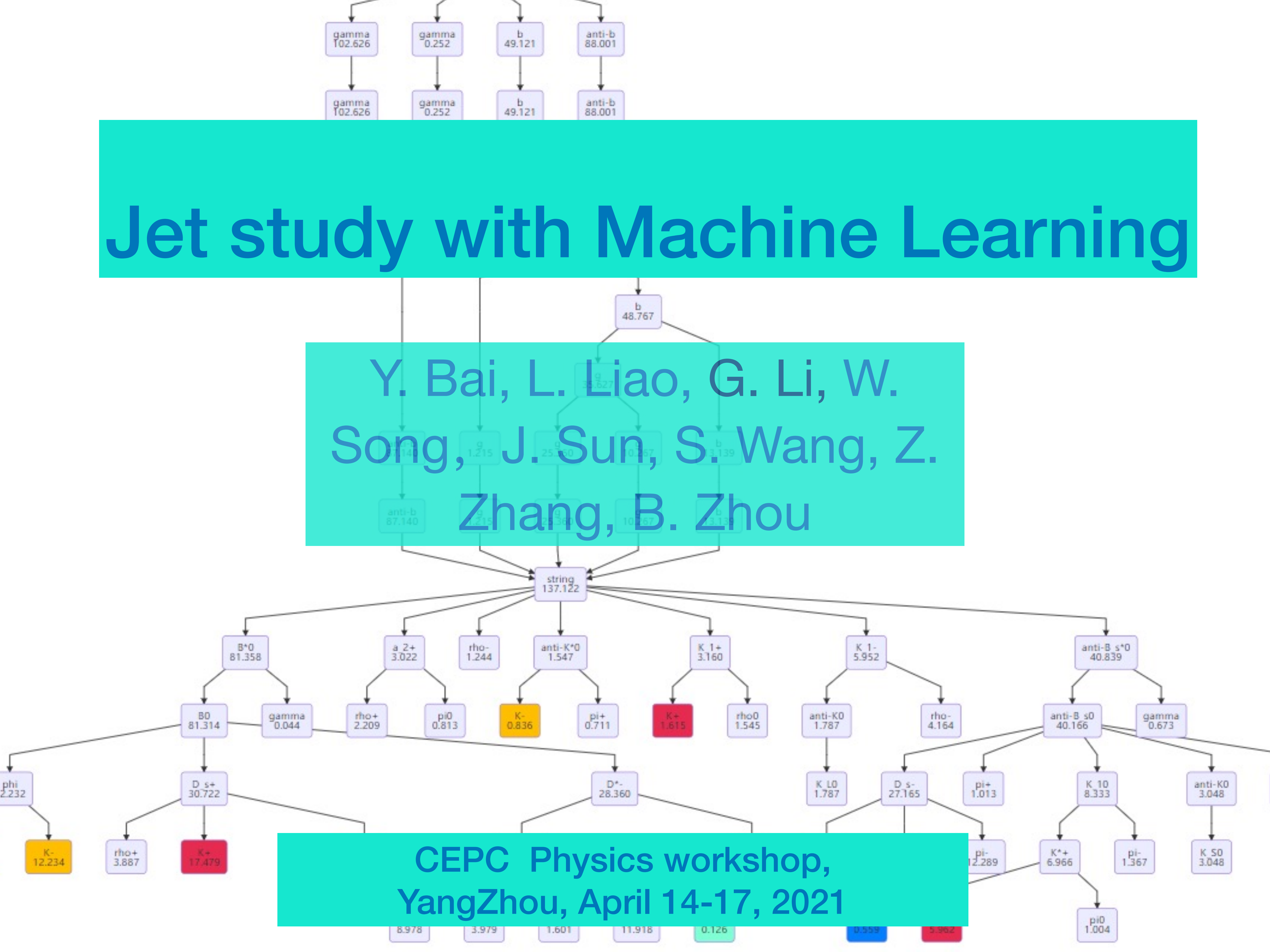
Jet study with Machine Learning

Y. Bai, L. Liao, G. Li, W. Song, J. Sun, S. Wang, Z. Zhang, B. Zhou

CEPC Physics workshop,
YangZhou, April 14-17, 2021

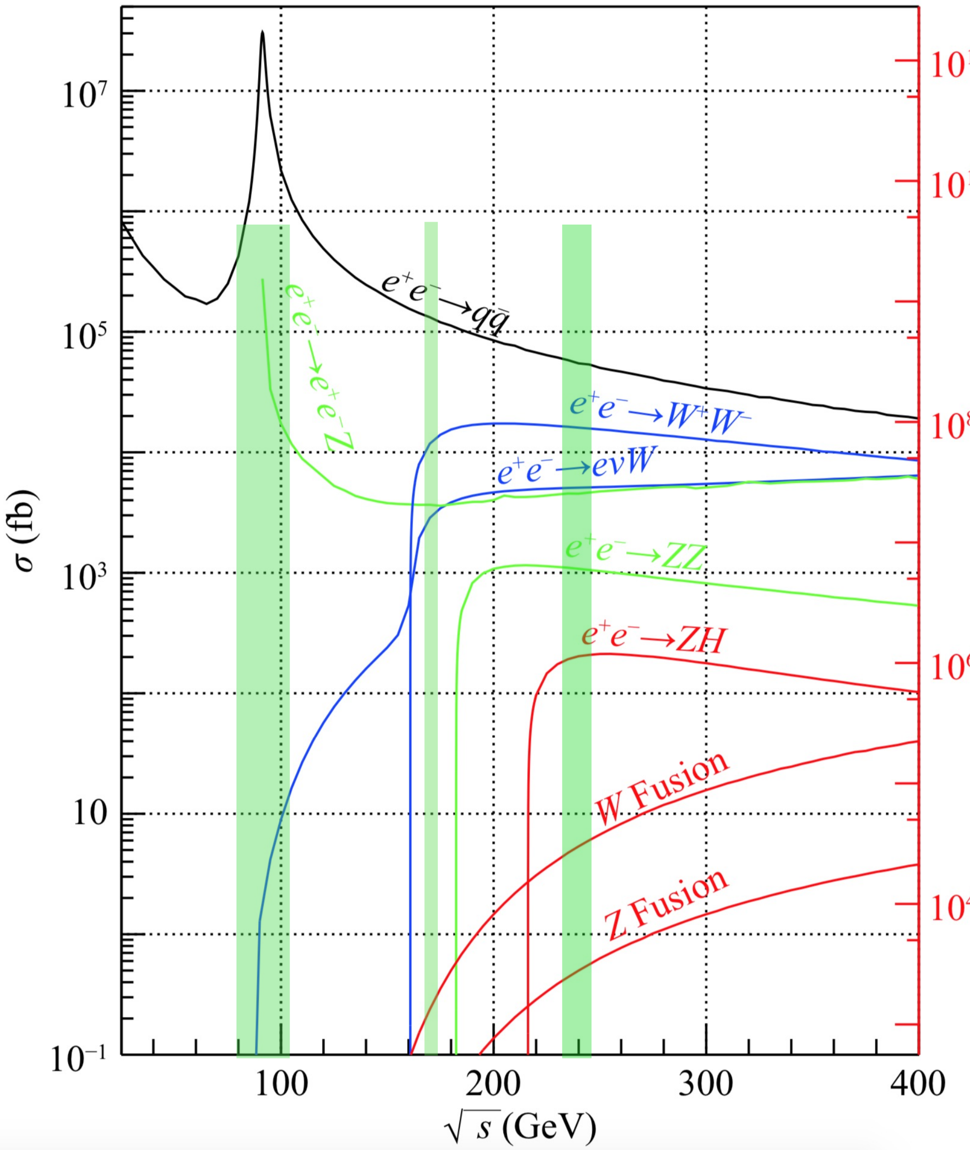


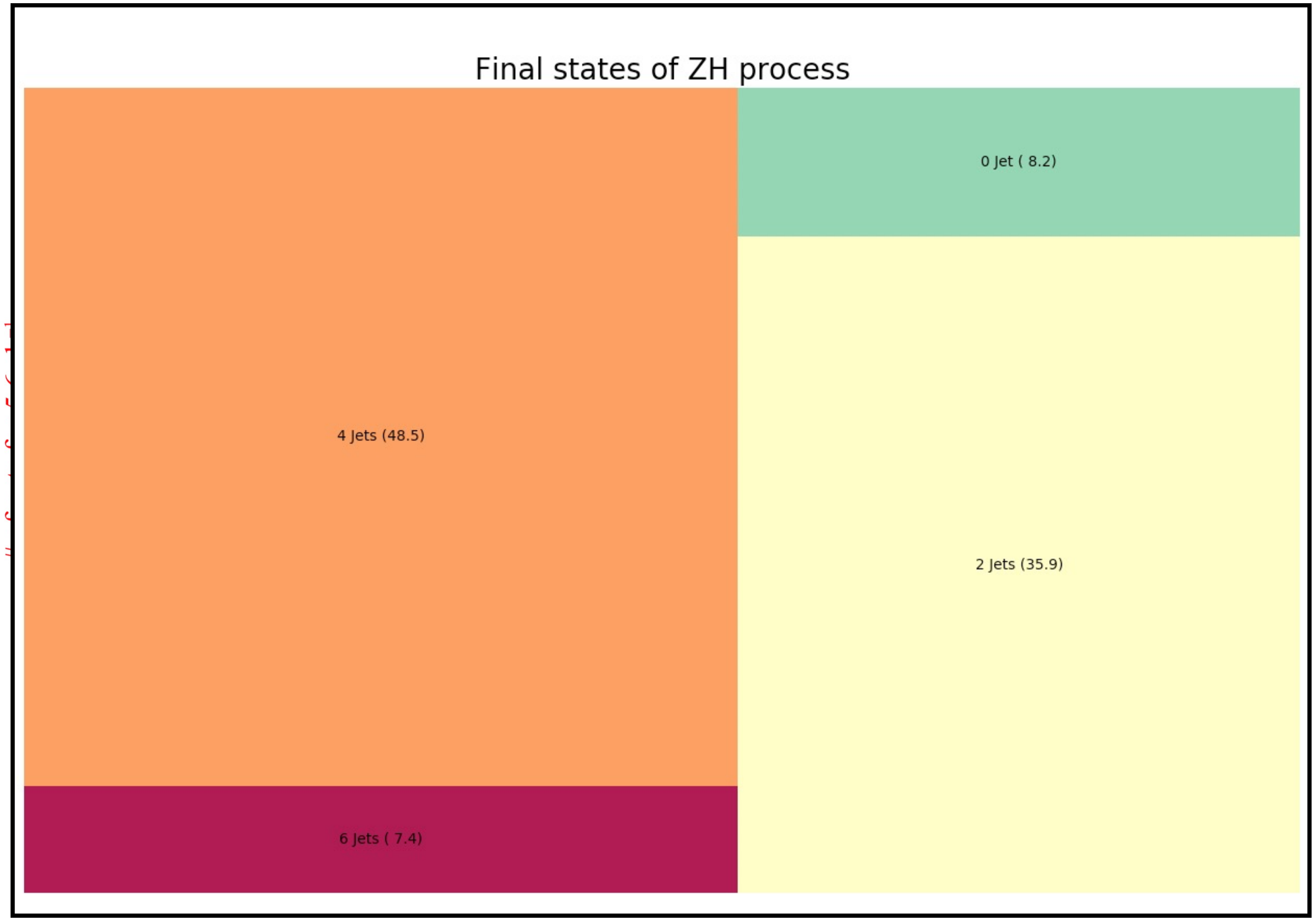
Outline

- Introduction
- Inputs and setup
- Preliminary results
- Summary and plan

Jet performance is curial for physics study at CEPC

CEPC: 1M Higgs, ~10^{12} Z, and 10^{8} W Pairs





- ☑ H/W/Z boson ~ 70% hadronic decay Br's
- ☑ ZH events > 90% final states contains jets

