Boosting Invisible Higgs Search with a Gluon Jet

Won Sang Cho,† Hyung Do Kim,† and Dongsub Lee†

1 Department of Physics and Astronomy and Center for Theoretical Physics, Seoul National University, Seoul 08826, Korea

We propose that tagging the jets emitted by initial state radiation (ISR) can help search invisible Higgs decays. Quark and gluon composition of ISR jets of the signal process is different from that of the background and we can enhance the signal-to-background ratio using flavor information of the ISR. By using jet substructure observables as input features, neural network-based event classifier is found to be very powerful for boosting the discrimination of gluon jet rich signal events from the other quark jet rich backgrounds. On top of the conventional missing transverse energy (MET) in mono-jet analysis, we show that Higgs produced from the gluon fusion constrains the invisible Higgs decay the most, over the vector boson fusion traditionally known as the most constraining, and the limit on the branching ratio of the Higgs invisible decay is significantly improved. We summarize with emphasizing that our method has wider implications in search for new resonances.

Introduction

Higgs discovery at the LHC in 2012 completed the Standard Model as a description of nature in terms of elementary particles and their interactions [1, 2]. At the same time, more than 95% of the universe is composed of dark matter and dark energy of which the properties are not known yet. Higgs boson is hard to produce at the colliders as it has the strongest interactions with the heaviest particle which decays instantaneously while it has the weakest coupling to the lightest particle which is stable. Therefore, if new sectors are connected to the visible section only through Higgs boson, it would be difficult to probe it. The possibility of having dark matter coupling to the Higgs boson is tantalising as dark matter can be probed at the colliders through Higgs production [3, 4].

While the LHC was mainly proposed as a discovery machine for the Higgs boson and new particles, the absence of new particles for a decade of running made the LHC as a precision machine for the Standard Model (SM) which can be compared with perturbative computation of QCD, and the precision measurement of Higgs couplings is also considered as one of the most important tasks new physics can be probed. As we measure the product of the Higgs production cross section and the branching ratio, the ultimate precision of the invisible branching ratio is tied up with the precise determination of Higgs coupling to the Standard Model gauge bosons and/or fermions. Thus the observation of nonzero branching ratio would be an indication of new physics, either modification of Higgs couplings compared to the Standard Model or the decay to the invisible particles. In this perspective the Higgs production in (weak) vector boson fusion (VBF) has been expected as the best channel at the LHC [7], and surveyed extensively [8–15].

In this letter, assuming the Higgs production cross section of the SM, we show that the limit on the branching ratio of invisible Higgs decay can be improved significantly if it is combined with ISR jet tagging for the most dominant Higgs production channel - gluon fusion Higgs (ggH) which has been considered not so useful so far. The rest of our paper consists of the description of 1) the dynamics of ggH production with high-pT jets at leading order, which leads gluon rich ISR jets, compared with the ones of V+jets processes, 2) current status of the search for invisible Higgs decay and the event generation and selection scheme for a mono-jet analysis in search for the invisible Higgs decays in ggH+jet channel, and 3) machine learning for optimal discrimination of the gluon-jet rich signal over the quark-jet rich background using jet substructures carrying gluon/quark flavor information. In the last stage, we demonstrate that a significant improvement can be achieved on the limit of invisible Higgs decays at the LHC, by using the machine learning optimal variables for profile likelihood ratio test. Conclusion is devoted to discussion of the results and future implications.

Physics

In search for the invisible Higgs decays, our main signal process is the ggH+jet production with H decaying invisibly, and the main background is estimated to be V(Z(ℓν)/W(ℓν))+jets. Table I shows the leading diagrams of the ggH and V+jet processes with an emission of ISR(s), for three different initial parton configurations (gg, gg, gq). The gluon lines in grey also represent extra emissions, and the V+jet(gg) diagram (bottom-center) from gg initial states is drawn in grey as it is subleading to the other 5 diagrams in αs.
For the V+jets process, the leading ISR jet is mostly quark jet. Even though there is non-negligible gluon jet contributions from 1) the emissions in subleading diagrams to the leading V+jet(gg,gq), and 2) the emission from the V+jet(qg) at leading order with suppression by small parton luminosity. However, the leading ISR jet is expected most likely to be a quark jet mainly because of the hierarchical parton luminosities between gg + gg and qg initial states, and the relative quark-vs-gluon jet ratio of a leading ISR jet is estimated as $\sigma_{Vq}/\sigma_{Vg} \approx 4$ for the V+jet process, as shown in Fig. 2 (red).

