

Jet tagging and event classification at CEPC

张冰洋，王书栋，廖立波，张兆领，宋维民，张晋，李刚等

中国科学院高能物理研究所

2023.06.10-12

2023年粒子物理实验计算软件与技术研讨会

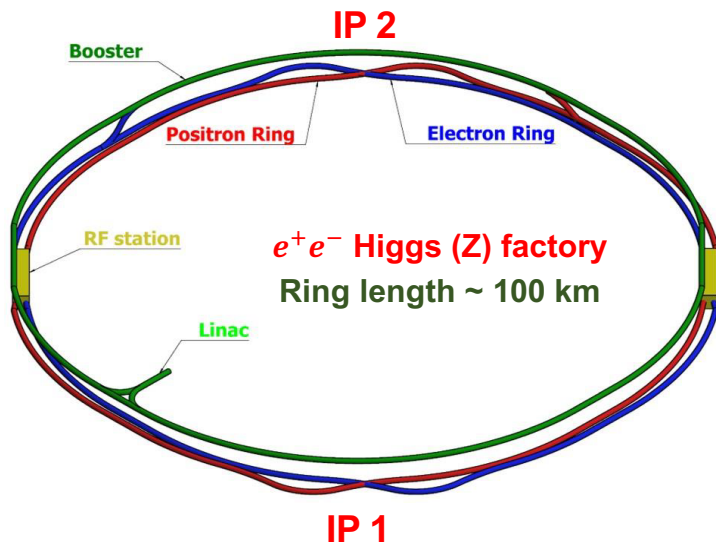


Outline

- Introduction
- Machine learning on jet tagging
- More than jet tagging: event classification
- Summary



- ❑ CEPC is an e^+e^- Higgs factory producing Higgs / W / Z bosons and top quarks, aims at discovering new physics beyond the Standard Model
- ❑ Proposed in 2012 right after the Higgs discovery
- ❑ Proposed to commence construction in ~ 2026 and start operation in 2030s.
- ❑ Upgrade: Super pp Collider (SppC) of $\sqrt{s} \sim 100$ TeV in the future.

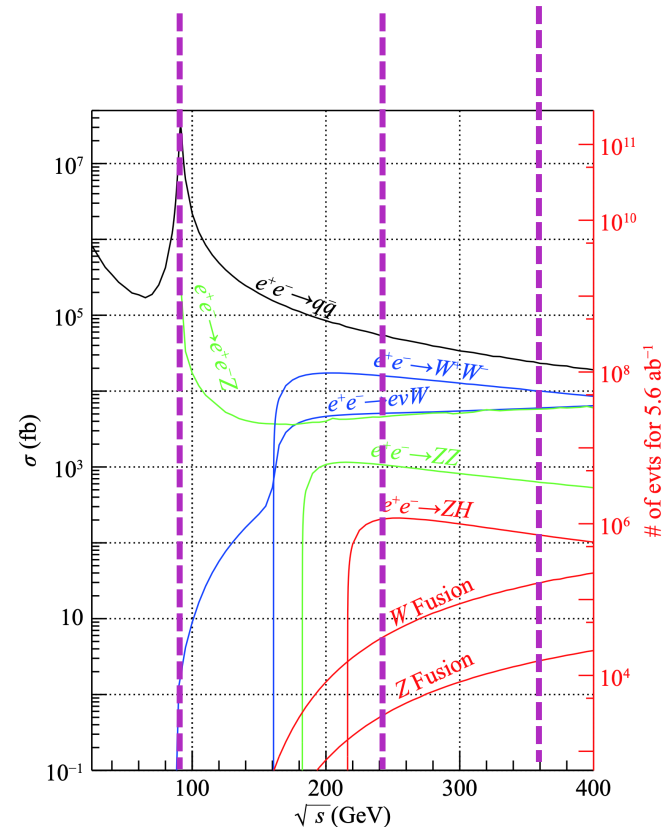


[arXiv:1809.00285](https://arxiv.org/abs/1809.00285), [arXiv:2203.09451](https://arxiv.org/abs/2203.09451)





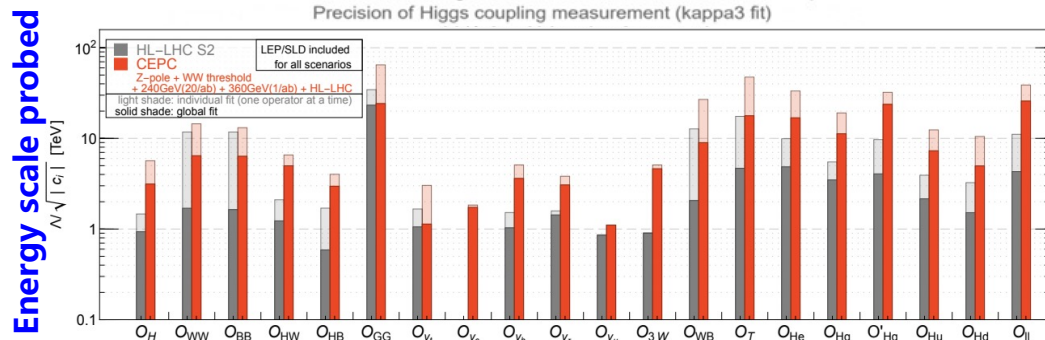
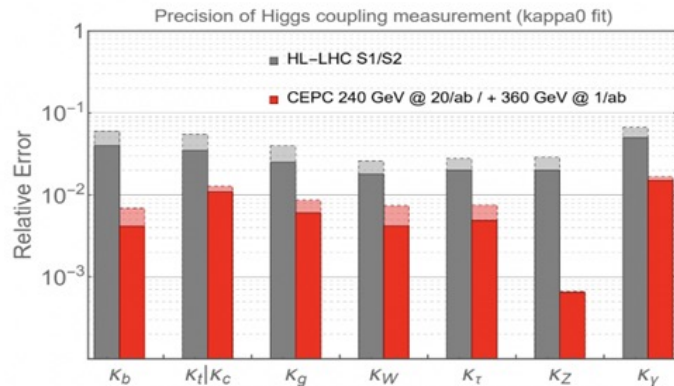
CEPC Operation mode		ZH	Z	W ⁺ W ⁻	ttbar
\sqrt{s} [GeV]		~ 240	~ 91.2	~ 160	~ 360
Run time [years]		7	2	1	-
CDR (30M W)	$L / \text{IP} [\times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}]$	3	32	10	-
	$\int L dt [\text{ab}^{-1}, 2 \text{ IPs}]$	5.6	16	2.6	-
	Event yields [2 IPs]	1×10^6	7×10^{11}	2×10^7	-
Run time [years]		10	2	1	5
Latest TDR (50M W)	$L / \text{IP} [\times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}]$	8.3	191.7	26.6	0.8
	$\int L dt [\text{ab}^{-1}, 2 \text{ IPs}]$	20	96	7	1
	Event yields [2 IPs]	4×10^6	4×10^{12}	5×10^7	5×10^5





