ABSTRACT

A SEARCH FOR RESONANT AND NON-RESONANT DI-HIGGS PRODUCTION IN THE $\gamma\gamma b\bar{b}$ CHANNEL USING THE ATLAS DETECTOR

Tyler James Burch, Ph.D.
Department of Physics
Northern Illinois University, 2020
Jahred Adelman, Director

This dissertation presents a search for resonant and non-resonant di-Higgs production in the $\gamma\gamma b\bar{b}$ final state using data from the ATLAS detector at the Large Hadron Collider (LHC). The search is performed on 36.1 fb^{-1} of data from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV collected in 2015 and 2016.

No significant excesses are observed in this search. The non-resonant analysis sets limits on the $HH \to \gamma\gamma b\bar{b}$ cross-section times branching ratio, with an upper observed (expected) limit of 0.73 (0.93) pb. The observed (expected) limits on the Higgs boson trilinear coupling at 95% Confidence Level (CL) are set at $-8.2 < \kappa_\lambda < 13.3$ ($-8.5 < \kappa_\lambda < 13.7$). A model-independent resonant search is also presented, setting limits on a generic scalar resonance under the narrow-width approximation. These limits cover mass hypotheses ranging from 260 GeV to 1000 GeV, and the observed (expected) limits set are 0.85 (0.92) pb at the lowest mass hypothesis to 0.13 (0.15) pb at the highest mass hypothesis.

Work toward the future of this analysis is presented. Improvements in photon identification are studied, investigating optimization through two approaches. First, through adding the moments of topological clusters as inputs, which show additional discriminating power.

Second, through using a multivariate approach to define photon identification, studying a Boosted Decision Tree (BDT) and a Neural Network (NN). Through these additional inputs and the employment of a BDT, an improvement of as much as 27% background rejection for the same signal efficiency as the current tight working point is shown.

Additionally, improvements to the analysis through studying the Vector Boson Fusion (VBF) production mode are shown. This provides handles on new couplings, and a dedicated signal region targeting this mode can improve overall Asimov significance. To define this signal region, a multiclass BDT is used, with classes to model the VBF HH production mode, along with gluon-gluon fusion HH production, as well as the dominant $\gamma\gamma$-continuum background, and ttH mono-Higgs background. By adding this signal region, a 9.7% improvement in Asimov significance is achieved using the full 140 fb^{-1} of Run 2 data.

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

MARCH 26 2020

A SEARCH FOR RESONANT AND NON-RESONANT DI-HIGGS PRODUCTION IN THE $\gamma\gamma b\bar{b}$ CHANNEL USING THE ATLAS DETECTOR

BY

TYLER JAMES BURCH

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

Dissertation Director:
Jahred Adelman

ACKNOWLEDGEMENTS

This work represents an amalgamation of six years of my life. Looking back on those six years, I've fundamentally changed in so many ways. I experienced memories that will last a lifetime, and tough times that I could not have made it through without a network of friends and family behind me. Thank you all for your guidance, your friendship, and your support.

First, I'd like to thank my advisor, Jahred Adelman. Your guidance has made me into the scientist I am today. I walked into this knowing very little about research, and your patience and wisdom shaped how I approach problems and my understanding of physics. I cannot thank you enough for the time you've poured into me and the places you've led me.

Next I'd like to thank Liza Brost, your mentorship was absolutely vital to my growth. The time you invested into me and care toward my success led me to be able to reach this point, and I cannot thank you enough for that. Beyond research, I could not have thrived at CERN without you, and I'm incredibly thankful for the support and guidance you provided in that chapter of my life.

Countless other scientists who have contributed to my growth and development. Specifically, I'd like to thank Jeremy Love, for the considerable guidance since the start of my time on ATLAS. I want to thank Walter Hopkins, for the countless conversations about physics and machine learning on the way to the coffee machine, and knowledge and insight that brought me. Beyond that, I'd like to thank the many other scientists I've been able to look up to along the way, for your mentoring and for the example you provided: Max Swiatlowski, Jonathan Long, Tova Holmes, Kate Whalen, Christopher Meyer, amongst many others.

I'd like to thank my friends who supported me during my time at CERN: John Myers, Lesya Horyn, Todd Seiss, Rob Mina, Elliot Parrish, Bingxuan Liu. Moving to the other side of the world was absolutely daunting, but you all made it an experience I won't ever forget. Also, thank you to Luca Buonocore, Francesca Forte, Roberto Cevenini, Gianmarco Pompeo, and Francesco Riccardi. When I moved into our apartment in Saint-Genis, I knew I'd found a place to live, but I didn't realize the Italian family I would inherit while there.

Pawel Klimek, Blake Burghgrave, Puja Saha, Brianna Stamas, Louis D'Eramo, and Dhiman Chakraborty thank you for contributing your thoughts, ideas, and input to my work. Also, thank you to all of the professors at NIU who taught me in a class or influenced me along the way, specifically Michael Eads and Stephen Martin. I could not have survived through classes, finals, and candidacy exams without the support of many other NIU graduate students, especially Mike McEvoy, Alex Halavanau, Marc Pavlik, Dan Faia, Andrew Fiedler, and Daniel Stange.

To the friends who have seen me grow up, who have been there from the beginning, and who have stuck through life with me, thank you. Aaron and Erin Spargur, Todd and Alex Grumley, Jacob Kreeb, Michael Hoelscher, and Lauren Ruser, being around you all has always felt like home, and having that throughout this chapter of my life was essential. Grant Knox, thank you for always being someone I could talk to about anything. Additionally to the many other friends who have spent time with me throughout the course of this degree, thank you, I could not have made it without the mental rest your friendship provided: John Mueth, Brendan Nelson, Joe Rossi, Naaim Siddiqi, and many others.

I'd like to thank Mike Rogier, the first instructor who made me excited about physics. All journeys have a start, and this is one that started in your classroom. I'm thankful for the spirit you brought to teaching, and to the confidence you showed in me.

Last, thank you to my family. Getting to this point has been an adventure. One that has taken me far and wide, and one that has taken me through highs and lows. Dad and Mom, thank you for being there with me through it all. Thank you for developing me into the man I am today. Colin, Aralynn, and Owen – one of the toughest sacrifices of this degree

has been having to watch you grow up from afar. I'm excited to see who you all grow up to be and hope that, in some shape, this inspires you to pursue your passions.

The beautiful thing about learning is nobody can take it away from you.
— B.B. King

DEDICATION

For Robert J. Burch and Jim L. Richardson, for the example you both set for me in work and in love.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure | Page

Figure Page

Figure Page

LIST OF APPENDICES

CHAPTER 1

THEORY

It's nice to be right sometimes.
— Peter Higgs

Humans have been concerned with the fundamental nature and constituents of matter since as early as eighth century BCE, when ancient Indian philosophers, proposed that the objects we experience are collections of particles too small to be seen [1]. In sixth century BCE, Greek philosophers such as Democritus and Leucippus proposed indivisible building blocks of nature known as "atoms." It wasn't until Dalton in the nineteenth century when scientific modeling of the atomic theory began to take shape [2]. Fundamental advances on the nature of atoms and their substructure were made over the next centuries by Mendeleev, Thomson, and Rutherford, among others. Over the course of the latter 1900s, particle theory advanced rapidly and many particles were discovered, forming what was known as the "particle zoo." In order to classify these particles, Gell-Mann and Zweig proposed the quark model in 1964 [3]. In the 1970s Glashow, Weinberg, and Salam put forth the electroweak theory, which gave a model for vector bosons, leading to successful mass predictions of the $W^{\pm}$ and Z bosons confirmed at the European Organization for Nuclear Research (CERN) in 1983 [4, 5]. The collective of these works is known as the Standard Model (SM) of particle physics, a term first used in 1975 [6]. It remains our best model for describing fundamental particles and their interactions[1], successfully predicting the top quark [8], tau neutrino [9], and Higgs boson [10, 11].

1.1 The Standard Model

The SM of particle physics aims to provide a description of the known fundamental particles and their interactions. A fundamental particle is one which cannot be further divided into smaller particles, thus having no substructure. These can be separated into two classes, fermions and bosons. Fermions are the particles which take up space and make up matter. In Section 1.1.1, the two categories of fermions, quarks and leptons, will be discussed. By contrast, interactions between particles are due to exchange of a gauge boson, which are discussed throughout Section 1.1.2.

The SM is a Yang-Mills theory [12], a gauge theory based on $SU(3)$ symmetry. Gauge theories are field theories in which the Lagrangian is invariant under local transformations (known as gauge invariance). The group of these transformations is known the symmetry group of the theory. Further, Yang-Mills theories are non-abelian gauge theories, meaning their symmetry group is non-commutative. Noether's theorem [13] states that every symmetry of nature implies a corresponding conservation law, and the invariance of many gauges of the SM point toward such conservation laws. In Section 1.1.2, the interactions of the SM are shown, each with a gauge invariance and associated force-carrying boson.

[1] While the SM has been extraordinarily successful in providing experimental predictions, it has several known shortcomings. Notably, the SM only explains 3 of the 4 fundamental forces, the exception being gravitation. It does not have an explanation for dark energy or a dark matter candidate. Neutrino flavor oscillation is not predicted by the SM, and the nature and origin of neutrino masses remains an open question [7]. The matter-antimatter asymmetry is also not addressed by the SM. Nevertheless, it represents our best collective understanding of particle physics.

1.1.1 Fermions

Fermions are half-integer spin particles which obey the Pauli Exclusion principle, and thus take up space and make up matter. In the SM, there are two types of fermions: quarks and leptons. Each of these can be broken down into three generations of particles, shown in Table 1.1.

Quarks

Quarks are fermions which combine to form hadrons. Quarks have several properties:

- Flavor - There are 6 flavors of quarks: up (u), down (d), strange (s), charm (c), bottom (b), and top (t) (from lightest to heaviest). Each quark also has a corresponding anti-quark: anti-up ($\bar{u}$), anti-down ($\bar{d}$), etc. These antiquarks have the same mass, lifetime, and spin as the corresponding quark, but have the opposite sign for their various charge values.

- Electric Charge - Quarks either have electric charge of $\frac{2}{3}$ (u, c, t) or $-\frac{1}{3}$ (d, s, b) in units of electron charge, while anti-quarks have the opposite charge. The electric charge of bound states formed by quarks have integer charge, the sum of the constituent quarks.

- Color Charge - Each quark has a degree of freedom known as *color charge*[2], which can be "red," "blue," or "green," as well as a corresponding anti-color. Color charge is a product of Quantum Chromodynamics (QCD), discussed in section 1.1.2.2, which has a distinct property that particles with color charge cannot be isolated, and must form bound states. This feature is known as *color confinement*. Bound states require a colorless configuration, which may be achieved either through a color-anticolor combination known as mesons, or an (anti)-RGB triplet known as (anti)-baryons.

[2]Quark color is a tool for describing a degree of freedom and has no connection to visible color.

Leptons

The remaining fermions are leptons, which differ from quarks in that they do not have QCD color. Each lepton generation contains one charged particle and a neutral counterpart, known as a neutrino. The charged leptons are the electron, muon, and tau (from lightest to heaviest). Table 1.1 also lists the leptons and neutrinos by generation as well as their masses. Each lepton has an associated antilepton as well: the positron, antimuon, and antitau, as well as their antineutrinos. The properties of the antileptons match their corresponding lepton exactly, except the electric charge and lepton number have opposite sign. Lepton number is a conserved quantity in particle interactions and is +1 for all standard leptons, -1 for all antileptons.

Table 1.1: Fermions in the SM, with their masses and charges listed [14].

	1st Generation	2nd Generation	3rd Generation	Charge
Quarks	Up (u) $m = 2.3^{+0.7}_{-0.5}$ MeV	Charm (c) $m = 1.275 \pm 0.025$ GeV	Top (t) $m = 173.2 \pm 0.7$ GeV	+2/3
	Down (d) $m = 4.8^{+0.5}_{-0.3}$ MeV	Strange (s) $m = 95 \pm 5$ MeV	Bottom (b) $m = 4.18 \pm 0.03$ GeV	-1/3
Leptons	Electron (e^-) $m = 511$ keV	Muon (μ^-) $m = 105.7$ MeV	Tau (τ^-) $m = 1.8$ GeV	-1
	Electron Neutrino (ν_e) $m < 2$ eV	Muon Neutrino (ν_μ) $m < 0.17$ MeV	Tau Neutrino (ν_τ) $m < 15.5$ MeV	0

Generally, the wave function of a fermion is given by $\psi(x)$, and is a spinor of four components. The motion is described by the relativistic version of the Schrödinger equation known as the Dirac equation, given as

$$(i\gamma'^{\mu}\partial_\mu - m)\Psi' = 0, \tag{1.1}$$

and interactions are given by the Dirac Lagrangian,

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi, \tag{1.2}$$

for a particle of mass m.

1.1.2 Bosons and Particle Interactions

In nature, there are four fundamental forces describing interactions: the strong nuclear force, the weak nuclear force, the electromagnetic force, and the gravitational force. The SM models the former three of these forces[3], describing each via a Quantum Field Theory (QFT) corresponding to the exchange of a gauge boson, a particle with integer spin that does not obey the Pauli Exclusion principle. A broad summary of the forces, their bosons, and their associated QFTs can be found in Table 1.2, and the following sections will describe the various theories relating to each interaction.

Table 1.2: List of forces described by the SM, as well as their associated field theory and gauge boson mediator. Relative strength is an order of magnitude estimate at a distance of 1 fm. The strength of the strong and weak force vary significantly with distance. By coupling constants only (neglecting distance), the weak force is comparable in strength to the EM force.

Force	Field Theory	Mediating Boson	Relative Strength
Strong Nuclear	Quantum Chromodynamics	Gluon (g)	1
Electromagnetism	Quantum Electrodynamics	Photon (γ)	10^{-3}
Weak Nuclear	Quantum Flavordynamics[4]	$W^{\pm}$, Z	10^{-8}

[3]The gravitational force is not modeled by the SM. It has relative strength of 10^{-37} at 1 fm. In theories of quantum gravity, it is mediated by the "graviton" (G), a massless spin-2 particle.

[4]Modern particle physics seldom uses the term quantum flavordynamics. At high energy scales, the weak force and electromagnetic force are interpreted through electroweak theory, which is discussed in Section 1.1.2.3.

1.1.2.1 Quantum Electrodynamics

Equation 1.1, the Dirac equation, is not gauge invariant. Performing a local gauge transformation on the aforementioned fermion spinor (denoted by ψ)

$$\psi(x) \rightarrow e^{-i\theta(x)}\psi(x), \tag{1.3}$$

for real-valued function $\theta(x)$, will transform the Dirac Lagrangian, Equation 1.2, like

$$\mathcal{L} \rightarrow \mathcal{L} - (\partial_\mu \theta)\bar{\psi}\gamma^\mu \psi. \tag{1.4}$$

In order to impose gauge invariance, a vector field can be introduced. Such a vector field must transform as

$$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \theta(x), \tag{1.5}$$

which has covariant derivative

$$D_\mu = \partial_\mu + iqA_\mu, \tag{1.6}$$

for a particle with charge q. This massless vector field corresponds to the electromagnetic field, and the covariant derivative describes the minimal coupling of the photon, its quanta, to fermions. Applying this vector field to the Dirac Lagrangian gives the Quantum Electrodynamics (QED) Lagrangian,

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.7}$$

With $F_{\mu\nu}$, the electromagnetic field tensor, as:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{1.8}$$

By this construction, the QED Lagrangian is gauge invariant, and symmetric under the transformation described in Equation 1.3. This transformation is equivalently thought of as multiplication by a unitary 1×1 matrix of the $U(1)$ group, meaning QED is symmetric under $U(1)$.

1.1.2.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the QFT corresponding to the strong nuclear interaction, and predicts how quarks and gluons interact. While QED corresponds to a $U(1)$ local gauge symmetry, QCD corresponds to a $SU(3)$ gauge group, where quarks transform in the fundamental **3** representation. This representation has eight 3×3 generators, corresponding to eight gluons, the mediator of the strong force. The dimensionality of the generators mean the wave function has three additional degrees of freedom known as color. Taking the convention of denoting these as "red" (r), "green" (g), and "blue" (b), the quark spinor can be denoted as a vector:

$$\psi = \begin{pmatrix} \psi_r \\ \psi_b \\ \psi_g \end{pmatrix}, \tag{1.9}$$

which makes a gauge transformation

$$\psi \to e^{i\theta} e^{-ig\boldsymbol{\lambda}\cdot\phi}\psi. \tag{1.10}$$

Here $\boldsymbol{\lambda}$ are the eight Gell-Mann matrices, while $\boldsymbol{\phi} = -\boldsymbol{a}/g$ for vector of length 8 $\boldsymbol{a}$, and strong coupling constant g. This corresponds to the $SU(3)_C$ group (C indicates "color").

In order to achieve invariance under said transformation, the derivative in the Dirac equation is replaced by the covariant derivative

$$D_\mu = \partial_\mu + ig\boldsymbol{\lambda} \cdot \boldsymbol{G}_\mu. \tag{1.11}$$

In this, $\boldsymbol{G}_\mu$ represents eight fields, corresponding to eight gluons with various color charge, the mediator of the strong force. This makes the QCD Lagrangian

$$\mathcal{L}_{QCD} = \bar{\psi}^j (i\gamma^\mu D_\mu^{jk} - m^{jk})\psi^k - \frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu}, \tag{1.12}$$

for mass m, indices j and k taking values 1 to 3, and gluon field tensor $F_{\mu\nu}$, given as

$$F_a^{\mu\nu} = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g f^{abc} \lambda^a G_\mu^b G_\nu^c. \tag{1.13}$$

Here f_{abc} are the structure constants associated with the $SU(3)$ group, and indices a, b, c represent the gluons, running from values 1 to 8. Of special note, this tensor contains triplet and quartic gluon self-coupling, a product of the gluon color charge.

1.1.2.3 Electroweak Interaction

In the SM, the weak and electromagnetic interactions are interpreted as manifestations of a single force. This unification occurs at energies above 100 GeV, and the combined force after unification is known as the electroweak force.

Electroweak interactions have four mediators, the $W^\pm$, Z, and γ, and subsequently four generators. The SM would require the masses of these generators to be symmetric: all equal to zero. The $W^\pm$ and Z, however, have non-zero masses ($m_{W^\pm} = \pm 80.379 \pm 0.012$ GeV and $m_Z = 91.1876 \pm 0.0021$ GeV). To reconcile this, a symmetry breaking is introduced, leaving the gauge group electromagnetically invariant. The SM of electroweak interactions by Glashow, Weinberg, and Salam [15–17] contains symmetry breaking:

$$SU(2) \times U(1) \to U(1)_{EM}. \tag{1.14}$$

The $SU(2)$ subgroup is weak isospin, where left-handed fermions are doublets, and right-handed fermions are all singlets. The $U(1)$ subgroup is weak hypercharge (Y), a conserved quantity. The symmetry breaking described requires the Higgs field, and is known as Electroweak Symmetry Breaking.

1.2 The Higgs Field and Higgs Boson

Consider the $U(1)$ gauge theory with a single gauge field (the photon). The Lagrangian is the second term of Equation 1.7,

$$\mathcal{L}_{QED} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{1.15}$$

with $F_{\mu\nu}$ defined in Equation 1.8. Adding a massive term to this equation would violate the established local gauge invariance, as the $U(1)$ invariance requires the photon to be massless. Extending this by adding a complex scalar doublet,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4, \end{pmatrix} \tag{1.16}$$

then the Lagrangian and gauge-invariant potential become:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + |D_\mu \phi|^2 - V(\phi), \tag{1.17}$$

with:

$$D_\mu = \partial_\mu + i\frac{g}{2}\tau \cdot W_\mu + i\frac{g'}{2}B_\mu Y \tag{1.18}$$

$$V(\phi) = \mu^2|\phi|^2 + \lambda(|\phi|^2)^2. \tag{1.19}$$

Where g is the SU(2) gauge coupling, g' is the U(1) gauge coupling, and Y is the U(1) hypercharge. The minima of this potential is known as the vacuum state, and for the minima to be finite, λ must be greater than 0. For $\mu^2 > 0$, the minima is located at 0, however, for $\mu^2 < 0$, the potential has minima at:

$$\phi^\dagger\phi = \frac{v^2}{2}, \tag{1.20}$$

where

$$v = \sqrt{-\mu^2/\lambda}. \tag{1.21}$$

This potential is shown (in 2 dimensions) in Figure 1.1.

These minima are degenerate. By convention, v is selected to be real, making $\phi_1 = \phi_2 = \phi_3 = 0$, and $\phi_4 = v$. In making this selection, electroweak symmetry is broken. Unitarity gauge is given by expanding about this minima, making this field:

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ H(x) + v \end{pmatrix} \tag{1.22}$$

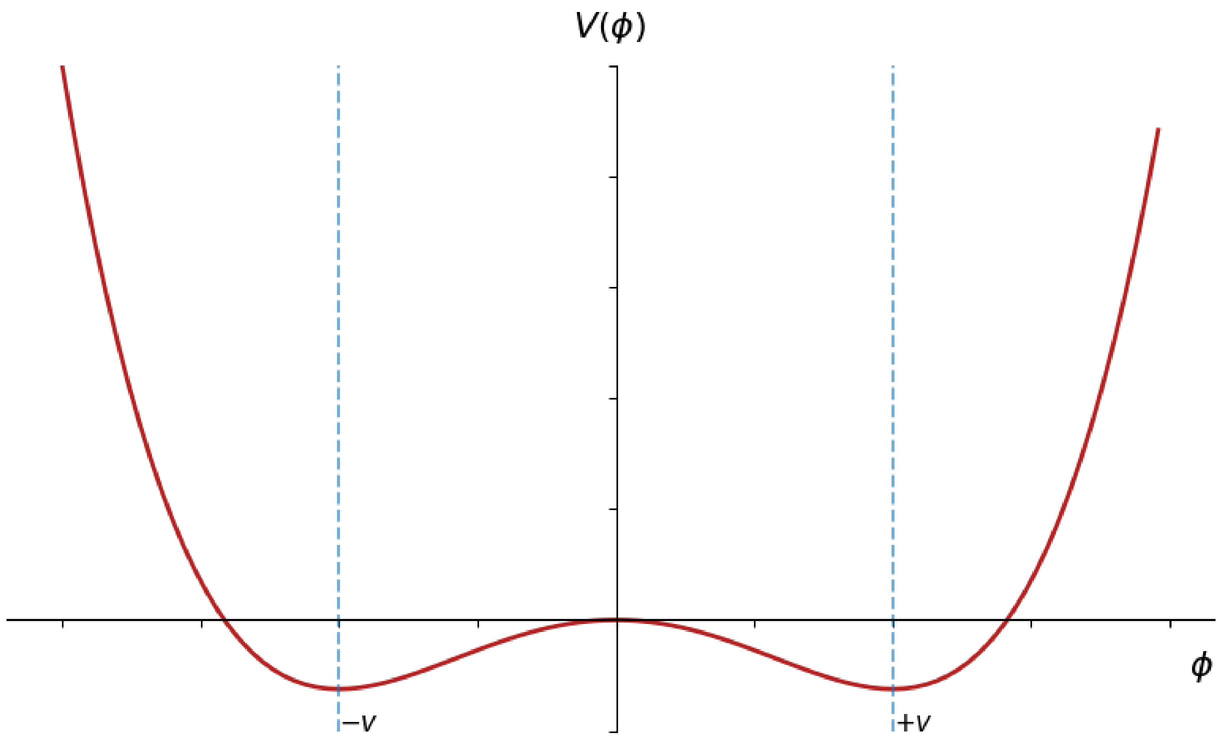


Figure 1.1: The one-dimensional Higgs potential where $\mu^2 < 0$. The minima of this curve, known as the Higgs vacuum expectation value (vev), is indicated.

In this formulation, $H(x)$ represents the Higgs field, and v is known as the vacuum expectation value (vev). The contribution to the gauge boson masses from the scalar kinetic energy term of the Lagrangian is then,

$$\frac{1}{2}(0, v)(\frac{1}{2}g\tau \cdot W_\mu + \frac{1}{2}g' B_\mu)^2 \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{1.23}$$

yielding two charged gauge fields and two neutral gauge fields,

$$W^{\pm}_{\mu} = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu) \tag{1.24}$$

$$Z^{\mu} = \frac{-g' B_\mu + gW^3_\mu}{\sqrt{g^2 + g'^2}} \tag{1.25}$$

$$A^{\mu} = \frac{gB_\mu + g' W^3_\mu}{\sqrt{g^2 + g'^2}}. \tag{1.26}$$

This provides a prediction the prediction of the gauge boson masses:

$$m^2_W = \frac{g^2 v^2}{4}, \qquad m^2_Z = \frac{(g^2 + g'^2)v^2}{4}, \qquad m_A = 0, \tag{1.27}$$

which agrees with experimental values when the vev is 246 GeV. In unitarity gauge, the Higgs potential takes the form:

$$V(h) = \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4, \tag{1.28}$$

indicating self-interactions with the Higgs boson. The second term gives rise to the trilinear coupling, while the third indicates the Higgs quartic coupling. These couplings imply the existence of hhh and $hhhh$ vertices, respectively. The Higgs boson mass can also be derived as $m_h = \sqrt{2\lambda} v$.

1.2.1 The Higgs Boson at the LHC

Higgs Boson Production

The dominant Higgs boson production modes are shown in Figure 1.2 as a function of the center-of-mass energy.

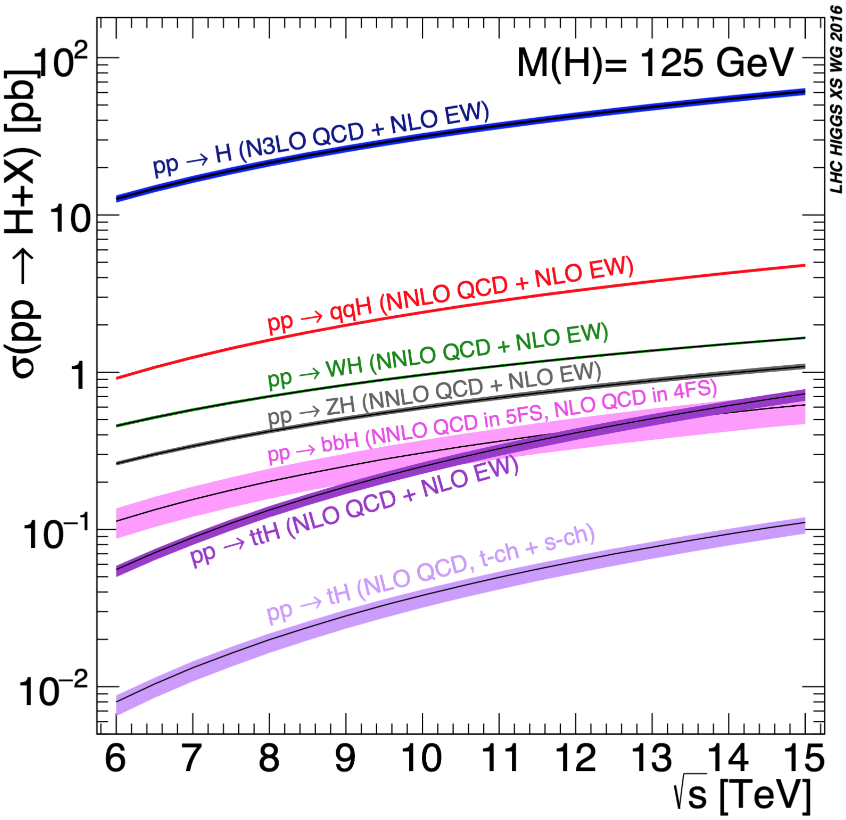


Figure 1.2: The cross-section of various Higgs boson production modes for a SM Higgs boson with $m_H = 125$ GeV as a function of the center-of-mass energy [18].

The most dominant production mode at the Large Hadron Collider (LHC) is gluon-gluon Fusion (ggF). At the current center-of-mass energy of 13 TeV, it makes up 87% of Higgs boson events. In this production mode, two initial state gluons produce the Higgs boson through a heavy quark loop.

Vector Boson Fusion (VBF) production is the second-most dominant Higgs boson production mode at the LHC, comprising 6.8% of Higgs boson events. In VBF production, incoming quarks radiate vector bosons, which interact to produce the Higgs boson. The radiating quarks are present in the final state, and possess a distinct forward signature.

The remaining major Higgs boson production modes at the LHC are vector boson associated production (WH and ZH), top associated production ($t\bar{t}H$ and tH), and $b\bar{b}$ fusion ($b\bar{b}H$).

Higgs Boson Decay

Due to the short lifetime of the Higgs boson, measurements and searches are performed by studying its decay products. The Higgs directly couples to all massive particles, and decays directly to any such massive particles provided conservation laws are followed. Additionally, it can decay to massless particles via virtual loops, but as a result, these branching ratios are relatively small. The SM Higgs branching ratios are shown in Table 1.3.

Table 1.3: Branching ratios for the SM Higgs boson [14].

Channel	Branching Ratio
$H \to bb$	57.5 ± 1.9
$H \to WW^*$	21.6 ± 0.9
$H \to gg$	8.56 ± 0.86
$H \to \tau^+\tau^-$	6.30 ± 0.36
$H \to c\bar{c}$	2.90 ± 0.35
$H \to ZZ^*$	2.67 ± 0.11
$H \to \gamma\gamma$	0.228 ± 0.011
$H \to Z\gamma$	0.155 ± 0.014
$H \to \mu^+\mu^-$	0.022 ± 0.001

1.3 Di-Higgs Production

Observing the trilinear coupling directly probes electroweak symmetry breaking, and provides a precision test electroweak theory. For a direct probe, production modes resulting in pair and triplet Higgs production must be studied[5]. The cross-section of such production

[5]Through loop corrections, mono-Higgs searches can provide an indirect handle on the trilinear coupling. ATLAS has published one such analysis with 80 fb^{-1}, found in Reference [19]. When making the assumption that all other Higgs couplings are as predicted by the SM, this analysis finds the observed (expected) 95% C.L. interval on the trilinear coupling to be $-3.2 < \kappa_\lambda < 11.9$ ($-6.2 < \kappa_\lambda < 14.4$).

modes, however, is quite small. Triplet production of Higgs bosons occurs at a rate 6 orders of magnitude less frequently than mono-Higgs production, and is not feasible to study at the LHC. Pair production is 3 orders of magnitude more rare than mono-Higgs production, but over the full LHC lifetime, a handle on the trilinear self-coupling may be achievable.

The analysis presented in this thesis is one such search for di-Higgs production, specifically where one of the produced Higgs bosons decays to a pair of photons, and the other to a $b\bar{b}$ pair.

1.3.1 Higgs Boson Pair Production in the Standard Model

In the standard model, several production modes pair produce Higgs bosons. ggF is the most dominant with a cross-section accounting for over 90% of di-Higgs production, described in Section 1.3.1.1. The next most dominant is VBF, which is responsible for just over 5% of di-Higgs production and is described in Section 1.3.1.2. The remaining production modes, VHH, $ttHH$, and $tjHH$ collectively comprise less than 5% of di-Higgs production, and have a cross-section too small to make a meaningful contribution to the presented analysis with the current LHC dataset.

1.3.1.1 Gluon-Gluon Fusion

The predominant di-Higgs production mode is ggF, in which the production is mediated by a quark loop. In the SM, this proceeds through two diagrams at tree level, where the quark loop either forms a box or a triangle. The former directly pair produces two Higgs bosons while the latter pair produces via a virtual Higgs boson, making it sensitive to the Higgs self-coupling, λ_{HHH}. These two diagrams interfere destructively. QCD corrections to this process are known to next-to-leading order (NLO), and can double the cross-section from the leading order (LO) value. Furthermore, next-to-next-to-leading order (NNLO) corrections are known in the large top mass limit[6] and represent a $\sim$20% increase from the NLO prediction. At 13 TeV, the cross-section for ggF production[7] is $31.05^{+2.2\%}_{-5.0\%} \pm 3.0\%$ fb [18].

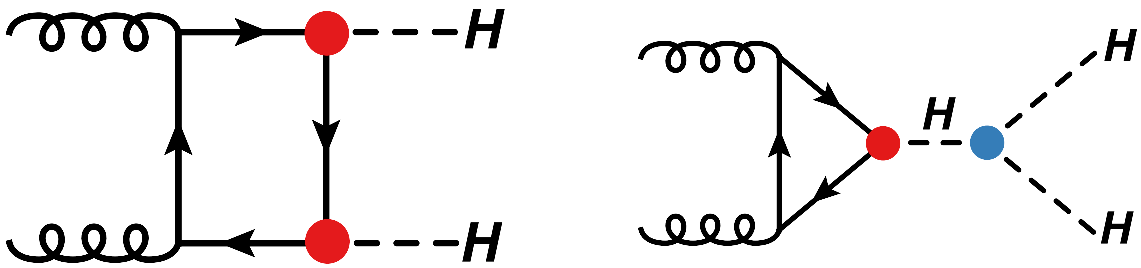


Figure 1.3: Tree level Feynman diagrams for non-resonant ggF HH production. The left is the "box" diagram, which is not sensitive to the Higgs trilinear coupling. The right is the "triangle" diagram, which is sensitive to the Higgs trilinear coupling, λ_{HHH}, via the interaction between the virtual Higgs boson and the pair produced di-Higgs system labeled by the blue vertex.

1.3.1.2 Vector Boson Fusion

The second-most dominant HH production mode is VBF ($qq \rightarrow HHqq$), in which two quarks scatter via the exchange of a virtual vector boson, $W^{\pm}$ or Z (generally denoted V), and from that boson, a di-Higgs system is produced. This proceeds via three Feynman diagrams at tree level, shown in Figure 1.4. At 13 TeV, the cross-section for VBF production is $1.73^{+0.03\%}_{-0.04\%} \pm 2.1\%$ fb [18].

The coupling denoted c_V is the interaction between two vector bosons and a Higgs boson. This coupling has constraints from mono-Higgs analyses, measured at $1.21^{+0.22}_{-0.21}$ the SM value [22]. The coupling denoted κ_λ is equal to λ/λ_{SM}, meaning that the VBF production

[6]The large top mass limit approximates m_t to be large compared to the partonic center-of-mass collision energy and m_H, in order to approximate QCD corrections to HH production. This approximation provides a poor description of of the total cross-section and kinematics [20].

[7]The first (upper and lower) uncertainties refer to scale uncertainties, the second is PDF + α_s uncertainty.

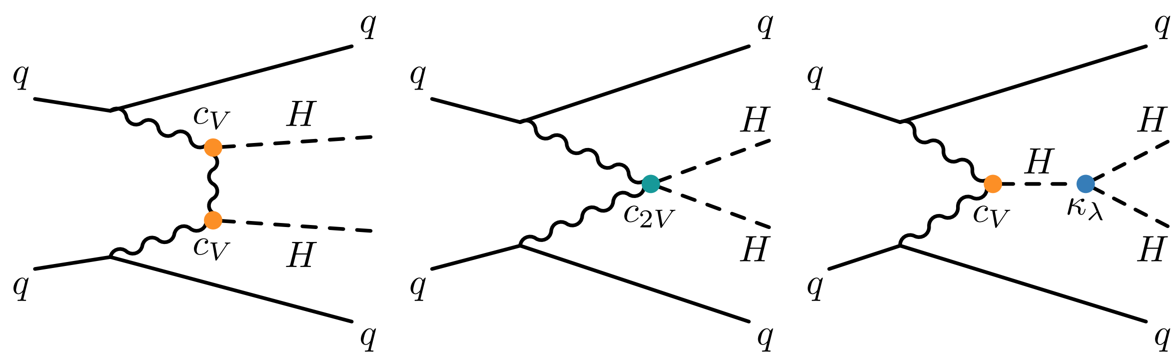


Figure 1.4: Tree level Feynman diagrams for non-resonant VBF HH production [21].

mode offers sensitivity to the Higgs trilinear coupling as well. Phenomenology studies at parton level have shown the sensitivity in the VBF mode to provide a competitive handle on κ_λ [23]. The coupling denoted c_{2V} is the interaction between two vector bosons and two Higgs bosons. VBF HH searches offer a unique handle on this vertex, previously entirely unconstrained.

Despite the low cross-section of VBF production, the unique kinematics make the channel appealing for searches. The diphoton VBF channel contributed to the 2012 single Higgs boson discovery [10]. The kinematics of VBF di-Higgs production are similar to that of mono-Higgs, with a forward, high-m_{jj} system. The major difference is that mono-Higgs searches take advantage of the low central jet activity by applying a veto. In $HH \to \gamma\gamma b\bar{b}$, however, the $H \to b\bar{b}$ (Higgs decay modes are discussed in Section 1.3.3) system causes central jet activity, making such a veto impossible.

1.3.2 Production Beyond the Standard Model

Due to the small cross-section of HH production, the current ATLAS dataset does not have sensitivity to SM production. An enhancement to the cross-section, however, may be noticeable in the dataset and would indicate the presence of physics Beyond Standard Model

(BSM). There are a variety of scenarios in which this could occur, both through non-resonant and resonant enhancements, discussed in Sections 1.3.2.1 and 1.3.2.2, respectively.

1.3.2.1 Non-Resonant BSM Production

Non-resonant enhancements to the HH cross-section could occur in two ways: coupling strengths which deviate from the SM prediction, or the presence of couplings not predicted by the SM.

$\boldsymbol{\kappa}_\lambda$

Deviations from the SM prediction of the Higgs trilinear coupling can cause enhancements to the cross-section, which may be observable at the current LHC dataset. The cross-section is minimized at 2.45 times the SM value, in which the destructive interference between the box and triangle diagrams is maximized. The cross-section for each production mode as a function of the coupling strength can be seen in Figure 1.5.

$\boldsymbol{c}_{2V}$

The VBF HH production mode offers sensitivity to the $hhVV$ vertex, and any deviations from SM expectations for this coupling would lead to a significant enhancement in the HH cross-section. At a factor of 2 times the standard model coupling strength, the cross-section increases by two orders of magnitude (shown in Figure 1.6), allowing for an enhancement visible in the current LHC dataset [25].

1.3.2.2 Resonant BSM Production

Several BSM theories predict new particles that would decay to a pair of Higgs bosons, which would result in a resonant enhancement to the HH cross-section. Notably, Two

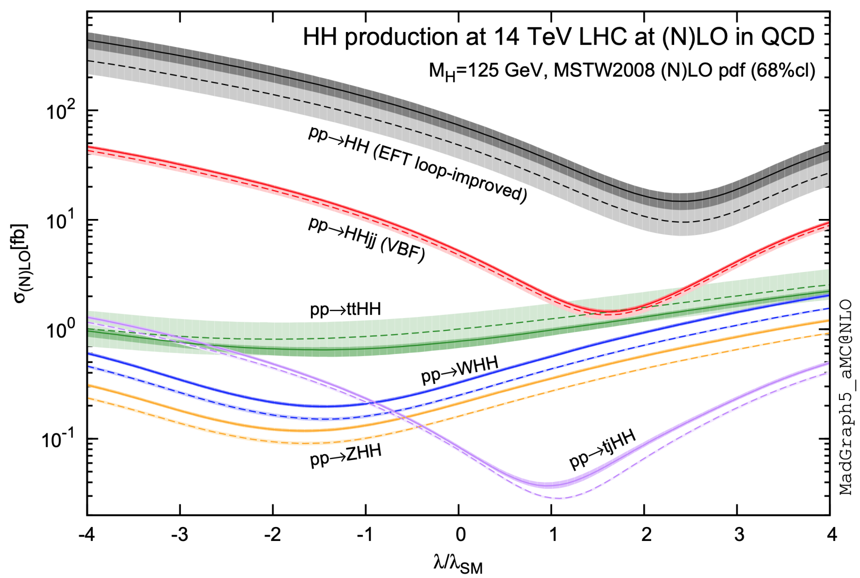


Figure 1.5: LO (dashed lines, light-color bands) and NLO (solid lines, dark-color bands) cross-sections for the various HH production modes as a function of the Higgs trilinear coupling, normalized to the standard model coupling value. Uncertainties represent scale and PDF uncertainties, added linearly [24].

Higgs Doublet Models (2HDMs) extend the Higgs sector with the presence of a heavy scalar doublet [26, 27]. Among these models are the minimal supersymmetric extension of the SM [28–30], twin Higgs models [31, 32], and composite Higgs models [33]. The analysis presented in this thesis is model-independent, searching for enhancements to the HH cross-section through the presence of a generic scalar particle.

ggF

In ggF production, the resonance would be present through a coupling to the triangle diagram, which subsequently decays into the HH pair. As a result, this resonance would be observed as an enhancement to the HH cross-section, as this production is not modeled in the SM. By conservation laws, its mass would be equivalent to the four-vector sum of the pair produced Higgs bosons, $m_X = m_{HH}$. A sample Feynman diagram for such a resonance

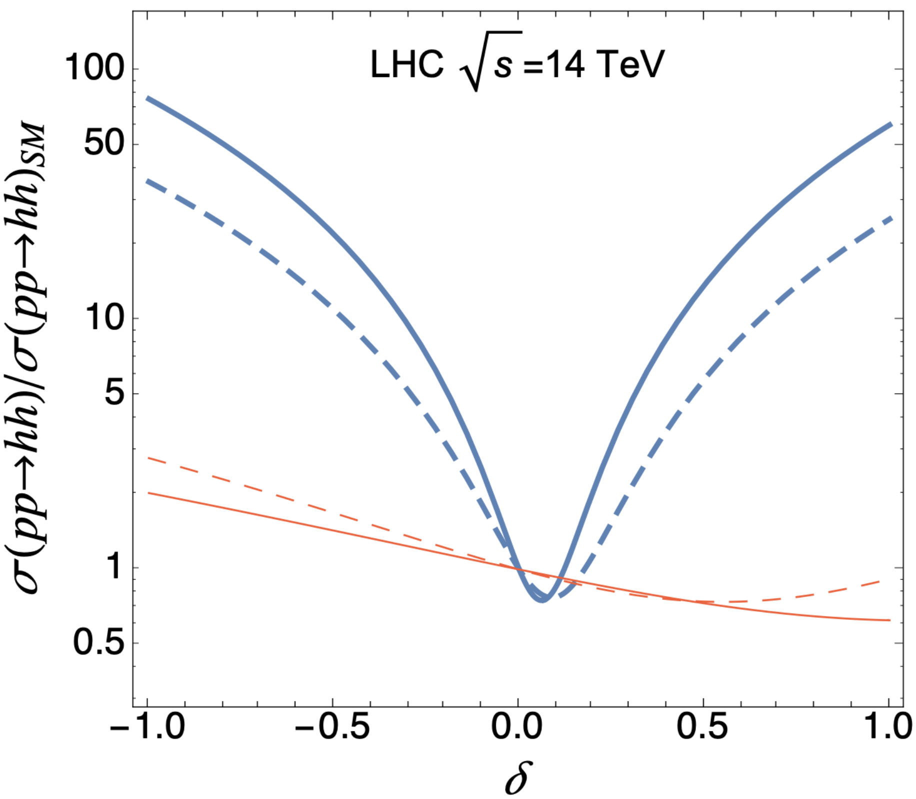


Figure 1.6: The VBF HH cross-section as a function of coupling deviation from prediction, in units of the SM value ($\delta = c - c_{SM}$ for coupling, c. I.e. $\delta = 0$ corresponds to the SM prediction, $\delta = 1$ is twice the SM prediction). The solid blue line shows deviations in the c_{2V} vertex, $\delta_{c_{2V}}$, while the red line shows deviations for the c_3 vertex, δ_{κ_λ} after applying broad acceptance cuts presented in Reference [25]. The dashed lines represent the cross-section after analysis cuts targeting VBF HH production.

can be seen in Figure 1.7. The presented analysis searches for this resonance as a generic scalar particle, in order to maintain model-independence.

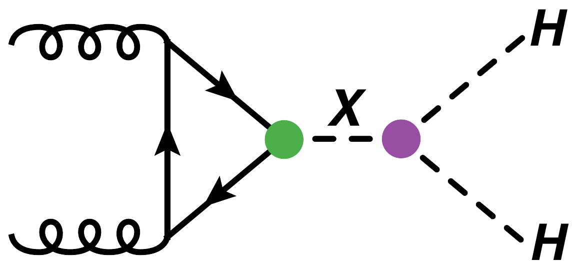


Figure 1.7: A Feynman diagram for resonant ggF HH production. In the analysis presented, X takes the form of a generic scalar to remain model-independent.

VBF

Similar to searches for resonant ggF Higgs boson pair production, VBF HH production involving an intermediate resonance may be considered. This kind of search would be complementary to the ggF searches, as in this case the vector bosons are coupling to the new resonance, which then decays to a pair of Higgs bosons, as depicted in Figure 1.8. However, this production mode is not particularly well studied (Reference [34] presents an analysis in the context of a model with warped extra dimensions), since its very small cross-section poses a question on the ability of the LHC to impose significant constraints through this type of search.

1.3.3 Di-Higgs Decay

In di-Higgs searches, properties of both decaying Higgs decays must be considered. Branching ratios for various combinations of HH decays are shown in Figure 1.9. The already small HH cross-section motivates searches to target high branching ratio decay modes, thus the three primary channels of study all incorporate the $H \to b\bar{b}$ decay. The $b\bar{b}b\bar{b}$ decay channel has highest possible branching ratio at 33.9%, but has significant challenges presented by

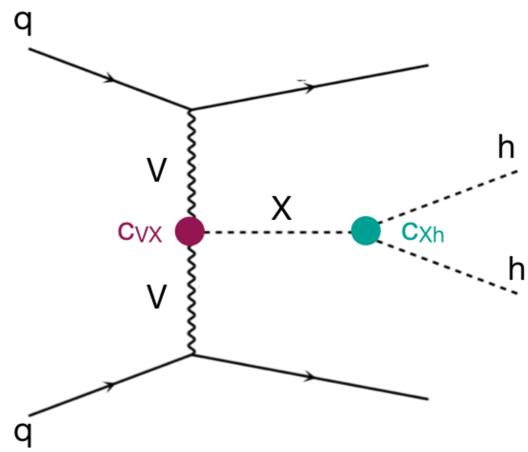


Figure 1.8: A Feynman diagram for resonant VBF HH production.

triggering and a large QCD background. The $b\bar{b}\tau\tau$ has the third largest branching ratio at 7.3%, and benefits from a much smaller SM background than the $b\bar{b}b\bar{b}$ channel, primarily $t\bar{t}$ and $Z \to \tau^+\tau^-$ + jets. Last the $b\bar{b}\gamma\gamma$ channel, presented in this thesis, has just a 0.3% branching ratio, but takes advantage of the excellent $H \to \gamma\gamma$ mass resolution, as well as the clean diphoton trigger and photon reconstruction efficiency.

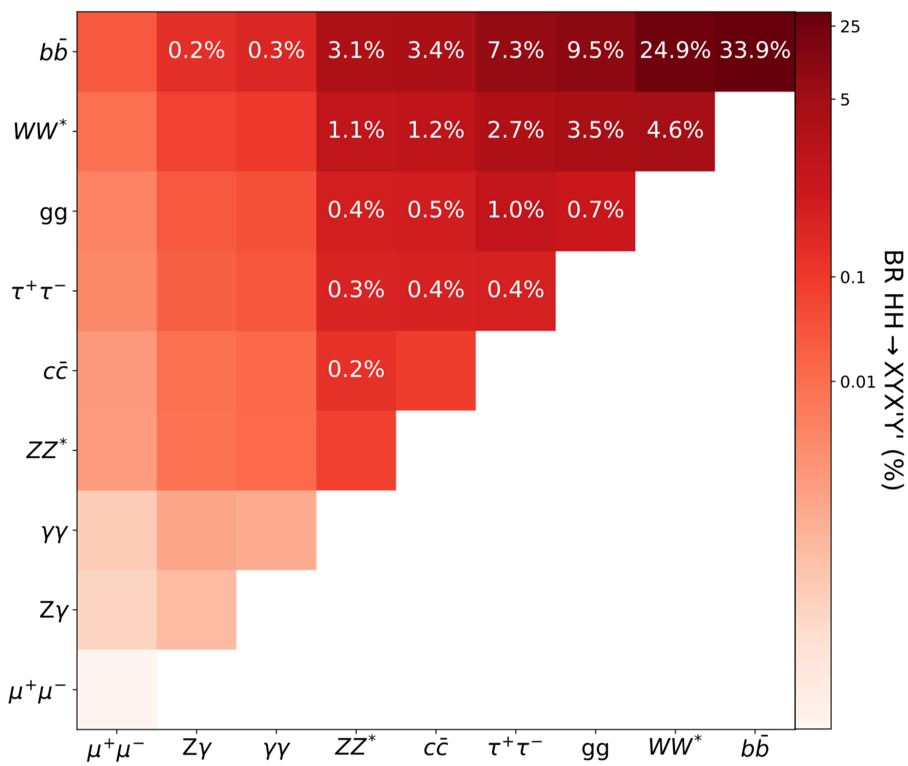


Figure 1.9: SM Higgs boson branching ratios for HH pairs to various final state combinations [35].

CHAPTER 2

THE LHC AND ATLAS EXPERIMENT

You can observe a lot by watching.
— Yogi Berra

2.1 The Large Hadron Collider

The LHC [36] is a hadron accelerator and collider 27 km in circumference located on the French-Swiss border. It collides beams of protons[1] at high rate and center-of-mass energy. In Run 2, which spanned from 2015 to 2018, the center-of-mass energy ($\sqrt{s}$) in the laboratory frame was 13 TeV (6.5 TeV per beam), and bunches of about 10^{11} protons collided every 25 ns.

Hydrogen atoms are the source of the protons the LHC collides. The protons are liberated from the atom using an electric field, then pass through several accelerators, which take advantage of radiofrequency (RF) cavities, to get to their final energy. First, the Linear Accelerator (Linac) accelerates the bunches to 50 MeV. The remaining pieces of the acceleration chain are circular accelerators, starting with the Proton Synchrotron Booster (PSB), which raises the energy to 1.4 GeV, then the Proton Synchrotron (PS) accelerates bunches to 26 GeV, and the Super Proton Synchrotron (SPS) accelerates them to 450 GeV before delivering to the LHC, where they are accelerated to their final energy of $\sqrt{s} = 13$ TeV.

Along the LHC, collisions occur at four interaction points, where detectors are situated. LHCb (Large Hadron Collider beauty) is a forward physics experiment designed to study B physics [37]. ALICE (A Large Ion Collider Experiment) is a heavy-ion detector studying

[1] The LHC also collides ions to study lead-lead (PbPb) as well as proton-lead interactions

nuclear physics [38]. The ATLAS [39] and Compact Muon Solenoid (CMS) [40] experiments are general-purpose detectors, designed to study a variety of physics. A schematic showing the location of these interaction points and detectors can be found in Figure 2.1.

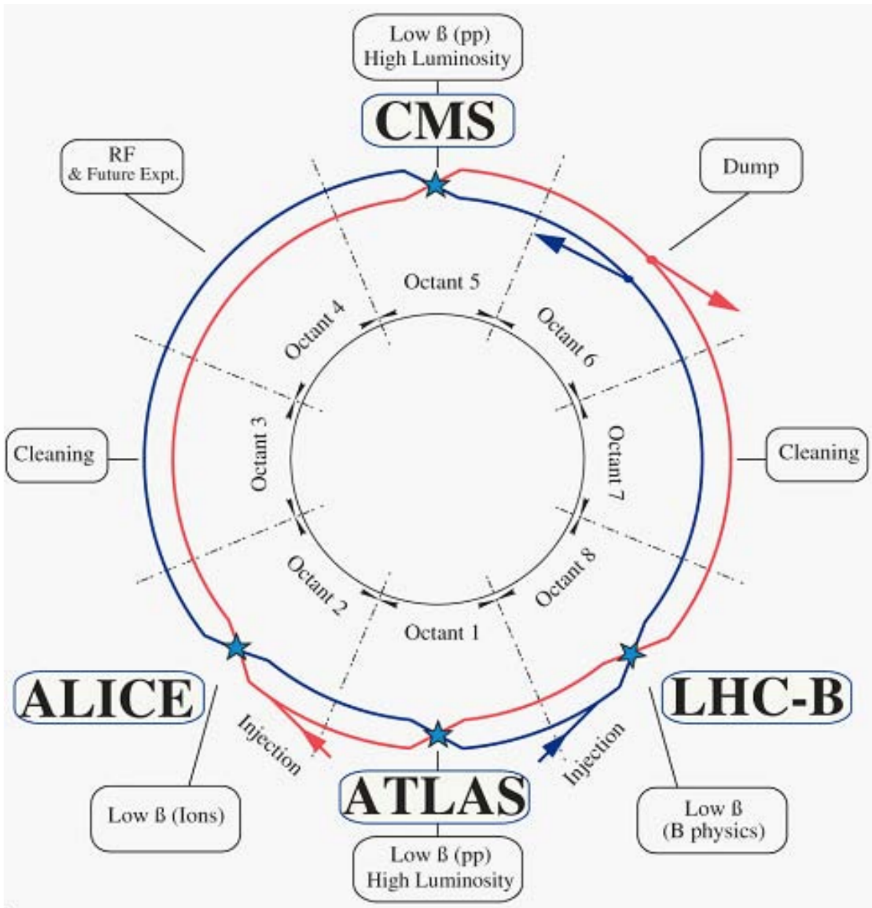


Figure 2.1: A schematic of the LHC with the various interaction points indicated [36].

2.2 The ATLAS Detector

ATLAS [39] is a general-purpose detector designed to reconstruct events from hadrons collided by the LHC. With respect to the Interaction Point (IP), it has forward-backward symmetric cylindrical geometry, and covers a solid angle of almost 4π. It is 44 m long, 25 m in diameter, weighs over 7,000 tons, and is located 100 m underground.

ATLAS is composed of several nested cylindrical subsystems. Nearest the IP is the Inner Detector (ID) (discussed in 2.2.3), measuring the trajectory of charged-particles as they bend through a 2 T magnetic field created by a solenoid which envelops the ID (discussed in 2.2.2). The ID is composed of three components: the silicon Pixel Detector, a semiconductor microstrip detector known as the SemiConductor Tracker (SCT), and a straw-tube tracker known as the Transition Radiation Tracker (TRT). Outside of the ID are the calorimeters (discussed in 2.2.4). ATLAS utilizes two calorimeters, a Liquid Argon (LAr) calorimeter to measure electromagnetic radiation and a scintillating tile calorimeter to measure hadronic radiation. Beyond the calorimeters is a 4 T toroidal magnetic field (also discussed in 2.2.2) and the Muon Spectrometer (discussed in 2.2.5), which tracks muon trajectories. ATLAS employs a two-stage triggering system, one hardware-based and one software-based, to identify and record interactions containing interesting physics by sending them to the Data Acquisition (DAQ) system (discussed in 2.2.6.) A schematic of the ATLAS detector can be found in Figure 2.2.

2.2.1 The ATLAS Coordinate System

ATLAS uses a right-handed coordinate system, with the IP defining the origin. The z-axis is along the beamline, the x-axis points toward the center of the LHC, and the y-axis points upwards, toward Earth's surface. Transverse quantities such as p_{T} and E_{T} are calculated in the $x - y$ plane. The half of the detector pointed toward the Saleve is known as the "A-side" and has positive z coordinates. The side with negative z, pointed toward the Jura, is denoted "C-side," where "B" is used to denote the barrel.

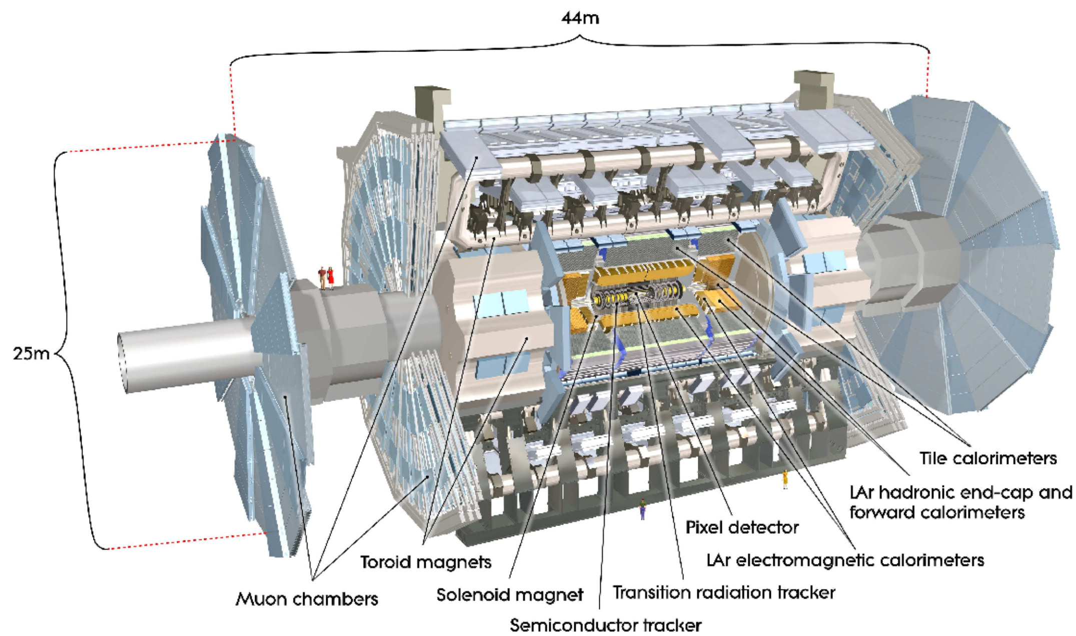


Figure 2.2: A cut-away schematic the ATLAS detector. Major components and dimensions are labeled [39].

Due to the shape of the detector, cylindrical coordinates are also useful for describing position. The azimuthal angle, ϕ, is measured around the beam axis, with $\phi = 0$ lying on the positive x-axis. The polar angle, θ, is the angle from the beam axis, where $\theta = 0$ is parallel to the beampipe. Commonly, the *pseudorapidity* is used to define position, defined as

$$\eta = -\ln[\tan(\theta/2)]. \tag{2.1}$$

This value is 0 perpendicular to the beamline, and infinite as it approaches the beam axis, as shown in Figure 2.3. Occasionally the *rapidity*, defined as

$$y = \frac{1}{2}\log\frac{E + p_z}{E - p_z}, \tag{2.2}$$

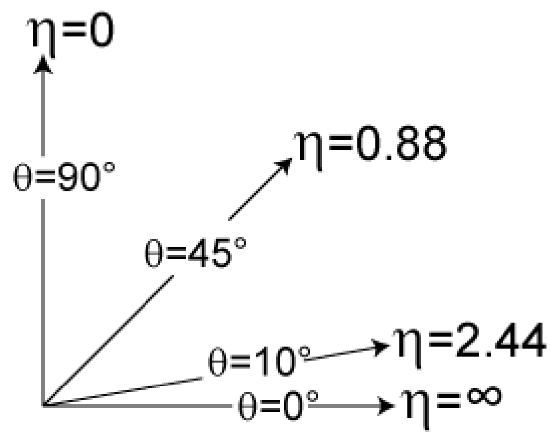


Figure 2.3: Pseudorapidity distribution shown on a polar grid.

is used to describe position as well. Rapidity and pseudorapidity are preferable to the polar angle due to their Lorentz invariance. For massless particles, the rapidity and pseudorapidity are equivalent.

To describe distance between particles, the angle space in pseudorapidity and azimuth is commonly used. It is also Lorentz invariant under longitudinal boosts and defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{2.3}$$

2.2.2 Magnet Systems

In order to measure the charge and momentum of particles in the detector, a magnetic field is used. The curvature of a charged particle as it traverses through a magnetic field can be used to derive the charge to mass ratio. The sign of the charge is determined from the bend of the track.

The magnet system in ATLAS is a hybrid system comprised of four superconducting magnets: a central solenoid, a barrel toroid, and two end-cap toroids. The barrel toroids define the dimensions of the ATLAS detector, extending 26 m along the beam axis, and have an outer diameter of 22 m [41]. A schematic of the components of the magnet system is shown in Figure 2.4.

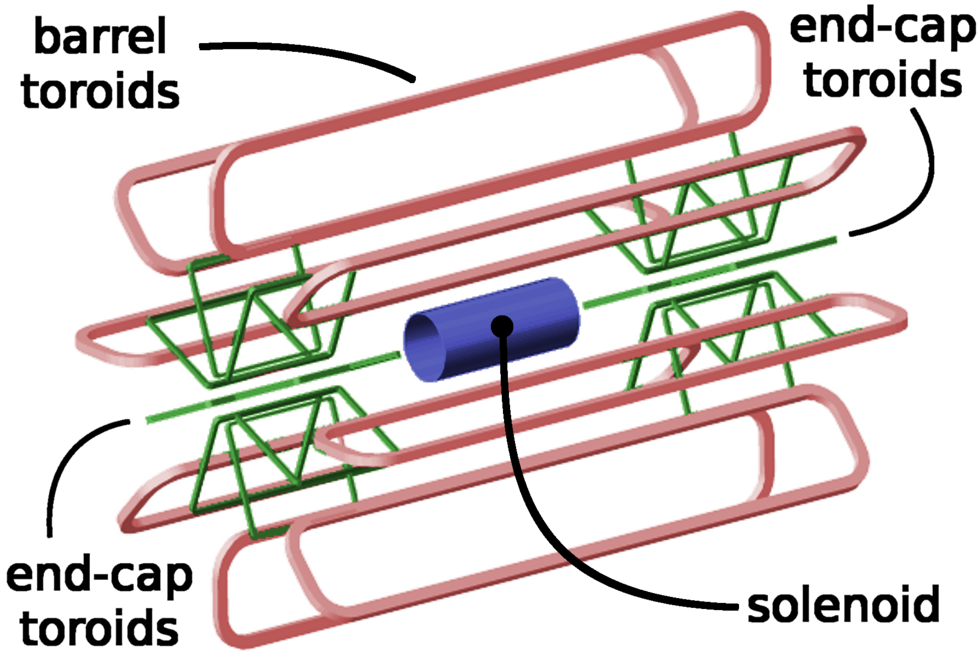


Figure 2.4: A schematic of the components of the magnet system in the ATLAS detector [42].

