Differential cross section measurement of the Higgs boson decay to a pair of τ leptons in proton-proton collisions at $\sqrt{s} = 13.6$ TeV at the Large Hadron Collider

by

Braden Allmond

B.S., Fort Hays State University, 2019

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Abstract

The Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC) has recently concluded its third run of data taking. It has produced the largest set of particle physics data to date, already being used in combination with previous data to push the boundaries of the standard model (SM). Since its discovery in 2012, the properties of the Higgs boson have been continually measured, refined, and scrutinized for deviations from the SM. This thesis presents blinded results of a differential cross section measurement of the Higgs boson decay to two tau leptons using proton-proton collision data collected by the CMS experiment at the CERN LHC at a center-of-mass energy of $\sqrt{s} = 13.6$ TeV totaling 172 fb^{-1}. The differential variables include the Higgs boson transverse momentum, number of jets, leading jet transverse momentum, and two new variables sensitive to the vector boson fusion production mode of the Higgs boson, namely dijet invariant mass and the dijet angular separation. These blinded measurements show an expected improvement on previous results by using new data, updated event triggers, and more refined background modeling techniques.

In addition, this thesis presents a summary of the SM and descriptions of the LHC and CMS suitable for a general audience. Work completed in service of the CMS collaboration is also described, including measurements of hadronic tau triggers and the thermal validation of prototype components for the MIP Timing Detector, a new subdetector to be installed for the CMS upgrade.

Differential cross section measurement of the Higgs boson decay to a pair of τ leptons in proton-proton collisions at $\sqrt{s} = 13.6$ TeV at the Large Hadron Collider

by

Braden Allmond

B.S., Fort Hays State University, 2019

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Approved by:

Major Professor
Ketino Kaadze

Copyright

© Braden Allmond 2026.

Abstract

The Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC) has recently concluded its third run of data taking. It has produced the largest set of particle physics data to date, already being used in combination with previous data to push the boundaries of the standard model (SM). Since its discovery in 2012, the properties of the Higgs boson have been continually measured, refined, and scrutinized for deviations from the SM. This thesis presents blinded results of a differential cross section measurement of the Higgs boson decay to two tau leptons using proton-proton collision data collected by the CMS experiment at the CERN LHC at a center-of-mass energy of $\sqrt{s} = 13.6$ TeV totaling 172 fb^{-1}. The differential variables include the Higgs boson transverse momentum, number of jets, leading jet transverse momentum, and two new variables sensitive to the vector boson fusion production mode of the Higgs boson, namely dijet invariant mass and the dijet angular separation. These blinded measurements show an expected improvement on previous results by using new data, updated event triggers, and more refined background modeling techniques.

In addition, this thesis presents a summary of the SM and descriptions of the LHC and CMS suitable for a general audience. Work completed in service of the CMS collaboration is also described, including measurements of hadronic tau triggers and the thermal validation of prototype components for the MIP Timing Detector, a new subdetector to be installed for the CMS upgrade.

Table of Contents

List of Figures . . . viii

List of Tables . . . x

Acknowledgements . . . xi

Preface . . . xiv

1 Introduction . . . 1

2 The Standard Model . . . 5

2.1 Leptons . . . 7

2.2 Quarks . . . 10

2.3 Force-carrying Bosons . . . 11

2.4 The Higgs Boson . . . 12

2.5 Permitted Interactions . . . 13

3 The Large Hadron Collider . . . 16

4 The Compact Muon Solenoid Detector . . . 22

4.1 Overview . . . 22

4.2 Coordinate System . . . 24

4.3 Subdetectors and Physics Objects . . . 24

4.4 Particle Flow and Hadronic Tau Reconstruction . . . 29

4.5 Trigger System . . . 32

4.6 Level 1 Trigger . . . 33

4.7 High Level Trigger . . . 34
4.8 Event Processing, Software, and Simulations . . . 35

5 The Higgs to Tau Tau Differential Run 3 Analysis . . . 38
5.1 Motivating Higgs Boson Studies . . . 38
5.2 Higgs Boson Status and Measurement Approaches . . . 39
5.3 Higgs to Tau Tau . . . 44
5.3.1 Datasets and Simulation . . . 47
5.3.2 Trigger Strategy . . . 48
5.3.3 Event Selection . . . 49
5.3.4 Background Estimation . . . 53
5.3.5 Backgrounds from MC . . . 54
5.3.6 Data-Driven Estimation . . . 54
5.3.7 Categorization . . . 67
5.3.8 Differential variables . . . 68
5.3.9 Unfolding . . . 69
5.3.10 Fiducial Region . . . 69
5.3.11 Corrections . . . 72
5.3.12 Statistical Methods . . . 74
5.3.13 Results . . . 75
5.3.14 Summary . . . 77

6 Support Projects at CMS . . . 79
6.1 Tau Trigger Efficiency Measurements . . . 79
6.2 The MIP Timing Detector at the High-Luminosity LHC . . . 81
6.3 The Minimum Ionizing Particle Timing Detector . . . 83
6.4 Conclusion . . . 89

Bibliography . . . 91

List of Figures

1.1 Illustration of the Bohr model 3

2.1 The Standard Model of fundamental particles 8
2.2 Higgs to Tau Tau decay 9
2.3 Organization of particle physics terms 13

3.1 Schematic of the LHC 17
3.2 Pileup profiles at CMS 20

4.1 CMS Overview 23
4.2 CMS Coordinate system 25
4.3 CMS dimuon data in Fireworks 26
4.4 Slice of CMS 28
4.5 Fundamental particles and corresponding physics objects 30
4.6 Hadronic tau decays in CMS 31
4.7 CMS trigger system 32
4.8 Computing tiers 35
4.9 High multiplicity event display 36
4.10 Simulated particle interaction 37

5.1 Summary of measured Higgs couplings 41
5.2 Common Higgs production modes 42
5.3 The Simplified Template Cross Section framework 43
5.4 Timelines of Higgs to tau tau decays at CMS and ATLAS 45
5.5 Pie chart of $\tau\tau$ final state pair fractions 46

5.6 Summarized Higgs to tau tau selection criteria 52
5.7 Possible object pairs and associated final state 52
5.8 Generic fake factor diagram 55
5.9 MuTau fake factor regions 58
5.10 QCD fake factors for MuTau 59
5.11 Extended transverse mass plot 61
5.12 W + jets fake factors for MuTau 62
5.13 Plot of transverse mass in the application region for MuTau 63
5.14 QCD fake factor regions for DiTau 64
5.15 QCD fake factor diagrams for DiTau 65
5.16 Jet fake event type abundance in ditau 67
5.17 Response Plot Example 70
5.18 Response Plot Higgs Transverse Momentum Comparison 71
5.19 Differential Cross Section Results 76
5.20 Comparison of results to Run 2 78

6.1 Comparison of tau trigger efficiency using DeepTau and ParticleNet 81
6.2 Comparison of tau trigger efficiency in data and MC 82
6.3 High-Luminosity LHC schedule 83
6.4 Reconstruction vertex selection with and without MTD 84
6.5 Overview of the MIP Timing Detector 85
6.6 Bethe-Bloche function 86
6.7 Experimental setup for thermal validation studies 87
6.8 Thermal validation of a BTL tray 88
6.9 Thermal validation of planned thermoelectric cooler operation 89

List of Tables

2.1 Forces and relative strengths 7

2.2 Tau decay modes and branching fractions 10

5.1 Branching ratios of the Higgs boson 40

5.2 Cross section values for a 125.09 GeV Higgs boson 42

5.3 Higgs boson channel final states and neutrinos 46

5.4 Offline kinematic requirements for ditau 50

5.5 Offline kinematic requirements for other final states 50

5.6 Kinematic requirements for jets 53

5.7 Table of boundaries for final state categories 68

5.8 Table of differential bin boundaries per variable 68

5.9 Fiducial selection thresholds per final state 72

5.10 Fiducial selection metrics 73

Acknowledgments

Isaac Newton once wrote "If I have seen further than others, it is by standing on the shoulders of giants." I would like to acknowledge all of the scientists and knowledge-seekers who came before me that have made this work possible, and also those who will come after to politely correct our ideas of the universe today.

I am very grateful to have been advised by Ketino Kaadze, who gave me countless opportunities to grow as a physicist and person. Keti went beyond teaching me particle interactions and data science, and guided me to build and navigate a successful project in a very large research collaboration. I would not be the scientist I am today without her support.

I would also like to thank Yurii Maravin for welcoming me into the broader CMS-KSU research group and lending his advice and experience to the analysis. His historical perspective in the development and future of CMS was very useful in understanding the scope and direction of current studies, bringing the "why" of what we do into a clear and achievable focus.

For my thesis and defense I am very grateful to the members of my committee, including Glenn Horton-Smith, Andrew Ivanov, and Victor Turchin, for their comments and guidance. I would like to express my gratitude toward Brett De Paola and Tim Bolton for serving as department head during my graduate studies and navigating the KSU physics department through uncertain times. Also in the KSU department, J. T. Laverty, Daniel Rolles, and Loren Greenman made an impact in my studies by taking the time to teach thoroughly, listen, or simply chat about physics at large. I am also especially grateful to Kim Coy, for her unwavering cheerfulness and support.

It would not have been possible to complete this work without the technical knowledge of the post-doctoral researchers at CMS. In particular, I had the pleasure of working with

Dennis Roy, Giulia Sorrentino, and Kyungwook Nam. More than their expertise and professionalism, I appreciate our group's postdocs for their willingness to teach, good humor, and seemingly infinite availability for a "quick chat." Dennis has been part of the Higgs to Tau Tau measurement from the start, and is an endless font of both particle physics and computing knowledge. He was instrumental in providing and servicing the tools that made this measurement possible.

Beyond KSU, I want to acknowledge everyone I met associated with CMS at CERN who would generously offer their time, expertise, and knowledge. I want to specifically thank Olivier Davignon, Silvia Goy Lopez, Marino Missiroli, Silvio Donato, Mia Tosi, Konstantin Androsov, Sam Harper, Karlis Dreimanis, Adi Bornheim, Xiaohu Sun, Andrea Cardini, Michal Bluj, and Izaak Neutelings, for their high level problem-solving abilities, their ideas, and their kind encouragement. I would also like to thank several CMS members who were graduate students at the same time as myself, including Guru, Amrutha, Botao, Filip, Muti, J.P., Andreas, Ang, Guillermo, Naila, and Athar—many of whom have already graduated and continue to make incredible contributions to science.

Aside from my professional connections, I count myself lucky to have a robust network of friends who believed in me. There are too many good memories to share in one place, so instead I write their names to acknowledge the joy they brought to my life during this work: James, Jo Lynn, Nathan, Rahasya, Jacob, Sajed, Sakhi, ML, Brie, Cee, Luke, Tessa, Holden, Kelly, Jared, Patrick C., Dene, Aditi, Laurens, Will, Nick, Rita, Taylor, Varun, and Patrick M. Roommates, neighbors, tour guides, gym partners, trivia team members, editors, and longtime pals, thank you for your friendship.

My mother, father, sister, and aunt always encouraged me to go as far as I could in science and I am grateful to them for their support on this long journey. For the last several years, I owe many thanks to my partner Annie, who has become the "life" in my work-life balance. I will forever be grateful for her help during this period of my life focused on the doctorate, and I am looking forward to our lives together post-PhD.

Finally, the pursuit of knowledge is a shared goal for us all, and in that vein, I would like to thank the reader for their interest in this subject.

AI Policy

This thesis and the contents therein were outlined, written, and typeset by a human. Generative machine learning tools based on large language models (ChatGPT, Claude, etc.) were not used in the creation of this work.

Preface

The goal of this thesis is to summarize my work at Kansas State University, and give the reader a fighting chance at understanding a part of the current scientific enterprise of particle physics. Many references were consulted, and parts of this thesis will almost certainly be out of date within a year of its completion. But that's what makes this field so exciting and worthwhile.

This thesis is organized so that someone who has no experience with particle physics can follow it throughout. We begin with a general introduction to physics at large, narrowing our questions to the realm of the very small and very energetic. Then we briefly describe the standard model, moving onto the experiments which allow us to make measurements of it. With this ground work established, we finally tackle the Higgs boson decay to two tau leptons. Finally, we conclude with a description of support work I undertook during this thesis.

We are all in a brave new information age, and the skill to reason and think critically about new and varied data is in higher demand now than it ever has been. Studying any topic with the goal of truly understanding a vast base of knowledge and synthesizing something to put on the heap is difficult but worthwhile. If you have questions that the references cannot satisfy, please reach out.

Chapter 1

Introduction

Particle physics today is waiting for the "next big thing." The Deep Underground Neutrino Experiment (DUNE) in the US is planned to be online in 5 years. Europe's enthusiasm for the Future Circular Collider (FCC) is growing. Experiments at the Large Hadron Collider (LHC) are gearing up for a brief upgrade period before the next round of data collection with the highest-yet-achieved instantaneous luminosity. Scientists are cozying up to and innovating truly advanced computing techniques to efficiently read the data they have collected. The public is becoming interested and developing fluency in the language of fundamental particle physics. It is an exciting and busy time in High Energy Physics (HEP).

In spite of, or maybe because of, this air of optimism and excitement for the next groundbreaking result, studies extending previous results continue. Cross sections for important processes are re-evaluated at higher collision energies; new strategies are tested on familiar final states; searches edge out until-now-unexplored phase-spaces. This may seem like small potatoes compared to tantalizing mysteries like the di-Higgs coupling, neutrino Yukawa couplings, or any flavor of Beyond the Standard Model extension. But, these studies are important precisely because they continue to test the familiar, and chip away at the almost-known. The "next big thing" may yet take some time to arrive. In the meanwhile, scientists have an obligation to eke every drop of information out of the data they have collected.

This thesis begins with a brief foundation in particle physics, followed by a description

of the collider and the detector. After, the differential cross section analysis of $H \to \tau\tau$ mentioned in the abstract is described in full, subdivided into major sections. Additional service work contributed to the experiment outside of the analysis work is described in the final chapter.

What is Physics?

The universe is very complicated. To date, there is no single theory that describes all phenomena at all scales. But, there are several frameworks of ideas that give precise, testable predictions. This is the heart of science: making a prediction, and comparing it to a measurement. Physics today is the subject of study that was once called "the natural sciences," and it primarily concerns itself with the dynamics of systems and fundamental properties of matter, such as mass and charge. In the last 100 years, a quantum revolution in physics led to transformative technologies like LCD screens, GPS, and X-ray imaging. Quantum mechanics investigates the very small world of atoms. Niels Bohr created the model of the atom that many are familiar with today, shown in Fig. 1.1, inspired by the orbits of planets[1]. This model, though useful, is incomplete. Over the next 50 years, as detector technologies outpaced theoretical developments, more and more particles were discovered that were not the familiar protons, neutrons, and electrons of Bohr's model.

Slightly before the birth of quantum mechanics, Albert Einstein shared his theory of general relativity with the world[2]. Among other predictions, he supposed a mass-energy equivalence for particles, namely $E = \sqrt{(mc^2)^2 + (pc)^2}$. In this equation, E represents the total kinetic and potential energy of a particle, m represents the mass, p momentum, and c the speed of light. An important prediction of Einstein's is that c is a constant that is the same everywhere, a speed limit for the universe, and is the speed that photons travel in vacuum, equivalent to 700 million miles per hour, 299 792 458 m/s, or approximately 3×10^{10} cm/s[3].

Above, quantum mechanics makes the cognitive leap that there are individual things that interact with each other, and general relativity provides useful rules and relations for how

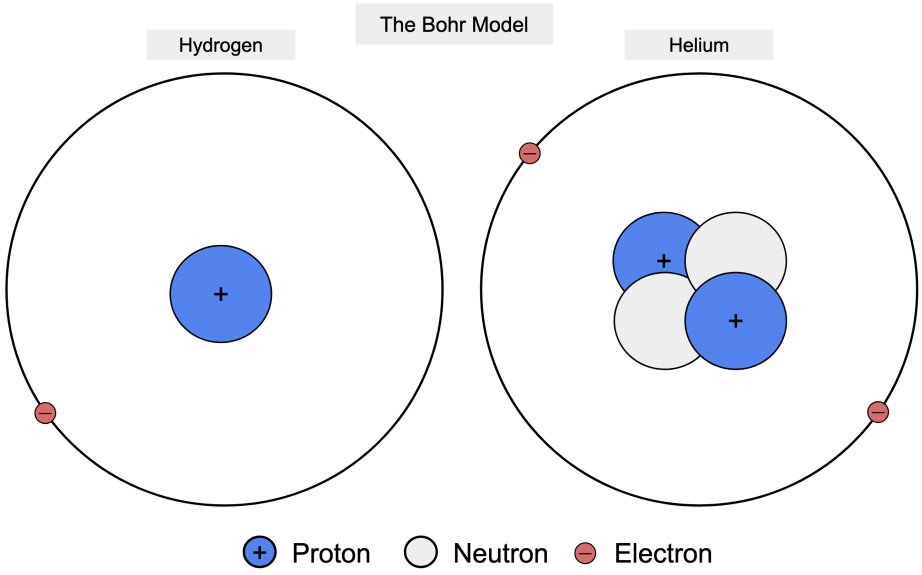


Figure 1.1: *The Bohr model depicted with a hydrogen atom and a helium atom. The model supposes a small, heavy nucleus of protons and neutrons as the center of an atom, orbited by electrons. In the Bohr model, an atom is 99% empty space. It also supposes that electrons appear in "orbital shells" around the nucleus, and transitions of electrons between the shells always result in the same emitted energy, in the form of a photon, characteristic of the energy shell transition. (Own work)*

those things may change. With these ideas, and 100 years of development, we can extract a single electron or proton, bombard it with photons (light from a laser), and predict then precisely measure how the particle will move or emit energy. We know this as sure as we know the sun will rise tomorrow. However, for all we do know now, there is even more we are still trying to understand.

We measure much stronger gravitational effects in the universe than we would expect from the amount of observable matter. We suppose this additional force comes from matter that does not interact with measurement instruments, and call it *dark matter*. We measure the universe to be expanding, and by thinking in reverse we assume that in the distant past the universe was contained in one very small, very hot point. This point exploded between 13 and 14 billion years ago in one "Big Bang," followed by a period of massive inflation. The exact dynamics and evolution of this system are lost to time, though it has left observable artifacts such as the Cosmic Microwave Background, and asymmetries in the types of matter in the universe.[4]

These questions, and more, are the realm of cosmology and HEP, the latter category of which this thesis falls under. Cosmology and HEP address these questions in different ways. In cosmology, one looks at the entire universe, peering intently into the night sky with highly advanced telescopes, and documenting the evolving properties of our universe. In HEP, we look at particles under conditions similar to the early universe recreated by powerful colliders and observed by building-sized detectors. Before those machines are introduced, let me introduce you to the Standard Model.

Chapter 2

The Standard Model

The Standard Model, or SM, is a framework of how particle interactions occur. It is a quantized field theory, meaning it is based in quantum mechanics (quantized), and describes physics at all points of space and time (a field). The majority of the SM was described theoretically by the mid 1970s[5;6]. In the nearly 50 years since, experimentalists created ever-larger experiments to find heavier and heavier particles predicted by the SM, as well as rarer and rarer interactions. Its crowning achievement was in 2012, with the joint discovery of the Higgs boson[7;8] by two experiments at the Large Hadron Collider (LHC), namely A Toroidal LHC Apparatus (ATLAS) and the Compact Muon Solenoid (CMS).

Because of the extremely small scale of particle physics, the unit "electron volt" or "eV" is used everywhere, normally with prefixes of kilo, mega, giga, and even terra (k, M, G, T). An eV is the amount of energy to accelerate an electron through a potential difference of one volt in a vacuum[5;6]. It is equivalent to 1.602×10^{-19} Joules of energy. Particle masses are typically expressed in units of eV/c^2, from the mass-energy equivalence relationship. Because c^2 appears so often in particle physics quantities, people normally work in *natural units*, where both c and $\hbar$ are set to 1, and evaluate the units of specific quantities from context.

Before introducing all the fundamental particles and their interactions, it is useful to say a few things about matter, energy, and forces. Matter, as I have used it until now,

includes all atoms, their subatomic constituents, and electric charges. Importantly, photons (the fundamental particles that we call "light") are not traditionally considered matter[5;6]. This is because photons transmit energy between electrically charged units of matter. This transfer of energy is effectively how pieces of matter "communicate" or pass information to other pieces of matter. In this school of thought, two electrons do not directly interact, but "feel" each others' electromagnetic field through the exchange of photons. This is a general principle of quantum field theory, and separates fundamental particles into two types: matter, and force mediators. In the case of light, electric current, and magnetic fields, the photon mediates the force.

The force-mediators of a field are called "bosons," and receivers of said force are called "fermions." In general, fermionic and bosonic properties refer to quantum numbers of any configuration of atoms, but in the context of fundamental particles these descriptors delineate between matter and force mediators. Fermions do not interact with one another except through a mediating boson. For example, in the SM, electrons do not push or pull electrons, rather, an electron interacts with a photon which interacts with another electron.

A force is another word for an interaction, simply indicating a piece of matter is being pushed or pulled in some direction. There are four types of forces observed in the universe: the strong force working inside the nucleus of atoms, the weak force which explains spontaneous radioactive decay, the electromagnetic force, and the gravitational force. Contrary to the other observed forces, the SM makes no inclusion or predictions about gravity[5;6]. Additionally, at the subatomic, high-energy scale where the SM is most often invoked, gravitational forces are vanishingly small compared to other forces as shown in Table 2.1. Consequently, there is currently no theorized or observed particle that mediates the gravitational force in the SM.

As shown in Fig. 2.1, there are 17 fundamental particles, represented by Greek or English characters. This number is not considering anti-particles, and, in some cases, a particle is its own antiparticle, such as the photon, further complicating the exact count of fundamental particles. The major groupings of the figure are the 12 fermions and 5 bosons, which are further categorized into 6 quarks, 6 leptons, 4 force-carrying bosons, and 1 scalar boson,

Table 2.1: *This table refers to approximate strengths at the scale of a bare hydrogen atom (proton). The strong force is set to 1, as it is the strongest among all forces. The electromagnetic force is characterized by the fine structure constant, α. The Weak nuclear force is approximately one-tenth the value of α, though it is difficult to quantify as the relative strength depends on the stability of the particle under consideration as well as the energy of the interaction. Finally, the gravitational force is a factor of 10^{-30} times less powerful than the electromagnetic force at this scale, based on naively using Newton's force of gravity. (Griffiths page 59, Chapter 2, adapted)*[6]

Force	Relative Strength	Associated Boson	Example Interaction
Strong	1	Gluon (g)	Binding quarks together to form protons
Electromagnetic	$\alpha = \frac{1}{137}$	Photon (γ)	Radio waves, TV screens, Lightning
Weak	$0.1\alpha \approx 1/1000$	W and Z	Nuclear decay
Gravitation	10^{-30}	-	Planetary motion, satelite orbits

each of which will be discussed in more detail below.

The SM changes slightly each year, with new estimations of particle mass, lifetime, and charges, so for any referenced particle parameter assume the Particle Data Group (PDG) is the source[9], unless otherwise supplemented with additional references.

2.1 Leptons

There are two types of leptons: charged, and uncharged. Charged leptons include the electron (e), the muon (μ), and the tau lepton (τ). Throughout this thesis tau leptons are typically referred to simply as "taus." Charged leptons carry one unit of fundamental charge, -1.602×10^{-19} Coulombs, or simply -1, and charged lepton antiparticles, the positron, anti-muon, and anti-tau, all have positive charge. As far as can be understood from the SM, the only difference between electrons, muons, and taus, are their mass. Electrons have a mass of 9.109×10^{-31} kg, or 0.511 MeV/c^2. Muons, with a mass of 105.6 MeV/c^2, are approximately 200 times more massive than electrons, and taus are approximately 3500 (15) times more massive than electrons (muons), having a mass of $1.776 GeV/c^2$. The electron is a stable fundamental particle, meaning it cannot decay into another particle, or at least such a decay has never been observed. Muons, being more massive than electrons, are able to decay, but only into an electron and two neutrinos. Taus, being the most massive of all, can decay

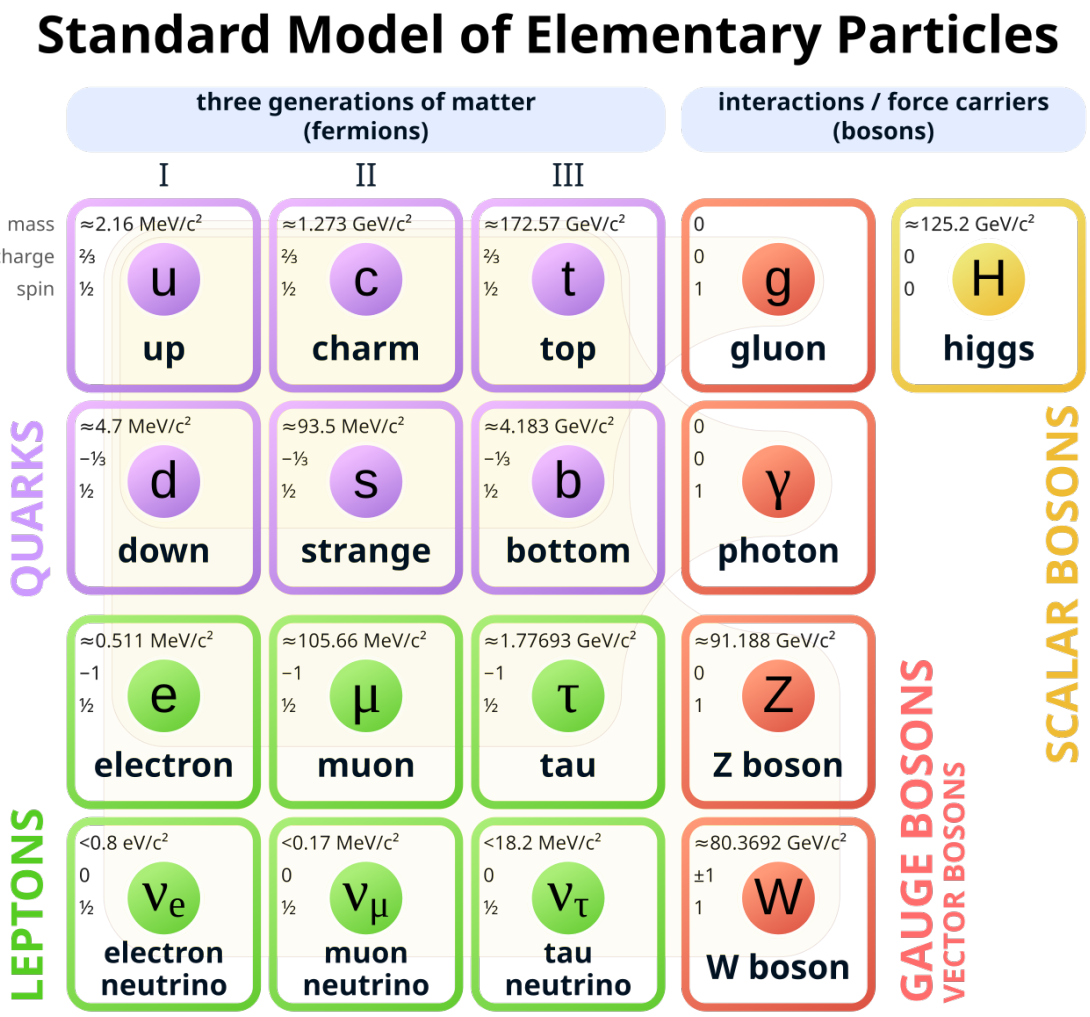


Figure 2.1: *The Standard Model of fundamental particles, organized into fermions (left) in three groupings roughly related to mass (generations), and bosons (right). Purple and green subcategorize the fermions into quarks and leptons. Vector bosons are shown in orange, while the Higgs boson is shown in yellow because it is the only scalar boson. In the background, different collections of bosons and fermions are grouped in a faint yellow to indicate which forces interact with which particles (mass values updated in 2024,* [10]*)*

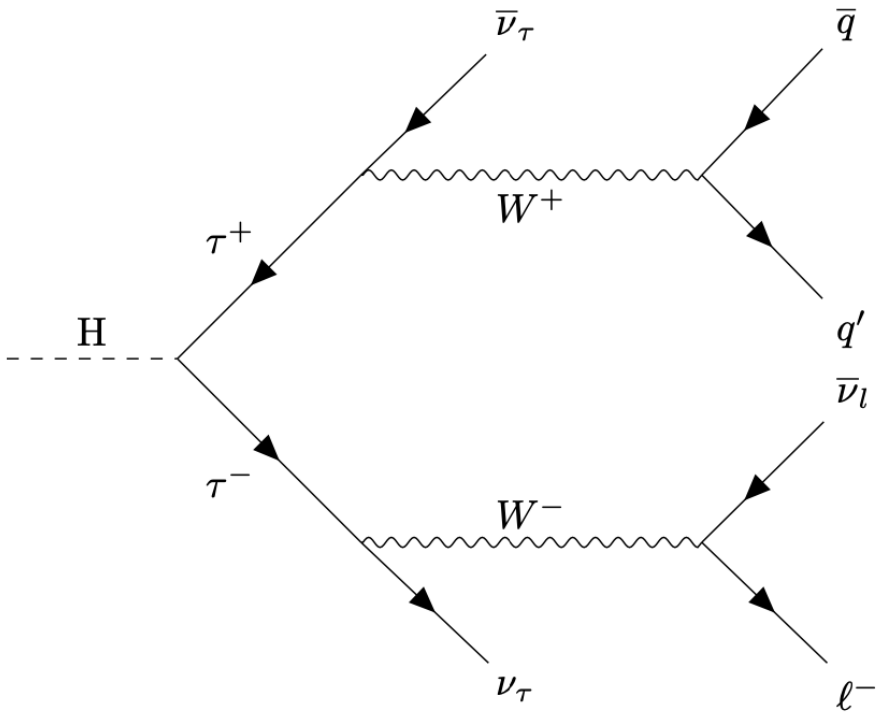


Figure 2.2: *A Feynman diagram of a Higgs boson decaying to two tau leptons, and the taus decaying hadronically (top) or leptonically (bottom). Both decays are through the weak force and subsequently involve neutrinos. Tau leptons inferred from hadronic decay products are called hadronic tau (τ_{h}) candidates*[11].

into muons, electrons, and mesons, as shown in Fig. 2.2 and Table 2.2. Mesons are pairs of quarks, the lightest of which is the pion family, with the charged pions ($\pi^{+/-}$) having a mass of 139.5 MeV/c^2 and the neutral pion (π^0) with a mass of 134.9 MeV/c^2. Notice how this mass is just above that of the muon, making muon-to-charged-pion decays inaccessible at rest as a consequence of mass-energy conservation.

