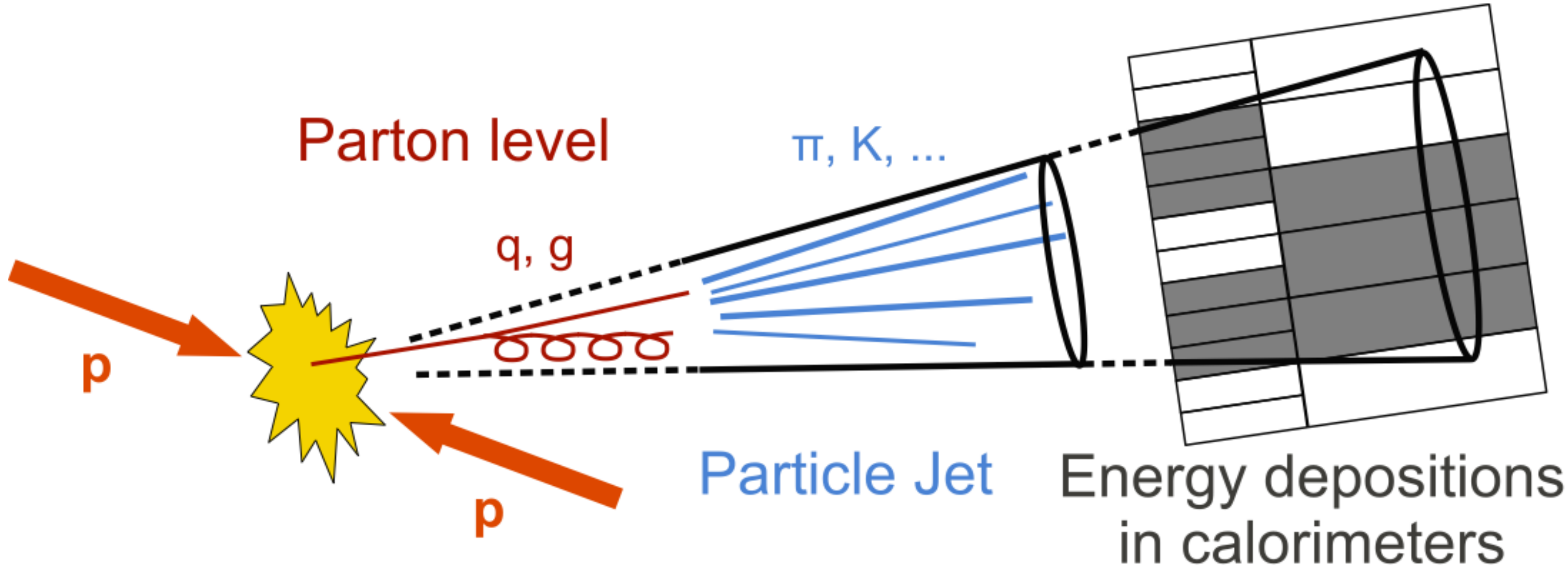
Jet Physics and Machine Learning:
Lecture 4 : ML for Jets & HEP

Sanmay Ganguly
IIT-Kanpur
sanmay@iitk.ac.in

28/02/2025

IMSc Spring School on High Energy Physics - 2025



The path we will take

Basic Motivation

- Why are we doing ML?
- The basic LHC operations with ML.
- Synergies through eye of ML.

The supervised ML

- Jet tagging : the most successful venture.
- Tracking & Calorimetery
- The Particle-Flow algorithm

Generative models

- Different paradigms
- Few physics cases.
- Interpretable ML & uncertainty estimation.

The anomaly detection saga

- The outliers & density estimate
- How to translate this in new signals?

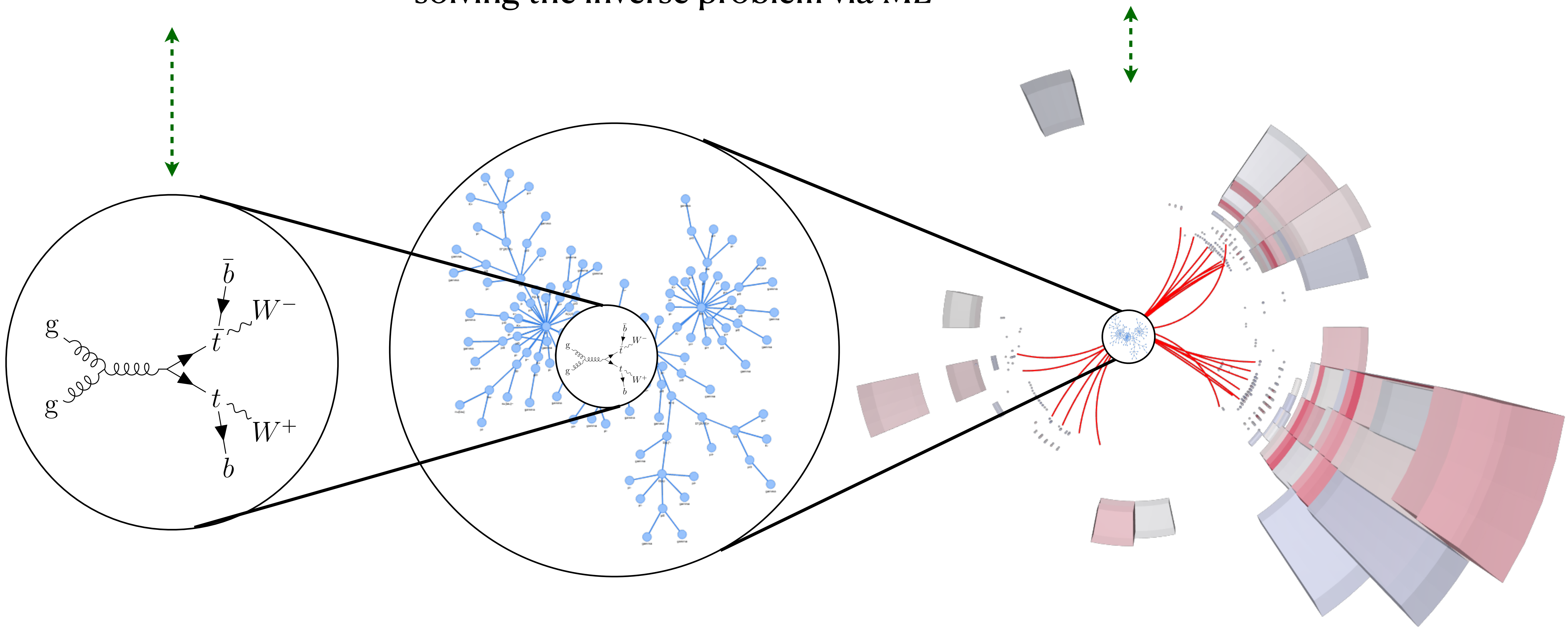
THE ROAD MAP

A broad strategy towards physics inference

Guess the Lagrangian

solving the inverse problem via ML

$$\frac{1}{\mathscr{L}}\frac{d^2N}{dp_T d\eta}$$

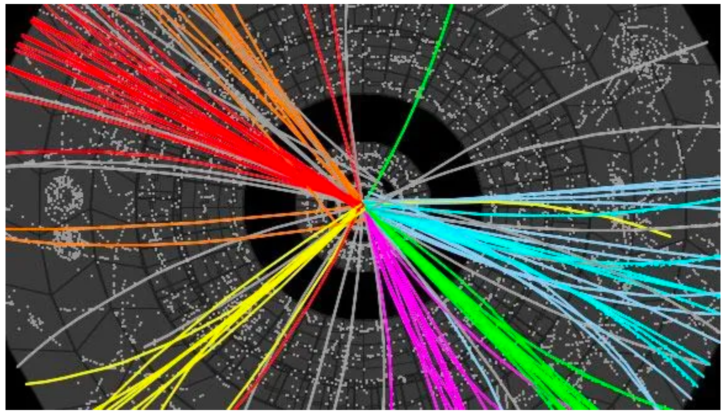


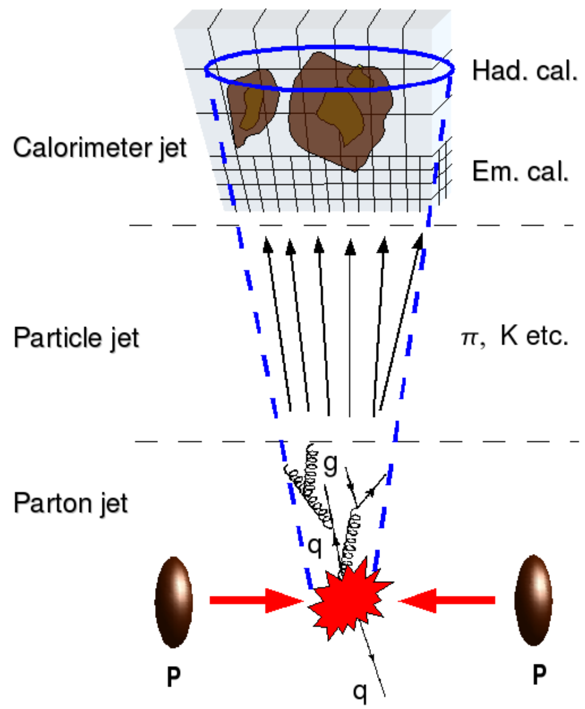
The physics at the core:
driven by the interaction between quantum fields, computed via perturbative or lattice techniques.

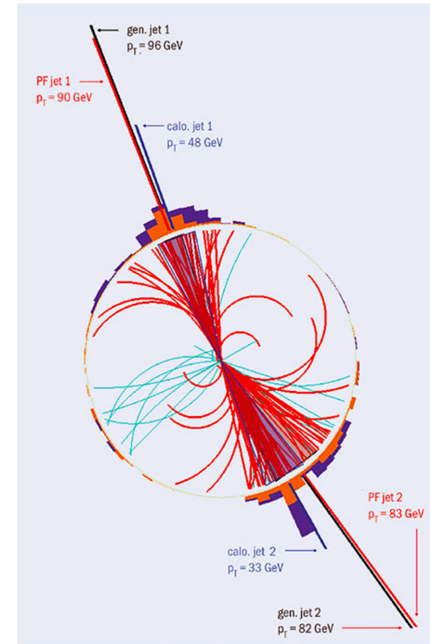
Final Particles
Produced via hadronization (no first principle analytic techniques are available)

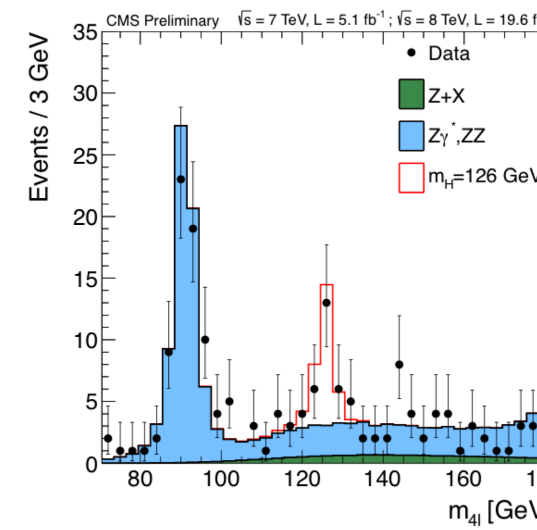
Detector output/Readout
Produced via hardware or simulation

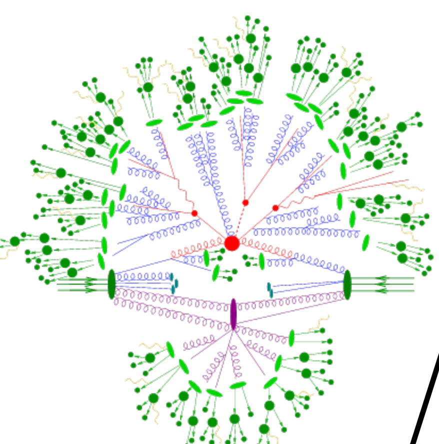
The collider program flow-chain









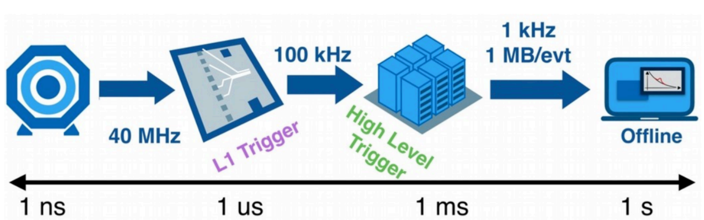


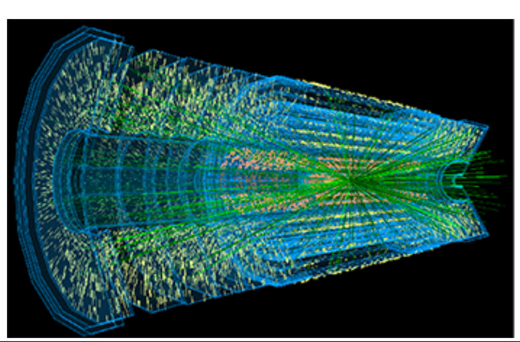
Trigger
Online process

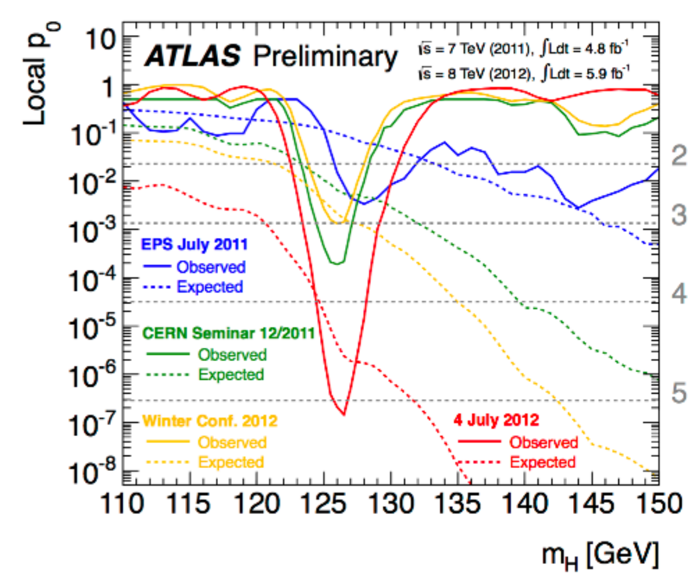
Object rec
PID + calib

Event rec
sig/bkg class.

Analysis
Unfolding/matching





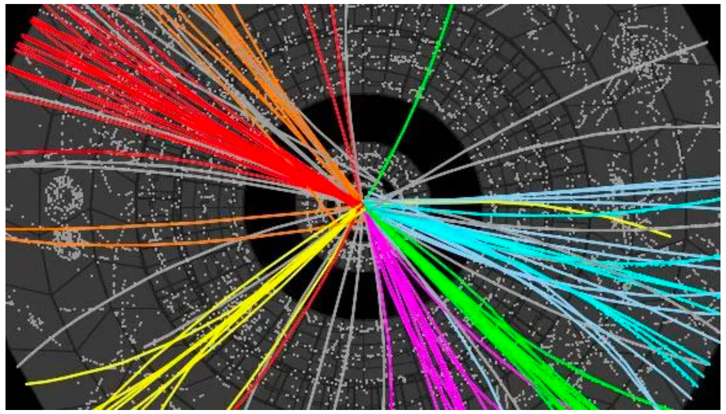


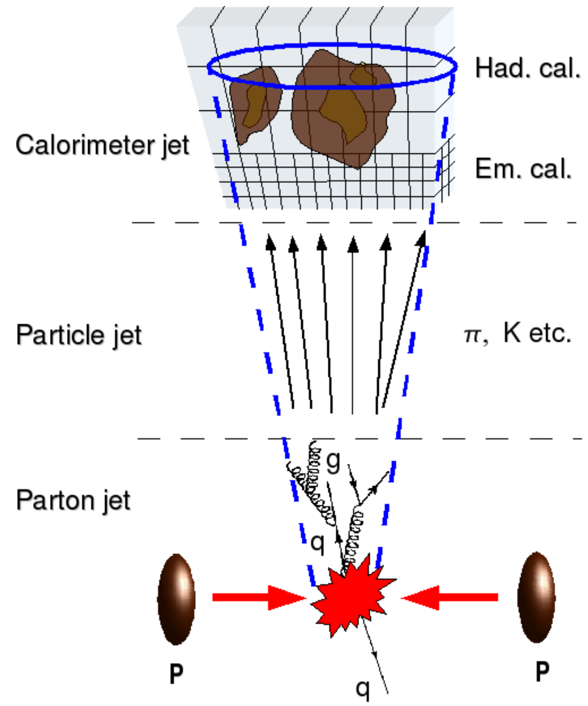
Inference

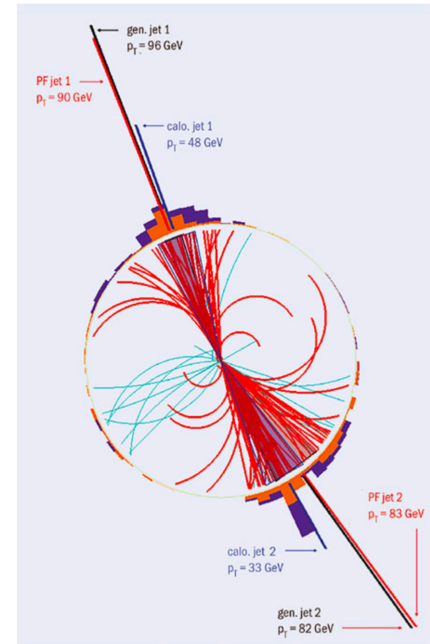
ME + PS generation
Detector Simulation

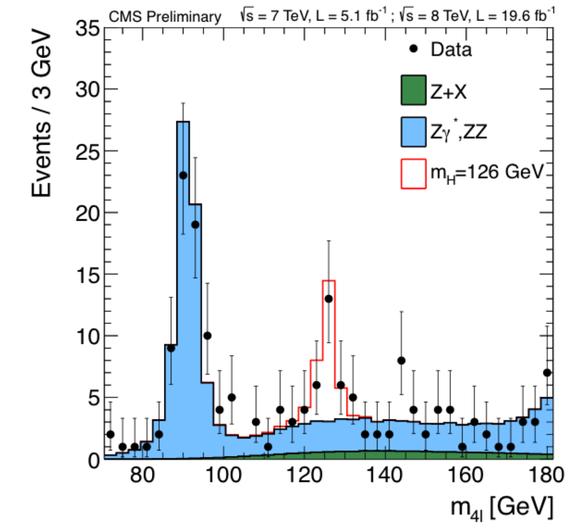
The collider program flow-chain

ML can (and will) play a role at every instance of this flow chain.







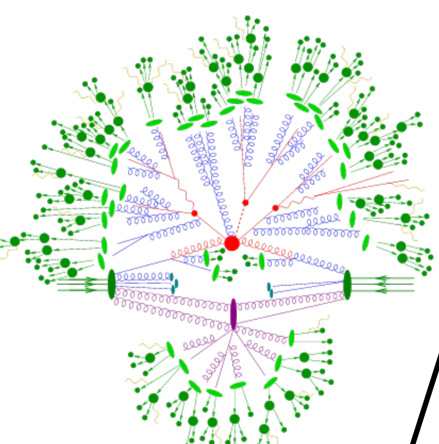


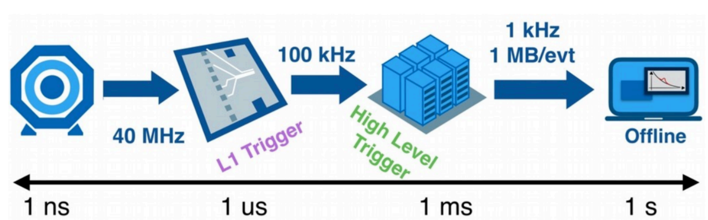
Trigger
Online process

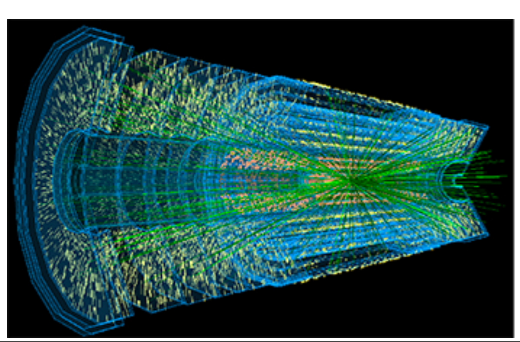
Object rec
PID + calib

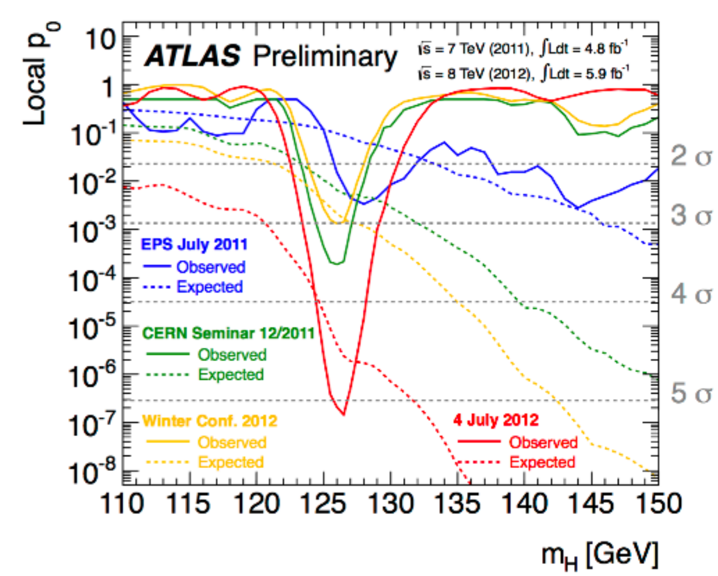
Event rec
sig/bkg class.

Analysis
Unfolding/matching





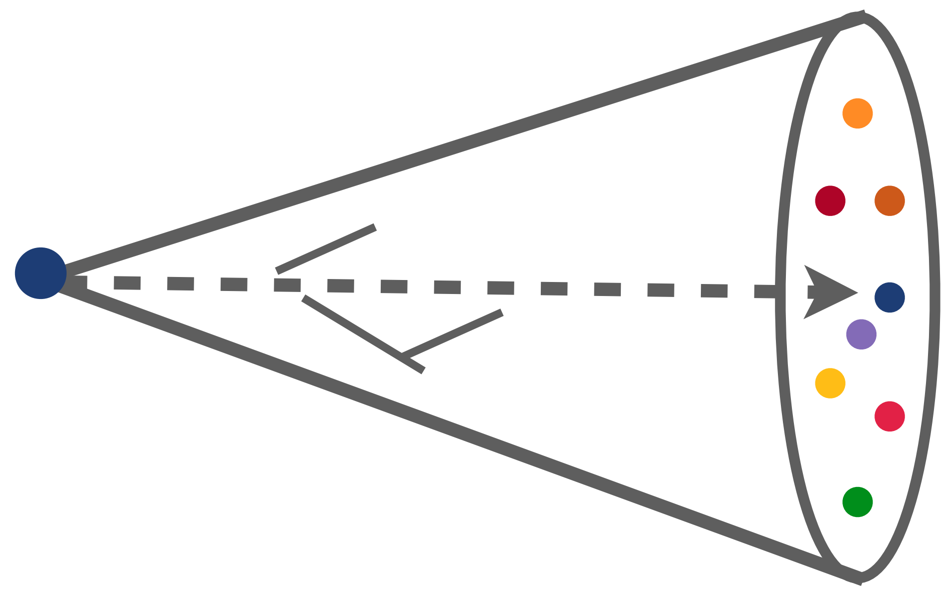




Inference

ME + PS generation
Detector Simulation

ML@Colliders : what's the broad task?

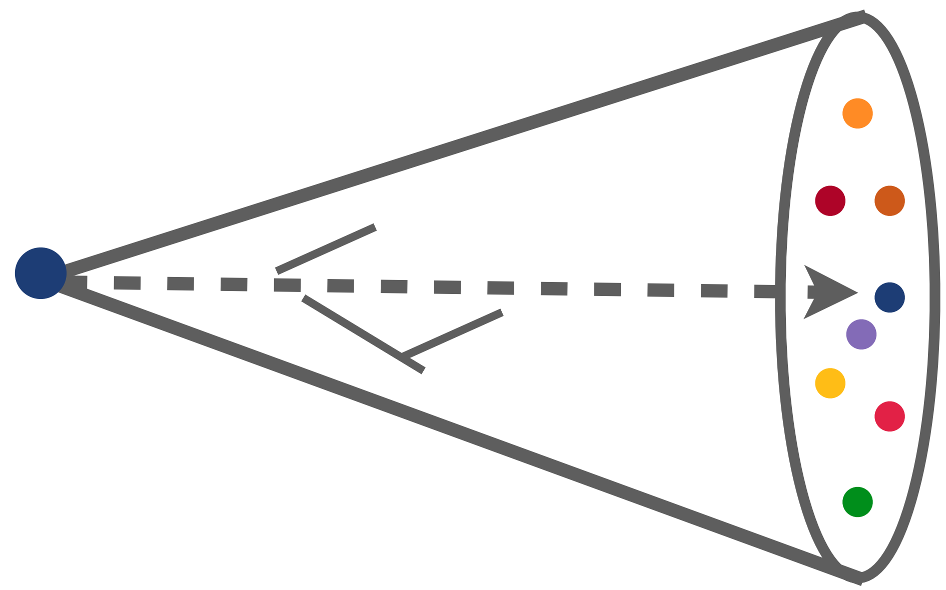
$f_{\{\theta\}}\Bigg($  $\Bigg) \longrightarrow \hat{y} = f_{\{\theta\}}(X)$

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

ML@Colliders : what's the broad task?

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

$f_{\{\theta\}}$ (



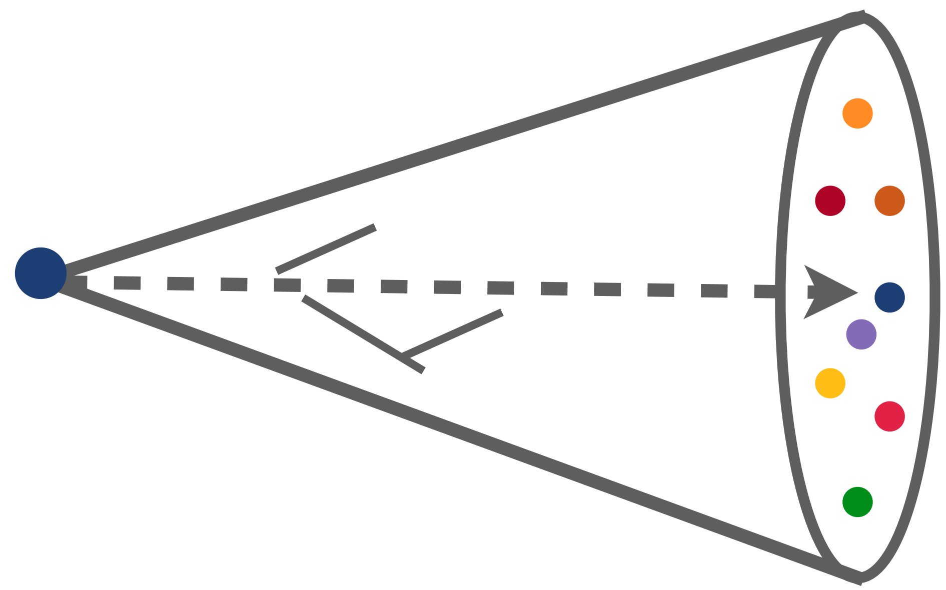
) $\rightarrow$ $\hat{y} = f_{\{\theta\}}(X)$

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

ML@Colliders : what's the broad task?

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

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

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

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

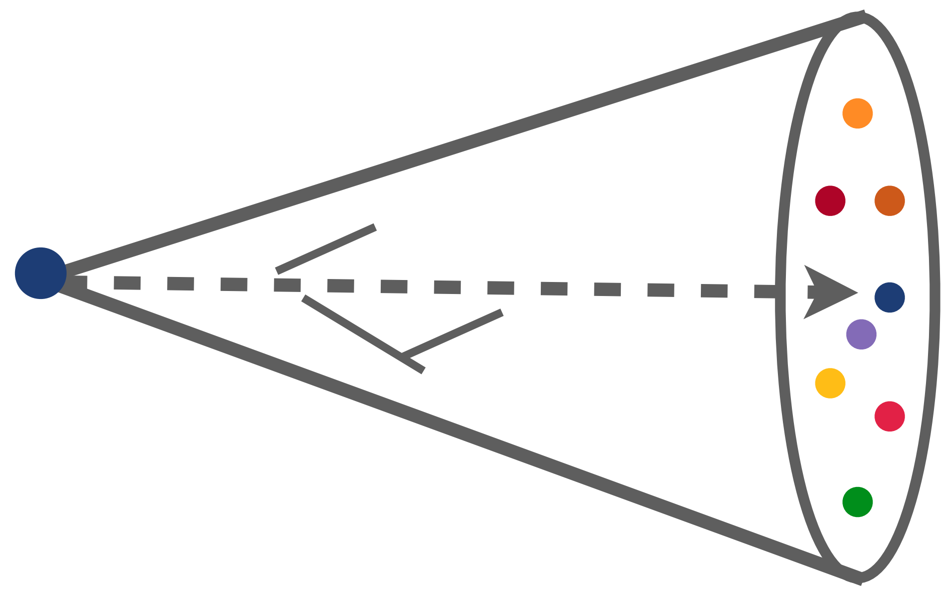
ML@Colliders : what's the broad task?

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.

$f_{\{\theta\}}$ (



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

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

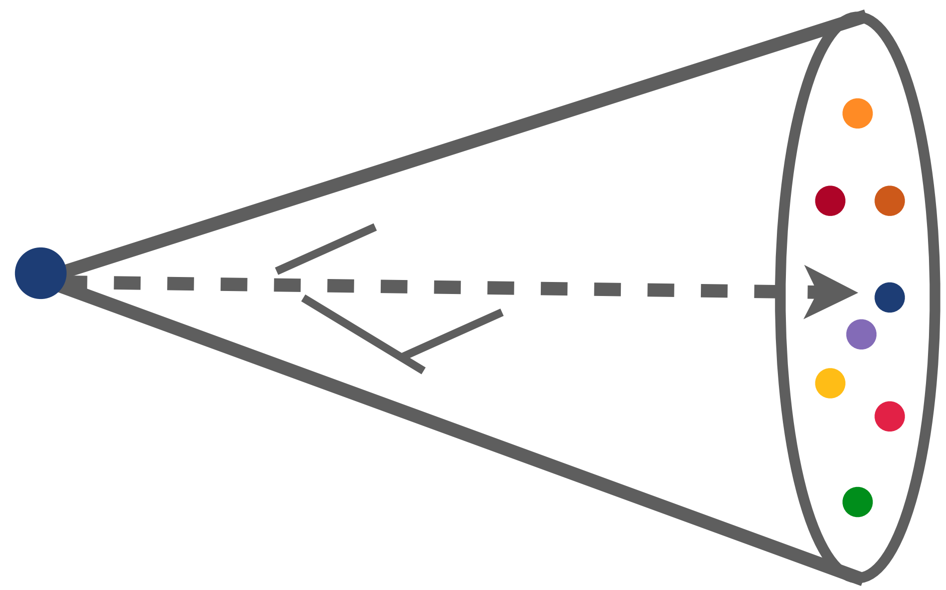
ML@Colliders : what's the broad task?

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.

$f_{\{\theta\}}$ (



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

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

Variation in data

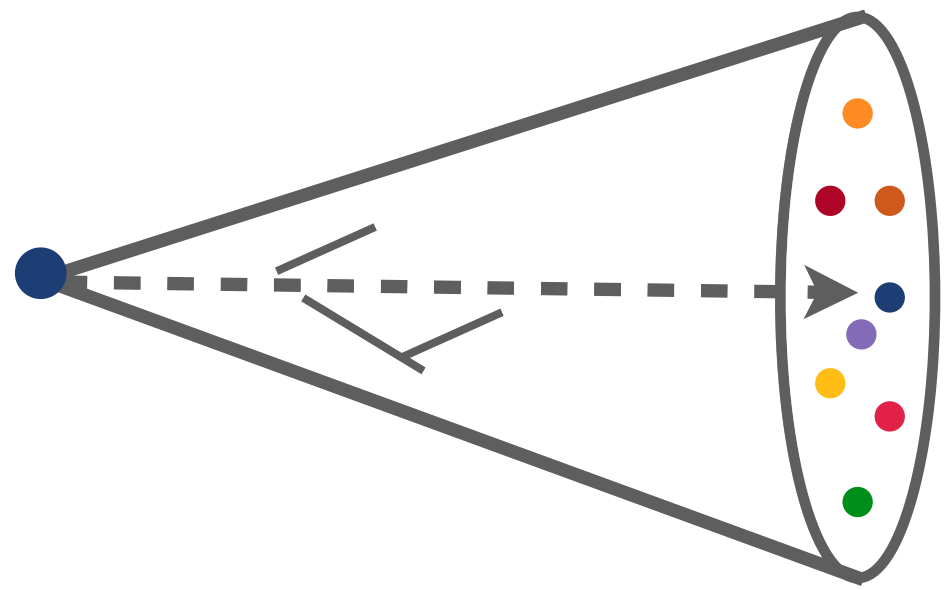
ML@Colliders : what's the broad task?

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.

$$f_{\{\theta\}} \Big(\quad \Big)$$



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

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

Variation in data

Unsupervised

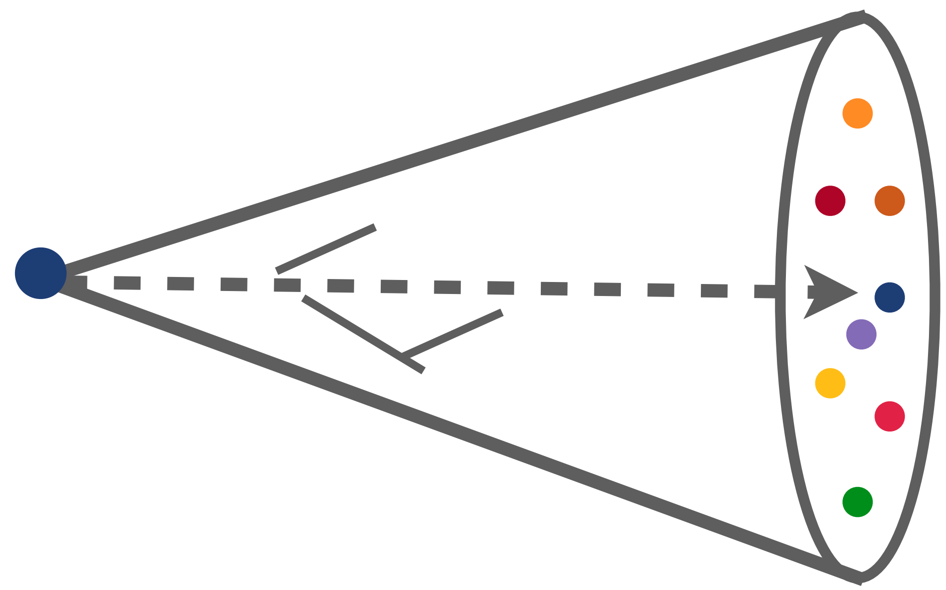
No-labels, the task is to figure out p(x) from which the data is drawn. e.g. VAE

ML@Colliders : 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.



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

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

Variation in data

Unsupervised **Semi-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)

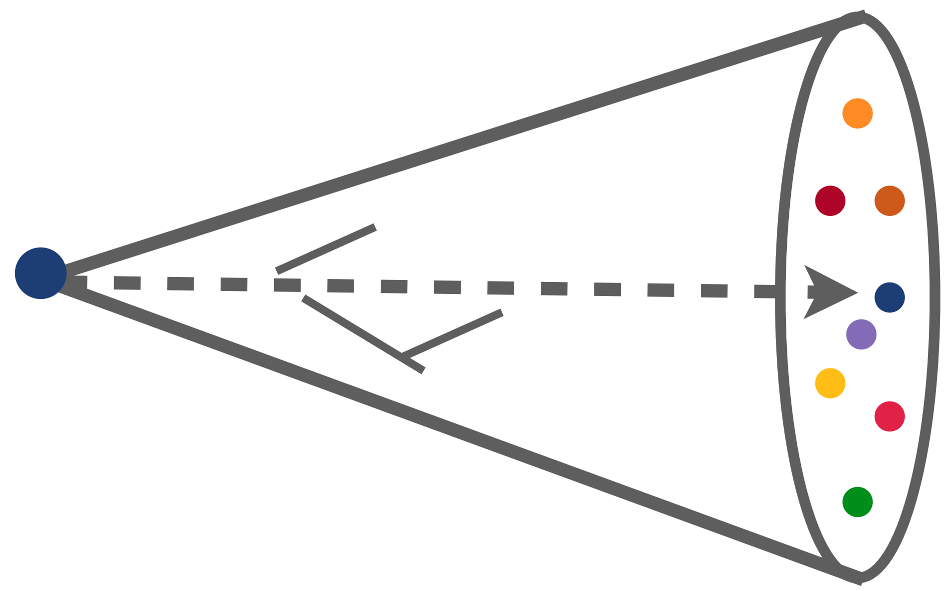
ML@Colliders : what's the broad task?

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.

$f_{\{\theta\}}$ (



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

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

Variation in data

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

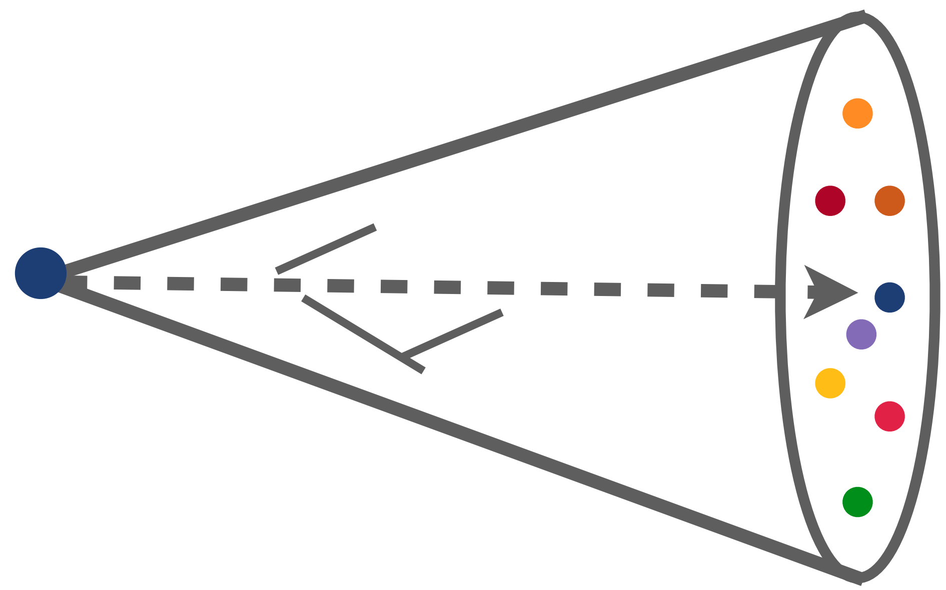
Partial labels. e.g. simulating : SM bkg vs many NP signals.

ML@Colliders : 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.

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

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

Variation in data

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.

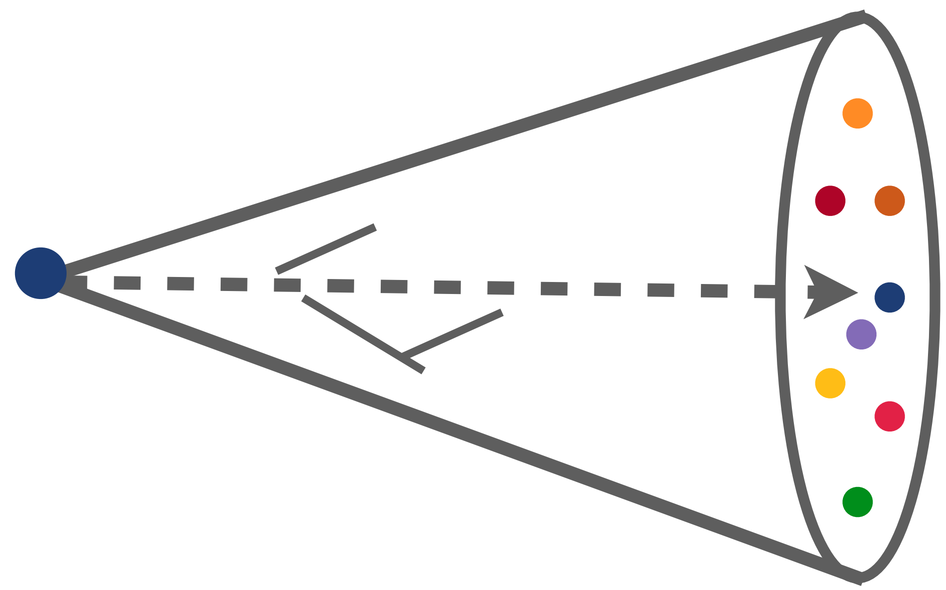
ML@Colliders : what's the broad task?

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.

$$f_{\{\theta\}} \Bigg($$



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

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

Self-supervised

Variation in data

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.

Looking the problem through ML lens

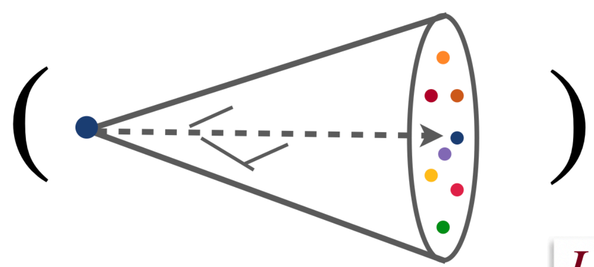
2. Choose a NN model (CNN/GNN/)

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

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

Uncertainty

Interpretability

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

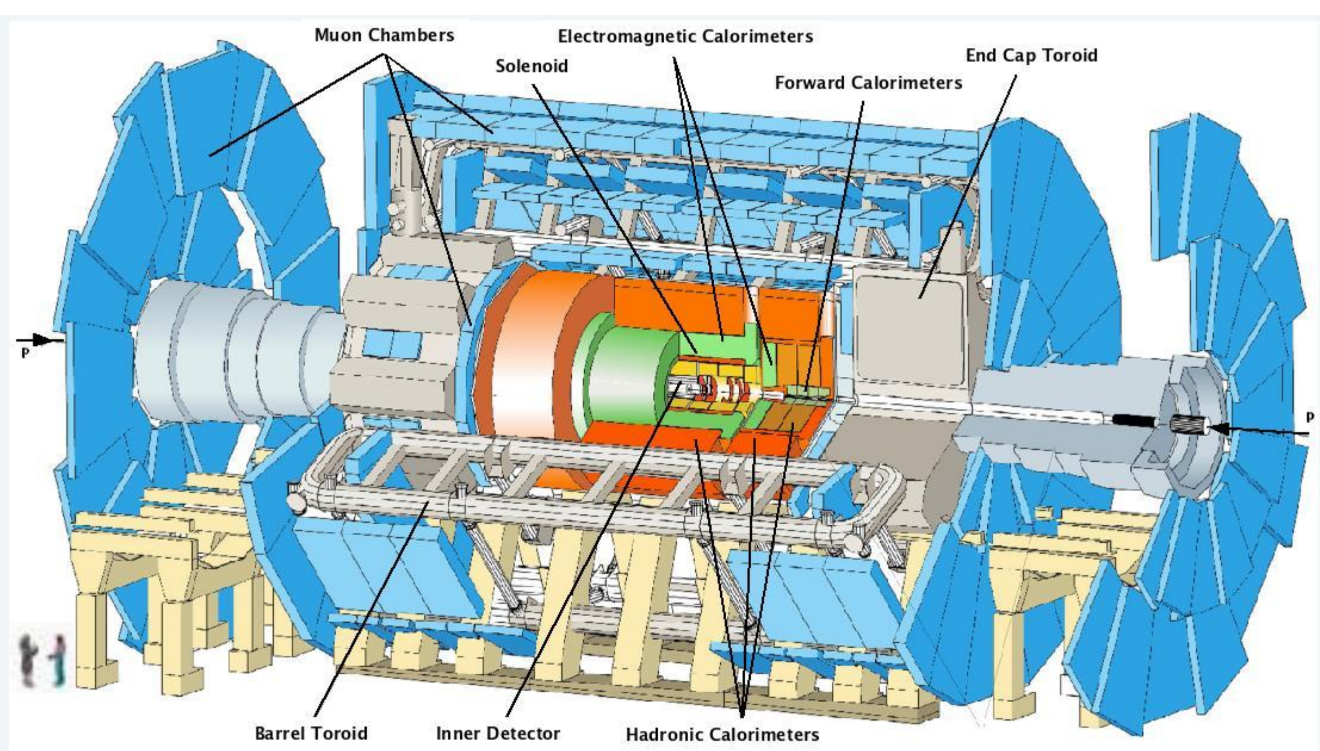
$$L(y, \hat{y}) \equiv L(\{\theta\})$$

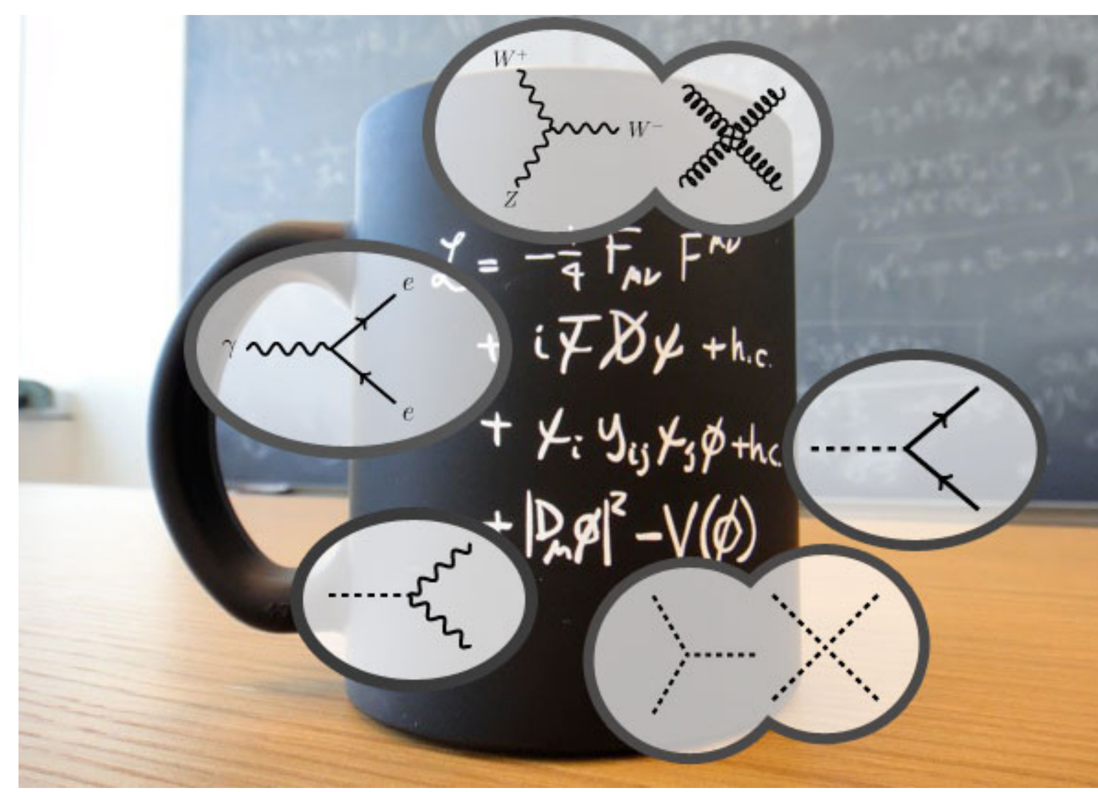
Self-supervised

Variation in data

- **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.

Figure from : https://atlas-public.web.cern.ch/





The path we will take

✓

Basic Motivation

- Why are we doing ML?
- The basic LHC operations with ML.
- Synergies through eye of ML.

The supervised ML

- Jet tagging : the most successful venture.
- Tracking & Calorimetery
- The Particle-Flow algorithm

Generative models

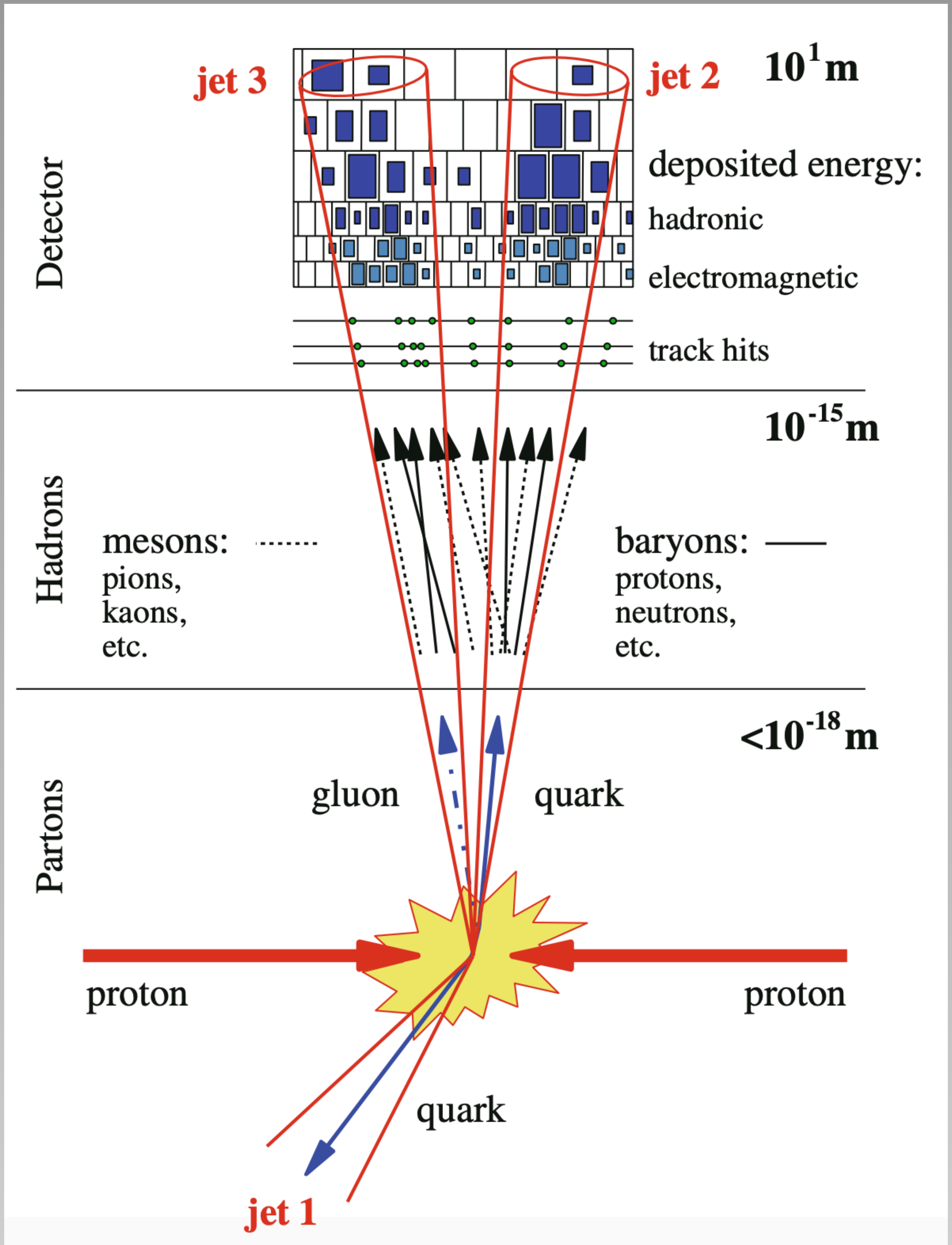
- Different paradigms
- Few physics cases.
- Interpretable ML & uncertainty estimation.

The anomaly detection saga

- The outliers & density estimate
- How to translate this in new signals?

THE ROAD MAP

Jet tagging (where it all started)



The Large Hadron Collider
Editors: Thomas Schörner-Sadenius

$$\{p_1, p_2, \ldots, p_n\}$$

Jet Algorithm (for CA, kT, anti-kT)

$$\{j_1, j_2, \ldots, j_k\}$$

$$\{p_1, p_2, \ldots, p_n\} = F(q)$$

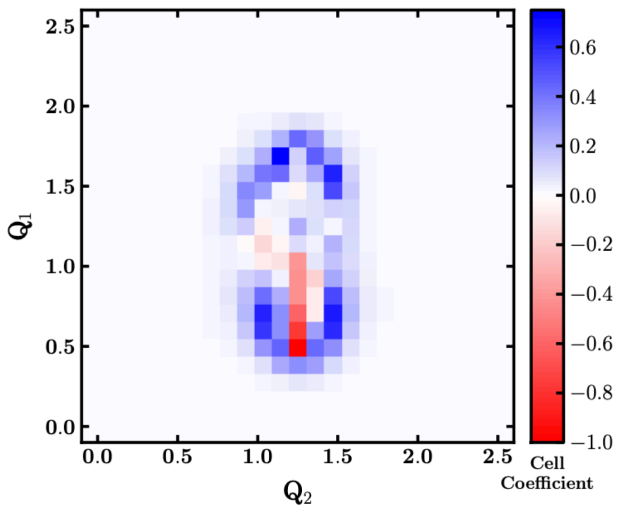
The forward problem is not computable from first principle

The question of jet tagging is how do we define the inverse problem?

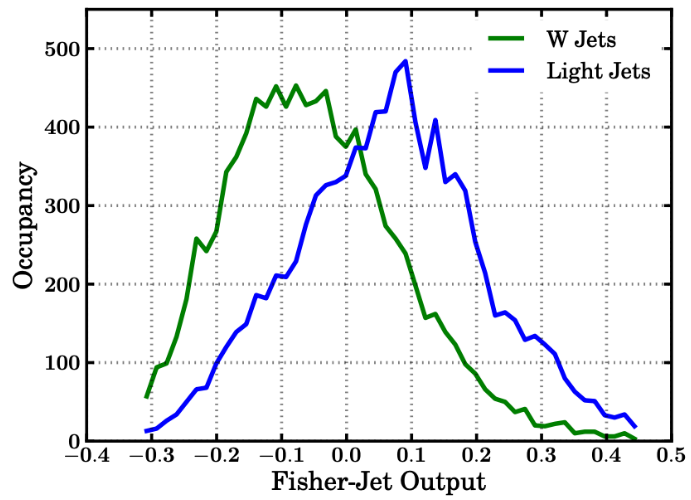
$$q = F^{-1}\Big(\{p_1, p_2, \ldots, p_n\}\Big) \ ?$$

Early jet tagging

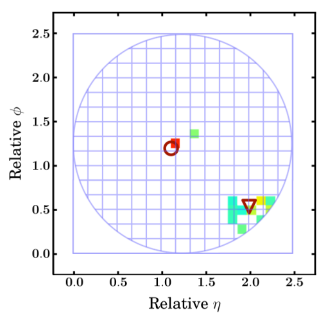
J. Cogan et-al JHEP 02 (2015) 118



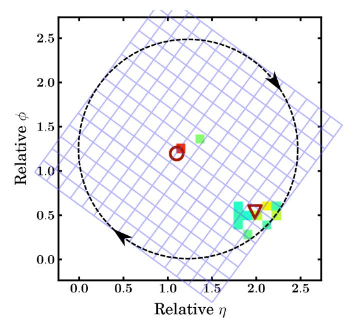
(a) Fisher-Jet



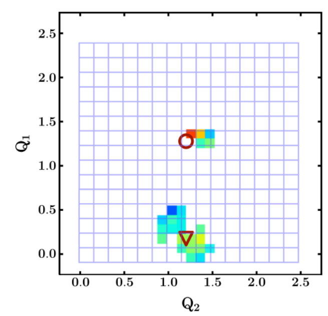
(b) Fisher-Jet Discriminant Output



(a) Jet-image prior to rotation

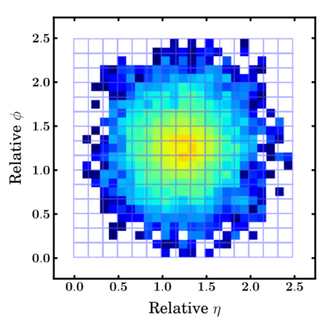


(b) Rotated pixel grid

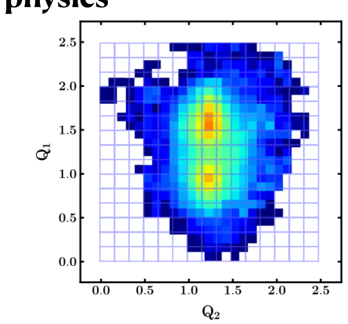


(c) Jet-image after projectio onto rotated grid, before transla tion

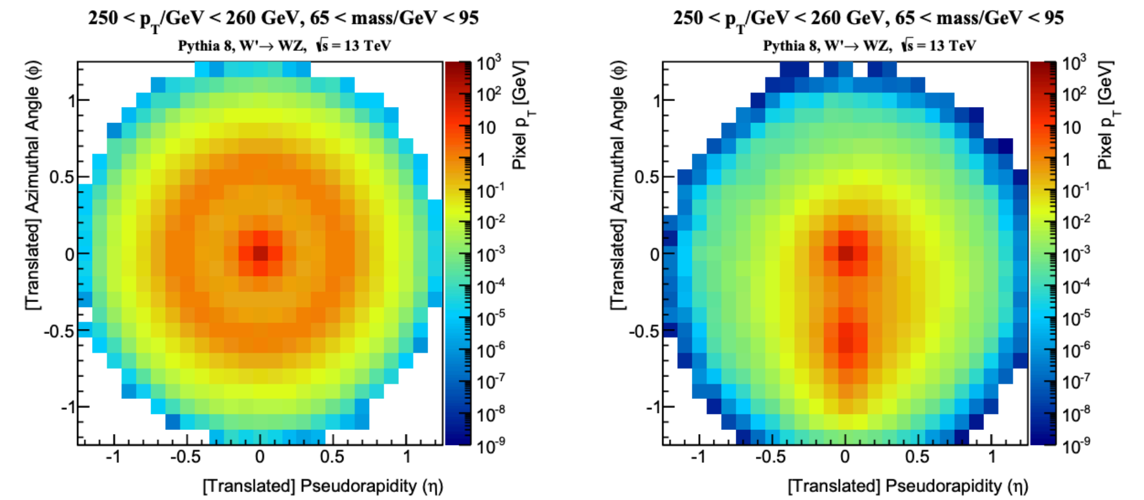
The first paper to discuss image pre-processing for jet physics



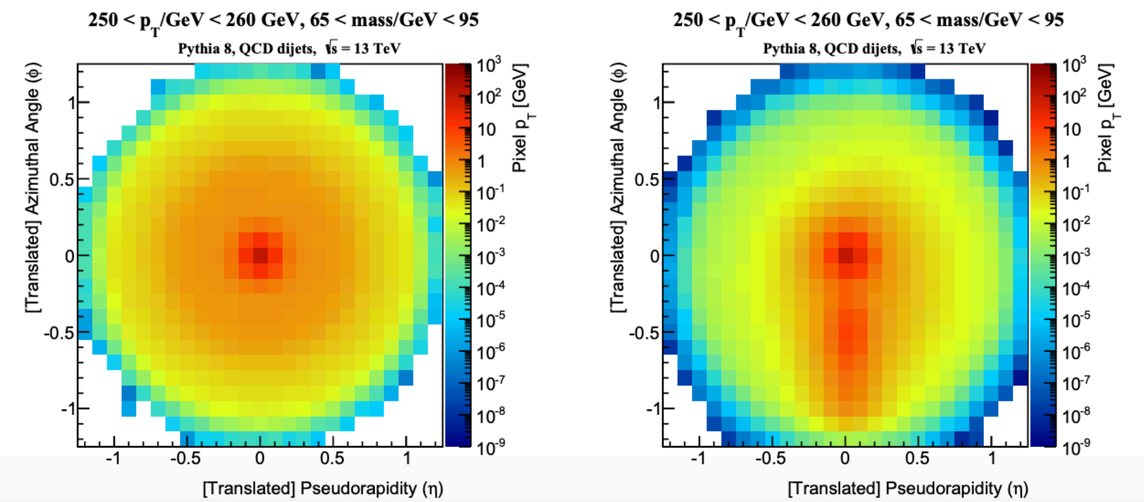
(d) Average jet-image, prior to rotation

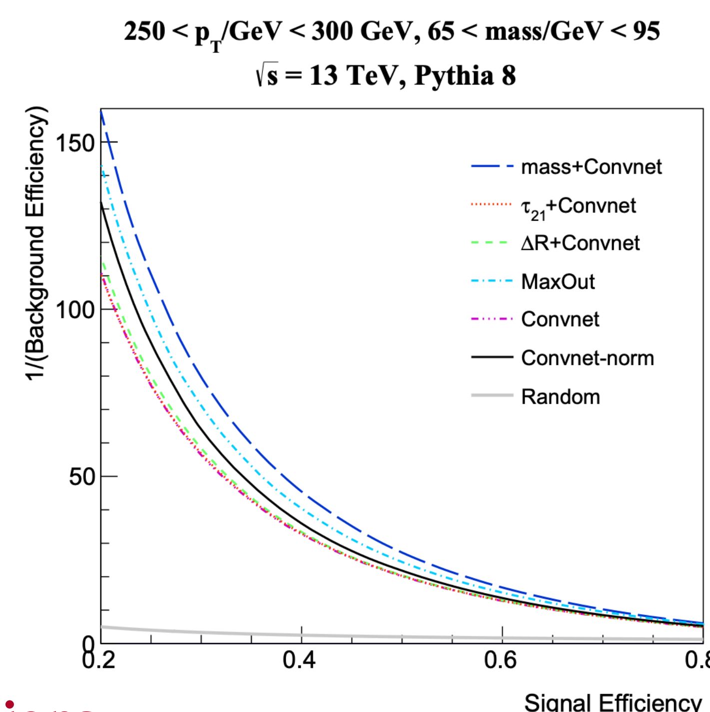


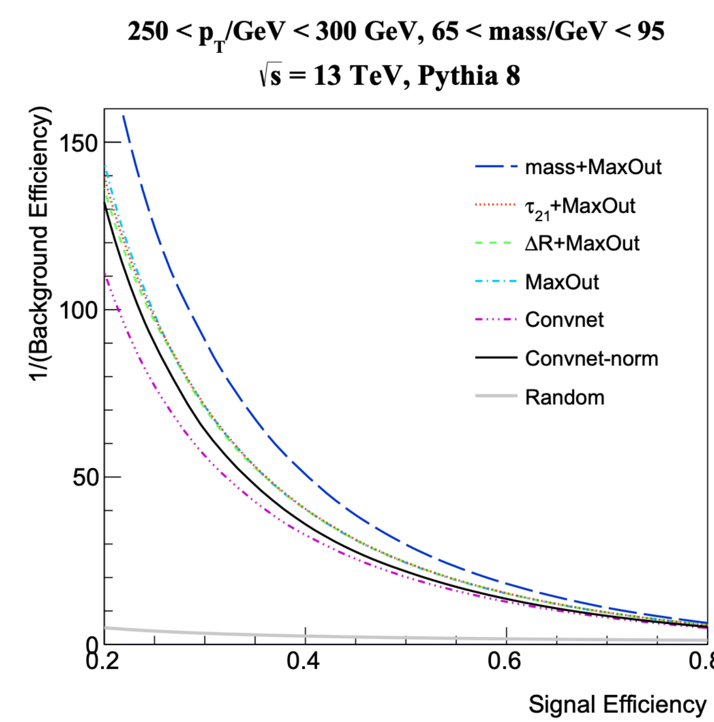
(e) Average jet-image, after pre processing



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







Similar methods were applied for particle identifications

Different possible representations

The energy deposition pattern can be thought as multi-layer image.

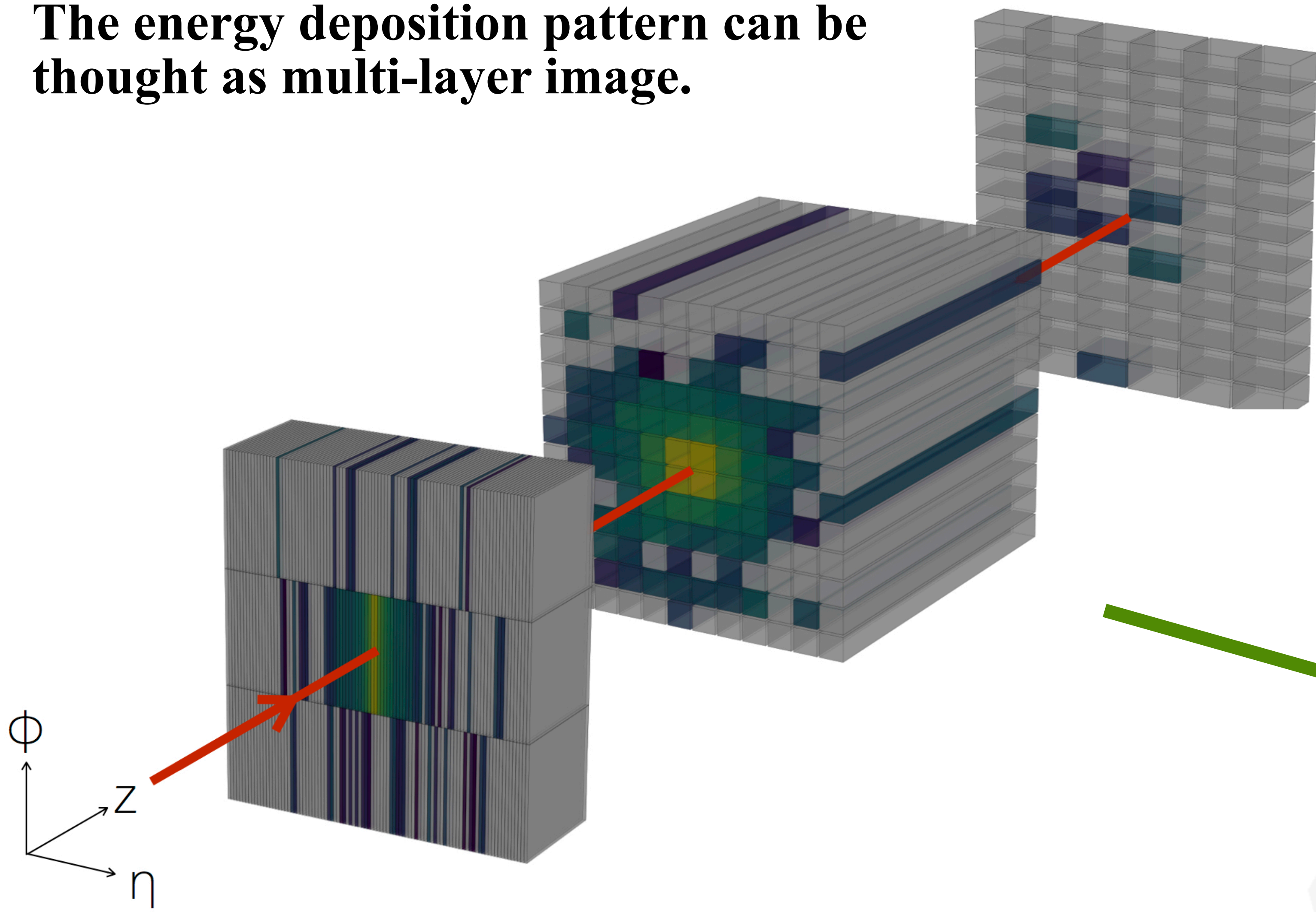
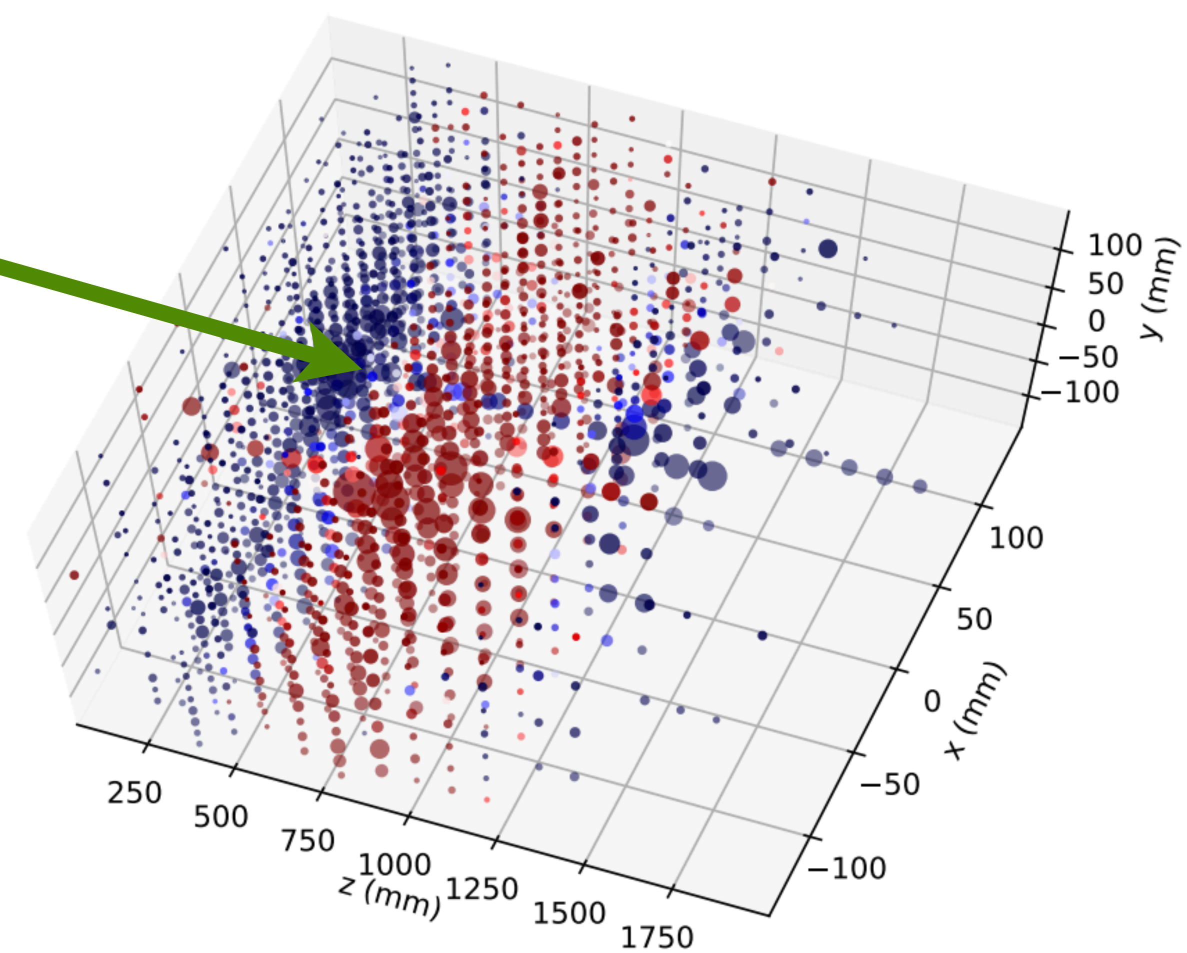


Image from 1705.02355

S. Qasim et-al EPJC 79 7 (2019) 608

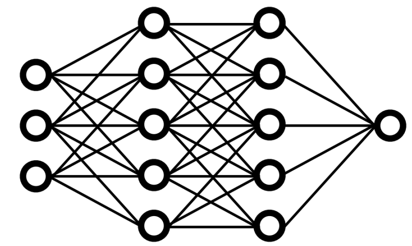


(a) Truth

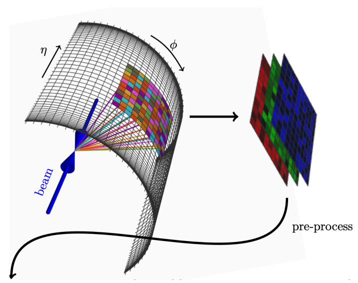
Soon point-cloud based methods became more popular.
Large sparsity in calorimeter shower were more efficient to treat.

