Higgs boson decay to light jets at the LHC

Linda M. Carpenter, Tao Han, Khalida Hendricks, Zhuoni Qian, and Ning Zhou

Higgs Boson Decay to Light Jets at the LHC

Linda M. Carpenter,[a] **Tao Han,**[b,c,d] **Khalida Hendricks,**[a] **Zhuoni Qian,**[b] **Ning Zhou**[c,d]

[a] *Department of Physics, Ohio State University 191 West Woodruff Ave., Columbus OH, 43210, U.S.A.*

[b] *Pittsburgh Particle Physics, Astrophysics, and Cosmology Center, Department of Physics and Astronomy, University of Pittsburgh, 3941 O'Hara St., Pittsburgh, PA 15260, U.S.A.*

[c] *Department of Physics, Tsinghua University, P.R. China*

[d] *Collaborative Innovation Center of Quantum Matter, Beijing, China*

E-mail: lmc@physics.osu.edu, than@pitt.edu, hendricks.189@osu.edu, zhq8@pitt.edu, nzhou2015@tsinghua.edu.cn

ABSTRACT: We study the Higgs boson (h) decay to two light jets at the 14 TeV High-Luminosity-LHC (HL-LHC), where a light jet (j) represents any non-flavor tagged jet from the observational point of view. The decay mode $h \to gg$ is chosen as the benchmark since it is the dominant channel in the Standard Model (SM), but the bound obtained is also applicable to the light quarks ($j = u, d, s$). We estimate the achievable bounds on the decay branching fractions through the associated production Vh ($V = W^{\pm}, Z$). Events of the Higgs boson decaying into heavy (tagged) or light (un-tagged) jets are correlatively analyzed. We find that with 3000 fb^{-1} data at the HL-LHC, we should expect approximately 1σ statistical significance on the SM $Vh(gg)$ signal in this channel. This corresponds to a reachable upper bound BR($h \to jj$) ≤ 4 BR$^{SM}(h \to gg)$ at 95% confidence level. A consistency fit also leads to an upper bound BR($h \to cc$) < 15 BR$^{SM}(h \to cc)$ at 95% confidence level. The estimated bound may be further strengthened by adopting multiple variable analyses, or adding other production channels.

KEYWORDS: Higgs boson, LHC.

1   Introduction

As we know for the Higgs detection at the LHC, $\gamma\gamma$ and ZZ were the discovery channels for the Standard Model (SM)-like Higgs boson (h) [1, 2]. Next came the WW decay channel, all have been measured with more than 5σ significance at Run I by both experiments ATLAS [3] and CMS [4]. While the ZZ, WW channels are tree-level processes, most directly related to the electroweak symmetry breaking (EWSB) with the coupling strength proportional to $M_{W,Z} \sim gv$, the Higgs coupling to the top quark is best inferred from its contribution to the production $gg \to h$ and the decay $h \to \gamma\gamma$ with a fitted accuracy of around 30% [5]. A direct measurement from Higgs and top associated production is yet to be established [6, 7]. For the lepton side, the challenging decay channel $h \to \tau^+\tau^-$ has also reached 5σ observation with a combined analysis of the two experiments [5]. With the upgrade of LHC to its higher center of mass energy at Run II and more accumulated data, the difficult mode $h \to b\bar{b}$ is expected to reach 5σ soon after several hundreds fb^{-1} at 14 TeV [8]. Thus, the Higgs couplings to the heaviest generation of fermions will soon be settled to the values expected from the Standard Model (SM) prediction at an accuracy of about 20% [9], and verifying the pattern of non-universal Yukawa couplings.

We next consider the LHC upgrade to a total integrated luminosity of 3000 fb^{-1} at 14 TeV (HL-LHC). While the precision measurements of those couplings will continue in the LHC experiments, it is imperative to seek other "rare decay" channels, in the hope of uncovering any deviations from the SM. Among the rare channels, it is perhaps most promising to observe the clean mode $gg \to h \to \mu^+\mu^-$ [10], despite the small decay branching fraction BR$(h \to \mu^+\mu^-) \sim 2 \times 10^{-4}$. A 5σ observation may be conceivable at the end of the run for HL-LHC with 3000 fb^{-1} [9], which would be of significant importance to establish the pattern of the Yukawa couplings by including a second generation fermion. For the other hadronic channels, it would be extremely challenging to make any measurements at the LHC due to the overwhelmingly large QCD backgrounds.[1]

The most promising production mechanism for the hadronic decay signal of the Higgs boson is

$$pp \to Vh, \quad \text{where} \quad V = W^{\pm}, \ Z. \tag{1.1}$$

With W/Z decaying leptonically to serve as effective triggers, the Higgs signal may be detected from the construction of its invariant mass of the hadronic products. To sufficiently suppress the large QCD backgrounds, it was proposed [14] to look for highly-boosted events for $h \to b\bar{b}$ against the leptonic W/Z. Studies on these processes at HL-LHC shows a $\approx 20\sigma$ (9σ) significance for the signal $Vh, h \to b\bar{b}$, with statistical (systematic added) uncertainty estimated [8]. Marching to the channel involving the second generation quarks, the sensitivity to $Vh, h \to c\bar{c}$

---

[1]Due to the much cleaner experimental environment, a lepton collider such as International Linear Collider (ILC) [11] or a circular e^+e^- collider [12, 13], running at the Zh threshold or higher energies, will give us much better sensitivity to the hadronic decays of the Higgs. The expected accuracy on $h \to gg$ and $h \to c\bar{c}$ will be 7% (2.3%) and 8.3% (3.1%) respectively, with the 250 GeV (1TeV) mission [11].

is significantly worse. Bounds are extrapolated in a recast study in Ref. [15] to be ~ 6.5 times the SM value (statistic errors assumed only). This is expected, given that $\mathrm{BR}(h \to b\bar{b})$ is ~ 20 times larger than $\mathrm{BR}(h \to c\bar{c})$, that expected b-tagging is twice as efficient as c-tagging, and that the dominant background $Vbb(cc)$ in the relevant kinematic region is about the same order. An interesting proposal to search for $h \to J/\psi + \gamma$ [16] does not seem to increase the observability for $hc\bar{c}$ coupling due to too low an event rate [17, 18]. Another study on $h + c$ associated production estimates a bound of order one on the SM charm Yukawa coupling at the end of the HL-LHC run [19].

It is natural to ask to what extent one would be able to search for other hadronic decays of the Higgs boson. We here quote the updated calculations of the branching fractions for the 125 GeV Higgs boson decay hadronically in the SM [20]

$$\mathrm{BR}(h \to b\bar{b}) = 58.2\%, \quad \mathrm{BR}(h \to c\bar{c}) = 2.89\%, \tag{1.2}$$

$$\mathrm{BR}(h \to gg) = 8.18\%, \quad \mathrm{BR}(h \to u\bar{u}, d\bar{d}, s\bar{s}) < 0.03\%. \tag{1.3}$$

While the decay rates to light quarks predicted in the SM would be too small to be observable, the decay to a pair of gluons, mediated via the heavy top quark, will be nearly three times larger than the $c\bar{c}$ channel. The experimental signatures for those channels would be to search for the un-tagged light jet pairs jj, which form a mass peak near the Higgs boson mass m_h. Obviously, the lack of a heavy-flavor tag makes background suppression difficult. We also point out that the event sample gets "contamination" from mis-tagged events of the leading decay $h \to b\bar{b}$ and $c\bar{c}$ as well. The individual event samples need to be correlatively quantified and treated.

In this work we set out to study Higgs decay to a pair of light un-tagged jets $h \to jj$, in the associated production channel as in Eq. (1.1). We will exploit the leptonic final state decays of the electroweak gauge bosons, and employ a hadronic tag for the Higgs boson while optimizing the mass reconstruction. We argue that a 1σ sensitivity of 1 (or 4) times the SM value with statistic (or systematic) uncertainties can be achieved for the case where the Higgs decays to un-tagged jets. This is achieved with a judicious choice of kinematic discriminants and a combination of the final state channels. Together with $h \to b\bar{b}$ and $h \to c\bar{c}$ studies, the un-tagged channel puts an independent dimension of bound in the space of branching ratios of Higgs decays to quarks and gluons. Assuming a well measured ggh coupling at the end of HL-LHC [9], the result further puts comparable but independent constraints on the light-quark Yukawa couplings. We also estimate that this channel may offer a better probe to the strange-quark Yukawa coupling.

This paper proceeds as follows, Section 2 specifies the signal and dominant background processes. Section 3 describes and presents the detailed analyses and gives the main results in terms of the cut-efficiency tables and figures. In the same section, we also study how to control the systematic errors for the large backgrounds. Section 4 describes an alternate search strategy based on momentum balance discriminants. Section 5 calculates the signal sensitivity and presents obtained constraints on Higgs couplings to quarks and gluons in a correlated manner, while Section 6 summarizes and concludes.

