Performance of jet flavor tagging and measurement of R_b using ParticleNet at CEPC

Libo Liao,

Gang LI, Weimin Song, Shudong Wang, and Zhaoling Zhang

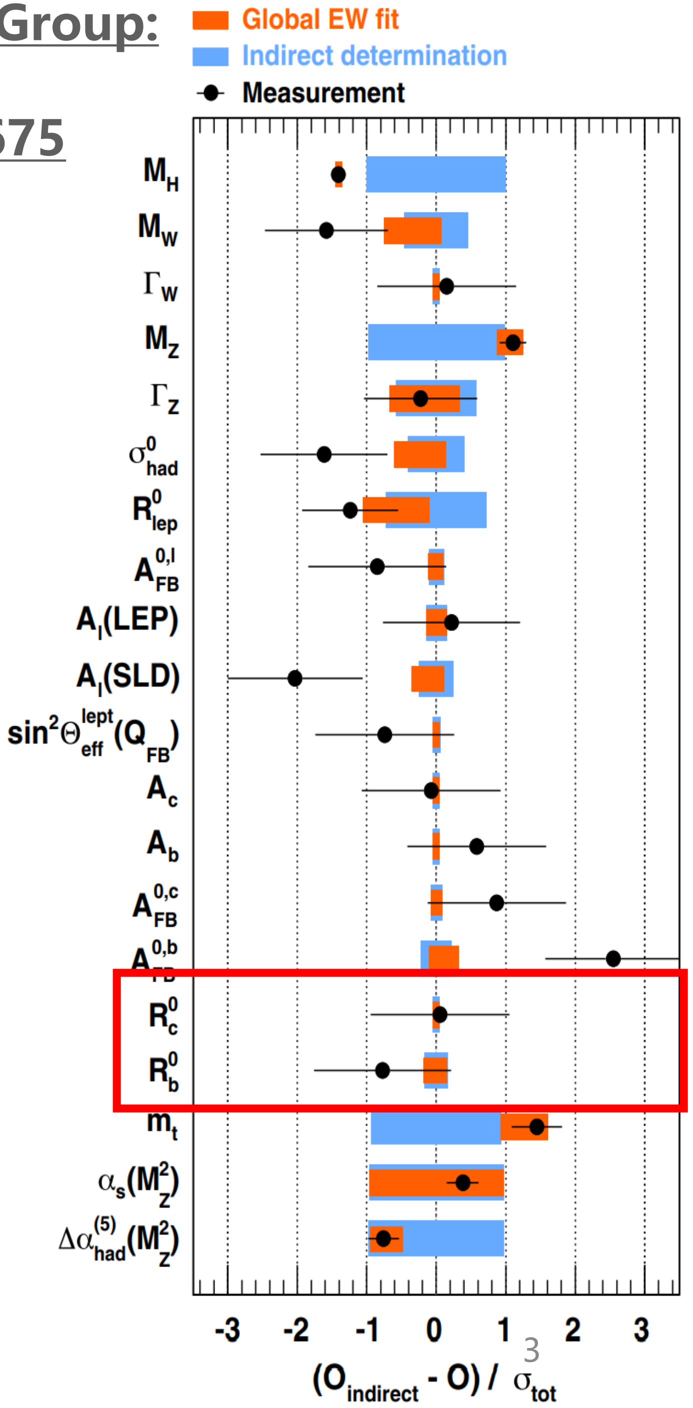
2022.10.28

Outline

- ➢Introduction
- ➢Deep learning architectures & Datasets
- ➢Evaluation metrics
- ➢Tagging
  - ➢Jet tagging
  - ➢Event tagging
- ➢R_b & R_c measurements
- ➢Conclusion

Introduction

Gfitter Group: EPJC78 (2018)675



- Relative decay width plays a key role in SM testing and experiment of precision measurement of Z boson

$$R_q = \Gamma_q/\Gamma_h$$

$$\Gamma(Z \to q\bar{q}) = 12\boxed{\Gamma_0}(\boxed{g_{Aq}^2}\boxed{R_{Aq}} + \boxed{g_{Vq}^2}\boxed{R_{Vq}})$$

- Status of R_b and R_c measurements in experiment and theory
  - Theoretical << Experimental