Topics of jet studies

- Flavor tagging
- Gluon identification
- Jet charge of heavy flavor
- Boson mass regression

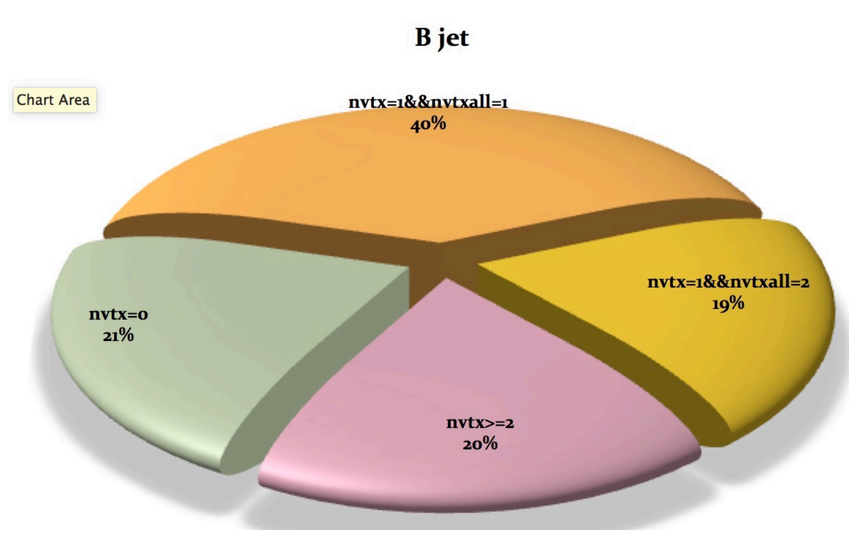
In a PFA detector, the complete information is a bunch of particles

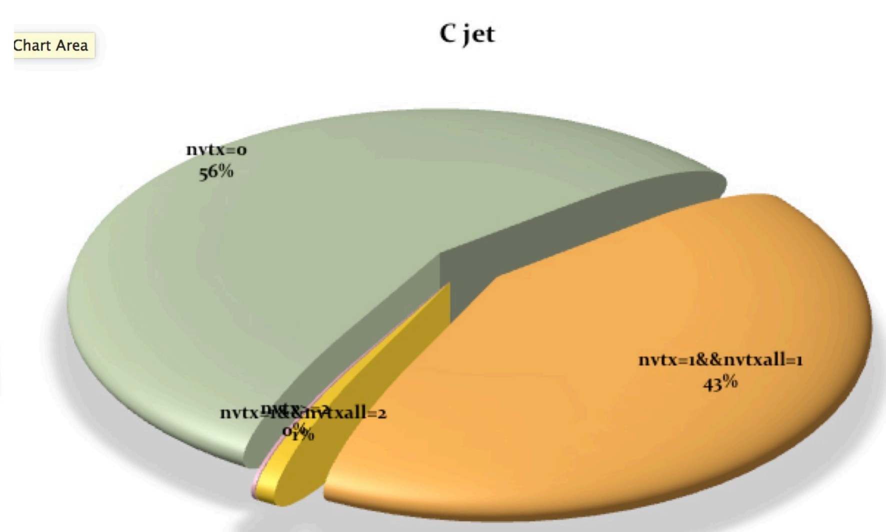
- **4 momenta**
- **Particle type**
- **Charge**
- **Impact parameters for tracks**

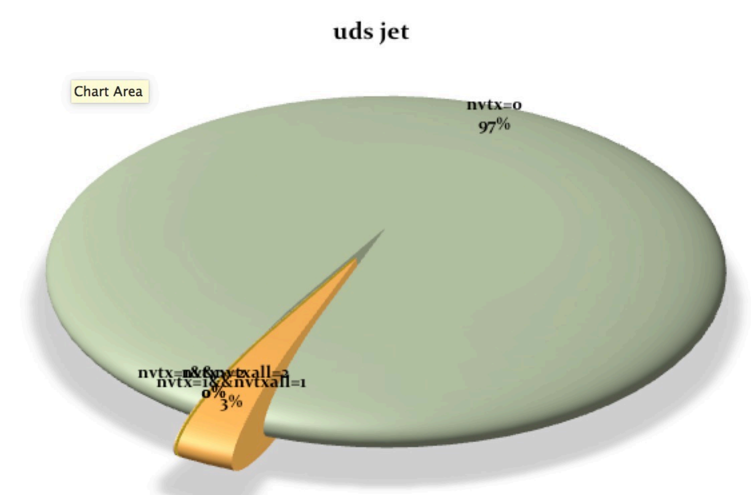
- ☑ **Conventional TMVA uses hand-eneneering high level features**
- ☑ **Here we will use “low” level information to realize so-called “end-to-end” machine learning**

CEPC jet tagging make use of high level features after vertex finding

	Total	nvtx==0	nvtx=1&& Nvtxall==1	nvtx==1&&nvtxall==2	Nvtx>=2
B	400 000	83 099	156 094	76 239	80 135
C	400 000	223 238	169 400	3 392	662
uds	400 000	382 522	10 511	171	106







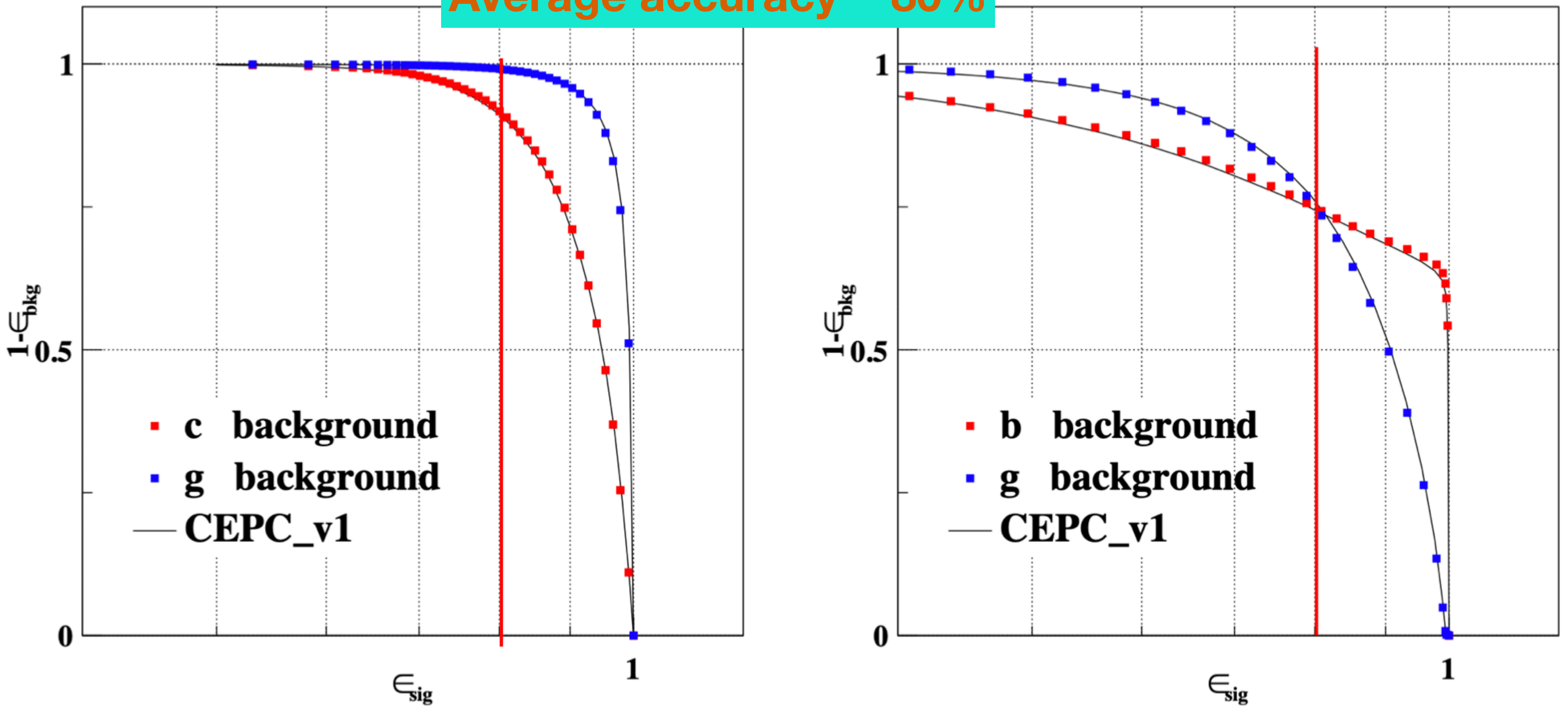
Input variables of each category

nvtx=0	trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr5sigma jprobz5sigma d0bprob d0cprob d0qprob z0bprob z0cprob z0qprob nmuon nelectron trkmass(17)
nvtx=1&&nvtxall=1	trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob d0bprob d0cprob d0qprob z0bprob z0cprob z0qprob trkmass nelectron nmuon(25)
nvtx=1&&nvtxall=2	trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob 1vtxprob vtxlen12all_jete vtxmassall (19)
Nvtx>=2	trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob vtxlen2_jete vtxsig2_jete vtxdirang2_jete vtxmom2_jete vtxmass2 vtxmult2 vtxlen12_jete vtxsig12_jete vtxdirang12_jete vtxmom_jete vtxmass vtxmult 1vtxprob(29)

