Boosted Higgs kinematics

Tin Hadži Veljković (mentored by dr.sc. Vuko Brigljević)
Deparment of physics, Faculty of natural sciences, Bijenička 32, Zagreb

Kinematics of a highly boosted Higgs boson plays a crucial role in the role in analysis of BSM (beyond standard model) particles. In this work we present a detailed analysis of boosted Higgs kinematics using two different samples.

I. INTRODUCTION

The Higgs boson is a scalar boson responsible for other particles gaining mass through the interaction with it. Though the existence of Higgs boson has been hypothesised in the early 60's, the actual detection did not occur until 2012 at LHC (CERN). Higgs boson is a short lived particle with a lifetime of order of $\sim 10^{-22}$ s. There are multiple decay modes, as shown in the figure 1. The latter figure is a plot of branching ratios against the Higgs mass - this is a remnant of the times when Higgs mass was not yet known. However, since we know that it is equal to about 125 GeV we can see from figure 1 that the dominant decay mode is $H \to b\bar{b}$. The initial detections of the Higgs boson were the decays into $W^{\pm}$ and Z bosons, as well into γ, even though they are less probable.

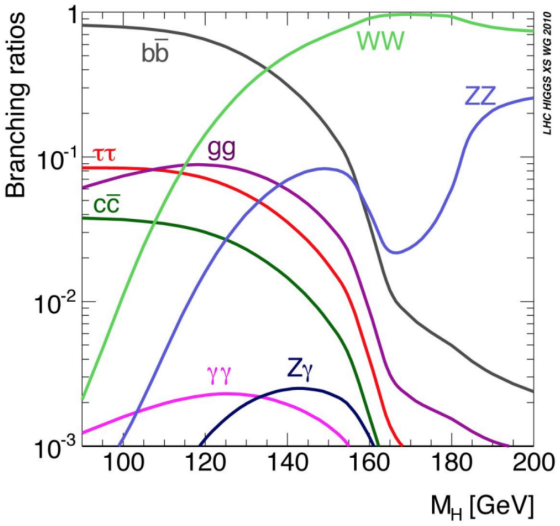

Figure 1. Branching ratios of Higgs boson with respect to Higgs mass.

The reason is twofold. On one side, the background for these decay channels (ZZ, $\gamma\gamma$ and WW) is smaller. On the other side, the mass resolution for the Higgs boson mass in the ZZ and $\gamma\gamma$ decays is also much better than in the bb final state. The $b\bar{b}$ decay is followed by a hadronisation process where hadron jet showers are produced, with a large number of final state particles. Also what makes detection through this decay channel hard is a fact that many other processes produce a similar particles (production of b quarks and B mesons). The process of hadronisation is schematically shown in the figure 2, along with the type of process that we shall be analysing in this paper. The Higgs bosons in our interest are produced in a proton-proton collisions.

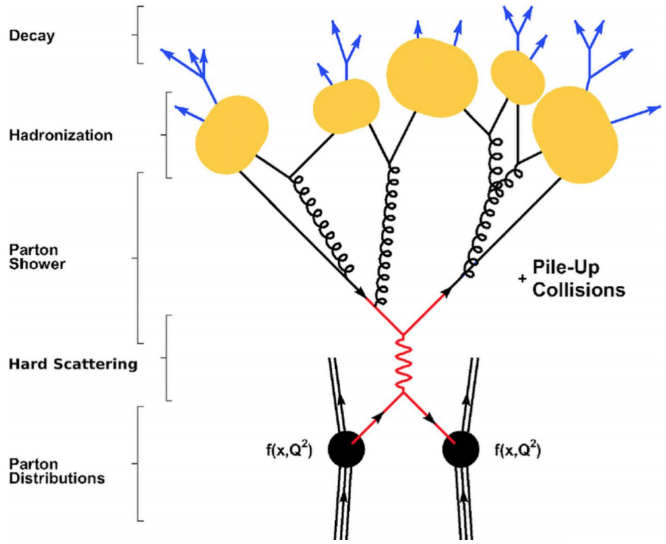

Figure 2. Graphical representation of hadronisation process.

A parton from one proton interacts with a parton from the other proton. In this basic interaction, referred to as the hard scattering (see Fig. 2), other particles can be created. We are interested here in processes in which one or more Higgs bosons are created in this elementary process, and the Higgs boson(s) decay into a pair of b quarks. b quarks propagate as free particles, but due to color confinement they hadronize - produce a directional stream of particles that we call *jet*. Jets are going to be one of the key objects of our interest. They are not, however, fundamental objects, but rather a method to cluster a group of particles using an algorithm of choice.

