UNIVERSITY OF CALIFORNIA SAN DIEGO

Observation of Higgs boson production in association with a top quark-antiquark pair in the diphoton decay channel

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Physics

by

Samuel James May

Committee in charge:

Professor Avraham Yagil, Chair
Professor Garrison Cottrell
Professor Aneesh Manohar
Professor Julian McAuley
Professor Frank Würthwein

2020

The dissertation of Samuel James May is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2020

DEDICATION

To my mom and dad, who have always encouraged my curiosity and supported my education.

EPIGRAPH

Sometimes science is more art than science.

—Rick Sanchez

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Many people, more than I can list here, have helped me in completing my PhD. I am grateful to everyone who has helped me along the way – my work would not have been possible without you.

My mom and dad deserve to be thanked first. They encouraged my education throughout my entire life and made me feel that my interests were worth pursuing. My cousin, Christopher Betancourt, also deserves to be thanked for getting me interested in particle physics in the first place.

Second, I would like to thank my undergraduate research advisors at UCLA: Jon Aurnou and Bob Cousins. Jon was my first research advisor and taught me how to think like a researcher. Bob gave me my introduction to particle physics, both in the classroom, where he taught the first course on particle physics I took, and in a research environment, where he mentored me on a project studying the statistical method of unfolding in the context of high energy physics (and taught me a great deal of statistics along the way).

Third, I would like to thank my thesis advisors, Frank Wuerthwein and Avi Yagil. Frank encouraged me to explore a wide range of projects, including ones not necessarily related to physics. Avi consistently challenged me to be a better physicist, and I learned more because of it. Claudio Campagnari also deserves to be thanked along with Frank and Avi, as he acted as my unofficial third advisor. Frank, Avi, and Claudio each helped me develop my mind as a physicist through a variety of projects in which they mentored me.

Fourth, I would like to thank "Surf and Turf" (SNT) group from UCSD, UCSB, Fermilab, Boston University, and the University of Nebraska. Everyone in SNT during my time at UCSD has helped me in some way, even if just indirectly. A few people deserve special thanks. Vince Welke, Dan Klein, Bobak Hashemi, and Mark Derdzinski helped me get started on my first project in grad school. Slava Krutelyov never failed to answer my questions about the fine details of CMSSW software. Philip Chang mentored me on my first foray into machine learning (and was always willing to discuss interesting papers & ambitious new ideas). I would also like to especially thank Nick Amin: not only was Nick the de facto expert on every SNT repository, he always went above and beyond when answering my many questions.

Last, I would like to thank those who specifically mentored me on my thesis project. Dominick Olivito and Giovanni Zevi Della Porta taught me how a physics analysis was done in CMS. They always made time to answer my questions, often taking time out of their evenings to do so. Bennett Marsh

helped in the initial stages of this project, and helped me in understanding the code used by the CMS Higgs to Gamma Gamma group. Frank Golf provided valuable insights and interesting discussions in the initial stages of this project. Avi and Claudio mentored me on this project over the course of two and a half years. The entire CMS Higgs to Gamma Gamma group must be thanked as well, as they not only provided feedback on my work, but contributed many of the common tools shared by all Higgs to Gamma Gamma analyses in CMS, without which this analysis would not have been possible. In particular, Shervin Nourbaksh, Seth Zenz, Julie Malcles, and Ed Scott, who acted as conveners of the Higgs to Gamma Gamma group during the time I worked on this thesis, provided valuable insight on my work. Meng Xiao, Andrei Gritsan, and Mehmet Ozgur Sahin developed the CP measurement component of the analysis and provided feedback on the analysis as a whole. Hualin Mei deserves to be thanked in particular: Hualin served as both a mentor and an (exceptionally reliable and hard-working) teammate to me in developing this analysis, often staying up late to answer my questions or discuss the finer details of the analysis.

Chapter 4 describes the Large Hadron Collider and Compact Muon Solenoid detector. The figures shown in Chapter 4 are taken from the following results: "Performance and track-based alignment of the Phase-1 upgraded CMS pixel detector", *CMS-CR-2017-256* (2017), "CMS Luminosity – Public Results", *https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults* (2020), "The CMS Experiment at the CERN LHC", *Journal of Instrumentation* (2008), "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *Journal of Instrumentation* (2014), and "The CMS trigger system", *Journal of Instrumentation* (2017), and were produced by other members of the CMS Collaboration.

Chapter 5 describes the event reconstruction pipeline in CMS with a particular focus on photon reconstruction. The figures shown in Chapter 5 are taken from the following publications: "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV", *Journal of Instrumentation* (2015), "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *Journal of Instrumentation* (2017), "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *Journal of Instrumentation* (2018), and "Measurements of Higgs boson properties in the diphoton decay channel at $\sqrt{s} = 13$ TeV", *CMS-PAS-HIG-19-015* (2020), and were produced by other members of the CMS Collaboration, particularly those involved in the CMS Higgs to Gamma Gamma working group.

Chapter 6 describes the $t\bar{t}H$ analysis documented in "Measurements of $t\bar{t}H$ production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel" *Phys. Rev. Lett.* 125 (2020), with a focus on the aspects to which I contributed most directly, but relies heavily from the work of other members of the CMS Collaboration and the CMS Higgs to Gamma Gamma working group, without whom this analysis would not have been possible. My primary individual contributions to this work included the following: implementation of the data-driven description of multi-jet and γ + jet backgrounds, studies of agreement between data and simulation, training and optimization of the deep neural networks and boosted decision trees used for signal region definition, and implementation of some systematic uncertainties. Figures 6.17,6.18, and 6.19 show the results of the signal and background models for the $t\bar{t}H$ analysis, and were produced by Hualin Mei. Figure 6.20 shows the impact of systematic uncertainties on the measurement of $\mu_{t\bar{t}H}$, and was produced by Hualin Mei. Figures 6.21 and 6.22 show the observed results of the $t\bar{t}H$ analysis, and were also produced by Hualin Mei. Figure 6.23 shows the result of the $t\bar{t}H$ CP measurement and was produced by Meng Xiao.

VITA

2016	B. S. in Physics, University of California, Los Angeles
2016-2017	Graduate Teaching Assistant, University of California San Diego
2018	M. S. in Physics, University of California San Diego
2020	Ph. D. in Physics, University of California San Diego

PUBLICATIONS

CMS Collaboration, "Measurements of Higgs boson properties in the diphoton decay channel at $\sqrt{s}$ = 13 TeV", *CMS-PAS-HIG-19-015*, 2020.

CMS Collaboration, "Measurements of $t\bar{t}H$ production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel", *Phys.Rev.Lett.* 125 (2020) 6, 061801.

Robert D. Cousins, **Samuel May**, and Yipeng Sun, "Should unfolded histograms be used to test hypotheses?", *arXiv:1607.07038*, 2016.

ABSTRACT OF THE DISSERTATION

Observation of Higgs boson production in association with a top quark-antiquark pair in the diphoton decay channel

by

Samuel James May

Doctor of Philosophy in Physics

University of California San Diego, 2020

Professor Avraham Yagil, Chair

This dissertation presents the first observation of Higgs boson production in association with a top quark-antiquark pair in the diphoton decay channel, with a significance of 6.6 standard deviations. The measurement is performed with a dataset of 13 TeV proton-proton collisions recorded by the Compact Muon Solenoid (CMS) detector at the CERN Large Hadron Collider (LHC), corresponding to an integrated luminosity of 137 fb^{-1}.

Chapter 1

Introduction

The standard model (SM) of particle physics is, to-date, the most successful theory in describing the known elementary particles of the universe and their interactions. It describes three of the four known fundamental forces: electromagnetic, weak, and strong; however, does not provide a description of gravity. Nearly every single experimental test of the standard model is in agreement with theory, with a few notable exceptions, described in greater detail in Sec. 2.3. One of the key features of the SM is the Higgs mechanism [89, 68, 88], which explains how particles obtain mass. An associated particle, the Higgs boson, is also predicted as a consequence of the Higgs mechanism. The Higgs boson was discovered by the ATLAS and CMS collaborations in 2012 [28, 34, 36] with data collected during Run 1 of the Large Hadron Collider (LHC). Since the discovery of the Higgs boson, characterizing its properties has remained one of the highest priorities of research in particle physics.

The results presented in this thesis describe the first observation of Higgs boson production in association with a top quark-antiquark pair ($t\bar{t}H$) in a single decay channel (in which the Higgs boson decays into a pair of photons) [50]. The observation is performed with data collected during Run 2 of the LHC with the CMS detector. Studying $t\bar{t}H$ production allows us to understand the interaction between the Higgs boson and the top quark, of particular interest from a theoretical point of view as many theories of physics beyond the standard model (BSM) may present themselves in the form of modified (relative to the SM prediction) interactions between the Higgs boson and the top quark [11]. This thesis is organized as

follows.

Chapter 2 provides an introduction to quantum field theory and the standard model of particle physics, with a focus on aspects related to the Higgs boson. It also describes the known shortcomings of the SM and how these shortcomings motivate measurements like that of $t\bar{t}H$.

Chapter 3 provides an introduction to the physics of proton-proton collisions, necessary for studying the Higgs boson at the LHC, which collides bunches of protons.

Chapter 4 gives an overview of the LHC and the CMS detector, a multi-purpose apparatus designed to study a wide variety of particles and their underlying physics.

Chapter 5 describes how the raw data from the CMS detector is reconstructed into high-level physics objects suitable for analysis, with a focus on aspects relevant to $H \to \gamma\gamma$ analyses.

Chapter 6 describes the $t\bar{t}H$ analysis documented in [50], with a focus on the aspects to which I contributed most directly.

Chapter 7 draws conclusions, provides perspective on how these results more broadly fit into the field particle physics, and speculates on future work which may build upon these results.

Chapter 2

Theory

2.1 Introduction

This section describes the standard model of particle physics, currently the best known description of the universe's fundamental particles and their interactions. Sec. 2.2 describes quantum field theory, the theoretical framework upon which the standard model (SM) is founded. Details of the SM are then described in Sec. 2.3, with a focus on the central role played by spontaneous symmetry breaking and the Higgs mechanism. Finally, shortcomings of the SM are discussed in Sec. 2.3.6, motivating the Higgs boson as a tool to search for new physics beyond the SM.

The SM is a quantum field theory which describes three of the four known fundamental forces and all known elementary particles. It describes the electromagnetic, strong, and weak interactions, but does not provide a description of gravity. The SM particles can be initially categorized into two groups, bosons and fermions, defined by their intrinsic angular momentum, called "spin".

Bosons are particles which have integer quantum numbers for spin, while fermions are particles which have half-integer quantum numbers for spin. Except for the Higgs boson, a spin-0 "scalar", all bosons in the SM have spin-1. Each of the three forces described by the SM are mediated by the spin-1 gauge bosons: the photon for the electromagnetic force, the $W^{\pm}$ and Z bosons for the weak force, and the eight gluons for the strong force.

The SM fermions all have spin-$1/2$ and can be further divided into two categories: leptons and

quarks. Quarks participate in the strong interaction, while leptons do not. Quarks also participate in the electromagnetic and weak interactions. There are both "up"-type (positively charged) and "down"-type (negatively charged) quarks, with three generations of each, giving six distinct quarks. Each quark also comes in three "color" varieties; however, the different colors of quarks are not experimentally distinct from one another. Leptons can also be further divided into two categories: those which interact with the electromagnetic force (electrons, muons, and taus) and those which interact only with the weak force (neutrinos). Furthermore, each particle in the SM has an accompanying *antiparticle* with opposite electric charge and parity, but otherwise identical physical properties. Table 2.1 summarizes the properties of the SM particles.

Table 2.1: Particle content of the SM, including names, symbol, spin, mass, and electric charge of each particle. Mass values taken from [56].

Particle	Symbol	Spin	Mass [GeV]	Electric Charge	Interactions
Higgs boson	H	0	125	0	
Z boson	Z	1	91.2	0	Weak
W boson	W	1	80.4	$\pm$ 1	Weak
Photon	γ	1	0	0	Electromagnetic
Gluon	g	1	0	0	Strong
Up quark	u	1/2	2.16×10^{-3}	2/3	Weak, Electromagnetic, Strong
Charm quark	c	1/2	1.27	2/3	Weak, Electromagnetic, Strong
Top quark	t	1/2	173	2/3	Weak, Electromagnetic, Strong
Down quark	d	1/2	4.67×10^{-3}	-1/3	Weak, Electromagnetic, Strong
Strange quark	s	1/2	0.093	-1/3	Weak, Electromagnetic, Strong
Bottom quark	b	1/2	4.18	-1/3	Weak, Electromagnetic, Strong
Electron	e	1/2	5.11×10^{-4}	-1	Weak, Electromagnetic
Muon	μ	1/2	0.106	-1	Weak, Electromagnetic
Tau	τ	1/2	1.78	-1	Weak, Electromagnetic
Electron neutrino	ν_e	1/2	$< 1.1 \times 10^{-9}$	0	Weak
Muon neutrino	ν_μ	1/2	$< 1.1 \times 10^{-9}$	0	Weak
Tau neutrino	ν_τ	1/2	$< 1.1 \times 10^{-9}$	0	Weak

2.2 Quantum Field Theory

2.2.1 Classical Field Theory

To begin to understand the Standard Model, a quantum field theory, it is helpful to first understand the classical notion of a field. A field is a physical quantity defined as a function of space and time. The physical quantity may be as simple as a scalar (e.g. the temperature at each point in space and time) or may be a vector (e.g. the electric field). More generally, the physical quantity is a tensor of arbitrary rank.

With the notion of a field defined, we may next ask how to use these fields to describe the behavior of the universe. In classical mechanics, this is achieved through constructing a Lagrangian density, $\mathcal{L}$ (referred to simply as the "Lagrangian"), as a function of one or more fields, $\phi(x)$, and their derivatives:

$$\mathcal{L} = \mathcal{L}(\phi, \partial_\mu \phi). \tag{2.1}$$

One of the fundamental concepts of classical mechanics is the principle of least action, which states that the action, S, defined as the time integral of the Lagrangian,

$$S = \int L\, dt = \int \mathcal{L}(\phi, \partial_\mu \phi)\, d^4x \tag{2.2}$$

for a given system will be minimized as the system evolves between two points in time [106]:

$$\delta S = 0. \tag{2.3}$$

Through inserting Eqn. 2.2 into Eqn. 2.3 and integrating by parts, one obtains the Euler-Lagrange equations of motion:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \tag{2.4}$$

Though the SM Lagrangian is vastly more complex than the toy Lagrangian of Eqn. 2.1, and is composed of quantum rather than classical fields, the same principle of least action allows us to describe the dynamics and interactions of the fundamental particles of the universe.

2.2.2 Quantum Mechanics

While classical mechanics provides a good description of our universe in some regimes, namely when we are dealing with large objects which move much more slowly than the speed of light, it cannot explain many of the observed phenomena in nature. A standard motivation for the need of quantum theory, is the so-called "ultraviolet catastrophe" [111], in which the classical prediction for the energy radiated by a blackbody, an object of some fixed temperature, diverges: the intensity of light radiated increases without bound as a function of frequency. The ultraviolet catastrophe led Planck to propose a solution whose core principle was the idea that light may be radiated only at specific energies – in other words, that the energy was *quantized*. This paved the way for the development of quantum mechanics, which proved very successful for describing blackbody radiation, developing models of the hydrogen atom, and many other applications.

However, one of the major shortcomings of quantum mechanics is the fact that it cannot describe the production or annihilation of particles [121]. More generally, quantum mechanics is incompatible with the theory of special relativity. This fundamental limitation of quantum mechanics motivated the development of theories which could explain the universe at both very small scales (i.e. "quantum") and at very high (relativistic) energies. Quantum field theory has emerged as the most successful theoretical framework for doing so.

2.2.3 The Klein-Gordon Field

To describe the universe in terms of quantum fields, it is helpful to examine a toy example: start with a classical field and make the necessary modifications to reinterpret the dynamical variables as quantum mechanical operators which obey the canonical commutation relations of quantum mechanics[1] (following the treatment in [106]). We then see that the allowed states of the resulting quantum field have a natural physical interpretation as particles.

In choosing a toy example for a quantum field theory, it is helpful to begin with a "derivation" [83] of the Schrödinger equation, which forms the basis of quantum mechanics. Beginning with the classical

[1] A full description of quantum mechanics is beyond the scope of this thesis. A description of quantum mechanical operators and the derivation of their canonical commutation relations can be found in many textbooks on quantum mechanics, e.g. [82].

energy-momentum relation

$$\frac{\mathbf{p^2}}{2m} + V = E, \tag{2.5}$$

one can promote the momentum and energy variables to quantum mechanical operators which act on the wave function Ψ, making the substitutions $\mathbf{p} \to -\mathbf{i}\hbar\nabla$ and $E \to i\hbar\partial/\partial t$, and obtain the Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi = i\hbar\frac{\partial\Psi}{\partial t}. \tag{2.6}$$

As one of the primary aims of quantum field theory is to provide a description of particles which is consistent with special relativity, it is natural to start with the relativistic energy-momentum relation

$$E^2 - \mathbf{p^2} = \mathbf{m^2}, \tag{2.7}$$

and again promote the momentum and energy variables to quantum mechanical operators. Doing so leads one to the *Klein-Gordon equation*, originally proposed to describe the behavior of relativistic electrons[2] [99, 79]:

$$-\frac{\partial^2\Psi}{\partial t^2} + \nabla^2\Psi = m^2\Psi. \tag{2.8}$$

However, we are still working in the context of the wave function for the dynamics of a single particle – in this paradigm we are still unable to describe the annihilation and pair production of particles. We will see that this is possible working in the field theory framework, so it is natural to next ask: what Lagrangian density will give rise to the Klein-Gordon equation? The Lagrangian for the classical Klein-Gordon field is given by

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{2}m^2\phi^2. \tag{2.9}$$

Rather than the Lagrangian formalism, it is often more convenient to work with the Hamiltonian formalism, in which a conjugate momentum density $\pi \equiv \partial L/\partial\dot{\phi}$ is used instead of the time-derivative of the field

[2] In fact, the Klein-Gordon equation does not provide a satisfactory description of relativistic electrons. It applies only to scalar (spin 0) particles, of which the Higgs boson is the only known example in nature.

variable, $\dot{\phi}$. The Hamiltonian density is then defined as

$$\mathcal{H} \equiv \sum \pi\dot{\phi} - \mathcal{L}. \tag{2.10}$$

See [70] for a description of the Hamiltonian formalism.

Returning to the classical Klein-Gordon field, the Hamiltonian density is given by:

$$\mathcal{H} = \frac{1}{2}\pi^2 + \frac{1}{2}(\nabla\phi)^2 + \frac{1}{2}m^2\phi^2. \tag{2.11}$$

The variables π and ϕ can then be promoted to quantum mechanical operators which obey the canonical commutation relations

$$[\phi(\mathbf{x}), \pi(\mathbf{y})] = i\delta^{(3)}(\mathbf{x}-\mathbf{y}) \tag{2.12}$$

$$[\phi(\mathbf{x}), \phi(\mathbf{y})] = [\pi(\mathbf{x}), \pi(\mathbf{y})] = \mathbf{0}. \tag{2.13}$$

Next, it is convenient to rewrite ϕ and π in terms of so-called ladder operators[3], $a_{\mathbf{p}}$ and $a^{\dagger}_{\mathbf{p}}$, defined implicitly as

$$\phi(x) = \int \frac{d^3p}{(2\pi)^3}\frac{1}{\sqrt{2\omega_{\mathbf{p}}}}(a_{\mathbf{p}} + a^{\dagger}_{-\mathbf{p}})e^{i\mathbf{p}\cdot\mathbf{x}}, \tag{2.14}$$

$$\pi(x) = \int \frac{d^3p}{(2\pi)^3}(-i)\sqrt{\frac{\omega_{\mathbf{p}}}{2}}(a_{\mathbf{p}} - a^{\dagger}_{-\mathbf{p}})e^{i\mathbf{p}\cdot\mathbf{x}}, \tag{2.15}$$

and with $\omega_{\mathbf{p}} \equiv \sqrt{|\mathbf{p}|^2 + \mathbf{m}^2}$. Combining the commutation relations with the definitions of the ladder operators, the Hamiltonian may be written [106] as

$$H = \int \frac{d^3p}{(2\pi)^3}\omega_{\mathbf{p}}\left(a^{\dagger}_{\mathbf{p}}a_{\mathbf{p}} + \frac{1}{2}\left[a_{\mathbf{p}}, a^{\dagger}_{\mathbf{p}}\right]\right). \tag{2.16}$$

Calculating the commutators of the Hamiltonian and the ladder operators, $[H, a^{\dagger}_{\mathbf{p}}] = \omega_{\mathbf{p}}a^{\dagger}_{\mathbf{p}}$ and $[H, a_{\mathbf{p}}] = -\omega_{\mathbf{p}}a_{\mathbf{p}}$, we obtain a natural physical interpretation. The operator $a^{\dagger}_{\mathbf{p}}$ acting on the ground state creates a

[3]The motivation for the use of the ladder operators can be found in any standard quantum mechnics textbook, e.g. [82]

state with momentum and energy given by $\mathbf{p}$ and $\omega_{\mathbf{p}}$, respectively – in other words, it creates a particle with momentum $\mathbf{p}$ and energy $\omega_{\mathbf{p}}$. Similarly, acting on this excited state with the operator $a_{\mathbf{p}}$ returns the system to the ground state – it annihilates a particle with momentum $\mathbf{p}$ and energy $\omega_{\mathbf{p}}$.

Although the fields describing the various particles in the SM are considerably more complex than the the scalar field in this example, the Klein-Gordon field still serves to illustrate a valuable point: the quantum field theory framework allows us to describe the creation and annihilation of particles with an energy-momentum relation that is consistent with special relativity. In particular, the Klein-Gordon Lagrangian (Eqn. 2.9) describes a field whose excitations are particles of spin-zero and mass m. More generally, the vast majority of fundamental particles are not spin-zero and we will need more complicated Lagrangians to describe their dynamics.

2.2.4 Spinor & Vector Fields

Other than the Higgs boson, all of the currently known fundamental particles are either spin-$\frac{1}{2}$ (fermions) or spin-1 (bosons). How can we move beyond the Klein-Gordon Lagrangian and construct Lagrangians for spin-$\frac{1}{2}$ and spin-1 particles? In general, the business of constructing Lagrangians in quantum field theory is not as rigorously motivated as in classical field theory, where Lagrangians are derived by the relation $L = T - U$ for a given physical system. Lagrangians in quantum field theory are usually motivated by writing down the most general Lagrangian which respects all of the symmetries of the physical system. Alternatively, we might choose a Lagrangian which yields the desired equations of motion.

The Lagrangian for spin-$\frac{1}{2}$ particles can be motivated by picking one whose resulting equations of motion are the Dirac equation, which Dirac showed describes the dynamics of spin-$\frac{1}{2}$ particles [65]. One such choice is the following, called the Dirac Lagrangian [106]:

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\gamma_\mu \partial_\mu - m)\psi, \tag{2.17}$$

whose resulting equation of motion is the Dirac equation

$$(i\gamma_\mu\partial_\mu - m)\psi(x) = 0. \tag{2.18}$$

In the preceding equations, ψ represents a two-component spinor with its adjoint $\bar{\psi} \equiv \psi^\dagger\gamma^0$ and the γ^μ a set of matrices which satisfy the anticommutation relation

$$\{\gamma_\mu, \gamma_\nu\} = 2g^{\mu\nu}, \tag{2.19}$$

with $g^{\mu\nu}$ the Minkowski metric.

The Lagrangian for spin-1 particles can be motivated by selecting a Lagrangian whose equations of motion are consistent with the dynamics with those of the photon, a familiar spin-1 particle. Such a Lagrangian is the Proca Lagrangian [83], which describes a four-component vector field A^μ

$$\mathcal{L}_{\text{Proca}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2A_\mu A^\mu. \tag{2.20}$$

The resulting field equation is then [83]

$$\partial_\mu F^{\mu\nu} + m^2A^\nu = 0, \tag{2.21}$$

which for the case of the photon (which is massless, $m = 0$) restores Maxwell's equations in empty space: $\partial_\mu F^{\mu\nu} = 0$. The Klein-Gordon, Dirac, and Proca Lagrangians form the basis from which the SM Lagrangian is constructed.

2.3 The Standard Model of Particle Physics

As previously mentioned, the standard model of particle physics is a quantum field theory which describes three of the four known fundamental forces: electromagnetic, weak, and strong. In particular, the SM is a gauge field theory, meaning its Lagrangian is invariant under certain local transformations. Gauge fields are discussed in greater detail in Sec. 2.3.1. In Sec. 2.3.2, we will see that imposing local

gauge invariance on the Dirac Lagrangian gives rise to quantum electrodynamics. Sec. 2.3.3 describes the strong interaction and Sec. 2.3.4 illustrates how spontaneous symmetry breaking and the Higgs mechanism allow for massive gauge fields, enabling us to describe the weak interaction in Sec. 2.3.5.

2.3.1 Gauge Fields

For an arbitrary Lagrangian made of a single field variable ψ, suppose we impose that its resulting field equations be invariant under the local phase transformation

$$\psi \rightarrow e^{iq\theta(x)}\psi. \tag{2.22}$$

This is deemed a "local" phase transformation as the phase θ is be a function of x^μ. In the case that θ is a constant, we deem this a "global" phase transformation. A Lagrangian that is invariant under the transformation in Eqn. 2.22 is said to be *gauge invariant*. More generally, theories which are invariant under gauge transformations are called *gauge theories*. As Sec. 2.3.2 will show, quantum electrodynamics is an abelian gauge theory under the symmetry group U(1), with a single gauge field. The SM as a whole is a non-abelian gauge theory under the symmetry group U(1) $\times$ SU(2) $\times$ SU(3), with a total of twelve gauge fields corresponding to the spin-1 bosons: the photon, the three massive weak bosons, and the eight gluons.

Gauge theories are particularly attractive from a theoretical standpoint for several reasons. First, demanding gauge invariance seems reasonable a priori – the transformation in Eqn. 2.22 is simply a change in the coordinate system we use to define the field ψ, and the physics of the universe should be independent of the particular choice of coordinates we use to describe it. Second, gauge theories have been proven to be renormalizable [115], also a reasonable requirement for a theory we hope will describe the universe.

Renormalization refers to the technique by which a quantum field theory is "cut off" above some very high energy scale Λ, above which the theory is assumed to no longer be valid. In general, this is motivated by the presence of infinities in perturbative calculations of decay rates and cross sections. Rather than assume these infinities render the Lagrangian a useless description of our universe, renormalization serves as a way of quantitatively applying the qualitative statement that the Lagrangian is a low-energy approximation of a more fundamental theory. By formalizing the idea that the theory is only valid up to a

certain energy scale, we are able to avoid the presence of infinities in the calculation of decay rates and cross sections.

2.3.2 Quantum Electrodynamics

Starting with the Dirac Lagrangian (Eqn. 2.17), suppose we impose that its resulting field equations must be invariant under a local phase transformation (as given by Eqn. 2.22). Initially, the Dirac Lagrangian is not invariant under the local phase transformation, as an extra term from the derivative of θ appears:

$$\mathcal{L} \to \mathcal{L} - q(\partial_\mu \theta)\bar{\psi}\gamma^\mu\psi. \tag{2.23}$$

The situation can be remedied with the introduction of a vector field A_μ which transforms under Eqn. 2.22 as

$$A_\mu \to A_\mu + \partial_\mu \theta. \tag{2.24}$$

The resulting Lagrangian,

$$\mathcal{L} = \mathcal{L}_{\text{Dirac}} + q\bar{\psi}\gamma^\mu\psi A_\mu \tag{2.25}$$

is gauge invariant as the second term in Eqn. 2.25 cancels exactly with the additional term from the transformation to the field A_μ in Eqn. 2.24.

Frequently this additional field is absorbed into the definition of a *covariant derivative*

$$\mathcal{D}_\mu \equiv \partial_\mu + iqA_\mu \tag{2.26}$$

which replaces the original definition of the derivative, and the resulting field equations are then invariant under local phase transformations, as desired.

The vector field A_μ which has been added to the Lagrangian implies the existence of an associated spin-1 particle. In principle, we must also include a free term for the field A_μ: it is natural to start with the Proca Lagrangian (Eqn. 2.20) which describes the dynamics of free spin-1 particles. It can be shown [83] that the mass term in the Proca Lagrangian is not invariant under Eqn. 2.24: this can be interpreted as a

requirement that this new vector field A_μ must be massless. The full Lagrangian becomes

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma_\mu \partial_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + q(\bar{\psi}\gamma^\mu\psi)A_\mu, \tag{2.27}$$

$$= \bar{\psi}(i\gamma_\mu \mathcal{D}_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{2.28}$$

with $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$, which can be identified as the Lagrangian for quantum electrodynamics. The field A_μ is associated with the photon, the constant q with the charge of the electron, the tensor $F^{\mu\nu}$ with the electromagnetic field strength, and interactions between photons and electrons with the trilinear term $q(\bar{\psi}\gamma^\mu\psi)A_\mu$.

It is instructive to reflect on the implications of demanding gauge invariance: we started with a spinor field characterized by the Dirac Lagrangian, as would be natural to do in attempting to describe the behavior of electrons. Next, by simply requiring that the equations of motion be invariant under changes in the coordinate system used to describe the field (i.e. demanding gauge invariance), we see that there must be an accompanying massless vector field which interacts with the spinor field, which we identify as the photon field. This is truly remarkable: we have inferred the existence of photons just by demanding that the behavior of electrons be independent of the choice of coordinate system used to describe their field.

2.3.3 Quantum Chromodynamics

As Sec. 3.1 details, inelastic scattering experiments in the 1960s gave strong evidence of the composite nature of protons. Zweig [122] and Gell-Mann [74] independently proposed a quark model to describe the composite nature, which initially implied that quarks violate the spin-statistics theorem. The remedy to this came in the proposal [80] that each quark comes in three different *colors*. More formally, this is the statement that quarks are assigned to the fundamental representation $SU(3)$, giving rise to a quantum number which has three states which we (arbitrarily) call *red*, *green*, and *blue*.

In attempting to construct the Lagrangian for quantum chromodynamics (QCD), which describes the strong interaction of quarks, we can again begin with the free Dirac Lagrangian for spin-1/2 particles (Eqn. 2.17). Given that we have three distinct colors of each quark, the free Lagrangian for a particular

flavor is actually a sum of three free Dirac Lagrangians. This is simplified with the notation

$$\psi = \begin{bmatrix} \psi_r \\ \psi_b \\ \psi_g \end{bmatrix}, \quad \bar{\psi} = [\bar{\psi}_r \; \bar{\psi}_b \; \bar{\psi}_g] \tag{2.29}$$

in which the spinor ψ from the original Dirac Lagrangian has now been promoted to a three-component column vector. The single-particle Dirac Lagrangian is invariant under global phase transformations; in other words, it has $U(1)$ invariance. Similarly, the three-particle Dirac Lagrangian has $U(3)$ invariance:

$$\psi \to U\psi, \quad \bar{\psi} \to \bar{\psi}U^{\dagger} \tag{2.30}$$

with U any unitary 3×3 matrix[4]. Whereas in the case of $U(1)$ symmetry, the invariance has the simple interpretation of a phase, the picture is more subtle for $U(3)$. It can be shown [83] that any unitary matrix can be written in the form

$$U = e^{i\theta} e^{i\lambda \cdot \mathbf{a}} \tag{2.31}$$

with

$$\lambda \cdot \mathbf{a} = \sum_{\mathbf{i=1}}^{\mathbf{8}} \lambda_{\mathbf{i}} \mathbf{a_i} \tag{2.32}$$

and the matrices a_i identified with the eight Gell-Mann matrices which are the generators of the group $SU(3)$. Following the development of the QED Lagrangian, we again impose the requirement that the Lagrangian not just be invariant under global transformations as described by Eqn. 2.30, but also local transformations. In other words, we want $\mathcal{L}$ to be invariant under local $SU(3)$ gauge transformations:

$$\psi \to S\psi, \qquad S \equiv e^{-ig\lambda \cdot \phi(\mathbf{x})} \text{ and } \phi \equiv -\frac{1}{g_s}\mathbf{a} \tag{2.33}$$

As in the case of QED, this can be accomplished through the definition of a covariant derivative

$$\mathcal{D}_\mu \equiv \partial_\mu + i g_s \lambda \cdot \mathbf{a}, \tag{2.34}$$

[4]A matrix U is said to be *unitary* if $U^{\dagger}U = 1$

resulting in the following Lagrangian which is now invariant under local gauge transformations:

$$\mathcal{L} = \bar{\psi}(i\gamma_\mu \mathcal{D}_\mu - m)\psi. \tag{2.35}$$

This time, we have introduced eight gauge fields $\mathbf{A}^\mu$, corresponding to the eight gluons.

Finally, we must account for the free gluon field. As before, the mass terms are excluded because they violate local gauge invariance. However, the field strength tensor for QED, $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$, cannot be directly generalized to QCD due to the fact that transformations of $SU(3)$ are non-Abelian. An additional term is required to restore local gauge invariance, resulting in the QCD field strength tensor

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g_s f^{abc} A^b_\mu A^c_\mu. \tag{2.36}$$

The term f^{abc} corresponds to the SU(3) structure constants and are defined by the commutation relation $[\lambda_a, \lambda_b] = i f^{abc} \lambda_c$. This has no analog in QED; it allows for self-interaction of gluons. The full QCD Lagrangian is then given by

$$\mathcal{L}_{\text{QCD}} = \left(\sum_f \bar{\psi}_f (i\gamma_\mu \mathcal{D}_\mu - m_f)\psi_f \right) - \frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu}, \tag{2.37}$$

where the sum over f corresponds to the different flavors of quarks, of which six have been experimentally observed.

Unlike in QED, in which the magnitude of the force associated with the free photon field *decreases* with distance, the magnitude of the strong force associated with the free gluon field *increases* as a function of distance. As a result, particles possessing color charge cannot exist as free particles and are instead confined to bound states of multiple particles which must always be colorless.

2.3.4 Spontaneous Symmetry Breaking & The Higgs Mechanism

We were able to derive the Lagrangians for the electromagnetic and strong interactions by starting with the Dirac Lagrangian describing free spin-1/2 particles and imposing the principle of local gauge invariance. This involved the introduction of additional vector fields, with which we are able to associate

the mediators of each force: the photon (for QED) and the eight gluons (for QCD). The fact that the mass term in the Proca Lagrangian is not locally gauge invariant implies that the mediators must be massless; conveniently, photons and gluons are indeed observed to be massless. Given the success of the method of imposing local gauge invariance for deriving the Lagrangians for the electromagnetic and strong interactions, it is natural to extend the method to the weak interaction. An immediate challenge, however, is the fact that the mediators of the weak interaction, the W and Z bosons, are not massless. Local gauge invariance can still be applied to the weak interaction, but it requires reinterpreting the original field variables in a form that allows us to expand about their ground state. By doing so, we find that symmetries in the original Lagrangian are broken because of the fact that the ground state does not share the symmetry of the original Lagrangian. This allows for locally invariant massive gauge fields, and as a consequence, implies the presence of a massive scalar particle, which we will identify with the Higgs boson.

Spontaneous symmetry breaking and the Higgs mechanism can be illustrated through a toy Lagrangian composed of a single complex field:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^*(\partial^\mu \phi) + \frac{1}{2}\mu^2(\phi^*\phi) - \frac{1}{4}\lambda^2(\phi^*\phi)^2, \qquad \phi \equiv \phi_1 + i\phi_2 \tag{2.38}$$

In this Lagrangian, the mass term $(1/2)\mu^2(\phi^*\phi)$ appears to have the wrong sign: naively, a positive coefficient implies that the particle associated with the ϕ field has an imaginary mass. Physically, this does not make sense. The subtlety lies in the fact that the Feynman calculus is a perturbative procedure, and must be performed by expanding about a system's ground state. Interpreting $\frac{1}{2}\mu^2(\phi^*\phi) - \frac{1}{4}\lambda^2(\phi^*\phi)^2$ as the *potential* term in the Lagrangian, we can expand about its minimum and apply the Feynman calculus. In contrast to previously considered fields, the minimum does not occur at $\phi_1 = \phi_2 = 0$, but rather is defined by the circle

$$\phi_{1_{\min}}^2 + \phi_{2_{\min}}^2 = \frac{\mu^2}{\lambda^2}. \tag{2.39}$$

Choosing $\phi_{1_{\min}} = \mu/\lambda$ and $\phi_{2_{\min}} = 0$, let us next rewrite the Lagrangian in terms of fields which can be treated as fluctuations about the vacuum state, defining

$$\eta \equiv \phi_1 - \frac{\mu}{\lambda}, \qquad \xi \equiv \phi_2. \tag{2.40}$$

In terms of these new fields, the Lagrangian becomes

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \mu^2 \eta^2 + \frac{1}{2}(\partial_\mu \xi)(\partial^\mu \xi) - \mu\lambda(\eta^3 + \eta\xi^2) - \frac{\lambda^2}{4}(\eta^4 + \xi^4 + 2\eta^2\xi^2) + \frac{\mu^4}{4\lambda^2}. \tag{2.41}$$

The original Lagrangian (Eqn. 2.38) was invariant under rotations in ϕ_1, ϕ_2 space[5]; however, this rotational symmetry is no longer manifest in the η, ξ space. The continuous SO(2) symmetry has been broken by the choice of a particular ground state. The particular ground state we chose, $\phi_{1_{\min}} = \mu/\lambda$ and $\phi_{2_{\min}} = 0$, is arbitrary: the system could just as easily choose any other ground state which satisfies Eqn. 2.40. For this reason, we say that the symmetry has been *spontaneously* broken.

Examining Eqn. 2.41, we can identify that the particle associated with the η field has mass $m_\eta = \sqrt{2}\mu$ and that the particle associated with the ξ field is massless. In fact, Goldstone's theorem [78] shows that the spontaneous breaking of a continuous global symmetry is associated with one or more massless scalar particles, referred to as *Goldstone bosons*.

Next, let us impose the condition of local gauge invariance on the original Lagrangian, Eqn. 2.38, demanding that it be invariant under transformations of the form $\phi \to e^{i\theta(x)}\phi$. As before, we can introduce a massless gauge field A^μ and replace derivatives with covariant derivatives to satisfy local gauge invariance. The Lagrangian becomes

$$\begin{aligned}\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \mu^2 \eta^2 + \frac{1}{2}(\partial_\mu \xi)(\partial^\mu \xi) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}\left(\frac{q\mu}{\lambda}\right)^2 A_\mu A^\mu \\ + q\Big[\eta(\partial_\mu \xi) - \xi(\partial_\mu \eta)\Big]A^\mu + q^2\frac{\mu}{\lambda}\eta(A_\mu A^\mu) + \frac{1}{2}q^2(\xi^2 + \eta^2)(A_\mu A^\mu) \\ - \mu\lambda(\eta^3 + \eta\xi^2) - \frac{\lambda^2}{4}(\eta^4 + \xi^4 + 2\eta^2\xi^2) + \frac{\mu}{\lambda}q(\partial_\mu \xi)A^\mu + \frac{\mu^4}{4\lambda^2}.\end{aligned} \tag{2.42}$$

The gauge field A^μ that we introduced to impose local gauge invariance now has a quadratic term $(1/2)(q\mu)\lambda)^2 A_\mu A^\mu$, which we can associate with a *massive* gauge boson. A mass term associated with the gauge field A^μ has appeared because of the fact that we have rewritten the Lagrangian in a form that allows us to expand about its ground state. In terms of the original ϕ_1 and ϕ_2 fields, no mass term for A^μ appears, but once a ground state has been selected (transforming to η and ξ fields) the gauge boson associated with

[5]More precisely, the original Lagrangian is invariant under SO(2)

A^μ acquires mass: spontaneous symmetry breaking generates masses for gauge bosons.

The Lagrangian of Eqn. 2.42 still presents some difficulties in its physical interpretation. There is a bilinear term proportional to $(\partial_\mu \xi)A^\mu$ which we would interpret as allowing for a ξ particle to suddenly become an A^μ gauge boson. This implies that we have not yet fully cast the Lagrangian in a form that makes its physical interpretation apparent and can be solved by choosing a particular gauge. The Lagrangian of Eqn. 2.38 is invariant under global U(1) phase transformations $\phi \to e^{i\theta}\phi$. If we choose $\theta = -\tan^{-1}(\phi_2/\phi_1)$, the transformed field ϕ' is real ($\phi'_2 = 0$), implying that $\xi = 0$: the problematic bilinear term has been eliminated by the choice of gauge.

We have shown that a gauge boson can acquire mass through the spontaneous breaking of a continuous global symmetry (the SO(2) symmetry of the complex scalar field ϕ). With a proper choice of gauge, we identify a real scalar field η and a massive scalar particle associated with this field. This process by which gauge bosons can acquire mass is known as the *Higgs mechanism* [89, 68] and the massive scalar is known as a Higgs boson.

2.3.5 Electroweak Interactions

The weak and electromagnetic interactions can be unified in a single electroweak interaction, originally developed by Glashow, Weinberg, and Salam [76, 117, 110]. The GWS theory of weak interactions begins with an $SU(2) \otimes U(1)$ gauge symmetry. The symmetry is broken spontaneously through the introduction of a scalar field, leading to the generation of masses for the gauge bosons of the $SU(2)$ component and leaving the gauge boson of the $U(1)$ symmetry massless. The former will be identified as the massive vector gauge bosons, the $W^\pm$ and the Z, while the latter be identified as the massless photon. The particle associated with the scalar field responsible for the spontaneous symmetry breaking will be identified as the Higgs boson.

As before, we demand that the Lagrangian be invariant under local gauge transformations, this time of the form $\phi \to e^{i\alpha^a\tau^a}e^{i\beta/2}\phi$ and define a covariant derivative for ϕ:

$$D_\mu\phi = (\partial_\mu - igA^a_\mu\tau^a - \frac{i}{2}g'B_\mu)\phi, \qquad \tau^a = \sigma^a/2 \tag{2.43}$$

where σ^a are the Pauli spin matrices, A^a_μ corresponds to the $SU(2)$ gauge bosons and B_μ corresponds to the $U(1)$ gauge boson. With a quartic potential for the scalar field ϕ, as in the example of Sec. 2.3.4, the field has a minimum defined by a circle in the ϕ_1, ϕ_2 plane (with $\phi = \phi_1 + i\phi_2$). The original $SO(2)$ symmetry of ϕ will be spontaneously broken when a particular ground state along this circle is chosen. Assuming a ground state of $\phi_1 = (1/\sqrt{2})v, \phi_2 = 0$, choosing the gauge $\alpha^1 = \alpha^2 = 0,\ \alpha^3 = \beta$, and rewriting the Lagrangian about this field configuration leads to the generation of masses for the bosons of the A^a_μ field and leaves the boson of the B_μ field massless. Expressing the Lagrangian in terms of the ground state, rather than the original field ϕ, we find that the following terms appear:

$$\Delta\mathcal{L} = \frac{1}{2}\frac{v^2}{4}\left[g^2(A^1_\mu)^2 + g^2(A^2_\mu)^2 + (g'B_\mu - gA^3_\mu)^2\right]. \tag{2.44}$$

Again we see that rewriting the Lagrangian in a way such that it can be expanded about its spontaneously chosen ground state breaks the original symmetry and gives rise to mass terms for the $SU(2)$ gauge field.

The original fields can be expressed in terms of their mass eigenstates, which will make the physical interpretation of this theory more transparent. It can be shown [106] that these eigenstates are

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(A^1_\mu \mp iA^2_\mu) \tag{2.45}$$

$$Z^0_\mu = \frac{1}{\sqrt{g^2+g'^2}}(gA^3_\mu - g'B_\mu) \tag{2.46}$$

$$A_\mu = \frac{1}{\sqrt{g^2+g'^2}}(g'A^3_\mu + gB_\mu) \tag{2.47}$$

The $W^\pm_\mu$ field has vector bosons of mass $m_W = gv/2$, the Z^0_μ field has a vector boson of mass $m_Z = \sqrt{g^2+g'^2}v/2$, and the A_μ field remains massless. The covariant derivative can be rewritten in terms of the mass eigenstates and the *weak mixing angle*, θ_w, defined as

$$\begin{pmatrix} Z^0 \\ A \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} A^3 \\ B \end{pmatrix}. \tag{2.48}$$

As the field A_μ will be identified as the electromagnetic vector potential, it is helpful to define e, which will

be identified as the electron charge

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}}. \tag{2.49}$$

In this notation the covariant derivative becomes

$$D_\mu \phi = \left[\partial_\mu - i\frac{g}{\sqrt{2}}(W_\mu^+ T^+ + W_\mu^- T^-) - i\frac{g}{\cos\theta_w} Z_\mu (T^3 - \sin^2\theta_w Q) - ieA_\mu Q \right] \phi. \tag{2.50}$$

We can next couple the $SU(2) \otimes U(1)$ gauge fields of the electroweak interaction to the leptons and quarks. As the W boson couples only to the left-handed helicity states of leptons and quarks, it is helpful to decompose the kinetic energy term for fermions into left- and right-handed components:

$$\bar{\psi} i\gamma^\mu \partial_\mu \psi = \bar{\psi}_L i\gamma^\mu \partial_\mu \psi_L + \bar{\psi}_R i\gamma^\mu \partial_\mu \psi_R, \tag{2.51}$$

so that ψ_L and ψ_R can couple differently to the gauge fields.

The left- and right-handed components of fermion fields couple differently to the gauge fields, they have different quantum numbers and consequently simple mass terms are forbidden by gauge invariance. Experimentally, we know that the fermions are not massless, so this poses a problem. Again, spontaneous symmetry breaking can remedy this and allow for fermions to acquire mass.

Assuming the scalar field ϕ undergoes spontaneous symmetry breaking ("acquires a vacuum expectation value"), we can add terms to the Lagrangian which describe interactions between ϕ and left- and right-handed components of fermions. For example, for the electrons:

$$\Delta\mathcal{L}_e = -\lambda_e \bar{E}_L \cdot \phi e_R + \text{ h.c.}, \qquad E_L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \tag{2.52}$$

Assuming the same ground state as before, $\phi = (0 \quad v/\sqrt{2})$, we obtain a mass term for the electron:

$$\Delta\mathcal{L}_e = \frac{-1}{\sqrt{2}} \lambda_e v \bar{e}_L e_R + \text{ h.c.}, \tag{2.53}$$

referred to as the Yukawa term and λ_e referred to as the Yukawa coupling. Yukawa terms for the other

leptons and the quarks can be obtained in a similar fashion, such that the mass of any fermion is given by

$$m_f = \frac{1}{\sqrt{2}}\lambda_f v. \tag{2.54}$$

Neither the magnitude of the Yukawa couplings, λ_f, nor the vacuum expectation value, v, are known a priori and must be measured experimentally.

One final subtlety of the electroweak interaction is that while the W boson couples to leptons only of the same generation, it can couple to quarks from different generations. This mixing is a result of the fact that the mass eigenstates of quarks are different from the weak isospin eigenstates. The mixing of these eigenstates is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [17, 100]:

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{2.55}$$

The CKM matrix is experimentally observed to be nearly diagonal, which has the physical consequence that W-mediated interactions between quarks of different generations are much weaker than those between quarks of the same generation.

