

University of California
Santa Barbara

Simulation of MTD Performance and Search for Rare Higgs Decays

A thesis submitted in partial satisfaction
of the requirements for the degree

Bachelor of Science
in
Physics

by

Jonathan Kasuke Guiang

Committee in charge:

Professor Claudio Campagnari, Chair
Professor David Stuart
Professor Jeffrey Richman

June 2019

The Bachelor's Honors Thesis of Jonathan Kasuke Guiang is approved.

Professor David Stuart

Professor Jeffrey Richman

Professor Claudio Campagnari, Committee Chair

May 2019

Simulation of MTD Performance and Search for Rare Higgs Decays

Copyright © 2019

by

Jonathan Kasuke Guiang

Acknowledgements

The research included in this thesis could not have been performed if not for the assistance, patience, and support of many individuals. I would first like to extend my deepest gratitude toward my advisor Dr. Claudio Campagnari for his mentorship and constant confidence over the course of my undergraduate studies. It was he who suggested that I study rare Higgs decays, the latter half of this thesis. He also introduced me to Dr. David Stuart, under whose advisement the former half of this thesis was performed. Claudio's support, generosity, and boundless wisdom has made me a better student and physicist.

I would additionally like to thank Dr. David Stuart. His versatile mind and constant effort to encourage learning led me to be more curious and thoughtful. Furthermore, his understanding and amiable treatment of his students promoted a positive environment for learning and practicing physics.

I would also like to extend my appreciation to Dr. Indara Suarez for believing in me and for her continual guidance as well as to Nick Amin, Bennett Marsh, and Sicheng Wang for their companionship and for helping me with anything and everything.

This research would not have been possible without the assistance of the CMS experimental collaboration who constructed the experimental apparatus and built the foundations for the data analysis.

Finally I would like to extend my deepest thanks to my parents Orlando and Cynthia Guiang without whose love, support and understanding I could never have completed this undergraduate degree, nor would I have had the confidence or inspiration to pursue research in physics.

Abstract

Simulation of MTD Performance and Search for Rare Higgs Decays

by

Jonathan Kasuke Guiang

Progress in experimental High Energy physics can be derived from efforts in two fundamentally-linked domains: detector design and data analysis. As such, this thesis covers them both. First, a 3D model of the MIP Timing Detector (MTD) endcap and barrel timing layers – a proposed addition to the Compact Muon Solenoid (CMS) for the High Luminosity LHC (HL-LHC) upgrade – was designed using OpenSCAD, whose modular construction allowed for a configurable model spanning several different detector designs. Particle trajectories from MinBias/ $t\bar{t}$ CMS Monte Carlo (MC) samples propagated through a known map of the CMS magnetic field in addition to the aforementioned 3D model were used to simulate detector performance for the purpose of completing simple, yet accurate studies – primarily concerning the efficiency of various detector configurations – conducted throughout the MTD’s technical design process. Second, a search for the $H \rightarrow \rho\gamma$ and $H \rightarrow \phi\gamma$ decays was performed through associated WH production. This search was motivated by the possibility of probing the Higgs boson couplings to light, flavored quarks, which could be enhanced by new physics. Using the full Run II 137 fb^{-1} pp collision dataset collected by CMS at the LHC, no significant excess of events was observed above background, as expected from the Standard Model. For each analysis, a Boosted Decision Tree (BDT) was trained on CMS background MC and a privately-generated signal sample. Then, an optimal BDT discriminant was determined, and the branching ratio exclusion limits $\mathcal{B}(H \rightarrow \phi\gamma) \leq 4.1 \times 10^{-3}$ and $\mathcal{B}(H \rightarrow \rho\gamma) \leq 7.0 \times 10^{-3}$ were observed to a 95% confidence level.

Contents

Abstract	v
1 Introduction	1
1.1 Particle Physics at the LHC	1
1.2 The Compact Muon Solenoid	2
2 MTD Simulation	4
2.1 Motivation	4
2.2 The Topolino Design	6
2.3 Rendering the Detector	6
2.4 Simulating Performance	8
3 Measurement of Rare Higgs Decays	13
3.1 Motivation	13
3.2 Event Selection	15
3.3 Backgrounds	17
3.4 Boosted Decision Tree	20
3.5 Results	26
A Additional Figures	28
B Technical Details	35
B.1 CMS Coordinate System	35
B.2 Kinematic Variable Definitions	35
B.3 Boosted Decision Trees	36
Bibliography	39

Chapter 1

Introduction

1.1 Particle Physics at the LHC

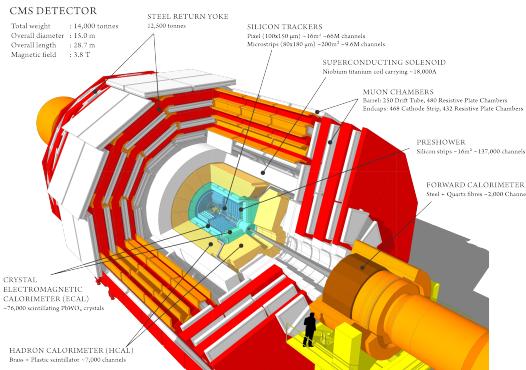
Over the last century, the Standard Model (SM) has been shown to be the most accurate description of the fundamental construction and operation of the universe, but it is still incomplete, giving rise to many theories including Supersymmetry (SUSY), a popular extension of the Standard Model. It predicts a partner particle to every SM particle, which would resolve some particularly bothersome issues with the Standard Model like fixing the mass of the Higgs Boson, the identity of Dark Matter, and many others. Assuming SUSY is real, supersymmetric particles are expected to appear in high-energy collisions at the Large Hadron Collider (LHC). However, direct evidence for SUSY, or for any other competing theory, has yet to be discovered. Therefore, the pursuit of direct discoveries of new physics – where “new physics” is hereafter used to refer to evidence for Beyond Standard Model (BSM) particles – is in a state of flux at the time of writing, and there is a growing doubt within the High Energy community that any discovery is likely to be made in the near future (i.e. within the next century). In response, CERN is investing in enhancing the LHC itself, first during a brief, two-

year (2019 to 2021) shutdown period, then during a much longer overhaul for the High-Luminosity LHC (HL-LHC) with the hope that upgraded software and hardware and an enormous increase in data from the HL-LHC will push the field forward. In addition, contributions from the LHC may come from the plethora of ongoing and future analyses that seek to both answer old questions and raise new, equally concerning ones. In the context of experiments at the LHC, raising a concerning question is as simple as poking another hole in the Standard Model, but the question is only useful if it is more answerable than questions like a seventy-percent deficit in the current description of the makeup of the universe (dark matter) or the possibility of the quantization of gravity at energy scales beyond comprehension (gravitons). Nevertheless, there is a future worth watching at the LHC.

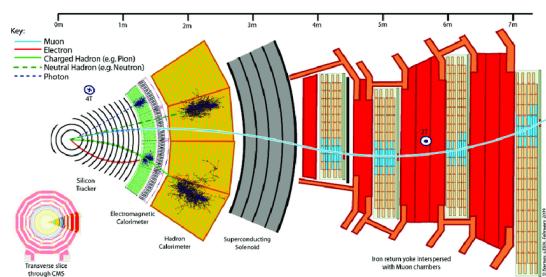
1.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is a general purpose detector at the LHC, meaning it probes for the existence of a variety of new physics. The detector’s namesake is a “compact” solenoid, the largest superconducting magnet ever built (with a length of 15m and diameter of 7m), that generates a 4T magnetic field strong enough to pull many high-energy, charged particles produced in a proton-proton collision towards CMS’s detector layers[1]. Particles first encounter the tracker, which is composed of several layers of silicon pixels and strip detectors. Essentially, a through-going particle leaves a cascading charge deposit that can be measured as an electrical signal. This results in an accurate measurement of the particle’s position as it passes through various sections of the tracker, which can then be reconstructed to give the particle’s entire trajectory. Track information provides fundamental insights like the charge and momentum of a particle, for instance, which can be inferred from the curvature of the particle’s trajectory. Having

passed through the tracker, particles then encounter the calorimeters, which will measure the final energies of emergent particles when possible. Energies of particles that interact by the electromagnetic force, electrons and photons, are measured by the Electromagnetic Calorimeter (ECAL) while those that interact by the strong force, hadrons, are measured by the Hadronic Calorimeter (HCAL). Finally, particles that escape the ECAL and HCAL – now only muons and weakly interacting particles – encounter the muon layer, where muons are tracked further, before leaving the bounds of CMS. Thus, the presence of neutrinos, which mostly pass through the detector undisturbed, and possibly yet-undiscovered particles can only be inferred. However, conservation of momentum is key here: invisible particles will show up as missing energy in comparing the sum of the measured particle momenta to the overall energy of the collision. Otherwise, BSM particles may have been produced inside of the detector, where they then decayed into a unique set (signature) of SM particles that are detectable by CMS, and they would register as an excess above SM prediction in an analysis of that signature.



(a) Cutaway view of CMS detector.



(b) Illustration of detector layer response.

Figure 1.1: Visualizations of the design and function of CMS.

Chapter 2

MTD Simulation

I begin with a description of the motivation behind the proposal of the Minimum Ionizing Particle (MIP) Timing Detector. Then, I describe the process of rendering one of the proposed designs for the detector’s endcap timing layer. Finally, I discuss the mathematics behind the simulation and show some of the resultant plots used during the detector’s technical design process.

2.1 Motivation

The HL-LHC upgrade will offer unprecedented luminosity in which one may find new physics, but it brings with it new complications that must be addressed by upgraded or additional detector hardware. Foremost among the new challenges posed by the HL-LHC is that higher luminosity means many more particles, which provides a far more complex picture to reconstruct for analysis. One such complexity is the increase in collision points, which is a key component of both offline and online reconstruction. With more particles colliding, collision points start to “pile up” on top of each other in space (this phenomena is aptly referred to as *pileup*). Consequently, when reconstructing collisions into discrete

“events,” one for each proton-proton collision, the reconstruction algorithm is unable to discern between one collision and many others that occurred in the same point in space. However, these collisions occur at different *times*, so adding timing information should certainly help reduce pileup. In fact, this has been studied and proven to be the case in simulations (Fig. 2.1). To this end, it has been proposed that a layer of silicon-based sensors called the MIP Timing Detector (MTD), with sufficient timing resolutions, be placed around the perimeter – surrounding the barrel and covering the endcap – of the CMS detector, but the design and approval of such an upgrade is no simple task. From its conceptual inception to the finalization of its technical design, questions continually arise, so clear, cogent answers are constantly required, but since the detector has yet to be built, computer simulations are necessary to address these important concerns regarding the design and projected performance of the MTD.