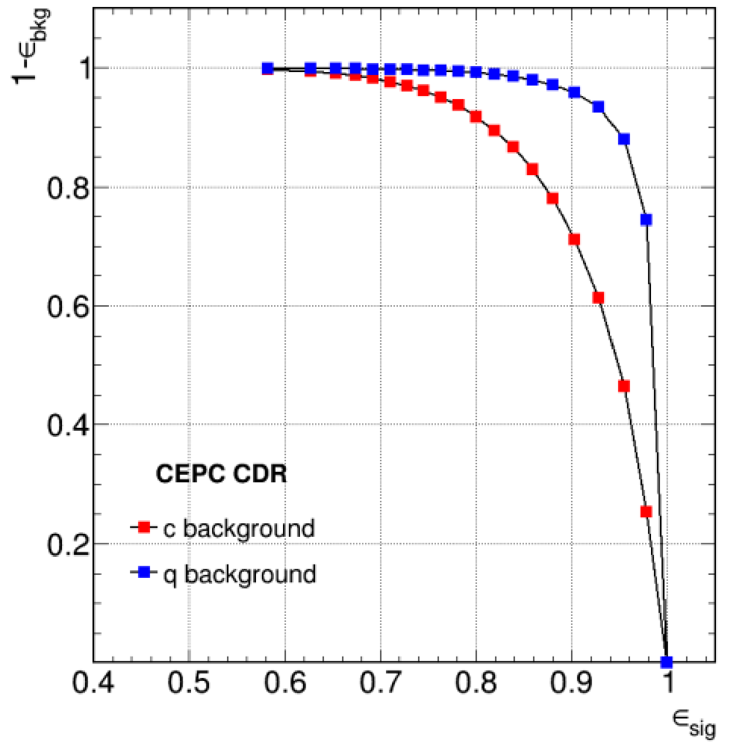
	Experiment	Theory
R_b	0.21629 ± 0.00066	0.21581 ± 0.00002
R_c	0.1721 ± 0.0030	0.17221 ± 0.00003

Introduction

- Jet: Key physics object
  - Vertexing->Clustering->Tagging
- Tagging methods
  - Cut-based->TMVA->Deep Learning
- CEPC baseline detector (TMVA/BDT)