2.3.6 Shortcomings of the Standard Model

Though the SM has proved to be compatible with nature in nearly every single experimental test, it is known to be an incomplete theory. Some of the most glaring shortcomings of the SM are summarized in this section, but it is by no means an exhaustive list of every problem.

First, the SM does not provide a description of gravity, and is (so far) incompatible with the theory of *general relativity*, currently the most successful theory of gravity. Second, the SM cannot explain dark matter & dark energy, with strong evidence for dark matter given by the inconsistency of galactic rotation curves with the amount of visible matter [109]. Third, the SM does not explain the observed dominance of matter over antimatter in the universe [64] – CP violation (which is allowed and observed in the SM) can account for some of this asymmetry, but not nearly enough to explain the observed asymmetry. Finally, the

SM is known to be self-inconsistent at very high energies, with the electromagnetic coupling and the Higgs self-coupling both diverging at arbitrarily high energies [77, 23]. This issue, called *quantum triviality*, implies that the SM is only valid up to some finite energy scale.

Given the abundance of issues with the SM, it is widely believed that the SM is a low-energy approximation of some more fundamental theory. With the stipulation that the SM is fundamentally an incorrect description of the universe, it is natural to search for the ways in which the SM does *not* correctly describe our universe. In this paradigm, the experimental success of the SM is puzzling – if the SM is wrong, why does every experimental test agree with its predictions? One of the more promising approaches is to test the SM at increasingly high energy scales, as is done through analysis of the 13 TeV center-of-mass collisions at the Large Hadron Collider (LHC). Among the countless measurements which can be made with the datasets from the LHC's numerous experiments, measuring the properties of the Higgs boson is a promising avenue to search for departures from the SM predictions given the Higgs unique role in mass generation and electroweak symmetry breaking.

2.3.7 The Higgs Boson

The SM hinges upon the existence of a scalar field ϕ which undergoes spontaneous symmetry breaking to acquire a vacuum expectation value, thereby allowing the gauge bosons and fermions to acquire mass. The scalar field must have a potential with minima that lie outside $\phi = 0$ in order for this to occur. As illustrated in Sec. 2.3.4, one such Lagrangian which leads to a vacuum expectation value is

$$\mathcal{L} = |D_\mu \phi|^2 + \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2. \tag{2.56}$$

The particle associated with this field will then have a mass given by

$$m_H = \sqrt{2}\mu = \sqrt{2\lambda}v. \tag{2.57}$$

A particle consistent with the Higgs boson was discovered in 2012 by the CMS and ATLAS collaborations [28, 34, 36], with its mass measured to be 125.35 ± 0.15 GeV [47].

Since its discovery, other measurements of the Higgs boson's properties have so far confirmed

that it is consistent with the SM Higgs boson. Multiple production modes of the Higgs boson have been experimentally confirmed at the LHC, with observations of Higgs boson production via gluon fusion and vector boson fusion [29, 39, 8] made during Run 1 of the LHC, and observations of Higgs boson production in association with a vector boson [30] or a top quark-antiquark pair [46] made during Run 2 of the LHC. A variety of expected decay modes of the Higgs boson have also been experimentally confirmed, with branching fractions consistent with the SM predictions. The $\gamma\gamma$, ZZ^*, $W^{\pm}W^{\mp *}$, $\tau^{\pm}\tau^{\mp}$ [29, 39, 8], and $b\bar{b}$ [30] decay modes have each been observed.

The Higgs Boson as a probe of new physics

While the discovery of the Higgs boson and the confirmation of its SM-like properties are some of the most successful experimental validations of the SM, they are also frustrating in some sense: the SM is known to be an incomplete description of our universe, for reasons detailed in the previous section. It is widely accepted that the SM is a low-energy approximation of some more fundamental theory of the universe. There are a variety of theories of physics beyond the standard model (BSM), for example, the theory of supersymmetry [103]. Direct searches for new physics have been continually performed at the LHC, but so far, there has been no conclusive evidence for the presence of any BSM physics. How can we reconcile these two facts: (1) there should be BSM physics and (2) there is no evidence of BSM physics at the LHC? One possibility is that new physics exists at masses which are beyond the energy reach of the LHC. In these scenarios, the presence of new physics might still manifest itself in the form of small deviations of the properties of the Higgs boson from the SM expectations. For example, a two-Higgs doublet model would result in modified coupling constants of the Higgs boson which would be observable at a future particle collider [97].

The top quark Yukawa coupling is of particular interest from a theoretical standpoint, as precise measurements of its value could give insight on the existence and energy scale of new physics [11]. Measurement of Higgs boson production in association with a top quark-antiquark pair is the best way to directly constrain the top quark Yukawa coupling at the LHC. One such measurement [50] is the focus of this dissertation.

Chapter 3

Physics of Proton-Proton Collisions

3.1 The Parton Model

The field of particle physics saw the discovery of a variety of new particles in the 1950s and 60s. At the time, their fundamental nature was unknown; however, due to their sheer number it seemed plausible that these particles, now known as mesons and baryons, were not elementary but composite. Zweig [122] and Gell-Mann [74] independently proposed that mesons and baryons were in fact composed of spin-1/2 particles which Gell-Mann coined "quarks". In this framework, mesons were bound states of a quark and an anti-quark while baryons were bound states of three quarks. More precisely, mesons and baryons are composed of their respective quarks, called valence quarks (which dictate the nucleon's quantum numbers), gluons (which mediate the strong force and bind the nucleon), and a sea of virtual quark-antiquark pairs [120]. The quark model of Zweig and Gell-Mann was initially met with some skepticism: it implied that quarks must have fractional charges of either 1/3 or 2/3 of the charge of the electron and that they violated the spin-statistics theorem. The quantum number color [80] was proposed to remedy the violation of the spin-statistics theorem and deep inelastic scattering experiments at SLAC [15, 67] gave strong indications of the composite structure of the proton.

In attempting to describe the nature of inelastic electron-proton scattering, Bjorken proposed that the cross section for such a process was determined not by the absolute energy of the collision, but instead depended on dimensionless kinematic ratios [12]. This phenomenon, known as Bjorken scaling, was

experimentally confirmed in experiments at SLAC [107]. Feynman interpreted Bjorken scaling in the following way [71]: hadrons behave as a collection of point-like constituents, or *partons*, which each carry some fraction of the hadron's total energy. Moreover, each of the partons can be described by a *parton distribution function* (PDF), which gives the probability density for a parton to carry a given fraction of the hadron's total momentum.

The parton model assumes that quarks exist as free particles within the hadron, which is a good approximation for high energy electron-proton scattering in which the interactions between the electron and parton occur on a very short time scale through electromagnetic interactions mediated by photons. This approximation breaks down for inelastic proton-proton scattering, in which the gluons which bind the proton together become relevant in the interaction and it is no longer valid to consider the partons as free particles. In this regime, Bjorken scaling is no longer applicable, and the PDFs become dependent on the magnitude of the momentum transfer (frequently denoted by Q^2). The PDF dependence on Q^2 cannot be calculated analytically; instead, the PDF can be measured through experiment at a particular value of Q^2 and then extrapolated to other values of Q^2 through the QCD evolution equations for parton densities [5, 66, 81], also called the DGLAP equations.

The PDFs for the proton, as determined by the MSTW group [102], are shown in Fig. 3.1 for $Q^2 = 10$ GeV and $Q^2 = 10^4$ GeV. Although the proton's quantum numbers are determined by its valence quarks, *uud*, there is a non-negligble fraction of the proton's momentum carried by both the gluons and the "sea" of quark-antiquark pairs.

3.2 Proton-Proton Collisions

3.2.1 Cross Sections

Some of the primary quantities we are interested in predicting and measuring in proton-proton collisions are *cross sections*, which can be thought of as a measure of the probability[1] of a specific process occurring. The QCD factorization theorem [57] allows the cross section for an arbitrary deep inelastic proton-proton collision to be written in terms of two components: a perturbatively calculable hard term and

[1] More precisely, a cross section is measured in units of distance squared. However, the characteristic cross section of a process is synonymous with its probability of occuring and it is more intuitive to describe cross sections in these terms.

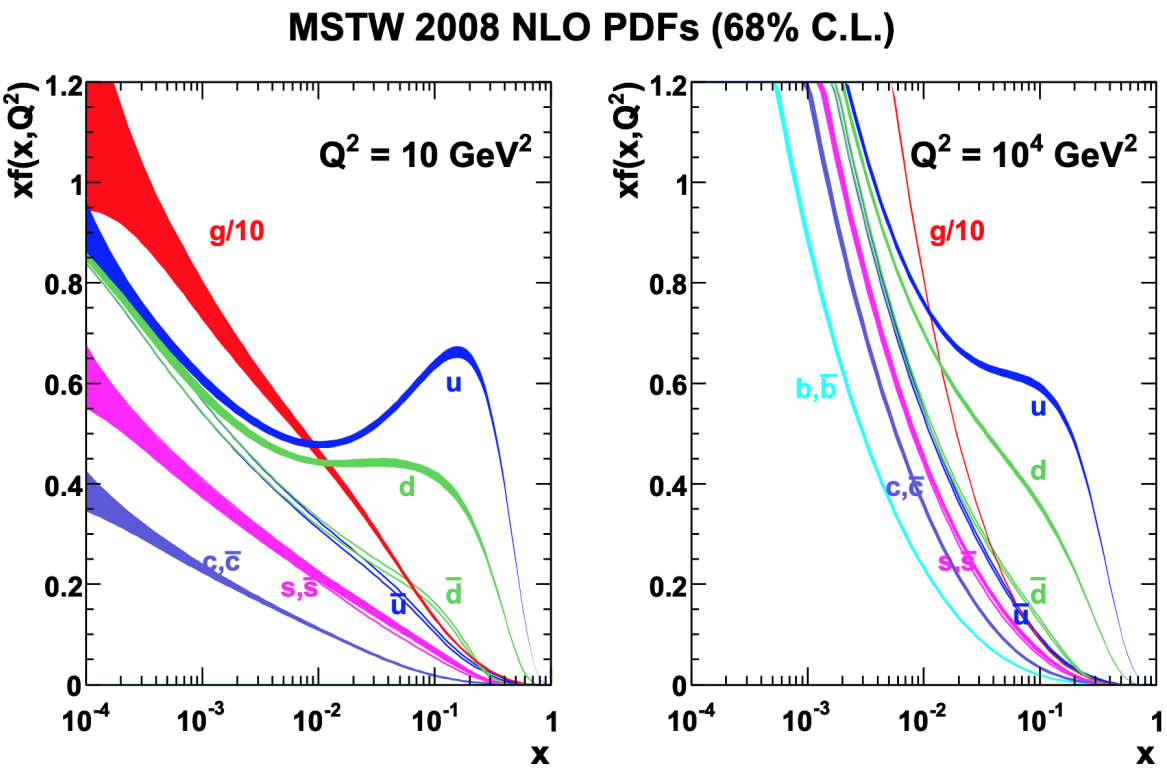

Figure 3.1: Parton distribution functions for the proton, shown for $Q^2 = 10$ GeV (left) and $Q^2 = 10^4$ GeV (right), as calculated by the MSTW group. The width of each band indicates the 68% C.L. Taken from [102].

a non-perturbative PDF. Thus, the cross section for $pp \to X + Y$ can be calculated in the following way:

$$\sigma(pp \to X + Y) = \sum_{i,j} \int dx_i dx_j f(x_i, Q^2) f(x_j, Q^2) \sigma(q_i q_j \to Y), \tag{3.1}$$

in which X may be any hadronic final state, and Y is an arbitrary final state for the inelastic scattering of two partons q_i and q_j. The sum is calculated over all partons and integrated over all possible momentum fractions for the PDFs of each parton. The hard term, $\sigma(q_i q_j \to Y)$, can be calculated perturbatively in QCD. In practice, these calculations are done through the use of Monte Carlo generators, described in greater detail in Sec. 3.3. Cross sections for typical processes of interest at pp collision experiments are shown as a function of the center-of-mass energy in Fig. 3.2.

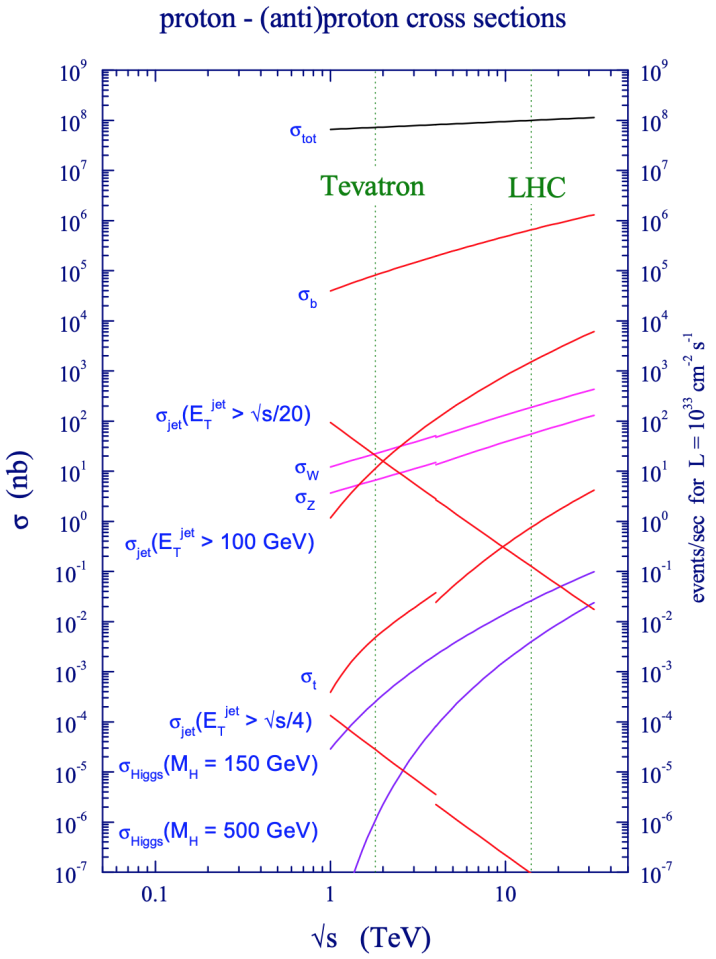

Figure 3.2: Cross sections for typical processes of interest in pp collision experiments, shown as a function of the center-of-mass energy, $\sqrt{s}$. Taken from [20].

3.2.2 Parton Showers, Hadronization, and Jets

High energy processes involving the strong interaction are very well-described by perturbative QCD calculations. However, at lower energies (less than or equal to about 1 GeV), the perturbative appraoch fails to provide an accurate description of the SM phenomena: the strong coupling α_s of QCD becomes close to unity, as shown in Fig. 3.3. When the coupling α_s nears unity, the perturbative approach

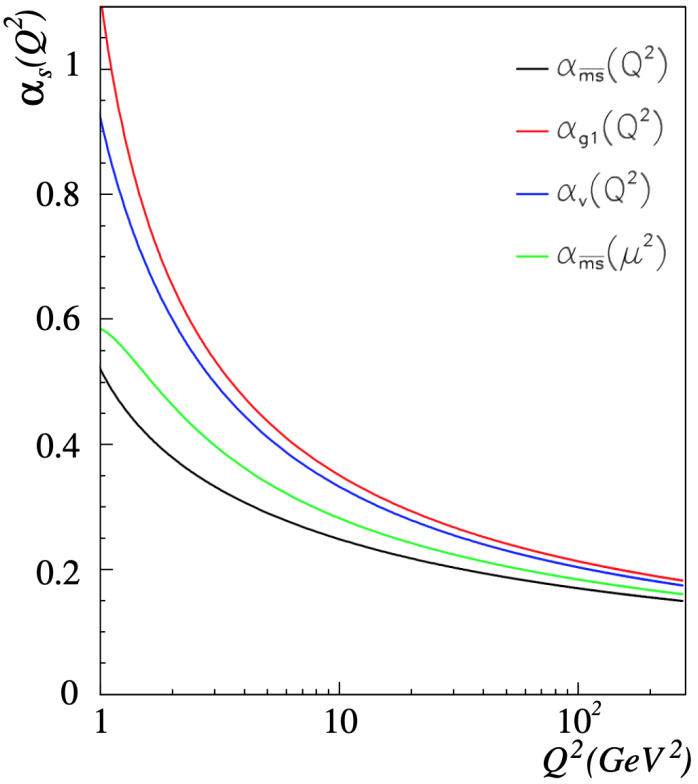

Figure 3.3: The strong coupling constant α_s of QCD as a function of Q^2. Different colored lines correspond to various renormalization schemes. Taken from [62].

fails for the following reason: perturbative expansions are made in powers of the coupling, so the coupling must be significantly less than one in order for a finite expansion to provide a good approximation. In describing phenomena like parton showers and hadronization, energy scales of $O(1)$ GeV are relevant, and a strictly perturbative calculation will not provide a satisfactory description.

A *parton shower* refers to the process by which a high energy parton stemming from the hard interaction produce showers of "soft" particles at lower energies. Typically, this is either gluon splitting,

in which a gluon converts into a quark-antiquark pair, or gluon radiation, in which a quark radiates a gluon. In practice, parton showers are modeled with Monte Carlo generators which utilize Sudakov form factors [114] and splitting functions to simplify calculations [90].

As discussed in Sec. 2.3.3, quarks and gluons are confined to bound states which must be colorless. Moreover, the potential energy of a hadron increases as a function of the distance between the partons. At a large enough distance, it becomes energetically favorable to break the original bound state in which they existed and instead form new hadrons. This process is called *hadronization*. In high energy collisions, quarks and gluons are often ejected from the hard interaction with high enough momentum for hadronization to occur. Frequently, the newly formed hadron will initiate a cascade of decays and gluon radiation, forming a cone of hadronic activity. This cone of particles stemming from the hadronization of a quark or gluon is called a *hadronic jet*. Hadronization cannot be adequately described through perturbative calculations alone, and instead phenomenological models like the Lund-String Model [7] are employed.

3.2.3 Underlying Event and Pileup

In a given bunch crossing, there is typically only one hard scattering interaction of interest from a physics point of view. In addition to this hard interaction, there are additional lower energy "soft" scattering interactions. The soft scattering may be due either to interactions between partons other than those involved in the hard scattering interaction or interactions between protons other than those involved in the hard scattering interaction. The former is called the *underlying event*, while the latter interactions are called *pileup* interactions. The modeling of underlying event and pileup is often performed through heuristic approaches which extrapolate directly from experimental collision data.

Though soft scattering interactions from underlying event and pileup are typically not of interest, it is still imperative to understand and adequately model them in order to study physics processes of interest. A large portion of the hadronic activity in an event at the LHC stems from these soft interactions and will effect, for example, the jet multiplicity and missing transverse momentum calculation in that event. Physics analyses often use the jet multiplicity and missing transverse momentum to identify regions of high signal purity (for example, an analysis searching for supersymmetric particles will typically select for events with high missing transverse momentum) – for these reasons, it is vital to understand the contribution of

underlying event and pileup to these distributions in order to properly model the targeted signal process and accurately estimate the relevant SM background processes.

Parton showers, hadronization, and underlying event are visually depicted for a hadron-hadron collision in Fig. 3.4.

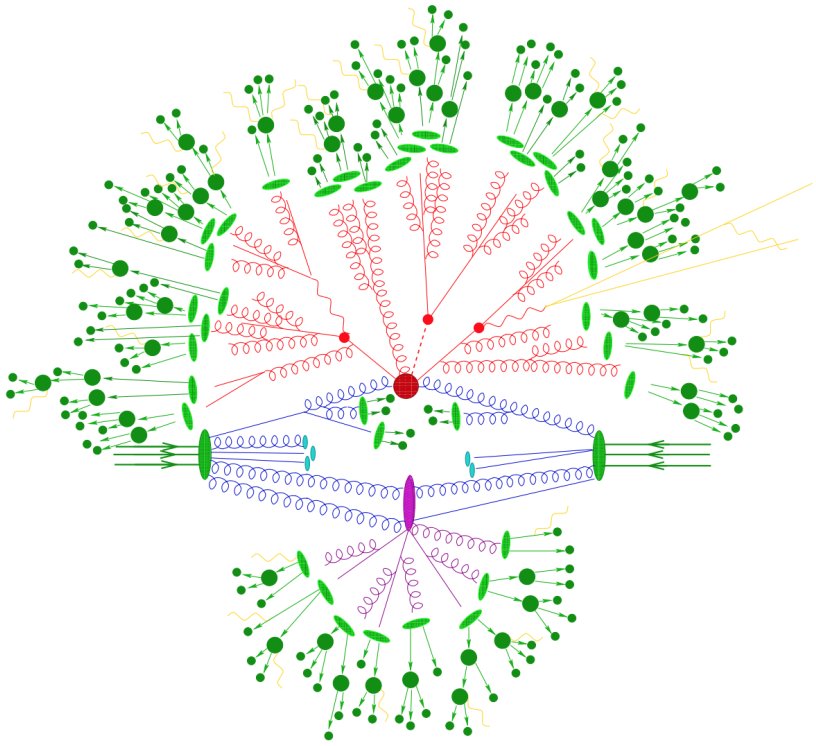

Figure 3.4: Schematic of a hadron-hadron collision. The red blob indicates the hard scattering interaction and the subsequent tree-like structure depicts parton showers, while the purple blob indicates an underlying event scattering interaction. Light green blobs depict hadronization, dark green blobs depict subsequent decays of those hadrons, and yellow lines depict soft Bremsstrahlung radiation. Taken from [90].

3.3 Monte Carlo Simulation

Simulations of proton-proton collisions are useful for a wide variety of applications. First, many physics analyses search for signal processes which are either very rare in the SM or not predicted to occur at all in the SM. For such processes, simulation is necessary to predict the kinematics and event yields. Second, simulation can help creating a description of the relevant SM background processes. It is often preferable to use data-driven procedures for describing backgrounds; however, this is not feasible for many

rare SM processes. Third, simulation can help instruct the development and event selection for a physics analysis. For example, simulation is used extensively within CMS in training and optimizing machine learning algorithms, which are used for tasks like jet flavor identification, discrimination between prompt and fake leptons or photons, and discrimination between signal and background processes.

Monte Carlo (MC) simulation of a given physics process typically undergoes three steps. First, the hard-scattering interaction is simulated using a generator like MADGRAPH [6]. These calculations are done through the use of perturbative expansions in powers of couplings (e.g. α_s for processes mediated by the strong interaction). This is not strictly a perturbative calculation however, as these generators also use the parton distribution functions of the protons as inputs, which cannot be derived through perturbative approaches. Generators like that of [6] simulate processes to next-to-leading-order (NLO) precision. These predictions are generally precise enough for the needs of most physics analyses, but it must be emphasized that these simulations are known a priori to be an incomplete description of the full SM phenomena.

As mentioned in Sec. 3.2.2, a perturbative approach is suitable for the hard-scattering interaction, but soft-scattering interactions, like parton showers and hadronization, occur at too low of energies for the perturbative approach to provide accurate results. Therefore, the output of parton-level generators like MADGRAPH are usually then interfaced with software like PYTHIA, which simulate the event all the way to the final state particles, including effects like parton showers, hadronization, and initial & final state radiation.

Finally, the event is framed not in terms of the final state particles and their properties, but rather in terms of the signatures they are expected to leave in a given detector. The detector for a given high energy physics experiment is, in general, very dynamic: detectors accrue radiation damage over time, components are subject to failure or faulty behavior, and upgrades may be implemented during periods in between data-taking. Given these considerations, a detailed model of the detector is created and implemented in software like GEANT [52].

Chapter 4

Compact Muon Solenoid

4.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a hadron accelerator and collider located in a 27 km underground tunnel on the French-Swiss border, near Geneva. Its primary physics goal is to reveal physics beyond the Standard Model [69]. The LHC hosts a variety of experimental collaborations, including the CMS experiment [33], the ATLAS experiment [27], the LHCb experiment [53], and the ALICE experiment [26]. The physics goals of the CMS and ATLAS experiments are identical, namely to search for the presence of new physics beyond the Standard Model and to make precision measurements of the properties of the Higgs boson. The LHCb experiment focuses on studies of CP violation and rare decays of b hadrons, while the ALICE experiment focuses on studying the strong interactions of QCD. The LHC collides both proton and lead (Pb) ion beams. The CMS, ATLAS, and LHCb experiments are designed to study physics from proton-proton collisions, while the ALICE experiment utilizes the Pb-Pb collision data. This thesis focuses on results obtained in proton-proton collisions with the CMS detector.

Given that the primary research goal of the LHC is to disover physics beyond the Standard Model, protons are the natural choice for the beam content, rather than electron-positron beams, as used in other colliders like the Large Electron Positron (LEP) collider [116]. Two primary advantages provided by proton beams as a disovery tool are the fact that protons are composite particles, rather than elementary particles like electrons, and the fact that proton beams can more easily be sustained at higher energies.

Protons are primarly composed of up quarks, down quarks, and gluons, with the subcomponents collectively referred to as partons. An interaction between two protons is then more precisely an interaction between two partons from each proton. The partons each carry some fraction of the proton's energy, effectively providing a range of collision energies. Since the energy scale of new physics is not precisely known, it is more desirable to have collisions occuring at a range of energies, as naturally occurs in proton collisions. The proton beams at the LHC are accelerated to energies of 6.5 TeV, providing a center of mass energy of 13 TeV. New physics which might exist at the TeV scale is then accessible, in principle, through the range of energies provided by proton-proton collisions. In contrast, in order to achieve sensitivity to possible new physics at a range of energies with electron-positron beams, one would need to manually change the collision energies. Broadly speaking, proton-proton colliders are better suited for discovery of particles whose exact mass is not known, while electron-positron colliders are better suited for precision measurements of particles whose mass is precisely known.

Proton beams are more easily sustained at high energies than electron-positron beams due to the fact that protons dissipate less energy through synchrotron radiation. The energy emitted per unit time per unit solid angle for an accelerated charged particle is inversely proportional to the particle's mass [94]:

$$P \propto \frac{1}{m^2} \tag{4.1}$$

Given that the mass of an electron is on the order of 1 MeV and the mass of a proton is on the order of 1 GeV, the mass dependence of Eqn 4.1 contributes a factor of $O(10^6)$ difference in the energy lost to synchrotron radiation between protons and electrons accelerated in a circular trajectory. The challenges associated with energy lost to synchrotron radiation in electron beams are reflected in the difference of center-of-mass energies achieved by the LHC (13 TeV) and LEP (209 GeV).

The LHC accelerates protons in groups known as bunches. The probability for any two protons to interact is extremely low; this is counteracted by increasing the number of protons contained in each bunch, N_b. N_b cannot be increased without bound and is limited by nonlinear beam-beam interactions which occur when two bunches collide with each other. At the LHC, the maximum attainable bunch size is $N_b = 1.15 \times 10^{11}$.

Although the large circumference (27 km) of the LHC results in longer and more expensive beam pipes, it provides several advantages for sustaining high energy collisions. Proton beams must be constrained to travel in a circular shape through the presence of a large magnetic field. The beams at the LHC require a magnetic field of over 8 T to stay on track; however, this requirement would be significantly higher with a smaller circumference collider – a larger circumference reduces the curvature and allows for higher energy beams at a given magnetic field strength. Sustaining such a large magnetic field also presents challenges: an extremely large current (over 10^4 A) is required to produce the magnetic field. For such a high current, superconducting magnets are required: niobium-titanium superconducting electromagnets are used in the LHC, and superconductors stable at higher magnetic fields are extremely expensive.

Even with billions of protons per bunch, only $O(10-100)$ pairs of protons will actually collide with each other in a given bunch crossing. Fig 4.1 shows the distribution of the number of proton-proton interactions, also referred to as the "pileup", in the CMS detector during Run 2 of the LHC. The bunches are spaced by a distance corresponding to a time of 25 ns between bunches. In general, the more closely bunches are grouped together, the more collisions can be recorded; however, the bunch spacing is limited by considerations like experiments' temporal resolution – bunches must be sufficiently separated to allow each experiment to distinguish between consecutive collisions.

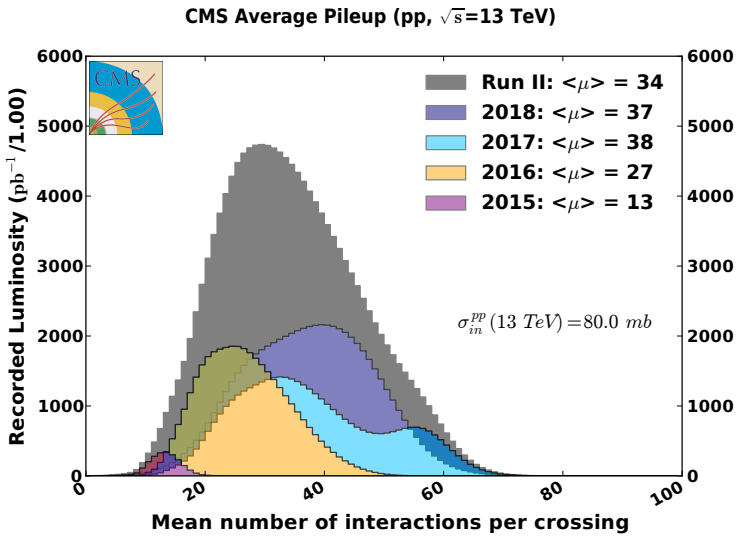

Figure 4.1: Mean number of interactions per bunch crossing recorded by the CMS detector during Run 2 of the LHC. Taken from [51].

During Run 2 of the LHC, the CMS detector recorded 150 fb^{-1} of data from proton-proton collisions. A subset of that data is verified to have stable detector performance and marked as usable for physics analysis, amounting to 137 fb^{-1} of data used in the analysis discussed in this work.

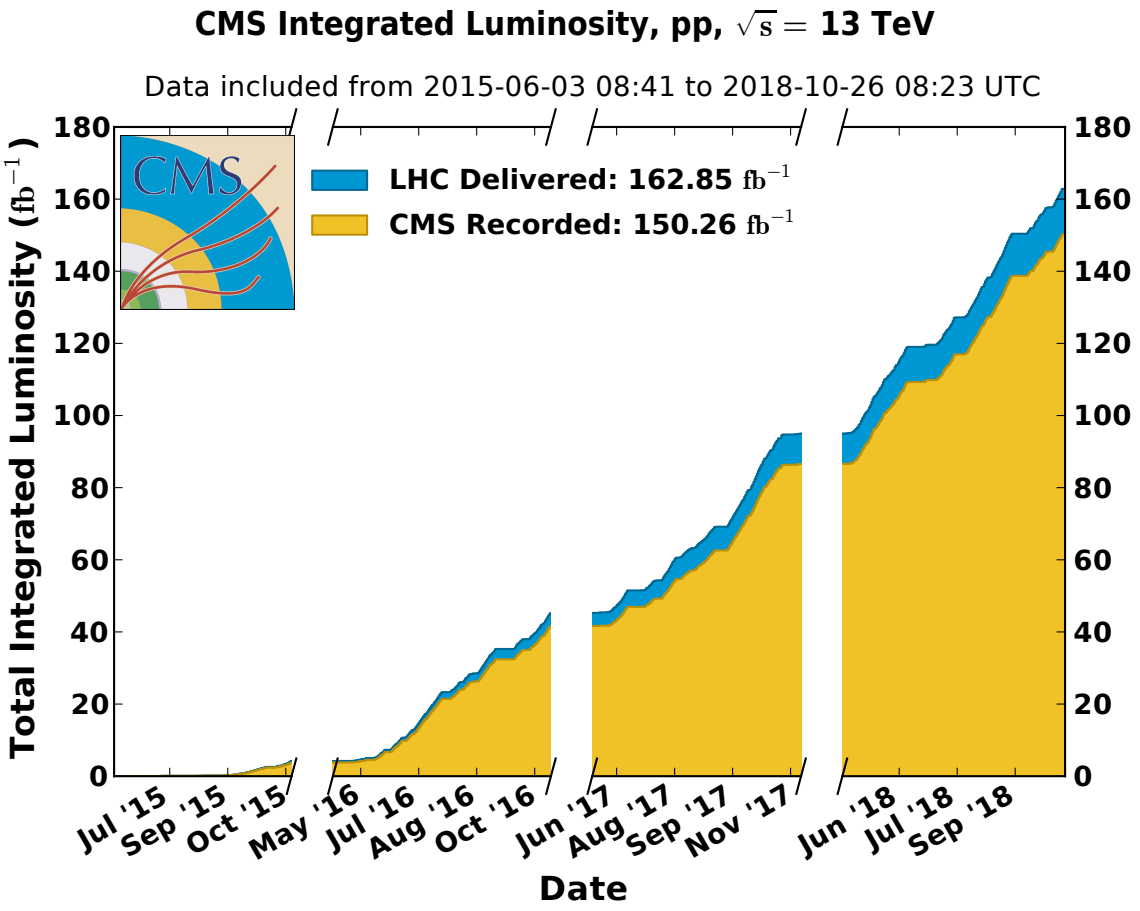

Figure 4.2: Total luminosity delivered by the LHC (blue) and total luminosity recorded by the CMS detector (yellow) during Run 2 of the LHC. Taken from [51].

4.2 The Compact Muon Solenoid Detector

The CMS detector was designed with two primary physics goals in mind. First, to study the properties of the Higgs boson; in particular, the nature of electroweak symmetry breaking for which the Higgs mechanism is responsible[1]. Second, to reveal signs of physics beyond the Standard Model which might be present at the TeV scale. This section discusses technical design aspects of the CMS detector and

[1]Note that during the design of the CMS detector, the Higgs boson was theorized to exist but had not yet been discovered.

how they support the physics goals of the CMS experiment.

4.2.1 General Design Concepts

The CMS detector is over 20m in length and nearly 15m in diameter – it is "compact" only in the context of the tremendous size of a detector needed to facilitate the physics goals for which it was designed. The various components of the CMS detector are shown in Fig. 4.3, with humans shown to illustrate the scale (banana not available).

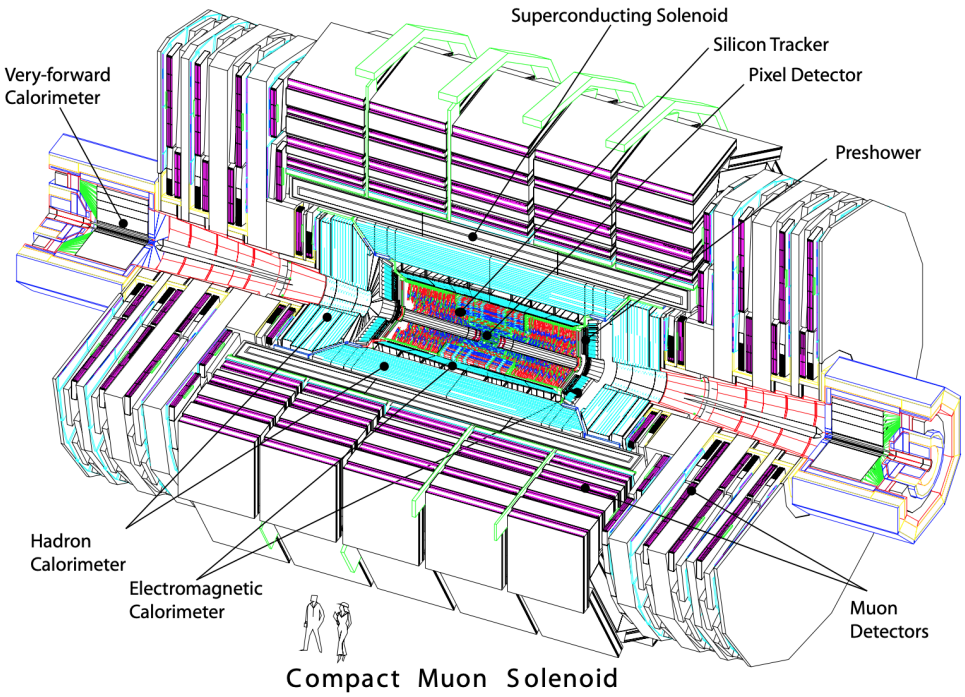

Figure 4.3: Schematic of the various components of the CMS detector. Taken from [33].

As its name suggests, the primary feature of the CMS detector is a superconducting solenoid providing a magnetic field of 4 T. The purpose of the magnetic field is to bend the path of charged particles originating from inelastic proton-proton interactions: the precise spatial resolution of the tracker and muon system allows one to determine a particle's radius of curvature, in turn allowing one to determine the particle's momentum. Excellent momentum resolution of charged particles supports nearly every physics analysis performed with the CMS detector, but is especially important for determining the invariant mass of heavy resonances (e.g. studying the properties of the Higgs boson in the $H \rightarrow ZZ^* \rightarrow 4l$ channel),

distinguishing between hadronic jets which originate from b quarks from those which originate from gluons or light-flavor quarks (e.g. the $H \to bb$ decay channel and searches for new physics involving final states with b-jets), and for precise resolution of the missing transverse energy (e.g. final states with neutrinos or searches for new physics involving final states with undetected dark matter or supersymmetry candidate particles).

The other major components of the detector include:

- The tracker, which allows for identification and measurement of the momenta of charged particles.
- The electromagnetic calorimeter (ECAL), which allows for identification and momenta measurement of electrons and photons, particularly important for $H \to \gamma\gamma$ physics.
- The hadronic calorimeter (HCAL), which assists in the identification and momentum resolution of charged hadrons and provides the only handle on measuring neutral hadrons.
- The muon system, which enables better momentum resolution of very high energy (O(TeV)) muons (while the tracker excels in providing good momentum resolution for lower energy, O(GeV) muons).

Each of these components is described in greater detail in the following subsections.

A design consideration common to multiple subdetector components is the goal of hermeticity: a fully hermetic detector is able to measure particles emerging in any direction from an inelatic collision. In other words, a hermetic detector has full coverage of the 4π steradians of solid angle surrounding the interaction point. The CMS detector is not fully hermetic, as it is practically impossible to measure particles which emerge parallel to the LHC's proton beams. Still, the CMS detector is able to measure very forward particles (with "forward" meaning "close to parallel with the beam axis"), aiding the nearly complete reconstruction of the final state of a given pp interaction, which is essential for resolution of the missing transverse energy.

To expand upon the concepts of hermeticity and the identification of forward particles, we must first introduce the coordinate system used to describe the CMS detector. Given the cylindrical shape of the detector, standard cylincdrical coordinates form the basis of the coordinate system: the $\hat{z}$-axis is defined as the axis along which the proton beams travel, and the $\hat{\phi}$ direction then coincides with the detector's

circular symmetry perpendicular to the beam axis. Instead of the typical polar angle $\hat{\theta}$, position is usually expressed in terms of pseudorapidity, defined in terms of θ as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{4.2}$$

A pseudorapidity of $\eta = 0$ corresponds to a direction perpendicular to the beam axis, while $\eta = \infty$ corresponds to a direction parallel to the beam axis. Pseudorapidity is convenient for a number of reasons, including the fact that it is nearly Lorentz invariant under boosts along the $\hat{z}$-axis. We say that it is "nearly" Lorentz invariant as this is only true for massless particles. However, at the LHC, the transverse momentum of a given particle is typically sufficiently larger than the mass ($p_T >> m$) such that the pseudorapidity is approximately Lorentz invariant.

Much of the reason for CMS's 20m of length in the direction of the beam axis is motivated by the goal of hermeticity. The forward calorimeter (described in greater detail in Sec. 4.2.5) provides coverage up to pseudorapidities of $|\eta| \leq 5.0$. Pairing the extensive range in pseudorapidities with the CMS detector's complete coverage in the $\hat{\phi}$-direction, the CMS detector is nearly hermetic, aiding the resolution of missing transverse energy and consequently the ability to infer the presence of undetected particles. The exact coverage of each of the detector subcomponents is discussed in greater detail in the following subsections.

4.2.2 Solenoid

The solenoid installed in the CMS detector is over 12 m in length and 6m in diameter, capable of providing a 4 T magnetic field. The purpose of such a strong magnetic field is to bend the trajectories of charged particles (as illustrated in Fig. 4.4, allowing CMS to measure their momentum, mass, and charge. Fig. 4.4 depicts the three major classes of particles which have their trajectories curved by the magnetic field produced by the solenoid: an electron (red), a charged hadron (green), and a muon (blue). Measurements of the momenta of electrons are also aided by the ECAL (described in Sec. 4.2.4), those of charged hadrons are also aided by the HCAL (described in Sec. 4.2.5), and those of muons are also aided by the muon system (described in Sec. 4.2.6).

The solenoid is installed around the the tracker (Sec. 4.2.3), the ECAL (Sec. 4.2.4), and the HCAL

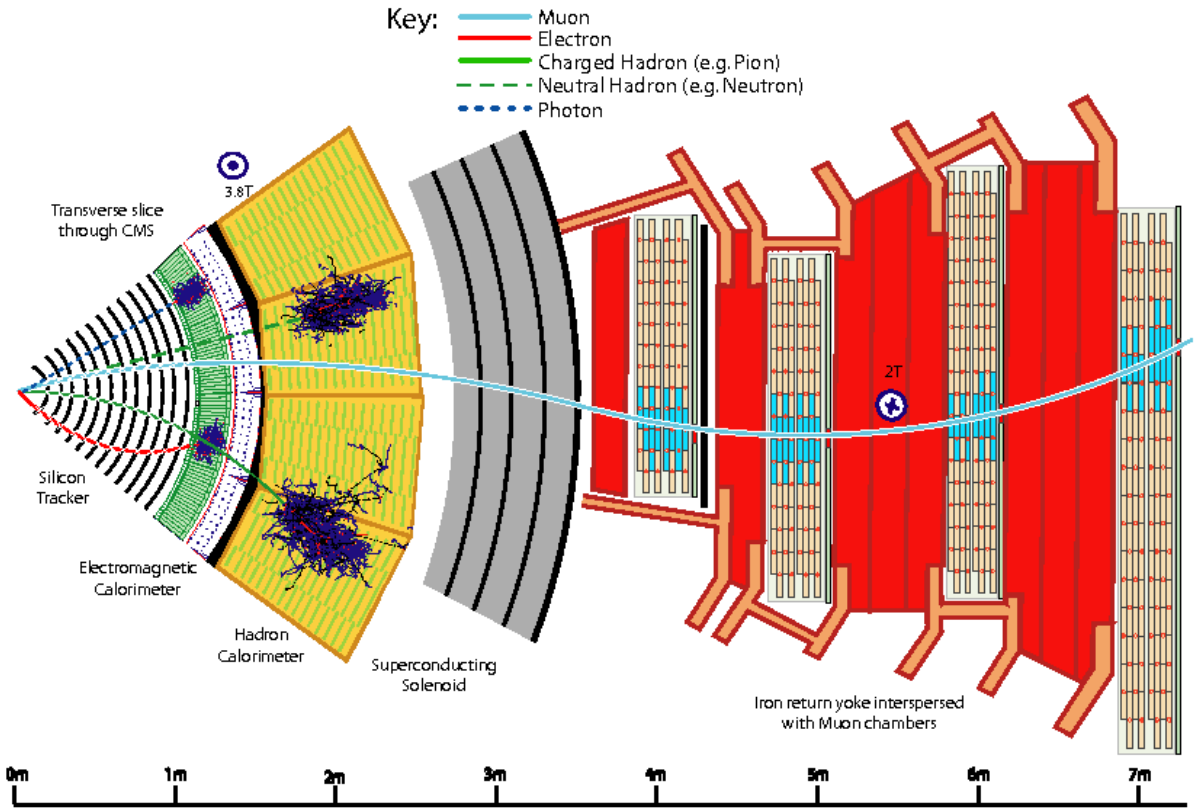

Figure 4.4: Depiction of a transverse slice of the CMS detector, along with trajectories of particles of different types. Taken from [42].

(Sec. 4.2.5) which is why the solenoid must be so large. In order to support the massive current (over 10^4 A) required for the magnetic field, the solenoid is constructed with superconducting Niobium-Titanium (NbTi), and its temperature must be kept sufficiently low to ensure superconductivity of the NbTi.

4.2.3 Tracker

The innermost component of the CMS detector is the silicon tracker, and its primary aim is to provide the precise reconstruction of charged particles and secondary vertices (an inelastic pp collision is deemed a "primary vertex" while decays of particles produced from a primary vertex are deemed "secondary vertices"). The tracker is nearly 6 m in length and 2.5 m in diameter, composed of an inner pixel detector with three layers ranging from 4-10 cm and an outer silicon strip tracker with ten layers ranging to 1.1 m. Both the pixel detector and the silicon strip tracker are accompanied by endcap disks on either end of the barrel, extending the pseudorapidity coverage to $|\eta| \leq 2.5$. Between the data-taking periods corresponding to 2016 and 2017, the inner pixel detector was upgraded [14], extending the coverage of the tracker up to $|\eta| \leq 3.0$. As shown in Fig. 4.5, the number of fake tracks, the impact parameter resolution, and the vertex resolution are each improved as well, resulting in an approximately 10% improvement in the b-tagging efficiency for a fixed fake rate.

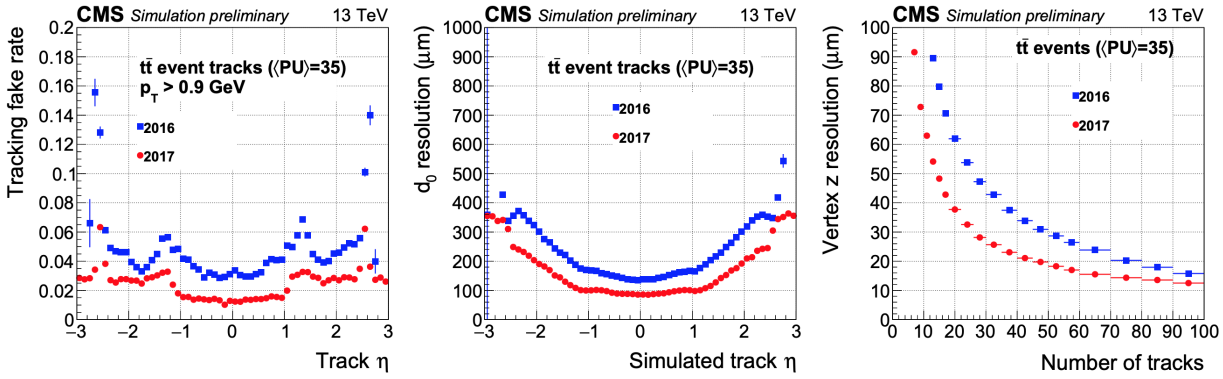

Figure 4.5: Comparison of tracker performance before and after the upgrade to the pixel detector, performed in between the 2016 and 2017 data-taking periods. Taken from [14].

The primary design considerations for the tracker include the following:

- Ability to reconstruct a large number of charged particles in each bunch crossing, with $O(1000)$

charged particles expected from a single bunch crossing at the LHC design luminosity of $\mathcal{L} = 10^{34}$ cm^{-2} s^{-1} (corresponding to about 20 individual pp interactions).

- Ability to reconstruct charged particles with precise temporal resolution, with bunch crossings separated by a distance corresponding to 25 ns.

