Measurement of R_b of Z hadronic decay

Hancen LU, Mohan PI, Zizi KANG, Bo LIU, Zhijun LIANG

Status:

✓ Samples: 2M for each flavor

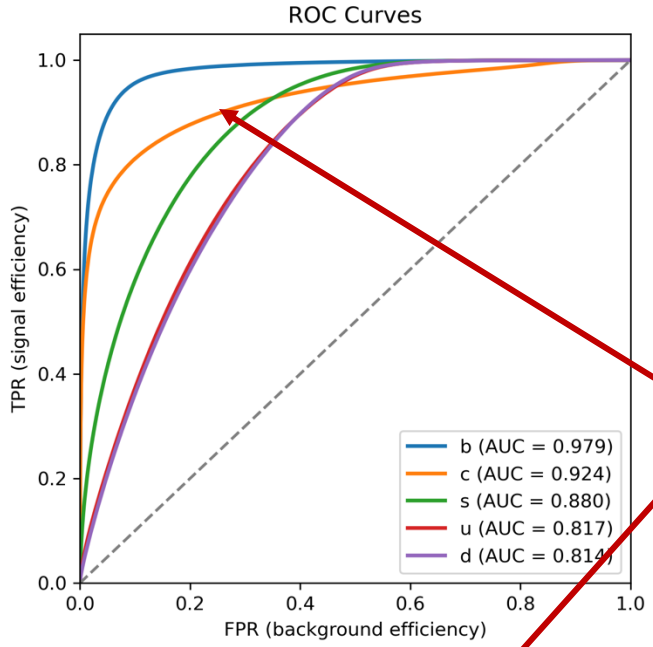
✓ Rb, Rc Calculation

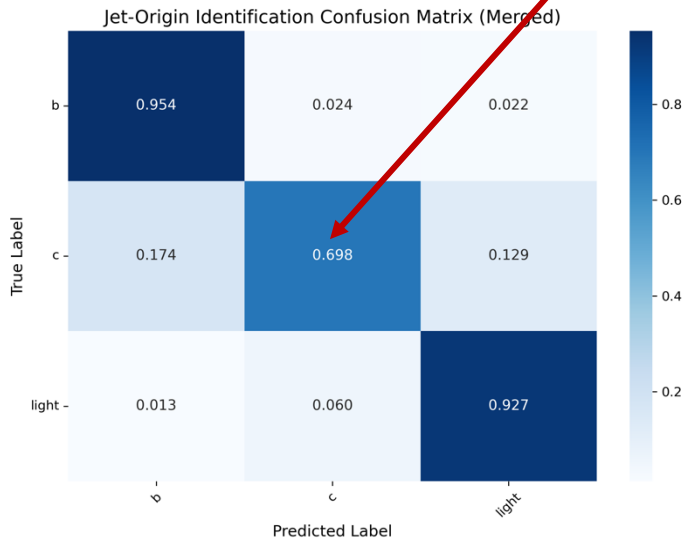
- Checking & improving results

	$\sigma_{R_b}(10^{-6})$	$\sigma_{R_c}(10^{-6})$	$\sigma_{R_q}(10^{-6})$	**flavor tagging method**
LEP+SLC	659	3015	-	-
FCC-ee	2.1(0.3)	-	-	-
Template fit	1.2	2.3	2.1	LCFIPlus
Double tag	1.3	1.4	-	Particle Net 10^{11} events
Double tag	51.8	66.4	-	ParT (PFOAna) 10^{11} events