Uncharged leptons are called neutrinos, because they have a very small mass (from the italian "-ino") and because they carry no electrical charge. In the 1970s, the SM originally described neutrinos as massless, but experiments in the early 2000s amended this picture with the discovery of neutrino oscillations[13], a behavior only possible with non-zero mass. Neutrinos are represented with the greek letter ν (nu). Typically, one talks about the flavor states of neutrinos in the SM, the electron neutrino, the muon neutrino, and the tau neutrino, as well as anti-neutrinos (the electron antineutrino, muon antineutrino, and tau

Table 2.2: *Table reproduced from Ref. 12. Decay modes and branching fractions ($\mathcal{B}$) of the tau lepton alongside the mesonic resonances primarily involved in hadronic tau lepton decays*[9].

Decay mode		$\mathcal{B}$[%]	Resonance [MeV]
Leptonic decays	35.2		
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$		17.8	
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$		17.4	
Hadronic decays	64.8		
$\tau^- \to h^- \nu_\tau$		11.5	
$\tau^- \to h^- \pi^0 \nu_\tau$		26.0	ρ (770)
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$		9.5	a_1 (1260)
$\tau^- \to h^- h^- h^+ \nu_\tau$		9.8	a_1 (1260)
$\tau^- \to h^- h^- h^+ \pi^0 \nu_\tau$		4.8	
Other hadronic modes		3.2	

antineutrino). Though they are a rich and exciting topic, for the purposes of this thesis the most important aspects of neutrinos is that they are present in charged lepton decays and are extremely difficult to detect. The implication of this is a reduced efficiency and energy resolution for experiments not directly detecting neutrinos, such as CMS.

2.2 Quarks

A quark—a name borrowed from one of the many gibberish words in James Joyce's work *Finnegans Wake*[14]—is a fundamental particle originally found inside protons and neutrons. Quarks are displayed in three generations, roughly grouping their mass, and have two primary characteristics: their "flavor," and their charge. The six flavor types are up, down, charm, strange, top, and bottom, abbreviated as u, d, c, s, t, and b. Up, charm, and top quarks are all said to be "up-type" quarks, and down, strange, and bottom quarks are said to be "down-type" quarks. Quarks have a fractional fundamental charge of 2/3 for up-type quarks, and -1/3 for down-type quarks. In the SM, all *observable* particles have integer charge, so the fractional charge of quarks is supported by experimental evidence of observed composite particles and particle decays[5;6].

Quarks have never been observed in isolation (i.e. as a single quark). Groupings of 3 quarks are called "baryons," and groupings of 2 are called "mesons." Exotic quark groupings of 4 or 5 are known to exist and are called "tetraquarks" and "pentaquarks," but are very rare and short-lived[15]. Any grouping of quarks may generally be called a "hadron." Examples of baryons are protons (uud) and neutrons (udd), and examples of mesons are the pion family, π^+ ($u\bar{d}$), π^- ($\bar{u}d$), π^0 ($\frac{(u\bar{u}+d\bar{d})}{\sqrt{2}}$), mentioned in the charged lepton section. Groupings of quarks are held together by gluons, the mediator of the strong nuclear force[5;6].

Quarks are another rich subject within the SM, forming the branch of "flavor" physics, but for the purposes of this thesis we will limit our scope of interest to proton-proton collisions and the resultant hadrons. When protons are pushed together with force sufficient to overcome the strong nuclear force binding its nucleus, their individual quarks (as well as gluons) become free to take part in any allowed interaction of the SM. Typically only 2 of the 6 quarks participate in this interaction, and the other 4 recombine with gluons in a cascade of particles through a process known as "hadronization," the result of which is a hadron shower. This is a consequence of quark-gluon confinement, the fact that quarks and gluons cannot appear in isolation in the SM[5;6].

2.3 Force-carrying Bosons

Care has been taken here to avoid the terms "gauge" and "vector" boson, as it is beyond the scope of this thesis, and treated more completely in standard particle physics texts (see Mandl and Shaw Chapter 11[5]). Suffice it to say that force-carrying bosons arise from any quantized field theory that is required to be invariant under local transformation, i.e. a theory stipulating that the universe follows the same laws of physics everywhere (invariant under local transformation). In addition, the conservation of charge (we only observe integer charges, and electrons do not lose their charge over time), and other quantum numbers introduces more symmetries in the SM. This ultimately results in 8 fundamental particles that mediate strong nuclear interactions, collectively called gluons and displayed in the SM with a lone g, and 4 fundamental particles mediating the unified electroweak force, denoted

as the γ, W^+, W^-, and Z bosons. Below energies of about 100 GeV the electroweak force becomes two distinct forces, the electromagnetic force and weak nuclear force, introduced previously. What is meant by "unification" is simply that the forces' interaction strengths become equivalent at some energy, meaning that if it is possible for a particle to decay via either the electromagnetic force or the weak nuclear force, there is an equal likelihood of either, other restrictions not withstanding. The 100 GeV threshold arises from the masses of the weak nuclear force mediating bosons, $W^{+/-} = 80.4\,GeV/c^2$ and $Z = 91.2\,GeV/c^2$, which effectively constrains the range of the force[5;6]. The photon, being massless, has an infinite range, and can interact with charged particles at any energy scale.

2.4 The Higgs Boson

The Higgs field is how the SM explains massive particles. Without it, there is no way to formulate the SM that includes mass terms without also including infinite values at some scale, which is undesirable and non-physical[5;6]. The mechanism by which other particles acquire mass from the Higgs field is called spontaneous symmetry breaking, and requires a non-zero ground state of a scalar field permeating the universe. All matter must traverse this field, and a particle's mass denotes how strongly it couples to the field. The Higgs boson is then the quanta of this scalar field, with a measured mass of $125.10 \pm 0.11\,GeV/c^2$, slightly above the electroweak sector. It is the most recently discovered fundamental particle of the SM, and its existence was predicted 40 years ahead of its discovery[5;6]. Importantly, the mass of the Higgs is not predicted from the SM, but once a mass is given then predictions of the Higgs coupling to other fermions can be made. Note that the Higgs field is not a force, and overall has "weaker" interactions than those mediated by the force-carrying bosons, making interactions involving a Higgs boson less common. More detail on the Higgs boson will be supplied in Chapter 5 in the course of motivating the study of Higgs to tau lepton decays.

The different terms used in sections 2 - 2.4 are neatly summarized in Fig. 2.3.

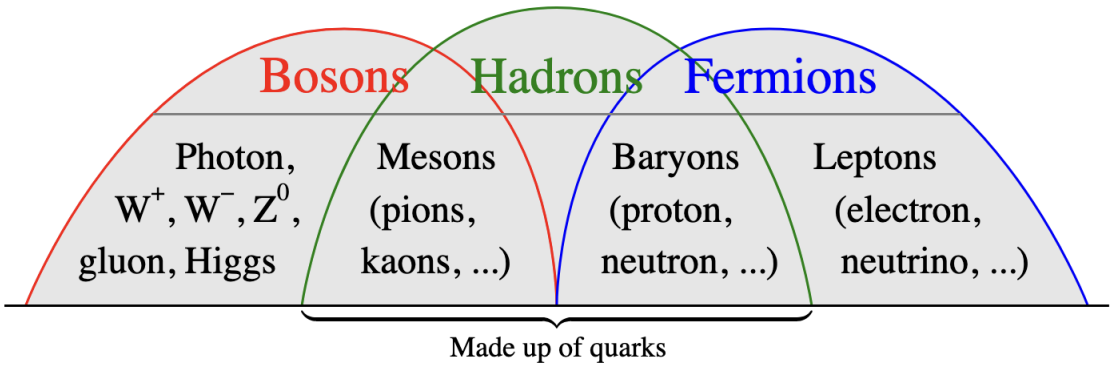


Figure 2.3: *This diagram shows three umbrella terms of particles and where additional terms fall within them, helpfully illustrating intersections and examples*[16].

.

2.5 Permitted Interactions

As mentioned, force-carrying fundamental particles are bosons that each mediate one force.

- Gluons mediate the strong nuclear force, interacting only with quarks and other gluons, not leptons or other force-carrying bosons.
- Photons mediate the electromagnetic force, interacting with any charged particle, including quarks, charged leptons, and charged bosons, not gluons or neutrinos.
- The weak force is divided into "neutral current" and "charged current" interactions, referring to the charge of the mediating boson.
  - Weak neutral interactions are mediated by the Z boson, and can interact with quarks, charged leptons, neutrinos, and other bosons, except for gluons and photons. Notably, many photon interactions with charged leptons could be replaced by the same interaction with the Z boson.
  - Weak charged interactions are mediated by W bosons, which are unique in the SM for allowing flavor changing. Through W boson interactions, and no other interaction, a top quark can become a bottom quark, a tau can become a muon or even transform into quarks as shown above. The W boson interacts with quarks, charged leptons, neutrinos, and other bosons except for the gluon.

- The Higgs boson, though not a force-carrying boson, interacts with all particles having mass, including quarks, charged leptons, and other bosons, except for massless photons and gluons and neutrinos, which acquire their mass through an as-yet-unknown mechanism. Notably the Higgs also has a pure self-interaction, which is not allowed for other bosons. The strength of interaction between fermions and scalar fields is termed a "Yukawa coupling." For example, one would say the Higgs fields' strongest Yukawa coupling to a lepton is through the tau, meaning the most common leptonic decay of the Higgs boson is to a pair of tau leptons.

I was taught a simple maxim for particle physics processes: If it is allowed, then it is mandatory. And, even with the numerous conservation laws of the SM, many things are allowed.

Unanswered Questions of the Standard Model

The SM as described above is in many ways quite complicated, but also surprisingly simple. In its current form, the SM accurately predicts many interactions and quantities of nature based on a handful of particles and symmetries. One might expect a much more complicated theory being necessary to account for the same wide array of behaviors, and yet, we achieve high-precision measurements and predictions for every type of fundamental particle with one single theory.

That said, there are features of the SM that do not explain themselves. There are apparently 3 generations of fermions, roughly grouped by mass, and separated by empirically observed quantum numbers. There is no obvious explanation for why the SM would stop at 3, instead of having 1 generation of fundamental particles with similar masses, or 5 generations with larger mass scales. Similarly, neutrinos are known to have mass, but it is much smaller than any other massive particle of the SM, a million times lighter than an electron. This wide disparity of masses is known as the hierarchy problem[17].

Separately, there are many physical phenomena that are not explained or predicted by the SM. As mentioned before, there is no gravitational force within the SM, and to date,

no experimental evidence for a fundamental particle related to a gravitational field. That said, there is abundant evidence that both normal matter and anti-matter interact with gravitational fields[18], as well as force mediating particles such as the photon through gravitational lensing[19]. Another glaring problem of the SM is its lack of prediction for a dark matter candidate. One would expect that a theory that makes predictions about matter might have insights beyond the standard matter we interact with daily. Although multiple beyond-the-standard-model (BSM) theories have been proposed and tested, such as the minimally supersymmetric standard model[20] and Randall-Sundrum models[21], none yet have found evidence for a candidate particle of dark matter.

These puzzles, among others, have been under active research for decades. In that time, experiments have steadily progressed to be larger and higher energy. The primary contribution of the U.S. in that time were discoveries of quarks at the Stanford Linear Accelerator Complex (SLAC), Berkeley National Lab (BNL), and Fermi National Lab (FNAL or Fermilab)[6;22]. Abroad, multiple countries composed the European Center for Nuclear Research (CERN), which discovered the W and Z bosons in the 1980s[23;24], and the Higgs boson in 2012.

This thesis does not present any answers to these fundamental questions. Instead, what is presented is a precision measurement of Higgs boson decays to tau leptons. This measurement seeks to improve the knowledge of the Higgs sector, increasing the precision of previous results. Taken together with other measurements of the Higgs boson, such results can reduce the number of BSM theories permitted with current experiments and further refine the shape of the Higgs potential through the strength of the Higgs self-interaction. Necessary components of detecting Higgs bosons are a sufficiently powerful accelerator, and a general-purpose detector, detailed in the following chapters.

Chapter 3

The Large Hadron Collider

In the following paragraphs I will introduce the basic facts of the Large Hadron Collider relevant to this thesis. There is of course much more beyond the basics, and I refer the interested reader to the technical design document and education material by Evans for further information[25;26].

The European Center for Nuclear Research, CERN, was founded in the early 1950s. Its goal was to pool scientific resources in Europe to study the fundamental building blocks and interactions of matter[27]. Over decades of operations and experiments, the amount of discoveries, number of member states, and collision energy all increased. Spurred by the observation of the W and Z bosons in the early 1980s at the Super Proton Synchrotron[23;24], CERN took on its most ambitious activity so far with the construction of the Large Electron-Positron collider, LEP[28]. It was a 27 km (19 mile) circumference ring built 100 meters (300 feet) underground, housing two opposing beamlines, and four experimental caverns. This experimental hall was used to conduct precise studies of the W and Z bosons for over a decade, eventually reaching a collision energy of 200 GeV. It was later repurposed for a proton-proton collider named the Large Hadron Collider (LHC).

The LHC powered on and quickly reached collision energies of 7 TeV, and later attained collision energies of 13.6 TeV[26], the highest of any collider in the world. At the time of the LHC construction, the next most powerful collider was the Tevatron at Fermilab, capable

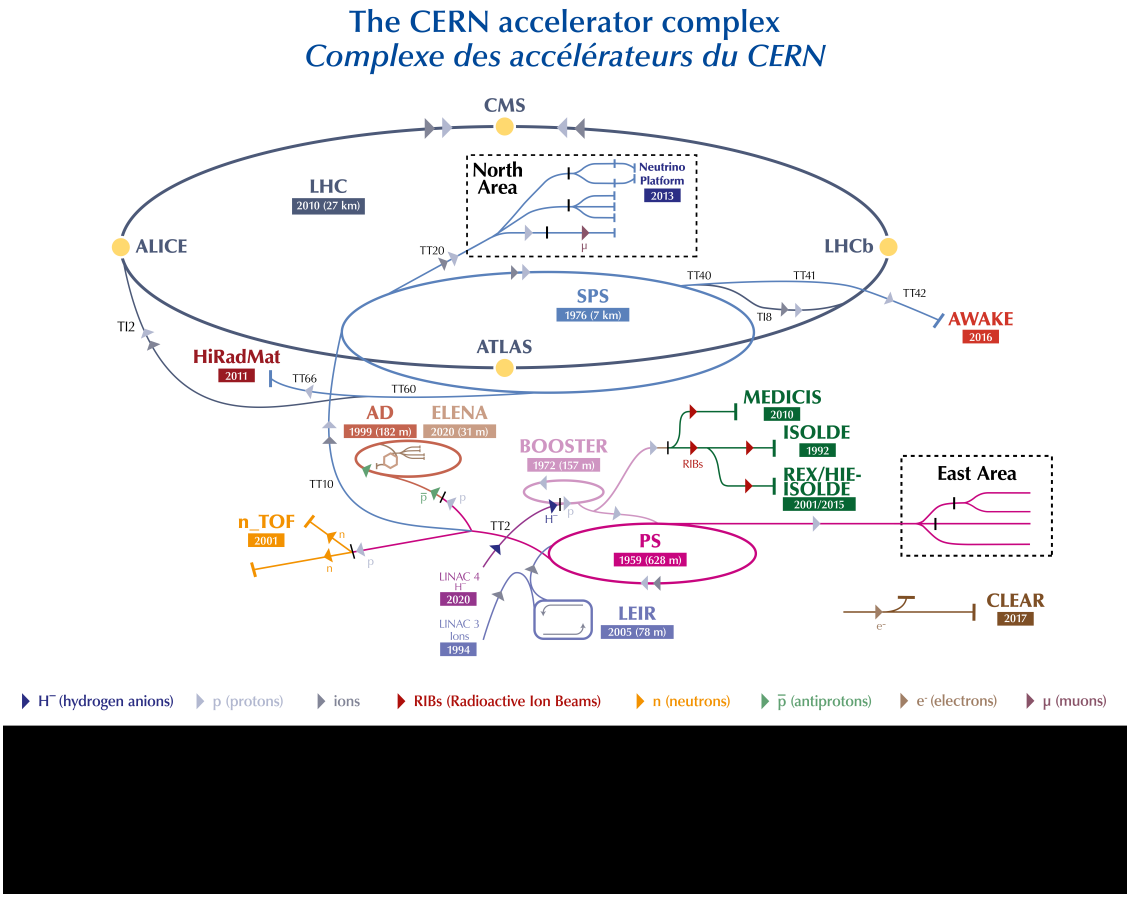


Figure 3.1: *This figure schematically shows the LHC (largest ring) and the smaller booster rings that lead to it. The year of construction is shown next to major rings, linear accelerators (LINACs), and experiments*[30].

of 2 TeV collisions[29]. This large increase in collision energy made the discovery of the Higgs boson possible, and also produced many uncommon top-quark-initiated processes. The relatively abundant source of Higgs bosons began a decade of Higgs discoveries and advancements, discussed further in Chapter 5. Besides Higgs physics, the LHC beamline and accelerator complex supports over forty other experiments related to nuclear science[30], as shown in Fig. 3.1. The major data collection periods of the LHC were from 2010-2012, 2016-2018, and 2022-2026, and are referred to as Run 1, Run 2, and Run 3, respectively.

Beam energies are the most limiting parameter of a collider experiment. Without sufficiently energetic collisions, massive particles cannot be produced due to the mass-energy equivalence relationship. After beam energy, the next most limiting factor is luminosity,

which is subcategorized into total luminosity and instantaneous luminosity. Total luminosity is determined by the instantaneous luminosity and the duration of time that an experiment runs, and is effectively the size of a collected dataset. Using the proton-proton interaction cross section of approximately 100 millibarns, 1 fb^{-1} equates to about 100 trillion proton-proton collisions.

This estimate comes from a fundamental "counting equation" in particle physics which makes luminosity much easier to understand:

$$\mathcal{L}\sigma = N \tag{3.1}$$

Above $\mathcal{L}$ is luminosity, σ is the cross section, and N is the number of particles. Note that N is a unitless quantity, meaning the units on the other two variables must cancel out.

Cross section, as used above, refers to the probability of a certain interaction taking place. A simple way to think about this is a fixed target experiment, where one person throws an object at a location. If the object or the location is very large, then the cross section is very high and results in an interaction between the two (a beach ball at the beach will likely hit the water). On the other hand, using a very small object or a very small location will likely not result in an interaction (flipping a dime into a cup can be very challenging without sufficient practice). Referring to these macroscopic interactions with "cross section" is a little odd, but it makes physical sense thanks to the direct relationship between size and interaction of macroscopic bodies.

Fundamental particles are very small and necessarily obey quantum mechanics, meaning their interactions are probabilistic. Nevertheless, cross section is measured in units of area as though the particles were hard bodies. The typical measurement is made in units of $10^{-24}\,cm^2$, also known as a *barn*. This unit is actually a historical inheritance from colliding Uranium atoms, and since fundamental particles are very much smaller than atoms, cross sections at modern colliders are typically measured in atto-, pico-, and femtobarns[31].

Since the units of cross section are fixed, the units of luminosity are inverse area. An intuitive explanation arises if one considers luminosity as the flux through a target multiplied

by proton density for one unit of time, i.e. $\frac{m^2}{s}\frac{1}{m^3}s$. This provides units of inverse length, and to get to inverse area the expression is squared, accounting for the fact that there are two proton beams, each the target of one another. Luminosity is often discussed in two ways, first in total luminosity, which is the ultimate size of a dataset, and in instantaneous luminosity, which is how quickly a dataset can be collected. The total Tevatron dataset size was about $10\,\mathrm{fb}^{-1}$, while the LHC has achieved over $500\,\mathrm{fb}^{-1}$ [32]. The combination of much higher collision energy and luminosity has created the largest particle physics datasets ever, probing energy regimes once only reachable by rare cosmic rays.

Instantaneous luminosity is primarily determined by the beam size and the bunch crossing rate, with further modifications from other more nuanced effects related to relativistic charged-beam interactions and crossing angles [33]. Here a "bunch," is a collection of protons. Since protons are much smaller than the precision of the magnetic fields which are used to control them, it is unlikely they will collide if they are aimed at one another. Subsequently, many protons, about 1 trillion per grouping, are bunched together and then prepared for collision with another bunch. Even with this large number of atoms concentrated in one spot, only a handful of each bunch will collide, the rest will simply pass through and continue circulating. Since collisions actually use up part of the beam, the instantaneous luminosity decays over time, which has implications for the trigger rate of experiments on the beamline [34].

The detector was designed with a collision frequency of 40 MHz, the inverse of which gives a 25 nanosecond window between each bunch crossing. Using the 27 km circumference of the LHC and approximating the proton bunch speed with the speed of light, one can find it takes 90 microseconds for a proton to traverse the entire beamline. This is a good approximation because the LHC is really 26659 meters in circumference, and the bunch speed is over 99.99% c [25]. Combining the 40 MHz bunch collision frequency with the 90 microsecond flight time for a single bunch, one arrives at a maximum value of 3600 proton bunches. Any more and the spacing between bunches would be smaller than 25 nanoseconds. In reality, the number of bunches containing protons is closer to 2800, where the remainder are left empty to improve collision efficiency. This bunch crossing rate results in an average instantaneous luminosity of $2.0 \times 10^{34} cm^{-1} s^{-1}$ [32], 200 times what was achieved at the Tevatron [35]. To date, the LHC

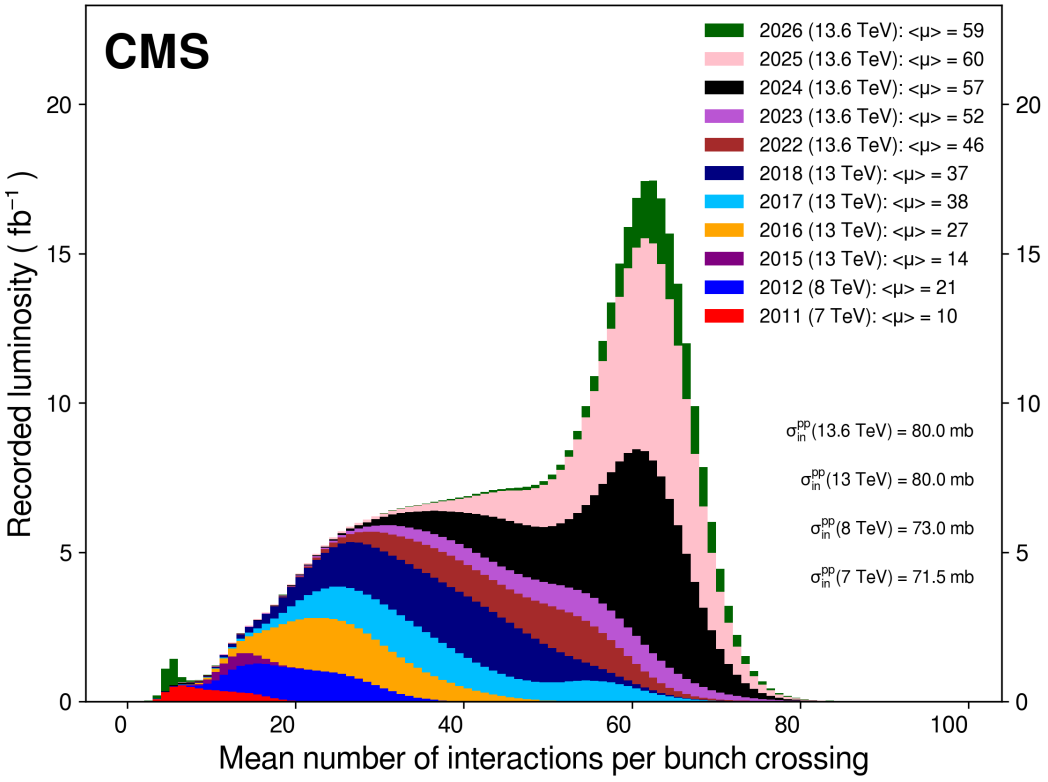


Figure 3.2: *Pileup profiles delivered by the LHC and collected by the Compact Muon Solenoid detector (CMS) for all data-taking years from 2010 to 2026*[32]*. Pileup is jargon used by particle physicists, and is equivalent to the displayed x-axis, "Mean number of interactions per bunch crossing."*

has produced nearly $500\,\mathrm{fb}^{-1}$ of data. The total of all hadron collider data collected in the world prior to the LHC was $11\,\mathrm{fb}^{-1}$[36], 10 of which was from the Tevatron.

As luminosity increases, the effect of multiple protons colliding per bunch crossing becomes more noticeable. Typically there is one collision of two atoms which produces the highest energy decay products, called the primary vertex (PV). Additional collisions between other atoms in the same bunch result in "softer" decay products. This is known as "pileup," because these softer collisions produce many relatively low energy particles that pile up on top of the high energy interactions and obscure them. As the beam collision angle tightened over time, the pileup profile of collected collisions has increased, as shown in Fig. 3.2.

Increasing pileup has implications for detector component radiation damage as well as

trigger rate which will be discussed in the next chapter. The current average pileup of 60 is manageable, but is an open challenge to resolve in the High Luminosity LHC upgrade (HL-LHC), which will reach pileup values between 140 and 200. Technology to mitigate this will be discussed in Chapter 6.

Chapter 4

The Compact Muon Solenoid Detector

There are four building-sized experiments directly on the beamline of the LHC. They are ATLAS (A Toroidal LHC Apparatus), LHCb (LHC beauty-experiment, another name for the bottom quark), ALICE (A Large Ion Collider Experiment), and CMS (The Compact Muon Solenoid). As its name suggests, ALICE was built to collect data on heavy ion collisions, such as lead, oxygen, and gold nuclei[37]. The LHCb is used for heavy ion collisions, but it was specifically constructed to improve measurements of bottom quarks and rare hadrons[38]. ATLAS and CMS are both general-purpose particle detectors[39;40], used primarily to capture events from proton-proton collisions, but also contributing to a wide variety of physics studies, including heavy ion measurements and lepton dipole momentum studies[41;42].

