The $Higgs \to b\bar{b}, c\bar{c}, gg$ measuremnet at CEPC

Yongfeng Zhu

advisor : Manqi Ruan

IHEP

2021 年 6 月 25 日

Motivation

- The g(Hcc), the second generation fermion Yukawa coupling, is one of the most important benchmark at CEPC, and its measurement strongly depend on the Higgs branching ratio measurement accuracy.

Contents

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

Sample

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

$llH(H \to b\bar{b}, H \to c\bar{c}, H \to gg)$ *accuracy measurement* at CEPC ($\sqrt{s} = 250$ *GeV* with integrated luminosity $5000fb^{-1}$)

Chinese Physics C Vol. 44, No. 1 (2020) 013001

Measurements of decay branching fractions of $H \to b\bar{b}/c\bar{c}/gg$ in associated $(e^+e^-/\mu^+\mu^-)H$ production at the CEPC*

Yu Bai(白羽)[1;1)] Chun-Hui Chen(陈春晖)[2;2)] Ya-Quan Fang(方亚泉)[3;3)] Gang Li(李刚)[3;4)]
Man-Qi Ruan(阮曼奇)[3;5)] Jing-Yuan Shi(史静远)[1,4] Bo Wang(王博)[1,5] Pan-Yu Kong(孔攀宇)[1,6]
Bo-Yang Lan(兰博扬)[1] Zhan-Feng Liu(刘站峰)[1]

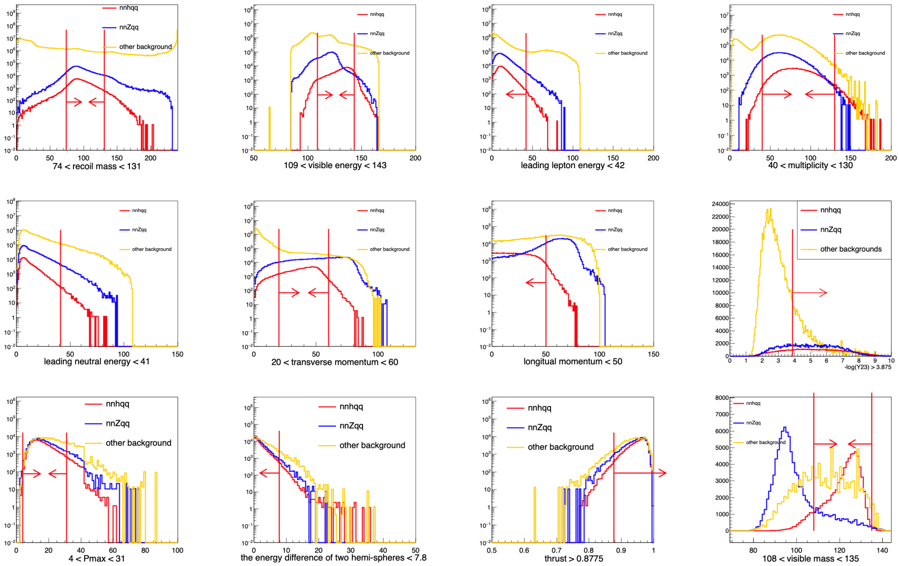
[1]School of Physics, Southeast University, Nanjing 210096, China
[2]Department of Physics and Astronomy, Iowa State University, Ames 50011-3160, USA
[3]Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China
[4]School of physics and astronomy, the University of Manchester, Oxford Rd, M13 9PL, UK
[5]Department of Physics, College of Sciences, Northeastern University, Shenyang 110004, China
[6]School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China

	$\mu^+\mu^-H$			e^+e^-H		
	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$
accuracy	1.1%	10.5%	5.4%	1.6%	14.7%	10.5%

for the detail : `https://arxiv.org/abs/1905.12903v2`

$\nu\nu H(H \to b\bar{b}, H \to c\bar{c}, H \to gg)$ accuracy measurement

The following cuts were used to identify $\nu\nu Hq\bar{q}$.