2 Signal and Background Processes

As discussed above, the promising channel in which to study the Higgs decay to light jets is the associated production with an electroweak gauge boson W or Z, which subsequently decays to leptons. Depending on the production mechanisms and the final states, we consider the following subprocesses

$$q\bar{q} \to W^{\pm}h \to \ell^{\pm}\nu + jj, \tag{2.1}$$

$$q\bar{q}, \ gg \to Zh \to \begin{cases} \ell^{+}\ell^{-} + jj, \\ \nu\bar{\nu} + jj, \end{cases} \tag{2.2}$$

where $\ell = e, \mu$ and $j = g$ or u, d, s. Practically, j is a gluon as expected in the SM. We thus generically denote the SM signal by $Vh(gg)$, whenever convenient. In our calculations, events are generated with MadGraph at the leading order, with "NNPDF23_nlo_as_0119_qed" as the PDF set [21]. For the $gg \to Zh$ process via the quark loops, we use Madgraph_NLO [22] and Madspin [23]. This channel contributes about $10\% - 20\%$ to the total Zh production rate. We apply an overall rescaling of QCD K-factors to the signal processes, to match the total NNLO QCD and NLO EW cross section results taken from summary of Higgs cross section working group [20]. The K-factors are about 2 and 1.2 for the gg and $q\bar{q}$, respectively. We include the finite masses for the fermions running in the loop in the gg initiated process. Some care is needed regarding the gg process because of its different transverse momentum (p_T) dependence and sensitivity to new physics contribution in the loop as discussed in Ref. [24]. In Fig. 1, we compare the Higgs boson transverse momentum distributions for the signal processes $q\bar{q} \to Zh$ and $gg \to Zh$. The $q\bar{q}$-initiated channel peaks at $p_{T(h)} \approx 50$ GeV, a typical mass scale associated with the final state particles of Zh. The gg-initiated channel peaks at around $p_{T(h)} \approx 150$ GeV, due to the top mass threshold enhancement. The differential cross section of gg drops faster than $q\bar{q}$ with increasing $p_{T(h)}$, due to the destructive interference between the triangle and box diagrams.

The Higgs is further decayed according to the branching ratios listed in Ref. [20]. Events are then showered and hadronized using PYTHIA6 [25], and run through DELPHES [26] for detector simulation and jet reconstruction. For the SM backgrounds, we mainly consider the dominant irreducible background process $V + jj$ at LO, where the V decays and contributes accordingly to the three signal channels. At the generator level, we apply some basic cuts on the jets to remove infrared and collinear divergences for the QCD background processes

$$p_{T(j)} > 20 \text{ GeV}, \quad |\eta_j| < 3, \quad R_{jj} > 0.4. \tag{2.3}$$

The hadronic jets are reconstructed with anti-kt jet algorithm with a cone size $R = 0.4$. In our future analyses, we will be considering a relatively boosted Higgs recoiling off of the vector boson. Therefore, to improve the simulation statistics, we also add a generator-level cut on the vector boson
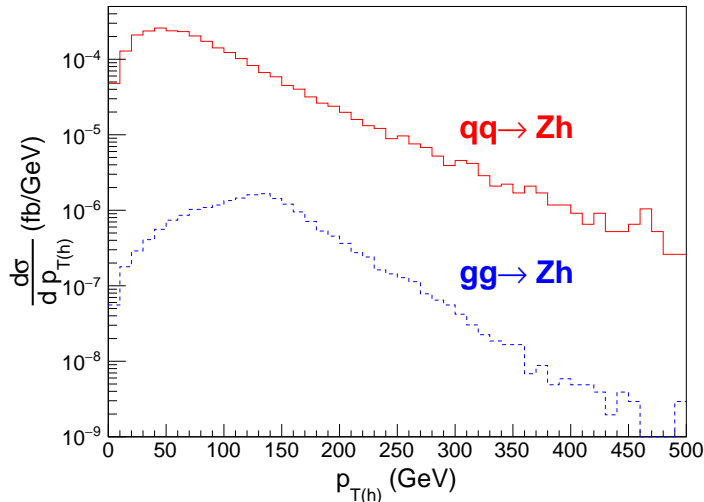
$$p_{T(V)} > 150 \text{ GeV}. \tag{2.4}$$

Figure 1. Higgs boson transverse momentum distribution for the signal processes $qq \to Zh$ (upper solid curve) and $gg \to Zh$ (lower dashed curve) at the 14 TeV LHC.

In Table 1 we give the cross sections used for our signal and background processes including the basic cuts in Eq. (2.3) and with various p_T thresholds for the vector boson. The first is the total cross section with no $p_{T(V)}$ cut, the second and third demand $p_{T(V)}$ cuts of 150 and 200 GeV respectively. No cuts on the final state leptons are applied for the table.

A few remarks are in order. Although we have re-weighted our events by the K-factors to account for the NNLO QCD/NLO EW corrections for the signal rate, we could not claim the theoretical accuracy to this level, in particular in the extreme kinematic region as in Eq. (2.4). Indeed, perturbative calculations tend to lead to harder p_T spectrum than our approximation via the final state radiation (FSR) or the initial state radiation (ISR) as in PYTHIA. On the other hand, the extended $p_{T(V)}$ spectrum is rather smooth and the good experimental measurement with the large data sample would make the imprecise knowledge from theory less crucial. As for the higher order corrections to the hadronic Higgs decay, we believe that the effects are not significant given the color-singlet nature of the Higgs boson and our rather inclusive treatment of the final-state jets, as discussed in detail in later sections.

3 Signal Selection

In further studying the signal characteristics in Eqs. (2.1) and (2.2), we categorize the channels according to the zero, one, or two charged leptons from the vector boson decays. In addition, the signal has two leading jets from the Higgs decay, with invariant mass of the Higgs boson. At high $p_{T(h)}$, the distance between the two hadronic jets can be estimated as

$$R_{jj} \approx \frac{1}{\sqrt{z(1-z)}} \frac{m_h}{p_{T(h)}}, \tag{3.1}$$

σ (fb)	cuts Eq. (2.3)	+ Eq. (2.4)	+ $p_{T(V)} > 200$ GeV
$q\bar{q} \rightarrow Zh \rightarrow \ell^+\ell^-\ gg$	3.5	0.39	0.17
$gg \rightarrow Zh \rightarrow \ell^+\ell^-\ gg$	0.71	0.20	6.2×10^{-2}
$q\bar{q} \rightarrow Zjj \rightarrow \ell^+\ell^-\ jj$	2.5×10^5	1.2×10^4	4.8×10^3
$q\bar{q} \rightarrow Wh \rightarrow \ell\nu\ gg$	20	2.3	0.99
$q\bar{q} \rightarrow Wjj \rightarrow \ell\nu\ jj$	2.5×10^6	1.0×10^5	3.9×10^4
$pp \rightarrow t\bar{t} \rightarrow \ell\nu jj b\bar{b}$	1.1×10^5	1.5×10^4	5.7×10^3
$q\bar{q} \rightarrow Zh \rightarrow \nu\nu\ gg$	11	1.2	0.50
$gg \rightarrow Zh \rightarrow \nu\nu\ gg$	2.1	0.60	0.18
$q\bar{q} \rightarrow Zjj \rightarrow \nu\nu\ jj$	7.4×10^5	3.6×10^4	1.4×10^4

Table 1. Cross sections in units of fb for signal and dominant background processes, with the parton-level cuts of Eq. (2.3), and boosted regions $p_{T(V)} > 150,\ 200$ GeV.

where $z,\ 1-z$ are the momentum fraction of the two jets. The LO parton-level distributions of three kinematic discriminants for the Zh channel, the transverse momentum $p_{T(Z)}$, the jet separation R_{jj}, and the di-jet invariant mass m_{jj}, are shown in Fig. 2, comparing the signal (solid) and dominant background (dashed), after the generator-level cuts as in Eqs. (2.3) and (2.4). Obviously, $p_{T(Z)}$ is singular for the QCD background as seen in Fig. 2(a). The two jet separation R_{jj} in Fig. 2(b) shows the either collinear feature from the parton splitting in the final state radiation (FSR) or back-to-back near π due to the initial state radiation (ISR) for the background process, and is narrowly populated near $2m_h/p_{T(h)}$ for the signal. The resonance bump near m_h is evident as in Fig. 2(c). Because of the small rate, the signal curves have been scaled up by a factor of 5000. We also show an event scatter plot in Fig. 2(d), where the (red) dense band with crosses presents the signal events and the (blue) dots show the background events. We see the strong correlation between the boosted $p_{T(Z)}$ and collimated jets with smaller R_{jj}.

