CEPC NOTE

CEPC_PHY_2021_002

January 29, 2021

The measurement of the $H \to bb$ signal strength at the CEPC

Zhu Yongfeng[a]

[a]*Institute of High Energy Physics*

Abstract

The main purpose of the Circular Electron Positron Collider (CEPC) is to study the properties of Higgs precisely. This note estimates the performance of the $e^+e^- \to ZH(Z \to q\bar{q}, H \to b\bar{b})$ benchmark measurement. With the full simulation analysis, the CEPC is expected to measure the signal strength to a relative accuracy of 0.37% with the signal efficiency of 36.73%. The left backgrounds mainly include the process $e^+e^- \to b\bar{b}$ and $e^+e^- \to ZZ \to 4\ quarks$.

E-mail address: zhuyf@ihep.ac.cn

1 Introduction

Since the Large Hadron Collider (LHC) had discovered the Higgs boson in 2012, the precise measurement of its properties becomes a common goal among the whole world. Several high energy physics experiments have been proposed, including CEPC, International Linear Collider (ILC) and Future Circular Collider e^+e^- (FCCee), and CLIC. The CEPC has a main ring circumference of 100 km and can be operated as a Z factory and a Higgs factory, it can also measure the mass and width of the W boson accurately. The CEPC is designed to operate at a center-of-mass energy of 240 GeV and accumulate $5.6ab^{-1}$ of integrated luminosity. With these setups, the CEPC is expected to produce one million Higgs bosons. The following table lists the branching ratio of the Higgs decay mode. The channel of $Higgs \to b\bar{b}$, which have the advantage of large branching ratio (more than 57%) and good b jet flavor tagging, becomes one of the most important channels for Higgs measurement. In this note, with the cut-flow method and official CEPC software and samples, the $Higgs \to b\bar{b}$ signal strength measurement is analyzed.

Decay mode	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to \tau^+\tau^-$	$H \to \mu^+\mu^-$	$H \to WW^*$	$H \to ZZ^*$	$H \to \gamma\gamma$	$H \to Z\gamma$	$H \to gg$
Branching ratio	57.7%	2.91%	6.32%	2.19×10^{-4}	21.5%	2.64%	2.28×10^{-3}	1.53×10^{-3}	8.57%

2 Samples and software chain

The SM Higgs bosons are mainly generated via the Higgsstrahlung and the vector boson fusion processes at the CEPC. In this analysis, the backgrounds are characterized according to the number of final state fermions at the Parton level. When CEPC operates at 240 GeV center-of-mass energy, the leading SM backgrounds are the 2-fermion and 4-fermion backgrounds. The 2-fermion backgrounds are the $q\bar{q}$, Bhabha, $\mu^+\mu^-$ and $\tau^+\tau^-$ processes; the 4-fermion backgrounds include the ZZ, the WW, the single W, the single Z, and the interfering processes. The latter is denoted as ZZorWW and ZorW process since the final state fermion combination allows multiple intermediate states and their interferences.

The CEPC baseline detector is a Particle Flow Oriented detector. It is composed of a low-material tracking system, a high granularity calorimeter system, and a 3-Tesla large radius solenoid that hosts both ECAL and HCAL inside. The software used to do simulation and reconstruction has been established. It uses the Whizard as the generator, the MokkaPlus for the full detector simulation, the Clupatra for tracking, and the Arbor for the PFA reconstruction. Based on the luminosity of the CEPC, an official MonteCarlo sample has been generated, simulated, and reconstructed.

3 Signal strength measurement at the CEPC

The samples at the CEPC could be classified into three classes, pure leptonic, semileptonic, and full hadronic according to the kinds of fermions in the Parton level. For the full hadronic, it can be further classified into 2-quark, 4-quark, and 6-quark. Since the signal, $e^+e^- \to ZH(Z \to q\bar{q}, H \to b\bar{b})$, only has four quarks in the Parton level, so it is straightforward and convenient to identify signal events in three steps.

- Firstly, separate the full hadronic from the pure(semi) leptonic samples.
- Secondly, separate 4-quark samples from that left from the first step.
- Thirdly, identify the signal from left backgrounds.

