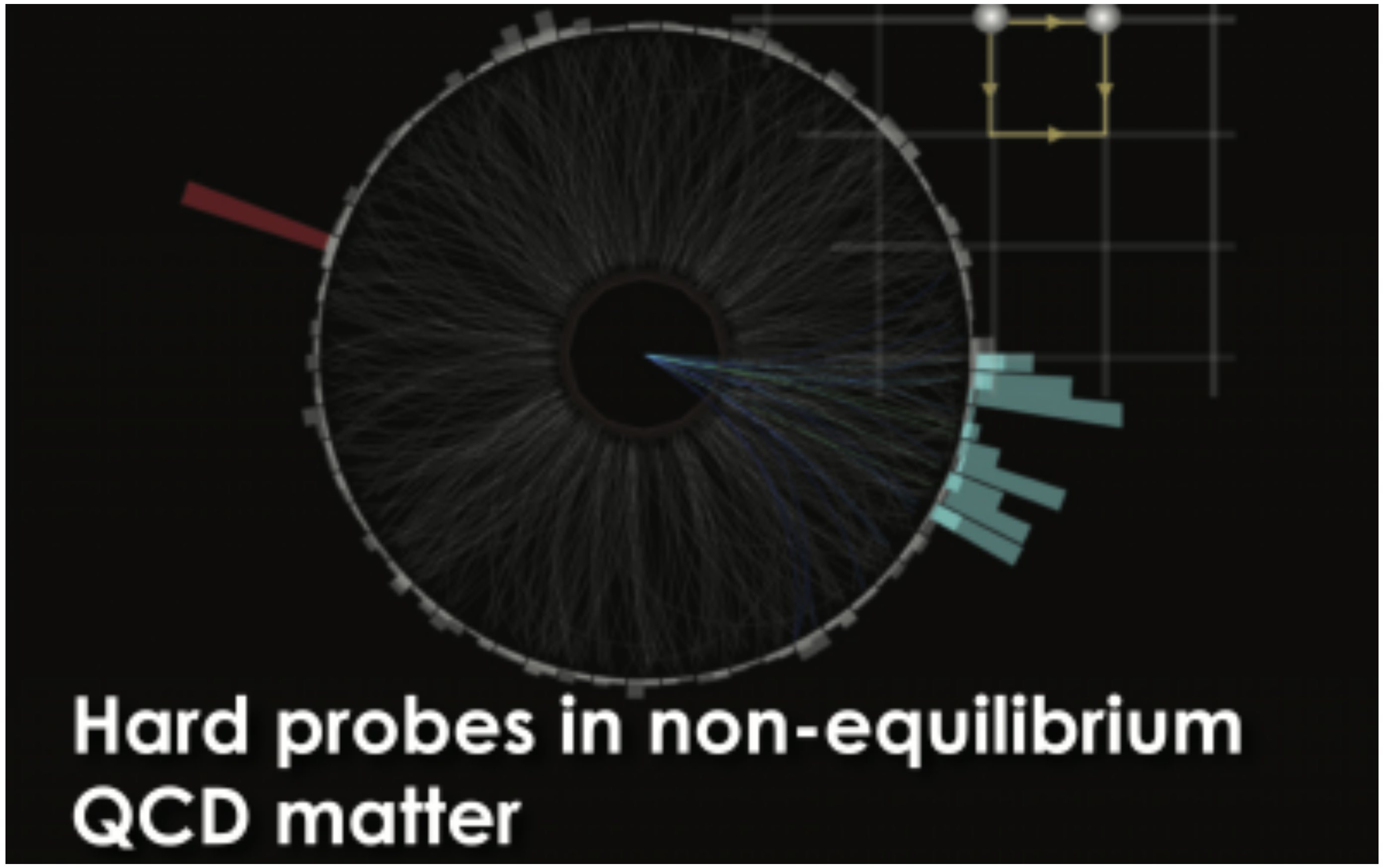
Jets at LHC and Future Colliders
The inevitable ML connection

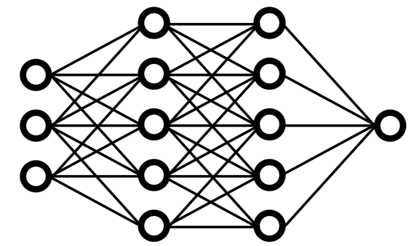
Sanmay Ganguly
IIT-Kanpur
sanmayphy@github.io
27th March, 2026

Part-2 : ML4Jets $\rightarrow$ Current Status & Future Prospects

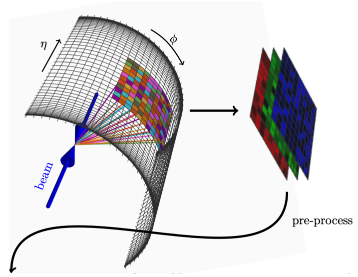


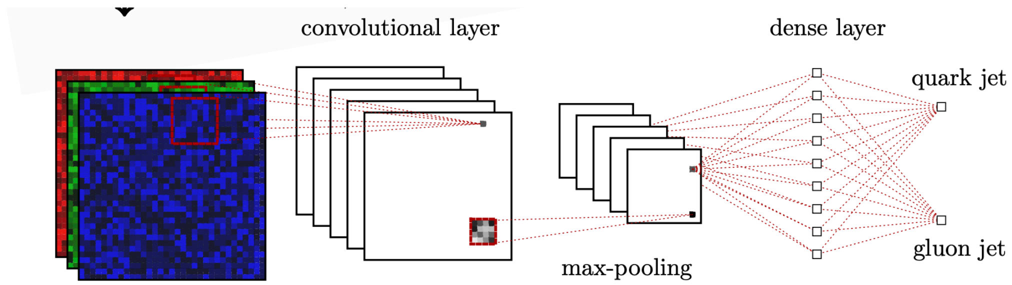
Data representation ⇔ NN correspondence

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

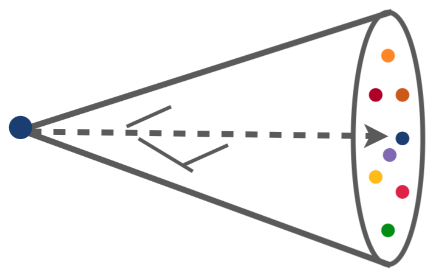


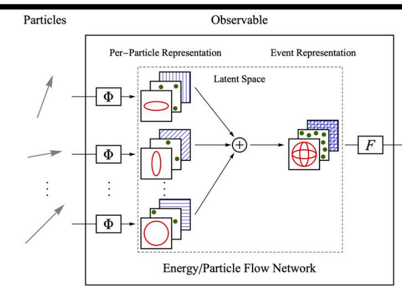
Ordered set
DNN



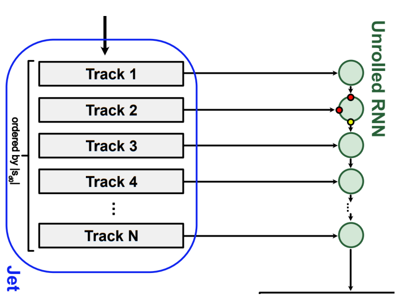


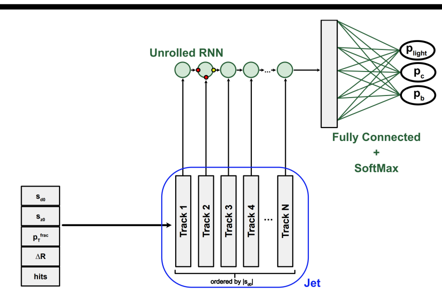
Grid
CNN



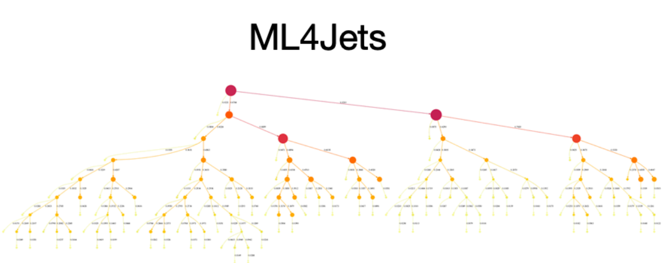


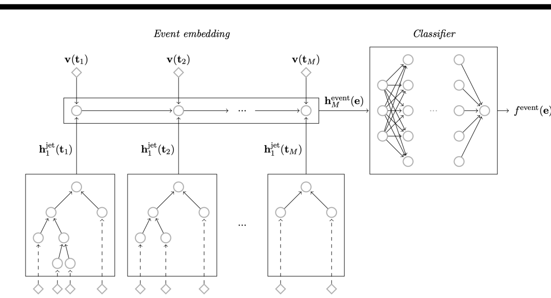
Unordered set
Deepest



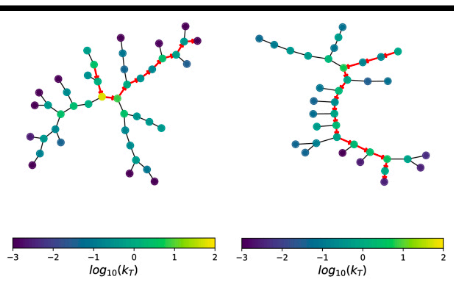


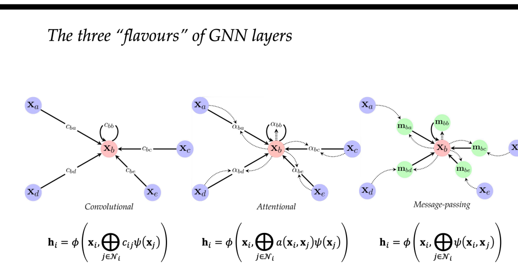
Sequential data
RNN





Tree structure
Deepset/GNN



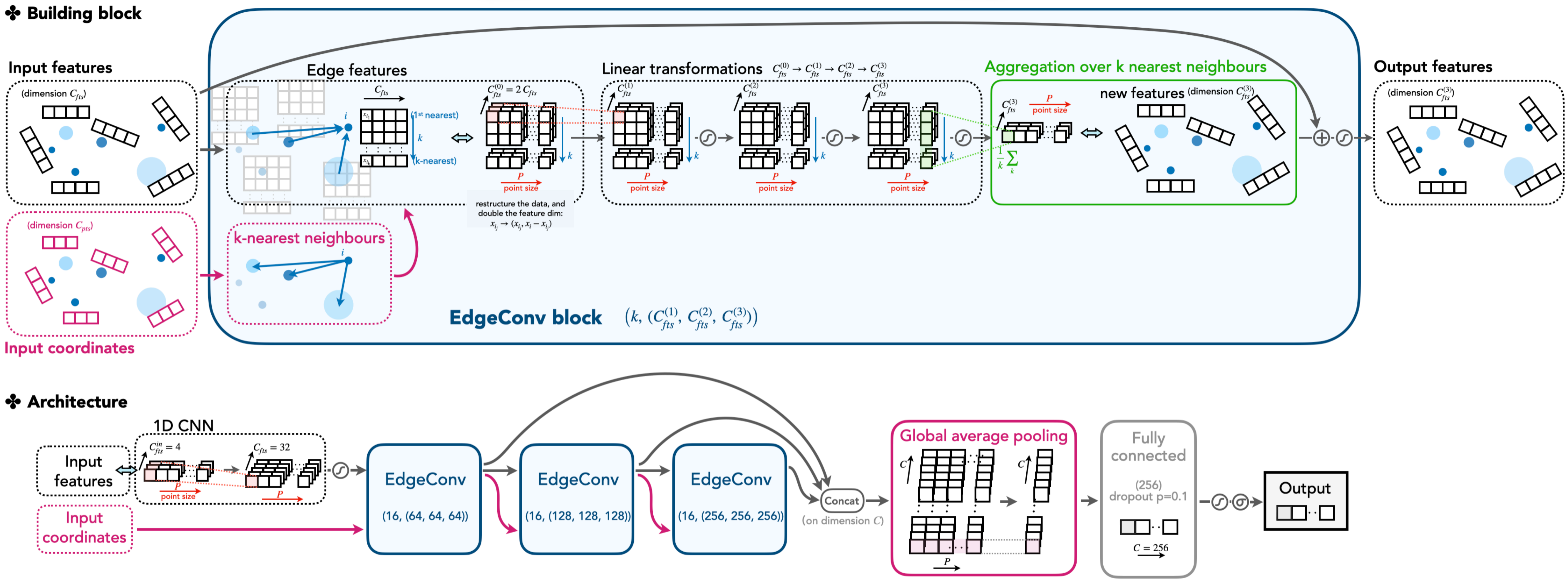


Graph
GNN

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

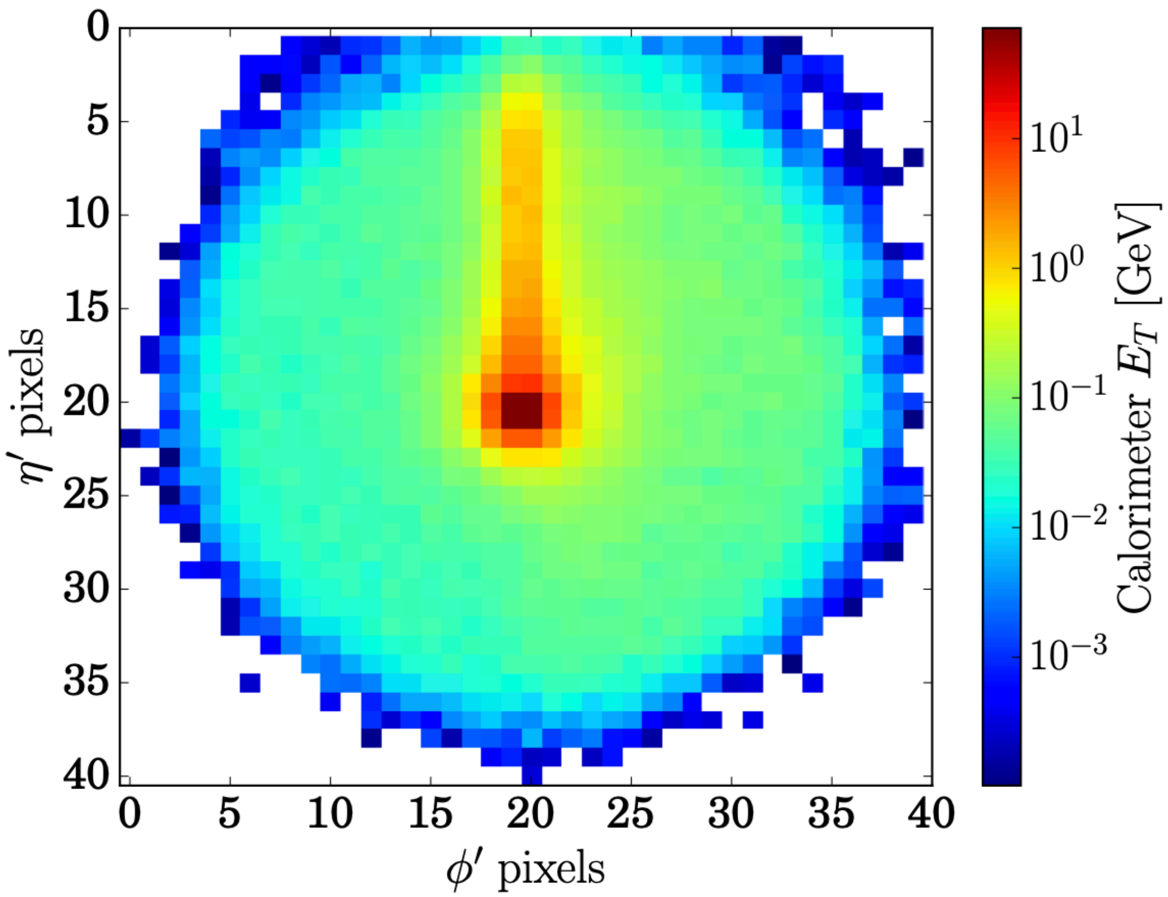
Computer Science > Computer Vision and Pattern Recognition

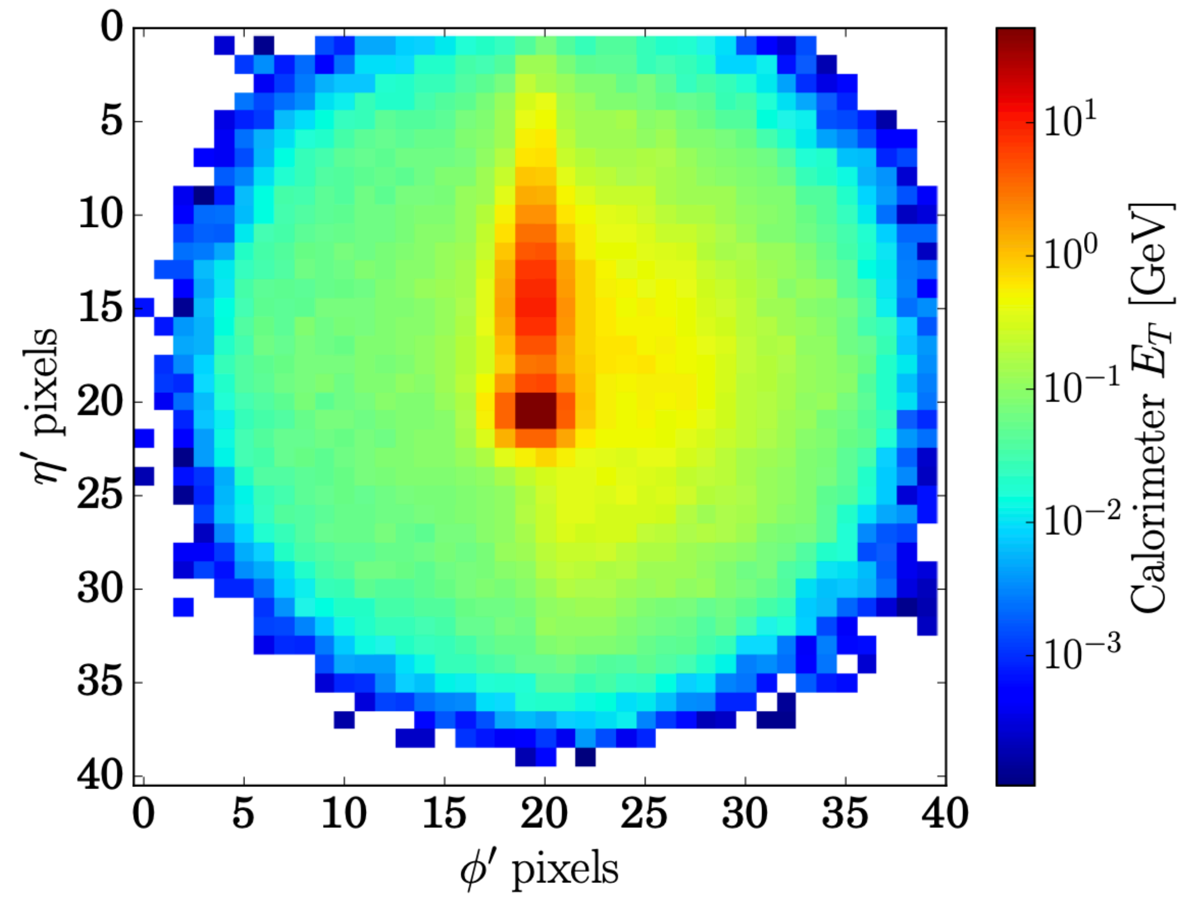
[Submitted on 24 Jan 2018 (v1), last revised 11 Jun 2019 (this version, v2)]

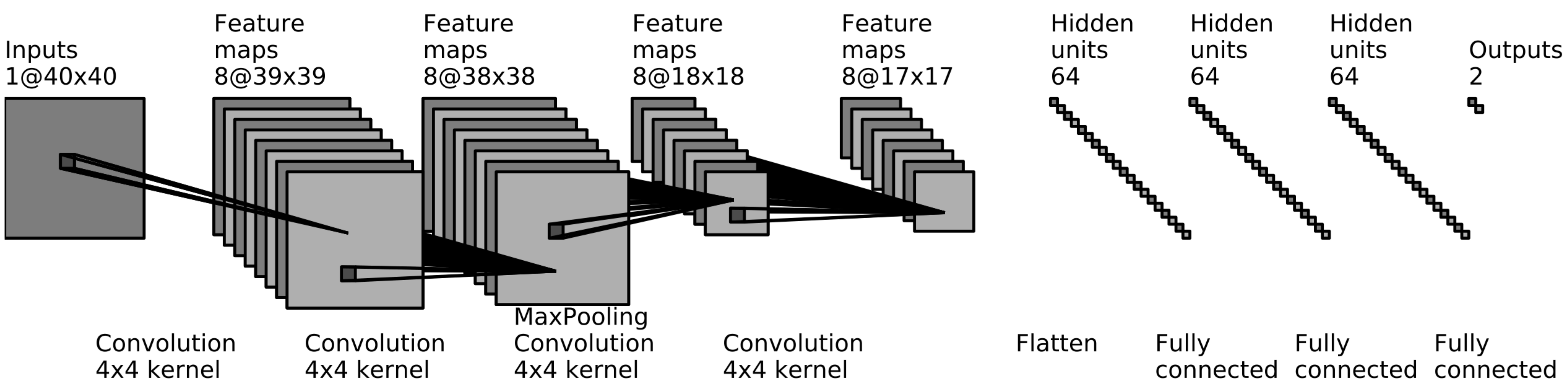
Dynamic Graph CNN for Learning on Point Clouds

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

Jet images & ML4Jets







Machine learning on sets

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

https://geometricdeeplearning.com/lectures/

Machine learning on sets

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

Neural network on a set

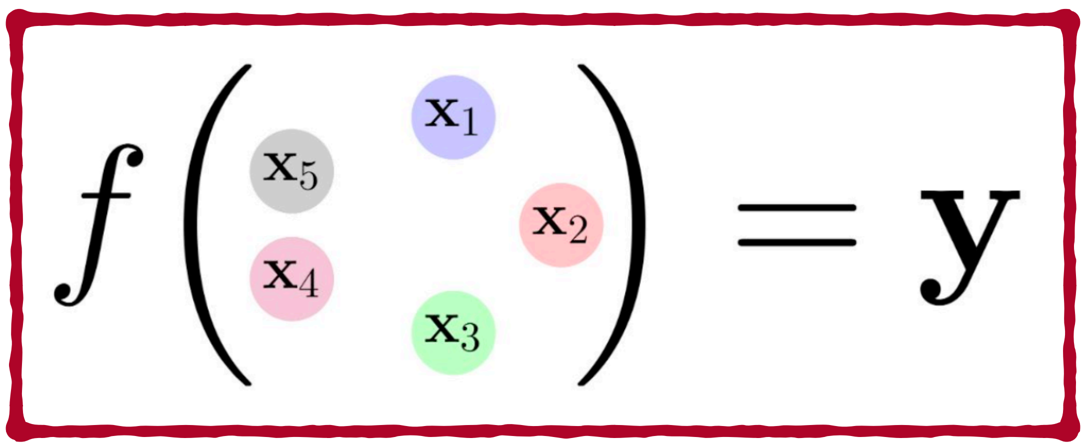
https://geometricdeeplearning.com/lectures/

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/

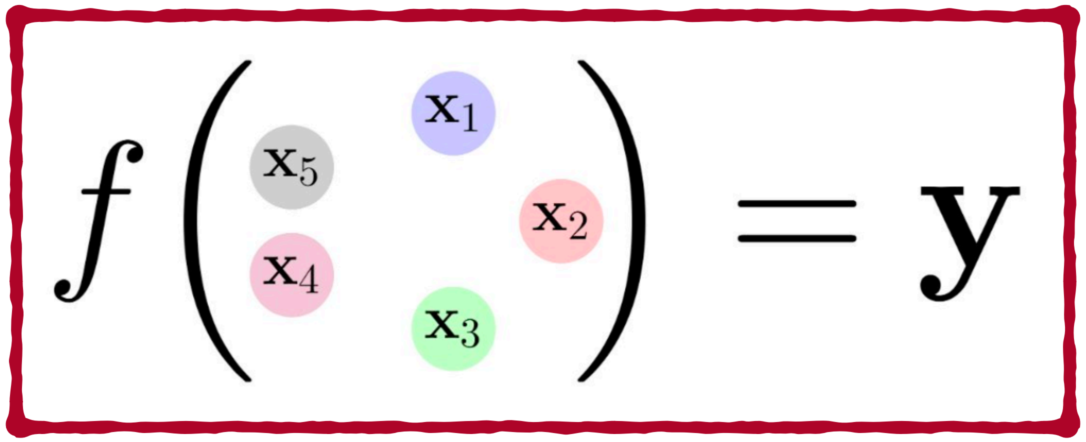


Machine learning on sets

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/



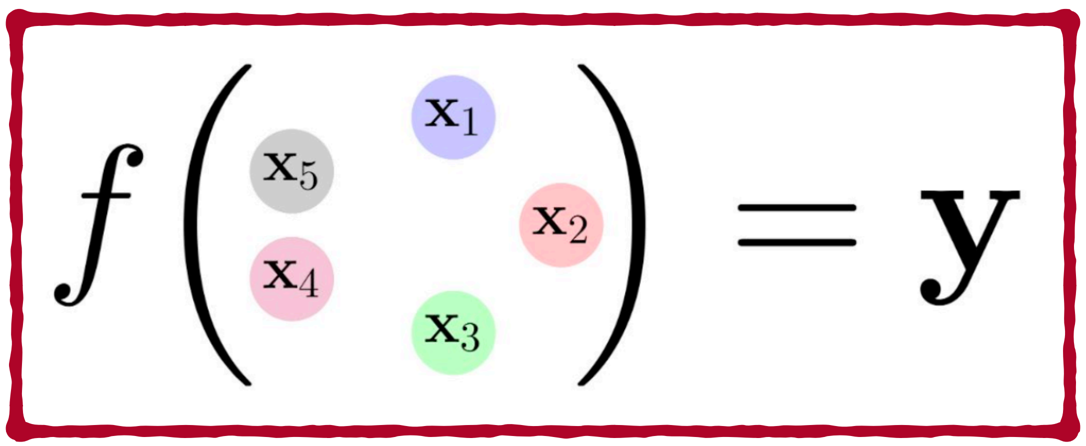
Basic required property : permutation invariance

Machine learning on sets

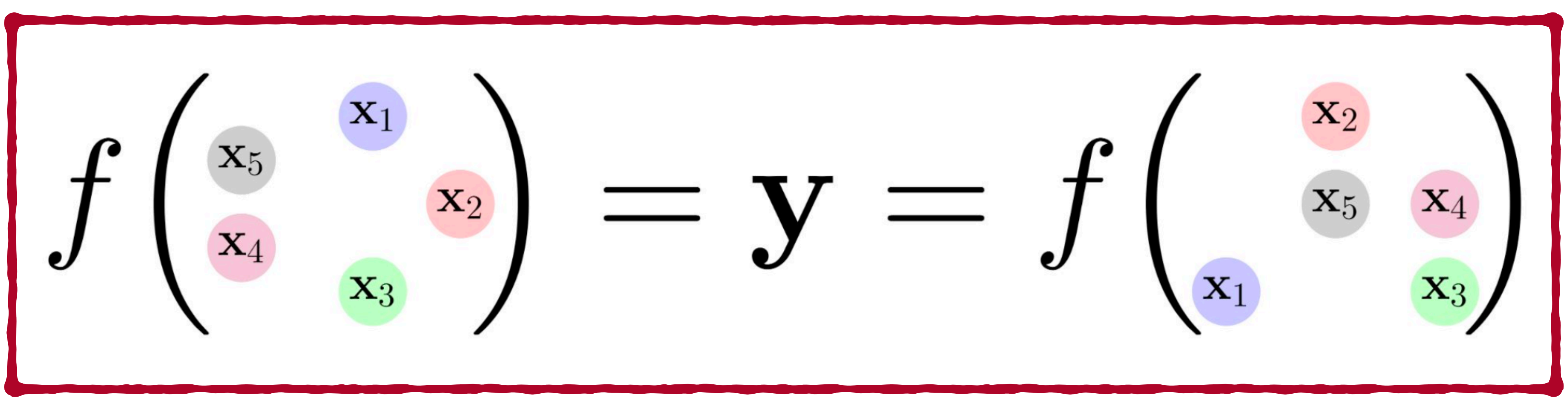
Let $x_1, x_2, \ldots, x_N \in \mathbb{R}^k$ be n pieces of data. This forms a set of cardinality N.

Neural network on a set

https://geometricdeeplearning.com/lectures/



Basic required property : permutation invariance



How the P.I. is achieved?

$$\mathbf{f}\Big(\mathbf{PX}\Big) = \mathbf{f}\Big(\mathbf{X}\Big)$$

Remember the permutation on a set?

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

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

$$\mathbf{x}_1 \xrightarrow{\psi} \mathbf{h}_1$$

$$\mathbf{x}_2 \xrightarrow{\psi} \mathbf{h}_2$$

$$\mathbf{x}_3 \xrightarrow{\psi} \mathbf{h}_3$$

$$\mathbf{x}_4 \xrightarrow{\psi} \mathbf{h}_4$$

$$\mathbf{x}_5 \xrightarrow{\psi} \mathbf{h}_5$$

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

https://geometricdeeplearning.com/lectures/

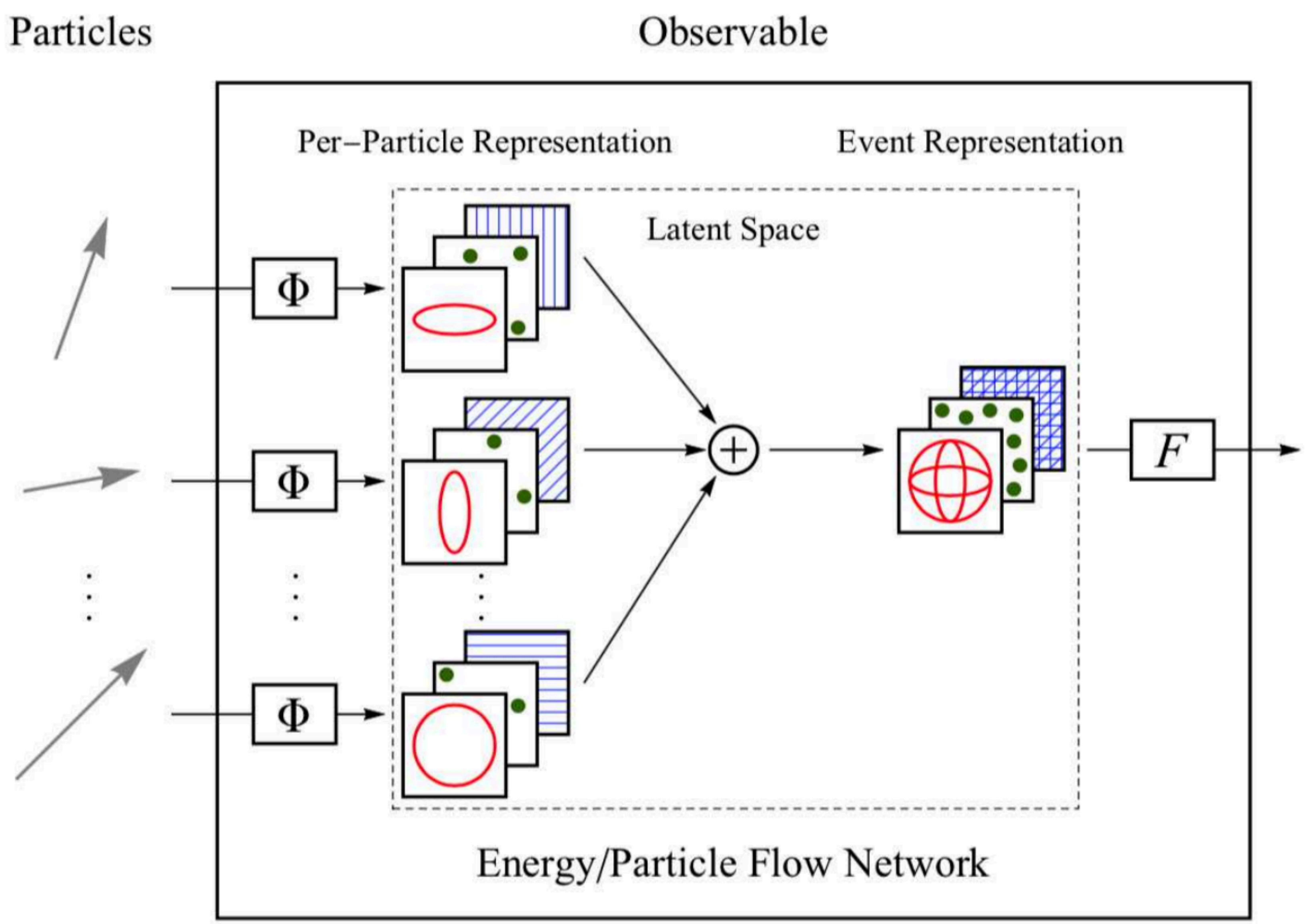
The permutation equivariant operation :

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

Example of deep-sets in HEP

[arXiv:1810.05165](https://arxiv.org/abs/1810.05165)

$$f(\{x_1, \ldots, x_M\}) = F\left(\sum_{i=1}^{M} \Phi(x_i)\right)$$



$$\text{EFN}\left(\{p_1^\mu, \ldots, p_M^\mu\}\right) = F\left(\sum_{i=1}^{M} z_i \Phi(\hat{p}_i)\right)$$

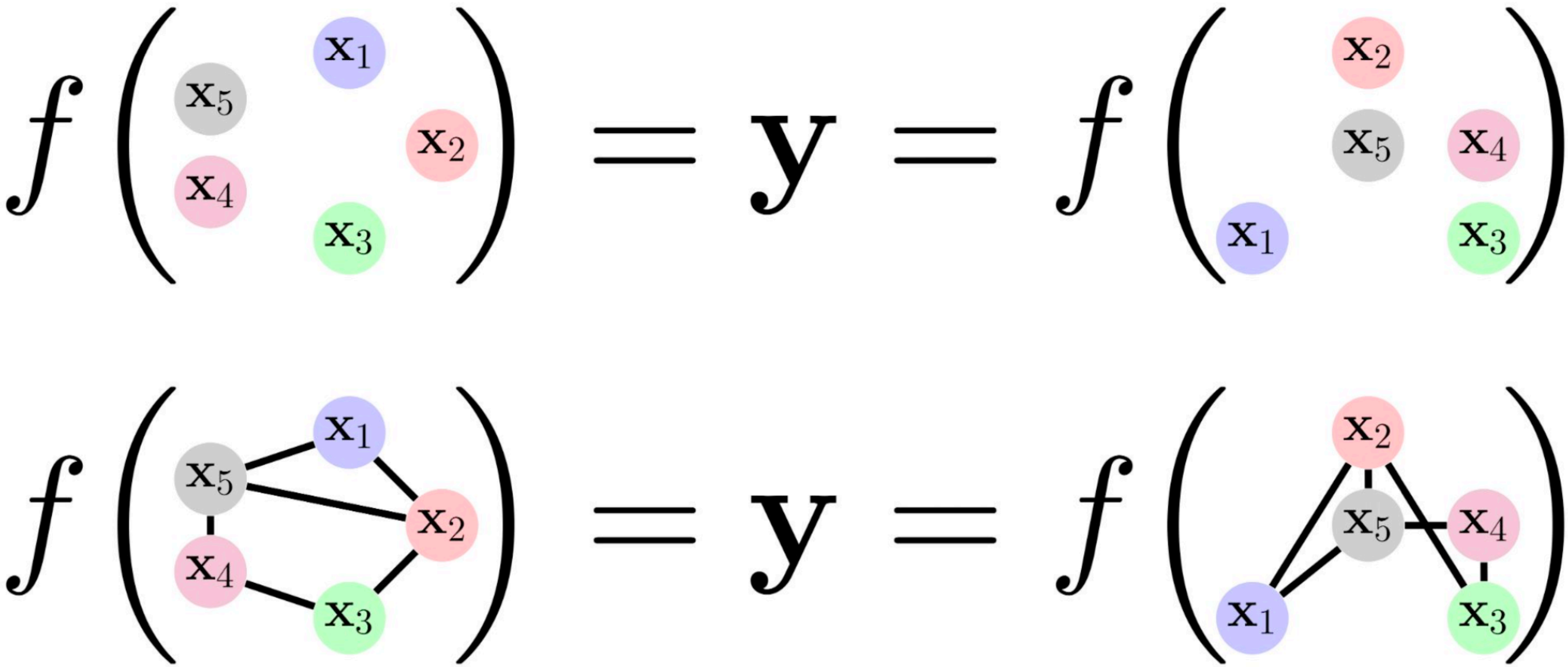
Manifestly IRC-safe latent space

$$\text{PFN}\left(\{p_1^\mu, \ldots, p_M^\mu\}\right) = F\left(\sum_{i=1}^{M} \Phi(p_i^\mu)\right)$$

Fully general latent space

What's the basic criteria of a GNN?

https://geometricdeeplearning.com/lectures/

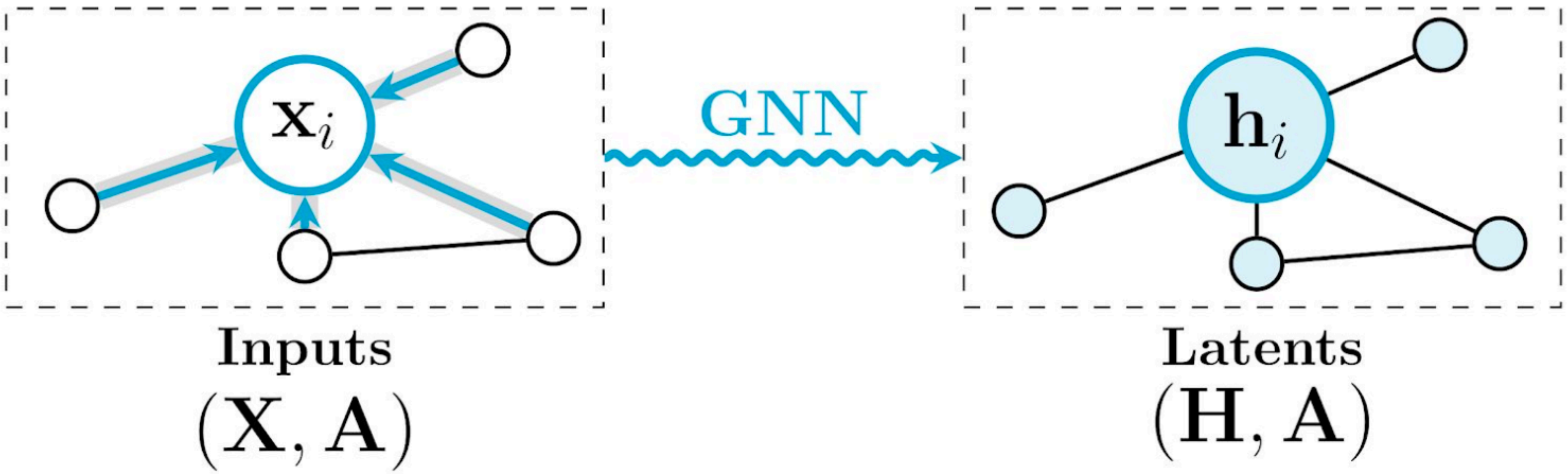


Invariance: $$f(\mathbf{PX}, \mathbf{PAP}^\top) = f(\mathbf{X}, \mathbf{A})$$

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

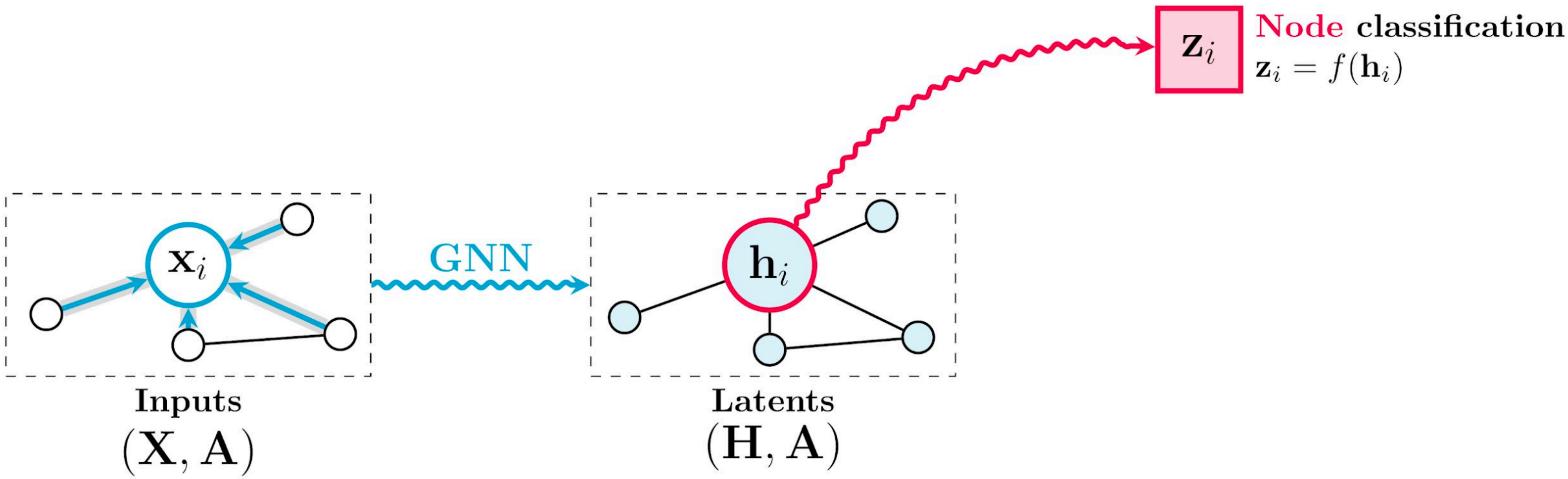
General methods of GNN

General methods of GNN



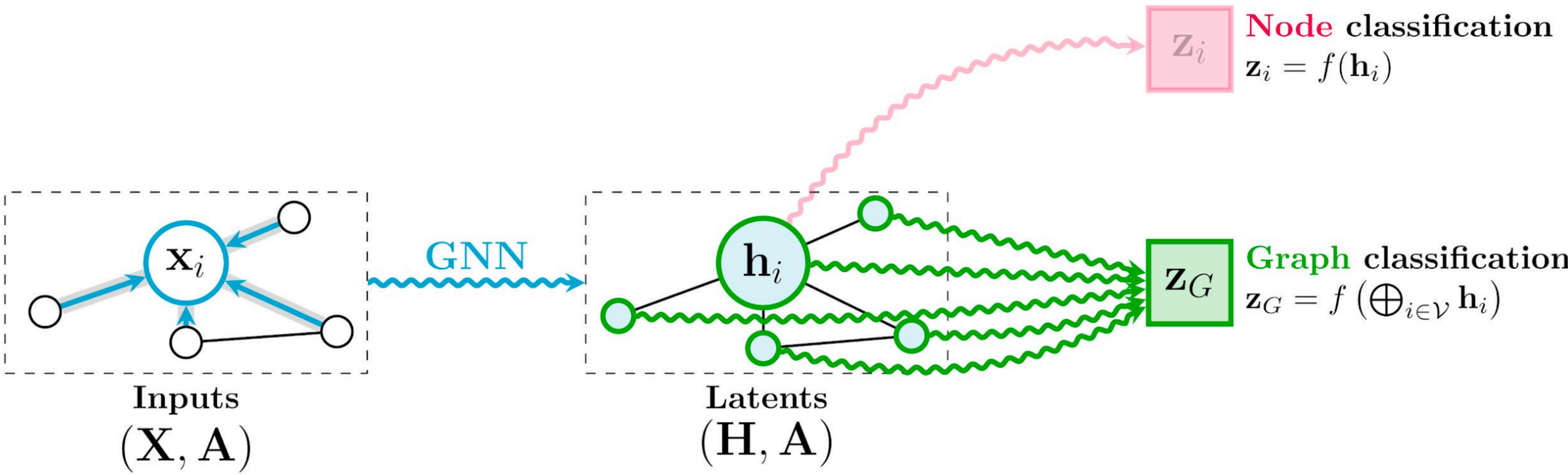
General methods of GNN

General methods of GNN



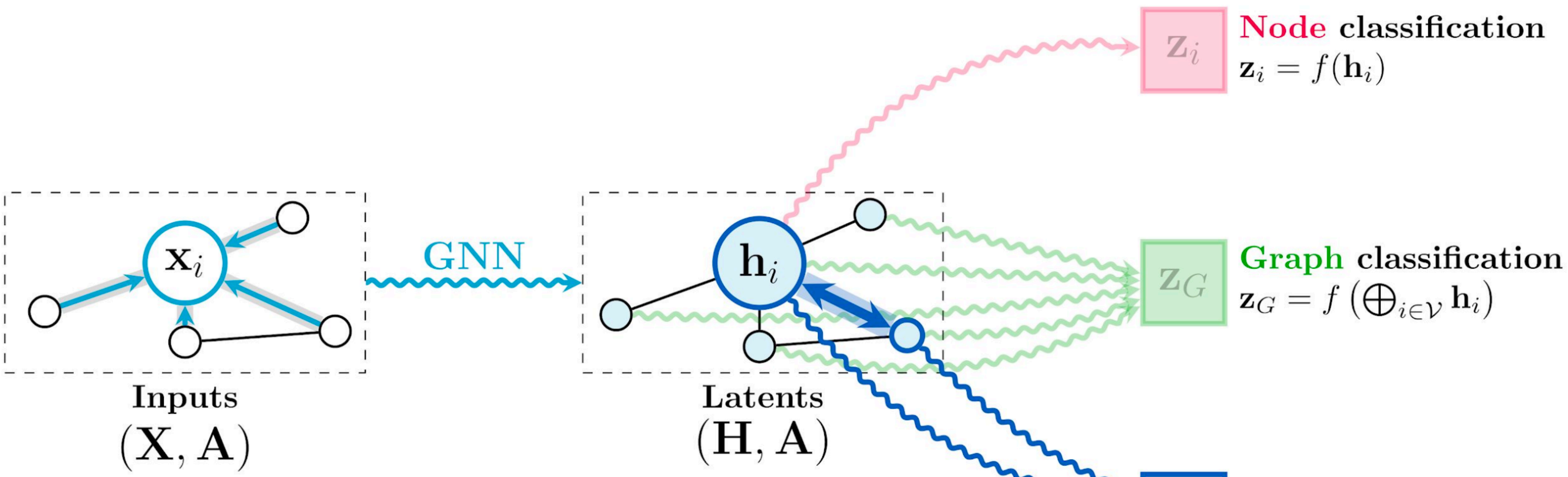
General methods of GNN

General methods of GNN



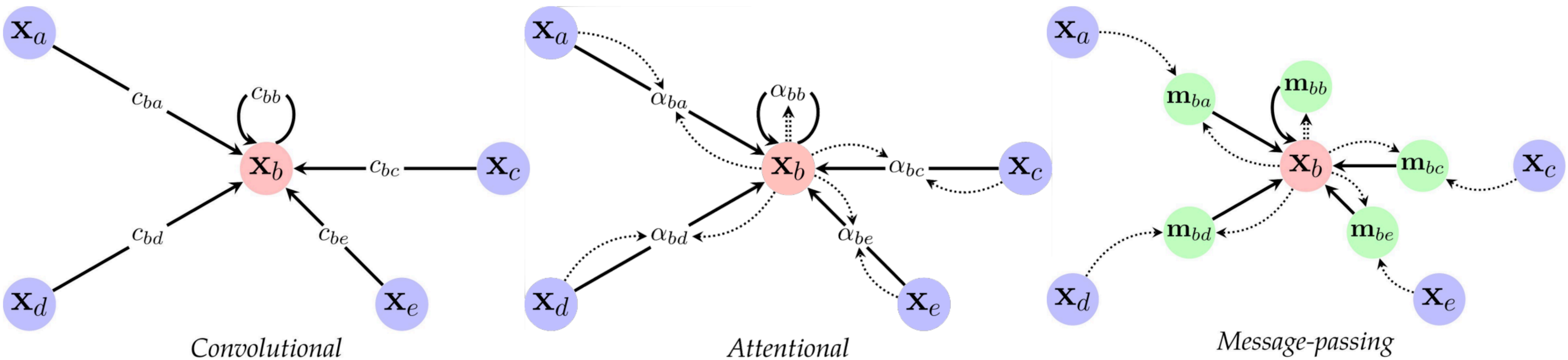
General methods of GNN

General methods of GNN



The different flavors of MPN

The three "flavours" of GNN layers



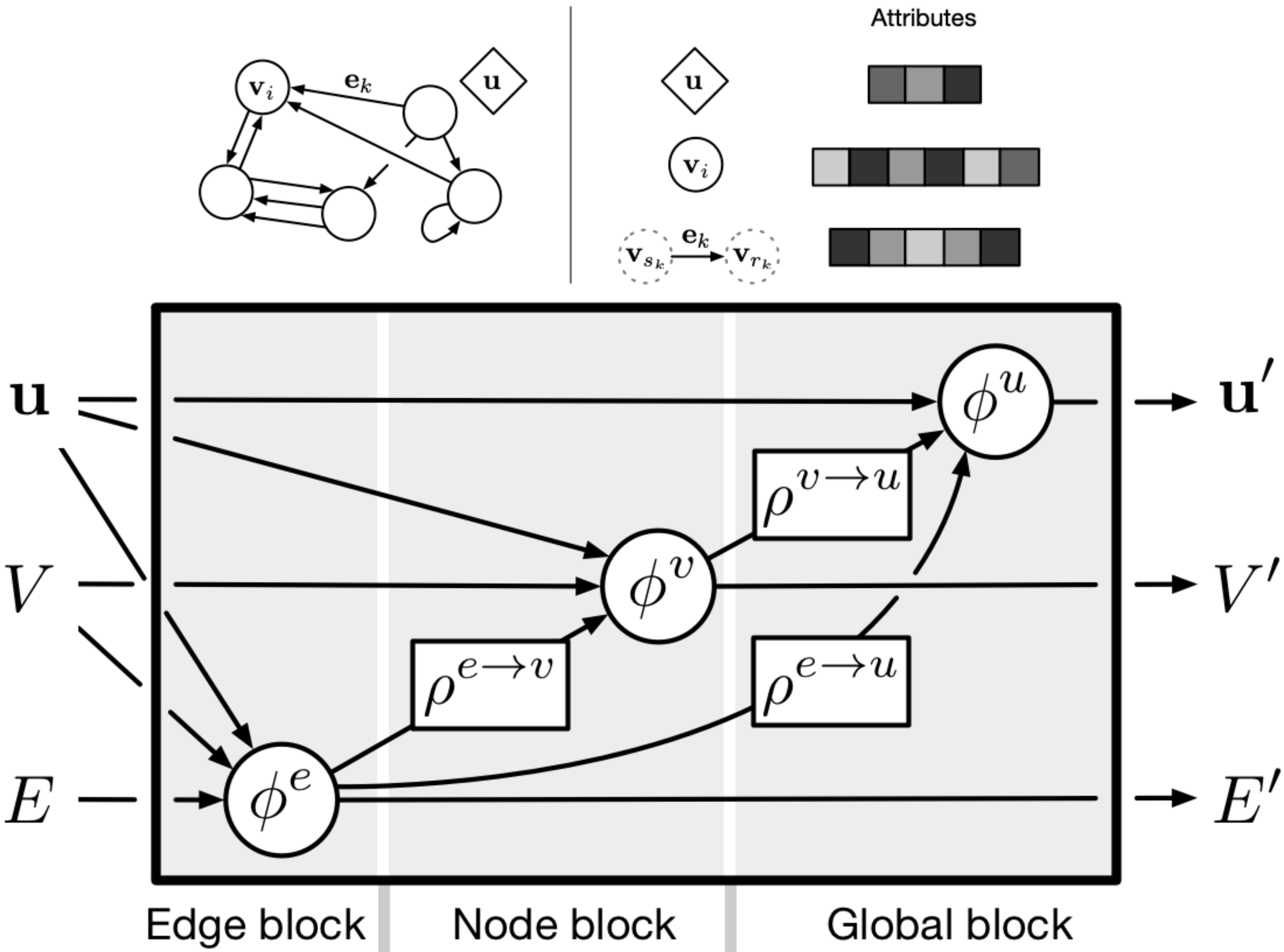
$$\mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j\in\mathcal{N}_i} c_{ij}\psi(\mathbf{x}_j)\right)$$

