MVA method in $ZH \rightarrow qq\gamma\gamma$ channel @CEPC

FANGYI GUO

MC samples and event reconstruction

MC samples:

- Signal: 240GeV CEPC_v4 full simulation $ee \rightarrow ZH \rightarrow qq\gamma\gamma$ samples
- Background: 240GeV CEPC_v4 full simulation, all kind of background processes, including 2 fermions, 4 fermions and ZH processes.

Event reconstruction: Self-write FSClasser processer (refer to Xuewei and Kunlin qqmumu code)

- 2 photon with the largest energy
- Force other visible parts as 2 jets (I'd rather call that "the rest").
- Missing system

∴ available variables: 4-vector of 2 photons, missing system and qq system("the rest system").

Event selection

Definition: γ_1/ γ_2 : photon with lower/ higher energy

$E_{\gamma 1} > 20GeV$

$30GeV < E_{\gamma 2} < 100GeV$

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

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

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

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

$|m_{miss}| < 50GeV$

$60GeV < m_{qq} < 120GeV$

Event 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

Remove the jet angle-relative variables, expand photon energy range, add $|m_{miss}|$ and m_{qq}

Event selection

Final efficiency and scaled event number for each process:

Process	sig	2f bhabha	2f mumu	2f tautau	2f nunu	2f qq	4f sw_l	4f sw_sl	4f sze_l	4f szeorsw_l	4f sze_sl	4f sznu_l
Eff	67.74%	0.21%	0.60%	0.02%	0.00%	0.36%	0.01%	0.09%	0.02%	0.01%	0.05%	0.00%
scaled	1178.104	292831.6	179806.3	6547.585	0	1086006	370.5766	12933.54	1259.33	171.8421	867.2706	0

Process	4f sznu_sl	4f ww_h	4f ww_l	4f ww_sl	4f zz_h	4f zz_l	4f zzorww_h	4f zzorww_l	4f zz_sl	eeh_X	tautauh_X	mumuh_X	nnh_X
Eff	0.00%	0.02%	0.00%	0.02%	0.01%	0.10%	0.02%	0.00%	0.06%	1.00%	0.15%	2.23%	0.00%
scaled	0	3770.378	0	5111.962	364.5684	564.7284	4467.587	0	1882.988	396.0027	55.22114	847.1086	0

Results in CDR (qqyy signal + 2f qq background, fast simulation):

- Sig: 53.09%, 824.38 events
- qq background: 0.01%, 26674.7 events

Main background:

- 2f bhabha 18.6%
- 2f mumu 11.4%
- 2f qq 69.1%
- 4f sw_sl 0.9%

MVA categorization

Considered variables:

- P, E, pT, $cos\theta$ of two photon
- P, E, pT, $cos\theta$, recoil mass, pTt, Pt* of di-photon system
- P, E, $cos\theta$ of missing system
- P, E, mass, recoil mass of qq system
- ΔP, ΔE, $\Delta\theta$, $\Delta\phi$ between two photon, $\gamma\gamma$-qq, $\gamma\gamma$-miss, qq-miss