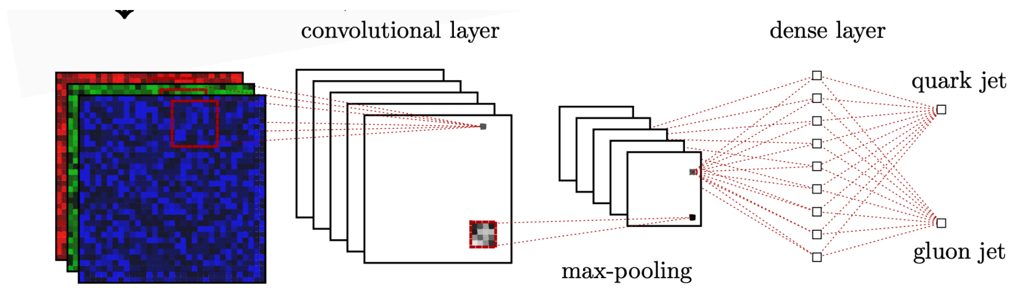
Data representation ⇔ NN correspondence

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

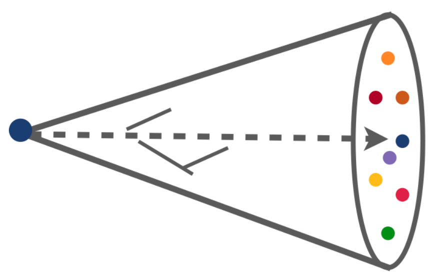


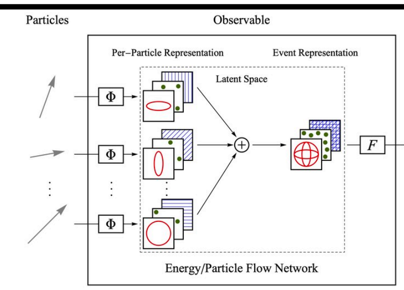
Ordered set
DNN



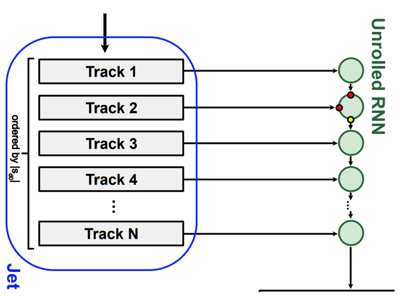


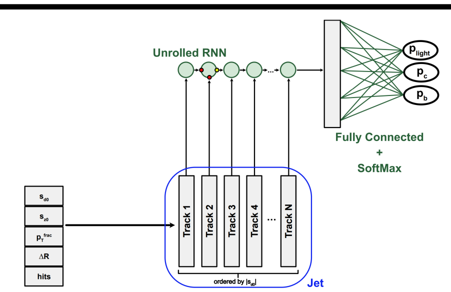
Grid
CNN



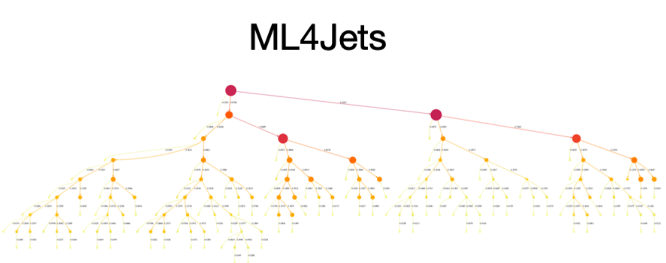


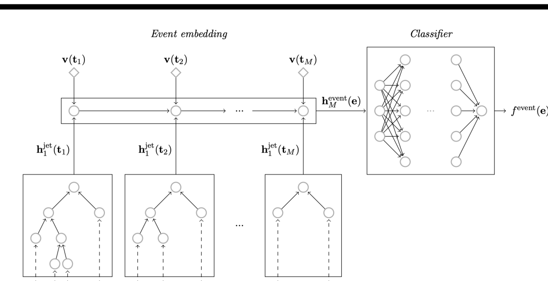
Unordered set
Deepest



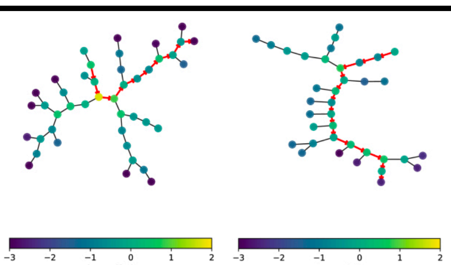


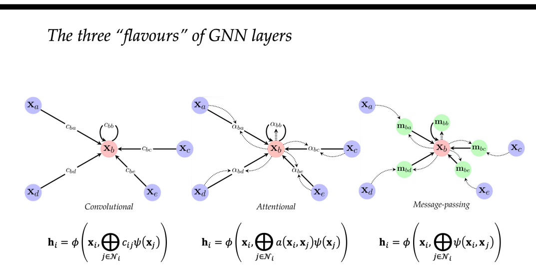
Sequential data
RNN





Tree structure
Deepset/GNN



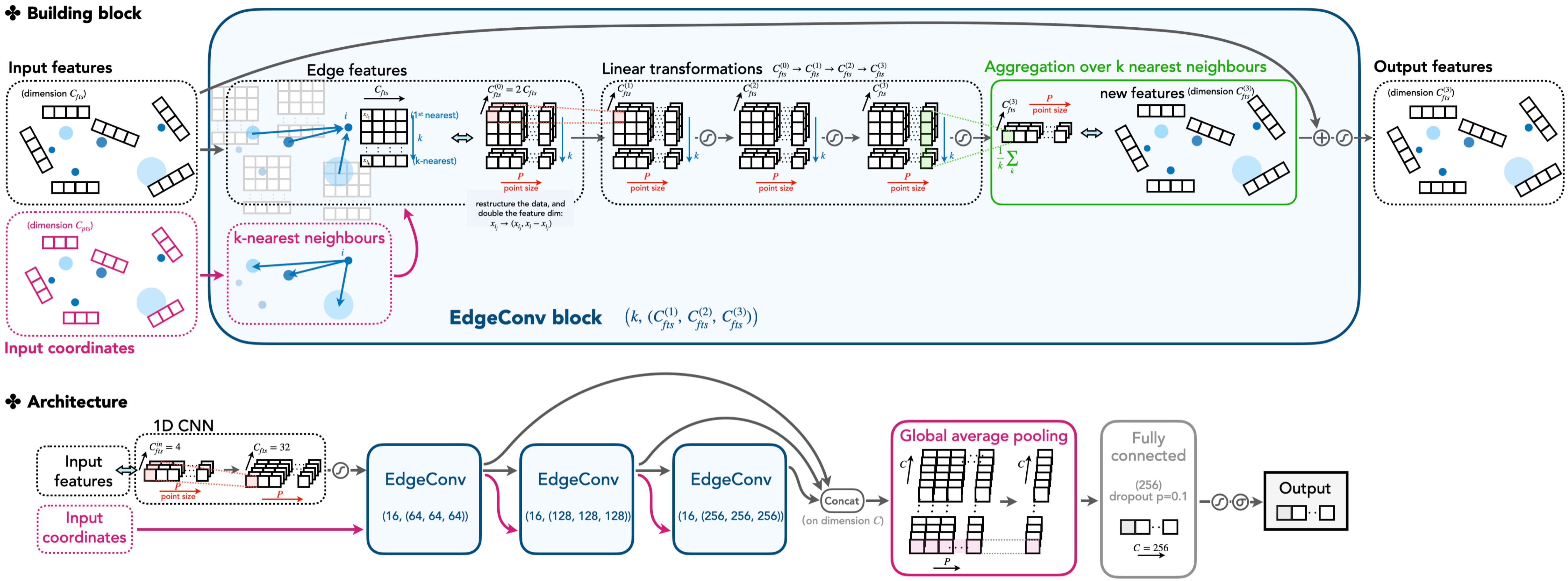


Graph
GNN

Object tagging

$$e'_{ijm} = \mathrm{ReLU}(\boldsymbol{\theta}_m \cdot (\mathbf{x}_j - \mathbf{x}_i) + \boldsymbol{\phi}_m \cdot \mathbf{x}_i),$$

Particle Net : 1902.08570
Huilin Qu, Loukas Gouskos



arXiv > cs > arXiv:1801.07829

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 24 Jan 2018 (v1), last revised 11 Jun 2019 (this version, v2)]

Dynamic Graph CNN for Learning on Point Clouds

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, Justin M. Solomon

Machine learning on sets

Let $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N} \in \mathbb{R}^k$ be n pieces of data. This forms a set of cardinality N.

https://geometricdeeplearning.com/lectures/

Machine learning on sets

Let $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N} \in \mathbb{R}^k$ be n pieces of data. This forms a set of cardinality N.

Neural network on a set

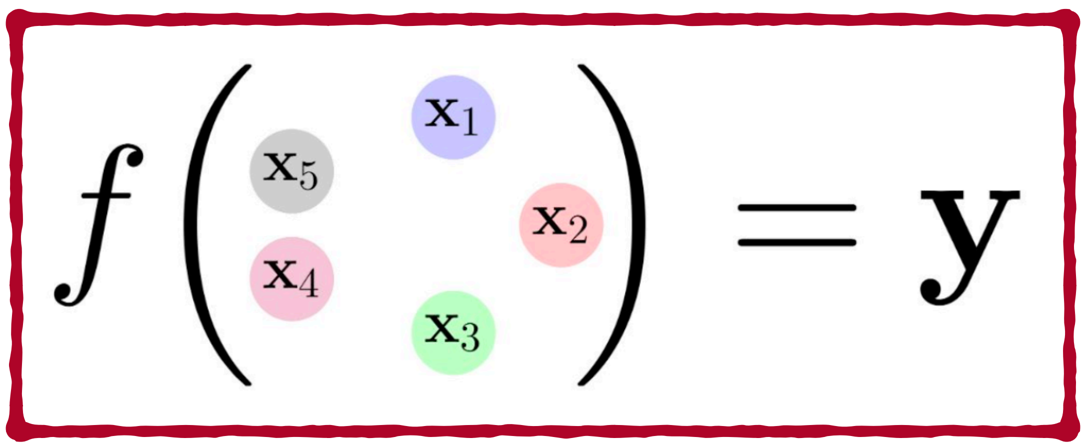
https://geometricdeeplearning.com/lectures/

Machine learning on sets

Let $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N} \in \mathbb{R}^k$ be n pieces of data. This forms a set of cardinality N.

Neural network on a set

https://geometricdeeplearning.com/lectures/

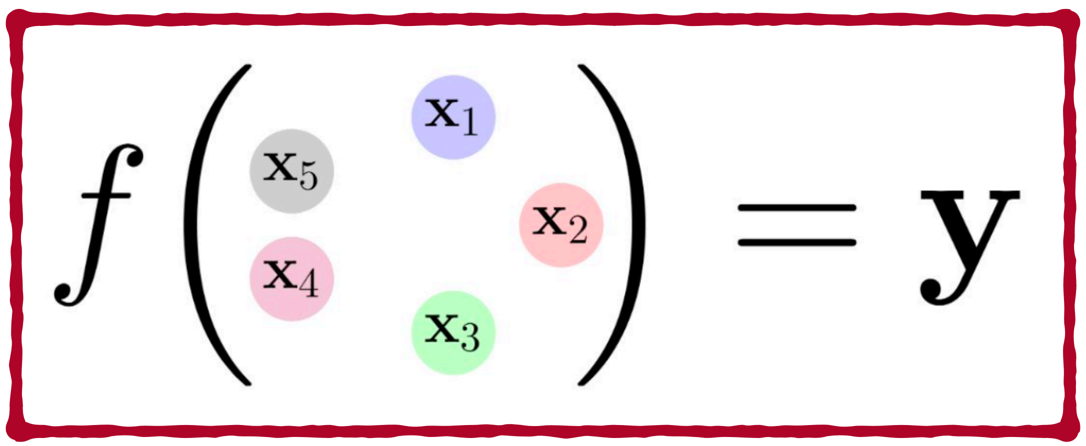


Machine learning on sets

Let $\mathbf{x_1, x_2, \ldots, x_N} \in \mathbb{R}^k$ be n pieces of data. This forms a set of cardinality N.

Neural network on a set

https://geometricdeeplearning.com/lectures/



Basic required property : permutation invariance

Machine learning on sets

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in \mathbb{R}^k$ be n pieces of data. This forms a set of cardinality N.

Neural network on a set

https://geometricdeeplearning.com/lectures/

$$\boxed{f\left(\begin{matrix} & \mathbf{x}_1 & \\ \mathbf{x}_5 & & \\ & & \mathbf{x}_2 \\ \mathbf{x}_4 & & \\ & \mathbf{x}_3 & \end{matrix}\right) = \mathbf{y}}$$

Basic required property : permutation invariance

$$\boxed{f\left(\begin{matrix} & \mathbf{x}_1 & \\ \mathbf{x}_5 & & \\ & & \mathbf{x}_2 \\ \mathbf{x}_4 & & \\ & \mathbf{x}_3 & \end{matrix}\right) = \mathbf{y} = f\left(\begin{matrix} & \mathbf{x}_2 & \\ & \mathbf{x}_5 & \mathbf{x}_4 \\ \mathbf{x}_1 & & \mathbf{x}_3 \end{matrix}\right)}$$

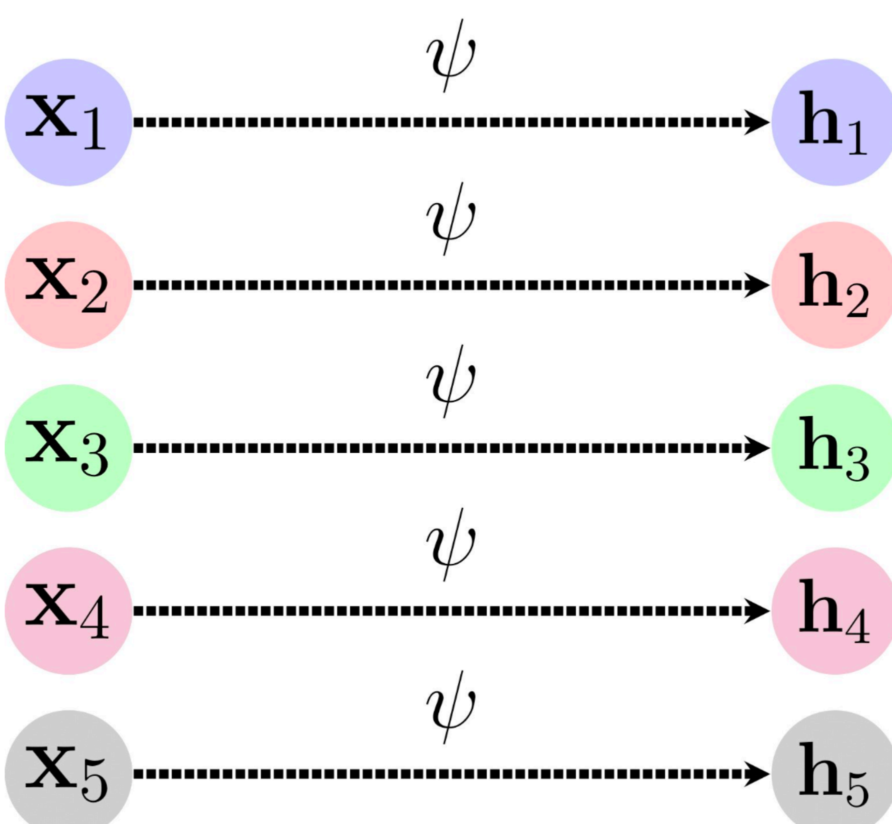
How the P.I. is achieved?

Remember the permutation on a set?

$$\mathbf{f}\left(\mathbf{PX}\right) = \mathbf{f}\left(\mathbf{X}\right)$$

$$\mathbf{P}_{(2,4,1,3)}\mathbf{X} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} — & \mathbf{x}_1 & — \\ — & \mathbf{x}_2 & — \\ — & \mathbf{x}_3 & — \\ — & \mathbf{x}_4 & — \end{bmatrix} = \begin{bmatrix} — & \mathbf{x}_2 & — \\ — & \mathbf{x}_4 & — \\ — & \mathbf{x}_1 & — \\ — & \mathbf{x}_3 & — \end{bmatrix}$$

$$\mathbf{h_i} = \psi(\mathbf{x_i})$$



$$f\left(X\right) = \phi\left(\bigoplus_{i \in V} \psi(X_i)\right)$$

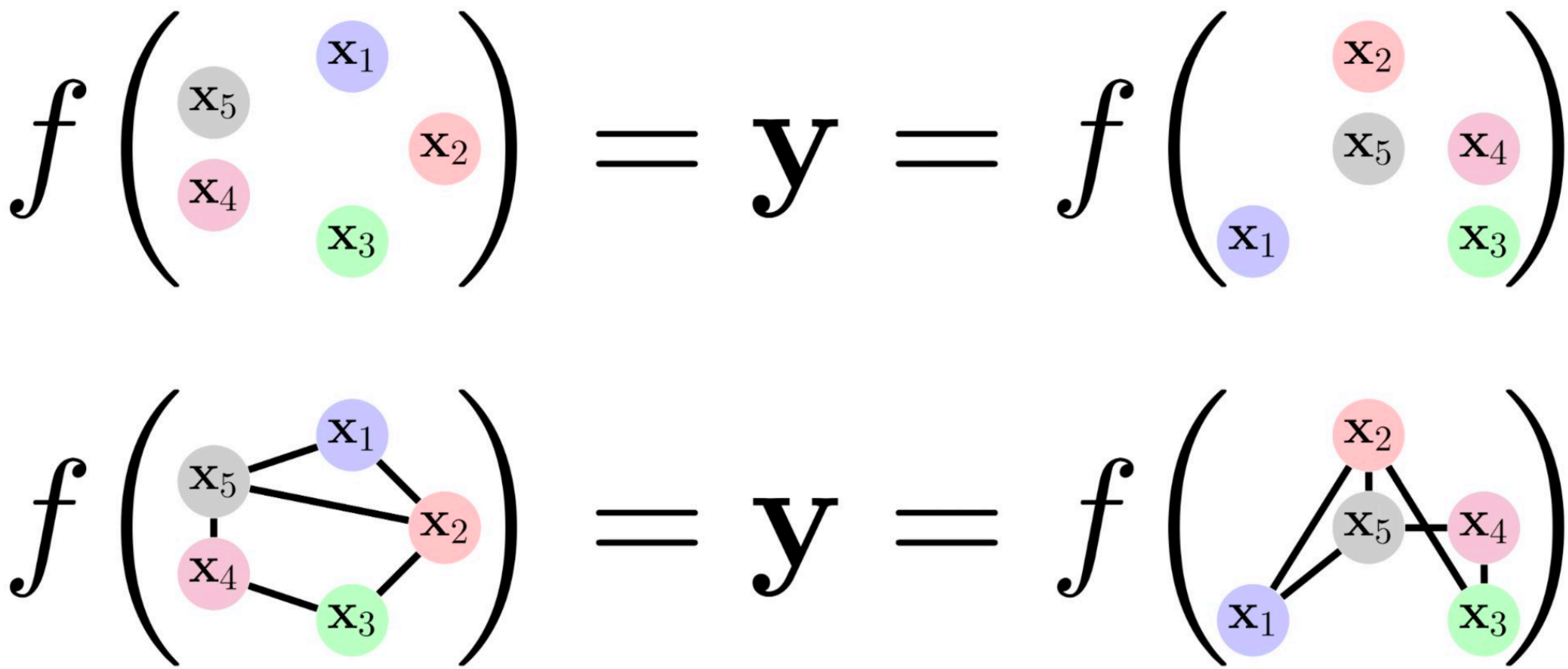
https://geometricdeeplearning.com/lectures/

The permutation equivariant operation :

$$\Theta = \lambda \mathbf{I} + \gamma \left(\mathbf{1}\mathbf{1}^\top\right) \text{ for } \lambda, \gamma \in \mathbb{R}.$$

What's the basic criteria of a GNN?

https://geometricdeeplearning.com/lectures/

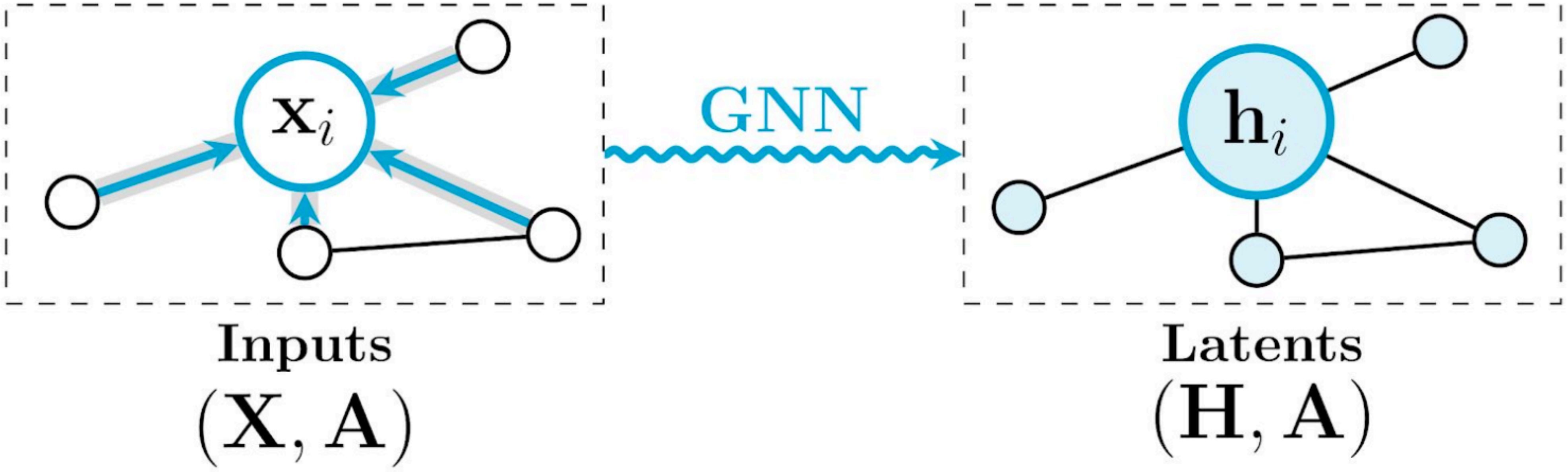


Invariance: $f(\mathbf{PX}, \mathbf{PAP}^\top) = f(\mathbf{X}, \mathbf{A})$

Equivariance: $f(\mathbf{PX}, \mathbf{PAP}^\top) = \mathbf{P}f(\mathbf{X}, \mathbf{A})$

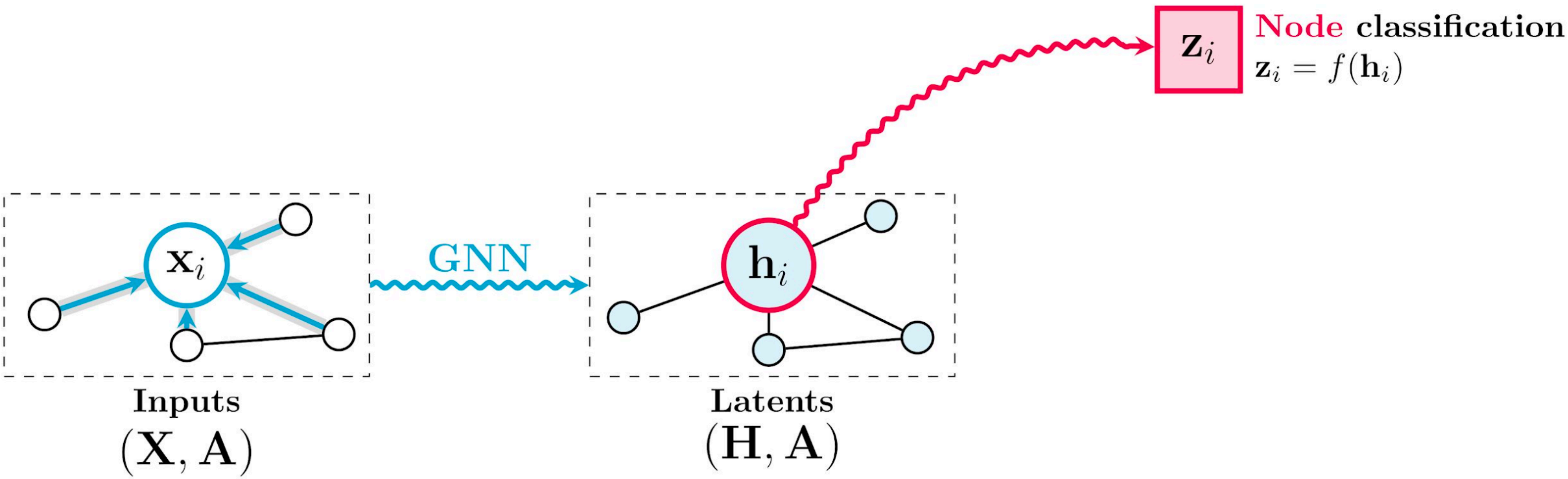
General methods of GNN

General methods of GNN



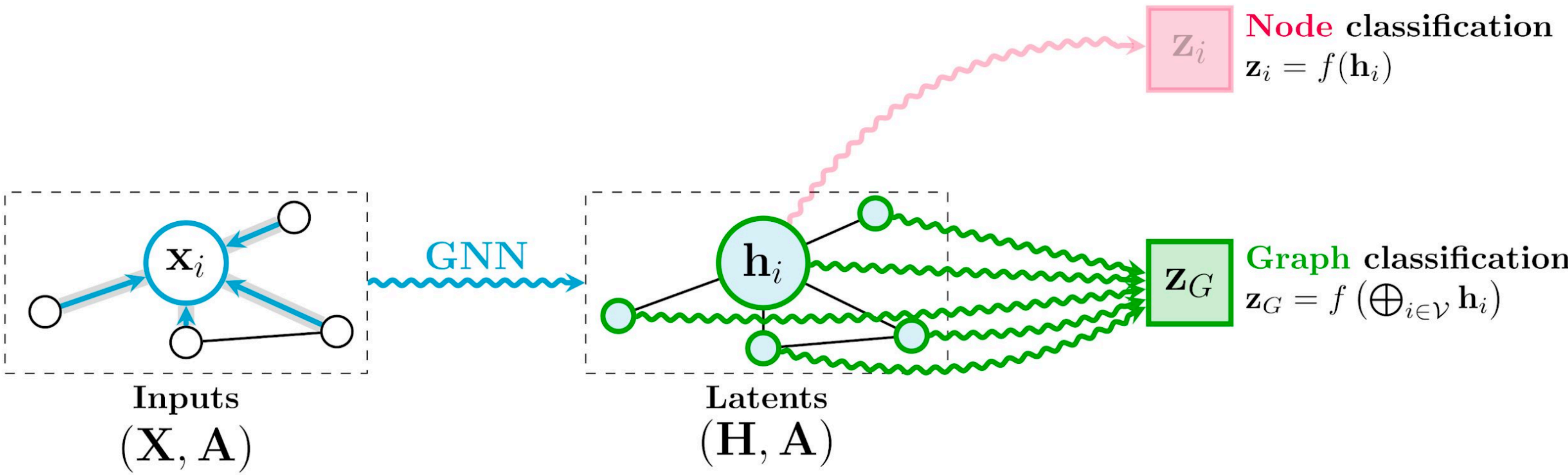
General methods of GNN

General methods of GNN



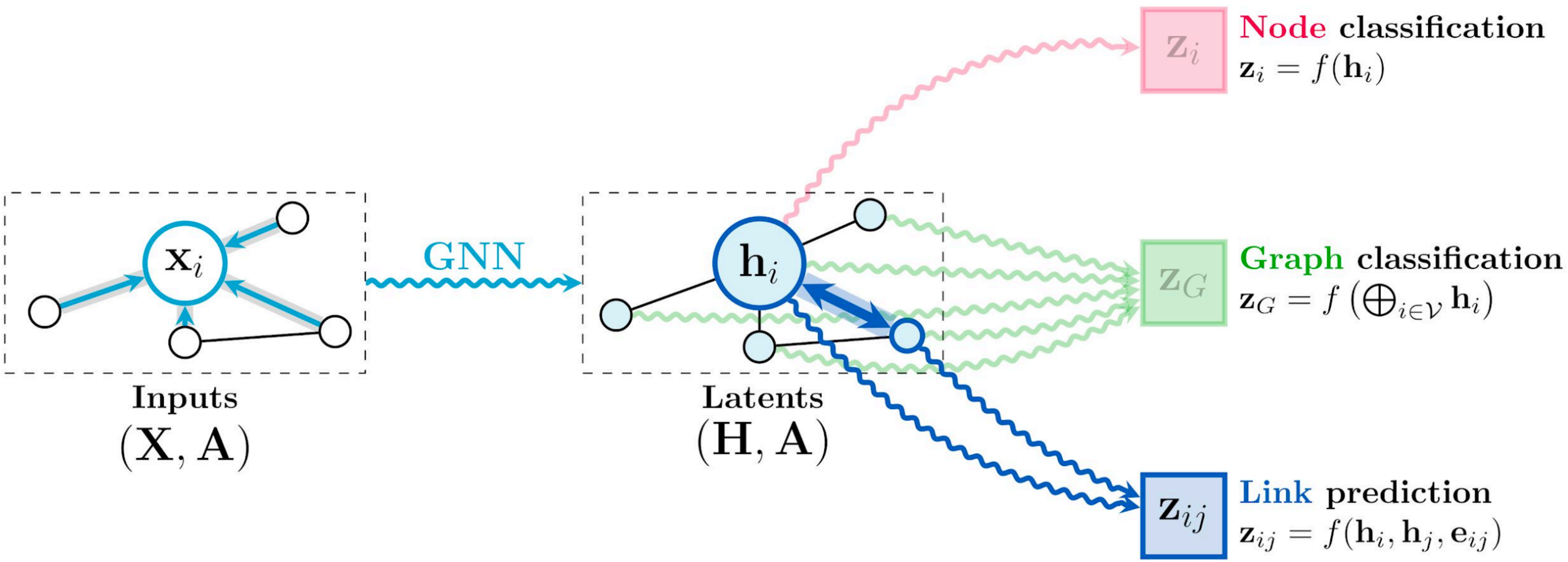
General methods of GNN

General methods of GNN



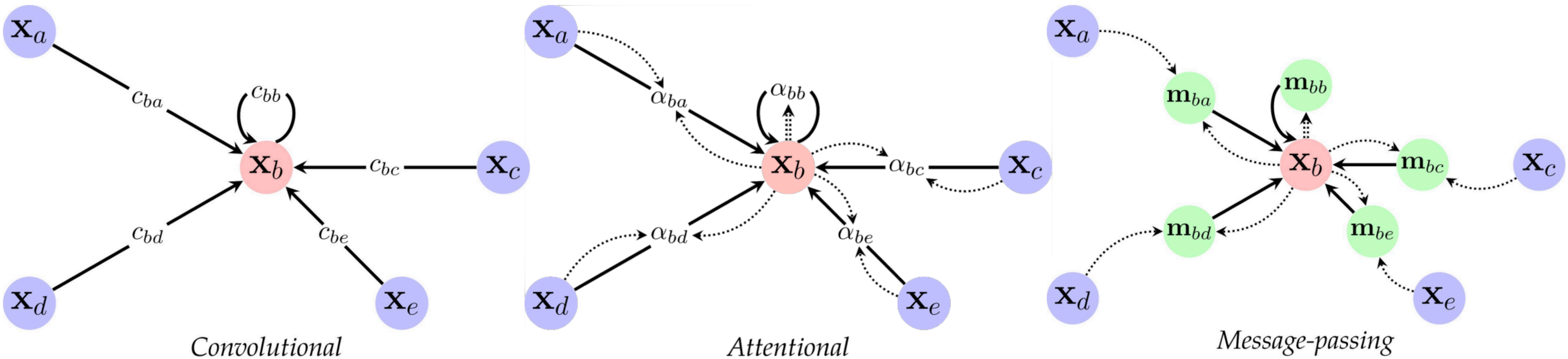
General methods of GNN

General methods of GNN



The different flavors of MPN

The three "flavours" of GNN layers



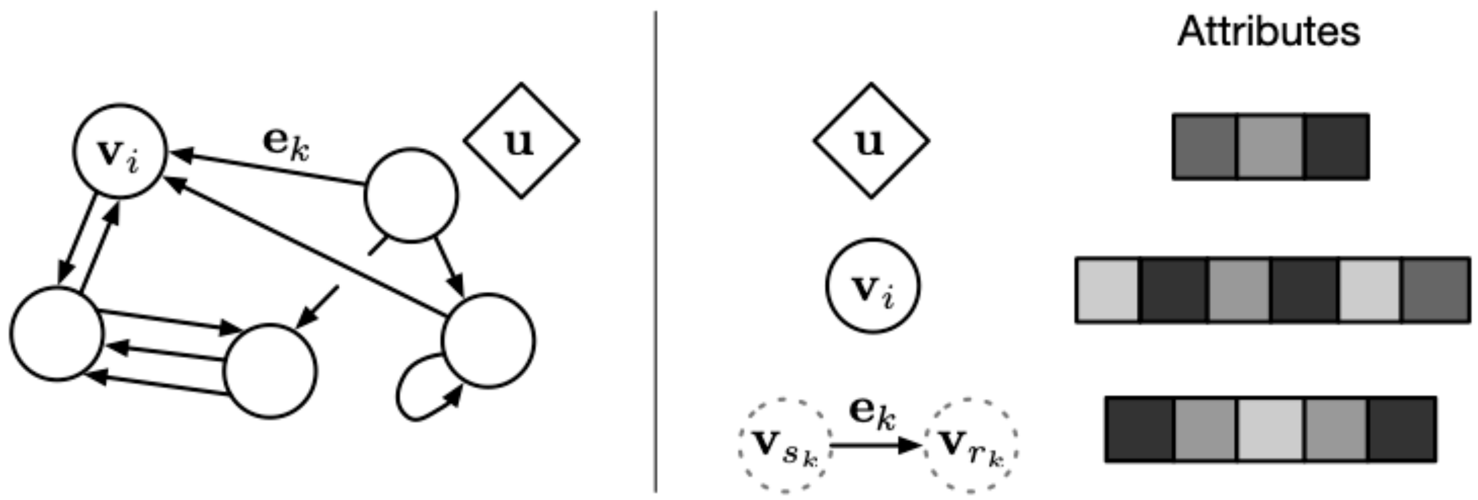
$$\mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j\in\mathcal{N}_i} c_{ij}\psi(\mathbf{x}_j)\right)$$

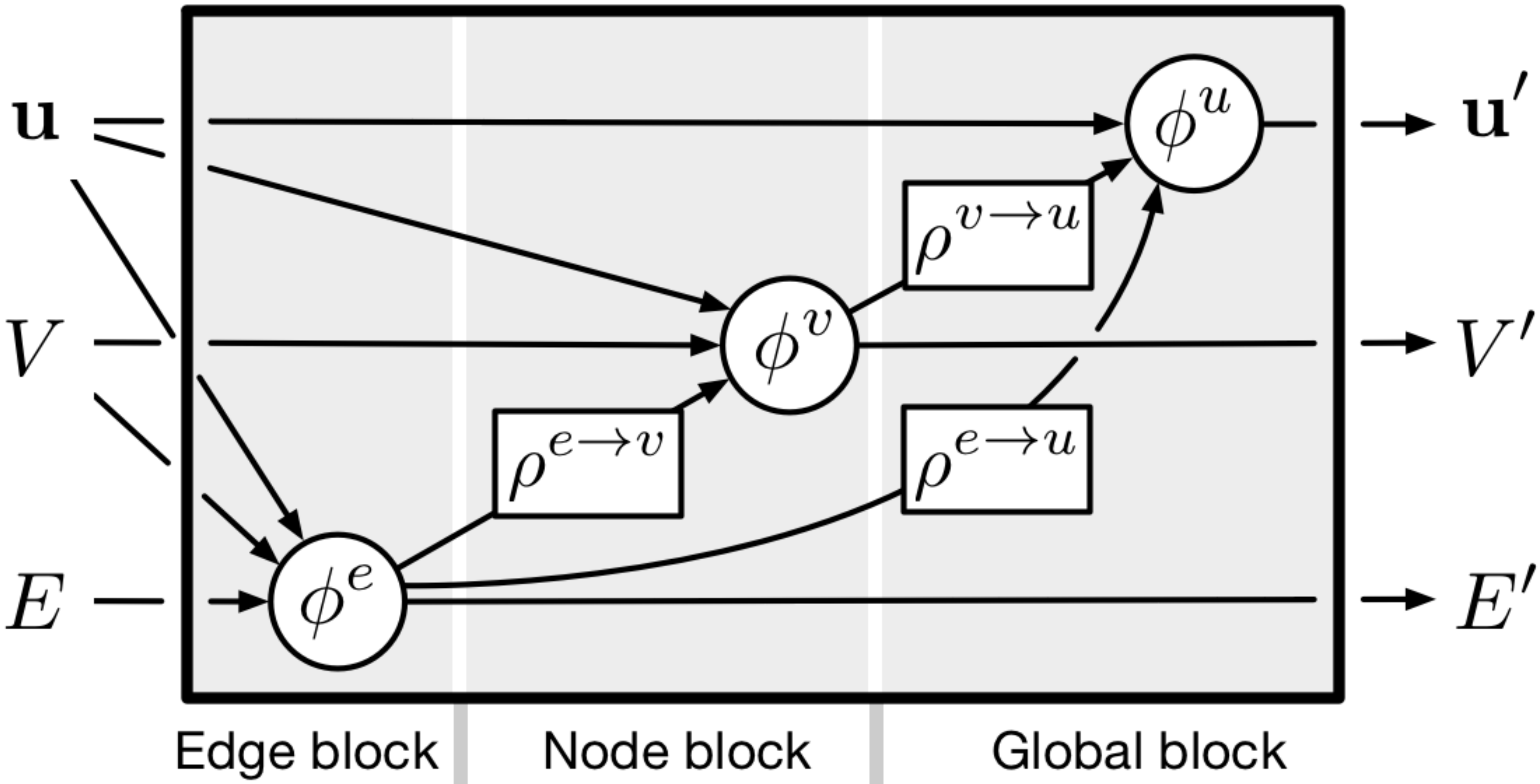
$$\mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j\in\mathcal{N}_i} a(\mathbf{x}_i, \mathbf{x}_j)\psi(\mathbf{x}_j)\right)$$

$$\mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j\in\mathcal{N}_i} \psi(\mathbf{x}_i, \mathbf{x}_j)\right)$$

The general GNN

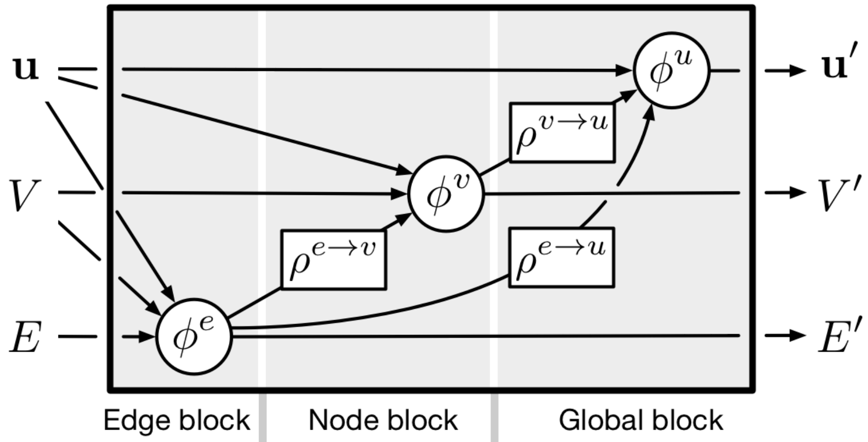
arXiv : 1806.01261



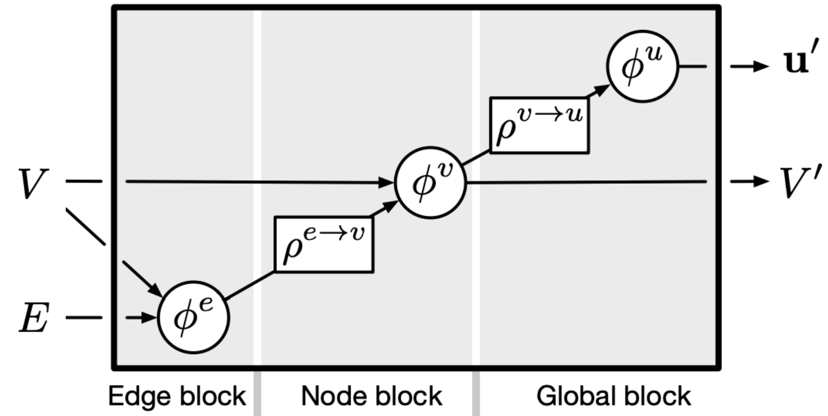


The general GNN

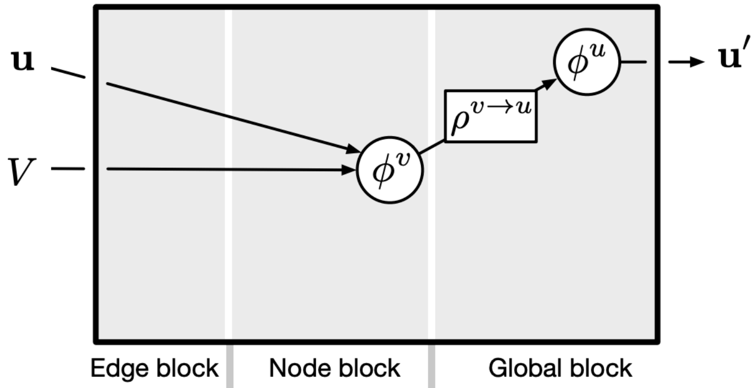
arXiv : 1806.01261



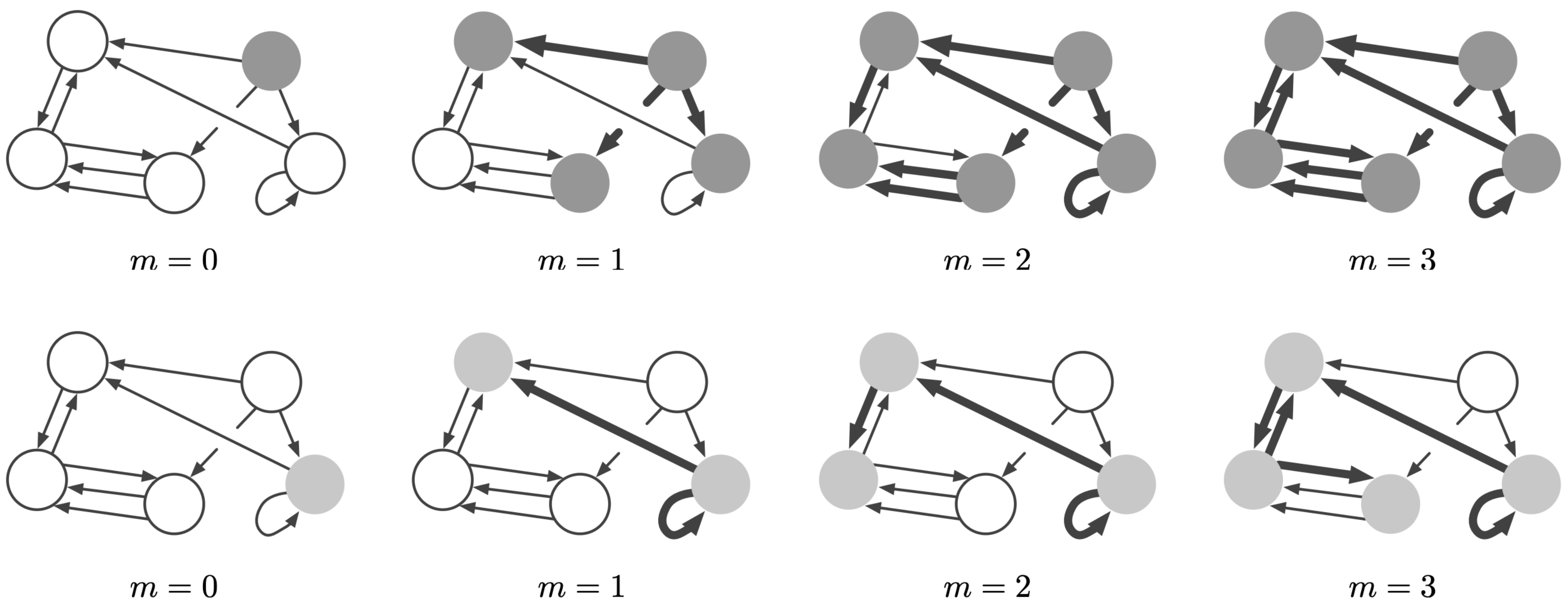
Full GN block



MPNN Layer



Deep-set layer



Application of sets & graphs in HEP

arXiv > hep-ex > arXiv:2203.12852

Search...

Help | Advar

High Energy Physics – Experiment

[Submitted on 23 Mar 2022 (v1), last revised 25 Mar 2022 (this version, v2)]

Graph Neural Networks in Particle Physics: Implementations, Innovations, and Challenges

Savannah Thais, Paolo Calafiura, Grigorios Chachamis, Gage DeZoort, Javier Duarte, Sanmay Ganguly, Michael Kagan, Daniel Murnane, Mark S. Neubauer, Kazuhiro Terao

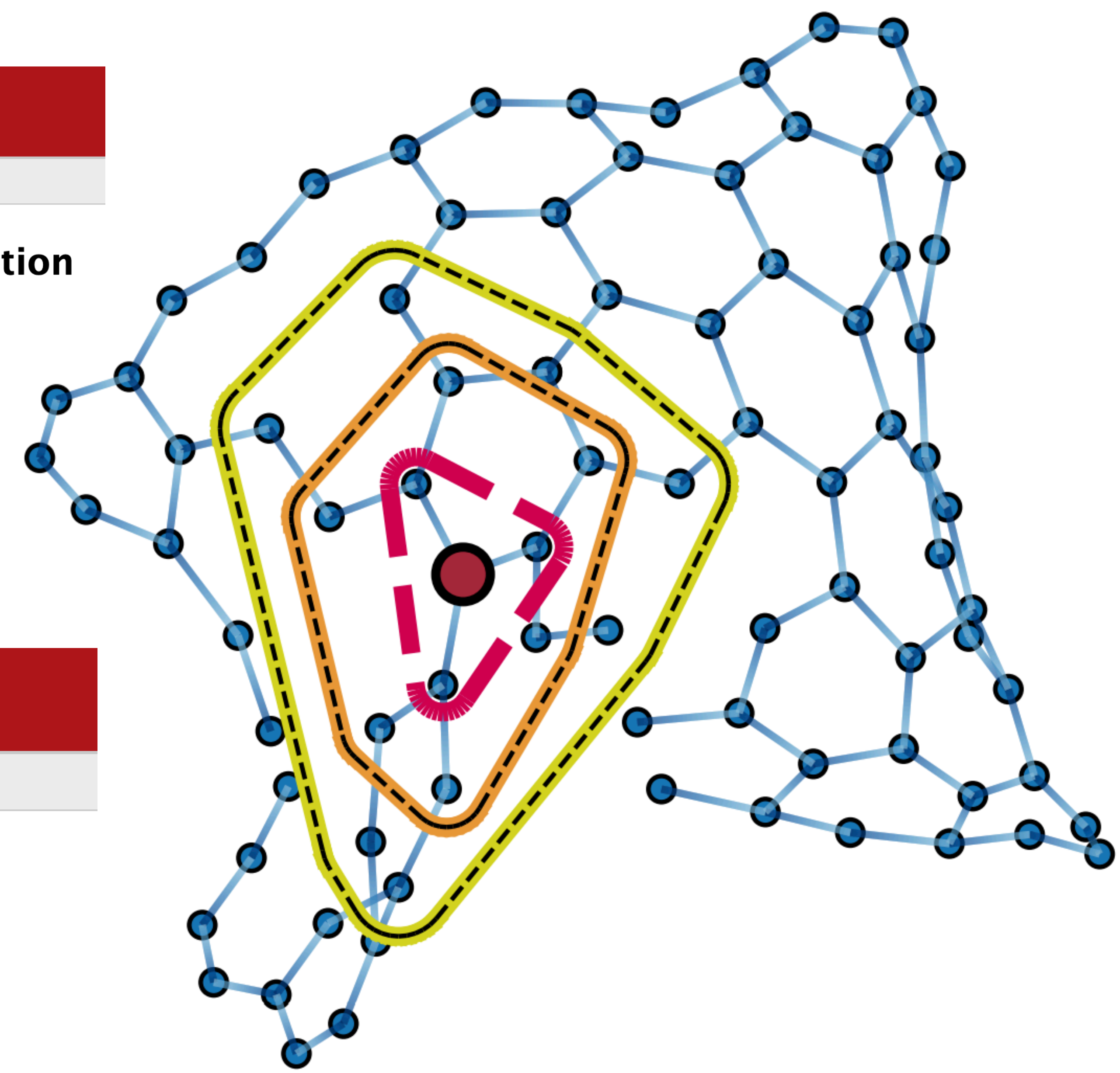
arXiv > hep-ph > arXiv:2012.01249

High Energy Physics – Phenomenology

[Submitted on 2 Dec 2020 (v1), last revised 7 Dec 2020 (this version, v2)]

Graph Neural Networks for Particle Tracking and Reconstruction

Javier Duarte, Jean-Roch Vlimant



arXiv > hep-ex > arXiv:2007.13681

High Energy Physics – Experiment

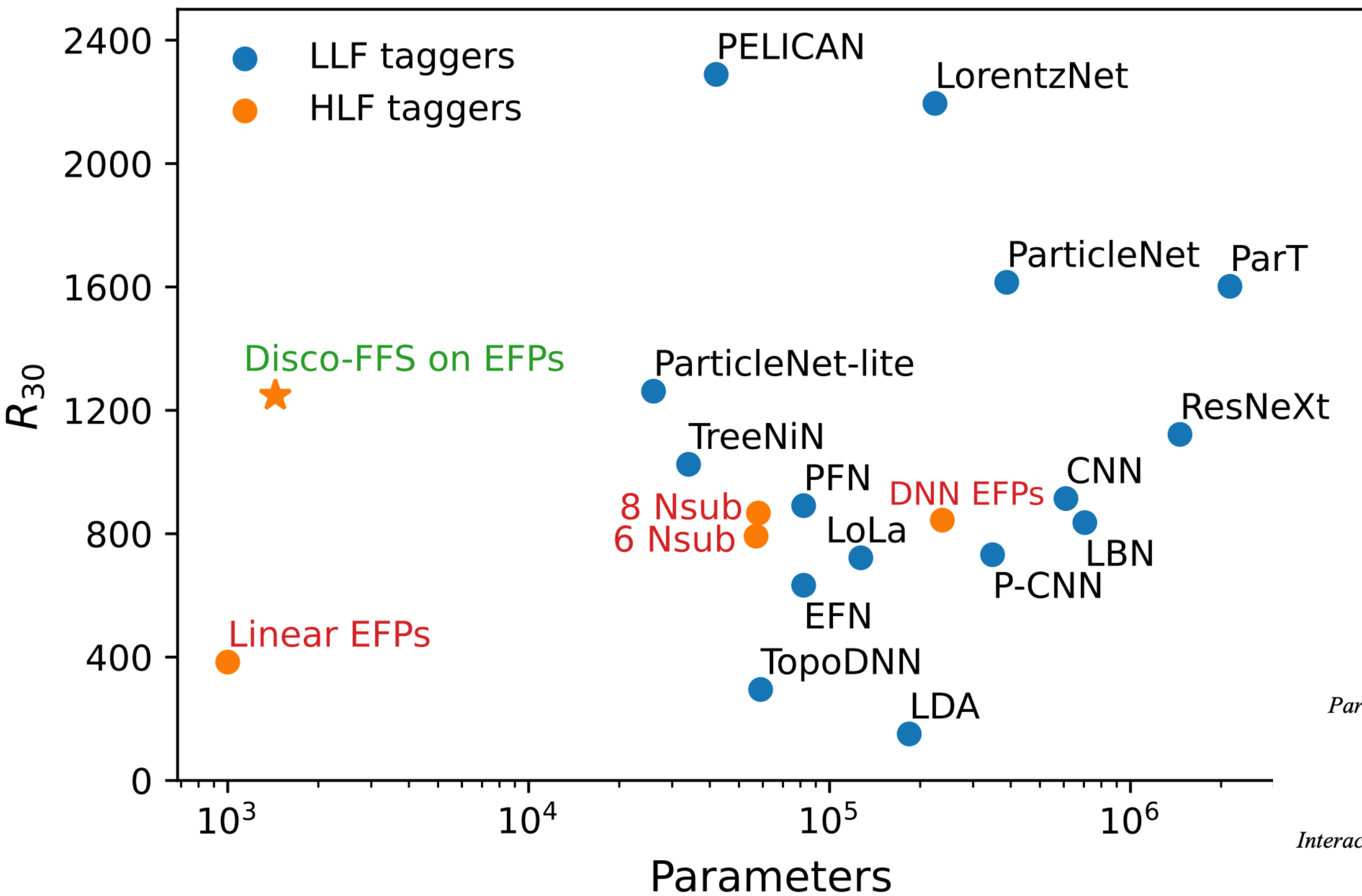
[Submitted on 27 Jul 2020 (v1), last revised 21 Oct 2020 (this version, v2)]

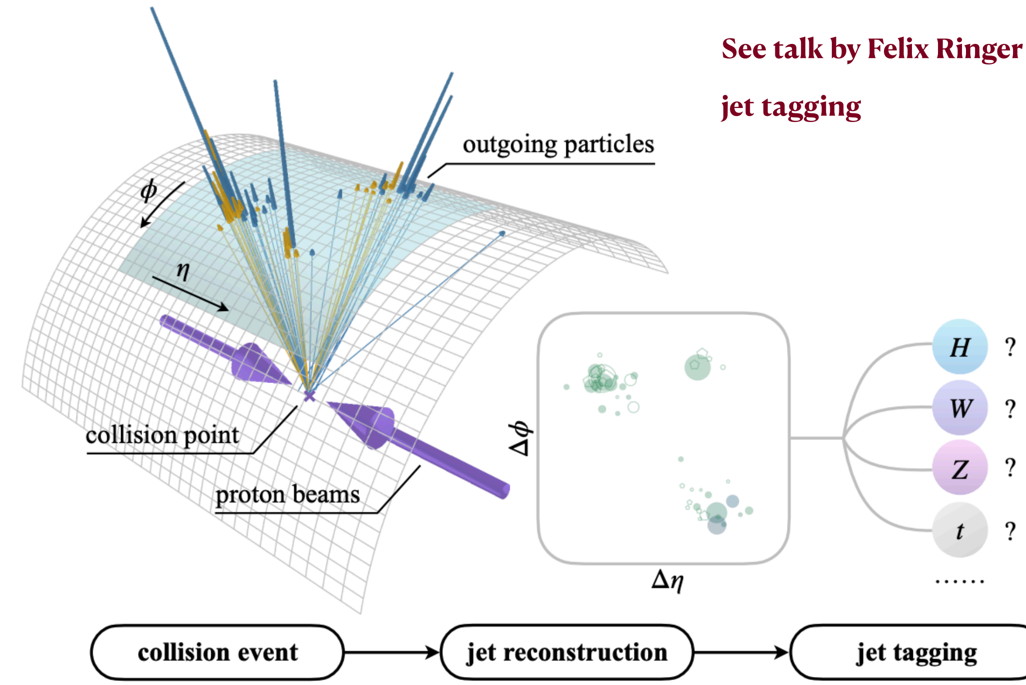
Graph Neural Networks in Particle Physics

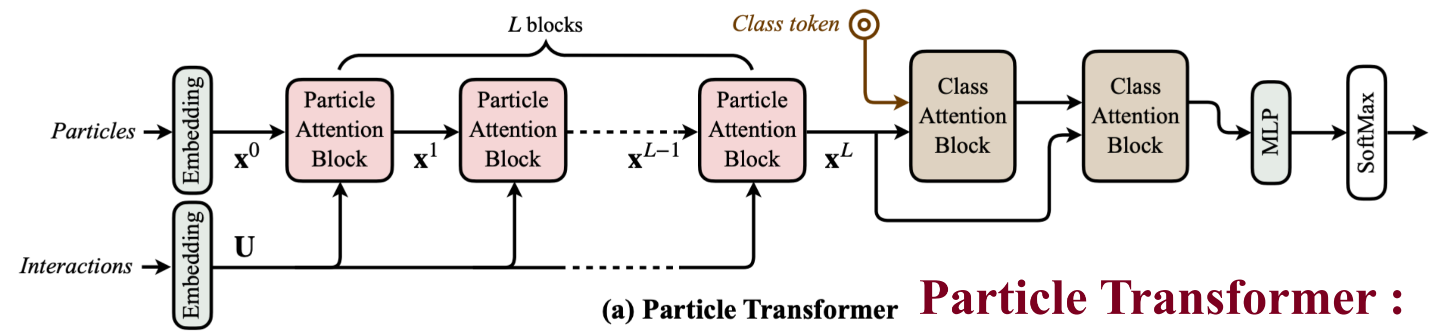
Jonathan Shlomi, Peter Battaglia, Jean-Roch Vlimant

Object tagging

See talk by Felix Ringer for ML based EIC jet tagging



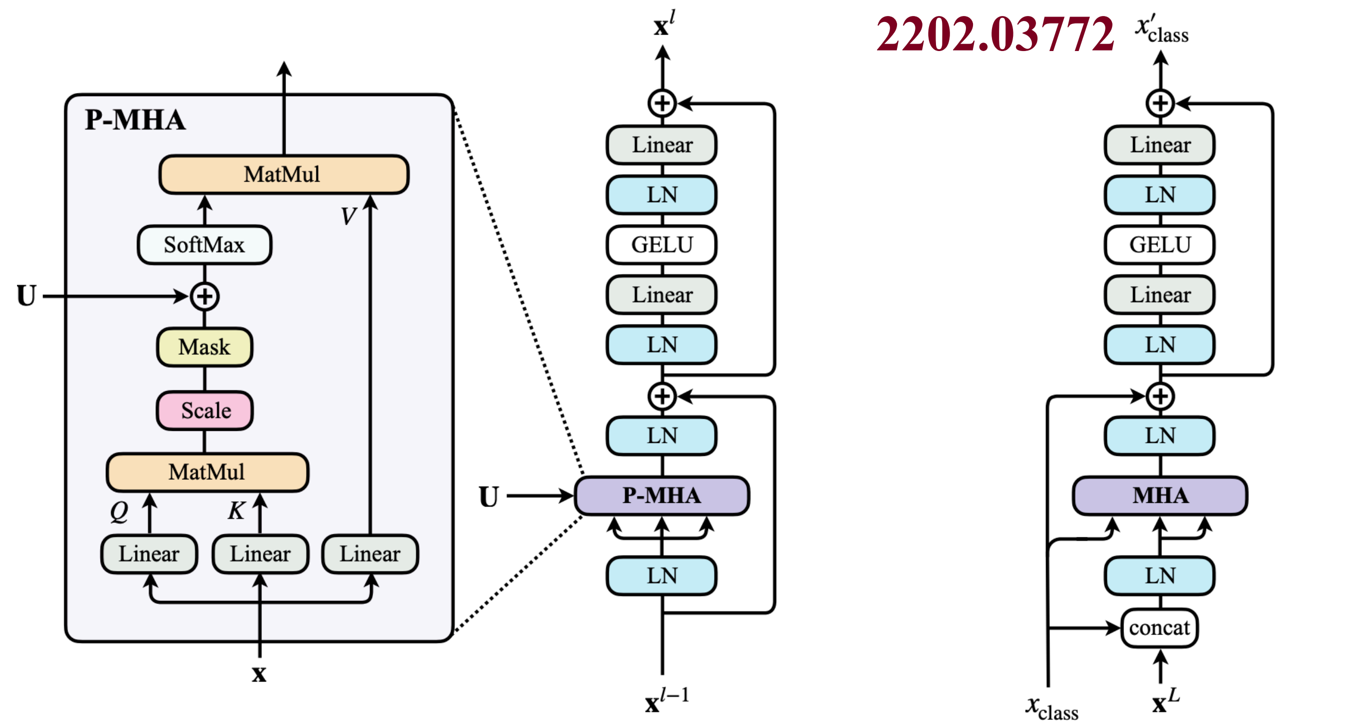




(a) Particle Transformer

Particle Transformer : 2202.03772

	Accuracy	AUC	$\text{Rej}_{50\%}$	$\text{Rej}_{30\%}$
P-CNN	0.930	0.9803	201 ± 4	759 ± 24
PFN	—	0.9819	247 ± 3	888 ± 17
ParticleNet	0.940	0.9858	397 ± 7	1615 ± 93
JEDI-net (w/ $\sum O$)	0.930	0.9807	—	774.6
PCT	0.940	0.9855	392 ± 7	1533 ± 101
LGN	0.929	0.964	—	435 ± 95
rPCN	—	0.9845	364 ± 9	1642 ± 93
LorentzNet	0.942	0.9868	498 ± 18	2195 ± 173
ParT	0.940	0.9858	413 ± 16	1602 ± 81
ParticleNet-f.t.	0.942	0.9866	487 ± 9	1771 ± 80
ParT-f.t.	**0.944**	**0.9877**	$\mathbf{691 \pm 15}$	$\mathbf{2766 \pm 130}$



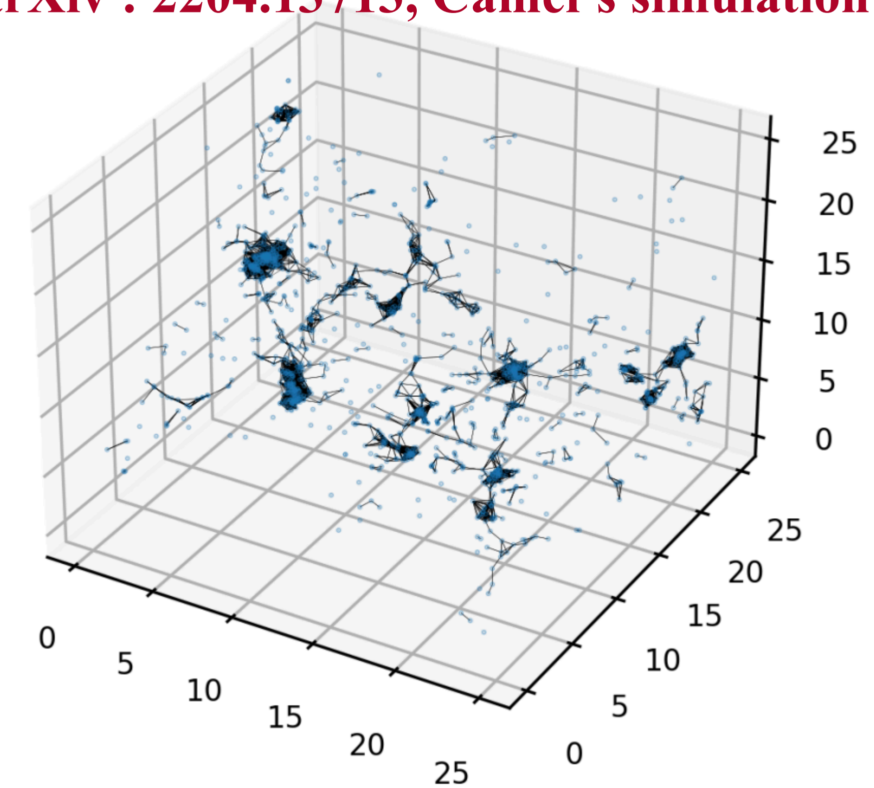
(b) Particle Attention Block

(c) Class Attention Block

Different physics : same ML problems

arXiv : 2204.13713, Camel's simulation

$$f_{\{\theta\}} \Big(\quad \Big)$$



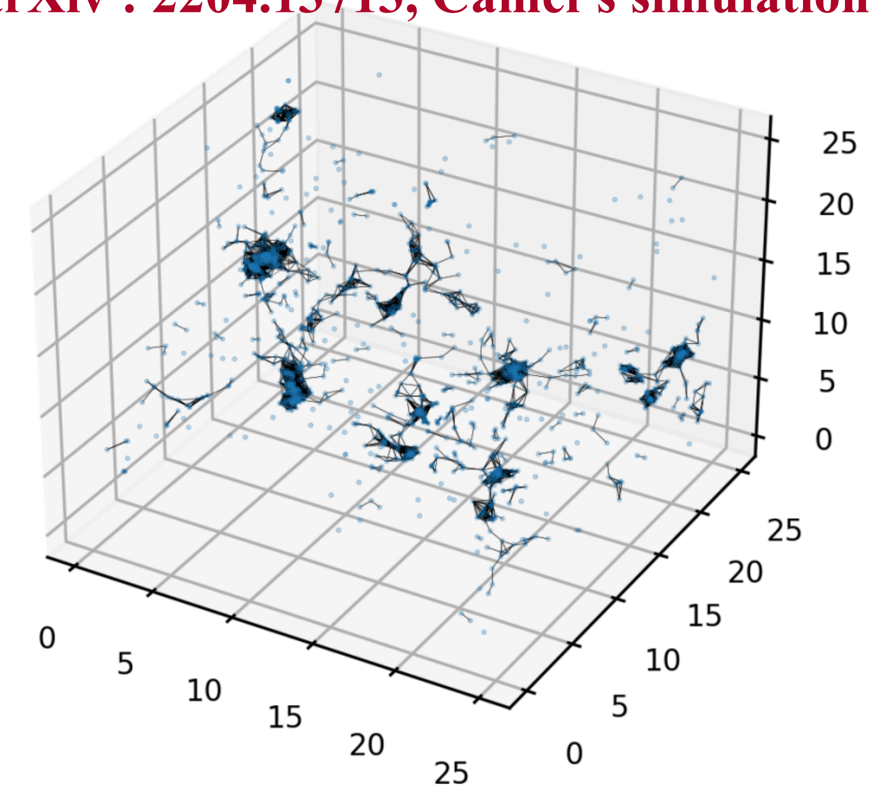
$$\longrightarrow \quad \hat{y} = f_{\{\theta\}}\,(X)$$

