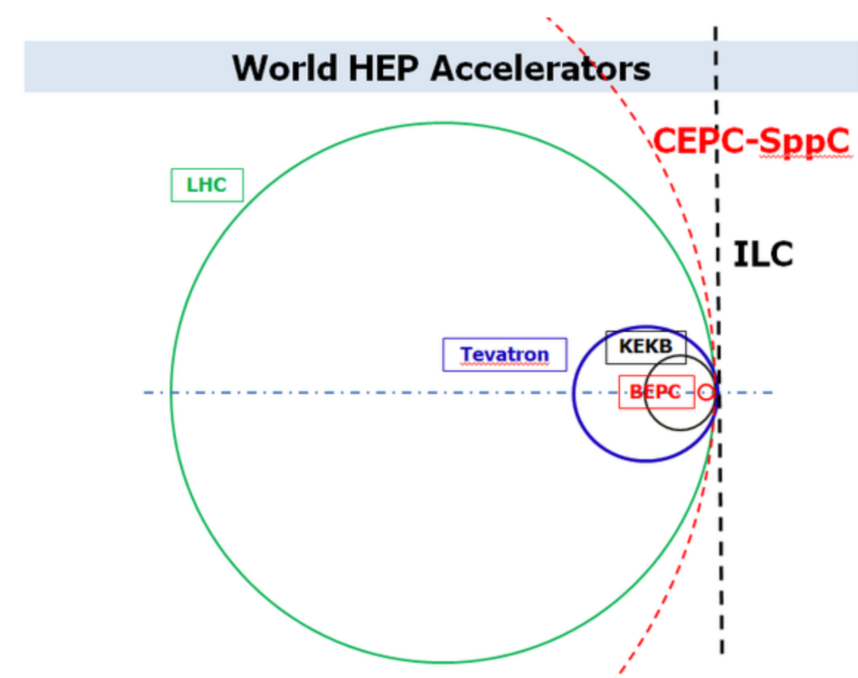
CEPC Workshop

Higgs Hadronic Decay Branch Ratio Measurement in CEPC

Yu Bai (from Southeast University, Nanjing)

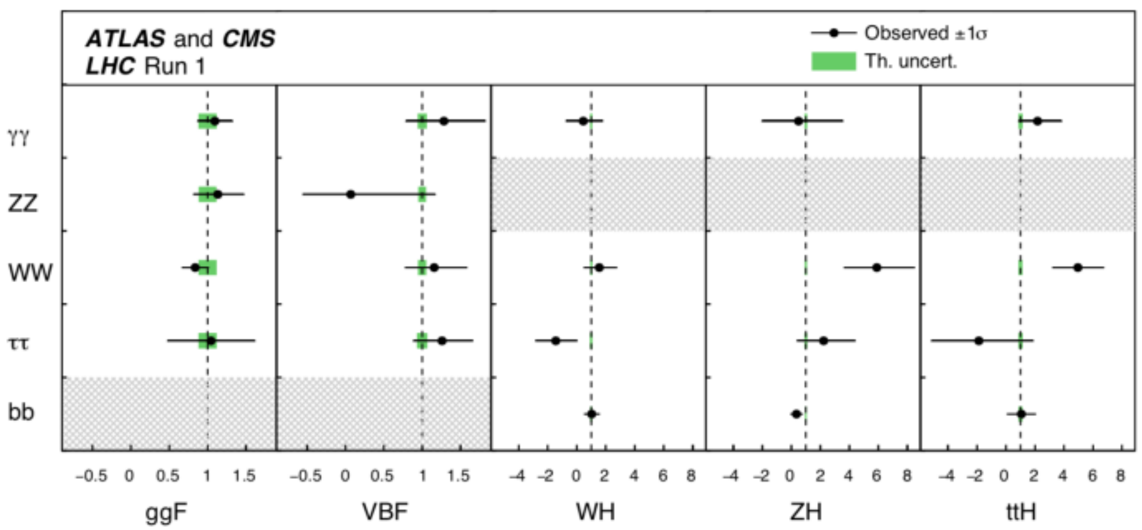
On Behalf of CEPC Physics-Software Study Group

May 25, 2018



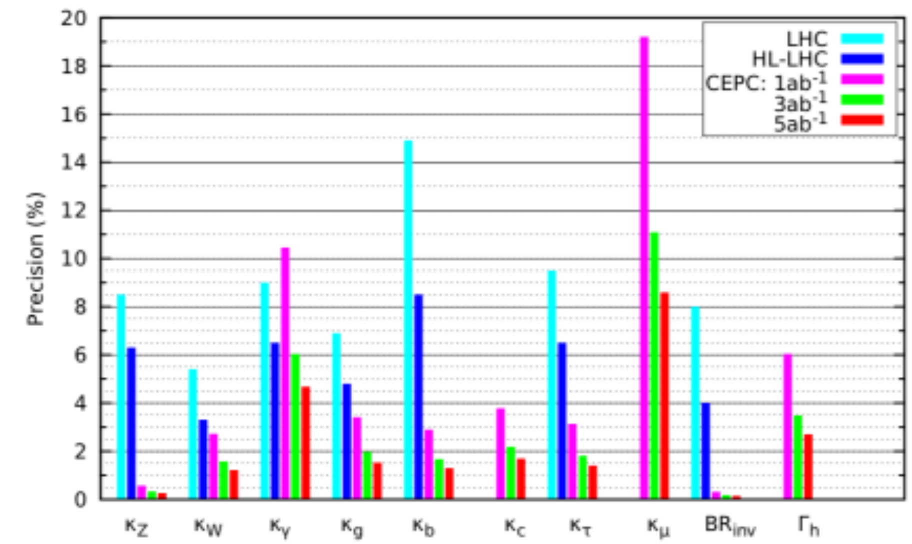
Introduction

Current Result in H->bb



$H \to b\bar{b}$	Tevatron	ATLAS Run 1	CMS Run 1	ATLAS Run 2	CMS Run 2
VH	1.6 ± 0.7	$0.52 \pm 0.32 \pm 0.24$	1.0 ± 0.5	$1.20 \pm 0.24 \pm 0.28$	1.2 ± 0.4
VBF	—	-0.8 ± 2.3	$2.8 \pm 1.4 \pm 0.8$	-3.9 ± 2.8	-3.7 ± 2.7
ttH	—	$1.4 \pm 0.6 \pm 0.8$	0.7 ± 1.9	$2.1 \pm 0.5 \pm 0.9$	$1.19 \pm 0.5 \pm 0.7$
Inclusive	—	—	—	—	2.3 ± 1.7
PDG Comb.	1.6 ± 0.7	0.6 ± 0.4	1.1 ± 0.5	1.2 ± 0.3	1.2 ± 0.4

An improvement of more than 1 order of magnitude in precision at CEPC

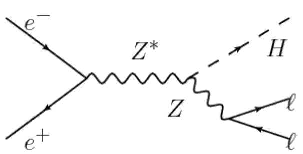
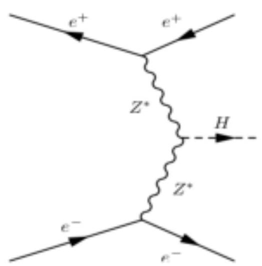
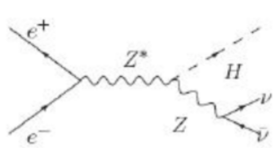
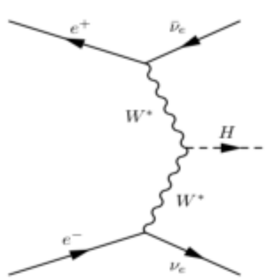
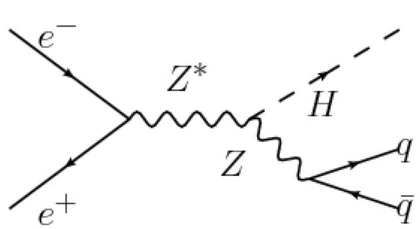


Higgs hadronic decay:
Benchmark channel to understand the performance in tracking, vertex finding, jet clustering and flavor tagging

