The $qqHc\bar{c}$ accuracy measurement at CEPC

Yongfeng Zhu

advisor : Manqi Ruan

IHEP

2021 年 6 月 2 日

Motivation

- The g(Hcc), the second generation fermion Yukawa coupling, is one of the most important benchmark at CEPC.
- The channel $qqH(H \to c\bar{c})$ has more significant statistics than that of $nnH(H \to c\bar{c})$.
- The color singlet identification has great impact on the g(Hcc) measurement.

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

- When CEPC operates at the $\sqrt{s} = 240\ GeV$ with integrated luminosity of $5600 fb^{-1}$, the $qqH(H \to c\bar{c})$ measurement accuracy($\frac{\sqrt{S+B}}{S}$) is 8.54%.
- The dependence of the current accuracy: color singlet identification.
    - With perfect jet clustering and jet matching, the $qqH(H \to c\bar{c})$ measurement accuracy can reach 5.49%.
    - With perfect flavor tagging algorithm, the $qqH(H \to c\bar{c})$ measurement accuracy can reach 2.14%.

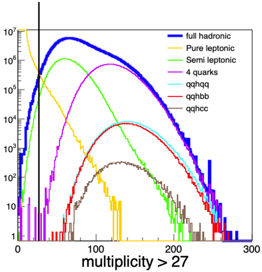
Sample :

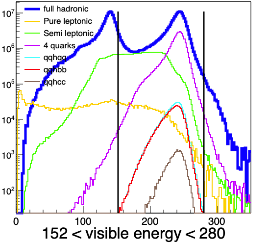
signal : $qqHc\bar{c}$ bkg : all SM backgrounds at CEPC with $5600fb^{-1}$

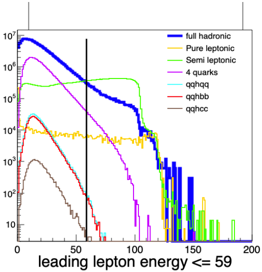
The cut chain method is used in this analysis, it has four steps.

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 $ZH(Z \to q\bar{q}, H \to q\bar{q})$.

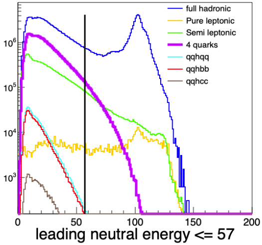
Firstly, finding the full hadronic samples from all samples.

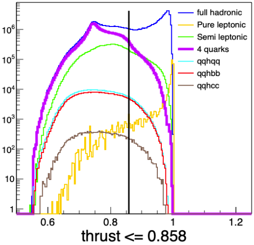


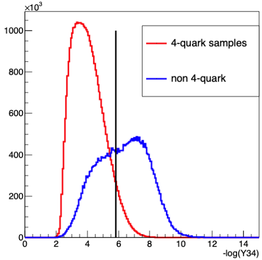




Secondly, finding 4-quark samples from the full hadronic samples.

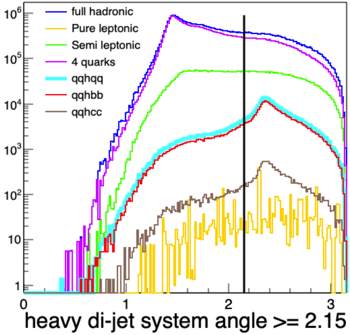


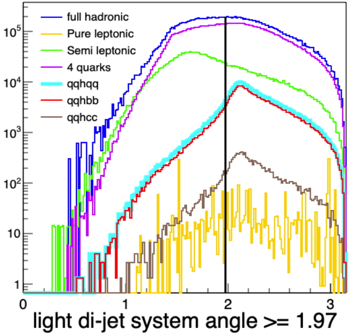




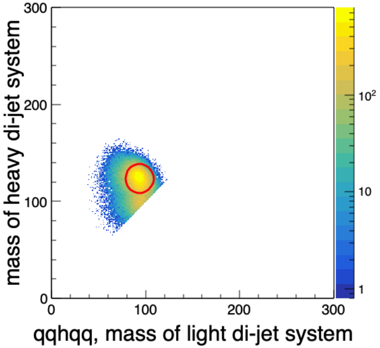
Thirdly, finding $ZH(Z \to q\bar{q}, H \to q\bar{q})$ from 4-quark samples.