The following table lists the cut-flow and entires of samples left in each class, the number in the bracket shows the left ratio compared to the last cut. The following subsections will introduce how these cuts suppress the background.

	pure leptonic	semi leptonic	full hadronic	two quarks	four quarks	qqhbb	bkg
Total generated	5.16×10^8	4.80×10^7	3.46×10^8	3.03001e+08	4.28649e+07	443320	9.09×10^8
$multiplicity >= 27$	3.33×10^6(0.64%)	4.78×10^7(99.57%)	3.45×10^8(99.63%)	3.02×10^8(99.58%)	4.29×10^7(100.00%)	443,320(100%)	3.95×10^8(43.46%
$visEn >= 152$	1.05×10^6(31.55%)	3.47×10^7(72.57%)	1.98×10^8(57.37%)	1.55×10^8(51.31%)	4.28×10^7(99.93%)	443,229(99.98%)	2.33×10^8(58.94%
$leadLepEn <= 58$	710, 807(67.71%)	1.62×10^7(46.63%)	1.92×10^8(97.05%)	1.49×10^8(96.49%)	4.24×10^7(99.06%)	442,781(99.90%)	2.08×10^8(89.41%
$leadNeuEn <= 57$	285, 495(40.16%)	1.43×10^7(88.28%)	1.19×10^8(62.04%)	7.81×10^7(52.26%)	4.09×10^7(96.43%)	441,908(99.80%)	1.33×10^8(63.92%
$thrust <= 0.858125$	13, 305(4.66%)	1.11×10^7(77.82%)	5.51×10^7(46.28%)	1.75×10^7(22.43%)	3.75×10^7(91.68%)	387,704(87.73%)	6.58×10^7(49.43%
$sum2b >= 1.2585$	229(1.72%)	179, 188(1.61%)	3.88×10^6(7.04%)	2.39×10^6(13.66%)	1.47×10^6(3.93%)	331,784(85.58%)	3.73×10^6(5.66%)
$(MB1 - 126.5)^2 + (MB2 - 91.5)^2 <= 24^2$	3(1.39%)	12,083(6.74%).	354,943(9.14%)	87,107(3.64%)	262,209(17.81%)	162,832(49.08%)	204197(5.47%)

3.1 Finding full hadronic samples

- Since the Parton would hadronize and fragment into a larger number of final state particles compared to leptons. So the number of final state particles would be an effective criterion to eliminate the pure leptonic, just shown as the left plot of Fig. 4. After setting the cut of $multiplicity >= 27$, only 0.64% pure leptonic samples left, while more than 99% semileptonic and more than 99% full hadronic samples left. The left pure leptonic samples mainly include processes with τ produced.

- The processes with neutrinos produced could be suppressed with visible energy. So some processes in pure leptonic and semileptonic with high energy neutrinos produced could be efficiently eliminated. Some events of the process $e^+e^- \rightarrow q\bar{q}$ can eliminate high energy photons, when these photons fly along the beam axis, the visible energy would be small. So the visible energy also can suppress some backgrounds $e^+e^- \rightarrow qq$, shown as the middle plot of Fig. 4. With the cut $visible\ energy >= 152$, about 31.55% of pure leptonic, 72.57% of semileptonic, and 57.37% of the full hadronic left (about 48.69% of two quarks samples been eliminated), while 99.93% of four-quarks samples kept.

- The pure leptonic and semileptonic backgrounds have energetic final state leptons compared to full hadronic, these backgrounds could be efficiently vetoed reduced by an up limit on the leading lepton energy($<= 58GeV$), shown as the right plot of Fig. 4. About 53.37% semileptonic and 32.29% pure leptonic been suppressed and more than 99% four-quarks kept.

The above three cuts eliminate 77% backgrounds and kept more than 99% signal.

