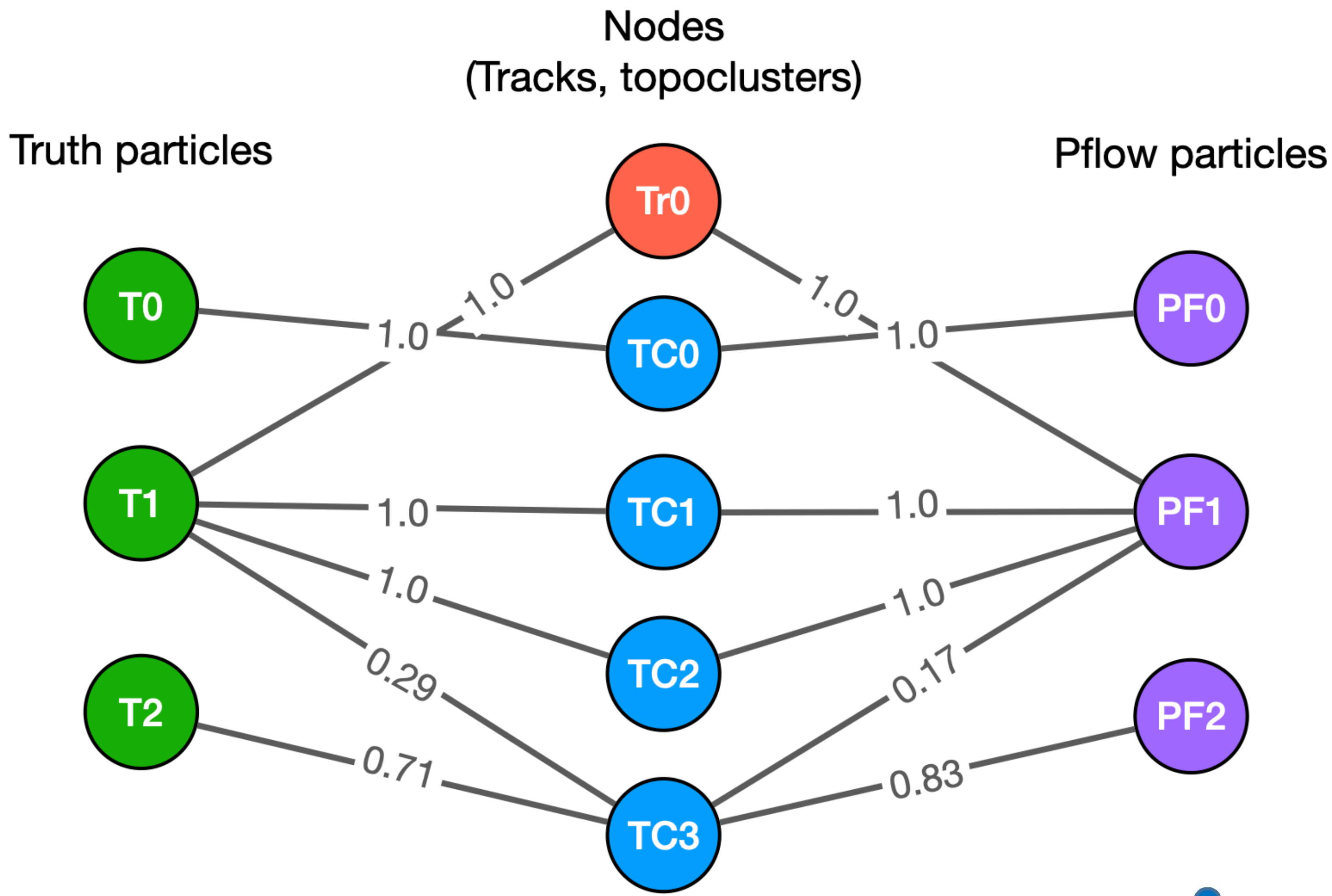
The PF reconstruction algorithm



What’s the core data structure?



What's the core data structure?




arXiv > hep-ph > arXiv:2309.17351

High Energy Physics – Phenomenology

[Submitted on 29 Sep 2023 (v1), last revised 19 Feb 2024 (this version, v2)]

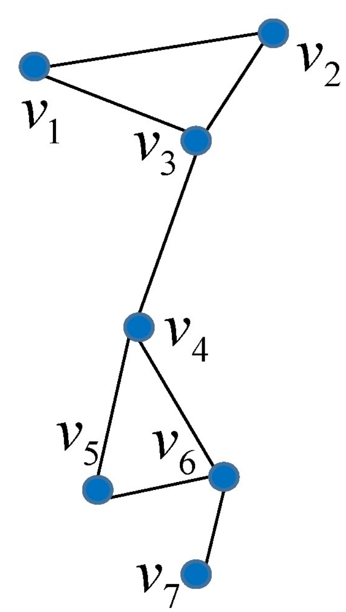
Hypergraphs in LHC Phenomenology -- The Next Frontier of IRC-Safe Feature Extraction