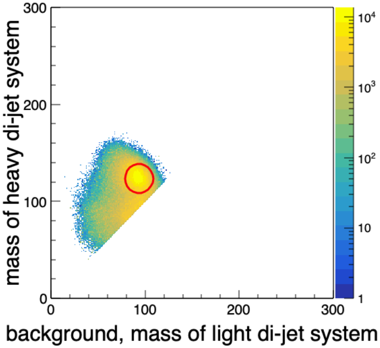
After finding 4-quark samples, the method of maximize $\chi^2 = \frac{(M_{12}-M_{B1})^2}{\sigma_{B1}^2} + \frac{(M_{34}-M_{B2})^2}{\sigma_{B2}^2}$ can be used to pair four jets into two di-jet systems.



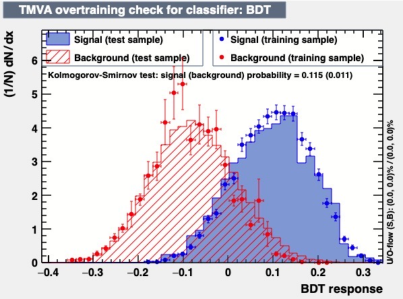


Then a circle can be used to find ZH events.





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	$8.01E8$	$1.95E7$	$9.07E6$	$5.08E7$	$6.39E6$	$2.18E7$	$1.12E6$	$9.10E8$	1.47428
multiplicity	20461	$3.04E8$	$1.46E7$	$3.37E6$	$4.85E7$	$6.00E6$	$1.81E7$	$1.08E6$	$3.96E8$	0.972244
visEn	20456	$1.54E8$	$1.30E7$	$1.66E6$	$4.00E7$	$4.25E6$	$1.80E7$	$8.28E5$	$2.32E8$	0.745363
LLepEn	20431	$1.50E8$	$5.16E6$	$8.01E5$	$3.09E7$	$3.66E6$	$1.78E7$	$7.82E5$	$2.09E8$	0.707026
LNeuEn	20363	$7.81E7$	$4,29E6$	$2.48E5$	$2.96E7$	$3,51E6$	$1.72E7$	$7.76E5$	$1.34E8$	0.567833
thrust	17749	$1.76E7$	$3.49E6$	$1.76E5$	$2.58E7$	$2.99E6$	$1.58E7$	$6.92E5$	$6.65E7$	0.45947
$-log(Y_{34})$	17320	$7.07E6$	$1.30E6$	$146E5$	$2.03E7$	$2.66E6$	$1.48E7$	$6.74E5$	$4.70E7$	0.395851
HJetA	12897	$3.74E6$	$5.71E5$	74874	$6.20E6$	$1.07E6$	$4.16E6$	467006	$1.63E7$	0.313
ZJetA	10867	$1.60E6$	$1.67E5$	44807	$2.97E6$	606051	$2.22E6$	377305	$7.99E6$	0.260208
circle	8232	623811	4828	19847	$1.52E6$	263460	1.27E06	228194	$3.92E6$	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 $qqHb\bar{b}$

	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 $qqHgg$

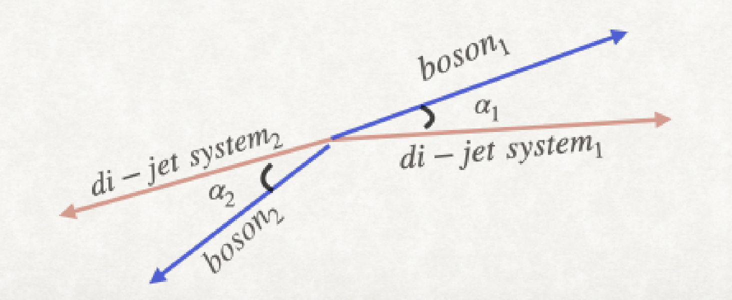
	qqHgg	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{\sqrt{S+B}}{S}$
BDT	15807	102685	0	22	145749	33732	128757	23612	434558	0.0424535

