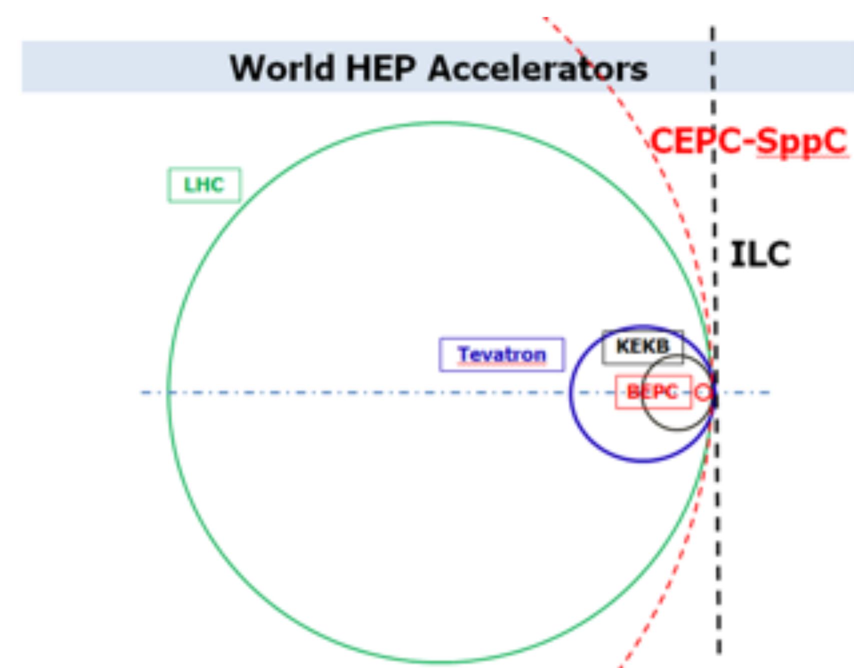
CEPC-SPPC Physics-Software Meeting

H->bb/cc/gg Branch Ratio Measurement in CEPC

Yu Bai (from Southeast University, Nanjing)

On Behalf of CEPC Physics-Software Study Group

August 30, 2016



Introduction: Benchmark Physics Process Measurement Overview

Physics Motivation

- Crucial to understand EW mechanism
  - Reason of fermions mass
  - Is there only one higgs boson?
- CEPC is ideal for such measurement
  - Clean background
  - Utility of recoil mass, free of higgs decay channel
  - High luminosity

One of the benchmark measurement in CEPC

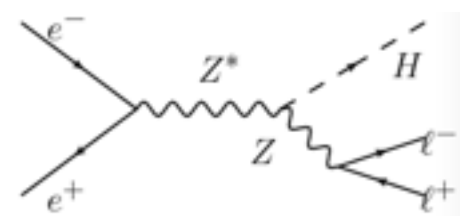
Physics Process	Measured Quantity	Critical Detector	Required Performance
$ZH \to \ell^+\ell^- X$	Higgs mass, cross section	Tracker	$\Delta(1/p_T) \sim 2 \times 10^{-5}$
$H \to \mu^+\mu^-$	$\mathrm{BR}(H \to \mu^+\mu^-)$		$\oplus 1 \times 10^{-3}/(p_T \sin\theta)$
$H \to b\bar{b},\ c\bar{c},\ gg$	$\mathrm{BR}(H \to b\bar{b},\ c\bar{c},\ gg)$	Vertex	$\sigma_{r\phi} \sim 5 \oplus 10/(p \sin^{3/2}\theta)\ \mu\mathrm{m}$
$H \to q\bar{q},\ VV$	$\mathrm{BR}(H \to q\bar{q},\ VV)$	ECAL, HCAL	$\sigma_E^{\mathrm{jet}}/E \sim 3-4\%$
$H \to \gamma\gamma$	$\mathrm{BR}(H \to \gamma\gamma)$	ECAL	$\sigma_E \sim 16\%/\sqrt{E} \oplus 1\%\ (\mathrm{GeV})$

Excellent flavor tagging capability required

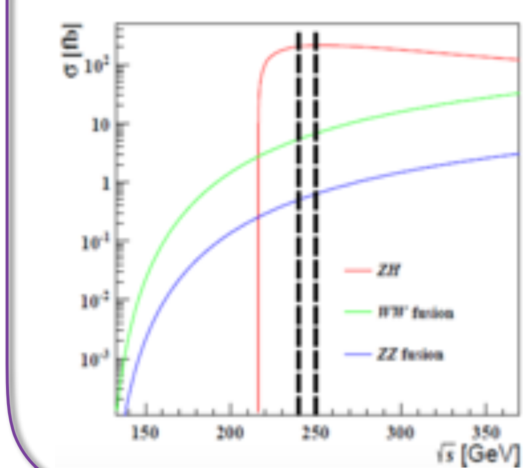
Introduction: Signal Process

- **ZH->ll+jj (Zhenxing Chen)**
  - ZH production with Z decay to muon pair, H decay to bb/cc quark or gluon pair
  - Very clean signal in muon pair invariant mass and recoil mass
- **ZH->vv+jj(Yulei Zhang, Dikai Li, Hao Liang)**
  - Via ZH(~86%) or WW fusion(~14%)
  - Clean background
- **ZH->Multi-jet (Boyang Li, Yu Bai)**
  - Both Z and Higgs decay hadronically
  - Much larger cross section than semi-leptonic channel

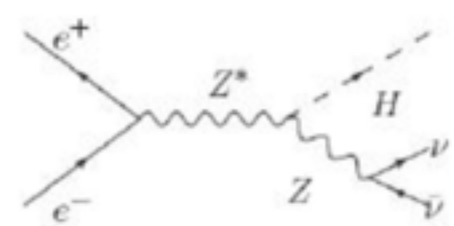
ZH->ll+jj

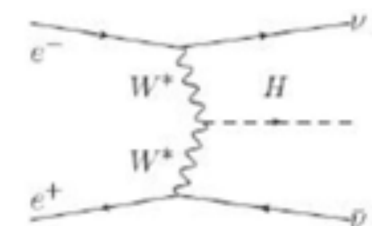


$\sqrt{s}$ = 250 GeV

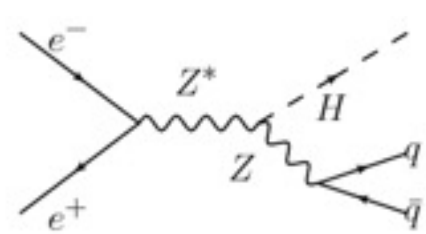


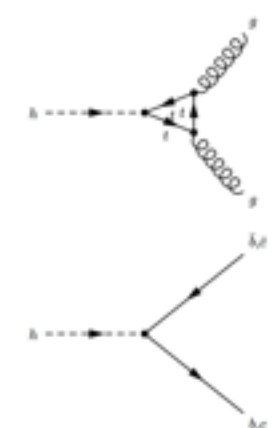
ZH->vv+jj





ZH->multi-jets





Integral luminosity of 5000 fb^{-1} is assumed in these study, corresponding to that of a few years of CEPC running

Analysis Outline

Event Preselection

	vvH Channel	llH Channel	qqH Channel
Dominant Background	WW/ZZ->semi-lep, quark pair production, Other Higgs Process	ZZ semi-lep, Other Higgs Processes	WW/ZZ hadronic, quark-pair production
Discriminator	Missing Energy, Jets pair invariant/recoil masses, jet multiplicity(yth-value).	Lepton pair invariant Mass, Lepton pair recoil Mass, Jets invariant Mass	Jet multiplicity, jet paring(invariant mass and angular distribution)

Flavor Tagging

LCFIPlus: MVA flavor weight, Template fit or other methods

The outcome currently interpreted in terms of $\sigma \times br$ with statistic uncertainty

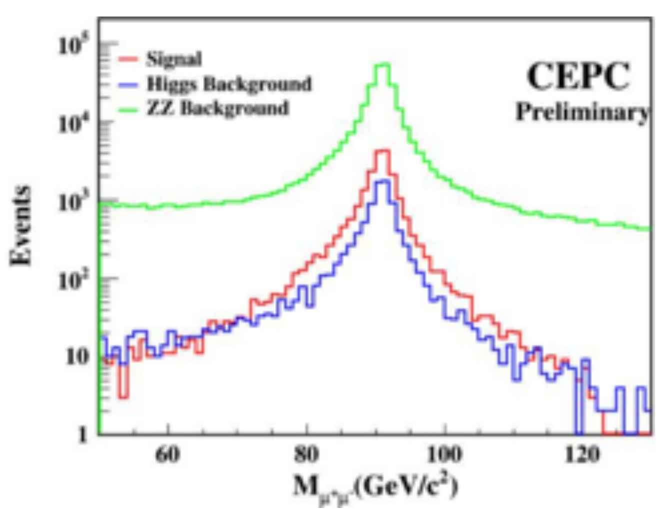
ZH->ll+jj Channel

Final states, 2 muons/electrons + 2 jets.