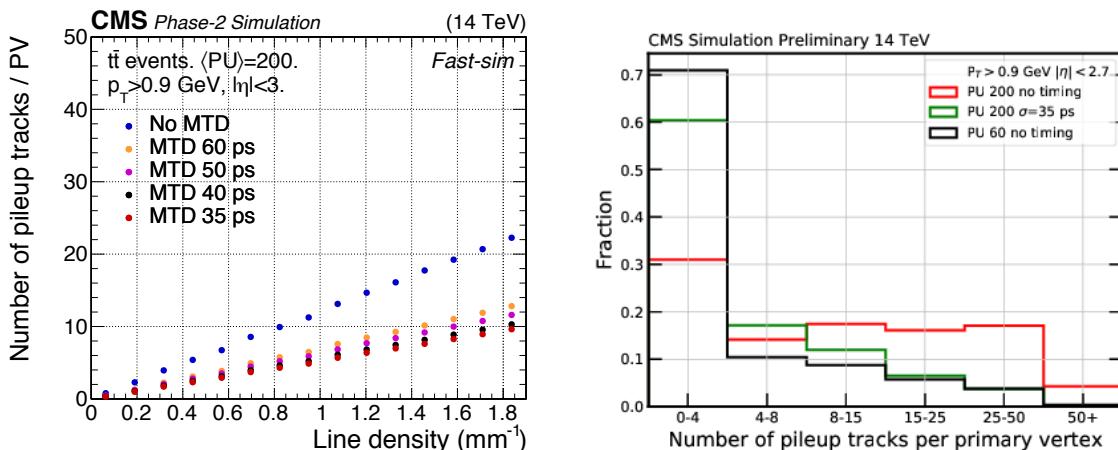


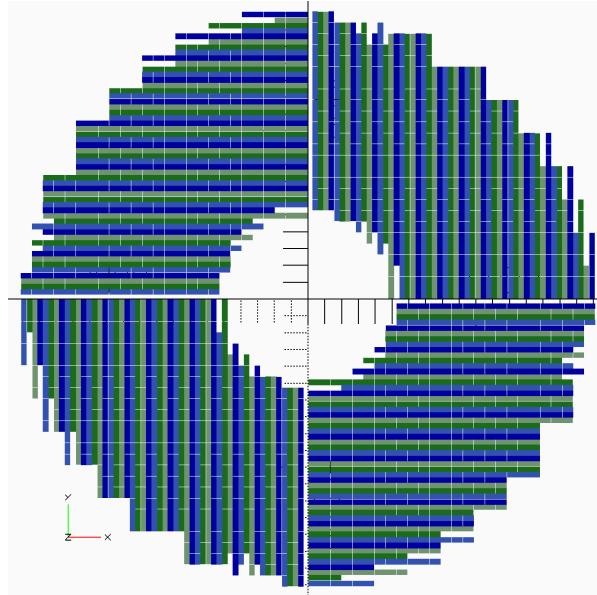
Figure 2.1: Plots from the TDR[2] demonstrating pileup reduction from timing information. (Left) Number of pileup tracks incorrectly associated with the hard interaction vertex as a function of the collision line density for different time resolutions. (Right) Distribution of the number of incorrectly associated tracks with the use of a 3σ (where $\sigma = 35 \text{ ps}$) selection on timing information and without use of timing information; the vertical axis is the fraction of primary vertices which have the number of pileup tracks shown on the horizontal axis associated to them.

2.2 The Topolino Design

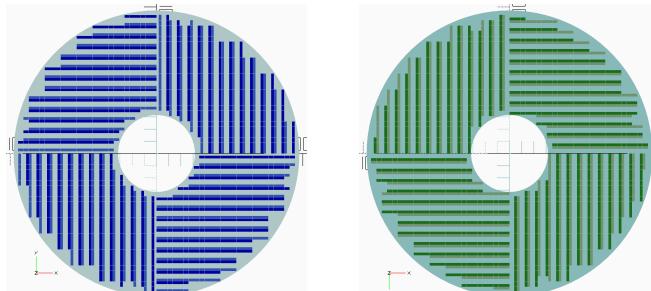
The barrel and endcap layers (BTL and ETL respectively) of the MTD have rotational symmetries – they are round – while the sensors are rectangular, so some careful consideration is required when looking to cover the CMS detector with a new sensor layer. For the BTL, the design is fairly simple: long trays of sensors may be laid along the axis of the barrel, maintaining its cylindrical symmetry. However, correctly fitting the rectangular sensors to the circular space allotted on the endcaps (which are essentially annuli) becomes a complicated exercise in geometric optimization, while including space for all of the required circuitry, wiring, and cooling systems imposes complicated physical constraints. Although the ETL offered unique complications, one design was arrived at – named “Topolino” (Mickey Mouse in Italian) by its inventor – where the ETL is divided into four ninety-degree wedges. The front of each wedge is tiled with parallel strips of sensors that continue up until the the edge of the endcap. The back of the same wedge is similarly covered, but the sensors are horizontally offset to cover the gaps left by the readout electronics on the front. Each wedge is covered in this way, then placed such that the sensors in each is perpendicular to its neighboring wedges. See Figure 2.2 for a detailed visualization of the proposed design.

2.3 Rendering the Detector

Detector physics simulation begins with a simulation of the detector, but for answers to questions about the MTD’s design and performance, verbose considerations of minute physical interactions are, at the moment, unnecessary. Therefore, though more complex tools offer more accurate physics, a simple rendering of each sensor’s position in space is sufficient. However, the simulation must also be configurable, so assembling the it by



(a) Rendered with no mounting disk to show coverage of electronics gaps.



(b) Front of one full disk. (c) Back of one full disk.

Figure 2.2: The “topolino” design.

hand, like most 3D-modeling software requires, was neither an entertaining, nor efficient solution. Instead, OpenSCAD, a C-like programming language that allows for simple, modular construction of three-dimensional models, was selected.

To algorithmically implement the Topolino design, fundamental rules that govern the layout had to be established. First, neither the sensors, nor the space allotted for their circuitry, could be allowed to hang over the perimeter of the endcap. Second, sensors are most easily assembled and placed as modules, so the detector must be tiled by *groups* of

sensors. Finally, the service modules – the “circuitry” for which space has already been allotted – are designed to service sensors on both sides, so sensors should (generally) be placed as such, resulting in a neat grid. With these rules in mind, an algorithm parses the x-axis in increments equal to the width of a sensor module, then places sensors by their lower, left-hand (closest to origin) corners, so long as their placement does not violate any of the previously stated rules. One wedge is tiled in this way where the sensors on its reverse side are placed in the same way but starting from a given displacement from the origin such that the holes left by the spacing for readout electronics are covered by sensors on the opposite side. The rest of the endcap is covered by simply taking this wedge, then placing orthogonally-rotated copies until the entire surface is covered. The images shown in Figure 2.2 were created by the algorithm described here.

2.4 Simulating Performance

2.4.1 Pre-processing

The three-dimensional model of the MTD can be exported as a Standard Tessellation Language (STL) file. During the export process, each independent object – say, a single sensor which is represented as a thin rectangular prism – is broken down into each of its constituent faces. Then, it undergoes the tessellation process, wherein the face considered is broken into some optimal number of constituent triangles (hereafter referred to as *polygons*). The *vertices* of each of these polygons are then written to the STL file, in a clockwise order relative to the origin. It is important to note that this guarantees that the *facet normal* vector – that is, a vector orthogonal to the surface of the polygon with a magnitude equal to the area of the polygon – always points outwards with respect to the origin.

2.4.2 Hit Detection Algorithm

Now, before the simulation itself is generally discussed, the algorithm by which an intersection between a point and a polygon is determined should first be outlined, since it is the core operation of the simulation but is otherwise only tangentially important to the overall process. Consider a triangle in three-space, namely a set of three vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ with facet normal \hat{n} , normalized such that $|\hat{n}| = 1$, pointing outwards relative to the origin. Then, suppose there exists a point \vec{a} somewhere in three-space (see Figure 2.3 for a visual representation of this setup). First, the coordinate system must be rotated into the plane of the polygon (in order to account for both the azimuthal symmetry of the BTL and radial symmetry of the ETL). One basis vector, arbitrarily chosen to be \hat{e}'_3 , is already given by the facet normal vector. Another, now chosen to be \hat{e}'_1 , is given by the original \hat{e}_3 vector (this is \hat{z} is the CMS coordinate system; see Appendix B.1 for more information). With two basis vectors determined, the third is simply given by the cross product $\hat{e}'_3 \times \hat{e}'_1$, assuming they are already normalized (again, see Figure 2.3). From this new set of basis vectors \hat{e}'_i , still represented in the original basis, a matrix R can be constructed to translate any arbitrary vector into the primed coordinate system:

$$\hat{e}'_i = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \rightarrow R \equiv \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix} \quad (2.1)$$

The point \vec{a} and vertices \vec{v}_i are translated into the primed coordinate system such that $R\vec{a} = \vec{a}'$ and $R\vec{v}_i = \vec{v}'_i$. Then, the \hat{e}'_3 coordinates of \vec{a}' and \vec{v}'_i are set to zero, ensuring that all points considered are in the same plane parallel to that spanned by the polygon. Now, the vector from the point \vec{a}' to the vertex \vec{v}'_i is defined as $\vec{d}'_i \equiv \vec{v}'_i - \vec{a}'$. Finally, the

following cross products are defined:

$$\vec{C}'_i \equiv \epsilon_{ijk}(\vec{d}'_j \times \vec{d}'_k) \quad (2.2)$$

In words, the vectors \vec{C}'_i are the cross products between adjacent vectors \vec{d}'_j , \vec{d}'_k taken in cyclical permutations, as dictated by the Levi-Cevita symbol ϵ_{ijk} . More importantly, however, each vector \vec{C}'_i must be *parallel* to the facet normal vector, which is \hat{e}'_3 in the primed coordinate system, if the point \vec{a} lies *inside* of the polygon. This is because the angles between each vector \vec{d}'_i must add up to 180 degrees inside the polygon or 360 degrees outside by geometric constraint, so the angle between two vectors \vec{d}'_i , \vec{d}'_j has an upper limit of 180 degrees inside of the triangle and 360 degrees outside. Thus, the cross product between these vectors will point one way (here, the positive \hat{e}'_3 direction) if the angle between them is less than 180 degrees or exactly anti-parallel (the $-\hat{e}'_3$ direction) otherwise, so if all \vec{C}'_i are greater than zero, \vec{a} must be inside of the polygon, otherwise, it can immediately determined that \vec{a} is outside. Put concisely, point \vec{a} intersects the polygon given by $\vec{v}_1, \vec{v}_2, \vec{v}_3$ and facet normal \hat{n} if and only if $\vec{C}'_i > 0$ where $i \in [1, 2, 3]$.

