MVA method in $ZH \to qq\gamma\gamma$ analysis in CEPC

Fangyi Guo, Yaquan Fang, Xinchou Lou

Institute of High Energy Physics, CAS

CEPC Physics Workshop

3rd July, 2019, Peking University

Content

Review of CEPC CDR

CEPC CDR release: Nov 2018, IHEP

Higgs$\rightarrow \gamma\gamma$ physics analysis:

- Design point at CEPC_v4, $\sqrt{s} = 240GeV$, $\mathcal{L} = 5.6ab^{-1}$
- Whizard 1.95 + MoccaC generator, dedicated fast simulation based on parametrized detector response.
- Considered $H \rightarrow \gamma\gamma$ signal and 2 fermion dominant background
- Result: $\delta(Br(H \rightarrow \gamma\gamma) \times \sigma(ZH))$=6.84% in 3 combined channel, and 9.84% in $q\bar{q}\gamma\gamma$ channel.
- Could be improved by applying MVA method

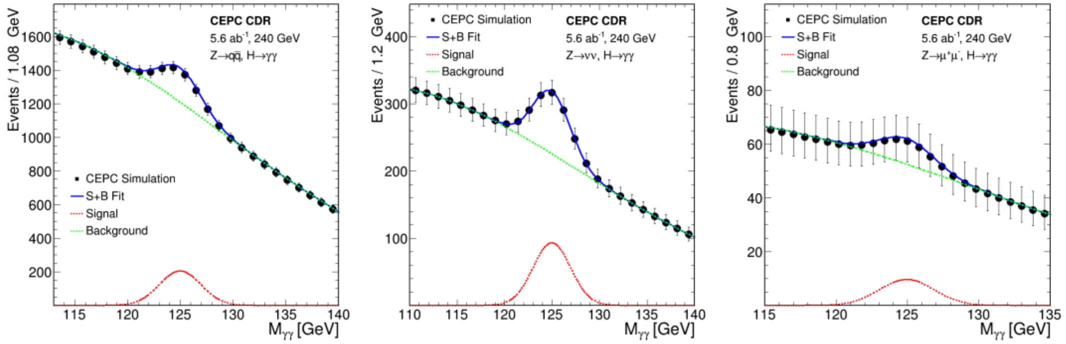


Figure1. di-photon invariant mass distributions in 3 sub-channel

Results

The combination of three sub-channel provides a final result of $\sigma(ZH) \times BR(H \rightarrow \gamma\gamma)$ measurement precision

Sub-channel	$q\bar{q}\gamma\gamma$	$l\bar{l}\gamma\gamma$	$\nu\bar{\nu}\gamma\gamma$	combined
precision	9.84%	23.7%	10.5%	6.84%

Arxiv: 1810.09037
Report in 2018 CEPC WS

MC sample and simulation

In order to keep consistent with CDR:

MC sample:

- Signal: $ee \to ZH \to q\bar{q}\gamma\gamma,$ 100k events.
- Background: $ee \to qq$+radiation photons, 20M events.

Simulation and event reconstruction

- Fast simulation: smear the objects with the resolution and efficiency with parametrized detector response, $\frac{\Delta E}{E} \sim \frac{16\%}{\sqrt{E}} \oplus 1\%$
- Event reconstruction: FSClasser FastJetClustering processer.

  reconstructed 2 on-shell photon with $m_{\gamma\gamma}$~125GeV and 2 jets

  define: γ_1/j_1 as photon/jet with lower energy, and γ_2/j_2 as higher energy one.