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

Different physics : same ML problems

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

arXiv : 2204.13713, Camel's simulation

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

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

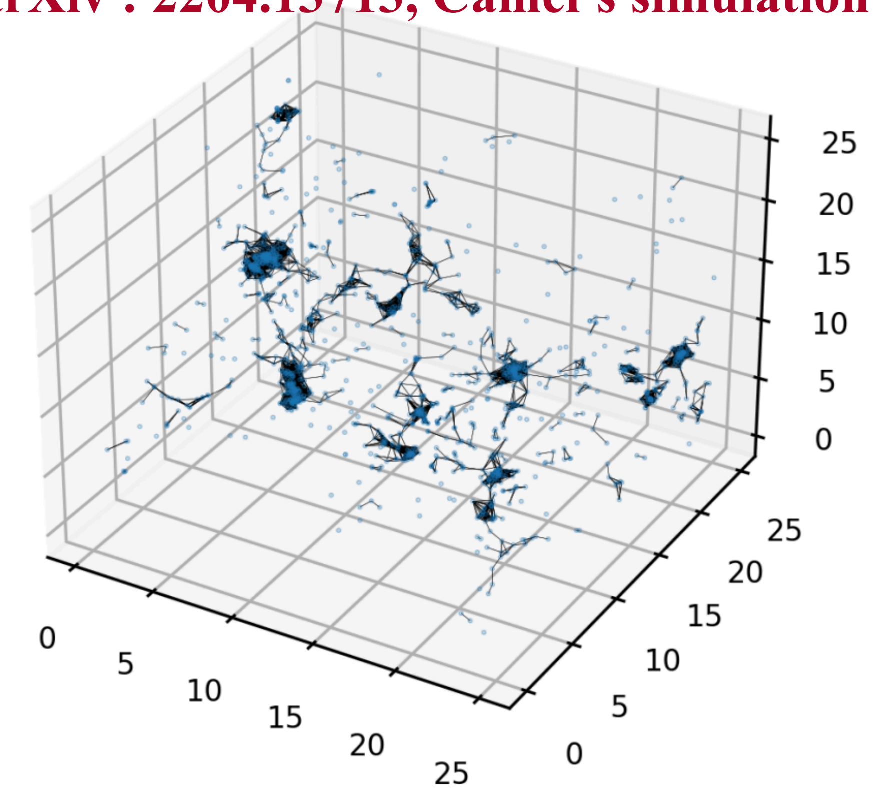
Different physics : same ML problems

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

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

arXiv : 2204.13713, Camel's simulation

$$f_{\{\theta\}} \left(\right.$$



$$\left. \right) \rightarrow \hat{y} = f_{\{\theta\}}(X)$$

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

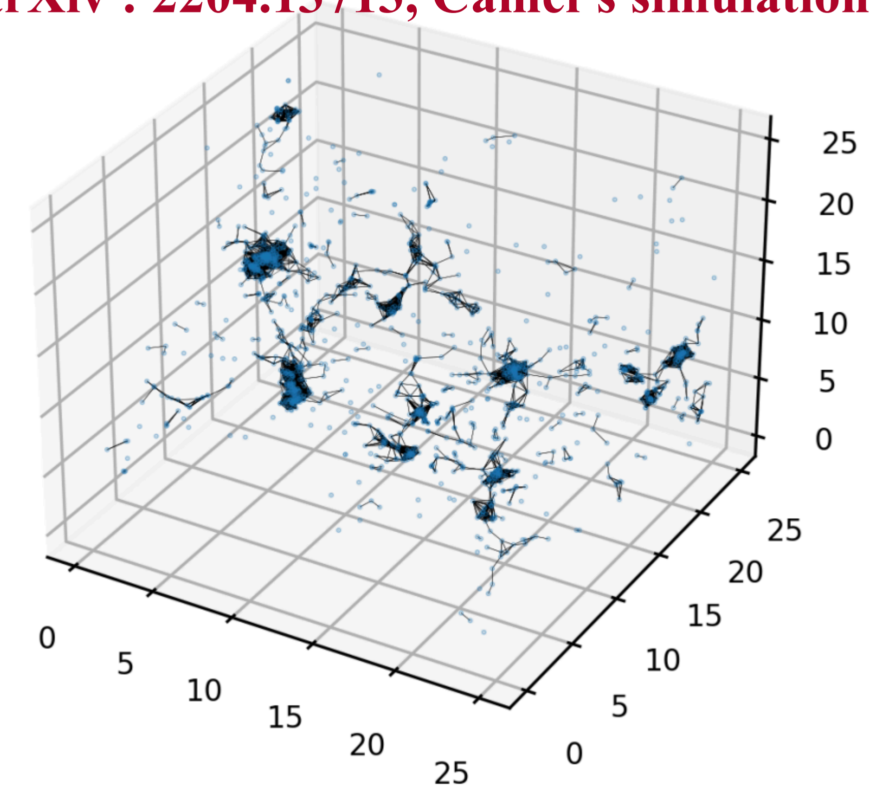
Different physics : same ML problems

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

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

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

Different physics : same ML problems

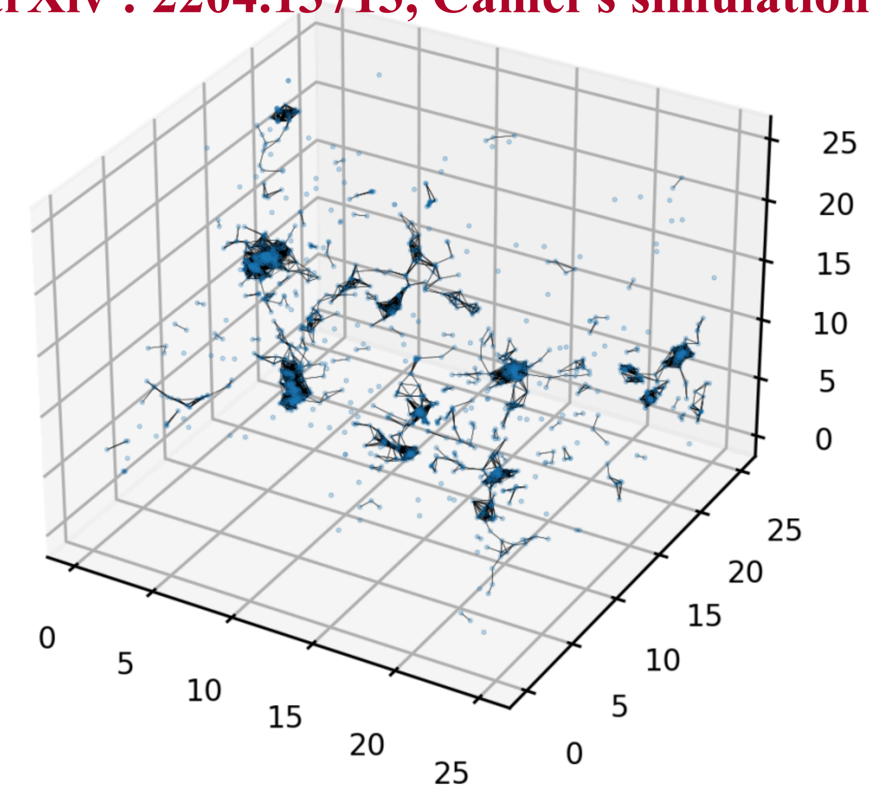
2. Choose a NN model (CNN/GNN/)

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

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



Variation in data

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

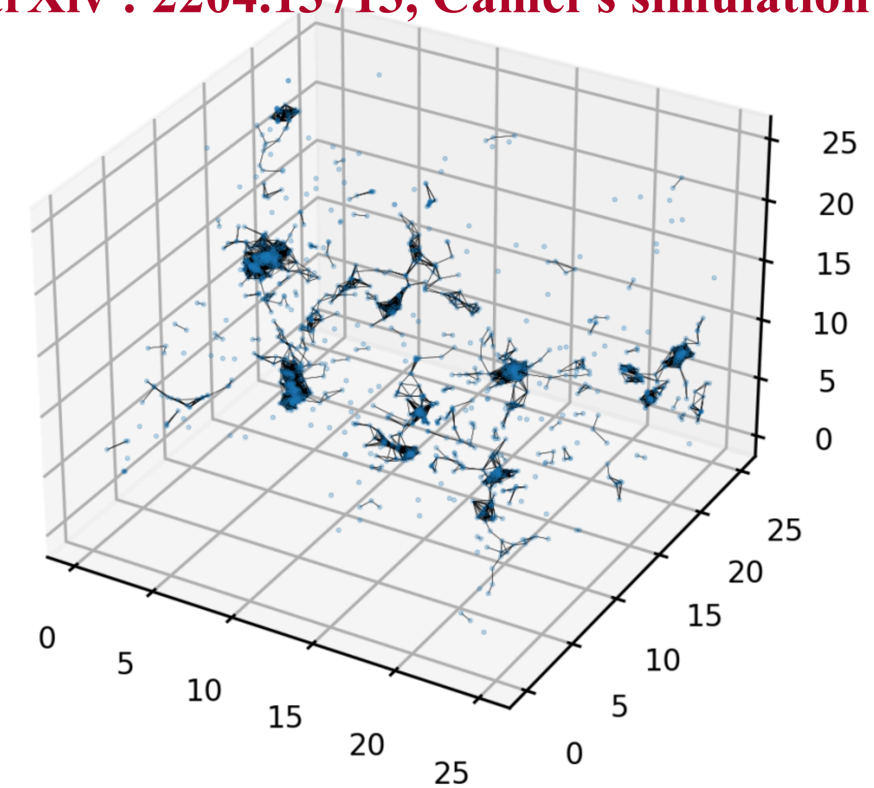
Different physics : same ML problems

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

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

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

Variation in data

Unsupervised

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

Different physics : same ML problems

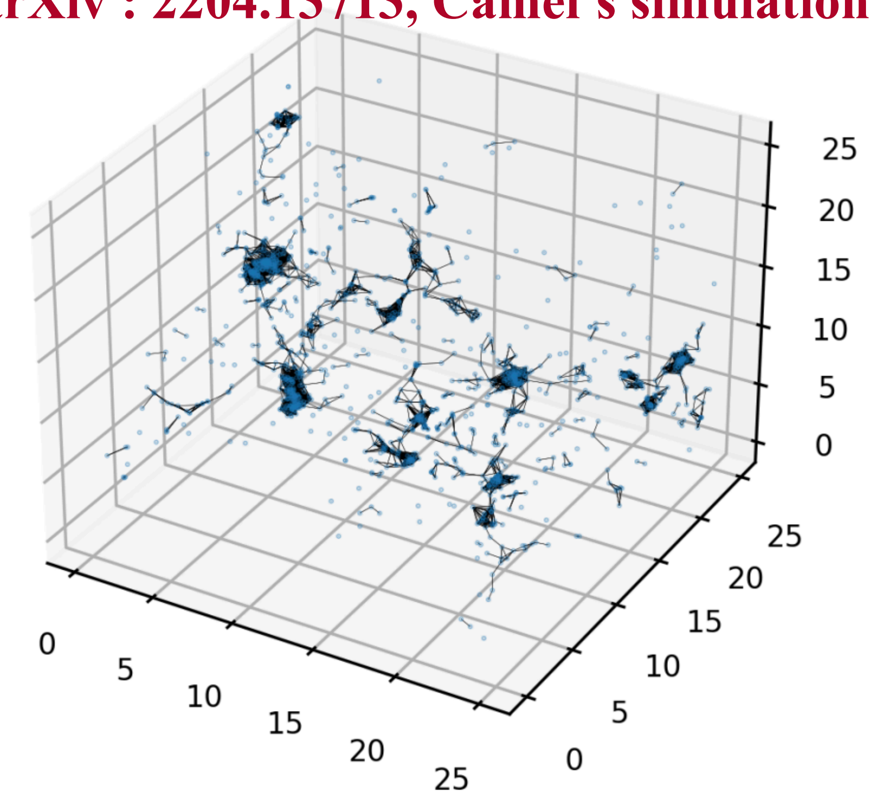
2. Choose a NN model (CNN/GNN/)

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

$$f_{\{\theta\}}\Bigg(\quad$$



$$\Bigg) \longrightarrow \hat{y} = f_{\{\theta\}}(X)$$

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

Variation in data

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)

Different physics : same ML problems

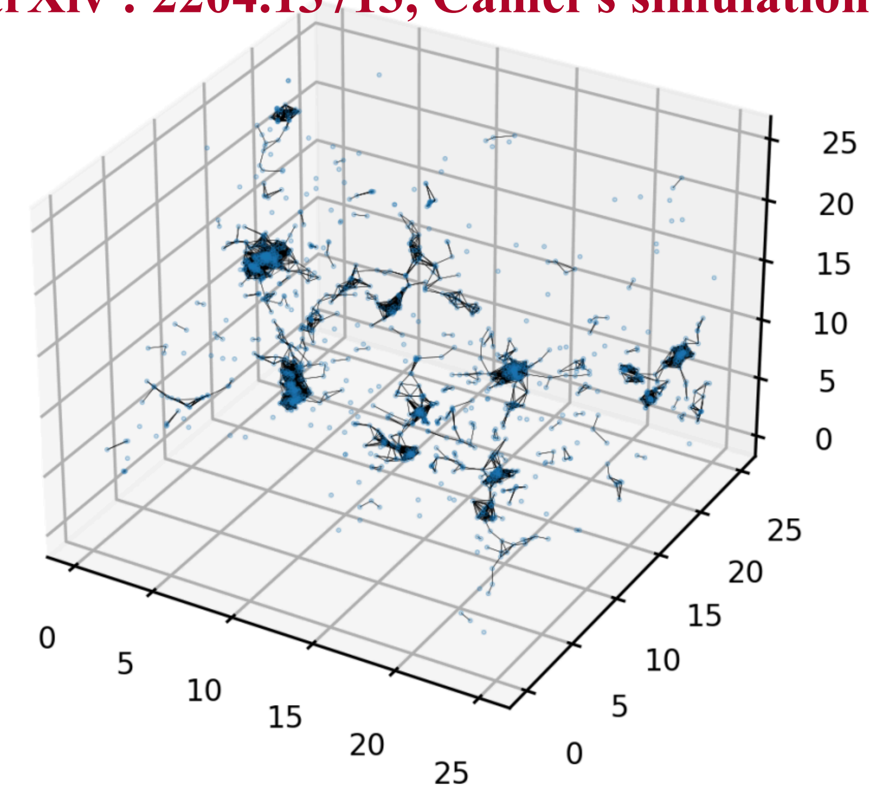
2. Choose a NN model (CNN/GNN/)

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

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.

Different physics : same ML problems

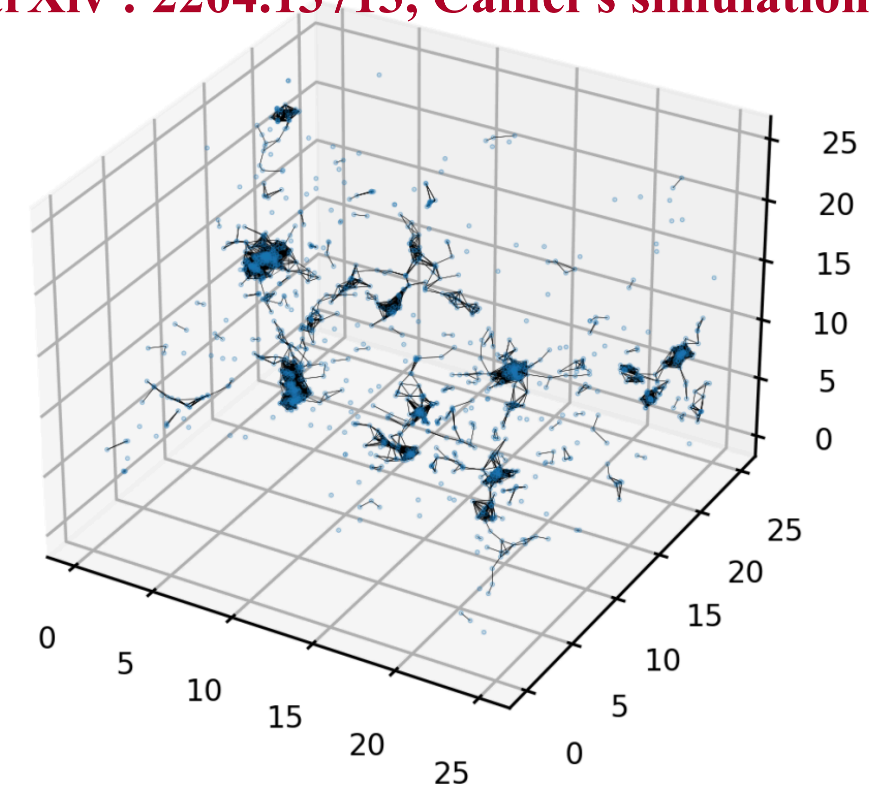
2. Choose a NN model (CNN/GNN/)

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

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

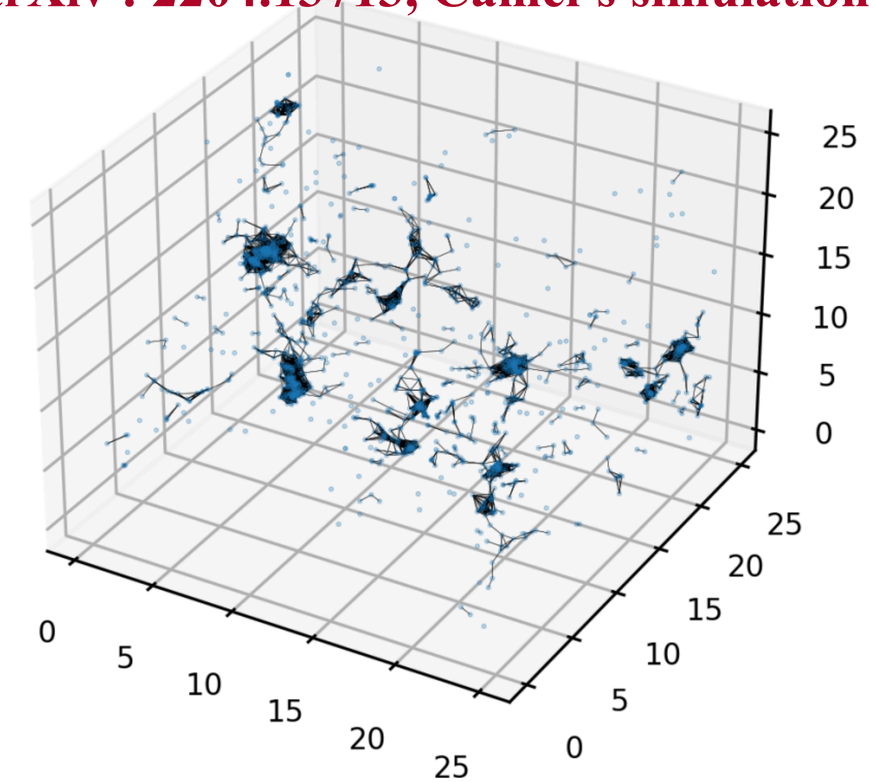
2. Choose a NN model (CNN/GNN/)

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

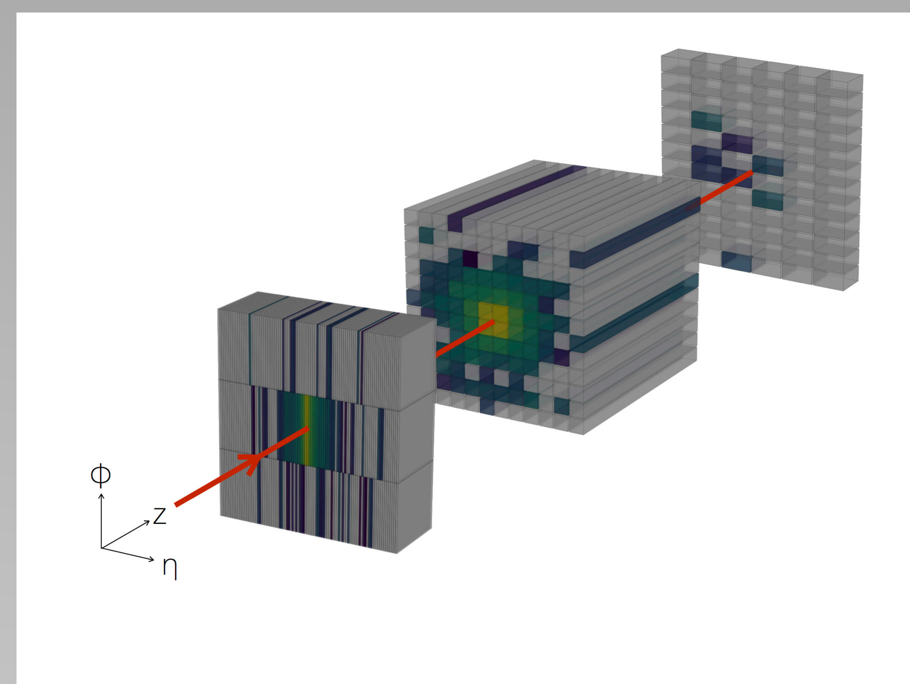
Partial labels. e.g. simulating : SM bkg vs many NP signals.

Supervised

Learning on all the well labeled data.

Calorimetery

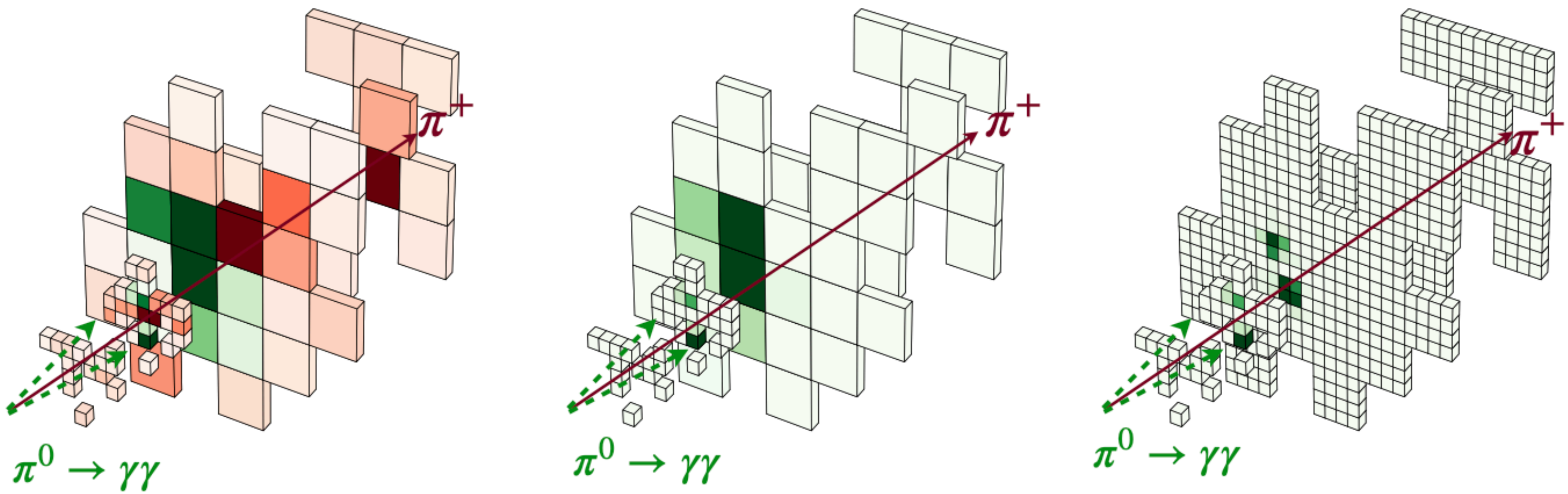
Image from 1705.02355

$f_{\{\theta\}}\,\Big($  $\Big)$

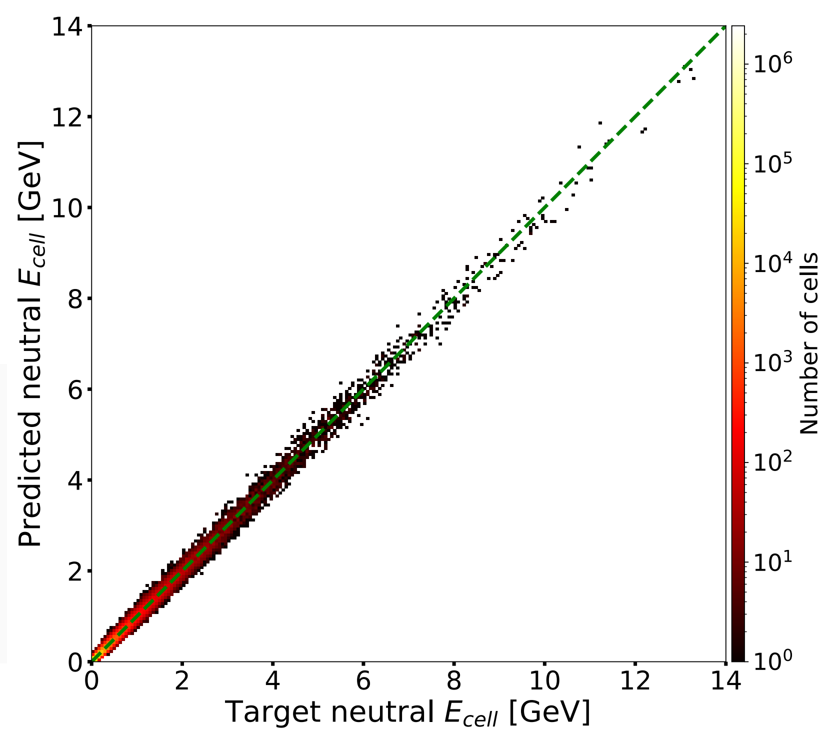
A 3-D view for topoclusters only

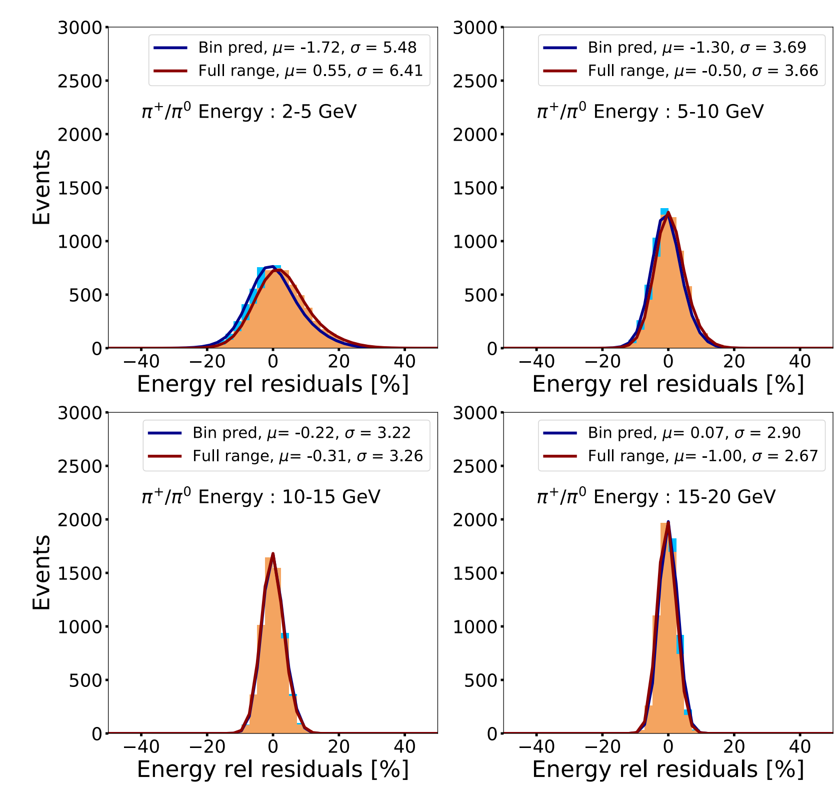
8 X 8 Low Res detector

32 X 32 High Res detector



The networks in general have good noise removal abilities.





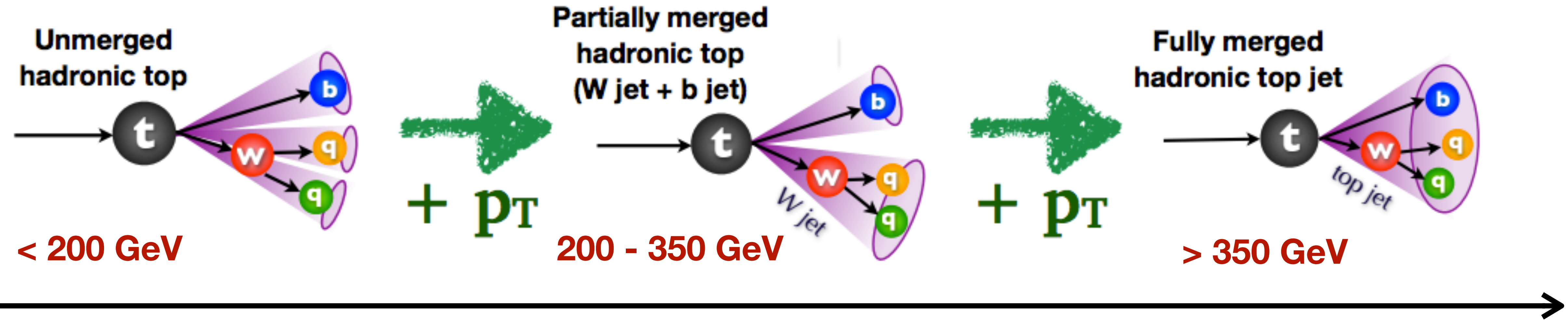
Towards a computer vision particle flow

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

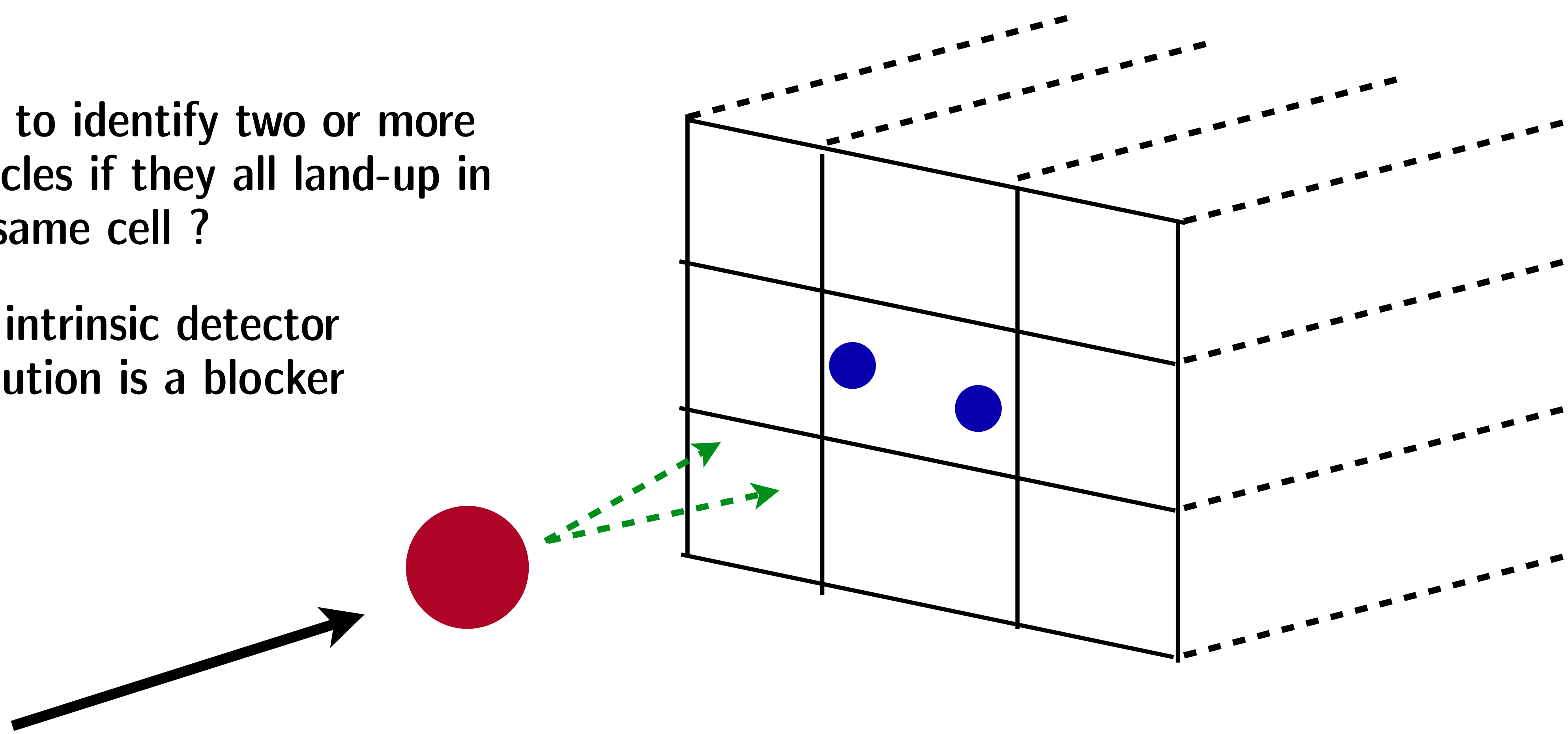
1341 Accesses | **13** Citations | **11** Altmetric | Metrics

When do intrinsic calorimeter sizes are limiting factors ?

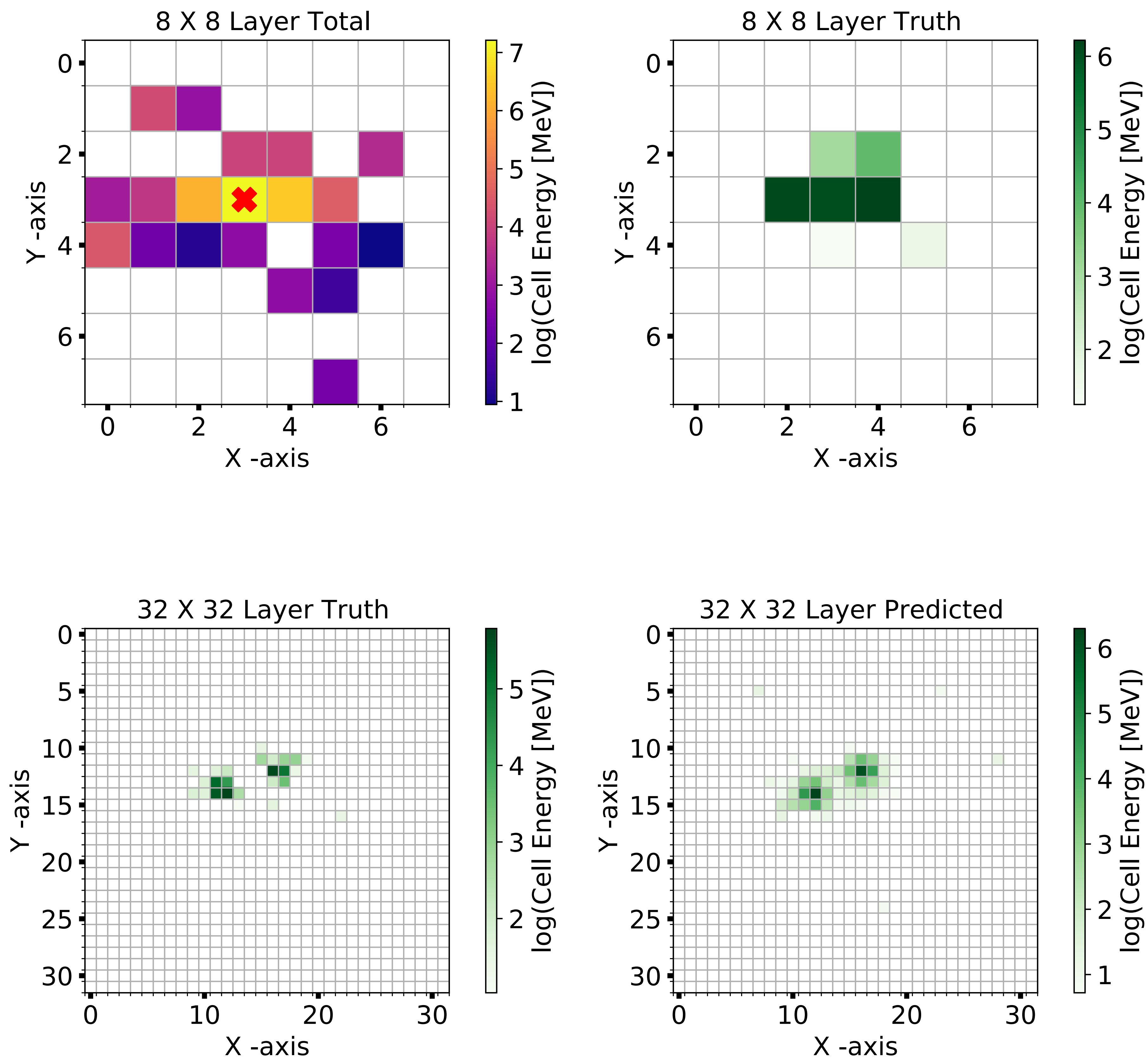


How to identify two or more particles if they all land-up in the same cell ?

The intrinsic detector resolution is a blocker

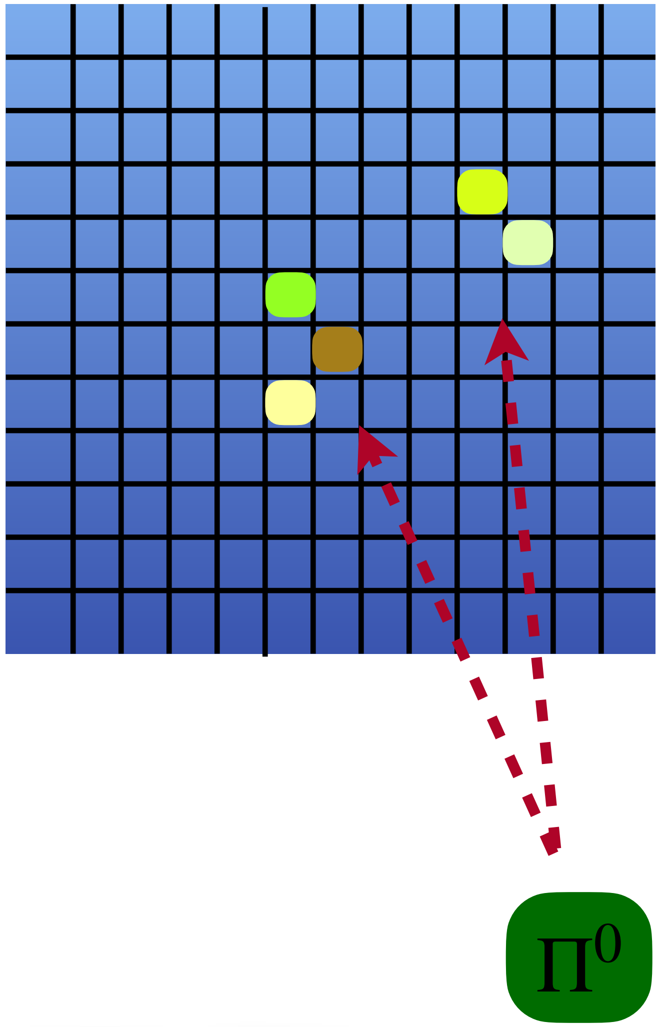


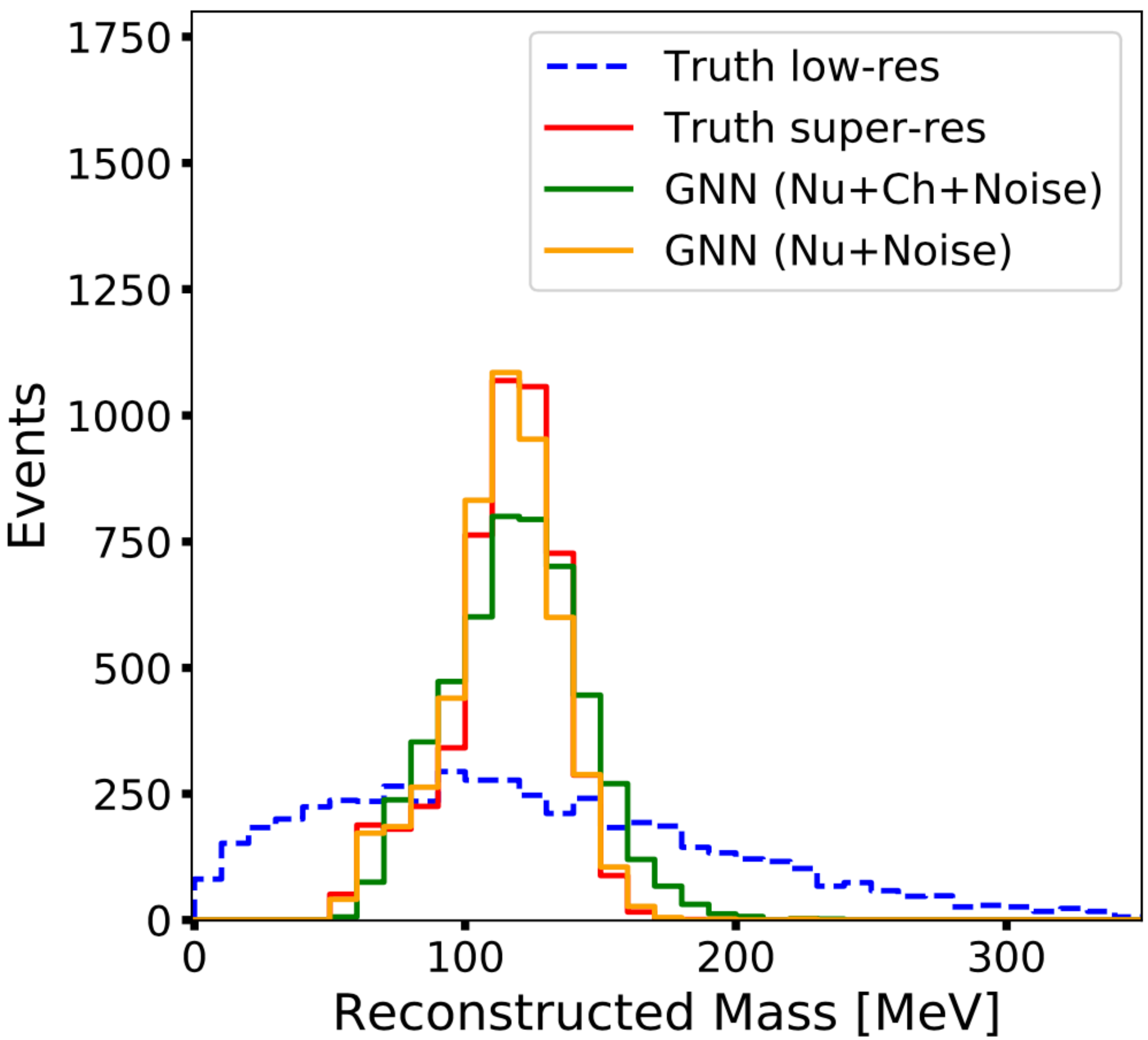
An event display for super-res prediction



The mass distribution

Invariant mass from reconstructed 4-vectors.





On the use of neural networks for energy reconstruction in high-granularity calorimeters *JINST* 16 (2021) 12, P12036
N. Akchurin (Texas Tech.), C. Cowden (Texas Tech.), J. Damgov (Texas Tech.), A. Hussain (Texas Tech.), S. Kunori (Texas Tech.)

How to GAN Higher Jet Resolution *SciPost Phys.* 13 (2022) 3, 064
Pierre Baldi[1], Lukas Blecher[2], Anja Butter[2], Julian Collado[1], Jessica N. Howard[3], Fabian Keilbach[2], Tilman Plehn[2], Gregor Kasieczka[4], and Daniel Whiteson[3]

Higgs and top physics reconstruction challenges and opportunities at FCC-ee *Eur.Phys.J.Plus* 137 (2022) 1, 39
Patrizia Azzi (INFN, Padua and Padua U.), Loukas Gouskos (CERN), Michele Selvaggi (CERN), Frank Simon (Munich, Max Planck Inst.)

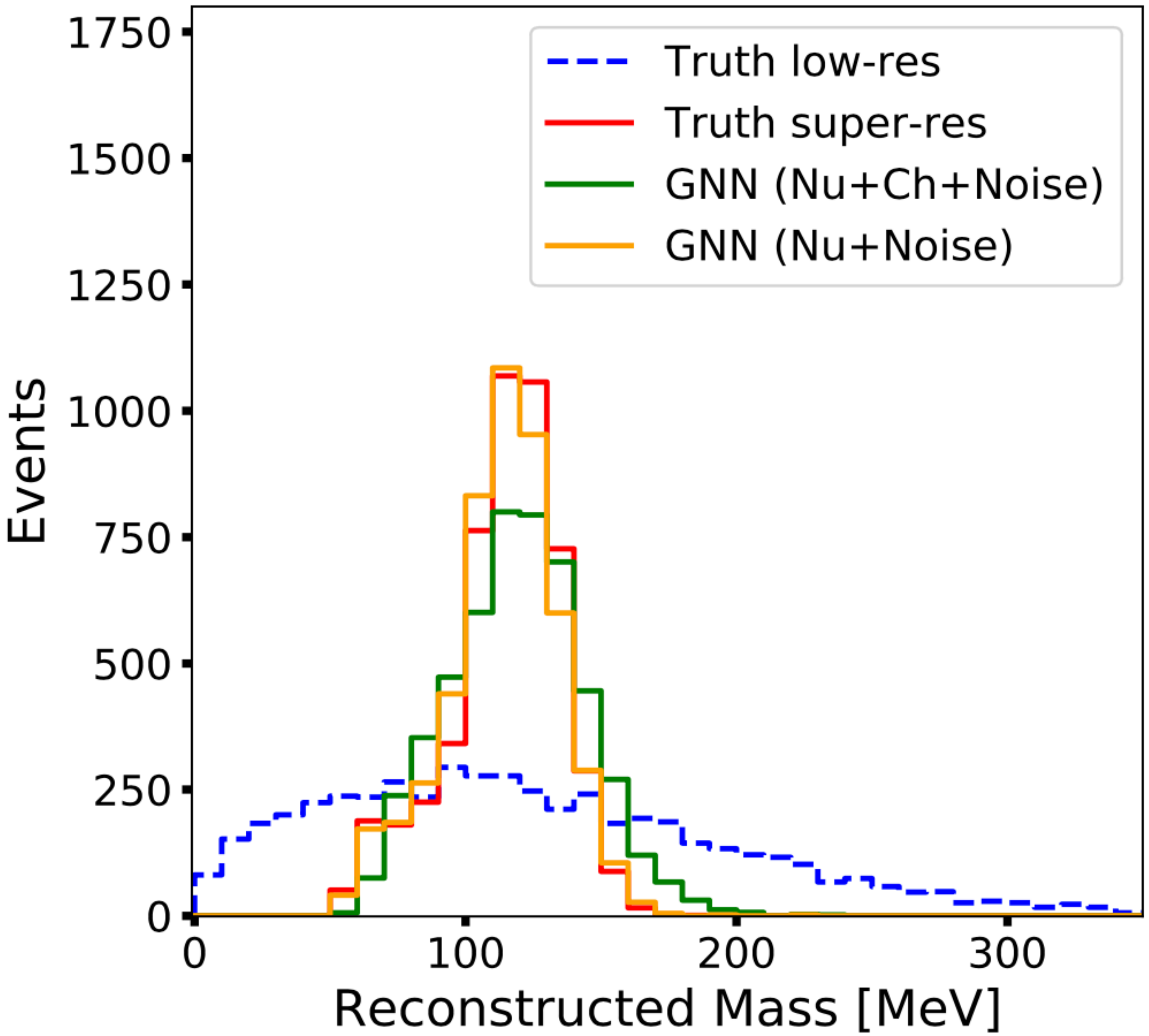
Test-beam and simulation studies towards RPWELL-based DHCAL #13
Dan Shaked-Renous (Weizmann Inst.), Fernando Domingues Amaro (Coimbra U.), Purba Bhattacharya (Weizmann Inst.), Amos Breskin (Weizmann Inst.), Maximilien Chefdeville (Annecy, LAPP) et al. (Aug 26, 2022) *JINST* 17 (2022) 12, P12008

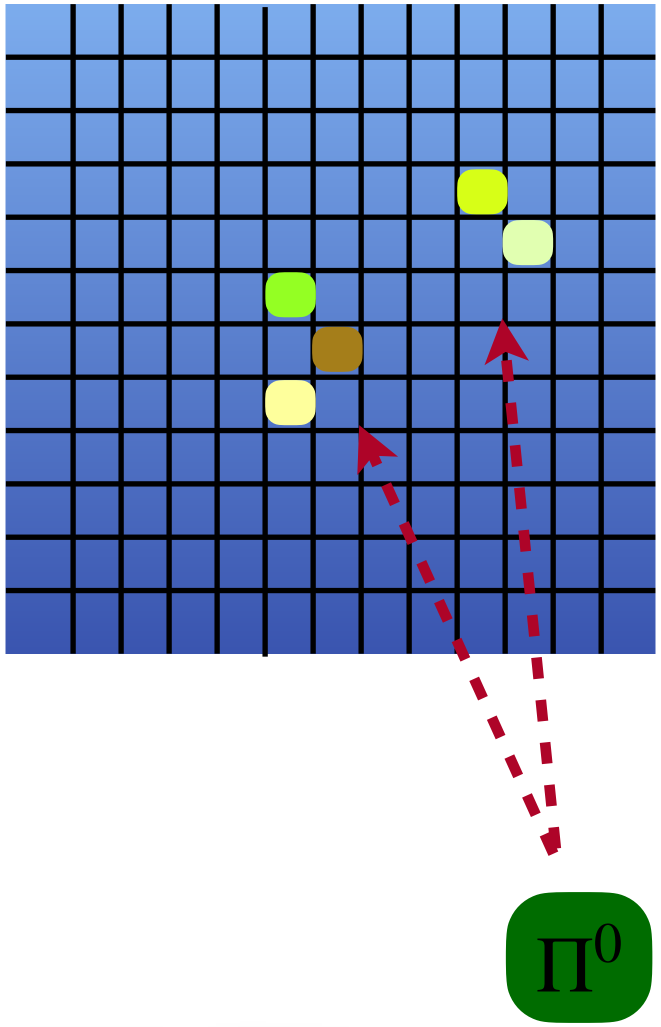
A high-granularity calorimeter insert based on SiPM-on-tile technology at the future Electron-Ion Collider *Nucl.Instrum.Meth.A* 1047 (2023) 167866
Miguel Arratia (UC, Riverside), Kenneth Barish (UC, Riverside), Liam Blanchard (UC, Riverside), Huan Z. Huang (Southern California U.), Zhongling Ji (Southern California U.) et al. (Aug 10, 2022)

The mass distribution

Invariant mass from reconstructed 4-vectors.

Let's think how we can bring this in the collimated object business.





On the use of neural networks for energy reconstruction in high-granularity calorimeters *JINST* 16 (2021) 12, P12036
N. Akchurin (Texas Tech.), C. Cowden (Texas Tech.), J. Damgov (Texas Tech.), A. Hussain (Texas Tech.), S. Kunori (Texas Tech.)

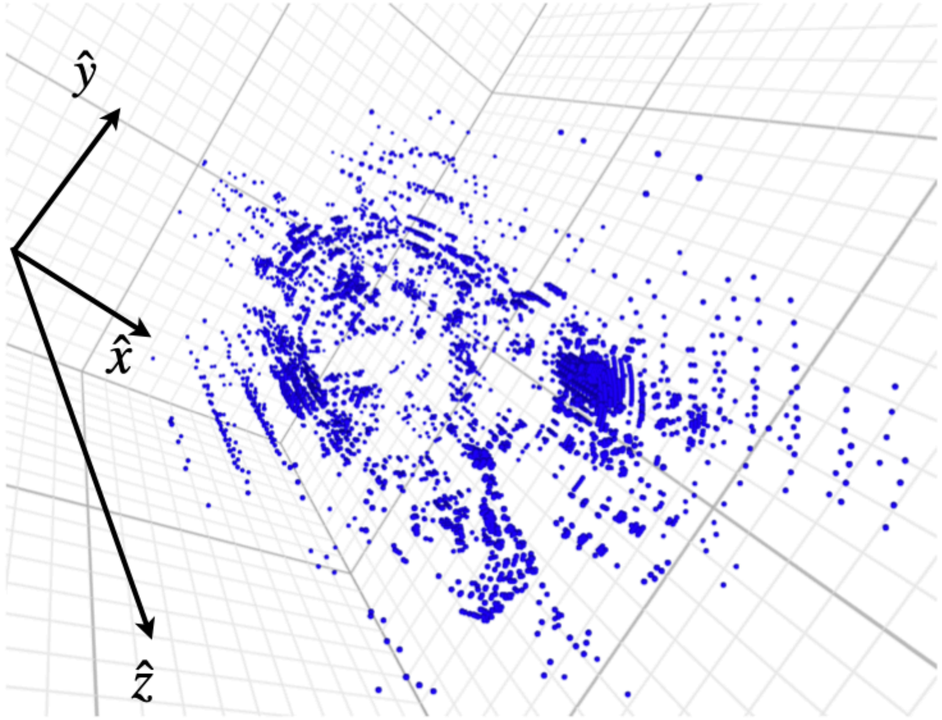
How to GAN Higher Jet Resolution *SciPost Phys.* 13 (2022) 3, 064
Pierre Baldi[1], Lukas Blecher[2], Anja Butter[2], Julian Collado[1], Jessica N. Howard[3], Fabian Keilbach[2], Tilman Plehn[2], Gregor Kasieczka[4], and Daniel Whiteson[3]

Higgs and top physics reconstruction challenges and opportunities at FCC-ee *Eur.Phys.J.Plus* 137 (2022) 1, 39
Patrizia Azzi (INFN, Padua and Padua U.), Loukas Gouskos (CERN), Michele Selvaggi (CERN), Frank Simon (Munich, Max Planck Inst.)

Test-beam and simulation studies towards RPWELL-based DHCAL #13
Dan Shaked-Renous (Weizmann Inst.), Fernando Domingues Amaro (Coimbra U.), Purba Bhattacharya (Weizmann Inst.), Amos Breskin (Weizmann Inst.), Maximilien Chefdeville (Annecy, LAPP) et al. (Aug 26, 2022) *JINST* 17 (2022) 12, P12008

A high-granularity calorimeter insert based on SiPM-on-tile technology at the future Electron-Ion Collider *Nucl.Instrum.Meth.A* 1047 (2023) 167866
Miguel Arratia (UC, Riverside), Kenneth Barish (UC, Riverside), Liam Blanchard (UC, Riverside), Huan Z. Huang (Southern California U.), Zhongling Ji (Southern California U.) et al. (Aug 10, 2022)

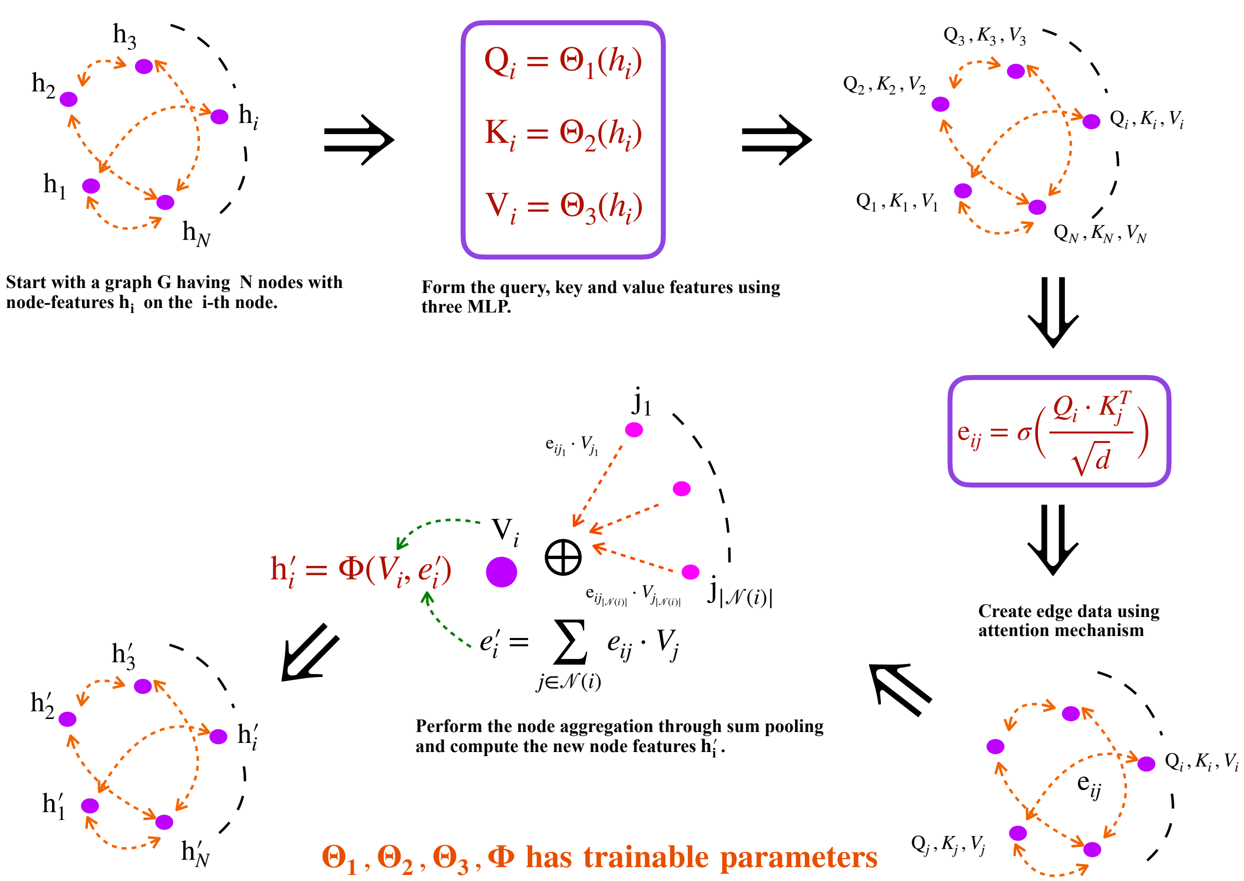
Pion identification within ATLAS

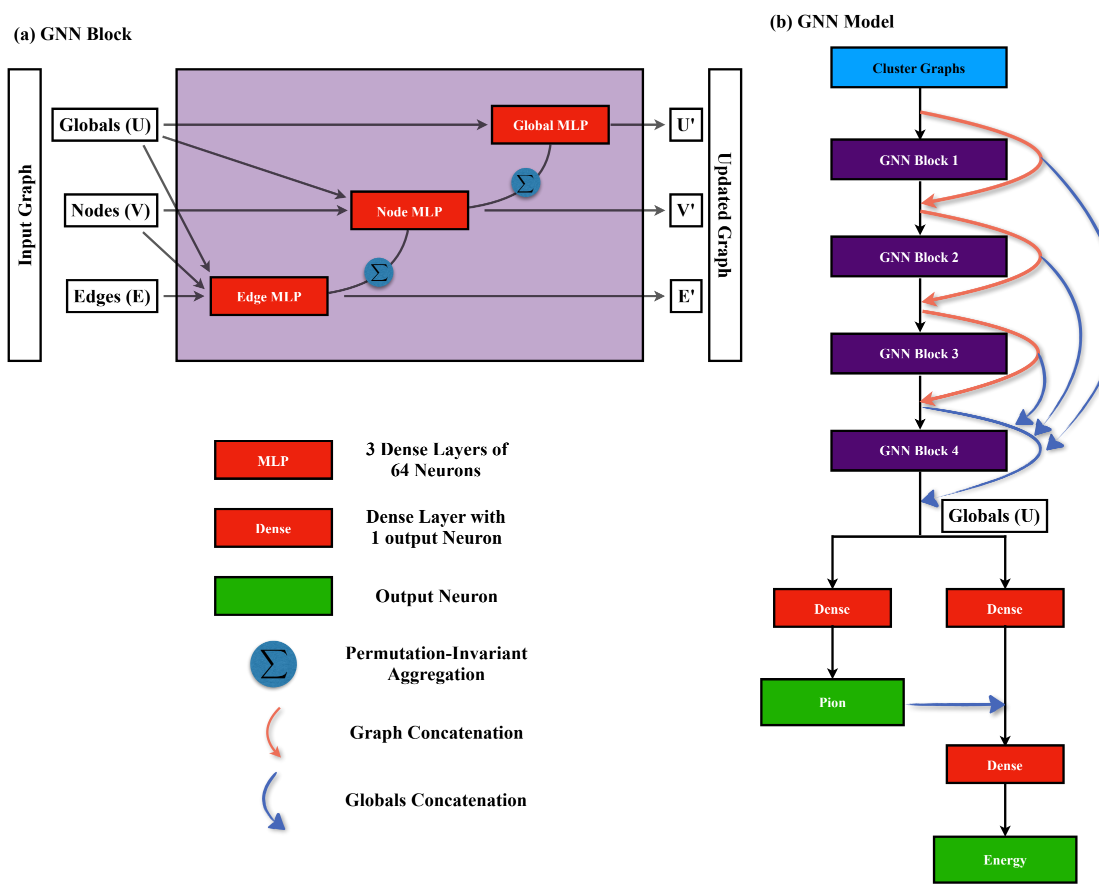


A classification & regression task is tested on ATLAS samples.
The calibrated topocluster cells are used to form images & P.C.

