The $H \to c\bar{c}$ accuracy measurement at CEPC

Yongfeng Zhu

advisor : Manqi Ruan

IHEP

2021 年 6 月 10 日

Motivation

- The $H \to c\bar{c}$ decay is the only promising channel to investigate the Higgs boson couplings to the second generation fermions.
- The g(Hcc), the second generation fermion Yukawa coupling, is one of the most important benchmark at CEPC.

decay mode	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$	$Z \to q\bar{q}$	$Z \to \nu\nu$	$Z \to l^+l^-$
branching ratio	57.7%	2.91%	8.57%	70%	20%	10%

Contents

- The cut chain for $qqHc\bar{c}$ accuracy measurement.
- The dependence of current accuracy.
  - color singlet identification (including jet clustering and jet matching)
  - flavor tagging performance.
- The cut chain for $\nu\nu Hc\bar{c}$ accuracy measurement.
- summary

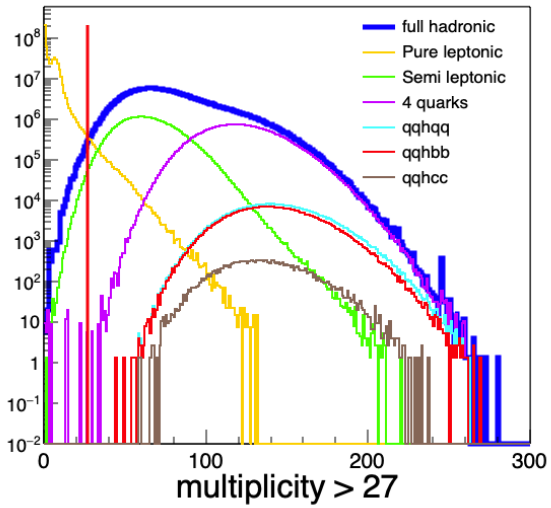
Sample :

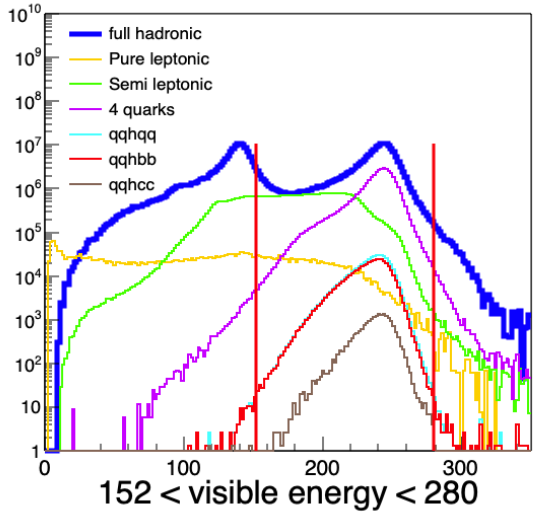
signal : $qqHc\bar{c}$ bkg : all SM backgrounds at CEPC with $5600 fb^{-1}$

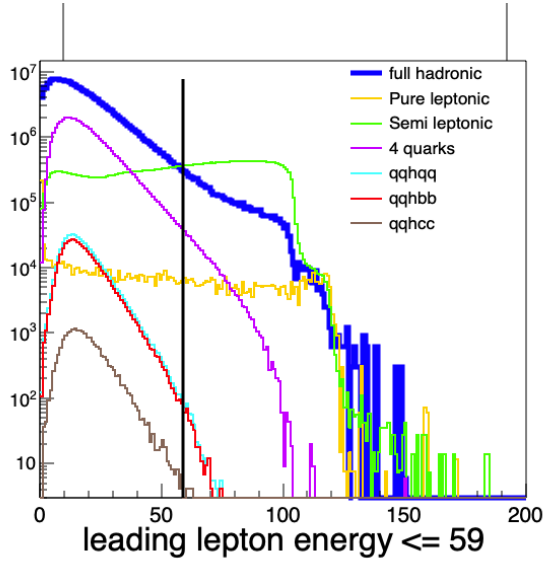
The cut chain method is used in this analysis, it has four steps.

1. Finding the full hadronic samples from all samples.
2. Finding 4-quark samples from the full hadronic samples.
3. Finding $ZH(Z \to q\bar{q}, H \to q\bar{q})$ from 4-quark samples.
4. Finding $qqHc\bar{c}$ from $ZH(Z \to q\bar{q}, H \to q\bar{q})$.

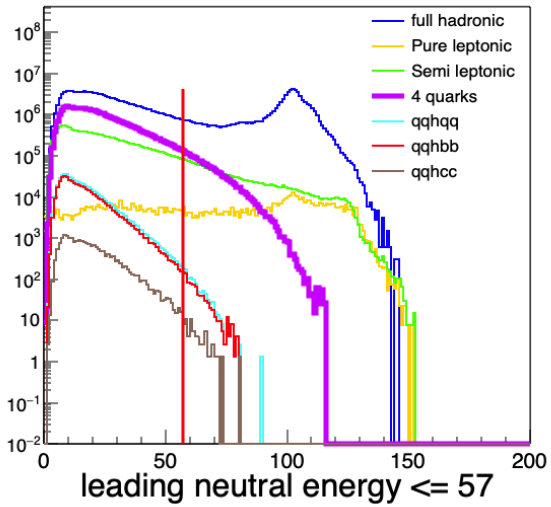
Firstly, finding the full hadronic samples from all samples.

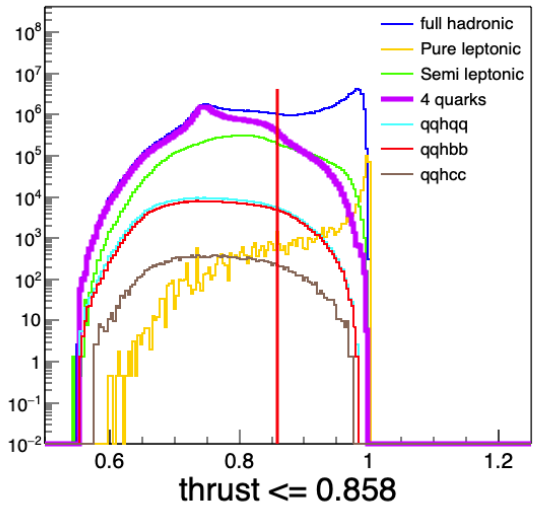


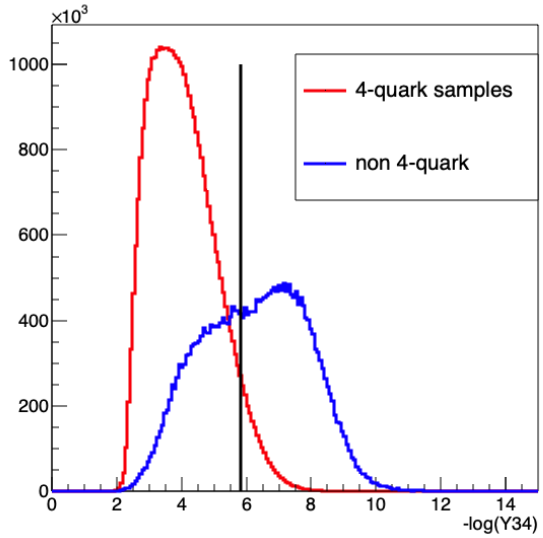




Secondly, finding 4-quark samples from the full hadronic samples.