Boosted regime

Kinematic variables that are suitable for detectors such as CMS are a transverse momentum and pseudorapidity. Here we define z-axis to be the axis of the particle beams, while x and y-axis form a plane perpendicular to the z-axis. Then, the transverse momentum is defined as:

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{1}$$

Since our detector is cylindrically shaped, the transverse momentum would correspond to a net momentum perpendicular to the axis of symmetry, since we choose the z-axis to be the beam axis. The second kinematic variable is rapidity[2]:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \tag{2}$$

However, the more widespread variable is pseudorapidity, which is an approximation for rapidity:

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \tag{3}$$

The angle θ in the definition above is the angle between particle three-momentum and the beam axis. In the limit when mass is negligible compared to the momentum, the rapidity and pseudorapidity are roughly the same. Even though pseudorapidity is not a bounded function, in any realistic detector the values of η for particles that we detect roughly do not exceed value of 3, the particles with higher values travel close to the beam axis are undetected. Graph of pseudorapidity is shown below.

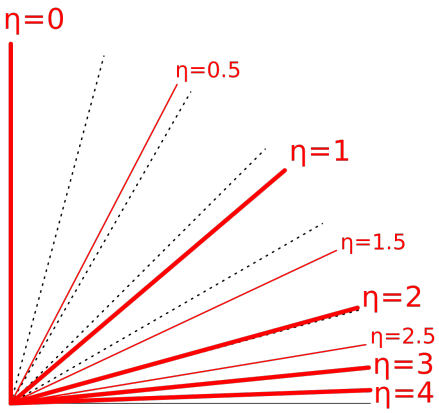

Figure 3. Pseudorapidity values for different travelling angles.

And finally, the last important kinematic variable that will be used often is ΔR defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{4}$$

where angle ϕ is the angle between x and y axis. Intuitively, ΔR is a quantity that roughly corresponds to an angle distance between two particles or jets. It is also worth noting that the set (p_T, η, ϕ, m) uniquely define a 4-vector since there is a bijection between that set and (E, p_x, p_y, p_z).

The main reason that makes boosted decays interesting is that we wish to test BSM model that includes decays such as X $\rightarrow$ HH, a decay of a hypothetical particle into two Higgs bosons. Such event is often called a di-Higgs decay. The BSM particle that we shall be investigating has a mass above 1 TeV, meaning that decay will produce two highly Lorentz boosted Higgs bosons, therefore, boosted analysis is necessary.

II. SIMULATIONS

Before analysing actual simulated data, we shall first present the simplified simulation of H $\rightarrow b\bar{b}$ decay, making use of Python and taking simple kinematical aspects into consideration. Afterwards, we shall explain the process of generating simulations of proton-proton collisions which we use for the most of analysis presented in this paper.

Python simulations

In this simple simulation, we model the distribution of ΔR between decay products of a heavy spin 0 particle (Higgs boson). We assume that the particle decays isotropically in its center of mass frame.

We have generated random orientations of the decay in CM frame by randomizing $\cos\theta \in [0, 1]$ and $\phi \in [0, 2\pi]$ where θ is the angle between the Higgs boson flight direction and the decaying particle in the Higgs boson center of mass frame. Then, the decay products were boosted using Lorentz boost matrix into the lab frame. The procedure was done for various values of Higgs p_T. Value of ΔR was then calculated and plotted against Higgs p_T. From theory we know that the most probable ΔR the distance is:

$$\Delta R_T = \frac{2m_{\mathrm{H}}}{p_T} \tag{5}$$

where m_{H} is the mass of Higgs. We wanted to test whether our simple simulation would reproduce such trend, so we plotted the density plot of ΔR between b quarks and the latter equation on the same graph (red line), as shown on figure 4. We can see a perfect correspondence between the theoretical value and the simplified simulation. It is interesting since the simulation is based purely on kinematical aspect, with no particle physics included.

Clearly, this predicts that the distance between two b quarks reduces with the increase of Higgs p_T. This is expected, however the b quarks are not particles in the final state since quarks hadronise and produce jets. One would expect the similar behaviour for jets as well, but to test that we need to use more physical simulations.