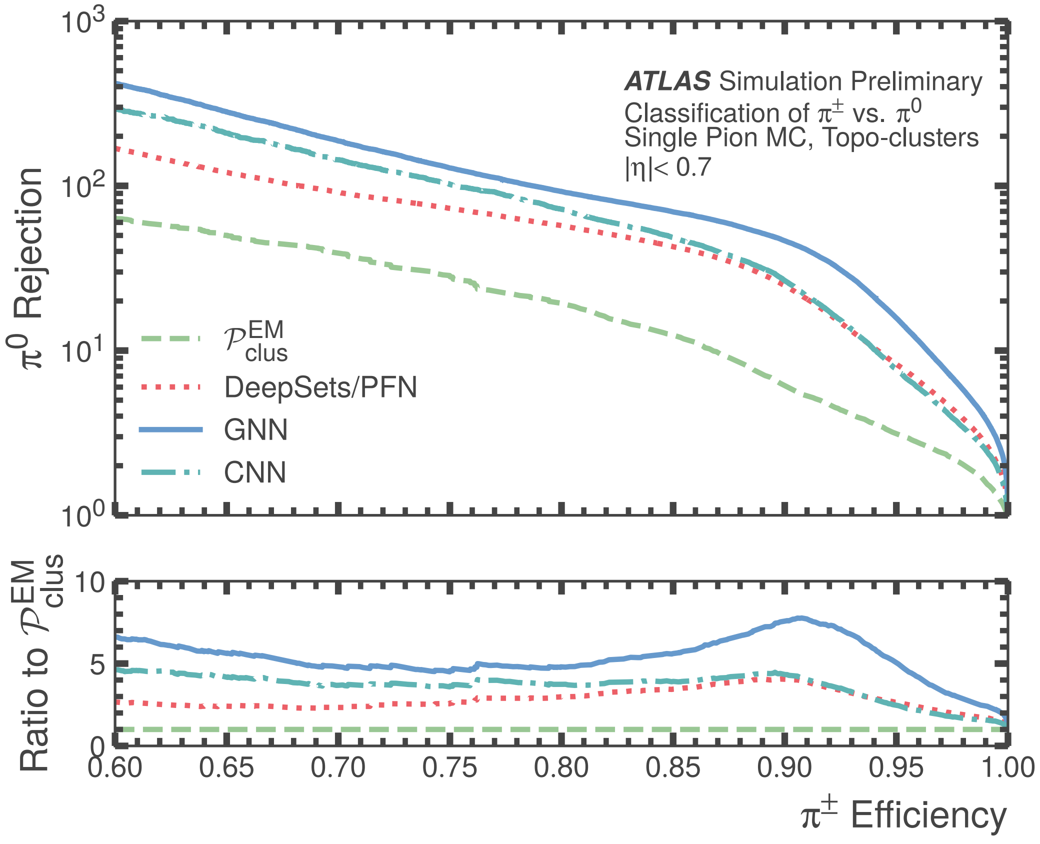
ATL-PHYS-PUB-2020-040

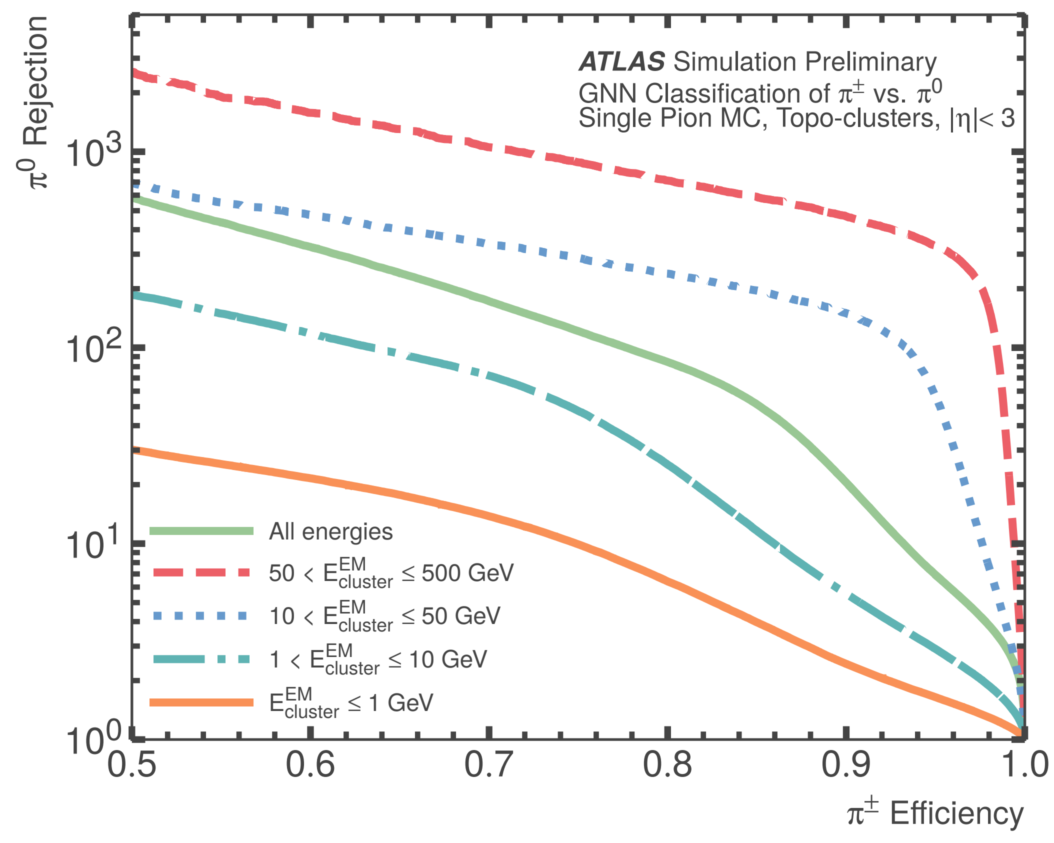
$$\mathcal{L} = (1-\alpha)\mathcal{L}_{\text{classification}} + \alpha\mathcal{L}_{\text{Regression}}$$

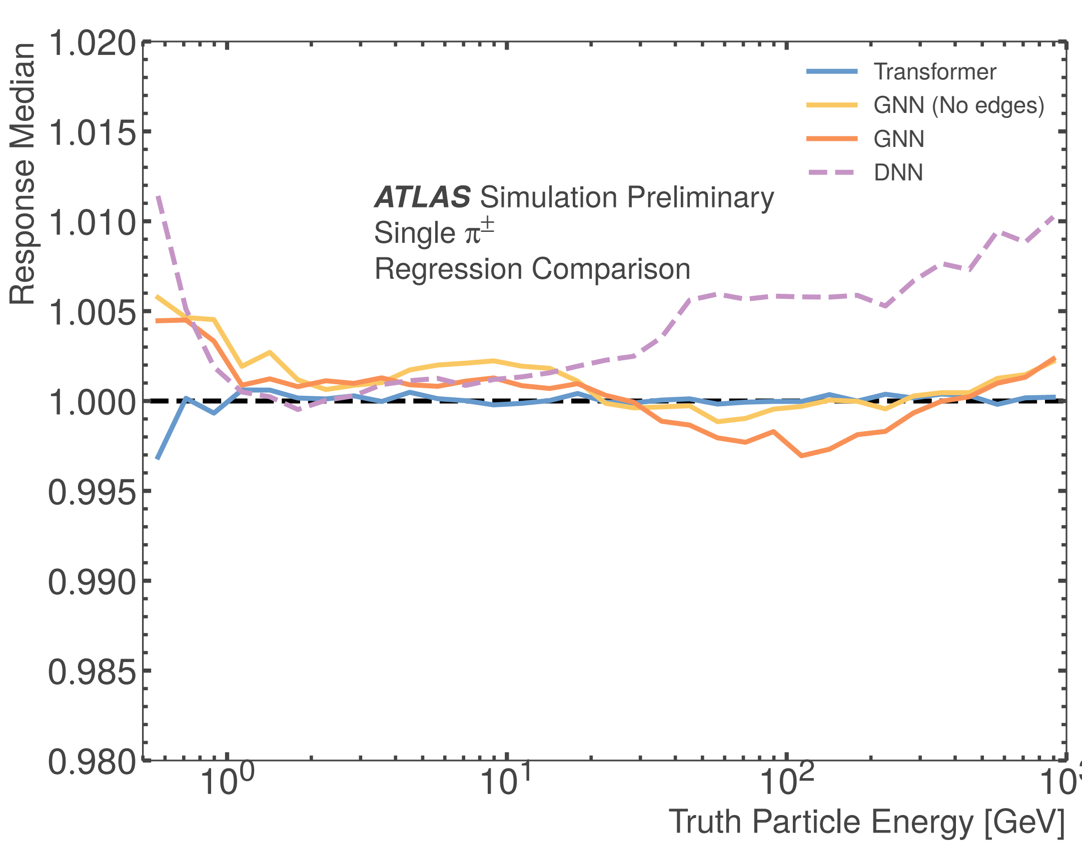


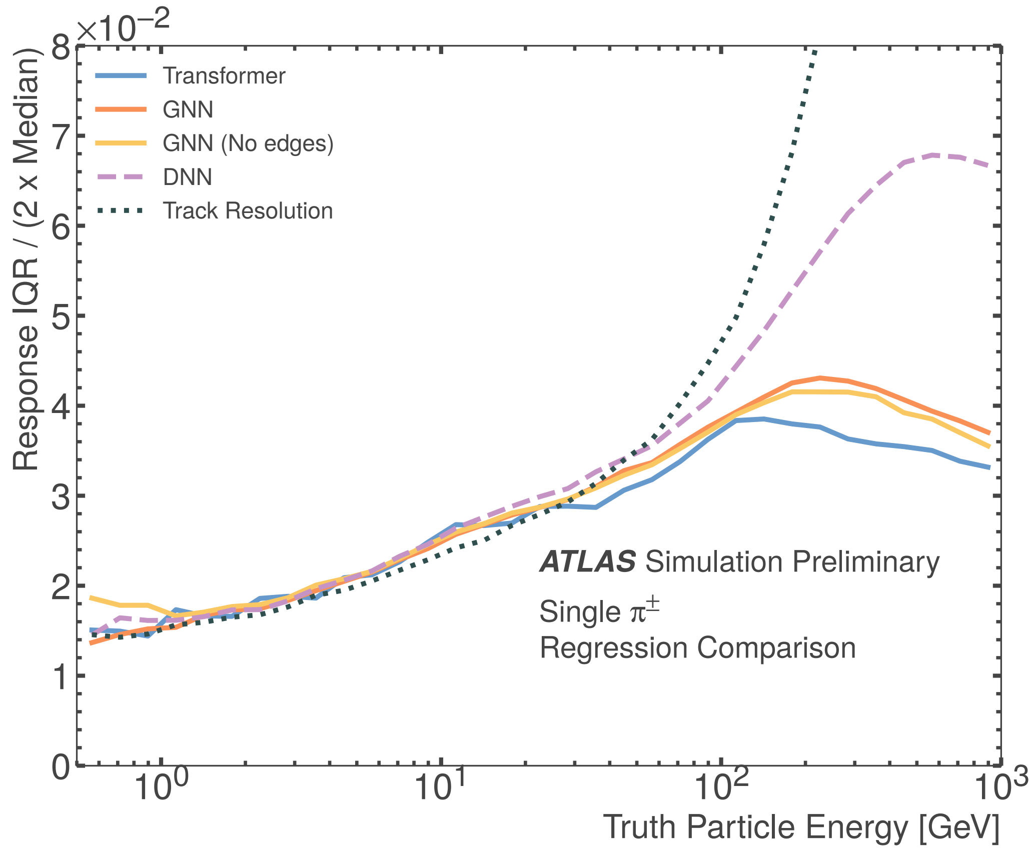


Pion identification within ATLAS

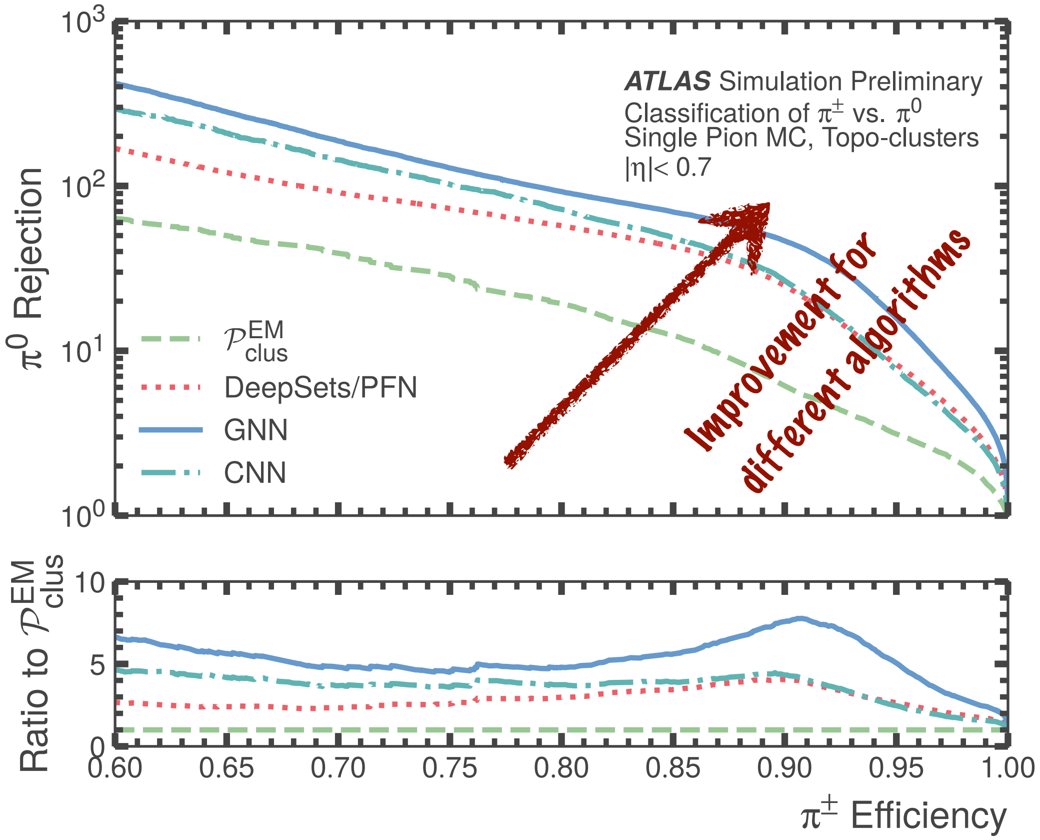


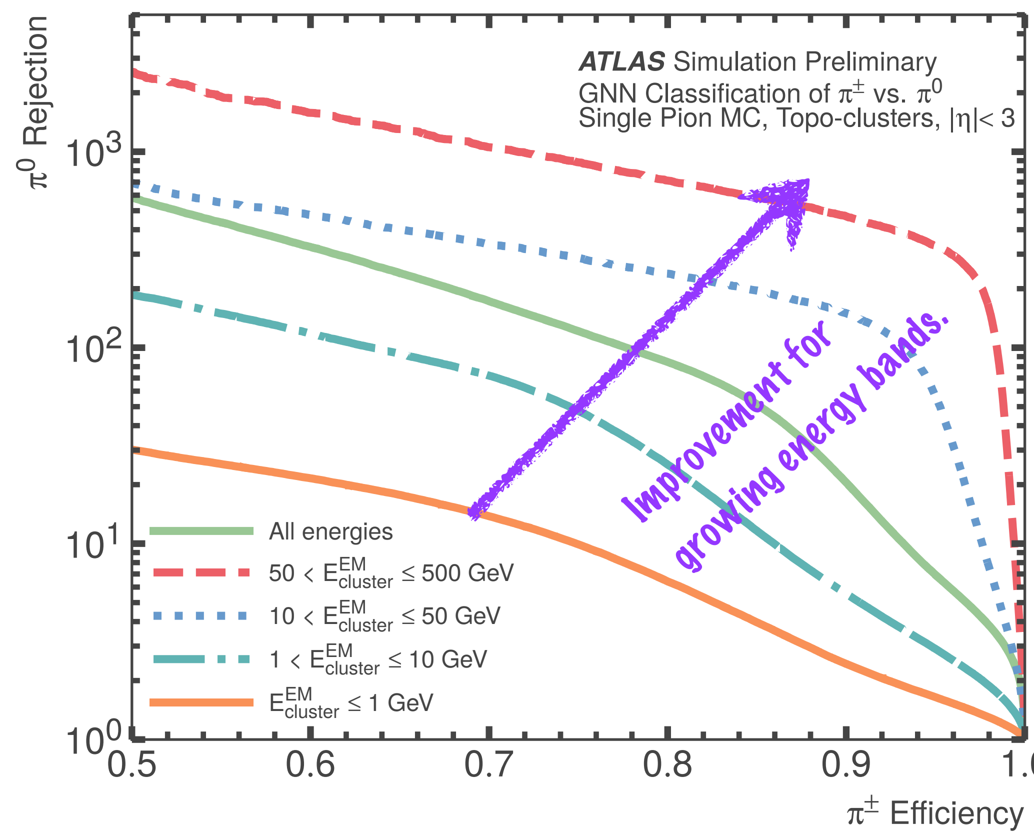


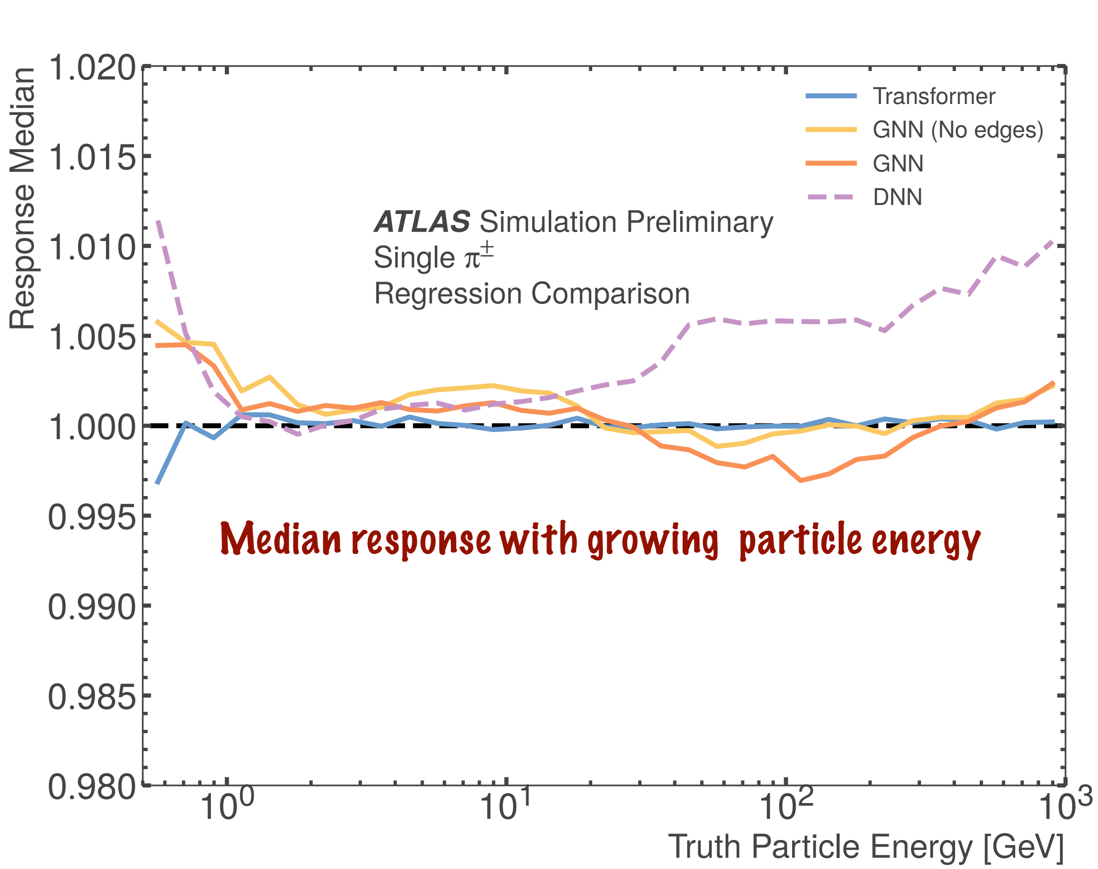


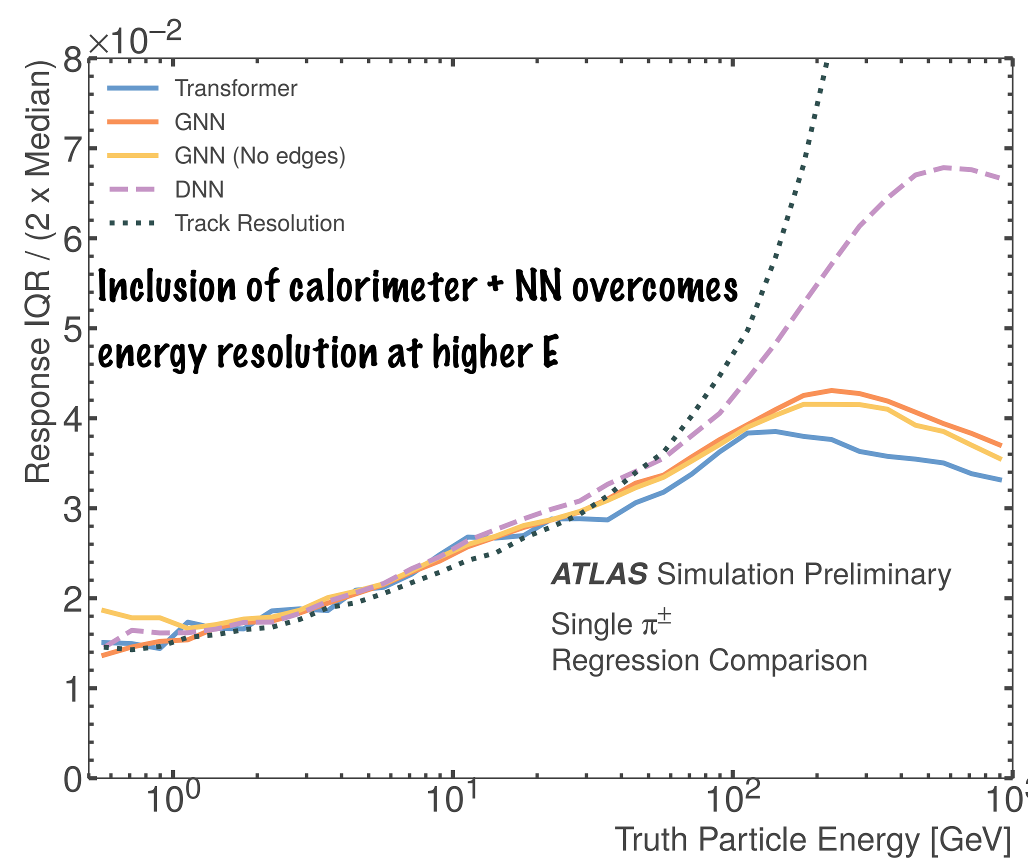


Pion identification within ATLAS

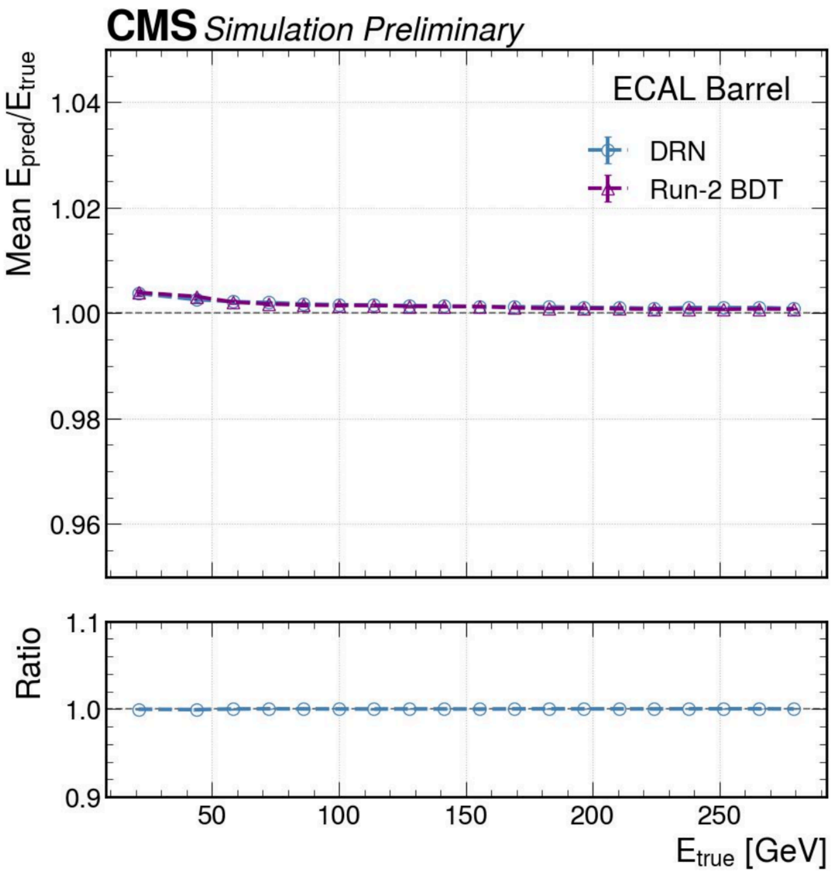


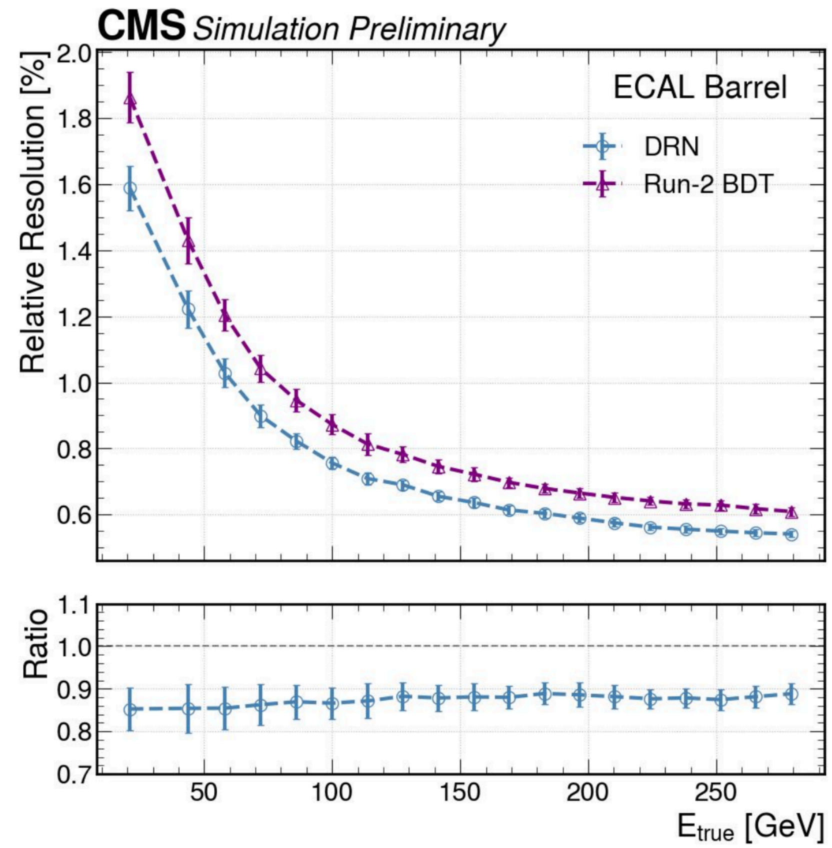




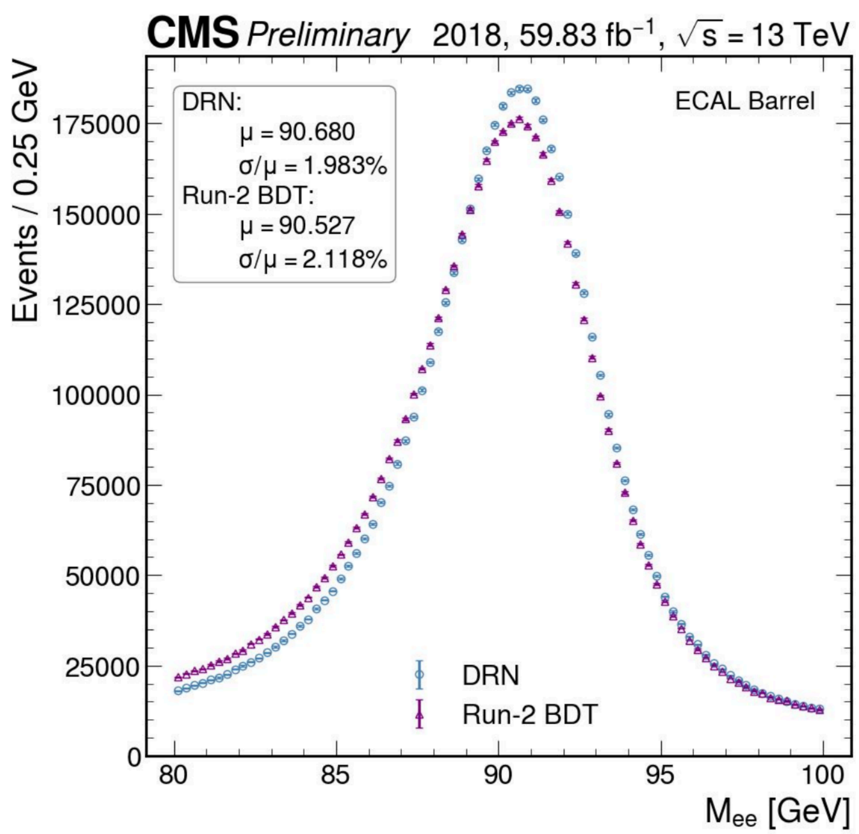


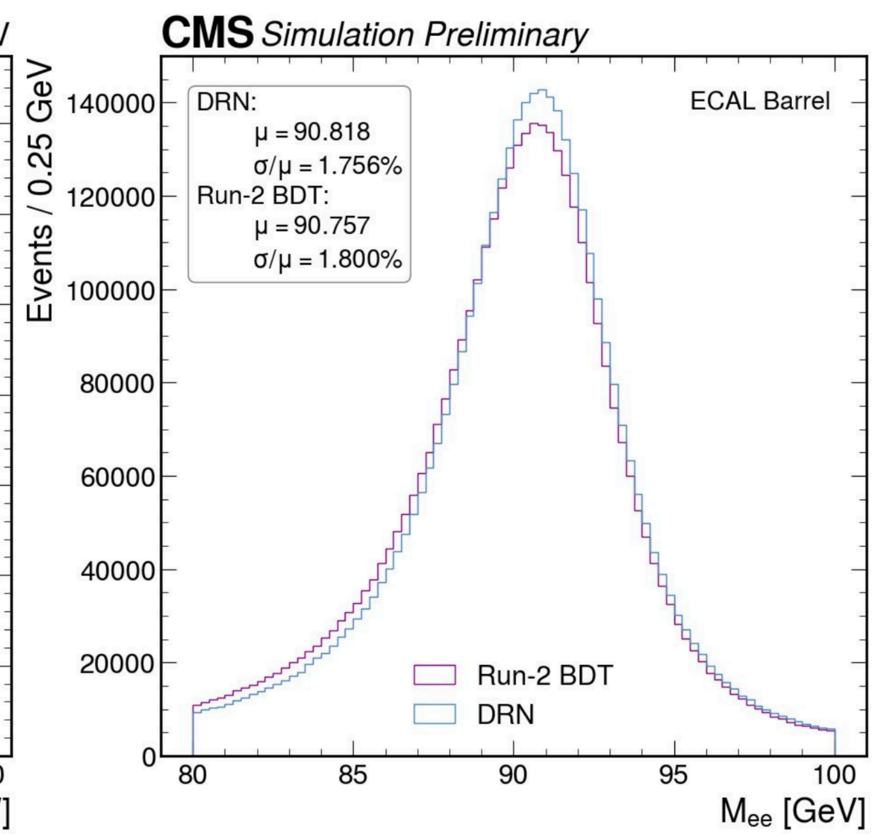
Electron identification within CMS





DRN is a dynamic GNN

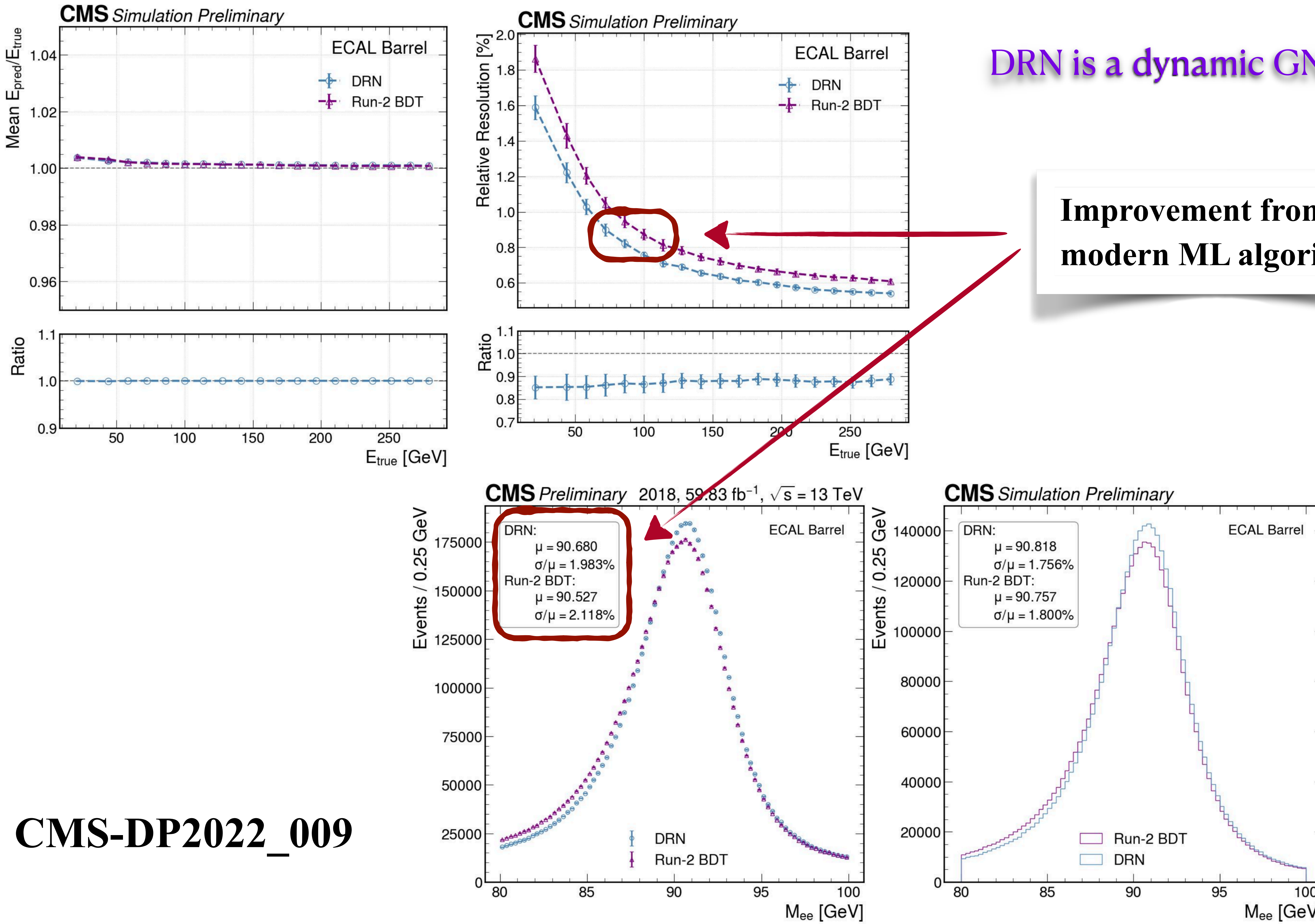




CMS-DP2022_009

Electron identification within CMS

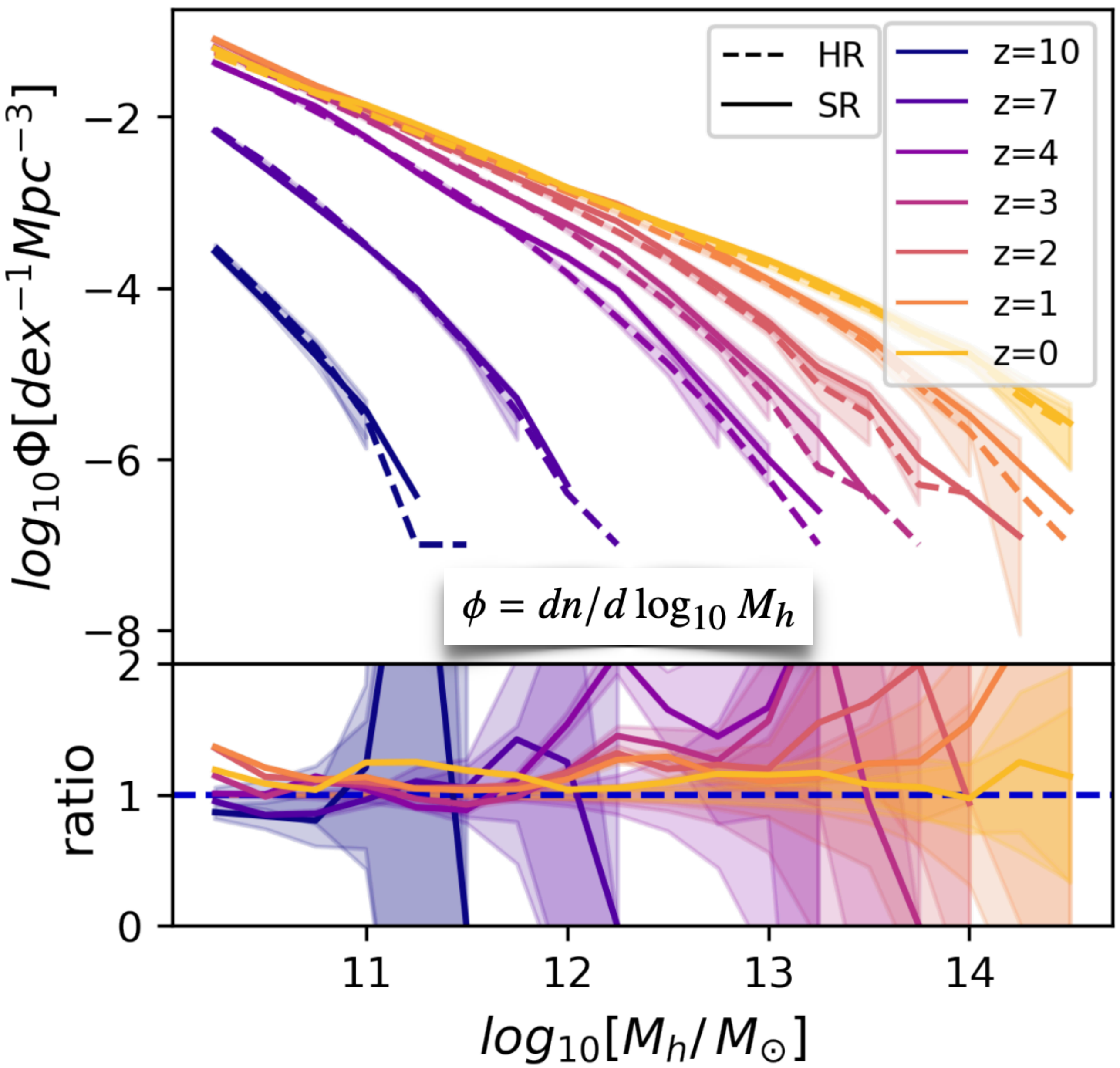
DRN is a dynamic GNN

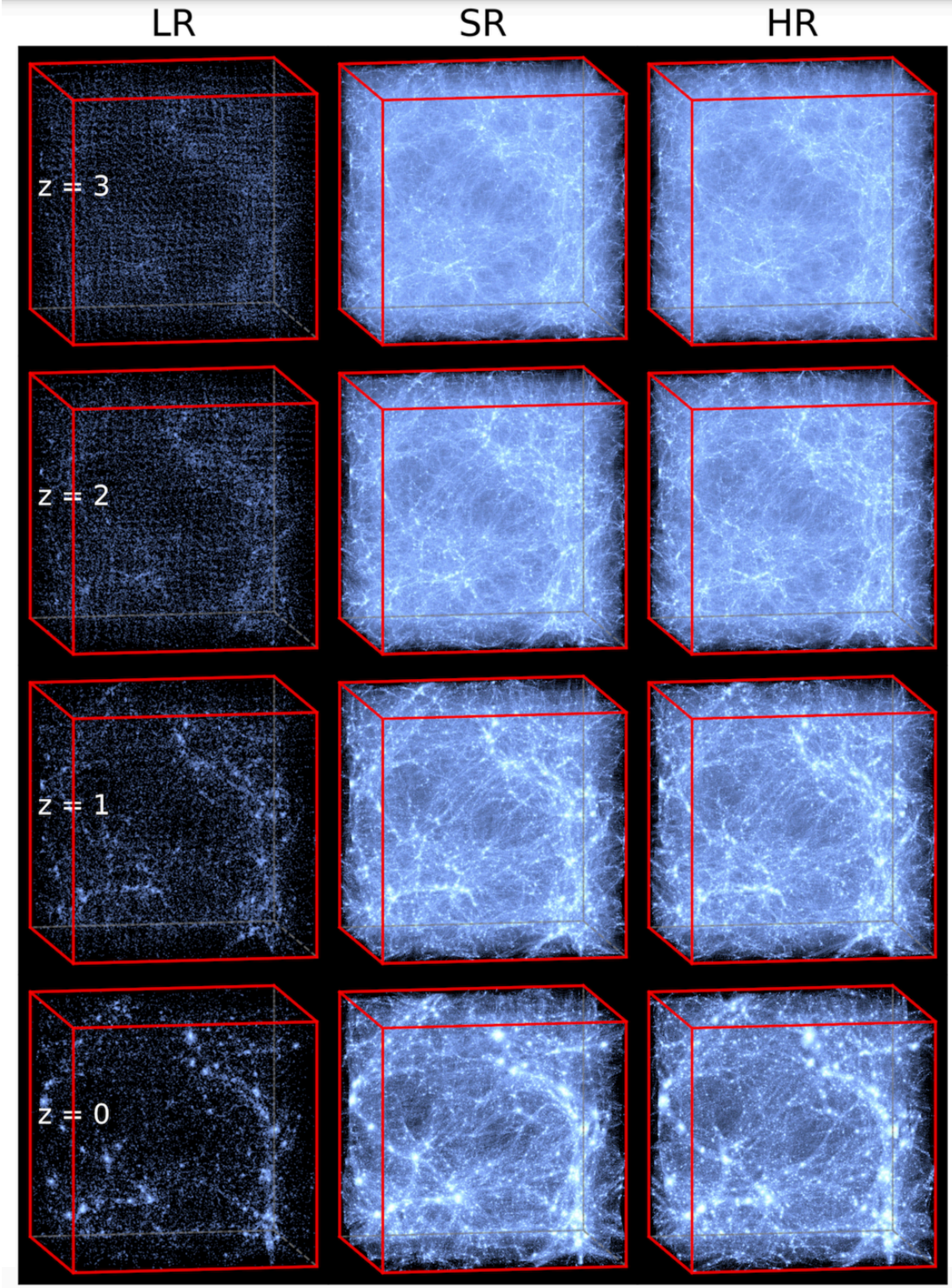


Improvement from a modern ML algorithm

CMS-DP2022_009

Super-resolution

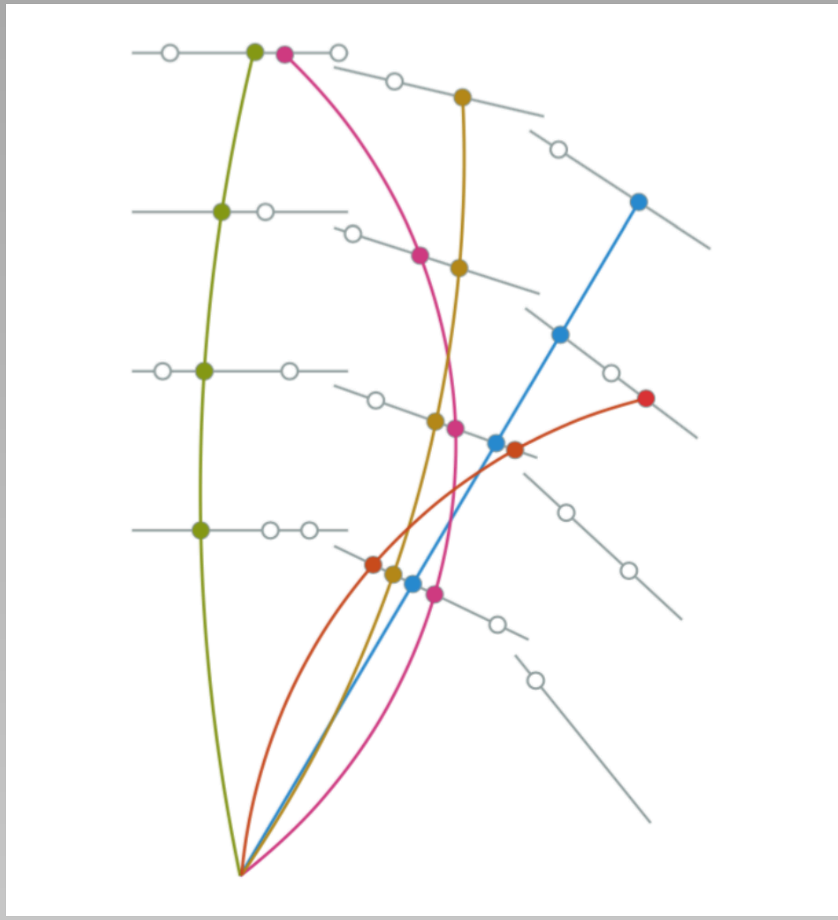




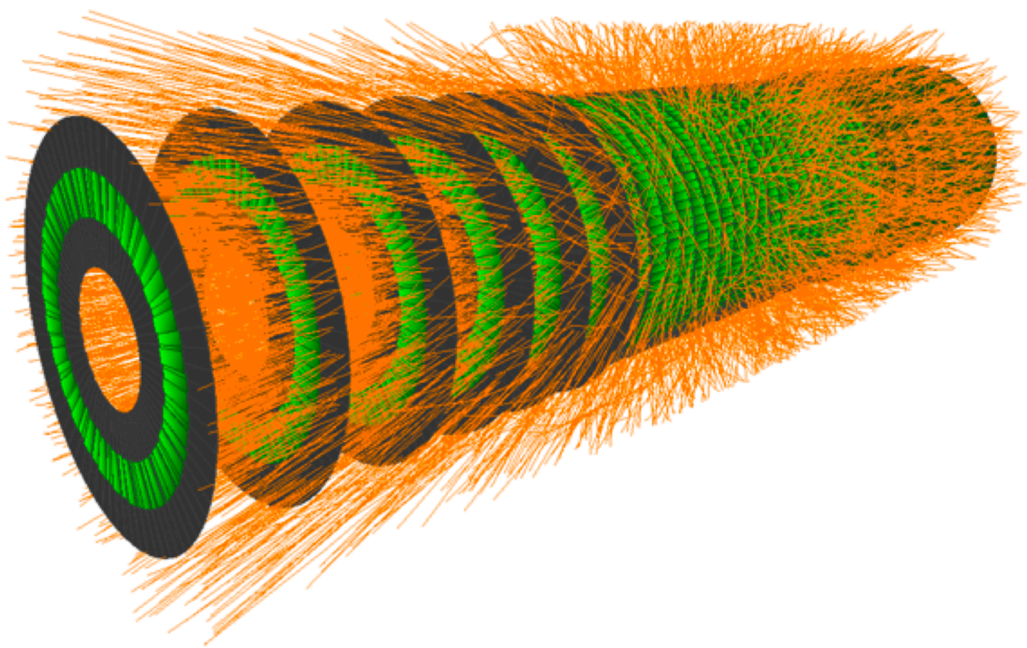
MNRAS 000, 000–000 (2022)

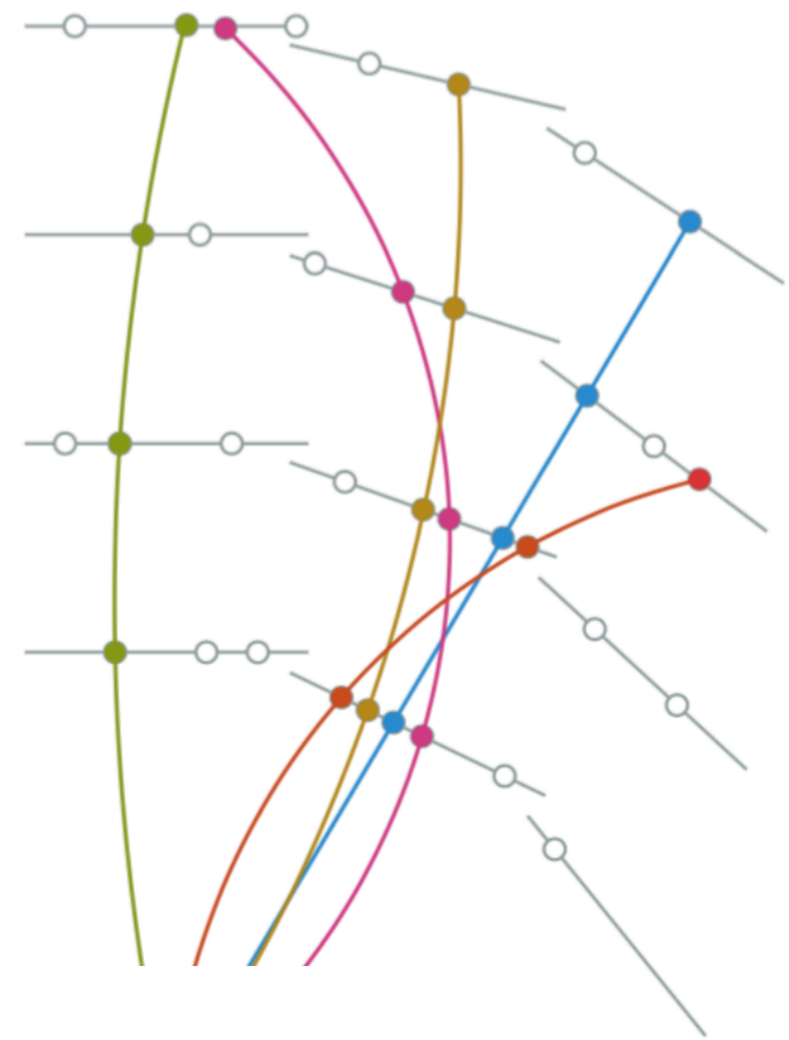
SR model is capable of generating merger histories that are solely dependent on on time-consistent LR input

Tracking

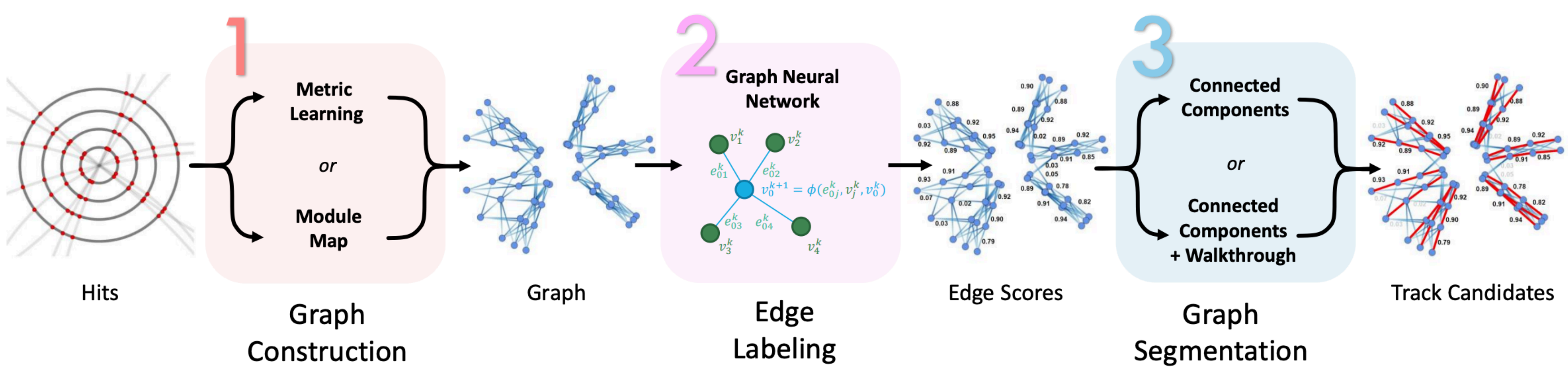
$f_{\{\theta\}}($  $)$

Tracking & ML

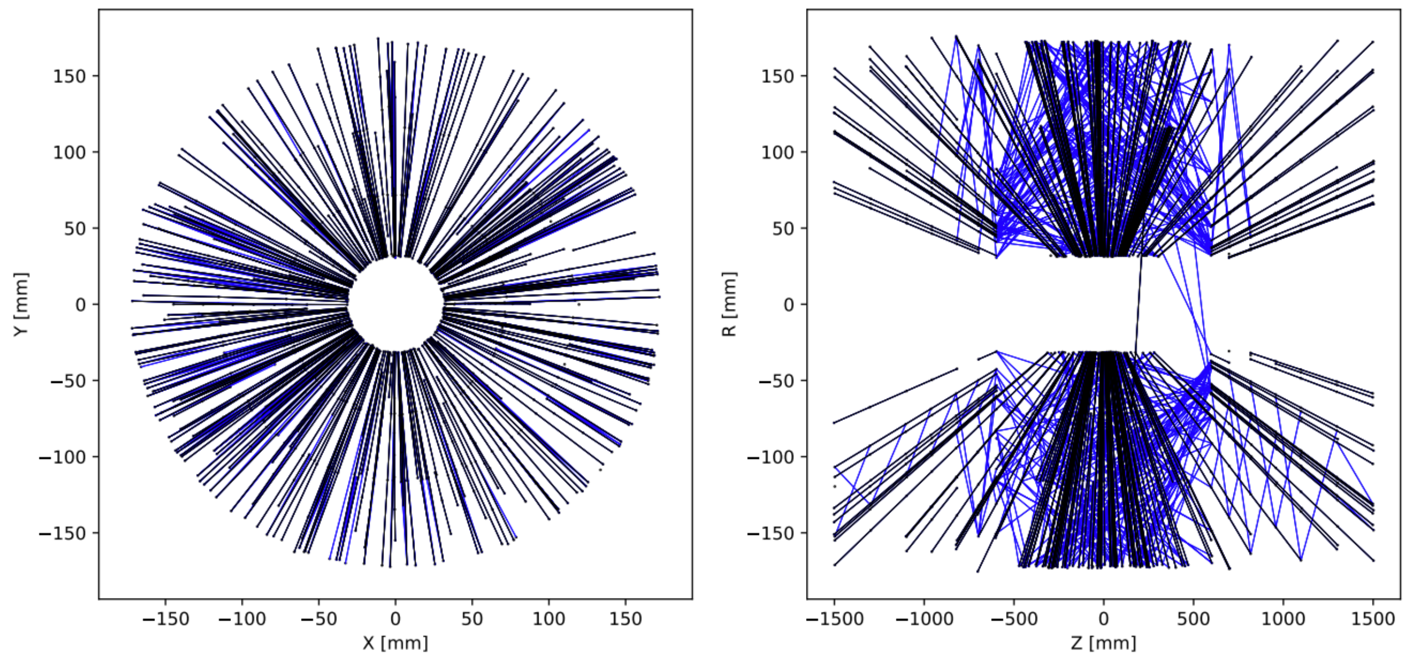


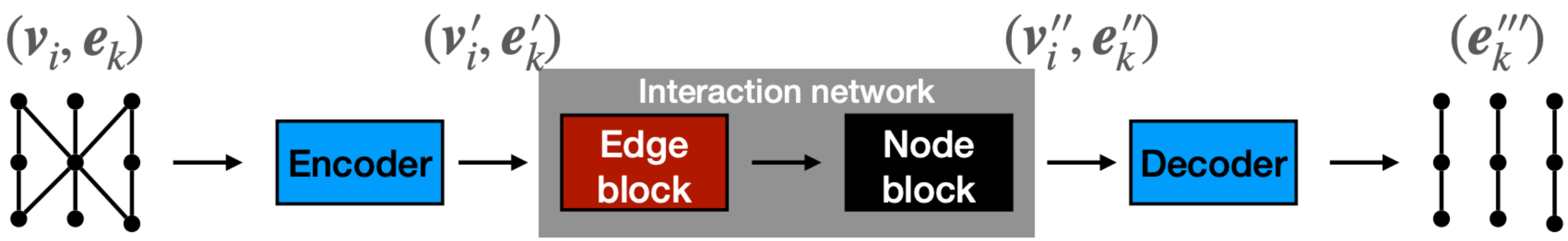


An exponentially large edge finding problem



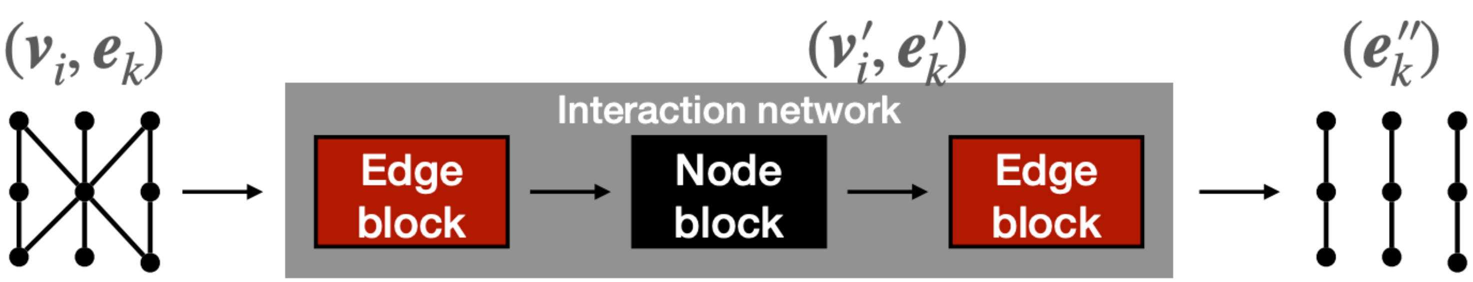
Tracking & ML





$\boldsymbol{e}'_k = \phi^e_1(\boldsymbol{e}_k)$ $\boldsymbol{e}''_k = \phi^e_2(\boldsymbol{e}'_k, \boldsymbol{v}'_{r_k}, \boldsymbol{v}'_{s_k})$ $\boldsymbol{v}''_i = \phi^v_2(\bar{\boldsymbol{e}}''_i, \boldsymbol{v}'_i)$ $\boldsymbol{e}'''_k = \phi^e_3(\boldsymbol{e}''_k)$

$\boldsymbol{v}'_i = \phi^v_1(\boldsymbol{v}_i)$ $\bar{\boldsymbol{e}}''_i = \rho^{e\to v}(E'_i)$

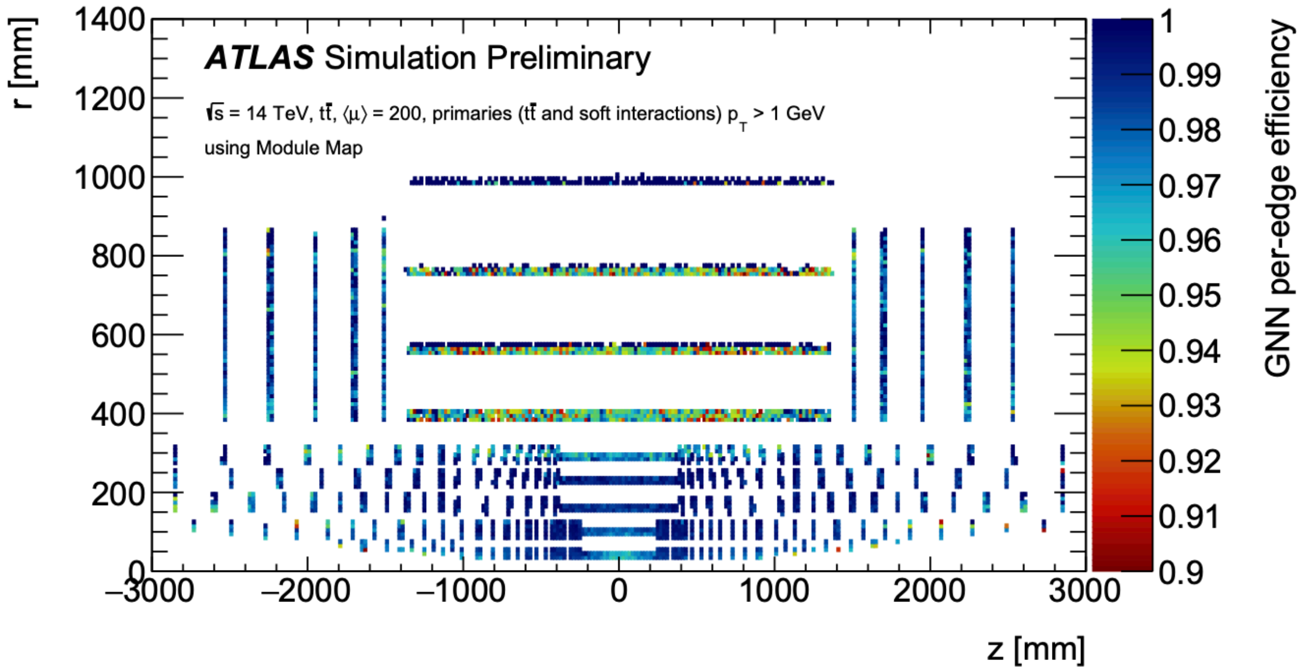


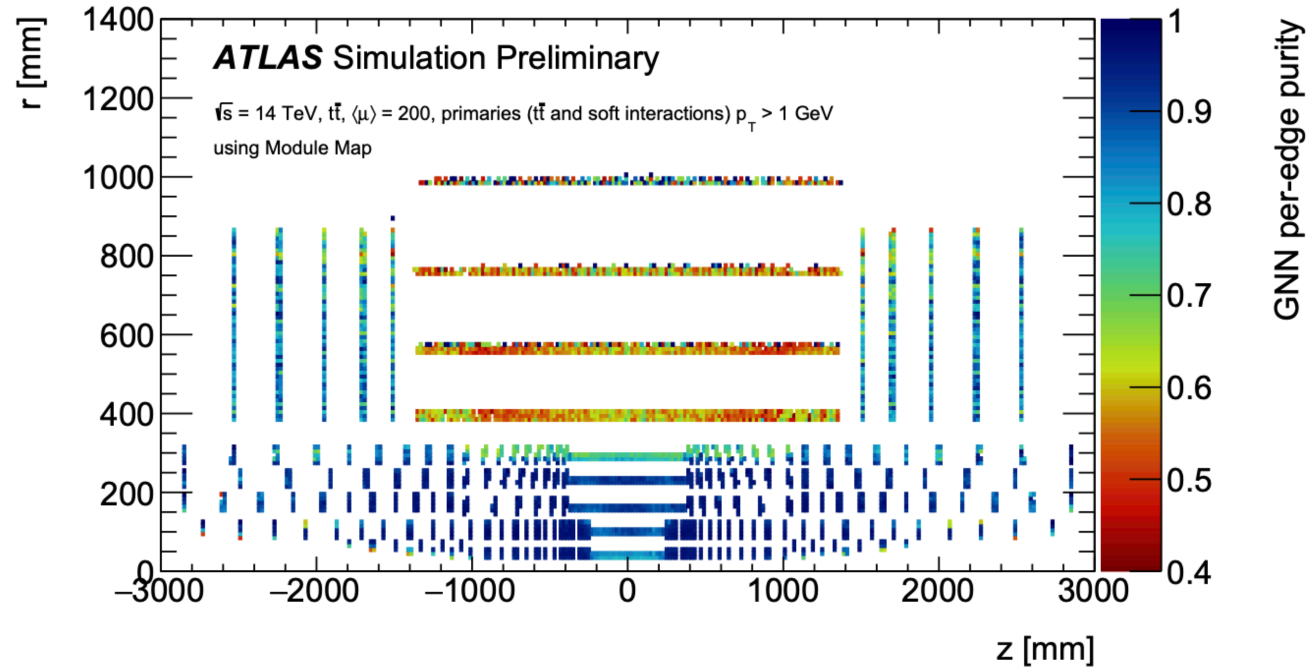
$\boldsymbol{e}'_k = \phi^e_2(\boldsymbol{e}_k, \boldsymbol{v}_{r_k}, \boldsymbol{v}_{s_k})$ $\boldsymbol{v}'_i = \phi^v_2(\bar{\boldsymbol{e}}'_i, \boldsymbol{v}_i)$ $\boldsymbol{e}''_k = \phi^e_2(\boldsymbol{e}'_k, \boldsymbol{v}'_{r_k}, \boldsymbol{v}'_{s_k})$

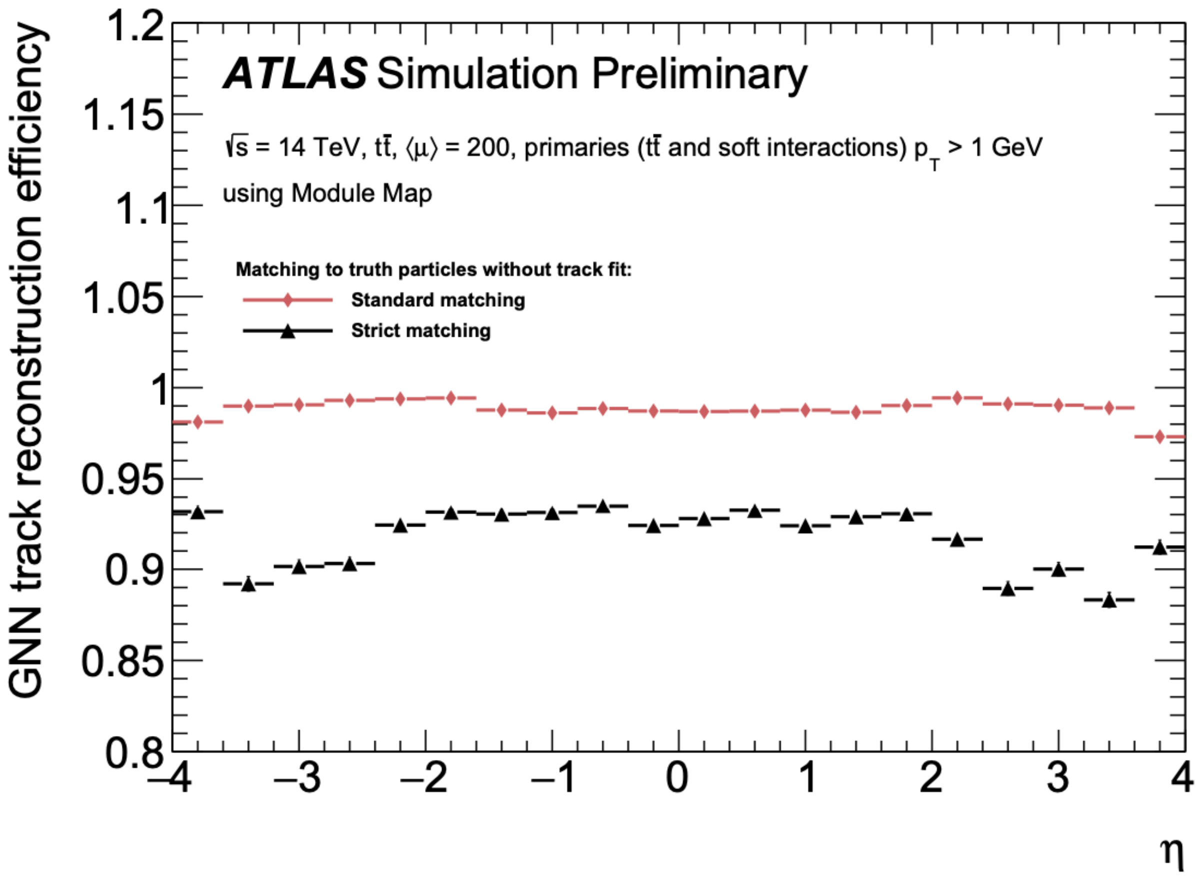
$\bar{\boldsymbol{e}}'_i = \rho^{e\to v}(E_i)$

