Higgs→bb/cc/gg measurement at the CEPC and corresponding optimization study

Yongfeng Zhu

advisor : Manqi Ruan

IHEP

2022.10.24

Motivation

- The measurement of $Higgs \to b\bar{b}/c\bar{c}/gg$ signal strength is important for Higgs coupling studies.
- The flavor tagging performance and color singlet identification has significant impact on the measurement accuracy.

Contents

- The relative accuracy of signal strength measurement of $\nu\nu H(H \to b\bar{b}, c\bar{c}, gg)$.
  - key performance : flavor tagging
- The relative accuracy of signal strength measurement of $qqH(H \to b\bar{b}, c\bar{c}, gg)$.
  - key performance : flavor tagging
  - key performance : color singlet identification
- summary

Sample

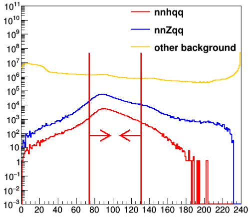
- all SM processes at CEPC ($\sqrt{s} = 240\,GeV$) with integrated luminosity of $5600\,fb^{-1}$

The analysis procedure can be divided into two steps:

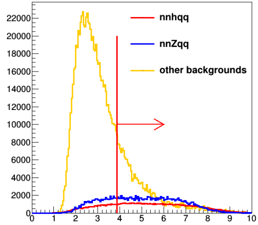
1. select signal events with cut flow
2. calculate the signal strength accuracy

The analysis process of $\nu\nu H$ channel.

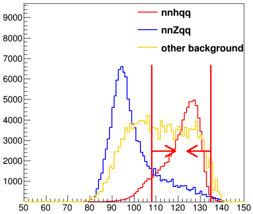
	$\nu\nu Hq\bar{q}/gg$	2f	SW	SZ	WW	ZZ	Mixed	ZH	$\frac{\sqrt{S+B}}{S}$ (%)
total	178890	8.01E8	1.95E7	9.07E6	5.08E7	6.39E6	2.18E7	961606	16.86
recoilMass (GeV) $\in(74, 131)$	157822	5.11E7	2.17E6	1.38E6	4.78E6	1.30E6	1.08E6	74991	4.99
visEn (GeV) $\in(109, 143)$	142918	2.37E7	1.35E6	8.81E5	3.60E6	1.03E6	6.29E5	50989	3.92
leadLepEn (GeV) $\in(0, 42)$	141926	2.08E7	3.65E5	7.24E5	2.81E6	9.72E5	1.34E5	46963	3.59
multiplicity $\in(40, 130)$	139545	1.66E7	2.36E5	5.24E5	2.62E6	9.07E5	4977	42751	3.29
leadNeuEn (GeV) $\in(0, 41)$	138653	1.46E7	2.24E5	4.72E5	2.49E6	8.69E5	4552	42303	3.12
Pt (GeV) $\in(20, 60)$	121212	248715	1.56E5	2.48E5	1.51E6	4.31E5	999	35453	1.37
Pl (GeV) $\in(0, 50)$	118109	52784	1.05E5	74936	7.30E5	1.13E5	847	34279	0.94
-log10(Y23) $\in(3.375, +\infty)$	96156	40861	26088	60349	2.25E5	82560	640	10691	0.76
InvMass (GeV) $\in(116, 134)$	71758	22200	11059	6308	77912	13680	248	6915	0.64
BDT $\in(-0.02, 1)$	60887	9140	266	2521	3761	3916	58	1897	0.47