  80% eff. & 90% purity in b-tagging

  60% eff. & 60% purity in c-tagging

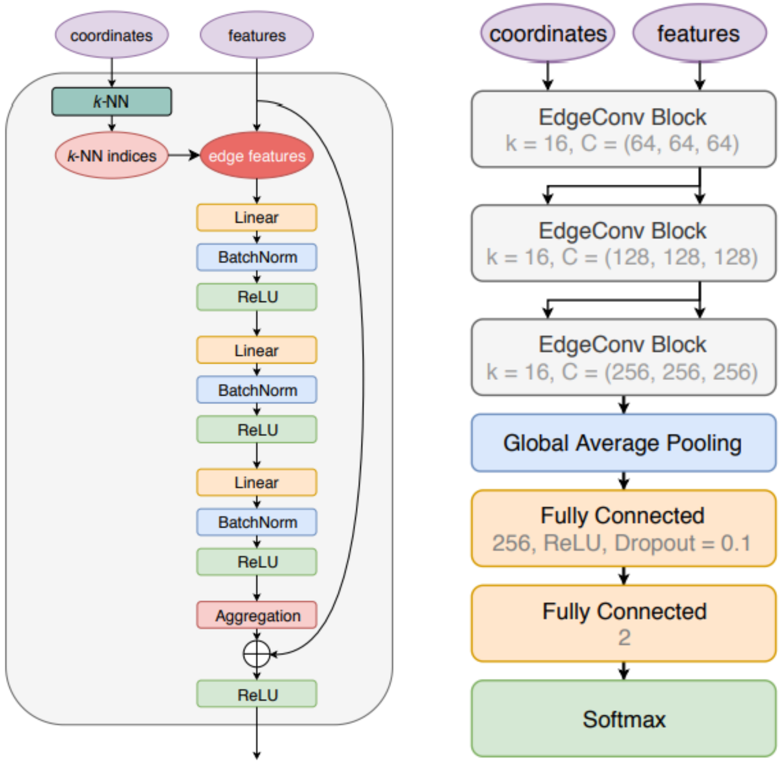


- To measure Rq, a new method introduced
  - Event tagging

Deep learning architectures

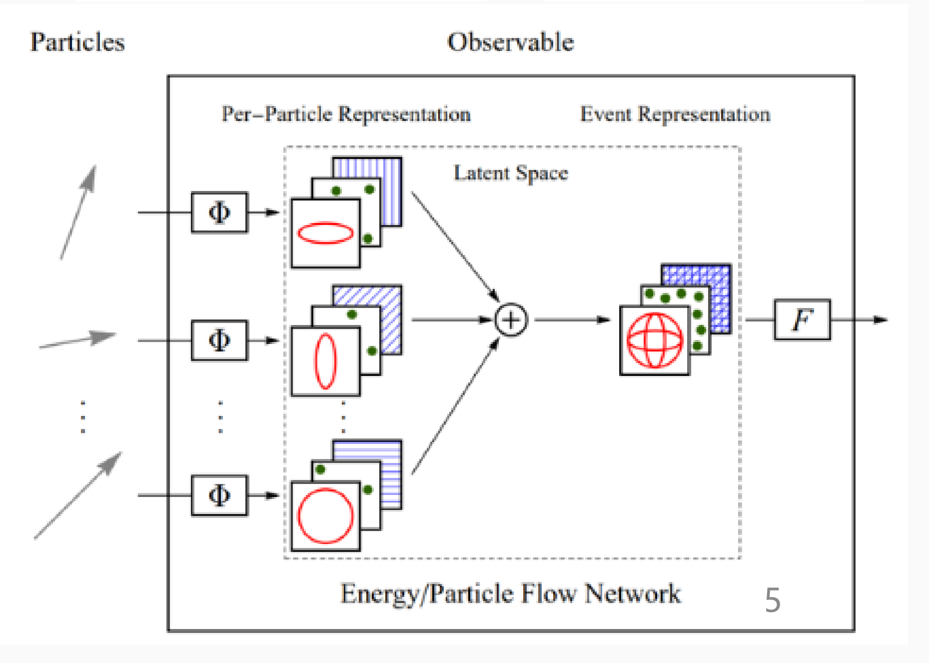
[Jet tagging via particle clouds]

- ParticleNet
  - Treating a jet as an unordered set of particles in space
  - Using permutation-invariant graph neural networks



[Energy flow networks: deep sets for particle jets]

- Particle Flow Network (PFN)
  - Based on "point clouds"
  - As a cross check



Datasets

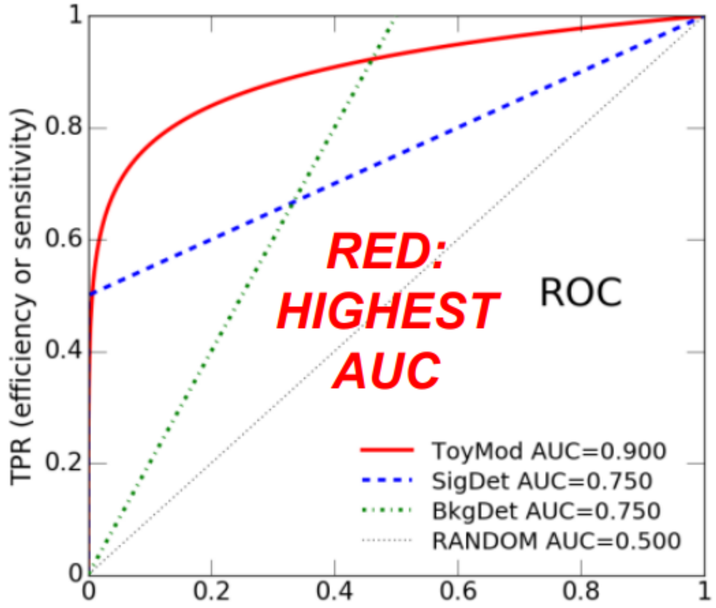
- Full simulation with CEPC baseline detector at Z-pole
- PID used as a feature by matching reconstruction and MC truth
- In flavor tagging
  - 900k jets for each flavor($b, c, o = uds$)
  - Clustered by $ee - kt$ into 2 jets
- In event tagging(same samples are used)
  - $Z \rightarrow b\bar{b}, c\bar{c}, o\bar{o}$, 450k events for each channel
  - The main background $Z \rightarrow \tau^{+}\tau^{-}$ considered
  - No jet clustering (directly classify events into different flavor)
- Train:validation:test = 7:1.5:1.5

Evaluation metrics

- Efficiency ϵ_s = TP/(TP+FN)
- Purity ρ_s = TP/(TP+FP)
- Accuracy = (TP+TN)/ALL
- ROC/AUC
- $\epsilon_s \times \rho_s$: between 0 and 1
  - The higher, the better
  - Proportional to 1/error2

$$(\Delta R_i)^2 \propto \frac{1}{\epsilon_i \rho_i}$$

TP (S_{sel}), FP (B_{sel}), **FN** (S_{rej}), TN (B_{rej})	**TP** (S_{sel}), **FP** (B_{sel}), FN (S_{rej}), TN (B_{rej})	TP (S_{sel}), **FP** (B_{sel}), FN (S_{rej}), **TN** (B_{rej})
$\mathbf{TPR} = \frac{\mathbf{TP}}{\mathbf{TP}+\mathbf{FN}}$	$\mathbf{PPV} = \frac{\mathbf{TP}}{\mathbf{TP}+\mathbf{FP}}$	$\mathbf{TNR} = \frac{\mathbf{TN}}{\mathbf{TN}+\mathbf{FP}} = \mathbf{1-FPR}$
HEP: "efficiency" $\epsilon_s = \frac{S_{\mathrm{sel}}}{S_{\mathrm{tot}}}$	HEP: "purity" $\rho = \frac{S_{\mathrm{sel}}}{S_{\mathrm{sel}}+B_{\mathrm{sel}}}$	HEP: "background rejection" $1-\epsilon_b = 1-\frac{B_{\mathrm{sel}}}{B_{\mathrm{tot}}}$