To suppress the huge QCD di-jet backgrounds, we must optimize the reconstruction of the Higgs mass. There are two common methods to reconstruct hadronic decays of Higgs boson depending on the kinematical configurations. One is the sub-structure (fat-jet) approach: an early example for Higgs search in $b\bar{b}$ channel was introduced in Ref. [14]. Because of the highly boosted nature of the Higgs boson, a fat-jet identified as the hadronic decay products of the Higgs boson is first selected. Various jet substructure observables and techniques such as mass-drop and filtering [14], pruning [27], trimming [28], N-subjettiness [29] etc. can be applied on the fat-jet to further improve the reconstruction of the invariant mass. For a recent review, see Ref. [30]. The other approach is to simply resolve the leading jets. This is the common practice when the Higgs is produced not far from the threshold, and the Higgs is identified as the sum of the two leading jets. Experimentally, the anti-kt jet algorithm, given its regular jet shape, gives good reconstruction of hadronic jets, and is the default hadronic jet reconstruction algorithm used at ATLAS/CMS. The $Vh(b\bar{b})$ search at
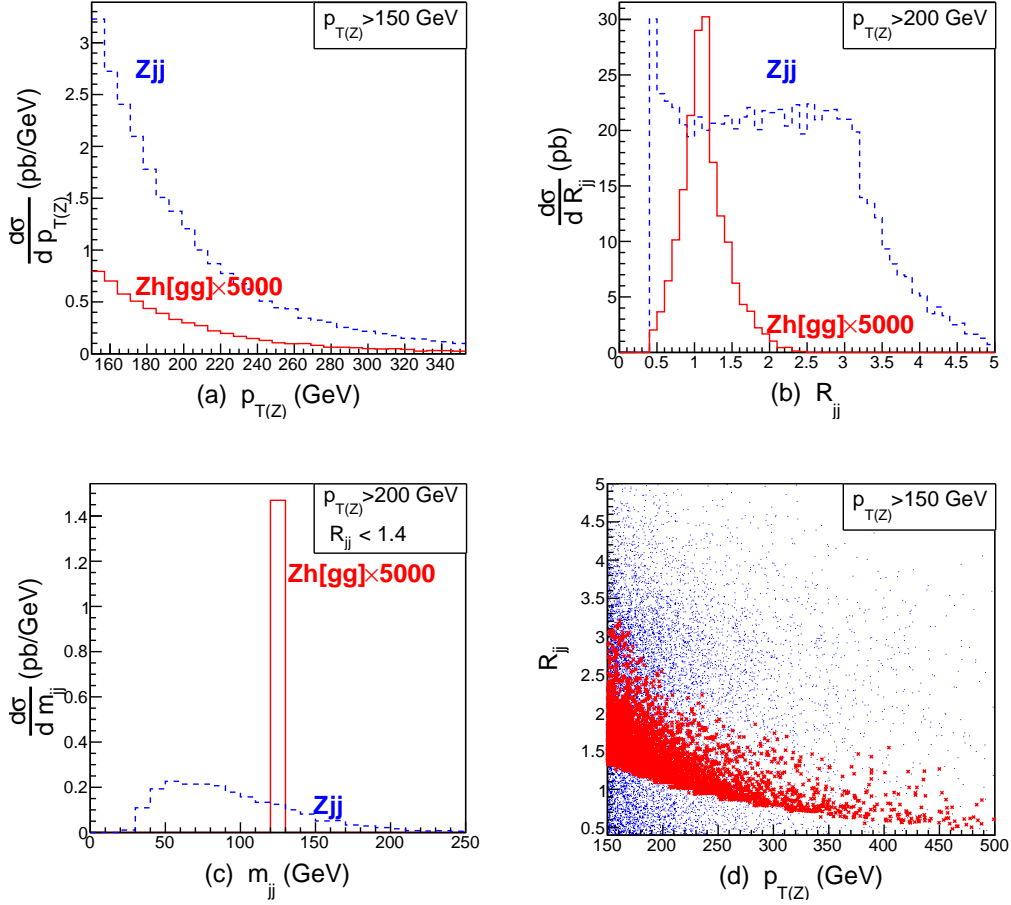
Figure 2. Kinematical distributions of the signal process $pp \to Zh, h \to gg$ (solid curves, scaled up by a factor of 5000) and the leading background $pp \to Zjj$ (dashed curves) for (a) $p_{T(Z)}$, (b) R_{jj}, (c) m_{jj}, and (d) event scatter plot in $R_{jj} - p_{T(Z)}$ plane, with the (red) dense band with crosses as the signal events and (blue) dots as the background. Generator level cuts of Eqs. (2.3) and (2.4) have been applied.

LHC is currently carried with the two resolved jet with anti-kt $R = 0.4$ method. In a recent analysis [31] the two methods are compared for the $Wh, h \to bb$ process for LHC14 in the kinematic region $200 \text{ GeV} < p_{T(h)} < 600 \text{ GeV}$. The resolved approach is better in the $200 \text{ GeV} < p_T < 300 \text{ GeV}$ range. The jet-substructure approach is significantly better in the $p_T > 600 \text{ GeV}$. The results are qualitatively expected, since the high p_T corresponds to a smaller cone-size of the fat-jet as argued in Eq. (3.1).

Since the signal events tend to populate near the kinematic threshold, we will exploit the resolved method with two hard jets. However, additional QCD radiations from the highly energetic jets are not negligible. Kinematically, it gives a reconstructed di-jet mass peak smeared towards lower value. Some related effects including the NLO correction is studied
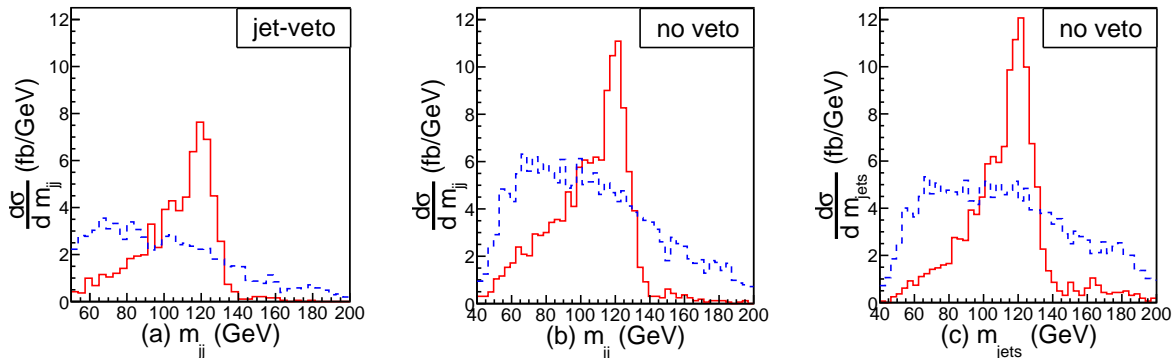
Figure 3. Invariant mass distributions m_{jj} of the signal process $pp \to Zh, h \to gg, Z \to \ell\ell$ (solid curves, scaled up by a factor of 5000) and the leading background $pp \to Zjj$ (dashed curves) for (a) with 2 jets only, (b) with 2 leading jets to reconstruct m_{jj}, (c) with 2 leading jets plus other jets together to reconstruct m_{jets}. All selection cuts as in Sec. 3.1 except for m_h cut are applied.

in Ref. [32]. We thus propose a modification of the two-jet-resolved method by including possible additional jets in the decay neighborhood – a "resolved Higgs-vicinity" method. After clustering the jets with anti-kt $\Delta R = 0.4$, two leading p_T jets are clustered as the "Higgs-candidate". Then additional jets j' are also clustered to the "Higgs candidate" in sequence of angular vicinity, whenever $R_{Hj'} \leq R_{\max}$. For the rest of the analyses, we choose

$$R_{\max} = 1.4. \tag{3.2}$$

The optimal method is to select events with two leading p_T jets that satisfy $R_{jj} \leq R_{\max}$, and add to the di-jet system any sub-leading jets within the distance $R_{\max}$. In practice, we find that including one additional hard radiation in the decay is sufficient. In Fig. 3 we compare several resolved-jet methods in their reconstruction of the Higgs mass, against the Vjj background. The central and hard jet requirements are $p_{T(j)} > 30$ GeV and $|\eta_j| < 2.5$. In Fig. 3(a), we reconstruct the Higgs with the two leading p_T jets and veto events with more than two central hard jets. As shown in the plot, the veto method removes the background most efficiently, the cut also reduces the signal significantly. Fig. 3(b) shows the 2jet-inclusive case, which is the same as (a) but does not veto additional jets. It improves the signal rate, but the signal mass peak is still smeared to the lower value. Fig. 3(c) is the "resolved Higgs-vicinity" method, which adds the additional hard jet, and sharpens the mass peak to help increase the overall $S/\sqrt{B}$ sensitivity.

We study the sensitivity to pile-up contamination of this reconstruction method. In Fig. 4, we compare it with the two jet resolved method adding pile-up samples in DELPHES. As expected, the additional-jet method is more sensitive to the pile-up jets, yet still retains a slight advantage even under pile-up value $\langle \mu \rangle = 140$ [33].

In the following, we describe the searches with the detailed signal and background analyses, for the channels with two, one and zero charged leptons, respectively. For simplicity, we
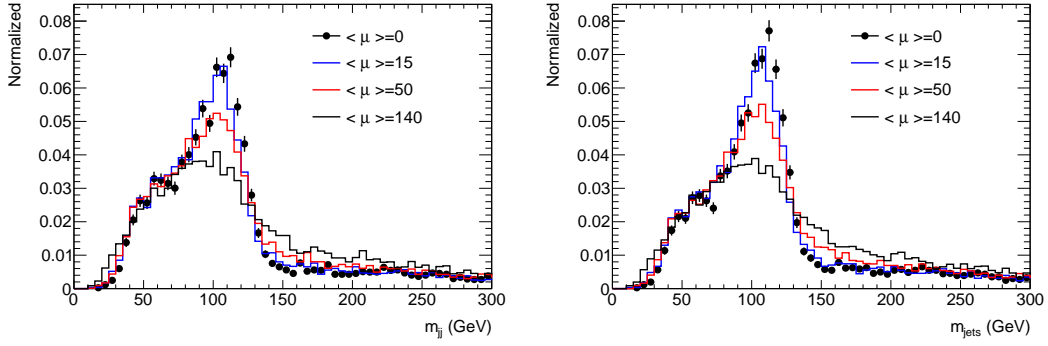
Figure 4. Invariant mass distributions constructed from (a) two-jet events and (b) three-jet events with different pile-up values $\langle \mu \rangle = 0, 15, 50, 140$, respectively.

use 2 jets reconstruction of the mass peak from now on.