- Precision Higgs, EW, flavor physics & QCD measurements at unprecedented precision
- BSM physics (e.g. dark matter, EW phase transition, SUSY, LLP, ...) up to ~ 10 TeV scale

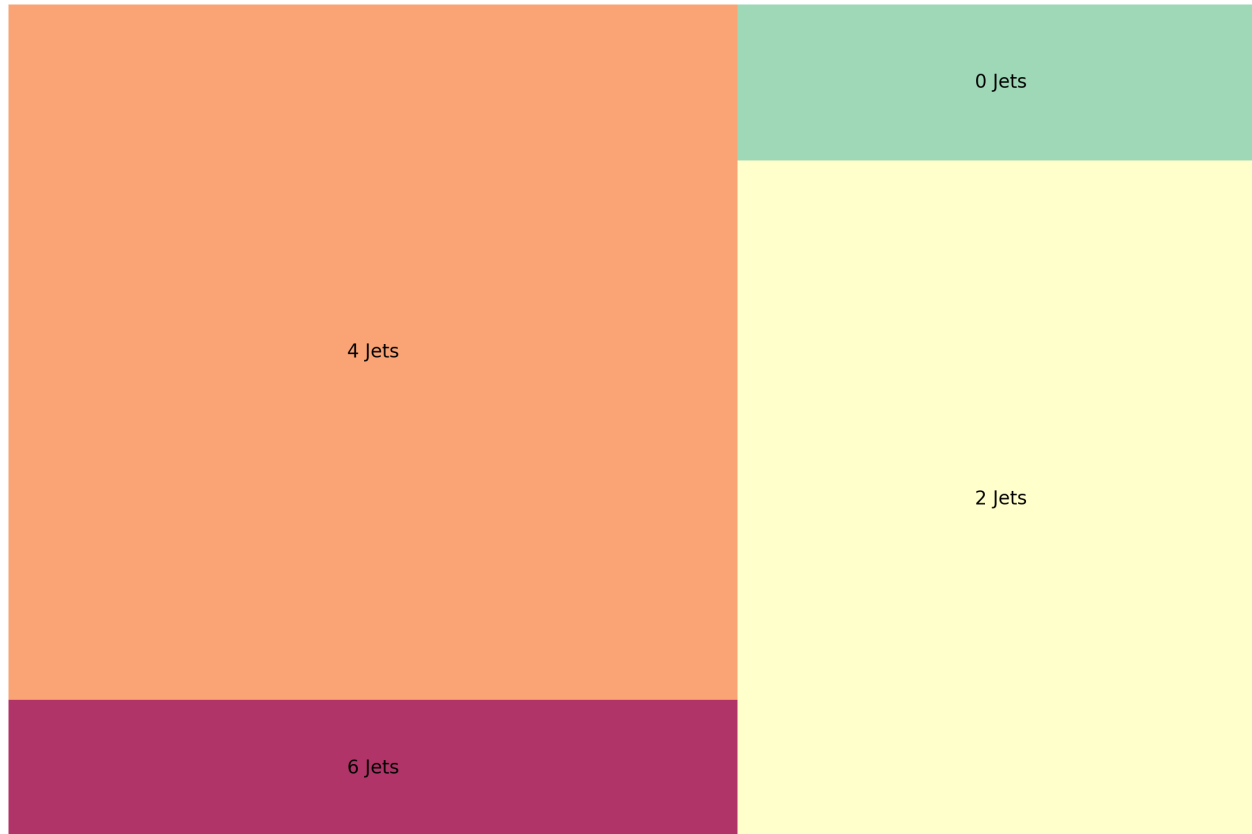
	240 GeV, 20 ab ⁻¹		360 GeV, 1 ab ⁻¹		
	ZH	vvH	ZH	vvH	eeH
inclusive	0.26%		1.40%	\	\
$H \rightarrow bb$	0.14%	1.59%	0.90%	1.10%	4.30%
$H \rightarrow cc$	2.02%		8.80%	16%	20%
$H \rightarrow gg$	0.81%		3.40%	4.50%	12%
$H \rightarrow WW$	0.53%		2.80%	4.40%	6.50%
$H \rightarrow ZZ$	4.17%		20%	21%	
$H \rightarrow \tau\tau$	0.42%		2.10%	4.20%	7.50%
$H \rightarrow \gamma\gamma$	3.02%		11%	16%	
$H \rightarrow \mu\mu$	6.36%		41%	57%	
$H \rightarrow Z\gamma$	8.50%		35%		
$\text{Br}_{\text{upper}}(H \rightarrow \text{inv.})$	0.07%				
Γ_H	1.65%		1.10%		