38 variables totally

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

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

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

For different background processes, calculate weighted average of them

MVA categorization

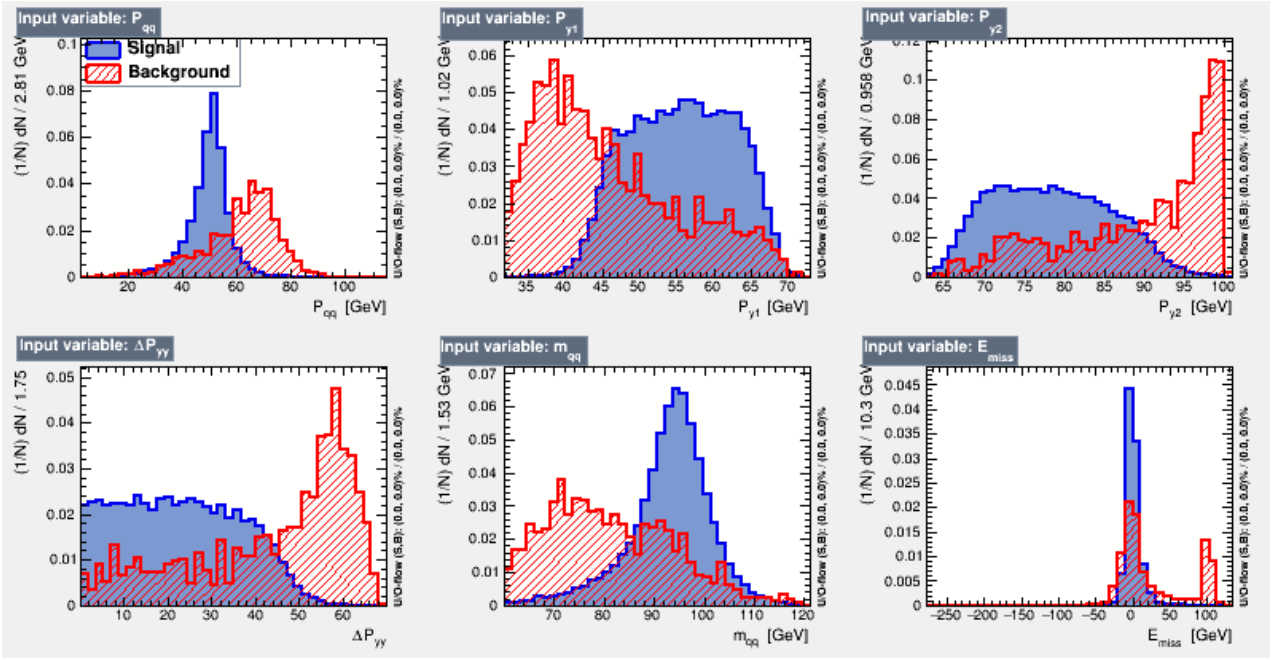
MVA variable: exclude high-relative variables and $m_{\gamma\gamma}$-relative variable:

Variable	Difination	Weighted average $\langle S^2 \rangle$
p_{qq}	Momentum of qq system	0.970
$\Delta P_{\gamma\gamma}$	ΔP of two photon	0.918
$P_{\gamma 1}, P_{\gamma 2}$	Momentum of two photon	0.864, 0.795
M_{qq}	Invariant mass of qq system	0.699
E_{miss}	Energy of missing system	0.675

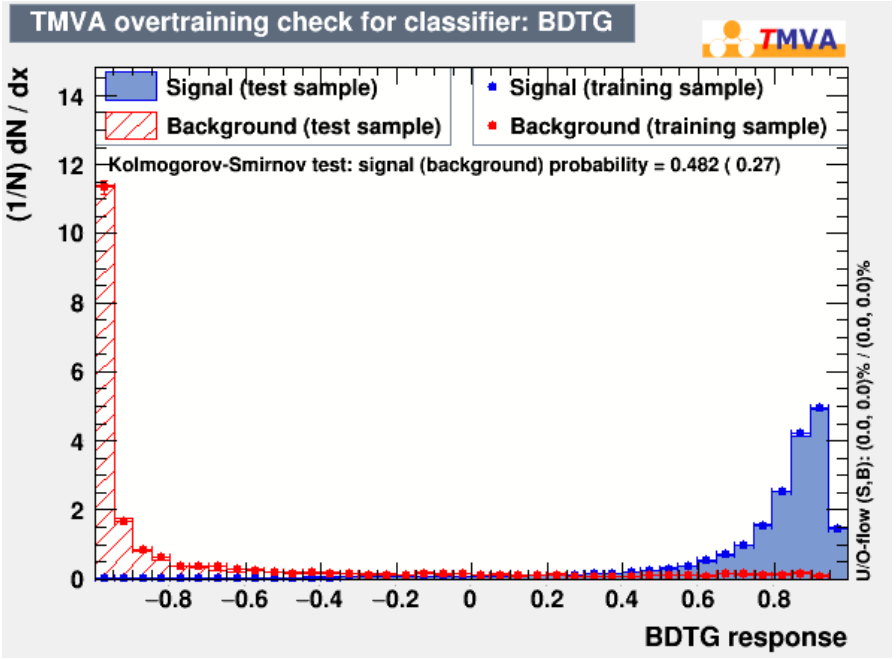
BDT training:

- Signal: ZH->qqyy
- Background: 2f bhabha, 2f mumu, 2f qq, 4f sw_sl
- Parameter: “BDTG”, “NTrees=900:nEventsMin=50:BoostType=Grad:Shrinkage=0.06:UseBaggedGrad:GradBaggingFraction=0.6:nCuts=20:MaxDepth=3”

