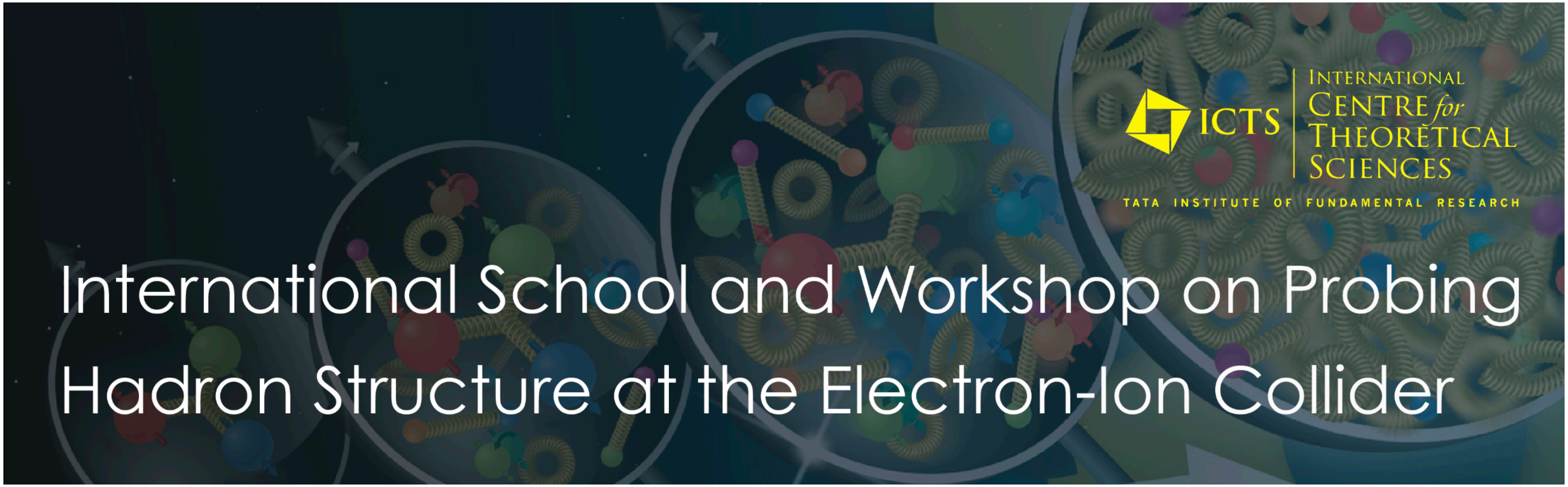
Machine Learning Knowledge exchange : from LHC to EIC



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

The collider experimental program

Simulation based approach :

Underlying theory ⇒ Physics simulation ⇒ Detector level observables ⇒ Physics inference

The real experiment :

The dice that nature rolls ⇒ Build the machine & do the experiment ⇒ Extract the unfolded distribution ⇒ Statistical inference

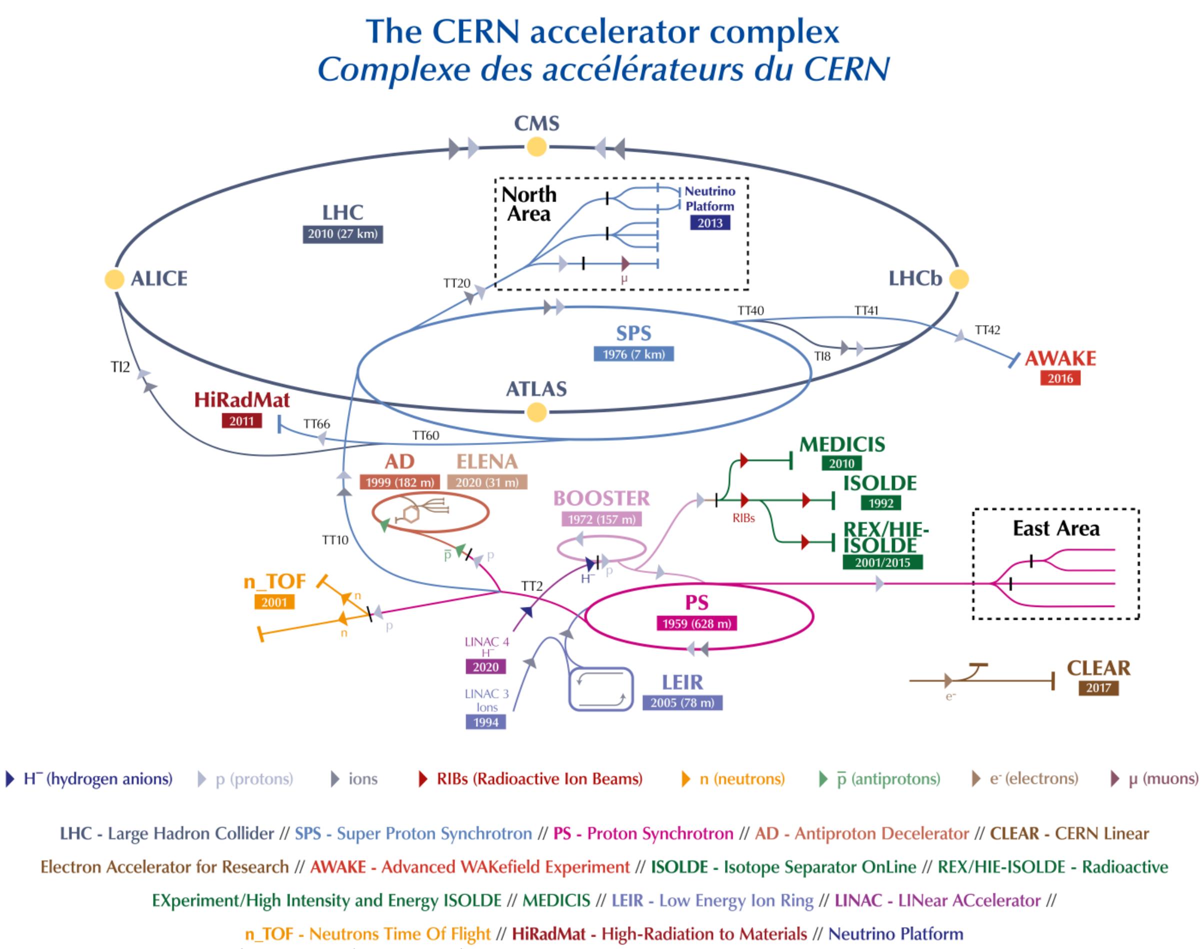
The collider experimental program

Simulation based approach :

Underlying theory ⇛ Physics simulation ⇛ Detector level observables ⇛ Physics inference

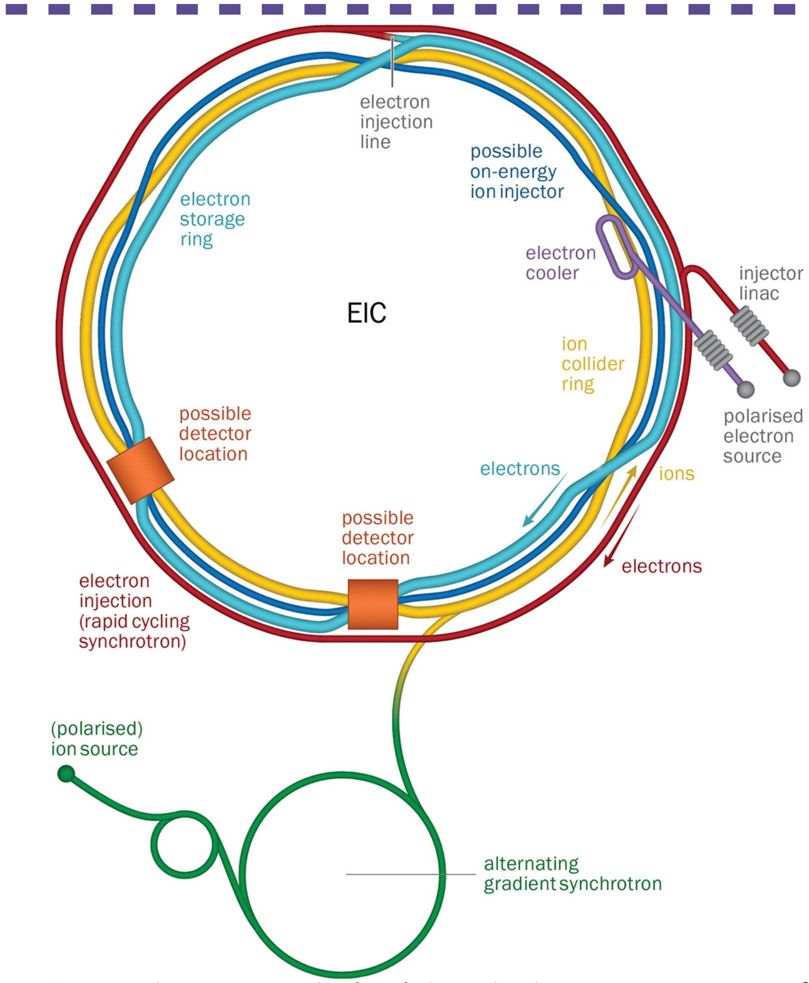
The real experiment :

The dice that nature rolls ⇛ Build the machine & do the experiment ⇛ Extract the unfolded distribution ⇛ Statistical inference



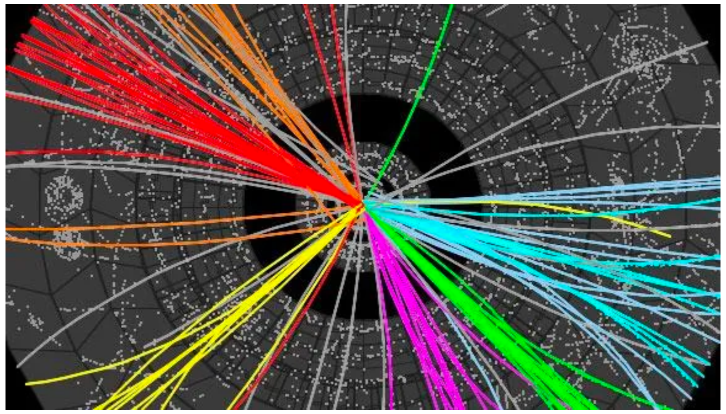
https://cds.cern.ch/images/CERN-GRAPHICS-2022-001-1

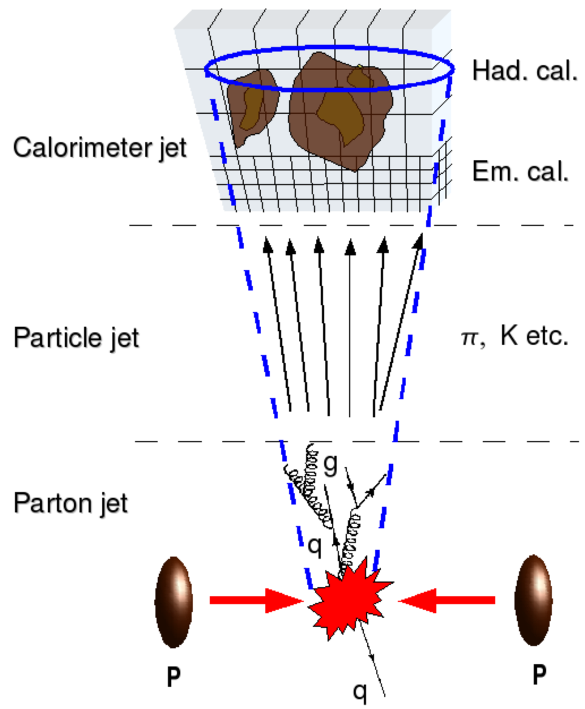
$\Longleftrightarrow$

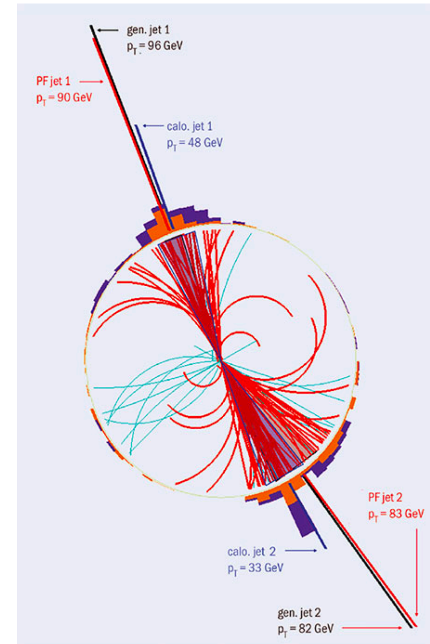


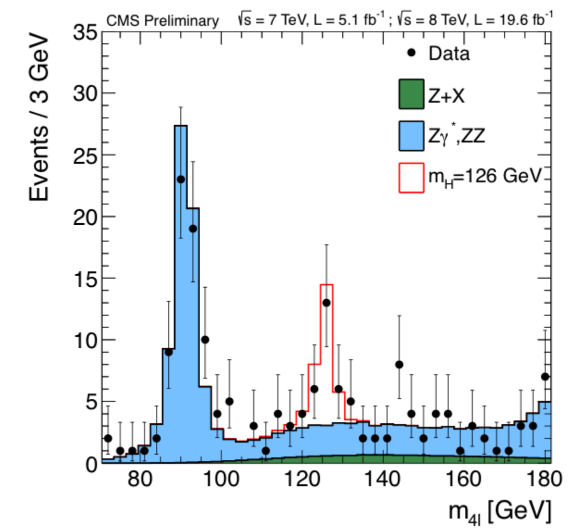
https://cerncourier.com/wp-content/uploads/2021/09/CCUSASupp21_EIC_fig1.jpg

The collider program flow-chain







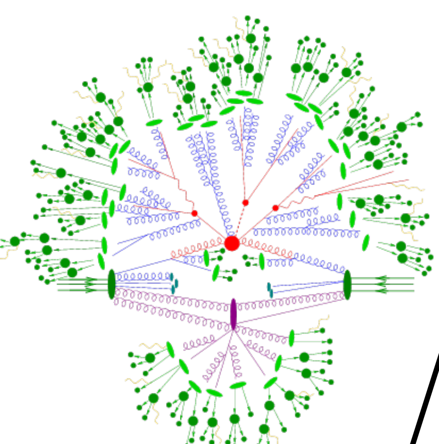


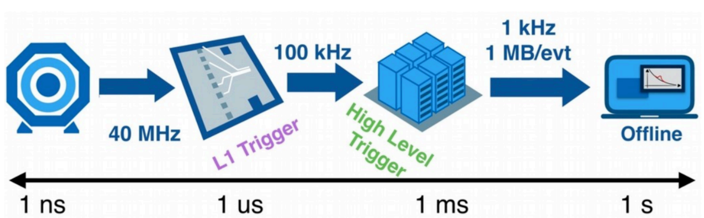
Trigger
Online process

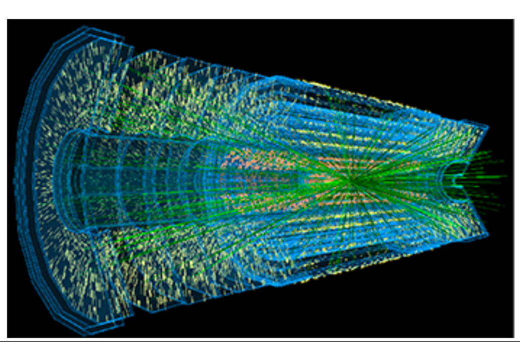
Object rec
PID + calib

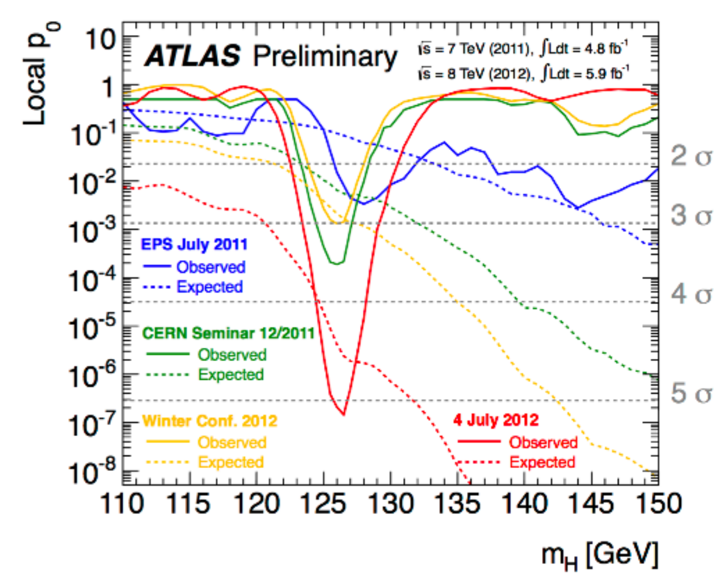
Event rec
sig/bkg class.

Analysis
Unfolding/matching







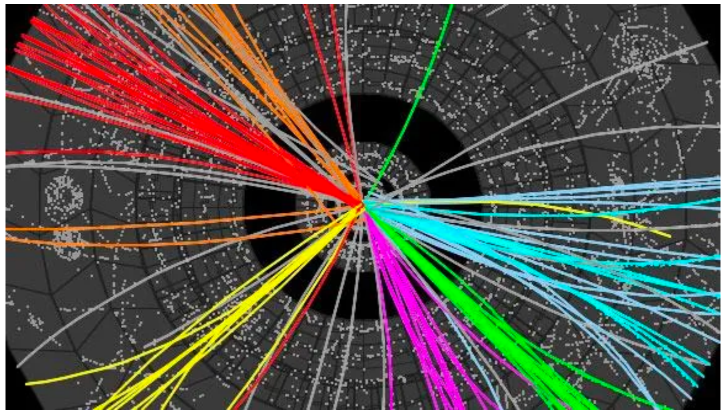


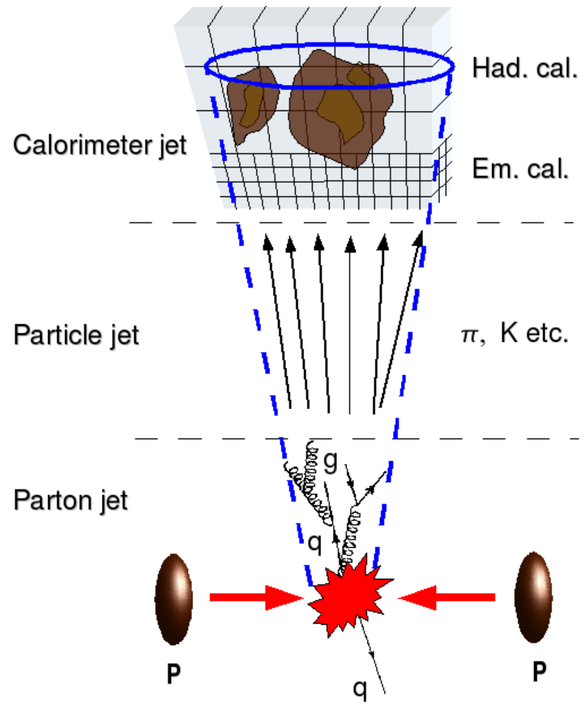
ME + PS generation
Detector Simulation

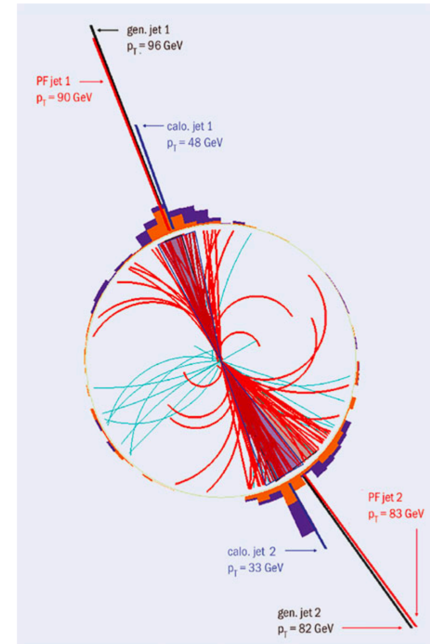
Inference

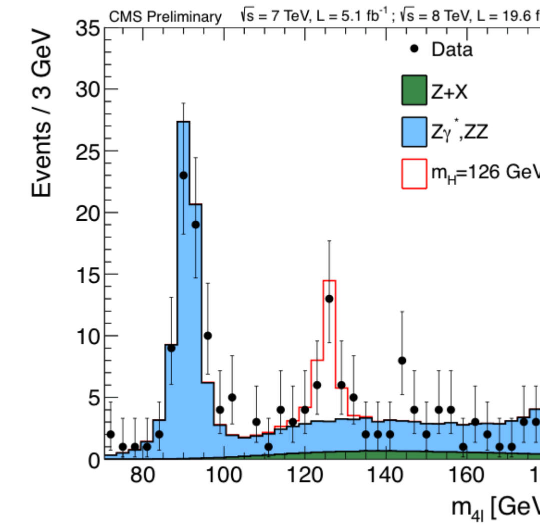
The collider program flow-chain

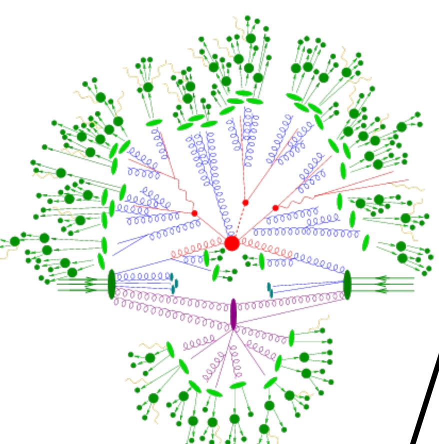
ML can (and will) play a role at every instance of this flow chain.









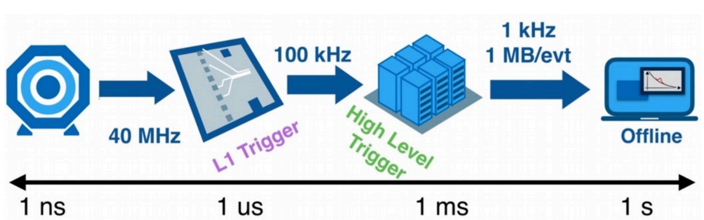


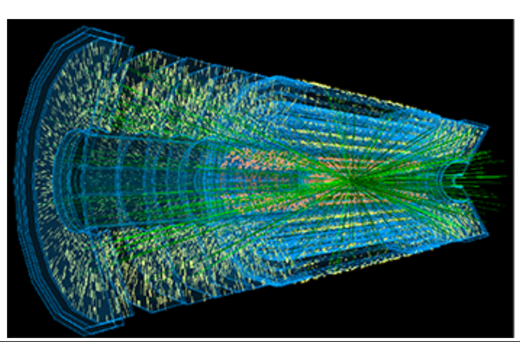
Trigger
Online process

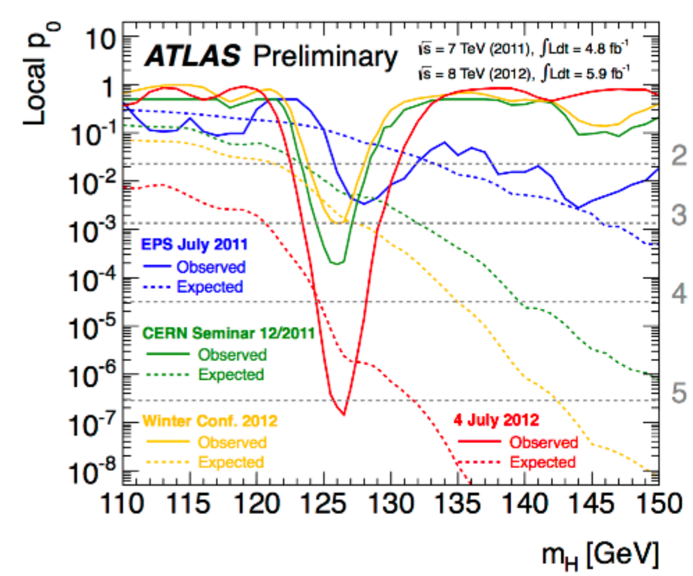
Object rec
PID + calib

Event rec
sig/bkg class.

Analysis
Unfolding/matching



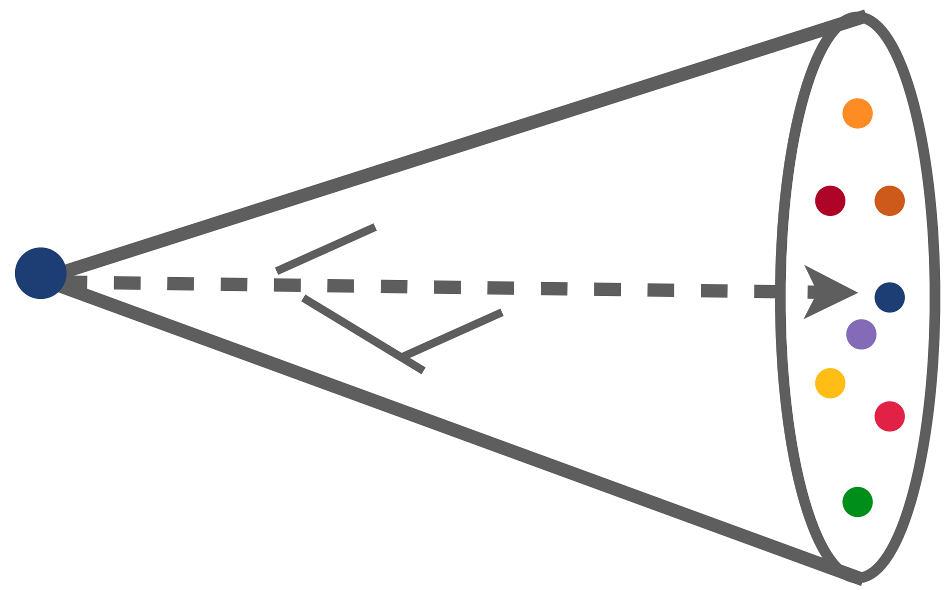




ME + PS generation
Detector Simulation

Inference

ML@Colliders : what's the broad task?

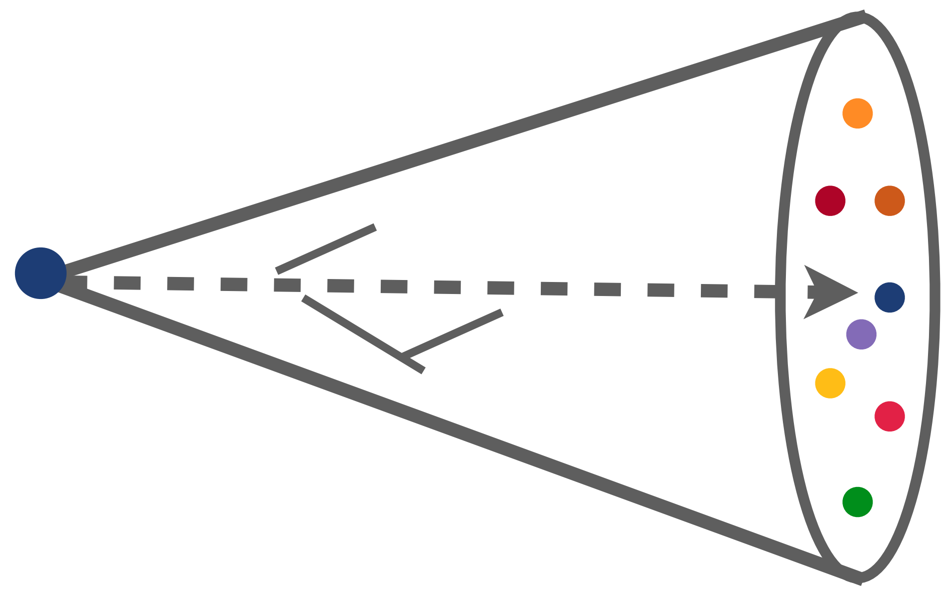
$$f_{\{\theta\}} \Bigg($$  $$\Bigg) \rightarrow \hat{y} = f_{\{\theta\}}(X)$$

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

ML@Colliders : what's the broad task?

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

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



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

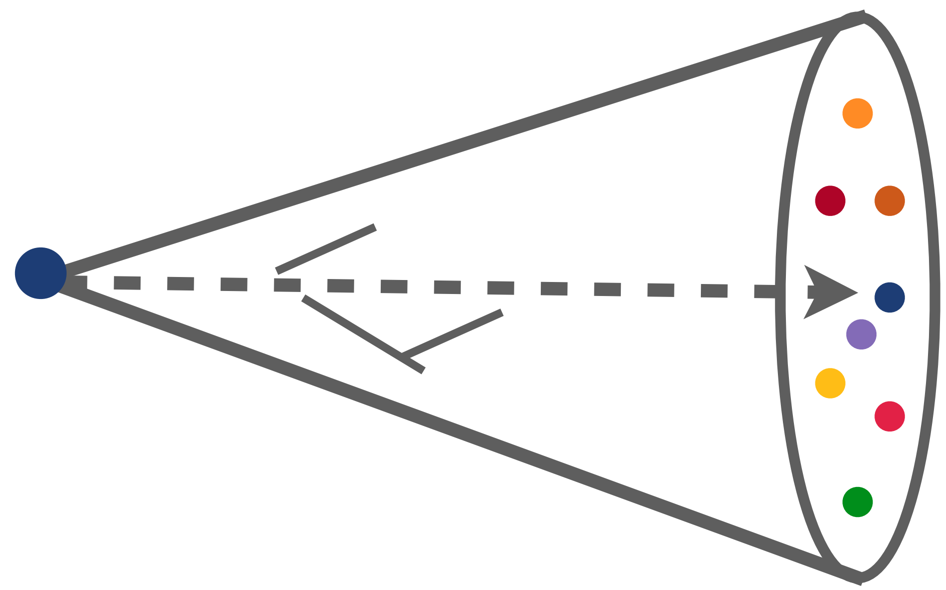
$$L(y, \hat{y}) \equiv L(\{\theta\})$$

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..)

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



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

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

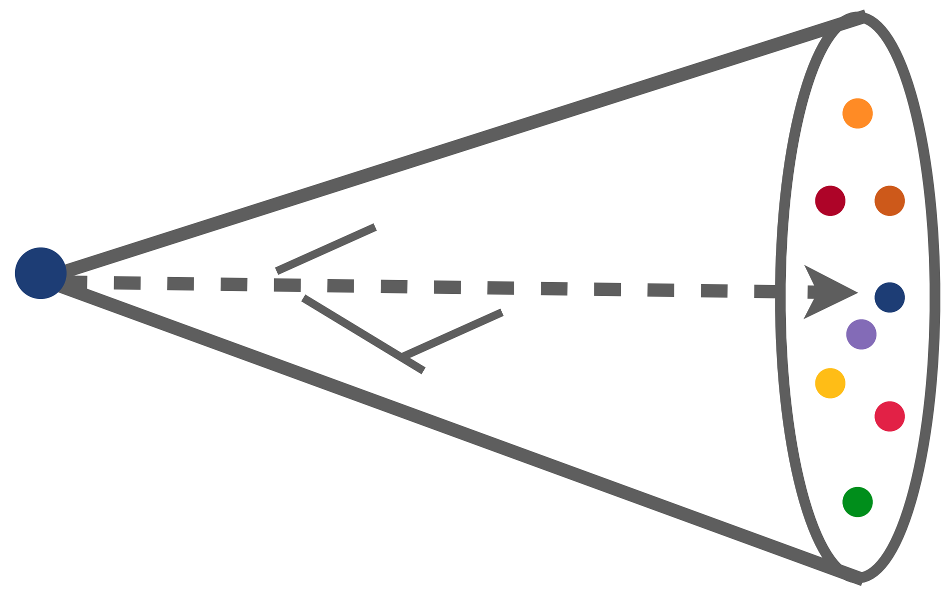
ML@Colliders : what's the broad task?

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

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

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

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



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

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

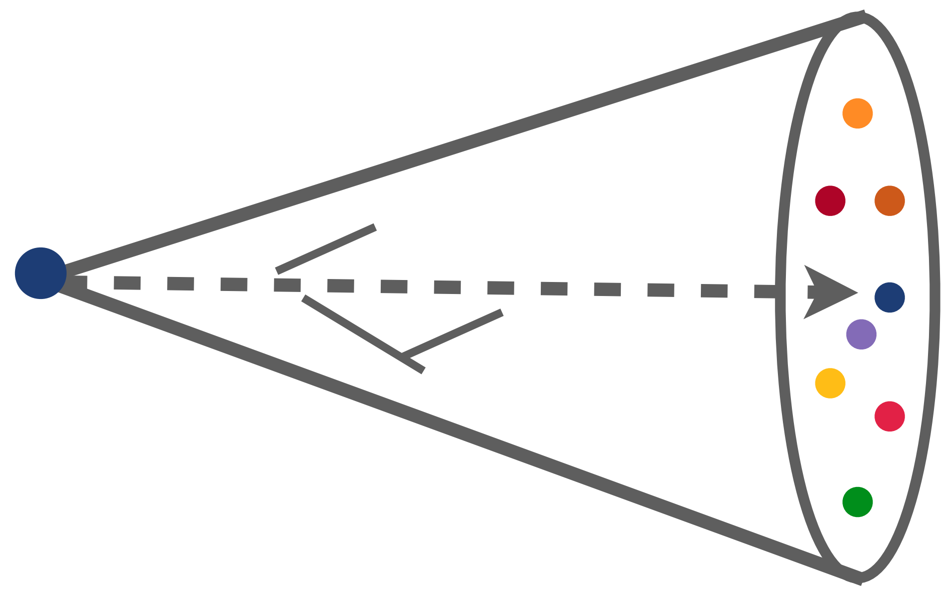
ML@Colliders : what's the broad task?

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.

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



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

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

Variation in data

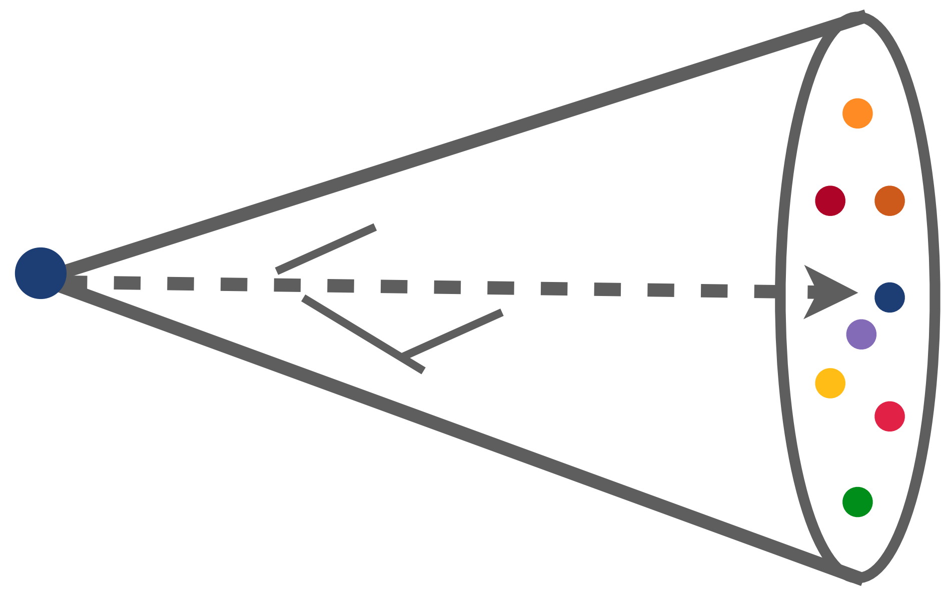
ML@Colliders : what's the broad task?

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.

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



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

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

Variation in data

Unsupervised

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

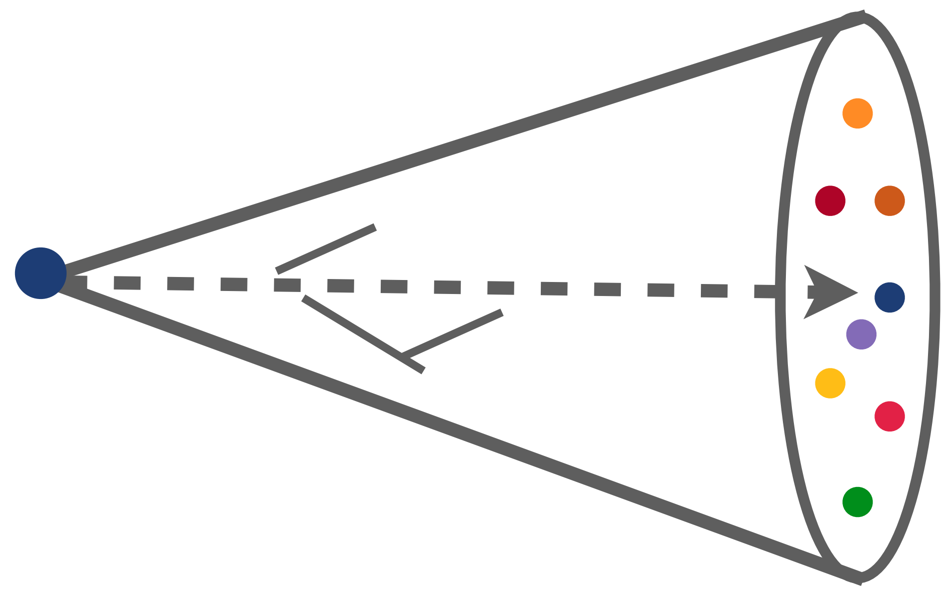
ML@Colliders : what's the broad task?

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.

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



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

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

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)

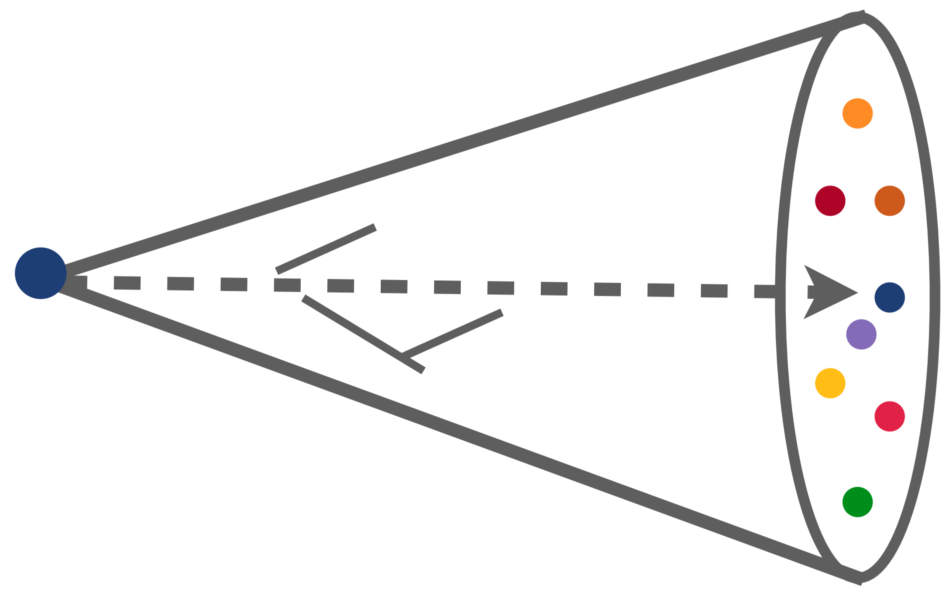
ML@Colliders : what's the broad task?

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.

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



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

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

Variation in data

Unsupervised	Semi-supervised	Weakly-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.

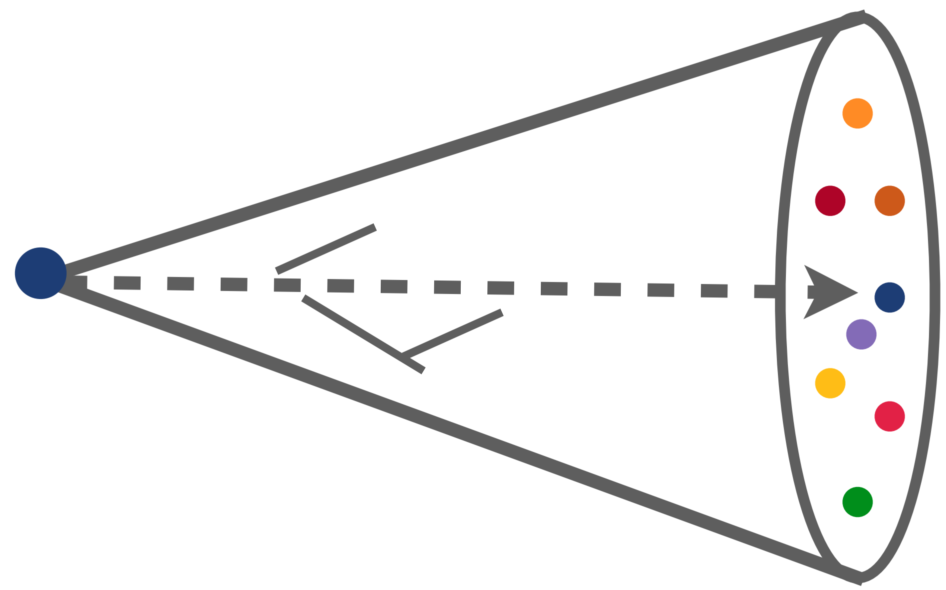
ML@Colliders : what's the broad task?

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.

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



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

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

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)

Weakly-supervised

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

Supervised

Learning on all the well labeled data.

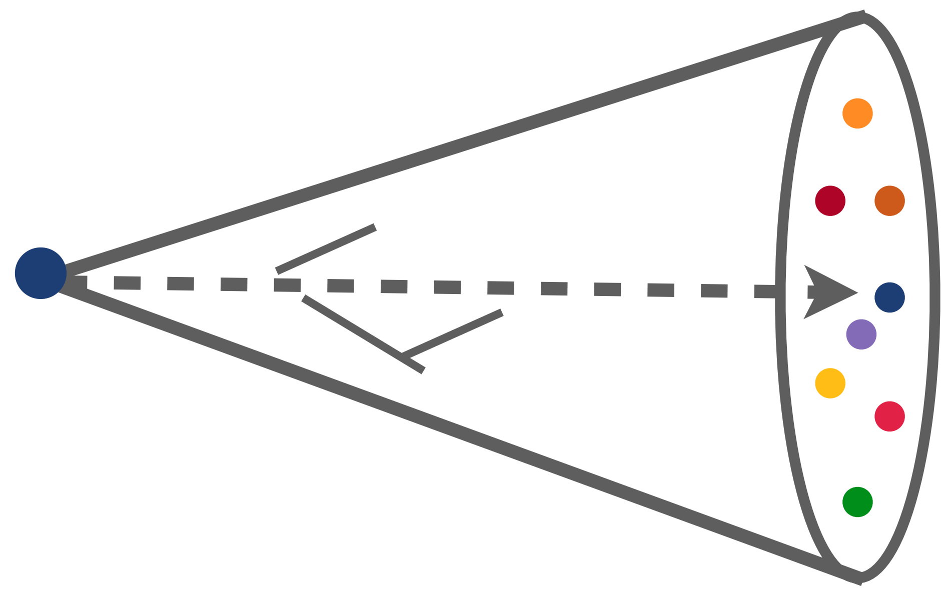
ML@Colliders : what's the broad task?

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.

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



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

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

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

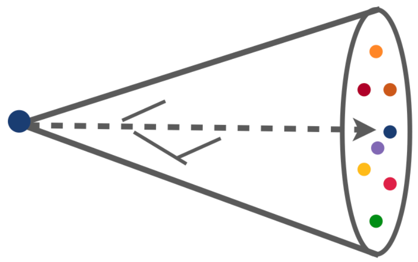
2. Choose a NN model (CNN/GNN/)

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

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

Uncertainty

Interpretability

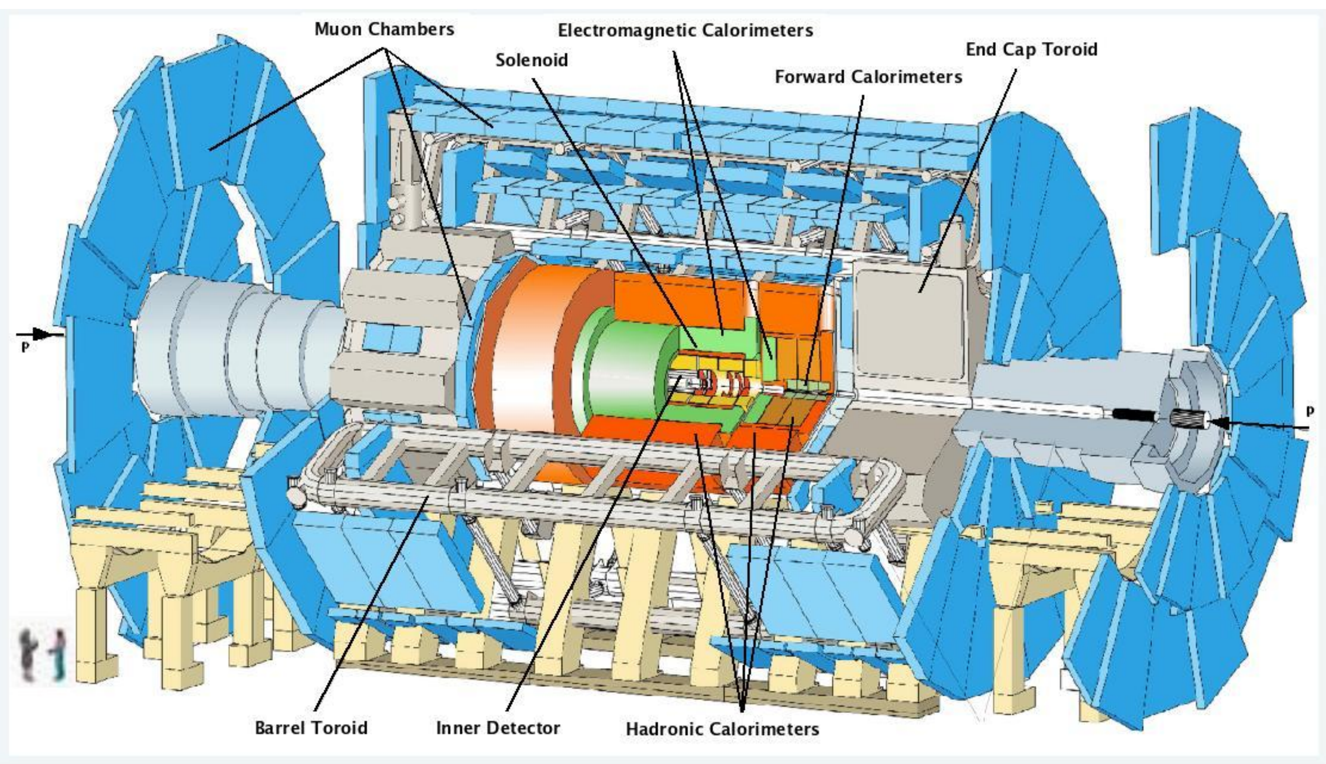
$f_{\{\theta\}}$ () $\rightarrow$ $\hat{y} = f_{\{\theta\}}(X)$

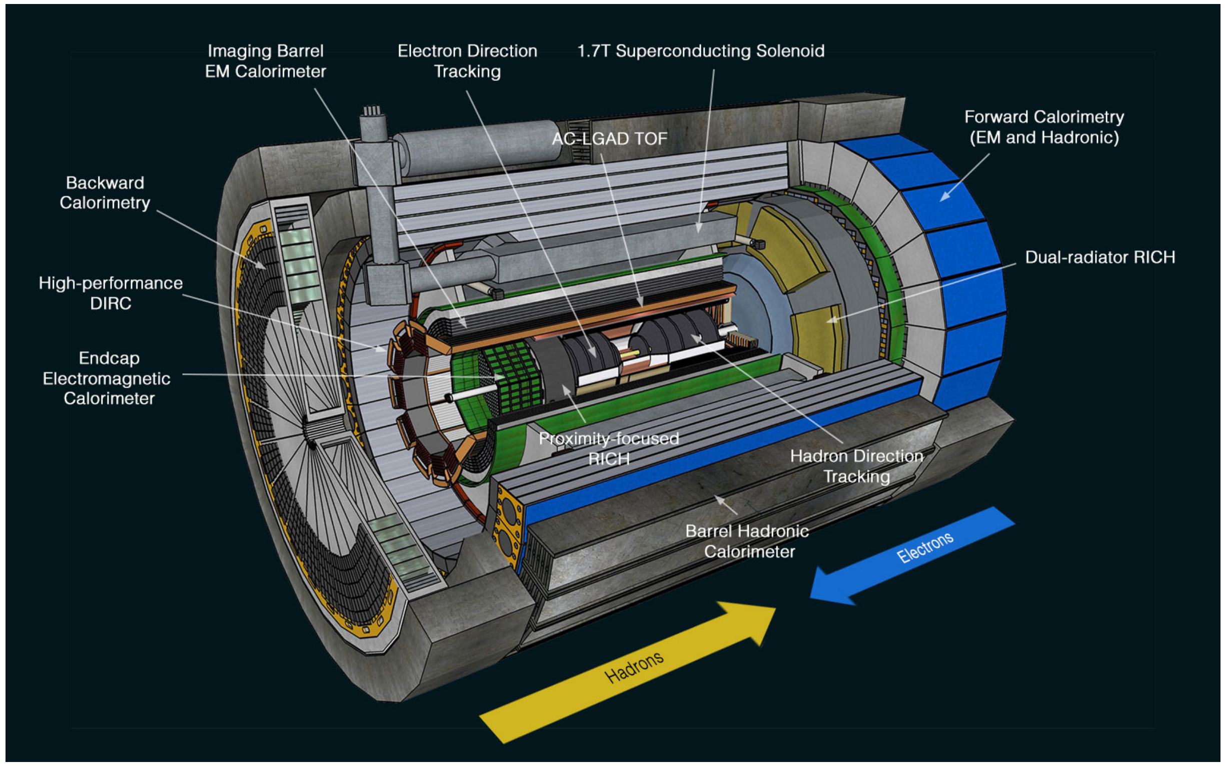
$$L(y, \hat{y}) \equiv L(\{\theta\})$$

Self-supervised

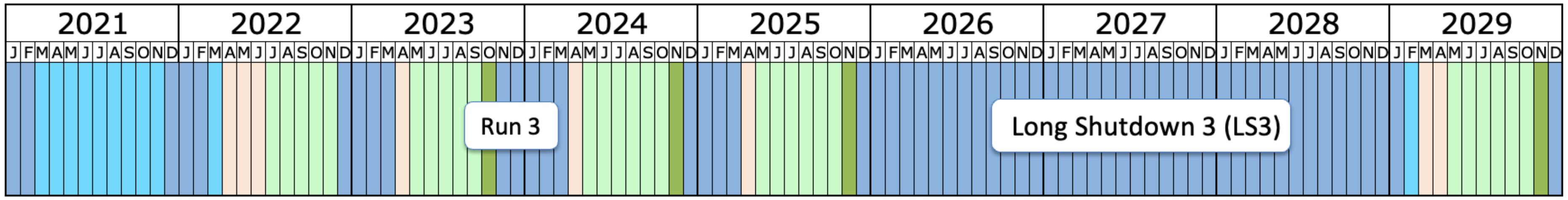
Variation in data

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.