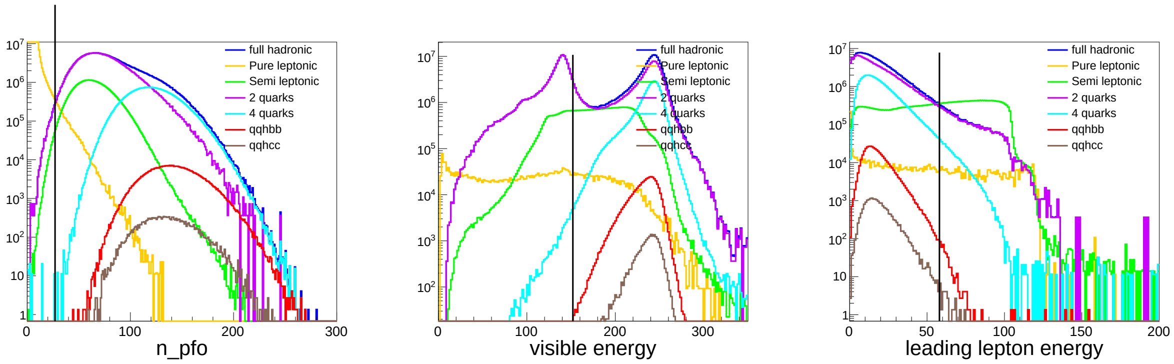

Figure 1: The distribution of multiplicity, visible energy and leading lepton energy on several classes. The left plot shows the distribution of multiplicity, the middle plot is visible energy and the right plot is leading lepton energy.

3.2 Finding 4-quarks samples

After suppressing the pure-leptonic and the semi-leptonic, the following cut-flow focus on selecting 4-quark samples.

- When the high energy photon, eliminated from the process $e^+e^- \to qq$, hit on the detector, the leading photon energy could be a good criterion to suppress the background $e^+e^- \to qq$. Since the detector's capability of reconstructing neutral particles is slightly worse, the high energy photon may be reconstructed as a neutral hadron. The energy of leading neutral particle, including photon and neutral hadron, could be used to suppress the background $e^+e^- \to qq$, shown as the left plot of Fig. 2. With the cut $leading\ neutral\ energy <= 57GeV$, about 47% 2-quark and 60% of pure leptonic been suppressed, while 96% of 4-quark kept.

- The next cut variable is thrust, which correlates with event shape. To evaluate the thrust of an event, one first determines the thrust axis n_T, which is the direction of maximum momentum flow. The thrust is then defined as the fraction of particle momentum flowing along the thrust axis, just as the following function, $T = max_{n_T}(\frac{1}{\sum_{j=1}^{N_{particles}} |P_j|} \sum_{i=1}^{N_{particles}} |P_i \cdot n_T|)$. For the process, $e^+e^- \to qq$, the two quarks would fly almost back to back(non-high energy initial state radiation) or along the same direction(the high energy initial state radiation fly along the opposite direction), so the value of thrust would close to one. For the process, $e^+e^- \to e^+e^-(\tau^+\tau^-)$, these two produced leptons would fly back to back, so the value of thrust would also close to one. While the samples with four quarks and six quarks tend to be fat, more like sphericity, so the value of thrust close to 0.5. After setting the cut $thrust < 0.858125$, only 22.43% of 2-quark and 4.66% of pure leptonic left, while around 91% 4-quark kept. The distribution of thrust for each class shown as the middle plot of Fig. 2.

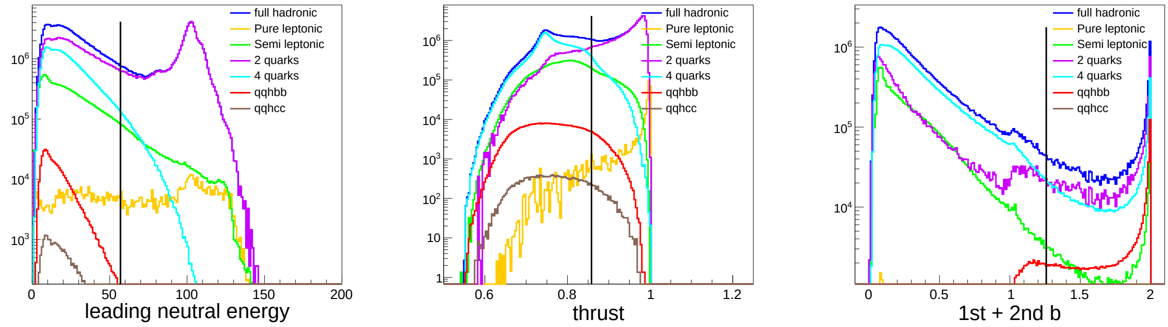

Figure 2: The distribution of leading neutral energy, thrust and $1st b + 2nd\ b$ on several classes.