74 < recoil mass < 131



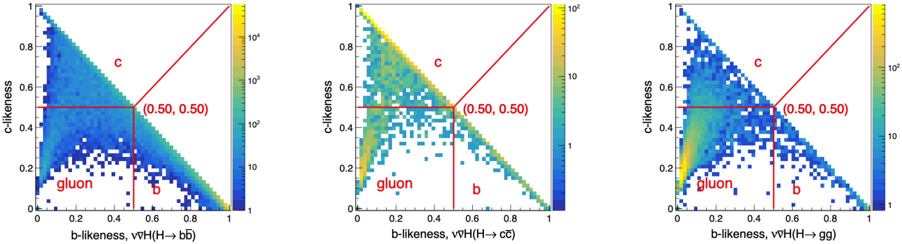
-log(Y23) > 3.875



108 < visible mass < 135

Optimized matrix

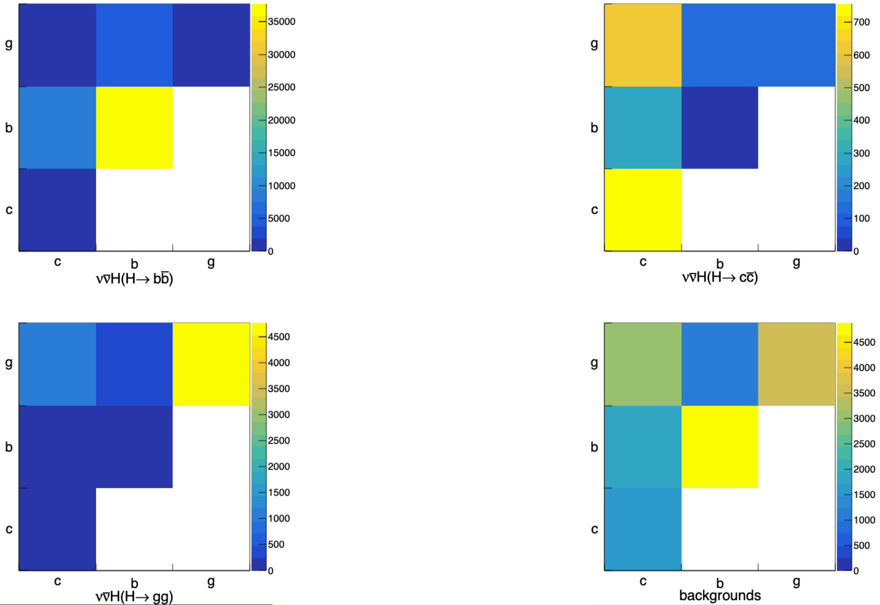
1. The b-likeness and c-likeness of two jets can be displaced in 2D graph.
2. The cut on b-likeness and c-likeness can be find to maximize the value of $eff(b \to b) + eff(c \to c) + eff(udsg \to udsg)$, the trace of flavor tagging matrix.



Optimized matrix :

true \ identified as	b	c	g
b	0.8675	0.0887	0.0437
c	0.1136	0.6263	0.2601
g	0.0411	0.1007	0.8582

events distribution based on optimized matrix :



$$-2\cdot log(\ell) = \sum_{i=1}^{i=6} \frac{[S_b\cdot N_{b,i} + S_c\cdot N_{c,i} + S_{light}\cdot N_{light,i} + N_{bkg,i} - N_i]^2}{N_i}$$

- S_b: the signal strength of $\nu\nu Hb\bar{b}$
- $N_{b,i}$: the event number of $\nu\nu Hb\bar{b}$ in *ith* bin
- N_i: the total event number in *i'th* bin of $\nu\nu Hb\bar{b}$, $\nu\nu H/c\bar{c}$, $\nu\nu Hgg$ and backgrounds
- $N_{bkg,i}$ is the expected event number in *ith* bin of backgrounds,
- similar for S_c, S_{light}, $N_{c,i}$, and $N_{light,i}$

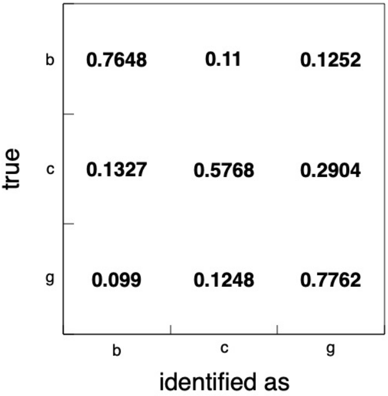
$$hessian\ matrix = \begin{bmatrix} \frac{\partial^2 log(\ell)}{\partial S_g \partial S_c} & \frac{\partial^2 log(\ell)}{\partial S_g \partial S_b} & \frac{\partial^2 log(\ell)}{\partial S_g \partial S_g} \\ \frac{\partial^2 log(\ell)}{\partial S_b \partial S_c} & \frac{\partial^2 log(\ell)}{\partial S_b \partial S_b} & \frac{\partial^2 log(\ell)}{\partial S_b \partial S_g} \\ \frac{\partial^2 log(\ell)}{\partial S_c \partial S_c} & \frac{\partial^2 log(\ell)}{\partial S_c \partial S_b} & \frac{\partial^2 log(\ell)}{\partial S_c \partial S_g} \end{bmatrix}$$

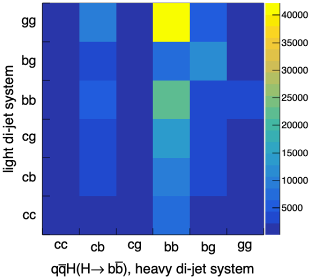
- The error covariance is obtained from the hessian matrix.
- The relative accuracy of signal strength is the square roots of the diagonal elements of the covariance matrix, it is 0.49%/5.75%/1.82% for $\nu\nu Hb\bar{b}/c\bar{c}/gg$.