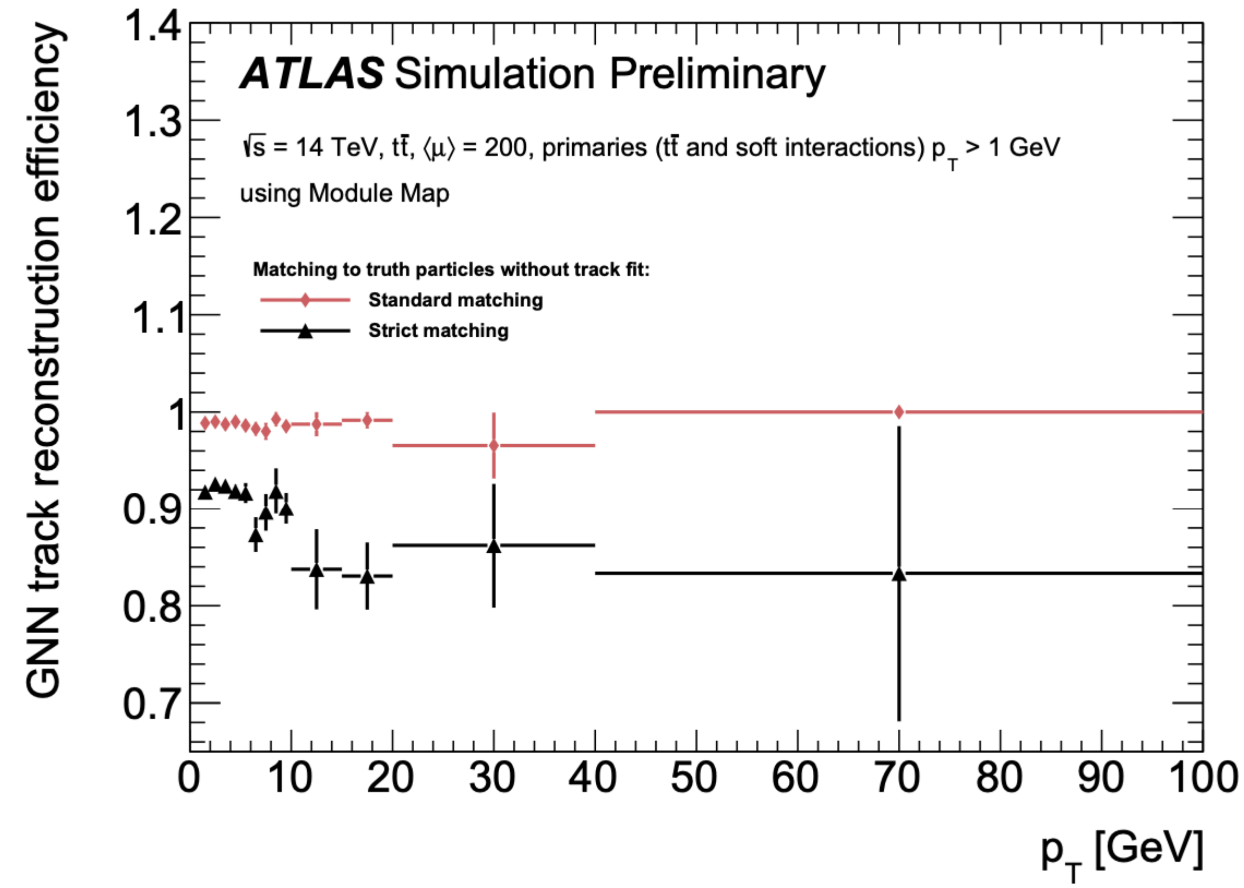
Tracking & ML : ATLAS

ATL-ITK-PROC-2022-006



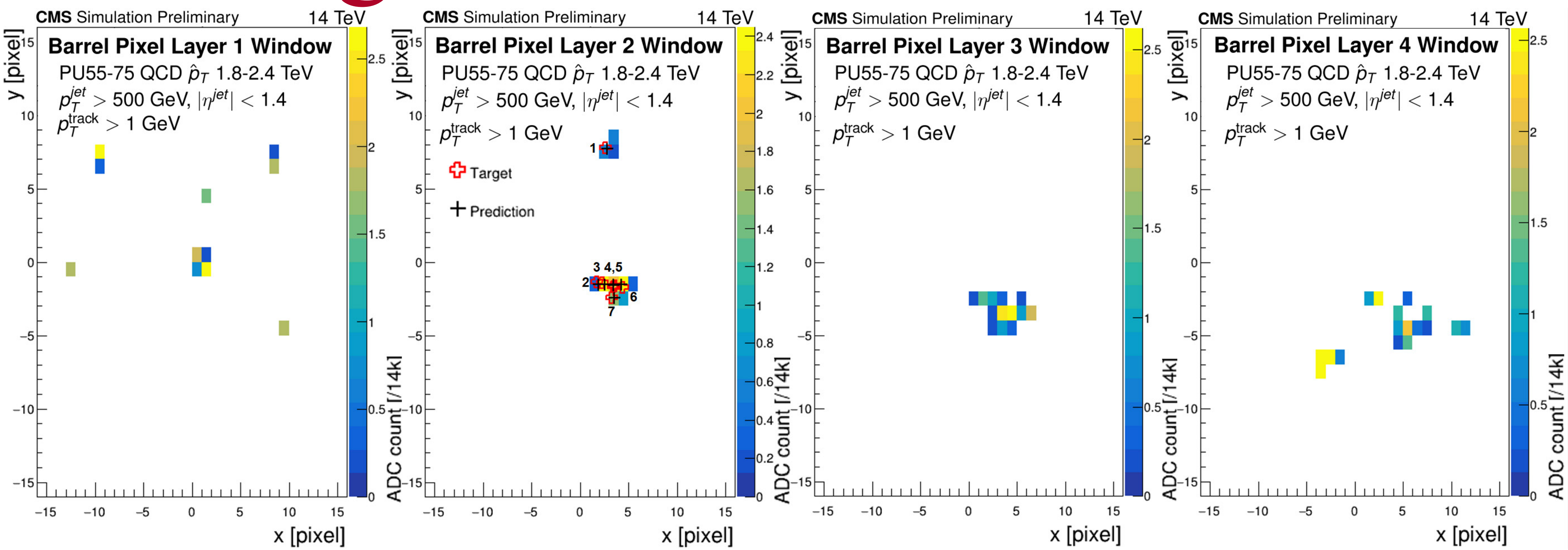


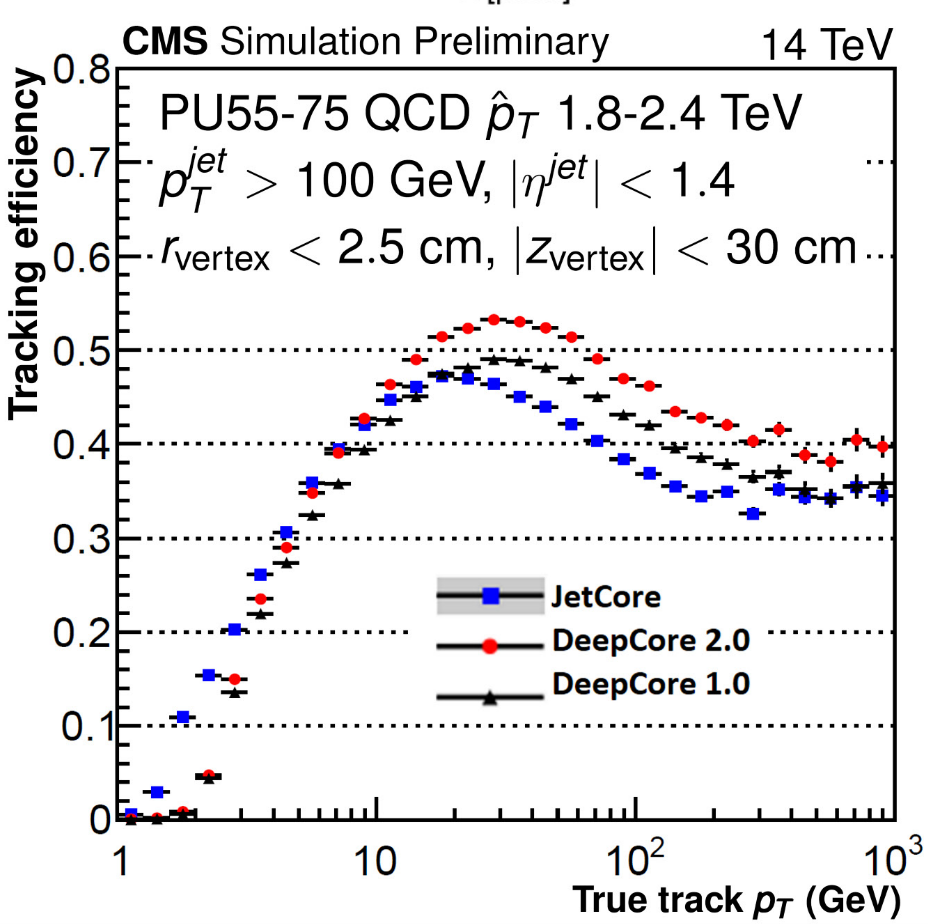


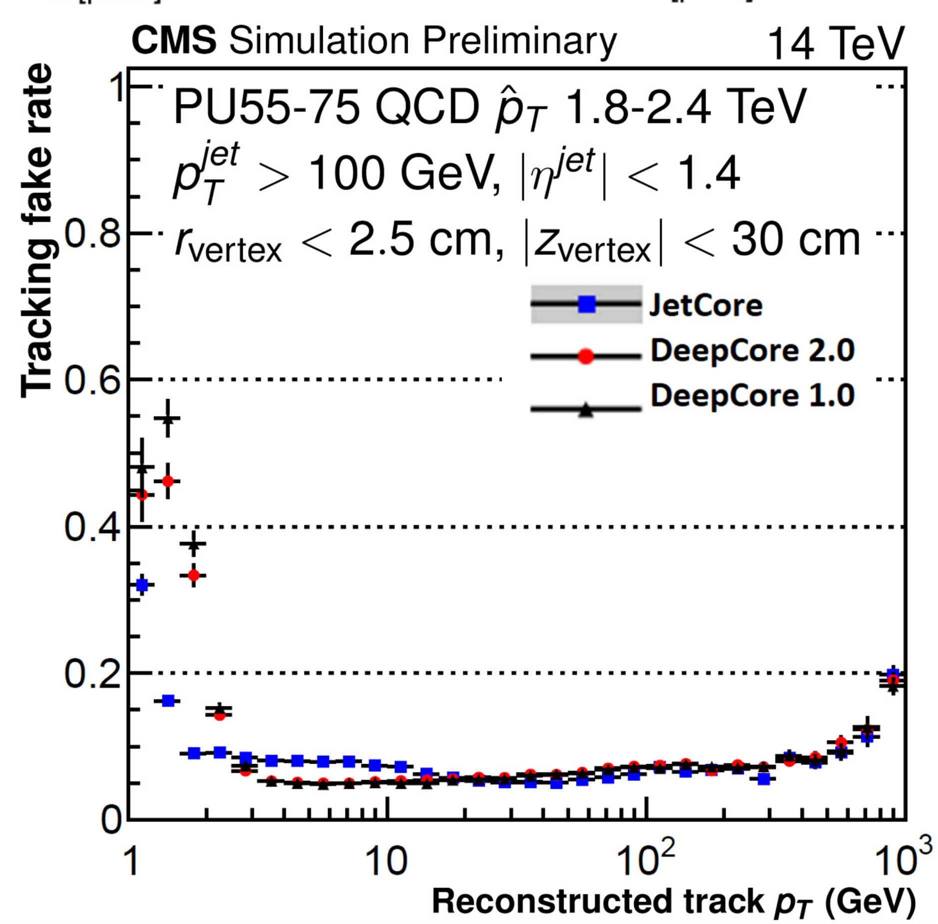


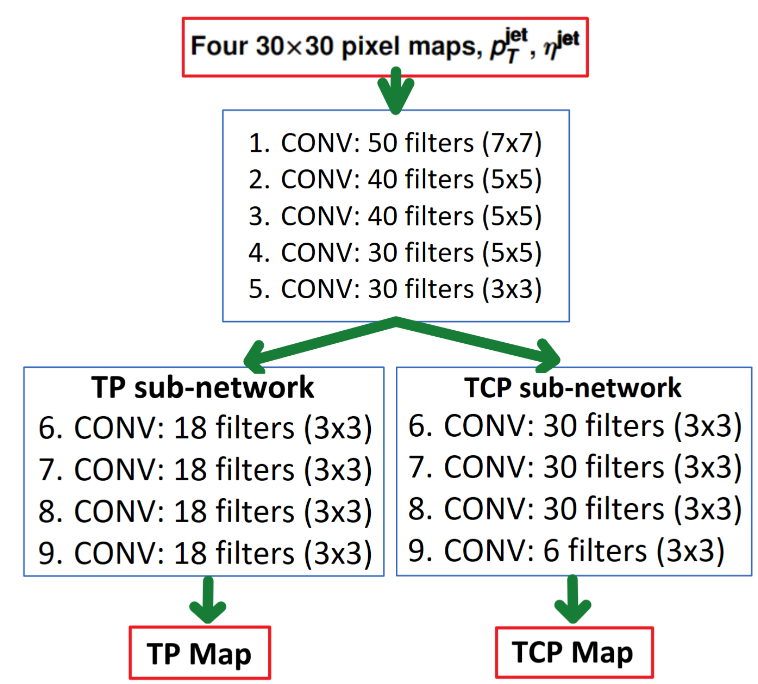
Tracking & ML : CMS

https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGRun3DeepCoreV2



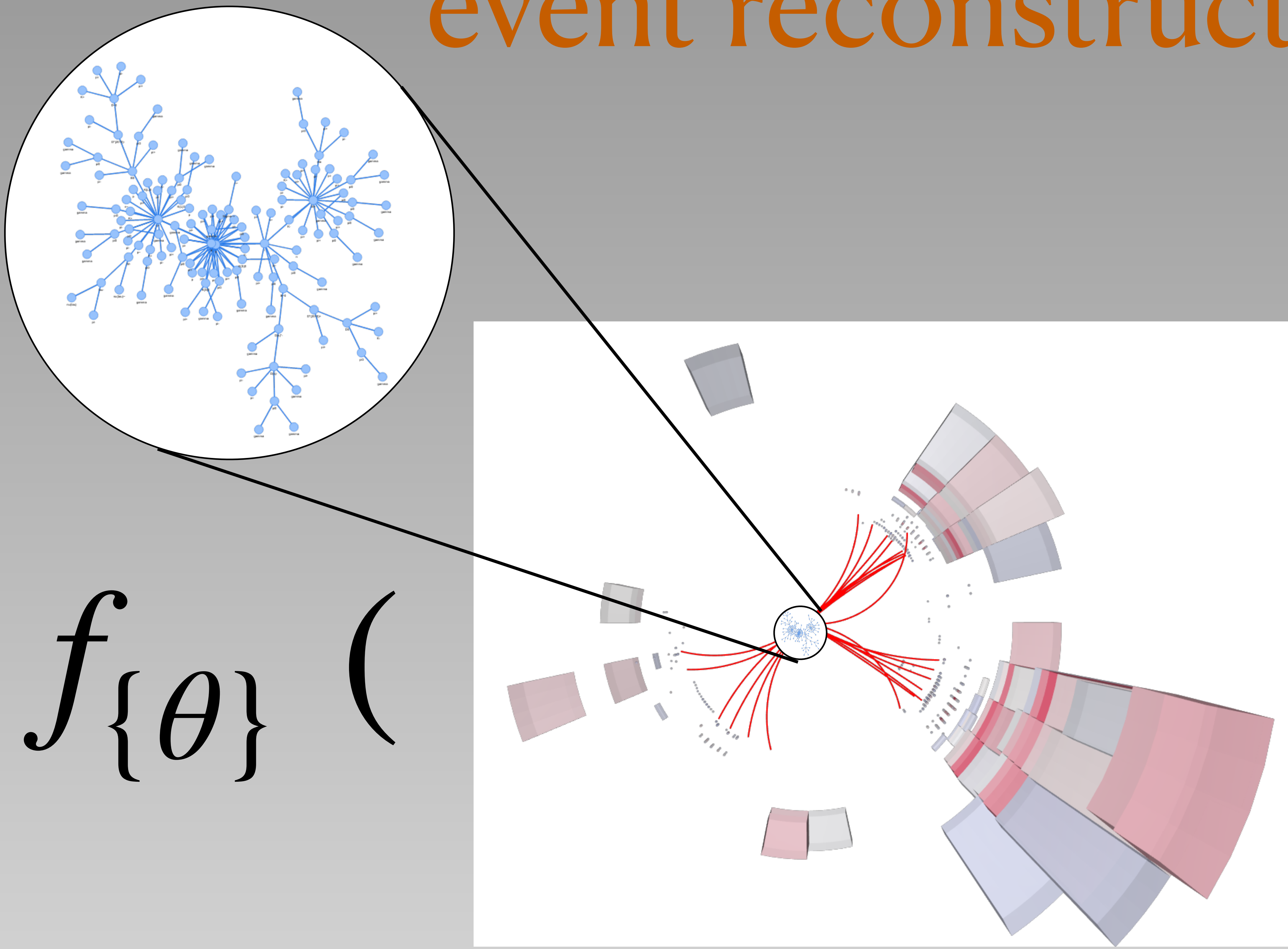






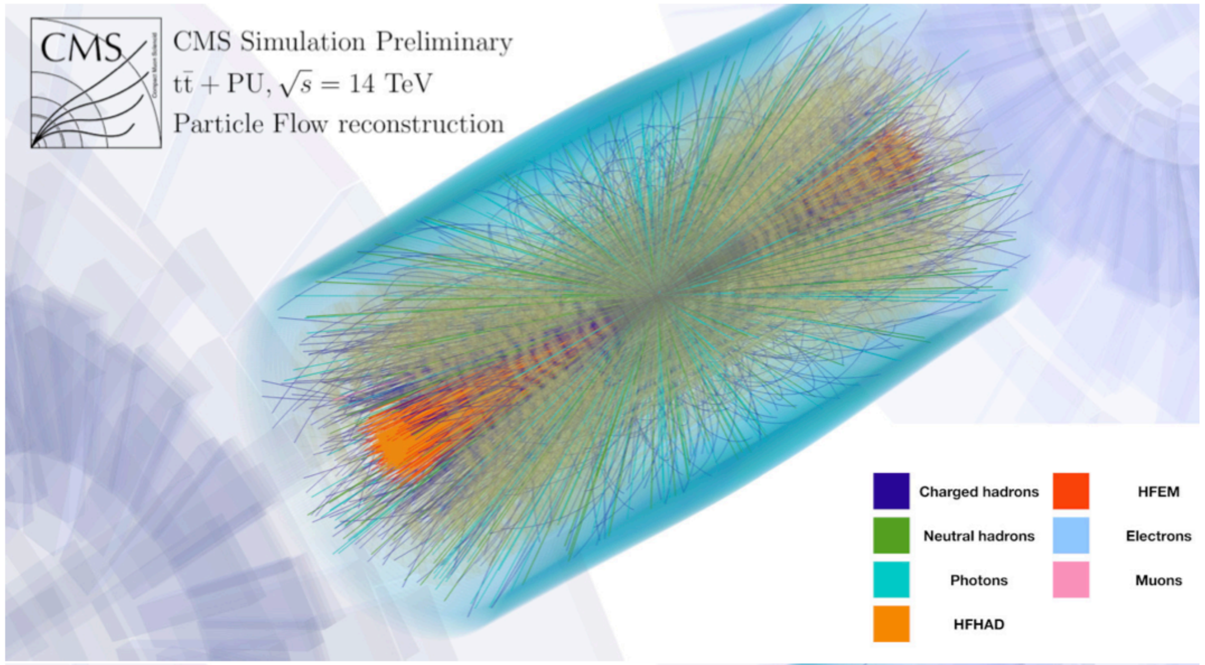
Particle-Flow event reconstruction

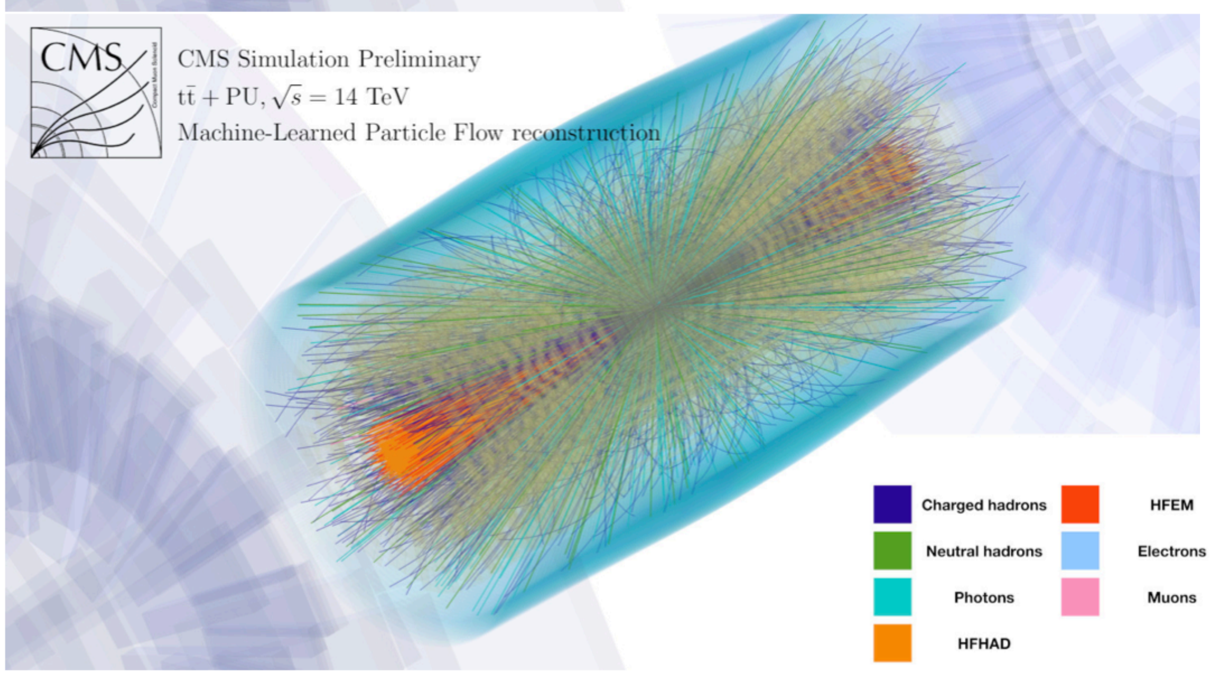
$f_{\{\theta\}}$ (

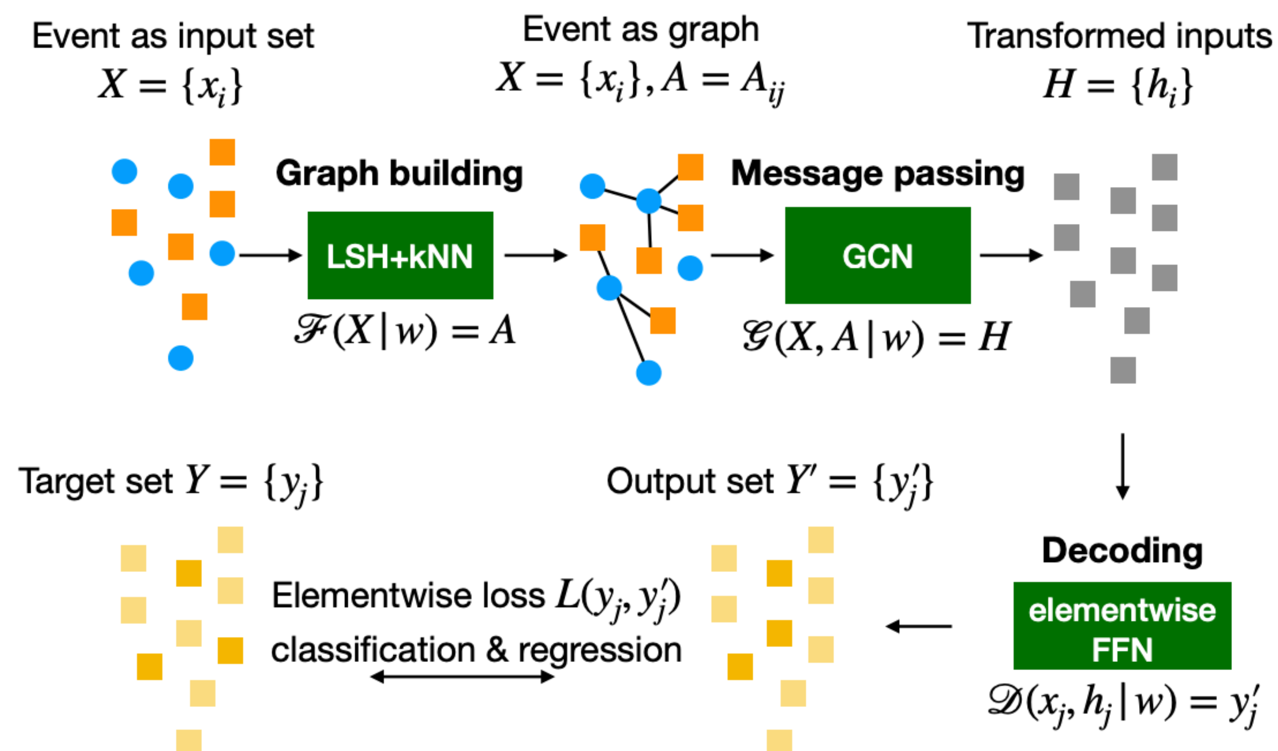


)

Full ML driven PFlow : MLPF







$x_i = [\text{type}, p_T, E_{ECAL}, E_{HCAL}, \eta, \phi, \eta_{outer}, \phi_{outer}, q, \ldots]$, type $\in$ {track, cluster}

$y_j = [\text{PID}, p_T, E, \eta, \phi, q, \ldots]$, PID $\in$ {none, charged hadron, neutral hadron, $\gamma, e^{\pm}, \mu^{\pm}$}

$h_i \in \mathbb{R}^{256}$

Trainable neural networks: $\mathscr{F}, \mathscr{G}, \mathscr{D}$

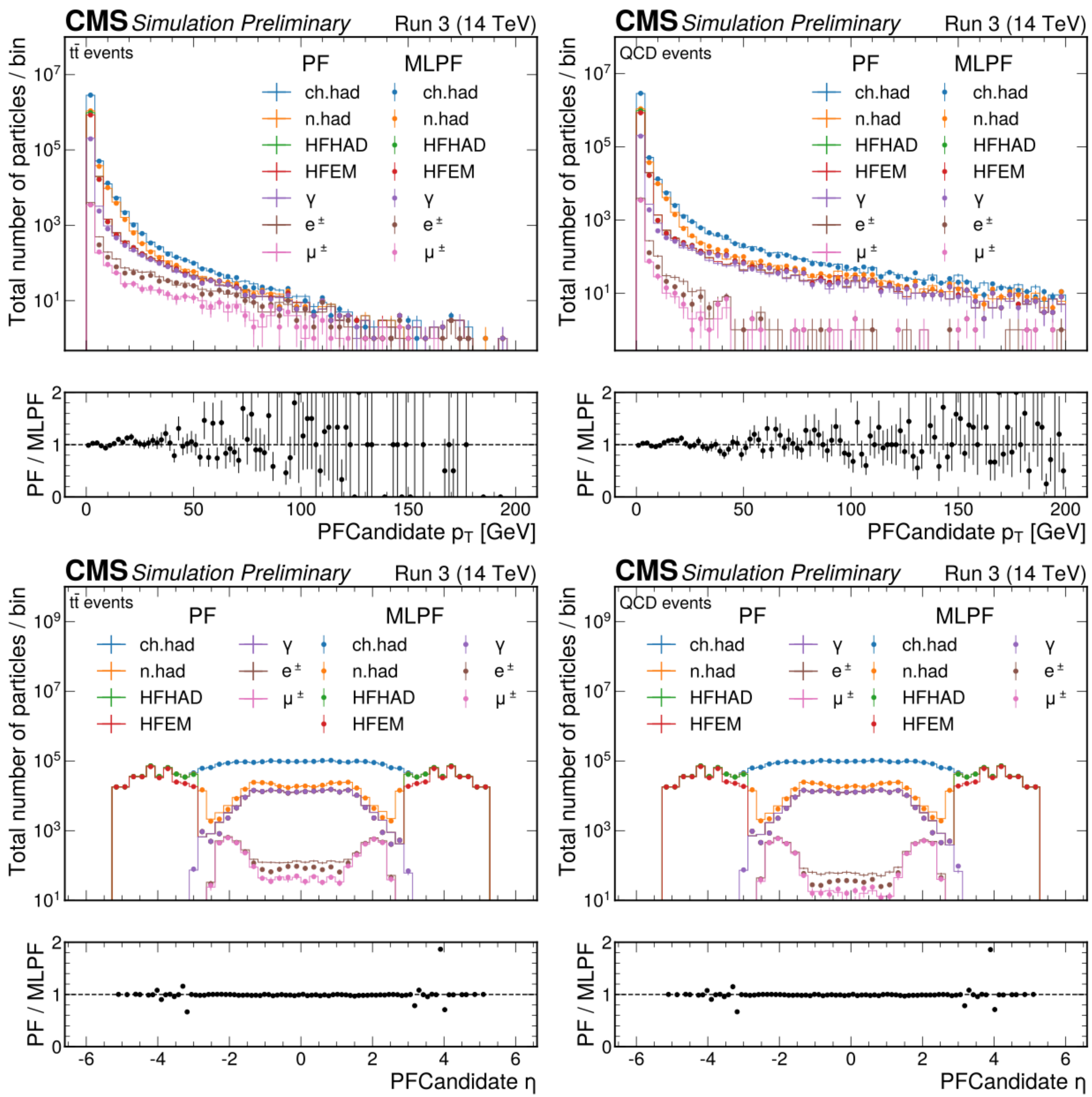
● - track, ■ - calorimeter cluster, ■ - encoded element

■ - target (predicted) particle, ■ - no target (predicted) particle

MLPF
Eur. Phys. J. C (2021) 81: 381
J. Pata et. al.

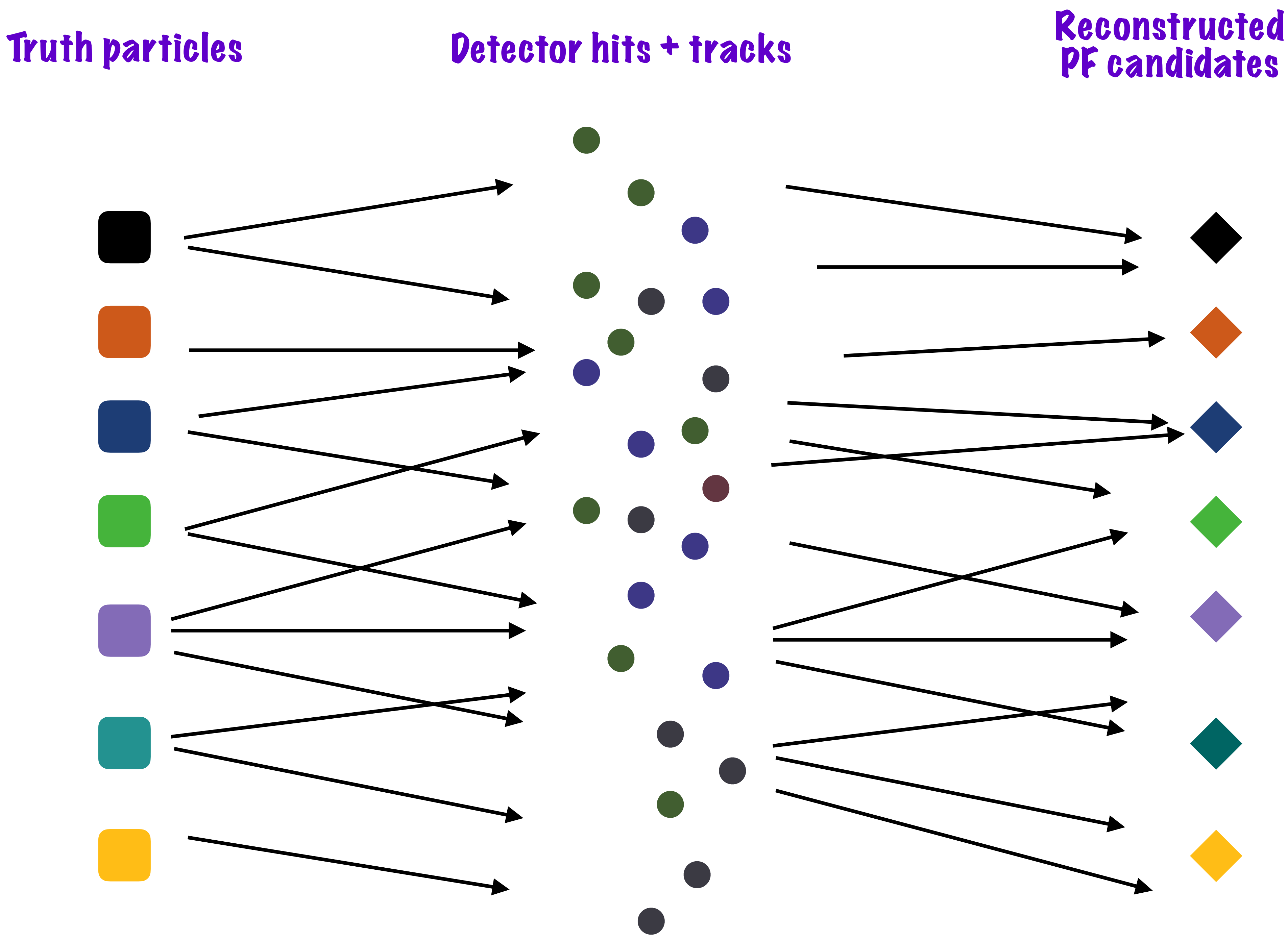
PF lepton, hadron, photon = F_{PF} (track hits + calo cells)

Combining track + calo for PFlow

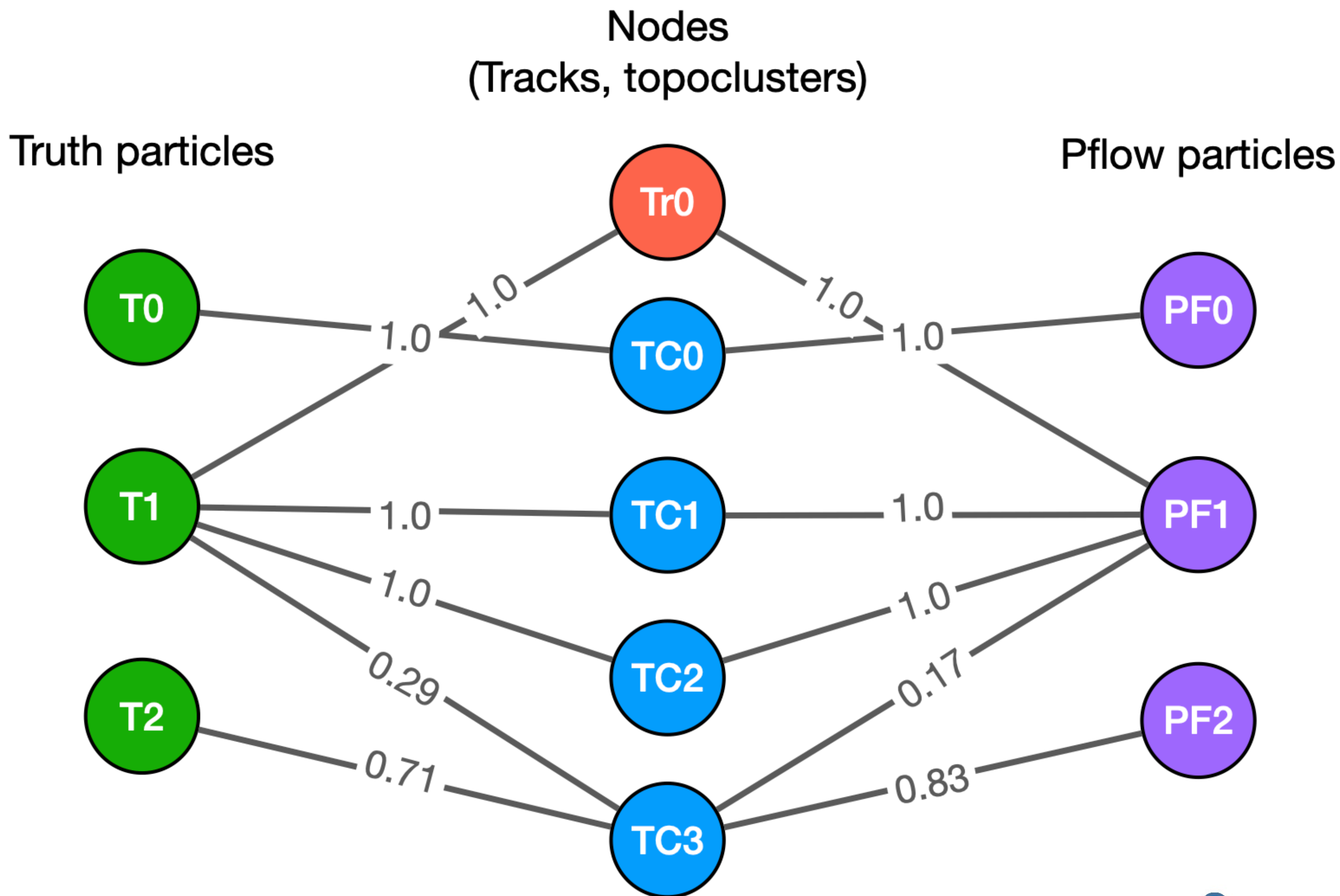


MLPF
arXiv : 2203.00330
J. Pata et. al.

What's the core data structure?

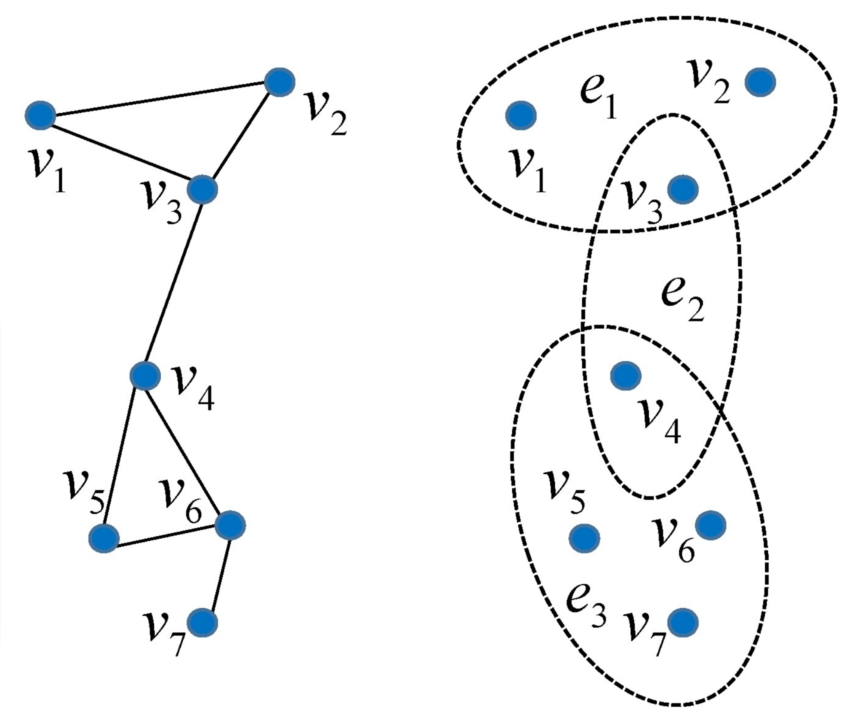


What's the core data structure?



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

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

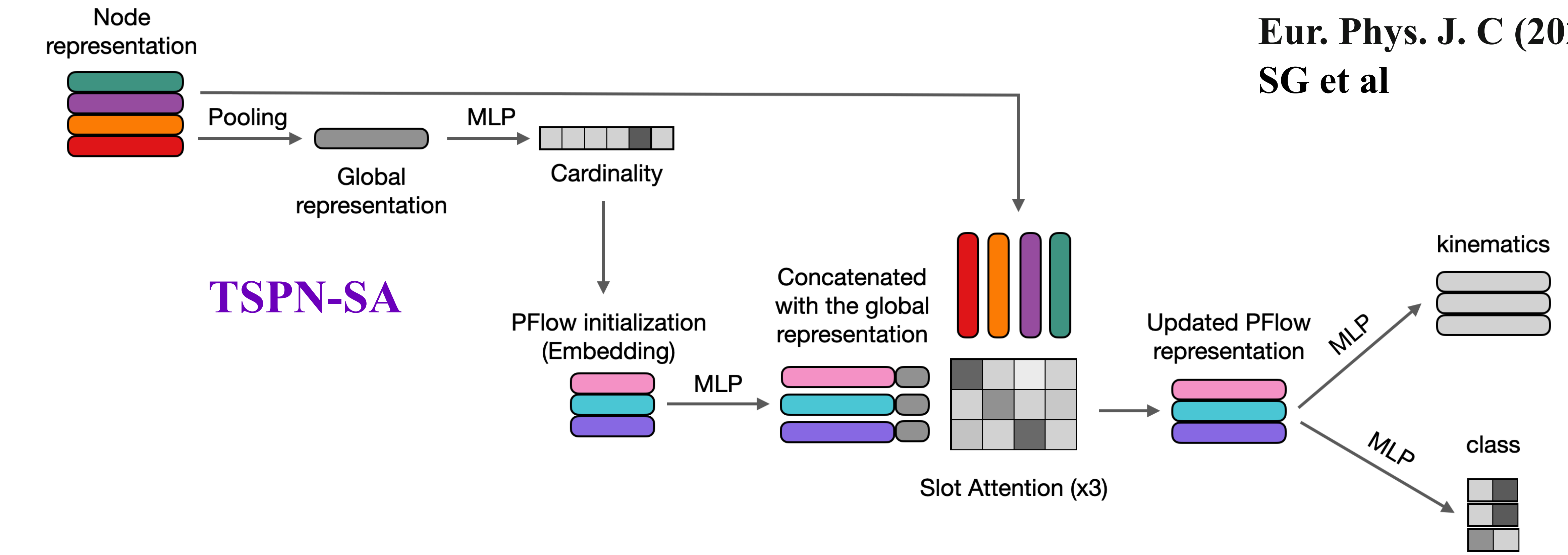


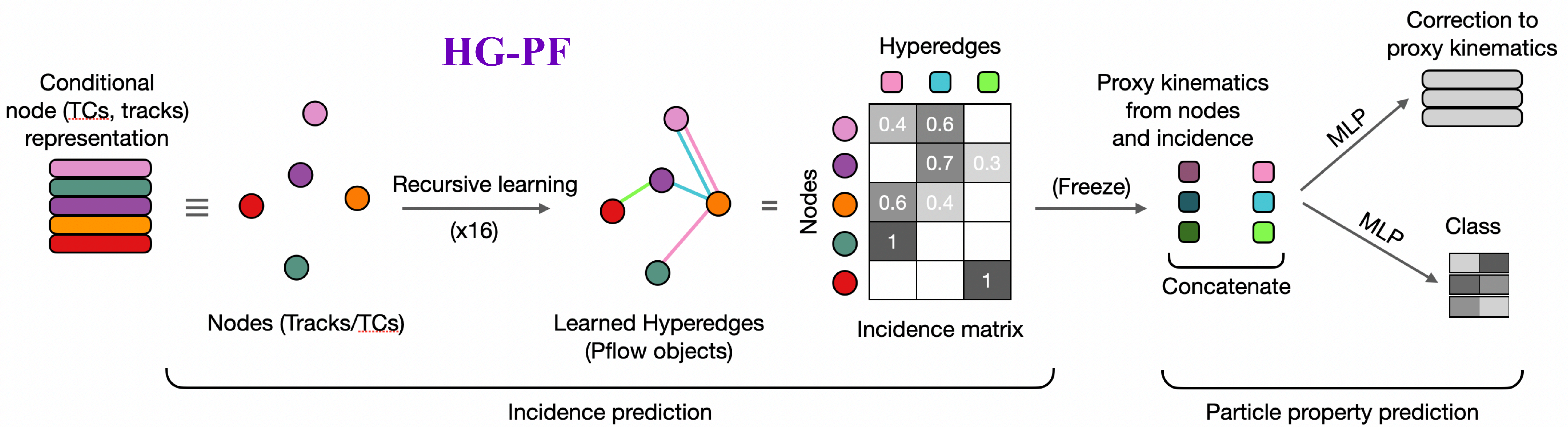
	e_1	e_2	e_3
v_1	1	0	0
v_2	1	0	0
v_3	1	1	0
v_4	0	1	1
v_5	0	0	1
v_6	0	0	1
v_7	0	0	1

(a) Simple graph (b) Hypergraph **G** (c) Incidence matrix **H**

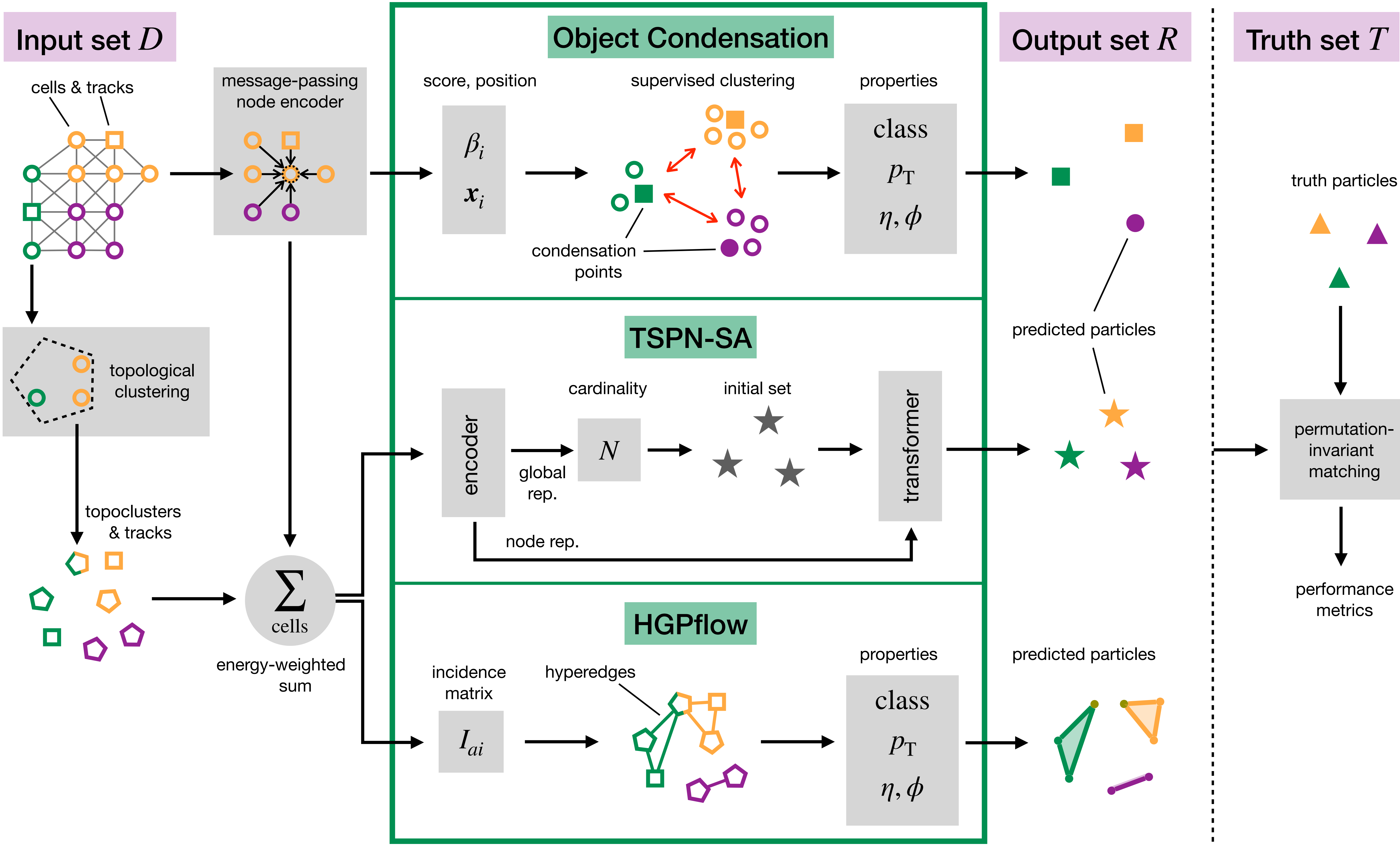
The new networks we tried

Eur. Phys. J. C (2023) 83:596
SG et al





The network flow comparisons



Design of the performance metrics

- Efficiency and fake rate
- Classification purity
- Particle angular and momentum resolution
- Jet-level quantities

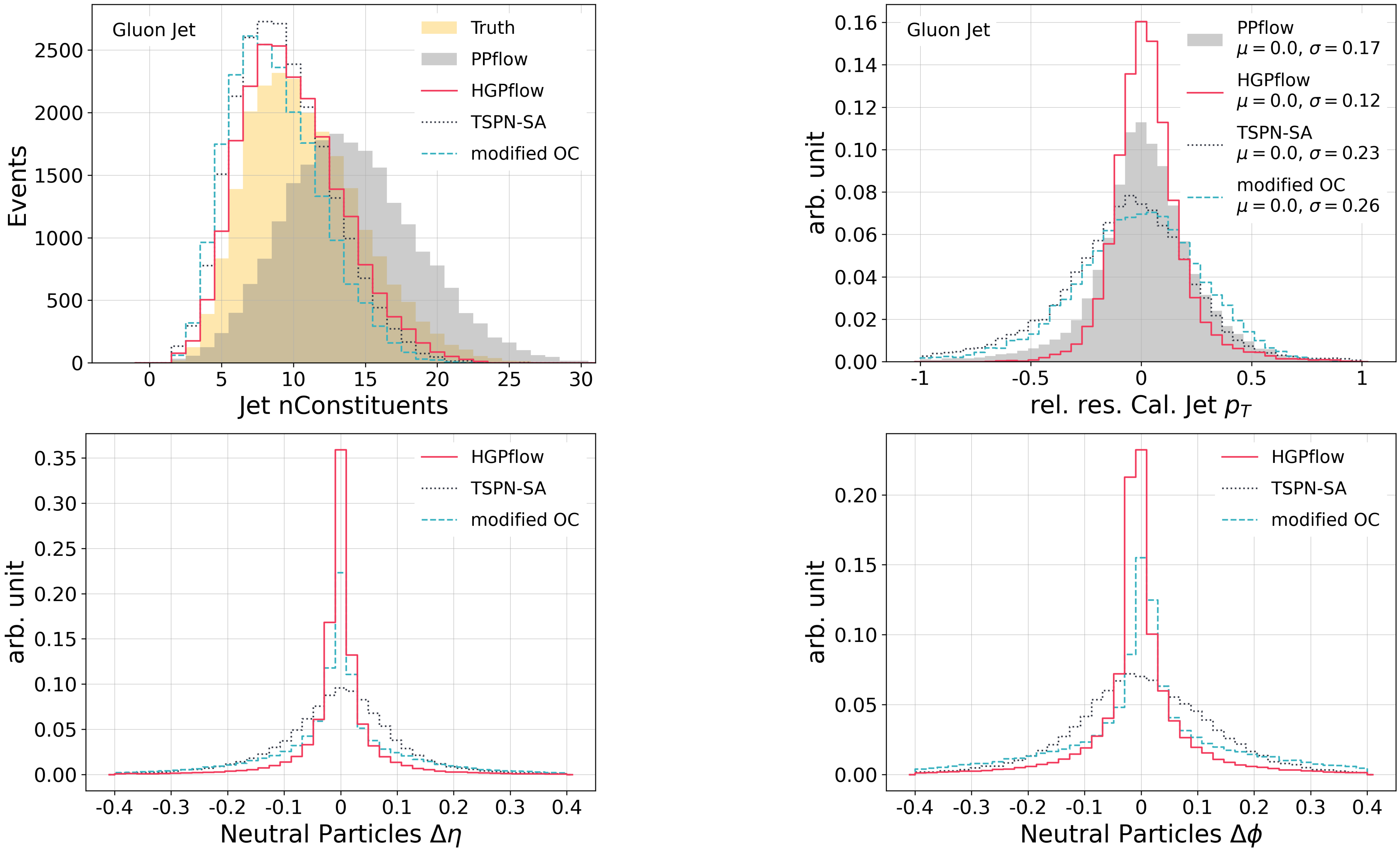
The efficiency and fake rate are defined as follows:

$$\varepsilon \equiv \frac{N(\text{matched pred})}{N(\text{targ})}, \quad f \equiv \frac{N(\text{unmatched pred})}{N(\text{pred})}$$

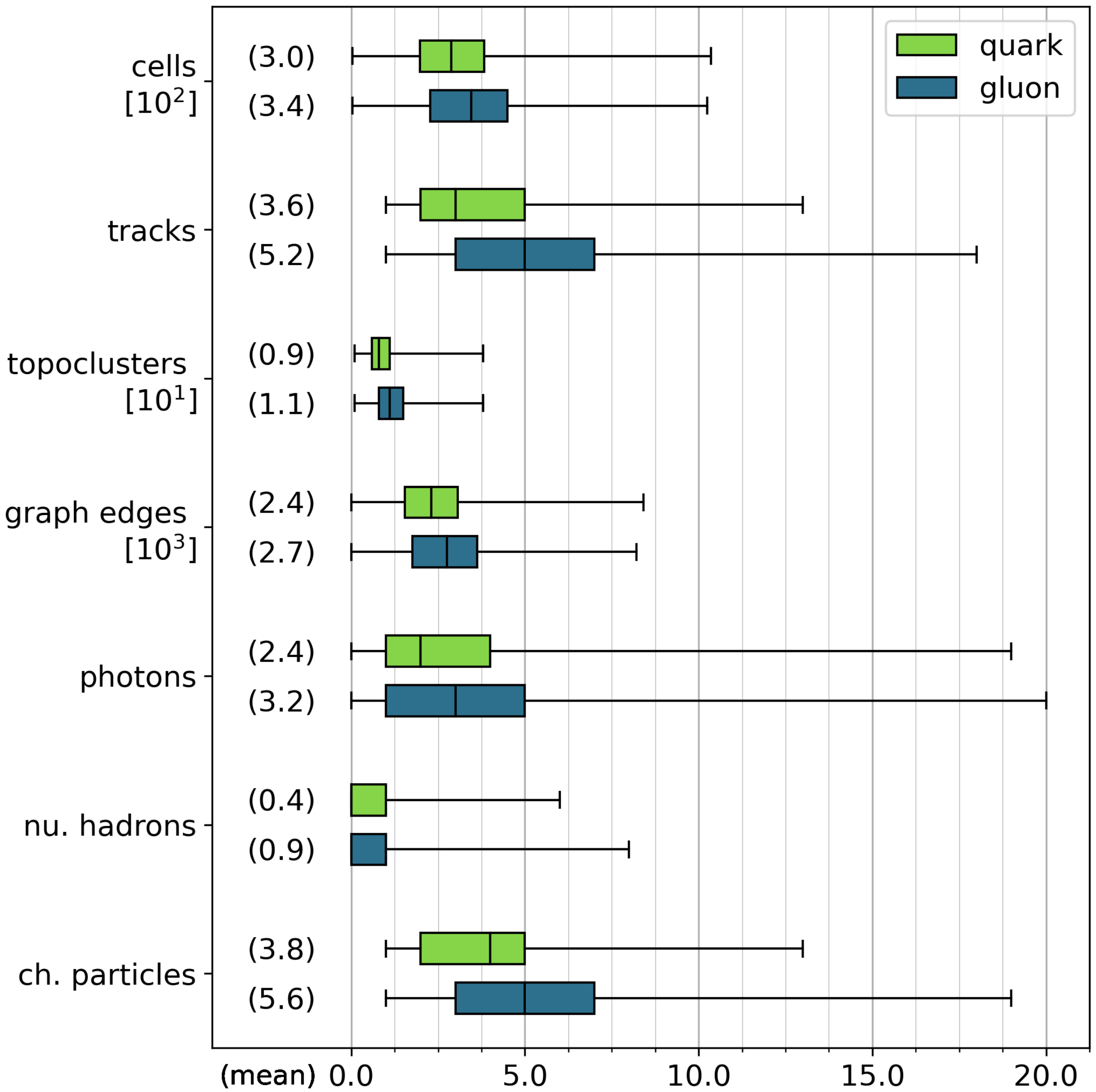
Set-to-set matching is done using the Hungarian distance :

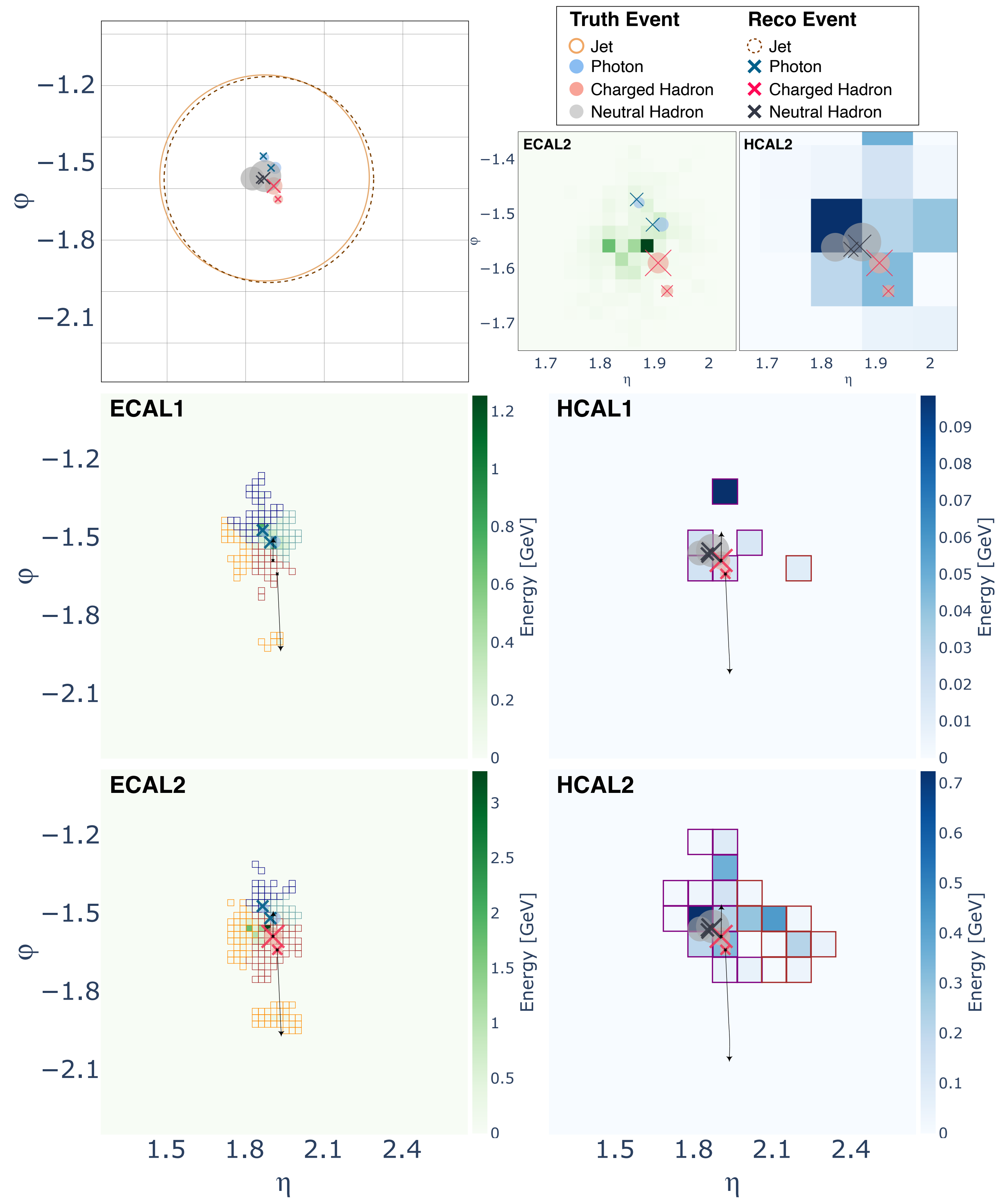
$$\sqrt{c_{p_T}(\Delta p_T/p_T^{truth})^2 + C_{\Delta R}(\Delta R)^2}$$

$$\mathcal{L}_{\text{hung}}(A, B) = \min_{\pi \in P} \sum_{a_i \in A} d(a_i, b_{\pi(i)})$$



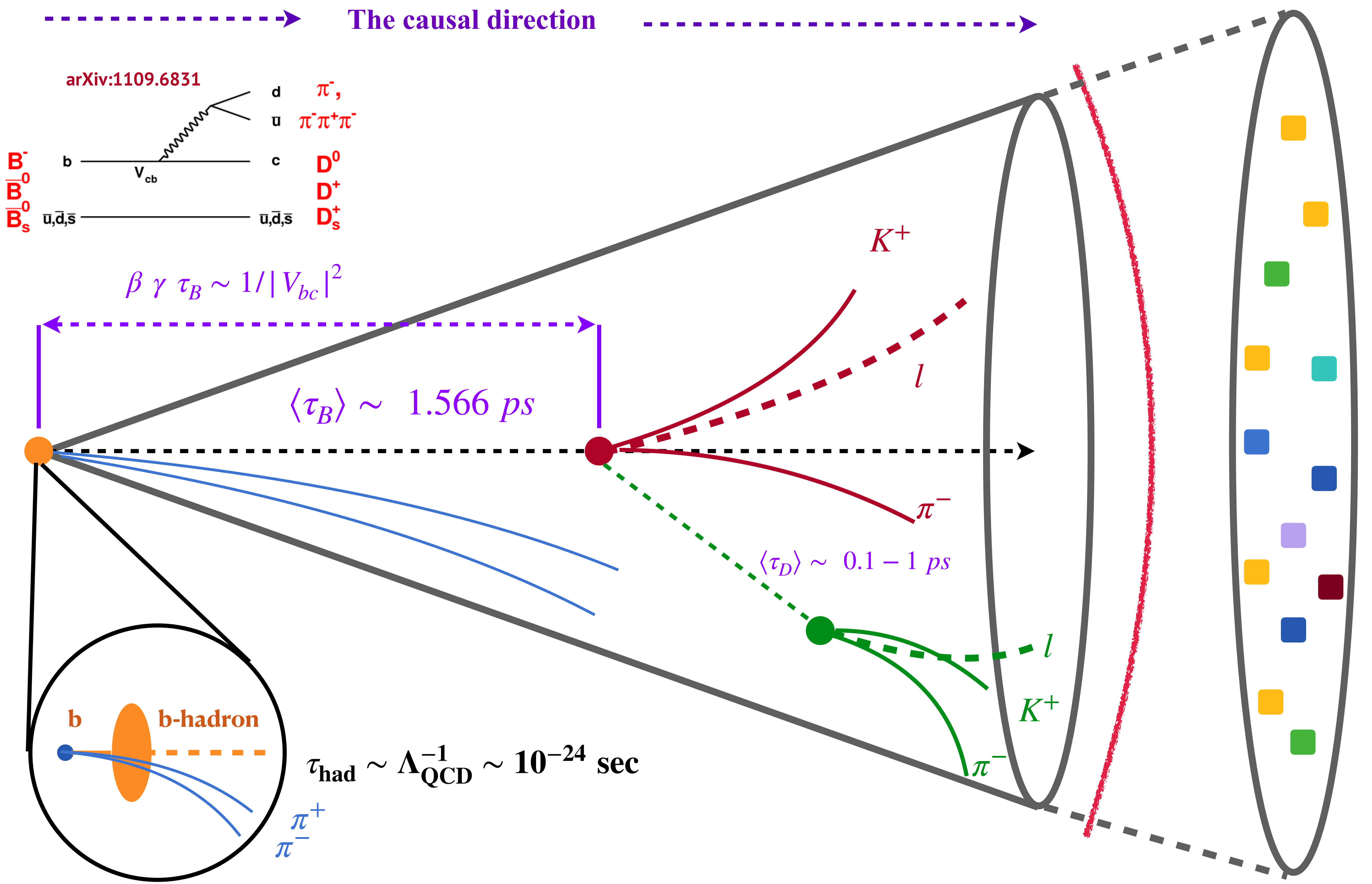
Data complexity & sample output



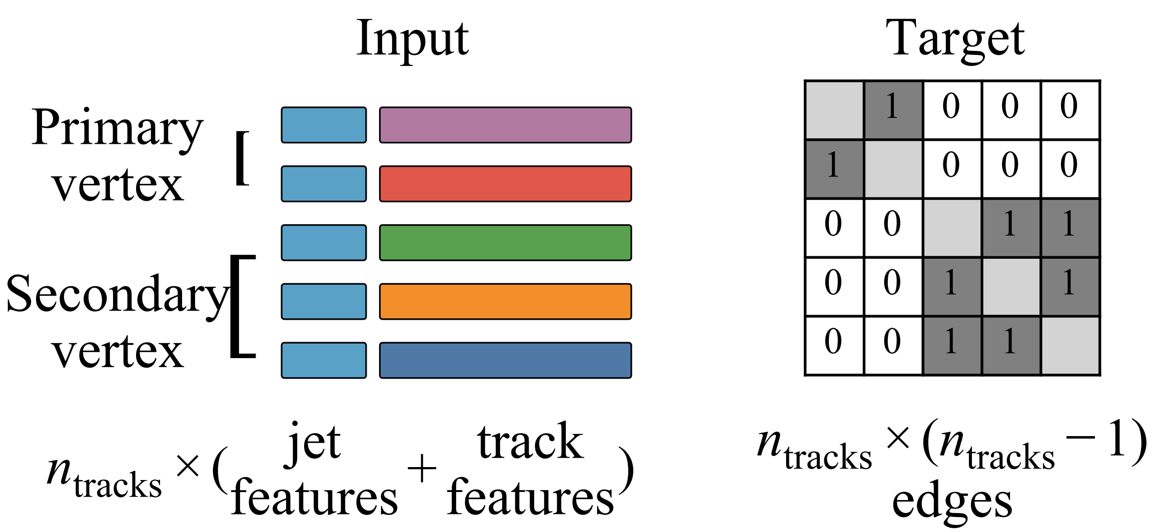


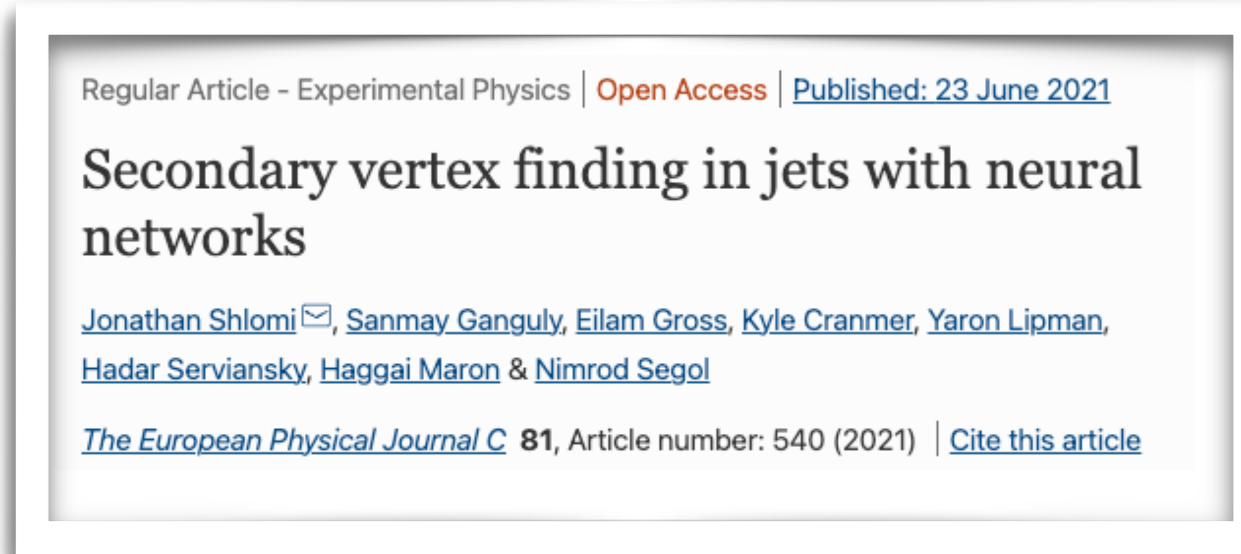
HG-PFlow seems to be doing a better job among the compared methods

Physics motivated ML: B-tagging



Set2Graph proposal for flavor-tagging

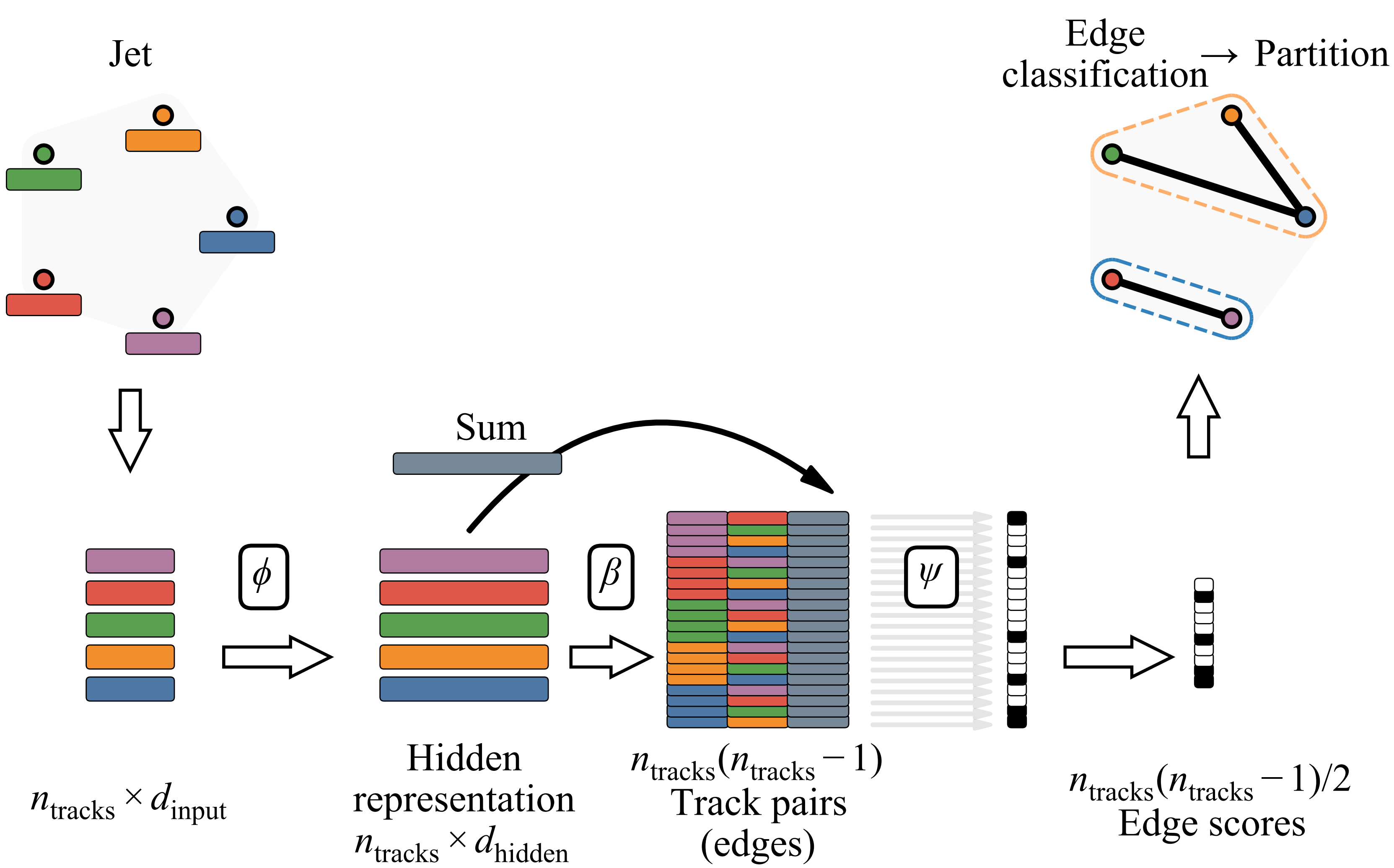



Regular Article - Experimental Physics | Open Access | Published: 23 June 2021

Secondary vertex finding in jets with neural networks

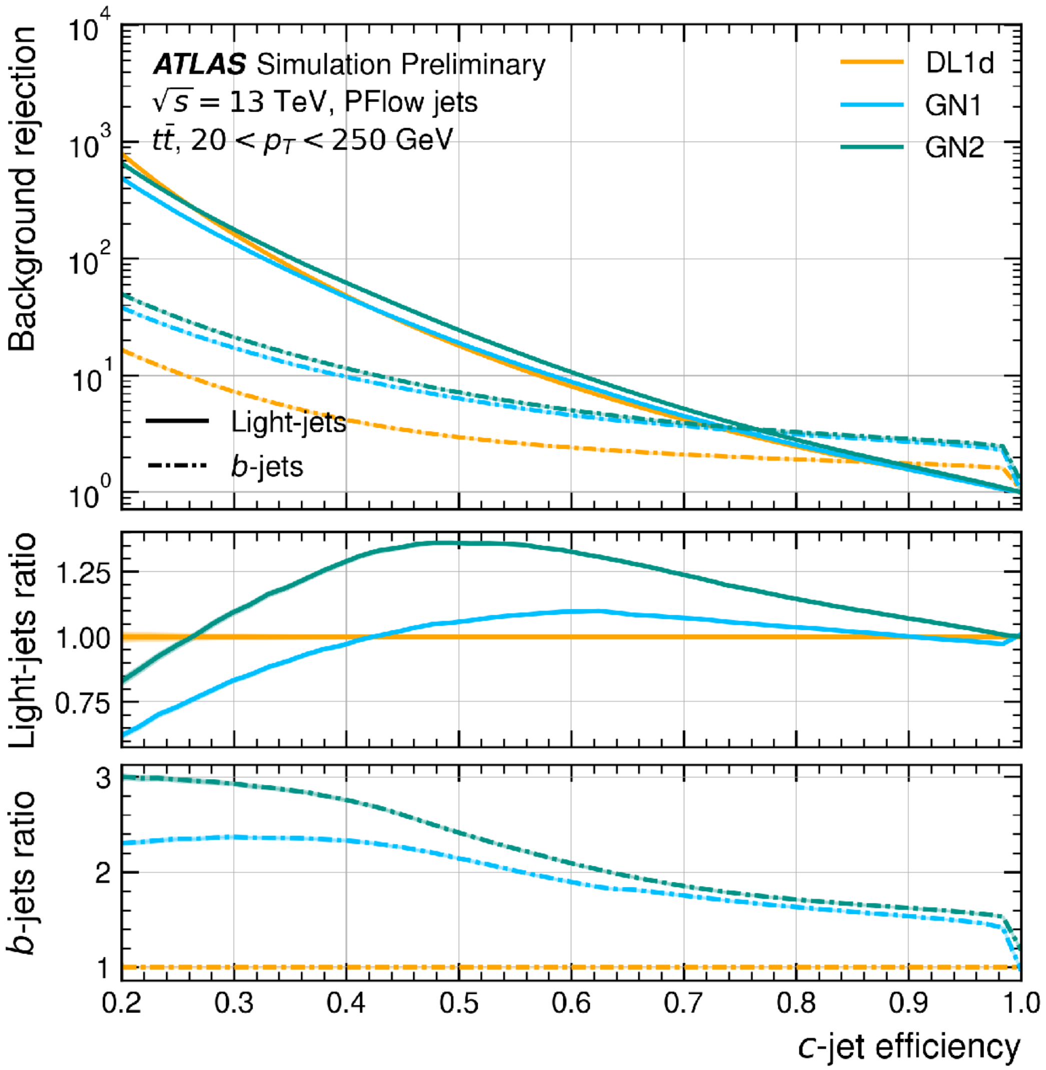
Jonathan Shlomi, Sanmay Ganguly, Eilam Gross, Kyle Cranmer, Yaron Lipman, Hadar Serviansky, Haggai Maron & Nimrod Segol