4.1 Overview

The CMS detector is interchangeably referred to as the CMS experiment or simply CMS. The detector is "Compact" with respect to other modern, general-purpose detectors such as ATLAS. CMS is 15 meters (49.2 ft) in diameter and 28.7 meters (94.1 ft) long, resulting in a total volume of 5071 m^3, the same volume as two Olympic swimming pools. In comparison,

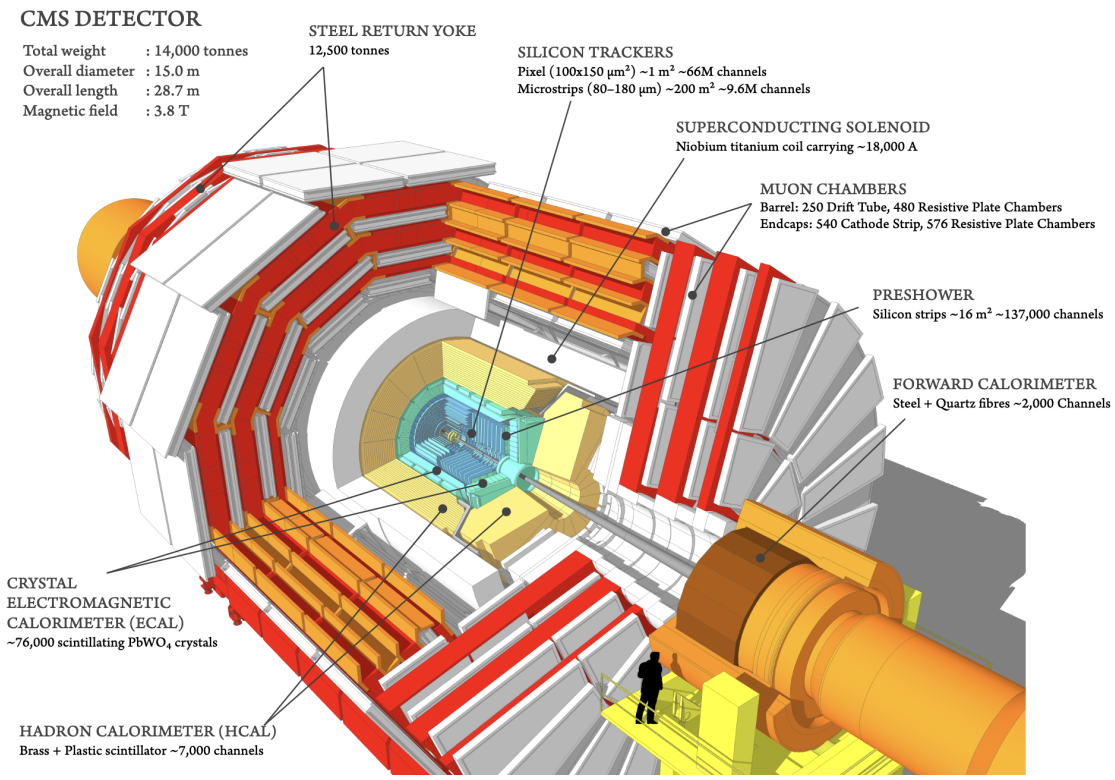


Figure 4.1: *Reference diagram of the CMS detector with a quarter section removed to show the arrangement of subdetectors*[43].

ATLAS is 60% longer and has a 66% larger diameter, resulting in a total volume of nine Olympic swimming pools, over 3x that of CMS.

The "Muon" is included explicitly in the name of CMS because there are three redundant muon detector technologies used in the detector, resulting in better than 99% muon identification efficiency in most of the detector. This was a critical component of the H $\rightarrow$ ZZ analysis (ZZ $\rightarrow \mu\mu\mu\mu$, ZZ $\rightarrow$ eeee, ZZ $\rightarrow$ ee$\mu\mu$) used to discover the Higgs boson[44].

Finally, the "Solenoid" is the central feature of CMS, placed in the middle of an exiting particle's trajectory to alter the course of charged particles mid-flight, giving another handle to distinguish particles from one another. It has an internal diameter of 6 meters and a field strength of 3.8 Tesla[40]. For comparison, a typical magnetic resonance imaging (MRI) machine used in hospitals has an internal diameter of 0.6 meters and uses field strength of 1.5 Tesla for patient imaging[45;46].

4.2 Coordinate System

CMS is cylindrically symmetric because most particle collisions result in particles with flight trajectories perpendicular to the beamline. A set of cylindrical coordinates that are invariant under relativistic transformations are (p_{T}, η, ϕ)[40;47]. Here, p_{T} is transverse momentum, meaning the momentum of a particle perpendicular to the beamline. Psuedorapidity, η, is defined as $\eta = -ln[tan\frac{\theta}{2}]$, where θ is the angle between p_{T} and the beamline as shown in Fig. 4.2[48]. Finally, ϕ is defined as the angle between the plane of the circular beamline and p_{T}, with $\phi = 0$ pointing towards the center of the LHC, as shown in Fig. 4.2.

Fireworks[49] is a useful tool developed by CMS to visually explore events. Fig. 4.3 shows one event from three different views to demonstrate key aspects of the coordinate system.

4.3 Subdetectors and Physics Objects

Many publications have been made with data collected by CMS, organized by topic online at Ref. 51, all of which must include a description of the detector. It was decided early in

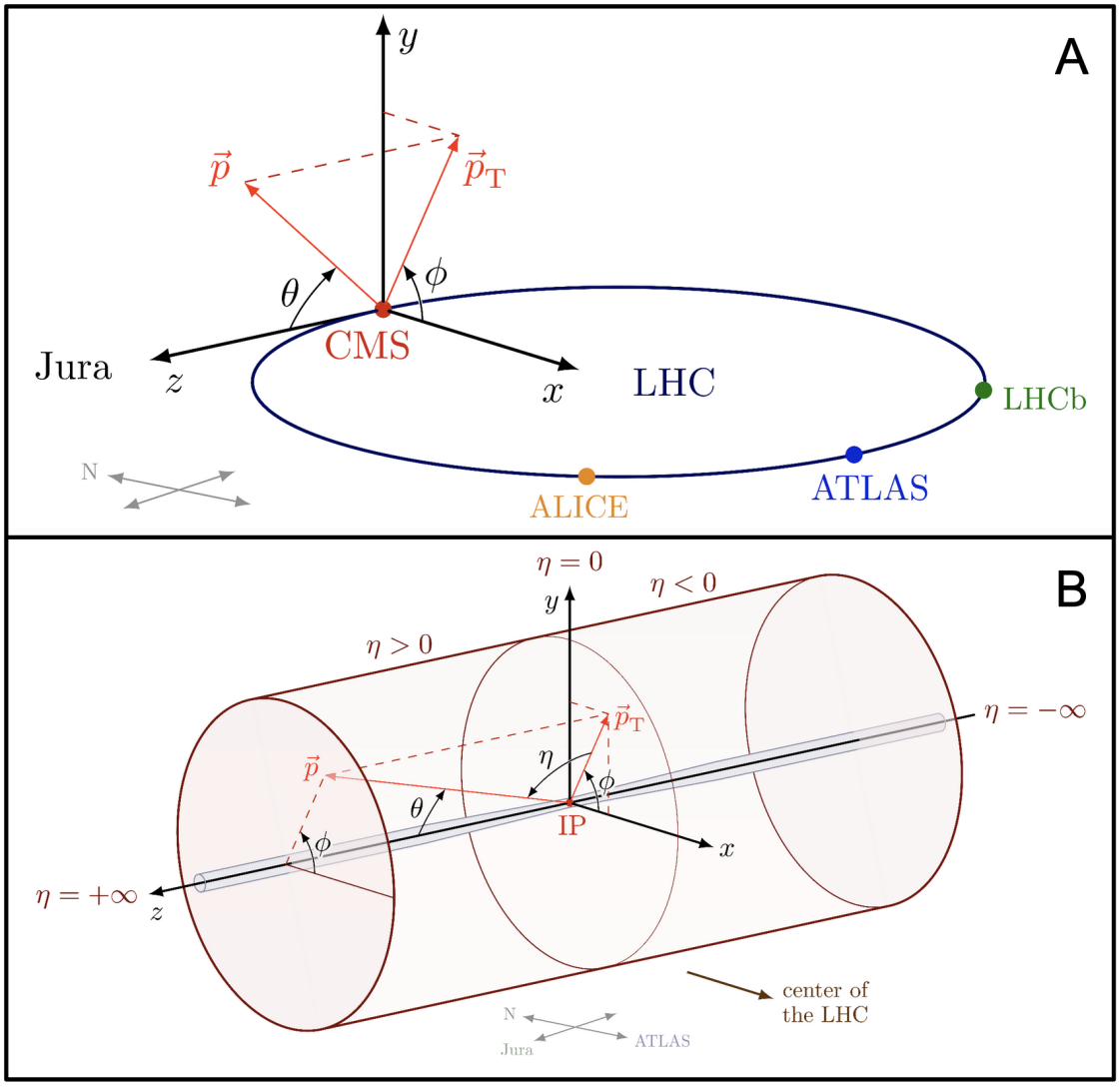


Figure 4.2: *A) Diagram of the CMS coordinate system with respect to the beamline, other detectors, and the Jura mountain range. B) Diagram of the CMS coordinate system in a cylindrical volume overlapping CMS, with component angles and coordinates added to put the usual coordinate system in full context*[47].

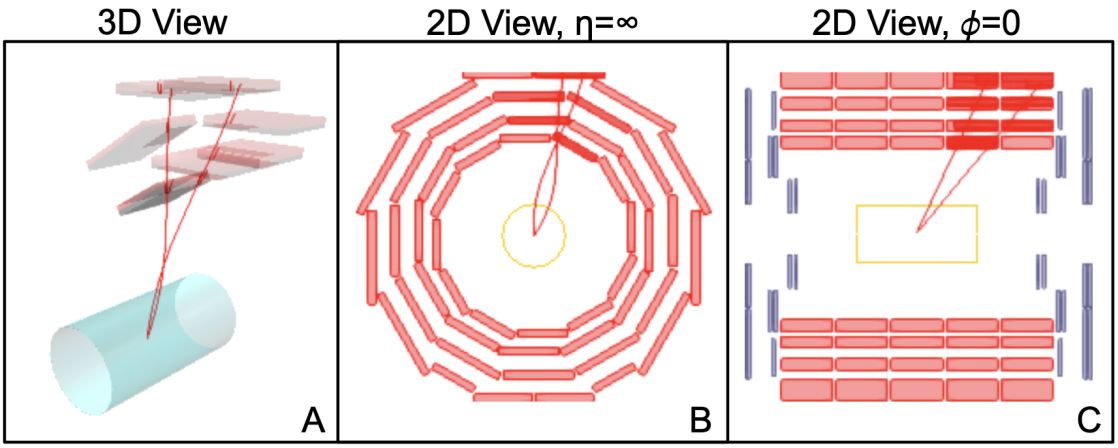


Figure 4.3: *Image made with Fireworks*[49] *and dimuon data from the CERN Open Data Portal*[50]*. A) A 3D image of two muons with track captured in the barrel region and hits detected in muon drift tubes. B) The same image but from the perspective where* η *is set to* ∞*, as though the viewer were standing in the beam line. C) The same image but from the perspective where* ϕ *is set to zero, as though the viewer were standing to the side of CMS, level with its center.*

the history of these publications that similar "boilerplate" text would be used in all papers to refer readers to standardized documentation, with the most relevant components to the analysis elaborated on as necessary. Following that fashion, here is an up-to-date version of that boilerplate text serving as an introduction, followed by elaboration on topics most relevant to this thesis:

The CMS apparatus[52;53] is a multipurpose, nearly hermetic detector, designed to trigger on[54–56] and identify electrons, muons, photons, and (charged and neutral) hadrons[57–59]. Its central feature is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are reconstructed using gas-ionization detectors interleaved with the layers of the steel flux-return yoke outside the solenoid. More detailed descriptions of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Refs. 52;53.

Events of interest are selected using a two-tiered trigger system. The first level (L1), composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of 4.5μs[54]. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to a few kHz before data storage[55;56].

Above, the term "hermetic" is used to mean completely covered, meaning that most particles cannot escape detection. A "scintillator" is a material, usually with crystalline structure, that converts energy into light. The terms "barrel" and "endcap" refer to the body of the cylinder and the flat circular ends perpendicular to the beamline, respectively.

It is useful to know which subdetector contributes to reconstruction of each "physics object." The primary objects are tracks, electrons, photons, muons, jets, b-tagged jets (b-

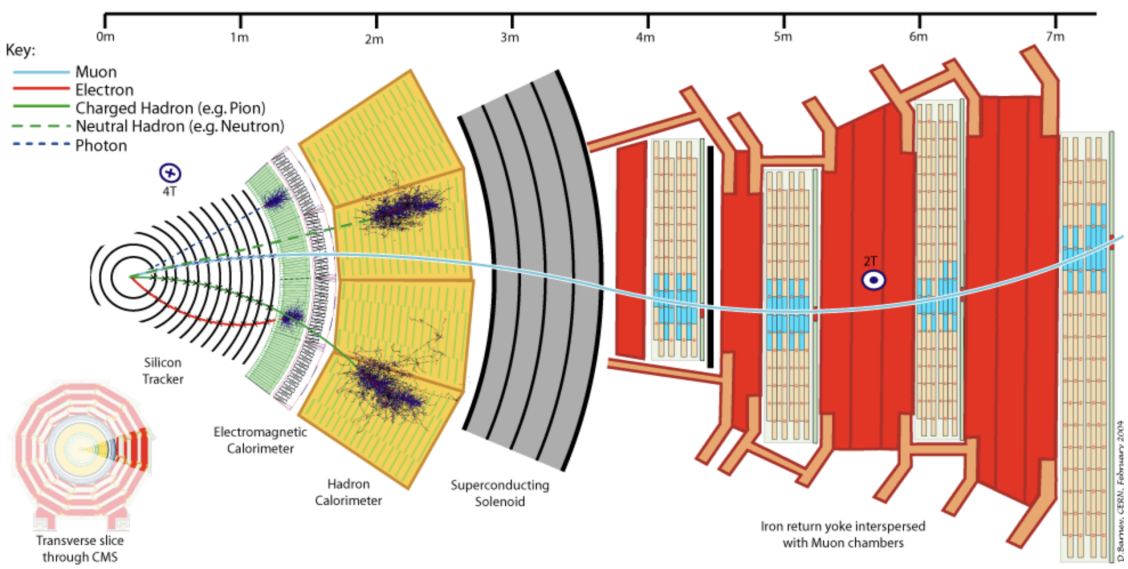


Figure 4.4: *CMS is sometimes referred to as a "cylindrical onion," having many layers used to detect and record measurements of different particle types. It is easiest to discuss the layered subdetectors from the innermost to the outermost: silicon tracker, electromagnetic calorimeter, hadronic calorimeter, magnetic solenoid, and three redundant muon detectors*[60].

jets), hadronic taus (τ_h), and missing transverse energy (MET). Using Fig. 4.4 and moving from the interaction vertex out, all charged particles leave tracks of energy in the tracker. The lightest charged and neutral particles, electrons and photons respectively, are then collected by the electromagnetic calorimeter and prevented from traveling further through the detector. After that, particles that interact with the strong force—both charged and uncharged hadrons—are captured in the hadronic calorimeter. These objects include light mesons like pions, which are used to reconstruct τ_h candidates, b-jets, and generic jets from uds-quark and gluon interactions.[53;55]

At this point, the only particle types that have not been stopped by the detector material are muons and neutrinos. This is because muons are too heavy to be stopped by the material of the electromagnetic calorimeter and too light to decay into even the lightest mesons. Neutrinos are nearly massless and only interact with the weak force, meaning they travel near the speed of light, interact with the detector materials negligibly, and escape detection almost entirely.

The central feature of CMS, the solenoid, is after the hadronic calorimeter. The region

contained by the solenoid has a magnetic field with the opposite sign of the region outside the solenoid, meaning that charged particle tracks will bend in the opposite direction of its initial trajectory. This makes it possible to identify the sign of charged particles, and improves the tracking resolution of muons, which are sampled by cathode strip chambers (CSC's, red in the figure) and drift tubes (DT's, blue in the figure). Muons are not stopped by the detector, and either continue outward in the plane of the LHC, downward into rock where they eventually decay, or upward toward the surface and into Earth's atmosphere.[53;55]

As mentioned, neutrinos are not captured or detected directly by CMS. There can be many neutrinos per interaction, but since they leave no tracks their individual directions are unknown. To determine the energy carried away from a collision by neutrinos, all particle tracks from the primary vertex are collected into a single vector. Since collisions should happen at rest, the missing neutrinos are assigned the opposite direction and magnitude of this vector, hence the term negative vectorial sum. The resulting "missing transverse energy" is abbreviated as MET, from E_T^{miss}.[53;55]

The above slice is of the barrel region of the CMS detector, and slightly different technologies are used for the endcaps of the detector. For example, Gas Electron Multipliers (GEMs), have been used to identify muons in the endcaps since the beginning of Run 3. This is because the endcaps must be significantly more radiation tolerant than the barrel region to capture the same types of particles. The exact cutoff between barrel, endcap, and the transition region between the two depends on the subdetector under consideration. In most cases, $0 < |\eta| < 1.7$ is considered the barrel, $1.7 < |\eta| < 2.3$ is considered the transition region, and $|\eta| > 2.3$ is considered the endcap region. One term that is sometimes used is the "forward region," which refers to high-η events. The different physics objects and their corresponding SM particles are summarized in Fig. 4.5.

4.4 Particle Flow and Hadronic Tau Reconstruction

The Particle Flow (PF) algorithm[61] is used to construct the basic physics objects in each event by combining information from the subdetectors of CMS as outlined above. The key

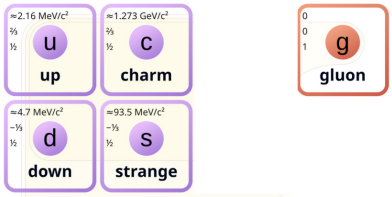
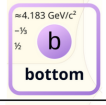
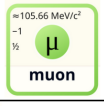
Standard Model Particle	Physics Object at CMS	Subdetector and explanation
	Electrons and photons	**ECAL + tracker** Photons do not produce tracks, allowing their separation from electrons
	Light jets or jets	**HCAL + tracker** Jets from light quarks have slightly different hadronization cone extents and activity than jets from bottoms or taus. In recent years, charm jet tagging has become possible at the software level.
	b-tagged jets or b jets	**HCAL + tracker** b quarks have a longer lifetime than light quarks, causing a displaced decay vertex, and resulting in their distinguishability from other objects
	Hadronic taus	**ECAL + HCAL + tracker** hadronic taus tend to have narrower and more isolated hadronic showers than normal jets, and have characteristic decay modes
	Muons	**DT + GEM + CSC + RPC + tracker** Muons are detected with multiple subdetectors to achieve high resolution
	MET	Neutrinos are not detected individually, but lack of energy conservation in the detector indicates their presence

Figure 4.5: *A table showing fundamental particles, their corresponding physics object at CMS, and a short description of subdetectors used for their reconstruction. In some sense, all subdetectors are used for all objects due to the Particle Flow algorithm progressively pruning events.*

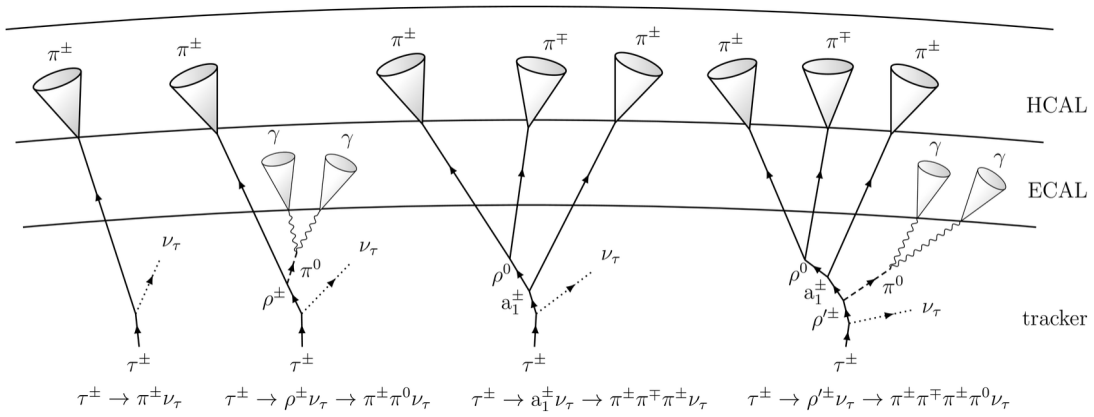


Figure 4.6: *A diagram showing (from bottom to top) a physics process, an initial tau lepton, intermediate resonance decay if applicable, and subsequent decay with relevant objects collected in the ECAL and HCAL for the 4 most probable hadronic tau decay modes*[67]*.*

element for jet reconstruction is the infrared and collinear safe anti-k_T algorithm[62;63], using a distance parameter of R=0.4, resulting in so called "AK4" jets. The jet momentum is the sum of its constituent particle momenta. In Run 3, the pileup per particle identification algorithm (PUPPI)[64;65] is used to reduce the contribution of extra tracks and calorimeter energy deposits from adjacent bunch crossings.

As noted in Chapter 2, taus can decay leptonically or hadronically. Electrons and muons decaying from taus cannot be reliably distinguished from electrons and muons from prompt electrons and muons, but hadronic jets from tau lepton decays can. Hadronic tau decays primarily proceed through a small number of mesons (see Table 2.2) resulting in characteristic and distinguishable decays. Because of these meson resonances, τ_h candidates tend to have more isolated jet contents than jets from light quark hadronization. This information is used by the Hadron Plus Strips (HPS) reconstruction algorithm[66], which combines 1 or 3 tracks with energy deposits in the calorimeters to identify tau decay modes. Both ECAL and HCAL are used because neutral pions (π^0) rapidly decay to a pair of photons captured in the ECAL, while heavier mesons are captured in the HCAL as shown in Fig. 4.6.

After reconstruction with HPS a deep neural network based multiclassifier named DeepTau[68] is used to refine the outputs and provide an overall identification (ID) score for how likely a τ_h candidate is to be genuine hadronic tau decay, an electron, a muon, or a light

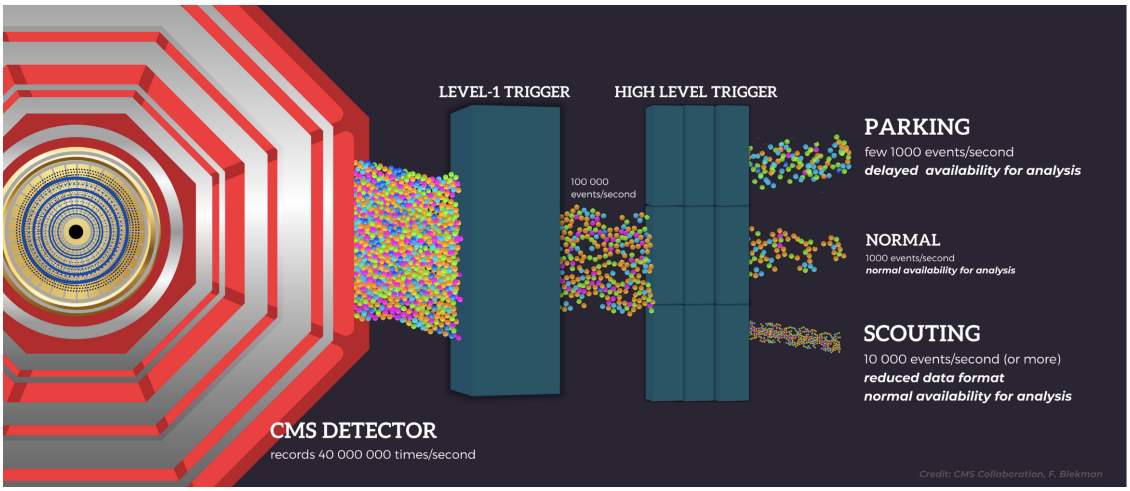


Figure 4.7: *A graphic illustrating the input rate of events to the two-tiered trigger system as well as the output datastreams*[73].

quark/pileup jet. Other universal jet tagging and identification algorithms and networks are also used in some analyses, such as ParticleNet or Universal Particle Tagger[69–71].

More detailed information about τ_h reconstruction can be found in Ref. 12, written and published in the course of completing this thesis.

4.5 Trigger System

Not all particle interactions are interesting for modern analyses. Many are well-understood thanks to the long history of particle physics experiments and collaborations, and the success of the SM. The challenge of working with the LHC is keeping up with the collision rate, and storing only the most promising events for further analysis. The LHC has a maximum collision rate of 40 MHz, which in practice is near 27 MHz due to empty bunch crossings. If all data could be stored at full resolution, this would be as much as 40 terrabytes per second (TB/s)[72], but CMS is limited by an event-building rate of 100 GB/s[53].

To cope with the high event rate and practical readout and storage limitations, a two-tiered trigger system was designed to bridge these 4 orders of magnitude, as shown in Fig. 4.7.

In recent years CMS has made a new type of datastream called "scouting data" that uses an ultra-minimal amount of event information, resulting in a large number of events suitable

for prospective studies on physics not captured by the normal trigger system. This analysis does not make use of scouting data, but does use "parked data," which is captured by the normal trigger system but processed from weeks to months after initial collection. Normal trigger data is processed with reconstruction algorithms immediately after collection.

4.6 Level 1 Trigger

The first step of the two-tiered trigger system is called the Level 1 (L1) trigger. The L1 trigger is hardware-based, loaded directly to field-programmable gate arrays (FPGAs) in a counting room 65 meters away from the CMS detector[74]. It has 4.5 microseconds of buffer time per event to decide whether it will be kept for further analysis. If an event is not stored, it is lost forever.

The L1 was designed to only use calorimeter and muon subdetectors. Even without information from the silicon pixel tracker, the L1 can build objects such as muons, jets, hadronic taus, and merged electron/photon candidates. With these, the L1 calculates simple counts, energy sums, correlated angular quantities, and isolation quantities.

These quantities are compared against a list of basic event requirements, which is dynamic and determined by the physics needs of the experiment. An example of an L1 trigger is "SingleMu_27," which requires one muon with $p_{\mathrm{T}} > 27\,\mathrm{GeV}$ and no restrictions on $|\eta|$. A more complicated L1 is "DoubleJet_120_er4p7," which requires two separate energy deposits in the hadronic calorimeter, both with $p_{\mathrm{T}} > 120 GeV$ and $|\eta| < 4.7$. For this analysis "L1_DoubleJet35_Mass_Min450_IsoTau45_RmOvlp" was built, which requires two jets with a minimum combined invariant mass of 450 GeV, one hadronic tau, and for all three objects to be well-separated as determined by the overlap removal algorithm "RmOvlp." It should be noted that a τ_{h} candidate at the L1 level is little more than a jet with loose isolation requirements, roughly requiring that the energy of an event is focused in one location of η–ϕ space, whereas jets may be more dispersed[12;55;56].

If an event satisfies the requirements of one L1 trigger it is forwarded to the High Level Trigger (HLT) for additional scrutiny[?] .

4.7 High Level Trigger

An HLT "path" is a sequence of algorithms and selections that performs a simplified event reconstruction during data taking and targets a physics process. There are about 800 different HLT paths in a proton-proton "menu" for Run 3 data taking. In the context of triggers, the term "online" is used to delineate between quantities calculated by trigger algorithms, as opposed to "offline" quantities calculated by full reconstruction algorithms after an event has been accepted for storage.

The input rate from L1 into the HLT is 110 kHz in Run 3, with a read out event rate of 5 kHz across multiple datastreams. Due to the lowered input rate, the HLT has comparatively much more time to make a decision, with an average event processing time of 500 milliseconds. This is practically an eternity compared to the amount of time allowed for an event at L1. If the L1 decision time is scaled up to one day, then the HLT has 27 years to make a decision. Subsequently, algorithms similar in complexity to offline reconstruction can be performed at the HLT, including object tagging with machine learning tools. For example, the L1 mentioned in the previous section was used to construct the HLT "HLT_DoublePFJets40_Mass500_MediumDeepTau45_L2NN_MediumDeepTau20_eta2p1" in 2022, which requires two jets, and two taus passing neural networks called L2NN and DeepTau, both of which were designed to reduce the background rate of jets misidentified as hadronic taus[12].

Modules which are run for many triggers are grouped into a sequence, which is run once during event processing and the outputs are used by all relevant HLTs. The paths must be designed with care to reuse objects where possible, and balance physics goals against available resources. Once an event is accepted by HLT, it is sent to the Tier 0 data center on the CERN campus.

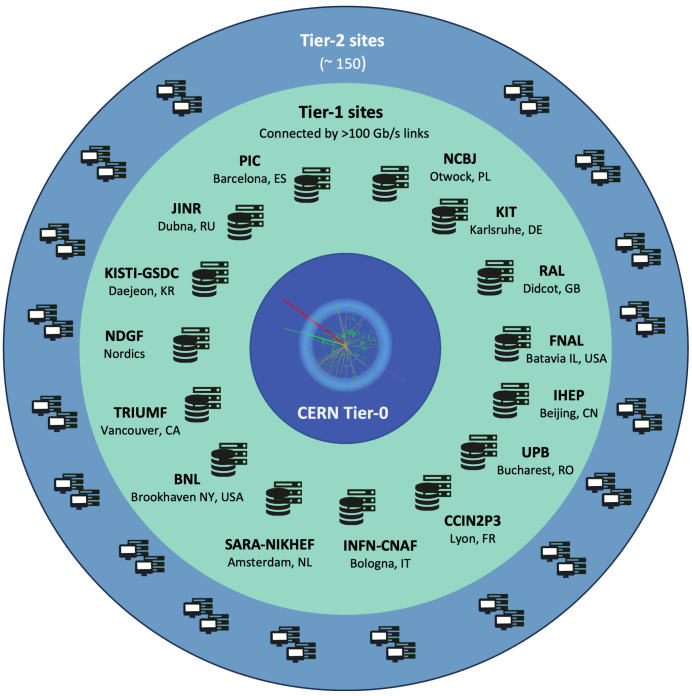


Figure 4.8: *Different processing tiers on the Worldwide LHC Computing Grid (WLCG). Tier 3 resources are not shown because they are not directly involved with WLCG operations*[76].