Review of the Analysis

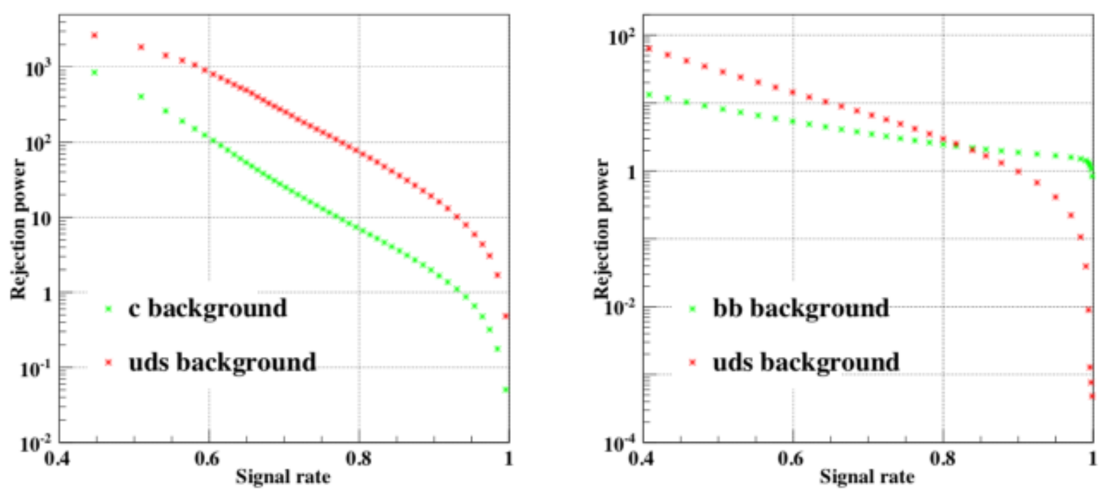
- **2016-2017: H->bb/cc/gg Analysis in qqH/ννH/eeH/μμH**
  - **Demonstrate the capability of flavor tagging and jet clustering in achieving the precision**
  - **Flavor tagging are implemented with template fit method**
  - **Will be included in white paper/CDR**
- **From summer in 2017, llH channel redone with new method:**
  - **2-D template fit replaced by a 3-D fit**
  - **Systematic uncertainty considered**
  - **Summarize in note, paper draft reviewed**
- **Latest update:**
  - **Redo analysis with 3T samples**
  - **New techniques to improve the performance of analysis**

Event Selection in llH/$\nu\nu$H/qqH

Signal:	Backgrounds:	Preselection:	Flavor Tagging
ZH->ll+jj  	**ZZ semi-leptonic($\mu\mu$jj)** **Single Z-leptonic(eejj)**	**Lepton Pair Invariant mass** **Lepton Recoil mass** **Jets Invariant mass** **Higgs Polar angle**	
ZH->$\nu\nu$+jj  	**ZZ semi-leptonic** **(one Z invisible decay)** **Single Z semi-leptonic** **WW/SW semi-leptonic**	**Missing Energy,p_T** **Jets invariant/recoil mass** **Jet Multiplicity(yth-value)** **Angle between jets** **MVA applied**	**Template Fit**
ZH->multi-jets 	**quark pair production** **ZZ/WW hadronic**	**Total Energy** **Jet Multiplicity** **Jet Paring(jet invariant mass and angular distribution)** **MVA applied**	

Results of llH/$\nu\nu$H/qqH

Performance of multi-variable based flavor tagging :

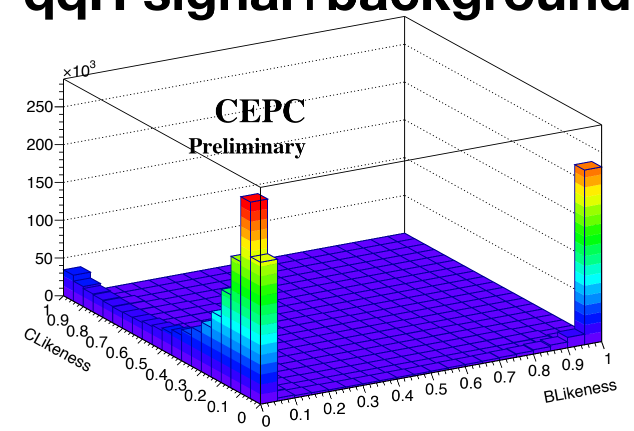


(a) (b)

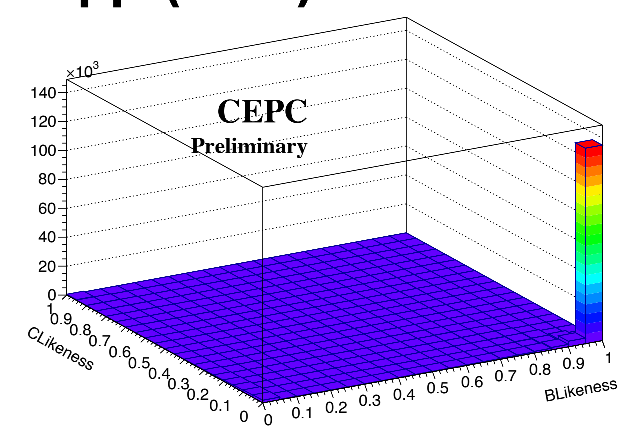
Template Fit:

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

qqH signal+background



qqH(->bb)



Results from full simulation study:

Table 6. Expected relative precision on $\sigma(ZH) \times$ BR for the $H \to b\bar{b}$, $c\bar{c}$ and gg decays from a CEPC dataset of 5 ab^{-1}.

Z decay mode	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$	Comments
$Z \to e^+e^-$	1.3%	14.1%	7.9%	CEPC study
$Z \to \mu^+\mu^-$	1.0%	10.5%	5.4%	CEPC study
$Z \to q\bar{q}$	0.4%	8.1%	5.4%	CEPC study
$Z \to \nu\bar{\nu}$	0.4%	3.8%	1.6%	CEPC study
Combined	0.3%	3.2%	1.5%	

Results in preCDR:

Decay mode	$\sigma(ZH) \times$ BR	BR
$H \to b\bar{b}$	0.28%	0.57%
$H \to c\bar{c}$	2.2%	2.3%
$H \to gg$	1.6%	1.7%

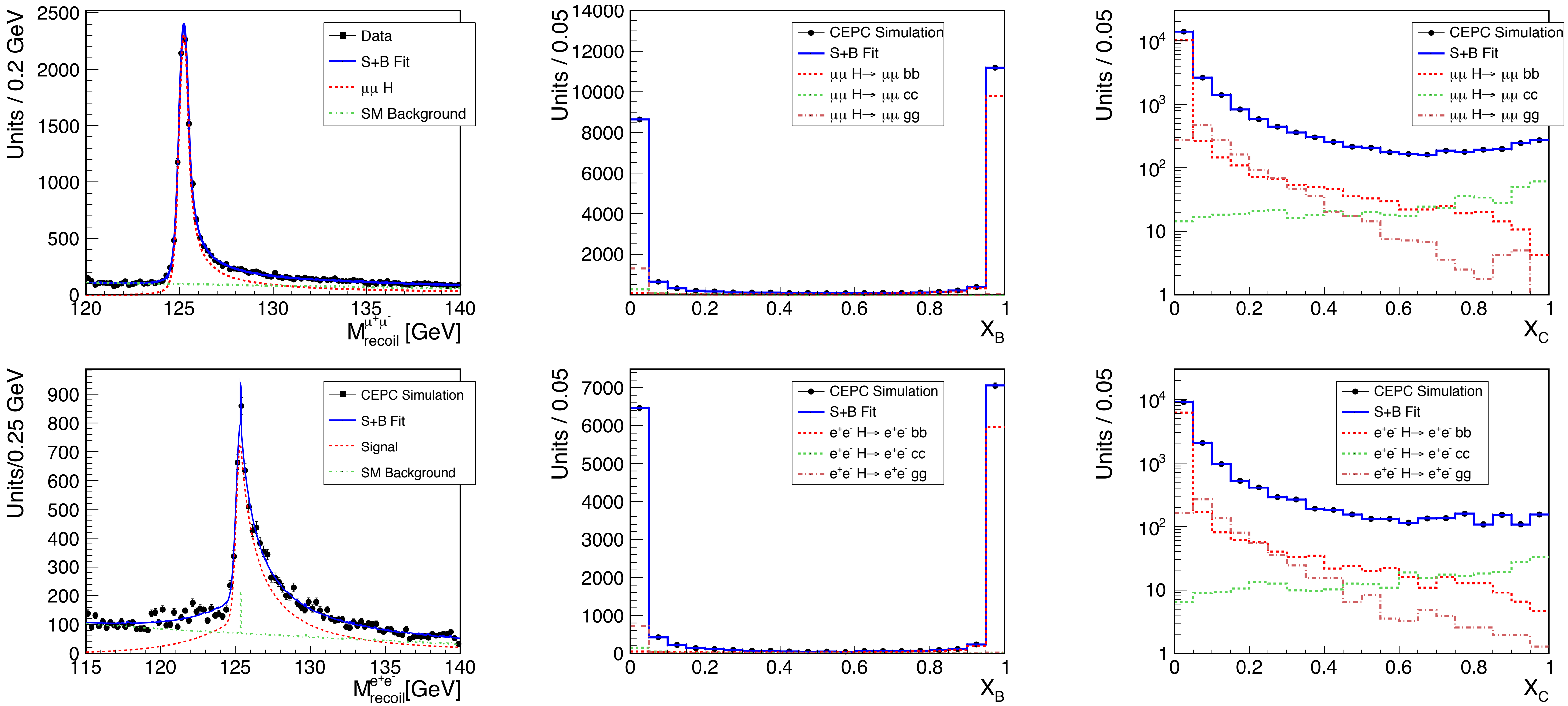
- Consistent with pre-CDR result
- Demonstrate the capability to achieve expected performance

llH Analysis with 3D Fit

- Dominant background in llH analysis estimated from ‘data’:
  - $\mu\mu$H channel: ZZ*/Zγ*-> $\mu\mu$qq
  - eeH channel: ee+qq
- Analysis independent of MC prediction of dominant backgrounds:
  - These backgrounds have different lepton pair recoil mass spectrum
  - Extract the backgrounds’ yield by including recoil mass spectrum in the fit