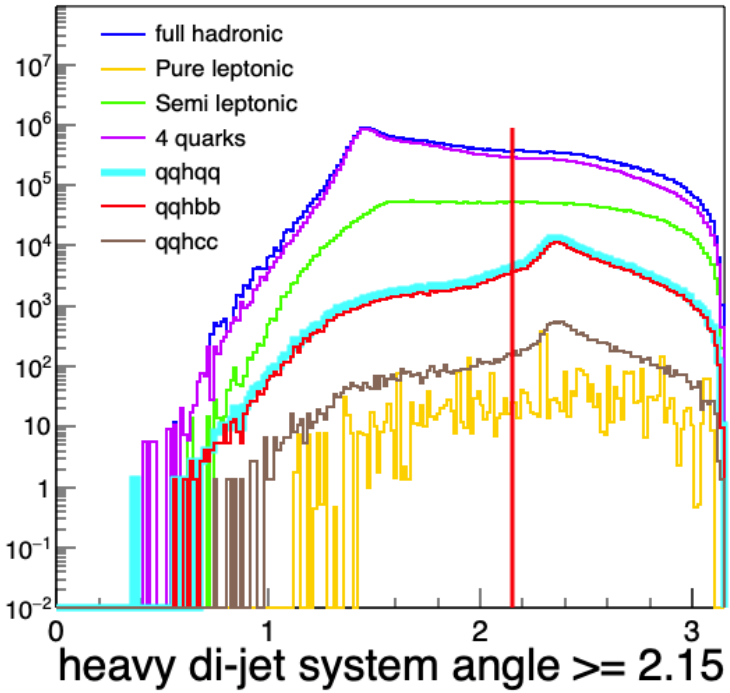


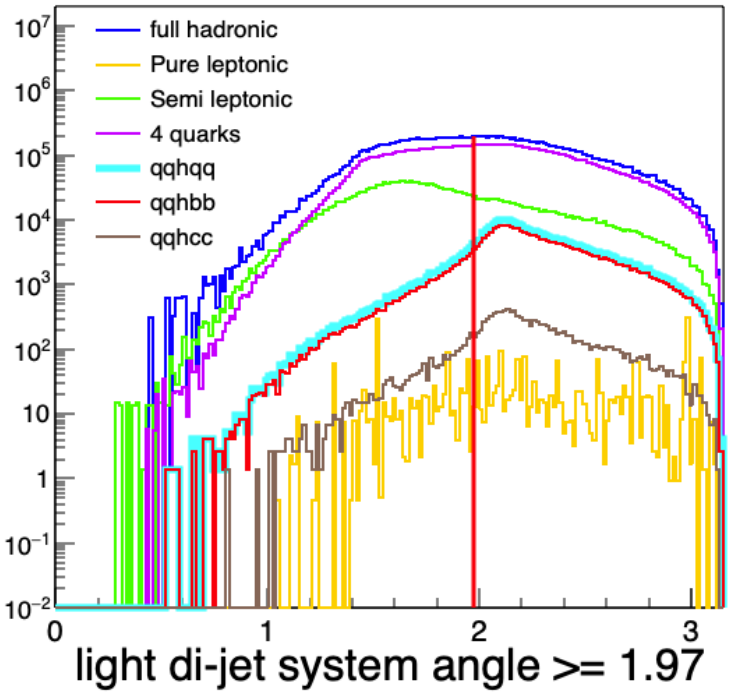




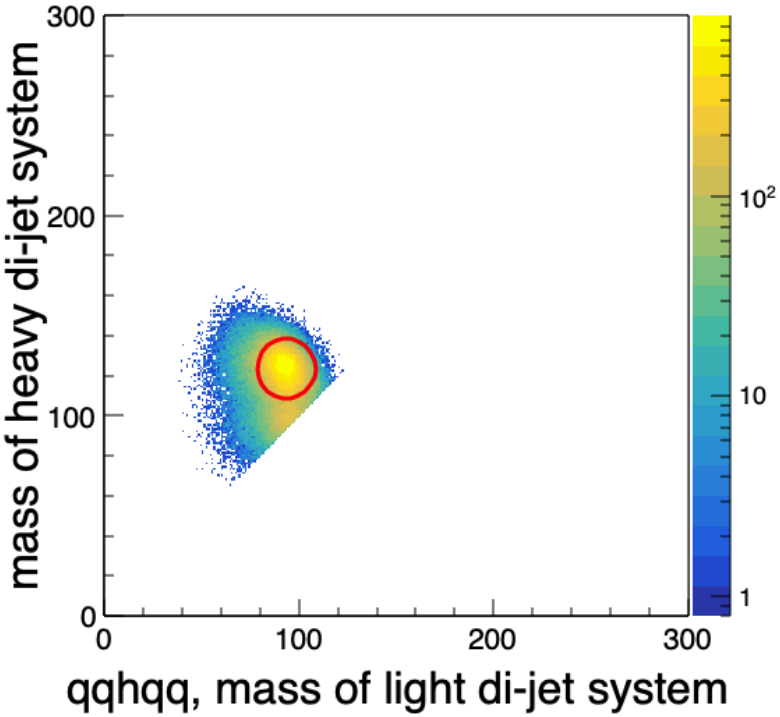
Thirdly, finding $ZH(Z \to q\bar{q}, H \to q\bar{q})$ from 4-quark samples.

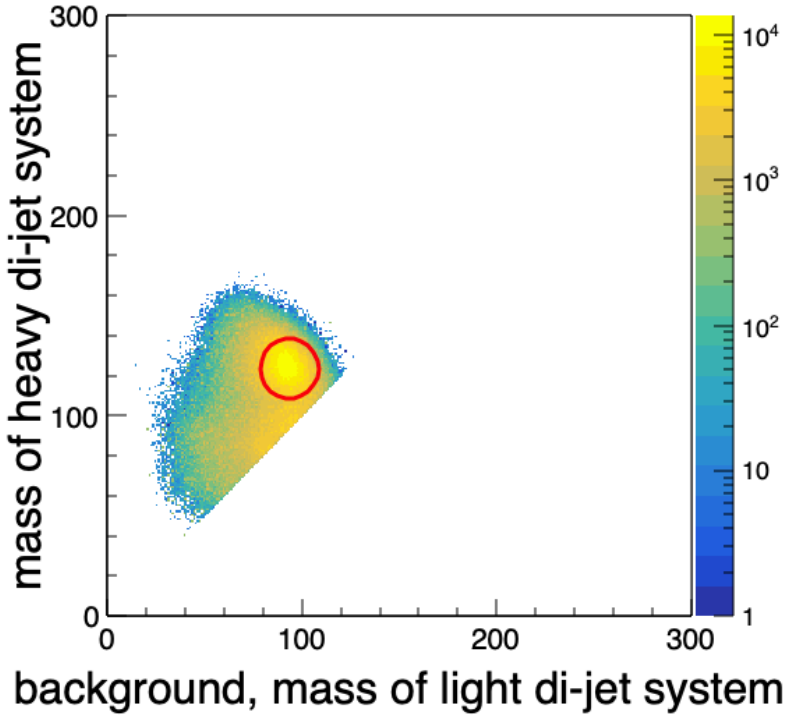
After finding 4-quark samples, the method of maximize $\chi^2 = \frac{(M_{12}-M_{B1})^2}{\sigma_{B1}^2} + \frac{(M_{34}-M_{B2})^2}{\sigma_{B2}^2}$ can be used to pair four jets into two di-jet systems.



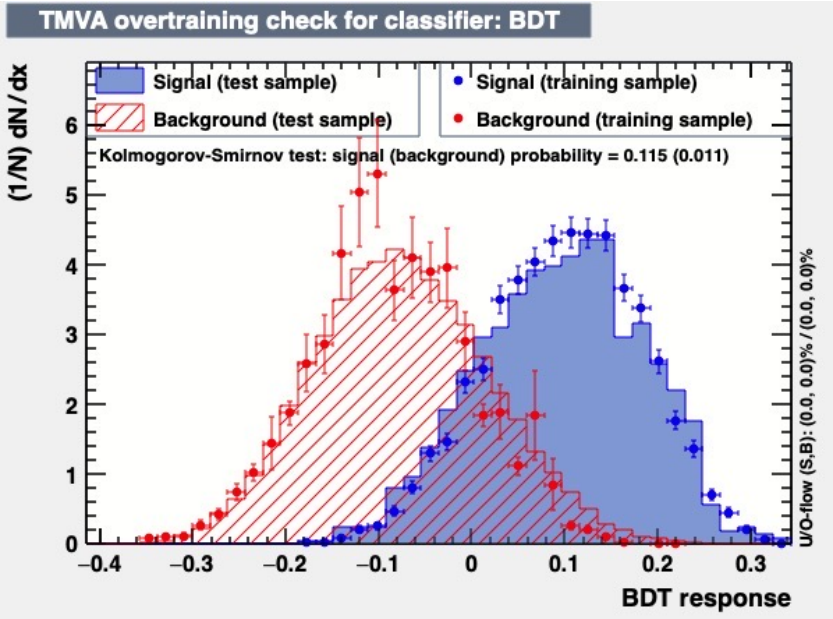


Then a circle can be used to find ZH events.