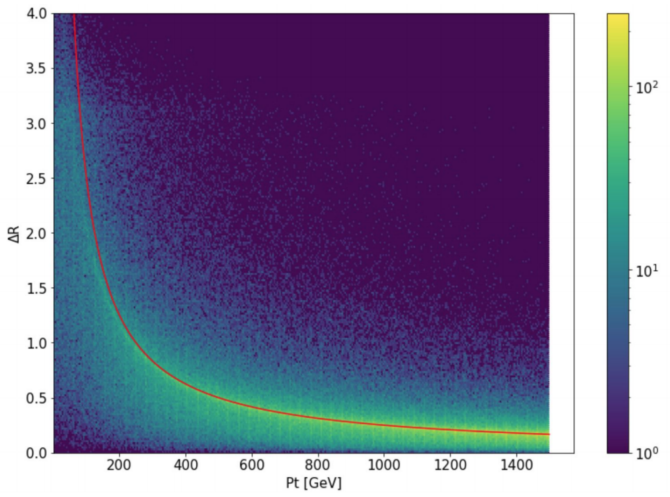

Figure 4. Density plot of ΔR between two quarks against Higgs p_T

Simulated data

In this step, we use fully simulated proton proton collisions in which H $\rightarrow b\bar{b}$ decays are produced. Such simulations are generated in three steps, which we shall briefly describe.

The first step is a Monte-Carlo simulation. POWHEG was used to compute the matrix element probabilities corresponding to the elementary processes considered in these events, while PYTHIA was used for to model the parton showering and hadronization processes which lead from the partons coming out of the hard process to the hadrons observed in the detector. This step simulates collision by computing all possible Feynman diagrams for a given process and recreates the stochastic outcome of such events. Briefly, this step recreates all the particles before they reach the detector, with their properties logged (such as energies, velocities, etc.)

The next step is simulation of the detectors response to the particles from the first step. Naturally, this depends on the detector that we are interested in. This process is extremely computation-heavy and usually takes the longest to complete, since all of the particles need to be propagated through the detector. This is difficult because the particles passing through a layer of detector dissipate energy in it and the propagation needs to be done iteratively.

The final step is reconstruction. So far, we have known exact information about the particles produced in the collision, as well as their energy deposition in the detector. In the final step we only use the signals from the second step and try to reconstruct the original collision event (produced particles, their 4-momenta, etc.). It is important to note that in this step the information from the first step. The same reconstruction algorithms are used to process real data measured by the detector.

The data that shall be analysed is in form of a Root tree file. Tree file contains a large number of collision events, along with the information about every particle, jet and lots of variables corresponding to them. It also contains generated jets (jets from the first step in the simulation as described below) and reconstructed jets (jets from the third step of the simulation). They will play a key role in our investigation of boosted Higgs regime.

III. ANALYSIS AND RESULTS

The analysis was done on two different samples. In the first sample, the Higgs boson is produced in association with a W^+ boson. In the second sample, the decay of a BSM particle into 2 Higgs bosons is simulated. Both samples had Higgs decaying into b quarks. First we shall present the results for the single Higgs production.

Single Higgs decay analysis

The Root tree for single Higgs production contains around 10^6 events, which is good for statistical analysis since the sample is significant. Initially we want to check whether the simulated Higgs decay will provide the results which we obtained in the previous section using simple python simulation. The plot on figure 5 shows the dependence of R between two b quarks with the respect to Higgs p_T. This was done in the following way - for a single Higgs decay, two b quarks (when writing two b quarks we mean b quark and anti-quark) were identified and by collecting their pseudorapidity and polar angle, we have calculated corresponding ΔR and filled the density histogram along with the corresponding Higgs p_T from that decay. Then, we have looped over all events in the tree and the final histogram was obtained.

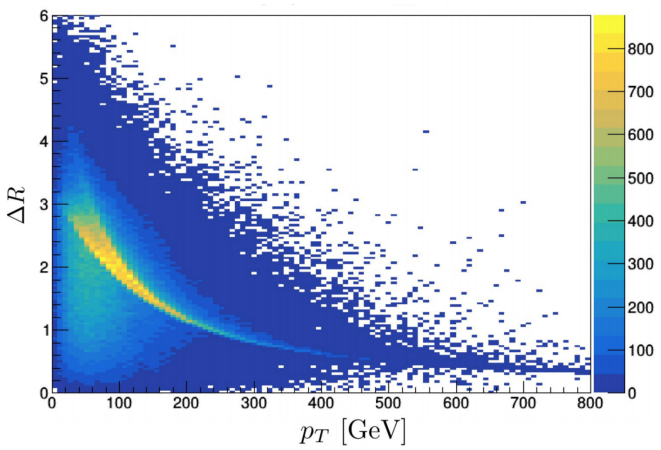

Figure 5. Density plot of ΔR between two quarks against Higgs p_T from single Higgs sample.