Central Solenoid

The central solenoid sits between the ID and the EM calorimeter and provides a 2 T magnetic field. This field is axially symmetric, and causes curvature in charged particles as they move through the ID. In order to minimize material before the calorimeters, it was designed to be thin, a single-layer coil wound with 1173 turns with Al-stabilized Nb/Ti conductor. It covers a length of 5.3 m along the beam axis and has an outer diameter of 2.63 m [43].

Toroids

The barrel toroid is made up of eight coils of a flat racetrack shape, which provide a symmetric radial magnetic field toward the beam axis, with a peak field of 3.9 T [44]. The two end-cap toroids are also composed of eight coils, with a length of 5 m, outer diameter of

10.7 m and an inner bore of 1.65 m, and providing a peak 4.1 T field. The end-cap toroids are inserted into the barrel toroid and line up with the center solenoid. The three-toroid design was chosen over a single toroid to maintain easy access to the core of the detector [45]. Together the toroids provide a magnetic field to the muon system, causing curvature in the muon trajectory.

2.2.3 Inner Detector

The ID [46] serves to identify the precise interaction point and trajectory of charged particles. The ID is segmented into barrel and end-cap sections along the beam axis. The barrel extends 1.6 m, covering $|\eta| < 1$, and the end-cap sections cover the remainder of the 7 m length. With this setup, the ID provides precision tracking up to $|\eta| = 2.5$. It is composed of three subsystems, the Pixel Detector, the SCT, and the TRT.

Pixel Detector

Nearest the IP is the Pixel Detector, which provides high-granularity, high-precision measurements, and is used to determine impact parameters.

Pixel sensors are reversely polarized diodes, created by an n+ implant on an n-doped silicon substrate with dimensions 50 µm × 400 µm, and a depth of 250 µm. The detector has 1,744 modules with 46,080 readout channels per modules, for a total of over 80 million pixels.

The innermost layer, the Insertable B-Layer (IBL) [47], was installed between Run 1 and Run 2, during the shutdown that lasted from 2013 to 2015. It takes advantage of smaller modules for finer granularity at 50 µm × 250 µm. Additionally, it is located closer to the interaction point, at a distance of $R = 33.25$ mm, compared to the previous first layer at $R = 50.5$ mm. The smaller distance between the beamline and first detector layer helps for finding displaced vertices from b-jets and photons converting into an e^+e^- pair.

SemiConductor Tracker

The SCT provides measurements at intermediate range from the IP, vital for momentum, impact parameter, and vertex position measurements. It is composed of a barrel region and two end-caps, covering the distance $R = 300$ mm to $R = 520$ mm from the IP. SCT modules are made up of four strip sensors, two on each side, oriented at an angle of 40 mrad from each other. The sensors are single-sided p-in-n microstrip detectors on silicon wafers. The barrel is constructed of 2112 modules in four layers that are equally spaced, and each end-cap is constructed of 988 modules over nine disks. Collectively, the SCT has about 6 million silicon strips.

Transition Radiation Tracker

The TRT is the outermost subdetector of the ID, spanning from a distance $R = 554$ mm to $R = 1082$ mm from the IP, and pseudorapidity range $|\eta| < 2.0$. The TRT contributes to the momentum measurement, aids in pattern recognition for tracking, and provides discrimination between electrons and hadrons. It is a proportional drift tube composed of 372,032 straws, each with a 4 mm inner diameter. In the barrel, the straws have a length of 144 cm and are arranged parallel to the beam axis. In the end-cap, straws are 37 cm long, and arranged into 18 wheels, radially from the beam axis.

The straws are filled with a gaseous mixture[2] composed of 70% Xe, 27% CO_2, and 3% O_2. As a charged particle penetrates the straw, it deposits energy into the gas and creates electron-ion pairs. The electrons drift to the center wire, while the ions drift to the inner wall of the straw. The drifting charge produces an electrical signal, which is then compared to a low and high threshold. The low threshold is used to measure drift time, which is used to derive position of incidence. The high threshold is used to identify large energy deposits from transition radiation X-rays, which is useful to discriminate charged pions from electrons.

[2]This was the design gas mixture, as described in the ID Technical Design Report, Reference [46]. During Run 2, straws with leaks were filled with 70% Ar gas, rather than Xe due to cost. Ar provides similar tracking performance as Xe, but a worse efficiency absorbing transition radiation X-rays.

For momentum measurement, the TRT provides measurements equivalent to a single point with precision of 50 µm. For tracking, it provides around 36 hits per track. For electron-hadron discrimination, it has a pion rejection factor of between 15 and 200 (depending on η) for pions with $p_{\mathrm{T}} > 20$ GeV, at an electron efficiency of 90%.

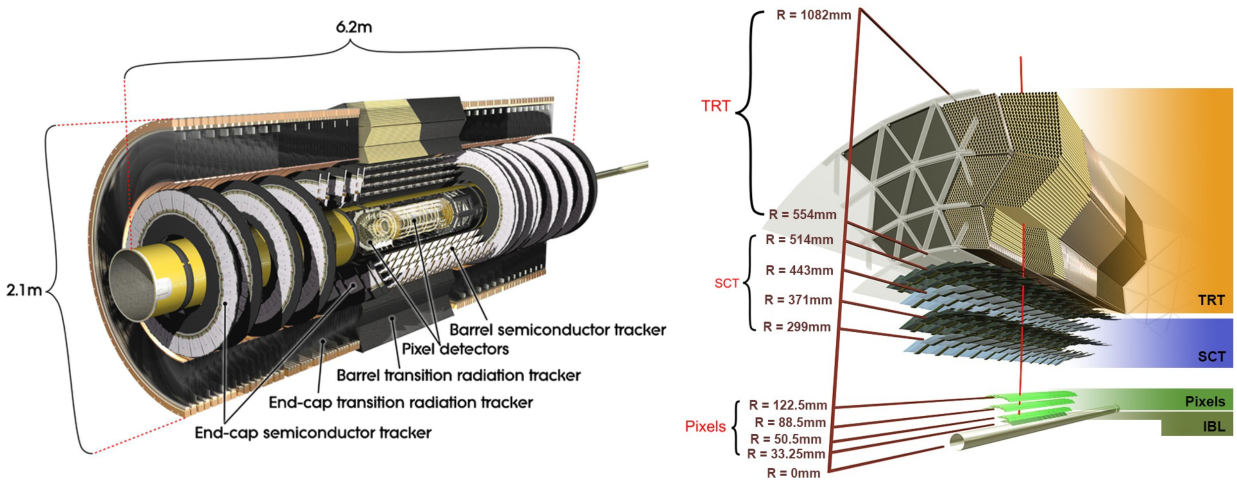


Figure 2.5: Schematic of the components of the ATLAS ID. A scale cutaway view of the ID with the Pixel and SCT (left) [48], and a layer-by-layer view of the ID subdetectors (right) [49].

2.2.4 Calorimeters

The calorimeter system in ATLAS serves to measure the energy of particles absorbed by their active material. ATLAS employs calorimeters targeting different types of particle showering: the LAr EM calorimeter targets EM showers from photons and electrons, while the hadronic Tile calorimeter targets hadronic showers. The EM calorimeter lies outside the ID and central solenoid, and the hadronic calorimeter is outside of the EM calorimeter. A schematic of the calorimeter can be found in Figure 2.6.

Electromagnetic Calorimeter

The electromagnetic calorimeter is a LAr sampling calorimeter. Upon interacting with

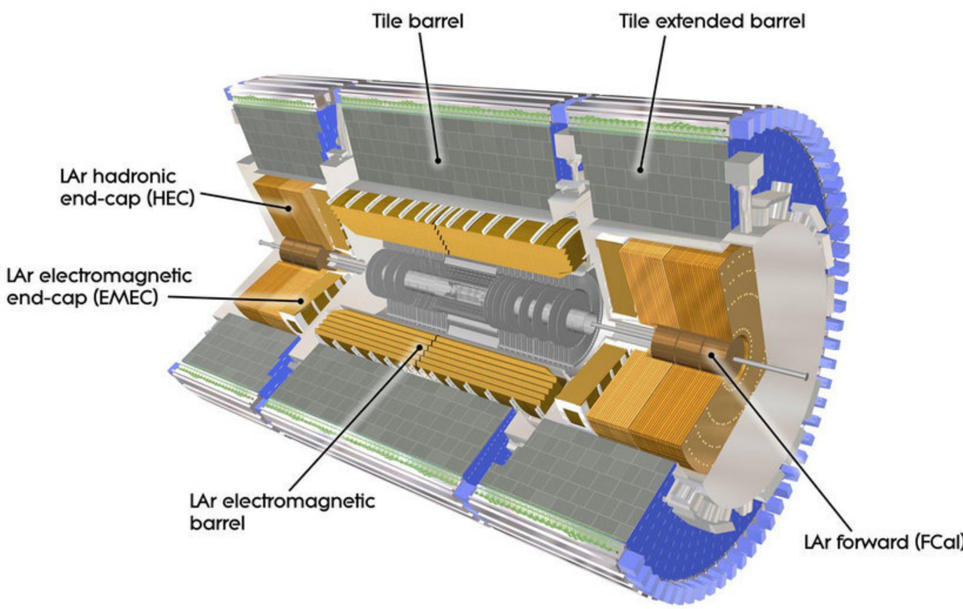


Figure 2.6: A schematic of the calorimeters in the ATLAS detector [39].

the LAr active material, photons will interact dominantly by creation of e^+e^- pairs above 5 MeV. Below that threshold, photons interact via the photoelectric or Compton effect. For electrons, interaction occurs via bremsstrahlung at energies above 1 MeV, causing the electron to radiate photons and decelerate. This continues for subsequent daughter particles and has a showering effect in the calorimeter [51].

The LAr EM calorimeter features absorbers with an "accordion" style geometry, and read-out electrodes. The accordion geometry provides gap-less azimuthal (ϕ) symmetry. The EM calorimeter is composed of three regions, the Electromagnetic Barrel (EMB) and two Electromagnetic End-Cap (EMEC), each with their own cryostat.

The EMB uses lead-stainless steel converters as the passive material and covers $|\eta| < 1.5$. It is split into two half-barrels, meeting at $|\eta| = 0$, each with 1024 converters and electrodes. Longitudinally, the EMB is divided into three layers. The first strip layer is very finely segmented in η to provide precision position information, with granularity 0.003×0.1 in

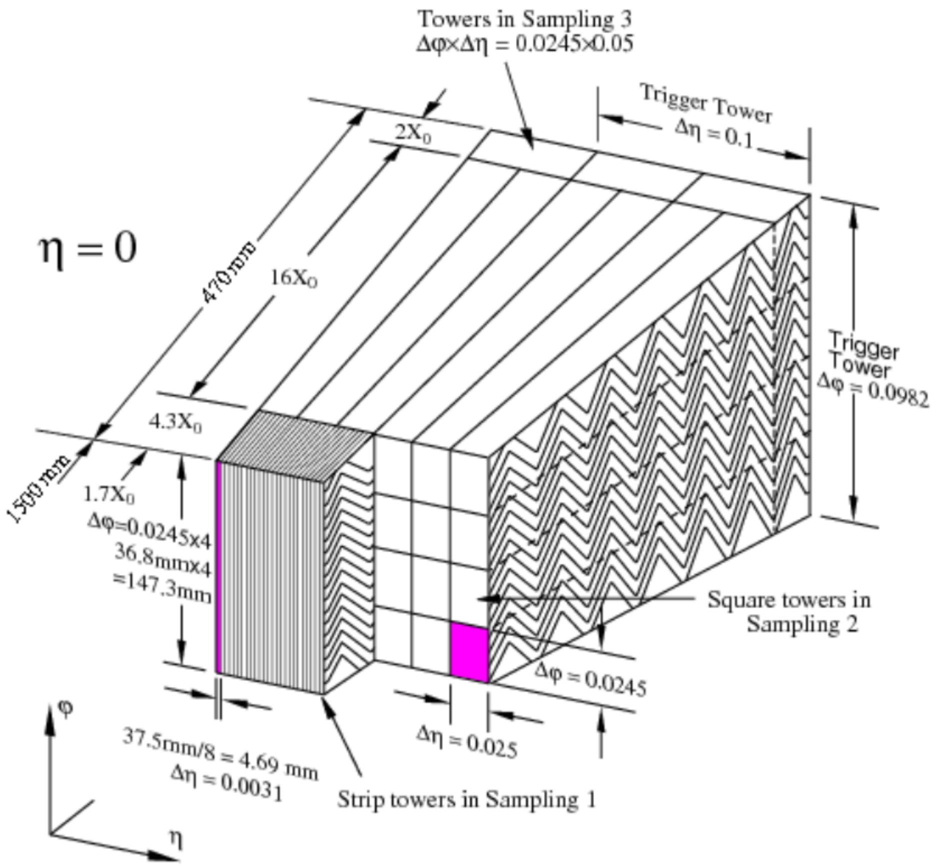


Figure 2.7: A sketch of the LAr EM calorimeter [50].

$\Delta\eta \times \Delta\phi$. The second layer has a wider segmentation, 0.025×0.025, and absorbs the bulk of the EM shower. The last layer has even lower granularity in η, 0.05×0.025, and serves to pick up the EM shower tails. In total, these three layers cover more than 22 radiation lengths (X_0) and have 101,760 total channels.

The EMEC also uses lead-stainless steel as the passive material and is comprised of two concentric wheels, an inner wheel covering $1.4 < |\eta| < 2.5$, and an outer wheel covering $2.5 < |\eta| < 3.2$, and covers > 24 X_0. Between the EMB and EMEC, a transition crack spans $1.37 < |\eta| < 1.52$, which is also used to service the detector.

In front of the calorimeter cryostats, a presampler is used to correct for energy lost before particles reach the calorimeters. The presampler uses thin copper electrodes and LAr as the active material, with a layer of 1 cm thickness in the barrel and 5 mm in the end-caps, and granularity of 0.025×0.1 in $\Delta\eta \times \Delta\phi$ [50].

A sketch of the composition of the EM calorimeter can be found in Figure 2.7.

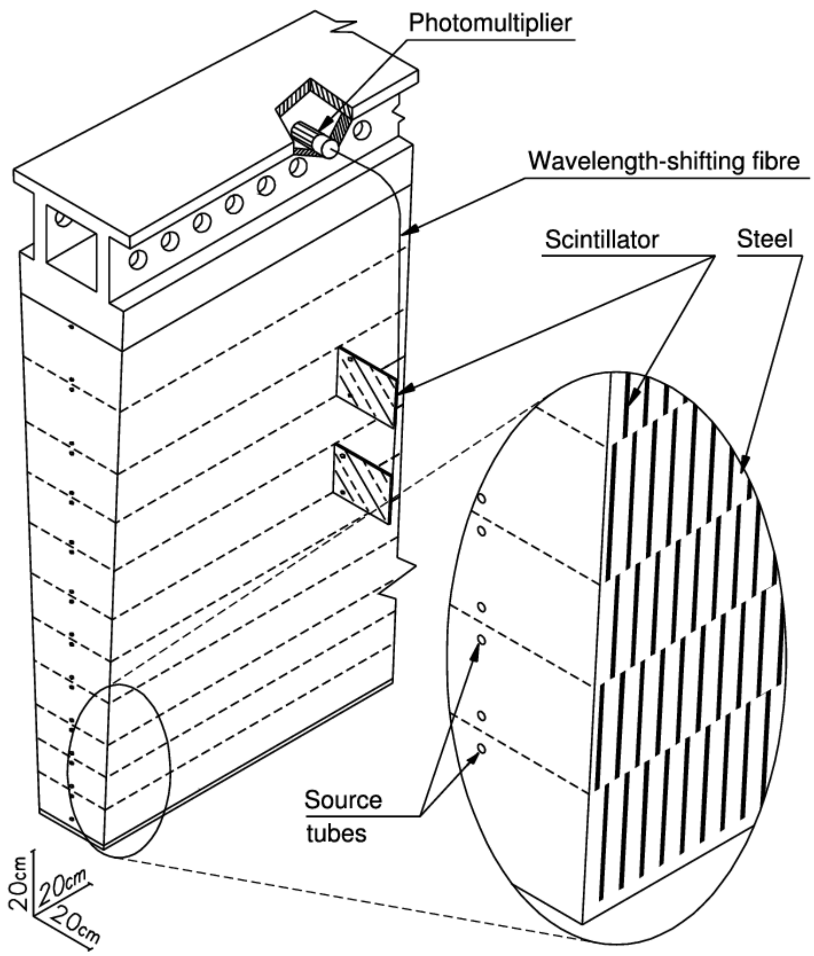


Figure 2.8: A sketch of the hadronic Tile calorimeter [52].

Hadronic Calorimeter

As hadrons interact with matter, their scattering produces secondary hadrons as well as protons and neutrons. These also decay and this cascade, known as a hadronic shower, will continue until the energy of the decay products is small enough to be halted by ionization energy loss or absorbed in a nuclear process.

The hadronic Tile Calorimeter is designed to measure the energy of particles by using these decays, employing steel as the absorber and scintillating tiles as the active material. The scintillators transmit light via fiber optics to photomultiplier tubes (PMTs). The hadronic calorimeter is composed of three regions, the central barrel, covering $|\eta| < 1.0$, and two extended barrels on either side of the central barrel, covering $|\eta| > 1.0$. Each of these three regions have 64 azimuthal modules, and 3 layers longitudinally. The first two layers have granularity 0.1×0.1 in $\Delta\eta \times \Delta\phi$, while the third is 0.2×0.1. The thickness covers more than 9.7 interaction lengths (λ). A sketch of the composition of the Tile Calorimeter can be seen in Figure 2.8.

A Hadronic End-Cap (HEC) also works to detect hadronic decays, covering pseudorapidity range of $1.4 < |\eta| < 3.2$. There are two wheels per end-cap, located behind the EMECs and sharing their cryostats. The HEC uses LAr as the active material, and copper plates as the passive material. Each end-cap features 32 wedge shaped modules, and two segments in depth, making four total samplings per end-cap [50].

Forward Calorimeter (FCal)

In order to get precise measurements of missing transverse energy (MET)[3], full η coverage is necessary. The FCal covers the most forward regions, covering $3.1 < |\eta| < 4.9$. It is three modules deep, covering 10 λ, the first module made of copper, targeting EM depositions, and the outer two made of tungsten targeting hadronic depositions.

[3]MET is not used in the present work, however is used in many ATLAS analyses. It aims to measure those particles which do not interact with the detector, such as neutrinos or many BSM signatures. Since the two colliding partons (defined in Section 3.1.1) cancel out each other's momentum, the initial transverse momentum is 0. Because of the law of conservation of momentum, the sum of momentum of the final state particles must also sum to 0. If the reconstructed particles do not, then the remainder is defined as MET.

2.2.5 Muon Spectrometer

The Muon Spectrometer (MS) [53] is the outermost layer of the ATLAS detector, aiming to track muons via independent triggering (for $|\eta| < 2.4$) and momentum measurement (for $|\eta| < 2.7$). Triggers based on the MS are combined with information from the ID to make the final trigger decision, so good timing resolution is imperative.

Timing information in the MS is collected in the Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGCs). These both feature timing resolution better than 4.5 ns. The RPCs cover the barrel ($|\eta| < 1.05$) region while the TGCs covers the end-cap ($1.0 < |\eta| < 2.4$).

Momentum information is provided through Al tubes 30 mm in diameter and filled with a 0.05 mm gold-plated tungsten wire in the center, called a Monitored Drift Tube Chamber (MDT). Three layers of Monitored Drift Tube Chambers (MDTs) cover the central region of the MS, at $|\eta| < 2.7$. In the forward region, $2.0 < |\eta| < 2.7$, the first MDTs layer is replaced by multi-wire proportional chambers with precision cathode strips, known as Cathode Strip Chambers (CSCs). A schematic of the MS is found in Figure 2.9.

2.2.6 Trigger and Data-Acquisition

Dataflow and storage are a major challenge for ATLAS. Bunches of protons interact every 25 ns, a rate of 40 MHz. On average, a raw event contains about 1.6 MB [55]; recording all of those interactions would not be possible, nor desirable since they are dominated by soft collisions processes which do not yield interesting physics. To isolate events which are desirable to record, ATLAS utilizes a two-stage trigger system [56].

First, the Level 1 (L1) trigger receives events at 40 MHz and makes quick decisions based on coarse information. This trigger system is hardware based, constructed out of custom

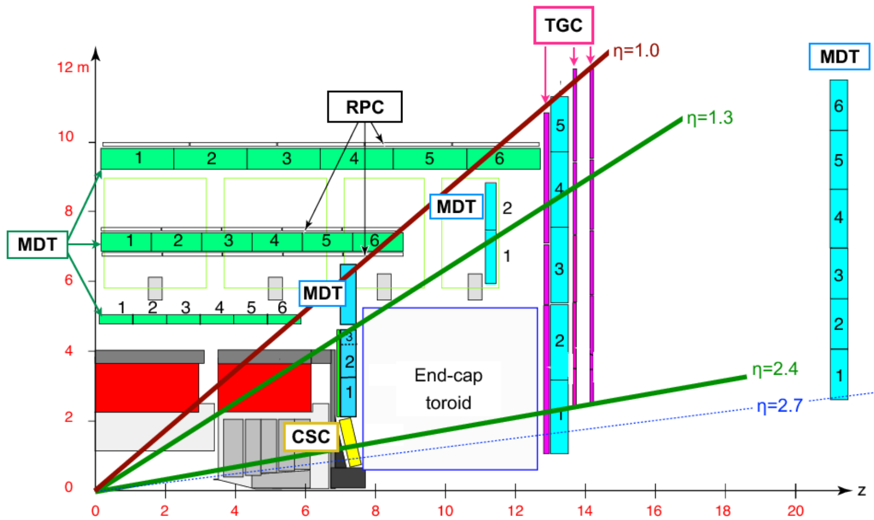


Figure 2.9: A schematic of a quarter of the MS. The described systems, the RPC, TGC, MDT, and CSC are labeled [54].

electronics, and targets specific features in the calorimeter or muon system. From hits in each detector, candidate physics objects (electrons, photons, muons, jets, taus, and MET) are formed, and events are accepted or rejected based on features of these candidates in order to get to an average event rate of between 75 and 100 kHz, a reduction factor of about 400. Additionally in this process, Regions of Interest (ROIs), cones in $\eta \times \phi$ containing interesting physics, are identified and passed down the trigger chain.

The second stage of the trigger is software-based, known the High-Level Trigger (HLT). On average, the HLT has 200 µs per event to make decisions. It uses the ROI provided from L1 to make decisions based on fast algorithms by reconstructing partial events within that region, reconstructing just 2%-6% of the total event data. At this stage, the event rate is reduced to around 3 kHz.

Events that pass this stage then are reconstructed with offline-like algorithms, allowing for triggering on sophisticated signatures (e.g. MET) and the final trigger decision is made using those objects. This produces a final output between 500 and 1000 Hz. Events selected by the HLT are saved to the offline system based on their trigger decision, known as data streams. A schematic of the pipeline for trigger decisions in conjunction with the DAQ system is shown in Figure 2.10.

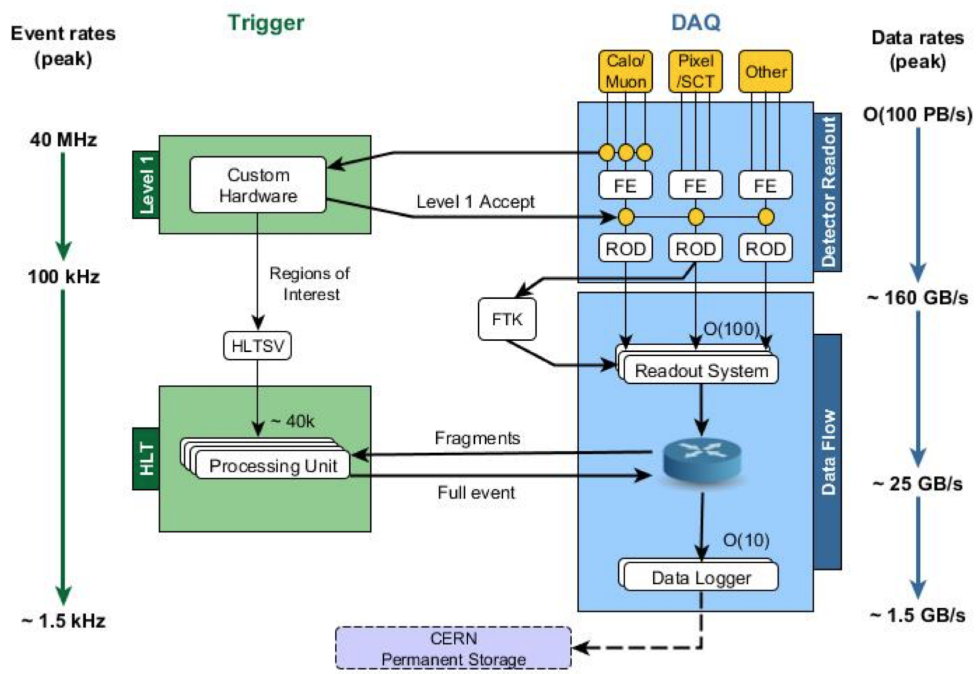


Figure 2.10: The ATLAS Trigger and DAQ systems during the Run 2 data taking period. Inputs from the various subsystems start at the top right, where they are then passed to the L1 trigger, shown in the upper left. Events passing that are sent to the HLT (lower left), while simultaneously data are passed from the Readout Drivers (RODs) to the Readout System (ROS) on L1 accepts. They are buffered and available for the HLT to sample. These accepted events are then sent to storage via the Data Logger. Event and data rates are shown for each step on the left and right, respectively [57].

CHAPTER 3

EVENT SIMULATION AND RECONSTRUCTION

In theory, there is no difference between theory and practice. In practice there is.
— Benjamin Brewster

3.1 Monte Carlo Simulation

To properly understand and interpret results, comparisons to theoretical predictions must be made. In the context of particle colliders, this means an understanding of both the underlying process as well as predictions of detector effects. In order to do this, ATLAS utilizes simulation based on Monte Carlo (MC) methods.

A MC simulation attempts to model a process by taking a function's underlying probability distribution and sampling it randomly. The steps of MC integration that compose the ATLAS simulation chain evolve serially, and each step is Markovian: the process is completely dependent *only* on the prior step, so sampling is performed on its posterior distribution. Producing MC from a set of subsequent Markovian processes is known as Markov Chain MC (MCMC), which is used to simulate complex processes, such as particle collisions.

The ATLAS simulation framework is designed to produce MCMC in this manner, evolving in steps of event generation (Section 3.1.1), parton showering (Section 3.1.2), hadronization (Section 3.1.3), detector simulation (Section 3.1.4), and finally reconstruction of physics objects (Section 3.2).

3.1.1 Event Generation

The first step of the ATLAS simulation is generating the hard-scatter process. When simulating collisions, parton distribution functions are used. A parton refers to one of the strongly interacting particles (quarks and gluons) that comprise protons. In *pp* collisions, these constituents are what actually interact, so distribution functions model this. These distribution functions describe the probability of finding a parton carrying a given fraction of the proton momentum.

Event generators use *matrix elements*, which take into account the partonic cross-section to a specified order approximation. The matrix element is returned in the *Les Houches Accord* format [58], which was designed to create a uniform event generator output. This format can then be interfaced to tools for parton shower and hadronization.

The ATLAS software contains many event generators. For the samples used in the presented analysis, the following generators are used:

Herwig++: A general-purpose event generator [59].

MadGraph: A matrix element generator, generating tree-level matrix elements for Lagrangian-based models [60].

Pythia: A general-purpose event generator [61].

Powheg: A NLO event generator that is built to overcome the problem of negatively weighted events [62].

Sherpa: A general-purpose simulation tool, containing all necessary components for a factorized description of scattering events [63].

3.1.2 Parton Showering

After interacting, the parton showering must then be simulated. Hard scatter processes emit QCD bremsstrahlung radiation in the form of gluons. Emitted gluons carry color charge, which also can further radiate. This process is known as parton showering, and generators approximate these higher-order QCD corrections through a chain of one-to-two parton branching. This is performed iteratively until reaching the non-perturbative regime, around 1 GeV.

Three event generators in the ATLAS software have parton showering built in: Pythia, Herwig, and Sherpa. Other event generators are able to interface with one of those three for parton showering and hadronization. These event generators order the parton shower to avoid divergences and double-counting. Pythia and Sherpa order showers such that the highest p_{T} emissions occur first. Herwig utilizes angular ordering, such that the widest emissions occur first, and proceed toward smaller angles.

In addition to the hard scatter process, these generators simulate the underlying event. This process accounts for effects due to rescattering or multiple parton interactions within the colliding protons.

3.1.3 Hadronization

Once partons have reached sufficiently low energies, they form hadrons. This is known as hadronization[1]. Hadronization is a non-perturbative process, that relies on phenomenological models tuned on data. Commonly used models are the Lund string model [64] (Pythia) or clustering models [65] (Herwig, Sherpa). ATLAS has several set tunes for each event

[1] Also “jet fragmentation.”

generator, as well as Parton Distribution Function (PDF) sets. In this work the following are used:

- The A14 Pythia8 tune (ATLAS 2014) [66]. Tunes for four separate LO PDF sets were performed, notably NNPDF23LO.
- The UEEE5 tune [67], implemented in Herwig, Pythia, and Sherpa. This accompanies hard collisions by Underlying Event (UE) simulation to mirror minimum bias interactions. In this work, the C10 PDF sets [68], derived by global analysis of hard scattering data in the general-mass framework of perturbative QCD are used with this tune.
- The H7-UE-MMHT tune [69] for Herwig, which provides a good description of the UE and double parton scattering.
- The H7-MMHT2014LO tune [70], also for Herwig, containing PDF sets of the proton at NLO, NNLO, and LO.

Through hadronization, simulation is agnostic of detector, describing just the particle interaction. A schematic of the evolution of a simulated event from event generation through hadronization is shown in Figure 3.1.

3.1.4 Detector Simulation

After hadronization, the model is a full detector-independent description of a process; next, the simulation must model how the particle will interact with the ATLAS detector. In order to do this, a component-level model of the ATLAS detector is implemented in Geant4 [72], and particles are propagated through this model. The energy deposition

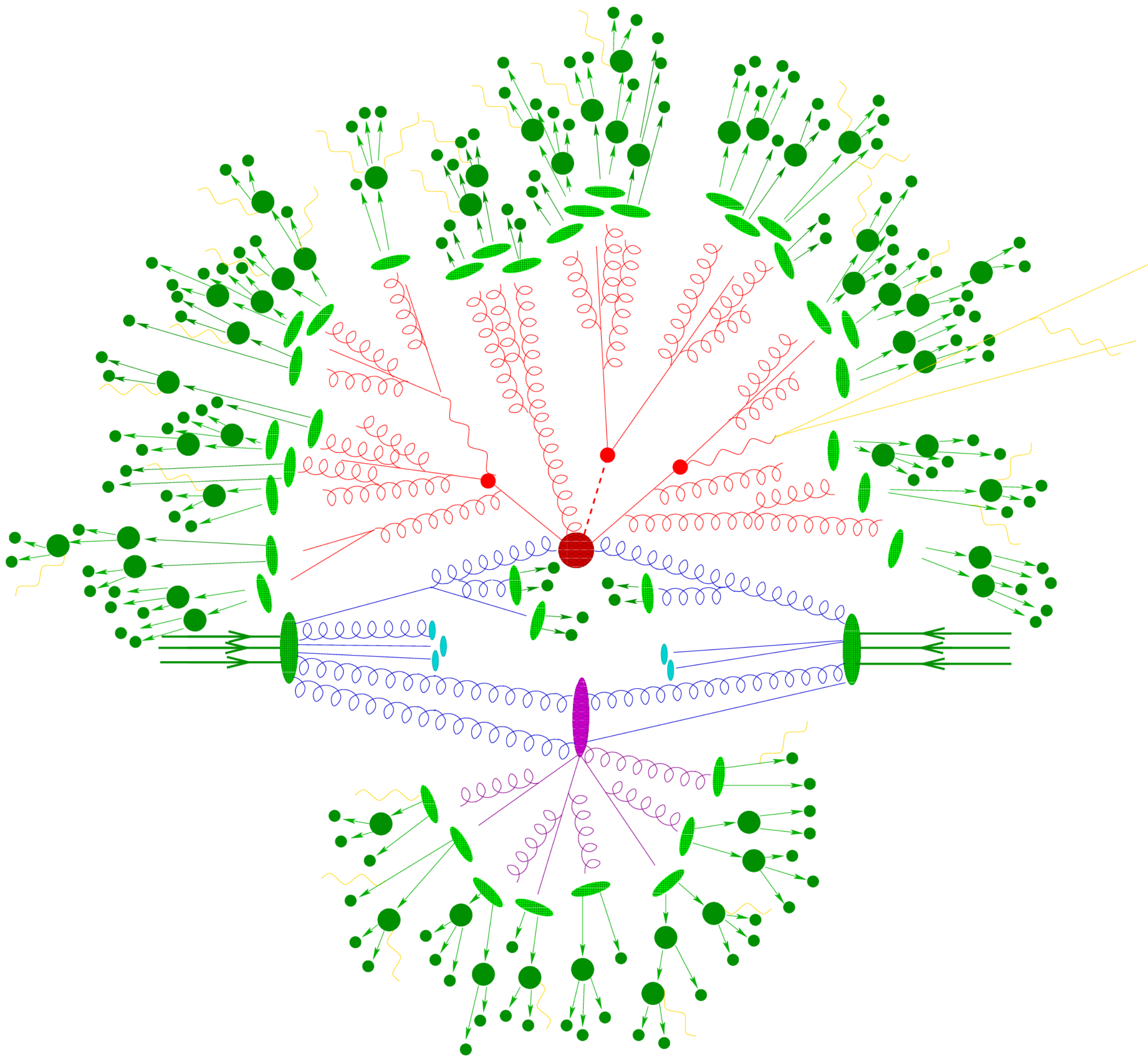


Figure 3.1: Schematic of the procession of a simulated hadron-hadron collision through hadronization. The red circle represents the matrix element, simulated by an event generator. The emitted red lines represent the bremsstrahlung radiation, which is simulated by a parton shower. The purple oval represents a secondary vertex. Hadronization is represented by the light green circles, in which hadrons are produced (dark green) as well as photon radiation (yellow) [71].

in each detector component is calculated stochastically and output as a collection of hits. The electrical response for these hits in each detector component is simulated, and output in a format matches that from actual data collection. From this, physics objects may be reconstructed via the same methods as real data, outlined in Section 3.2.

While this method is the most accurate method of simulation, it is CPU-intensive. Methods of faster simulation have been developed to mitigate this problem. The Atlfast-II (AFII) simulation reduces simulation time by an order of magnitude through using FastCaloSim for calorimeter simulation [73]. FastCaloSim uses parameterizations of electromagnetic showers to directly deposit energy rather than explicit simulation.

3.2 Object Reconstruction

Both the full simulation chain and actual data collection result in a set of hits within the ATLAS detector. Reconstruction is the method by which these hits are defined into final physics objects useful for analysis. This is performed independently for each type of signature. The following sections discuss the reconstruction of objects used in the analysis presented in this thesis. For the Higgs decays studied, the final state physics objects are a pair of photons (discussed in Section 3.2.1) and a pair of jets (discussed in Section 3.2.2). To account for energy losses in jets, corrections using muons (discussed in Section 3.2.3) are applied. In addition to the objects used in this analysis, ATLAS reconstructs tau leptons and MET (corresponding to neutrino signatures in SM analyses, as well as a number of BSM signatures), not discussed here.

3.2.1 Electromagnetic Signatures: Electrons and Photons

Interaction

Photons and electrons interact with the LAr calorimeter, leaving an electromagnetic signature in a group of neighboring cells known as *showering*. Their reconstruction is performed in parallel due to the similarity in signature.

EM signatures can be the result of three different sources:

- Electrons: Being a charged lepton, in addition to interactions in the EM calorimeter, electrons interact with the ID, leaving a track that can be extrapolated to the EM cluster.

- Converted Photons: Photons that pair produce electron-positron pairs in the detector prior to the calorimeter have a EM cluster matched to a secondary vertex. The rate of photons that convert varies with η, as more detector material is traversed at higher angles. At low $|\eta|$, about 20% of photons convert, where above $|\eta| = 2.3$, up to 65% convert [74].

- Unconverted Photons: Unconverted photons do not interact with the ID, and thus are not matched to an electron track or secondary vertex.

As a result of bremsstrahlung radiation, an electron can lose a substantial portion of its energy as it traverses the detector material. In this process, the electron radiates a photon, which itself can radiate an electron-positron pair. Should these occur in the beampipe or ID, there may be multiple track candidates from the same electron. The process by which these depositions become final analysis objects proceed in steps of cluster reconstruction, track reconstruction, supercluster reconstruction, then construction of the final analysis object.

Cluster Reconstruction

The algorithm to build electron and photon candidates is a dynamic algorithm, building variable-sized clusters. This is in contrast to previous methods, which use a fixed "sliding window" definition [75]. The algorithm starts from a *seed* cell, which is identified by a signal larger than a high signal threshold relative to underlying electronic noise, $S\sigma_{noise}$. The seed and its neighbors are added to the cell, known as a *proto-cluster*. Then, all neighbor[2] cells are scanned for a signal larger than a moderate threshold, $N\sigma_{noise}$. The subsequent neighbors of cells meeting this criteria are added to the cluster, and themselves are evaluated if they pass the moderate signal threshold, provided that they pass a cell filter defined by $P\sigma_{noise}$. In the case where two clusters have an overlapping cell, the clusters are merged. ATLAS uses default values of $S = 4, N = 2, P = 0$ in this algorithm. The choice of $P = 0$ means that no filter is applied, and candidates passing the $2\sigma_{noise}$ have all neighboring cells added to the cluster. This method of clustering is known as *topological* (topo) clustering [76].

For EM objects, only energy in the EM calorimeter is used[3], denoted EM energy. Clusters considered for must pass a EM fraction (f_{EM}) criteria, in which the EM energy must be greater than 50% of the total cluster energy. This cut rejects ∼60% of pile-up clusters with no affect on true electron topo-clusters. Clusters passing this criteria are known as EM topo-clusters.

Track Reconstruction

Tracks are constructed out of hits within the ID, in which three dimensional space-points are constructed out of clusters. Three sets of space-points are formed into track seeds, which then undergo stages of pattern recognition, ambiguity resolution, and TRT extension. Track candidates with $p_{\mathrm{T}} > 400$ MeV are fit with the ATLAS Global χ^2 fitter [77].

[2]In the same sampling layer, neighboring means cells are adjacent. In adjacent layers, neighboring cells must have overlap in the (η, ϕ) plane.

[3]In the crack in the EM calorimeter, $1.37 < |\eta| < 1.63$, the presampler and scintillator between the calorimeter cryostats are used.

For tracks with at least four silicon hits that match to an EM cluster, a secondary fit is performed using a Gaussian-sum filter [78]. This filter is is a generalization of the Kalman filter [79], taking into account non-linearities resulting from bremsstrahlung radiation. The separation between the cluster barycenter and the position of the track extrapolated from the perigee to the calorimeter must satisfy[4] $|\eta_{cluster} - \eta_{track}| < 0.05$, and $-0.10 < q \cdot (\phi_{track} - \phi_{cluster}) < 0.05$ for reconstructed track charge, q.

If multiple tracks are matched to a cluster, they are ranked. First, priority is given to tracks with pixel hits over those with just SCT hits. Second, tracks are ordered in ΔR to the cluster. If multiple tracks have small ΔR differences, then the track with the most pixel hits is selected.

Supercluster Reconstruction

The object ultimately defined as an electron or photon is a supercluster, which is a grouping of topo-clusters. Superclusters are formed in a two-stage process.

First, topo-clusters are tested for use as a seed for the supercluster. Topo-clusters are sorted by E_{T}, then sequentially tested to be seed candidates. An electron seed candidate must have $E_{\mathrm{T}} \geq 1$ GeV and a matching track; a photon seed must have $E_{\mathrm{T}} \geq 1.5$ GeV.

Second, the topo-clusters near the identified seed are selected to be added to the supercluster. These extra topo-clusters, called satellite clusters, can arise from either splitting or through bremsstrahlung radiation. A window of $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$ around the barycenter is considered. A wider window of $\Delta\eta \times \Delta\phi = 0.125 \times 0.300$ is used for electrons if the best-matched track for the seed cluster is the same best-matched track for the satellite cluster. For photons that have converted, a cluster is added if its best-matched track belongs to the conversion vertex matched to the seed. The supercluster reconstruction logic is shown in Figure 3.2.

[4]This requirement is asymmetric because clusters can measure energy from radiated photons that tracks miss.

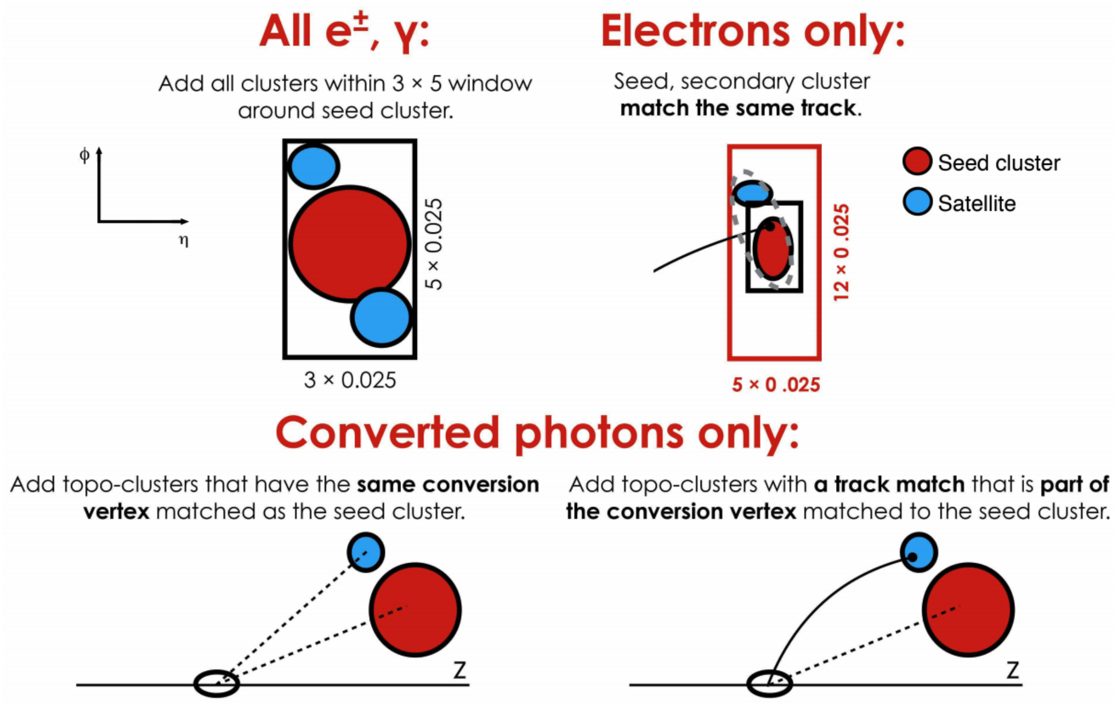


Figure 3.2: Pictorial representation of supercluster algorithm for EM objects [74]. The resolution of 0.025 is due to the granularity of the EM calorimeter.

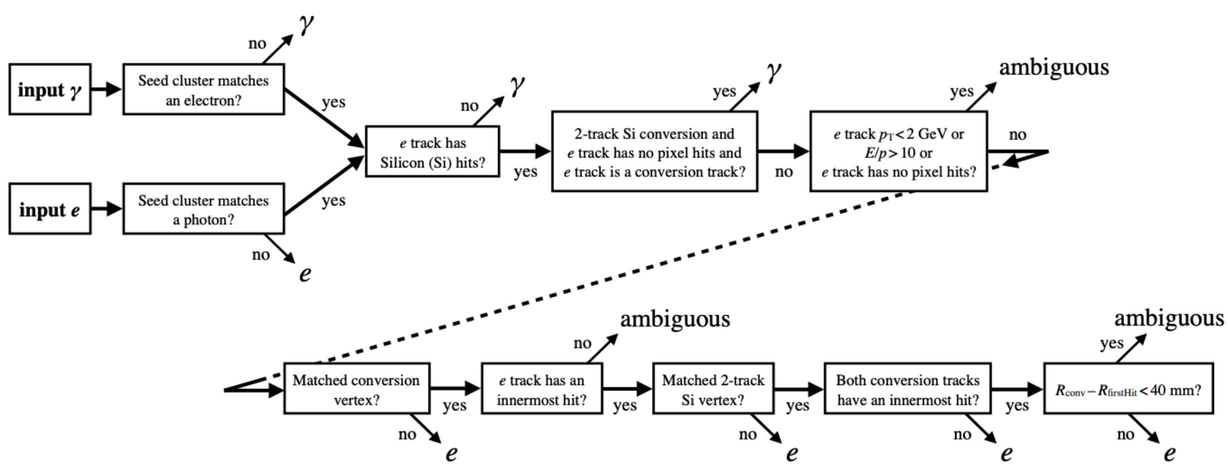


Figure 3.3: Logic flowchart for superclusters that are reconstructed as both an electron and a photon [74].

Analysis Objects

Once superclusters are constructed, they make up the final analysis objects, but due to

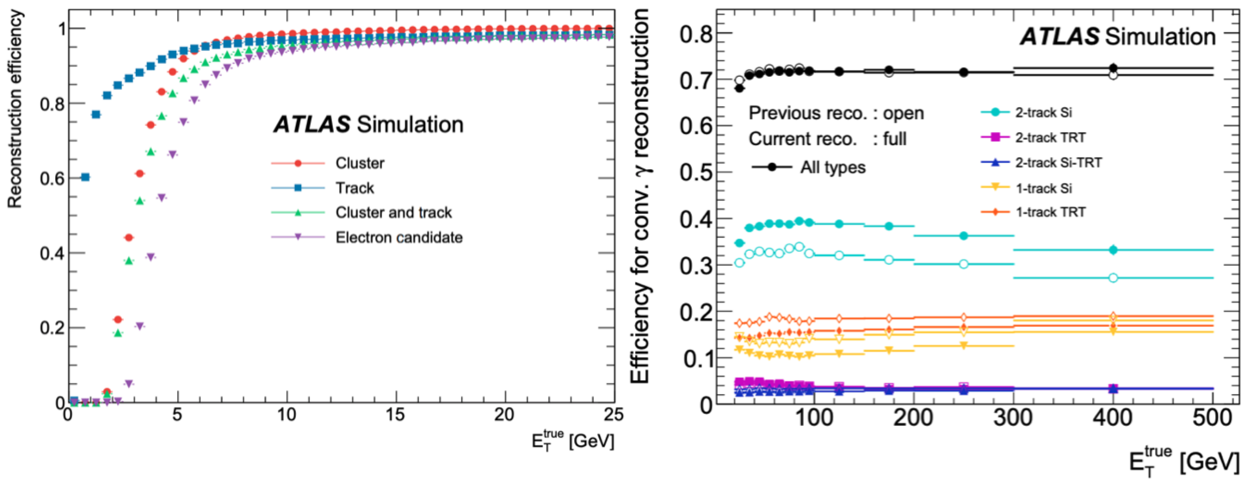


Figure 3.4: The reconstruction efficiency as a function of E_{T} for clusters, track, cluster and track, and electrons (left) and the various photon conversion types (right) [74].

the logic of supercluster formation, a seed could be both an electron and photon candidate. In these cases, after calibration and positron correction, the logic shown in Figure 3.3 is applied. In cases denoted ambiguous, both a photon and electron are formed, and these are marked as "ambiguous" for analysis-specific requirements. The electron and converted photon reconstruction efficiency is shown in Figure 3.4 as a function of truth E_{T}.

3.2.2 Jets

As discussed in Section 1.1.1, strongly interacting particles cannot exist independently due to color confinement. Due to this, partons created in pp collisions shower, a process where they produce additional strongly interacting particles through splitting and radiation. This creates a proliferation of nearly collinear subsequent partons. Ultimately, once they've reached a sufficiently low energy, about 1 GeV, they hadronize in order to create a colorless state, such as mesons or baryons that deposit energy into the calorimeters. This spray of interactions all proceeding in the same general direction as the initial parton is known as a "jet."

In principle, jets can be built from any set of 4-vector objects. The standardized jet definition [80] requires that jets be:

- Simple to implement in both experimental analysis and theoretical calculation.
- Defined at any order of perturbation theory, and yield a finite cross-section.
- Yield a cross-section relatively insensitive to hadronization.

For this work, they are constructed via topological clusters. The topological clustering logic described in detail in 3.2.1 can also be applied to the hadronic calorimeter and jets can be formed out of these topo-clusters or EM topo-clusters [76]. Two classes of jet definitions exist. The first is cluster-based, which defines the jet by combining sets of four-vector objects until the distance between objects surpasses a stopping condition. The second is cone-based, which define jets as the sum of momenta within a defined radius, realizing jets as energy flow in a specified direction. ATLAS uses one such cone-based algorithm, known as the Anti-k_t algorithm [81]. The cone is built from topological clusters using a fixed radius[5], $R = 0.4$.

Since all fragmenting strongly interacting particles appear as jets in the detector, an important process is *flavor tagging*, which aims to classify the type of particle that produced a given jet. Paramount to the presented analysis is the process of b-tagging[6], the method of classifying jets which come from b-quarks. There are several algorithms to perform this classification, notably a BDT named MV2 [82] and a feed-forward deep NN[7] known as DL1 [82]. From this score, b-tagging working points are established and calibrated based on their identification efficiency.

[5]In analyses that probe high mass regimes, two jets may be overlapping and merge to produce one single large radius jet. This is known as a "boosted" topology, where the radius parameter is typically $R = 1.0$. The presented analysis is not sensitive at high mass, thus such topologies are not considered.

[6]In addition to b-tagging, ATLAS has also produced discriminators for charm jets (c-tagging).

[7]See Appendix A for an explanation of these multivariate techniques.

3.2.3 Muons

Muons are a minimum-ionizing particles, leave very little energy in the calorimeter, and are not stopped by any of the material in ATLAS. Thus, muons are reconstructed using a combination of track-based information from the ID and the MS. In the ID tracks for muons are identical to other charged particles, outlined in Section 3.2.1. In the MS, each detector region (i.e. MDTs, CSCs, etc.), builds track segments from aligned hits in the bending plane of the detector, found via a Hough transform [83]. Track candidates are formed by fitting together hits from different track segments via a combinatorial search outlined in Reference [84]. Then an overlap removal algorithm finds the optimum track assignment. Hits are fitted with a global χ^2 fit. Afterward, a hit recovery and subsequent track refit are performed if necessary.

Ultimately four muon types are defined [84]:

- Combined (CB) muon: Tracks reconstructed in the ID are formed into a combined track using a global refit incorporating hits from both subdetectors.

- Segment-tagged (ST) muon: An ID track is extrapolated to a local track segment in the MDT or CSC, accounting for muons that only cross one layer of MS chambers.

- Calorimeter-tagged (CT) muon: An ID track matched to a energy deposit in the calorimeter compatible with a minimum-ionizing particle. This muon type compensates for portions of the MS that are only partially instrumented, near $|\eta| \approx 0$.

- Extrapolated (ME) muons: These use only track information from the MS, where parameters must be loosely in agreement with the location of the IP. ME muons are useful for acceptance in $2.5 < |\eta| < 2.7$, which is uncovered by the ID.

From these muons, four selections are created (*loose*, *medium*, *tight*, and *high-p_T*). These are defined based on varying requirements on the type of muon considered, number of hits, χ^2 fit, η, q/p significance, and p_{T}. The efficiencies for the *loose* and *medium* selections are $> 98\%$, and the *tight* selection ranges from 90% to 98% efficient. Figure 3.5 shows these efficiencies as a function of η.

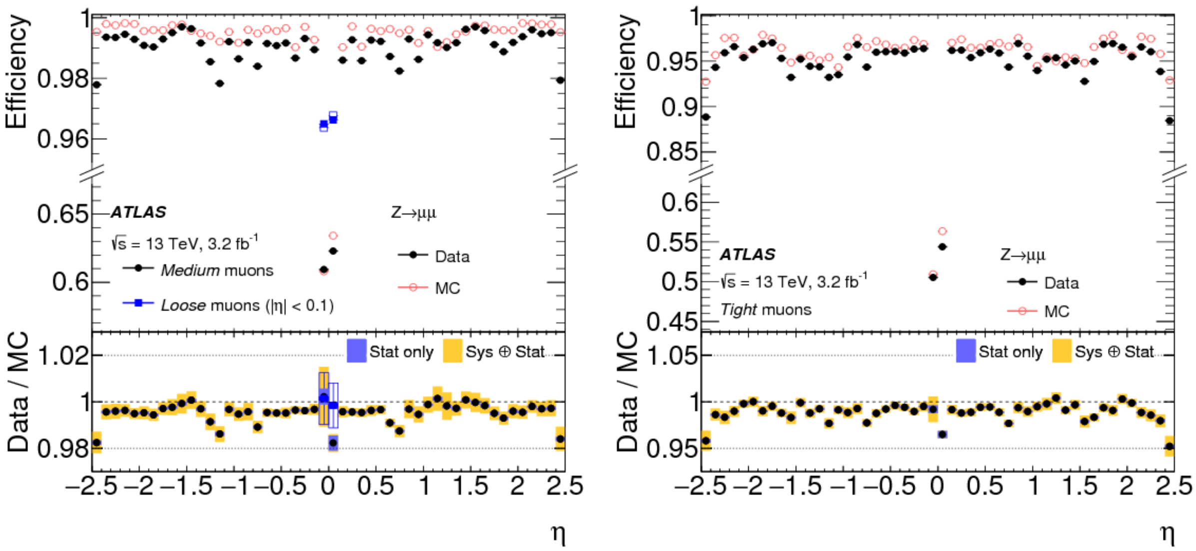


Figure 3.5: The reconstruction efficiency for muons with $p_{\mathrm{T}} > 10$ GeV as a function of η. The left panel shows the *loose* selection, the right shows the *tight* selection [84].

CHAPTER 4

PHOTON IDENTIFICATION OPTIMIZATION

I've found you've got to look back at the old things and see them in a new light.
— John Coltrane

Photon identification in ATLAS is carried out through applying cuts to 9 shower shape variables that explain the longitudinal and lateral showering development of electromagnetic interactions in the calorimeter. These include widths of energy depositions in the EM calorimeter (w_{η_2}, w_{s3}), energy ratios of depositions in the EM calorimeter (R_η, R_ϕ, F_{side}), the energy difference and ratios of the strips (ΔE, E_{ratio}), and variables describing the energy that deposits past the EM calorimeter, in the hadronic calorimeter, called leakage (R_{had}, R_{had_1}). A schematic of all these variables can be found in Figure 4.1. Additionally, a table describing these variables can be found in Table 4.1. Two sets of cuts are defined, the *tight* and *loose* working points.

Cuts on these variables are derived in bins of $|\eta|$. The intervals are 0.0, 0.6, 0.8, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37, with a bin defined between each step (e.g. $[0.0, 0.6)$, $[0.6, 0.8)$, etc.) to account for the geometry of the calorimeter and the effects upstream material has on the shower shapes. The region $[1.37, 1.52)$ is omitted, due to the transition crack between where the EMB and EMEC lies. In this region, there is considerable material before the EM calorimeter, and is not considered for precision photon identification. The most recent iteration of the tight photon identification introduced binning in p_{T}, where bins are defined at 25, 30, 40, 50, 60, 80, 100, 125, 150, 175, 250, 500, and 1000 GeV. This provides increased signal efficiency at low p_{T}, and improved background rejection at high p_{T}.

This chapter describes an optimization to photon identification through two approaches. First, through adding additional discriminating variables to be used as inputs. This is done

Category	Description	Name	*loose*	*tight*
Acceptance	$\|\eta\| < 2.37$, with $1.37 < \|\eta\| < 1.52$ excluded	–	✓	✓
Hadronic leakage	Ratio of E_T in the first sampling layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $\|\eta\| < 0.8$ or $\|\eta\| > 1.37$)	R_{had_1}	✓	✓
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \|\eta\| < 1.37$)	R_{had}	✓	✓
EM Middle layer	Ratio of 3×7 $\eta \times \phi$ to 7×7 cell energies	R_η	✓	✓
	Lateral width of the shower	w_{η_2}	✓	✓
	Ratio of 3×3 $\eta \times \phi$ to 3×7 cell energies	R_ϕ		✓
EM Strip layer	Shower width calculated from three strips around the strip with maximum energy deposit	$w_{s\,3}$		✓
	Total lateral shower width	$w_{s\,\mathrm{tot}}$		✓
	Energy outside the core of the three central strips but within seven strips divided by energy within the three central strips	F_{side}		✓
	Difference between the energy associated with the second maximum in the strip layer and the energy reconstructed in the strip with the minimum value found between the first and second maxima	ΔE		✓
	Ratio of the energy difference associated with the largest and second largest energy deposits to the sum of these energies	E_{ratio}		✓

Table 4.1: List of discriminating variables used in the present photon identification [86].

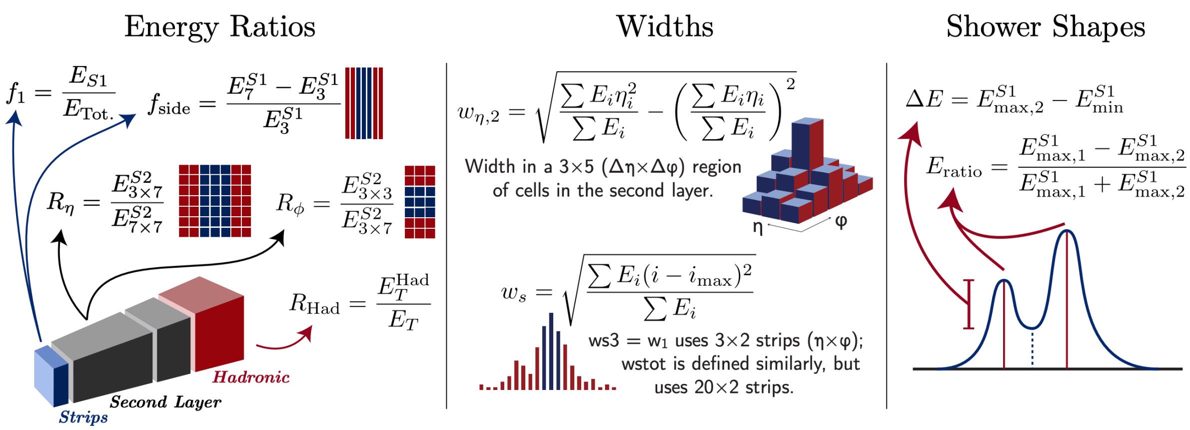


Figure 4.1: Schematic of the shower shape variables used in the present cut-based photon identification [85].

through investigating the use of observables of topological clusters, described in Section 4.2. Second, alternative methods of classification are studied, replacing the current univariate cut-based approach with a multivariate algorithm. This is shown in Section 4.3, in which a BDT and a NN are studied.

4.1 MC Samples

In optimization of photon identification, the aim is to discriminate prompt photons from various background sources. The predominant background is from decays inside of hadronic jets which result in non-prompt photons, such as $\pi^0 \to \gamma\gamma$.

For these studies, a sample of leading-order γ+jet events from $qg \to q\gamma$ and $q\bar{q} \to g\gamma$ hard scattering, as well as events with prompt photons from quark fragmentation in QCD dijet events is used. For background non-prompt photons in jets, a jet-jet sample with all tree-level $2 \to 2$ QCD processes is used, with γ+jet events containing prompt photons removed. Both samples are generated using Pythia8 with the `A14` tune and `NNPDF2.3LO` PDF set. The simulation matched the detector conditions in the 2015 and 2016 data taking period. In both cases, an inclusive OR of loose single photon triggers is applied, with thresholds running as low as 10 GeV[1].

4.1.1 Event Selection

Selected events are required to have a photon with $|\eta| < 2.37$, excluding the crack region ($1.37 < |\eta| < 1.52$), and $p_{\mathrm{T}} > 25$ GeV. In each event, the photon with the highest p_{T} is considered. For the signal single photon sample, MC truth information is used to require the selected photon to be a prompt photon (*truth-matched*), for the jet-jet background a prompt photon veto is applied to the selected photon (*truth-vetoed*).

4.2 Topological Clusters

Photon reconstruction is performed through a method known as "topological clustering," which is described in Section 3.2.1. In the calorimeter, energy depositions are clustered via this method, then these topo-clusters are merged into superclusters, which are ultimately the detector signature which may be defined as the photon. The reconstructed observables of these topological clusters, or *moments*, can be used to characterize them. This section will discuss the use of these topological cluster moments in photon identification to better reject background from non-prompt photons.