Finally, finding $qqHc\bar{c}$ from $ZH(Z \to q\bar{q}, H \to q\bar{q})$.



the full cut chain shown in the following table :

	qqHcc	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{sqrtS+B}{S}$
total	20461	8.01*E*8	1.95*E*7	9.07*E*6	5.08*E*7	6.39*E*6	2.18*E*7	1.12*E*6	9.10*E*8	1.47428
multiplicity	20461	3.04*E*8	1.46*E*7	3.37*E*6	4.85*E*7	6.00*E*6	1.81*E*7	1.08*E*6	3.96*E*8	0.972244
visEn	20456	1.54*E*8	1.30*E*7	1.66*E*6	4.00*E*7	4.25*E*6	1.80*E*7	8.28*E*5	2.32*E*8	0.745363
LLepEn	20431	1.50*E*8	5.16*E*6	8.01*E*5	3.09*E*7	3.66*E*6	1.78*E*7	7.82*E*5	2.09*E*8	0.707026
LNeuEn	20363	7.81*E*7	4,29*E*6	2.48*E*5	2.96*E*7	3,51*E*6	1.72*E*7	7.76*E*5	1.34*E*8	0.567833
thrust	17749	1.76*E*7	3.49*E*6	1.76*E*5	2.58*E*7	2.99*E*6	1.58*E*7	6.92*E*5	6.65*E*7	0.45947
$-log(Y_{34})$	17320	7.07*E*6	1.30*E*6	146*E*5	2.03*E*7	2.66*E*6	1.48*E*7	6.74*E*5	4.70*E*7	0.395851
HJetA	12897	3.74*E*6	5.71*E*5	74874	6.20*E*6	1.07*E*6	4.16*E*6	467006	1.63*E*7	0.313
ZJetA	10867	1.60*E*6	1.67*E*5	44807	2.97*E*6	606051	2.22*E*6	377305	7.99*E*6	0.260208
circle	8232	623811	4828	19847	1.52*E*6	263460	1.27E06	228194	3.92*E*6	0.240869
BDT	2905	18336	0	15	9590	7561	18318	1850	58577	0.08535

ILC : https://link.springer.com/article/10.1140/epjc/s10052-013-2343-8

for $qqH\bar{b}\bar{b}$

	qqHbb	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{\sqrt{S+B}}{S}$
BDT	127482	50425	0	22	5653	37532	4938	5793	104367	0.00377706

for *qqHgg*

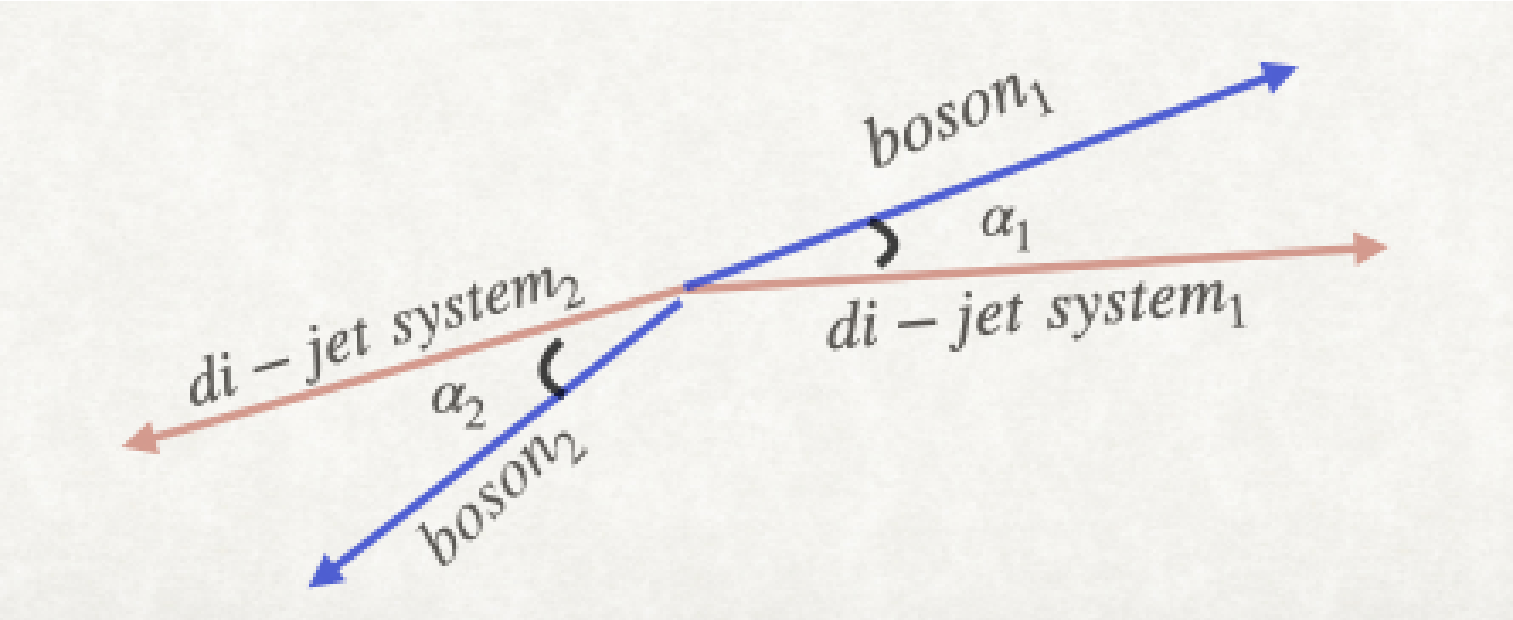
	qqHgg	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{\sqrt{S+B}}{S}$
BDT	15664	101768	0	37	144995	32607	128464	22816	430688	0.0426498

The dependence of the current accuracy :

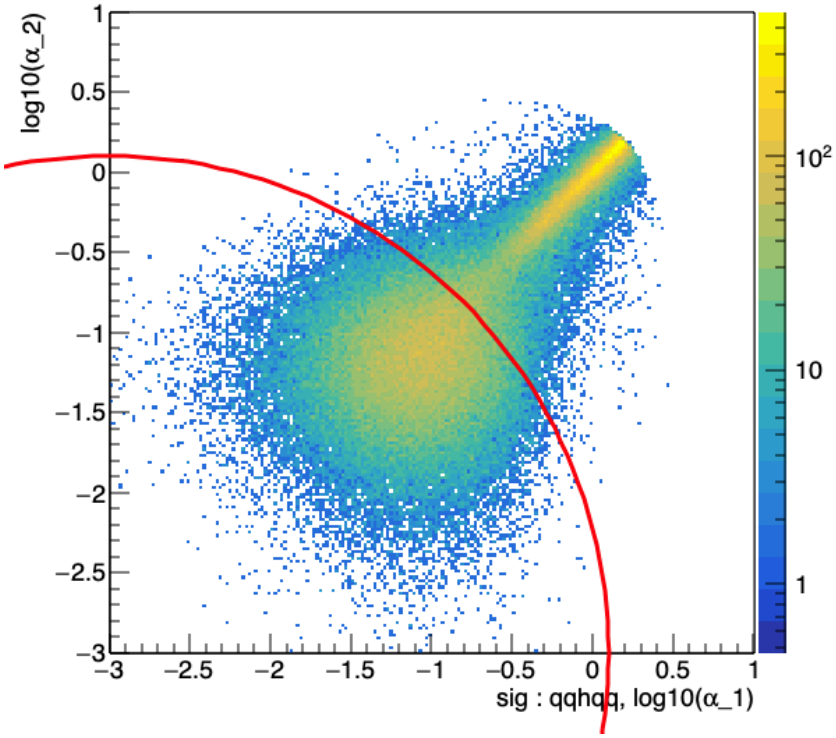
- color singlet identification (including jet clustering and jet matching)
- flavor tagging performance
- others have not been analyzed