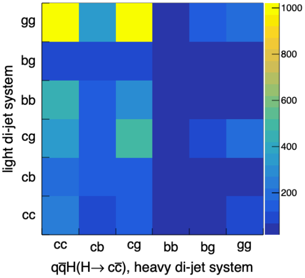
The analysis process of $q\bar{q}H$ channel.

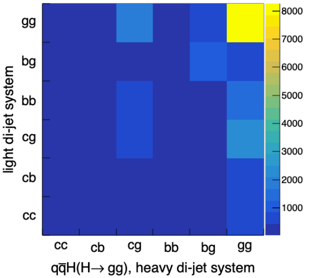
total	qqHqq/gg 527488	2f 8.01E8	SW 1.95E7	SZ 9.07E6	WW 5.08E7	ZZ 6.39E6	Mixed 2.18E7	ZH 613008	$\frac{\sqrt{S+B}}{S}$(%) 5.71
multiplicity $\in(27,+\infty)$	527488	3.04E8	1.46E7	3.37E6	4.85E7	6.00E6	1.81E7	577930	3.77
leadLepEn $\in(0,59)$	527036	2, 98E8	6.76E6	2.44E6	3.93E7	5.40E6	1.79E7	531411	3.65
visEn $\in(199,278)$	510731	1.21E8	1.29E6	551105	2.14E7	3.06E6	1.71E7	180571	2.52
leadNeuEn $\in(0,57)$	509623	5.68E7	716161	168030	2.04E7	2.93E6	1.65E7	176387	1.94
thrust $\in(0,0.86)$	460535	7.81E6	473732	132126	1.88E7	2.60E6	1.54E7	167863	1.47
$-log(Y_{34})$ $\in(0,5.8875)$	451468	4.90E6	181432	119836	1.74E7	2.40E6	1.45E7	165961	1.40
HiggsJetsA $\in(2.18,2\pi)$	326207	2.83E6	110156	58613	4.54E6	870276	3.74E6	96560	1.08
ZJetsA $\in(1.97,2\pi)$	279030	1.37E6	33491	37101	2.39E6	496611	2.00E6	74005	0.93
ZHiggsA $\in(2.32,2\pi)$	274530	1.32E6	17026	33847	2.28E6	468340	1.91E6	69620	0.92
circle	268271	1.20E6	10193	31567	2.13E6	424514	1.79E6	65434	0.90
BDT $\in(0.02,1)$	192278	378300	40	307	271436	141446	244126	30022	0.57

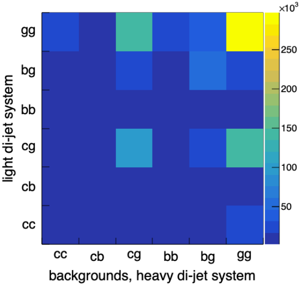
optimized matrix











The signal strength accuracy is 0.35%/7.74%/3.96% for $q\bar{q}Hb\bar{b}/c\bar{c}/gg$.

Z decay mode	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$
$Z \to e^+e^-$	1.57%	14.43%	10.31%
$Z \to \mu^+\mu^-$	1.06%	10.16%	5.23%
$Z \to q\bar{q}$	0.35%	7.74%	3.96%
$Z \to \nu\bar{\nu}$	0.49%	5.35%	1.77%
combination	0.27%	4.03%	1.56%

key performance : flavor tagging

$$M_{mig} = \frac{Tr_{mig} - Tr_{opt}}{Tr_{I} - Tr_{opt}} \cdot (M_{I} - M_{opt}) + M_{opt}, Tr_{mig} >= Tr_{opt}$$

$$M_{mig} = \frac{Tr_{mig} - Tr_{opt}}{Tr_{1/3} - Tr_{opt}} \cdot (M_{1/3} - M_{opt}) + M_{opt}, Tr_{mig} < Tr_{opt}$$