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

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

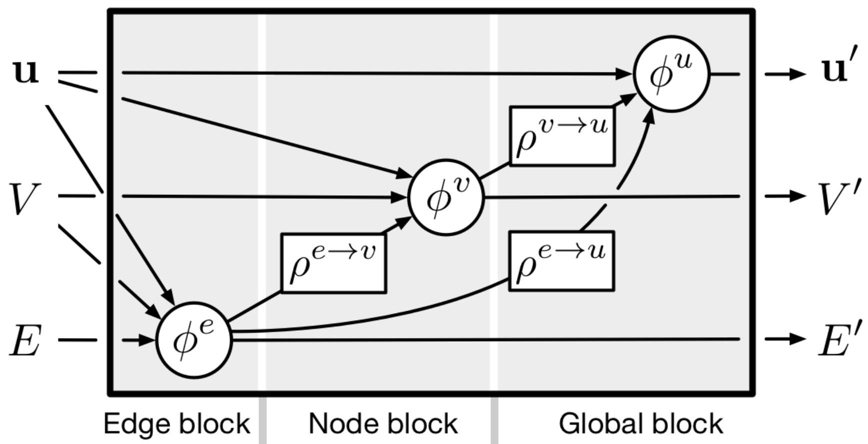
The general GNN

arXiv : 1806.01261

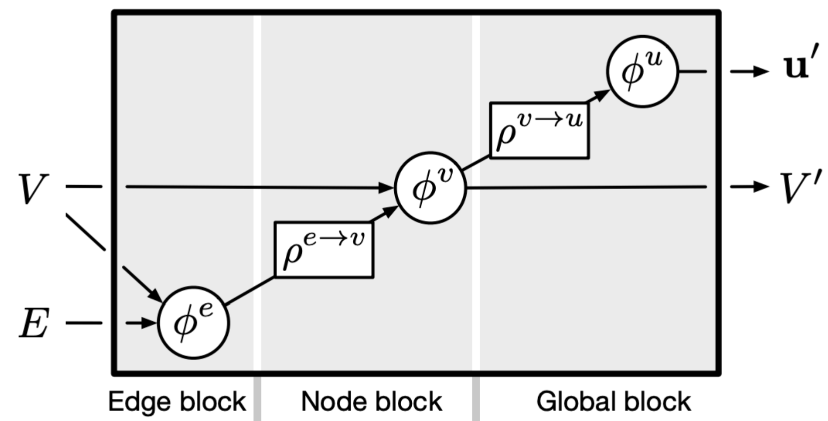


The general GNN

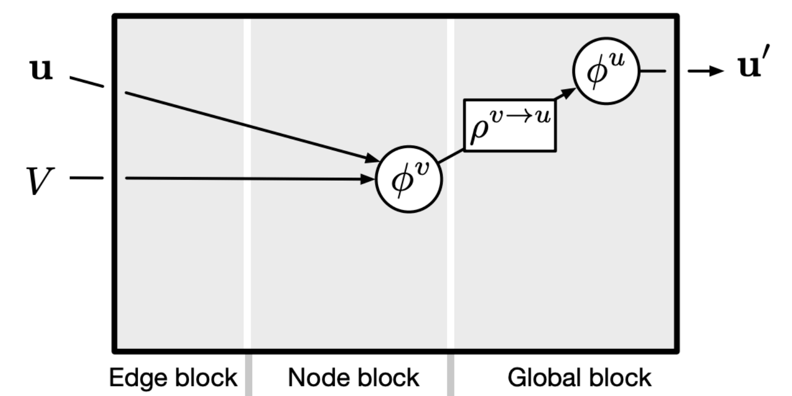
arXiv : 1806.01261



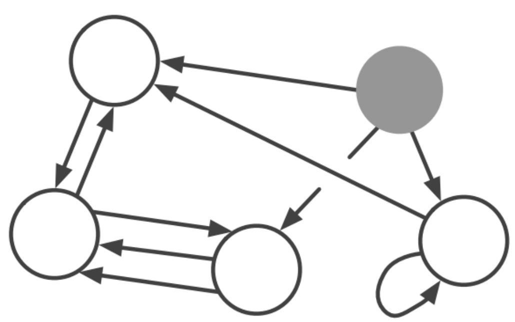
Full GN block



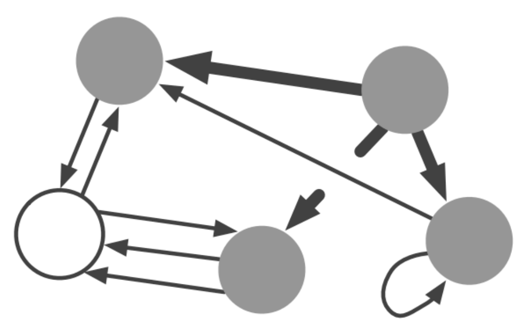
MPNN Layer



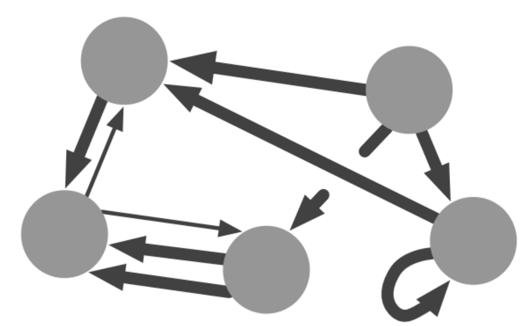
Deep-set layer



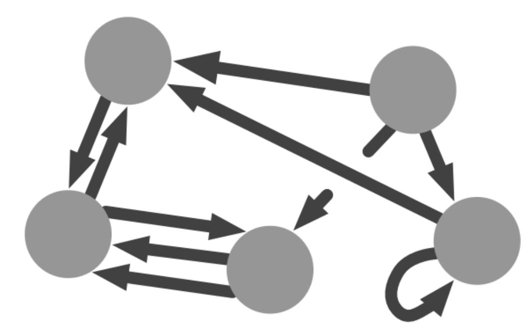
$m = 0$



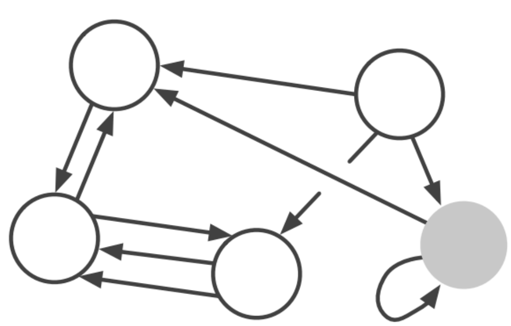
$m = 1$



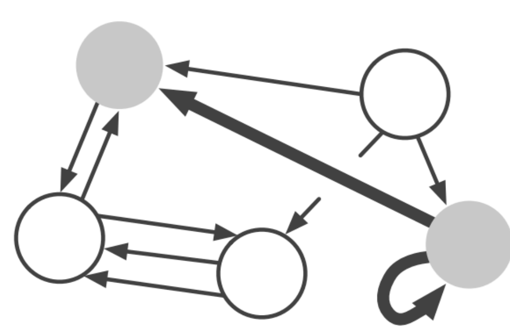
$m = 2$



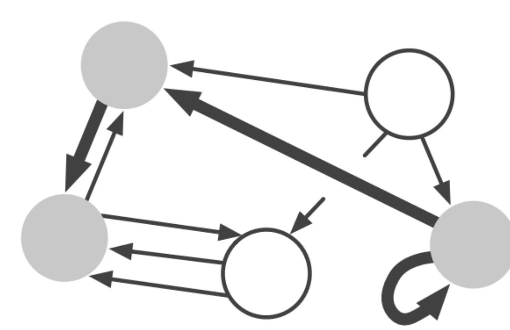
$m = 3$



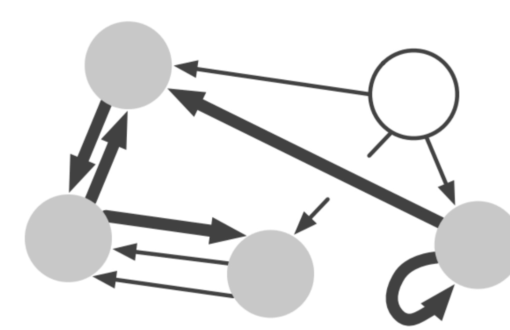
$m = 0$



$m = 1$



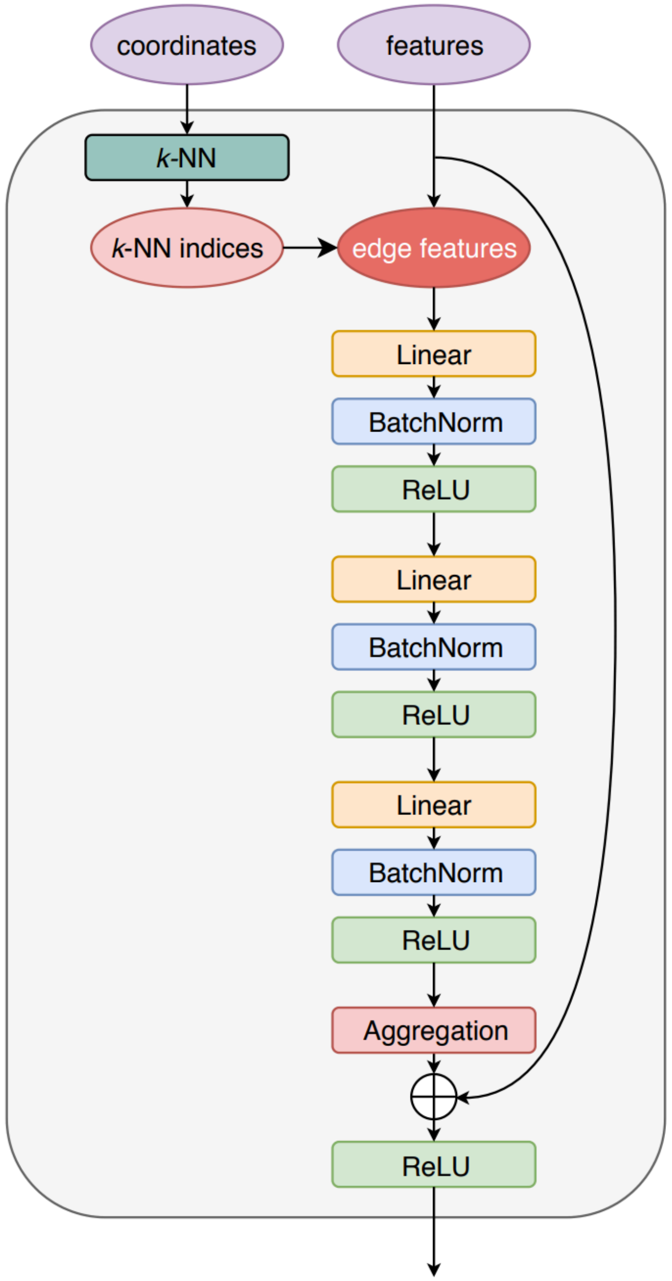
$m = 2$

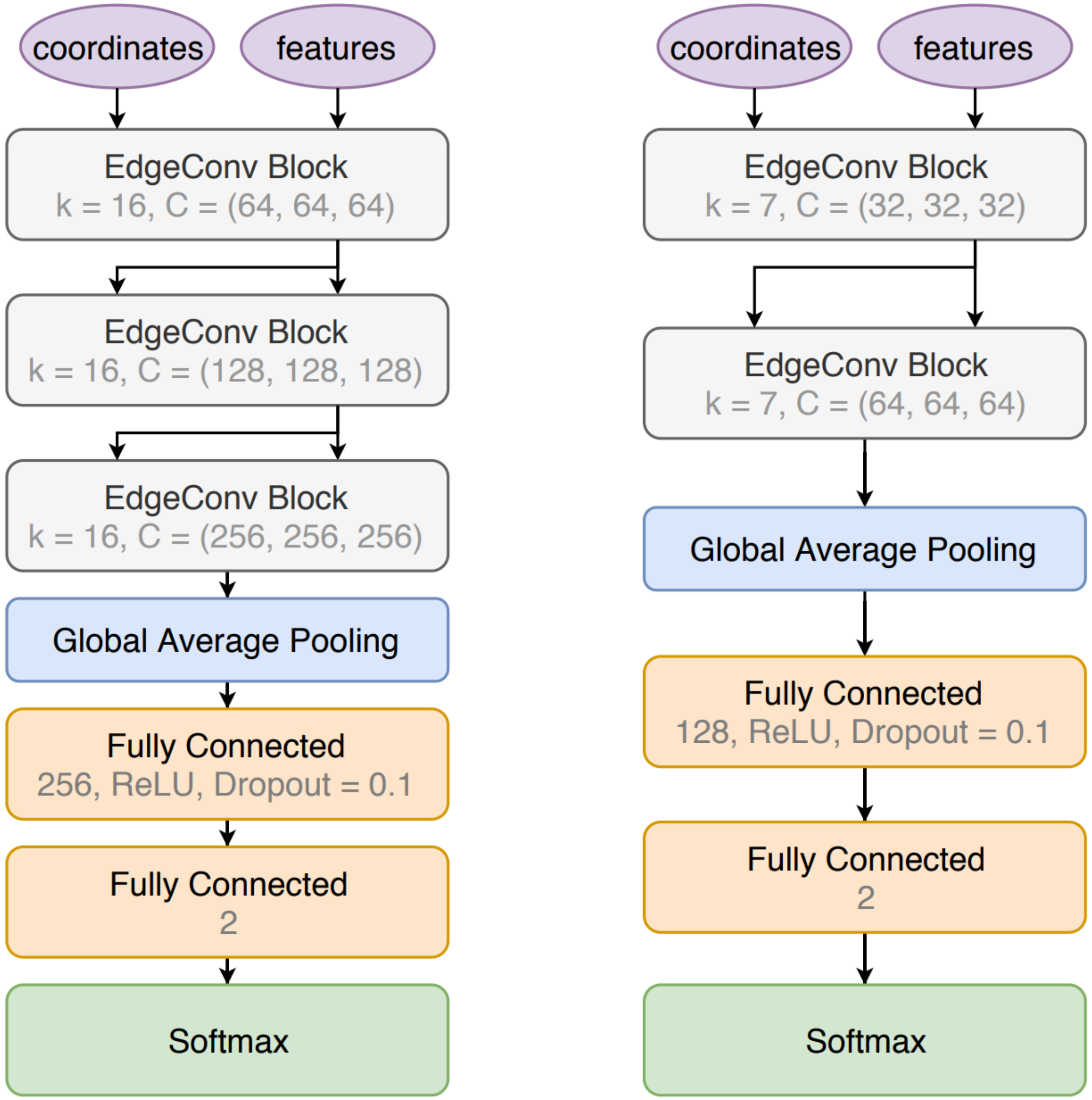


$m = 3$

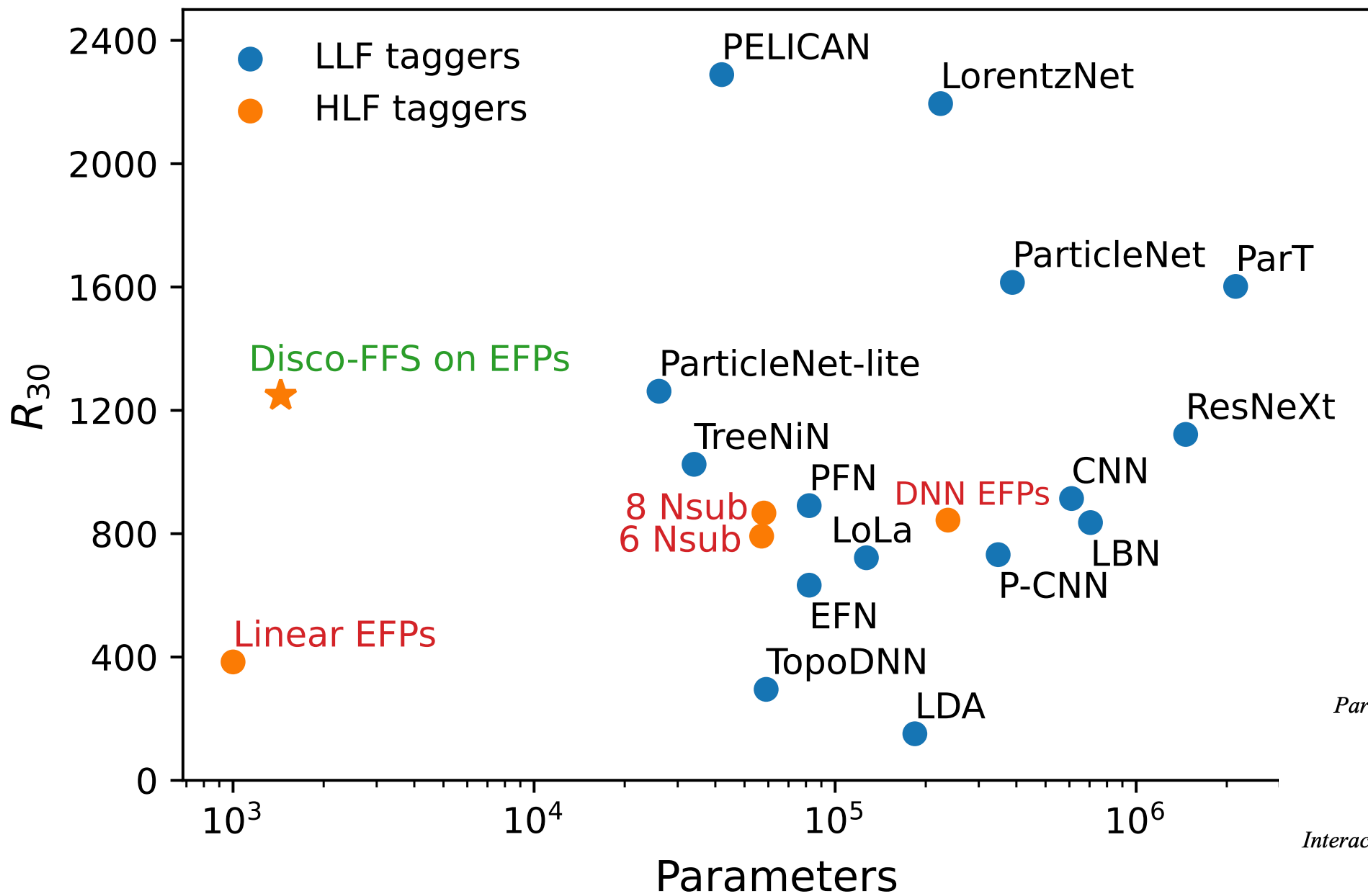
An example GNN in use

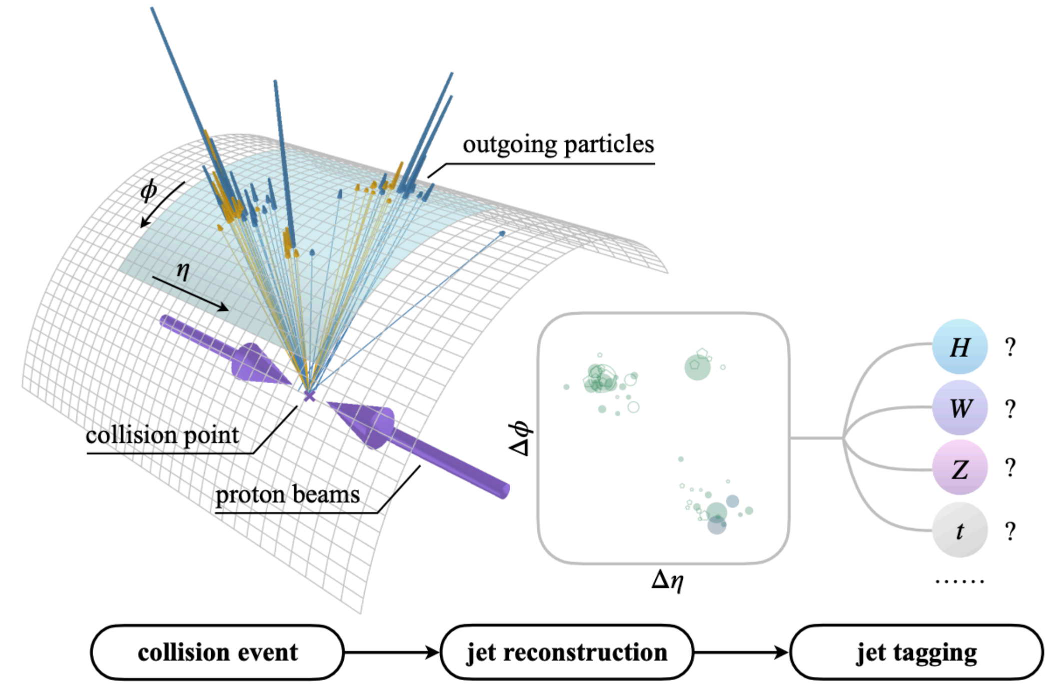
https://arxiv.org/pdf/1902.08570

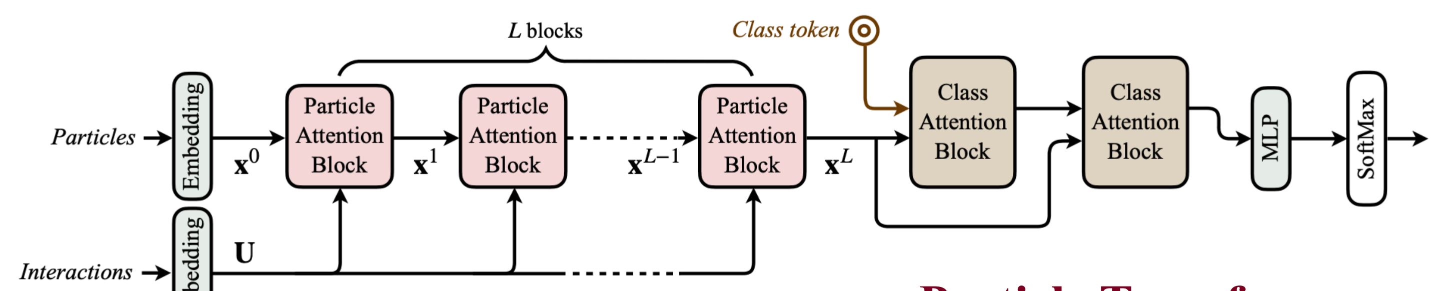




Object tagging

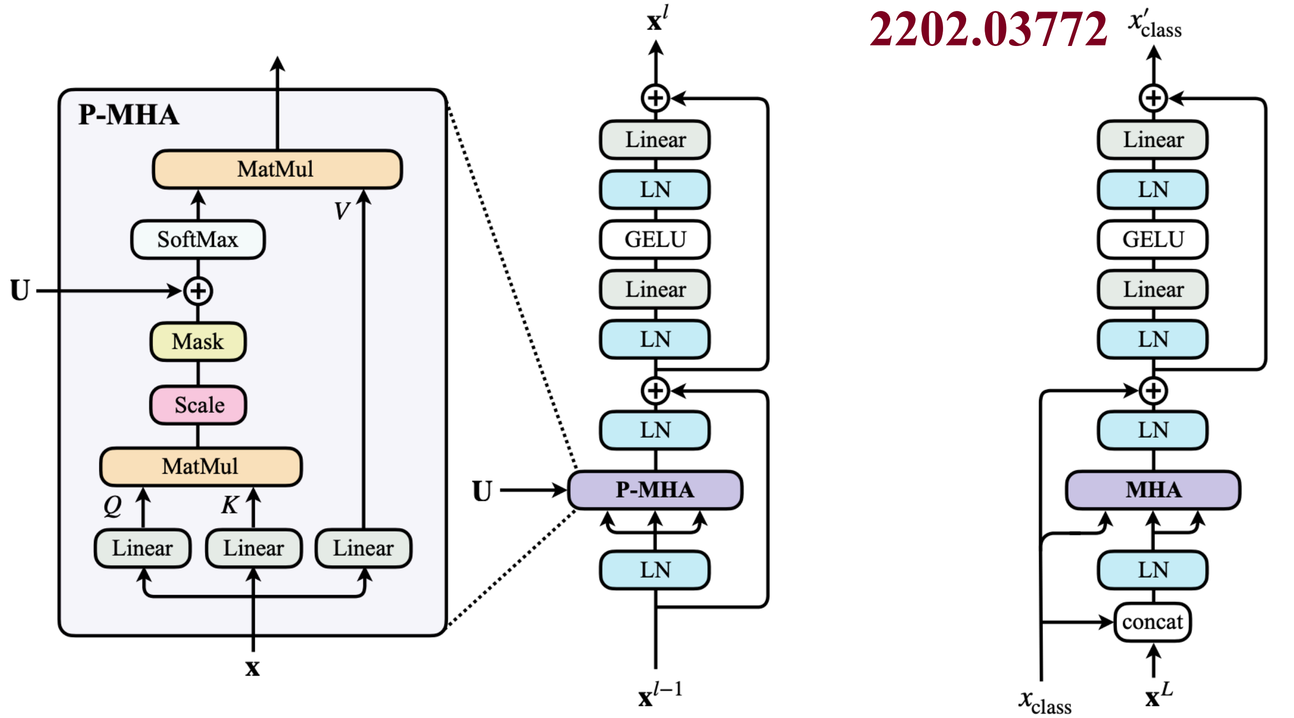






(a) Particle Transformer

Particle Transformer : 2202.03772



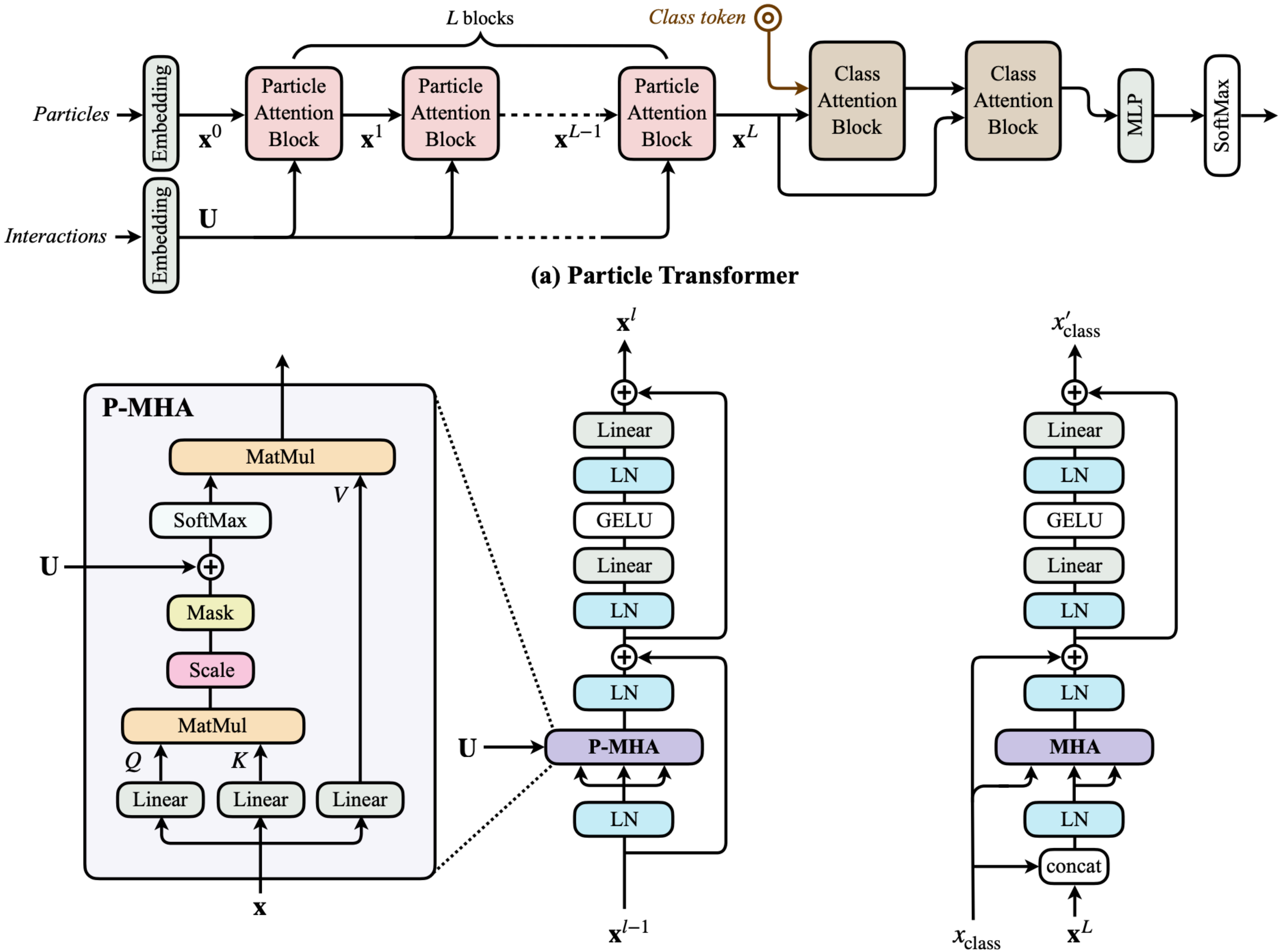
(b) Particle Attention Block

(c) Class Attention Block

	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**	$\mathbf{691 \pm 15}$	$\mathbf{2766 \pm 130}$

A modern classifier : ParT

Particle Transformer : 2202.03772



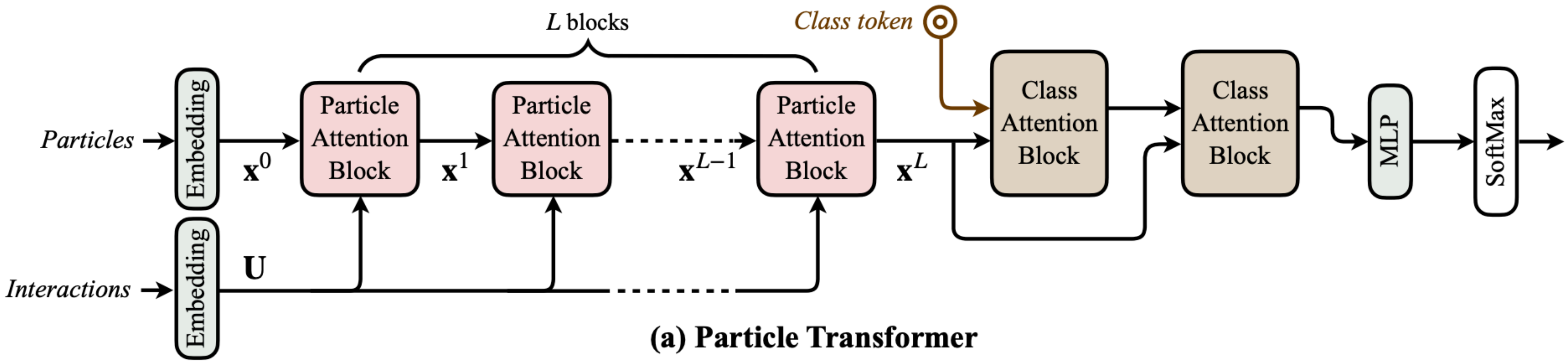
(a) Particle Transformer

(b) Particle Attention Block

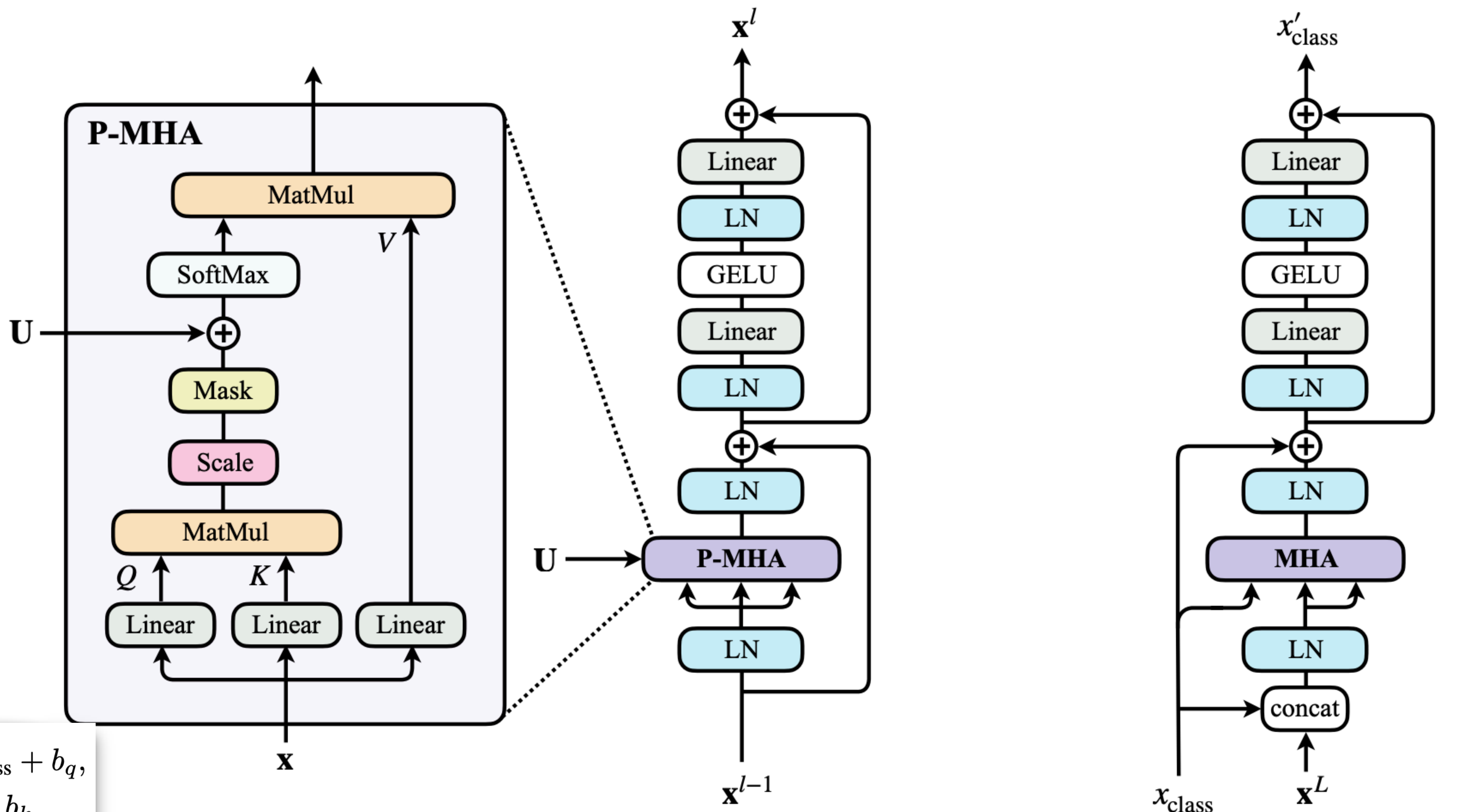
(c) Class Attention Block

A modern classifier : ParT

Particle Transformer : 2202.03772



(a) Particle Transformer

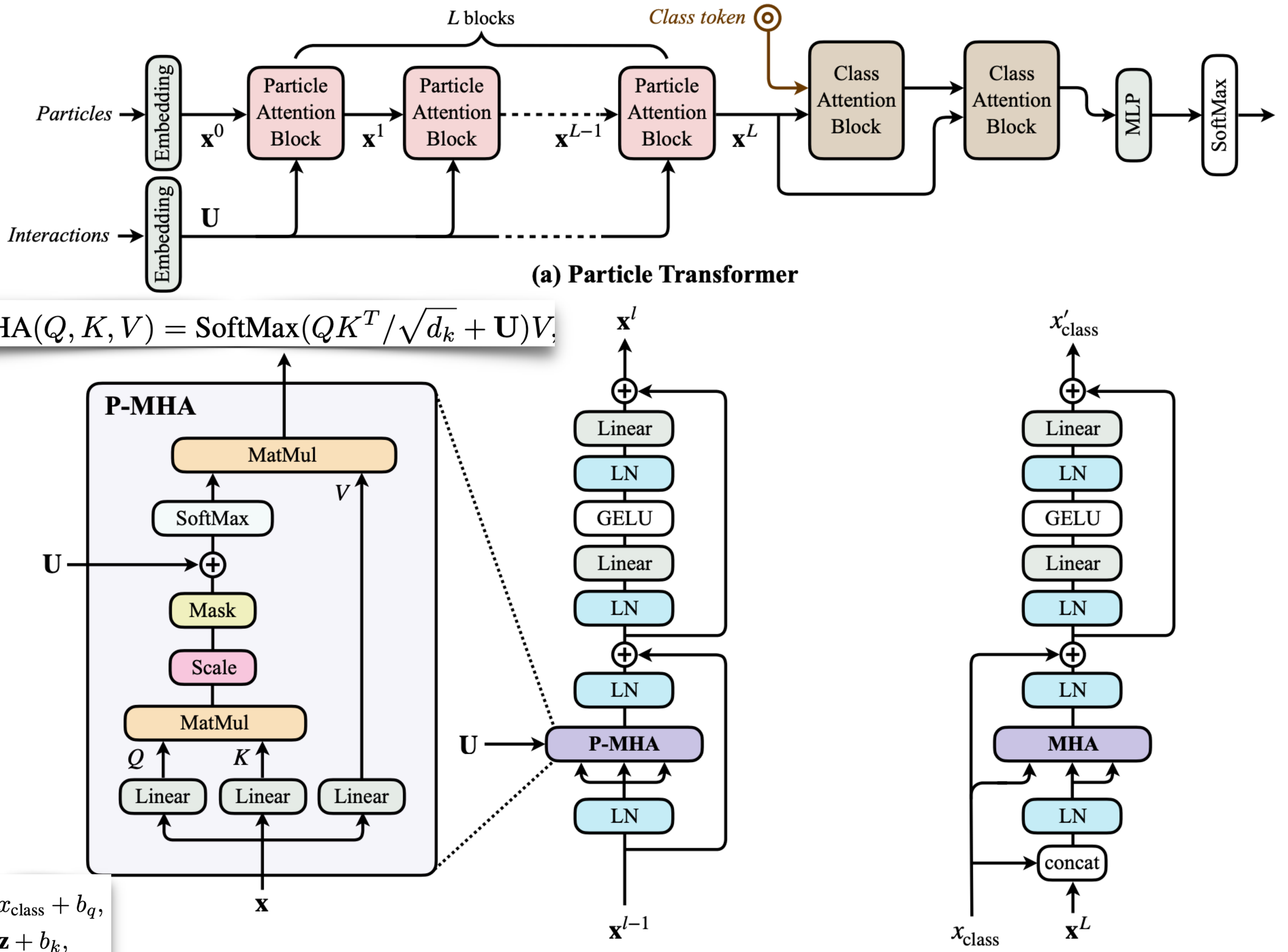


$$Q = W_q x_{\text{class}} + b_q,$$
$$K = W_k \mathbf{z} + b_k,$$
$$V = W_v \mathbf{z} + b_v,$$

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

A modern classifier : ParT

Particle Transformer : 2202.03772



(a) Particle Transformer

$$\text{P-MHA}(Q, K, V) = \text{SoftMax}(QK^T/\sqrt{d_k} + \mathbf{U})V,$$

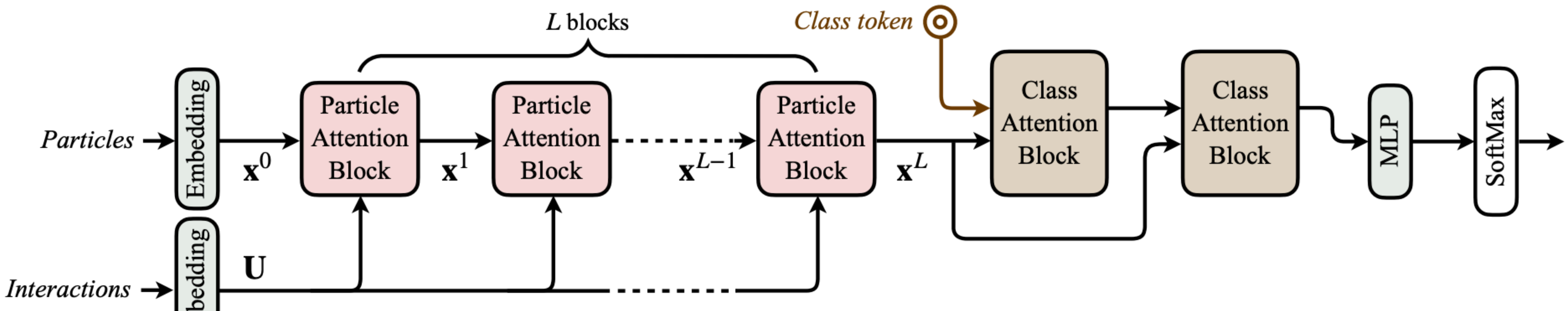
$$Q = W_q x_{\text{class}} + b_q,$$
$$K = W_k \mathbf{z} + b_k,$$
$$V = W_v \mathbf{z} + b_v,$$

(b) Particle Attention Block

(c) Class Attention Block

A modern classifier : ParT

Particle Transformer : 2202.03772



(a) Particle Transformer

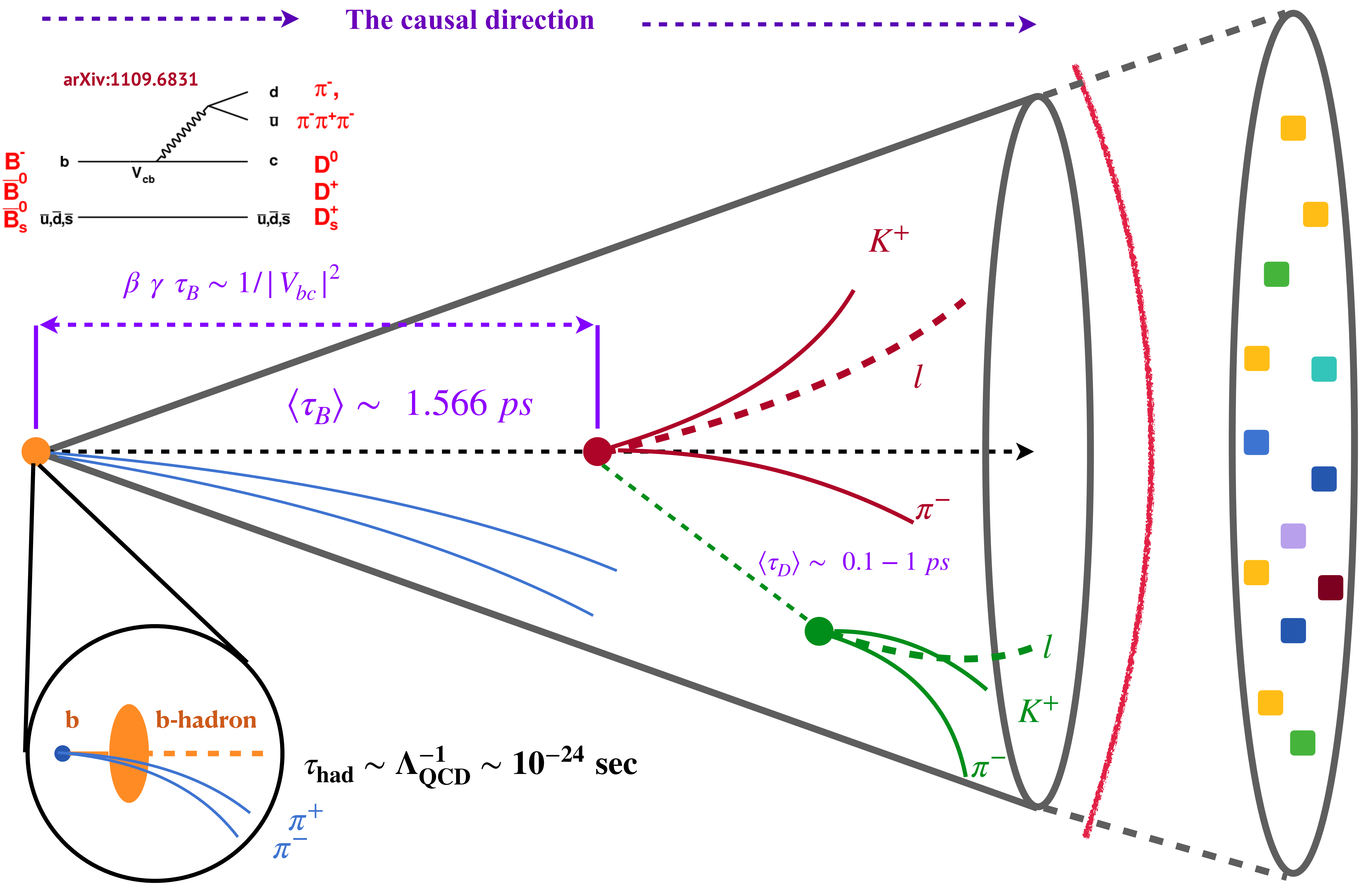
Category	Variable	Definition	JetClass	Top	QG_{exp}	QG_{full}
Kinematics	$\Delta\eta$	difference in pseudorapidity η between the particle and the jet axis	✓	✓	✓	✓
	$\Delta\phi$	difference in azimuthal angle ϕ between the particle and the jet axis	✓	✓	✓	✓
	$\log p_T$	logarithm of the particle's transverse momentum p_T	✓	✓	✓	✓
	$\log E$	logarithm of the particle's energy	✓	✓	✓	✓
	$\log \frac{p_T}{p_T(\text{jet})}$	logarithm of the particle's p_T relative to the jet p_T	✓	✓	✓	✓
	$\log \frac{E}{E(\text{jet})}$	logarithm of the particle's energy relative to the jet energy	✓	✓	✓	✓
	ΔR	angular separation between the particle and the jet axis ($\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$)	✓	✓	✓	✓
Particle identification	`charge`	electric charge of the particle	✓	—	✓	✓
	`Electron`	if the particle is an electron (`|pid|==11`)	✓	—	✓	✓
	`Muon`	if the particle is an muon (`|pid|==13`)	✓	—	✓	✓
	`Photon`	if the particle is an photon (`pid==22`)	✓	—	✓	✓
	`CH`	if the particle is an charged hadron (`|pid|==211 or 321 or 2212`)	✓	—	✓	✓[a]
	`NH`	if the particle is an neutral hadron (`|pid|==130 or 2112 or 0`)	✓	—	✓	✓[b]
Trajectory displacement	$\tanh d_0$	hyperbolic tangent of the transverse impact parameter value	✓	—	—	—
	$\tanh d_z$	hyperbolic tangent of the longitudinal impact parameter value	✓	—	—	—
	σ_{d_0}	error of the measured transverse impact parameter	✓	—	—	—
	σ_{d_z}	error of the measured longitudinal impact parameter	✓	—	—	—

$$Q = W_q x_{\text{class}} + b_q,$$
$$K = W_k \mathbf{z} + b_k,$$
$$V = W_v \mathbf{z} + b_v,$$

$\mathbf{x}$ $\mathbf{x}^{l-1}$ x_{class} $\mathbf{x}^L$

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

Physics motivated ML: B-tagging



Are all these useful ?