3.3 Finding $e^+e^- \to ZH(Z \to qq, H \to bb)$

The above analysis has tried its best to keep signal and suppress backgrounds. But compared to the kept signal, the background is still sufficient, the magnitude of semileptonic and full hadronic background is two orders of the signal.

- Since the b jet flavor tagging performing well and the branching ratio of $Z \to b\bar{b}$ is small compared to $Higgs \to b\bar{b}$, the value of b likelihood could be a critical criterion to further suppress the background. Ask the sum of two leading b likelihood larger than 1.2585, only 1.61% of semileptonic and 13.66% of 2-quark samples left, while 85.58% of signal survive, just shown as the right plot of Fig. 2.

- For signal, the invariant mass of two leading b jet could close to 125 GeV($Higgs \to two\ leading\ b\ jets$) or 91 GeV($Z \to two\ leading\ b\ jets$). So the next cut would focus on the invariant mass of two leading b jets. Fig. 3 shows the invariant mass distribution of signal and background, the vertical axis is about the invariant mass of two leading b jets (named as MB1), while the horizontal axis is invariant mass of the other two jets(named as MB2). Based on the invariant mass distribution, the highlight point of signal and a radius circles signal as much as possible can be found. With this circle, the ultimate signal strength accuracy could reach 0.37%.

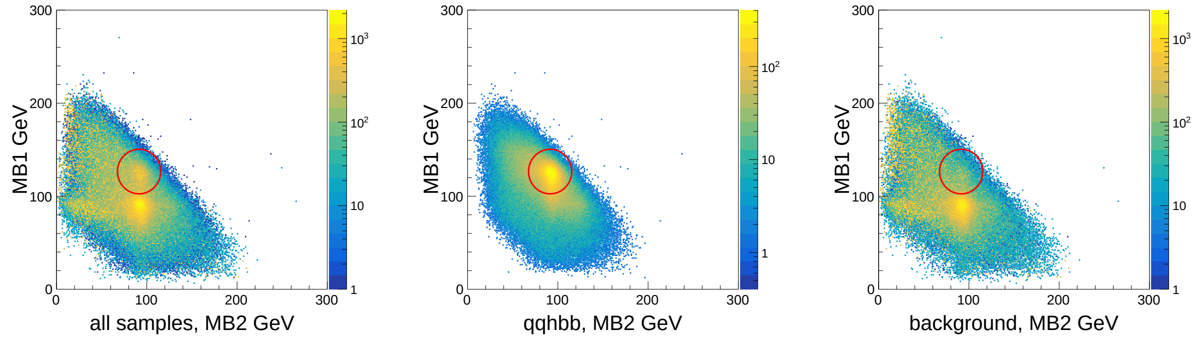

Figure 3: The distribution of invariant mass of two leading b jets and other two jets. The left plot containing signal and background, while the middle plot is about signal and right plot about background.

4 Conclusion

In this note, the expected signal strength accuracy of the channel $e^+e^- \rightarrow ZH(Z \rightarrow q\bar{q}, Higgs \rightarrow b\bar{b})$ has been studied. Based on the characteristics of final state particles of backgrounds, a cut chain has been established. With cut-flow, the signal strength accuracy could reach 0.37% with the signal efficiency of 36.73%. The following table shows the detail of the cut-flow. The cut $multiplicity >= 27$ could reduce almost all pure leptonic, $visible\ energy >= 152$ could reduce the events with high energy neutrinos produced as well as two quarks samples with high energy ISR flying along the beam axis. Since the single W and single Z would produce high-energy leptons, these two class samples could be vetoed efficiently with the up limit on leading lepton energy. The events with final state particles flying almost back to back would have thrust value close to one, while the thrust value of signal $e^+e^- \rightarrow ZH \rightarrow qqbb$ would close to 0.5, the cut $thrust <= 0.858125$ could suppress more than 50% backgrounds and keep 86% signal. The most crucial cut is the sum of two leading b jets likelihood since the signal aims at $Higgs \rightarrow b\bar{b}$ while the backgrounds have low branching ratio decay to $b\bar{b}$. This criterion can suppress 94.5% backgrounds and keep 88% signal.