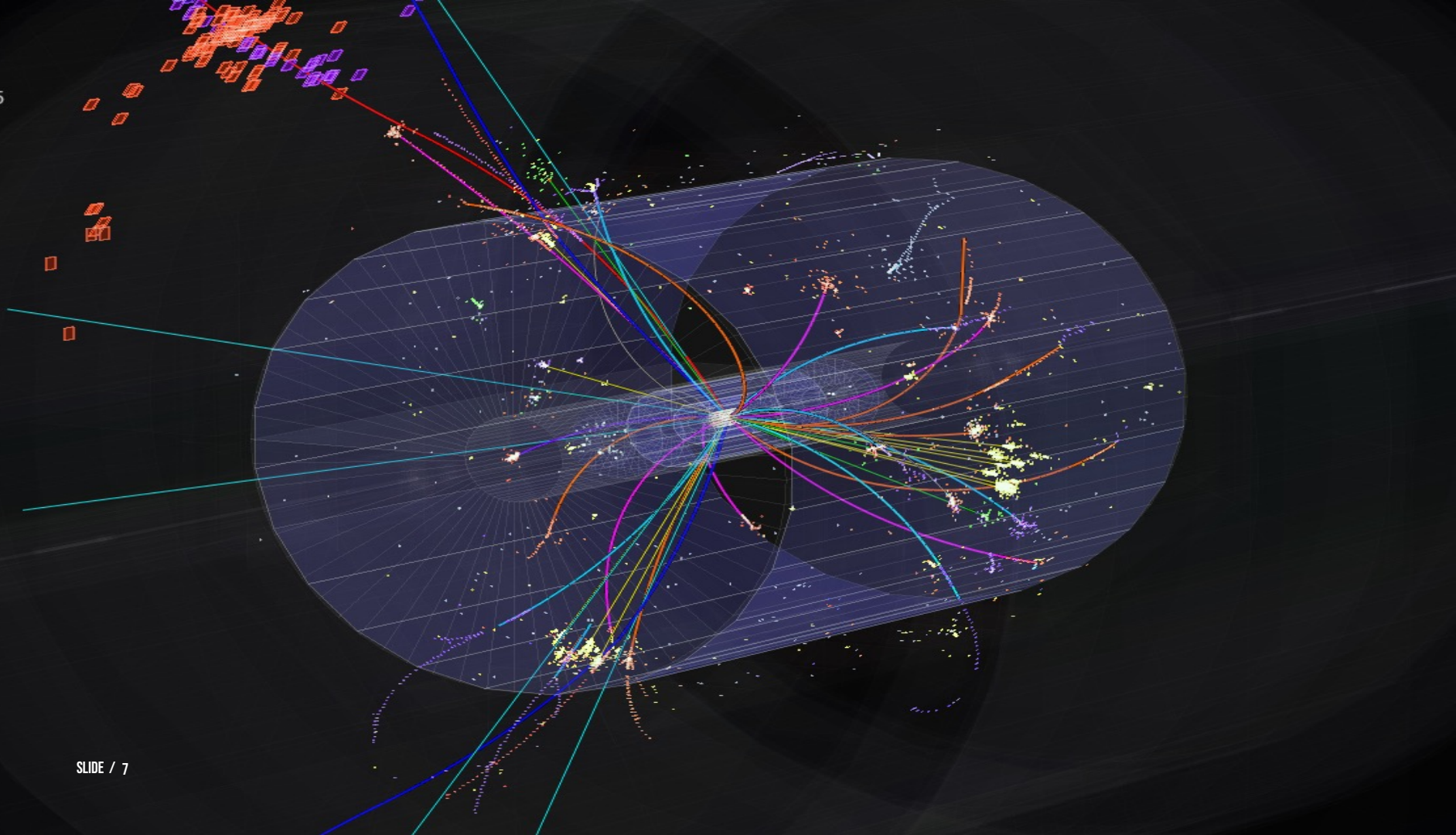
arXiv:2205.08553

CEPC can reveal new physics at energy ~ 10 TeV or higher

Jet is important physics objects



>90% of ZH events contains jets



Jet tagging setup

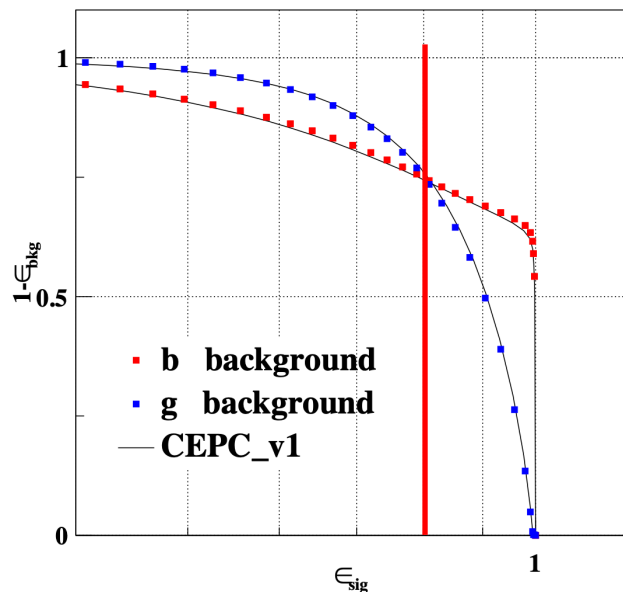
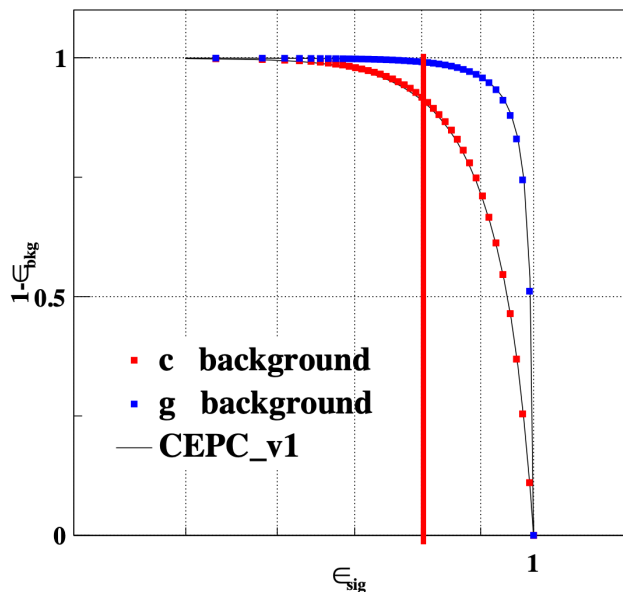
- 91 GeV, Z pole
- $Z \rightarrow bb$
- $Z \rightarrow cc$
- $Z \rightarrow qq$ (uu,dd,ss)
- WHIZARD generator
- CEPCSW full simulation and reconstruction
- Jet Clustering with ee-kt algorithm
- 450k events for each class (900k jets)

Jet tagging with BDT: feature engineering

nvtx=0	trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr5sigma jprobrz5sigma d0bprob d0cprob d0qprob z0bprob z0cprob z0qprob nmuon nelectron trkmass(17)
nvtx=1&&nvtxall=1	trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobrz 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 jprobrz 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 jprobrz 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)

Lots of work!

Performance: ROC used to evaluate how good the algorithm



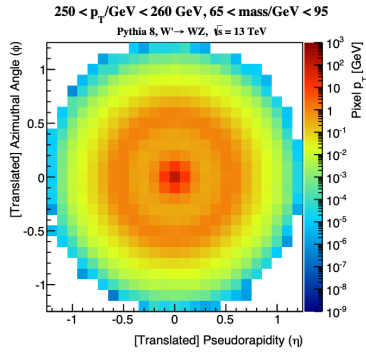
Average accuracy 80%

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

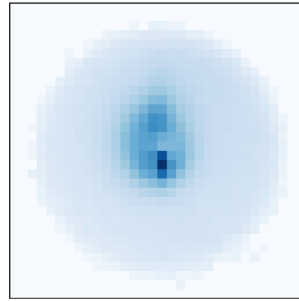
80% c-tagging eff. : Reject 75% b and 75% o jets