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

nature > nature communications > articles > article

Article | Open access | Published: 14 January 2026

Transforming jet flavour tagging at ATLAS

The ATLAS Collaboration

S. Ganguly et al

Nature Communications **17**, Article number: 541 (2026) | Cite this article

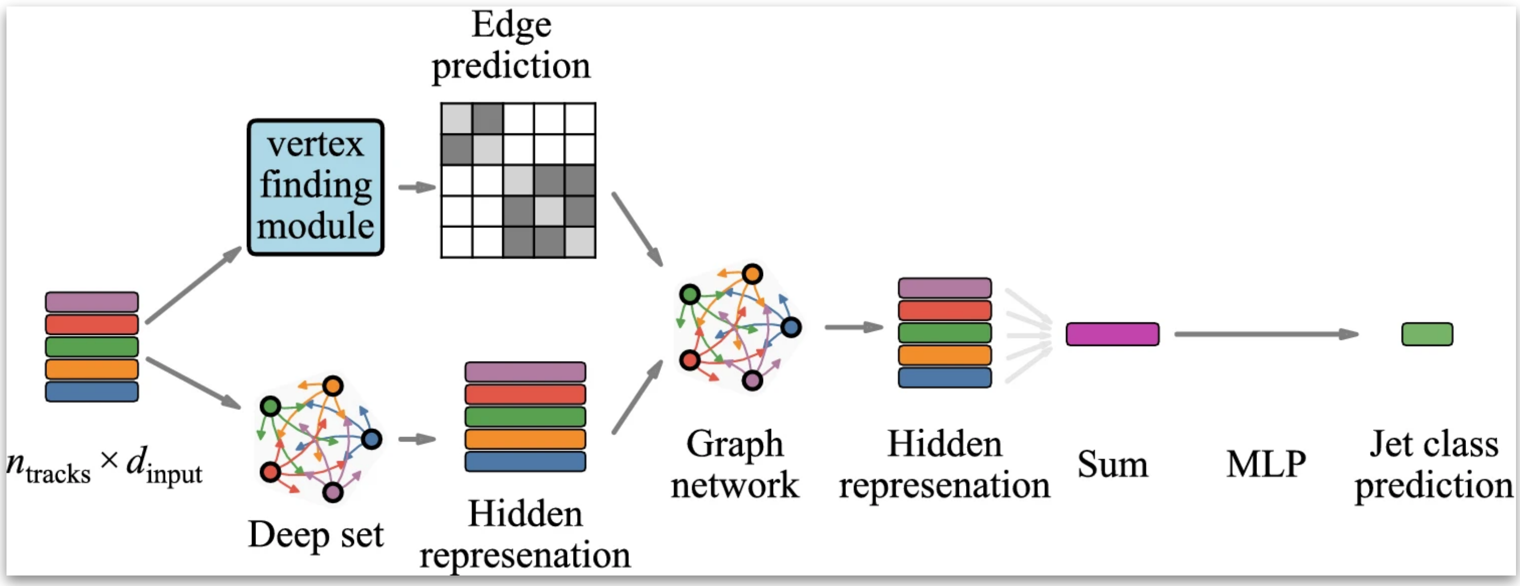
Are all these useful ?

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



nature › nature communications › articles › article

Article | Open access | Published: 14 January 2026

Transforming jet flavour tagging at ATLAS

The ATLAS Collaboration

S. Ganguly et al

Nature Communications **17**, Article number: 541 (2026) | Cite this article

Are all these useful ?

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

nature > nature communications > articles > article

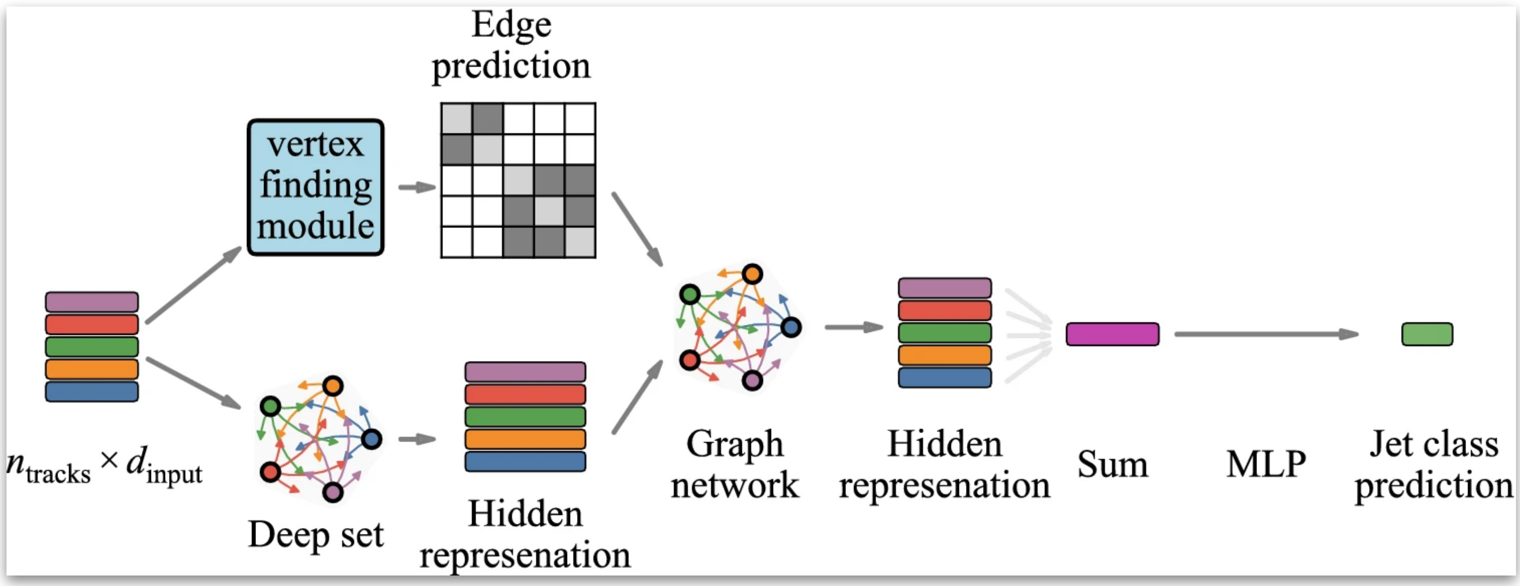
Article | Open access | Published: 14 January 2026

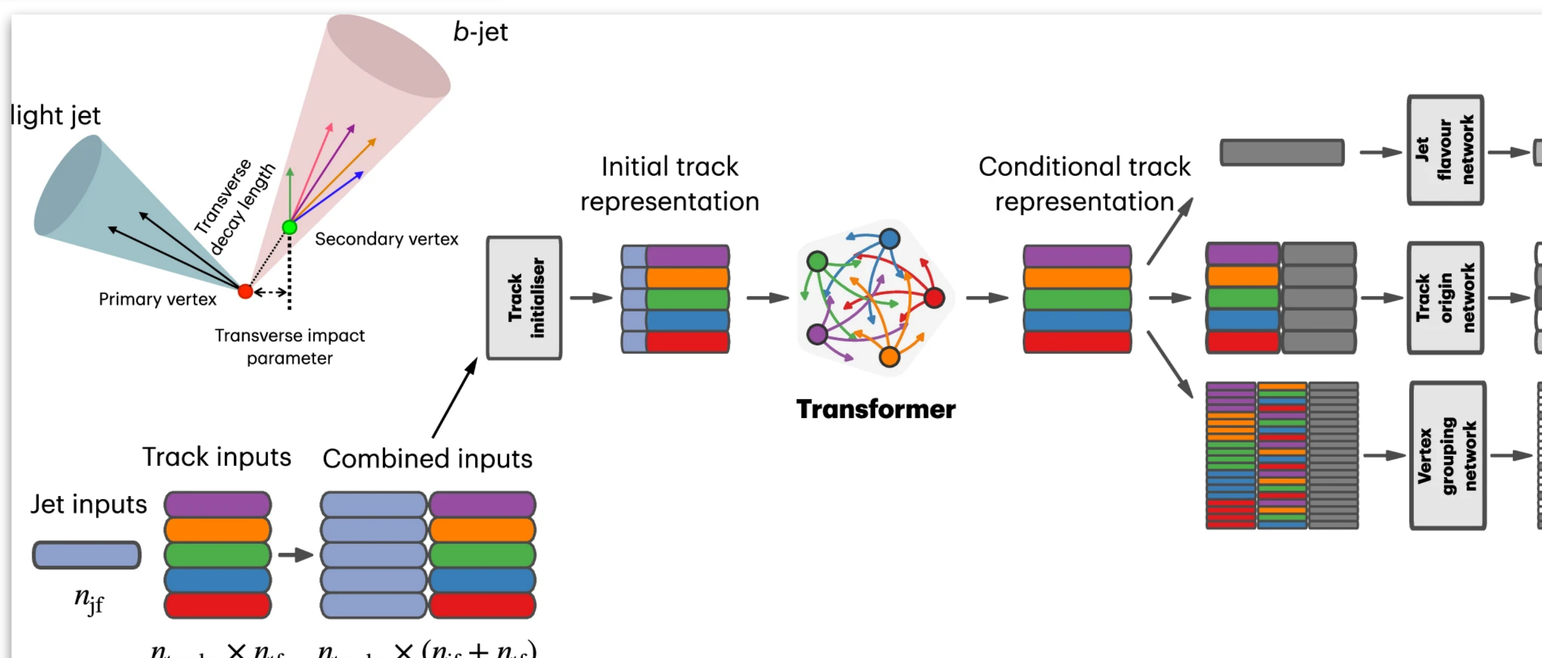
Transforming jet flavour tagging at ATLAS

The ATLAS Collaboration

S. Ganguly et al

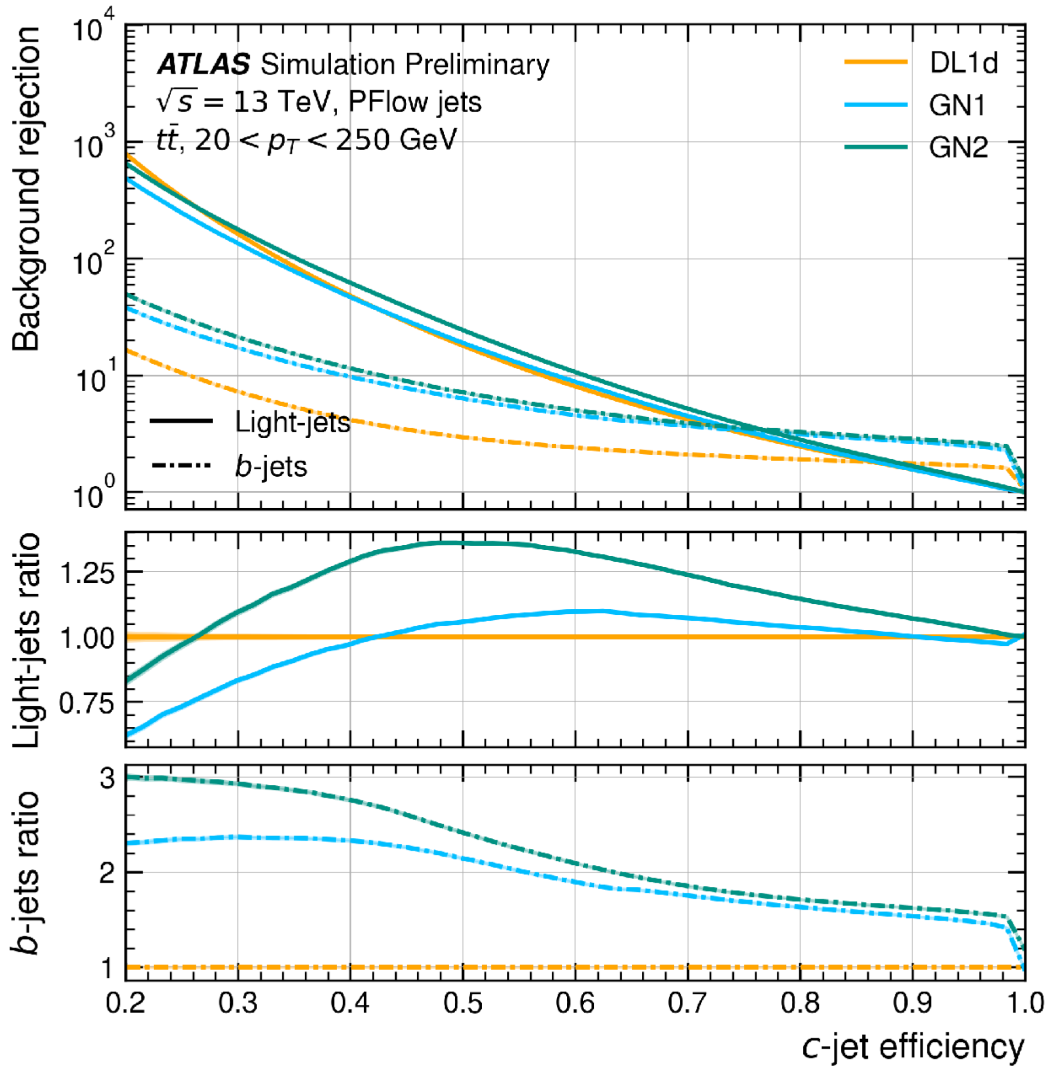
Nature Communications 17, Article number: 541 (2026) | Cite this article

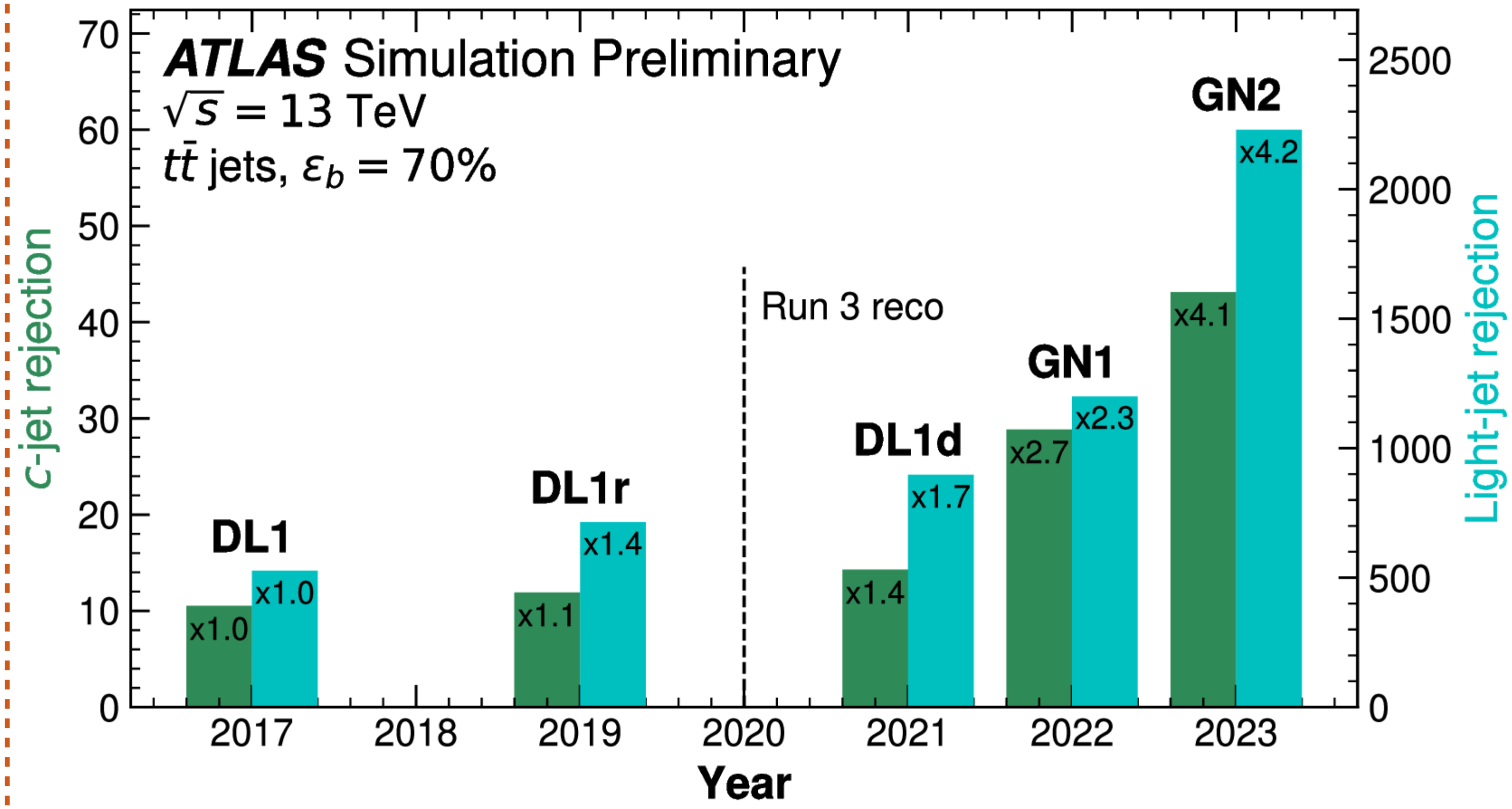


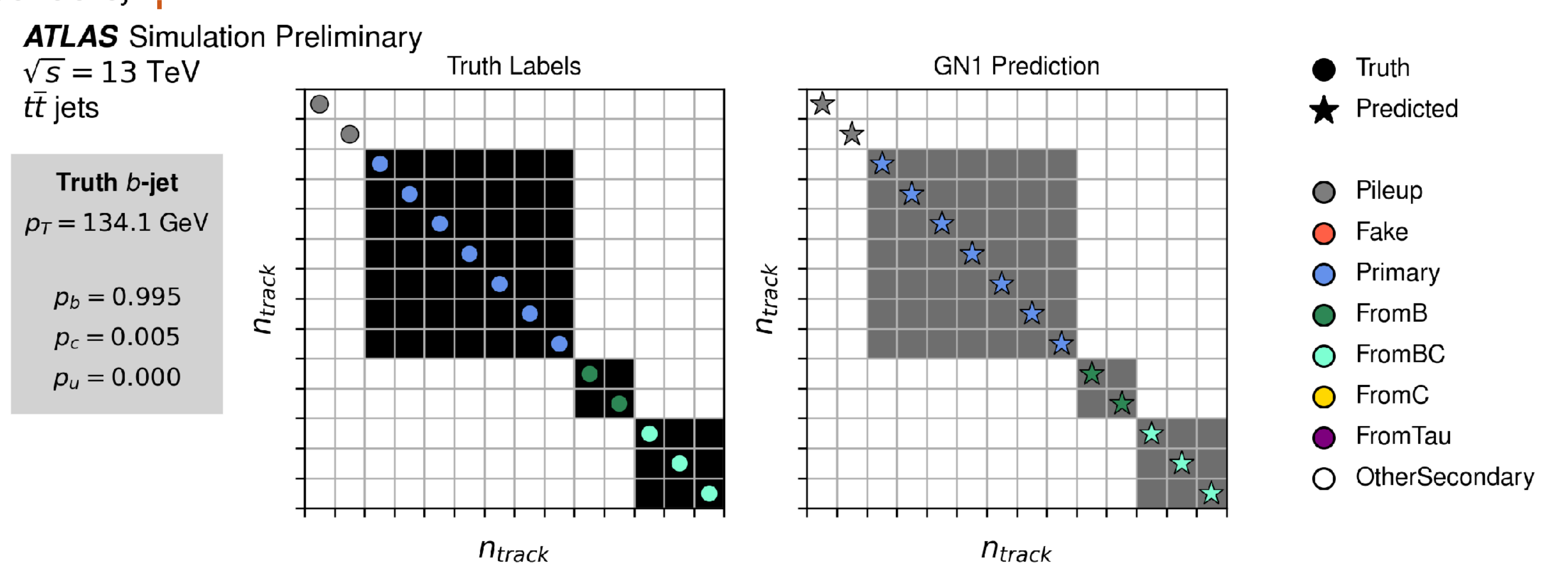


Set2Graph model within ATLAS

FTAG-2023-001





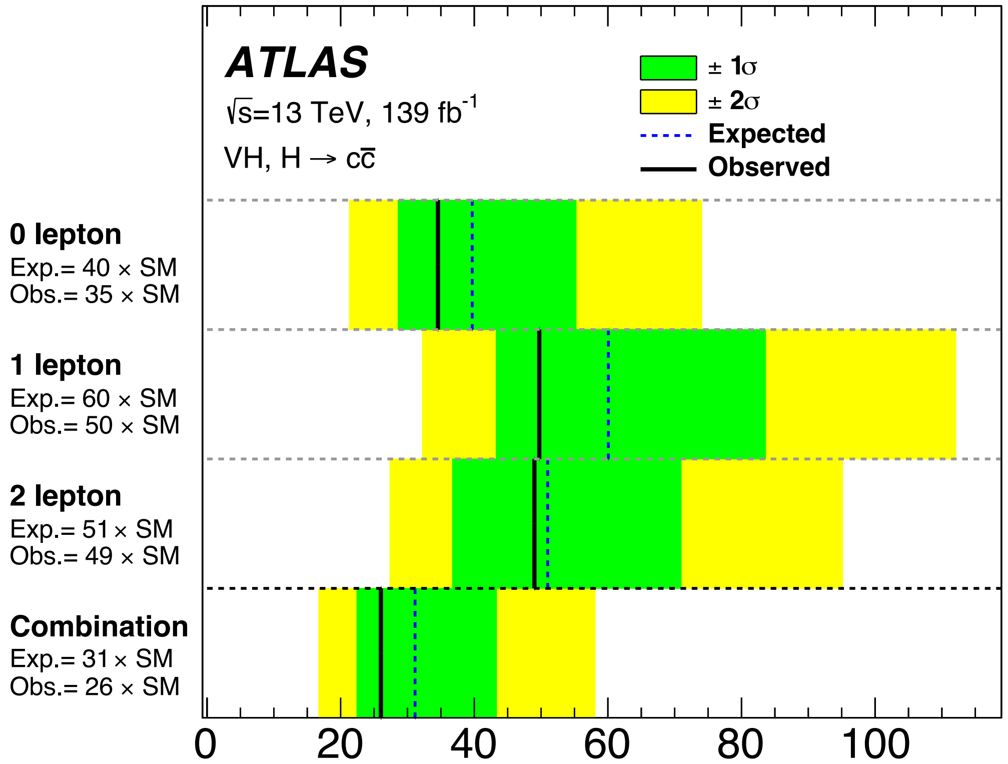


Sizable improvement over the current DL1r algorithm.

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

Direct physics application of the taggers

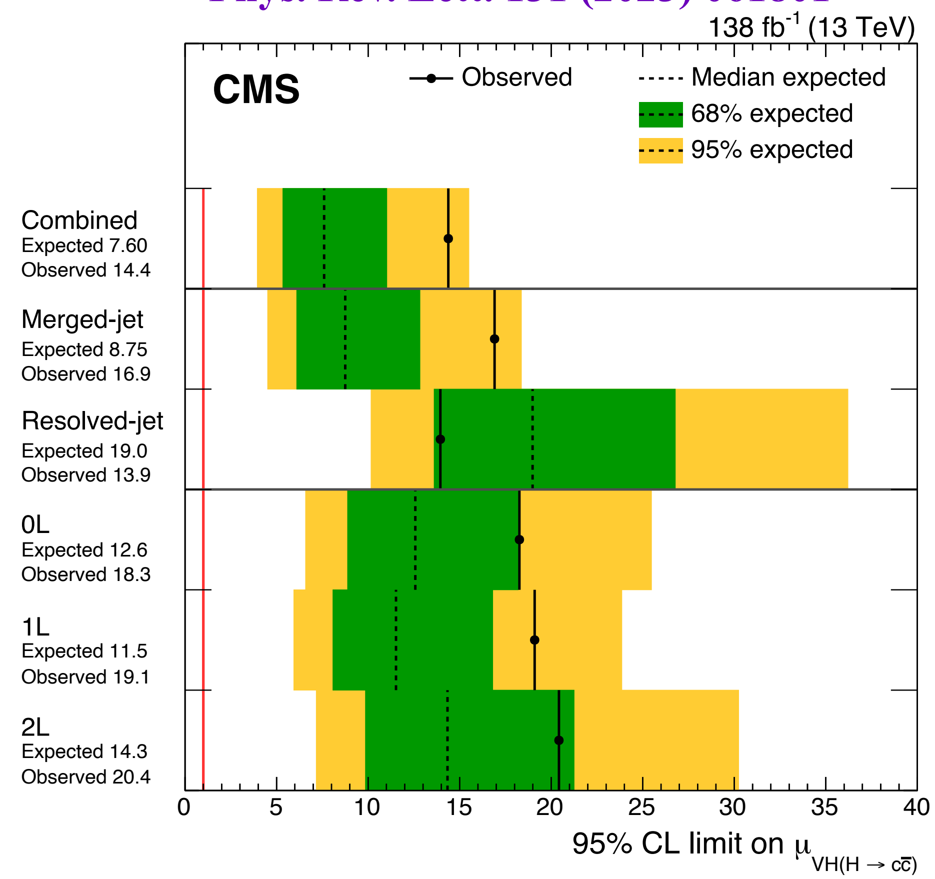
Eur. Phys. J. C (2022) 82:717

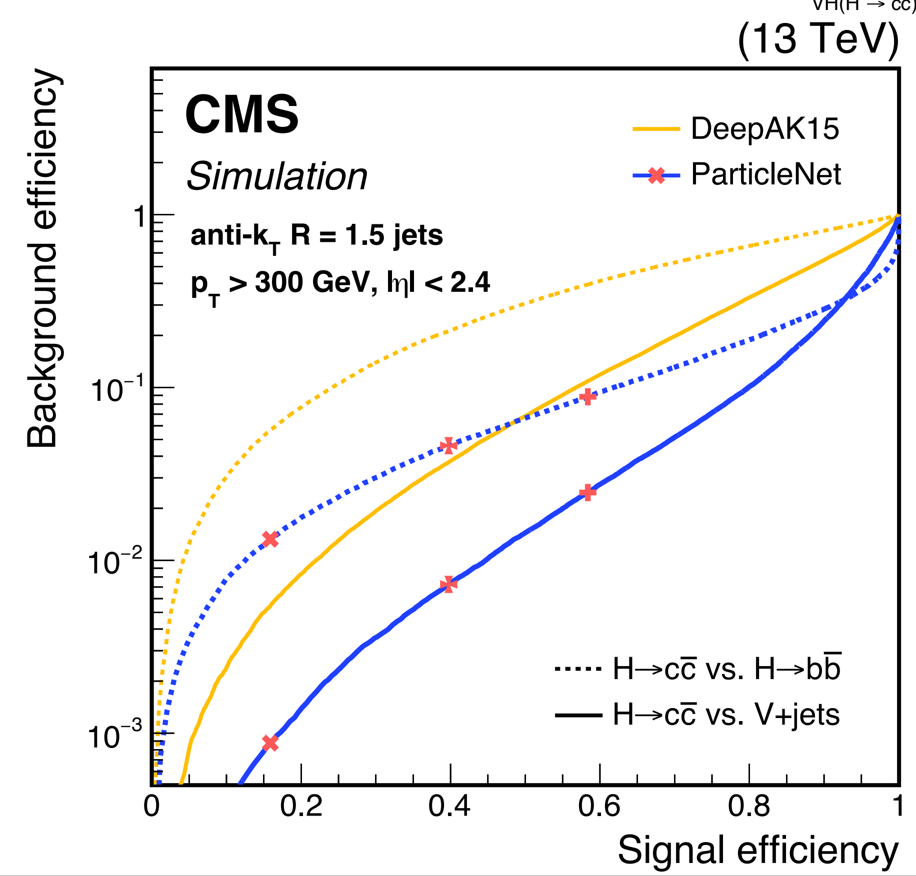


DL1r based taggers

Future iteration will use GN2

Phys. Rev. Lett. 131 (2023) 061801





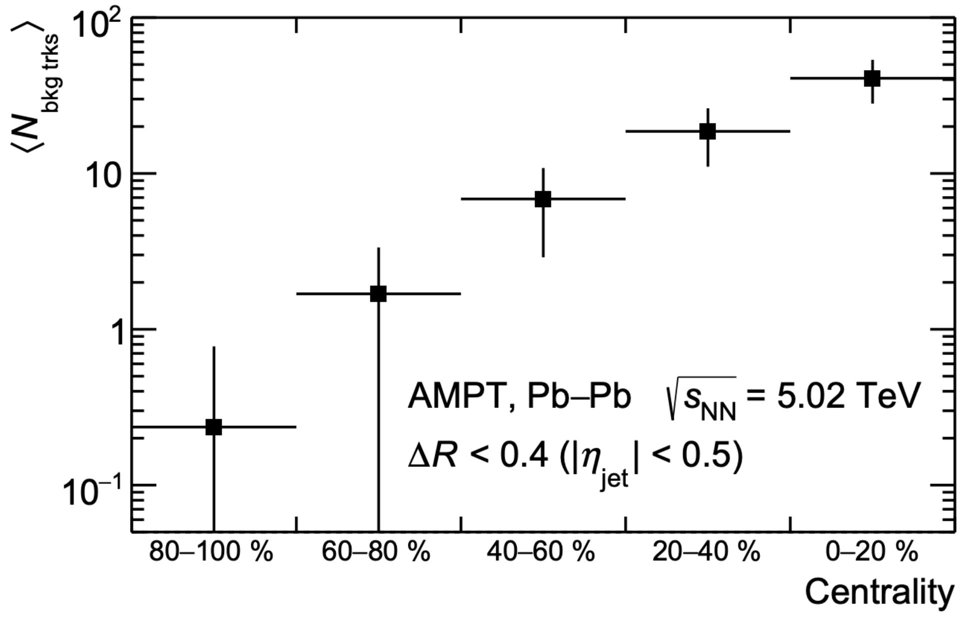
ATLAS bound : $|\kappa_c| < 8.5$

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?**

Same architecture in Heavy Ion

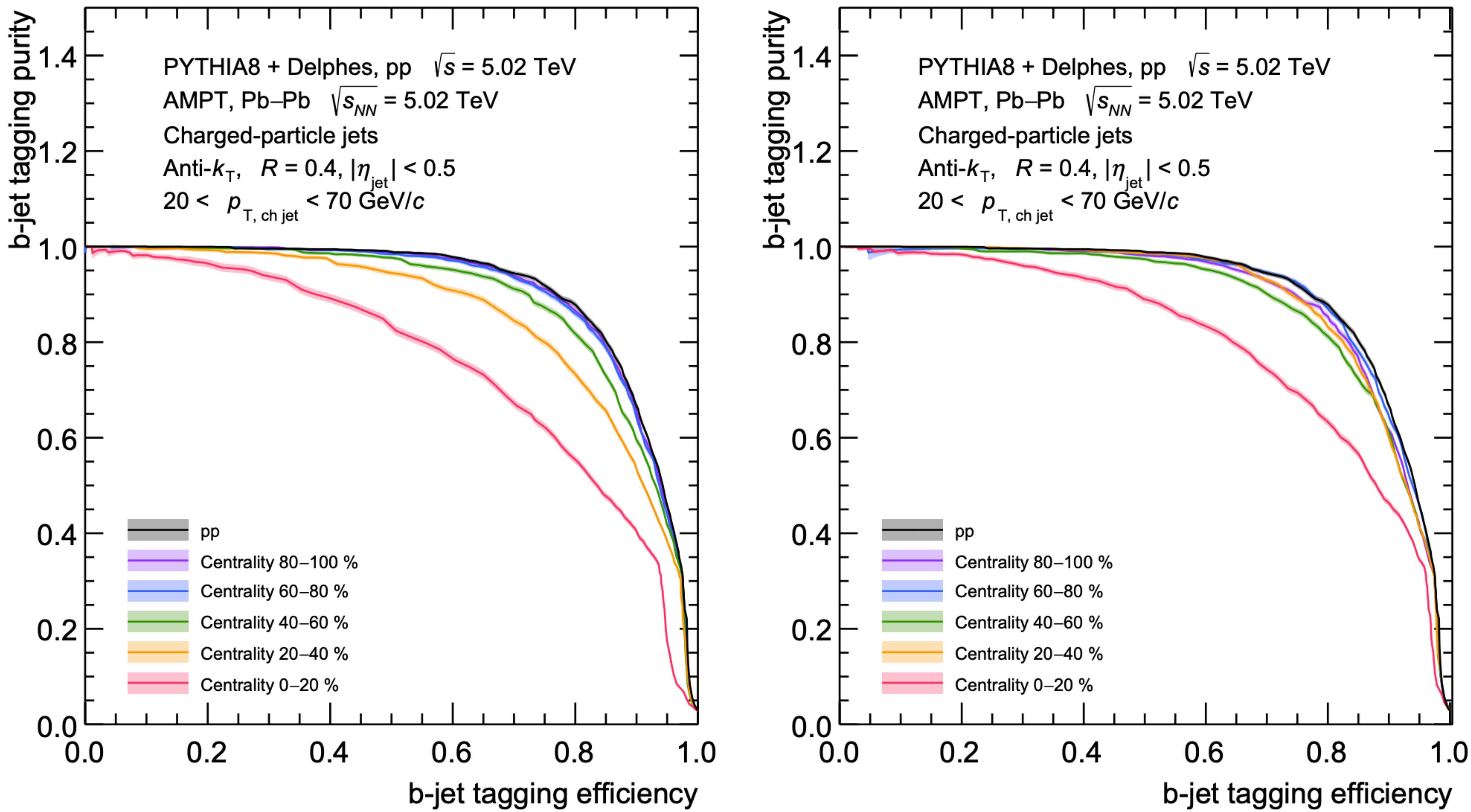


Jet input	Description
$p_{T,\text{ jet}}$	Charged-particle jet transverse momentum
η_{jet}	Jet pseudorapidity
ϕ_{jet}	Jet azimuthal angle
m_{jet}	Jet invariant mass
Track input	**Description**
p_{T}	Track transverse momentum
η	Track pseudorapidity
ϕ	Track azimuthal angle
q	Track charge
d_{xy}	Signed transverse impact parameter
d_z	Signed longitudinal impact parameter

Track origin	Description
Background	Background particles from AMPT
FromC	Charm quark or hadron decay
FromB	Beauty quark or hadron decay
Primary	Primary vertex except for c/b origins
OtherSecondary	Secondary vertices except for c/b origins