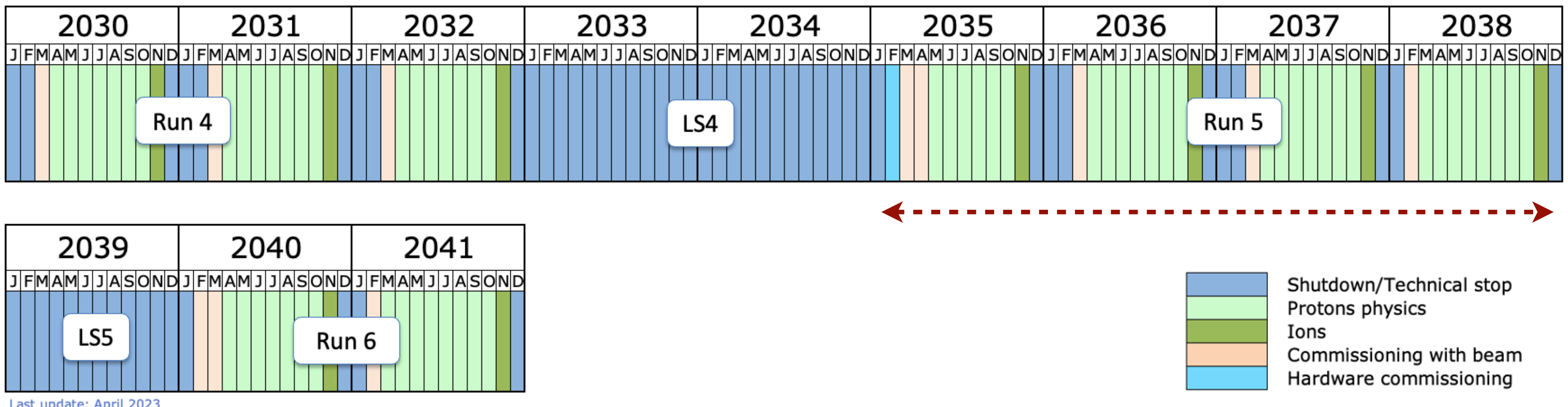


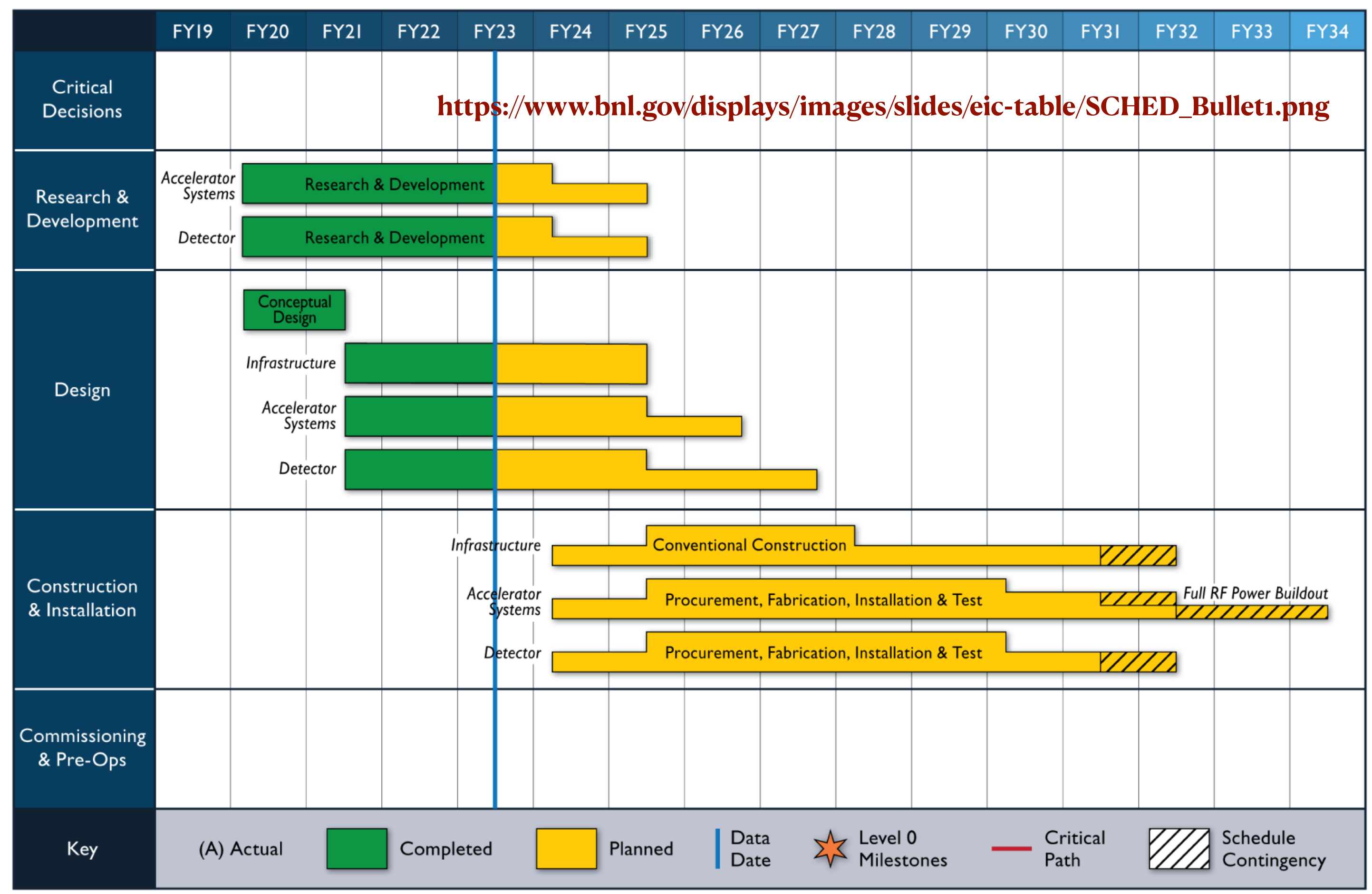


We have time for a coffee together

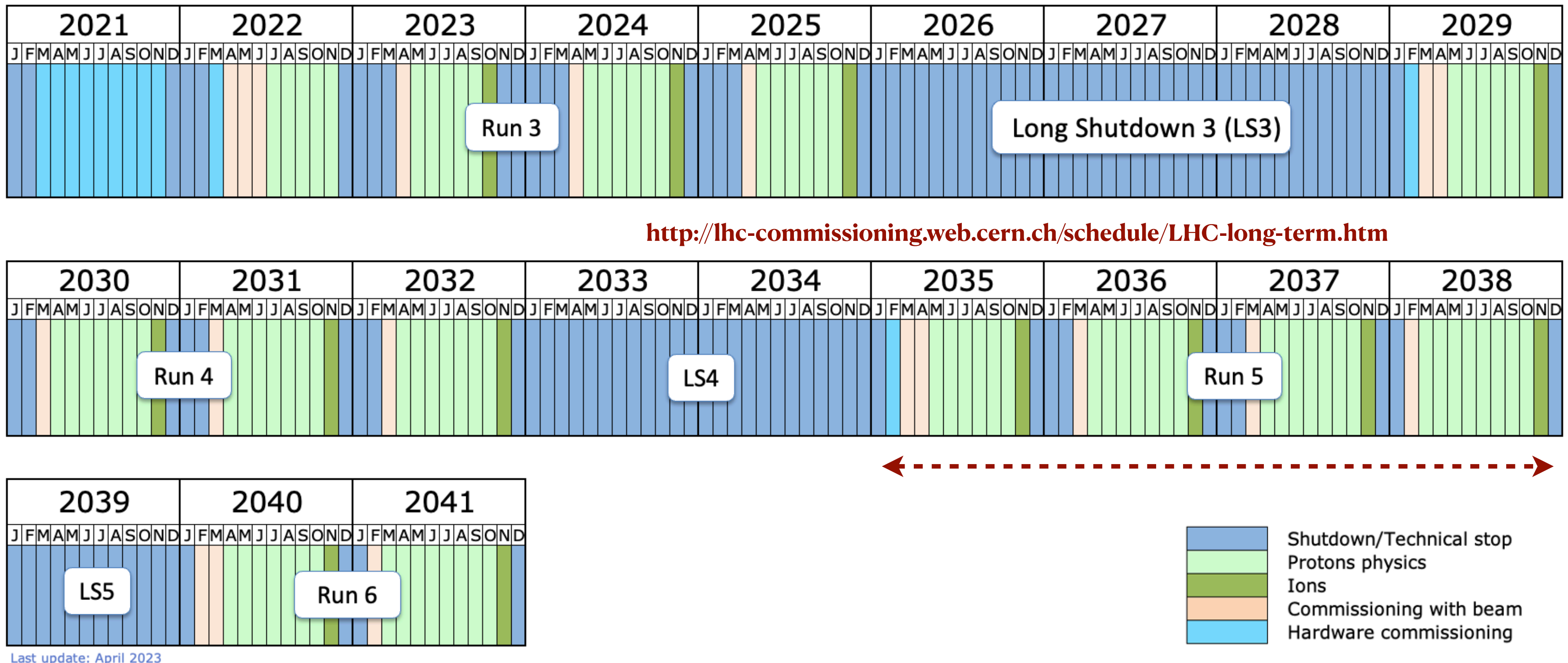


http://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm





We have time for a coffee together

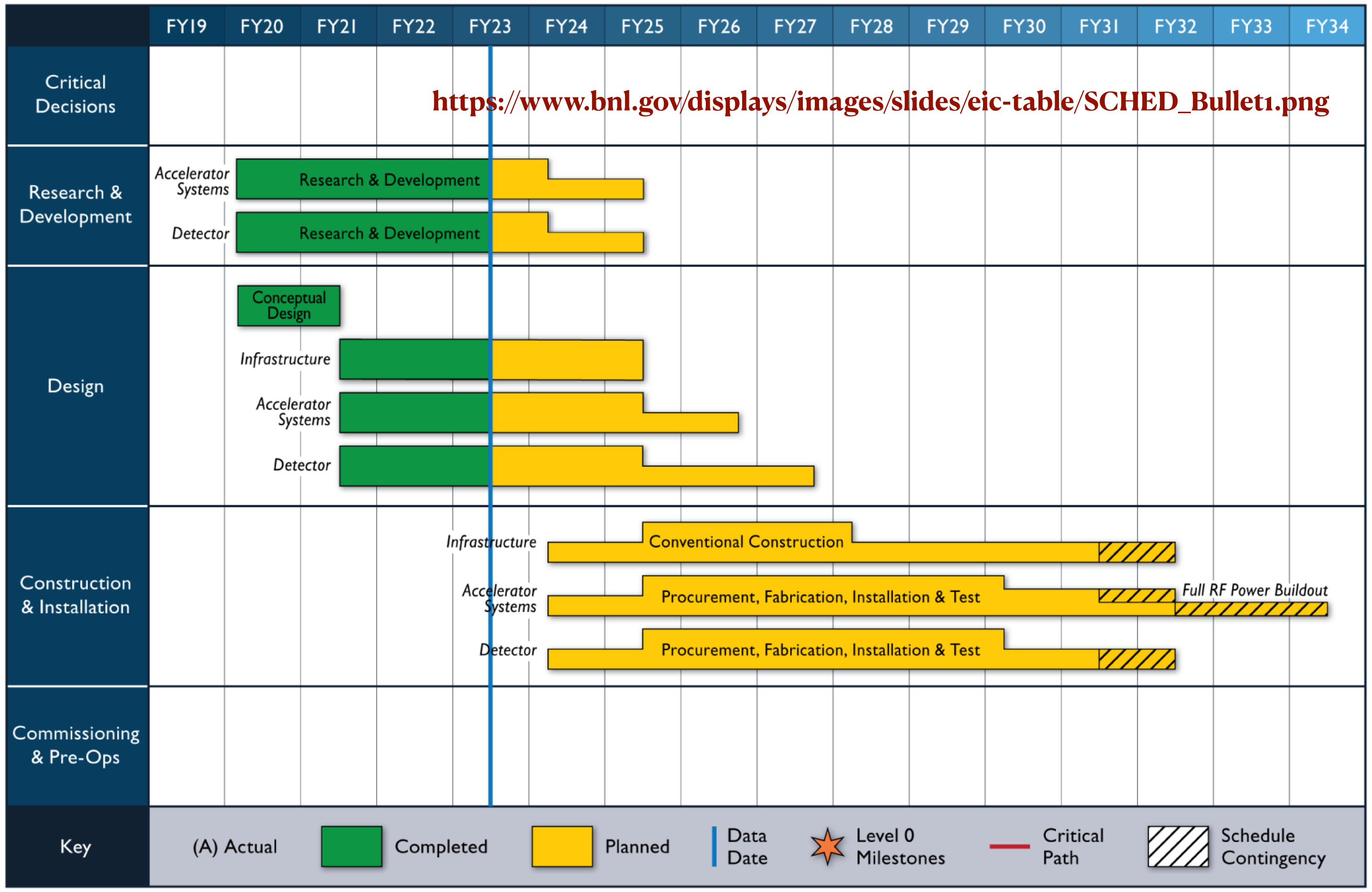


http://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm

There will be a decade of overlap between LHC and EIC.

The probe of strong interaction program will have overlapping physics goals.

Through the ML lens the formulation of the problems will be closer.

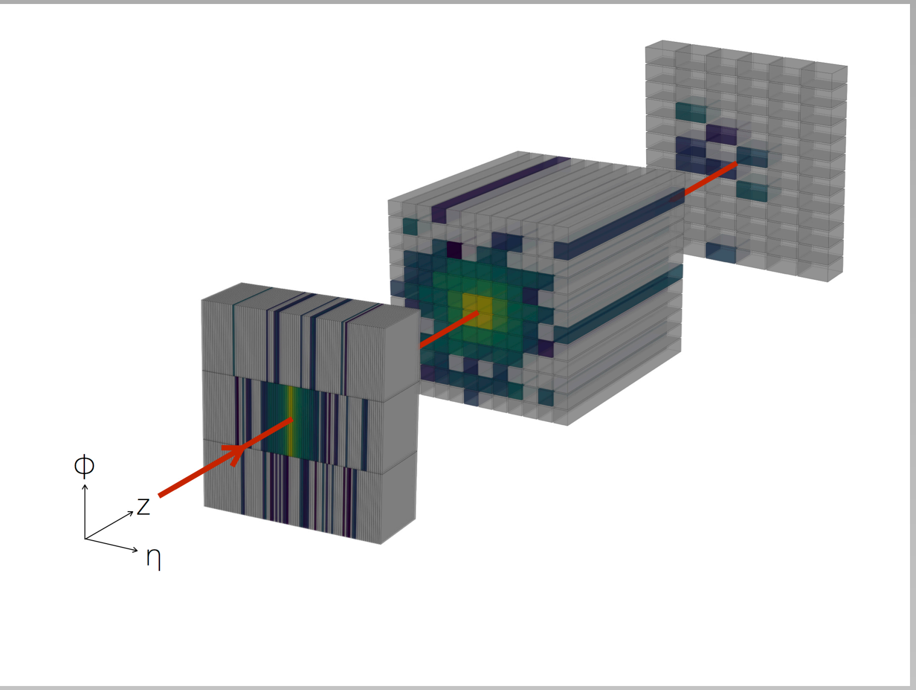


https://www.bnl.gov/displays/images/slides/eic-table/SCHED_Bullet1.png

Calorimetery

Image from 1705.02355

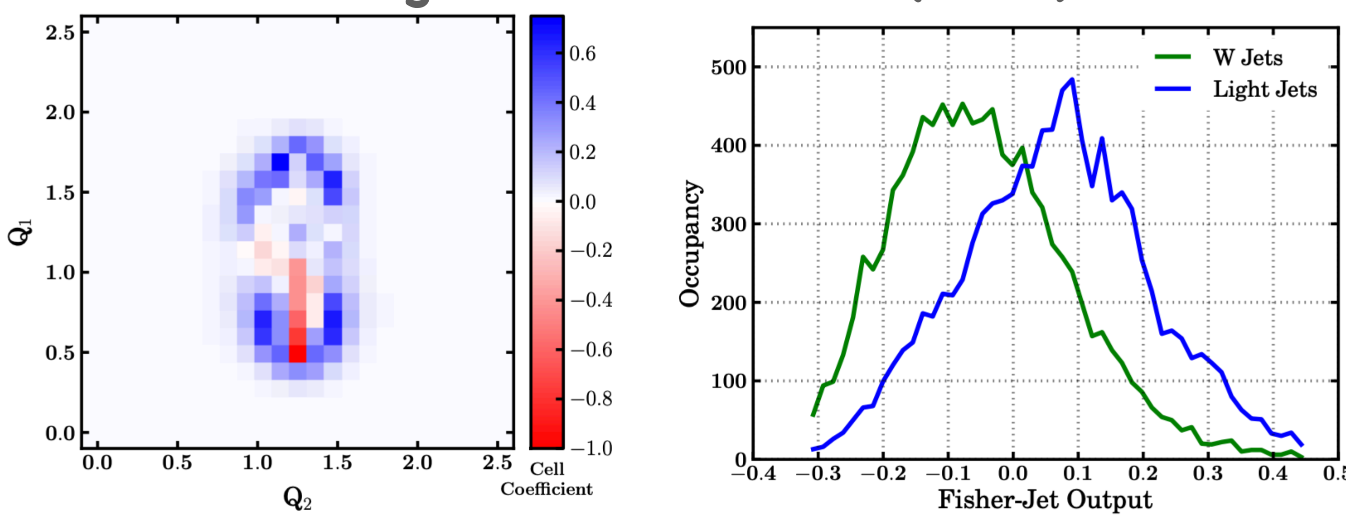
$f_{\{\theta\}}\Big($



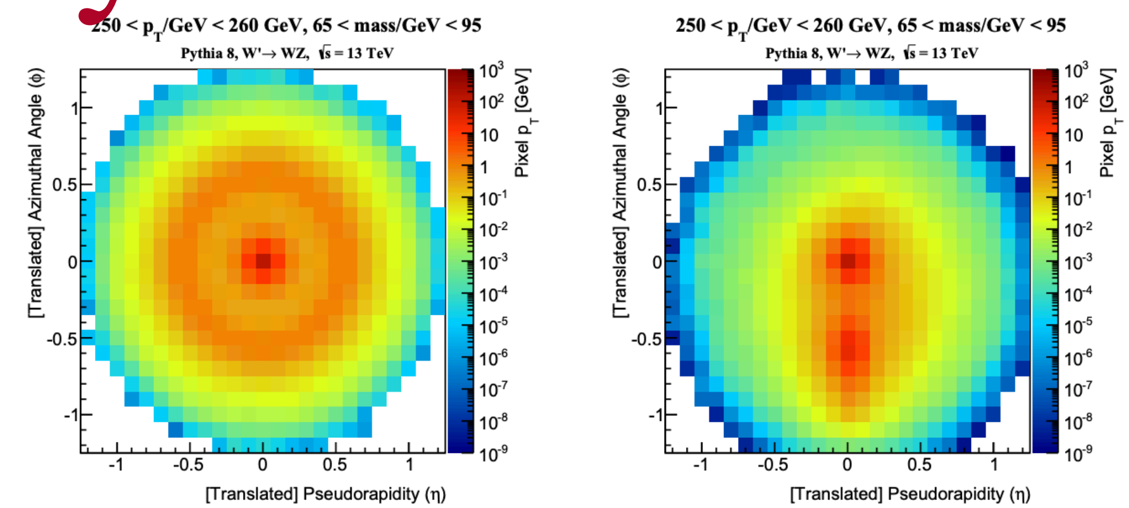
$\Big)$

Calorimetry + ML early works

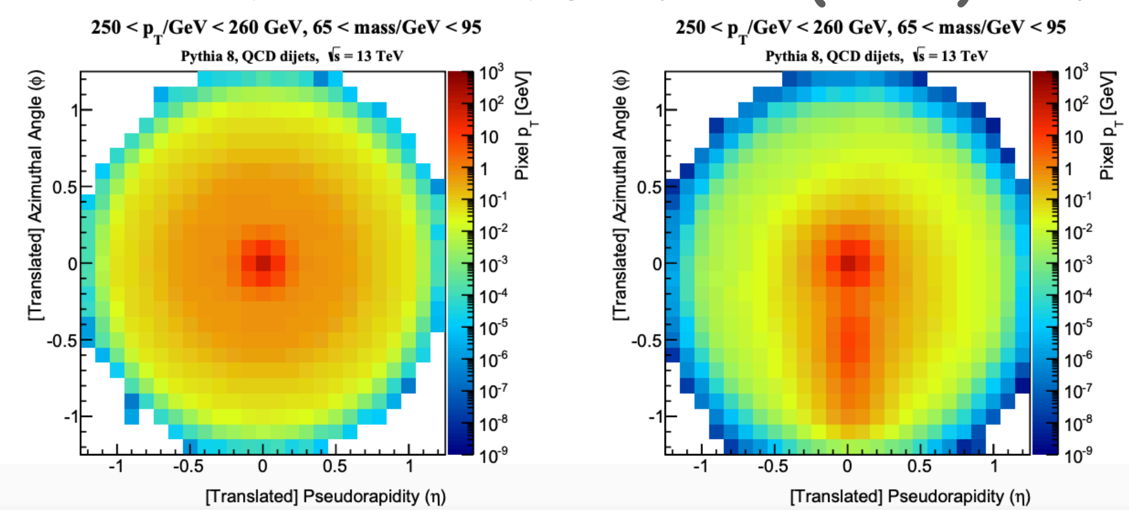
J. Cogan et-al JHEP 02 (2015) 118

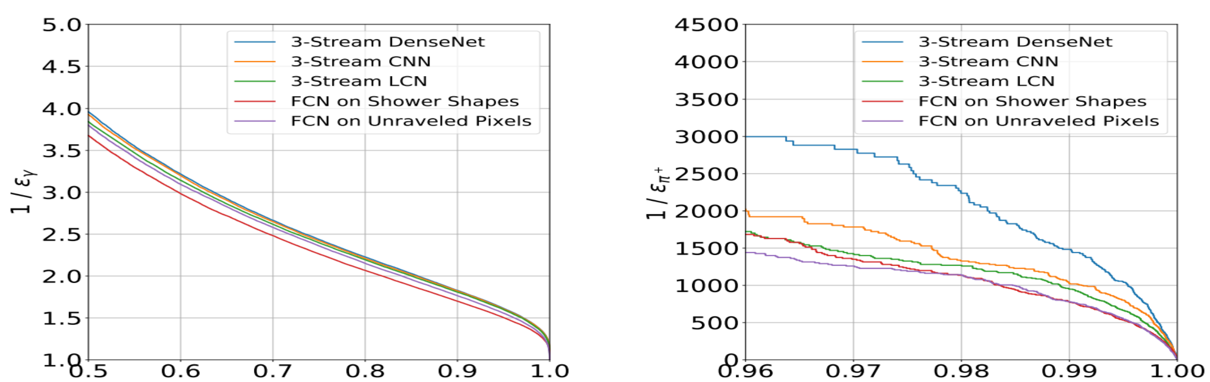


(a) Fisher-Jet
(b) Fisher-Jet Discriminant Output



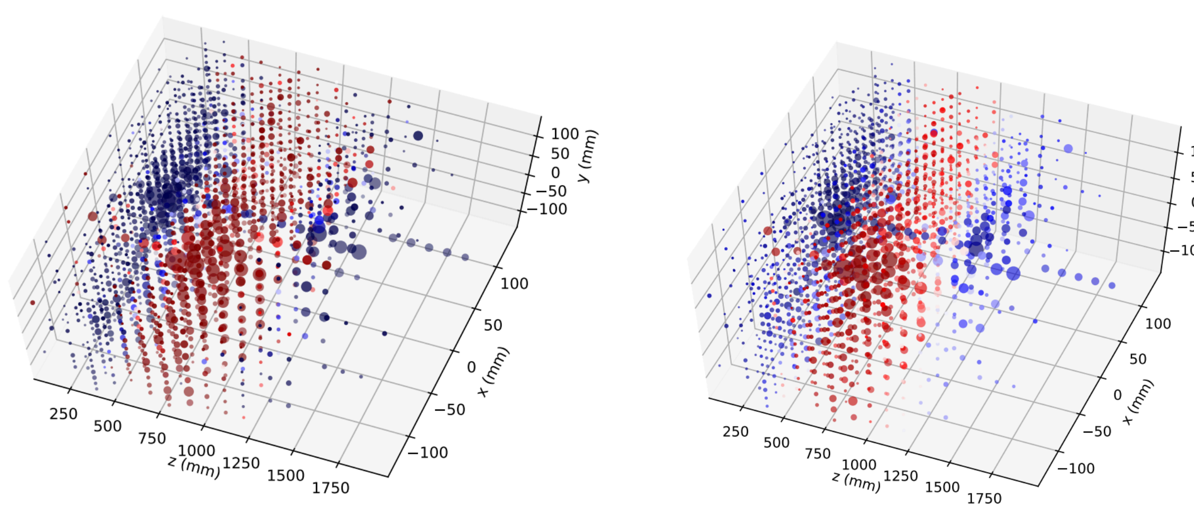
L. De Oliveira et-al JHEP 07 (2016) 069





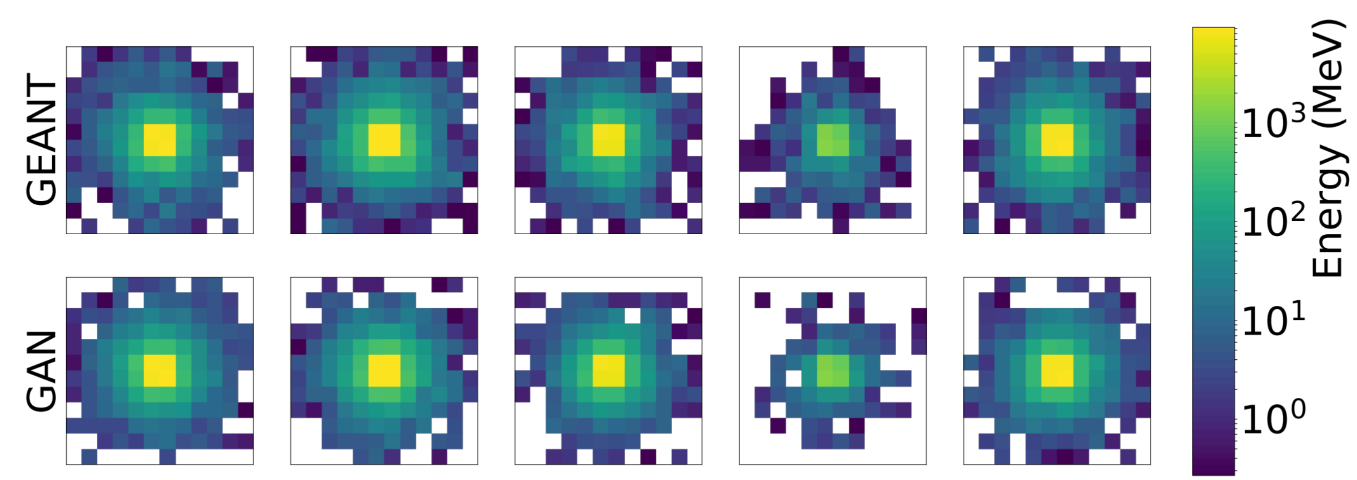
L. De Oliveira et-al NIM-A 951v162879

(a) ROC curves for e^+-γ classification
(b) ROC curves for e^+-π^+ classification



(a) Truth
(b) Reconstructed

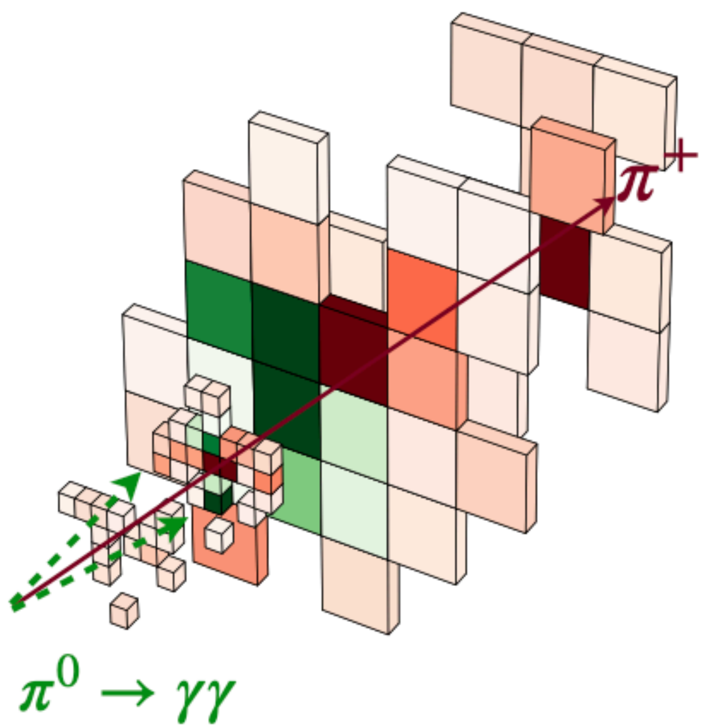
S. Qasim et-al EPJC 79 7 (2019) 608

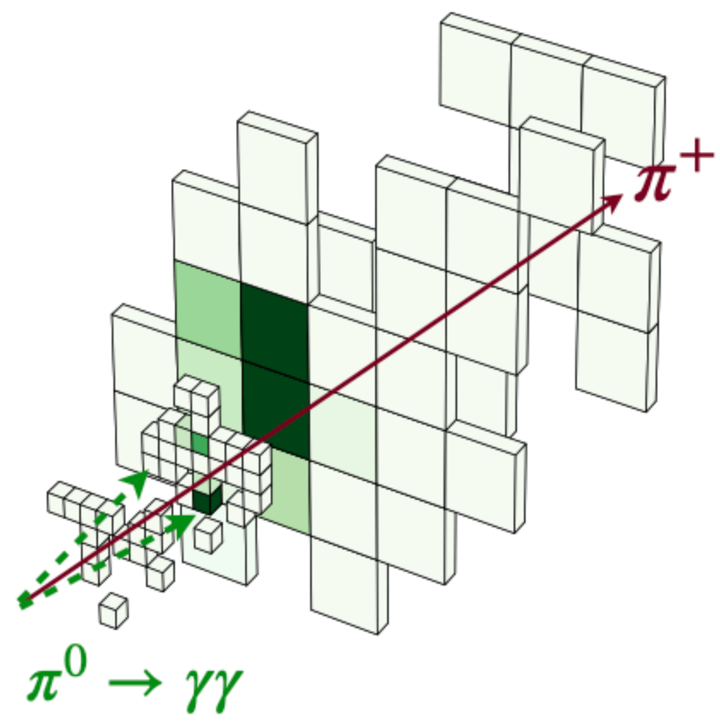


M. Paganini et-al PRD 97 (2018) 1, 014021

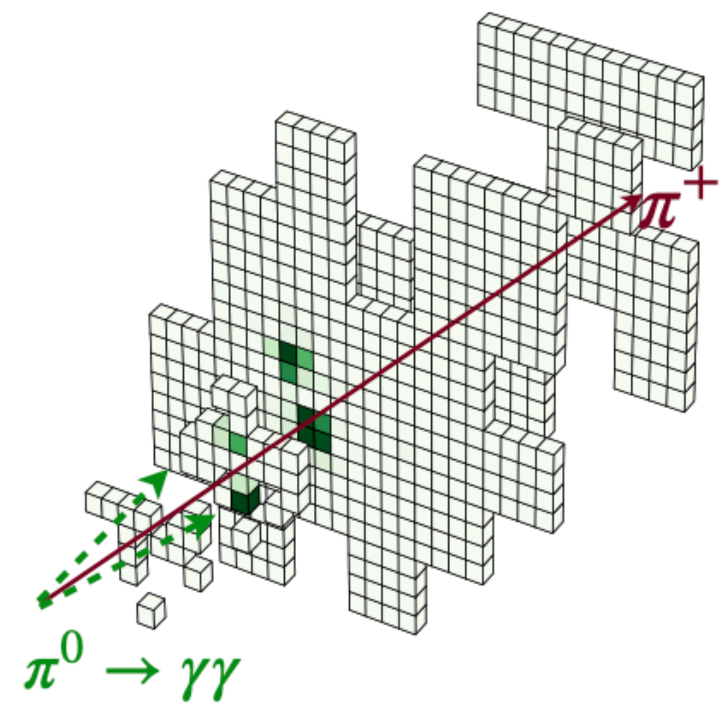
A 3-D view for topoclusters only

8 X 8 Low Res detector

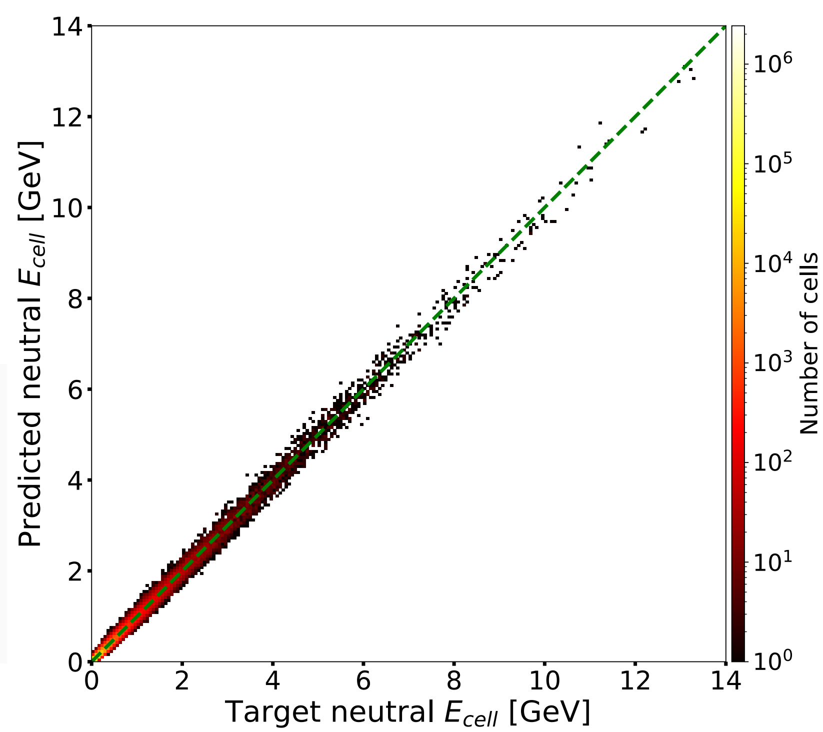


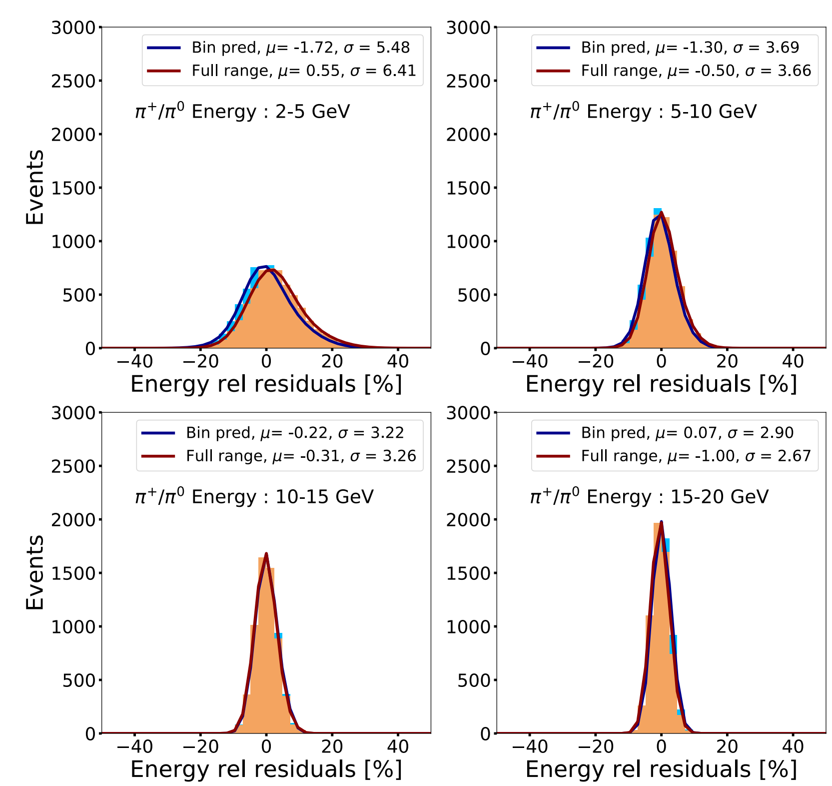


32 X 32 High Res detector



The networks in general have good noise removal abilities.





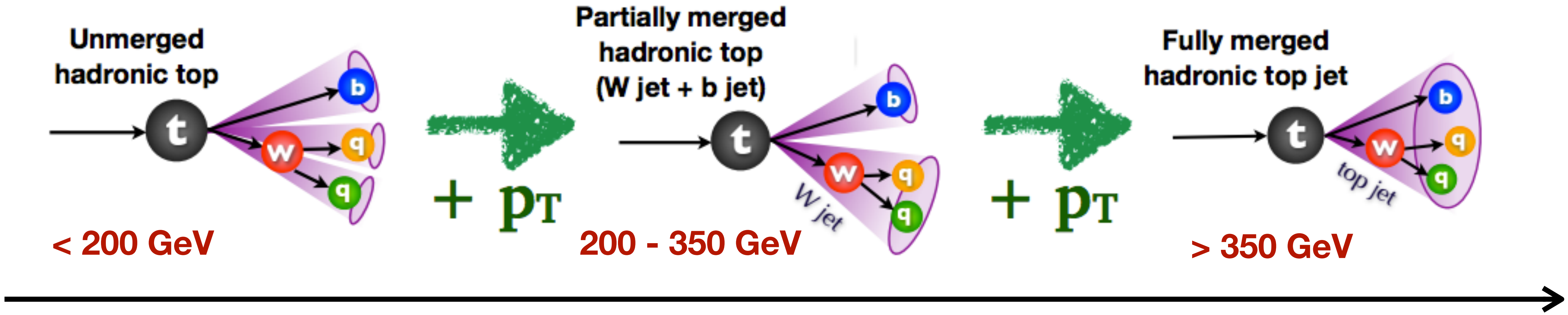
Towards a computer vision particle flow

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

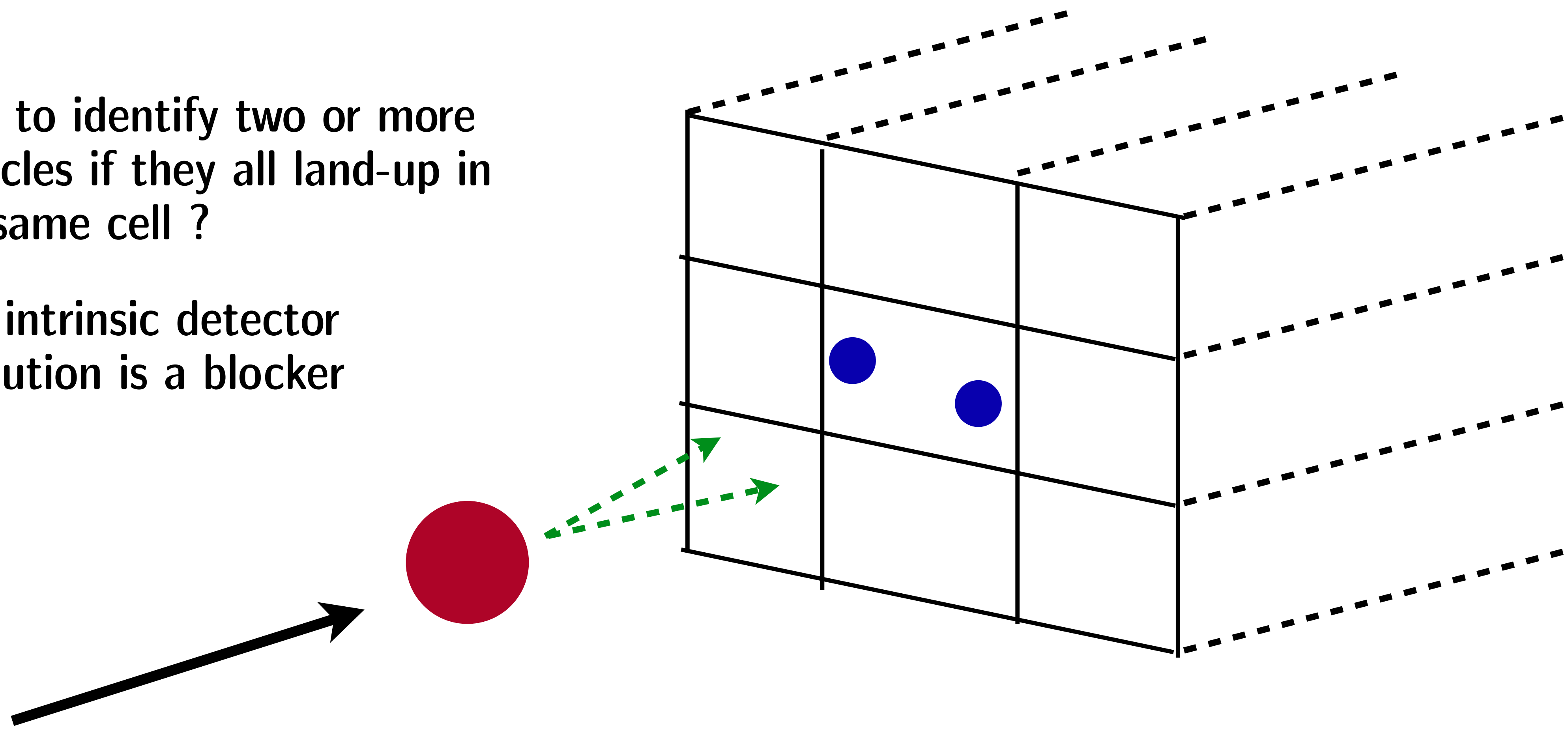
1341 Accesses | **13** Citations | **11** Altmetric | Metrics

When do intrinsic calorimeter sizes are limiting factors ?

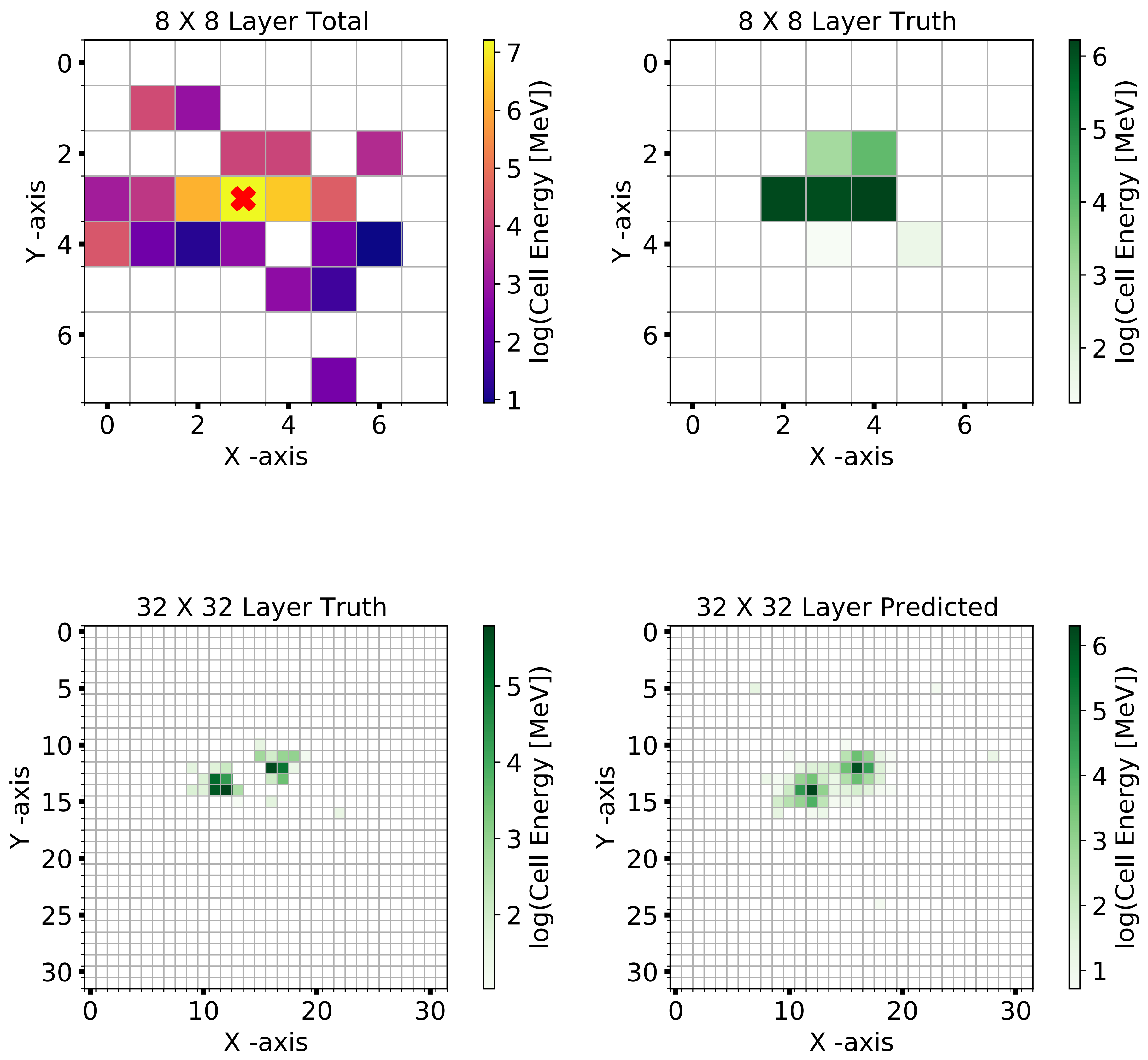


How to identify two or more particles if they all land-up in the same cell ?

The intrinsic detector resolution is a blocker

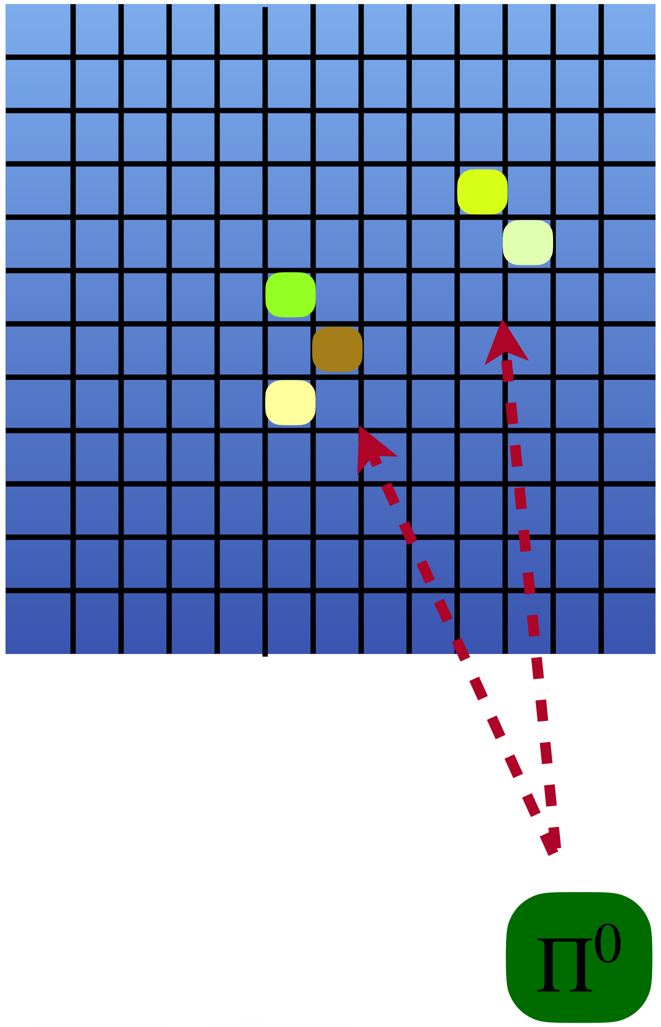


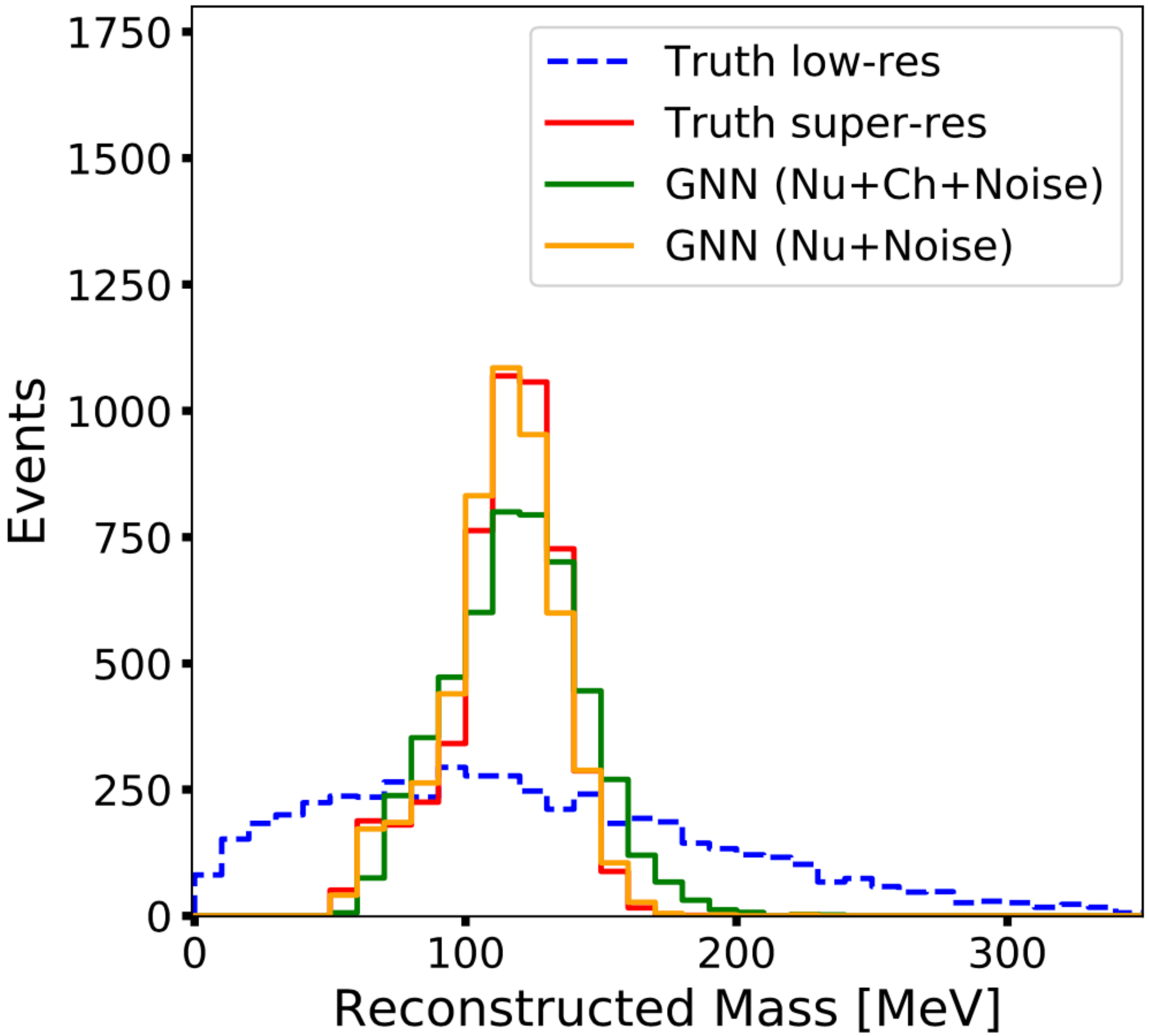
An event display for super-res prediction



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.)