3.1 $\ell^+\ell^- + jj$ channel

For the two-lepton channel, we simulate the signal processes as in Eq. (2.2) with $Z \to \ell^+\ell^-$, $h \to gg$. We require exactly one pair of charged leptons $\ell^\pm = e^\pm$ or $\mu^\pm$, same flavor, opposite charge, along with at least two energetic jets. The dominant background is by far from $Z + jj$. The two leading p_T jets are required to be close by having a separation less than $R_{\max} = 1.4$, and an invariant mass between 95 and 150 GeV. They satisfy the following acceptance cuts

- 2 leptons with $p_{T(l)} > 30$ GeV and $|\eta_l| < 2.5$

- $p_{T(\ell\ell)} > 200$ GeV

- at least 2 jets with $p_{T(j)} > 30$ GeV and $|\eta_j| < 2.5$

- $R_{j_1 j_2} < 1.4$

- 95 GeV$< m_h < 150$ GeV

The di-jet mass window around m_h is chosen to optimize the $S/\sqrt{B}$ at HL-LHC. Table 2 shows the efficiency of applying the sequence of cuts. The overall efficiencies are about 14%, 7.6%, for the $q\bar{q}$, gg initiated signal processes, respectively, and about 1.9% for the background process. We would like to point out that from only the statistical sense, the signal sensitivity $S/\sqrt{B}$ would not be notably increased from the generator level results to that with final cuts. However, the fact that the background is reduced by around two orders of magnitude helps to control the systematic uncertainties, as we will discuss later.

cut eff (%)	$q\bar{q} \to Zh \to \ell^+\ell^- gg$	$gg \to Zh \to \ell^+\ell^- gg$	$q\bar{q} \to Zjj \to \ell^+\ell^- jj$
σ (fb)	3.9×10^{-1}	2.0×10^{-1}	1.2×10^4
2 leptons	59%	52%	40%
≥ 2 jets	51%	49%	32%
$70 < m_{ll} < 110$	50%	49%	31%
$p_{T(\ell\ell)} > 200$ GeV	26%	23%	16%
$R_{j_1 j_2} < 1.4$	21%	12%	5.3%
$95 < m_h < 150$ GeV	14%	7.6%	1.9%
final (fb)	5.4×10^{-2}	1.5×10^{-2}	2.4×10^2

Table 2. The consecutive cut efficiencies for signal $\ell^+\ell^-\ jj$ and dominant background processes at the LHC.

3.2 $\ell^\pm + \not{E}_T + jj$ channel

For the one-lepton channel, we look at signal process in Eq. (2.1) with $W \to \nu\ell$, $h \to gg$. The dominant backgrounds are $W + jj$ and $t\bar{t}$. Similar to the last section, the acceptance cuts are

- one lepton $p_{T(\ell)} > 30$ GeV and $|\eta_\ell| < 2.5$

- $p_{T(\nu\ell)} > 200$ GeV, $\not{E}_T > 30$ GeV

- at least 2 jets with $p_{T(j)} > 30$ GeV and $|\eta_j| < 2.5$

- $R_{j_1 j_2} < 1.4$

- 95 GeV $< m_h < 150$ GeV.

The W transverse momentum $p_{T(\nu\ell)}$ can be reconstructed from the charged lepton plus the missing transverse momentum $\not{E}_T$. Table 3 shows the cut-flow at various stages of the cuts applied. The overall efficiencies are about 18% for the $q\bar{q}$ initiated signal process, and about 2.5%, 2.5% for the Wjj, $t\bar{t}$ background processes, respectively.

3.3 $\not{E}_T + jj$ channel

The zero-lepton channel is studied with signal processes as in Eq. (2.2) with $Z \to \nu\nu$, $h \to gg$. The dominant background again mainly is $Z + jj$. Similar to the above, the cuts acceptance are

- lepton veto with $p_{T(\ell)} > 30$ GeV $|\eta_\ell| < 2.5$

- $\not{E}_T > 200$ GeV

- at least 2 jets with $p_{T(j)} > 30$ GeV $|\eta_j| < 2.5$

- $R_{j_1 j_2} < 1.4$

cut eff (%)	$q\bar{q} \to Wh \to \ell\nu gg$	$q\bar{q} \to Wjj \to \ell\nu jj$	$t\bar{t} \to \ell\nu jj b\bar{b}$
σ (fb)	2.3	1.0×10^5	1.5×10^4
$\not{E}_T > 30$ GeV	94%	87%	93%
1 lepton	72%	52%	62%
$p_{T(\ell\nu)} > 200$ GeV	39%	24%	26%
≥ 2 jets	35%	20%	22%
$R_{j_1 j_2} < 1.4$	27%	6.8%	11%
$95 < m_h < 150$ GeV	18%	2.5%	2.5%
final (fb)	4.1×10^{-1}	2.5×10^3	3.7×10^2

Table 3. The consecutive cut efficiencies for signal $\ell^{\pm} \not{E}_T\ jj$ and dominant background processes at the LHC.

cut eff (%)	$q\bar{q} \to Zh \to \nu\nu gg$	$gg \to Zh \to \nu\nu gg$	$q\bar{q} \to Zjj \to \nu\nu jj$
σ (fb)	1.2	6.0×10^{-1}	3.6×10^4
$\not{E}_T > 200$ GeV	49%	44%	42%
≥ 2 jets	45%	43%	35%
$R_{j_1 j_2} < 1.4$	36%	25%	12%
$95 < m_h < 150$ GeV	23%	15%	4.5%
final (fb)	2.7×10^{-1}	8.9×10^{-2}	1.6×10^3

Table 4. The consecutive cut efficiencies for signal $\not{E}_T\ jj$ and dominant background processes at the LHC.

- 95 GeV $< m_h < 150$ GeV.

The $\not{E}_T$ is essentially from $p_{T(Z)}$. Table 4 shows the cut-flow at various stages of the cuts applied. The overall efficiencies are about 23%, 15%, for the $q\bar{q}$, gg initiated signal processes, respectively, and about 4.5% for the background process.

Results presented in the above three sections have been double checked by other approaches.

3.4 Background control

As calculated earlier and presented in the previous tables, the signals for $h \to gg$ in the SM associated with W/Z to leptons at the 3000 fb^{-1} HL-LHC may lead to sizable event rates, with about 200 events for the $\ell^+\ell^-$ channel, 1300 events for the $\ell^{\pm}\nu$ channel, and 1200 events for the $\nu\nu$ channel, respectively. However, the difficulty is the overwhelmingly large SM background, with a signal-to-background ratio at the order of 10^{-4}. As such, one must be able to control the systematic errors to sub-percent in order to reach statistically meaningful result. This is an extremely challenging job, and one would not be able conclude without real data to show the detector performances. On the other hand, there are ideas to shoot
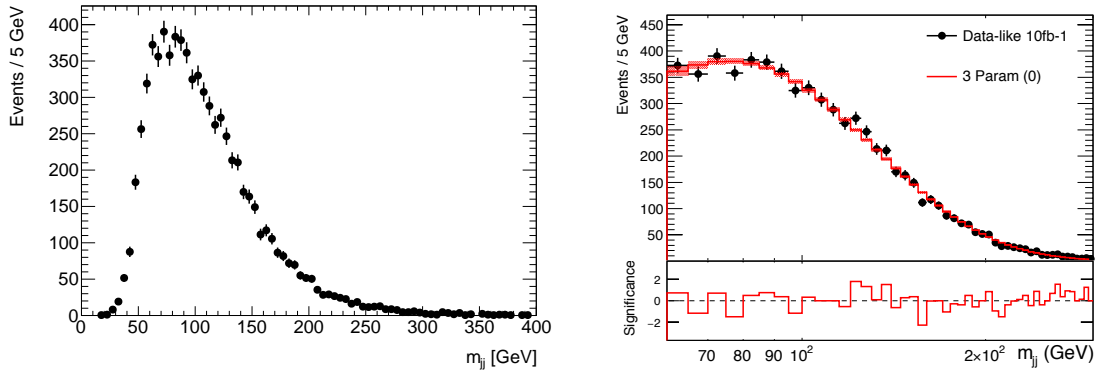
Figure 5. Invariant mass distribution m_{jj} for $Z(\ell^+\ell^-)$+jets at the 14 TeV LHC for (a) MC simulated events normalized to 10 fb^{-1}, and (b) fitted spectrum from three-parameter ansatz function in Eq. (3.3) range from 60 GeV to 300 GeV (solid curve).

at the goal. Here we adopt one of the commonly considered methods and demonstrate our expectations.