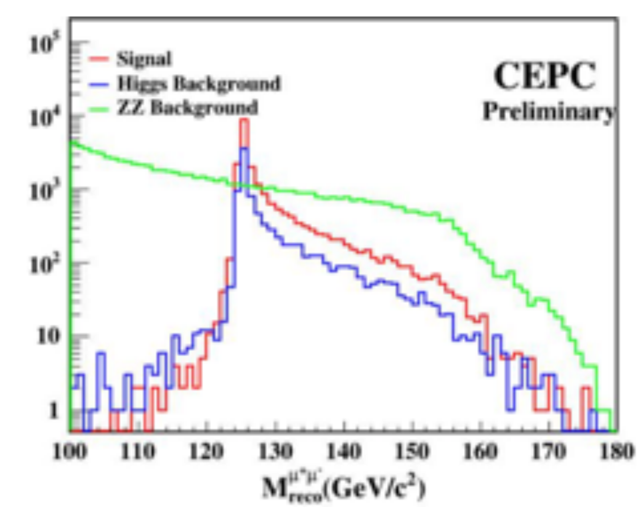
Using invariant mass of muon pair, jet pair, and the recoil mass of muon pair

Recoil mass study see: http://indico.ihep.ac.cn/event/5592/contribution/7/material/slides/0.pdf

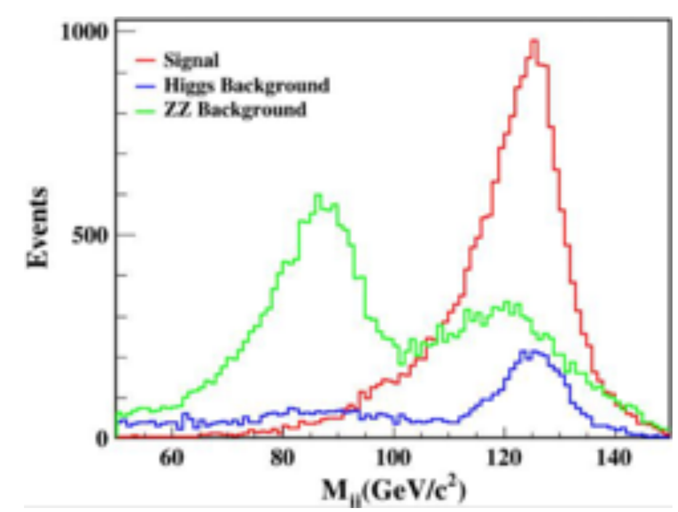
Lepton pair invariant mass between 70 - 105 GeV



Lepton pair recoil mass between 120 - 150 GeV(120 - 160 GeV for ee channel)



jet pair recoil mass between 105 - 135 GeV



Results of Event Selection

	signal	higgs bkg	non-higgs bkg
ee channel	9.42k	1.87k	11.1k
$\mu\mu$ channel	16.5k	3.34k	7.81k

$\nu\nu$+jj Channel, Datasets and Event Pre-selection

Datasets:

Name	Statistics	weight	Note
$\nu\nu H$	5000fb^{-1}	1	Full simulation
$(qq, e^+e^-, \mu^+\mu^-)H$	5000fb^{-1}	1	Full simulation
$\tau^-\tau^+H$	0	0	Not available
2fermions/4fermions	500fb^{-1}	10	Fast simulation

Generator: Whizard, Simulation arbor
Major SM background are with qqln, qqnn and qq events

Event Pre-selection:

- Number of particles(PFO) >=20
- Visible Energy between 110 and 150 Ge
- Isolated electron and isolated muon veto
- y_{12} between 0.15 and 1.0, y_{23}<0.06, y_{34}<0.008
- cos θ between -0.98 and -0.4,
- BDT Cut

Event yields after Cut flow

Cut Definition	*Sig.*	*qq*	*qqnn*	*qqln*	*xxh*
FSClasser output	148955	25M	183687	3698817	63194
$N_{\text{PFO}(E>0.4\text{GeV})} > 20$	148808	23M	163088	3439927	58882
$110 < E_{\text{total}} < 150$	132561	10M	125878	705357	34215
$P_T > 19$	126006	34198	116314	627602	32300
Isolation lepton veto	123586	33775	115867	327206	23773
$100 < M_{\text{inv}} < 135$	117845	9506	10420	162511	21277
$70 < M_{\text{rec}} < 125$	111886	7521	10045	110426	20458
$0.15 < y_{12} < 1$	111353	7405	9702	101797	19983
$y_{23} < 0.06$	105078	6644	8456	69313	14495
$y_{34} < 0.008$	100117	6504	7878	58532	6899
$-0.98 < \cos(\theta^{(2\text{jets})}_{\text{included}}) < -0.4$	97277	5178	5365	33293	6273
$BDT > -0.01$	76666	344	118	69	1594
Significance			**265.20**		
Efficiency			**51.5%**		

Signal and xxh background ~ 5000/fb

qq,qqnn,qqln ~ 500/fb

Signal yields: 76.6k, background: 6.9k

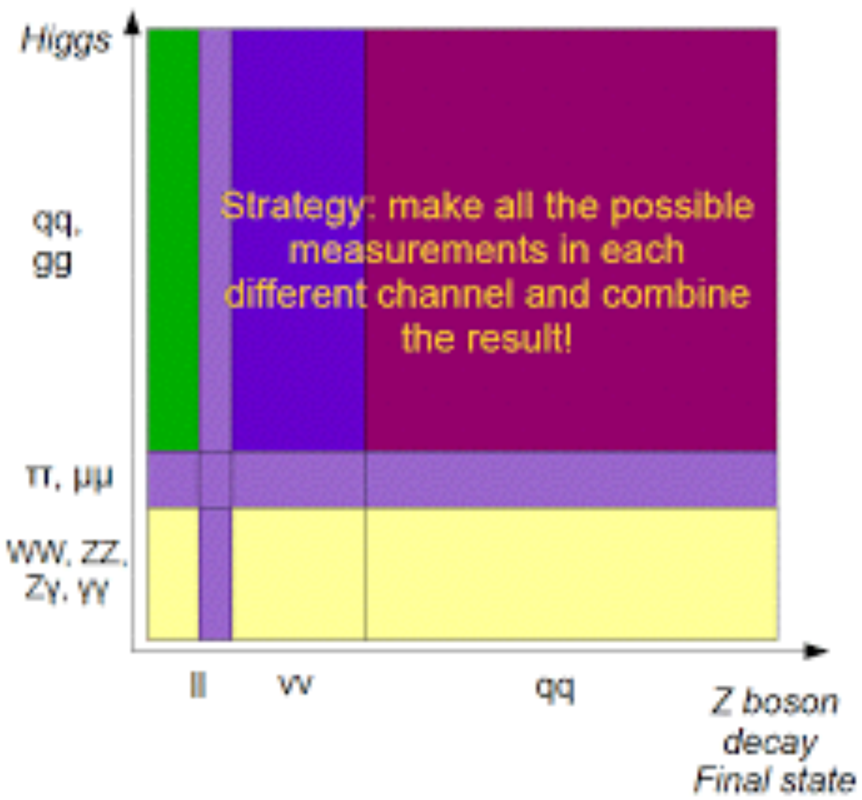
Very high signal/background ratio

Much better than ILC Results:

		Signal	Background
Likelihood ratio	$LR > 0.165$	6293	10940

Multi-jets Channel

Majority decay FS in HZ production
Large Statistics



Datasets:

- **Main background: Irreducible background from WW/ZZ and quark pair process**
- **Full simulation sample used**
- **Quark pair sample are filtered due to its huge cross section and low rate to passing the event selection**

With 4 jets in final states, which two jets should be used to do template fit (from higgs) is ***not*** straightforward.