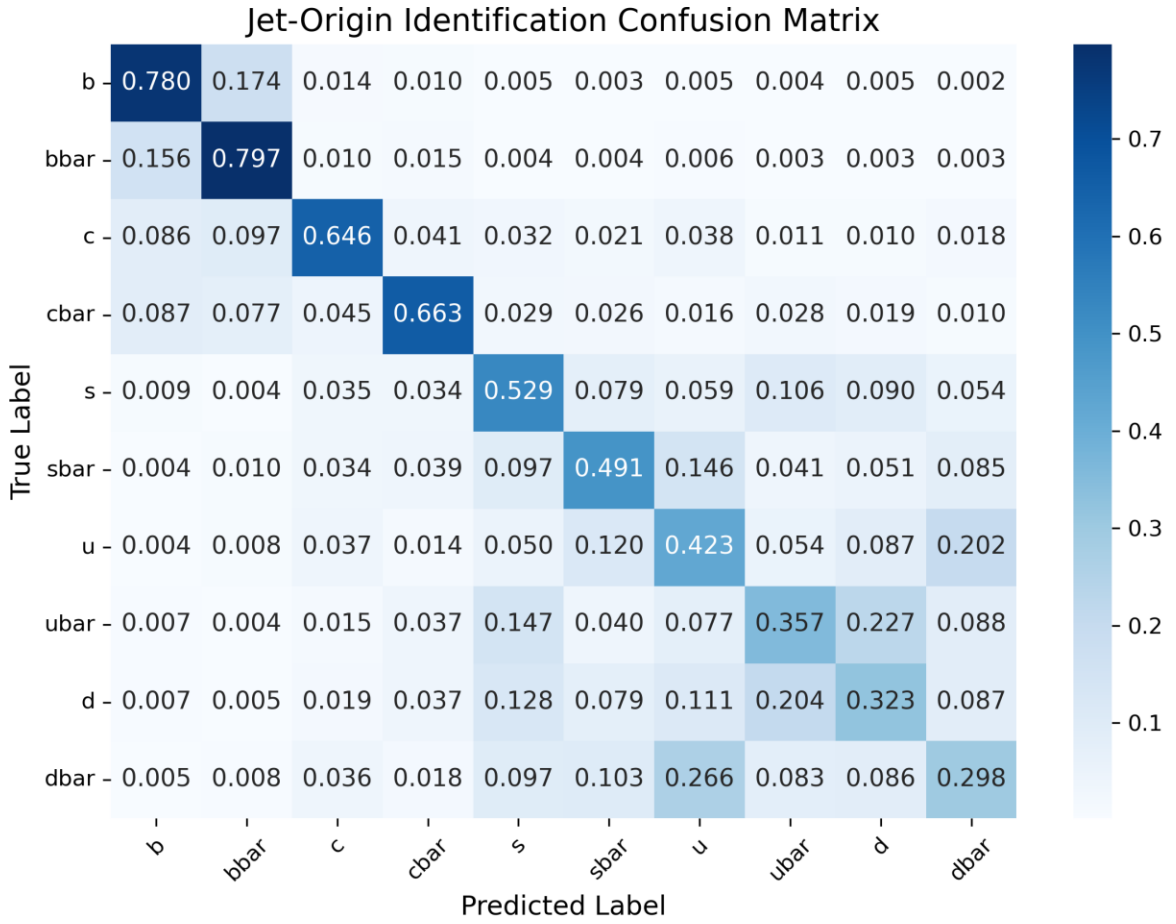
Lots of room for further improvement

Model performance





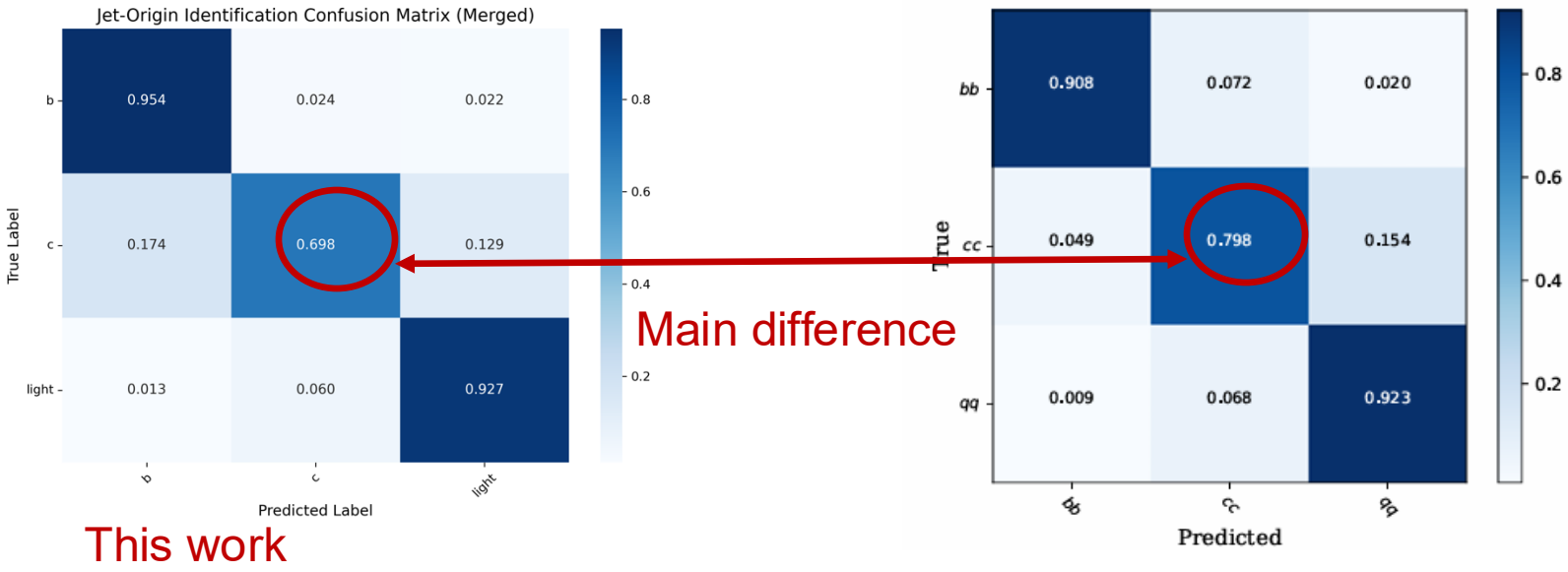
Strange



Comparing with Liao's study (arXiv:2208.13503v4)

	This work	arXiv:2208.13503v4
Vertex Geo	Baseline (Ladder + stitching)	Backup (Ladder)
Algorithm	Particle Transformer	Particle Net
Training Set Size	450k per channel (b c q)	1M per flavor Total 11M)
Training Set Type	Jets	Jets
Number of Samples	1e11 Z hadronic decay	1e11 Z hadronic decay
Method	Double Tagging	Double Tagging

Confusion Matrix



This work

Method: Double tagging

Parameters:

- $N^d_{i,\text{obs}}$: Events whose jets are tagged as flavor $i = b, c, q$.
- ϵ_{ij} : Efficiency matrix (3×3) from confusion matrix, $j = b, c, q$ is true label.
- $N_{h,\text{pro}}$: Total events of Z hadronic decay.

Considering $R_q = 1 - R_b - R_c$, the linear function is:

$$N^d_{i,\text{obs}} = N_{h,\text{pro}} \left[R_b \left(\epsilon^2_{ib} - \epsilon^2_{iq} \right) + R_c \left(\epsilon^2_{ic} - \epsilon^2_{iq} \right) + \epsilon^2_{iq} \right]$$

then