By comparison of the latter plot with the python simulation, it is clear that the results are in agreement.

There were few difficulties that needed working around. One would expect to see only two quarks from the Higgs decay, but when inspecting the Root file, there were at least 4 quarks. This raised a question which quark should be used and what do other quarks mean. By further investigation of the Root file this issue was resolved as it turns out that the other quarks correspond to the same physical quarks, but they were listed separately in the Root file due to internal bookkeeping.

We have seen the prediction for the distance between quarks, but as we know they hadronize and we need to observe the hadrons produced in the process to be able to make predictions about the final state particles which pass through the detector. From the Root tree file we see that b quarks produce B mesons. We are interested in their kinematics since they follow up on the quarks. One would expect that for highly boosted quarks (low ΔR), the mesons would minimally deviate from the quarks trajectory, which would imply that ΔR between mesons should roughly be same as the one between quarks. For that purpose we have investigated ΔR between mesons compared to the ΔR between the quarks that produced those mesons. If our predictions were to be true, we should have a peak in density plot on a line with a slope equal to 1. In figure 6 we can see such a clear confirmation of our predictions, meaning that the B mesons do not deviate appreciably from the trajectory of the b quarks.

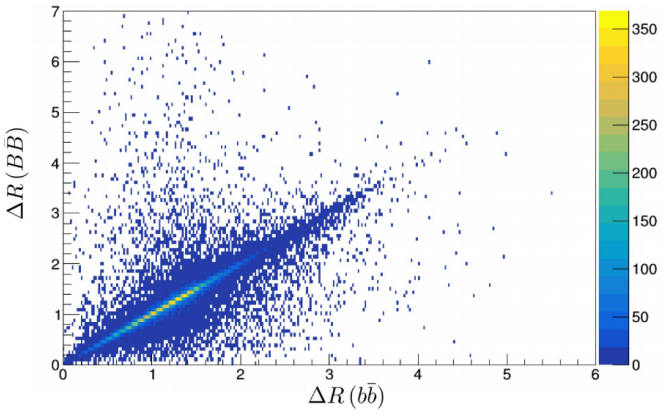

Figure 6. Density plot of ΔR between B mesons against ΔR between corresponding b quarks.

This is especially true for lower values of ΔR, since those quarks are considerably boosted, therefore the mesons will deviate even less from the initial trajectory. At very low values of ΔR we don't have so many entries, but this is due to the momenta spectrum of the decay.

After inspection of the B mesons, we shall move onto the analysis of jets. In the Root tree file we distinguish two types of reconstructed jets: narrow jets and fat jets. However, the Root file contains also generated jets, the jets that were obtained by clustering the generated particles coming out of the proton proton interaction in the Monte-Carlo simulation, while the previous two types of jets were reconstructed in the third step of the data obtaining process. We shall often use generated jets as they present the upper limit of the precision of measurement.

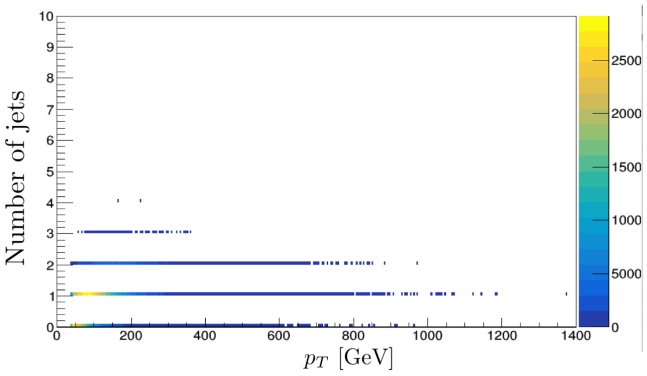

Figure 7. Density plot of number of jets around Higgs ($\Delta R < 1.0$) against Higgs p_T.