Event Selection and Cut flow

筛选条件	四喷注组合并去除轻子	可见能量>206GeV	y_{34} > 0.007	喷注粒子数>9	$\Delta\theta > 0.92$	X>0.21	BDT>-0.19
所有信号	493947	459972	393979	371240	318163	236652	211281
信号－$b\bar{b}$	413299	381470	325137	305982	261808	197510	177447
信号－$c\bar{c}$	19362	18690	15976	14903	12610	9562	8324
信号－gg	61286	59812	52866	50355	43745	29580	25510
所有背景	75M	50.6M	26.4M	21.4M	13.55M	3.15M	1.52M
希格斯背景	299534	109529	100813	82281	71987	38579	32653
四费米子强子衰变	36.83M	28.19M	21.32M	18.78M	12.16M	2.2M	1.08M
两夸克	23.86M（250M）	20.37M	5.207M	2.601M	1.315M	907188	405567
四费米子半轻子衰变	14.72M	1.967M	218394	27487	4745	3012	580

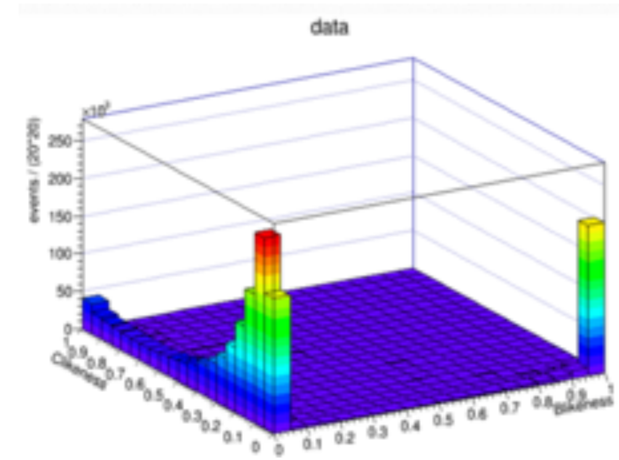
See Boyang's talk with detail

After event selection, the dominant background are WW/ZZ hadronic decay

Better than ILC Results: Likelihood ratio $LR > 0.375$ 13726 166807

Template Fit and ToyMC Results

$$L_{qq} = \frac{qq\ pair}{qq\ pair\ +\ neither\ is\ q} = \frac{x_q^1\, x_q^2}{x_q^1 x_q^2 + (1-x_q^1)(1-x_q^1)}\ (qq = bb, cc)$$



CEPC Results

	e^+e^-H	$\mu^+\mu^-H$	ννH (2500 fb^{-1})	qqH	combined
r_{bb}	1.000±0.013	1.000±0.009	0.997±0.006	1.000±0.003	1.000±0.002
r_{cc}	1.002±0.118	0.993±0.095	1.008±0.038	1.000±0.120	1.00±0.02
r_{gg}	0.998±0.106	0.991±0.070	1.020±0.019	1.001±0.189	1.00±0.01

Large uncertainty due to SM bkg

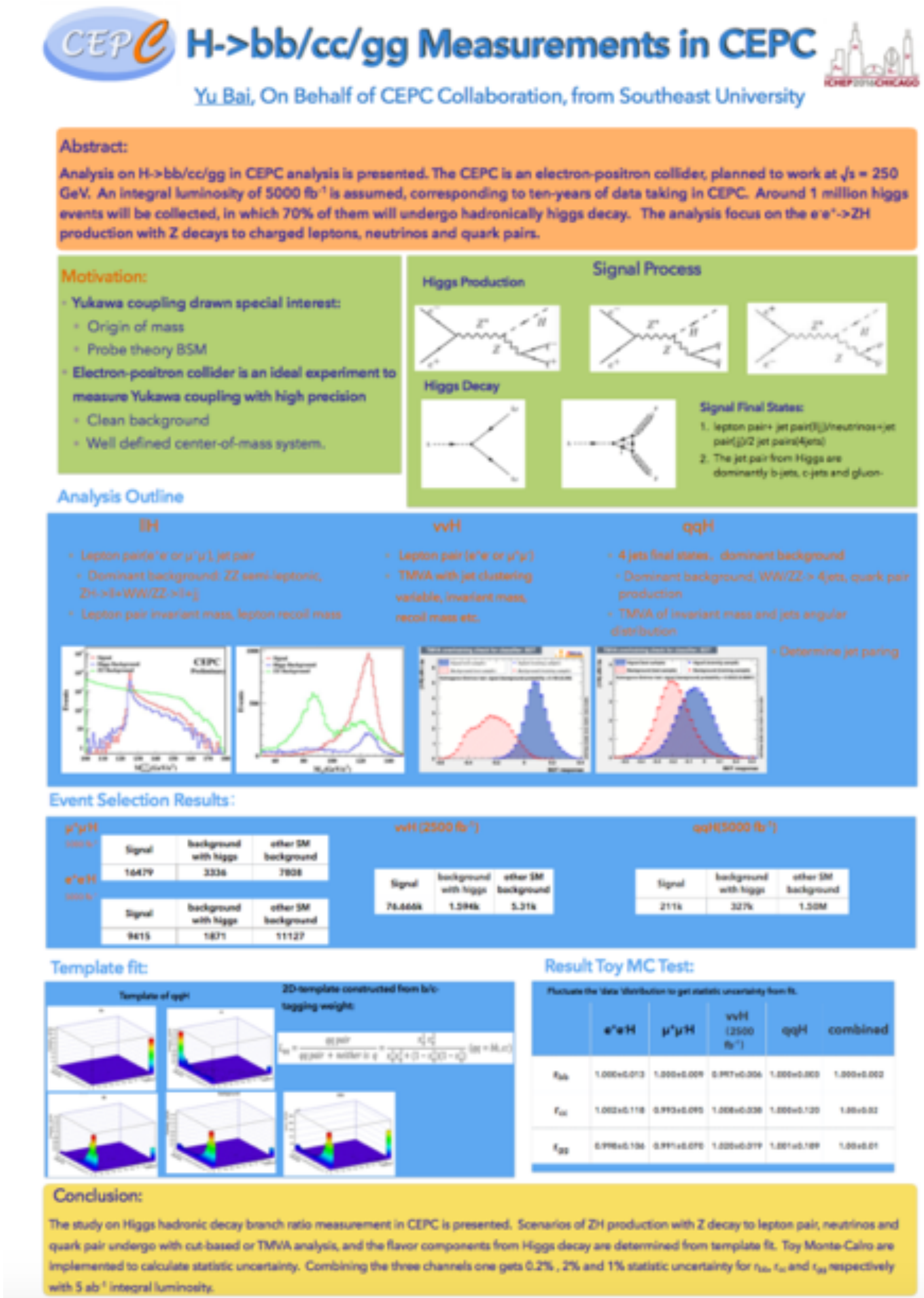
Comparable results in llH channel

Much better result in ννH channel

Much worse results(for r_c and r_g) in qqH channel

ILC Results

	$\nu\bar{\nu}H$	$q\bar{q}H$	e^+e^-H	$\mu^+\mu^-H$	Comb.
$r_{b\bar{b}}$	1.00 ± 0.02	1.00 ± 0.01	1.00 ± 0.04	1.00 ± 0.03	1.00 ± 0.01
$r_{c\bar{c}}$	1.02 ± 0.11	1.01 ± 0.10	1.02 ± 0.27	1.01 ± 0.23	1.02 ± 0.07
r_{gg}	1.02 ± 0.14	1.02 ± 0.13	1.05 ± 0.33	1.02 ± 0.24	1.02 ± 0.09



Presented at ICHEP 2016

Thanks for your hard working!