ROC: Receiver Operating Characteristic Curve

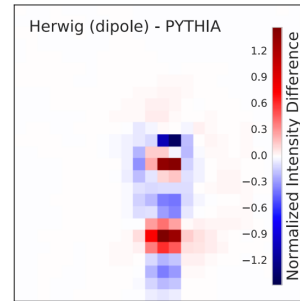
Jet as images



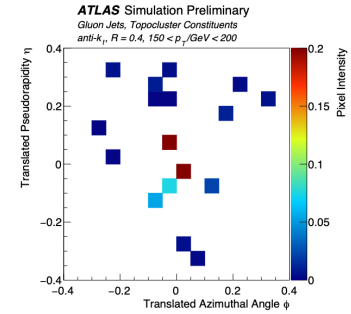
1511.05190



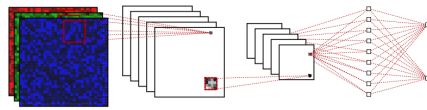
1603.09349



1609.00607



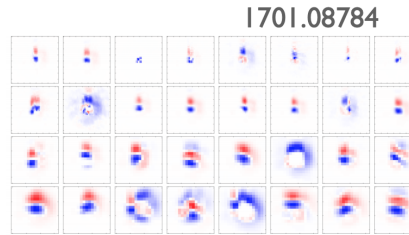
ATL-PHYS-PUB-2017-017



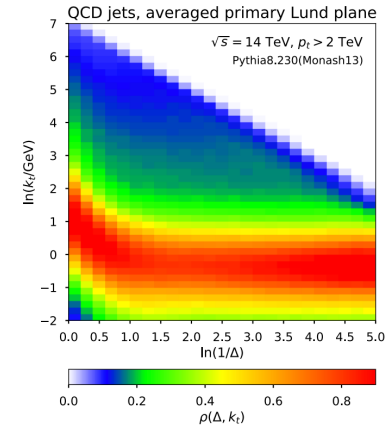
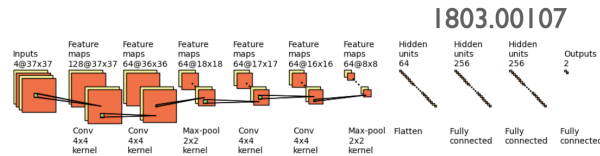
red = transverse momenta of charged particles

green = the transverse momenta of neutral particles

blue = charged particle multiplicity

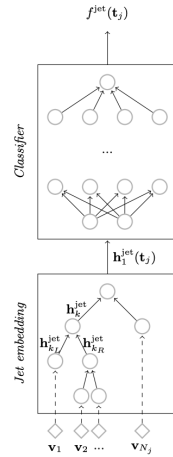
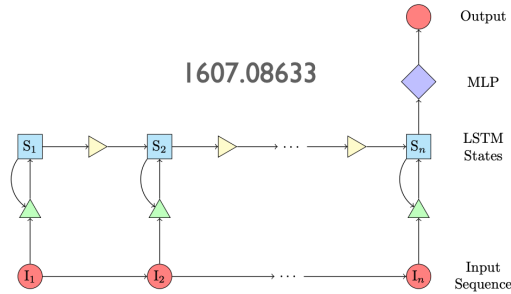


1612.01551

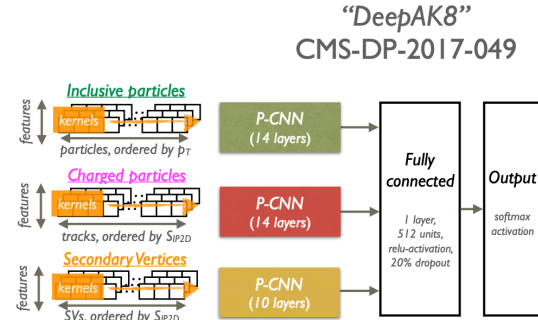


1807.04758

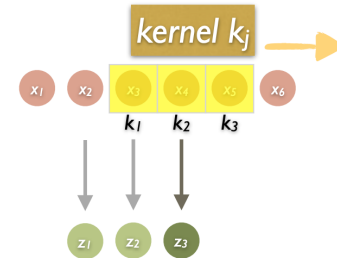
Jet as sequence



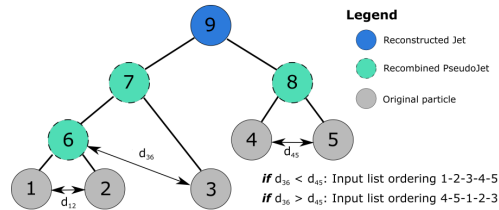
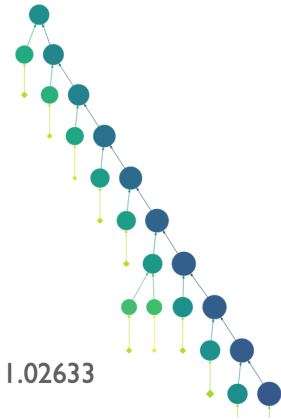
1702.00748



P-CNN:



1711.02633



1711.09059

A jet

Unordered, variable number of particles

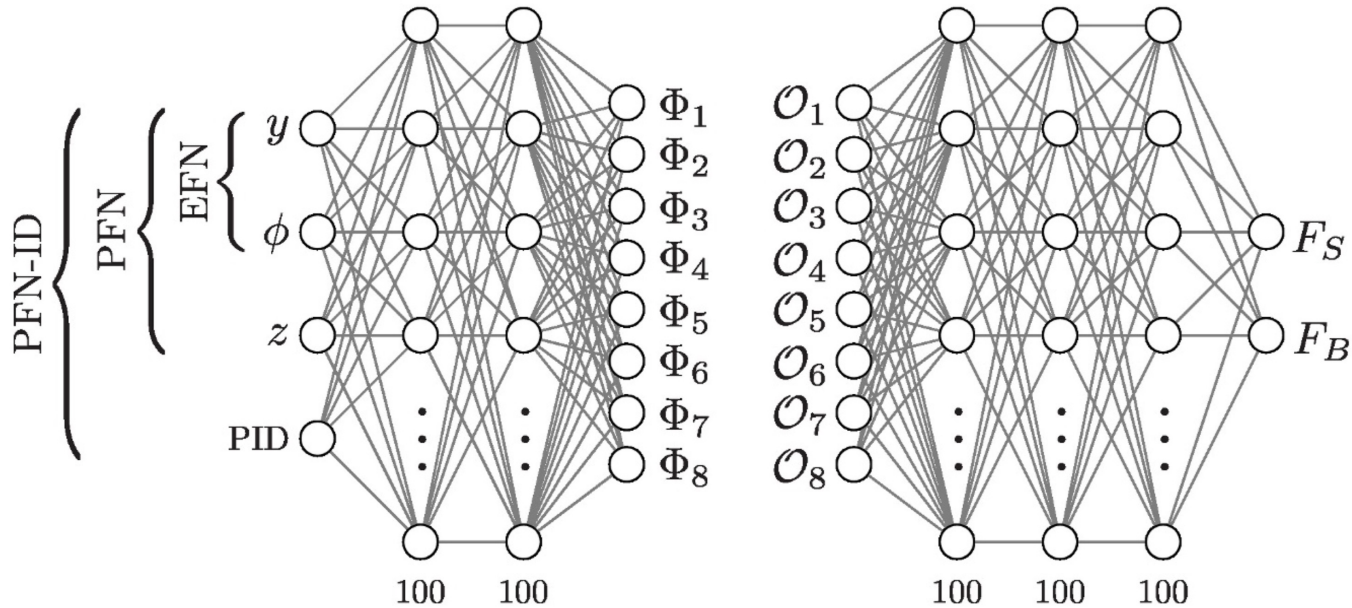
Permutation invariant
Equivariant

$$J(\{p_1^\mu, \dots, p_M^\mu\}) = J(\{p_{\pi(1)}^\mu, \dots, p_{\pi(M)}^\mu\})$$

p_i^μ represents *all* the particle properties:

Mapping

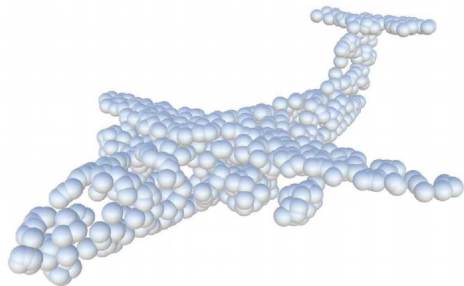
DNN



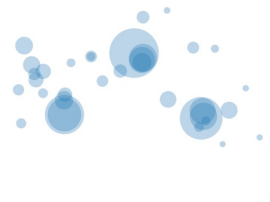
EFN & PFN

Jet as sets or cloud

Point cloud vs particle cloud



- Point cloud
 - points are intrinsically unordered
 - primary information:
 - 3D coordinates in the xyz space

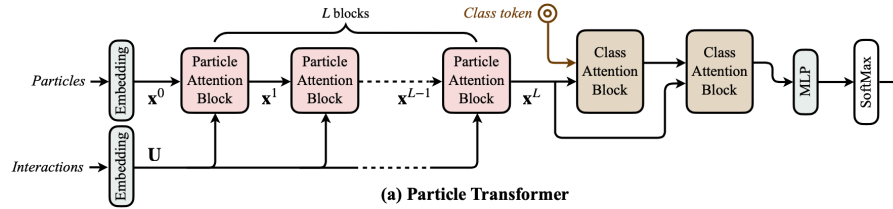


- Particle cloud
 - particles are intrinsically unordered
 - primary information:
 - 2D coordinates in the η - φ space
 - **but also additional “features”**:
 - energy/momenta
 - charge/particle type
 - track quality/impact parameters/etc.

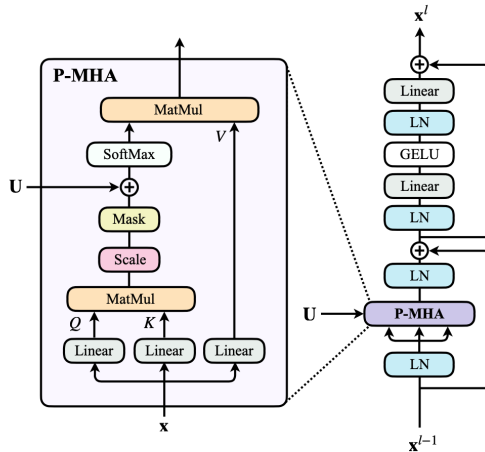
Particle Transformer

ParT architecture

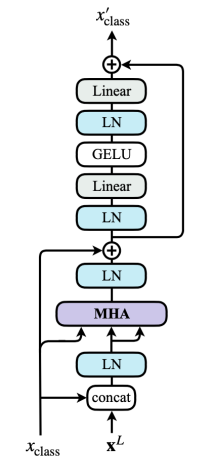
H.Ou et al. [arXiv:2202.03772](https://arxiv.org/abs/2202.03772), proceedings of 39th ICML, Vol.162



(a) Particle Transformer



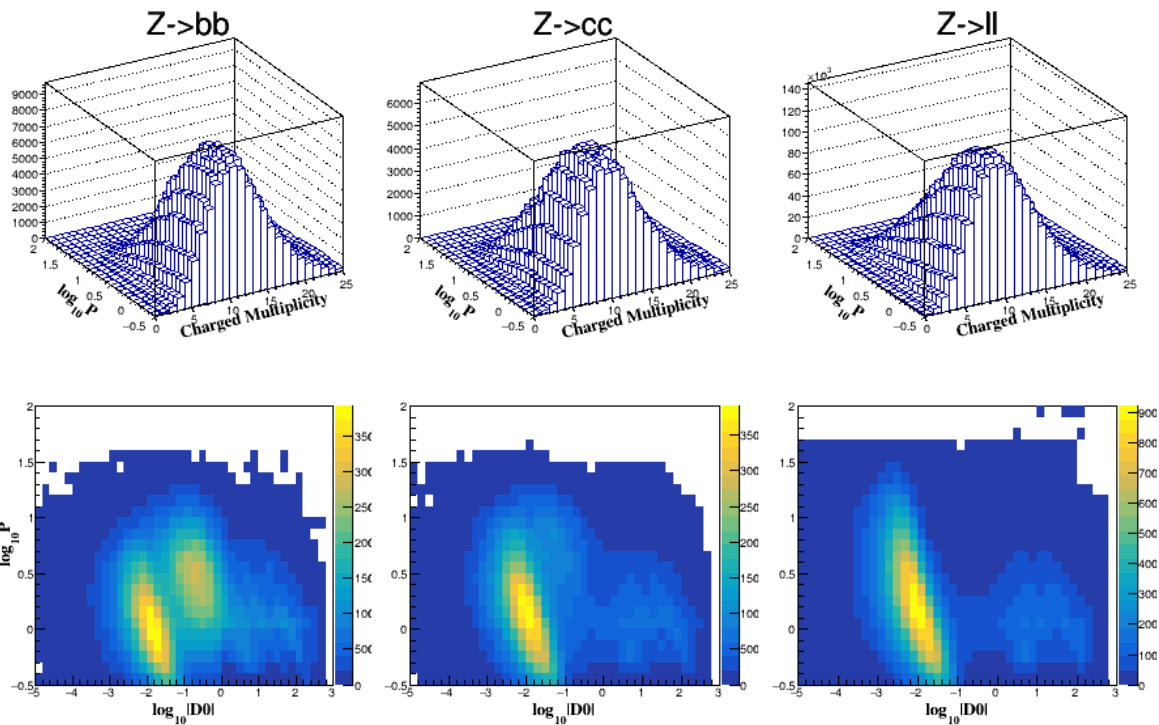
(b) Particle Attention Block



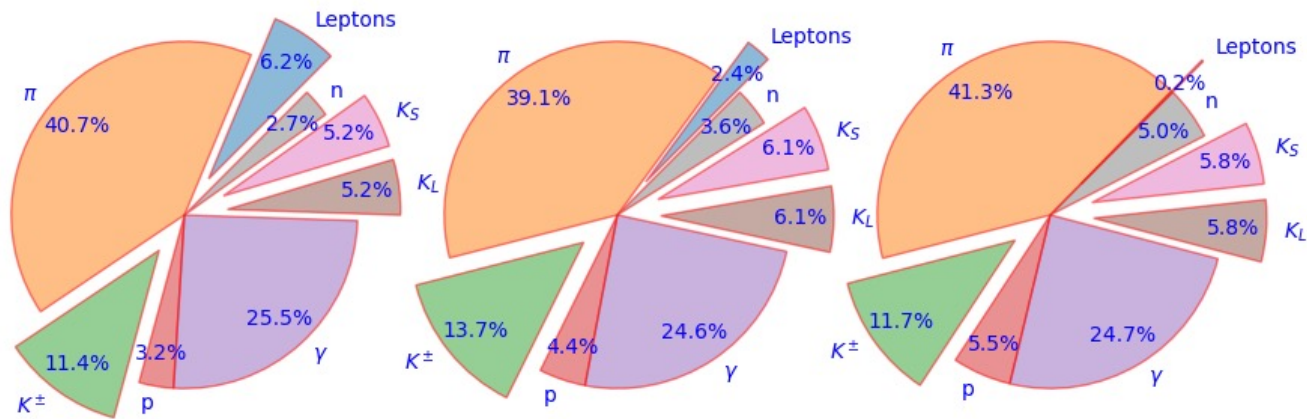
(c) Class Attention Block

Multi-head attention introduced

Visualizing features



Visualizing features



Particle identification

Comparing the performances of all algorithms

Algorithm	ParticleNet	PFN	DNN	BDT	GBDT	gcforest	XGBoost
Accuracy	0.872	0.850	0.788	0.776	0.794	0.785	0.801