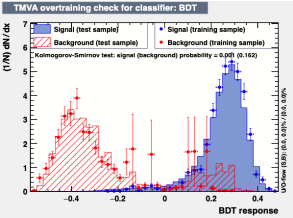
The full cut chain shown in the following table.

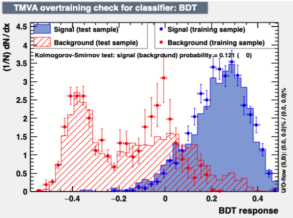
	$\nu\nu Hq\bar{q}$	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{\sqrt{S+B}}{S}$ (%)
total	178890	8.01*E*8	1.95*E*7	9.07*E*6	5.08*E*7	6.39*E*6	2.18*E*7	961606	9.10*E*8	16.86
recoilMass	157822	5.11*E*7	2.17*E*6	1.38*E*6	4.78*E*6	1.30*E*6	1.08*E*6	74991	6.19*E*7	4.99
visEn	142918	2.37*E*7	1.35*E*6	8.81*E*5	3.60*E*6	1.03*E*6	6.29*E*5	50989	3.13*E*7	3.92
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	3.59
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	3.29
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	3.12
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	1.37
Pl	118109	53680	1.08*E*5	74936	729604	1.14*E*5	789	34279	1.11*E*6	0.94
Y23	82035	38771	13084	50195	109007	65822	633	4040	281554	0.74
Pmax	78135	31835	6639	45952	87414	59800	334	3747	235724	0.72
EnDif	76739	31108	6264	44853	79715	57687	320	3306	223256	0.71
bThrust	75822	31016	5948	43599	73986	55312	320	2769	212953	0.70
InvMass	68348	24551	3169	6994	38567	11634	198	1882	86998	0.58

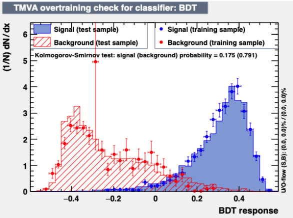
- recoilMass : the recoil mass of final state particles
- visEn : visible energy of final state particles
- Npfo : number of final state particles
- leadLepEn : leading lepton energy,
- leadNeuEn : leading neutral energy
- Pt : the transverse momentum of all final state particles
- Pl : the longitudinal momentum of all final state particles
- thrust, Y23
- EnDif : the energy difference of two hemispheres
- Pmax : the maximum transverse momentum among final state particles
- InvMass : visible mass

The BDT was used to identify $\nu\nu H(H \to b\bar{b}, H \to c\bar{c}, H \to gg)$.

- # of final state particles
- visible energy of final state particles
- leading lepton energy, leading neutral energy
- thrust, Y23
- the transverse momentum of all final state particles
- the longitudinal momentum of all final state particles
- the maximum transverse momentum among final state particles
- the angle of two jets
- # of charge particles
- visible mass
- recoil mass of final state particles
- the b-likeness of each jet
- the b-likeness of each jet
- the energy difference of two hemispheres







$\nu\nu Hb\bar{b}$ $\nu\nu Hc\bar{c}$ $\nu\nu Hgg$

	signal	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{\sqrt{S+B}}{S}$ (%)
$\nu\nu Hc\bar{c}$	1868	5181	120	693	3578	1345	33	1467	12420	6.39
$\nu\nu Hb\bar{b}$	47928	7632	0	940	27	1478	0	305	10385	0.50
$\nu\nu Hg\bar{g}$	5568.31	726	210	1273	2222	1772	0	789	6995.22	2.01

key performance : 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	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