The reconstructed narrow jets were clustered with anti-k$_T$ algorithm[1], as well as the fat jets. Narrow jets represent a group of particles of width of roughly $\Delta R < 0.4$, compared to the fat jets whose $\Delta R < 0.8$ - a larger group of particles are being clustered together. Since Higgs decays into two b quarks, one would expect two narrow jets originating from the quarks. This is true, at least for the lower values of p_T. At higher transveral momentuma, the distance between quarks (i.e ΔR) is being gradually reduced. Therefore, at some point, the distance between two jets shall become smaller than $\Delta R < 0.4$, and the two narrow will be clustered as one. This is an interesting kinematical limit to observe, so we have investigated the dependence of number of jets with the respect to Higgs p_T. This is plotted in the figure 7. The number of jets that are plotted is the number of jets that have $\Delta R < 1.0$ with Higgs. From the plot below it is clear that around 600 GeV entries for 2 jets rapidly start diminishing as 2 separate jets start merging into a single jet.

Now we have a general idea of what to expect for Higgs decay in the boosted regime. We now wish to determine the precision and efficiency of the Higgs mass. On the histogram on figure 6. we see a Higgs mass spectrum calculated using the invariant mass of two narrow jets. It is clear that the spectrum has a spread, although the peak is at around 125 GeV. The tail on the left side of the graph is partially due to neutrinos. As we are dealing with reconstructed jets, the neutrinos are not being clustered since we cannot detect them. This is the reason for undervalued tail in the histogram. However, it is also possible to overestimate the mass. For example, particles from the underlying event (proton-proton collision) might get clustered, resulting in larger jet mass. One way to understand whether this tail is really coming from neutrinos is to manually cluster the neutrinos by

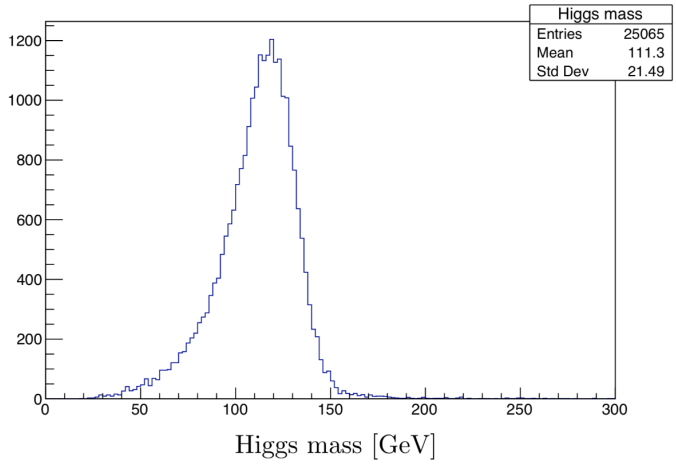

Figure 8. Higgs mass spectrum obtained by narrow jet reconstruction.

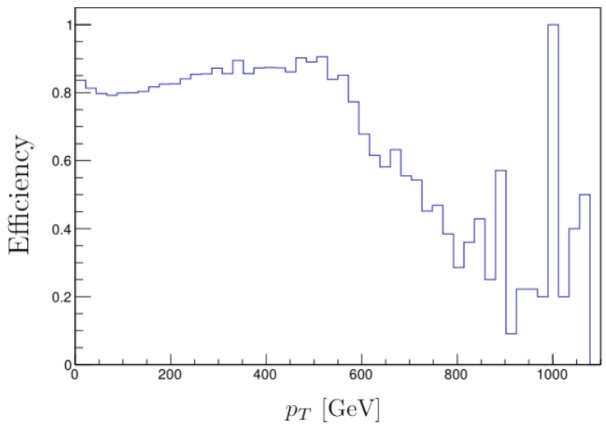

Figure 9. Efficiency of narrow jets with respect to Higgs p_T.

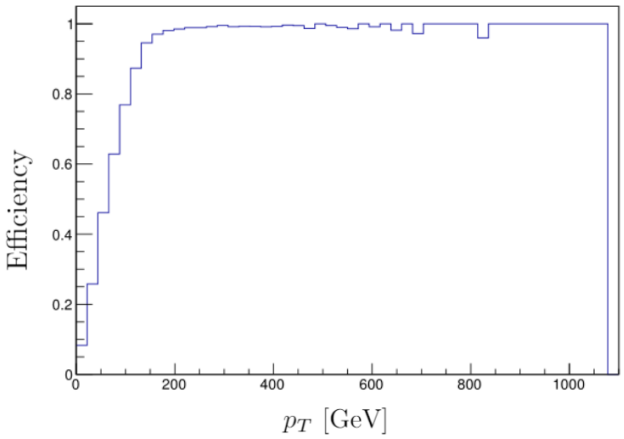

Figure 10. Efficiency of fat jets with respect to Higgs p_T.

adding their 4-vector to jets which might help diminishing the lower tail of the latter distribution. This can only be done with the generated particles (not reconstructed) and we shall try to implement this later.