Introduction — Jet tagging in CDR

DNN, xgboost, … similar performance

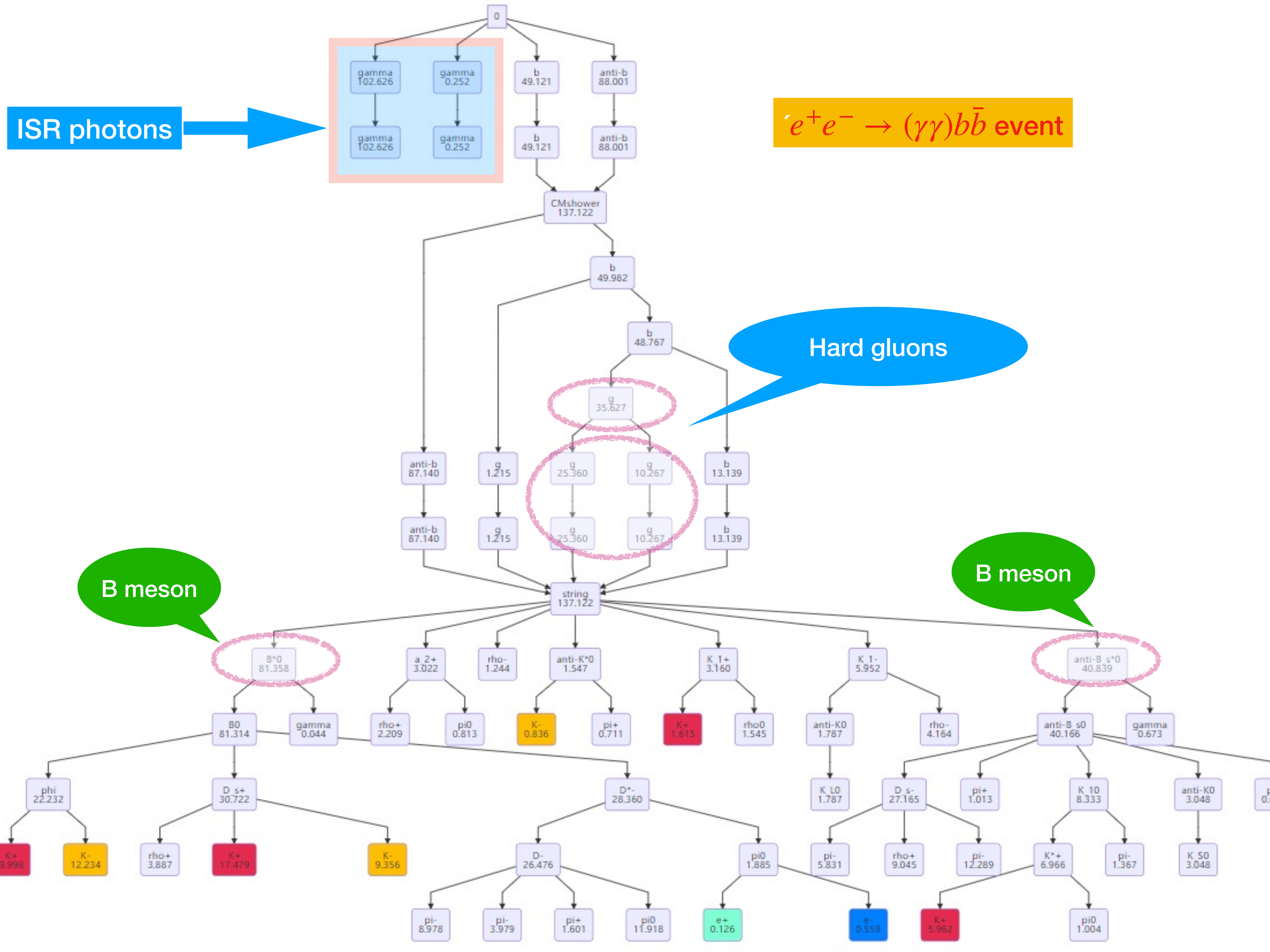
Average accuracy ~ 80%



Receiver Operating Characteristic Curve (ROC)

80% b-tagging eff. : Reject 90% c and 99% o jets

80% c-tagging eff. : Reject 75% b and 75% o jets

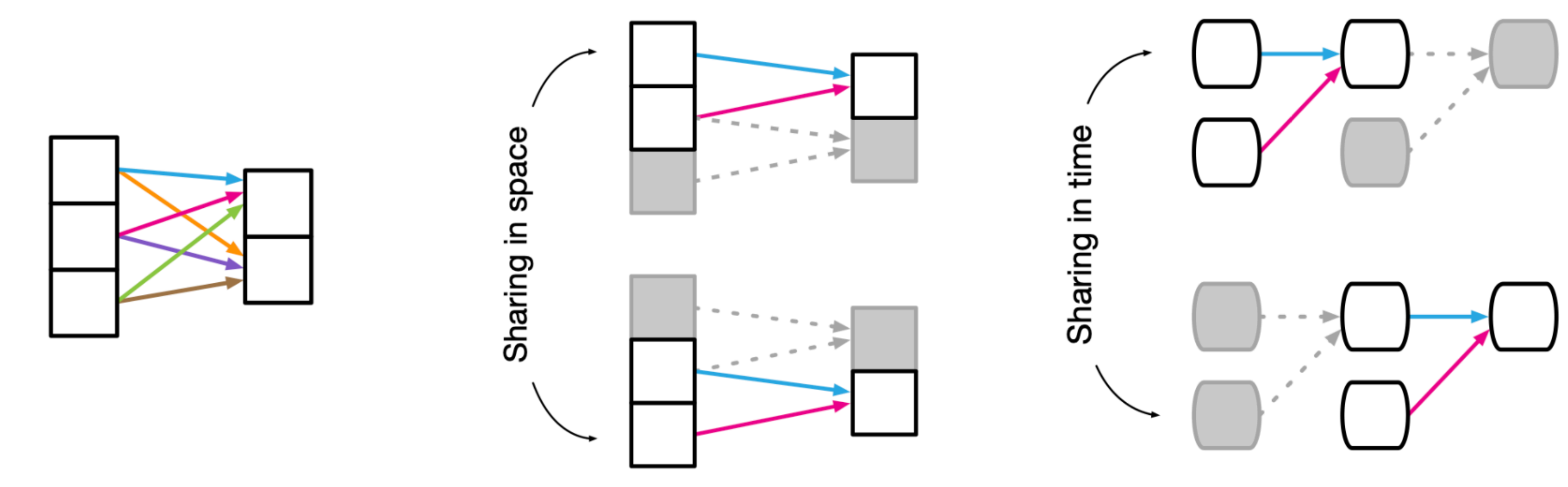


ISR photons
$e^+e^- \to (\gamma\gamma)b\bar{b}$ event
Hard gluons
B meson
B meson
0
gamma 102.626
gamma 0.252
b 49.121
anti-b 88.001
gamma 102.626
gamma 0.252
b 49.121
anti-b 88.001
CMshower 137.122
b 49.982
b 48.767
g 35.627
anti-b 87.140
g 1.215
g 25.360
g 10.267
b 13.139
anti-b 87.140
g 1.215
g 25.360
g 10.267
b 13.139
string 137.122
B*0 81.358
a_2+ 3.022
rho- 1.244
anti-K*0 1.547
K_1+ 3.160
K_1- 5.952
anti-B_s*0 40.839
B0 81.314
gamma 0.044
rho+ 2.209
pi0 0.813
K- 0.836
pi+ 0.711
K+ 1.615
rho0 1.545
anti-K0 1.787
rho- 4.164
anti-B_s0 40.166
gamma 0.673
phi 22.232
D_s+ 30.722
D*- 28.360
K_L0 1.787
D_s- 27.165
pi+ 1.013
K_10 8.333
anti-K0 3.048
K+ 9.998
K- 12.234
rho+ 3.887
K+ 17.479
K- 9.356
D- 26.476
pi0 1.885
pi- 5.831
rho+ 9.045
pi- 12.289
K*+ 6.966
pi- 1.367
K_S0 3.048
pi- 8.978
pi- 3.979
pi+ 1.601
pi0 11.918
e+ 0.126
e- 0.559
K+ 5.962
pi0 1.004