For the two lepton and $\not{E}_T$ channel, the dominant background is the SM $Z+jj$ production. With current selection, the two jet invariant mass spectrum is smoothly decreasing within a range of $[60, 300]$ GeV and our signal region lies between 95 GeV and 150 GeV. Making use of the well-measured side-bands, the estimation of background contribution in the signal region could be obtained directly from a fit to the m_{jj} distribution. We generated $Z+$jets samples with MadGraph generator corresponding to 10 fb^{-1} and passed the events through PYTHIA and DELPHES to simulate the parton shower and ATLAS detector effect. We adopt a parameterization ansatz to fit the distribution in the m_{jj} range from 60 GeV to 300 GeV

$$f(z) = p_1(1 - z)^{p_2} z^{p_3}, \tag{3.3}$$

where p_i are free parameters and $z = m_{jj}/\sqrt{s}$. This ansatz is found to provide a satisfactory fit to the generated $Z+$jets MC simulation at 14 TeV, as shown in Fig. 5.

In order to estimate the uncertainty of background determination for 3000 fb^{-1} integrated luminosity, we take this three-parameter function in Eq. (3.3) as the baseline to generate the data-like spectrum following Poisson fluctuation. Figure 6 shows the generated spectra for 300 fb^{-1} and 3000 fb^{-1}. We fit these spectra with three-parameter, four-parameter and five-parameter functions within the range of $[60, 300]$ GeV but excluding the signal region $[95, 150]$ GeV. The fitting results and uncertainties are summarized in Figure 7 and Table 5. Besides the three-parameter function, four-parameter and five-parameter functions are tested as below

$$f(z) = p_1(1 - z)^{p_2} z^{p_3 + p_4 \log(z)}, \quad f(z) = p_1(1 - z)^{p_2} z^{p_3 + p_4 \log(z) + p_5 \log^2(z)}. \tag{3.4}$$

We also vary the fitting range from $[60, 300]$ GeV to $[70, 250]$ GeV and $[80, 200]$ GeV to test the stability, which are summarized in Table 6. If we consider the variation due to this fitting
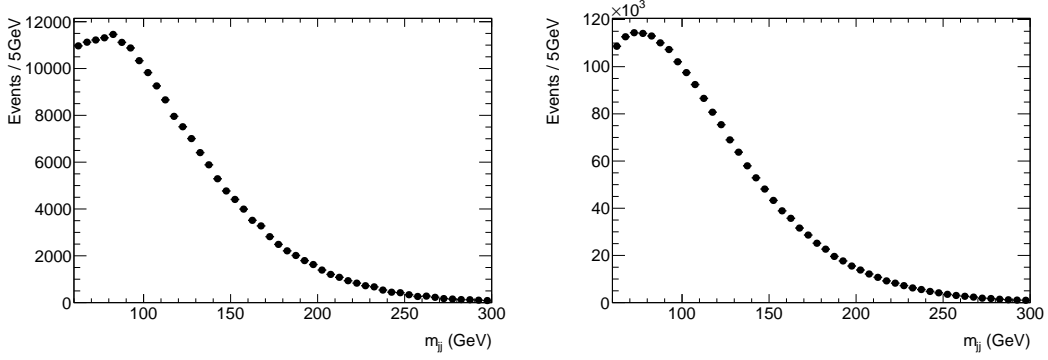
Figure 6. Generated distribution from three-parameter ansatz function in Eq. (3.3) for m_{jj} with (a) 300 fb^{-1}, (b) and 3000 fb^{-1} (right).
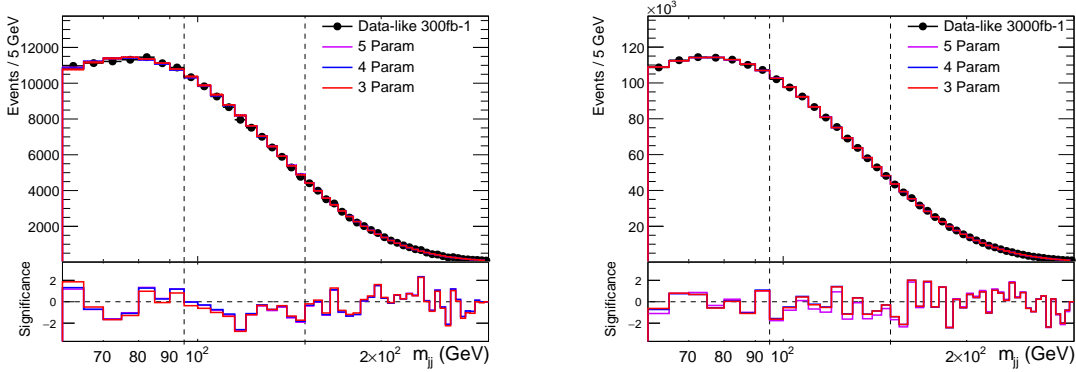

Figure 7. Fitted results for 300 fb^{-1} (left) and 3000 fb^{-1} (right).

Background	300 fb^{-1}	3000 fb^{-1}
Expectation	8.29×10^4	8.26×10^5
3-parameter	$(8.39 \pm 0.05) \times 10^4$	$(8.28 \pm 0.01) \times 10^5$
4-parameter	$(8.38 \pm 0.05) \times 10^4$	$(8.27 \pm 0.01) \times 10^5$
5-parameter	$(8.39 \pm 0.04) \times 10^4$	$(8.29 \pm 0.01) \times 10^5$
Uncertainty	1.32%	0.21%

Table 5. Fitted results for the background rates from various fitting functions as in Eqs. (3.3) and (3.4).

range as another source of systematics, the uncertainty of background estimation of $Z(\ell\ell)$+jets for 3000 fb^{-1} is 0.33%. The uncertainty considered here includes the fitting uncertainty, fitting function variation and fitting range variation, which is largely depending on the statistics of side-band region. The background uncertainty from fitting is dominated by the statistics of side-band regions, which is proportional to the background yield. It is reasonable to assume the jet mass spectrum to be independent of the leptonic decay details of the vector bosons. To

3000 fb^{-1}	True	[60, 300] GeV	[70, 250] GeV	[80, 200] GeV
3-parameter	8.26×10^5	$(8.28 \pm 0.01) \times 10^5$	$(8.26 \pm 0.03) \times 10^5$	$(8.27 \pm 0.05) \times 10^5$

Table 6. Fitted results for the background rate from various fitting ranges by the fitting function in Eq. (3.3).

the first-order estimation, the uncertainties of $Z(\nu\nu)$+jets and $W(\nu\ell)$+jets are comparable at the order of 0.1%. We thus summarize the systematic percentage uncertainties for the three leptonic channels as

$$Z(\ell^+\ell^-) + jj : \ 0.33\%; \quad W(\ell^\pm\nu) + jj : \ 0.10\%; \quad Z(\nu\nu) + jj : \ 0.13\%. \tag{3.5}$$

As seen for example in Table 3 for the one-lepton channel, the $t\bar{t}$ background is subdominant yet not negligible. There are other smaller and non-negligible processes such as semi-leptonic decays of di-boson, which are not included in our current studies since they would not change our conclusions. Full simulation and control shall be required on all the relevant processes once the data is available. Similar to the comment at the end of Sec. 2, we could only claim the theoretical accuracy at the LO for the background processes, given the complicated kinematical acceptance. However, we argue that for our purpose of estimating the signal sensitivity, it suffices to say that the di-jet invariant mass distribution for backgrounds is smooth in the signal region, fitted with simple functions as done above.

4   Alternative Discriminants with Missing Energies

We note that a momentum balance discriminant has been proposed in Ref. [34] as a useful kinematic variable in processes where a new resonant particle is produced in association with a SM vector boson radiated in an initial state, $pp \to R + V$. The transverse momenta of these states should balance

$$p_T^R - p_T^V = 0. \tag{4.1}$$

Due to detector effects and radiation, the measured momentum balance is not perfect and it is particularly more severe for the background since the QCD processes tend to have larger radiation. This is a useful kinematic discriminant between the signal and background [34]. However it is not applicable whenever there is missing energy in the event. In fact, the definition of the missing transverse energy in an event is the negative of the vector sum of the visible p_T. In the above example it offers only a tautology for the momentum balance discriminant. We offer, in the case of events with significant missing energy, a new discriminant to capture the kinematic features of the event. We define this discriminant by calculating the scalar sum of the transverse momenta of the visible particles in the event, and then subtracting the missing transverse energy

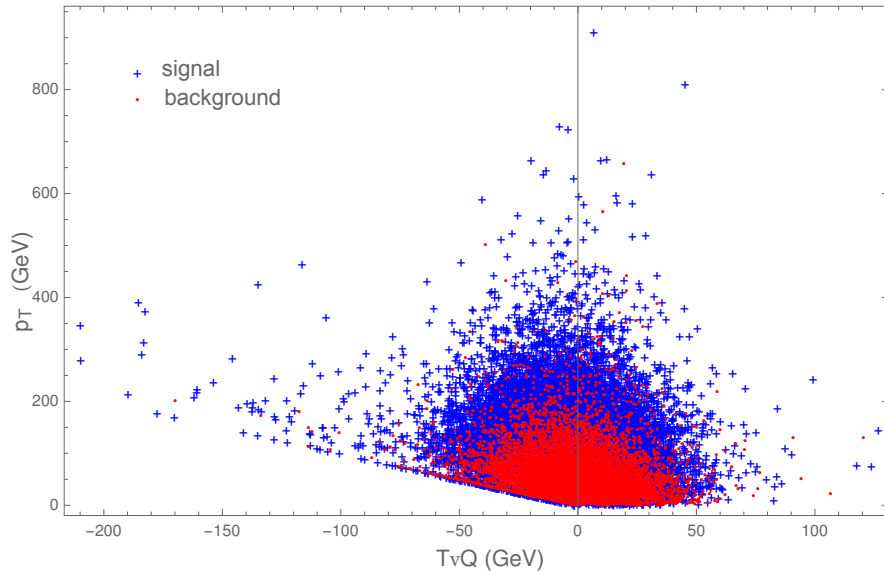$$TvQ \equiv \Sigma_i |p_{Ti}| - |\rlap{/}{E}_T|. \tag{4.2}$$

Figure 8. Scatter plot of 10000 events for the signal (blue crosses) and background (red dots) in the visible $p_T - TvQ$ plane.