Partha Konar, Vishal S. Ngairangbam, Michael Spannowsky

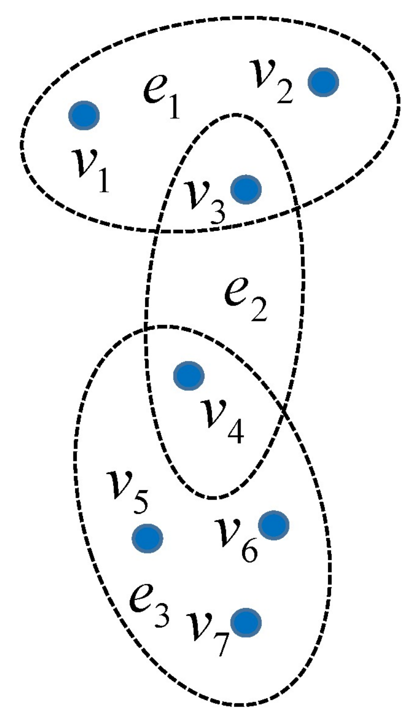
In this study, we critically evaluate the approximation capabilities of existing infra-red and collinear (IRC) safe feature extraction algorithms, namely Energy Flow Networks (EFNs) and Energy-weighted Message Passing Networks (EMPNs). Our analysis reveals that these algorithms fall short in extracting features from any N-point correlation that isn't a power of two, based on the complete basis of IRC safe observables, specifically C-correlators. To address this limitation, we introduce the Hypergraph Energy-weighted Message Passing Networks (H-EMPNs), designed to capture any N-point correlation among particles efficiently. Using the case study of top vs. QCD jets, which holds significant information in its 3-point correlations, we demonstrate that H-EMPNs targeting up to N=3 correlations exhibit superior performance compared to EMPNs focusing on up to N=4 correlations within jet constituents.

https://www.mdpi.com/2072-4292/9/5/506

Learning Flow is essentially learning the incidence matrix of a Hypergraph.