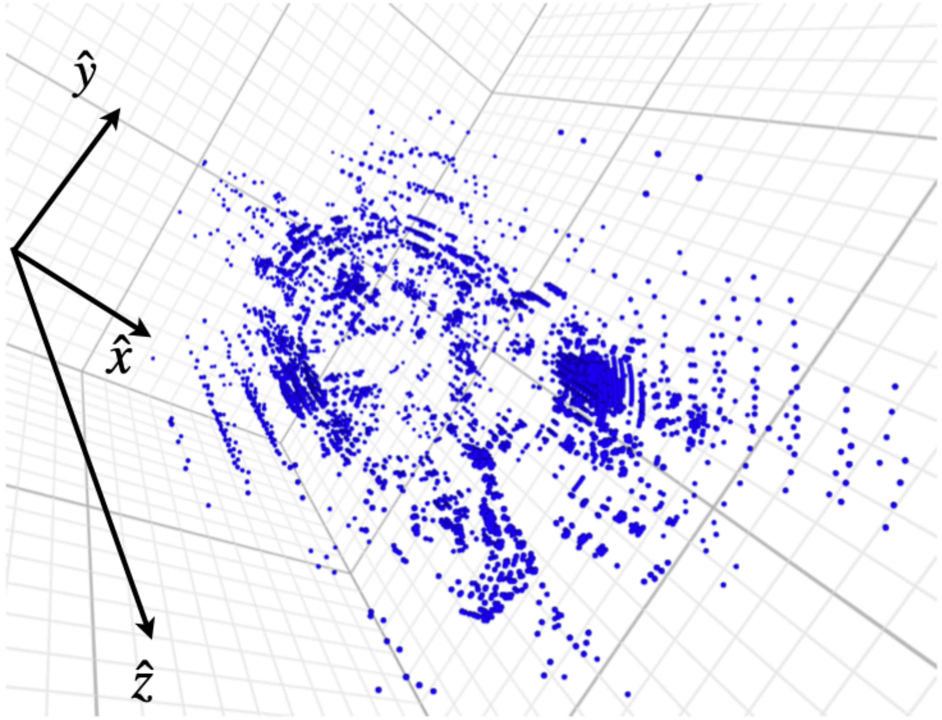
How to GAN Higher Jet Resolution
SciPost Phys. 13 (2022) 3, 064
Pierre Baldi[1], Lukas Blecher[2], Anja Butter[2], Julian Collado[1], Jessica N. Howard[3], Fabian Keilbach[2], Tilman Plehn[2], Gregor Kasieczka[4], and Daniel Whiteson[3]

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)

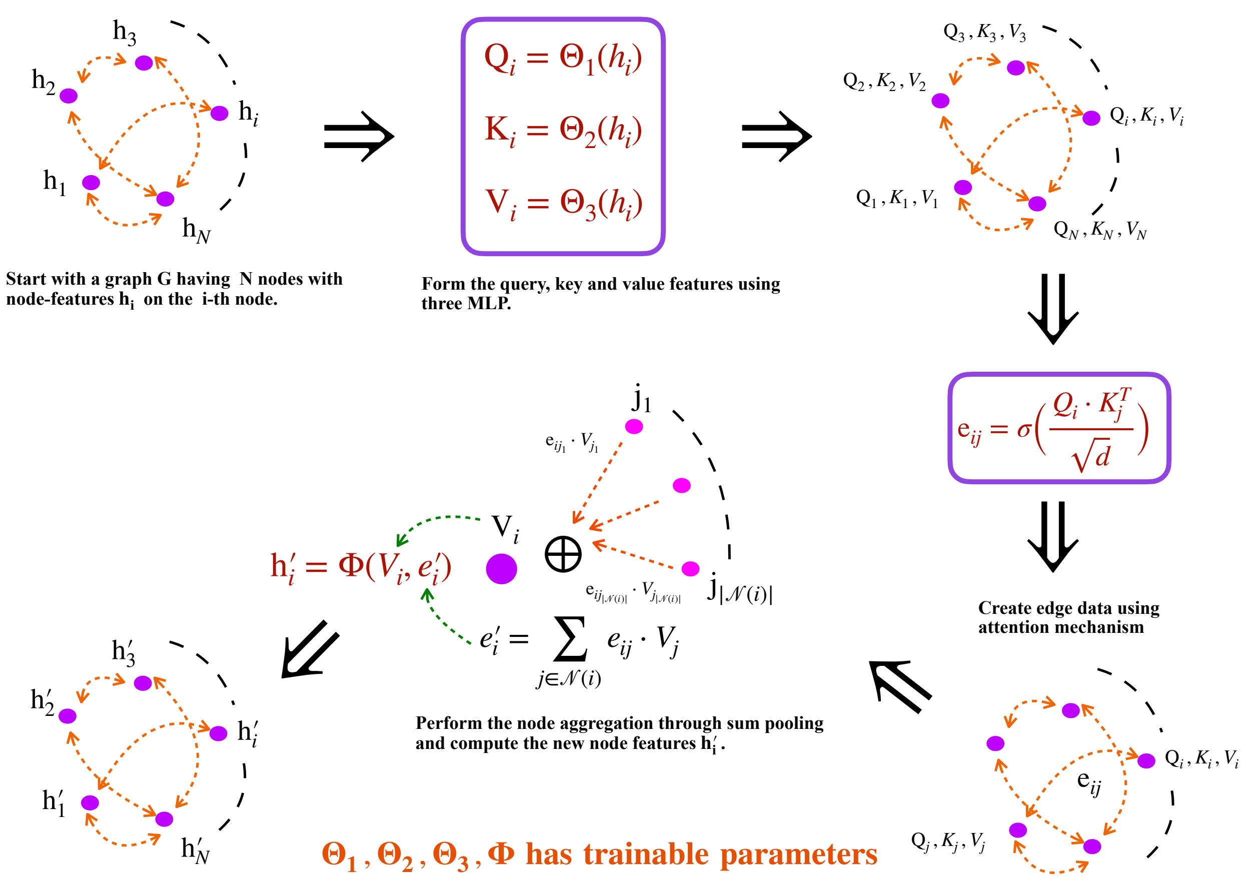
Pion identification within ATLAS

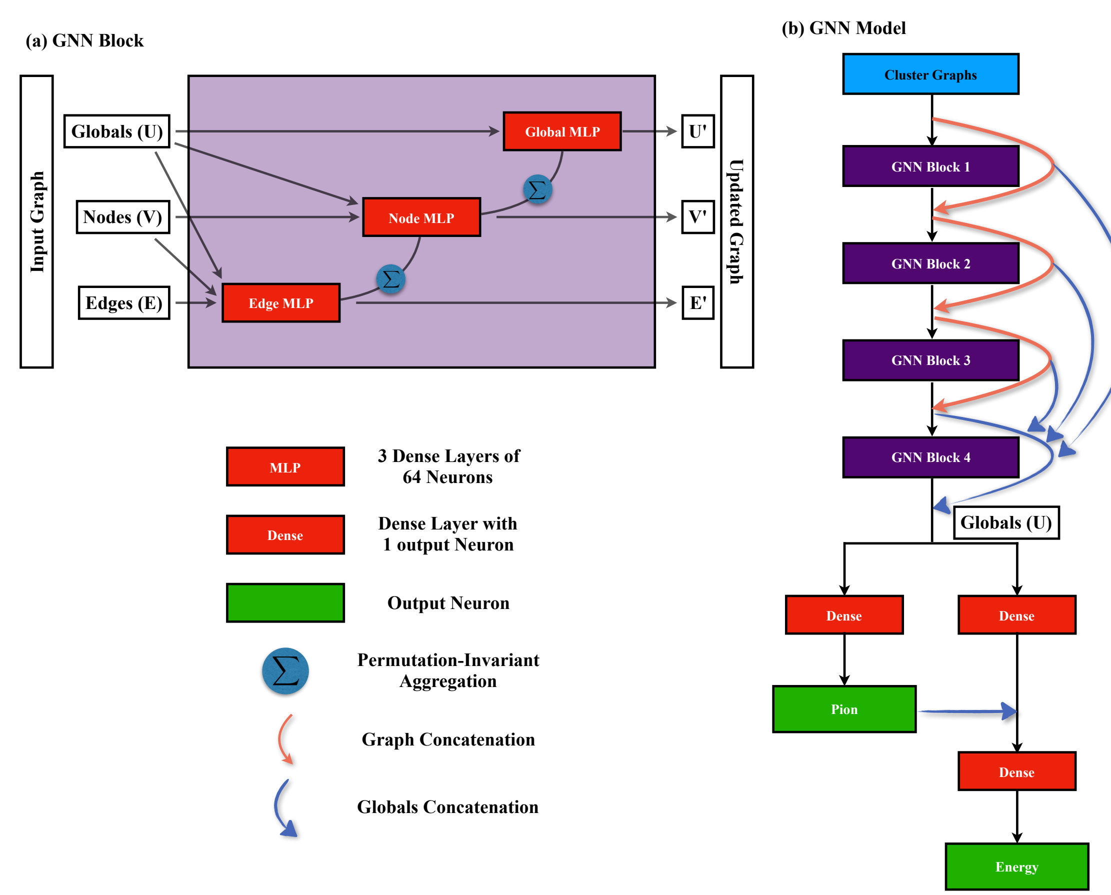


A classification & regression task is tested on ATLAS samples.
The calibrated topocluster cells are used to form images & P.C.

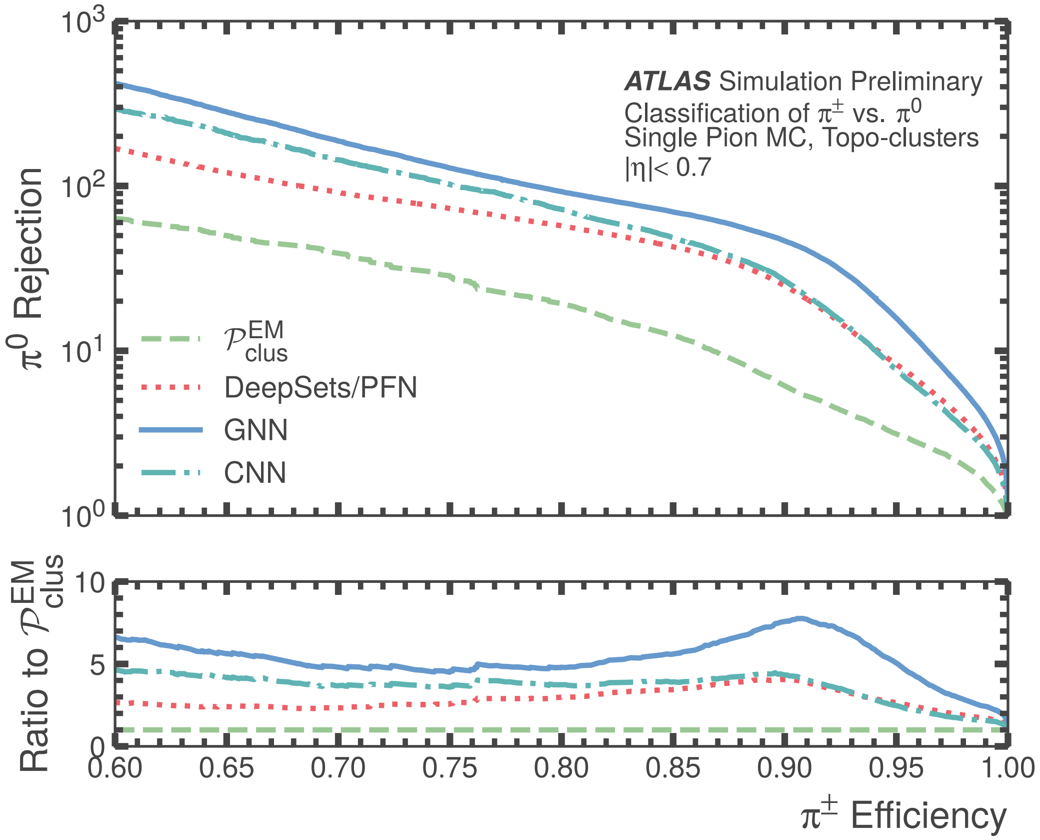
ATL-PHYS-PUB-2020-040

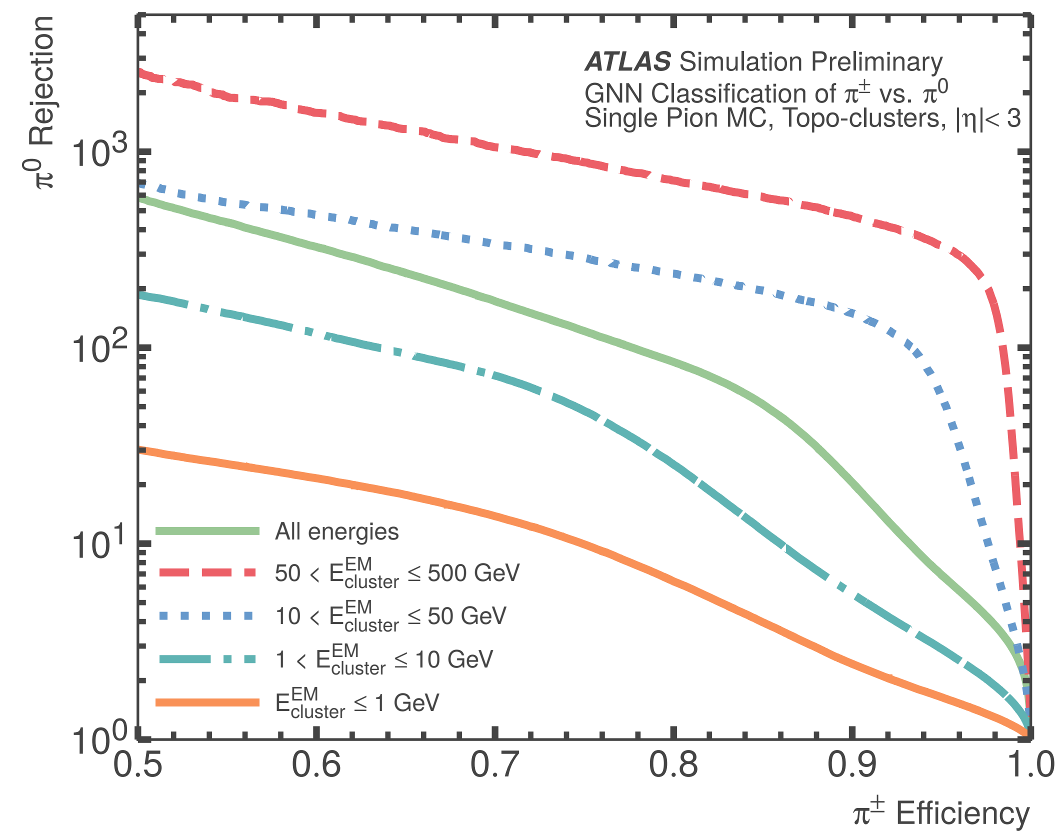
$$\mathcal{L} = (1-\alpha)\mathcal{L}_{\text{classification}} + \alpha\mathcal{L}_{\text{Regression}}$$