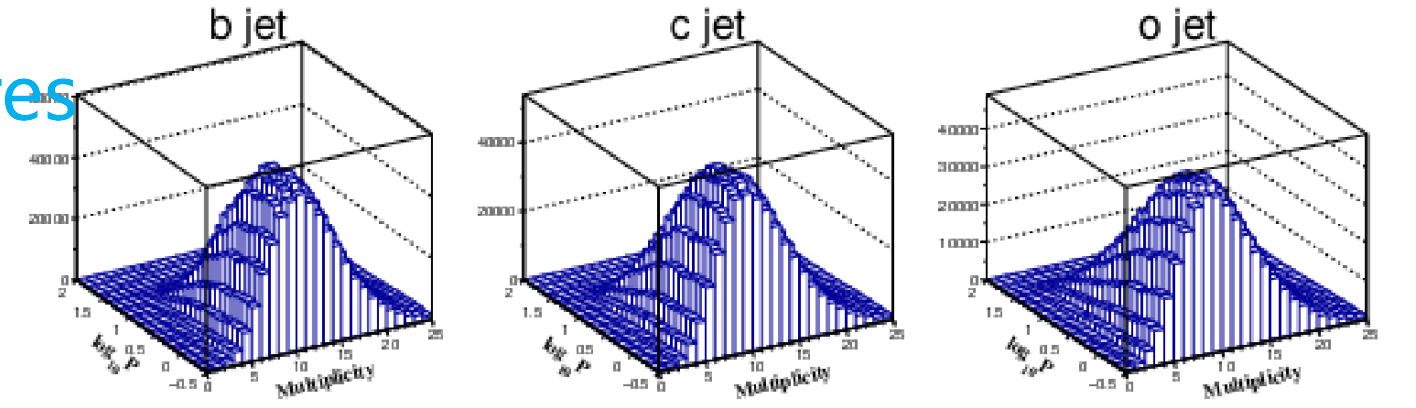


Andrea Valassi : ROC's, AUC's and alternatives in HEP and other domains

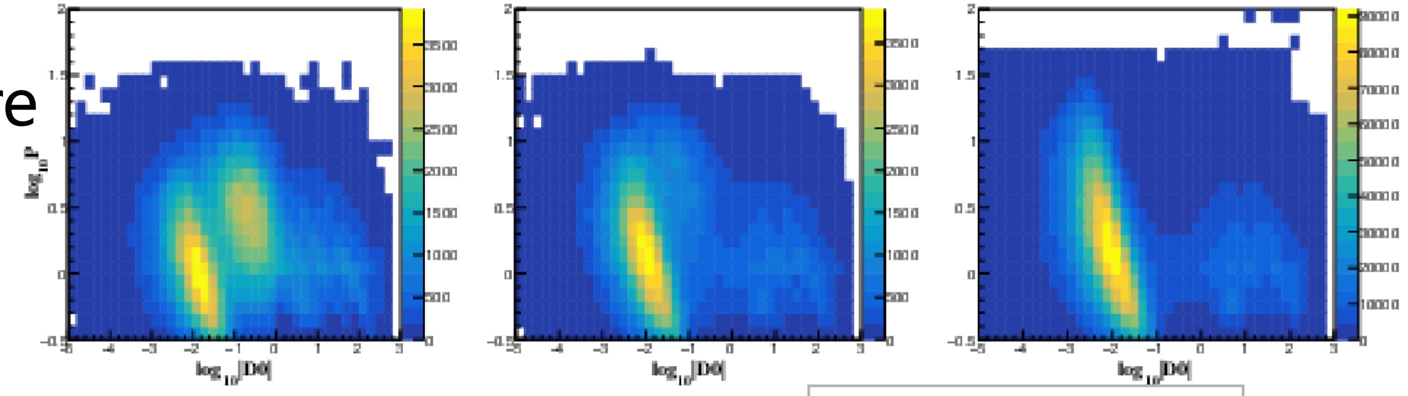
Jet tagging

Jet features

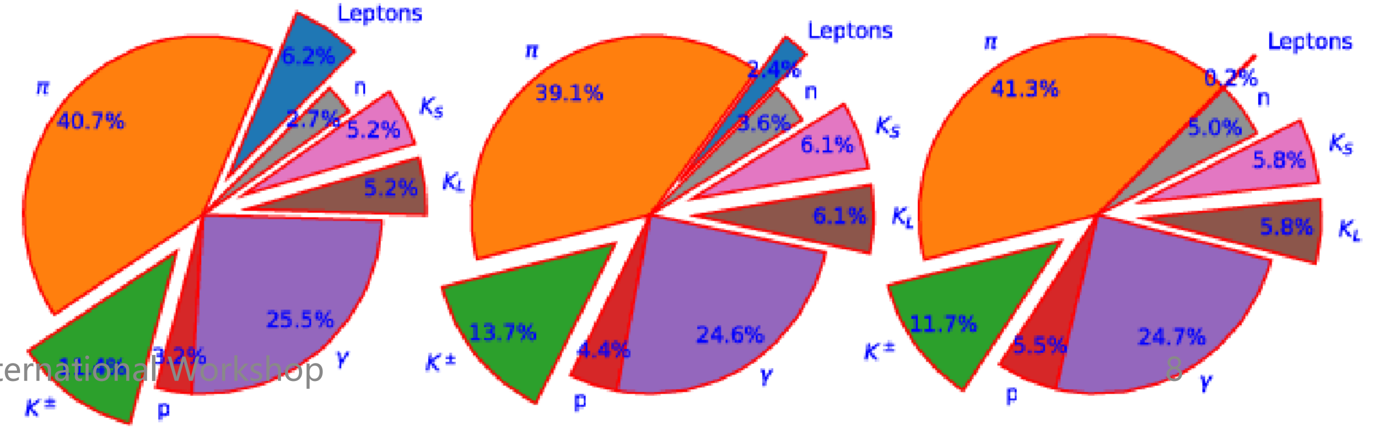
- (a) Multiplicity
  - $b > c > o$
- (b) Impact parameters
  - Larger impact parameters, more energetic tracks in b
