Boosted $H \to \tau\tau$ and the top Yukawa coupling

Gabriel Lynch[a], Jordan Webster[b]

[a]*Department of Physics, 5720 S. Ellis Ave, Chicago, IL 60637, USA.*
[b]*HEP Division, Argonne National Laboratory, 9700 S.Cass Avenue, Argonne, IL 60439, USA.*

Abstract

Since the discovery of the Higgs boson, there has been much work done to obtain precise measurements of its couplings. Results are so far consistent with the Standard Model. However, boosted Higgs decays have yet to be fully explored. Such decays provide one method to study possible beyond the Standard Model modifications to the Higgs couplings that are hidden at low momentum. A preliminary detector level analysis is done of boosted Higgs events against a $Z+$jets background targeting the $H \to \tau\tau$ leptonic decay. Projections are done to estimate the integrated luminosity needed to observe the Higgs in this channel, and to exclude one benchmark model of modified top-Yukawa couplings. Results indicate that searches for new physics in this channel become interesting once the LHC has collected 1100 fb^{-1} of data.

Email addresses: `gplynch@uchicago.edu` (Gabriel Lynch), `jwebster@anl.gov` (Jordan Webster)

1. Introduction

The discovery of the Higgs boson in 2012 [1] has made it possible to measure how closely its properties match predictions made by the Standard Model (SM). One of the properties of interest is the top Yukawa coupling, an indicator of how strongly the Higgs interacts with the top quark. A precise measurement of this coupling constant could provide an avenue for exploring effects from new physics.

The $t\bar{t}H$ channel provides one approach for making a measurement. However, [2, 3] offers the $H \to \tau\tau$ as an appealing alternative to $t\bar{t}H$ for both observing Higgs production and testing for possible new physics models involving gluon fusion Higgs production. These new models, detailed in [2], involve modifying the SM Higgs coupling with the top loop with a new, non-SM coupling coming from some new physics. These modifications can be expressed through coefficients c_t and κ_g, for the SM and BSM contributions, respectively. The coupling coefficient $(\kappa_g, c_t) = (0, 1)$ corresponds to an SM top loop coupling. A model with $(\kappa_g, c_t) = (0.5, 0.5)$ has equal contributions to gluon fusion Higgs production from a top quark loop and new physics. The overall Higgs production rate is identical to that in the SM, but the boosted momentum of the Higgs boson takes on a slightly different shape that can be observed with enough data.

In this analysis, we aim to improve upon the conclusions of [2] by incorporating a detector level simulation, as well as introducing more involved statistical analysis in order to make better estimates of sensitivity in this channel.

2. Event Generation

In this analysis we target the $H \to \tau\tau$ decay. We have limited this analysis to the leading signal and background processes that meet this condition. We limit our focus to the dominant $Z+$ jets background. The background process $t\bar{t}+$jets is also non-negligible in this channel, but is ignored for this preliminary analysis. Its inclusion is left as a future project.

All events are generated using `MadGraph5_aMC@NLO` [4]using the included HEFT model and showered using `Pythia6` [5]. The only generator level event selections applied in addition to the default run card were to require $p_T > 10$ GeV for both the leptons and the jets in the final state of the event. The frugality of event selection at this stage helps to ensure that the event generation is not biased towards events that will pass our selections. Although 40 million $Z+$ jets events were generated as a background, jet-parton matching done by `Pythia6` effectively reduce this number by a factor of 2. This process removes events duplicates caused by overlap between the event generation and the parton shower. Table 1 summarizes the event generation parameters used in this analysis.

Process	**No. Events**
`p p > Z j, (z > ta+ ta-)`	$2 \cdot 10^7$
`p p > H j, (H > ta+ ta-)`	$5 \cdot 10^7$

Table 1: Number of events generated with each MadGraph5 command.