EightEquation Method

In jet pair production:

$$n_{b1} = \epsilon_{b1}^{b} N_b + \epsilon_{b1}^{c} N_c + \epsilon_{b1}^{l} N_l + N_{b1}^{bkg}$$
$$n_{b2} = \epsilon_{b2}^{b} N_b + \epsilon_{b2}^{c} N_c + \epsilon_{b2}^{l} N_l + N_{b2}^{bkg}$$
$$n_{b1,b2} = \epsilon_{b1}^{b}\epsilon_{b2}^{b}(1+\rho_{b1,b2}^{b}) N_b + \epsilon_{b1}^{c}\epsilon_{b2}^{c}(1+\rho_{b1,b2}^{c}) N_c + \epsilon_{b1}^{l}\epsilon_{b2}^{l}(1+\rho_{b1,b2}^{l}) N_l + N_{b1,b2}^{bkg}$$
$$n_{c1} = \epsilon_{c1}^{b} N_b + \epsilon_{c1}^{c} N_c + \epsilon_{c1}^{l} N_l + N_{c1}^{bkg}$$
$$n_{c2} = \epsilon_{c2}^{b} N_b + \epsilon_{c2}^{c} N_c + \epsilon_{c2}^{l} N_l + N_{c2}^{bkg}$$
$$n_{c1,c2} = \epsilon_{c1}^{b}\epsilon_{c2}^{b}(1+\rho_{c1,c2}^{b}) N_b + \epsilon_{c2}^{c}\epsilon_{c2}^{c}(1+\rho_{c1,c2}^{c}) N_c + \epsilon_{c1}^{l}\epsilon_{c2}^{l}(1+\rho_{c1,c2}^{l}) N_l + N_{c1,c2}^{bkg}$$
$$n_{b1,c2} = \epsilon_{b1}^{b}\epsilon_{c2}^{b}(1+\rho_{b1,c2}^{b}) N_b + \epsilon_{b1}^{c}\epsilon_{c2}^{c}(1+\rho_{b1,c2}^{c}) N_c + \epsilon_{b1}^{l}\epsilon_{c2}^{l}(1+\rho_{b1,c2}^{l}) N_l + N_{b1,c2}^{bkg}$$
$$n_{c1,b2} = \epsilon_{c1}^{b}\epsilon_{b2}^{b}(1+\rho_{c1,b2}^{b}) N_b + \epsilon_{c1}^{c}\epsilon_{b2}^{c}(1+\rho_{c1,b2}^{c}) N_c + \epsilon_{c1}^{l}\epsilon_{b2}^{l}(1+\rho_{c1,b2}^{l}) N_l + N_{c1,b2}^{bkg}$$

n_{b1}…. observed number in each category

N_b N_c N_g … numbers to work out

ε … Tagging efficiencies

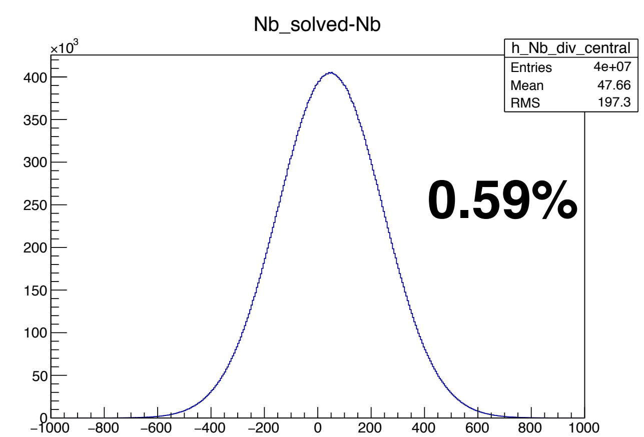
ρ… Correlation between the 1st and 2nd jets tagging efficiency, typically of the order 0.1%-1%, but can be very large (~100%-1000%) for light jets

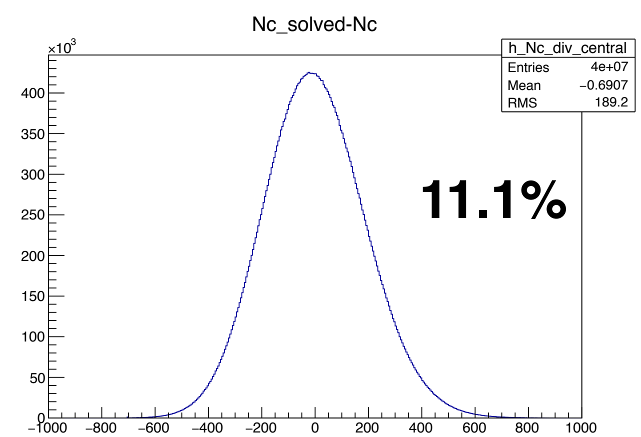
Mathematically similar to *P.L.B 401(1997) 163-175* (High precision R_b measurement in ALEPH)

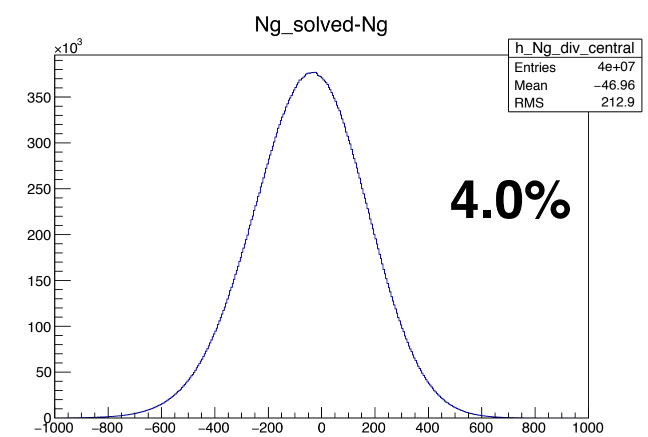
b/c tagging efficiencies from data-driven.
No need to measure b/c tagging efficiency previously.

Systematic uncertainty of the method can be characterized by the parameter presumably set.

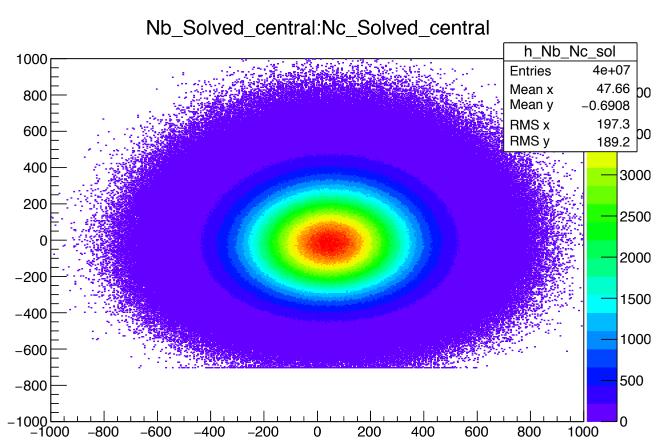
Systematic Uncertainty

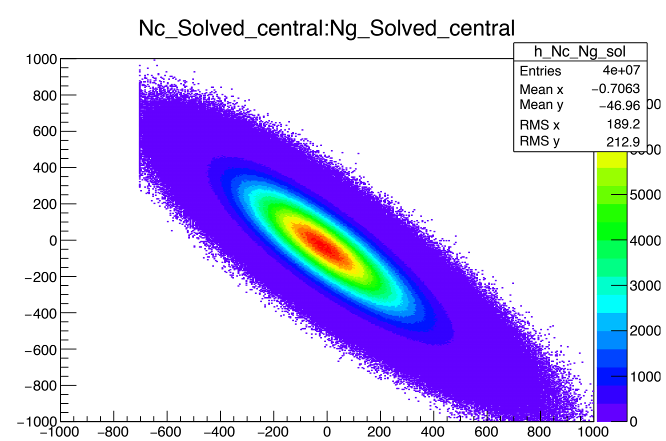


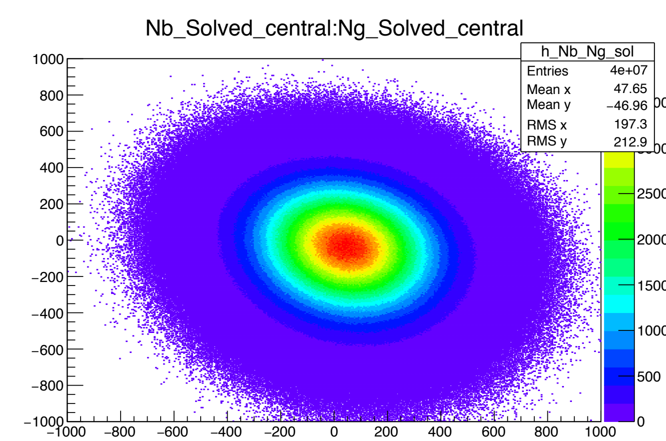




Central value deviate from 0 due to some correlation term are ignored.