In general, moments are given at order n for a calorimeter cell variable, v_{cell} as

$$\langle v^n_{\mathrm{cell}} \rangle = \frac{\sum_{\{i|E^{\mathrm{EM}}_{\mathrm{cell},i}>0\}} w^{\mathrm{geo}}_{\mathrm{cell},i} E^{\mathrm{EM}}_{\mathrm{cell},i} v^n_{\mathrm{cell},i}}{\sum_{\{i|E^{\mathrm{EM}}_{\mathrm{cell},i}>0\}} w^{\mathrm{geo}}_{\mathrm{cell},i} E^{\mathrm{EM}}_{\mathrm{cell},i}}. \tag{4.1}$$

[1]The triggers applied are HLT_g*_loose

The moments use EM scale cell signals, $E^{\mathrm{EM}}_{\mathrm{cell},i}$, and the summation is performed over i cells in a cluster. To restrict moment calculation to in-time signals, $E^{\mathrm{EM}}_{\mathrm{cell},i}$ must be greater than 0. The geometric signal weights are given by $w^{\mathrm{geo}}_{\mathrm{cell},i}$, which are introduced by cluster splitting. If proto-clusters have multiple local signal maxima with $E^{\mathrm{EM}}_{\mathrm{cell}} > 500$ MeV in the EM sampling layers, the cluster is split. Cells which are neighbors to multiple maxima are assigned in a weighted proportion to the two highest energy clusters, $E^{\mathrm{EM}}_{\mathrm{clus},1}$ and $E^{\mathrm{EM}}_{\mathrm{clus},2}$, after splitting the original topo-cluster. For cells of distances d_1 and d_2 to the center of the maxima, these are given by:

$$w^{\mathrm{geo}}_{\mathrm{cell},1} = \frac{E^{\mathrm{EM}}_{\mathrm{clus},1}}{E^{\mathrm{EM}}_{\mathrm{clus},1} + rE^{\mathrm{EM}}_{\mathrm{clus},2}} \tag{4.2}$$

$$w^{\mathrm{geo}}_{\mathrm{cell},2} = 1 - w^{\mathrm{geo}}_{\mathrm{cell},1} \tag{4.3}$$

$$r = \exp(d_1 - d_2). \tag{4.4}$$

For cells not split between maxima, this value is 1. The used moments are all of order $n = 1$ (centroids) or $n = 2$ (spreads) [76].

Number of Topo-Clusters

As described in Section 3.2.1, superclusters are made up of one or more topo-clusters. Multiple topo-clusters in a supercluster can be a result of cluster splitting or bremsstrahlung radiation. In over 90% of photon events, there is only 1 topo-cluster associated to the photon supercluster, shown in Figure 4.2. As a result, the quantities studied as inputs are all of the topo-cluster with the largest total energy (summed over constituent cells). Additionally, the number of topo-clusters is on average fewer for prompt photons than fakes, so this value is used as an input into the classification.

Shape Information

The size of the topo-cluster indexed i is parameterized by moments in r_i and λ_i, which are defined as:

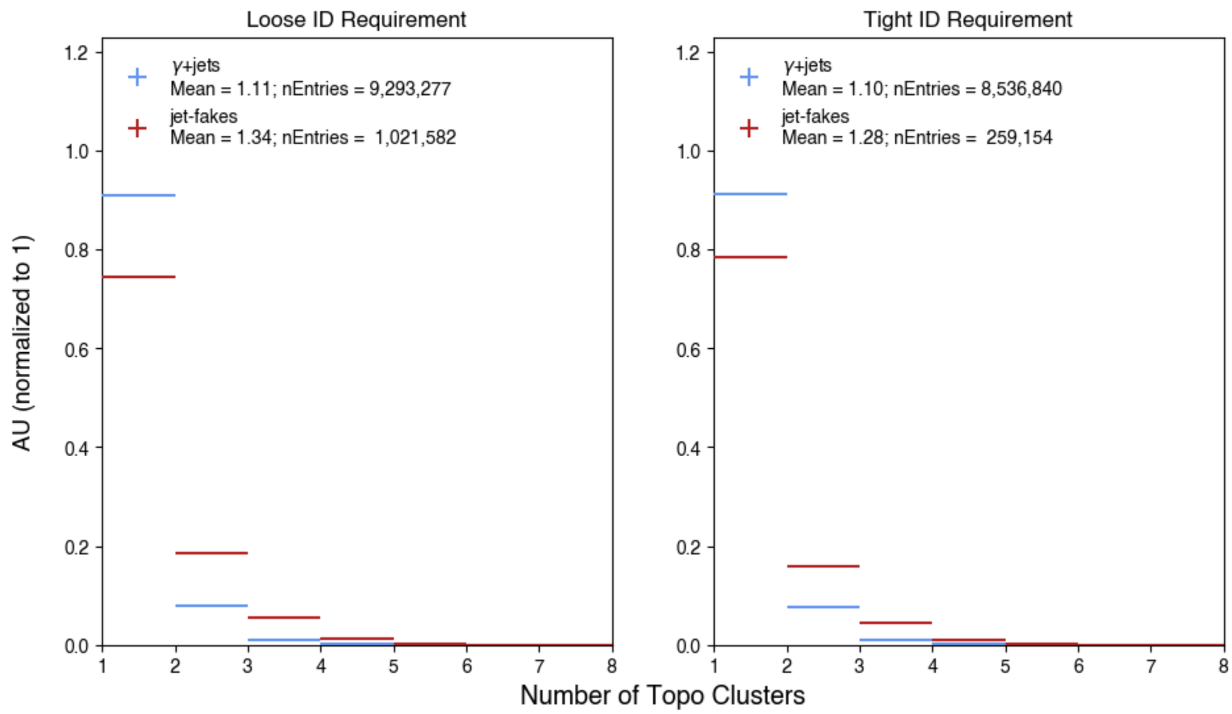


Figure 4.2: Number of topo-clusters in unconverted photon superclusters in truth-matched γ+jet events and truth-vetoed jet fake events. Both prompt and fake photons predominantly just have 1 topo-cluster per supercluster. This is more true for prompt photons, which on fewer topo-clusters on average than fakes.

$$r_i = |(\vec{x_i} - \vec{c}) \times \vec{s}| \qquad \text{(radial distance to shower axis)} \tag{4.5}$$

$$\lambda_i = (\vec{x_i} - \vec{c}) \cdot \vec{s} \qquad \text{(longitudinal distance to shower center of gravity)}, \tag{4.6}$$

where $\vec{s}$ is the shower axis and $\vec{c}$ is the center of gravity. The center of gravity, $\vec{c}$, is calculated using Equation 4.1, using the first moments of the Cartesian coordinates specifying calorimeter cell centers. The locations are in the frame of reference of the detector with the IP at $(x, y, z) = (0, 0, 0)$. The shower axis measures the direction of flight of the particle and is given by:

$$\mathcal{C}_{uv} = \frac{1}{\mathcal{W}} \sum_{\{i|E^{\mathrm{EM}}_{\mathrm{cell},i}>0\}} (w^{\mathrm{geo}}_{\mathrm{cell,i}}, E^{\mathrm{EM}}_{\mathrm{cell},i})^2 (u_i - \langle u \rangle)(v_i - \langle v \rangle) \tag{4.7}$$

for every permutation of $u, v \in \{x, y, z\}$ and

$$\mathcal{W} = \sum_{\{i|E^{\text{EM}}_{\text{cell},i}>0\}} (w^{\text{geo}}_{\text{cell,i}}, E^{\text{EM}}_{\text{cell},i})^2. \tag{4.8}$$

Then, the values of $\mathcal{C}_{uv}$ can be used to create a 3×3 matrix, $\mathbf{C} = [\mathcal{C}_{uv}]$. The shower axis, $\vec{s}$, is the eigenvector of $\boldsymbol{C}$ closest to the direction $\vec{c}$ from the IP to the center of gravity of the topo-cluster. If the angular distance between $\vec{c}$ and $\vec{s}$ is larger than 20 degrees, then $\vec{c}$ is used as the shower axis.

Through this parameterization, the topo-cluster can be analyzed as a spheroid, with the second moments in r and λ as the extensions; $\sqrt{\langle \lambda^2 \rangle}$ is the semi-major axis in depth (along the shower axis), $\sqrt{\langle r^2 \rangle}$ is the semi-major axis in width. The distribution of these two parameters are shown for the leading topo-cluster in Figures 4.4 and 4.5, after applying the current loose and tight photon identification criteria. Differences remaining in these distributions after applying the current identification is information the current criteria does not capture, and can bring additional discriminating power.

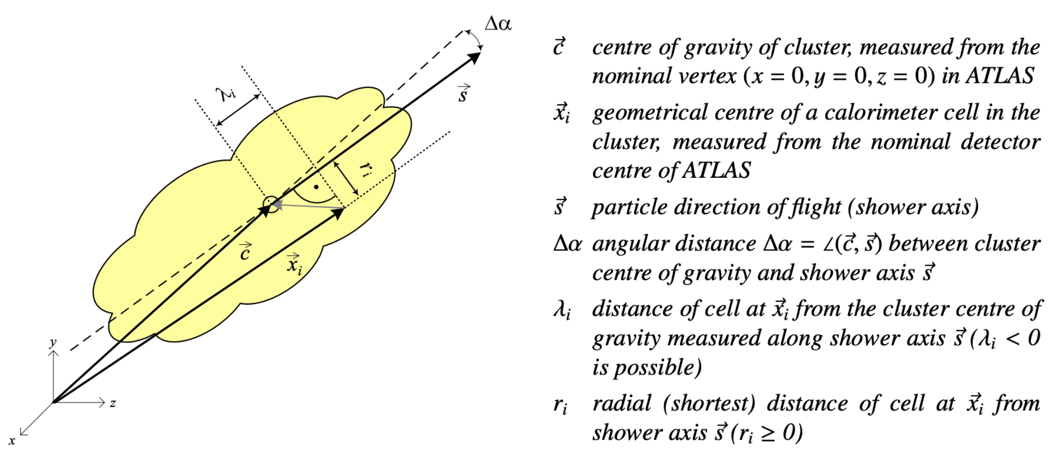


Figure 4.3: The geometric moments and relevant parameters for topo-clusters [87].

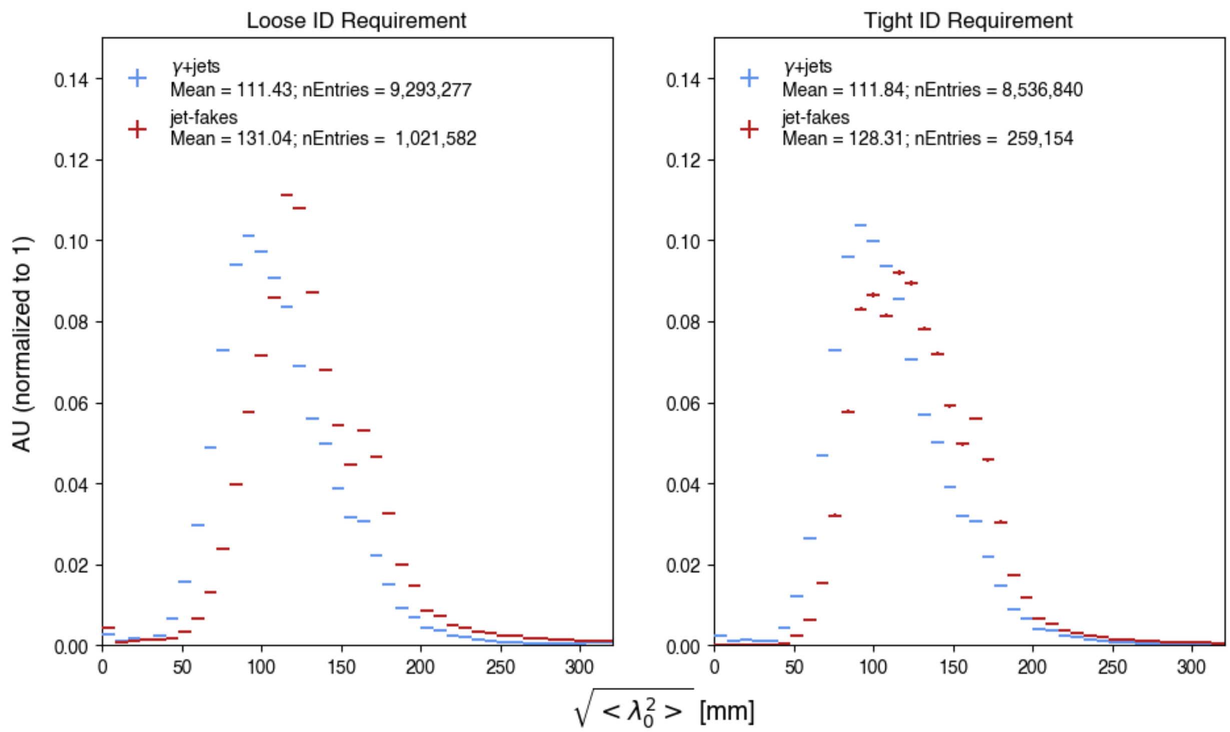


Figure 4.4: The distribution of the semi-major axis in depth, $\sqrt{\langle\lambda^2\rangle}$, for the leading topo-cluster in truth-matched γ+jet events and truth-vetoed jet fake events.

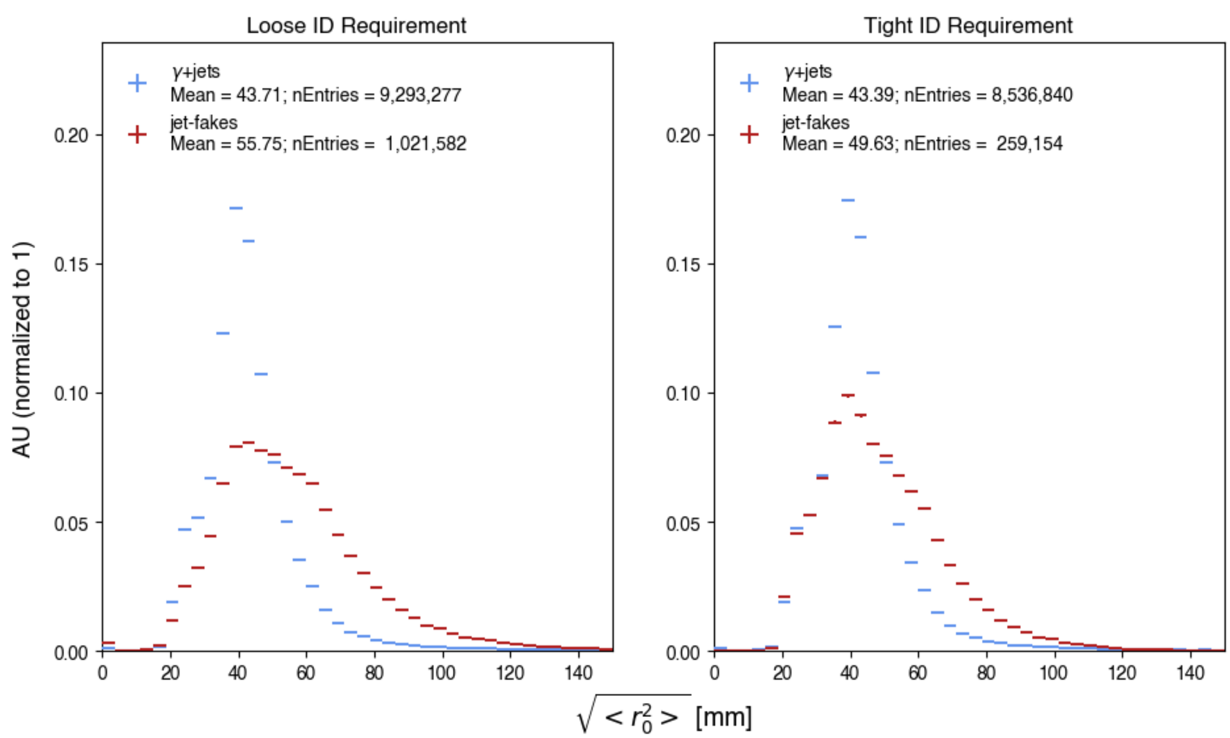


Figure 4.5: The distribution of the semi-major axis in width, $\sqrt{\langle r^2 \rangle}$, for the leading topo-cluster in truth-matched γ+jet events and truth-vetoed jet fake events.

Location Information

The location of the topo-cluster is calculated from the first moments of the three Cartesian coordinates, describing the position of $\vec{c}$. Due to azimuthal symmetry in the detector, the key locational component of topo-clusters is the depth in the calorimeter. Thus, this location can be described via λ_{clus}, the distance of the center of gravity from the front face of the calorimeter. This distribution is shown in Figure 4.6.

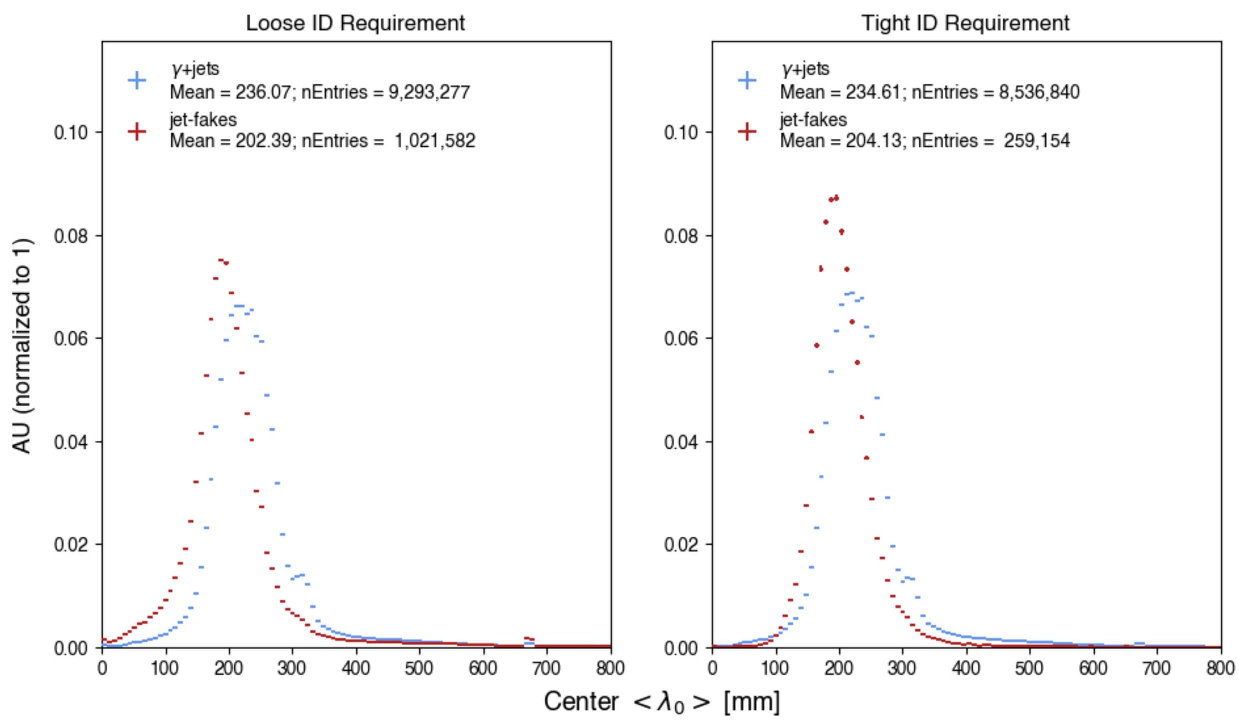


Figure 4.6: The distribution of the centroid depth ($< \lambda_0 >$) of the leading topo-cluster in truth-matched γ+jet events and truth-vetoed jet fake events.

A visualization of the geometric moments, both location and shape can be found in Figure 4.3, with relevant spacial parameters (e.g. $\vec{c}$, $\vec{s}$, etc.) shown and defined.

Additional Moments

Beyond the shape and location information, there are several other moments that can be used. Two are considered as inputs, cluster isolation and EM probability.

The signal thresholds built into the topological clustering algorithm are constructed to suppress noise, but lead to signal losses at the perimeter of clusters. The isolation moment, f_{iso}, measures the degree of isolation of a cluster. It is constructed as a weighted fraction of the sampling layer energy (E_s^{EM}) of non-clustered neighbor cells on the perimeter of the topo-cluster in the sampling layer. It is given by

$$f_{\text{iso}} = \frac{\sum_{S\in\{\text{samplings with} E_s^{\text{EM}}>0\}} E_s^{\text{EM}} N_{\text{cell},s}^{\text{noclus}} / N_{\text{cell},s}^{\text{neighbor}}}{\sum_{S\in\{\text{samplings with} E_s^{\text{EM}}>0\}}}. \tag{4.9}$$

Here, $N_{\text{cell},s}^{\text{noclus}} / N_{\text{cell},s}^{\text{neighbor}}$ is the ratio of unclustered perimeter cells to the total number of perimeter cells for a given cluster. This is summed across all sampling layers with positive energy, s. For the signal and background samples, the isolation is shown using the existing loose and tight identification criteria in Figure 4.7.

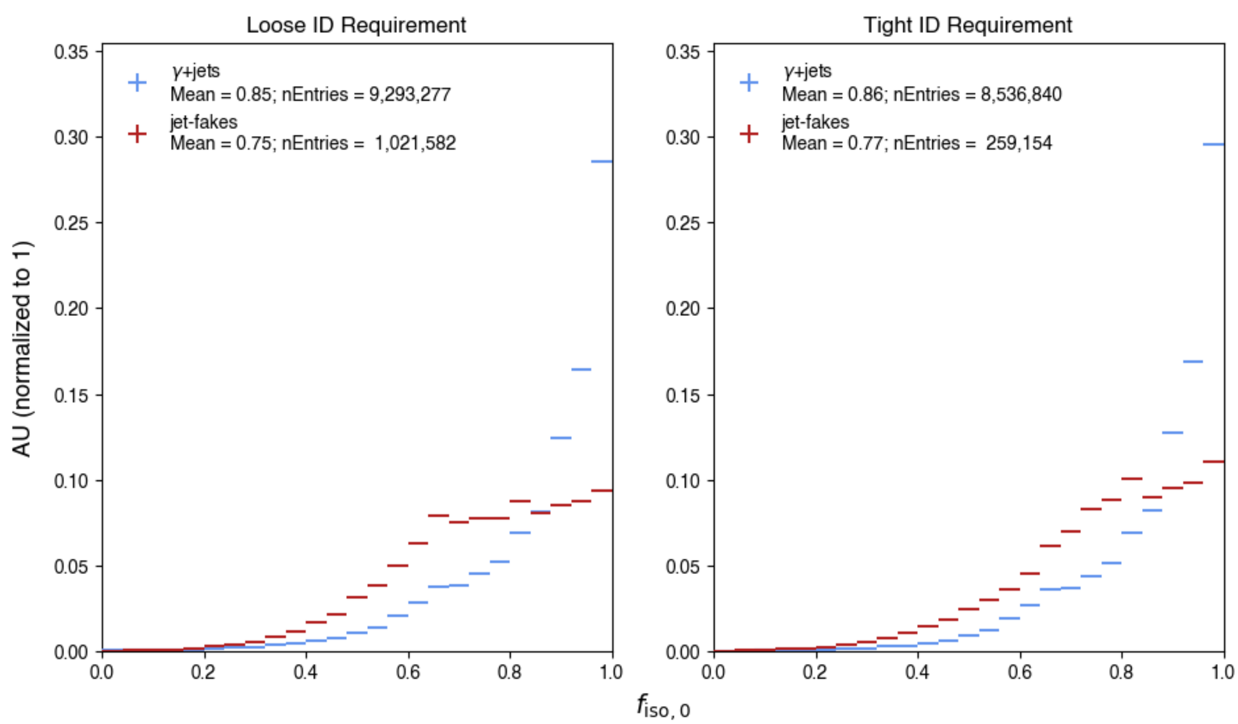


Figure 4.7: The distribution of the isolation moment of the leading topo-cluster in truth-matched γ+jet events and truth-vetoed jet fake events

Energy losses in topo-clusters are corrected through a method known as Local Hadronic Calibration (LCW) [88]. Since energy losses differ between hadronic and electromagnetic deposits, LCW includes a step where the likelihood that a topo-cluster originates from an electromagnetic shower is evaluated. The likelihood is defined through measuring the efficiency for detecting an EM-like cluster in bins of four other topo-cluster observables, defined by

$$\mathfrak{O}_{\text{clus}}^{\text{class}} = \{E_{\text{clus}}^{\text{EM}}, \eta_{\text{clus}}, \log_{10}(\rho_{\text{clus}}/\rho_0) - \log_{10}(E_{\text{clus}}^{\text{EM}}/E_0), \log_{10}(\lambda_{\text{clus}}/\lambda_0)\} \tag{4.10}$$

where $\rho_0 = 1$ MeV mm^{-3}, $E_0 = 1$ MeV, $\lambda_0 = 1$ mm. These bins (denoted $ijkl$) can then be used to define the likelihood, $\mathcal{P}_{\text{clus}}^{\text{EM}}$, in each bin as

$$\mathcal{P}_{\text{clus}}^{\text{EM}}(E_{\text{clus}}^{\text{EM}}, \eta_{\text{clus}}, \rho_{\text{clus}}/E_{\text{clus}}^{\text{EM}}, \lambda_{\text{clus}}) \mapsto \mathcal{P}_{\text{clus},ijkl}^{\text{EM}} = \frac{\epsilon_{ijkl}^{\pi^0}}{\epsilon_{ijkl}^{\pi^0} + 2\epsilon_{ijkl}^{\pi^\pm}}, \tag{4.11}$$

where $\epsilon_{ijkl}^{\pi^0(\pi^\pm)}$ is the efficiency, defined as

$$\epsilon_{ijkl}^{\pi^0(\pi^\pm)} = \frac{N_{ijkl}^{\pi^0(\pi^\pm)}}{N_{ij}^{\pi^0(\pi^\pm)}} \tag{4.12}$$

for $N_{ijkl}^{\pi^0(\pi^\pm)}$ topo-clusters in a given bin $ijkl$ and $N_{ij}^{\pi^0(\pi^\pm)}$ topo-clusters in bin ij of the $(E_{\text{clus}}^{\text{EM}}, \eta_{\text{clus}})$ phase space. This likelihood, $\mathcal{P}_{\text{clus}}^{\text{EM}}$, by definition is bounded from 0 to 1, and considered as an input. This distribution is shown in Figure 4.8, with both the current loose and tight photon identification criteria applied.

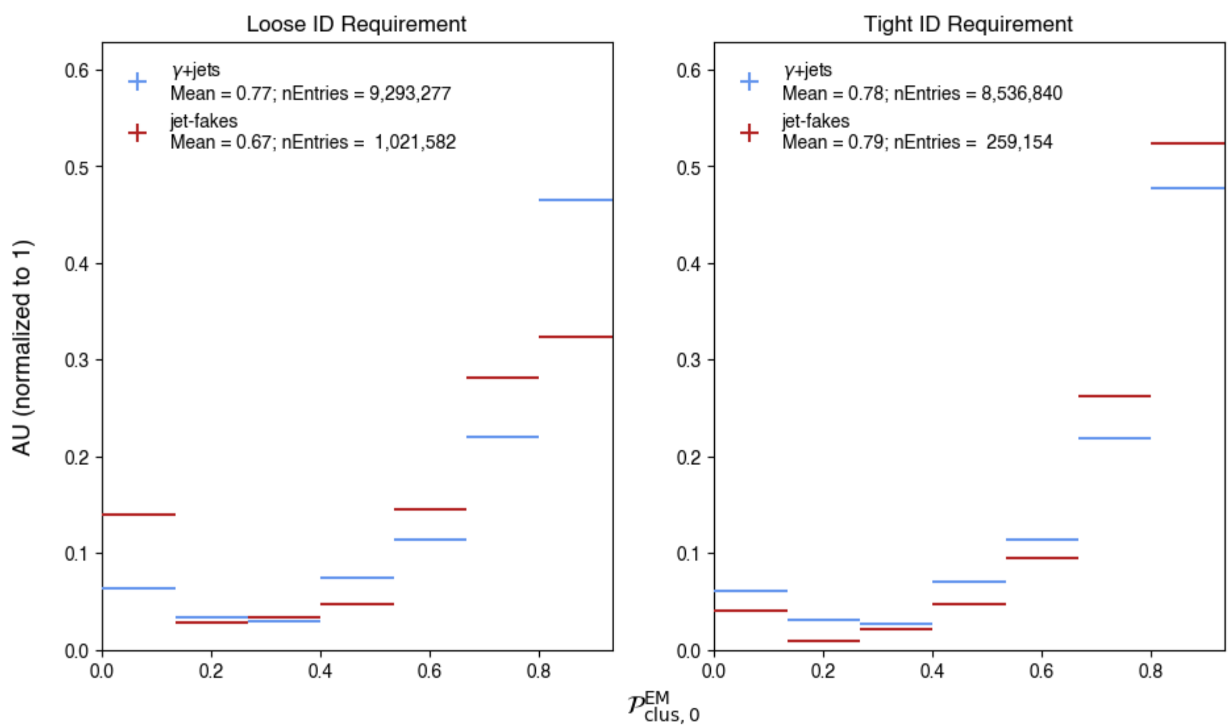


Figure 4.8: The distribution of the EM probability moment of the leading topo-cluster in truth-matched γ+jet events and truth-vetoed jet fake events

4.2.1 Data-MC Comparison

In order to ensure that the topological clusters are properly modeled in the MC simulation, the variables investigated are compared to data. To do so, the present tight photon ID is applied to both the MC simulation and data. Data from the 2015+2016 data-taking period is considered, collected using a set of single-photon triggers with loose identification requirements matching those applied to the MC samples, running as low as 10 GeV. As before, photons in MC must correspond to a truth photon. Comparisons for $\sqrt{\lambda_0^2}$ and $\sqrt{r_0^2}$ are shown in Figure 4.9 and 4.10, respectively. Comparisons for the remaining input variables considered are included into Appendix B for both converted and unconverted photons.

In general, the MC models the unconverted photons well, with the exception of the isolation, particularly in the most central bin and beyond the crack. For converted photons, $\sqrt{r_0^2}$ modeling differs between data and MC, and isolation differs considerably, particularly in the most central bin. On average converted photons have more topo-clusters per photon object, meaning they suffer more from cluster splitting, which influences these effects.

Correlations

The correlations of the topo-cluster variables to each other, to the shower shape variables, and to the isolation working points are evaluated. The correlations between variables are checked since highly correlated do not bring new information, and thus are not useful for adding to photon identification methods. Photon identification and isolation must be uncorrelated in order to define the regions used for 2x2D sideband methods [89], as correlations can affect the purity estimation. These correlations are shown for both the γ+jet in Figure 4.11 and for the jet-fakes sample in Figure 4.12. The added variables have low correlation to the isolation working points with the exception of the topo-cluster isolation, which is $\sim$40% correlated to the working points. Additionally the topo-cluster variables have some correlation to one another, particularly the ones with shape information, but in general are comparable to the level of correlation between the shower shape variables.

4.2.2 Incorporating into Photon Identification

The topological cluster variables were added to the existing shower shape variable-based photon identification. The cut-based approach is reoptimized including the new variables, and new cuts are derived in each $|\eta|$-p_{T} bin. To compare improvement across bins, the figure of merit used is the improvement in background rejection for the same signal efficiency as the current method, given as

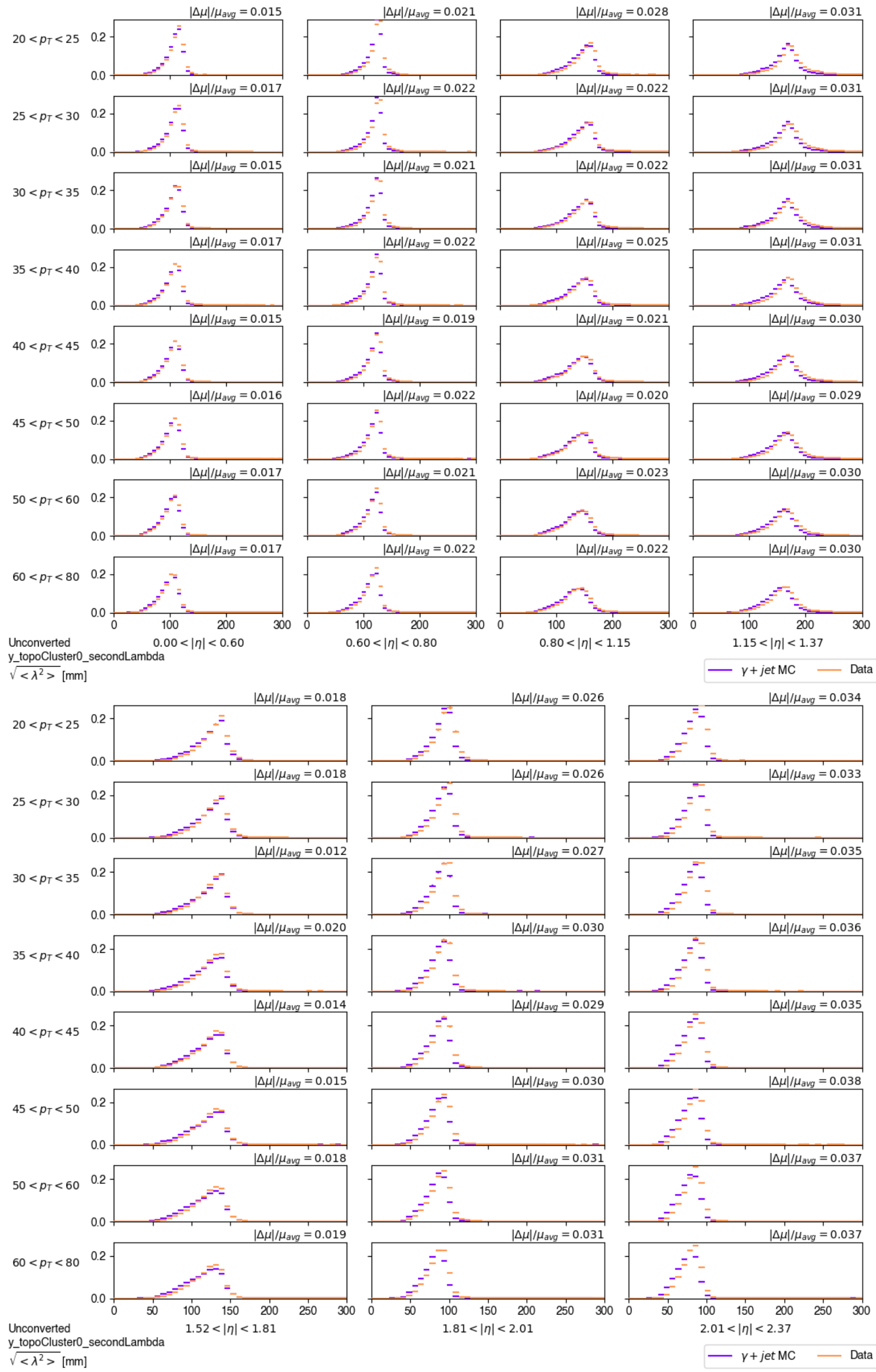


Figure 4.9: Data-MC comparison of the semi-major axis in depth of the leading topo-cluster ($\sqrt{\lambda_0^2}$) for unconverted photons. The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}$ is shown.

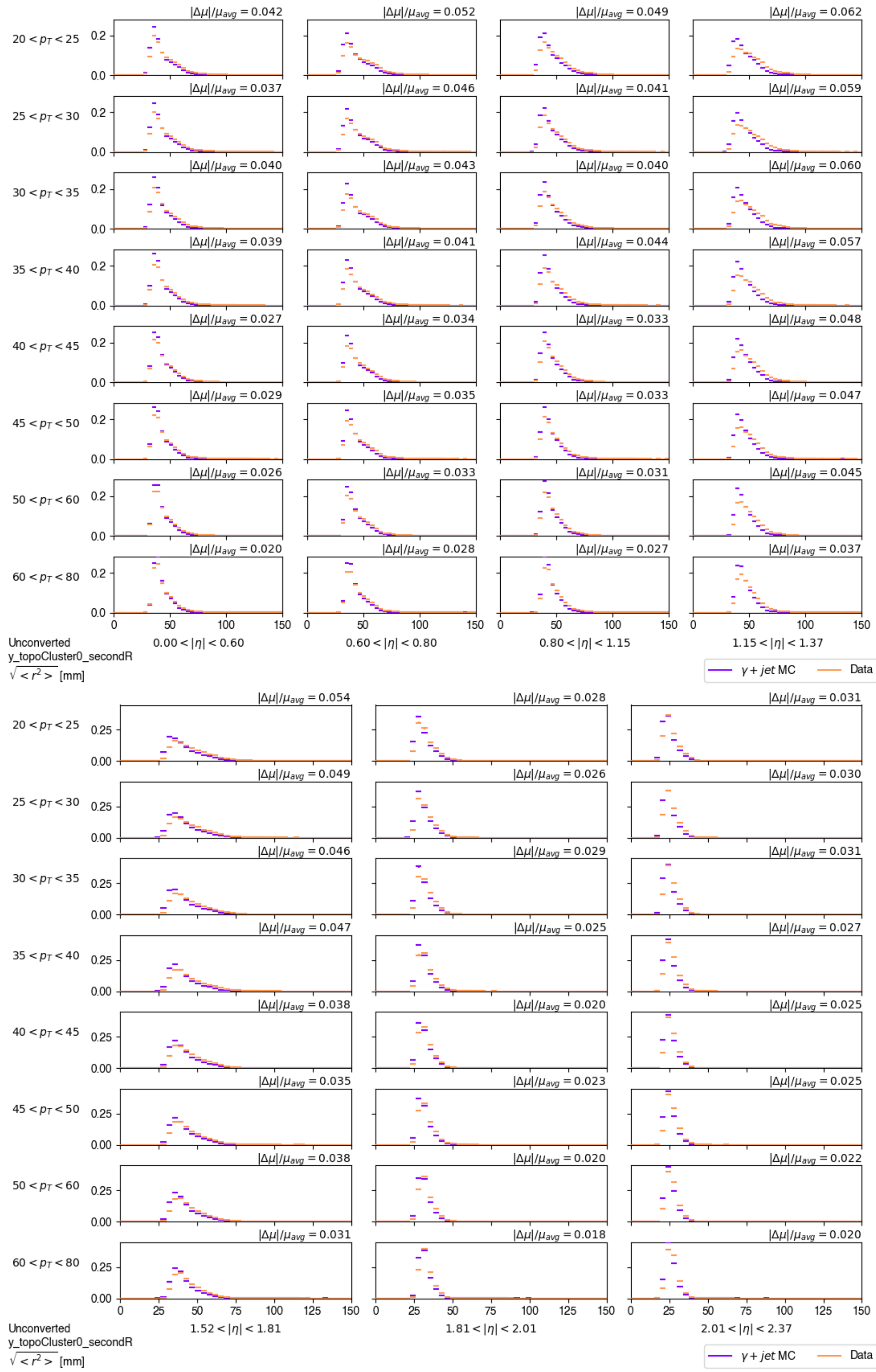


Figure 4.10: Data-MC comparison of the semi-major axis in width of the leading topo-cluster ($\sqrt{r_0^2}$) for unconverted photons.The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}$ is shown.

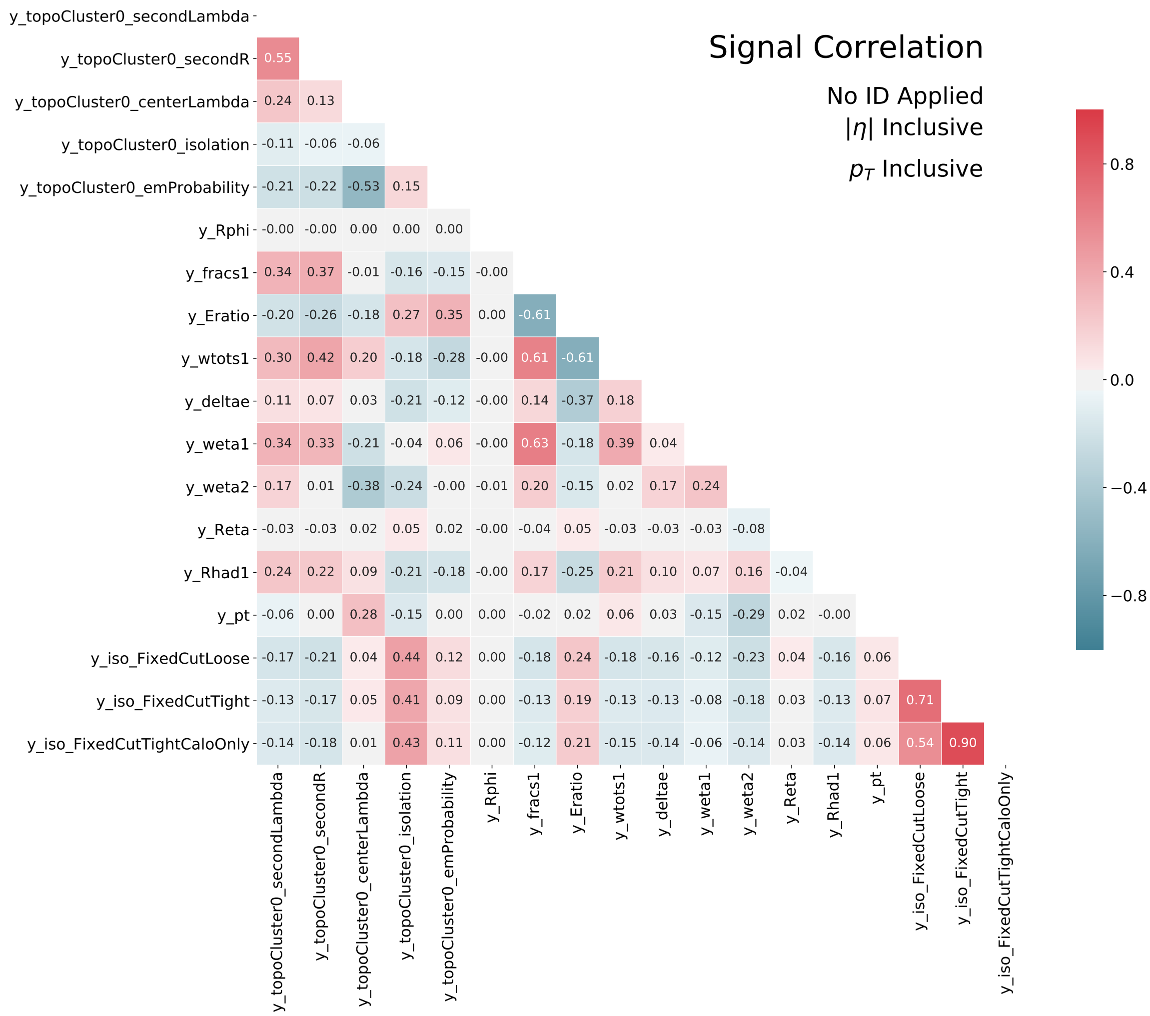


Figure 4.11: Correlation values between the shower shape variables, topological cluster variables, and isolation working points for the γ+jet sample.

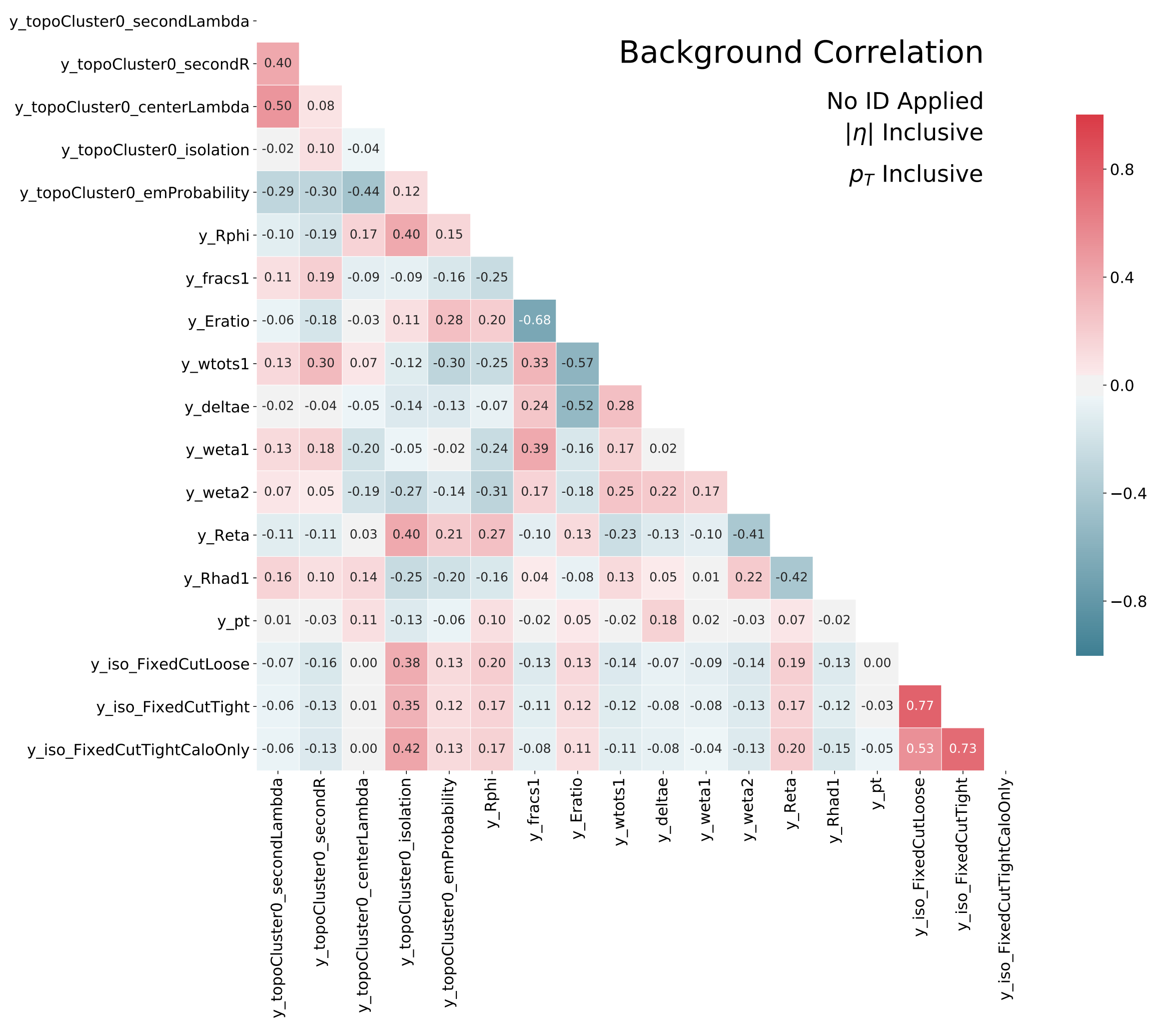


Figure 4.12: Correlation values between the shower shape variables, topological cluster variables, and isolation working points for the jet-fakes sample.

$$Z_{\text{rel}} = \frac{(1-\epsilon_{bkg,i}) - (1-\epsilon_{bkg,0})}{(1-\epsilon_{bkg,0})} = \frac{\epsilon_{bkg,0} - \epsilon_{bkg,i}}{1-\epsilon_{bkg,0}}, \tag{4.13}$$

for relative gain Z_{rel} (the figure of merit), and background efficiency ϵ_{bkg} (thus, background rejection is $(1-\epsilon_{bkg})$, where index i indicates the reoptimized photon identification, and index 0 indicates the shower shape-only identification. This value is calculated bin-wise in $|\eta|$-p_{T} and shown in Figure 4.14. In principle, adding additional variables to a cut-based method should only improve optimization if the search over possible cut values is exhaustive. However, in several bins, worse performance is found by incorporating new variables. This can be the result of two causes:

- In a high dimensional space, such a grid search becomes computationally prohibitive, so a Genetic Algorithm[2] [91] is used to find an approximate solution. Increasing dimensionality increases the probability to find local minima that differ from the global minimum, and thus suboptimal solutions.

- Statistical fluctuations. Models are trained on an orthogonal subset of events than they are evaluated on, and statistical fluctuations in these samples can influence reported gain, particularly in low stats bins. Figure 4.13 shows the number of signal and background events in each $|\eta|$-p_{T} bin.

As shown in Figure 4.14, by adding these variables as inputs, as much as a 12% improvement in background rejection for the same signal efficiency can be achieved.

[2]The Genetic Algorithm used is a standard method implemented in `TMVA`, and the details of implementation can be found in Reference [90].

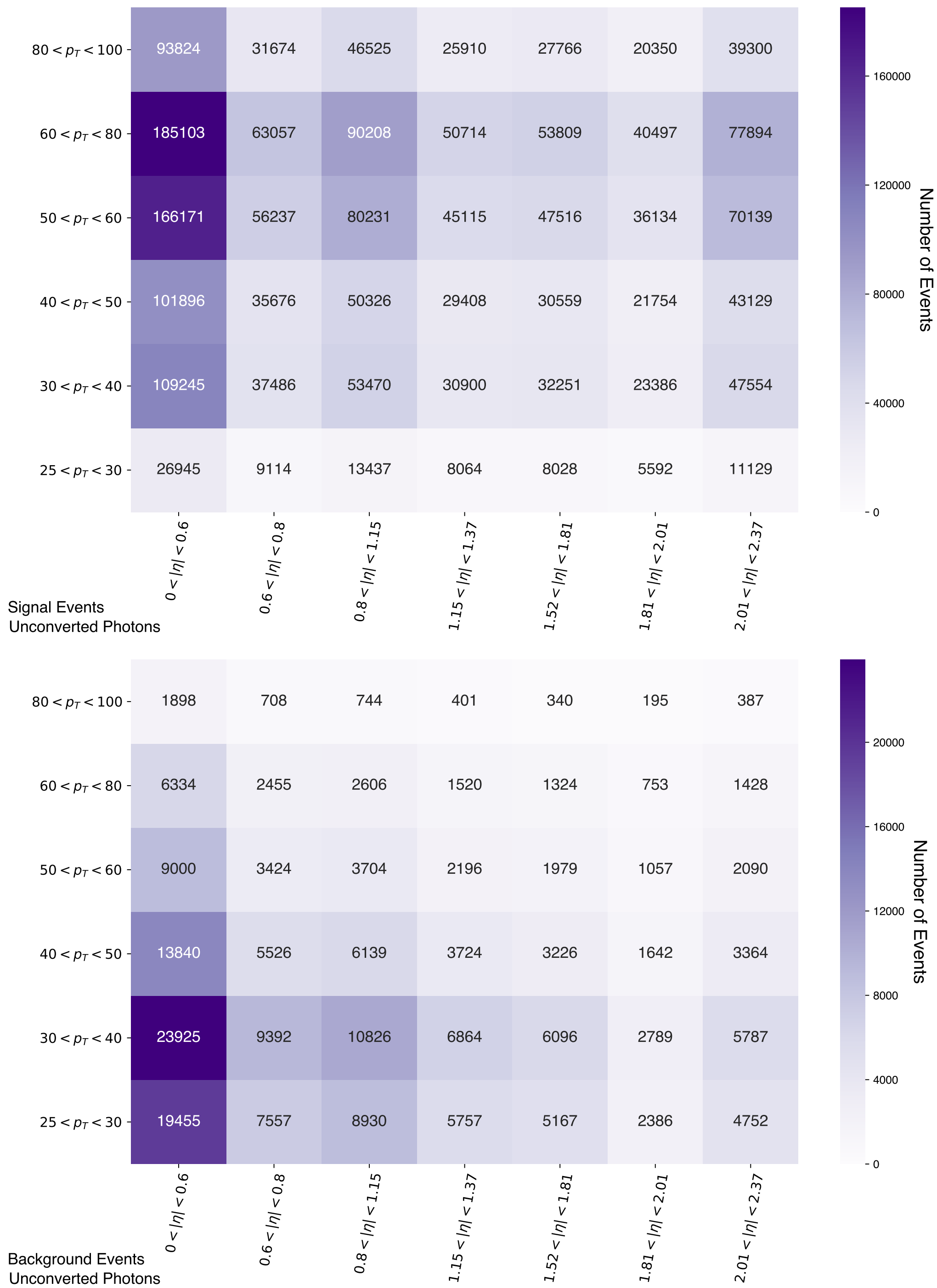


Figure 4.13: Number of training events used in the model for unconverted photons in each $|\eta|$-p_{T} bin for the signal γ+jets sample (top), and the background jet fakes sample (bottom). The preselection described in Section 4.1 is used.

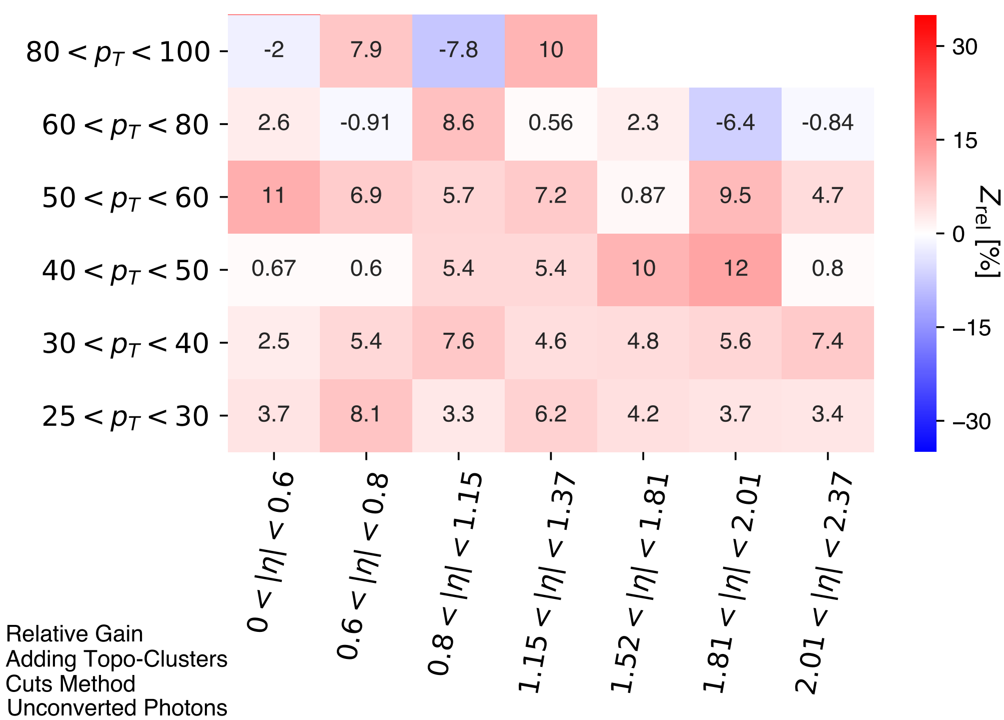


Figure 4.14: The relative gain in background rejection, $Z_{\rm rel}$, for each $|\eta|$-$p_{\rm T}$ bin at 85% signal efficiency by adding topo-cluster moment variables into the cut-based photon identification menu. Bins with fewer than 400 background events are masked.

4.3 Multivariate Analysis Techniques

The current photon identification working points are defined by a univariate rectangular cuts method. Cut optimization is performed through a 9-dimensional scan over the shower shape variables. In order to better reject background, a multivariate approach to defining these working points has been investigated. Multivariate Analysis (MVA) methods are described in detail in Appendix A, and this section will discuss the improvements brought to photon identification through implementing them. The methods investigated are a BDT, described in Section 4.3.1, and a NN, described in Section 4.3.2.

4.3.1 Boosted Decision Tree

A gradient boosted BDT was employed as an alternative classification algorithm. These models are built using `TMVA` [90]. In training, performance suffered heavily from the lack of background training statistics when segmenting into $|\eta|$-p_{T} bins. To navigate this problem, rather than training an independent model for each $|\eta|$-p_{T} bin, p_{T} inclusive models were trained. As shown in Figures 4.11 and 4.12, the input variables have low correlation to p_{T}, and thus this does not bias the model. This provided better performance in high p_{T} bins. In low p_{T} bins, it provided better performance overall, but in a few bins had equivalent or negligibly worse performance. The percent improvement, as defined using the same metric defined in Equation 4.13, is shown in Figure 4.15, with the bin-wise trained BDT used as the baseline.

To see the improvement over the cut-based method, the BDT is compared to the current photon identification, using the same figure of merit (Equation 4.13), where the efficiency denoted i now represents the BDT model, and the efficiency denoted 0 represents the current

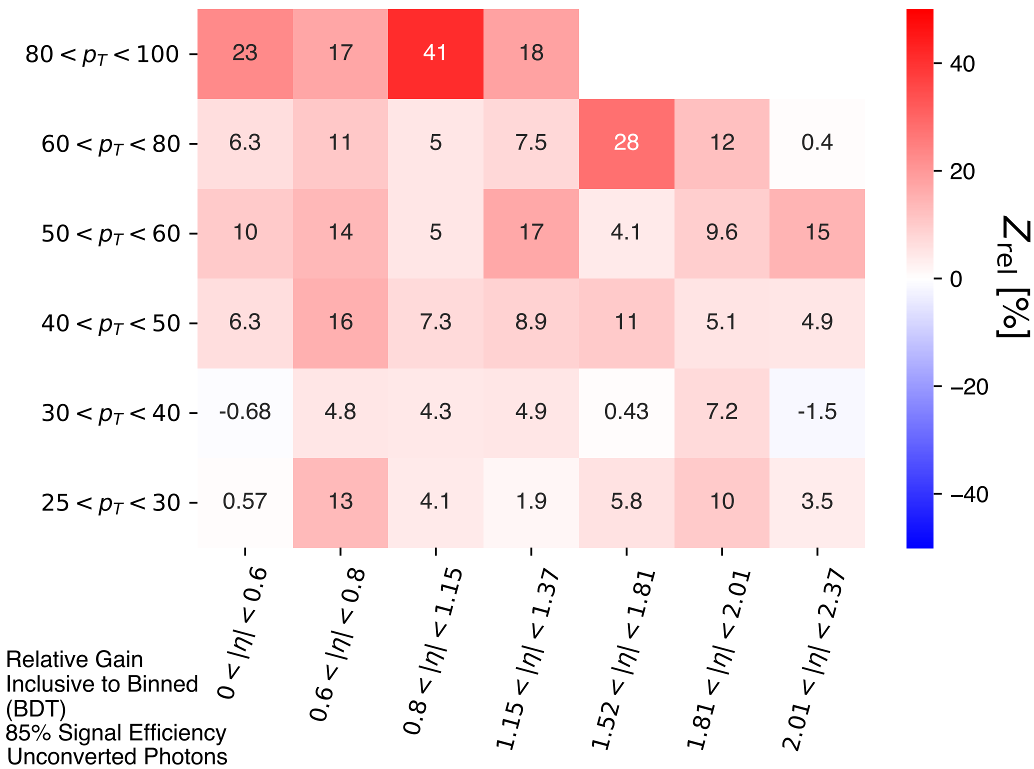


Figure 4.15: The relative gain in background rejection, Z_{rel}, for each $|\eta|$-p_{T} bin at 85% signal efficiency by using a p_{T}-inclusive trained BDT compared to a p_{T}-binned BDT. Only shower-shape variables are included as inputs. Bins with fewer than 400 background events are masked.

cut-based model. This model performs better in the majority of bins, by as much as 22%, but performs worse in all of the outer $|\eta|$ bins where there are poor background statistics, shown in Figure 4.16.

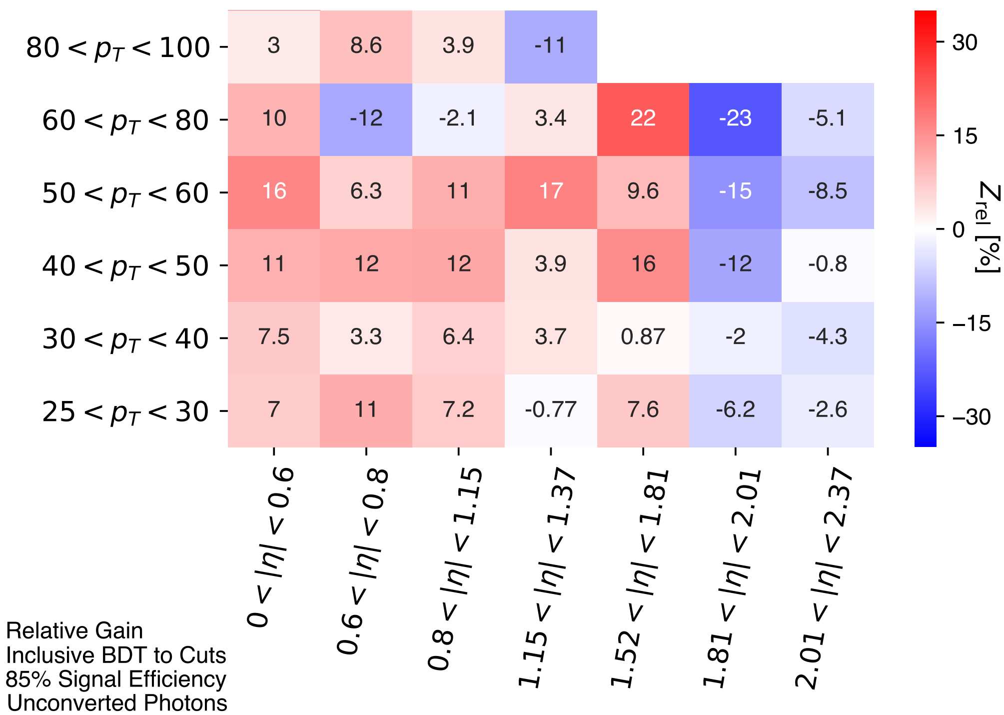


Figure 4.16: The relative gain in background rejection, Z_{rel}, for each $|\eta|$-p_{T} bin at 85% signal efficiency by using a BDT compared to the cut-based photon identification menu. Only shower-shape variables are included as inputs. Bins with fewer than 400 background events are masked.

