Vertex Identification Optimization in the H→γγ Decay Channel

Julia Gonski (Advisers: Pasquale Musella, André David)

I. Introduction

In the reconstruction of Higgs to diphoton events, invariant mass resolution is primarily defined by two factors: energy resolution of the detector, and resolution of the opening angle between the two photons. To keep the opening angle resolution degrading mass resolution, the location of the Higgs production vertex must be known within 1cm. Enhancing the efficiency of vertex identification would therefore increase the feasibility of the Higgs mass constraint in future analyses, by improving mass resolution and consequently the signal to background ratio. Additionally, one of the challenges of the expected 2014 proton energy increase at the Large Hadron Collider will be pile up contribution, specifically in the task of vertex identification. The increase in pile up and luminosity thus provides additional motivation for improvement on current Higgs vertex identification methods.

II. Vertex Identification and Time of Flight Discriminant

Higgs to diphoton vertex analysis is unique in that the final photons do not leave tracks in the detector unless they convert to electron-positron pairs. This separates the task of vertex identification into two categories, for converted and unconverted photons. In the case of unconverted photons, the only information available for vertex determination is related to kinematics of the calorimeter supercluster. Additional information about the photon can be found from its time of flight in the detector, motivating the use of a time of flight discriminant in vertex identification. The discriminant is constructed in such a way that the location of the vertex in the beamline and the time travelled in the detector before collision do not bias the final identification, and is studied assuming a time resolution of 0.01ns.

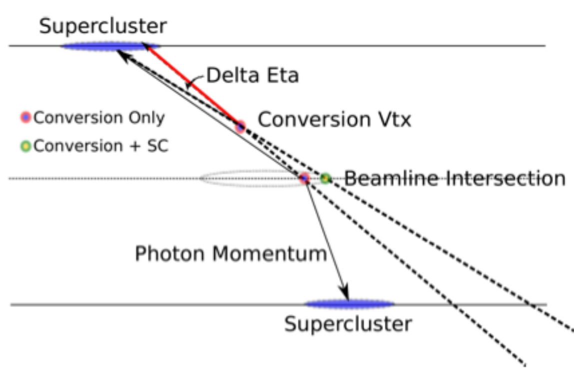

Figure 1: Vertex identification diagram for converted and unconverted photons.

The time of flight discriminant was constructed by examining the diphoton system of equations in the detector (equation 1), and solving for the time of flight of each photon as a function of vertex position. The final discriminant is given in equation 2, where V* is the true vertex, V is a reconstructed vertex, ΔTOF is the difference between time of flights of two photons from a given vertex, and the time of flight is calibrated for vertices in the center of the detector. Figure 2 shows the geometry of the diphoton decay as it pertains to the time of flight calculation.

$$(t_1 - t_0)^2 = R^2 + (z_1 - z_0)^2$$
$$(t_2 - t_0)^2 = R^2 + (z_2 - z_0)^2 \qquad (1)$$

$$\Delta\Delta TOF = \Delta TOF(V^*) - \Delta TOF(V) \qquad (2)$$

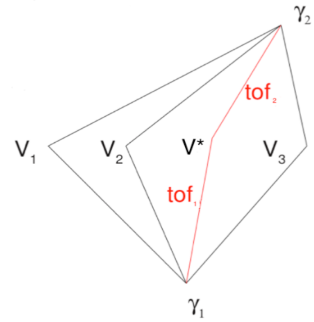

Figure 2: Time of flight diagram, with true vertex V* and numbered reconstructed vertices.

III. Boosted Decision Tree

Vertex identification in Higgs to diphoton events is done using a multivariate method known as a boosted decision tree (BDT). A boosted decision tree takes a set of discriminating variables, along with a set of spectator variables for benchmarking, and 'learns' how to distinguish correct from incorrect vertex identifications using the discriminating variables provided. In this analysis, the time of flight metric was added as a discriminating variable, and a series of BDT trainings were run to obtain the best possible discriminating power. All training was done over Powheg Monte Carlo samples with a Higgs mass of 125 GeV, √s = 14 TeV, and 50 pile up events, with both gluon gluon fusion and vector boson fusion production. A table of discriminating and spectator variables used for this analyiss can be seen in Table 1, below.