- (c) The weighted fractions of particles
  - Far more energetic leptons in b
  - Slightly more energetic K in c



(a) Log(P) vs. Multiplicity



(b) Log(P) vs. log(D0)



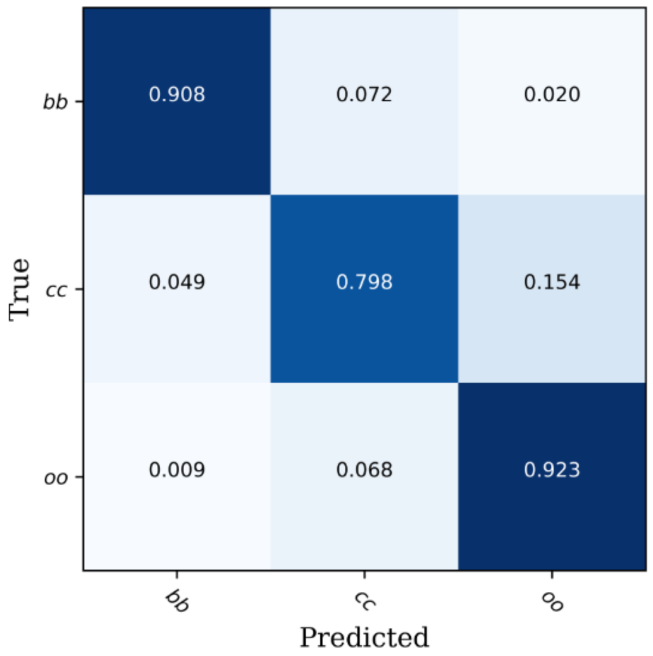
Jet tagging

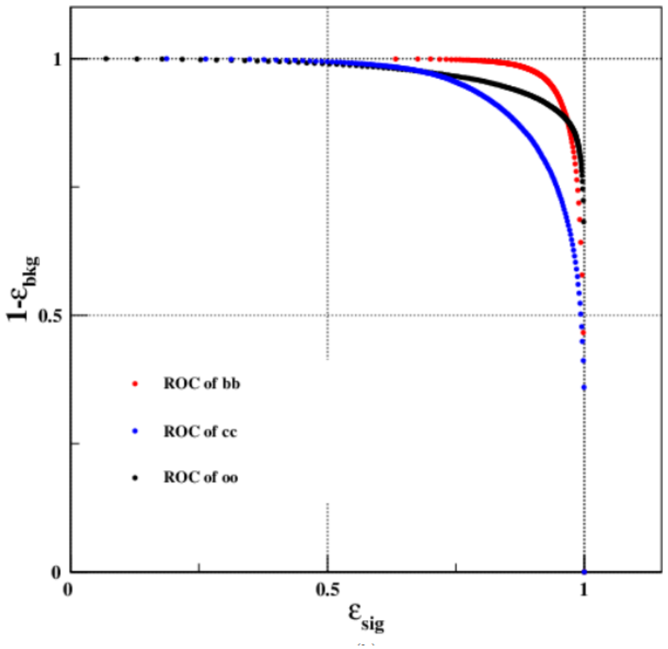
Fan Yang: Flavor Tagging Using Machine Learning