4.8 Event Processing, Software, and Simulations

Once data is collected it is stored permanently to magnetic tape at the Tier 0, then distributed to Tier 1 sites for event reconstruction and storage, as shown in Fig. 4.8. Tier 2 sites use Tier 1 files for analysis specific reprocessing or planned re-reconstruction campaigns. Analysis groups use these files to study physics processes, typically running additional skimming and processing algorithms with local resources at regional Tier 3 sites supported by universities[75;76].

All these computing resources are necessary because of the size and complexity of data collected by modern particle physics experiments. Fig. 4.9 shows how busy an event at CMS can be, with numerous tracks and calorimeter hits. To properly combine this information in a reproducible and shareable way, standardized algorithms have been built over time, and saved to the CMS Software library (CMSSW)[77]. This library provides modules and tools for reconstruction at CMS, including standardized workflows run on all tiers of computing.

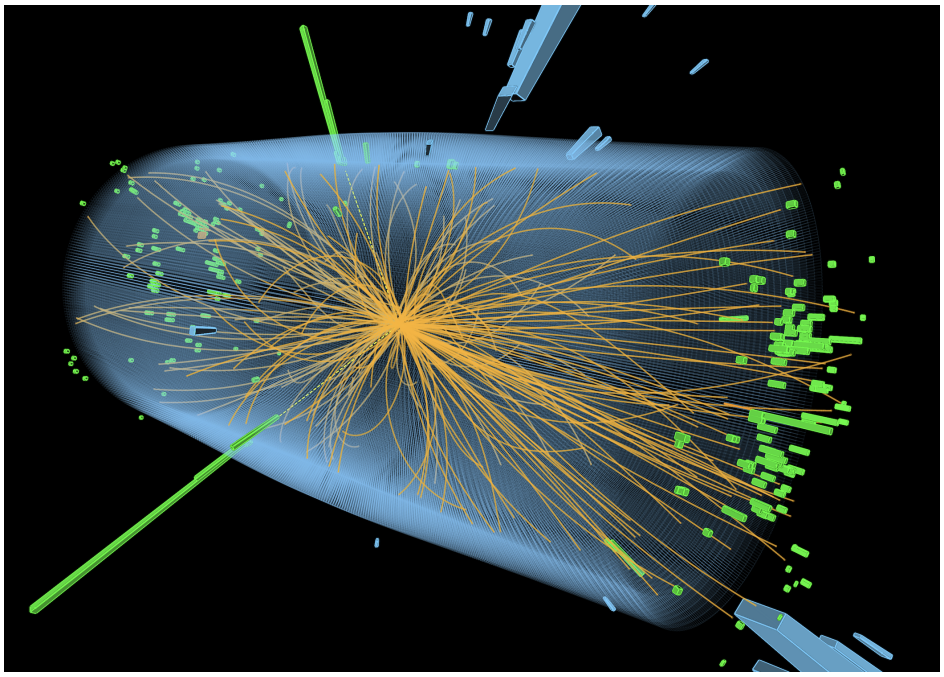


Figure 4.9: *Event display of data, used in this context to show the high multiplicity of tracks and detector hits*[78].

It is versioned to continue working with changing detector conditions and reconstruction algorithm improvements. Typically initial data processing is performed with a frozen version of CMSSW as soon as the data is collected, then reprocessing campaigns are undertaken at a later time.

In addition to collected data, numerous simulated datasets are produced. Each simulated dataset is produced for a specific physics process using Monte Carlo (MC) algorithms to seed initial conditions within a specified range of parameters. These samples are referred to as simulation or MC. Three key components of simulated samples are their generator, their hadronization framework, and a rendering of the detector. Some currently popular generators include aMCNLO, POWHEG, and JHUGEN, which have different approaches to modeling particular processes[79–81]. The most common hadronization framework used at CMS is PYTHIA, but HERWIG and SHERPA are also used in some contexts[82–84]. The detector material and layout are always interfaced with GEANT4, precisely modeling component positions and thicknesses[85].

Simulating particle interactions allows physicists a chance to know truth-level information

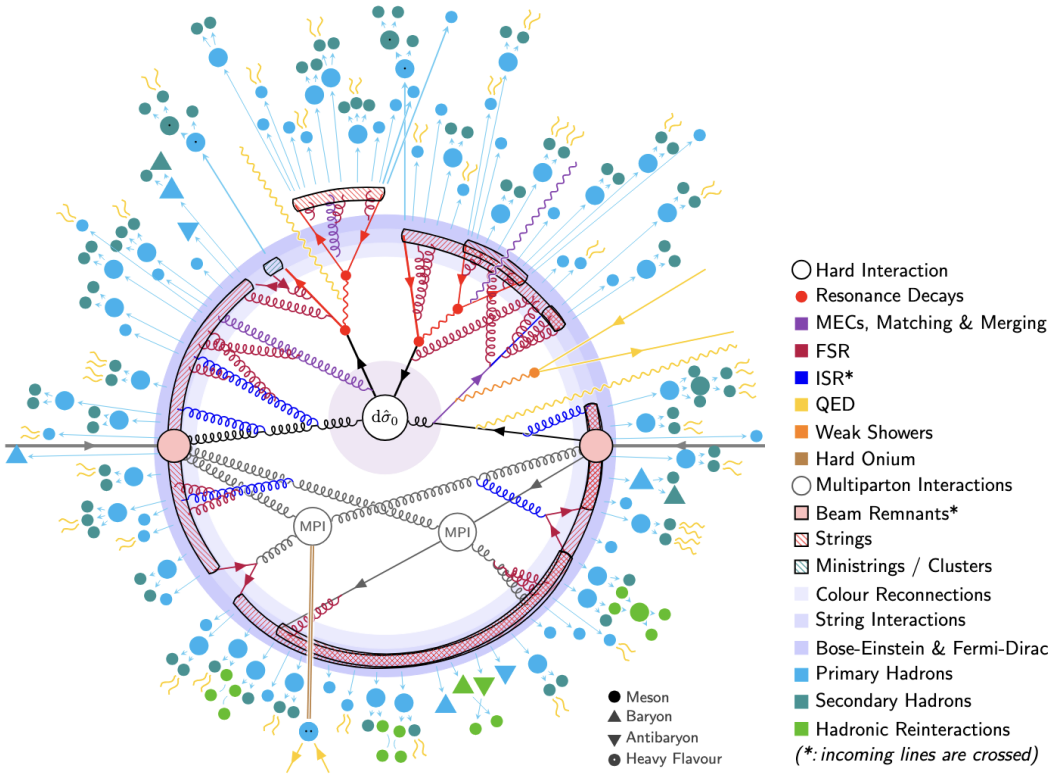


Figure 4.10: *Simulated particle interaction and subsequent decays of a proton-proton to $t\bar{t}$ event, modeled by PYTHIA* [82].

about a single collision. In real data, it is not possible to know intermediate steps of a particle decay. Simulating many possible interactions allows us to arrive at statistically supported scenarios. Fig. 4.10 is included to emphasize the complexity of particle provenance, only possible with simulation.

Chapter 5

The Higgs to Tau Tau Differential Run 3 Analysis

5.1 Motivating Higgs Boson Studies

Being the newest accessible fundamental particle, the field naturally has the most to learn from the Higgs boson. The Higgs boson is connected to all other massive particles, and a key ingredient in electroweak symmetry breaking. Without the Higgs field, there would be no weak mixing of massive bosons, and we would not live in the universe as we know it. The Higgs boson offers so many ways to study physics that were unreachable before its discovery that it is sometimes called "the Higgs portal" in presentations and media for the public.

The major studies of the Higgs boson can be broken into 3 categories, with three different questions:

1. Does the Higgs boson decay with the proportions predicted by the SM? Are there as many decays of $H \to ZZ$ and others as there should be?

2. Do observed quantum numbers match predicted values? Is the Higgs boson polarized, does it violate charge-parity conservation, and does it violate flavor conservation?

3. What is the shape of the Higgs potential? Is there more than one Higgs boson, and

does the Higgs boson interact with itself as predicted?

In the last decade, most studies have been dedicated to questions of type one and two. There are still studies underway for rare couplings of the Higgs boson and improved precision of the same measurements, but most properties have been validated to agree with the SM, or, at the very least, do not disagree enough to be noticeable. Significant research has been dedicated to the di-Higgs or self-coupling interaction in the $HH \to b\bar{b}\tau^-\tau^+$ channel, which may have enough data to be measured in Run 3, and will be measured extensively in Run 4 (2030-2036).

It is an exciting time to be a particle physicist. New analysis techniques, improved reconstruction algorithms, and better trigger strategies can all improve physics results, and analysts are pushing data to the limits of the SM's current knowledge. By improving precision of these measurements the number of alternative hypotheses that fit the data shrink, and the remaining problems of the SM—including naturalness, mass hierarchy, baryon asymmetry—become more apparent. This is why people study the Higgs boson, because it takes us to the edge of the SM, and may even offer answers to the longest standing problems in the field.

5.2 Higgs Boson Status and Measurement Approaches

The most basic type of particle physics experiment is called a counting experiment, where the analysts count the number of times a type of event occurred in a dataset[86]. This count is converted to a cross section using the fundamental counting equation (Eq. 3.1) and the luminosity of the dataset. This cross section can then be compared to the SM prediction. Often, for ease of communication between experiments, the measured cross section is normalized to the SM value and the ratio is reported as the "signal strength." The primary difficulty of a counting experiment is determining which events to count as signal and which to count as background.

As discussed in Chapter 2, all massive particles interact with the Higgs field, meaning all massive particles have interactions with the Higgs boson in the SM. Each type of Higgs

Table 5.1: *Branching ratios (BRs) of the Higgs boson. The "status" column is listed as "Observed" if the decay has 5-sigma significance or "Evidence" for 3-sigma significance. $H \to gg$ is listed as "Indirect Evidence" because the inverse process, gluon-gluon fusion, has been observed and it would be almost impossible to control backgrounds to observe $H \to gg$ directly. The rare decays $H \to c\bar{c}$ and $H \to Z\gamma$ are listed as "Limits Set" because they are predicted by the SM, but so far only enough data has been collected to measure an upper limit of their production. The total of percentages reported in the table is 97.3%, meaning 2.7% is left to other rarer decays*[88].

Decay	BR	Status
$b\bar{b}$	58.2%	Observed
W^-W^+	21.4%	Observed
gg	8.2%	Indirect Evidence
$\tau^-\tau^+$	6.3%	Observed
$c\bar{c}$	2.9%	Limits Set
ZZ	2.6%	Observed
$\gamma\gamma$	0.23%	Observed
$Z\gamma$	0.15%	Limits Set
$\mu^-\mu^+$	0.029%	Evidence

boson decay is called a "channel"[87]. The original Higgs boson discovery was from a combination of 5 channels, $H \to \gamma\gamma$, $H \to ZZ$, $H \to W^-W^+$, $H \to \tau^-\tau^+$, and $H \to b\bar{b}$, with $H \to \gamma\gamma$ and $H \to ZZ$ being the most precise and easiest to capture with CMS[7;8]. After the initial discovery from combined measurements and due celebration, analysts refined their methods and collected more data to produce single-channel measurements of the Higgs boson. Since its discovery, parameters of the Higgs boson have continually been measured to determine if the observed particle is consistent with SM predictions, as shown in Table 5.1.

Fig. 5.1 neatly summarizes the state of observed Higgs boson decays, including the single-channel measurements of channels mentioned previously, a measurement of $H \to \mu^-\mu^+$, and the coupling of the Higgs boson to the top quark, all of which show agreement with the SM[89]. Note that the Higgs boson cannot decay directly to a top quark pair because the top quark is more massive than the Higgs boson.

Rare decays, such as $H \to c\bar{c}$ and $H \to Z\gamma$ are being searched for, and are expected to be observed within the lifetime of the LHC. Besides decays, the production mechanism of the Higgs boson at the LHC is another handle used to judge consistency with the SM. The primary production modes are gluon-gluon fusion (ggF or ggH), vector boson fusion (VBF),

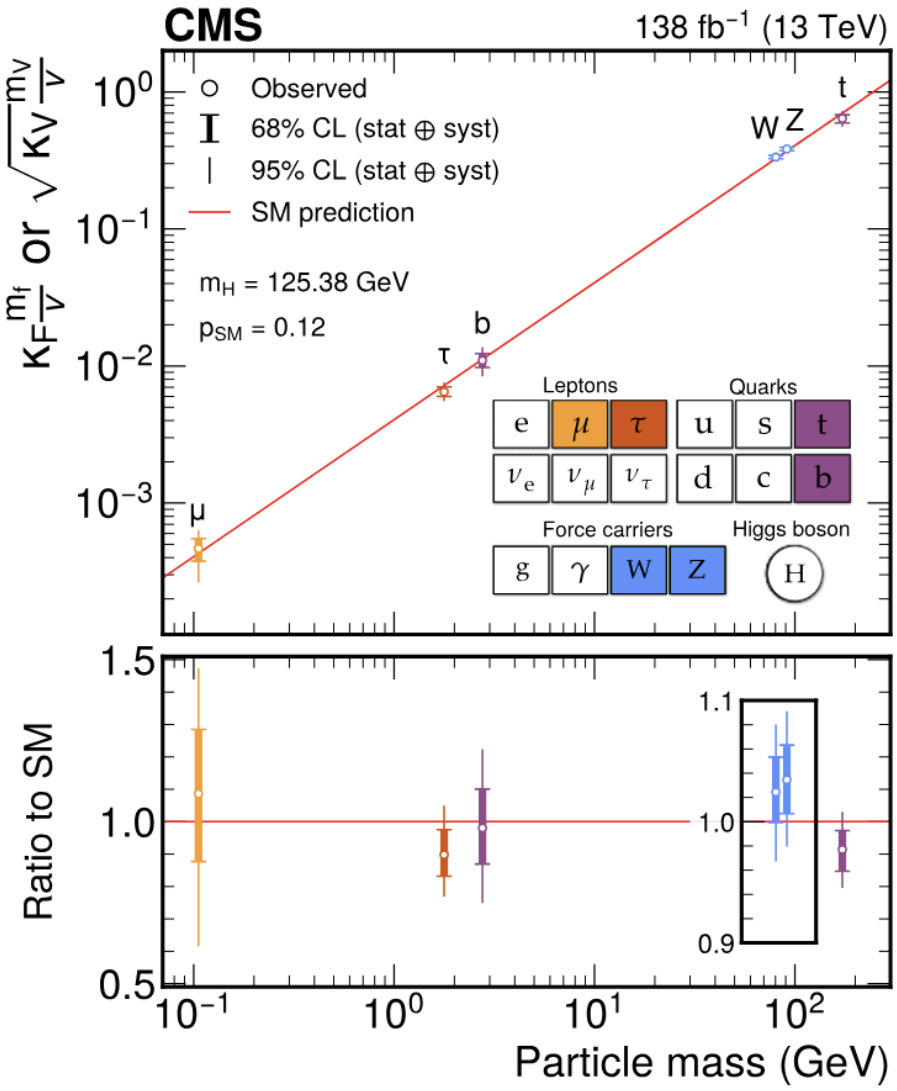


Figure 5.1: *A graph of couplings vs particle mass, with a diagonal drawn to show the SM prediction. The* $H \to \gamma\gamma$ *decay is not shown on the graph because* γ*'s are massless and couple to the Higgs boson indirectly through virtual W bosons in next-to-leading order interactions*[89].

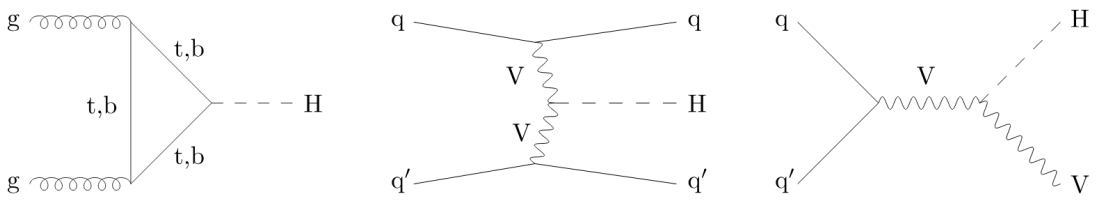


Figure 5.2: *Feynman diagrams for the three leading production modes of the Higgs boson. In VBF production two quarks are present in the final state, resulting in two high energy back-to-back jets*[89].

Table 5.2: *Cross section (XS) values for signal mode production of a 125.09 GeV Higgs boson*[90]*. Note that the cross section of ggH has increased by 7.5% from the Run 2 to Run 3 collision energy.*

Production Mode	XS (13 TeV) [pb]	XS (13.6 TeV) [pb]	% of Total (13.6 TeV)
ggH	48.1	51.7	86.9%
VBF	3.8	4.1	7.0%
VH	2.3	2.4	4.1%
Other	1.2	1.3	2.0%
Total	55.4	59.5	-

associated vector boson production (VH), and associated production with top and bottom quarks (ttH, tH, bbH, or bH)[87]. Production from ggH is dominant in most channels, followed by VBF, VH, and then all other associated production modes, as shown in Table 5.2.

The production mode is used to further subcategorize Higgs boson measurements, making SM measurements more precise. To improve translatability of results between experiments and maximize physics reach while minimizing model dependence, the Simplified Template Cross Section (STXS) framework was made[91], as shown in Fig. 5.3.

Key variables used in the STXS framework are the number of jets (N_{jets}) in the event, the Higgs boson transverse momentum (p_T^{H}), and the dijet invariant mass (m_{jj}). The number of jets is self-explanatory, though it should be noted that VBF events are expected to be reconstructed with 2 jets at a minimum. Since the Higgs boson is not measured directly, p_T^{H} is the combined transverse momentum of its identified decay products. A value of p_T^{H} greater than 200 GeV is typically considered "boosted" meaning it recoiled against a jet in the event, causing the Higgs boson decay products to be collimated, appearing closer together in the detector. These boosted signal events are more easily distinguishable from most types

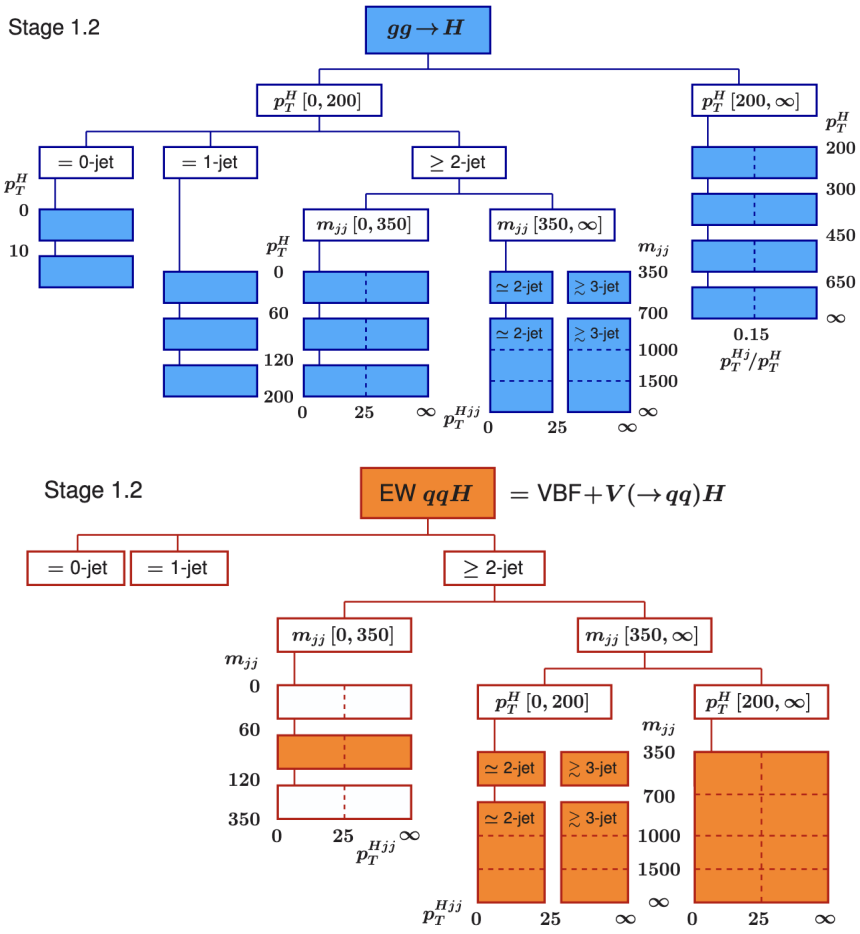


Figure 5.3: *The STXS framework for ggH and VBF production*[91].

of backgrounds. Similarly, in VBF events m_{jj} is determined by combining hadronic jets, with higher m_{jj} values resulting in more distinguishable events. These variables are used together to define bins of the STXS framework, each of which is effectively its own counting experiment, albeit with shared uncertainties with other bins.

Instead of using STXS to combine multiple counting experiments, a differential measurement of a single variable can be made of all signal processes. This is more model-independent because no assumptions are made about the production mechanism. It is simply a comparison of available data to an inclusive signal prediction.

5.3 Higgs to Tau Tau

The recent history of the $\mathrm{H} \to \tau\tau$ channel measurements is shown in Fig. 5.4. After the discovery of the Higgs boson in 2012 with Run 1 data[92], $\mathrm{H} \to \tau\tau$ was observed in 2017 with partial Run 2 data[93], followed by a differential cross section measurement in 2021 with full Run 2 data[94]. This was followed a short time later with a STXS measurement, as well as detailed analyses of charge-parity violation in Higgs boson decays and anomalous contributions in Higgs boson production, all with full Run 2[95–97]. In all studies, $\mathrm{H} \to \tau\tau$ has been found to agree with SM predictions. It is notable that of the Higgs boson channels, $\mathrm{H} \to \tau\tau$ has the highest precision measurements in VBF production, followed closely by $\mathrm{H} \to \mathrm{WW}$. Most recently, in 2026, the $\mathrm{H} \to \tau\tau$ channel has set the strongest limits on charge-parity violation to date by combining Run 2 and early Run 3 measurements[98].

The remainder of this chapter describes the ongoing Run 3 $\mathrm{H} \to \tau\tau$ differential measurement. The analysis measures the Higgs boson cross section as a function of p_T^{H}, N_{jets}, leading jet transverse momentum (p_{T}^{j}), m_{jj}, dijet eta difference ($\Delta\eta_{\mathrm{jj}}$), and dijet phi difference ($\Delta\phi_{\mathrm{jj}}$). These variables were chosen because they are sensitive to properties of the Higgs boson, and can be combined with the results of other channels. A more precise measurement than Run 2 is produced by exploiting detector improvements, additional data, and updated event triggers.

Since the Higgs boson can decay into two tau leptons, and the tau itself can decay

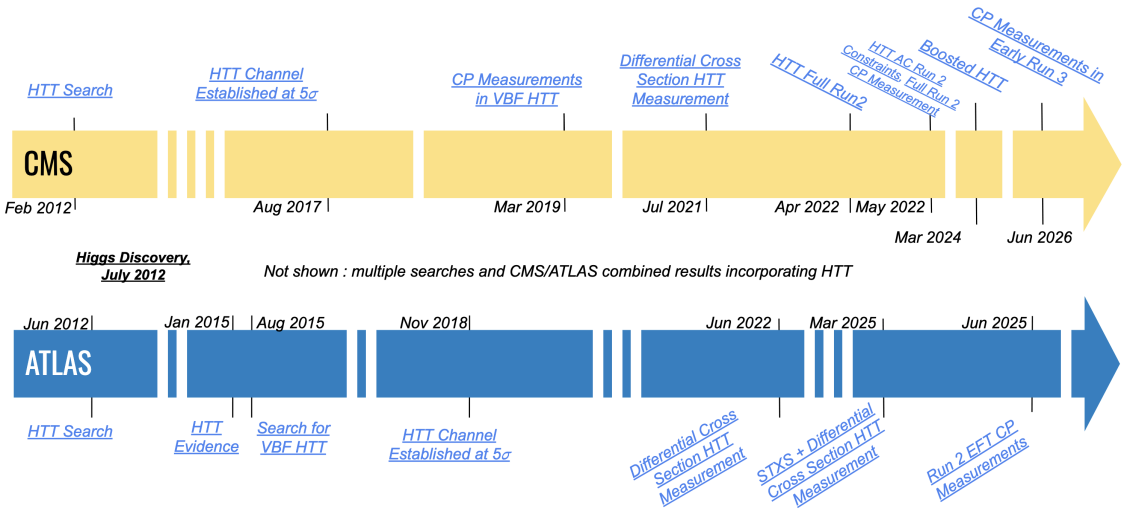


Figure 5.4: *Timelines of measurements for Higgs to tau tau decays at the CMS*[92–99] *and ATLAS*[100–106] *experiments.*

leptonically or hadronically, there are a total of six possible final states: $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$, $e\mu$, $\mu\mu$, and ee. Only the first four are used in this analysis because the $\mu\mu$ and ee have small branching fractions and dominant backgrounds from copious Z decays. In Fig. 5.5, one can see the relative abundance of each possible final state pair. Additionally, this figure illustrates the dominance of the $\tau_h\tau_h$ final state. Any change to τ_h candidate reconstruction affects three of the four final states, and affects the $\tau_h\tau_h$ final state twice as much as the $\mu\tau_h$ and $e\tau_h$ final states. This thesis focuses on those three final states because the $e\mu$ channel was primarily studied by another collaborator.

It is instructive also to show the final states of the mainline Higgs boson analysis channels at CMS and their corresponding number of neutrinos, as seen in Table 5.3, which illustrates the challenge of missing energy in the $H \to \tau\tau$ channel.

Combining the information of Fig. 5.5 and Table 5.3, one can see that $\tau_h\tau_h$ final state has approximately the same branching fraction as $\mu\tau_h$ and $e\tau_h$ combined. The $\tau_h\tau_h$ final state also has the largest number of neutrinos present in the event. This makes it the most challenging final state of the analysis due to the lower object resolution from missing energy, but potentially the most useful due to the abundance of events.

In the time between the Run 2 version of this analysis and now, there have been several

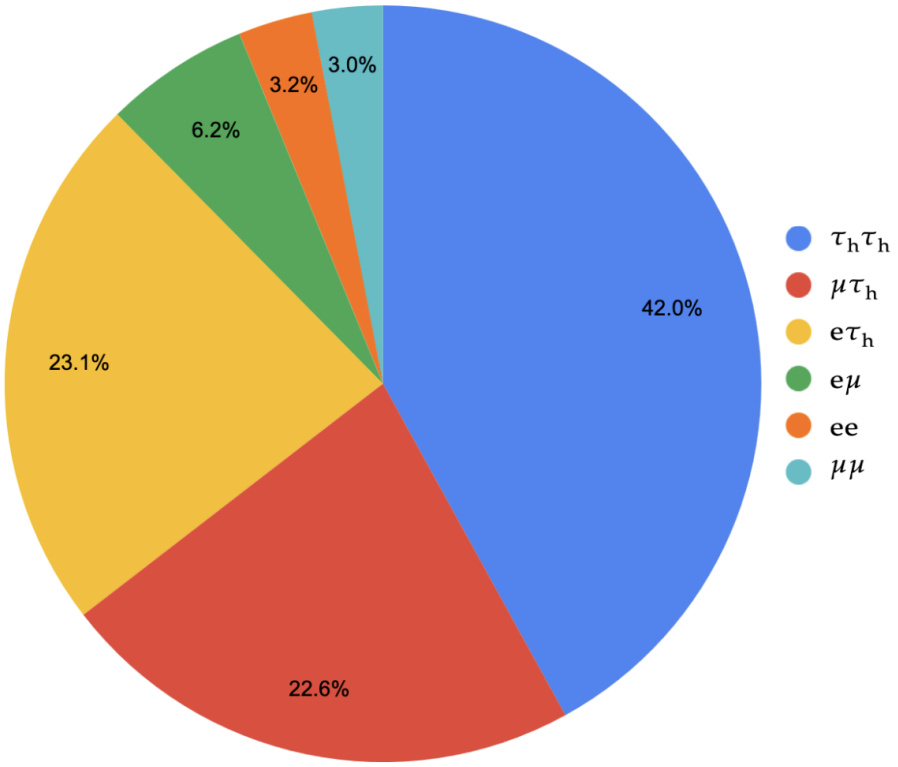


Figure 5.5: *This chart shows the theoretical branching ratio for each possible $\tau\tau$ decay.*

Table 5.3: *Higgs boson channels with a differential measurement, considered final states, and corresponding number of neutrinos. Of the 5 channels below,* $H \to \tau\tau$ *and* $H \to b\bar{b}$ *are the only ones published at CMS with hadronic reconstruction.*

Channel	Final States	nNeutrinos
$H \to \tau\tau$	$\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$, $e\mu$	4, 3, 3, 2
$H \to b\bar{b}$	2b + 0, 1, or 2 ℓ	0, 1, 2
H $\to$ ZZ (Leptonic)	4μ, 4e, 2e2μ	0, 0, 0
H $\to$ WW (Leptonic)	1e1μ, 2e, 2μ	2, 2, 2
H $\to \gamma\gamma$ (Leptonic)	$\gamma\gamma$	0

changes online and offline that have improved the overall object quality for every analysis using hadronic objects.