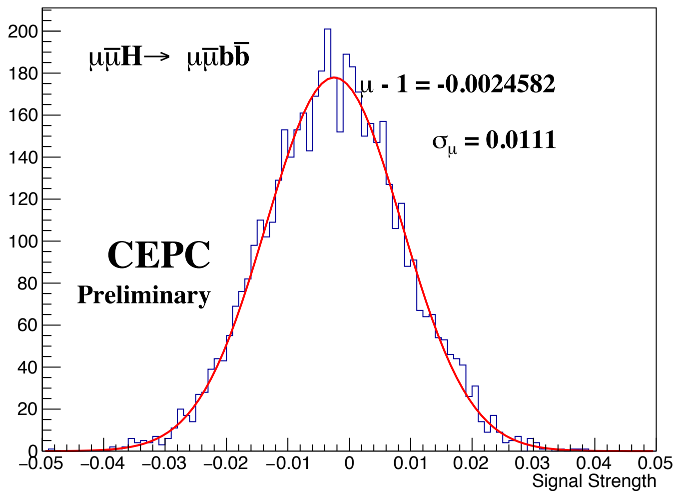
3D Fit

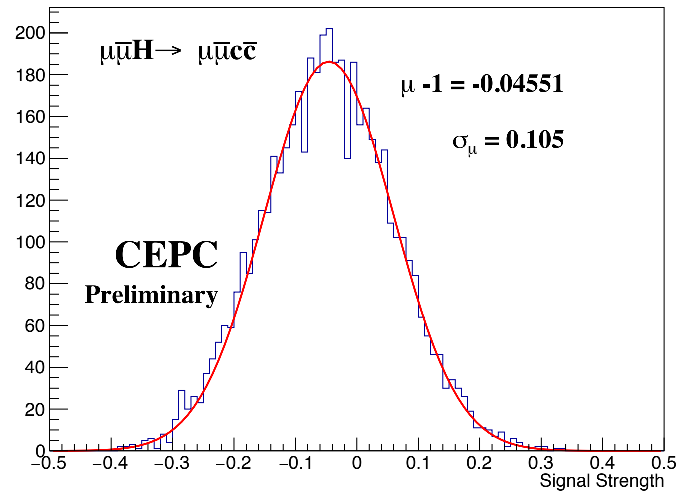
$$PDF^{3D}(X_B, X_C, M_{recoil}) = PDF^{flavor}(X_B, X_C) \times PDF^{recoil\_mass}(M_{recoil})$$

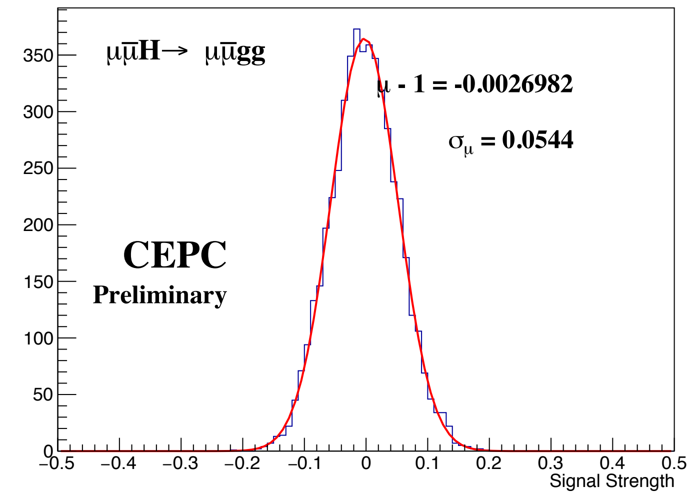


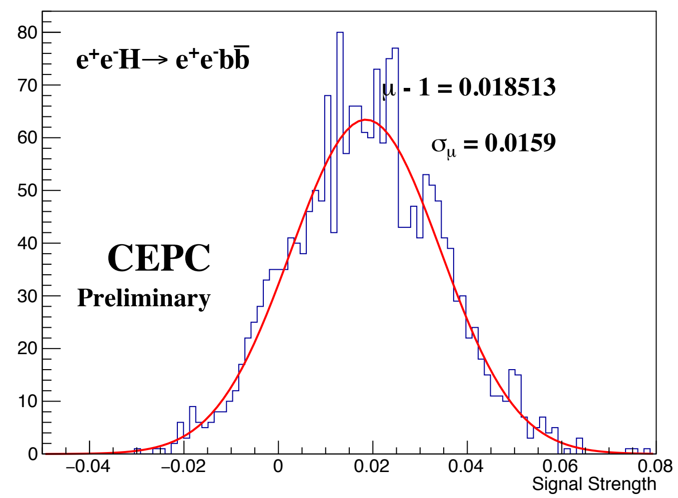
- Recoil mass of signal: Crystal ball + double side exponential
- Recoil mass of dominant background: 1 order Chebychev polynomial
- Background and signal model describe the simulated data well

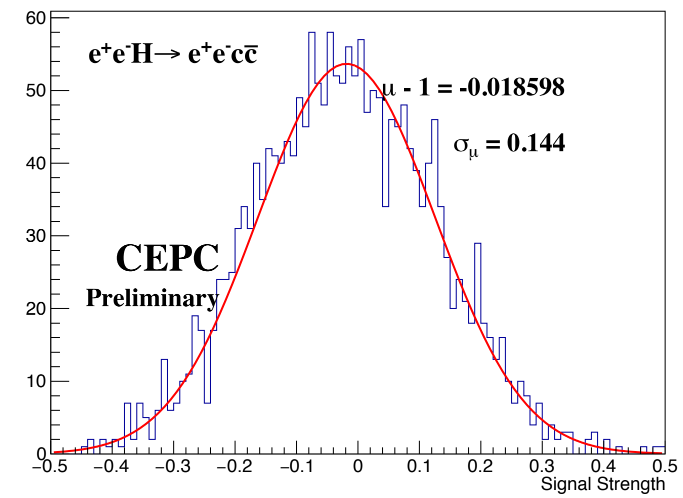
Statistic Uncertainties

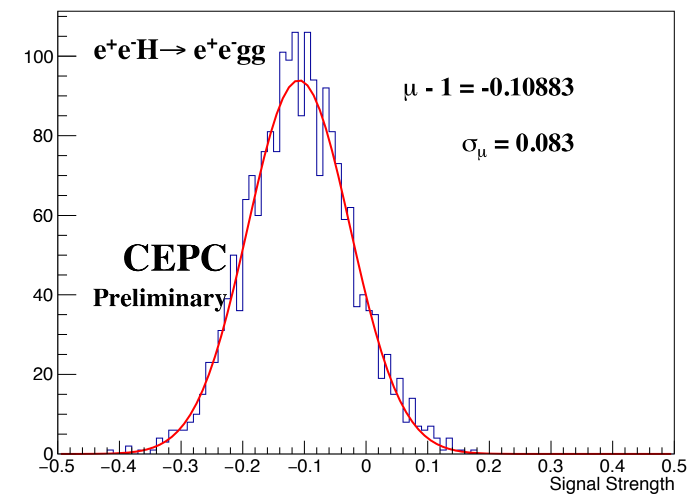








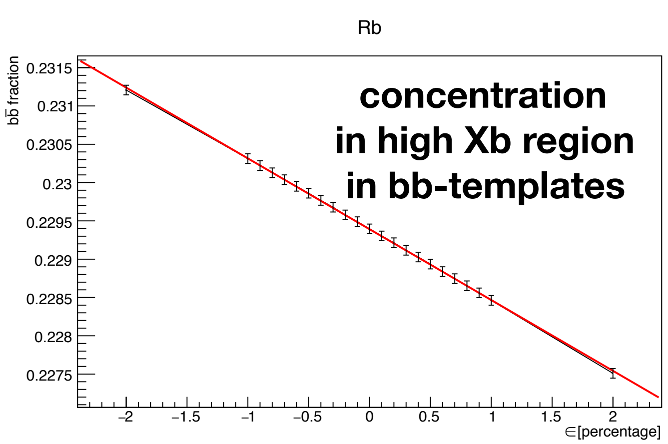


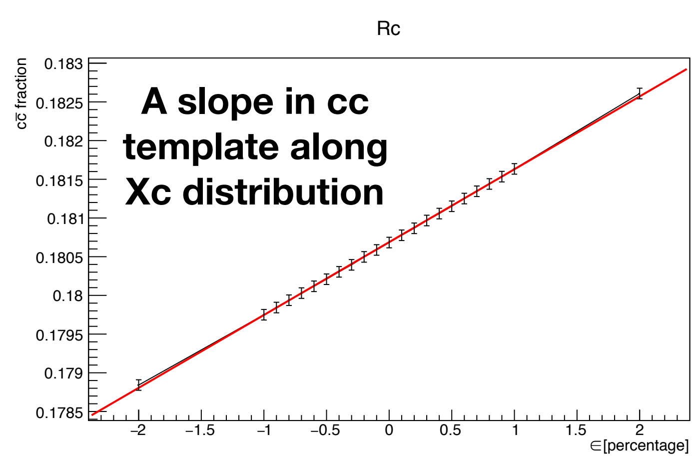


- ToyMC generate data fluctuate according to statistic uncertainty
- Roughly get the same statistic uncertainty as before