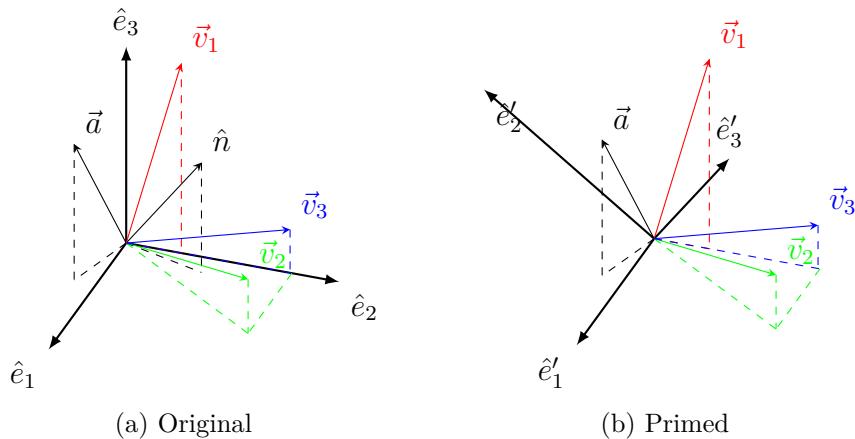


Figure 2.3: Coordinate systems used.

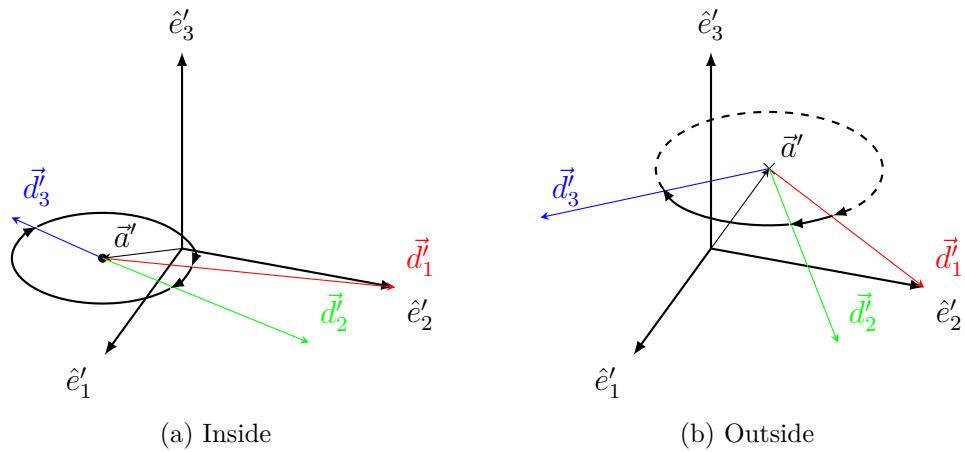


Figure 2.4: Illustration of vectors from hit position to vertices.

2.4.3 Post-processing

The STL representation of the MTD and a text file containing the kinematics of many thousands of simulated particle trajectories is now input to a Python program that operates as follows: for each trajectory in the trajectory file, parse over every polygon in the STL file and check if the trajectory intersects the polygon using the algorithm described previously; if intersection, save some kinematics of the trajectory including the hit position; else, continue. This is, in essence, how many “ray tracing” algorithms function[3], which calculate effects like lighting for computer graphics. Ray tracing is a slow process, which is why graphics seen in movies are pre-rendered and far more detailed and realistic than graphics computed in real time, though recent developments are making real-time, brute-force ray tracing more feasible. However, such technology was neither present nor necessary for this simulation. Instead, the Condor[4] system on the Tier 2 computing center at UC San Diego was used to run the algorithm over approximately 250,000 trajectories. As a result, plots accurate up to the millimeter scale may be produced to study the efficiency of the entire MTD for a dynamic range of geometries. An example of the simulation’s output is shown in Figure 2.5.

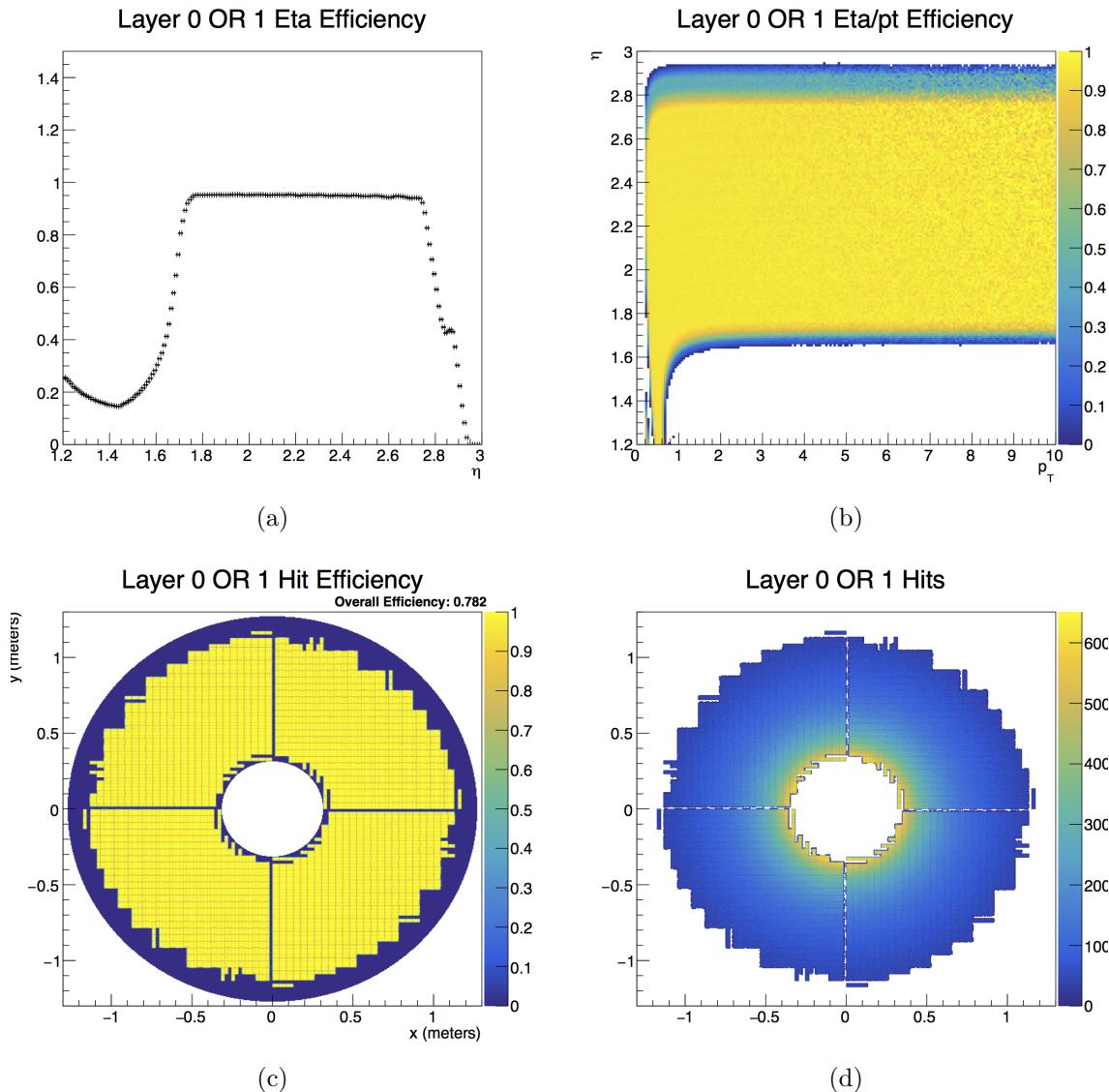


Figure 2.5: Efficiency plots produced using simulation data for the ETL with a 1190mm outer radius and 315mm inner radius following the Topolino design.

Chapter 3

Measurement of Rare Higgs Decays

I begin with a brief discussion of the motivation behind looking for rare Higgs decays like $H \rightarrow \rho\gamma$ and $H \rightarrow \phi\gamma$. In addition, I justify the search via the associated WH production mode rather than gluon-gluon fusion. Then, I outline the event selection methods and relevant backgrounds. Finally, I describe the Boosted Decision tree we used in detail, followed by the results of the analysis.

3.1 Motivation

The Standard Model branching ratios for the $H \rightarrow \phi\gamma$ and $H \rightarrow \rho\gamma$ decays are expected to be 2.31×10^{-6} and 1.68×10^{-5} respectively[5]. Put simply, they are *very* rare - so rare, in fact, that they have never been directly observed. However, some theories[6, 7, 8] suggest that there exist yet-undiscovered processes that enhance the rate of either of these decays. This implies that this measurement is experimentally exciting, because any significant measurement of the production of ρ or ϕ mesons from the decay of a Higgs boson would be direct evidence of the existence of new physics as well as a probe into the coupling between the Higgs boson and light, flavored quarks which remains

unmeasured at the time of writing. At the same time, if no ρ or ϕ mesons are found, then the upper limit of the branching ratio for either decay mode can be pushed farther back, depending on the detector's sensitivity to the signature.

Now, ideally, we would search for these decays through a Higgs boson produced by gluon-gluon fusion (Fig. 3.1) that then decays to a photon and a ϕ or ρ meson. However, CMS does not have a dedicated trigger for detecting pions or kaons (the final-state products of the ϕ and ρ mesons respectively), so the only possibly detectable particle would be the photon. However, since the Higgs mass is 125 GeV, the photon and meson both have a momentum of approximately 60 GeV, so the photon is well below the threshold of the CMS single photon trigger of 200 GeV[9]. Therefore, we chose an associated production mode, namely, WH (Fig. 3.2), where $W \rightarrow \ell^\pm \nu_\ell$. We can easily trigger on the leptons produced by the W boson, and the Higgs comes for free. At the same time, we lose a *lot* of signal, since this process is far more rare due to a lower production cross-section compounded with a lower branching ratio for the specific W to leptons decay signature leading to a loss in cross-section by a factor of approximately 146 (more detail in Section 3.5). This means that we get a large loss of sensitivity compared to a search through gluon-gluon fusion, but alternative associated productions (ZH , for instance) face similar challenges.

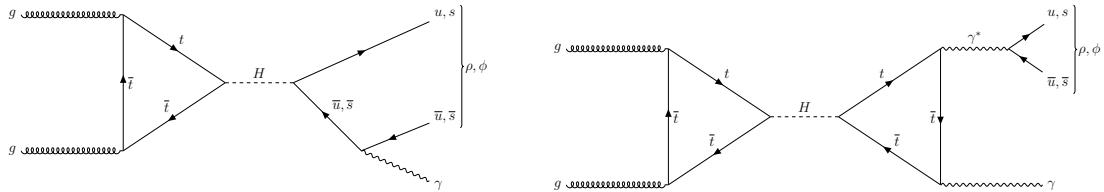


Figure 3.1: Feynman diagram[10] for primary mechanism of the production of the Higgs boson, namely through gluon-gluon fusion, at the LHC where $H \rightarrow \rho\gamma$ or $H \rightarrow \phi\gamma$. (Left) Direct coupling between the Higgs boson and light, flavored quarks. (Right) Higgs decay to light, flavored quarks through top quark loop and virtual photon.