ArXiv: [2208.13503](https://arxiv.org/abs/2208.13503)

Impacts on physics study

tag	$\epsilon_S(\%)$	$\epsilon \times \rho$			
		LCFIPlus	XGBoost	ParticleNet	PFN
<i>b</i>	60	-	-	0.589	0.596
	70	-	-	0.694	0.689
	80	-	0.747	0.780	0.763
	90	0.72	0.713	0.810	0.752
	95	-	0.609	0.721	0.645
<i>c</i>	60	0.36	-	0.548	0.485
	70	-	-	0.589	0.497
	80	-	0.345	0.584	0.467
	90	-	0.292	0.516	0.402
	95	-	0.251	0.451	0.348

Roughly estimation of c-jet

$\text{sqrt}(0.584/0.345)=1.3$

Statistical uncertainty can be reduced by 30%

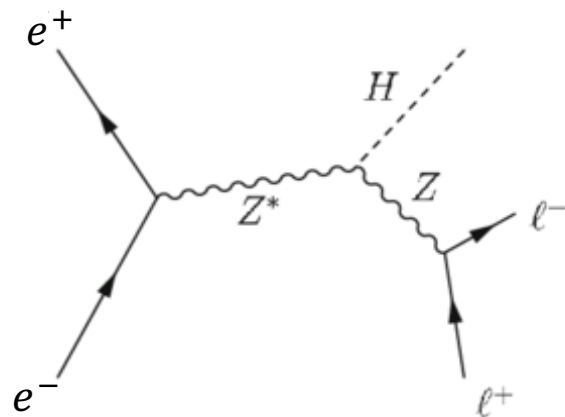
$$\frac{1}{(\Delta\sigma_s)^2} = \frac{1}{\sigma_s} \mathcal{L}_{\epsilon_s \rho} = \frac{1}{\sigma_s^2} S_{\text{tot}} \epsilon_s \rho$$

Could be more ambitious: classify events

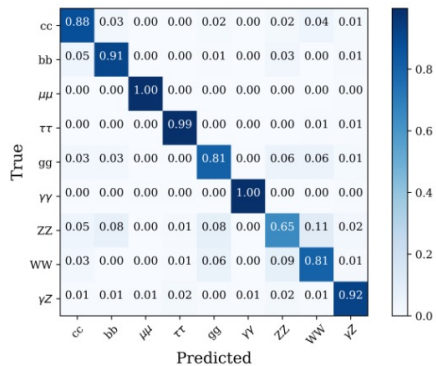
- Most important objective : $e^+e^- \rightarrow ZH$
- 4 Z decay modes

$ee, \mu\mu, \tau\tau, qq$

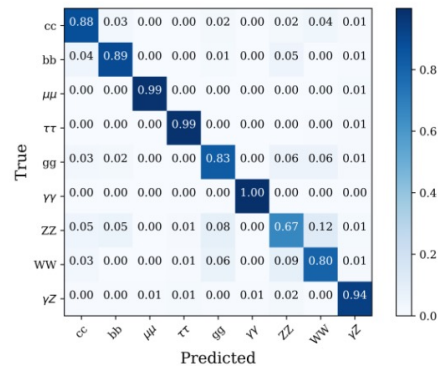
- 10 (or 9) Higgs decay modes
 $ss, cc, bb, \mu\mu, \tau\tau, \gamma\gamma, gg, WW, ZZ, \gamma Z$
- 40 in total
- 9 background processes
- 400k events for each



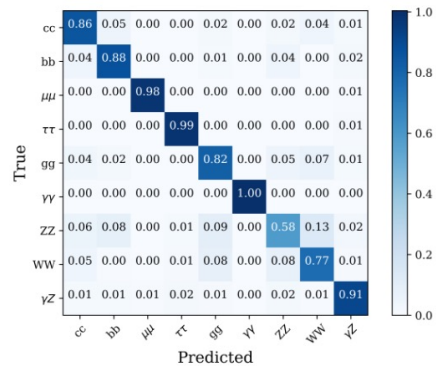
Confusion matrices of 9-classification



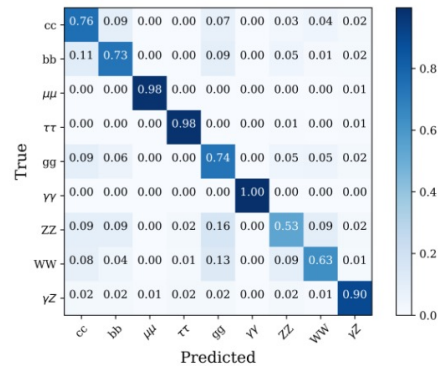
(a)



(b)



(c)

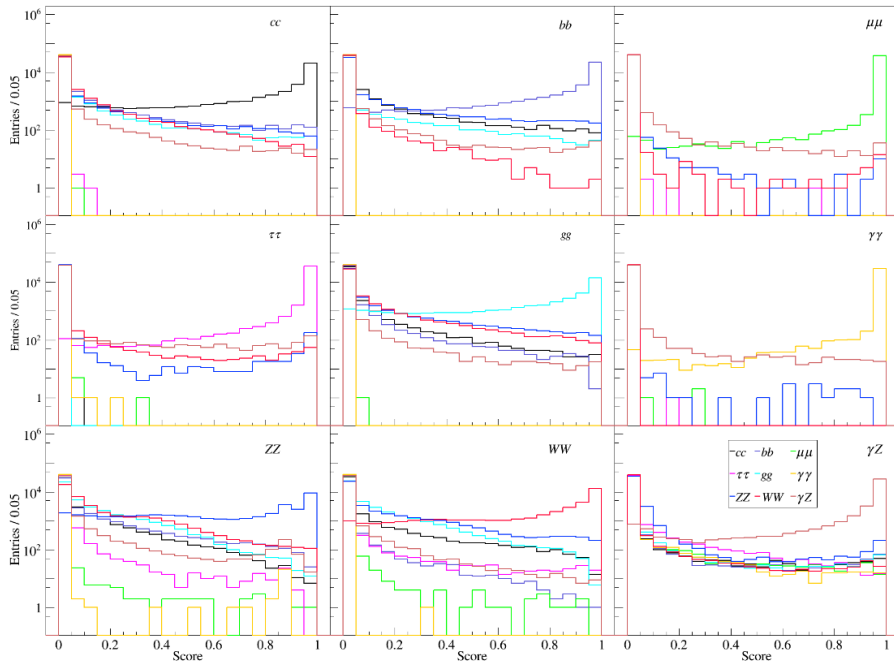


(d)

