How the friendship between HEP and Cosmology can be made deeper over a cup of ML?

Frontiers of Particle Physics 2024, CHEP

Sanmay Ganguly
Indian Institute of Technology, Kanpur
sanmay@iitk.ac.in

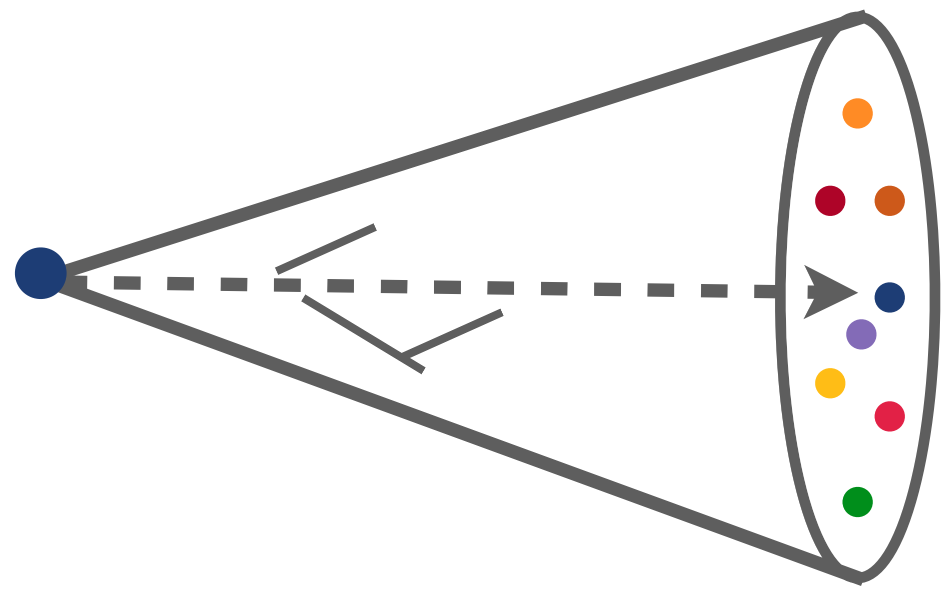
10/08/2024

ML@Natural-Science : what's the broad task?

2. Choose a NN model (CNN/GNN/)

1. Decide the right representation of the data (images/graphs/trees..)

3. With a defined learning task, compute the loss function.

$f_{\{\theta\}}\,($  $) \rightarrow \hat{y} = f_{\{\theta\}}\,(X)$

$$L(y,\hat{y}) \equiv L(\{\theta\})$$

Self-supervised

Variation in data

Unsupervised	Semi-supervised	Weakly-supervised	Supervised
No-labels, the task is to figure out p(x) from which the data is drawn. e.g. VAE	Noisy labels. estimate : p(s-enriched)/p(s-depleted)	Partial labels. e.g. simulating : SM bkg vs many NP signals.	Learning on all the well labeled data.

Different physics : same ML problems

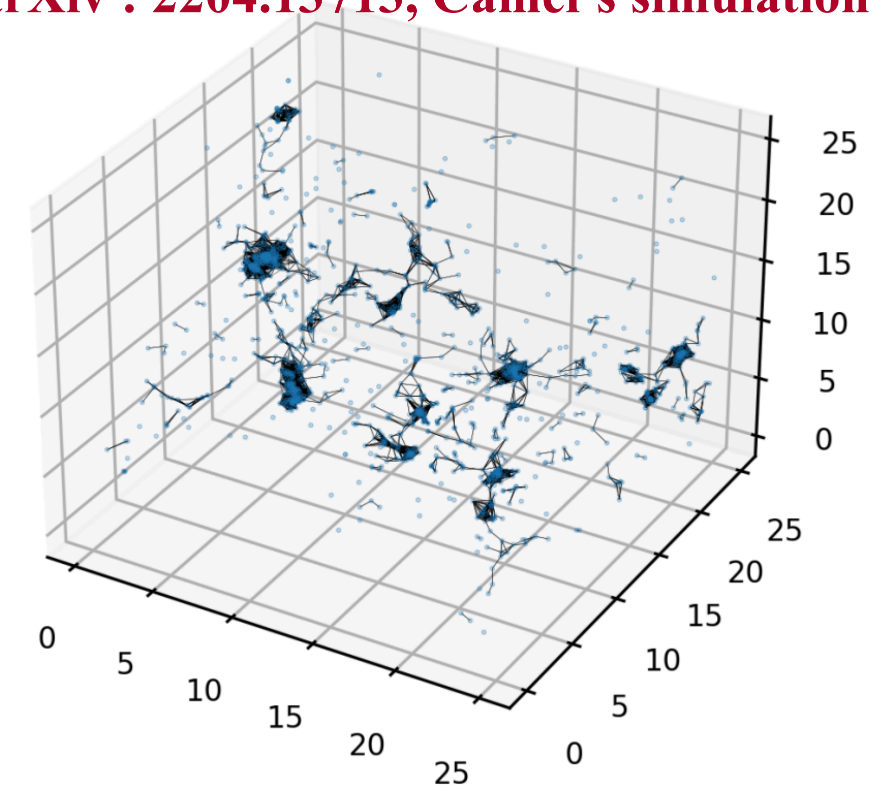
2. Choose a NN model (CNN/GNN/)

1. Decide the right representation of the data (images/graphs/trees..)

3. With a defined learning task, compute the loss function.

arXiv : 2204.13713, Camel's simulation

$$f_{\{\theta\}} \Big(\quad \Big) \rightarrow \hat{y} = f_{\{\theta\}}(X)$$



$$L(y, \hat{y}) \equiv L(\{\theta\})$$

Variation in data

Self-supervised

Unsupervised

No-labels, the task is to figure out p(x) from which the data is drawn. e.g. VAE

Semi-supervised

Noisy labels. estimate :

p(s-enriched)/p(s-depleted)

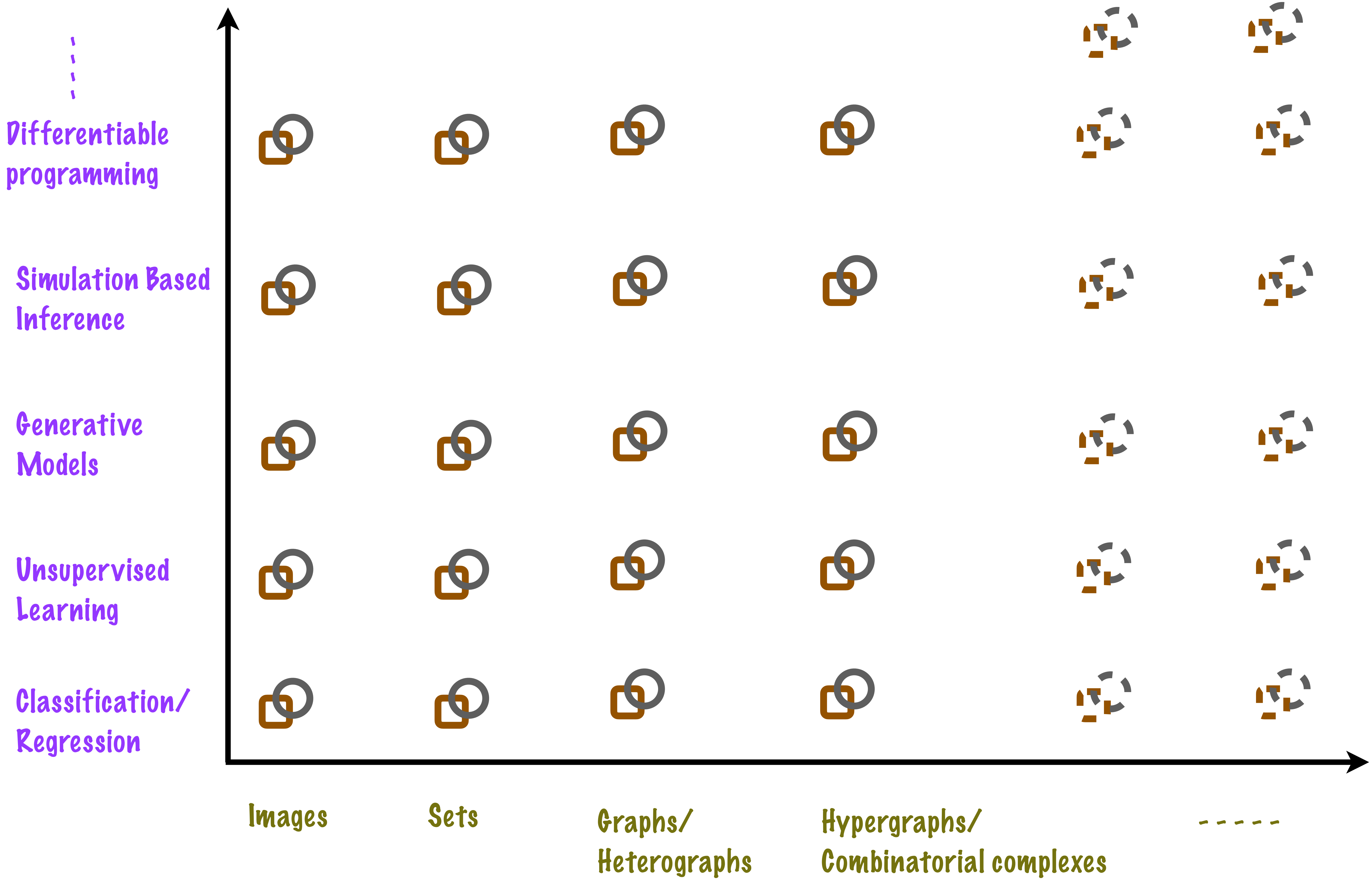
Weakly-supervised

Partial labels. e.g. simulating : SM bkg vs many NP signals.

Supervised

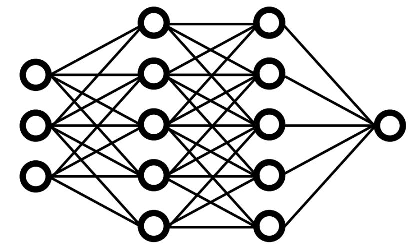
Learning on all the well labeled data.

This talk is about ……

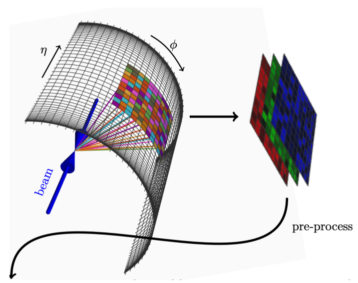


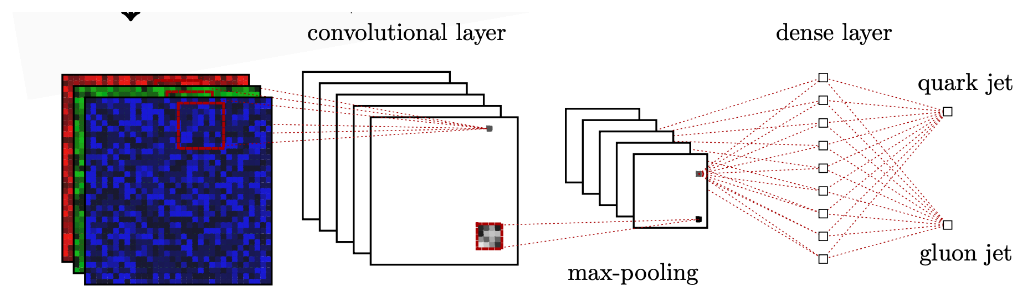
Data representation ⇔ NN correspondence

$$J = \{p_1^\mu, p_2^\mu, \dots\}$$

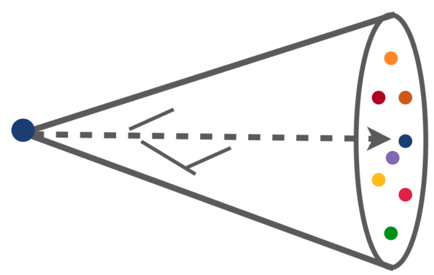


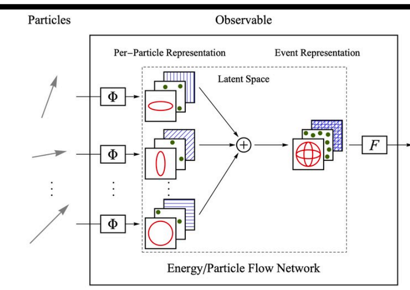
Ordered set
DNN



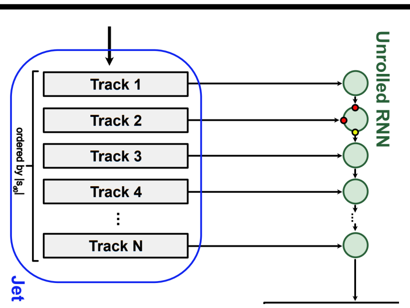


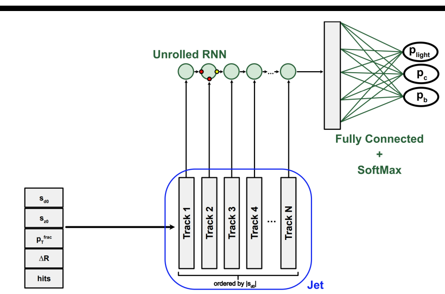
Grid
CNN



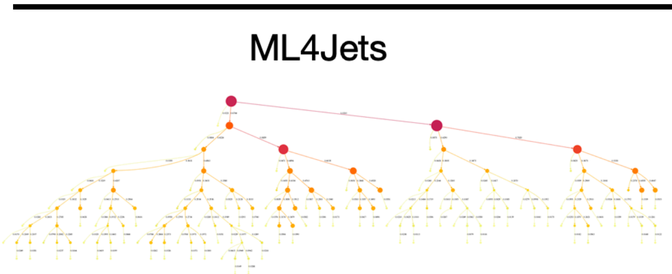


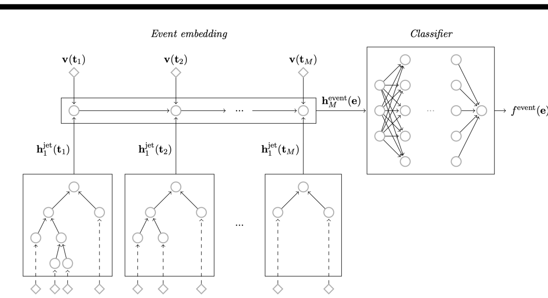
Unordered set
Deepest



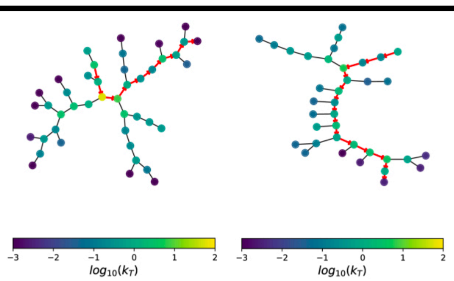


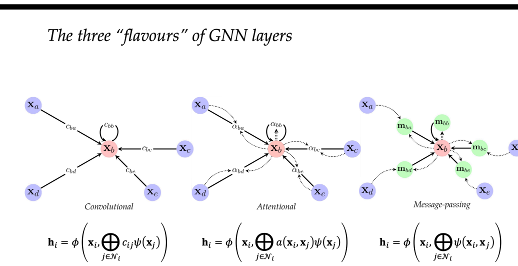
Sequential data
RNN





Tree structure
Deepset/GNN

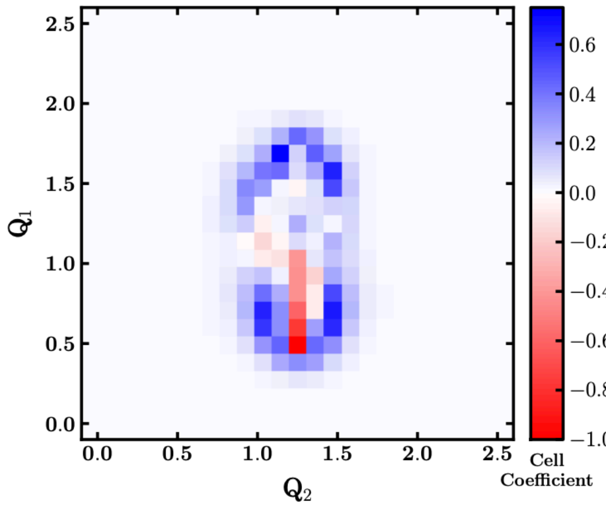




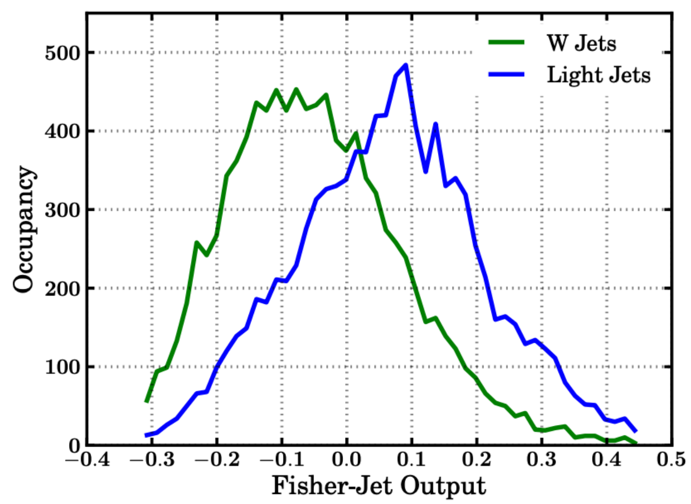
Graph
GNN

Early jet tagging

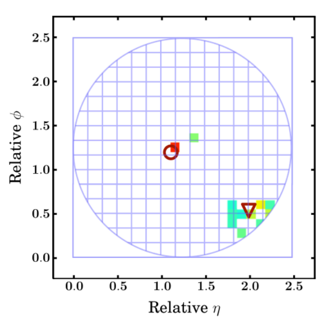
J. Cogan et-al JHEP 02 (2015) 118



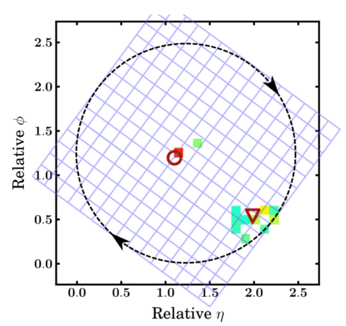
(a) Fisher-Jet



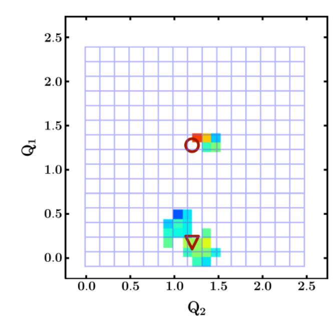
(b) Fisher-Jet Discriminant Output



(a) Jet-image prior to rotation

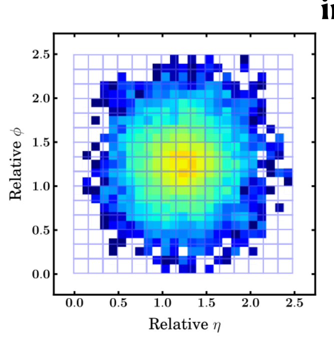


(b) Rotated pixel grid

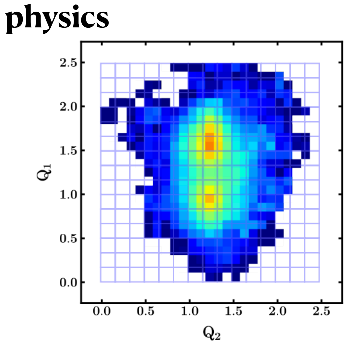


(c) Jet-image after projectio onto rotated grid, before transla tion

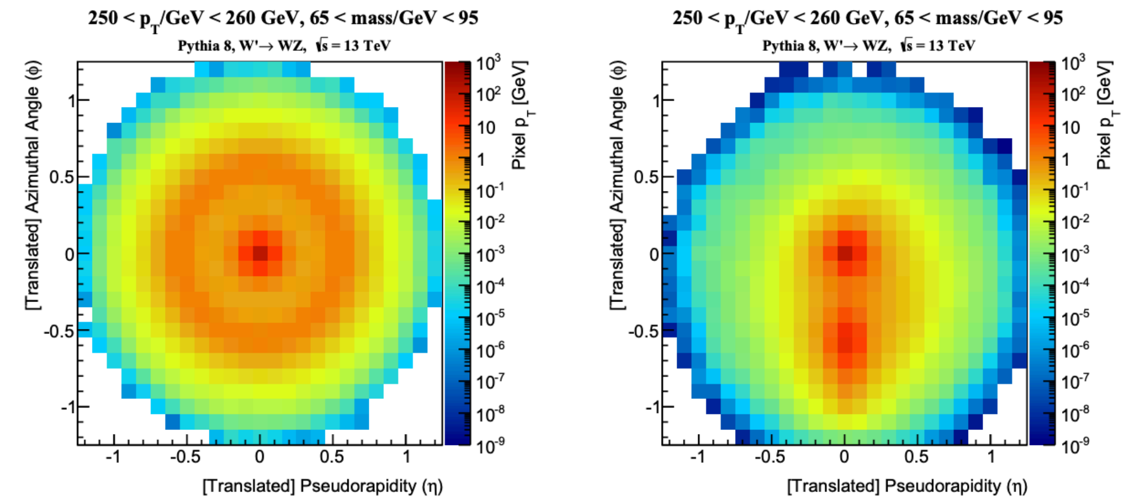
The first paper to discuss image pre-processing for jet physics