Systematic Uncertainty

- Flavor tagging systematic uncertainty directly caused by the bias in templates
- Flavor tagging systematic uncertainty are estimated in the scenario calibration with $\mu\mu$qq
- The precision of calibration are limited by $\mu\mu$qq statistic uncertainty and the knowledge of its flavor components
- Typical bias are considered in each template





Output of flavor fraction linear to typical bias: linearity can be used to extract uncertainty

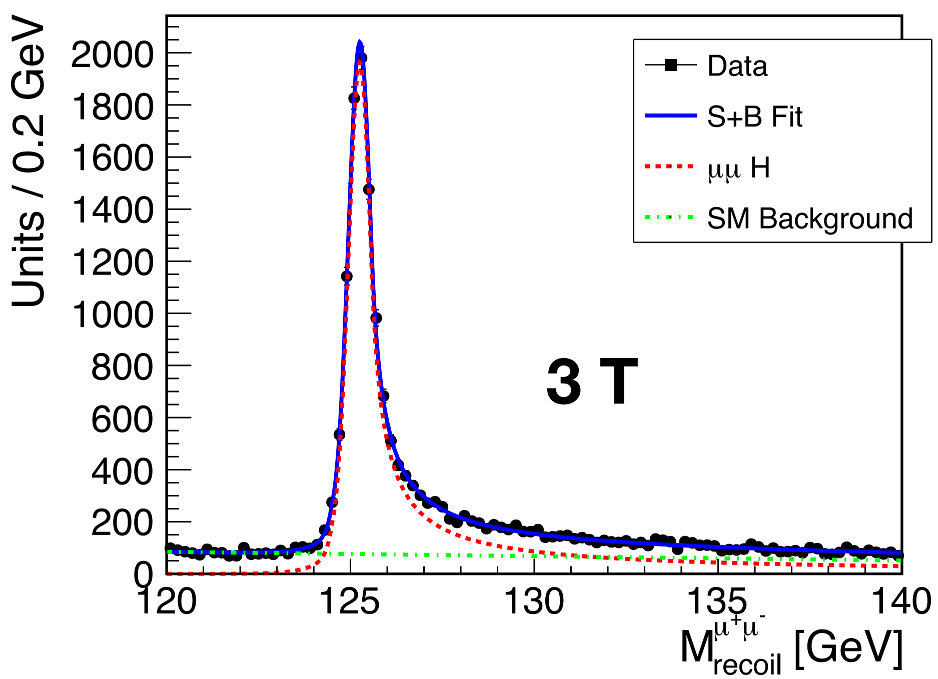
Only a methodology study. Need to be fulfilled with calibration in real data

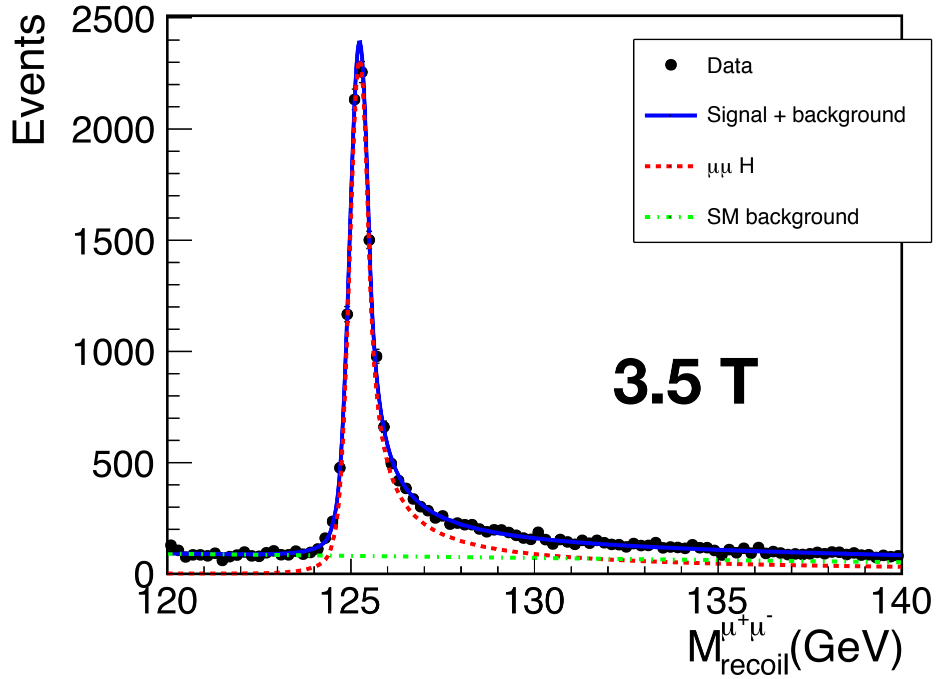
Systematic uncertainty in llH channel

	$\mu^+\mu^- H$			$e^+e^- H$		
	1.11%	10.5%	5.44%	1.59%	14.4%	8.3%
	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$
Fixed Background	-0.17% +0.06%	+4.1% -4.2%	7.6%	-0.17% +0.06%	+4.1% -4.2%	7.6%
Event Selection	+0.68% -0.20%	+0.43% -1.1%	+0.71% -1.7%	+0.68% -0.20%	+0.43% -1.1%	+0.71% -1.7%
Flavor Tagging	0.67%	10.4%	1.1%	0.67%	10.4%	1.1%
Non uniformity	0.016%			0.016%		
Combined	+0.96% -0.72%	+11.2% -11.3%	+7.7% -7.9%	+0.96% -0.72%	+11.2% -11.3%	+7.7% -7.9%

eeH uncertainty extrapolate from mumuH

Analysis with 3T sample





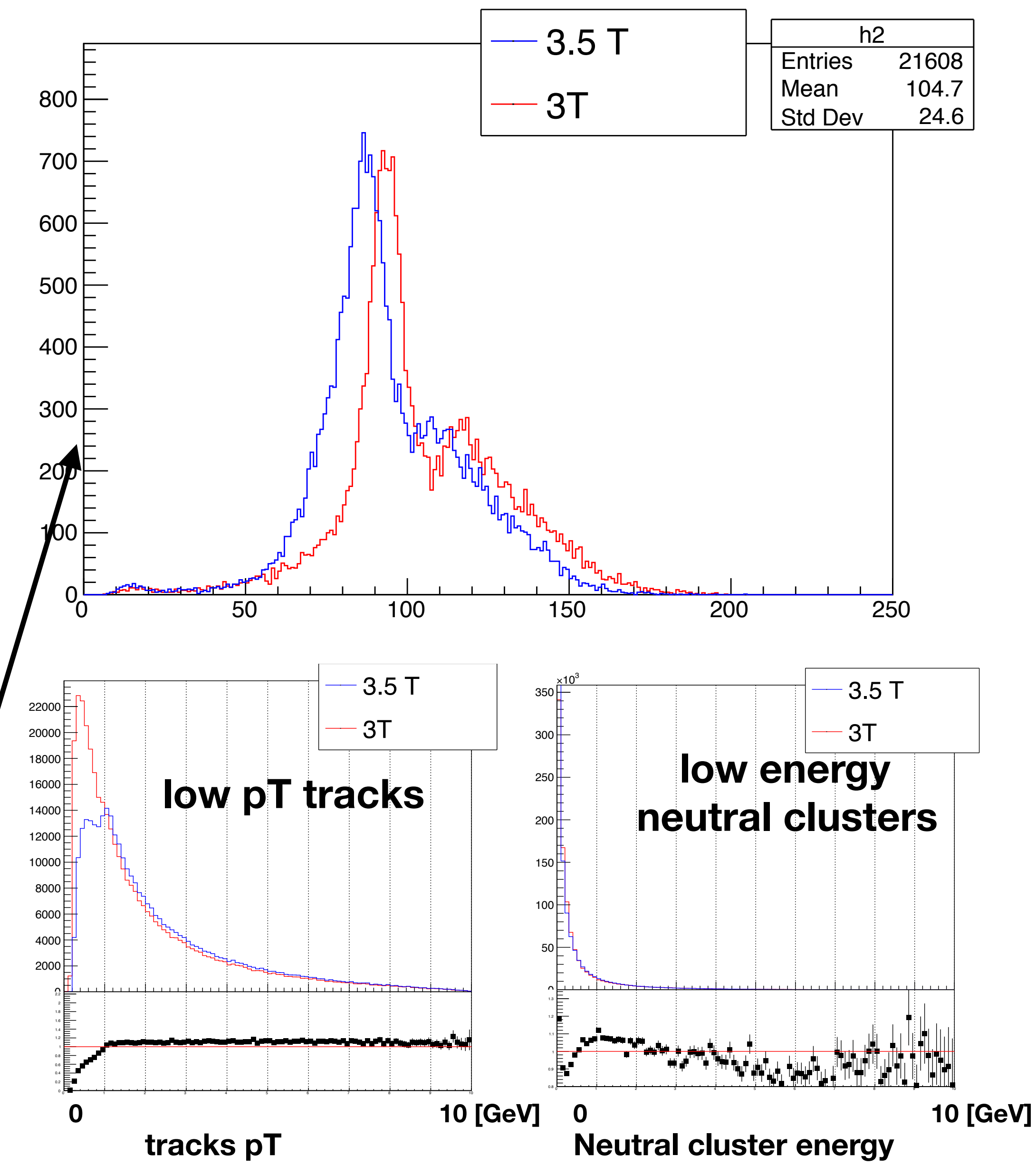
	value	error	central value
nHbb	1.095×10^4	139	10806
nHcc	542	60.1	497
nHgg	1413	90.2	1472
sigma	0.323	0.0065	

	value	error	central value
nHbb	1.123×10^4	132	11188.2
nHcc	513	55.8	497
nHgg	1545	84.3	1472
sigma	0.269	0.0054	