- Minimal interaction of photons with the tracker material, as precise measurements of photons are vital to studying Higgs physics in the H $\rightarrow \gamma\gamma$ decay channel.

The first two considerations are in direct conflict with the third consideration: a tracker with high granularity and fast response implies large power density of electronics, which requires efficient cooling. This increases the material budget of the tracker, increasing the chances of bremsstrahlung and photon conversions, which in turn degrade the ECAl's photon energy resolution. An acceptable compromise providing both execllent tracking and excellent photon resolution was achieved with the tracker design depicted in Fig. 4.6.

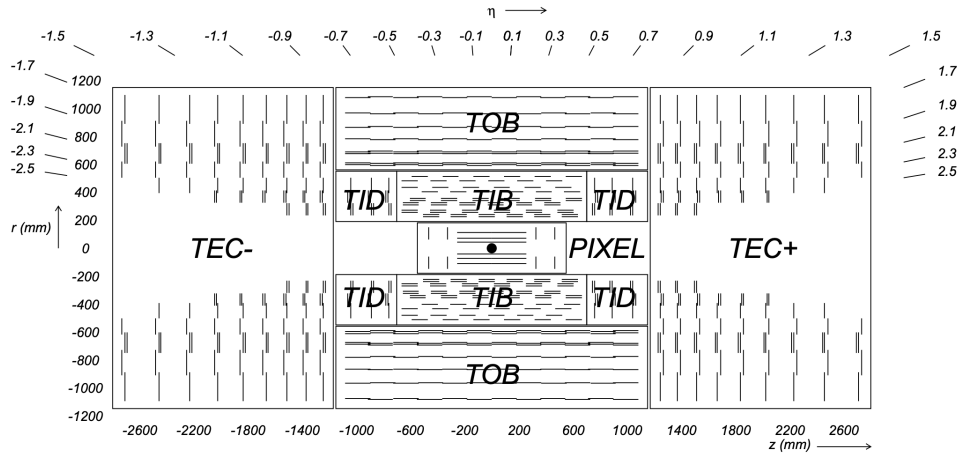

Figure 4.6: Schematic of the CMS tracker from a cross-sectional viewpoint. TIB, TOB, TID, and TEC represent the tracker inner barrel, tracker outer barrel, tracker inner disk, and tracker endcap components, respectively. Taken from [33].

The material budget for the CMS tracker is shown in Fig. 4.7, showing the thickness of the tracker material in terms of the characteristic radiation lengths X_0 (for electromagnetic particles, e.g. electrons and photons) and characteristic nuclear interaction lengths λ_I (i.e. for hadrons). The tracker thickness in terms of both radiation lengths and nuclear interaction lengths is lowest in the most central part of the barrel and

increases in the more forward components, accounting for one of the reasons that CMS achieves better energy resolution for very central particles.

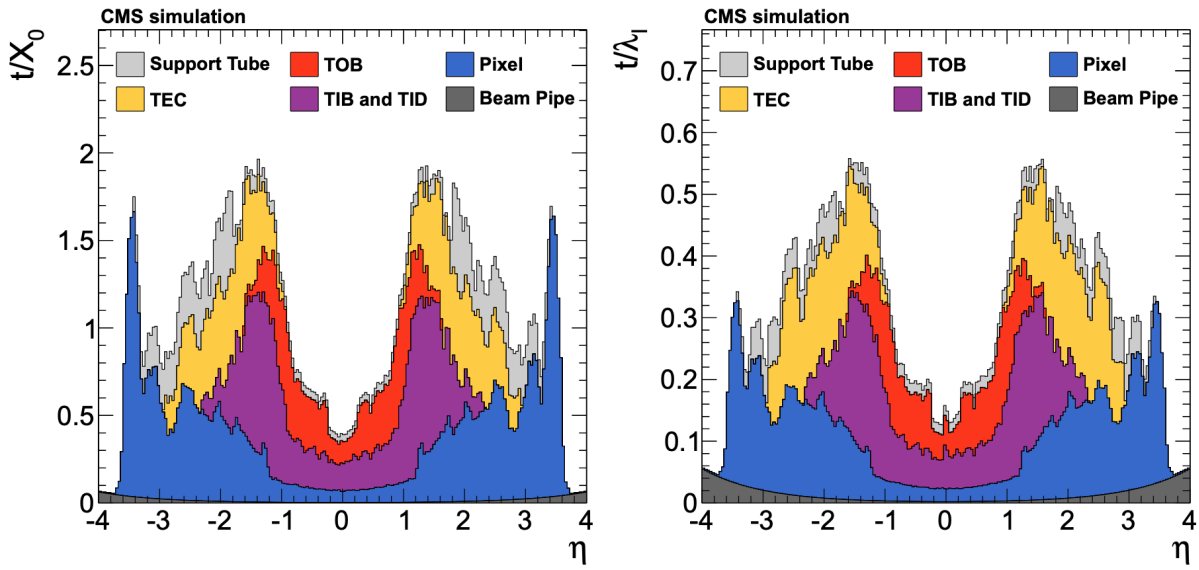

Figure 4.7: The material budget for the CMS tracker shown for both the characteristic radiation lengths of electromagnetic interactions (left) and the characteristic nuclear interaction lengths of hadronic interactions (right), with the contributions of each of the tracker subcomponents shown individually. Taken from [37].

The tracking effieciency achieved by the CMS tracker is shown in Fig. 4.8, for muons, pions, and electrons as a function of their transverse momentum. In general, the tracker achieves higher efficiency for muons than for electrons or pions, as electrons are more likely to emit radiation via bremsstrahlung and charged pions may undergo nuclear interactions with the tracker material. Energy resolution of high p_{T} ($O(TeV)$) muons is assisted by the muon system, as shown in Fig. 4.10.

4.2.4 Electromagnetic Calorimeter

The primary goal of the CMS electromagnetic calorimeter is to preicsely measure the momenta of photons, especially important for studying the properties of the Higgs boson in the $\mathrm{H} \rightarrow \gamma\gamma$ channel. Along with the tracker, the ECAL also assists in measurements of electrons which radiate a significant fraction of their energy through via bremsstrahlung as they pass through the detector. Other charged particles, like charged pions and muons, interact relatively negligibly with the ECAL, as they emit a much smaller fraction of their energy via bremsstrahlung (due to the fact that they are much more massive than electrons).

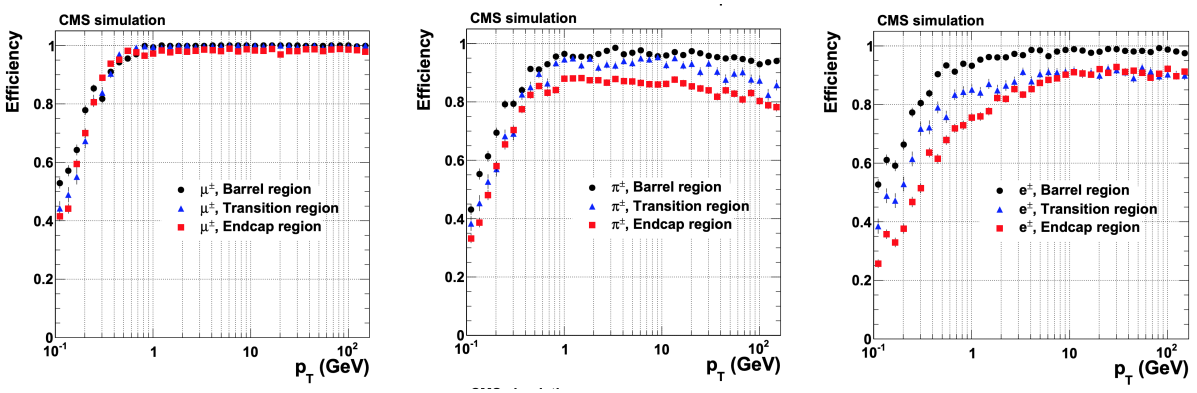

Figure 4.8: Tracking efficiency as a function of p_{T} for muons (left), charged pions (middle), and electrons (right), shown separately for the barrel (black), transition region (blue), and endcap (red). Taken from [37].

The CMS ECAL is composed of over 6×10^4 lead tungstate ($PbWO_4$) crystals in the barrel and over 7×10^3 crystals in each of the endcaps, with the barrel component (EB) providing coverage up to $|\eta| \leq 1.479$ and the endcap components providing coverage from $1.479 \leq |\eta| \leq 3.0$ [33].

Lead tungstate is an attractive choice for the ECAL crystals for several reasons, including its high density (8.28 g/cm^3), short radiation length (0.89 cm), and small Moliere radius (2.2 cm), defined as the average size of a cylinder containing 90% of an incident photon or electron's energy. The short radiation length allows the CMS ECAL to be compact, and the small Moliere radius allows for better spatial resolution of incident photons and electrons. The former is important from a practical and financial point of view, as the ECAL must be placed within the HCAL and solenoid, while the latter is important as better spatial resolution allows for better diphoton and dielectron invariant mass resolution (assisting with idenitfying $\mathrm{H} \to \gamma\gamma$ and $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ events).

Two additional attractive properties of lead tungstate include its short scintillation time and resistance to radiation damage. The scintillation time must be on the order of the bunch crossing time at the LHC (25 ns) so that ECAL deposits from consecutive crossings can be distinguished from each other (a short bunch crossing time is desirable as it results in increased integrated luminosity). Indeed, about 80% of light is emitted within 25 ns within the CMS ECAL [32], allowing for high temporal resolution in the high luminosity conditions of the LHC. Due to the high particle flux, radiation damage to detector components is inevitable; this results in wavelength-dependent loss of light transmission [32]. Although

lead tungstate is particularly radiation-hard, the damage must still be tracked and corrected for by injecting laser light and monitoring the transparency of crystals.

Further details of the reconstruction of photons are described in Sec. 5.4.

4.2.5 Hadronic Calorimeter

The CMS Hadronic Calorimeter (HCAL) is particularly important for measuring the momenta of neutral hadrons, the only detector subcomponent which is able to do so. It also assists in the momenta measurements of charged hadrons, though the tracker is typically much more effective for this purpose. Precisely measuring the momenta of hadrons allows for good energy resolution of hadronic jets – this is important for constructing a reliable estimate of the missing transverse momentum in a given event. Conversely, poor resolution of jet energies would result in poor resolution of the missing transverse momentum, degrading the experiment's ability to identify events in which there is true missing transverse momentum, either from neutrinos or yet-to-be discovered particles which do not interact with the CMS detector. The CMS detector is typically able to achieve a mometum resolution of around 10% [32] for hadronic jets, using a combination of information from the various detector subcomponents.

The CMS HCAL is composed of two primary components: barrel (HB), covering $|\eta| \leq 1.4$ and endcap (HE), covering $1.3 \leq |\eta| \leq 3.0$. The barrel component also contains a "tail catcher" placed outside the solenoid (HO), which covers up to $|\eta| \leq 1.26$. Finally, a forward calorimeter (HF) specializes in measuring very forward particles, up to $|\eta| \leq 5$ [32]. The various components of the HCAL and their pseudorapidity coverage are depicted in Fig. 4.9.

Conceptually, the HCAL aims to measure hadron energies by placing a high density of atomic nuclei, with which hadrons are likely to undergo nuclear (i.e. strong) interactions. The products of these nuclear interactions are then measured with plastic scintillators and the HCAL is then able to reconstruct the energy of the original hadron.

In practice, this is achieved through brass alloy absorber plates with plastic scintillators interspersed between them [32]. To decrease the probability that hadrons "punch through" [10] the HCAL and leave the detector unmeasured, multiple layers of brass and plastic scintillators are utilized, with 17 layers total.

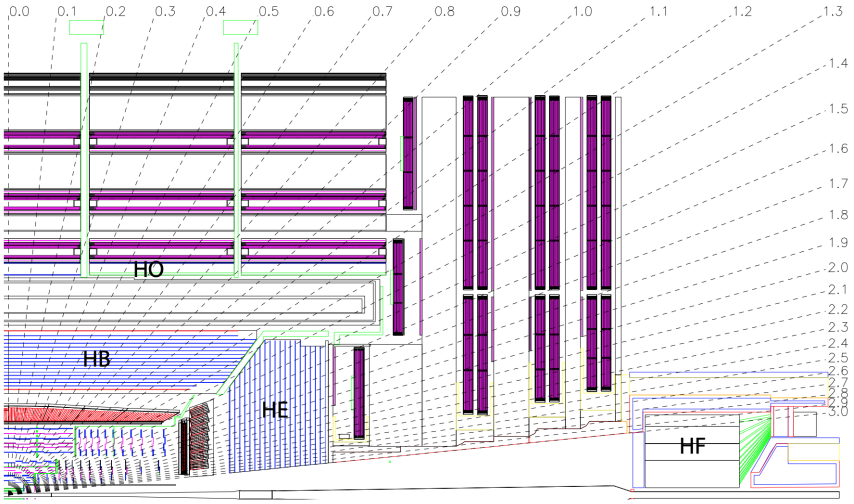

Figure 4.9: Schematic of subcomponents of the CMS HCAL, along with their pseudorapidity coverage. Taken from [33].

4.2.6 Muon System

The outermost detector subcomponent of the CMS detector is the muon system, placed outside the solenoid. The primary goals of the muon system are to identify the presence of muons, measure their momenta, and provide the ability to trigger on events with muons [33]. Installing a second tracker outside the solenoid would be ideal, but in practice, this would be far too expensive. A more economical solution is a detector using gas-filled chambers, exploiting the fact that muons traversing through the gas will ionize it.

The muon system is made of three components:

1. Drift tube (DT) chambers, which cover the region $|\eta| \leq 1.2$.

2. Cathode strip chambers (CSC), which cover the region $0.9 \leq |\eta| \leq 2.4$.

3. Resitive plate chamber (RPC) system, which is installed in both the barrel and endcap regions, covering $|\eta| \leq 1.6$.

The DT chambers are well-suited to the barrel, where the muon rate and background rate are lower, while the CSCs, having better radiation resistance [33], are better suited to the endcaps, where the muon rate and background rate are higher. The RPCs specialize in providing the ability to trigger on events with

muons: although their position resolution is coarser than that of the DT chambers or CSCs, they provide excellent time resolution, allowing consecutive bunch crossings to be distinguished from one another.

The muon system is especially helpful in assisting in the measurement of high p_{T} O(TeV) muons. As charged particles' energies are primarily determined through the curvature of their tracks, this presents a challenge for especially high energy particles: higher energy particles curve less and there is greater uncertainty on their respective momentum measurements. Fig. 4.10 shows the momentum resolution for muons as a function of p_{T}. While both the tracker and muon system are capable of measuring the momenta of muons of a wide range of energies, the muon system provides the greatest improvement to the overall resolution at very high p_{T}.

Beyond just improving the momentum resolution of high energy muons, the muon system is useful in the measurement of lower energy muons as well. As the muon system and the tracker provide independent measurements of muons, this allows for fault-finding and cross-checks of each detector component.

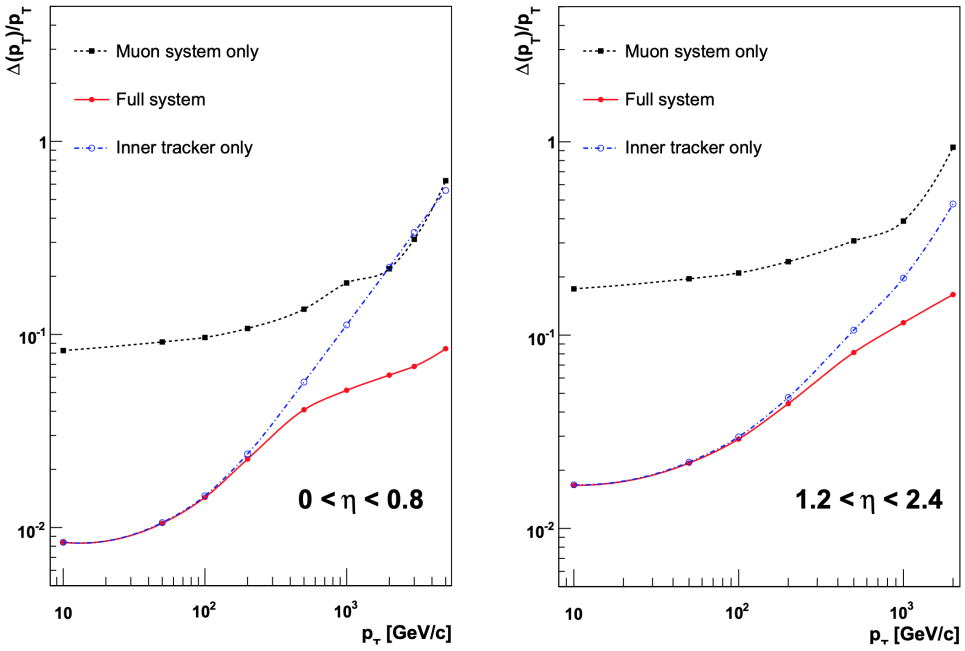

Figure 4.10: Fractional momentum resolution for muons reconstructed by the CMS detector, shown for reconstructions using the inner tracker only (blue), the muon system only (black), and the combination of measurements from both subdetectors (red). The muon system improves significantly the momentum resolution of O(TeV) muons. Taken from [33].

4.2.7 Trigger System

At the LHC's design luminosity of 10^{34} cm^{-2} s^{-1}, the pp interaction rate is greater than 1 GHz [43]. It is neither necessary nor feasible to store the data of each of the more than 10^9 events per second, as the vast majority of pp interactions are neither interesting nor useful in light of the goals of the CMS physics program. It is not feasible in the sense that the CMS readout electronics impose an upper limit on the event rate of about 100 kHz, such that the vast majority of events must be thrown away. Is is not necessary in the sense that most physics processes of interest have cross sections many orders of magnitude smaller than the nominal pp interaction cross section.

In order to select only the most interesting events to store for later analysis, a two-tiered trigger system is employed by the CMS detector. The first level (L1), is implemented on custom hardware, and reduces the event rate by a factor of about 10^4, from around 1 GHz to around 100 kHz. The second level (HLT), is implemented in software, and further reduces the event rate to a typical rate of 400 Hz.

L1 Trigger

The L1 trigger combines information from the ECAL, HCAL, and muon system to decide with a fixed latency of 4 μs of a collision if the event should be accepted or not. A schematic overview of the trigger system is shown in Fig. 4.11. Events which pass the L1 trigger are then evaluated by the high-level trigger system to make a final decision on if the full event data will be stored.

High-Level Trigger

In contrast to the L1 trigger, the high-level trigger (HLT) system accepts events for storage based off a more complete, nearly offline-quality reconstruction of the event. Details of event reconstruction, e.g. the particle flow algorithm, are described in Sec. 5.2. Near offline-quality event reconstruction is achieved in practice through the use of a "processor farm", a system of over 10^4 CPUs working in parallel to efficiently reconstruct each event. The HLT takes significantly longer to "think" about each event, with an average processing time on the order of 100 ms per event (compare to 4 μs for L1).

The high-level trigger paths used for the $t\bar{t}H$ ($H \to \gamma\gamma$) analysis are the following:

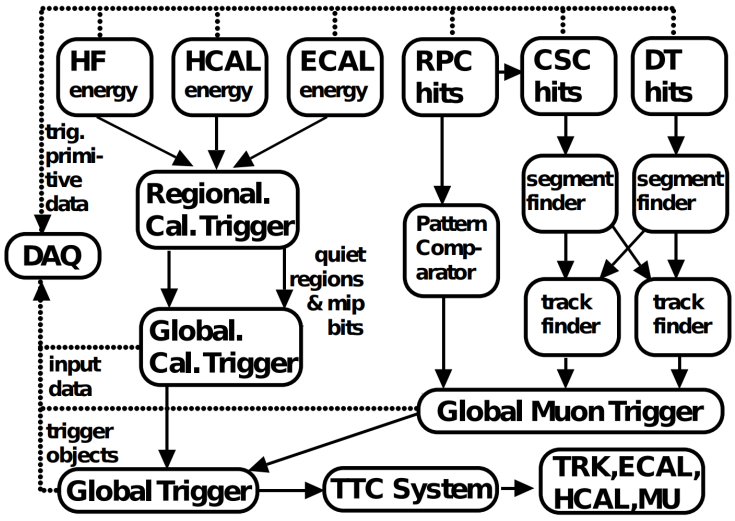

Figure 4.11: Schematic overview of the CMS L1 trigger system. Information from the calorimeters are first processed regionally and then a global calorimeter decision (GCT) is made. Similarly, information from the various components of the muon system are first processed regionally and then at a global level (GMT). The information from the global calorimeter trigger and global muon trigger are combined in a single global trigger (GT) which performs the final decision on whether to store the event. Taken from [43].

- 2016: `HLT_Diphoton30_18_R9Id_OR_IsoCaloId_AND_HE_R9Id_Mass90*`
- 2017: `HLT_Diphoton30_22_R9Id_OR_IsoCaloId_AND_HE_R9Id_Mass90*`
- 2018: `HLT_Diphoton30_22_R9Id_OR_IsoCaloId_AND_HE_R9Id_Mass90*`

Conceptually, each of these triggers requires the presence of two photons with leading (subleading) transverse momenta of 30 (18/22) GeV, imposes requirements on the photons' shower shape variables (described in further detail in Sec. 5.4), and requires a diphoton invariant mass of at least 90 GeV. The photon selection requirements described in Sec. 5.4 are defined to be similar (and slightly stricter) than those of the trigger paths listed here. Still, the efficiency of the trigger in simulation does not necessarily match that in data. The efficiency is measured in data with $Z \to e^+e^-$ events and the efficiency in simulation is accordingly corrected as a function of the transverse energy, pseudorapdidity, and shower shape variable R_9 (defined in Sec. 5.4), with the uncertainty on the scale factor taken as a systematic uncertainty (described in Sec. 6.8).

4.3 Acknowledgements

The figures shown in Chapter 4 are taken from the following results: “Performance and track-based alignment of the Phase-1 upgraded CMS pixel detector”, *CMS-CR-2017-256* (2017), “CMS Luminosity – Public Results”, *https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults* (2020), “The CMS Experiment at the CERN LHC”, *Journal of Instrumentation* (2008), “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *Journal of Instrumentation* (2014), and “The CMS trigger system”, *Journal of Instrumentation* (2017), and were produced by other members of the CMS Collaboration.

Chapter 5

Event Reconstruction and Selection

5.1 Introduction

The raw data recorded by the CMS detector for a single bunch crossing is typically not yet suitable for high-level physics analysis. Several layers of abstraction transform the data from the raw detector readout into high-level physics objects. The first layer of abstraction uses the particle flow (PF) algorithm [41], which combines information from each of the CMS subdetectors (the tracker, ECAL, HCAL, and muon system) in an attempt to reconstruct every particle in the event (PF candidates). This first step is common to most physics analyses performed within the CMS experiment and is described in Section 5.2. Physics objects are then refined further by placing quality requirements on the PF candidates. This second step is often analysis-specific and the details are dictated by the individual needs of the given analysis. The details of physics object definition specific to the $t\bar{t}H$ analysis are described in Sections 5.4–5.7.

5.2 The Particle Flow Algorithm

The PF algorithm forms the basis of event reconstruction for almost all physics analyses in CMS. It attempts to individually reconstruct every reconstructable particle (i.e. all particles except neutrinos) in a given event, doing so by combining information from each of the CMS subdetectors. The fundamental

inputs to the PF algorithm are tracks, originating from the tracker and muon system, and calorimeter clusters, obtained from the ECAL and HCAL. Combining each of these pieces of information results in more accurate reconstruction of individual particles (and by extension, hadronic jets). Notably, jets built from PF candidates contain 95–97% of the jet energy, compared 60–80% for jets built solely from calorimeter clusters. The angular resolution of jets is also improved by a factor of 2–3 [9].

Tracks are reconstructed in three stages using a procedure based on Kalman-Filtering. First, seeds are generated using a small number of hits compatible with a track. Second, compatible hits in other tracker layers along the trajectory of this track are identified. Finally, a fit is performed with each of the hits in order to determine the properties associated with the candidate particle: origin, transverse momentum, and direction. In order to prevent the identification of fictitious tracks, a set of strict quality criteria are imposed upon the candidate tracks. Tracks are required to have a minimum amount of transverse energy ($p_{\mathrm{T}} > 0.9$ GeV), must be seeded from hits in at least two consecutive layers in the pixel detector, must have at least 8 total hits, and may be missing hits in at most 1 layer. While these requirements allow the tracking algorithm to maintain a low misidentification rate, they also exclude around 20-30% of charged hadron tracks with $p_{\mathrm{T}} > 1$ GeV. Charged hadron tracks frequently do not pass the track requirements because of their high probability to undergo nuclear interactions with the beam pipe or detector material before reaching the outer tracker. Muons with $p_{\mathrm{T}} > 1$ GeV, on the other hand, have a negligible probability of interacting before reaching the outer tracker and consequently have a much higher tracking efficiency of 99%.

Clusters in the ECAL and HCAL are built by first identifying a seed, a cell with energy larger than a certain threshold and also larger than the energy of neighboring cells. The ECAL and HCAL cells are of a size such that a typical particle interacting with either calorimeter will leave its energy distributed across multiple neighboring cells. If two particles enter a calorimeter close to each other, some cells may receive energy contributions from both particles. To address scenarios like this, topological clusters are formed by joining cells that are "once-removed" from the seed (i.e. they share at least one neighbor with a seed) with an energy larger than a certain threshold, not necessarily the same as the seed threshold. The energy of each cluster within a topological cluster is then determined by a maximum-likelihood fit to a sum of Gaussians. The number of Gaussians is the number of seeds in the topological cluster and the parameters

to fit are the energy of each cluster (the amplitude of each Gaussian), A_i, and the position (η_i, ϕ_i) of each cluster. The initial values for the energy and position of each cluster are chosen as the energy and position of the corresponding seed. The width for each Gaussian is fixed and depends on the specific calorimeter.

A link algorithm then combines compatible tracks and clusters, using the full set of information to reconstruct five different types of particles:

- **Muons:** reconstruction based on tracks from both the inner tracker and the muon system.
- **Electrons:** tracks for electrons are reconstructed with a Gaussian-sum filter (GSF) that allows for sudden loss of energy due to bremsstrahlung. GSF tracks linked to an ECAL supercluster are then chosen as electron candidates.
- **Photons:** reconstruction based on ECAL superclusters which are *not* linked to GSF tracks.
- **Charged hadrons:** reconstruction based on tracks that are linked to both ECAL and HCAL clusters.
- **Neutral hadrons:** reconstruction based on HCAL clusters which are *not* linked to tracks or ECAL superclusters.

The PF algorithm does not attempt to distinguish between different types of charged or neutral hadrons.

5.3 Vertex Reconstruction

For data collected by the CMS detector during Run II of the LHC, the mean number of of primary interactions was $\mu = 29$. The primary vertex is taken to be the one with the largest value of the sum of the squares of the transverse momenta of the physics objects [58]. In other words, it is chosen as:

$$\arg\max_{i \in I} f(i), \qquad f(i) \equiv \sum_{j=1}^{N_i} (p_{\mathrm{T}}^{j})^2 \tag{5.1}$$

where the sum runs over the N_i physics objects associated with the i-th vertex.

For many $\mathrm{H} \to \gamma\gamma$ analyses, this prescription of choosing the primary vertex is suboptimal, as it relies on charged tracks linked to the primary vertex and as photons are neutral particles, they do not leave

tracks. However, for the $t\bar{t}H$ analysis in which additional jets and leptons are expected in the final state, this choice of primary vertex is found to be the correct choice for $> 99\%$ of $t\bar{t}H$ events, so no further vertex selection criteria are employed.

5.4 Photon Reconstruction

The photons selected for use in this analysis are initially taken from the PF photon candidates described in Sec. 5.2. In further reconstruction of photons for $H \to \gamma\gamma$ analyses, there are several challenges to overcome.

First, the energy estimate provided by the PF algorithm generally has inherent bias. The bias in the photon energy estimates are corrected for with a regression technique which utilizes $Z \to e^+e^-$ events in which the electrons have been reconstructed as photons. The procedure, described in Sec. 5.4.2 exploits the fact that the mass of the Z boson is known [55] to good precision, ensuring that the $m_{e^+e^-}$ distribution is centered around m_Z. An additional smearing procedure corrects for the differences in energy resolution between data and simulation. This step is vital as simulation, rather than actual data, is used to construct the models of SM Higgs boson production modes, including $t\bar{t}H$.

A second challenge in $H \to \gamma\gamma$ analyses is distinguishing between "prompt" photons and "fake" photons. Prompt photons are those produced in the decay of the Higgs boson or from the primary hard inelastic scattering process and are typically the objects of interest of physics analyses. Fake photons are those produced in hadronic jets, usually through the decay $\pi^0 \to \gamma\gamma$, and are (typically) of less interest. Prompt photons tend to be more isolated in the detector and tend to have different shower shapes in the ECAL. These differences are exploited through the used of a BDT trained to distinguish between prompt and fake photons.

The photon ID BDT, described in Sec. 5.4.4, uses a variety of high-level variables describing the photon's kinematics, shower shape, and isolation. The shower shape & isolation variables and the photon ID BDT are used to select photons of interest for the $t\bar{t}H$ analysis, rejecting as many fake photons as possible while retaining a high efficiency on prompt photons. As simulation is used to construct the models of SM Higgs boson production modes, it is important that the shower shape & isolation variables

and the photon ID BDT are well-described in simulation. A chained quantile regression method, described in Sec. 5.4.3, corrects the distributions of these features in simulation. The multitude of high-level variables used to describe photons are defined in Sec 5.4.1.

5.4.1 Variable Definitions

The variables defined in this section are used for studying the scale & resolution of photon energy reconstruction, discriminating between prompt and fake photons, selecting the photons to be used in analysis, or a combination of the three.

General

- *Conversion-safe electron veto*: a flag rejecting the photon candidate if there is a track with at least one hit in the inner layer of the pixel detector pointing to the photon supercluster *and* the track is not matched to a vertex.
- *Pixel seed veto*: a flag rejecting the photon candidate if any track with at least two hits points to the photon supercluster. In general, the pixel seed veto provides a more severe rejection of electrons but excludes a larger fraction of photons.
- ρ: the median energy density per unit area in the event.

The pixel seed veto is, in general, much stricter than the conversion-safe electron veto in rejecting electrons imitating photons. For most $H \to \gamma\gamma$ analyses, electrons are not a large source of fake photons and so the conversion-safe electron veto is used for its greater efficiency on real photons. However, the $t\bar{t}H$ analysis has a significant background component coming from $t\bar{t} + X$ events in which an electron from a $W \to e\nu_e$ decay is reconstructed as a photon. To target this background, the pixel seed veto is employed (as a training variable in the BDT used to define signal regions).

Shower shape variables

- $E_{2\times2}/E_{5\times5}$: the ratio of energies between 2×2 and 5×5 matrices of ECAL crystals. The 2×2 matrix is defined as that containing the two most energetic crystals, the 5×5 matrix is defined as

that centered on the supercluster seed crystal.

- $cov_{i\eta i\phi}$: the covariance of the crystal values of the 5×5 matrix centered on the supercluster seed crystal.

- $\sigma_{i\eta i\eta}$: the standard deviation along the η direction of the electromagnetic shower (expressed in terms of crystal cells).

- R_9: $E_{3\times3}/E_{\mathrm{SC}}$, where $E_{3\times3}$ is the 3×3 crystal matrix centered on the supercluster seed crystal and E_{SC} is the total energy of the supercluster.

- σ_η: the standard deviation of crystal η values in the supercluster, with each crystal's contribution weighted by the logarithm of its energy.

- σ_ϕ: the standard deviation of crystal ϕ values in the supercluster, with each crystal's contribution weighted by the logarithm of its energy.

- *Preshower* σ_{RR}: the standard deviation of the shower spread in the x and y directions of the preshower detector (defined only in the endcap).

The shower shape variables are useful in both the regression of photon energy and the discrimination between prompt photons and hadronic jets misidentified as photons ("fake" photons).

Isolation variables

- I_{ph}: the transverse energy sum of all other PF photons in a cone size $R = 0.3$ around the photon candidate.

- $I_{\mathrm{ch,\ sel}}$: the transverse energy sum of all PF charged hadrons in a cone size $R = 0.3$ around the photon candidate, measured with respect to the selected vertex.

- $I_{\mathrm{ch,\ wst}}$: the transverse energy sum of all PF charged hadrons in a cone size $R = 0.3$ around the photon candidate, measured with respect to the worst-fit vertex.

- H/E: the energy sum from the HCAL towers within a cone of $R = 0.15$ around the supercluster, divided by the energy of the supercluster.

- I_{tk}: the transverse momentum sum of all tracks in a cone size $R = 0.3$ around the photon candidate. Tracks within an inner cone size $R = 0.04$ are *not* included in this sum, effectively making the sum over a "hollow" cone.

The isolation variables are useful in discriminating between prompt photons and hadronic jets misidentified as photons.

5.4.2 Energy Scale & Resolution Corrections

The energy measurements of individual ECAL channels are first corrected, as described in Sec. 4.2.4. Once the ECAL energy measurements are calibrated, a multivariate regression technique [38] is used to further correct the energy of photon candidates. After the multivariate regression is applied, further energy scaling is applied to correct for any time or position dependent bias effects in the energy measurements in data. Lastly, a smearing procedure is applied to simulation such that the energy resolution from simulation matches that in data.

The multivariate regression attempts to correct for many systematic sources of bias in the supercluster energy measurement. For example, the supercluster may not capture all of the electromagnetic shower from a given photon (thereby underestimating its energy). These sources typically relate to the details of the ECAL geometry (i.e. the geometrical arrangement of crystals and voids between crystals) and the probability of interaction with detector material before reaching the ECAL. Among the training inputs for the regressor are the supercluster coordinates (η and ϕ), shower shape variables, information about the seed crystal of the supercluster, and variables (ρ and number of vertices) describing the pileup conditions of the event.

The regressor attempts to predict the form of the probability distribution function for $E_{\text{true}}/E_{\text{raw}}$. The functional form for the probability distribution function for $E_{\text{true}}/E_{\text{raw}}$ is chosen as a Gaussian with two power law tails. The regressor simultaneously predicts the true energy and the uncertainty in the energy measurement for a given photon, by returning values of the parameters for the functional form. The true energy, E_{true} is taken as the most probable value of the probability distribution function returned by the regressor. The energy resolution is determined from the width of the probability distribution function. The sum of probability distribution functions returned by the regressor are compared to the actual $E_{\text{true}}/E_{\text{raw}}$

distribution (in simulation) in Fig. 5.1.

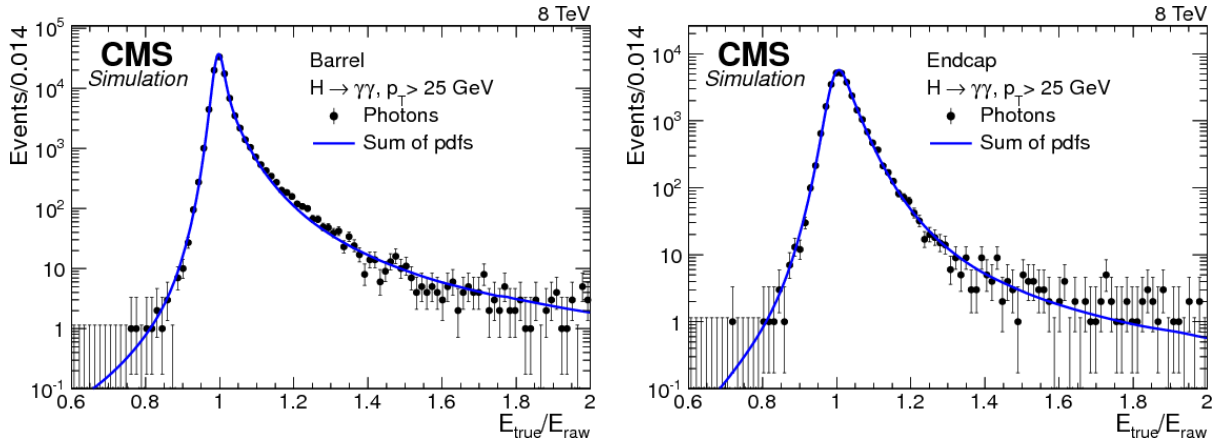

Figure 5.1: Sum of probability distribution functions returned by the regressor (blue) compared with the actual $E_{\text{true}}/E_{\text{raw}}$ distribution in simulation (black). Taken from [38]

.

Next, an energy scaling procedure is applied on data to correct any non-uniformity in time or position (η) in the energy measurements. Sources of bias include damage to the ECAL due to radiation, for example. Since the detector response changes as a function of η (radiation damage is not uniform in η) and as a function of time (damage is "cumulative"), the scale correction is derived in bins of η and run number. This procedure exploits the known mass of the Z boson by using an analytic fit to the invariant mass of electrons reconstructed as photons in $Z \to e^+e^-$ events in data and simulation. The functional form in the fit is the convolution of a Breit-Wigner [16] and a Crystal Ball function, with the Crystal Ball modeling both the calorimeter resolution effects and losses due to bremsstrahlung. The parameters of the Breit-Wigner are fixed to the Particle Data Group values [55] of m_Z and Γ_Z. The scale correction is calculated from the difference in the mass peaks between data and simulation:

$$\Delta P = \frac{m_{\text{data}} - m_{\text{MC}}}{m_Z} \tag{5.2}$$

While the Z mass peak initially varies by a few percent as a function of η and run number, the peak is stable after applying the scale corrections to data.

Finally, a smearing procedure is applied to the energy measurements in simulation to ensure that the energy resolution in simulation matches that observed in data. The additional smearing applied to

simulation is a Gaussian function and its properties are determined by a fit to the Z invariant mass peak. The smearings are derived with the same bins as the energy scales.

The regression, scales, & smearings are validated by comparing the invariant mass distribution of electrons reconstructed as photons in $Z \to e^+e^-$ events between data and simulation. Fig. 5.2 shows that excellent agreement between data and simulation is achieved for all three years of data-taking. An

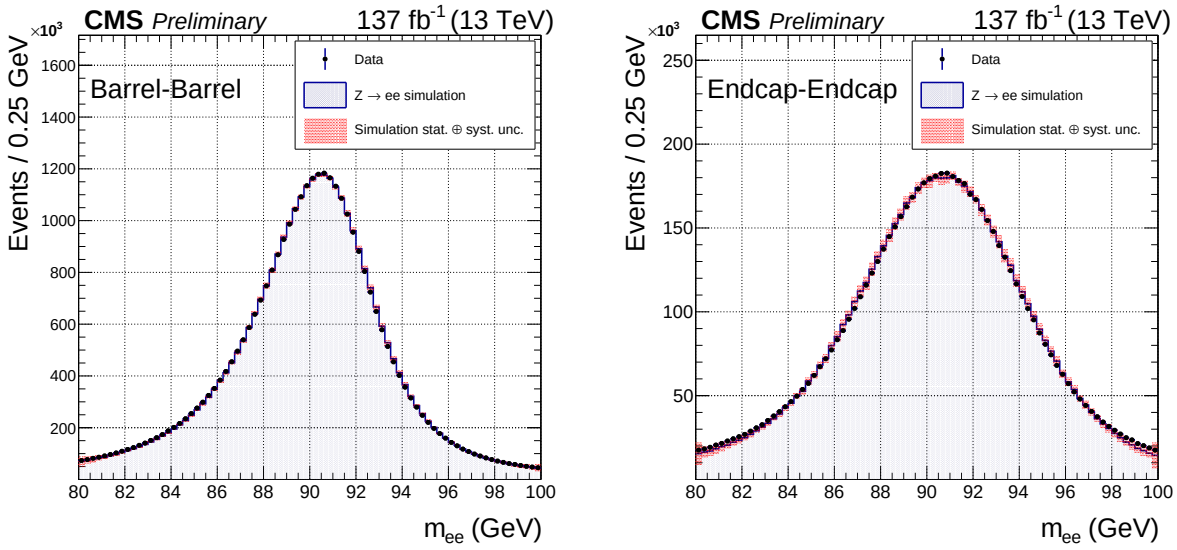

Figure 5.2: Validation of photon energy regression, scales, and smearings: comparisons of m_{ee} distributions in $Z \to e^+e^-$ events. Taken from [49].

additional, "residual", scale correction is derived simultaneously with the smearings. The residual scale correction, applied on top of the run-dependent scale corrections previously described, corrects for any remaining differences between the central value of Z mass peak in data and simulation (which match by construction of the run-dependent scale corrections) and the central value of the Z mass peak known from the well-measured mass of the Z. The run-dependent scale corrections ensure agreement between data and simulation, while the residual scale corrections ensure agreement between data, simulation and the known mass of the Z. Each of the run-dependent scale corrections, residual scale corrections, and smearing corrections range from approximately 1–3%.

5.4.3 Shower Shape & Isolation Corrections

The shower shape and isolation variables that are used in training the photon ID BDT show disagreement between data and simulation. Because of the disagreement in the input variables, disagreement between data and simulation in the photon ID BDT score is also observed. The distributions of these variables in simulation are corrected with a chained quantile regression method [119]. Each variable is corrected using separately trained BDTs, each trained to predict the conditional shape of the cumulative distribution function in both data and simulation. The value in simulation is replaced by the value corresponding to the same point on the cumulative distribution function in data. The BDTs take the photon kinematics, ρ, and the variables that have already been corrected as inputs. The variables that have already been corrected are given as additional inputs in order to better preserve the correlations between the input variables in data. After applying this method, good agreement between data and simulation in the photon ID BDT output is achieved for all three years, as seen in Fig. 5.3.

5.4.4 Photon Identification BDT

A common challenge to all $H \to \gamma\gamma$ analyses is the discrimination between prompt and fake photons.

- **Prompt photons:** photons which are external lines in the Feynman diagram of the primary hard inelastic scattering process of the event.
- **Fake photons:** all other photons. Primarily composed of hadronic jets in which a $\pi^0 \to \gamma\gamma$ decay results in the jet being misidentified as a photon.

Broadly speaking, distinguishing between the two is an easy problem. Prompt photons tend to be *isolated* in the detector, meaning there are few particles in close physical proximity. Fake photons tend not to be isolated, as they are overwhelmingly hadronic jets and therefore typically accompanied by a shower of hadronic activity. However, the characteristic scale for the cross sections of multi-jet production is many orders of magnitude larger than the characteristic scale for the cross sections of Higgs boson production. While the vast majority of hadronic jets can easily be distinguished from prompt photons, the "tails of

the distribution", in which the electromagnetic activity of a hadronic jet may be quite isolated, provide a challenging background.

A binary classifier BDT is trained to distinguish between the two cases, helping further reduce the contribution of fake photons to the background. The BDT is trained on simulation of $\gamma+$jets events. Signal events are prompt photons, taken as reconstructed photons which are matched to a generator-level photon from the hard inelastic scattering process The matching procedure is done by requiring a maximum ΔR between the reconstructed and generator-level photons. Background events are taken as all other reconstructed photons in the event. These are overwhelmingly populated by hadronic jets misidentified as photons. The BDT is trained with the (previously defined) shower shape and isolation variables. The photon ID BDT is validated with two methods, both exploiting the tag-and-probe method. The first uses $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ events in which electrons are reconstructed as photons and the second uses $\mathrm{Z} \to \mu^+\mu^-\gamma$ events in which the Z decays to two muons and one of the muons radiates a photon. Good agreement between data and simulation is found with both methods. Fig. 5.3 shows the agreement in $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ events.

5.4.5 Selection Criteria

As described in Sec. 4.2.7, the data-taking rate imposes a formidable challenge on identifying events of interest. Events in data will only enter the analysis provided they pass one of the HLTdiphoton triggers used for this analysis (described in Sec. 4.2.7). The trigger is not applied on simulation, so the preselection requirements are chosen to be slightly more stringent than those of the trigger: events (in data or simulation) passing the preselection requirements are a subset of events passing the HLT trigger.

The photon with the highest transverse momentum ("leading") is required to have $p_\mathrm{T} > 35$ GeV, and the photon with the second highest transverse momentum ("subleading") is required to have $p_\mathrm{T} > 25$ GeV. To ensure that the $m_{\gamma\gamma}$ distribution has a smooth shape, "sliding" p_T requirements of $p_\mathrm{T}/m_{\gamma\gamma} > 1/3(1/4)$ are imposed for the leading (subleading) photon. Without the sliding p_T requirements, the $m_{\gamma\gamma}$ distribution may be subject to features like peaks at lower $m_{\gamma\gamma}$: since individual photon p_T is positively correlated with $m_{\gamma\gamma}$, fixed p_T requirements reject a greater fraction of low $m_{\gamma\gamma}$ events. It is preferable to avoid these features to ensure that the $m_{\gamma\gamma}$ distribution may be fit by simple analytic functions (the background estimation method, described later, relies on this assumption).

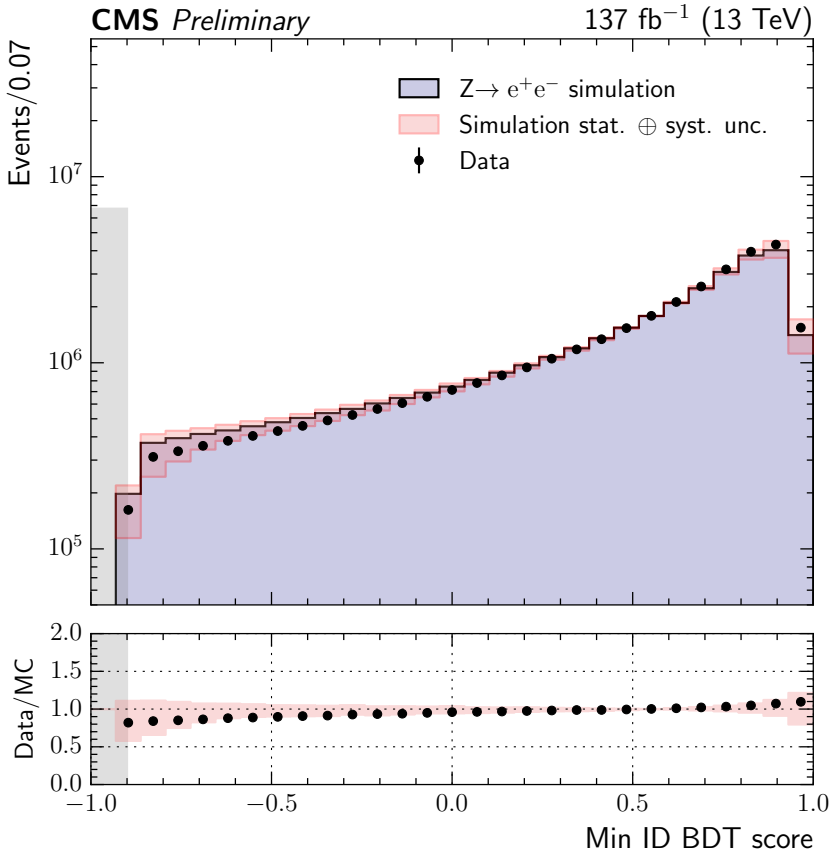

Figure 5.3: Validation of the photon ID BDT in $Z \to e^+e^-$ events: comparison of distributions in data and simulation. Taken from [49].