Algorithm	ParticleNet	PFN	DNN	BDT	GBDT	gcforest	XGBoost
Accuracy	0.876	0.850	0.788	0.776	0.794	0.785	0.801

- At least 9% improvement in ParticleNet at global accuracy
  - Richer information
  - Strong inductive bias
- The performance of b-tagging and o-tagging are much better than c-tagging
- ParticleNet is better than the PFN
  - Consistent with the study

    [Jet tagging via particle clouds]





tag	ParticleNet		PFN	
	Efficiency	AUC	Efficiency	AUC
b	0.908	0.986	0.870	0.979
c	0.798	0.951	0.765	0.930
o	0.923	0.974	0.911	0.966

Physics impacts of jet tagging

- Take LCFIPlus & XGBoost(CDR baseline) as reference
  - ParticleNet & PFN are better than the baseline, especially in c-tagging
- Statistical uncertainty can be improved
  - roughly 30%(sqrt(0.597/0.345)) in counting c jets

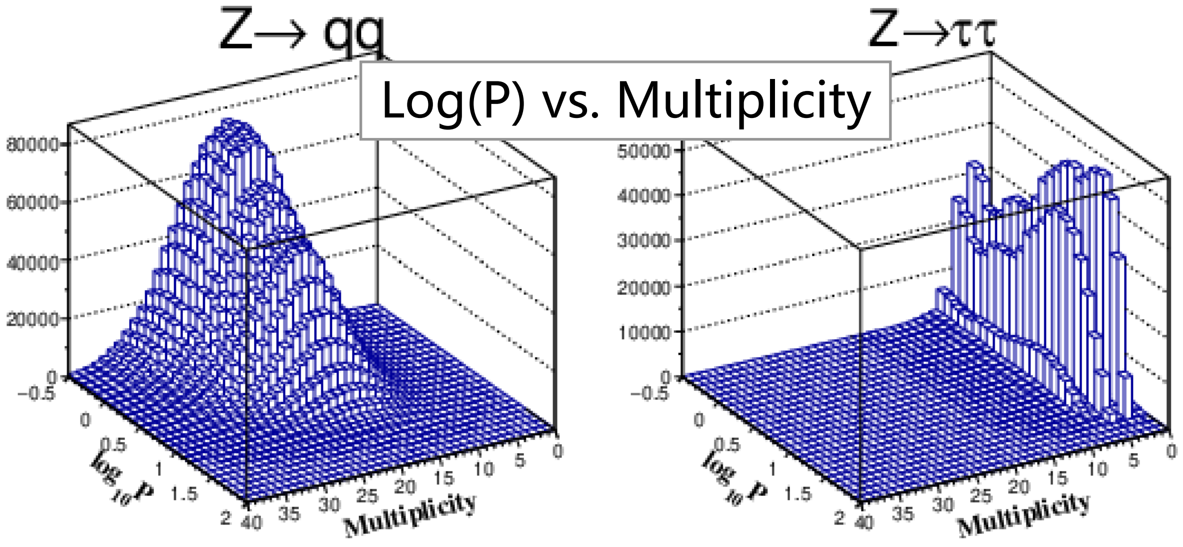
tag	$\epsilon_S(\%)$	$\epsilon \times \rho$			
		LCFIPlus	XGBoost	ParticleNet	PFN
b	80	-	0.747	0.786	0.763
	90	0.72	0.713	0.821	0.752
c	60	0.36	-	0.554	0.485
	70	-	-	0.605	0.497
	80	-	0.345	0.597	0.467
	90	-	0.292	0.532	0.402