true / identified as

	b	c	g
b	1	0	0
c	0	1	0
g	0	0	1

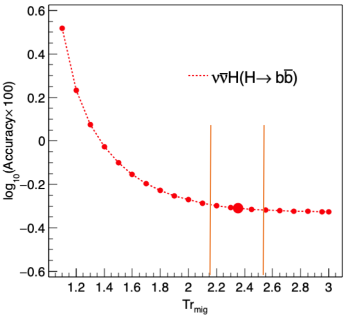
perfect

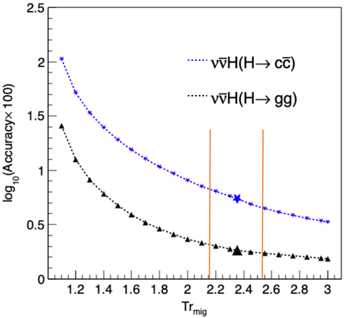
true / identified as

	b	c	g
b	1/3	1/3	1/3
c	1/3	1/3	1/3
g	1/3	1/3	1/3

none

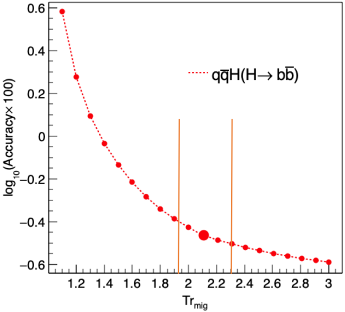
$\nu\bar{\nu}H$

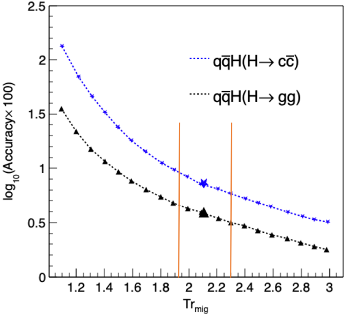




- The perfect flavor tagging performance can improve the $\nu\bar{\nu}H(H\to b\bar{b}/c\bar{c}/gg)$ signal strength accuracy by 2%/63%/13%.
- If the values of vertex detector parameters, including material budget, inner radius, and spatial resolution, is 0.5/2 times compared to the CEPC baseline, the Tr_{mig} will changes from 2.35 to 2.54/2.16 accordingly. The detail can be found in the following pages.

$q\bar{q}H$





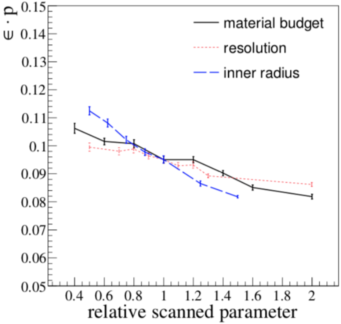
- The perfect flavor tagging performance can improve the $\nu\bar{\nu}H(H \to b\bar{b}/c\bar{c}/gg)$ signal strength accuracy by 35%/122%/181%.
- If the values of vertex detector parameters is 0.5/2 times compared to the CEPC baseline, the Tr_{mig} will changes from 2.12 to 2.31/1.93 accordingly.

The dependency of flavor tagging performance on vertex detector design.

Z. Wu et al 2018 JINST 13 T09002

Table 1. The baseline design parameters of the CEPC vertex system.

	R(mm)	Z(mm)	single-point resolution(μm)	material budget
Layer 1	16	62.5	2.8	$0.15\%/X_0$
Layer 2	18	62.5	6	$0.15\%/X_0$
Layer 3	37	125.0	4	$0.15\%/X_0$
Layer 4	39	125.0	4	$0.15\%/X_0$
Layer 5	58	125.0	4	$0.15\%/X_0$
Layer 6	60	125.0	4	$0.15\%/X_0$



the right plot : the correlation between c-tagging efficiency times purity and vertex detector parameters

The dependency of Tr_{mig} on vertex detector parameters.

In $\nu\bar{\nu}H$ channel,

$$Tr_{mig} = 2.35 + 0.05 \cdot log_2 \frac{R^0_{material}}{R_{material}} + 0.04 \cdot log_2 \frac{R^0_{resolution}}{R_{resolution}} + 0.10 \cdot log_2 \frac{R^0_{radius}}{R_{radius}}.$$

$R^0_{material}$: the default material budget $R_{material}$: the modified material budget

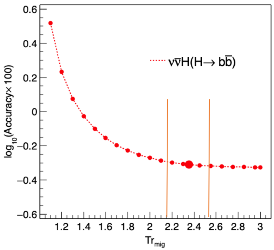
In $q\bar{q}H$ channel,

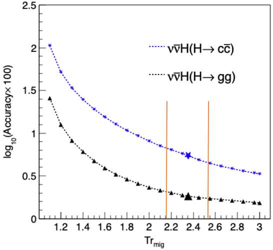
$$Tr_{mig} = 2.12 + 0.05 \cdot log_2 \frac{R^0_{material}}{R_{material}} + 0.04 \cdot log_2 \frac{R^0_{resolution}}{R_{resolution}} + 0.10 \cdot log_2 \frac{R^0_{radius}}{R_{radius}}.$$

Table 2. Reference geometries.