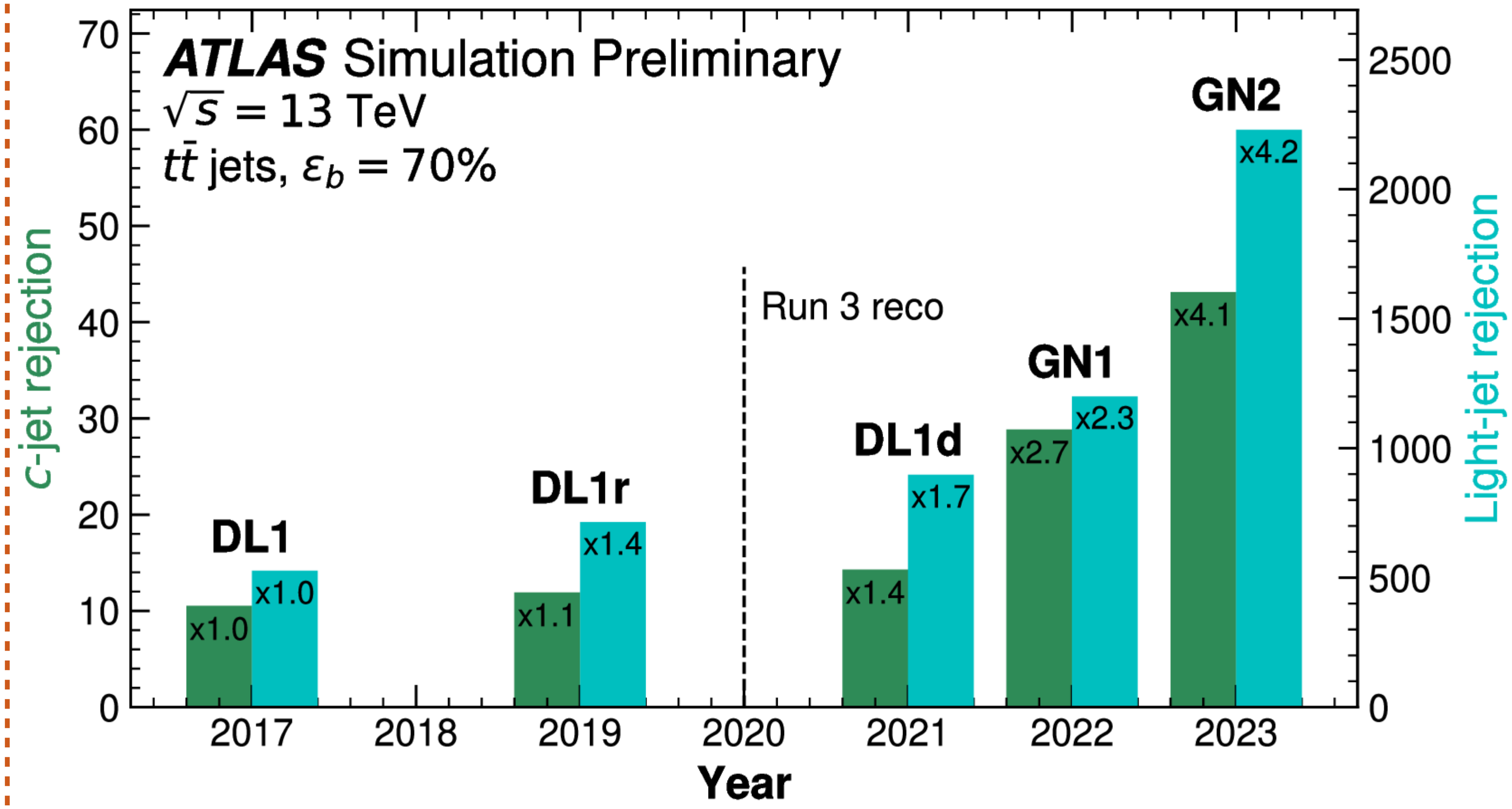
The European Physical Journal C **81**, Article number: 540 (2021) | Cite this article

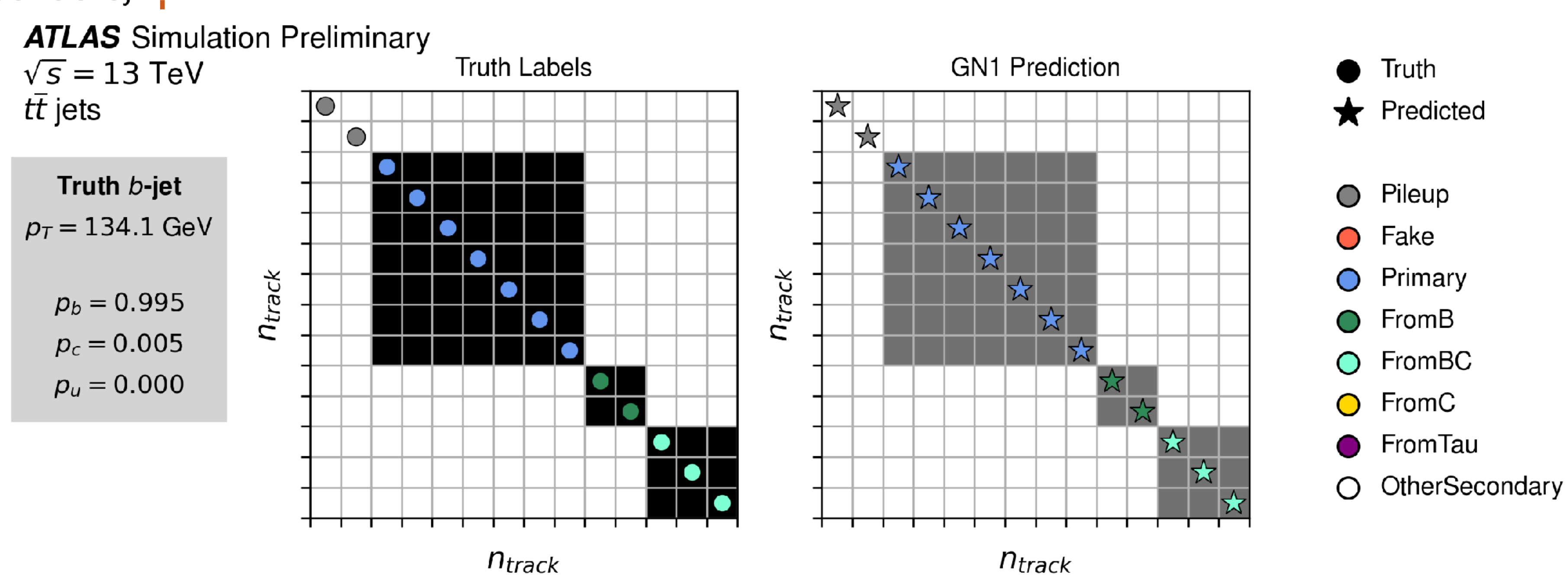


Set2Graph model within ATLAS

FTAG-2023-001





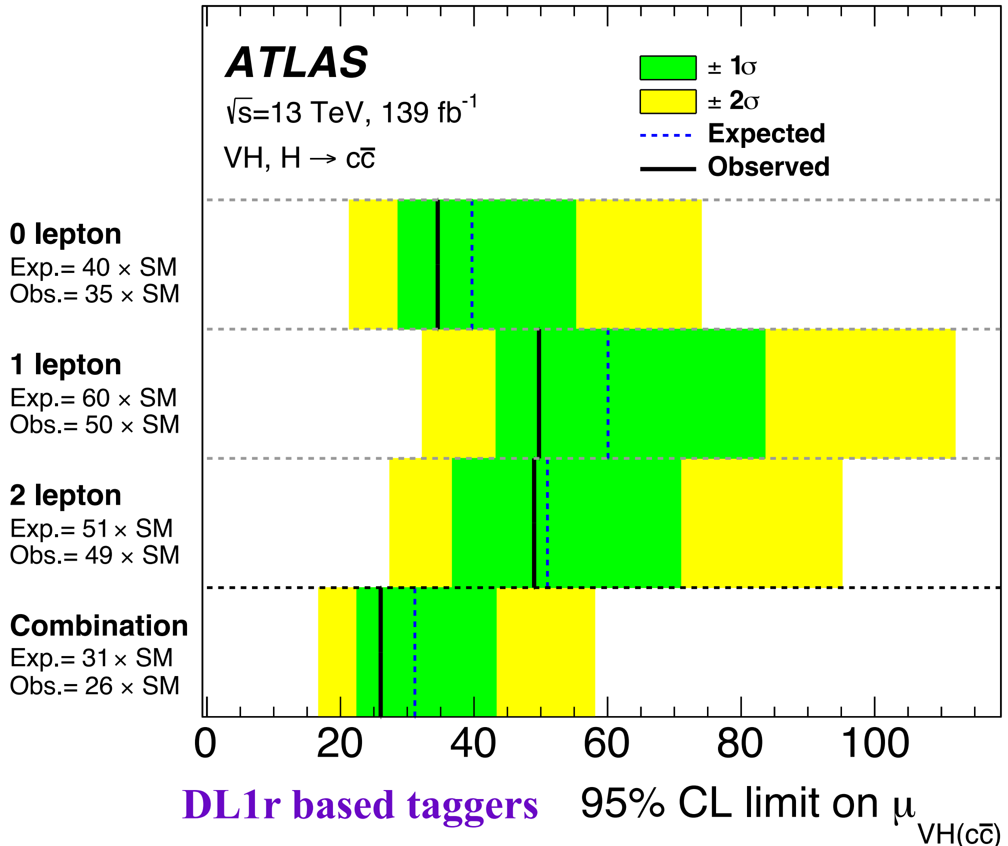


Sizable improvement over the current DL1r algorithm.

For a c-tagging working point ~ 30%, a significant gain in Rejection rate is obtained.

Direct physics application of the taggers

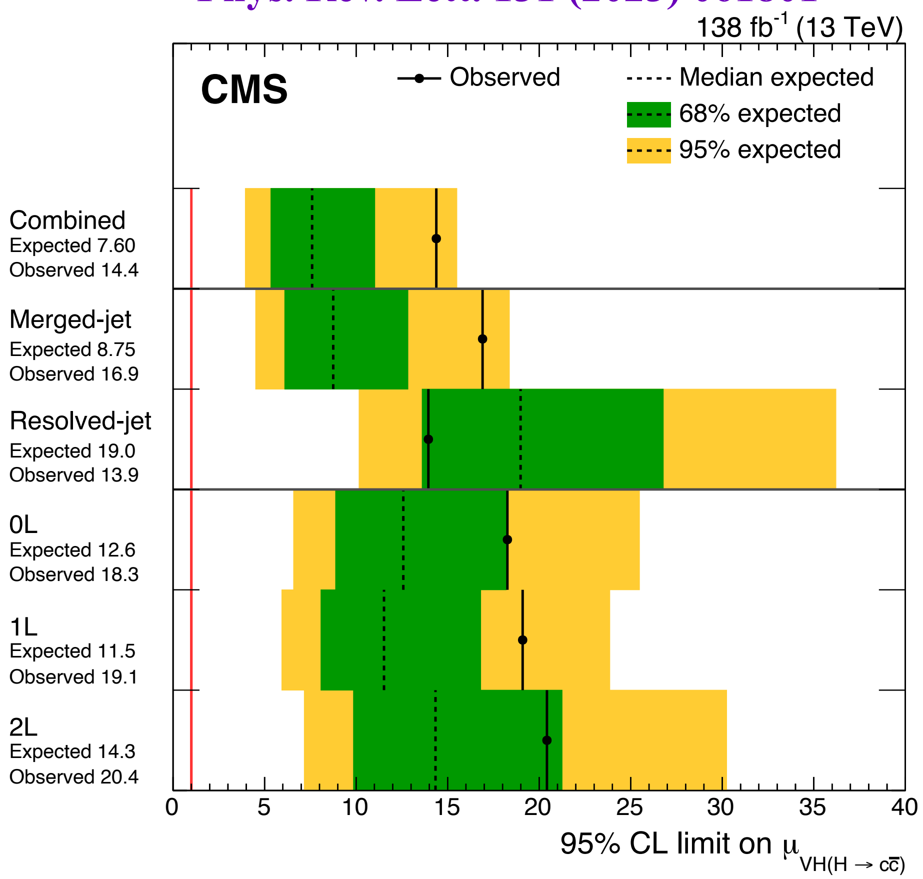
Eur. Phys. J. C (2022) 82:717



DL1r based taggers

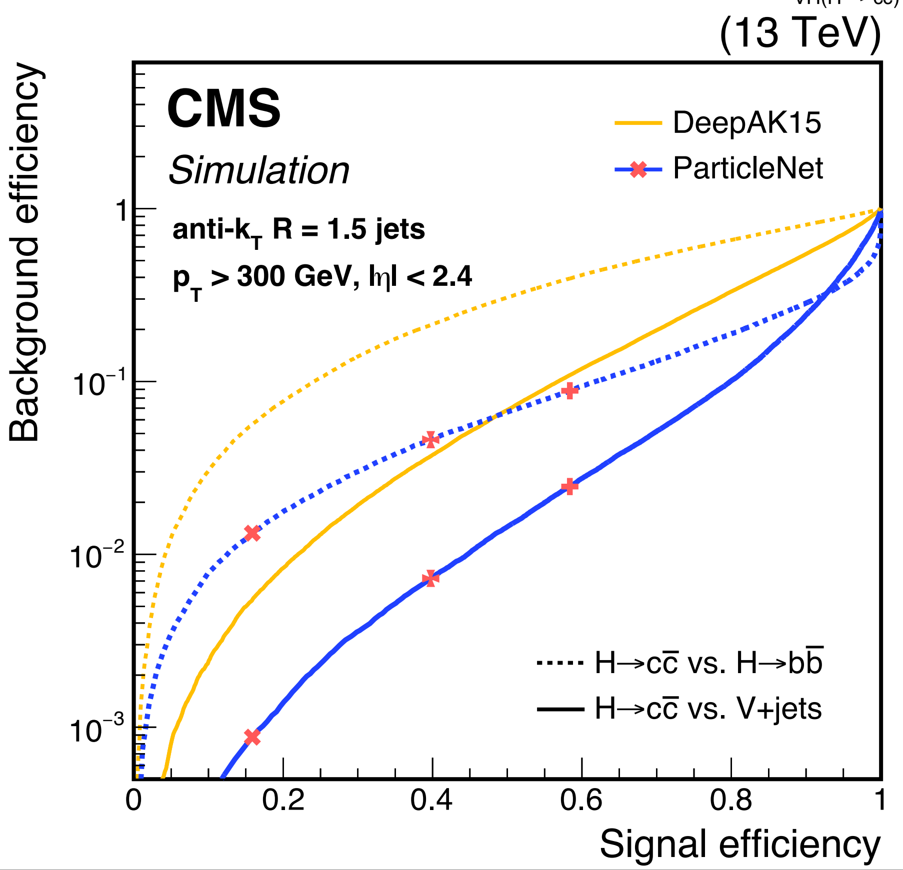
Future iteration will use GN2

Phys. Rev. Lett. 131 (2023) 061801



ATLAS bound : $|\kappa_c| < 8.5$

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?

Some other examples of physics injection

arXiv > hep-ph > arXiv:2204.12231

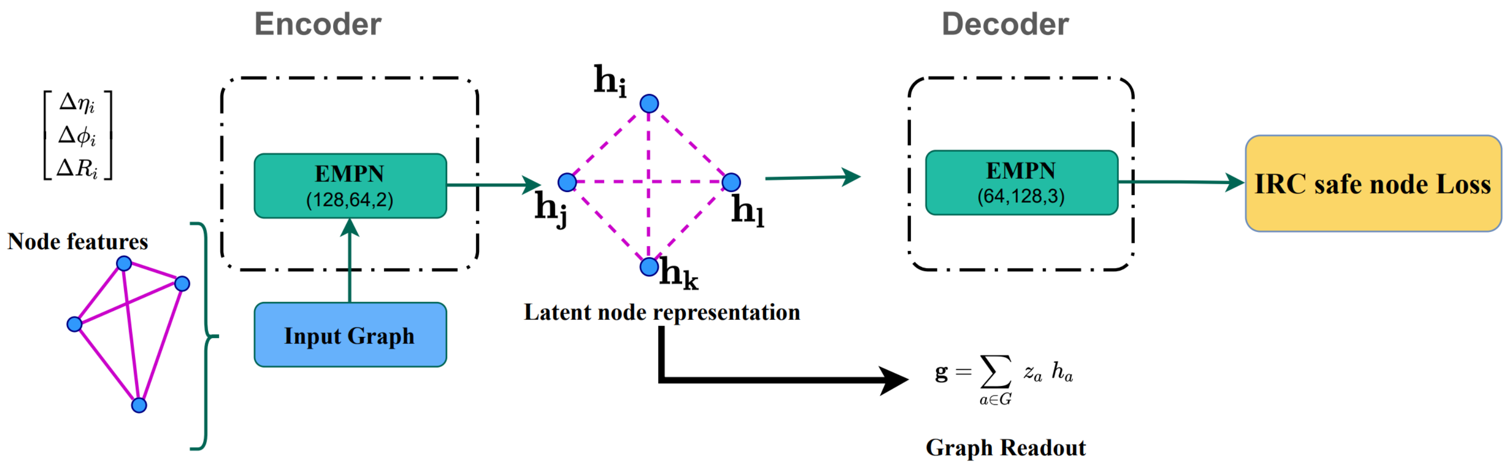
High Energy Physics – Phenomenology

[Submitted on 26 Apr 2022 (v1), last revised 31 Jul 2022 (this version, v2)]

IRC-safe Graph Autoencoder for unsupervised anomaly detection

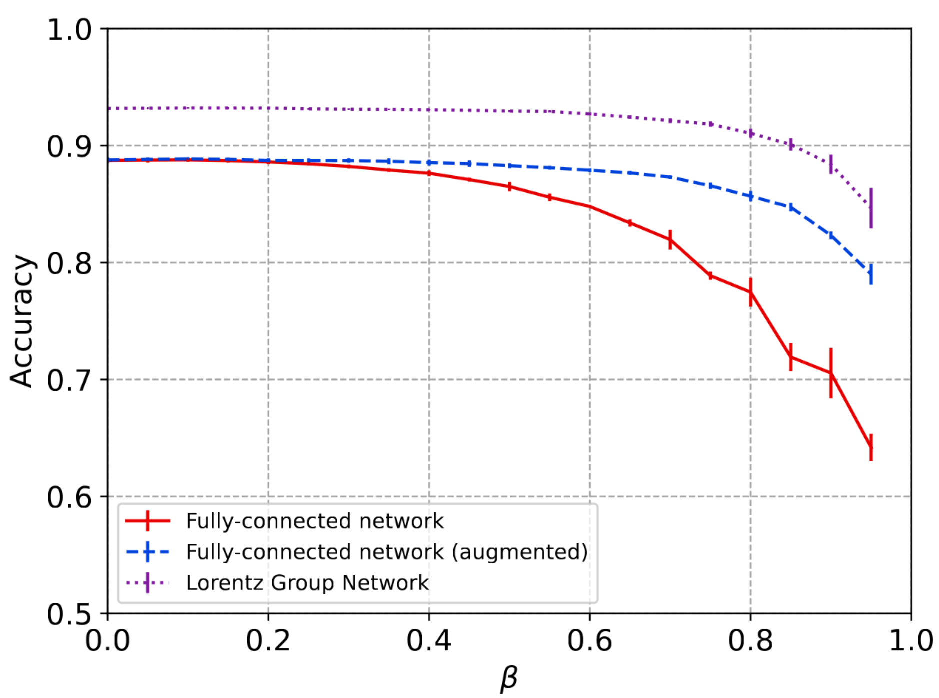
Oliver Atkinson, Akanksha Bhardwaj, Christoph Englert, Partha Konar, Vishal S. Ngairangbam, Michael Spannowsky

$$\omega_q^{(\mathcal{N}[i])}\ \hat{\Phi}^{(0)}(\hat{p}_i, \hat{p}_q) = \omega_r^{(\mathcal{N}[i])}\ \hat{\Phi}^{(0)}(\hat{p}_i, \hat{p}_r) + \omega_s^{(\mathcal{N}[i])}\ \hat{\Phi}^{(0)}(\hat{p}_i, \hat{p}_s)\,.$$



arXiv:2006.04780 : A Bogatskiy et al

arXiv:2203.06153 : SG et al



$$m_{ij}^l = \phi_e\left(h_i^l, h_j^l, \psi(\|x_i^l - x_j^l\|^2), \psi(\langle x_i^l, x_j^l\rangle)\right)$$

$$h_i^{l+1} = h_i^l + \phi_h(h_i^l, \sum_{j\in[N]} w_{ij} m_{ij}^l),$$

The path we will take

Basic Motivation

- Why are we doing ML?
- The basic LHC operations with ML.
- Synergies through eye of ML.

The supervised ML

- Jet tagging : the most successful venture.
- Tracking & Calorimetery
- The Particle-Flow algorithm

Generative models

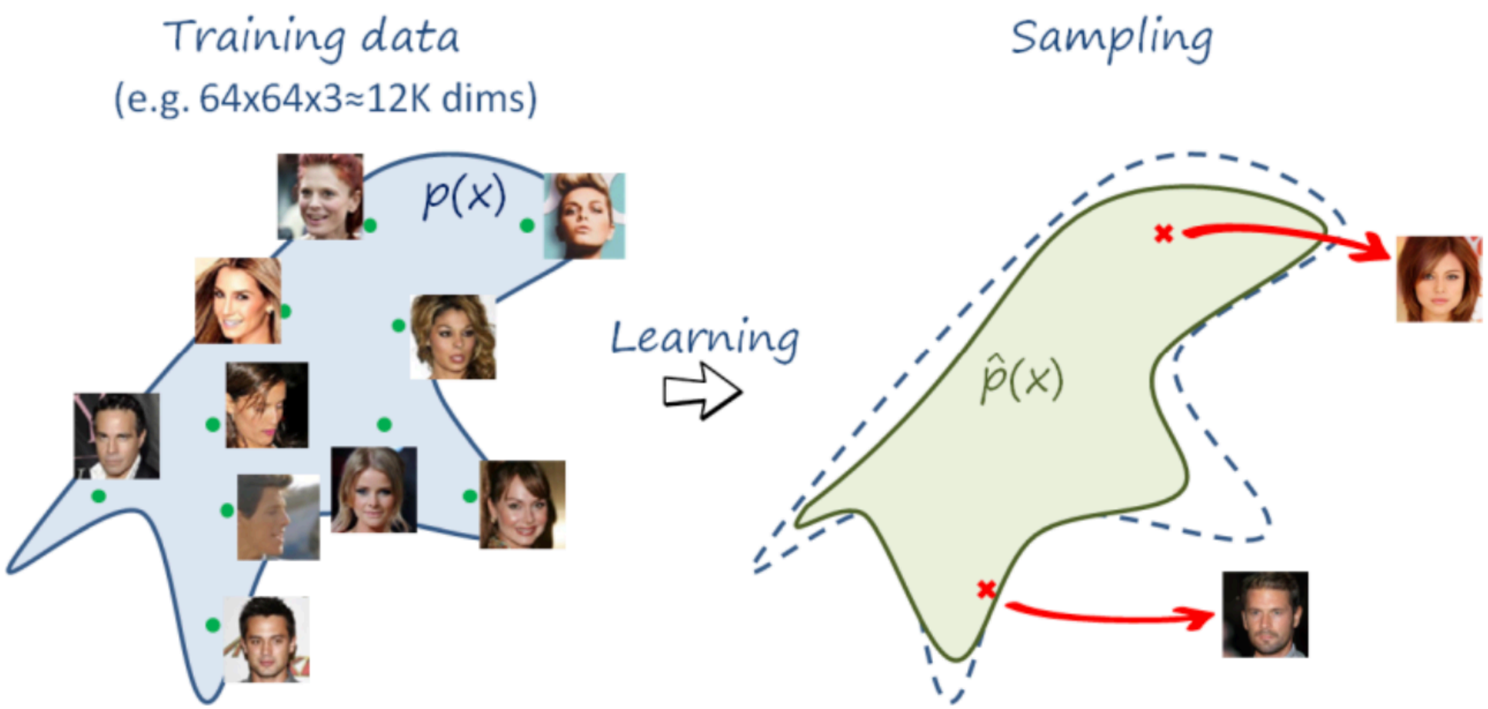
- Different paradigms
- Few physics cases.
- Interpretable ML & uncertainty estimation.

The anomaly detection saga

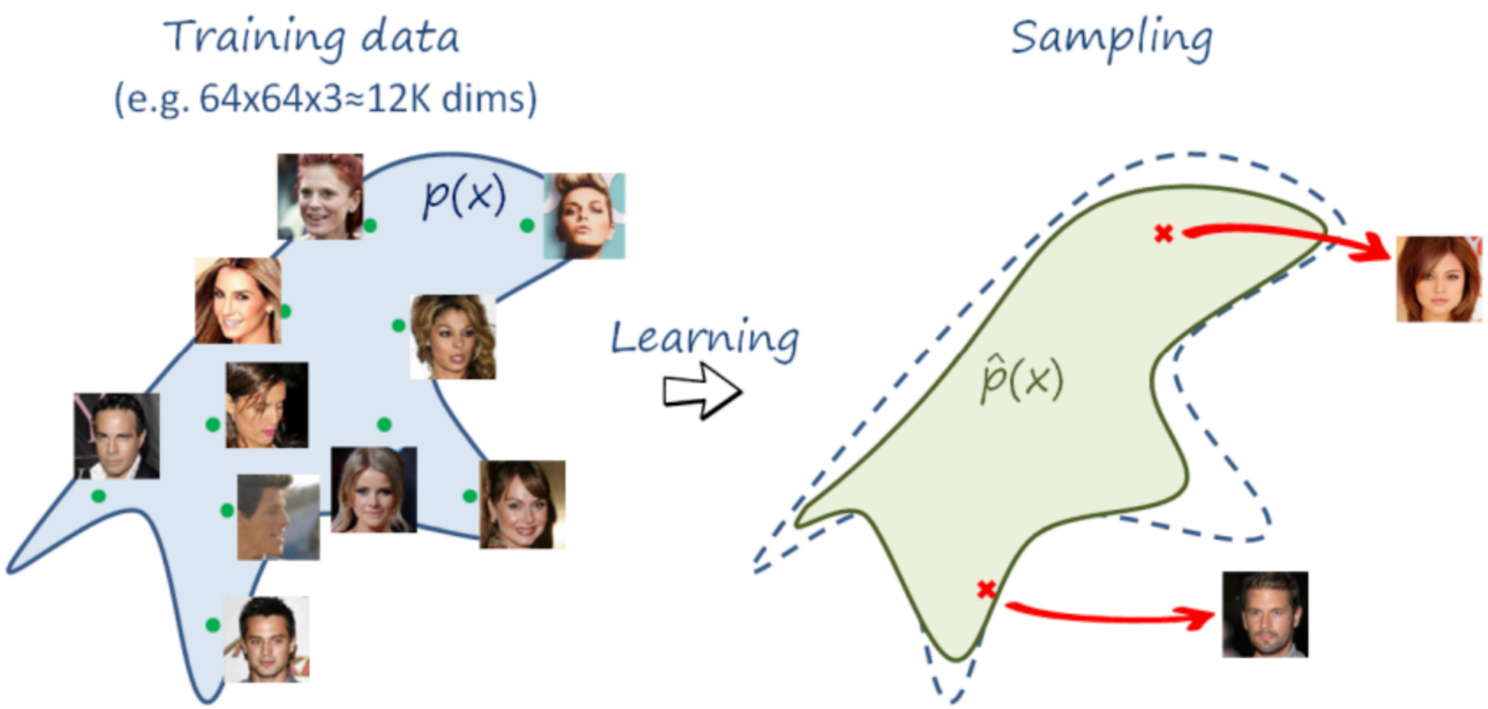
- The outliers & density estimate
- How to translate this in new signals?

THE ROAD MAP

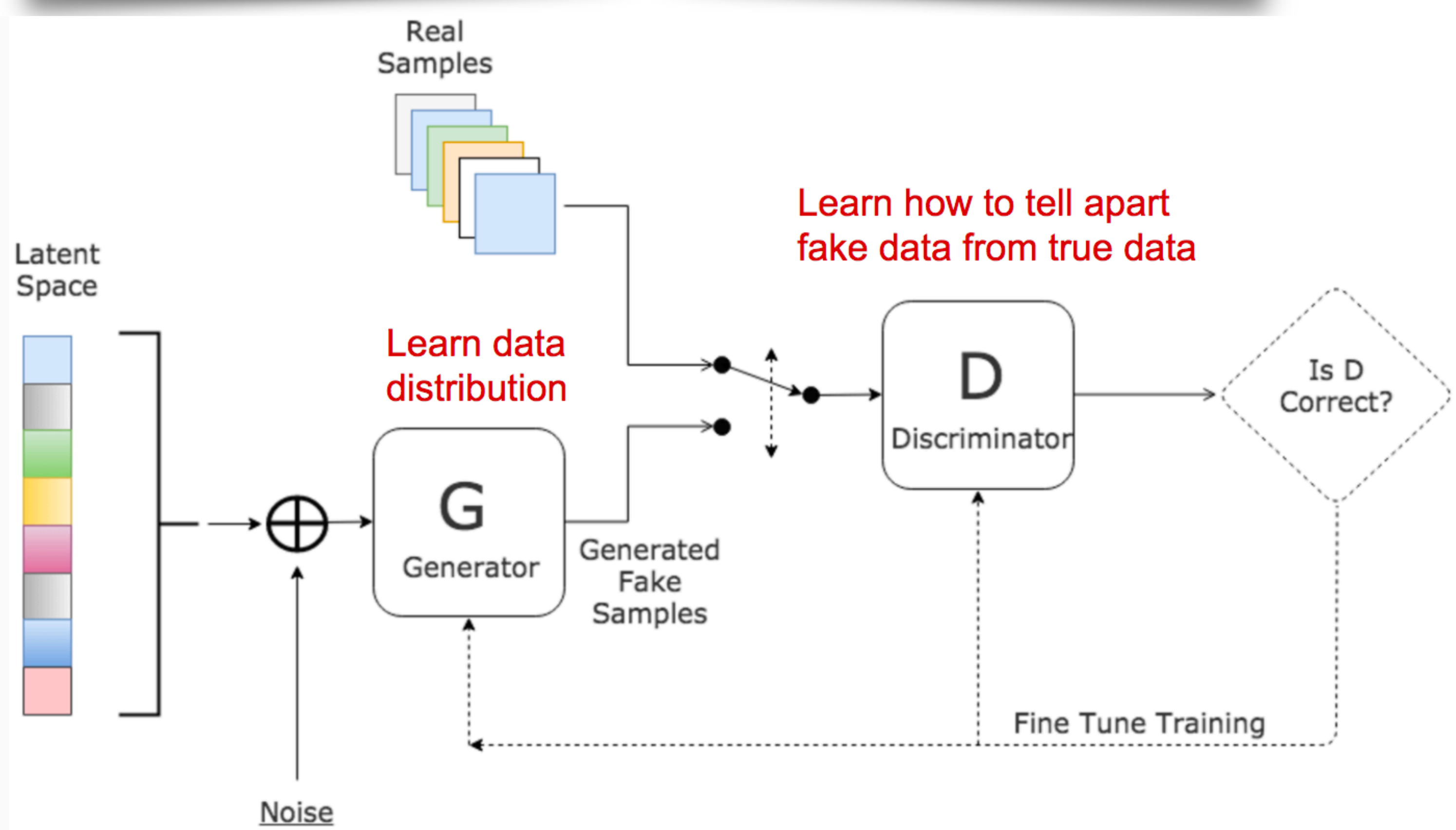
Generative models : what are they?



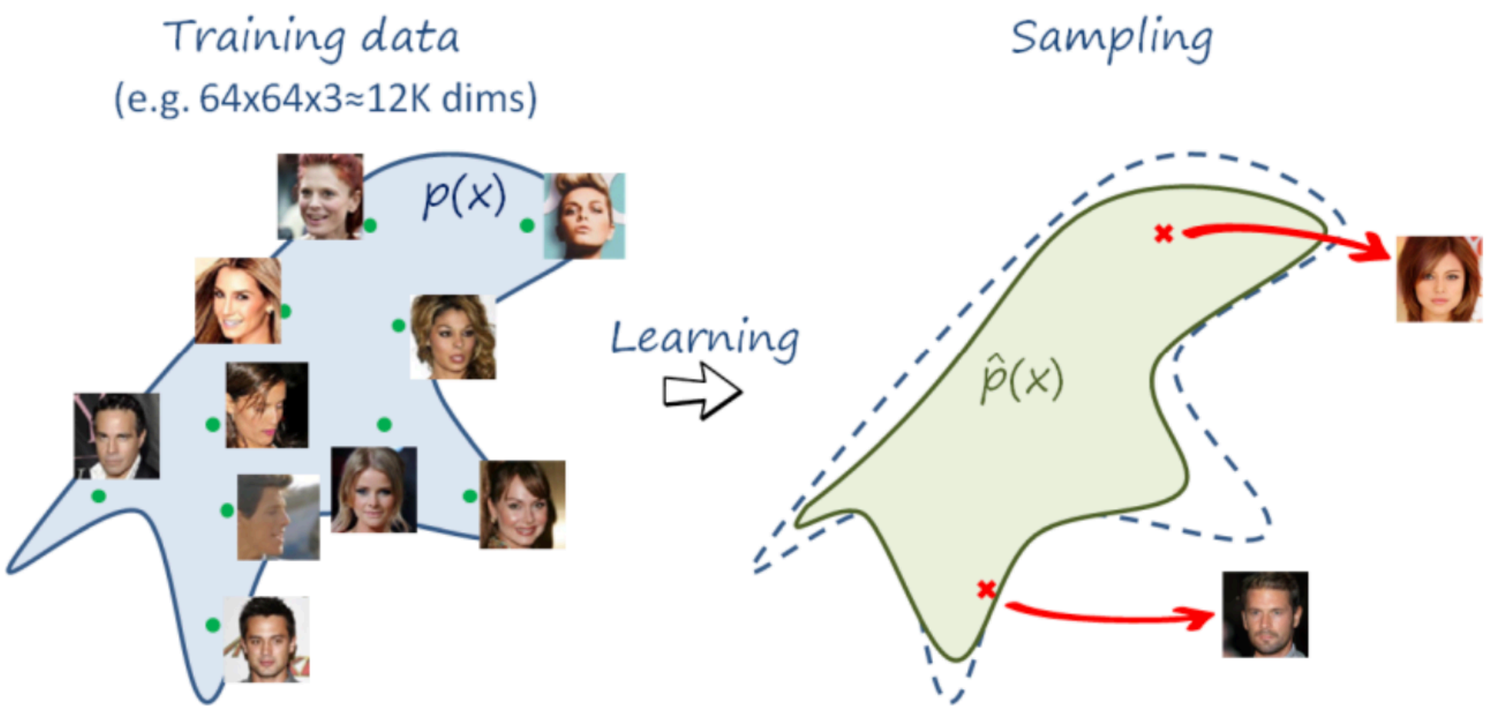
Generative models : what are they?

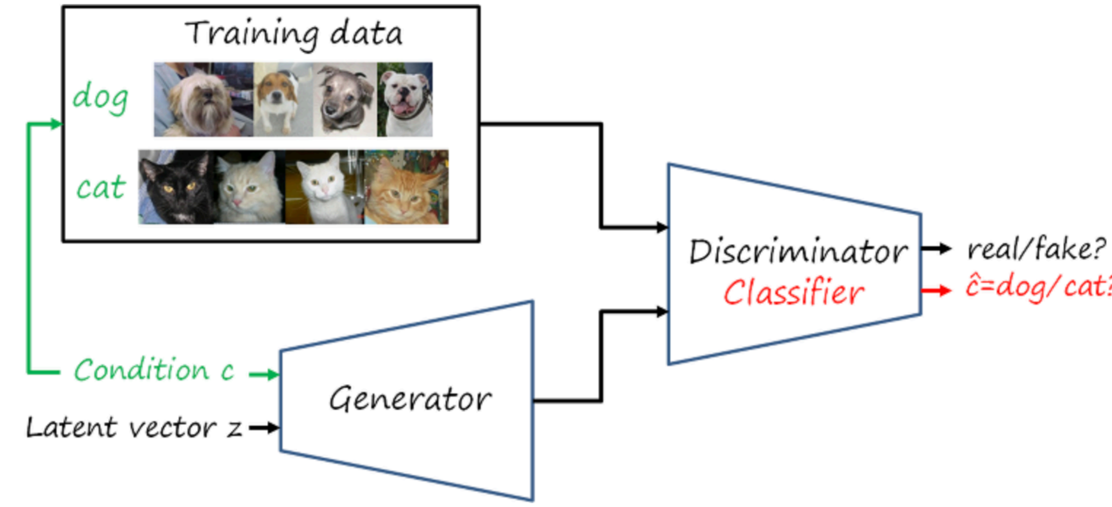


$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x\sim p_{data}(x)} \log D_{\theta_d}(x) + \mathbb{E}_{z\sim p(z)} \log\left(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)\right]$$

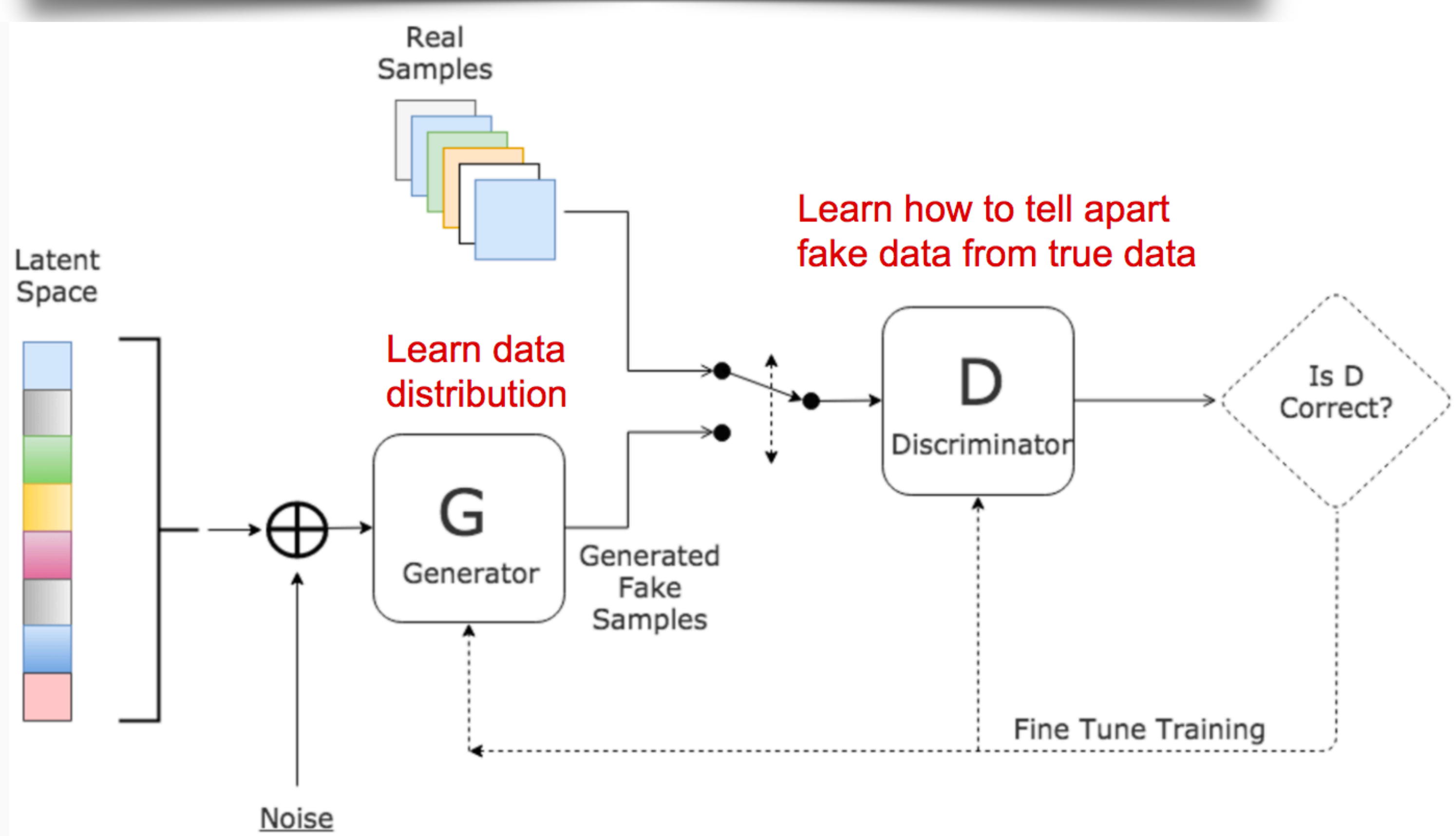


Generative models : what are they?

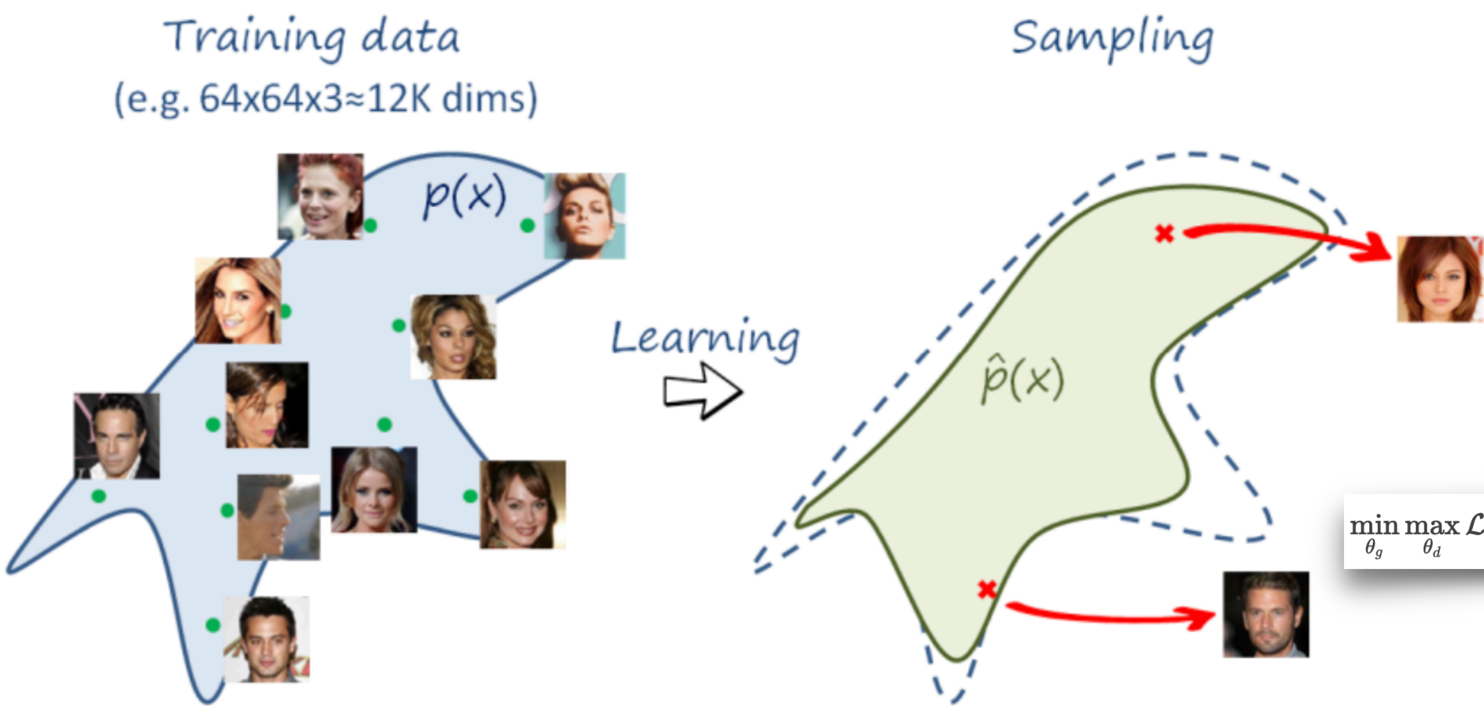


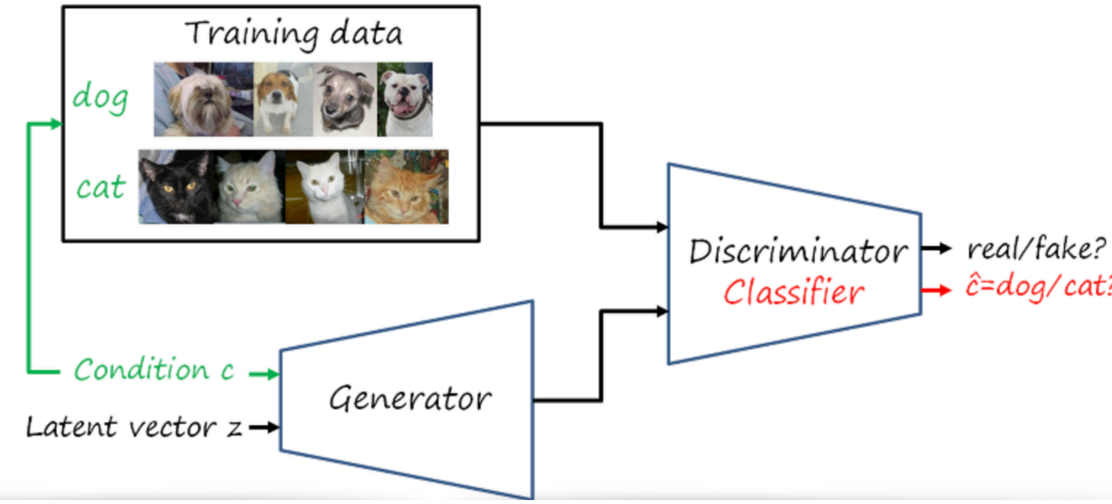


$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}(x)} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log \left(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right)\right) \right]$$



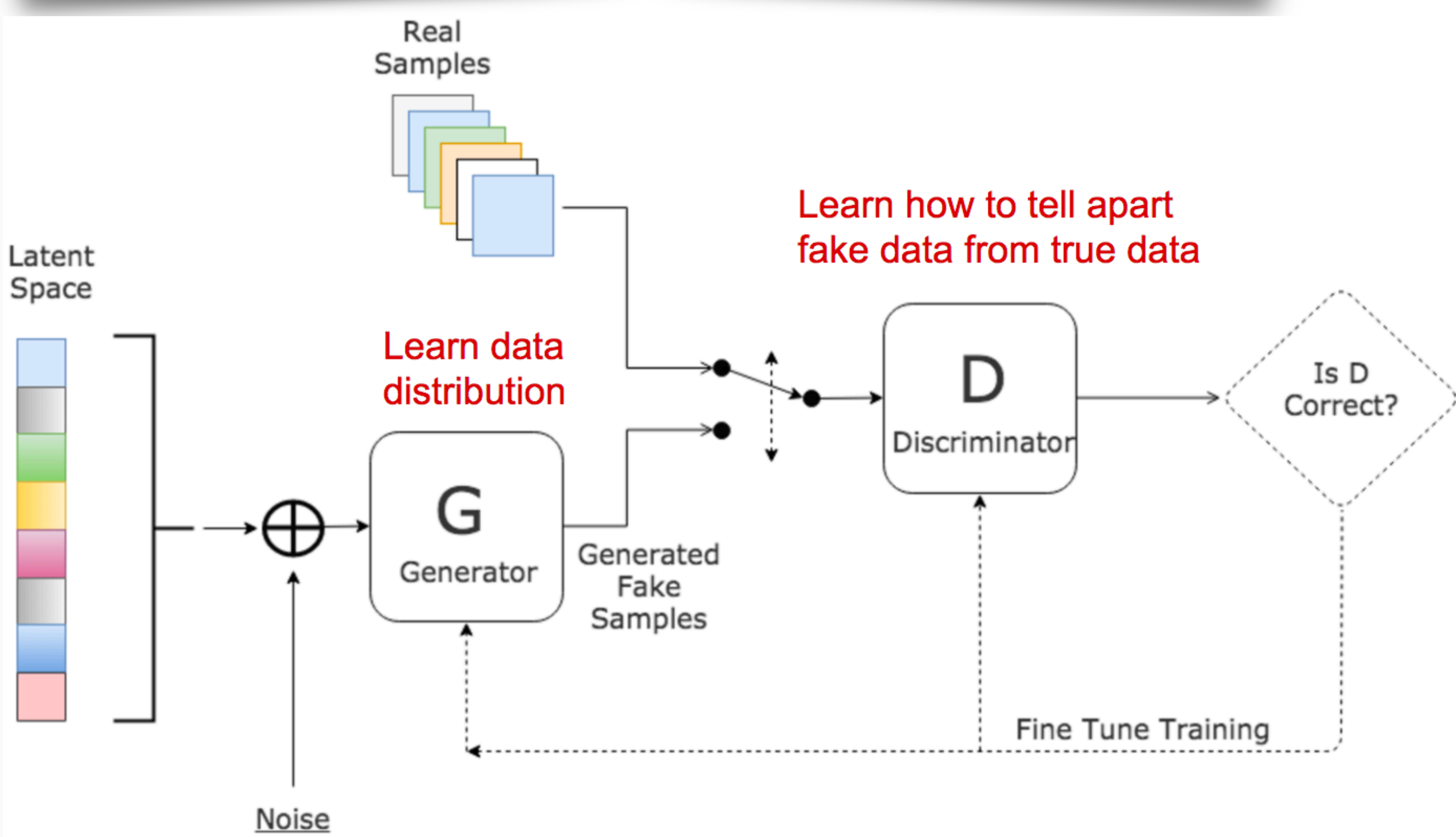
Generative models : what are they?



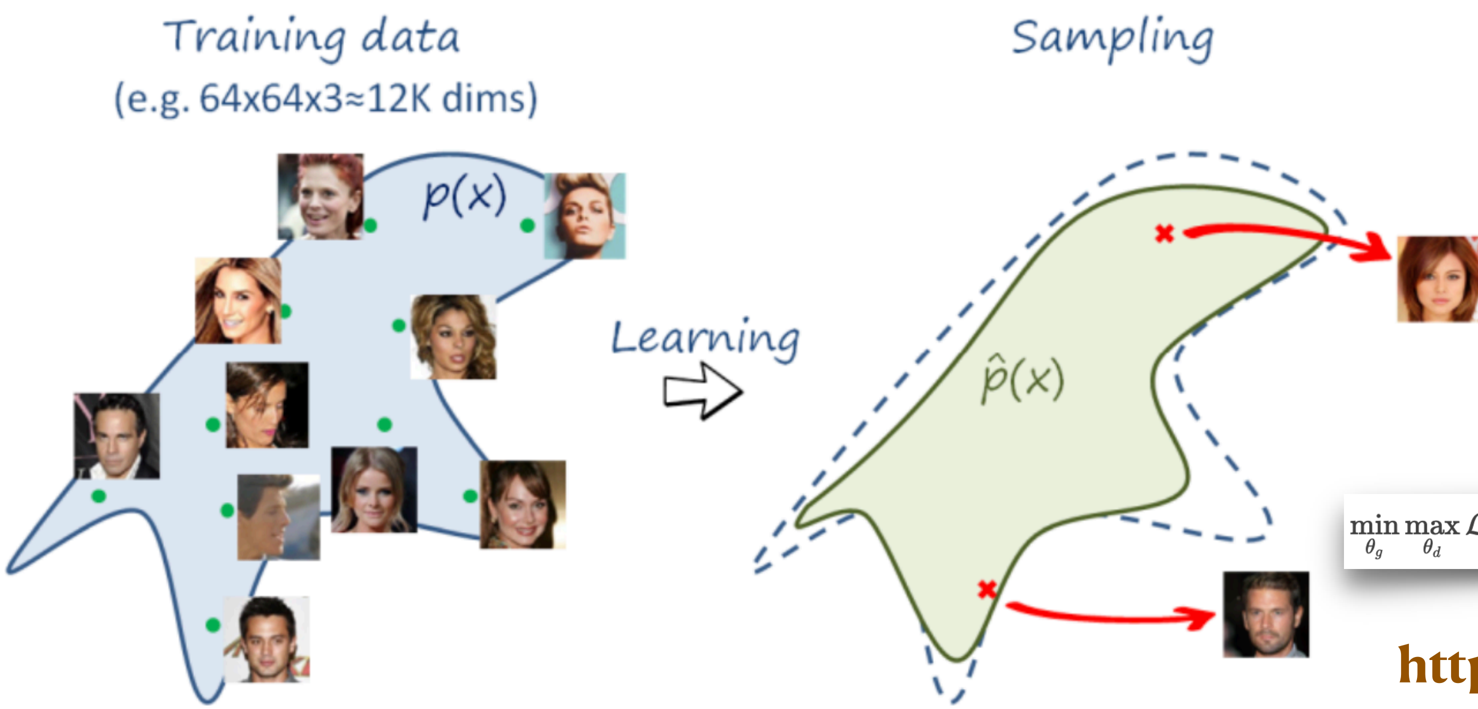


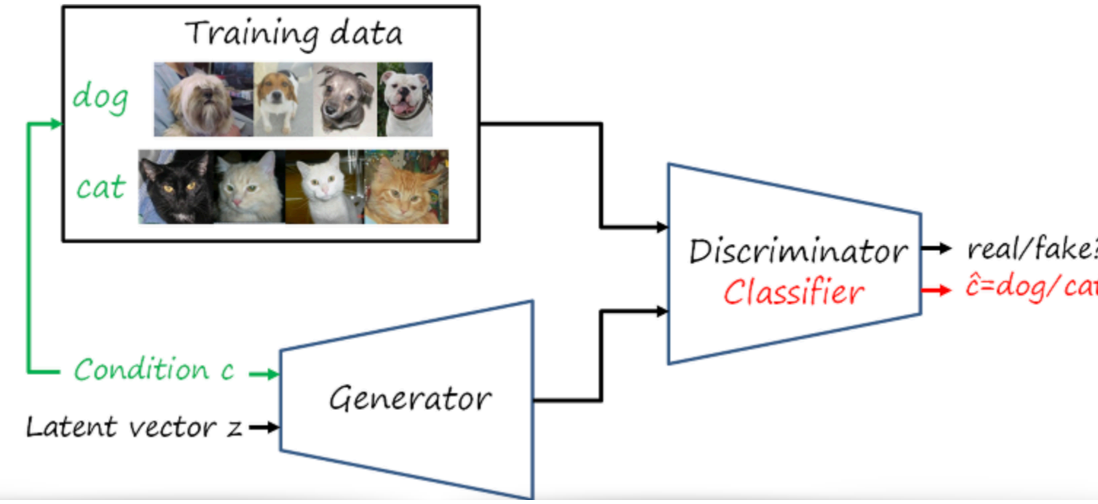
$$\min_{\theta_g}\max_{\theta_d} \mathcal{L}_{\text{cGAN}}(\theta_g,\theta_d) = \min_{\theta_g}\max_{\theta_d} \left[\mathbb{E}_{c,x\sim p_{data}(c,x)} \log D_{\theta_d}(c,x) + \mathbb{E}_{c\sim p_{data}(c),z\sim p_z(z)} \log\left(1 - D_{\theta_d}\left(c, G_{\theta_g}(c,z)\right)\right)\right]$$

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



Generative models : what are they?

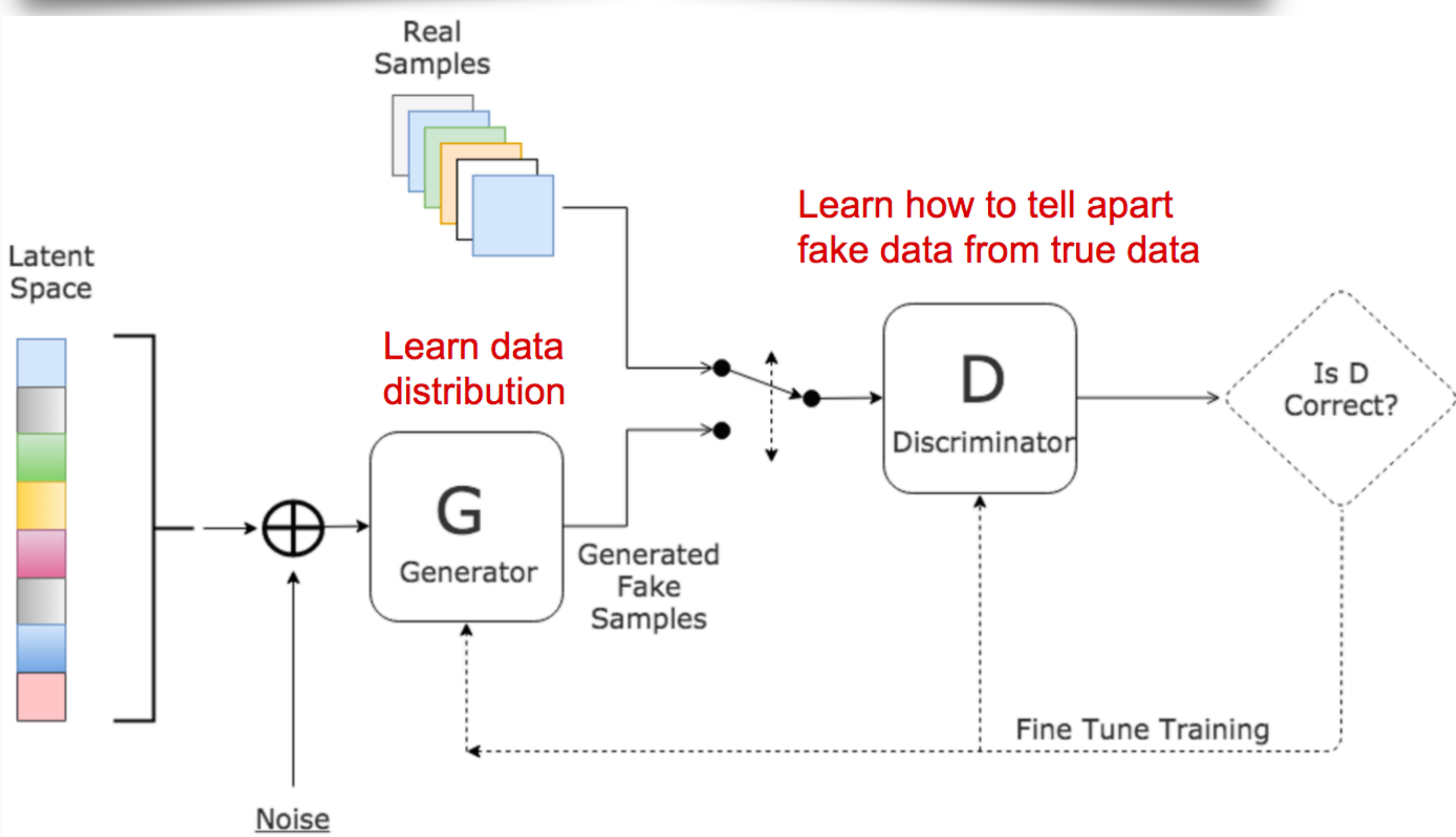




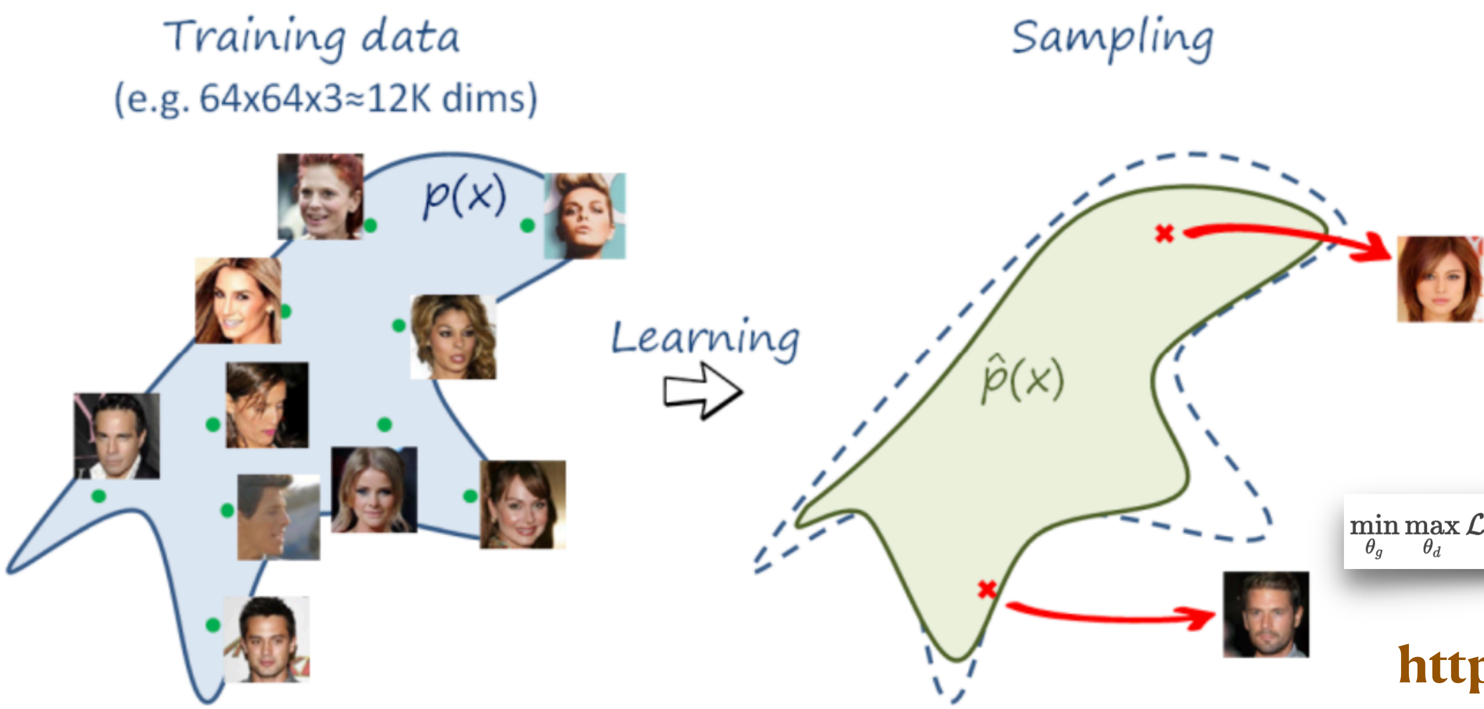
$$\min_{\theta_g}\max_{\theta_d} \mathcal{L}_{\mathrm{cGAN}}(\theta_g,\theta_d) = \min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{c,x\sim p_{data}(c,x)} \log D_{\theta_d}(c,x) + \mathbb{E}_{c\sim p_{data}(c),z\sim p_z(z)} \log\left(1-D_{\theta_d}\left(c,G_{\theta_g}(c,z)\right)\right)\right]$$

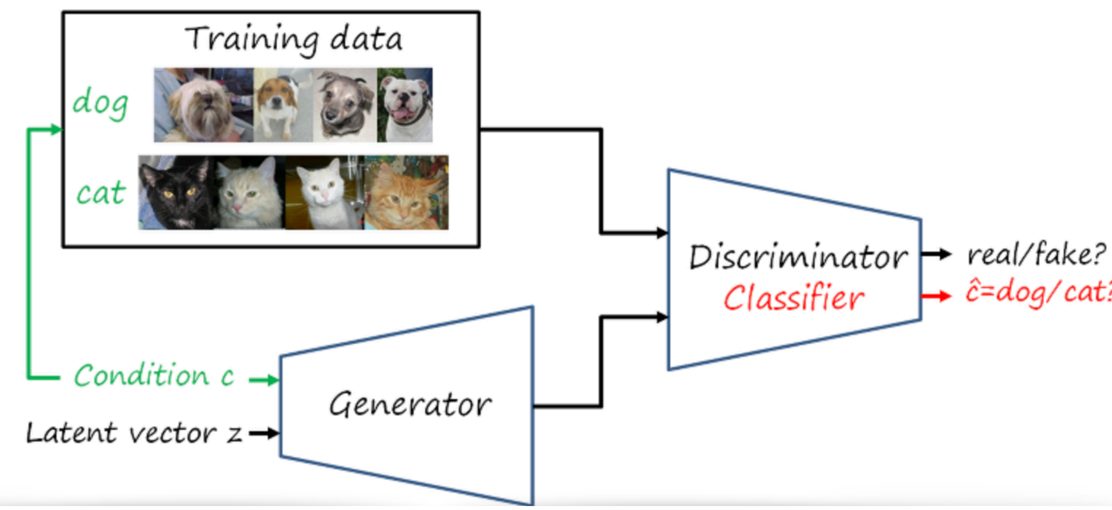
http://www.lherranz.org/2018/08/07/imagetranslation/

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



Generative models : what are they?