- DeepTau neural network improvement from v2p1 to v2p5[107]
- Improved jet and MET objects using PUPPI[108]

Additionally, I was involved in changes to the $\tau_h\tau_h$ trigger strategy of $H \to \tau\tau$ analyses specifically. In addition to this differential analysis, these changes also benefit other $H \to \tau\tau$-

like analyses by improving the trigger efficiency and enlarging the space of event types captured.

- Online trigger neural network integration (DeepTauHLT, and ParticleNet@HLT)[12], described in more detail in Chapter 6
- Increased trigger phase space coverage (DiTau+Jet, VBF+SingleTau), described in Sec. 5.3.2

Two changes were made that simplified this Run 3 analysis with respect to the previous version.

- Inclusion of the 0j category in the $\tau_h\tau_h$ final state
- Using the FastMTT[109] algorithm for p_T^H calculation instead of correcting the p_T^H variable, described in Sec. 5.3.8

5.3.1 Datasets and Simulation

The analysis uses data collected in 2022, 2023 and 2024, corresponding to 34.7, 27.8 and 109.1 fb^{-1} respectively, resulting in a total integrated luminosity of 171.6 fb^{-1}[110–113]. After collection the data were centrally processed by CMS into the NanoAOD file format[114]. Each of the final states used has a different datastream associated with it: $\tau_h\tau_h$ uses the Tau dataset, $\mu\tau_h$ uses the SingleMuon dataset, and $e\tau_h$ uses the EGamma dataset. Each datastream is defined by a primary HLT, and several cross triggers are also used in the analysis, discussed in the next section.

The signal processes, ggH, VBF, VH, and ttH, were generated with the POWHEGv2 generator[80;115–117] at next-to-leading-order (NLO) precision with additional corrections from the MINLO generator for VH processes[118]. Event filters are enforced at the generator level that reject events with p_T below trigger thresholds, allowing for the enhancement of the number of events produced in the sample and reduction of statistical error. The ttH signal sample is used both as a signal and background, as some events in the sample contribute to genuine $H \to \tau\tau$ signal, and the cross section is split accordingly.

The major backgrounds involving $\tau\tau$ production include Z boson and photon production with associated jets ($Z/\gamma + \text{jets}$), quantum chromodynamic (QCD) jets misidentified as τ_h candidates, W boson production with associated jets ($W + \text{jets}$), $t\bar{t}$ production (TT), single top production with associated jets (ST), diboson production of WW, ZZ, WZ (VV, collectively), and $H \rightarrow WW$ production where the W bosons decay to taus. In all channels, $Z/\gamma + \text{jets}$ is a dominant background and QCD and $W + \text{jets}$ are combined to form the "Jet Fakes" background described in more detail below. Other processes are minor backgrounds, though TT, ST, and VV become relevant in events with 2 or more jets as they are genuine sources of hadronic tau decays that are not centered at the Z or W boson resonance mass.

All simulated events were hadronized using PYTHIA8.3[119;120], using the CP5 tune parton showering and hadronization model derived on CMS data. Interactions with the CMS detector are simulated with a digital model in GEANT4[85;121;122].

5.3.2 Trigger Strategy

Hadronic objects like τ_h's can be difficult to separate from other hadronic backgrounds such as light jets and pileup jets. The primary distinction between τ_h's and other hadronic objects is the amount of activity in a narrow cone around the decaying τ lepton's original path. Hadronic decay products from a τ lepton tend to be closer together and are said to be "more isolated" whereas hadronic decay products from light jets are more diffuse. Still, the difference is subtle, and can be missed in the busy detector environment.

Subsequently, triggers for τ_h objects have relatively high p_T thresholds. For example, triggers to capture a single isolated muon at CMS start at $p_T = 24$ GeV, whereas a trigger for a single isolated τ_h starts at $p_T = 180$ GeV. To capture a usable number of τ_h's, two objects are required at the same time, resulting in the DiTau trigger with a still relatively high threshold of $p_T = 35$ GeV for both τ_h's. This online threshold is 5 GeV lower than the regional isolation based-triggers of Run 2, due to the integration of a neural network tagger in the HLT[12]. In 2024, this threshold was lowered another 5 GeV for all online τ_h candidates by an improved online neural network, discussed in more detail in Chapter 6.

In addition to this new online tagger improving τ_h identification generally, we included two new cross triggers for $\tau_h\tau_h$ in Run 3. The first is the DiTau+Jet trigger, which can require slightly softer τ_h's of 30 GeV by simultaneously requiring a high-p_T jet of 60 GeV. This 3-object combination results in a trigger that captures fewer events overall compared to the main $\tau_h\tau_h$ trigger, but is able to capture events inaccessible to the main trigger.

Another trigger we used, which I implemented with the direction of Olivier Davignon and other trigger experts at CMS, was the VBF+DiTau trigger. Similar to DiTau+Jet, the idea is to add high-p_T jets to reduce the total trigger rate, and then lower the online τ_h thresholds, accepting events that are otherwise missed by the primary trigger. The VBF+DiTau trigger initially required two jets with $p_T > 40$ GeV, an $m_{jj} > 500$ GeV, one τ_h with $p_T > 45$ GeV, and another with $p_T > 20$ GeV. Within a year of its implementation, the online requirement of the second τ_h candidate was removed, changing the VBF+DiTau trigger into the VBF+Tau trigger. This trigger in particular resulted in a 25% improvement in signal strength for the VBF STXS bins with $m_{jj} > 600$ GeV.

These changes to the trigger strategy significantly improve the $\tau_h\tau_h$ channel of the analysis, which is trigger-limited compared to other final states. In $\mu\tau_h$ the majority of events come from the SingleMuon trigger, meaning the trigger improvements only affect the 15% of events which come from the $\mu + \tau$ cross-trigger. In contrast, since all τ_h candidates in $\tau_h\tau_h$ must pass the trigger first, any improvement at that level directly improves event acceptance.

5.3.3 Event Selection

Events must pass a set of complementary HLTs and basic kinematic selection offline depending on the final state. The trigger and corresponding offline thresholds are summarized in Tables 5.4 & 5.5. The $\tau_h\tau_h$ final state benefits from improved online τ_h tagging algorithms compared to Run 2, as well as an extended trigger strategy covering a larger phase space. In $\mu\tau_h$ and $e\tau_h$ the cross triggers account for 15% of events captured, emphasizing the importance of low kinematic thresholds at the trigger level to capture large amounts of signal.

Table 5.4: *Offline kinematic requirements for τ_h candidates in the $\tau_h\tau_h$ final state and jets when collected by a trigger. In 2024, new τ_h triggers resulted in a lower offline selection for events from the DiTau and DiTau+Jet triggers.*

Trigger	$p_T^{\tau_h^1}$ [GeV]	$p_T^{\tau_h^2}$ [GeV]	nJet	$p_T^{j_1}$	$p_T^{j_2}$	m_{jj}
2022 and 2023						
DiTau	≥ 40	≥ 40				
DiTau + Jet	≥ 35	≥ 35	≥ 1	≥ 65		
2024						
DiTau	≥ 35	≥ 35				
DiTau + Jet	≥ 30	≥ 30	≥ 1	≥ 65		
VBF + Tau	≥ 50	≥ 25	≥ 2	≥ 65	≥ 65	≥ 900

Table 5.5: *Offline kinematic requirements in the $\mu\tau_h$ and $e\tau_h$ final states.*

Final state	Trigger	τ_h p_T [GeV]	μ p_T [GeV]	e p_T [GeV]
$\mu\tau_h$	Single muon	≥ 25	≥ 25	
	MuTau	≥ 32	≥ 20	
$e\tau_h$	Single electron	≥ 25		≥ 31
	ETau	≥ 35		≥ 25

Events are selected by identifying electrons, muons, and τ_h candidates passing a loose selection criteria. These "loose" objects are then checked against a dilepton veto which reduces the background contribution from genuine $Z/\gamma + \text{jets}$ events decaying to two electrons or two muons. These vetoed events are later used in a control region to determine corrections to event MET and recoil. After the veto a tighter selection is applied and pairs of objects are created. All object pairs are required to have opposite electric charge, and be separated in η–ϕ space. This separation is measured by the angular distance, $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, and required to be greater than 0.5 between the two objects. If both objects have the same η, $\Delta R = 0.5$ is equivalent to being separated by 28.6deg in the detector. Objects with separation less than this distance have poorly defined isolation and identification variables for standard algorithms, making them unsuitable for this analysis.

Event selection differs by final state, with additional signal region cuts on derived variables applied to $\mu\tau_h$ and $e\tau_h$. The transverse mass variable, m_T, is used in $\mu\tau_h$ and $e\tau_h$ to reduce background from electroweak ($W + \text{jets}$) processes. The variable is defined in

Eqn. 5.1:

$$m_{\mathrm{T}} = \sqrt{2p_T^{lep}p_T^{miss}\left(1-\cos(\phi_{lep}-\phi_{miss})\right)} \tag{5.1}$$

Note that this is the transverse mass of a two-particle system, as opposed to the definition of transverse mass for a single particle as used by the PDG and implemented in the ROOT programming language[48;123].

The impact parameter is the distance from the primary interaction point[6]. This is a 3D quantity, but it is typically separated into transverse and longitudinal components, D_{xy} and D_z. The isolation variable used for electrons and muons is relative isolation defined from Particle Flow candidates, or "PFRelIso." It is defined in Eqn. 5.2, where the sum of charged hadrons, photons, neutral hadrons, and pileup is calculated within a ΔR cone less than 0.3 for electrons and 0.4 for muons, and normalized by the p_{T} of the electron or muon. Lower values are more isolated because there is less hadronic activity in the ΔR cone.[124]

$$I_{e,\mu} = \frac{1}{p_{\mathrm{T}}}\sum_{R<X}\left[p_{\mathrm{T}}^{\mathrm{ch.had.}} + \max(p_{\mathrm{T}}^{\mathrm{photons}} + p_{\mathrm{T}}^{\mathrm{neut.had.}} - 0.5p_{\mathrm{T}}^{\mathrm{pileup}}, 0)\right] \tag{5.2}$$

The loose and tight selections are summarized in Fig. 5.6 below. Note that "PFRelIso03" and "PFRelIso04" are the same isolation variables just defined. Object IDs are also used for the baseline selection of muons and electrons. The "Medium ID" for muons is better than 99% efficient for simulated decays from W and Z bosons[58]. For electrons, the conversion veto ("ConvVeto") rejects non-prompt electrons coming from photon conversion, the lost hits requirement has a similar purpose, and the final branch is an MVA used for particle electron ID with greater than 90% efficiency[57]. Possible object pairs and their associated final state are also summarized in Fig. 5.7. However, most of the time there are only two objects in an event passing the tight selection criteria, making the final state assignment straightforward.

Each jet in the event is re-evaluated after finding a final state pair. Jets overlapped with electrons, muons, or τ_{h} candidates selected as final state objects are no longer considered,

Common to all Final States

Pass Lepton Vetoes + Opposite Sign + dR > 0.5 for objects in pair + matched to trigger objects with dR < 0.5 cone

If two or more final state pairs are possible, the pair with the most isolated objects is selected

Object	Object Requirements
τ_h	D_z [cm] < 0.2
μ	PFRelIso04 < 0.15, D_{xy} [cm] < 0.045, D_z [cm] < 0.2, pass Medium ID
e	PFRelIso03 < 0.15, D_{xy} [cm] < 0.045, D_z [cm] < 0.2, pass ConvVeto, LostHits ≤ 1, pass mvaNoIso WP90

Channel	DeepTau vsJet	DeepTau vsMu	DeepTau vsEle	Lepton Iso	n Medium PNet b Jets	m_T [GeV]	Δζ [GeV]
$\tau_h\tau_h$	M	VL	VVL	-	-	-	-
$\mu\tau_h$	M	T	VVL	< 0.15	0	< 60	-
$e\tau_h$	M	VL	T	< 0.15	0	< 60	-

Figure 5.6: *The top half of the figure shows the common selection for all final states separated by physics object. The bottom half are additional selections specific to each final state. The* DeepTau *columns refer to a selection for the τ_h candidates, where L, M, and T correspond to Loose, Medium, and Tight working points, respectively, with V meaning Very.*

Pair Passing Loose Sel.	"Event Possibly Categorized as [FS]" = O, "Not Possibly Categorized as [FS]" = X							Best Pair Information		
Event Type	Tau Tau	Mu Tau	E Tau	E Mu	Mu Mu	E E	Branch Veto Fired	HTT FS pair	nPossiblePairs	Analysis SR?
Tau Tau	O	X	X	X	X	X	X	Tau Tau	1	Y
Mu Tau	X	O	X	X	X	X	X	Mu Tau	1	Y
E Tau	X	X	O	X	X	X	X	E Tau	1	Y
E Mu	X	X	X	O	X	X	X	E Mu	1	Y
Mu Mu	X	X	X	X	O	X	X	Mu Mu	1	N, CR
E E	X	X	X	X	X	O	X	E E	1	N, CR
E-Mu-Tau + Triples										
E Mu Tau	X	O	O	O	X	X	O Mu Tau, O E Tau, X E Mu	E Mu	1	Y
Tau Tau Tau	O	X	X	X	X	X	X (Vetos on Mu/E)	Tau Tau	2	Y
Mu Mu Mu	X	X	X	X	O	X	O	Mu Mu	2	N, CR
E E E	X	X	X	X	X	O	O	E E	2	N, CR
Other Combinations										
Mu Tau Tau	O	O	X	X	X	X	X Mu Tau, O Tau Tau	Mu Tau	2	Y
Mu Mu Tau	X	O	X	X	O	X	O Mu Tau, X Mu Mu	Mu Mu	1	N, CR
E Tau Tau	O	X	O	X	X	X	X E Tau, O Tau Tau	E Tau	2	Y
E E Tau	X	X	O	X	X	O	O E Tau, X E E	E E	1	N, CR
E Mu Mu	X	X	X	O	O	X	O E Mu, O Mu Mu	Mu Mu	1	N, CR
E E Mu	X	X	X	O	X	O	O E mu, O E E	E E	1	N, CR

Figure 5.7: *A table of all possible 2- and 3-object combinations of electrons, muons, and τ_h candidates. The 2-object combinations are in the first rows. An "O" is used to show the pairing can be considered for the final state indicated by the column. The selected final state is shown in the next column, based on prioritizing signal events while requiring a dilepton veto. The possible number of pairs and whether the event is used in a signal region or control region is also shown in the final columns.*

Table 5.6: *Kinematic requirements for jet objects.*

p_T	$\lvert\eta\rvert$
$> 30\,\text{GeV}$	< 4.7
$> 50\,\text{GeV}$	> 2.5 and < 3.0

as they are redundant objects that should not be double-counted. Jets reconstructed in uninstrumented regions of the detector are also removed. The remaining jets must pass the kinematic requirements described in Table 5.6 to be considered in the analysis. This selection removes soft jets from pileup, and the tighter selection in the $2.5 < |\eta| < 3.0$ reduces the background of jets from the transition region of the detector.

Jets are considered b-tagged if they have an identification score from ParticleNet [69] greater than an era-dependent threshold value. Events in $\mu\tau_h$ and $e\tau_h$ with b-jets are vetoed. The veto is not enforced in $\tau_h\tau_h$ because it was found not to reduce background events significantly.

5.3.4 Background Estimation

The backgrounds considered in this analysis are the same for all channels, though their abundance varies per channel due to the different HLT path and selection criteria in each final state. The backgrounds $Z/\gamma + \text{jets}$, $W + \text{jets}$, TT, ST, VV, and HWW are modeled with MC, while QCD is derived from control regions of data. The $W + \text{jets}$ background is also derived from data in the $\mu\tau_h$ and $e\tau_h$ channels, although the $W + \text{jets}$ MC is still used in some steps of the derivation. Below, the processes modeled by MC will be described together as the methods are the same across final states. For the QCD and $W + \text{jets}$ background estimation, the general methods applicable to all final states will be introduced, and then the detailed implementation for $\mu\tau_h$ will be described. The method is implemented in a very similar way for $e\tau_h$.

Notably, another method to model $Z/\gamma + \text{jets}$ is the so-called "Embedding" method. A full description can be found in Ref. 125. The method was not used for this analysis because the technical updates to implement it in Run 3 were not finalized at the time of this analysis' completion.

5.3.5 Backgrounds from MC

For each MC sample, the events are scaled by the luminosity and the process cross section. After this initial scaling, additional weights accounting for detector efficiency, such as trigger scale factors (SFs, example derivation given in Chapter 6), are used to scale the total number of events and account for residual differences between data and MC. Depending on the sample, further reweighting is performed to account for discrepancies arising from known mismodeling, such as event recoil, and $t\bar{t}$ NNLO corrections. A list of corrections is described in Section 5.3.11.

Other than the considerations above, events from MC follow the same selection and categorization steps as events from data. A subset of events in the MC samples that are accounted for by data-driven estimation are removed using generator-level information to match the event type to avoid double-counting. These events pass the final state selection, but contain a τ_{h} candidate in the final state pair which cannot be matched at generator level. In effect, this means the MC models all genuine SM backgrounds, as well as backgrounds arising from light leptons being misidentified as τ_{h} candidates or "lepton fakes."

5.3.6 Data-Driven Estimation

The dominant background in this analysis is jets misidentified as τ_{h} candidates, called "jet fakes." The jets in this background arise primarily from QCD in $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, and from both QCD and $\mathrm{W}+\mathrm{jets}$ in $\mu\tau_{\mathrm{h}}$ and $e\tau_{\mathrm{h}}$. It is computationally expensive to generate a sufficiently large QCD or $\mathrm{W}+\mathrm{jets}$ MC samples because the background arises from misreconstructed objects by definition. Meaning simply producing QCD events will not guarantee true jets are generated which pass the DeepTau ID used in the signal region. To overcome this, analyses use control regions of the collected data to model the background in the signal region, hence the term "data-driven estimation." By directly using data, this technique's available statistics also scales with the collected luminosity, and has detector effects already present in the data that would normally need to be accounted for with additional SFs applied to MC.

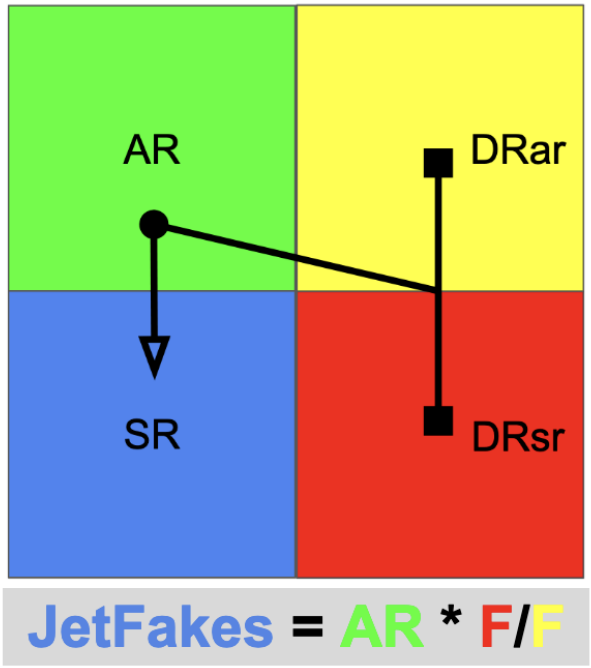


Figure 5.8: *Diagram showing a generic description of a fake factor's derivation and application. Note that the black diagram in the center can be interpreted as an operator. The vertical bar with two squares on the right side is the fake factor ratio, the diagonal line in the center connecting to a black dot indicates the application of the FF, and the left side vertical arrow points to the region the background estimate is intended for.*

In H $\rightarrow \tau\tau$-like analyses, this data-driven estimation technique is called the "Fake Factor" (FF) method, and is an example of an ABCD method[126]. This method has been used in some form in every H $\rightarrow \tau\tau$ and $Z \rightarrow \tau\tau$ analysis, and is described well in Refs. 93, 127, 98, & 95. A simple, generic FF is shown with Fig. 5.8. Using two variables of interest, four orthogonal regions are created. Three of the four regions are used to create an estimate of a process occurring in the remaining region, denoted with the black symbol. Typically, one wants an estimate of background in the signal region (SR). To do this, we use two determination regions (DR's) dominated by the desired background to find a ratio (sometimes called a transfer function), which is the FF. Events are weighted from the application region (AR) with the FF to estimate the amount of background in the SR. The DR's have selections which make them either SR-like, or AR-like, hence the labels DRsr and DRar, alternatively written with subscripts as DR_{sr} and DR_{ar}.

The AR, DRsr, and DRar have at least one selection inverted with respect to the SR,

defining regions abundant (or enriched) in the targeted background process, introducing a bias. Corrections for these biases are derived in a similar fashion for each channel, such that the jet fakes estimation is broadly summarized in five steps. The steps are similar between final states and backgrounds, with the main difference being what selections are made to define the SR, AR, and DR's.

1. Find the ratio of $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ of the fake object in two DR's which are abundant in either QCD or W + jets events, and fit a function to this ratio

2. Perform a closure check of this FF in the region it was derived, and correct the distribution as a function of p_{T} of the other object in the final state pair (the other τ_{h} in $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, μ in $\mu\tau_{\mathrm{h}}$, e in $e\tau_{\mathrm{h}}$)

3. Correct the bias due to differences in selection between the SR and DR's (the bias varies by final state)

4. Determine the composition of the AR, finding the fraction of events which are QCD or W + jets (other jet fakes backgrounds are accounted for by MC and are removed)

5. Finally, weight events from the AR using the FF as well as the event fraction found in the previous step. These weighted events are the final "Jet Fakes" background to be used in the SR

In the steps above, several components remain the same for $\mu\tau_{\mathrm{h}}$, $e\tau_{\mathrm{h}}$, and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$.

- All τ_{h} candidates must pass the loosest DeepTau vs Jet WP — in SR-like regions they must also pass the medium WP and in AR-like regions they must fail the medium WP

- Fit functions are usually one line and one constant, and sometimes two lines and one constant when that model better fits the data. In either case, the lines and constants are "connected" by an error function to smoothly transition between the regions:

```
(ROOT)
const + line = [0] + [1]*(x-[2])*(erf(-999*(x-[2]))+1)
const + line + line = [0] + [1]*(x-[2])*(erf(-999*(x-[2]))+1)
                      + [3]*(x-[4])*(erf(-999*(x-[4]))+1)
```

- FFs are always a function of $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ and the number of jets (N_{jets})
- Closure corrections to FFs are derived in the same region as the FF
- FFs are derived separately for the primary trigger (Single Muon, Single Electron, or DiTau) and cross-triggers (MuTau, ETau, or DiTau+Jet). This is necessary because cross-triggers cover a relatively small phase space, but have different isolation requirements already imposed at the trigger level
- Minor contributions of jet fakes from MC ($\mathrm{Z}/\gamma + \mathrm{jets}$, VV, TT, ST, HWW, Signal) are subtracted from the data in the DR's before FFs or corrections are calculated
- DR's for QCD estimation must have a same-sign (SS) final state object pair, while opposite-sign (OS) final state objects are required in the SR and AR
- DR's for $\mathrm{W} + \mathrm{jets}$ estimation require $m_{\mathrm{T}} > 60\,\mathrm{GeV}$, while the SR and AR require $m_{\mathrm{T}} < 60\,\mathrm{GeV}$ in the $\mu\tau_{\mathrm{h}}$ and $e\tau_{\mathrm{h}}$ final states
- Background estimation is performed separately for each year, but eras within years are combined to fully exploit available statistics (i.e. 2022 is 2022 preEE and 2022 postEE, and similarly for 2023)

The remaining sections will describe the FF derivation and correction procedure for the $\mu\tau_{\mathrm{h}}$ final state. Due to the similarity of the steps for the $\mu/\mathrm{e} + \tau_{\mathrm{h}}$ channels, we fully describe $\mu\tau_{\mathrm{h}}$ and only list the minor differences in $e\tau_{\mathrm{h}}$.

Fake Factor for the MuTau Final State

In $\mu\tau_h$ there are two dominant sources of jet fakes, QCD and W + jets. The QCD FF and corrections are derived first to have an accurate prediction of QCD contamination in the W + jets DR's during the W + jets FF derivation. The requirement of an SS pair targets a region dominated by QCD. An SS-to-OS bias correction is derived in the anti-isolated regions to account for this difference. In the QCD DR's the isolation of the μ is required to be between 0.02 and 0.15. The region 0.0 to 0.02 is excluded in the FF derivation to avoid events with well-isolated muons and faked τ_h candidates from other background processes, like Z/γ + jets and W + jets. A correction is derived to account for this difference. Throughout the derivation, simulated samples are used to subtract contributions from non-QCD background, including W + jets. The relevant regions in are depicted in Fig. 5.9.

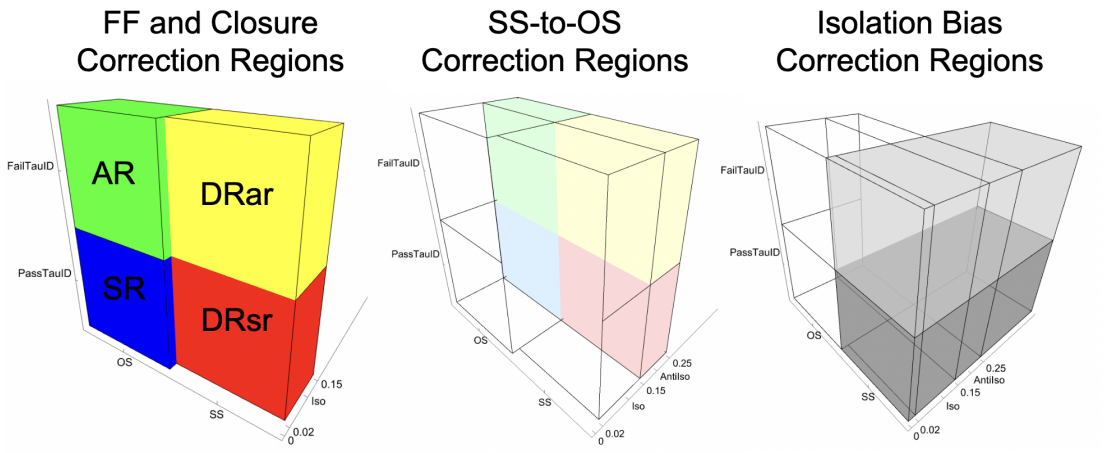


Figure 5.9: *Diagrams showing regions relevant to the FF and closure correction derivation (left), SS-to-OS bias correction derivation (middle), and isolation bias correction derivation (right).*

The FF is a fit of the ratio of events in DRsr and DRar, as a function of $p_T^{\tau_h}$, shown in Fig. 5.10. A closure test is performed with the FF, using DRar as the AR, resulting in a prediction of the background in DRsr. This prediction is compared to the real distribution of events in DRsr as a function of the μ leg. This is sometimes called the "unfakeable leg," because the μ must pass isolation requirements similar to the SR, and is assumed to be a genuine muon.