- No obvious change in the fitted results
- The recoil mass spectrum change as expected: 0.323/0.269 ≈ 3.5/3.0, mumu recoil mass get width

3T sample disagreement in soft region

	3T	3.5T	Eff Ratio
Filter	4.5%		
FSClasser	Unknown		
cos θ_Z	68.0%	68.6%	0.99
cos $\theta_{\mu\mu}$	89.1%	89.1%	1.00
$M_{\mu\mu}$	88.4%	89.7%	1.00
M_recoil	44.8%	45.3%	0.99
2J+Lep_Veto	97.1%	98.4%	0.99
JetnPFO	99.3%	97.4%	1.02
cos θ_{JJ}	91.9%	92.4%	0.99
M_{JJ}	94.5%	85.8%	1.10
y-value	93.2%	94.0%	0.99



- It seems the disagreement is from low energy neutral clusters
- Problems in survey, will be cleared soon

Technic development : Convolutional NN in Jets

- We use CNN to separate H->qq and H-> ZZ*/WW*->qqqq

Take image of Energy Distribution

65×65 pixels in detectable region

64 5×5 filters
4×4 pooling

14×14 pixels
64 images

64 5×5 filters
2×2 pooling

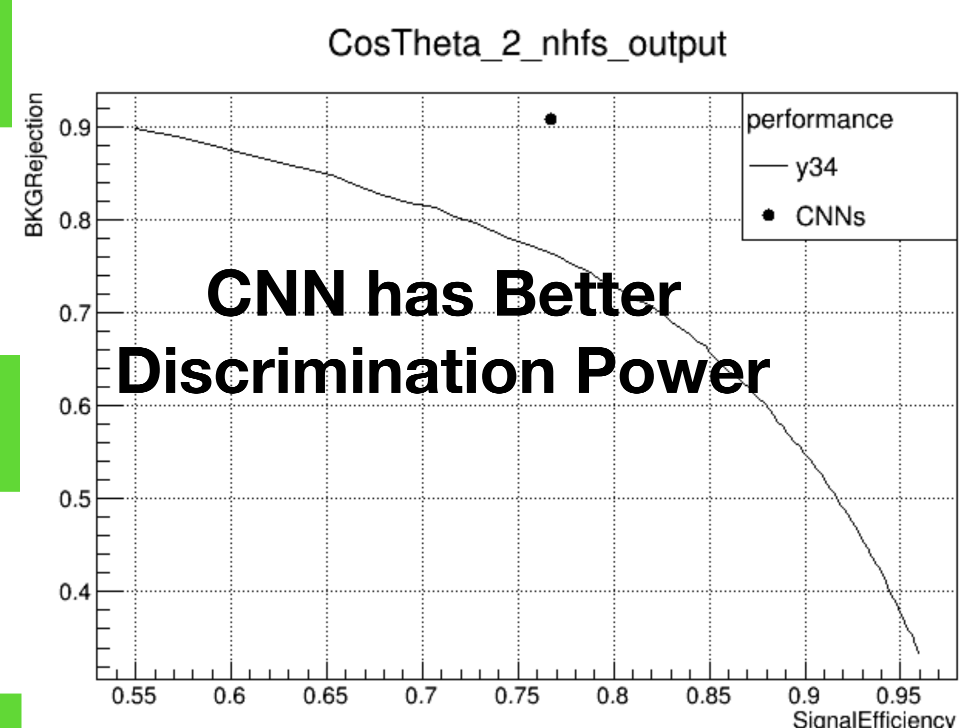
5×5 pixels
64 images

64 5×5 filters
2×2 pooling

64 Single Pixels images

64×128 Condense layer

Logistic Regression to 2 classes



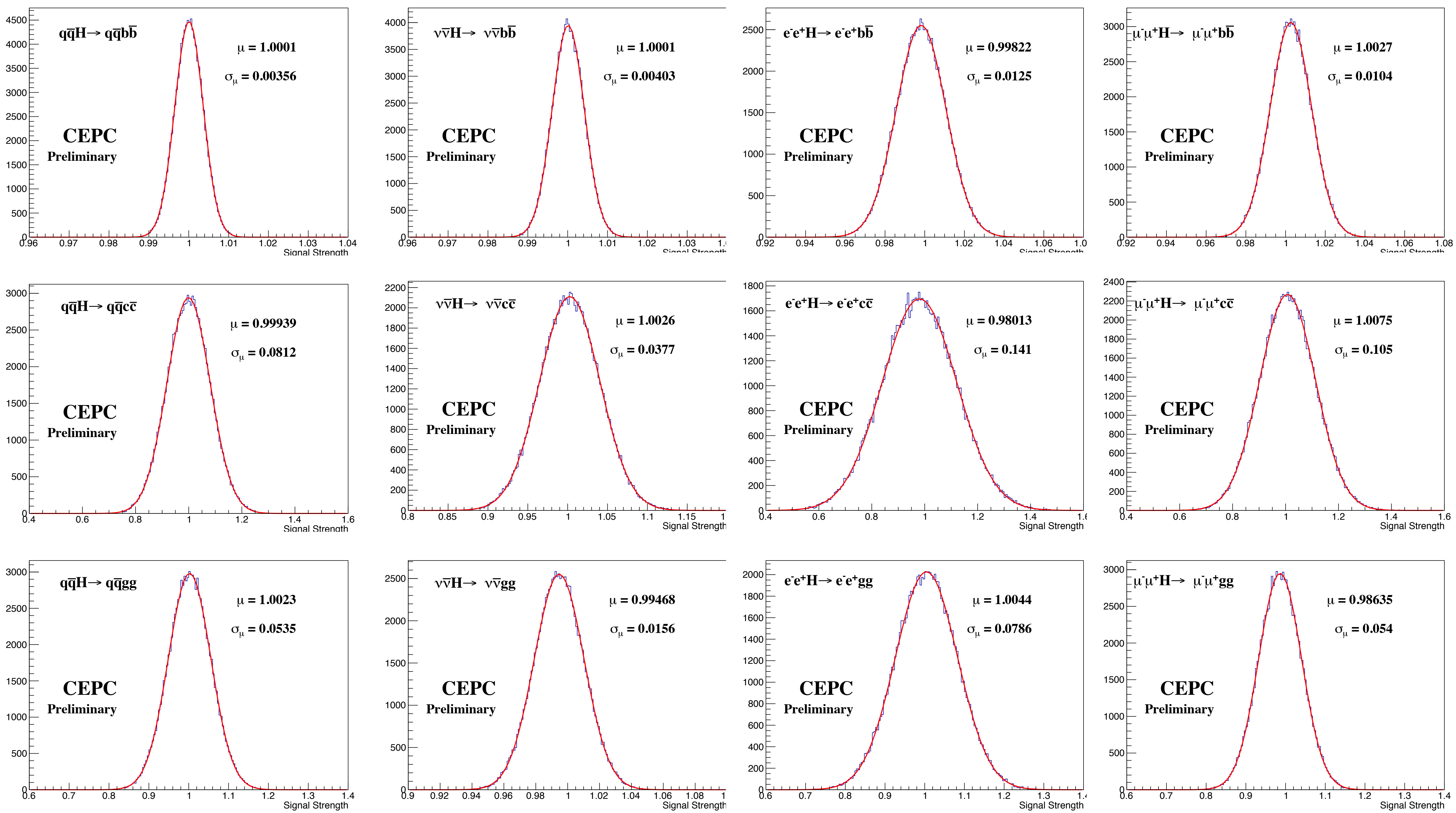
- **CNN provides new to extract character in event with jets**
- **Should be useful need to know the multiplicity of jets: Higgs->qq, QCD measurements etc.**
- **Would like to know other types of character can be extracted: try to use in qqH channel**