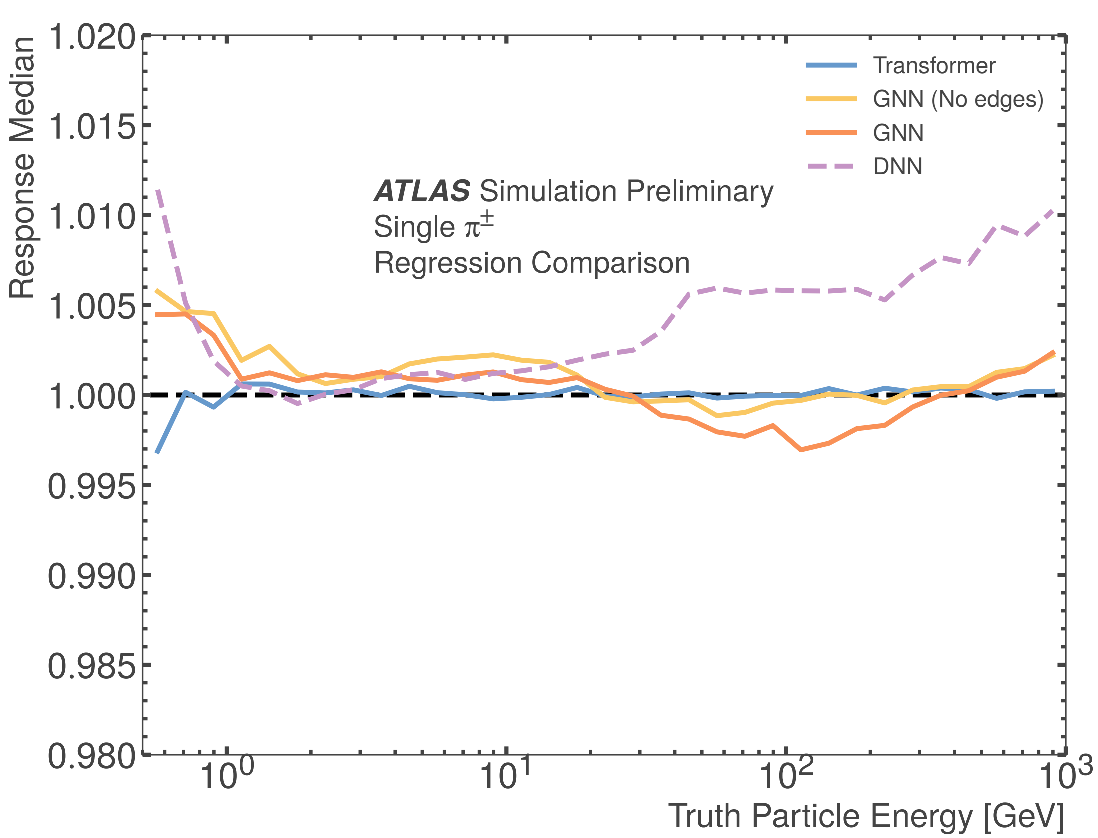


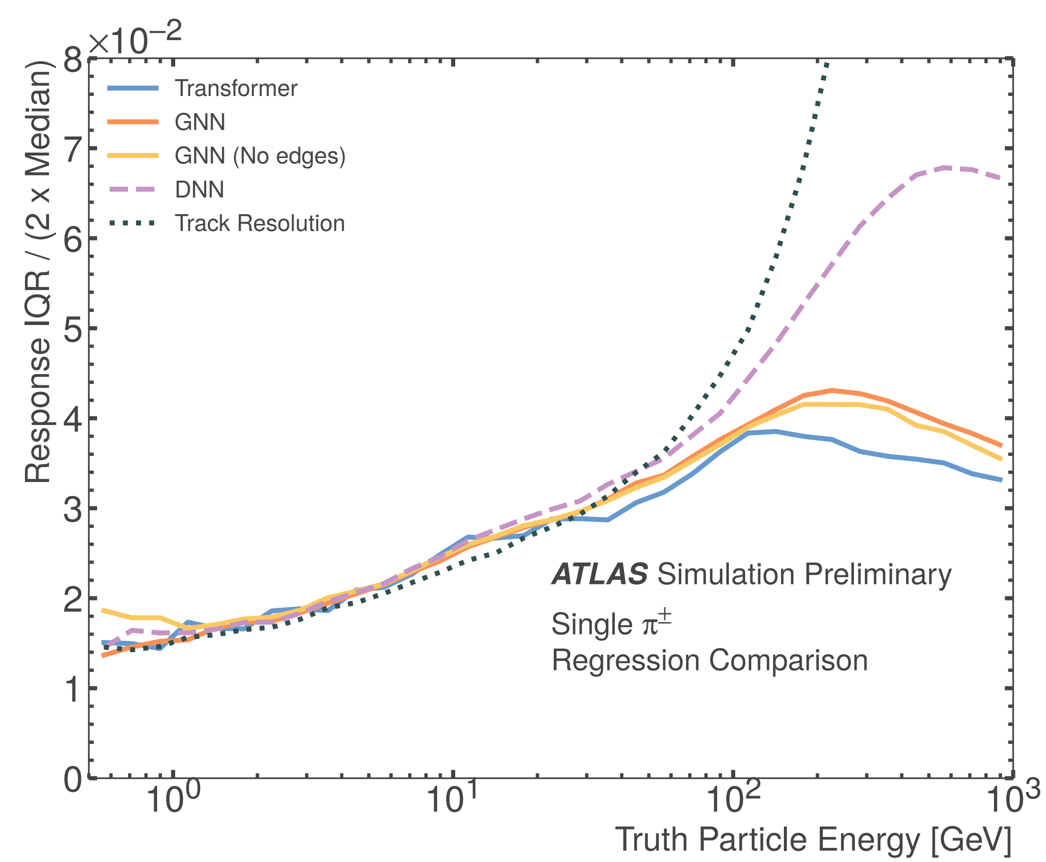


Pion identification within ATLAS

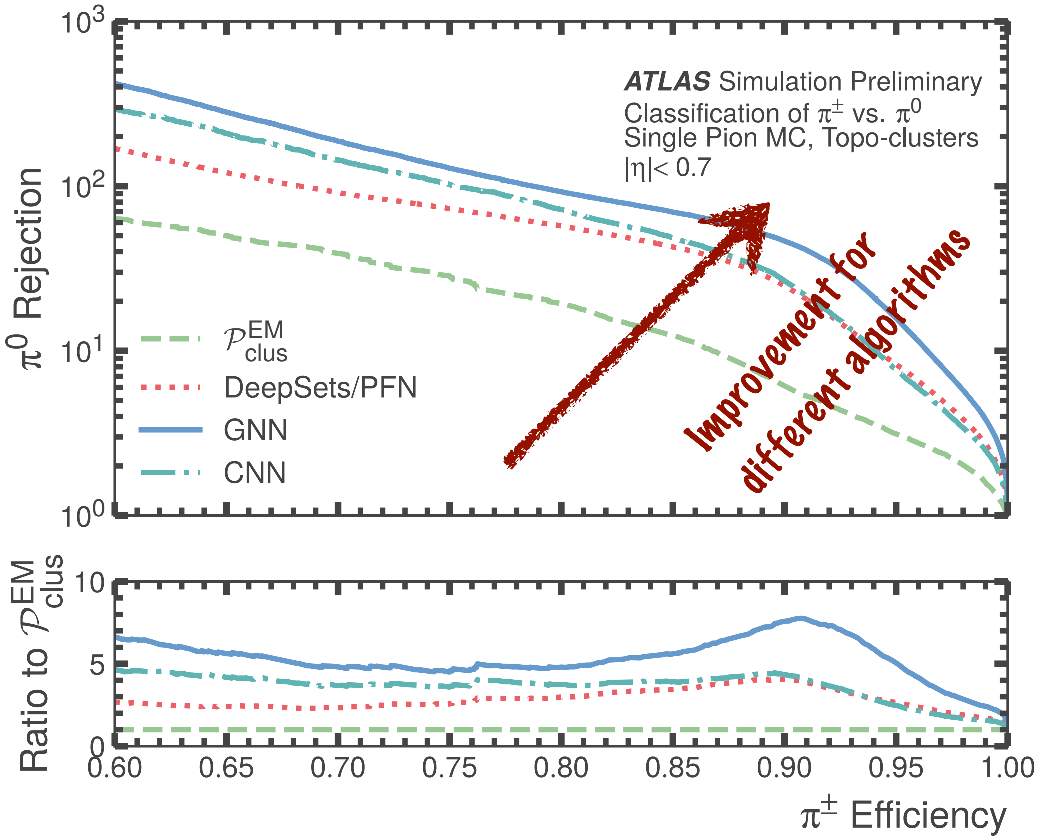


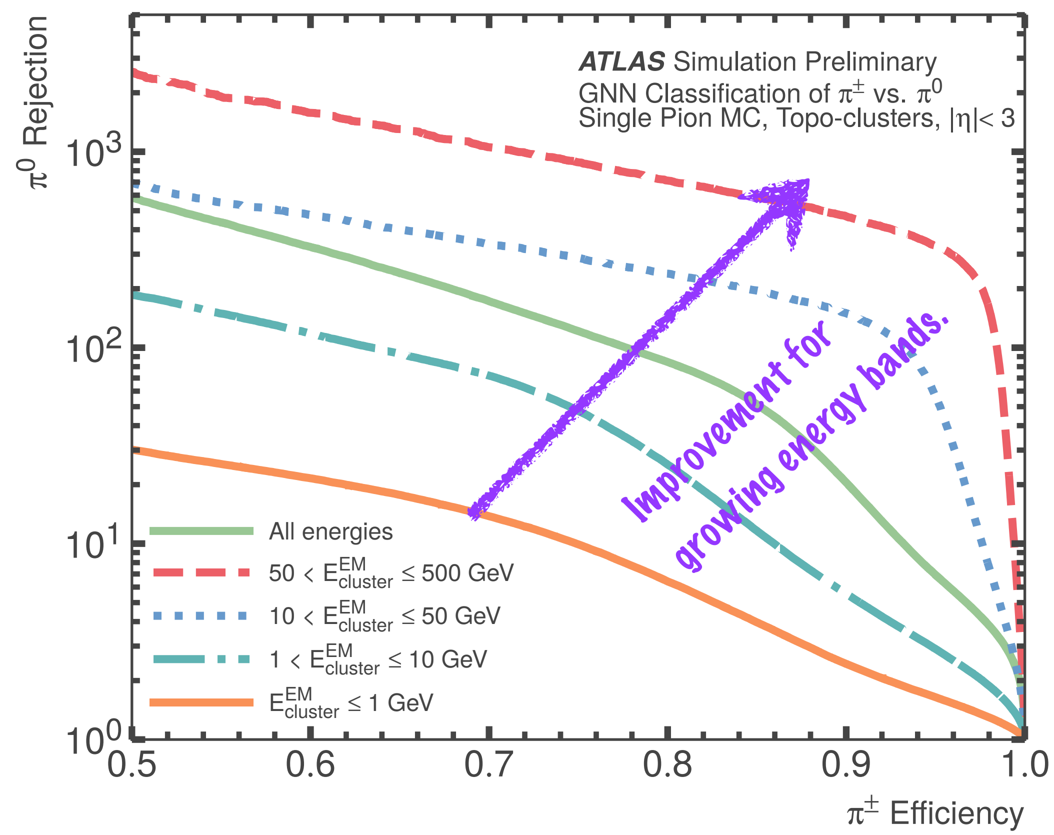


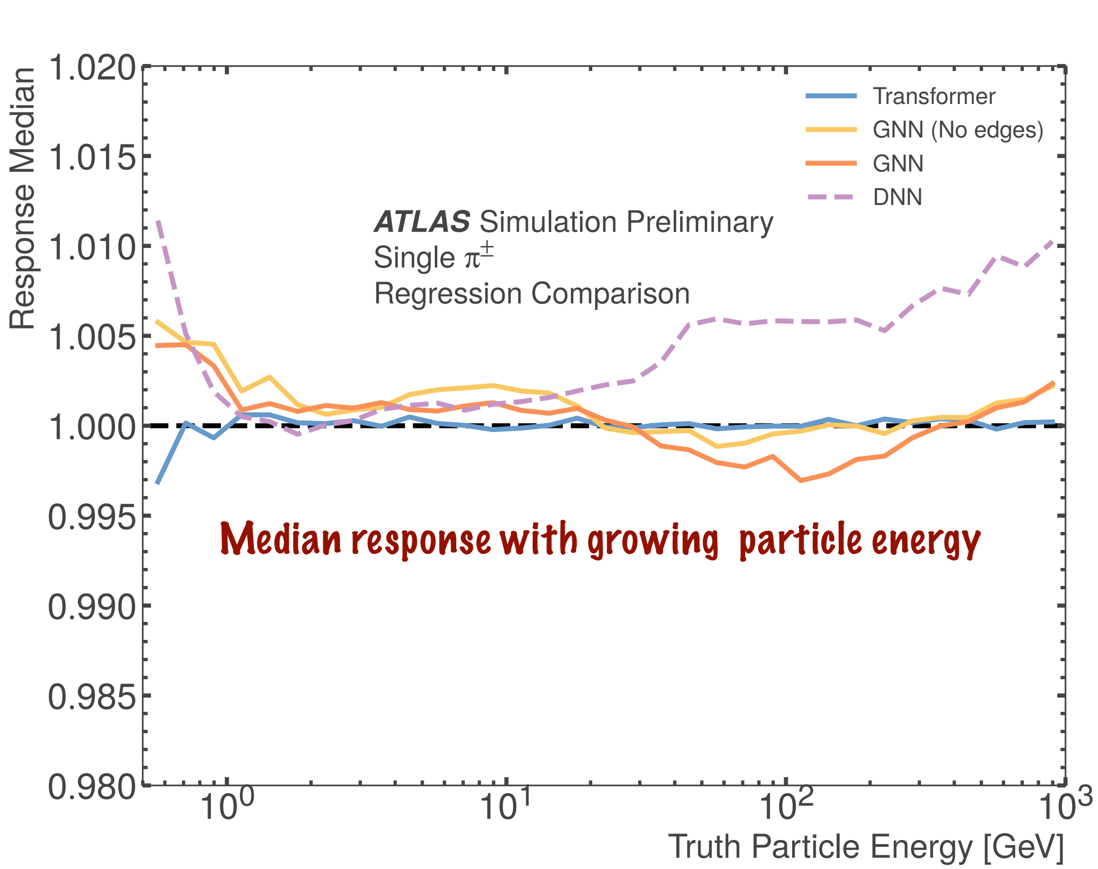


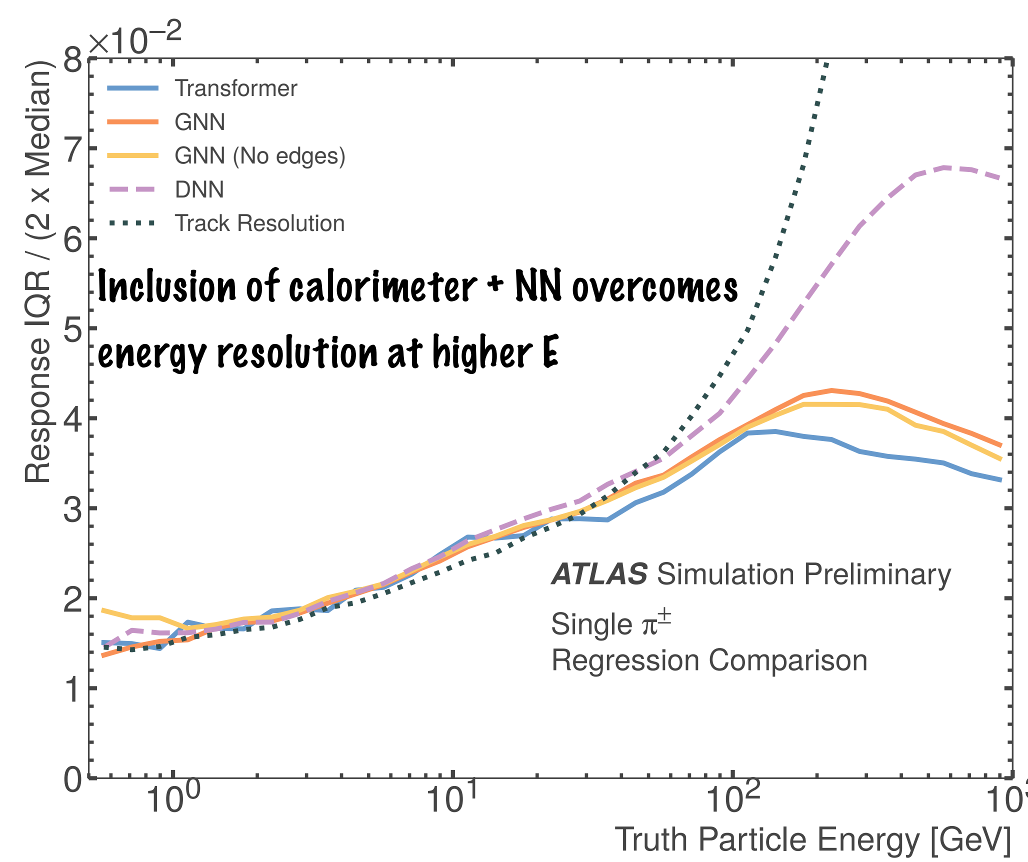


Pion identification within ATLAS

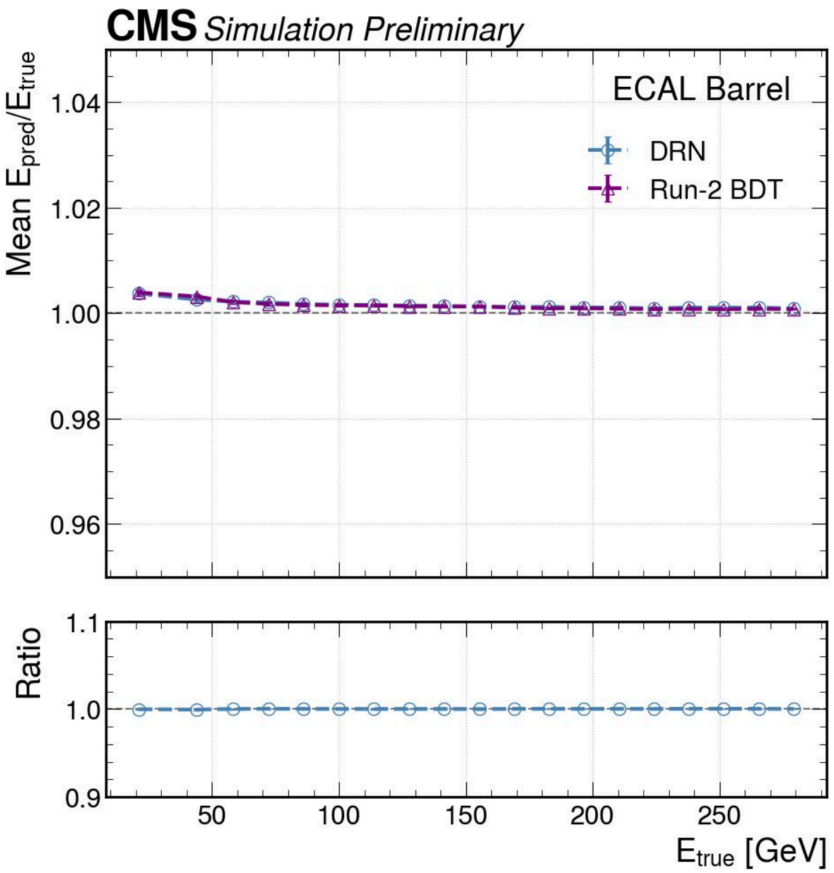


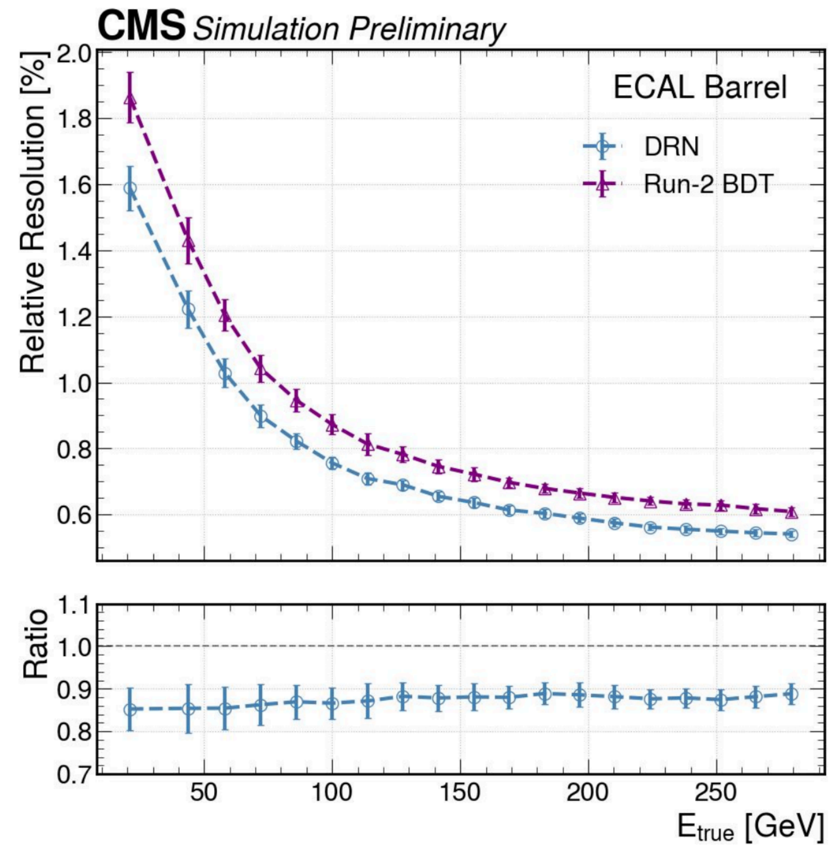




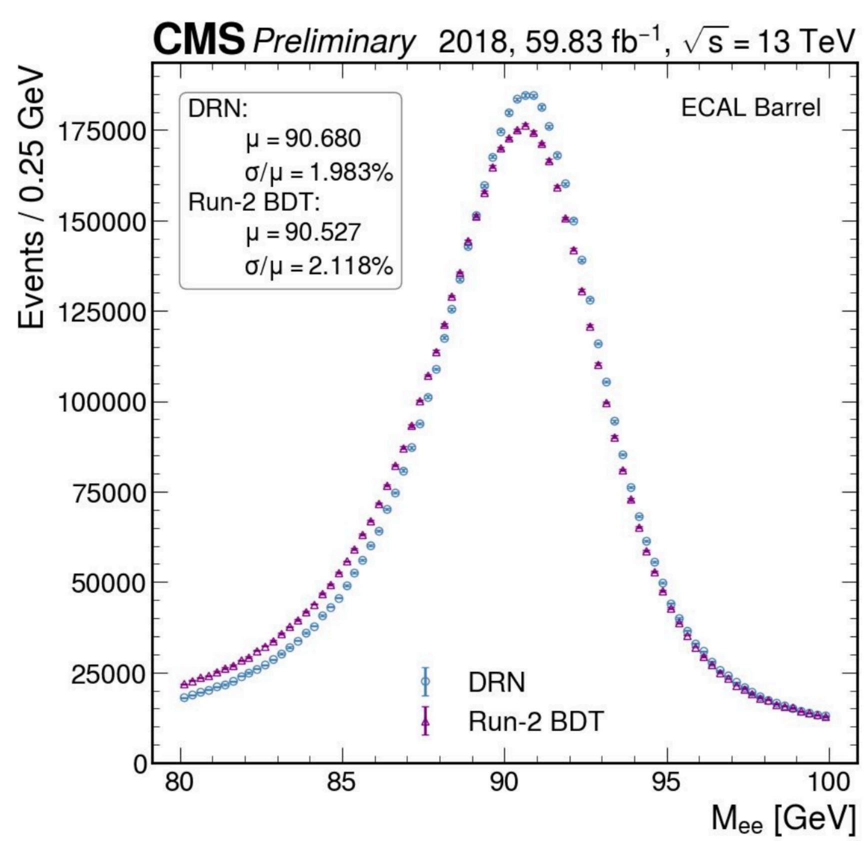


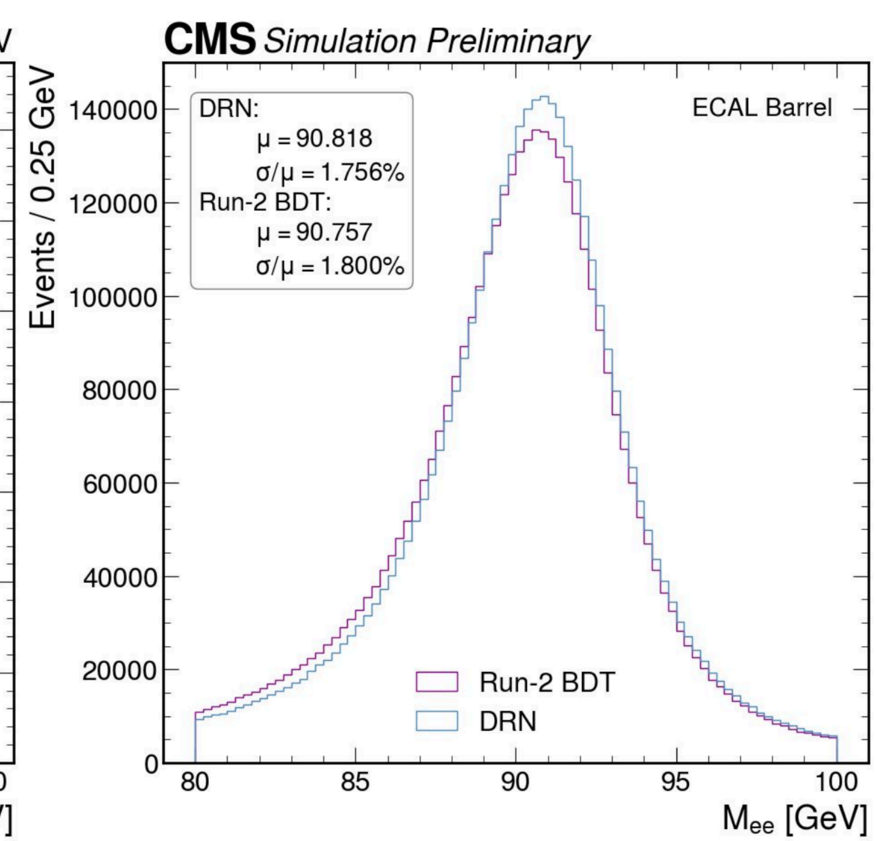
Electron identification within CMS





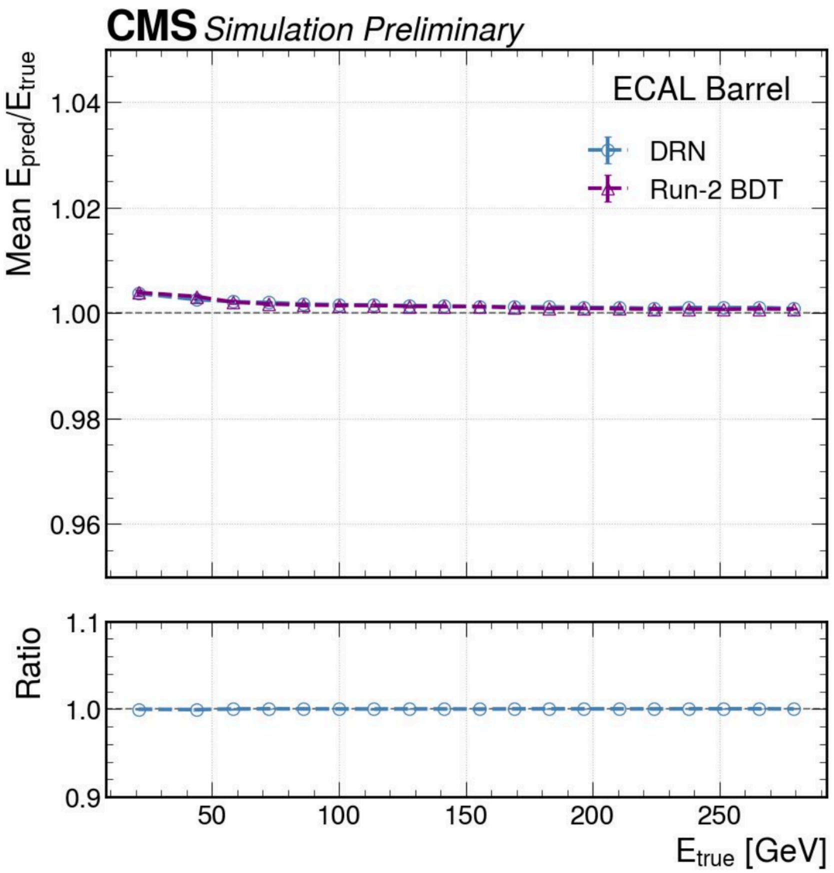
DRN is a dynamic GNN

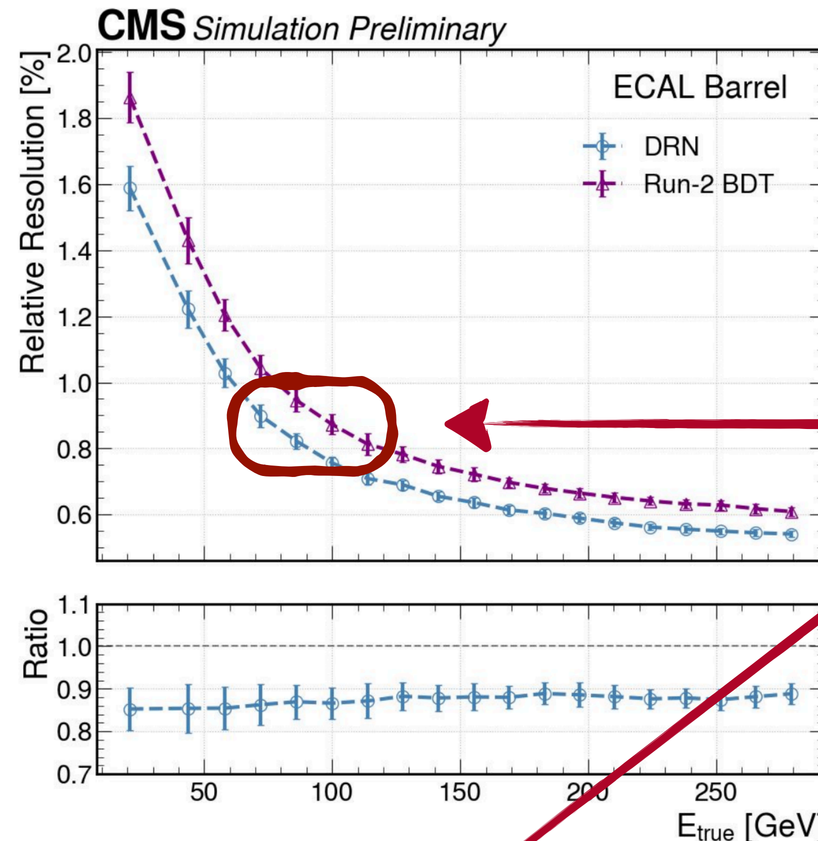




CMS-DP2022_009

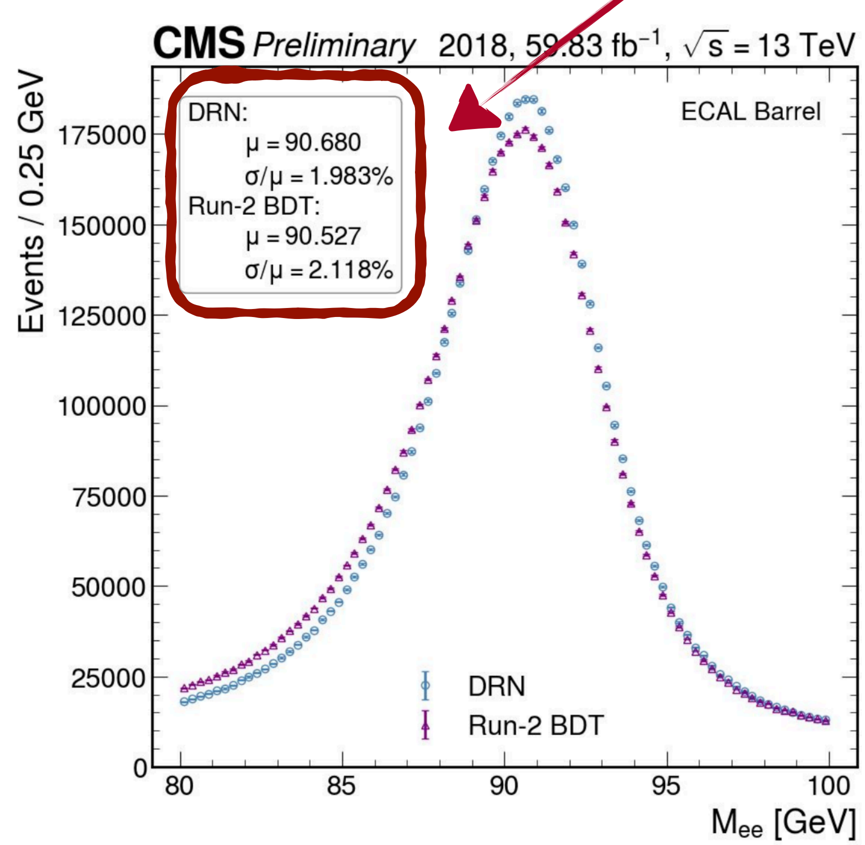
Electron identification within CMS

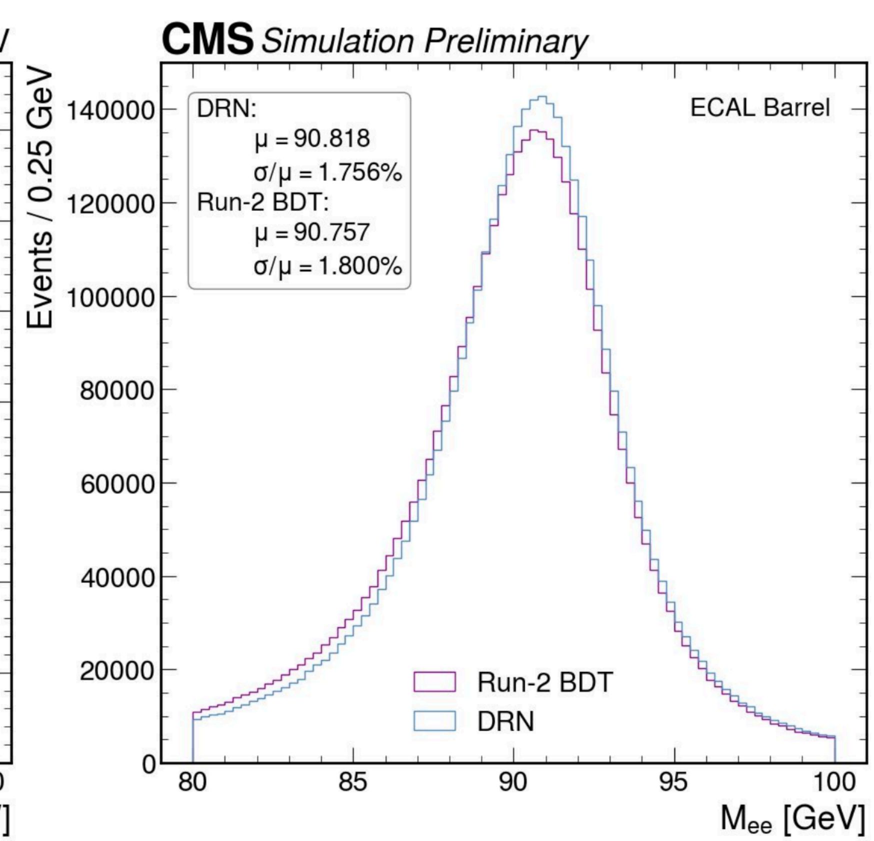




DRN is a dynamic GNN

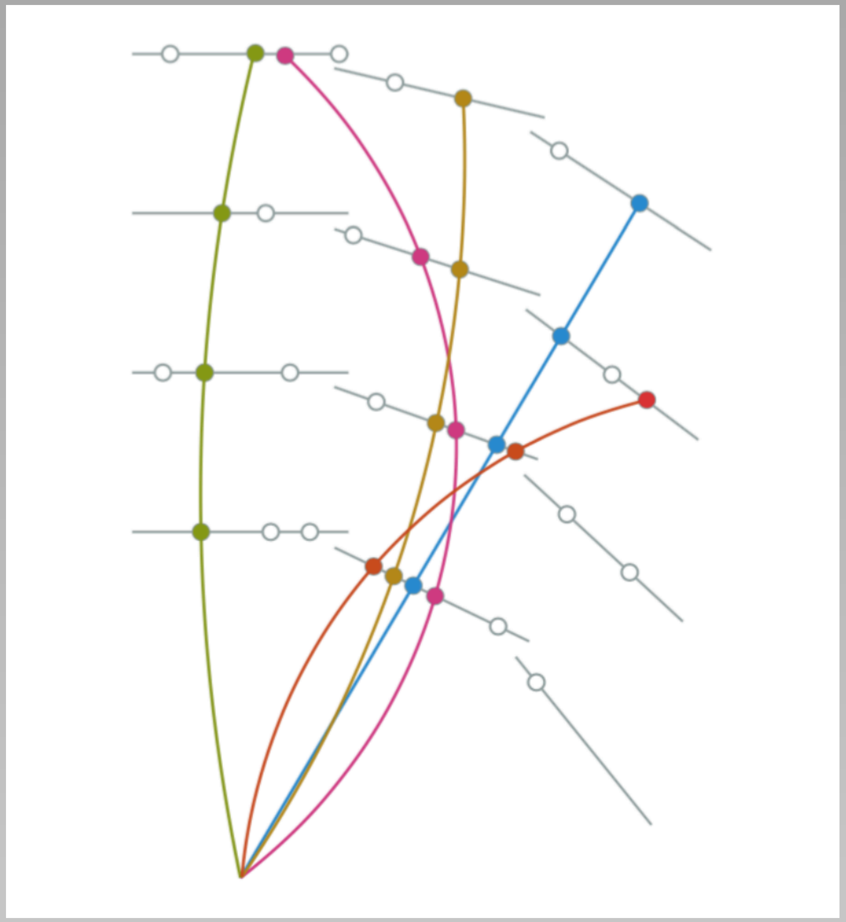
Improvement from a modern ML algorithm



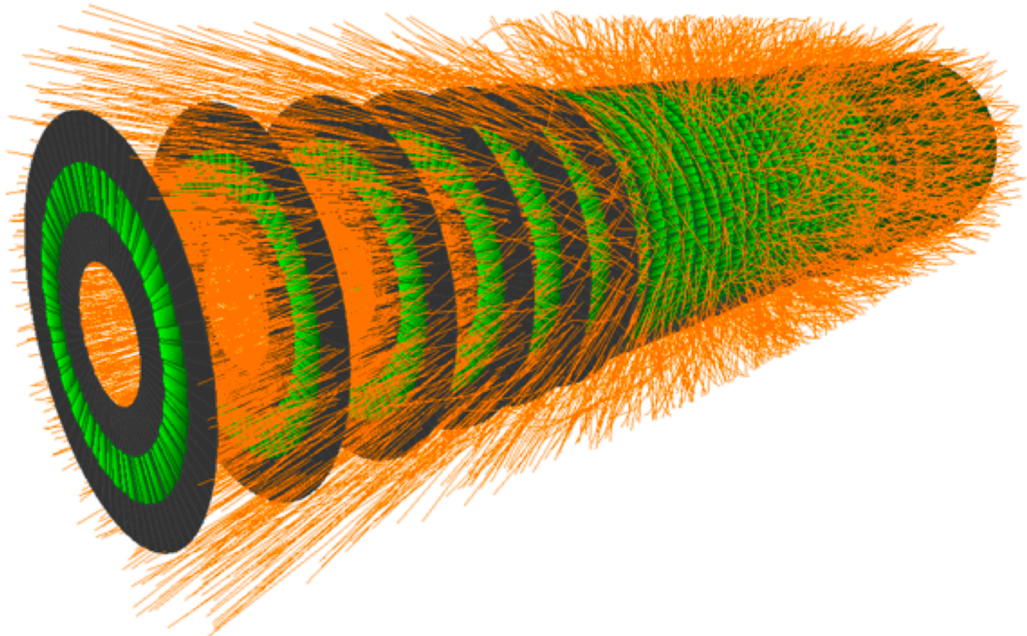


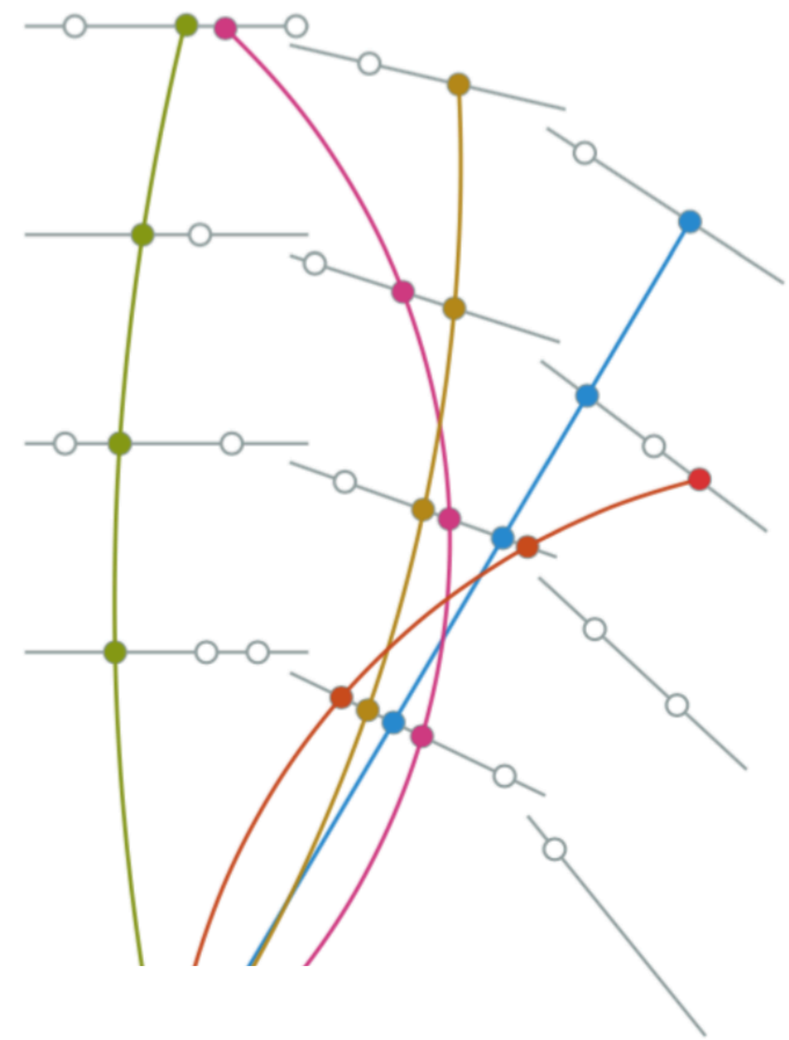
CMS-DP2022_009

Tracking

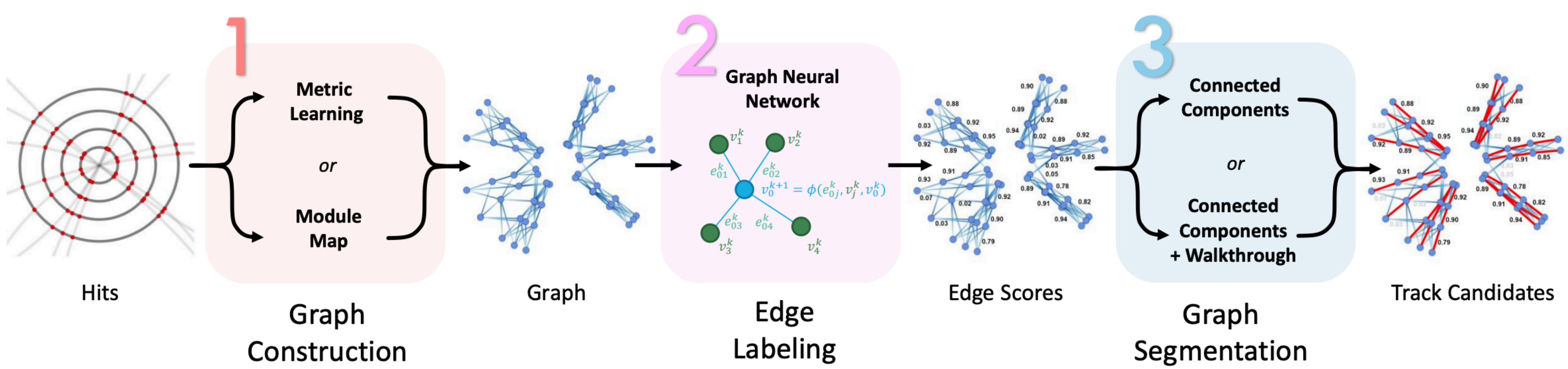
$$f_{\{\theta\}} \Bigg(\quad$$  $$\Bigg)$$

Tracking & ML

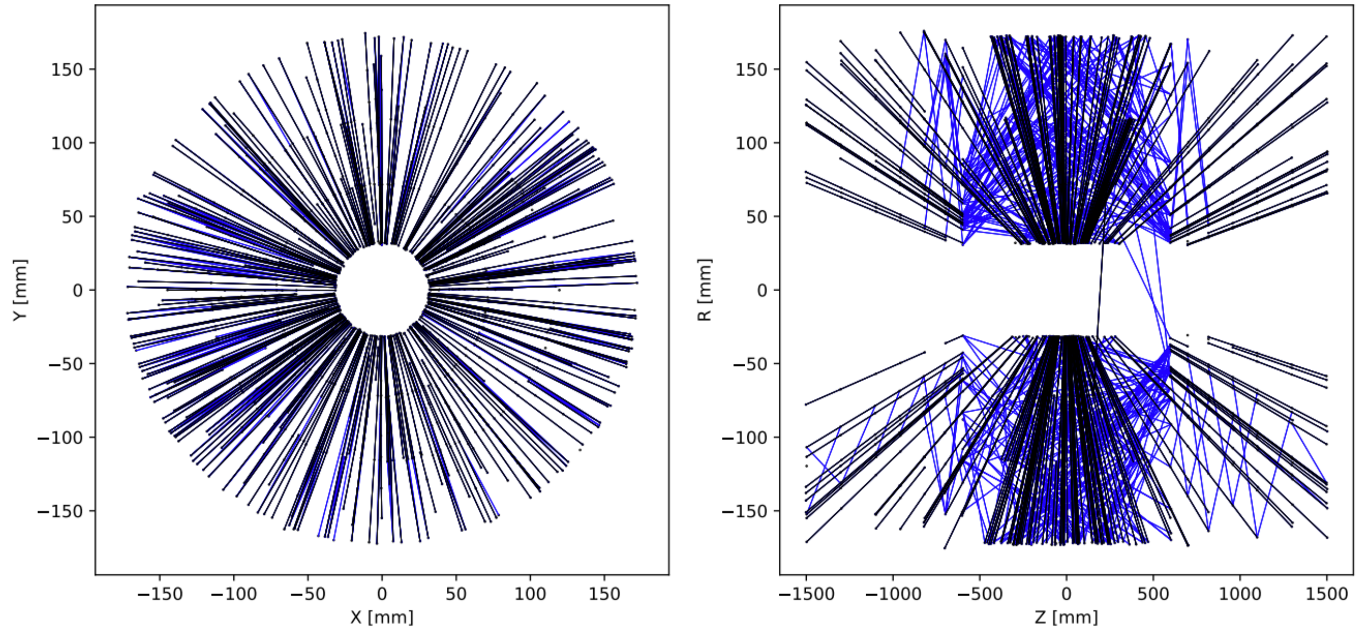


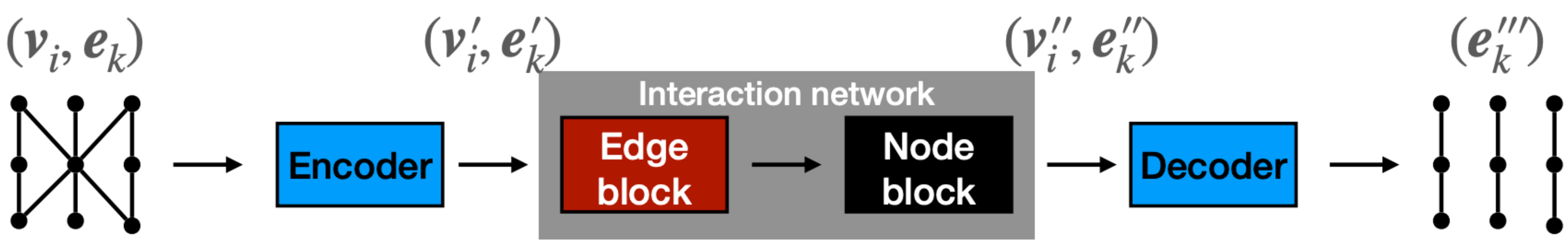


An exponentially large edge finding problem



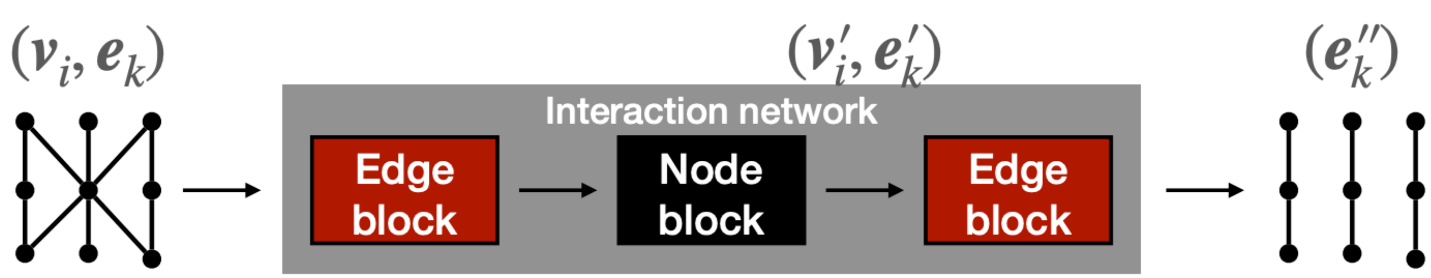
Tracking & ML





$e_k' = \phi_1^e(e_k)$ $\quad e_k'' = \phi_2^e(e_k', v_{r_k}', v_{s_k}')$ $\quad v_i'' = \phi_2^v(\bar{e}_i'', v_i')$ $\quad e_k''' = \phi_3^e(e_k'')$

$v_i' = \phi_1^v(v_i)$ $\quad \bar{e}_i'' = \rho^{e \to v}(E_i')$

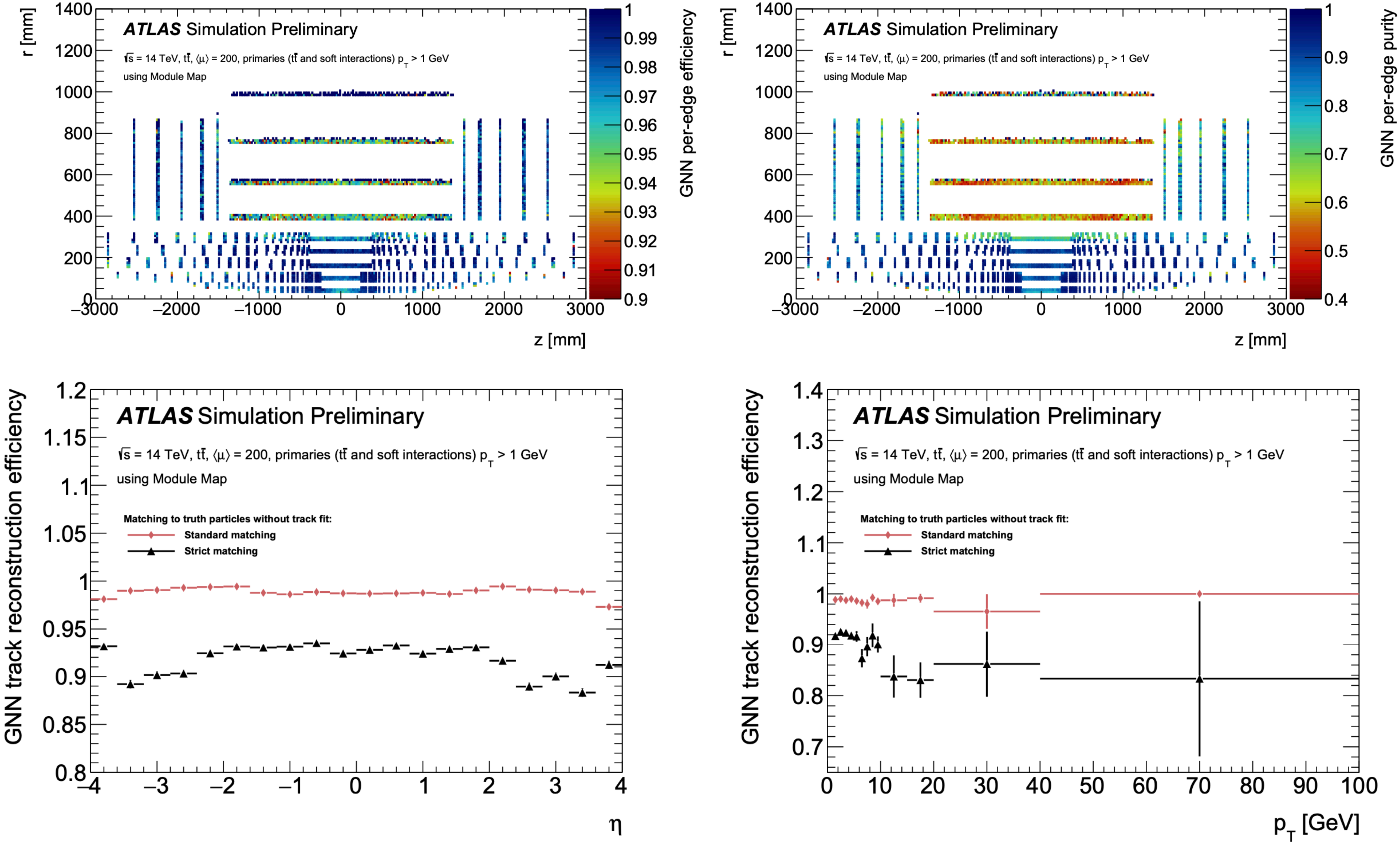


$e_k' = \phi_2^e(e_k, v_{r_k}, v_{s_k})$ $\quad v_i' = \phi_2^v(\bar{e}_i', v_i)$ $\quad e_k'' = \phi_2^e(e_k', v_{r_k}', v_{s_k}')$

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

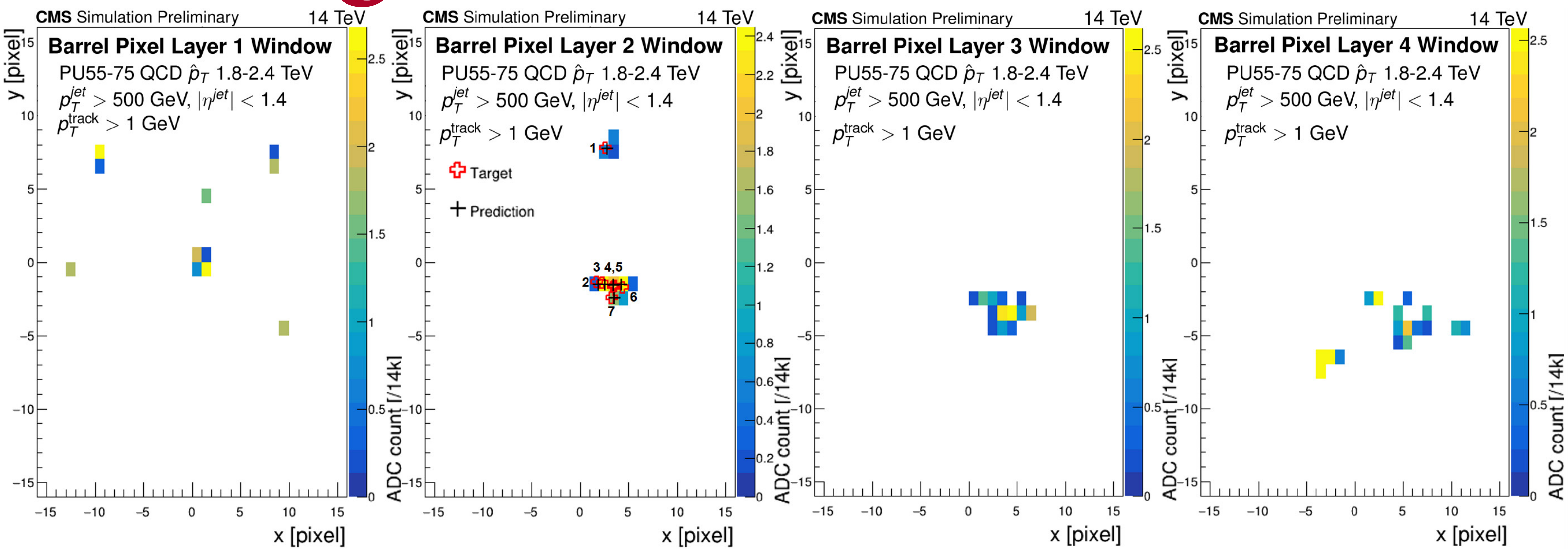
Tracking & ML : ATLAS

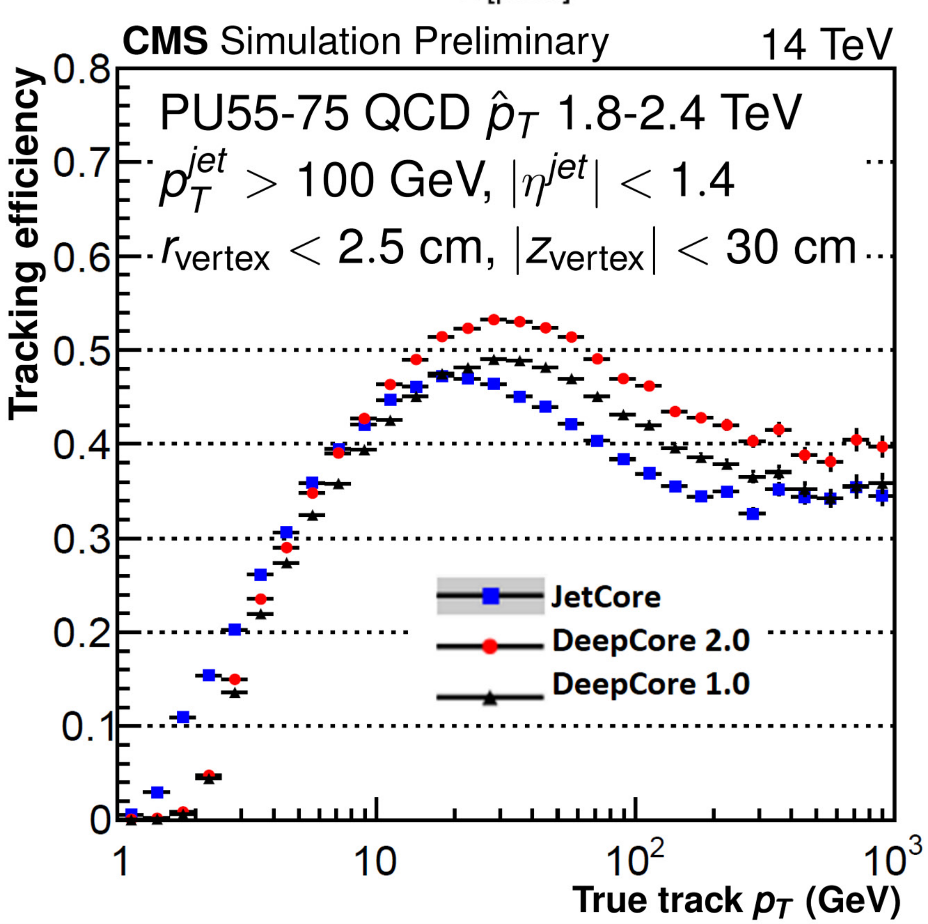
ATL-ITK-PROC-2022-006

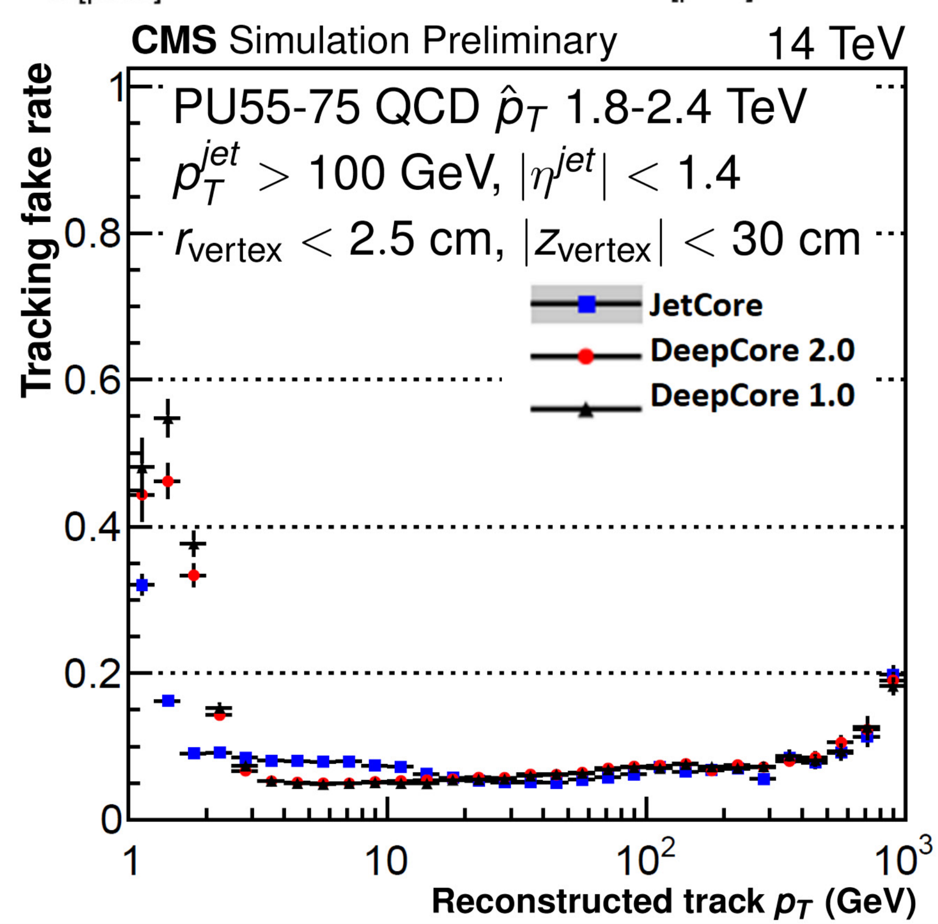


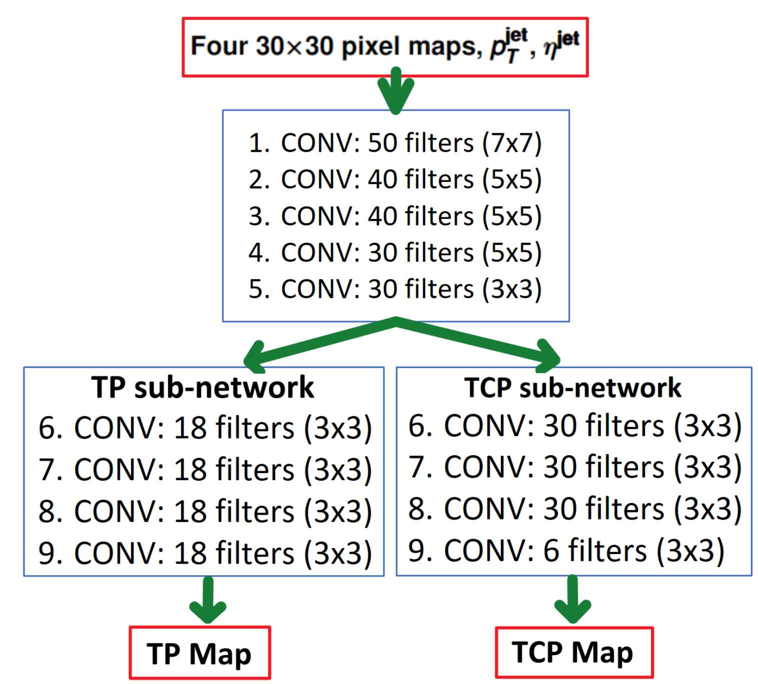
Tracking & ML : CMS

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



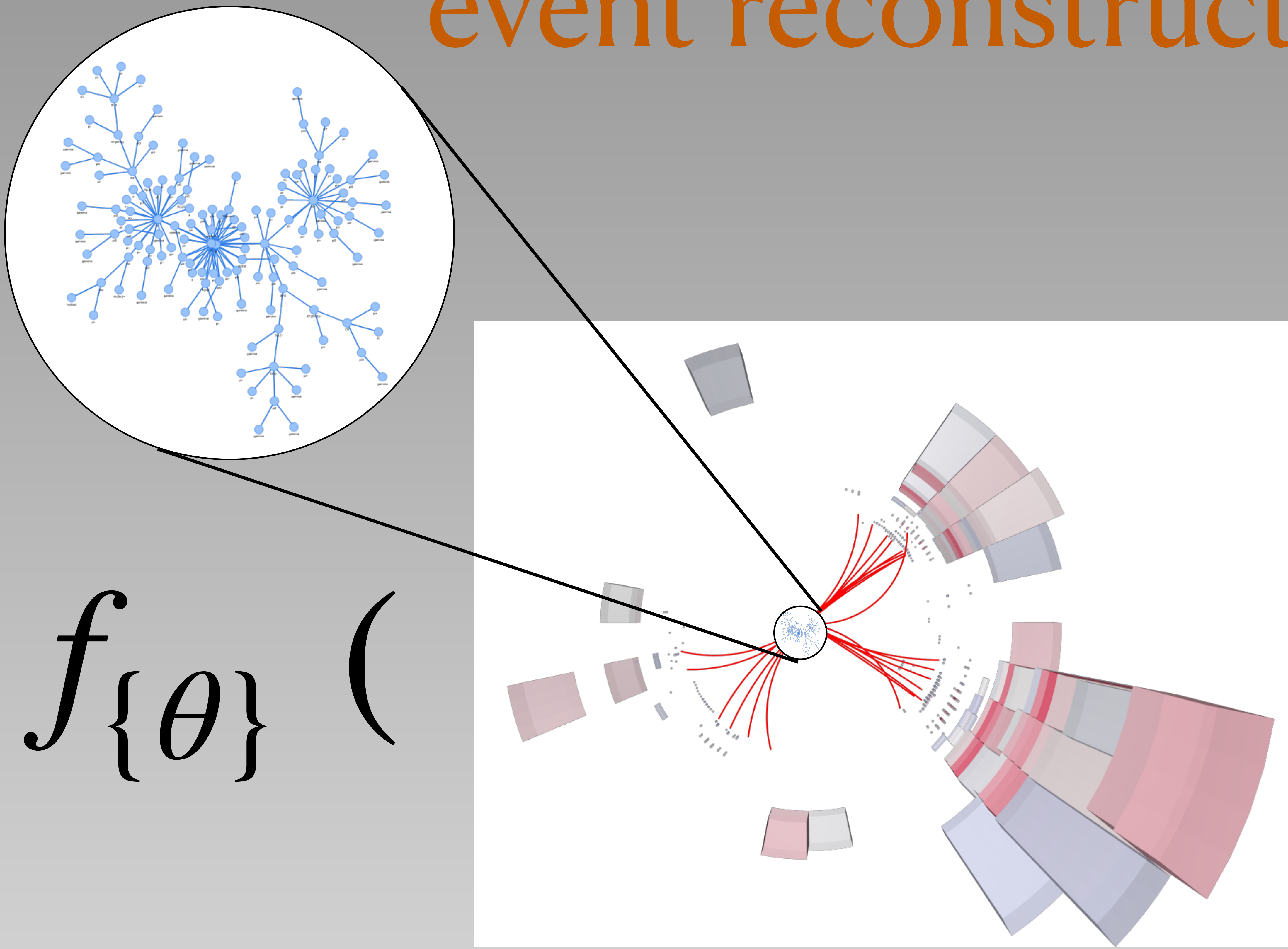






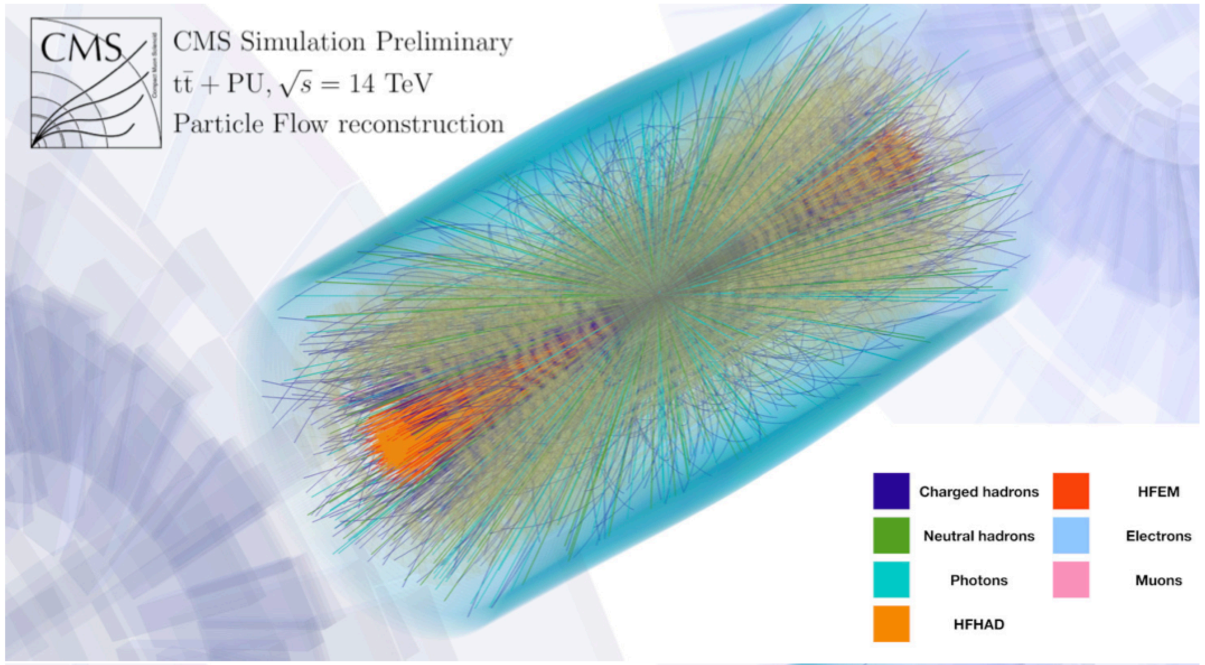
Particle-Flow event reconstruction

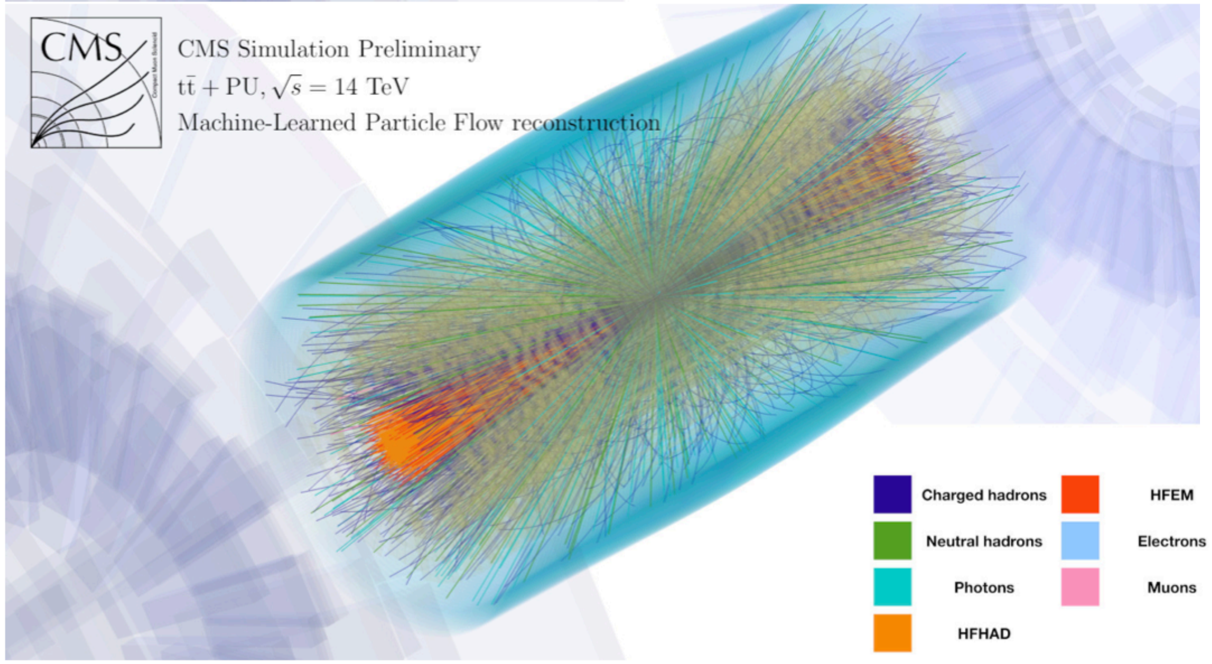
$f_{\{\theta\}}\Big($

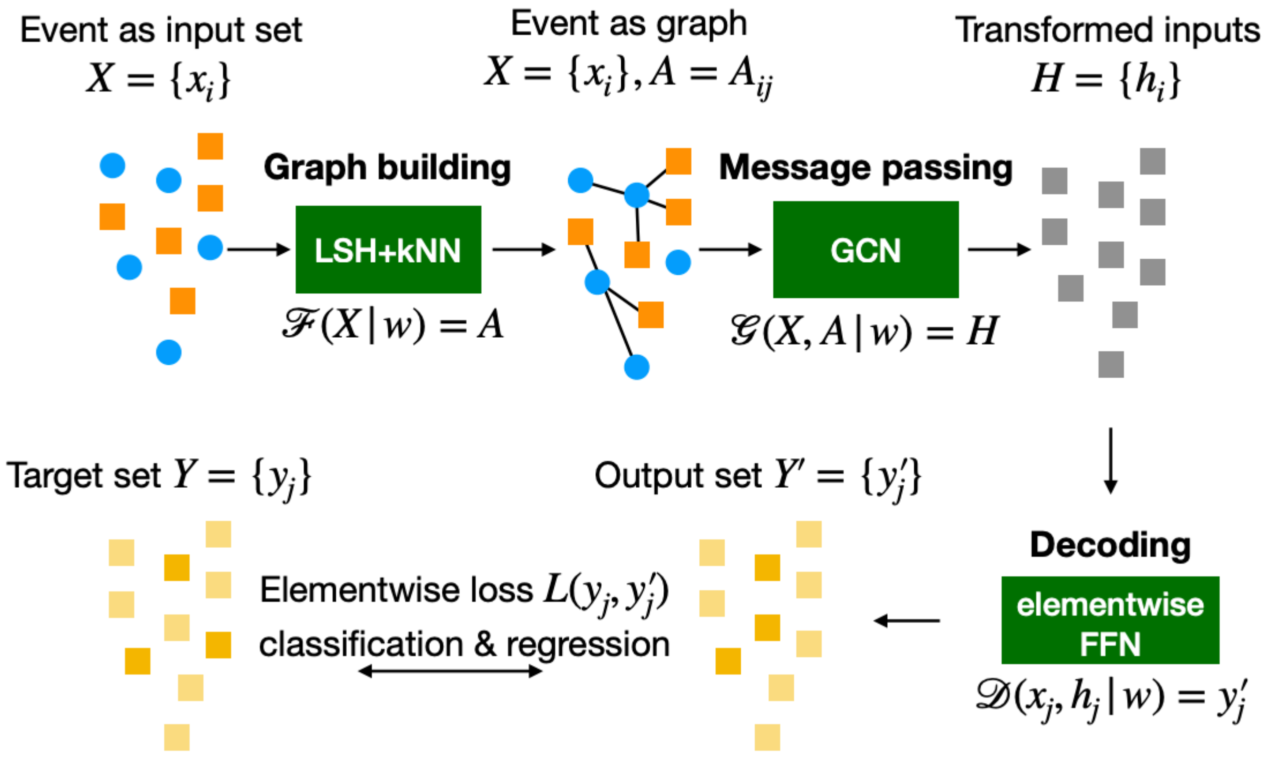


$\Big)$

Full ML driven PFlow : MLPF







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

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

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

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

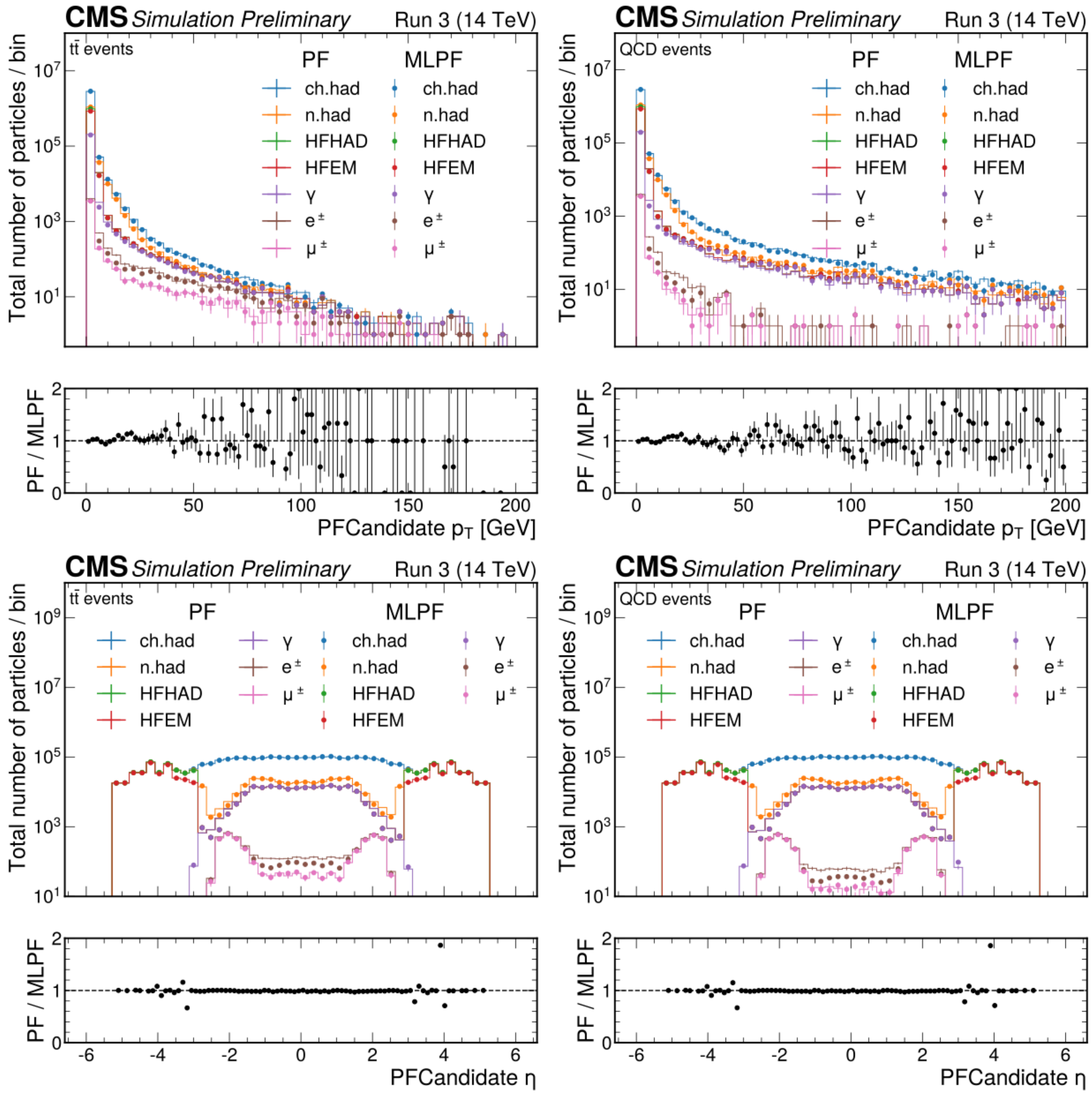
● - track, ■ - calorimeter cluster, ■ - encoded element

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

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

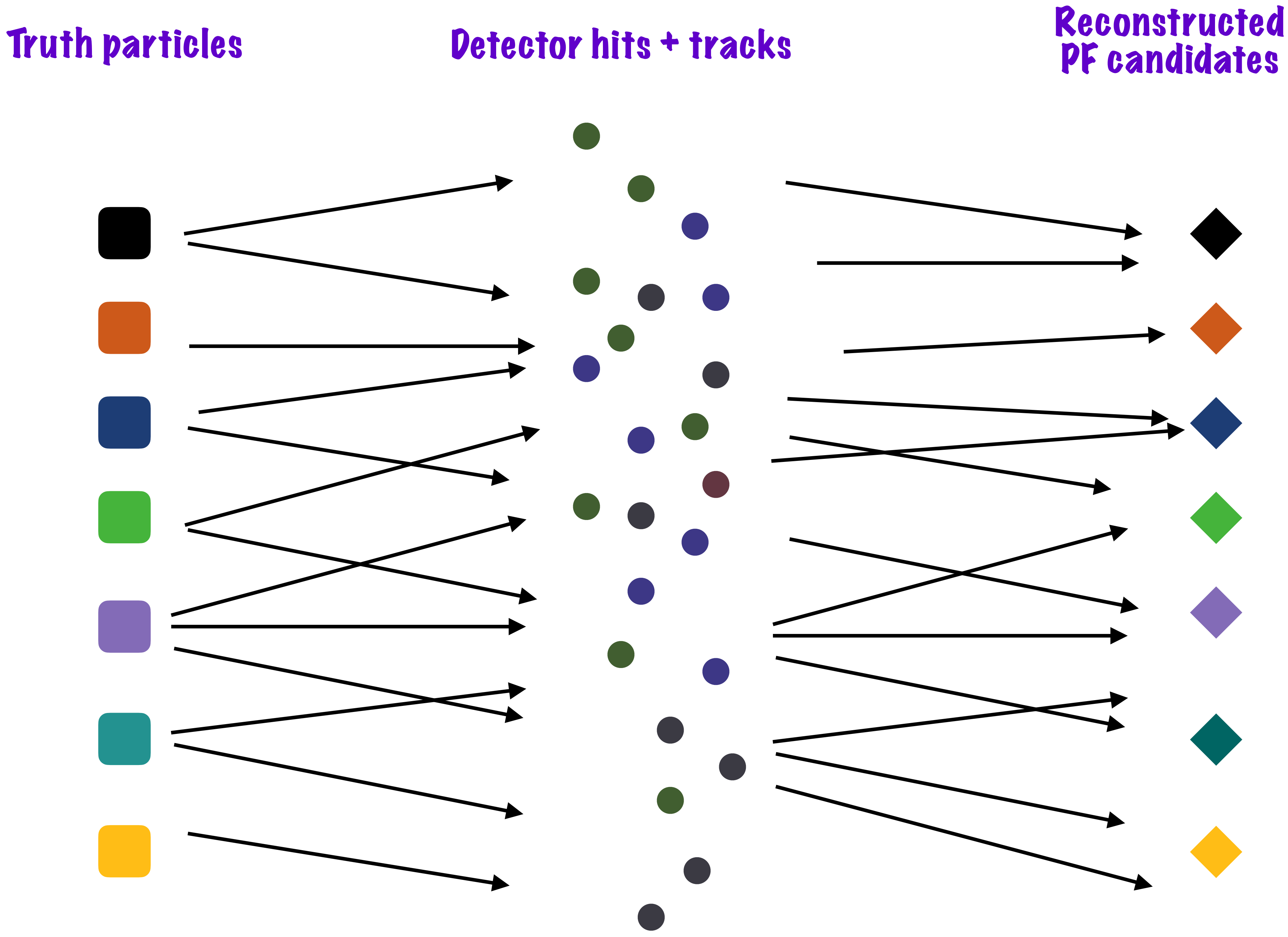
PF lepton, hadron, photon = F_{PF} (track hits + calo cells)

Combining track + calo for PFlow

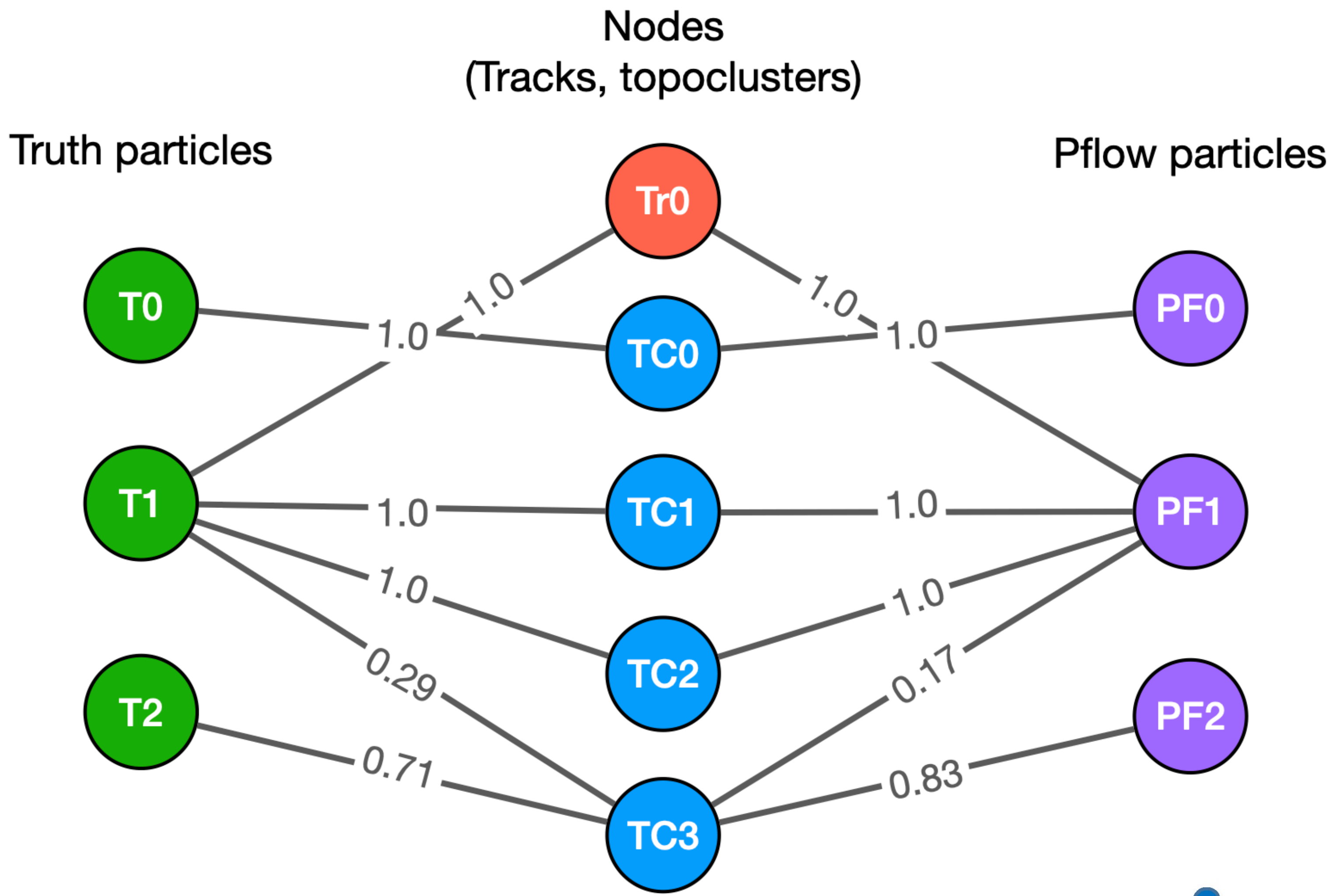


MLPF
arXiv : 2203.00330
J. Pata et. al.