	qqHbb	2f	SW	SZ	WW	ZZ	Mixed	ZH	Total Bkg	$\frac{\sqrt{S+B}}{S}(\%)$
Total generated	443,319	801,152,062	19,517,399	9,072,948	50,826,211	6,389,426	21,837,925	697,174	9.09493×10^8	6.80439
$N_{pfo} >= 27$	443,319	304,093,523	14,620,421	3,376,517	48,504,072	5,995,412	18,063,959	662,098	3.95316×10^8	4.48745
$visible\ energy >= 152$	443,229	155,699,349	12,954,417	1,665,294	40,026,041	4,248,857	18,013,808	405,125	2.33013×10^8	3.44726
$leading\ lepton\ energy <= 58$	442,781	149,987,734	5,009,976	789,199	30,713,952	3,650,556	17,840,111	356,253	2.08348×10^8	3.26337
$leading\ neutral\ energy <= 56$	441,908	78,343,205	4,137,216	237,258	29,438,527	3,495,739	17,179,048	351,112	1.33182×10^8	2.61584
$thrust <= 0.858125$	387,704	17,523,392	3,380,240	166,789	25,687,314	2,984,550	15,775,189	319,038	6.58365×10^7	2.09898
$\sum 1st + 2nd\ b\ likelihood >= 1.2585$	331,784	2,391,805	4,566	24,723	123,000	989,363	116,627	80,173	3.73026×10^6	0.607459
$(MB1 - 126.5)^2 + (MB2 - 91.5)^2 <= 24^2$	162,832	87,020	0	999	10,287	73,788	13,635	18,455	204, 184	0.372051

5 Appendix A : Refining the cut-flow of $ZH(Z \rightarrow q\bar{q}, H \rightarrow b\bar{b})$ finding

Total generated	qqHbb	2f	SW	SZ	WW	ZZ	Mixed	ZH	Total Bkg	$\frac{\sqrt{S+B}}{S(\%)}$
	442,687	801,152,083	19,517,399	9,072,946	50,826,211	6,389,422	21,837,924	697,805	9.09494×10^8	6.81
N_{pfo} >= 136	256229	4399917	1800	1692	5438721	985691	4590934	147016	1.55658×10^7	1.55
visible energy >= 186	255411	4187355	1050	829	5384959	977109	4548862	143640	1.52438×10^7	1.54
leading lepton energy <= 3.5	121699	214578	75	0	67322	6826	57638	83795	430234	0.61
leading neutral energy <= 28.5	113494	164440	0	0	48425	4993	40236	76403	334497	0.59
thrust <= 0.861875	103405	10816	0	0	31930	2557	26714	73365	145382	0.48
$\sum$ *1st + 2nd b likelihood* >= 0.858125	99270	3208	0	0	1033	1206	1180	19354	25981	0.36