Category:	No conversions	≥1 conversions
Discriminating Variables:	• ptasym • ptbal • sumpt2 • absTOF • pho1.Eta • pho2.Eta	• ptasym • ptbal • sumpt2 • $pull_{conv}$ • absTOF • pho1.Eta • pho2.Eta
Spectator Variables:	• diphoton p_T • photon1 p_T • photon2 p_T • # vertices • # pile up	

Table 1: Boosted decision tree variables.

IV. Results and Discussion

Preliminary performance results indicated a lower efficiency increase than expected, propmpting further training to examine the performance of the algorithm in various regions of the detector, each corresponding to a different diphoton decay geometry. Events were classified based on the location of the final photon superclusters, and split into three categories: both photons in the barrel, both photons in the endcaps, or one photon in each. To study the capacity for signal and background separation in each region, the width of the incorrect matching curve was examined as a function of Δη. These plots, shown in Figure 3, demonstrate the widening of the background curve as the photon separation increases, allowing for greater discrimination at high Δη. They also demonstrate a lower discrimination for decay geometries with endcap photons, as seen from the lack of curve separation in the endcap-endcap plot.

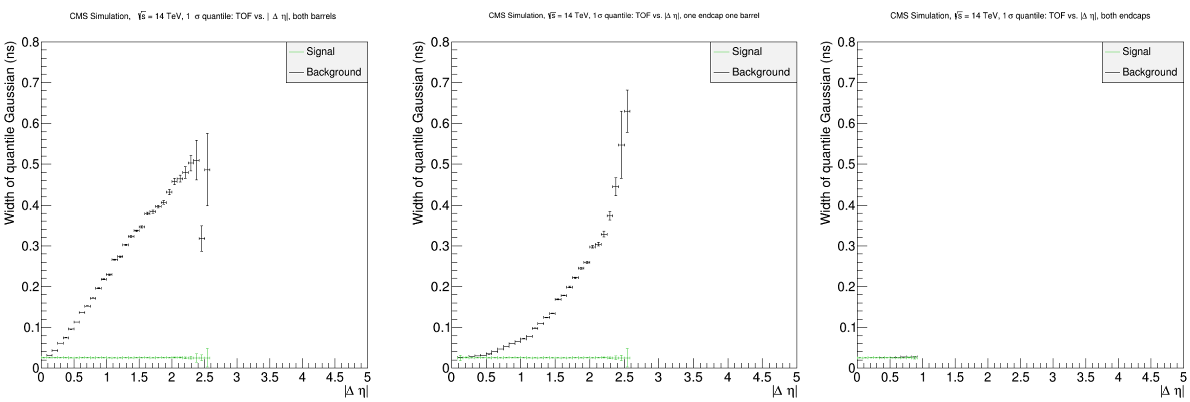

Figure 3: 1σ quantile plots, for barrel barrel, barrel endcap, and endcap endcap (left to right).

Such differences exhibited by the diphoton decay system in various detector regions motivated the individualized training of the BDT for each specific region. In such trainings, performance is measured via ROC curves, which plot background rejection as a function of signal efficiency. To measure efficiency improvements, the ROC curve from time of flight training is compared to a previous ROC curve from the 14 TeV high pile up sample that does not use time of flight in discrimination. As predicted by the above geometry studies, the time of flight training did not provide efficiency increases for events with endcap photons. However, a significant efficiency increase is seen for the barrel-barrel diphoton case. All ROC curve comparisons can be seen in Figure 4, below.