Comments for FN, CNN, RNN, and GN

arXiv:1806.01261v3

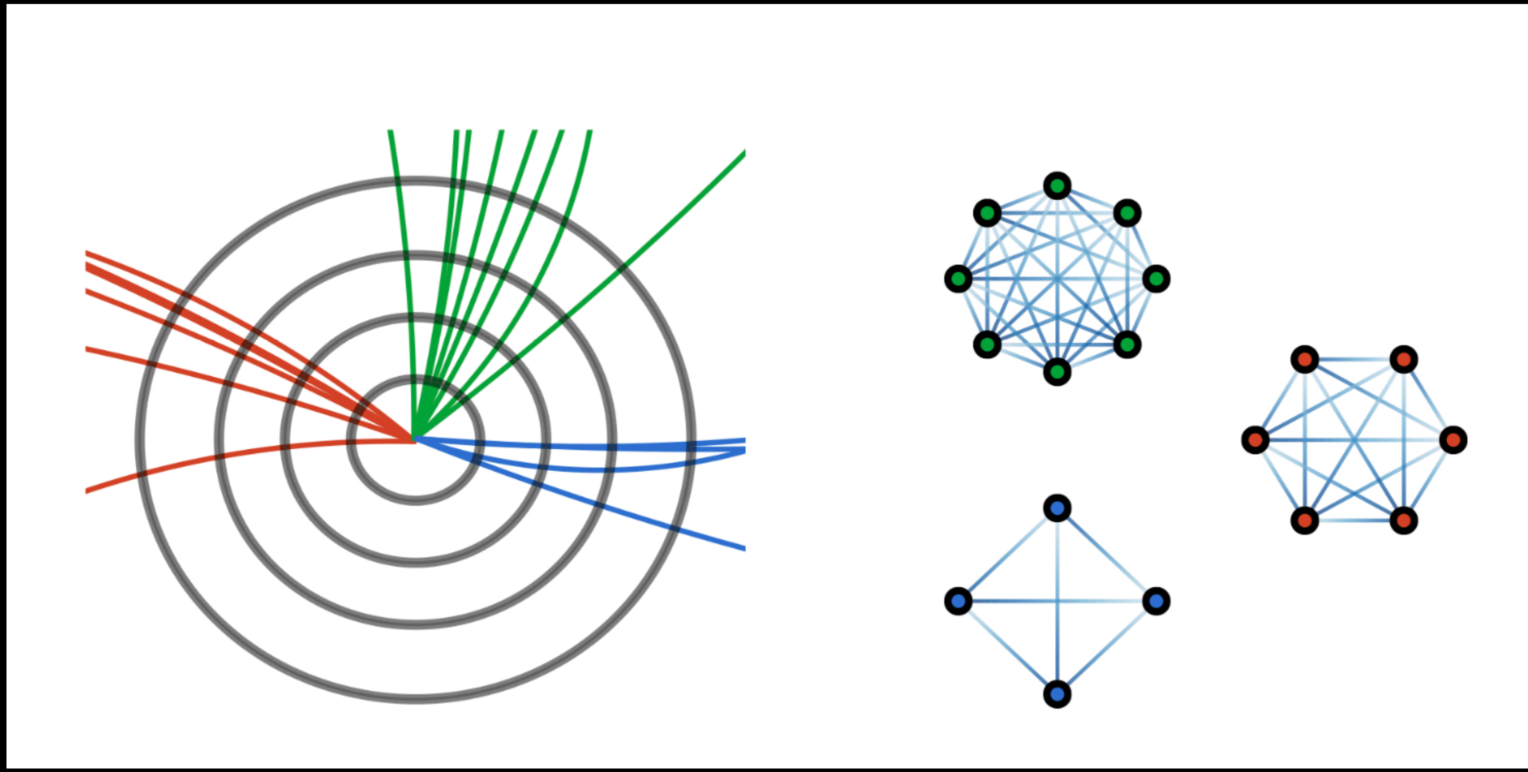
- FN: no reuse, no isolation of information
- CNN: locality and translation invariance, very effective for processing natural image data. Jet doesn't satisfy translation invariance
- RNN: The rule reused over each step, temporal invariance (similar to a CNN's translational invariance in space)
- GN： a natural representation for systems described entities whose order is undefined or irrelevant; in particular, their relational inductive bias does not come from the presence of something, but rather from the absence



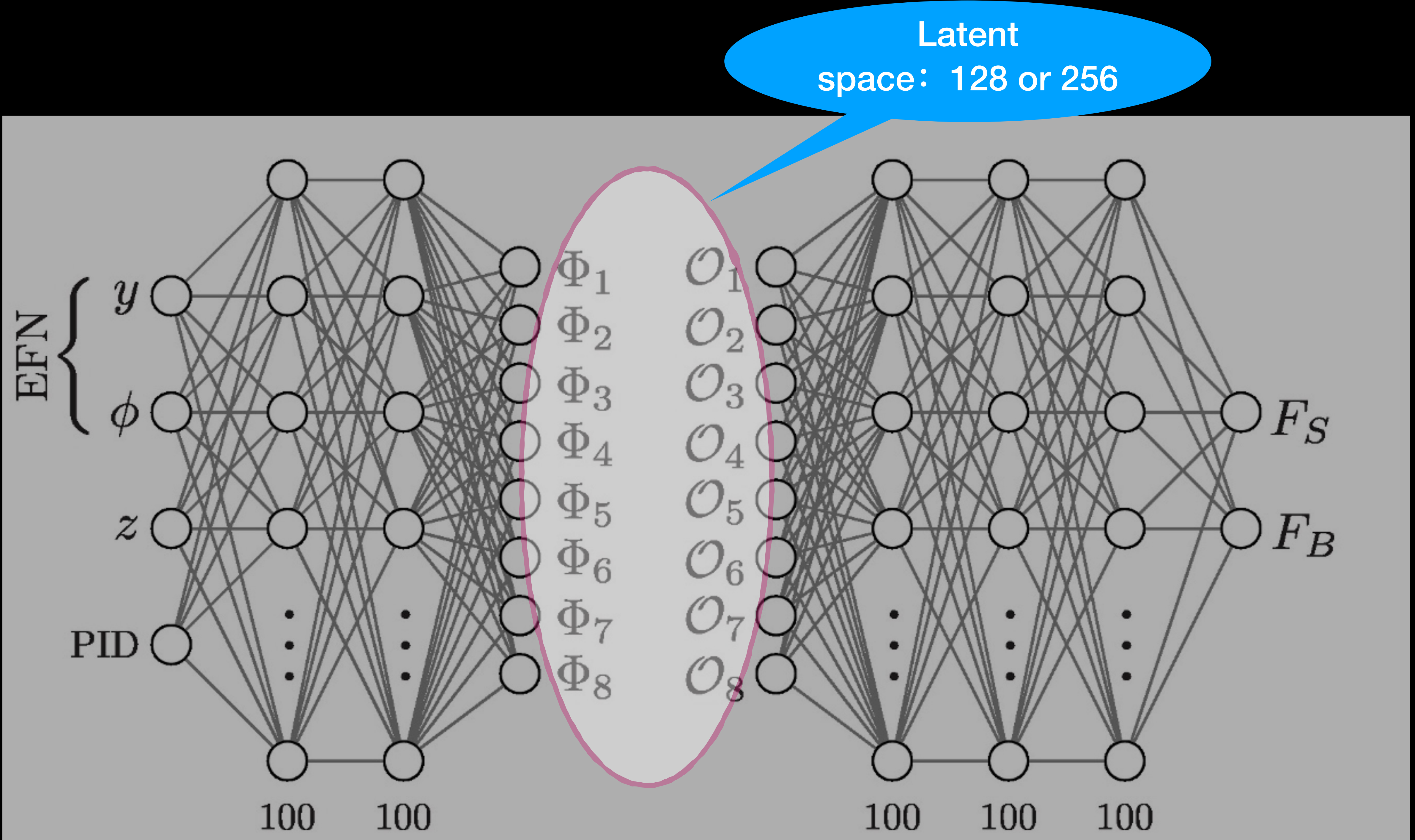
(a) Fully connected (b) Convolutional (c) Recurrent

Graph Network (GN)

- No algorithm works generally good everywhere
- Jet related studies need some algorithms with strong relational inductive bias in this field
- Graph
  - Nodes : momenta, impact parameters, PID, charge, …
  - Edges : invariant mass, angles, …
  - Graph level attributes



Architecture of network



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

Collider observables decomposed into per-particle maps Φ and functions F

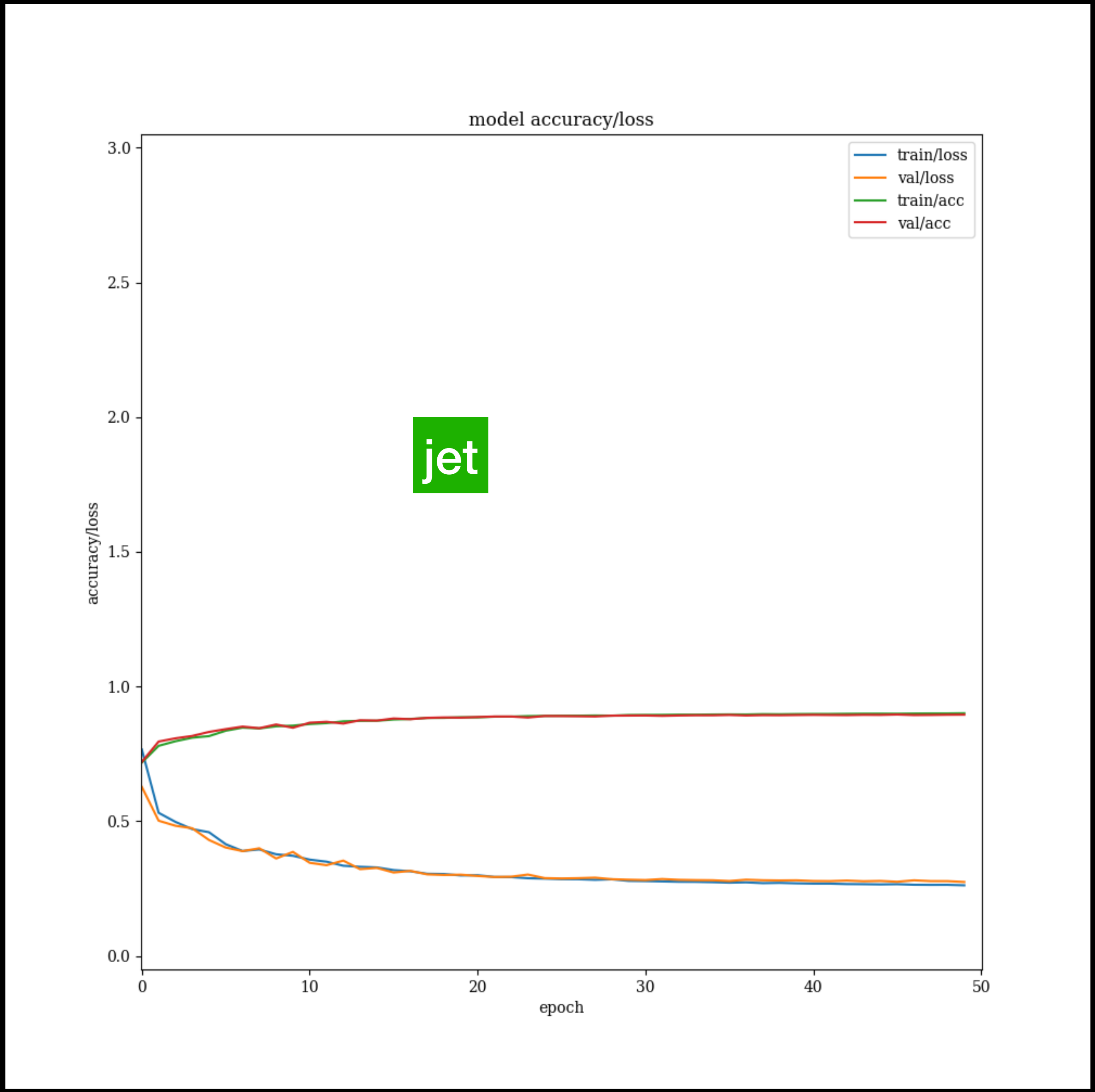
Observable $\mathcal{O}$		Map Φ	Function F
Mass	m	p^μ	$F(x^\mu) = \sqrt{x^\mu x_\mu}$
Multiplicity	M	1	$F(x) = x$
Track Mass	m_{track}	$p^\mu \mathbb{I}_{\text{track}}$	$F(x^\mu) = \sqrt{x^\mu x_\mu}$
Track Multiplicity	M_{track}	$\mathbb{I}_{\text{track}}$	$F(x) = x$
Jet Charge [72]	$\mathcal{Q}_\kappa$	$(p_T, Q\, p_T^\kappa)$	$F(x, y) = y/x^\kappa$
Eventropy [74]	$z \ln z$	$(p_T, p_T \ln p_T)$	$F(x, y) = y/x - \ln x$
Momentum Dispersion [93]	p_T^D	(p_T, p_T^2)	$F(x, y) = \sqrt{y/x^2}$
C parameter [94]	C	$(\lvert\vec{p}\rvert, \vec{p} \otimes \vec{p}/\lvert\vec{p}\rvert)$	$F(x, Y) = \frac{3}{2x^2}[(\mathrm{Tr}\, Y)^2 - \mathrm{Tr}\, Y^2]$