arXiv:2506.22691

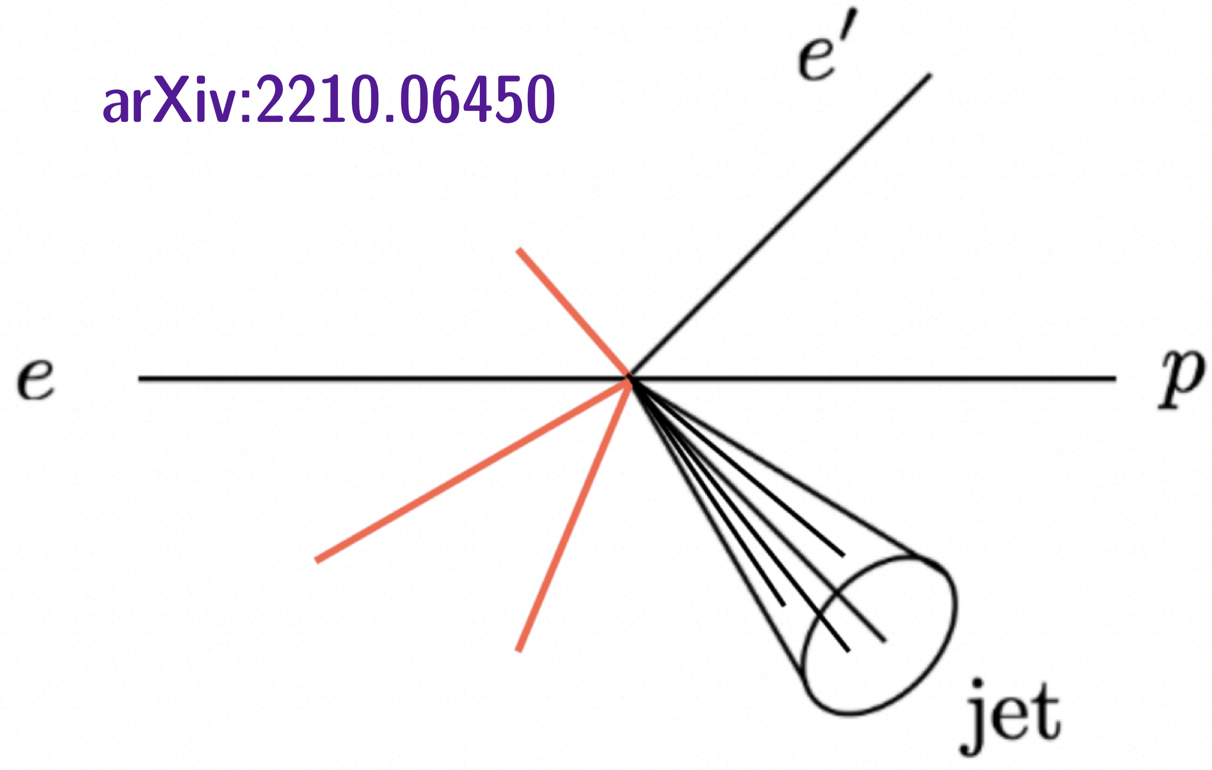
Same architecture in Heavy Ion

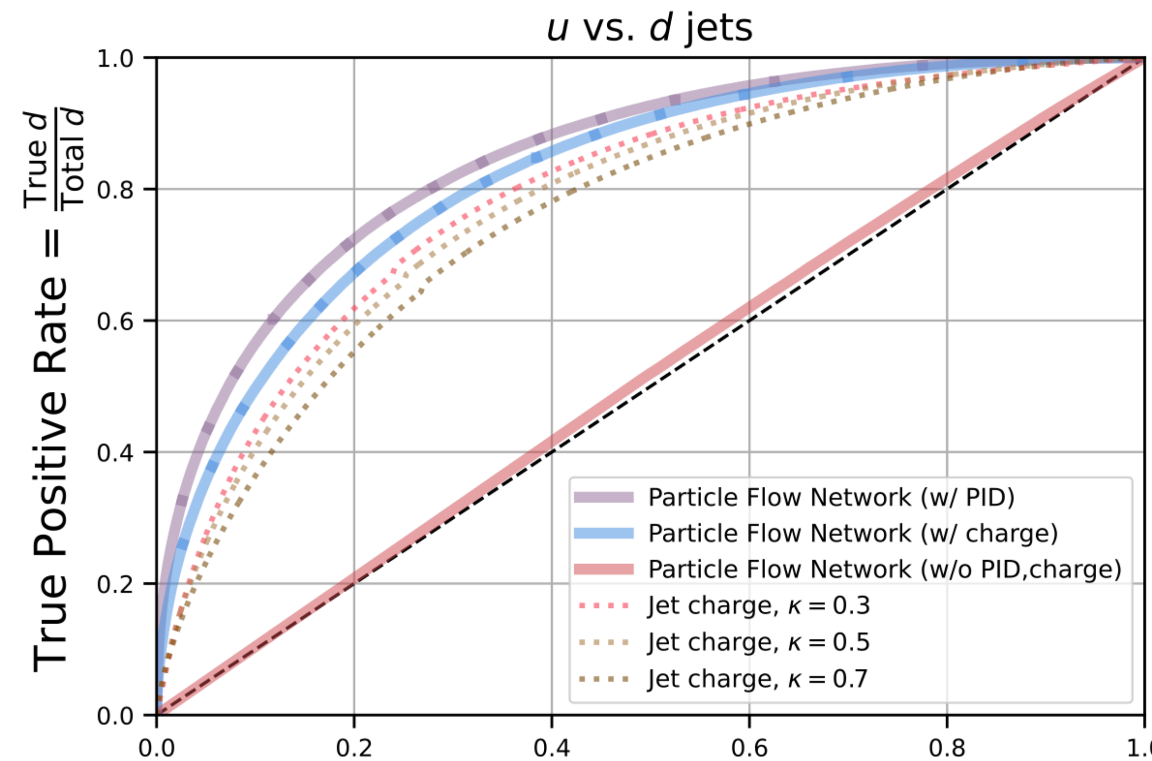


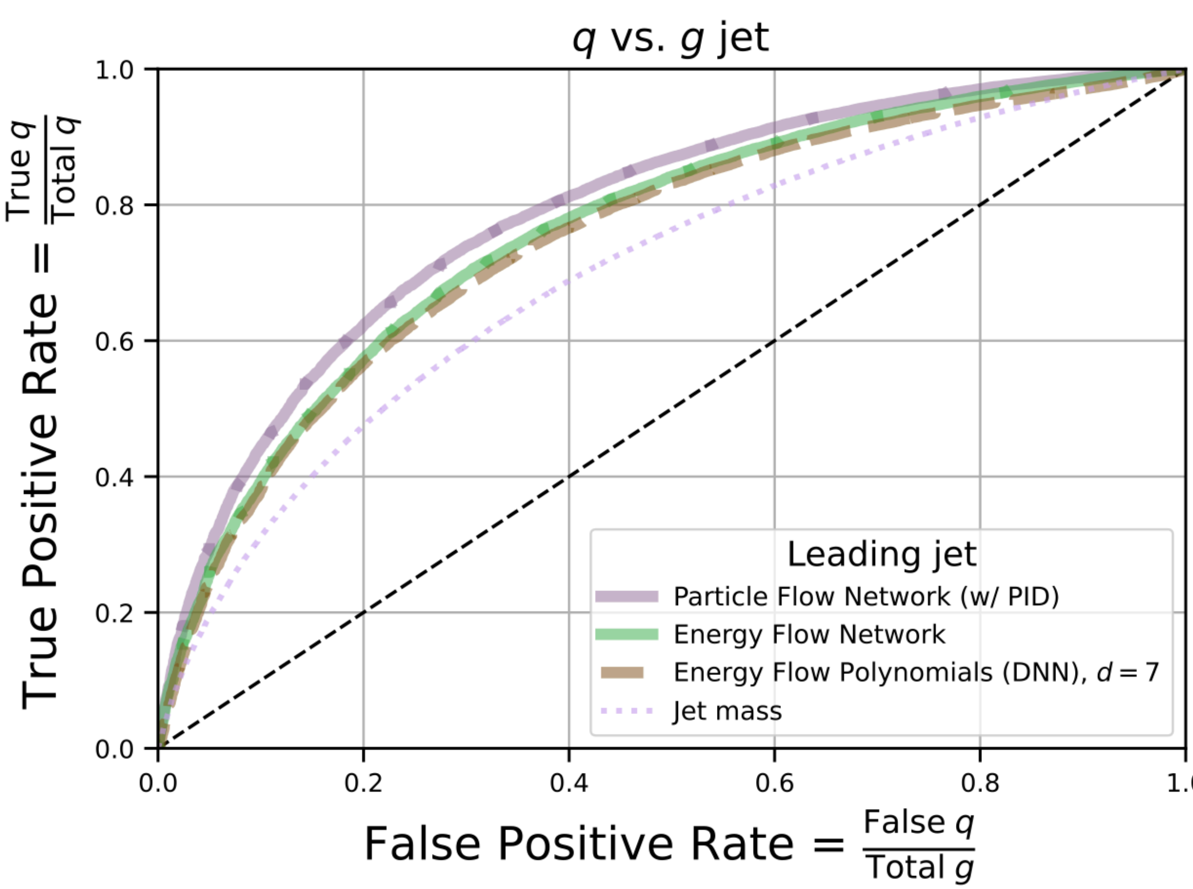
arXiv:2506.22691

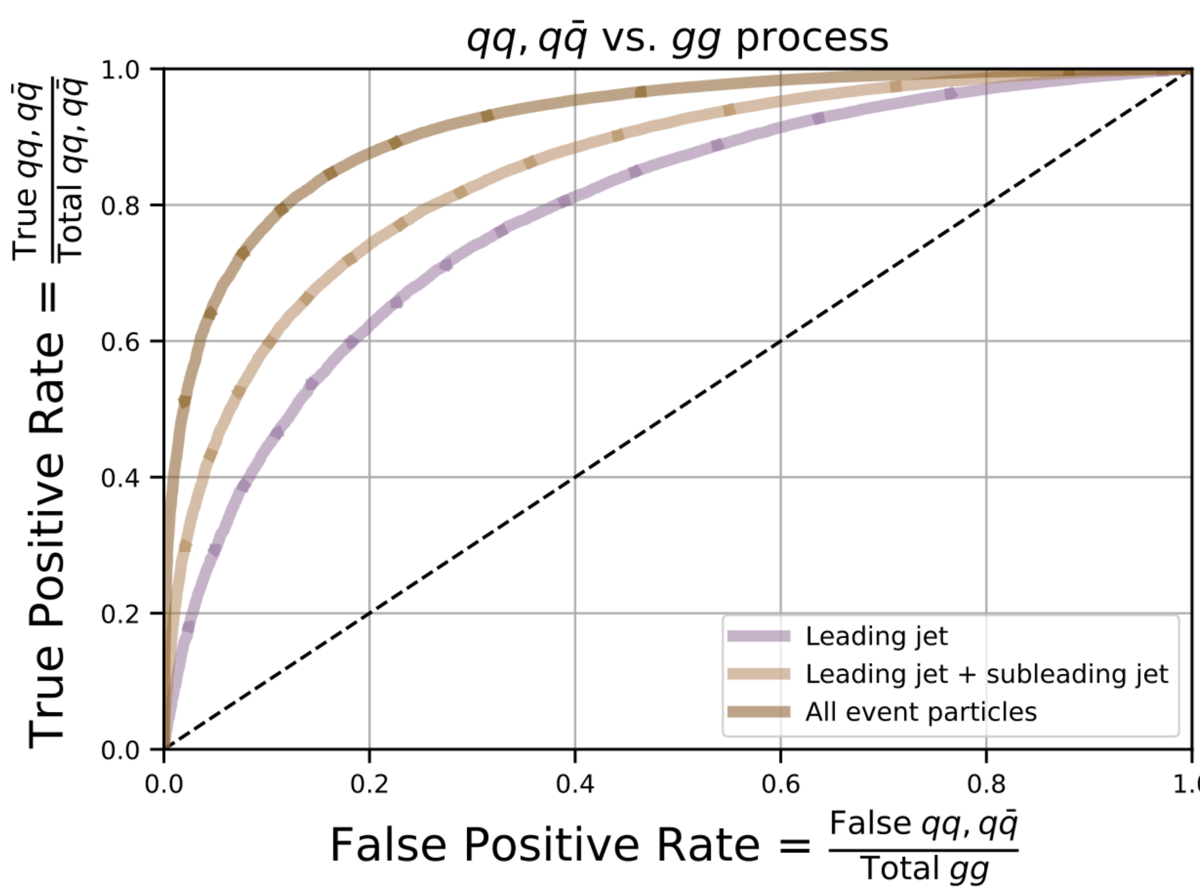
Classification performance at EIC

arXiv:2210.06450





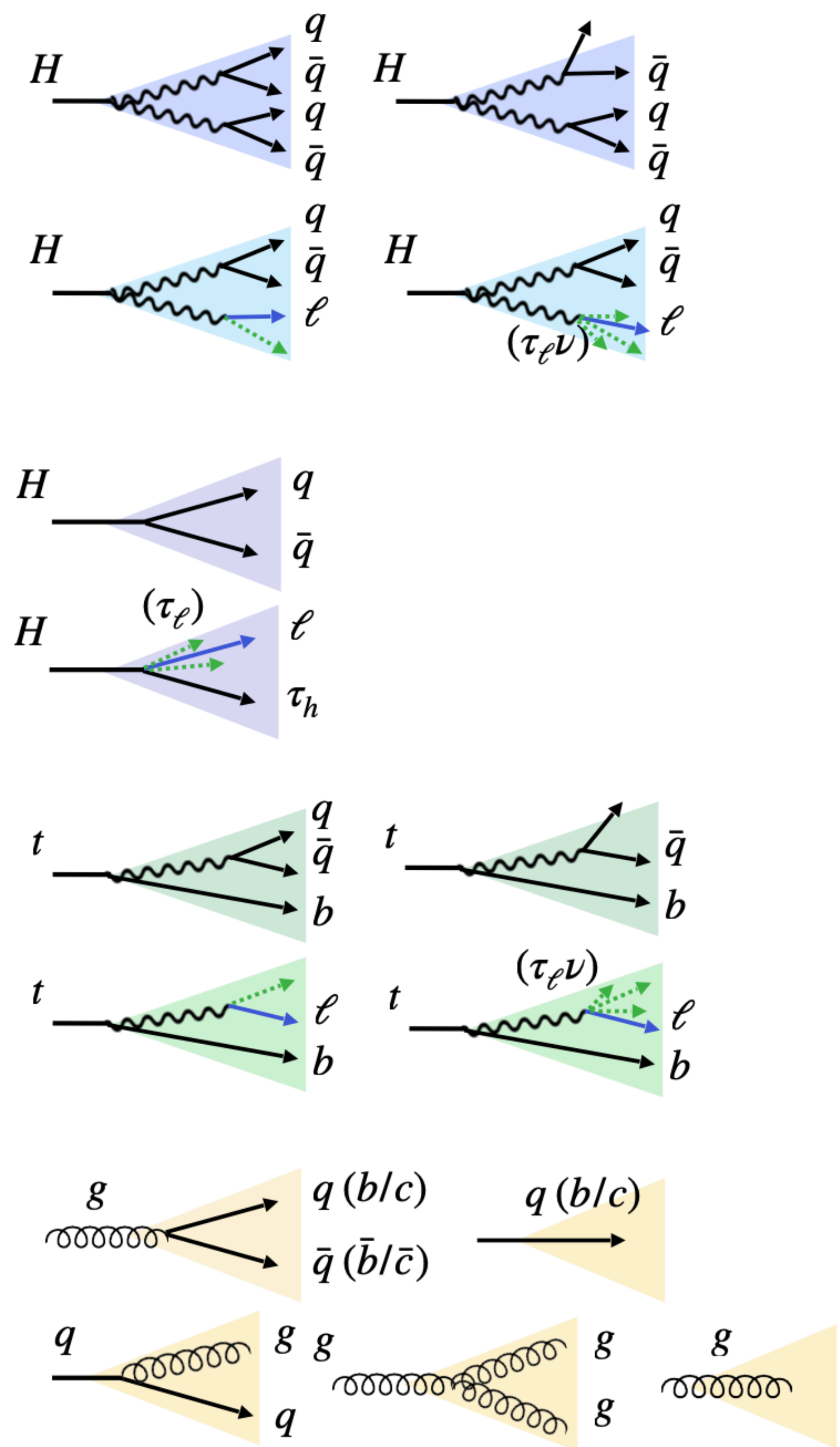




Even more grand : GloParT

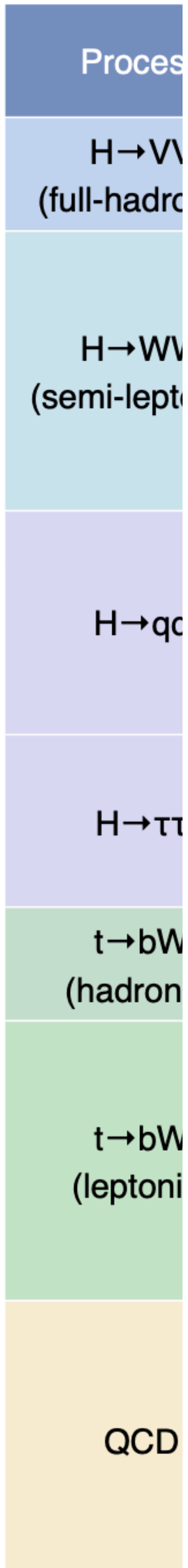
CMS PAS HIG-23-012

Process	Final state/ prongness	heavy flavour	# of classes
H→VV (full-hadronic)	qqqq	0c/1c/2c	3
	qqq		3
H→WW (semi-leptonic)	eνqq	0c/1c	2
	μνqq		2
	τ_eνqq		2
	τ_μνqq		2
	τ_hνqq		2
H→qq		bb	1
		cc	1
		ss	1
		qq (q=u/d)	1
H→ττ	$\tau_e\tau_h$		1
	$\tau_\mu\tau_h$		1
	$\tau_h\tau_h$		1
t→bW (hadronic)	bqq	1b + 0c/1c	2
	bq		2
t→bW (leptonic)	beν	1b	1
	bμν		1
	bτ_eν		1
	bτ_μν		1
	bτ_hν		1
QCD		b	1
		bb	1
		c	1
		cc	1
		others (light)	1

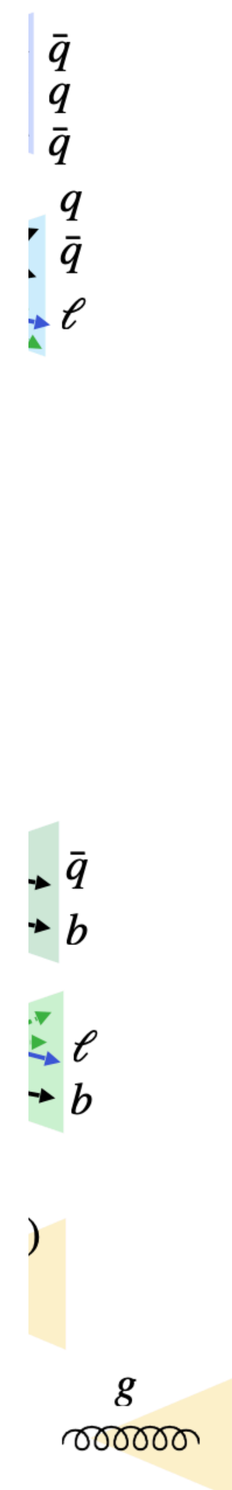


Even more grand : GloParT

CMS PAS HIG-23-012



Variable	Definition
charged PF candidates	
$\log p_{\mathrm{T}}$	logarithm of the particle p_{T}
$\log E$	logarithm of the particle energy
$\Delta\eta(\mathrm{jet})$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi(\mathrm{jet})$	difference in azimuthal angle between the particle and the jet axis
$\|\eta\|$	absolute value of the particle pseudorapidity
q	electric charge of the particle
`isMuon`	true if the particle is identified as a muon
`isElectron`	true if the particle is identified as an electron
`isChargedHadron`	true if the particle is identified as a charged hadron
`pvAssociationQuality`	flag related to the association of the track to the primary vertices
`lostInnerHits`	quality flag of the track related to missing hits on the pixel layers
χ^2/dof	χ^2 value of the trajectory fit normalized to the number of degrees of freedom
`qualityMask`	quality flag of the track
d_z	longitudinal impact parameter of the track
d_z/σ_{d_z}	significance of the longitudinal impact parameter
d_{xy}	transverse impact parameter of the track
$d_{xy}/\sigma_{d_{xy}}$	significance of the transverse impact parameter
η_{rel}	pseudorapidity of the track relative to the jet axis
$p_{\mathrm{T,rel}}$ ratio	track momentum perpendicular to the jet axis, divided by the magnitude of the track momentum
$p_{\mathrm{par,rel}}$ ratio	track momentum parallel to the jet axis divided by the magnitude of the track momentum
$d_{3\mathrm{D}}$	signed 3D impact parameter of the track
$d_{3\mathrm{D}}/\sigma_{3\mathrm{D}}$	signed 3D impact parameter significance of the track
`trackDistance`	distance between the track and the jet axis at their point of closest approach
Neutral PF candidates	
$\log p_{\mathrm{T}}$	logarithm of the particles p_{T}
$\log E$	logarithm of the particles energy
$\Delta\eta(\mathrm{jet})$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi(\mathrm{jet})$	difference in azimuthal angle between the particle and the jet axis
$\|\eta\|$	absolute value of the particle pseudorapidity
`isPhoton`	true if the particle is identified as a photon
`isNeutralHadron`	true if the particle is identified as a neutral hadron
For SVs within the jet cone	
$\log p_{\mathrm{T}}$	logarithm of the SV p_{T}
m_{SV}	invariant mass of the tracks associated with the SV
$\Delta\eta(\mathrm{jet})$	difference in pseudorapidity between the SV and the jet axis
$\Delta\phi(\mathrm{jet})$	difference in azimuthal angle between the SV and the jet axis
$\|\eta\|$	absolute value of the SV pseudorapidity
N_{tracks}	number of tracks associated with the SV
χ^2/dof	χ^2 value of the SV fit normalized to the number of degrees of freedom
$d_{2\mathrm{D}}$	signed 2D impact parameter (i.e., in the transverse plane) of the SV
$d_{2\mathrm{D}}/\sigma_{2\mathrm{D}}$	signed 2D impact parameter significance of the SV
$d_{3\mathrm{D}}$	signed 3D impact parameter of the SV
$d_{3\mathrm{D}}/\sigma_{3\mathrm{D}}$	signed 3D impact parameter significance of the SV



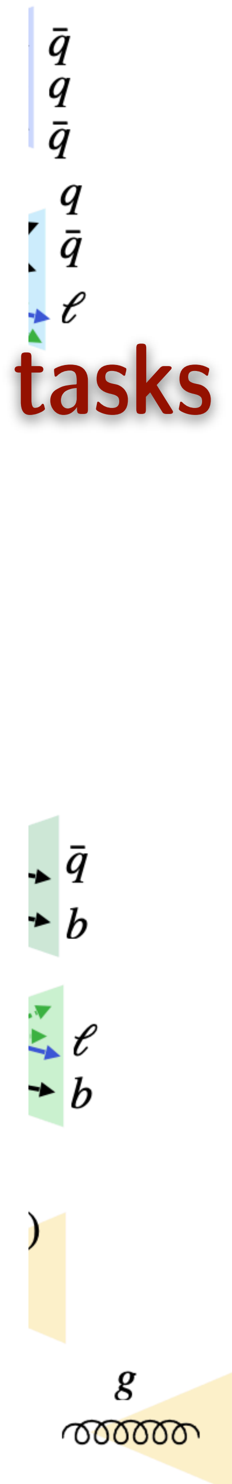
Even more grand : GloParT

CMS PAS HIG-23-012

Proces
H→VV (full-hadr
H→WW (semi-lept
H→qq
H→ττ
t→bW (hadron
t→bW (leptoni
QCD

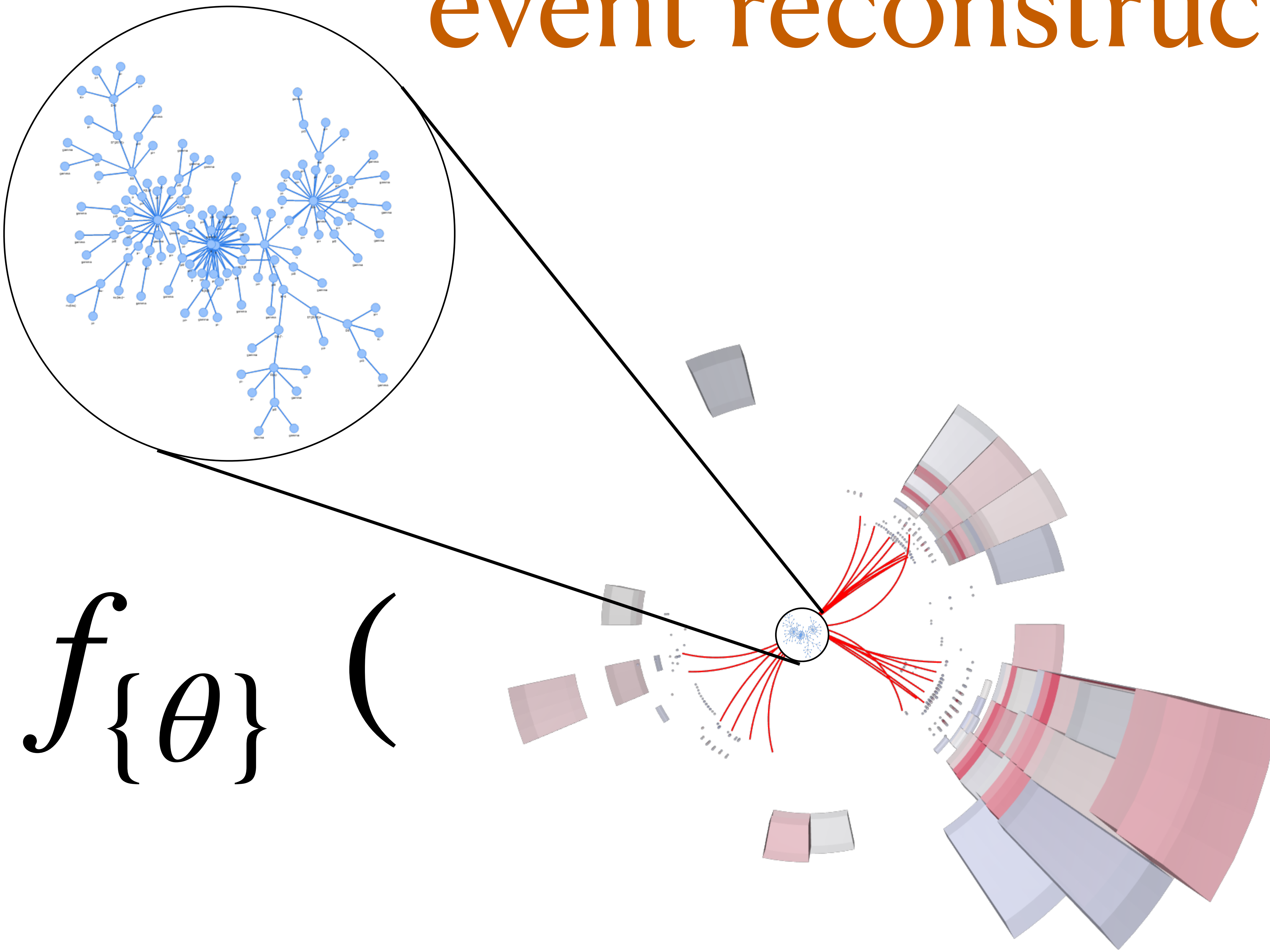
Variable	Definition
charged PF candidates	
$\log p_{\mathrm{T}}$	logarithm of the particle p_{T}
$\log E$	logarithm of the particle energy
$\Delta\eta(\text{jet})$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi(\text{jet})$	difference in azimuthal angle between the particle and the jet axis
$\|\eta\|$	absolute value of the particle pseudorapidity
q	electric charge of the particle
`isMuon`	true if the particle is identified as a muon
`isElectron`	true if the particle is identified as an electron
`isChargedHadron`	true if the particle is identified as a charged hadron
`pvAssociationQuality`	flag related to the association of the track to the primary vertices
`lostInnerHits`	quality flag of the track related to missing hits on the pixel layers
χ^2/dof	χ^2 value of the trajectory fit normalized to the number of degrees of freedom
`qualityMask`	quality flag of the track
d_z	longitudinal impact parameter of the track
d_z/σ_{d_z}	significance of the longitudinal impact parameter
d_{xy}	transverse impact parameter of the track
$d_{xy}/\sigma_{d_{xy}}$	significance of the transverse impact parameter
η_{rel}	pseudorapidity of the track relative to the jet axis
$p_{\mathrm{T,rel}}$ ratio	track momentum perpendicular to the jet axis, divided by the magnitude of the track momentum
$p_{\text{par,rel}}$ ratio	track momentum parallel to the jet axis divided by the magnitude of the track momentum
$d_{3\mathrm{D}}$	signed 3D impact parameter of the track
$d_{3\mathrm{D}}/\sigma_{3\mathrm{D}}$	signed 3D impact parameter significance of the track
`trackDistance`	distance between the track and the jet axis at their point of closest approach
Neutral PF candidates	
$\log p_{\mathrm{T}}$	logarithm of the particles p_{T}
$\log E$	logarithm of the particles energy
$\Delta\eta(\text{jet})$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi(\text{jet})$	difference in azimuthal angle between the particle and the jet axis
$\|\eta\|$	absolute value of the particle pseudorapidity
`isPhoton`	true if the particle is identified as a photon
`isNeutralHadron`	true if the particle is identified as a neutral hadron
For SVs within the jet cone	
$\log p_{\mathrm{T}}$	logarithm of the SV p_{T}
m_{SV}	invariant mass of the tracks associated with the SV
$\Delta\eta(\text{jet})$	difference in pseudorapidity between the SV and the jet axis
$\Delta\phi(\text{jet})$	difference in azimuthal angle between the SV and the jet axis
$\|\eta\|$	absolute value of the SV pseudorapidity
N_{tracks}	number of tracks associated with the SV
χ^2/dof	χ^2 value of the SV fit normalized to the number of degrees of freedom
$d_{2\mathrm{D}}$	signed 2D impact parameter (i.e., in the transverse plane) of the SV
$d_{2\mathrm{D}}/\sigma_{2\mathrm{D}}$	signed 2D impact parameter significance of the SV
$d_{3\mathrm{D}}$	signed 3D impact parameter of the SV
$d_{3\mathrm{D}}/\sigma_{3\mathrm{D}}$	signed 3D impact parameter significance of the SV

One can do both classification & regression tasks



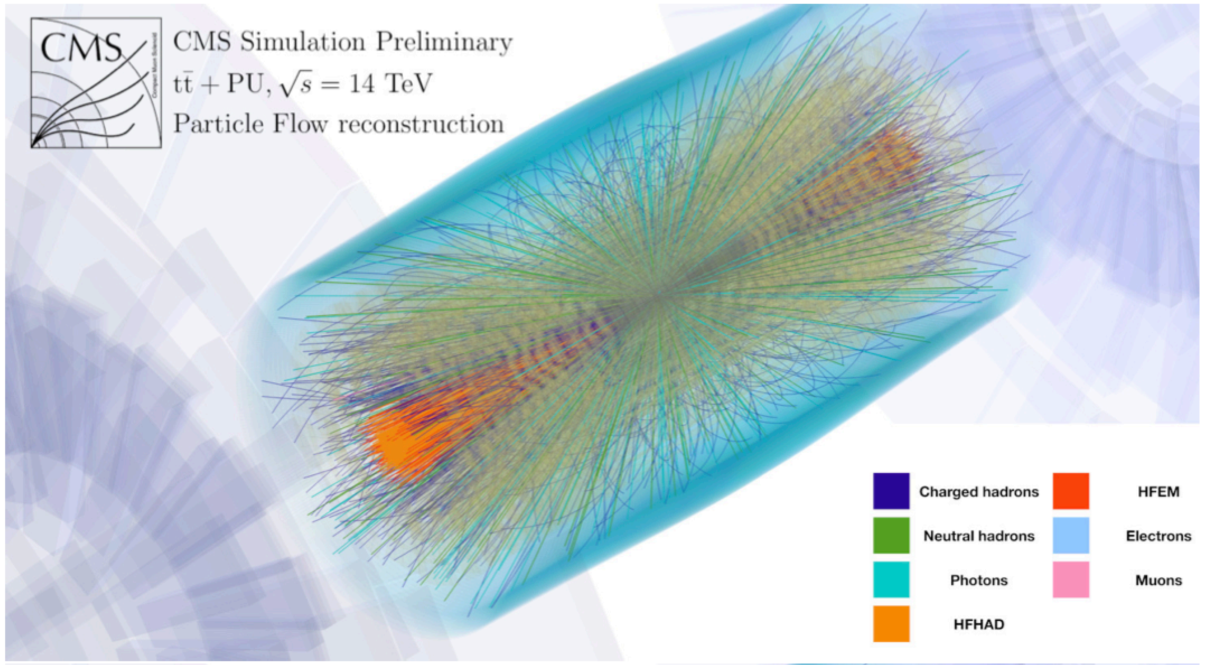
Particle-Flow event reconstruction

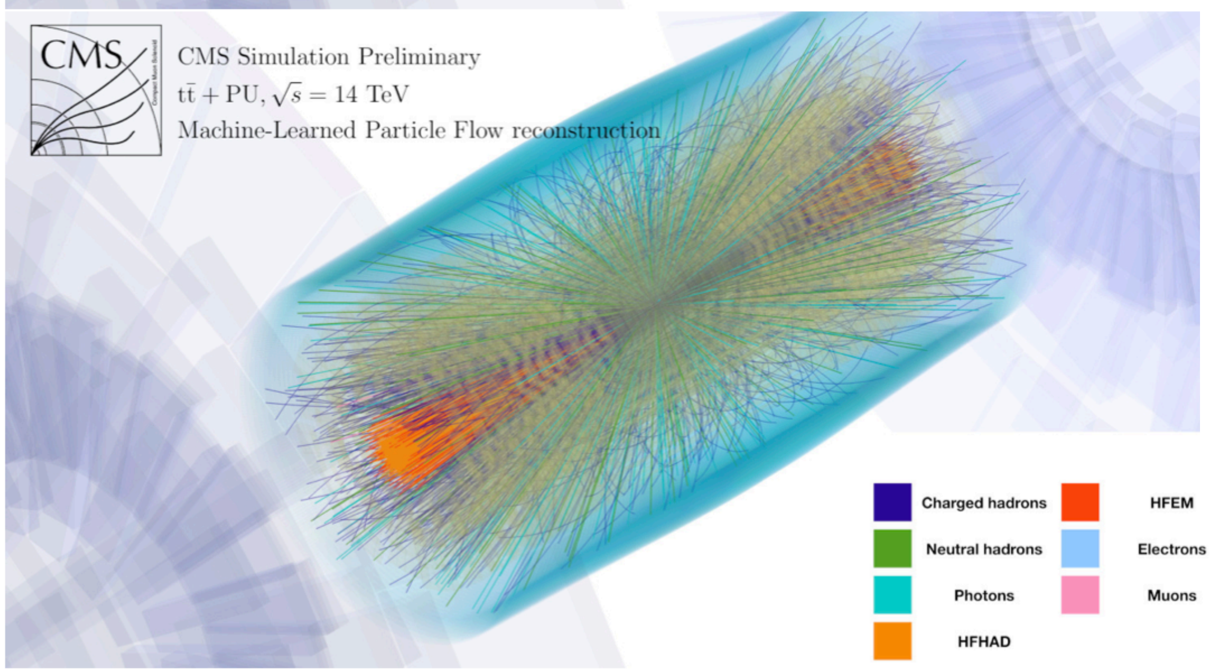
$$f_{\{\theta\}}\Big($$

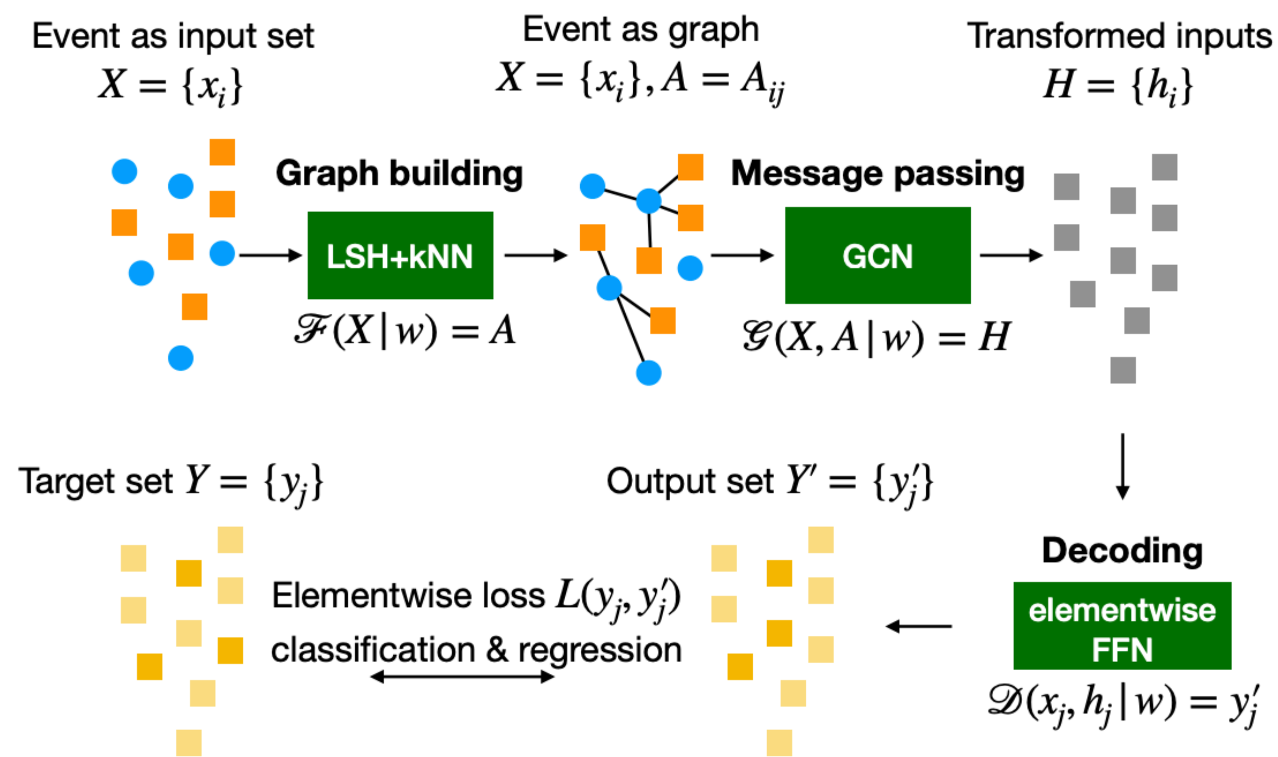


$$\Big)$$

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: $\mathscr{F}, \mathscr{G}, \mathscr{D}$

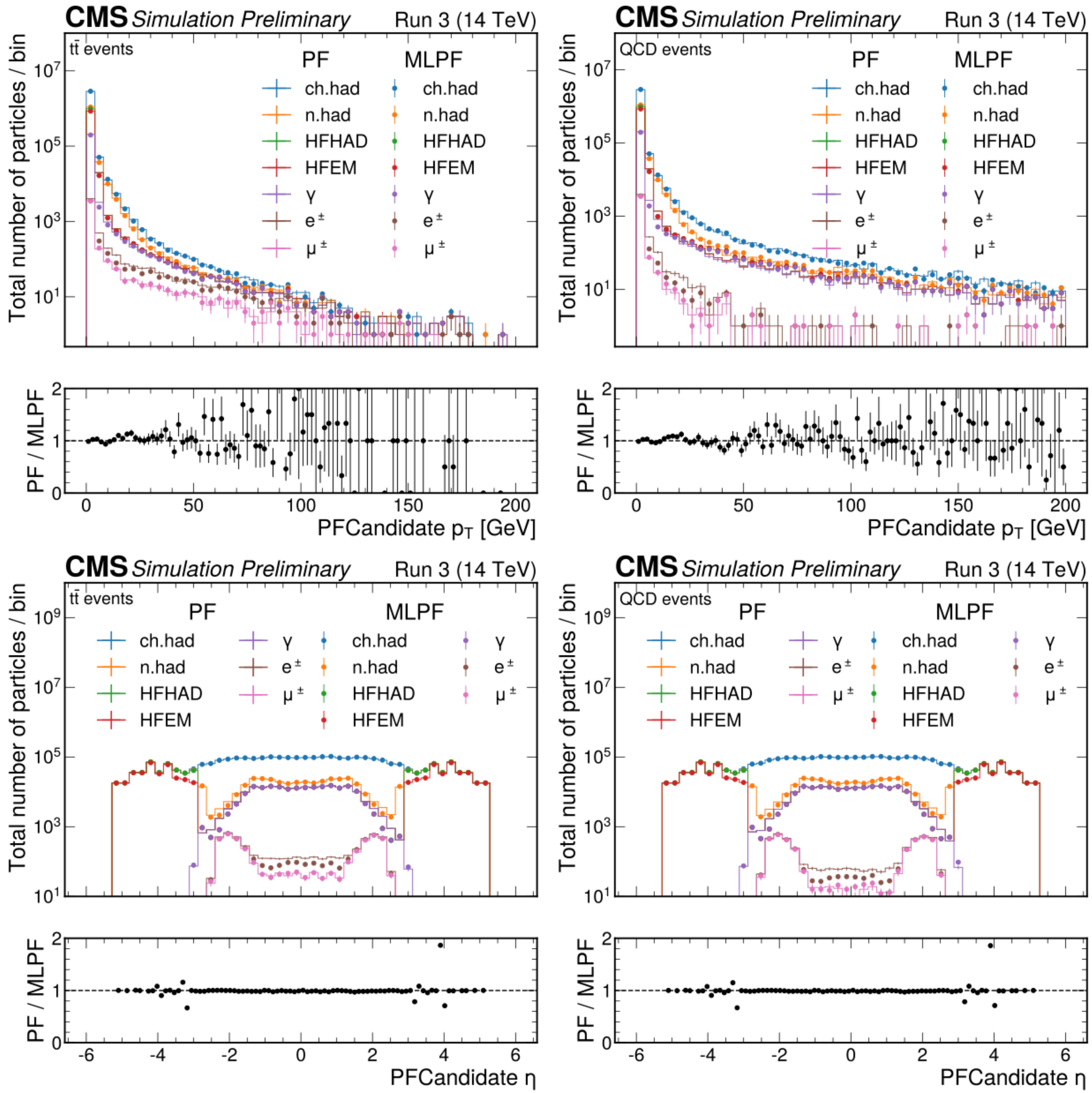
● - track, ■ - calorimeter cluster, ■ - encoded element

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

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

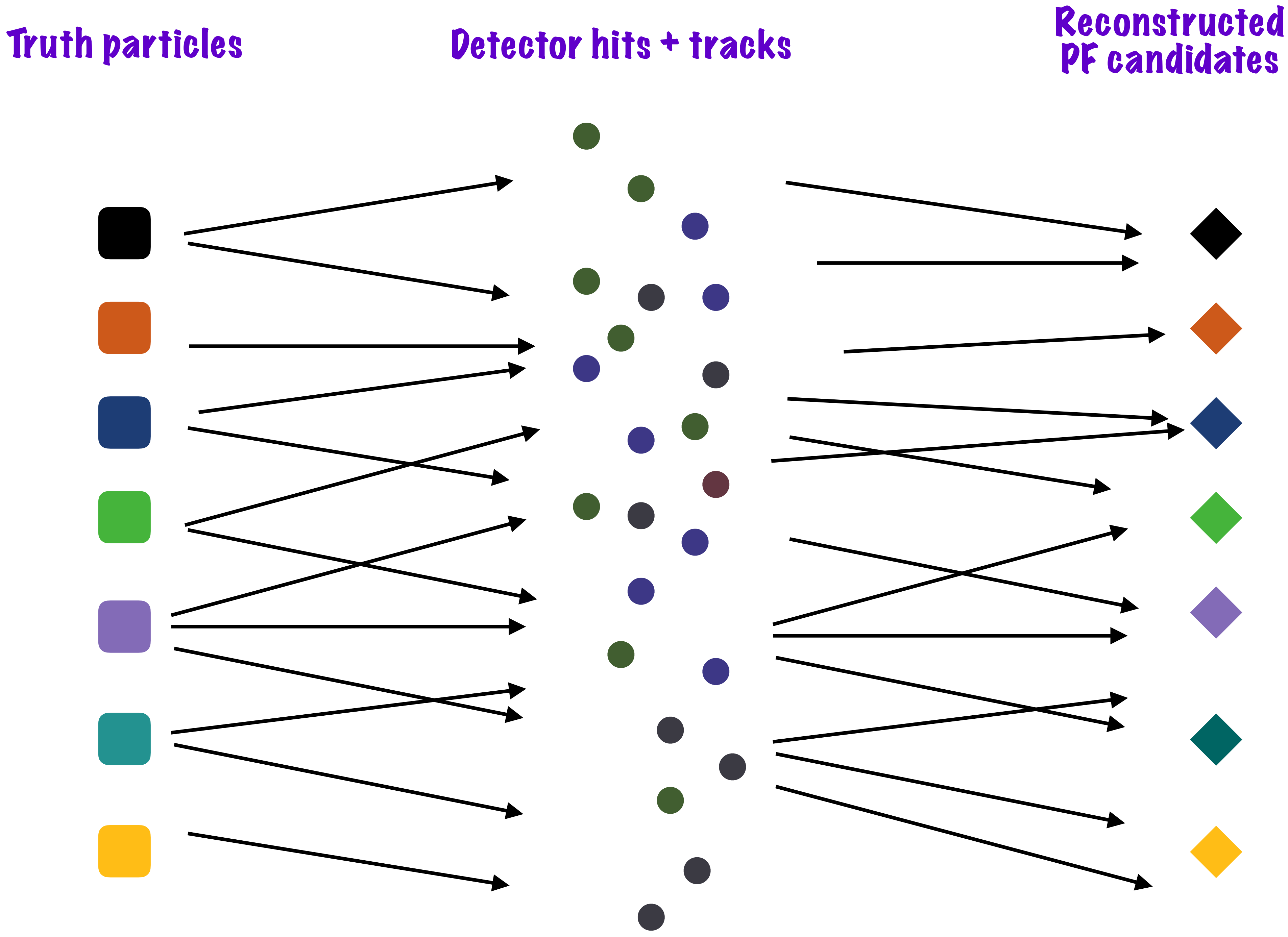
PF lepton, hadron, photon = F_{PF} (track hits + calo cells)

Combining track + calo for PFlow

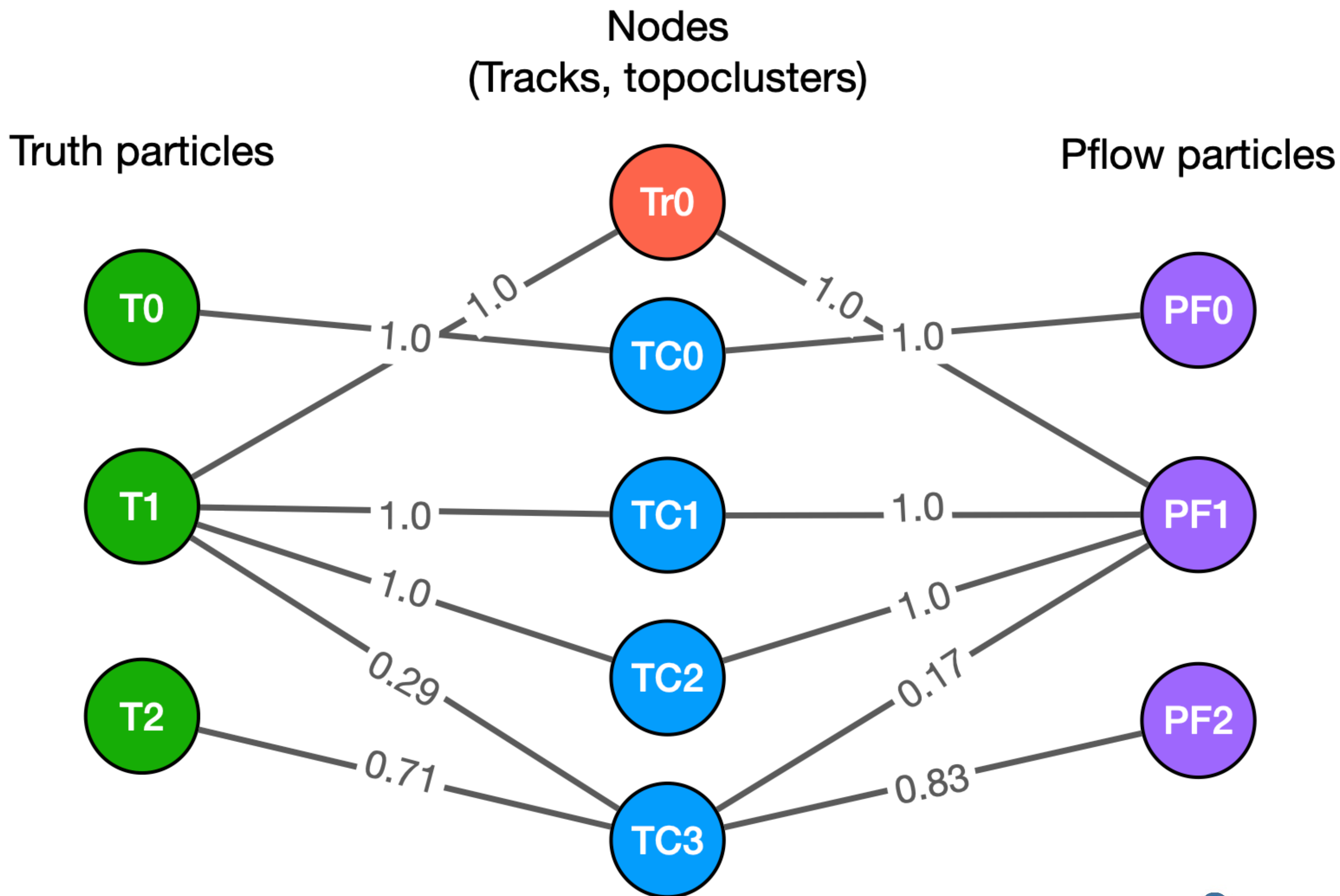


MLPF
arXiv : 2203.00330
J. Pata et. al.

What's the core data structure?

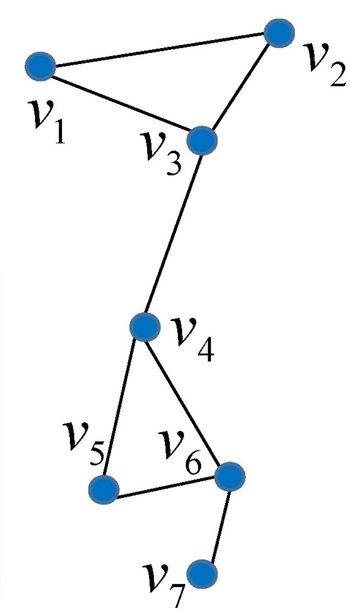


What's the core data structure?