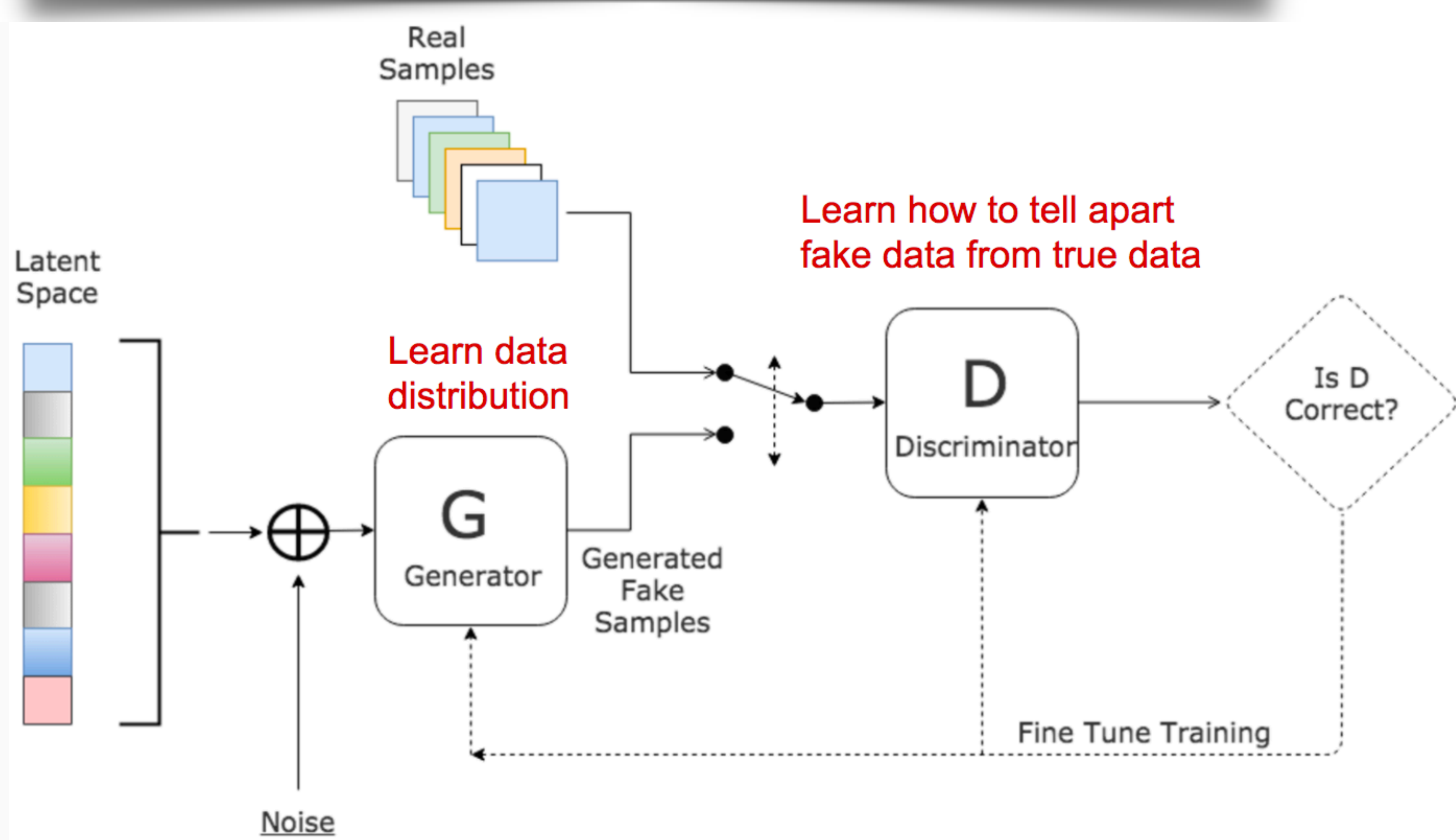


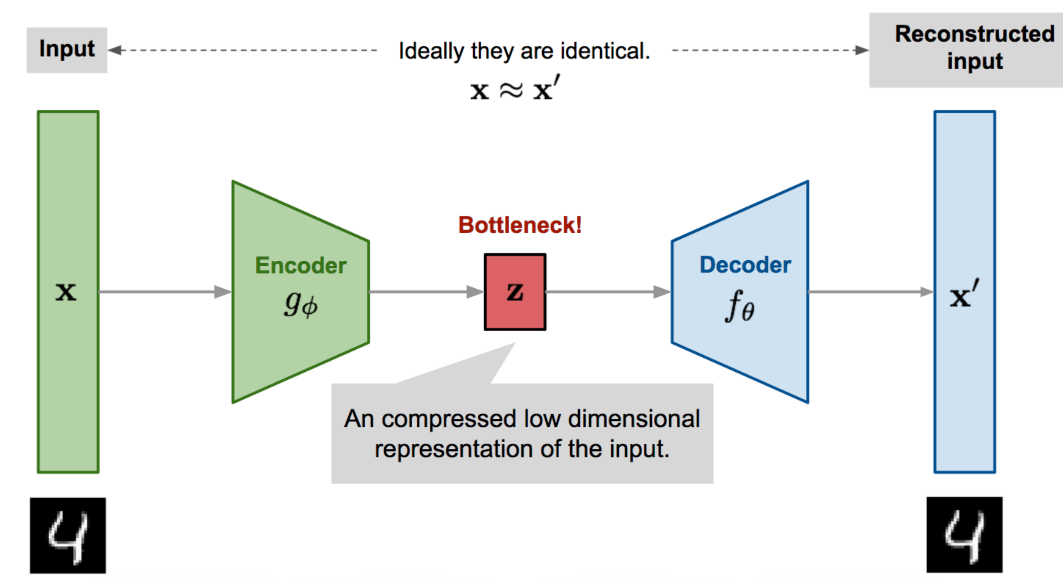


$$\min_{\theta_g}\max_{\theta_d} \mathcal{L}_{\mathrm{cGAN}}(\theta_g,\theta_d) = \min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{c,x\sim p_{\mathrm{data}}(c,x)}\log D_{\theta_d}(c,x) + \mathbb{E}_{c\sim p_{\mathrm{data}}(c),z\sim p_z(z)}\log\left(1-D_{\theta_d}\left(c,G_{\theta_g}(c,z)\right)\right)\right]$$

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

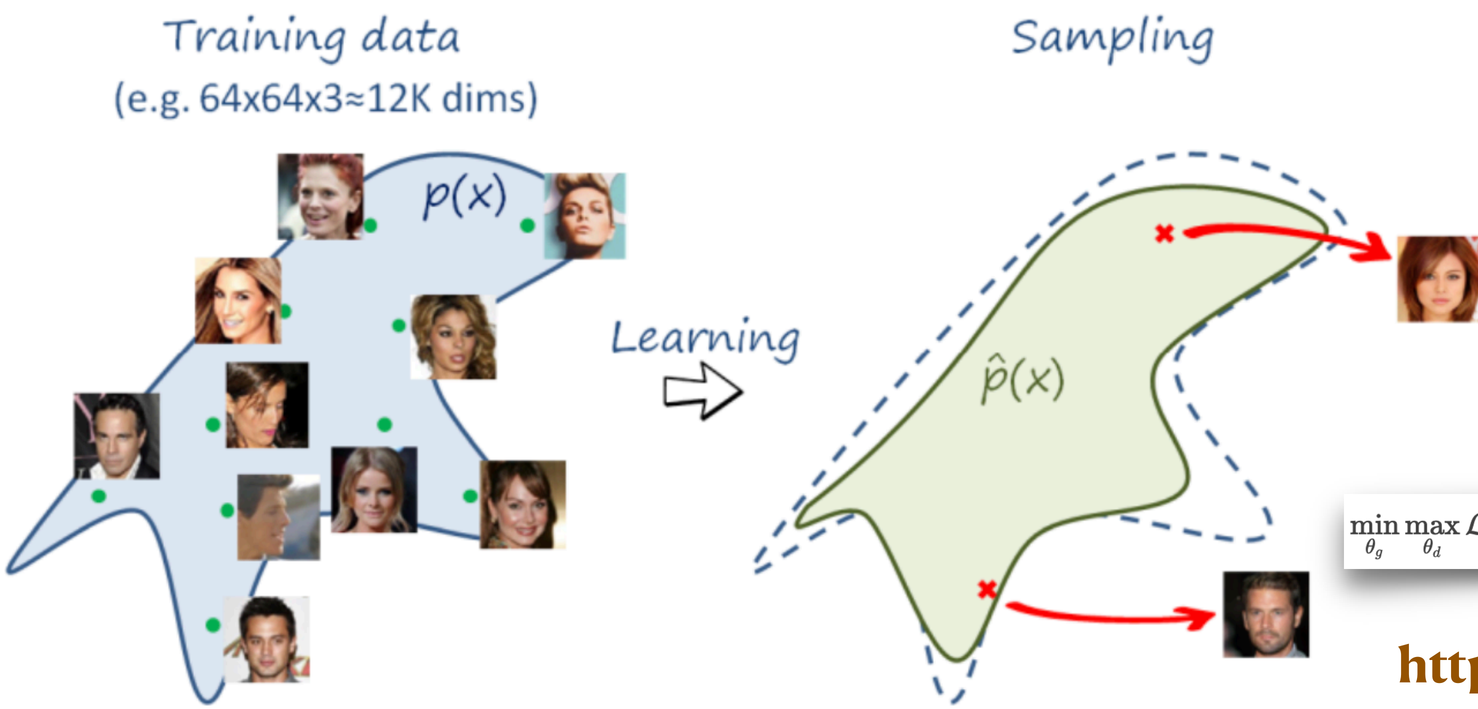
$$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p_{data}(x)}\log D_{\theta_d}(x) + \mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)\right]$$

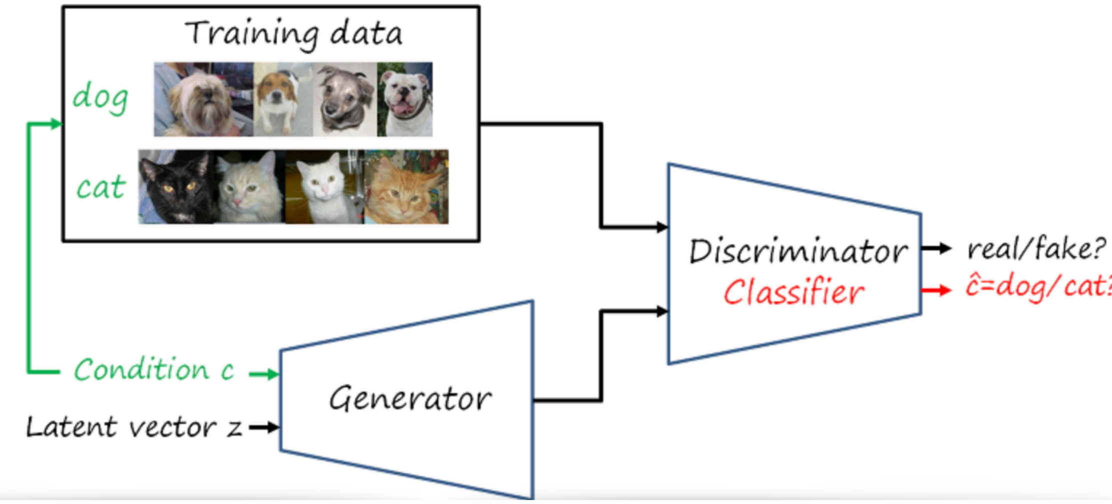




$$L_{\mathrm{AE}}(\theta,\phi) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}^{(i)} - f_\theta(g_\phi(\mathbf{x}^{(i)})))^2$$

Generative models : what are they?

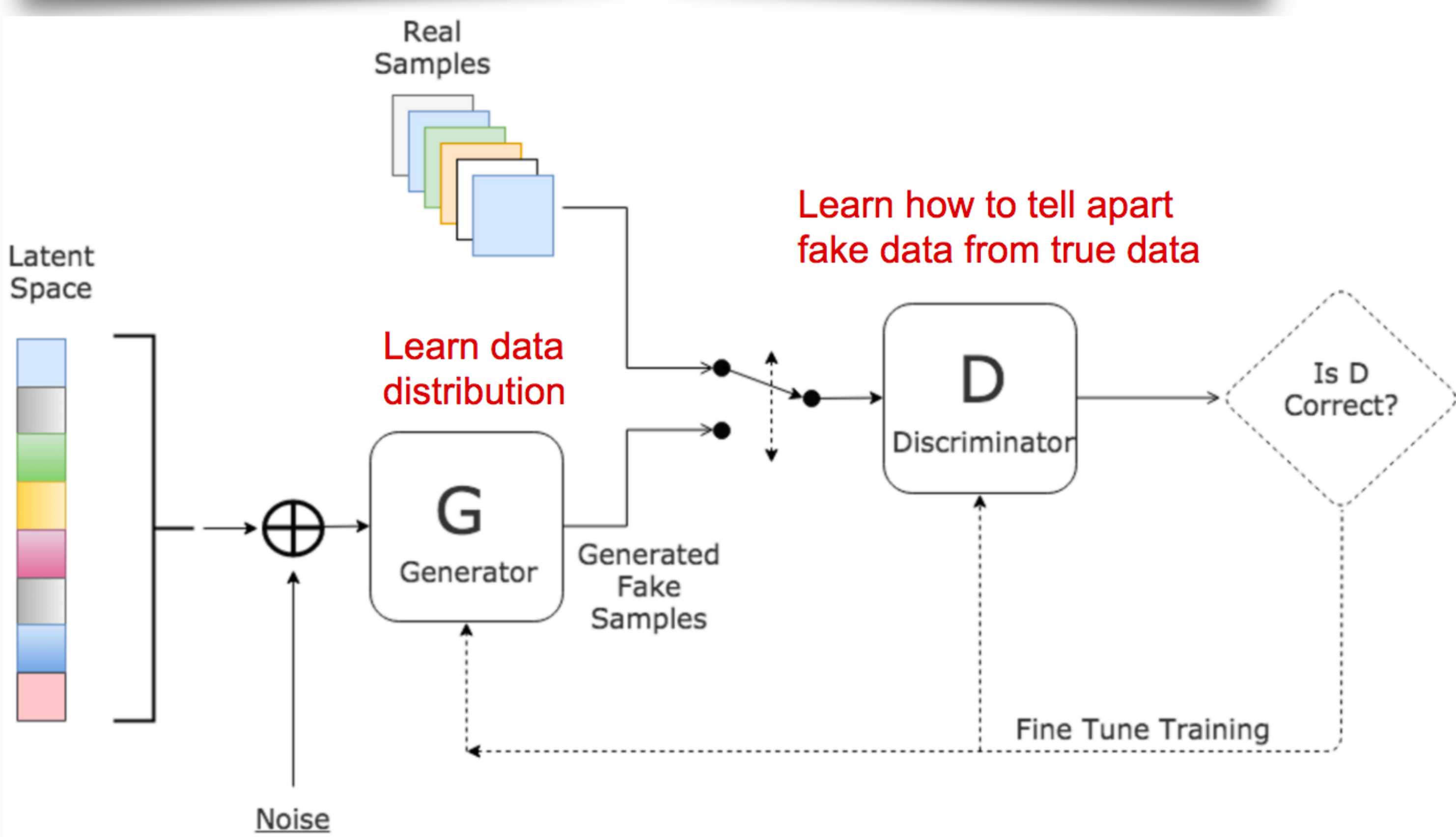


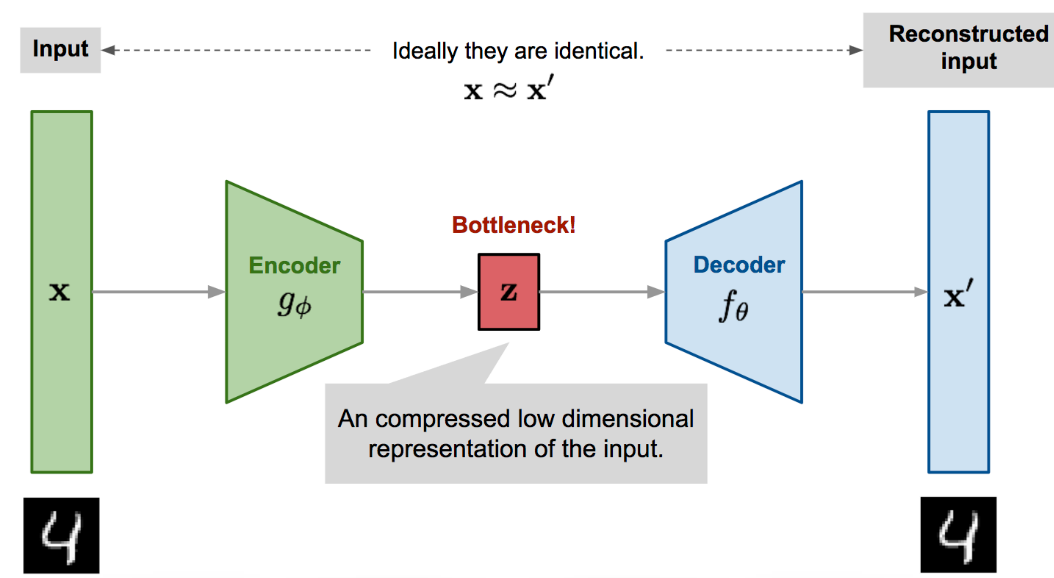


$$\min_{\theta_g}\max_{\theta_d} \mathcal{L}_{\mathrm{cGAN}}(\theta_g,\theta_d) = \min_{\theta_g}\max_{\theta_d} \left[\mathbb{E}_{c,x\sim p_{\mathrm{data}}(c,x)} \log D_{\theta_d}(c,x) + \mathbb{E}_{c\sim p_{\mathrm{data}}(c),z\sim p_z(z)} \log\left(1-D_{\theta_d}\left(c, G_{\theta_g}(c,z)\right)\right)\right]$$

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

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

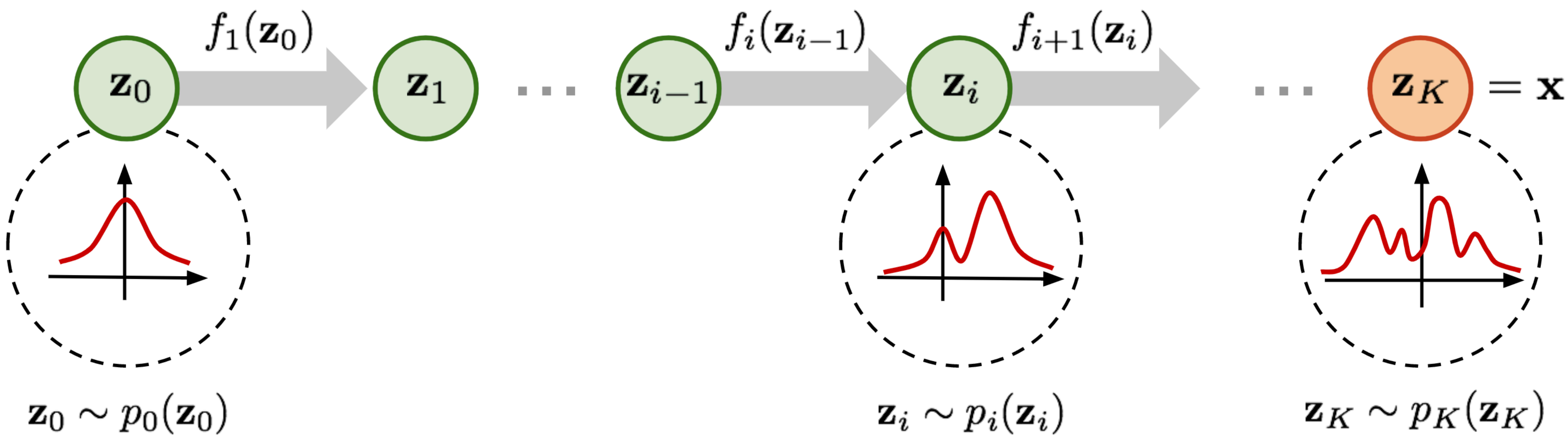




$$L_{\mathrm{AE}}(\theta,\phi) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}^{(i)} - f_\theta(g_\phi(\mathbf{x}^{(i)})))^2$$

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

Learn the PDF through bijections

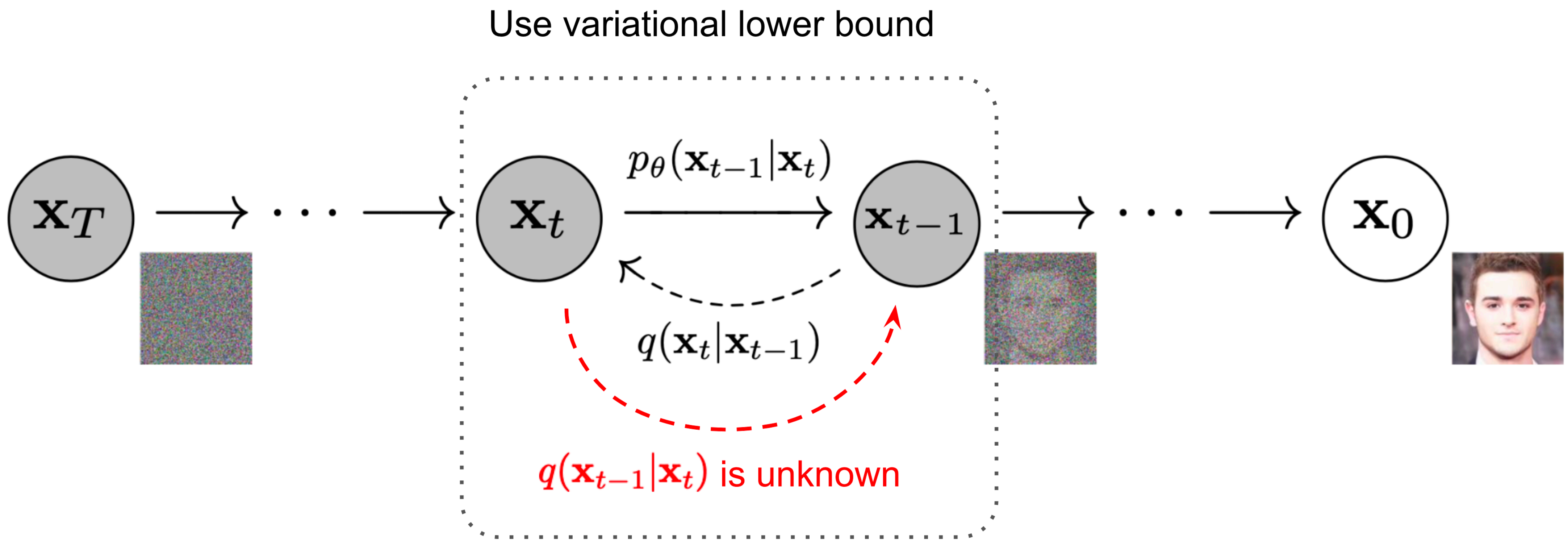


$$\mathbf{z} \sim \pi(\mathbf{z}), \mathbf{x} = f(\mathbf{z}), \mathbf{z} = f^{-1}(\mathbf{x})$$
$$p(\mathbf{x}) = \pi(\mathbf{z}) \left|\det \frac{d\mathbf{z}}{d\mathbf{x}}\right| = \pi(f^{-1}(\mathbf{x})) \left|\det \frac{df^{-1}}{d\mathbf{x}}\right|$$

$$\begin{aligned}
p_i(\mathbf{z}_i) &= p_{i-1}(f_i^{-1}(\mathbf{z}_i)) \left|\det \frac{df_i^{-1}}{d\mathbf{z}_i}\right| \\
&= p_{i-1}(\mathbf{z}_{i-1}) \left|\det \left(\frac{df_i}{d\mathbf{z}_{i-1}}\right)^{-1}\right| \\
&= p_{i-1}(\mathbf{z}_{i-1}) \left|\det \frac{df_i}{d\mathbf{z}_{i-1}}\right|^{-1} \\
\log p_i(\mathbf{z}_i) &= \log p_{i-1}(\mathbf{z}_{i-1}) - \log \left|\det \frac{df_i}{d\mathbf{z}_{i-1}}\right|
\end{aligned}$$

$$\mathbf{x} = \mathbf{z}_K = f_K \circ f_{K-1} \circ \cdots \circ f_1(\mathbf{z}_0)$$
$$\begin{aligned}
\log p(\mathbf{x}) = \log \pi_K(\mathbf{z}_K) &= \log \pi_{K-1}(\mathbf{z}_{K-1}) - \log \left|\det \frac{df_K}{d\mathbf{z}_{K-1}}\right| \\
&= \log \pi_{K-2}(\mathbf{z}_{K-2}) - \log \left|\det \frac{df_{K-1}}{d\mathbf{z}_{K-2}}\right| - \log \left|\det \frac{df_K}{d\mathbf{z}_{K-1}}\right| \\
&= \ldots \\
&= \log \pi_0(\mathbf{z}_0) - \sum_{i=1}^{K} \log \left|\det \frac{df_i}{d\mathbf{z}_{i-1}}\right|
\end{aligned}$$

How about create some noise & do it?



Forward diffusion

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \quad q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

Reverse diffusion

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \quad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Generative models : the popular species

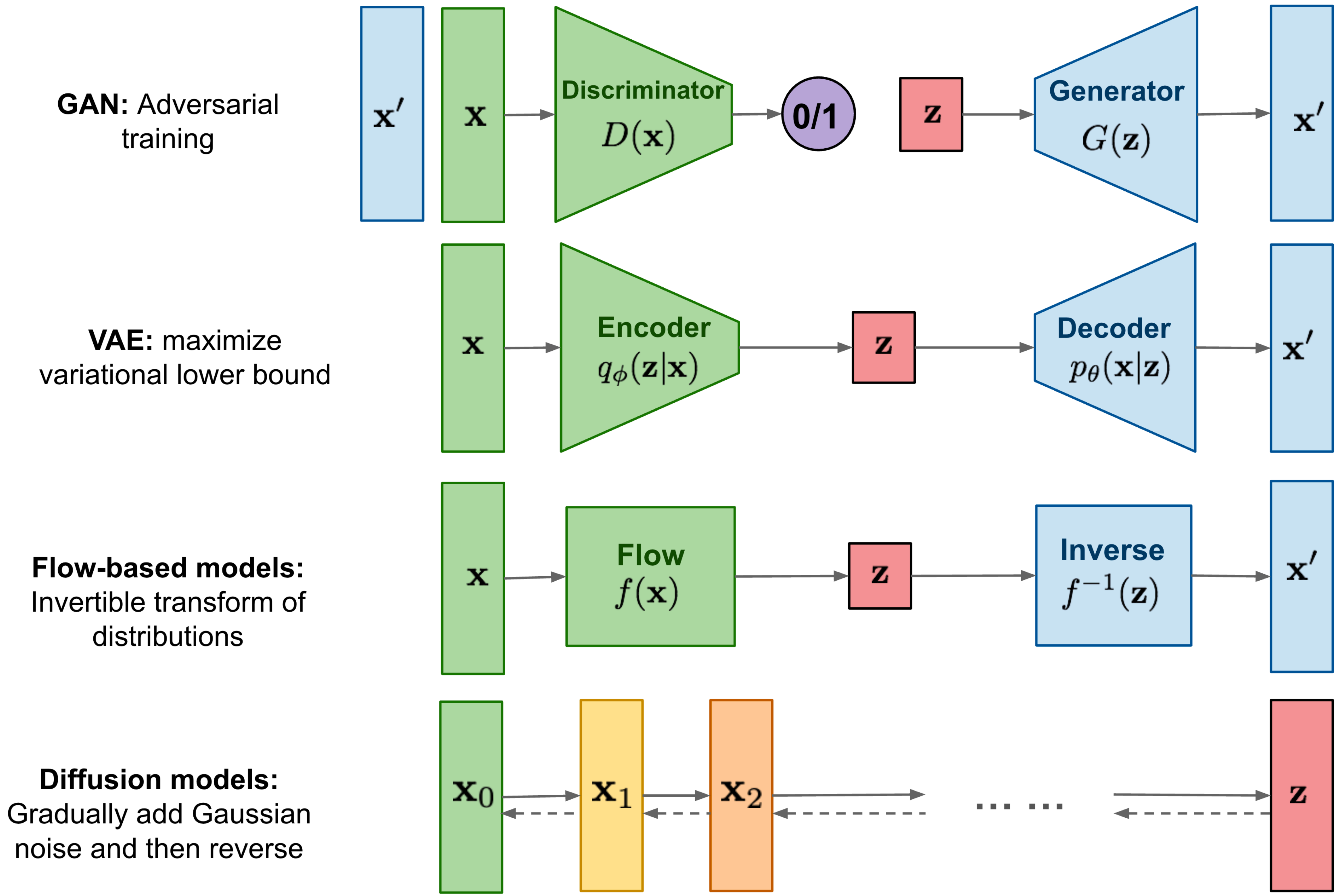
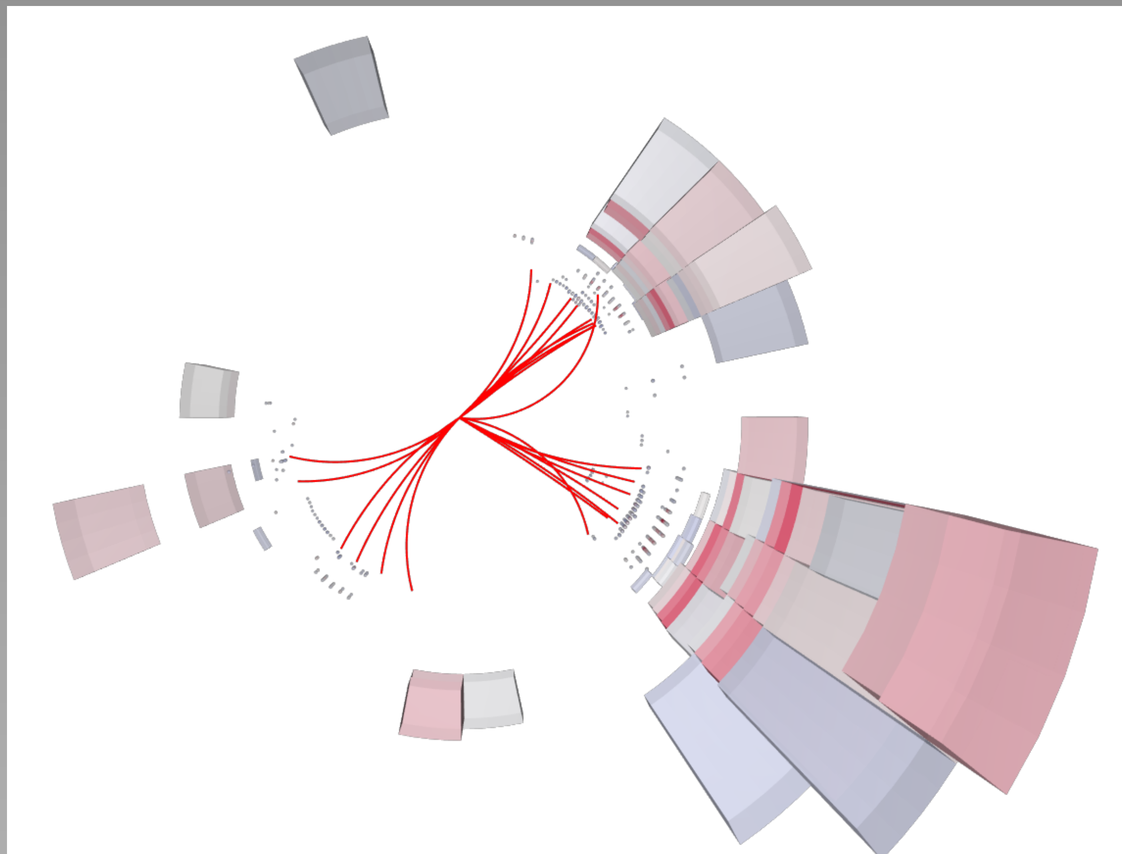
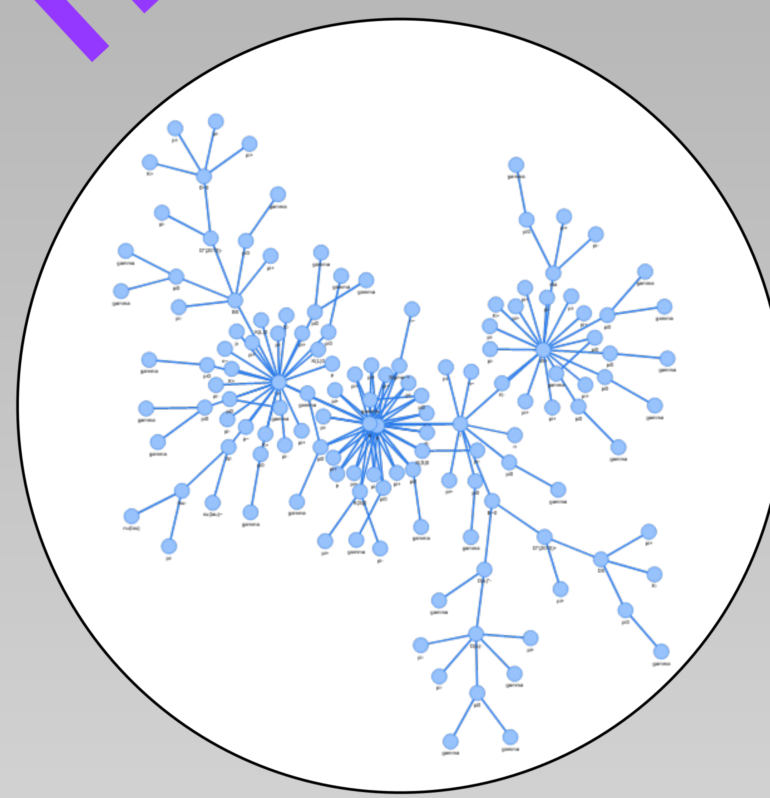


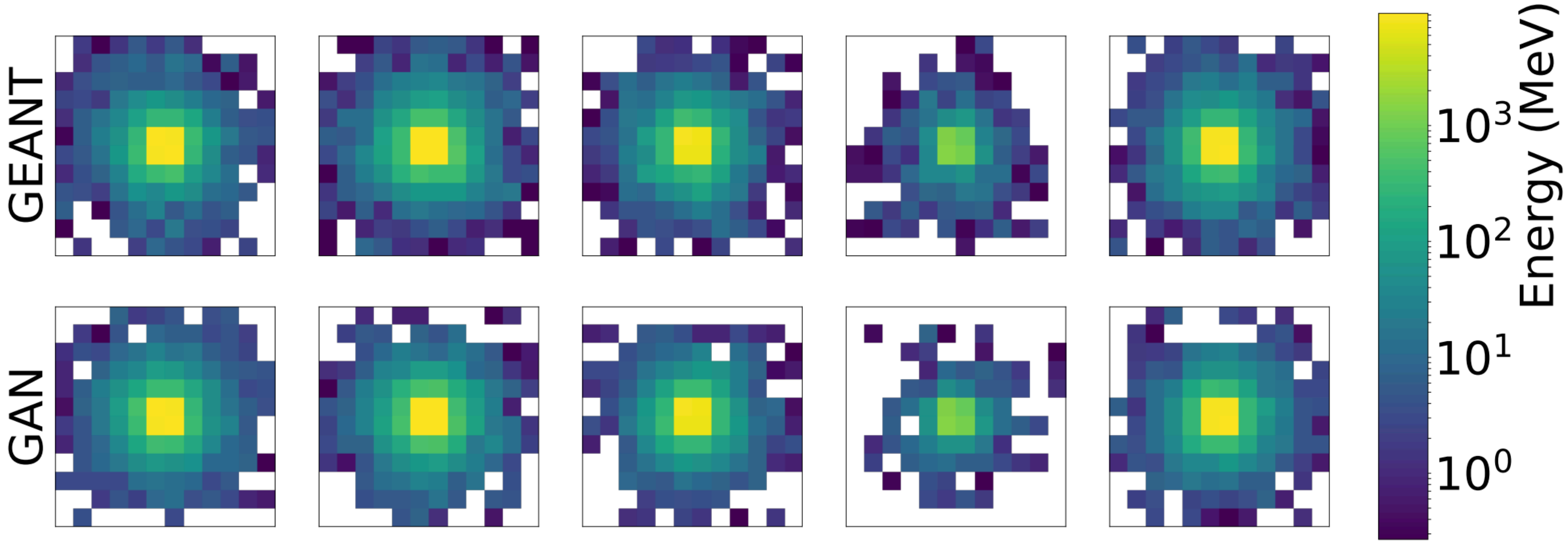
Fig from : https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

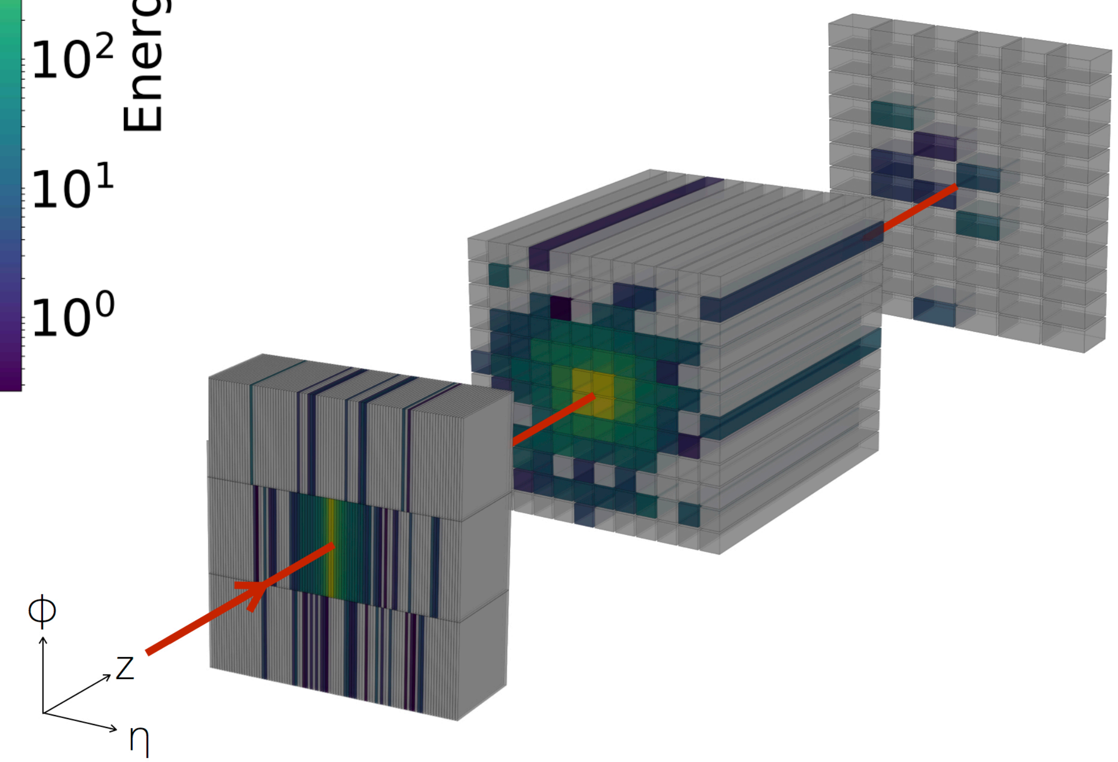
Generative models for calorimeter simulations



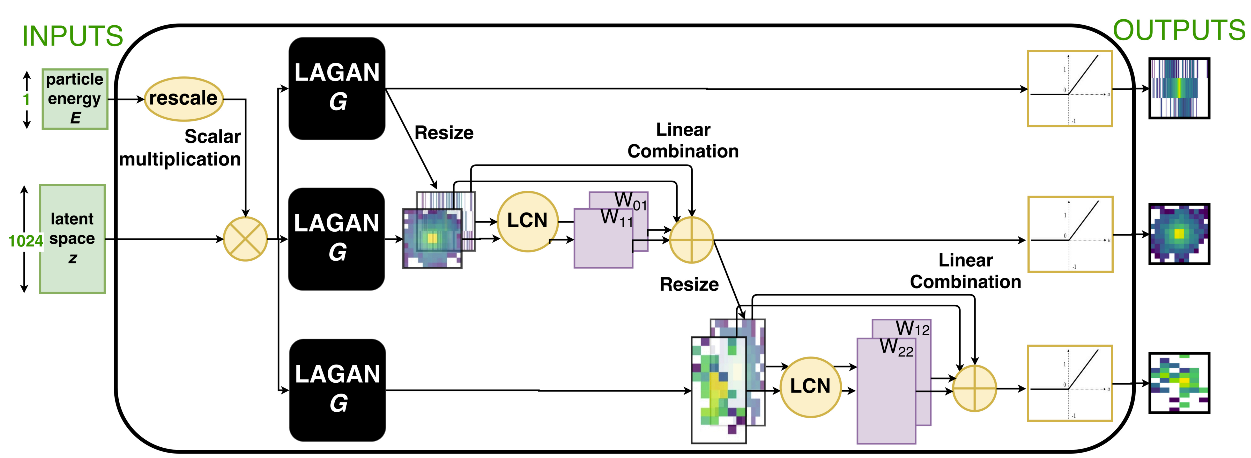
$f_{\{\theta\}}\big($  $, z \big)$

Detector simulation using ML

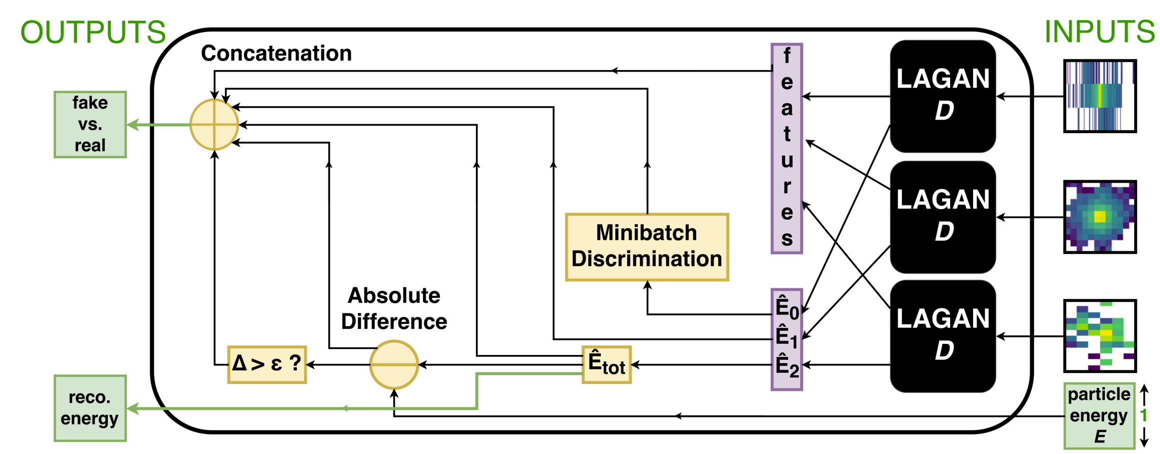




CaloGAN 1705.02355
[Michela Paganini], [Luke de Oliveira], [Benjamin Nachman]

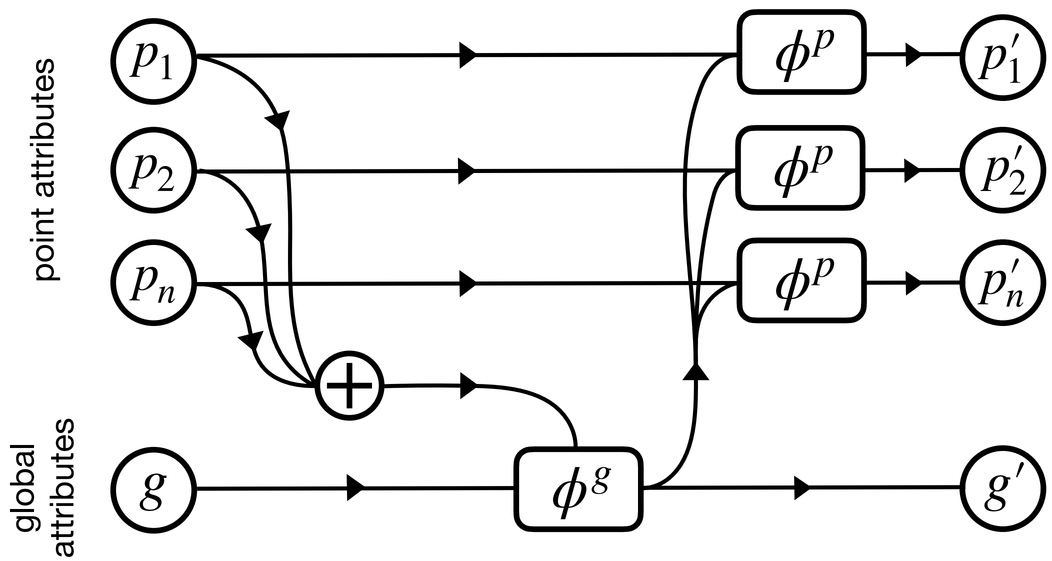


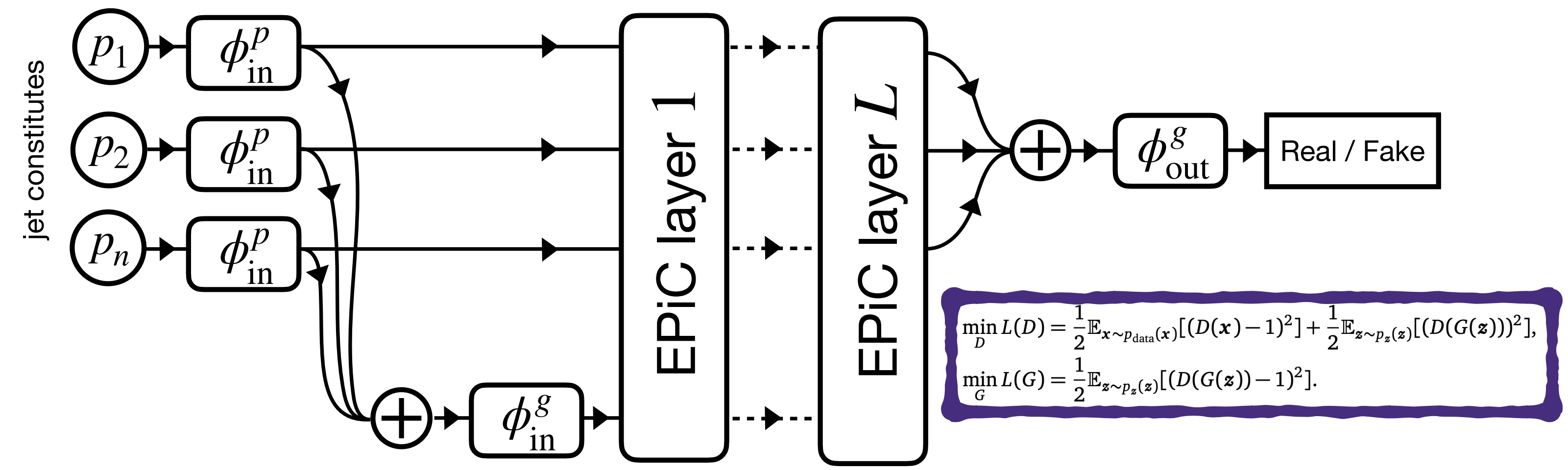
Generator



Discriminator

Detector simulation using ML

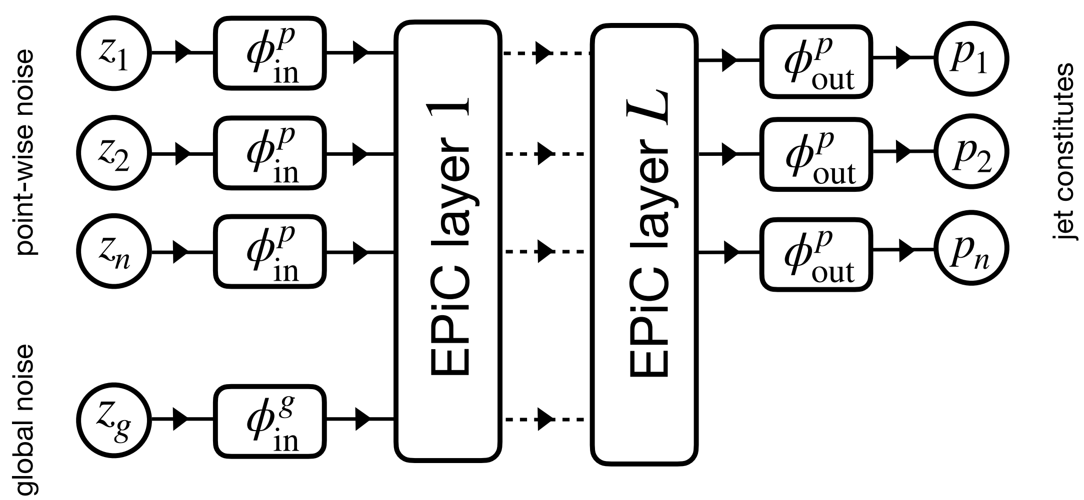




$$\min_D L(D) = \frac{1}{2}\mathbb{E}_{x\sim p_{\mathrm{data}}(x)}[(D(x)-1)^2] + \frac{1}{2}\mathbb{E}_{z\sim p_z(z)}[(D(G(z)))^2],$$

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

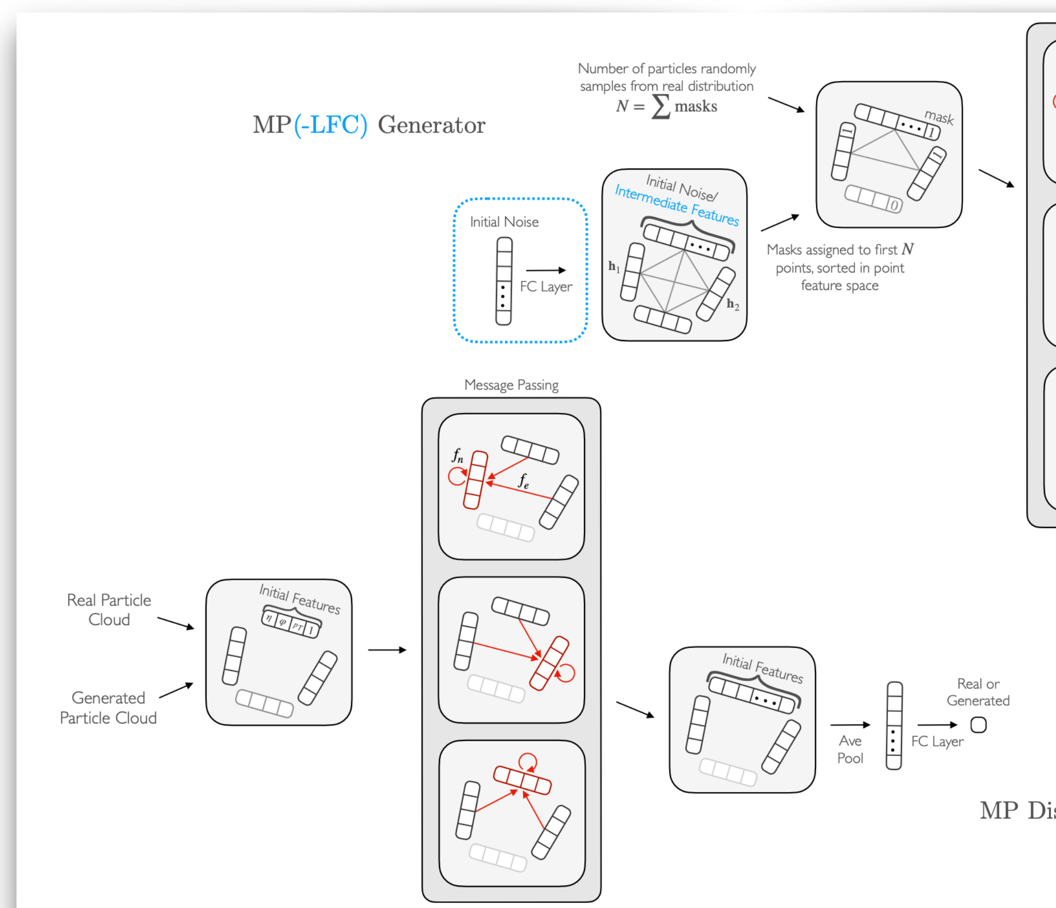
EPiC-GAN : SciPost Phys. 15, 130 (2023) Erik Buhmann, Gregor Kasieczka, Jesse Thaler



2305.10475

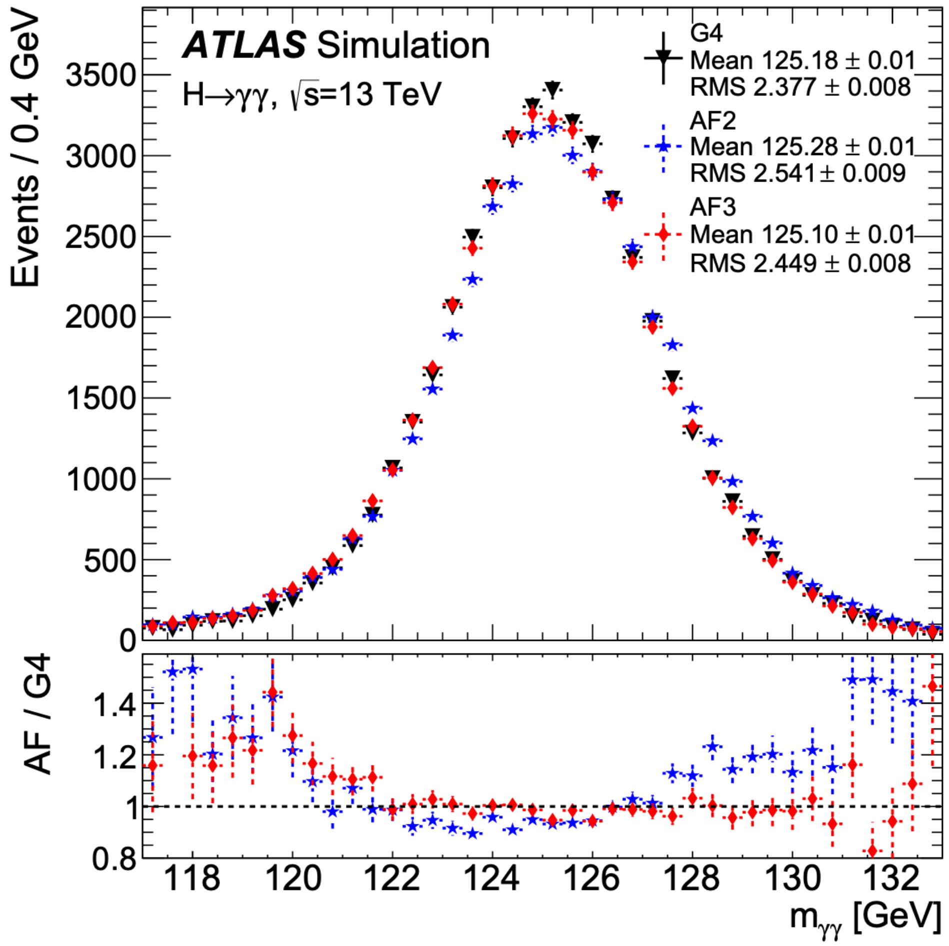
Jet Diffusion versus JetGPT — Modern Networks for the LHC

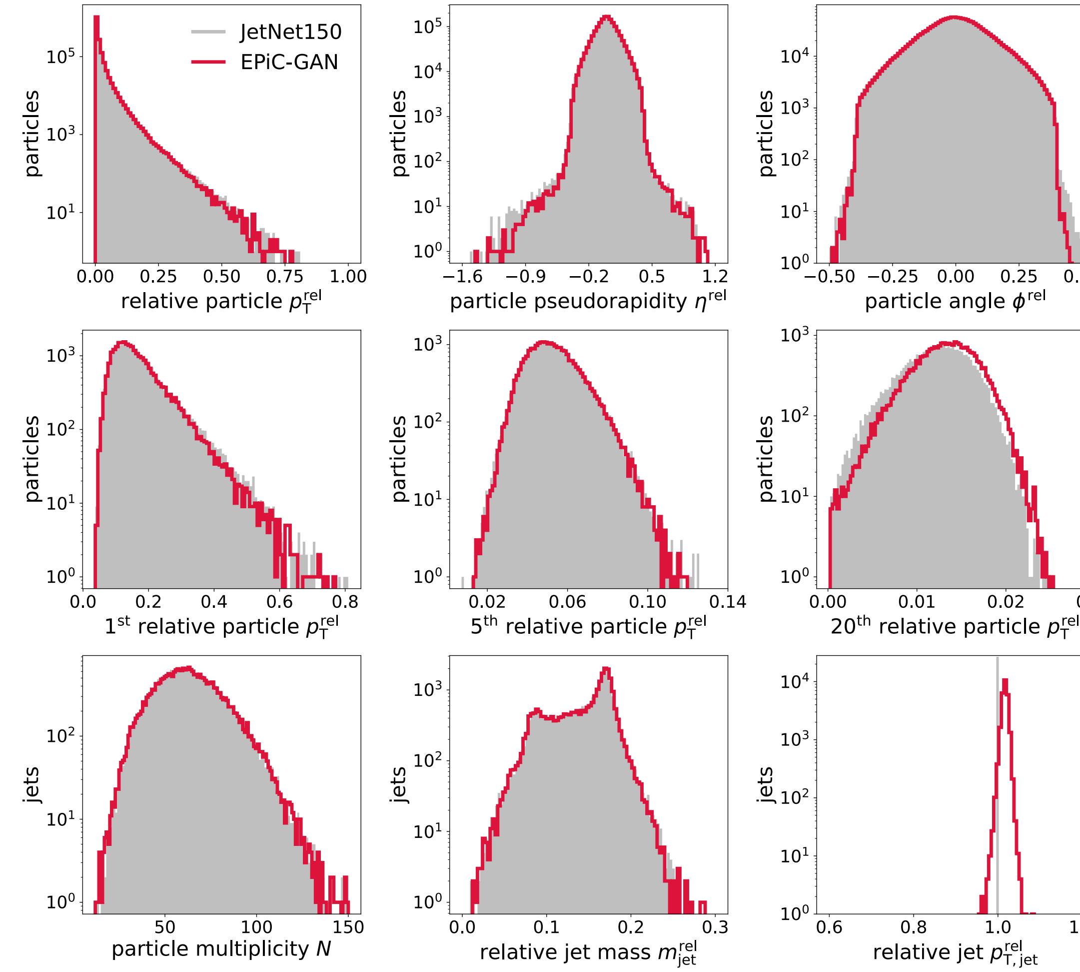
Anja Butter[1,2], Nathan Huetsch[1], Sofia Palacios Schweitzer[1], Tilman Plehn[1], Peter Sorrenson[3], and Jonas Spinner[1]



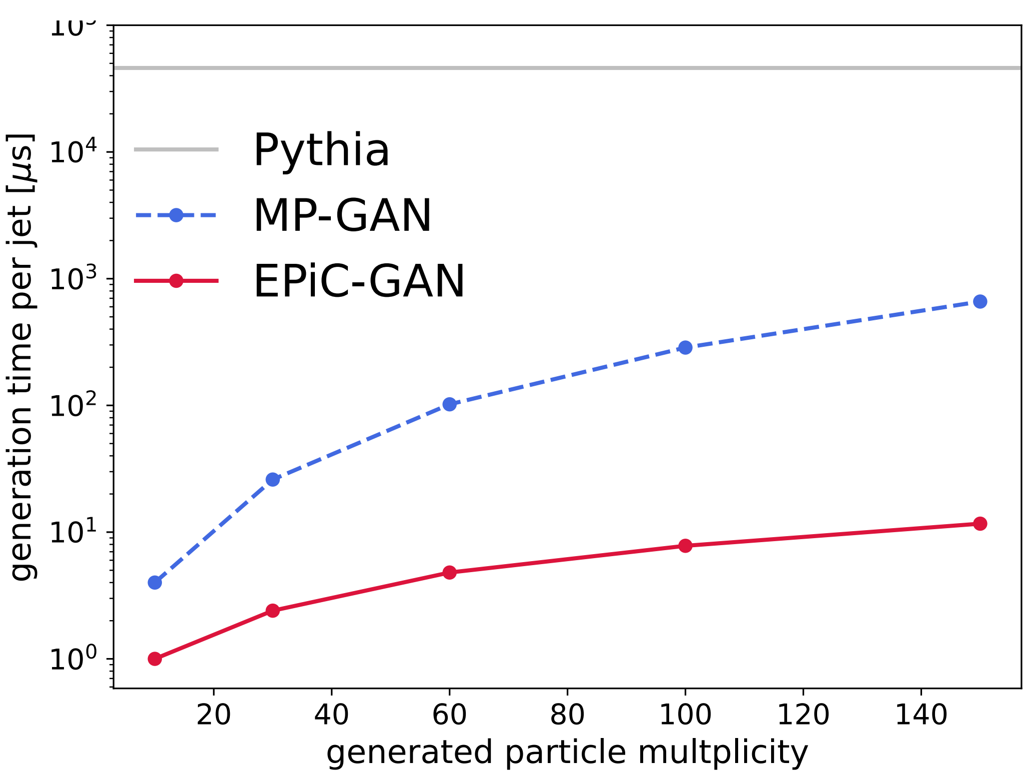
MPGAN :
arXiv : 2106.11535
R. Kansal et al