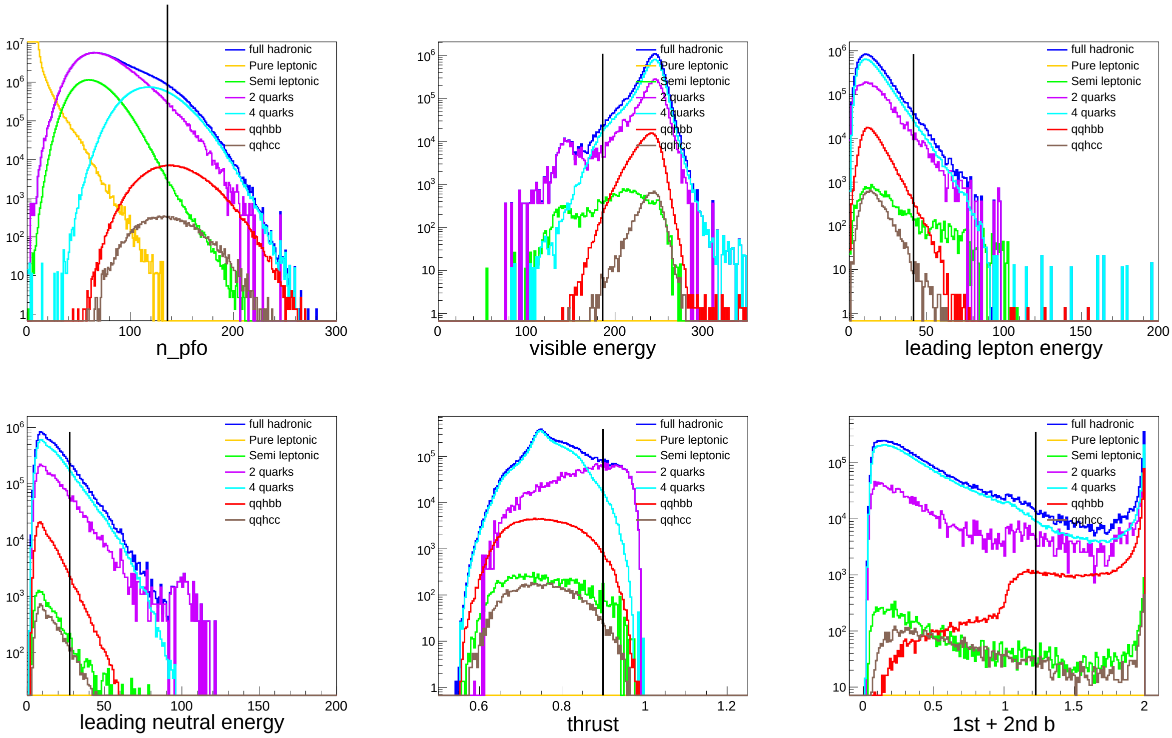

Figure 4: Ask qqhbb as signal from beginning, the distribution of multiplicity, visible energy, leading lepton energy, leading neutral energy, thrust and sum of leading two jets b likelihood on several classes.

If set qqhbb as signal from the begining, the final accuracy is 0.36%, the left signal entries is 99270 and background 25981.