4.3.2 Neural Network

In addition to a BDT, a NN was studied for photon classification, considering the same shower shape variables. The model topology[3] used is shown in Figure 4.17, it has an input

[3] A description of NN topologies, activation functions, and other associated terminology can be found in Appendix A.

layer of size 9 to match the number of shower shape variables, which is densely connected to a hidden layer of size 6, with ReLU activation functions. This layer is densely connected to an output with a sigmoid activation function. A dropout [92] of 20% is used on the hidden layer while training to reduce overtraining. This produces a total of 67 trainable parameters. Models with additional layers and with a wider hidden layer were investigated but found to not provide any additional improvement in background rejection, so were rejected.

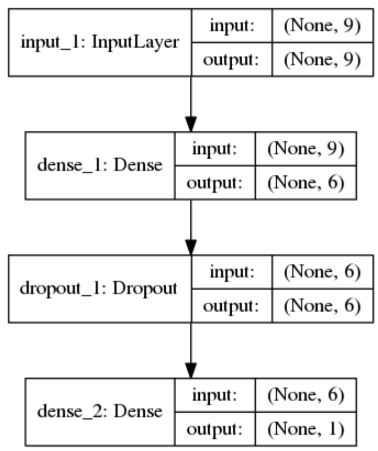


Figure 4.17: A diagram of the NN topology used for photon identification. Inputs to this NN are the shower shape variables.

The NN performed better at low p_{T} than the inclusively trained BDT by as much as 30%, but worse at high p_{T}. This comparison can be seen in Figure 4.18. A p_{T}-inclusive training approach was considered for the NN as well, however provided a worse performance for the majority of $|\eta|$-p_{T} bins, shown in Figure 4.19.

4.3.3 MVA With Topo-Cluster Variables

The BDT was studied incorporating the topo-cluster moments as inputs. The relative gain when doing so is shown in Figure 4.20, as compared to the current tight photon identification

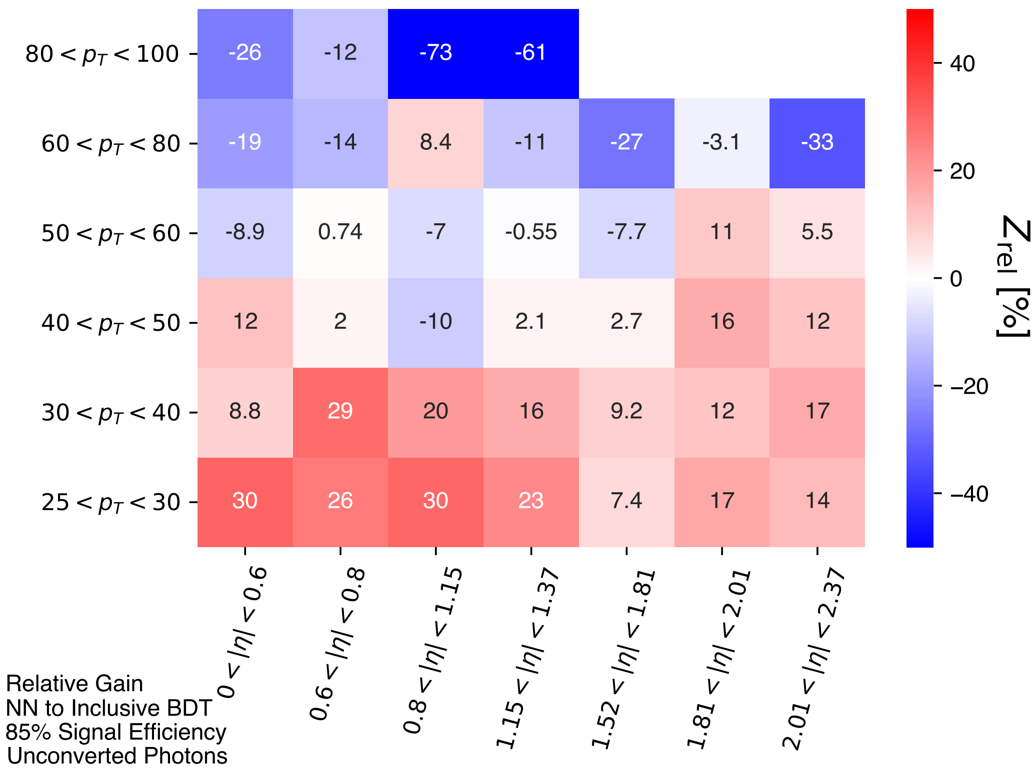


Figure 4.18: The relative gain in background rejection, Z_{rel}, for each $|\eta|$-p_{T} bin at 85% signal efficiency by using a NN trained in each $|\eta|$-p_{T} bin compared to a BDT trained inclusively. Negative values can be interpreted as the BDT having better performance than the NN. Only shower-shape variables are included as inputs for both models. Bins with fewer than 400 background events are masked.

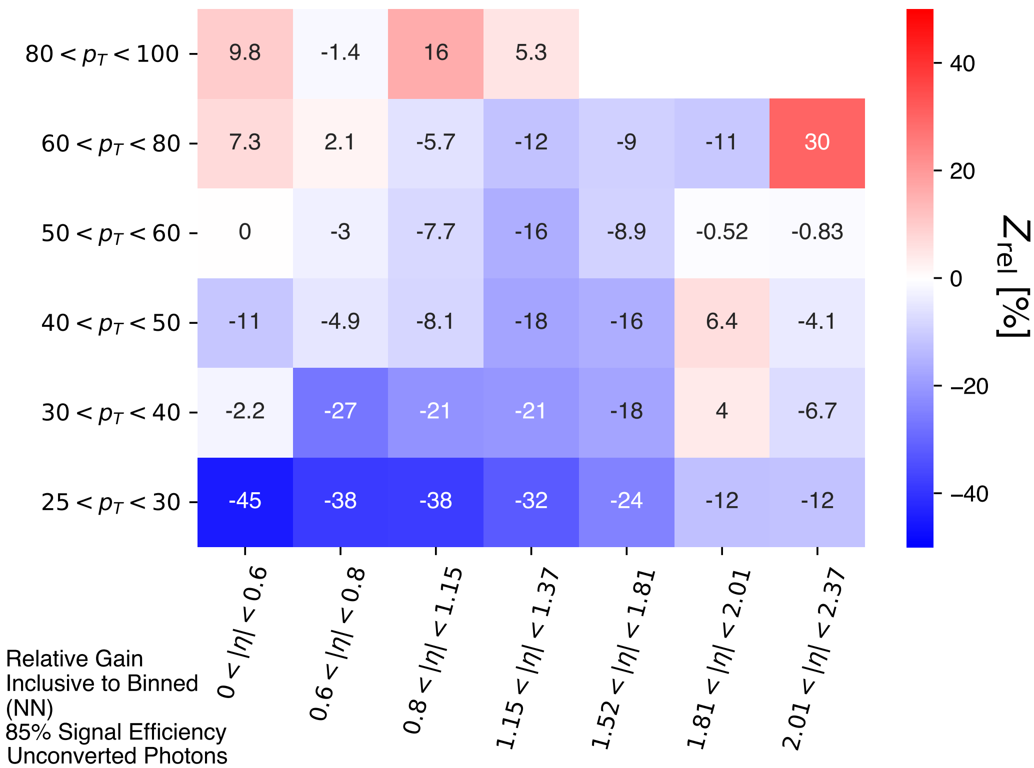


Figure 4.19: The relative gain in background rejection, Z_{rel}, for each $|\eta|$-p_{T} bin at 85% signal efficiency by using a NN trained inclusively compared to one trained in each $|\eta|$-p_{T} bin. Negative values can be interpreted as the bin-wise training having better performance than the inclusive training. Only shower-shape variables are included as inputs for both models. Bins with fewer than 400 background events are masked.

working point using the cut-based method. Through applying this optimization, as much as a 27% improvement is shown, and in most $|\eta|$-p_{T} bins, a gain of greater than 10% is shown.

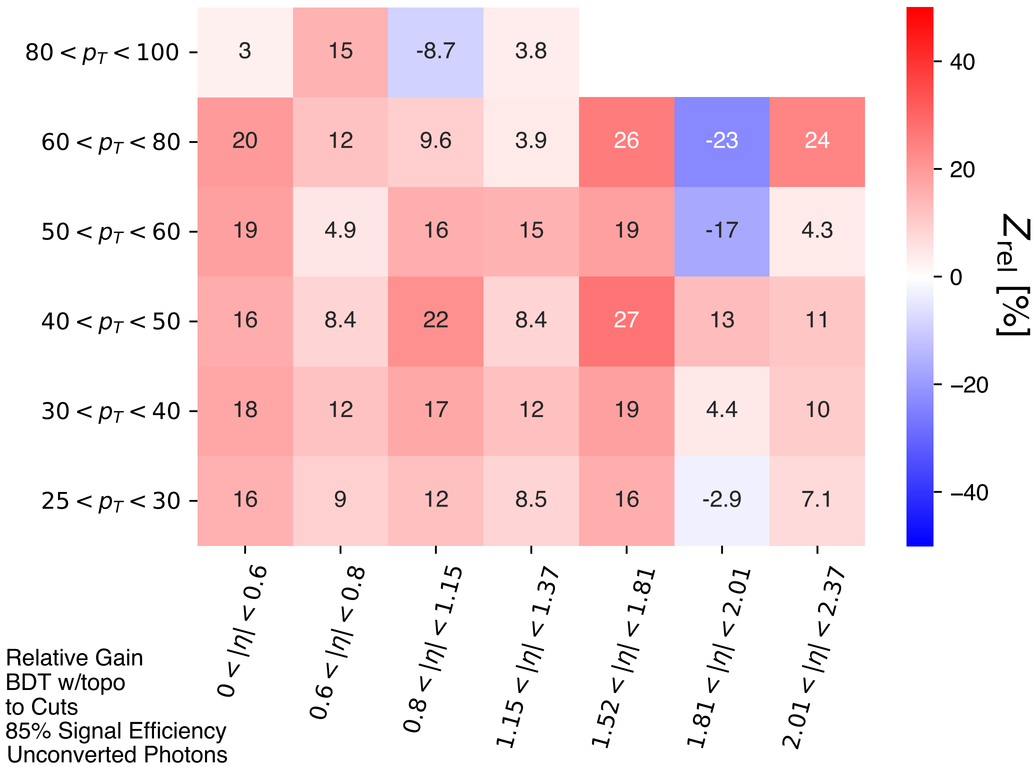


Figure 4.20: The relative gain in background rejection, Z_{rel}, for each $|\eta|$-p_{T} bin at 85% signal efficiency by using a BDT with the topo-cluster variables as input in addition to the shower shape variables compared to the current tight photon identification cut-based method. Bins with fewer than 400 background events are masked.

4.4 Conclusion

Through adding topo-cluster moments as inputs and by switching to a BDT, an improvement of greater than 10% in background rejection for the same signal efficiency as the tight working point is demonstrated in most $|\eta|$-p_{T} bins, and as much as a 27% improvement is shown. While the comparison between data and MC modeling for the topo-cluster moments already demonstrates reasonably good agreement, proper calibration must be performed before full implementation.

The presented analysis in this thesis, studying $HH \to \gamma\gamma b\bar{b}$ production, has two photons in the final state. These per-photon improvements can lead to a improvement, since the signal selection is sensitive to the square of the photon identification efficiency. The analysis is statistics limited, the signal significance effectively scales as the number of signal events, so the ability to loosen the signal selection will directly improve the final analysis sensitivity.

CHAPTER 5

SEARCH FOR DI-HIGGS PRODUCTION IN THE $\gamma\gamma b\bar{b}$ CHANNEL

Science is magic that works.
— Kurt Vonnegut, Cat's Cradle

5.1 Analysis Overview

This analysis searches for resonant and non-resonant di-Higgs production, where one Higgs boson decays to a pair of photons, and the other decays to a $b\bar{b}$ pair. The motivation for searching for di-Higgs production is outlined in Section 1.3. Di-Higgs production provides a direct probe of the Higgs trilinear coupling. This decay channel, $\gamma\gamma b\bar{b}$, is appealing to study due to the clean diphoton trigger, relatively small backgrounds, and excellent diphoton mass resolution. Enhancements to the di-Higgs rate may be observable in the current Run 2 dataset, and realized through a resonance or non-resonantly. The resonant search looks for a generic scalar, and thus is model independent, however this scalar may be realized as a heavy Higgs in a 2HDM [26]. Non-resonantly, enhancements can occur through the presence of BSM couplings.

Previous results were presented with an integrated luminosity of 20 fb^{-1} at $\sqrt{s} = 8$ TeV representing ATLAS dataset in 2012. In this, a 2.4σ excess was observed [93]. This was followed by a search encompassing early Run 2 data, just that taken in 2015, corresponding to 3.2 fb^{-1} at $\sqrt{s} = 13$ TeV [94]. This search used the same object selection as the Run 1 analysis, and a common selection between the resonant and non-resonant analysis. No significant excesses were seen.

The presented analysis expands on the prior analyses, using Run 2 data from 2015-2016 corresponding to 36.1 fb^{-1}. A new event selection is defined to be sensitive to the Run 2 conditions. Furthermore, the previous cut-and-count approach has been replaced by unbinned profile likelihood fits to the signal and background. In the resonant analysis, this is performed on the $m_{\gamma\gamma b\bar{b}}$ distribution after cutting on $m_{\gamma\gamma}$ to isolate $H \to \gamma\gamma$ events. In the non-resonant analysis the fit is performed on the $m_{\gamma\gamma}$ spectrum. Chapter 6 will then present future updates to this analysis.

5.2 Data and Simulated Samples

5.2.1 Data

The analysis presented uses *pp* collision data recorded by the ATLAS experiment at the LHC in 2015 and 2016, the first half of "Run 2." After detector and data quality requirements, the total collected data corresponds to 36.1 fb^{-1}. An unprescaled loose diphoton trigger is used for this search, `HLT_g35_loose_g25_loose`, which triggers on diphoton events where the leading (subleading) photon candidate p_{T} is greater than 35 (25) GeV. At HLT, these candidates must pass loose photon identification criteria. Using bootstrap methods presented in reference [95] the efficiency of the diphoton trigger is found to be 99.4% for events that pass the event selection described in Section 5.4.

5.2.2 Simulated Samples

Simulation is generated as outlined in Chapter 3 in order to compare theoretical predictions to data. Independent samples are made for the signal hypotheses considered (outlined in Section 5.2.2.1), as well as for the predominant backgrounds (outlined in Section 5.2.2.2).

5.2.2.1 Signal Samples

This analysis considers two signal sources: non-resonant di-Higgs production and resonant production of Higgs boson pairs through a scalar resonance. In addition to SM non-resonant signal samples, samples with varied κ_λ are produced. This section describes the MC samples used to model these various signal processes.

Non-resonant SM Signal Samples

The non-resonant di-Higgs signal sample corresponding to SM ggF HH production is generated at NLO in QCD. The model was developed within the Cosmology, Particle Physics, and Phenomenology (CP3) theory group [96], and is known as $\mathrm{FT}_{\mathrm{approx}}$. It generates events in a Higgs Effective Field Theory (EFT) by using the aMC@NLO method [97].

An event reweighting is applied to account for top-quark mass dependence. Considering the interference between the box and triangle diagrams, this is an important effect (documented, for example, in Reference [98]). All exact one-loop amplitudes and known exact two-loop amplitudes are used. For the NNLO normalization, these effects are of order 5%.

The sample used is showered using Herwig++ [59], using the `UEEE5` tune and the `CTEQ6L1` PDF set.

Non-resonant Varied κ_λ Signal Samples

Samples for non-resonant di-Higgs production with varied κ_λ were generated at values $\kappa_\lambda = 0, \pm1, \pm2, \pm4, \pm6, \pm10$. These samples were generated at LO, then reweighted to NNLO through a K-factor, taking the ratio of the cross-sections of the $\mathrm{FT}_{\mathrm{approx}}$ SM non-resonant signal sample (at NNLO) and the $\kappa_\lambda = 1$ sample (at LO).

For generic κ_λ, an interpolation function can be used to predict the amplitude, and thus the signal yield. The amplitude can be written as

$$A(k_t, \kappa_\lambda) = k_t^2 B + k_t \kappa_\lambda T, \tag{5.1}$$

where B is the contribution from the box diagram (with two th vertices, thus squared) and T is the contribution from the triangle diagram (containing one th and one hhh vertex). Squaring this amplitude and varying κ_λ (consider values 0, 1, and 10) gives

$$|A(k_t, \kappa_\lambda)|^2 = k_t^4 |B|^2 + k_t^2 \kappa_\lambda^2 |T|^2 + k_t^3 \kappa_\lambda (B^* T + B T^*) \tag{5.2}$$

$$|A(1,0)|^2 = |B|^2 \tag{5.3}$$

$$|A(1,1)|^2 = |B|^2 + |T|^2 + (B^* T + B T^*) \tag{5.4}$$

$$|A(1,10)|^2 = |B|^2 + 100|T|^2 + 10(B^* T + B T^*). \tag{5.5}$$

These equations can then be combined to express the squared amplitude in terms of the amplitudes of the varied reference samples for any arbitrary κ_λ and k_t,

$$|A(k_t, \kappa_\lambda)|^2 = k_t^2 \left[\frac{90k_t^2 + 9\kappa_\lambda^2 - 99k_t\kappa_\lambda}{90} |A(1,0)|^2 + \frac{100k_t\kappa_\lambda - 10\kappa_\lambda^2}{90} |A(1,1)|^2 + \frac{\kappa_\lambda^2 - k_t\kappa_\lambda}{90} |A(1,10)|^2 \right]. \tag{5.6}$$

Resonant Signal Samples

The resonant samples were generated using a model of a heavy scalar decaying to two Higgs bosons, with the narrow-width approximation for the $X \to HH$ decay, requiring that the width is much smaller than the mass (Γ = 10 MeV is used, which is ∼0 considering resolution effects). A gg-initiated state is generated at NLO using aMC@NLO, and the scalar is forced to decay to two SM-like Higgs bosons with $m_H = 125$ GeV. Showering is performed in Herwig++ using the `CT10 NLO` PDF set and `UEEE5` tuned parameters. An event filter is used, which selects subsets of the full event generator output, requiring the $\gamma\gamma b\bar{b}$ decay channel[1]. Samples are generated at the following scalar mass hypotheses (in GeV): 260, 275, 325, 350, 400, 450, 500, 750, 1000.

5.2.2.2 Background Samples

The dominant background in this analysis are the non-resonant $m_{\gamma\gamma}$ continuum events. Additionally, mono-Higgs boson production is a subdominant background to this analysis, which is composed of several production modes. This section describes the samples used to model these background processes.

$\gamma\gamma$-Continuum

$\gamma\gamma$-continuum processes including two photons and jets are generated at LO using Sherpa. This MC sample is used predominantly for shape information, since normalization is taken through a data-driven approach using the diphoton sidebands. Jets faking photons are also modeled from the data sidebands using a template fit, described in Section 5.5.1.

Mono-Higgs Production

The following mono-Higgs production channels are considered: gluon-gluon fusion (ggH),

[1]The `XtoVVDecayFilter` is used for this

vector-boson fusion (VBF H) Z-associated production (ZH), W-associated production (WH), single and top quark pair associated production (tH and $t\bar{t}H$, respectively), and bottom quark pair associated production (bbH). Rates for these samples are taken from the 2016 Yellow Report [99]. The generators for these samples vary, and are listed in Table 5.1, along with the cross-section times Branching Ratio (BR) for each sample.

Table 5.1: Mono-Higgs boson backgrounds considered, including generator and PDF used to produce each sample. Cross-section for each sample is shown, multiplied by the $H \to \gamma\gamma$ BR of 0.00227. Cross-sections assume $\sqrt{s} = 13$ TeV and $m_H = 125$ GeV.

Process	Generator+PDF	$\sigma \times \mathcal{B}$ $(H \to \gamma\gamma)$ [pb]
ggH	Powheg + Pythia8 PDF4LHC15	0.1101404
VBF H	Powheg + Pythia8 PDF4LHC15	0.00857833
W^-H	Powheg + Pythia8 PDF4LHC15	0.00120605
W^+H	Powheg + Pythia8 PDF4LHC15	0.00190226
ZH	Powheg + Pythia8 PDF4LHC15	0.001724519
$ggZH$	Powheg + Pythia8 PDF4LHC15	0.000278529
$t\bar{t}H$	Pythia8 A14 NNPDF23LO	0.001149755
bbH (positive)	aMC@NLO + Pythia8 A14 NNPDF23LO	0.00110390
bbH (negative)	aMC@NLO + Pythia8 A14 NNPDF23LO	-8.95378e-05
tH (t-channel)	MadGraph + Pythia8 A14 CT10ME	0.00016857
tH (W-associated)	aMC@NLO + Pythia8 UEEE5 CTEQ6L1 CT10ME	3.44359e-05

5.3 Object Definition

5.3.1 Photons

Photons selected for this analysis are converted or unconverted photons which pass tight ID cuts (outlined in Section 3.2.1). Photons must have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.37$. Additionally, the crack region of $1.37 < |\eta| < 1.52$ is rejected.

Isolation criteria are set on $E_{\mathrm{T}}^{\mathrm{iso}}$ and $p_{\mathrm{T}}^{\mathrm{iso}}$. $E_{\mathrm{T}}^{\mathrm{iso}}$ is the calorimeter-based isolation, the sum of energy of all topological clusters in $\Delta R = 0.2$ of a photon candidate. $p_{\mathrm{T}}^{\mathrm{iso}}$ is the track-based isolation, the sum of the transverse momenta of all tracks with $p_{\mathrm{T}} > 1$ GeV within $\Delta R = 0.2$ of the candidate. Photon candidates must have $E_{\mathrm{T}}^{\mathrm{iso}}$ less than 6.5% of their transverse energy, and $p_{\mathrm{T}}^{\mathrm{iso}}$ less than 5% of their transverse energy[2]. This requirement is 98% efficient.

Photons must pass an ambiguity tool that reduces the large $e \to \gamma$ fake rates from converted photons. This differentiates overlapping selections by making additional requirements on silicon hits and conversion rates [86].

The leading photon p_{T} distribution is shown in Figure 5.1, the subleading is shown in Figure 5.2. In these figures, the histograms are split by the MC truth flavor label of the $H \to b\bar{b}$ decay products. The γj, $j\gamma$, and jj fakes are taken from not-tight, not-isolated data and further described in Section 5.5.1.

Since photons do not leave tracks in the detector, the standard Primary Vertex (PV), chosen by the highest track $\sum p_{\mathrm{T}}^2$, is often incorrect. To select the PV, a NN is used. This NN uses the following inputs:

- $(z_{\mathrm{common}} - z_{\mathrm{vertex}})/(\sigma_z)$. z_{vertex} is the position of the primary vertex. z_{common} is defined by the weighted mean of extrapolated photon trajectories. For converted photons,

[2]Known as the FixedCutLoose working point.

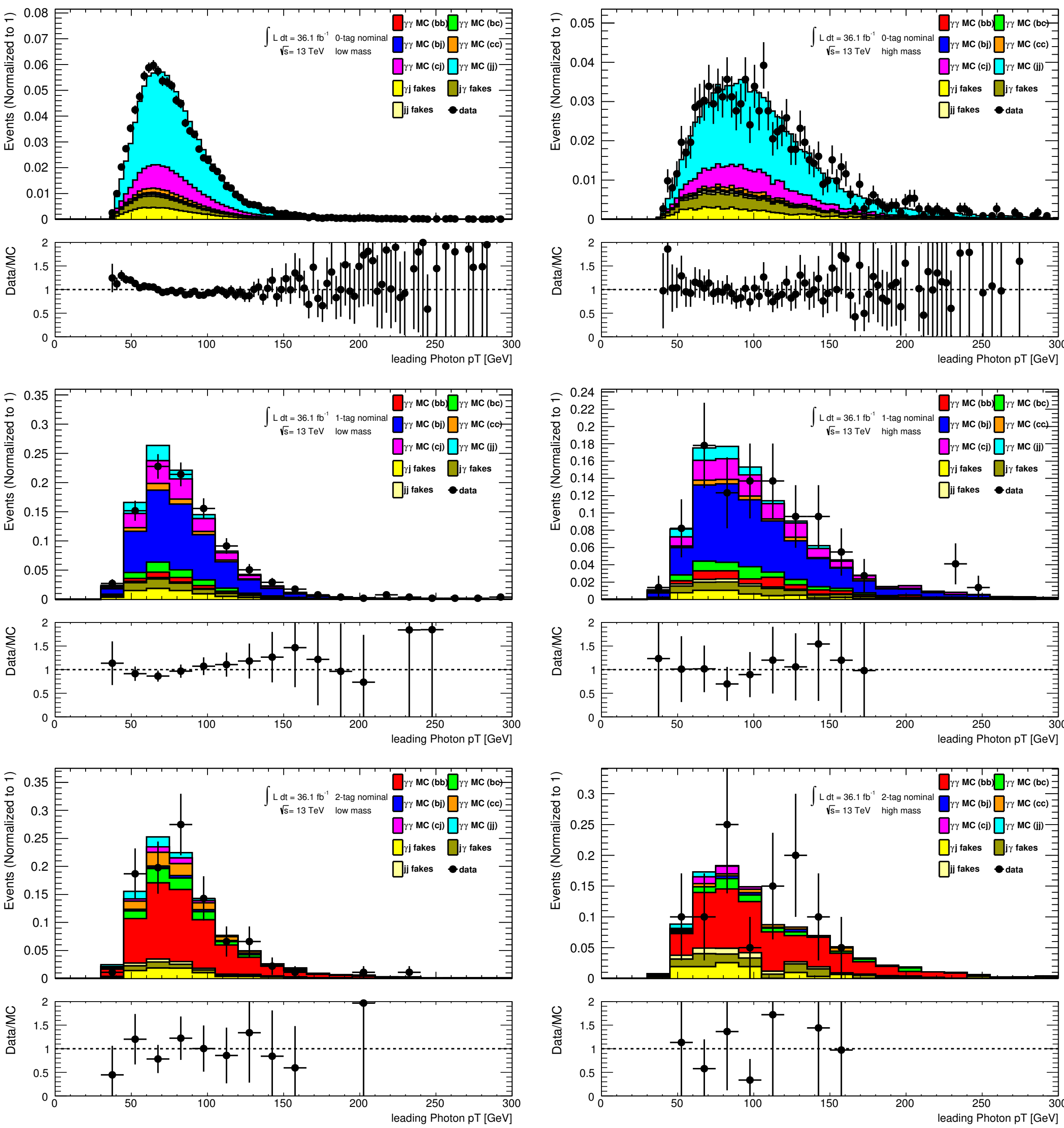


Figure 5.1: Leading photon p_{T} by b-tagging category. The low (left) and high (right) mass selections are shown. Both data and MC are normalized such that the integral is 1. Histograms are split by the flavor of the $H \to b\bar{b}$ decay products, where labels are taken from MC truth information. The γj, $j\gamma$, and jj fakes are taken from not-tight, not-isolated data and further described in Section 5.5.1.

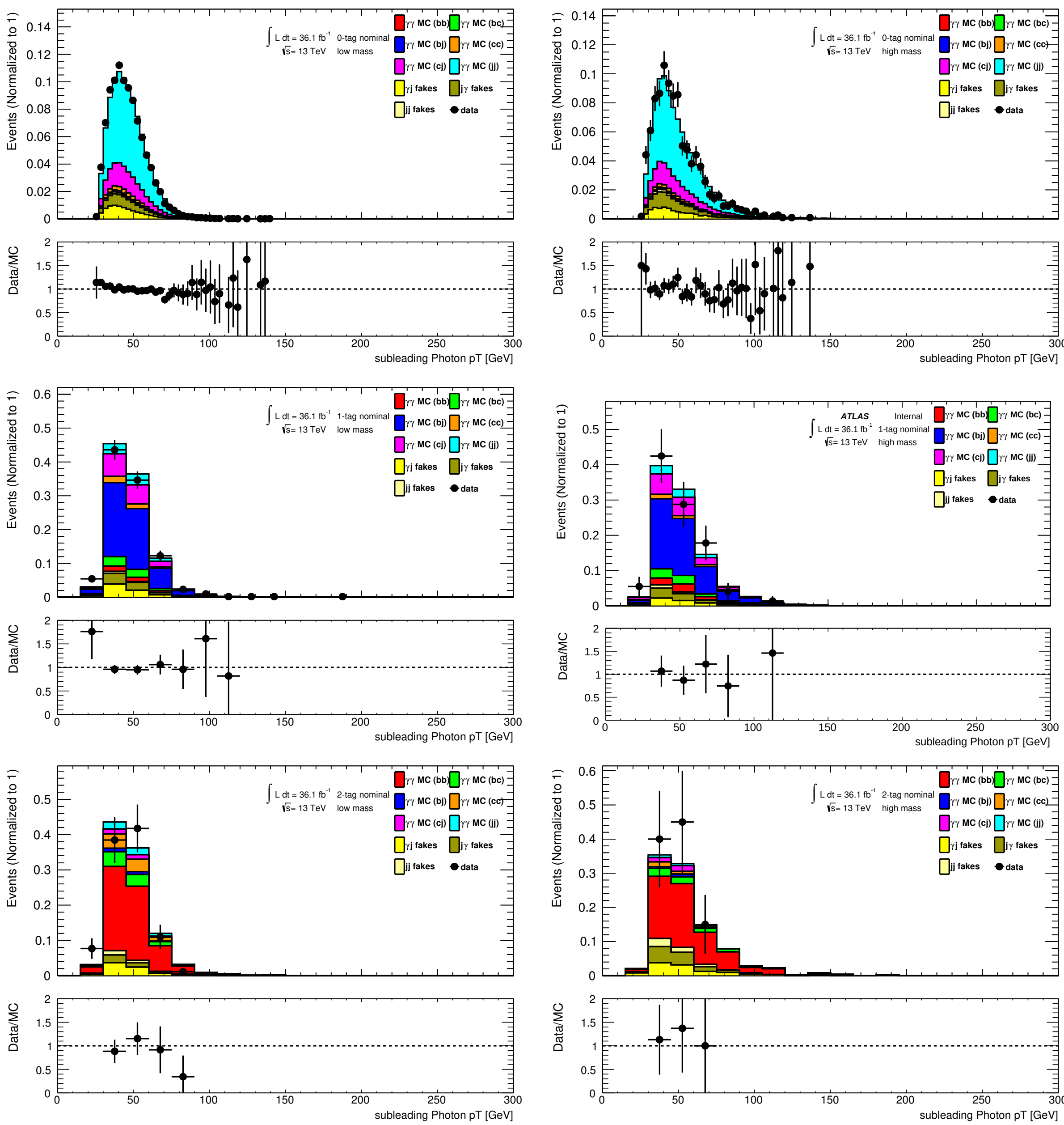


Figure 5.2: Subleading photon p_{T} by b-tagging category. The low (left) and high (right) mass selections are shown. Both data and MC are normalized such that the integral is 1. Histograms are split by the flavor of the $H \to b\bar{b}$ decay products, where labels are taken from MC truth information. The γj, $j\gamma$, and jj fakes are taken from not-tight, not-isolated data and further described in Section 5.5.1.

this is replaced by the direction derived by connecting the conversion location to the cluster barycenter in the first sampling layer in the EM calorimeter. σ_z is the associated uncertainty.

- $\log(\sum p_{\mathrm{T}})$, the scalar sum of transverse momenta of tracks associated to the vertex.
- $\log(\sum p_{\mathrm{T}}^2)$, the squared scalar sum of transverse momenta of tracks associated to the vertex.
- $\Delta\phi(\gamma\gamma, \mathrm{PV})$, the azimuthal angle between the $\gamma\gamma$ system and the system defined by the vector sum of tracks associated to the vertex. ϕ of the PV is defined as the ϕ vector sum of tracks associated to the vertex.

After selecting this PV, tracking and calorimeter variables are recomputed with respect to the new PV. This method is typically between 70% and 100% efficient at selecting the correct PV, depending on sample and selection.

5.3.2 Electrons

Electrons are used to reject jets originating from leptons. Electrons are required to have $p_{\mathrm{T}} > 10$ GeV, and $|\eta| < 2.47$ and outside of the crack-region ($1.37 < |\eta| < 1.52$). They must pass medium identification and loose isolation requirements [100], which are set on both calorimeter and tracking variables and are 99% efficient.

5.3.3 Jets

Jets are reconstructed with the `FastJet` [101] package. They are defined using the anti-k_t algorithm with $R = 0.4$, using jets with energy corrected for pile-up effects and direction readjusted to the primary vertex (EM-JES jets). Jets considered must have $p_{\mathrm{T}} > 25$ GeV and be within the tagging region, $|\eta| < 2.5$. The following additional requirements are applied[3]:

- Jets with $p_{\mathrm{T}} < 50$ GeV and $|\eta| < 2.4$ must pass the "default" Jet Vertex Tagger (JVT) requirement, with $|\mathrm{JVT}| \geq 0.59$ [102].
- Events that contain a jet passing the "LooseBad" cut are vetoed. This criteria is defined in Reference [103], and target fake jets resulting from detector noise and non-collision background.
- Jets with $\Delta R < 0.4$ from a tight, isolated photon are vetoed.
- Jets with $\Delta R < 0.2$ from a selected electron are vetoed.

In order to reconstruct the 4-body mass, $m_{\gamma\gamma jj}$, in cases where there is only 1 b-tagged jet, an additional candidate jet is selected through the use of a BDT. The inputs to this BDT are: jet and dijet p_{T}, dijet mass, jet and dijet η, dijet $|\Delta\eta|$, booleans for passing the 77% and 85% b-tagging WP, and the ranking of the jets in terms of:

- Highest jet p_{T}.
- Lowest $\Delta m = m_{bb} - m_H$ when summed with the b-tagged jet.
- Highest dijet p_{T}, when summed with the b-tagged jet.

The jet with the highest BDT score is selected as the pairing when constructing $m_{\gamma\gamma jj}$. The performance is better at high mass hypotheses, since the signal shape is more pronounced from the background than at low masses.

Candidate jets are selected in the following way:

[3]This selection is similar to that of the default jet selection for ATLAS $H \to \gamma\gamma$ analyses. The $|\eta| < 2.5$ is an additional requirement imposed atop that selection for b-tagging.

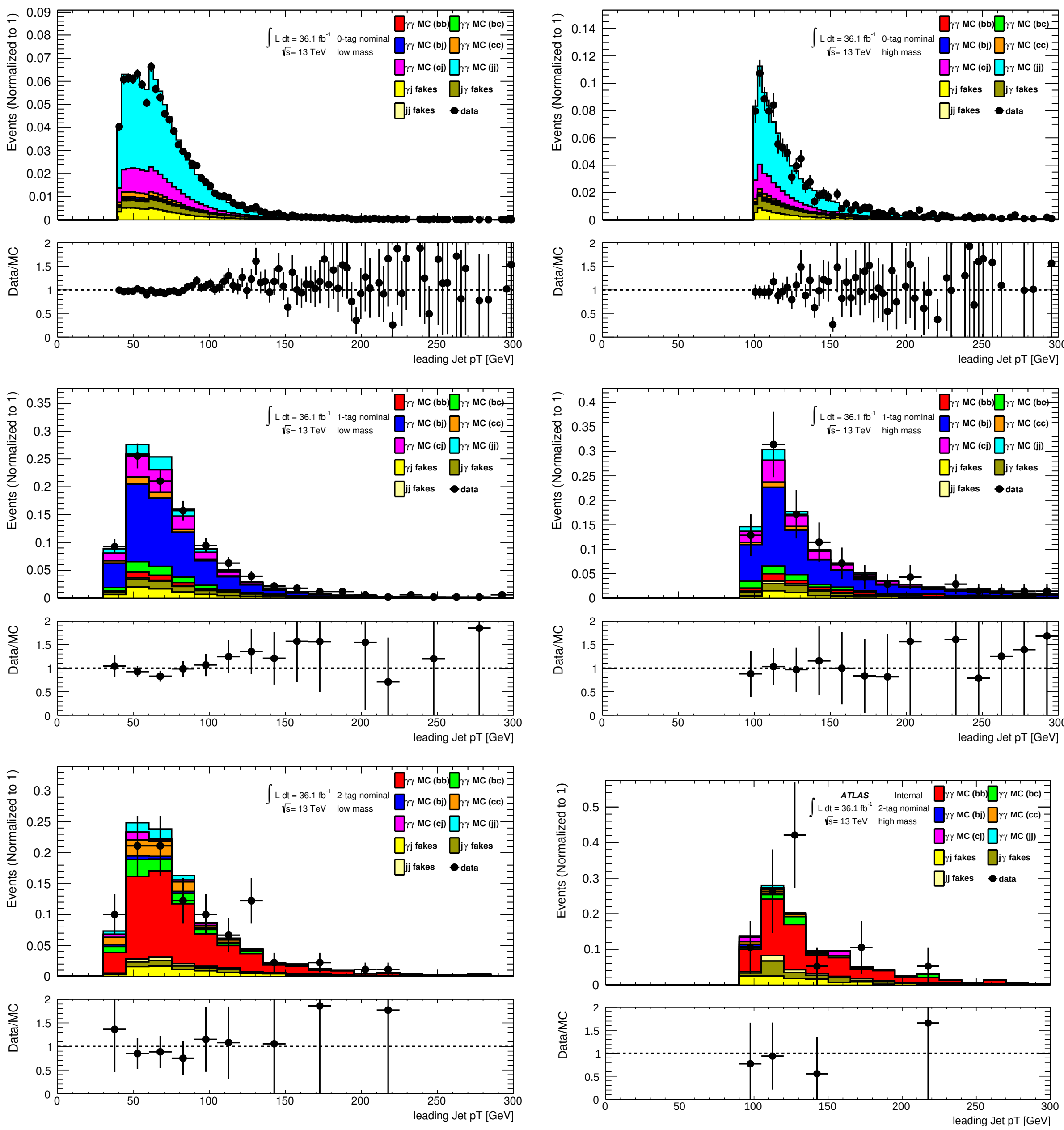


Figure 5.3: Leading jet p_{T} by b-tagging category. The low (left) and high (right) mass selections are shown. Both data and MC are normalized such that the integral is 1. Histograms are split by the flavor of the $H \to b\bar{b}$ decay products, where labels are taken from MC truth information. The γj, $j\gamma$, and jj fakes are taken from not-tight, not-isolated data and further described in Section 5.5.1.

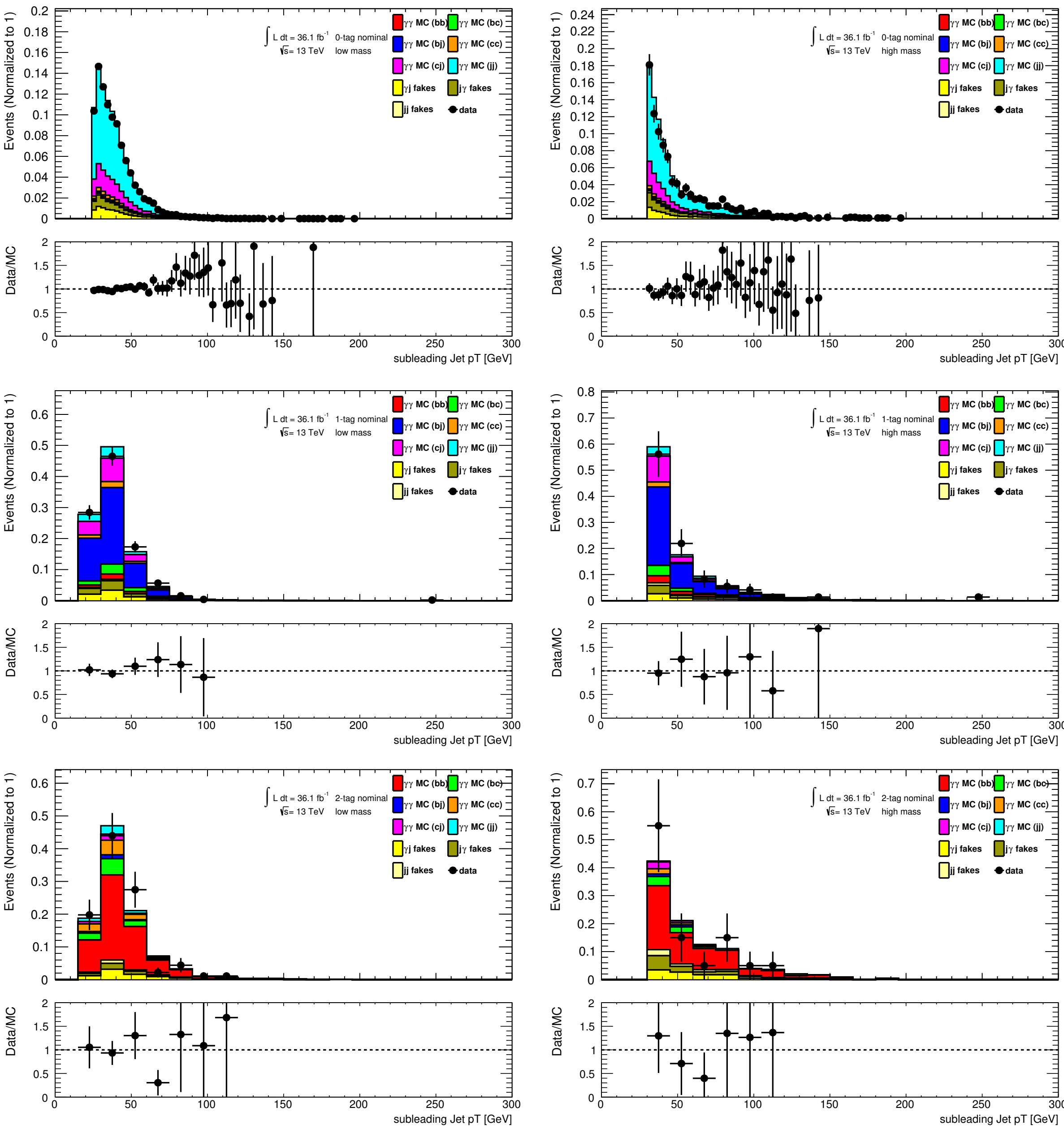


Figure 5.4: Subleading jet p_{T} by b-tagging category. The low (left) and high (right) mass selections are shown. Both data and MC are normalized such that the integral is 1. Histograms are split by the flavor of the $H \to b\bar{b}$ decay products, where labels are taken from MC truth information. The γj, $j\gamma$, and jj fakes are taken from not-tight, not-isolated data and further described in Section 5.5.1.

- In the 0-tag control region, the two leading jets are the candidate jets.
- In the 1-tag region, one of the jets is the b-tagged jet, the other is selected via a BDT
- In the 2-tag region, the two b-tagged jets are the candidate jets.

The leading jet p_{T} distribution is shown in Figure 5.3, the subleading jet p_{T} is shown in Figure 5.4. In these figures, the histograms are split by the MC truth flavor label of the $H \to b\bar{b}$ decay products. The γj, $j\gamma$, and jj fakes are taken from not-tight, not-isolated data and further described in Section 5.5.1.

5.3.4 Muons

In jet reconstruction, a muon-in-jet correction is used to improve resolution for b-jets. This accounts for leptonic decays involving muons which do not entirely deposit their energy in the EM or hadronic calorimeters. Muons must pass the ATLAS default medium requirements[4] [104], have $p_{\mathrm{T}} > 4$ GeV, $|\eta| < 2.7$, $|d_0|$ significance less[5] than 3, $|z_0|$ less than 0.5 mm.

For correcting the b-jets, 4-momenta of all muons within the jet's ΔR cone are added to the b-jet. Adding this correction shifts the m_{bb} distribution closer to the Higgs mass and provides a $5-6\%$ improvement in signal acceptance for each resonant mass point. Figure 5.5 shows the number of muons added to each jet, typically 0 or 1 muon.

[4] The medium criteria only uses CB and ME muons (defined in Section 3.2.3) and is middle between high selection efficiency and signal purity

[5] Both $|d_0|$ and $|z_0|$ requirements are with respect to the primary vertex.

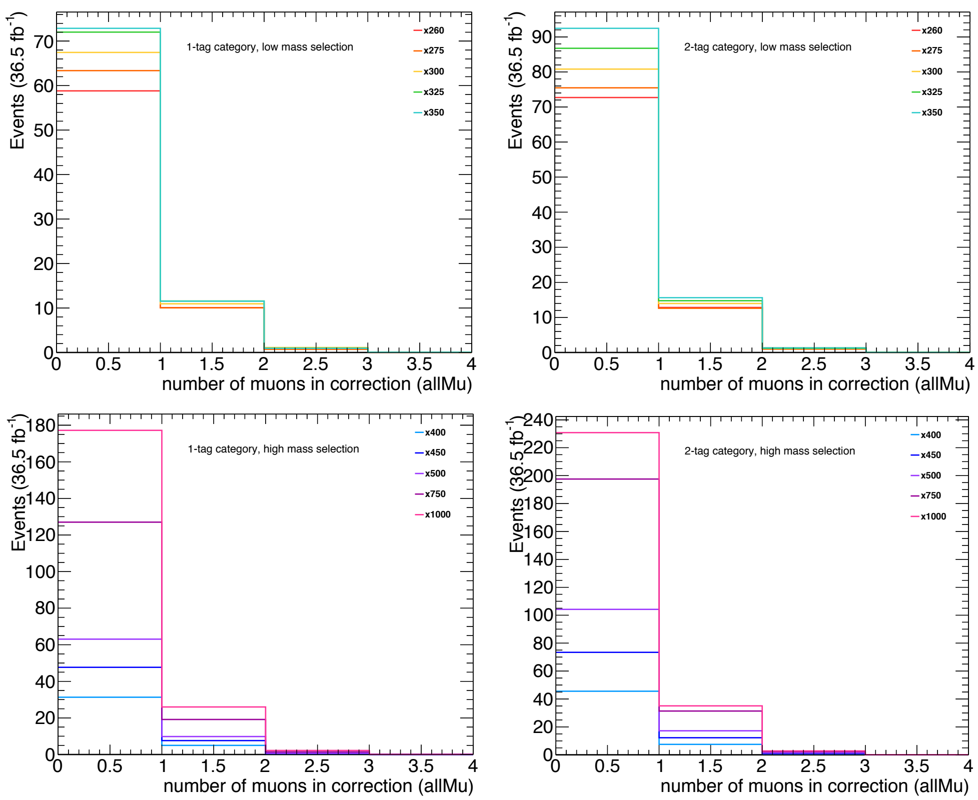


Figure 5.5: Number of muons added to each jet, shown for each b-tagging category with the low and high mass selections applied.

5.4 Event Selection

Preselection

For all samples, a preselection common to all $H \to \gamma\gamma$ analyses is applied [105]. To target this decay, events must pass the diphoton trigger, which requires at least two photons passing *loose* identification. One of these photons must have $E_{\mathrm{T}} > 35$ GeV and the other must have $E_{\mathrm{T}} > 25$ GeV. This is equivalent to the `g35_loose_g25_loose` trigger used on data. After this requirement, p_{T} cuts are applied, requiring the leading (subleading) photon p_{T} to be greater than 35% (25%) the diphoton invariant mass. Additionally the diphoton invariant mass must be in the window of 105 to 160 GeV.

Signal Selection

In addition to the preselection requirements, a further requirement of two jets within the b-tagging region ($|\eta| < 2.5$) is added to the preselection in order to target the $H \rightarrow b\bar{b}$ decay. Events are separated into 0, 1, and 2-tag regions based on b-tagging criteria. Events with 3 or more jets passing the 70% efficient b-tagging Working Point (WP) are vetoed from this analysis in order to maintain orthogonality with the $HH \rightarrow b\bar{b}b\bar{b}$ analysis. The 2-tag signal region requires exactly two jets passing the 70% efficient b-tagging WP. If an event fails this requirement, but has one b-jet passing the 60% efficient b-tagging WP, it is added to the 1-tag signal region. If it fails both of these criteria The 1 and 2-tag regions are used as signal regions, the 0-tag is used as a control region. This logic flow is shown in Figure 5.6.

After selection, overlapping low and high mass selections are created. In the resonant search, the low mass selection is for signal hypotheses of 500 GeV and below; the high mass selection is used for 500 GeV and above. In the non-resonant search the high mass hypothesis is used, and in the κ_λ scan the low mass hypothesis is used. These selections are defined by:

- **Low Mass:** Leading jet $p_{\mathrm{T}} > 40$ GeV, subleading jet $p_{\mathrm{T}} > 25$ GeV, and $80 < m_{bb}/\mathrm{GeV} < 140$.
- **High Mass:** Leading jet $p_{\mathrm{T}} > 100$ GeV, subleading jet $p_{\mathrm{T}} > 30$ GeV, and $90 < m_{bb}/\mathrm{GeV} < 140$.

Due to the poor mass resolution of the $b\bar{b}$ system, a scale factor is applied, assuming the decay is consistent with the signal hypothesis, a SM Higgs boson. The 4-vector of the system is rescaled by a factor of $m_H/m_{b\bar{b}}$ (with $m_H \equiv 125.09$ GeV). This scaling improves the $m_{\gamma\gamma b\bar{b}}$ resolution as shown in Figure 5.7.

Last, in only the resonant analysis, a tight diphoton mass cut is applied, isolating a window around the Higgs boson mass that contains 95% of signal events. This corresponds to $m_H \pm 4.7$ GeV for the low mass selection and $m_H \pm 4.3$ GeV for the high mass selection.

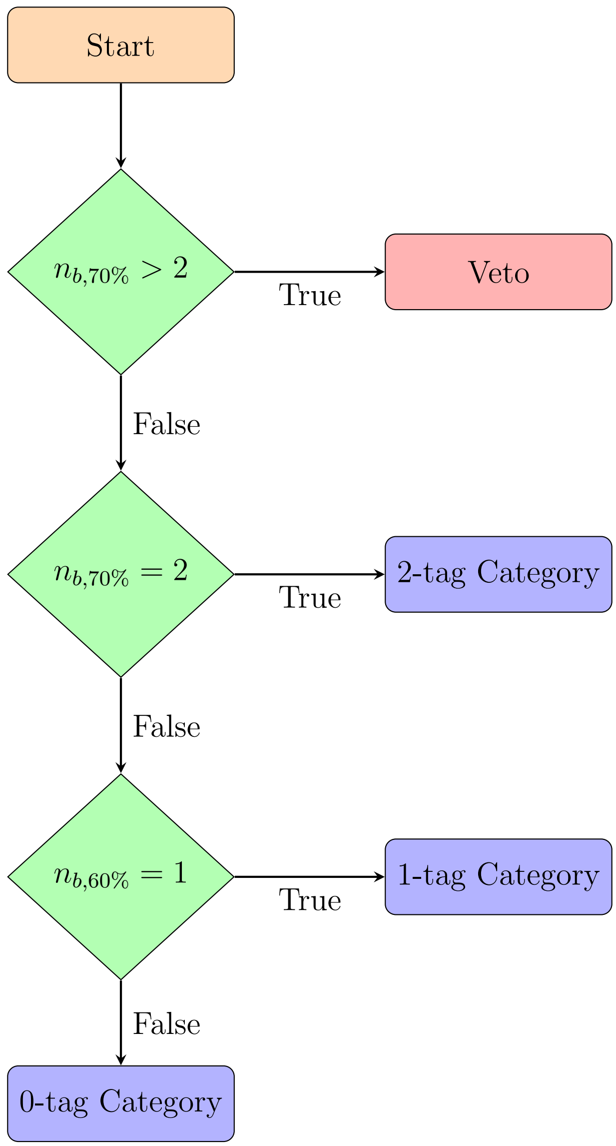
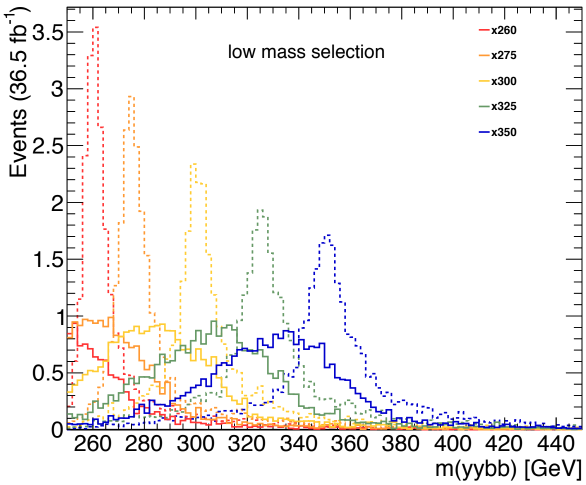
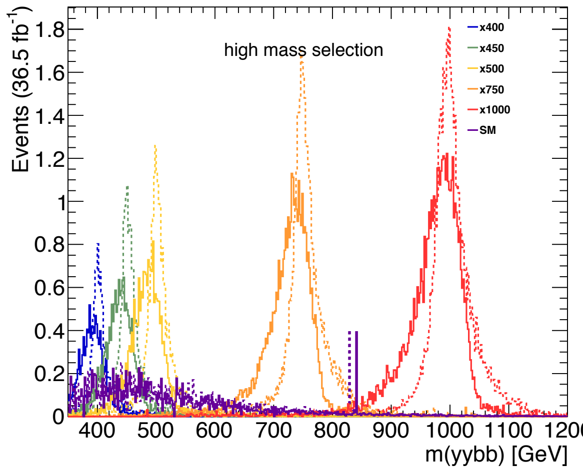


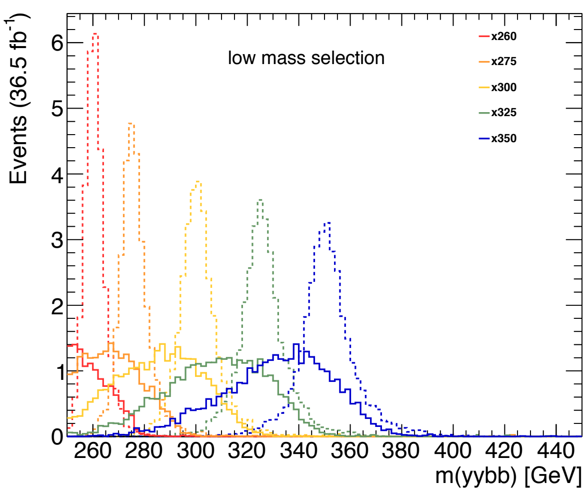
Figure 5.6: Logic flowchart for analysis b-tagging category definitions. The 1 and 2-tag categories are signal regions, the 0-tag category is a control region. Events with more than 2 b-tagged jets are rejected to maintain event orthogonality with the $HH \to b\bar{b}b\bar{b}$ analysis.



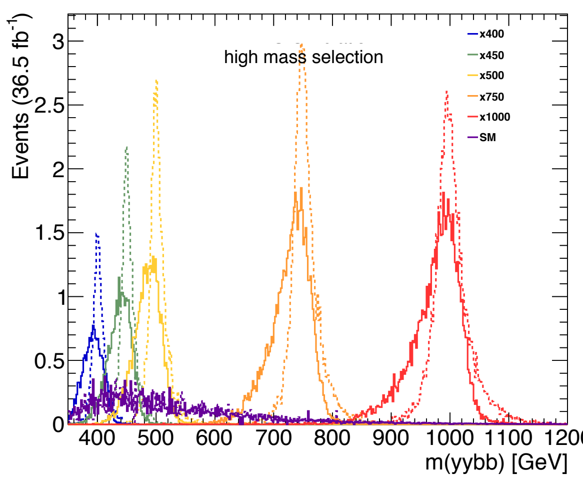
(a) Low Mass, 1-tag selection



(b) High Mass, 1-tag selection



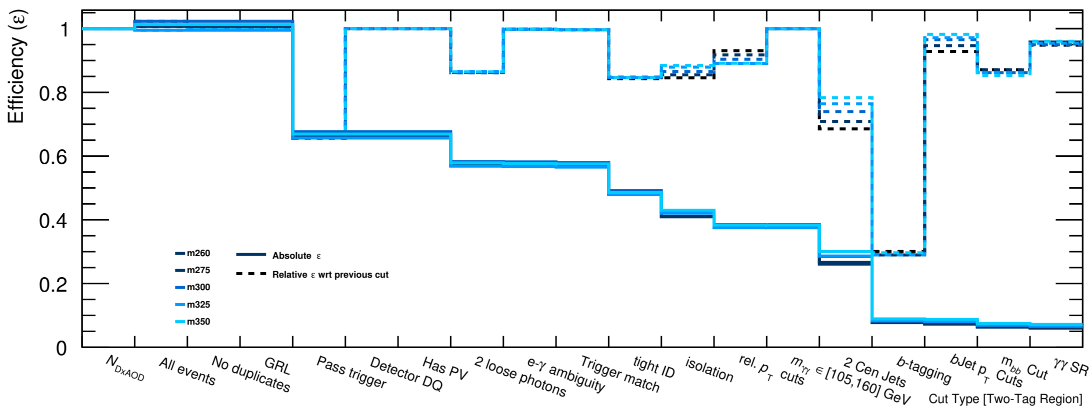
(c) Low Mass, 2-tag selection



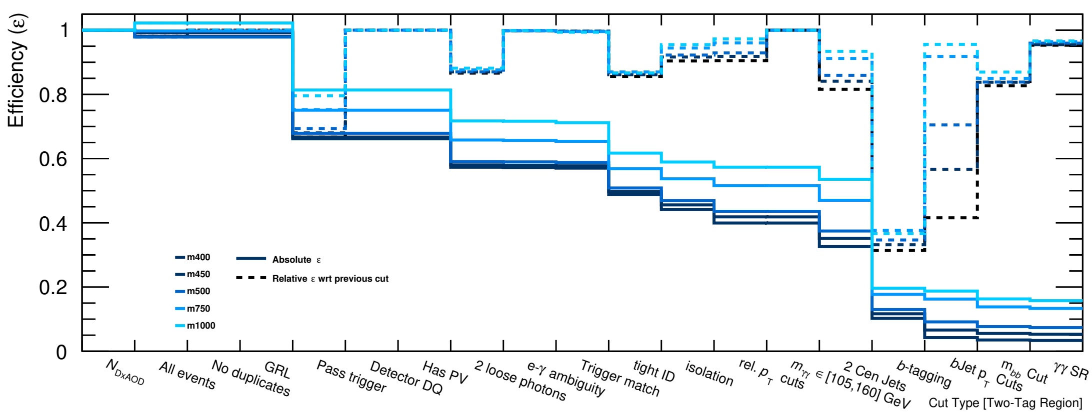
(d) High Mass, 2-tag selection

Figure 5.7: Reconstructed $m_{\gamma\gamma b\bar{b}}$ for the signal MC samples. The solid lines show the distribution before applying the dijet mass constraint, the dashed lines have the dijet mass constraint imposed. Low and High, 1 and 2-tag selections are shown.

The absolute and relative signal efficiency for the 2 b-tag region of the resonant analysis is shown in Figure 5.8. The yield and efficiency for each successive cut for the non-resonant, 2 b-tag region are in Tables 5.2 and 5.3 for the low and high mass selections, respectively. The background yield for the resonant and non-resonant selections are shown in Figure 5.4. Further cutflow studies (including the yield and efficiency tables for the 1 and 0 b-tag catgeories) can be found in Appendix D.



(a) Low Mass Selection



(b) High Mass Selection

Figure 5.8: Absolute (solid line) and relative (dotted line) efficiencies for the 2 b-tag resonant signal samples to pass each cut in the low (a) and high (b) mass selection. A description of the cuts is listed in the caption of Table 5.2.

Table 5.2: Cutflow for non-resonant $HH \rightarrow \gamma\gamma b\bar{b}$, 2 tag category (low mass selection). Cuts are applied successively from top to bottom. The "yield" is in number of events, the error is the absolute error on number of events, and the efficiency presented is the absolute efficiency. From top to bottom, the cuts are as follows: Number of events in the original file (N_{xAOD}), number of events in the derivation (N_{DxAOD}), total events code, events after duplicate removal, events in the Good Run List (GRL), events passing diphoton trigger requirements, events passing detector data quality cuts, events with a primary vertex selected, events with 2 photons, events that pass the $e-\gamma$ ambiguity requirements, events that pass trigger matching, events meeting photon isolation criteria, events passing relative p_{T} thresholds, events within the [105, 160] GeV $m_{\gamma\gamma}$ window defined, events with 2 central jets, events passing b-tagging requirements, events where the b jets meet criteria, events passing the $m_{b\bar{b}}$ criteria, and last, events in the 120-130 GeV $\gamma\gamma$ signal region. The last cut is not used for the non-resonant analysis. Cuts through the $m_{\gamma\gamma}$ window are the standardized $H \rightarrow \gamma\gamma$ preselection.

Cuts	2 b-tag		
	Yield	Error	Efficiency (%)
N_{xAOD}	3.147	0.027	–
N_{DxAOD}	3.147	0.019	–
All events	3.181	0.024	100.0
No duplicates	3.181	0.024	100.0
GRL	3.181	0.024	100.0
Pass trigger	2.184	0.021	68.7
Detector DQ	2.184	0.021	68.7
Has PV	2.184	0.021	68.7
2 *loose photons*	1.894	0.019	59.5
$e-\gamma$ *ambiguity*	1.891	0.019	59.5
Trigger match	1.885	0.019	59.3
tight ID	1.617	0.018	50.8
isolation	1.476	0.017	46.4
rel. p_T	1.362	0.017	42.8
$m_{\gamma\gamma} \in [105, 160]$ *GeV*	1.362	0.017	42.8
2 *Cen Jets*	1.149	0.016	36.1
$b-$*tagging*	0.373	0.008	11.7
bJet p_T	0.369	0.008	11.6
m_{bb} window	0.318	0.007	10.0
$\gamma\gamma$ *SR*	0.305	0.007	9.6

Table 5.3: Cutflow for non-resonant $HH \to \gamma\gamma b\bar{b}$, 2 tag category (high mass selection). A description of the cuts is listed in the caption of Table 5.2.