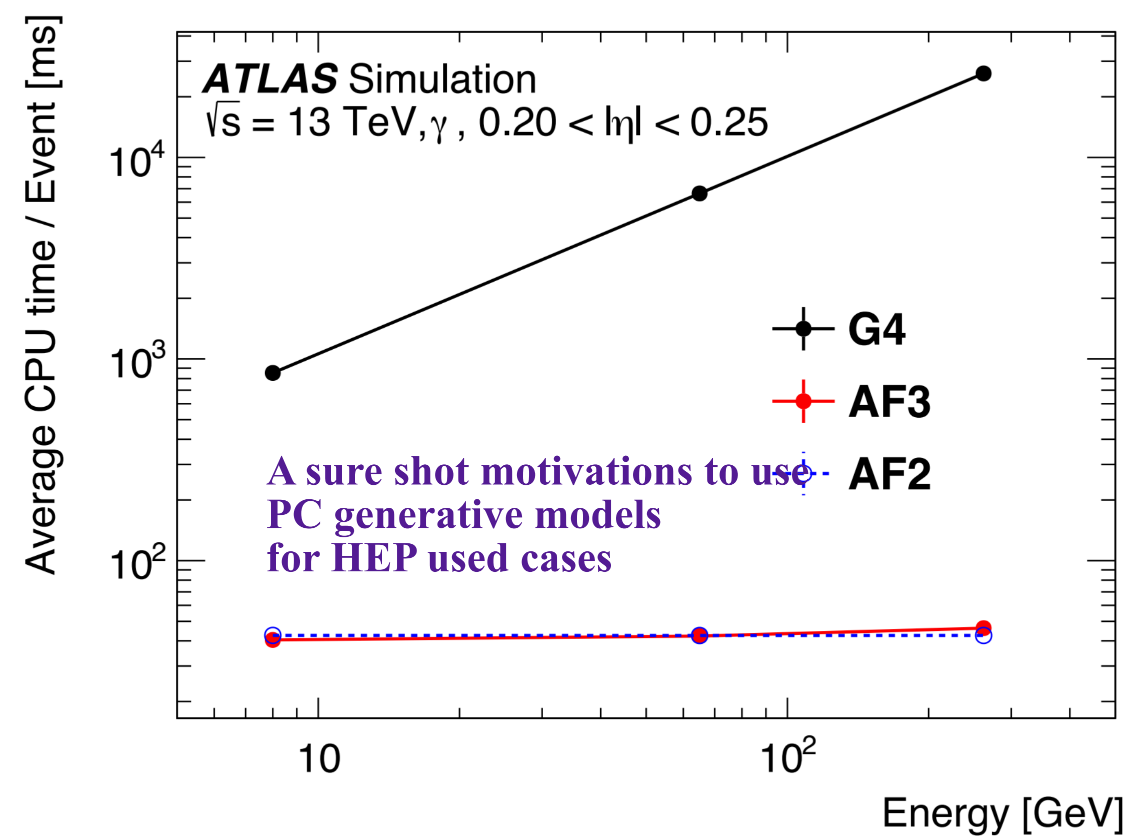
The major gain





arXiv : 2109.02551

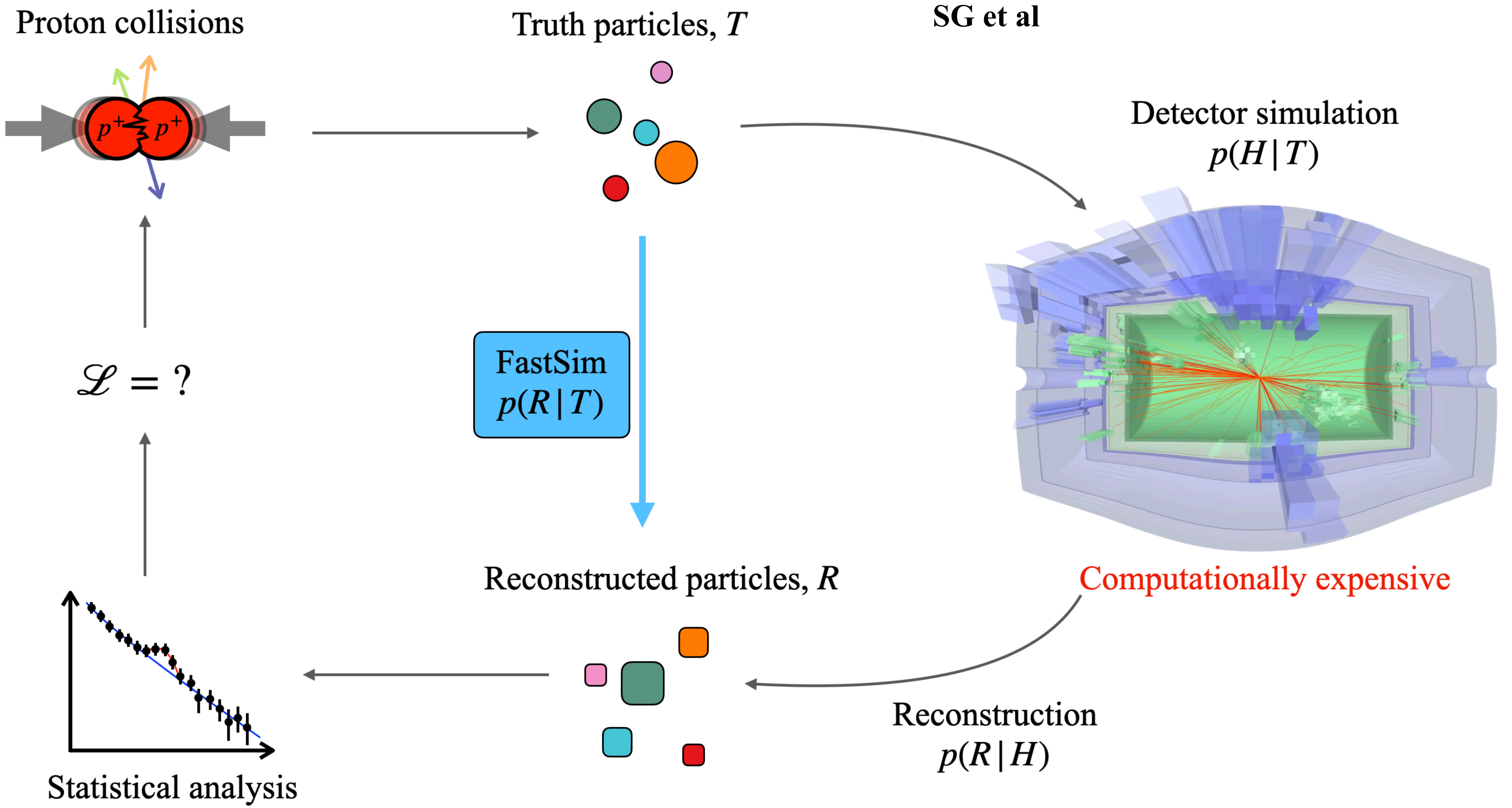




A generative model for Particle-flow

***Mach. Learn.: Sci. Technol.* 4 (2023) 045036**

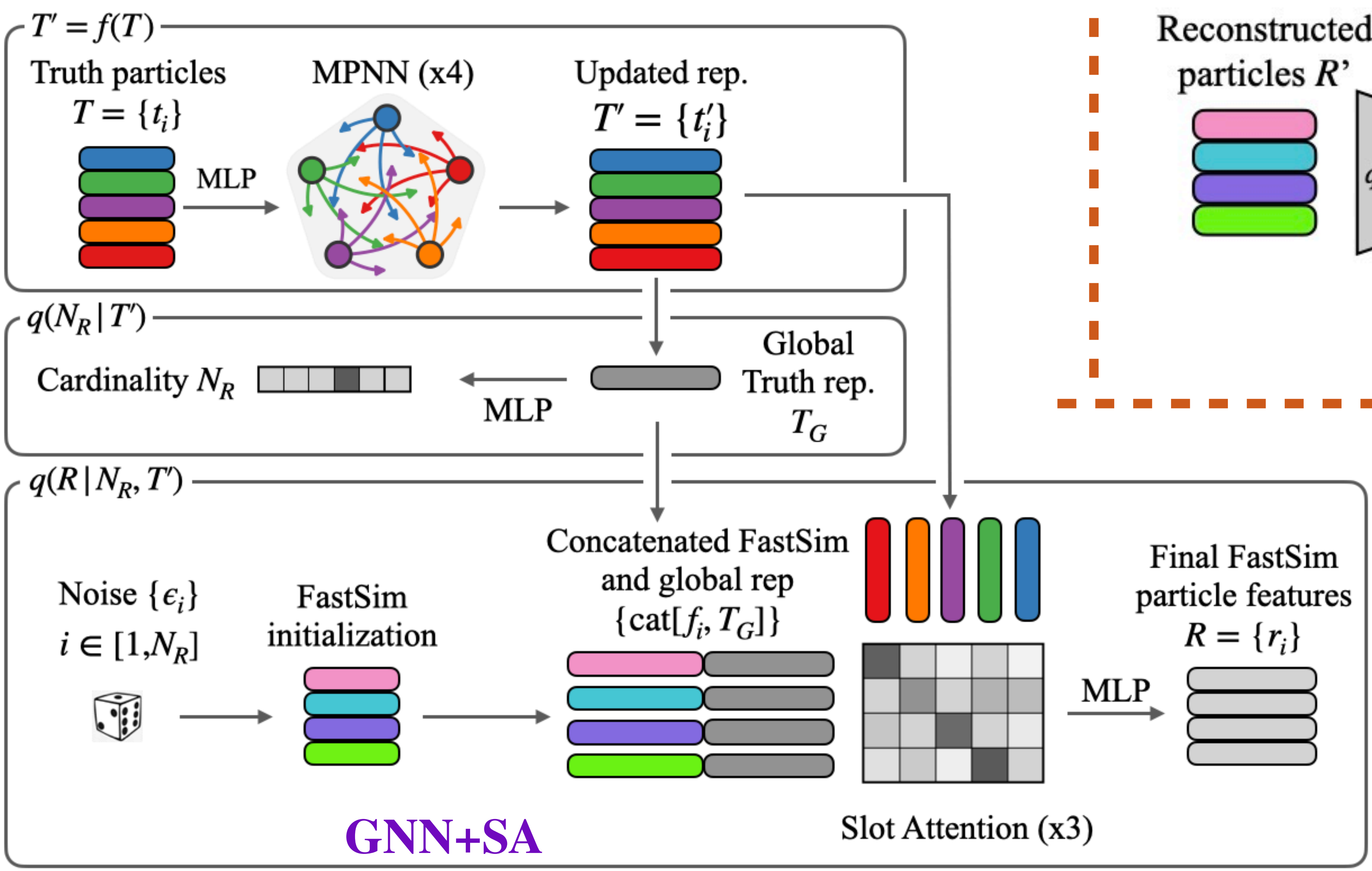
SG et al

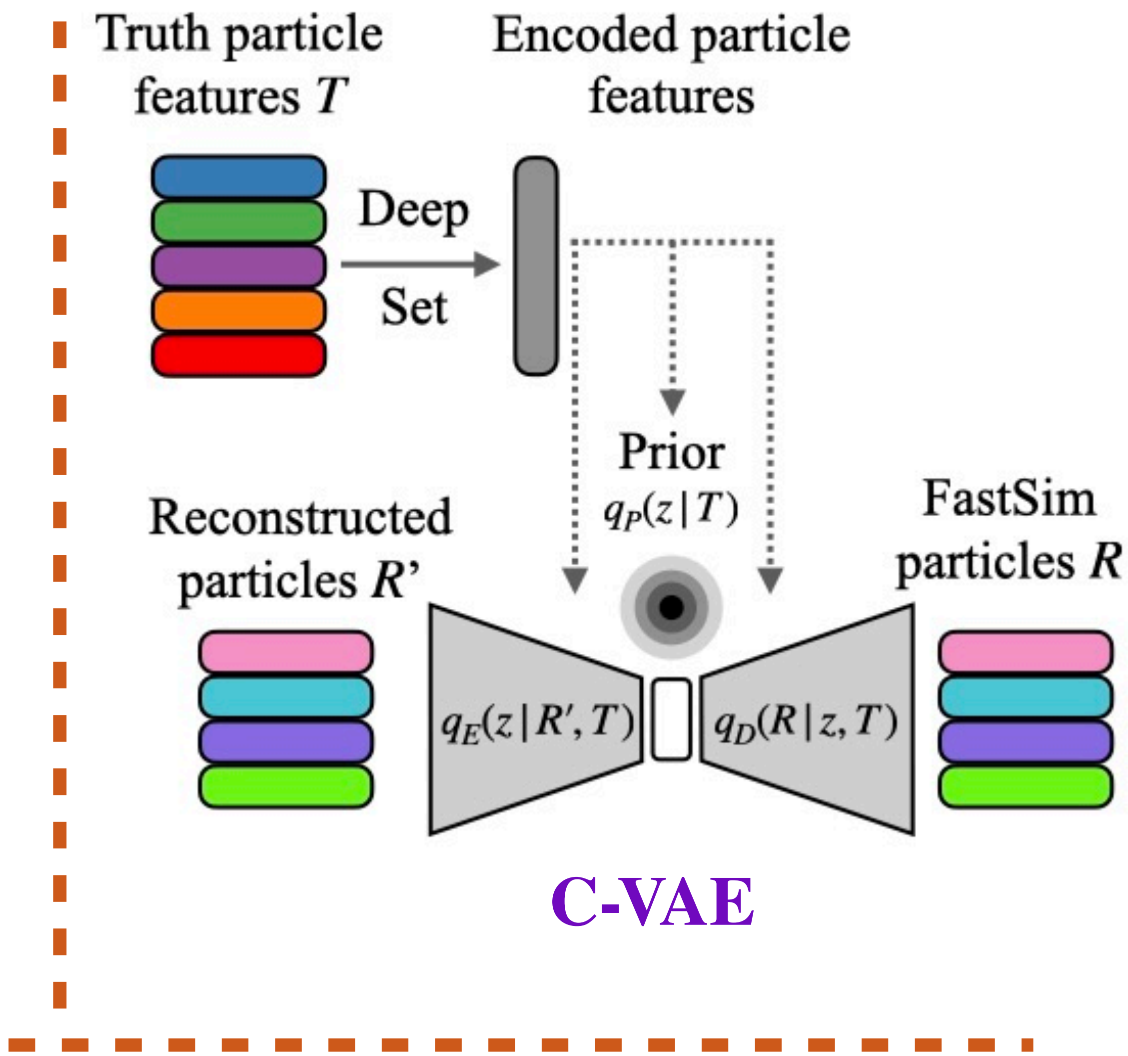


$$R \sim p(R|T) = \int \mathrm{d}H\, \delta(R(H) - R) p_{\mathrm{sim}}(H|T).$$

The task of constrained set generation

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

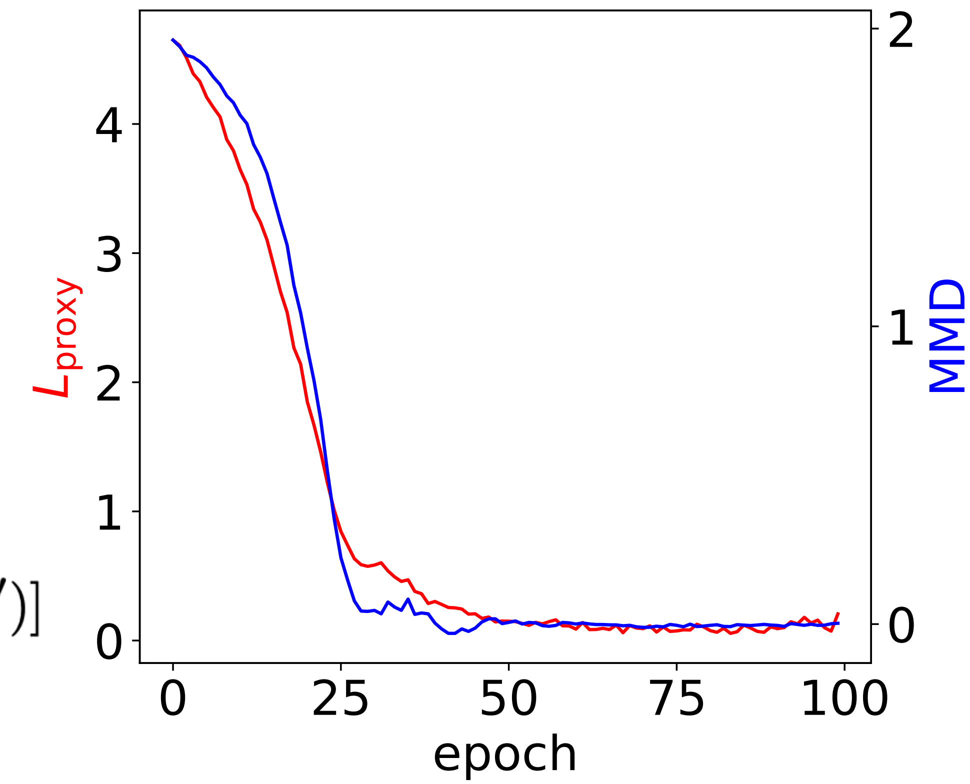




The task of constrained set generation

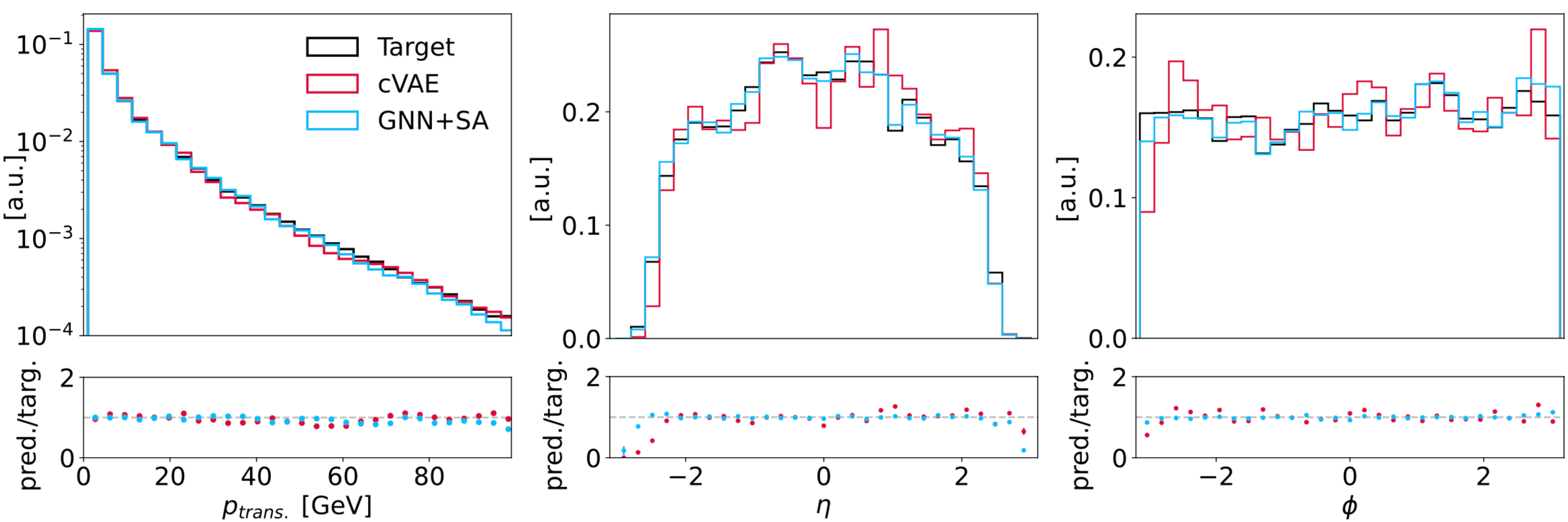
The cVAE training is done by optimizing negative evidence lower bound (ELBO) loss :

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



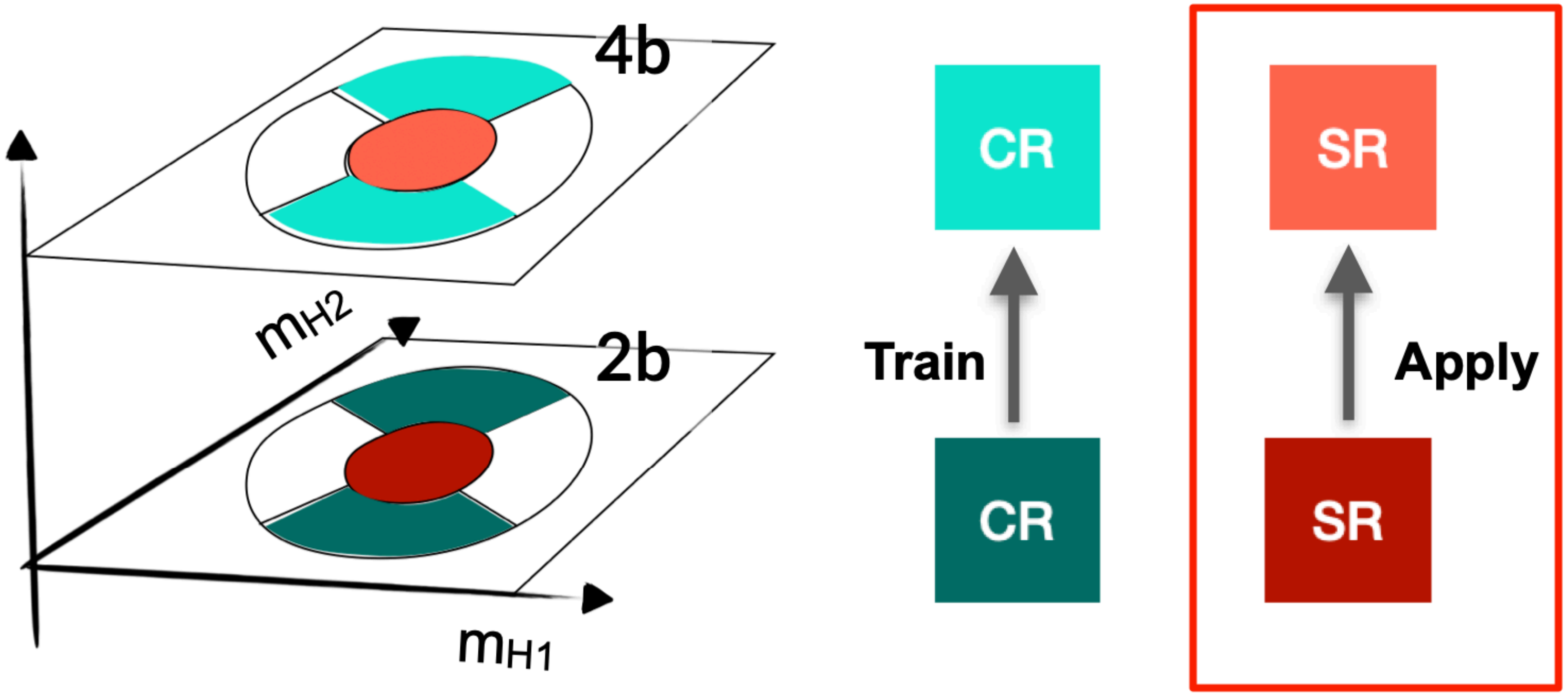
For GNN+SA, we try a regular Hungarian loss and also MMD (maximum mean discrepancy) :

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



Generating smooth backgrounds

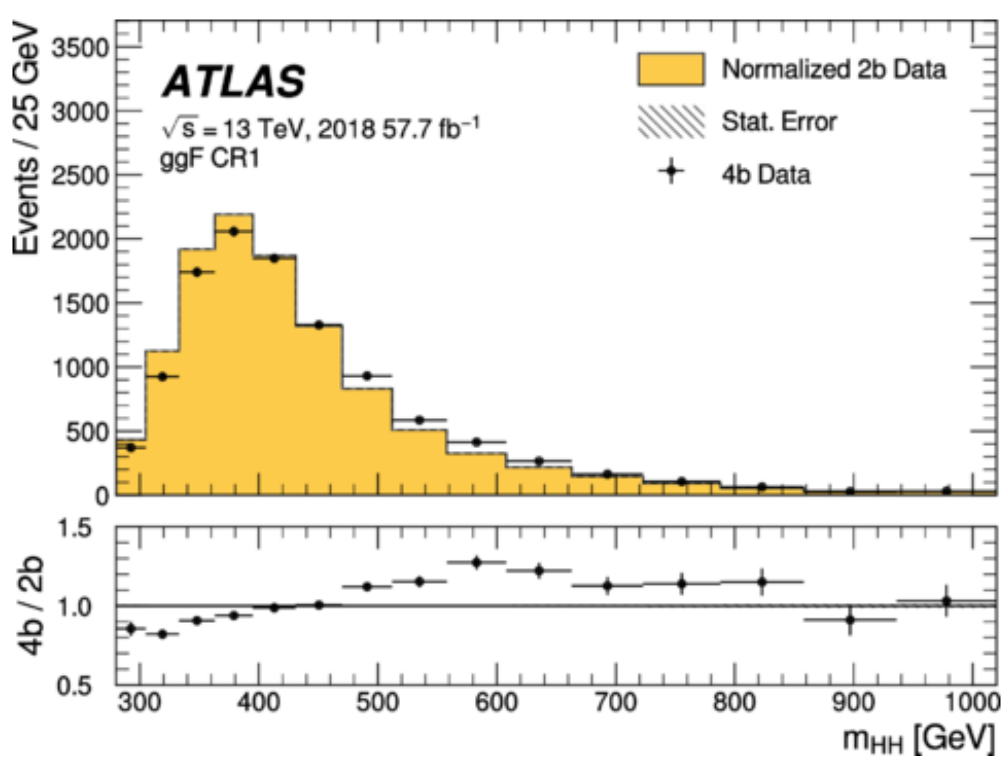
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HDBS-2019-29/

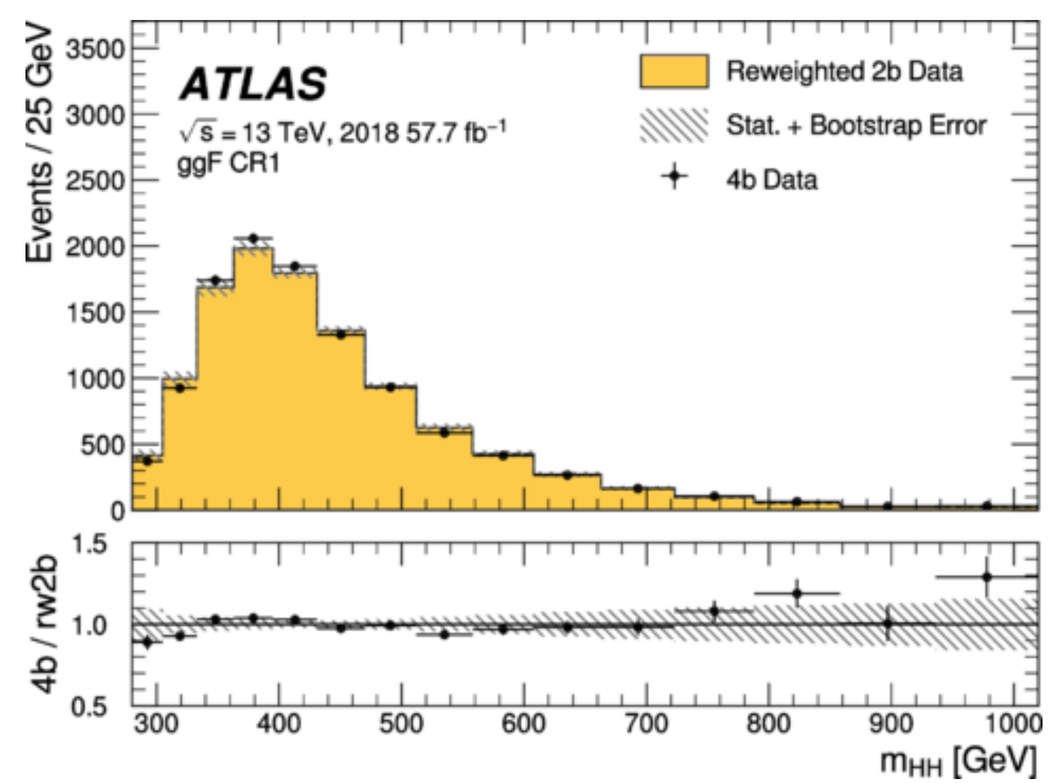


How do we want to learn?

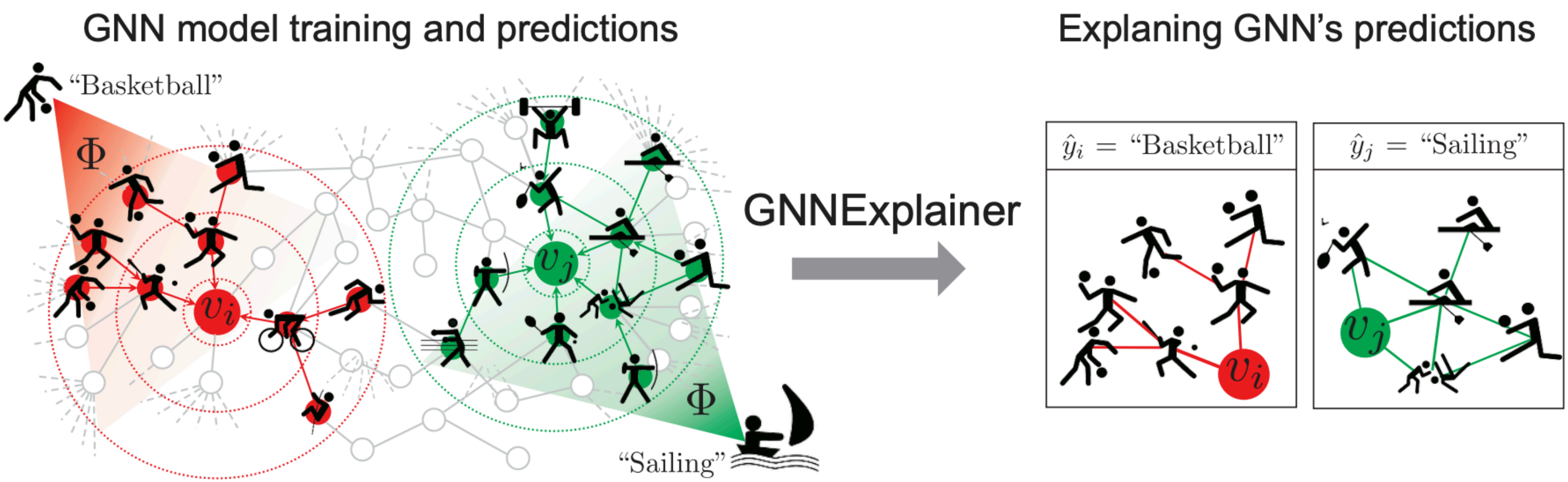
$$p_{4b} = \boxed{w(x)} \cdot p_{2b}(x),\ x \in \mathbb{R}^d$$

Can we learn w(x) using a flow method?





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

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

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.

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

Explainability for MLPF

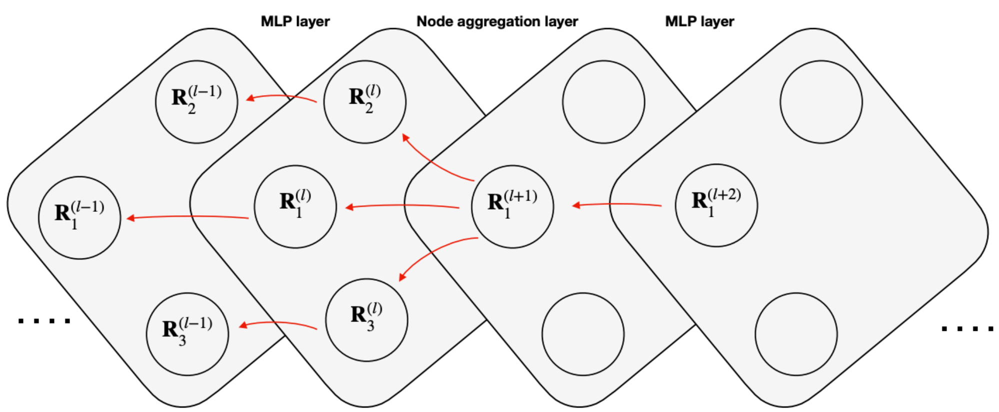
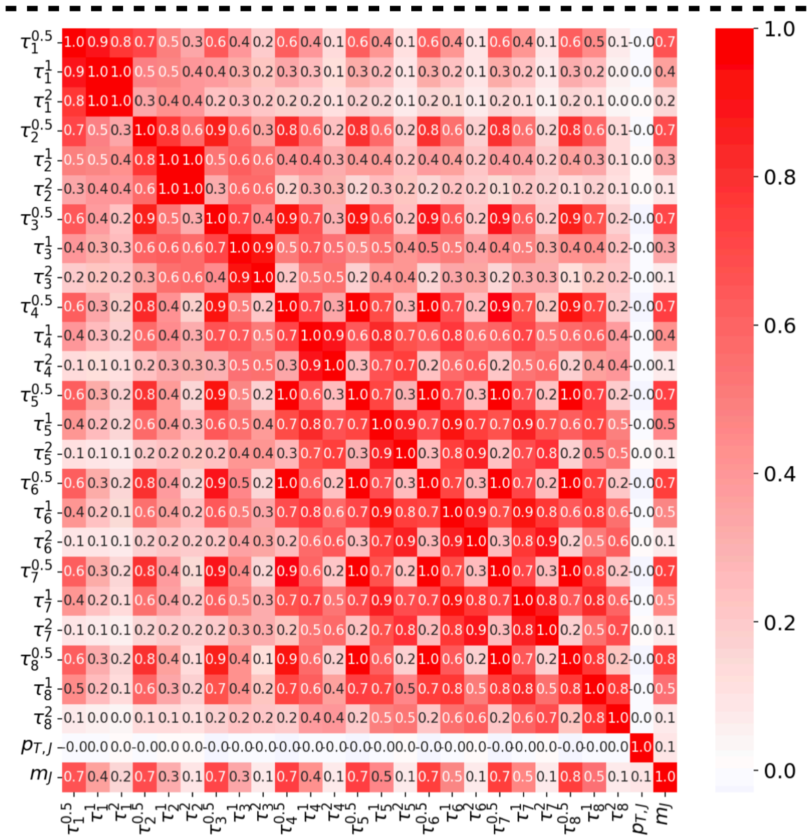


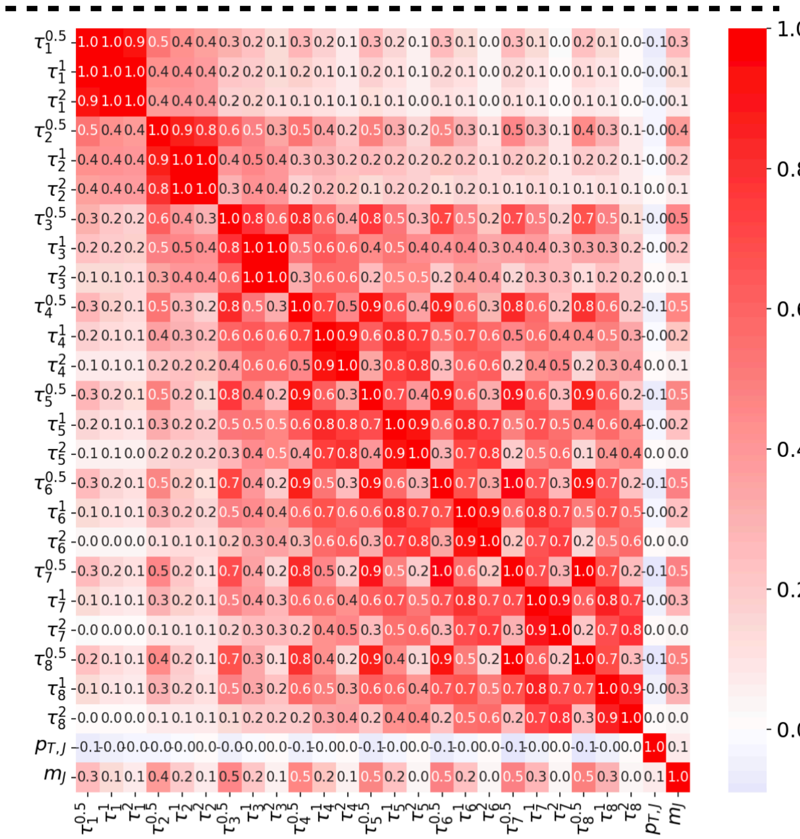
Figure 1: The flow of R-scores of node 1 across the different layers in MLPF. For MLP layers, the redistribution of R-scores follows the standard LRP rules [35, 36]. For the aggregation step in the message passing layer, the redistribution follows Equation 3. We only show three nodes for simplicity.

Feature correlation for top tagging.

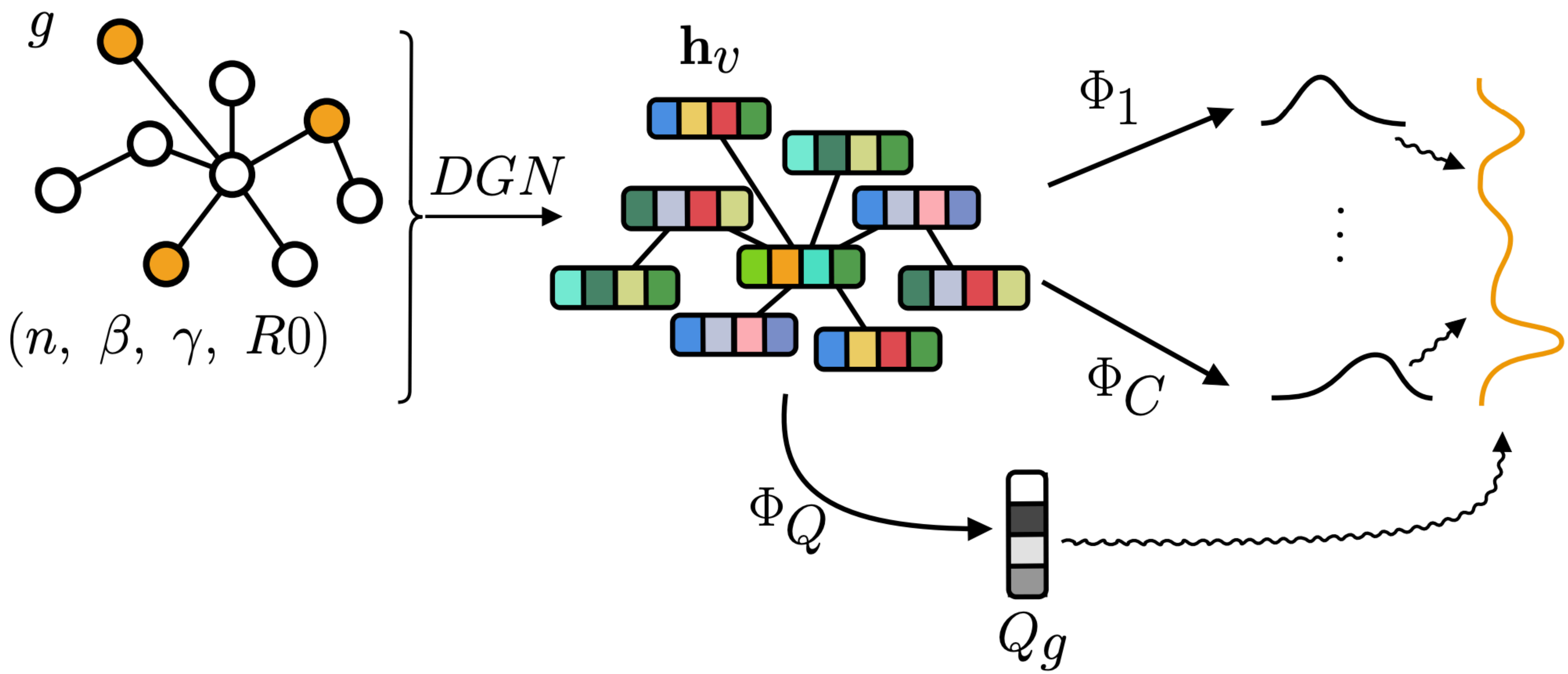
arXiv 2210.04371

Ayush Khot, Mark S. Neubauer, Avik Roy





Major thrust in immediate future : Uncertainty

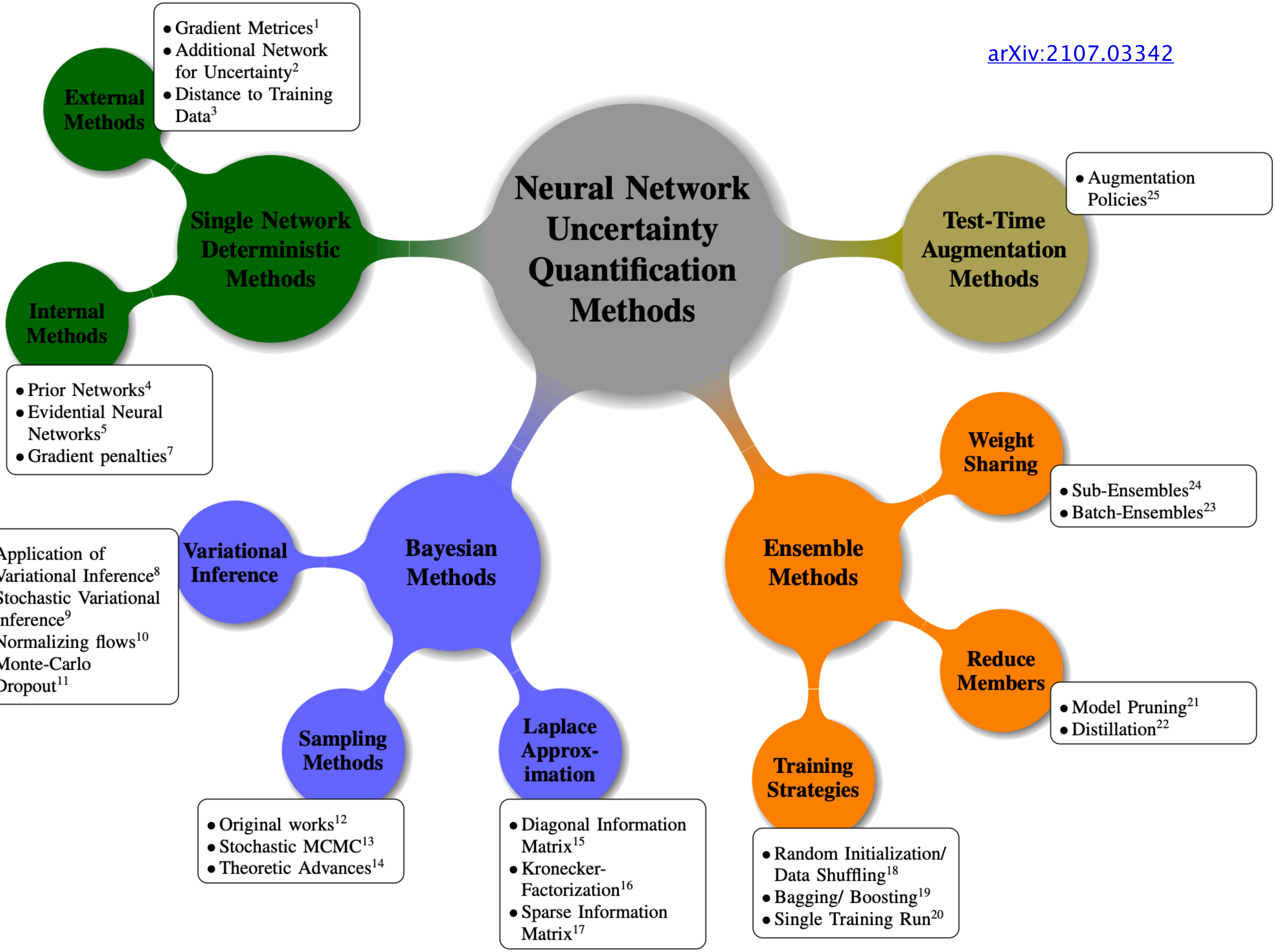


Reliable uncertainty estimation on ML based predictions are crucial for HEP
Only few Bayesian methods have been tested naively.

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

Major thrust in immediate future : Uncertainty

arXiv:2107.03342



The path we will take

Basic Motivation

- Why are we doing ML?
- The basic LHC operations with ML.
- Synergies through eye of ML.

The supervised ML

- Jet tagging : the most successful venture.
- Tracking & Calorimetery
- The Particle-Flow algorithm

Generative models

- Different paradigms
- Few physics cases.
- Interpretable ML & uncertainty estimation.

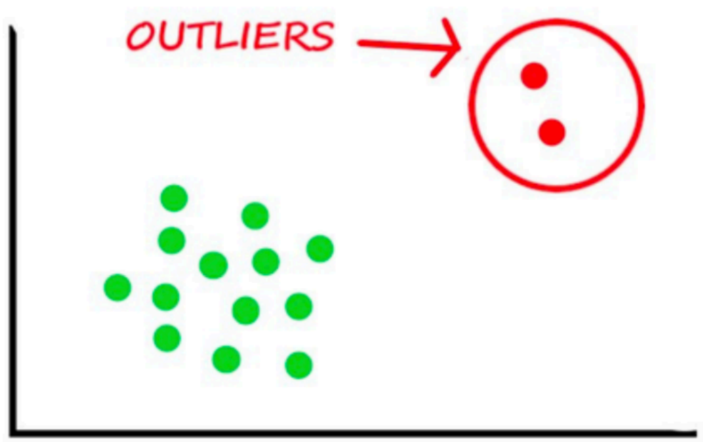
The anomaly detection saga

- The outliers & density estimate
- How to translate this in new signals?

THE ROAD MAP

Anomaly detection : the general types

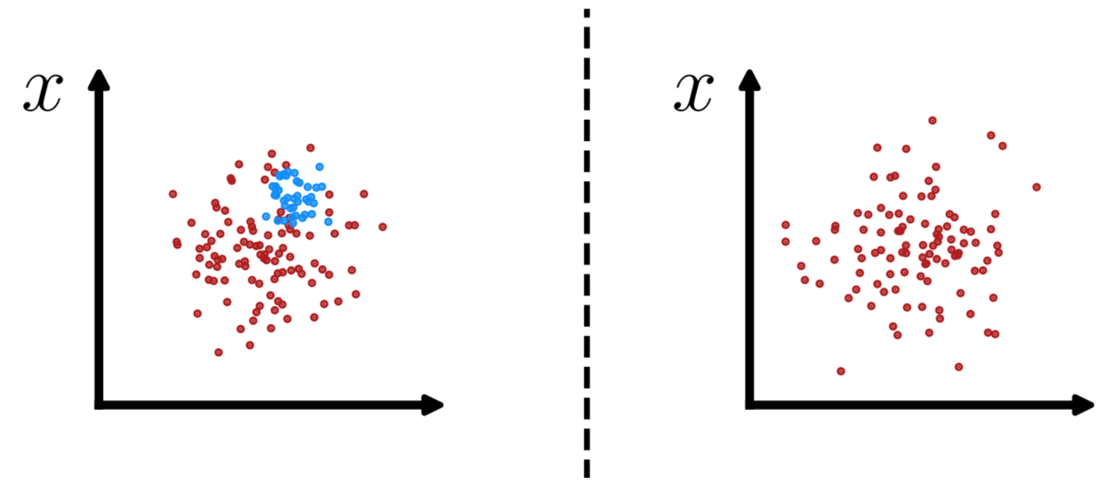
a. Outlier detection ("point anomalies")



low $p(x)$

[Autoencoders]
Farina, Nakai & **DS** 1808.08992
Heimel et al 1808.08979
Cerri et al 1811.10276
…
Density estimation
Caron, Hendriks, Verheyen 2106.10164
…

b. Overdensity detection ("group anomalies")



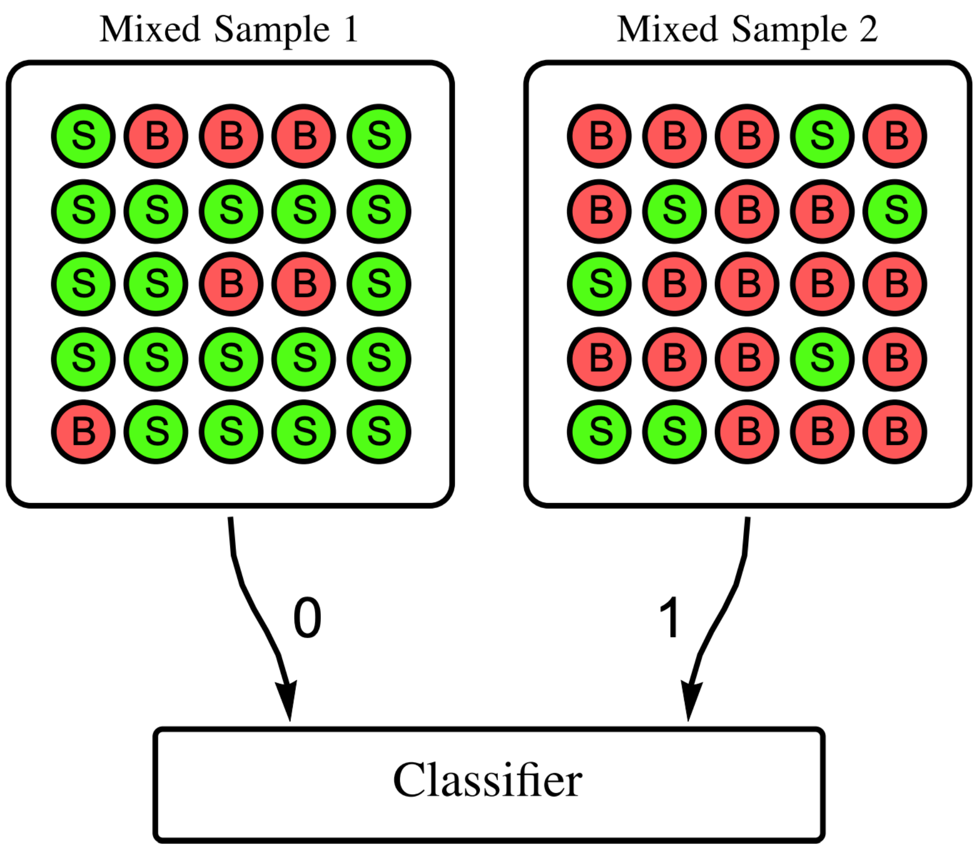
high $\frac{p_{data}(x)}{p_{bg}(x)}$

Data vs bg test statistic
D'Agnolo et al 1806.02350,1912.12155, 2111.13633

Enhanced bump hunts
CWoLa Hunting [Collins, Howe & Nachman 1805.02664, 1902.02634]
ANODE [Nachman & **DS** 2001.04990]
CATHODE [**DS+** Hallin et al 2109.00546, 2210.14924]
CURTAINS [Raine et al 2203.09470]

Slide taken from David Shih

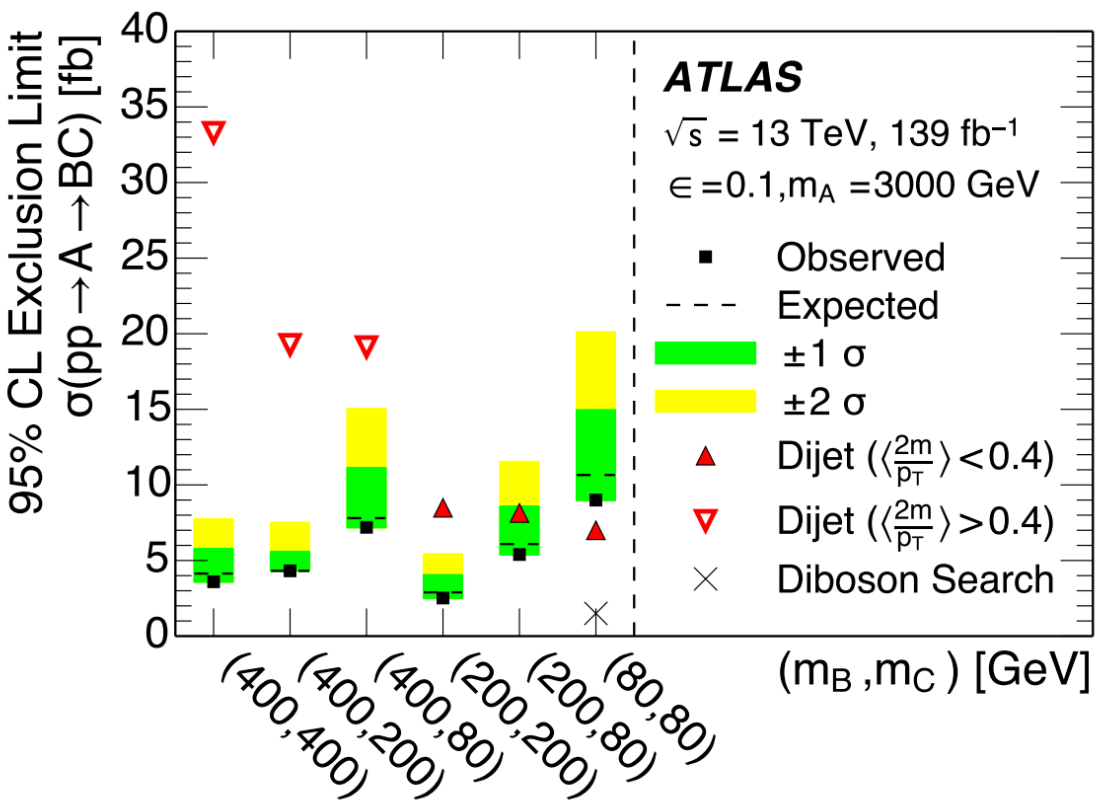
CWoLa : a weakly supervised method

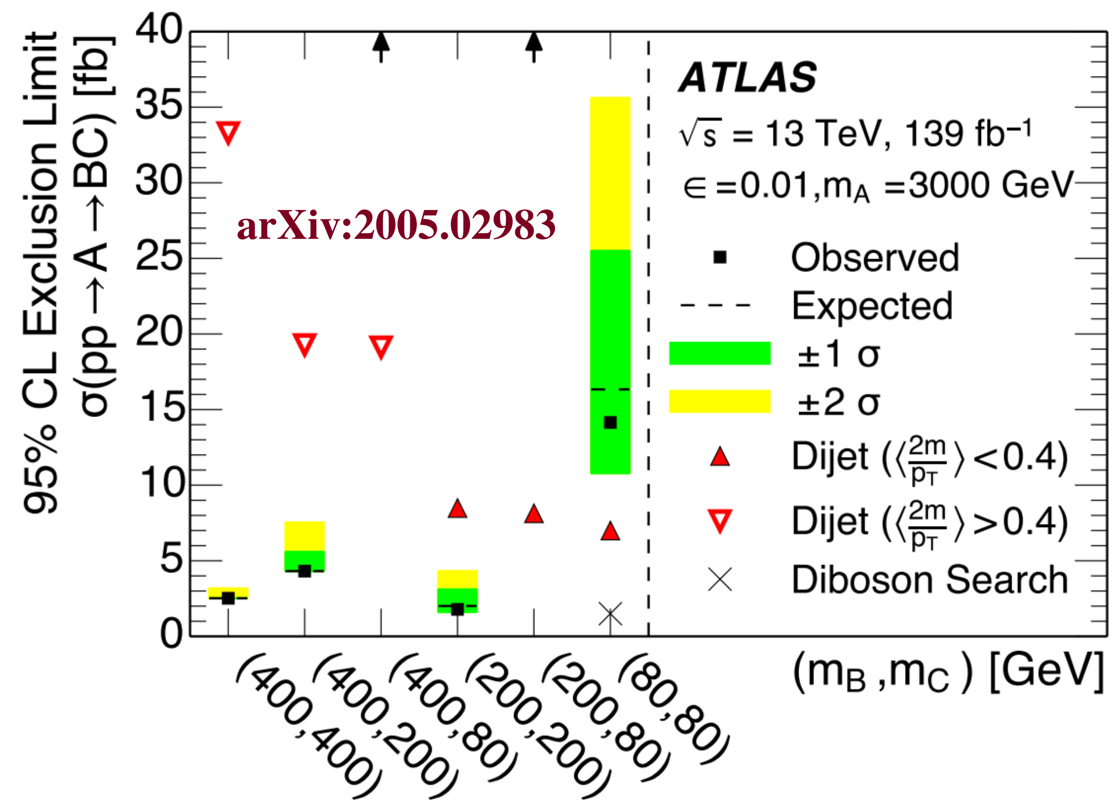


$$L_{M_1/M_2} = \frac{p_{M_1}}{p_{M_2}} = \frac{f_1\, p_S + (1-f_1)\, p_B}{f_2\, p_S + (1-f_2)\, p_B} = \frac{f_1\, L_{S/B} + (1-f_1)}{f_2\, L_{S/B} + (1-f_2)};$$

$$\partial_{L_{S/B}} L_{M_1/M_2} = (f_1 - f_2)/(f_2 L_{S/B} - f_2 + 1)^2 > 0.$$

arXiv:1708.02949





ABCD background estimation with ML

arXiv > hep-ph > arXiv:2007.14400

High Energy Physics – Phenomenology

[Submitted on 28 Jul 2020]

ABCDisCo: Automating the ABCD Method with Machine Learning

Gregor Kasieczka, Benjamin Nachman, Matthew D. Schwartz, David Shih

$$N_{A,b}^{\text{predicted}} \equiv \frac{N_{B,a}N_{C,a}}{N_{D,a}}.$$

$$\begin{aligned} N_{A,\ell} &= N_\ell \Pr(f \geq f_c \ \text{ and } \ g \geq g_c|\ell) \\ N_{B,\ell} &= N_\ell \Pr(f \geq f_c \ \text{ and } \ g < g_c|\ell) \\ N_{C,\ell} &= N_\ell \Pr(f < f_c \ \text{ and } \ g \geq g_c|\ell) \\ N_{D,\ell} &= N_\ell \Pr(f < f_c \ \text{ and } \ g < g_c|\ell), \end{aligned}$$

Find a ML driven way to find the variables f & g

Single DisCo

$$\mathcal{L}[f(X)] = \mathcal{L}_{\text{classifier}}[f(X), y] + \lambda \, \text{dCorr}^2_{y=0}[f(X), X_0],$$

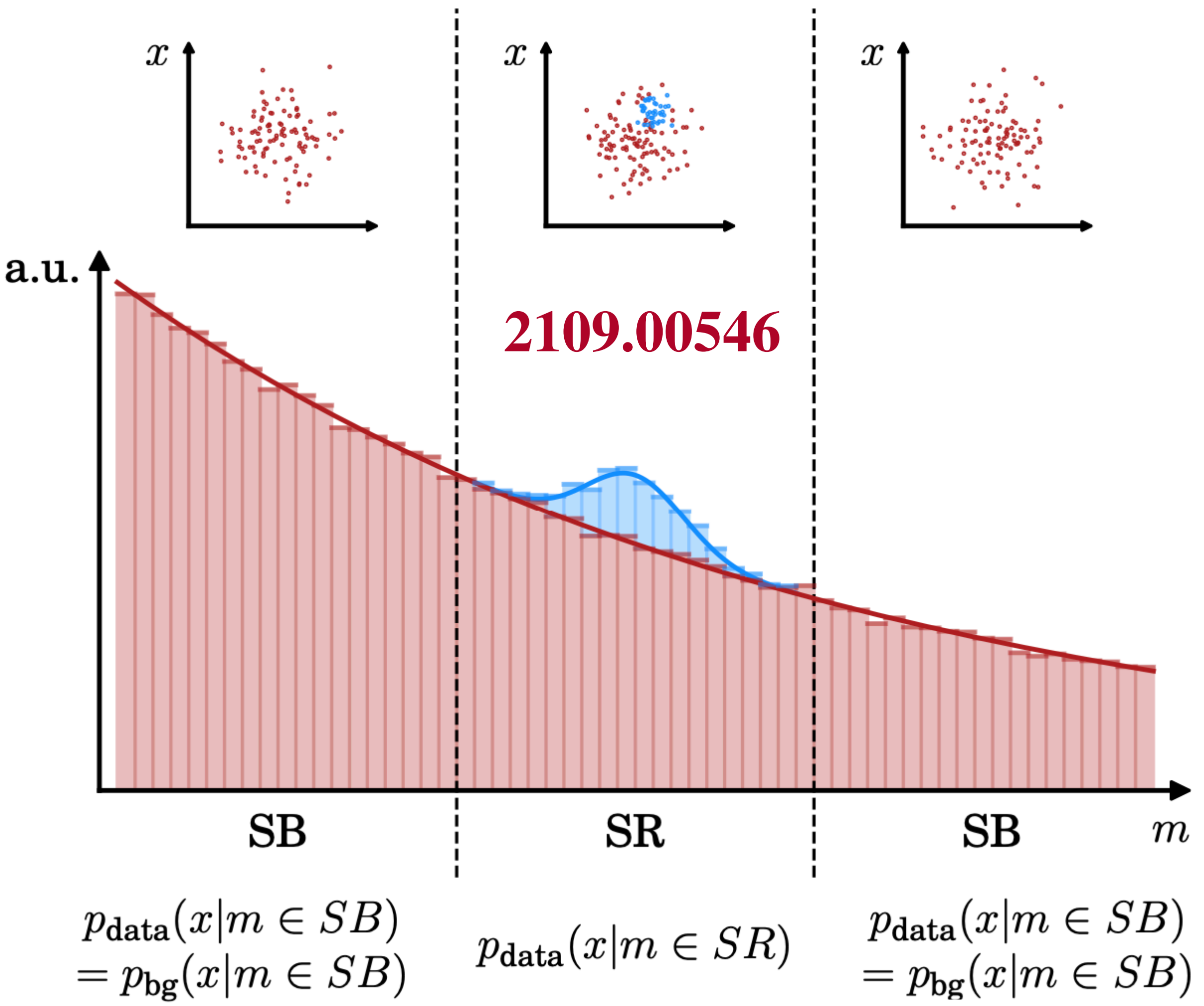
Double DisCo

$$\mathcal{L}[f, g] = \mathcal{L}_{\text{classifier}}[f(X), y] + \mathcal{L}_{\text{classifier}}[g(X), y] + \lambda \, \text{dCorr}^2_{y=0}[f(X), g(X)],$$

$$\begin{aligned} \text{dCov}^2[f, g] = &\Big\langle |f - f'| \times |g - g'| \Big\rangle \\ &+ \Big\langle |f - f'| \Big\rangle \times \Big\langle |g - g'| \Big\rangle - 2\Big\langle |f - f'| \times |g - g''| \Big\rangle \end{aligned}$$

$$\text{dCorr}^2[f, g] = \frac{\text{dCov}^2[f, g]}{\text{dCov}[f, f]\, \text{dCov}[g, g]}$$

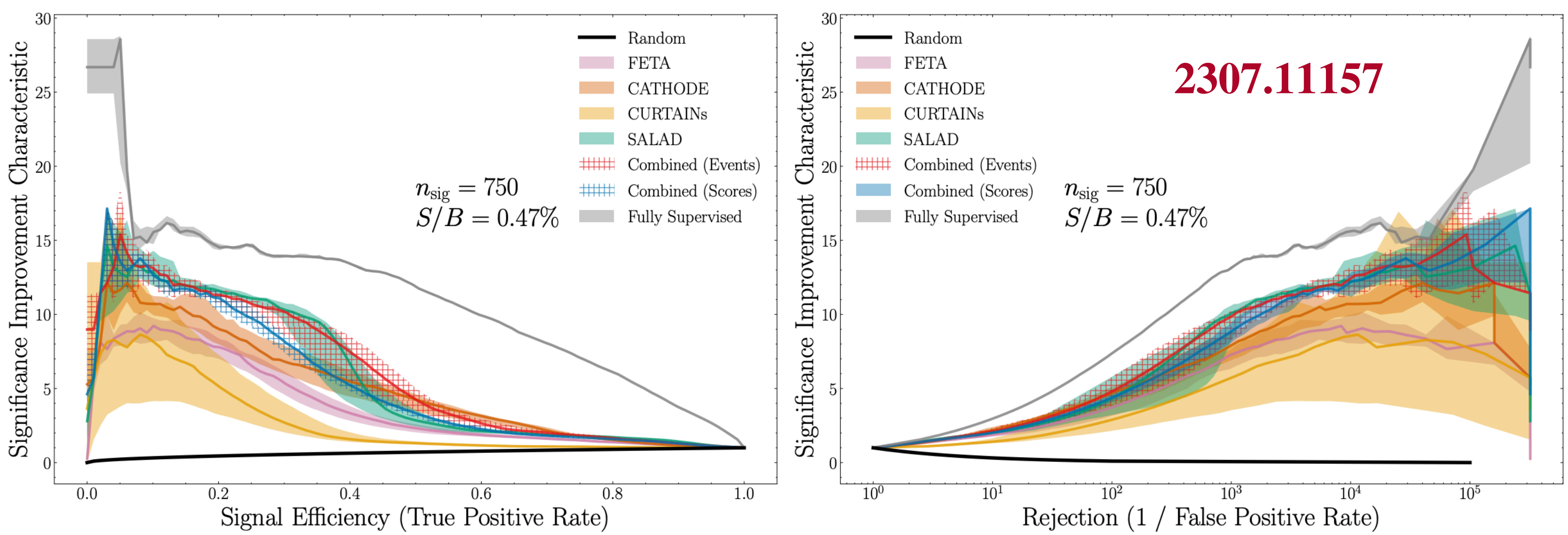
Anomaly detection with density estimation



Main goal : find $R(x) = \dfrac{p_{\mathrm{data}}(x)}{p_{\mathrm{bg}}(x)}$ **in the signal region, estimating from SB.**

Train two NF separately in SR and SB to learn $p_{\mathrm{data}}(x\,|\,m \in \mathrm{SR})$ and $p_{\mathrm{b}}(x\,|\,m \in \mathrm{SB})$

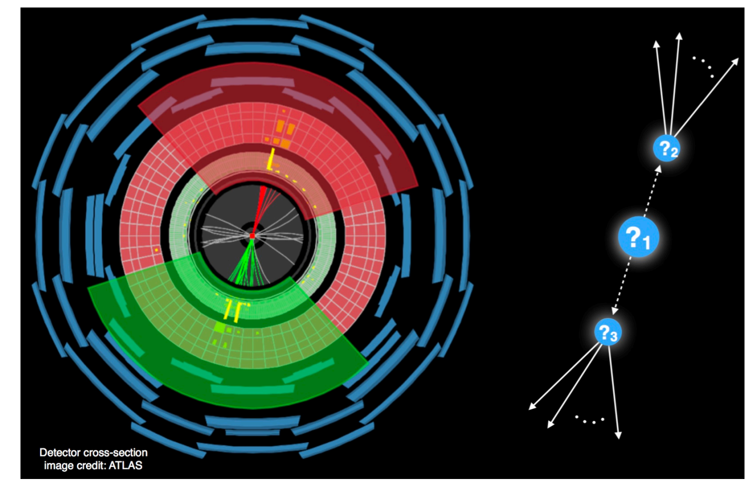
This is ANODE method. On top of this add CWoaLa to create CATHODE.



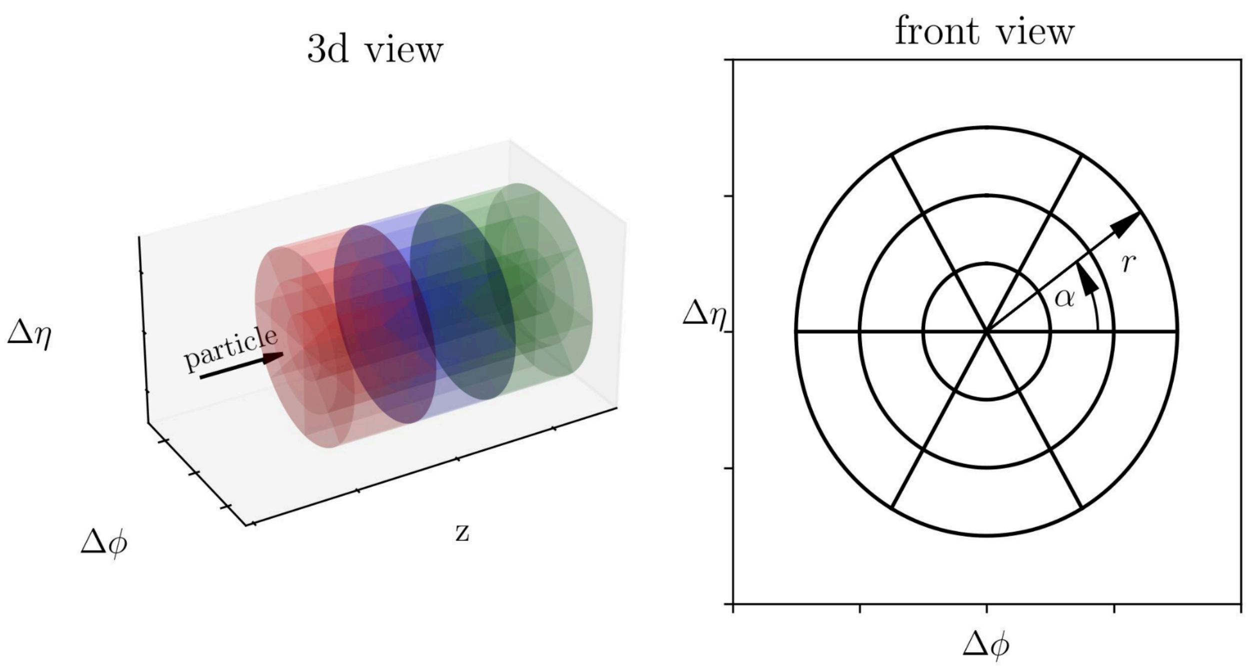
Open datasets for testing your idea!

https://zenodo.org/records/3547722

https://lhco2020.github.io/homepage/

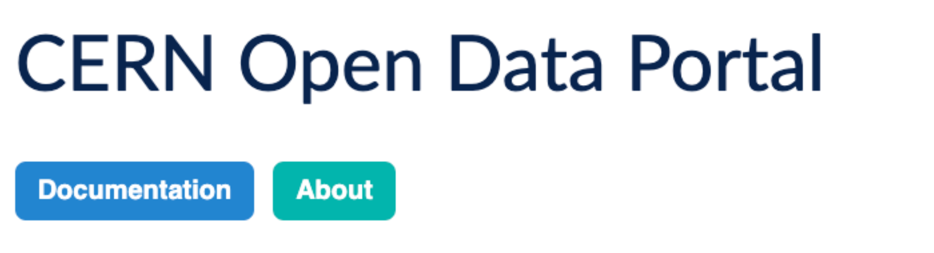


LHC Olympics



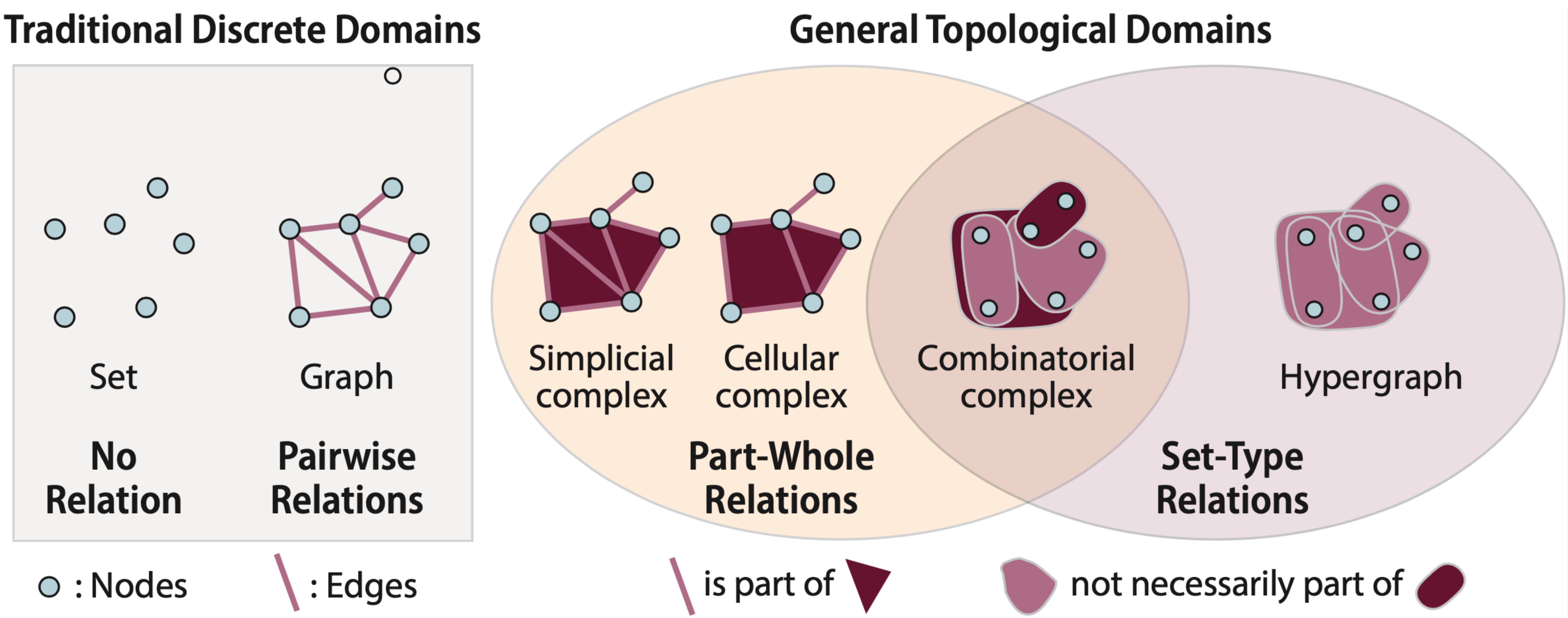
Calo Challenge

https://calochallenge.github.io/homepage/



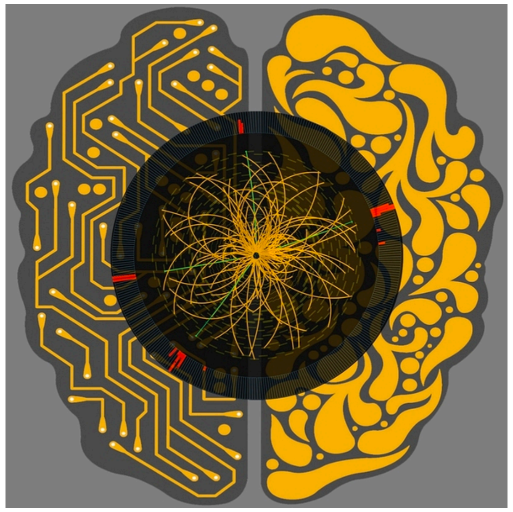
An example of next frontiers

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



Let’s formulate the questions

Image: FermiLab

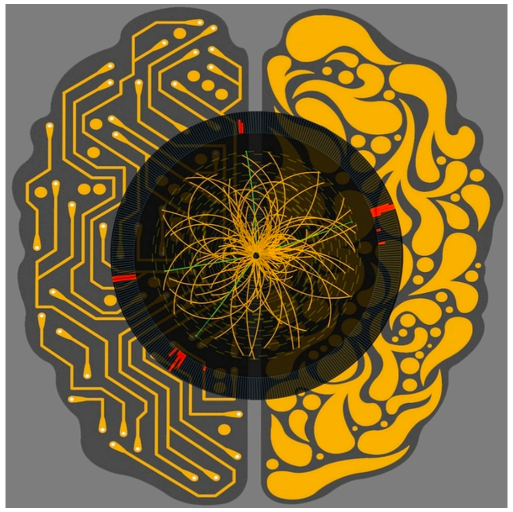


- ☑ ML is here to stay with HEP.
- ☑ When looked through the lens of ML, what’s are the core questions to answer for these soft signals?
- ☑ Interpretability and uncertainty estimation is a corner stone which we should emphasize.
- ☑ The collider community should talk with mathematicians/comp-sc and other branches of natural science who are using the similar methods and exchange ideas.

https://iml-wg.github.io/HEPML-LivingReview/

Let’s formulate the questions

Image: FermiLab



☑ML is here to stay with HEP.

☑When looked through the lens of ML, what’s are the core questions to answer for these soft signals?

☑Interpretability and uncertainty estimation is a corner stone which we should emphasize.

☑The collider 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/

THANK YOU