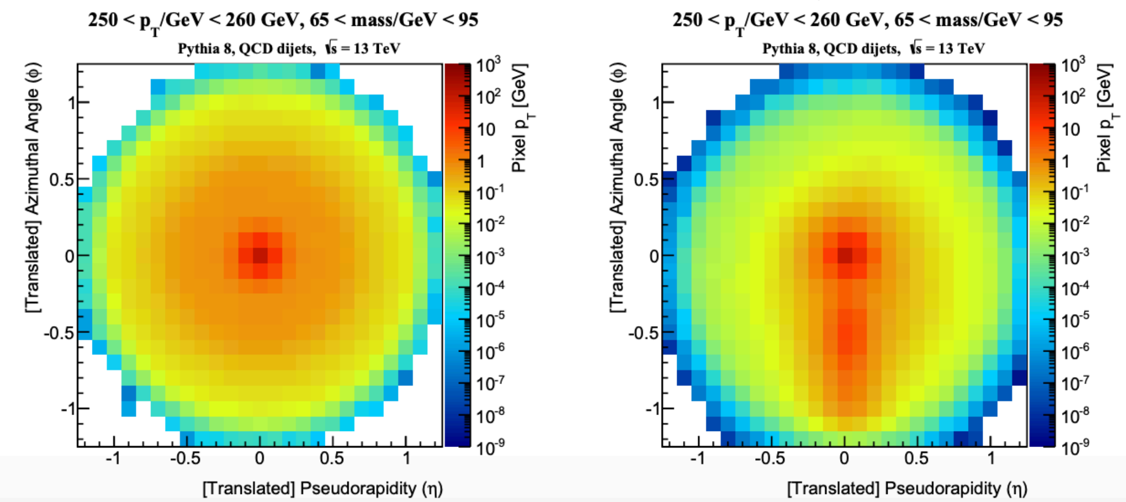
(d) Average jet-image, prior to rotation

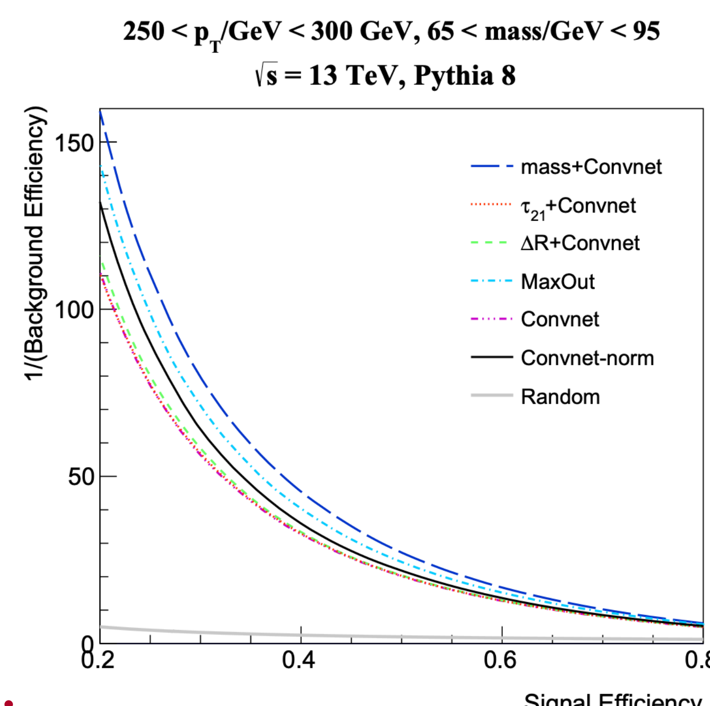


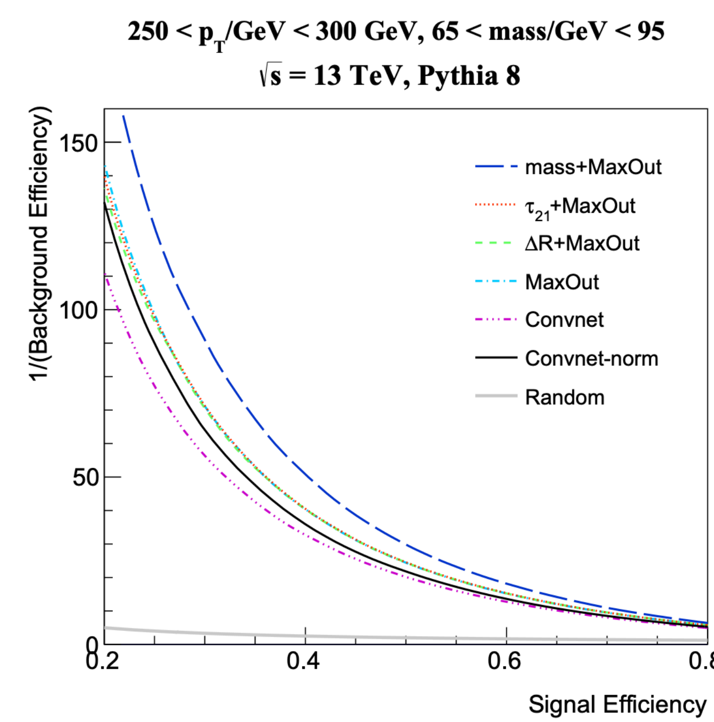
(e) Average jet-image, after pre processing



L. De Oliveira et-al JHEP 07 (2016) 069

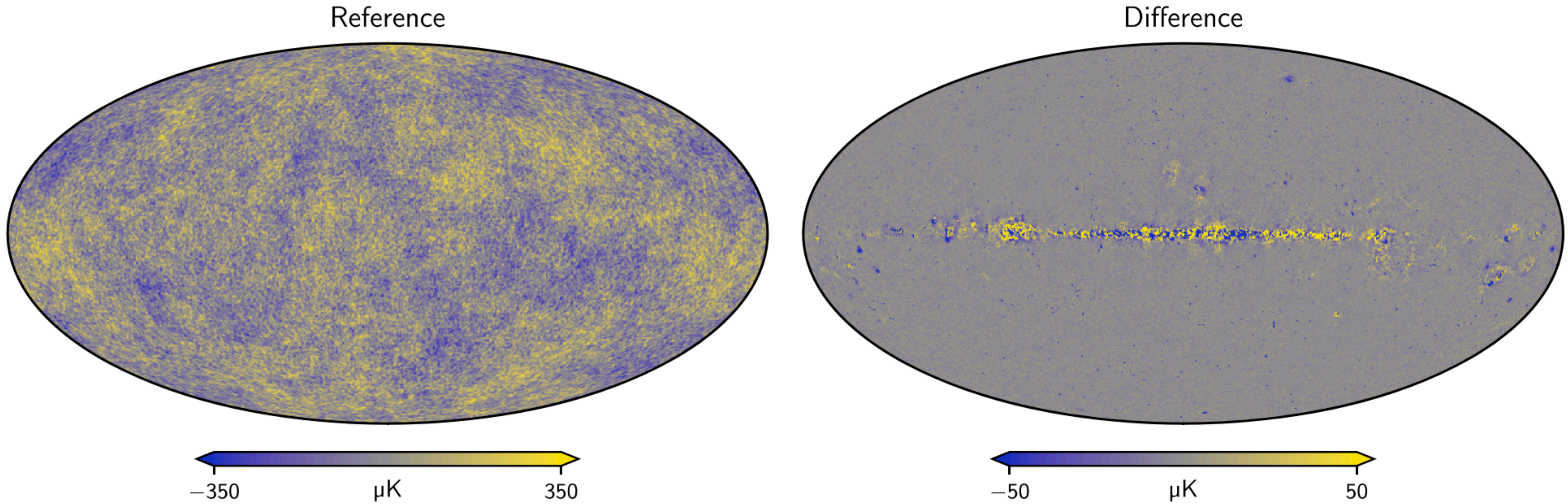




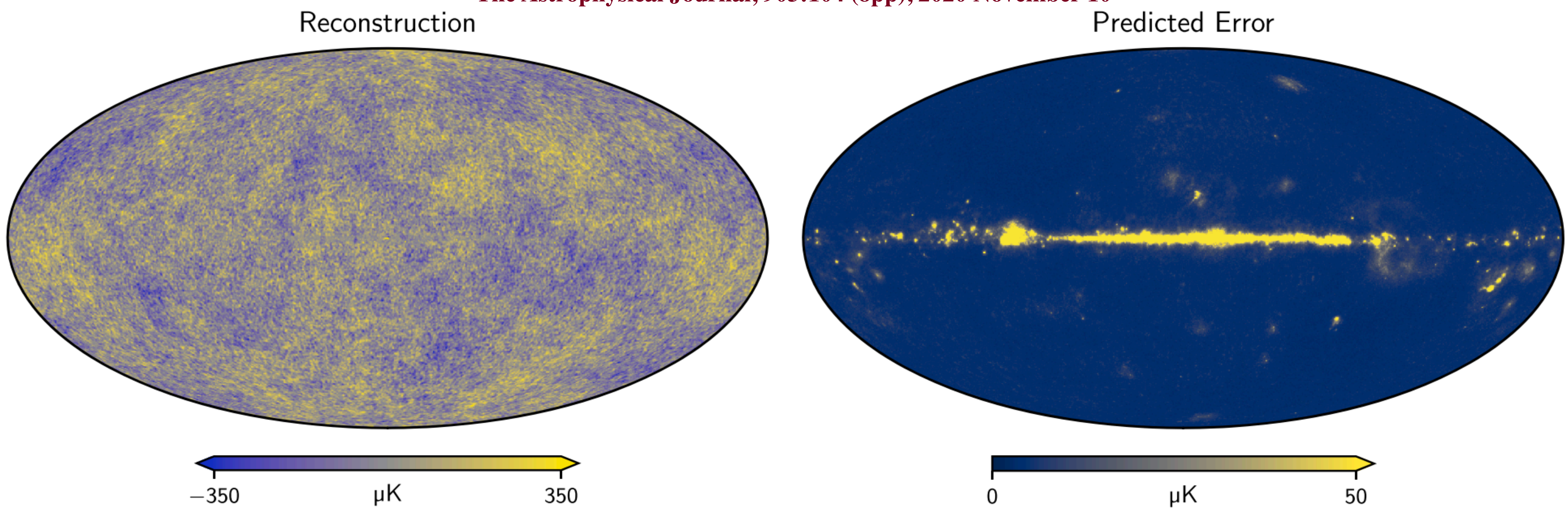


Similar methods were applied for particle identifications

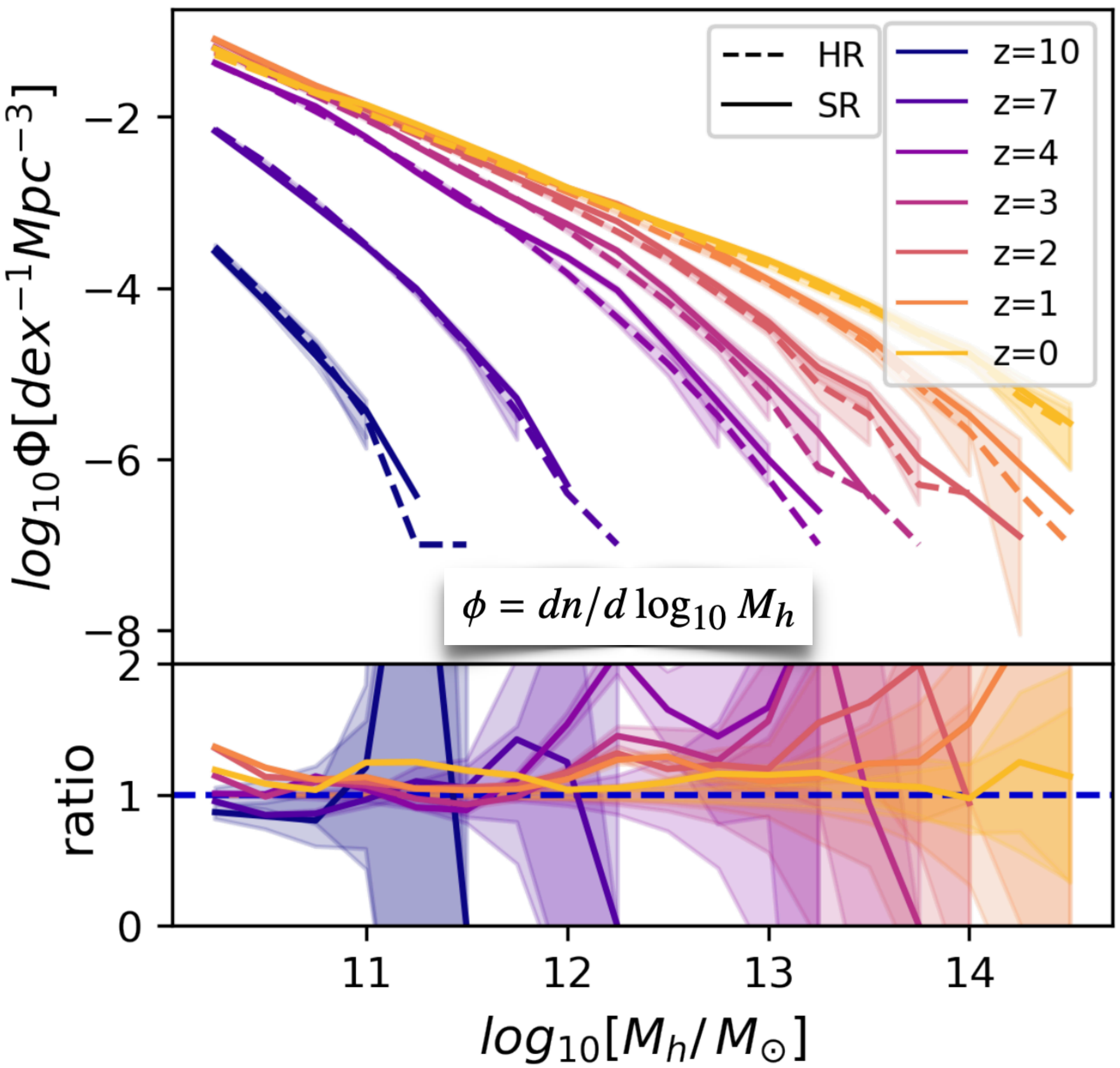
CMB spectrum cleaning using NN

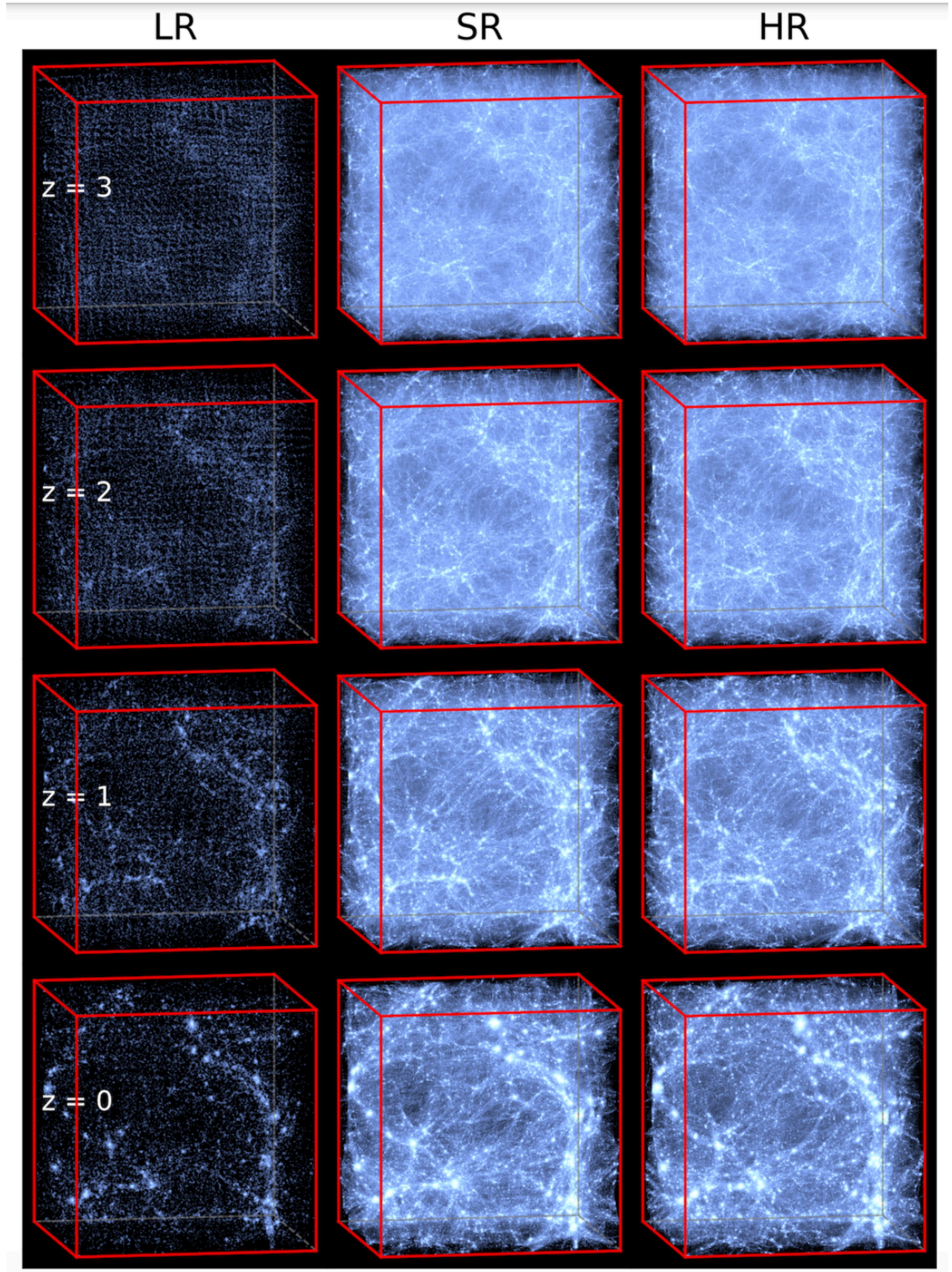


The Astrophysical Journal, 903:104 (8pp), 2020 November 10



Super-resolution



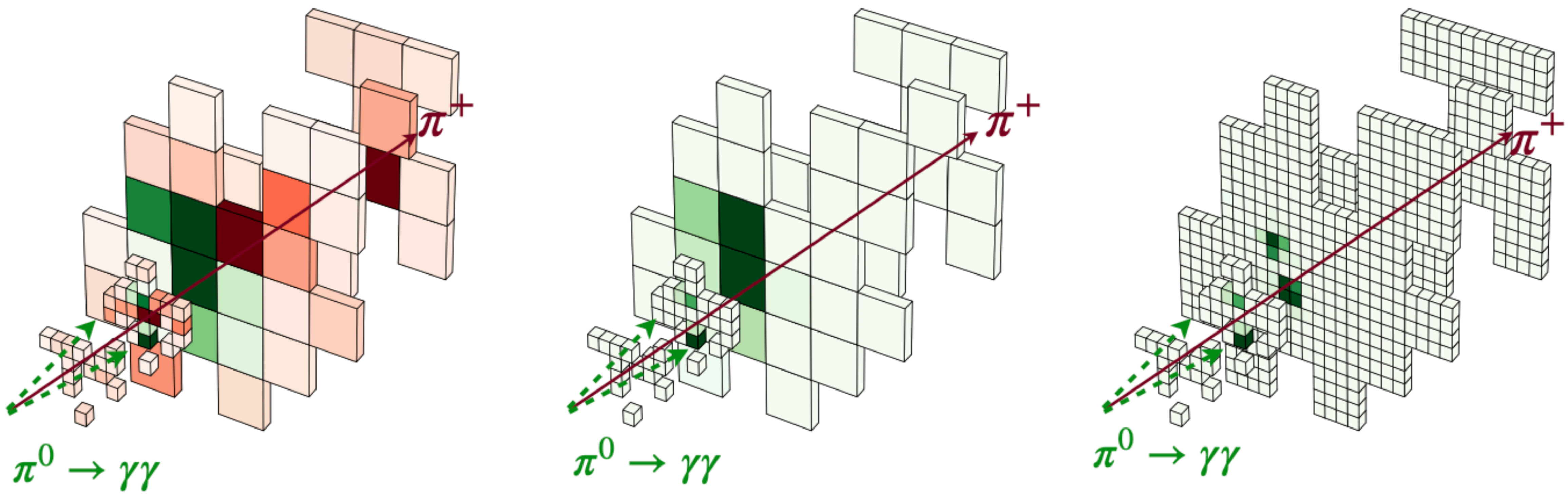


MNRAS 000, 000–000 (2022)