Cuts	2 b-tag		
	Yield	Error	Efficiency (%)
N_{xAOD}	3.147	0.027	–
N_{DxAOD}	3.147	0.019	–
$All\ events$	3.181	0.024	100.0
$No\ duplicates$	3.181	0.024	100.0
GRL	3.181	0.024	100.0
$Pass\ trigger$	2.184	0.021	68.7
$Detector\ DQ$	2.184	0.021	68.7
$Has\ PV$	2.184	0.021	68.7
$2\ loose\ photons$	1.894	0.019	59.5
$e - \gamma\ ambiguity$	1.891	0.019	59.5
$Trigger\ match$	1.885	0.019	59.3
$tight\ ID$	1.617	0.018	50.8
$isolation$	1.476	0.017	46.4
$rel.\ p_T$	1.362	0.017	42.8
$m_{\gamma\gamma}\ \in\ [105, 160]\ GeV$	1.362	0.017	42.8
$2\ Cen\ Jets$	1.149	0.016	36.1
$b - tagging$	0.373	0.008	11.7
$bJet\ p_T$	0.219	0.005	6.9
m_{bb} window	0.183	0.005	5.8
$\gamma\gamma\ SR$	0.175	0.005	5.5

Table 5.4: Background yield in the 2 b-tag region split by process for the non-resonant and resonant analyses. The non-resonant analysis requires passing of all cuts but the $m_{\gamma\gamma}$ SR constraint, and is denoted by the "Yield" column. The resonant analysis includes the additional $m_{\gamma\gamma}$ constraint and is denoted "SR Yield."

2-tag Category	Low Mass				High Mass			
Background	Yield	Error	SR Yield	Error	Yield	Error	SR Yield	Error
$\gamma\gamma$-continuum	109.79		21.03		23.58		4.03	
ggH	0.315	0.035	0.306	0.035	0.074	0.013	0.073	0.013
VBF H	0.042	0.004	0.041	0.004	0.007	0.002	0.006	0.002
W^-H	0.004	0.001	0.004	0.001	0.002	0.001	0.001	0.001
W^+H	0.007	0.002	0.006	0.002	0.002	0.001	0.002	0.001
ZH	0.388	0.009	0.378	0.009	0.095	0.005	0.094	0.005
$ggZH$	0.129	0.007	0.128	0.007	0.048	0.004	0.048	0.004
ttH	1.363	0.015	1.331	0.015	0.353	0.008	0.343	0.008
bbH (positive)	0.016	0.024	0.014	0.023	-0.005	0.005	-0.005	0.005
bbH (negative)	-0.004	0.001	-0.005	0.001	0.000	0.000	0.000	0.000

5.5 Modeling

In both the resonant and non-resonant analyses, a full signal+background fit is performed. For the non-resonant analysis, the $m_{\gamma\gamma}$ spectrum is fitted, while for the resonant analysis $m_{\gamma\gamma b\bar{b}}$ is fitted. This section describes the parameterization to model the signal and background of each analysis.

5.5.1 Background Composition

The $\gamma\gamma$-continuum background contains contributions from both diphoton events, as well as those where jets fake photons, denoted $j\gamma$, γj, or jj, where the leading, subleading, or both photons have been misidentified jets, respectively. To establish the contributions of each of these sources, a 2x2D sideband method [106] is used. This method varies the photon identification and isolation criteria to create a signal region and 15 control regions, then expresses the yield, identification efficiency, and isolation efficiency for both photons and jets in each region. Using this, along with the correlations between isolation distributions, a global minimization procedure is used to calculate the $\gamma\gamma$, $j\gamma$, γj, and jj yield separately for the 1 and 2 b-tag categories. In all cases, the $\gamma\gamma$ contribution is dominant, representing 80-90% of events. This method is described in detail in Appendix C. The relative contributions are shown in Figure 5.9.

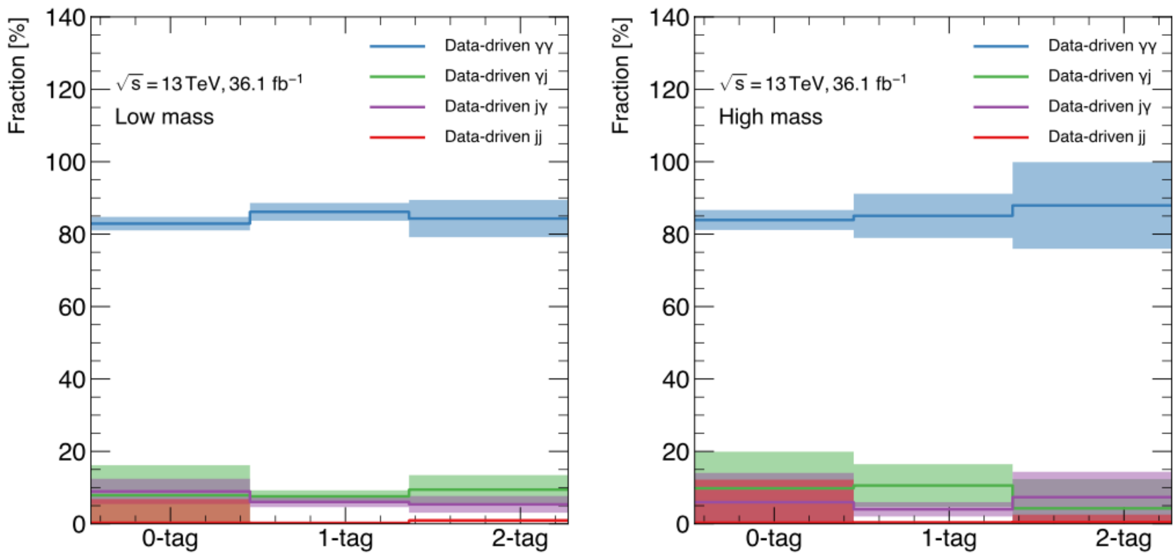


Figure 5.9: Expected contributions from $\gamma\gamma$, γj, $j\gamma$ and jj to the 2015+2016 data for the low mass (left) and high mass (right) selections.

In order to get expected shapes of the $\gamma\gamma$, γj, $j\gamma$ and jj contributions, the 2x2D method is applied bin-wise in $m_{\gamma\gamma}$ (bins defined in width of 2.5 GeV between 105 and 160 GeV). First, a template for each category is produced using the high stats $\gamma\gamma$-continuum sample. For the $\gamma\gamma$ contribution, this requires both photons passing both tight identification and FixedCutLoose isolation. For the γj, $j\gamma$ and jj, this requires photons which pass loose requirements, but do not fall into the $\gamma\gamma$ category.

These templates from MC are then compared to the data-driven 2x2D method and reweighted to better match data. The template and 2x2D templates are separately fitted and their ratio is used as a smooth reweighting function in $m_{\gamma\gamma}$. Figures 5.10 and 5.11 show this for the low and high mass selection respectively, in the 0-tag region. There are insufficient statistics to perform this in the 1 and 2 b-tag regions, so the same reweighting functions from the 0-tag region are applied. For the final continuum $\gamma\gamma$ background prediction, the proportions shown in Figure 5.9 are used. The fractional contribution is shown in Table 5.5. The normalization is chosen such that the sideband region ($105 < m_{\gamma\gamma} < 120$ GeV and $130 < m_{\gamma\gamma} < 160$ GeV) contribution is equal to the contribution in data.

Table 5.5: Expected fractional contributions from $\gamma\gamma$, γj, $j\gamma$ and jj to the events passing the nominal selection in each of the 0-, 1- and 2-tag categories.

	0-Tag		1-Tag		2-Tag	
	Low mass	High mass	Low mass	High mass	Low mass	High mass
$\gamma\gamma$	0.99 ± 0.08	0.85 ± 0.01	0.85 ± 0.02	0.84 ± 0.09	0.85 ± 0.05	0.68 ± 0.22
γj	0.00 ± 0.04	0.07 ± 0.01	0.07 ± 0.02	0.07 ± 0.07	0.08 ± 0.04	0.09 ± 0.14
$j\gamma$	0.00 ± 0.03	0.07 ± 0.01	0.07 ± 0.02	0.08 ± 0.04	0.06 ± 0.03	0.18 ± 0.13
jj	0.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.01	0.02 ± 0.01	0.05 ± 0.06

5.5.2 Modeling in the Non-resonant Analysis

5.5.2.1 Signal Modeling

A double-sided Crystal Ball function [107], a Gaussian core with power-law tails, is used to describe the shape of the $HH \to \gamma\gamma b\bar{b}$ signal. Model parameters are taken from the simulated signal sample. Figure 5.12 shows the parameterization using the high mass selection, and Figure 5.13 shows the parameterization with the low mass selection.

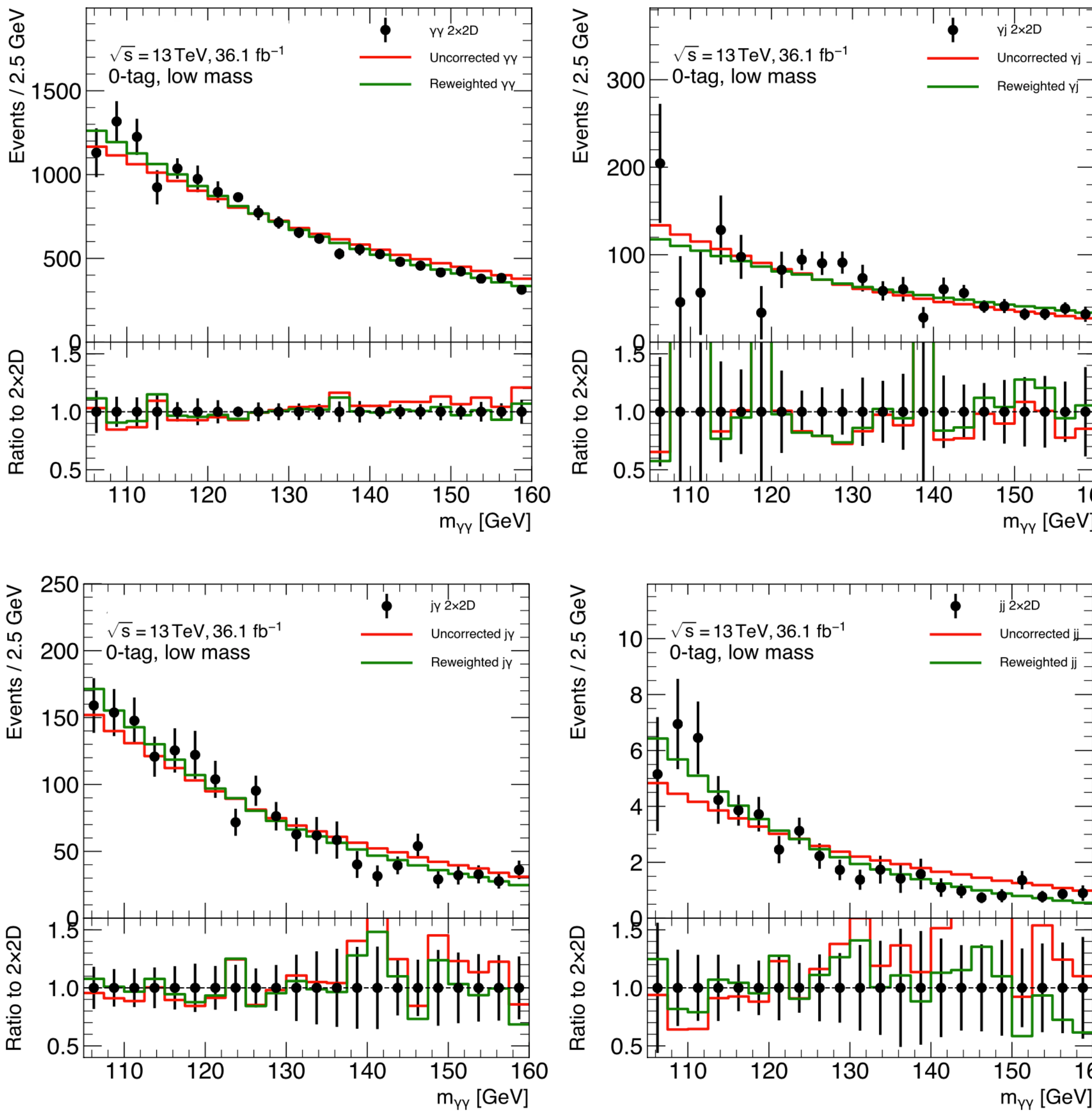


Figure 5.10: Template distributions for each of the $\gamma\gamma$, γj, $j\gamma$ and jj components from Sherpa before (red) and after reweighting (green) are compared to the output of the 2x2D method for 0-tag events with the low mass selection.

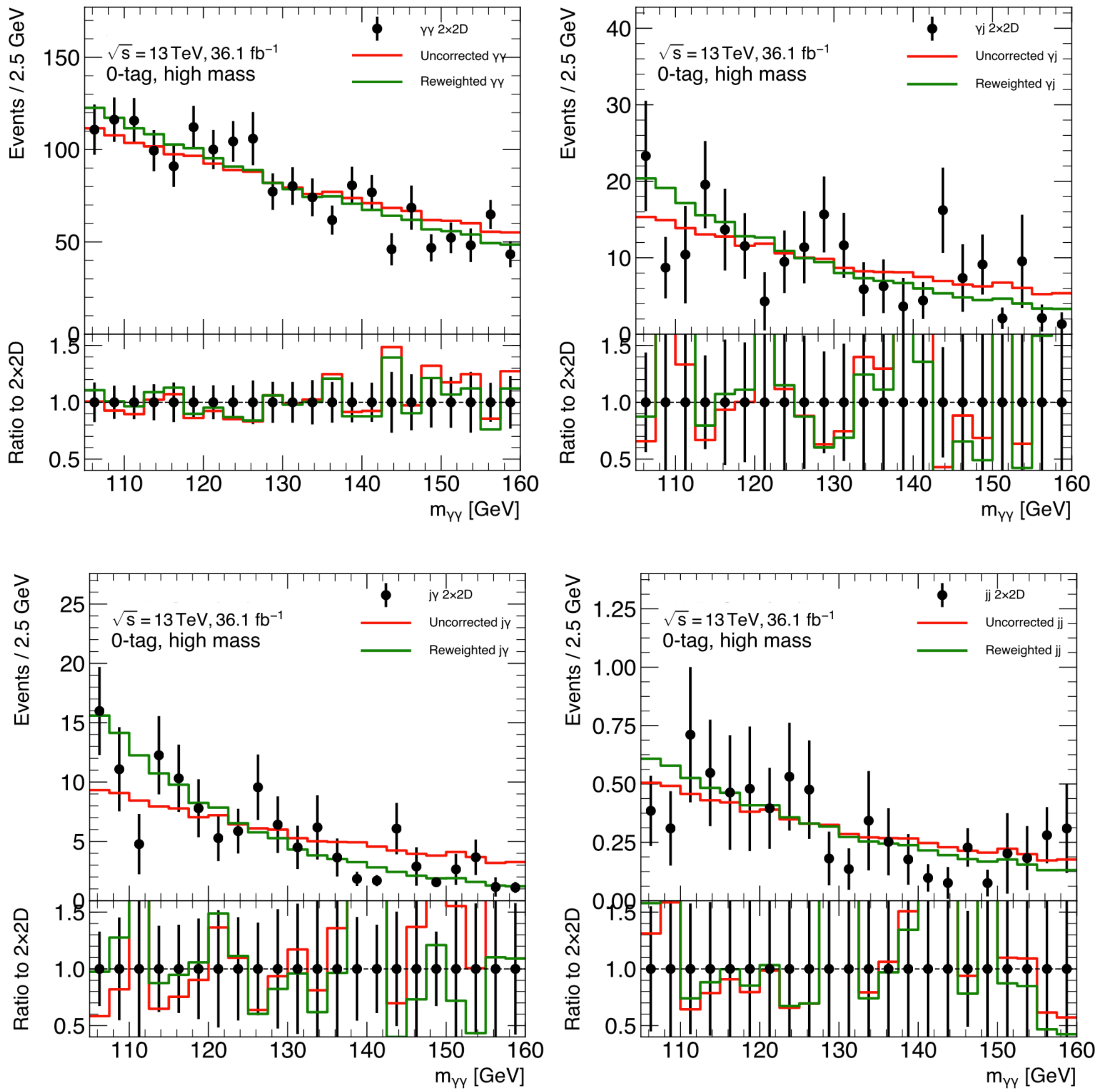


Figure 5.11: Template distributions for each of the $\gamma\gamma$, γj, $j\gamma$ and jj components from Sherpa before (red) and after reweighting (green) are compared to the output of the 2x2D method for 0-tag events with the high mass selection.

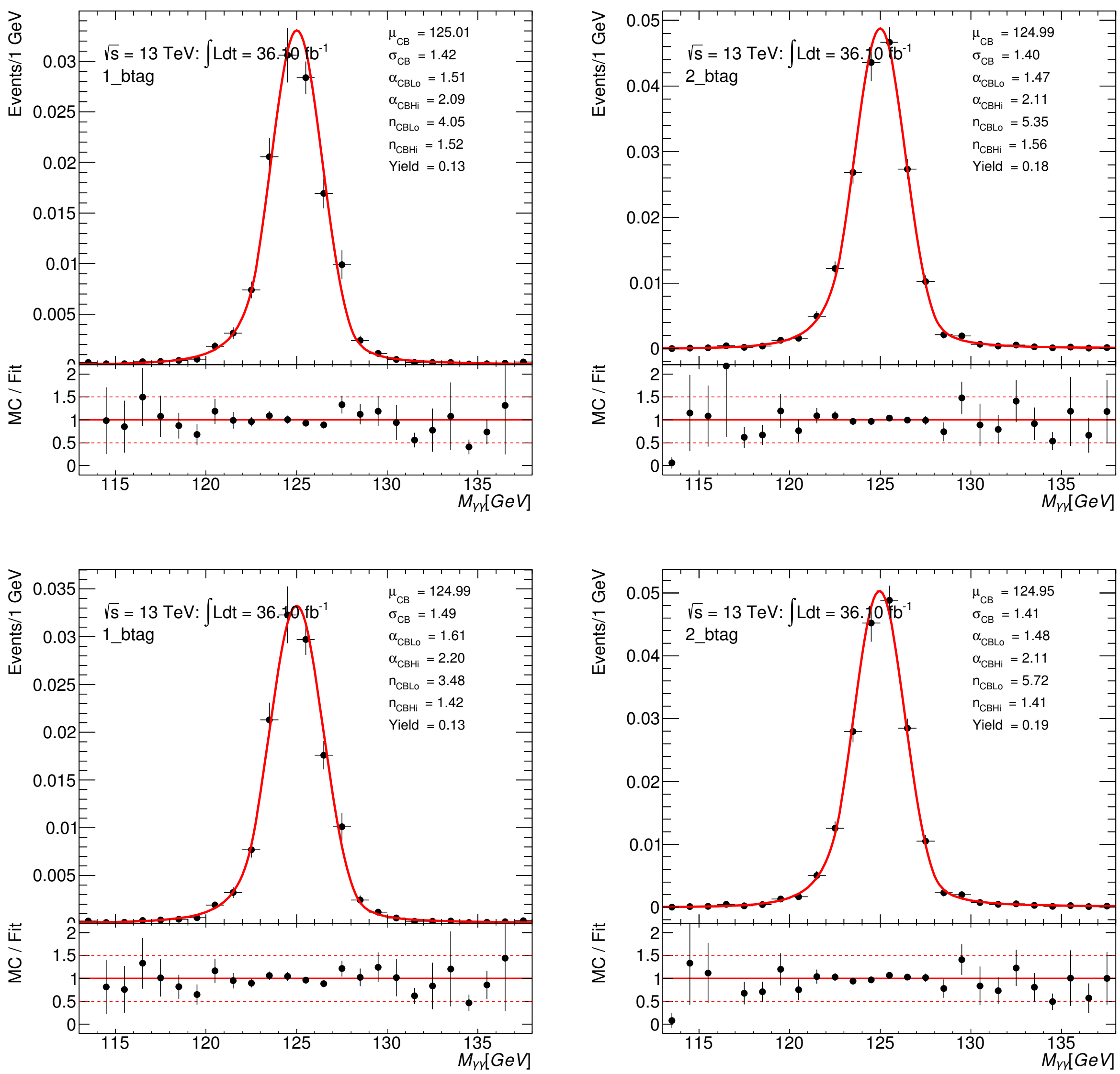


Figure 5.12: Parameterization of the $m_{\gamma\gamma}$ distribution from the SM di-Higgs simulated samples at a mass of 125 GeV in the 1-tag (left) and 2-tag (right) categories, with (top) and without the signal MC reweighting to full NLO (bottom). Events were required to pass the high mass selection.

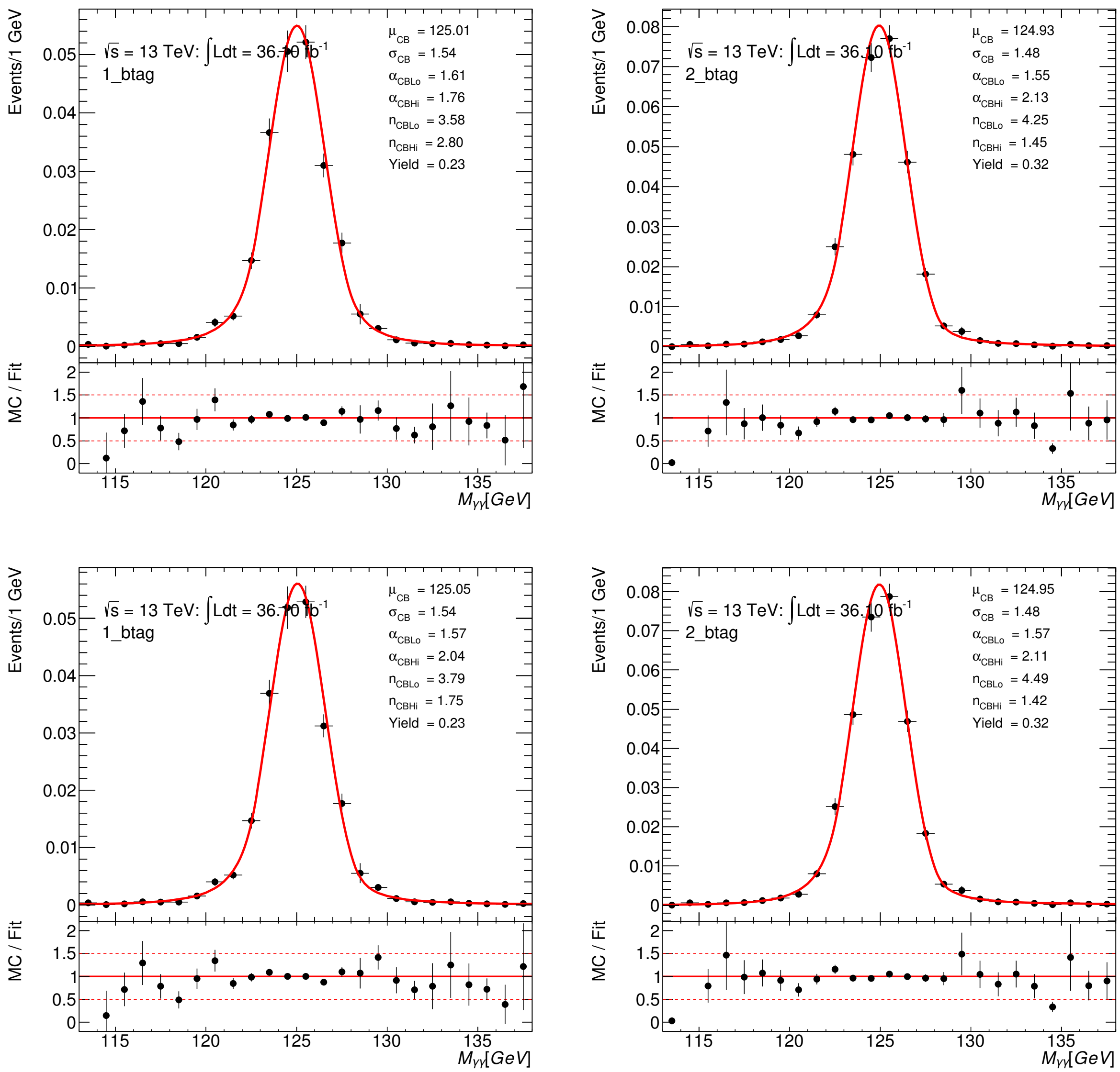


Figure 5.13: Parameterization of the $m_{\gamma\gamma}$ distribution from the SM di-Higgs simulated samples at a mass of 125 GeV in the 1-tag (left) and 2-tag (right) categories, with (top) and without the signal MC reweighting to full NLO (bottom). Events were required to pass the low mass selection.

5.5.2.2 Background Modeling

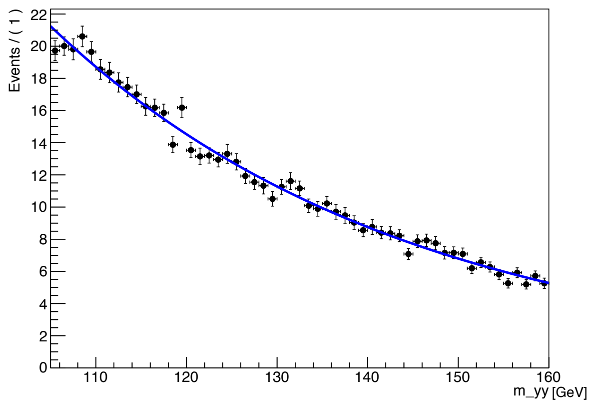
The background for the non-resonant analysis is composed of a model for the smoothly-falling $\gamma\gamma$-continuum, and a model for single Higgs boson production.

For the $\gamma\gamma$-continuum background, a functional form is fitted to data. The fit function is selected through a *spurious signal* test [108], which evaluates potential bias from this procedure. This test performs signal-plus-background fits to just the background, evaluating the extracted signal in the 121 GeV $< m_{\gamma\gamma} <$ 129 GeV window, which is taken as the bias. This procedure is repeated for each considered function, and the function with the fewest spurious signal events is selected as the background function. If functions have the same bias, then by an Occam's Razor argument, the function with fewer free parameters is selected. The considered functions for this test are as follows:

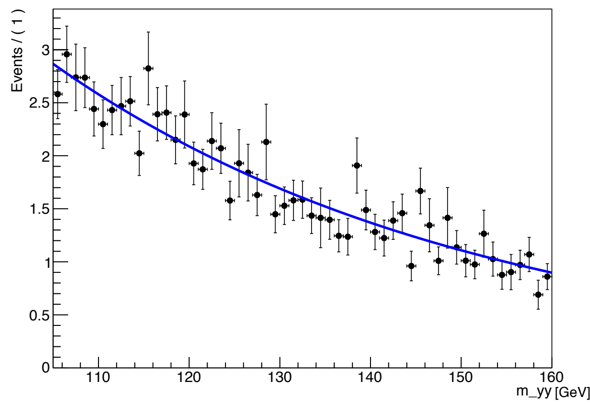
- Exponential, consider orders $N = 1, 2, 3$. Form: $Exp^N(m_{\gamma\gamma}; \theta^{bkg}) = \exp(\sum_{j=0}^{N} \theta_j^{bkg} \cdot m_{\gamma\gamma}^j)$
- Bernstein, consider orders $N = 3, 4, 5$. Form: $Bernstein^N(m_{\gamma\gamma}; \theta^{bkg}) = \sum_{j=0}^{N} \theta_j^{bkg} b_{j,N}$ for $b_{j,N} = C_n^i x^j (1-x)^{N-j}$, where $x = (m_{\gamma\gamma} - 100)/60$
- Dijet of form: $dijet(m_{\gamma\gamma}; \theta^{bkg}) = \theta_1^{bkg}(1 - m_{\gamma\gamma})^{\theta_2^{bkg}} m_{\gamma\gamma}^{\theta_3^{bkg}}$

Through this test, the first-order exponential function is found to have the lowest bias, and is selected as the $\gamma\gamma$-continuum background model. This fit is shown for both selections in the 1 and 2-tag categories in Figure 5.14.

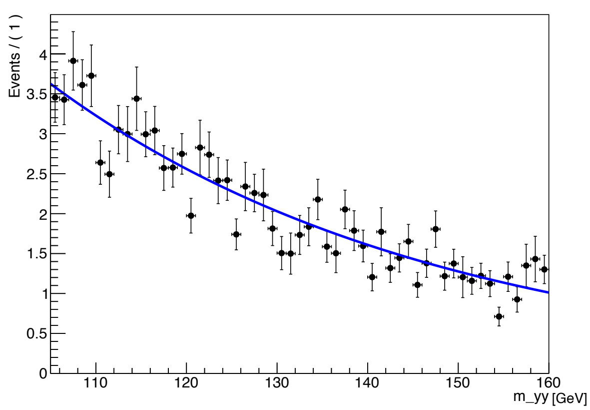
The single Higgs boson background is modeled through a double-sided Crystal Ball function, where parameters are taken from simulation. These fits are shown in Figure 5.15.



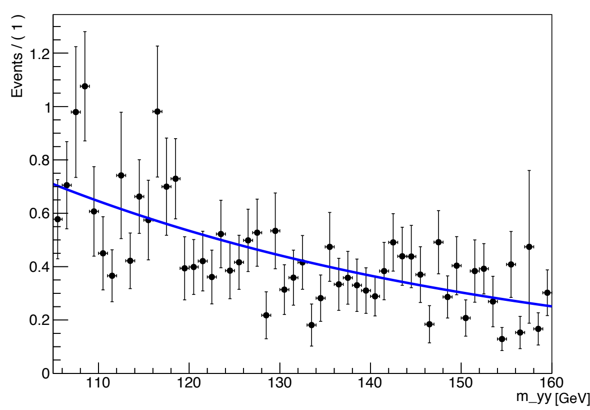
(a) Low mass selection, 1 b-tag



(b) High mass selection, 1 b-tag



(c) Low mass selection, 2 b-tag



(d) High mass selection, 2 b-tag

Figure 5.14: Background-only fit to the $m_{\gamma\gamma}$ distribution using the low and high-mass selections, for the 1 and 2 b-tag categories using the first-order exponential function.

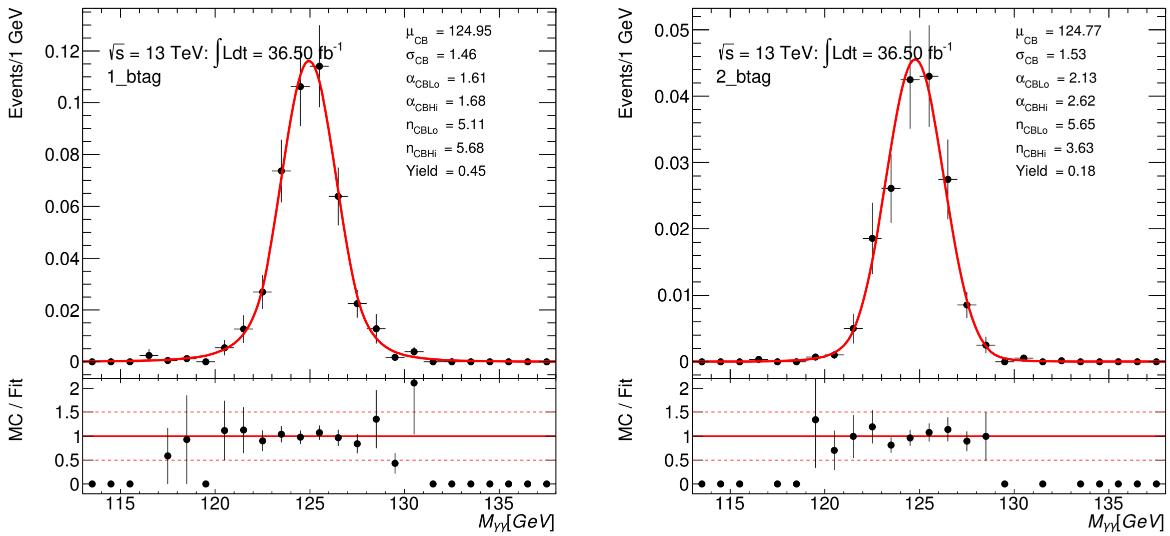


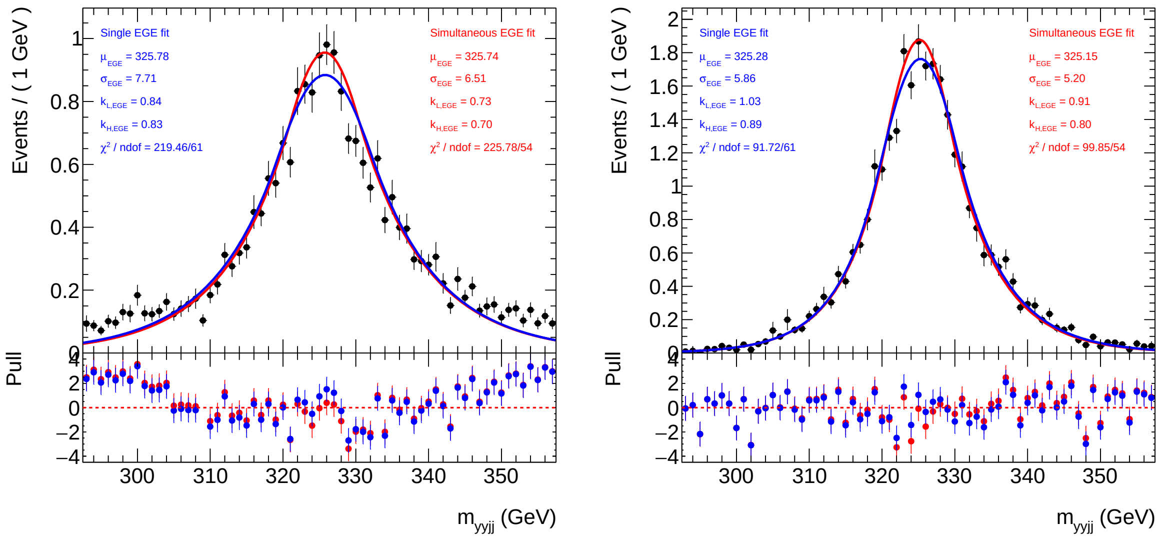
Figure 5.15: Parameterization of the $m_{\gamma\gamma}$ distribution from the SM single-Higgs simulated samples at a mass of 125 GeV in the 1-tag (left) and 2-tag (right) categories. Events were required to pass the high mass selection.

5.5.3 Modeling in the Resonant Analysis

5.5.3.1 Signal Modeling

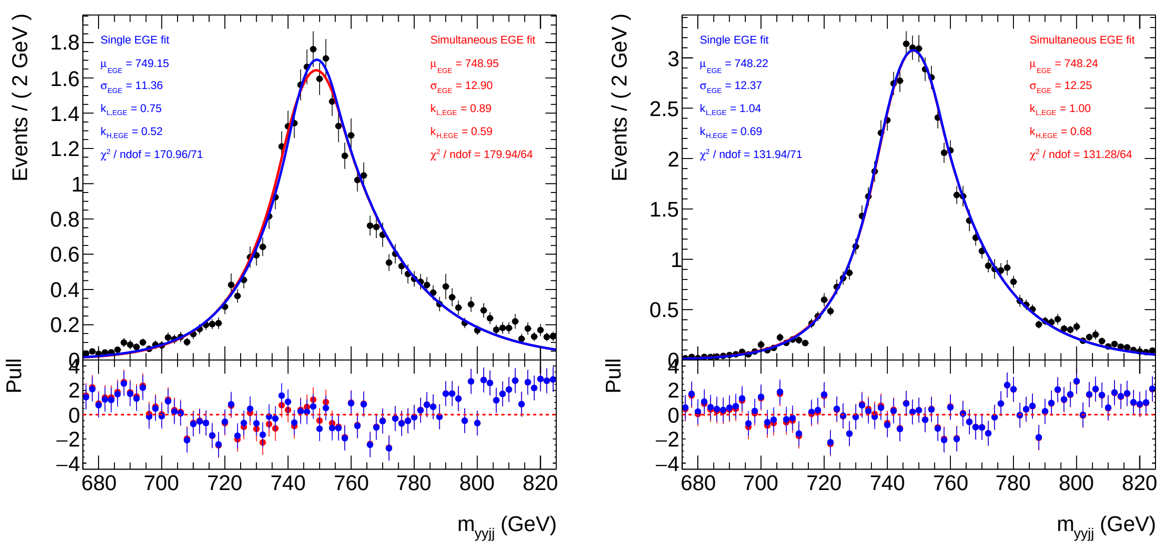
For the resonant analysis, several functional forms were considered, including `ExpGaussExp`[6], Crystal Ball + Gaussian, Crystal Ball + Voigt, and a Double Sided Crystal Ball. The model with the fewest parameters that converged was selected. This model was the `ExpGaussExp`, a function using a gaussian core with exponential tails. The mass hypotheses were fitted simultaneously, which yielded more stable fits. Sample fits for two mass hypotheses, one high mass (750 GeV) and one low mass (325 GeV), are shown in Figure 5.16.

[6] Also called a "double-shouldered GaussExp."



(a) 325 GeV, 1 b-tag, low mass selection

(b) 325 GeV, 2 b-tag, low mass selection



(c) 750 GeV, 1 b-tag, high mass selection

(d) 750 GeV, 2 b-tag, high mass selection

Figure 5.16: Parameterization with an `ExpGaussExp` of the $m_{\gamma\gamma b\bar{b}}$ distribution from one low mass (325 GeV) and one high mass (750 GeV) resonant di-Higgs sample. The mass hypotheses, selection used, and b-tag category are indicated.

5.5.3.2 Background Modeling

Similar to the non-resonant analysis, a spurious signal test (described in Section 5.5.2.2) is performed to select the background fit function. The fit is performed in both the 1 and 2-tag categories, where the same generalized function is used to correlate uncertainties, however the parameter values can differ between categories. Different fit functions are selected for the low and high mass categories, where the low mass fit must contain a characteristic peak, and the high mass fit must be monotonically decreasing. For the low mass region (using the loose selection), the following functions were considered:

- Novosibirsk: $P(x) = \exp(-0.5^{\ln(q_y)^2/\Lambda^2+\Lambda^2})$ with $q_y = 1 + \Lambda(x - x_0)/\sigma \times \frac{\sinh(\Lambda\sqrt{\ln(4)})}{\Lambda\sqrt{\ln(4)}}$.
- Modified Gamma: a Gamma distribution where the shape and scale parameters are replaced by linear functions of the mass hypothesis. This adds 2 additional degrees of freedom (5 total).
- Modified Landau: a Landau distribution where the shape parameter is replaced by a linear function of the mass hypothesis. This adds 1 additional degree of freedom (3 total).

For the high mass region (using the tight selection), the following functions were considered:

- Exponential (x): $Exp[x](m_{\gamma\gamma b\bar{b}}; p_0) = \exp(p_0 \cdot m_{\gamma\gamma b\bar{b}})$
- Exponential (x^2): $Exp[x^2](m_{\gamma\gamma b\bar{b}}; p_0) = \exp(p_0 \cdot m^2_{\gamma\gamma b\bar{b}})$
- Inverse polynomial (x^{-2}): $Inv\ Poly[x^{-2}](m_{\gamma\gamma b\bar{b}}; p_0) = 1 + p_0/(m^2_{\gamma\gamma b\bar{b}})$
- Inverse polynomial (x^{-3}): $Inv\ Poly[x^{-3}](m_{\gamma\gamma b\bar{b}}; p_0) = 1 + p_0/(m^3_{\gamma\gamma b\bar{b}})$
- Power Law: $Power\ Law(m_{\gamma\gamma b\bar{b}}; p_0) = m^{p_0}_{\gamma\gamma b\bar{b}}$

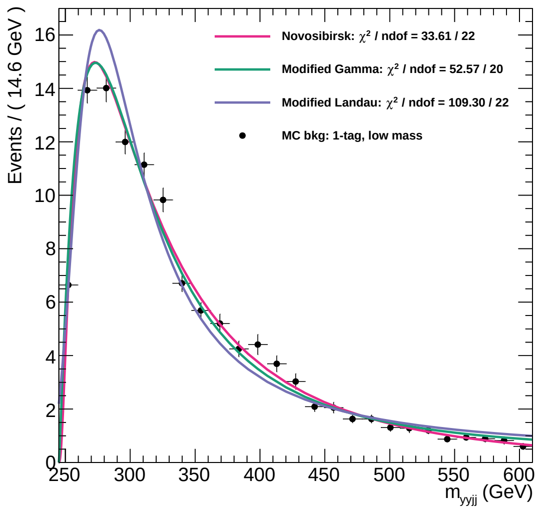
The number of spurious signal events were obtained by performing S+B fits in the $m_{\gamma\gamma b\bar{b}}$ spectrum separately in the 1 and 2-tag categories. For the low mass region, the fits were performed from 245 GeV to 485 GeV, for the high mass region the fits were performed from 335 GeV to 1140 GeV, which correspond to 95% signal coverage of the resonant masses considered. The results of the 1-tag spurious signal test are shown in Table 5.6, and the 2-tag results are shown in Table 5.7, with the systematic uncertainty on the signal amplitude in terms of spurious signal N_{spur} and its ratio to the statistical uncertainty on the fitted number of signal events Z_{spur} is its ratio the fitted number of signal events.

The Novosibirsk function is selected for the low mass region and the exponential function is selected for the high mass region. For the 2 b-tag category, more complicated functions have slightly smaller N_{spur}, but failed to fit high-mass distributions well in pseudodata trials. Background-only fits to using these functions are shown for the 1 and 2 b-tag categories in Figure 5.17.

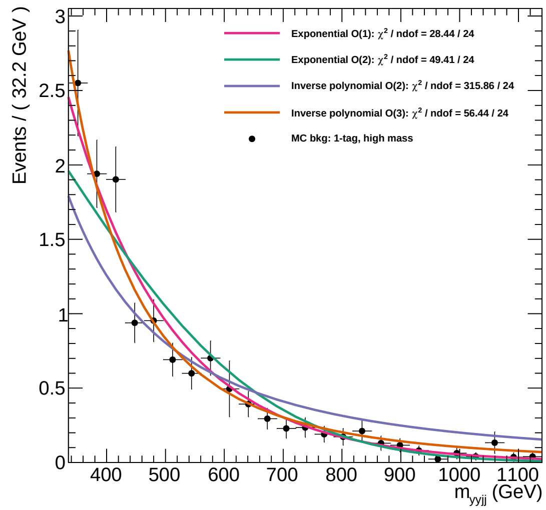
Table 5.6: Spurious signal test results for the 1 b-tag category. N_{spur} is the spurious signal amplitude, Z_{spur} is its ratio the fitted number of signal events. χ^2/ndof values is obtained from a background-only fit. Parameters gives the number of floating parameters in the given function.

Low Mass				
Model	Maximum Z_{spur} [%]	Maximum N_{spur}	Parameters	χ^2/ndof
Novosibirsk	33.64	2.06	3	36.7/22
Modified Gamma	53.67	3.42	5	45.2/20
Modified Landau	96.93	7.54	3	128.3/22
High Mass				
Model	Maximum Z_{spur} [%]	Maximum N_{spur}	Parameters	χ^2/ndof
Exp (x)	63.66	0.89	1	32.5/26
Exp (x^2)	73.64	0.97	1	53/9/26
Inv. Poly (x^{-2})	355.52	1.51	1	354.7/26
Inv. Poly (x^{-3})	146.35	1.03	1	68.8/26
Power Law	78.00	0.96	1	45.0/26

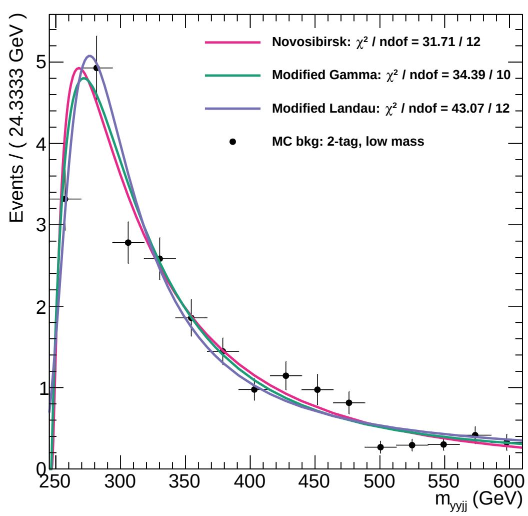
For the low-mass fit, both the signal and background fit functions have a peak shape, which can result in a bias in the extracted cross-section. As a means to stabilize the background



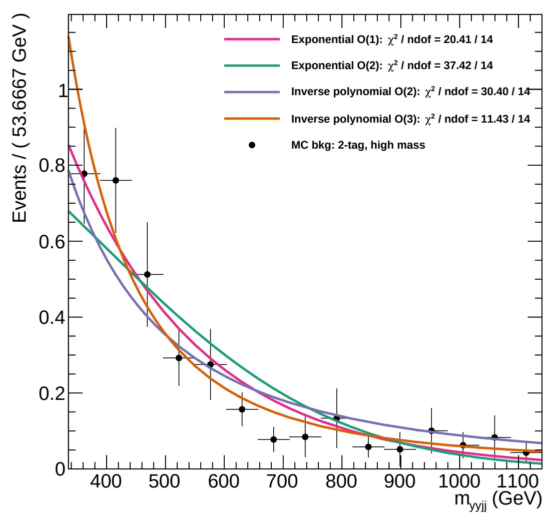
(a) 1 b-tag, low mass



(b) 1 b-tag, high mass



(c) 2 b-tag, low mass



(d) 2 b-tag, high mass

Figure 5.17: Background-only fit to the $m_{\gamma\gamma b\bar{b}}$ distribution. The b-tag category and selection used are indicated.

Table 5.7: Spurious signal test results for the 2 b-tag category. N_{spur} is the spurious signal amplitude, Z_{spur} is its ratio to the fitted number of signal events. χ^2/ndof values is obtained from a background-only fit. Parameters gives the number of floating parameters in the given function.

Low Mass				
Model	Maximum Z_{spur} [%]	Maximum N_{spur}	Parameters	χ^2/ndof
Novosibirsk	56.97	0.58	3	33.9/22
Modified Gamma	82.05	0.77	5	36.3/20
Modified Landau	65.62	1.09	3	45.8/22
High Mass				
Model	Maximum Z_{spur} [%]	Maximum N_{spur}	Parameters	χ^2/ndof
Exp (x)	98.94	0.21	1	36.4/26
Exp (x^2)	123.18	0.27	1	59.0/26
Inv. Poly (x^{-2})	88.33	0.25	1	42.9/26
Inv. Poly (x^{-3})	116.84	0.13	1	22.2/26
Power Law	72.46	0.17	1	25.9/26

fit, values of the shape parameters are estimated by fitting to the background MC, then allowed to float within the likelihood of the template fit. A signal bias test is performed through injecting pseudodata, performing signal+background fits, and attempting to recover the signal. A bias is found for resonances below 400 GeV, so a correction is applied to resolve this. The correction is m_X dependent and less than ± 0.05 pb everywhere, and a ± 0.02 pb uncertainty in this correction is applied to the extracted cross-section.

5.6 Systematic Uncertainties

As a result of the small HH cross-section, the sensitivity of this analysis is dominated by statistical uncertainties. The following sections will discuss systematic uncertainties, both theoretical (Section 5.6.1) and experimental (Section 5.6.2).

5.6.1 Theoretical Uncertainties

This section describes theoretical uncertainties on the mono-Higgs and di-Higgs samples. In each of these, alternate models of parton showering and hadronization were considered and differences were found to be negligible.

Mono-Higgs: Theory uncertainties for the mono-Higgs boson production cross-section are estimated through varying renormalization and factorization scales. Values are taken from the CERN Yellow Report 4 [99]. The uncertainty on the WH and ZH scales are fully correlated, and the PDFs between gluon-gluon (ggH, ttH) and quark-quark (VBF H, WH, ZH) production modes are also fully correlated. The scale uncertainties reach $^{+20\%}_{-24\%}$ for bbH, and the remaining production modes reach $^{+5.8\%}_{-9.2\%}$. Uncertainties due to the PDF and running of the QCD coupling constant (α_S) are also considered, reaching $\pm 3.6\%$. These values are shown in Table 5.8. Production associated with heavy-flavor jets is considered as follows:

- ggH, WH, VBF H: A 100% uncertainty is assigned, motivated by previous studies of heavy-flavor production with top quark pairs (ggH) [109] and W boson production in association with b-jets (WH, VBF H) [110].
- ZH and $t\bar{t}H$: No heavy-flavor uncertainty assigned. The dominant contribution is accounted for in the LO process.

A $H \to \gamma\gamma$ BR uncertainty of $^{+2.9\%}_{-2.8\%}$ and a $H \to b\bar{b}$ BR uncertainty of $\pm 1.7\%$ [18] are taken into account as well.

SM Di-Higgs: The SM HH signal samples are impacted by the same theory uncertainties as the mono-Higgs, as described above. For the NNLO cross-section, the scale effects are 4-8%, and the PDF + α_S uncertainty is 2-3%. An additional 5% uncertainty is added from the infinite top-quark mass approximation [111].

Table 5.8: SM single Higgs scale and PDF percent uncertainties are presented, by production mode. Values are taken from the CERN Yellow Report 4 [99]. The uncertainty on the WH and ZH scales are fully correlated. The PDFs between gluon-gluon (ggH, ttH) and quark-quark (VBF H, WH, ZH) production modes are also fully correlated. An additional uncertainty of +2.9/-2.84% applies to the $H \rightarrow \gamma\gamma$ branching ratio.

	Scale		PDF+α_S	
Prod. Mode	Value Up	Value Down	Value Up	Value Down
ggH	+4.6%	-6.7%	+3.2%	-3.2%
VBF H	+0.4%	-0.3%	+2.1%	-2.1%
WH	+0.5%	-0.7%	+1.9%	-1.9%
ZH	+3.8%	-3.0%	+1.6%	-1.6%
ttH	+5.8%	-9.2%	+3.6%	-3.6%
bbH	+20.1%	-23.9%	(included at left)	

Resonant Di-Higgs: For the resonant analysis, scale and PDF uncertainties are neglected. The SM non-resonant HH production is a background with uncertainty $^{+7\%}_{-8\%}$, and interference between the resonant HH signal and non-resonant HH production is neglected.

5.6.2 Experimental Uncertainties

Luminosity: Beam-separation scans in 2015 and 2016 are used to derive the systematic uncertainty in the integrated luminosity of 2.1%. This methodology is outlined in Reference [112].

Simulation: Calibration uncertainties apply to photons and jets used in the analysis, for all samples except the $\gamma\gamma$-continuum, which is estimated from data. The uncertainties are propagated through the analysis chain, and relevant observables are constructed. Then changes in peak location and width (m_{peak} and σ_{peak}, respectively), as well as signal yield in $m_{\gamma\gamma}$ (non-resonant) or $m_{\gamma\gamma b\bar{b}}$ (resonant) are extracted. In the resonant analysis, uncertainties are evaluated at each mass hypothesis, and the maximum value is taken.

Yield: Photon identification and isolation affect the diphoton selection efficiency. Uncertainties in peak location are primarily due to photon energy scale, and are about 0.2-0.6% for both mono- and di-Higgs samples. Uncertainties in width are primarily due to photon energy resolution, and are 5-14%.

Jets: Uncertainties in Jet Energy Scale (JES) and Jet Energy Resolution (JER) affect the $m_{b\bar{b}}$ acceptance [113]. Uncertainties in flavor tagging can miscategorization of events in the tagging categories. A summary of the most dominant systematic uncertainties can be found in Table 5.10.

Signal Model: The spurious signal, described in Section 5.5.2.2, is used as an uncertainty on the number of signal events. The values for each analysis category can be found in Table 5.9.

Table 5.9: Uncertainty on total number of signal events in each analysis category, as derived from the spurious signal test.

Analysis	Category	Uncertainty on Number of Events
Non-resonant	1-tag	0.25
	2-tag	0.63
Resonant	Loose	0.21
	Tight	0.89

Cross-Section: In the resonant analysis, an m_X correction is applied to the signal cross-section with a ± 0.02 pb uncertainty, described in Section 5.5.3.2.

Table 5.10: Summary of dominant systematic uncertainties affecting expected yields in the resonant and non-resonant analyses. For the non-resonant analysis, uncertainties in the Higgs boson pair signal and SM single-Higgs-boson backgrounds are presented. For the resonant analysis, uncertainties on the Higgs boson pair signal for the loose and tight selections are presented. These are evaluated at every mass hypothesis and the maximum is taken, to be conservative (described in Section 5.7.1). Sources marked '-' and other sources not listed in the table are negligible by comparison. No systematic uncertainties related to the continuum background are considered, since this is derived through a fit to the observed data.

Source of systematic uncertainty		% effect relative to nominal in the 2-tag (1-tag) category: Non-resonant analysis		Resonant analysis: BSM HH	
		SM HH signal	Single-H bkg	Loose selection	Tight selection
Luminosity		±2.1 (±2.1)	±2.1 (± 2.1)	±2.1 (±2.1)	±2.1 (±2.1)
Trigger		±0.4 (±0.4)	±0.4 (± 0.4)	±0.4 (±0.4)	±0.4 (±0.4)
Pile-up modelling		±3.2 (±1.3)	±2.0 (± 0.8)	±4.0 (±4.2)	±4.0 (±3.8)
Photon	identification	±2.5 (±2.4)	±1.7 (± 1.8)	±2.6 (±2.6)	±2.5 (±2.5)
	isolation	±0.8 (±0.8)	±0.8 (± 0.8)	±0.8 (±0.8)	±0.9 (±0.9)
	energy resolution	-	-	±1.0 (±1.3)	±1.8 (±1.2)
	energy scale	-	-	±0.9 (±3.0)	±0.9 (±2.4)
Jet	energy resolution	±1.5 (±2.2)	±2.9 (± 6.4)	±7.5 (±8.5)	±6.4 (±6.4)
	energy scale	±2.9 (±2.7)	±7.8 (± 5.6)	±3.0 (±3.3)	±2.3 (±3.4)
Flavor tagging	b-jets	±2.4 (±2.5)	±2.3 (± 1.4)	±3.4 (±2.6)	±2.5 (±2.6)
	c-jets	±0.1 (±1.0)	±1.8 (±11.6)	-	-
	light-jets	<0.1 (±5.0)	±1.6 (± 2.2)	-	-
Theory	PDF+α_S	±2.3 (±2.3)	±3.1 (± 3.3)	n/a	n/a
	Scale	+4.3 (+4.3)	+4.9 (+ 5.3)	n/a	n/a
		−6.0 (−6.0)	+7.0 (+ 8.0)	n/a	n/a
	EFT	±5.0 (±5.0)	n/a	n/a	n/a

5.7 Results

5.7.1 Statistical Treatment

The test statistic used in the statistical analysis is the profile likelihood ratio

$$\tilde{q}_{\mu} = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta}(\mu))} & 0 \leq \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \tag{5.7}$$

where θ is the nuisance parameters and μ is the parameter of interest (POI). This analysis is statistics limited, so toy experiments are generated following the frequentist recommendation. Global observables are randomized around the corresponding nuisance parameter values fitted on the data on the data for the given signal hypothesis (μ), except spurious signal, which is randomized around 0. The statistical framework is based on `RooFit` [114].

The non-resonant analysis performs a unbinned fit to the $m_{\gamma\gamma}$ spectrum simultaneously in the 1 and 2 b-tag regions using the fit functions described in Section 5.5.2. The high-mass selection is used for setting limits on the di-Higgs cross-section, the low mass is used for setting limits on κ_{λ}. The POI is the cross-section for non-resonant HH production in picobarns, constrained to lie between 0 and 100.

The resonant analysis performs a signal + background fit to $m_{\gamma\gamma b\bar{b}}$ simultaneously in the 1 and 2 b-tag regions, using the fit functions described in Section 5.5.3. No excesses are observed so limits are set. Fitting is performed independently in the 1 and 2 b-tag categories; no correlation is assumed between fit parameters.

5.7.2 Limits on Non-Resonant Production

The tight selection is used in the non-resonant analysis to set a 95% CL upper limit on non-resonant ggF Higgs boson pair production. The data and background-only fit on the $m_{\gamma\gamma}$ spectrum is shown in Figure 5.18. The observed and expected limits are shown in Table 5.11 in both absolute value and as a multiple of the SM production cross-section. The observed $\pm 1\sigma$ values are shown as well. This upper limit is shown in Figure 5.19.

Table 5.11: Non-resonant limits in terms of the Standard Model HH cross-section.

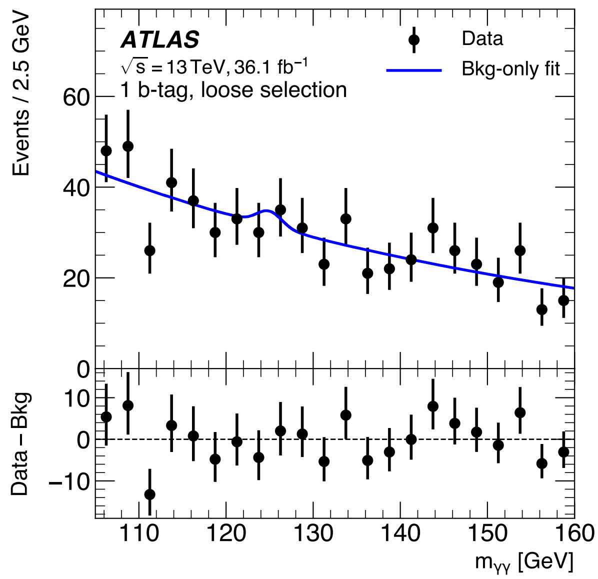
	Observed	Expected	-1σ	$+1\sigma$
$\sigma_{gg\to HH}$ [pb]	0.73	0.93	0.66	1.3
Multiple of σ_{SM}	22	28	20	40

5.7.3 Limits on Resonant Production

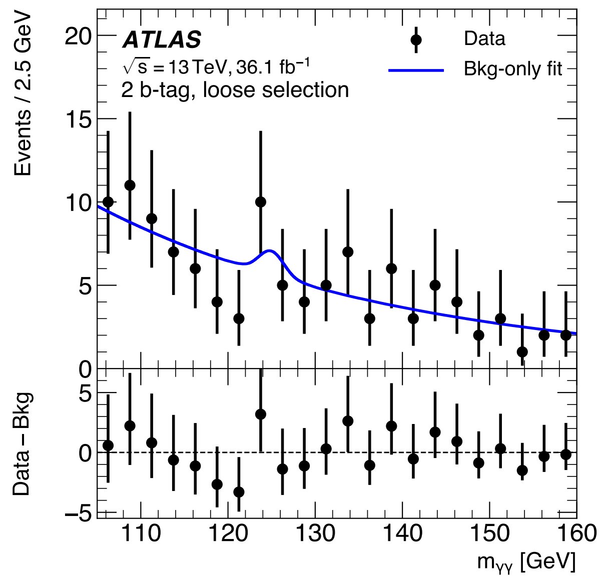
The 95% CL limits on resonant di-Higgs production are shown in Figure 5.20. This limit uses the loose selection at $m_X \leq 500$ GeV and the tight selection at $m_X \geq 500$ GeV. No significance excesses are found. Expected exclusion limits range from 1.14 pb to 0.12 pb, observed limits range from 0.90 pb to 0.15 pb.

5.7.4 Limits on the Trilinear Higgs Coupling

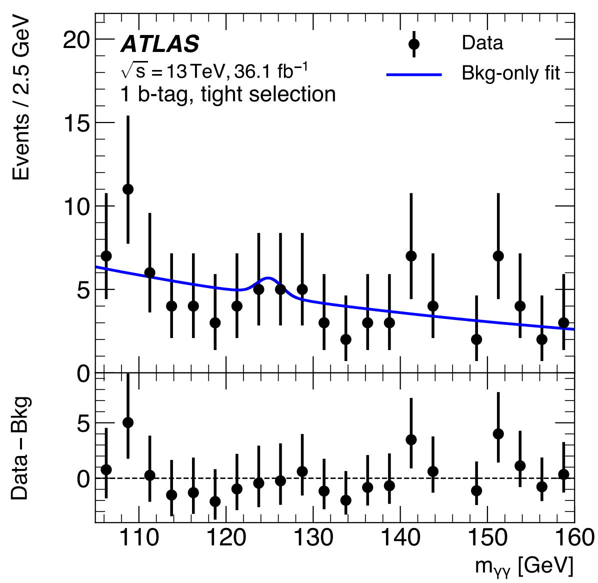
Limits are set on κ_λ using the asymptotic formula. The exclusion as a function of κ_λ is shown in Figure 5.21, corresponding to $-8.2 < \kappa_\lambda < 13.3$.