We can reconstruct the Higgs mass by using narrow jets and fet jets. Natural question that arises is which method is better. The answer to this question is not unique, as the best method depends on the kinematics involved. We shall present the efficiency of reconstructing Higgs for narrow and fat jets with respect to Higgs p_T. Efficiency presented here is defined as following. One histogram of Higgs p_T distribution is filled with an entry only if we are able to find two separate narrow jets originating from the b quarks. Requirement is that the distance between quarks and jets cannot be grater than $\Delta R = 0.4$ for narrow jets, and $\Delta R = 0.8$ from both quarks for fat jets, as we expect only one fat jet. Also, we require that the same jet is matched for both quarks in the case of narrow jets. Second histogram is filled for every Higgs entry. Finally, the two histograms are divided and the resulting y-axis corresponds to the probability of reconstructing jets for a given Higgs p_T. In figure 9 we can see the efficiency dependence of narrow jets with respect to Higgs p_T and on figure 10 we see the same plot, but only for the fat jets. From these plots we can conclude that the efficiency at low Higgs transverse momentum is better if we use narrow jets, which drastically drops around 550 GeV. Here, narrow jets are mostly clustered as one jet, which reduces efficiency drastically as we are not able to identify two separate jets. However, quite the opposite is true for fat jets - the efficiency rapidly increases with the increasing Higgs p_T. This is also easy to understand, at lower values of p_T, it is possible that we cannot find one fat jet that would correspond to both quarks, as quarks can be spread out one from another (or even back-to-back at very low values of p_T).

Di-Higgs decay analysis

For this analysis we have used another sample, a BSM sample that contains decay X→ HH. We can see the mass spectrum for the X particle in the figure 11. We can see a peak around 2 TeV and due to that we expect the resulting two Higgs bosons to be highly Lorentz boosted. In this section we will analyze the efficiencies of narrow and fat jets for Higgs detection and we will inspect the differences in mass resolution in the case of manual clustering of neutrinos into fat jets compared with the resolution with neutrinos disregarded.

First, let's consider the efficiency of narrow jets in the di-Higgs sample. In figure 12 we can see two plots with efficiencies of narrow jets for both Higgs bosons separately. They have a similar trend compared one to another, but that is expected as the process is symmetric in a way that there is no preference towards decay products. Comparing this to the plot on figure 9, we see a similar drop-off in efficiency around 550 GeV. Efficiencies at lower Higgs p_T are lower in the di-Higgs sample, due to the fact that

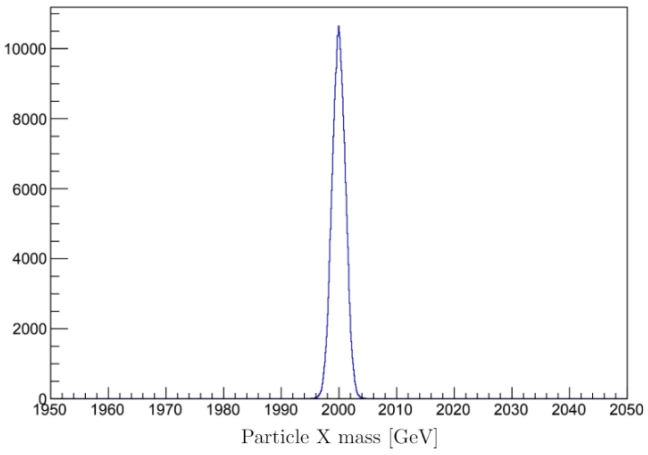

Figure 11. Mass spectrum of BSM particle X.

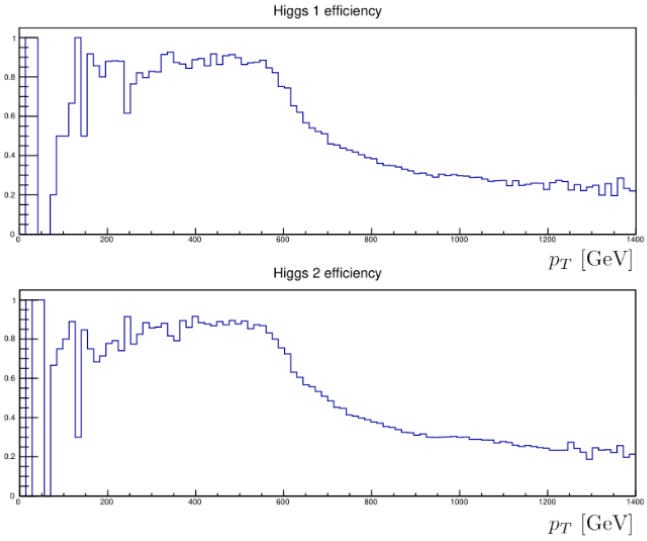

Figure 12. Narrow jets efficiency for both Higgs bosons in the di-Higgs decay with respect to Higgs p_T.