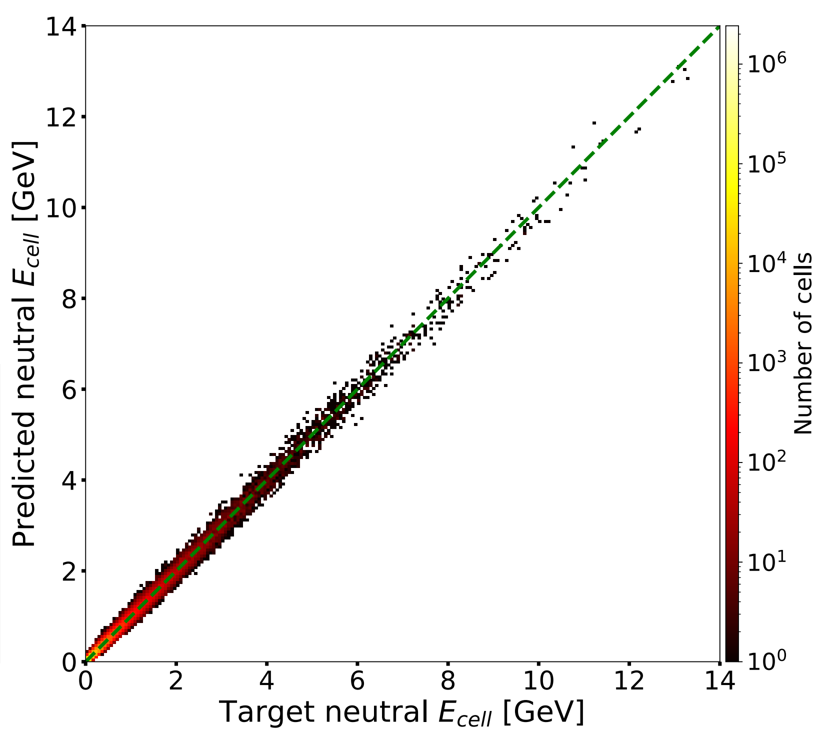
SR model is capable of generating merger histories that are solely dependent on on time-consistent LR input

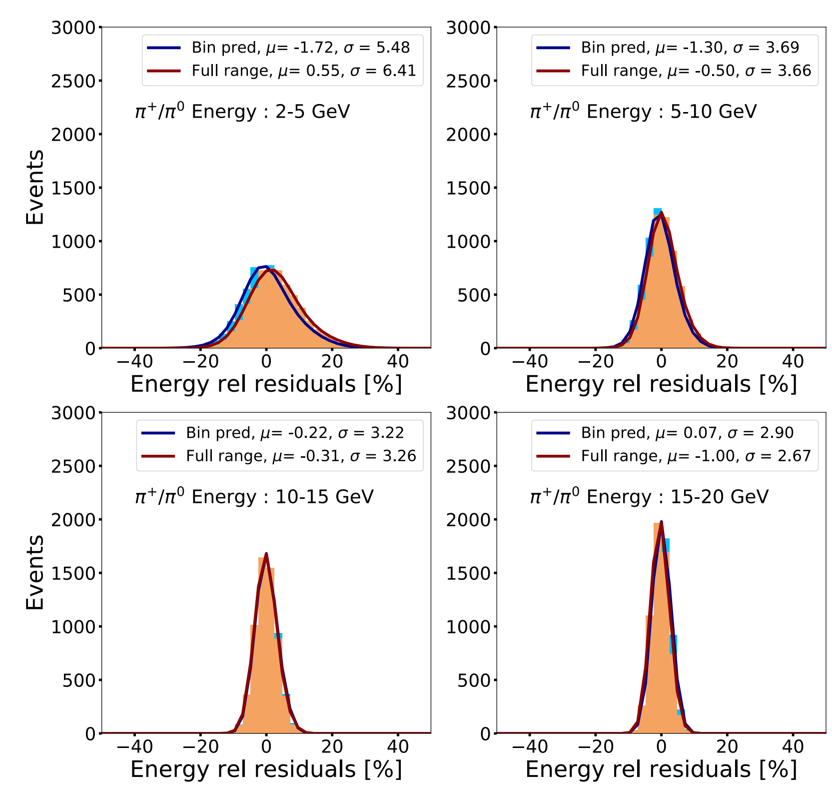
A super-resolution case for HEP

8 X 8 Low Res detector | 32 X 32 High Res detector



The networks in general have good noise removal abilities.





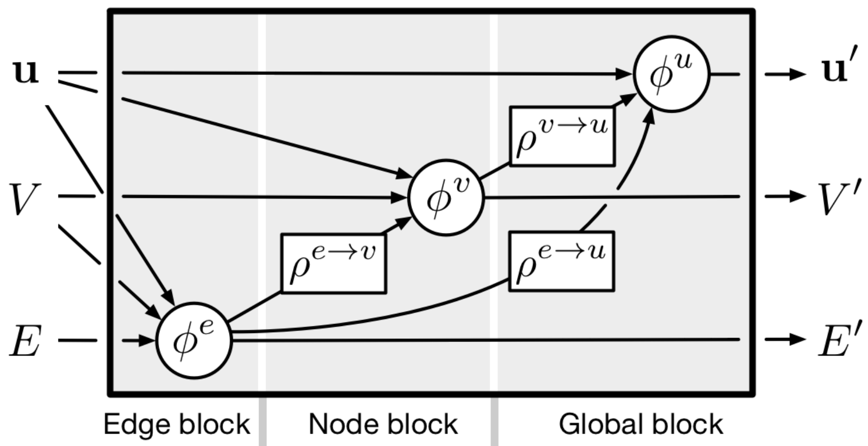
Towards a computer vision particle flow

Francesco Armando Di Bello, Sanmay Ganguly, Eilam Gross, Marumi Kado, Michael Pitt, Lorenzo Santi & Jonathan Shlomi

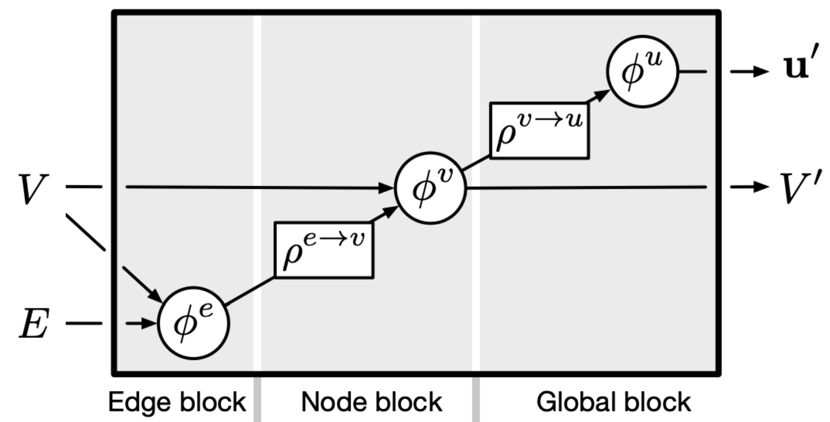
The European Physical Journal C **81**, Article number: 107 (2021) | Cite this article

The general GNN

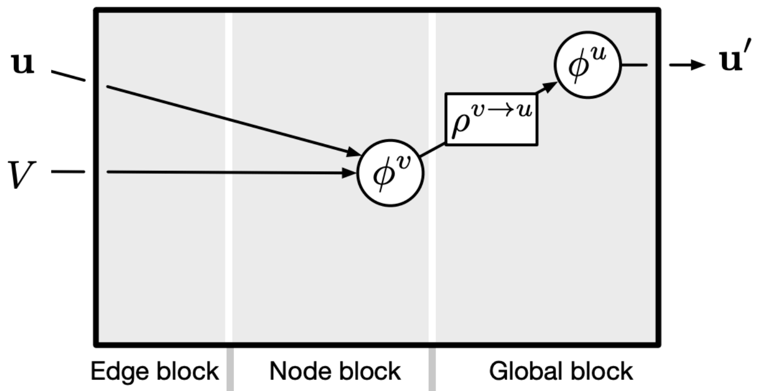
arXiv : 1806.01261



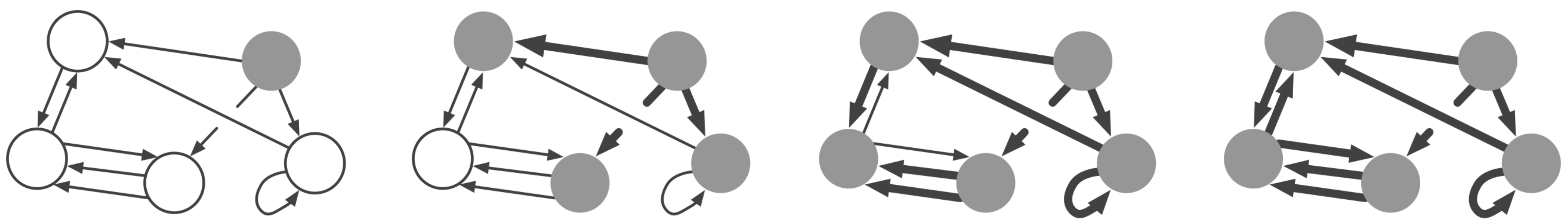
Full GN block



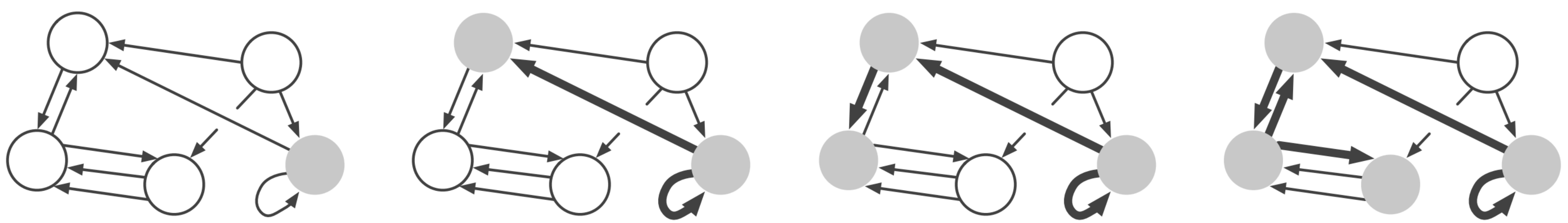
MPNN Layer



Deep-set layer



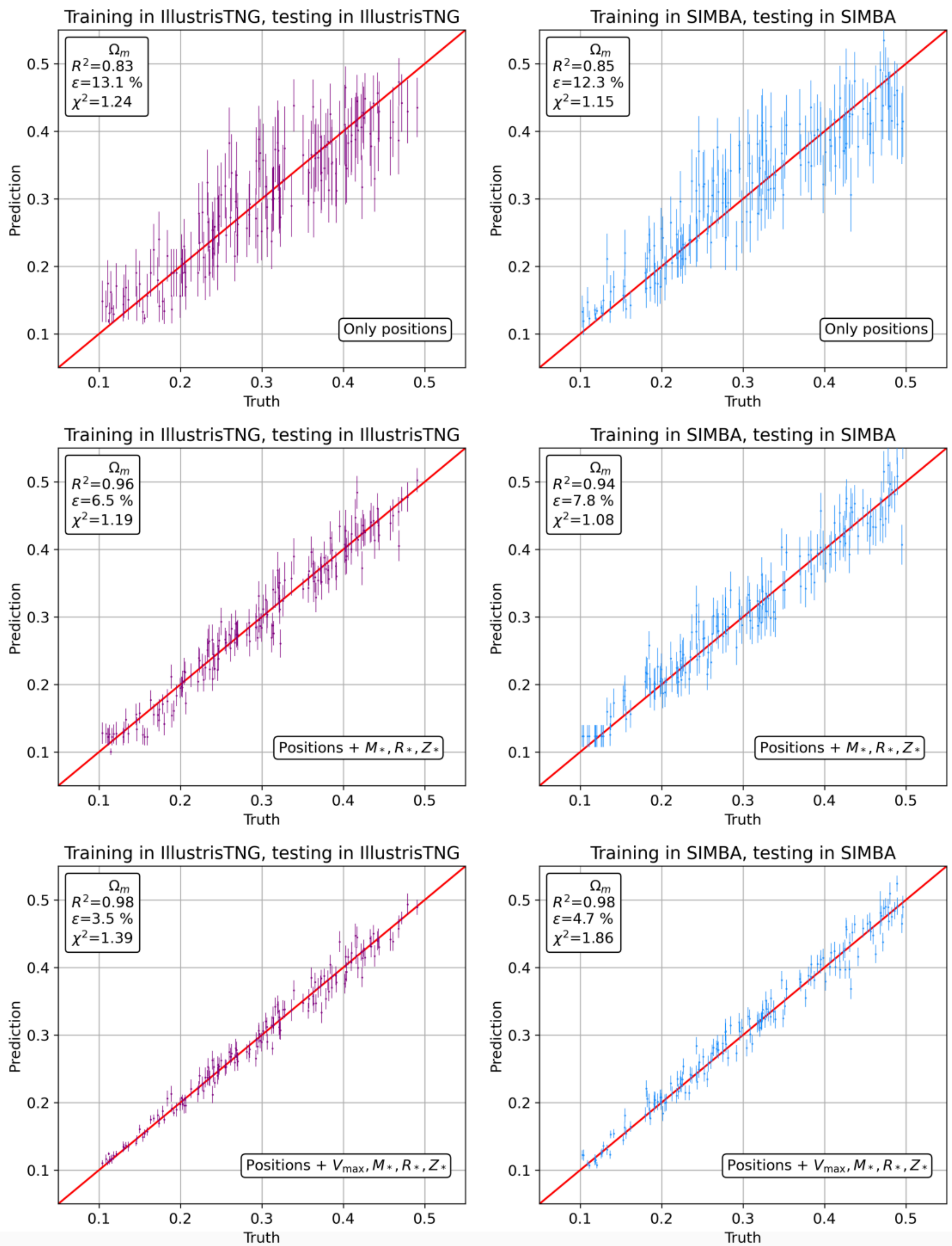
$m = 0$ $m = 1$ $m = 2$ $m = 3$

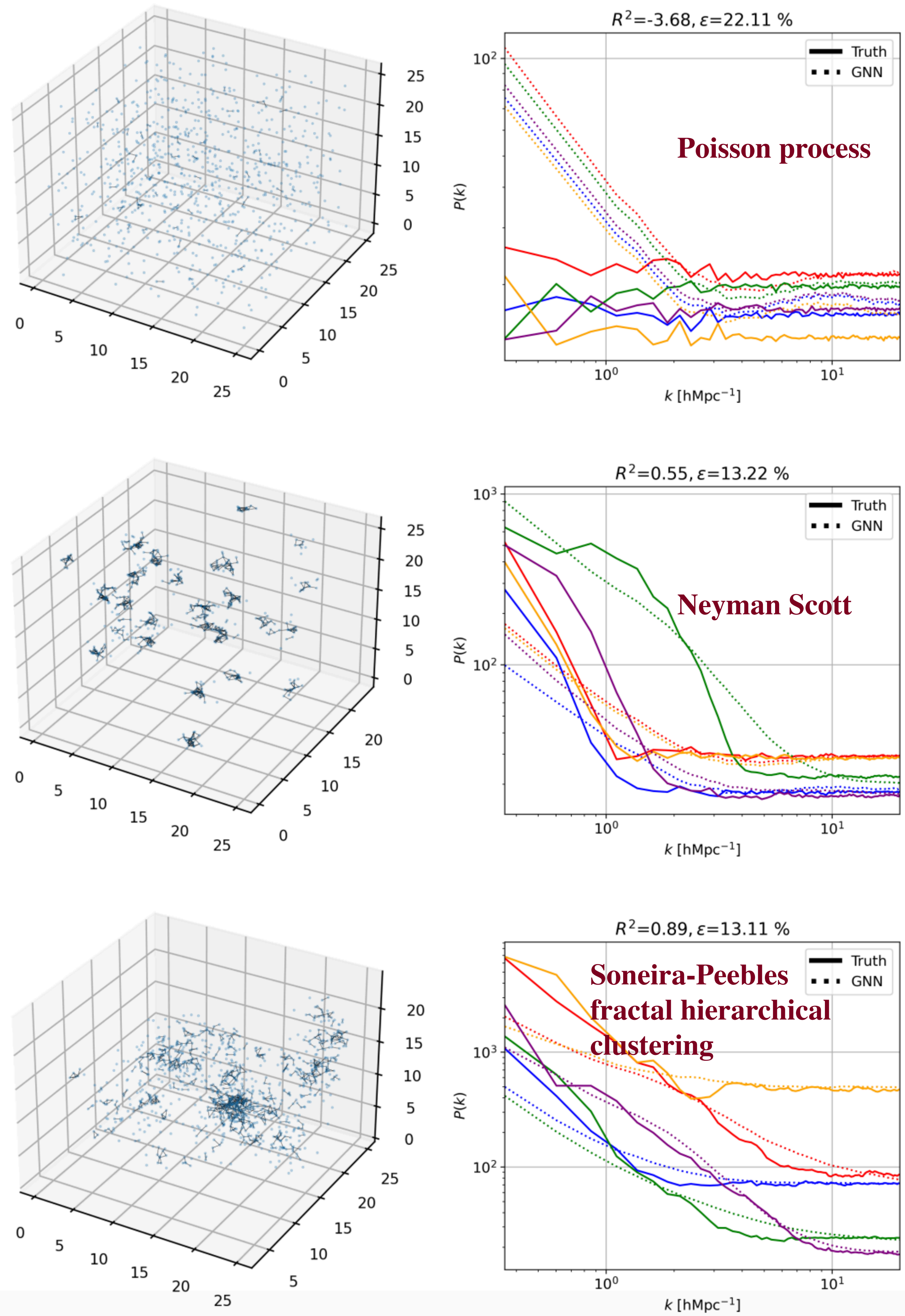


$m = 0$ $m = 1$ $m = 2$ $m = 3$

GNN in Cosmology

arXiv : 2204.13713

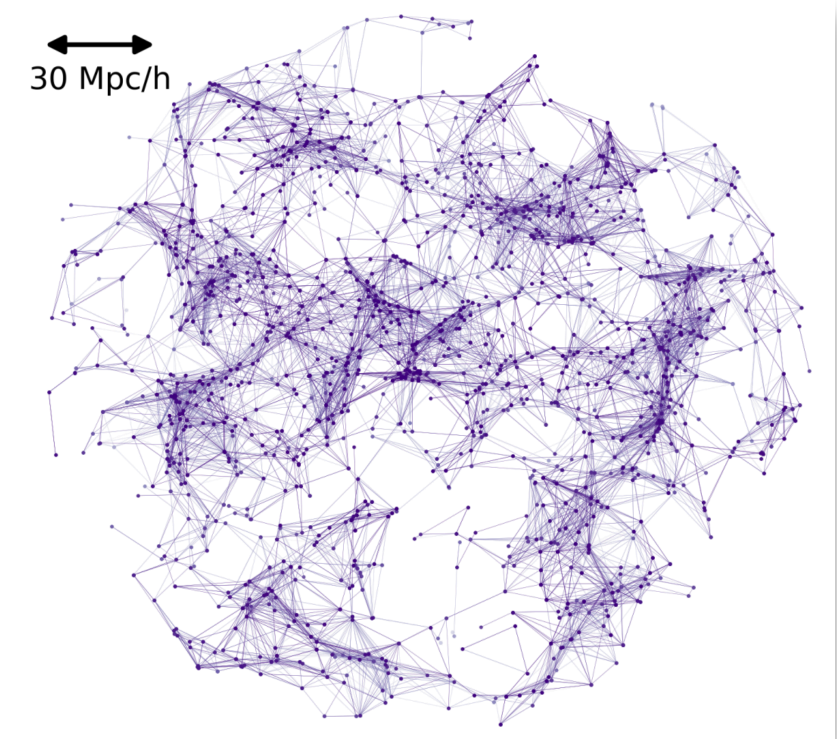




GNN in Cosmology : LOS velocity

Type	Feature	Symbol
Node	3D position	p
Node	Stellar mass	m_*
Node	Star formation rate	m_{SFR}
Global	Number of galaxies in a graph	N_g