Photons must have $R_9 > 0.8$, $I_{\text{ch, sel}} < 20$ GeV. If a photon has $p_T > 14$ GeV and $H/E < 0.15$, it must also satisfy $I_{\text{ch, sel}}/p_T < 0.3$. Additional requirements on the isolation variables are imposed as a function of the photon location (barrel vs. endcap) and R_9, summarized in Table 5.1.

Table 5.1: Photon preselection requirements. Values are chosen to be slightly more stringent than the HLT requirements.

		H/E	$\sigma_{i\eta i\eta}$	I_{ph}	I_{tk}
Barrel	$0.5 < R_9 < 0.85$	< 0.08	< 0.015	< 4.0 GeV	< 6.0 GeV
	$R_9 \geq 0.85$	< 0.08	–	–	–
Endcap	$0.5 < R_9 < 0.9$	< 0.08	< 0.035	< 4.0 GeV	< 6.0 GeV
	$R_9 \geq 0.9$	< 0.08	–	–	–

5.5 Jet Reconstruction

As discussed in Sec. 3.2.2, color confinement prevents the existence of free quarks and gluons. A quark or gluon produced at the LHC typically undergoes hadronization and presents itself in the detector as a collection of collimated particles. Jets are built from PF candidates, using the anti-k_T clustering algorithm [18, 19] with a distance parameter of 0.4. The input PF candidates have charged hadron subtraction (CHS) applied, meaning charged hadrons associated with vertices other than the primary vertex of that event are removed. CHS reduces the contribution of particles originating from pileup vertices. Once the jets are built from PF candidates, three steps are taken to correct the jets' energies.

First, a pileup offset correction is applied to remove additional jet contributions from pileup not removed by CHS. The pileup contributions not remove by CHS are primarily charged hadrons not matched to a good vertex and PF photons. The individual jet energies are corrected by a multiplicative factor derived in simulation, parametrized in bins of jet area (A), ρ, p_T, and η. Typical values for the pileup offset corrections are shown in Fig. 5.4.

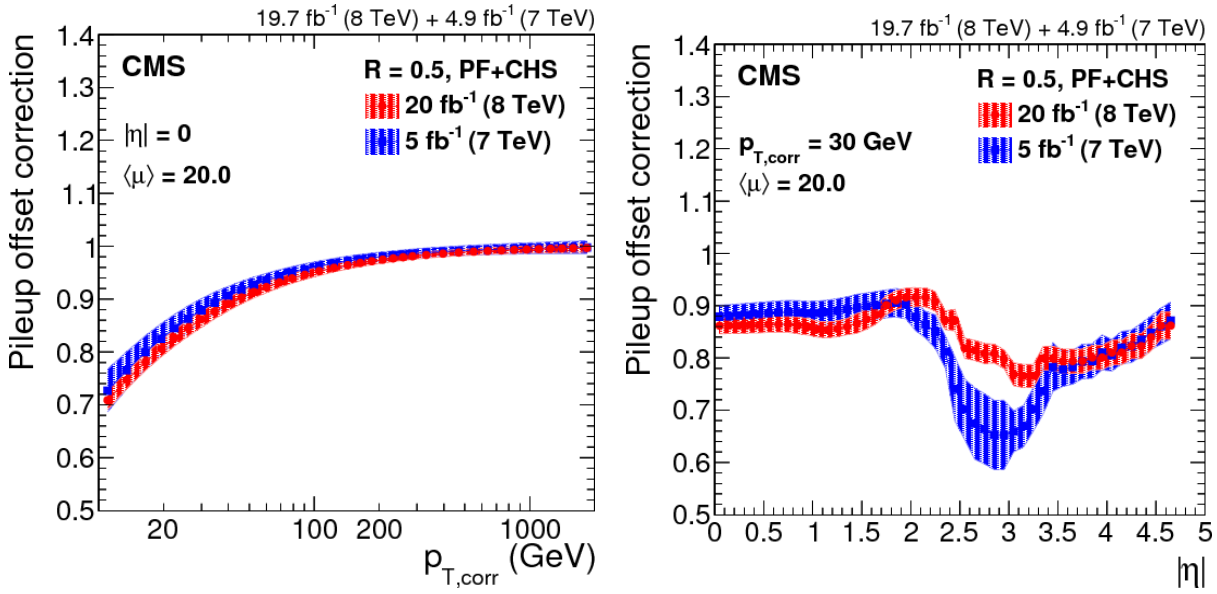

Figure 5.4: Pileup offset correction values as a function of jet p_T (left) and jet $|\eta|$ (right). Taken from [40].

Second, jet energy scale corrections designed to correct for the detector response to jets are derived in simulation, again in bins of jet area (A), ρ, p_T, and η. The goal of this step is to correct the reconstructed

jet energy to match that of the true jet energy (only available in simulation). Typical values for the jet energy scale corrections are shown in Fig. 5.5.

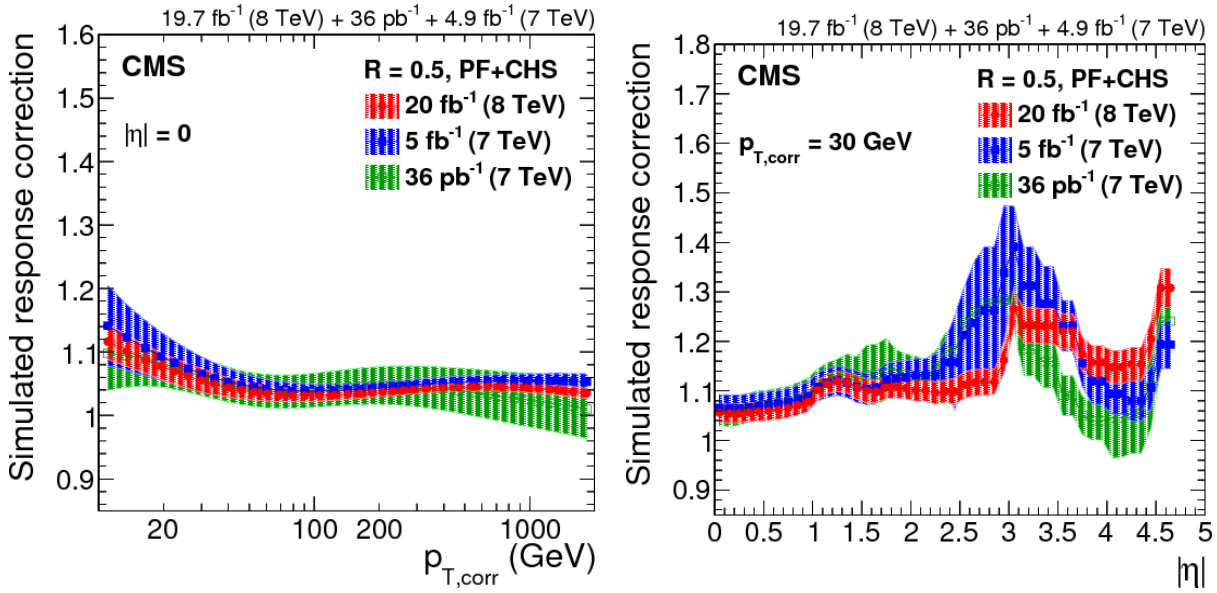

Figure 5.5: Jet energy scale correction values as a function of jet p_T (left) and jet $|\eta|$ (right). Taken from [40].

Third, remaining differences between data and simulation are corrected with a residual correction applied to data, derived as a function of p_T and η. A variety of event topologies (γ + jets, $Z \to e^+e^-$ + jets, $Z \to \mu^+\mu^-$ + jets, and di-jet) are used to derive these corrections. In each topology, the underlying strategy is the same: exploit the momentum conservation in the transverse plane betwen a well-measured reference object and the jet to be corrected. The fact that the reference object (γ, $Z \to e^+e^-$, $Z \to \mu^+\mu^-$, a well-measured central jet) is well-measured allows us to infer the true energy of the jet to be corrected. The full details of these procedures are described in Ref. [40].

Jets used in the $t\bar{t}H$ analysis are first corrected with the procedures described in this section. They are further required to have $p_T > 25$ GeV and $|\eta| < 2.4$ and must pass a loose pileup jet ID criteria. The loose pileup jet ID criteria is based on a BDT designed to discriminate between jets originating from pileup interactions and those originating from the primary vertex in the event. The BDT is trained with variables describing the jets' shape as well as additional track information. Jets are finally also required to not be overlapping with any photons or leptons in the event, requiring $\Delta R(\text{jet}, \text{photon/lepton}) > 0.4$.

b-Tagged Jets

Hadronic jets at the LHC typically result from the hadronization of either a quark or gluon (with the exception of top quarks, which decay before they are able to hadronize). Jets originating from a light-flavor quark (u,d,s) or a gluon are typically indistinguishable in the CMS detector. However, jets originating from c or b quarks often have distinguishing features. While hadrons containing only light-flavor quarks (as would typically be produced by the hadronization of a light-flavor quark or a gluon) often reach the calorimeters before decaying, hadrons containing b quarks tend to decay on a length scale of a few millimeters when produced at typical LHC energies. Hadrons containing charm quarks frequently decay even sooner than this. The resolution of the tracker is sufficient to distinguish the vertices of these decays, called "secondary vertices", from the primary vertices in the event.

Jet flavor tagging algorithms attempt to exploit information about the secondary vertices associated with a given jet to determine the flavor of the quark (or gluon) it originated from. Machine learning algorithms are often used to classify jet flavor, using information about the secondary vertices, tracks, and pf candidates associated with a given jet. Recently, algorithms built with deep neural networks have shown significantly improved jet flavor tagging performance over more traditionally used methods, such as those based on boosted decision trees [87]. The DeepCSV [44] algorithm is one such DNN-based tagger. For a given jet, the algorithm assigns multiple flavor scores, indicating its degree of certainty that the jet originated from a quark of that flavor. DeepCSV outputs scores corresponding to its degree of certainty that the jet originated from a b quark, c quark, light flavor quark (u,d,s) or gluon, and a $b\bar{b}$ pair (four scores). The performance of DeepCSV (purple) and other commonly used jet flavor algorithms is shown in Fig. 5.6.

Jet flavor tagging is particularly useful for the $t\bar{t}H$ analysis, as two b quarks are produced in the decay of the $t\bar{t}$ pair. As the multi-jet, $\gamma+$jets, and $\gamma\gamma+$jets backgrounds primarily feature jets originating from light flavor quarks or gluons, the ability to select b-tagged jets allows for rejection of a significant component of the overall background.

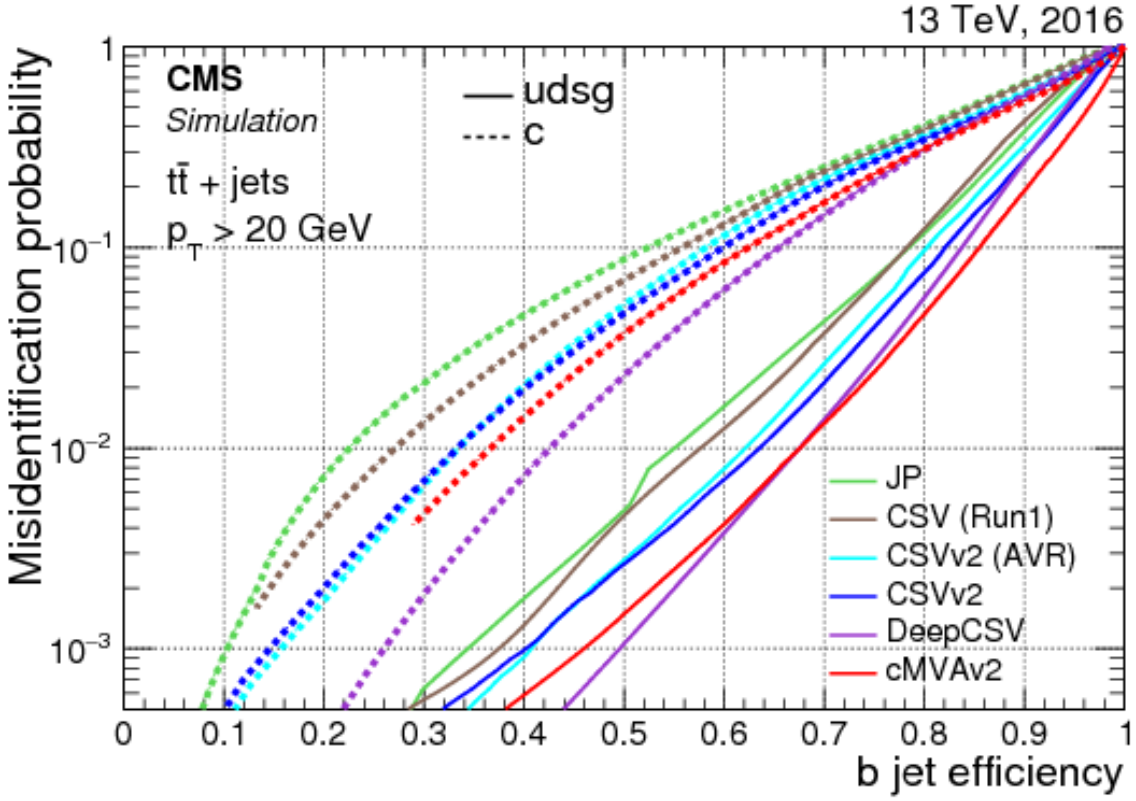

Figure 5.6: Misidentification rate as a function of b-tagging efficiency, shown for b vs. c jet discrimination (dotted lines) and b vs. light jet discrimination (solid lines). Taken from [44].

5.6 Lepton Reconstruction

Muons and electrons identified by the PF algorithm form the starting point for the leptons to be used in analysis.

Muons

Muons are required to have $p_{\mathrm{T}} > 5$ GeV and $|\eta| < 2.4$. Next, a requirement is made on the mini-isolation of the muon, defined as

$$I_{\mathrm{mini}} = \frac{I}{p_{\mathrm{T}}}, \tag{5.3}$$

where the isolation I is taken as the sum of all other PF candidate energies in a cone of size $R = 0.4$ around the muon. The isolation is corrected for contributions from pileup. The $\mathrm{t\bar{t}H}$ analysis requires $I_{\mathrm{mini}} < 0.25$ to mitigate the contribution of hadronic jets misidentified as muons.

Electrons

Electrons are required to have $p_T > 10$ GeV, $|\eta| < 2.5$, and additionally must not be in the ECAL barrel-endcap gap of $|\eta| = [1.4442, 1.566]$. A BDT-based electron ID criteria is also employed. The BDT is trained to distinguish prompt electrons from hadronic jets misidentified as electrons, and is trained with a variety of variables describing the electron's isolation, impact parameter, and kinematics. The invariant mass of electrons with each photon in the event is required to have a difference of greater than 5 GeV with the mass of the Z, in order to reject $Z \to e^+e^-$ events in which one of the electrons is reconstructed as a photon.

Finally, all leptons are required to not be overlapping with the photons in the event, requiring $\Delta R(\text{lepton}, \text{photon}) > 0.2$.

5.7 Missing Transverse Momentum Reconstruction

Neutrinos and other weakly interacting particles cannot be directly detected by the CMS detector. Instead, their presence must be inferred through the presence of missing transverse momentum, p_T^{miss} (often colloquially denoted as E_T^{miss}). The E_T^{miss} in the event is first computed as the negative vector sum of all of the PF candidates in the event. The E_T^{miss} is then corrected according to the jet energy corrections for all of the jets in the event.

5.8 Acknowledgements

The figures shown in Chapter 5 are taken from the following publications: "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV", *Journal of Instrumentation* (2015), "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *Journal of Instrumentation* (2017), "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *Journal of Instrumentation* (2018), and "Measurements of Higgs boson properties in the diphoton decay channel at $\sqrt{s} = 13$ TeV", *CMS-PAS-HIG-19-015* (2020), and were produced by other members of the CMS Collaboration, particularly those involved in the CMS Higgs to Gamma Gamma

working group.

Chapter 6

$t\bar{t}H\ (H \to \gamma\gamma)$ Analysis

6.1 Introduction

Since the Higgs boson was first observed in 2012 by the CMS and ATLAS collaborations [28, 34, 36], characterizing its properties has remained one of the highest priorities of the LHC research program. The Standard Model predicts values for many properties of the Higgs boson, including the strength of its coupling to the other elementary particles. Physics beyond the Standard Model, such as mechanisms of mass generation other than spontaneous symmetry breaking, could modify these coupling strengths. Consequently, precise measurements of the Higgs boson's coupling to elementary particles are of great interest: any deviation from the Standard Model prediction could be indicative of the presence of new physics.

6.1.1 The Top Quark Yukawa Coupling

The coupling of the Higgs boson to the top quark, called the top quark Yukawa coupling, is of particular interest from a theoretical standpoint. Specifically, the top quark Yukawa coupling could help give an indication about the scale of new physics [11].

A primary means of constraining the top quark Yukawa coupling is through the measurement of

the $t\bar{t}H$ production cross section, which is proportional its square:

$$\sigma_{t\bar{t}H} \propto y_t^2. \tag{6.1}$$

The dominant tree-level diagram for $t\bar{t}H$ production is shown in Fig. 6.1.

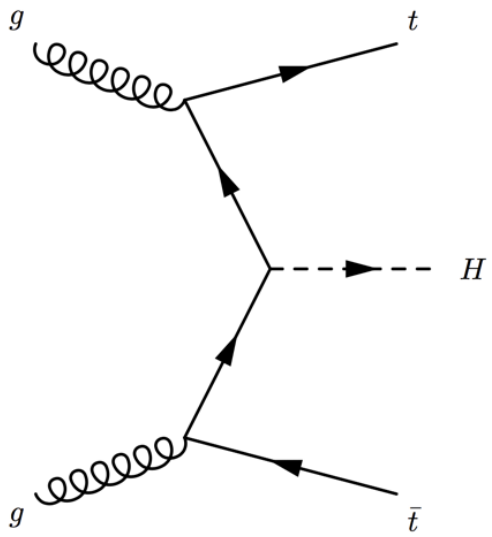

Figure 6.1: Tree-level production of a Higgs boson in association with a top quark-antiquark pair.

This is in fact the best method of *directly* constraining the top quark Yukawa coupling at the LHC. Indirect constraints on y_t come from measurements of Higgs boson production via gluon fusion and $H \to \gamma\gamma$ decay, both of which proceed primarily through a top quark loop, as in Fig. 6.2. However, these constraints are indirect because they make the assumption that no other BSM particles also contribute to the loops.

Complementary methods of constraining y_t include extraction from the shapes of kinematic distributions in $t\bar{t}$ events [48], reporting a measured value of $y_t = 1.16^{+0.24}_{-0.35}$. This measurement does, however, rely on the aforementioned assumptions about the $H \to \gamma\gamma$ branching ratio, and is said to be an indirect constraint on the top Yukawa coupling.

6.1.2 Landscape of $t\bar{t}H$ Measurements

The first observation of the $t\bar{t}H$ process was reported in 2018 by the CMS experiment [46], using 36 fb^{-1} of data from pp collisions at $\sqrt{s} = 13$ TeV. The first observation of $t\bar{t}H$ production in a single H decay channel ($H \to \gamma\gamma$) was reported by the CMS collaboration [50], with the ATLAS collaboration reporting a similar result shortly after [31]. The CMS observation of $t\bar{t}H$ ($H \to \gamma\gamma$) is detailed in the following sections of this thesis.

6.2 Overview of Analysis Strategy

A measurement of the cross section times branching fration of $t\bar{t}H$ ($H \to \gamma\gamma$) production is performed by defining regions of the data which are highly enriched in $t\bar{t}H$ events. These regions, called "signal regions", are constructed through a set of requirements placed on all candidate events. The requirements consist of two components: (1) a "loose preselection", which selects events with at least some of the expected decay products of the $t\bar{t}H$ system and (2) a selection based on the output of a binary classification algorithm (called "BDT-bkg"), trained to separate $t\bar{t}H$ from the SM background processes. The loose preselection aims to maintain a high signal efficiency and defines the phase space in which BDT-bkg is trained. The BDT-bkg algorithm is trained on MC simulation of signal and background, as well as a data-driven description of some backgrounds. After training the BDT-bkg algorithm, signal regions are constructed by placing requirements on the output of BDT-bkg (on top of the preselection requirements). Within these signal regions, signal and background models are constructed and a measurement of the $t\bar{t}H$ cross section is calculated by performing a simultaneous fit to events in all of the signal regions.

6.2.1 The $H \to \gamma\gamma$ Decay Mode

The analysis targets $t\bar{t}H$ events in which the Higgs boson decays into two photons ($H \to \gamma\gamma$). As the photon is a massless particle, it does not couple directly to the Higgs boson. Instead, the dominant process through which the Higgs boson decays to two photons involves a top quark loop, as shown in Figure 6.2. The $H \to \gamma\gamma$ branching ratio is quite small ($\approx 0.2\%$) in comparison to other commonly studied decay modes, but the diphoton channel presents several key advantages.

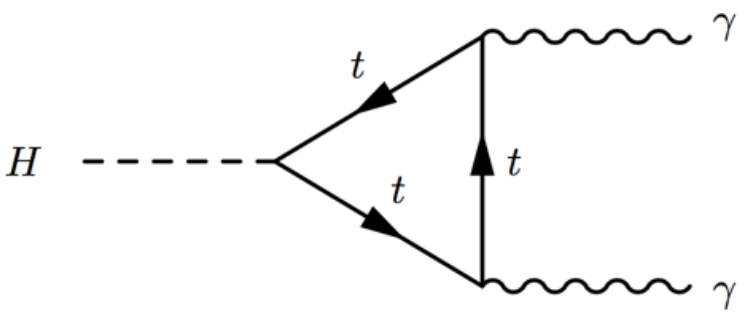

Figure 6.2: Feynman diagram of H $\rightarrow \gamma\gamma$ decay proceeding via a top quark loop.

First, the CMS ECAL provides excellent energy resolution (σ_E/E) for reconstructed photons, which ranges from 1-5% [35]. In general, photons with smaller absolute values of pseudorapidity and higher values of the shower shape variable R_9 (defined in Section 5.4) are reconstructed with better energy resolution. The resulting resolution of the invariant mass of diphoton pairs then ranges from 1-2% for events considered in this analysis. The mass resolution for other Higgs decay modes is typically much worse. The CMS observation of the H $\rightarrow b\bar{b}$ decay mode, for example, achieved a mass resolution of 10-13% [45]. The excellent mass resolution of the diphoton channel contributes to its competitive sensitivity – the SM background follows a steeply falling distribution as a function of increasing diphoton invariant mass, while H $\rightarrow \gamma\gamma$ events are clustered around m_{H} with a resolution of 1-2%. The narrow peak around m_{H} allows for greater discrimination against the SM background processes.

The second advantage of the diphoton decay channel is the relatively small SM background. At the LHC, final states with photons or leptons are significantly rarer than final states with hadrons. Each photon or lepton in a final state introduces another factor of the fine structure constant, $\alpha \approx 1/137$, in the cross section times branching ratio for a given process. A crude, back-of-the-envelope estimate suggests that final states with N leptons plus photons have characteristic cross sections times branching ratios a factor of α^{-N} ($\alpha \approx 137$) times smaller than characteristic cross sections times branching ratios for all-hadronic final states.

Finally, the diphoton decay channel has relatively low systematic uncertainties in comparison to other Higgs decay modes. This is due in part to the fact that photons are well-measured by the CMS ECAl and in part to the fact that the background is estimated directly from data (described in Sec. 6.7.2). The fact that the uncertainty on measurements in the H $\rightarrow \gamma\gamma$ channel are dominated by statistical uncertainties

casts it as the "golden channel" for future Higgs studies in Run 3 of the LHC; the statistical uncertainty on measurements scales roughly with the inverse square root of the luminosity, while the systematic uncertainty remains roughly constant as a function of luminosity.

6.2.2 Estimates of Expected Sensitvity

In developing the $t\bar{t}H$ analysis, the following question frequently arises: will strategy A or strategy B give a better result? "Strategy" may refer to a machine learning algorithm, a description of the SM background, etc. "Better" is a somewhat subjective matter, but the primary measurement of the $t\bar{t}H$ analysis is that of $\mu_{t\bar{t}H}$, defined as the ratio of the observed cross section times branching fraction of $t\bar{t}H$ ($H \to \gamma\gamma$) and the predicted cross section times branching fraction:

$$\mu_{t\bar{t}H} = \frac{(\sigma_{t\bar{t}H} \times \mathcal{B}(H \to \gamma\gamma))_{obs}}{(\sigma_{t\bar{t}H} \times \mathcal{B}(H \to \gamma\gamma))_{SM}}. \tag{6.2}$$

Therefore, "better" is taken to mean "giving a smaller expected uncertainty on the measurement of $\mu_{t\bar{t}H}$". The superlative is framed in terms of the expected uncertainty, rather than the observed uncertainty, as the analysis is developed in a blinded fashion to avoid introducing bias. The full calculation of $\mu_{t\bar{t}H}$ involves extensive computing (and human) time, so a simplified measure of the expected uncertainty is used. For these purposes, we use Z_A [59], defined as

$$Z_A(s,b) = \sqrt{2\left((s+b)\ln(1+\frac{s}{b}) - s\right)}, \tag{6.3}$$

where s is the number of signal events and b is the number of background events. In the limit $s << b$, the Taylor expansion of Z_A gives the familiar $s/\sqrt{b}$:

$$Z_A(s,b) = \frac{s}{\sqrt{b}}\left[1 + O\left(\frac{s}{b}\right)\right]. \tag{6.4}$$

The signal yield s is estimated by fitting a Gaussian function to the $m_{\gamma\gamma}$ distribution of signal MC events and integrating the fitted function over the signal mass window. The background yield b is estimated by counting the yield of all events in the $m_{\gamma\gamma}$ sidebands (100 GeV $< m_{\gamma\gamma} <$ 120 GeV, 130 GeV $< m_{\gamma\gamma} <$ 180 GeV) and

scaling to the size of the signal mass window. The signal mass window is taken to be $m_{\mathrm{H}} \pm 1.645 \times \sigma_{\mathrm{eff}}$, with σ_{eff} taken from the fitted Gaussian. It is not necessary to employ more sophisticated estimates of the signal and background yields (as is done in Sec. 6.7) because Z_A is only used in comparing the performance of two strategies; the absolute value of Z_A is not relevant. When quoting an improvement obtained by using a specific method, the value is taken as the percentage difference between the maximum Z_A values obtained by the two methods.

6.3 Preselection

Events passing the diphoton preselection, described in Sec. 5.4.5, are eligible to enter one of two exclusive channels. The hadronic channel targets $\mathrm{t\bar{t}H}$ events in which the $\mathrm{t\bar{t}}$ pair decays fully hadronically, while the leptonic channel targets events in which at least one of the top (anti-)quarks decays leptonically. The channels are defined to be orthogonal through selections on the number of leptons in the event: the leptonic channel requires at least one lepton and the hadronic channel requires exactly zero leptons.

Each channel then places an additional set of requirements on the events which enter. The hadronic channel requires at least three jets, one of which must be identified as originating from a b quark. Jets are identified as originating from a b quark using the DeepCSV b score, with a working point that corresponds to a 10% misidentification rate for jets originating from light quarks or gluons. The leptonic channel requires at least one jet, with no requirement on the jet flavor.

6.4 Background Description

The analysis uses two methods of estimating the contribution of background processes (i.e., those other than $\mathrm{t\bar{t}H}$). The first estimates the background directly from data by fitting events in the $m_{\gamma\gamma}$ sidebands, defined as $m_{\gamma\gamma} \in [100, 115] \cup [135, 180]$ GeV, by fitting a variety of functional forms. The second estimates the background from individual descriptions of each background process. These descriptions are taken primarily from simulations of each process; however, data-driven descriptions of some processes are also utilized: the multi-jet and $\gamma+$jets backgrounds are described with a sample of events in data from the low photon ID sideband (described in Sec. 6.4.2). The first method is referred to as the "discrete profiling

method", while the second is referred to as the "MC description" of the background.

The discrete profiling method is used to estimate the background in the final statistical analysis and is described in Sec. 6.7.2. The MC description of the background is used only in designing and optimizing the cuts on the BDT-bkg algorithms and is described in this section.

6.4.1 Challenges of MC Description

The events entering the hadronic or leptonic preselection are dominated by Standard Model processes other than $t\bar{t}H$. Precise knowledge of exactly which processes enter the preselection and their relative contributions to the overall background is not strictly necessary, as the background is modeled from events in data (described in Sec. 6.7.2) when performing the measurement of $\mu_{t\bar{t}H}$. However, the BDT-bkg algorithms are designed to distinguish between $t\bar{t}H$ and the SM background processes; to this end, an accurate description of the background is desirable. Note that events in data cannot be used to both model the background in the measurement of $\mu_{t\bar{t}H}$ and in training the BDT-bkg algorithm, as this would bias the measurement. The starting point for the background description used in training the BDT-bkg algorithms is simulation of the relevant SM processes. In the hadronic channel, the dominant backgrounds at preselection level are the multi-jet, $\gamma+$jets, and $\gamma\gamma+$jets processes; those same processes, as well as $t\bar{t}+$jets, $t\bar{t}+\gamma+$jets, $t\bar{t}+\gamma\gamma$, and $V+\gamma$ dominate for the leptonic channel. The $m_{\gamma\gamma}$ distribution for events in data and simulation are shown in Fig. 6.3. The exact yields and relative contributions of all considered background processes are shown in Table 6.1.

In both channels, the overall yield from the background description given by simulation is somewhat smaller than what is observed in data. This underprediction can be primarily attributed to a poor MC description of the multi-jet and $\gamma+$jets processes, as the following section illustrates (and remedies).

6.4.2 Data-Driven Description of Multi-jet and $\gamma+$jets Backgrounds

To better understand the reason for the large discrepancies between data and simulation shown in Fig. 6.3 and Table 6.1, it is helpful to study the discrepancy as a function of the event kinematics. The distribution of minimum photon ID, defined as the smaller of the two photon ID BDT scores in the event, illustrates that the underprediction from simulation stems primarily from low values of minimum photon

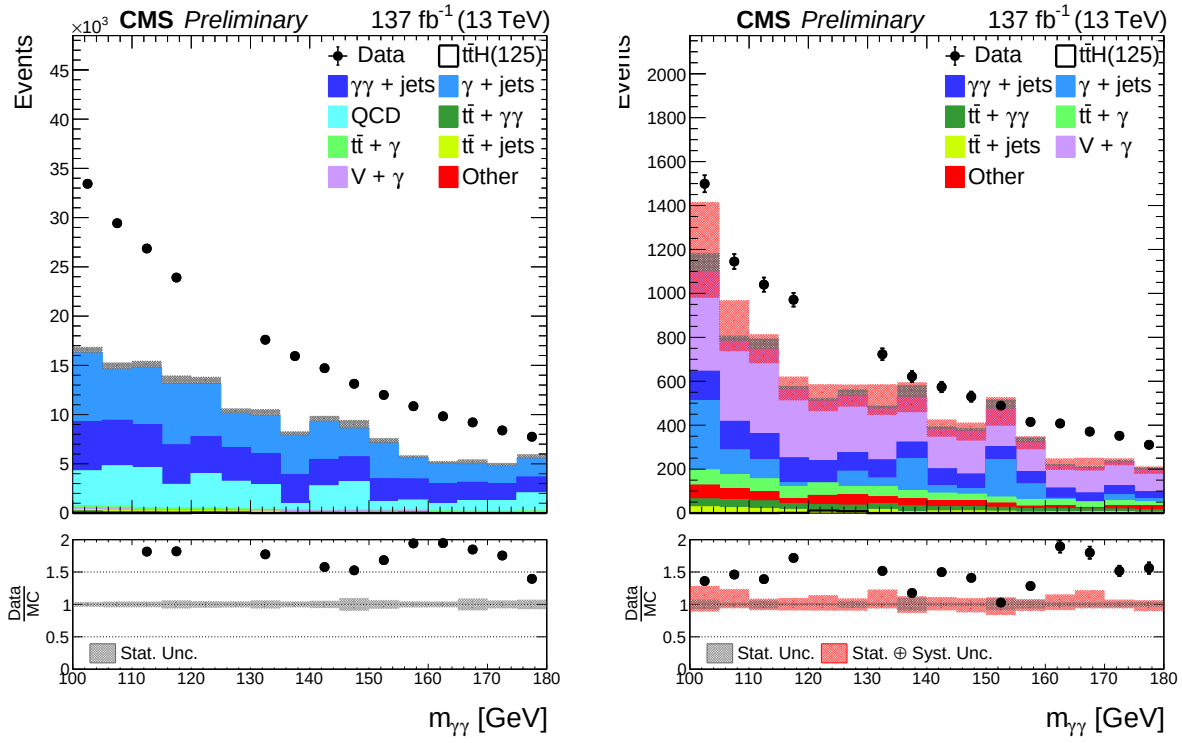

Figure 6.3: Diphoton invariant mass distributions for events from data and simulation entering the hadronic (left) and leptonic (right) channel preselections. Events in data are blinded in the region $m_{\gamma\gamma} \in [120, 130]$.

ID, as shown in Fig. 6.4.

The underprediction from simulation is especially pronounced at lower values of minimum photon ID, while the agreement becomes better for higher values. As the photon ID BDT discriminates between prompt and fake photons, we expect the lower end of the distribution to have a larger contribution from fake photons. Thus, Fig. 6.4 suggests that simulation provides an inadequate description of events with at least one fake photon.

To improve the description of events with at least one fake photon (i.e. multi-jet and $\gamma+$jets events), a sample of events from an orthogonal region of data is used in place of the simulation when training BDT-bkg. This region, the "low photon ID sideband", is defined identically to the preselection except for the cut on minimum photon ID: the preselection requires minimum photon ID > -0.7, while the low photon ID sideband requires $-0.9 <$ minimum photon ID < -0.7. The preselection and low photon ID sideband are depicted in Fig. 6.5. Replacing the MC description of multi-jet and $\gamma+$jets with the

Table 6.1: Yields and fraction of total background by process for the hadronic (left) and leptonic (right) channel preselections. Backgrounds not explicitly shown in Fig. 6.3 are consolidated in the "Other" category.

Process	Yield	$\mathcal{F}$ of bkg
$\gamma\gamma+$jets	40972.68 $\pm$ 75.87	0.32
$\gamma+$jets	52434.13 $\pm$ 1960.51	0.41
Multi-jet	29277.57 $\pm$ 3566.18	0.23
$t\bar{t}+\gamma\gamma$	642.31 $\pm$ 35.09	0.01
$t\bar{t}+\gamma+$jets	1538.53 $\pm$ 68.77	0.01
$t\bar{t}$	997.19 $\pm$ 74.15	0.01
Drell-Yan	265.24 $\pm$ 47.48	0.00
$t+\gamma$	170.79 $\pm$ 26.66	0.00
V + γ	1237.50 $\pm$ 39.10	0.01
$t\bar{t}W$	3.13 $\pm$ 0.17	0.00
$t\bar{t}Z$	3.41 $\pm$ 0.15	0.00
VV	55.06 $\pm$ 4.49	0.00
tV	135.61 $\pm$ 7.57	0.00
tHq	5.53 $\pm$ 0.00	0.00
tHW	1.44 $\pm$ 0.00	0.00
ggH	199.20 $\pm$ 1.80	0.00
VH	22.94 $\pm$ 0.22	0.00
VBF	18.78 $\pm$ 0.24	0.00
All bkg.	128029.09 $\pm$ 4072.22	1.00
Data	233060.00 $\pm$ 482.76	1.82
$t\bar{t}H$	48.03 $\pm$ 0.32	0.00

Process	Yield	$\mathcal{F}$ of bkg
$\gamma\gamma+$jets	1067.40 $\pm$ 13.85	0.15
$\gamma+$jets	1070.57 $\pm$ 236.09	0.15
Multi-jet	482.43 $\pm$ 343.91	0.07
$t\bar{t}+\gamma\gamma$	313.58 $\pm$ 8.39	0.04
$t\bar{t}+\gamma+$jets	542.93 $\pm$ 11.70	0.08
$t\bar{t}$ + Jets	159.21 $\pm$ 6.15	0.02
Drell-Yan	220.56 $\pm$ 39.78	0.03
$t+\gamma$	37.43 $\pm$ 13.28	0.01
V + γ	3081.67 $\pm$ 56.46	0.43
$t\bar{t}W$	3.61 $\pm$ 0.18	0.00
$t\bar{t}Z$	4.75 $\pm$ 0.17	0.00
VV	51.79 $\pm$ 4.12	0.01
tV	61.17 $\pm$ 4.88	0.01
tHq	1.83 $\pm$ 0.00	0.00
tHW	0.79 $\pm$ 0.00	0.00
ggH	4.90 $\pm$ 0.25	0.00
VH	10.67 $\pm$ 0.14	0.00
VBF	0.74 $\pm$ 0.04	0.00
All bkg.	7138.36 $\pm$ 423.60	1.00
Data	9450.00 $\pm$ 97.21	1.32
$t\bar{t}H$	22.36 $\pm$ 0.21	0.00

data-driven description relies on the assumption that events in the low photon ID sideband are exlusively multi-jet and $\gamma+$jets events. Simulation indicates that $> 95\%$ of events in the low photon ID sideband are multi-jet or $\gamma+$jets events.

An immediate challenge in making the replacement of MC description of multi-jet and $\gamma+$jets $\rightarrow$ data-driven description of multi-jet and $\gamma+$jets is the fact that the minimum photon ID for these events and the minimum photon ID for events in the preselection are disjoint, by definition. This is problematic because of the fact that minimum photon ID is used as a training feature for the BDT-bkg algorithms. Training with an unaltered minimum photon ID distribution would lead the BDT to eliminate all of these background events with a single cut at the value of minimum photon ID that defines the sideband. To make the data-driven sample feasible for use in training the BDT-bkg algorithm, its minimum photon ID distribution should be altered such that it resembles the expected distribution of multi-jet and $\gamma+$jets

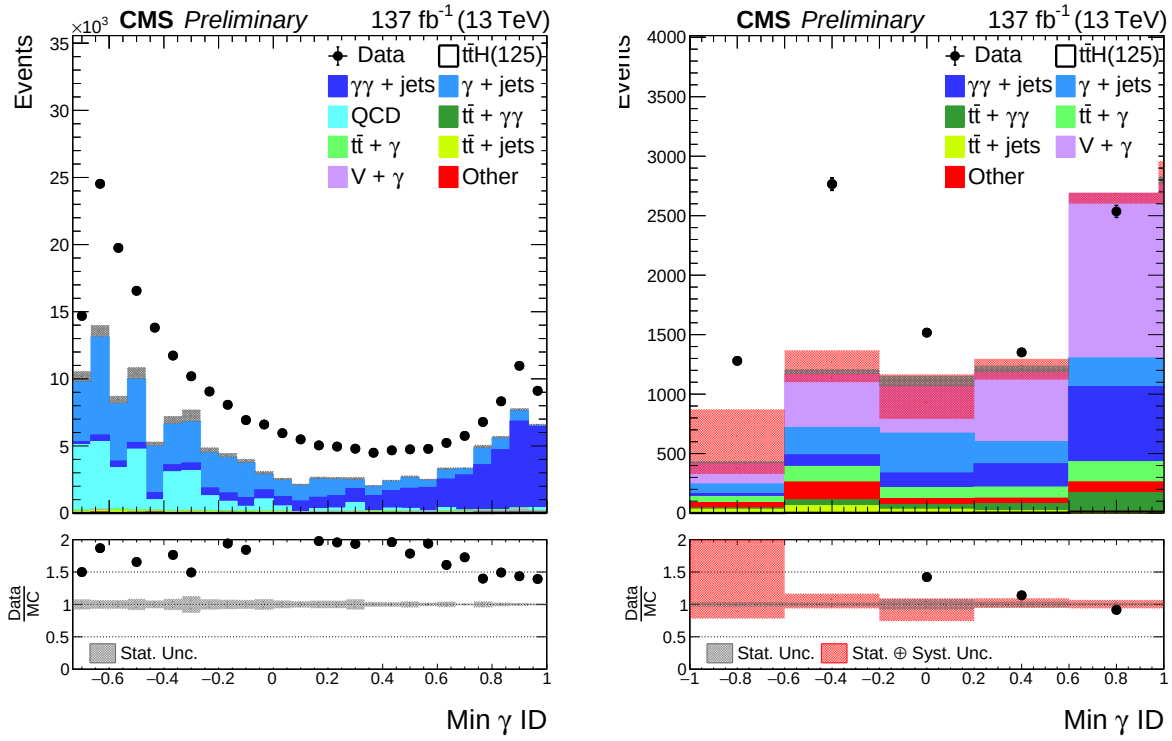

Figure 6.4: Minimum photon ID distributions for events from data and simulation entering the hadronic (left) and leptonic (right) channel preselections. Events in data are blinded in the region $m_{\gamma\gamma} \in [120, 130]$.

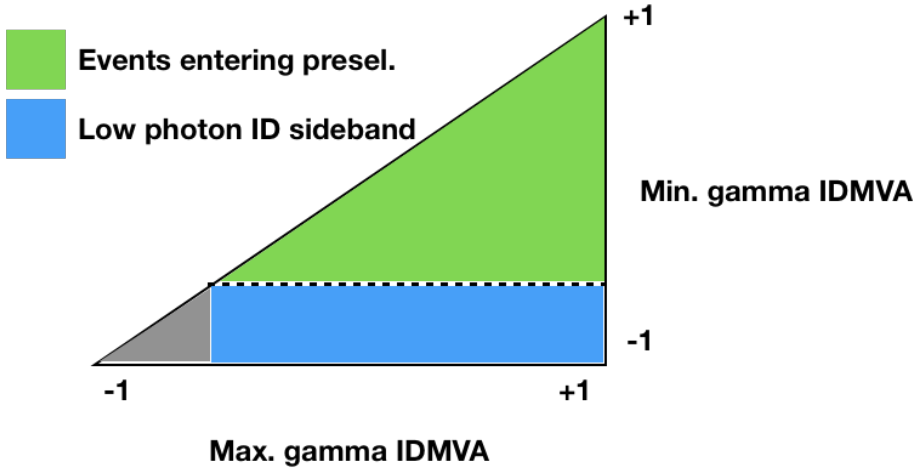

Figure 6.5: Depiction of the relationship between preselection (green) and low photon ID sideband (blue).

events in the preselection region.

To generate the proper minimum photon ID distribution for events in the data-driven sample,

the minimum photon ID score of each event is replaced by a randomly drawn value from a probability distribution function that describes the expected distribution of multi-jet and $\gamma+$jets in the preselection. The procedure is simplified by assuming that for events in the low photon ID sideband the photon with the lower photon ID score is always a fake photon. This assumption is always true for multi-jet events (which have two fake photons), and from simulation, is found to be true for $> 95\%$ of $\gamma+$jets events. Under this assumption, the expected distribution of minimum photon ID for multi-jet and $\gamma+$jets events in the preselection region can be approximated by the probability distribution function of photon ID for fake photons, called the "fake pdf". The fake pdf is derived from simulation using the photon ID distribution of photons which are identified as fakes at generator-level. For ease of drawing random values from this pdf, a histogram of the fake pdf is fitted with a seventh-order polynomial, shown in Fig. 6.6. However,

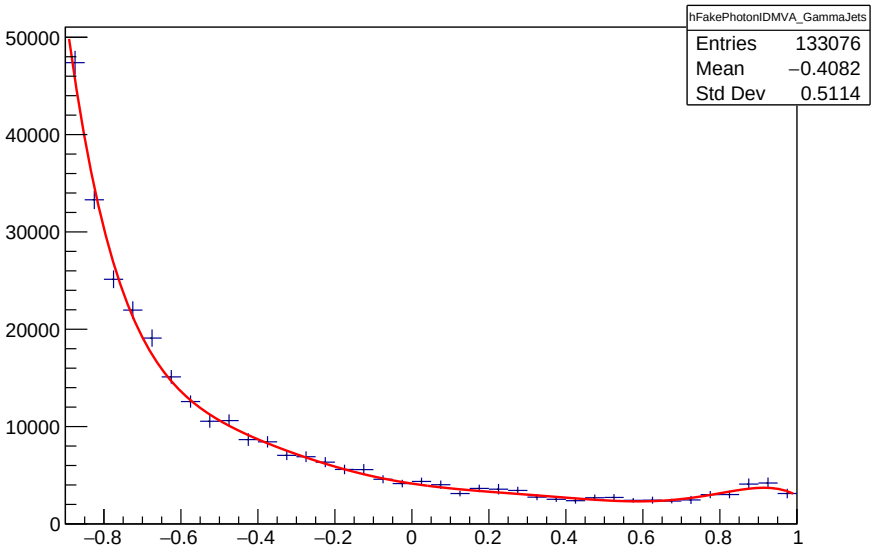

Figure 6.6: Histogram of photon ID for fake photons in simulation (blue) and resulting seventh-order polynomial.

as minimum and maximum photon ID are strongly correlated by construction, the maximum photon ID distribution for events from the low photon ID sideband will also be different than the maximum photon ID distribution for events from the preselection. In general, events from the low photon ID sideband will have lower maximum photon ID scores than events from the preselection. To address the differences in maximum photon ID, an additional weight is applied to each event to correct for the fact that the

distribution is biased towards lower values of maximum photon ID:

$$w = \frac{\int_B^{\max \gamma\,\mathrm{ID}} \text{fake pdf}}{\int_A^B \text{fake pdf}} \tag{6.5}$$

$$A \equiv \text{minimum value of photon ID in the low photon ID sideband} = -0.9 \tag{6.6}$$

$$B \equiv \text{preselection cut value on minimum photon ID} = -0.7 \tag{6.7}$$

Qualitatively, the term of the numerator of Eqn. 6.5 increases the contribution of events from the low photon ID sideband with high values of maximum photon ID. The term in the denominator is simply an overall normalization factor. After applying the per-event weight, the overall normalization of the data-driven sample is determined with a simultaneous fit to data of the minimum and maximum photon ID distributions in the preselection. The normalization of $\gamma\gamma+$jets is also allowed to float in this fit, while the normalization of all other background processes are taken to be fixed.

Table 6.2: Results of binned fit of diphoton templates in the hadronic preselection, with template for prompt/prompt taken from MC simulation and template for fake/fake and fake/prompt taken from the data-driven description.

Template	Initial Fraction	Fitted Fraction	Scale
γ + jets (fake/prompt)	0.68	0.73 ± 0.00	1.07
$\gamma\gamma$ + jets (prompt/prompt)	0.18	0.25 ± 0.00	1.42

The minimum and maximum photon ID distributions are shown pre-/post-fit in Fig. 6.7 and the results of the fit are shown in Table 6.2. The agreement with data is significantly improved when using the data-driven description of multi-jet and $\gamma+$jets in place of the MC description, as can be seen by comparing Fig. 6.4 and Fig. 6.7. This improvement is consistently seen in other distributions as well, including the jet and b-jet multiplicities shown in Fig. 6.8.