Event selection adjustment

Release the event selection criteria for the further MVA method

Selection in CDR:

$E_{\gamma 1} > 35GeV$

$35GeV < E_{\gamma 2} < 96GeV$

$cos\theta_{\gamma\gamma} >$ -0.95, $cos\theta_{jj}$>-0.95

$pT_{\gamma 1} > 20GeV, pT_{\gamma 1} > 30GeV$

$110GeV < m_{\gamma\gamma} < 140GeV$

$125GeV < E_{\gamma\gamma} < 145GeV$

$min|cos\theta_{\gamma j}|$<0.9

Present:

$E_{\gamma 1} > 25GeV$

$35GeV < E_{\gamma 2} < 96GeV$

$cos\theta_{\gamma\gamma} >$ -0.95, $cos\theta_{jj}$>-0.95

$pT_{\gamma 1} > 20GeV, pT_{\gamma 1} > 30GeV$

$110GeV < m_{\gamma\gamma} < 140GeV$

$E_{\gamma\gamma} > 120GeV$

$\min|cos\theta_{\gamma j}|$<0.9

Old	signal	background
Total efficiency	53.08%	0.010%
Scaled to 5.6 ab^{-1}	923.13	29875.6

New	signal	background
Total efficiency	53.34%	0.010%
Scaled to 5.6 ab^{-1}	927.65	31587.6

MVA method

Considered variables:

- P, E, pT, $cos\theta$ of two photon and 2 jets
- P, E, pT, $cos\theta$, recoil mass, pTt, Pt* of di-photon system
- P, E, mass, recoil mass, $cos\theta$ of jj system
- ΔP, ΔE, $\Delta\phi$ between two photon, $\gamma\gamma$-qq
- Cosine angle between 2 photon, 2 jets, 1 photon and 1 jet, $\gamma\gamma$ and jj system.
- Minimum ΔR between any photon and jet

Totally 42 variables

Pt*: Di-photon P projected perpendicular to the di-photon thrust axis.(similar as pTt but replace pT with P)

pTt = $|(\overrightarrow{P_1} + \overrightarrow{P_2}) \times \frac{\overrightarrow{P_1}-\overrightarrow{P_2}}{|\overrightarrow{P_1}-\overrightarrow{P_2}|}|$

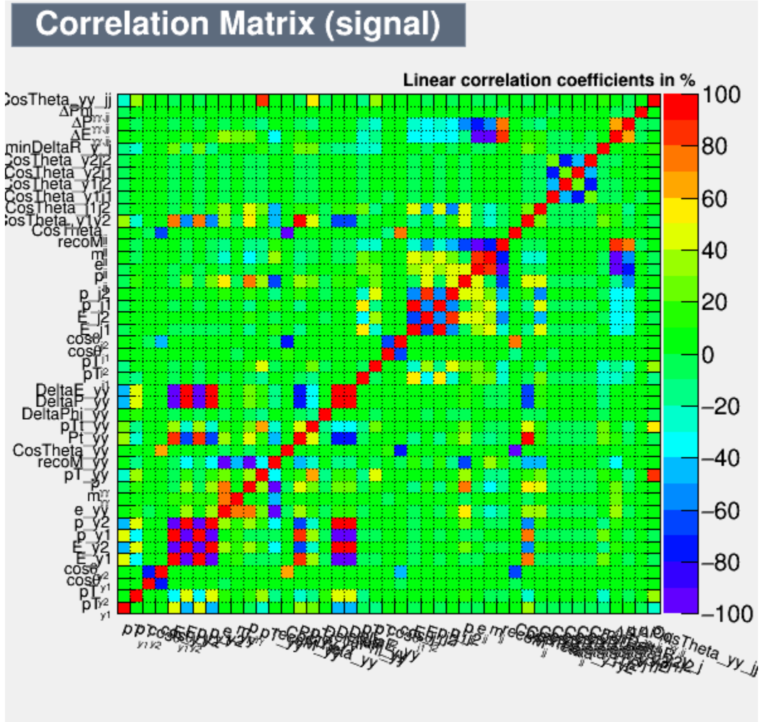
Separation power:

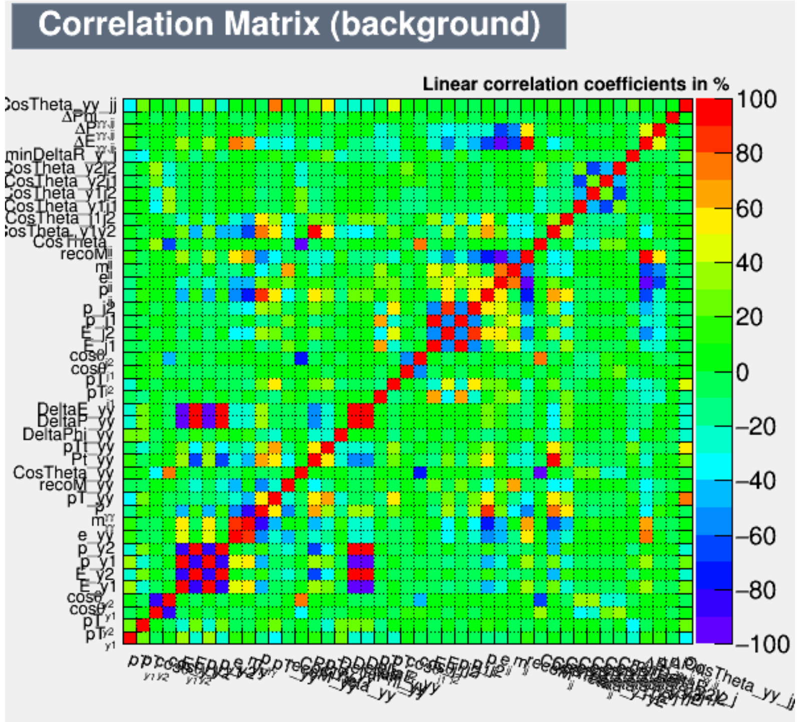
$$< S^2 >= \frac{1}{2}\int \frac{(\hat{y}_s(y) - \hat{y}_b(y))^2}{\hat{y}_s(y) + \hat{y}_b(y)} dy.$$

y: discriminating variable
$\hat{y}_s(y)$and$\hat{y}_B(y)$: the distributions of the variable for signal and background samples

MVA method

Variable correlation matrix





First step: remove high $m_{\gamma\gamma}$-related variable $|Corr_{v-m_{\gamma\gamma}}|$<30%