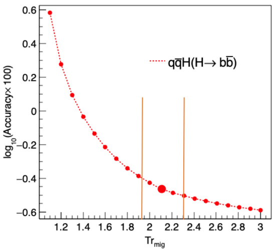
	Scenario A (Aggressive)	Scenario B (Baseline)	Scenario C (Conservative)
Material per layer/X_0	0.075	0.15	0.3
Spatial resolution/µm	1.4 - 3	2.8 - 6	5 - 10.7
R_{in}/mm	8	16	23
Tr_{mig} for $q\bar{q}H$	2.31	2.12	1.93
Tr_{mig} for $\nu\bar{\nu}H$	2.54	2.35	2.16

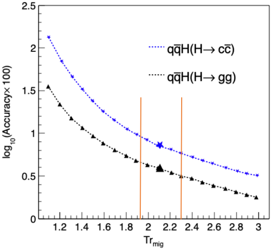
$\nu\bar{\nu}H$



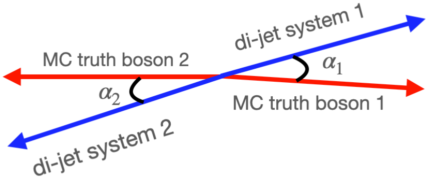


qqH



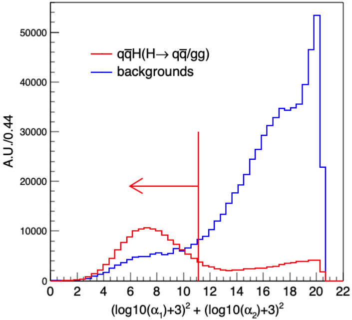


key performance : color singlet identification
(i.e. jet clustering and jet matching)



We use the angle between reconstructed boson and MC truth boson to evaluate the CSI performance.

note : The CSI evaluator in this report is just a demonstrator.



	$qqHb\bar{b}$	$qqHc\bar{c}$	$qqHgg$
w.o. alpha cut	0.35%	7.74%	3.96%
w.i. alpha cut	0.33%	4.37%	2.08%
improved by	6%	77%	90%

Summary :

- The total signal strength of $H \to b\bar{b}, c\bar{c}, gg$ can be measured to a relative accuracy of 0.27%/4.03%/1.56%, combining all four different channels of $\mu\mu H$, eeH, $\nu\nu H$ and qqH.

Z decay mode	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$
$Z \to e^+e^-$	1.57%	14.43%	10.31%
$Z \to \mu^+\mu^-$	1.06%	10.16%	5.23%
$Z \to q\bar{q}$	0.35%	7.74%	3.96%
$Z \to \nu\bar{\nu}$	0.49%	5.35%	1.77%
combination	0.27%	4.03%	1.56%

- The flavor tagging and color singlet identification (CSI) are the critical performances for these benchmarks. Their impact on the anticipated physics reach is evaluated.
  - for $\nu\nu H$ channel
    - The flavor tagging is critical for the $\nu\bar{\nu}H(H \to b\bar{b}/c\bar{c}/gg)$ measurement. Using an ideal flavor tagging, the anticipated accuracy could be improved by 2%/63%/13%.
  - for qqH channel
    - With perfect flavor tagging, the anticipated accuracy of $q\bar{q}H(H \to b\bar{b}/c\bar{c}/gg)$ could be improved by 35%/122%/181%.
    - If we can quantify the CSI performance and select the events with good CSI performance, the $q\bar{q}H(H \to b\bar{b}/c\bar{c}/gg)$ accuracy can be improved by 6%/77%/90%.
- A good color singlet identification, or even a reliable color singlet identification performance evaluator at reconstruction level, is highly appreciated.

https://arxiv.org/abs/2203.01469

Many thanks !

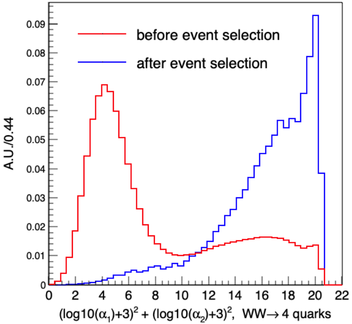
Back Up

systematic uncertainties

We categorize the leading systematic uncertainties into three groups:

- The first group includes the reconstructed energy/momentum scale of the physics objects, which are significantly smaller than the statistical uncertainties.
- The second group are those comparable to the statistical uncertainty, especially the integrated luminosity.
- The third group are those that can be significantly larger than the statistical uncertainty, including CSI and the jet configuration.

The detailed discussion can be found in `https://arxiv.org/abs/2203.01469`.



before event selection
after event selection
A.U./0.44
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0 2 4 6 8 10 12 14 16 18 20 22
(log10(α1)+3)2 + (log10(α2)+3)2, WW→ 4 quarks