(a) Simple graph

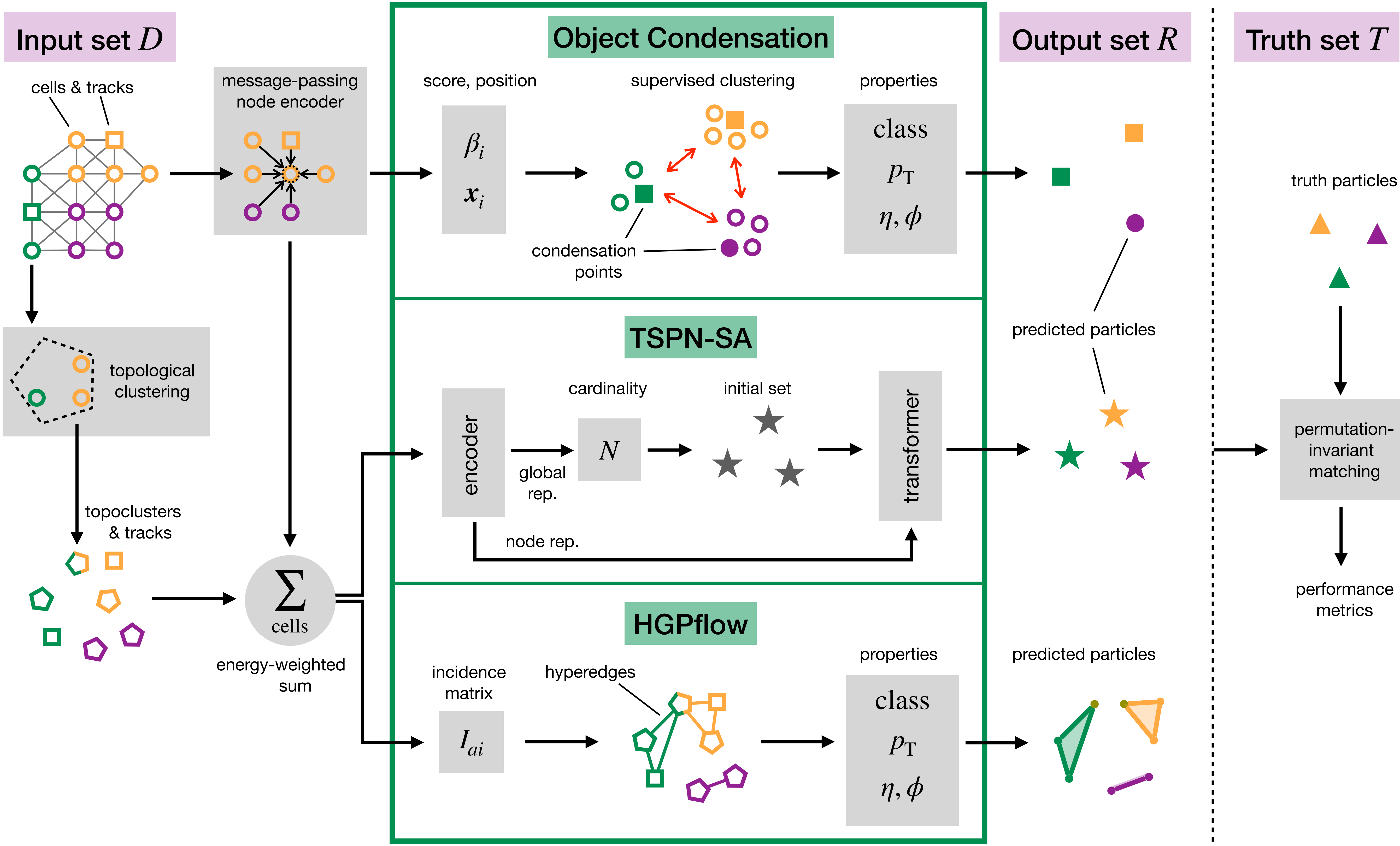


(b) Hypergraph **G**

	e_1	e_2	e_3
v_1	1	0	0
v_2	1	0	0
v_3	1	1	0
v_4	0	1	1
v_5	0	0	1
v_6	0	0	1
v_7	0	0	1