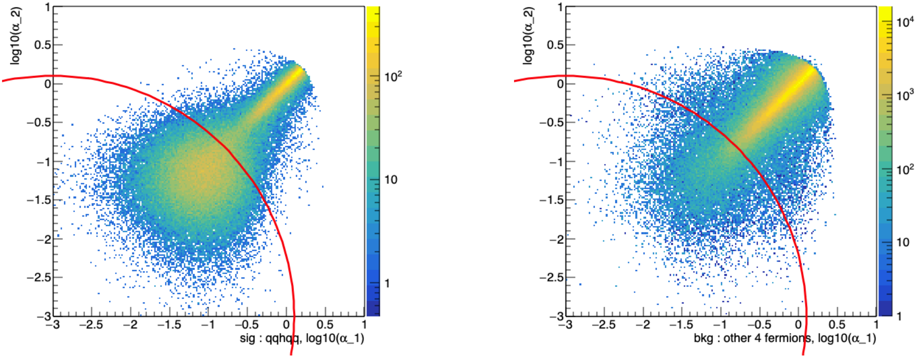
The dependence of the current accuracy : color singlet identification

- jet clustering and jet matching
- flavor tagging performance
- others have not been analyzed

jet clustering and jet matching

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 red curve : $.(log10(\alpha_1)+3)^2+(log10(\alpha_2)+3)^2 <= 3.1^2$ used to select events with good jet clustering and jet matching

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 0.1064 even though without flavor tagging algorithm.
2. With optimized flavor tagging performance, the accuracy can reach 0.0548824.

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

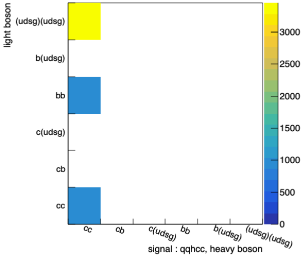
perfect flavor tagging

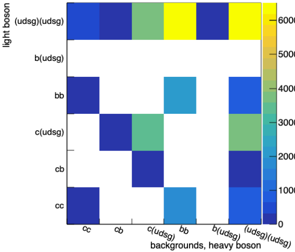
	c	b	udsg
udsg	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
b	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
c	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

non flavor tagging

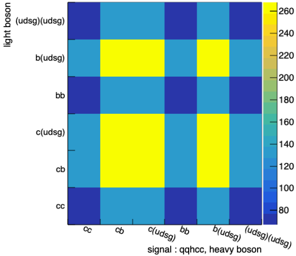
signal : $qqHc\bar{c}$ after alpha cut background : 4-quark samples after alpha cut

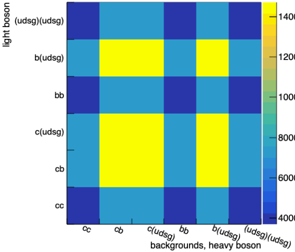
perfect flavor tagging





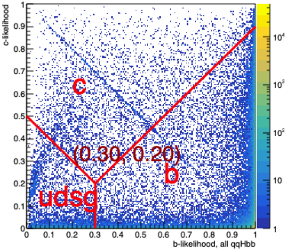
non flavor tagging

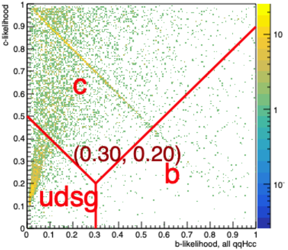


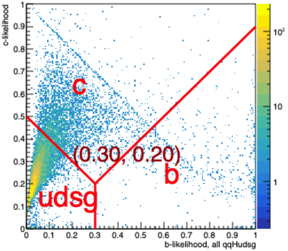


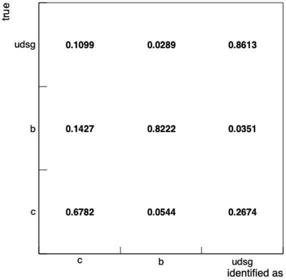
Optimized matrix

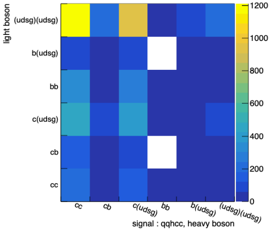
1. Suppose that the alpha cut has exclude the events with bad jet clustering and bad jet matching.
2. Then the b-likeliness and c-likeliness of two jets from heavy di-jet system, corresponds to Higgs for $qqHq\bar{q}$, can be displaced in 2D graph.
3. The cut on b-likeliness and c-likeliness can be find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