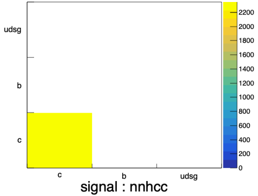
non flavor tagging

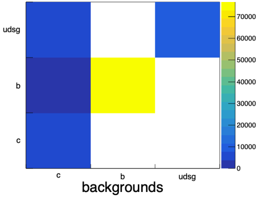
	$\nu\nu Hc\bar{c}$	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{\sqrt{S+B}}{S}$ (%)
InvMass	2833	24551	3950	7525	46912	12735	198	70661	166536	14.53
BDT	2345	17994	270	3195	5620	4908	99	66752	98841	13.56

the input variables for BDT just as variables shown in page 7, exclude flavor tagging information

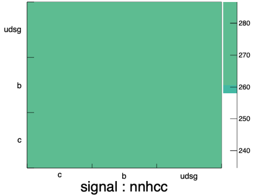
see $\nu\nu Hc\bar{c}$ as signal, and other samples as bkg

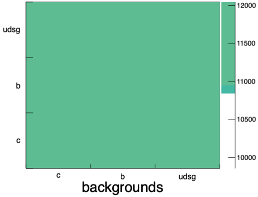
perfect flavor tagging





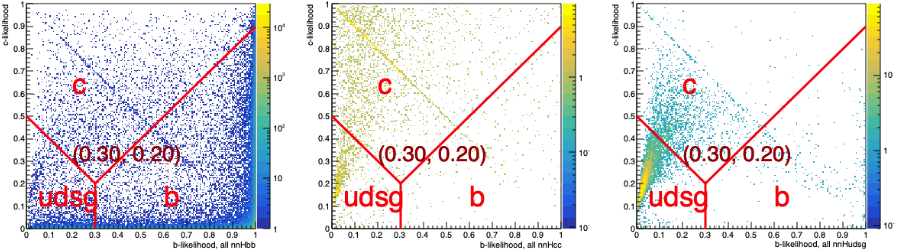
non flavor tagging

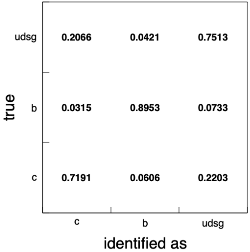


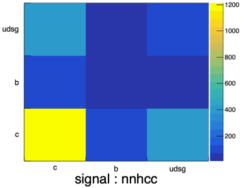


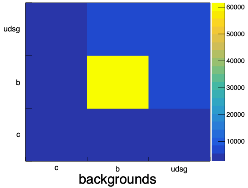
Optimized matrix

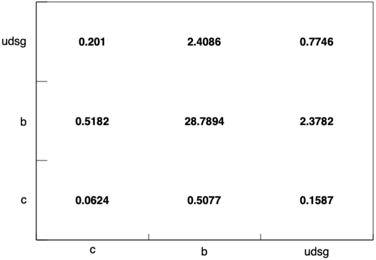
1. The b-likeness and c-likeness of two jets can be displaced in 2D graph.
2. The cut on b-likeness and c-likeness can be find to maximize the value of $eff(b \to b) + eff(c \to c) + eff(udsg \to udsg)$, the trace of flavor tagging matrix.











the accuracy for every bin

- accuracy : $\frac{\sqrt{S+B}}{S}$
- accuracy for $\nu\nu Hc\bar{c}$: 6.24%
- 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 → perfect flavor tagging

the changing procedure of flavor tagging performance matrix :

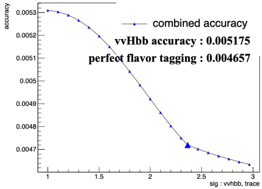
$$temp\ matrix = \frac{x - trace_I}{trace_T - trace_I} \cdot (T - I) + I \qquad (trace_I \le x \le trace_T)$$