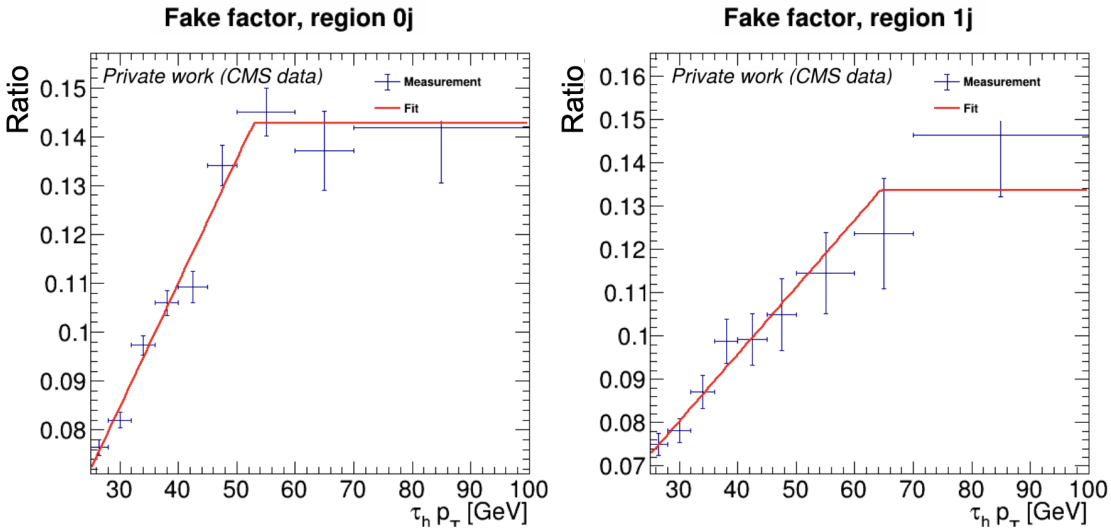


Figure 5.10: *Plots of the raw fake factor and fit for the QCD $\mu\tau_{\mathrm{h}}$ DRsr/DRar, as a function of $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ for $N_{jets} = 0$ (left) and $N_{jets} > 0$ (right).*

The isolation bias correction is derived using DR's without any isolation requirements, as shown in the rightmost diagram of Fig. 5.9. A comparison is made between the real and predicted I_μ distributions, where the predicted distribution is found by applying the FF and closure correction first derived in the isolated DR's to the DRar without any isolation requirements. A fit is made to the ratio of this comparison, and the result is the isolation bias correction.

The SS-to-OS correction is derived in an "anti-isolated" ("aiso") region, defined with $0.15 < I_\mu < 0.25$, and shown in the middle diagram of Fig. 5.9. To derive the SS-to-OS correction, an FF and closure correction are derived with the anti-isolated DR's, and then used with the anti-isolated AR to estimate the amount of background in the anti-isolated SR. The difference between the predicted and actual background from QCD in the anti-isolated SR is fit with a linear function and taken as the SS-to-OS correction.

The final QCD FF combines the four components described above: the core FF, the closure correction, the isolation bias correction, and the SS-to-OS bias correction, as described in Eqn.5.3, referencing subsequent equations to describe the nested terms.

$$FF_{QCD}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}}, p_{\mathrm{T}}^{\mu}, m_{vis}, I_{\mu}, N_{jet}) = FF_{core}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}}) \times CC(p_{\mathrm{T}}^{\mu}) \times Iso.\,Bias.(I_{\mu}) \times SS\text{-}to\text{-}OS(m_{vis}) \tag{5.3}$$

$$FF_{core}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}}) = \mathrm{Fit}\left[\frac{DRsr(p_{\mathrm{T}}^{\tau_{\mathrm{h}}})}{DRar(p_{\mathrm{T}}^{\tau_{\mathrm{h}}})}\right] \tag{5.4}$$

$$\begin{aligned} CC(p_{\mathrm{T}}^{\mu}) &= \mathrm{Fit}\left[\frac{DRsr(p_{\mathrm{T}}^{\mu})}{DRsr^{est.}(p_{\mathrm{T}}^{\mu})}\right] \\ \text{where} \quad DRsr^{est.} &= FF_{core}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}}) \times DRar \end{aligned} \tag{5.5}$$

$$\begin{aligned} Iso.\,Bias.(I_{\mu}) &= \mathrm{Fit}\left[\frac{DRsr_{noiso}(I_{\mu})}{DRsr_{noiso}^{est.}(I_{\mu})}\right] \\ \text{where} \quad DRsr_{noiso}^{est.} &= FF_{core}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}}) \times CC(p_{\mathrm{T}}^{\mu}) \times DRar^{noiso} \end{aligned} \tag{5.6}$$

$$\begin{aligned} SS\text{-}to\text{-}OS(m_{vis}) &= \mathrm{Fit}\left[\frac{SR_{aiso}(m_{vis})}{SR_{aiso}^{est.}(m_{vis})}\right] \\ \text{where} \quad SR_{aiso}^{est.} &= FF_{aiso}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}}) \times CC_{aiso}(p_{\mathrm{T}}^{\mu}) \times AR_{aiso} \\ \text{where} \quad FF_{aiso} &= \mathrm{Fit}\left[\frac{DRsr^{aiso}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}})}{DRar^{aiso}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}})}\right] \end{aligned} \tag{5.7}$$

The calculation of the W + jets FF is relatively simple in comparison to the QCD FF. For the W + jets DR's, the m_{T} selection is inverted (requiring $m_{\mathrm{T}} > 60\,\mathrm{GeV}$) to capture a region dominated by fakes from W + jets, as shown in the extended m_{T} plot of Fig. 5.11. Note that the W + jets regions (SR, AR, DRsr, and DRar) all use OS pairs. Additionally, the I_{μ} region is expanded to cover $0.0 < I_{\mu} < 0.15$, alleviating the need for another isolation bias correction.

Before calculating the W + jets FF, an estimate of QCD is calculated in the region m_{T}

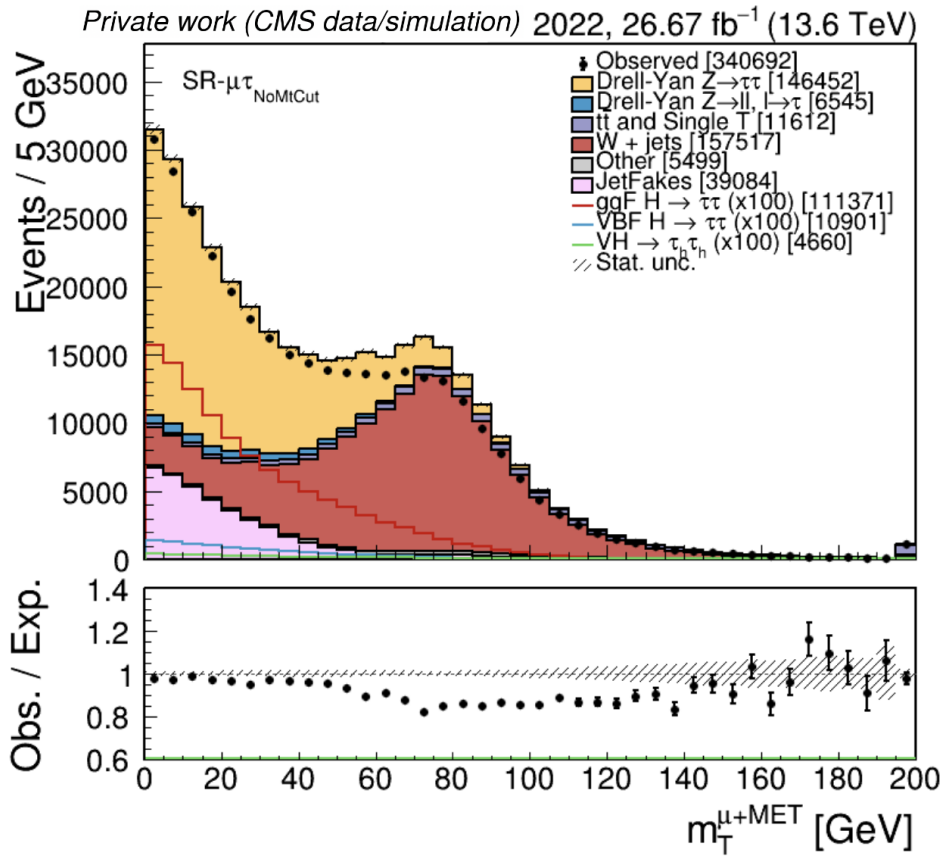


Figure 5.11: *This plot shows the m_{T} variable without the SR requirement at 60 GeV, highlighting the dominant WJ contribution in the region above 60 GeV. Note that this plot is from an early iteration of the analysis, is for demonstration only, and is missing several corrections that are otherwise applied in the analysis.*

$> 60\,\mathrm{GeV}$ with SS pairs, and $0.0 < I_{\mu} < 0.15$. A fit is made to the ratio of DRar and DRsr events as a function of $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$. No closure correction is calculated for this FF as it is only used to predict the QCD background contribution in the DR of the W + jets FF.

The W + jets FF and closure correction are calculated in the same way as before, but now using the W + jets DR's. A bias correction is calculated to account for the inversion of the m_{T} selection in the W + jets FF DR's. It is derived entirely from W + jets MC; no data or other background samples are used. Similar to previous corrections, we fit the ratio of events in W + jets MC minus other backgrounds to predicted W + jets events from the FF, and use the fit as the correction.

Several equations describe explicitly what variables and terms are used to derive the

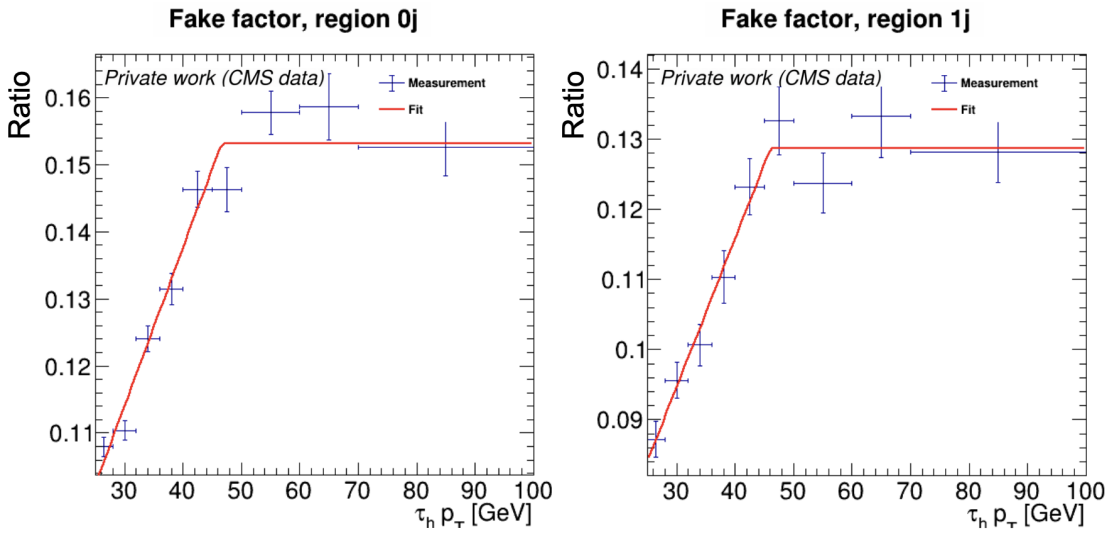


Figure 5.12: *Plots of the raw fake factor and fit for the* $\mathrm{W}+\mathrm{jets}$ $\mu\tau_\mathrm{h}$ *DRsr/DRar, as a function of* $p_\mathrm{T}^{\tau_\mathrm{h}}$ *for* $N_{jets}=0$ *(left) and* $N_{jets}>0$ *(right).*

$\mathrm{W}+\mathrm{jets}$ FF and the m_T-bias correction.

$$FF_{WJ}(p_\mathrm{T}^{\tau_\mathrm{h}}, p_\mathrm{T}^{\mu}, m_\mathrm{T}) = FF_{core}(p_\mathrm{T}^{\tau_\mathrm{h}}) \times CC(p_\mathrm{T}^{\mu}) \times m_\mathrm{T}\text{-}bias(m_\mathrm{T}) \tag{5.8}$$

$$\begin{aligned} m_\mathrm{T}^{bias}(m_\mathrm{T}) &= \mathrm{Fit}\left[\frac{DRar_{nomt}^{WJMC}(m_\mathrm{T})}{DRsr_{nomt}^{WJMC,est.}(m_\mathrm{T})}\right] \\ \text{where}\quad DRsr_{nomt}^{WJMC,est.}(m_\mathrm{T}) &= FF_{nomt}^{WJMC}(\tau_\mathrm{h}) \times CC_{nomt}^{WJMC}(p_\mathrm{T}^{\mu}) \times DRar_{nomt}^{WJMC} \\ \text{where}\quad CC_{nomt}^{WJMC}(p_\mathrm{T}^{\mu}) &= \mathrm{Fit}\left[\frac{DRsr_{nomt}^{WJMC}(p_\mathrm{T}^{\mu})}{FF_{nomt}^{WJMC}(p_\mathrm{T}^{\tau_\mathrm{h}}) \times DRar_{nomt}^{WJMC}}\right] \\ \text{where}\quad FF_{nomt}^{WJMC}(p_\mathrm{T}^{\tau_\mathrm{h}}) &= \mathrm{Fit}\left[\frac{DRar_{nomt}^{WJMC}(p_\mathrm{T}^{\tau_\mathrm{h}})}{DRsr_{nomt}^{WJMC}(p_\mathrm{T}^{\tau_\mathrm{h}})}\right] \end{aligned} \tag{5.9}$$

After the QCD and $\mathrm{W}+\mathrm{jets}$ FFs and corresponding corrections are derived, they are combined in a weighted sum and applied to events in the AR, yielding a data-driven estimation of the jet fakes background. The weights for the sum are the fractional bin content of QCD and $\mathrm{W}+\mathrm{jets}$ events in the m_T distribution of AR, shown in Fig. 5.13. There, $\mathrm{W}+\mathrm{jets}$ bin content is determined from MC, and QCD bin content is Observed Data minus All MC Backgrounds.

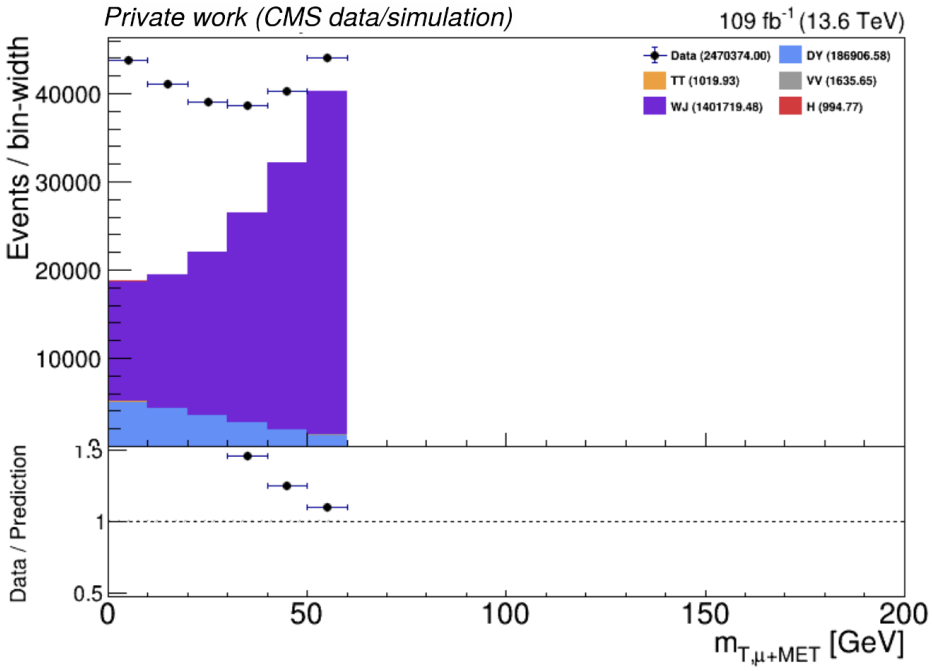


Figure 5.13: *This plot shows the m_T variable in the AR, where f_{QCD} and f_{WJ} are determined.*

$$FF = f_{QCD}FF_{QCD} + f_{WJ}FF_{WJ}$$
$$JetFakes = AR \times FF \tag{5.10}$$

Fake Factor Considerations for Other Final States

The $e\tau_h$ FFs are derived analogously to the description given for the $\mu\tau_h$ FFs, with the differences being the light lepton (e), HLT paths (SingleElectron, ETau), and isolation variable used (I_e). In $e\tau_h$, the isolation requirements for the DR's is $0.05 < I_e < 0.15$. Aside from the variables, the methods are exactly the same. The equations described in the above section are reused here, where all muon variables are replaced by corresponding electron variables.

In the $\tau_h\tau_h$ channel, QCD is the dominant source of jet fakes, with W + jets only being a minor background and modeled with MC. This channel also has the additional complication of two τ_h candidates, meaning more opportunities for jets to fake τ_h candidates, and no unfakeable, well-isolated lepton leg to rely on. Similar to the μ/e + τ_h channels, the $\tau_h\tau_h$

QCD estimation has eight total regions. These can be visualized as a cube or two connected squares, such as in the $\mu\tau_h$ channel. However, a more helpful mental model for $\tau_h\tau_h$ can be found by rearranging the regions based on the number of jet fakes, as shown in Fig. 5.14.

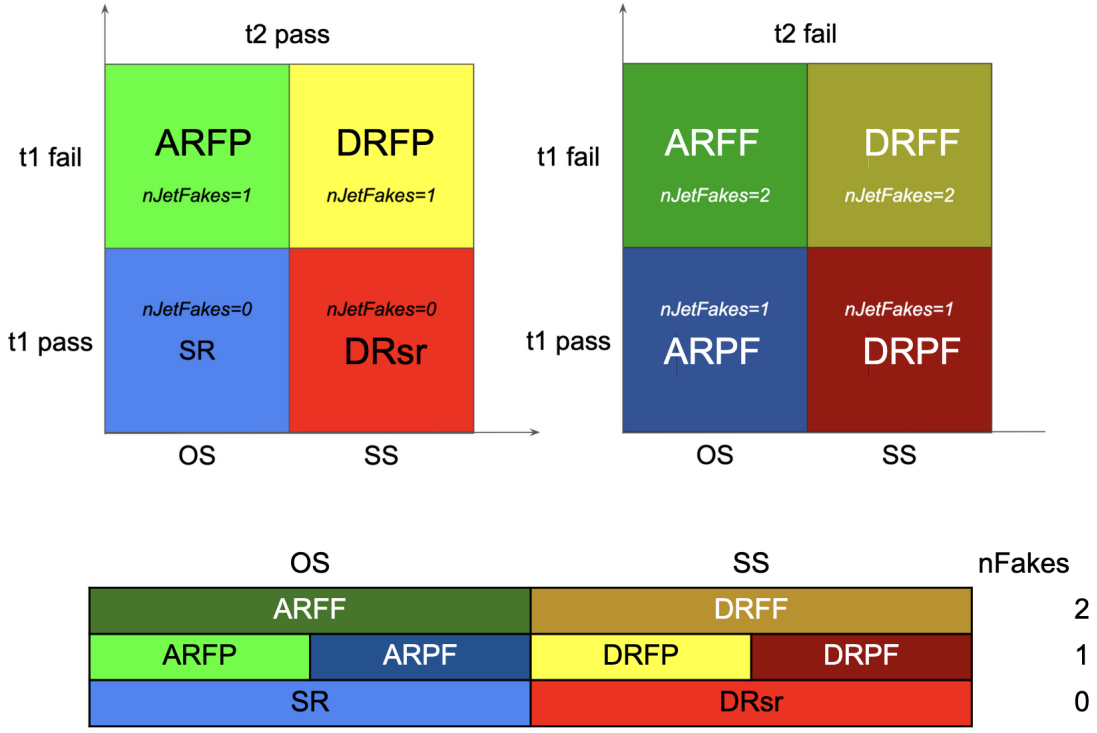


Figure 5.14: *Two representations of the eight regions used to derive the $\tau_h\tau_h$ QCD FF. The upper two squares represent the regions similarly to the $\mu\tau_h$ figure, whereas the lower figure represents the same regions but ordered by the number of τ_h candidates failing* DeepTau *ID, shortened as nJetFakes or nFakes.*

With this figure we introduce some new notation to better identify regions since "aiso" cannot be used. A label like "ARFP" indicates an application region where the leading τ_h candidate failed (F) DeepTau ID and the subleading τ_h candidate passed (P). Sometimes abbreviations like "FFtoPP" are used to indicate the ratio DRFF to DRsr, where DRsr is used interchangeably with "PP."

From here, we extend previous methods and derive an FF and closure correction for each type of event: one where the leading τ_h candidate is a jet fake (FPtoPP), one where the subleading τ_h candidate is a jet fake (PFtoPP), and one where both are jet fakes (FFtoPP). The relevant regions are depicted in the top three images of Fig. 5.15, where the bottom two

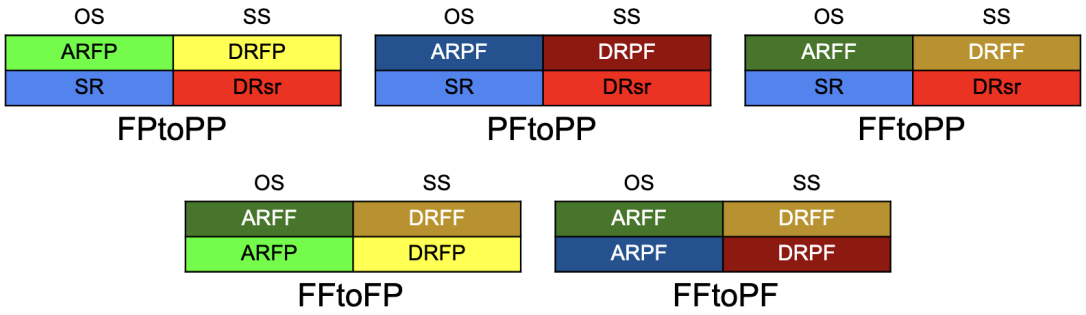


Figure 5.15: *Representations of the different combinations of regions used to derive $\tau_h\tau_h$ QCD FFs. The upper three are used to derive the nominal FFs, and the lower two are used to derive FFs used in the SS-to-OS bias correction.*

are later used for the SS-to-OS correction.

Several equations are given below to explicitly describe the contributions of each FF and all corresponding corrections. Note that the variable $p_T^{\tau_h\tau_h}$ is $p_{T,ll}$ where the two legs considered are the τ_h candidates.

$$FF_{FPtoPP}(p_T^{\tau_h^1}, p_T^{\tau_h^2}, p_T^{\tau_h\tau_h}, N_{jet}) = FF_{core}^{FPtoPP}(p_T^{\tau_h^1}) \times CC^{FPtoPP}(p_T^{\tau_h^2}) \times SS\text{-}to\text{-}OS^{PF}(p_T^{\tau_h\tau_h}) \tag{5.11}$$

$$FF_{PFtoPP}(p_T^{\tau_h^1}, p_T^{\tau_h^2}, p_T^{\tau_h\tau_h}, N_{jet}) = FF_{core}^{PFtoPP}(p_T^{\tau_h^2}) \times CC^{PFtoPP}(p_T^{\tau_h^1}) \times SS\text{-}to\text{-}OS^{FP}(p_T^{\tau_h\tau_h}) \tag{5.12}$$

$$FF_{FFtoPP}(p_T^{\tau_h^1}, p_T^{\tau_h^2}, p_T^{\tau_h\tau_h}, N_{jet}) = FF_{core}^{FFtoPP}(p_T^{\tau_h^1}) \times CC^{PFtoPP}(p_T^{\tau_h^2}) \times SS\text{-}to\text{-}OS^{FP}(p_T^{\tau_h\tau_h}) \times SS\text{-}to\text{-}OS^{PF}(p_T^{\tau_h\tau_h}) \tag{5.13}$$

$$
\begin{aligned}
SS\text{-}to\text{-}OS^{FP}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}\tau_{\mathrm{h}}}) &= \mathrm{Fit}\left[\frac{ARFP(p_{\mathrm{T}}^{\tau_{\mathrm{h}}\tau_{\mathrm{h}}})}{ARFP^{est.}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}\tau_{\mathrm{h}}})}\right] \\
\text{where} \quad ARFP^{est.} &= FF^{FFtoFP}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2}) \times CC^{FFtoFP}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1}) \times ARFF \\
\text{where} \quad FF^{FFtoFP} &= \mathrm{Fit}\left[\frac{DRFP(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2})}{DRFF(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2})}\right]
\end{aligned}
\tag{5.14}
$$

$$
\begin{aligned}
SS\text{-}to\text{-}OS^{PF}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}\tau_{\mathrm{h}}}) &= \mathrm{Fit}\left[\frac{ARPF(p_{\mathrm{T}}^{\tau_{\mathrm{h}}\tau_{\mathrm{h}}})}{ARPF^{est.}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}\tau_{\mathrm{h}}})}\right] \\
\text{where} \quad ARPF^{est.} &= FF^{FFtoPF}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1}) \times CC^{FFtoPF}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2}) \times ARFF \\
\text{where} \quad FF^{FFtoPF} &= \mathrm{Fit}\left[\frac{DRPF(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1})}{DRFF(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1})}\right]
\end{aligned}
\tag{5.15}
$$

$$
\begin{aligned}
FF_{core}^{FPtoPP}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1}) = \mathrm{Fit}\left[\frac{DR_{sr}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1})}{DRFP(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1})}\right], \quad FF_{core}^{PFtoPP}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2}) = \mathrm{Fit}\left[\frac{DR_{sr}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2})}{DRPF(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2})}\right] \\
FF_{core}^{FFtoPP}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1}) = \mathrm{Fit}\left[\frac{DR_{sr}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1})}{DRPF(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1})}\right]
\end{aligned}
\tag{5.16}
$$

$$
\begin{aligned}
CC^{FPtoPP}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2}) = \mathrm{Fit}\left[\frac{DR_{sr}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2})}{DR_{sr}^{FPest.}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2})}\right], \quad CC^{PFtoPP}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1}) = \mathrm{Fit}\left[\frac{DR_{sr}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1})}{DR_{sr}^{PFest.}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^1})}\right] \\
CC^{FFtoPP}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2}) = \mathrm{Fit}\left[\frac{DR_{sr}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2})}{DR_{sr}^{FFest.}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}^2})}\right] \quad \text{where} \quad DR_{sr}^{XXest.} = FF^{XXtoPP}(p_{\mathrm{T}}^{\tau_{\mathrm{h}}}) \times DRXX \\
\text{where} \quad XX = \{FP, PF, FF\}
\end{aligned}
\tag{5.17}
$$

With the FFs and corrections now derived, they are used to predict the QCD background contribution to the SR. Each FF is only used with its respective application region, as described in Eqn. 5.18.

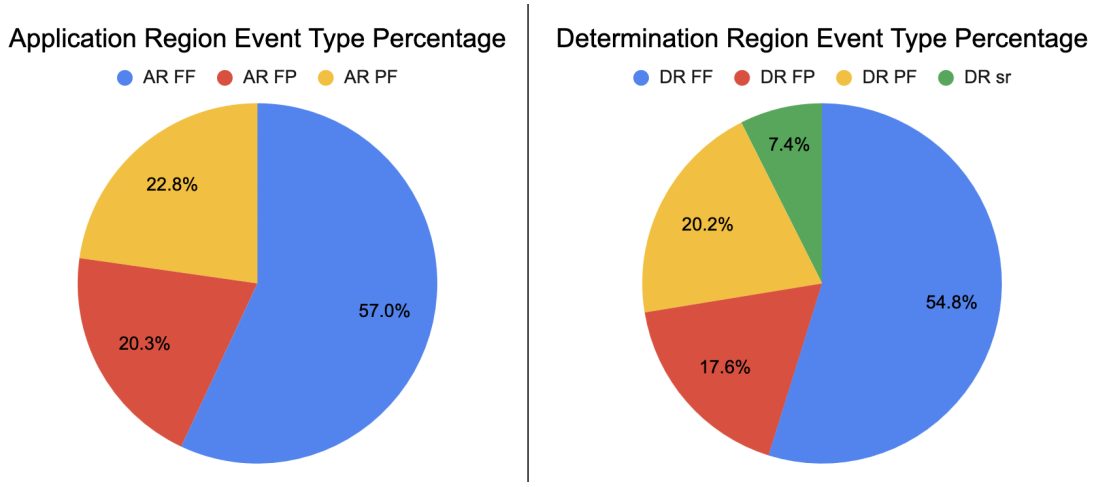


Figure 5.16: *Pie charts showing the types of events present in the application regions and determination regions in the $\tau_h\tau_h$ final state for 2024.*

$$JetFakes = FF_{FPtoPP} \times ARFP + FF_{PFtoPP} \times ARPF + FF_{FFtoPP} \times ARFF \quad (5.18)$$

The fake factor method can be technically challenging to setup and validate due to the sheer volume of regions to test. In recent years efforts have been made to measure FFs with neural networks instead of performing the method by hand as done in this analysis. For $\tau_h\tau_h$ it is instructive to note that the majority of events have both τ_h candidates being faked by a jet, meaning the FF is primarily determined by those types of events, with corrections from the other two types at the level indicated in Fig. 5.16.

5.3.7 Categorization

The variable of interest in this analysis is the mass of the dilepton system ($m_{\tau\tau}$) with bins of: [50, 70, 90, 110, 130, 150, 170, 210, 250, inf] GeV. It is calculated using FastMTT[109], a maximum likelihood based estimation for the mass of a $\tau\tau$ system considering missing energy due to neutrinos. As mentioned previously, FastMTT is also used to calculate the Higgs boson transverse momentum.