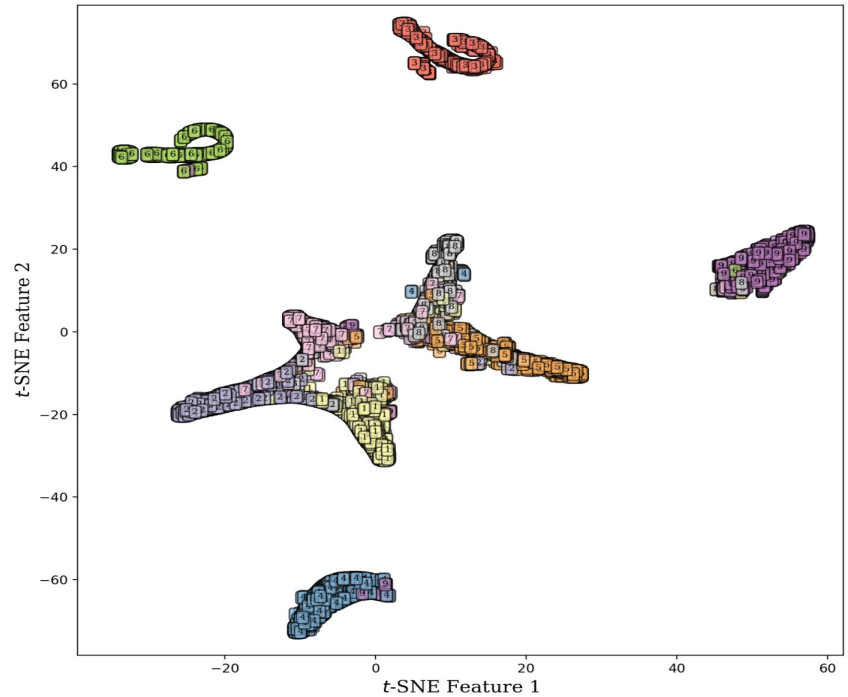
ParticleNet

9-classification

Distribution of score



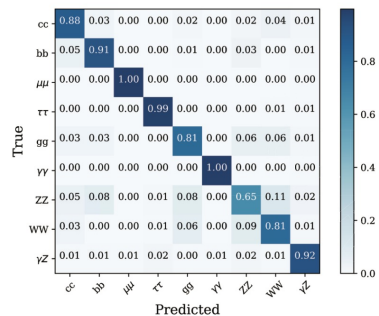
Dimensional reduction (t-SNE)



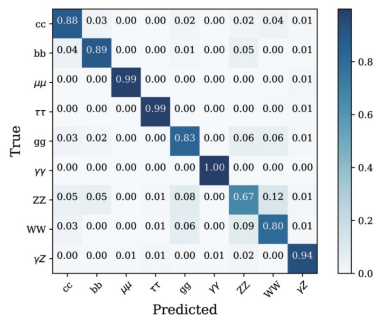
ParticleNet

4x9 signals

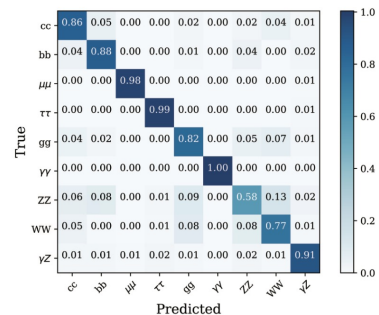
ParT



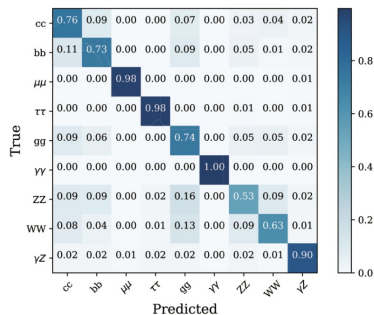
(a)



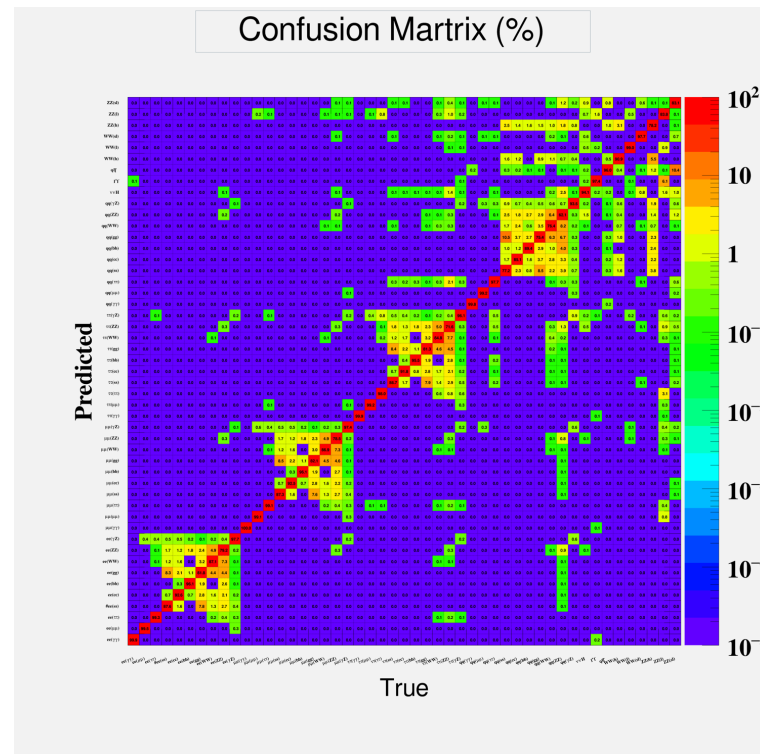
(b)