Summary

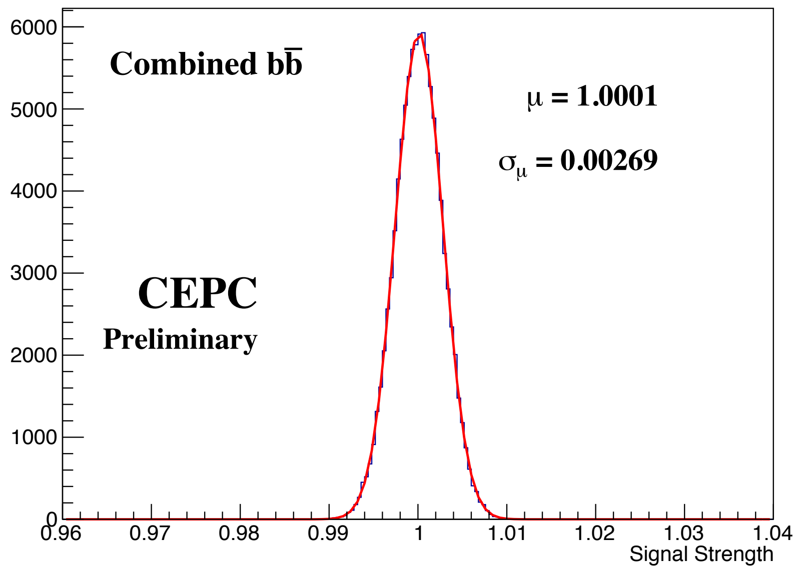
- Study of Higgs hadronic decay measurement in eeH/$\mu\mu$H/$\nu\nu$H/qqH channel are done with full simulation sample, demonstrating the capability to achieve expected performance in CEPC
- eeH/$\mu\mu$H redone with less MC-dependent way, systematic uncertainties studied in terms of methodology
- Analysis with 3 Tesla magnetic has been studied. No significant change in final results but need to find out the reason of disagreement with previous sample
- New analysis technology are developed and it is helpful to the analysis in future.

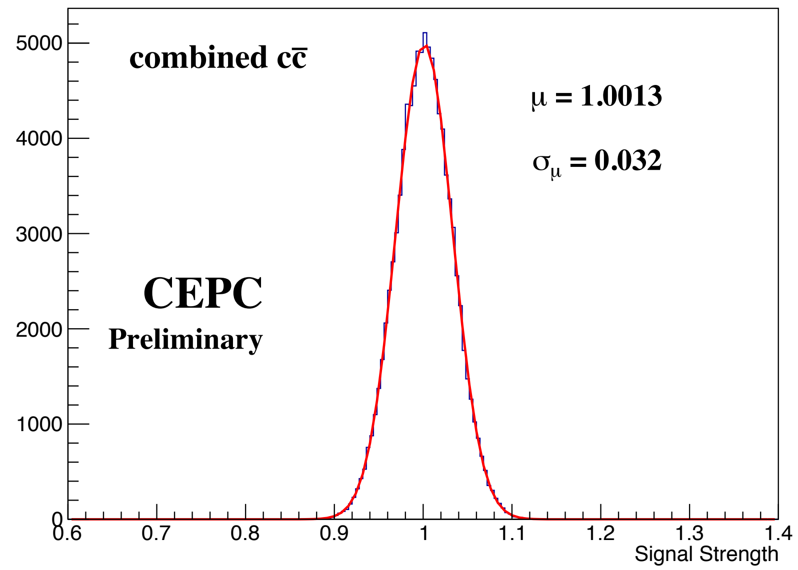
Back Up

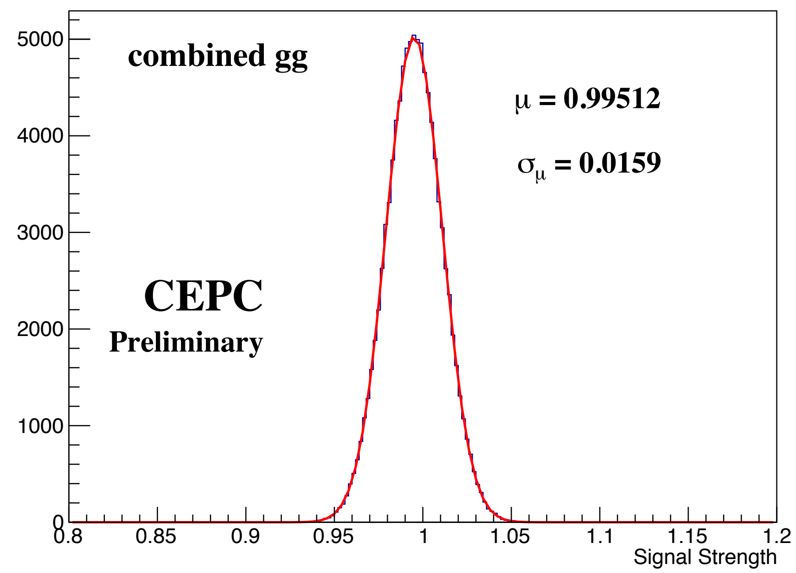
Uncertainty of qqH/$\nu\nu$H/eeH/$\mu\mu$H



Combined Fit of qqH/$\nu\nu$H/eeH/$\mu\mu$H







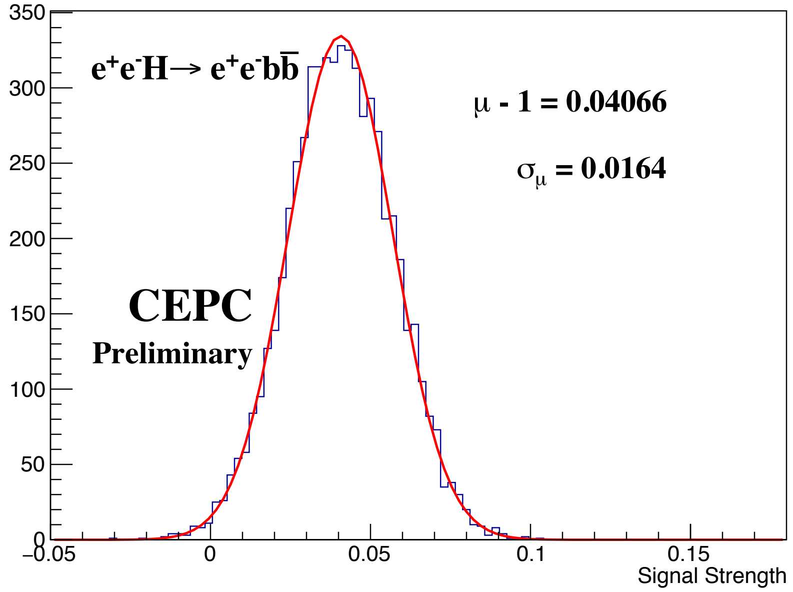
3D Fit Result-$\mu\mu H$

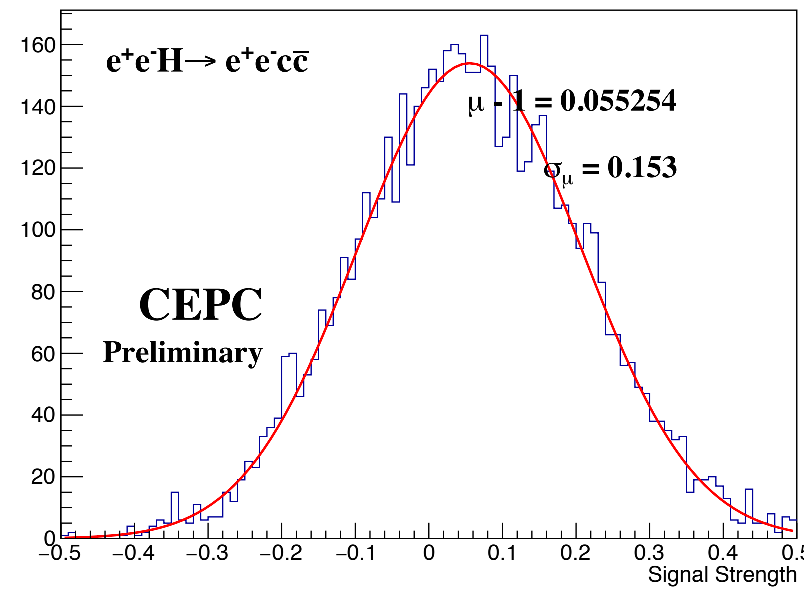
```
FCN=-233039 FROM HESSE     STATUS=OK            131 CALLS        684 TOTAL
                     EDM=0.000150386    STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                                INTERNAL      INTERNAL
 NO.   NAME      VALUE            ERROR       STEP SIZE       VALUE
  1  C            1.56571e-02   7.64113e-03   1.04870e-03  -1.17248e+00
  2  a           -1.02613e+00   2.97467e-02   7.60479e-05  -1.02794e-01
  3  a1          -3.35677e-01   2.41562e-02   1.44428e-03  -3.35677e-01
  4  mean         1.25238e+02   5.49756e-03   3.02347e-04   1.53414e-01
  5  n            9.28731e-01   2.66452e-02   9.30339e-05   2.89863e-01
  6  nHbb         1.11897e+04   1.32936e+02   1.52103e-04   1.19208e-01
  7  nHcc         4.94687e+02   5.59896e+01   3.04892e-04  -1.12265e+00
  8  nHgg         1.50519e+03   8.46824e+01   4.29695e-04  -4.09559e-01
  9  nbkg         1.55884e+02   2.88976e+01   1.26565e-03  -8.90024e-01
 10  nzzsl_mu_bb   1.55457e+03   8.94563e+01   3.09297e-04  -5.89637e-01
 11  nzzsl_mu_cc   1.56988e+03   6.19408e+01   2.51995e-04  -5.84381e-01
 12  nzzsl_mu_uds   4.63830e+03   8.96344e+01   6.29296e-04  -5.65984e-01
 13  sigma        2.69879e-01   5.43925e-03   1.51060e-04  -6.12975e-01
                               ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR= 13    ERR DEF=0.5
ELEMENTS ABOVE DIAGONAL ARE NOT PRINTED.
 5.858e-05
-2.806e-05  8.849e-04
-9.440e-06  1.615e-04  5.835e-04
 7.919e-07 -8.649e-05 -9.962e-06  3.022e-05
 1.813e-05  6.713e-04  2.332e-04 -5.644e-05  7.100e-04
 4.042e-02 -4.443e-01 -6.539e-01  2.807e-02 -9.041e-01  1.767e+04
-6.706e-03 -3.640e-02 -6.717e-02  3.880e-03 -1.235e-01 -1.472e+02  3.136e+03
-2.760e-03 -5.241e-02 -1.421e-01  6.385e-03 -2.942e-01  5.379e+02 -7.552e+02  7.174e+03
-1.884e-03  3.416e-03 -3.791e-02  3.034e-04  1.579e-02 -9.437e+01 -2.146e+01 -8.302e+01  8.377e+02
-4.032e-02  4.555e-01  6.874e-01 -2.917e-02  9.379e-01 -6.036e+03 -2.457e+02 -9.486e+02 -9.091e+01  8.005e+03
 8.055e-03  4.149e-02  8.677e-02 -4.521e-03  1.507e-01 -2.226e+02 -1.427e+03 -3.805e+01 -8.248e+01  1.232e+02  3.837e+03
 3.190e-03  3.507e-02  1.303e-01 -5.063e-03  2.223e-01 -5.088e+02  3.275e+02 -3.107e+03 -3.058e+02  7.553e+02 -5.760e+02
 3.190e-03  3.507e-02  1.303e-01 -5.063e-03  2.223e-01 -5.088e+02  3.275e+02 -3.107e+03 -3.058e+02  7.553e+02 -5.760e+02  8.035e+03
 1.818e-05 -7.546e-05 -7.696e-06  1.682e-05 -3.590e-05  4.671e-02  5.908e-03  1.612e-02 -2.876e-04 -4.886e-02 -7.249e-03
 1.818e-05 -7.546e-05 -7.696e-06  1.682e-05 -3.590e-05  4.671e-02  5.908e-03  1.612e-02 -2.876e-04 -4.886e-02 -7.249e-03 -1.267e-02  2.959e-05
```