This is a version of a momentum balance discriminant, referred as TvQ (Transverse event Quality). Since the missing momentum in an event is defined by the negative of the vector sum $|\Sigma_i \vec{p}_{Ti}|$, the quantity TvQ is the difference between the scalar and vector sums of the visible p_T in the event. TvQ tends to be small when the observable particles are a highly collimated collinear bunch, while it takes a large value when the observable particles spread out and when $R + V$ production is near the kinematical threshold.

It would be more intuitive to look at the signal and background in a two dimensional space of discriminants. Consider the $\not{E}_T$ signal from $pp \to Zh \to \nu\nu \; gg$. We plot the event population in the $p_{T(jj)} - TvQ$ plane as shown in Fig. 8. We see that in the signal sample (blue crosses), regions of large visible p_T correlate with the zero value of TvQ. Events with high boost, and therefore columnated Higgs decay products, correlate with lower values of TvQ as predicted. The QCD background sample Z+jets (red dots), on the other hand, tends to further spread out.

Another simple discriminant, somewhat correlated with TvQ for the Zh final state is a transverse angular variable, ϕ_{Zh} defined as the angle between the missing transverse energy vector and the vector sum of the visible p_T. This is clearly motivated since we expect the Z and h states to be nearly back to back in the event, in contrast to the QCD multiple jet events. We examined the selective cuts (-30 GeV $< TvQ < 10$ GeV) or ($\pi - 0.5 < \phi_{Zh} < \pi + 0.5$) and found them effective in separating the signal from the backgrounds. In exploiting more kinematical variables in some treatment like Boosted-decision-Tree technique (BDT) or Neural Networks (NN), those discriminative variables may be taken into consideration.

σ (fb)	$\ell^+\ell^- + jj$	$\ell^\pm + \not{E}_T + jj$	$\not{E}_T + jj$	combined
Vh signal	7.0×10^{-2}	4.1×10^{-1}	3.6×10^{-1}	
Vjj background	2.4×10^2	2.5×10^3	1.6×10^3	
$\mathcal{S}$	0.25	0.61	0.49	0.82
$\mathcal{S}_{\text{sys}}$	0.09	0.17	0.17	0.26

Table 7. Signal significance achieved from each channel and combined results for both statistics and systematics dominance.

5 Results and Discussion

5.1 Signal significance

As we see from the cut-flow tables 2-4, the Vjj backgrounds are dominant. We calculate the signal statistical significance as

$$\mathcal{S} = \frac{N_{\text{sig}}}{\sqrt{N_{\text{bkg}}}}, \tag{5.1}$$

with the statistical uncertainty of the dominant background as the only uncertainty. The combined significance of the $Vh(gg)$ signal is shown in Table 7. The three leptonic channels from the V decays give comparable contributions. The two-charged-lepton channel has the smallest signal strength, but cleaner in signal identification. The one and zero-charged-lepton channels show good reconstruction and contribute better sensitivities. Adding the $0, 1, 2$ charged-lepton channels, the pure statistical estimation gives a 0.82σ significance, which indicates how challenging an observation of the SM $Vh(gg)$ signal could be.

When the signal rate and S/B is small, one must worry about the systematic uncertainties for the measurements. As discussed in length in Sec. 3.4, we rely on the precision side-band fit to control the systematics in the signal region near $m_{jj} \sim m_h$. If ϵ_B is the fitted background percentage uncertainty, we then assume the systematic error to be $\epsilon_B \times N_{\text{bkg}}$. We thus present a different significance dominated by the systematics, defined as

$$\mathcal{S}_{\text{sys}} = \frac{N_{\text{sig}}}{\epsilon_B \times N_{\text{bkg}}}, \tag{5.2}$$

As shown in Sec. 3.4, with 3000 fb^{-1} of data and m_{jj} signal mass window taken as $95 - 150$ GeV, we have $\epsilon_B = 0.33\%,\ 0.10\%,\ 0.13\%$ for the two, one and zero lepton channels, respectively. The results with this significance estimation are also shown in Table 7. The outcome is worse than the statistical-error-only treatment. We would also hope the further reduction of non-statistic uncertainties with more dedicated background fitting schemes, once real data is available from experiments.

5.2 Bounds on the branching fractions and correlations with $h \to b\bar{b},\ c\bar{c}$

The interpretation of these results to bound on individual Higgs decay channels needs further discussion. Thus far, we have only simulated $h \to gg$ as the Higgs decay channel,

Table 8. Flavor tagging efficiency				Table 9. Fraction of SM decay channels			
ϵ_{ai}	b-quark	c-quark	$j = g, u, d, s$	e_{ai}	$h \to b\bar{b}$	$h \to c\bar{c}$	$h \to jj$
b-tag	70%	20%	1.25%	bb-tag	99.6%	0.4%	0%
c-tag	13%	19%	0.50%	cc-tag	90.4%	9.6%	0%
un-tag $j'j'$	17%	61%	98.25%	un-tag j'	16%	10%	74%

since it dominates the SM branching fraction of the Higgs decay to light jets. Practically, however, contributions from mis-tagged $h \to b\bar{b}$, $h \to c\bar{c}$, and possible light-quark pairs are all accumulated in the events and should be taken into account correlatively. Thus, the signal we have been searching for in this study really is $h \to j'j'$ where j' is an "un-tagged jet" including possible b, c and j (g, u, d, s) contributions.

Listed in Table 8 are the working points for the tagging/mis-tagging efficiencies assuming that different observable event categories listed as different rows are un-correlated. For instance, a b quark will be tagged as a b with a probability of $\epsilon_{bb} = 70\%$, and mis-tagged as a c and an un-tagged j' with $\epsilon_{cb} = 13\%$ and $\epsilon_{j'b} = 17\%$, and so on. Here the subscript a denotes the jet-tagged flavor category, and i denotes the parton as the source channel. The numbers are the same as in Category "c-tagging I" of Table 1 in Ref. [15], as reasonable estimates for the experimental performance at the 14 TeV LHC, and for consistency of later comparison. We extend to the double-tagged event categories with corresponding Higgs branching fraction channels as,

$$e_{ai} = \frac{\epsilon_{ai}^2 \times (\mathrm{BR})_i}{\sum_j \epsilon_{aj}^2 \times (\mathrm{BR})_j}. \tag{5.3}$$

We show in Table 9 the percentage contributions of these decay channels $h \to ii$ in each experimentally tagged category a. For instance, a pair of un-tagged jets in category $j'j'$ will have a probability of 74% from the SM Higgs decay to a pair of gluons, and 16% or 10% from $b\bar{b}$ or $c\bar{c}$, respectively. With the current tagging efficiency, we translate the significance 0.82σ on $\mathrm{BR}(h \to jj)$ to the un-tagged signal category $\mathrm{BR}(h \to j'j')$ by rescaling as

$$\mathcal{S}_{j'} = \frac{\mathcal{S}_j}{e_{j'j}} = \frac{0.82\sigma}{74\%} = 1.1\sigma, \tag{5.4}$$

that accounts for mis-tagged $b\bar{b}$, $c\bar{c}$ contributions as well. In other words, if an observation of $h \to j'j'$ were made in the future LHC run, the interpretation for individual channels would be based on Table 9, with updated tagging efficiencies.