What's the core data structure?

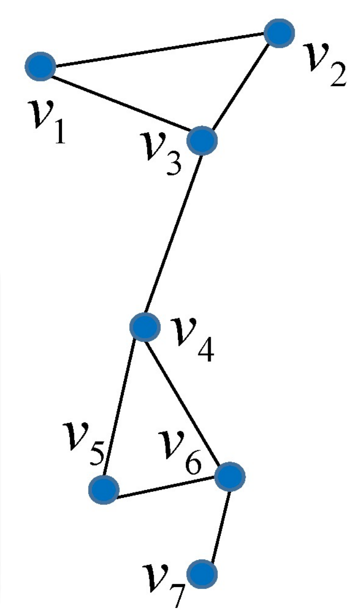


What's the core data structure?

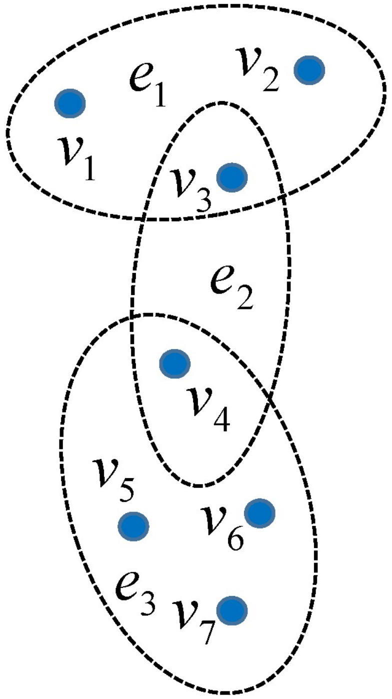


https://www.mdpi.com/2072-4292/9/5/506

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



(a) Simple graph



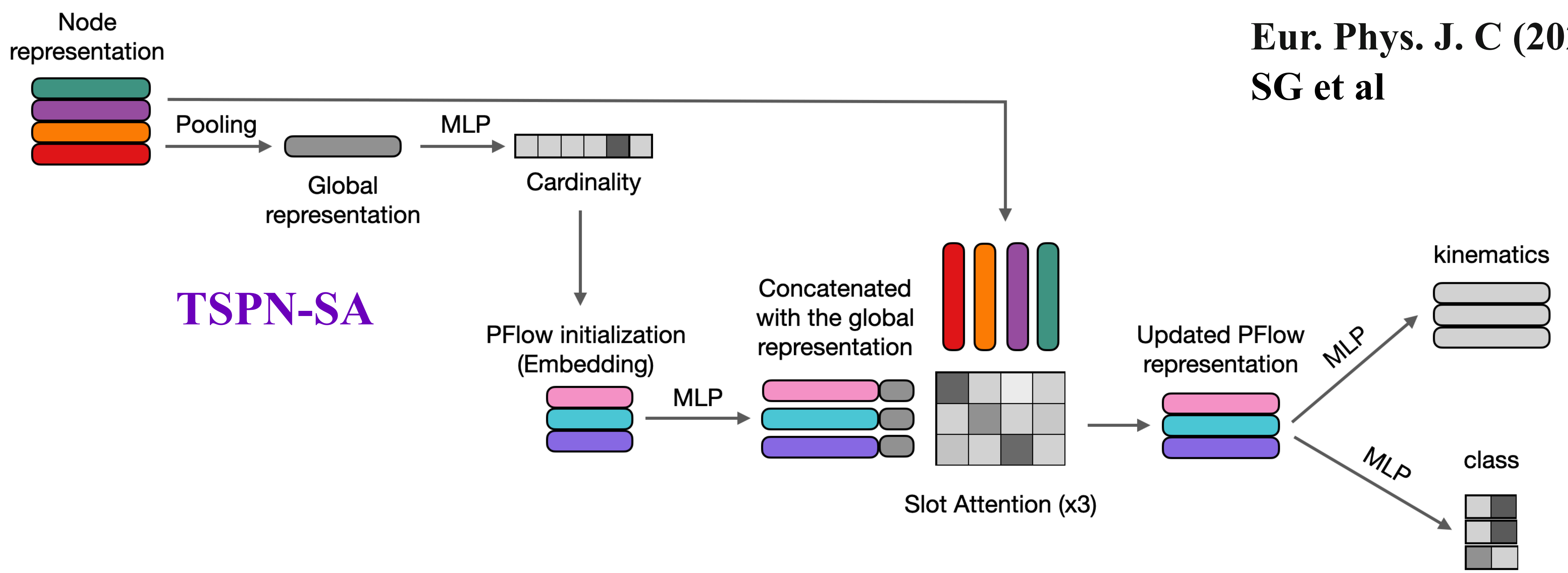
(b) Hypergraph **G**

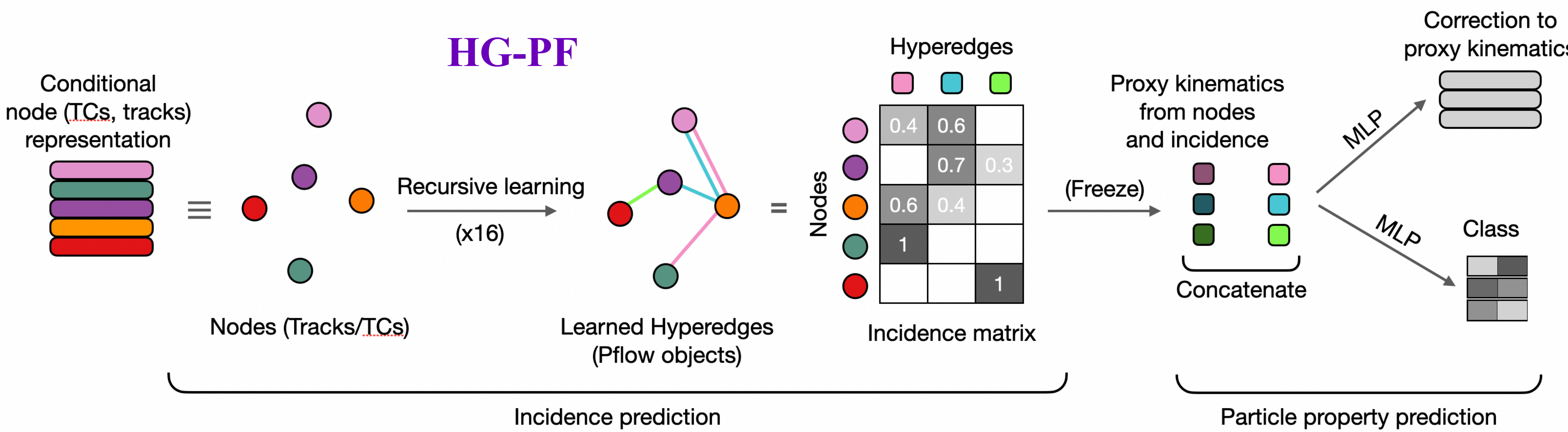
	e_1	e_2	e_3
v_1	1	0	0
v_2	1	0	0
v_3	1	1	0
v_4	0	1	1
v_5	0	0	1
v_6	0	0	1
v_7	0	0	1

(c) Incidence matrix **H**

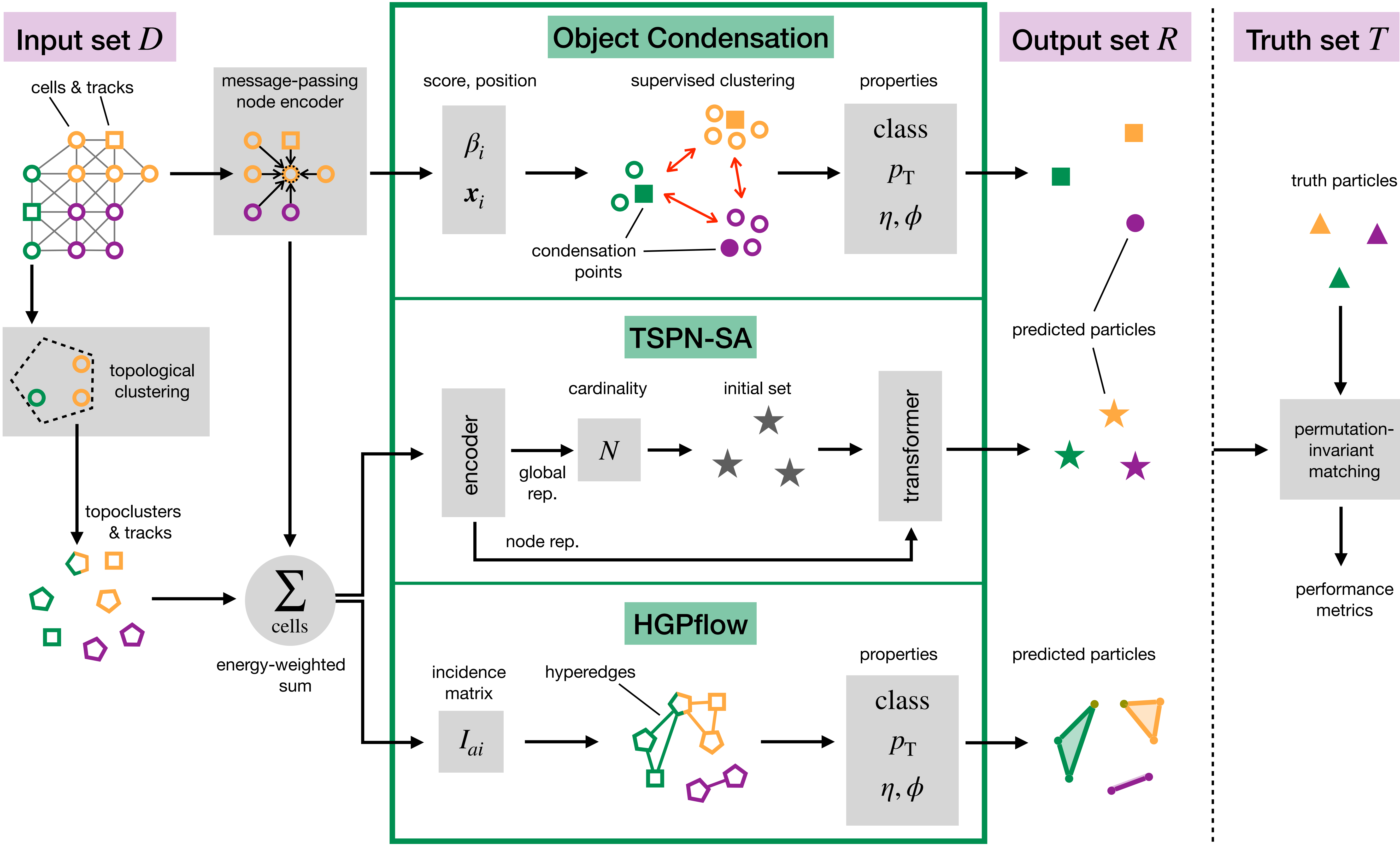
The new networks we tried

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

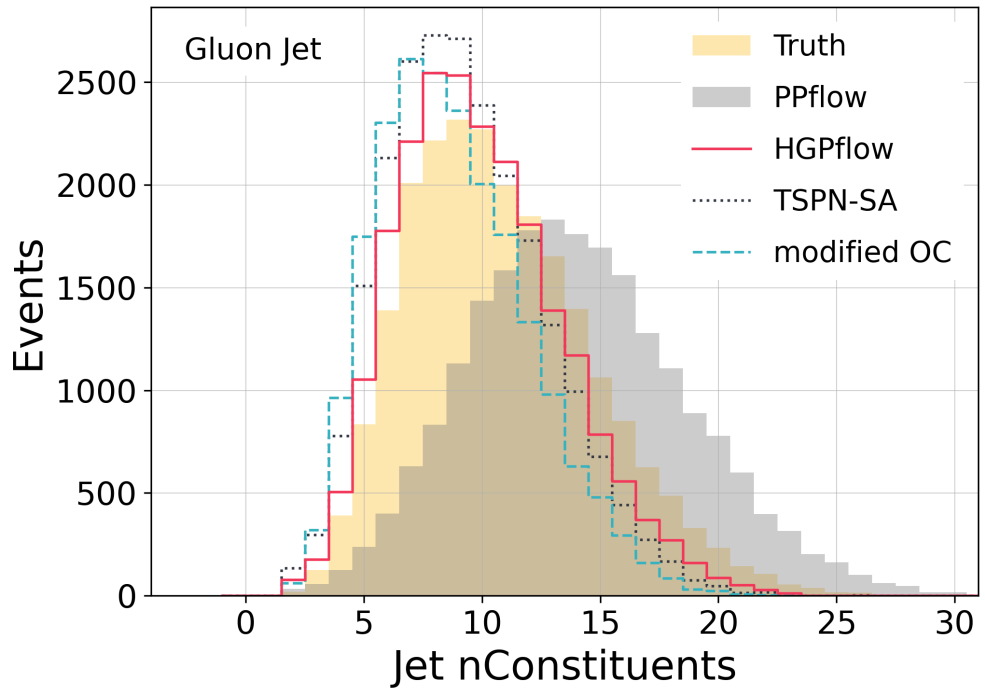
The efficiency and fake rate are defined as follows:

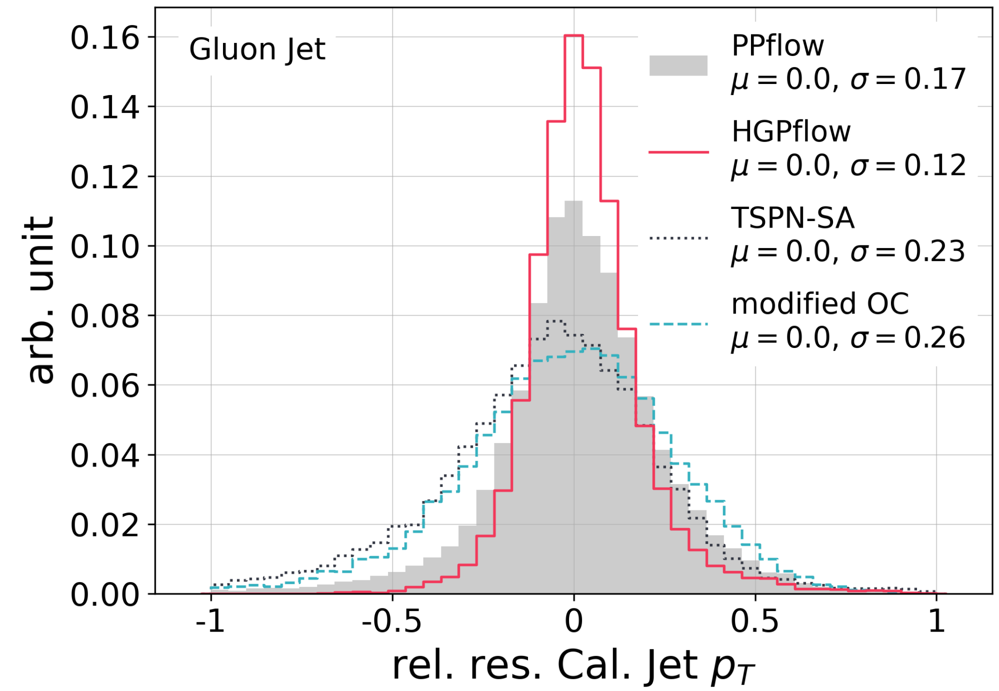
$$\varepsilon \equiv \frac{N(\text{matched pred})}{N(\text{targ})}\ ,\ f \equiv \frac{N(\text{unmatched pred})}{N(\text{pred})}$$

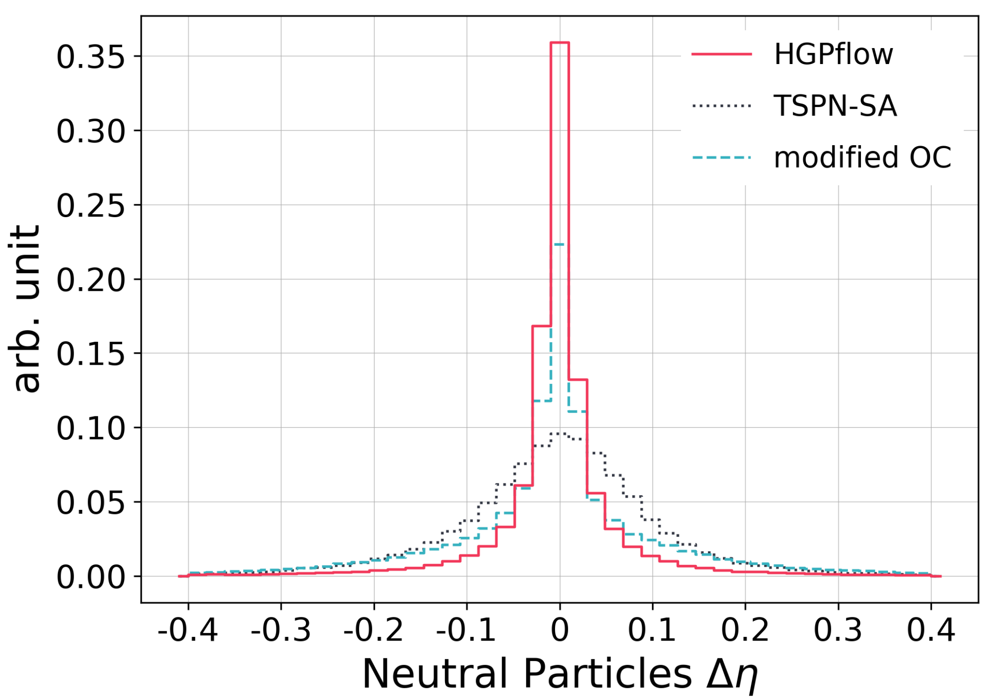
Set-to-set matching is done using the Hungarian distance :

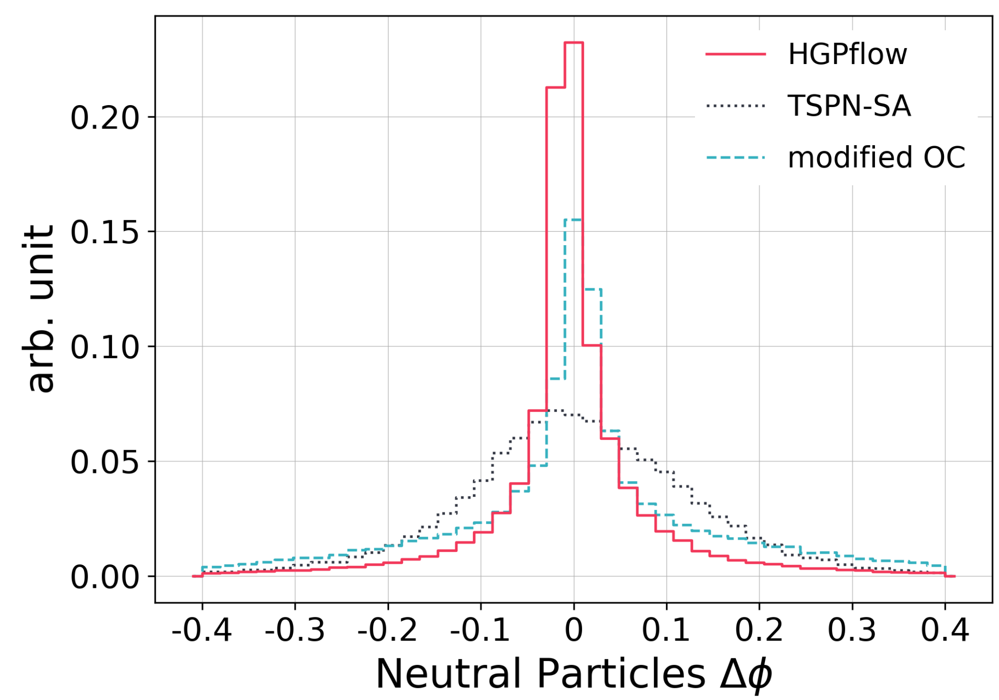
$$\sqrt{\mathbf{c_{p_T}(\Delta p_T/p_T^{truth})^2 + C_{\Delta R}(\Delta R)^2}}$$

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

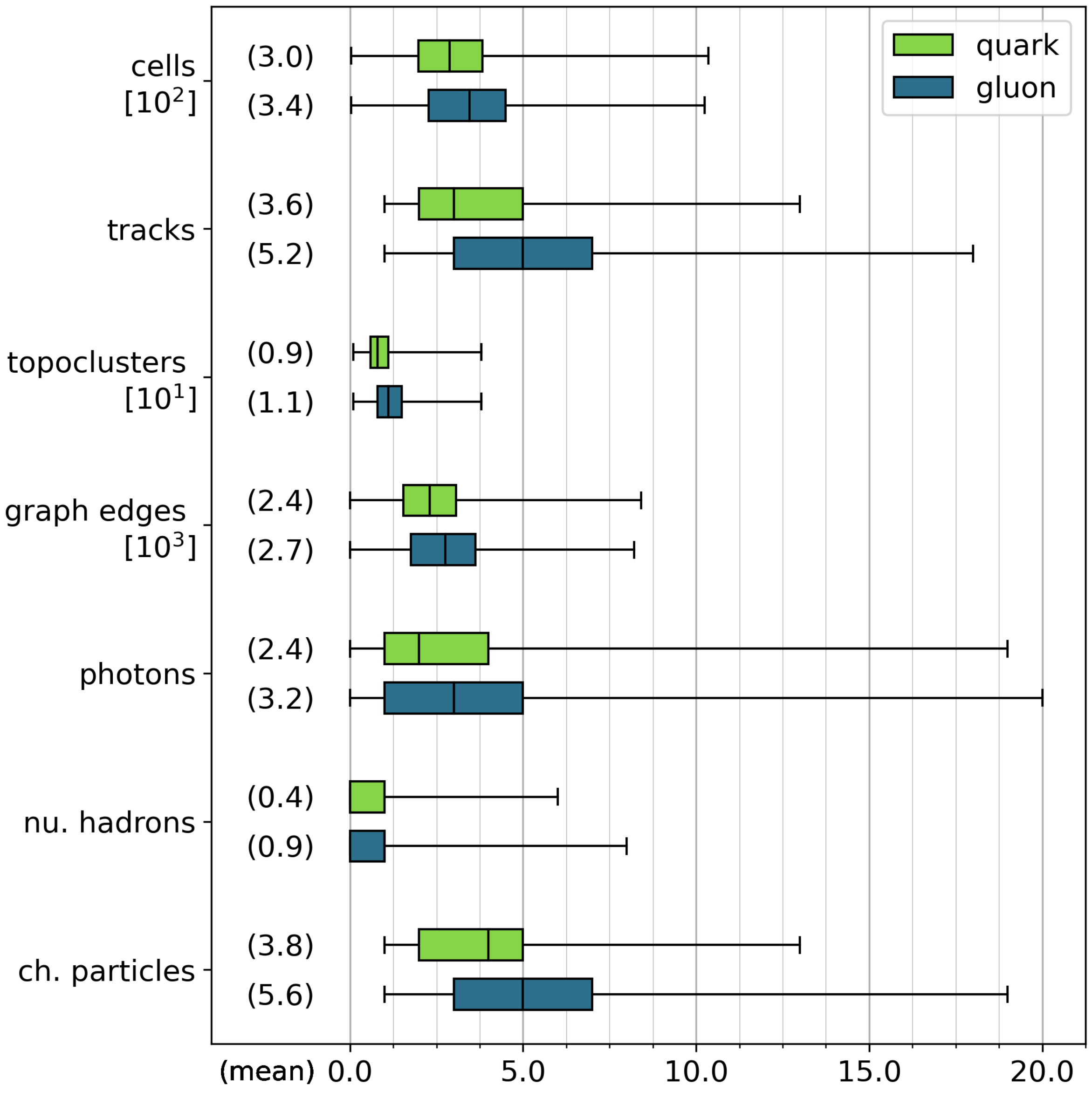


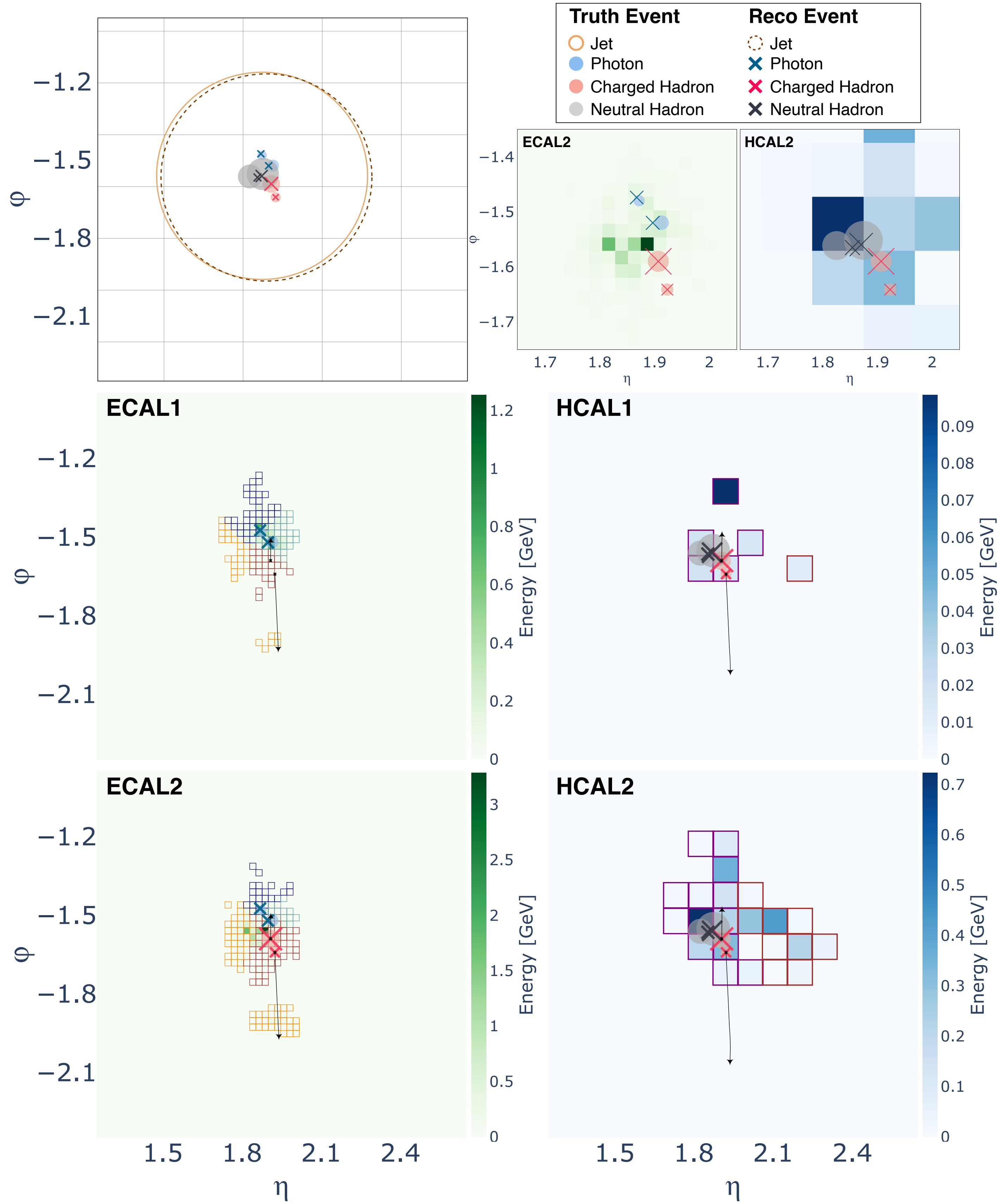






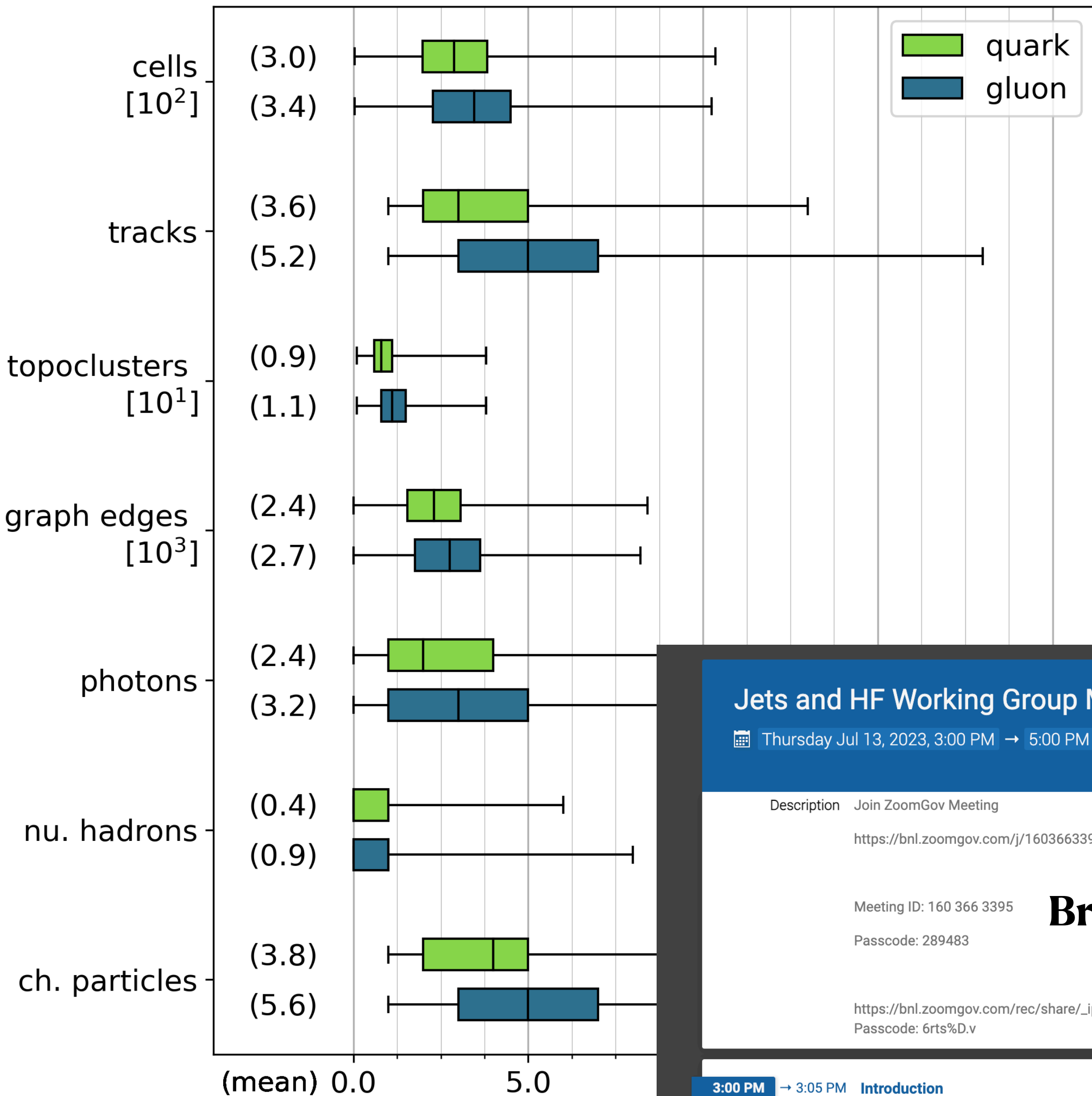
Data complexity & sample output

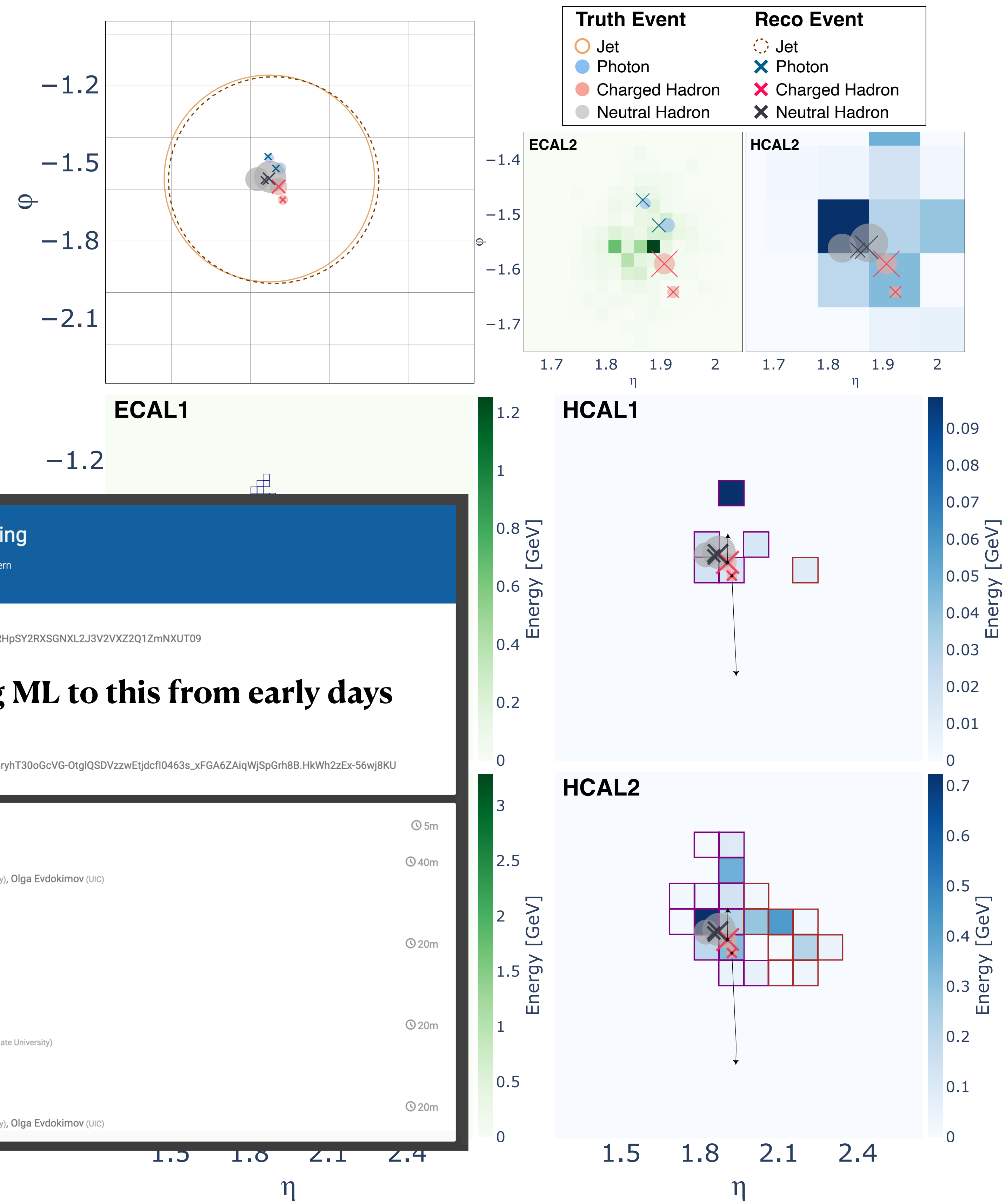




HG-PFlow seems to be doing a better job among the compared methods

Data complexity & sample output





Jets and HF Working Group Meeting

Thursday Jul 13, 2023, 3:00 PM → 5:00 PM US/Eastern

Description Join ZoomGov Meeting

https://bnl.zoomgov.com/j/1603663395?pwd=RHpSY2RXSGNXL2J3V2VXZ2Q1ZmNXUT09

Meeting ID: 160 366 3395

Passcode: 289483

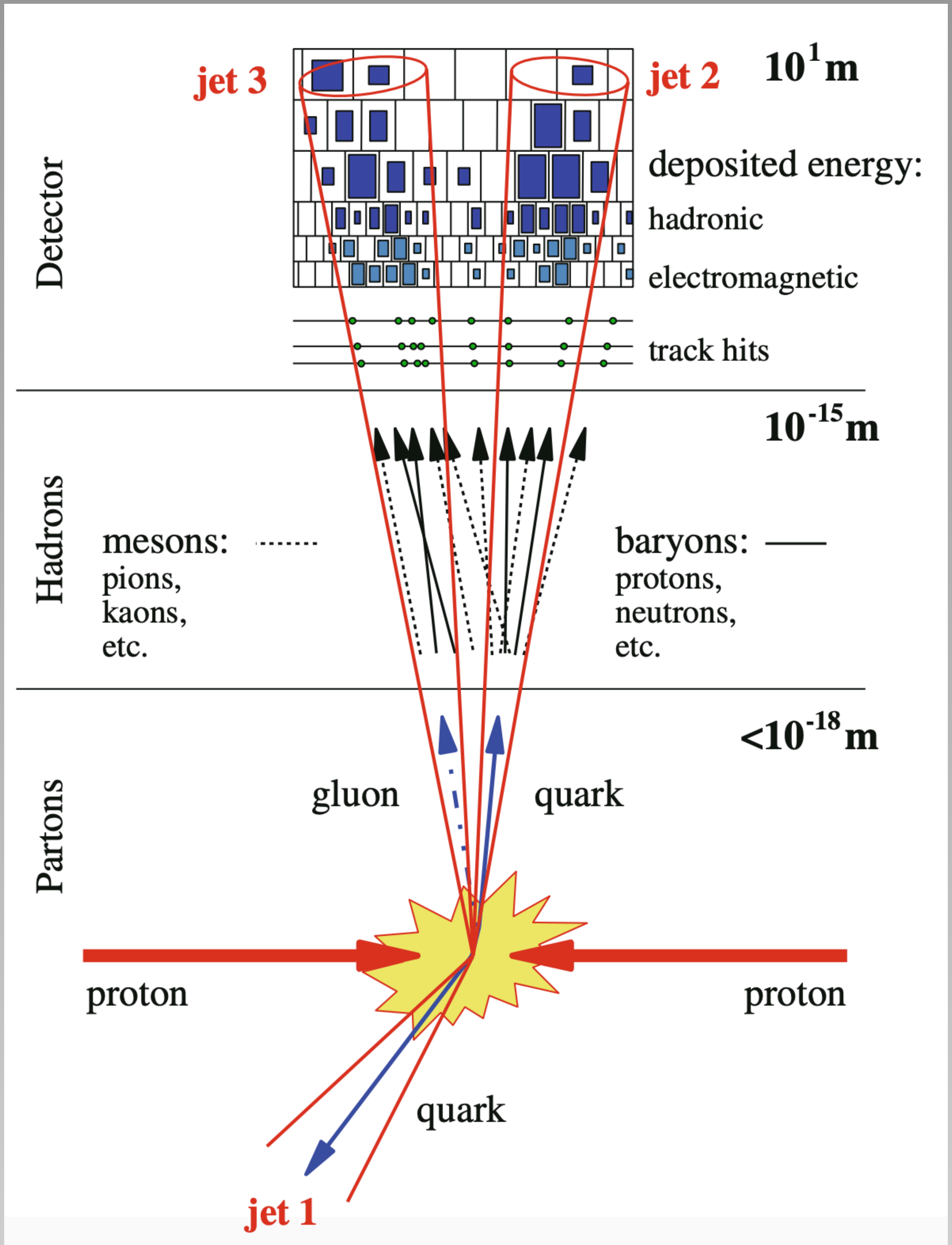
https://bnl.zoomgov.com/rec/share/_ipKV0vlU6ryhT30oGcVG-OtglQSDVzzwEtjdcfI0463s_xFGA6ZAiqWjSpGrh8B.HkWh2zEx-56wj8KU
Passcode: 6rts%D.v

3:00 PM → 3:05 PM	Introduction	5m
3:05 PM → 3:45 PM	Particle Flow in CMS Speakers: Brian Page (Brookhaven National Laboratory), Olga Evdokimov (UIC) PFCMS_MAN_2307...	40m
3:45 PM → 4:05 PM	PF implementation survey Speaker: Derek Anderson (Iowa State University) EpicParticleFlow_S...	20m
4:05 PM → 4:25 PM	PF experience @ sPHENIX Speaker: Antonio Carlos Oliveira da Silva (Iowa State University) ParticleFlow_ePIC_...	20m
4:25 PM → 4:45 PM	Benchmark Discussion Speakers: Brian Page (Brookhaven National Laboratory), Olga Evdokimov (UIC)	20m

Bring ML to this from early days

HG-PFlow seems to a better job among t methods

Jet tagging



The Large Hadron Collider
Editors: Thomas Schörner-Sadenius

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

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

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

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

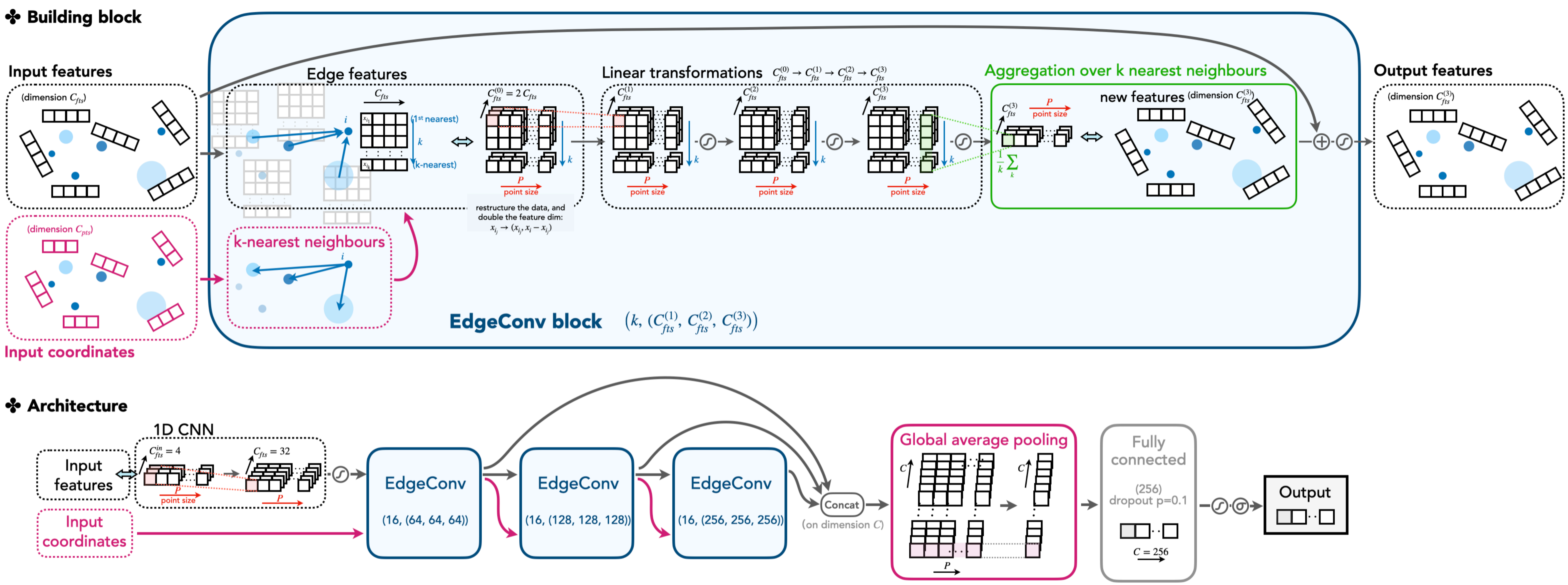
The forward problem is not computable from first principle

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

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

Object tagging

Particle Net : 1902.08570
Huilin Qu, Loukas Gouskos



arXiv > cs > arXiv:1801.07829

Computer Science > Computer Vision and Pattern Recognition

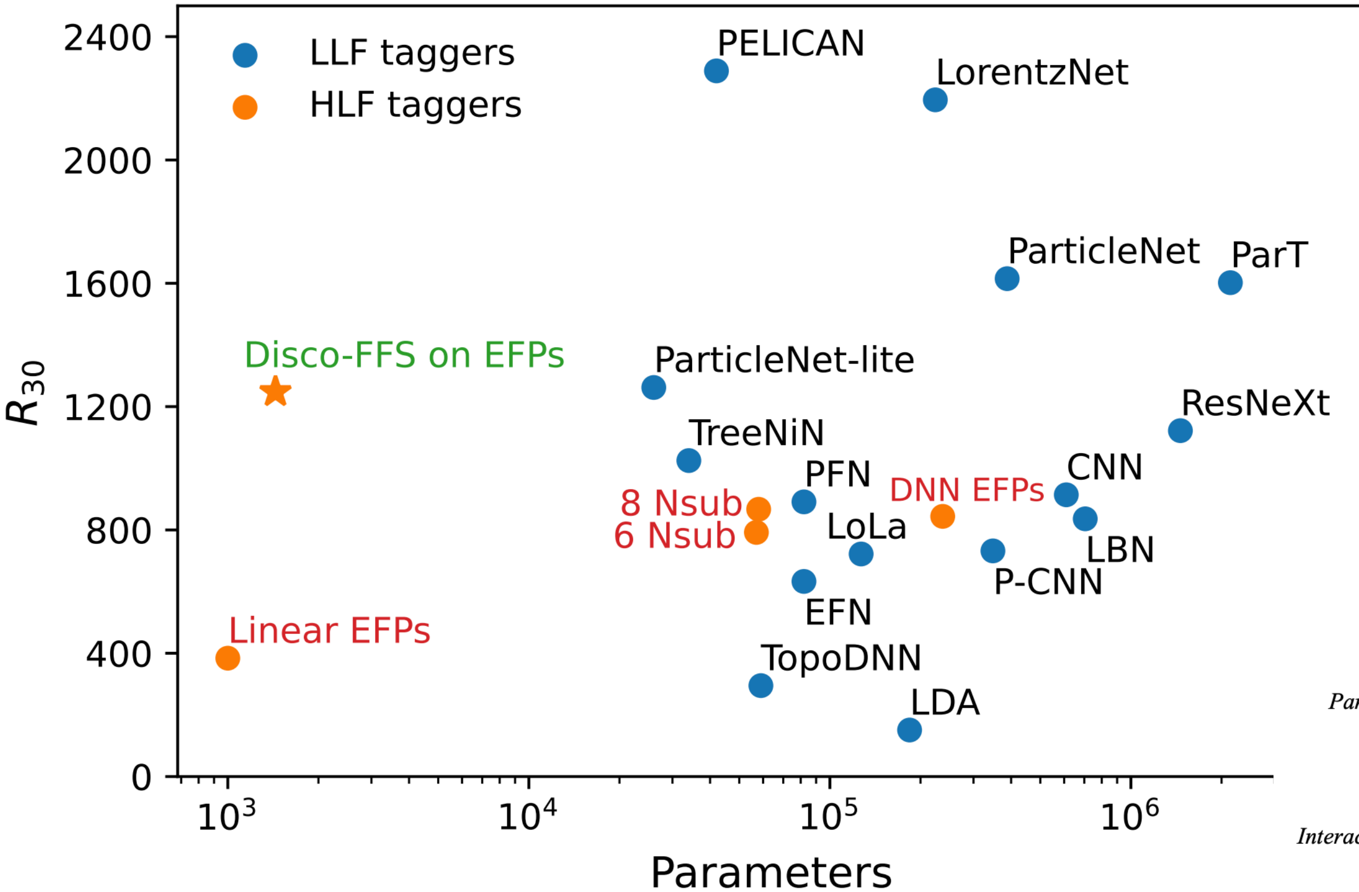
[Submitted on 24 Jan 2018 (v1), last revised 11 Jun 2019 (this version, v2)]

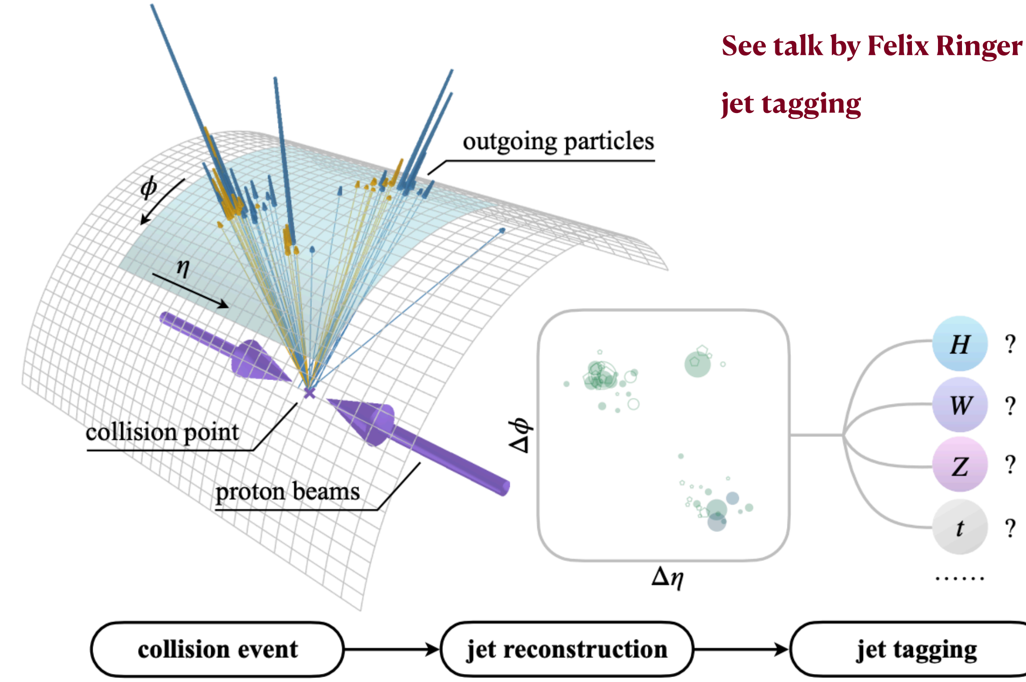
Dynamic Graph CNN for Learning on Point Clouds