Second step: remove high co-related variables $|Corr_{v_1-v_2}|$<40%

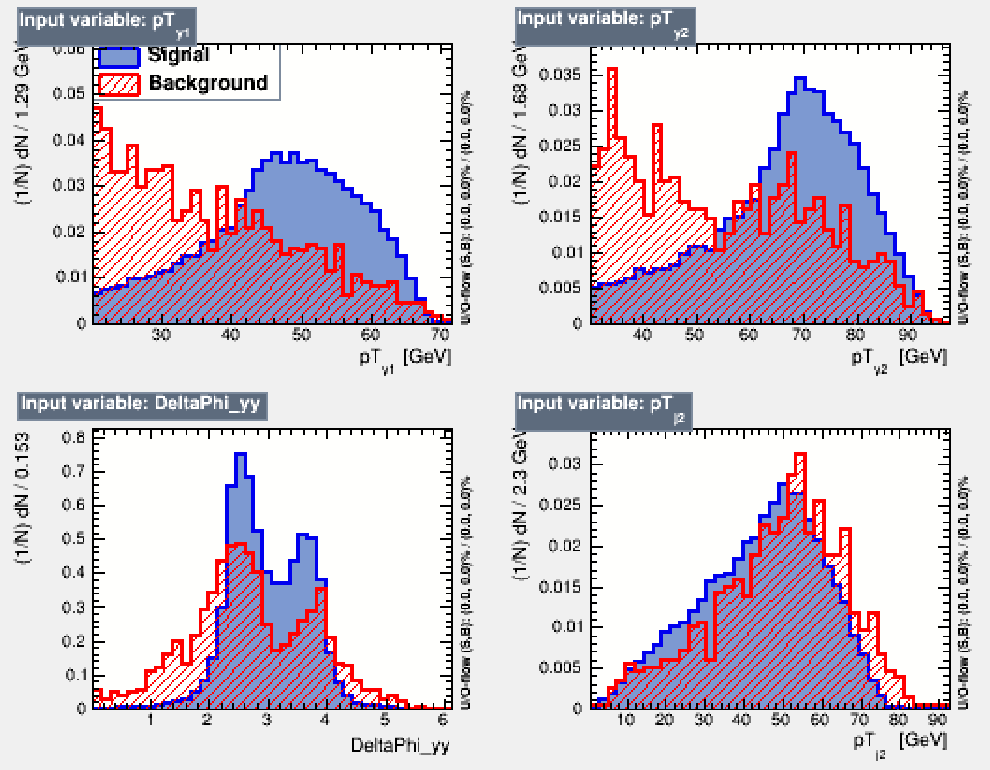
MVA method

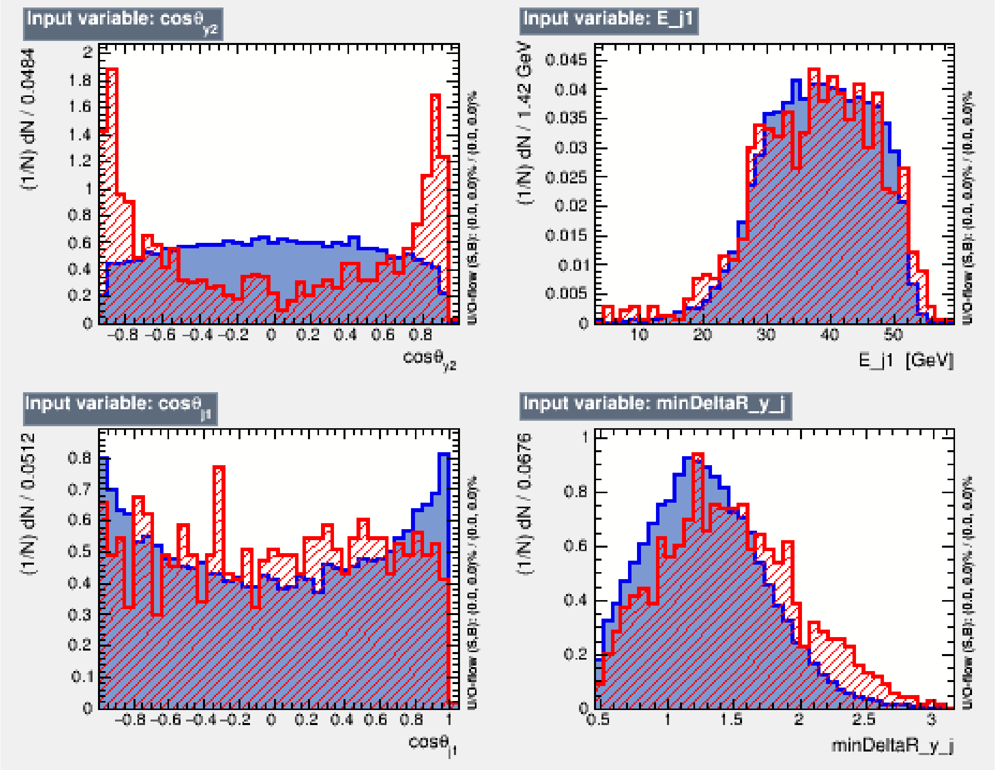
Remaining variables:

Variable	Definition	$\langle S^2 \rangle$
$pT_{\gamma 1}$ / $pT_{\gamma 2}$	pT of $\gamma 1$ or $\gamma 2$	0.39 / 0.30
$cos\theta_{\gamma 2}$	Cosine polar angle of $\gamma 2$	0.39
$\lvert\Delta\Phi_{\gamma\gamma}\rvert$	$\lvert\Delta\Phi\rvert$ between 2 photon	0.30
$min\Delta R_{\gamma,j}$	Minimum ΔR between photon and jet	0.09
$cos\theta_{j1}$	Cosine polar angle of $j1$	0.08
pT_{j2}	pT of $j2$	0.08
E_{j1}	Energy of $j1$	0.03
$cos\theta_{\gamma 2-j2}$	Cosine value of the angle between $\gamma 2$ and $j2$	0.03
$cos\theta_{\gamma 1-j1}$	Cosine value of the angle between $\gamma 1$ and $j1$	0.02
$\lvert\Delta\Phi_{\gamma\gamma-jj}\rvert$	$\lvert\Delta\Phi\rvert$ between $\gamma\gamma$ and jj system	0.01
$cos\theta_{\gamma\gamma-jj}$	Cosine value of the angle between $\gamma\gamma$ and jj system	0.01

BDT training parameter:
“BDTG”,“NTrees=900:nEventsMin=50:BoostType=Grad:Shrinkage=0.06:UseBaggedGrad:GradBaggingFraction=0.6:nCuts=20:MaxDepth=3”