For the ggH process, the leading ISR jet is more gluon like. There exist two enhancement factors of the ggH process over the leading quark ISR process, ggH(gg). In the diagrams of the ggH(gg) in Table I, gluon is emitted as the leading ISR from the gluon parton or from the top loop, and all of the results end up with the enhancement of QCD color factor ratio, $C_A/C_F$ in the cross section compared to the ggH(qg) where $C_A$ ($C_F$) is the quadratic Casimir of the adjoint (fundamental) representation. Furthermore, compared to the quark ISR from the ggH(qg), the gluon ISR from the ggH(gg) is more likely to be emitted in central region with small rapidity as in Fig. 1(d) due to the balanced momentum profile of the gg initial states. Although the parton luminosity function in Fig. 1 is larger for the gg state compared to the gg (by a factor of 2 at $\sqrt{s} \sim 100$ GeV), all in the leading ISR in the ggH becomes more gluon jet like, and the relative gluon-vs-quark jet ratio of a leading ISR jet is estimated to be $\sigma_{Hg}/\sigma_{Hq} \approx 1.8$ as in Fig. 2 (blue) especially for the conventional event selection cuts requiring a central leading jet in $|\eta^j| < 2.4 \sim 2.5$.

Currently, the best search channels for the invisible Higgs decays are the VBF and Higgsstrahlungs (VH) [11, 12, 14], as they contain additional characteristic features associated with the Higgs. In spite of its dominant production rate much larger than the sub-dominant VBF by a factor of 7 at $\sqrt{s} = 13$ TeV, the ggH production with invisible Higgs decays constrains much milder in lack of such an additional feature to be identified.

However, like as the kinematics of the associated two quark jets in forward region from the VBF, now the ggH process also has such a unique property accompanied - the gluon-like ISR jets in central region originated from the gluon fusion dynamics. In this regard, if some relevant variables containing the flavor information of ISR jets are employed, one can obtain the most stringent constraints from the ggH channel for the invisible Higgs decays, as is demonstrated in the next sections.

**Analysis** In search for the invisible Higgs decays in $E_T^{miss}+jets$ signature via the ggH channel, samples are generated by Monte Carlo simulated pp collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV at the LHC, for 36 fb$^{-1}$ using MadGraph5 aMC@NLO v2.6.2 [16] interfaced with Pythia v8.235 [17] for hadronisation and fragmentation. Delphes v.3.4.1 is used for detector simulation [18]. The signal process (ggH+jets) is generated at LO taking into account finite top mass effects [19] under the SM Higgs hypothesis with $M_H = 125$ GeV, and backgrounds are generated at NLO in QCD. We use FxFx scheme with $k_T$-algorithm and $\Delta R = 1$ for jet merging [20]. For jet clustering, FastJet v3.2.1 [21] is used with anti-$k_T$ algorithm with $\Delta R = 0.4$, and CT10NLO [22] is used for parton distribution function.

Among the relevant background processes - $W(Z(\nu\bar{\nu}), W(l\bar{\nu})+jets$, Diboson, top quarks, $Z/\gamma \rightarrow \ell\ell$, QCD multijets, where the leptons ($\ell$) in $W/Z/\gamma$ decays are mis-
identified, we only included the most dominant background contribution - V+jets (Z+jets:W+jets ~ 1 : 0.5) while the others take just O(1)% level for the event selection criteria as follows [12]:

- \( p_T^{4j} > 100 \text{GeV}, \mid \eta^{4j} \mid < 2.5, E_T^{miss} > 200 \text{GeV}, \min_{j \in \text{jets}} \Delta \phi(p_T^{miss}, p_T^{j}) \geq 0.5, N_{\text{jet}} \geq 1. \)

The 1st (2nd) cut on the transverse momentum (pseudorapidity) of the leading jet is imposed to suppress all of the backgrounds, the 3rd cut on the missing transverse energy is mainly to reduce the QCD and top quarks, and the 4th cut with the missing transverse momentum, \( p_T^{miss} \) suppresses the QCD multijets very efficiently [13].

There also exist contributions from other Higgs productions, VBF and VH with yield rates (ggH:VBF:VH ~ 70:20:10%). However as the leading jets from VBF are most likely quark jets opposed to the gluonic leading jets in the ggH, we checked that the VBF can be easily separated from the ggH by using new flavor information in addition to the traditional forward jet tagging. As for the VH which also has quark jet like leading ISRs according to the same argument with the V+jets, it can have additional selection criteria [23] for identifying jets from hadronically decaying vector bosons. In this regard, to demonstrate the main idea without making event selection scheme too complicated, we simply consider the ggH as the only signal versus the V+jets as the main background in this analysis, without loss of consistency in applying the flavor information for discrimination of gluon-jet rich ggH signal from general quark-jet rich backgrounds.