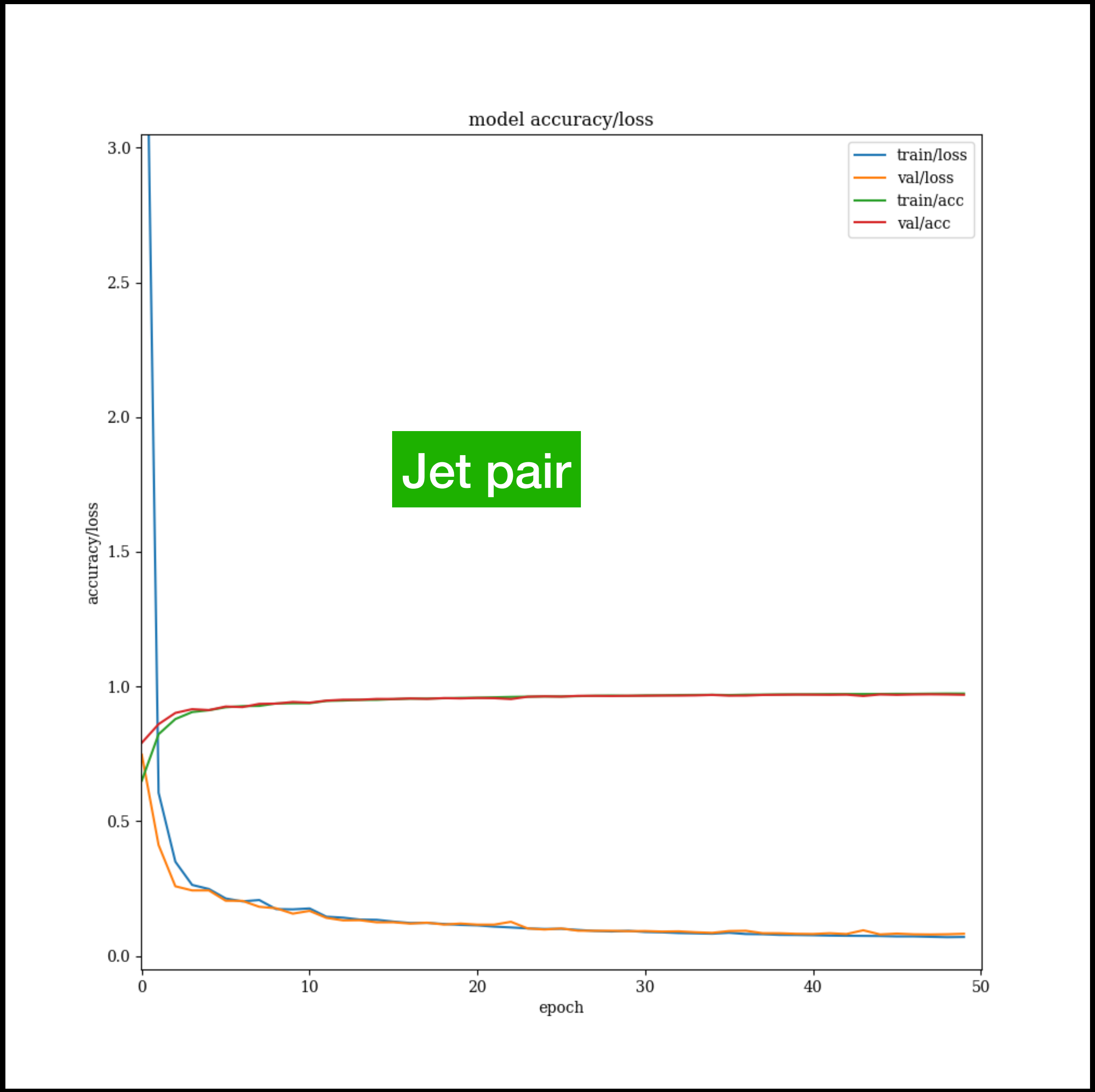
Inputs and setup

- Input:
  - 300 k Z—> bb, cc, and other jets events
  - 100 k vvH —> bb, cc, gg event
- Same fast simulation configuration
- Using fastjet/ee-kt algorithm to force all particles to 2 jets
- Train: validation: test = 8:1:1