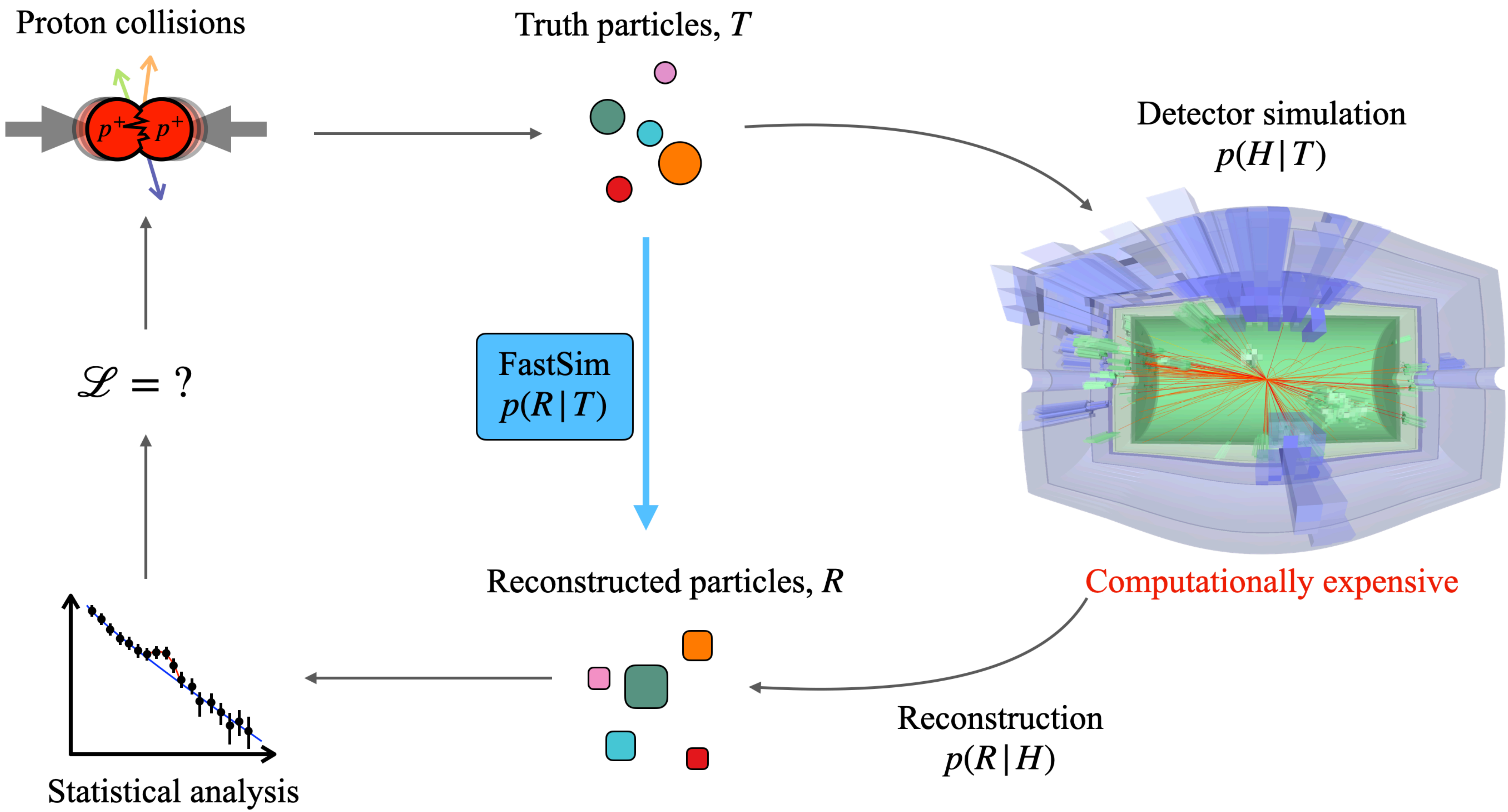
(c) Incidence matrix **H**

The network flow comparisons



A generative model for Flow

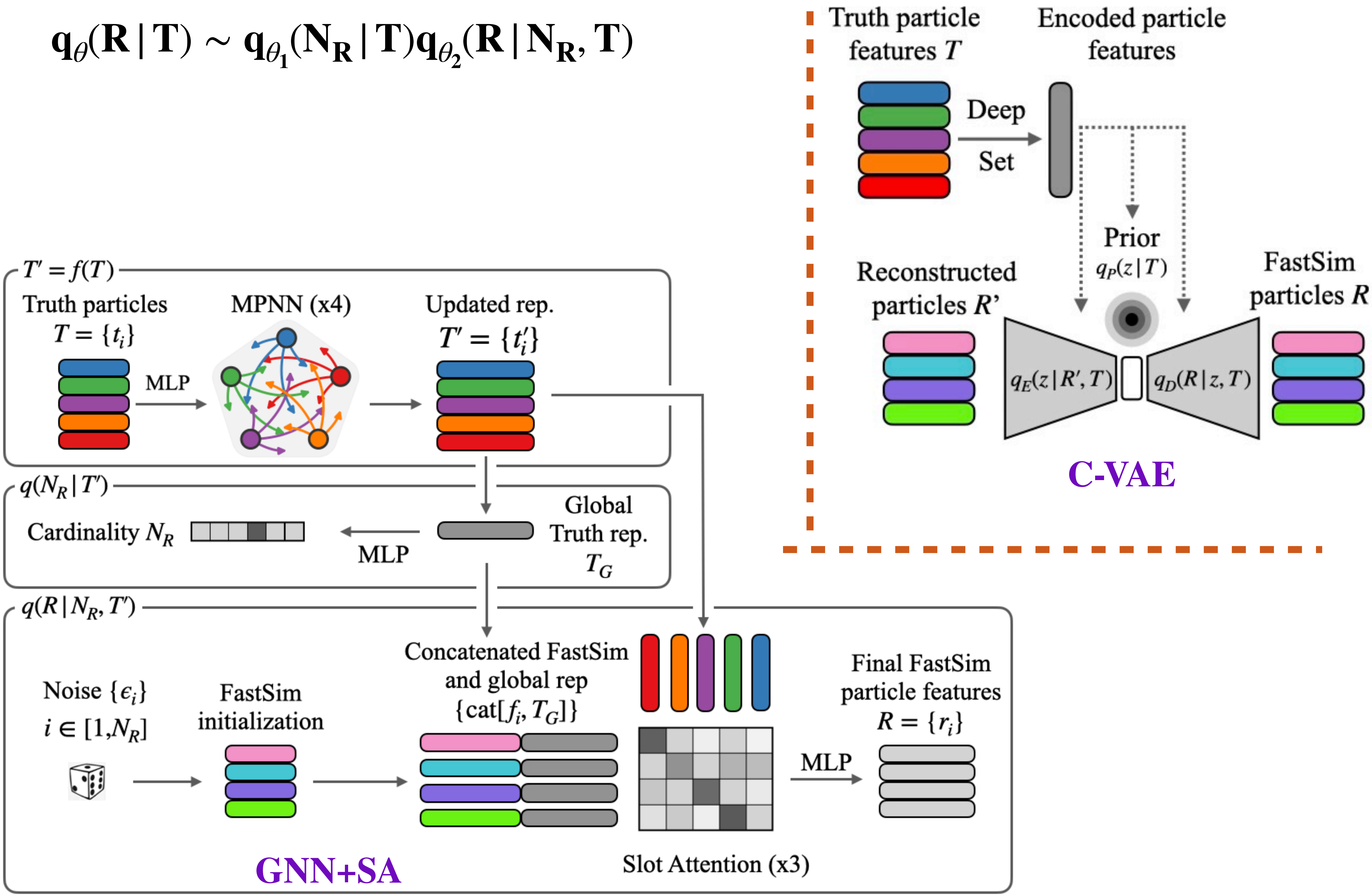
arXiv : 2211.06406



$$R \sim p(R|T) = \int \mathrm{d}H\, \delta(R(H) - R)\, p_{\mathrm{sim}}(H|T).$$

The task of constrained set generation

$$\mathbf{q}_{\theta}(\mathbf{R}\,|\,\mathbf{T}) \sim \mathbf{q}_{\theta_1}(\mathbf{N_R}\,|\,\mathbf{T})\mathbf{q}_{\theta_2}(\mathbf{R}\,|\,\mathbf{N_R},\mathbf{T})$$