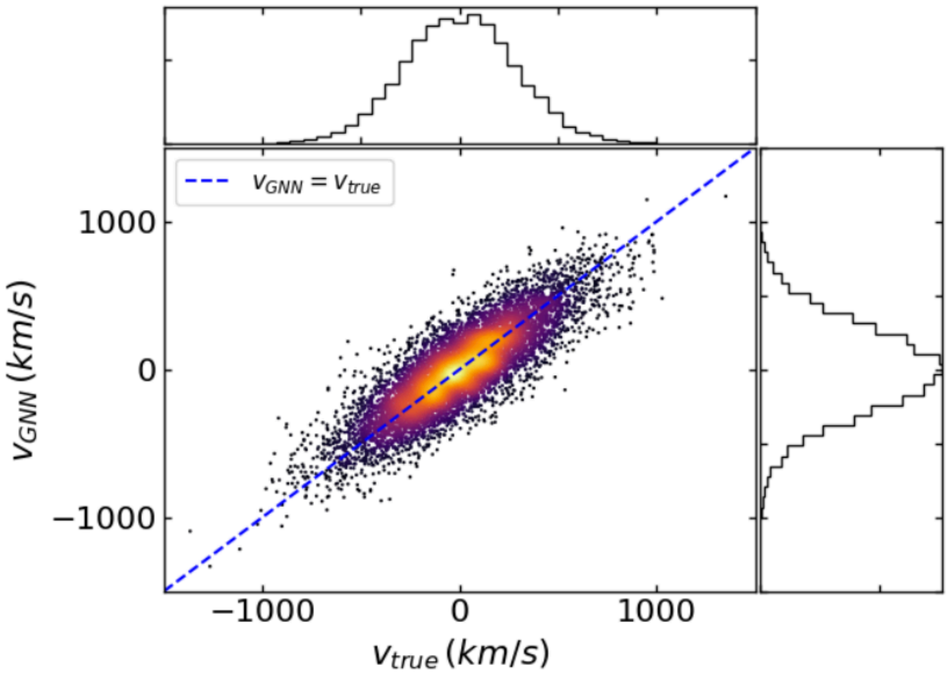
$$\frac{\Delta T_{\mathrm{kSZ}}}{T_{\mathrm{CMB}}} = -\sigma_{\mathrm{T}} \int n_{\mathrm{e}} \left(\frac{\boldsymbol{v} \cdot \hat{\boldsymbol{n}}}{c}\right) \mathrm{d}l \approx -\tau \left(\frac{\boldsymbol{v} \cdot \hat{\boldsymbol{n}}}{c}\right)$$

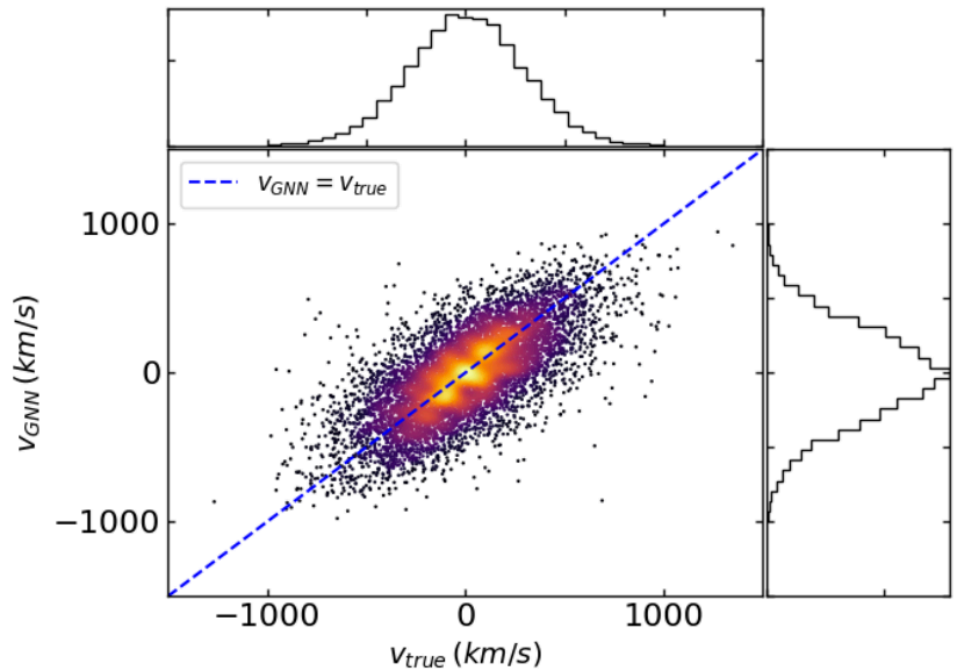


A 6% improvement was achieved by including galaxy masses.

GNN can overcome the shortcoming of explicit galaxy bias.

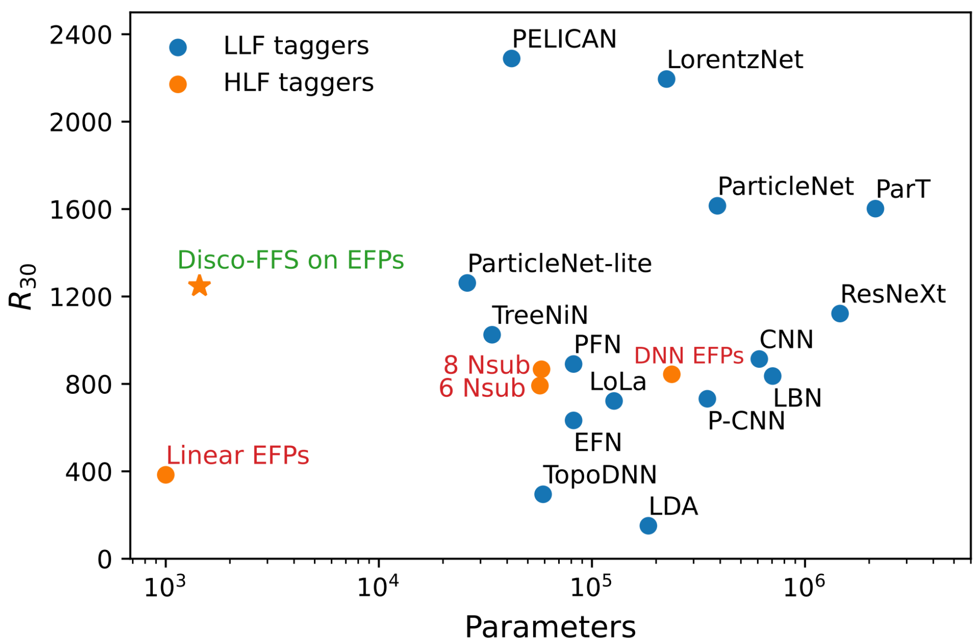
(Required by perturbation theory based methods)

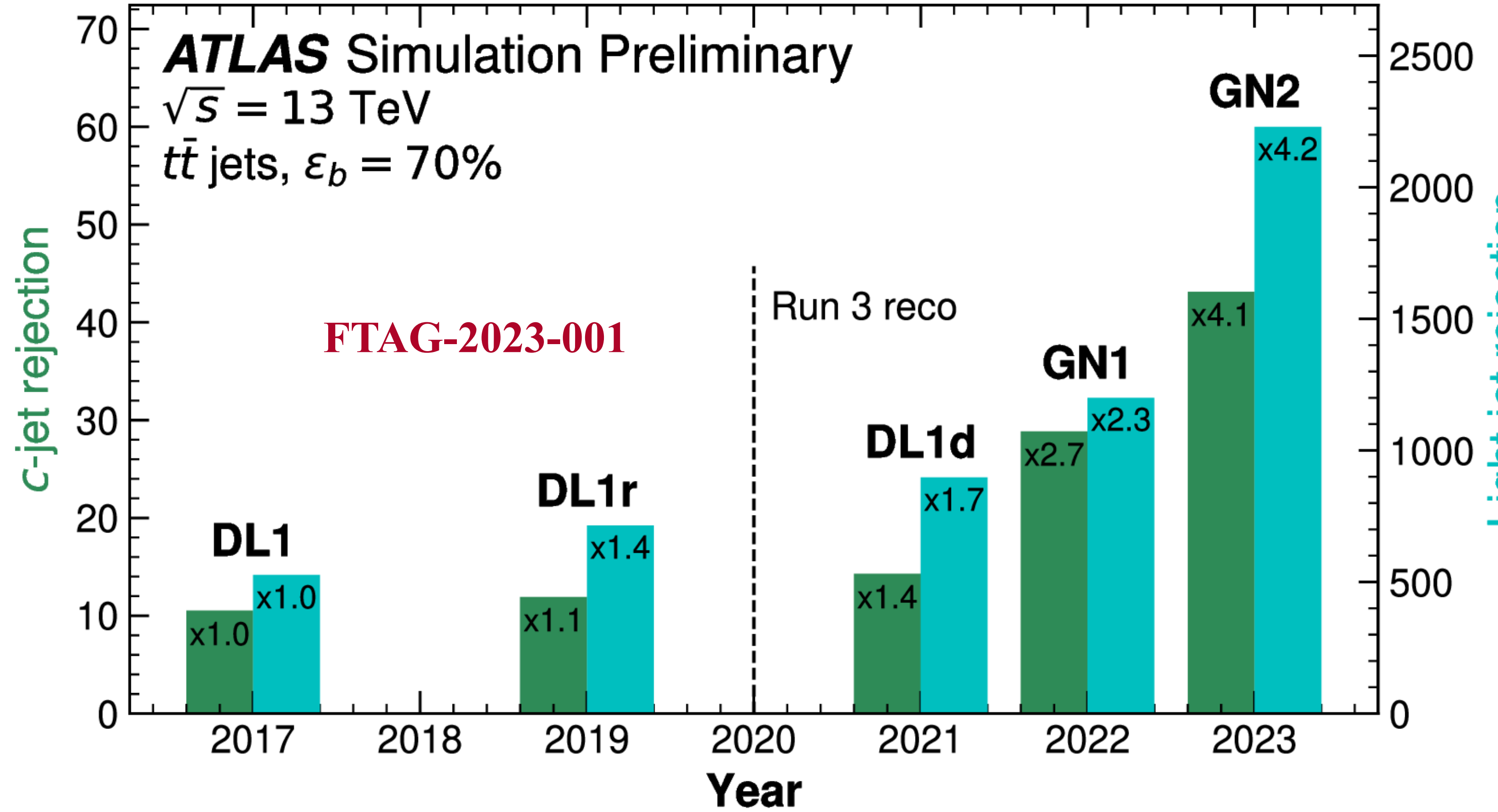


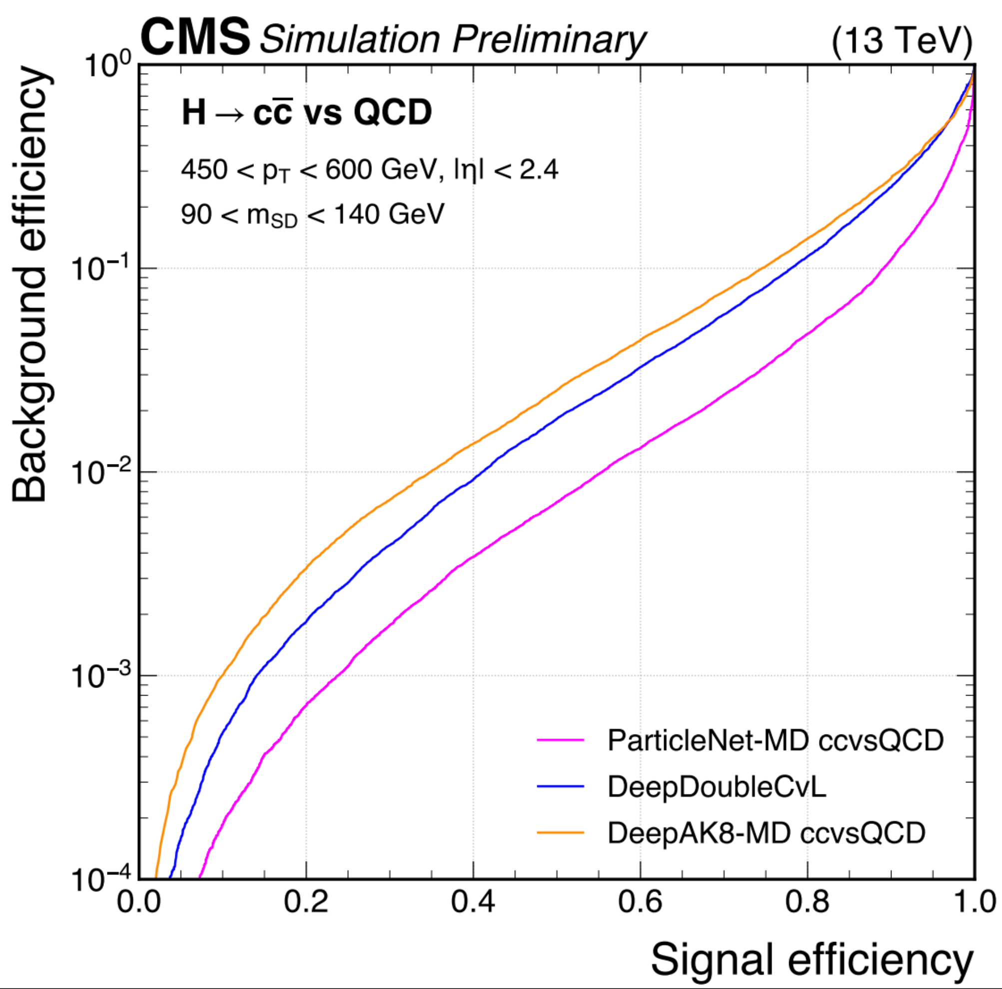


arXiv : 2402.1239

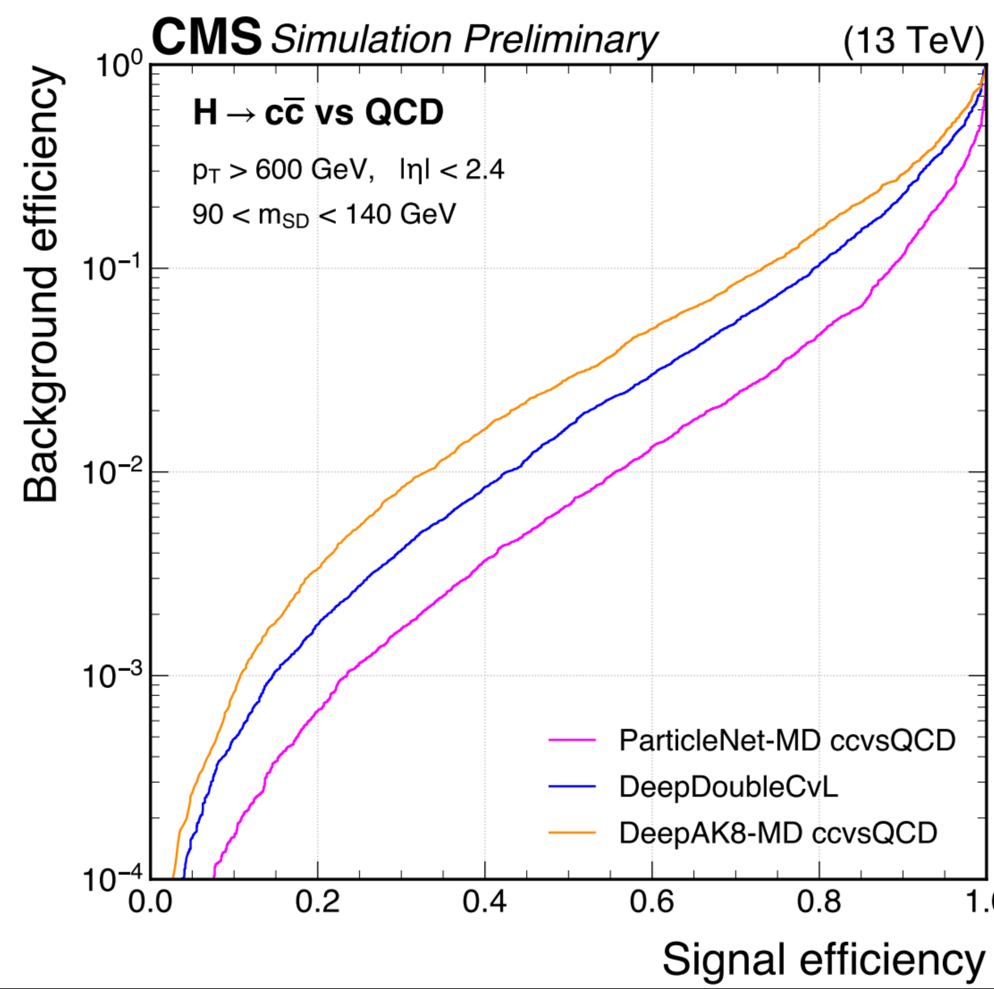
GNN in HEP





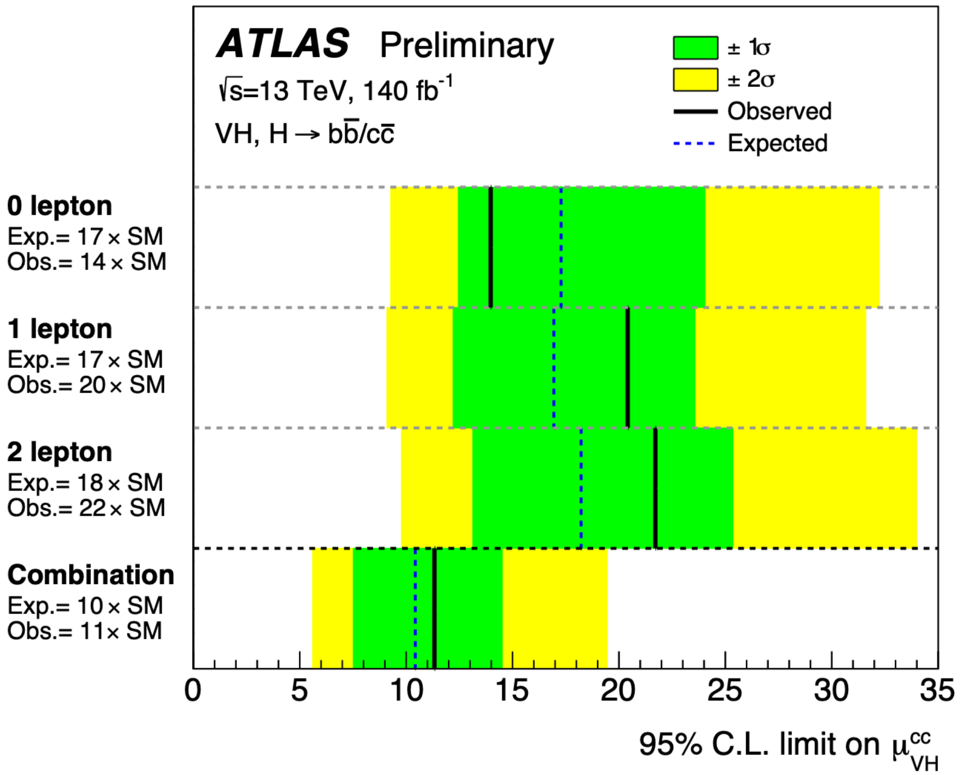


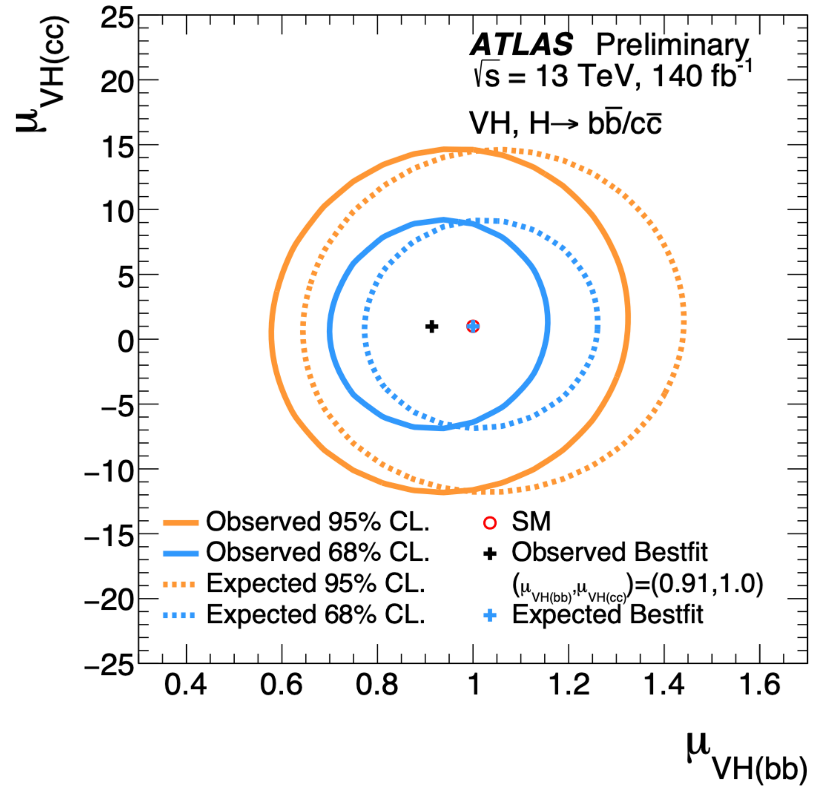
CERN-CMS-BTV-22-001



Jet tagging is revolutionized by GNN

Direct physics application of the taggers





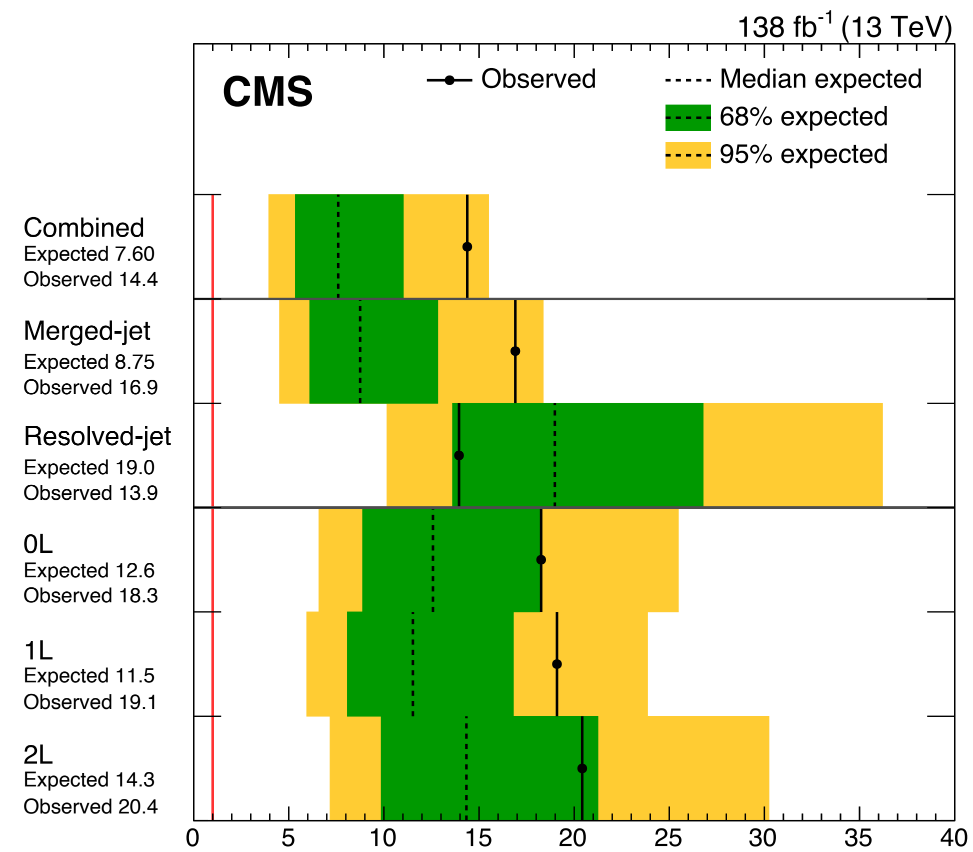
Future iteration will use GN2

ATLAS bound : $|\kappa_c| < 4.2$

CMS bound : $1.1 < |\kappa_c| < 5.5$

Future direction of tagger improvement:

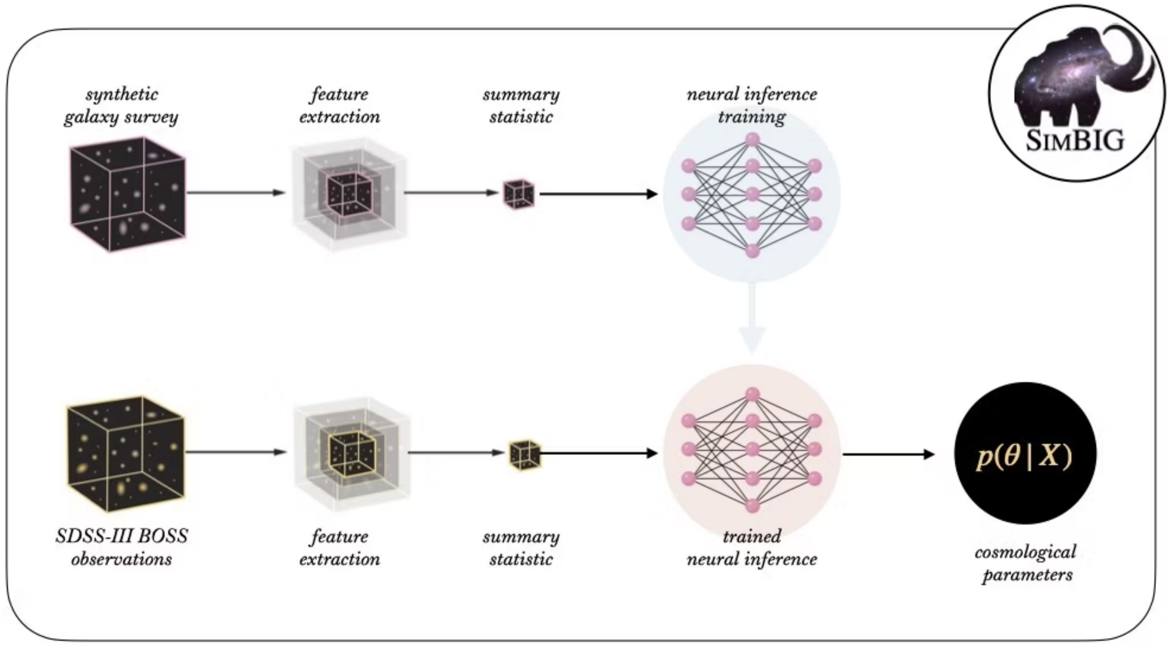
1. **Explainable taggers on heterogeneous pc**
2. **A systematic uncertainty extraction.**
3. **How much universal taggers can be made across topologies?**

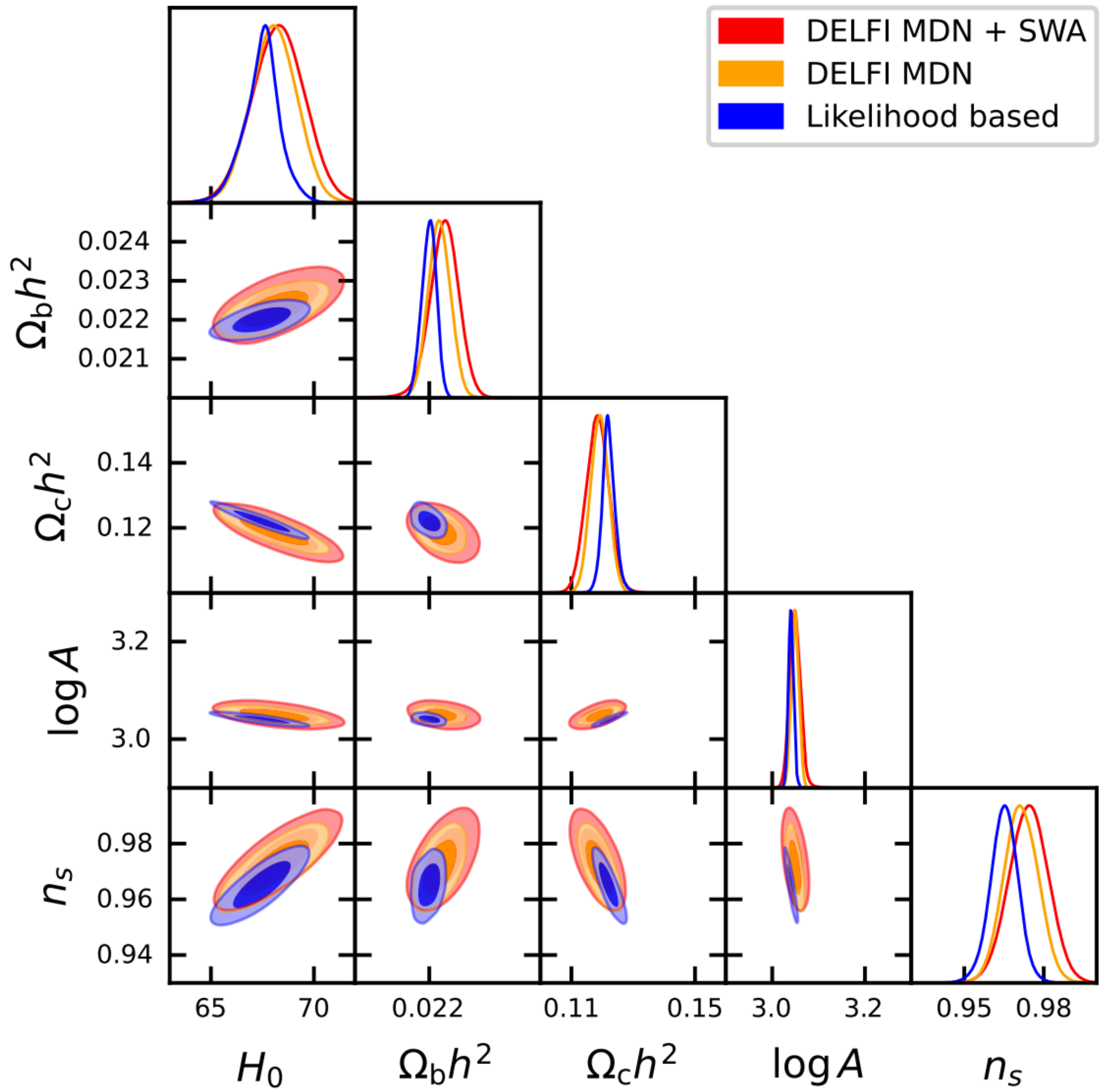


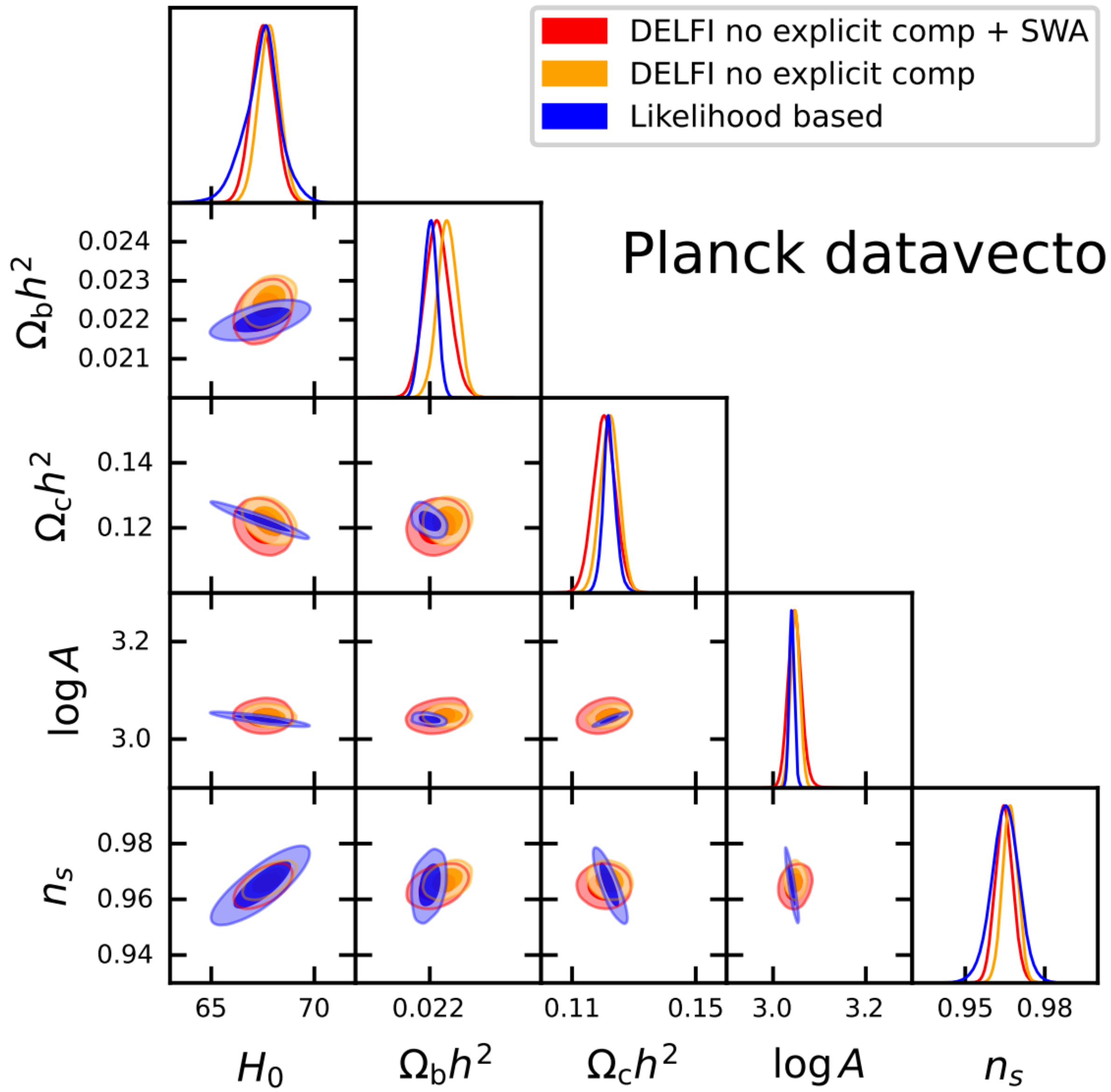
Phys. Rev. Lett. 131 (2023) 061801

SBI in Cosmology

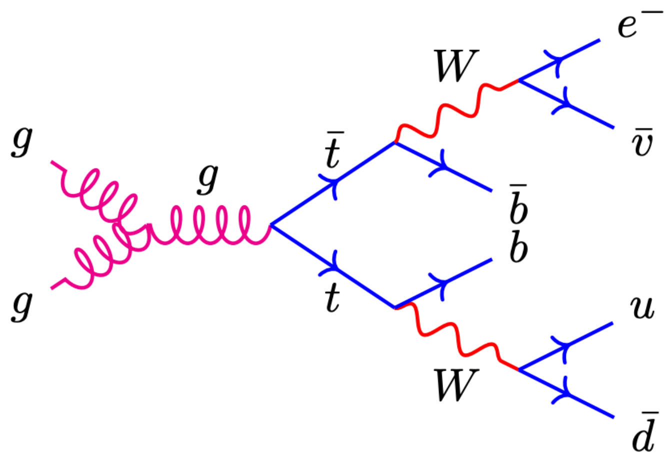
Likelihood free inference is state of the art