We are interested in events in which we have both a high p_T jet and two leptons in the final state. The dileptonic final state is an important feature of these events, as it provides for a cleaner collinear mass reconstruction of both the Higgs and the Z boson. `Delphes3` [6] is configured with the existing ATLAS card distributed with `MadGraph5_aMC@NLO` and is used as a fast detector simulation. A detector level simulation accounts for particles lost in the detector, fake jets and leptons, and the detector resolution, all of which can affect the number of events passing the event selection. Including this stage of simulation allows for a more realistic estimation of integrated luminosity required for a sensitive result. The center of mass energy for this analysis is set to $\sqrt{s} = 14$ TeV to match the design energy of the LHC.

3. Event Selection

The selections applied in this analysis are summarized below; a more complete description of the of the selections can be found in the appendix. The preliminary cuts are borrowed from Section IV of [2] with minor adjustments to better suite a detector level analysis. A comparison of Table III from [2] and Table 2 in this paper highlights these adjustments.

When a particle escapes the detector without being observed, the measured energy in the transverse plane becomes unbalanced. The negative vector sum of the transverse energy is referred to as the missing transverse energy, denoted by $\not{p}_T$. In an event in which we see a highly boosted Higgs, we can also expect the resultant leptons to be highly boosted. In such a scenario, this missing energy can be projected onto the lepton four-vectors

Selection	$H \to \tau\tau$	$Z + \text{jets}$
Initial cross-section	1077	178505
$n_l = 2$	25.71	246.5
$m_{ll} > 15$ GeV	25.63	242.8
$p_{T,H}^{rec} > 200$ GeV	20.33	236.1
$n_j = 1$ with $p_T > 200$ GeV	0.6429	63.75
$n_b = 0$	0.5949	59.03
$\not{p}_T$ inside leptons	0.5681	57.48
$m_{ll} < 70$ GeV	0.5074	48.53

Table 2: Selection efficiencies times cross section for each selection. Cross sections are in fb. The Z bosons decay to $\tau^+\tau^-$.

for a good approximation of the invisible neutrino momentum vectors. The selection "$\not{p}_T$ inside leptons" says that we require the $\not{p}_T$ four-vector to be between the two lepton four-vectors. This guarantees that the projection can take place. This cut also significantly reduces the contribution from $H \to WW^*$ leptonic decays, as these events tend to see the leptons and $\not{p}_T$ back to back. This criterion is important as the neutrino four-vectors can be vectorially summed with the lepton four-vectors to give an approximate Higgs four-vector.

These selections were applied using the `ROOT` data analysis framework, starting with the output from `Delphes3`. Each selection was applied successively in the order shown in Table 2, with the cut flow recorded for each selection. After the final $m_{ll} < 70$ GeV selection, $S/\sqrt{B}$ for 300 fb^{-1} is calculated as 1.262.

4. SM Higgs Analysis

In order to make a measurement of the Standard Model Higgs mass, it is necessary to reconstruct its four-momentum. As explained above, this analysis makes use of a collinear approximation, where the neutrinos produced from tau decay are assumed to be collinear with the leptons. By requiring $\not{p}_T$ to be inside the two leptons, we are able to use the lepton momentum vectors as a basis onto which we project $\not{p}_T$:

$$\not{p}_T = p_{T,\nu_1} + p_{T,\nu_2}, \qquad p_{\nu_i} = a_i \cdot p_{\ell_i} \tag{1}$$

This projection gives us an approximate neutrino momentum for each neutrino produced along with the leptons. Summing these vectors gives us the collinear reconstruction of the Higgs four-vector, the mass of which provides a good approximation for the Higgs mass. In the $Z+$ jets background, this collinear mass reconstruction approximately represents the reconstructed mass of the Z boson.