N_c and N_g are statistically strong correlated

- Reasonable statistic uncertainty
- Uncertainty can be reduced by optimization(for Nc less than 10% without bkg and 11% with bkg)

Systematic Uncertainty

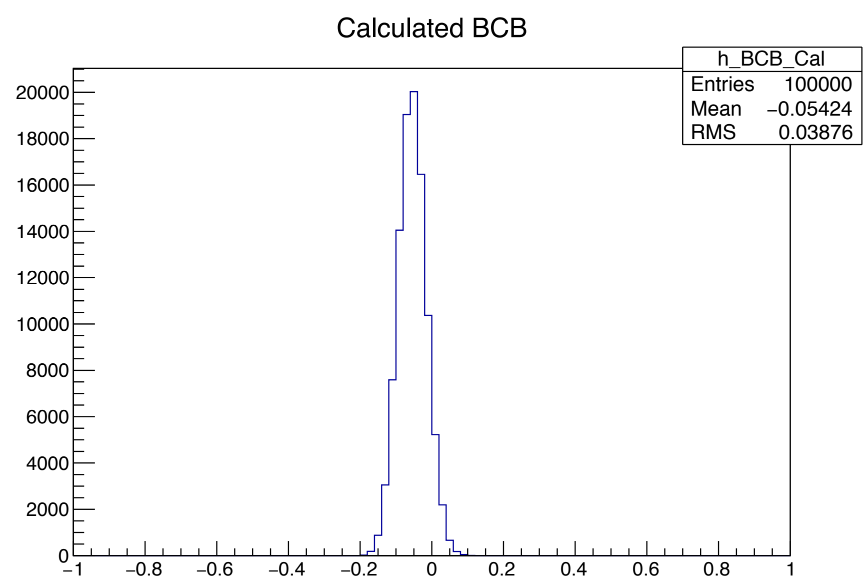
Impact of Correlation to the Results:

去掉的关联	Nb	B误差	Nc	C误差
\	33677	0	1703	0
CL	33675	-6.5E-05	1612	-0.05322
CBB	33677	-8.9E-06	1692	-0.00639
BCL	33677	-5.9E-06	1694	-0.00513
CBC	33677	-3E-06	1698	-0.00286
BB	33723	0.001375	1700	-0.00174
BCC	33677	-3E-06	1700	-0.0017
BC	33676	-3.9E-05	1703	4.7E-05
CBL	33677	2.97E-06	1707	0.002566
CC	33677	5.94E-06	1711	0.004885
BL	33525	-0.00452	1713	0.005755
CB	33677	8.91E-06	1716	0.007757
BCB	33682	0.000143	1900	0.115514

Assuming 30% sys.uncertainty on BL,CL and 10% that of BCB

Systematic uncertainty from other correlation terms estimated from bias

For simplify one can only take the correlation terms with largest impact into account.
But should use large MC sample to determine that.



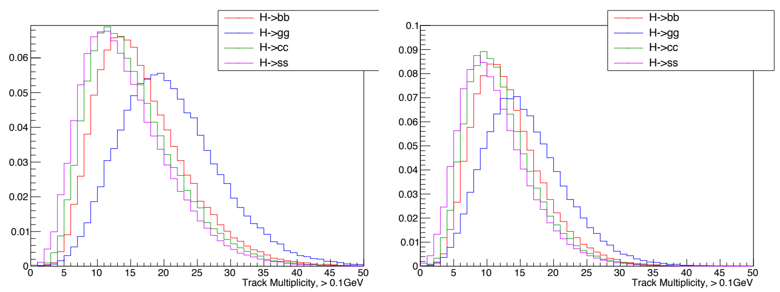
MC statistic limit is also a source of systematic uncertainty.
Can be reduced by enlarge MC statistics

Uncertainty		Nb	Nc	Ng
Statisitic		0.59%	11.1%	4.0%
Systematic uncertainty	BL	0.14%	0.18%	0.84
	CL	0.002%	2.2%	0.7%
	BCB	0.01	8.3%	2.8%
	other	0.14%	0.04%	0.89%
All		0.62%	14.0%	5.1%

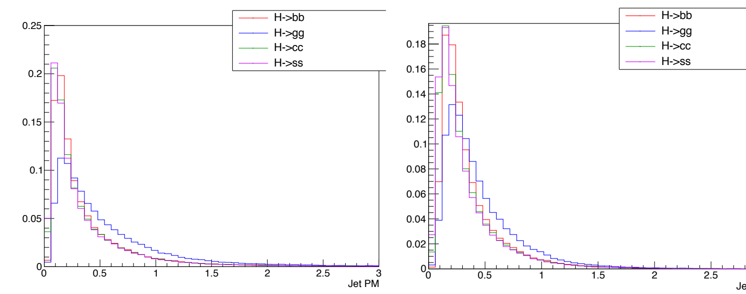
Gluon/Quark Jet Seperation

- **Gluon jets fraction is crucial to H->bb/cc/gg analysis**
- **Complex phenomenon of gluon jets have impact on other final state measurement**
- **Good performance on b/c-tagging but not consider gluon-quark jet separation yet**

Track multiplicity in Jet(already in LCFIPlus)

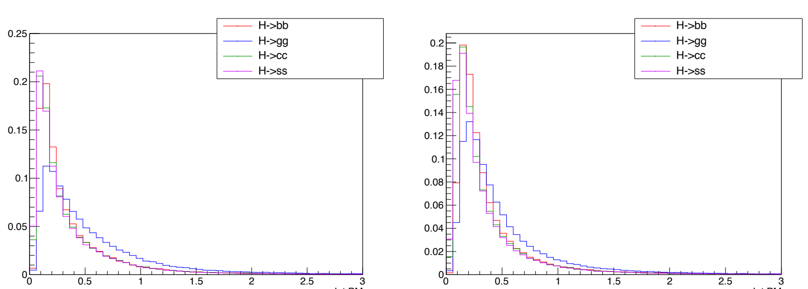


Jet Energy Momentum

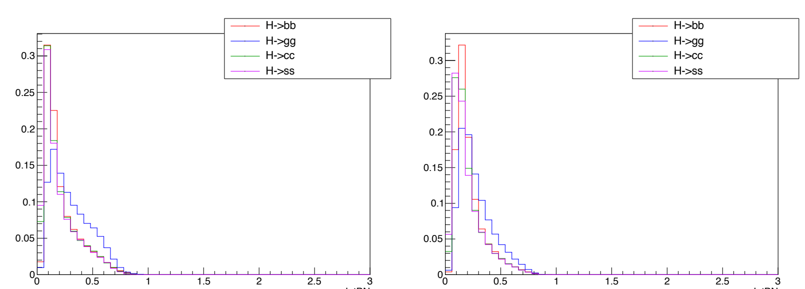


E_{pfo}*DeltaR(PFO to Jet)

Jet p_T Momentum



Jet Broadening



Discrimination power shown. Need to implement to LCFIPlus

Radial Moment, or Girth : $g = \sum_{i \in jet} \frac{p_T^i}{p_T^{jet}} |r_i|$

$$B_{jet} = \frac{\sum_i |\vec{p}_i \times \hat{n}_{jet}|}{\sum_i |\vec{p}_i|} = \frac{\sum_i |\vec{k}_T i|}{\sum_i |\vec{p}_i|}$$

Summary

- Very nice work have been done on the CEPC benchmark measurement on H->bb/cc/gg analysis, as milestone. Something to do in each channel
    - ZH->ll+jj, analysis with all SM background and apply more cuts to enhance signal/background ratio and reduce statistic uncertainty
    - In vv+jj channel, full simulation SM background should be used
    - In qqH->4jets channel, need to understand why ILC results are so different
- Methods other than template fit are studied
    - Inspiration from high precision R_b measurement
    - Statistic uncertainty are estimated; systematic uncertainty estimated in reasonable way
    - Impacts of gluon splitting on flavor tagging need to be study
- Flavor tagging will include gluon/quark jet separation to fix to the measurement