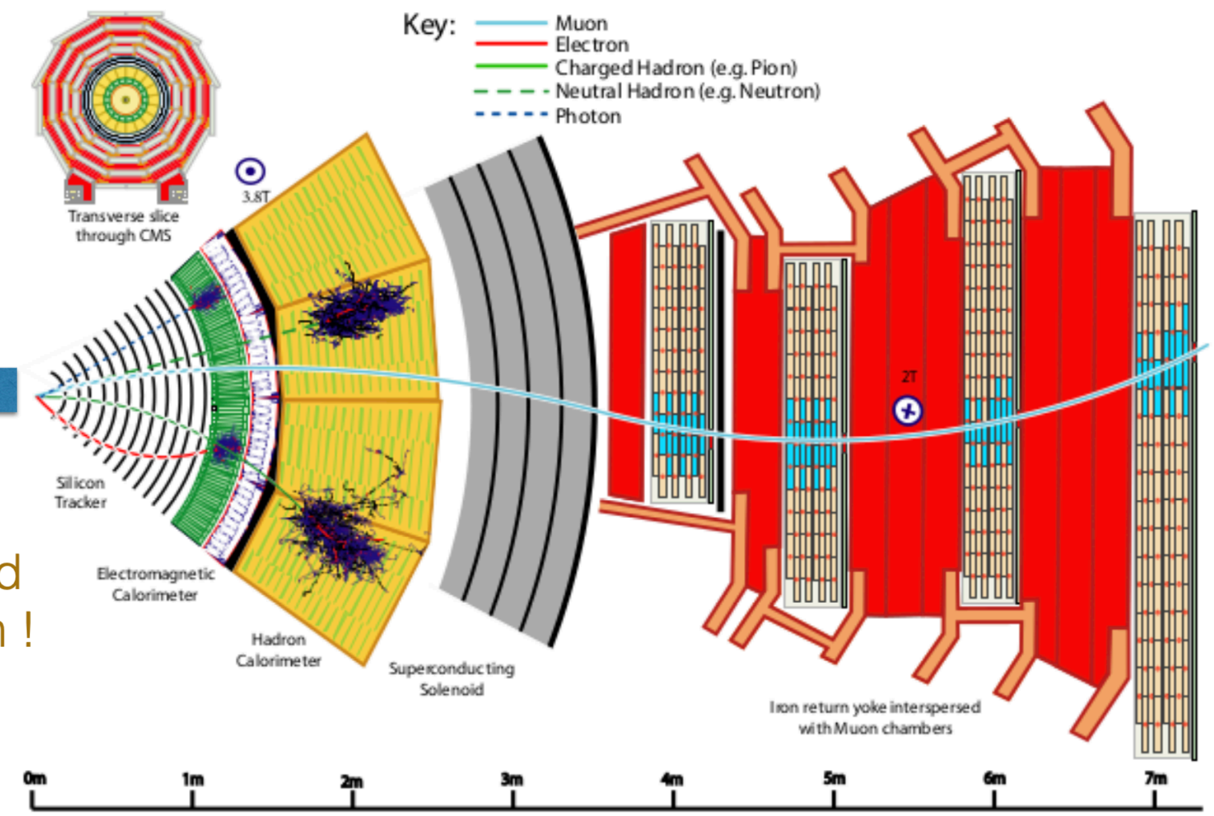


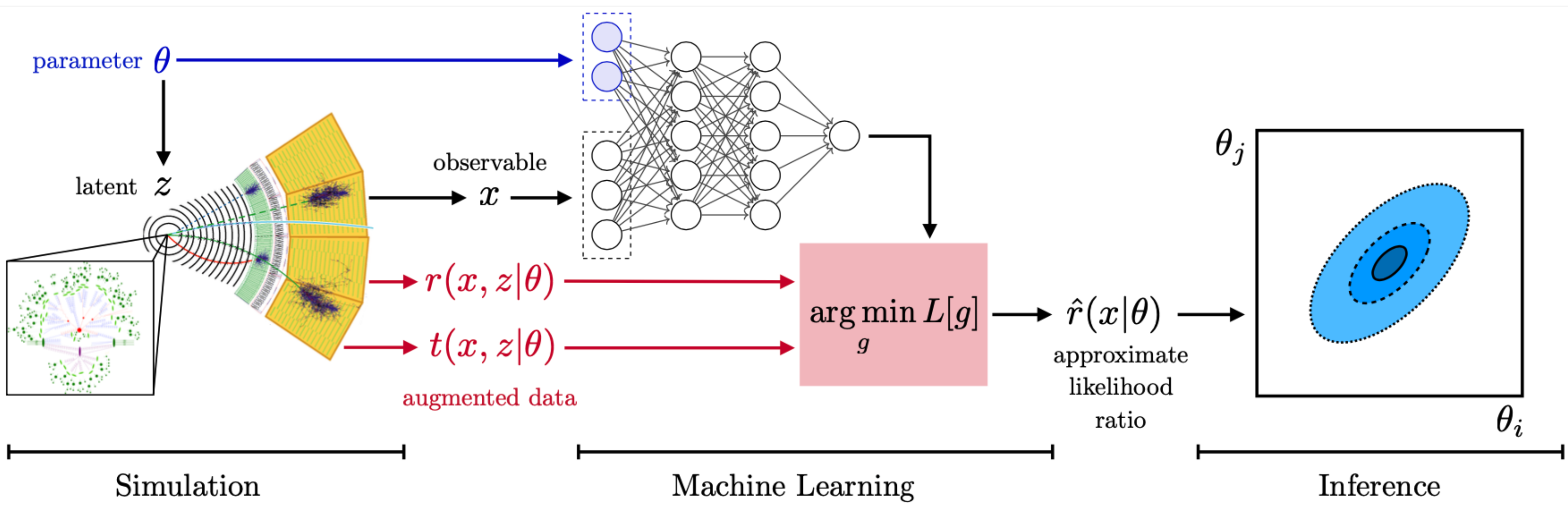


SBI in HEP

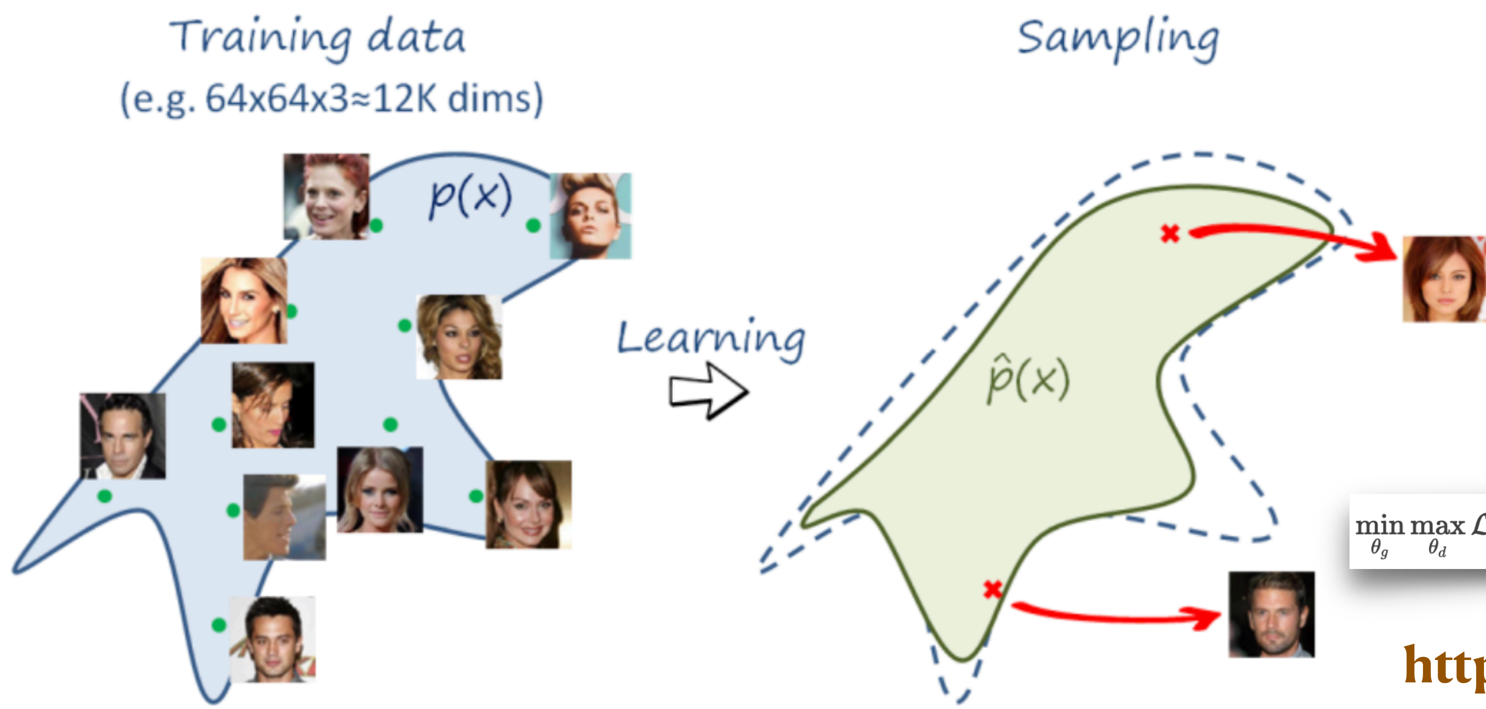


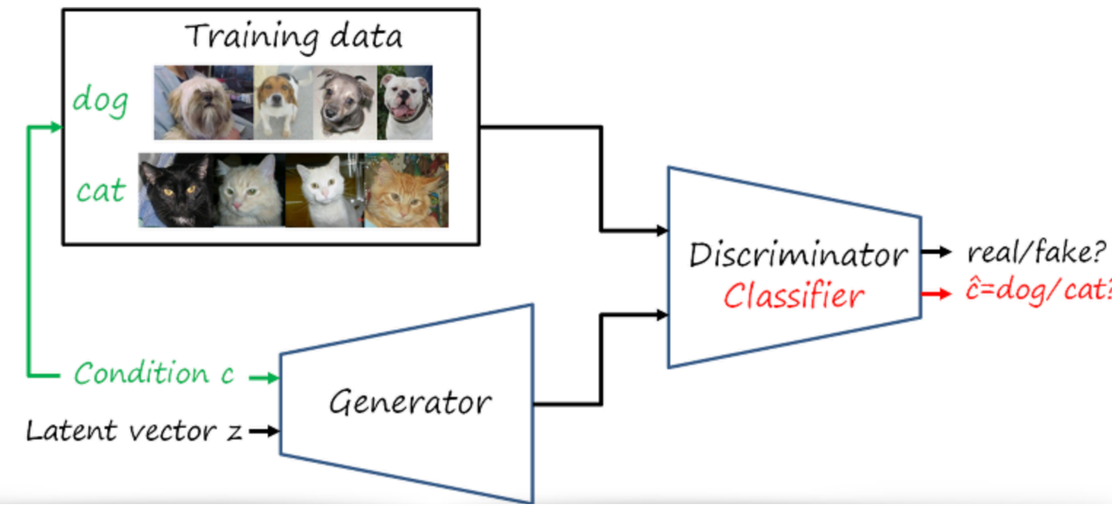
Underconstrained inversion problem !





Generative models : what are they?

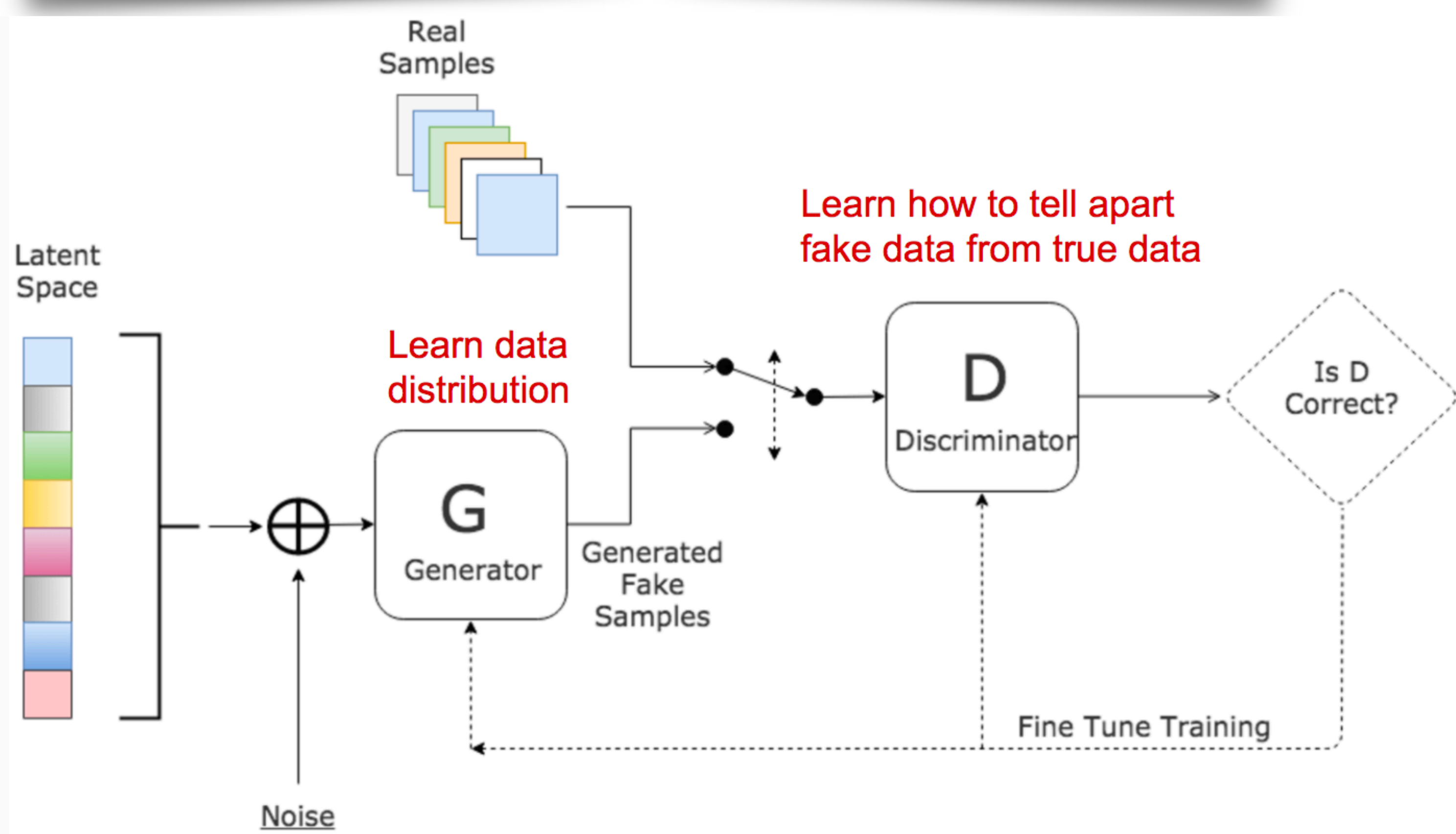


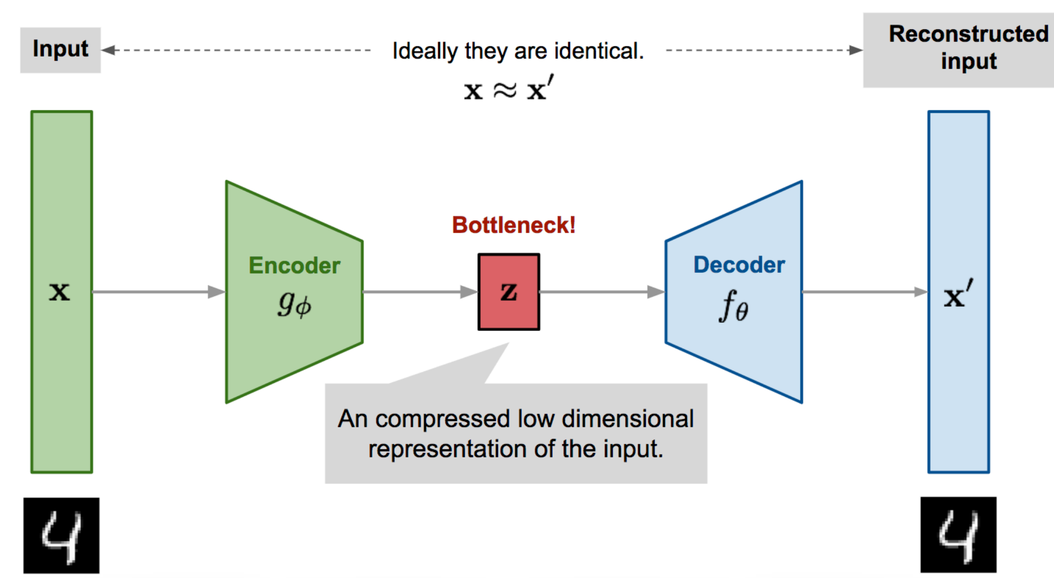


$$\min_{\theta_g}\max_{\theta_d}\mathcal{L}_{\text{cGAN}}\left(\theta_g,\theta_d\right)=\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{c,x\sim p_{\text{data}}(c,x)}\log D_{\theta_d}\left(c,x\right)+\mathbb{E}_{c\sim p_{\text{data}}(c),z\sim p_z(z)}\log\left(1-D_{\theta_d}\left(c,G_{\theta_g}\left(c,z\right)\right)\right)\right]$$

http://www.lherranz.org/2018/08/07/imagetranslation/

$$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p_{data}(x)}\log D_{\theta_d}\left(x\right)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}\left(z\right)\right)\right)\right]$$





$$L_{\text{AE}}(\theta,\phi)=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}^{(i)}-f_\theta(g_\phi(\mathbf{x}^{(i)})))^2$$

ttps://lilianweng.github.io/posts/

Generative models : the popular species

GAN: Adversarial training

VAE: maximize variational lower bound

Flow-based models: Invertible transform of distributions

Diffusion models: Gradually add Gaussian noise and then reverse

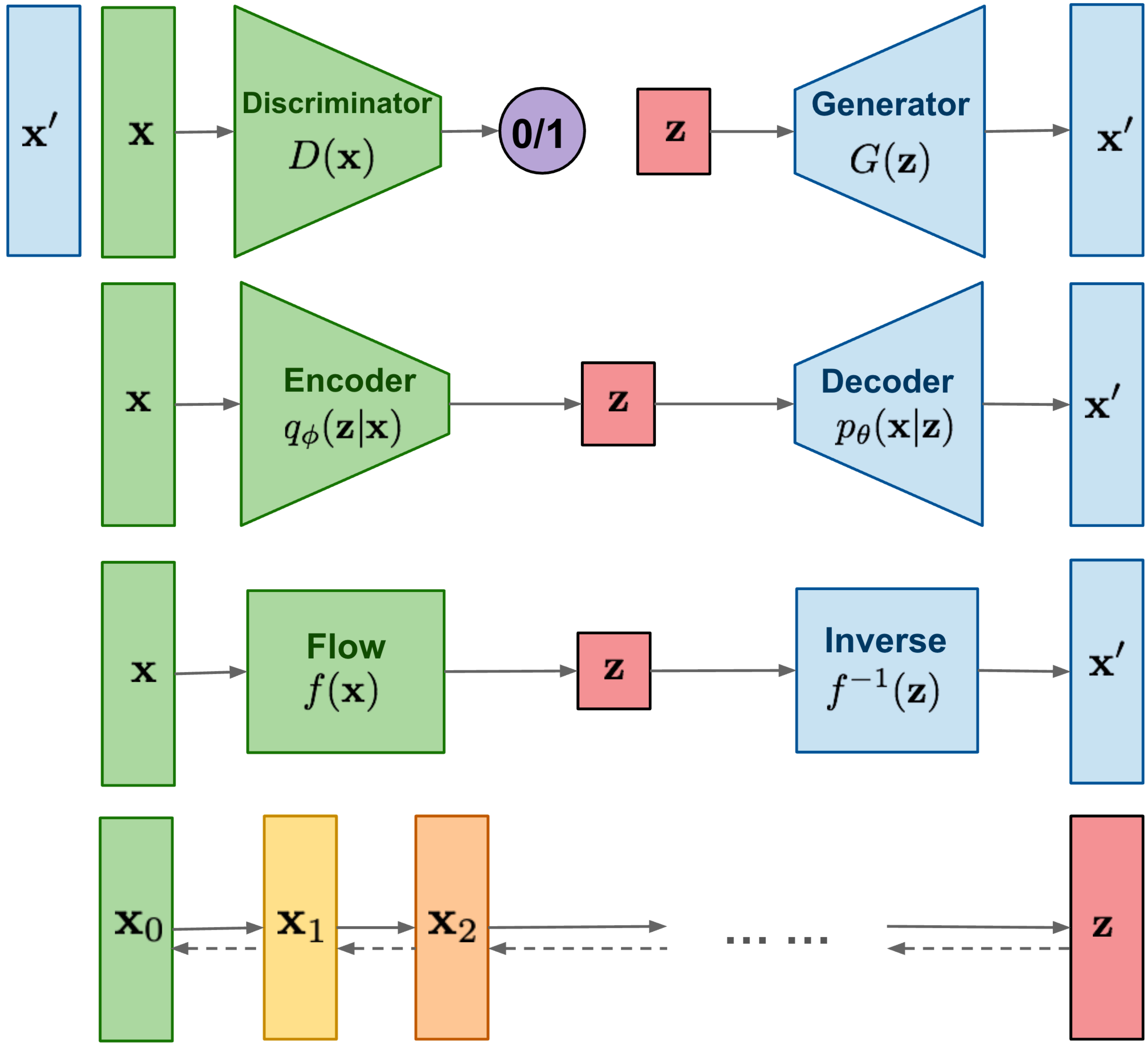


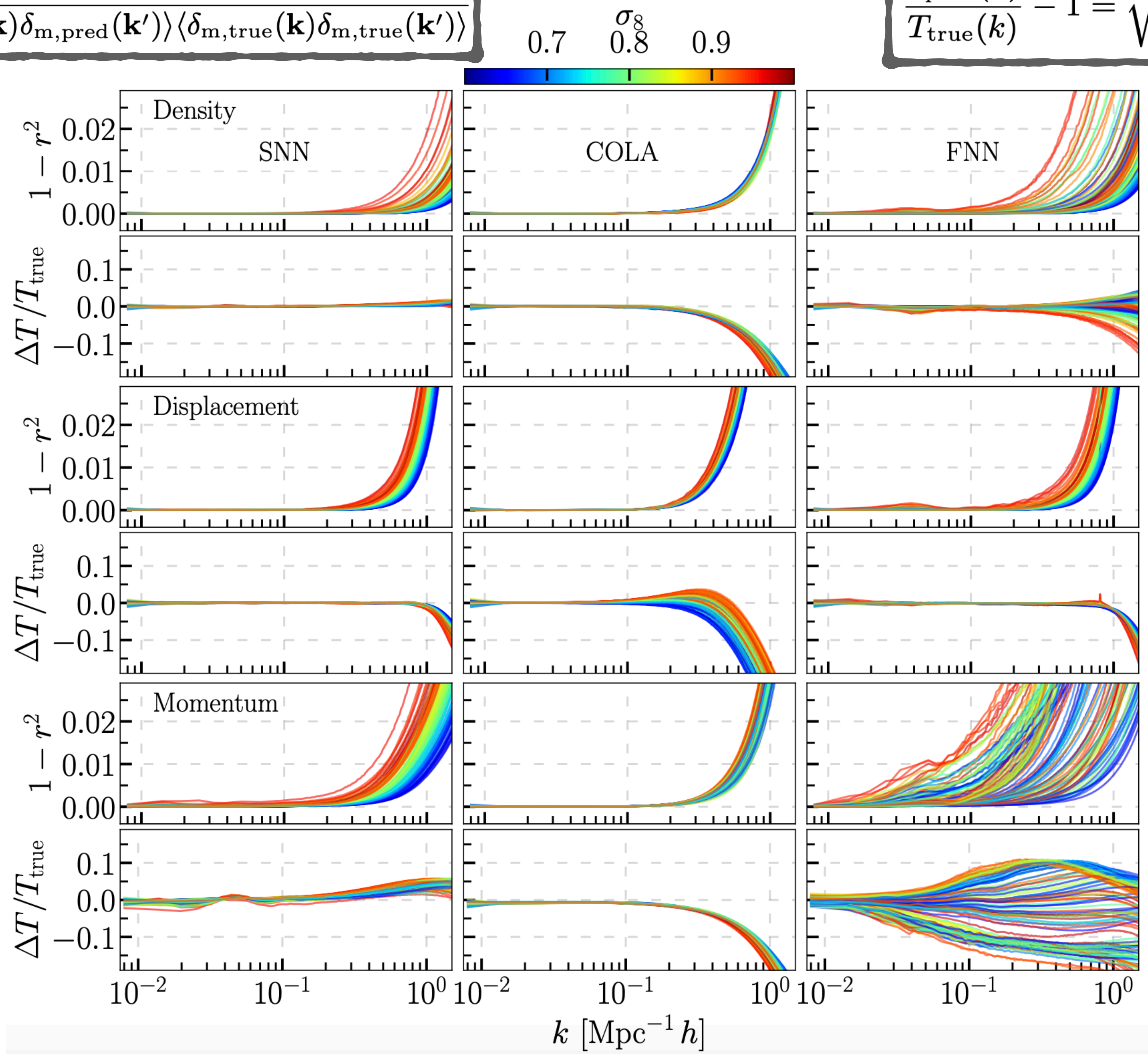
Fig from : https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

This universe doesn't exist

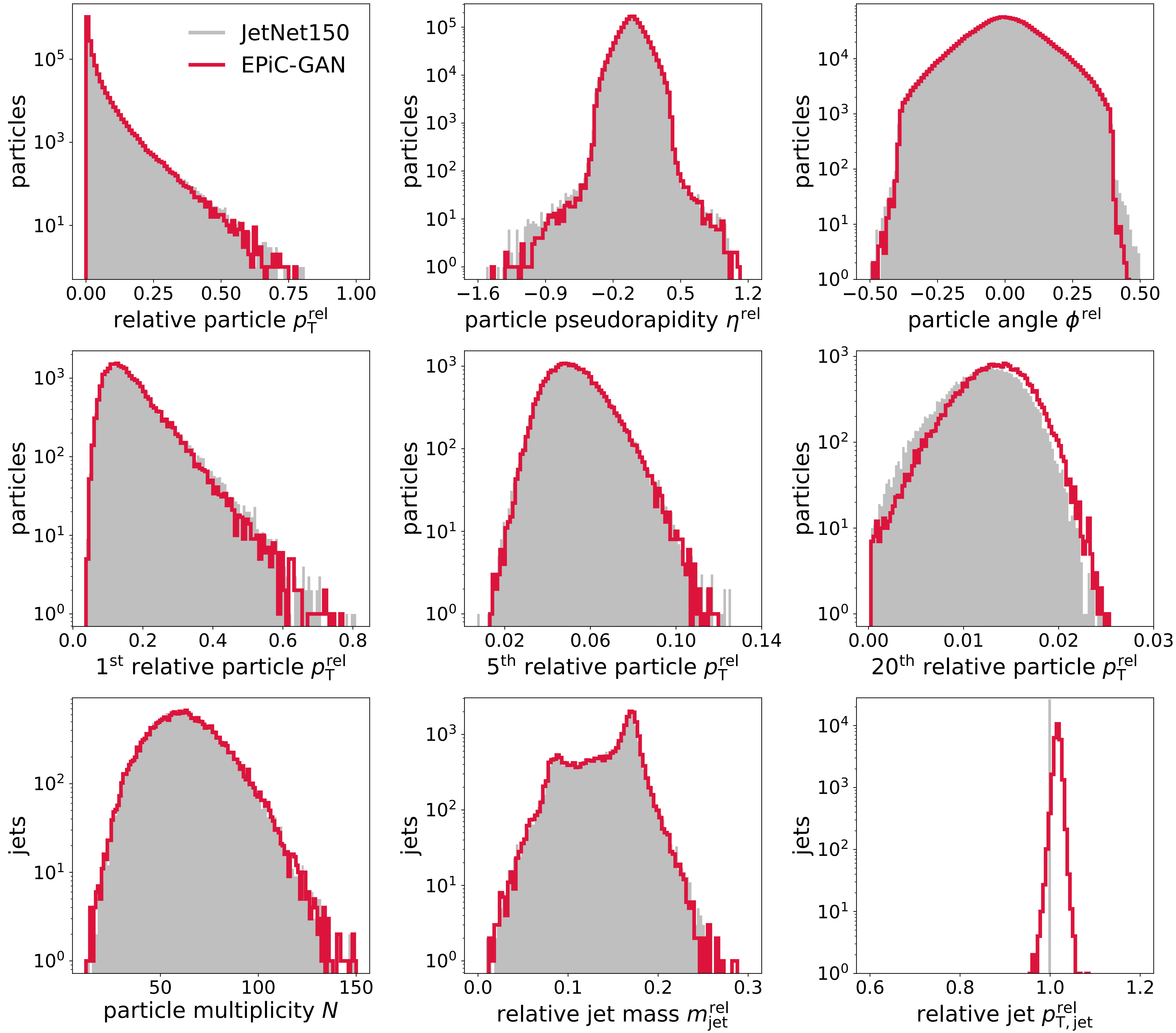
arXiv : 2206.04594

$$r(k) = \frac{\langle \delta_{\mathrm{m,pred}}(\mathbf{k})\delta_{\mathrm{m,true}}(\mathbf{k}')\rangle}{\sqrt{\langle \delta_{\mathrm{m,pred}}(\mathbf{k})\delta_{\mathrm{m,pred}}(\mathbf{k}')\rangle\langle \delta_{\mathrm{m,true}}(\mathbf{k})\delta_{\mathrm{m,true}}(\mathbf{k}')\rangle}}$$