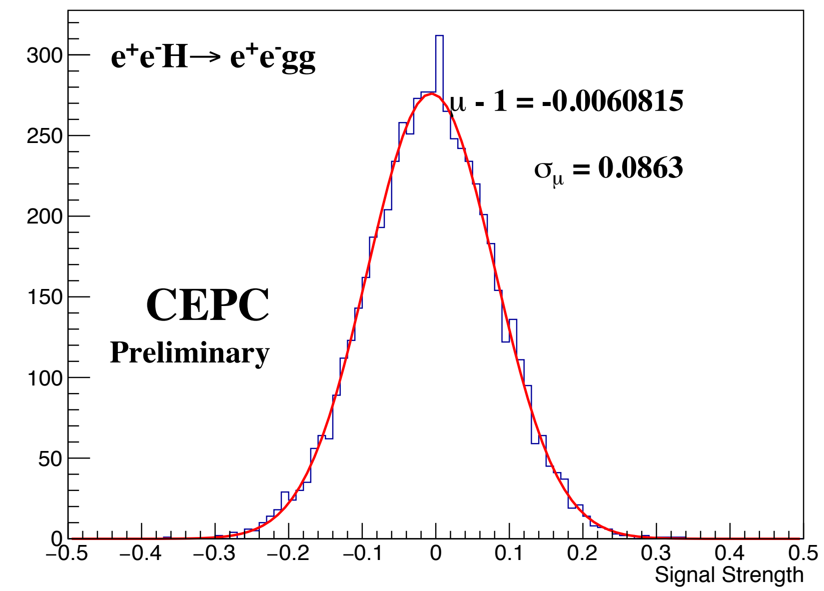
3D Fit Result-eeH

```
FCN=-143088 FROM HESSE     STATUS=OK            115 CALLS          709 TOTAL
                     EDM=0.000239403    STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                                  INTERNAL      INTERNAL
 NO.   NAME      VALUE            ERROR       STEP SIZE       VALUE
  1  a           -2.77463e-01   2.94531e-02   9.34960e-06  -2.77499e-02
  2  a1          -5.43966e-01   2.49963e-02   2.76246e-04  -5.75156e-01
  3  mean         1.25306e+02   2.32734e-02   1.37162e-04   2.30756e-01
  4  n            1.39175e+00   8.87463e-02   3.01172e-04   6.36640e-01
  5  nHbb         6.84780e+03   1.15456e+02   9.91464e-05  -3.20758e-01
  6  nHcc         2.91319e+02   4.55160e+01   2.44970e-04  -1.22833e+00
  7  nHgg         8.31608e+02   7.99969e+01   3.61313e-04  -7.31102e-01
  8  nsig         7.13414e-01   1.13396e-02   7.11706e-04  -6.91426e-01
  9  nzzsl_mu_bb   1.23165e+03   8.78586e+01   2.41494e-04  -7.05396e-01
 10  nzzsl_mu_cc   1.17387e+03   5.35245e+01   1.87426e-04  -7.27288e-01
 11  nzzsl_mu_uds   4.04222e+03   8.80655e+01   9.52480e-05  -6.38334e-01
 12  sigma        2.70013e-01   1.88887e-02   2.57037e-04  -6.12564e-01
                               ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR= 12    ERR DEF=0.5
ELEMENTS ABOVE DIAGONAL ARE NOT PRINTED.
 8.675e-04
 9.084e-05  6.250e-04
-5.093e-04 -2.606e-05  5.418e-04
 1.938e-03  6.561e-04 -6.578e-04  7.890e-03
-1.192e-01 -3.896e-01 -2.148e-02 -2.252e+00  1.333e+04
-1.262e-02 -6.956e-02 -3.097e-03 -2.534e-01 -4.325e+00  2.072e+03
-3.586e-02 -3.652e-01  1.300e-02 -9.345e-01  6.238e+02 -5.922e+02  6.403e+03
-1.245e-05 -2.664e-05  2.097e-05  1.791e-04 -4.520e-01 -3.895e-02 -1.017e-01  1.289e-04
 1.207e-01  4.173e-01  2.215e-02  2.325e+00 -6.245e+03 -2.251e+02 -9.681e+02  4.630e-01  7.722e+03
 1.234e-02  9.547e-02  1.955e-03  3.079e-01 -2.262e+02 -1.098e+03 -5.345e+01  4.636e-02  1.487e+02  2.865e+03
 2.445e-02  3.172e-01 -7.329e-03  7.814e-01 -6.112e+02  2.899e+02 -3.796e+03  8.485e-02  8.132e+02 -4.116e+02  7.756e+03
-3.705e-04 -1.474e-05  3.619e-04 -2.220e-04 -1.456e-01 -9.821e-03 -1.095e-02  5.349e-05  1.487e-01  9.952e-03  1.010e-02
-3.705e-04 -1.474e-05  3.619e-04 -2.220e-04 -1.456e-01 -9.821e-03 -1.095e-02  5.349e-05  1.487e-01  9.952e-03  1.010e-02  3.572e-04
```

Results with exponential background shape in eeH channel







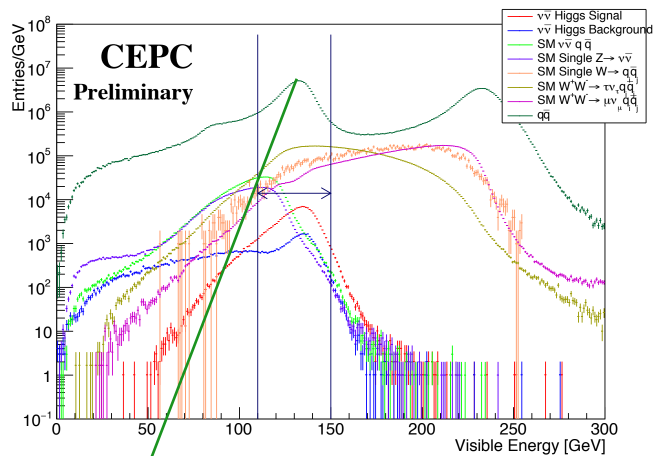
- much smaller deviation in H->cc/gg channel, but much larger deviation in H->bb channel