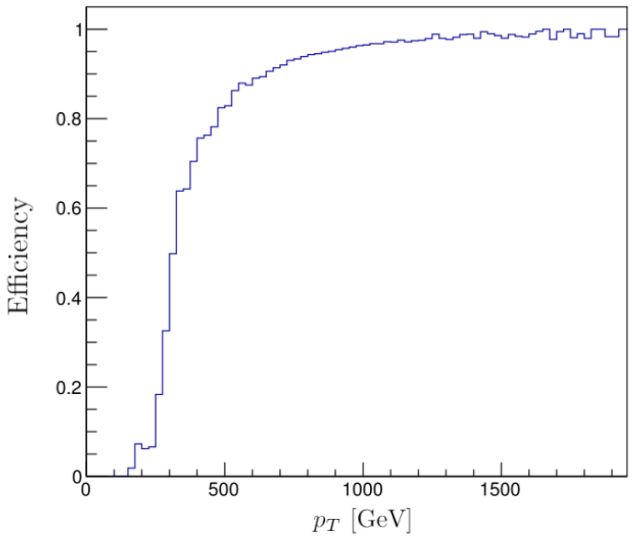

Figure 13. Fat jets efficiency in the di-Higgs decay with respect to Higgs p_T.

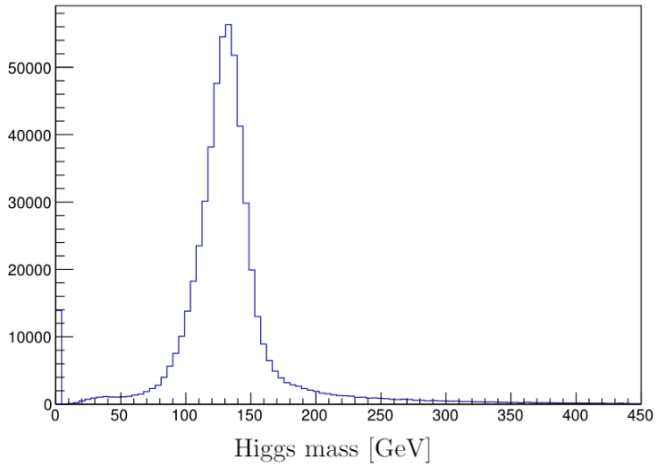

Figure 14. Mass spectrum of Higgs boson for reconstructed fat jets.

decay of X produces highly boosted Higgs bosons. We will also check the trend of the fat jet efficiency for the di-Higgs sample, as shown in the figure 13. Again, this is similar to the efficiency in the figure 10, which presented the efficiency of fat jets for the single Higgs decay.

The last aspect that we shall consider is the differences of mass resolution mentioned earlier in this section. First we need to check whether the generated AK8 jets include neutrinos or not. We do that by comparing the mass spectrum with the mass spectrum of reconstructed fat jets. If both have a tail on the left side, this implies that the neutrinos are not clustered. From figures 14 and 15 we clearly see previous statement to be true. Now, we wish to increase resolution obtained by the generated AK8 jets by manually clustering neutrinos to the jet. We do that by searching for neutrinos in the vicinity of $\Delta R <$ 0.8 from the generated fat jet and then adding them to the fat jet 4-vector. We use the new vector to find the invariant mass, i.e Higgs mass. Result of this calculation is shown in figure 16. What we can see is a obvious loss of the tail on the left side of the spectrum. This was previously expected, as now we underestimate Higgs mass appreciably less due to neutrinos getting clustered in the jet.

Finally, we will try to match generated AK8 jets with their corresponding reconstructed fat jets. We assign generated AK8 jet to the reconstructed fat jet by searching for closest AK8 jet to the fat jet.