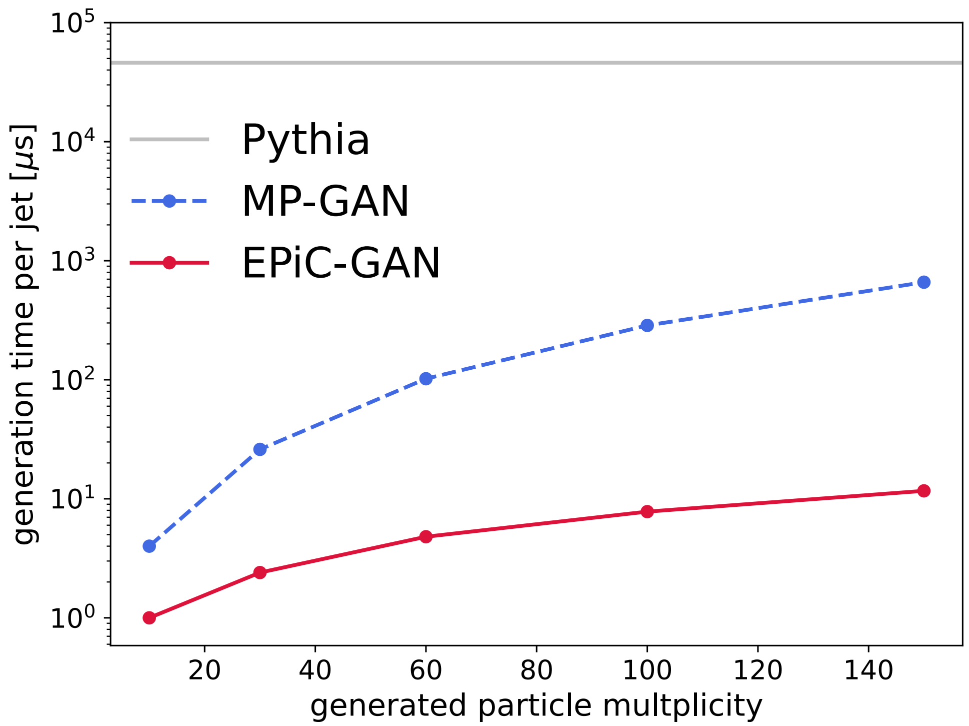
$$\frac{T_{\mathrm{pred}}(k)}{T_{\mathrm{true}}(k)} - 1 = \sqrt{\frac{P_{\mathrm{mm,pred}}(k)}{P_{\mathrm{mm,true}}(k)}} - 1$$

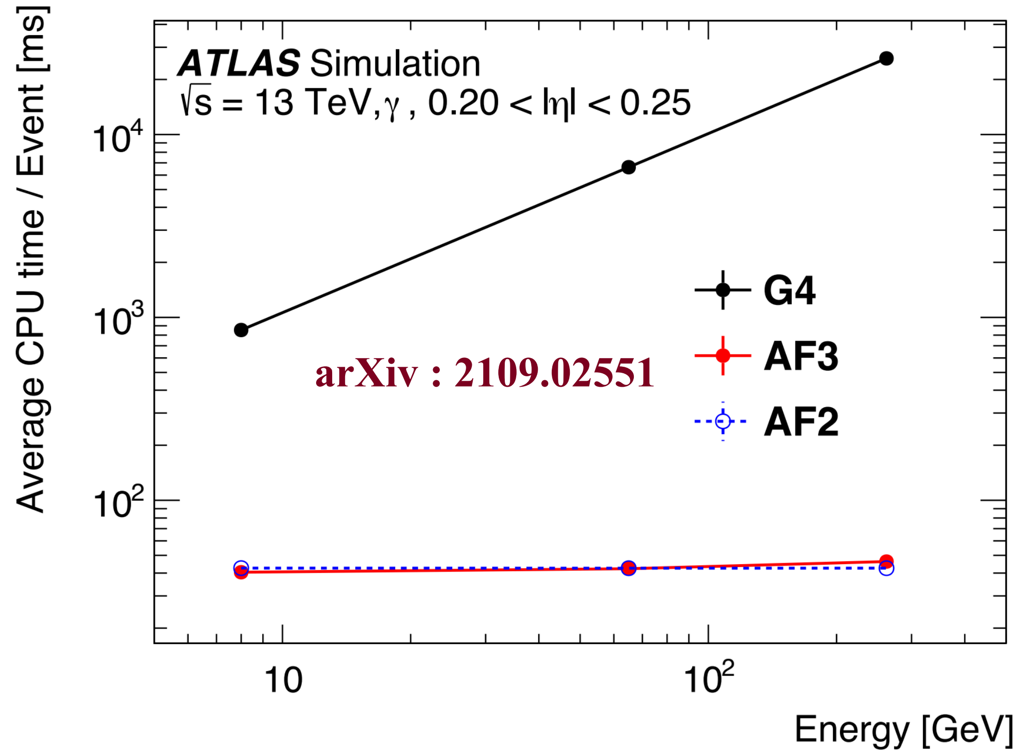


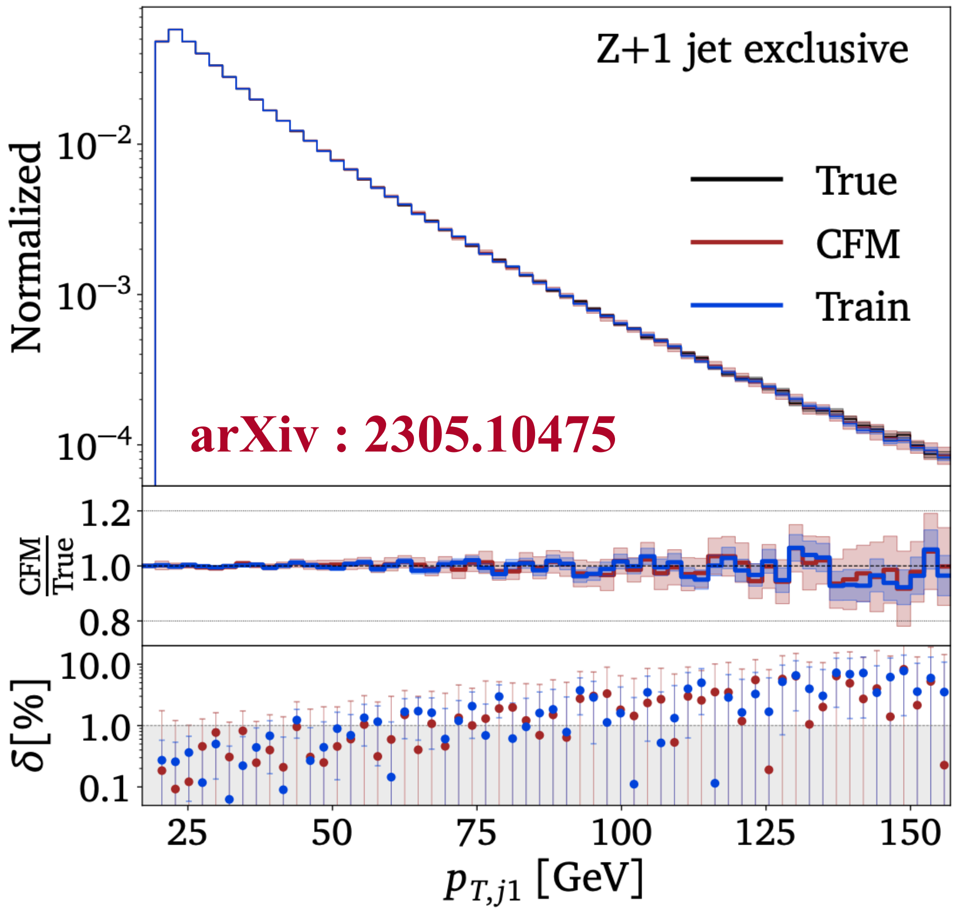
The performance

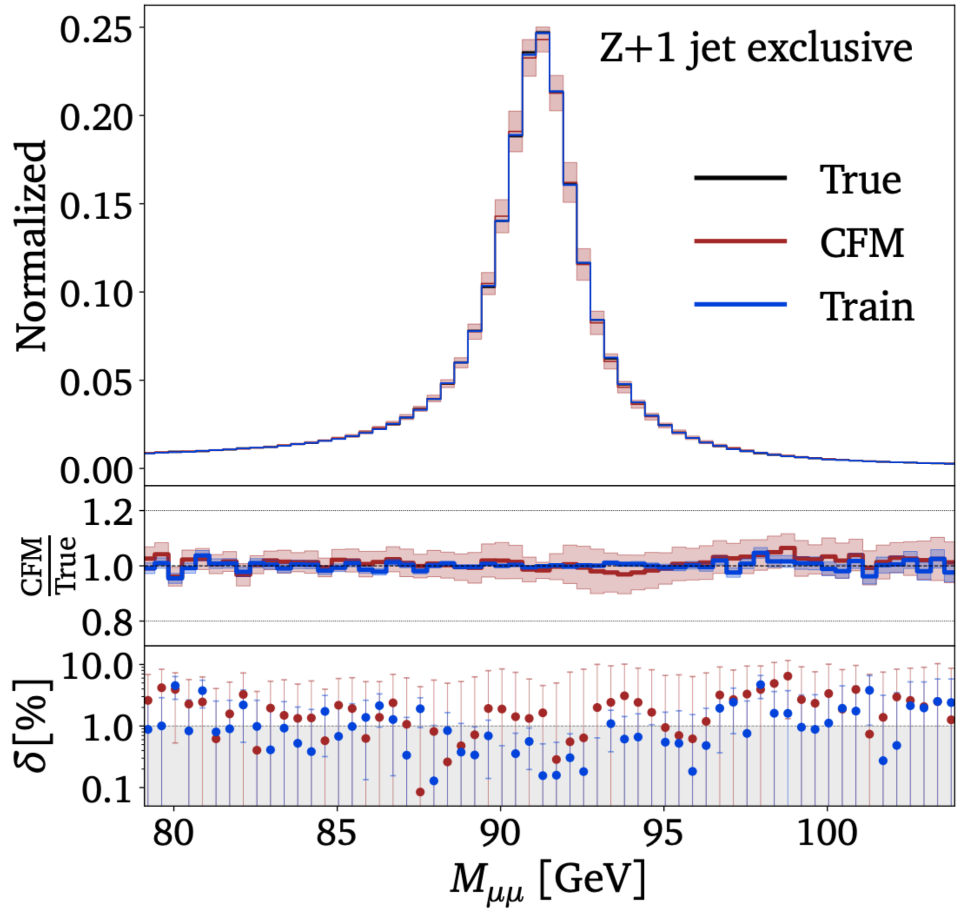


The major gain





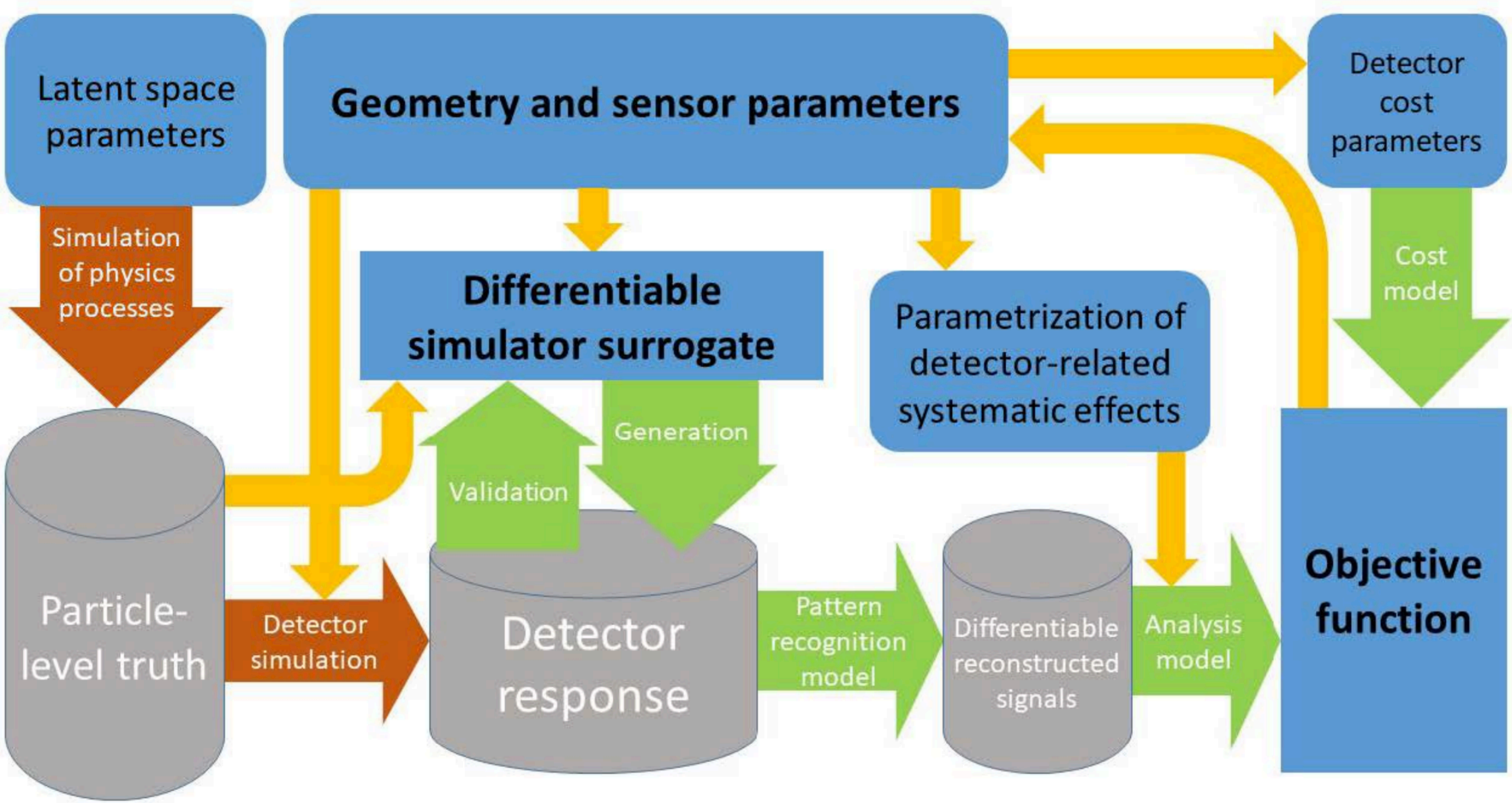




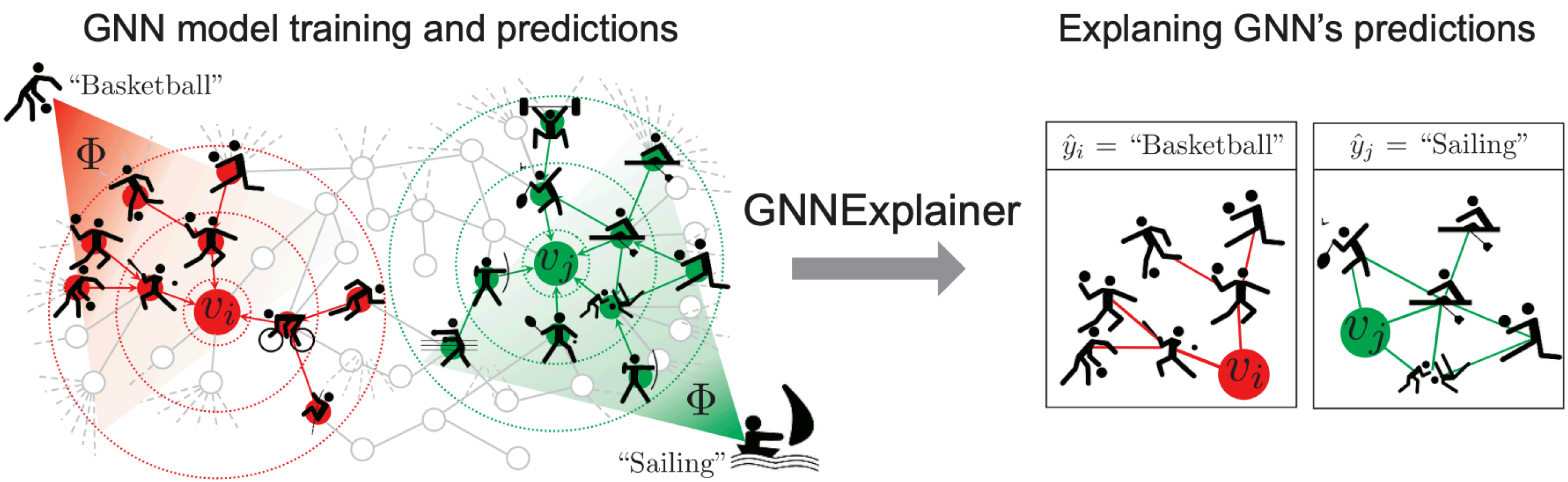
Differentiable programming

Differentiable programming allows us to configure our analysis optimization in learnable

https://mode-collaboration.github.io/



Major thrust in immediate future : Interpretability



Interpretability is a key issue and efforts are ongoing to map the NN explainability to first principle physics intuition

Interpretability : an example attempt

$$\mathbf{R}_j^{(l)} = \sum_k \frac{x_j A_{jk}}{\sum_m x_m A_{mk}} \mathbf{R}_k^{(l+1)} \tag{3}$$

where $\mathbf{R}_j^{(l)}$ represent the R-scores of the features of node j at layer l, while the quantity $x_j A_{jk}$ models the extent to which node j at layer l, with activation x_j, contributes to the relevance of node k at layer $l+1$, where A is the adjacency matrix.

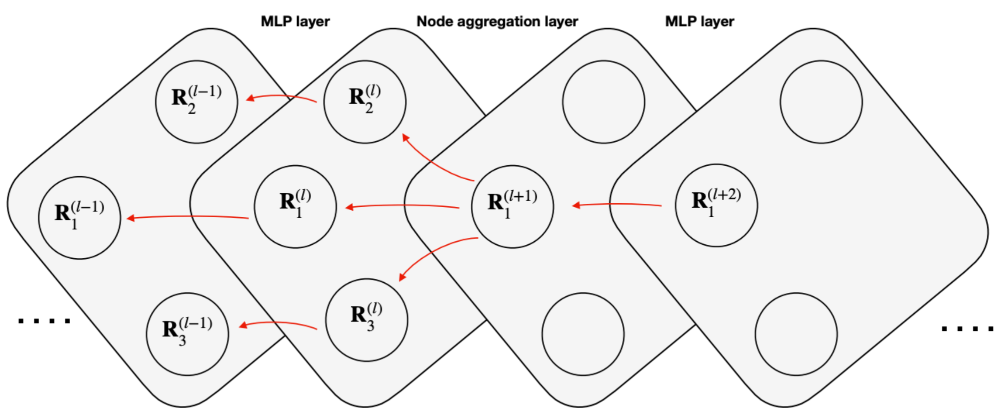


Figure 1: The flow of R-scores of node 1 across the different layers in MLPF. For MLP layers, the redistribution of R-scores follows the standard LRP rules [35, 36]. For the aggregation step in the message passing layer, the redistribution follows Equation 3. We only show three nodes for simplicity.