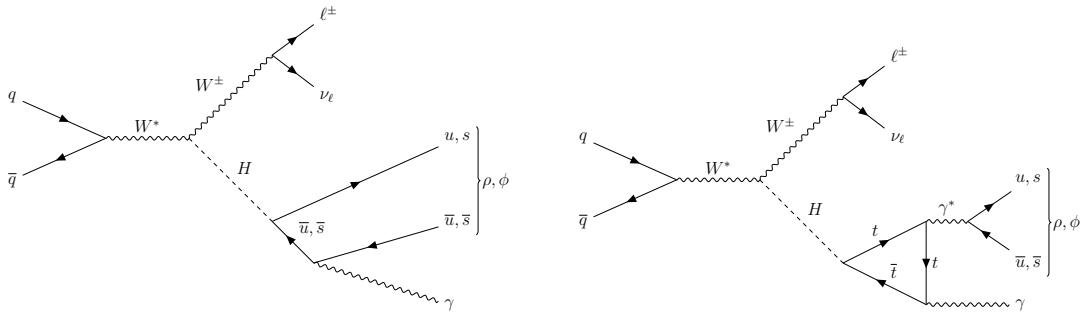


Figure 3.2: Feynman diagram[10] for primary mechanism of the associated WH production at the LHC studied in this analysis. (Left) Direct coupling between the Higgs boson and light, flavored quarks. (Right) Higgs decay to light, flavored quarks through top quark loop and virtual photon.

3.2 Event Selection

3.2.1 Data Aquisition

The analysis begins with data based on a sample of proton-proton collisions collected by the Compact Muon Solenoid (CMS) detector in the LHC. “Interesting” events are selected by the first level of the CMS trigger system which uses information from the detector’s calorimeters and muon detectors to select events for analysis in a fixed time interval of less than $4 \mu\text{s}$. These events are then further processed by a high-level trigger processor farm, which decreases the event rate from around 100 kHz to less than 1 kHz, before the data is stored. Finally, the particle-flow algorithm reconstructs and identifies all particles from the events selected by the CMS trigger system. With the data properly processed and promptly reconstructed “online,” further analysis can be carried out “offline.”

3.2.2 Baseline Selection

To start offline analysis, we first apply a baseline selection on data and Monte Carlo samples to filter out particularly irrelevant events. First, we require at one “good” lepton, qualified as:

- $p_T(\ell^\pm) > 35 \text{ GeV}$
- $|\eta(\ell^\pm)| < 2.4$
- $ID(\ell^\pm) = \text{Medium POG ID}$
- $I_{mini}(e^\pm) < 0.1$
- $I_{mini}(\mu^\pm) < 0.2$

We also require one good photon, qualified as:

- $p_T(\gamma) > 30 \text{ GeV}$
- $|\eta(\gamma)| < 2.5$
- $ID(\gamma) = \text{Medium POG ID}$
- $I_{rel}(\gamma) < 0.06$
- $\Delta R(\gamma, \text{any } e^\pm) > 0.2$

where if two or more good ℓ or γ candidates are found, the candidate with the highest p_T is selected. Additionally, we have a tight second lepton veto to filter out Z events. Finally, we require two, oppositely charged hadrons. This is complicated by the fact that CMS does not distinguish pions and kaons in reconstruction, but we begin by requiring one good hadron (h) candidate pair (h^+, h^-), which are assumed to be the daughters of some unidentified mother meson (M), qualified as:

- $p_T(h^\pm) > 35 \text{ GeV}$
- $\eta(h^\pm) < 2.4$
- h^+, h^- from primary vertex
- $I_{rel}(M) < 0.06$
- $\Delta R(h^+, h^-) < 0.1$

where $I_{rel}(M) = \frac{\max(I(h^+), I(h^-))}{p_T(M)}$. Now, as mentioned previously, the exact identity of these hadrons is ambiguous. They are saved, by default, as pions, so their four-momenta are all constructed using the pion's mass contribution to the energy component. We add these to form the ρ -candidate four-momentum and designate it as the "pion hypothesis." Then, we manually set the energy-components of the hadron four-momenta using the kaon mass, add them to form the ϕ -candidate four-momentum, and designate it as the "kaon hypothesis." If two or more of either hypothesis is found, we select the hypothesis that is closest to the true, respective mass. This leads to the possibility of biasing the data, but we avoid this later by rejecting any events that had more than one meson candidate. Finally, for 2018 data and MC, we exclude any events with e, γ in the HEM region, where a portion of the Hadronic Calorimeter was malfunctioning for a particular period of time.

3.3 Backgrounds

Our final state particle signature is a single isolated photon, lepton, and a reconstructed ρ or ϕ candidate. However, there are a number of processes besides the Higgs decays we are interested in that could present the same signature to the detector.

First, We may get a real, prompt photon in W events where a photon was radiated by one of the initial-state quarks (Fig. 3.2). We might also get a fake photon from misidentified jets in W events. Finally, from Z events, we can get fake photons from misidentified electrons.

At the same time, fake mesons may be produced by any two tracks that happen to both be isolated from all particles other than each other and are generally “meson-like” in that they pass the baseline selections for hadrons described in the previous section.

Now, these background signatures are superficially identical to signal, but they are generally kinematically different. For the fake and real photons from W events, we see that the ΔR between the reconstructed meson and photon (Fig. 3.3(a)) is consistently close to π , meaning they are often back-to-back as we would expect from a photon recoiling against some jet that is being identified as a meson candidate. For signal, we see that the distribution is more varied. For the fake photons from misidentified electrons from Z events, we see that the invariant mass of the lepton-photon system (Fig. 3.3(b)) peaks at the Z mass (approximately 91 GeV[11]) for the Drell Yan sample, which we expect because it models a quark-antiquark pair annihilating to form a Z boson that then decays to leptons. For the other backgrounds and signal, this is obviously not the case, so no obvious peak is evident. Finally, fake mesons can be differentiated in two different ways. First, the ΔR between the two hadrons (Fig. 3.3(c)) is particularly discriminatory in the ϕ analysis, since the kaons from real signal will have large, practically parallel momenta. Last, the invariant mass of the hadron pair (Fig. 3.3(d)) will be close to the true meson mass for signal – again this is particularly true for ϕ due to its narrow mass distribution – whereas background is a more flat distribution with a small peak in the bulk of the true meson mass from real mesons (not from the signal we are interested in). For the analysis, then, we may make cuts on all of the aforementioned quantities to distinguish signal from background. Where exactly we make these cuts is another

question, which address through a machine learning approach discussed in the following section.

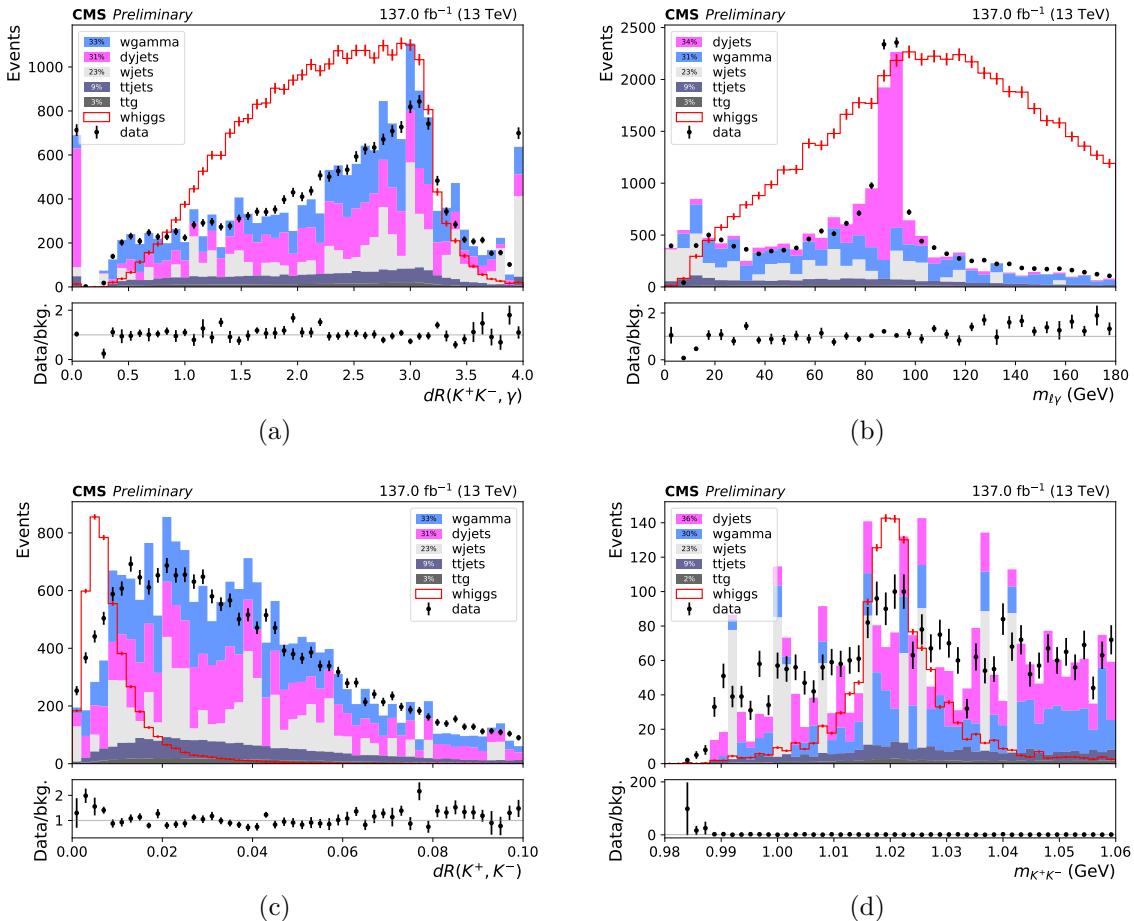


Figure 3.3: Discriminatory quantities for differentiating between signal (red) and background (stacked), where background is scaled to data (black) and signal is scaled to background.

3.4 Boosted Decision Tree

3.4.1 Training

Prior to training, we made the following cuts that prevent over-training the BDT by essentially pre-training it to cut on each respective variable:

- $1.0 < m_{K^+K^-} < 1.04 \text{ GeV}$ or $0.5 < m_{\pi^+\pi^-} < 1.1 \text{ GeV}$
- $80 < m_{\ell\gamma} < 95 \text{ GeV}$

where each was chosen based on the width of their distributions. Then, because BDT's are vulnerable to low statistics, we also added several extraneous, orthogonal datasets that were unnecessary for the greater analysis, but served as useful, accurate background shapes.