Results

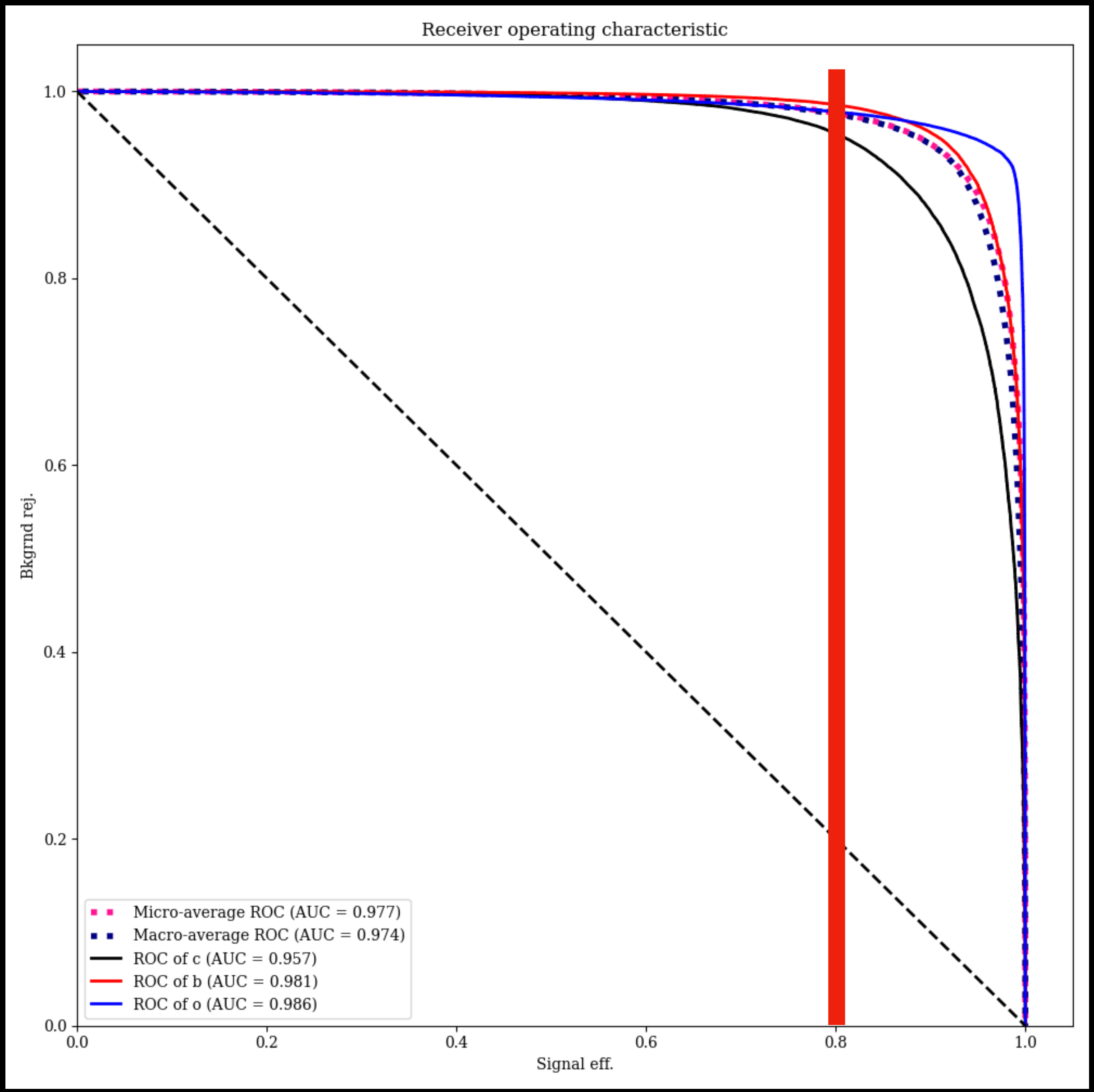
Training process
Loss and accuracy

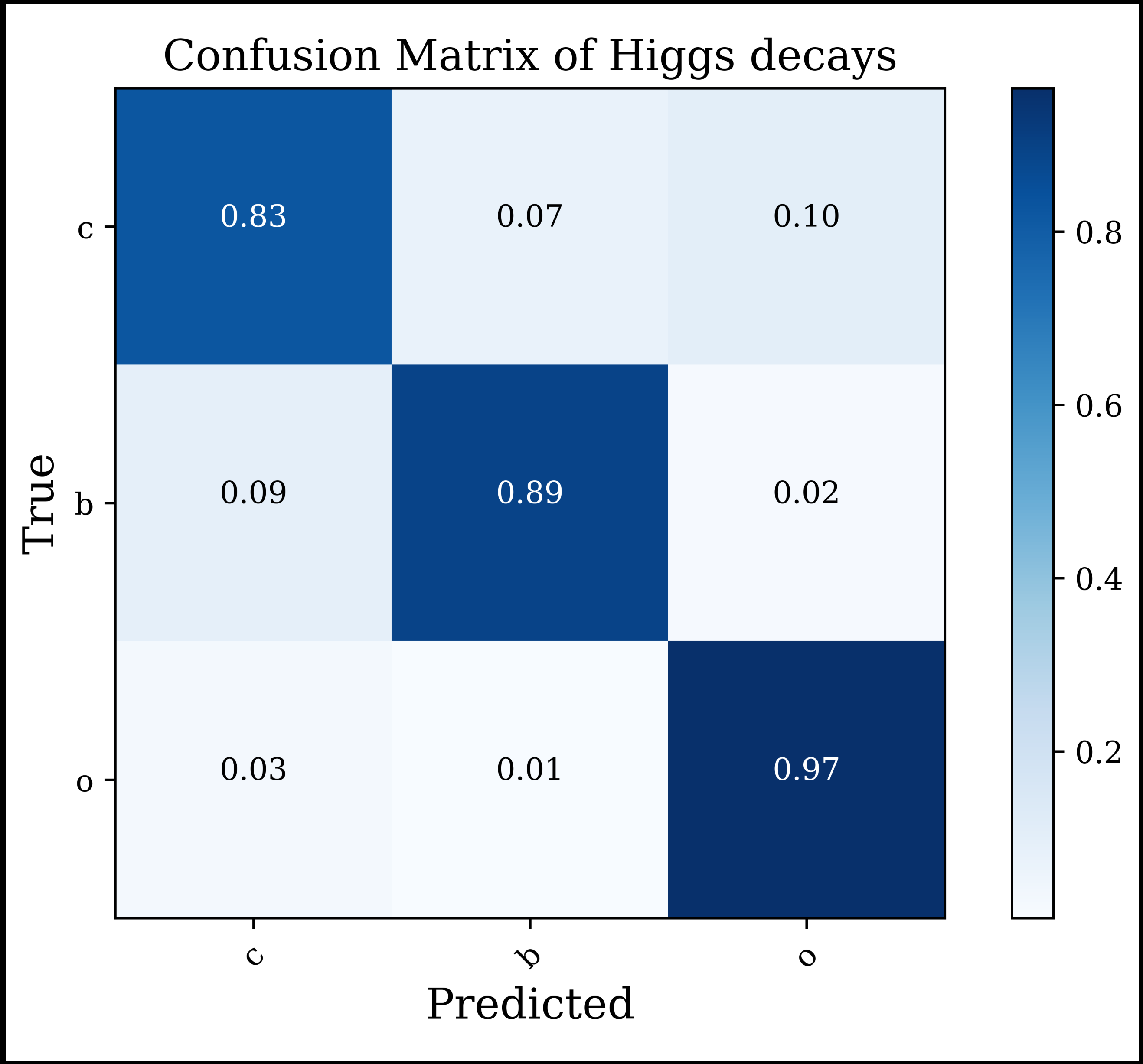




Single jet tagging

Averaged accuracy is 89%, CDR: 80%

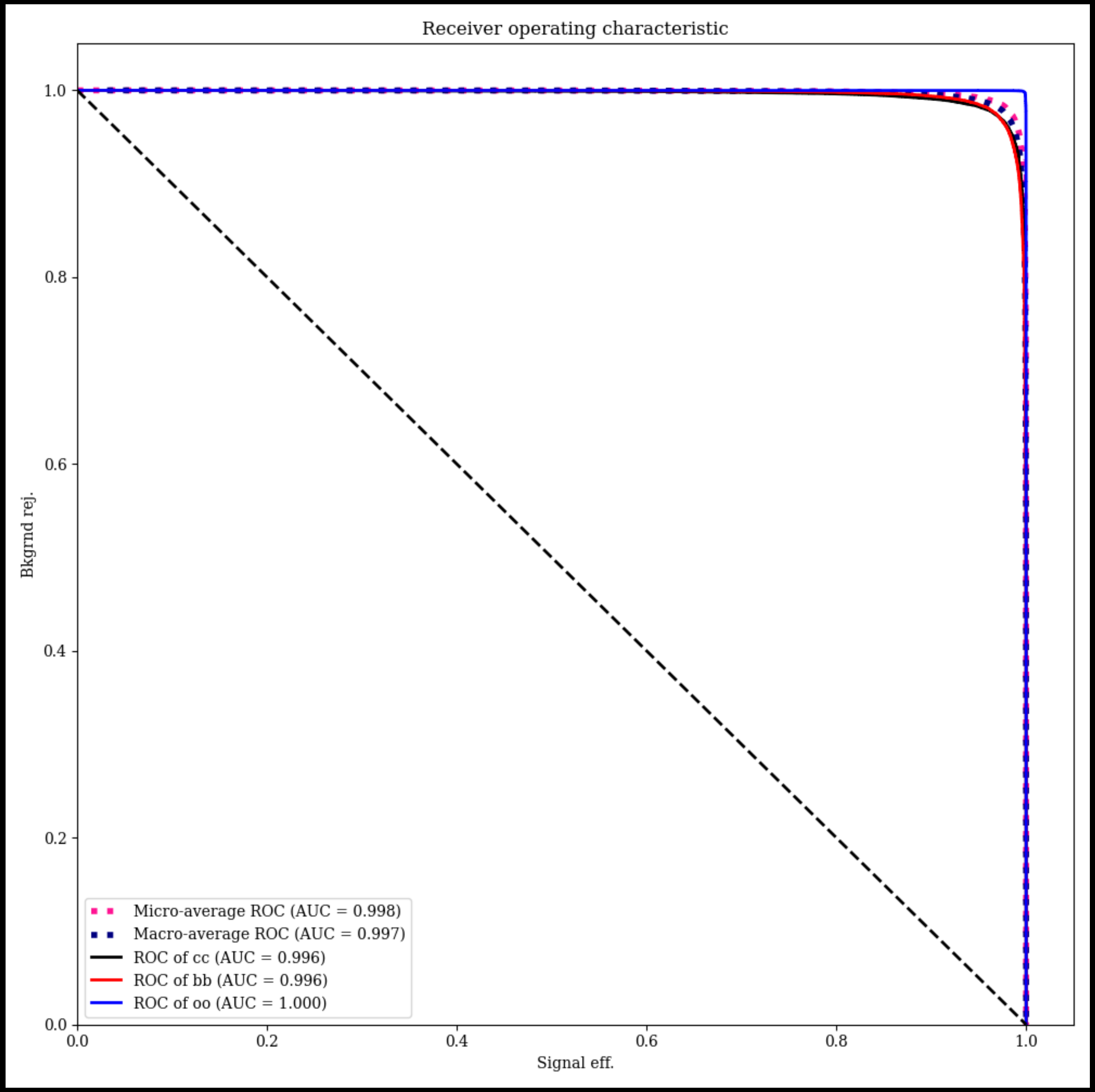