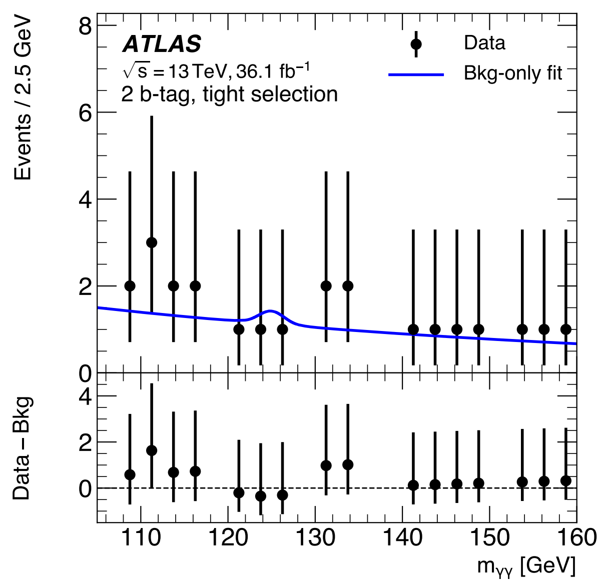
(a)



(b)



(c)



(d)

Figure 5.18: Data and background-only fits for each of the analysis categories in the non-resonant analysis. The background fit is composed of the smoothly-falling continuum $\gamma\gamma$ background as well as the single Higgs boson production background, peaking around 125 GeV.

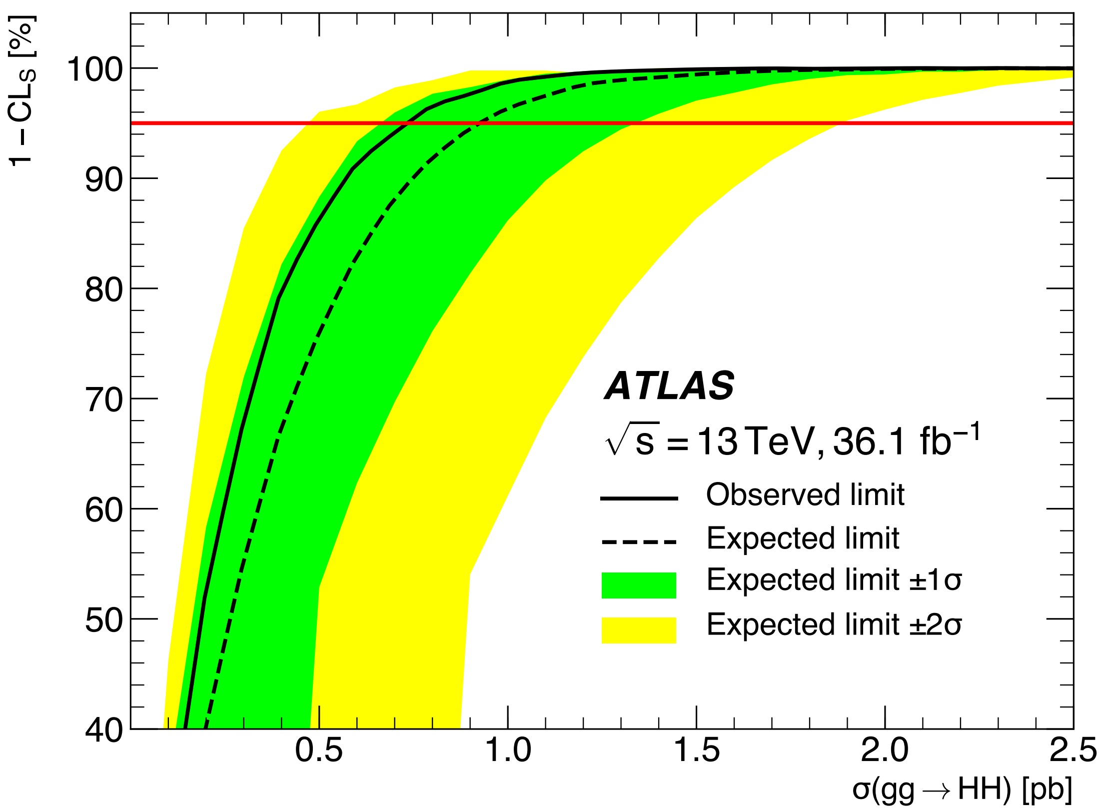


Figure 5.19: The expected and observed 95% CL limits on the non-resonant HH production cross-section, $\sigma_{gg\to HH}$. These limits are set using the tight analysis selection. The 95% confidence level is indicated by the red horizontal line.

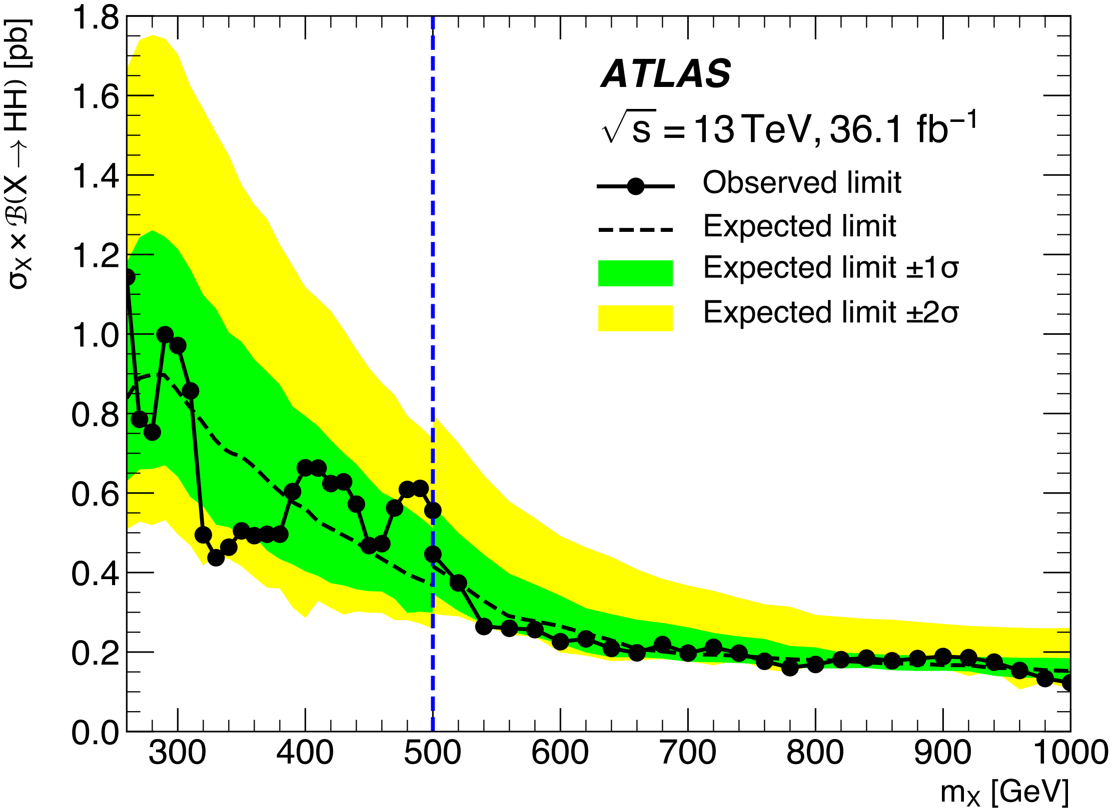


Figure 5.20: The expected and observed 95% CL limits on the resonant HH production cross-section, $\sigma_X \times \mathcal{B}(X \rightarrow HH)$, as a function of m_X. The loose selection is used in the low mass regime, $m_X \leq 500$ GeV, while the tight selection is used in the high mass regime, $m_X \geq 500$ GeV. The transition point between these selections is indicated by the dashed line.

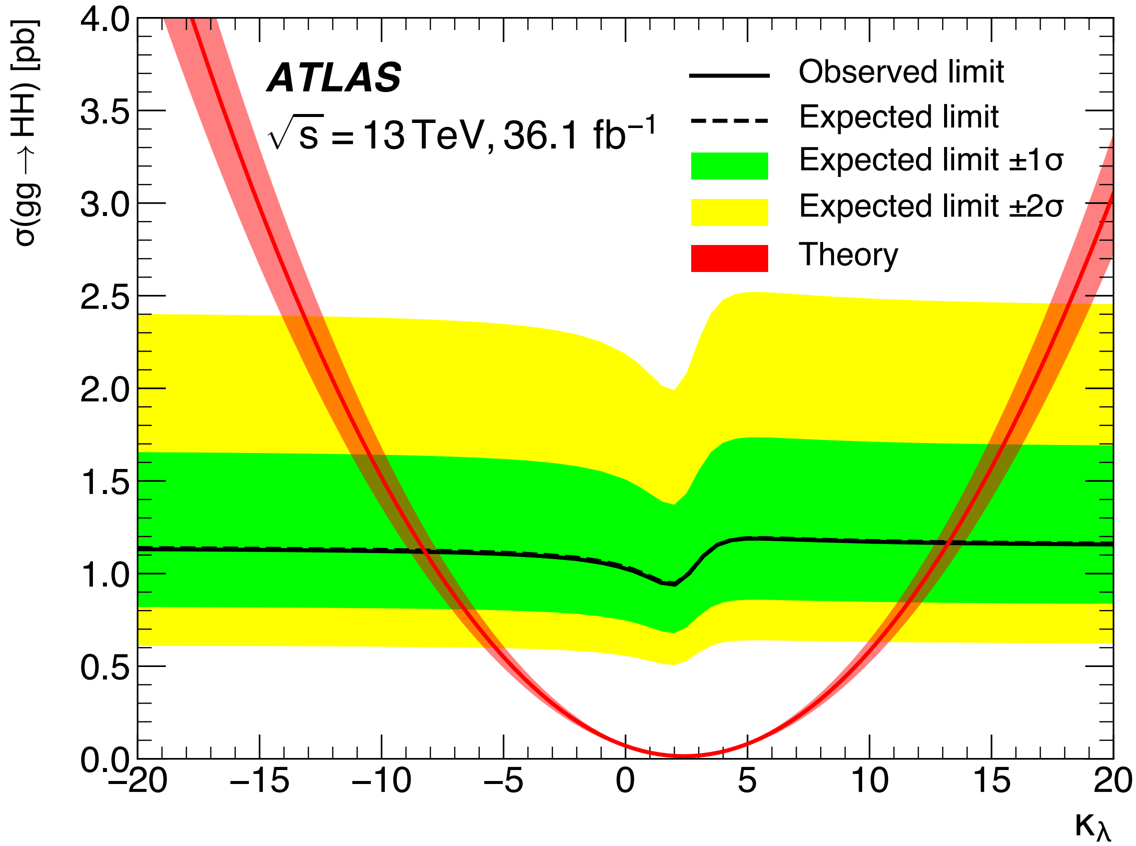


Figure 5.21: The expected and observed 95% CL limits on the non-resonant HH production cross-section as a function of κ_λ. The red line indicated the predicted HH cross-section for varied κ_λ, assuming all other couplings are equivalent to SM prediction, the band includes theoretical uncertainty.

CHAPTER 6

FUTURE IMPROVEMENTS - VECTOR BOSON FUSION HH PRODUCTION

Do not be too timid and squeamish about your actions. All life is an experiment. The more experiments you make the better.
— Ralph Waldo Emerson

To improve this analysis, a signal model for the Vector Boson Fusion production mode has been designed and will be implemented into this analysis using the full Run 2 dataset. This signal model is integrated into the analysis through a separate analysis category that is enriched in VBF events, defined via cuts on multiclass BDT scores.

This chapter outlines the approach to defining the signal selection for the VBF production mode. Section 6.1 describes the MC simulation produced to model VBF $HH \to \gamma\gamma b\bar{b}$ production. Section 6.2 outlines improvements in signal selection for the analysis, starting with a BDT used to define analysis categories focused on ggF HH production (Section 6.2.1), followed by a separate multiclass BDT used to define a VBF-enriched signal category (Section 6.2.2). Adding such a category takes advantage of the unique VBF topology to improve the overall Asimov significance, and also provides sensitivity to additional couplings, most notably the c_{2V} coupling corresponding to the interaction between the two vector bosons and two Higgs bosons, which can lead to significant increases in the HH cross–section, described in Section 1.3.1.2.

6.1 Simulated Samples

6.1.1 Non-resonant VBF HH Samples

The signal MC samples[1] for VBF HH production are generated at LO using MadGraph5_aMC@NLO 2.6.0, using cards presented in [25]. The process generated is $pp \to HHqq$. The dominant production in this sample is VBF HH production, but contains contributions from VHH hadronic and Higgsstrahlung production. While these additional production modes are not VBF HH production, they do not overlap with the ggF HH sample, since the generated process differs. The generated VBF sample requires pp interactions producing two Higgs bosons and two additional jets, where the ggF sample requires the production of two Higgs bosons without additional jets. The `NNPDF 2.3 LO` PDF set [115] is used in the matrix element, interfaced to Herwig 7.0.4 using the `H7-UE-MMHT` tune for underlying events and the `H7-MMHT2014LO` tune for parton shower and hadronization.

6.1.2 Non-resonant ggF HH Samples

The ggF HH sample produced in these studies matches the non-resonant SM signal sample described in Section 5.2.2.1, however is produced for the full Run 2 dataset, with independent samples produced to match running conditions in the 2017 and 2018 data taking periods in addition to the 2015+2016 sample.

[1] Samples with varied coupling strengths were also produced, described in Appendix F.

6.1.3 Background Samples

The same $\gamma\gamma$-continuum processes described in Section 5.2.2.2 are the predominant background in these studies. Dominant mono-Higgs backgrounds described in Section 5.2.2.2 are considered as well, ttH, ZH, and ggH. Both of these have been produced to match the conditions of the full Run 2 data taking period.

Additionally, Not-Tight, Not-Isolated (NTNI) data, which fails photon identification, isolation, or both, is considered to model γj, $j\gamma$, and jj contributions. The yield is proportionally scaled to the template fits, described in Section 5.5.1. Last, a sample of top-quark pair production associated with 2 photons, $tt\gamma\gamma$, is used. This sample is produced at LO using MadGraph, showered with Pythia8 using the `A14` tune and the `NNPDF23LO` PDF set.

6.2 Event Selection

To improve the event selection presented in Section 5.4 events are evaluated by a BDT trained to target HH production, described in Section 6.2.1. Then the VBF enriched signal region is constructed, described in Section 6.2.2. Events considered have the following preselection cuts are applied:

- $N_{\text{cen jets}} < 6$, to help reject ttH production (hadronic top decays).
- $N_{\text{lep}} = 0$, to help reject ttH production (leptonic top decays).
- $N_{85\ \%\ \text{b-tags}} >= 2$.
- $N_{70\ \%\ \text{b-tags}} < 3$, to maintain orthogonality with the $HH \to b\bar{b}b\bar{b}$ analysis.

6.2.1 BDT for ggF Production

A BDT has been trained in order to maximize signal efficiency for ggF HH production[2]. Following preselection, events are segmented into two regions, a high mass region with $m_{X^*} > 350$ GeV targeting the standard model signal, and a low mass region with $m_{X^*} < 350$ GeV targeting BSM signals.

In each region, a separate BDT is trained using a benchmark HH signal against a combination of $\gamma\gamma$, ttH, ggH, and ZH backgrounds. In the high mass region, the standard model HH sample is used as signal, while in the low mass region, the $\kappa_\lambda = 6$ sample is used as signal.

The following inputs are used:

- The $p_T/m_{\gamma\gamma}$, η, ϕ of the two photons.
- The p_T, η, ϕ, and pseudo-continuous b-tagging score of the first two jets.
- The MET and the ϕ angle of the MET.
- The p_T, η, ϕ, and mass of the $H \rightarrow b\bar{b}$ candidate.
- The H_T (scalar p_T sum of all jets).

where,

- The p_T of the two photons are scaled by $m_{\gamma\gamma}$ to prevent the BDT from learning the $m_{\gamma\gamma}$ variable. There is no equivalent scaling done for the two jets, since fitting is performed in $m_{\gamma\gamma}$, not $m_{b\bar{b}}$.

[2]This BDT has been developed within the full $HH \rightarrow \gamma\gamma b\bar{b}$ analysis group. While contributions were made to early ggF BDTs, the focus of the work presented is on the VBF selection. The ggF BDT is constructed to target di-Higgs production, and is used as a prerequisite for selection into the VBF category. Further, the signal categories defined with this BDT are used as a benchmark to evaluate improvement in significance, so a description is included.

- Jets are ordered by pseudo-continuous b-tagging score, and then by p_T in case of a tie. The $H \to b\bar{b}$ candidate is then formed by the two highest ranked jets by simply summing their four-vectors.

- All vectors are rotated by the same ϕ angle so that the ϕ of the leading photon is equal to zero. This effectively removes one degree of freedom for the BDT to learn.

After training, two categories are created in each mass region by creating cuts on the BDT output. The category boundaries are optimized to give the best Asimov number counting significance [116], given by

$$Z = \sqrt{2[(s+b)\log(1+s/b) - s]}. \tag{6.1}$$

The signal, s, is the sum of both ggF and VBF HH contributions in a category. The background, b, is the sum of the considered backgrounds described in Section 6.1.3: $\gamma\gamma$-continuum processes, NTNI data to model jets faking photons, leading mono-Higgs production modes (ttH, ZH, and ggH), and $tt\gamma\gamma$. The four signal categories (two mass regions, each with two categories) are added in quadrature for the final Asimov significance of 0.415 using the full 140 fb^{-1} collected in Run 2.

6.2.2 BDT for VBF Production

On top of the ggF BDT selection, a dedicated VBF-enriched analysis category is built. Events which have already passed the ggF BDT based selection and have at least four jets are considered the VBF category. The four jet requirement is imposed in order to reconstruct VBF-based variables, since two jets are needed for the $H \to b\bar{b}$ decay, and another two are needed for the VBF jets. The VBF jets are then selected as the highest m_{jj} pairing, excluding $H \to b\bar{b}$ candidate jets.

These events are then evaluated by a multiclass BDT, which has independent classes for VBF HH, ggF HH, $\gamma\gamma$-continuum, and ttH. This model was chosen due to large ttH contamination in the VBF enriched category by binary classifiers considering just $\gamma\gamma$-continuum backgrounds. Ultimately, an event selected for the VBF category if it passes the BDT score threshold for VBF HH, and fails the score threshold for the other three classes. A flowchart of this logic is shown in Figure 6.1.

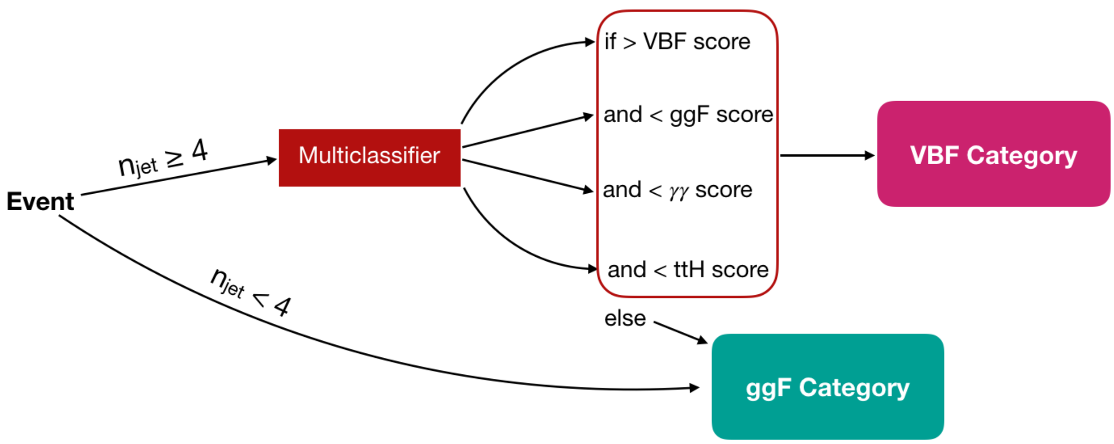


Figure 6.1: Logic flow for an event to be selected into the VBF-enriched signal region. The starting set of candidates (denoted "event") requires passing preselection and at least one of the ggF-based signal categories. If an event is assigned to the ggF categories, it is separated into 4 categories by the BDT described in Section 6.2.1.

6.2.2.1 Input Selection

A set of 26 variables were considered for this BDT, listed in Appendix E. Broadly, they describe the kinematics of various physics objects in the event, as well as event shape variables [117] and variables to reconstruct the W and top mass, which aim to suppress the ttH background.

For dimensionality reduction, a pruning procedure was performed.

1. To remove redundant variables, those with a Pearson correlation value above 0.85 were removed. This brought the list from 26 to 19, pruning several event shape variables.

2. After pruning those variables and retraining, low-impact variables were removed. The Pearson correlation for each variable to the 4 class BDT probability distributions were calculated. Those 4 correlation values were summed, and variables with a sum less than 1.0 were pruned. This brought the set from 19 to 10 input variables.

After this pruning procedure, the 10 final inputs were:

- **VBF-targeted:** $\Delta\eta_{jj}$, m_{jj}.
- **$\gamma\gamma$-suppressing:** $m_{\gamma\gamma b\bar{b}}$, $\Delta R_{\gamma\gamma}$, $\Delta R_{b\bar{b}}$, the best b-tagging WP for the $H \rightarrow b\bar{b}$ jets.
- **ttH-suppressing:** Transverse Sphericity ($S_\perp$), Planar Flow (Pf), $p_{\mathrm{T}}^{\mathrm{balance}}$.

$S_\perp$ is given in Reference [117], Planar Flow is given in Reference [118], and $p_{\mathrm{T}}^{\mathrm{balance}}$ is given by:

$$p_{\mathrm{T}}^{\mathrm{balance}} = \frac{|\vec{p}_T^{\gamma_1} + \vec{p}_T^{\gamma_2} + \vec{p}_T^{j_1} + \vec{p}_T^{j_2}|}{|\vec{p}_T^{\gamma_1}| + |\vec{p}_T^{\gamma_2}| + |\vec{p}_T^{j_1}| + |\vec{p}_T^{j_2}|}. \tag{6.2}$$

Distributions for each these variables after applying a 4-jet requirement can be found in Appendix E.

6.2.2.2 Model Optimization

This section outlines two optimizations performed in the definition of the VBF-enriched category. First, the optimization of the model's main hyperparameters. Second, the selection of cuts on the BDT scores to define the VBF signal category. For these optimizations, an orthogonal sample of events must be selected to ensure the model is generalizable. A subset of 50% of the data is used for training, 25% for testing, and 25% for model selection[3].

Hyperparameter Optimization

The maximum BDT depth is selected by training models at increasing depth and selecting the depth when the accuracy on the testing set ceases to improve, a depth of 14. After that point, the accuracy on the training set flattens, while it continues to increase on the testing set, leading to overtraining. The same procedure is performed for the number of trees, evaluating 10, 25, 50, 75, 100, 150, 200, and 300 trees. This parameter flattens for both training and testing, so a safe choice of 200 trees is used. The accuracy as a function of these parameters is shown in Figure 6.2.

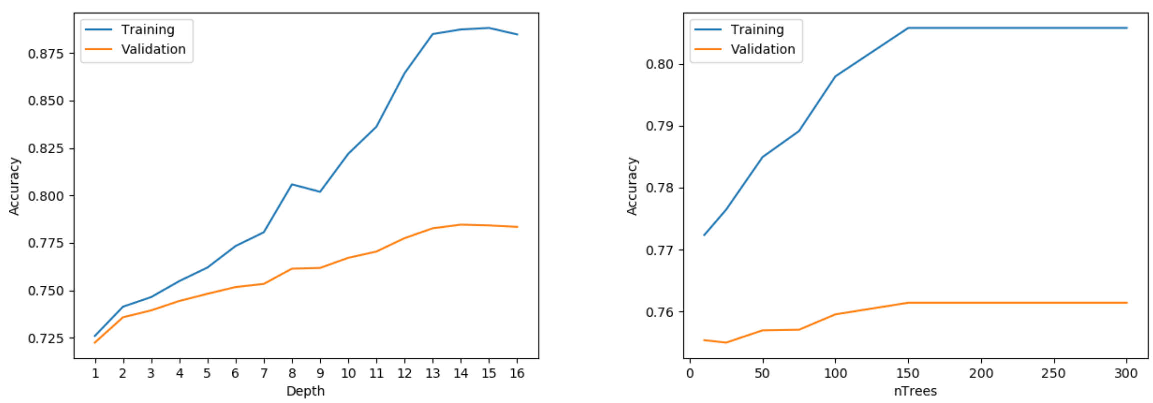


Figure 6.2: The BDT accuracy as a function of hyperparameter values. The left is the maximum tree depth, which is chosen to be 14. The right shows the number of trees, which is chosen to be 200. The "validation" set is the events used for testing, and orthogonal to the training set.

BDT Score Cuts

Evaluating an event on this multiclass BDT outputs four probabilities, one for each

[3]This split is done using the ATLAS global event number. A modulo 4 operation is performed on the event number, if the result is 0 or 1, it is used in the training sample. If the result is 2, it is used in the testing sample. If the result is 3, it is used in the model selection sample.

class. These are denoted p_{class} for each target class, e.g. p_{VBF} for the class targeting VBF HH production. In order to set the thresholds for each of these probabilities, an optimization procedure is performed. A 4-dimensional scan over the probability scores is performed, in steps of 0.03. At each step, the Asimov significance, given in Equation 6.1, is calculated for each category and summed in quadrature. The contributions to the signal and background components of this equation match that used to evaluate the Asimov significance in Section 6.2.1.

Through this optimization procedure, the thresholds shown in Table 6.1 are found. The output probability distributions for each class on each sample is shown in Figure 6.3, with the optimized thresholds overlaid.

Table 6.1: The optimized thresholds for each BDT score used to define the VBF-enriched category. The cut threshold of 0.00 on the VBF HH class means that this class is not cut on.

Class	Target	Selection Logic
0	VBF HH	$p_{\text{VBF}} > 0.00$
1	ggF HH	$p_{\text{ggF}} < 0.72$
2	$\gamma\gamma$-Continuum	$p_{\gamma\gamma} < 0.06$
3	ttH	$p_{ttH} < 0.81$

The contribution of each signal and background process to each category is shown in Figure 6.4. The categorization for each signal process as well as the dominant $\gamma\gamma$-continuum background is shown in Figure 6.5. The VBF HH category is dominated by ggF production due to the large difference in cross-sections, however the VBF HH sample is predominantly classified into the VBF-targeted signal region. Adding these five categories in quadrature gives an Asimov significance of 0.459, a 9.7% improvement over not defining such a category.

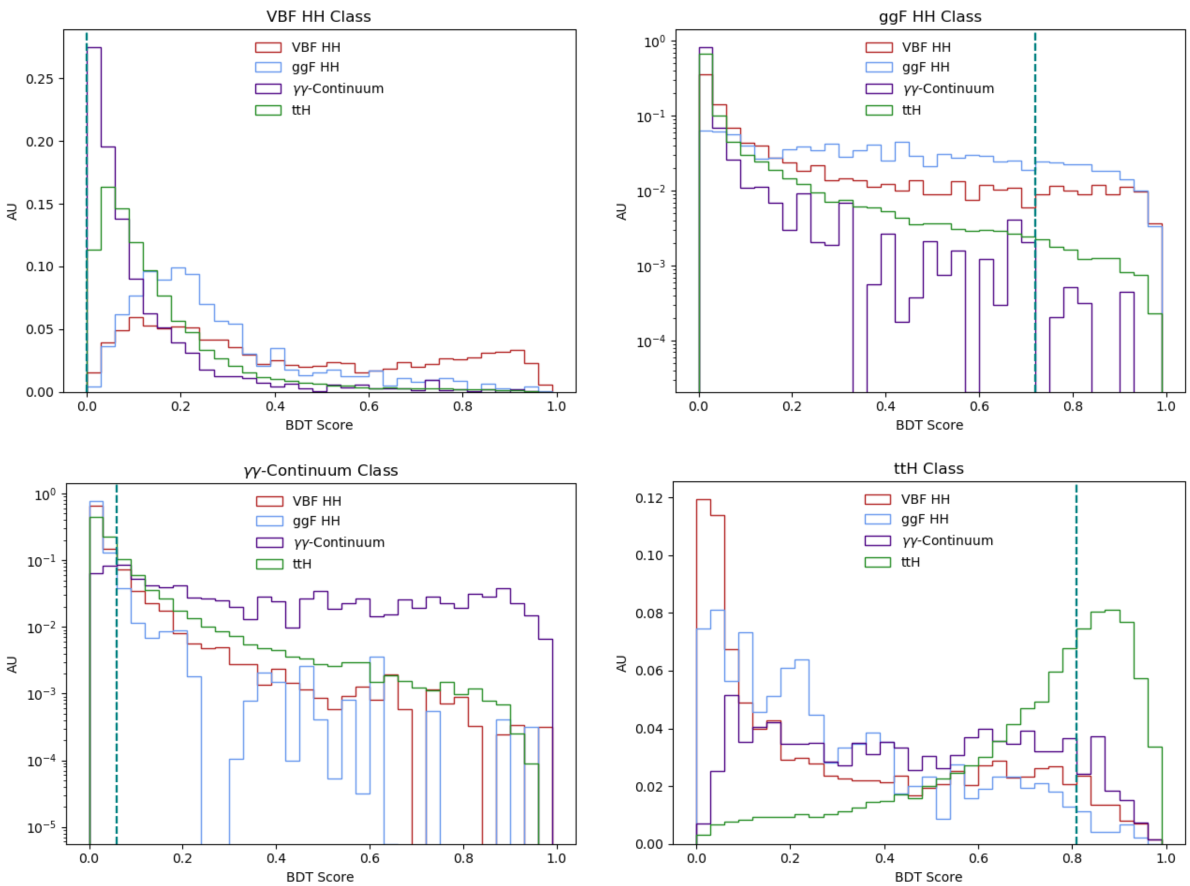


Figure 6.3: The BDT score distribution for each class, evaluated on the testing VBF HH, ggF HH, $\gamma\gamma$-continuum, and ttH samples. The ggF HH and $\gamma\gamma$-continuum distributions are shown on a log scale. A vertical line showing the cut threshold is shown on each distribution.

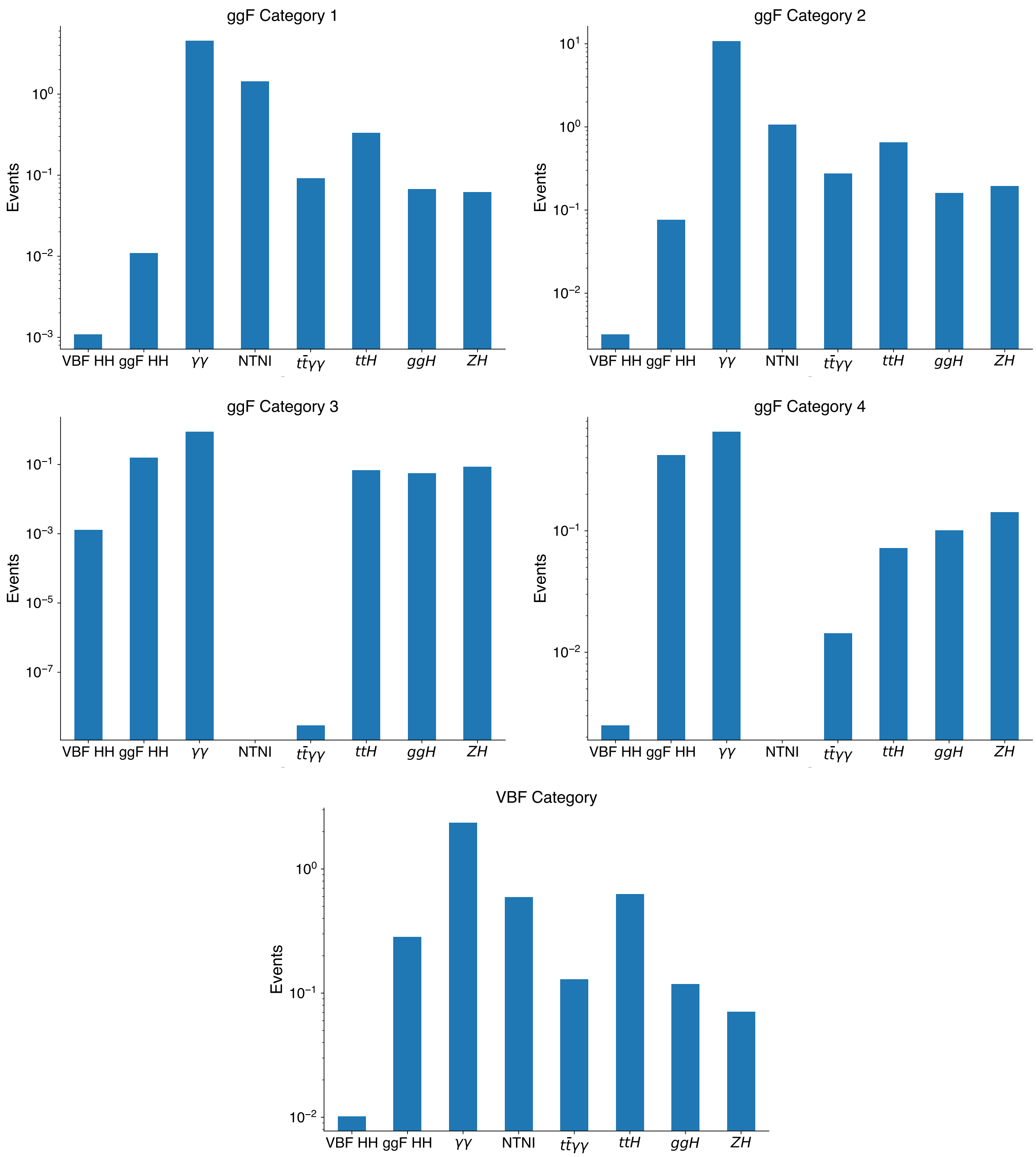


Figure 6.4: The contribution of signal and leading background samples to each analysis category after adding the optimized VBF-enriched category. VBF and ggF HH are the two signal production modes considered. $\gamma\gamma$ is the diphoton continuum background, $t\bar{t}\gamma\gamma$ is top-quark pair production associated with 2 photons, and NTNI is not-tight not-isolated data (scaled using a template fit). The three dominant mono-Higgs backgrounds, ttH, ggH, and ZH are shown. The ggF categories correspond to (1) $m_{X^*} < 350$ GeV, loose b-tag (2) $m_{X^*} < 350$ GeV, tight b-tag (3) $m_{X^*} \geq 350$ GeV, loose b-tag (4) $m_{X^*} \geq 350$ GeV, tight b-tag.

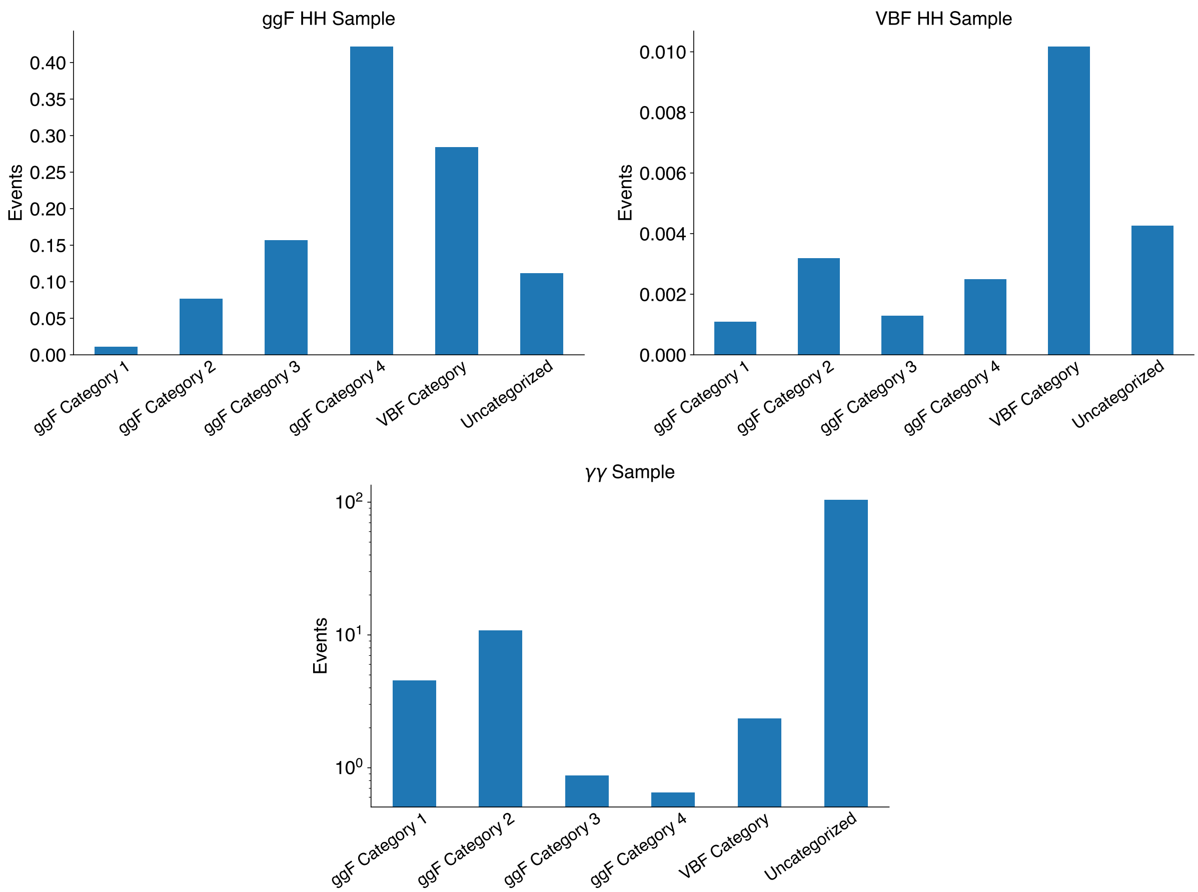


Figure 6.5: The categorization of the VBF HH, ggF HH, and $\gamma\gamma$-continuum samples. The "Uncategorized" bin are events which do not fall into any of the analysis categories. The $\gamma\gamma$-continuum plot is shown with log-scale. The ggF categories correspond to (1) $m_{X^*} < 350$ GeV, loose b-tag (2) $m_{X^*} < 350$ GeV, tight b-tag (3) $m_{X^*} \geq 350$ GeV, loose b-tag (4) $m_{X^*} \geq 350$ GeV, tight b-tag.

CHAPTER 7

CONCLUSION

One never notices what has been done; one can only see what remains to be done.
— Marie Curie

A long-term goal of the LHC physics program is understanding Electroweak Symmetry Breaking. Measurement of the Higgs trilinear coupling probes the nature of the Higgs potential and provides a precision test of electroweak theory. Studying di-Higgs production provides a direct handle on the Higgs trilinear coupling, and the $\gamma\gamma b\bar{b}$ decay channel is an interesting channel due to the high $H \to b\bar{b}$ branching ratio as well as the excellent $H \to \gamma\gamma$ mass resolution and trigger efficiency.

A search for resonant and non-resonant di-Higgs production in the $\gamma\gamma b\bar{b}$ final state has been presented. This search was performed using 36.1 fb^{-1} of $\sqrt{s} = 13$ TeV data from pp collisions collected by the ATLAS detector in 2015-2016, the first two years of Run 2. The non-resonant search set limits on the $HH \to \gamma\gamma b\bar{b}$ cross-section times branching ratio, with an upper observed (expected) limit of 0.73 (0.93) pb. The observed (expected) 95% CL constraint on the Higgs boson trilinear coupling is set between $-8.2 < \kappa_\lambda < 13.3$ ($-8.5 < \kappa_\lambda < 13.7$). A model-independent resonant search has been presented, setting limits on a generic scalar resonance between 260 and 1000 GeV under the narrow-width approximation. These observed (expected) limits range from 0.85 (0.92) pb at the lowest mass hypothesis to 0.13 (0.15) pb to the highest mass hypothesis.

This analysis is sensitive to the square of the photon identification efficiency, and in the future of this search, improvements to photon identification can allow a loosening of the signal selection, which can help the overall significance. By using topological cluster moments as inputs into the current cut-based model, an improvement of as much as 12% in background rejection for the same signal efficiency is found. Replacing the current cut-based model with a p_{T}-inclusively trained BDT leads to an improvement of 22% in background rejection. By combining both optimization strategies, an improvement of 27% is shown.

Direct improvements to the analysis have also been presented, studying the VBF HH production mode. This provides sensitivity to different couplings compared to prior ggF-only searches, notably the interaction between two Higgs bosons and two vector bosons. Additionally, constraints on di-Higgs production can be improved by incorporating a signal region targeting VBF production. This signal region is defined through the use of a multiclass BDT with a dedicated class to differentiate the production mode from ggF HH production, as well as classes to differentiate from the dominant $\gamma\gamma$-continuum background, and ttH mono-Higgs background. Through defining this VBF enriched signal region, a 9.7% improvement in Asimov significance is achieved.

Appendices

APPENDIX A

MULTIVARIATE ANALYSIS AND MACHINE LEARNING

It's tough to make predictions, especially about the future.
— Danish Proverb

Machine Learning (ML) can be separated into two broad categories: supervised and unsupervised learning. Supervised learning aims to learn a mapping from inputs to outputs, utilizing labeled data for training. This can be employed in both classification and regression problems. Unsupervised learning works from unlabeled data in order to infer information and patterns about the dataset. The ML methods used in this thesis all fall under the former label. Through the simulation methods (outlined in Section 3.1), the physics process present in any given MC sample are known, and this truth information is leveraged as the "label" of the data, making the problem suitable for supervised ML.

More specifically, the tasks presented in this thesis fall under the problem known as classification, where a model is trained to map inputs to outputs $y \in \{1, \ldots, C\}$ for C classes. Commonly $C = 2$, known as binary classification, trying to isolate signal from background.

The "learning" process in ML is an updating of internal weights and biases in order to minimize of some *loss function*. Table A.1 lists loss functions typically used with various tasks commonly approached by ML methods (this list is certainly not exhaustive).

Table A.1: A list of common tasks in which ML is employed, and loss functions typically minimized in setting biases and weights for ML models.

Task	Common Loss Function
Binary Classification	Binary Cross-Entropy
Multiclass Classification	Categorical Cross-Entropy
Regression	Mean Squared Error

A.1 General Principles and Terminology

This section will overview some general concepts relevant to ML methods and referenced within this thesis.

Hyperparameters

When developing ML algorithms, tuning is generally done by manipulating the "hyperparameters" of the algorithm. Algorithms are defined as a generic set of rules (functions, cuts, weights, etc.) which adapt as they "learn" from data. The properties that define the parameters, or rules, of the algorithm itself (its topology, learning rate, etc.) are known as hyperparameters. Hyperparameters are fixed before the learning process begins. For most hyperparameters, there is no a priori assumption of the best values, and optimization through successive model trainings with differing values is required.

Training and Testing Sets

A major concern when developing a fitting model is known as *overfitting*, in which a model is overconstrained to model the particular data it is trained on, and thus not generalizable and not useful for inference. To avoid overfitting, before training data is often split into subsamples. The training sample is used to fit the model, setting the weights and biases. Once fit, the model is applied to the testing set to evaluate performance[1]. Significant differences in performance between the training and testing sample is an indication of overfitting, while a similar performance suggests a model is generalizable to new data.

A.2 Algorithms

While there are many ML algorithms, in the work presented in this thesis, two are used: Boosted Decision Trees and Neural Networks. These are outlined in Sections A.2.2 and A.2.2. As a contrast, a more simple univariate approach to classification, rectangular cuts, is often used and is checked as a baseline. This method is described in Section A.2.1.

[1] In some cases, a third set of data known as a "validation" set is used. In this case, the training data is used for initial fitting, the validation set is used for hyperparameter optimization, and the test set is used to evaluate performance of the final model.

A.2.1 Rectangular Cuts

Although not ML, one classical univariate method for signal selection is known as rectangular cuts. In this, a feature of the data set is cut at a single point, and values either above or below that are taken as the signal. Performing this procedure along many different input variables yields a signal region that mathematically appears like a N-dimensional hypercube, where N is the number of variables cut upon.

Optimization of a rectangular cuts method usually involves calculating a figure of merit (FOM) (such as categorical accuracy or Asimov significance) at each possible cut along the N-dimensional space. As a result, cut selection can be dependent on the granularity of the scan; an appropriately wide binning is important to not overtrain the model on fluctuations.

A.2.2 Boosted Decision Tree

The fundamental idea of a decision tree is a familiar one. Input data is split based on discriminating features, each level of the tree creating a new split (called node) motivated by best separating data from the target class from background data. This procedure is repeated until a stopping condition is met, typically either minimum number of samples in a node or a maximum tree depth is achieved (both hyperparameters). Once final splits have been made, the predominant sample type present in a node is assigned to all of the data in that node. A pictorial diagram of this logic can be found in Figure A.1.

Single decision trees do not typically do a good job at classification tasks, so the principle of *boosting* is used in order to improve the classification accuracy. Boosting is a type of ensemble method, which are a class of methods that combine the output of many weak learners (classifier with small correlation to truth) to form one strong learner (classifier

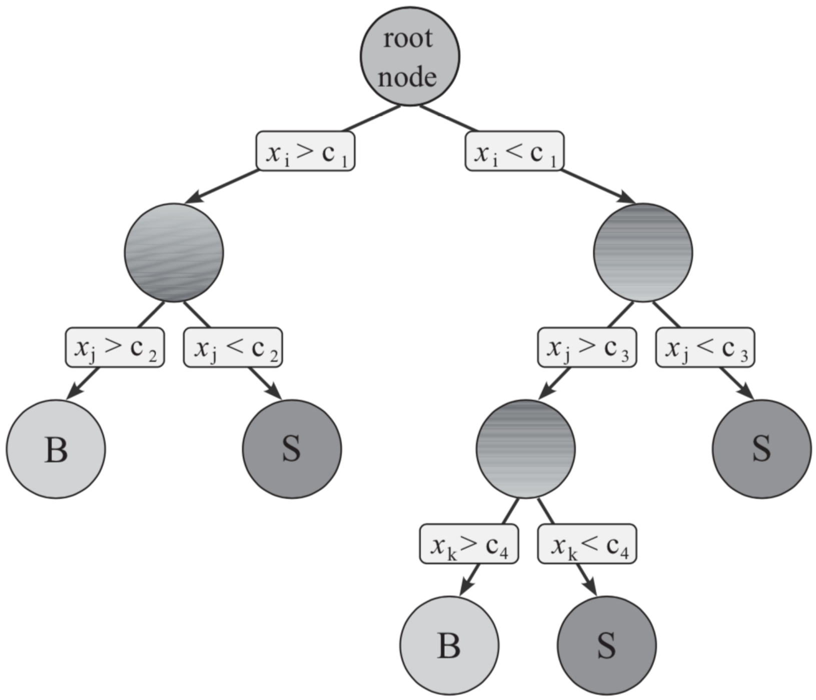


Figure A.1: A diagram of a single decision tree, where data is split on properties x_N. The presented tree has a maximum depth of 3, and is classified into either signal (S) or background (B) in stopping nodes [119].

well-correlated to truth). In boosting, single decision trees are trained, then misclassifications within that tree are identified and up-weighted in the next iteration. All of the trees are ultimately combined to create the final strong learner, called a Boosted Decision Tree (BDT). By this construction, cutting on a BDT score gives a more flexible solution space than that of a rectangular cuts method or a single decision tree. There are various different boosting methods, notably "Adaboost" [120] and "Gradient Boosting" [121], which differ by the definition of their loss function.

The BDTs in this thesis are implemented through several different frameworks. `ROOT` [122] has a native toolkit called `TMVA` [90] that includes a BDT implementation with flexible options for boosting types and hyperparameters. `XGBoost` [123] is an open-source library that performs gradient boosting, and is highly optimized for fast training and good performance.

A.2.3 Neural Networks

An artificial Neural Network (NN)[2] is a nonlinear machine learning algorithm that attempts to perform a task via a combination of adaptive nonlinear basis functions ($h(x)$) where parameter values and weights (w) are learned in the training procedure[3]. The resulting function is:

$$y(x) = w_0^2 + \sum_{m=1}^{M} \left[w_m^2 \cdot h\big(w_{0m}^1 + \sum_{k=1}^{D} w_{km}^1 x_k)\right] \tag{A.1}$$

Where D is the number of inputs and M is the number of basis functions. An illustration of a NN with $D = 4$ and $M = 5$ is shown in Figure A.2. Here h is known as an *activation function*, and given a summation over enough of these activation functions, any possible function can be approximated [124]. The choice of activation function is motivated by both the task as well as the type of data. A few common activation function choices are:

- The sigmoid function, given by $h(t) = 1/(1 + e^{-t})$. The range of this function is $[0, 1]$, so it is typically used to map predicted values to probabilities, thus used on output layers.

- The hyperbolic tangent function, given by $h(t) = (e^t - e^{-t}/(e^t + e^{-t}))$. This function is also used to map to outputs, but unlike the sigmoid is centered at zero, and has a monotonically increasing derivative.

[2]Historically, these have been abbreviated ANN for "Artificial Neural Network," however to avoid ambiguity with modern "Adversarial Neural Networks," the abbreviation NN is used.

[3]The name "Neural Network" comes from such models being inspired by neurological structures. At each neuron, signals are received by dendrites, passed along an axon to produce some output, which is passed along through axon terminals. Many of these neurons come together to process sophisticated information. This loosely mirrors the topology of an NN, where basis functions parallel neurons.

- The rectified linear unit (ReLU), given by $h(t) = \max(0, x)$, in other words, returning 0 below $x = 0$ and $y = x$ otherwise. This converges quickly and is typically used in layers other than the output.

- The softmax function, given by $h(t) = e^{t_i} / \sum_j e^{t_j}$, for class i in a multiclassification problem with j outputs.

- The linear function, given by $h(t) = t$, commonly used on outputs for regression problems.

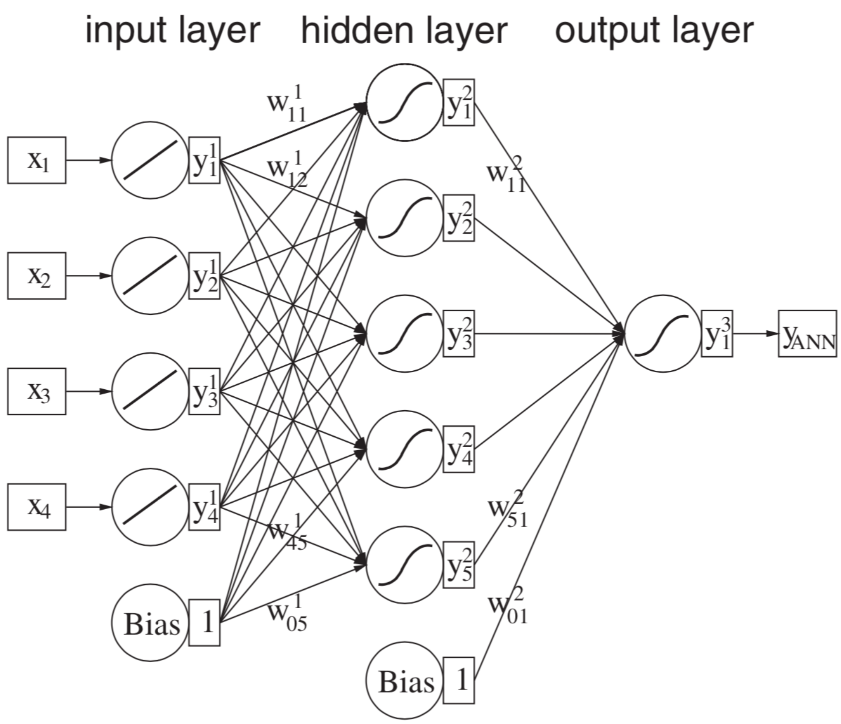


Figure A.2: A sample NN with $D = 4$ and $M = 5$ [119].

In modern research, various advanced NN topologies are being leveraged for domain specific tasks[4], but the domain of problems in this thesis are classification problems, so

[4]E.g. Convolutional NNs used for spatially dependent data and leveraged in computer vision problems, Recurrent NNs are used for sequential data and leveraged in text-prediction problems, Generative Adversarial Networks are leveraged for generative modeling, amongst others.

the topology used is a simple *feed-forward* network, also commonly known as a *multilayer perceptron*. Feed forward means that data proceeds from strictly from input to output (left to right in Figure A.2), with no closed internal cycles.

In the training of feed-forward NNs, weights are adjusted to minimize the loss function through a backpropagation [125]. Since the error at any layer is dependent on the subsequent layer, errors flow backward in the network. Thus, backpropagation works backward through the network, from the output layer of the network to the input layer, computing the gradient of the loss function with respect to each weight. Networks are then trained via iterative steps of calculating an output (forward phase) and evaluating the error terms (backward phase). Specifically in each *epoch* of the network's training, it calculates an output on a *batch* of data, and then updates weights and biases through backpropagation (where number of epochs and batch size are hyperparameters of the network).

The concept of dropout [92] is also employed in the neural networks built in this thesis. In backpropagation, each parameter is updated in order to minimize the loss function, dependent on the behavior of the whole network. This can cause units to compensate for one another, leading to cross-correlations between nodes. These correlations, known as co-adaptations, reduce the ability of a model to generalize. Dropout, a regularization technique, attempts avoid this through randomly removing a percentage of nodes on each update cycle. By making the presence of other nodes unreliable, this reduces the likelihood for co-adaptations, and leads to better model generalizability.

The NNs in this thesis are constructed using the `Keras` Application Program Interface (API), an open-source library that provides a modular interface to construct NNs, through high-level bindings to a backend software. In the cases of the NNs used in this thesis, the `Tensorflow` [126] backend was used.

APPENDIX B

TOPOLOGICAL CLUSTER DATA-MC COMPARISONS

In addition to the comparisons shown in the body of the work, the full set of comparisons between data and MC for the topo-cluster variables used as inputs to the models presented are shown here. The number of topo-clusters is shown for unconverted photons in Figure B.1, and in Figure B.5 for converted photons.

The distributions for $\sqrt{\lambda_0^2}$ and $\sqrt{r_0^2}$ for unconverted photons is shown in Section 4.2.1. The centroid depth, isolation, and EM probability of the leading topo-cluster for unconverted photons are shown in Figures B.2, B.3, and B.4, respectively. For converted photons, the leading topo-cluster $\sqrt{\lambda_0^2}$ is shown in Figure B.6, $\sqrt{r_0^2}$ is shown in Figure B.7, centroid depth is shown in Figure B.8, isolation is shown in Figure B.9, and EM probability is shown in Figure B.10. Plots are binned according to the p_{T} and η bins used in the photon ID menu.

In general, the MC models the unconverted photons well, with the exception of the isolation, particularly in the most central bin and beyond the crack. For converted photons, $\sqrt{r_0^2}$ modeling differs between data and MC, and isolation differs considerably, particularly in the most central bin. On average converted photons have more topo-clusters per photon object, meaning they suffer more from cluster splitting, which influences these effects.

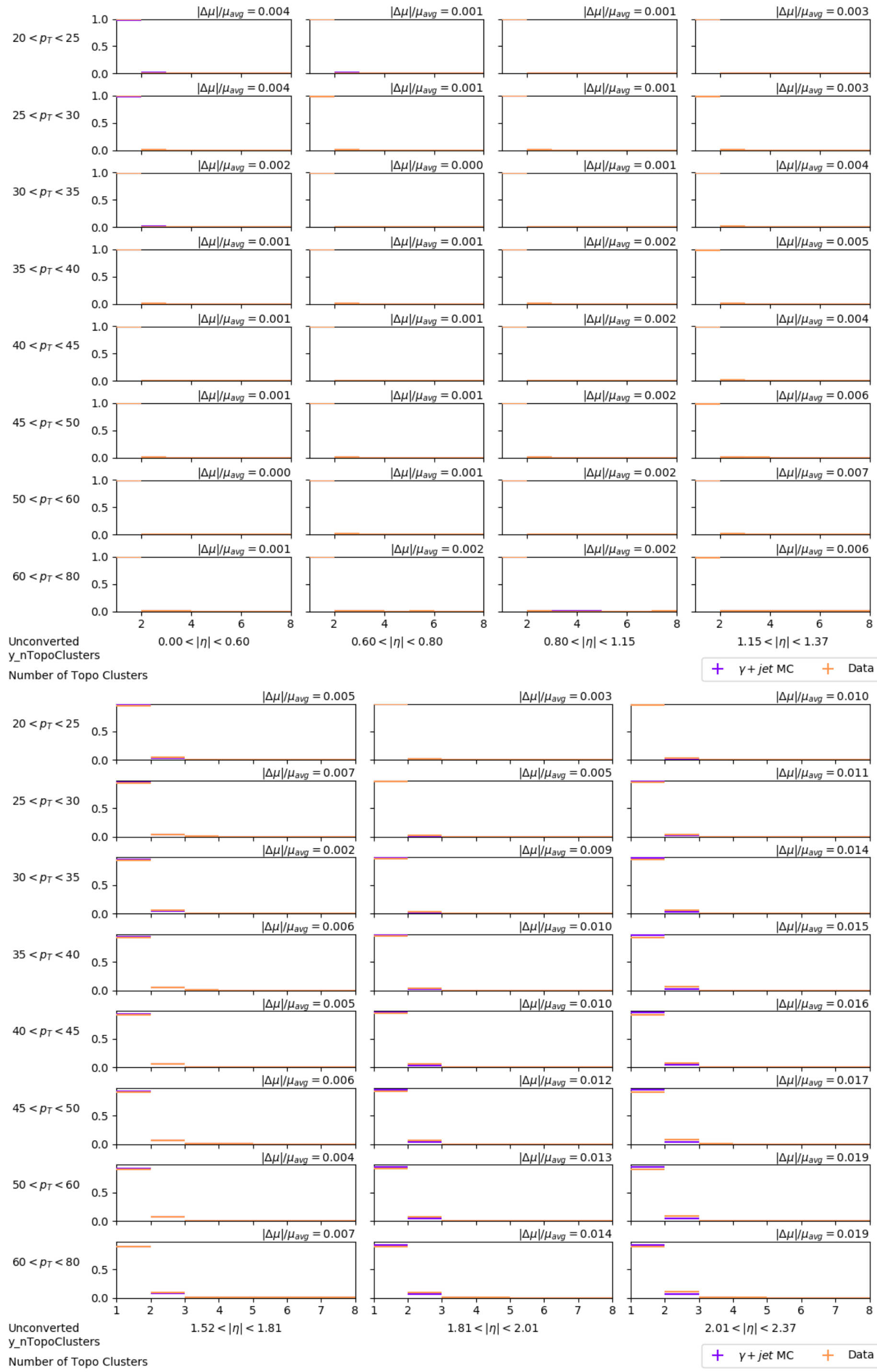


Figure B.1: Data-MC comparison of the number of topo-clusters for converted photons. The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}|$ is shown.

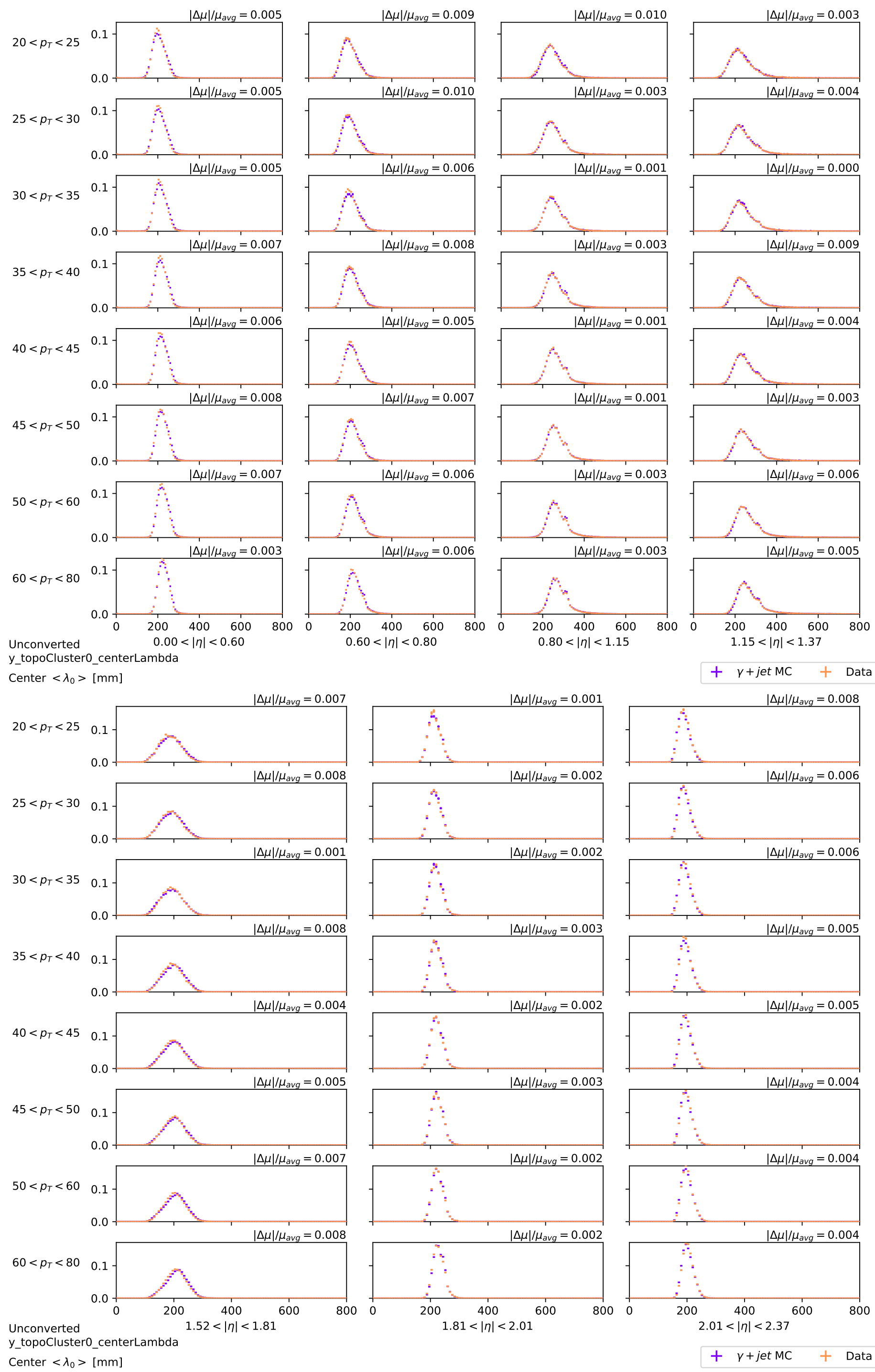


Figure B.2: Data-MC comparison of the centroid depth of the leading topo-cluster ($< \lambda_0 >$) for unconverted photons. The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}|$ is shown.