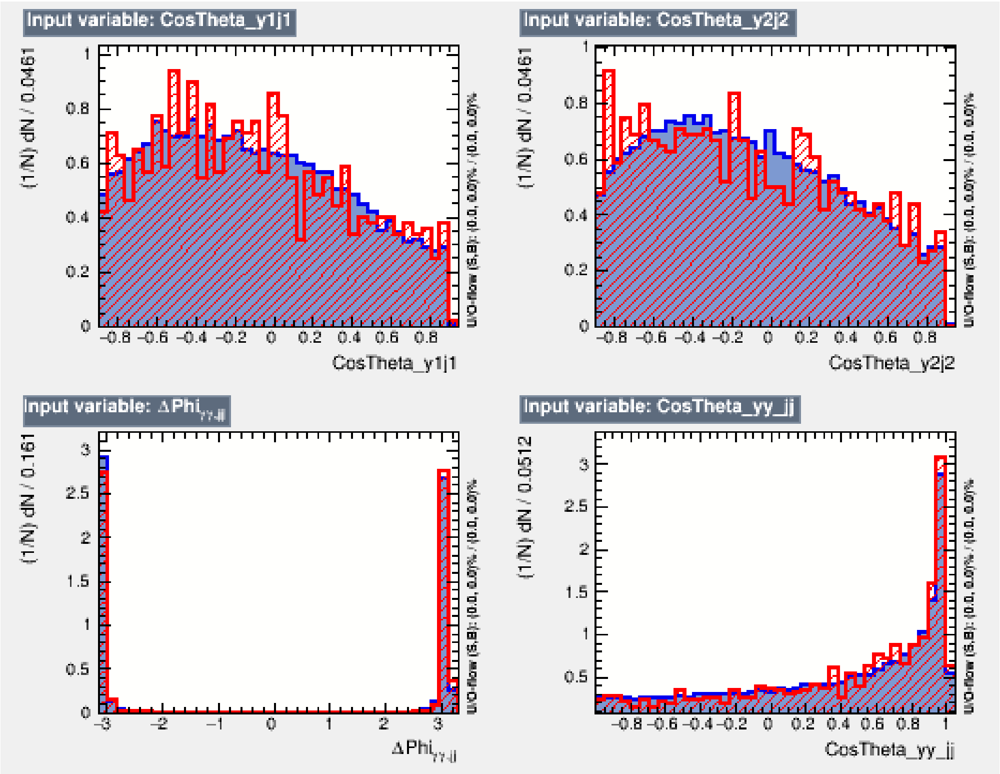
MVA method

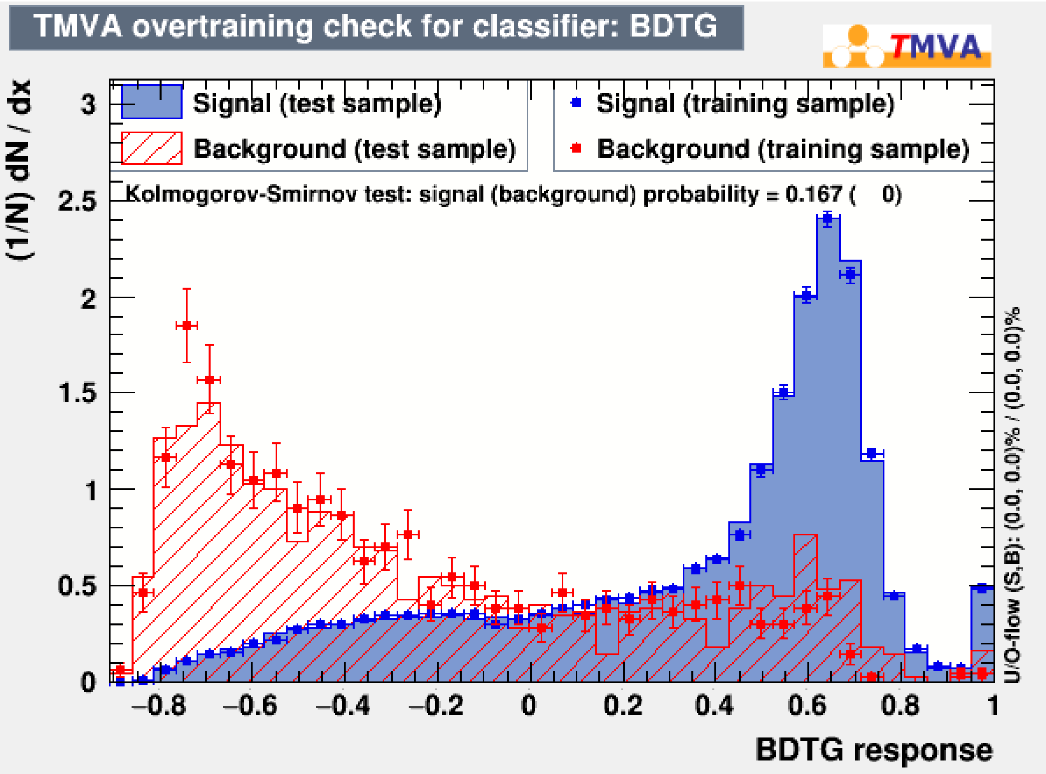




MVA variables distribution for signal(blue) and background(red)

MVA method





MVA variables distribution(left) and BDT response(right) for signal(blue) and background(red)