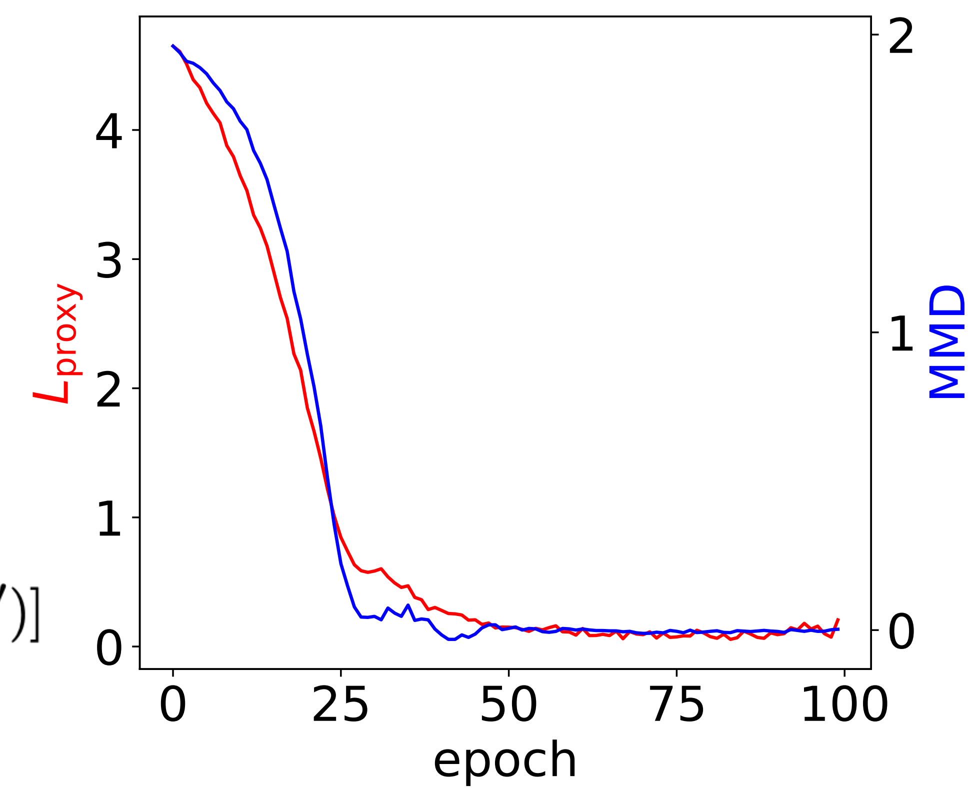
The task of constrained set generation

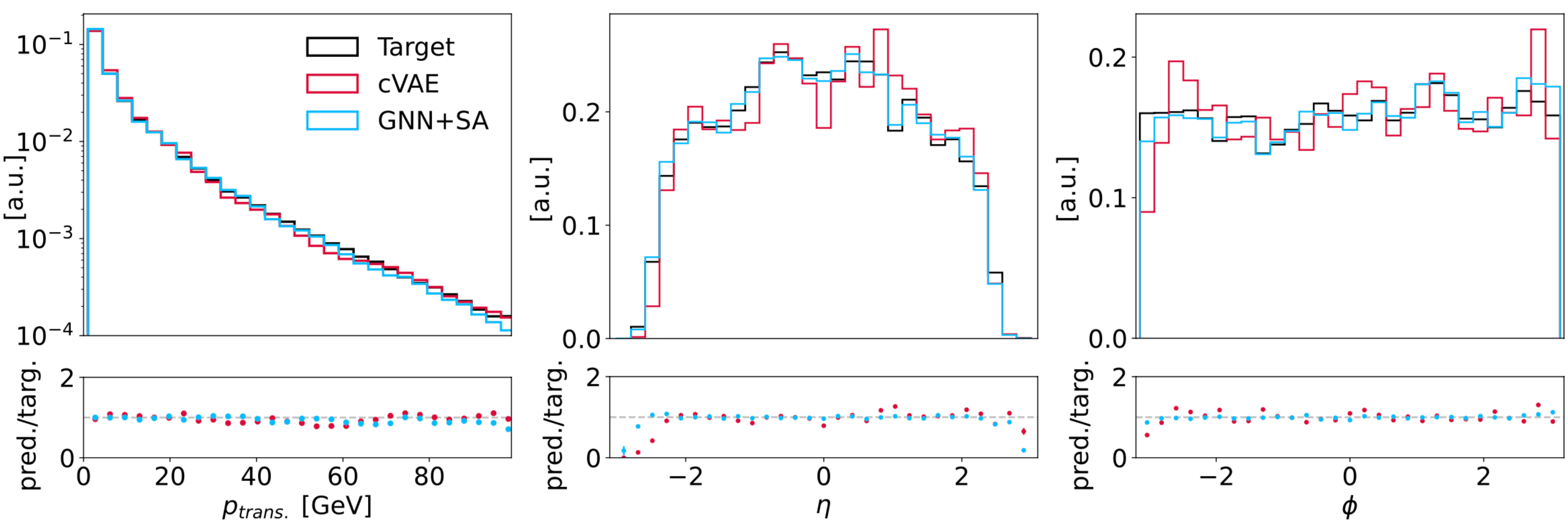
The cVAE training is done by optimizing negative evidence lower bound (ELBO) loss :

$$L = -\mathbb{E}_{T,R}\mathbb{E}_{z\sim q_E(z|R,T)} \log \frac{q_D(R|z,T)q_P(z|T)}{q_E(z|R,T)}$$
$$= -\mathbb{E}_{T,R}\mathbb{E}_z \log q_D(R|z,T) + D_{\mathrm{KL}}(q_E(z|R,T)||q_P(z|T))$$