MVA categorization



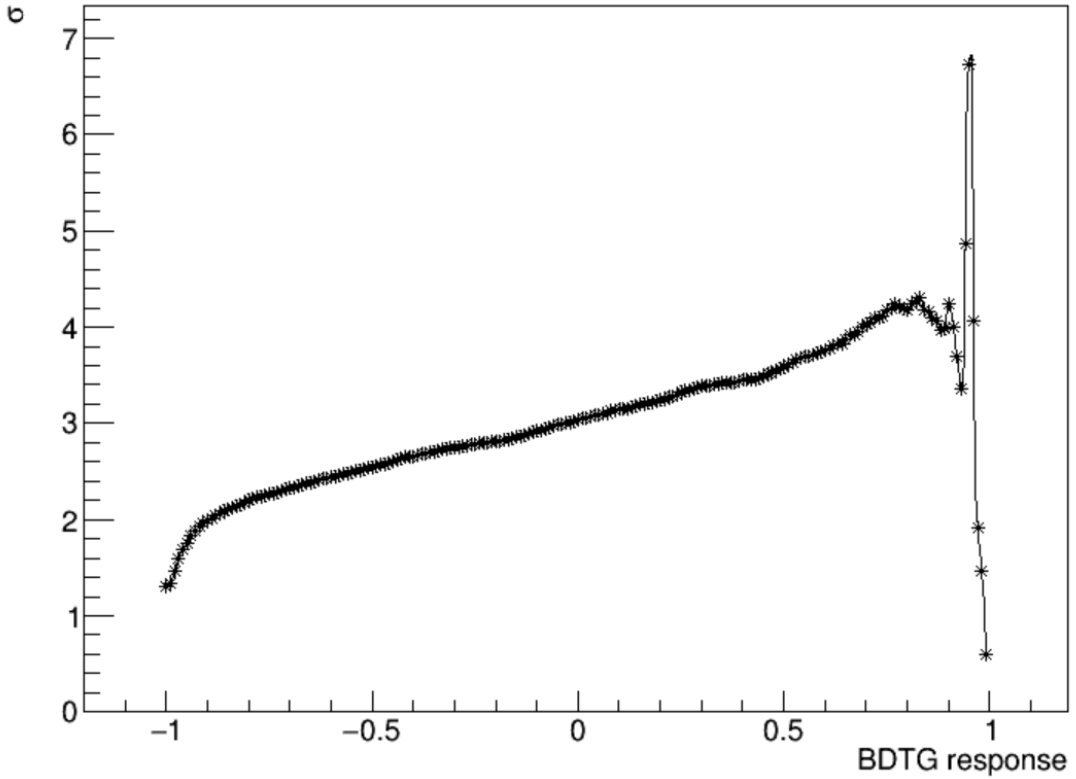
Input variable distribution



BDTG response

MVA categorization

Categorization: maximum $\sigma = N_{sig}/\sqrt{N_{sig} + N_{bkg}}$



Kcut: BDTG=0.83
Tight category: BDTG>0.83
  Nsig: 608
  Nbkg: 19350
  significance: 4.31
Loose category: BDTG<0.83
  Nsig: 414
  Nbkg: 593225
  significance: 0.54
Combined significance: 4.34

Conclusion

Present results is worse than CDR due to:

- Update the MC to full simulation sample, so photon energy in background process increased largely.
- Changed the FSClasser processer, the reconstruction progress is different, and jet 4-vector is not available.

MVA method:

- Increase signal significance from <2 to 4.34

Next step:

- Use more MC statistics and fit $m_{\gamma\gamma}$ distribution to get $\sigma \times Br$ precision. (may be larger than 10%).