As is customary, we define the signal strength for a decay channel $h \to ii$ as

$$\mu_i = \frac{\mathrm{BR}(h \to ii)}{\mathrm{BR}^{\mathrm{SM}}(h \to ii)}, \tag{5.5}$$

where we consider $ii = b\bar{b}$, $c\bar{c}$, and jj. Assuming each category is statistically independent and following Gaussian statistics. We combine the three categories to get the three dimensional

contour constraint on $\{\mu_b, \mu_c, \mu_j\}$ correlatively based on the relation

$$
\begin{aligned}
\mathcal{S}^2 > \sum_a \chi_a^2 &= \sum \frac{(x_a - \overline{x}_a)^2}{\sigma_a^2} \\
&= \sum_a \frac{(\sum_i \epsilon_{ai}^2 \mathrm{BR}_i N_{\mathrm{sig}}^{prod} - \sum_i \epsilon_{ai}^2 \mathrm{BR}_i^{\mathrm{SM}} N_{\mathrm{sig}}^{prod})^2}{(\sqrt{N_{\mathrm{bkg}}})^2} \\
&= \sum_a \frac{(\sum_i e_{ai} \, \mu_i - 1)^2}{(1/\mathcal{S}_a)^2}
\end{aligned}
\tag{5.6}
$$

where $\mathcal{S}_a$ is the significance from each category identified by experiments, and e_{ai} are the double efficiencies from each decay channel i in category a given in Table 9.[2] We take $\mathcal{S}_a = (11,\ 1.35,\ 1.1\ (0.35))$ for the three categories, assuming only statistical errors with $3000\ \mathrm{fb}^{-1}$ data. The first number is from Table 12 in the ATLAS MC study [8], making use of "One+Two-lepton" combined sensitivity. The second number comes from Fig. 2(a) of Ref. [15], the extrapolated study on the same MC dataset assuming the same tagging efficiency. Assuming most of the sensitivity on μ_c comes from the double c-tagged category, we likewise rescale the number with $e_{c'c}$ and a $\sqrt{2}$ since they consider $2 \times 3000\ \mathrm{fb}^{-1}$ data from two experiments. The third number is from our current "Zero+One+Two-lepton" un-tagged jets study, with the number in parenthesis including the systematic error. The fully correlated signal strengths are plotted in Fig. 9, for (a) a 3-dimensional contour in (μ_b, μ_c, μ_j) at 1σ, (b) the projected contour on the $\mu_j - \mu_c$ plane with statistical error only, and (c) with systematical error dominance. The shadowed contour regions are the projection of the 3D contour (μ_b, μ_c, μ_j) onto the μ_c-μ_j plane at 1σ and 2σ, and the solid ovals are for a fixed value $\mu_b = 1$. Allowing μ_b to float, the contour regions are slightly larger than the ovals. We note that certain values of the parameter space plane are excluded when $\mathrm{BR}(h \to bb) + \mathrm{BR}(h \to cc) + \mathrm{BR}(h \to jj) > 1$ and where our SM production assumption breaks down. This is represented in the plots by the gray shaded region. The 95% Confidence Level (CL) global upper bounds (approximately 2σ) on the branching fractions with statistical errors (systematic errors) for $3000\ \mathrm{fb}^{-1}$ with respect to the SM predictions can be obtained as

$$
\mathrm{BR}(h \to jj) \leq 4\ (9) \times \mathrm{BR}^{SM}(h \to gg),
\tag{5.7}
$$

$$
\mathrm{BR}(h \to c\bar{c}) < 15 \times \mathrm{BR}^{SM}(h \to c\bar{c}),
\tag{5.8}
$$

Although this bound on the $h \to gg$ channel is not nearly as strong as that from the production fit $gg \to h$ assuming the SM value, our study and results lay out the attempt of the search for the direct decay of the Higgs boson to gluons and the light quarks. The result for $c\bar{c}$ is comparable with the best existing extrapolations [15, 35], although adding the un-tagged category slightly improve the constraints on the c-quark Yukawa coupling, as expected.

Further improvements can be made by including the production of the vector boson fusion (VBF) [36] and $t\bar{t}h$ [37]. They are the sub-leading contributions to the $h \to jj$ study at Run

---

[2]Note the different efficiencies defined in Tables 8 and 9, with the normalizations $\sum_a \epsilon_{ai} = 1$ in categories, and $\sum_i e_{ai} = 1$ in channels.
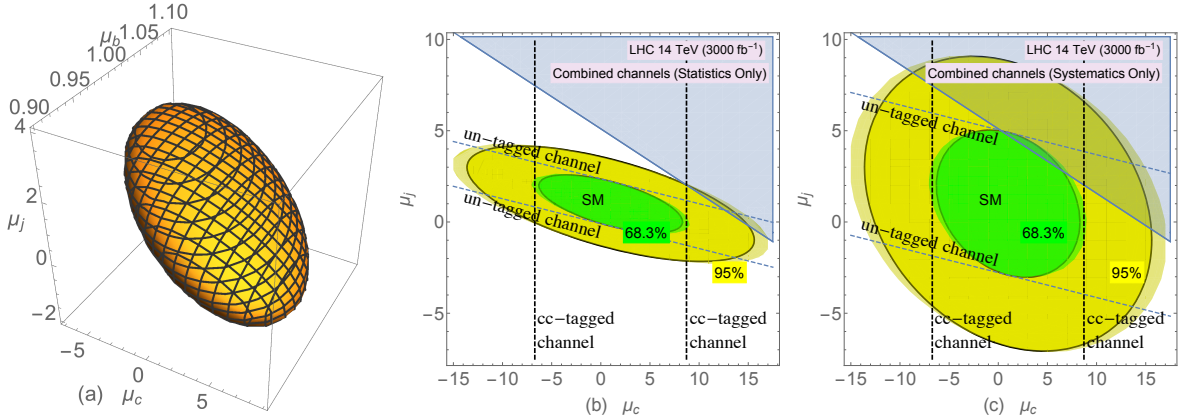
Figure 9. Signal strengths in correlated regions for (a) 1σ contour in 3-dimension (μ_b, μ_c, μ_j), (b) and (c) contours in μ_c-μ_j plane, for statistics only and including systematic uncertainties, respectively. The shadowed contour regions are the projection of the 3D contour (μ_b, μ_c, μ_j) onto the μ_c-μ_j plane at 1σ and 2σ, and the solid ovals are for fixing $\mu_b = 1$. The grey triangle area at the upper right corner is unphysical $\mathrm{BR}(h \to bb) + \mathrm{BR}(h \to cc) + \mathrm{BR}(h \to jj) > 1$.

I and become more important production channels at Run II [38]. Our study includes for simplicity only double-tagged categories, and single b or c tagged categories can be further included as done in the recast by Ref. [39]. Statistics can be further improved by analysis with likelihood fitting, BDT, etc. once data is available.

5.3   Bounds on light-quark Yukawa couplings

So far, possible contributions from light quarks (u, d, s) have been ignored in accordance with the SM expectation. The bound on $h \to jj$ in Eq. (5.7) can be translated into those for the light quark Yukawa couplings. Assuming the SM ggh coupling, and varying one light quark Yukawa y_q at a time, we translate our bound on μ_j to the Yukawa couplings for light quarks u, d, s by scaling the branching fraction with $\mu_q \propto y_q^2$. Our results of the bounds on the Yukawa couplings are shown in Table 10. There have been attempts to probe the light quark Yukawa couplings in the literature [35, 40–44]. Recent studies on the inclusive Higgs production and its spectra of $p_{T(h)}$ and y_h claim various improved constraints on the couplings [35, 41], compared to constraints from a global fit [45]. The upper bounds from our study of Higgs decay to light jets are comparable to those derived from the Higgs production kinematics, as also shown in Table 10, and thus provide complementary information to the existing approaches. The SM Yukawa coupling for the light quarks are taken to be proportional to their individual $\overline{\mathrm{MS}}$ running masses, which we evolve with N^{4}LO QCD from the PDG definition at 2 GeV [46] to the Higgs mass at 125 GeV.[3]

---

[3] There is however a discrepancy of the values used for the light quark running mass in the literature [35, 40, 43]. For instance, using the input from PDG, we find the strange quark mass to be $\overline{m}_s(m_h) = 53$ MeV, whereas the other adopted values are 83 MeV at the scale m_h [43], $\overline{m}_s(m_h/2) \sim 48$ MeV [40], and $\overline{m}_s(m_h/2) \sim 57$ MeV [35].

$\mathcal{L}(\mathrm{fb}^{-1})$	$\overline{\kappa}_u$ (κ_u)	$\overline{\kappa}_d$ (κ_d)	$\overline{\kappa}_s$ (κ_s)
300 (un-tagged $j'j'$)	1.2 (2600)	1.2 (1200)	1.2 (61)
3000 (un-tagged $j'j'$)	0.65 (1500)	0.65 (680)	0.65 (34)
Current Global Fits [45]	0.98 (2200)	0.97 (1000)	0.70 (37)
300 [41]	0.36 (820)	0.41 (430)	
3000 [35]			0.58 (30)

Table 10. Extrapolated upper bounds at 95% CL on the light-quark Yukawa couplings $\overline{\kappa}_q = y_q/y_b^{\mathrm{SM}}$ ($\kappa_q = y_q/y_q^{\mathrm{SM}}$) for $q = u, d, s$.

6 Summary and Conclusions

We have carried out a detailed study of the Higgs boson decay to light un-tagged jets in the vector boson associated channel $pp \to Vh$, with $h \to gg$ and $V = W^{\pm}$, Z decaying to leptons at the 14 TeV HL-LHC with 3000 fb^{-1}. To differentiate the di-jet signal from the huge SM QCD backgrounds, we have maximized the signal sensitivity by combining searches in the 0, 1 and 2-leptonic decay channels of the vector bosons. We used MadGraph, PYTHIA, and DELPHES for the signal and background simulations. Our findings can be summarized as follows.

- In Sections 3.1-3.3, we optimized the kinematical cuts according to the individual signal channels to enhance the $S/\sqrt{B}$ as well as S/B. The boosted kinematics for the di-jet signal has the advantage to improve S/B, while to keep the $S/\sqrt{B}$ roughly the same. We proposed the "di-jet-vicinity" Higgs mass reconstruction method as seen in Fig. 3, and tested its effectiveness against the pile-up effects as in Fig. 4.

- In Sec. 3.4, we studied in great detail on how to control the systematic errors by making use of the side-bands with a few fitting functions. We found that with 3000 fb^{-1}, it is conceivable to achieve the sub-percent level systematic uncertainties, as given in Eq. (3.5). It would be crucially important to take advantage of the large statistics and to keep the systematics under control.