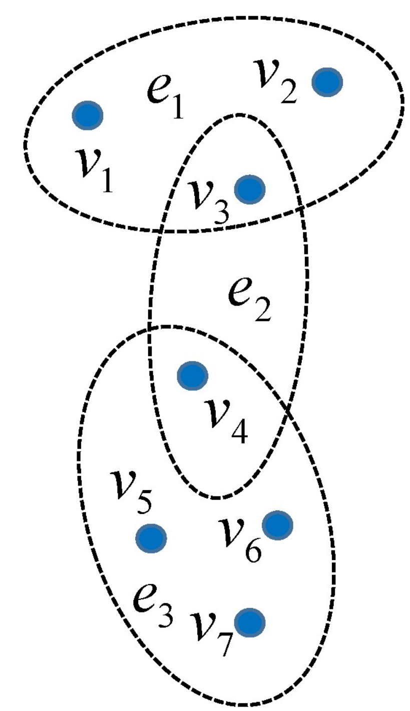


https://www.mdpi.com/2072-4292/9/5/506

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



(a) Simple graph



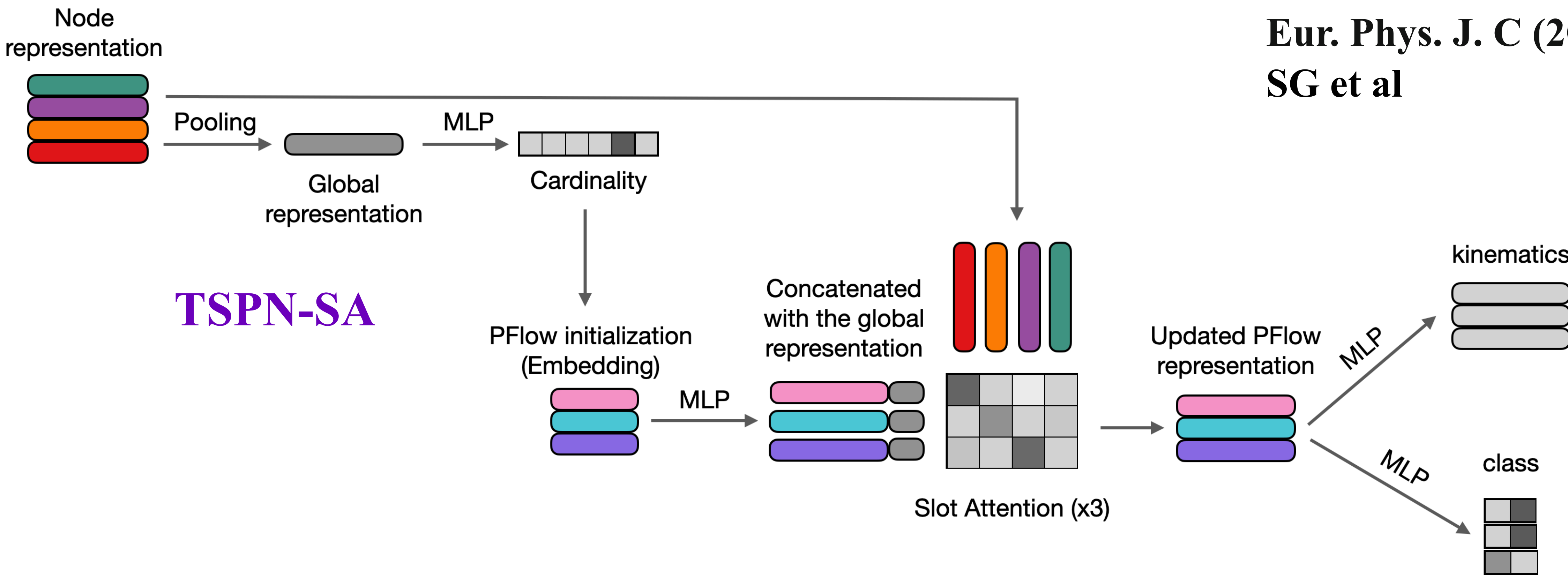
(b) Hypergraph **G**

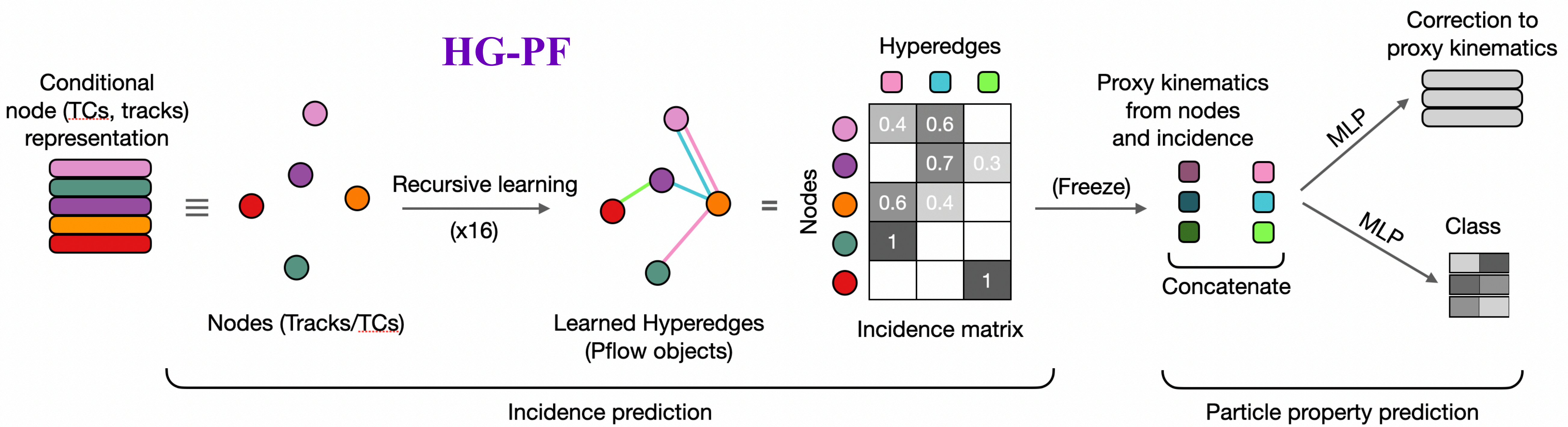
	e_1	e_2	e_3
v_1	1	0	0
v_2	1	0	0
v_3	1	1	0
v_4	0	1	1
v_5	0	0	1
v_6	0	0	1
v_7	0	0	1

(c) Incidence matrix **H**

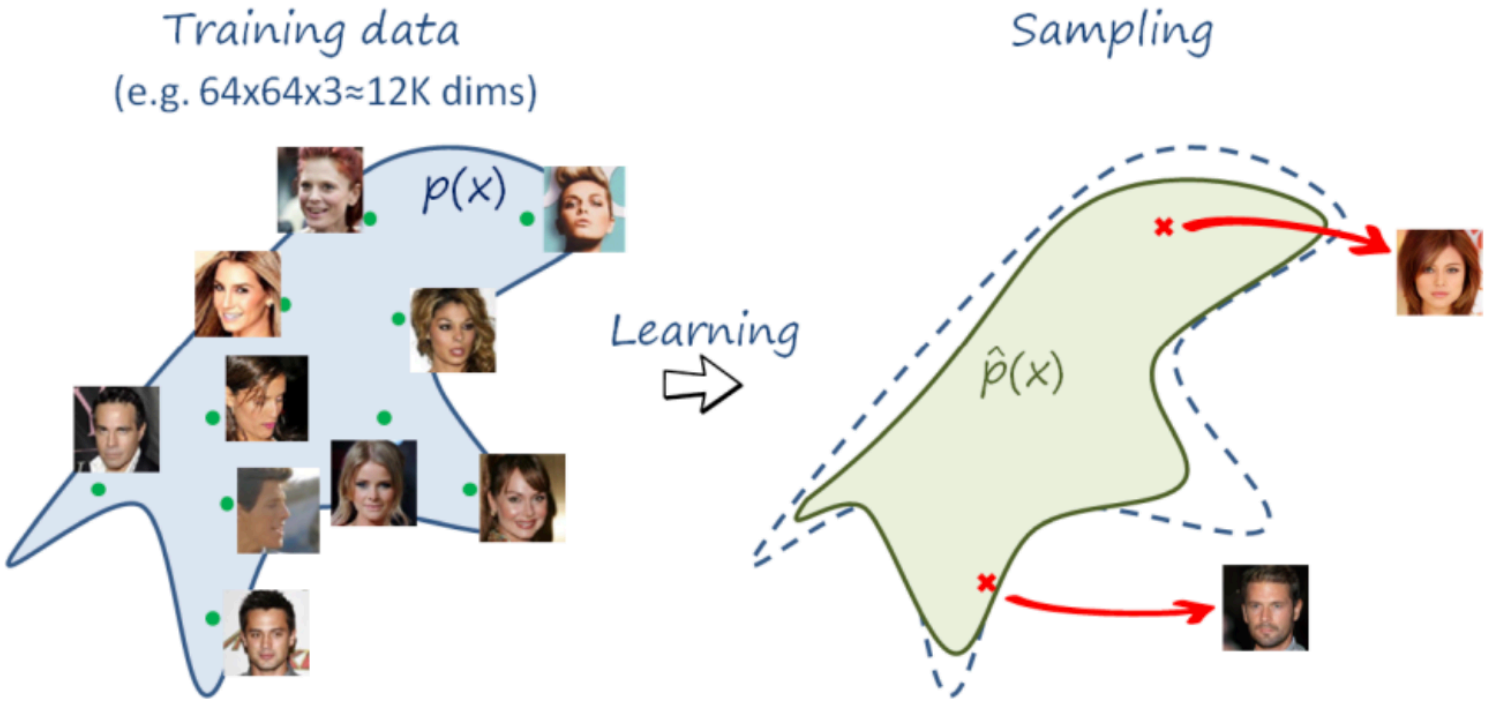
The hypergraph network

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

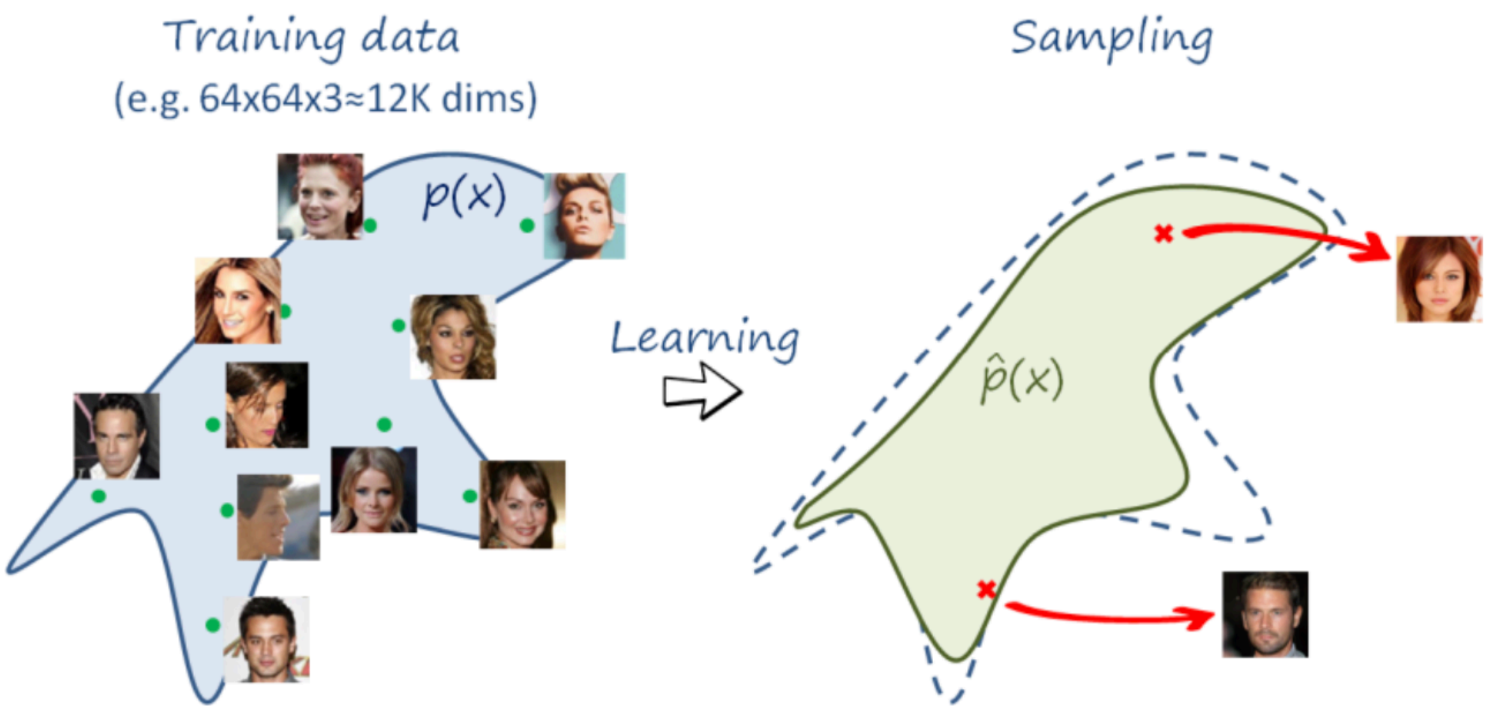




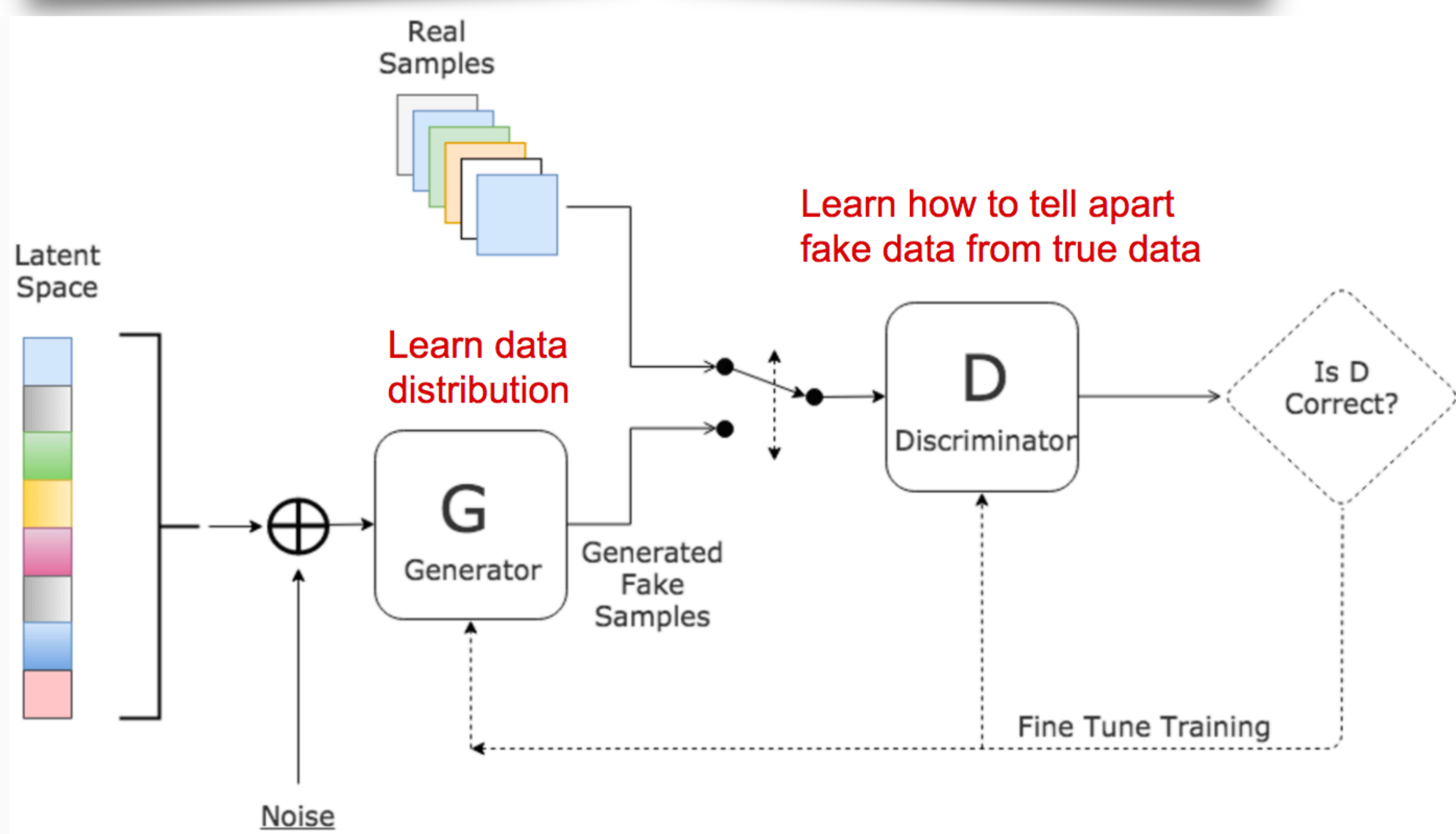
Generative models : what are they?



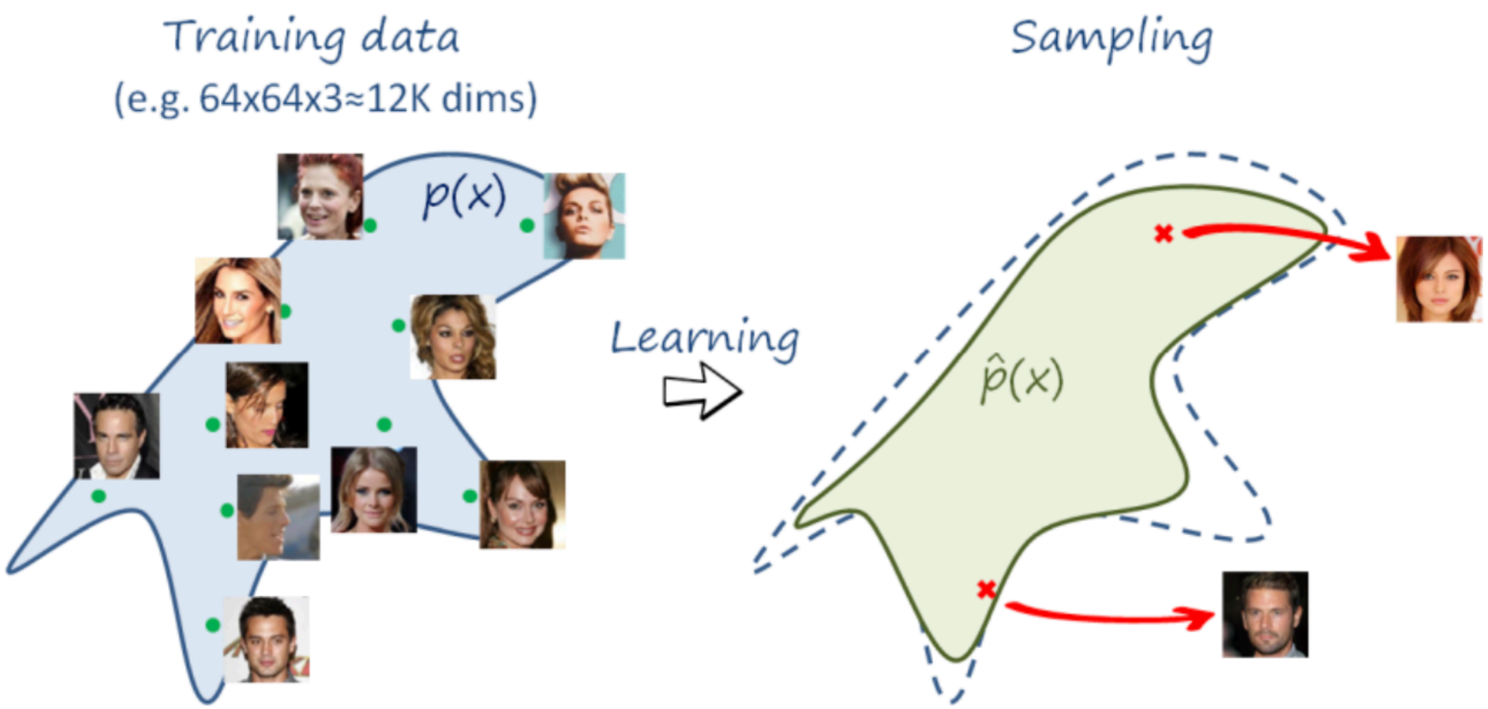
Generative models : what are they?

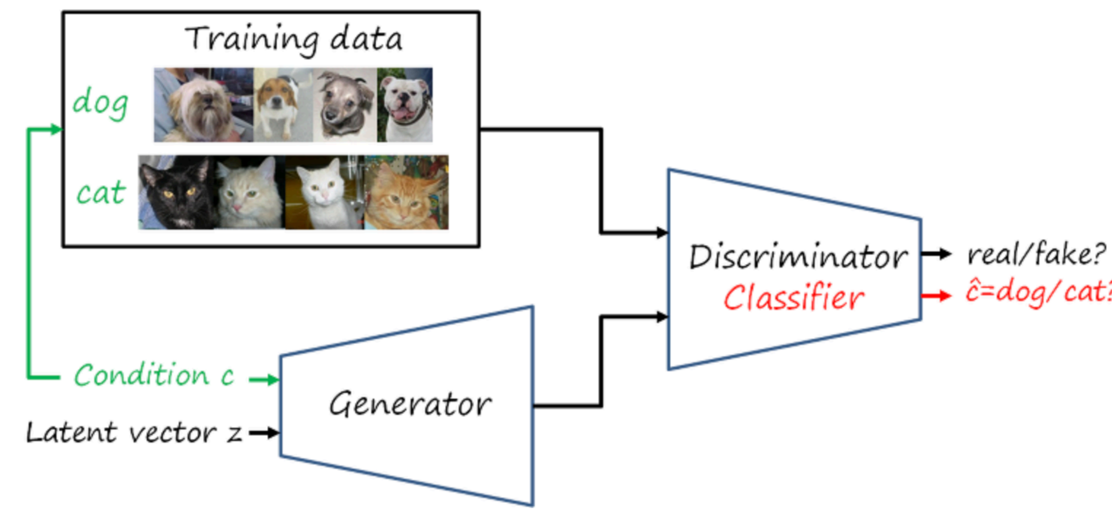


$$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p_{data}(x)}\log D_{\theta_d}(x)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)\right]$$

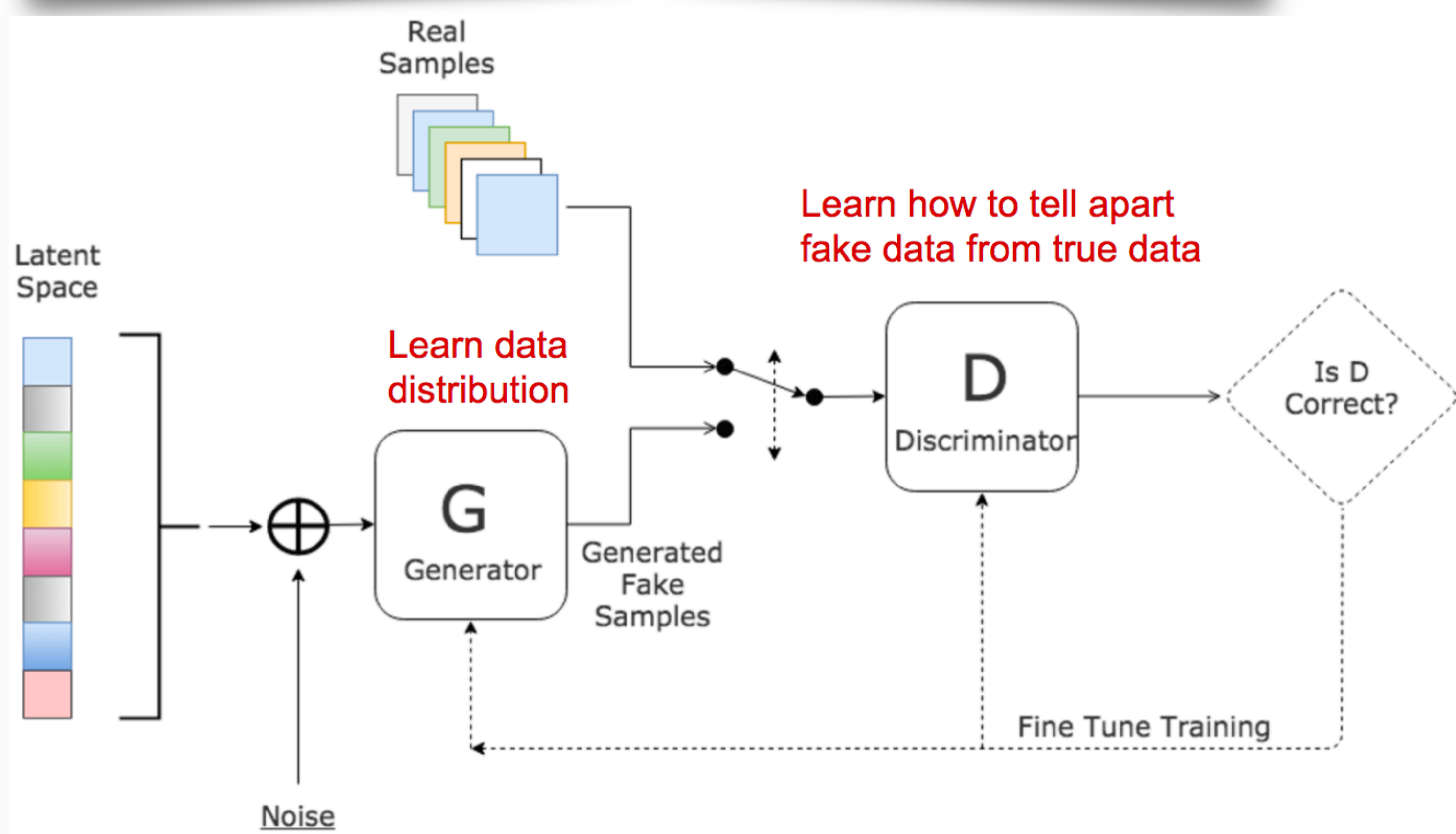


Generative models : what are they?

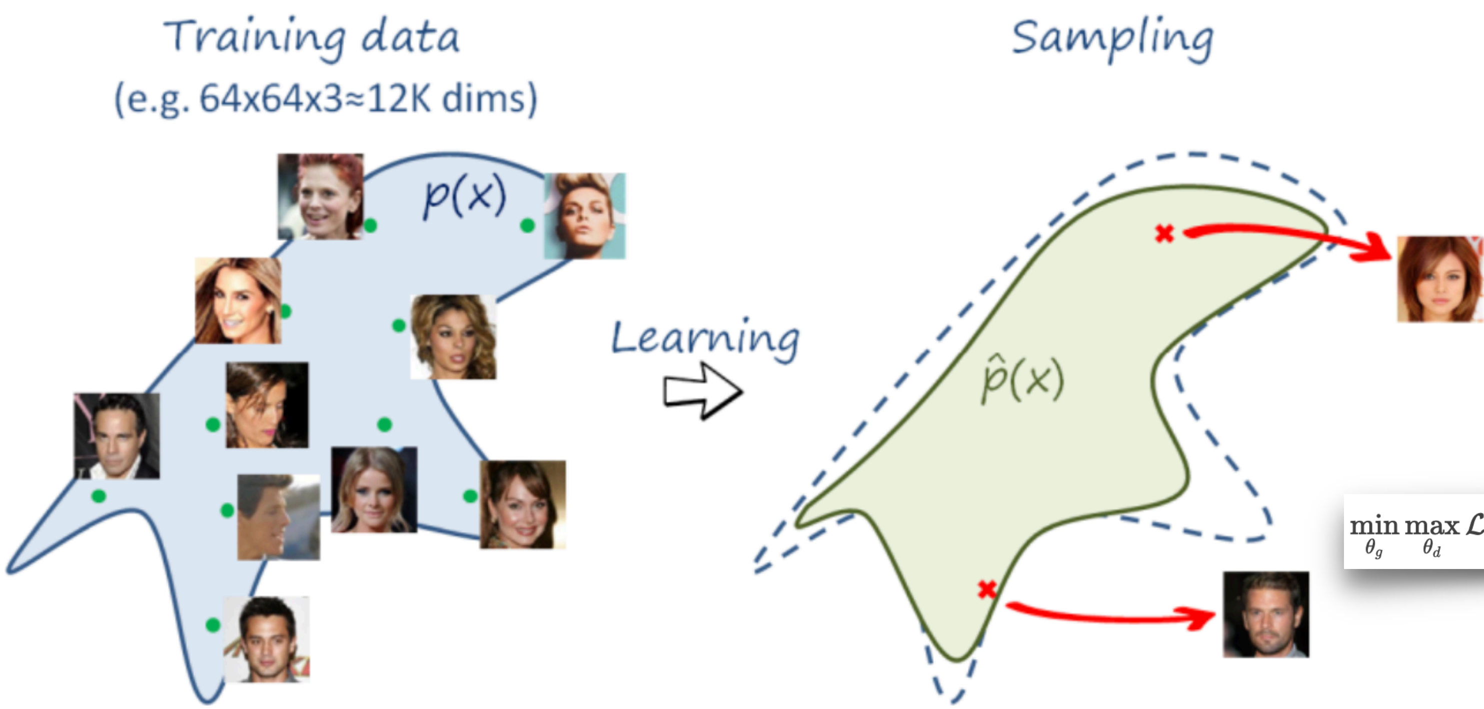


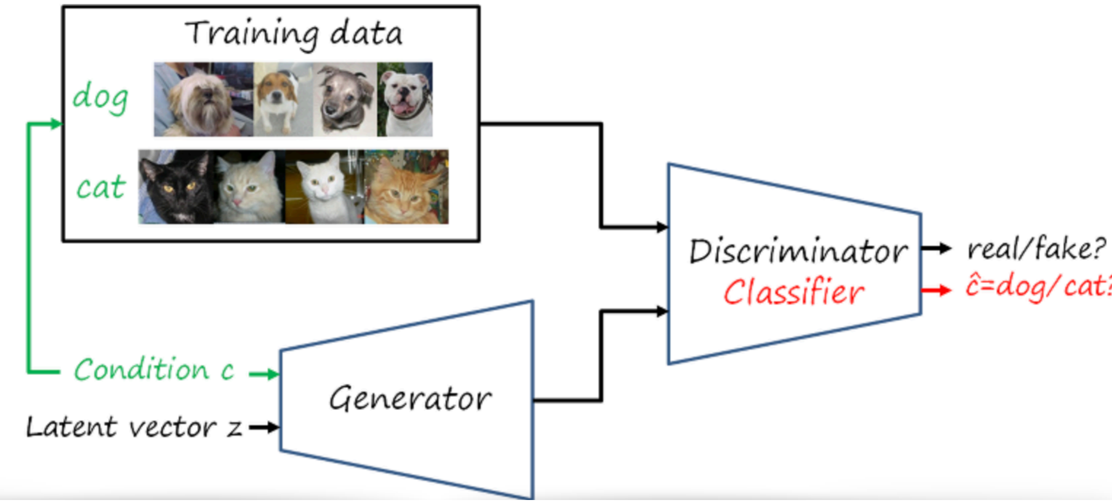


$$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p_{data}(x)}\log D_{\theta_d}(x)+\mathbb{E}_{z\sim p(z)}\log\left(1-D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)\right]$$



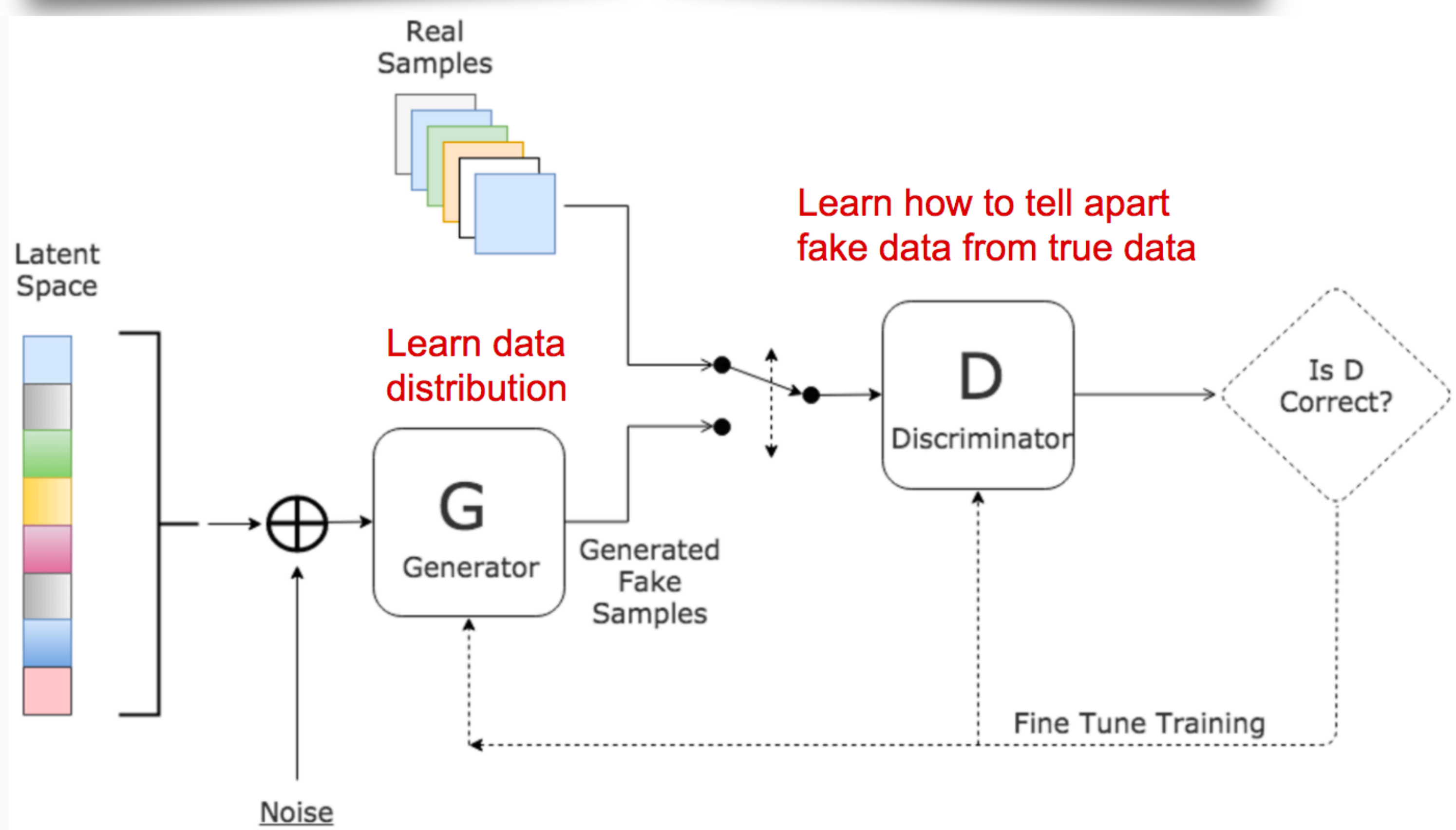
Generative models : what are they?



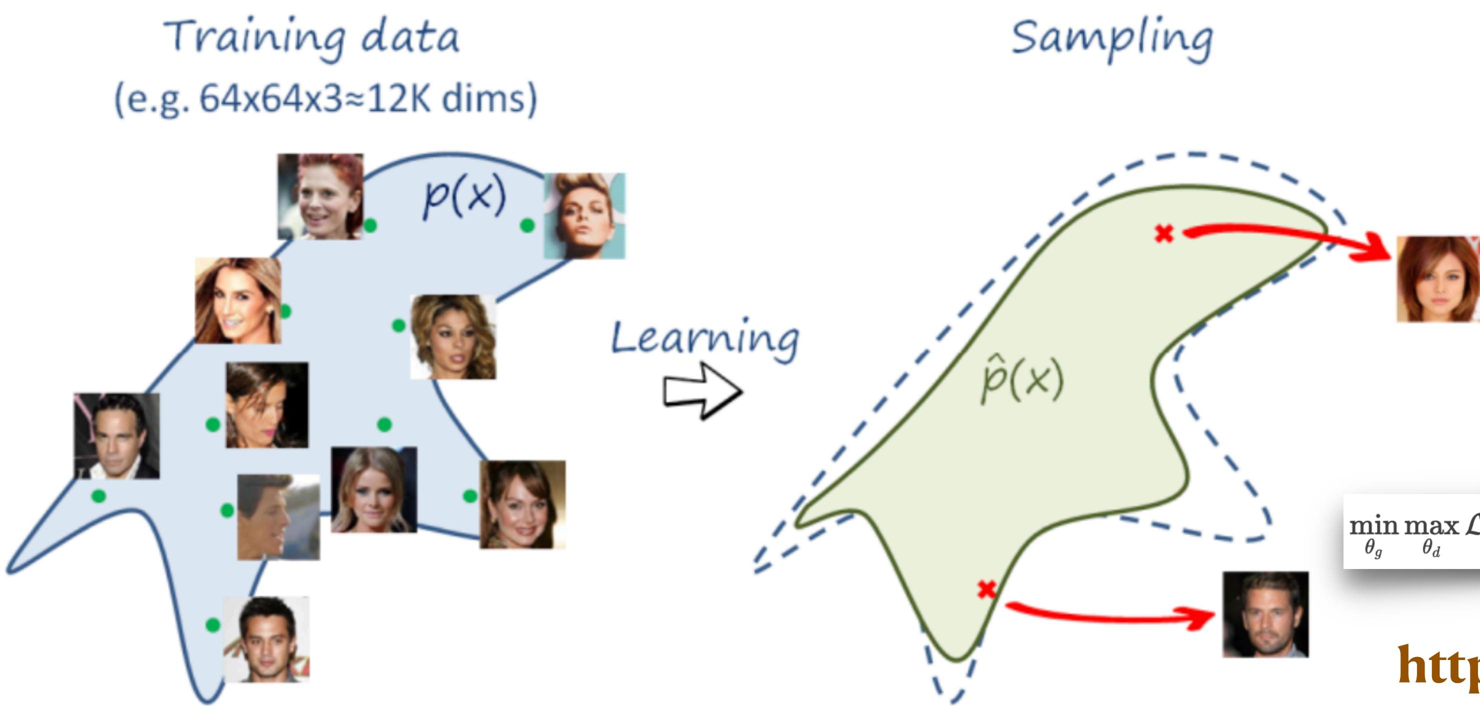


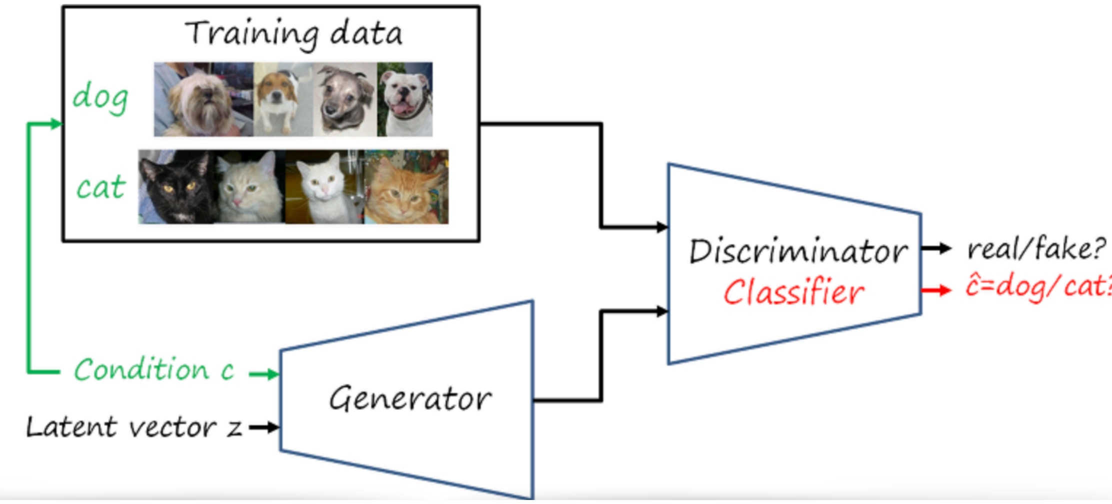
$$\min_{\theta_g}\max_{\theta_d} \mathcal{L}_{\text{cGAN}}(\theta_g,\theta_d) = \min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{c,x\sim p_{data}(c,x)}\log D_{\theta_d}(c,x) + \mathbb{E}_{c\sim p_{data}(c),z\sim p_z(z)}\log\left(1-D_{\theta_d}\left(c, G_{\theta_g}(c,z)\right)\right)\right]$$

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



Generative models : what are they?

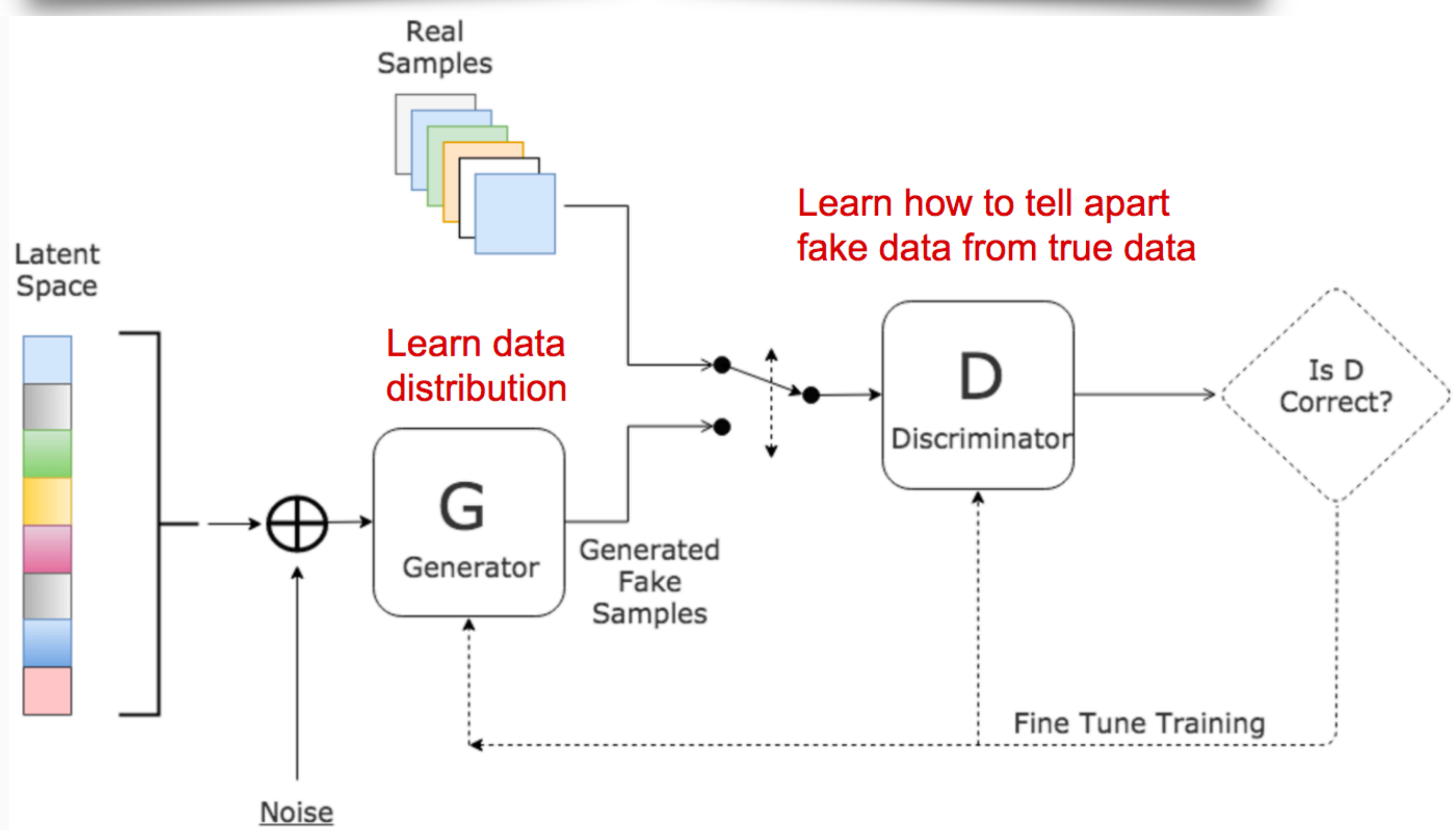




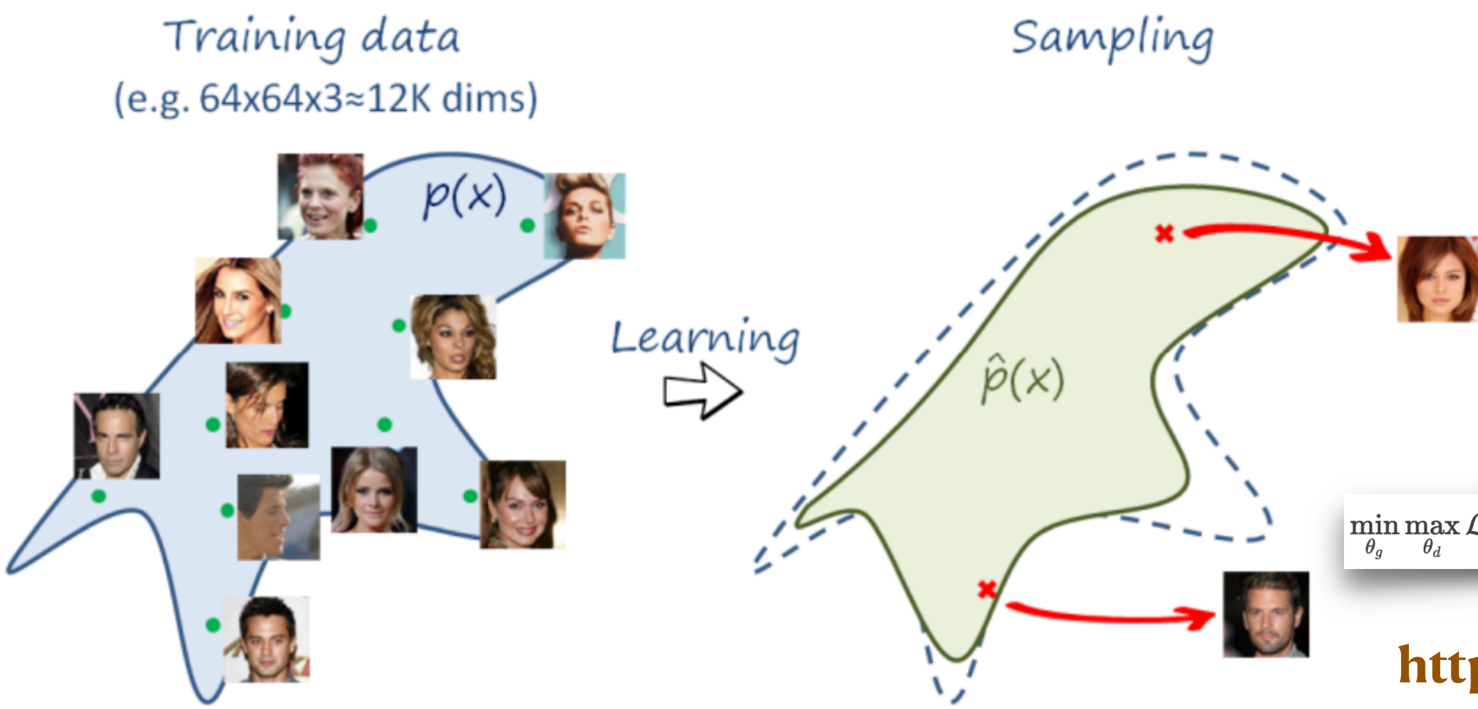
$$\min_{\theta_g}\max_{\theta_d} \mathcal{L}_{\mathrm{cGAN}}(\theta_g,\theta_d) = \min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{c,x\sim p_{\mathrm{data}}(c,x)} \log D_{\theta_d}(c,x) + \mathbb{E}_{c\sim p_{\mathrm{data}}(c),z\sim p_z(z)} \log\left(1-D_{\theta_d}\left(c,G_{\theta_g}(c,z)\right)\right)\right]$$

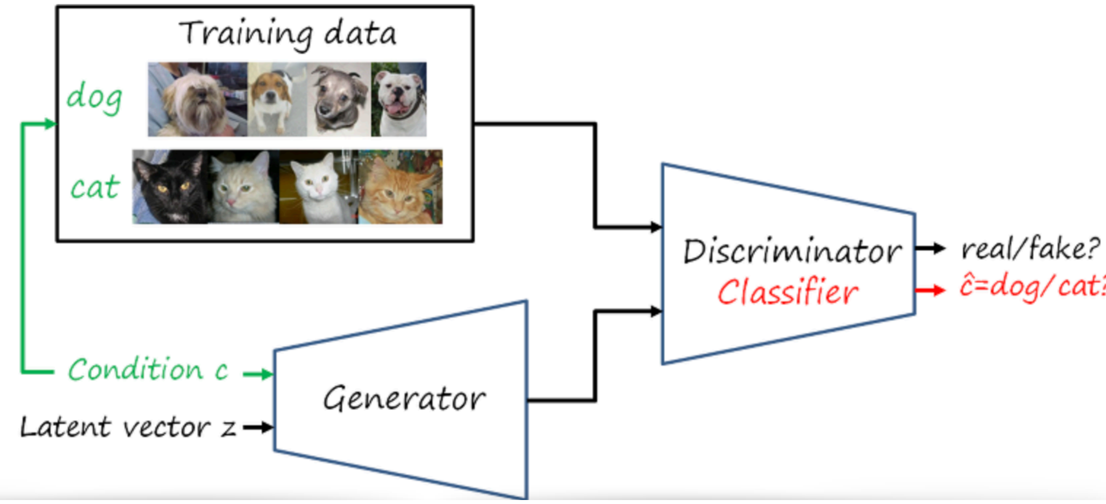
http://www.lherranz.org/2018/08/07/imagetranslation/

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



Generative models : what are they?