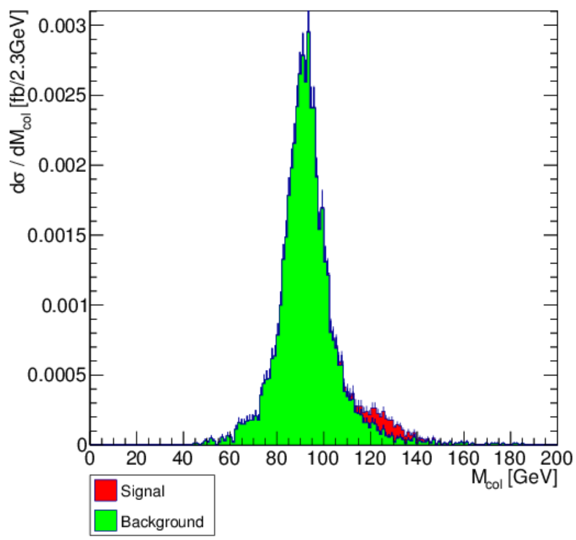

Figure 1: Collinear mass distribution normalized to cross section. The signal distribution has been scaled by a factor of 5.

We perform a negative log likelihood test in order to search for the Higgs signal on top of the $Z+$ jets background. Our null hypothesis is a background only collinear mass shape, whereas our H1 hypothesis is the background plus signal shape. An "Asimov dataset" [7], or a data set that replicates the shape of the collinear mass distribution for the H1 hypothesis, is constructed. The data is fitted following the procedure outlined in [7, 8] and and a signal strength parameter is measured using the asymptotic approximation. This method of fitting provides us with a convenient way to estimate both the uncertainty and sensitivity of the measurement.

4.1. Results

The signal and background collinear mass distributions are shown in Figure 1. A peak due to the background events is visible at around 90 GeV, close to the mass of the Z boson. We see another peak near 125 GeV, the mass of the Standard Model Higgs. The position of these peaks are more evident in Figure 2, where the distributions have been normalised to unity.

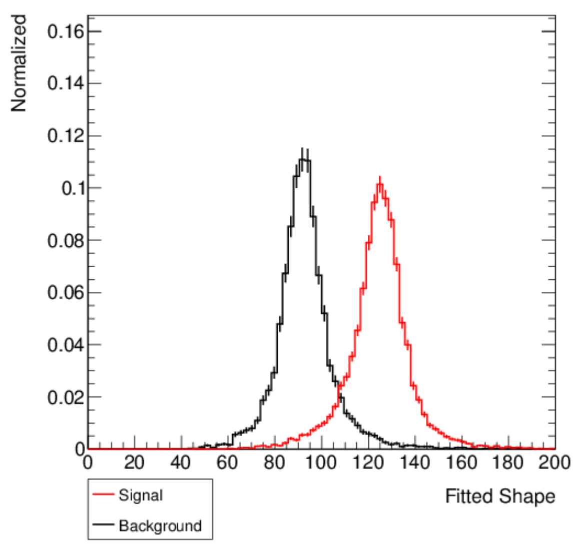

Figure 2: Signal and background peaks normalized to unity.

Figure 3 shows the negative log likelihood curve corresponding to 1 fb^{-1} of data. Assuming no systemic error, we can exclude the background only hypothesis at a confidence level 0.316σ sigma. Scaling this analysis for larger luminosities we project that this channel is sensitive to a 5σ discovery of the Higgs at approximately 250 fb^{-1} of collected data.

Systematic error is modeled by introducing signal and background rate uncertainties in the likelihood fit. In Figure 3, we can see that 10% systematic error increases the uncertainty in our measurement by a factor of two at high integrated luminosities where the measurement is not limited by statistics.