- We may reach about 1σ combined significance for the un-tagged di-jet channel, as shown in Table 7 and in Eq. (5.4). We also considered the correlation with mis-tagged events from $h \to b\bar{b}$, $c\bar{c}$ channels, as discussed in Sec. 5.2

- Assuming the SM Vh production, our results can be translated to upper bounds on the branching fractions of 4 and 15 times the SM values for BR($h \to gg$) and BR($h \to c\bar{c}$), respectively, at 95% CL, seen in Eqs. (5.7) and (5.8).

- Exploiting our results, indirect upper bounds on light-quark Yukawa couplings can be extracted, as summarized in Table 10, and compared with the currently existing literature.

- We pointed out that there are other variables to explore. Kinematic discriminants like TvQ and ϕ_{Zh} as discussed in Sec. 4 may be among them. In the hope to improve the simple cut-based analyses, multiple variable methods like BDT and NN would be promising. Addition of other production channels such as VBF and $t\bar{t}h$ will also help to strengthen the bounds.

After the Higgs boson discovery and initial measurements for the SM-like properties at the LHC Run I and Run II, it is imperative at the HL-LHC to tackle the more challenging channels with the rare Higgs decays. Our studies on the Higgs decay to the light un-tagged jets would hopefully serve as an initial proposal among the future efforts.

Acknowledgments

References

[1] **ATLAS** Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1–29, [arXiv:1207.7214].

[2] **CMS** Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30–61, [arXiv:1207.7235].

[3] **ATLAS** Collaboration, G. Aad et al., *Observation and measurement of Higgs boson decays to WW* with the ATLAS detector*, Phys. Rev. **D92** (2015), no. 1 012006, [arXiv:1412.2641].

[4] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states*, JHEP **01** (2014) 096, [arXiv:1312.1129].

[5] **ATLAS, CMS** Collaboration, G. Aad et al., *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, JHEP **08** (2016) 045, [arXiv:1606.02266].

[6] **CMS** Collaboration, V. Khachatryan et al., *Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method*, Eur. Phys. J. **C75** (2015), no. 6 251, [arXiv:1502.02485].

[7] **ATLAS** Collaboration, G. Aad et al., *Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **05** (2016) 160, [arXiv:1604.03812].

[8] **ATLAS** Collaboration, *A study of standard model higgs boson production in the decay mode h bb in association with a w or z boson for high luminosity lhc running*, tech. rep., ATL-PHYS-PUB-2014-011, July, 2014.

[9] **ATLAS** Collaboration, *Projections for measurements of Higgs boson signal strengths and coupling parameters with the ATLAS detector at a HL-LHC*, Tech. Rep. ATL-PHYS-PUB-2014-016, CERN, Geneva, Oct, 2014.

[10] T. Han and B. McElrath, *h to mu+ mu- via gluon fusion at the LHC*, *Phys. Lett.* **B528** (2002) 81–85, [`hep-ph/0201023`].

[11] D. Asner, T. Barklow, C. Calancha, K. Fujii, N. Graf, H. E. Haber, A. Ishikawa, S. Kanemura, S. Kawada, M. Kurata, A. Miyamoto, H. Neal, H. Ono, C. Potter, J. Strube, T. Suehara, T. Tanabe, J. Tian, K. Tsumura, S. Watanuki, G. Weiglein, K. Yagyu, and H. Yokoya, *Ilc higgs white paper*, [`arXiv:1310.0763`].

[12] **FCC-ee study** Collaboration, M. Koratzinos, *FCC-ee accelerator parameters, performance and limitations*, *Nucl. Part. Phys. Proc.* **273-275** (2016) 2326–2328, [`arXiv:1411.2819`].

[13] **CEPC-SPPC Study Group** Collaboration, *CEPC-SPPC Preliminary Conceptual Design Report. 1. Physics and Detector, http://inspirehep.net/record/1395734,* .

[14] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *AIP Conf. Proc.* **1078** (2009) 189–191, [`arXiv:0809.2530`].

[15] G. Perez, Y. Soreq, E. Stamou, and K. Tobioka, *Prospects for measuring the Higgs coupling to light quarks*, [`arXiv:1505.06689`].

[16] G. T. Bodwin, F. Petriello, S. Stoynev, and M. Velasco, *Higgs boson decays to quarkonia and the Hc̄c coupling*, *Phys. Rev.* **D88** (2013), no. 5 053003, [`arXiv:1306.5770`].

[17] **ATLAS** Collaboration, G. Aad et al., *Search for Higgs and Z Boson Decays to J/ and (nS) with the ATLAS Detector*, *Phys. Rev. Lett.* **114** (2015), no. 12 121801, [`arXiv:1501.03276`].

[18] **ATLAS** Collaboration, *Search for the Standard Model Higgs and Z Boson decays to $J/\psi\,\gamma$: HL-LHC projections*, Tech. Rep. ATL-PHYS-PUB-2015-043, CERN, Geneva, Sep, 2015.

[19] I. Brivio, F. Goertz, and G. Isidori, *Probing the Charm Quark Yukawa Coupling in Higgs+Charm Production*, *Phys. Rev. Lett.* **115** (2015), no. 21 211801, [`arXiv:1507.02916`].

[20] **LHC Higgs Cross Section Working Group** Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [`arXiv:1610.07922`].

[21] **NNPDF** Collaboration, R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo, *Parton distributions with QED corrections*, *Nucl. Phys.* **B877** (2013) 290–320, [`arXiv:1308.0598`].

[22] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [`arXiv:1405.0301`].

[23] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, *JHEP* **03** (2013) 015, [`arXiv:1212.3460`].

[24] C. Englert, M. McCullough, and M. Spannowsky, *Gluon-initiated associated production boosts higgs physics*, `arXiv:1310.4828`.

[25] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [`hep-ph/0603175`].

[26] **DELPHES 3** Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057, [`arXiv:1307.6346`].

[27] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, *Phys. Rev.* **D80** (2009) 051501, [`arXiv:0903.5081`].

[28] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084, [`arXiv:0912.1342`].

[29] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015, [`arXiv:1011.2268`].

[30] S. Sapeta, *QCD and Jets at Hadron Colliders*, *Prog. Part. Nucl. Phys.* **89** (2016) 1–55, [`arXiv:1511.09336`].

[31] J. M. Butterworth, I. Ochoa, and T. Scanlon, *Boosted Higgs $\to b\bar{b}$ in vector-boson associated production at 14 TeV*, *Eur. Phys. J.* **C75** (2015), no. 8 366, [`arXiv:1506.04973`].

[32] A. Banfi and J. Cancino, *Implications of qcd radiative corrections on high-pt higgs searches*, `arXiv:1207.0674`.

[33] **ATLAS** Collaboration, *Expected pileup values at the HL-LHC*, Tech. Rep. ATL-UPGRADE-PUB-2013-014, CERN, Geneva, Sep, 2013.

[34] C. Shimmin and D. Whiteson, *Boosting low-mass hadronic resonances*, *Phys. Rev.* **D94** (2016), no. 5 055001, [`arXiv:1602.07727`].

[35] F. Bishara, U. Haisch, P. F. Monni, and E. Re, *Constraining light-quark yukawa couplings from higgs distributions*, `arXiv:1606.09253`.

[36] **CMS** Collaboration, *Search for the standard model higgs boson produced through vector boson fusion and decaying to b bbar*, `arXiv:1506.01010`.

[37] **ATLAS** Collaboration, *Search for the standard model higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s}$=8 tev with the atlas detector*, `arXiv:1604.03812`.

[38] **CMS** Collaboration, *VBF H to bb using the 2015 data sample*, .

[39] G. Perez, Y. Soreq, E. Stamou, and K. Tobioka, *Constraining the charm yukawa and higgs-quark coupling universality*, `arXiv:1503.00290`.

[40] Y. Zhou, *Constraining the Higgs boson coupling to light quarks in the HZZ final states*, *Phys. Rev.* **D93** (2016), no. 1 013019, [`arXiv:1505.06369`].

[41] Y. Soreq, H. X. Zhu, and J. Zupan, *Light quark yukawa couplings from higgs kinematics*, `arXiv:1606.09621`.

[42] G. Bonner and H. E. Logan, *Constraining the Higgs couplings to up and down quarks using production kinematics at the CERN Large Hadron Collider*, `arXiv:1608.04376`.

[43] Y. Meng, Z. Surujon, A. Rajaraman, and T. M. P. Tait, *Strange Couplings to the Higgs*, *JHEP* **02** (2013) 138, [`arXiv:1210.3373`].

[44] F. Yu, *Phenomenology of Enhanced Light Quark Yukawa Couplings and the $W^{\pm}h$ Charge Asymmetry*, `arXiv:1609.06592`.

[45] A. L. Kagan, G. Perez, F. Petriello, Y. Soreq, S. Stoynev, and J. Zupan, *Exclusive Window onto Higgs Yukawa Couplings*, *Phys. Rev. Lett.* **114** (2015), no. 10 101802, [`arXiv:1406.1722`].

[46] **Particle Data Group** Collaboration, K. A. Olive et al., *Review of Particle Physics*, *Chin. Phys.* **C38** (2014) 090001.