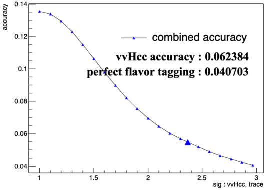
T : matrix with perfect flavor tagging

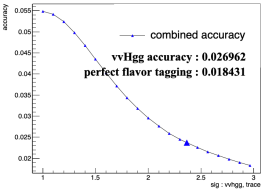
I : matrix with non flavor tagging

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

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





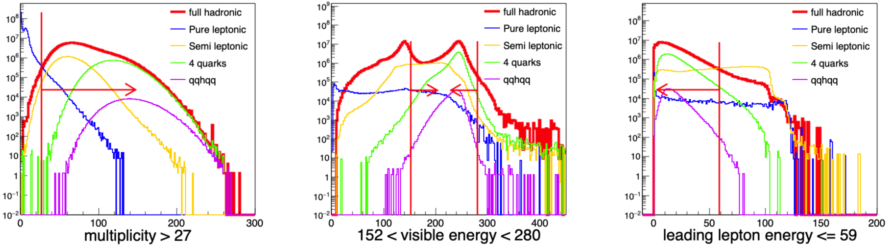


relative accuracy (%)	$nnHb\bar{b}$	$nnHc\bar{c}$	$nnHgg$
BDT	0.50	6.39	2.01
cut based flavor tagging	0.52	6.24	2.70
perfect flavor tagging	0.46	4.07	1.84

$qqH(H \rightarrow b\bar{b}, H \rightarrow c\bar{c}, H \rightarrow gg)$ *accuracy measurement*

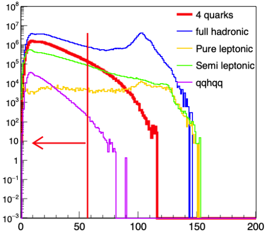
1. Finding the full hadronic samples from all samples.
2. Finding 4-quark samples from the full hadronic samples.
3. Finding $ZH(Z \to q\bar{q}, H \to q\bar{q})$ from 4-quark samples.
4. Finding $qqH(H \to b\bar{b}, H \to c\bar{c}, H \to gg)$ from $ZH(Z \to q\bar{q}, H \to q\bar{q})$.

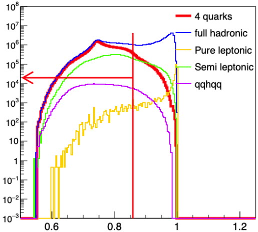
Firstly, finding the full hadronic samples from all samples.

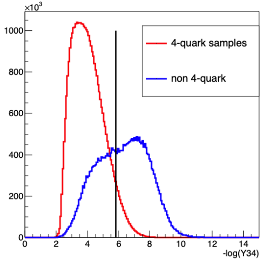


tuned to 4-quark processes finding

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

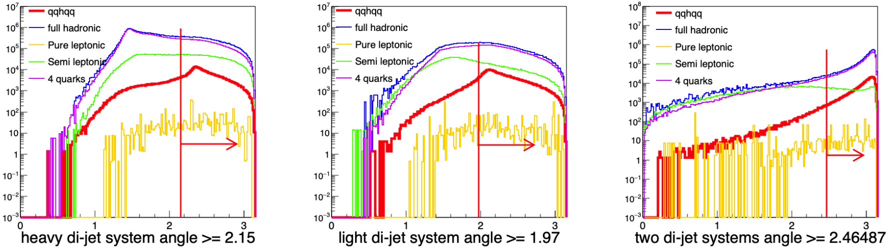




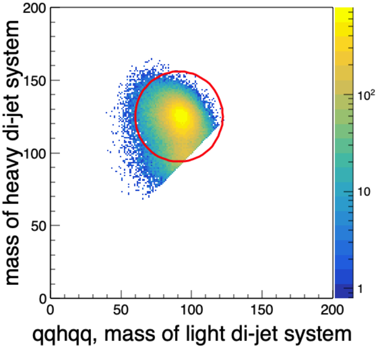


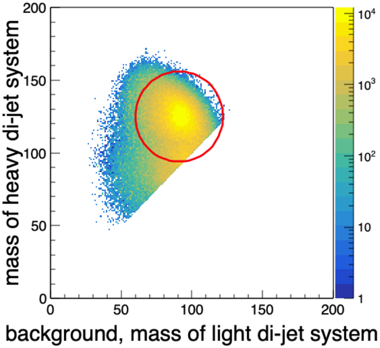
Thirdly, finding $ZH(Z \to q\bar{q}, H \to q\bar{q})$ from 4-quark samples.