We selected the XGBoost python package, a well-known implementation of gradient-boosted decision trees. Twenty features were selected from those saved during the baseline selection step based on their merit as variables that are reasonably uncorrelated to the reconstructed Higgs boson mass. The unweighted distributions for all input features are shown in Fig. A.1 for $H \rightarrow \phi\gamma$ and Fig. A.2 for $H \rightarrow \rho\gamma$, but they are also listed below divided into categories for brevity and later reference:

(i.) Missing Transverse Energy (\cancel{E}_T):

- $p_T(\cancel{E}_T)$
- $\varphi(\cancel{E}_T)$

(ii.) Basic Kinematics:

- $p_T(\gamma), p_T(h^+h^-)$

- $\eta(\gamma), \eta(h^+h^-), \eta(\ell^\pm)$
- $\varphi(\gamma), \varphi(h^+h^-), \varphi(\ell^\pm)$
- $\Delta R(\gamma, \ell^\pm), \Delta R(h^+h^-, \gamma), \Delta R(h^+, h^-)$

(iii.) Masses:

- $m_{\ell\gamma}, m_{h^+h^-}$

(iv.) MELA “Magic” Angles:

- Φ, Φ_1
- $\cos\theta_1, \cos\theta_2, \cos\theta^*$

Now, there are some important clarifications to make here. First, we scaled all p_T quantities that are correlated to the Higgs and input into the BDT by the Higgs mass. Second, φ refers to the azimuthal angle (see Fig. B.1) which is not to be confused with the ϕ meson. Finally, we calculated the quantities listed under (iv.) using MELA[12] (all angles used are illustrated in Fig. 3.4).

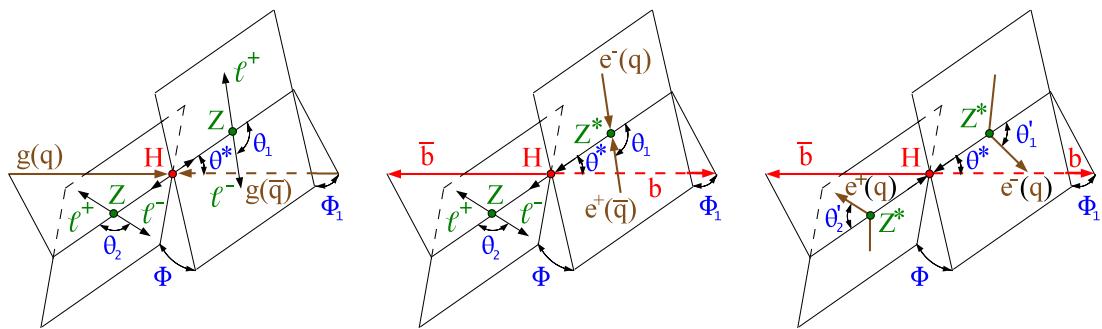


Figure 3.4: Diagram due to Anderson et. al. [13] of Higgs rest frame angles used for BDT training. The center diagram is most relevant under the following replacements: Z, Z^* to W, W^* , b, \bar{b} to $\rho/\phi, \gamma$; ℓ^+, ℓ^- to $e^-/\mu^-, \nu_e/\nu_\mu$.

With these features properly defined, we ran the BDT for 200 training rounds with the following model hyperparameters selected to maximize BDT efficiency without over-training (See Fig. B.2 for hyperparameter definitions):

- `objective` = 'binary:logistic'
- `eta` = 0.1
- `max_depth` = 3
- `verbosity` = 1
- `nthread` = 12
- `eval_metric` = "auc"
- `subsample` = 0.6
- `alpha` = 8.0
- `gamma` = 2.0
- `lambda` = 1.0
- `min_child_weight` = 1.0
- `colsample_bytree` = 1.0

3.4.2 Performance and Validation

We determined satisfactory performance by evaluating the BDT's ROC curves for testing and training (Fig. 3.6 for the ϕ analysis and Fig. A.4 for the ρ analysis) as well as the sanity of the BDT's feature rankings (Fig. 3.5). We also checked for background sculpting by two methods. First, we looked at the plot of the BDT scores versus the

reconstructed Higgs mass (Fig. A.6). A correlation between high BDT scores and the known Higgs mass (125 GeV) would indicate that the BDT was simply learning and cutting on the reconstructed Higgs mass and thus sculpting the background. Second, we directly evaluated the background and signal distributions before and after making a tight cut ($D > 0.9$) on the BDT discriminant (Fig. A.5), where an artificial peak of the background inside of the signal region would directly show that the BDT was sculpting the background. We required that the BDT show no signs of bias by either of these two metrics.

cut	gain	cover	weight	cut	gain	cover	weight
$I_{rel}(K^+K^-)$	64.524145	335.861234	51	$I_{rel}(\pi^+\pi^-)$	484.455590	2568.905639	82
$\Delta R(K^+, K^-)$	32.763722	223.291627	116	$m_{\ell\gamma}$	180.493965	1592.108795	101
$m_{\ell\gamma}$	23.003519	235.263912	84	$\Delta R(\pi^+, \pi^-)$	131.068580	1392.192105	238
$p_T(K^+K^-)$	21.203773	166.753031	117	$m_{\pi^+\pi^-}$	110.852849	1934.547136	107
$\Delta R(K^+K^-, \gamma)$	16.087510	201.926715	84	$p_T(\pi^+\pi^-)$	107.280293	1191.653109	222
$m_{K^+K^-}$	15.973606	283.408161	67	$\Delta R(\pi^+\pi^-, \gamma)$	69.180979	1048.392044	98
$\cos\theta^*$	14.002168	153.376312	25	$\Delta R(\gamma, \ell)$	54.764187	1889.223927	64
$\Delta R(\gamma, \ell)$	13.281328	204.025649	26	$p_T(\gamma)$	51.278799	1535.699391	72
$p_T(\gamma)$	11.414138	134.733902	33	$\cos\theta^*$	49.505357	1429.279040	44
$\cos\theta_1$	7.189837	245.472520	47	$\cos\theta_1$	45.809139	1961.926026	54

Figure 3.5: Top ten variables as ranked by the BDT by gain. All input variables are reconstruction-level Monte Carlo data, and all p_T variables are scaled by $m_{h^+h^-\gamma}$.

3.4.3 Comparison to Cut-Based Approach

Based on the following basic study, we found that a Boosted Decision Tree (BDT) trained on Monte Carlo simulations of signal and background performed 10% to 15% better than traditional, cut-based methods. First, we determined an optimal BDT working point by sampling the BDT’s ROC curve (generated using testing data and predictions) at each defined threshold value, then calculating an expected significance (σ) defined as:

$$\sigma = \sqrt{2(s+b)\ln(1+s/b) - 2s} \quad (3.1)$$

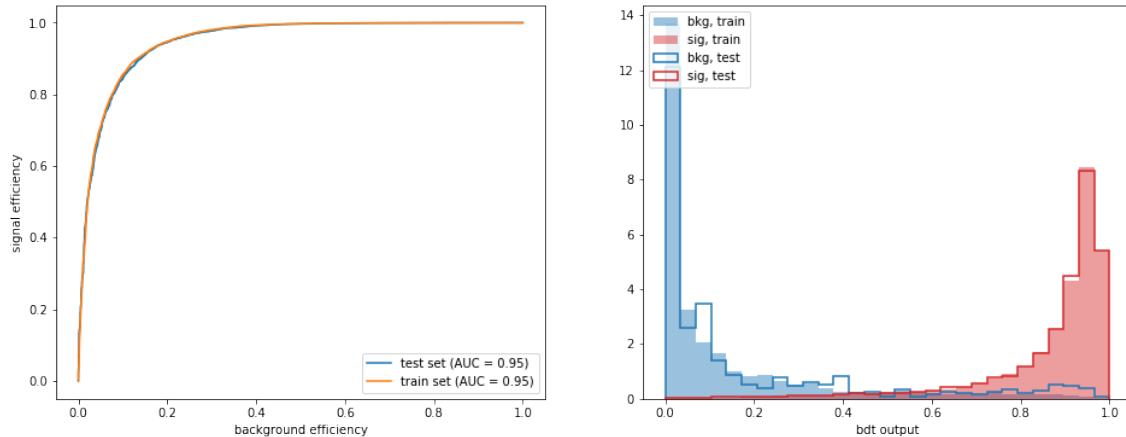


Figure 3.6: Left: ROC curve showing BDT testing (blue) and training (orange) performance. Right: Background (blue) and signal (red) distributions versus BDT score for testing (outline) and training (filled). ($H \rightarrow \phi\gamma$ analysis.)

where s and b are the number of signal and background events for a given threshold value, and the optimal BDT working point is then given by the maximum σ value. Next, we defined a general cut-based approach by the following cuts:

- $1.0 < m_{K^+K^-} < 1.04$ GeV or $0.5 < m_{\pi^+\pi^-} < 1.1$ GeV
- $\Delta R(K^+, K^-) < 0.015$ or $\Delta R(\pi^+, \pi^-) < 0.04$
- $p_T(h^+h^-) > 25$ GeV
- $p_T(\gamma) > 40$ GeV
- $I_{rel}(h^+h^-) < 0.01$

Finally, we calculated the false-positive and true-positive rates of the cut-based methods and plotted them against the BDT's ROC curve. The results of this study are plotted in Fig. A.3. Note that the terms BDT “score,” “threshold,” and “discriminant” have been and will be used interchangeably throughout the remainder of this paper with the understanding that they are the same quantity $D \in [0, 1]$.

3.4.4 Optimization for Data

With the BDT validated, we proceeded to calculate a proper BDT working point for cutting on data. First, we noticed that the Monte Carlo samples were not completely modeling the data in the signal region of the BDT discriminant (Fig. 3.7). Thus, we concluded that we could not sample the BDT testing ROC curve for an effective working point. Instead, we fed the BDT ROC curve data as background in addition to signal Monte Carlo. We then sampled this distribution, calculated expected significance again using Eq. 3.1, and took the best BDT working point to be at the threshold with maximum expected significance (Fig. A.7). This calculation for both the ρ and ϕ analysis resulted in a fairly large cut on the BDT discriminant. In particular, the optimal working point calculated for the ρ analysis blew up towards one as is evident in the right-hand plot in Fig. A.7. As the deadline to complete the analysis grew closer, we decided to simply choose a more sensible point on the graph for the ρ analysis, namely $D = 0.83$.