As the MC description of the background is used only for designing and optimizing the cuts on the BDT-bkg algorithms, the bottom-line test of the merit of the data-driven description is BDT-bkg performance. To this end, we compare the performance of the BDT-bkg algorithm trained with the MC description of $\gamma+$jets (including the MC description of multi-jet events was found to degrade performance because of the extremely few number of events from the multi-jet simulation samples) with the performance of the BDT-bkg algorithm trained with the data-driven description of multi-jet and $\gamma+$jets, using Z_A (defined

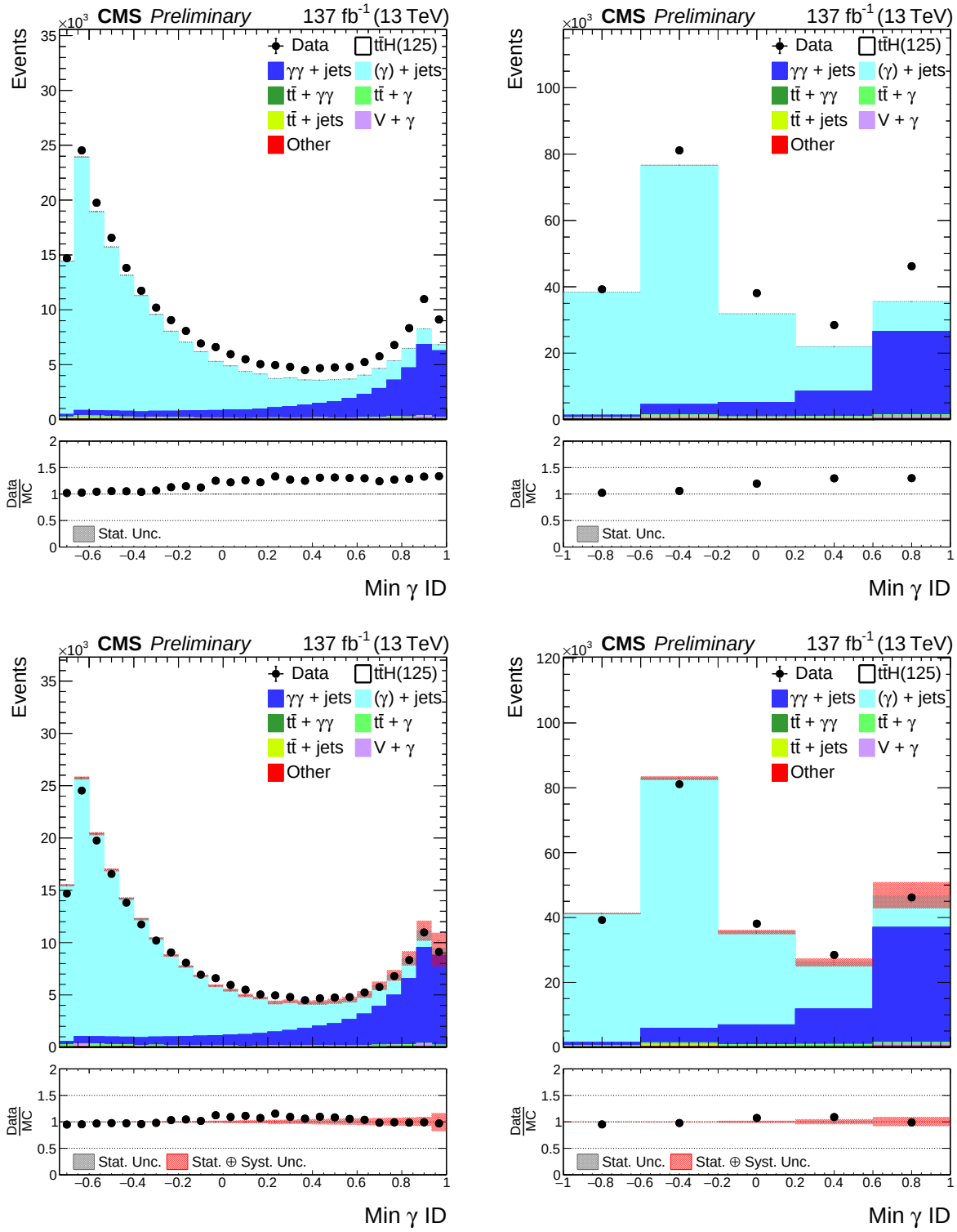

Figure 6.7: Distributions of minimum (left) and maximum (right) photon ID in the hadronic preselection before (top) and after (bottom) fitting the normalization of the data-driven description of multi-jet and $\gamma+$jets and the MC description of $\gamma\gamma+$jets.

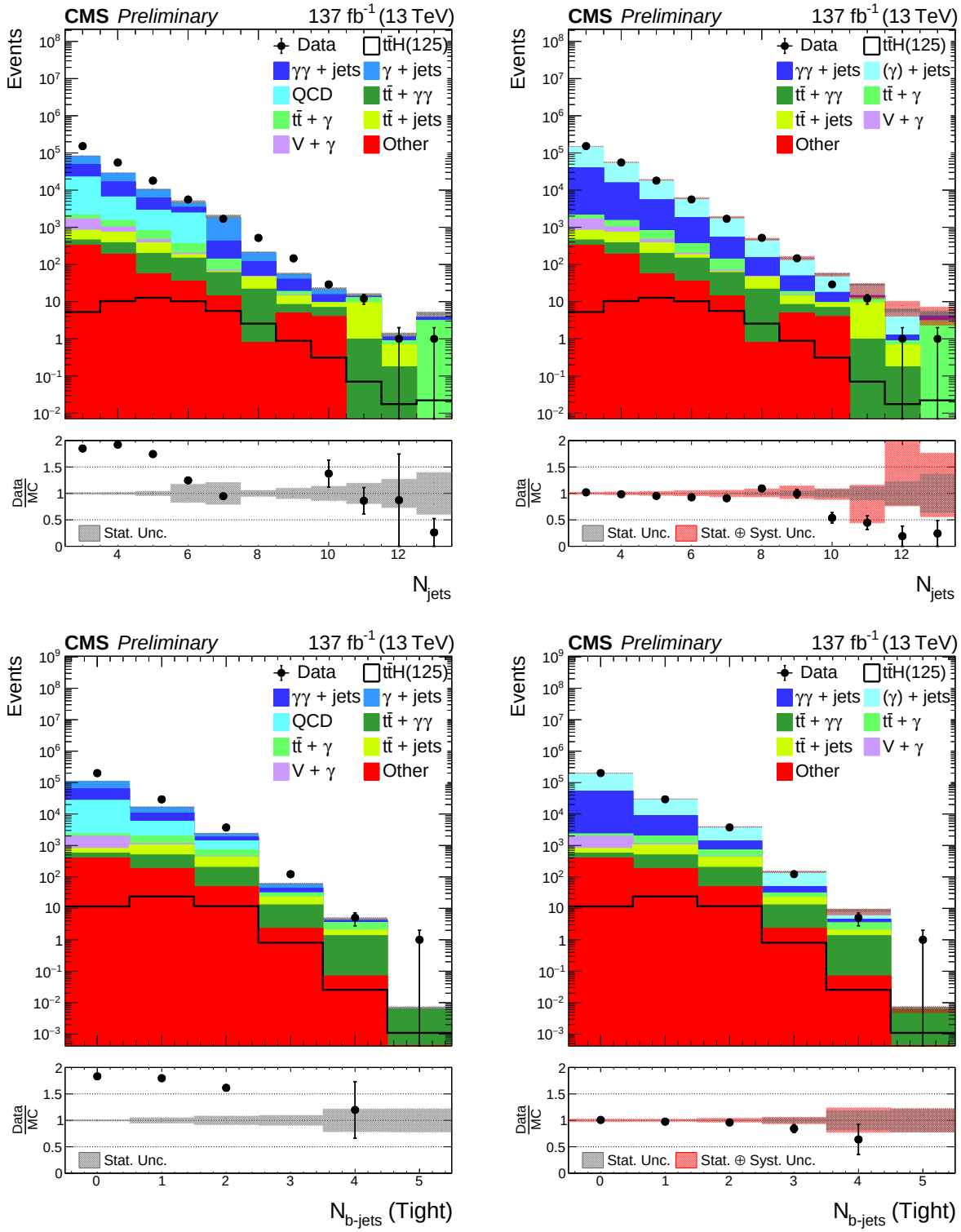

Figure 6.8: Agreement between data and MC description of background for jet multiplicity (top) and b-jet multiplicity (bottom), shown with both the MC description of multi-jet and $\gamma+$jets (left) and the data-driven description of multi-jet and $\gamma+$jets (right).

in Sec. 6.2.2) as an estimate of the expected significance obtained when using either BDT. Z_A is shown for each BDT as a function of the number of $t\bar{t}H$ events in Fig. 6.9. The improvement obtained by replacing the MC description of $\gamma+$jets with the data-driven description of multi-jet and $\gamma+$jets, taken as the percentage difference between the maximum Z_A values obtained with either method, is about 8%. The data-driven

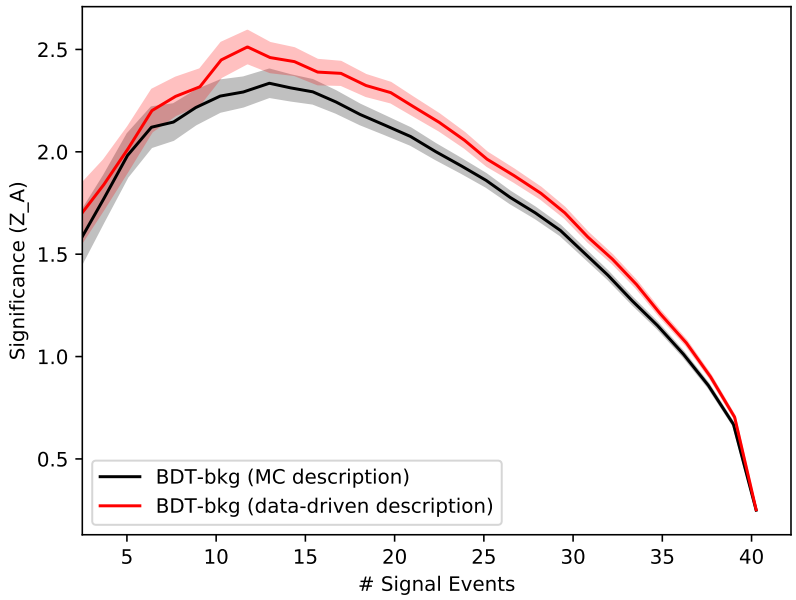

Figure 6.9: Expected significance (Z_A) shown as a function of the number of $t\bar{t}H$ events passing a given cut on BDT-bkg for versions of BDT-bkg trained with the MC description of $\gamma+$jets (black) and the data-driven description of multi-jet and $\gamma+$jets (red). Shaded bands show the $\pm1\sigma$ statistical uncertainty in Z_A.

description of multi-jet and $\gamma+$jets events is not employed in the leptonic channel as it was not found to significantly improve BDT-bkg performance. This is likely due to the combination of two factors:

1. Events in the leptonic low photon ID sideband are not dominated by multi-jet and $\gamma+$jets events to the same extent as in the hadronic channel; $t\bar{t}+$jets and $t\bar{t}+\gamma+$jets events play a significant role as well.

2. The multi-jet and $\gamma+$jets processes make up a smaller overall fraction of the background contribution in the leptonic preselection; there is a lower ceiling on improvements to BDT-bkg performance from improved description of these processes.

6.5 Machine Learning Algorithms

The analysis strategy relies on the use of a binary classification algorithm to create regions of high $t\bar{t}H$ purity. A boosted decision tree (BDT) is chosen as the binary classification algorithm. Deep neural networks (DNNs) were also explored; however, these were found to outperform BDTs only when a very high number of training examples are available (discussed in further detail in Sec. 6.5.2). The BDT-bkg algorithms are trained with the XGBOOST [22] framework, using MC simulation of $t\bar{t}H$ for the signal events and a combination of MC simulation and data-driven descriptions of the relevant background processes as background. In order to avoid bias, the BDT-bkg algorithms are trained and optimized on completely separate samples from those used in the measurement of $\mu_{t\bar{t}H}$. The various background processes are weighted according to their cross sections, while the weights of signal events are scaled such that the total number of signal events is equal to the total number of background events. The signal weights are scaled in such a fashion to avoid issues in training related to imbalanced classes. The features used in training the BDT-bkg algorithms are discussed in the following subsections.

6.5.1 High-Level Features

In order to effectively separate $t\bar{t}H$ from the SM backgrounds, we construct a number of high-level physics variables which are expected to have discriminating power between the two. The full list of features is shown in Table 6.3. Plots showing the distributions for both data and simulation for each input feature are shown in Appendix A.

The physics motivations for the inclusion of each feature are enumerated below:

1. Photon Kinematics

   - Photon p_{T} divided by $m_{\gamma\gamma}$: prompt photons tend to have higher transverse momentum than fake photons from hadronic jets. The photon p_{T} is normalized by $m_{\gamma\gamma}$ to prevent the BDT from learning m_{H}.
   - Photon η: prompt photons tend have smaller $|\eta|$ than fake photons.
   - Photon pixel seed veto: $t\bar{t}+X$ events often have one or more electrons from leptonically

Table 6.3: High level features used in training BDTs. The fourth jet kinematics are given as input features only to the BDT-bkg in the hadronic channel, while the lepton kinematics are given as input features only to the BDT-bkg in the leptonic channel.

Input Features to BDTs

Category	Features		
Photon Kinematics	γ_1 $p_T/m_{\gamma\gamma}$	γ_1 η	γ_1 Pixel Seed Veto
	γ_2 $p_T/m_{\gamma\gamma}$	γ_2 η	γ_2 Pixel Seed Veto
	Max γ ID MVA	Min γ ID MVA	
Jet Kinematics	Jet 1 p_T	Jet 1 η	Jet 1 b-tag score
	Jet 2 p_T	Jet 2 η	Jet 2 b-tag score
	Jet 3 p_T	Jet 3 η	Jet 3 b-tag score
	Jet 4 p_T	Jet 4 η	Jet 4 b-tag score
	Max b-tag score	2nd max b-tag score	
	N_{jets}	H_T	
DiPhoton Kinematics	$p_T^{\gamma\gamma}/m_{\gamma\gamma}$	$Y_{\gamma\gamma}$	$\lvert\cos(\Delta\phi)_{\gamma\gamma}\rvert$
	$\Delta R_{\gamma\gamma}$	$\lvert\cos(\text{helicity angle}(\theta))\rvert$	
Lepton Kinematics	lepton p_T	lepton η	$N_{\text{leptons (tight ID)}}$
Event-level Kinematics	E_T^{miss}		

decaying W bosons which are identified as photons at reco-level. The pixel seed veto helps identify these events.

- Photon ID MVA: the primary means of separating between prompt and fake photons. Described in further detail in Sec. 5.4.4.

2. Jet Kinematics

   - Jet p_{T}: jets from $t\bar{t}H$ events tend to have higher transverse momentum than those from background processes (both multi-jet + X and $t\bar{t}+X$), as the $t\bar{t}$ system recoils against the Higgs.
   - Jet η: jets from $t\bar{t}H$ tend to be more central than those from multi-jet + X events.
   - Jet b-tag scores: jets from $t\bar{t}H$ have higher b-tag scores than those from non-$t\bar{t}$ backgrounds, as two b-jets are expected 2 from the $t\bar{t}$ decay.
   - Number of jets: we expect at least 6 (4) jets in a $t\bar{t}H$ event in the Hadronic (Leptonic) channel.
   - H_{T}: $t\bar{t}H$ events tend to have higher values of H_{T}, due to the fact that we expect higher p_T of individual jets as well as more total jets in the event.

3. DiPhoton Kinematics

- DiPhoton p_{T} divided by $m_{\gamma\gamma}$: the recoil of the Higgs against the $\mathrm{t\bar{t}}$ system results in higher expected values for the diphoton momentum. The DiPhoton p_{T} is normalized by $m_{\gamma\gamma}$ to prevent the BDT from learning m_{H}.
- DiPhoton Y: the DiPhoton rapidity is expected to be closer to zero for $\mathrm{t\bar{t}H}$ events.
- DiPhoton ΔR: the angle between the two photons is expected to be smaller for $\mathrm{t\bar{t}H}$ events due to the fact that the Higgs tends to be boosted from recoil against the $\mathrm{t\bar{t}}$ system.
- Helicity angle (θ): defined by boosting to the rest frame of the DiPhoton pair and calculating the angle between the photons in that frame. Since the SM Higgs is a scalar, we expect a uniform distribution in $\cos(\theta)$ for $\mathrm{t\bar{t}H}$ (at generator-level), while backgrounds may peak closer to 1.
- $\cos(\Delta\phi)$ of DiPhoton pair: similar argument to above.

4. Lepton Kinematics

- Lepton p_{T}: prompt leptons tend to have higher p_{T} than fake leptons from hadronic jets.
- Lepton $|\eta|$: prompt leptons tend to be more central than fake leptons.
- Number of leptons passing tight ID: again, helpful in discriminating between prompt and fake leptons. The preselection requires that leptons pass medium ID, the number of leptons passing tight ID is given as an input to the BDT.

5. Event-level Kinematics

- $E_{\mathrm{T}}^{\mathrm{miss}}$: in the Leptonic channel, we expect nonzero $E_{\mathrm{T}}^{\mathrm{miss}}$ in $\mathrm{t\bar{t}}$ events due to the neutrino from the $\mathrm{W} \to l\nu$ decay, while no $E_{\mathrm{T}}^{\mathrm{miss}}$ is expected for multi-jet + X events. In the Hadronic channel, this may also be useful in identifying leptonically decaying $\mathrm{t\bar{t}H}$ events in which the lepton is not reconstructed.

6.5.2 Deep Neural Networks for $\gamma\gamma+$ jets and $\mathrm{t\bar{t}}+\gamma\gamma$ Backgrounds

The dominant SM backgrounds in regions of high $\mathrm{t\bar{t}H}$ purity (i.e. similar to the signal regions) are: $\gamma\gamma+$ jets and $\mathrm{t\bar{t}}+\gamma\gamma$ in the hadronic channel, and $\mathrm{t\bar{t}}+\gamma\gamma$ in the leptonic channel. To further reduce these

backgrounds, additional methods and variables designed to specifically target these backgrounds were explored. Among the most successful of these methods were deep neural networks (DNNs) trained to reject these backgrounds specifically. Another notable susccessful tool is the top tagger BDT, discussed in Sec. 6.5.3.

The DNNs exploit low-level information in each event which is lost in the process of summarizing the event in terms of the high-level features described in Sec. 6.5.1. An example of such low-level information is the azimuthal angle ϕ of the reconstructed jets and leptons. Any physics process should obey azimuthal symmetry in the CMS detector – any value of ϕ is equally likely for any physics object from any physics process. For this reason, the ϕ value of a given jet or lepton on its own provides no means to discriminate between $t\bar{t}H$ and other SM backgrounds. However, the ϕ value of a jet or lepton may provide discriminatory power when considered in the context of the rest of the event: angles between jets and leptons likely have different distributions in $t\bar{t}H$ events and SM background events. Providing this type of low-level input to a BDT would be of limited use, as making a splitting on the ϕ value of a jet or lepton is not useful on its own. Deep neural networks were then explored as an algorithm which could potentially make better use of such low-level features.

Training Features

The high-level features described in Sec. 6.5.1 do not retain the full information of the original event. To provide a more complete description of each event, the four-vectors of the leading jets and leptons ("physics objects") are given as inputs to the DNNs. The four-vector includes the physics object's p_{T}, η, ϕ, and total energy (E). In addition to the four-vectors, four jet flavor scores and lepton ID flags, indicating whether a given lepton is a muon or electron are provided for each physics object. In total, for up to the leading eight (six) jets in each event and the leading zero (two) leptons in each event in the Hadronic (Leptonic) channel, the following nine features are provided for each physics object:

- Four-vector: p_T, η, ϕ, E
- 4 DeepCSV scores: b, c, $udsg$, bb
- Lepton ID flag: 0 for muons, 1 for electrons

In the hadronic channel, there are no leptons and so the Lepton ID flag is omitted, resulting in 8 features per physics object. In the leptonic channel, jets (leptons) are assigned "dummy" values for the lepton ID flag (DeepCSV scores) of -2. The "dummy" values are necessary as each physics object that is input to the DNN must have the same number of input features. Photons are not included in the list of physics objects to prevent the DNN from learning m_{H}.

In addition to the low-level object features, a set of high-level features, shown in Table 6.4, is given as inputs to the DNN to allow it to learn correlations between the physics objects and the rest of the event, including the diphoton kinematics and missing transverse momentum.

Table 6.4: High level features used in training DNNs. The lepton kinematics features are only given as inputs to the leptonic channel DNN.

High-level features for DNNs

Category	Features			
Photon Kinematics	γ_1 $p_{\mathrm{T}}/m_{\gamma\gamma}$	γ_1 η	γ_1 ϕ	γ_1 Pixel Seed Veto
	γ_2 $p_{\mathrm{T}}/m_{\gamma\gamma}$	γ_2 η	γ_2 ϕ	γ_2 Pixel Seed Veto
	Max γ ID MVA	Min γ ID MVA		
Jet Kinematics	Max b-tag score	2nd max b-tag score		
	N_{jets}			
DiPhoton Kinematics	$p_{\mathrm{T}}^{\gamma\gamma}/m_{\gamma\gamma}$	$Y_{\gamma\gamma}$	$\Delta R_{\gamma\gamma}$	
Lepton Kinematics	$N_{\mathrm{leptons\ (tight\ ID)}}$			
Event-level Kinematics	$E_{\mathrm{T}}^{\mathrm{miss}}$	$E_{\mathrm{T}}^{\mathrm{miss}}\phi$		

Architecture

Arguably the simplest approach would be to feed the low-level and high-level features to a fully-connected deep neural network. However, the question arises of how to organize the physics objects in the event: how does one identify the "first" jet in an event? One solution is to order the physics objects by their p_{T}, meaning that the first jet in the event is the one with the highest p_{T}. This strategy is far from perfect: in some $\mathrm{t\bar{t}H}$ events the first jet may be a b jet from a top decay, while in others the first jet may be a c jet from a W decay. A long-short term memory (LSTM) architecture [91] is employed to address this challenge. The physics objects are treated as a one-dimensional sequence (ordered by p_{T}) that is given to the LSTM. This choice of architecture is motivated by other successful applications of LSTMs to physics objects, notably the DeepCSV and DeepJet architectures which classify jet flavor in part from a one-dimensional

sequence of PF candidates given to an LSTM. The output of the LSTM network is then merged with the high-level features in a set of fully-connected layers, as depicted in Fig. 6.10.

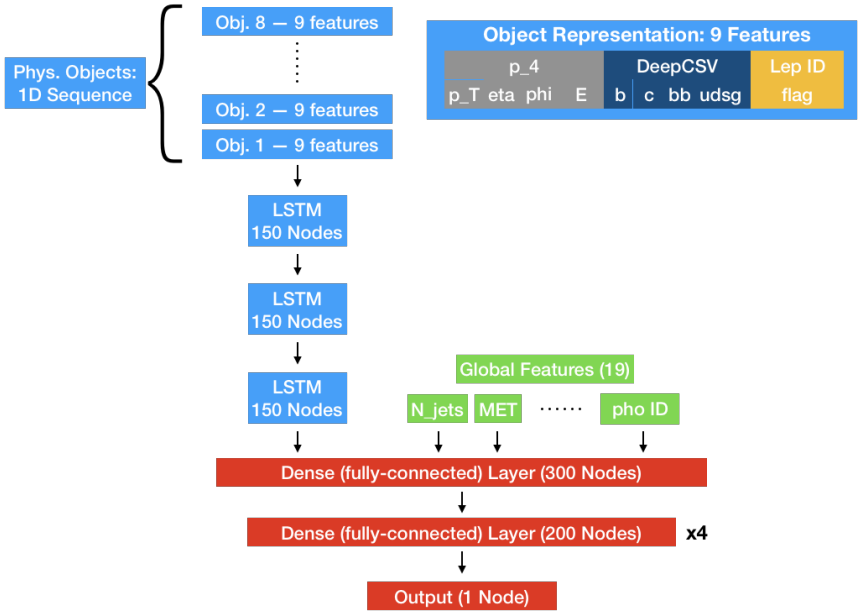

Figure 6.10: Schematic of deep neural network architecture, shown for the leptonic channel.

Training Details

The DNN is implemented in `keras` [24] with `tensorflow` [104] backend. It uses the Adam optimizer [98] with a learning rate of of 10^{-3} and with a binary cross-entropy loss function. It is trained with an early-stopping procedure in which the batch size is increased over the course of training [112]. Starting with a batch size of 1024, the DNN is trained until the improvement in 1-AUC, as calculated on the validation set, after each epoch is less than 1%, at which point the batch size is quadrupled. This procedure is repeated until the batch size exceeds 50,000, at which point it is capped and the training is stopped when the validation AUC ceases to improve. The batch size is capped due to technical limitations of the GPU: large batch sizes take up a large amount of memory. DNN hyperparameters are summarized in Table 6.5.

A number of regularization methods are employed, which were found to improve performance and/or convergence speed during training. First, training features are preprocessed with a "Z-score"

Table 6.5: Hyperparameters for the deep neural networks used in both the hadronic and leptonic channels.

Hyperparameter	Value(s)
Number of nodes (fully connected layers)	300, 200, 200, 200, 200
Number of nodes (LSTM layers)	150, 150, 150
L2-normalization constraint ("maxnorm")	3
Dropout rate	0.1
Learning rate	10^{-3}
Batch momentum	0.99
Activation function (LSTM)	hyperbolic tangent
Activation function (fully-connected layers)	exponential linear unit
Activation function (output layer)	sigmoid

normalization procedure, subtracting the mean and dividing by the standard deviation of each feature such that all features have zero mean and unit variance. In addition to the Z-score transformation, input features with units of GeV are given in terms of their logarithm, with the log taken before the Z-score transformation. The aim of preprocessing is to provide a uniform scale for all input features, as this results in faster convergence during training and even improved performance [101]. In the same spirit as feature preprocessing, batch normalization [93] is applied between the fully-connected layers of the DNN, normalizing each layer's inputs. Batch normalization is not applied in between the LSTM layers. Both feature preprocessing and batch normalization resulted in faster convergence and improved performance of the DNNs. In addition, dropout [113] is applied between the fully-connected layers in order to reduce overfitting and improve performance. Between layers which apply both batch normalization and dropout, batch normalization is applied first and dropout is applied second.

Performance

Three separate DNNs are trained:

- Hadronic channel: $t\bar{t}H$ vs. $\gamma\gamma+$jets
- Hadronic channel: $t\bar{t}H$ vs. $t\bar{t}+\gamma\gamma$
- Leptonic channel: $t\bar{t}H$ vs. $t\bar{t}+\gamma\gamma$

The output of each DNN is shown for both data and simulation in Fig. 6.11. The DNNs are used as inputs to the BDT-bkg, rather than in place of the BDT-bkg because superior performance from DNNs was only

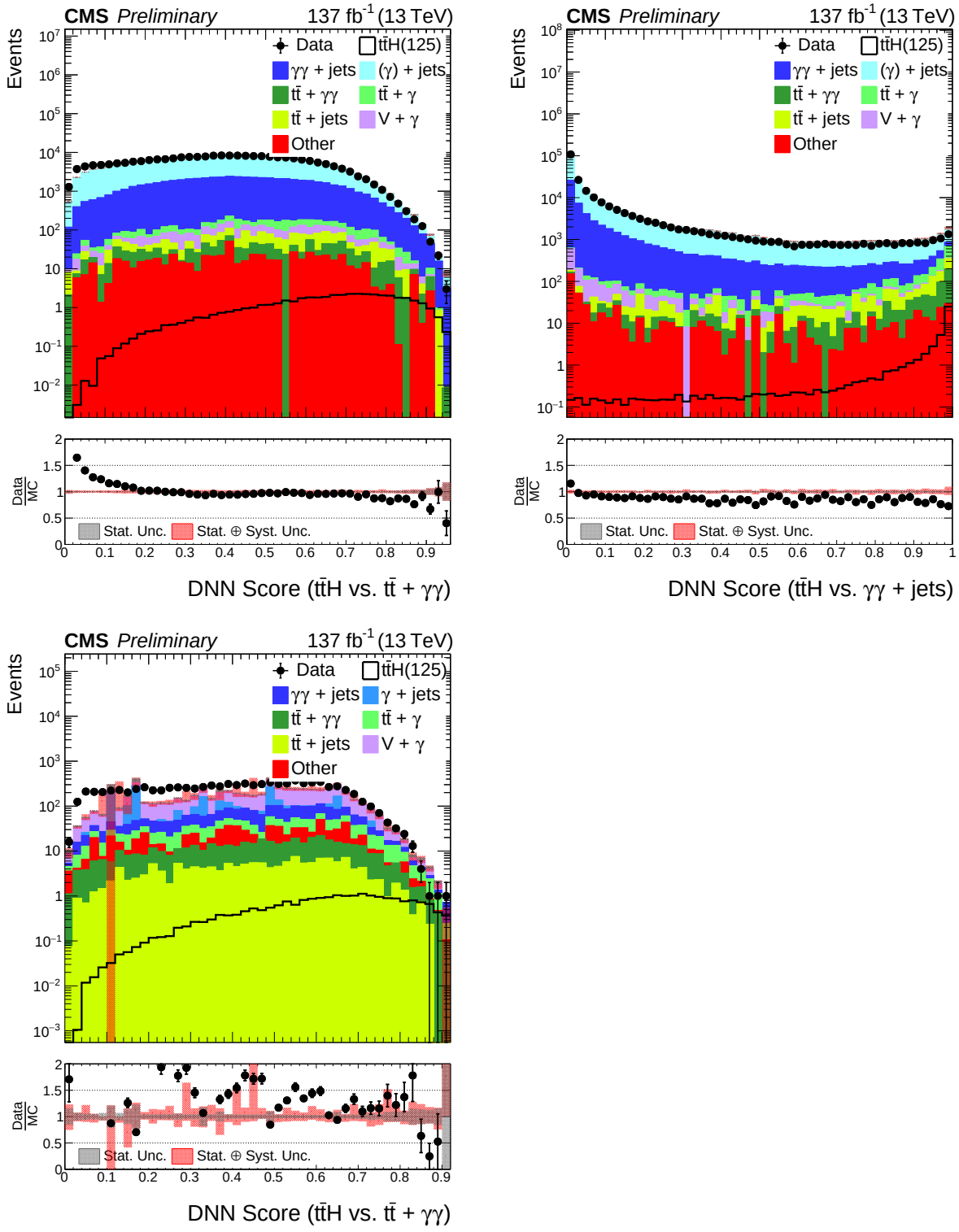

Figure 6.11: Agreement between data and MC description of background for the various DNNs used as input features to BDT-bkg, for the hadronic channel (top) and the leptonic channel (bottom).

observed in the case of a high number of events available for training. The simulation samples describing $t\bar{t}H$, $\gamma\gamma+$jets, and $t\bar{t}+\gamma\gamma$ processes each have a high number of individual events ($\geq 10^5$) passing the preselection requirements, so DNNs are trained to distinguish between these processes. The improvement brought to each channel by the DNNs is shown in Fig. 6.12. The improvement in expected sensitivity in

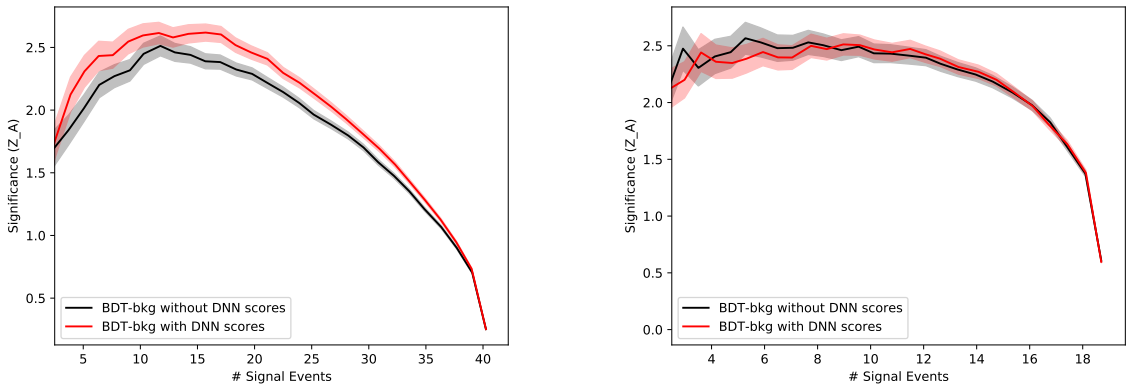

Figure 6.12: Expected significance (Z_A) shown as a function of the number of $t\bar{t}H$ events passing a given cut on BDT-bkg for versions of BDT-bkg trained with (red) and without (black) the DNN scores as training features. Shaded bands show the $\pm 1\sigma$ statistical uncertainty in Z_A. The background yield is estimated from events in data in the $m_{\gamma\gamma}$ sidebands.

the hadronic channel is about 10%. The improvement in the leptonic channel is significantly smaller than the statistical uncertainty in Z_A. The fact that the improvement is greater in the hadronic channel is likely attributed to the fact that it benefits from the DNN trained against the $\gamma\gamma+$jets background. Not only is this the largest background in the $t\bar{t}H$ hadronic channel signal regions, but the simulation sample has the largest number of events entering the preselection, allowing for aggressive DNN training with lower risk of overfitting.

6.5.3 Top Tagger BDT

The dominant backgrounds in the hadronic channel at both preselection level and signal region level are the multi-jet, $\gamma+$jets, and $\gamma\gamma+$jets processes. An obvious difference between these processes and $t\bar{t}H$ is the fact that there are two top (anti-)quarks in the latter, while there are none in the former. This motivates the use of methods which can identify the presence of top quarks as a tool for further rejecting the multi-jet, $\gamma+$jets, and $\gamma\gamma+$jets backgrounds.

Table 6.6: Input features used in training the Top Tagger BDT.

Category	Features		
Single Jet Quantities	p_T	mass	
	DeepCSV b	DeepCSV c vs. light	DeepCSV c vs. b
	ptD	axis1	multiplicity
Di-Jet Quantities	ΔR(j,j)	m_{jj}	
Tri-Jet Quantities	ΔR(b,W)	m_{jjj}	

The chosen method is a top tagger BDT, originally developed in a search for supersymmetric partners of the top quark [25] and later updated for $\mathrm{t\bar{t}H}$. The BDT is trained using XGBOOST [22]. The BDT takes jet triplets as inputs, with triplets that are matched as coming from a top quark (using generator truth-level information) designated as signal and all other triplets designated as background. Jets are required to have $p_T > 25$ GeV and $|\eta| < 2.4$. In addition, the jets are cleaned such that they are not overlapping with leptons. The truth matching enforces the following additional requirements:

- $|m_{jjj} - m_t| < 80$ GeV
- All three reco jets are matched to generator-level quarks from a hadronically decaying top (ΔR(jet, quark) $<$ 0.4).

The training features are shown in Table 6.6. The training features are defined as follows:

- DeepCSV scores: for each jet, three DeepCSV quantities are provided: the b-tag score, and the c-tag score, given in terms of c vs. light and c vs. b.
- ptD, axis1, multiplicity: standard quark-gluon discrimination variables. The fragmentation function is defined as ptD $\equiv \frac{\sqrt{\sum p_{\mathrm{T}_i}^2}}{\sum p_{\mathrm{T}_i}}$, axis1 is the jet shape variable describing the jet's long axis, and multiplicity provides the number of constituents in the jet.

The jet in the triplet with the highest DeepCSV b score is labeled as the b-jet, while the other two jets are labeled as W-jet 1 and 2, with $p_{\mathrm{T}}(\mathrm{W}_{j1}) > p_{\mathrm{T}}(\mathrm{W}_{j2})$.

The ouput of the top tagger BDT is shown for both data and simulation in Fig. 6.13. Similar to the DNN scores, the top tagger BDT is given as an additional training feature to BDT-bkg. The improvement in expected sensitivity gained by adding the top tagger to BDT-bkg is shown in Fig. 6.14 and is about 5%.

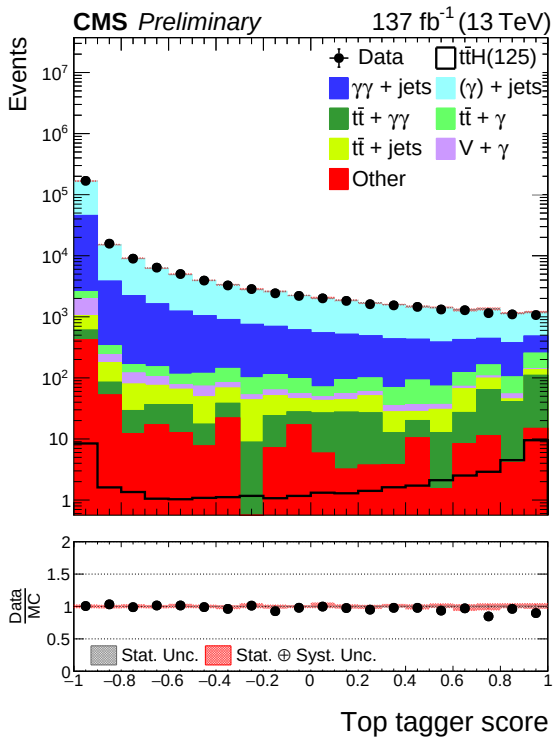

Figure 6.13: Agreement between data and MC description of background for the top tagger BDT score.

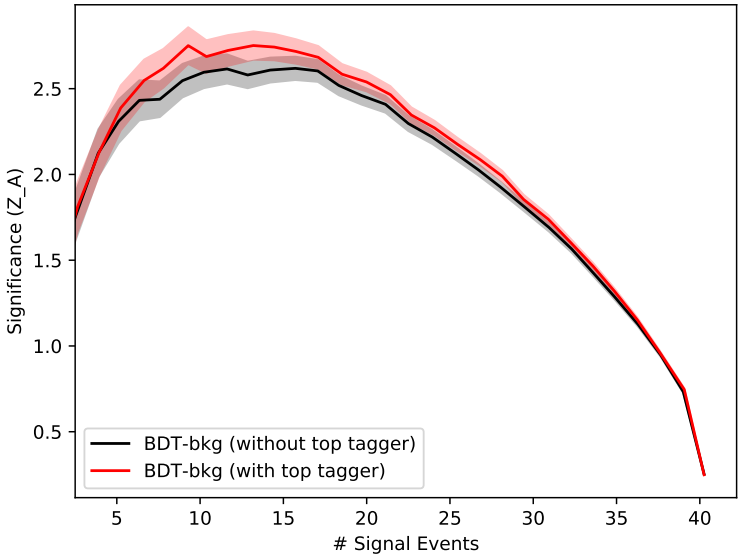

Figure 6.14: Expected significance (Z_A) shown as a function of the number of tt̄H events passing a given cut on BDT-bkg for versions of BDT-bkg trained with (red) and without (black) the top tagger BDT as a training feature. Shaded bands show the $\pm 1\sigma$ statistical uncertainty in Z_A. The background yield is estimated from events in data in the $m_{\gamma\gamma}$ sidebands.

6.5.4 BDT-bkg

The final BDT-bkg algorithms for each channel use the high-level features listed in Table 6.3, the DNN scores described in Sec. 6.5.2, and the top tagger BDT score described in Sec. 6.5.3 as the training features. The outputs of the BDT-bkg algorithms are shown in Fig. 6.15, where agreement between data and simulation is observed within statistical and systematic uncertainties.

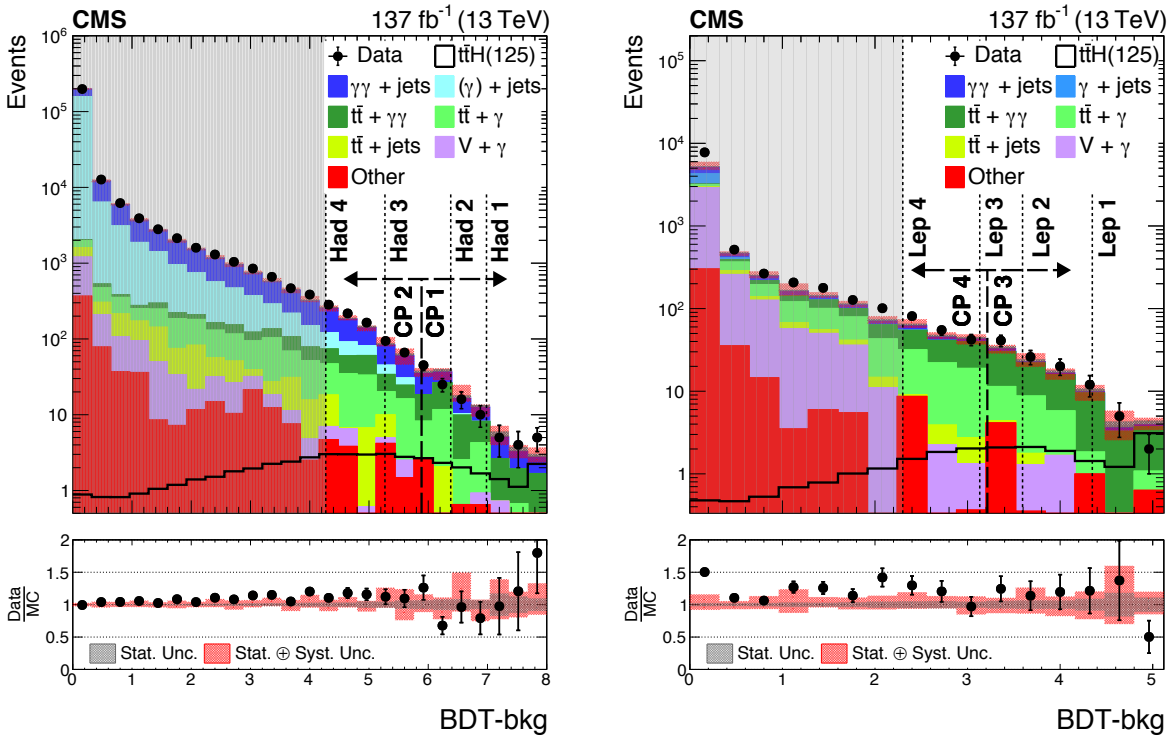

Figure 6.15: Output of the BDT-bkg algorithm for the hadronic channel (left) and the leptonic channel (right). Events from the $m_{\gamma\gamma}$ sidebands are shown for both data and simulation. The statistical (statistical $\oplus$ systematic) uncertainties in simulation are shown with black (red) shaded bands. The thinly dashed lines show the boundaries of each signal region used for the cross section measurement, while the thickly dashed lines show the boundaries of signal regions used for a measurement of the CP structure. Events in the gray shaded region are discarded. Taken from [50].

Validation in $t\bar{t}Z$ Events

As an additional check of the agreement between data and simulation in the output of the BDT-bkg algorithms, a validation in a control region targeting $t\bar{t}Z$ ($Z \to e^+e^-$) events is performed. The rationale for using $t\bar{t}Z$ events is the following: $t\bar{t}Z$ and $t\bar{t}H$ should have very similar kinematic distributions for

most components. Therefore, BDT-bkg should similarly assign high scores to $t\bar{t}Z$ events as it does for $t\bar{t}H$ events and a region of high $t\bar{t}Z$ purity should be present at high scores. Because the cross section times branching fraction of $t\bar{t}Z$ ($Z \to e^+e^-$) is much higher than that of $t\bar{t}H$ ($H \to \gamma\gamma$), this control region provides a test of the agreement between data and simulation for high scores of BDT-bkg with much smaller statistical uncertainty than obtained in the $m_{\gamma\gamma}$ sidebands. Lastly, it is important to emphasize that this check is qualitative in nature; no formal estimate of the compatibility between the distributions in data and simulation is performed.

The selection for the $t\bar{t}Z$ control region is the same as the preselection, with the exception that the cut on the conversion-safe electron veto is inverted to select for $Z \to e^+e^-$ events. The reconstructed "diphoton pairs" in these events are then primarily composed of electrons that were reconstructed as photons.

Additional cuts are next applied on top of the preselection in order to increase the $t\bar{t}Z$ purity:

- $|m_Z - m_{\gamma\gamma}| < 10$ GeV
- $N_{\text{jets}} \geq 5(3)$ for hadronic (leptonic)
- $N_{\text{b-jets}} \geq 2$, using the tight (medium) working point for hadronic (leptonic)

One subtlety in validating the BDT-bkg performance in $t\bar{t}Z$ events is the fact that BDT-bkg is specifically trained to reject events in which electrons are reconstructed as photons through use of the pixel seed veto (defined in Sec. 5.4.1). This challenge is addressed by manually changing the value of the pixel seed veto for these events before evaluating their score with the BDT-bkg algorithm: although most of these events fail the pixel seed veto, BDT-bkg is told that they all pass the pixel seed veto. This hard-coding of the pixel seed veto ensures that $t\bar{t}Z$ events are not assigned lower scores due to the fact that the typical diphoton pair in these events is suspiciously electron-like.

Comparisons of BDT-bkg between data and simulation are shown for events entering the $t\bar{t}Z$ control regions in Fig. 6.16.

The purity of $t\bar{t}Z$ events is much higher in the leptonic channel, where good agreement between data and simulation is observed at high BDT-bkg scores. Good agreement is also observed in the hadronic

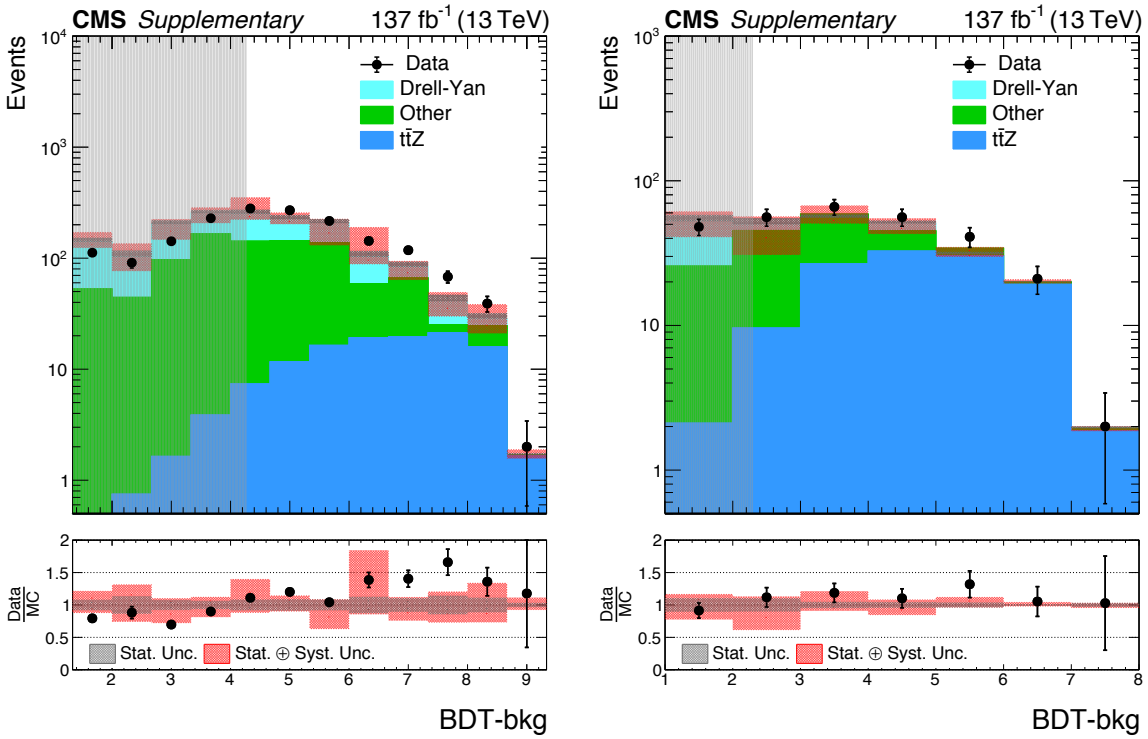

Figure 6.16: Output of the BDT-bkg algorithm for the hadronic channel (left) and the leptonic channel (right) in the $t\bar{t}Z$ control region. The statistical (statistical $\oplus$ systematic) uncertainties in simulation are shown with black (red) shaded bands. Events in the gray shaded region are discarded. Taken from [50].

channel of the $t\bar{t}Z$ control region, though the purity of $t\bar{t}Z$ events is lower. Disagreement is present at lower values of BDT-bkg; however, these events do not enter the analysis selection.