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

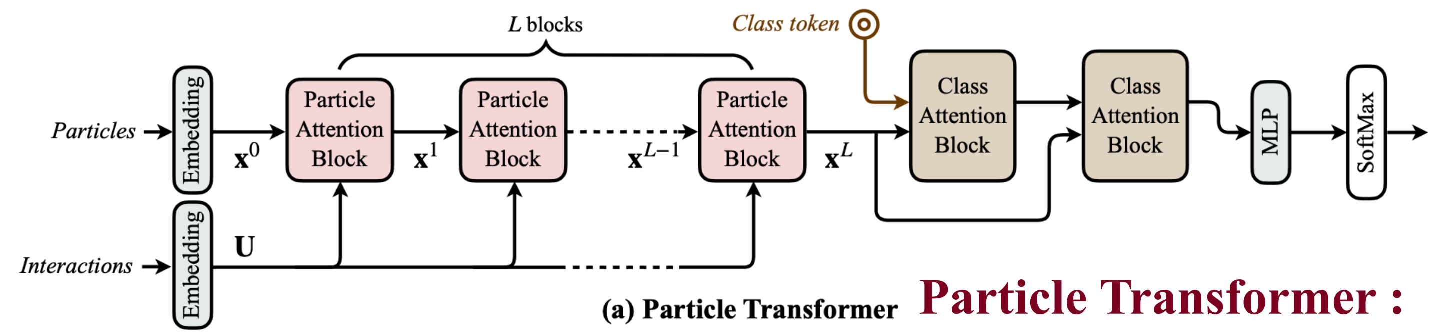
Object tagging

See talk by Felix Ringer for ML based EIC jet tagging



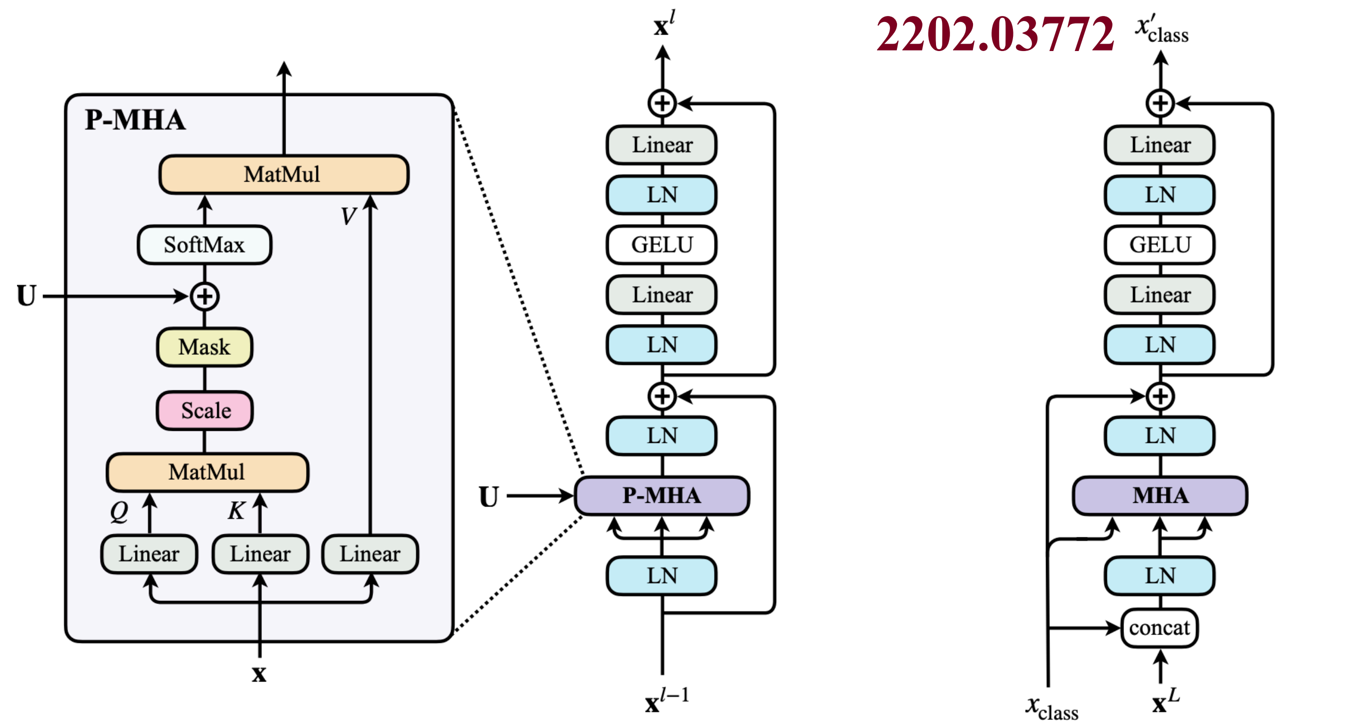


	Accuracy	AUC	$\mathrm{Rej}_{50\%}$	$\mathrm{Rej}_{30\%}$
P-CNN	0.930	0.9803	201 ± 4	759 ± 24
PFN	—	0.9819	247 ± 3	888 ± 17
ParticleNet	0.940	0.9858	397 ± 7	1615 ± 93
JEDI-net (w/ $\sum O$)	0.930	0.9807	—	774.6
PCT	0.940	0.9855	392 ± 7	1533 ± 101
LGN	0.929	0.964	—	435 ± 95
rPCN	—	0.9845	364 ± 9	1642 ± 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**	$\mathbf{691 \pm 15}$	$\mathbf{2766 \pm 130}$



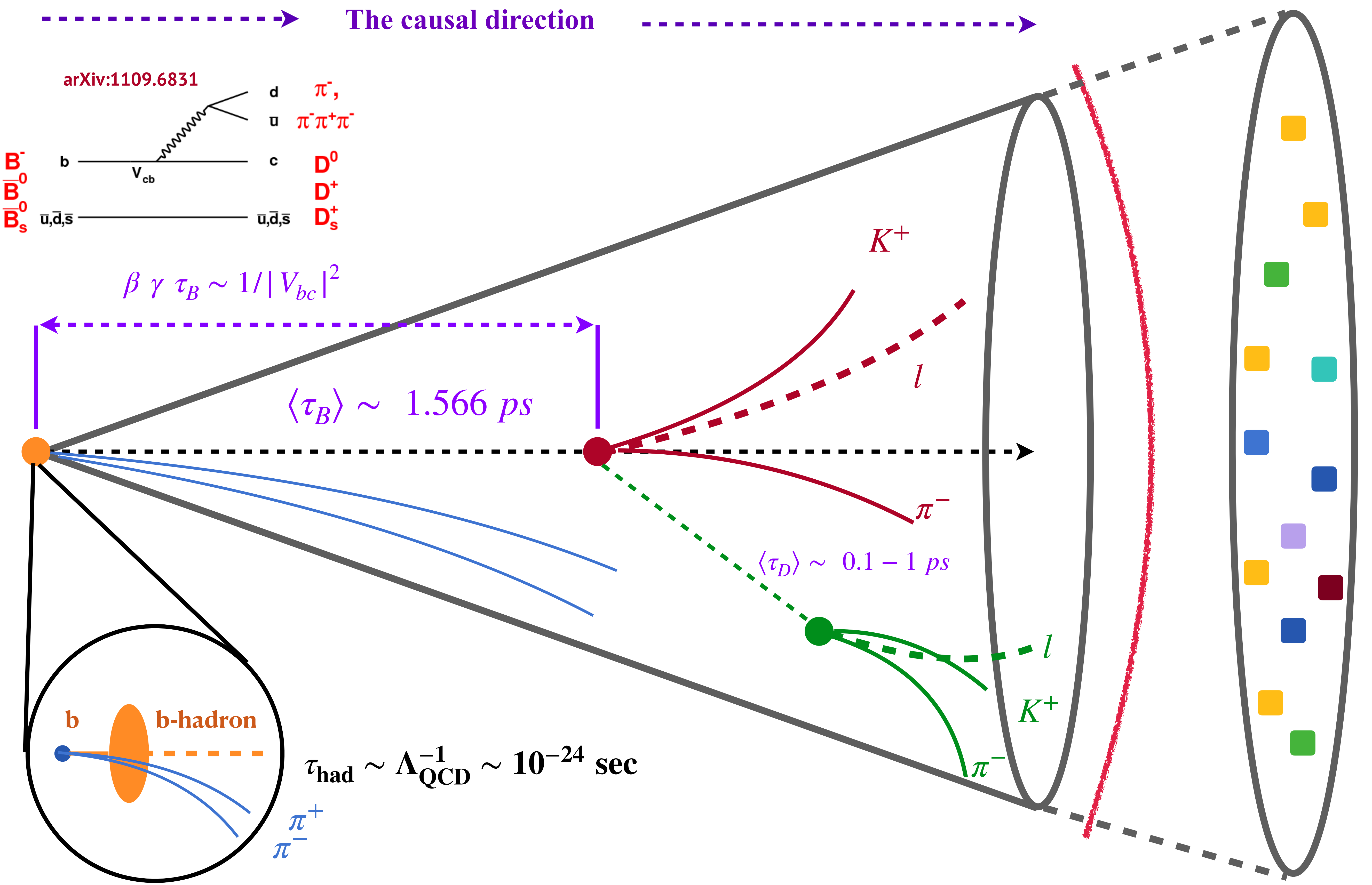
(a) Particle Transformer

Particle Transformer : 2202.03772

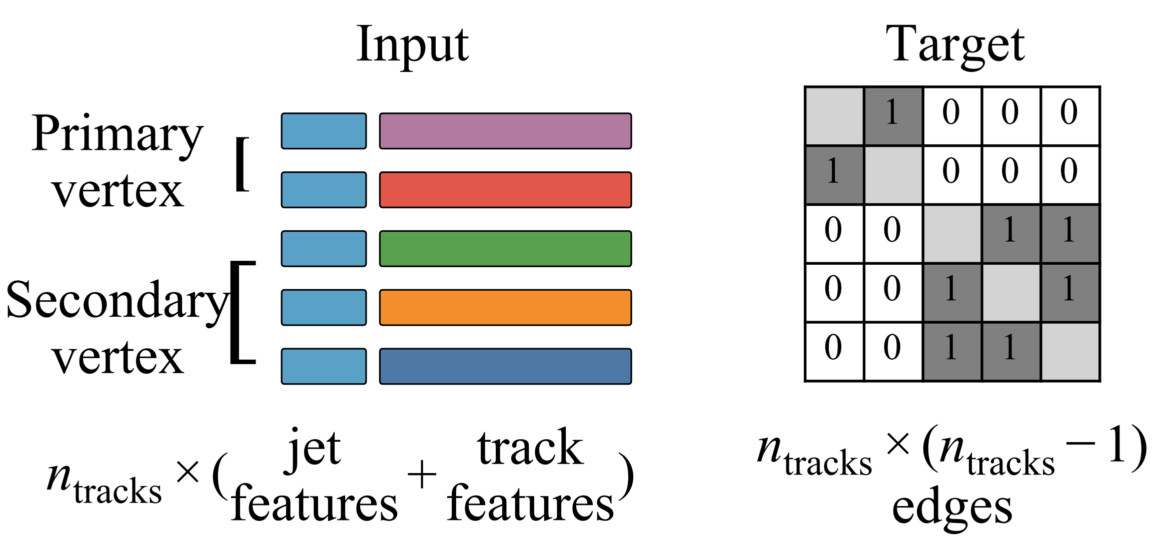


(b) Particle Attention Block (c) Class Attention Block

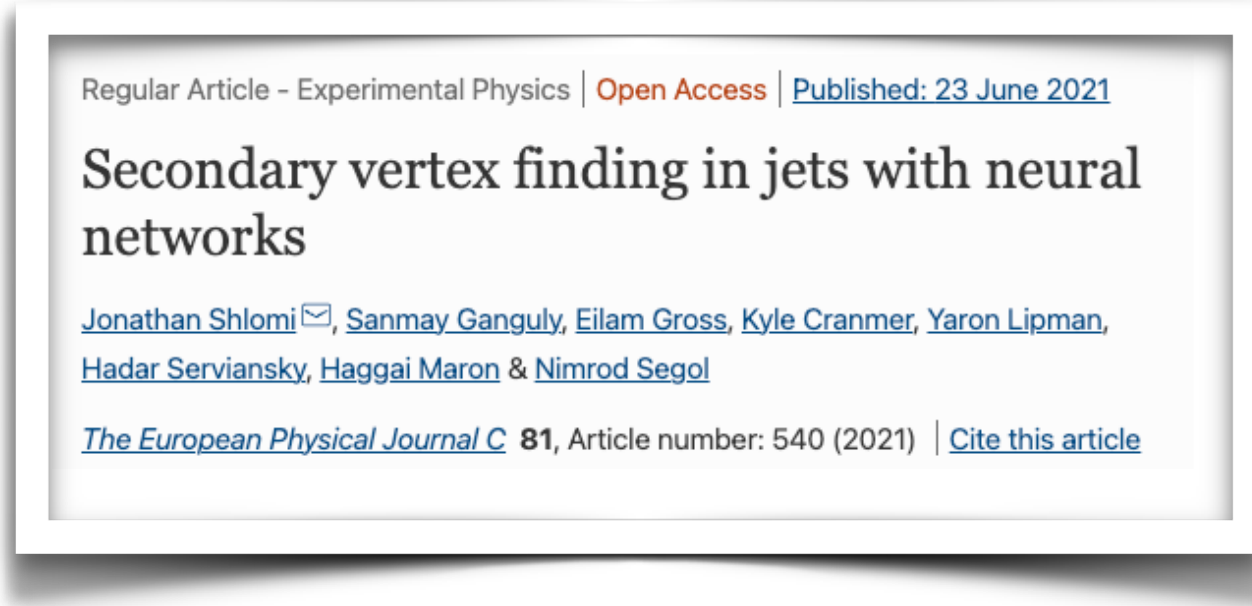
Anatomy of heavy-quark hadronization



Set2Graph proposal for flavor-tagging



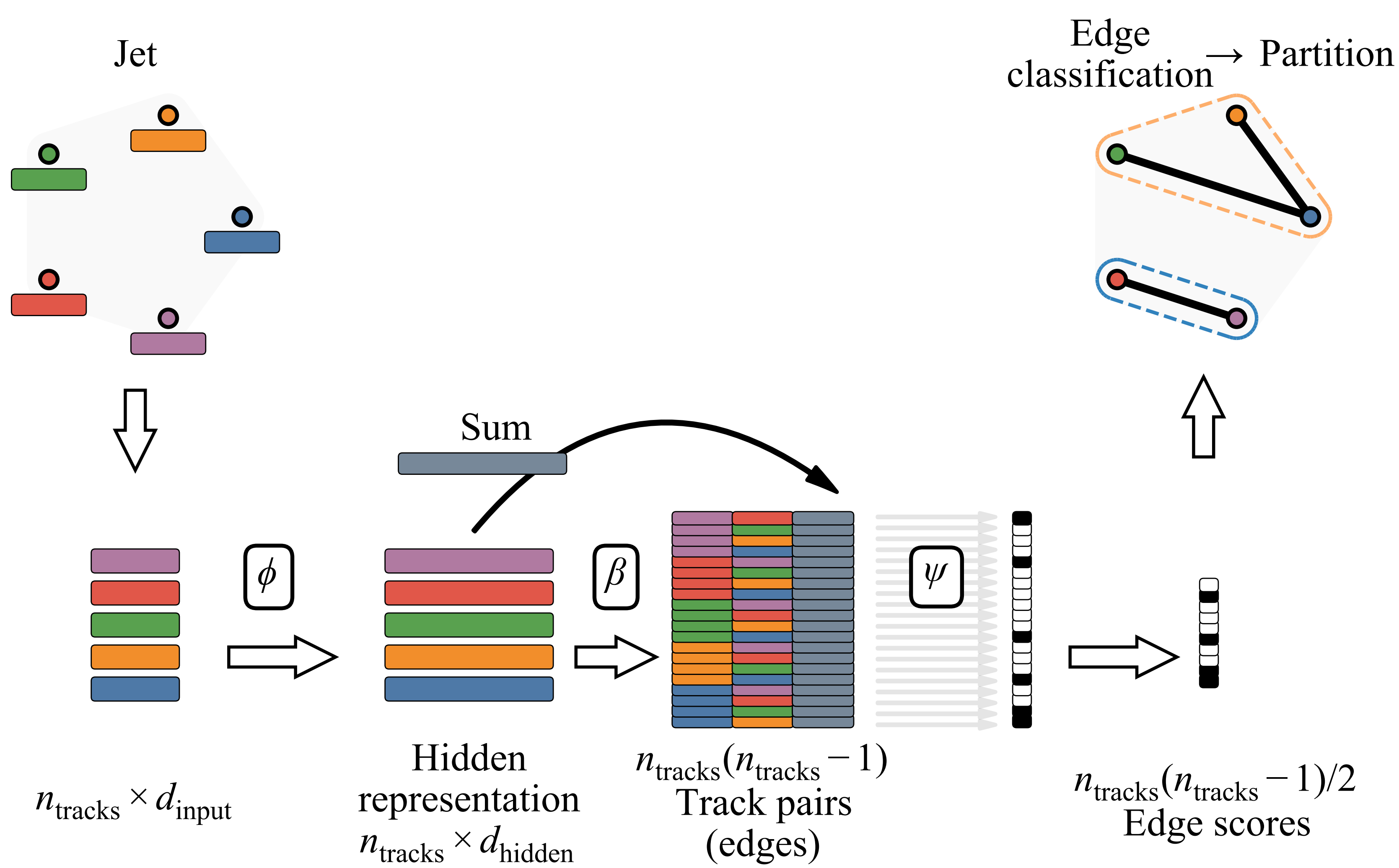
	1	0	0	0
1		0	0	0
0	0		1	1
0	0	1		1
0	0	1	1	


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

Secondary vertex finding in jets with neural networks

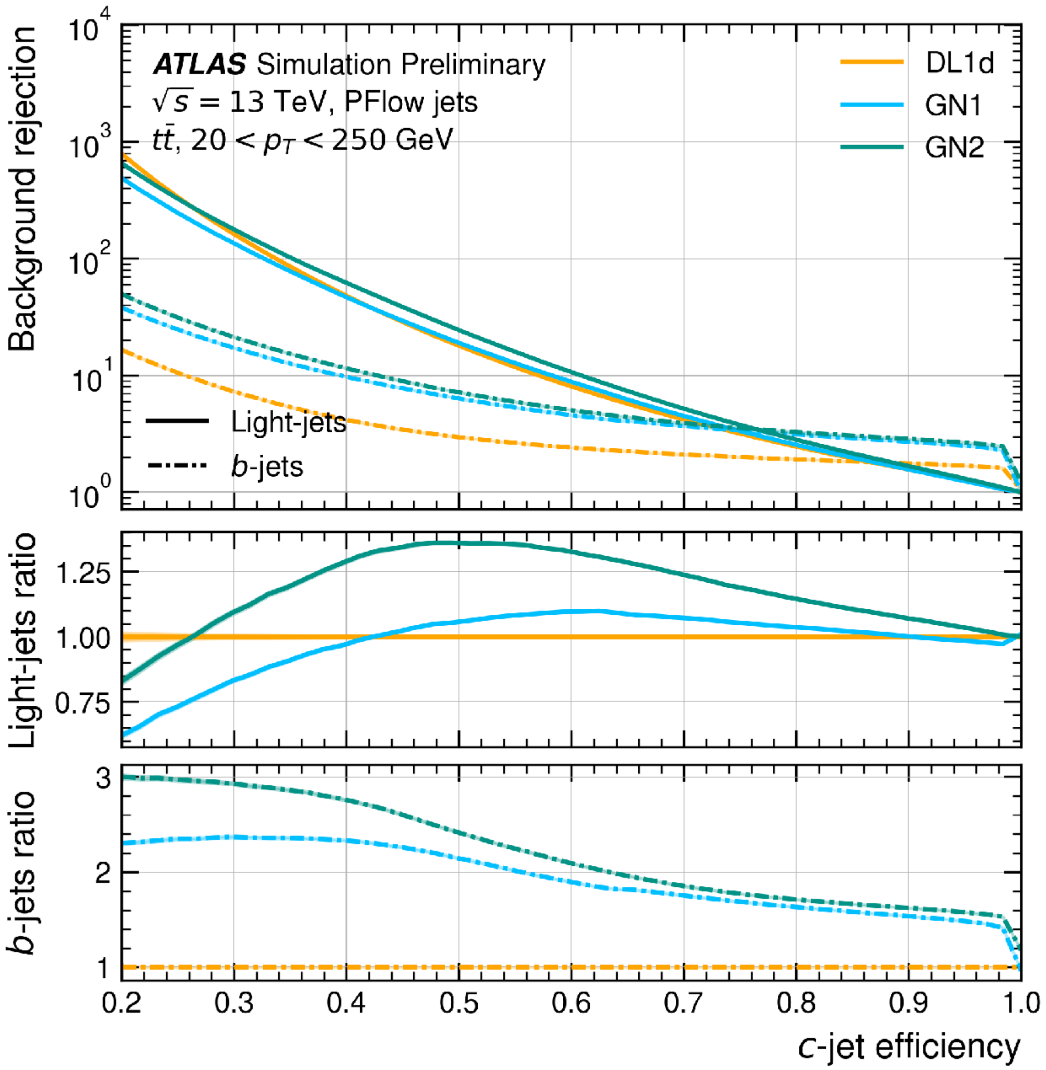
Jonathan Shlomi, Sanmay Ganguly, Eilam Gross, Kyle Cranmer, Yaron Lipman, Hadar Serviansky, Haggai Maron & Nimrod Segol

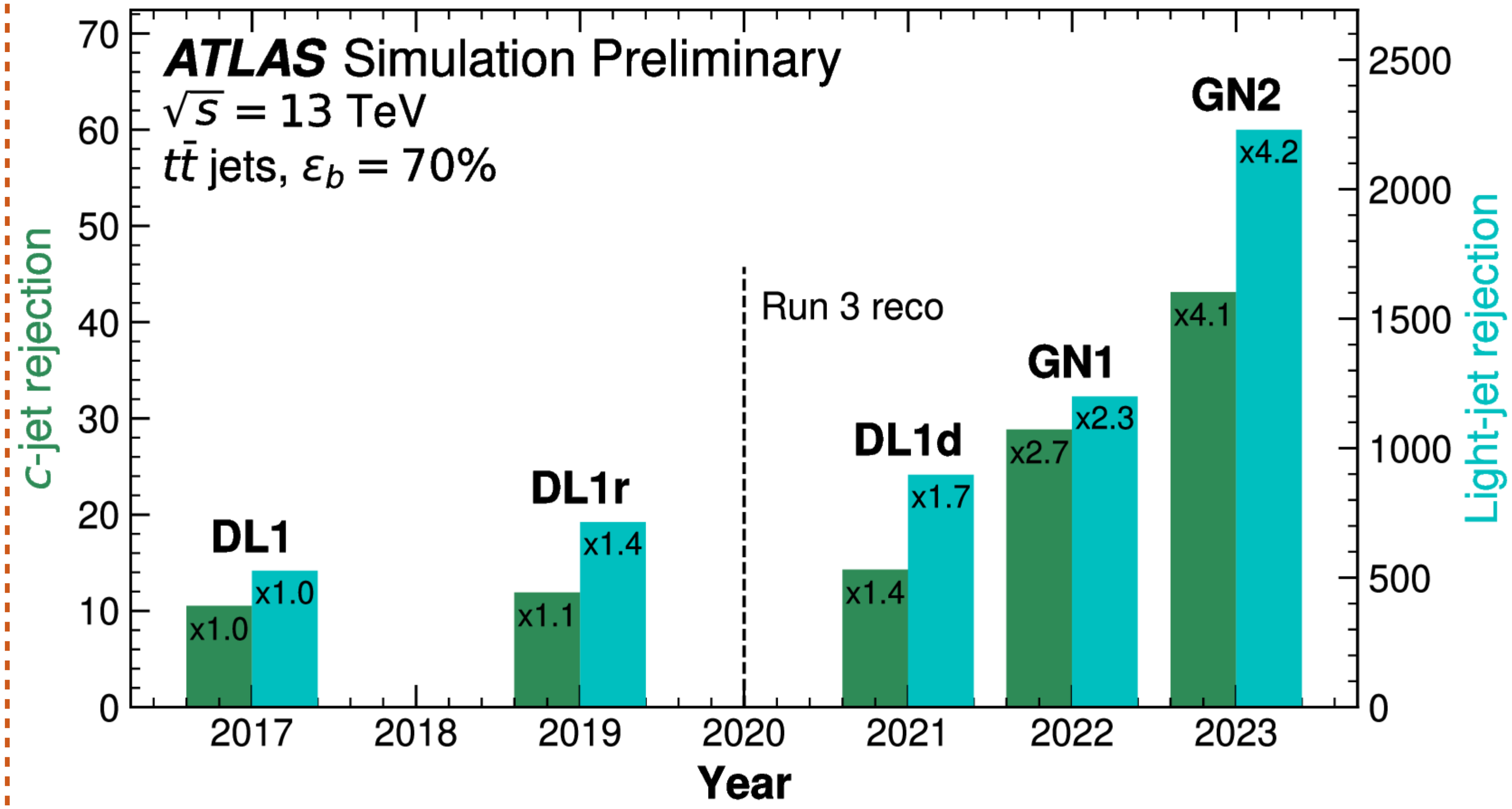
The European Physical Journal C **81**, Article number: 540 (2021) | Cite this article



Set2Graph model within ATLAS

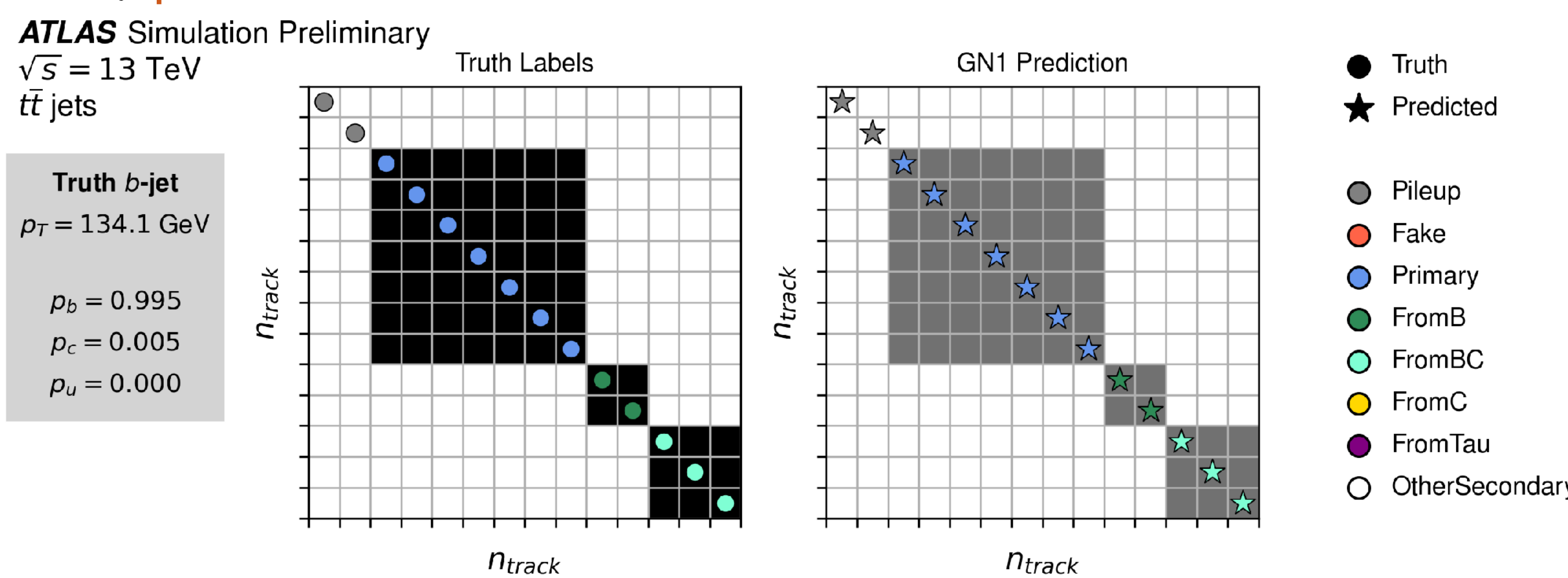
FTAG-2023-001





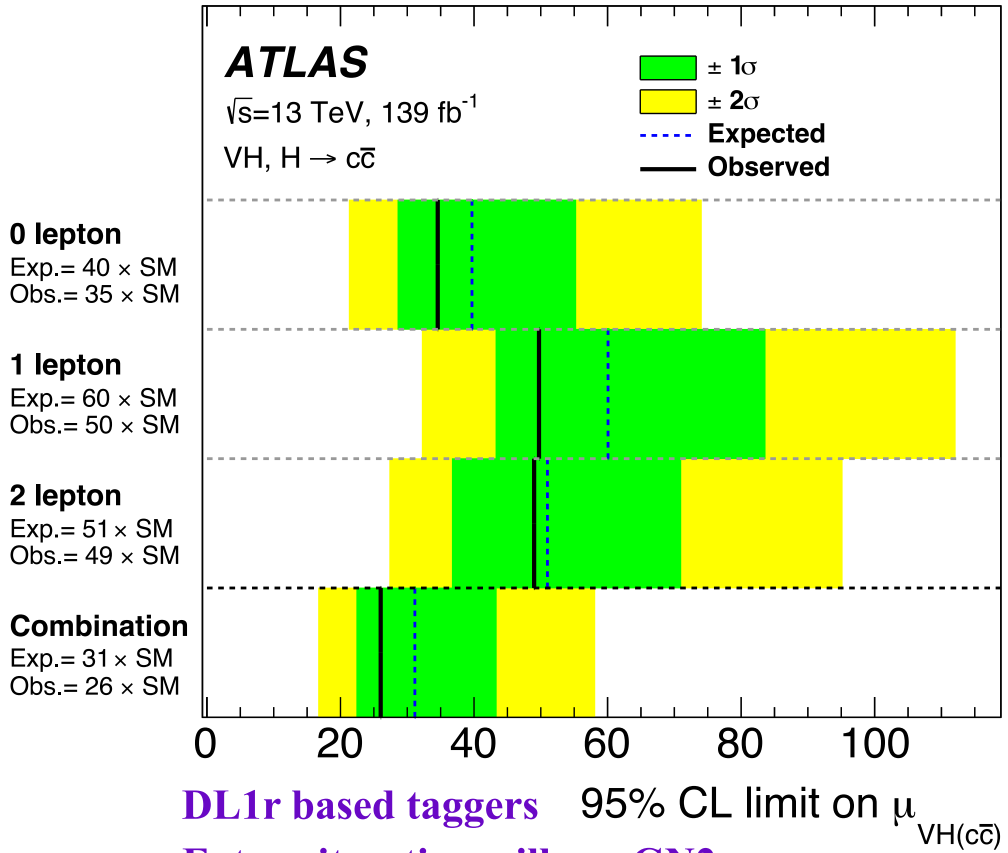
Sizable improvement over the current DL1r algorithm.

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



Direct physics application of the taggers

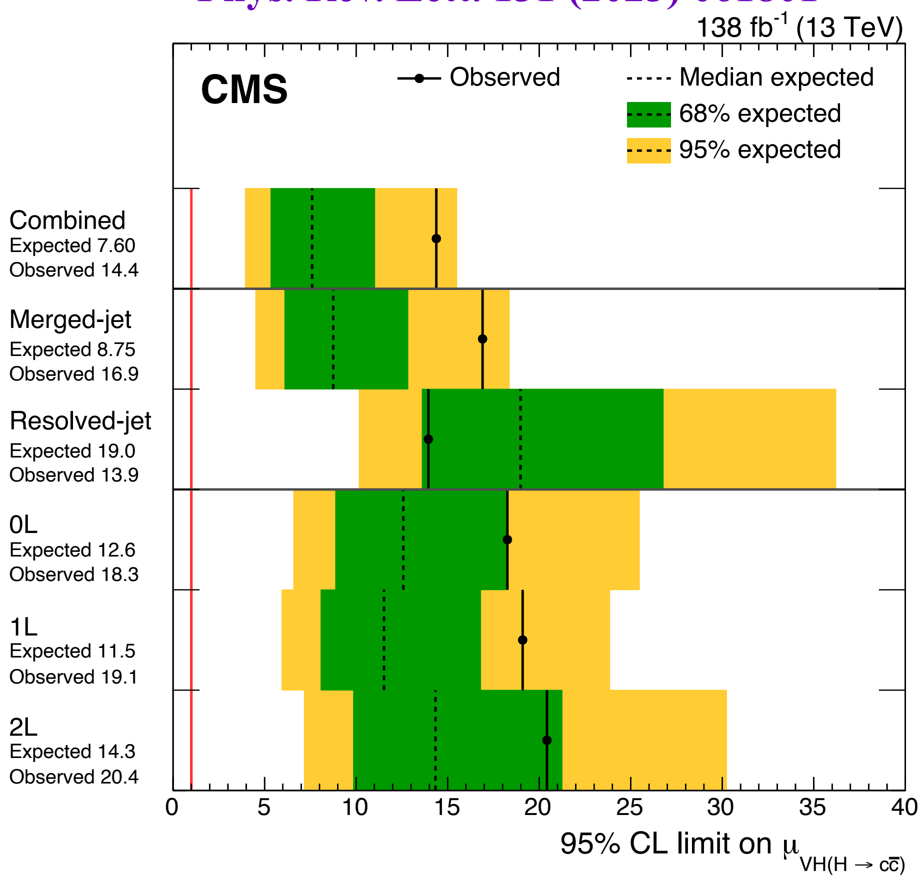
Eur. Phys. J. C (2022) 82:717



DL1r based taggers

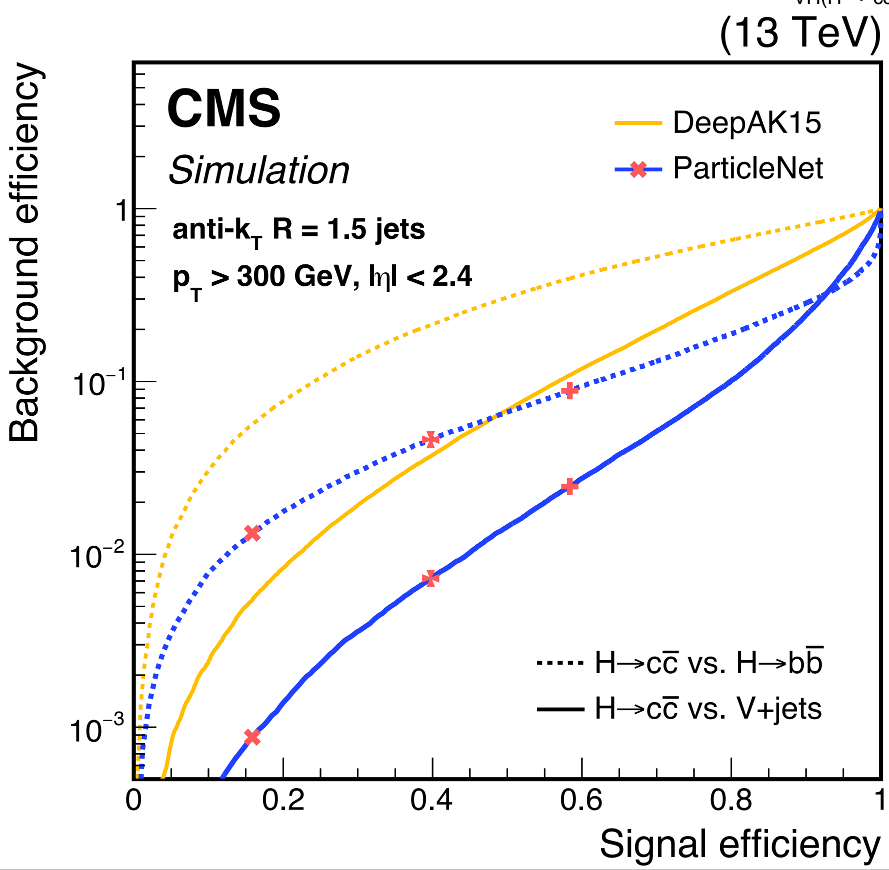
Future iteration will use GN2

Phys. Rev. Lett. 131 (2023) 061801



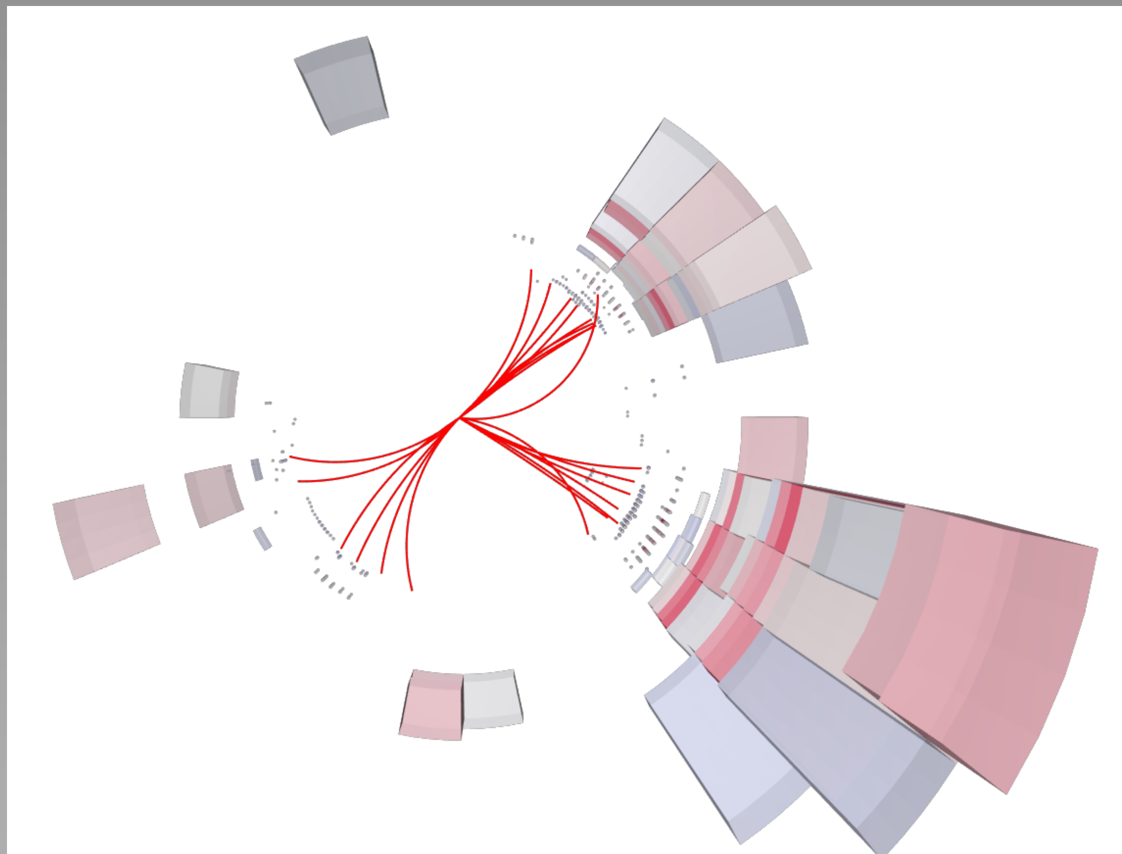
ATLAS bound : $|\kappa_c| < 8.5$

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

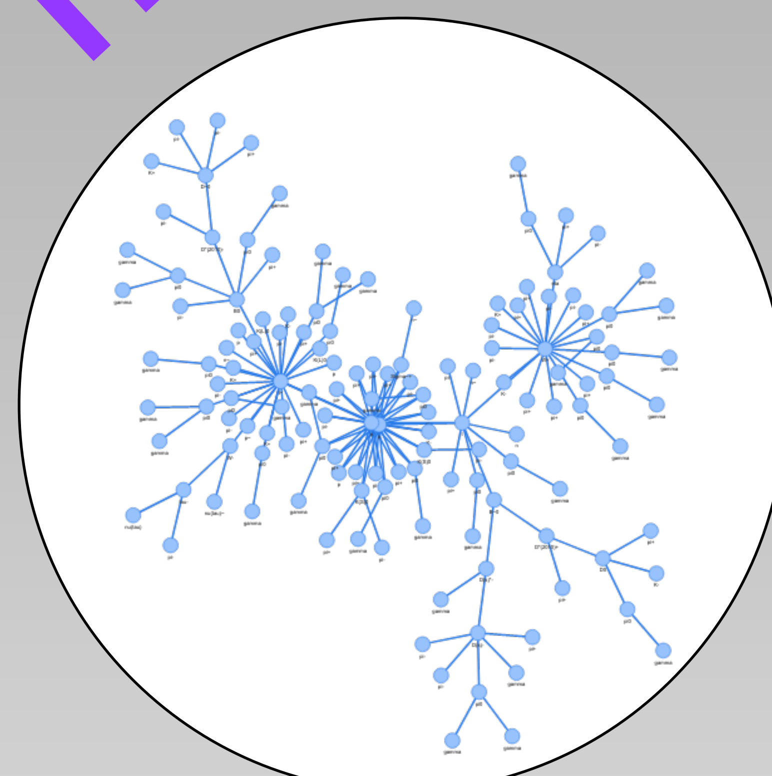


Future direction of tagger improvement:

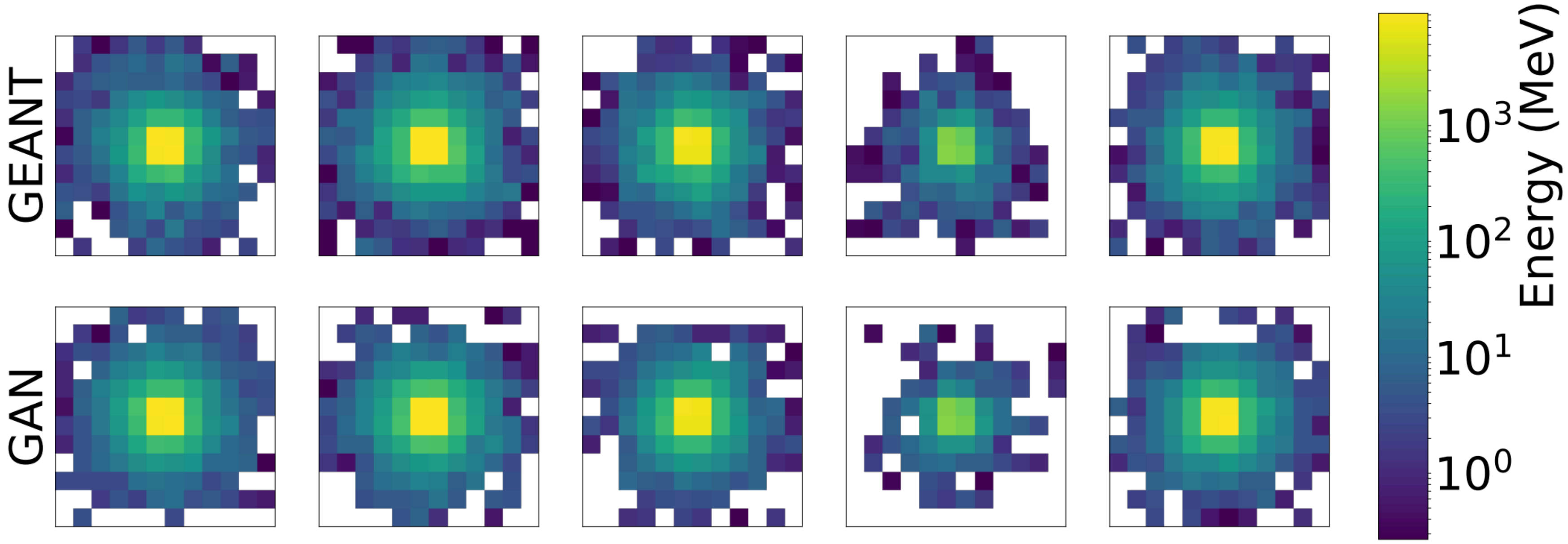
1. Explainable taggers on heterogeneous pc
2. A systematic uncertainty extraction.
3. How much universal taggers can be made across topologies?