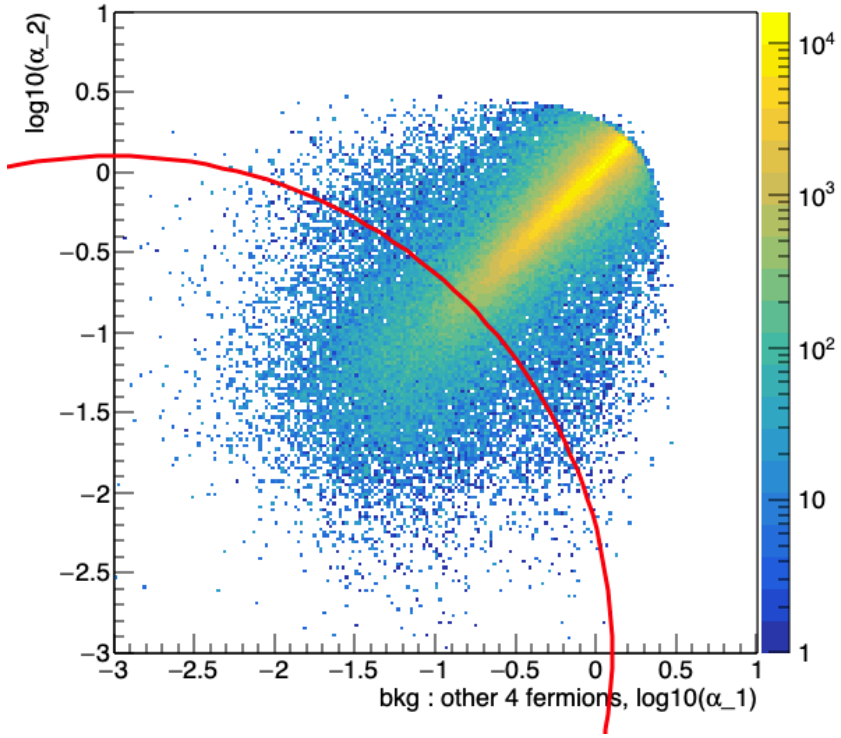
color singlet identification

For di-boson event, there are two MC truth bosons and two di-jet systems, the variable $\alpha_i = angle(di-jet\ system_i, truth\ boson_i), (i = 1, 2)$ is used to characterize the performance of jet clustering and jet matching.



the α variable is just a cheated variable used to characterize the performance of color singlet identification





the red curve : $(log10(\alpha_1)+3)^2+(log10(\alpha_2)+3)^2 <= 3.1^2$ used to select events with good color singlet identification

α detail :

`https://link.springer.com/article/10.1140/epjc/s10052-019-6719-2`

	qqHcc	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{sqrtS+B}{S}$
circle	8232	623811	4828	19847	1.52*E*6	263460	1.27E06	228194	3.92*E*6	0.240869
alpha	5401	15168	117	2437	80477	34510	72854	119181	309579	0.1064

1. With alpha cut $(log10(\alpha_1)+3)^2+(log10(\alpha_2)+3)^2 <= 3.1^2$, the $qqHc\bar{c}$ measurement accuracy can reach 10.64% even though without flavor tagging algorithm.
2. With optimized flavor tagging performance, the accuracy can reach 5.49%.

flavor tagging

introduce the flavor tagging performance matrix :

eff \ to / true	c	b	udsg
udsg	udsg to c	udsg to b	udsg to udsg
b	b to c	b to b	b to udsg
c	c to c	c to b	c to udsg

to : identified as

	c	b	udsg
udsg	0	0	1
b	0	1	0
c	1	0	0

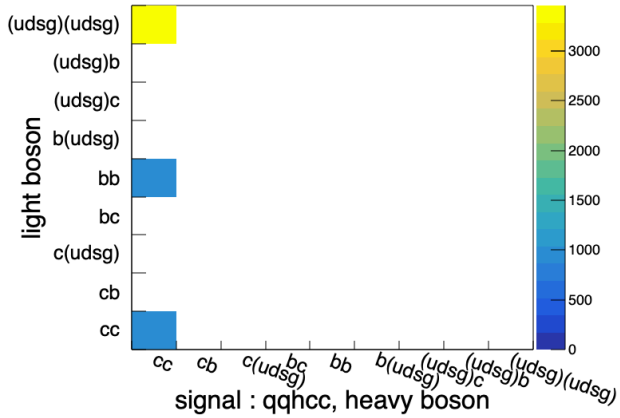
perfect flavor tagging

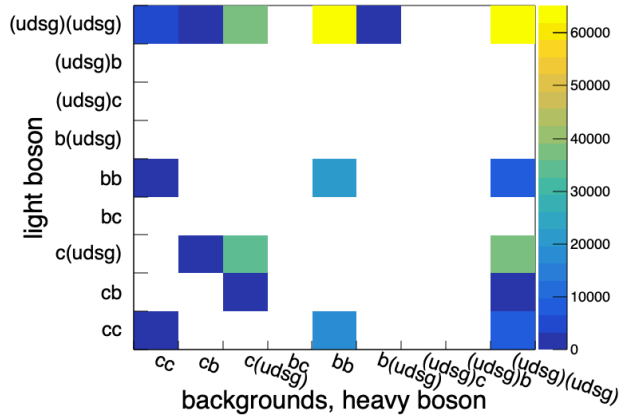
	c	b	udsg
udsg	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
b	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
c	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

non flavor tagging

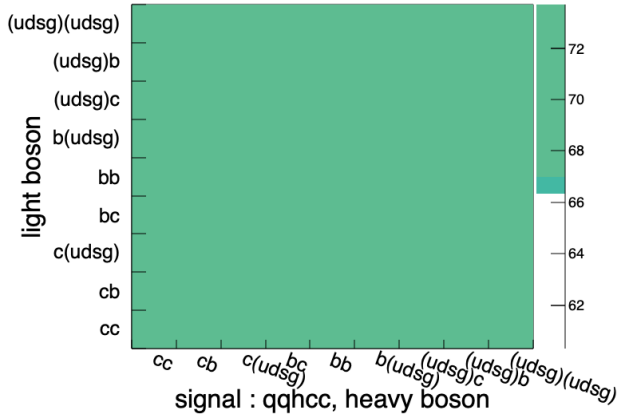
signal : $qqHc\bar{c}$ after alpha cut background : 4-quark samples after alpha cut