6 Appendix B : finding $ZH(Z \to q\bar{q}, H \to c\bar{c})$

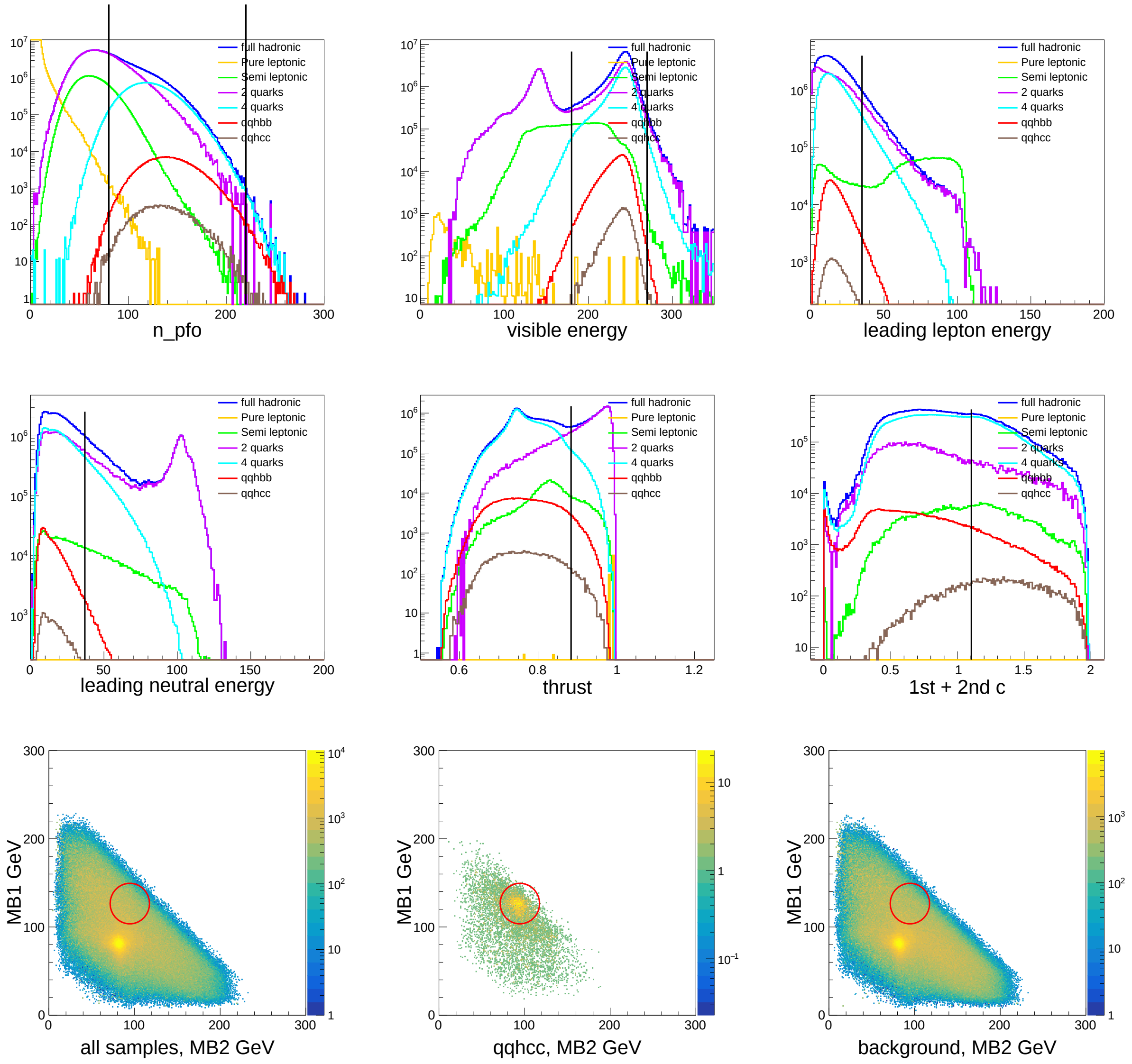

Figure 5: The distribution of multiplicity, visible energy, leading lepton energy, leading neutral energy, thrust, the sum of leading two c jets likeliness on several classes.

	qqHcc	2f	SW	SZ	WW	ZZ	Mixed	ZH	Total Bkg	$\frac{\sqrt{S+B}}{S}$ (%)
Total generated	20,496	801,152,062	19,517,399	9,072,948	50,826,211	6,389,426	21,837,925	1,119,997	9.10×10^8	147.18
$220 >= N_{pfo} >= 80$	20,384	103,633,636	2,294,799	448,803	25,597,117	3,569,366	17,585,144	866,112	1.54×10^8	60.88
$270 >= visible\ energy >= 180$	20,291	63,485,025	1,629,637	239,348	22,630,947	3,100,281	17,320,400	715,355	1.09×10^8	51.49
$leading\ lepton\ energy <= 35$	19,011	53,880,105	202,978	47,864	18,979,861	2,632,937	15,475,906	618,748	9.18×10^7	50.41
$leading\ neutral\ energy <= 37$	16,605	24,652,632	34,730	4,850	13,564,580	2,053,222	11,166,128	562,806	5.20×10^7	43.45
$thrust <= 0.884375$	15,630	5,880,067	28,932	3,557	13,105,545	1,933,283	10,824,073	537,806	3.23×10^7	36.38
$\sum 1st + 2nd\ c\ likelihood >= 1.1045$	10,724	1,452,521	6,474	618	4,465,603	509,405	3,705,982	112,999	1.02×10^7	29.87
$(MB1 - 126.5)^2 + (MB2 - 94.5)^2 <= 23^2$	4,660	106,237	15	44	346,400	37,267	325,937	14,293	830, 193	19.61