6.6 Event Categorization

The scores of the BDT-bkg algorithms, shown in Fig. 6.15, are used to define the signal regions in which the $t\bar{t}H$ cross section measurement is performed. The signal region boundaries are chosen such that the expected signifiance of the measurement is maximized. They are determined with the following iterative procedure:

1. Determine the N cut values of BDT-bkg score that correspond to $N+1$ intervals evenly spaced in $t\bar{t}H$ efficiency. N is chosen as 100.

2. For each BDT-bkg cut value x, divide events into two regions: $[x_{\text{min}}, x]$ and $[x, x_{\text{max}}]$.

3. Within each region, create parametric models of the signal and background distributions as a function of $m_{\gamma\gamma}$.

   - The $t\bar{t}H$ signal model is estimated from simulation by fitting a Double Crystal Ball function [105] to the $m_{\gamma\gamma}$ distribution.
   - The background model is estimated from the MC description of the background by fitting an exponential function to the $m_{\gamma\gamma}$ distribution.
   - Other standard model Higgs boson production modes are included in the background model, and as the signal, are estimated by fitting a Double Crystal Ball function.
   - Likelihood functions are then constructed for (1) signal + background scenario and (2) background-only scenario. The expected significance σ is calculated as (more detail on this is given in Sec. 6.9.1):

$$\sigma = \sqrt{-2\Big(\log[L_{\text{S+B}}(m_{\gamma\gamma})] - \log[L_{\text{B}}(m_{\gamma\gamma})]\Big)} \tag{6.8}$$

4. If the splitting at x into two signal regions improves the expected significance by more than 2%, the procedure is then repeated iteratively within each signal region. The procedure is terminated when an additional splitting fails to improve the expected significance by at least 2%.

The optimization procedure results in four signal regions for each channel, with the values of BDT-bkg defining each region shown with the thinly dotted lines in Fig. 6.15. The signal and background modeling in the signal region optimization procedure is similar to what is done in the final statistical analysis, described in Sec. 6.7, but does not use the same level of rigor in selecting functional forms. This simplified method is chosen for the optimization for the sake of computing speed and is expected to influence the final boundary selection negligibly.

In order to avoid introducing bias in the result, the $t\bar{t}H$ signal yields are estimated using separate simulation samples from those used in the final analysis and the background yields are estimated using the MC description of the background, rather than events from data.

6.7 Signal & Background Models

The $t\bar{t}H$ cross section measurement is extracted by performing a maximum likelihood fit of the signal and background models to the diphoton invariant mass distribution ($m_{\gamma\gamma}$) observed in data. This fit, described in full detail in Sec. 6.9, relies on the construction of reliable models of the signal and background processes, described in this section.

6.7.1 Signal Models

Models of signal ($t\bar{t}H$) and other standard model Higgs boson production modes (which are considered as backgrounds for the $t\bar{t}H$ cross section measurement) are built as a function of $m_{\gamma\gamma}$, using a Double Crystal Ball plus Gaussian function. A separate fit is performed for each signal region in each channel. Additionally, the fits are performed independently for each year of data-taking: 2016, 2017, and 2018, with the final signal model taken as the sum of the signal models for each of the three years, scaling the normalization of the signal model for each year by the appropriate luminosity. Signal fits are performed separately by year in order to capture the changes in $m_{\gamma\gamma}$ resolution in each year, due to the evolving CMS

ECAL. As the mass of the Higgs boson, m_{H}, is not precisely known, the fit parameters of the signal models are modeled as linear functions of m_{H}. The m_{H} dependence is determined by fitting signal models with simulation samples corresponding to three different values of m_{H}: 120, 125, and 130 GeV. With three years, two channels and four signal regions per channel, this results in 24 signal models per Higgs boson production mode. Some representative signal models are shown for $\mathrm{t\bar{t}H}$ in Fig. 6.17.

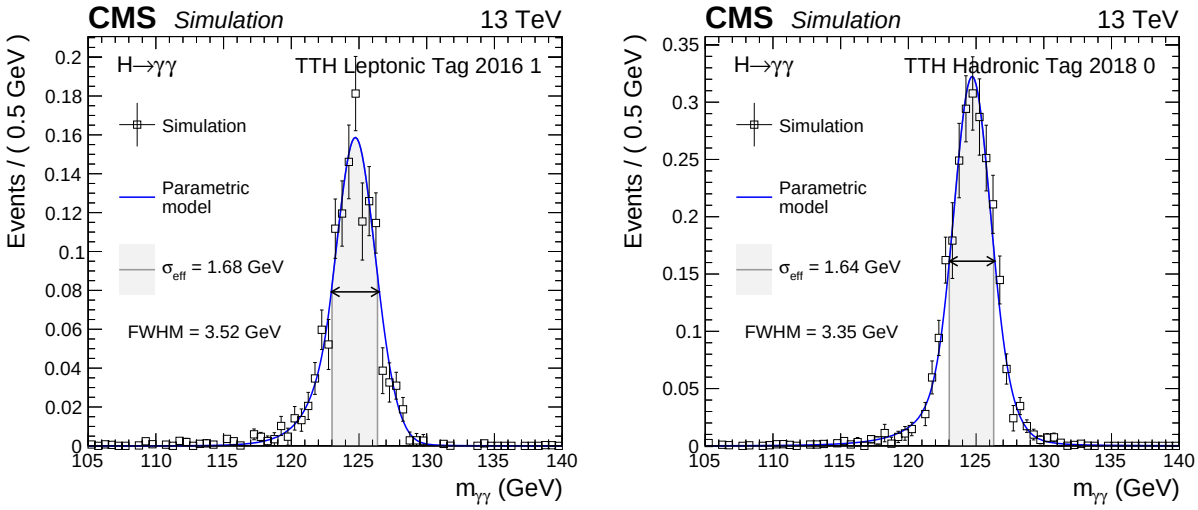

Figure 6.17: Fitted signal models for simulation of $\mathrm{t\bar{t}H}$ production, shown for leptonic tag 1 in 2016 (left) and hadronic tag 0 in 2018 (right).

6.7.2 Background Models

The background model in each category represents the smoothly falling spectrum of events in $m_{\gamma\gamma}$ arising from processes other than Higgs boson production. The exact shape of this spectrum is not known, so a variety of functional forms are used to fit the $m_{\gamma\gamma}$ distribution. Moreover, different choices for the functional form will generally result in different predictions for the background yield under the m_{H} peak. For this reason, the choice of functional form used to describe the smoothly falling background is treated as a discrete nuisance parameter. This strategy is known as the "discrete profiling method", first decribed in Ref. [61].

There are four families of functions considered for the background fits:

1. Exponential

$$f_N(x) = \sum_{i=0}^{N} a_i \exp(-b_i x) \tag{6.9}$$

2. Power Law

$$f_N(x) = \sum_{i=0}^{N} a_i x^{-b_i} \tag{6.10}$$

3. Bernstein polynomial

$$f_N(x) = \sum_{i=0}^{N} a_i \binom{N}{i} x^i (1-x)^{N-i} \tag{6.11}$$

4. Laurent series

$$f_N(x) = \sum_{i=0}^{N} a_i x^{-4+\Sigma_{j=0}^{i}(-1)^j j} \tag{6.12}$$

The a_i and b_i are the parameters to be fitted in each case. In general, as the order N of each family of function is increased, the function gains more tunable parameters and can better fit any arbitrary distribution. In order to determine the optimal order N of each function that is considered, an F-test [72] is employed to assess the improvement in goodness-of-fit brought by using a higher-order function in the context of the increase in function complexity; a higher-order function is selected only if the improvement is greater than some threshold, chosen to penalize more complex functions. The final set of functions and their respective orders considered for the background models are shown for a few representative signal regions in Fig. 6.18. Unlike the fits for the signal models, the fits for the background models are performed inclusively for all three years of data-taking. The best-fit function for each signal region is taken as the nominal value of the background. The final background models, along with uncertainties, are shown in Fig. 6.19 for the same signal regions as shown in Fig. 6.18.

6.8 Systematic Uncertainties

As the background is estimated from data (and not simulation), the uncertainties associated with the background yield are either statistical in nature or associated with the details of the fitting procedure. The latter uncertainties, which are systematic in nature, are those associated with estimating the background through a fit to events in the $m_{\gamma\gamma}$ sidebands and are treated through the discrete profiling method mentioned

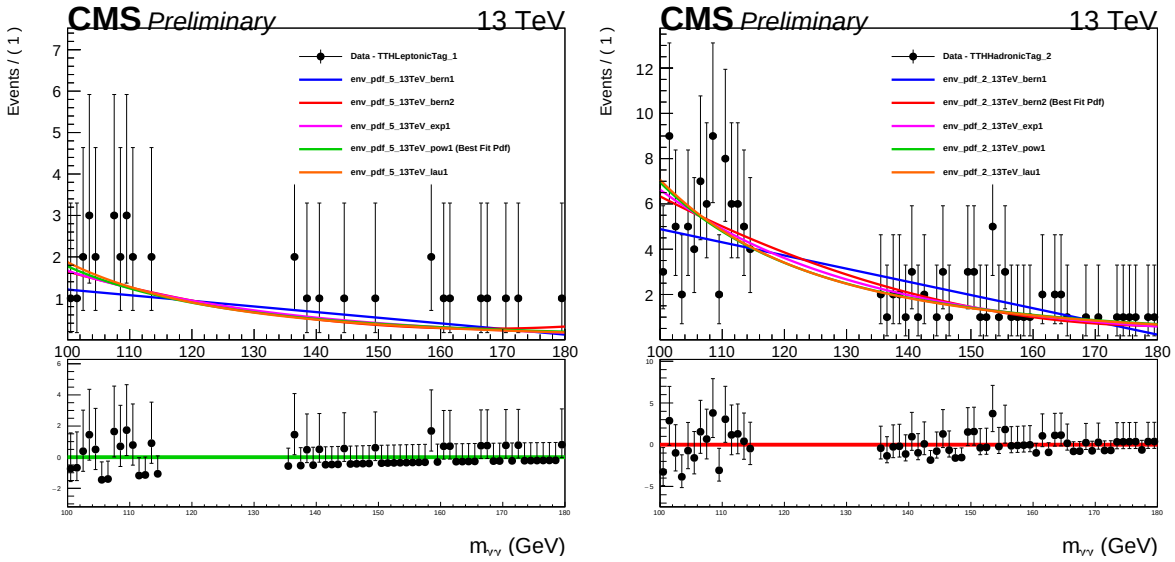

Figure 6.18: Families of functions considered for the background model, shown for leptonic tag 1 (left) and hadronic tag 2 (right).

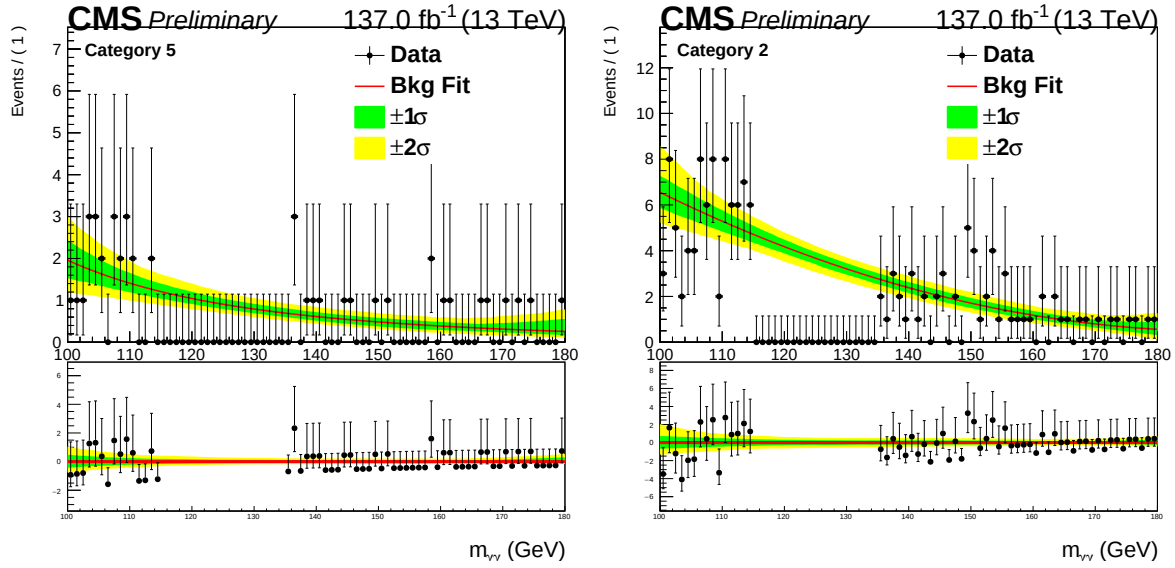

Figure 6.19: Families of functions considered for the background model, shown for leptonic tag 1 (left) and hadronic tag 2 (right).

in Sec. 6.7.2.

In contrast to the relatively simple uncertainties associated with the background model, there are many sources of uncertainty may affect either the yield or $m_{\gamma\gamma}$ shape of the Standard Model Higgs boson production modes (often both).

Those affecting only the yield are modeled with log-normal pdfs, which approximate the behavior of a Gaussian pdf in most cases but have one distinct advantage. A Gaussian pdf is unbounded, meaning the quantity modeled by the pdf could, in principle, take on any value with non-zero probability. When modeling the uncertainty of positive definite quantities like cross sections, luminosities, and efficiencies, the possibility of a negative value is unphysical. A log-normal pdf is defined for a parameter θ as

$$p(\theta) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left[-\frac{1}{2}\left(\frac{\ln(\theta/\hat{\theta})}{\ln\kappa}\right)^2\right]\frac{1}{\theta}, \tag{6.13}$$

where $\hat{\theta}$ is the most probable value and κ is analogous to the width σ of a Gaussian.

The sources are divided into theoretical uncertainties, those relating to imperfect knowledge of Higgs boson production in the Standard Model, and experimental uncertainties, those related specifically to the CMS detector and the methods used to reconstuct each event.

6.8.1 Theoretical Uncertainties

Each theoretical uncertainty described below is calculated separately for each Higgs boson production mode. Generally, the uncertainty for $t\bar{t}H$ has the largest effect on the measurement of $\mu_{t\bar{t}H}$, though the uncertainties for other production modes can be significant as well. Uncertainties are either calculated as an overall uncertainty on the normalization, a "flat uncertainty", or as individual variations on a per-event basis. The flat uncertainties affect only normalization, while the individual uncertainties may also modify kinematics and result in event migration between signal regions.

- *Strong coupling constant* (α_s): flat uncertainty in the value of the coupling constant of the strong force, α_s. The magnitude is taken following the PDF4LHC prescription [95], and is about 2% for $t\bar{t}H$.

- *PDF (parton density function)*: uncertainty due to imperfect knowledge of the structure of the proton. Two distinct PDF uncertainties are considered: a flat uncertainty, computed following the PDF4LHC prescription [95, 85], and per-event PDF weight variations, taken from the NNPDF3.0 PDF set [54] with the `MC2Hessian` procedure [21]. The flat uncertainty is about 3% for $t\bar{t}H$ and the per-event PDF weight uncertainties are typically $\leq 1\%$.

- *QCD scale*: the uncertainty in the renormalization and factorization scales. Values are taken following the recommendations of [86] and is nearly 10% for $t\bar{t}H$, making it the single largest systematic uncertainty.

- $H \to \gamma\gamma$ *branching fraction*: estimated to be around 2% [86].

- ggH *contamination*: the standard model predictions of Higgs boson production via gluon fusion (ggH) are not reliable in the $t\bar{t}H$ regime with a high number of jets. Three distinct sources contribute to this uncertainty:

    1. *Parton shower*: the uncertainty in the gluon fusion yield at a high number of jets (i.e. the uncertainty in the parton shower modeling) is taken from the difference between the jet multiplicity in simulation and that observed in data for fully leptonic $t\bar{t}+$jets events, where the dominant production mode is via gluon fusion.

    2. *Gluon splitting modeling*: the uncertainty in gluon splitting to b quarks is taken from the difference between data and simulation in the ratio $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}j\bar{j})$.

    3. *Statistical*: uncertainty in the ggH estimate due to limited number of simulated events in the high-jet regime.

6.8.2 Experimental Uncertainties

Like the theoretical uncertainties, experimental uncertainties may either be described globally as a flat uncertainty or on a per-event basis. Additionally, the per-event uncertainties may either modify the central weight of the event (i.e. normalization) or the shape of the $m_{\gamma\gamma}$ distribution. The uncertainties which affect the shape of the $m_{\gamma\gamma}$ distribution are accounted for by performing separate fits of the signal models for the up and down variations of each uncertainty source. These then manifest themselves as uncertainties in the fitted parameters of the Gaussian and Double Crystal Ball functions used to model each Higgs boson production mode.

The uncertainty sources affecting the shape of the $m_{\gamma\gamma}$ distribution are:

- *Photon energy scale & resolution*: the uncertainty associated with the corrections derived for the

photon energy scales and resolution, described in Sec. 5.4.2, is estimated by varying the shower shape variable R_9, the electron ID criteria, and the preselection E_T requirement. The variations from each of these sources are added in quadrature with the statistical uncertainty to give the total uncertainty. Separate uncertainties are considered for both the scale and the resolution and each source is additionally split into contributions from low R_9, high R_9 $\odot$ barrel, endcap.

- *Residual p_T dependence of scale corrections*: photon energy scale corrections are derived in $Z \to e^+e^-$ events with $p_T \sim 45$ GeV but applied in $H \to \gamma\gamma$ events with $p_T \sim 60$ GeV, which may introduce error. This uncertainty is conservatively estimated as the magnitude of the correction itself, translating to a 0.1% uncertainty in the overall photon energy scale.

- *Differences between electrons and photons*: nearly all corrections, smearings, scale factors, etc. are derived on electrons in $Z \to e^+e^-$ events, but applied on photons. Several differences between electrons and photons (and their reconstruction in the CMS detector) are used to estimate the uncertainty:

    1. *Modeling of the material budget*: in general, electrons shower earlier than photons when passing through the CMS detector. The uncertainty in the material between the interaction point and the ECAL then translates to an additional source of uncertainty.

    2. *Non-uniformity of light collection*: differences in the light collection efficiency (LCE) along the length of ECAL crystals result in a different response to electrons and photons (again due to the fact that electrons shower earlier than photons). This uncertainty is estimated using the LCE model described in [4], derived from optical simulation [75].

- *Shower shape corrections*: the shower shape corrections described in Sec 5.4.3 may effect the photon energy scale. The uncertainty is estimated by comparing the energy scale before and after the application of corrections.

The remaining sources of uncertainty affect only the overall normalization of a given process, and include:

- *Shape of the b-tagging discriminant*: the b-tagging discriminant is corrected in simulation by a continuous reshaping factor (derived as a function of p_{T}, η, and jet flavor) such that the distributions between data and simulation agree. The uncertainty in the reshape factor for a given event is calculated as described in [44] and has an impact of about 4% on the $\mathrm{t\bar{t}H}$ signal strength measurement.

- *Integrated luminosity*: the total uncertainty in the integrated luminosity is estimated to be about 2% [1, 2, 3].

- *Trigger Scale Factor*: the efficiency of the HLT triggers used for this analysis are calculated using the tag-and-probe method. Simulation is then corrected for this trigger efficiency, with the uncertainty in the efficiency taken as a systematic uncertainty.

- *Diphoton Preselection Scale Factor*: the efficiency of the diphoton preselection is calculated for both data and simulation with the tag-and-probe method and a scale factor is derived from the ratio between data and simulation. The uncertainty in this scale factor (binned by barrel/endcap and low/high R_9) is then used to calculate the associated systematic uncertainty.

- *Photon Identification BDT Score*: the uncertainty in the photon ID BDT score is assumed to stem from the limited size of the training sample used to derive the corrections for its inputs, the shower shape and isolation variables. The magnitude of this uncertainty is estimated by splitting the original training sample in half and deriving two sets of corrections, taking the uncertainty as the difference between the two trainings.

- *Jet Energy Scale & Resolution*: the scaling and smearing factors derived for individual jets each have associated uncertainties, which are propagated to the final result by varying all factors up/down by their uncertainty.

- E_T^{miss}: there are four individual uncertainty sources associated with the calculation of the $E_{\mathrm{T}}^{\mathrm{miss}}$:

  - Jet energy scale: as described above.

  - Jet energy resolution: as described above.

- Photon energy scale: the energy scales of pf photons used in the calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$ are varied within their uncertainties, with the result propagated through the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation.
- Unclustered pf candidate energy scale: the energy scales of pf candidates not clustered within a jet are varied in the same way.

- *Lepton ID and Isolation*: both electrons and muons have scale factors derived with the tag-and-probe method to account for differences in efficiency between data and simulation, with the uncertainty in these scale factors dictating the resulting systematic uncertainty.

6.8.3 Impact of Systematic Uncertainties

By far, the largest systematic uncertainty is the uncertainty in the QCD renormalization and factorization scales, with an impact of about 10% on $\mu_{\mathrm{t\bar{t}H}}$. Other large uncertainties include those associated with the shape of the b-tagging discriminant, the integrated luminosity, the parton density function, and the $\mathrm{H} \to \gamma\gamma$ branching ratio. The impacts of the dominant systematic uncertainties on $\mu_{\mathrm{t\bar{t}H}}$ are shown in Fig. 6.20.

6.9 Results

6.9.1 Statistical Analysis

The measured parameters of interest (POIs), namely $\mu_{\mathrm{t\bar{t}H}}$, are extracted by constructing a likelihood function which depends on these POIs and finding their values which maximize the likelihood function. The likelihood function expresses the probability of the observed data, given the prediction taken from the signal and background model. More precisely,

$$\mathcal{L}(\text{data} \mid \mu_{\mathrm{t\bar{t}H}}, \vec{\theta}) = \mathcal{L}\left(\text{data} \;\middle|\; \left[S(\mu_{\mathrm{t\bar{t}H}}, \vec{\theta}) + B(\vec{\theta})\right] \times C(\vec{\theta})\right), \tag{6.14}$$

where $\vec{\theta}$ is the vector of nuisance parameters (i.e. those described in Sec. 6.8) which are typically modeled as log-normal distributions (Eqn. 6.13).

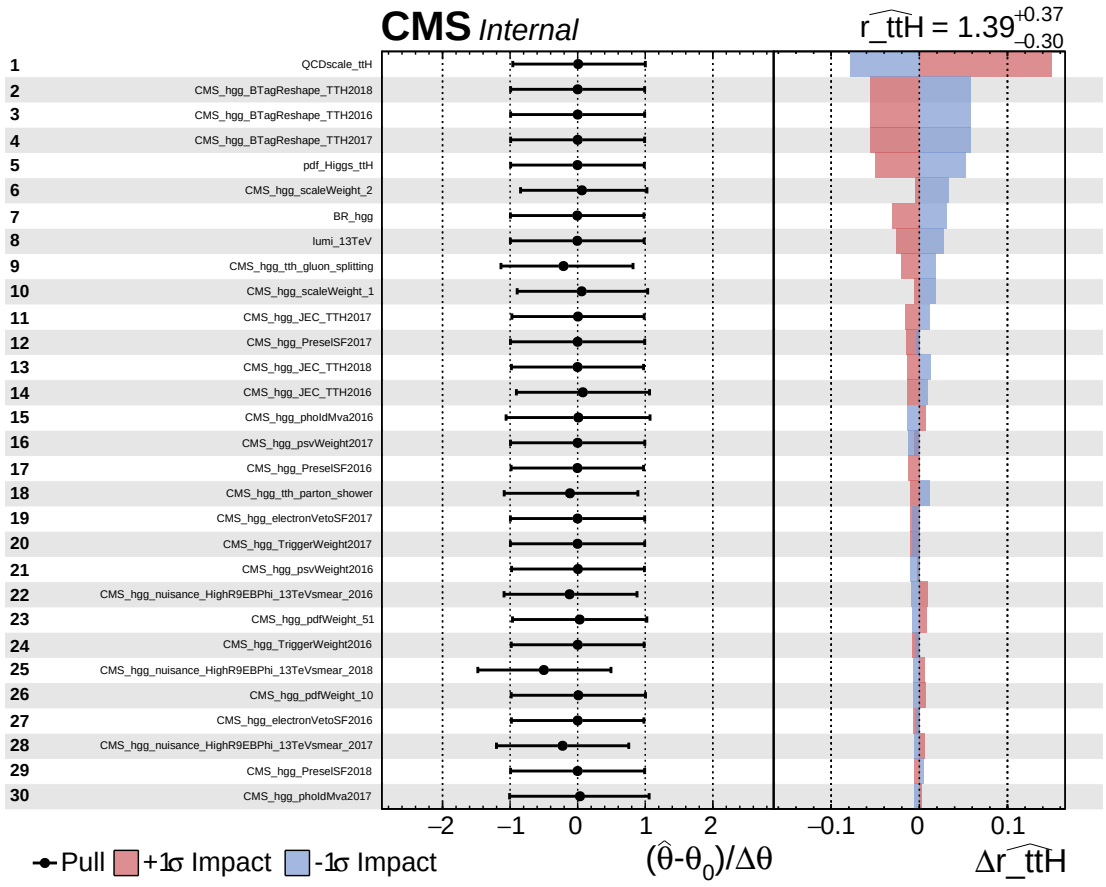

Figure 6.20: Impacts of the dominant systematic uncertainties on the measurement of $\mu_{t\bar{t}H}$.

The fit is performed simultaneously in all signal regions; in other words, the likelihood function is a product of the likelihood functions for each signal region:

$$\mathcal{L}(\text{data} \mid \mu_{\text{t}\bar{\text{t}}\text{H}}, \vec{\theta}) = \prod_{i=1}^{N_{\text{SR}}} \mathcal{L}_i\left(\text{data}_i \;\middle|\; \left[S_i(\mu_{\text{t}\bar{\text{t}}\text{H}}, \vec{\theta}) + B_i(\vec{\theta})\right] \times C(\vec{\theta})\right), \tag{6.15}$$

where $N_{\text{SR}} = 8$ is the total number of signal regions and data_i, S_I, and B_i are the observed data, signal model, and background model in the i-th signal region, respectively.

Moreover, the likelihood function is discretized into bins of 0.25 GeV in the [100, 180] GeV region. The likelihood function in a particular signal region is then

$$\mathcal{L}_i\left(\text{data}_i \;\middle|\; \left[S_i(\mu_{\text{t}\bar{\text{t}}\text{H}}, \vec{\theta}) + B_i(\vec{\theta})\right] \times C(\vec{\theta})\right) = \prod_{j=1}^{N_{\text{bins}}} \text{Poisson}\left(n_{i,j} \;\middle|\; \lambda_{i,j}\right) \times C(\vec{\theta}), \tag{6.16}$$

with $N_{\text{bins}} = 320$ the total number of bins per signal region, $n_{i,j}$ the number of observed data events in the j-th bin of the i-th signal region, and $\lambda_{i,j}$ the expected number of events in that bin

$$\lambda_{i,j} = S_{i,j}(\mu_{\text{t}\bar{\text{t}}\text{H}}, \vec{\theta}) + B_{i,j}(\vec{\theta}), \tag{6.17}$$

and Poisson indicates the standard Poisson distribution

$$\text{Poisson}(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}. \tag{6.18}$$

The bin size of 0.25 GeV is chosen with the characteristic diphoton mass resolution of 1.5-2 GeV in mind – the bin size is sufficiently smaller than the resolution that the information lost by binning the data is negligible.

In practice, -2 times the natural logarithm of the likelihood function is nicer to work with from a numerical optimization point of view, and it is this quantity, referred to as the "log-likelihood", that is actually minimized in the fit:

$$2\text{NLL} = -2\ln(\mathcal{L}). \tag{6.19}$$

In general, the fitted value $\hat{\mu}$ of a POI μ is called the maximum likelihood estimate (MLE) of μ.

The log-likelihood also has desirable qualities for purposes of assessing the uncertainty on fitted POIs. In particular, we may be interested in how much more likely a particular value of a POI μ is than its MLE $\hat{\mu}$, i.e. the uncertainty on the fitted value. To this end, it is helpful to study the quantity

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\vec{\theta}}})}{L(\hat{\mu}, \hat{\hat{\vec{\theta}}})} \tag{6.20}$$

where $\hat{\hat{\vec{\theta}}}$ and $\hat{\hat{\vec{\theta}}}$ are the ML values of $\vec{\theta}$ for μ and $\hat{\mu}$, respectively. The quantity $\lambda(\mu)$ is called the "profile likelihood ratio", and -2 times the logarithm of this quantity is called the "log-likelihood ratio". A convenient property of the log-likelihood ratio is the fact that in the case of a single POI, it approximately follows a χ^2 distribution with one degree of freedom [60]. For this reason, taking the square root of $\lambda(\mu)$ gives the Gaussian significance Z associated with μ [60], where

$$Z \equiv \Phi^{-1}(1-p), \tag{6.21}$$

with Φ the Gaussian quantile function and p the p-value. The frequentist interpretation of p is the following: in the limit of an infinite number of repeated, indepedent experiments in which the true value of the POI is $\hat{\mu}$, a value more extreme than μ would be obtained in p percent of these. The Gaussian significance Z can be interpreted in the following way: a Gaussian-distributed variable found Z standard deviations away from its mean value has an associated p-value of p.

Within this framework, we express the uncertainty on $\hat{\mu}$ in terms of the values of μ corresponding to a 68% (1 standard deviation) CL[1], namely the values of μ which give $\lambda(\mu) = 1$. Another value of μ of particular interest is $\mu = 0$, corresponding to the case of the background-only hypothesis. The associated significance $Z = \sqrt{\lambda(0)}$ is said to be the significance with which the signal has been observed, with $Z = 5$ taken as the threshold for claiming discovery.

[1]The choice of a 68% CL as the default for expressing uncertainties is somewhat arbitrary, and could easily be chosen as some other value.

6.9.2 Cross Section, Signal Strength, & Significance

The observed diphoton mass distributions in the eight signal regions are nicely summarized in a couple plots in Fig. 6.21, which shows the weighted and unweighted sums of the distributions from each signal region. In the case of the weighted sum, the distribution from each signal region is weighted by the factor $S/(S+B)$, giving higher weight to regions with higher purity. S and B are the signal and background yields, defined as the total number of $H \to \gamma\gamma$ events and the total number of non-resonant background events, respectively.

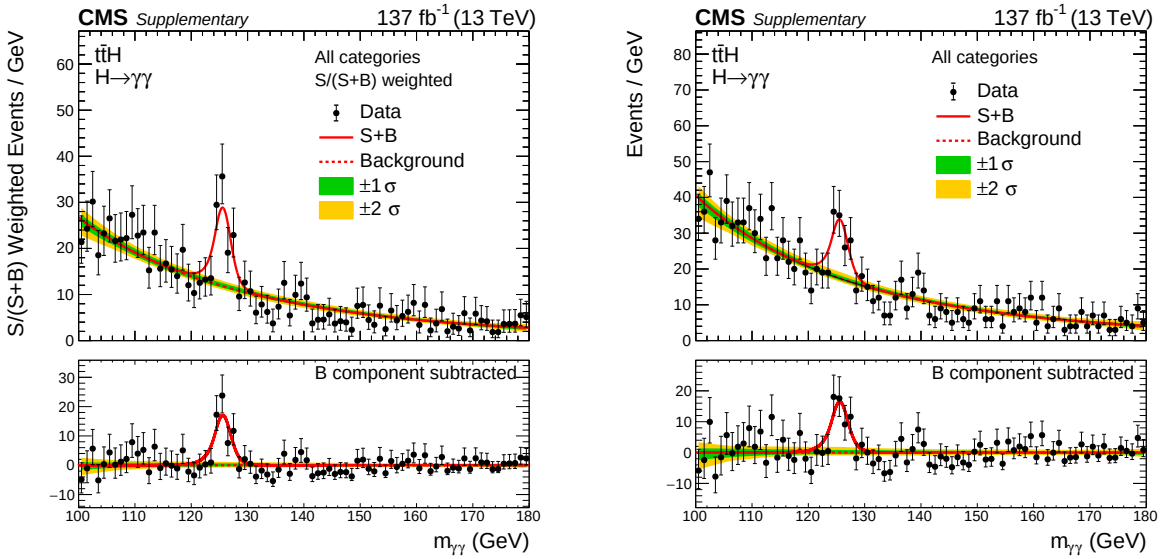

Figure 6.21: Weighted (left) and unweighted (right) sum of observed diphoton mass distributions for all of the signal regions. Events from each signal region are weighted by the respective $S/(S+B)$ of that category in the case of the weighted sum. Taken from [50].

The diphoton mass distributions for each of the eight signal regions are shown individually in Appendix B.

The observed MLE of $\mu_{t\bar{t}H}$ is obtained from minimization of 2NLL of the likelihood function defined in Sec. 6.9.1 and is found to be 1.38. The 68% CL for $\hat{\mu}_{t\bar{t}H}$ is obtained from constructing the log-likelihood ratio defined in Eqn. 6.20 as a function of $\mu_{t\bar{t}H}$ and is found to be $1.38^{+0.36}_{-0.29}$. The log-likelihood ratio is shown in Fig. 6.22.

The observed cross-section times branching fraction of the $t\bar{t}H$ ($H \to \gamma\gamma$) process is found to be $\sigma_{t\bar{t}H}\mathcal{B}(H \to \gamma\gamma) = 1.56^{+0.34}_{-0.32}$ fb, while the SM prediction is $\sigma_{t\bar{t}H}\mathcal{B}(H \to \gamma\gamma) = 1.13^{+0.08}_{-0.11}$ fb.

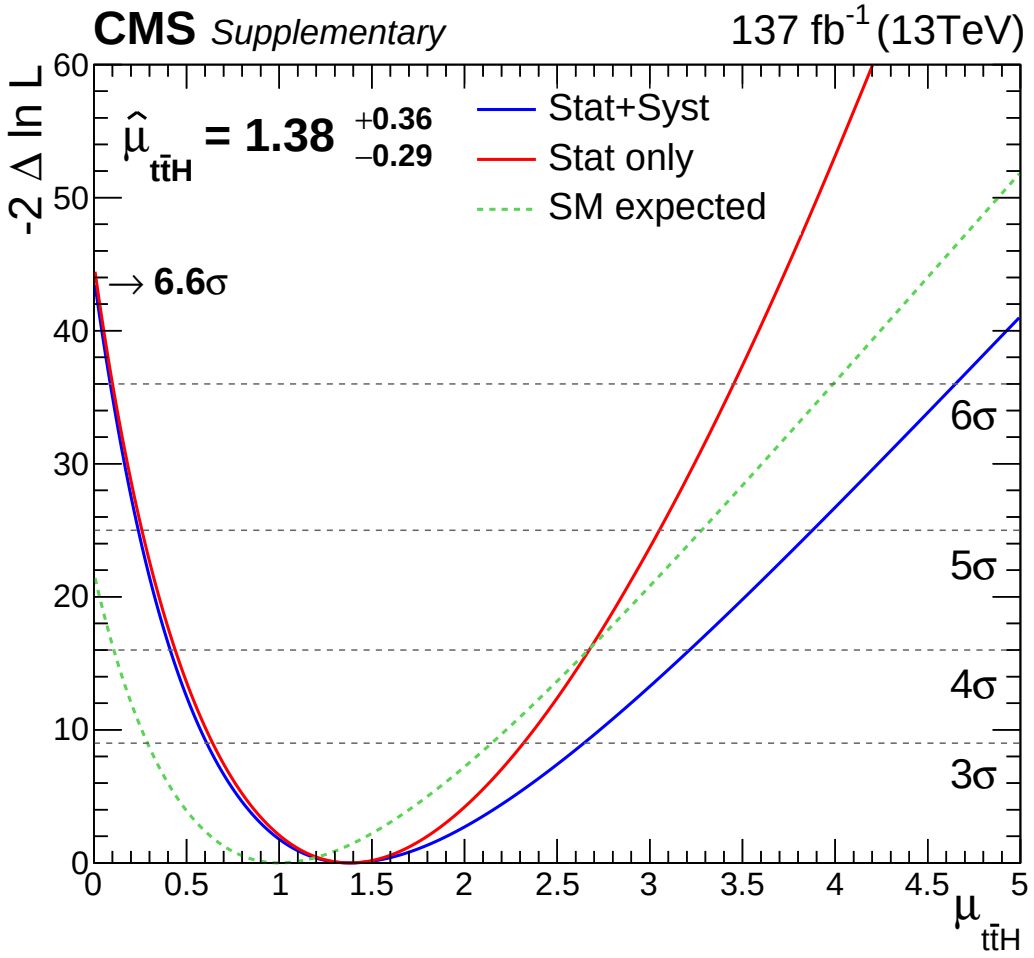

Figure 6.22: Log-likelihood ratio for $\mu_{t\bar{t}H}$. The expected distribution, assuming the SM signal strength $\mu_{t\bar{t}H} = 1$, is shown in the green dotted line. The observed distribution is shown with full uncertainties (only statistical uncertainty) in the blue (red) lines. Taken from [50].

In addition to the MLE of $\mu_{t\bar{t}H}$ and its uncertainty, we are interested in the significance of the observation: the difference between the log-likelihood ratio evaluated at the MLE of $\mu_{t\bar{t}H} = \hat{\mu}_{t\bar{t}H}$ and $\mu_{t\bar{t}H} = 0$, the case of the background-only hypothesis. The observed significance, relative to the background-only hypothesis, is 6.6 standard deviations, while the expected significance is 4.7 standard deviations. With a discovery threshold of 5 standard deviations, we are able to claim observation of the $t\bar{t}H$ ($H \to \gamma\gamma$) process. The observed and expected results for the cross section, signal strength, and significance are shown in Table 6.7.

Table 6.7: Expected and observed values of the cross section times branching fraction ($\sigma_{t\bar{t}H}\mathcal{B}(H \to \gamma\gamma)$), signal strength ($\mu_{t\bar{t}H}$), and significance.

Quantity	Expected Value	Observed Value
$\sigma_{t\bar{t}H}\mathcal{B}(H \to \gamma\gamma)$	1.13 fb	$1.56^{+0.34}_{-0.32}$ fb
$\mu_{t\bar{t}H}$	1.00	$1.38^{+0.36}_{-0.29}$
Significance	4.7σ	6.6σ

6.9.3 CP Measurement

In addition to measuring the cross section and signal strength of the $t\bar{t}H$ ($H \to \gamma\gamma$) process, the CP structure of the tree-level top quark Yukawa (Htt) coupling can also be tested. The SM predicts that the Htt coupling is purely CP-even; any non-zero CP-odd component of the coupling would be an indication of new physics.

A parametrization of the CP structure of the Htt amplitude can be given in terms of CP-even and CP-odd components [84]:

$$A(\text{Htt}) = -\frac{m_t}{v}\bar{\psi}_t\left(\kappa_t + i\tilde{\kappa}_t\gamma_5\right)\psi_t, \tag{6.22}$$

where κ_t and $\tilde{\kappa}_t$ represent the CP-even and CP-odd couplings, respectively, and v represents the SM Higgs field VEV. As the SM predicts a purely CP-even Htt coupling, this implies the SM prediction of $\kappa_t = 1$ and $\tilde{\kappa}_t = 0$. The fractional CP-odd component is defined as

$$f_{\text{CP}}^{\text{Htt}} = \frac{|\tilde{\kappa}_t|^2}{|\kappa_t|^2 + |\tilde{\kappa}_t|^2}\ \text{sign}(\tilde{\kappa}_t/\kappa_t), \tag{6.23}$$

and this is the physical observable constrained by the $t\bar{t}H$ ($H \to \gamma\gamma$) analysis [50].

The CP measurement is performed by starting with signal categories defined by the BDT-bkg algorithm detailed in Sec. 6.5. Next, BDTs, referred to as the $\mathcal{D}_{0-}$ discriminants, are trained to distinguish between CP-even and CP-odd scenarios for the Htt coupling. This is achieved by first simulating $t\bar{t}H$ samples with anomalous couplings, including samples of pure CP-even (SM-like), pure CP-odd, and a mixture of the two. These samples are generated at leading order with the JHUGEN 7.0.2 software package and reweighted with the MELA matrix element library [84, 73, 13, 92]. The $\mathcal{D}_{0-}$ BDTs are trained on the kinematic features described in Sec. 6.5.1, with a separate BDT for the hadronic and the leptonic channels, just as for the BDT-bkg algorithm. In both the hadronic and leptonic channels, two signal categories are formed with requirements on the output of BDT-bkg, with the boundaries shown in Fig. 6.15. Each of these signal categories is further divided into three signal regions, chosen to maximize the expected sensitivity to f_{CP}^{Htt}, giving 12 total signal categories for the CP measurement.

As for the cross section and signal strength measurements, f_{CP}^{Htt} is constrained with a simultaneous fit to the diphoton invariant mass spectrum in all 12 signal categories. Fig. 6.23 shows the results of this fit, which are consistent with the SM prediction of $f_{CP}^{Htt} = 0$. The observed (expected) constraint on the CP structure of the Htt coupling is $f_{CP}^{Htt} = 0.00 \pm 0.33 (0.00 \pm 0.49)$ at 68% CL. The observed (expected) significance with which the pure CP-odd model is excluded is $3.2\sigma (2.6\sigma)$. An additional systematic uncertainty is introduced for the CP measurement to account for potential differences in kinematic distributions obtained through the JHUGEN generator and the MADGRAPH generator used to model the SM processes, though the uncertainty in the measurement is still dominated by the statistical uncertainty.

Thus, the measurement of the CP structure of the Htt coupling is found to be consistent with the SM value of $f_{CP}^{Htt} = 0$.

6.10 Acknowledgements

Chapter 6 describes the $t\bar{t}H$ analysis documented in "Measurements of $t\bar{t}H$ production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel" *Phys. Rev. Lett.* 125 (2020), with a focus on the aspects to which I contributed most directly,

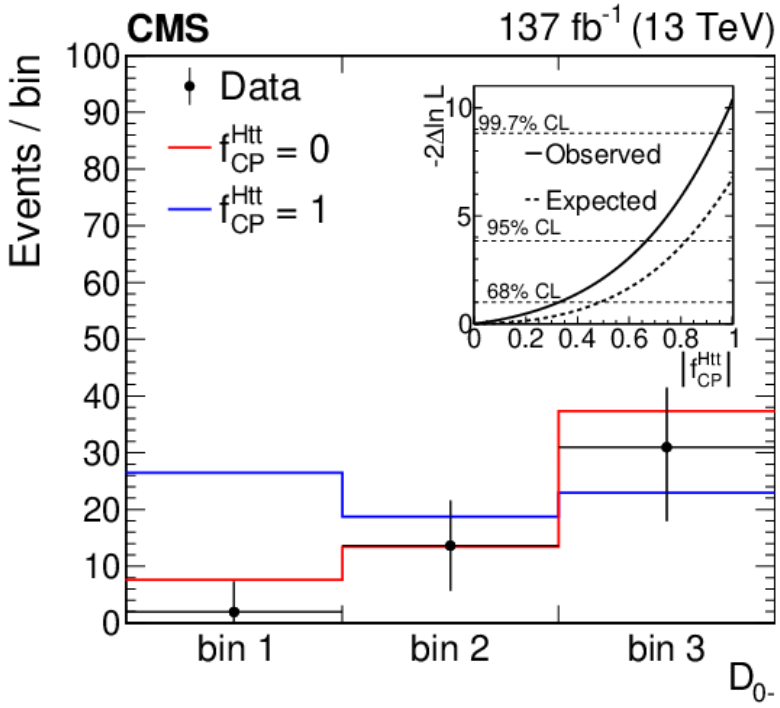

Figure 6.23: Distribution of events, weighted by S/(S+B), selected for the CP measurement of the Htt coupling. Events from both BDT-bkg categories in both the hadronic and leptonic channels are shown in each $\mathcal{D}_{0-}$ bin. The background contribution is subtracted from each bin. The likelihood scan for f_{CP}^{Htt} is displayed in the inner panel. Taken from [50].

but relies heavily from the work of other members of the CMS Collaboration and the CMS Higgs to Gamma Gamma working group, without whom this analysis would not have been possible. My primary individual contributions to this work included the following: implementation of the data-driven description of multi-jet and γ + jet backgrounds, studies of agreement between data and simulation, training and optimization of the deep neural networks and boosted decision trees used for signal region definition, and implementation of some systematic uncertainties. Figures 6.17,6.18, and 6.19 show the results of the signal and background models for the $t\bar{t}H$ analysis, and were produced by Hualin Mei. Figure 6.20 shows the impact of systematic uncertainties on the measurement of $\mu_{t\bar{t}H}$, and was produced by Hualin Mei. Figures 6.21 and 6.22 show the observed results of the $t\bar{t}H$ analysis, and were also produced by Hualin Mei. Figure 6.23 shows the result of the $t\bar{t}H$ CP measurement and was produced by Meng Xiao.

Chapter 7

Conclusion

Measurements of the production cross section and signal strength of Higgs boson production in association with a top quark-antiquark pair in the diphoton decay channel were presented. With an observed significance of 6.6 standard deviations, this is the first observation of $t\bar{t}H$ in a single decay channel of the Higgs boson. The observed cross section times branching fraction of $1.56^{+0.34}_{-0.32}$ fb is compatible with the SM prediction of $1.13^{+0.08}_{-0.11}$ fb, and the observed signal strength of $1.38^{+0.36}_{-0.29}$ is compatible with the SM prediction of unity.