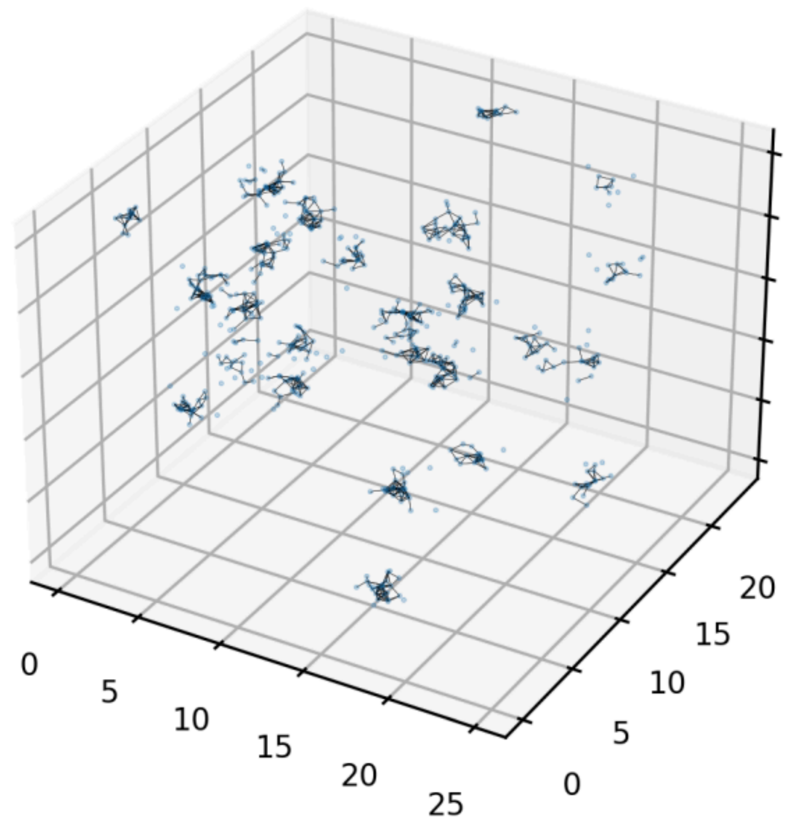
For GNN+SA, we try a regular Hungarian loss and also MMD (maximum mean discrepancy) :

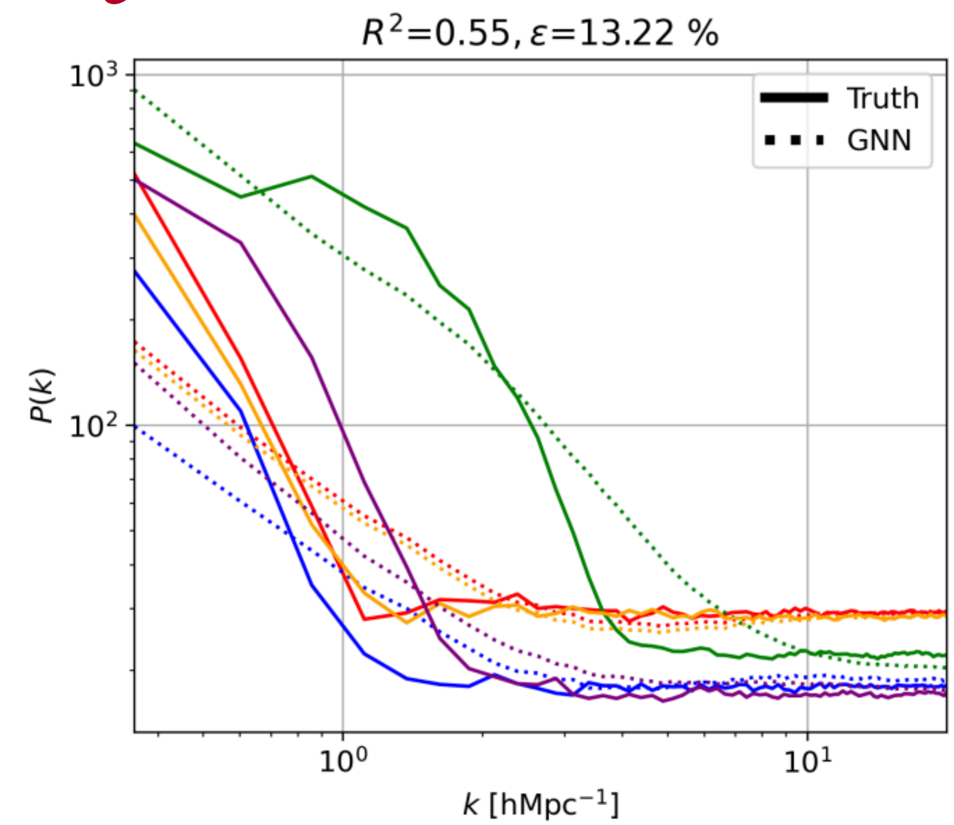
$$\mathrm{MMD}^2 = \mathbb{E}_{(x\sim p,x'\sim p)}[k(x,x')] + \mathbb{E}_{(x\sim q,x'\sim q)}[k(x,x')] - 2\mathbb{E}_{(x\sim q,x'\sim p)}[k(x,x')]$$