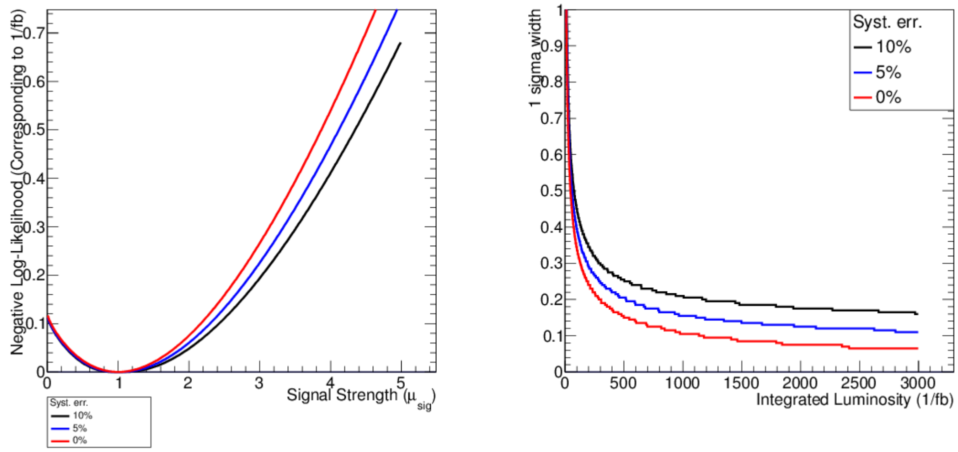

Figure 3: Left: The negative log likelihood fit the background plus signal shape versus a background only hypothesis. The intersection with the y-axis is the square of the z-value for the distribution shape. Right: 1σ width as a function of integrated luminosity.

5. Exclusion of New Physics

We attempt to distinguish between a hypothesis that assumes Standard Model cross sections and one where these cross sections have been modified by a BSM parameter κ_g, introduced in Section 1. When κ_g is between zero and one, and c_t is equal to $1-\kappa_g$, the Higgs cross section at low p_T ranges is scaled below the SM value, while at high p_T ranges it is scaled above the SM value. The effect of this is to keep the overall inclusive Higgs production rate the same while modifying the shape of the distribution. In p_T bins above 200 GeV, we are able to distinguish the κ_g model point distribution from that of the SM Higgs. The significance of κ_g is explained in greater detail in [2].

In this analysis, we set $(c_t, \kappa_g) = (0.5, 0.5)$. We begin by selecting only events with $|\mathrm{M}_{\mathrm{col}} - m_H| < 15$ GeV. This improves our signal to background ratio by roughly a factor of 10. We then scale the cross section in high p_T bins according to the scaling factors of the $(c_t, \kappa_g) = (0.5, 0.5)$ model point. These are detailed in Table V of [2]. We again apply a negative log likelihood test with our null hypothesis being the SM distribution, and our H1 hypothesis being the distribution for a model point of $\kappa_g = 0.5$. Distributions for the two hypotheses are shown in the left plot of Figure 4.

5.1. Results

This fitting procedure provides us with a useful way to distinguish between p_T distributions. In the lower p_T bins the difference in shape is small. As such, it is difficult to distinguish between hypotheses. We expect the difference in shape to more significant in higher p_T ranges [2], but there are also fewer events passing the event selection at higher p_T.

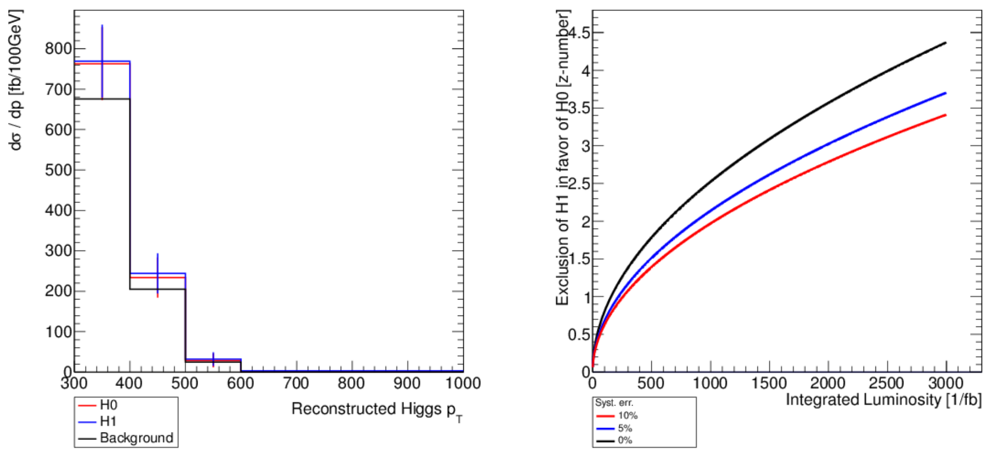

Figure 4: Left: Reconstructed Higgs p_T distributions for the H0 and H1 hypotheses stacked on top of the Z+jets background at 300 fb^{-1}. There is insufficient data to make a fit in higher p_T ranges. Right: Exclusion of H1 as a function of luminosity.