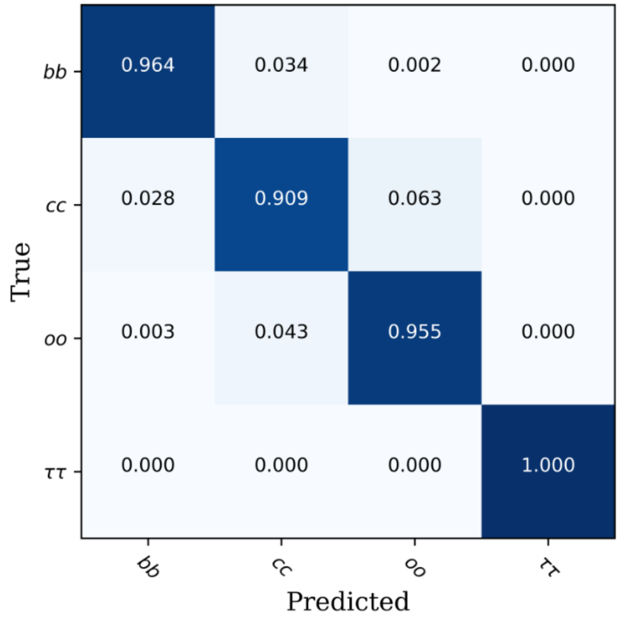
Applied in R_q measurement

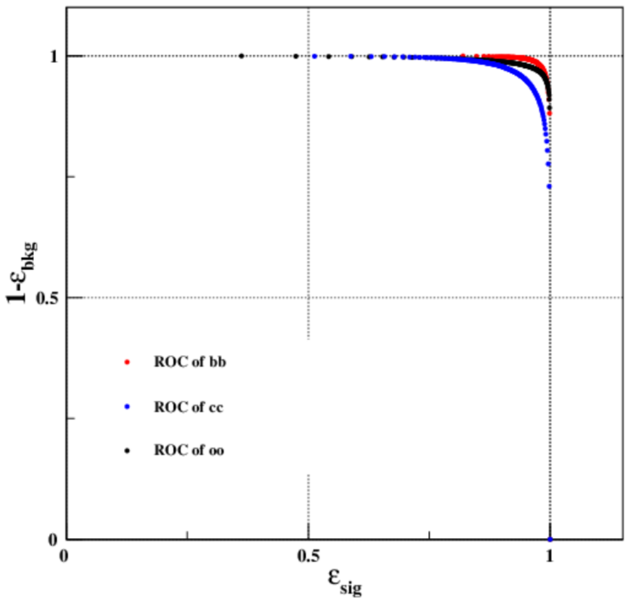
Event tagging

- The multiplicity versus momenta of tracks
  - Few tracks and more energetic in $\tau^+\tau^-$
  - Easy to discriminate $q\bar{q}$ & $\tau^+\tau^-$
- Confusion matrix almost diagonal

tag	ParticleNet		PFN	
	Efficiency	AUC	Efficiency	AUC
b	0.964	0.998	0.930	0.993
c	0.909	0.990	0.832	0.976
o	0.955	0.995	0.945	0.992







Next: R_q measurement with two methods

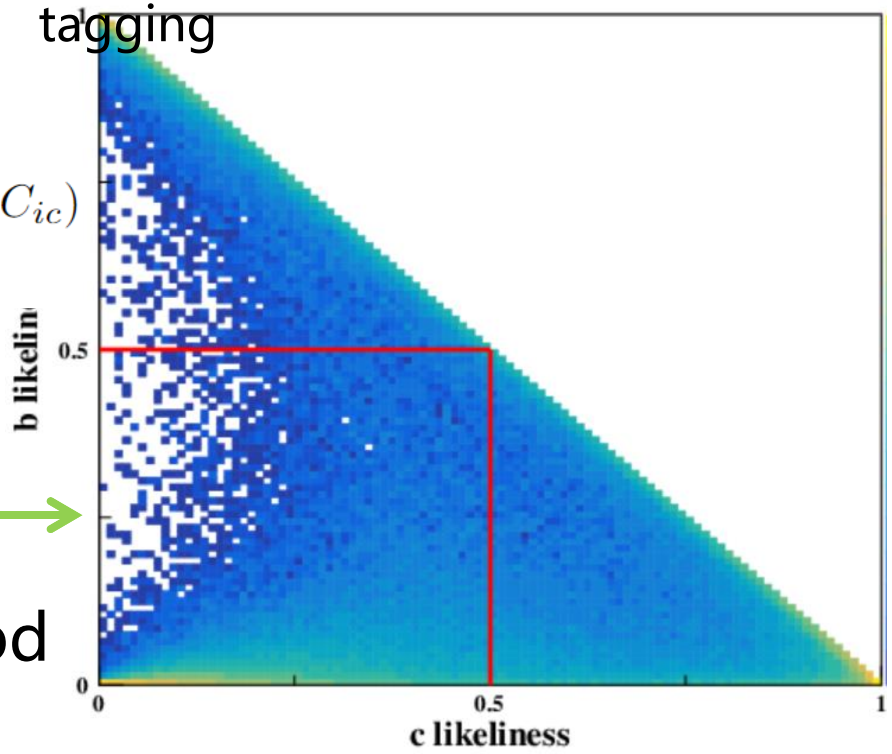
R_b & R_c measurement

$$N_s^{i,\mathrm{obs}} = 2N^{h,\mathrm{pro}} \cdot (R_b\varepsilon_{ib} + R_c\varepsilon_{ic} + R_o\varepsilon_{io}) \ ,$$

$$N_d^{i,\mathrm{obs}} = N^{h,\mathrm{pro}} \cdot [R_b\varepsilon_{ib}^2(1+C_{ib}) + R_c\varepsilon_{ic}^2(1+C_{ic}) + R_o\varepsilon_{io}^2(1+C_{io})] \ ,$$