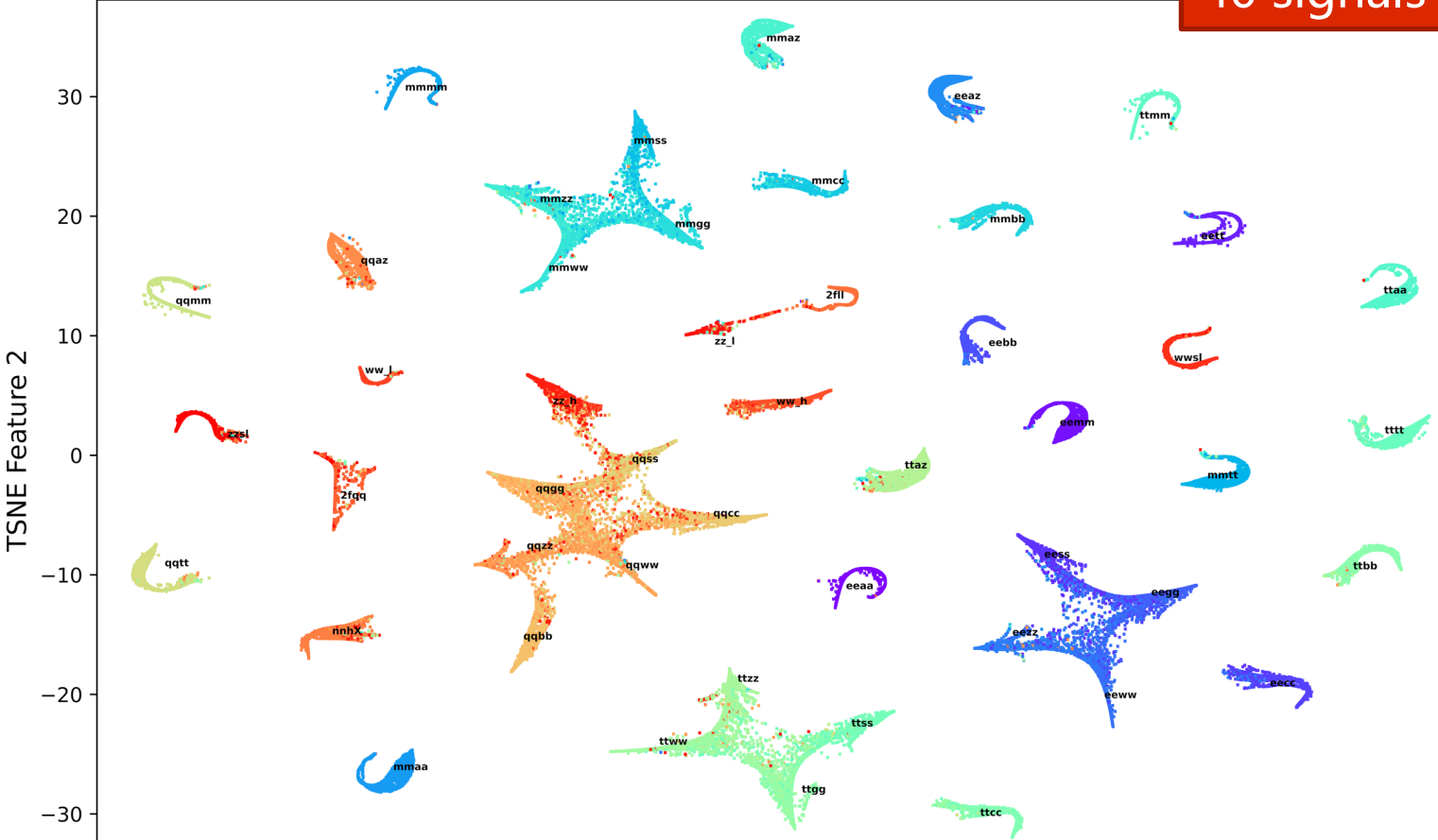
(c)



(d)



40 signals + 9 bkgs



ParT

Combination is going on

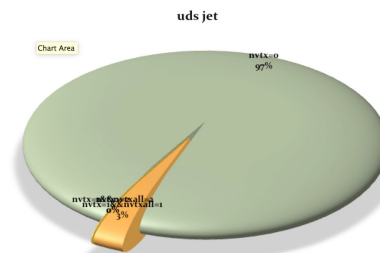
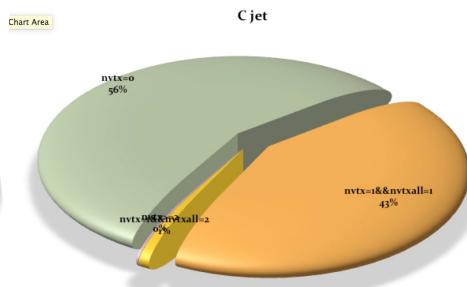
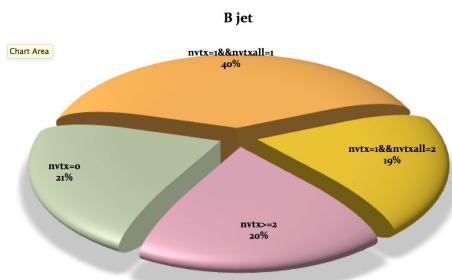
Summary

- Jet is important physics object at CEPC
- Jet tagging studied with various many ML method at CEPC
- Event classification is promising and efficient analysis method
- More topics: clustering, counting, jet mass, jet charge, ...
- BESIII study, unsupervised clustering, ...

Extras

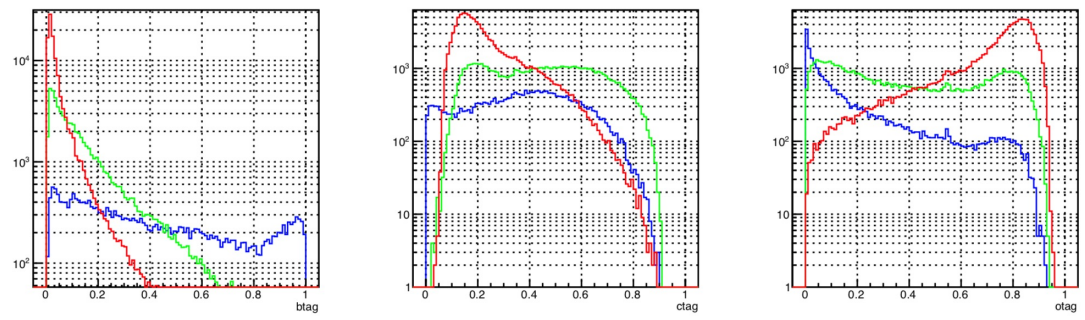
Check the # of vertexes

	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

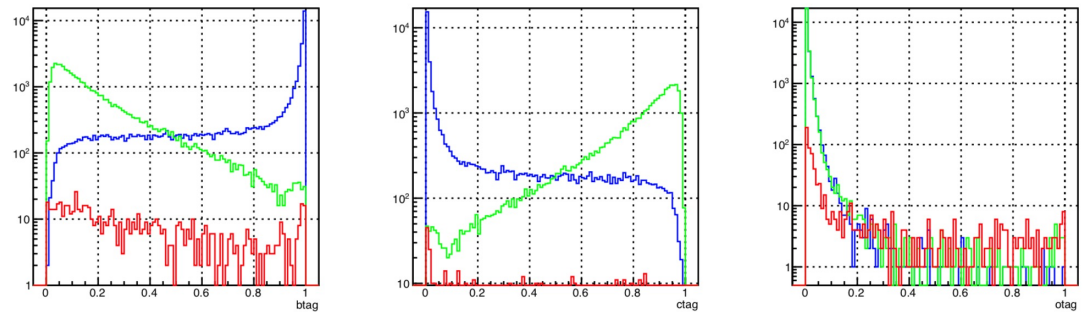


b, c, and q likeliness

Category 1



Category 2



$btag+ctag+otag=1$, so 2 outputs independent