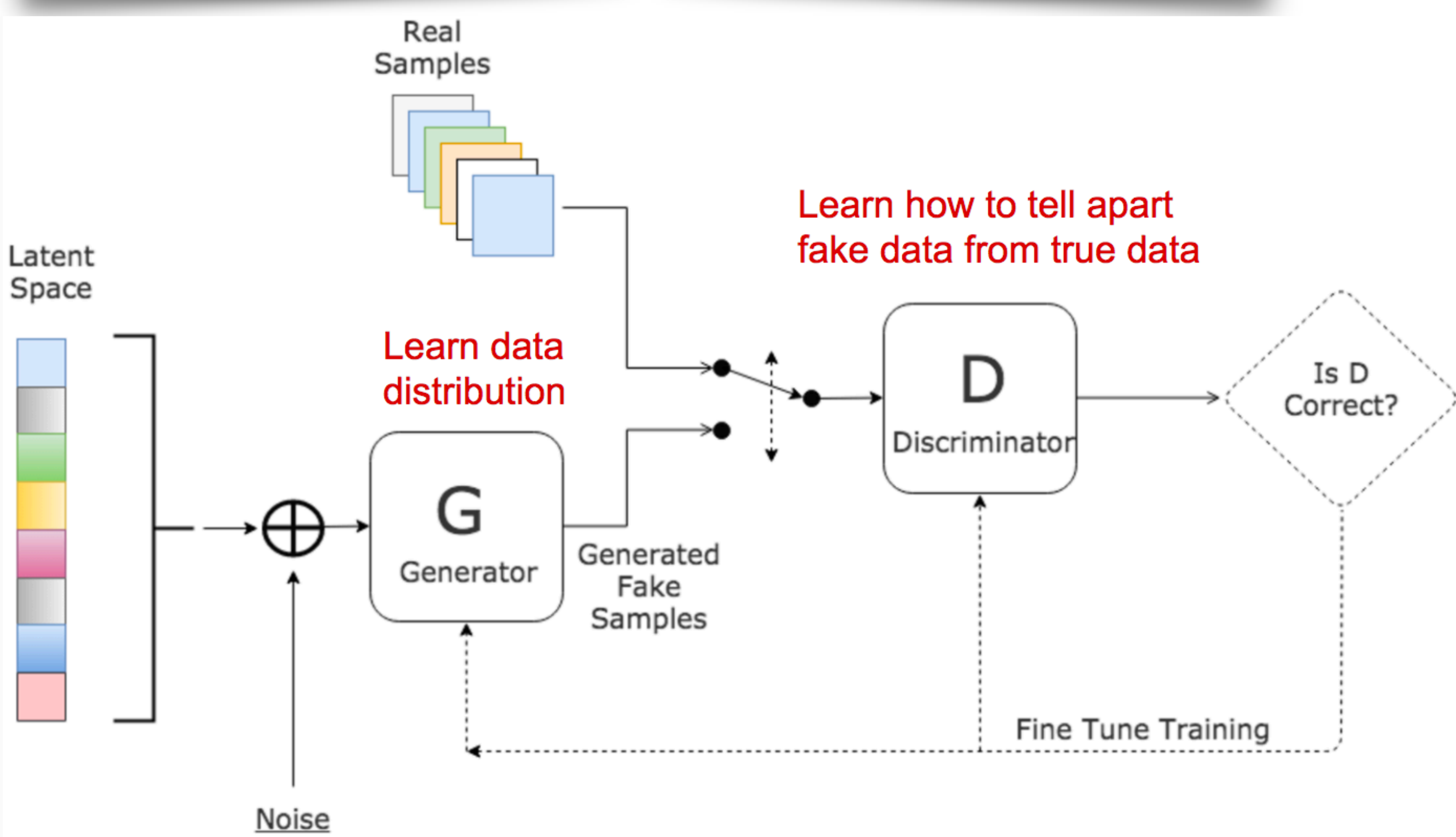


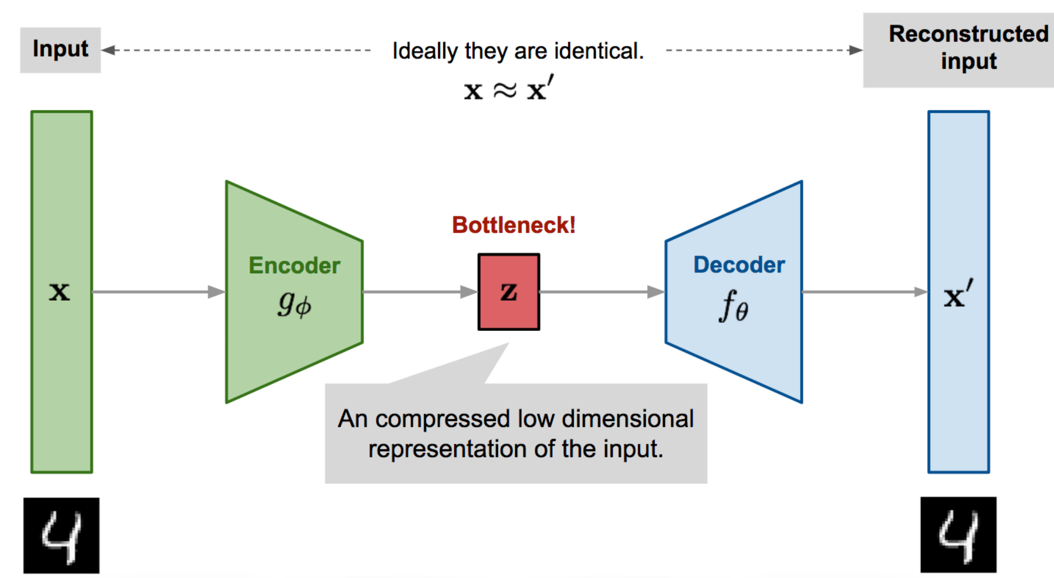


$$\min_{\theta_g}\max_{\theta_d} \mathcal{L}_{\mathrm{cGAN}}(\theta_g, \theta_d) = \min_{\theta_g}\max_{\theta_d} \left[\mathbb{E}_{c,x\sim p_{\mathrm{data}}(c,x)} \log D_{\theta_d}(c, x) + \mathbb{E}_{c\sim p_{\mathrm{data}}(c), z\sim p_z(z)} \log\left(1 - D_{\theta_d}\left(c, G_{\theta_g}(c, z)\right)\right)\right]$$

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

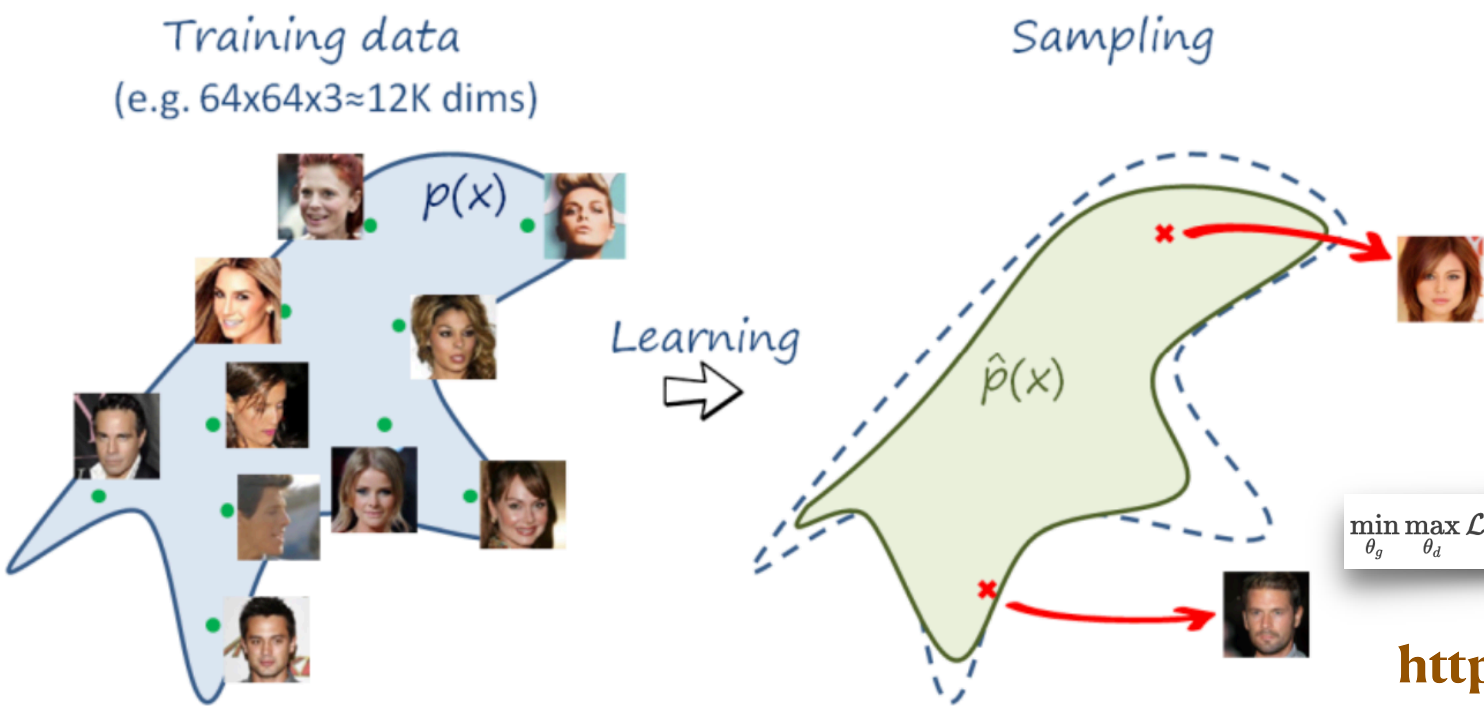
$$\min_{\theta_g}\max_{\theta_d} \left[\mathbb{E}_{x\sim p_{data}(x)} \log D_{\theta_d}(x) + \mathbb{E}_{z\sim p(z)} \log\left(1 - D_{\theta_d}\left(G_{\theta_g}(z)\right)\right)\right]$$

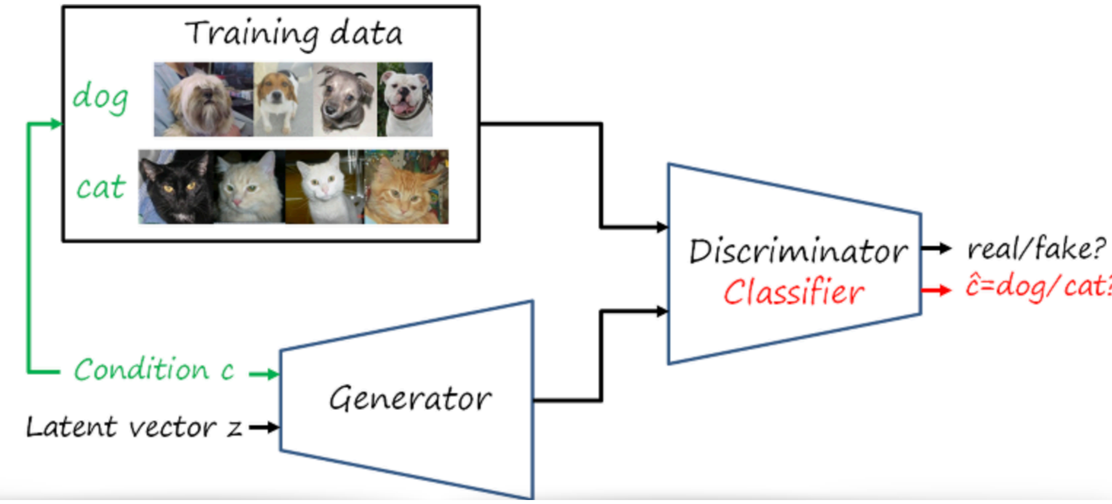




$$L_{\mathrm{AE}}(\theta, \phi) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}^{(i)} - f_\theta(g_\phi(\mathbf{x}^{(i)})))^2$$

Generative models : what are they?

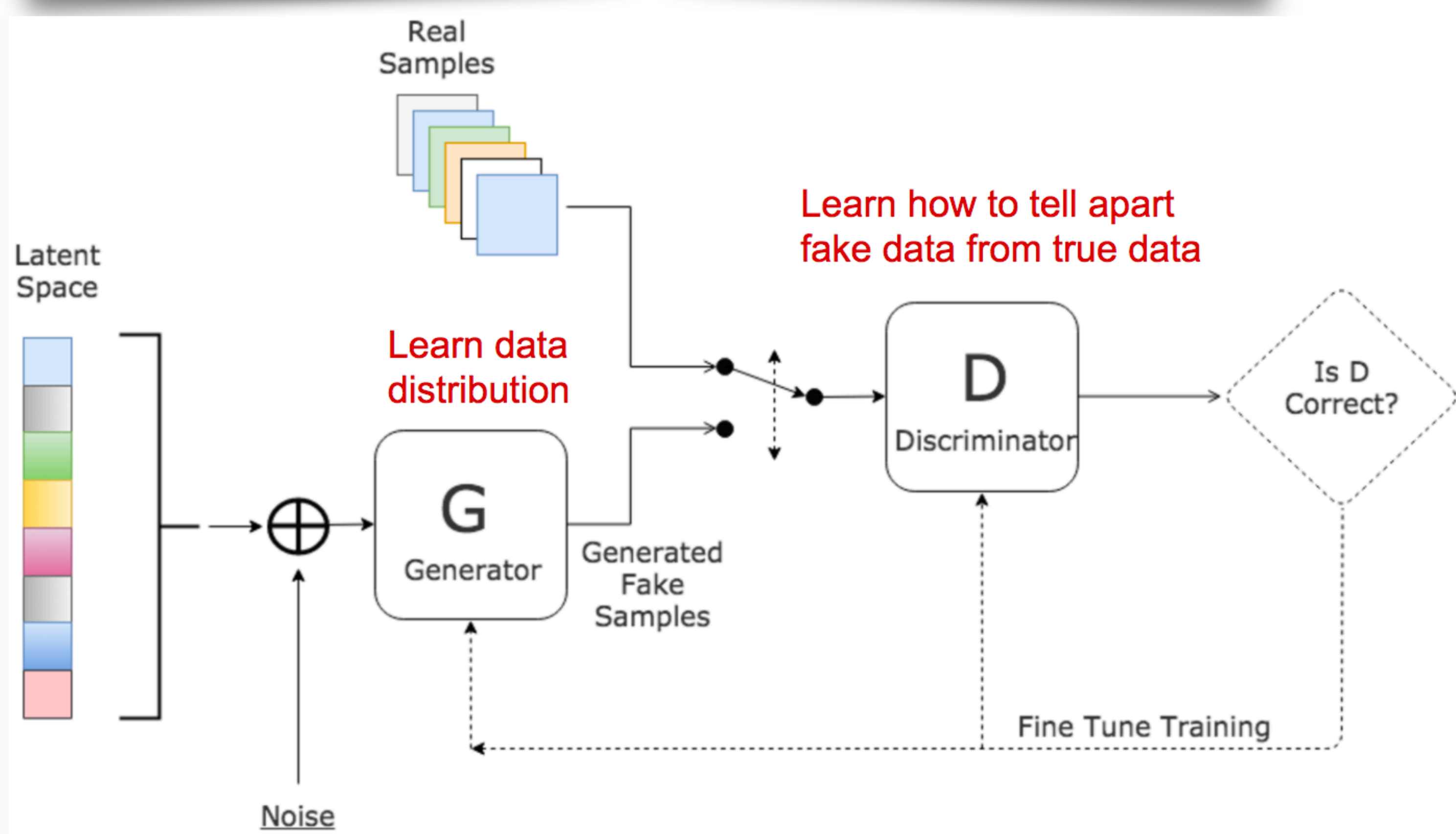


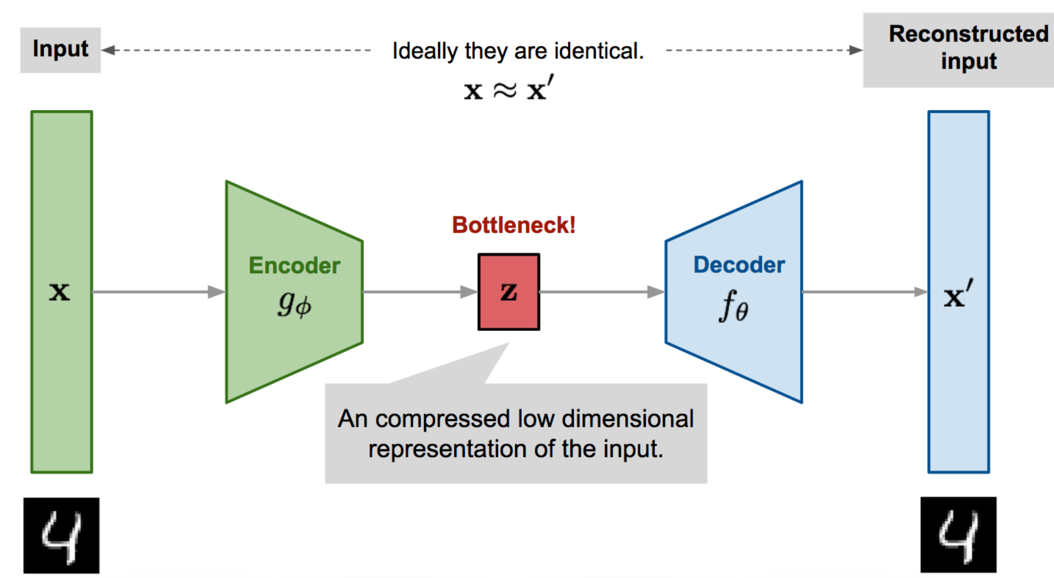


$$\min_{\theta_g}\max_{\theta_d}\mathcal{L}_{\mathrm{cGAN}}\left(\theta_g,\theta_d\right)=\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{c,x\sim p_{\mathrm{data}}(c,x)}\log D_{\theta_d}\left(c,x\right)+\mathbb{E}_{c\sim p_{\mathrm{data}}(c),z\sim p_z(z)}\log\left(1-D_{\theta_d}\left(c,G_{\theta_g}\left(c,z\right)\right)\right)\right]$$

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

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

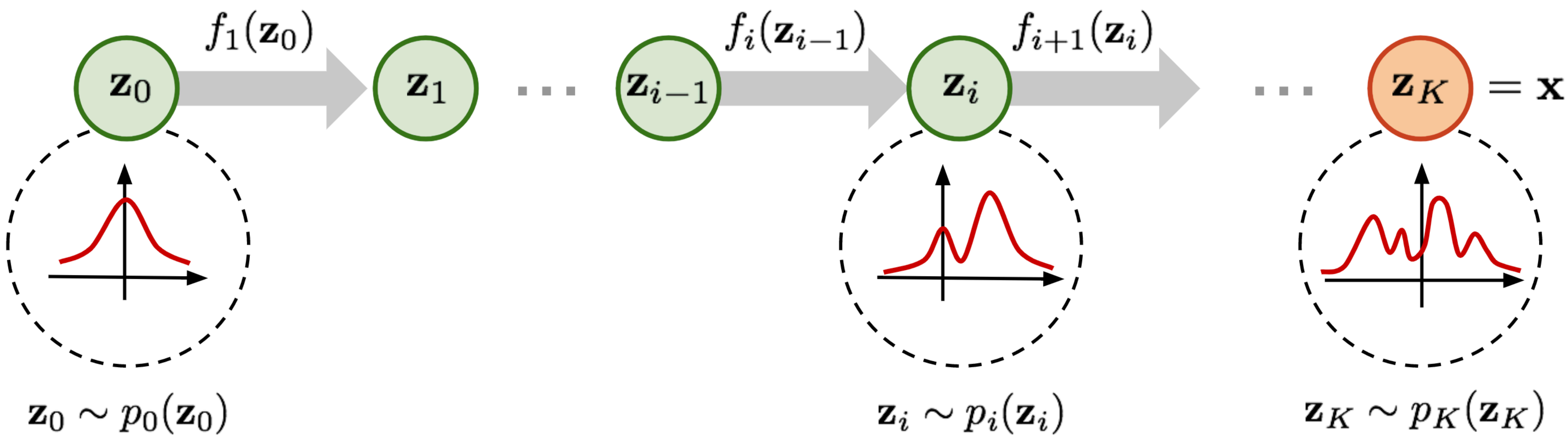




$$L_{\mathrm{AE}}(\theta,\phi)=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}^{(i)}-f_\theta(g_\phi(\mathbf{x}^{(i)})))^2$$

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

Learn the PDF through bijections

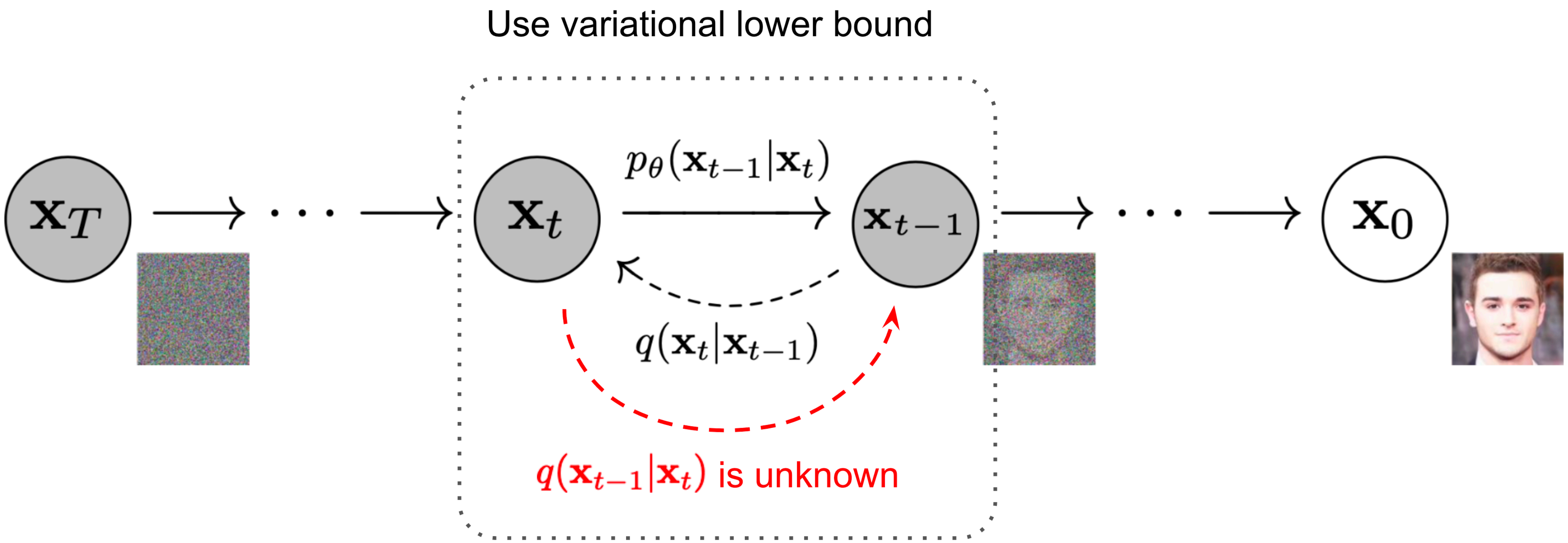


$$\mathbf{z} \sim \pi(\mathbf{z}), \mathbf{x} = f(\mathbf{z}), \mathbf{z} = f^{-1}(\mathbf{x})$$
$$p(\mathbf{x}) = \pi(\mathbf{z}) \left| \det \frac{d\mathbf{z}}{d\mathbf{x}} \right| = \pi(f^{-1}(\mathbf{x})) \left| \det \frac{df^{-1}}{d\mathbf{x}} \right|$$

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

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

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

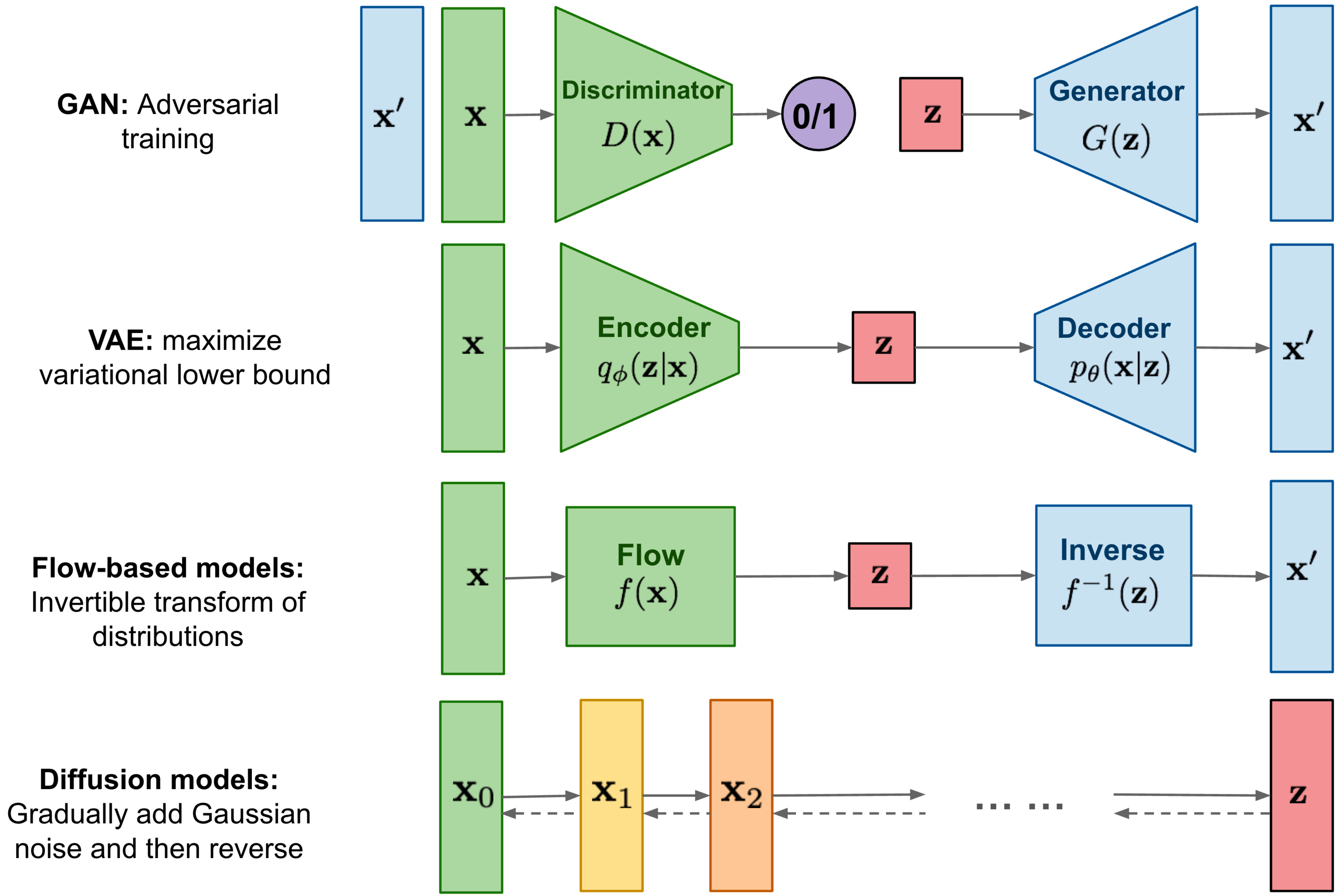
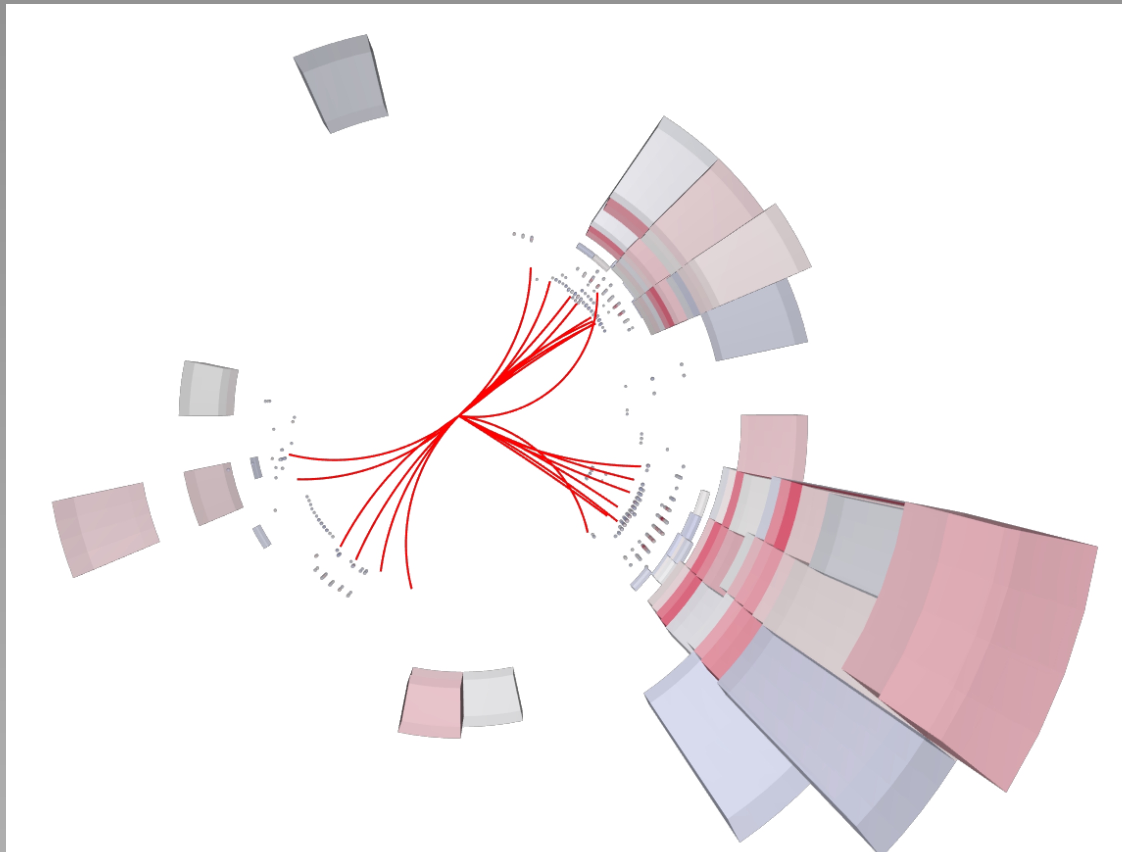
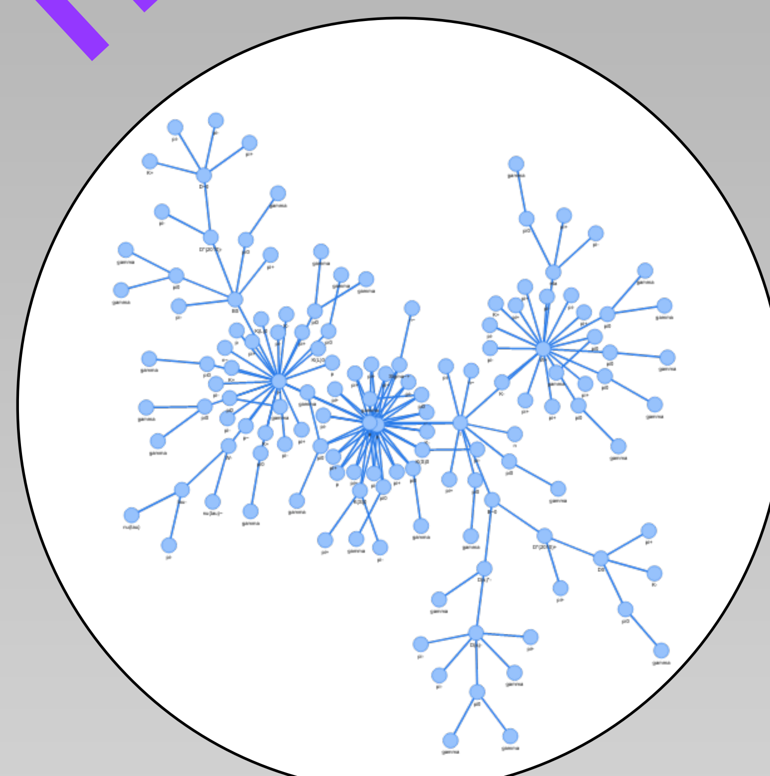


Fig from : https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

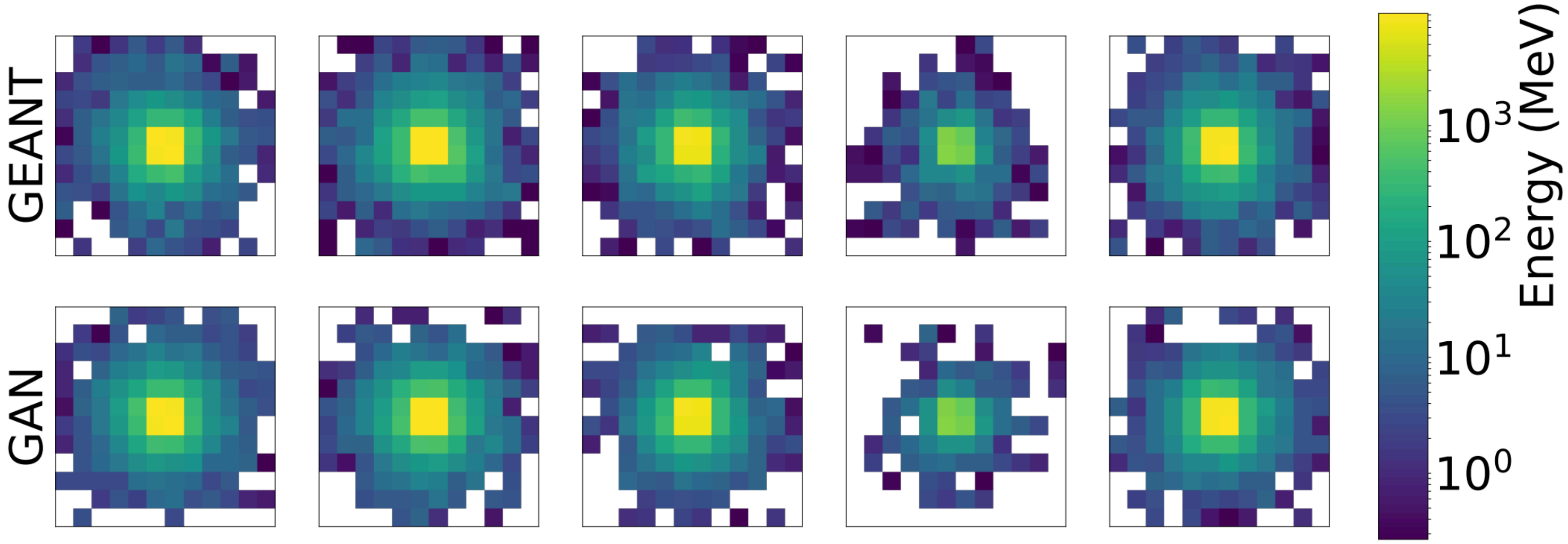


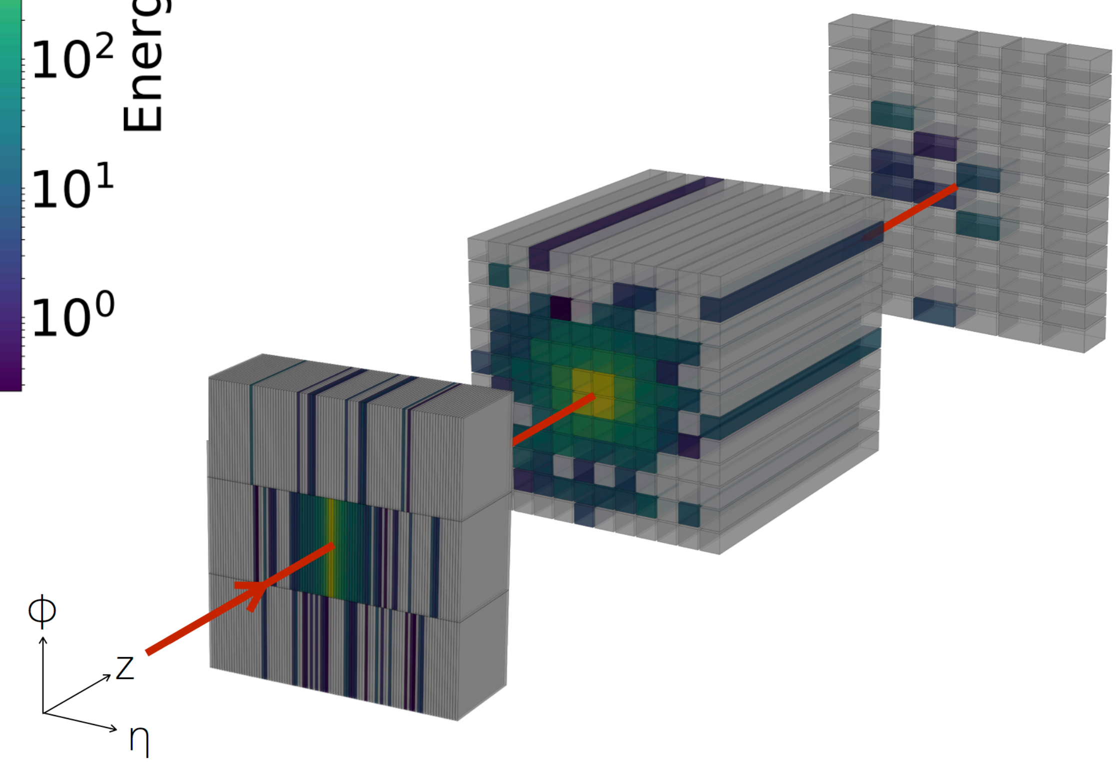
Generative models for calorimeter simulations

$$f_{\{\theta\}}\left(\quad , z \right)$$

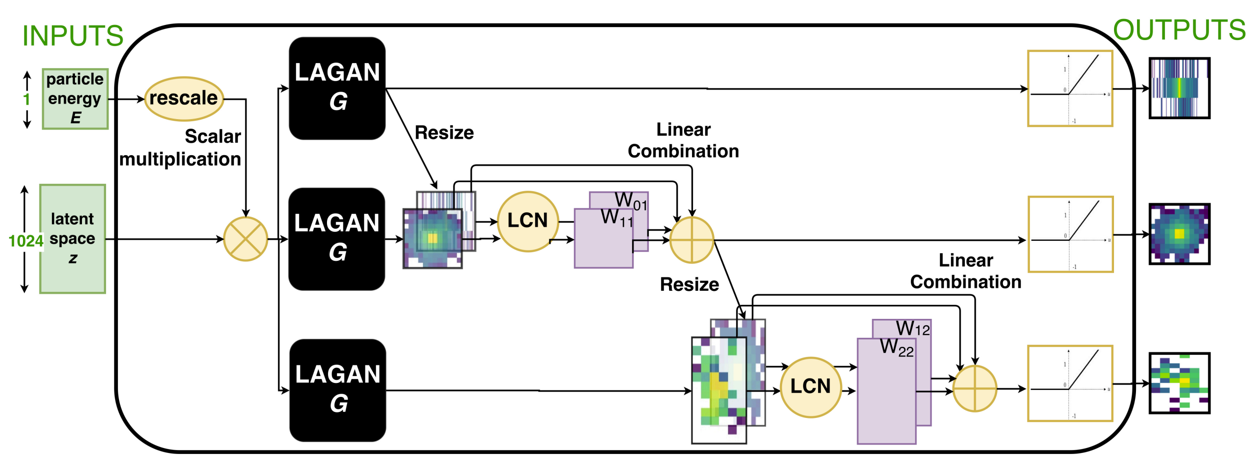


Detector simulation using ML

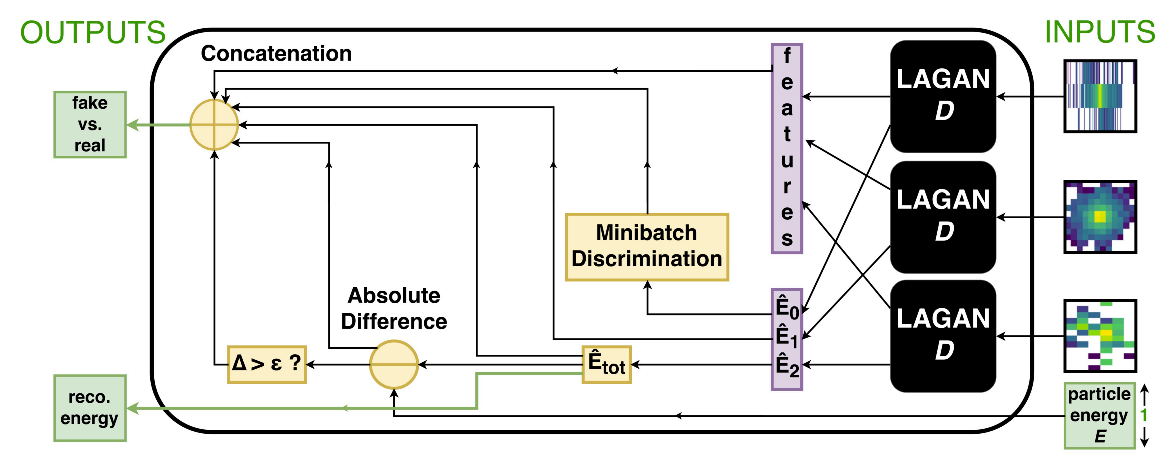




CaloGAN 1705.02355
[Michela Paganini](), [Luke de Oliveira](), [Benjamin Nachman]()