**Method** We use a set of jet substructure variables [24], say \( S_{\text{jet}} \), in our analysis as the following,

- \( S_{\text{jet}} = \{ n_{\text{tk}} \text{ (track multiplicity)} [26], \text{Girth} [26, 27], \text{Broadening} [28], \text{EEC} \text{ (energy-energy correlation)} [29] \text{ with } \beta = 0.2 [30], \text{RMS-}p_T [29], \} \),

which contain the information on jet flavors. It can also be extended to include more raw data, e.g. jet images [31] [32] for deep learning. Among the five jet substructure variables used, Girth reflects a fatness/radius of a jet, as gluon jets tend to have more showers and be fatter than quark jets by the color factor enhancement, \( C_A(q \to gg)/C_F(q \to gg) \). Such a property can be checked in the Girth distribution of the leading jet from ggH and V+jet processes in Fig 3.

Jet substructure observables have been used to build a jet tagger, \( P_{q/g}(S_{\text{jet}}) \), while the kinematic observables, such as reconstructed four-momenta of jets have been used to build an event classifier, \( P_{S/B}(\{p_T^{\text{jet}...}\}) \). However, as can be seen from \( d^2\sigma/dp_T^{\text{jet}}dy^{\text{jet}} \) in Fig 1b, the flavor of a jet can have a correlation with kinematic information depending on the scattering process. This observation motivates us to build \( P_{S/B}(\{p_T^{\text{jet}...}\} \cup S_{\text{jet}}) \), rather than a factorized classifier, \( P_{S/B}(\{p_T^{\text{jet}...}\}) \cup P_{q/g}(S_{\text{jet}}) \).

For \( S = \{ \text{ggH (V+jets)} \} \) process, Fig 3 shows the normalized distribution of (a) \( E_T^{miss} \), (b) Girth, (c) \( P_{S/B}(S_{\text{jet}}) \), and (d) \( P_{S/B}(S_{\text{jet}} \cup \{E_T^{miss}, p_T^{\text{jet}}, \eta^{\text{jet}}\}) \). The two event classifiers \( P_{S/B} \) in Fig 3c and 3d are obtained by training neural networks with 2-3 layers each with 200-300 nodes with the specified input features. We used one million event samples with Keras [33] for building and training the neural network models. It is noticeable that the event classifier using the set of jet substructures alone can provide much better separation of signal and background compared to the one \( E_T^{miss} \) as in Fig 3a. Combining them all we get the best separation as is clearly seen in Fig 3d.

**Result** The result obtained up to now can be used to discover invisible Higgs decay or put constraints on the invisible branching ratio of Higgs. After selecting the events with the criteria, we performed the profile likelihood ratio test following the procedure in [34] with the four template distributions in Fig.3.

The likelihood function is given,

\[
\mathcal{L} = \prod_{i=1}^{N_{\text{bin}}} \frac{n_{i}}{n_{i}} e^{-n_{i}} \times \frac{1}{2\pi} e^{-\frac{1}{2}(\theta_1^2 + \theta_2^2)},
\]

where \( n_i \) is the number of events (or pseudo events) in
The $i$-th bin, and $\hat{n}_i$ is the number of expected events with branching ratio parameter $\mu = \frac{\sigma}{\sigma_{SM}} \times BR(h \rightarrow inv)$, i.e.,

$$\hat{n}_i = \mu N_i P_s(i)(1 + f_s)^{\theta_s} + N_b P_b(i)(1 + f_b)^{\theta_b}.$$  \hspace{1cm} (2)

Here the $P_s(b)(i)$ is the expected event rate in $i$-th bin, given the total number of events $N_s(b)$ survived the cut, and the $\theta_s(b)$ in the Poisson and prior probabilities denotes a nuisance parameter associated to the systematic uncertainty $f_{s(b)}$ of signal (background). We take the systematic uncertainties to be 10% and varied in 5-20% range as a global variation of event rates in signal and background distributions. Signal cross section is taken from [35] and we derive the acceptance on the criteria for the simulated events. For the background process we take both cross section and acceptance from the MC simulator, and introduce a K-factor to cover the other sources of background, and to reproduce the expected limit using missing transverse energy [14].