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.





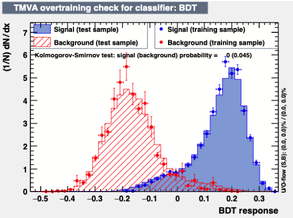
the full cut chain shown in the following table :

	qqHqq	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{\sqrt{S+B}}{S}$ (%)
total	527488	8.01E8	1.95E7	9.07E6	5.08E7	6.39E6	2.18E7	613008	9.09E8	5.71
multiplicity	527488	3.04E8	1.46E7	3.37E6	4.85E7	6.00E6	1.81E7	577930	3.95E8	3.77
visEn	527281	1.54E8	1.30E7	1.66E6	4.00E7	4.25E6	1.80E7	320888	2.32E8	2.89
LLepEn	526854	1.50E8	5.16E6	8.01E5	3.09E7	3.66E6	1.78E7	275612	2.08E8	2.74
LNeuEn	525737	7.81E7	4, 29E6	2.48E5	2.96E7	3, 51E6	1.72E7	270709	1.33E8	2.20
thrust	463900	1.76E7	3.49E6	1.76E5	2.58E7	2.99E6	1.58E7	245858	6.60E7	1.76
$-log(Y_{34})$	454993	7.07E6	1.30E6	146E5	2.03E7	2.66E6	1.48E7	237888	4.72E7	1.52
HJetA	334789	3.74E6	5.71E5	74874	6.20E6	1.07E6	4.16E6	145114	1.60E7	1.20
ZJetA	283229	1.60E6	1.67E5	44807	2.97E6	606051	2.22E6	104944	7.71E6	1.00
ZHA	271757	1.28E6	17528	33793	2.36E6	490831	1.94E6	80537	6, 21E6	0.94
circle	263723	1.14E6	7565	30675	2.14E6	428375	1.80E06	66285	5.62E6	0.92

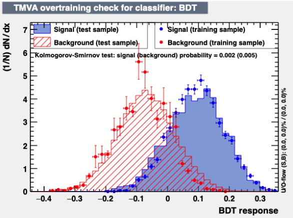
The circle cut can be optimized for different Higgs decay channels.

Finally, finding $qqH(H \rightarrow b\bar{b}, c\bar{c}, gg)$ from $ZH(Z \rightarrow q\bar{q}, H \rightarrow q\bar{q})$ with BDT.

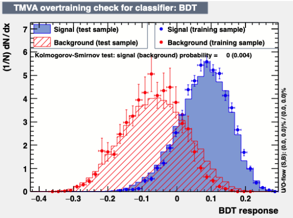
- # of final state particles
- visible energy of final state particles
- leading lepton energy, leading neutral energy
- thrust, Y23, Y34
- the transverse momentum of all final state particles
- the minimal angle among four jets
- the invariant mass of light di-jet system
- the invariant mass of heavy di-jet system
- # of charge particles in each jet
- # of particles in heavy di-jet system
- the b-likeness of each jet
- the b-likeness of each jet



qqhbb



qqhcc



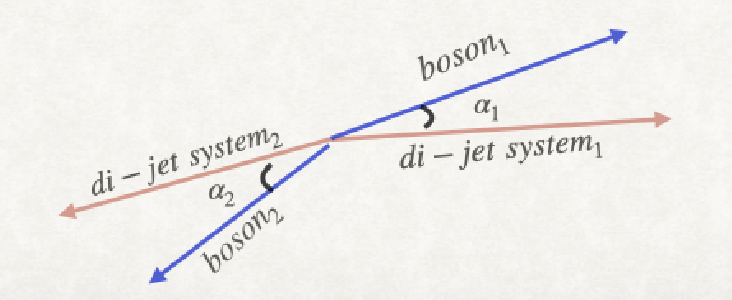
qqhgg

	signal	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{\sqrt{S+B}}{S}$ (%)
qqhbb	172667	89849	0	67	6996	66009	7134	10124	180182	0.34
qqhcc	3330	27455	0	22	15523	11290	27640	6532	88464	9.09
qqhgg	19819	155861	0	52	214696	51555	189740	37254	649159	4.12