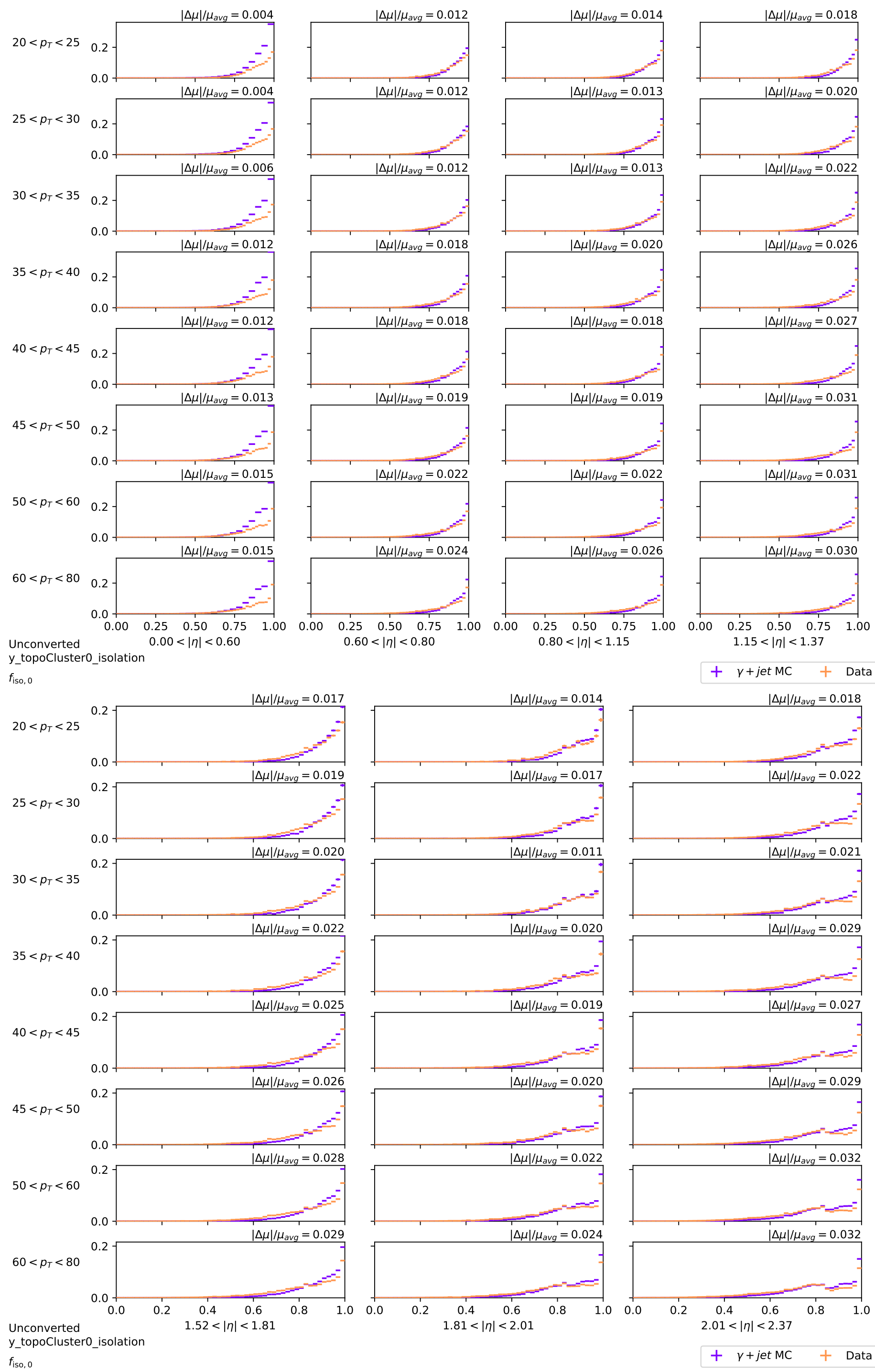


Figure B.3: Data-MC comparison of the isolation of the leading topo-cluster (f_{iso}) for unconverted photons. The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}|$ is shown.

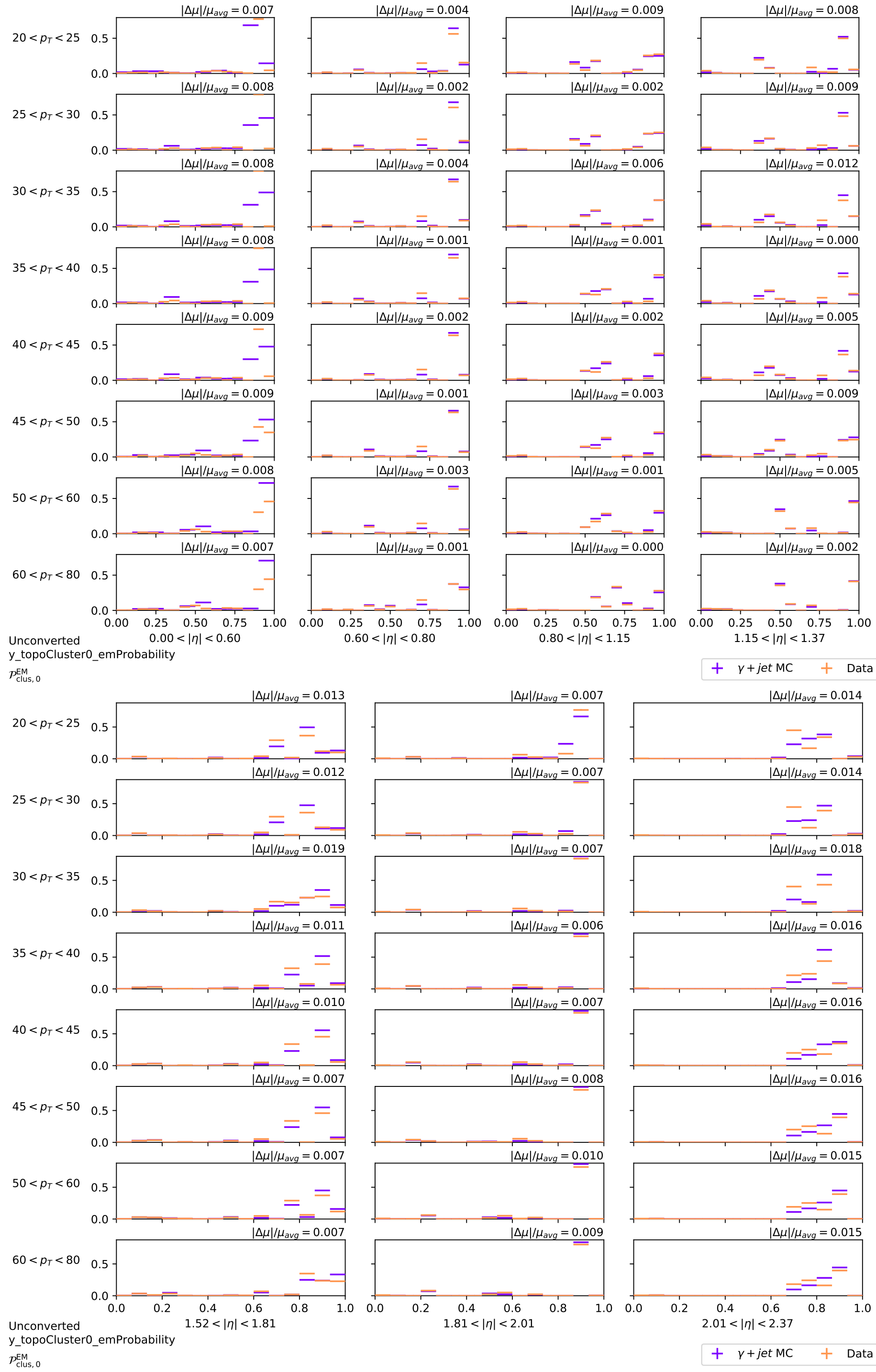


Figure B.4: Data-MC comparison of the EM probability of the leading topo-cluster ($\mathcal{P}^{\mathrm{EM}}_{\mathrm{clus}}$) for unconverted photons. The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}|$ is shown.

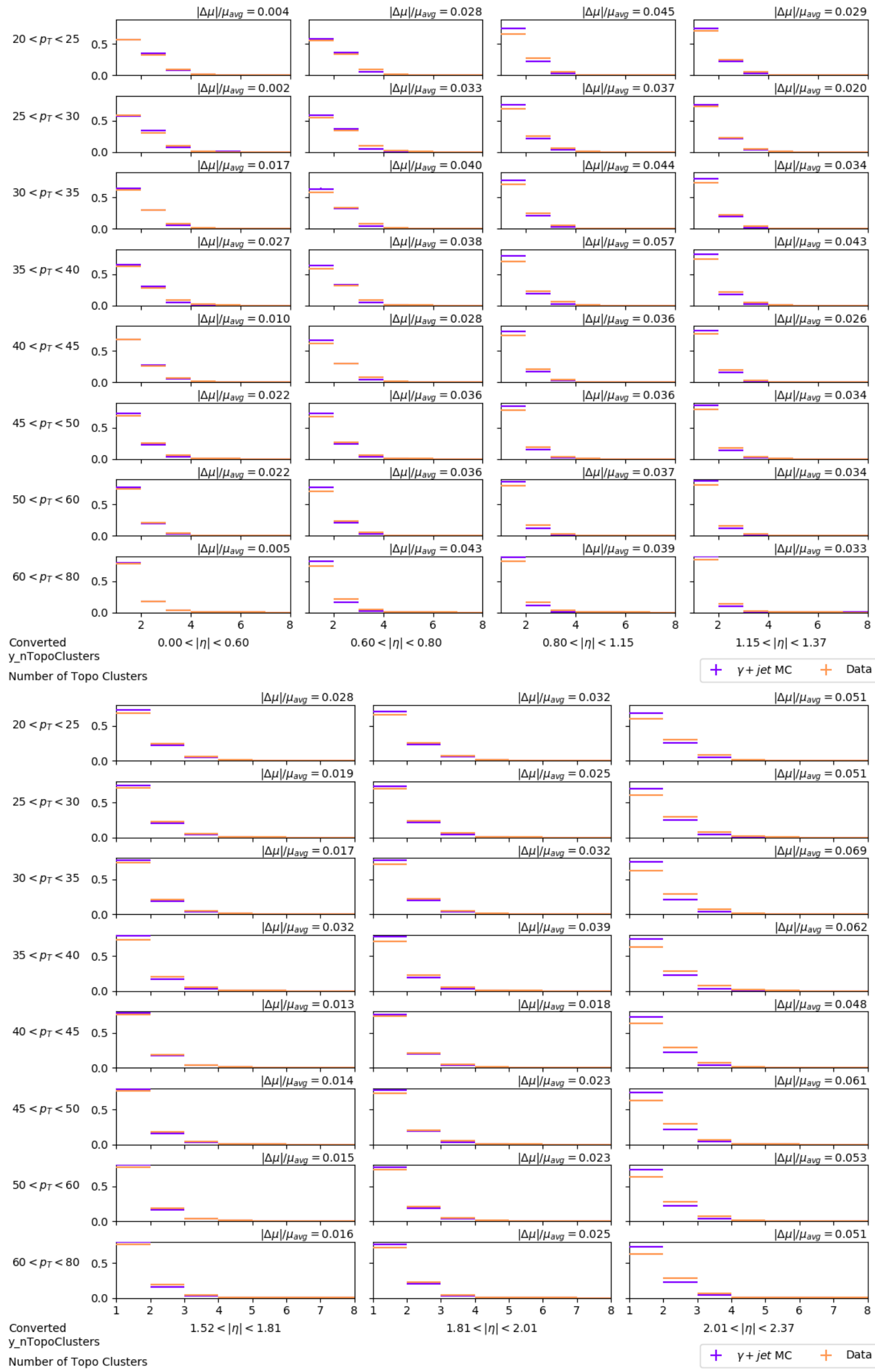


Figure B.5: Data-MC comparison of the number of topo-clusters for converted photons. The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}|$ is shown.

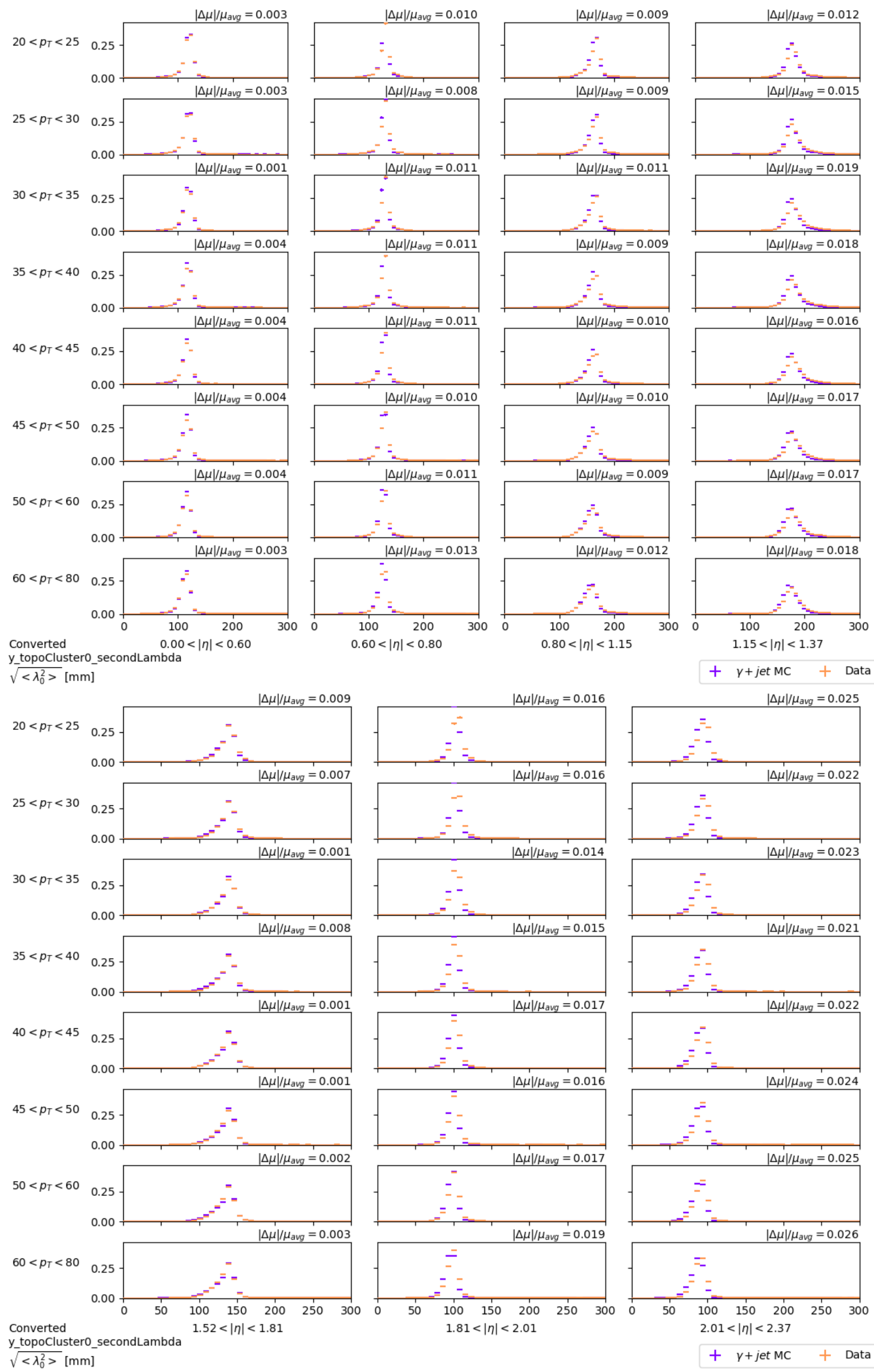


Figure B.6: Data-MC comparison of the semi-major axis in depth of the leading topo-cluster ($\sqrt{\lambda_0^2}$) for converted photons. The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}|$ is shown.

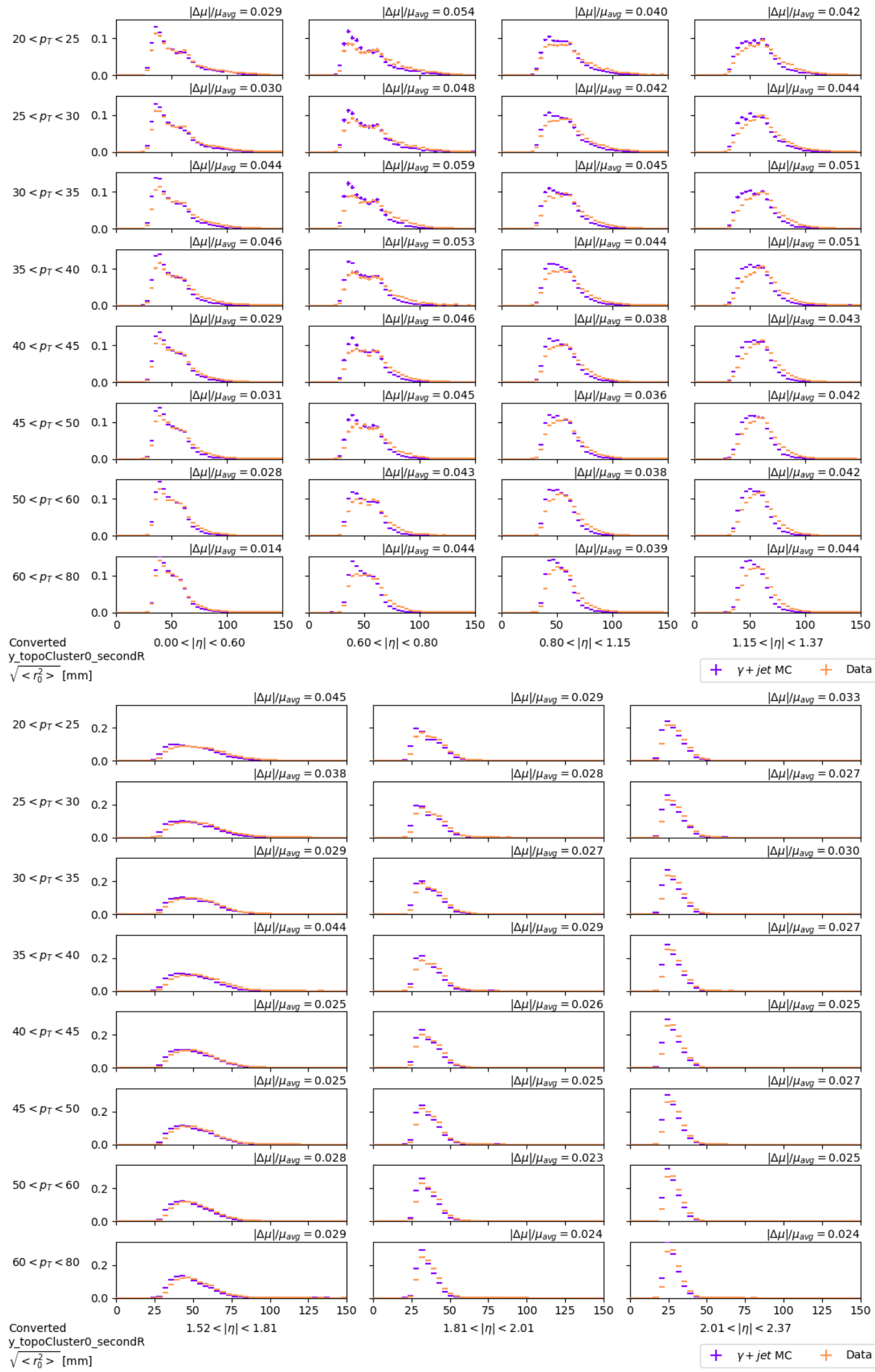


Figure B.7: Data-MC comparison of the semi-major axis in width of the leading topo-cluster ($\sqrt{r_0^2}$) for converted photons.The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}|$ is shown.

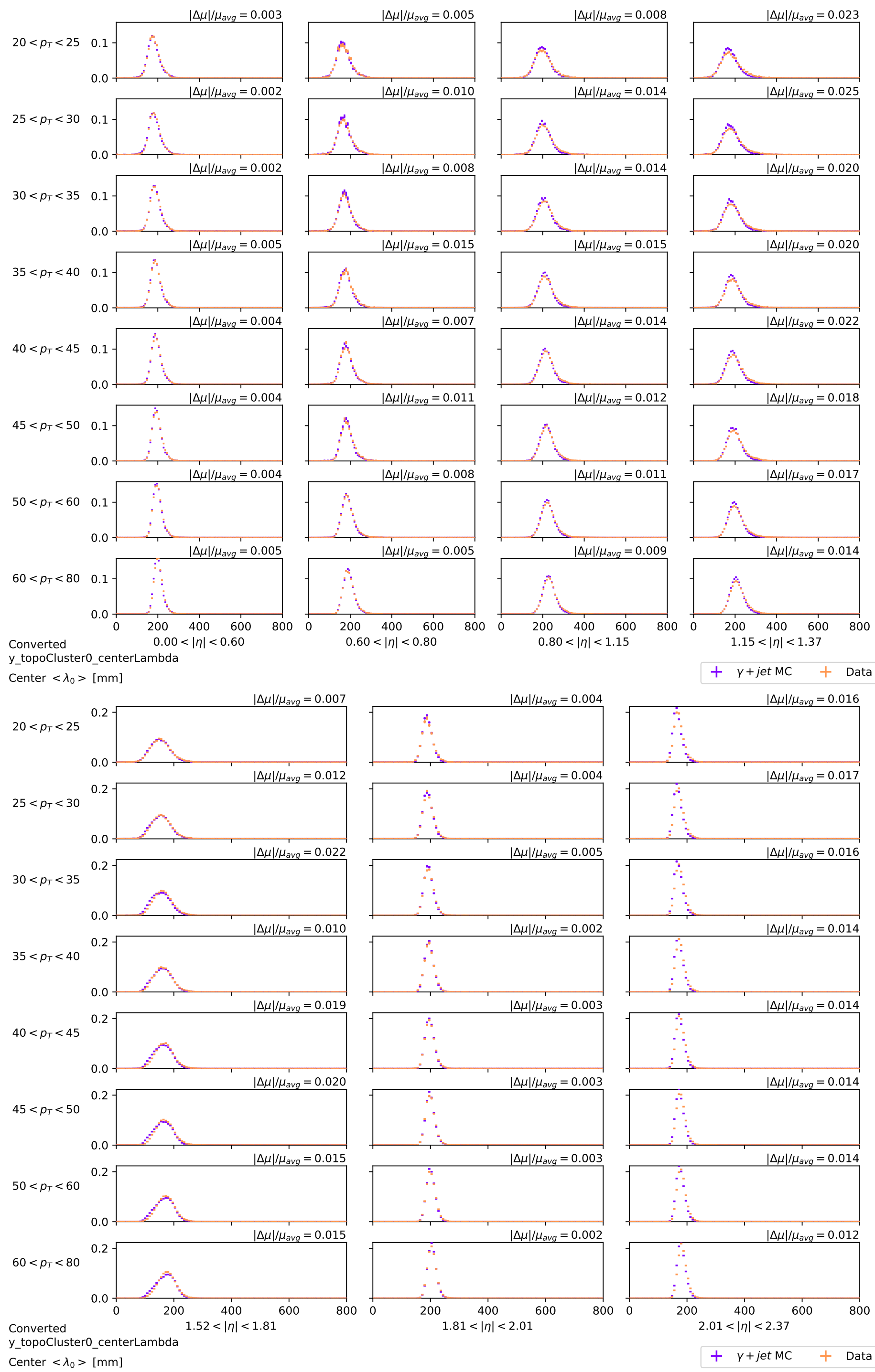


Figure B.8: Data-MC comparison of the centroid depth of the leading topo-cluster ($< \lambda_0 >$) for converted photons. The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}|$ is shown.

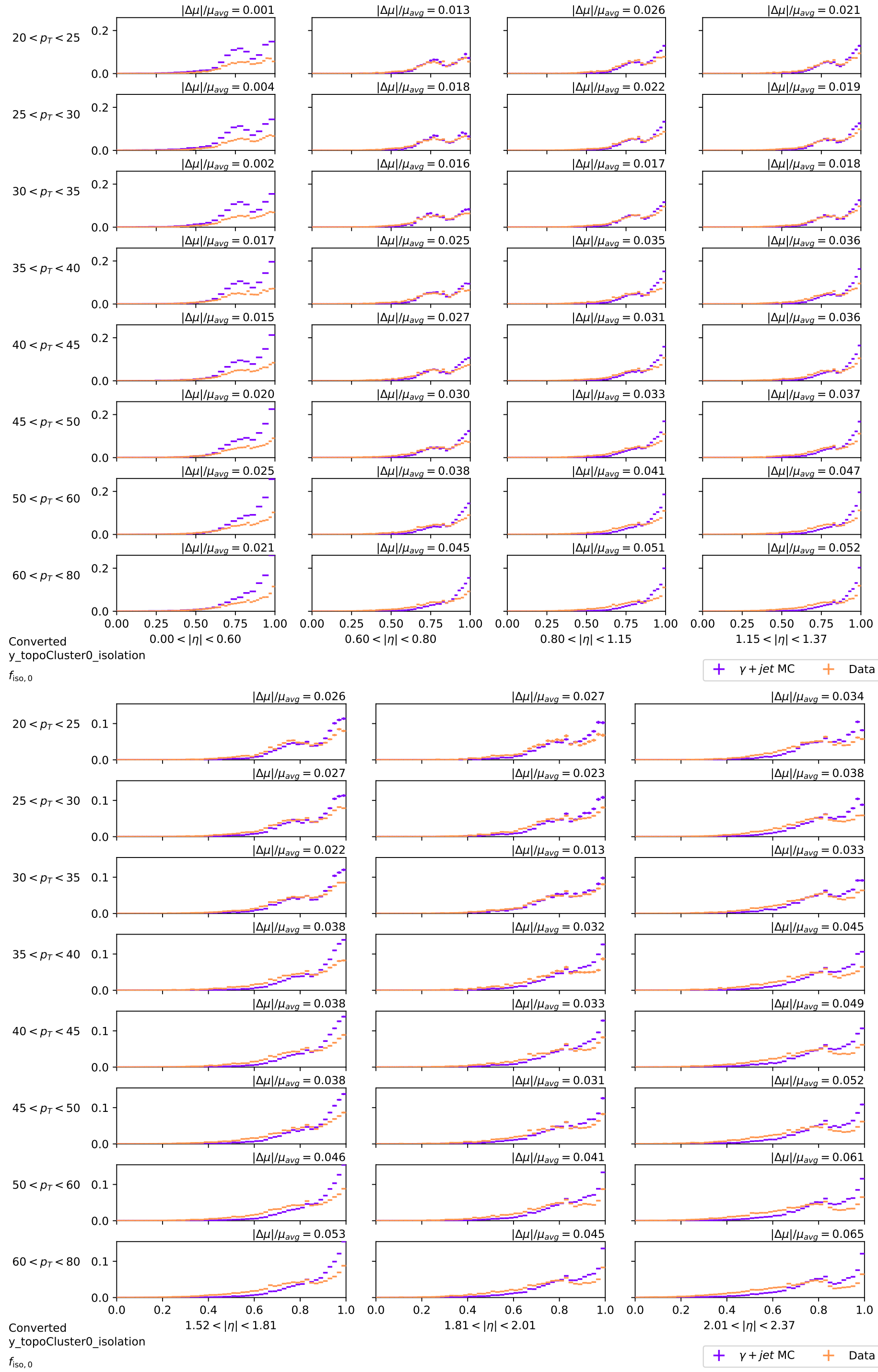


Figure B.9: Data-MC comparison of the isolation of the leading topo-cluster (f_{iso}) for converted photons. The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}|$ is shown.

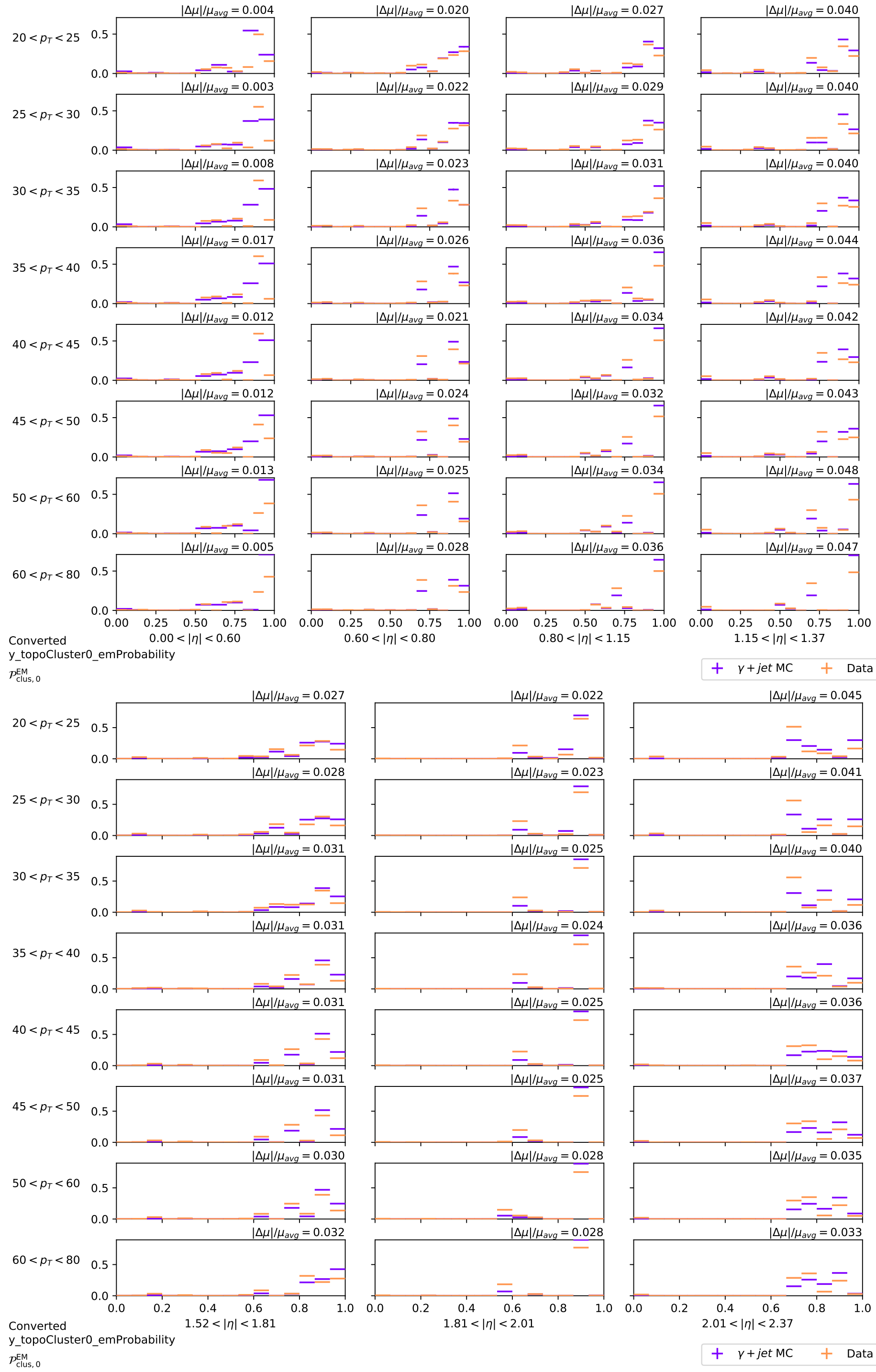


Figure B.10: Data-MC comparison of the EM probability of the leading topo-cluster ($\mathcal{P}^{\text{EM}}_{\text{clus}}$) for converted photons. The current tight identification is applied to both samples, and the MC requires truth-matched photons. The absolute difference of the distribution means normalized to their average mean, $|\Delta\mu|/\mu_{avg}|$ is shown.

APPENDIX C

2X2D SIDEBAND METHOD

The 2x2D sideband method[1] defines a $\texttt{Loose}'$ identification preselection, using all the standard photon ID shower shape cuts, except on F_{side}, w_{s3}, ΔE, and E_{ratio}. Events that pass this criteria are input into the 2x2D method, and split into 16 orthogonal subsamples, based on if each photon passes or fails tight identification criteria or isolation criteria. This creates 15 background regions where one or both of the photons fails one or both of the identification and isolation criteria. The signal region is the category where both photons pass both identification and isolation criteria. In each region, the yield can be expressed in terms of signal and background yields for the whole L'L' sample, along with identification and isolation efficiencies for photons and jets and the correlations between the isolation distributions of the two fake photons in dijet events as follows:

$$\begin{aligned} N_{AA} &= W_{\gamma\gamma}^{\mathrm{L'L'}} \varepsilon_{I1}\varepsilon_{T1}\varepsilon_{I2}\varepsilon_{T2} \\ &+ W_{\gamma j}^{\mathrm{L'L'}} \varepsilon_{I1}\varepsilon_{T1} f'_{I2} f_{T2}\xi_{j2} \\ &+ W_{j\gamma}^{\mathrm{L'L'}} \varepsilon_{I2}\varepsilon_{T2} f_{I1} f_{T1}\xi_{j1} \\ &+ W_{jj}^{\mathrm{L'L'}} f'_{I1} f'_{T1} f'_{I2} f'_{T2}\xi_{Ijj}\xi_{j1}\xi_{j2} \end{aligned} \tag{C.1}$$

$$\begin{aligned} N_{AB} &= W_{\gamma\gamma}^{\mathrm{L'L'}} \varepsilon_{I1}\varepsilon_{T1}\left(1-\varepsilon_{I2}\right)\varepsilon_{T2} \\ &+ W_{\gamma j}^{\mathrm{L'L'}} \varepsilon_{I1}\varepsilon_{T1}\left(1-f'_{I2}\xi_{j2}\right) f_{T2} \\ &+ W_{j\gamma}^{\mathrm{L'L'}} f_{I1} f_{T1}\left(1-\varepsilon_{I2}\right)\varepsilon_{T2}\xi_{j1} \\ &+ W_{jj}^{\mathrm{L'L'}} f'_{I1} f'_{T1}\left(1-f'_{I2}\xi_{Ijj}\xi_{j2}\right) f'_{T2}\xi_{j1} \end{aligned} \tag{C.2}$$

[1] This method is described in detail in Reference [89]

$$\begin{aligned} N_{BA} &= W_{\gamma\gamma}^{\mathrm{L'L'}} \left(1-\varepsilon_{I1}\right) \varepsilon_{T1}\varepsilon_{I2}\varepsilon_{T2} \\ &+ W_{\gamma j}^{\mathrm{L'L'}} \left(1-\varepsilon_{I1}\right) \varepsilon_{T1} f'_{I2} f_{T2} \xi_{j2} \\ &+ W_{j\gamma}^{\mathrm{L'L'}} \left(1-f_{I1}\xi_{j1}\right) f_{T1}\varepsilon_{I2}\varepsilon_{T2} \\ &+ W_{jj}^{\mathrm{L'L'}} \left(1-f'_{I1}\xi_{Ijj}\xi_{j1}\right) f'_{T1} f'_{I2} f'_{T2}\xi_{j2} \end{aligned} \tag{C.3}$$

$$\begin{aligned} N_{BB} &= W_{\gamma\gamma}^{\mathrm{L'L'}} \left(1-\varepsilon_{I1}\right)\left(1-\varepsilon_{I2}\right) \varepsilon_{T1}\varepsilon_{T2} \\ &+ W_{\gamma j}^{\mathrm{L'L'}} \left(1-\varepsilon_{I1}\right)\left(1 f'_{I2}\xi_{j2}\right) \varepsilon_{T1} f_{T2} \\ &+ W_{j\gamma}^{\mathrm{L'L'}} \left(1-\varepsilon_{I2}\right)\left(1-f_{I1}\xi_{j1}\right) \varepsilon_{T2} f_{T1} \\ &+ W_{jj}^{\mathrm{L'L'}} \left(1-f'_{I1}\xi_{j1}-f'_{I2}\xi_{j2}+f'_{I1}f'_{I2}\xi_{Ijj}\xi_{j1}\xi_{j2}\right) f'_{T1} f'_{T2} \end{aligned} \tag{C.4}$$

$$\begin{aligned} N_{AC} &= W_{\gamma\gamma}^{\mathrm{L'L'}} \varepsilon_{I1}\varepsilon_{T1}\varepsilon_{I2}\left(1-\varepsilon_{T2}\right) \\ &+ W_{\gamma j}^{\mathrm{L'L'}} \varepsilon_{I1}\varepsilon_{T1} f'_{I2}\left(1-f_{T2}\xi_{j2}\right) \\ &+ W_{j\gamma}^{\mathrm{L'L'}} f_{I1} f_{T1}\varepsilon_{I2}\left(1-\varepsilon_{T2}\right)\xi_{j1} \\ &+ W_{jj}^{\mathrm{L'L'}} f'_{I1} f'_{T1} f'_{I2}\left(1-f'_{T2}\xi_{j2}\right)\xi_{Ijj}\xi_{j1} \end{aligned} \tag{C.5}$$

$$
\begin{aligned}
N_{CA} &= W_{\gamma\gamma}^{\mathrm{L'L'}}\varepsilon_{I1}\left(1-\varepsilon_{T1}\right)\varepsilon_{I2}\varepsilon_{T2} \\
&+ W_{\gamma j}^{\mathrm{L'L'}}\varepsilon_{I1}\left(1-\varepsilon_{T1}\right)f'_{I2}f_{T2}\xi_{j2} \\
&+ W_{j\gamma}^{\mathrm{L'L'}}f_{I1}\left(1-f_{T1}\xi_{j1}\right)\varepsilon_{I2}\varepsilon_{T2} \\
&+ W_{jj}^{\mathrm{L'L'}}f'_{I1}\left(1-f'_{T1}\xi_{j1}\right)f'_{I2}f'_{T2}\xi_{Ijj}\xi_{j2} \qquad \text{(C.6)}
\end{aligned}
$$

$$
\begin{aligned}
N_{AD} &= W_{\gamma\gamma}^{\mathrm{L'L'}}\varepsilon_{I1}\varepsilon_{T1}\left(1-\varepsilon_{I2}\right)\left(1-\varepsilon_{T2}\right) \\
&+ W_{\gamma j}^{\mathrm{L'L'}}\varepsilon_{I1}\varepsilon_{T1}\left(1-f'_{I2}-f_{T2}+f'_{I2}f_{T2}\xi_{j2}\right) \\
&+ W_{j\gamma}^{\mathrm{L'L'}}f_{I1}f_{T1}\xi_{j1}\left(1-\varepsilon_{I2}\right)\left(1-\varepsilon_{T2}\right) \\
&+ W_{jj}^{\mathrm{L'L'}}\xi_{j1}f'_{I1}f'_{T1}\left(1-\xi_{Ijj}f'_{I2}-f'_{T2}+f'_{I2}f'_{T2}\xi_{j2}\xi_{Ijj}\right) \qquad \text{(C.7)}
\end{aligned}
$$

$$
\begin{aligned}
N_{DA} &= W_{\gamma\gamma}^{\mathrm{L'L'}}\left(1-\varepsilon_{I1}\right)\left(1-\varepsilon_{T1}\right)\varepsilon_{I2}\varepsilon_{T2} \\
&+ W_{\gamma j}^{\mathrm{L'L'}}\left(1-\varepsilon_{I1}\right)\left(1-\varepsilon_{T1}\right)f'_{I2}f_{T2}\xi_{j2} \\
&+ W_{j\gamma}^{\mathrm{L'L'}}\left(1-f_{I1}-f_{T1}+f_{I1}f_{T1}\xi_{j1}\right)\varepsilon_{I2}\varepsilon_{T2} \\
&+ W_{jj}^{\mathrm{L'L'}}\left(1-f'_{I1}\xi_{Ijj}-f'_{T1}+f'_{I1}f'_{T1}\xi_{j1}\xi_{Ijj}\right)f'_{I2}f'_{T2}\xi_{j2} \qquad \text{(C.8)}
\end{aligned}
$$

$$\begin{aligned} N_{BC} &= W_{\gamma\gamma}^{\mathrm{L'L'}}\left(1-\varepsilon_{I1}\right)\varepsilon_{T1}\varepsilon_{I2}\left(1-\varepsilon_{T2}\right) \\ &+ W_{\gamma j}^{\mathrm{L'L'}}\left(1-\varepsilon_{I1}\right)\varepsilon_{T1}f'_{I2}\left(1-f_{T2}\xi_{j2}\right) \\ &+ W_{j\gamma}^{\mathrm{L'L'}}\left(1-f_{I1}\xi_{j1}\right)f_{T1}\varepsilon_{I2}\left(1-\varepsilon_{T2}\right) \\ &+ W_{jj}^{\mathrm{L'L'}}\left(1-f'_{I1}\xi_{Ijj}\xi_{j1}\right)f'_{T1}f'_{I2}\left(1-f'_{T2}\xi_{j2}\right) \end{aligned} \tag{C.9}$$

$$\begin{aligned} N_{CB} &= W_{\gamma\gamma}^{\mathrm{L'L'}}\varepsilon_{I1}\left(1-\varepsilon_{T1}\right)\left(1-\varepsilon_{I2}\right)\varepsilon_{T2} \\ &+ W_{\gamma j}^{\mathrm{L'L'}}\varepsilon_{I1}\left(1-f'_{I2}\xi_{j2}\right)\left(1-\varepsilon_{T1}\right)f_{T2} \\ &+ W_{j\gamma}^{\mathrm{L'L'}}f_{I1}\left(1-\varepsilon_{I2}\right)\left(1-f_{T1}\xi_{j1}\right)\varepsilon_{T2} \\ &+ W_{jj}^{\mathrm{L'L'}}f'_{I1}\left(1-f'_{T1}\xi_{j1}\right)\left(1-f'_{I2}\xi_{Ijj}\xi_{j2}\right)f'_{T2} \end{aligned} \tag{C.10}$$

$$\begin{aligned} N_{CC} &= W_{\gamma\gamma}^{\mathrm{L'L'}}\varepsilon_{I1}\left(1-\varepsilon_{T1}\right)\varepsilon_{I2}\left(1-\varepsilon_{T2}\right) \\ &+ W_{\gamma j}^{\mathrm{L'L'}}\varepsilon_{I1}\left(1-\varepsilon_{T1}\right)f'_{I2}\left(1-f_{T2}\xi_{j2}\right) \\ &+ W_{j\gamma}^{\mathrm{L'L'}}f_{I1}\left(1-f_{T1}\xi_{j1}\right)\varepsilon_{I2}\left(1-\varepsilon_{T2}\right) \\ &+ W_{jj}^{\mathrm{L'L'}}f'_{I1}f'_{I2}\left(1-f'_{T1}\xi_{j1}-f'_{T2}\xi_{j2}+f'_{T1}f'_{T2}\xi_{j1}\xi_{j2}\right)\xi_{Ijj} \end{aligned} \tag{C.11}$$

$$
\begin{aligned}
N_{BD} &= W_{\gamma\gamma}^{\mathrm{L'L'}}\left(1-\varepsilon_{I1}\right)\varepsilon_{T1}\left(1-\varepsilon_{I2}\right)\left(1-\varepsilon_{T2}\right) \\
&+ W_{\gamma j}^{\mathrm{L'L'}}(\left(1-\varepsilon_{I1}\right)\varepsilon_{T1}\left(1-f'_{I2}-f_{T2}+f'_{I2}f_{T2}\xi_{j2}\right) \\
&+ W_{j\gamma}^{\mathrm{L'L'}}\left(1-f_{I1}\xi_{j1}\right)f_{T1}\left(1-\varepsilon_{T2}\right)\left(1-\varepsilon_{I2}\right) \\
&+ W_{jj}^{\mathrm{L'L'}}f'_{T1}\left(1-f'_{I2}-f'_{T2}-f'_{I1}\xi_{j1}+f'_{I2}f'_{T2}\xi_{j2}+f'_{T2}f'_{I1}\xi_{j1}+f'_{I2}f'_{I1}\xi_{Ijj}\xi_{j1}-f'_{I1}f'_{I2}f'_{T2}\xi_{Ijj}\xi_{j1}\xi_{j2}\right)
\end{aligned} \tag{C.12}
$$

$$
\begin{aligned}
N_{DB} &= W_{\gamma\gamma}^{\mathrm{L'L'}}\left(1-\varepsilon_{I1}\right)\left(1-\varepsilon_{T1}\right)\left(1-\varepsilon_{I2}\right)\varepsilon_{T2} \\
&+ W_{\gamma j}^{\mathrm{L'L'}}\left(1-\varepsilon_{I1}\right)\left(1-\varepsilon_{T1}\right)\left(1-f'_{I2}\xi_{j2}\right)f_{T2} \\
&+ W_{j\gamma}^{\mathrm{L'L'}}\left(1-f_{I1}-f_{T1}+f_{I1}f_{T1}\xi_{j1}\right)\left(1-\varepsilon_{I2}\right)\varepsilon_{T2} \\
&+ W_{jj}^{\mathrm{L'L'}}\left(1-f'_{I1}-f'_{T1}-f'_{I2}\xi_{j2}+f'_{I1}f'_{T1}\xi_{j1}+f'_{T1}f'_{I2}\xi_{j2}+f'_{I1}f'_{I2}\xi_{Ijj}\xi_{j2}-f'_{I1}f'_{T1}f'_{I2}\xi_{Ijj}\xi_{j1}\xi_{j2}\right)f'_{T2}
\end{aligned} \tag{C.13}
$$

$$
\begin{aligned}
N_{CD} &= W_{\gamma\gamma}^{\mathrm{L'L'}}\varepsilon_{I1}\left(1-\varepsilon_{T1}\right)\left(1-\varepsilon_{I2}\right)\left(1-\varepsilon_{T2}\right) \\
&+ W_{\gamma j}^{\mathrm{L'L'}}\varepsilon_{I1}\left(1-\varepsilon_{T1}\right)\left(1-f'_{I2}-f_{T2}+f'_{I2}f_{T2}\xi_{j2}\right) \\
&+ W_{j\gamma}^{\mathrm{L'L'}}f_{I1}\left(1-f_{T1}\xi_{j1}\right)\left(1-\varepsilon_{I2}\right)\left(1-\varepsilon_{T2}\right) \\
&+ W_{jj}^{\mathrm{L'L'}}f'_{I1}\left(\left(1-f'_{T1}\xi_{j1}\right)\left(1-f'_{I2}\xi_{Ijj}\right)-f'_{T2}\left(1-f'_{T1}\xi_{j1}\right)\left(1-f'_{I2}\xi_{j2}\xi_{Ijj}\right)\right)
\end{aligned} \tag{C.14}
$$

$$
\begin{aligned}
N_{DC} &= W_{\gamma\gamma}^{\mathrm{L'L'}} \varepsilon_{I2} \left(\left(1-\varepsilon_{T2}\right)\left(1-\varepsilon_{I1}\right) - \varepsilon_{T1}\left(1-\varepsilon_{T2}\right)\left(1-\varepsilon_{I1}\right) \right) \\
&+ W_{\gamma j}^{\mathrm{L'L'}} f'_{I2} \left(\left(1-\varepsilon_{I1}\right)\left(1-f_{T2}\xi_{j2}\right) - \varepsilon_{T1}\left(1-\varepsilon_{I1}\right)\left(1-f_{T2}\xi_{j2}\right) \right) \\
&+ W_{j\gamma}^{\mathrm{L'L'}} \varepsilon_{I2} \left(\left(1-\varepsilon_{T2}\right)\left(1-f_{I1}\right) - f_{T1}\left(1-f_{I1}\xi_{j1}\right)\left(1-\varepsilon_{T2}\right) \right) \\
&+ W_{jj}^{\mathrm{L'L'}} \left(\left(1-f'_{T2}\xi_{j2}\right)\left(1-f'_{I1}\xi_{Ijj}\right) - f'_{T1}\left(1-f'_{T2}\xi_{j2}\right)\left(1-f'_{I1}\xi_{j1}\xi_{Ijj}\right) \right) f'_{I2}
\end{aligned}
\tag{C.15}
$$

$$
\begin{aligned}
N_{DD} &= W_{\gamma\gamma}^{\mathrm{L'L'}} \left(1-\varepsilon_{T1}-\varepsilon_{T2}+\varepsilon_{T1}\varepsilon_{T2}+\varepsilon_{I2}\left(1-\varepsilon_{T1}-\varepsilon_{T2}+\varepsilon_{T1}\varepsilon_{T2}\right)+\varepsilon_{I1}\left(-\left(1-\varepsilon_{I2}\right)\left(1-\varepsilon_{T1}\right)+\varepsilon_{T2}\left(1-\varepsilon_{I2}\right)\left(1-\varepsilon_{T1}\right)\right)\right) \\
&+ W_{\gamma j}^{\mathrm{L'L'}} \left(1-\varepsilon_{T1}-f_{T2}+\varepsilon_{T1}f_{T2}+f'_{I2}\left(1-\varepsilon_{T1}-f_{T2}\xi_{j2}+\varepsilon_{T1}f_{T2}\xi_{j2}\right)+\varepsilon_{I1}\left(\left(1-\varepsilon_{T1}\right)\left(1-f'_{I2}\right)+f_{T2}\left(1-f'_{I2}\xi_{j2}\right)\left(1-\varepsilon_{T1}\right)\right)\right) \\
&+ W_{j\gamma}^{\mathrm{L'L'}} \left(1-\varepsilon_{T2}-f_{T1}+\varepsilon_{T2}f_{T1}+\varepsilon_{I2}\left(1-f_{T1}-\varepsilon_{T2}+\varepsilon_{T2}f_{T1}\right)+f_{I1}\left(\left(1-\varepsilon_{I2}\right)\left(1-f_{T1}\xi_{j1}\right)+\varepsilon_{T2}\left(1-\varepsilon_{I2}\right)\left(1-f_{T1}\xi_{j1}\right)\right)\right) \\
&+ W_{jj}^{\mathrm{L'L'}} \left(1-f'_{T1}-f'_{T2}+f'_{T1}f'_{T2}+f'_{I2}\left(1-f'_{T1}-f'_{T2}\xi_{j2}+f'_{T1}f'_{T2}\xi_{j2}\right)+f'_{I1}\left(-\left(1-f'_{I2}\xi_{Ijj}\right)\left(1-f'_{T1}\xi_{j1}\right)+f'_{T2}\left(1-f'_{I2}\xi_{Ijj}\xi_{j2}\right)\left(1-f'_{T1}\xi_{j1}\right)\right)\right)
\end{aligned}
\tag{C.16}
$$

where

- ε_{I1} and ε_{I2} are the efficiencies of the isolation criteria for the leading and subleading photons respectively
- ε_{T1} and ε_{T2} are the `Tight` identification efficiencies for the leading and subleading photons respectively
- f_{I1} and f'_{I2} are the isolation fake rates for the jets in γj and $j\gamma$ events respectively
- f_{T1} and f_{T2} are the `Tight` identification fake rates for the jets in γj and $j\gamma$ events respectively
- f'_{I1} and f'_{I2} are the isolation fake rates for the leading and subleading jets in jj events
- f'_{T1} and f'_{T2} are the `Tight` identification fake rates for the leading and subleading jets in jj events

Table C.1: Isolation and identification efficiencies for true photons used as input to the 2x2D sideband method for the measurement of the inclusive diphoton purity of the sample passing the full selection, for the various selections. Efficiencies are determined with respect to the leading and subleading photon candidates of true diphoton events that pass the full selection except the isolation and tight identification criteria: the photons must instead pass the `Loose`′ identification requirements. The uncertainty arises from the limited MC statistics.

Selection		Isolation		Identification
Low mass 0-tag	ε_{I1}	0.957 ± 0.014	ε_{T1}	0.960 ± 0.013
	ε_{I2}	0.907 ± 0.028	ε_{T2}	0.948 ± 0.016
Low mass 1-tag	ε_{I1}	0.955 ± 0.014	ε_{T1}	0.962 ± 0.012
	ε_{I2}	0.906 ± 0.028	ε_{T2}	0.946 ± 0.017
Low mass 2-tag	ε_{I1}	0.955 ± 0.014	ε_{T1}	0.964 ± 0.011
	ε_{I2}	0.906 ± 0.028	ε_{T2}	0.945 ± 0.017
High mass 0-tag	ε_{I1}	0.961 ± 0.012	ε_{T1}	0.962 ± 0.011
	ε_{I2}	0.912 ± 0.025	ε_{T2}	0.949 ± 0.015
High mass 1-tag	ε_{I1}	0.954 ± 0.015	ε_{T1}	0.958 ± 0.013
	ε_{I2}	0.900 ± 0.029	ε_{T2}	0.948 ± 0.016
High mass 2-tag	ε_{I1}	0.960 ± 0.015	ε_{T1}	0.972 ± 0.010
	ε_{I2}	0.907 ± 0.029	ε_{T2}	0.943 ± 0.018

- ξ_{j1} and ξ_{j2} are the correlation between the identification and isolation fake rates for the leading and subleading jets in jj events

- ξ_{Ijj} is the isolation correlation factor between the jets in jj events

The 16 equations have 19 unknowns, six of which are constrained to a predetermined range:

- ε_{T1}, ε_{T2}, ε_{I1} and ε_{I2} are are determined from Monte Carlo simulation of diphoton events (see Table C.1).

- ξ_{j1} and ξ_{j2} are constrained to the range 1 ± 0.01

The other 13 unknowns are the outputs of the 2x2D method, derived by performing a global minimization procedure. Using these outputs, the observed yield from $\gamma\gamma$, $j\gamma$, γj, and jj events can be calculated for each signal and control region.

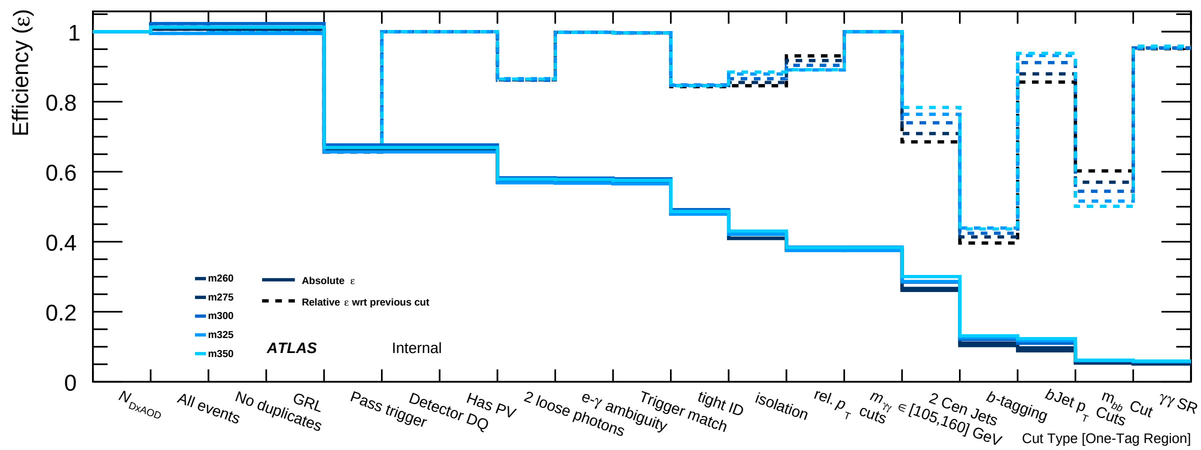
APPENDIX D

CUTFLOW STUDIES

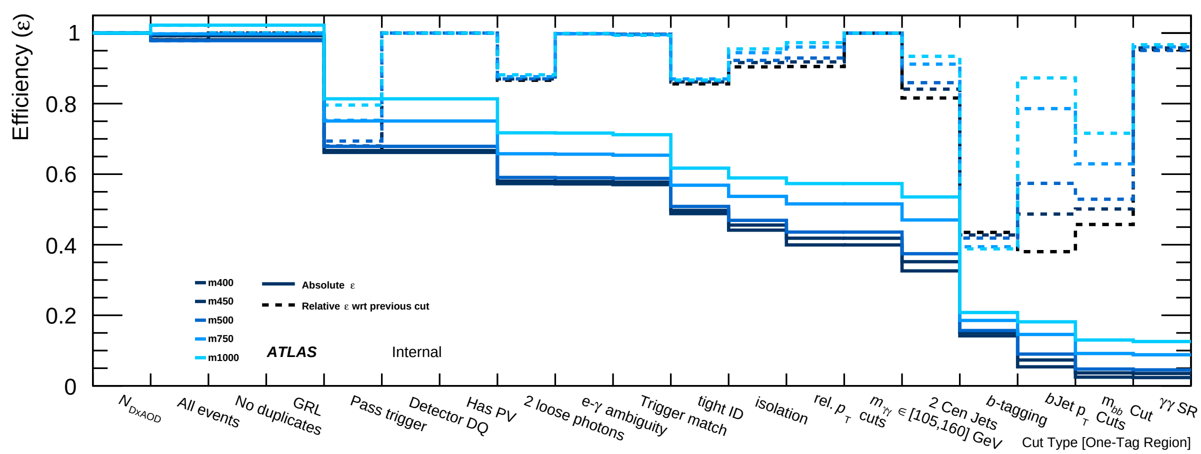
This appendix shows further studies on signal selection efficiency.

Resonant Signal Yield

The relative and absolute selection efficiency after each cut for the resonant samples is shown in Figure D.1 for the 1 b-tag category, and in Figure D.2 for the 0 b-tag category, for both the low and high mass selections, as indicated.

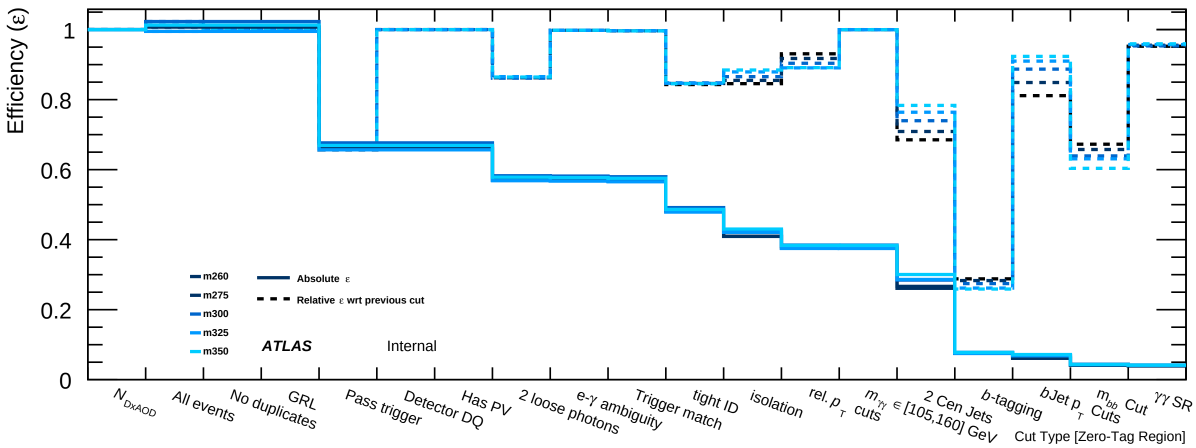


(a) Low Mass Selection

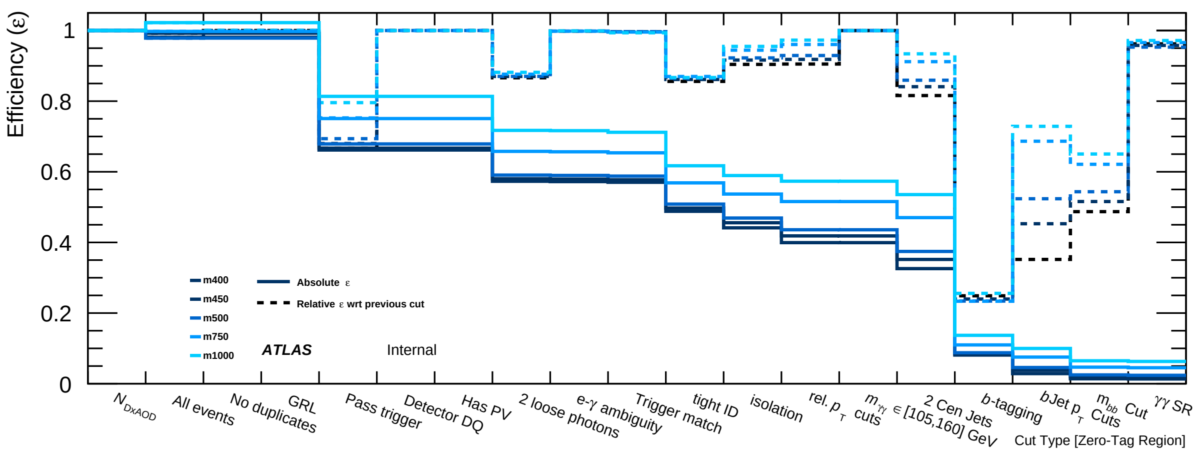


(b) High Mass Selection

Figure D.1: Absolute (solid line) and relative (dotted line) efficiencies for the 1 b-tag resonant signal samples to pass each cut in the low (a) and high (b) mass selection



(a) Low Mass Selection



(b) High Mass Selection

Figure D.2: Absolute (solid line) and relative (dotted line) efficiencies for the 0 b-tag resonant signal samples to pass each cut in the low (a) and high (b) mass selection

Table D.1: Cutflow for non-resonant $HH \rightarrow \gamma\gamma b\bar{b}$, 1 tag category (low mass selection)

Cuts	1 b-tag		
	Yield	Error	Efficiency
N_{xAOD}	3.147	0.027	–
N_{DxAOD}	3.147	0.019	–
All events	3.181	0.024	100.0
No duplicates	3.181	0.024	100.0
GRL	3.181	0.024	100.0
Pass trigger	2.184	0.021	68.7
Detector DQ	2.184	0.021	68.7
Has PV	2.184	0.021	68.7
2 *loose photons*	1.894	0.019	59.5
$e - \gamma$ *ambiguity*	1.891	0.019	59.5
Trigger match	1.885	0.019	59.3
tight ID	1.617	0.018	50.8
isolation	1.476	0.017	46.4
rel. p_T *cuts*	1.362	0.017	42.8
$m_{\gamma\gamma} \in [105, 160]\ GeV$	1.362	0.017	42.8
2 *Cen Jets*	1.149	0.016	36.1
$b - tagging$	0.490	0.011	15.4
bJet p_T *Cuts*	0.475	0.011	14.9
m_{bb} *Cut*	0.229	0.006	7.2
$\gamma\gamma$ *SR*	0.219	0.006	6.9

Non-Resonant Signal Yield

The yield and efficiency for each successive analysis cut are shown in Tables D.1, D.2, D.3, and D.4 for the 1 b-tag low mass, 1 b-tag high mass, 0 b-tag low mass, and 0 b-tag high mass selections respectively.