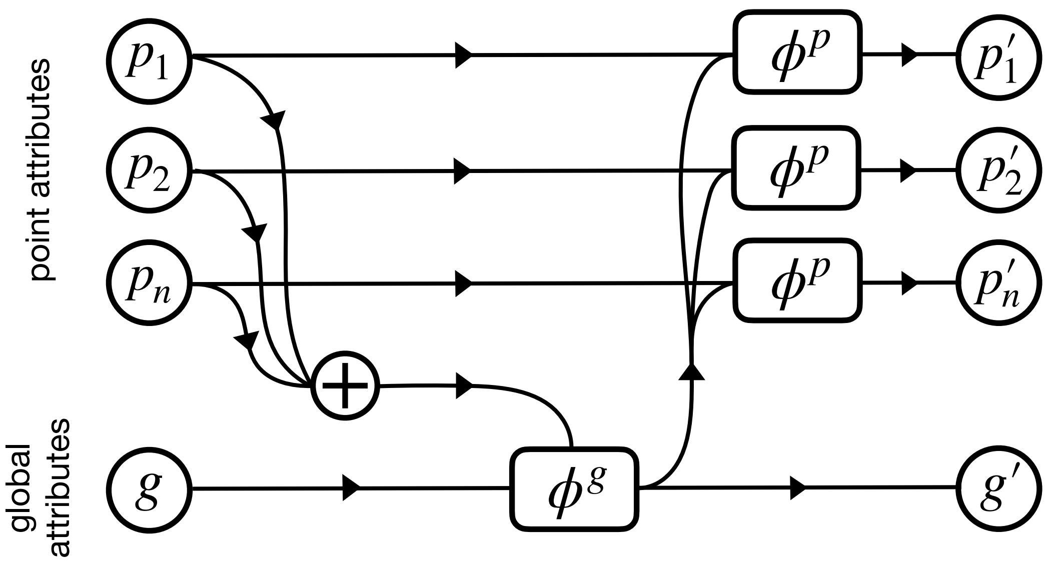


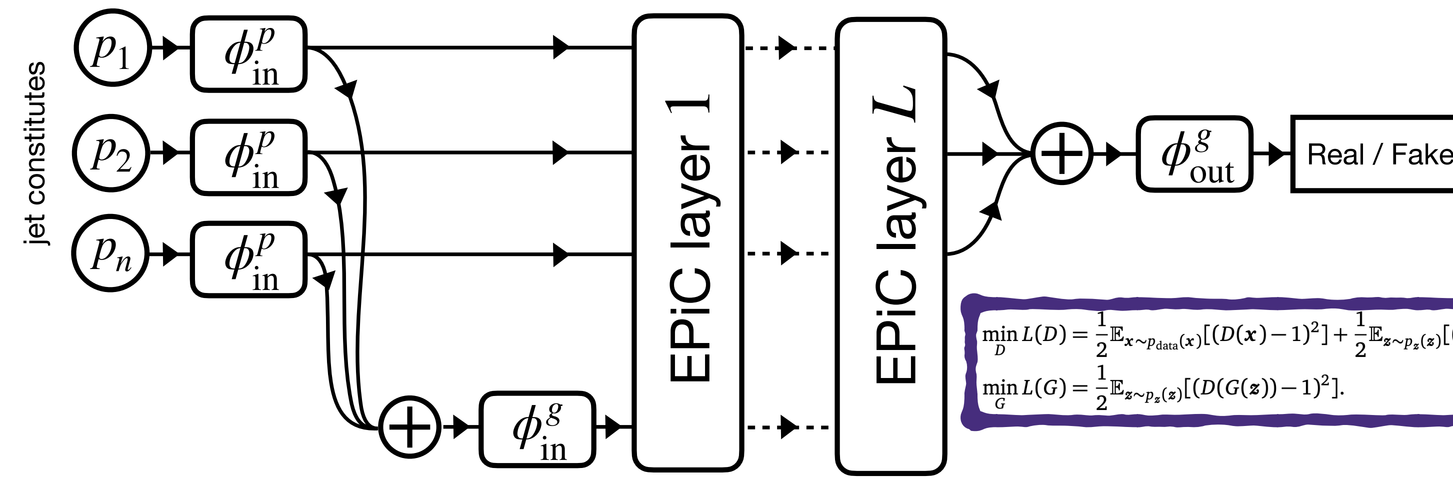
Generator



Discriminator

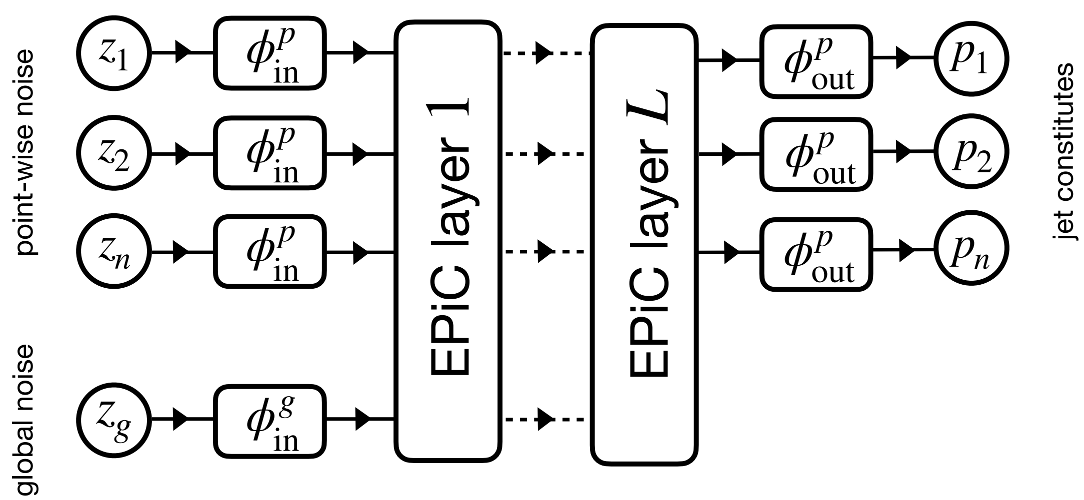
Detector simulation using ML





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

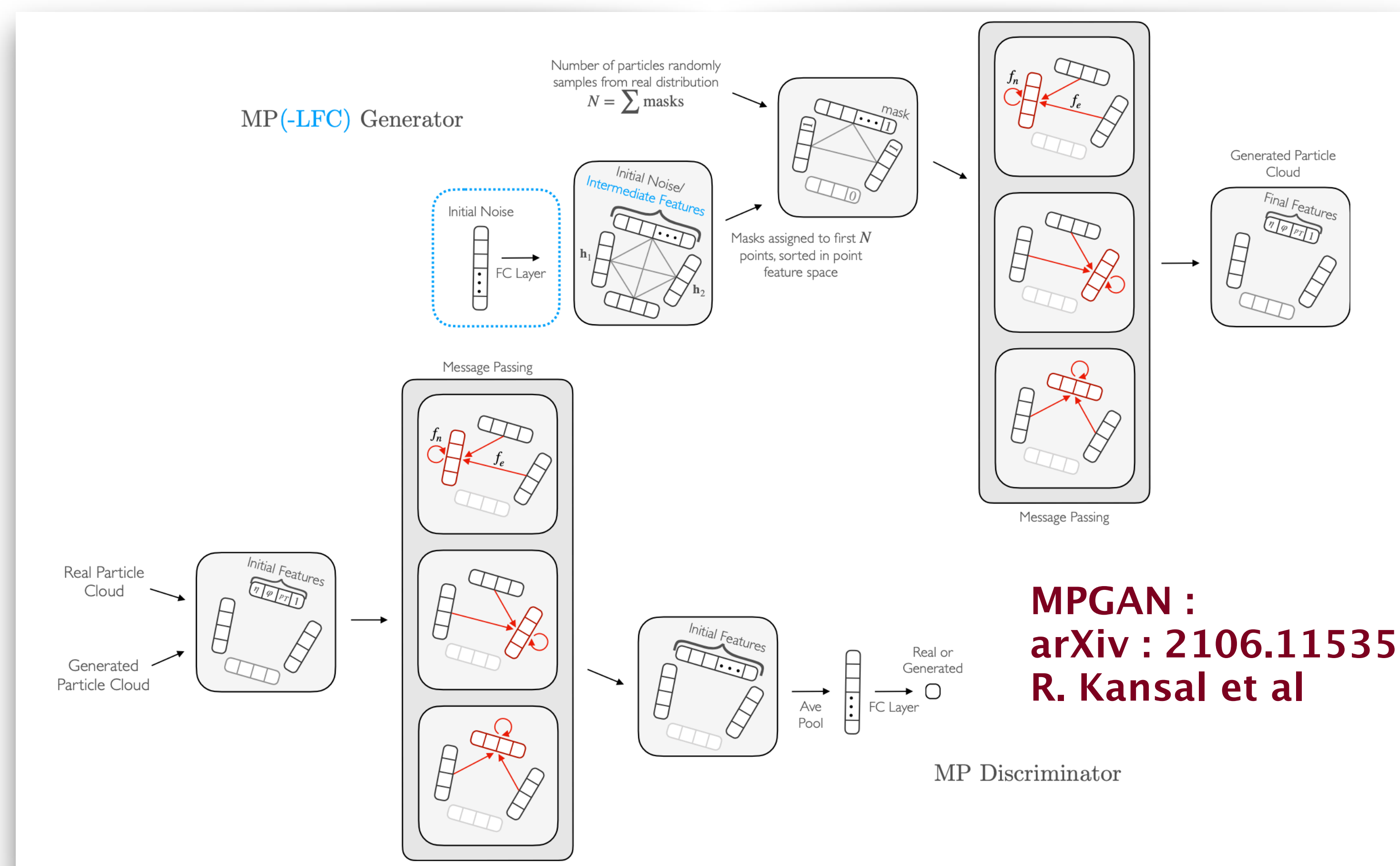
EPiC-GAN : SciPost Phys. 15, 130 (2023) Erik Buhmann, Gregor Kasieczka, Jesse Thaler



2305.10475

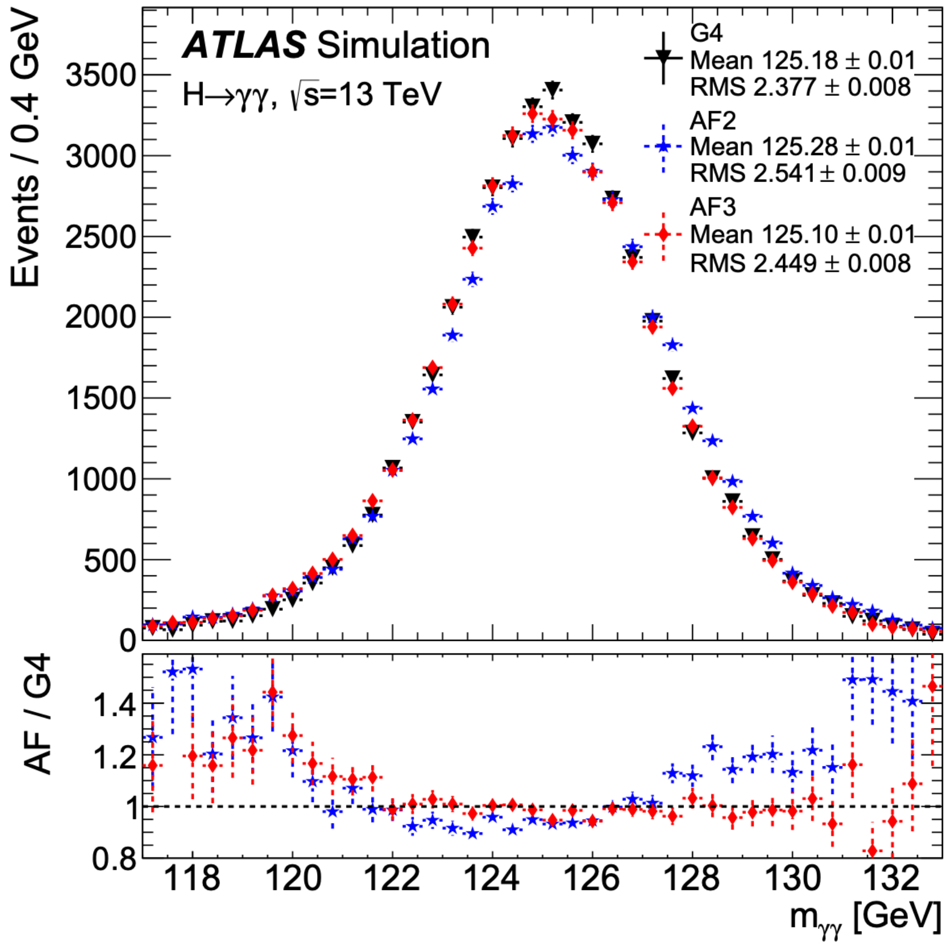
Jet Diffusion versus JetGPT — Modern Networks for the LHC

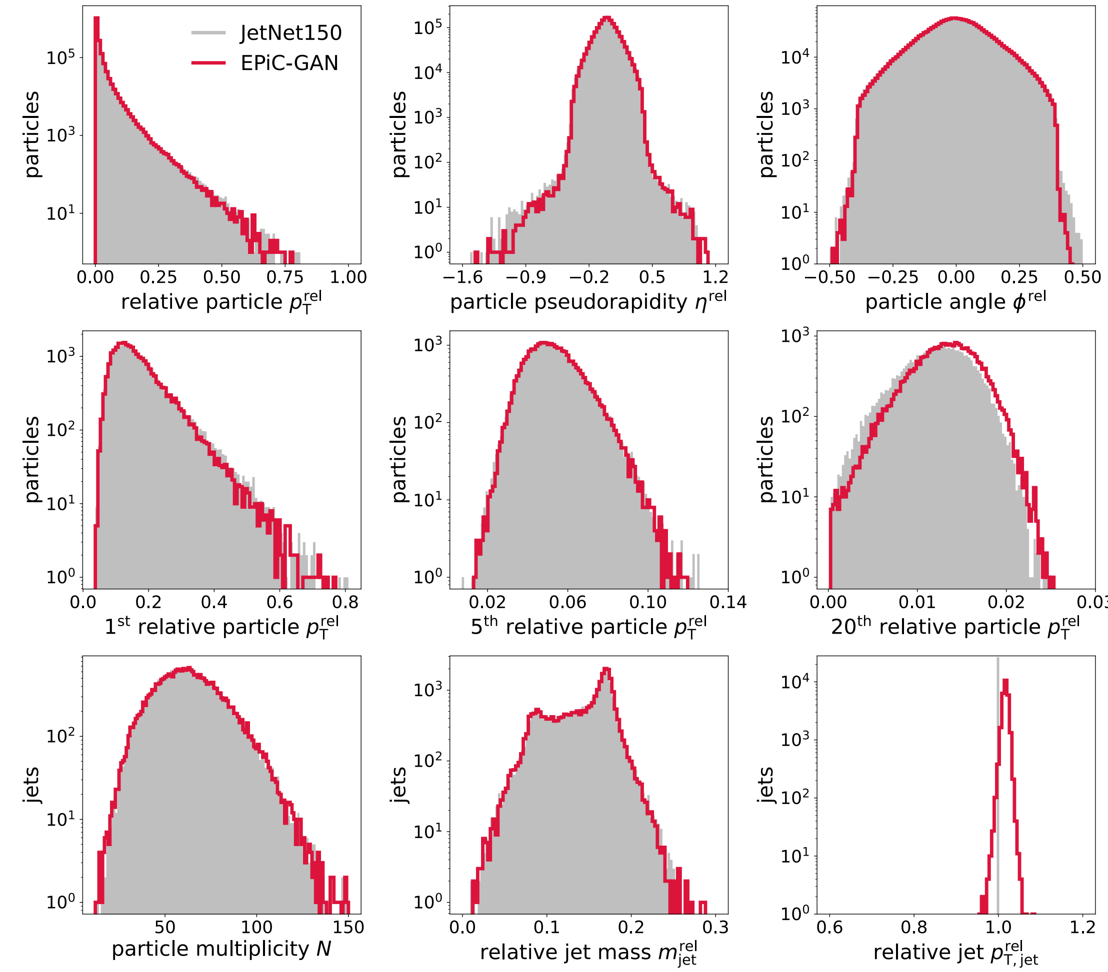
Anja Butter[1,2], Nathan Huetsch[1], Sofia Palacios Schweitzer[1], Tilman Plehn[1], Peter Sorrenson[3], and Jonas Spinner[1]



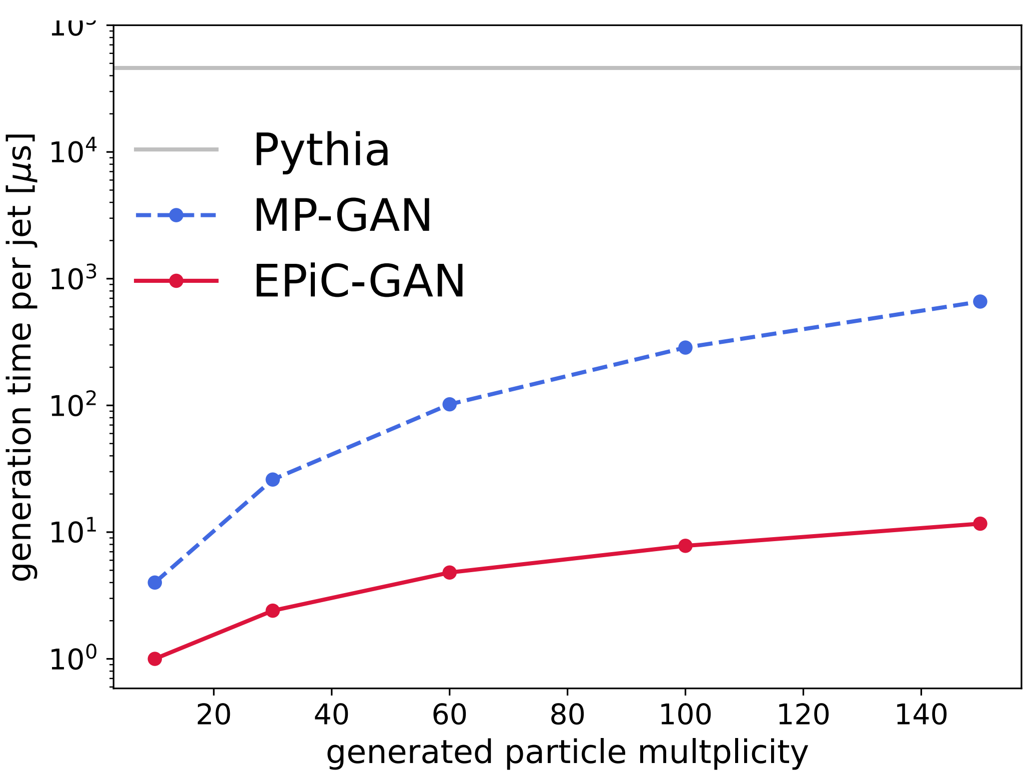
MPGAN : arXiv : 2106.11535 R. Kansal et al

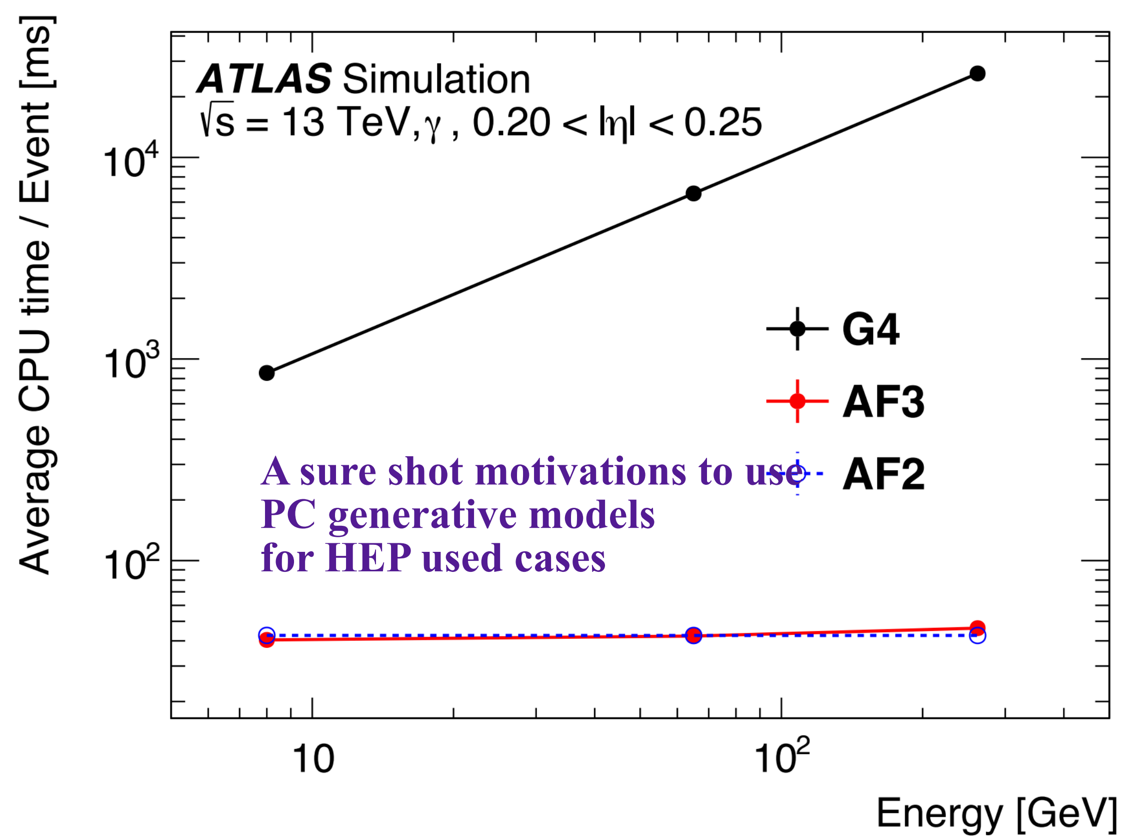
The major gain





arXiv : 2109.02551



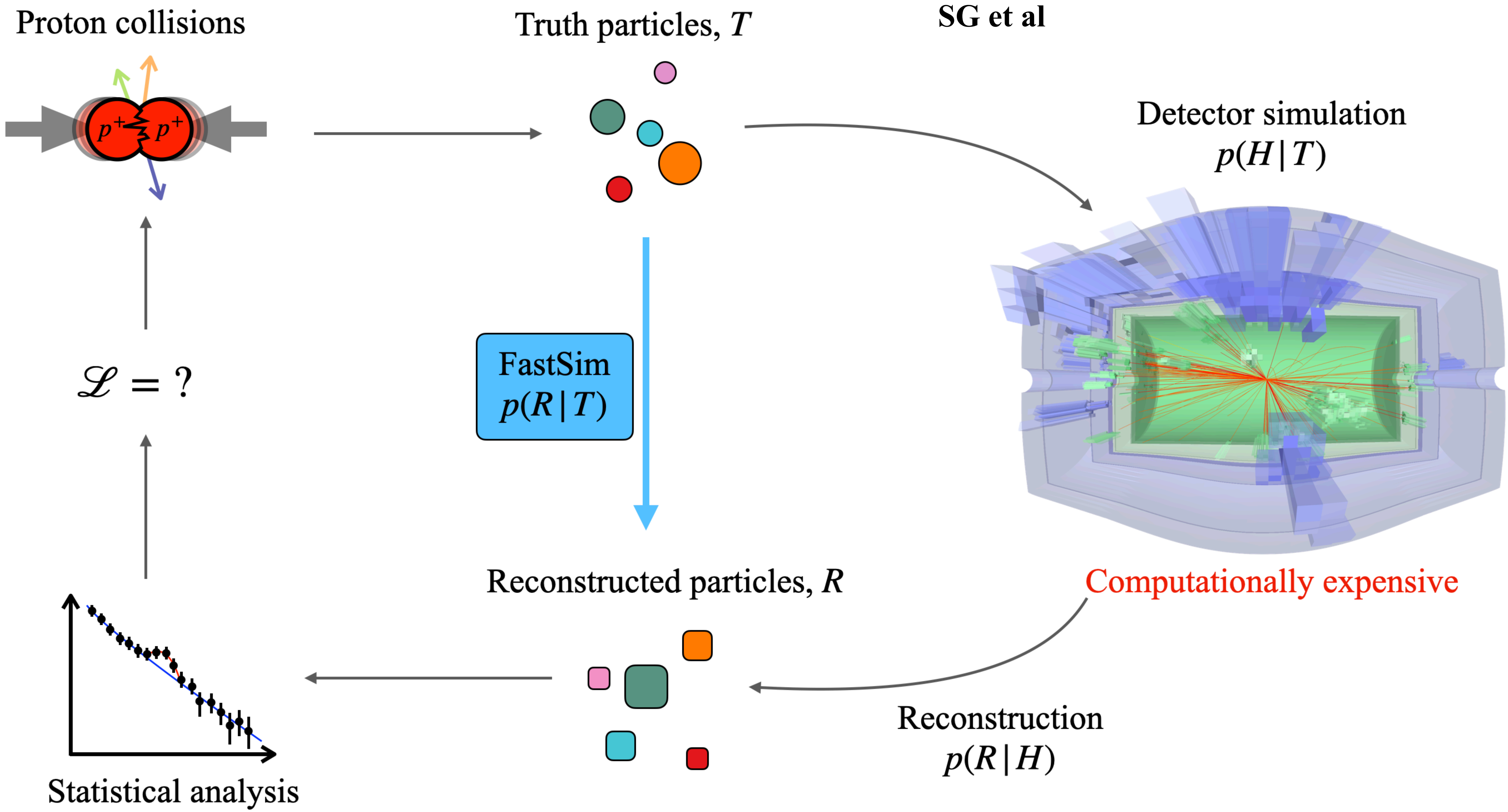


A sure shot motivations to use PC generative models for HEP used cases

A generative model for Particle-flow

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

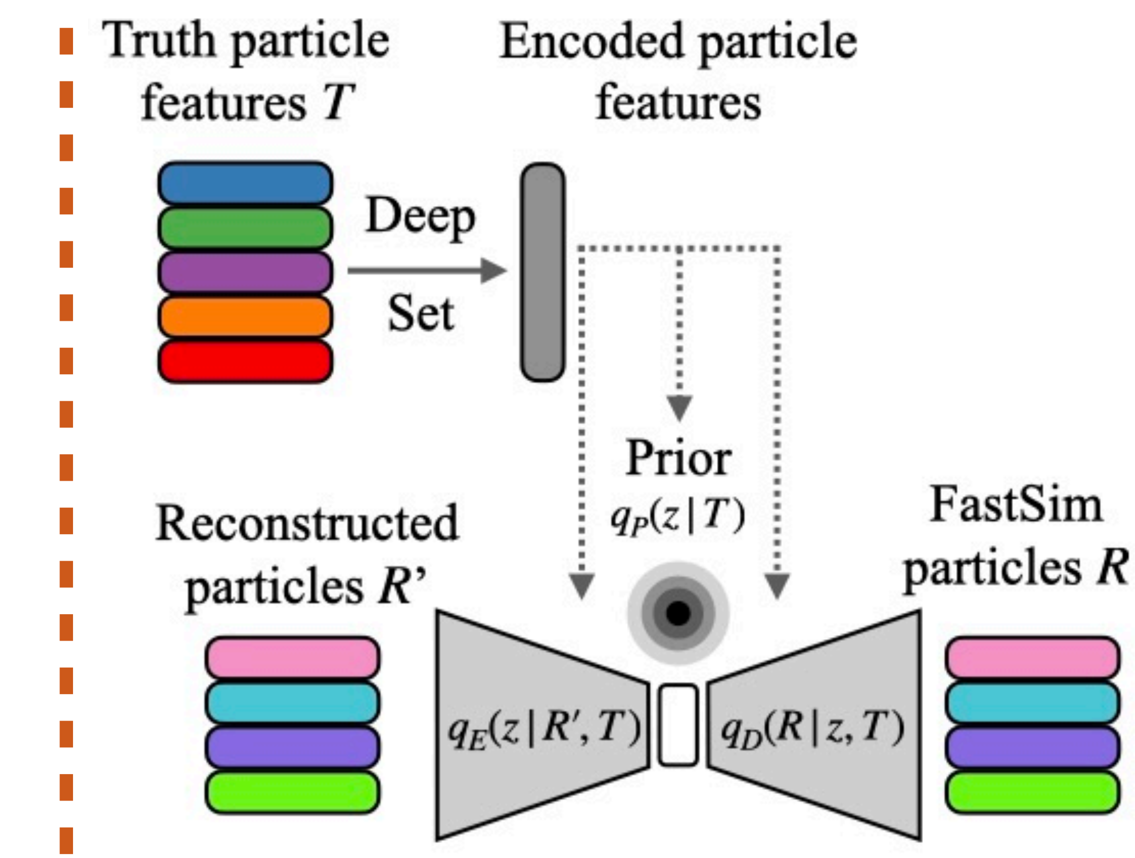
SG et al



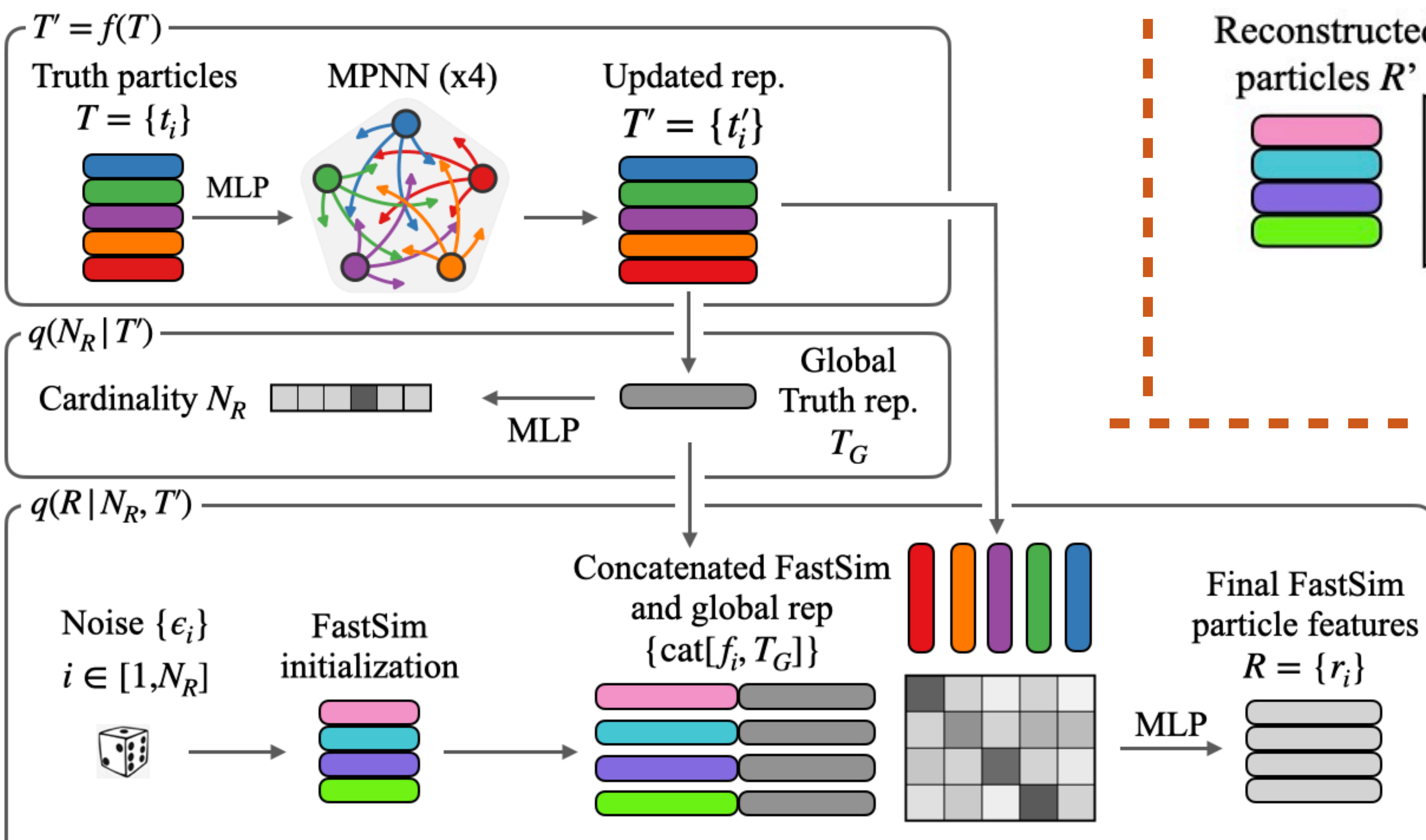
$$R \sim p(R|T) = \int \mathrm{d}H\, \delta(R(H) - R)\, p_{\mathrm{sim}}(H|T).$$

The task of constrained set generation

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



C-VAE

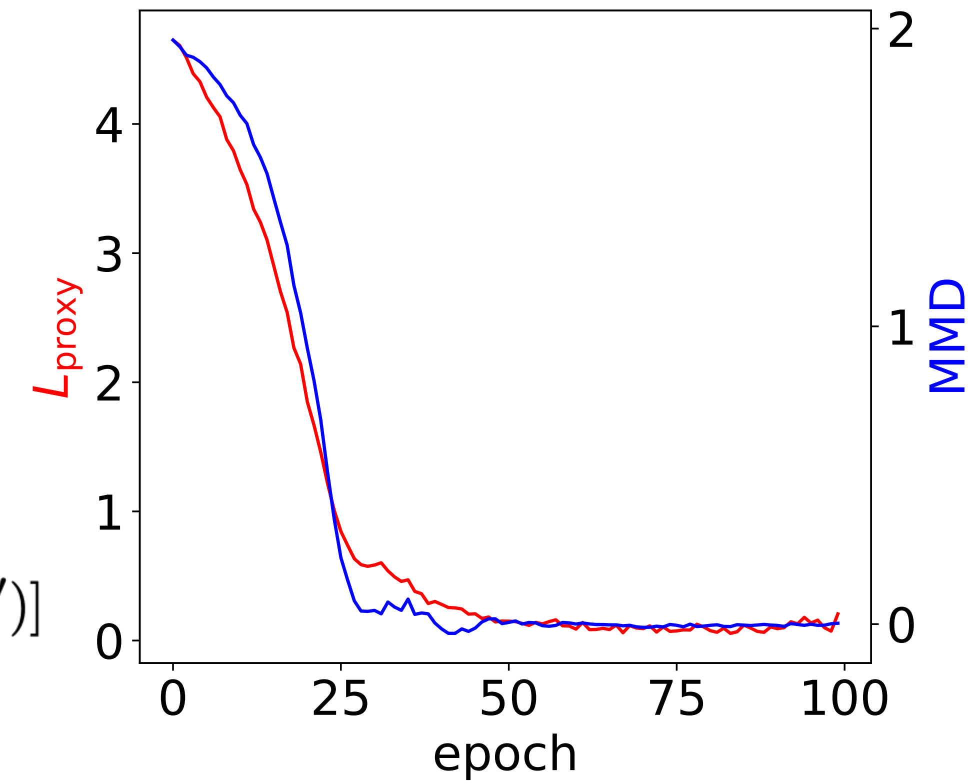


GNN+SA

The task of constrained set generation

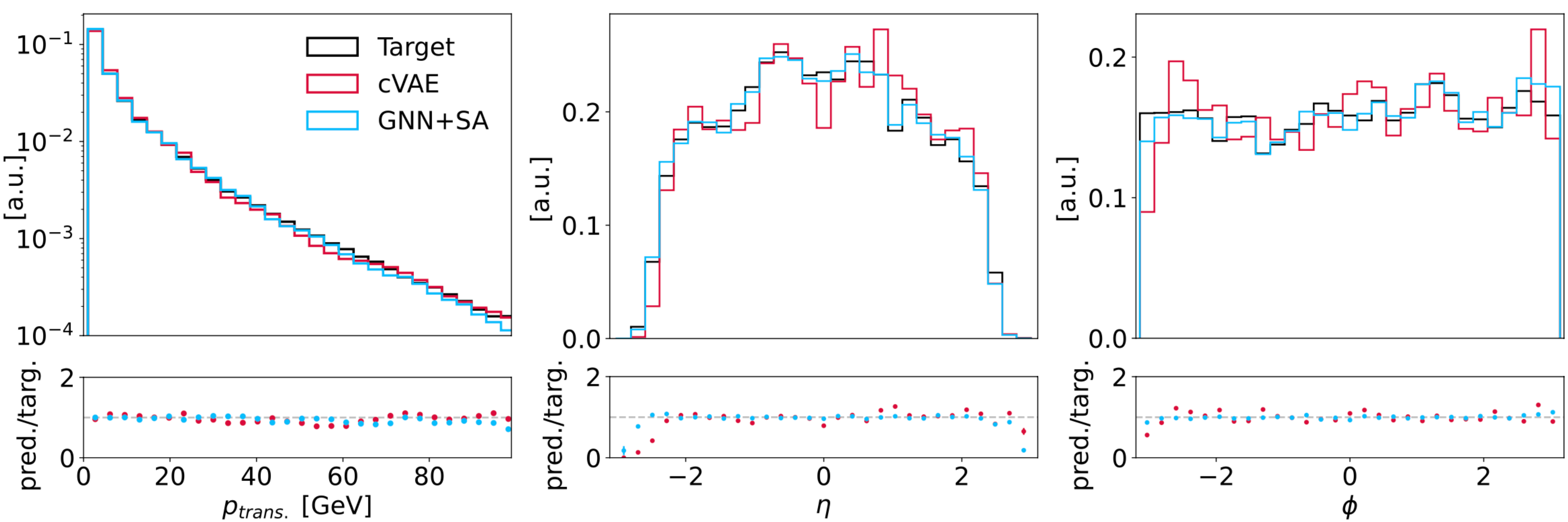
The cVAE training is done by optimizing negative evidence lower bound (ELBO) loss :

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

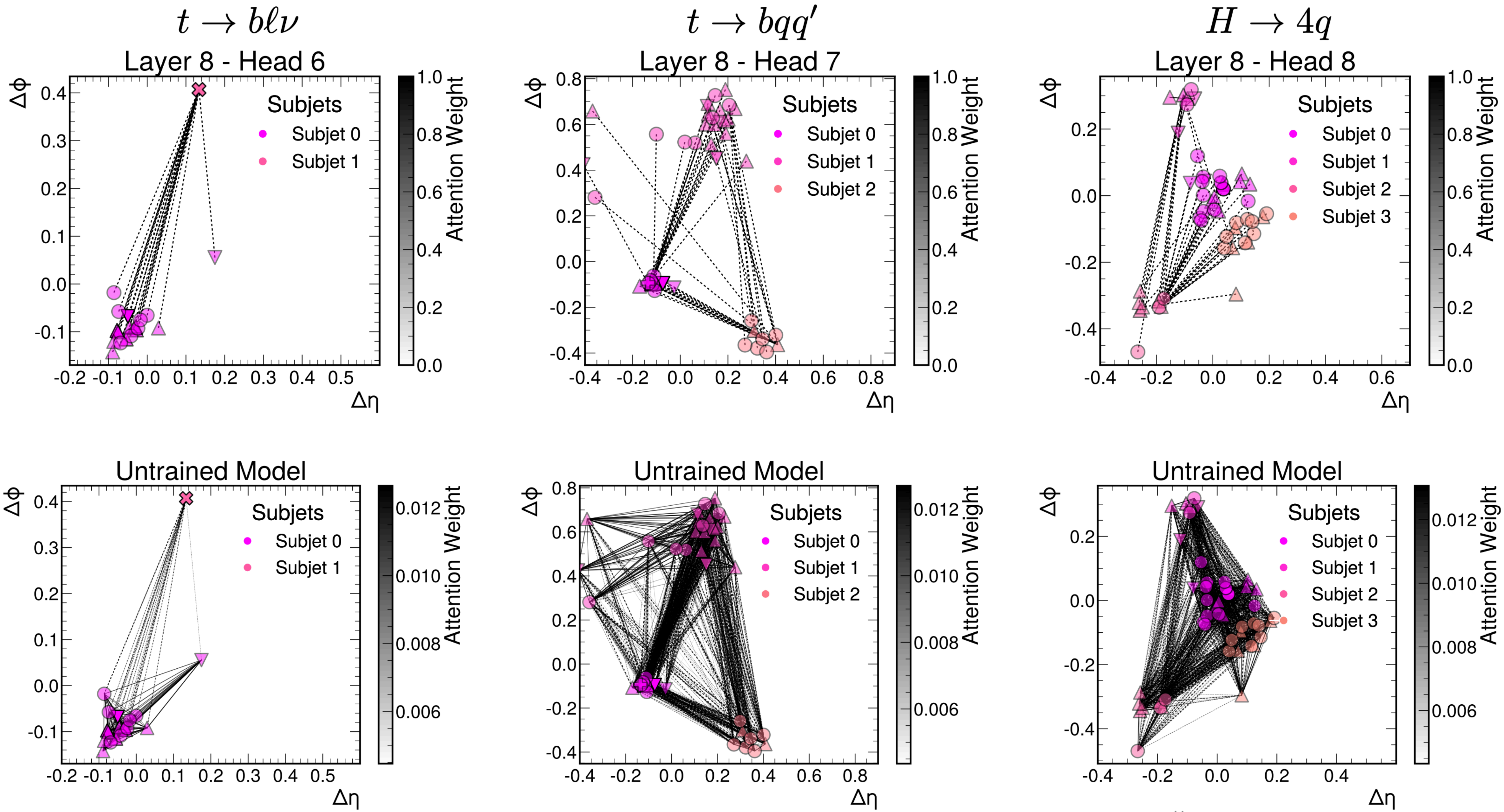


For GNN+SA, we try a regular Hungarian loss and also MMD (maximum mean discrepancy) :

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



In the direction of interpretation

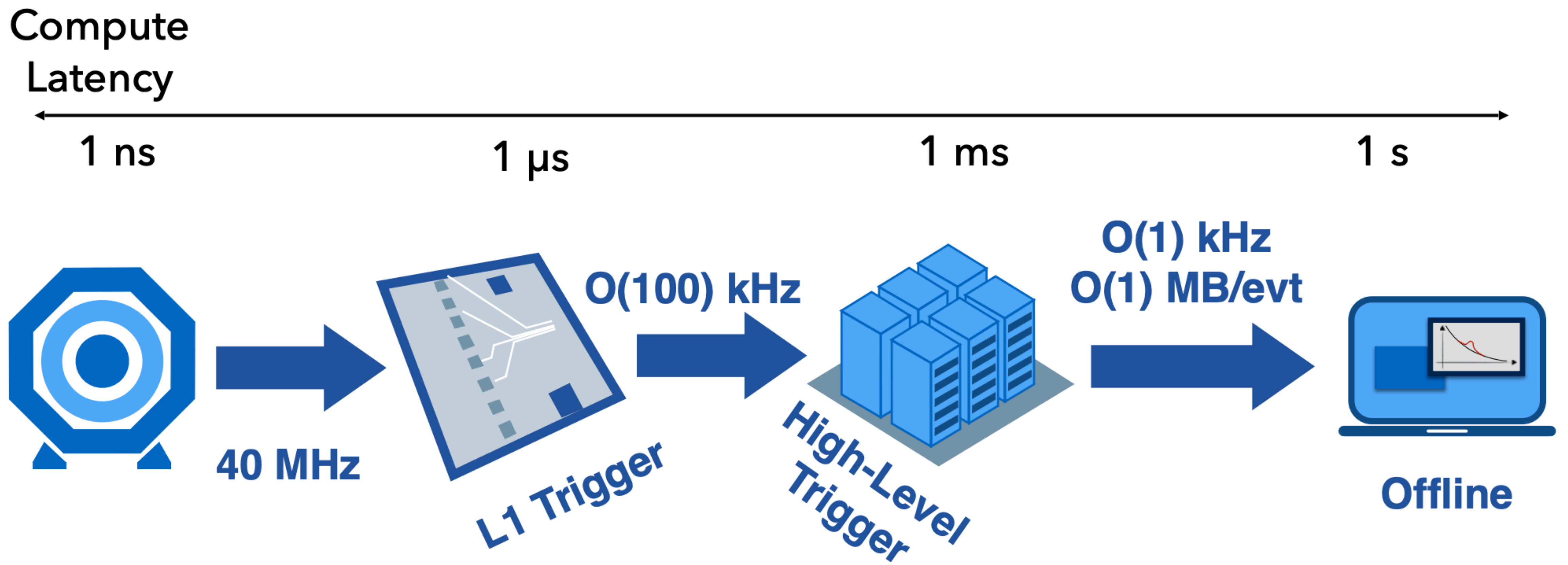


arXiv:2412.03673

✖: muon
▲:charged hadron
▼: neutral hadron
●: photon
✚: electron

For leptonic top decay, the attention score is more on lepton.
For hadronic top decay, the attention score is more on subjets equally.

Fast ML for jets



ASICs

FPGAs

CPUs
GPUs

CPUs
GPUs
FPGAs

Exabyte-scale datasets

Challenges:

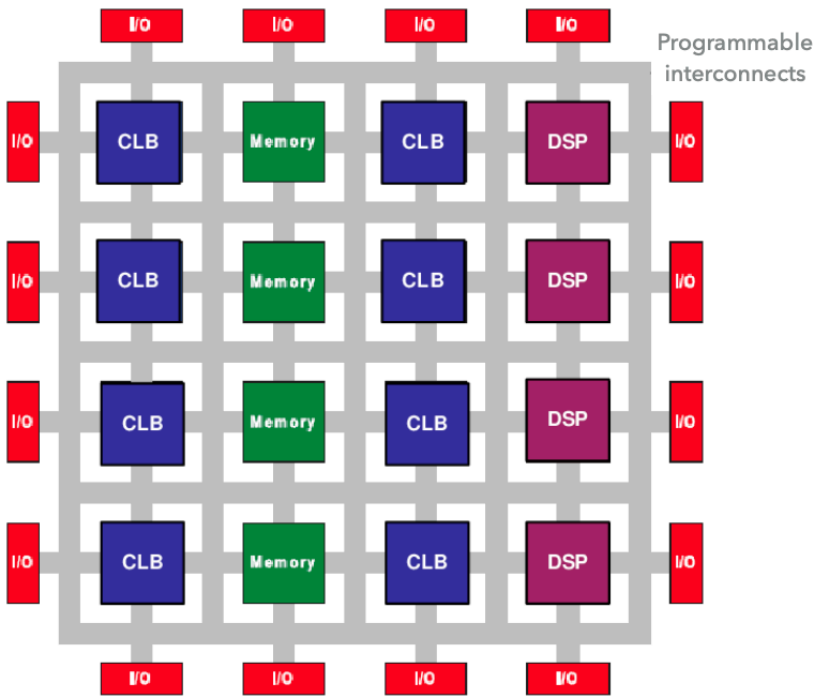
Each collision produces $O(10^3)$ particles

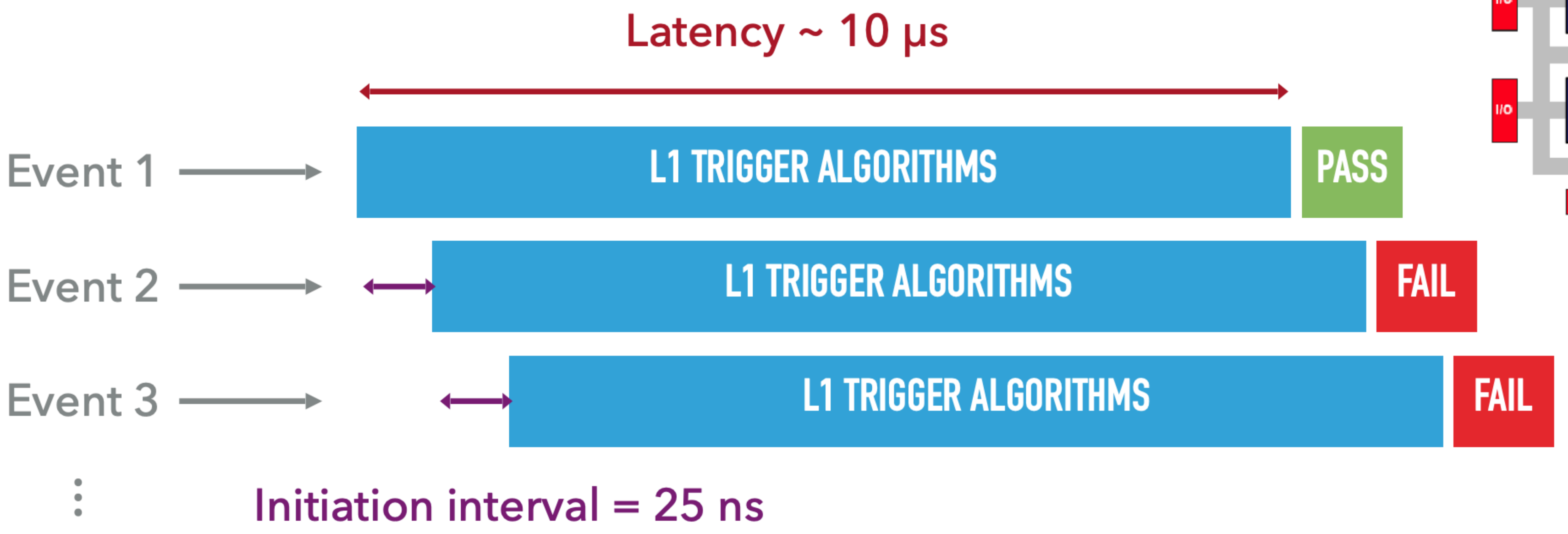
The detectors have $O(10^8)$ sensors

Extreme data rates of O(100 TB/s)

The hard part for fast ML

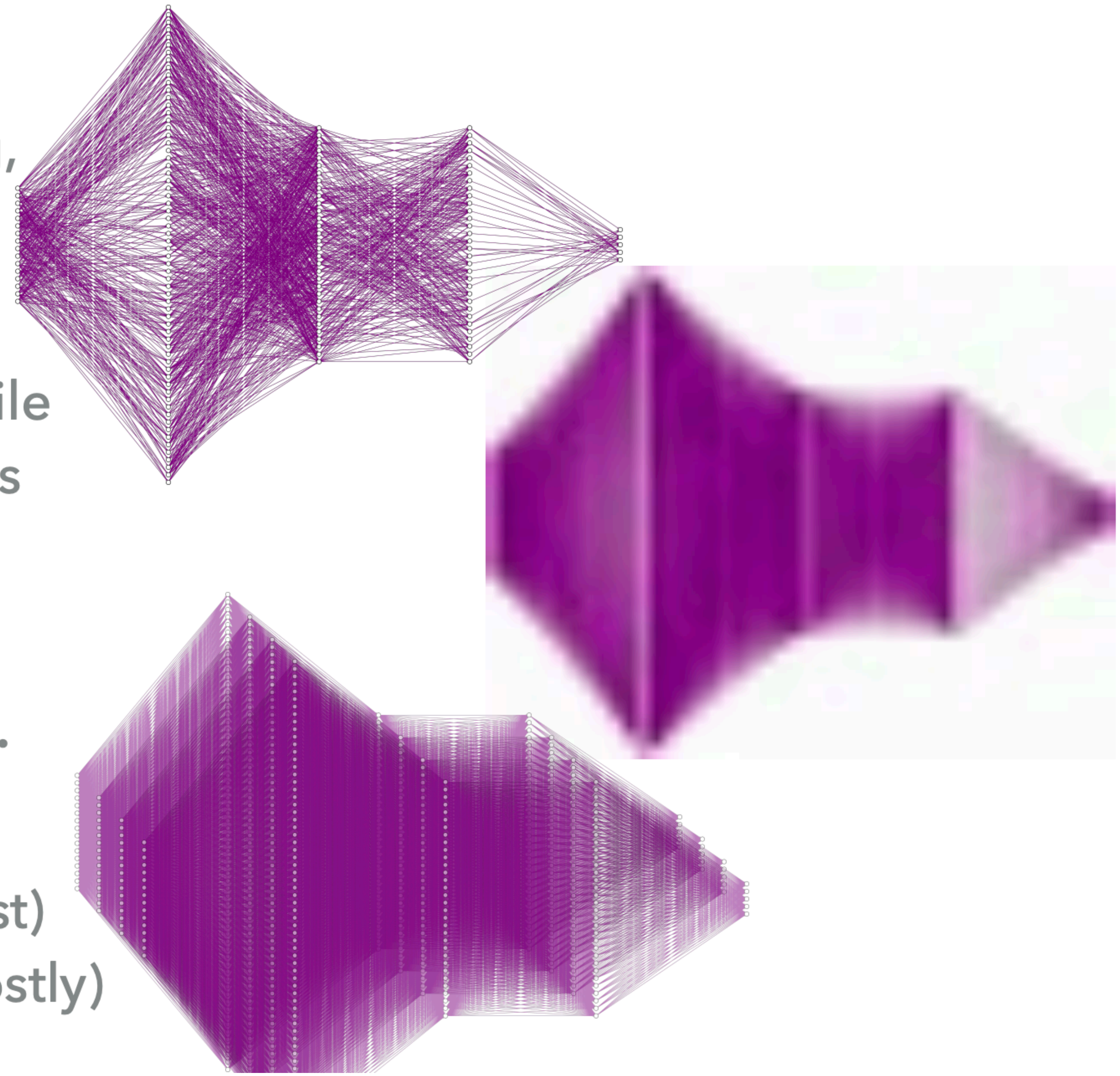
- Reconstruct all events and reject 98% of them in O(10) μs
  - Algorithms have to be <1 μs and process new events every 25 ns
- Latency necessitates all **FPGA** design
  - Algorithms have to fit on <1 FPGA
- How can we satisfy these constraints?