Although the measurements of $t\bar{t}H$ production are so far compatible with the SM predictions, more precise measurements are necessary to determine whether the interactions of the Higgs boson are truly compatible with those predicted by the SM. Many BSM theories predict deviations from the SM couplings at a percent level [118], while the precision of the measurement presented in this thesis is around an order of magnitude higher. As the uncertainty of this measurement is still heavily statistically-dominated, simply repeating the analysis with the larger datasets expected from Run 3 of the LHC and the HL-LHC will improve our ability to judge whether the Higgs couplings are indeed SM-like. The $H \rightarrow \gamma\gamma$ decay channel will especially benefit from increased luminosity, due to its low systematic uncertainties relative to other decay modes, like the decay of the Higgs boson to bottom quarks. Beyond increasing the integrated luminosity of the datasets, the sensitivity of this measurement can be improved in multiple ways, including: through further study of advanced machine learning algorithms used to identify signal-like events, such as

those described in Sec. 6.5.2, through the continued development of creative methods for improving the description of the SM background processes, like that of Sec. 6.4.2, and through the use of methods to decrease the experimental systematic uncertainties associated with the measurement, such as the chained quantile regression method utilized to improve the agreement between simulation and data, as described in Sec. 5.4.3.

The strategies developed in this analysis are broadly applicable to measurements other than just that of $t\bar{t}H$, especially measurements involving $H \to \gamma\gamma$ in the final state. In particular, a similar strategy may be adopted for searches for new physics with $H \to \gamma\gamma$ in the final state, such as a search for the Higgs boson acting as a flavor-changing neutral current [108] in decays of the top quark to a Higgs boson and a light-flavor quark. The strategies may be similarly adopted to measurements of other SM Higgs production modes, such as that of double-Higgs production, which may be a sensitive probe to the Higgs self-coupling [63]. In the same spirit that new physics may present itself in modified interactions of the top quark and the Higgs boson [11], it might also present itself in a modified Higgs self-coupling [96]. Precision measurements of the properties of the Higgs boson will continue to test the limits of the standard model's accuracy and provide a complementary approach to direct searches for new physics.

The results from Run 3 and the HL-LHC will provide unprecedented precision on the properties of the Higgs boson. Either the results will continue to be compatible with the SM predictions, giving us further validation of one of the most successful theories in all of physics, or the results will show disagreement with the SM predictions, giving the field of particle physics a clear area to focus on in the goal of discovering new physics beyond the standard model.

Appendix A

Plots of input features to BDT-bkg

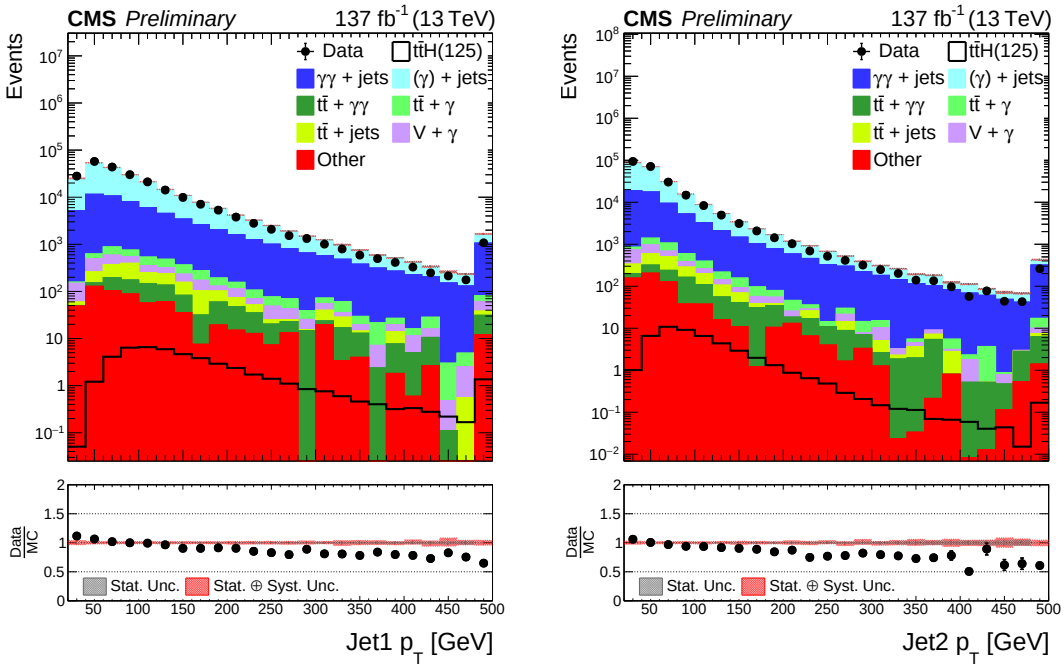

Figure A.1: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ hadronic channel.

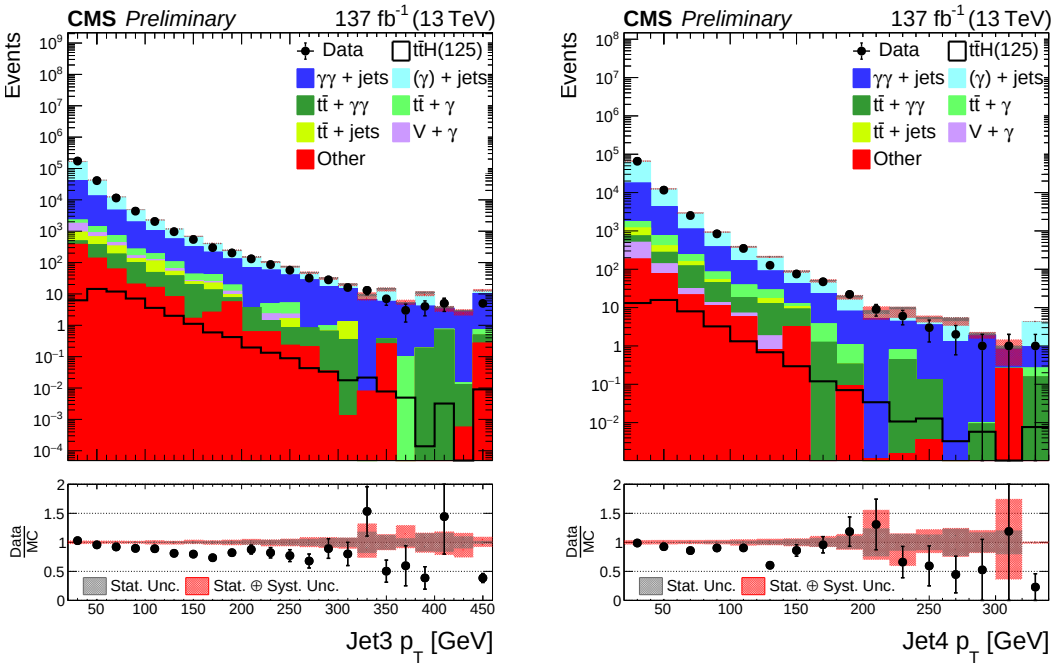

Figure A.2: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ hadronic channel.

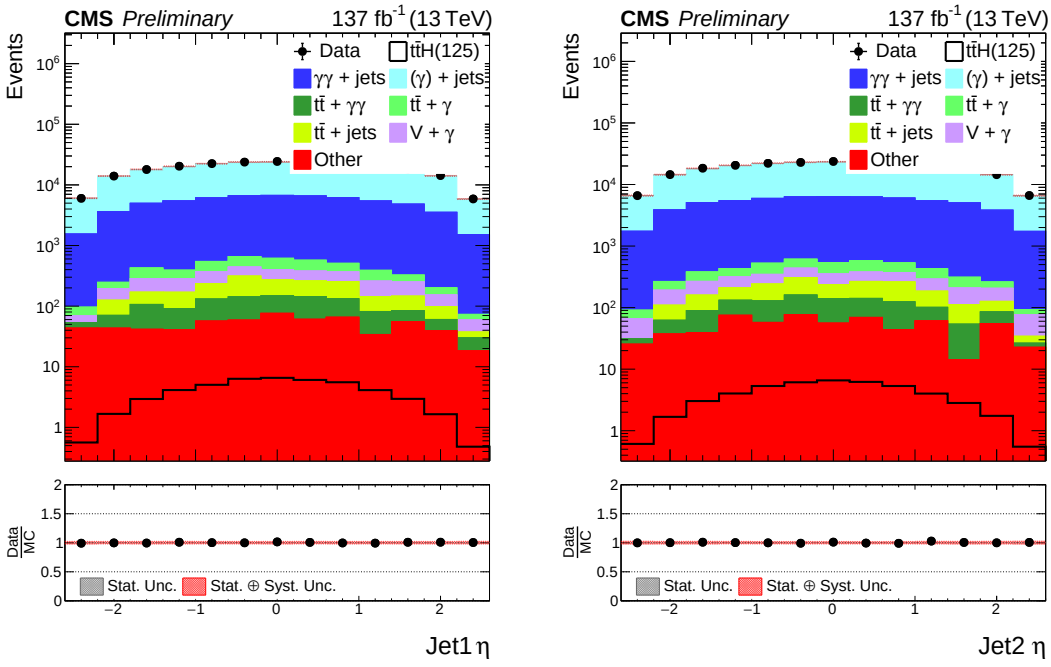

Figure A.3: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ hadronic channel.

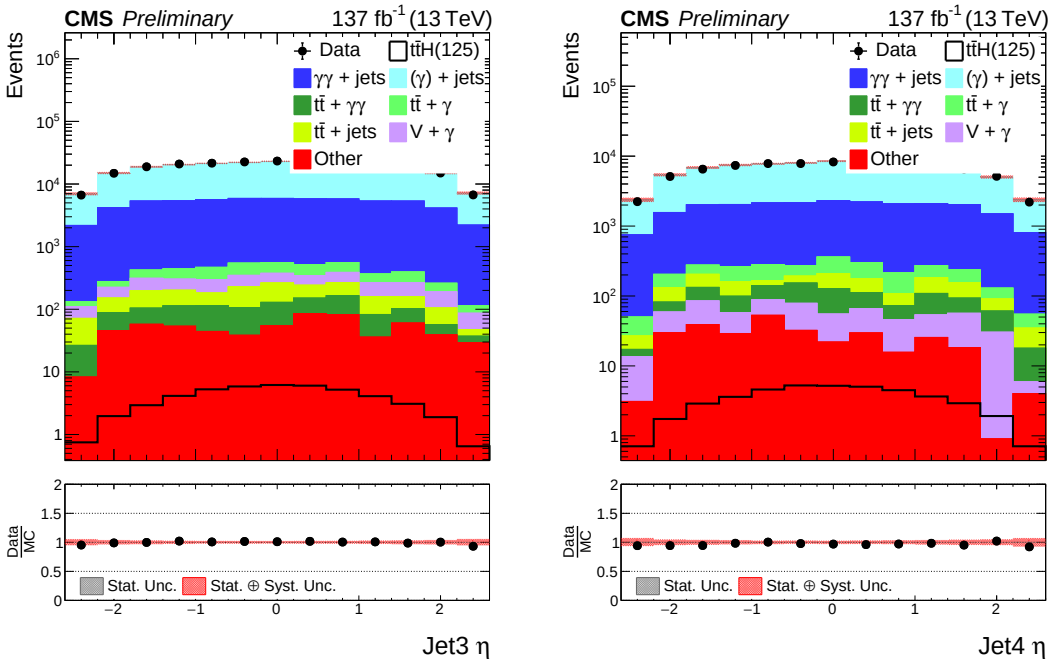

Figure A.4: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ hadronic channel.

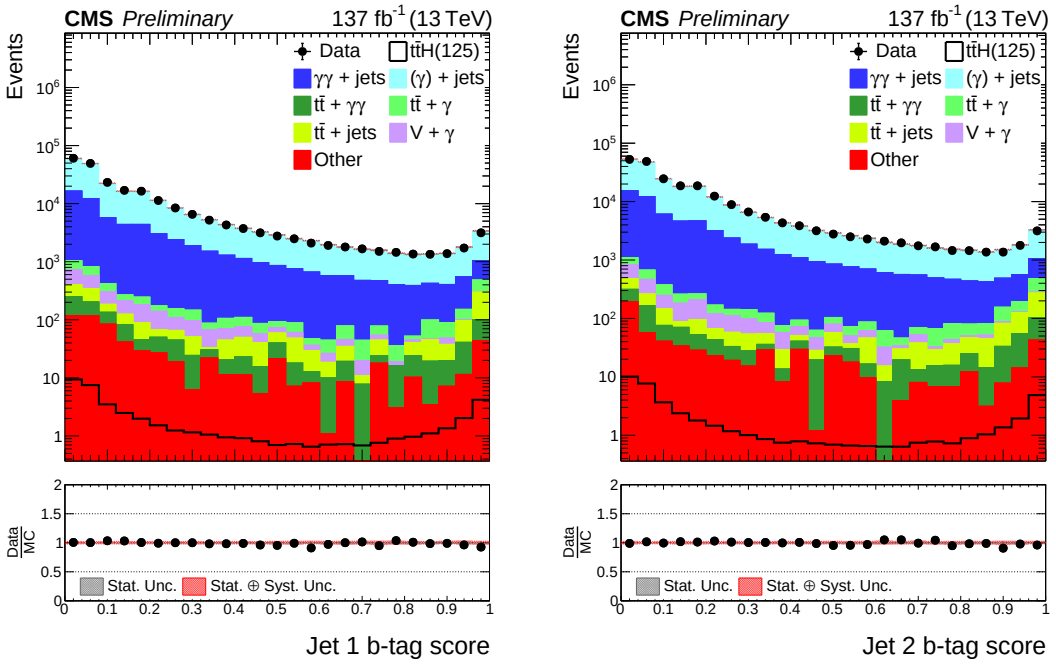

Figure A.5: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ hadronic channel.

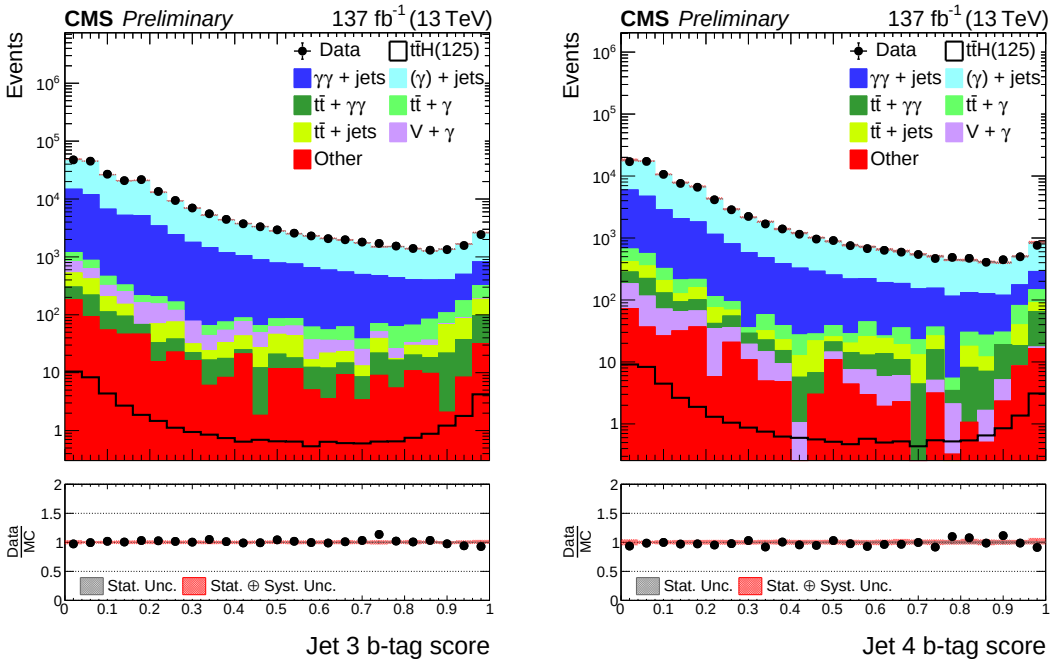

Figure A.6: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ hadronic channel.

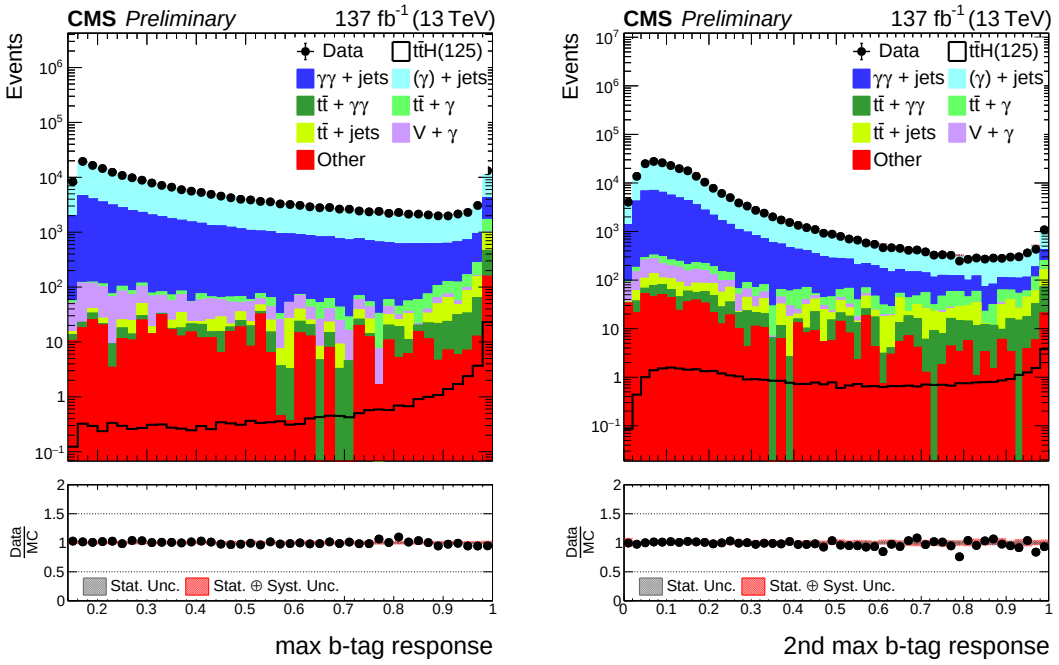

Figure A.7: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ hadronic channel.

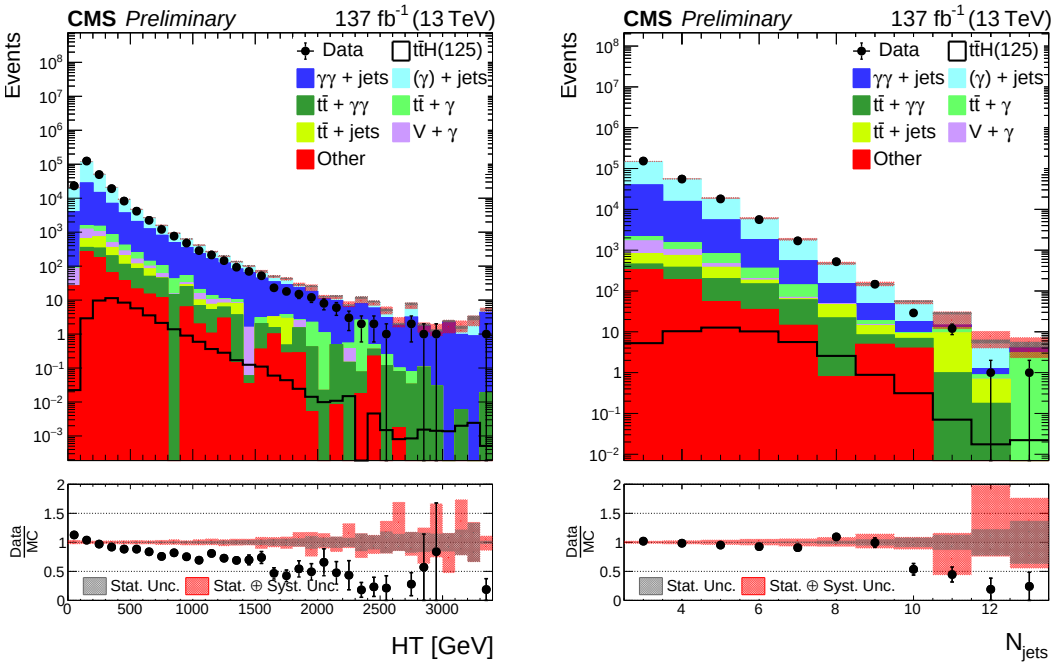

Figure A.8: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ hadronic channel.

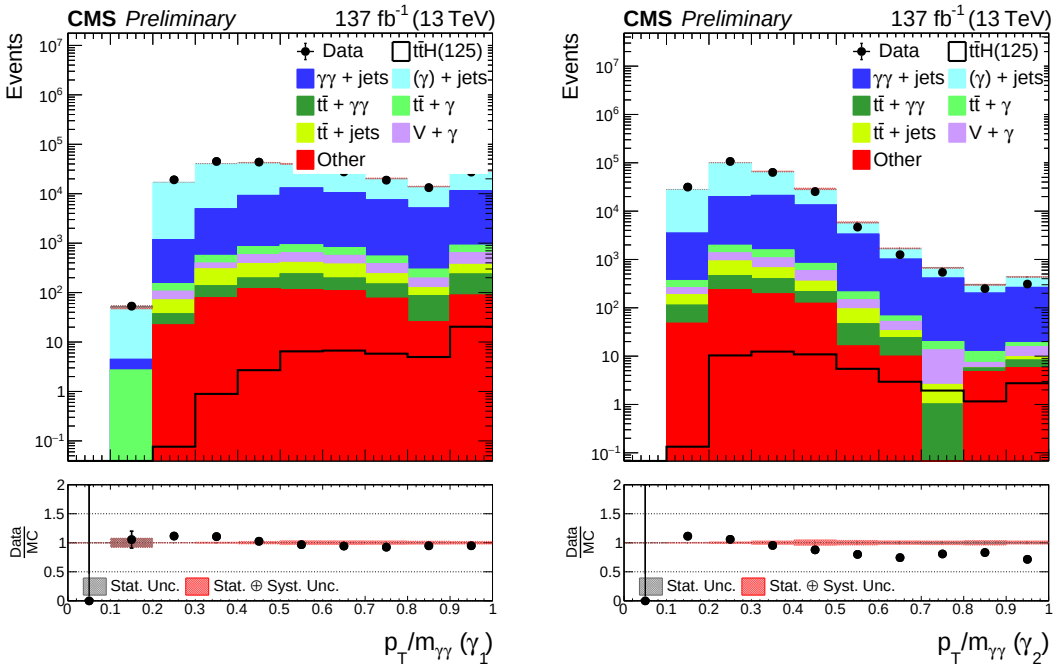

Figure A.9: Agreement between data and simulation for the photon kinematics input features to the BDT-bkg algorithm in the tt̄H hadronic channel.

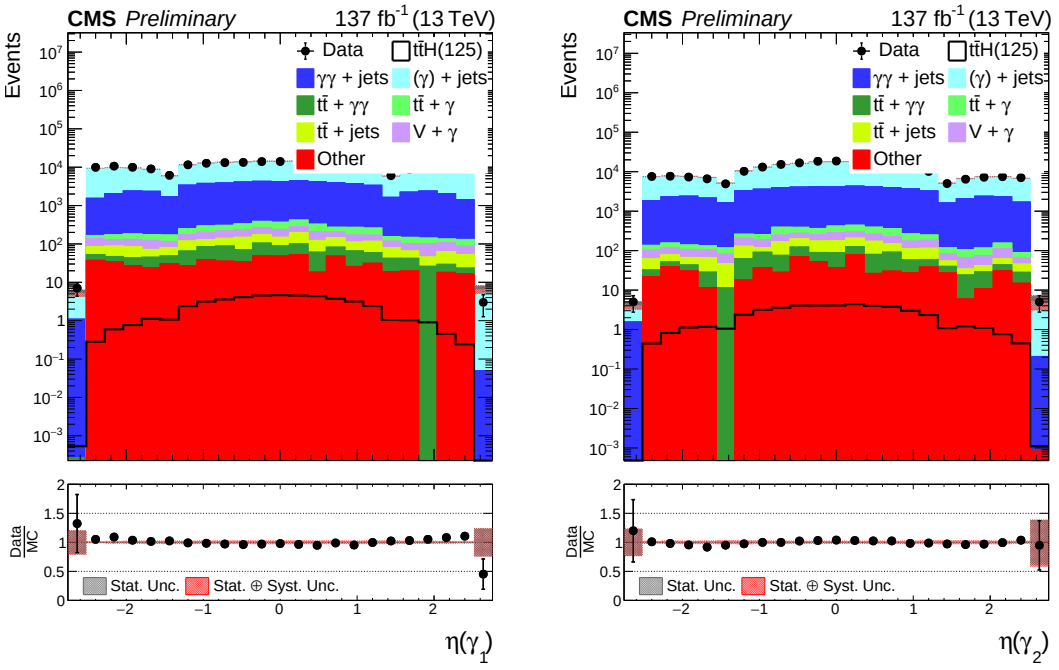

Figure A.10: Agreement between data and simulation for the photon kinematics input features to the BDT-bkg algorithm in the tt̄H hadronic channel.

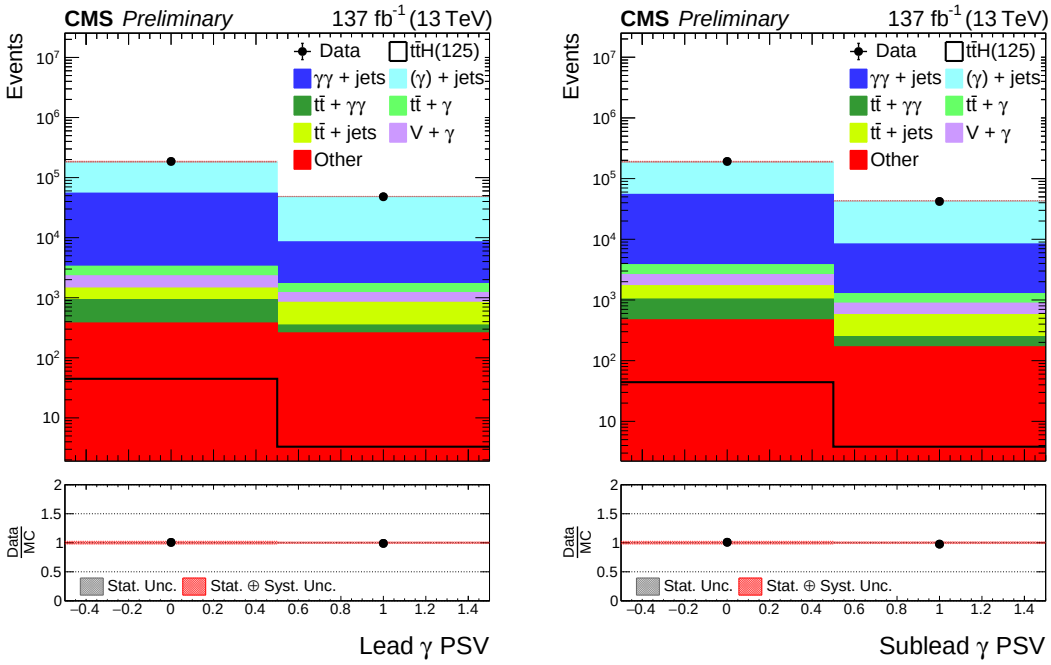

Figure A.11: Agreement between data and simulation for the photon kinematics input features to the BDT-bkg algorithm in the tt̄H hadronic channel.

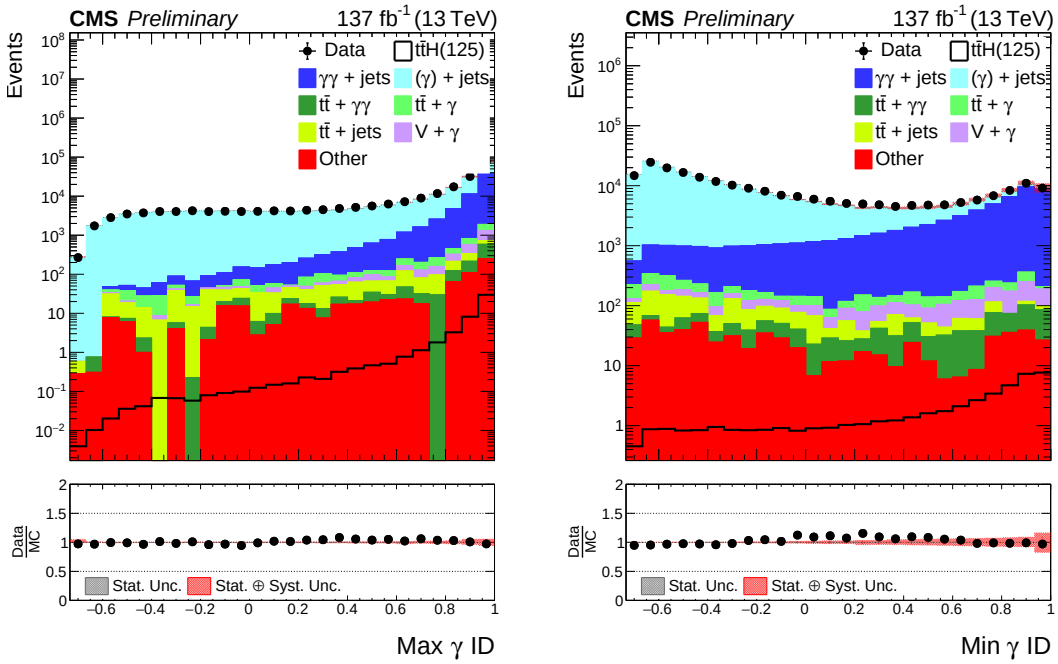

Figure A.12: Agreement between data and simulation for the photon kinematics input features to the BDT-bkg algorithm in the tt̄H hadronic channel.

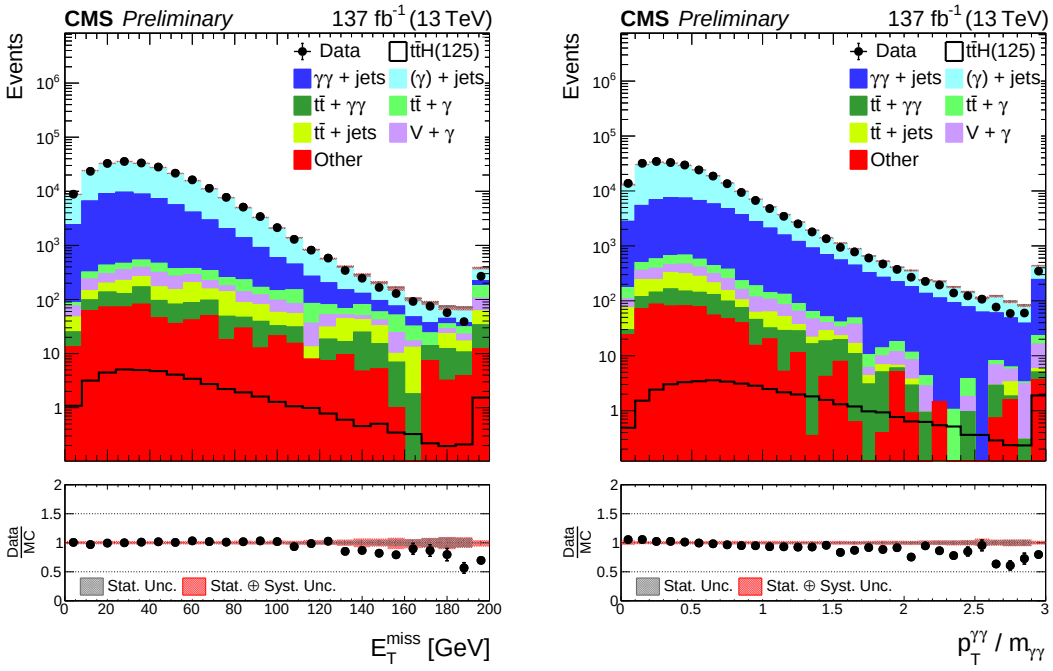

Figure A.13: Agreement between data and simulation for the diphoton kinematics input features to the BDT-bkg algorithm in the tt̄H hadronic channel.

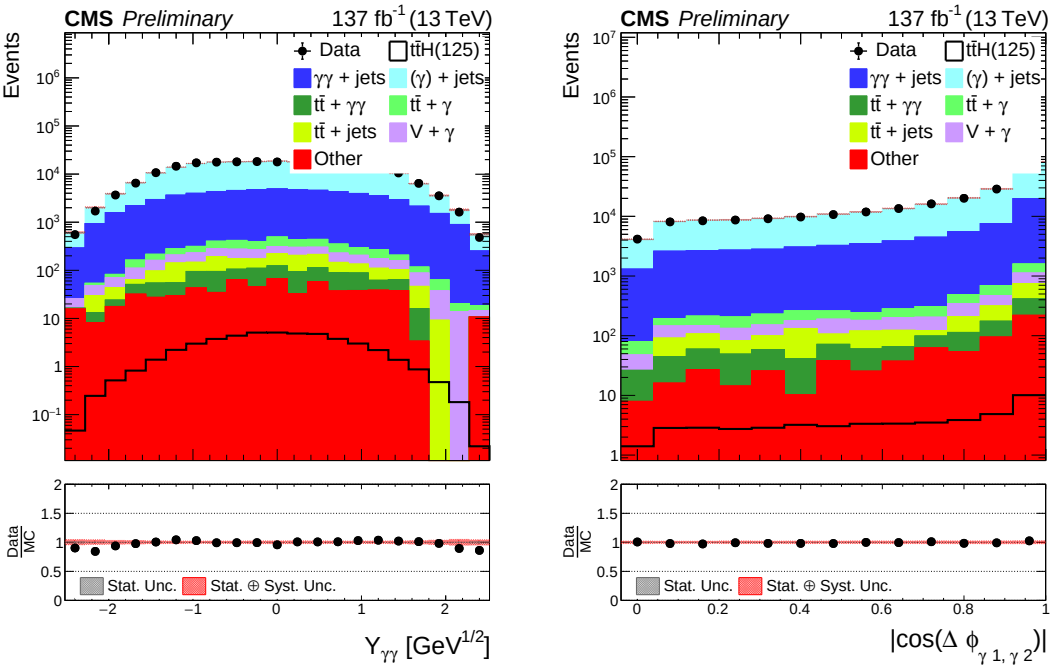

Figure A.14: Agreement between data and simulation for the diphoton kinematics input features to the BDT-bkg algorithm in the tt̄H hadronic channel.

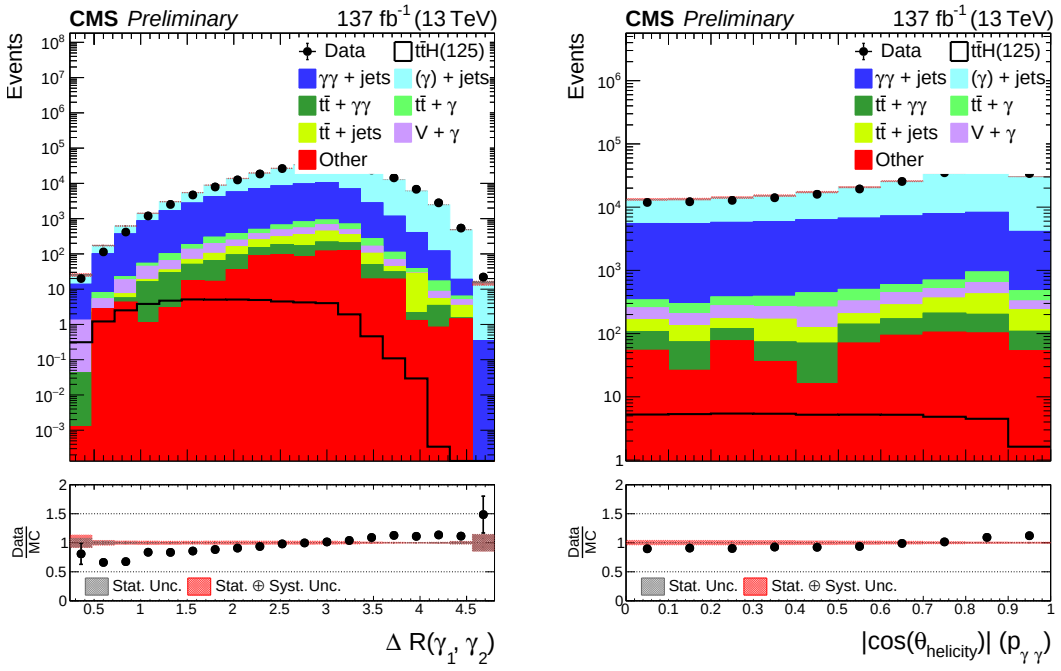

Figure A.15: Agreement between data and simulation for the diphoton kinematics input features to the BDT-bkg algorithm in the tt̄H hadronic channel.

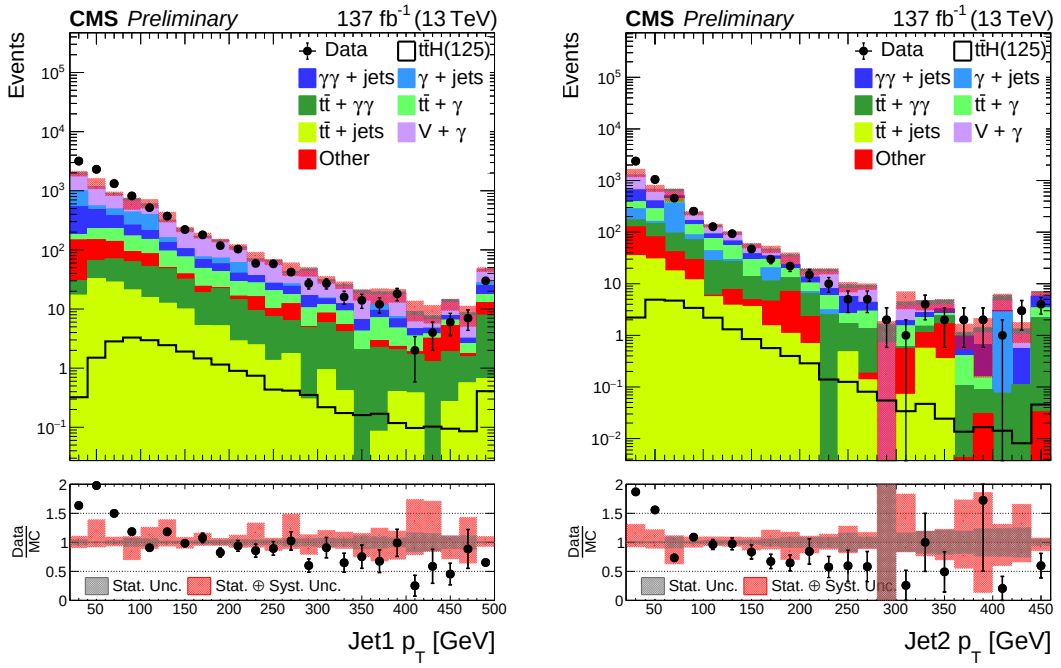

Figure A.16: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

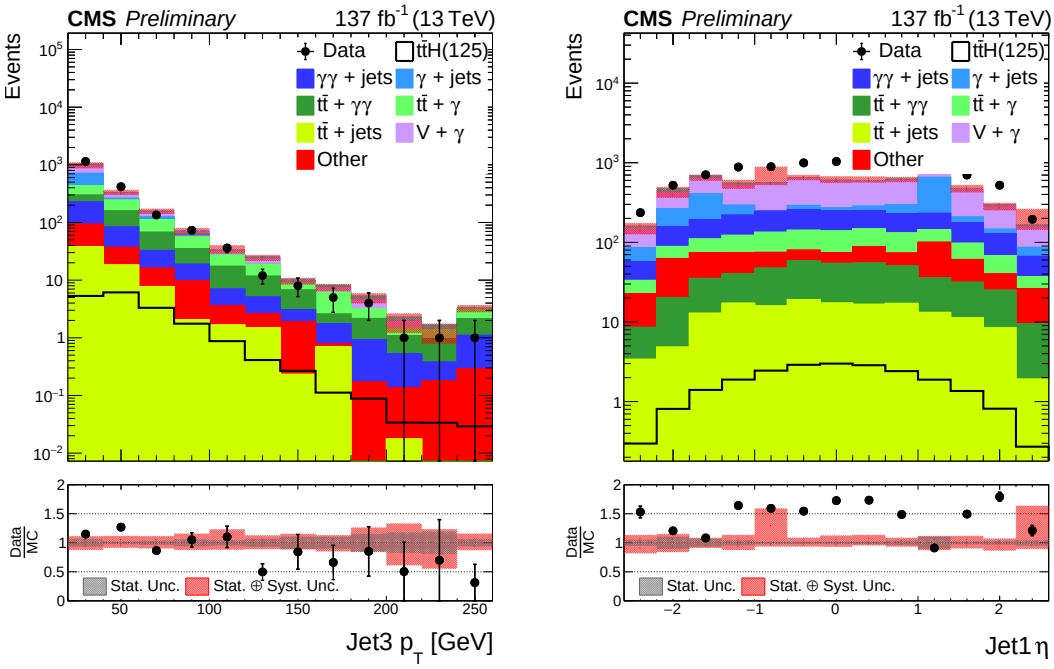

Figure A.17: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ leptonic channel.

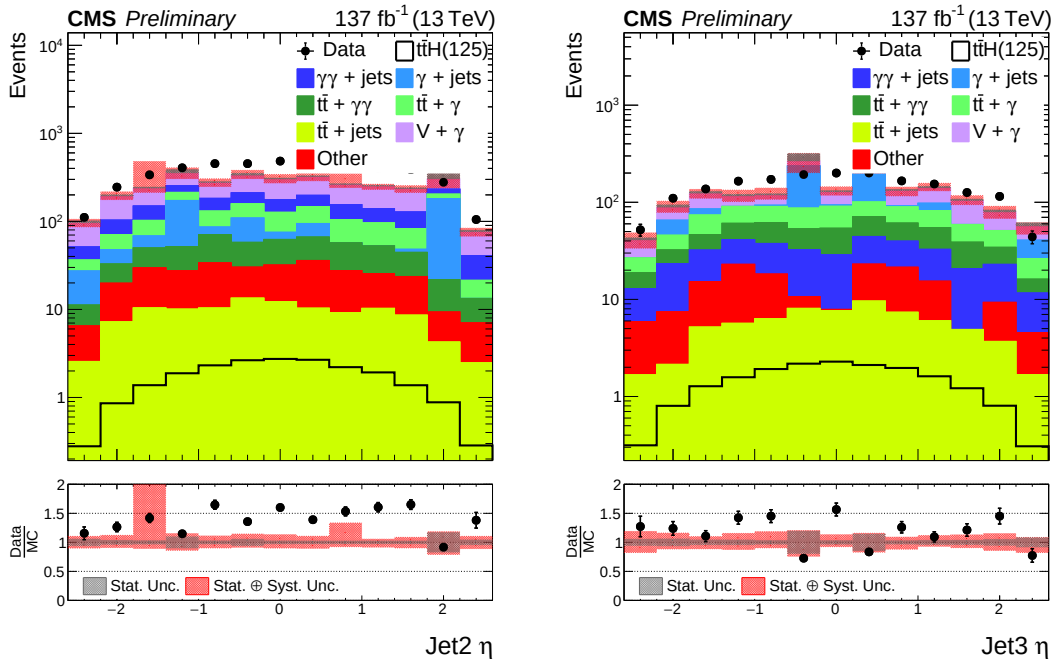

Figure A.18: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ leptonic channel.

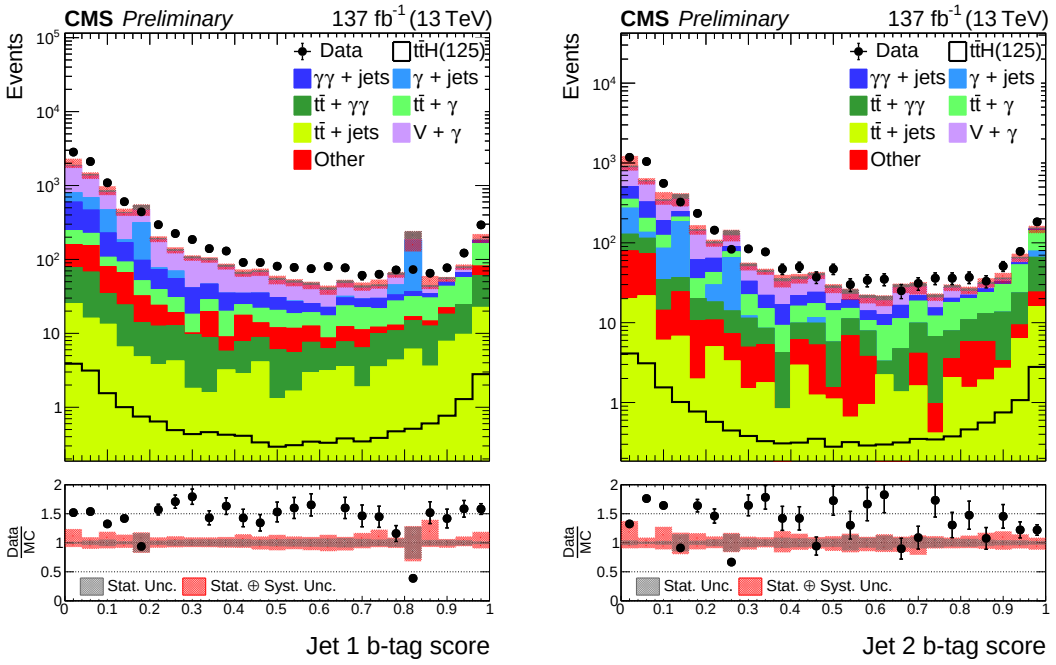

Figure A.19: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ leptonic channel.

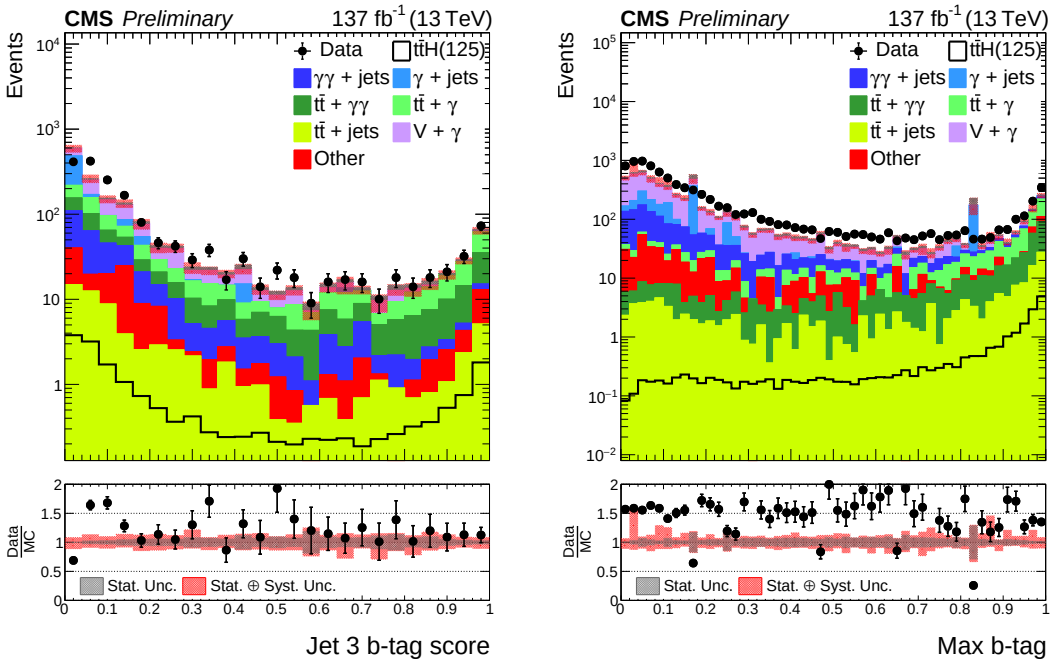

Figure A.20: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the $t\bar{t}H$ leptonic channel.