Backup

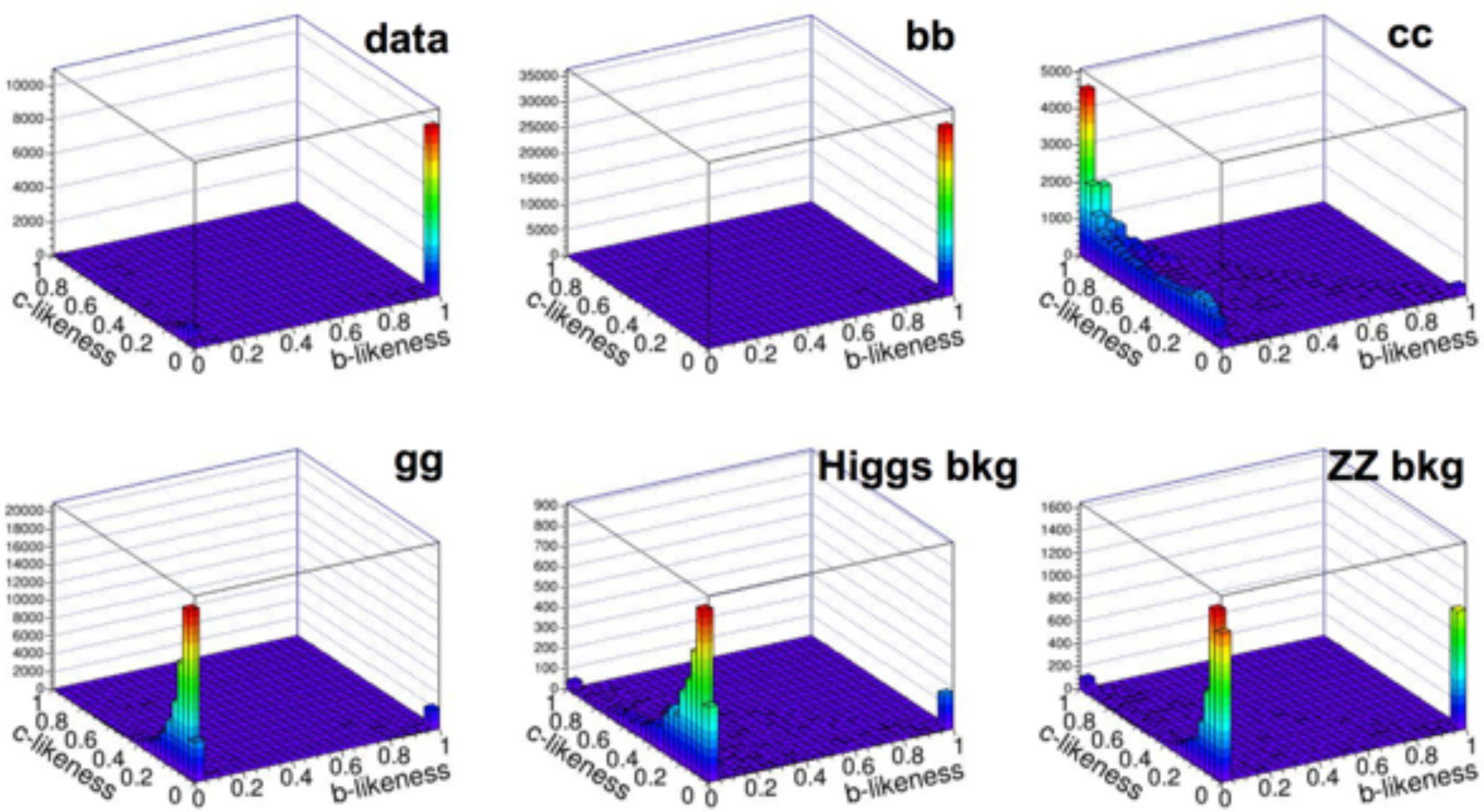
Template Fit in ll+jj Channel

Templates and data

Category	$Signal$	ZH	ZZ
Pre-selection	21300	9486	381678
$120GeV < M^{\mu^+\mu^-}_{recoil} < 150GeV$	20582	8678	35305
$70GeV < M_{\mu^+\mu^-} < 105GeV$	20121	8340	26525
$105GeV < M_{jj} < 135GeV$	16479	3336	7732

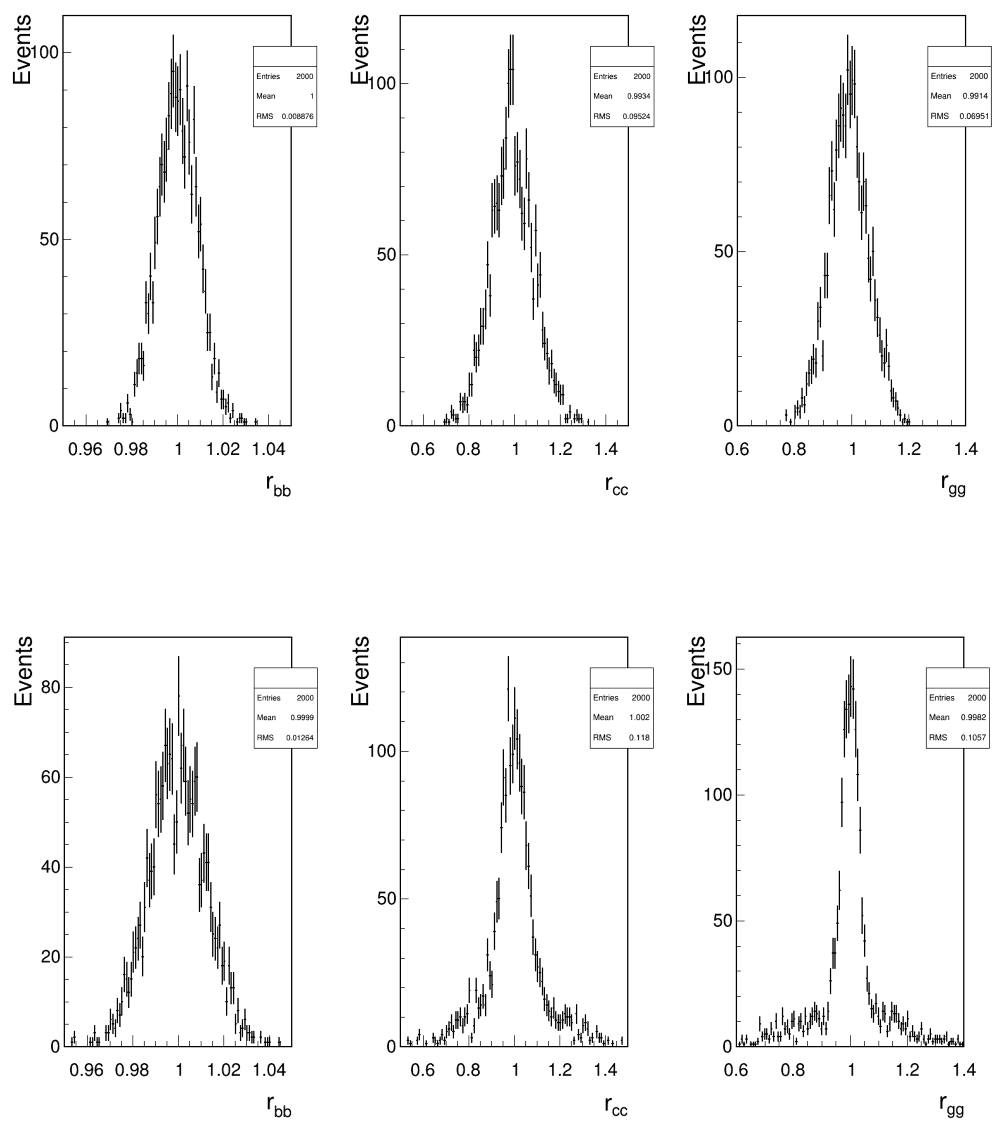
Fit target

$$L_{qq} = \frac{qq\ pair}{qq\ pair\ +\ neither\ is\ q} = \frac{x_q^1\, x_q^2}{x_q^1 x_q^2 + (1 - x_q^1)(1 - x_q^1)}\ (qq = bb, cc)$$