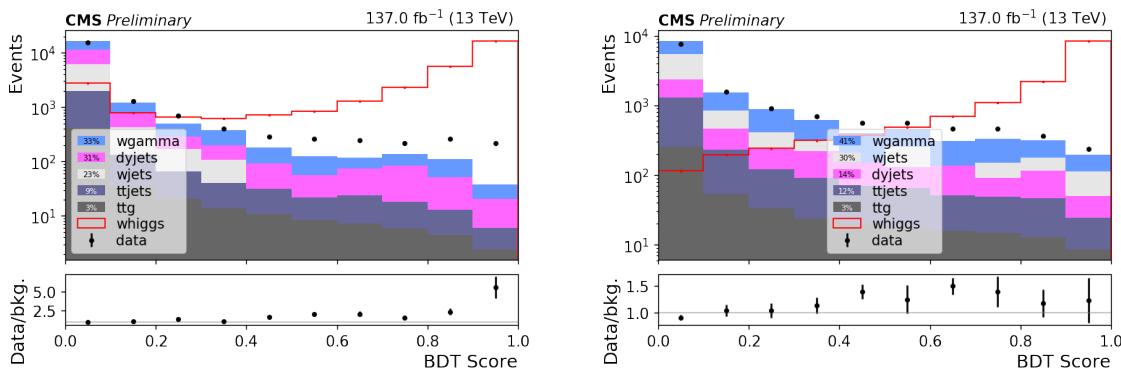


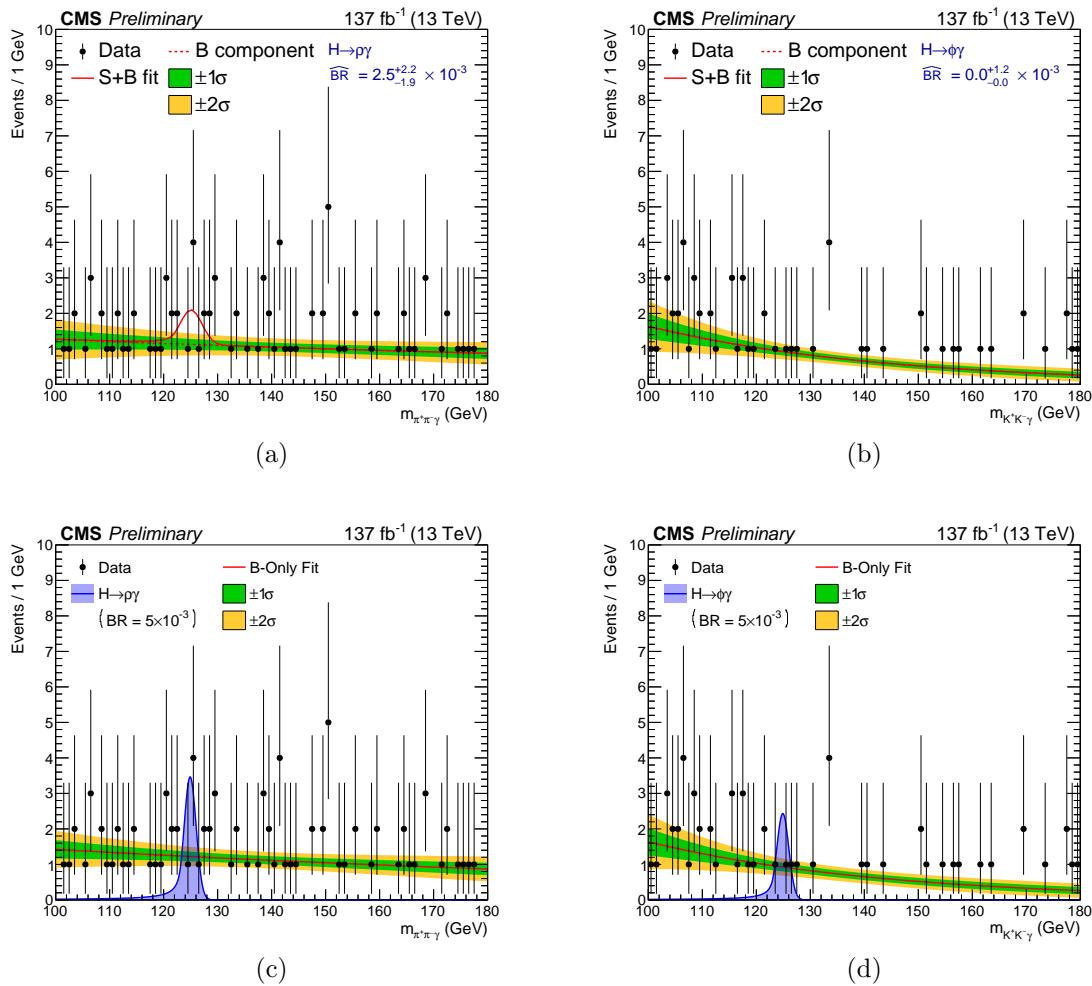
Figure 3.7: Stacked BDT discriminant distribution plot by sample. (Left) $H \rightarrow \phi\gamma$ analysis. (Right) $H \rightarrow \rho\gamma$ analysis.

3.5 Results

After cutting on the best BDT working point described in the previous section, we fit the background to an exponential function and the signal to a single Crystal Ball function. Then, we used the HiggsCombine tool’s asymptotic limits calculations[14] to generate the plots in Fig. 3.9 as well as the exclusion limits listed in Fig. 3.8 (at the 95% confidence level) taking a 20% uncertainty – a rough, but conservative estimate made because we did not have time to perform detailed studies of the uncertainty – on the signal acceptance. We see that our result is within one order of magnitude of that produced by the ATLAS Collaboration in their study[15] of these same decays, but through gluon-gluon fusion production using a dedicated trigger. However, a weaker result was expected, as briefly prefaced in Section 3.1. Although we had approximately four times the amount of data from the full Run II 137 fb^{-1} dataset, the much smaller cross-section of the associated WH production compared to gluon-gluon fusion already results in a reduction in cross-section by a factor of about 32 (cross-sections for both processes were sourced from this TWiki[16]). On top of that, the branching ratio[17] for $W \rightarrow e^\pm/\mu^\pm + \nu_{e,\mu}$ is approximately 22%, reducing the effective WH cross-section, which gives an overall reduction in cross-section by a factor of 146. Therefore, we conclude that our results are reasonable within the limits of the sensitivity of the CMS detector to the $H \rightarrow \rho\gamma$ and $H \rightarrow \phi\gamma$ decays through the associated WH production.

Branching Ratio	Expected	Observed
$\mathcal{B}(H \rightarrow \phi\gamma)$	$(5.2^{+3.0}_{-1.8}) \times 10^{-3}$	4.1×10^{-3}
$\mathcal{B}(H \rightarrow \rho\gamma)$	$(4.1^{+2.3}_{-1.3}) \times 10^{-3}$	7.0×10^{-3}

Figure 3.8: Exclusion limits for the two analyses covered by this paper.

Figure 3.9: Final fits for the $H \rightarrow \rho\gamma$ (left) and $H \rightarrow \phi\gamma$ (right) analyses.

Appendix A

Additional Figures

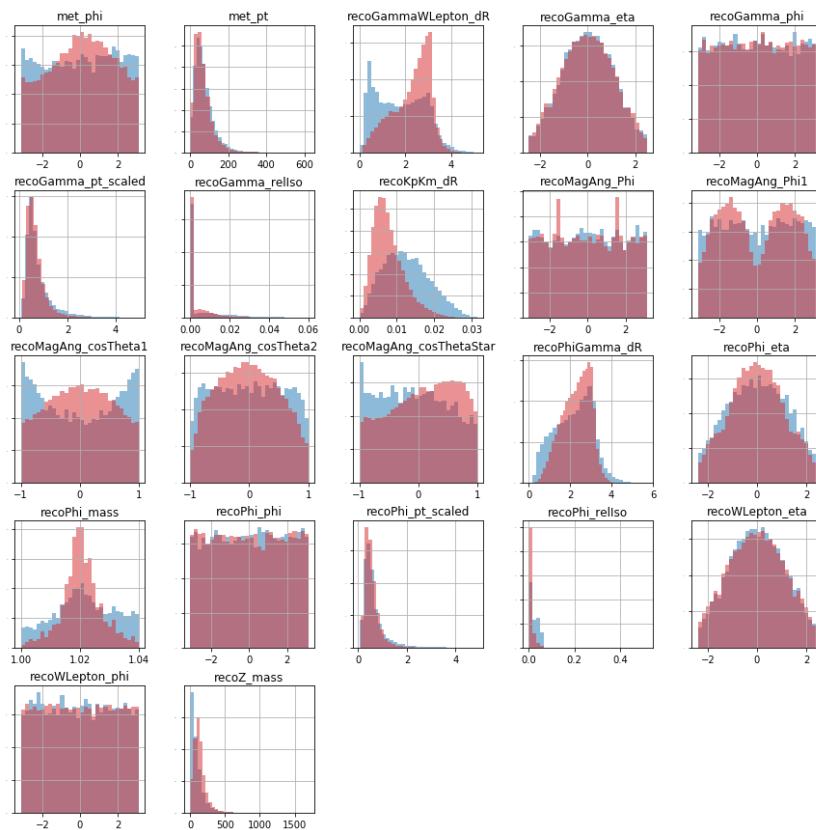


Figure A.1: Preliminary histograms for every feature used to train the BDT. These were drawn before any weights are applied. ($H \rightarrow \phi\gamma$ analysis.)

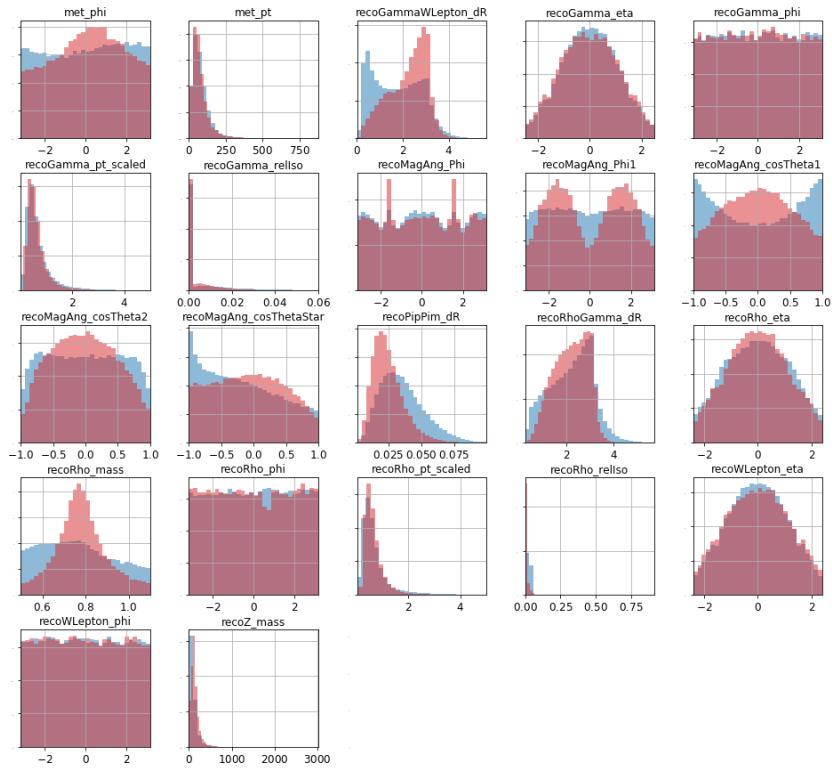


Figure A.2: Preliminary histograms for every feature used to train the BDT. These were drawn before any weights are applied. ($H \rightarrow \rho\gamma$ analysis.)