- Double tagging:
  - Neglect the correlation of jets
  - Choose the working point
  - Solved 6 equations by the least square method

b likelihood vs. c likelihood in jet tagging



- Inverse matrix:
  - To solve the right equations
  - Calculate $R_q = \frac{N_q}{\Sigma_i N_i}$

$$\begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \\ N_3 \end{pmatrix}$$

R_b & R_c measurement

	σ_{R_b}	σ_{R_c}	σ_{R_o}
LEP+SLD	659	3015	-
Template fit	1.2	2.3	2.1
Double tag	1.3	1.4	-
Inverse matrix	1.4	1.4	-

- References
  - LEP+SLD:
    a) Limited by statistics & flavor tagging
  - Template fit:
    a) Much larger statistics & flavor tagging in CEPC baseline
- Results
  - Double tag & Inverse matrix:
    a) Statistic of 10^{11} Z bosons, same as template fit
    b) Comparable in R_b
    c) Improved more than 60% in R_c measurement

Systematic uncertainties of R_b & R_c measurement

- To be dominant in future ee colliders
  - Efficiency
    a) Arise by MC models
    b) Reduced by orders of magnitudes, since much improved knowledge on the production and decay of B & D mesons
  - Correlation between 2 jets
    a) Estimated from $2.7 \times 10^{-4} - 6.7 \times 10^{-3}$ at LEP/SLD
    b) Can be reduced when tagging efficiency is enhanced
    c) Disappear in inverse matrix method since no jet clustering
  - DL model

Summary of preliminary systematics study

- Reducing the systematics is very challenging
- MC should be tuned carefully to reduce systematics
- Systematics of DL architectures can be reduced by increasing sample statistics – comparable with data

~10M samples used for systematic study

	range	$\Delta R_b(10^{-4})$	$\Delta R_c(10^{-4})$
Momenta resolution($\delta p/p$)	10%	0.30	0.34
Detector coverage($\|\cos\theta\|$)	$0.994 - 0.995$	0.69	0.86
Impact parameters($\delta z/z, \delta r/r$)	1%	0.27	2.5
Number of EdgeConv blocks	2-4	2.4	3.9
Parameter of k-NN	14-18	2.8	4.5
Number of units of fully connected layer	226-286	1.2	5.4

Conclusion

- Two novel deep learning methods are used to enhance the performance of jet flavor tagging
  - Significant improvement in jet tagging, especially for c tagging
  - Maximize the usage of information in a jet/event
  - Strong inductive bias
- R_q measurement is taken to demonstrate the physics impacts
  - Statistical uncertainty improved 60+% in R_c measurement
  - Preliminary systematic uncertainties studies show that they are under control, but very challenging for these kind of precision measurements
  - Correlation between jets disappear in inverse matrix method

Thank you!

Comparison of the performance of ParticleNet with three alternative models

TABLE II: Performance comparison on the top tagging benchmark dataset. The ParticleNet, ParticleNet-Lite, P-CNN and ResNeXt-50 models are trained on the top tagging dataset starting from randomly initialized weights. For each model, the training is repeated for 9 times using different randomly initialized weights. The table shows the result from the median-accuracy training, and the standard deviation of the 9 trainings is quoted as the uncertainty to assess the stability to random weight initialization. Uncertainty on the accuracy and AUC are negligible and therefore omitted. The performance of PFN on this dataset is reported in Ref. [52], and the uncertainty corresponds to the spread in 10 trainings.

	Accuracy	AUC	$1/\varepsilon_b$ at $\varepsilon_s = 50\%$	$1/\varepsilon_b$ at $\varepsilon_s = 30\%$
ResNeXt-50	0.936	0.9837	302 ± 5	1147 ± 58
P-CNN	0.930	0.9803	201 ± 4	759 ± 24
PFN	-	0.9819	247 ± 3	888 ± 17
ParticleNet-Lite	0.937	0.9844	325 ± 5	1262 ± 49
ParticleNet	**0.940**	**0.9858**	$\mathbf{397 \pm 7}$	$\mathbf{1615 \pm 93}$