After matching, the ratio of the mass obtained by the AK8 generated jet with the neutrino included and mass of reconstructed fat jet is calculated and plotted in figure 17. With this, we conclude our analysis of Higgs proper-

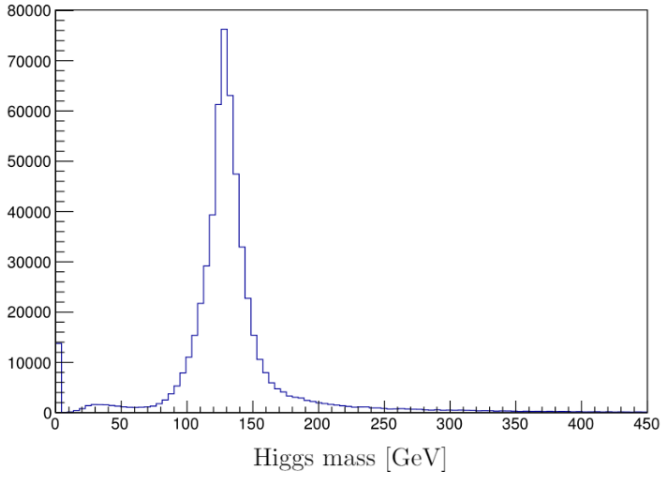

Figure 15. Mass spectrum of Higgs boson for generated AK8 jets (generated fat jets).

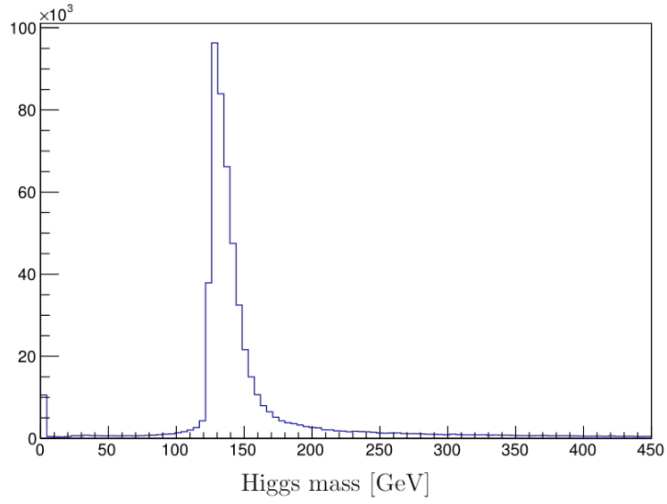

Figure 16. Mass spectrum of Higgs boson for generated AK8 jets (generated fat jets).

ties in the boosted regime.

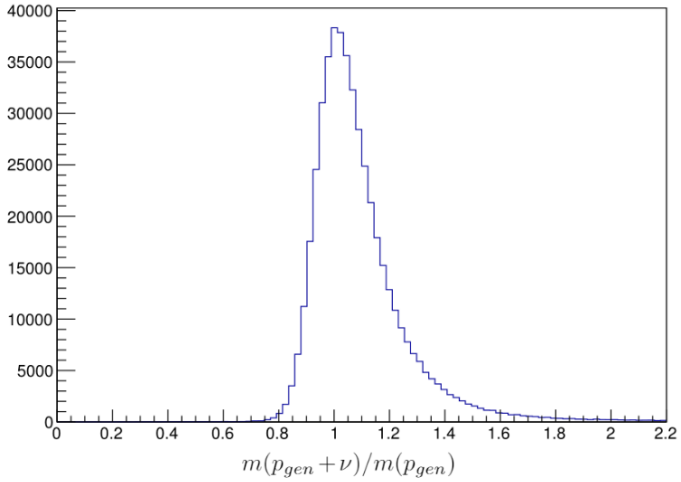

Figure 17. Ratio of Higgs masses obtained by reclustered AK8 generated jet and reconstructed fat jet.

IV. CONCLUSION

In goal to analyze properties of Higgs boson in the boosted regime, we began our research by constructing simple Python simulations. Afterwards, we used actual simulations for proton-proton collisions in which Higgs boson was produced. Finally, we presented the usage of such analysis for a BSM (beyond standard model) sample in which massive particle X decayed into two Higgs bosons (di-Higgs decay). Using these samples we calculated mass spectra, properties of generated and reconstructed narrow and fat jets, efficiencies and researched ways to improve the mass resolution. Further research would implement concepts such as b-tagging and mass regression using machine learning techniques to improve mass resolution even further.

[1] M. Cacciari, *The anti-k_t jet clustering algorithm*, arXiv:0802.1189

[2] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013