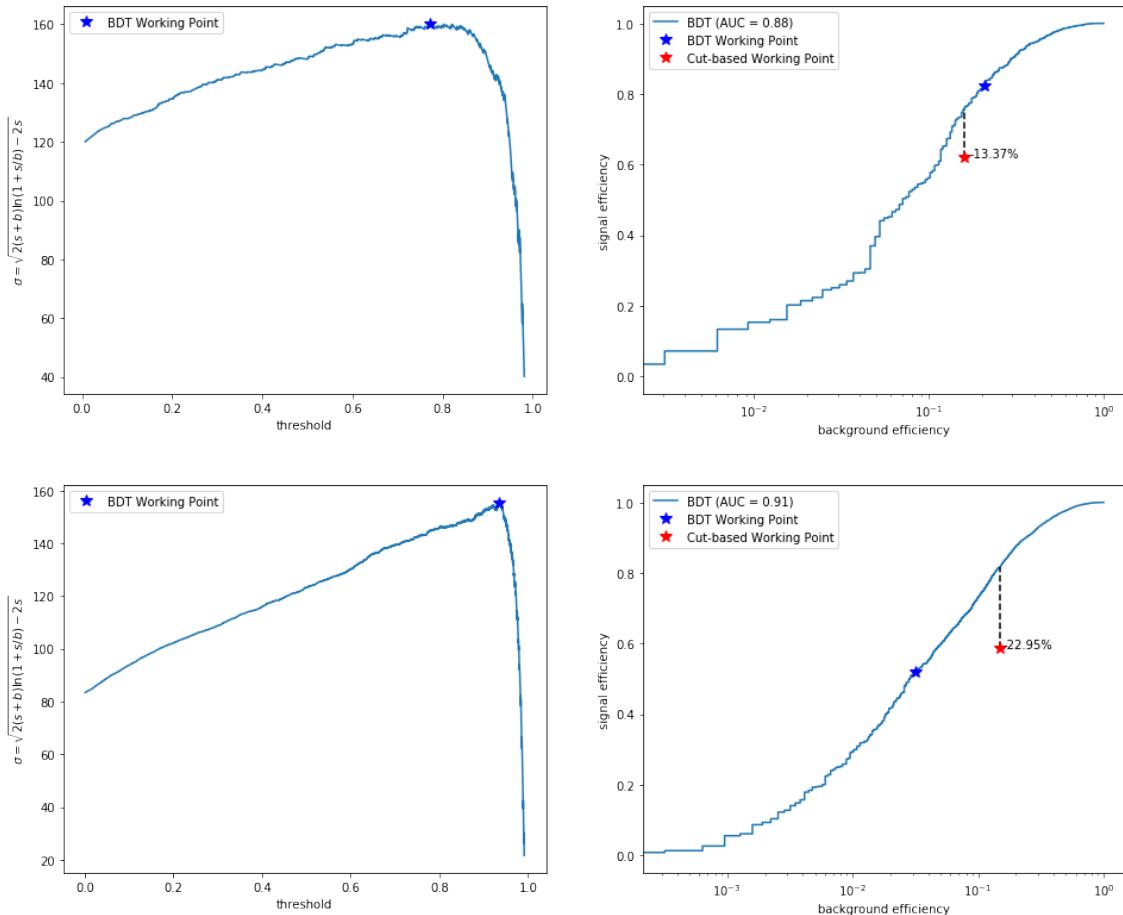


Figure A.3: (Left) Expected significance (σ) versus BDT threshold value. (Right) Comparison between an optimal BDT working point and cut-based method plotted on top of the BDT's ROC curve. (Top) $H \rightarrow \phi\gamma$. (Bottom) $H \rightarrow \rho\gamma$.

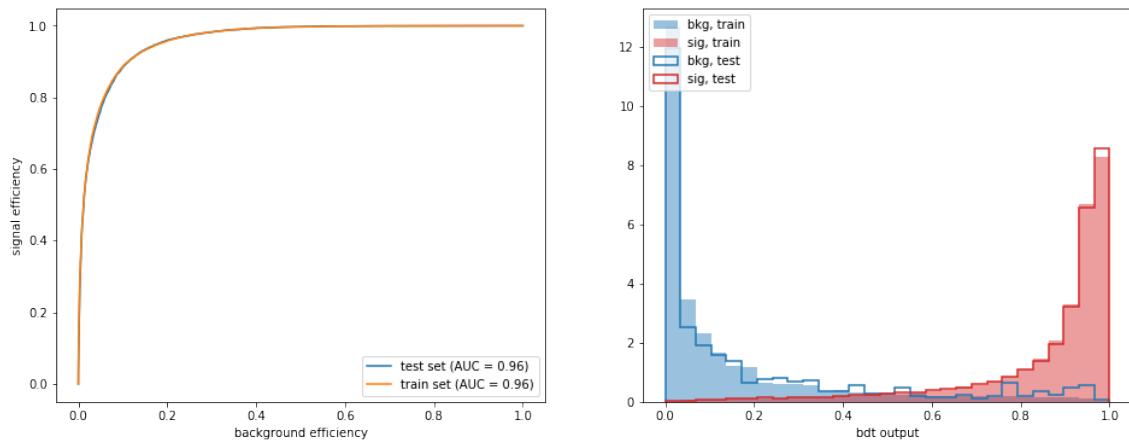


Figure A.4: Left: ROC curve showing BDT testing (blue) and training (orange) performance. Right: Background (blue) and signal (red) distributions versus BDT score for testing (outline) and training (filled). ($H \rightarrow \rho\gamma$ analysis.)

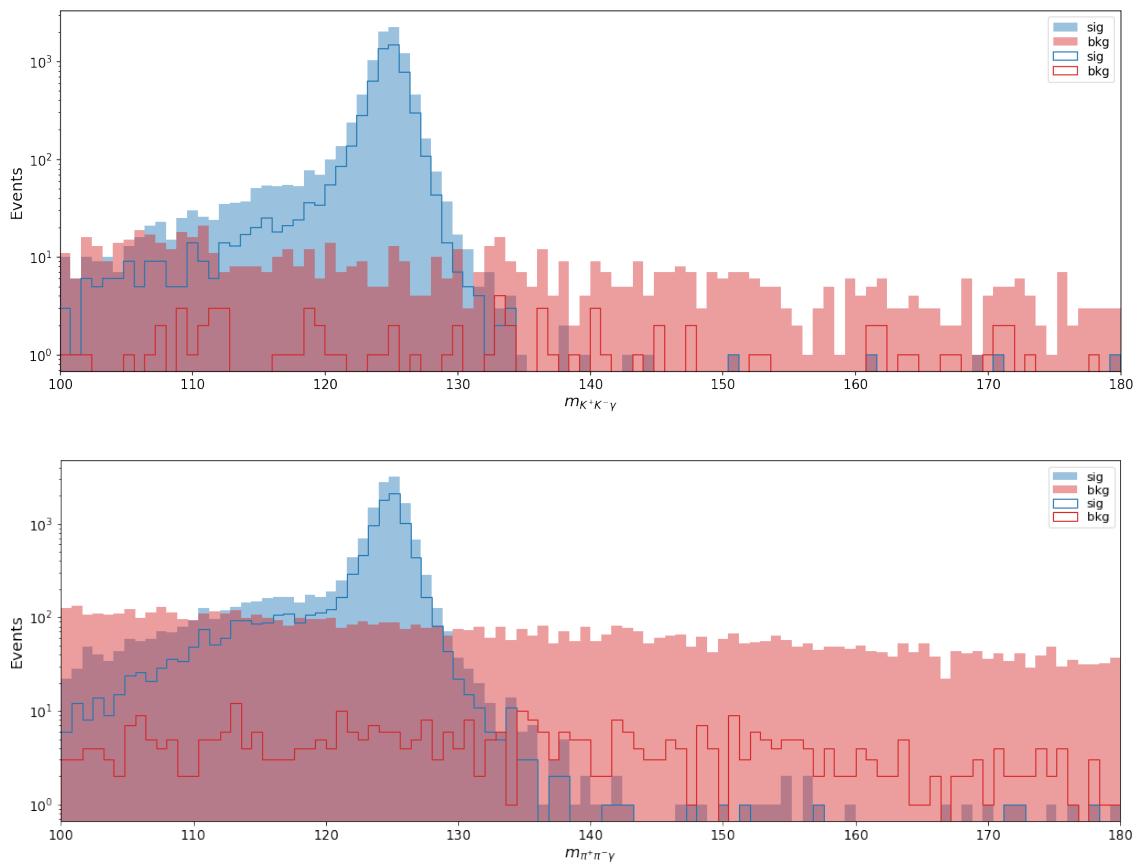


Figure A.5: Reconstructed Higgs mass distribution for signal (blue) and background (red) before (filled) and after (outline) requiring $D > 0.9$. (Top) $H \rightarrow \phi\gamma$. (Bottom) $H \rightarrow \rho\gamma$.

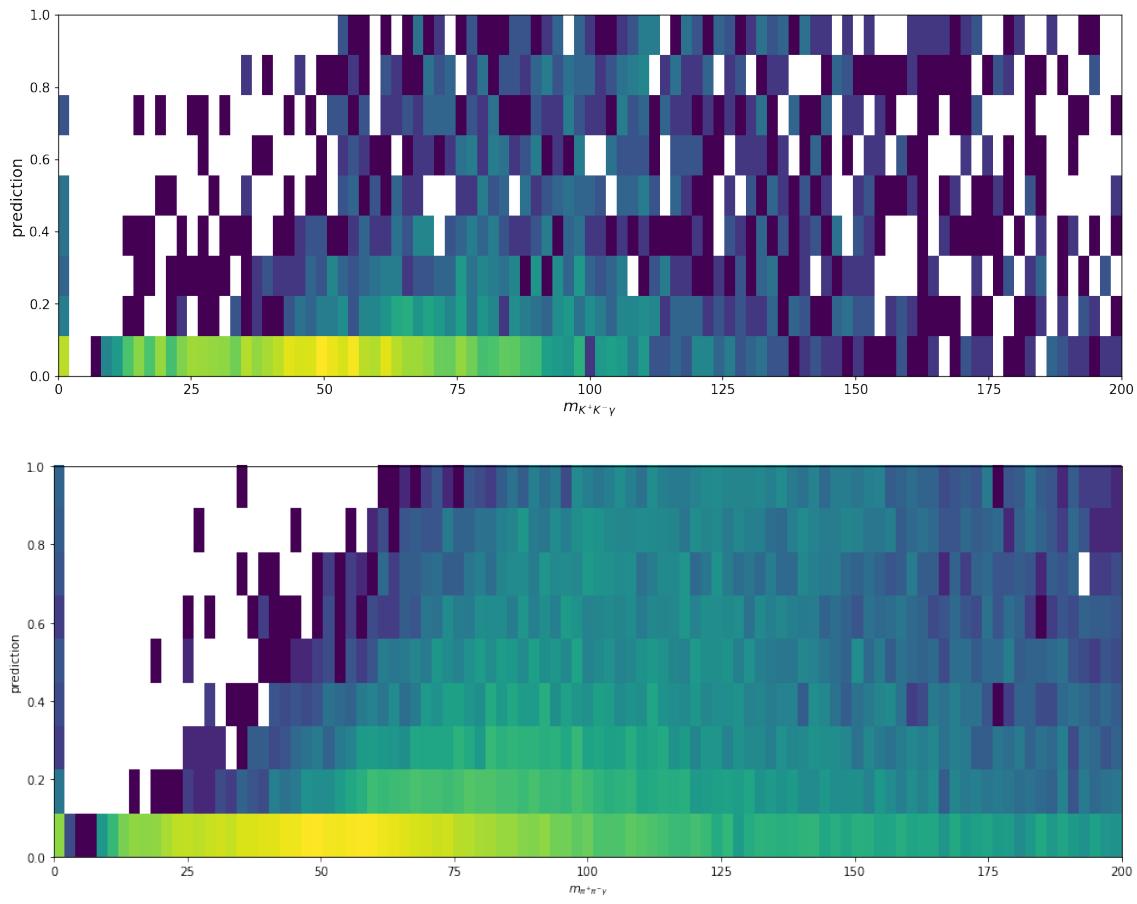


Figure A.6: Plot of the BDT score versus the reconstructed Higgs mass. A heavy correlation between high score and the true Higgs mass would suggest background is being sculpted. (Top) $H \rightarrow \phi \gamma$. (Bottom) $H \rightarrow \rho \gamma$.