$|Corr_{m_{\gamma\gamma}-BDTout}|$ is 5.4% in signal and 17% in background

2-D signal and background fit

Fit the sample in 2 dimension

- Di-photon invariant mass $m_{\gamma\gamma}$

  Signal: Gaussian PDF

  Background: 2nd polynomial exponential PDF

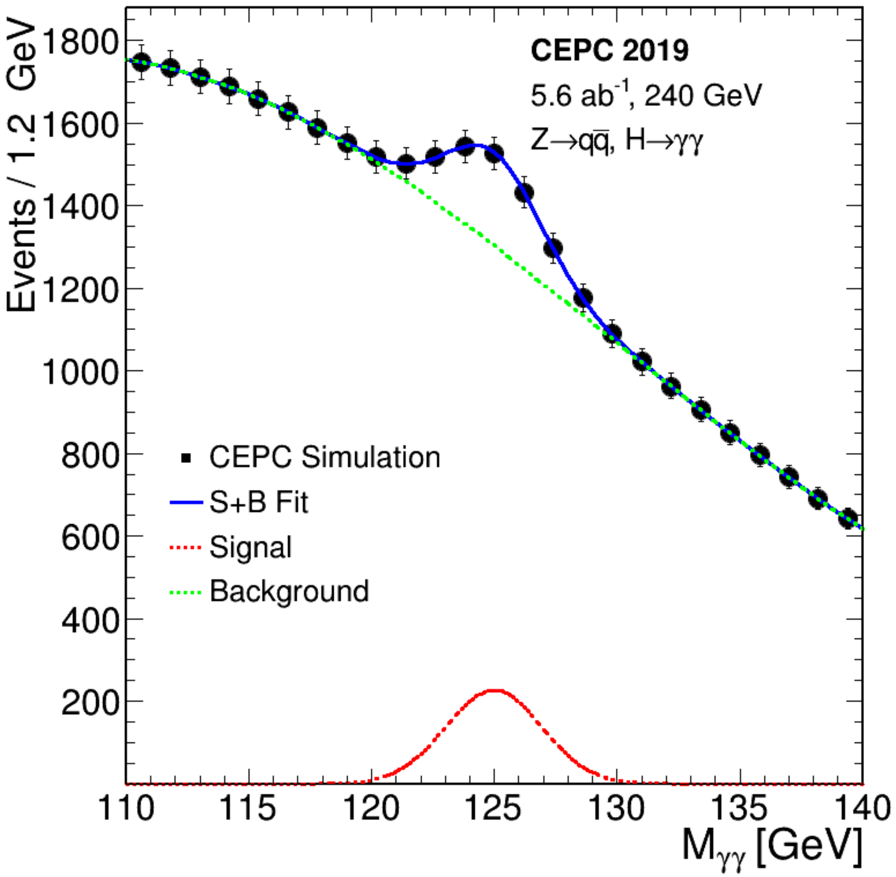
  $$PDF_{m_{\gamma\gamma}} = \mu \times N_{sig}^{SM} \times PDF_{sig} + N_{bkg} \times PDF_{bkg}$$

- BDT response ($\left|Corr_{m_{\gamma\gamma}-BDT}\right|$ <20%)

  Binned PDF for signal and background

  $$PDF_{BDT} = \mu \times N_{sig}^{SM} \times PDF_{sig} + N_{bkg} \times PDF_{bkg}$$

- $PDF_{2D} = PDF_{m_{\gamma\gamma}} \times PDF_{BDT}$



Fangyi Guo, guofangyi18@mails.ucas.ac.cn

Results

Fit result in $ee \to ZH \to qq\gamma\gamma$ channel

$$\mu = 1.000 \pm 0.066$$

	$qq\gamma\gamma$ channel	3 Combined channel
Pre MVA (CDR result)	9.84%	6.80%
After MVA	6.56%	5.39%
Improvement	33%	21%

*combined results are based on MVA $qq\gamma\gamma$ + no MVA $(ll\gamma\gamma + \nu\nu\gamma\gamma)$

FCC-ee case:

- 3% @240GeV, 10ab^{-1}, based on CMS ECal resolution, TLEP physics, 2013
- 9% @240GeV, 5ab^{-1}, FCC-ee CDR, 2018

Future work

MVA method in $ZH \to ll\gamma\gamma, ZH \to \nu\nu\gamma\gamma$ channel

- Estimation: if the same improvement(30%) could be reached in these two channel

  $ll\gamma\gamma$ channel: 23.7% ➡ 16.6%

  $\nu\nu\gamma\gamma$ channel: 10.5% ➡ 7.4%

  combined results after MVA: ~4.7%

Simulation sample

- Update the fast simulation sample to full simulation

  Signal: broaden the $m_{\gamma\gamma}$ distribution, lose ~7.5% events

  Background: include more high energy photon, might be excluded by reconstruction algorithm.

Conclusion

MVA method improvement in CEPC $H \to \gamma\gamma$ analysis

- $qq\gamma\gamma$ channel has been tested: 33% improvement in one sub-channel, reaching the result $\delta(Br \times \sigma) = 6.56\%$
- Prospect: 4.7% combined precision could be reached after MVA in 3 channels

Next step towards TDR

- Full simulation samples are ready, but a accurate reconstruction algorithm is necessary
- Based on previous study, a ~20% decrease would appear after converting to full simulation

Back up

Correlation matrix (after removing $m_{\gamma\gamma}$-related variables)