Assuming no systematic error, we can exclude the $(c_t, \kappa_g) = (0.5, 0.5)$ model point with 95% confidence with approximately 600 fb^{-1} of data. With 10% systematic error, in order to exclude the $\kappa_g = 0.5$ hypothesis with a 95% confidence level we require approximately 1100 fb^{-1} of data. The full result is summarized in the right plot in Figure 4.

6. Conclusion

Boosted $H \to \tau\tau$ decays are studied at detector level using a basic event selection similar to the selection suggested in [2]. The boosted $H \to \tau\tau$ channel becomes sensitive enough for SM Higgs observation at an integrated luminosity of 250 fb^{-1}. As such, this channel is a feasible alternative to $t\bar{t}H$ for Higgs measurements in the near future. Measurements of the top Yukawa coupling and excluding more extreme models of BSM physics becomes feasible near an integrated luminosity of 1100 fb^{-1}.

Future analysis that introduces multivariate event selection and the inclusion of the $H \to WW^*$ and $t\bar{t}$+jets processes may improve this outlook. The boosted $H \to \tau\tau$ channel suitable for a long term analysis of the Higgs top loop interaction, but $t\bar{t}H$ remains preferable for the near future.

7. Acknowledgments

We thank S. Chekanov for help with event generation and management. We also thank Reinhard Schwienhorst for useful discussion regarding `MadGraph5_aMC@NLO` and the HEFT model.

References

[1] ATLAS collaboration, G. Aad *et al.*, Phys. Lett. **B716** (2012), 1207.7214.

[2] M. Schlaffer *et al.*, Eur. Phys. J. C **74** (2014), 1405.4295v2.

[3] M. Schlaffer, (2015), 1405.4295v2.

[4] J. Alwall *et al.*, JHEP **79** (2014), 1405.0301.

[5] T. Sjostrand, S. Mrenna, and P. Z. Skands, JHEP **0605** (2006), 0603175.

[6] DELPHES3 collaboration, J. d. Favereau *et al.*, JHEP **1405** (2014), 1307.6346.

[7] G. Cowan, K. Cranmer, E. Gross, and O. Vitelss, Eur. Phys. J. C **71** (2011), 1007.1727.

[8] G. Cowan, *Statistical Data Analysis* (Oxford Science Publications, 1998), chap. 6.

Appendix A. Selection Descriptions

We provide a brief explanation of the initial selections used in this paper.

1. $n_l = 2$ selects only those events that have exactly two oppositely charged leptons in the final state. We further require that each lepton have p_T ¿ 10 GeV. Although our generator level cuts should guarantee there are no low p_T leptons, this extra requirement was added for redundancy.
2. $m_{ll} = 2$ selects those events with a dilepton mass greater than 15 GeV. The dilepton mass is the sum of the two lepton masses.
3. $p_{T,H}^{rec} > 200$ GeV requires all events to have a Higgs or Z boson with transverse momentum greater than 200 GeV.
4. $n_j = 1$ with $p_T > 200$ GeV requires each event to have exactly one very high p_T jet. This ensures that the Higgs is boosted.
5. $n_b = 0$ requires that there be no jets produced from b quark hadronization.
6. "$\not{p}_T$ inside leptons" requires the invisible decay products to be highly collimated with the end state leptons. This allows for an accurate reconstruction of the Higgs four-vector.

A more detailed discussion of these cuts can be found in [2].