$$R_b \left(\epsilon^2_{ib} - \epsilon^2_{iq} \right) + R_c \left(\epsilon^2_{ic} - \epsilon^2_{iq} \right) = \frac{N^d_{i,\text{obs}}}{N_{h,\text{pro}}} - \epsilon^2_{iq}$$

With two variables and three functions, we can solve the equations with ***least square method***:

$$\begin{bmatrix} \hat{R}_b \\ \hat{R}_c \end{bmatrix} = \left(A^T A \right)^{-1} A^T \mathbf{b}$$

and Cov matrix:

$$\text{Cov}\left(\hat{R}_b, \hat{R}_c \right) = \sigma^2 \left(A^T A \right)^{-1}$$

where

$$\sigma^2 = \frac{\| \mathbf{b} - A\hat{\mathbf{R}} \|^2}{3 - 2}$$

Scale

$$\sigma_{\text{scaled}} = \frac{\sigma_{\text{original}}}{\sqrt{k}}, \quad k = \frac{N_{h,\text{obs, scaled}}}{N_{h,\text{obs, original}}}$$

Result

calc scripts:
`/hpcfs/cepc/higgsgpu/luhancen/RbCalc`

	$\sigma_{R_b}(10^{-6})$	$\sigma_{R_c}(10^{-6})$	$\sigma_{R_q}(10^{-6})$	**flavor tagging method**
LEP+SLC	659	3015	-	-
FCC-ee	2.1(0.3)	-	-	-
Template fit	1.2	2.3	2.1	LCFIPlus
Double tag	1.3	1.4	-	Particle Net 10^{11} events
Double tag	51.8	66.4	-	ParT (PFOAna) 10^{11} events

The statistical uncertainties (10^{-6}) we calculate are based on the least square method.