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

color singlet identification

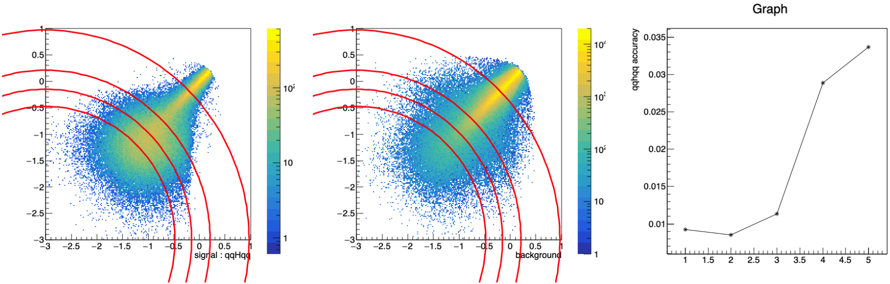
For di-boson event, there are two MC truth bosons and two di-jet systems, the variable $\alpha_i = angle(di-jet\ system_i, truth\ boson_i), (i = 1, 2)$ is used to characterize the performance of jet clustering and jet matching.



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

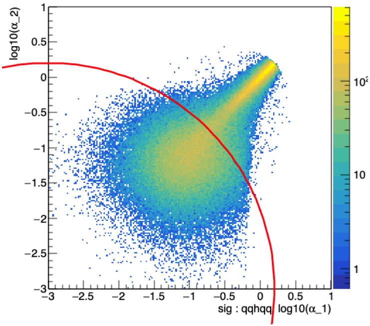
	qqHqq	2f	SW	SZ	WW	ZZ	Mixed	ZH	total bkg	$\frac{\sqrt{S+B}}{S}$ (%)
circle	263723	1.14E6	7565	30675	2.14*E*6	428375	1.80E06	66285	5.62*E*6	0.92

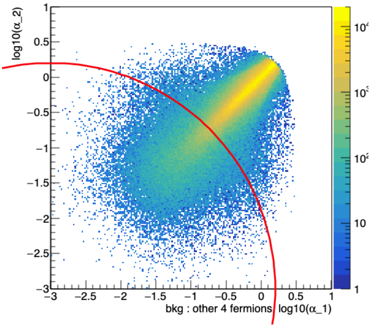
After the above cut, instead of BDT, the distribution of $log10(\alpha_1)$ and $log10(\alpha_2)$ shown as following plots.



Four red curve correspond to $log10(\alpha_1)$ = 2.525, 2.85, 3.21 and 3.97, respectively.

	bin_1	bin_2	bin_3	bin_4	bin_5
qqhqq	51245	52607	51244	51013	50954
bkg	171825	148041	287875	2.13E6	2.89E6
accuracy	0.92%	0.85%	1.14%	2.89%	3.37%





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

	qqhbb	qqhcc	qqhgg	background
circle	215416	10236	31274	5.62E6
alpha	119357	6185	16178	497386
relative accuracy (%)	0.67	12.92	4.94	

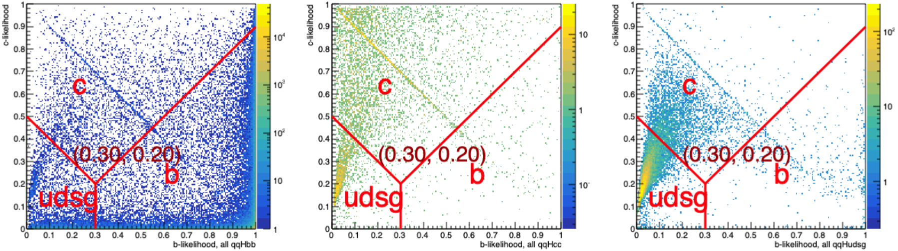
1. With alpha cut $(log10(\alpha_1)+3)^2+(log10(\alpha_2)+3)^2 <= 3.1^2$, the $qqHc\bar{c}$ measurement accuracy can reach 12.92% even though without flavor tagging algorithm.
2. With optimized flavor tagging performance, the accuracy can reach 5.87%.