Back up

Cutflow of full and fast simulation

	Full sim qq bkg		Fast qq background		Fast qq background in CDR	
generated	1999052		19999930		20000000	
$qq\gamma\gamma$					13914611	69.573%
$E_{\gamma 1}$>35GeV	597378	29.883%	3668805	18.344%	120726	0.868%
35GeV<$E_{\gamma 2}$<96GeV	339636	56.854%	1748931	47.670%	55583	46.041%
$cos\theta_{jj}$>-0.95	-	-	-	-	44012	79.182%
$cos\theta_{\gamma\gamma}$>-0.95	202029	59.484%	1310484	74.931%	36794	83.600%
$pT_{\gamma 1}$>20GeV	96901	47.964%	504704	38.513%	22481	61.100%
$pT_{\gamma 2}$>30GeV	65109	67.191%	334196	66.216%	11733	52.191%
110GeV<$m_{\gamma\gamma}$<140GeV	13145	20.189%	32405	9.696%	4316	36.785%
125GeV<$E_{\gamma\gamma}$<145GeV	10808	82.221%	28263	87.218%	3912	90.639%
min$\lvert cos\theta_{\gamma j}\rvert$<0.9	-	-	-	-	1972	50.409%
$\lvert m_{miss}\rvert < 50GeV$	7799	72.160%	22758	80.522%	-	-
$60GeV < m_{qq} < 120GeV$	7165	91.871%	21259	93.413%	-	-
		0.358%		0.106%		0.010%
scaled to 5.6 ab-1	1086006		322073.9		30335	

Full separation power

	2f bhabha	2f mumu	2f qq	4f sw_sl	Weighted average
p_yy:	0.75385	1.06076	1.13479	0.859414	1.053073
p_qq:	0.765603	0.832624	1.04833	0.959733	0.970241
DeltaP_yy:	0.321369	0.856973	1.09323	0.506039	0.917546
DeltaE_yy:	0.321369	0.856973	1.09323	0.506039	0.917546
p_y2:	0.29227	0.65993	1.05895	0.310316	0.864281
e_y2:	0.20038	0.540294	1.05085	0.223973	0.827164
p_y1:	0.210274	0.657831	0.978941	0.458889	0.794697
e_y1:	0.164858	0.55058	0.967278	0.307266	0.764657
m_qq:	0.44561	0.517551	0.792798	1.06643	0.698867
e_miss:	0.350578	0.780536	0.739491	1.15457	0.675137
m_recoil_yy:	0.537937	0.531498	0.730609	0.50496	0.670071
DeltaP_qq_miss:	0.519106	0.52347	0.568037	0.997732	0.557357
p_miss:	0.216456	1.01997	0.501995	1.69641	0.517882
pT_y1	0.23197	0.727028	0.550151	0.550405	0.511103
DeltaP_yy_miss:	0.486795	0.581791	0.488341	1.1114	0.503872
Pt_yy:	0.428326	0.465372	0.508513	0.691513	0.490142
DeltaE_yy_miss:	0.481676	0.556644	0.465391	1.29921	0.485728
cosTheta_miss:	0.731802	0.511869	0.394722	0.094157	0.468459
DeltaE_qq_miss:	0.228706	0.602659	0.44238	1.56398	0.430134

	2f bhabha	2f mumu	2f qq	4f sw_sl	Weighted average
pT_yy:	0.438645	0.625374	0.367543	0.594398	0.412157
pT_y2	0.36605	0.507384	0.397118	0.659951	0.406108
DeltaTheta_yy:	0.45254	0.369316	0.360132	0.470588	0.37931
pTt_yy:	0.68706	0.422187	0.23929	0.491125	0.345721
cosTheta_y2:	0.423651	0.821288	0.159472	0.799547	0.289683
DeltaE_yy_qq:	0.376607	0.406401	0.204614	1.16529	0.267654
m_recoil_qq:	0.35757	0.313374	0.214431	1.38454	0.262052
DeltaPhi_yy:	0.706649	0.32096	0.122419	0.366745	0.256004
cosTheta_yy:	0.488468	0.67515	0.104619	0.587656	0.245391
cosTheta_y1:	0.287657	0.757538	0.060569	0.730569	0.188137
DeltaTheta_yy_qq:	0.12795	0.580083	0.12312	1.37322	0.186589
e_qq:	0.189498	0.216011	0.118639	1.32119	0.152879
e_yy:	0.266564	0.128205	0.122165	0.226928	0.150624
DeltaP_yy_qq:	0.155248	0.179825	0.075807	0.800358	0.108473
DeltaPhi_yy_qq:	0.113687	0.239043	0.042155	1.07514	0.086511
DeltaTheta_yy_miss:	0.055216	0.057437	0.017506	0.071642	0.029546
DeltaTheta_qq_miss:	0.060357	0.057784	0.0145	0.101661	0.028714
DeltaPhi_yy_miss:	0.027305	0.018937	0.012849	0.056217	0.016596
DeltaPhi_qq_miss:	0.02634	0.02218	0.009229	0.075023	0.01444