For the four profile likelihood ratio tests with/without the new features of jet flavors, now we obtain the upper limits in 95% of confidence level on the branching ratio of invisible Higgs decays, for the integrated luminosity 36 fb$^{-1}$ at the LHC, as in Fig.4 (right panel), and compare the result with the existing experimental result (left panel) [14]. From the left in the right panel of Fig.4 the limits on the Higgs invisible branching ratio are drawn for the features of (traditional) missing transverse energy, Girth, all jet substructure variables, and the combination of missing transverse energy and all jet substructure variables. From the left in the left panel of Fig.4 three columns are the experimental results which did not use jet flavor information for the channels, 1) gluon fusion, 2) vector boson fusion, and 3) combined, from the right to the left. The result of the rightmost (in the left panel) can be compared and matched with the result of the leftmost (in the right panel) as they are using the same features under the same selection criteria in the same production channel. The result shows that the limit of Higgs invisible decays can be significantly improved down to 5% if jet flavor information from the gluon fusion is employed, which can be compared to 60% which used only MET.

It is also noticeable that the jet substructures alone provides much stronger constraints than the missing transverse energy in Higgs invisible search. Moreover, combining the features in two kinds, we end up with the result much better than the one obtained from vector boson fusion, down to 5% on the limit, only from gluon fusion. Projected to future LHC, TABLE II also shows the medians of the expected limit can be as small as 2% for 300 fb$^{-1}$ or even smaller (less than 1%) for 3 ab$^{-1}$. It is also expected the limit can be significantly improved again if it is combined with the results from vector boson fusion and other processes.

### TABLE II: Expected upper limit on $\frac{\sigma \times BR(H \rightarrow inv)}{\sigma_{SM}}$ in three integrated luminosity points. Central values are derived with 10% systematic uncertainties on both signal and background. We vary both systematic uncertainties by $f_{s,b}=5\%$ and $f_{s,b}=20\%$.

| $L$   | $E_{miss}^{\text{median}}$ | $G_{j_1}$ | DNN(jet sub.) | DNN(all) |
|-------|-----------------------------|-----------|---------------|----------|
| 36fb$^{-1}$ | 60.2$^{+1.5\%}_{-1.3\%}$ | 20.4$^{+0.3\%}_{-1.0\%}$ | 8.3$^{+0.7\%}_{-0.4\%}$ | 5.2$^{+0.1\%}_{-0.2\%}$ |
| 300fb$^{-1}$ | 20.9$^{+0.3\%}_{-1.0\%}$ | 7.1$^{+0.1\%}_{-0.3\%}$ | 2.9$^{+0.0\%}_{-0.1\%}$ | 1.8$^{+0.0\%}_{-0.1\%}$ |
| 3ab$^{-1}$   | 6.6$^{+0.1\%}_{-0.3\%}$  | 2.2$^{+0.0\%}_{-0.1\%}$ | 0.9$^{+0.0\%}_{-0.0\%}$ | 0.6$^{+0.0\%}_{-0.0\%}$ |

FIG. 4: Upper limit in 95% of confidence level(C.L.) on $\frac{\sigma \times BR(H \rightarrow inv)}{\sigma_{SM}}$ with the integrated luminosity 36fb$^{-1}$. Here we used 5,000 ensembles of pseudo data set which consists of background events only. Result on the left panel shows the result from the experiment [14].
Though more dedicated analysis is necessary to obtain a
firm number for the expected limit on the Higgs invisible
decays with sophisticated treatment of systematic
uncertainties in real data analysis, the exercise we did in
this Letter strongly suggests that 1% precision for Higgs
invisible branching ratio at the end of the LHC running
with 3 ab$^{-1}$ is a plausible expectation from the gluon
fusion solely.

**Conclusion**

The flavor information of ISR jets can be used to enhance the signal when the gluon and quark composition of ISR jet is different for signal and background. Focusing on the gluon-jet rich leading ISR in central region from gluon fusion Higgs production versus the quark-jet rich background processes, we apply this idea to the invisible Higgs search and show that the invisible Higgs from gluon fusion which suffers from huge irreducible $V$-$j$-jets backgrounds, can be the best channel to constrain the invisible Higgs branching ratio by exceeding the limit given by the other channels, vector boson fusion or vector-associated. The physics and methods in this analysis can also be applied to a broad range of new resonance productions induced by gluon fusion, and decays to non-QCD particles, in search for di-Higgs/toponium/exotics with lepton/photons/MET final states, concurrently with their irreducible background processes mostly containing quark jet dominant ISR.

This work was supported by the National Research Foundation of Korea (NRF), Grants No. 2017R1A2B1010749. WC and DL were also supported by the NRF funded by the Ministry of Science and ICT, Grant No. NRF-2017R1C1B2011048. Parts of this study have been presented at ICHEP 2018 in Seoul [3].