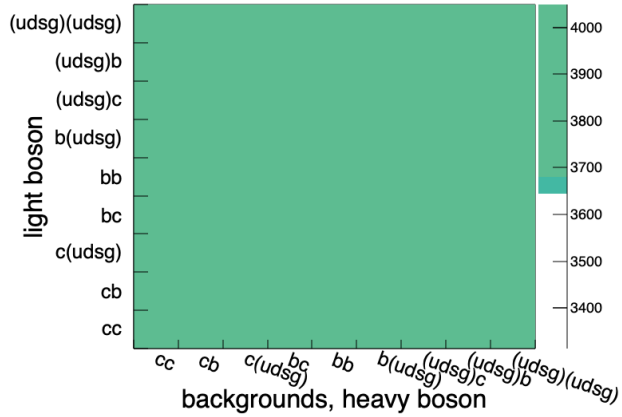
perfect flavor tagging





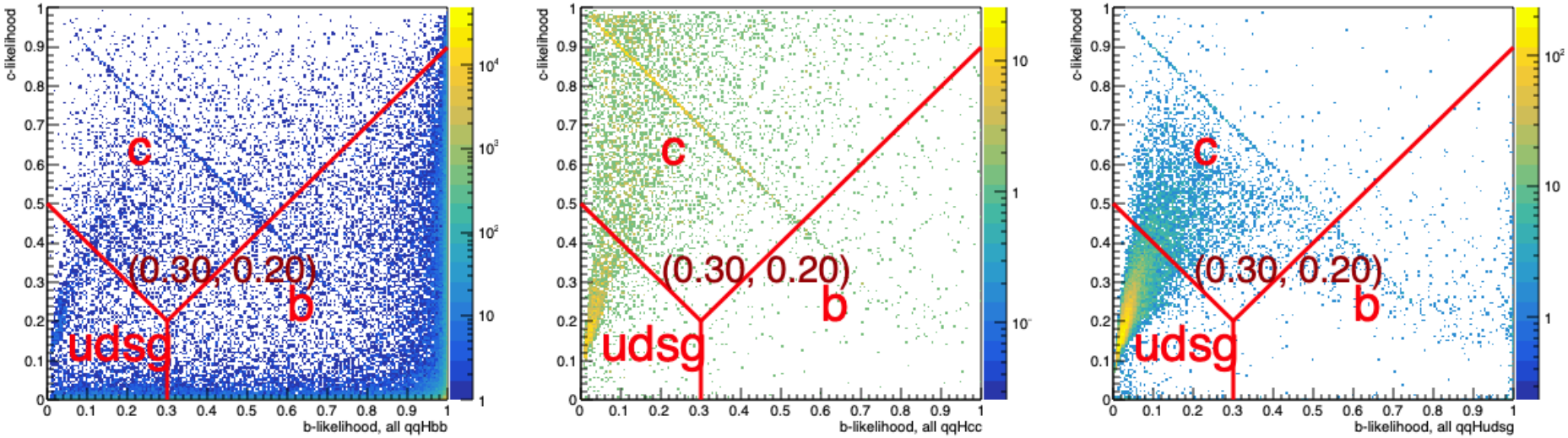
non flavor tagging

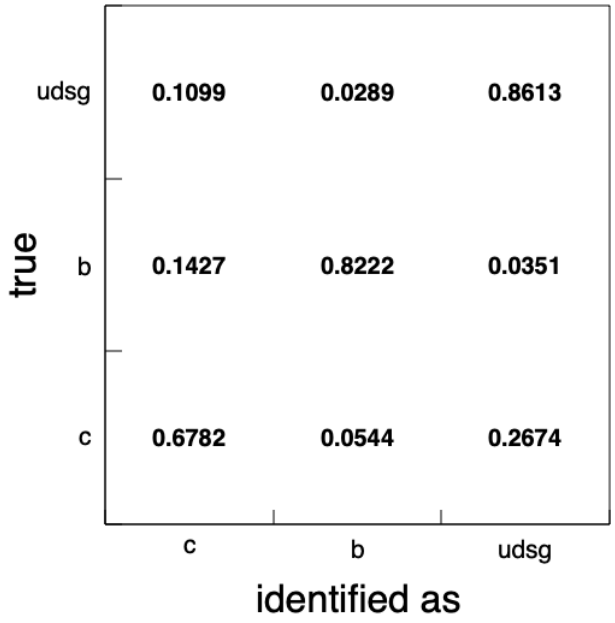


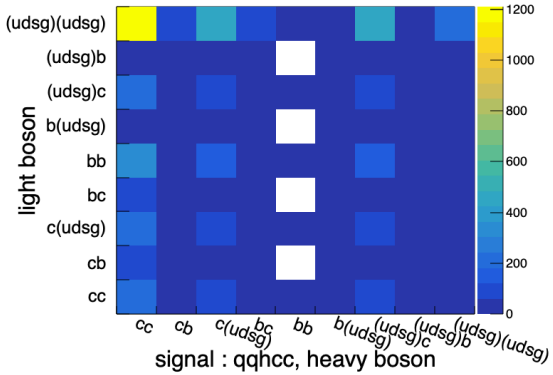


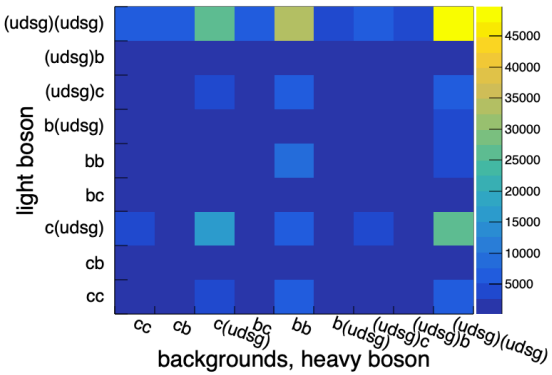
Optimized matrix

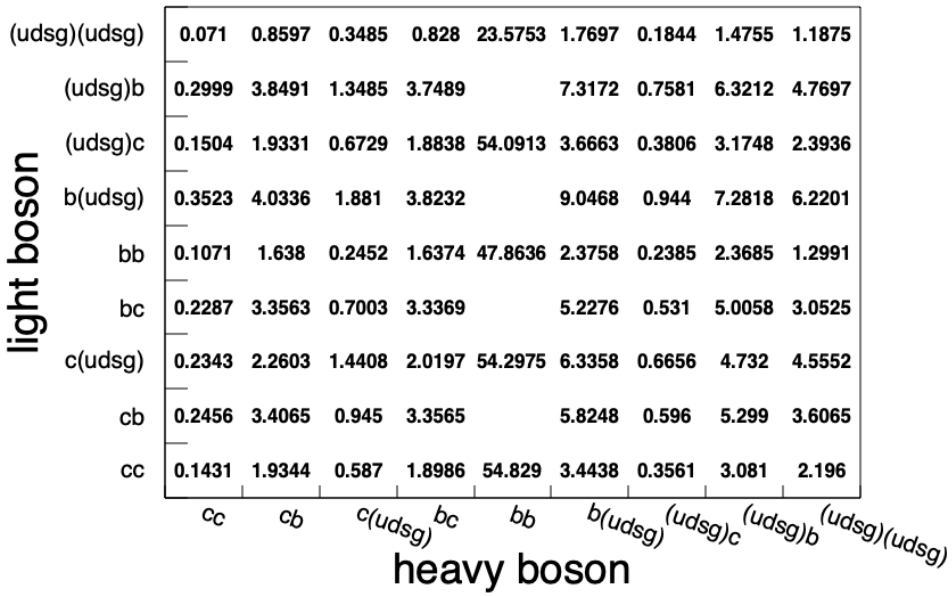
1. Suppose that the alpha cut has exclude the events with bad jet clustering and bad jet matching.
2. Then the b-likeness and c-likeness of two jets from heavy di-jet system, corresponds to Higgs for $qqHq\bar{q}$, can be displaced in 2D graph.
3. The cut on b-likeness and c-likeness can be find to maximize the value of $eff(b \to b) + eff(c \to c) + eff(udsg \to udsg)$, the trace of flavor tagging matrix.