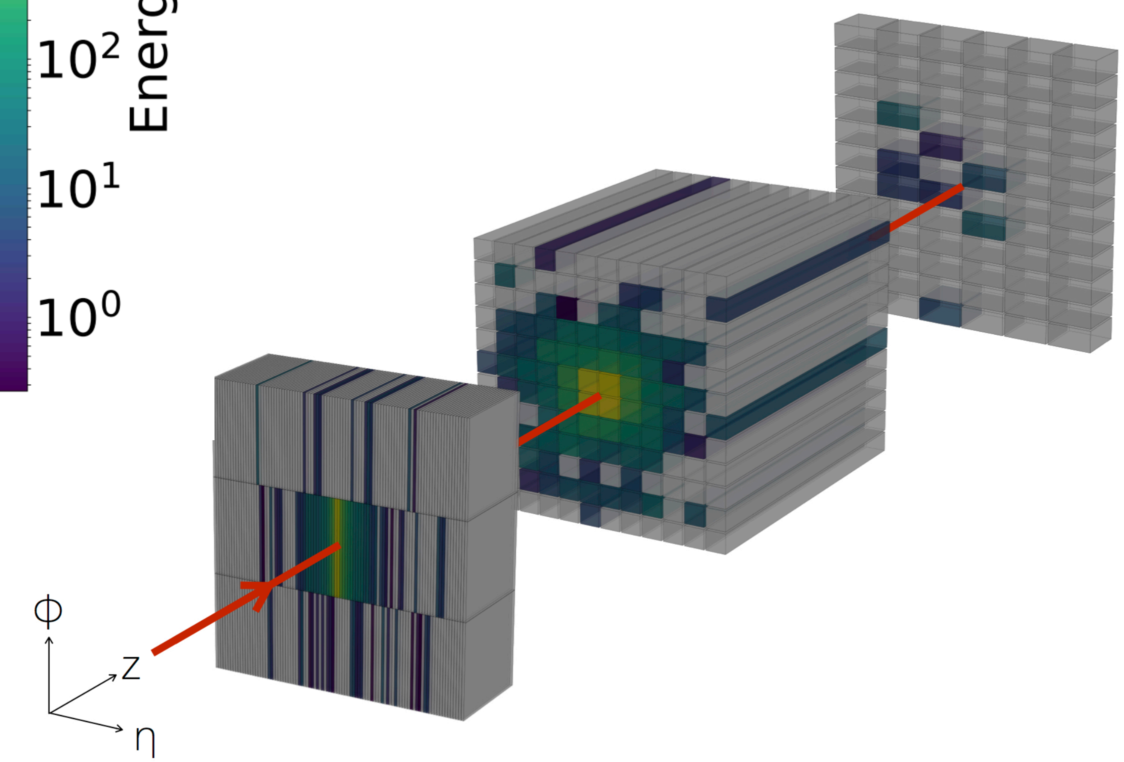


Generative models for calorimeter simulations

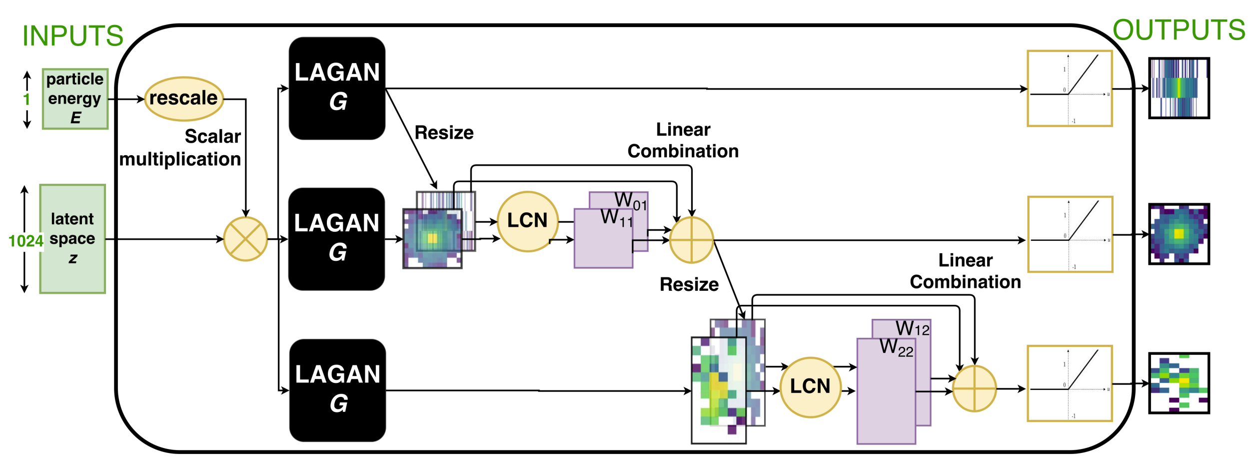
$f_{\{\theta\}}($  $, z)$

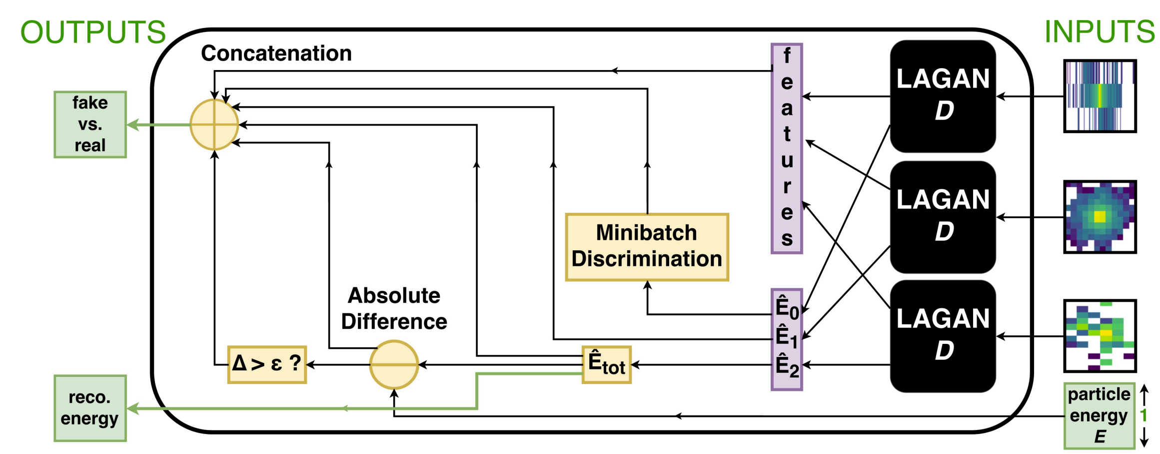
Detector simulation using ML





CaloGAN 1705.02355
Michela Paganini, Luke de Oliveira, Benjamin Nachman

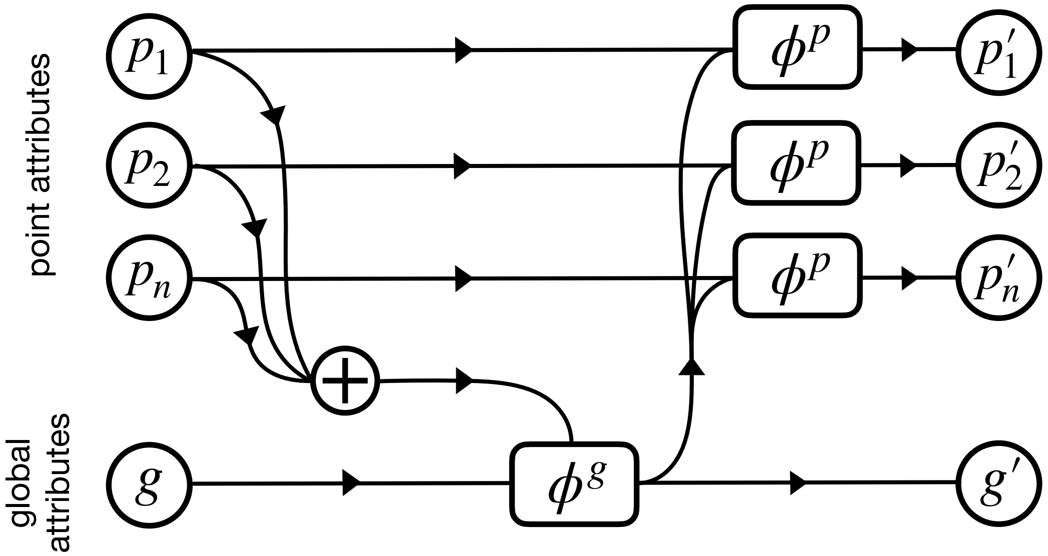


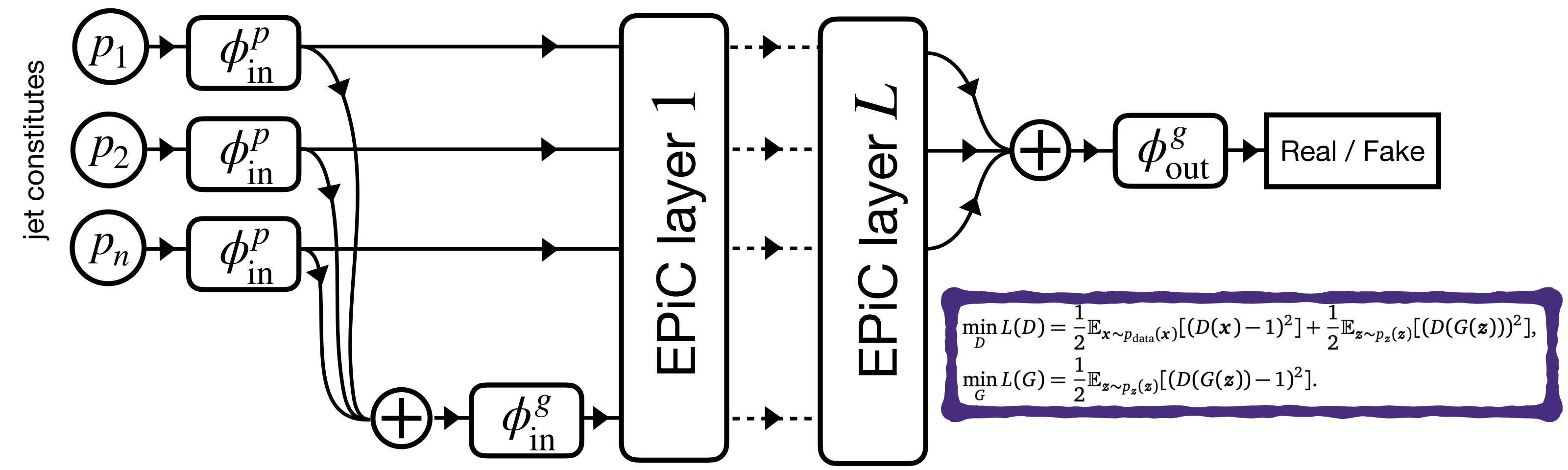


Generator

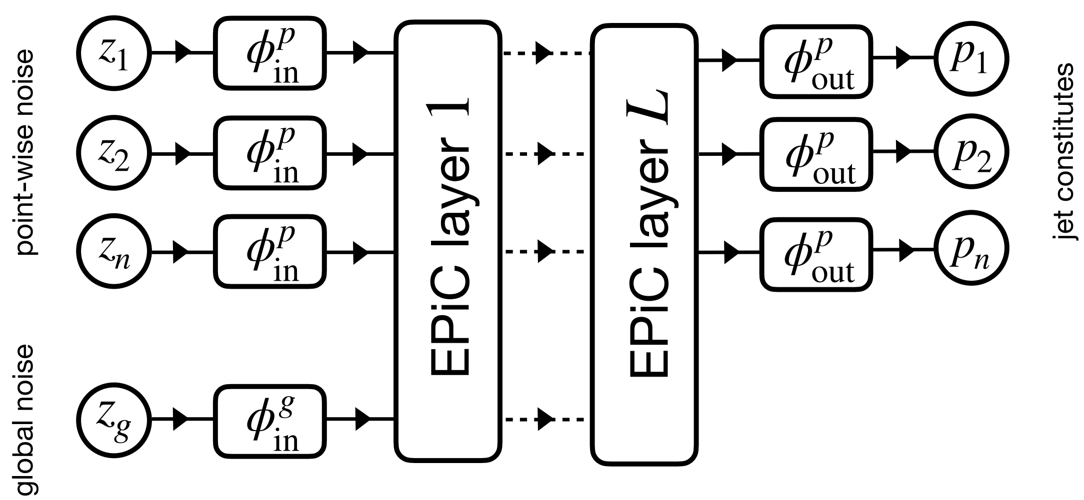
Discriminator

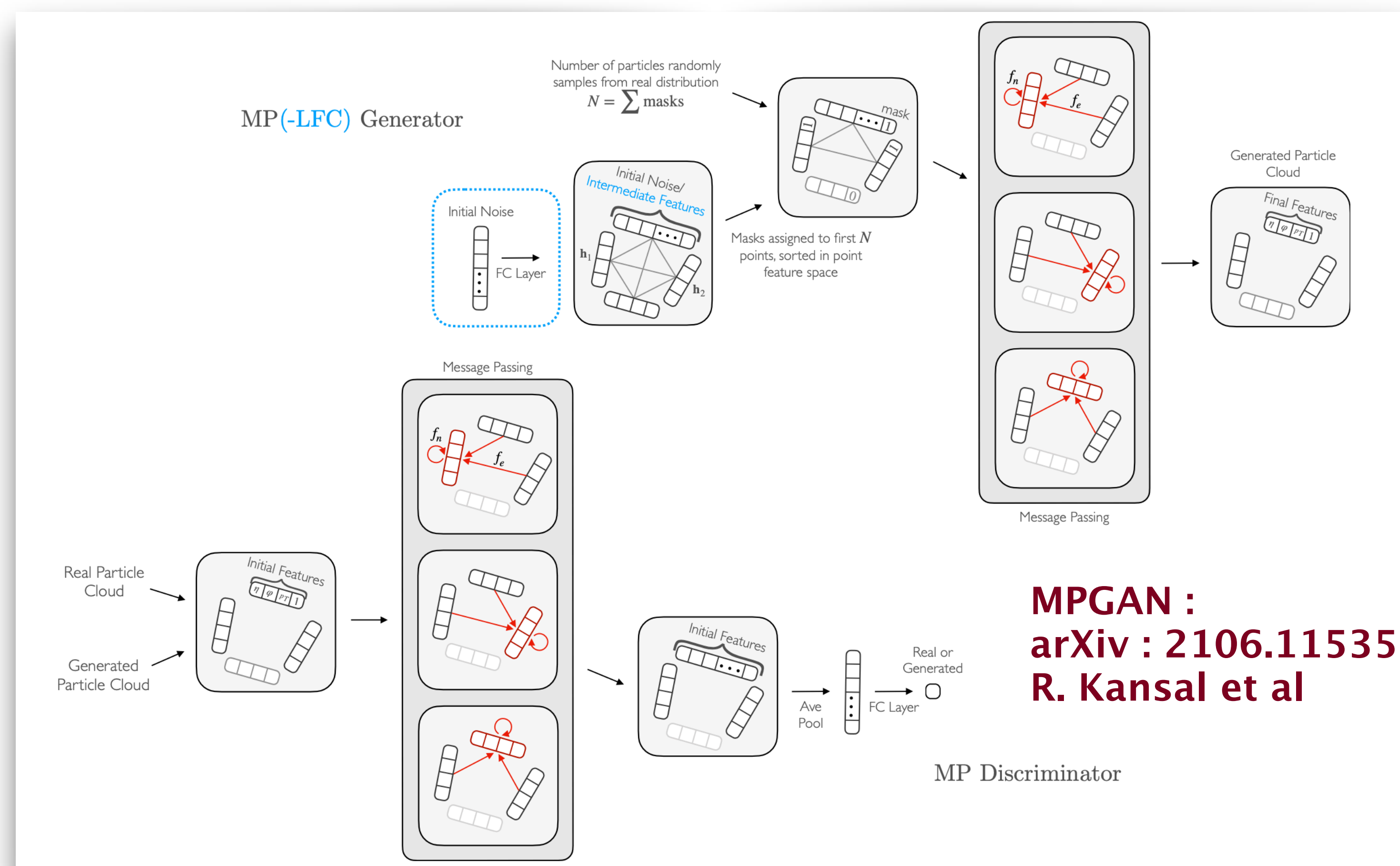
Detector simulation using ML





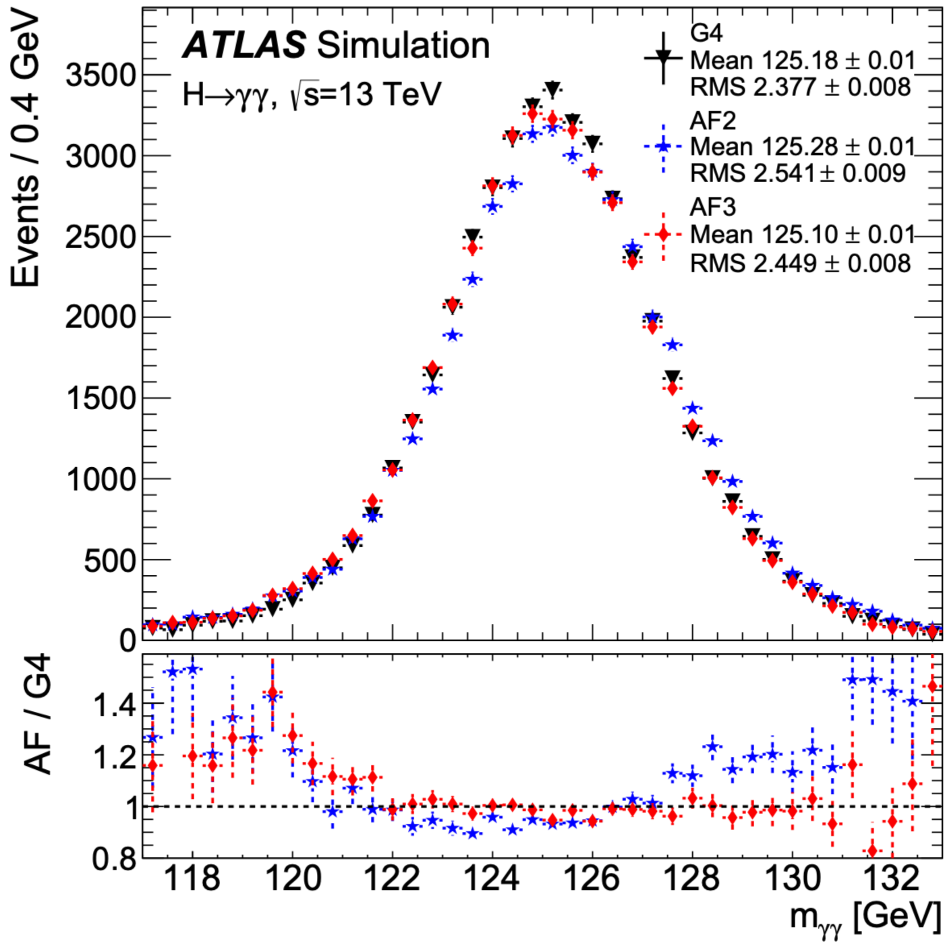
EPiC-GAN : SciPost Phys. 15, 130 (2023) Erik Buhmann, Gregor Kasieczka, Jesse Thaler

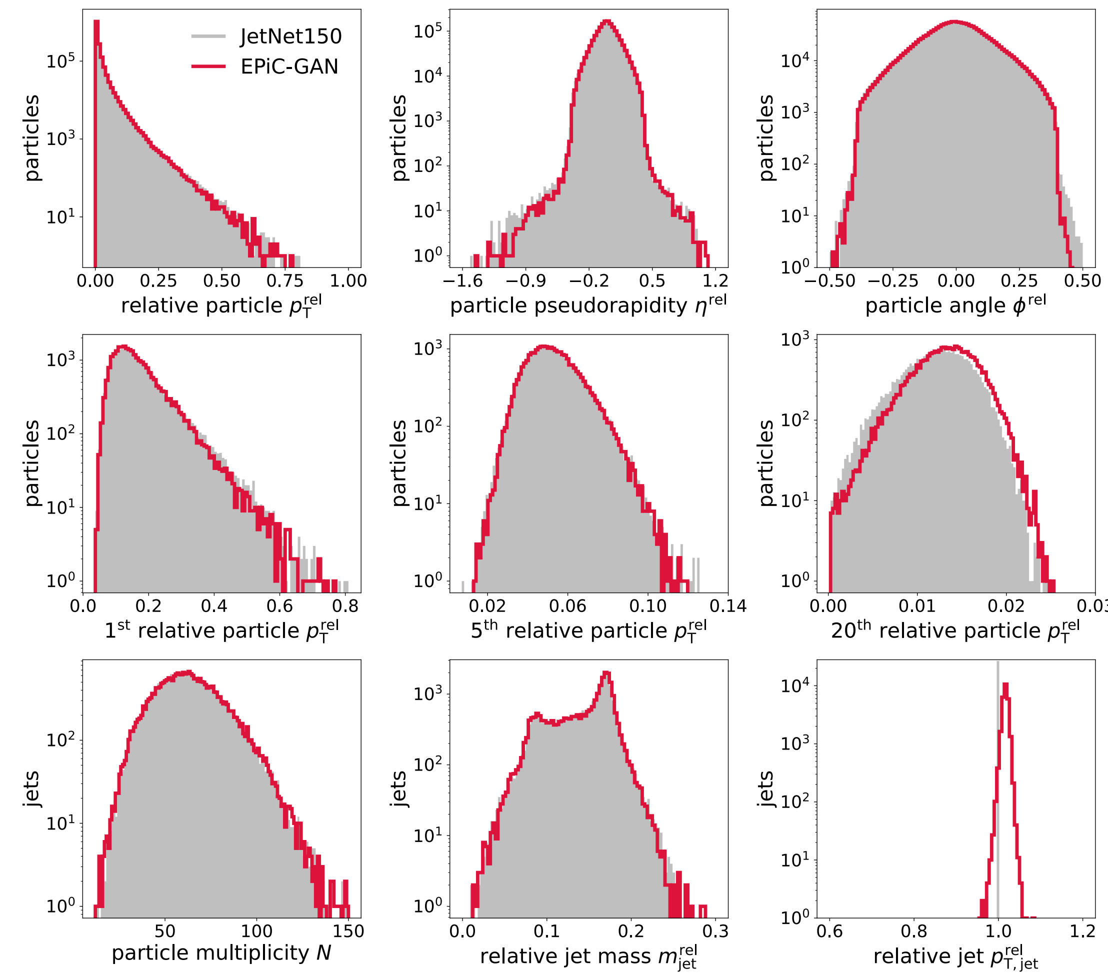




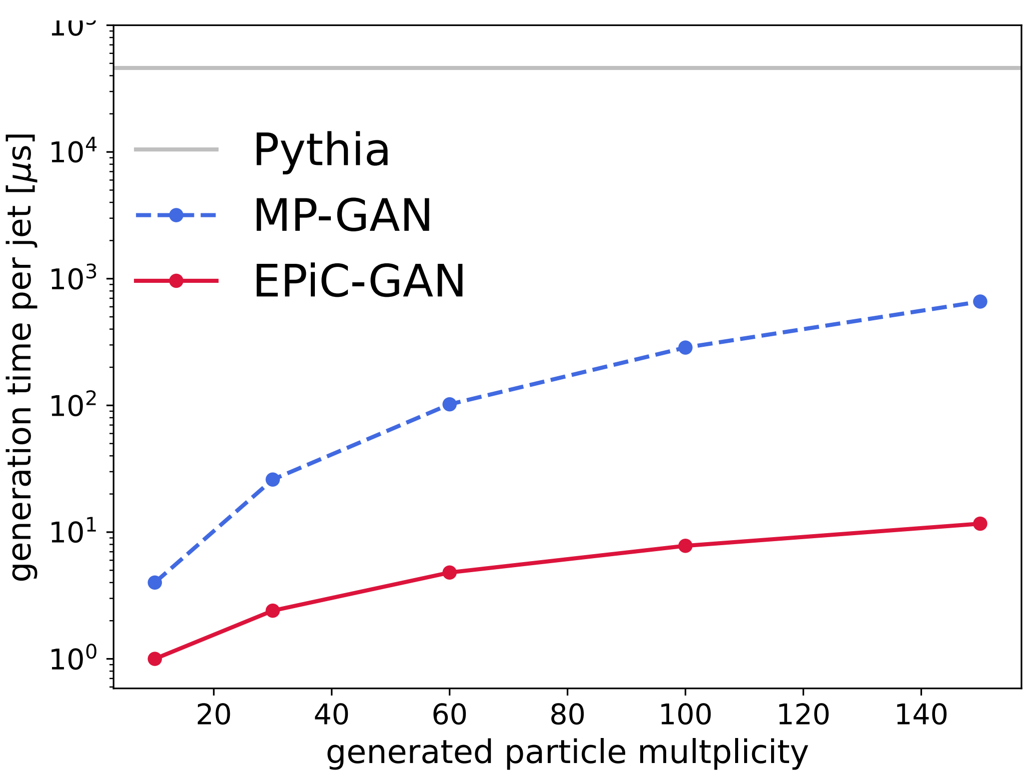
MPGAN : arXiv : 2106.11535 R. Kansal et al

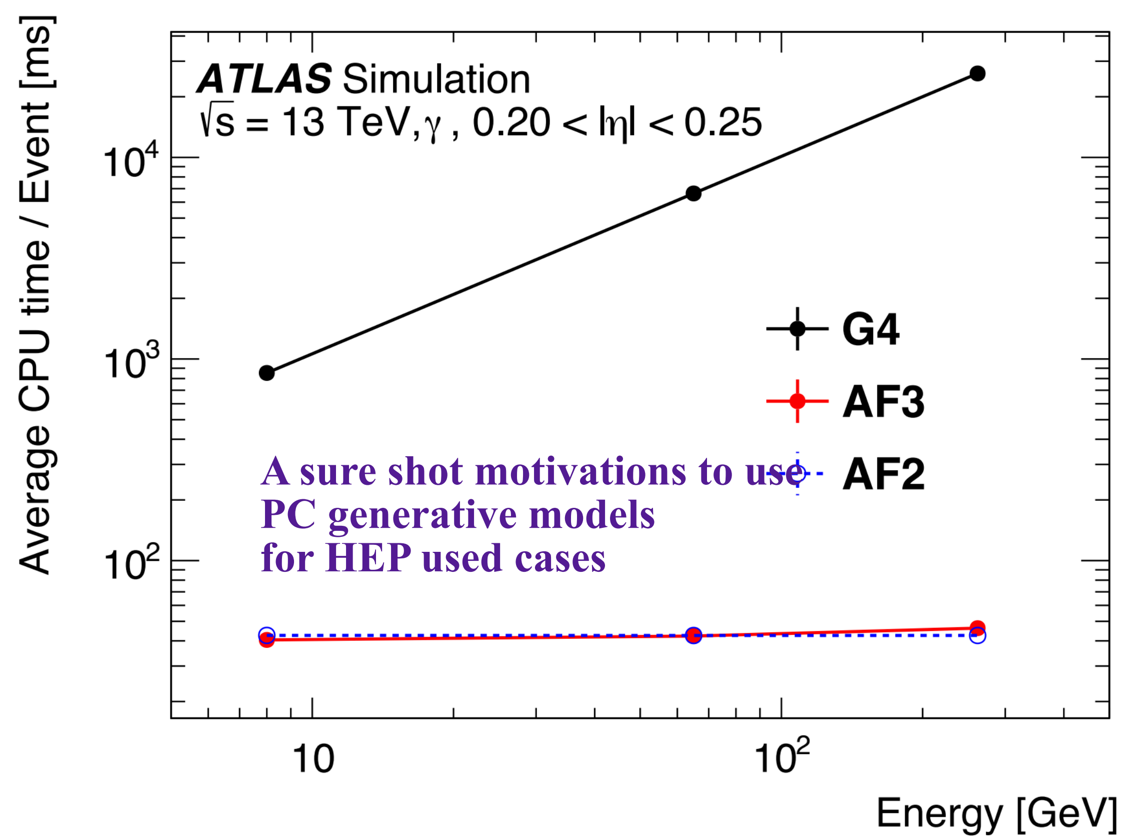
The major gain





arXiv : 2109.02551



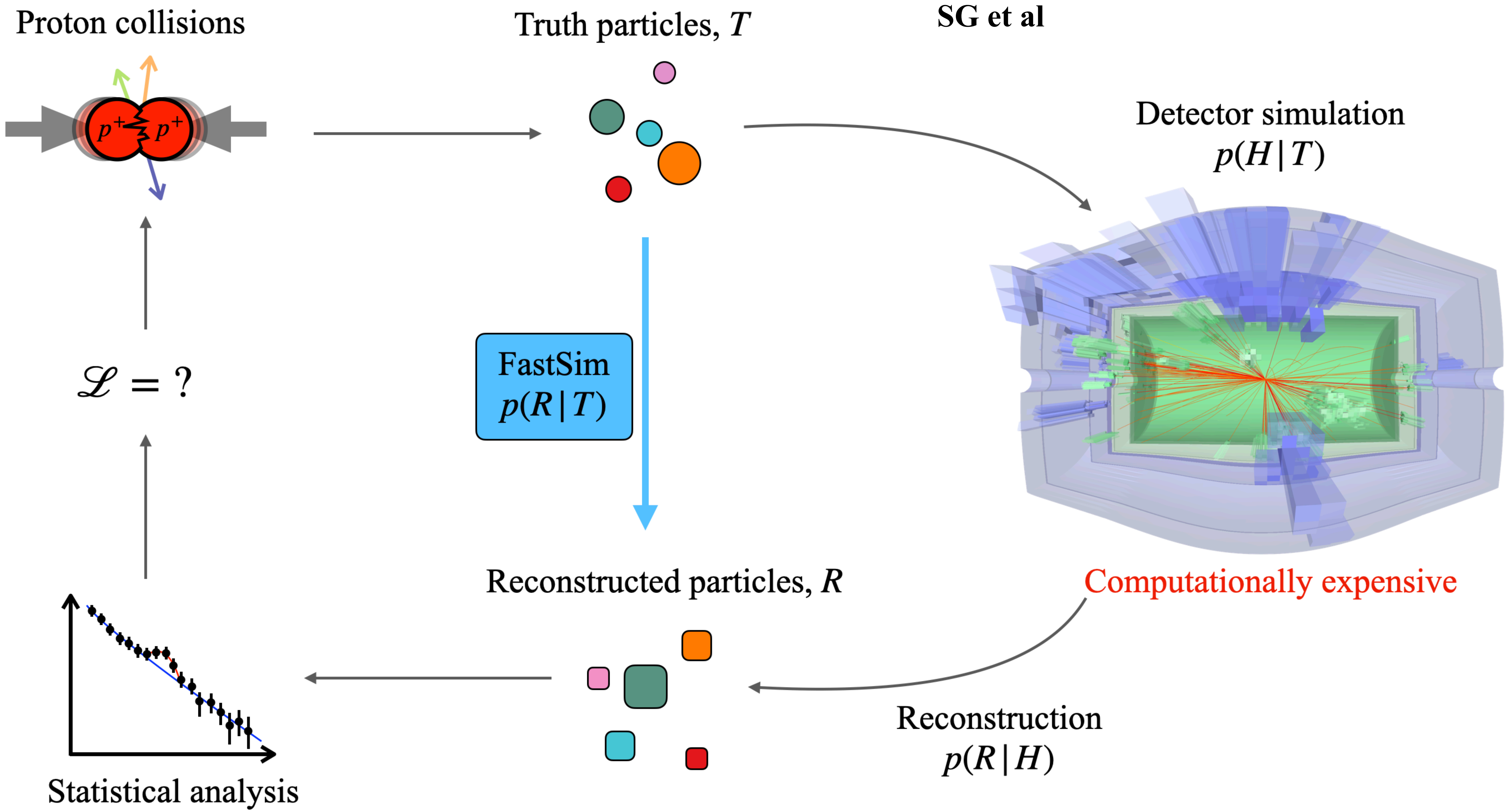


A sure shot motivations to use
PC generative models
for HEP used cases

A generative model for Particle-flow

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

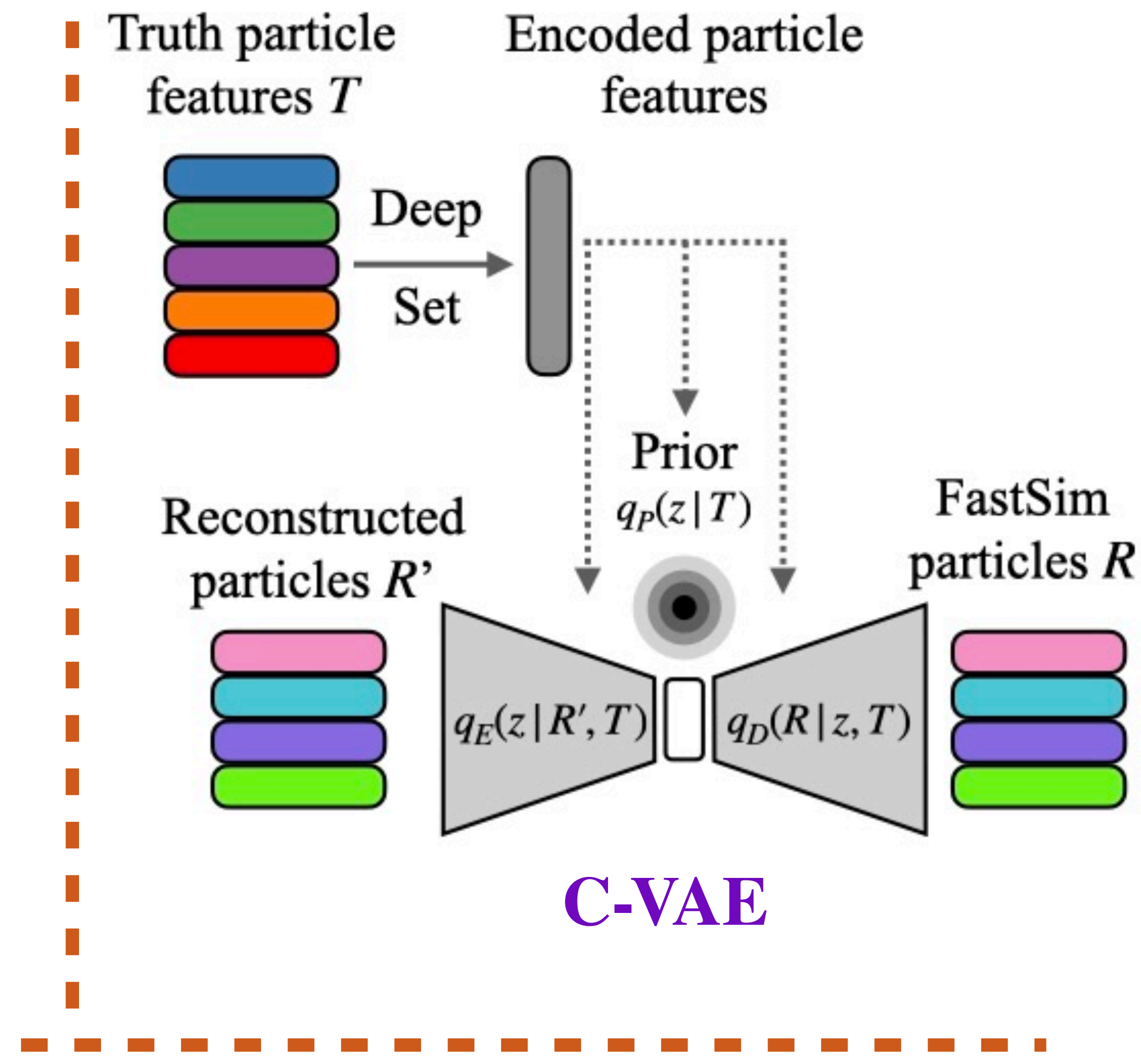
SG et al

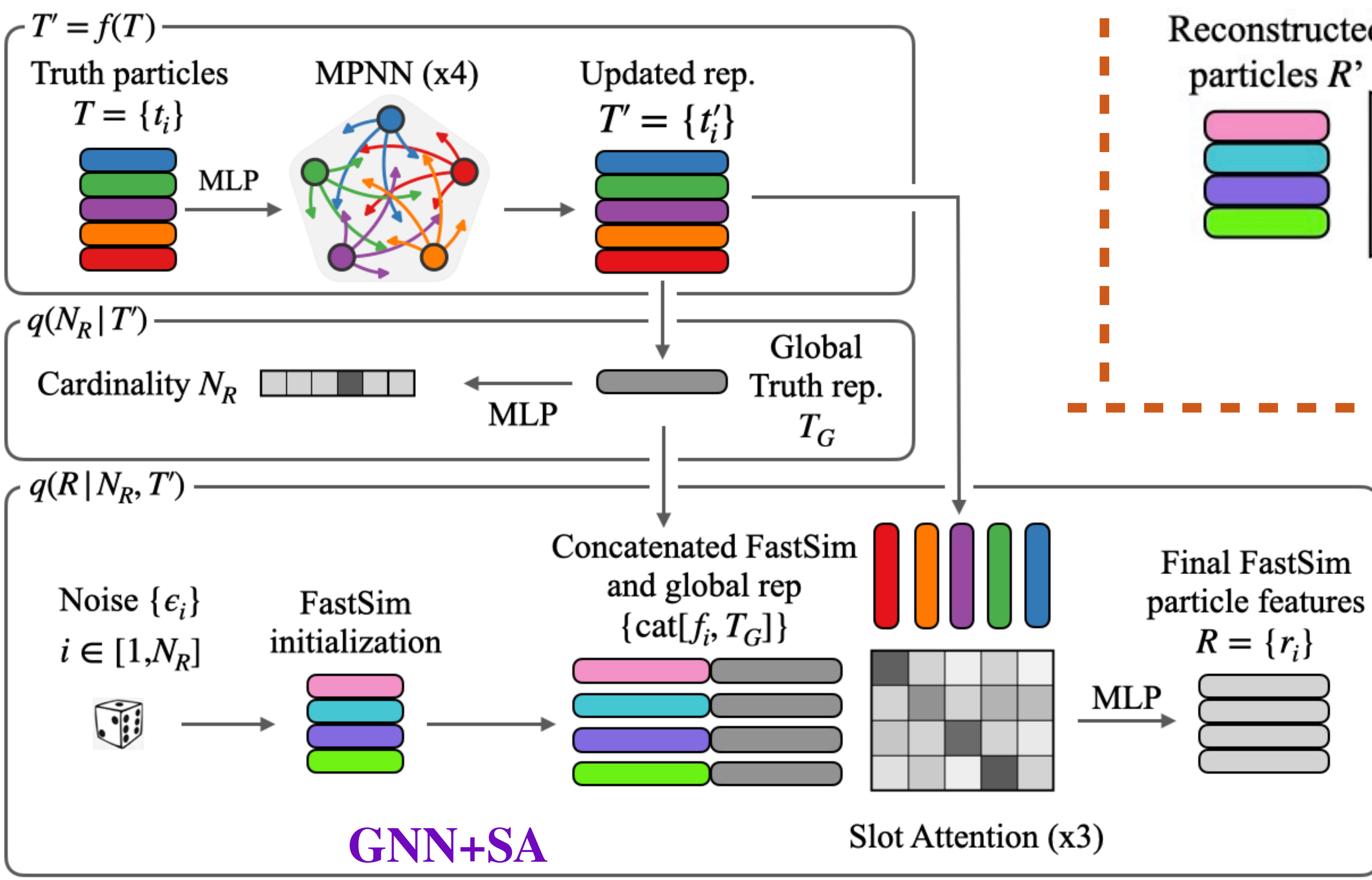


$$R \sim p(R|T) = \int \mathrm{d}H\, \delta(R(H) - R)\, p_{\mathrm{sim}}(H|T).$$

The task of constrained set generation

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





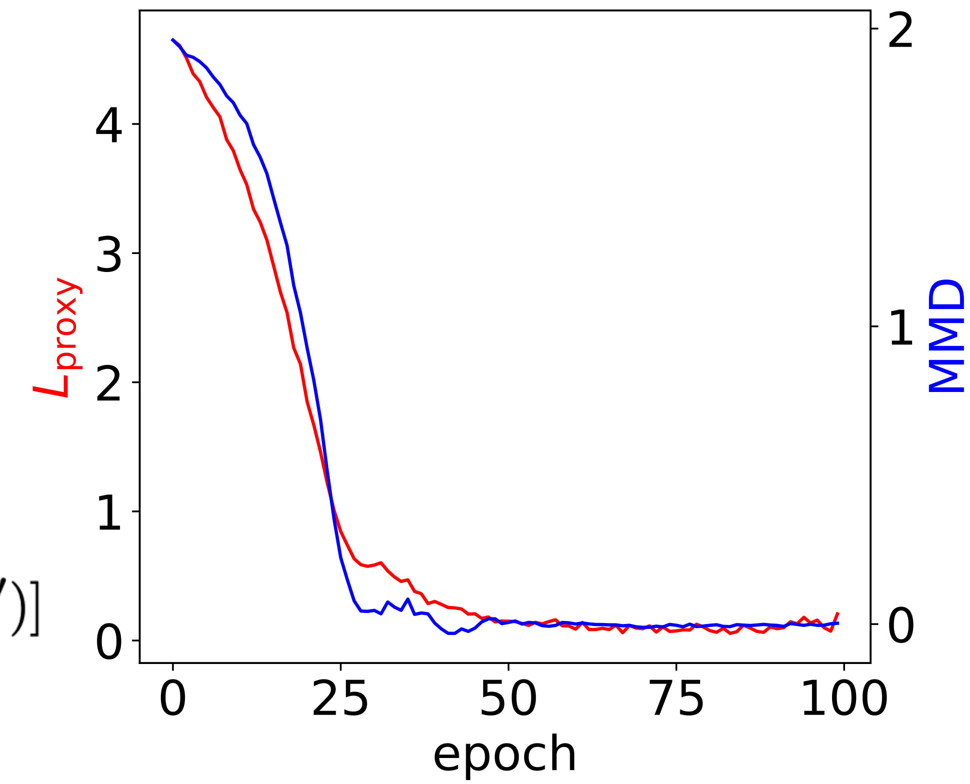
The task of constrained set generation

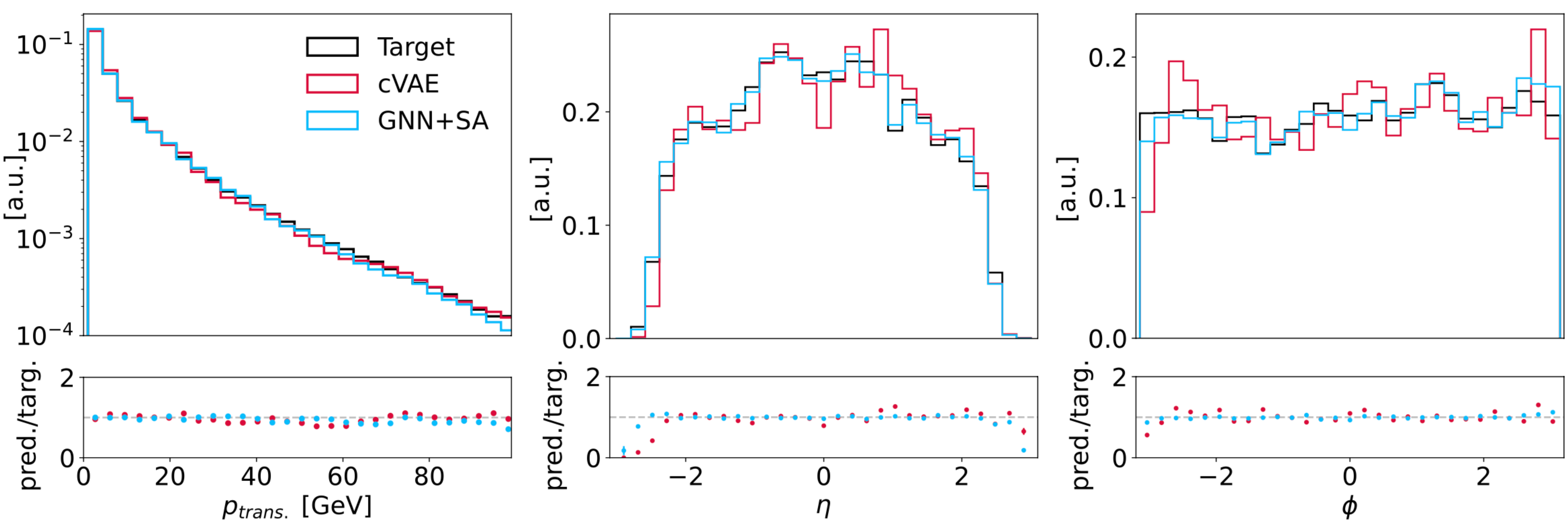
The cVAE training is done by optimizing negative evidence lower bound (ELBO) loss :

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

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

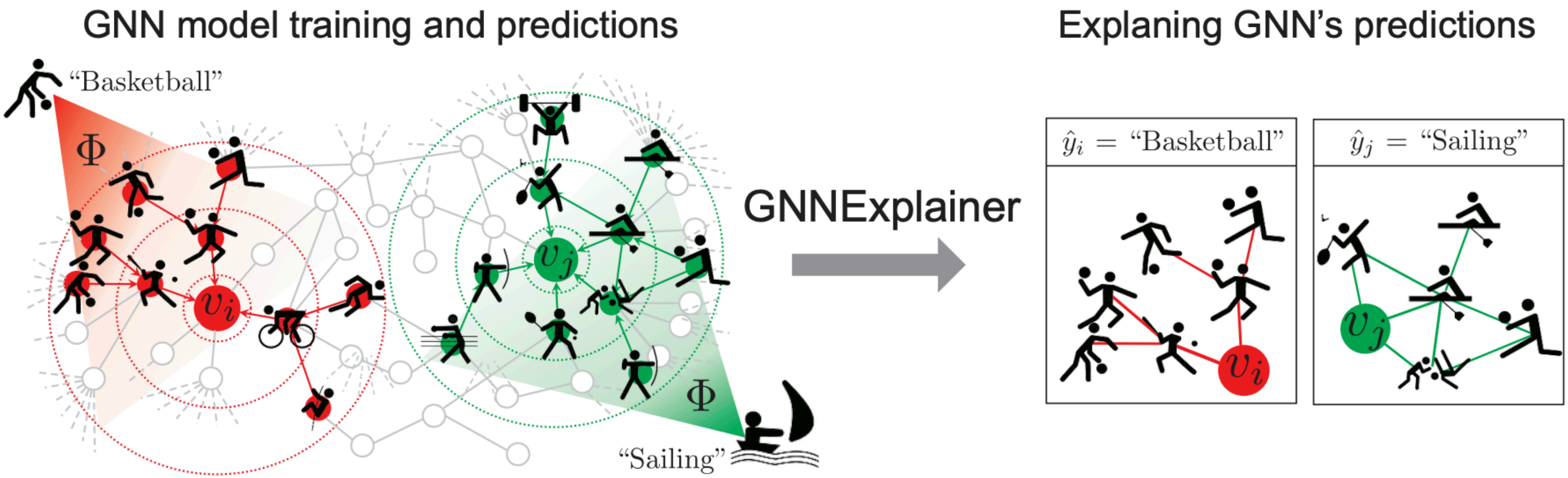
$$\mathrm{MMD}^2 = \mathbb{E}_{(x\sim p,x'\sim p)}[k(x,x')] + \mathbb{E}_{(x\sim q,x'\sim q)}[k(x,x')] - 2\mathbb{E}_{(x\sim q,x'\sim p)}[k(x,x')]$$





What's brewing now?

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

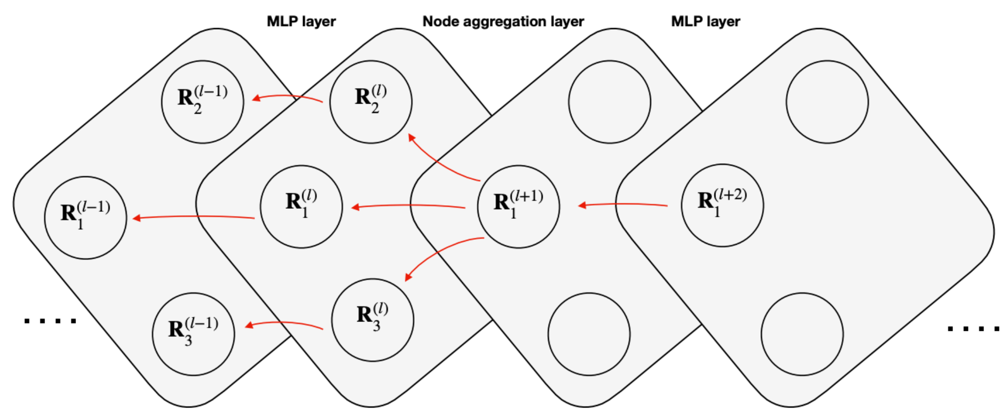
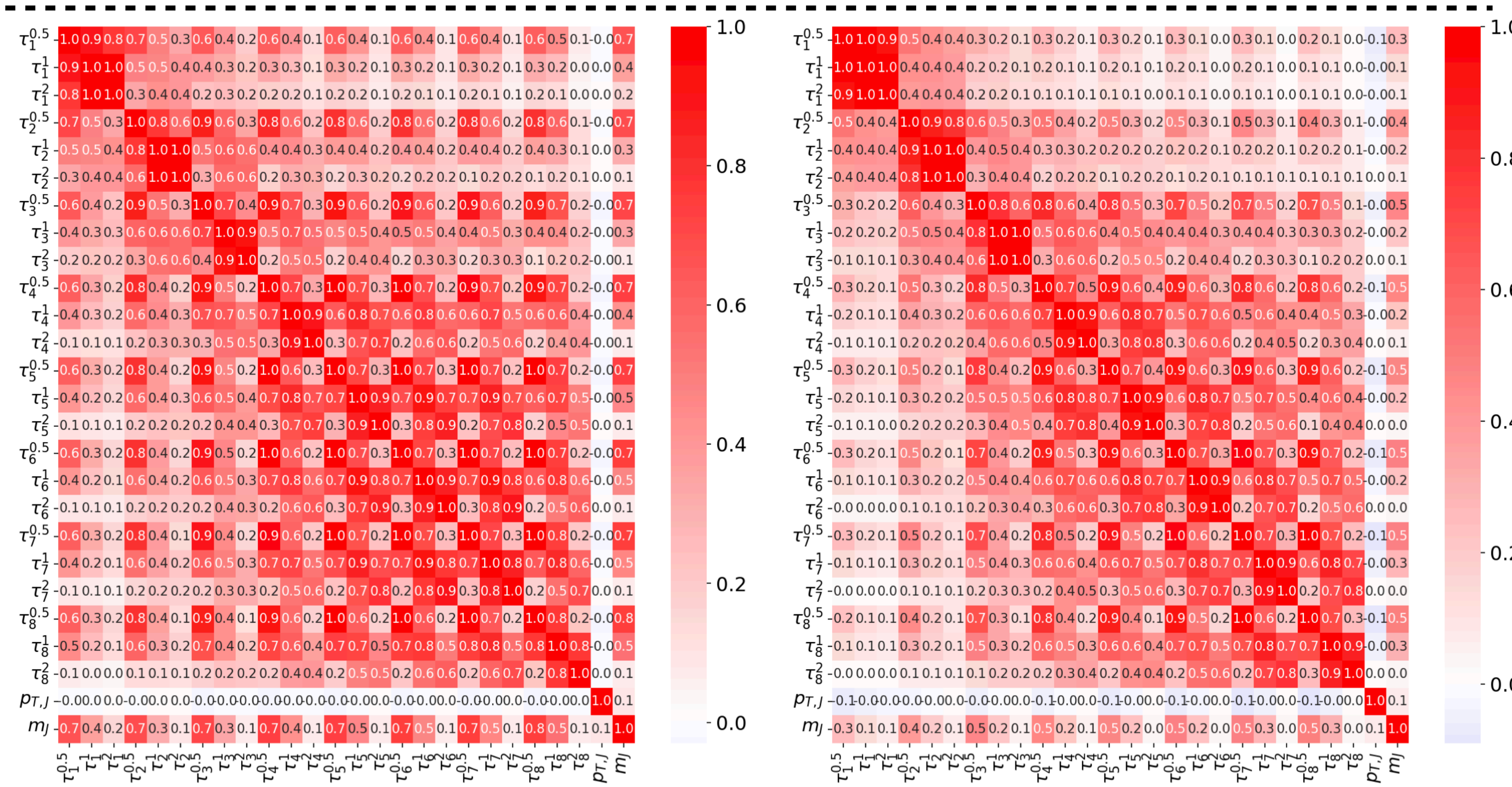


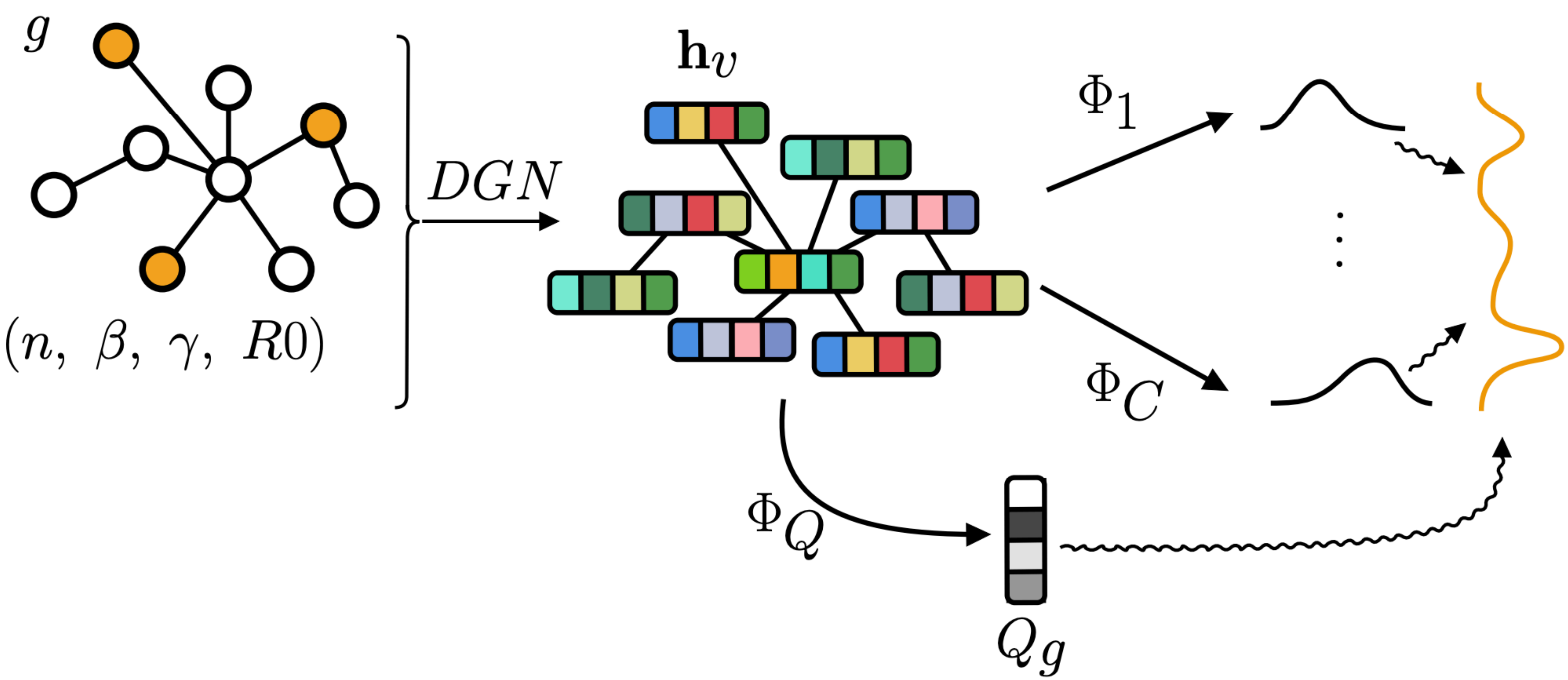
Figure 1: The flow of R-scores of node 1 across the different layers in MLPF. For MLP layers, the redistribution of R-scores follows the standard LRP rules [35, 36]. For the aggregation step in the message passing layer, the redistribution follows Equation 3. We only show three nodes for simplicity.