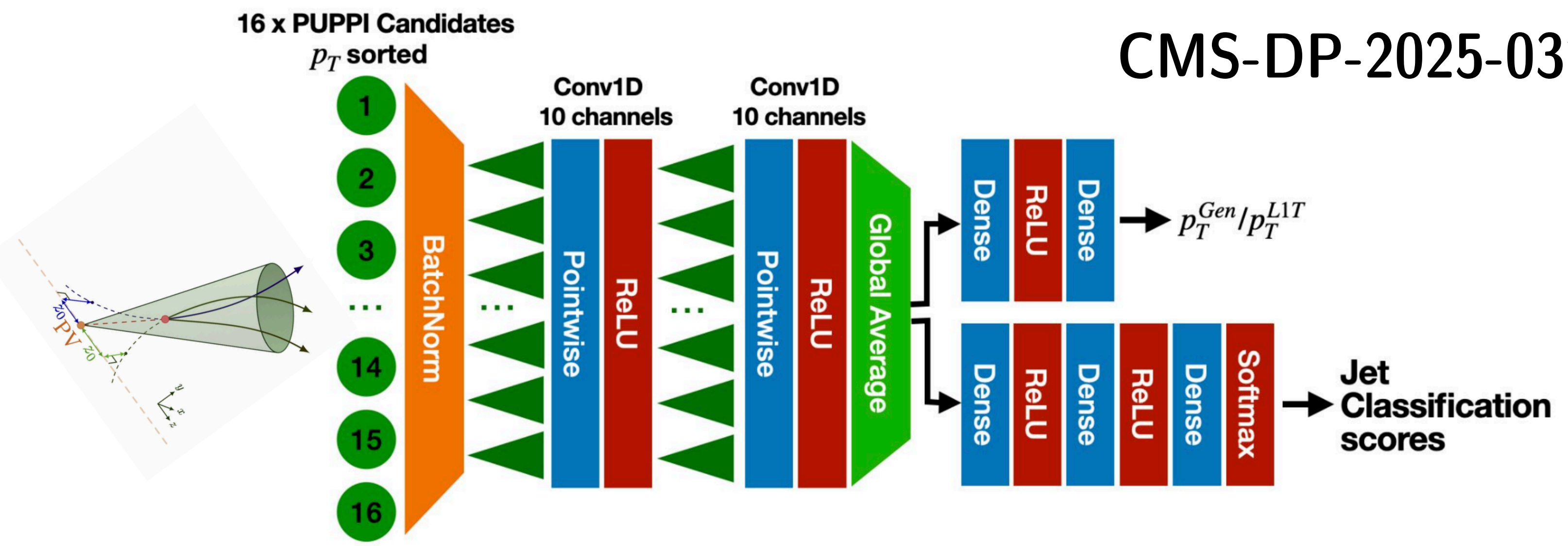
The hard part for fast ML

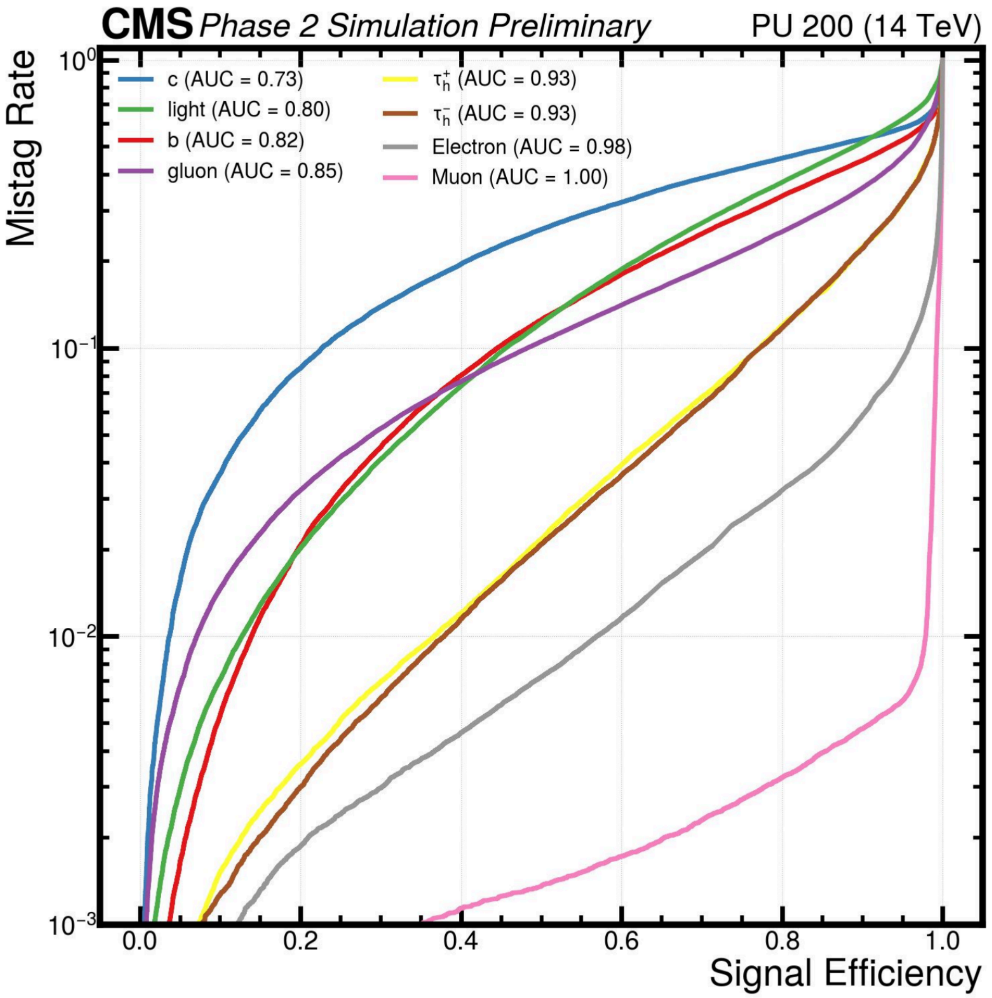
- **Codesign**: intrinsic development loop between ML design, training, and implementation
- Pruning
  - Maintain high performance while removing redundant operations
- Quantization
  - Reduce precision from 32-bit floating point to 16-bit, 8-bit, …
- Parallelization
  - Balance parallelization (how fast) with resources needed (how costly)



Deploy an ML on FPGA

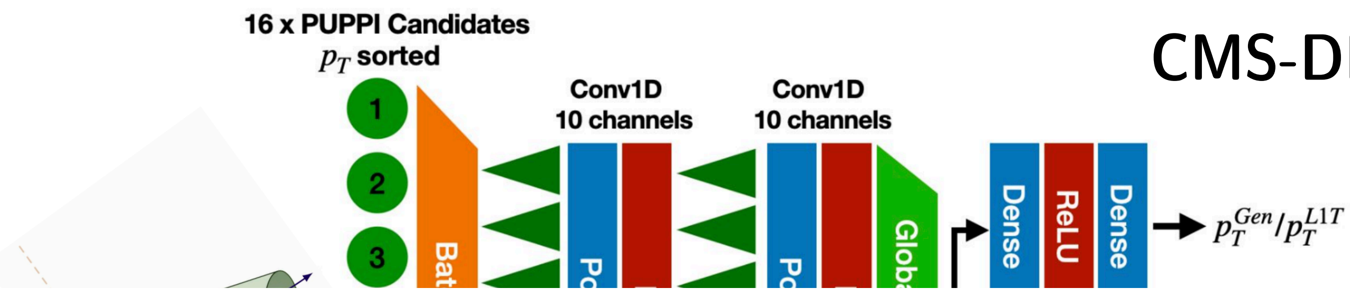
CMS-DP-2025-032

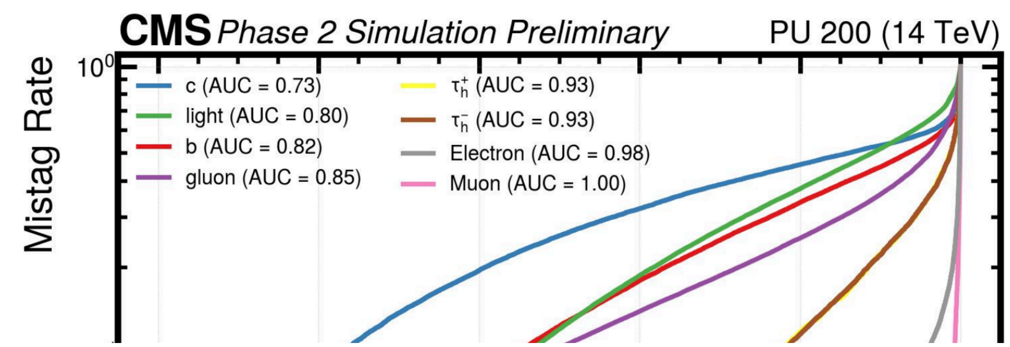


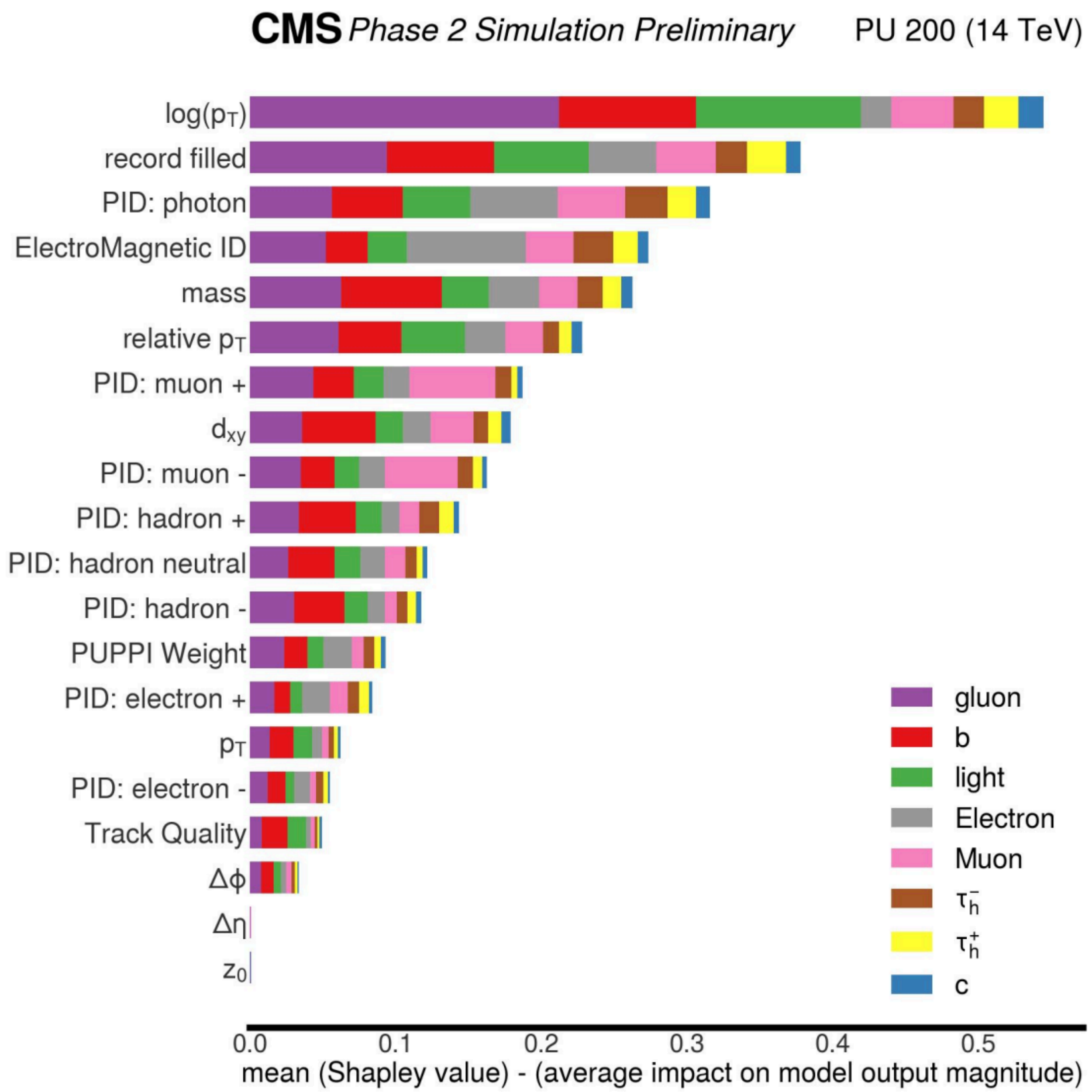


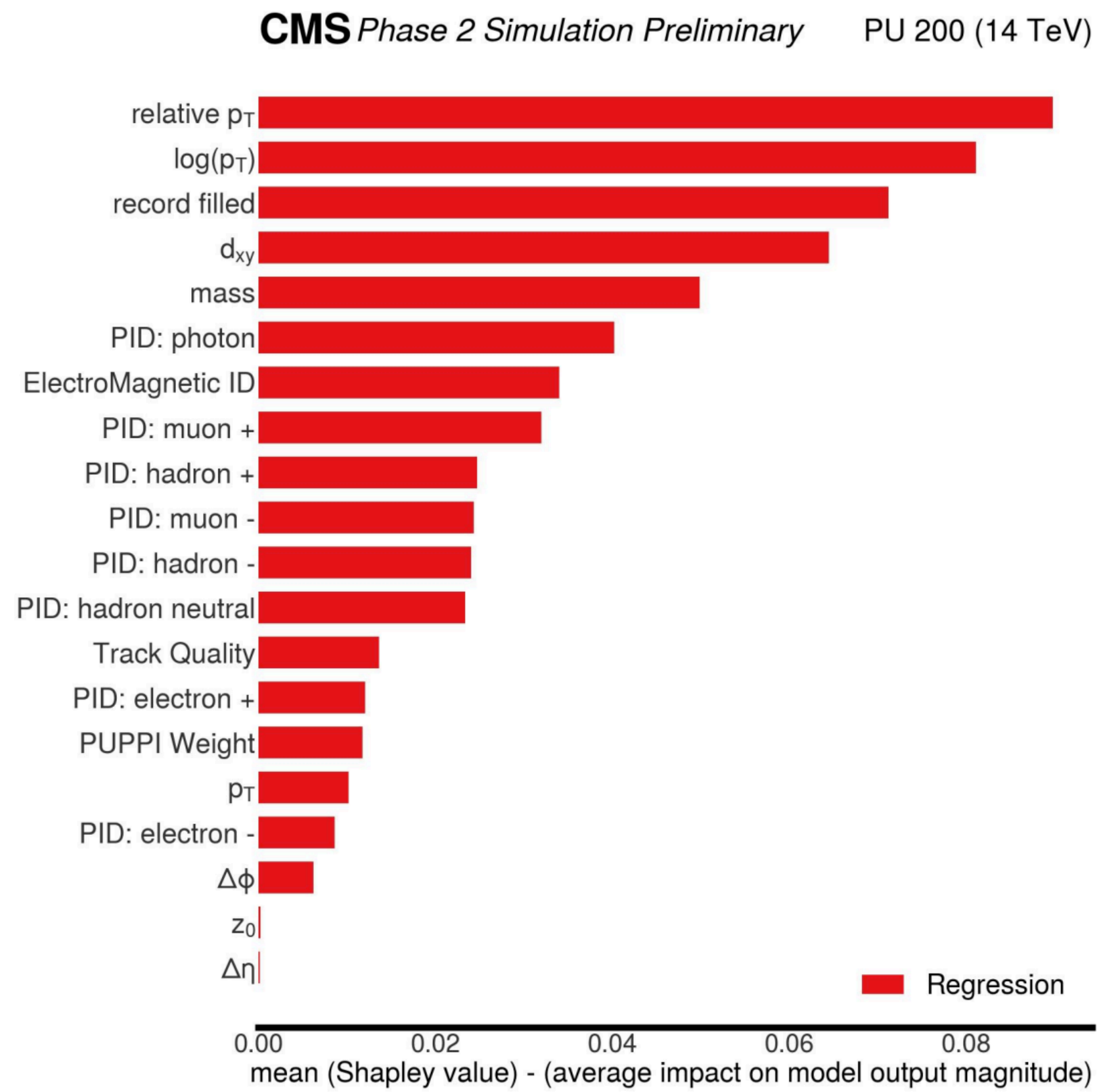
Deploy an ML on FPGA

CMS-DP-2025-032



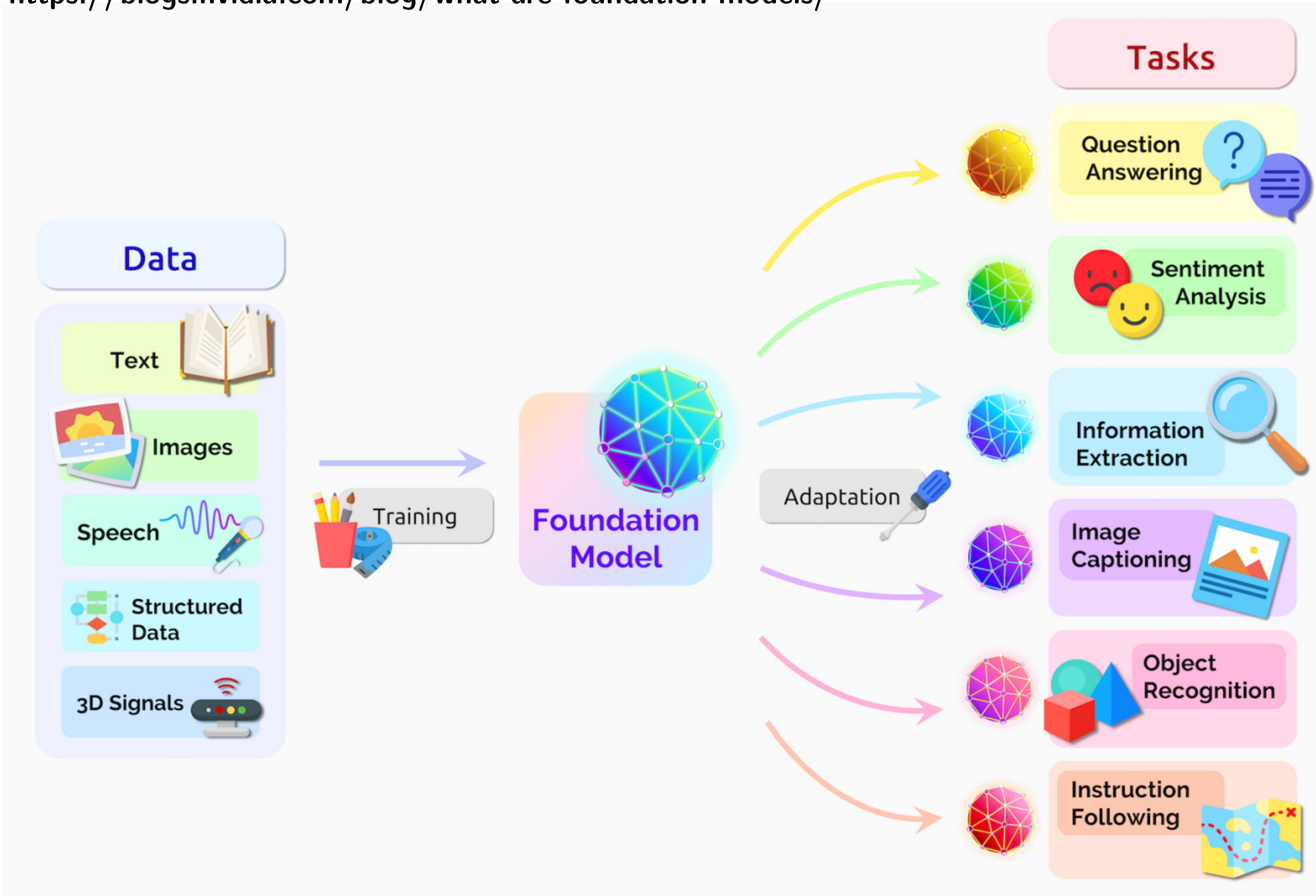




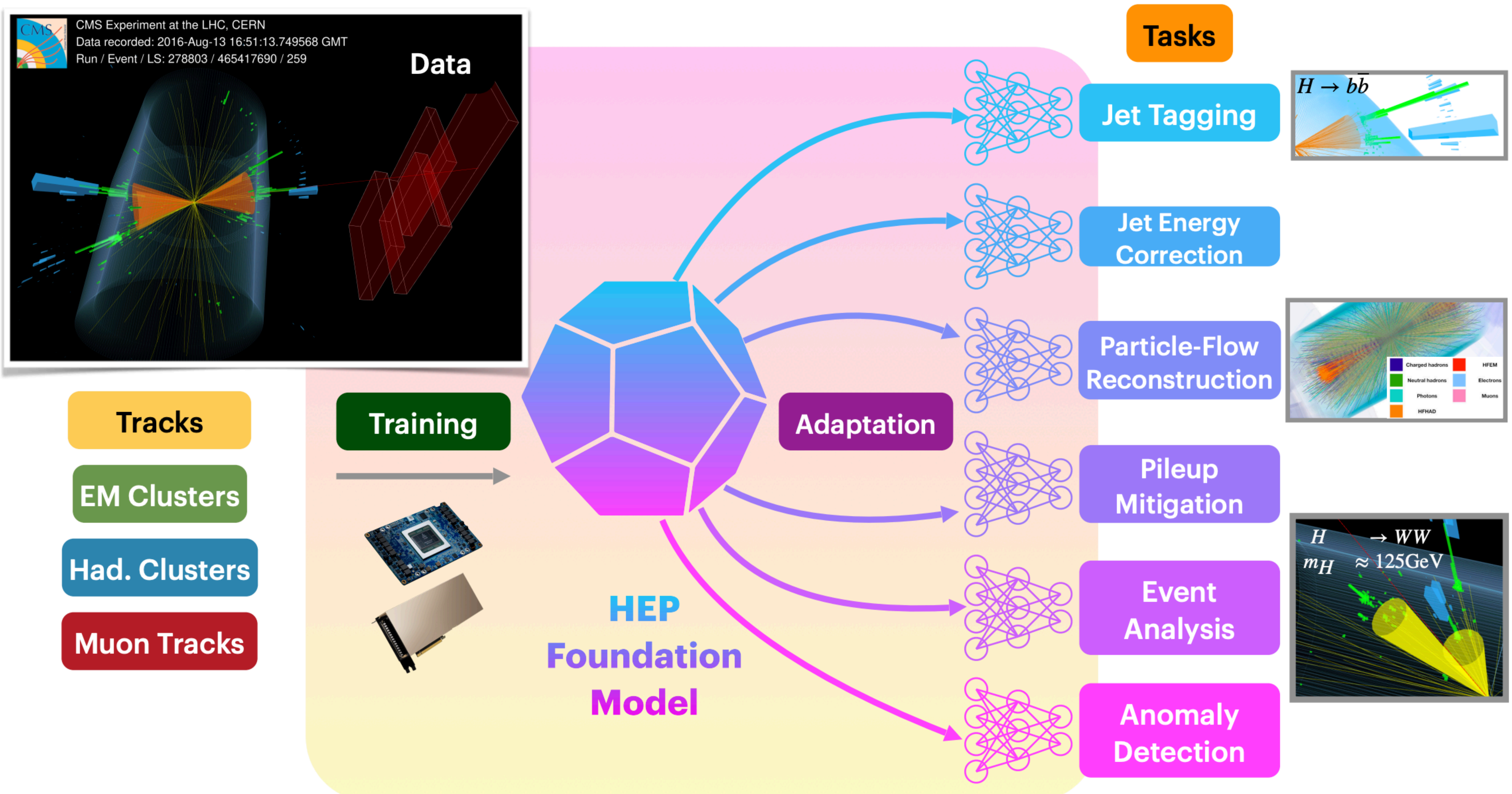


HEP and Jet foundation model?

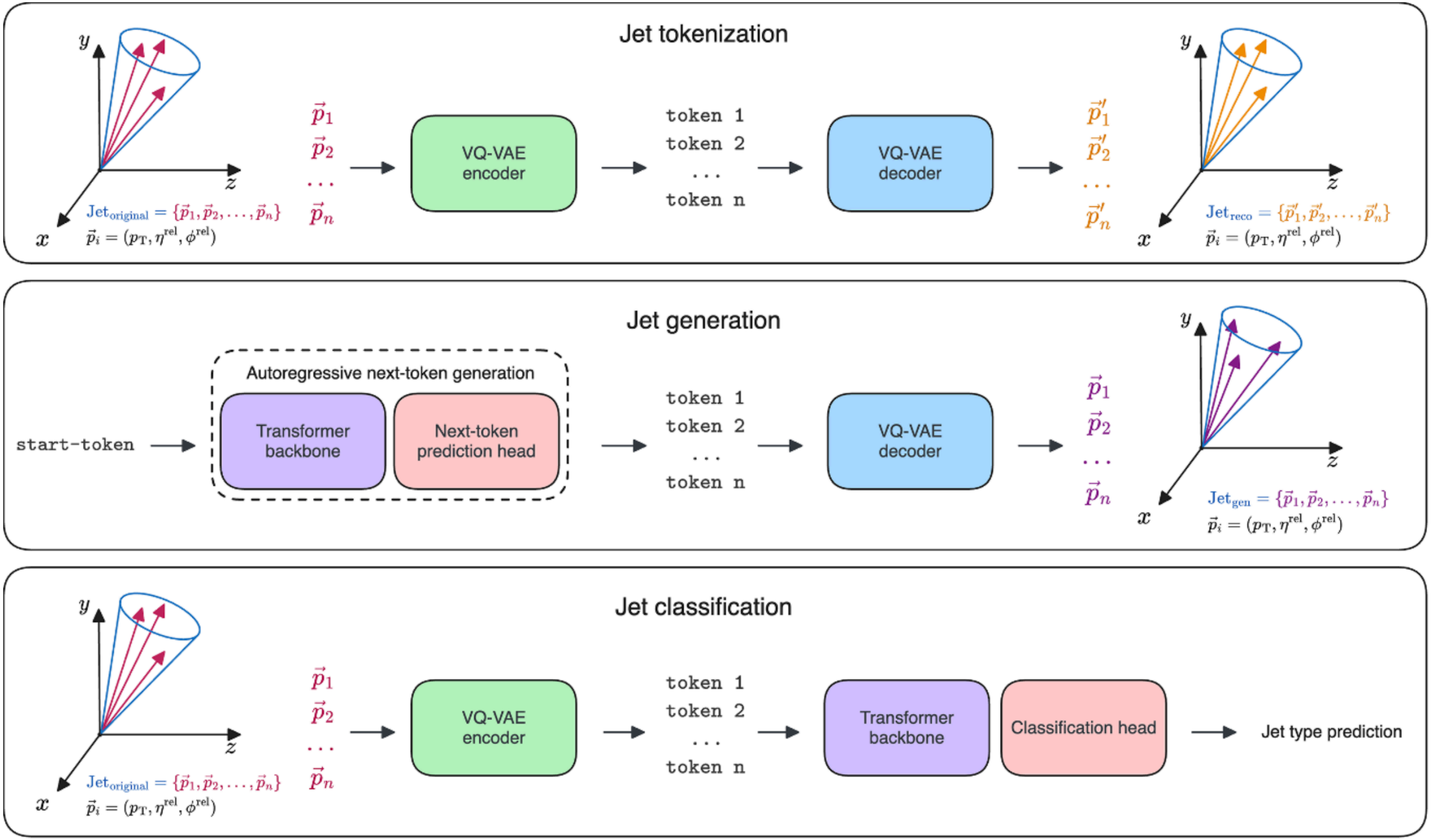
https://blogs.nvidia.com/blog/what-are-foundation-models/



HEP and Jet foundation model?



Omnijet - alpha

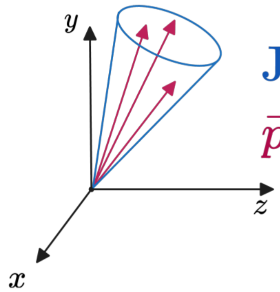


Mach. Learn.: Sci. Technol. 5 (2024) 035031

Omnijet - alpha

Mach. Learn.: Sci. Technol. 5 (2024) 035031

OmniJet-α [1] uses **next token prediction as a pre-training task** for a jet tagger

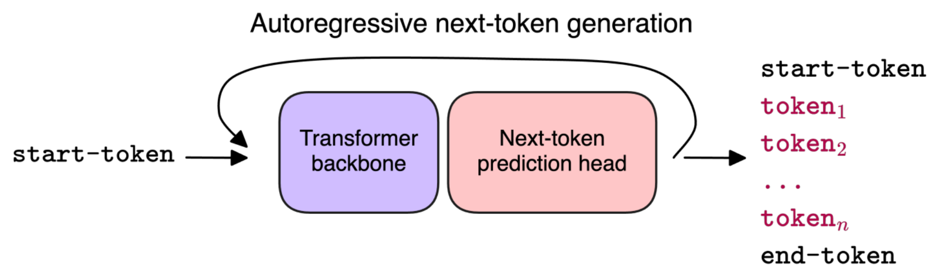


$$\text{Jet} = \{\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_n\}$$
$$\vec{p}_i = (p_\mathrm{T}, \Delta\eta, \Delta\phi)$$

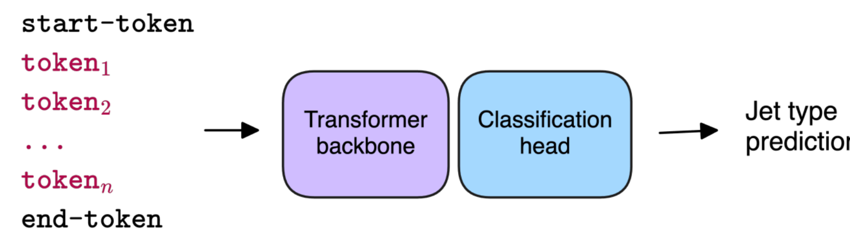
Constituents are tokenized with a VQ-VAE [2-4]

$$\text{Jet} = \{\texttt{start-token}, \texttt{token}_1, \ldots, \texttt{token}_n, \texttt{end-token}\}$$
$$\texttt{token}_i = \text{integer value} \in [1, \ldots, 8192]$$

Self-supervised pre-training of transformer backbone on generative task (next token prediction)



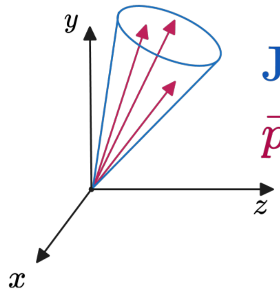
Fine-tuning to classification task:
Swap model head and use pre-trained backbone



Omnijet - alpha

Mach. Learn.: Sci. Technol. 5 (2024) 035031

OmniJet-α [1] uses **next token prediction as a pre-training task** for a jet tagger

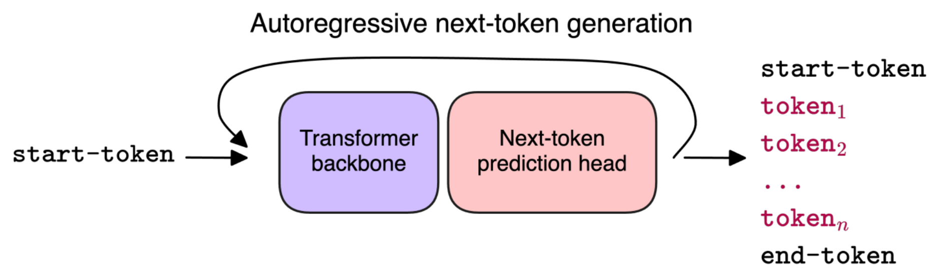


$$\text{Jet} = \{\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_n\}$$
$$\vec{p}_i = (p_\mathrm{T}, \Delta\eta, \Delta\phi)$$

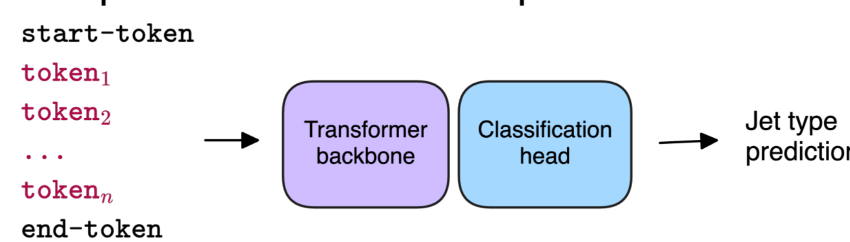
Constituents are tokenized with a VQ-VAE [2-4]

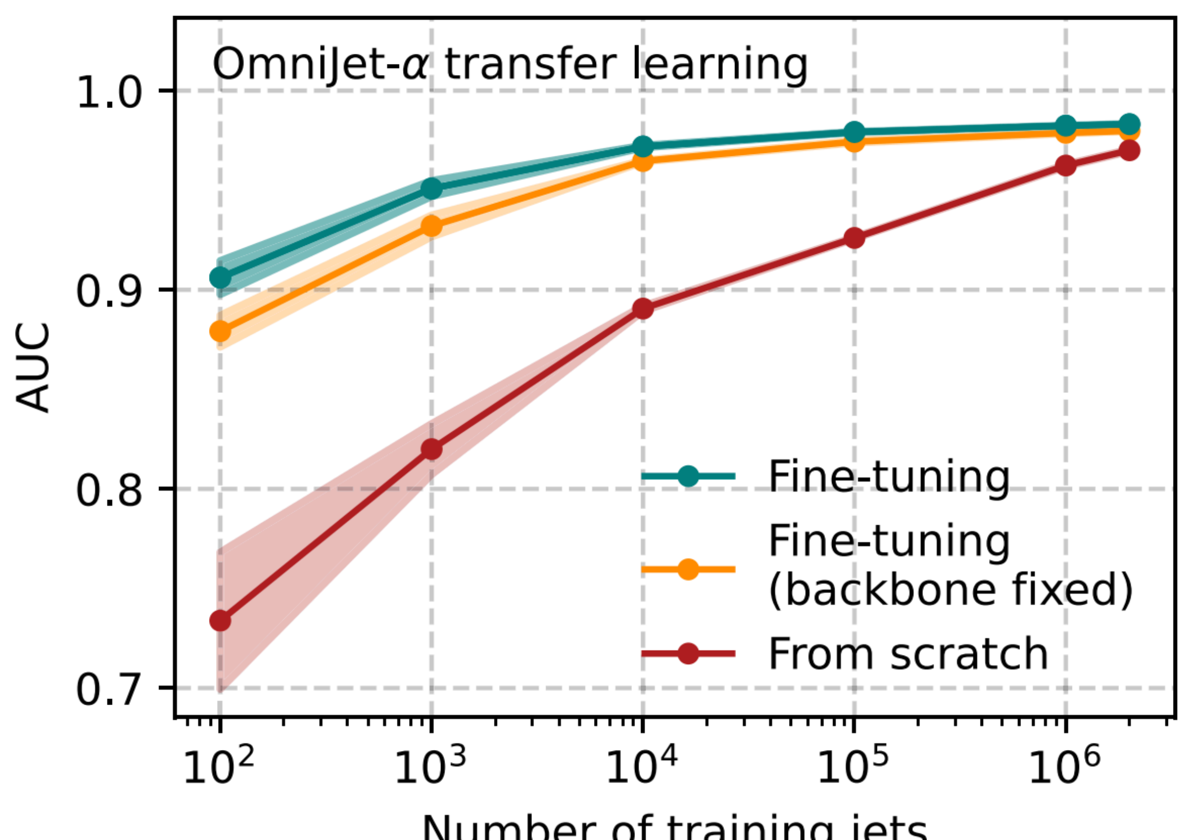
$$\text{Jet} = \{\texttt{start-token}, \texttt{token}_1, \ldots, \texttt{token}_n, \texttt{end-token}\}$$
$$\texttt{token}_i = \text{integer value} \in [1, \ldots, 8192]$$

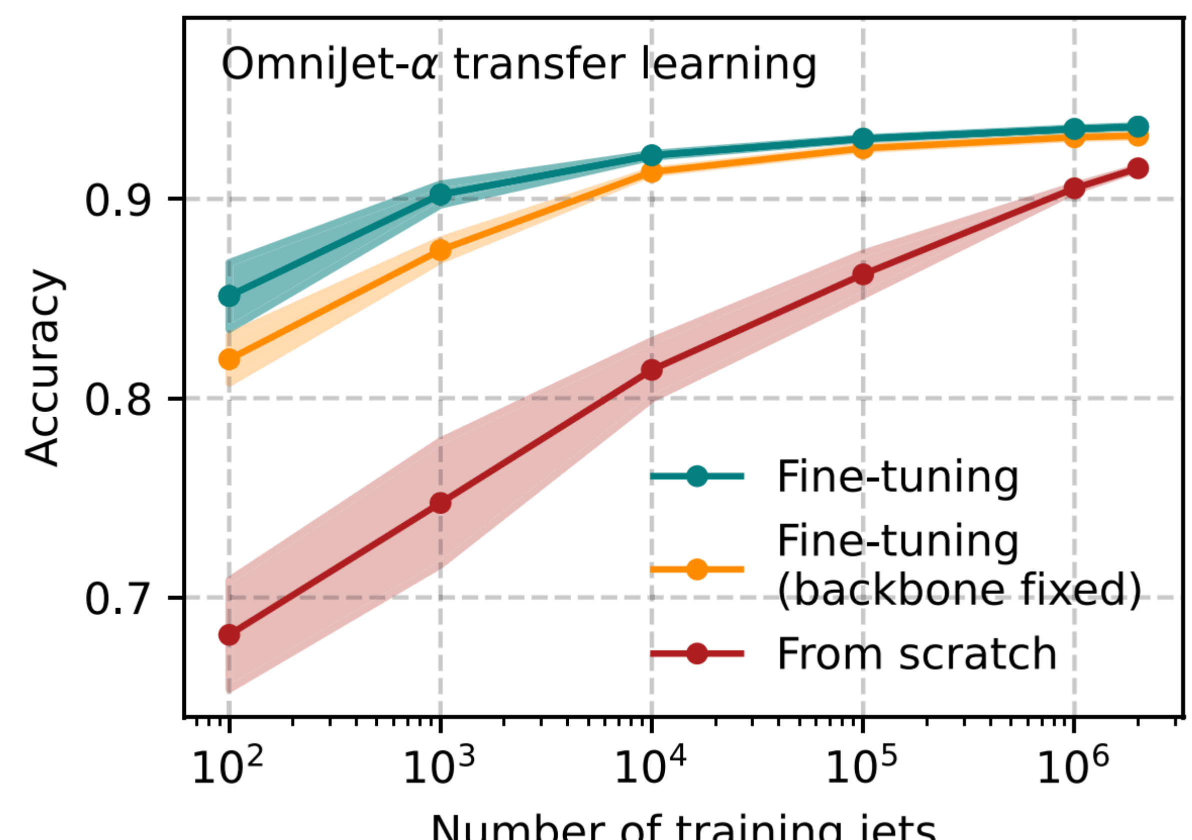
Self-supervised pre-training of transformer backbone on generative task (next token prediction)



Fine-tuning to classification task:
Swap model head and use pre-trained backbone

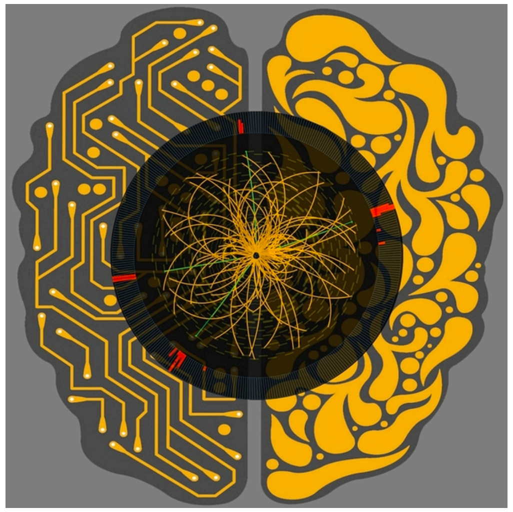






Let's have more fun with ML + Jets

Image: FermiLab

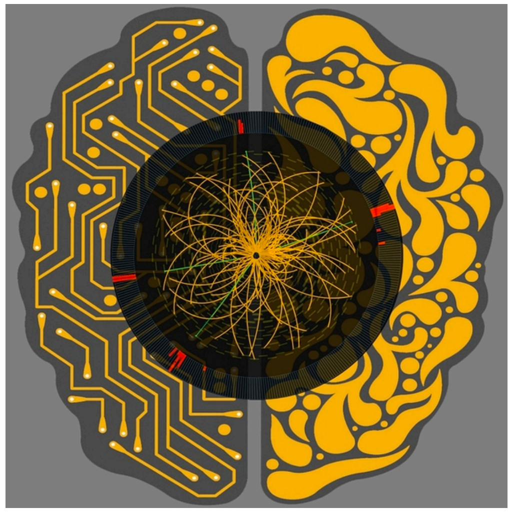


- ☑ML is here to stay with HEP.
- ☑When looked through the lens of ML, what's are the core questions to answer for these soft signals?
- ☑Interpretability and uncertainty estimation is a corner stone which we should emphasize.
- ☑Future collider program will depend heavily on all the ML methods we discussed + future R&D.
- ☑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 have more fun with ML + Jets

Image: FermiLab

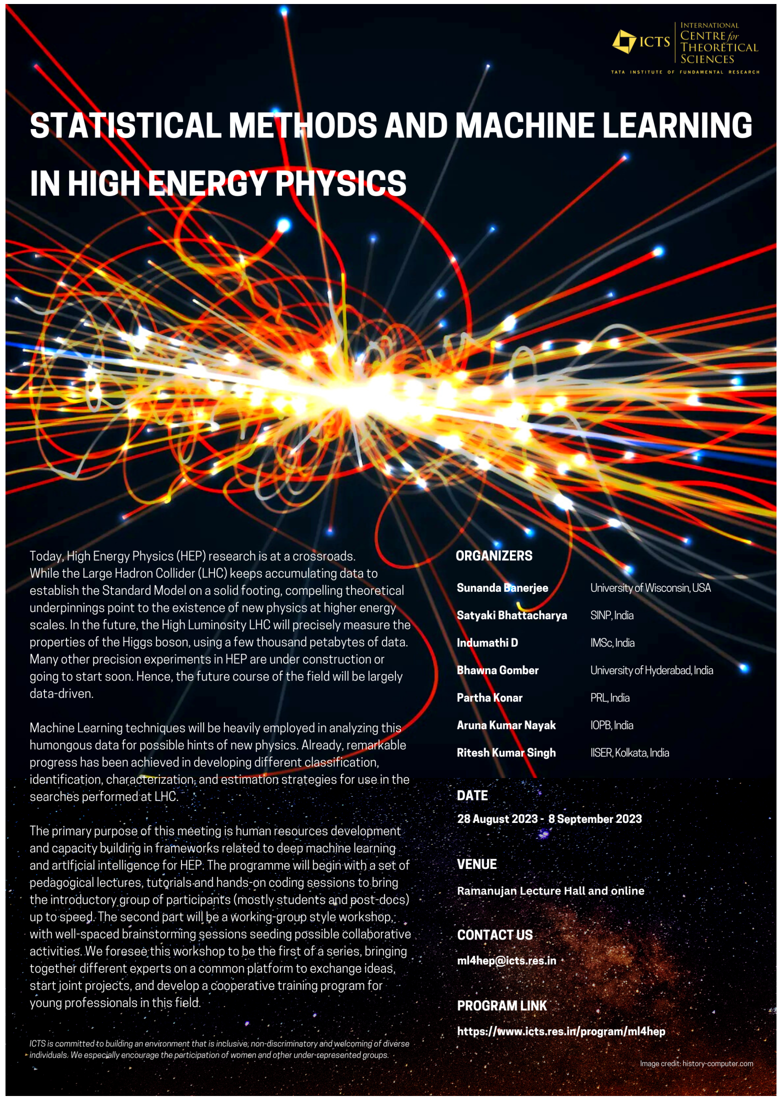


- ☑ ML is here to stay with HEP.
- ☑ When looked through the lens of ML, what’s are the core questions to answer for these soft signals?
- ☑ Interpretability and uncertainty estimation is a corner stone which we should emphasize.
- ☑ Future collider program will depend heavily on all the ML methods we discussed + future R&D.
- ☑ 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

The annual HEP-ML schools



This year the venue is TIFR, Mumbai