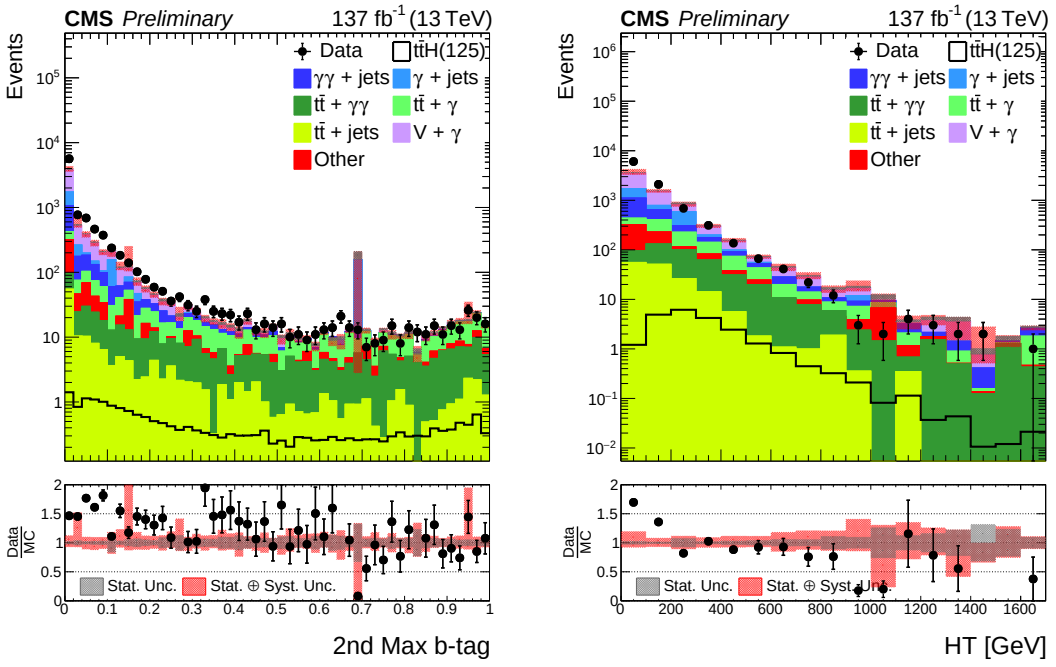

Figure A.21: Agreement between data and simulation for the jet kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

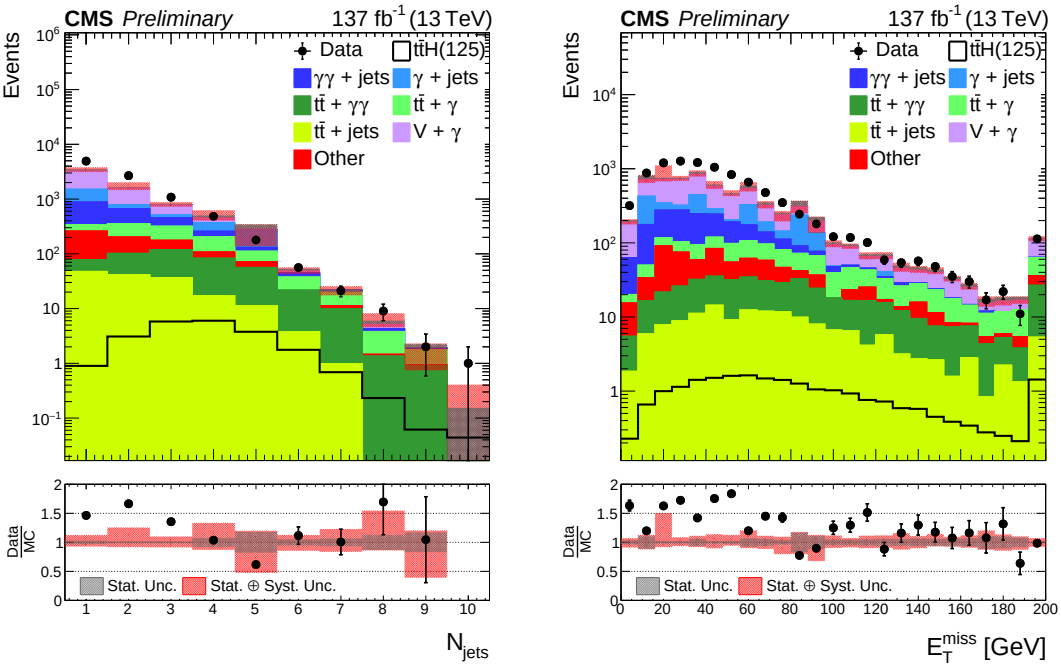

Figure A.22: Agreement between data and simulation for the event-level kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

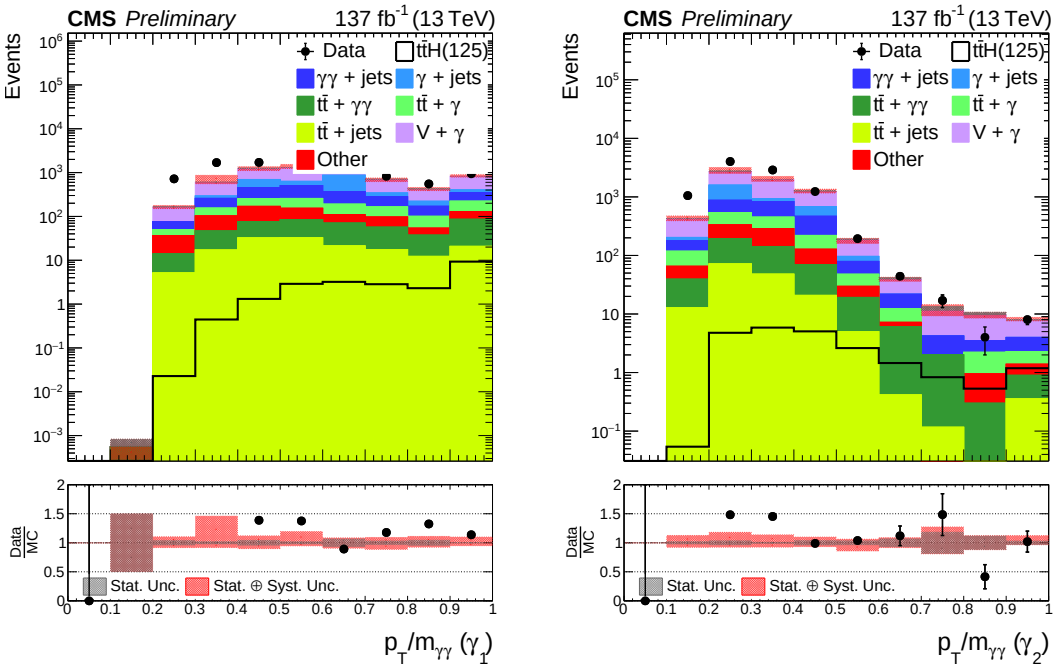

Figure A.23: Agreement between data and simulation for the photon kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

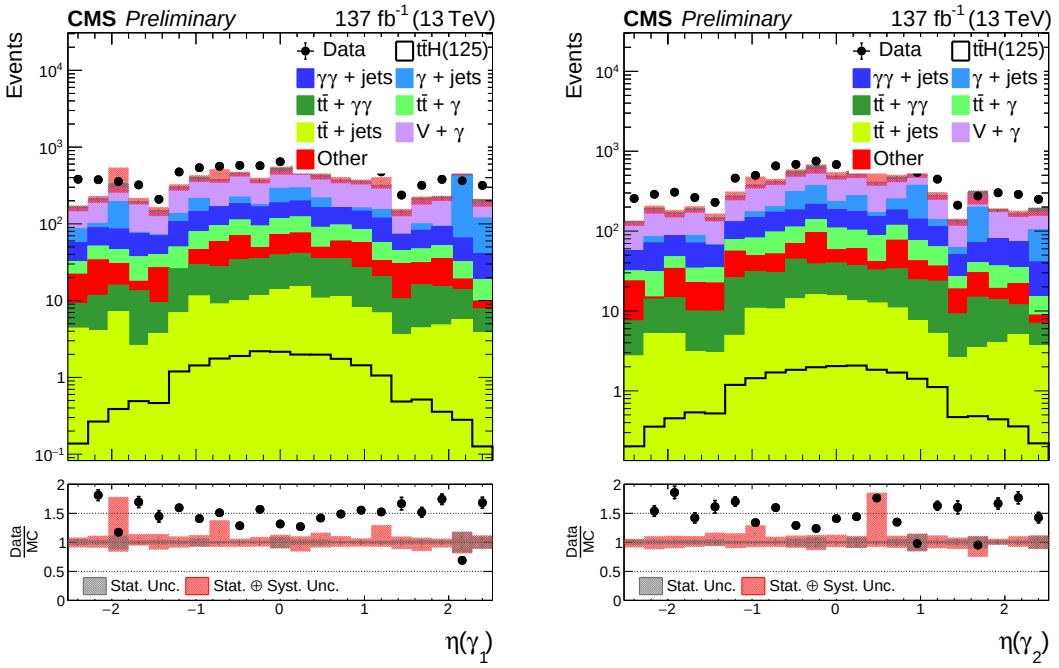

Figure A.24: Agreement between data and simulation for the photon kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

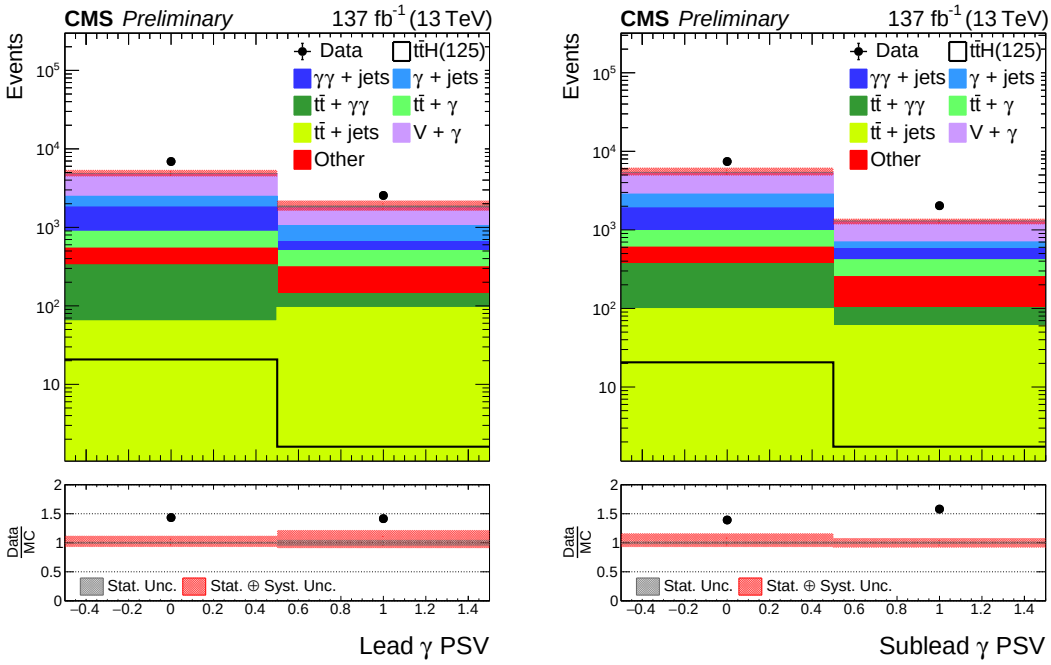

Figure A.25: Agreement between data and simulation for the photon kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

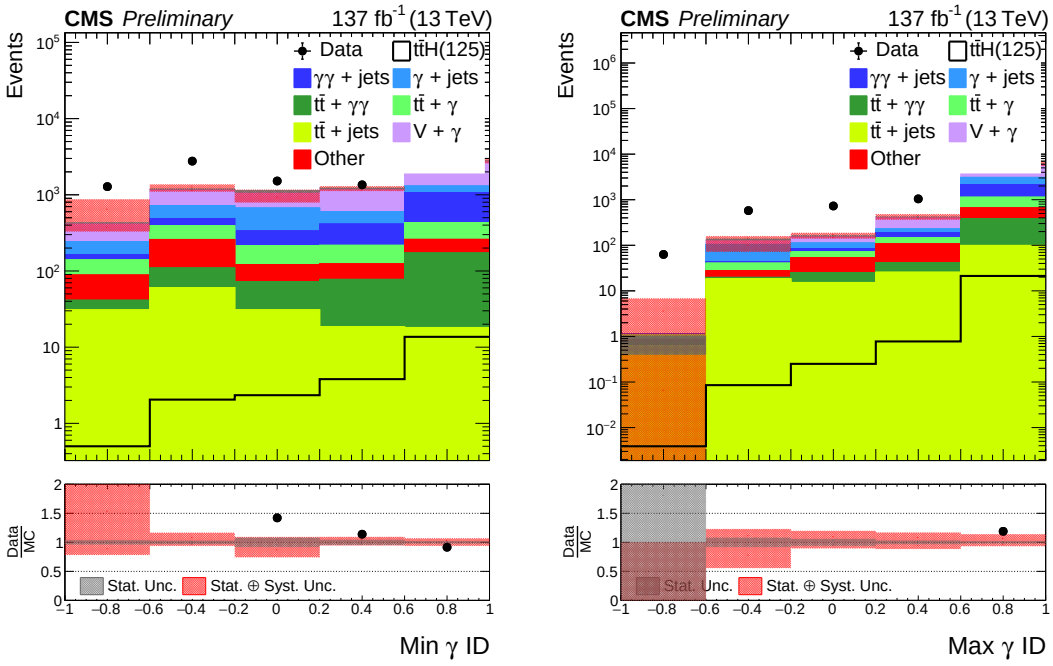

Figure A.26: Agreement between data and simulation for the photon kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

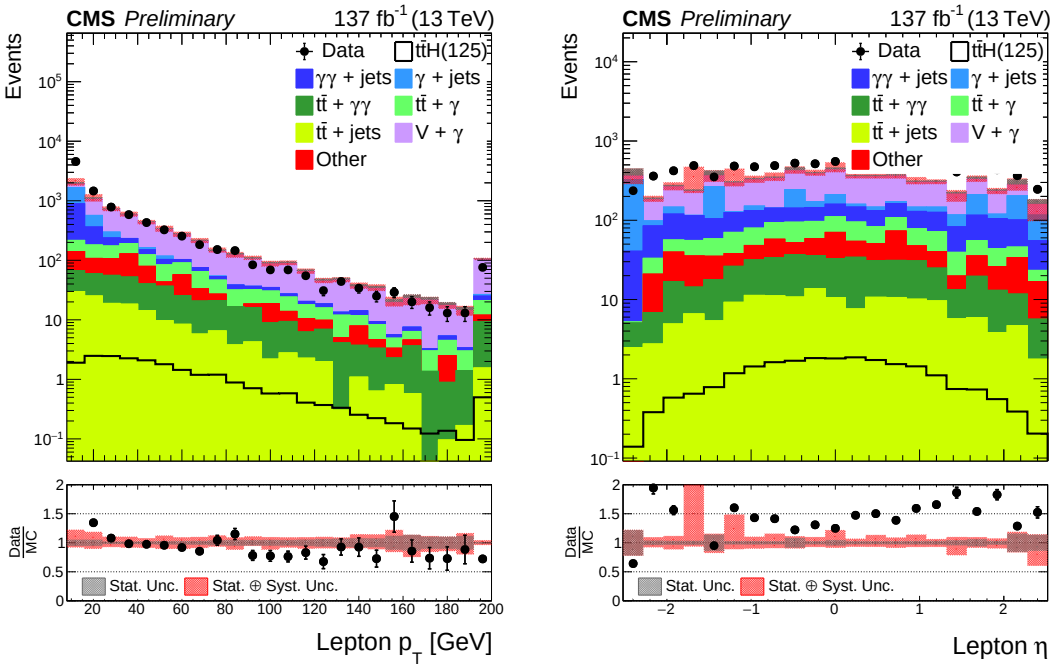

Figure A.27: Agreement between data and simulation for the lepton kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

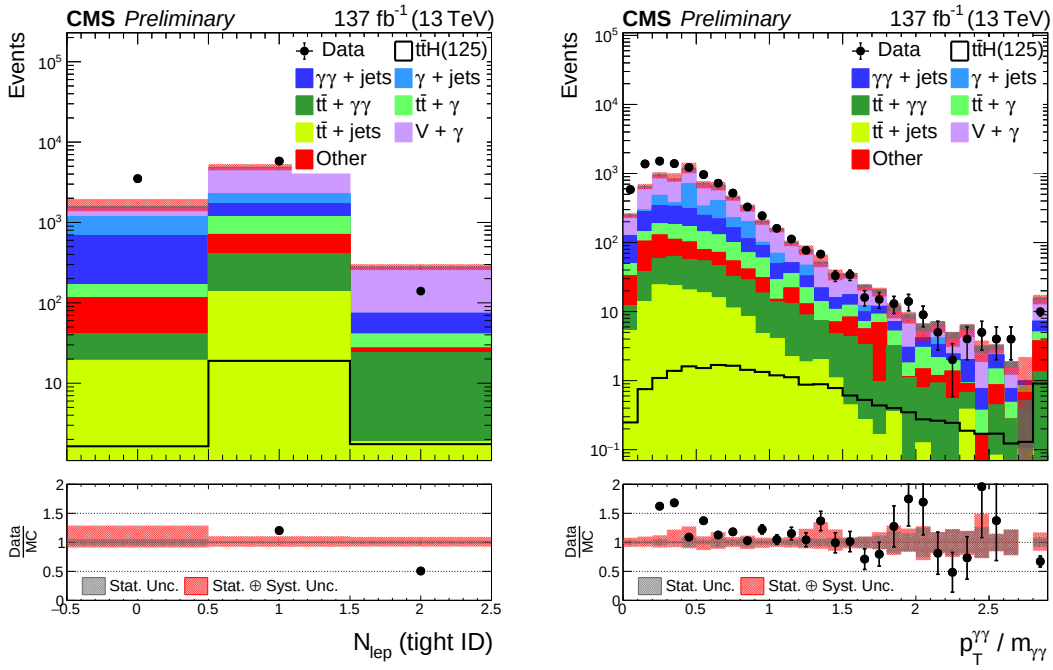

Figure A.28: Agreement between data and simulation for the diphoton kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

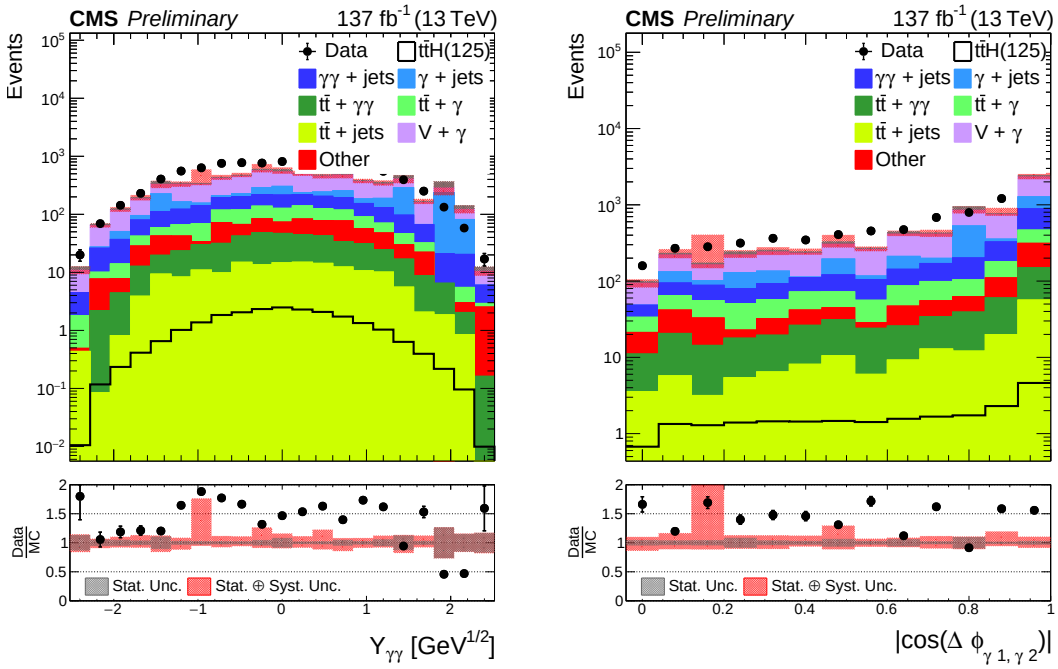

Figure A.29: Agreement between data and simulation for the diphoton kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

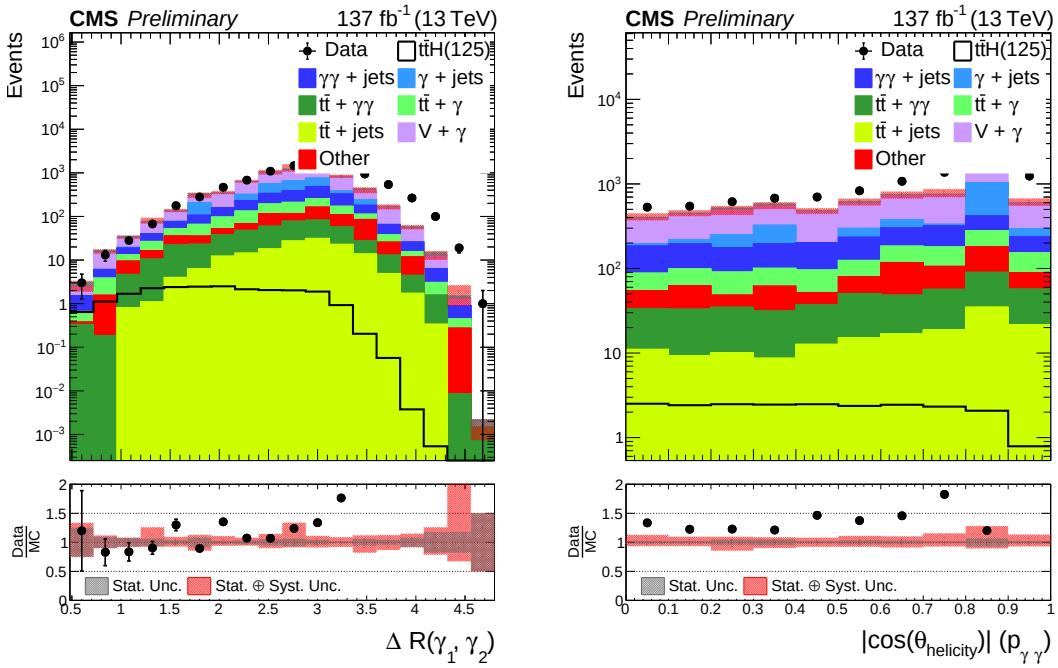

Figure A.30: Agreement between data and simulation for the diphoton kinematics input features to the BDT-bkg algorithm in the tt̄H leptonic channel.

Appendix B

Observed Diphoton Mass Distributions

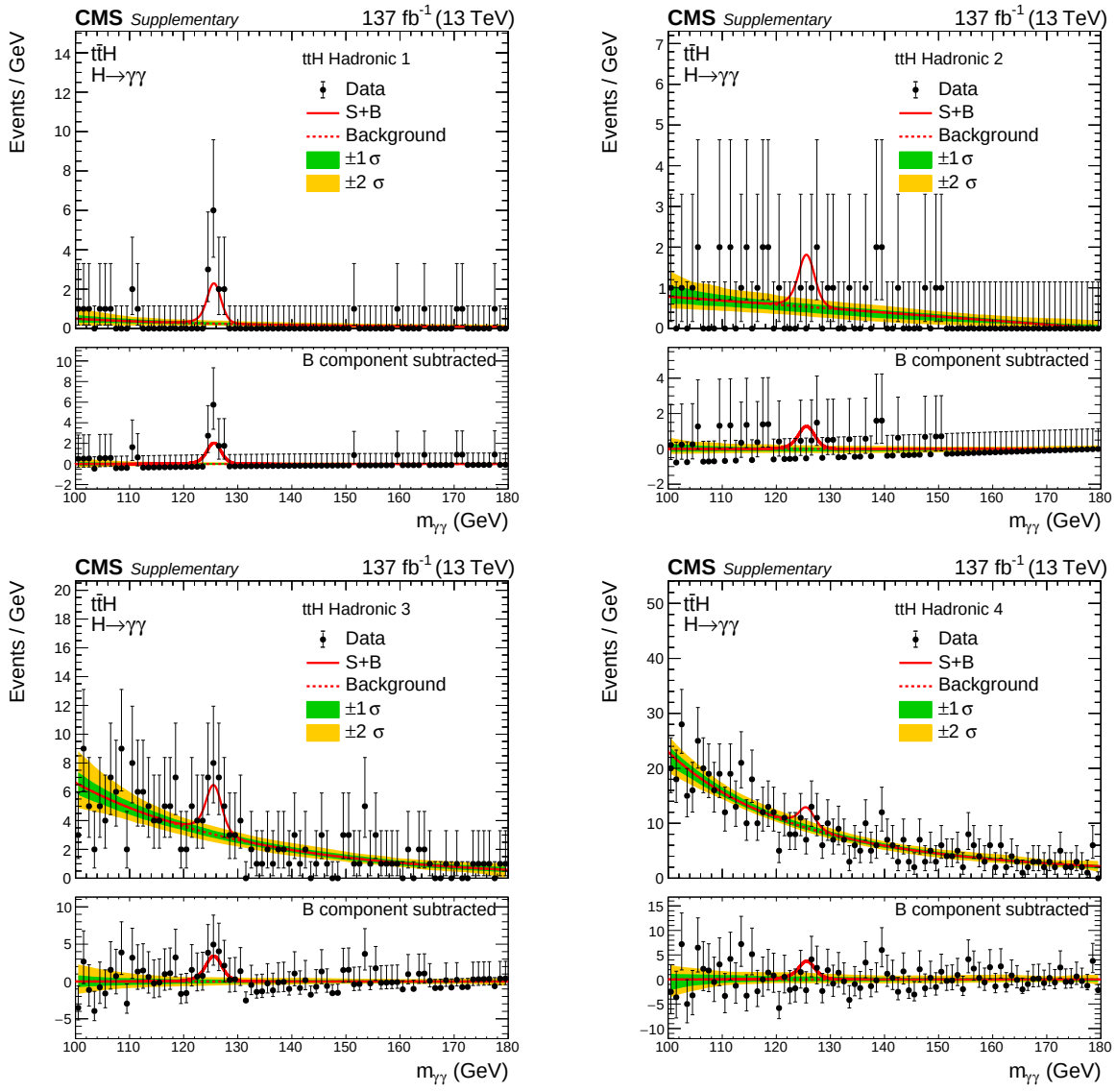

Figure B.1: Observed diphoton mass distributions for the hadronic channel signal regions.

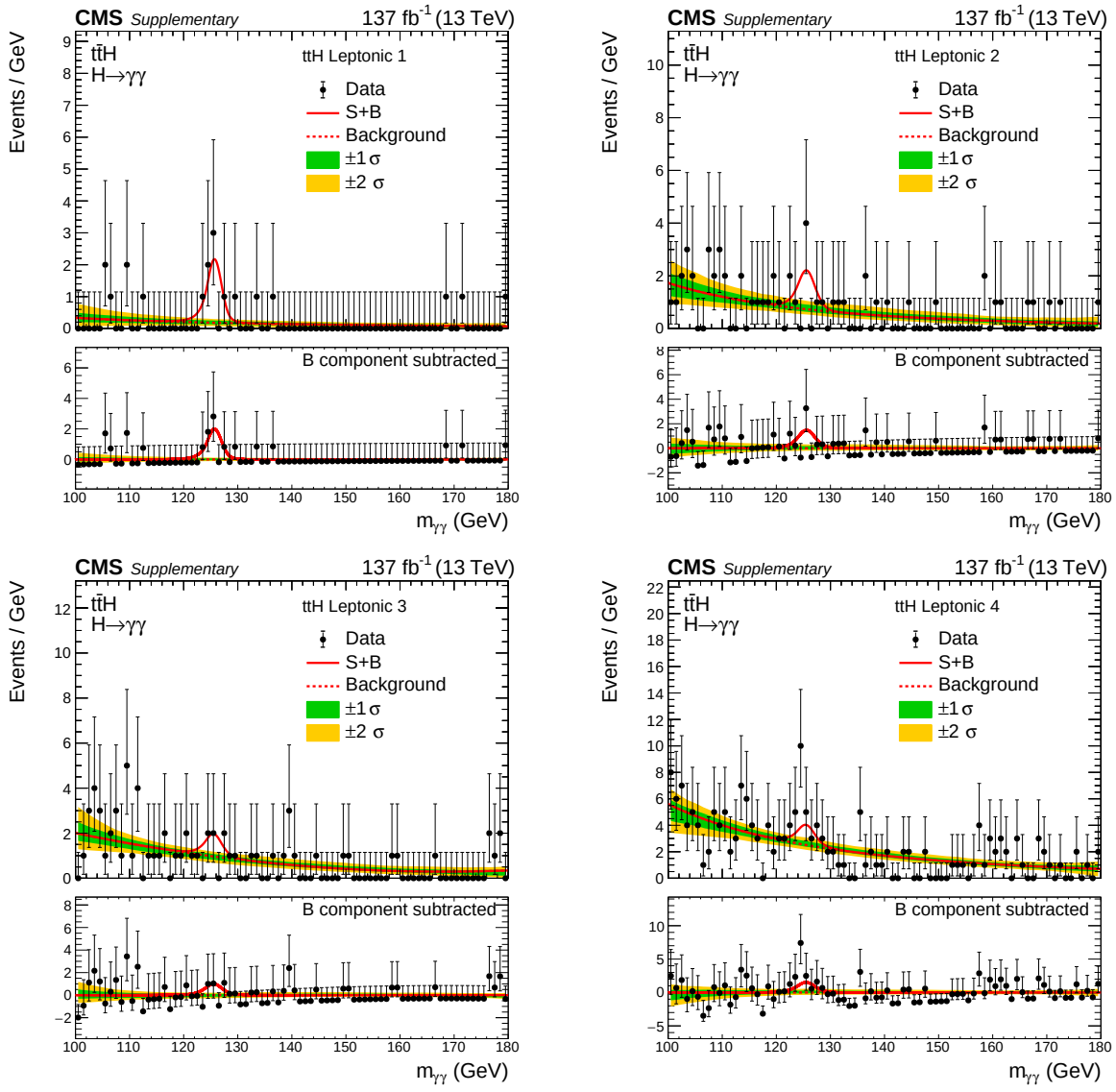

Figure B.2: Observed diphoton mass distributions for the leptonic channel signal regions.

Bibliography

[1] CMS Luminosity Measurements for the 2016 Data Taking Period. Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.

[2] CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-17-004, CERN, Geneva, 2018.

[3] CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-18-002, CERN, Geneva, 2019.

[4] T. Adams et al. Beam test evaluation of electromagnetic calorimeter modules made from proton-damaged $PbWO_4$ crystals. *JINST*, 11(04):P04012, 2016.

[5] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, 126:298–318, 1977.

[6] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[7] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.

[8] ATLAS and CMS Collaborations. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 08:045, 2016.

[9] Florian Beaudette. The cms particle flow algorithm. In Jean-Claude Brient, Roberto Salerno, and Yves Sirois, editors, *Proceedings of the CHEF2013 Conference*. LLR, April 2014.

[10] Ivan Belotelov, Alexander Golunov, Igor Golutvin, N V Gorbunov, Vladimir Karjavin, Yuri Kiryushin, Alexey Kamenev, Serguei Khabarov, Vadim Khabarov, Gleb Meshcheryakov, K P Moissenz, P V Moissenz, Sergey Movchan, Vladimir Palichik, Victor Perelygin, Sergey Shmatov, Dmitry Smolin, and Anatoli Zarubin. Electromagnetic Secondaries and Punchthrough Effects in the CMS ME1/1. Technical Report CMS-NOTE-2006-034, CERN, Geneva, Feb 2006.

[11] Fedor Bezrukov and Mikhail Shaposhnikov. Why should we care about the top quark Yukawa coupling? *J. Exp. Theor. Phys.*, 120:335–343, 2015. [Zh. Eksp. Teor. Fiz.147,389(2015)].

[12] J.D. Bjorken and Emmanuel A. Paschos. Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon. *Phys. Rev.*, 185:1975–1982, 1969.

[13] Sara Bolognesi, Yanyan Gao, Andrei V. Gritsan, Kirill Melnikov, Markus Schulze, Nhan V. Tran, and Andrew Whitbeck. On the spin and parity of a single-produced resonance at the LHC. *Phys. Rev. D*, 86:095031, 2012.

[14] Valeria Botta. Performance and track-based alignment of the Phase-1 upgraded CMS pixel detector. Technical Report CMS-CR-2017-256, CERN, Geneva, Sep 2017.

[15] Martin Breidenbach, Jerome I. Friedman, Henry W. Kendall, Elliott D. Bloom, D.H. Coward, H.C. DeStaebler, J. Drees, Luke W. Mo, and Richard E. Taylor. Observed behavior of highly inelastic electron-proton scattering. *Phys. Rev. Lett.*, 23:935–939, 1969.

[16] G. Breit and E. Wigner. Capture of Slow Neutrons. *Phys. Rev.*, 49:519–531, 1936.

[17] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963.

[18] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_{T} jet clustering algorithm. *JHEP*, 04:063, 2008.

[19] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet user manual. *Eur. Phys. J. C*, 72:1896, 2012.

[20] John M. Campbell, J.W. Huston, and W.J. Stirling. Hard Interactions of Quarks and Gluons: A Primer for LHC Physics. *Rept. Prog. Phys.*, 70:89, 2007.

[21] Stefano Carrazza, Stefano Forte, Zahari Kassabov, Jose Ignacio Latorre, and Juan Rojo. An Unbiased Hessian Representation for Monte Carlo PDFs. *Eur. Phys. J. C*, 75(8):369, 2015.

[22] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.

[23] R.Sekhar Chivukula and Nick J. Evans. Triviality and the precision bound on the Higgs mass. *Phys. Lett. B*, 464:244–248, 1999.

[24] François Chollet. Keras. `https://keras.io`, 2015.

[25] CMS Collaboration. Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at sqrt(s) = 13 TeV. 2017.

[26] ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[27] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[28] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the detector at the LHC. *Phys. Lett. B*, 716:1, 2012.

[29] ATLAS Collaboration. Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC. *Phys. Lett. B*, 726:88–119, 2013. [Erratum: Phys.Lett.B 734, 406–406 (2014)].

[30] ATLAS Collaboration. Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector. *Phys. Lett. B*, 786:59–86, 2018.

[31] ATLAS Collaboration. *CP* Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \to \gamma\gamma$ with the ATLAS Detector. *Phys. Rev. Lett.*, 125(6):061802, 2020.

[32] CMS Collaboration. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. *CERN-LHCC-2006-001, CMS-TDR-8-1*, 2006.

[33] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[34] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012.

[35] CMS Collaboration. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8:P09009, 2013. [JINST8,9009(2013)].

[36] CMS Collaboration. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ = 7 and 8 TeV. *JHEP*, 06:081, 2013.

[37] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(CMS-TRK-11-001. CERN-PH-EP-2014-070. CMS-TRK-11-001):P10009. 80 p, May 2014. Comments: Replaced with published version. Added journal reference and DOI.

[38] CMS Collaboration. Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV. *JINST*, 10(08):P08010, 2015.

[39] CMS Collaboration. Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. *Eur. Phys. J. C*, 75(5):212, 2015.

[40] CMS Collaboration. Jet energy scale and resolution in the cms experiment in pp collisions at 8 tev. *Journal of Instrumentation*, 12(02):P02014âĂŞP02014, Feb 2017.

[41] CMS Collaboration. Particle-flow reconstruction and global event description with the cms detector. *Journal of Instrumentation*, 12(10):P10003, Oct 2017.

[42] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(CMS-PRF-14-001. CMS-PRF-14-001-004. 10):P10003. 82 p, Jun 2017. Replaced with the published version. Added the journal reference and DOI. All the figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/PRF-14-001 (CMS Public Pages).

[43] CMS Collaboration. The CMS trigger system. *JINST*, 12(01):P01020, 2017.

[44] CMS Collaboration. Identification of heavy-flavour jets with the cms detector in pp collisions at 13 tev. *Journal of Instrumentation*, 13(05):P05011âĂŞP05011, May 2018.

[45] CMS Collaboration. Observation of Higgs boson decay to bottom quarks. *Phys. Rev. Lett.*, 121(12):121801, 2018.

[46] CMS Collaboration. Observation of $t\bar{t}H$ production. *Phys. Rev. Lett.*, 120(23):231801, 2018.

[47] CMS Collaboration. A measurement of the Higgs boson mass in the diphoton decay channel. *CMS-PAS-HIG-19-004*, 10 2019.

[48] CMS Collaboration. Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic distributions in the dilepton final state at $\sqrt{s} = 13$ TeV. *CMS-PAS-TOP-19-008*, 5 2020.

[49] CMS Collaboration. Measurements of Higgs boson properties in the diphoton decay channel at $\sqrt{s}$ = 13 TeV. *CMS-PAS-HIG-19-015*, 7 2020.

[50] CMS Collaboration. Measurements of $t\bar{t}H$ Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel. *Phys. Rev. Lett.*, 125(6):061801, 2020.

[51] CMS Collaboration. *CMS Luminosity – Public Results*, 2020 (accessed September 9, 2020).

[52] GEANT4 Collaboration. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[53] LHCb Collaboration. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[54] NNPDF Collaboration. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[55] Particle Data Group Collaboration. Review of Particle Physics. *Phys. Rev. D*, 98(3):030001, 2018.

[56] Particle Data Group Collaboration. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[57] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989.

[58] D Contardo, M Klute, J Mans, L Silvestris, and J Butler. Technical Proposal for the Phase-II Upgrade of the CMS Detector. Technical Report CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02, Geneva, Jun 2015. Upgrade Project Leader Deputies: Lucia Silvestris (INFN-Bari), Jeremy Mans (University of Minnesota) Additional contacts: Lucia.Silvestris@cern.ch, Jeremy.Mans@cern.ch.

[59] G. Cowan. Discovery sensitivity for a counting experiment with background uncertainty. Technical report, Royal Holloway, London, 2012.

[60] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[61] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies. Handling uncertainties in background shapes: the discrete profiling method. 2014.

[62] Alexandre Deur, Stanley J. Brodsky, and Guy F. de Teramond. The QCD Running Coupling. *Nucl. Phys.*, 90:1, 2016.

[63] Stefano Di Vita, Christophe Grojean, Giuliano Panico, Marc Riembau, and Thibaud Vantalon. A global view on the Higgs self-coupling. *JHEP*, 09:069, 2017.

[64] Michael Dine and Alexander Kusenko. The Origin of the matter - antimatter asymmetry. *Rev. Mod. Phys.*, 76:1, 2003.

[65] Paul A.M. Dirac. The quantum theory of the electron. *Proc. Roy. Soc. Lond. A*, 117:610–624, 1928.

[66] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.

[67] H.C. DeStaebler(SLAC) J. Drees(SLAC) Guthrie Miller(SLAC) Luke W. Mo(SLAC) Richard E. Taylor(SLAC) Martin Breidenbach(MIT LNS) Jerome I. Friedman(MIT LNS) George C. Hartmann(MIT LNS) Henry W. Kendall(MIT LNS) Elliott D. Bloom(SLAC), D.H. Coward(SLAC). High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees. *Phys. Rev. Lett.*, 23:930–934, 1969.

[68] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.

[69] Lyndon Evans and Philip Bryant. LHC machine. *Journal of Instrumentation*, 3(08):S08001–S08001, aug 2008.

[70] A. Fetter and J. Walecka. *Theoretical Mechanics of Particles and Continua*. 2003.

[71] Richard P. Feynman. Very high-energy collisions of hadrons. *Phys. Rev. Lett.*, 23:1415–1417, 1969.

[72] R. Fisher. On the interpretation of χ^2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society*, 85(1):87–94, 1922.

[73] Yanyan Gao, Andrei V. Gritsan, Zijin Guo, Kirill Melnikov, Markus Schulze, and Nhan V. Tran. Spin Determination of Single-Produced Resonances at Hadron Colliders. *Phys. Rev. D*, 81:075022, 2010.

[74] Murray Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8:214–215, 1964.

[75] F.X. Gentit. Litrani: A General purpose Monte Carlo program simulating light propagation in isotropic or anisotropic media. *Nucl. Instrum. Meth. A*, 486:35–39, 2002.

[76] Sheldon L. Glashow. The renormalizability of vector meson interactions. *Nucl. Phys.*, 10:107–117, 1959.

[77] M. Gockeler, R. Horsley, V. Linke, Paul E.L. Rakow, G. Schierholz, and H. Stuben. Is there a Landau pole problem in QED? *Phys. Rev. Lett.*, 80:4119–4122, 1998.

[78] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. Broken Symmetries. *Phys. Rev.*, 127:965–970, 1962.

[79] W. Gordon. Der Comptoneffekt nach der Schrodingerschen Theorie. *Z. Phys.*, 40:117–133, 1926.

[80] O.W. Greenberg. Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons. *Phys. Rev. Lett.*, 13:598–602, 1964.

[81] V.N. Gribov and L.N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.

[82] David Griffiths. *Introduction to Quantum Mechanics*. 2005.

[83] David Griffiths. *Introduction to elementary particles*. 2008.

[84] Andrei V. Gritsan, Raoul Röntsch, Markus Schulze, and Meng Xiao. Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques. *Phys. Rev. D*, 94(5):055023, 2016.

[85] LHC Higgs Cross Section Working Group. Handbook of lhc higgs cross sections: 3. higgs properties. 2013.

[86] LHC Higgs Cross Section Working Group. Handbook of lhc higgs cross sections: 4. deciphering the nature of the higgs sector, 2016.

[87] Daniel Guest, Julian Collado, Pierre Baldi, Shih-Chieh Hsu, Gregor Urban, and Daniel Whiteson. Jet flavor classification in high-energy physics with deep neural networks. *Physical Review D*, 94(11), Dec 2016.

[88] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, 1964.

[89] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.

[90] Stefan Höche. Introduction to parton-shower event generators. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, pages 235–295, 2015.

[91] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[92] Fabrizio Caola Yanyan Gao Andrei V. Gritsan Christopher B. Martin Kirill Melnikov Markus Schulze Nhan V. Tran Andrew Whitbeck Yaofu Zhou Ian Anderson, Sara Bolognesi. Constraining Anomalous HVV Interactions at Proton and Lepton Colliders. *Phys. Rev. D*, 89(3):035007, 2014.

[93] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.

[94] John David Jackson. *Classical electrodynamics; 2nd ed.* Wiley, New York, NY, 1975.

[95] Amanda Cooper-Sarkar Albert De Roeck Joel Feltesse Stefano Forte Jun Gao Sasha Glazov Joey Huston Zahari Kassabov Ronan McNulty Andreas Morsch Pavel Nadolsky Voica Radescu Juan Rojo Robert Thorne Jon Butterworth, Stefano Carrazza. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016.

[96] Shinya Kanemura, Shingo Kiyoura, Yasuhiro Okada, Eibun Senaha, and C.P. Yuan. New physics effect on the Higgs self-coupling. *Phys. Lett. B*, 558:157–164, 2003.

[97] Shinya Kanemura, Yasuhiro Okada, Eibun Senaha, and C.-P. Yuan. Higgs coupling constants as a probe of new physics. *Phys. Rev. D*, 70:115002, 2004.

[98] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

[99] O. Klein. Quantentheorie und funfdimensionale Relativitatstheorie. *Z. Phys.*, 37:895–906, 1926.

[100] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[101] Yann LeCun, Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop*, pages 9–50, London, UK, 1998. Springer-Verlag.

[102] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63:189–285, 2009.

[103] Stephen P. Martin. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.*, 21:1–153, 2010.

[104] Paul Barham-Eugene Brevdo Zhifeng Chen Craig Citro Greg S. Corrado Andy Davis Jeffrey Dean Matthieu Devin Sanjay Ghemawat Ian Goodfellow Andrew Harp Geoffrey Irving Michael Isard Yangqing Jia Rafal Jozefowicz Lukasz Kaiser Manjunath Kudlur Josh Levenberg Dan Mane Rajat Monga Sherry Moore Derek Murray Chris Olah Mike Schuster Jonathon Shlens Benoit Steiner Ilya Sutskever Kunal Talwar Paul Tucker Vincent Vanhoucke Vijay Vasudevan Fernanda Viegas Oriol Vinyals Pete Warden Martin Wattenberg Martin Wicke Yuan Yu Xiaoqiang Zheng Martin Abadi, Ashish Agarwal. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[105] Mark Joseph Oreglia. *A study of the reactions* $\psi' \to \gamma\gamma\psi$. PhD thesis, Stanford University, 1980. SLAC Report SLAC-R-236.

[106] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[107] J. Prentki and J. Steinberger, editors. *Proceedings, 14th International Conference on High-Energy Physics (ICHEP 68): Vienna, Austria, 28 Aug-5 Sep, 1968*, Geneva, 1968. CERN.

[108] Snowmass 2013 Top quark working group. Snowmass 2013 top quark working group report, 2013.

[109] V.C. Rubin, N. Thonnard, and Jr. Ford, W.K. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. *Astrophys. J.*, 238:471, 1980.

[110] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968.

[111] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3 2014.

[112] Samuel L. Smith, Pieter-Jan Kindermans, Chris Ying, and Quoc V. Le. Don't decay the learning rate, increase the batch size, 2017.

[113] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[114] V.V. Sudakov. Vertex parts at very high-energies in quantum electrodynamics. *Sov. Phys. JETP*, 3:65–71, 1956.

[115] Gerard 't Hooft. Renormalizable Lagrangians for Massive Yang-Mills Fields. *Nucl. Phys. B*, 35:167–188, 1971.

[116] Thomas Taylor and Daniel Treille. *The Large Electron Positron Collider (LEP): Probing the Standard Model*, volume 27, pages 217–261. 2017.

[117] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[118] Energy Frontier Higgs Boson working group. Working Group Report: Higgs Boson. In *Community Summer Study 2013: Snowmass on the Mississippi*, 10 2013.

[119] Eleftherios Spyromitros Xioufis, William Groves, Grigorios Tsoumakas, and Ioannis P. Vlahavas. Multi-label classification methods for multi-target regression. *CoRR*, abs/1211.6581, 2012.

[120] Tung-Mow Yan and Sidney D. Drell. *The Parton Model and Its Applications*, pages 227–243. 2015.

[121] A. Zee. *Quantum field theory in a nutshell*. 11 2003.

[122] G. Zweig. *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, pages 22–101. 2 1964.