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

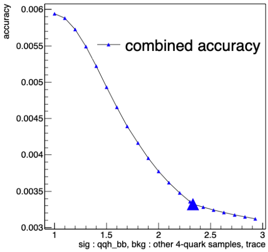
key performance : flavor tagging

Optimized matrix

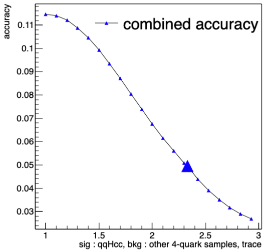
1. Suppose that the alpha cut has select the events with good color singlet identification.
2. Then the b-likeness and c-likeness of two jets from heavy di-jet system, corresponds to Higgs for $qqHq\bar{q}$, can be displaced in 2D graph.



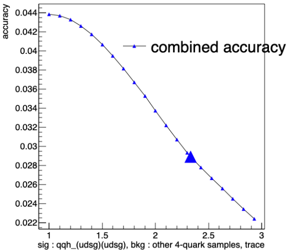
the variation of $qqHq\bar{q}$ measurement accuracy with trace



qqhbb



qqhcc



qqhgg

relative accuracy (%)	$qqHb\bar{b}$	$qqHc\bar{c}$	$qqHgg$
BDT	0.34	9.10	4.13
alpha	0.39	5.87	3.63
alpha & perfect flavor tagging	0.31	2.73	2.89

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%/4.46%/1.69%, combining all four different channels of $\mu\mu H$, eeH, $\nu\nu H$ and qqH.

	$\mu^+\mu^- H$	$e^+e^- H$	$\nu\nu H$	qqH	combined
$H \to b\bar{b}$	1.1%	1.6%	0.50%	0.34%	0.27%
$H \to c\bar{c}$	10.5%	14.7%	6.39%	9.10%	4.46%
$H \to gg$	5.4%	10.5%	2.01%	4.13%	1.69%

- 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 $H \to c\bar{c}$ measurement. Using an ideal flavor tagging, the anticipated accuracy could be improved by 57% (baseline/ideal : 6.39%/4.07%).
  - for qqH channel
    - The CSI dominant the precision for $H \to c\bar{c}$ measurement. If we can quantify the CSI performance and select halve of the statistic, corresponding to relatively good CSI, the $H \to c\bar{c}$ accuracy can already be improved by 55% (baseline/good CIS : 9.10%/5.87%).
    - With good CSI and perfect flavor tagging, the anticipated accuracy could be improved by 233% (baseline/ideal : 9.10%/2.73%).
- A good color singlet identification, or even a reliable color singlet identification performance evaluator at reconstruction level, is highly appreciated.

Thanks !

Backup

ILC result

Eur. Phys. J. C (2013) 73:2343
DOI 10.1140/epjc/s10052-013-2343-8

THE EUROPEAN PHYSICAL JOURNAL C

Special Article - Tools for Experiment and Theory

A study of measurement precision of the Higgs boson branching ratios at the International Linear Collider

Hiroaki Ono[1,a], **Akiya Miyamoto**[2]

[1] School of Life Dentistry at Niigata, Nippon Dental University, 1-8 Hamaura-cho, Chuo-ku, Niigata 951-8580, Japan
[2] High Energy Accelerator Research Organization, 1-1 Oho, Tsukuba 305-0801, Japan