- accuracy : $\frac{\sqrt{S+B}}{S}$
- accuracy for $qqHc\bar{c}$: 5.49%
- accuracy for every bin
- combined accuracy : $\frac{bin_i \cdot bin_j}{\sqrt{bin_i \cdot bin_i + bin_j \cdot bin_j}}$ iterate for each pair of bins

changing flavor tagging performance :

$$non\ flavor\ tagging \rightarrow perfect\ flavor\ tagging$$

the changing procedure of flavor tagging performance matrix :

$$temp\ matrix = \frac{x - trace_I}{trace_T - trace_I} \cdot (T - I) + I \qquad (trace_I \leq x \leq trace_T)$$

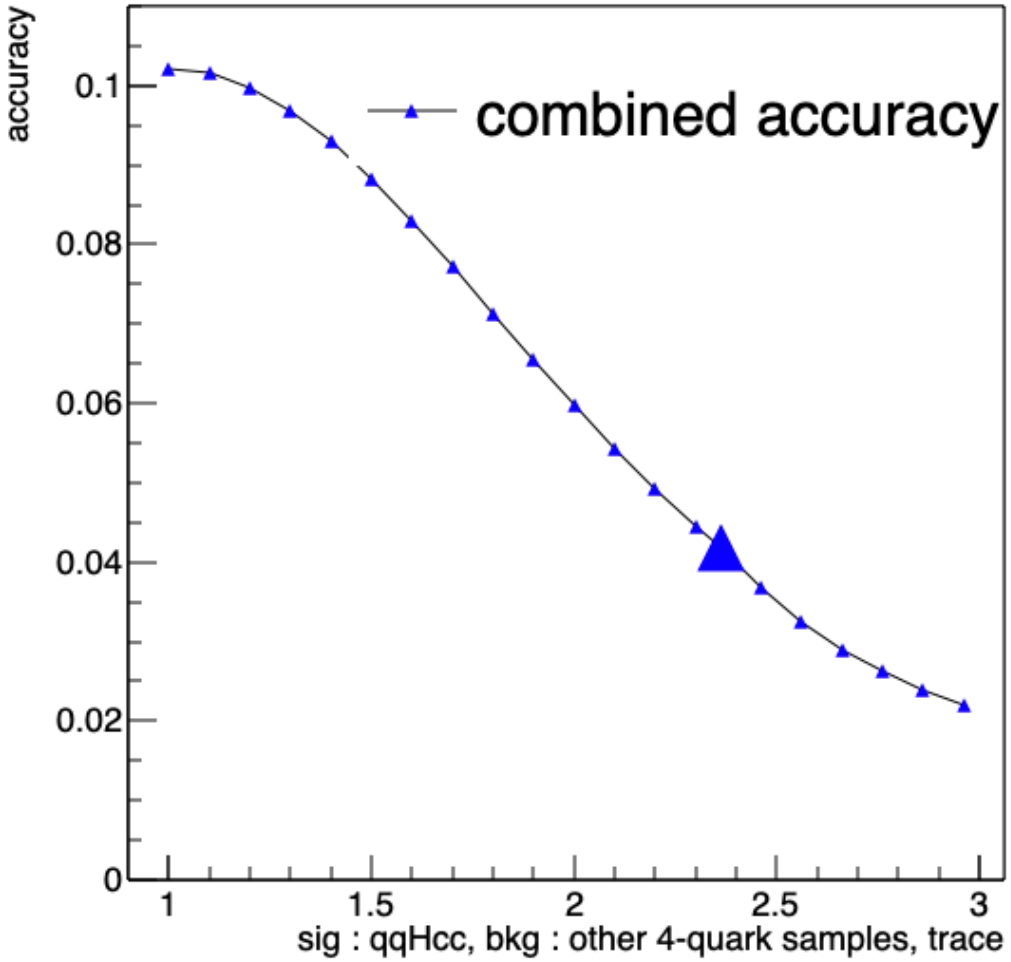
T : matrix with perfect flavor tagging

I : matrix with non flavor tagging

$trace_I, trace_T$: the trace of matrix I and T

The x value and flavor tagging performance matrix have a one to one relation.

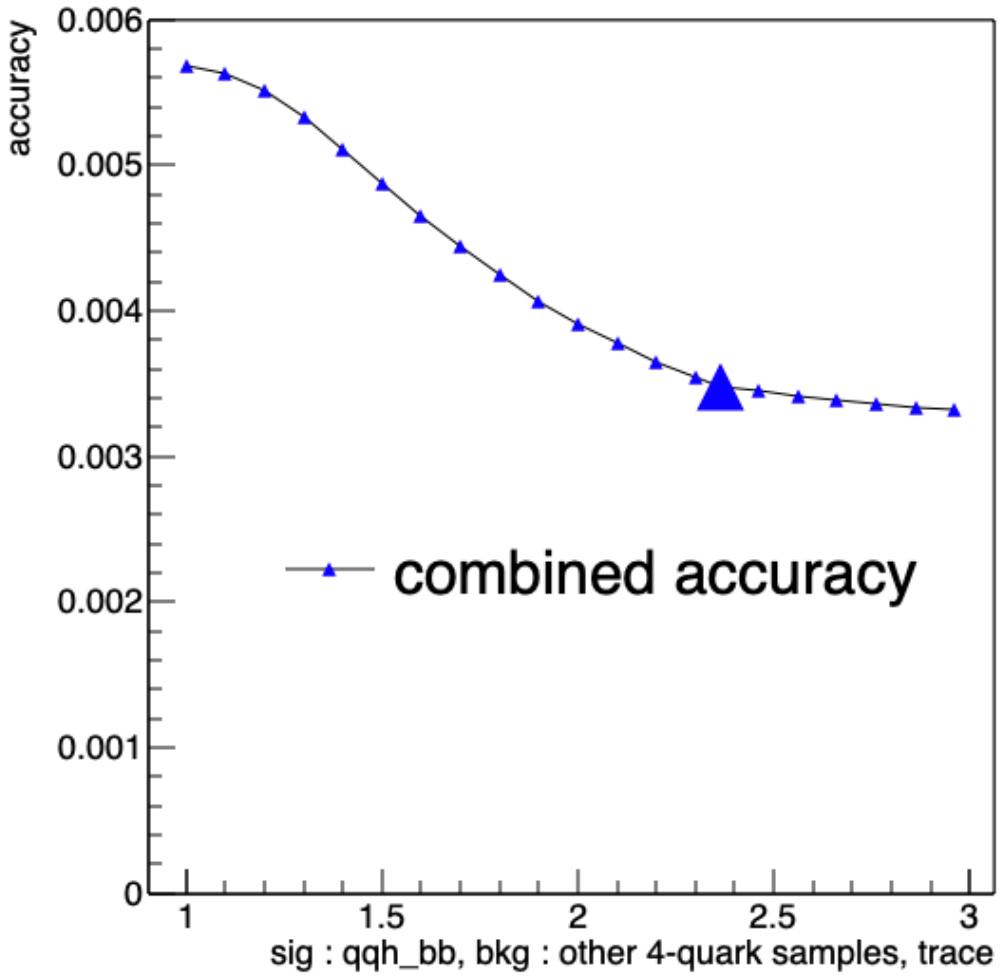
the variation of $qqHc\bar{c}$ measurement accuracy with trace



with good color singlet identification : 5.49%

with perfect flavor tagging : 2.14%

the variation of $qqHb\bar{b}$ measurement accuracy with trace



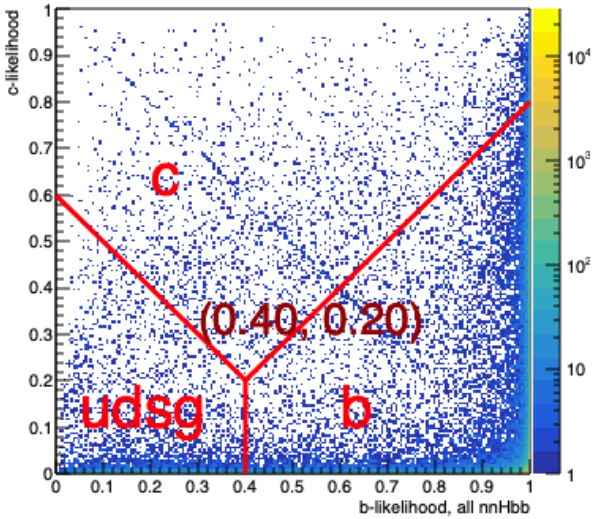
with good color singlet identification : 0.37%