To enhance analysis sensitivity we subdivide each final state. The $\tau_h\tau_h$, $\mu\tau_h$, and $e\tau_h$ are

divided into three categories using the $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$. For $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$, the subleading τ_{h} candidate p_{T} is used. The boundaries are defined in Table 5.7, and are referred to in the text as low, medium, or high $p_{\mathrm{T}}^{\tau_{\mathrm{h}}}$ regions. Notably, the lowest p_{T} τ_{h} candidates in the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final state in the Run 2 analysis were 40 GeV.

Table 5.7: *Table of boundaries for final state categories.*

Final State	Low τ_{h} p_{T} [GeV]	Medium τ_{h} p_{T} [GeV]	High τ_{h} p_{T} [GeV]
$\tau_{\mathrm{h}}\tau_{\mathrm{h}}$	25-50	50-70	≥ 70
$\mu\tau_{\mathrm{h}}$	25-50	50-70	≥ 70
$e\tau_{\mathrm{h}}$	25-50	50-70	≥ 70

5.3.8 Differential variables

To perform the differential measurement, the events are split further using the reconstruction-level variables p_T^{H}, number of jets, leading jet p_{T}, m_{jj}, and $\Delta\eta_{\mathrm{jj}}$. The categories follow the same divisions as the Run 2 analysis, to allow for easy comparison. For the m_{jj} and $\Delta\eta_{\mathrm{jj}}$ quantities, the binnings come from Run 2 H $\rightarrow$ ZZ differential analyses. The bin boundaries are shown in Table 5.8.

Table 5.8: *Table of differential bin boundaries per variable. The "01j" entries for m_{jj} and $\Delta\eta_{\mathrm{jj}}$ account for events that do not have 2 or more jets. Similarly, the 0-30 bin of leading jet p_{T} captures events with zero jets because the lowest p_{T} jets in the analysis are 30 GeV.*

p_T^{H} [GeV]	nJets	Leading Jet p_{T} [GeV]	m_{jj} [GeV]	$\Delta\eta_{\mathrm{jj}}$	$\Delta\phi_{\mathrm{jj}}$
0-45	0	0-30	01j	01j	01j
45-80	1	30-60	0-75	0-0.7	$[-\pi, -\pi/2]$
80-120	2	60-120	75-120	0.7-1.6	$[-\pi/2, 0]$
120-200	3	120-200	120-180	1.6-3.0	$[0, \pi/2]$
200-350	≥ 4	200-350	180-300	3.0-5.0	$[\pi/2, \pi]$
350-450		≥ 350	300-500	≥ 5.0	
≥ 450			500-1000		
			≥ 1000		

5.3.9 Unfolding

Of the Higgs boson production modes, the subdominant VBF, WH, ZH, and ttH are grouped together as one mode called xH. This is done because the statistics of any individual subdominant mode is not large enough to perform a differential measurement. The result is two modes, ggH, and xH.

In addition to the reconstruction-level splitting categorization, signal samples are also split by generator-level bins of the same variables. In this way, we are able to determine which generator-level bins contribute to our reconstruction-level observations, and control for detector resolution effects. This generator-level to reconstruction-level comparison is most often made with a "response" plot, which is a representation of the migration matrix of generator-level quantities to reconstruction-level bins. Response plots are calculated per differential variable per era of data-taking. An example plot of $\mu\tau_h$ $\Delta\eta_{jj}$ is shown in Fig. 5.17. It is observed that $\mu\tau_h$ is typically the final state with the most diagonal response, followed by $e\tau_h$, and then $\tau_h\tau_h$. This is driven by the resolution of the light lepton and the number of neutrinos present in the final state.

Additionally, it is observed that events with zero jets have noticeably poorer response plots in $\tau_h\tau_h$ than events with at least one jet, particularly noticeable for $p_T^H < 60$ GeV. This is due to reconstruction inefficiency in MET for zero-jet events, and misidentification of one jet events as zero jet events. In Run 2, this poor response was corrected as function of $Reco/Gen[p_T]$. A similar method was pursued in Run 3, but it was determined that FastMTT could be used to calculate p_T^H resulting in better response plots, replacing the p_T^H correction altogether. Fig. 5.18 demonstrates the response in $\tau_h\tau_h$ for the three treatments. The FastMTT p_T^H is used in all final states, and is directly comparable with the p_T^H differential variable of other analyses.

5.3.10 Fiducial Region

The measurement is performed within a fiducial volume described in Table 5.9. Signal events from ggH or xH are required to pass the fiducial selection, and those failing are treated

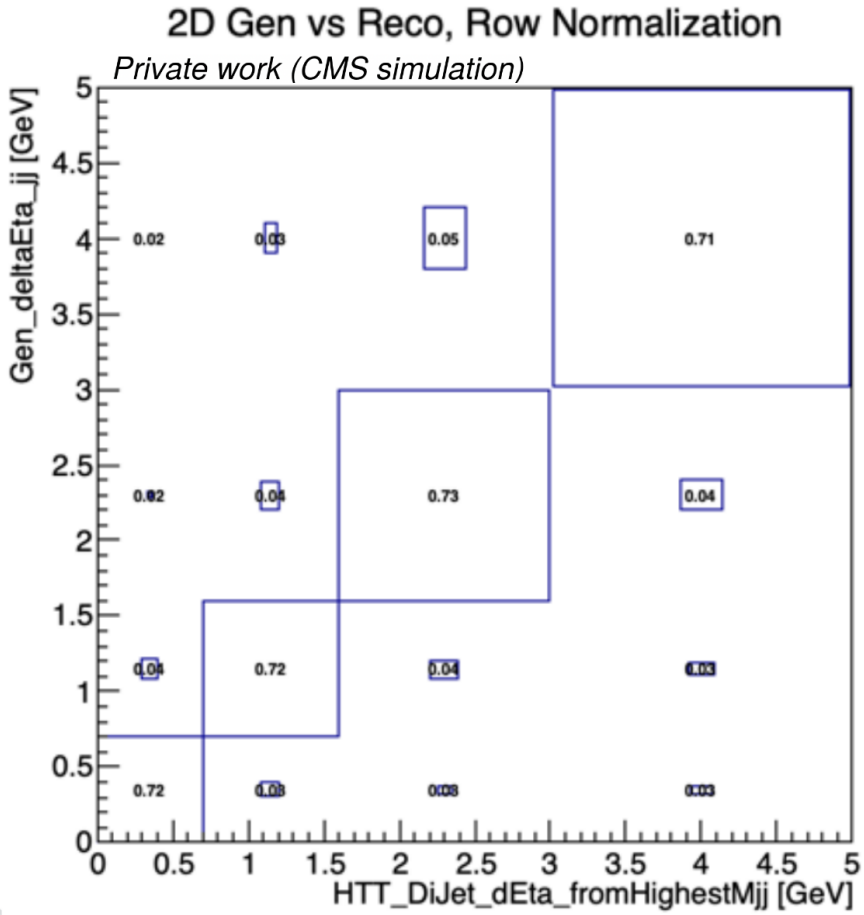


Figure 5.17: *A response plot of the $\Delta\eta_{\mathrm{jj}}$ variable for $\mu\tau_{\mathrm{h}}$ in 2024. This plot shows good alignment between generator-level and reconstruction-level quantities as evidenced by values near 1 along the diagonal, and symmetric lower values on the off-diagonals. This plot is row-normalized, meaning the displayed reconstructed values for each row are normalized to the total number of events in that row of the generator-level variable.*

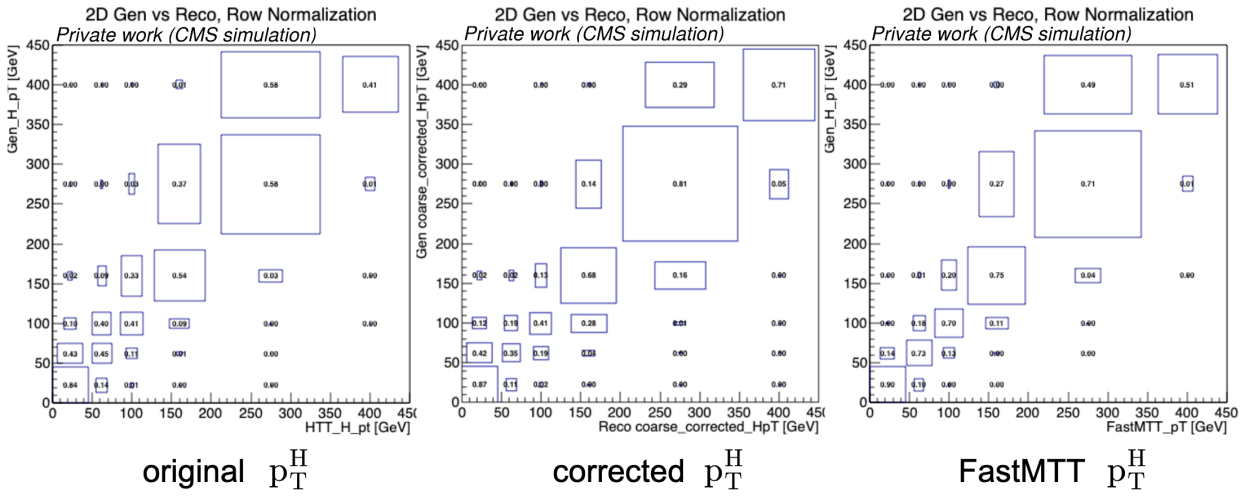


Figure 5.18: *Response plots of the p_T^H variable for $\tau_h\tau_h$ in 2024 under different treatments. The leftmost plot shows the original response of p_T^H, which is noticeably skewed upward. The middle plot shows p_T^H with a correction function applied, derived in a similar manner as the Run 2 correction, which improves the symmetry of the response but leaves lower p_T^H values mostly unaffected. The rightmost plot shows the p_T^H variable calculated with* FastMTT*, which significantly improves the lowest values of p_T^H on the diagonal. The majority of events have $p_T^H < 80$ GeV, meaning it is more important to have a diagonal response at lower values, achieved with* FastMTT *in the rightmost plot.*

as background. Background events are not required to pass the fiducial selection. The selections are defined per final state and are made at generator-level with similar kinematic requirements as previously described that are, in general, looser to account for the resolution differences between generator-level and reconstruction-level.

The fiducial region is meant to clearly define the signal region of the analysis in terms of kinematic variables. To define this region, we collect the number of events passing generator-level selection only (N^{Gen}), passing reconstruction-level selection only (N^{Reco}), and passing both ($N^{Gen\&Reco}$), then construct the ratios $N^{Gen\&Reco}/N^{Reco}$ and $N^{Gen\&Reco}/N^{Gen}$. In a perfect detector, both of these ratios would equal one, indicating in the first quantity that no generator-level cuts remove events that would otherwise pass reconstruction, and indicating in the second quantity that no reconstruction-level cuts remove events that otherwise pass the generator-level requirements. In the real world, due to resolution losses from imperfect reconstruction, we expect reconstruction-level requirements, especially particle ID and isolation, to reduce the $N^{Gen\&Reco}/N^{Gen}$ ratio. Values of the ratio per final state are shown in

Table 5.9: *This table describes the fiducial selection used for each final state on generator-level objects variables. The generator-level τ_h candidates are produced by combining all of the visible daughter particles of the true τ lepton, which excludes energy from neutrinos.*

$H \to \tau\tau$ (e.g., $H \to WW$ is treated as non-fiducial)	
$\tau_h\tau_h$	Generated $\tau_h\tau_h$ decay Generated leading $p_T^{\tau_h} > 25\,\text{GeV}$ and $\lvert\eta_{\tau_h}\rvert < 2.5$ Generated subleading $p_T^{\tau_h} > 20\,\text{GeV}$ and $\lvert\eta_{\tau_h}\rvert < 2.5$
$\mu\tau_h$	Generated $\mu\tau_h$ decay Generated $p_T^{\mu} > 20\,\text{GeV}$ and $\lvert\eta_{\mu}\rvert < 2.4$ Generated $p_T^{\tau_h} > 20\,\text{GeV}$ and $\lvert\eta_{\tau_h}\rvert < 2.5$ Generated $m_T(\mu, p_T^{\text{miss}}) < 60\,\text{GeV}$
$e\tau_h$	Generated $e\tau_h$ decay Generated $p_T^{e} > 25\,\text{GeV}$ and $\lvert\eta_{e}\rvert < 2.5$ Generated $p_T^{\tau_h} > 20\,\text{GeV}$ and $\lvert\eta_{\tau_h}\rvert < 2.5$ Generated $m_T(e, p_T^{\text{miss}}) < 60\,\text{GeV}$

Table 5.10.

The generator-level objects are hadronic taus, muons, or electrons which have a Higgs boson as a parent particle. In the case of the muons and electrons, a small ΔR cone of 0.3 is made around the location of the generator-level lepton and all photons within this cone are included in the generator-level lepton's p_T determination. These are called "Dressed Leptons". The generator-level τ_h candidates are produced by combining all of the visible daughter particles of the true τ lepton, which excludes energy from neutrinos at generator-level. Dressed leptons and generator-level τ_h candidates are used because they more accurately correspond to the offline objects used at the reconstruction level. Most τ_h candidates are found to have p_T within 5 GeV of the generator-level value, and most muons and electrons are found to have p_T within 1 GeV of the generator-level value. Generator-level jets are AK4 and required to have $p_T > 30\,\text{GeV}$ and $|\eta| < 4.7$.

5.3.11 Corrections

Various corrections are applied to events from MC samples to improve the overall agreement to data. Standard corrections specific to simulated samples used in CMS include top

Table 5.10: *Ratios related to the fiducial selection, the middle being purity, and the right being fiducial selection efficiency.*

Final State	$p = N^{Gen\&Reco}/N^{Reco}$	$\epsilon = N^{Gen\&Reco}/N^{Gen}$
$\tau_h\tau_h$	99%	4%
$\mu\tau_h$	98%	26%
$e\tau_h$	97%	15%

quark p_T reweighting and next-to-next-to-leading order parton showering corrections for ggH production. A pileup correction derived with minimum bias data is used in all CMS analyses.

The reconstructed energy of τ_h candidates, electrons, and muons in data differs slightly from that in MC. To improve the data-MC agreement, energy scale corrections are used. These are multiplicative factors applied to the p_T and the mass of the reconstructed τ_h candidate in MC, or η bins in the case of electrons and muons. These corrections scale the energy of the reconstructed objects in simulation to match the observation in data. Correction derivations are standardized and provided by internal groups at CMS.

Similarly, trigger efficiencies are evaluated in data and MC and SFs are derived to correct residual differences from changes in data taking conditions. Efficiencies for triggers used in this analysis are all derived similarly in control regions isolating one object and measuring an unbiased trigger efficiency of the other. A description of the derivation for triggered τ_h candidates is given in the next chapter.

The $p_{T,ll}$ spectrum of $Z/\gamma + jets$ is known to be mismodeled. Corrections are derived in a $Z \to \mu\mu$ control region to improve the data-MC agreement as a function of the dilepton system's p_T. This correction is applied to all $Z \to \ell\ell$ processes, including those where a tau lepton decays hadronically.

The recoiling of a boson against jets and MET is corrected in MC using events in data containing $Z \to \mu\mu$ or $Z \to ee$ with jets. The momentum vector of the dilepton system is decomposed into several components to isolate the recoil of the vector boson, and correct MC to match data in the control region. This correction is applied to all MC processes in the analysis using bosons, including signal, $Z/\gamma + jets$, and $W + jets$.

5.3.12 Statistical Methods

The results in this note were derived with the CMS statistical analysis tool COMBINE[128], which is based on the RooFit[129] and RooStats[130] frameworks. The methods used to extract the results are the same as those used in the Run 2 H $\rightarrow \tau\tau$ differential measurement[99], and the Run 2 H $\rightarrow$ WW differential measurement[131], namely a differential unfolding.

The problem of unfolding can be cast algebraically as $\mathbf{n} = R\mathbf{s} + \mathbf{b}$, where $\mathbf{n}$, $\mathbf{s}$, and $\mathbf{b}$ are vectors of the total number of events per bin, the number of signal events per bin, and the number of background events per bin, respectively. The factor R describes the efficiency of reconstructing a signal event in the correct bin, which accounts for bin migrations due to measurement resolution. In the single bin case R is one, and in multibin problems R is a matrix. Therefore, R is called the response matrix and describes the detector smearing of generated-level information from one bin into multiple reconstruction-level bins.

We perform a simultaneous fit of all reconstruction-level bins, which determines the signal strength modifiers of the generated-level bins by using the extra information encoded in the response matrix. This process is equivalently described as maximizing a likelihood function of the form:

$$\mathcal{L}\left(\mu;\theta\right) = Pois\big(n_i; \sum_{j}^{fid} \epsilon_{i,j}\mu_j L_i \sigma_j^{fid} + f^{nonfid}(\theta)\sigma^{fid}L_i + B_i(\theta)\big)\mathcal{C}(\theta)K(\mu) \tag{5.19}$$

Above, μ is the overall signal strength modifier, and μ_j is the signal strength modifier for bin j. Signal strength was previously defined as the ratio with respect to the SM expectation for a given signal process, meaning $\mu = 1.1$ indicates 10% more events being measured than the SM would predict.

The nuisance parameters, another term for uncertainties, are represented by θ, which are denoted both as inputs to the response matrix $\epsilon_{i,j}$ as well as constraint terms, $\mathcal{C}$. The expression "Poisson(n; k)" denotes the Poisson probability of observing n events under the expectation k. Here, n_i is the number of observed events, and k is the combined signal-

plus-background estimation from MC, both in bin i. The i^{th} bin referred to above is one reconstruction bin of FastMTT, in one reconstruction bin of a differential variable, for a single tau-p_{T} category, of one final state, in one era. This alternatively could be written explicitly as a product where one index refers to each of those. The signal in the signal-plus-background term comes from the product of the $\epsilon_{i,j}$ matrix and the generator-level signal strength modifier of the j^{th} bin, summed over j generator-level bins (and multiplied by luminosity and cross-section), resulting in the i^{th} reconstruction-level bin signal contribution. Signal events falling outside of the fiducial region are treated as background, indicated by the term f^{nonfid}.

5.3.13 Results

The two quantities of interest are the signal strength parameter (μ) and its uncertainty. When blind, the Asimov dataset is used, which has $\mu = 1$ by definition, though the uncertainties are still informative. These values represent a ratio to the measured cross section in the detector. We convert signal strengths into fiducial cross sections. What is meant by "fiducial" is the cross section accepted by a perfect CMS detector, e.g. after all corrections are applied and after unfolding is performed to remove smearing from reconstruction.

The fiducial region is defined to be well-overlapped with the phase space covered in the analysis. A fiducial efficiency is defined as $\epsilon_{fid} = N^{Gen\&Reco}/N^{Gen}$, and is measured for each final state in Table 5.10. It is a measure of how much phase space is removed by generator-level cuts after reconstruction-level cuts are applied.

These values are combined with the standard counting equation, $N = \mathcal{L}\sigma$, which becomes:

$$\sigma_{fid} = \frac{N}{\mathcal{L}\epsilon_{fid}} \tag{5.20}$$

As seen in the top plot of Fig. 5.19, the blind inclusive cross section is calculated to be 1356 $\pm$172 fb ($\pm$84 fb stat. and $\pm$150 fb syst.). This estimate is from MC passed through the entire analysis pipeline and will be compared to a measurement with real data once the

analysis is unblinded.

The blind signal strength and uncertainty values from all final states combined are translated to a visual representation in Fig. 5.19.

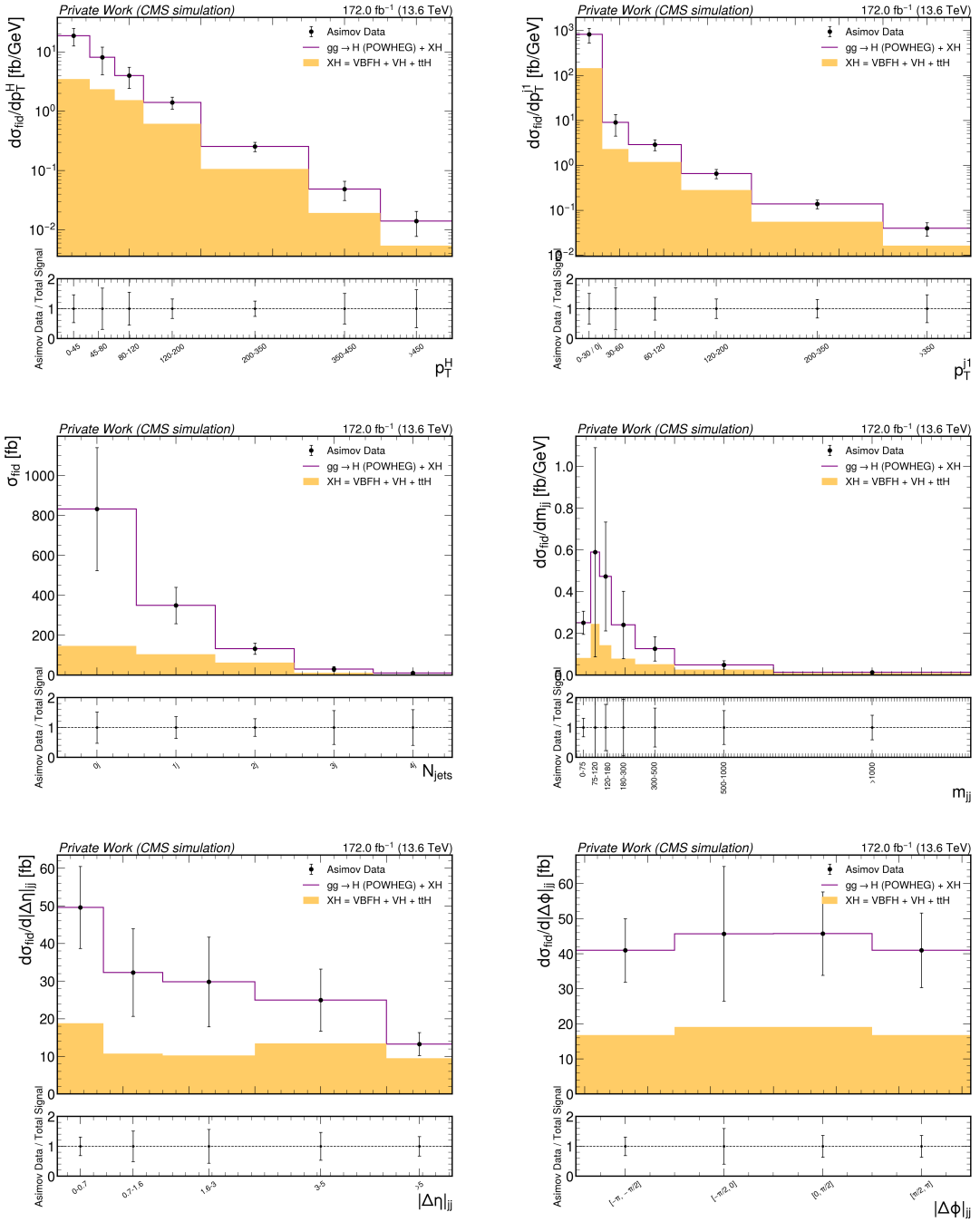


Figure 5.19: *Plots of the blind signal strength and uncertainty translated to a differential cross section.*

The results are extracted from all categories in all final states using COMBINE. The samples from 2022, 2023, and 2024 are used together, and the results are currently blinded, meaning the "Asimov" dataset is used in place of real data. The name comes from the sci-fi work "Franchise" by Isaac Asimov, where voting is done by a single person representative of an entire population[132]. In practical terms, the Asimov dataset in this analysis is the combination of signal and background MC where signal strength is set to 1, the SM expectation, because the $H \to \tau\tau$ signal has been observed previously.

Both statistical and systematic uncertainties are included. When compared to Run 2 blind results as shown in Fig. 5.20, expected improvement in the inclusive result is 40%. Part of this improvement is from using a larger dataset. Additional improvements are seen in the low p_T^H bins, attributable to usage of the FastMTT algorithm in Run 3 to compute the p_T^H variable. A notable improvement is seen in the zero jet bin, which in the Run 2 analysis did not include the $\tau_h\tau_h$ final state. Improvements in the low and high p_T range of the leading jet are likely attributable to better jet resolution in Run 3 and improved pileup mitigation techniques.

5.3.14 Summary

An analysis of the Higgs boson decay to two tau leptons was presented based on proton-proton collision data collected by the CMS experiment at the CENR LHC at a center-of-mass energy of $\sqrt{s} = 13.6$ TeV and totaling $172\,\text{fb}^{-1}$. Differential cross section measurements were made in Higgs boson transverse momentum, number of jets, leading jet transverse momentum, dijet invariant mass, and dijet angular separation. These are the first differential cross section measurements for dijet variables in $H \to \tau\tau$ at CMS. The results are currently blinded, but expected results with the Asimov dataset shows strong improvement over previous results.

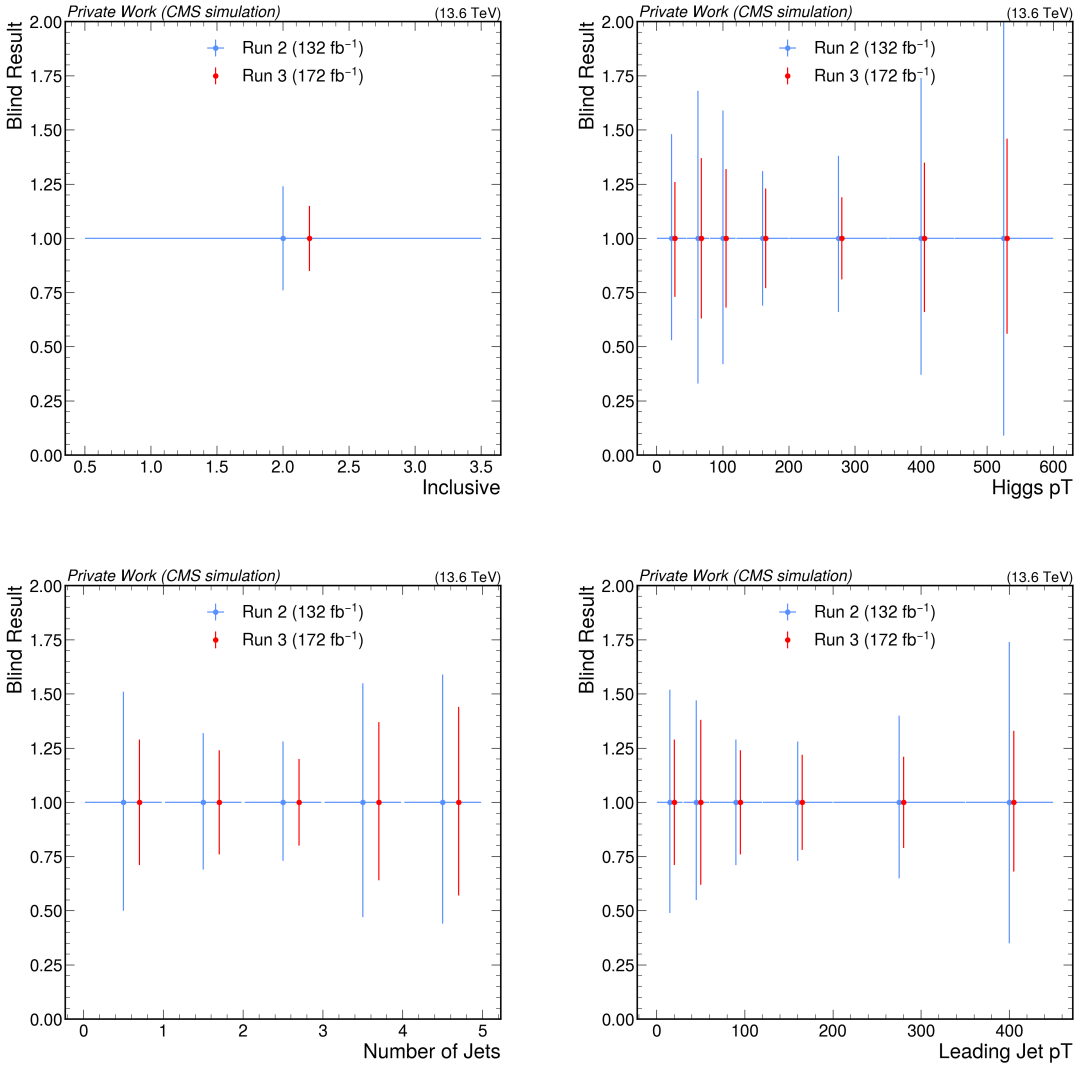


Figure 5.20: *Comparisons of the total uncertainty on the blind result per differential bin of the variables used in Run 2. The results from Run 3 generally improved, notably in zero jet and low p_T^{H} bins. Comparisons are not made for m_{jj} and $\Delta\eta_{\mathrm{jj}}$ because these were not measured in the Run 2 analysis.*

Chapter 6

Support Projects at CMS

Science is more collaborative now than it has ever been in the past. Using machines the size of buildings requires extreme calibration and coordination efforts. The CMS experiment is composed of over 6000 contributing members, including career scientists, technicians, and students from more than 60 countries[133]. All of these members work with CMS data but have different levels of experience, and pursue different analyses. Nevertheless, every result at CMS comes from the same detector, so it is a shared responsibility to develop new ideas, calibrate the detector as its response changes, and organize all the work into papers, reports, and meeting agendas.