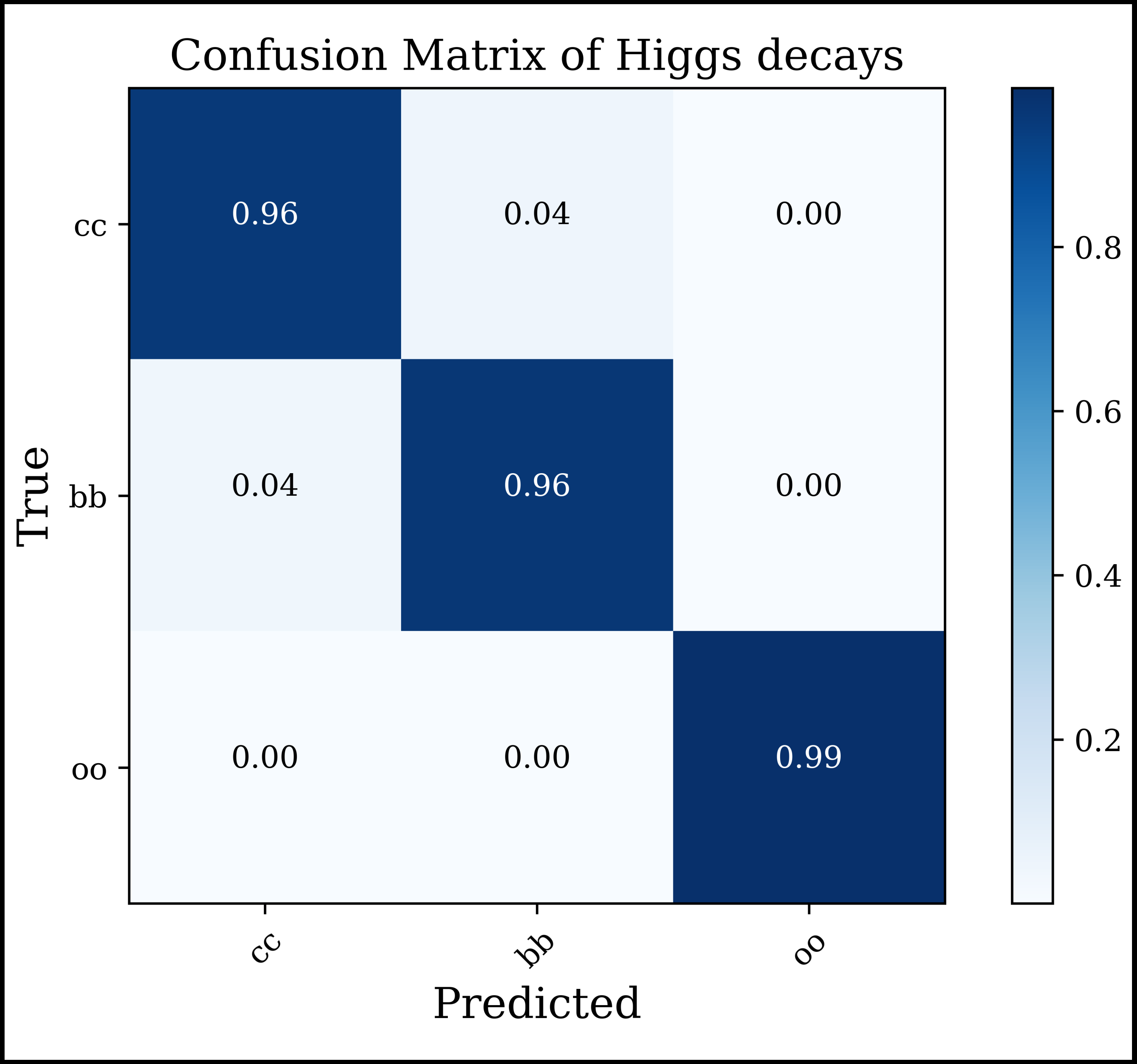


c-tagging rejection power still > 90% for 80% efficiency

Jet pair tagging

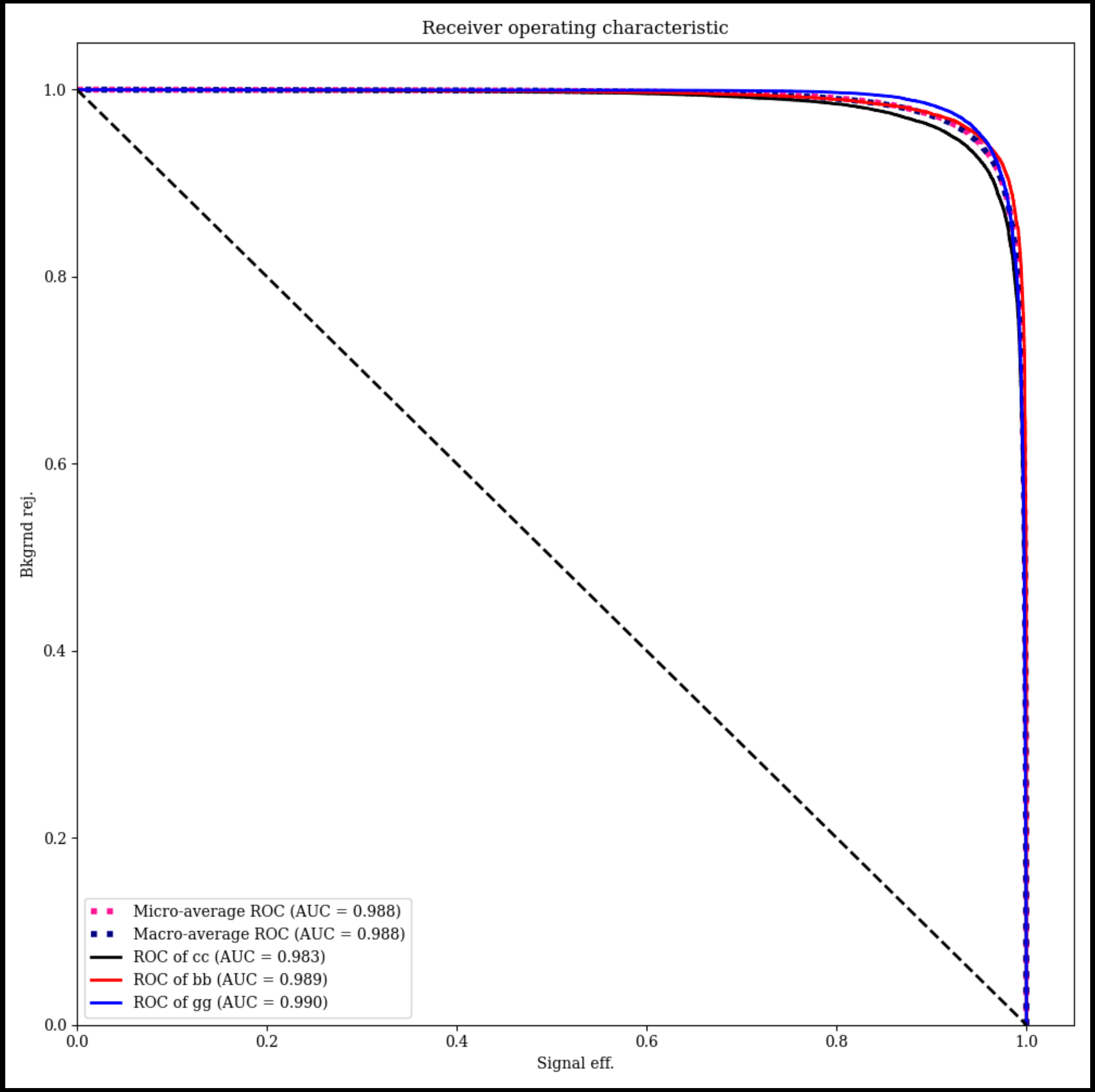
Averaged accuracy is 97%, good news for Rb & Rc measurement