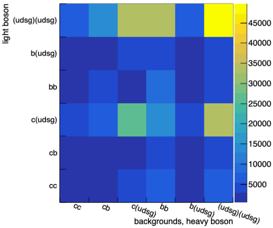


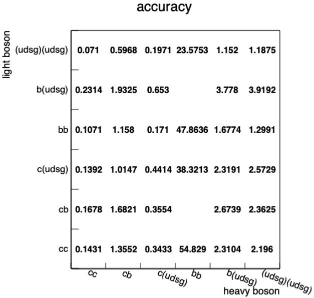












- accuracy : $\frac{\sqrt{S+B}}{S}$
- accuracy for $qqHc\bar{c}$: 5.49%
- accuracy for every bin
- combined accuracy : $\frac{bin_i \cdot bin_j}{\sqrt{bin_i \cdot bin_i + bin_j \cdot bin_j}}$ iterate for each pair of bins

changing flavor tagging performance :

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

the changing procedure of flavor tagging performance matrix :

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

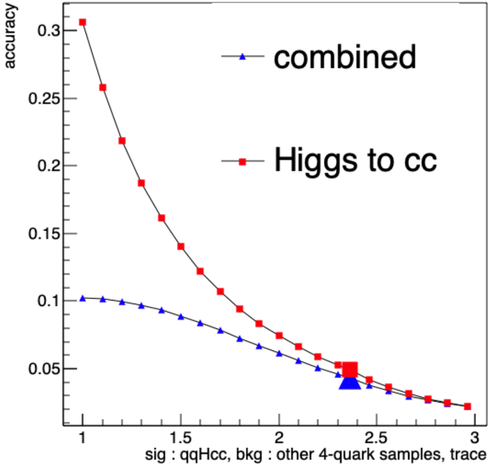
T : matrix with perfect flavor tagging

I : matrix with non flavor tagging

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

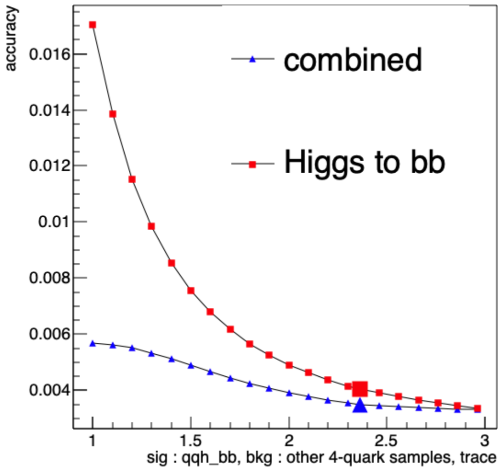
The x value and flavor tagging performance matrix have a one to one relation.

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



with perfect jet clustering and jet matching : 5.49%
with perfect flavor tagging : 2.14%

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



with perfect jet clustering and jet matching : 0.37%
with perfect flavor tagging : 0.33%

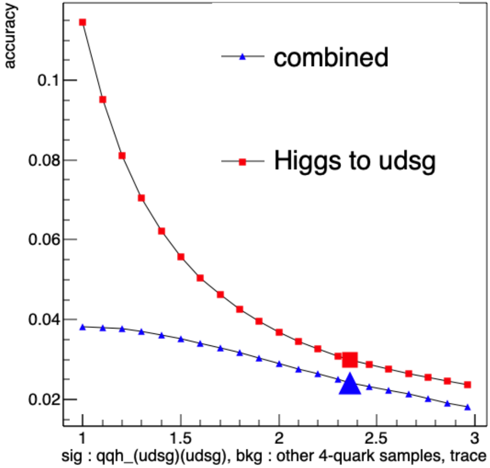
Summary :

- At present, the $qqHc\bar{c}$ measurement accuracy is 8.54%.
- The color singlet identification, including jet clustering, jet matching and flavor tagging, has great impact on $qqHc\bar{c}$ accuracy measurement.
  - With perfect jet clustering and jet matching, the $qqHc\bar{c}$ measurement accuracy can reach 5.49%.
  - With perfect jet clustering, jet matching and flavor tagging, 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 influence of jet clustering and jet matching, 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%