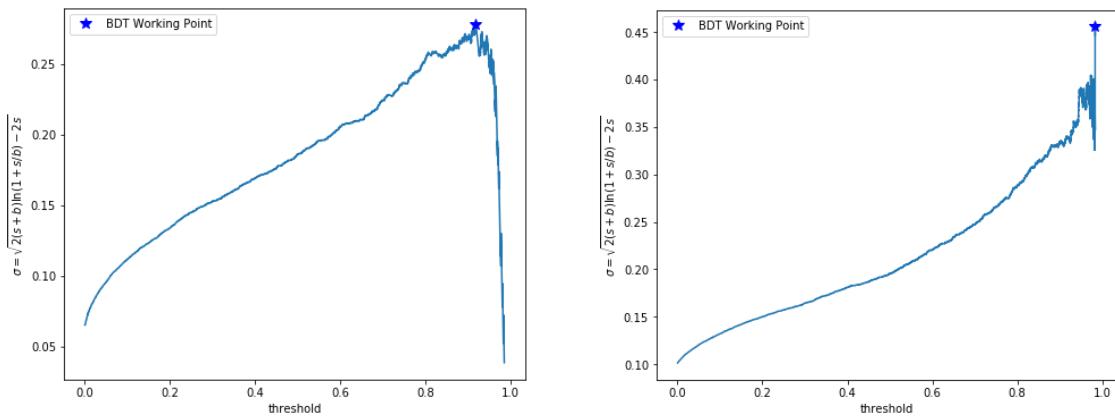


Figure A.7: Expected significance (σ) versus BDT threshold. (Left) $H \rightarrow \phi + \gamma$ analysis. (Right) $H \rightarrow \rho + \gamma$ analysis.

Appendix B

Technical Details

B.1 CMS Coordinate System

In the CMS Coordinate system, the z -axis is aligned along the beampipe, therefore running parallel to the barrel and perpendicular to the endcap of the detector. The x -axis points radially towards the center of the LHC, and the y -axis points orthogonal to x and z . The azimuthal angle φ is defined about the z -axis, as in cylindrical coordinates, and the pseudo-rapidity η is defined as

$$\eta \equiv -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (\text{B.1})$$

where θ is the angle between the particle three-momentum and positive z -axis. See Figure B.1 for a visualization of this coordinate system.

B.2 Kinematic Variable Definitions

The following quantities are defined using the CMS Coordinate System defined in Appendix A.1. First, it is useful to define “transverse” components (denoted as A_T for a

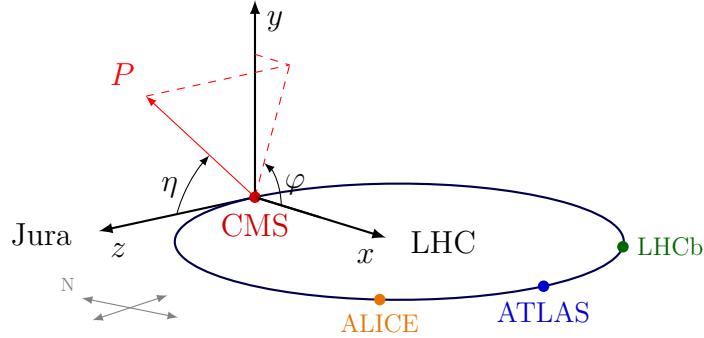


Figure B.1: Diagram of CMS coordinate system.

three or four-vector quantity A) to be the spatial x and y -components of a given variable. Additionally, ΔR between two particles x_1, x_2 is defined as

$$\Delta R(x_1, x_2) = \sqrt{\Delta\varphi(x_1, x_2)^2 + \Delta\eta(x_1, x_2)^2} \quad (\text{B.2})$$

where $\Delta\eta(x_1, x_2) = \eta(x_2) - \eta(x_1)$, and $\Delta\varphi(x_1, x_2) = |\varphi(x_2) - \varphi(x_1)|$ is the *inner* angle between the x_1 and x_2 three-momenta), such that $\Delta\varphi \in [0, 2\pi]$. Then, the isolation I of a particle x is defined as the sum of transverse momentum p_T in a cone of $\Delta R < 0.4$ [18] around that particle. Finally, the “relative” isolation I_{rel} of a particle x is defined as:

$$I_{rel}(x) = \frac{I(x)}{p_T(x)} \quad (\text{B.3})$$

B.3 Boosted Decision Trees

A Boosted Decision Tree (BDT) is simply an extension of a decision tree, which is can be visualized as follows: starting from a single “node,” two “branches” emerge; at the end of each branch, there is another node that also joins two branches, and so on. In this

picture, each node is a binary decision (i.e. $A > B$, $C == D$, etc.) and the two branches joined by each node represent the path “traveled” by data should it pass or fail – hence *two* branches – the condition. The tree may be optimized by recursively adding nodes and adjusting the “splits” (the exact value over which the input is split) at each node until some given condition is met such that an optimal configuration is reached. This is essentially an accurate representation of the classical, “cut-based” analysis that has been performed in High Energy for decades, where physicists filter out everything they are not looking for (background) with the hope that what they are looking for (signal) remains after the surrounding noise is cleared. A similar (both in practice and results) optimization of this technique is achieved by a combination of educated insight and trial and error. Thus, the computational method of the decision tree fits into the workflow of a High Energy physics analysis, but a decision tree is not an exceptionally powerful classifier on its own. However, they can be “boosted” by creating several decision trees, collecting their output, and calculating a value called a “determinant” (henceforth referred to as D), forming a stronger classifier from the collection of weak ones. The exact process of constructing and optimizing a BDT is dependent on the particular algorithm used and is beyond the scope of this paper.

Variable	Definition
<code>objective</code>	Learning objective. ('binary:logistic': logistic regression for binary classification, output probability)
<code>eta</code>	Step size shrinkage used in update to prevents overfitting. (range: [0, 1])
<code>max_depth</code>	Maximum depth of a tree. Increasing this value will make the model more complex and more likely to overfit. (range: [0, ∞])
<code>verbosity</code>	Verbosity of printing messages. Valid values are 0 (silent), 1 (warning), 2 (info), 3 (debug).
<code>nthread</code>	Number of parallel threads used to run XGBoost.
<code>eval_metric</code>	Evaluation metrics for validation data. ('auc': Area under the curve)
<code>subsample</code>	Subsample ratio of the training instances. (range: (0, 1])
<code>alpha</code>	L1 regularization term on weights. Increasing this value will make model more conservative.
<code>gamma</code>	Minimum loss reduction required to make a further partition on a leaf node of the tree. The larger <code>gamma</code> is, the more conservative the algorithm will be. (range: [0, ∞])
<code>lambda</code>	L2 regularization term on weights. Increasing this value will make model more conservative. Normalized to number of training examples.
<code>min_child_weight</code>	Minimum sum of instance weight (hessian) needed in a child. The larger <code>min_child_weight</code> is, the more conservative the algorithm will be.
<code>colsample_bytree</code>	Subsample ratio of columns when constructing each tree. Subsampling occurs once for every tree constructed.

Figure B.2: XGBoost BDT hyperparameters[19].

Bibliography

- [1] CMS Collaboration, *CMS, the Compact Muon Solenoid: Technical proposal*, .
- [2] <http://stuart.physics.ucsb.edu/tdr.pdf>. Accessed: 2019-05-26.
- [3] P. Rademacher, *Ray tracing: graphics for the masses*, *Crossroads* **3** (1997), no. 4 3–7.
- [4] D. Thain, T. Tannenbaum, and M. Livny, *Distributed computing in practice: the condor experience.*, *Concurrency - Practice and Experience* **17** (2005), no. 2-4 323–356.
- [5] M. König and M. Neubert, *Exclusive radiative higgs decays as probes of light-quark yukawa couplings*, *Journal of High Energy Physics* **2015** (2015), no. 8 1–38.
- [6] M. König and M. Neubert, *Exclusive radiative higgs decays as probes of light-quark yukawa couplings*, *Journal of High Energy Physics* **2015** (2015), no. 8 1–38.
- [7] A. L. Kagan, G. Perez, F. Petriello, Y. Soreq, S. Stoynev, and J. Zupan, *Exclusive window onto higgs yukawa couplings*, *Physical review letters* **114** (2015), no. 10.
- [8] G. Perez, Y. Soreq, E. Stamou, and K. Tobioka, *Prospects for measuring the higgs boson coupling to light quarks*, *Physical Review D* **93** (2016), no. 1.
- [9] CMS Collaboration, *The cms high level trigger*, *The European Physical Journal C - Particles and Fields* **46** (2006), no. 3 605–667.
- [10] J. P. Ellis, *Tikz-feynman: Feynman diagrams with tikz*, *Computer Physics Communications* **210** (2017), no. C 103–123.
- [11] LEP Collaboration, *Precise determination of the z resonance parameters at lep: "zedometry"*, *European Physical Journal C* (2001).
- [12] <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MELAProject>. Accessed: 2019-05-06.

- [13] I. Anderson, S. Bolognesi, F. Caola, Y. Gao, A. V. Gritsan, C. B. Martin, K. Melnikov, M. Schulze, N. V. Tran, A. Whitbeck, and Y. Zhou, *Constraining anomalous hvv interactions at proton and lepton colliders*, *Physical Review D* **89** (2014), no. 3.
- [14] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *The European Physical Journal C* **71** (2011), no. 2 1–19.
- [15] ATLAS Collaboration, *Search for exclusive higgs and z boson decays to and with the atlas detector*, *Journal of High Energy Physics* **2018** (2018), no. 7 1–37.
- [16] <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV2014>. Accessed: 2019-05-29.
- [17] LEP Collaboration, *Measurement of the $e + e^- \rightarrow w + w^-$ cross section and w decay branching fractions at lep*, *The European Physical Journal C* **52** (2007), no. 4 767–785.
- [18] https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideEgammaIsolation#Isolation_Algorithms_by_Subdetec. Accessed: 2019-05-26.
- [19] <https://xgboost.readthedocs.io/en/latest/parameter.html>. Accessed: 2019-04-30.