GNN in other fields of physics





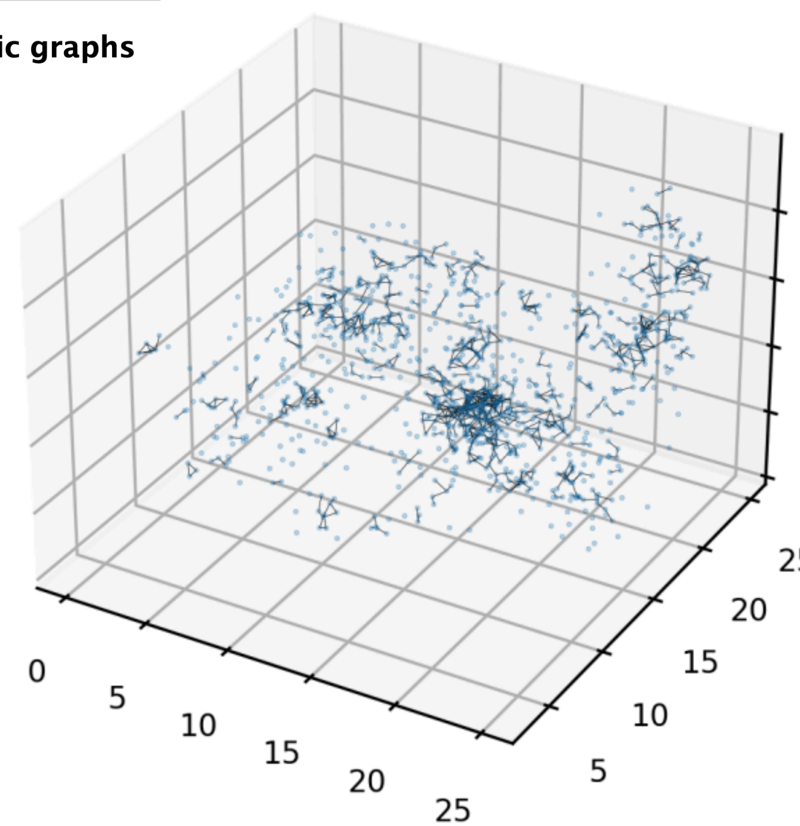
arXiv > astro-ph > arXiv:2204.13713

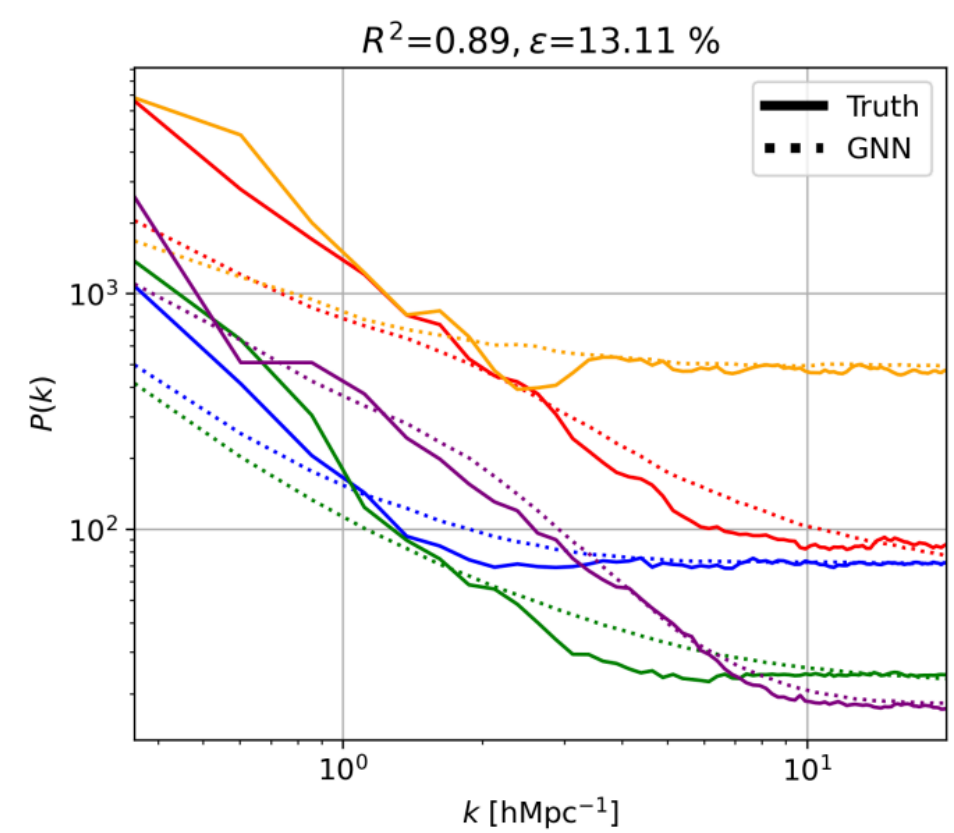
Astrophysics > Cosmology and Nongalactic Astrophysics