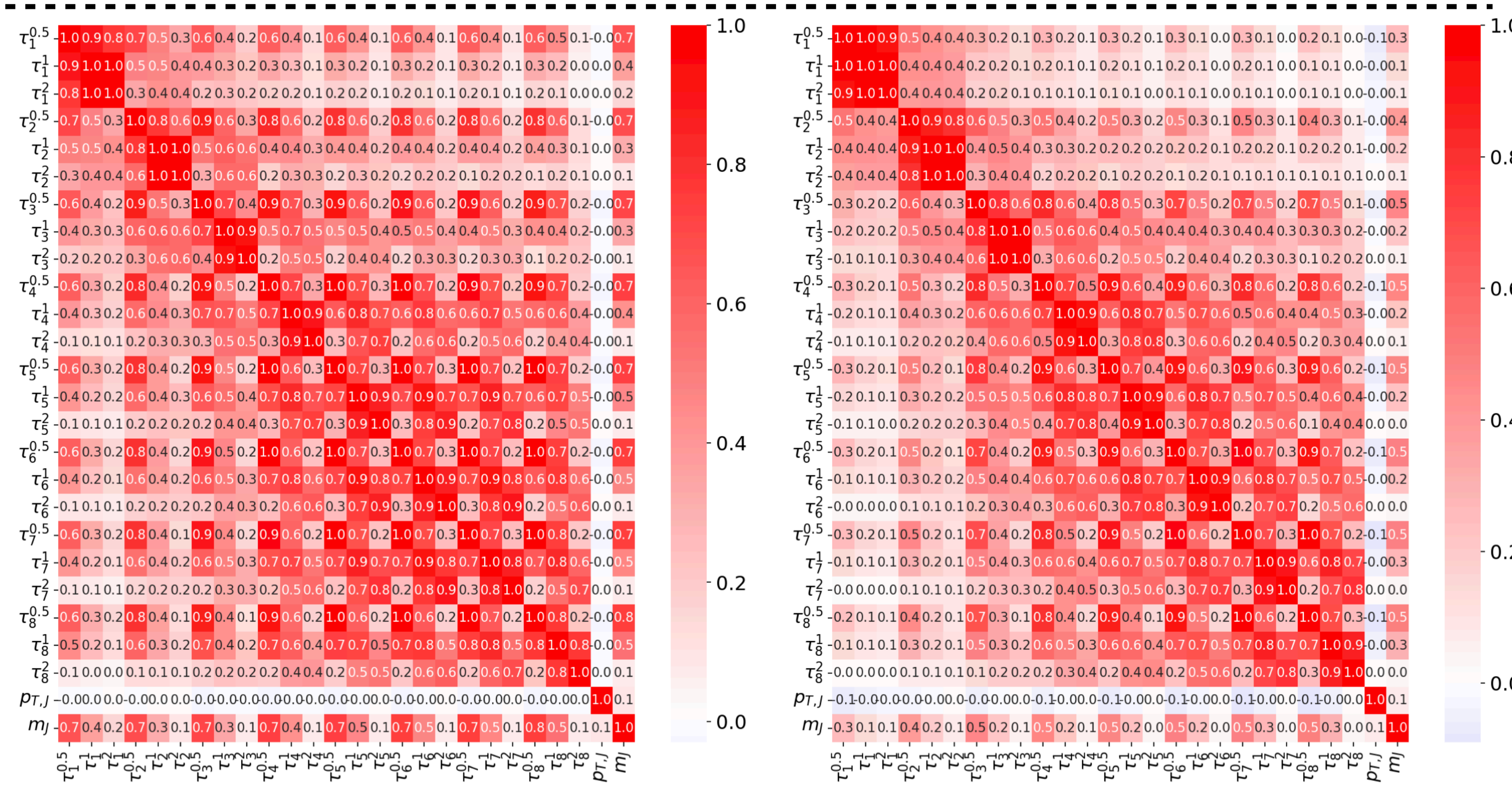
Neur IPS 2021. F. Mokhtar, R. Kansal et al

Explainability for MLPF

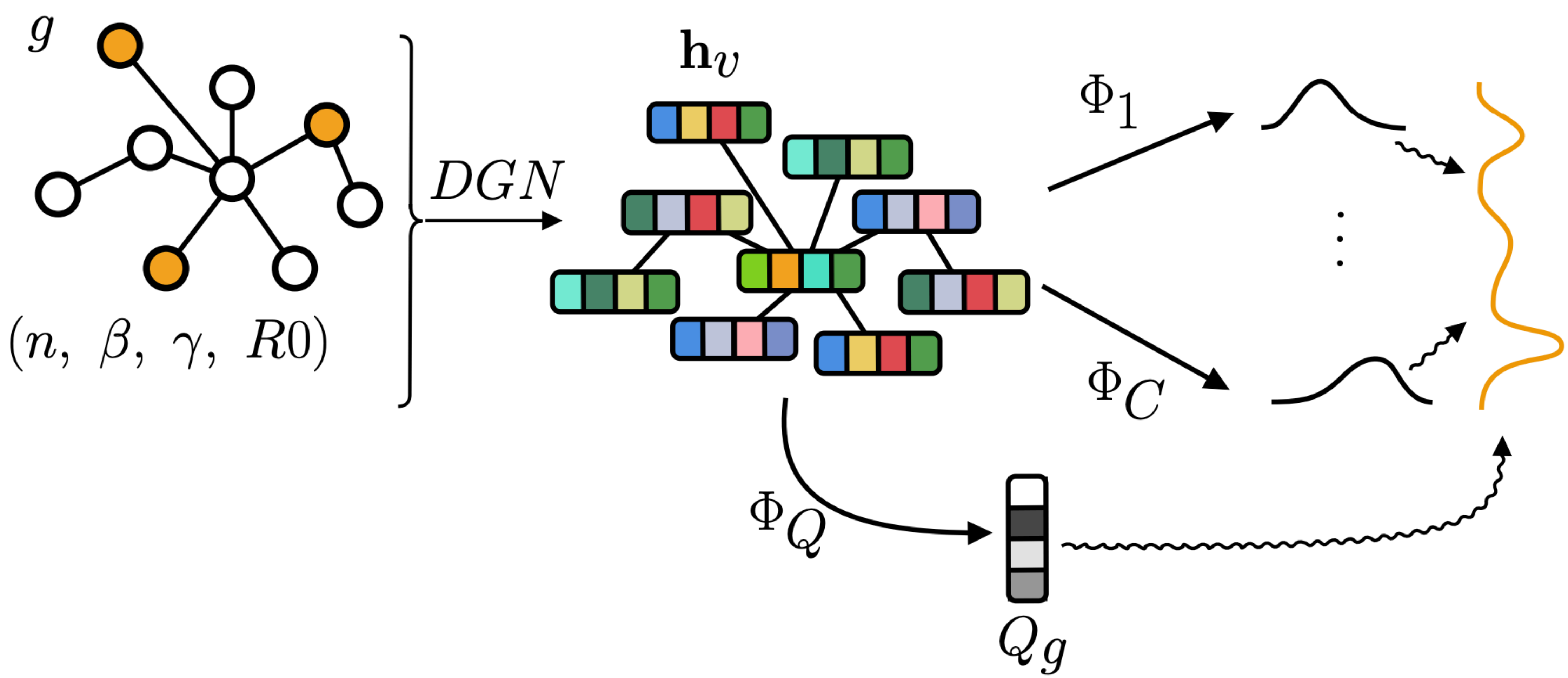
Feature correlation for top tagging.

arXiv 2210.04371

Ayush Khot, Mark S. Neubauer, Avik Roy



Major thrust in immediate future : Uncertainty

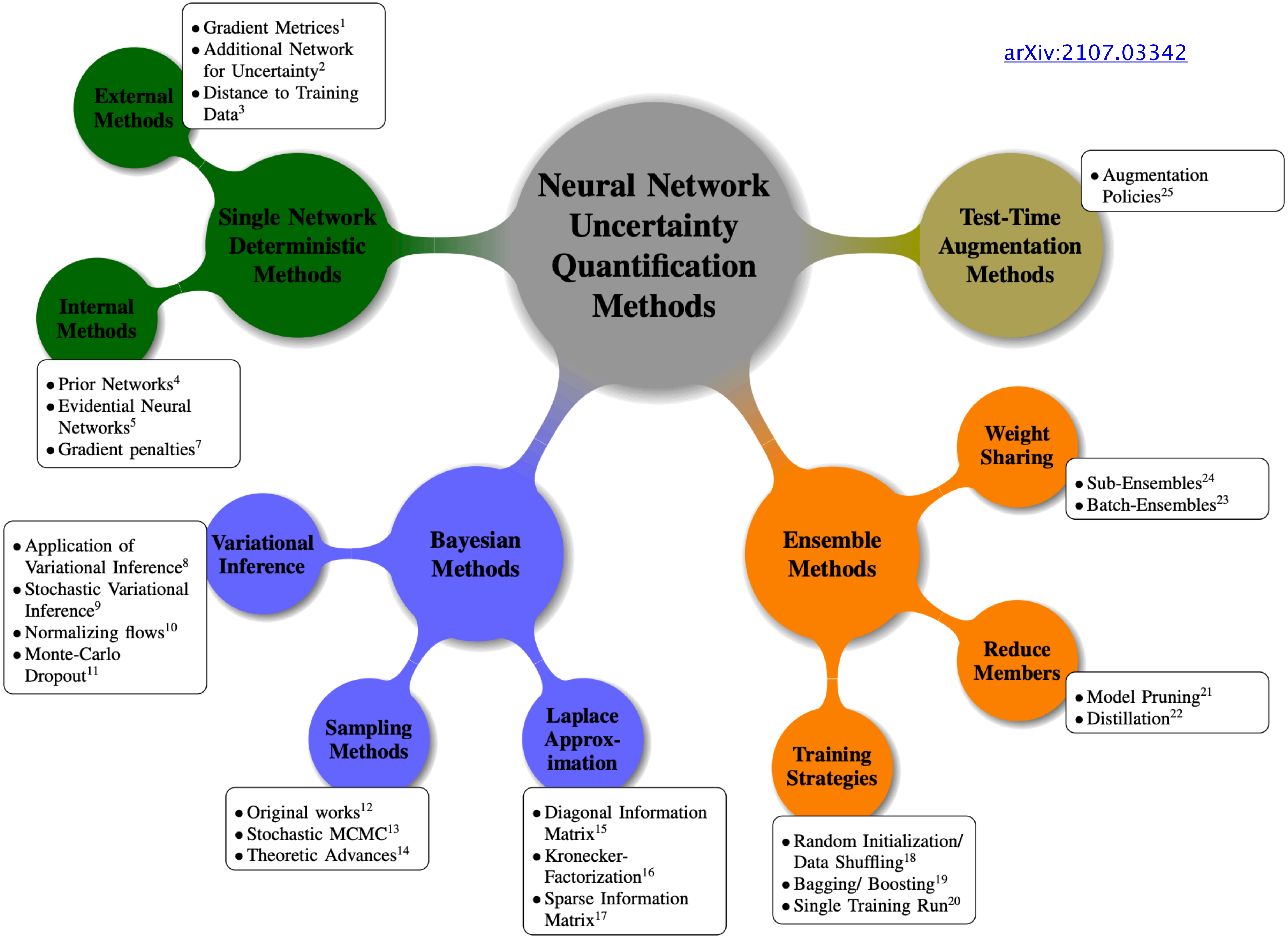


Reliable uncertainty estimation on ML based predictions are crucial for HEP
Only few Bayesian methods have been tested naively.

Can we decompose and correlate the aleatoric and epistemic uncertainties with the underlying physics?

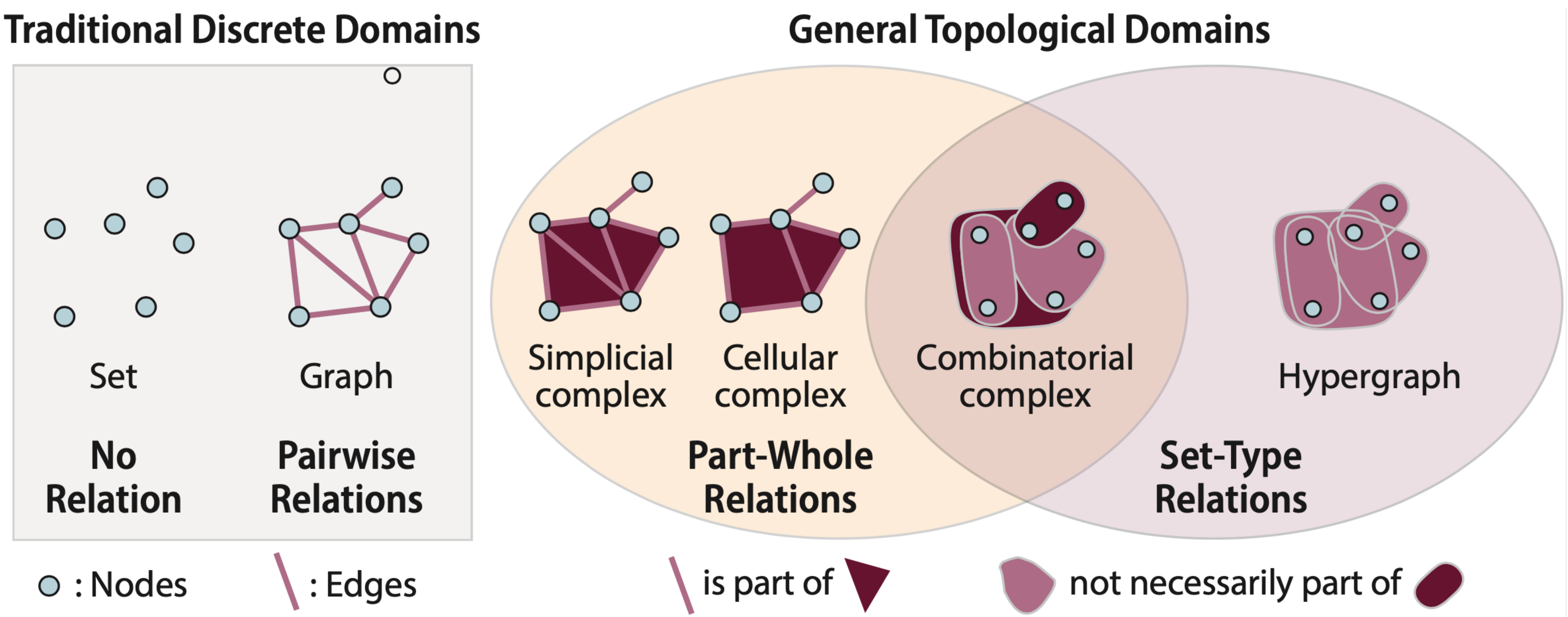
Major thrust in immediate future : Uncertainty

arXiv:2107.03342

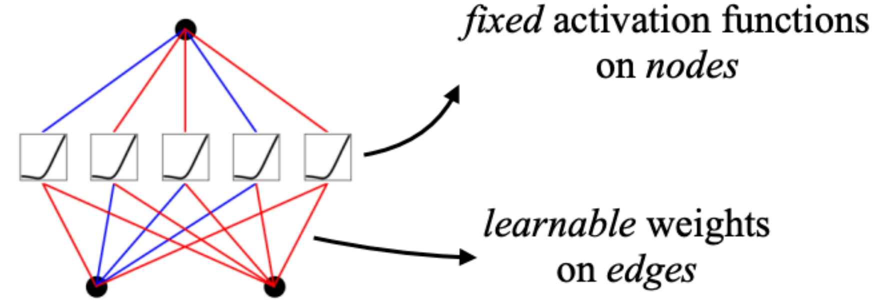
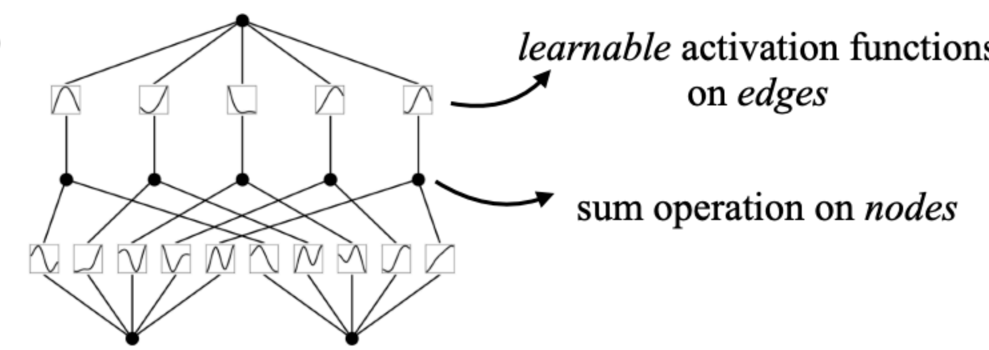
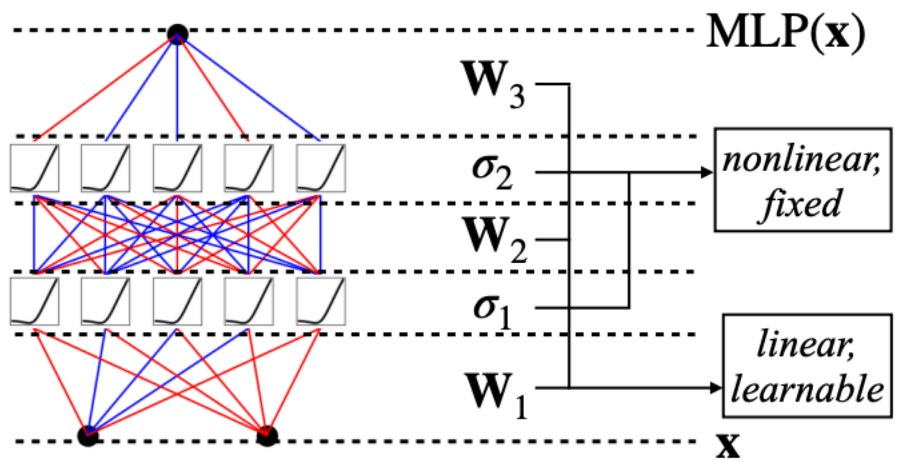
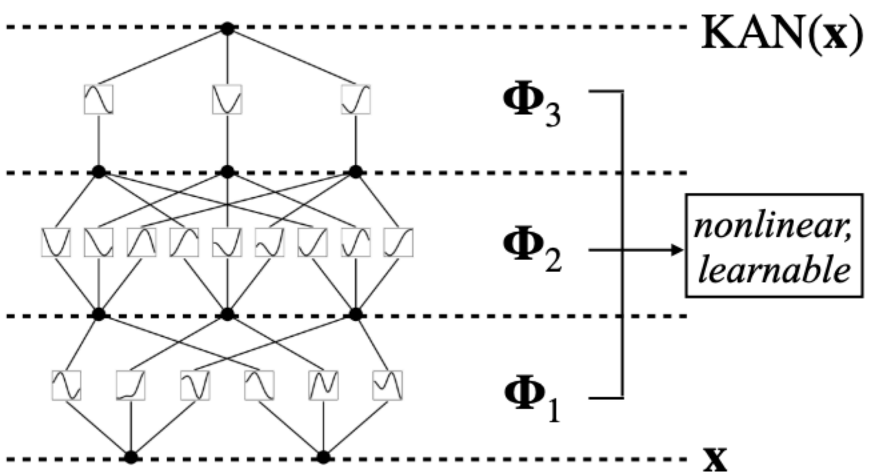


An example of next frontiers

https://pyt-team.github.io/toponetx/



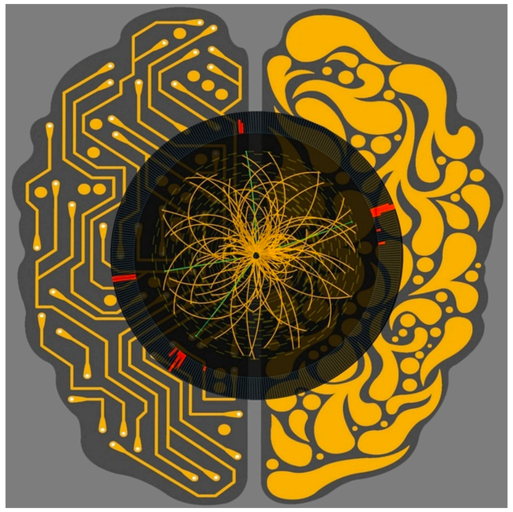
An example of next frontiers

Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a) 	(b) 
Formula (Deep)	$\mathrm{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\mathrm{KAN}(\mathbf{x}) = (\mathbf{\Phi}_3 \circ \mathbf{\Phi}_2 \circ \mathbf{\Phi}_1)(\mathbf{x})$
Model (Deep)	(c) 	(d) 

Better interpretability through KAN ?

Let's formulate the questions

Image: FermiLab



- ☑ **ML is here to stay with HEP/Cosmology and other branches of natural sciences.**
- ☑ **When looked through the lens of ML, it's about finding the right inductive bias for a prob. distribution**
- ☑ **Interpretability and uncertainty estimation is a corner stone which we should emphasize.**
- ☑ **The HEP community should talk with mathematicians/comp-sc and other branches of natural science who are using the similar methods and exchange ideas.**

https://iml-wg.github.io/HEPML-LivingReview/
https://github.com/georgestein/ml-in-cosmology

THANK YOU