Full Simulated Higgs and ZZ events

Template Fit in ll+jj Channel: Toy MC Check

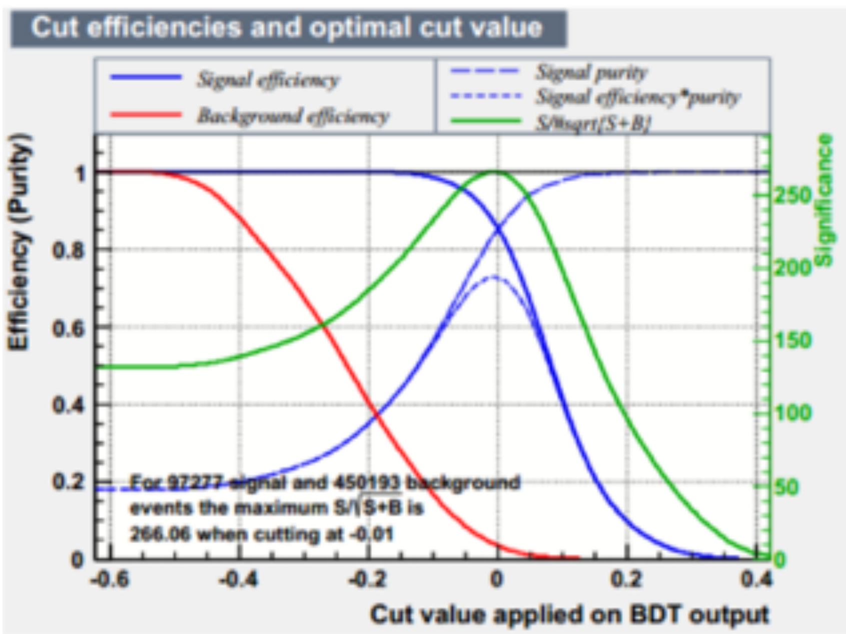


$\mu\mu$ **channel**

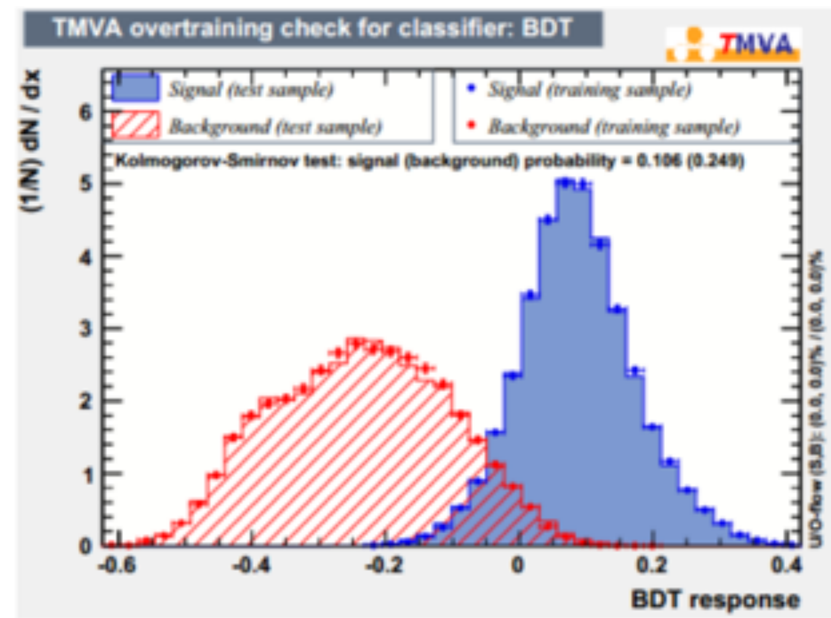
ee channel

Fit works stable

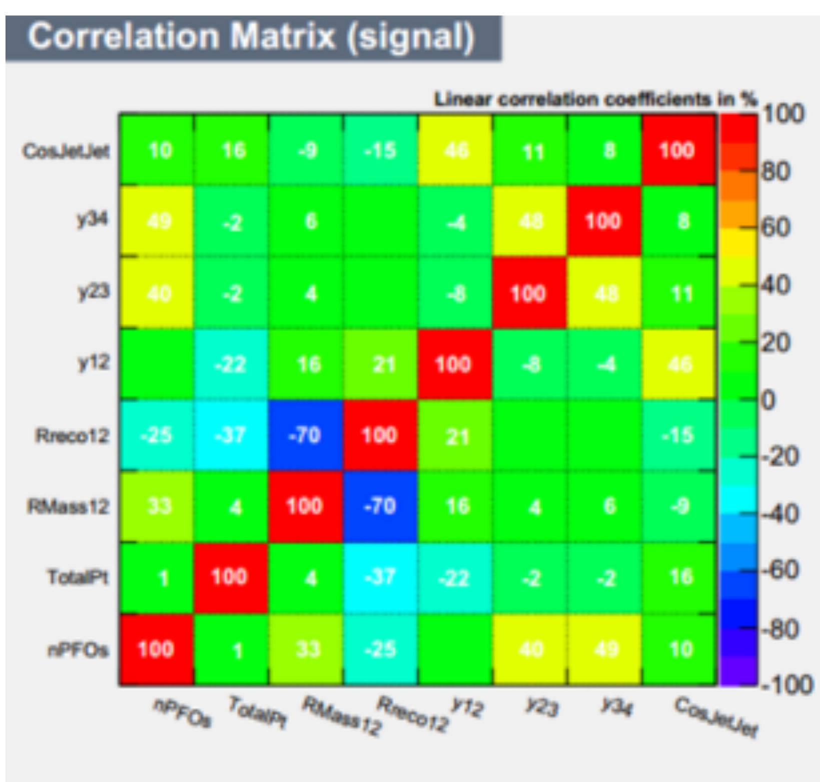
MVA in νν+jj channel

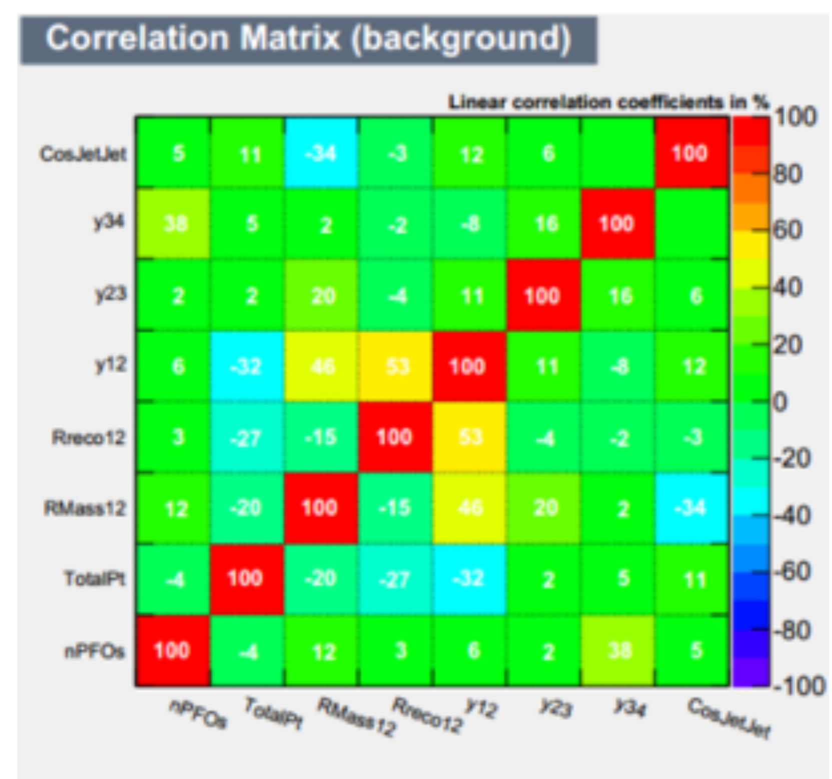


(a) Cut efficiencies and optimal cut value.

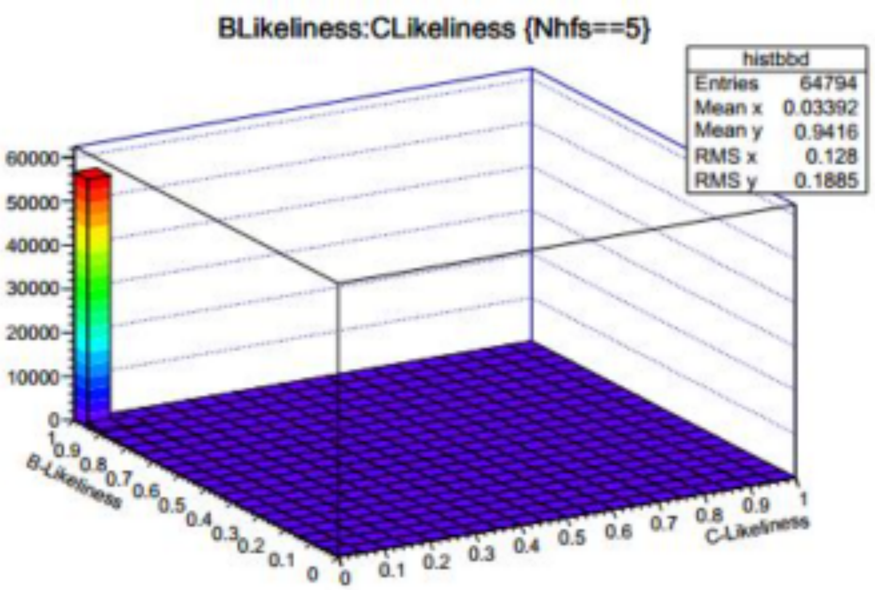


(b) Overtraining check for classifier.