[Submitted on 28 Apr 2022 (v1), last revised 8 Feb 2023 (this version, v2)]

Learning cosmology and clustering with cosmic graphs

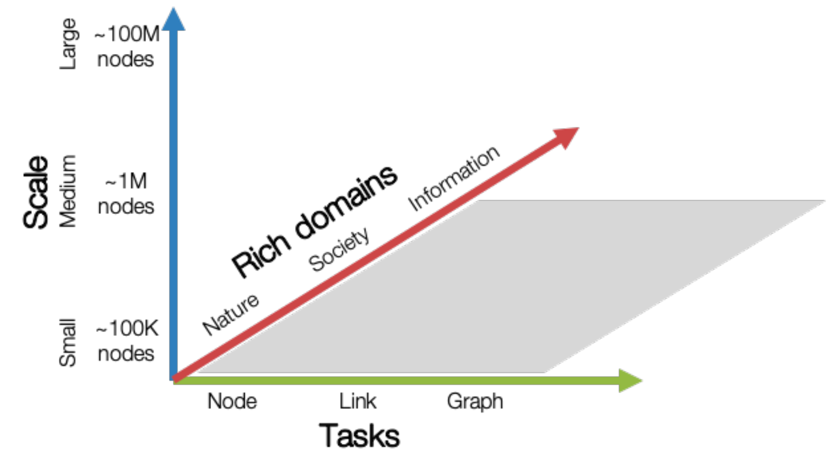
Pablo Villanueva-Domingo, Francisco Villaescusa-Navarro





Open software support

OPEN GRAPH BENCHMARK



PyG
torchgeometricco.slack.com

85K messages and files are hidden.
Messages and files sent more than 90 days ago are available on paid plans. See upgrade options

PyTorch Geometric Tutorial Project

The PyTorch Geometric Tutorial project provides **video tutorials and** **Colab notebooks** for a variety of different methods in PyG:

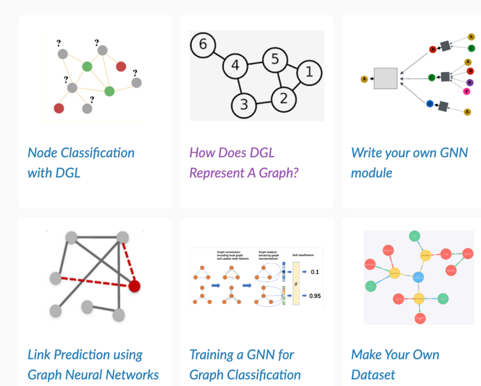
1. Introduction [YouTube, Colab]
2. PyTorch basics [YouTube, Colab]
3. Graph Attention Networks (GATs) [YouTube, Colab]
4. Spectral Graph Convolutional Layers [YouTube, Colab]
5. Aggregation Functions in GNNs [YouTube, Colab]
6. (Variational) Graph Autoencoders (GAE and VGAE) [YouTube, Colab]
7. Adversarially Regularized Graph Autoencoders (ARGA and ARGVA) [YouTube, Colab]
8. Graph Generation [YouTube]
9. Recurrent Graph Neural Networks [YouTube, Colab (Part 1), Colab (Part 2)]
10. DeepWalk and Node2Vec [YouTube (Theory), YouTube (Practice), Colab]
11. Edge analysis [YouTube, Colab (Link Prediction), Colab (Label Prediction)]
12. Data handling in PyG (Part 1) [YouTube, Colab]
13. Data handling in PyG (Part 2) [YouTube, Colab]
14. MetaPath2vec [YouTube, Colab]
15. Graph pooling (DiffPool) [YouTube, Colab]

DEEP GRAPH LIBRARY

DGL
deep-graph-library.slack.com

68K messages and files are hidden.
Messages and files sent more than 90 days ago are available on paid plans. See upgrade options

A Blitz Introduction to DGL



Much more advanced social communities