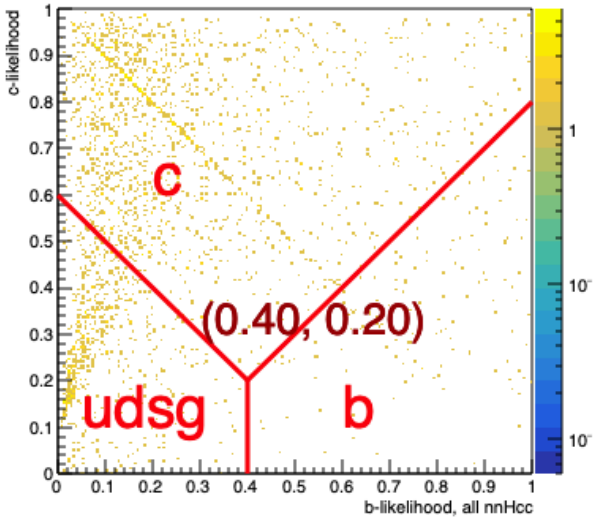
with perfect flavor tagging : 0.33%

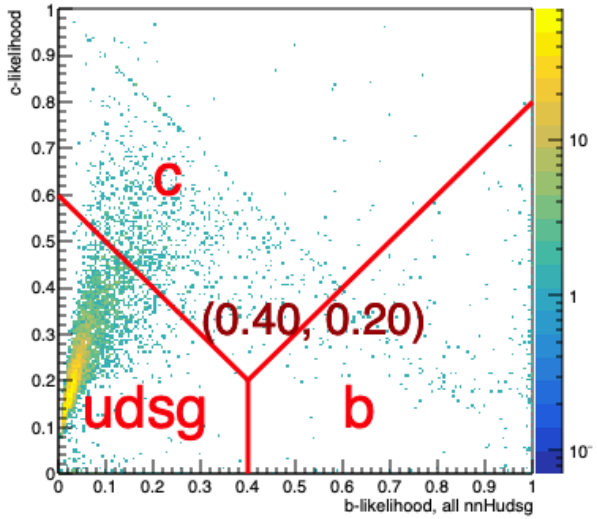
$\nu\nu Hc\bar{c}$ accuracy measurement

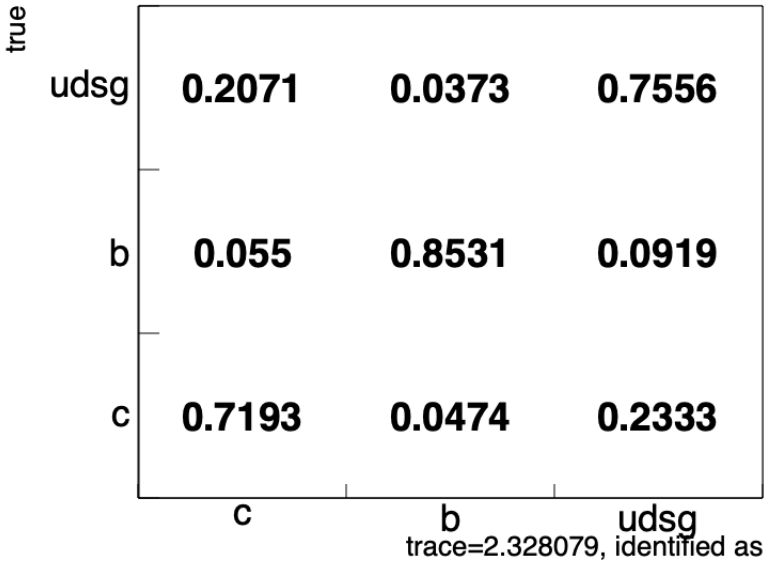
For nnHcc accuracy measurement, finding nnHqq firstly, then get its accuracy with the method of optimized flavor tagging matrix.

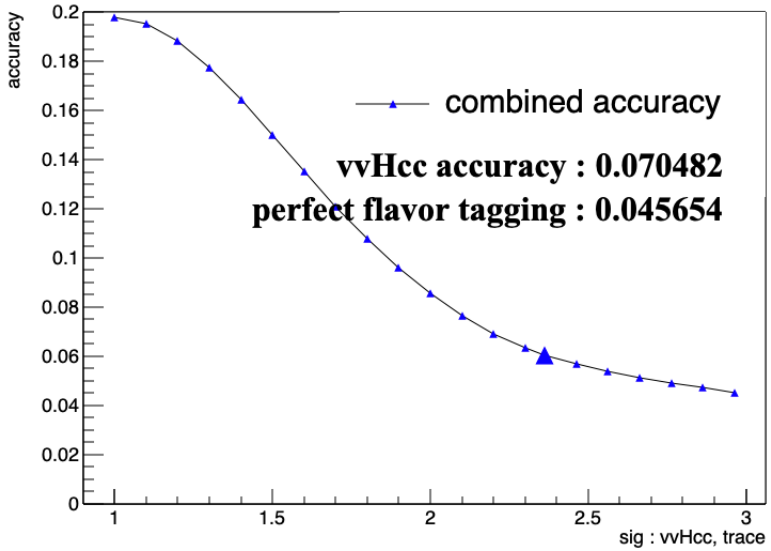
	nnHqq	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{sqrtS+B}{S}$
total	178890	8.01*E*8	1.95*E*7	9.07*E*6	5.08*E*7	6.39*E*6	2.18*E*7	961606	9.10*E*8	0.168624
recoilMass	157822	5.11*E*7	2.17*E*6	1.38*E*6	4.78*E*6	1.30*E*6	1.08*E*6	74991	6.19*E*7	0.0499366
visEn	142918	2.37*E*7	1.35*E*6	8.81*E*5	3.60*E*6	1.03*E*6	6.29*E*5	50989	3.13*E*7	0.0392443
leadLepEn	141926	2.08*E*7	3.65*E*5	7.24*E*5	2.81*E*6	9.72*E*5	1.34*E*5	46963	2.59*E*7	0.0359349
Npfo	139545	1.66*E*7	2.36*E*5	5.24*E*5	2.62*E*6	9.07*E*5	4977	42751	2.09*E*7	0.0329016
LeadNeuEn	138653	1.46*E*7	2.24*E*5	4.72*E*5	2.49*E*6	8.69*E*5	4552	42303	1.86*E*7	0.0312835
Pt	121212	248715	1.56*E*5	2.48*E*5	1.51*E*6	4.31*E*5	999	35453	2.63*E*6	0.0136992
Pl	118109	53680	1.08*E*5	74936	729604	1.14*E*5	789	34279	1.11*E*6	0.00940262
Y23	82035	38771	13084	50195	109007	65822	633	4040	281554	0.00735026
Pmax	78135	31835	6639	45952	87414	59800	334	3747	235724	0.00716999
MassDif	76739	31108	6264	44853	79715	57687	320	3306	223256	0.00713734
bThrust	75822	31016	5948	43599	73986	55312	320	2769	212953	0.00708735
InvMass	68348	24551	3169	6994	38567	11634	198	1882	86998	0.00576667
BDT	63462	13811	45	1854	1132	2913	33	686	20476	0.00456526