Table 1 Summary of the $\nu\bar{\nu}H$ channel background reduction assuming $\mathscr{L} = 250\ \mathrm{fb}^{-1}$ with $P(e^-, e^+) = (-0.8, +0.3)$

CM energy	250 GeV			350 GeV
Cut names	Condition	Sig.	Bkg.	Condition
Generated		19360	44827100	
Missing mass	$80 < M_{miss} < 140$ GeV	15466	6214050	$50 < M_{miss} < 240$ GeV
Transverse visible momentum	$20 < P_T < 70$ GeV	13727	549340	$10 < P_T < 140$ GeV
Longitudinal visible momentum	$P_L < 60$ GeV	13342	392401	$P_L < 130$ GeV
# of charged tracks	$N_{chd} > 10$	12936	374877	$N_{chd} > 10$
Maximum track momentum	$P_{max} < 30$ GeV	11743	205038	$P_{max} < 60$ GeV
Y_{23} value	$Y_{23} < 0.02$	7775	74439	$Y_{23} < 0.02$
Y_{12} value	$0.2 < Y_{12} < 0.8$	7438	62584	$0.2 < Y_{12} < 0.8$
Di-jet mass	$100 < M_{jj} < 130$ GeV	6691	19061	$100 < M_{jj} < 130$ GeV
Likelihood ratio	$LR > 0.165$	6293	10940	$LR > 0.395$
Significance (Efficiency)	$S/\sqrt{S+B}$	47.9 (32.5 %)		$S/\sqrt{S+B}$

Table 2 Summary of $q\bar{q}H$ channel background reduction assuming $\mathscr{L} = 250\ \mathrm{fb}^{-1}$ with $P(e^-, e^+) = (-0.8, +0.3)$

CM energy	250 GeV			350 GeV
Cut names	Condition	Sig.	Bkg.	Condition
Generated		52507	45904900	
χ^2	$\chi^2 < 10$	32447	2608980	$\chi^2 < 10$
# of charged tracks	$N_{chd} > 4$	25281	1120950	$N_{chd} > 4$
Y_{34} value	$-\log(Y_{34}) > 2.7$	25065	1002125	$-\log(Y_{34}) > 2.7$
thrust	thrust < 0.9	24688	935950	thrust < 0.85
thrust angle	$\lvert\cos\theta_{thrust}\rvert < 0.9$	21892	696201	$\lvert\cos\theta_{thrust}\rvert < 0.9$
Higgs jets angle	$105° < \theta_H < 160°$	20062	622143	$70° < \theta_H < 120°$
Z di-jet mass	$80 < M_Z < 100$ GeV	16359	411863	$80 < M_Z < 100$ GeV
H di-jet mass	$105 < M_H < 130$ GeV	16359	411863	$105 < M_H < 130$ GeV
Likelihood ratio	$LR > 0.375$	13726	166807	$LR > 0.15$
Significance (Efficiency)	$S/\sqrt{S+B}$	32.3 (26.1 %)		$S/\sqrt{S+B}$

	Higgs Mass	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$
CEPC	125 GeV	57.7%	2.91%	8.57%
ILC	120 GeV	65.7%	3.6%	5.5%

CEPC CDR result

Table 6. Expected relative precision on $\sigma(ZH)\times\mathrm{BR}$ for the $H\rightarrow b\bar{b}$, $c\bar{c}$ and gg decays from a CEPC dataset of $5.6\,\mathrm{ab}^{-1}$.

Z decay mode	$H\rightarrow b\bar{b}$	$H\rightarrow c\bar{c}$	$H\rightarrow gg$
$Z\rightarrow e^+e^-$	1.3%	12.8%	6.8%
$Z\rightarrow \mu^+\mu^-$	1.0%	9.4%	4.9%
$Z\rightarrow q\bar{q}$	0.5%	10.6%	3.5%
$Z\rightarrow \nu\bar{\nu}$	0.4%	3.7%	1.4%
combination	0.3%	3.1%	1.2%