Background Yield

The background yield in the 1 and 0 b-tag categories is shown in Figures D.5 and D.6, respectively.

Table D.2: Cutflow for non-resonant $HH \rightarrow \gamma\gamma b\bar{b}$, 1 tag category (high mass selection)

Cuts	1 b-tag		
	Yield	Error	Efficiency
N_{xAOD}	3.147	0.027	–
N_{DxAOD}	3.147	0.019	–
$All\ events$	3.181	0.024	100.0
$No\ duplicates$	3.181	0.024	100.0
GRL	3.181	0.024	100.0
$Pass\ trigger$	2.184	0.021	68.7
$Detector\ DQ$	2.184	0.021	68.7
$Has\ PV$	2.184	0.021	68.7
$2\ loose\ photons$	1.894	0.019	59.5
$e-\gamma\ ambiguity$	1.891	0.019	59.5
$Trigger\ match$	1.885	0.019	59.3
$tight\ ID$	1.617	0.018	50.8
$isolation$	1.476	0.017	46.4
$rel.\ p_T\ cuts$	1.362	0.017	42.8
$m_{\gamma\gamma}\ \in\ [105, 160]\ GeV$	1.362	0.017	42.8
$2\ Cen\ Jets$	1.149	0.016	36.1
$b-tagging$	0.492	0.011	15.5
$bJet\ p_T\ Cuts$	0.247	0.009	7.8
$m_{bb}\ Cut$	0.127	0.004	4.0
$\gamma\gamma\ SR$	0.120	0.004	3.8

Table D.3: Cutflow for non-resonant $HH \to \gamma\gamma b\bar{b}$, 0 tag category (low mass selection)

Cuts	0 b-tag		
	Yield	Error	Efficiency
N_{xAOD}	3.147	0.027	–
N_{DxAOD}	3.147	0.019	–
$All\ events$	3.181	0.024	100.0
$No\ duplicates$	3.181	0.024	100.0
GRL	3.181	0.024	100.0
$Pass\ trigger$	2.184	0.021	68.7
$Detector\ DQ$	2.184	0.021	68.7
$Has\ PV$	2.184	0.021	68.7
$2\ loose\ photons$	1.894	0.019	59.5
$e - \gamma\ ambiguity$	1.891	0.019	59.5
$Trigger\ match$	1.885	0.019	59.3
$tight\ ID$	1.617	0.018	50.8
$isolation$	1.476	0.017	46.4
$rel.\ p_T\ cuts$	1.362	0.017	42.8
$m_{\gamma\gamma} \in [105, 160]\ GeV$	1.362	0.017	42.8
$2\ Cen\ Jets$	1.149	0.016	36.1
$b - tagging$	0.268	0.007	8.4
$bJet\ p_T\ Cuts$	0.257	0.007	8.1
$m_{bb}\ Cut$	0.157	0.006	4.9
$\gamma\gamma\ SR$	0.149	0.005	4.7

Table D.4: Cutflow for non-resonant $HH \to \gamma\gamma b\bar{b}$, 0 tag category (high mass selection)

Cuts	0 b-tag		
	Yield	Error	Efficiency
N_{xAOD}	3.147	0.027	–
N_{DxAOD}	3.147	0.019	–
$All\ events$	3.181	0.024	100.0
$No\ duplicates$	3.181	0.024	100.0
GRL	3.181	0.024	100.0
$Pass\ trigger$	2.184	0.021	68.7
$Detector\ DQ$	2.184	0.021	68.7
$Has\ PV$	2.184	0.021	68.7
$2\ loose\ photons$	1.894	0.019	59.5
$e - \gamma\ ambiguity$	1.891	0.019	59.5
$Trigger\ match$	1.885	0.019	59.3
$tight\ ID$	1.617	0.018	50.8
$isolation$	1.476	0.017	46.4
$rel.\ p_T\ cuts$	1.362	0.017	42.8
$m_{\gamma\gamma} \in [105, 160]\ GeV$	1.362	0.017	42.8
$2\ Cen\ Jets$	1.149	0.016	36.1
$b - tagging$	0.268	0.007	8.4
$bJet\ p_T\ Cuts$	0.121	0.004	3.8
$m_{bb}\ Cut$	0.062	0.003	2.0
$\gamma\gamma\ SR$	0.061	0.003	1.9

Table D.5: Background yield in the 1 b-tag region split by process for the non-resonant and resonant analyses. The non-resonant analysis requires passing of all cuts but the $m_{\gamma\gamma}$ SR constraint, and is denoted by the "Yield" column. The resonant analysis includes the additional $m_{\gamma\gamma}$ constraint and is denoted "SR Yield."

1-tag Category	Low Mass				High Mass			
Background	Yield	Error	SR Yield	Error	Yield	Error	SR Yield	Error
$\gamma\gamma$-continuum								
ggH	1.77	0.067	1.749	0.067	0.325	0.028	0.318	0.028
$VBFH$	0.158	0.011	0.151	0.011	0.027	0.004	0.026	0.004
W^-H	0.055	0.004	0.054	0.003	0.009	0.001	0.008	0.001
W^+H	0.073	0.005	0.071	0.005	0.016	0.002	0.015	0.002
ZH	0.409	0.01	0.4	0.009	0.087	0.004	0.084	0.004
$ggZH$	0.109	0.006	0.107	0.006	0.034	0.003	0.033	0.003
ttH	2.693	0.021	2.611	0.021	1.187	0.015	1.141	0.014
bbH (positive)	0.404	0.063	0.391	0.062	0.01	0.012	0.008	0.012
bbH (negative)	-0.024	0.003	-0.024	0.003	-0.001	0.001	-0.001	0.001

Table D.6: Background yield in the 0 b-tag region split by process for the non-resonant and resonant analyses. The non-resonant analysis requires passing of all cuts but the $m_{\gamma\gamma}$ SR constraint, and is denoted by the "Yield" column. The resonant analysis includes the additional $m_{\gamma\gamma}$ constraint and is denoted "SR Yield."

0-tag Category	Low Mass				High Mass			
Background	Yield	Error	SR Yield	Error	Yield	Error	SR Yield	Error
$\gamma\gamma$-continuum								
ggH	61.309	0.423	59.561	0.416	9.116	0.148	8.849	0.146
$VBFH$	6.921	0.064	6.725	0.063	1.269	0.027	1.239	0.027
W^-H	3.755	0.034	3.646	0.033	0.585	0.014	0.568	0.013
W^+H	5.259	0.05	5.103	0.049	0.857	0.02	0.834	0.019
ZH	5.519	0.036	5.361	0.036	1.173	0.018	1.136	0.018
$ggZH$	1.508	0.023	1.477	0.023	0.479	0.013	0.464	0.012
ttH	3.311	0.024	3.213	0.024	1.137	0.014	1.094	0.014
bbH (positive)	0.468	0.08	0.450	0.078	0.017	0.009	0.017	0.009
bbH (negative)	-0.039	0.003	-0.039	0.003	-0.001	0.000	-0.001	0.000

APPENDIX E

VARIABLES FOR VBF BDT

The full list of variables considered for the BDT used to define the VBF-enriched signal region is as follows:

VBF-targeted:

$\Delta\eta_{jj}$, m_{jj}, n_{jet}, $p_{T,\gamma\gamma}/m_{\gamma\gamma}$, $p_{T,jjb\bar{b}}$, $p_{T,jjbb}/p_{T,b\bar{b}}$[1].

$\gamma\gamma$-suppressing:

$m_{\gamma\gamma b\bar{b}}$, $\Delta R_{\gamma\gamma}$, $\Delta R_{b\bar{b}}$, $\Delta R_{\gamma\gamma,b\bar{b}}$, and the best b-tagging WP for the $H \to b\bar{b}$ jets.

ttH-suppressing:

H_T, $m_{jj,b1}$, $m_{jj,b2}$, $m_{jj,W1}$, $m_{jj,W2}$. Event shape variables (given in Reference [117]) - aplanority, aplanarity, sphericity, spherocity, sphericityT, planarity, circularity, planarFlow.

The pruning procedure described in Section 6.2 removed the following variables:

Pruning Step	Variables Removed
Variables with Pearson correlation $r_{xy} > 0.85$	$p_{T,jjb\bar{b}}$, $p_{T,\gamma\gamma}/m_{\gamma\gamma}$, sphericity, spherocity, planarity, circularity
Variables with $\sum_{p=0}^{3} r_{x,p} < 1.0$ for BDT-class probabilities, p	n_{jet}, $p_{T,jjbb}/p_{T,b\bar{b}}$, $\Delta R_{\gamma\gamma,b\bar{b}}$, aplanarity, H_T, $m_{jj,b1}$, $m_{jj,b2}$, $m_{jj,W1}$

Distributions for all considered variables are shown in Figure E.1 through E.5, in which each histogram has been normalized such that its integral is unity. Preselection and a 4-jet requirement has been applied.

[1]In these variables jj refers to the VBF-candidate jets, while $b\bar{b}$ refers to the $H \to b\bar{b}$ jets

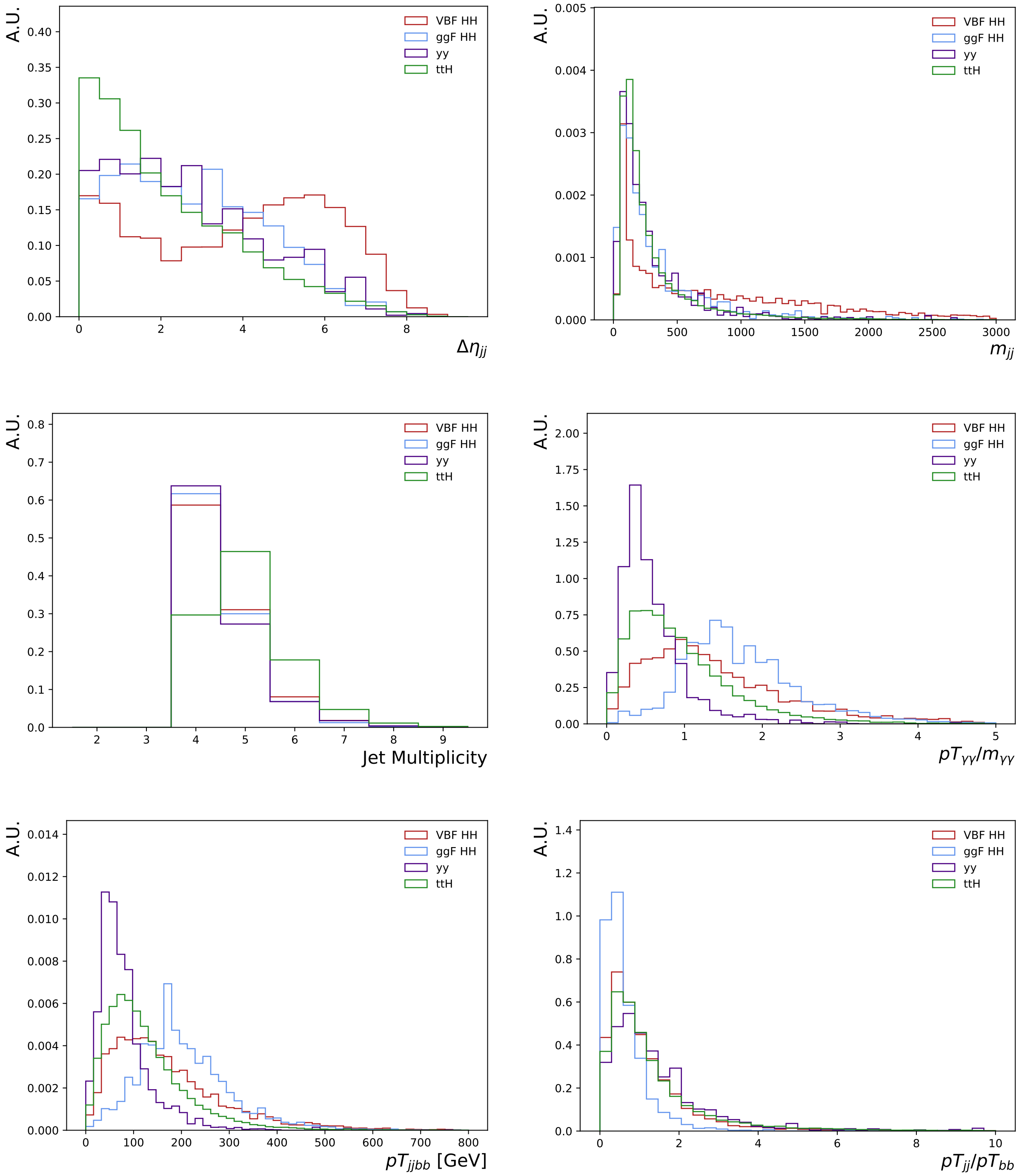


Figure E.1: Distributions of variables for targeting VBF HH production considered for use the VBF BDT

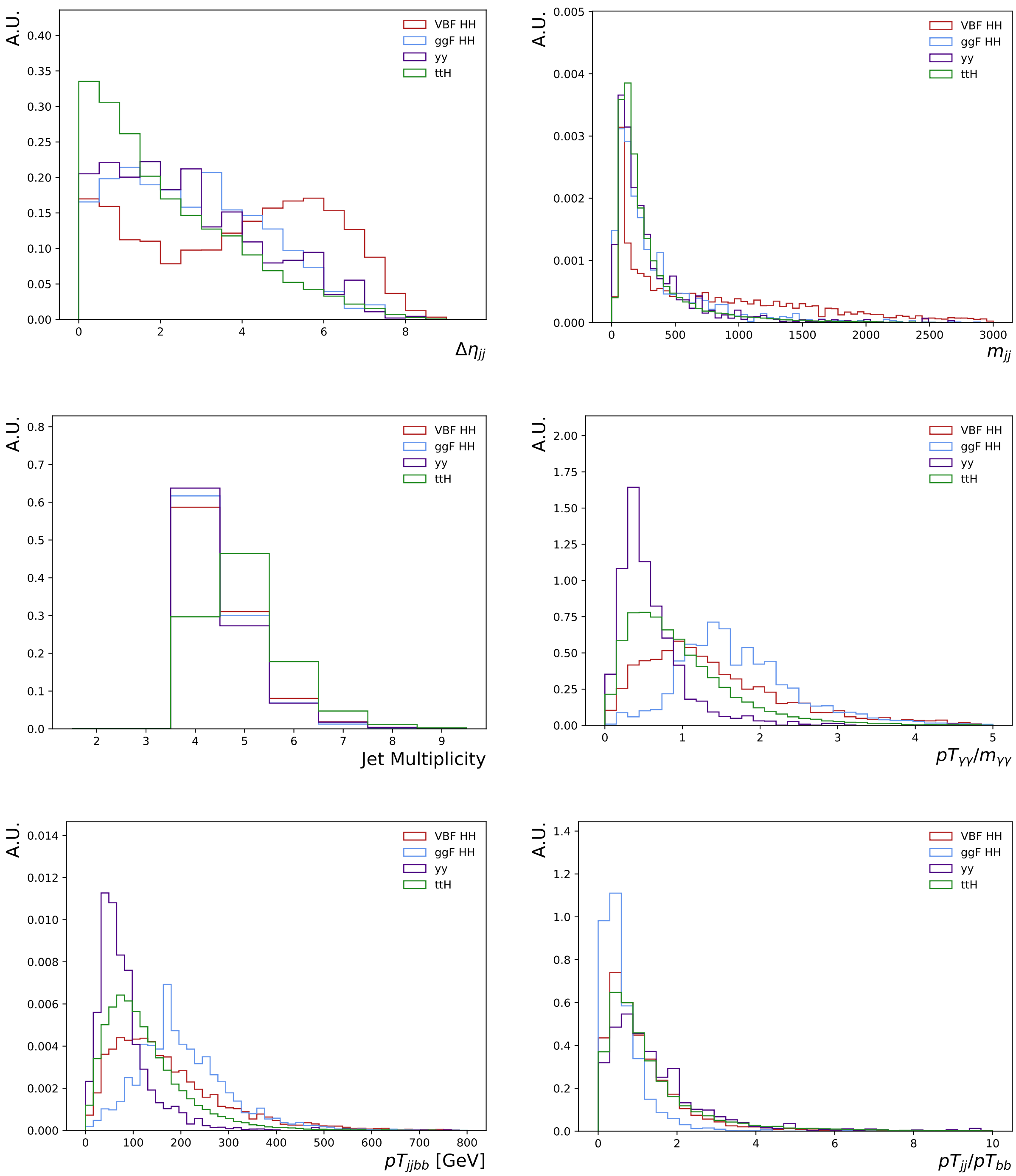


Figure E.2: Distributions of variables for targeting VBF HH production considered for use the VBF BDT

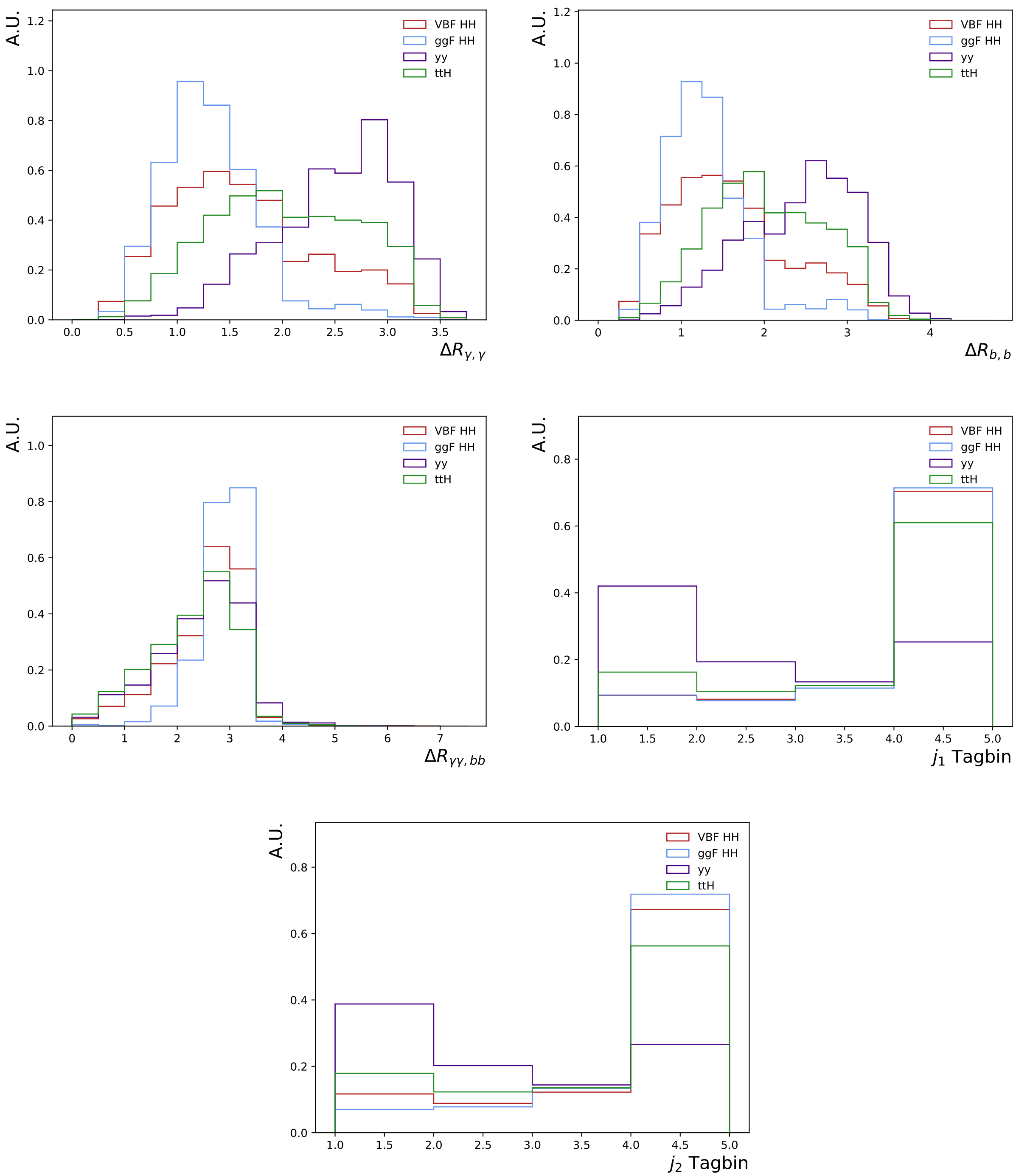


Figure E.3: Distributions of variables for suppressing the $\gamma\gamma$ background considered for use the VBF BDT

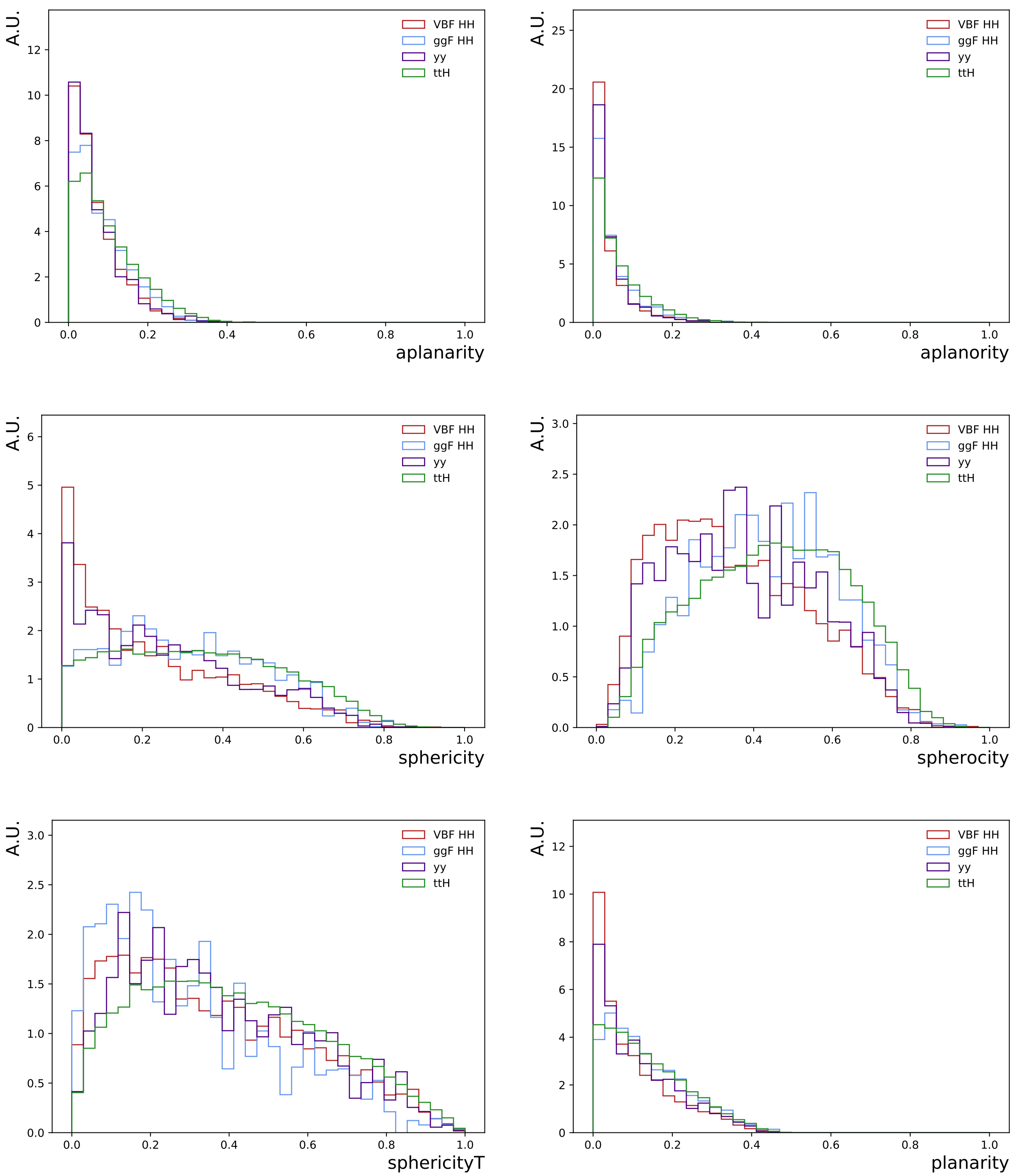


Figure E.4: Distributions of event shape for suppressing the ttH background considered for use the VBF BDT

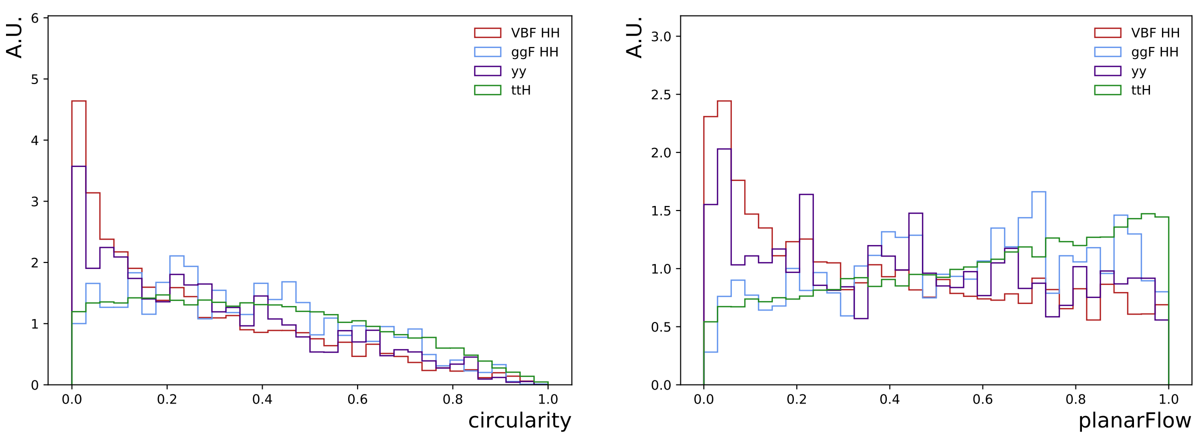


Figure E.4: Distributions of event shape for suppressing the ttH background considered for use the VBF BDT

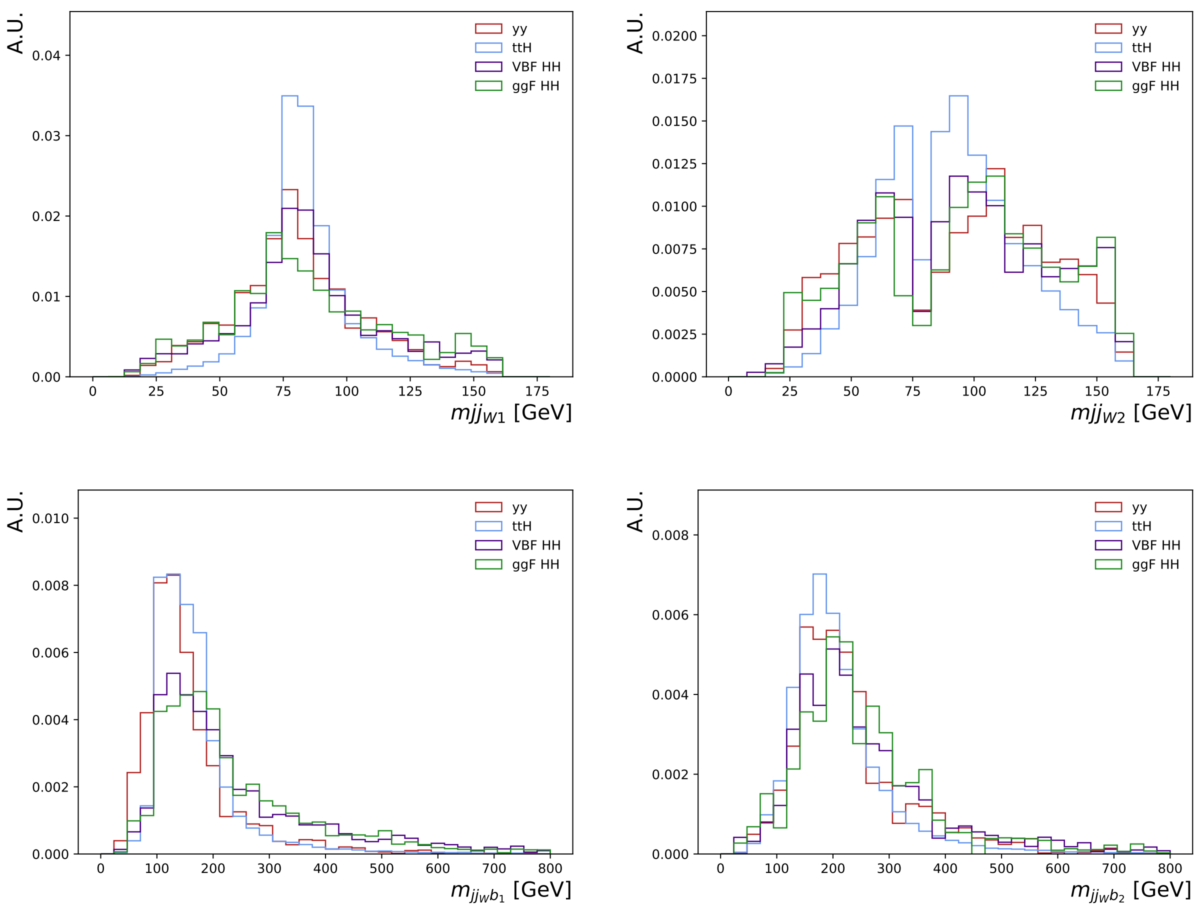


Figure E.5: Distributions of W and top mass variables for suppressing the ttH background considered for use the VBF BDT

APPENDIX F

VBF VARIED COUPLING SAMPLES

In addition to the SM VBF HH sample used in Chapter 6, samples have been produced for the various coupling values shown in Table F.1. These samples were generated using the same generators and showering as described in Section 6.1.1. The values of κ_λ,c_{2V}, and c_v are all 1 in the SM. The values selected for production aim to vary each coupling value enough that interpolation may be performed, and one value was produced near where limits could be expected to be set ($\kappa_\lambda = 10$ and $c_{2V} = 4$).

Table F.1: Grid of coupling values used in production of VBF HH samples.

κ_λ	c_{2V}	c_v
1	0	0.5
1	1	0.5
0	0	1
1	0	1
1	0.5	1
1	0	0.5
1	1	1
1	1.5	1
1	2	1
1	4	1
0	1	1
2	1	1
10	1	1
1	1	1.5

Relevant distributions for validation of these samples can be seen in Figure F.1. Plots are shown at parton level, with a basic jet selection applied to isolate the VBF jets similar to that done at reconstruction level. Jets are considered if they have $p_{\mathrm{T}} > 25$, then the $m_{b\bar{b}}$ pair is selected as the two jets with $m_{b\bar{b}}$ closest to 125 GeV. The VBF jets are selected as the highest m_{jj} pair after removing jets used in the $m_{b\bar{b}}$ pairing. Additional detail on the validation for these samples is outlined in Reference [127].

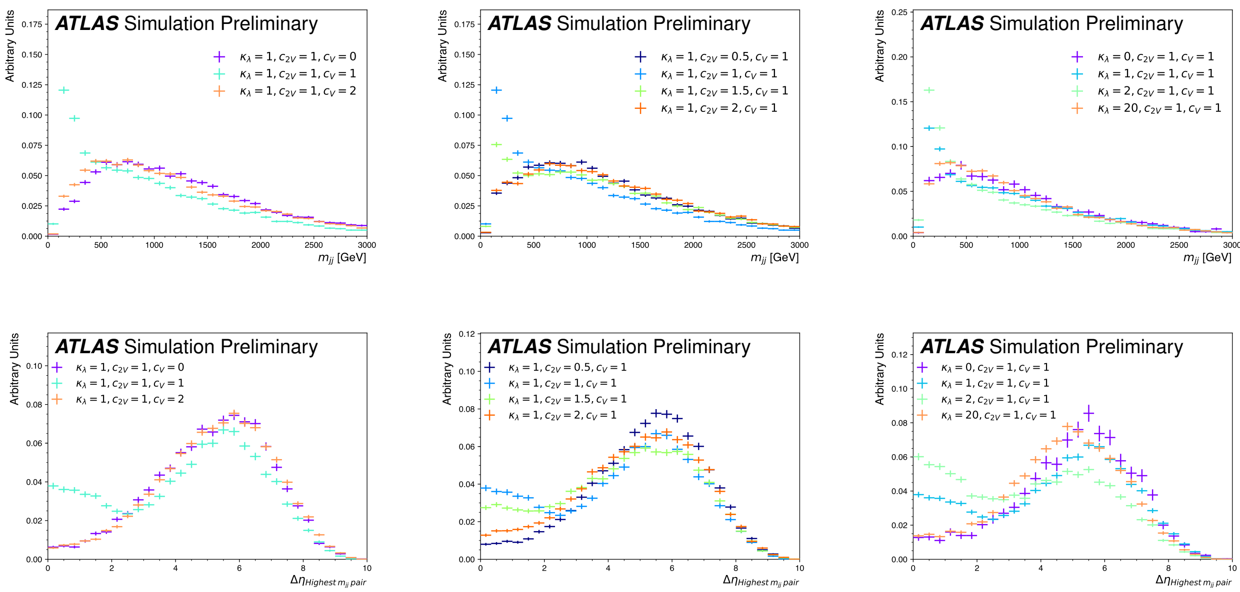


Figure F.1: Validation plots at parton level for the VBF HH signal samples. The highest m_{jj} pair is selected as a proxy for the VBF jets. The top row shows the invariant mass distribution of this system, the bottom depicts the $\Delta\eta$ between these jets. Left to right, these plots vary the c_V, c_{2V}, and κ_λ strength. The peak at low m_{jj} values and shoulder at low $\Delta\eta$ is a product of the VHH hadronic and Higgsstrahlung contributions in the samples.

BIBLIOGRAPHY

[1] M. Gangopadhyaya. *Indian atomism: history and sources.* Atlantic Highlands, N.J.: Humanities Press, 1981.

[2] H. Hartley. "John Dalton, F.R.S. (1766-1844) and the Atomic Theory-A Lecture to Commemorate his Bicentenary". *Proceedings of the Royal Society of London. Series B, Biological Sciences* 168.1013 (1967), pp. 335–359. ISSN: 00804649. URL: http://www.jstor.org/stable/75700.

[3] G. Zweig. *An SU_3 model for strong interaction symmetry and its breaking; Version 1.* Tech. rep. CERN-TH-401. Geneva: CERN, Jan. 1964. URL: https://cds.cern.ch/record/352337.

[4] G. Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV". *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91177-2.

[5] G. Arnison et al. "Experimental observation of lepton pairs of invariant mass around 95 GeV/c2 at the CERN SPS collider". *Physics Letters B* 126.5 (1983), pp. 398–410. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)90188-0.

[6] T. Y. Cao and J. T. Cushing. "Conceptual Developments of 20th Century Field Theories". 1997.

[7] Y. Fukuda et al. "Evidence for Oscillation of Atmospheric Neutrinos". *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. URL: https://link.aps.org/doi/10.1103/PhysRevLett.81.1562.

[8] F. Abe et al. "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab". *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631.

DOI: 10.1103/PhysRevLett.74.2626.
URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[9] M. L. Perl et al.
"Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation".
Phys. Rev. Lett. 35 (22 Dec. 1975), pp. 1489–1492.
DOI: 10.1103/PhysRevLett.35.1489.
URL: https://link.aps.org/doi/10.1103/PhysRevLett.35.1489.

[10] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC".
Physics Letters B 716.1 (2012), pp. 1–29. ISSN: 0370-2693.
DOI: https://doi.org/10.1016/j.physletb.2012.08.020. URL:
http://www.sciencedirect.com/science/article/pii/S037026931200857X.

[11] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". *Physics Letters B* 716.1 (2012), pp. 30–61.
ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021. URL:
http://www.sciencedirect.com/science/article/pii/S0370269312008581.

[12] C. N. Yang and R. L. Mills.
"Conservation of Isotopic Spin and Isotopic Gauge Invariance".
Phys. Rev. 96 (1 Oct. 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191.
URL: https://link.aps.org/doi/10.1103/PhysRev.96.191.

[13] E. Noether. "Invariante Variationsprobleme". ger.
Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse 1918 (1918), pp. 235–257.
URL: http://eudml.org/doc/59024.

[14] Particle Data Group. "Review of Particle Physics". *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[15] S. L. Glashow. "The renormalizability of vector meson interactions". *Nucl. Phys.* 10 (1959), pp. 107–117. DOI: 10.1016/0029-5582(59)90196-8.

[16] S. Weinberg. "A Model of Leptons". *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[17] A. Salam and J. C. Ward. "Weak and electromagnetic interactions". *Il Nuovo Cimento (1955-1965)* 11.4 (Feb. 1959), pp. 568–577. ISSN: 1827-6121. DOI: 10.1007/BF02726525. URL: https://doi.org/10.1007/BF02726525.

[18] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector" (2016). DOI: 10.2172/1345634,10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[19] ATLAS Collaboration. *Constraint of the Higgs boson self-coupling from Higgs boson differential production and decay measurements*. Tech. rep. ATL-PHYS-PUB-2019-009. Geneva: CERN, Mar. 2019. URL: https://cds.cern.ch/record/2667570.

[20] S. Dawson, S. Dittmaier, and M. Spira. "Neutral Higgs boson pair production at hadron colliders: QCD corrections". *Phys. Rev.* D58 (1998), p. 115012. DOI: 10.1103/PhysRevD.58.115012. arXiv: hep-ph/9805244 [hep-ph].

[21] ATLAS Collaboration. *Search for the $HH \to b\bar{b}b\bar{b}$ process via vector boson fusion production using proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2019-030. Geneva: CERN, July 2019. URL: http://cds.cern.ch/record/2682801.

[22] ATLAS Collaboration.
Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment.
Tech. rep. ATLAS-CONF-2018-031. Geneva: CERN, July 2018.
URL: http://cds.cern.ch/record/2629412.

[23] E. Arganda, C. Garcia-Garcia, and M. J. Herrero. "Probing the Higgs self-coupling through double Higgs production in vector boson scattering at the LHC".
Nucl. Phys. B945 (2019). [Nucl. Phys.B945,114687(2019)], p. 114687.
DOI: 10.1016/j.nuclphysb.2019.114687. arXiv: 1807.09736 [hep-ph].

[24] R. Frederix et al.
"Higgs pair production at the LHC with NLO and parton-shower effects".
Phys. Lett. B732 (2014), pp. 142–149. DOI: 10.1016/j.physletb.2014.03.026.
arXiv: 1401.7340 [hep-ph].

[25] F. Bishara, R. Contino, and J. Rojo.
"Higgs pair production in vector-boson fusion at the LHC and beyond".
Eur. Phys. J. C77.7 (2017), p. 481. DOI: 10.1140/epjc/s10052-017-5037-9.
arXiv: 1611.03860 [hep-ph].

[26] T. D. Lee. "A Theory of Spontaneous T Violation".
Phys. Rev. D 8 (4 Aug. 1973), pp. 1226–1239. DOI: 10.1103/PhysRevD.8.1226.
URL: https://link.aps.org/doi/10.1103/PhysRevD.8.1226.

[27] G. Branco et al. "Theory and phenomenology of two-Higgs-doublet models".
Physics Reports 516.1 (2012). Theory and phenomenology of two-Higgs-doublet models, pp. 1–102. ISSN: 0370-1573.
DOI: https://doi.org/10.1016/j.physrep.2012.02.002. URL:
http://www.sciencedirect.com/science/article/pii/S0370157312000695.

[28] S. Dimopoulos and H. Georgi. "Softly broken supersymmetry and SU(5)". *Nuclear Physics B* 193.1 (1981), pp. 150–162. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(81)90522-8. URL: http://www.sciencedirect.com/science/article/pii/0550321381905228.

[29] A. Djouadi et al. "Fully covering the MSSM Higgs sector at the LHC". *Journal of High Energy Physics* 2015.6 (2015), p. 168. DOI: 10.1007/JHEP06(2015)168. URL: https://doi.org/10.1007/JHEP06(2015)168.

[30] A. Djouadi and J. Quevillon. "The MSSM Higgs sector at a high MSUSY: reopening the low tan β regime and heavy Higgs searches". *Journal of High Energy Physics* 2013.10 (2013), p. 28. DOI: 10.1007/JHEP10(2013)028. URL: https://doi.org/10.1007/JHEP10(2013)028.

[31] Z. Chacko et al. "Natural little hierarchy from a partially goldstone twin Higgs". *Journal of High Energy Physics* 2006.01 (Jan. 2006), pp. 126–126. DOI: 10.1088/1126-6708/2006/01/126. URL: https://doi.org/10.1088%2F1126-6708%2F2006%2F01%2F126.

[32] W. Ma, C.-X. Yue, and Y.-Z. Wang. "Pair production of neutral Higgs bosons from the left-right twin Higgs model at the ILC and LHC". *Phys. Rev. D* 79 (9 May 2009), p. 095010. DOI: 10.1103/PhysRevD.79.095010. URL: https://link.aps.org/doi/10.1103/PhysRevD.79.095010.

[33] J. Mrazek et al. "The other natural two Higgs doublet model". *Nuclear Physics B* 853.1 (2011), pp. 1–48. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2011.07.008. URL: http://www.sciencedirect.com/science/article/pii/S0550321311003804.

[34] G. Brooijmans et al.
"Les Houches 2013: Physics at TeV Colliders: New Physics Working Group Report" (2014). arXiv: 1405.1617 [hep-ph].

[35] J. Alison et al. "Higgs boson pair production at colliders: status and perspectives". *Double Higgs Production at Colliders Batavia, IL, USA, September 4, 2018-9, 2019.* Ed. by B. Di Micco et al. 2019. arXiv: 1910.00012 [hep-ph].

[36] L. Evans and P. Bryant. "LHC Machine". *JINST* 3 (2008), S08001.
DOI: 10.1088/1748-0221/3/08/S08001.

[37] The LHCb Collaboration. "The LHCb Detector at the LHC".
Journal of Instrumentation 3.08 (Aug. 2008), S08005–S08005.
DOI: 10.1088/1748-0221/3/08/s08005.
URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08005.

[38] The ALICE Collaboration. "The ALICE experiment at the CERN LHC".
Journal of Instrumentation 3.08 (Aug. 2008), S08002–S08002.
DOI: 10.1088/1748-0221/3/08/s08002.
URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08002.

[39] The ATLAS Collaboration.
"The ATLAS Experiment at the CERN Large Hadron Collider".
Journal of Instrumentation 3.08 (Aug. 2008), S08003–S08003.
DOI: 10.1088/1748-0221/3/08/s08003.
URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08003.

[40] The CMS Collaboration. "The CMS Experiment at the CERN LHC".
JINST 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[41] *ATLAS magnet system: Technical Design Report, 1.* Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/338080.

[42] J. Goodson. 2012. URL: http://www.jetgoodson.com/images/thesisImages/magnetSystems.png.

[43] *ATLAS central solenoid: Technical Design Report.* Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331067.

[44] J. P. Badiou et al. *ATLAS barrel toroid: Technical Design Report.* Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331065.

[45] The ATLAS Collaboration. *ATLAS end-cap toroids: Technical Design Report.* Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331066.

[46] The ATLAS Collaboration. *ATLAS inner detector: Technical Design Report, 1.* Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331063.

[47] M. Capeans et al. *ATLAS Insertable B-Layer Technical Design Report.* Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. Sept. 2010. URL: https://cds.cern.ch/record/1291633.

[48] The ATLAS Collaboration. "ATLAS pixel detector electronics and sensors". *Journal of Instrumentation* 3.07 (July 2008), P07007–P07007. DOI: 10.1088/1748-0221/3/07/p07007. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F07%2Fp07007.

[49] *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV.* Tech. rep. ATL-PHYS-PUB-2015-018. Geneva: CERN, July 2015. URL: http://cds.cern.ch/record/2037683.

[50] *ATLAS liquid-argon calorimeter: Technical Design Report.* Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331061.

[51] K. Kleinknecht. *Detectors for Particle Radiation; 2nd ed.* Cambridge: Cambridge Univ. Press, 1998. URL: https://cds.cern.ch/record/366030.

[52] *ATLAS tile calorimeter: Technical Design Report.* Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331062.

[53] *ATLAS muon spectrometer: Technical Design Report.* Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331068.

[54] "Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s} = 8$ TeV". *Eur. Phys. J.* C75 (2015), p. 120. DOI: 10.1140/epjc/s10052-015-3325-9. arXiv: 1408.3179 [hep-ex].

[55] A. Outreach. "ATLAS Fact Sheet : To raise awareness of the ATLAS detector and collaboration on the LHC". 2010. URL: https://cds.cern.ch/record/1457044.

[56] "ATLAS high-level trigger, data acquisition and controls: Technical design report" (2003).

[57] W. Panduro Vazquez. "The ATLAS Data Acquisition system in LHC Run 2". *J. Phys. Conf. Ser.* 898.3 (2017), p. 032017. DOI: 10.1088/1742-6596/898/3/032017.

[58] E. Boos et al. “Generic User Process Interface for Event Generators”. *Physics at TeV colliders. Proceedings, Euro Summer School, Les Houches, France, May 21-June 1, 2001.* 2001. arXiv: hep-ph/0109068 [hep-ph]. URL: http://lss.fnal.gov/archive/2001/conf/fermilab-conf-01-496-t.pdf.

[59] M. Bahr et al. “Herwig++ Physics and Manual”. *Eur. Phys. J.* C58 (2008), pp. 639–707. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[60] J. Alwall et al. “MadGraph 5 : Going Beyond”. *JHEP* 06 (2011), p. 128. DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph].

[61] T. Sjöstrand et al. “An Introduction to PYTHIA 8.2”. *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[62] S. Frixione, P. Nason, and C. Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[63] E. Bothmann et al. “Event Generation with Sherpa 2.2”. *SciPost Phys.* 7.3 (2019), p. 034. DOI: 10.21468/SciPostPhys.7.3.034. arXiv: 1905.09127 [hep-ph].

[64] B. Andersson et al. “Parton fragmentation and string dynamics”. *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7. URL: http://www.sciencedirect.com/science/article/pii/0370157383900807.

[65] T. Sjostrand and P. Z. Skands. "Multiple interactions and the structure of beam remnants". *JHEP* 03 (2004), p. 053. DOI: 10.1088/1126-6708/2004/03/053. arXiv: hep-ph/0402078 [hep-ph].

[66] The ATLAS Collaboration. *ATLAS Pythia 8 Tunes to 7 TeV Data.* Tech. rep. ATL-PHYS-PUB-2014-021. Geneva: CERN, Nov. 2014. URL: https://cds.cern.ch/record/1966419.

[67] S. Gieseke, C. Röhr, and A. Siódmok. "Colour reconnections in Herwig++". *The European Physical Journal C* 72.11 (2012), p. 2225. DOI: 10.1140/epjc/s10052-012-2225-5. URL: https://doi.org/10.1140/epjc/s10052-012-2225-5.

[68] H.-L. Lai et al. "New parton distributions for collider physics". *Phys. Rev. D* 82 (7 Oct. 2010), p. 074024. DOI: 10.1103/PhysRevD.82.074024. URL: https://link.aps.org/doi/10.1103/PhysRevD.82.074024.

[69] J. Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". *The European Physical Journal C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. URL: https://doi.org/10.1140/epjc/s10052-016-4018-8.

[70] L. A. Harland-Lang et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". *Eur. Phys. J.* C75.5 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[71] S. Höche. "Introduction to parton-shower event generators". *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*. 2015, pp. 235–295. DOI: 10.1142/9789814678766_0005. arXiv: 1411.4085 [hep-ph].

[72] S. Agostinelli et al. “GEANT4: A Simulation toolkit”. *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[73] T. Yamanaka. “The ATLAS calorimeter simulation FastCaloSim”. *J. Phys. Conf. Ser.* 331 (2011), p. 032053. DOI: 10.1088/1742-6596/331/3/032053.

[74] The ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data” (2019). arXiv: 1908.00005 [hep-ex].

[75] W. Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance.* Tech. rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003. Geneva: CERN, Apr. 2008. URL: https://cds.cern.ch/record/1099735.

[76] The ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. *Eur. Phys. J.* C77 (2017), p. 490. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex].

[77] T. G. Cornelissen et al. “The global χ^2 track fitter in ATLAS”. *Journal of Physics: Conference Series* 119.3 (July 2008), p. 032013. DOI: 10.1088/1742-6596/119/3/032013. URL: https://doi.org/10.1088%2F1742-6596%2F119%2F3%2F032013.

[78] *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung.* Tech. rep. ATLAS-CONF-2012-047. Geneva: CERN, May 2012. URL: https://cds.cern.ch/record/1449796.

[79] R. Frühwirth. “Application of Kalman filtering to track and vertex fitting”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450.

ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(87)90887-4. URL: http://www.sciencedirect.com/science/article/pii/0168900287908874.

[80] J. E. Huth et al. "Toward a standardization of jet definitions". FERMILAB-CONF-90-249-E (Dec. 1990), 7 p. URL: https://cds.cern.ch/record/217490.

[81] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-ktjet clustering algorithm". *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. DOI: 10.1088/1126-6708/2008/04/063. URL: https://doi.org/10.1088%2F1126-6708%2F2008%2F04%2F063.

[82] *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. Tech. rep. ATL-PHYS-PUB-2017-013. Geneva: CERN, July 2017. URL: https://cds.cern.ch/record/2273281.

[83] J. Illingworth and J. Kittler. "A survey of the hough transform". *Computer Vision, Graphics, and Image Processing* 44.1 (1988), pp. 87–116. ISSN: 0734-189X. DOI: https://doi.org/10.1016/S0734-189X(88)80033-1. URL: http://www.sciencedirect.com/science/article/pii/S0734189X88800331.

[84] The ATLAS Collaboration. "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV". *Eur. Phys. J.* C76.5 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].

[85] J. Saxon. "Discovery of the Higgs Boson, Measurements of its Production, and a Search for Higgs Boson Pair Production". Presented 13 06 2014. July 2014. URL: https://cds.cern.ch/record/1746004.

[86] The ATLAS Collaboration. "Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data". *Eur. Phys. J.* C76.12 (2016), p. 666. DOI: 10.1140/epjc/s10052-016-4507-9. arXiv: 1606.01813 [hep-ex].

[87] "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". *The European Physical Journal C* 77.7 (July 2017), p. 490. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5004-5. URL: https://doi.org/10.1140/epjc/s10052-017-5004-5.

[88] T. Barillari et al. *Local Hadronic Calibration*. Tech. rep. ATL-LARG-PUB-2009-001-2. ATL-COM-LARG-2008-006. ATL-LARG-PUB-2009-001. Geneva: CERN, June 2008. URL: https://cds.cern.ch/record/1112035.

[89] G. Aad et al. "Measurement of isolated-photon pair production in *pp* collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". *JHEP* 01 (2013), p. 086. DOI: 10.1007/JHEP01(2013)086. arXiv: 1211.1913 [hep-ex].

[90] A. Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis" (2007). arXiv: physics/0703039 [physics.data-an].

[91] J. Sadeghi, S. Sadeghi, and S. T. A. Niaki. "Optimizing a hybrid vendor-managed inventory and transportation problem with fuzzy demand: An improved particle swarm optimization algorithm". *Information Sciences* 272 (2014), pp. 126–144. ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2014.02.075. URL: http://www.sciencedirect.com/science/article/pii/S0020025514001819.

[92] N. Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting".

Journal of Machine Learning Research 15 (2014), pp. 1929–1958.
URL: http://jmlr.org/papers/v15/srivastava14a.html.

[93] The ATLAS Collaboration. "Search For Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State using pp Collision Data at $\sqrt{s} = 8$ TeV from the ATLAS Detector". *Phys.Rev.Lett.* 114.8 (2015), p. 081802. DOI: 10.1103/PhysRevLett.114.081802. arXiv: 1406.5053 [hep-ex].

[94] The HSG1 working group. "Search for resonant and enhanced non-resonant di-Higgs production in the gamma gamma bbbar channel with 3.2 fb-1 of data at 13 TeV" ().

[95] The ATLAS Collaboration. "Performance of the ATLAS Trigger System in 2015". *Eur. Phys. J.* C77.5 (2017), p. 317. DOI: 10.1140/epjc/s10052-017-4852-3. arXiv: 1611.09661 [hep-ex].

[96] F. Maltoni, E. Vryonidou, and M. Zaro. "Top-quark mass effects in double and triple Higgs production in gluon-gluon fusion at NLO". *Journal of High Energy Physics* 2014.11 (2014), p. 79. DOI: 10.1007/JHEP11(2014)079. URL: https://doi.org/10.1007/JHEP11(2014)079.

[97] S. Frixione and B. R. Webber. "Matching NLO QCD computations and parton shower simulations". *Journal of High Energy Physics* 2002.06 (June 2002), pp. 029–029. DOI: 10.1088/1126-6708/2002/06/029. URL: https://doi.org/10.1088%2F1126-6708%2F2002%2F06%2F029.

[98] J. Grigo, J. Hoff, and M. Steinhauser. "Higgs boson pair production: Top quark mass effects at NLO and NNLO". *Nuclear Physics B* 900 (2015), pp. 412–430. ISSN: 0550-3213.

DOI: https://doi.org/10.1016/j.nuclphysb.2015.09.012. URL: http://www.sciencedirect.com/science/article/pii/S0550321315003314.

[99] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector" (2016). DOI: 10.2172/1345634,10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[100] The ATLAS Collaboration. "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV". *Eur. Phys. J. C* 79.arXiv:1902.04655. 8 (Feb. 2019), 639. 40 p. DOI: 10.1140/epjc/s10052-019-7140-6. URL: http://cds.cern.ch/record/2657964.

[101] M. Cacciari, G. P. Salam, and G. Soyez. "FastJet user manual". *The European Physical Journal C* 72.3 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. URL: https://doi.org/10.1140/epjc/s10052-012-1896-2.

[102] The ATLAS Collaboration. "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector". *Eur. Phys. J.* C76.11 (2016), p. 581. DOI: 10.1140/epjc/s10052-016-4395-z. arXiv: 1510.03823 [hep-ex].

[103] *Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector*. Tech. rep. ATLAS-CONF-2015-029. Geneva: CERN, July 2015. URL: https://cds.cern.ch/record/2037702.

[104] S. Rettie. *Muon identification and performance in the ATLAS experiment*. Tech. rep. ATL-PHYS-PROC-2018-052. Geneva: CERN, June 2018. DOI: 10.22323/1.316.0097. URL: https://cds.cern.ch/record/2626330.

[105] The ATLAS Collaboration.
"Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector".
Phys. Rev. D98 (2018), p. 052005. DOI: `10.1103/PhysRevD.98.052005`.
arXiv: `1802.04146 [hep-ex]`.

[106] G. Aad et al. "Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector".
Phys. Rev. D83 (2011), p. 052005. DOI: `10.1103/PhysRevD.83.052005`.
arXiv: `1012.4389 [hep-ex]`.

[107] The ATLAS Collaboration.
"Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector".
Phys. Rev. D98 (2018), p. 052005. DOI: `10.1103/PhysRevD.98.052005`.
arXiv: `1802.04146 [hep-ex]`.

[108] The ATLAS Collaboration.
"Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector".
Phys. Rev. D90.11 (2014), p. 112015. DOI: `10.1103/PhysRevD.90.112015`.
arXiv: `1408.7084 [hep-ex]`.

[109] The ATLAS Collaboration. "Study of heavy-flavor quarks produced in association with top-quark pairs at $\sqrt{s} = 7$ TeV using the ATLAS detector".
Phys. Rev. D 89 (7 Apr. 2014), p. 072012. DOI: `10.1103/PhysRevD.89.072012`.
URL: `https://link.aps.org/doi/10.1103/PhysRevD.89.072012`.

[110] The ATLAS Collaboration.
"Measurement of the cross-section for W boson production in association with b-jets in

pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector".
Journal of High Energy Physics 2013.6 (June 2013), p. 84. ISSN: 1029-8479.
DOI: 10.1007/JHEP06(2013)084.
URL: https://doi.org/10.1007/JHEP06(2013)084.

[111] S. Borowka et al. "Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading Order with Full Top-Quark Mass Dependence".
Phys. Rev. Lett. 117 (1 June 2016), p. 012001.
DOI: 10.1103/PhysRevLett.117.012001.
URL: https://link.aps.org/doi/10.1103/PhysRevLett.117.012001.

[112] The ATLAS Collaboration. "Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC". *Eur. Phys. J.* C76.12 (2016), p. 653.
DOI: 10.1140/epjc/s10052-016-4466-1. arXiv: 1608.03953 [hep-ex].

[113] The ATLAS Collaboration. *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV.*
Tech. rep. ATL-PHYS-PUB-2015-015. Geneva: CERN, July 2015.
URL: https://cds.cern.ch/record/2037613.

[114] W. Verkerke and D. P. Kirkby. "The RooFit toolkit for data modeling".
eConf C0303241 (2003). [,186(2003)], MOLT007.
arXiv: physics/0306116 [physics].

[115] R. D. Ball et al. "Parton distributions with LHC data".
Nucl. Phys. B867 (2013), pp. 244–289. DOI: 10.1016/j.nuclphysb.2012.10.003.
arXiv: 1207.1303 [hep-ph].

[116] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics".
Eur. Phys. J. C71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554.

DOI: 10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z. arXiv: 1007.1727 [physics.data-an].

[117] The ATLAS Collaboration. "Measurement of event shapes at large momentum transfer with the ATLAS detector in pp collisions at $\sqrt{s} = 7\,\text{TeV}$". *Eur. Phys. J. C* 72 (2012), p. 2211. DOI: 10.1140/epjc/s10052-012-2211-y. arXiv: 1206.2135 [hep-ex].

[118] L. G. Almeida et al. "Substructure of high-p_T Jets at the LHC". *Phys. Rev.* D79 (2009), p. 074017. DOI: 10.1103/PhysRevD.79.074017. arXiv: 0807.0234 [hep-ph].

[119] O. Behnke et al., eds. *Data analysis in high energy physics.* Weinheim, Germany: Wiley-VCH, 2013. ISBN: 9783527410583, 9783527653447, 9783527653430. URL: http://www.wiley-vch.de/publish/dt/books/ISBN3-527-41058-9.

[120] Y. Freund and R. E. Schapire. "A desicion-theoretic generalization of on-line learning and an application to boosting". *Computational Learning Theory.* Ed. by P. Vitányi. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995, pp. 23–37. ISBN: 978-3-540-49195-8.

[121] H. Drucker and C. Cortes. "Boosting Decision Trees". *NIPS*. 1995.

[122] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework". *Nucl. Instrum. Meth.* A389 (1997), pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

[123] T. Chen and C. Guestrin. "XGBoost: A Scalable Tree Boosting System". *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge*

Discovery and Data Mining. KDD ’16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. URL: http://doi.acm.org/10.1145/2939672.2939785.

[124] K. Hornik, M. Stinchcombe, and H. White. “Multilayer feedforward networks are universal approximators”. English (US). *Neural Networks* 2.5 (Jan. 1989), pp. 359–366. ISSN: 0893-6080. DOI: 10.1016/0893-6080(89)90020-8.

[125] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. “Learning representations by back-propagating errors”. *Nature* 323 (1986), pp. 533–536.

[126] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.* Software available from tensorflow.org. 2015. URL: http://tensorflow.org/.

[127] X. Sun et al. *Validation of signal Monte Carlo event generation in searches for Higgs boson pairs with the ATLAS detector*. Tech. rep. ATL-COM-PHYS-2019-120. Geneva: CERN, Feb. 2019. URL: https://cds.cern.ch/record/2660207.