When trained over all detector regions, the time of flight training still exhibits a moderate improvement over the previous training done with 14 TeV high pile up samples. The inclusive ROC curve comparison can be seen in Figure 5. From this result, it is evident that despite decay geometry dependence, the addition of a time of flight discriminant to the BDT training yields an overall increase in vertex identification efficiency. Future trainings with different set of discriminating variables may improve performance in the endcaps, enhancing inclusive performance as a result.

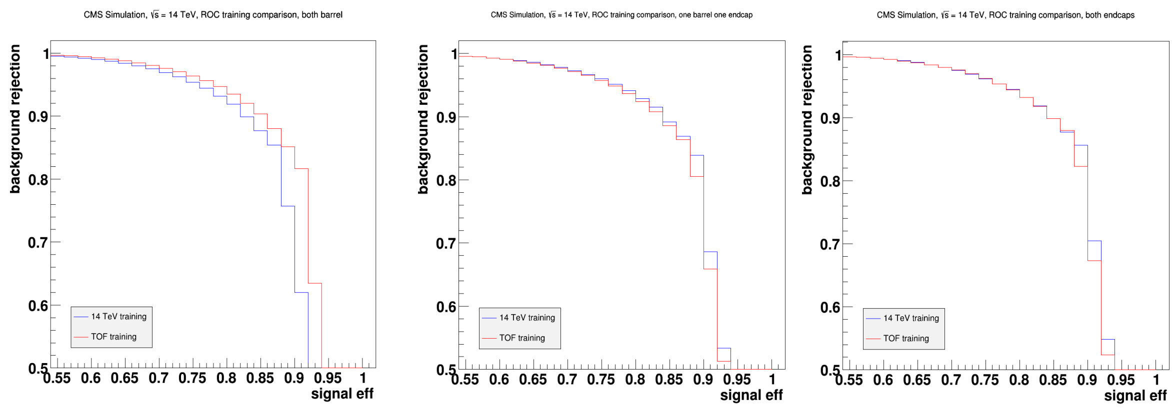

Figure 4: ROC comparisons, for barrel barrel, barrel endcap, and endcap endcap (left to right).

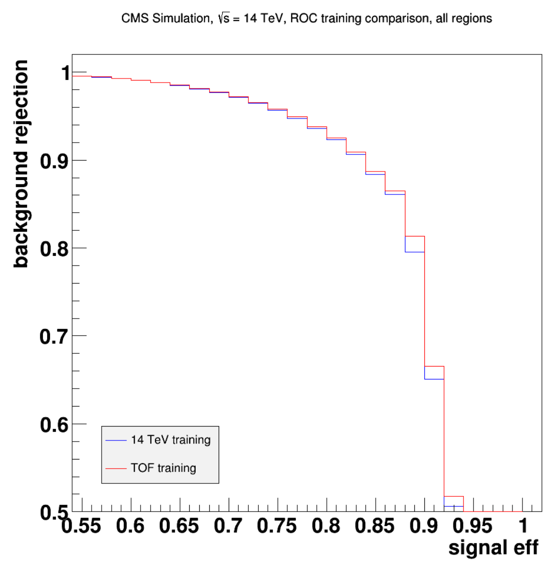

Figure 5: Inclusive ROC comparison curve for time of flight training.

V. Conclusions & Future Studies

A study of vertex identification efficiency in the Higgs to gamma gamma channel has been performed using boosted decision tree multivariate classification. The algorithm uses a photon time of flight discriminant and assumes a time resolution of 0.01 nanoseconds. Since the current time resolution in CMS is 0.1ns, a high precision timing layer would have to be installed for functionality of this method in future analyses. The analysis reveals dependency of vertex identification efficiency on diphoton decay geometry in the barrel and endcap detector regions. However, preliminary efficiency increases suggest the possibility of extrapolation of this method for all regions and higher pile up conditions. Future use of this algorithm would require additional statistics in Monte Carlo samples to achieve better efficiencies in the 2014 dataset, as well as further boosted decision tree training to eliminate detector region dependency.