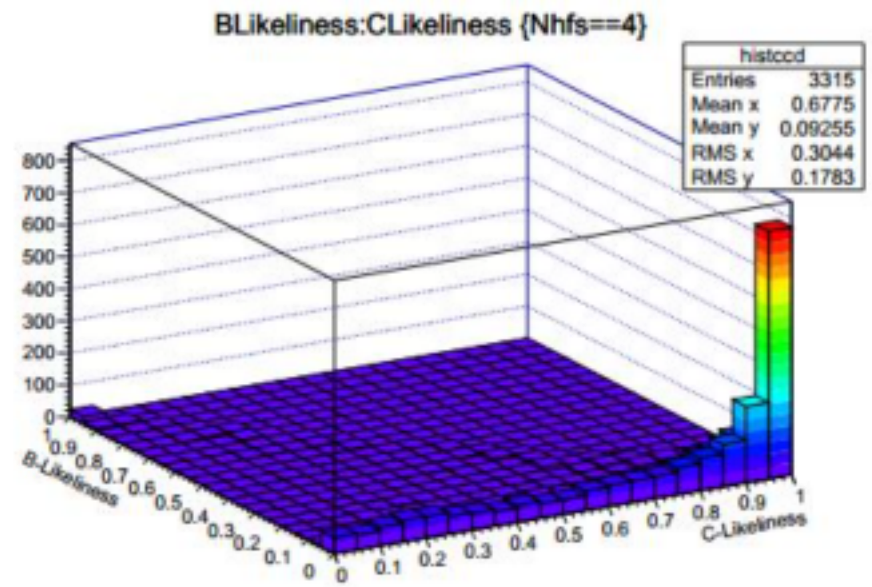




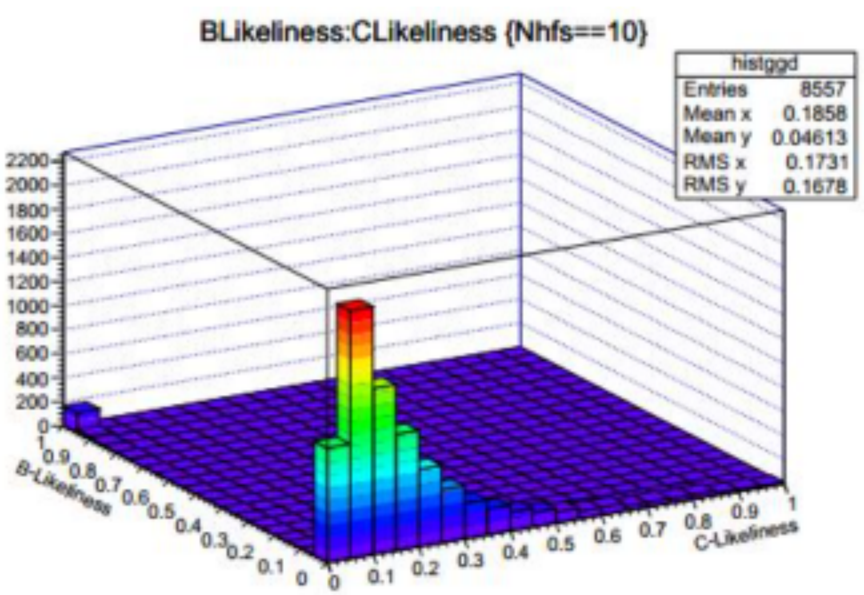
Template fit in vv+jj channel



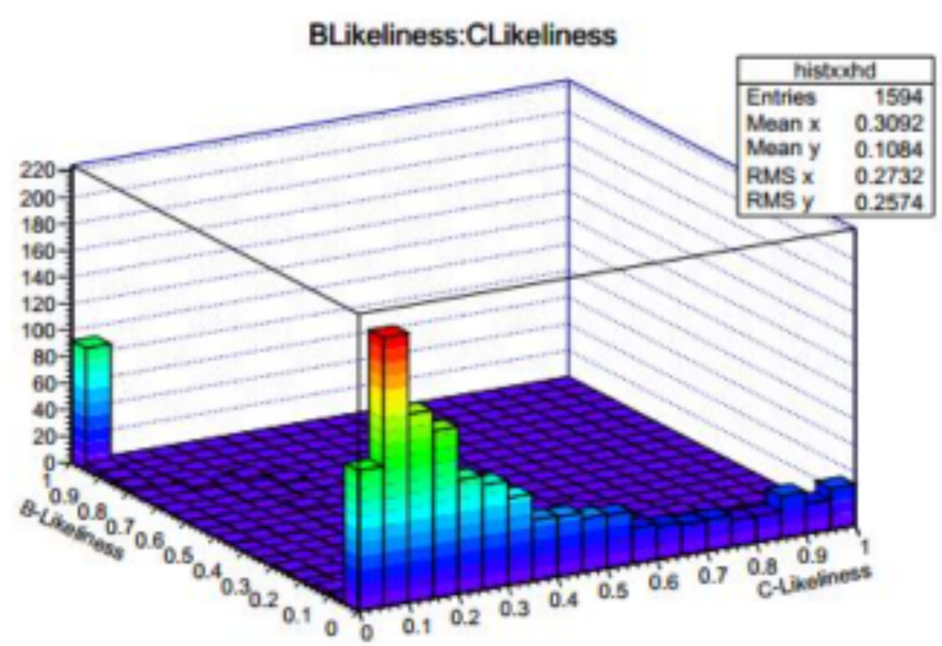
(a) The B-C Likeliness plot for $H \to b\bar{b}$



(b) The B-C Likeliness plot for $H \to c\bar{c}$

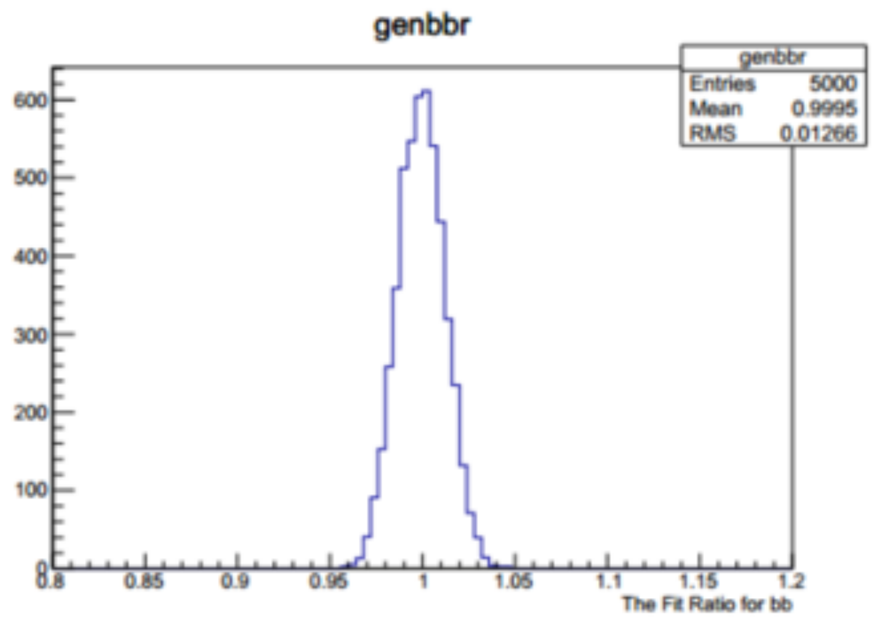


(c) The B-C Likeliness plot for $H \to g\bar{g}$



(d) The B-C Likeliness plot for $\ell\ell H$, $\nu\nu H$, qqH backgrounds

ToyMC Results in νν+jj channel



Statistics

Template ~ 2000/fb

Data ~ 500/fb

Non-Higgs background not included