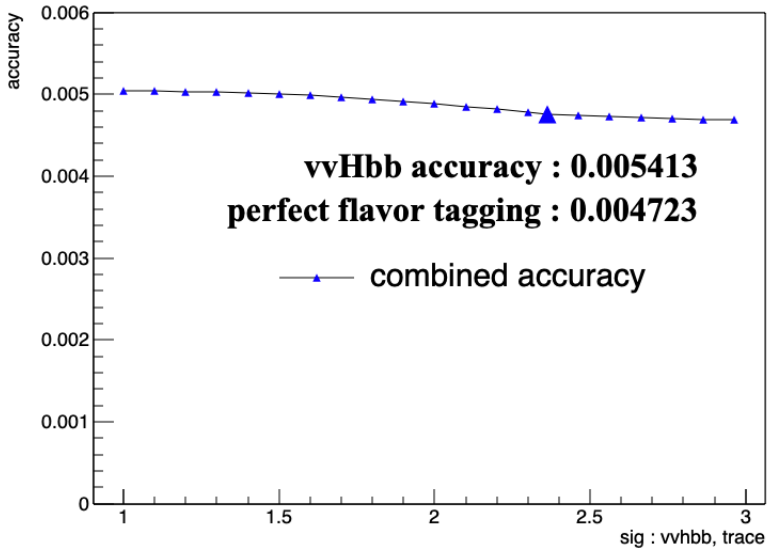


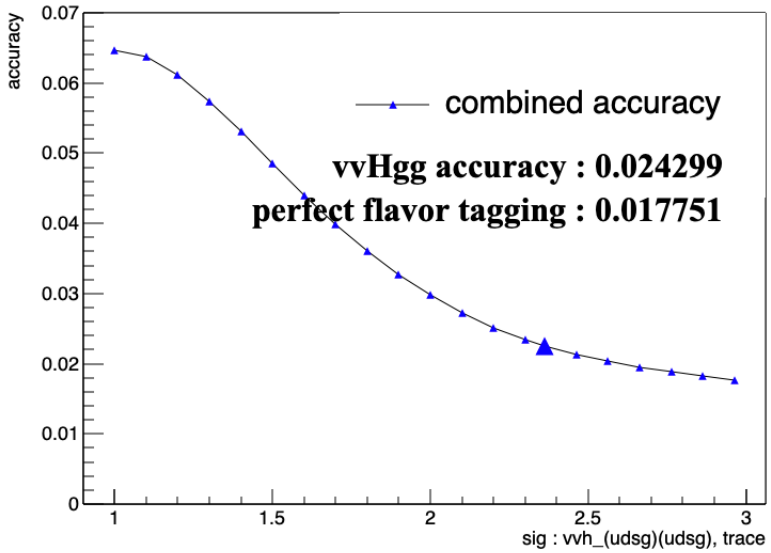





true
udsg
b
c
0.2071
0.0373
0.7556
0.055
0.8531
0.0919
0.7193
0.0474
0.2333
c
b
udsg
trace=2.328079, identified as


accuracy
combined accuracy
vvHcc accuracy : 0.070482
perfect flavor tagging : 0.045654
sig : vvHcc, trace


accuracy
vvHbb accuracy : 0.005413
perfect flavor tagging : 0.004723
combined accuracy
sig : vvhbb, trace


accuracy
combined accuracy
vvHgg accuracy : 0.024299
perfect flavor tagging : 0.017751
sig : vvh_(udsg)(udsg), trace

comparition

		$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$
$Z \to q\bar{q}$	CEPC ($5600 fb^{-1}$)	0.38%	8.53%	4.26%
	ILC ($250 fb^{-1}$, P(e^-, e^+)=(-0.8, +0.3))	1.5%	10.2%	13.1%
$Z \to \nu\nu$	CEPC ($5600 fb^{-1}$)	0.54%	7.05%	2.43%
	ILC ($250 fb^{-1}$, P(e^-, e^+)=(-0.8, +0.3))	1.7%	11.2%	13.9%

ILC : https://link.springer.com/article/10.1140/epjc/s10052-013-2343-8

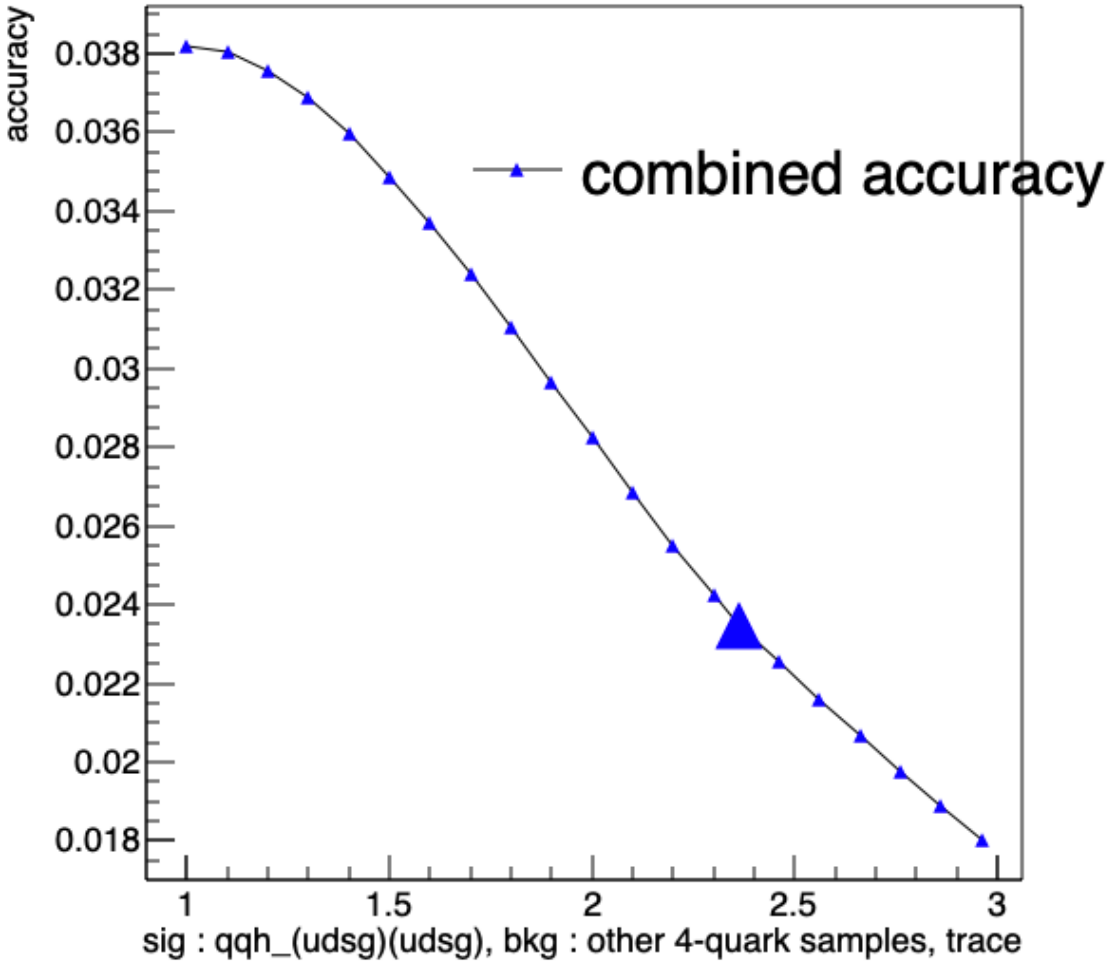
Summary :

- At present, the $qqHc\bar{c}$ measurement accuracy is 8.54%, and $\nu\nu Hc\bar{c}$ is 7.05%.
- The color singlet identification, including jet clustering, jet matching, and flavor tagging, have great impact on $qqHc\bar{c}$ accuracy measurement.
  - With good color singlet identification, the $qqHc\bar{c}$ measurement accuracy can reach 5.49%.
  - With good color singlet identification and perfect flavor tagging performance, the $qqHc\bar{c}$ measurement accuracy can reach 2.14%.
- Next, we need to optimize the cut chain and find a reconstructed variable which can describe the color singlet identification, instead of the variable α.

Thanks !

Backup

the variation of *qqHgg* measurement accuracy with trace



with perfect jet clustering and jet matching : 2.97%

with perfect flavor tagging : 2.36%