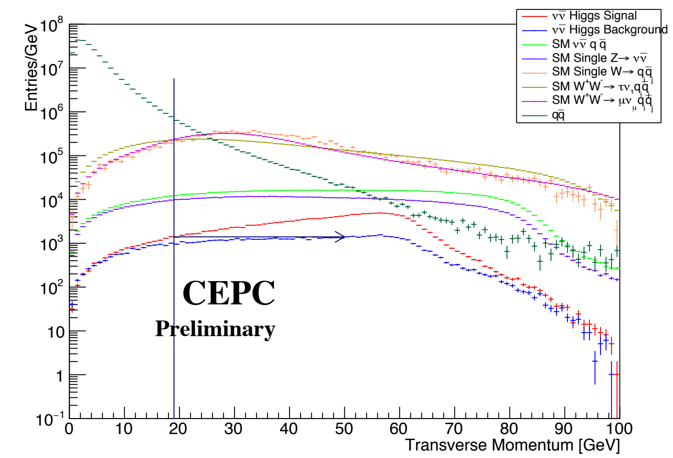
Event Selection of nnh

Event Pre-selection:

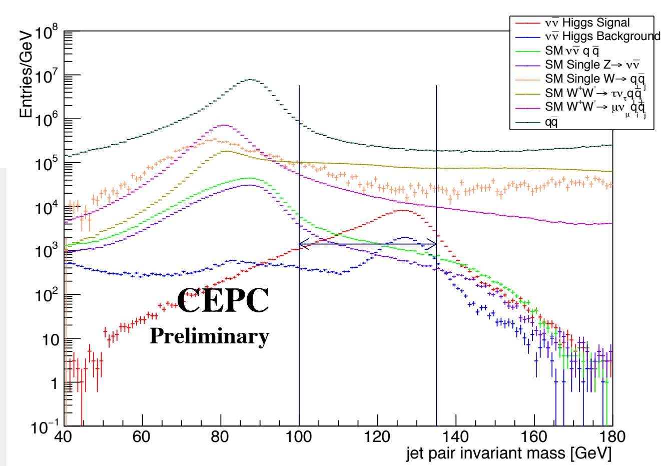
- Number of particles(PFO) >=20
- Visible Energy between 110 and 150 Ge
- Isolated electron and isolated muon veto
- y_{12} between 0.15 and 1.0, y_{23}<0.06, y_{34}<0.008
- cos θ between -0.98 and -0.4,
- BDT Cut(redone at 0.07)

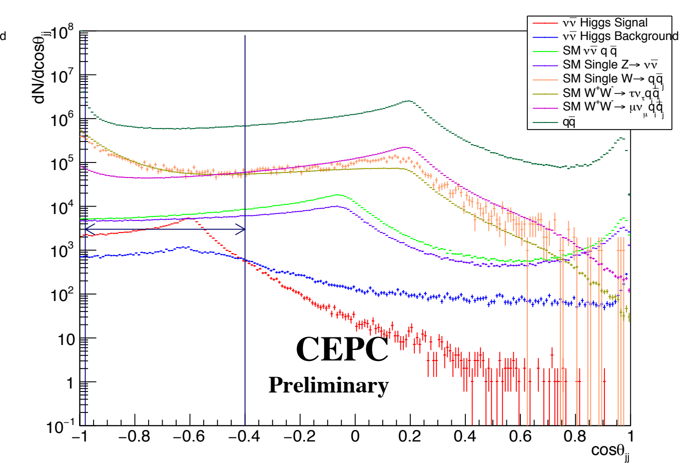
signal	Signal eff	higgs bkg	non-higgs bkg
85.8k	49.2%	1.96k	22.88k

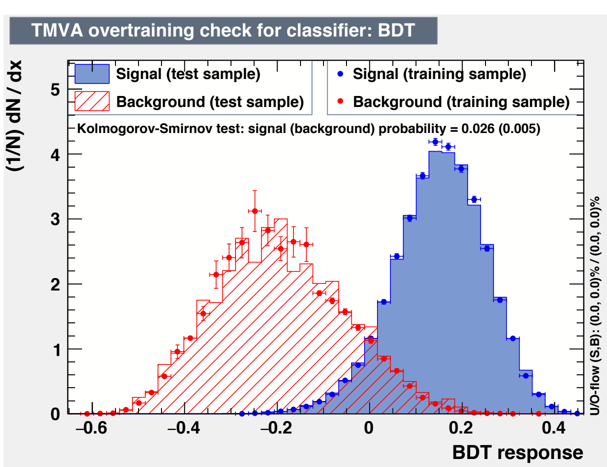


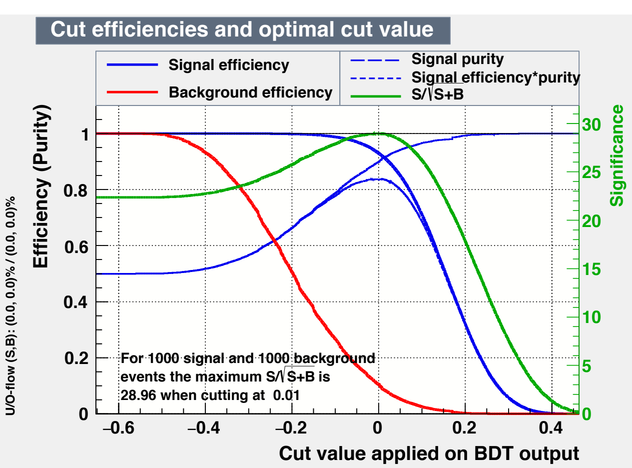


qq events peak in signal region due to radiation return



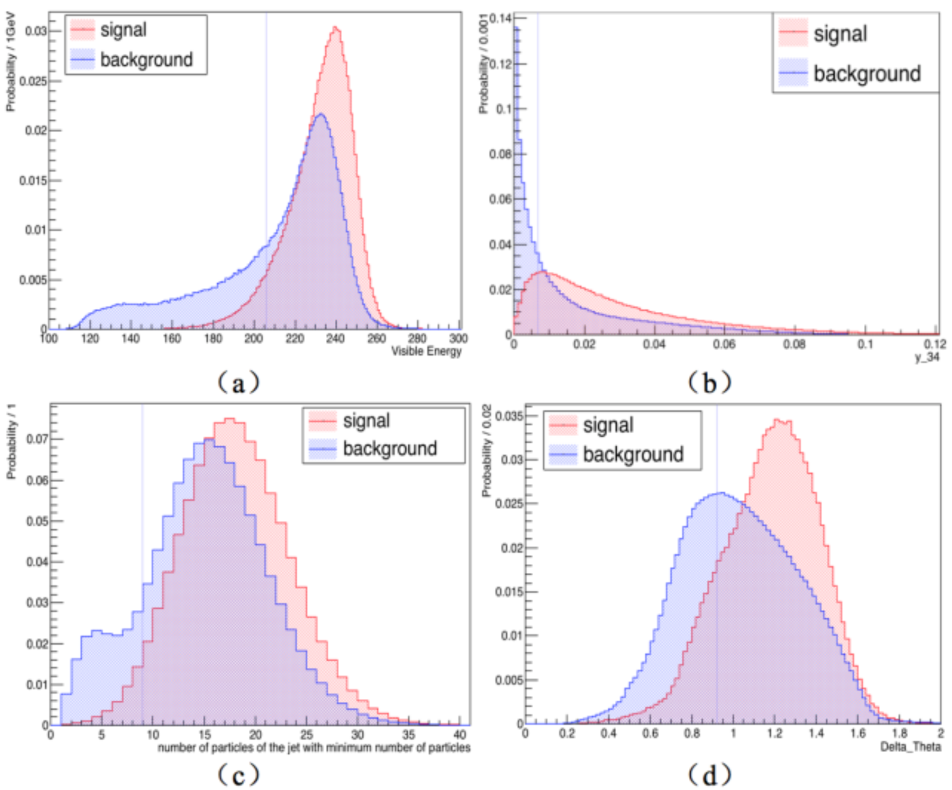






Event Selection of qqh

Basic Selection

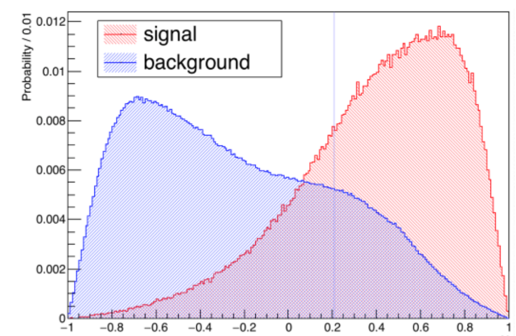


Jet Paring

$$\chi^2 = \min_{i\neq j\neq a\neq b} \frac{\left(M_{ij}-M_H\right)^2}{\sigma_H^2} + \frac{(M_{ab}-M_Z)^2}{\sigma_Z^2}$$

$$X = \frac{\chi_W^2-\chi_{ZH}^2}{\chi_W^2+\chi_{ZH}^2}$$

χ_W defined similar to χ_{ZH}



After X cut, BDT Analysis with input variables like Hmass, Zmass KMasses(masses of jet pairs after kinematic fit), y_{45}, H/Z $\cos\theta$, sphericity was applied.

Classified according to gluon and gamma radiation, increase statistics

signal	Signal eff	higgs bkg	4f-hadronic	qq	4f-semileptonic
211.2k	42.8%	32.6k	1.08M	405.6k	0.58k