Original result (5×10^6 events):
$R_b = 22.808 \pm 0.733\%$
$R_c = 15.434 \pm 0.939\%$

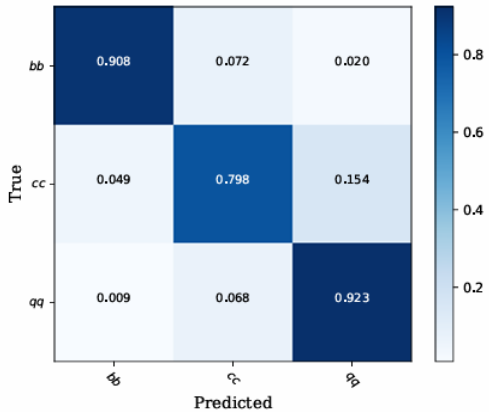
Scaled result (10^{11} events):
$R_b = 22.808 \pm 0.00518\%$
$R_c = 15.434 \pm 0.00664\%$

Items	**Entries**
$b\bar{b}$ pairs	1025097
$c\bar{c}$ pairs	432679
$q\bar{q}$ pairs	2640431
Unmatched pairs	901794
Total events	5000001

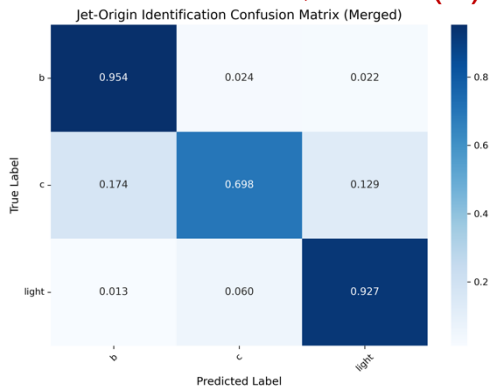
Events selected:

Channel	**Entires**	**Ratio**
Zbb	1096570	21.93 %
Zcc	854779	17.10 %
Zdd	1096990	21.94 %
Zss	1096510	21.93 %
Zuu	855152	17.10 %

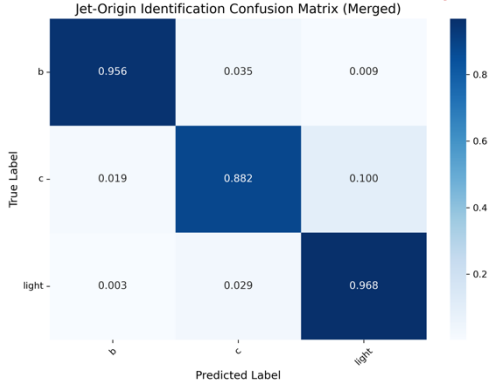
Confusion Matrix, PN



Confusion Matrix, ParT (Z)



Confusion Matrix, ParT (Higgs)



ToyMC Result:

Alg	Total Events	Ave. Rb	Ave. Rc	Std. Rb	Std. Rc
PN	5e6	22.243	16.014	0.646	0.732
ParT (Z)	5e6	22.833	15.455	0.865	1.108
ParT (Higgs)	5e6	22.058	16.646	0.251	0.272
ParT (Higgs)	2e7	22.060	16.643	0.249	0.269
ParT (Full Sim Z)	5e6	22.808	15.434	0.733	0.939
Ref	-	21.93	17.10	-	-

Some Conclusion:

- Confusion Matrix contributes the majority of the errors
- These errors are from least square method, having nothing to do with number of total events

Possible issues:

- Merging categories
- Convert Higgs (240GeV) to Z (91.2GeV)
- Least square method

Todo:

- Optimizing tagging algorithm
- Figuring out what leads to the bias of mean value
- Trying to reduce the statistical error
- Considering removing the limits Rb + Rc + Rq = 1