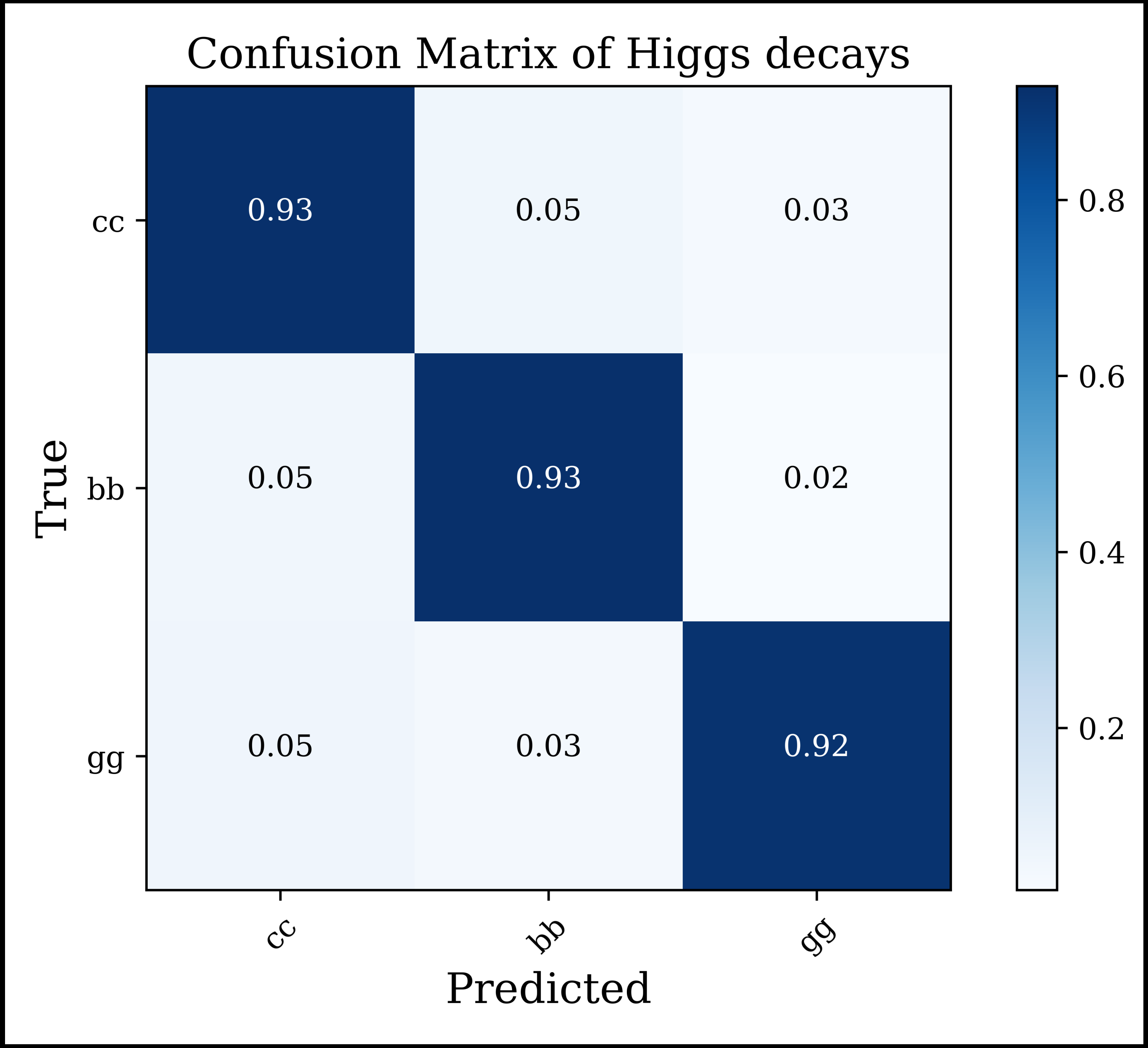




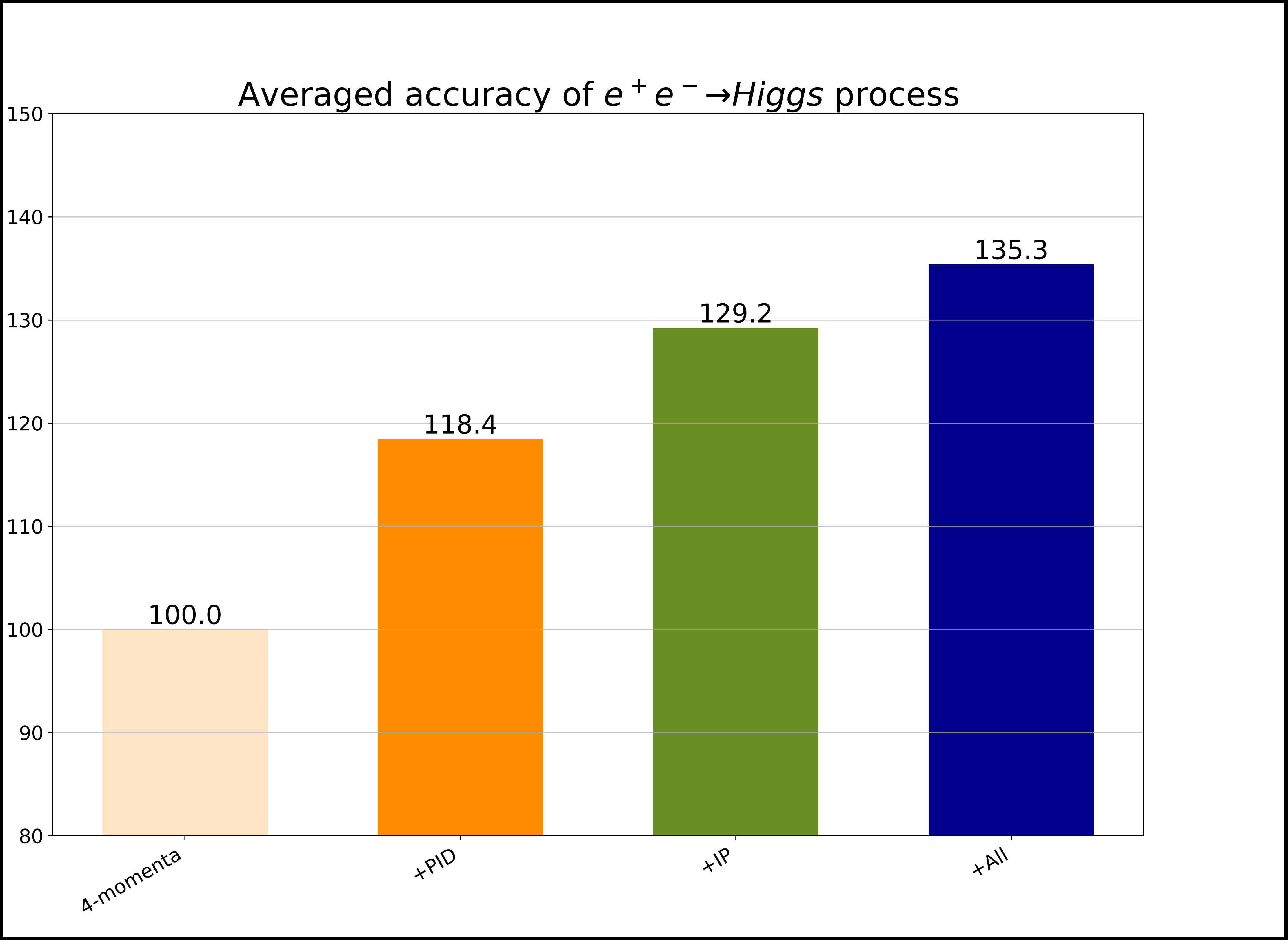
Gluon pair tagging

Averaged accuracy is 92.5%





Performance dependence on E/p/PID/impact parameters



Summary and plan

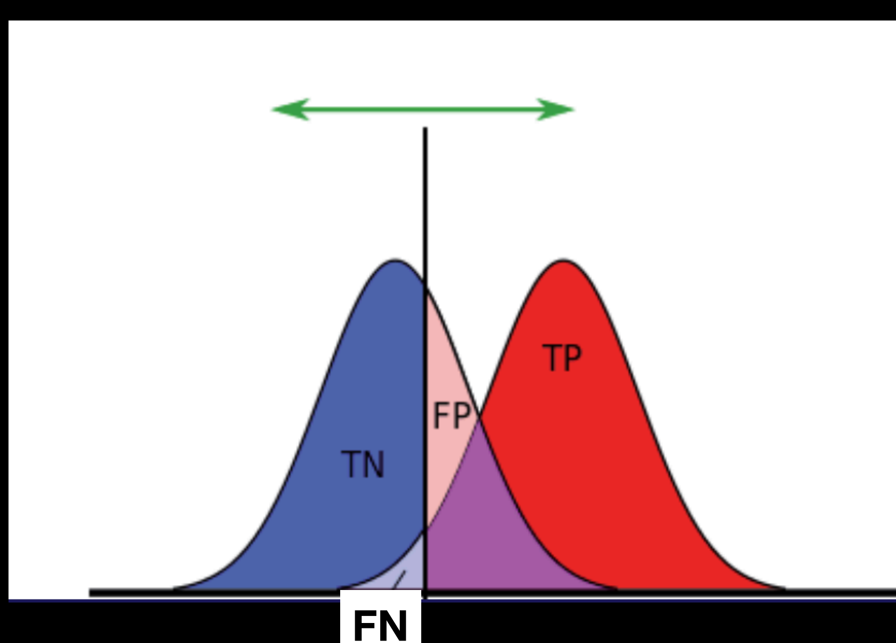
- ML based on graph representation shows very promising performance for jet study @ fast sim level
- Jet charge and boson (jet-pair) mass
- Try other graph network method, such as DGCNN
- More study to make the “black box” transparency
- Move to full simulation and apply to analysis
- And detector optimization …

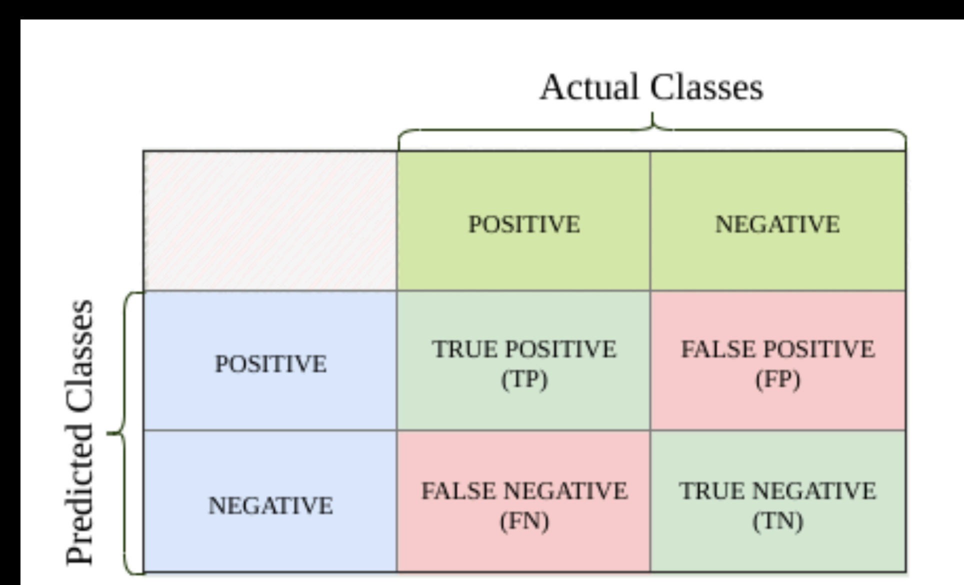
The end

Thanks a lot

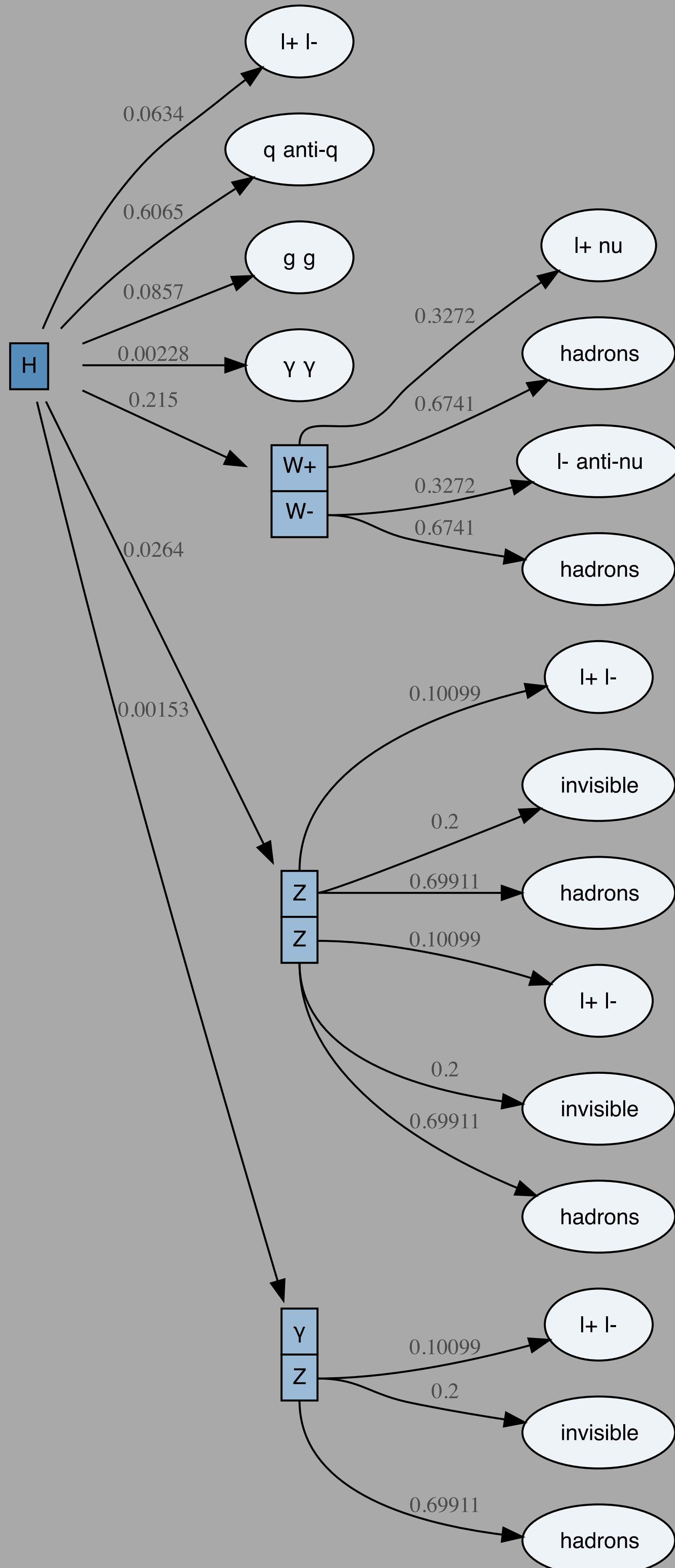
Termnologies

- Tagging efficiency: accuracy in ML
- ROC : Receiver Operating Characteristics Curve, mainly for binary classification,
  - In HEP it is Rejection rate vs. Tagging efficiency (FN rate vs. TP rate)
- AUC : Area under the ROC
- Confusion matrix
  - it is the efficiency matrix when neglecting SM backgrounds





		Actual Classes	
		POSITIVE	NEGATIVE
Predicted Classes	POSITIVE	TRUE POSITIVE (TP)	FALSE POSITIVE (FP)
	NEGATIVE	FALSE NEGATIVE (FN)	TRUE NEGATIVE (TN)



H
l+ l-
0.0634
q anti-q
0.6065
g g
0.0857
γ γ
0.00228
0.215
W+
W-
l+ nu
0.3272
hadrons
0.6741
l- anti-nu
0.3272
hadrons
0.6741
0.0264
Z
Z
l+ l-
0.10099
invisible
0.2
hadrons
0.69911
l+ l-
0.10099
invisible
0.2
hadrons
0.69911
0.00153
γ
Z
l+ l-
0.10099
invisible
0.2
hadrons
0.69911

DeepSets theorem

Deep Sets

[1703.06114]

Manzil Zaheer[1,2], Satwik Kottur[1], Siamak Ravanbhakhsh[1], Barnabás Póczos[1], Ruslan Salakhutdinov[1], Alexander J Smola[1,2]
[1] Carnegie Mellon University [2] Amazon Web Services

Feature space

Variable length

Deep Sets Theorem [63]. *Let* $\mathfrak{X} \subset \mathbb{R}^d$ *be compact,* $X \subset 2^{\mathfrak{X}}$ *be the space of sets with bounded cardinality of elements in* $\mathfrak{X}$*, and* $Y \subset \mathbb{R}$ *be a bounded interval. Consider a continuous function* $f : X \to Y$ *that is invariant under permutations of its inputs, i.e.* $f(x_1, \ldots, x_M) = f(x_{\pi(1)}, \ldots, x_{\pi(M)})$ *for all* $x_i \in \mathfrak{X}$ *and* $\pi \in S_M$*. Then there exists a sufficiently large integer* ℓ *and continuous functions* $\Phi : \mathfrak{X} \to \mathbb{R}^\ell$, $F : \mathbb{R}^\ell \to Y$ *such that the following holds to an arbitrarily good approximation:*[1]

Permutation invariance

Latent space

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

General parametrization for a function of sets