Feature correlation for top tagging.

arXiv 2210.04371
Ayush Khot, Mark S. Neubauer, Avik Roy



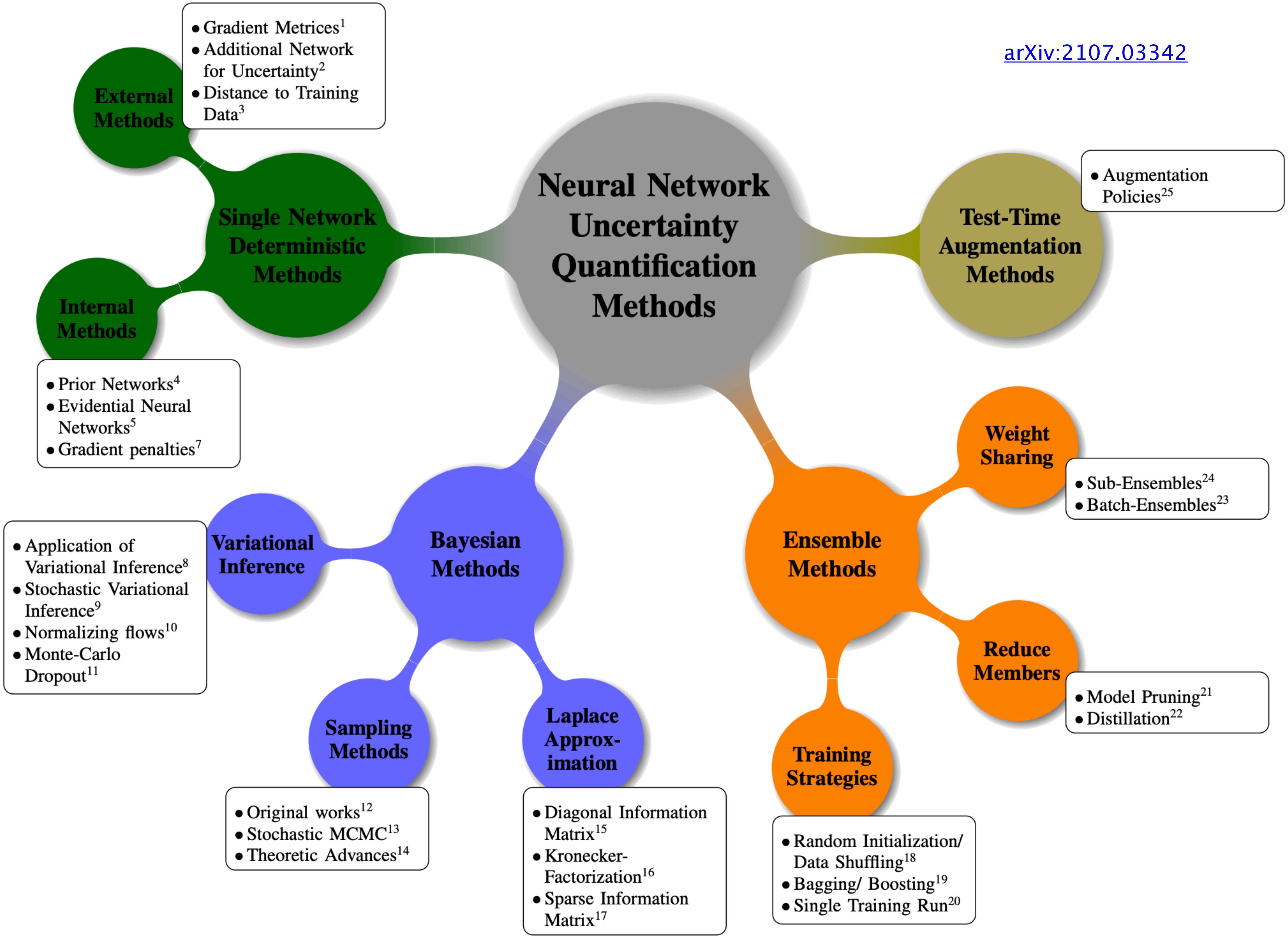
Major thrust in immediate future : Uncertainty



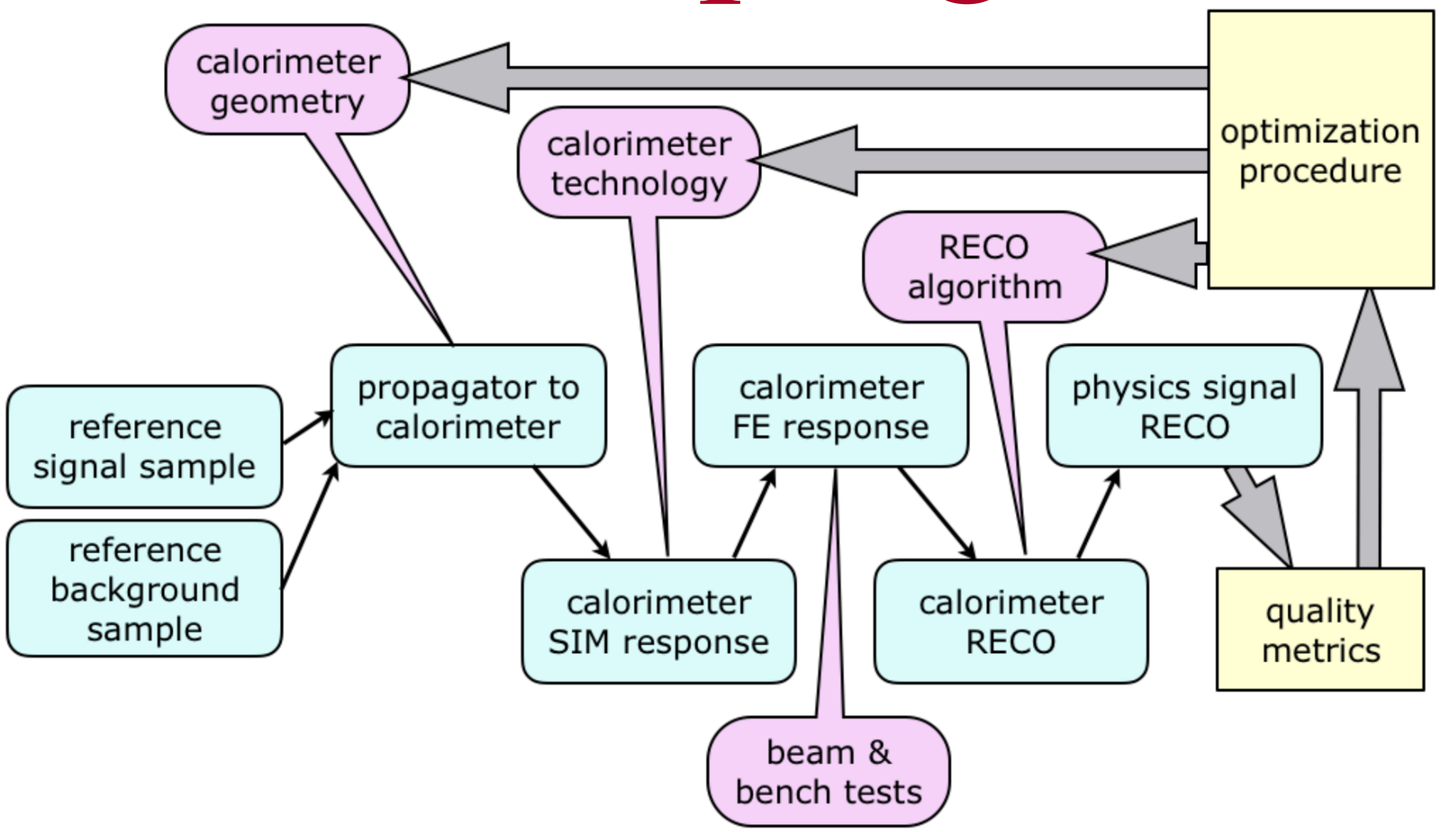
Reliable uncertainty estimation on ML based predictions are crucial for HEP
Only few Bayesian methods have been tested naively.

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

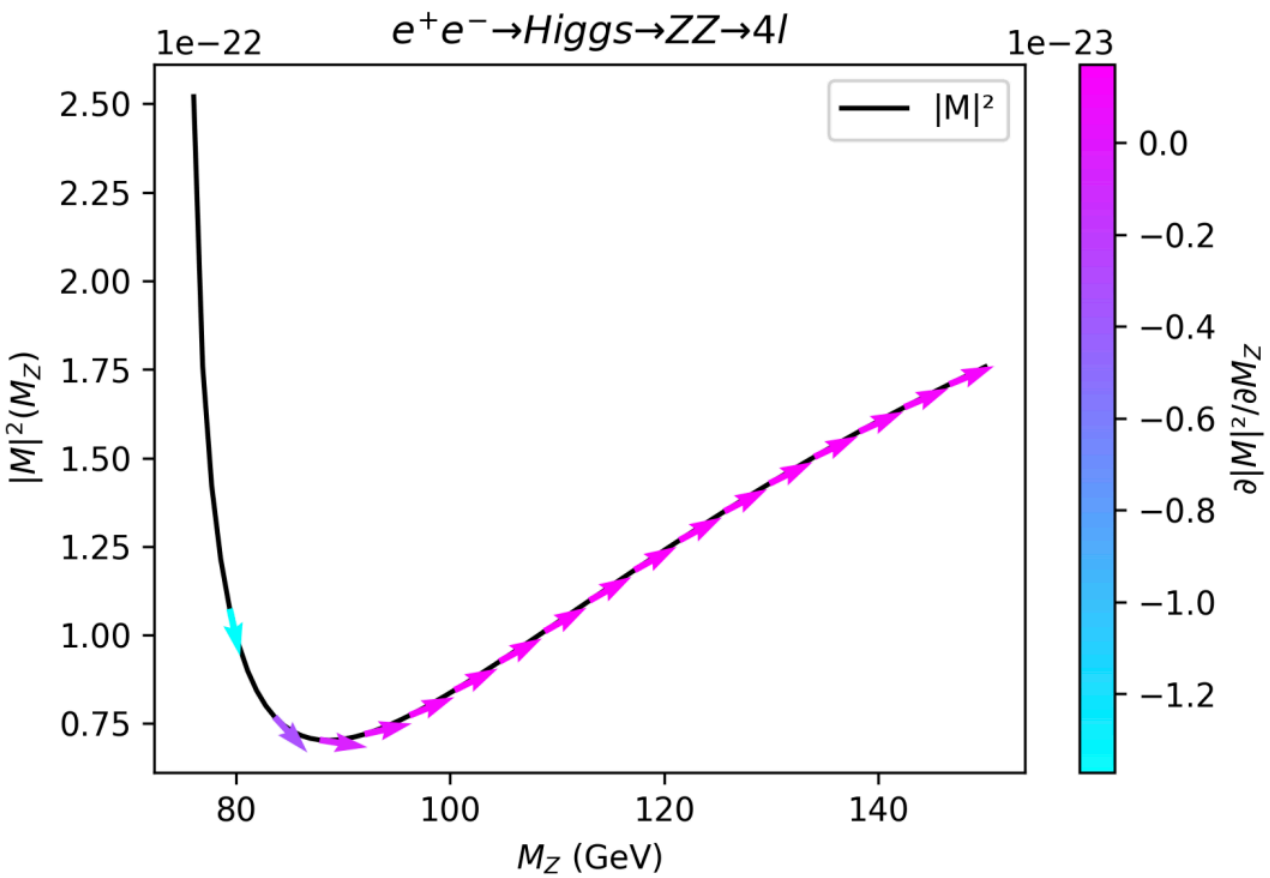
Major thrust in immediate future : Uncertainty



Differential programming in HEP



MODE
JENAA
iris hep
APPEC
Third MODE Workshop on
Differentiable Programming for Experiment Design
Princeton University
24-26 July, 2023




```
generate p p > t t~, t > b udsc udscx , t~ > b~ udsc udscx
output madjax generated_ttbar
set auto_update 0
```

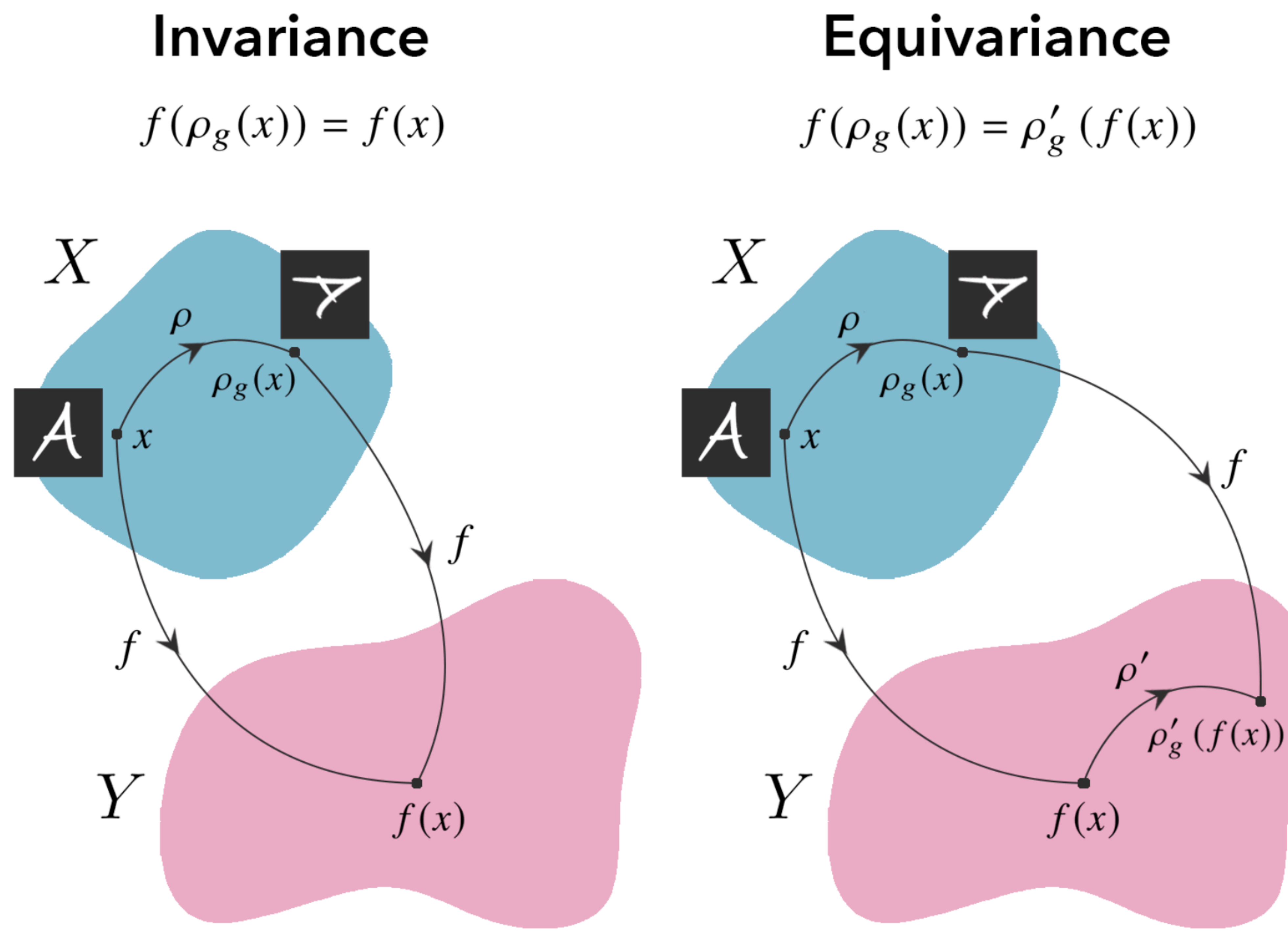
arXiv > hep-ph > arXiv:2203.00057
High Energy Physics – Phenomenology
[Submitted on 28 Feb 2022]
Differentiable Matrix Elements with MadJax
Lukas Heinrich, Michael Kagan

2. Evaluation:

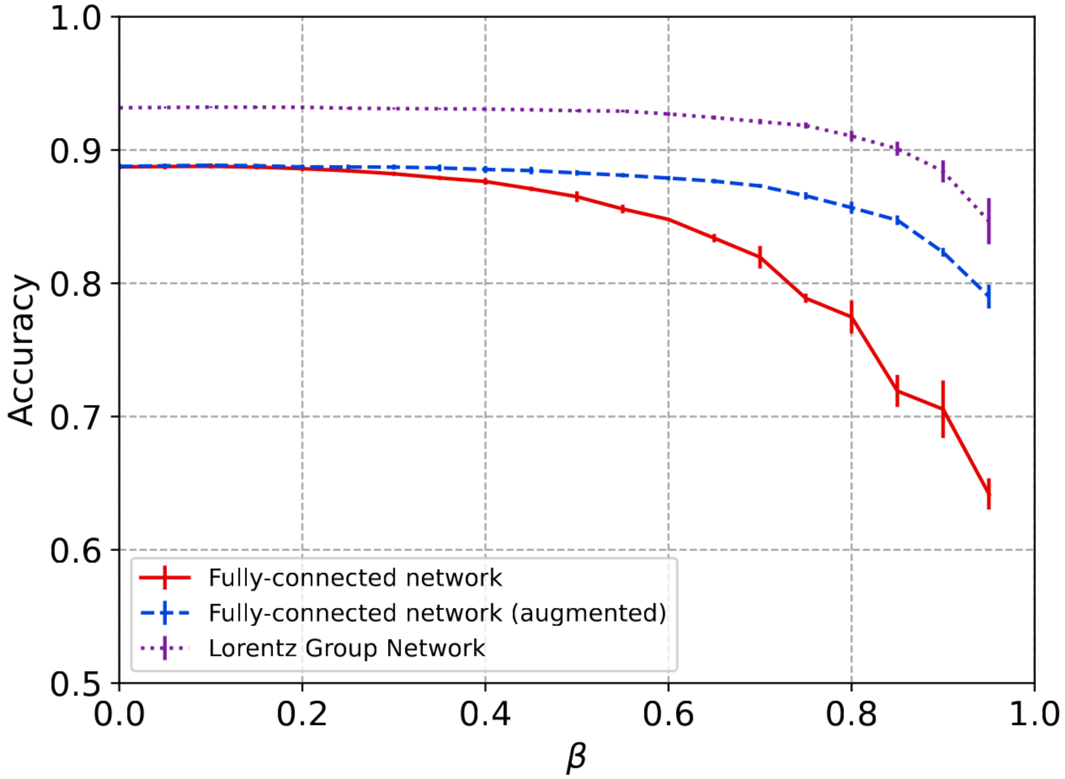
```
import madjax
mj = madjax.MadJax('generated_ttbar')
E_cm = 14000 #GeV
process = 'Matrix_1_gg_ttx_t_budx_tx_bxdux'
matrix_element = mj.matrix_element(E_cm,process)

parameters = ('mass',6): 173.0 #set top mass
phasespace_coords = [0.1]*14 #14D phasespace

val, grad = matrix_element(parameters,phasespace_coords)
grad[('mass', 6)] #gradient wrt top mass
```

Symmetry equivariant networks

arXiv:2203.06153 : SG et al



Symmetry equivariant networks

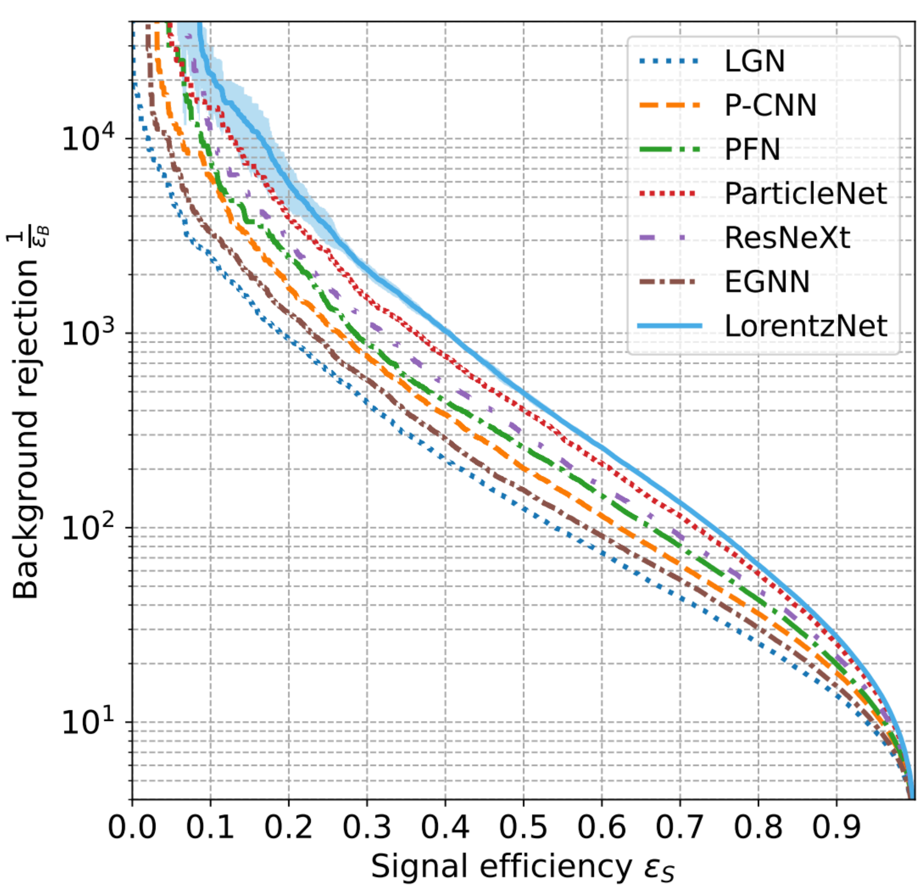


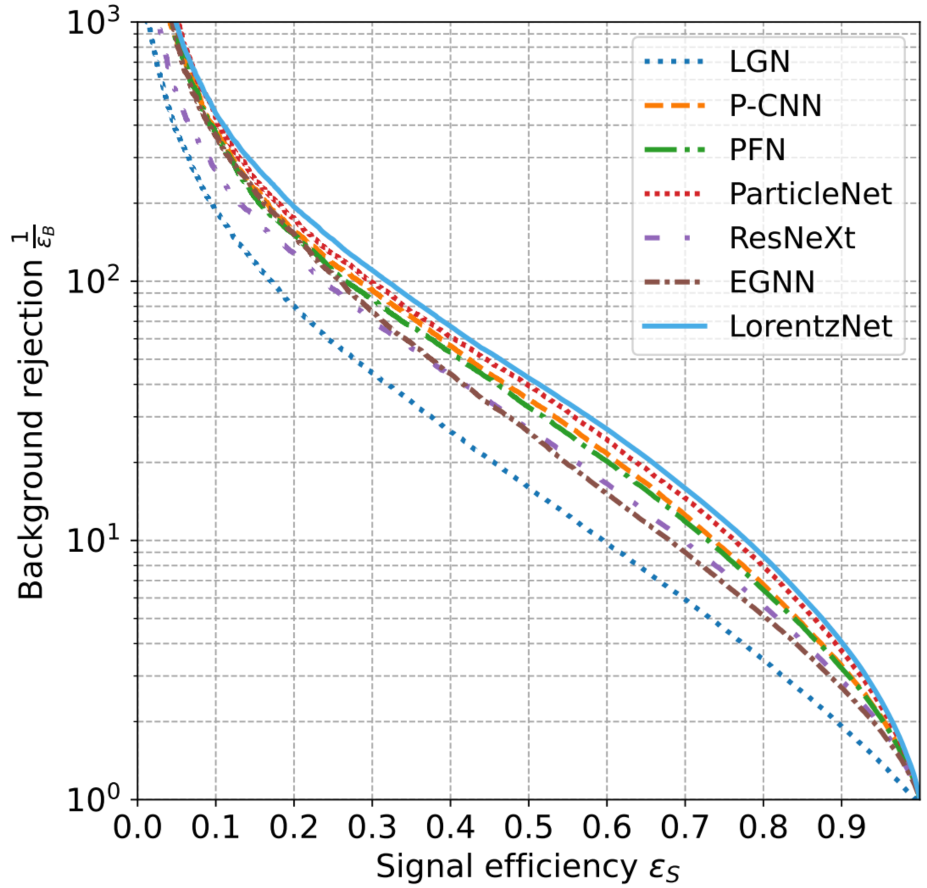
arXiv:2006.04780 : A Bogatskiy et al

arXiv:2203.06153 : SG et al

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

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

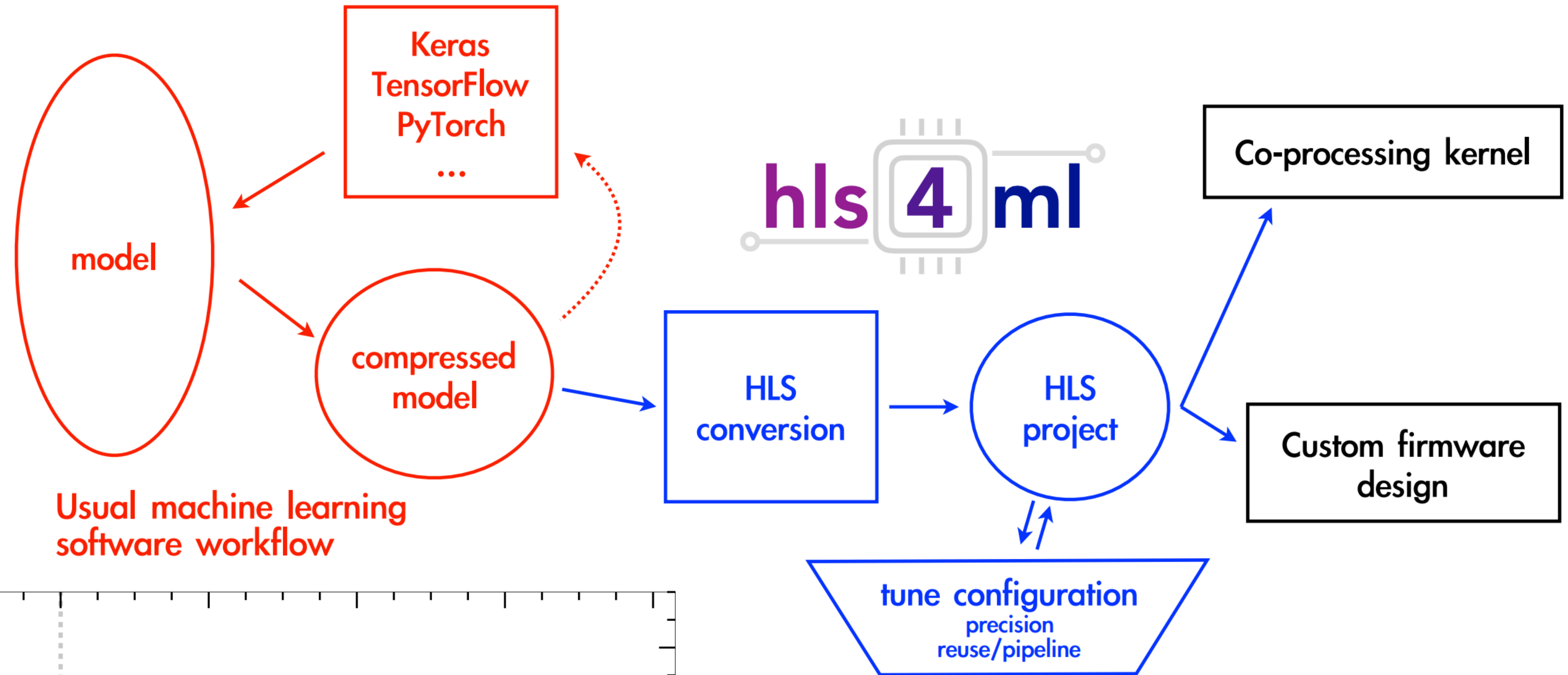


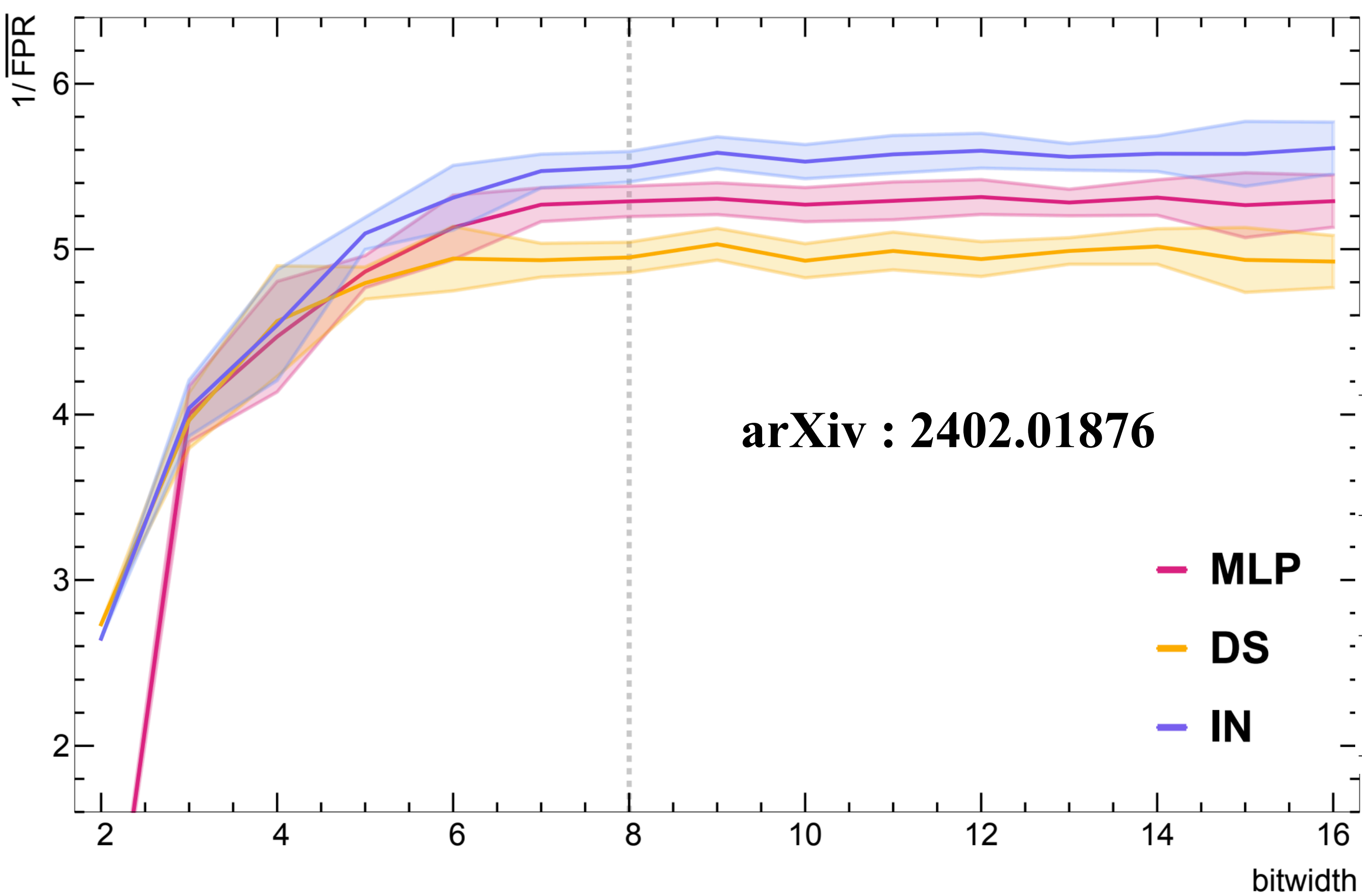


LG equivariant GNN :
arXiv 2201.08187

ML on FPGA

arXiv : 1804.06913



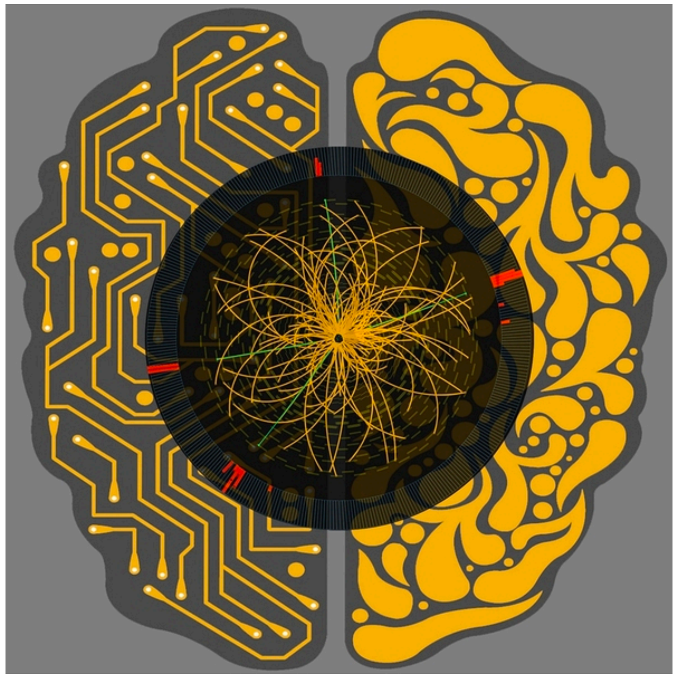


arXiv : 2402.01876

Architecture	Constituents	Parameters	FLOPs	Accuracy	AUC g	AUC q	AUC W	AUC Z	AUC t
MLP	8	26,826	53,162	$64.6 \pm 0.1\%$	0.84	0.88	0.90	0.88	0.92
DS	8	3,461	36,805	$64.0 \pm 0.3\%$	0.84	0.88	0.90	0.88	0.92
IN	8	3,347	37,232	$64.9 \pm 0.2\%$	0.84	0.88	0.91	0.89	0.92
MLP	16	20,245	40,485	$68.4 \pm 0.3\%$	0.87	0.89	0.91	0.90	0.94
DS	16	3,461	71,109	$69.4 \pm 0.2\%$	0.87	0.89	0.93	0.92	0.94
IN	16	3,347	140,432	$70.8 \pm 0.2\%$	0.88	0.90	0.94	0.92	0.94
MLP	32	24,101	48,197	$66.2 \pm 0.2\%$	0.90	0.89	0.89	0.88	0.94
DS	32	3,461	139,717	$75.9 \pm 0.1\%$	0.91	0.91	0.96	0.95	0.95
IN	32	7,400	109,556	$75.8 \pm 0.3\%$	0.91	0.91	0.96	0.95	0.95

Take away

Image: FermiLab

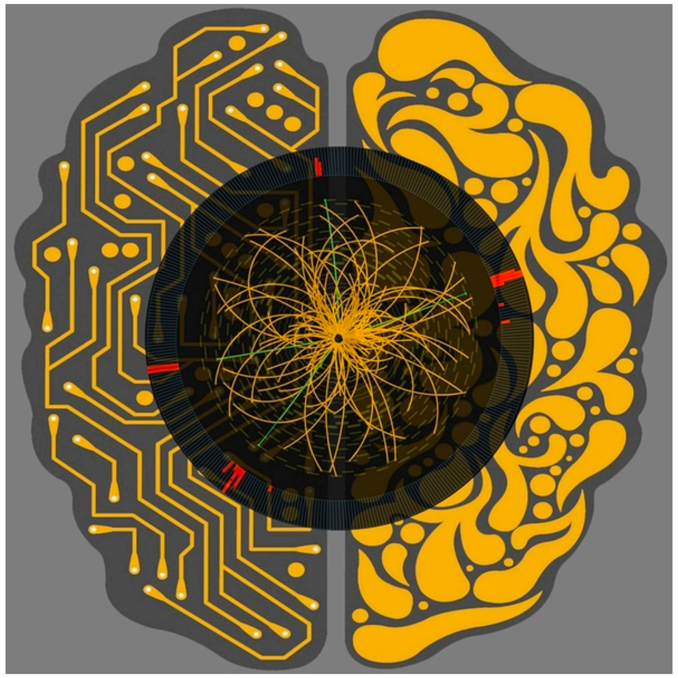


- ☑ ML is here to stay with HEP.
- ☑ When looked through the lens of ML, LHC and EIC are not that far.
- ☑ Interpretability and uncertainty estimations are two key aspects where we the HEP-ML people need to emphasize.
- ☑ Need to keep a close connection with the comp-sc/math community with the latest developments and contribute if possible.
- ☑ Symmetry equivariance and geometric DL methods might play a key role in this field.
- ☑ Didn't want to talk about an elephant entering the room : QML (but should track it).

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

Take away

Image: FermiLab



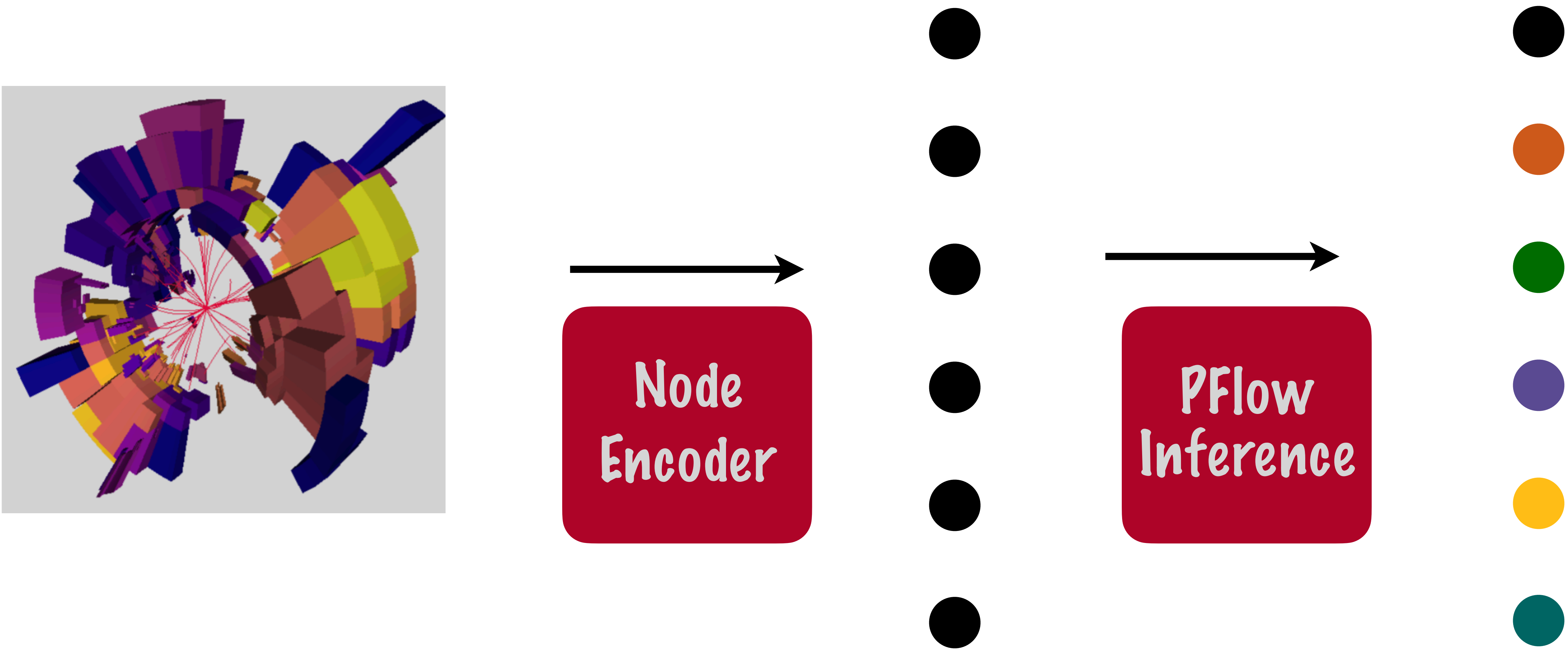
- ☑ ML is here to stay with HEP.
- ☑ When looked through the lens of ML, LHC and EIC are not that far.
- ☑ Interpretability and uncertainty estimations are two key aspects where we the HEP-ML people need to emphasize.
- ☑ Need to keep a close connection with the comp-sc/math community with the latest developments and contribute if possible.
- ☑ Symmetry equivariance and geometric DL methods might play a key role in this field.
- ☑ Didn't want to talk about an elephant entering the room : QML (but should track it).

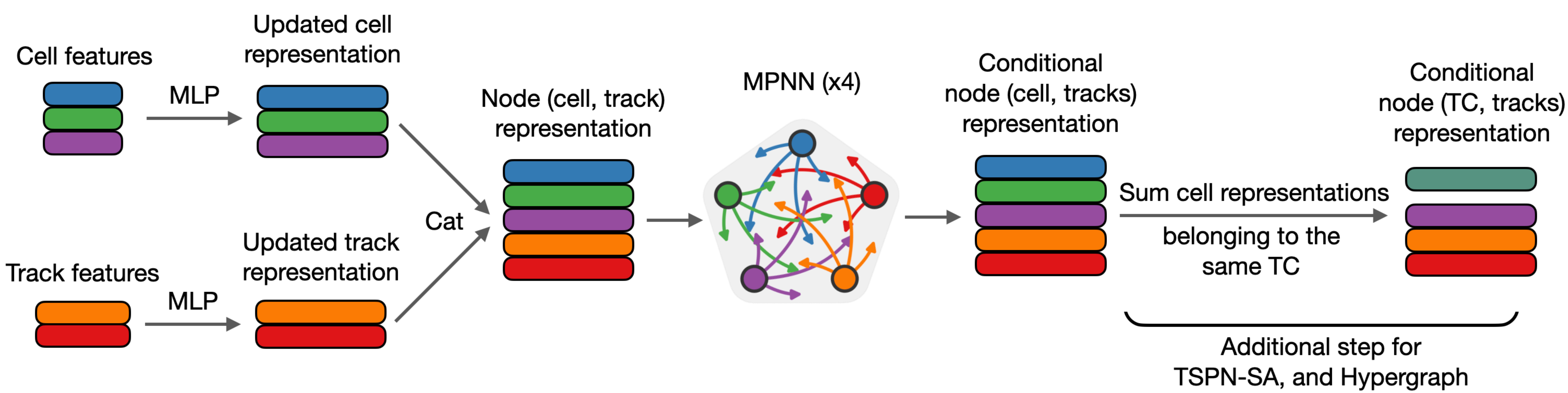
https://iml-wg.github.io/HEPML-LivingReview/

THANK YOU


BACKUP

The general workflow

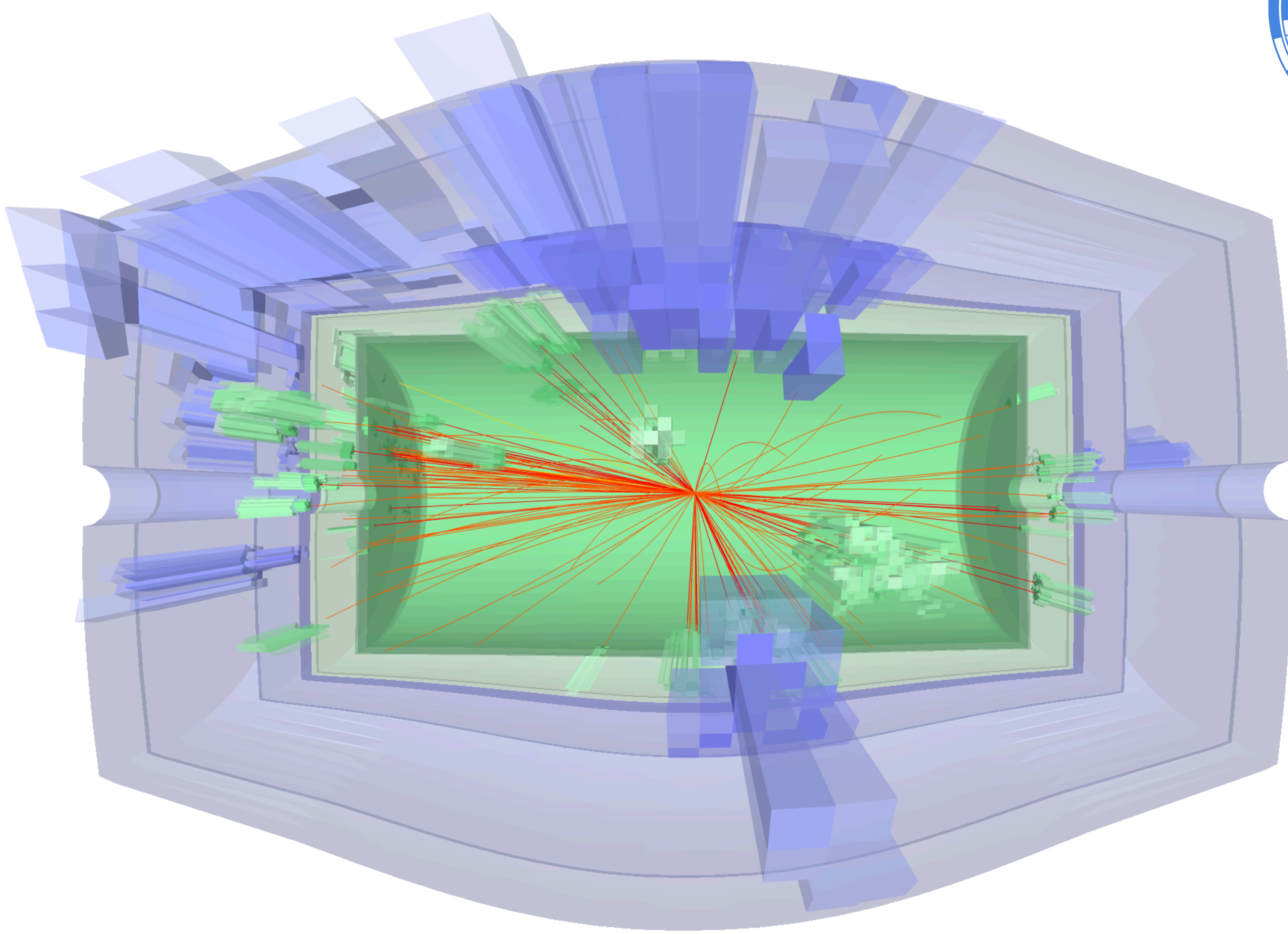


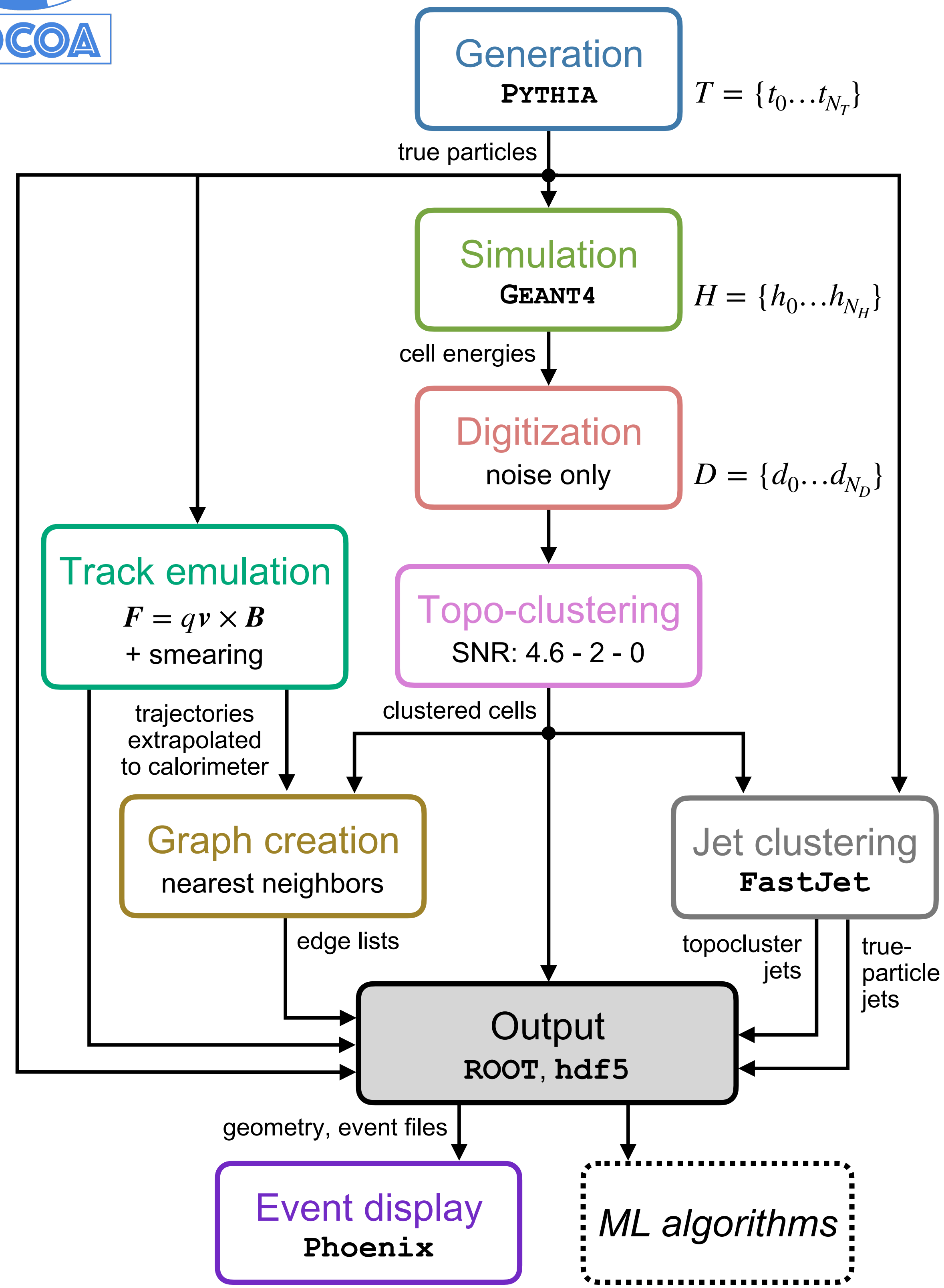


The COCOA

Mach. Learn.: Sci. Technol. 4 035042

https://cocoa-hep.readthedocs.io/en/latest/

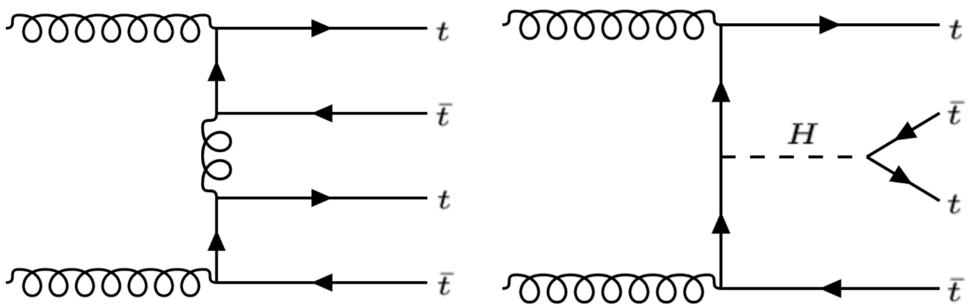




COnfigurable CalOrimeter simulation for AI

- ☑ A complete hermetic geometry with full GEANT simulation.
- ☑ PYTHIA-8 based ME/PS & Hadronization
- ☑ FASTJET integration is inbuilt.
- ☑ Comes with an ATLAS style pPFlow.

Event construction using NN



$$\sigma_{t\bar{t}t\bar{t}} = 22.5^{+4.7}_{-4.3}(\text{stat})\,^{+4.6}_{-3.4}(\text{syst})\text{ fb} = 22.5\,^{+6.6}_{-5.5}\text{ fb.}$$

Eur. Phys. J. C 83 (2023) 496

Operators	Expected C_i/Λ^2 [TeV $^{-2}$]	Observed C_i/Λ^2 [TeV $^{-2}$]
O^1_{QQ}	[-2.4, 3.0]	[-3.5, 4.1]
O^1_{Qt}	[-2.5, 2.0]	[-3.5, 3.0]
O^1_{tt}	[-1.1, 1.3]	[-1.7, 1.9]
O^8_{Qt}	[-4.2, 4.8]	[-6.2, 6.9]