6.1 Tau Trigger Efficiency Measurements

The Tau Physics Object Group (POG) is separated into three subgroups: the algorithms subgroup, the calibration and quality monitoring subgroup, and the trigger subgroup. Part of my service to CMS was organizing the tau trigger subgroup, responsible for managing the triggers in the tau data stream and producing routine efficiency measurements. These were made primarily with the tag-and-probe method[66], which triggers on an isolated muon and defines a phase space targeting $Z \to \tau\tau \to \mu\tau_h$ events. Since only the muon is triggered, the τ_h is not biased and the events are a suitable source for defining trigger efficiency. Using

events from the $\tau_h\tau_h$ trigger would introduce a bias in measuring the trigger efficiency because the trigger itself would be used to define the sample.

For events passing the IsoMu_24 HLT, a muon with $p_T > 24\,\text{GeV}$ and $|\eta| < 2.1$ is required. The muon (tag) is required to pass ID at the medium working point for muons[58]. Events with a second muon passing the loose ID are rejected to suppress events from the $Z \to \mu\mu$ process. The offline muon must be matched to the triggered muon within a cone $\Delta R < 0.5$. For the τ_h candidate (probe), $p_T > 20\,\text{GeV}$ and $|\eta| < 2.1$ are required. The τ_h must pass working points of the DeepTau ID specified per object: very loose for electrons, tight for muons, and medium for jets. A $\mu\tau_h$ pair is formed from the objects. Pairs with overall charge equal to zero are considered signal, and pairs with charge equal to ± 1 are considered to be background where the τ_h is a misidentified jet (i.e. JetFakes from QCD or W + jets). These events are subtracted from the signal events as a basic data-driven background approximation method. The visible mass of the pair is calculated and required to be between 40 and 80 GeV, corresponding with the visible mass peak of the Z boson. Transverse mass (m_T) is required to be less than 30 GeV to suppress background from W + jets.[134]

This selection produces a sample of $\mu\tau_h$ events where the τ_h leg is unbiased. Subsequently, this sample can be used to measure the efficiency of multiple hadronic tau triggers. As mentioned, using the $\tau_h\tau_h$ trigger directly would produce a biased sample, so "monitoring" $\mu\tau_h$ triggers with standard muon legs and exactly the same τ_h leg requirement as signal triggers are used instead. Since there are two steps to the trigger system efficiency can be defined with respect to the L1, the HLT, or the combined L1 and HLT, defined in the following equation. To measure the efficiency of the monitoring trigger, the offline τ_h leg must be matched in a ΔR cone of 0.5 to the online τ_h object captured by the monitoring trigger, denoted below with $Match\ \tau^{HLT}$:

$$\text{L1 + HLT Efficiency} = \frac{\text{IsoMu_24 + Offline Selection + Monitoring Trigger + Match } \tau^{\text{HLT}}}{\text{IsoMu_24 + Offline Selection}} \tag{6.1}$$

The results from data are compared to MC, and the ratio is taken to be the trigger SF,

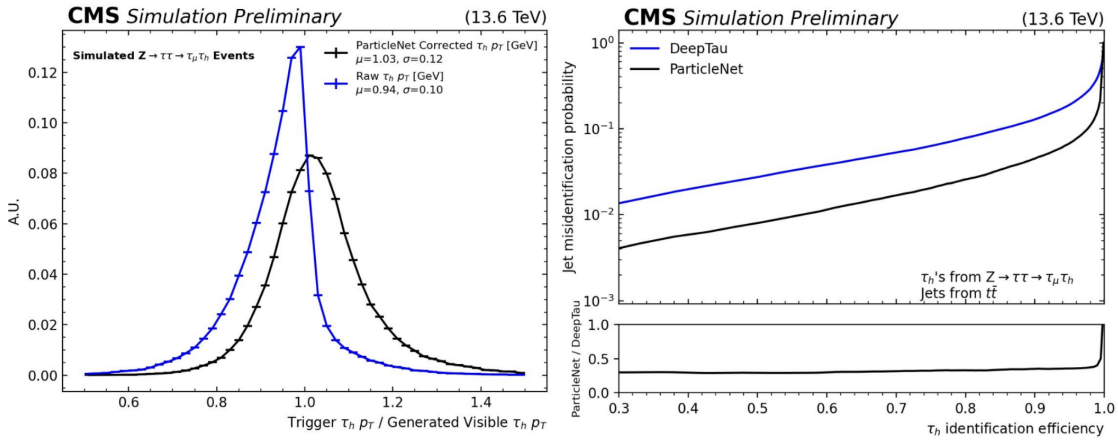


Figure 6.1: *On the left, a comparison between* τ_{h} *candidates captured by* DeepTau *and* ParticleNet *triggers in simulated* $\mathrm{Z}/\gamma+\mathrm{jets}$ *samples showing that* τ_{h} *candidates from the* ParticleNet *trigger have a reconstructed* p_{T} *that matches the object's generated* p_{T} *more often than* τ_{h} *candidates from the* DeepTau *trigger. On the right, a comparison showing a receiver operating characteristic (ROC) curve for the* DeepTau *and* ParticleNet *triggers; true* τ_{h} *candidates are sourced from the* $\mathrm{Z}/\gamma+\mathrm{jets}$ *sample and jets passing Medium offline ID and faking true* τ_{h} *candidates are sourced from a* $t\bar{t}$ *sample. Better performance on this curve is indicated by lower mis-ID rate and higher efficiency.*[134]

accounting for residual differences between data and MC conditions. In 2024 a new set of hadronic tau triggers were made using ParticleNet[69] because they were found to be more efficient than the online DeepTau triggers, as shown in Figs. 6.1 & 6.2.

These trigger efficiency measurements are crucial for any analysis using taus passing a trigger. The performance plot above shows a 10% increase in efficiency in $\mathrm{Z}\to\tau\tau\to\mu\tau_{\mathrm{h}}$ events, which have a similar phase space to $\mathrm{H}\to\tau\tau\to\mu\tau_{\mathrm{h}}$ events. This 10% increase in efficiency is roughly equivalent to an additional $10\,\mathrm{fb}^{-1}$ of data, or 9% of the 2024 dataset.

6.2 The MIP Timing Detector at the High-Luminosity LHC

The experiment changes over time, sometimes dramatically. Currently CMS is upgrading the detector to receive a higher luminosity beam from the LHC by 2030. The so-called HL-LHC will be at the same energy as the current beam, but have twice the proton density. This

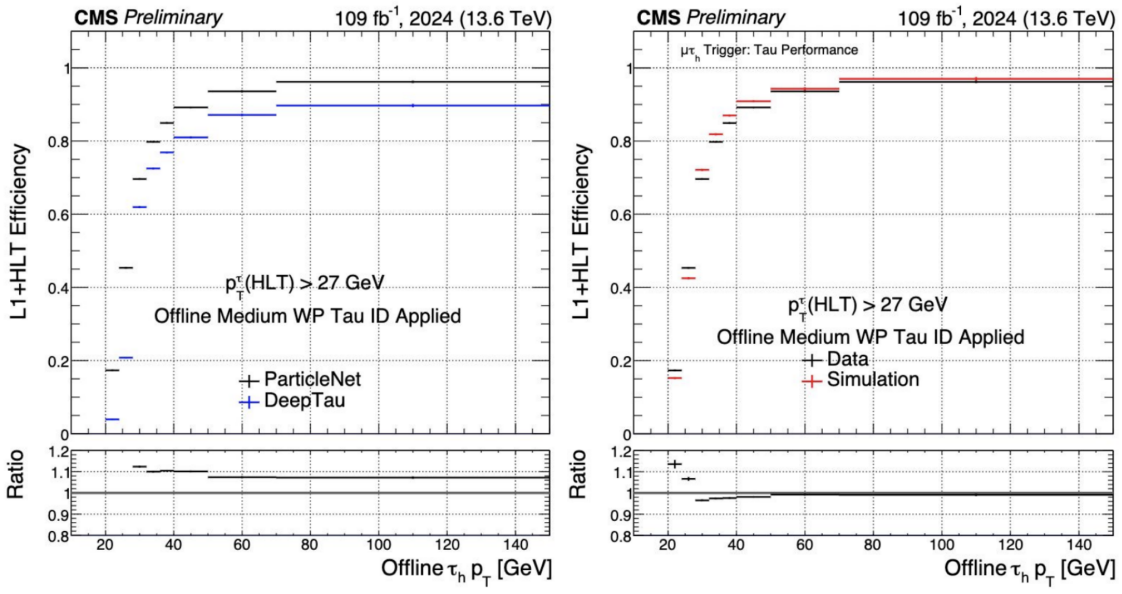


Figure 6.2: *On the left, τ_h candidates captured by* DeepTau *and* ParticleNet *triggers in 2024 data as a function of τ_h p_T, showing that τ_h candidates from the* ParticleNet *trigger have a higher efficiency. On the right, the* ParticleNet *trigger events from data are compared to a simulated* $Z/\gamma + \text{jets}$ *sample*[134]

will increase the number of interactions, especially soft interactions captured as pileup. The beam upgrade will result in 3 attobarns (3000 fb^{-1}) of data by 2040[36;135], as indicated in Fig. 6.3.

In Run 3, CMS operated with an average pileup ranging from 40 to 60 interaction vertices per particle collision. In 2030, at the beginning of Run 4, the expected average pileup is 120-140, with an increase to 200 envisioned by Run 5[36]. To overcome this doubling and tripling of pileup and maintain Run-3-like resolution, the CMS detector will undergo an extensive upgrade. Major upgrades include a completely new tracker extended to $|\eta| = 4$, an enhanced L1 trigger system capable of 1 MHz event rate and using track information, high granularity calorimeters designed specifically for detector endcaps to capture high-eta physics processes, and a completely new subdetector providing timing information to events captured at CMS.

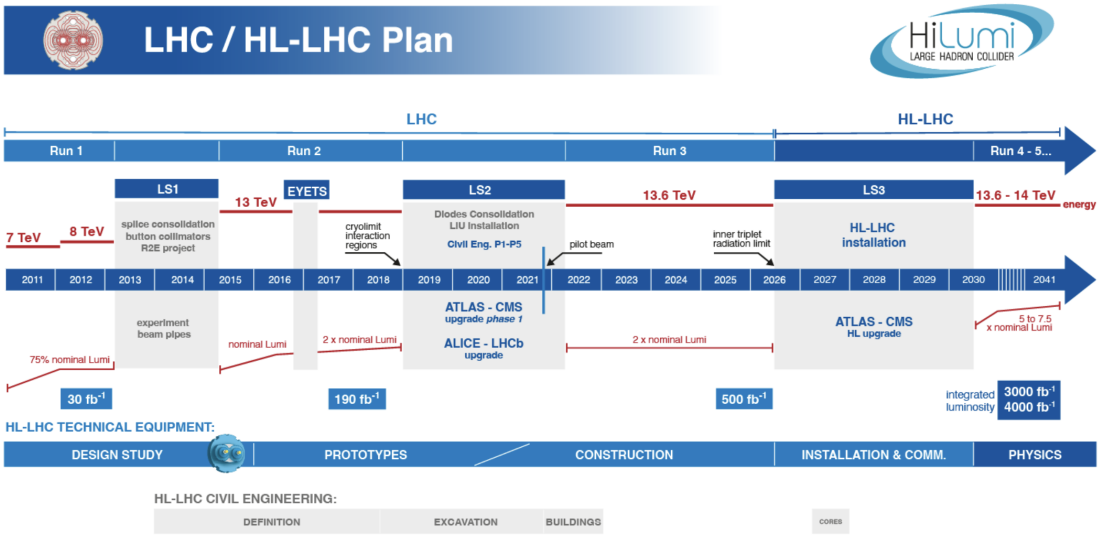


Figure 6.3: *Schedule of the High-Luminosity LHC in the context of the existing LHC and detector history*[136].

6.3 The Minimum Ionizing Particle Timing Detector

CMS collects events in 3D space in 45 nanosecond increments, which is enough information to reconstruct events with high spatial resolution in Run 3. In the busy environment of Run 4, the Minimum Ionizing Particle Timing Detector (MTD) provides CMS with a 4th dimension to separate events with: time. Using time-of-flight (TOF) information and track position together reduces the number of possible originating vertices. With MTD, Run-3-like resolution can be recovered in the Run 4 environment by providing 30-60 ps timing resolution to separate pileup tracks from primary interactions[137], as illustrated in Fig. 6.4.

The MTD is split into a two components, the barrel timing layer (BTL) capturing events with $|\eta|$ from 0 to 1.5, and the endcap timing layer (ETL), capturing events with $|\eta|$ up to 3.0. BTL and ETL use different technologies to deliver TOF information due to the different radiation environments they must tolerate. The principle component of BTL is a scintillating crystal, cerium-doped lutetium yttrium orthosilicate (LYSO:Ce), that converts energy from particle interactions into light collected by silicon photomultiplier (SiPM) pads and then converted to a TOF measurement. A key advantage of this technology is its relatively low cost and weight for the amount of detector coverage (38 m^2 or 410 ft^2) and resolution[138].

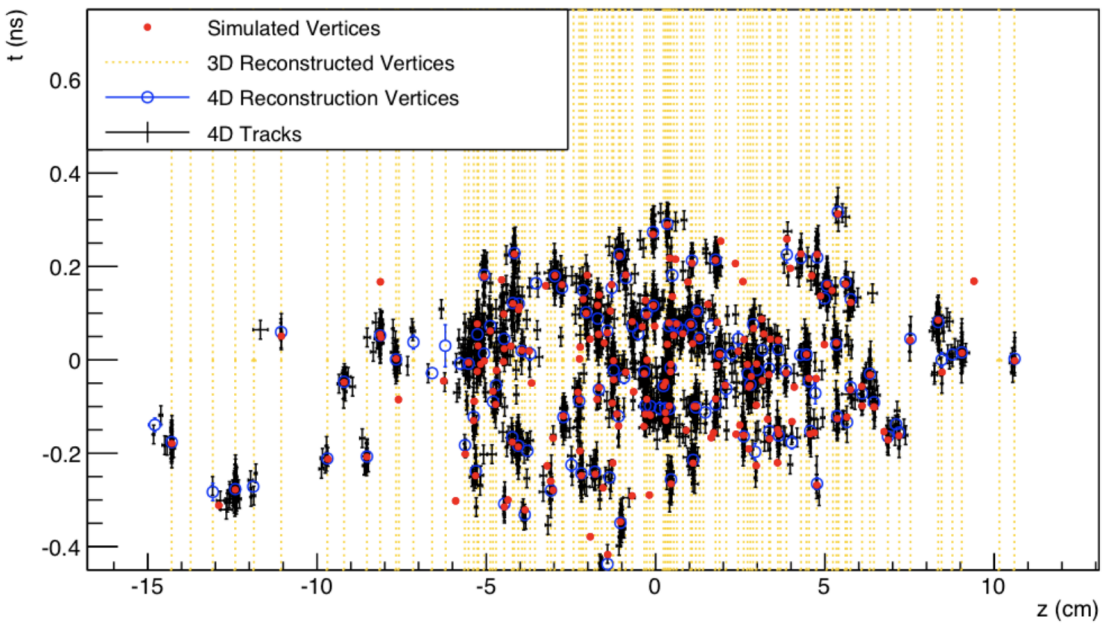


Figure 6.4: *An image showing simulated vertices in space and time for different reconstruction strategies. Using only spatial information (dashed lines) many events are overlapping or too near each other to be resolved. Using time information also (black crosses and blue circles), many simulated vertices can be resolved*[138].

TOF measurements are made with minimum ionizing particles (MIPs), the most common being muons and charged pions. Energy loss through particle interaction is dependent on particle velocity, and MIPs are sufficiently energetic to induce pair-production and cascade photoprotoduction when interacting with the LYSO:Ce scintillator. For relativistic particles that are not electrons, the energy loss is modeled by several competing processes described by the Bethe-Bloche equation, shown in Fig. 6.6[139].

However, SiPMs systematically suffer from a background called "dark current," which is jargon for electrical noise that is detected as a light signal without a MIP inducing it. This background arises from random thermal electrons and can be reduced by lowering the temperature of the components. BTL was designed to be passively cooled with liquid CO_2 to $-35°C$, with an additional active cooling of $-10°C$ from Peltier coolers. Also called thermo-electric coolers or TECs, these components convert an electric current between two conductors into a temperature change. This additional $10°C$ of cooling reduces the background from dark current by a factor of two[140].

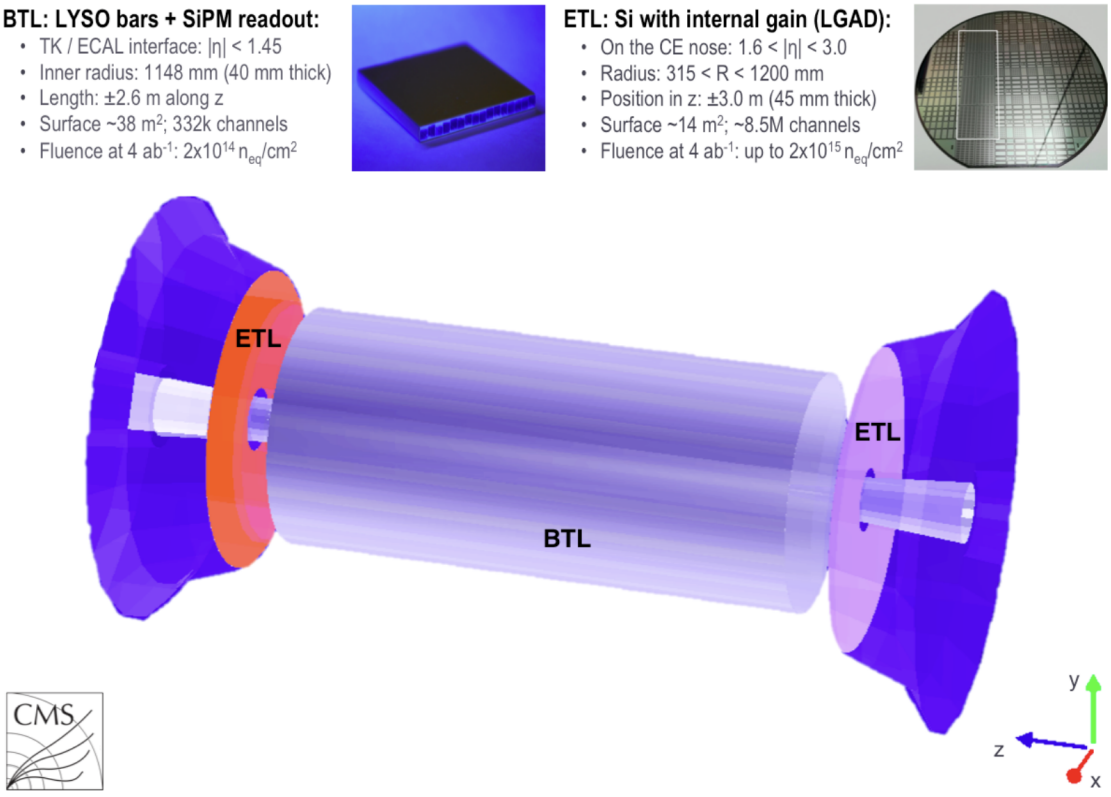


Figure 6.5: *Overview of the MIP Timing Detector, showing the Barrel Timing Layer and Endcap Timing Layer, highlighting the different technologies used in each subdetector*[138].

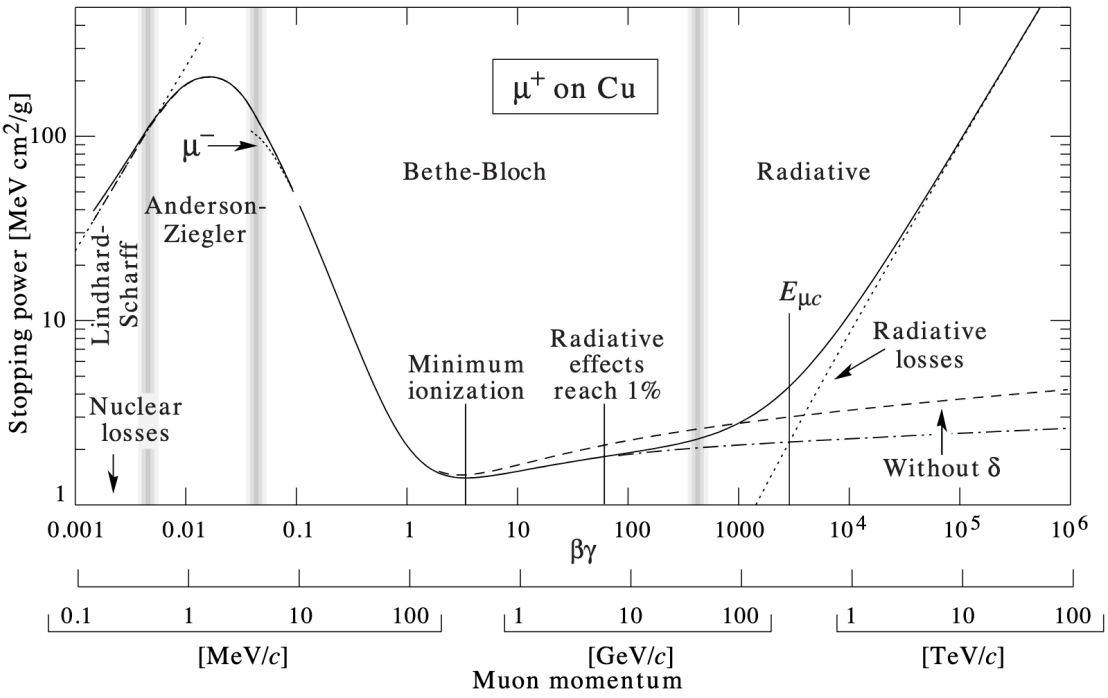


Figure 6.6: *A diagram of the Bethe-Bloche function for muons bombarding copper. The function depends on the number of atoms on the target, but atoms of similar mass have similarly behaved distributions, and the x-axis is similar for charged pions as muons. Near the middle of the curve is the minimum ionizing energy, where the particle travels through materials with minimum losses. Notably, even up to 100 GeV p_{T} muons have stopping power of only 3 MeV cm^2/g, meaning they are negligibly affected by material with a characteristic energy loss*[139].

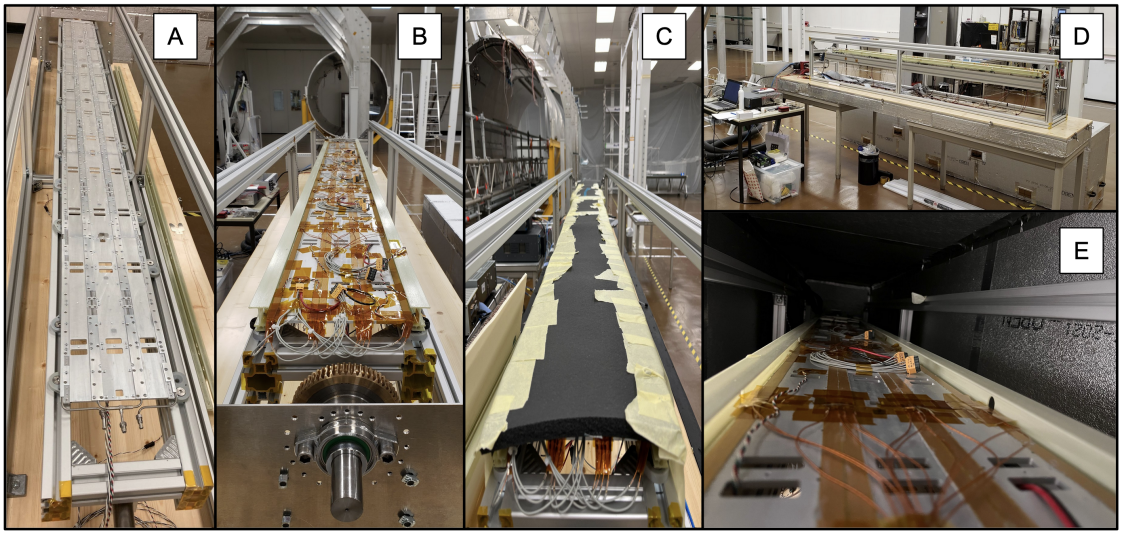


Figure 6.7: *The setup described in the text, located in the Tracker Integration Facility (TIF) at CERN. Image A shows the unmodified tray, and B shows all dummy resistors and sensors attached. Image C shows the tray now with insulating Airex material added, similar to its final conditions when mounted on CMS. Image D shows the entire setup, with an insulating cover in the background, and image E shows what the setup looks like with the insulating cover in place for cooling tests.*[141]

The BTL is segmented into 72 units 2 meters in length called "trays." Each tray holds 6 "concentrator cards" or "boards," which themselves hold 6 modules of LYSO:Ce crystals bonded to SiPMs. Together, these components contribute 600 W of power dissipated over the entire tray. A single tray of BTL was produced and thermally validated from $20°C$ to $-35°C$ to ensure feasibility before final system implementation. Only the tray, passive cooling, and one board were available during the time of testing. To simulate the full thermal load of the other 5 boards, strip resistors or "heaters" were attached to the aluminum tray using thermal paste, and measured with room temperature detectors (RTDs) connected to a readout system. The detectors were placed uniformly along the length and width of the tray, with additional sensors placed at the CO_2 inlet and return to measure temperature change. The setup is shown in Fig. 6.7[141].

The setup was fairly simple in this configuration consisting only of resistors which could be powered up to 600 W, and the CO_2 cooling system. The goal with this setup was to demonstrate that the cooling tray could reach and maintain temperatures near $-35°C$ even

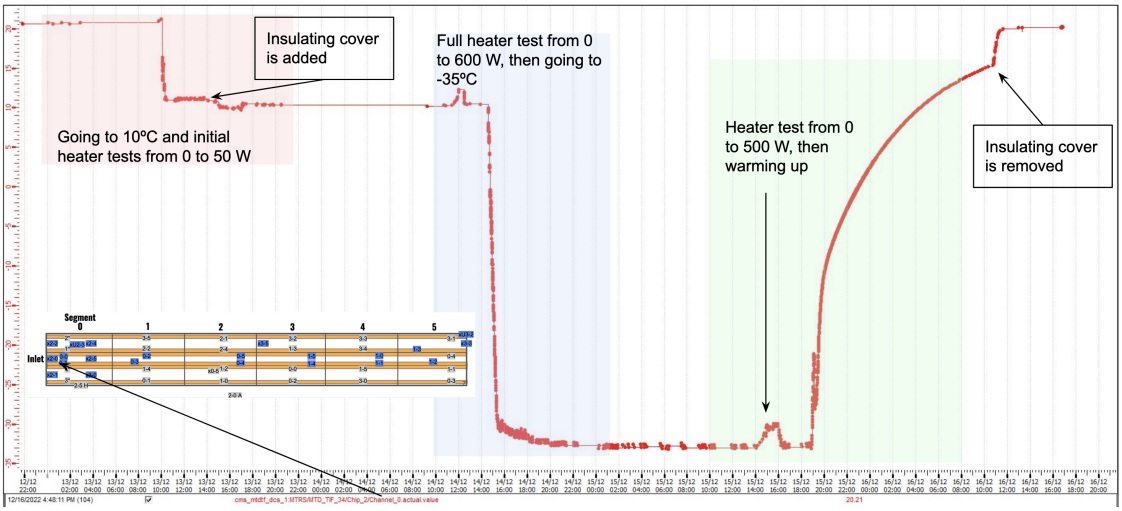


Figure 6.8: *Graph of temperature near the CO_2 inlet for the duration of the thermal validation described in the text*[141].

under thermal load. To operate the setup and collect data the insulating cover was put in place, and dry air was pumped into the enclosed volume to reduce the humidity. This was done to avoid condensation occurring on the tray from the passive CO_2 cooling. Before cooling, the heaters were tested at full power to demonstrate an expected temperature change of about $2°C$. The cooling was turned on for 12 hours prior to data taking, ensuring a near constant temperature across the entire tray. The heaters were then turned on to full power, which was reduced to 500 W due to the temperature dependence of resistivity in copper. Still, with this data it was determined that passive cooling would allow the setup to operate near $-35°C$ even under thermal load. This test measurement, shown in Fig. 6.8 yielded important experience with the experimental setup that contributed to the success of later validations.

Using the same setup as before, a prototype board with modules was added to the tray, replacing one section of the heaters. The validation was performed again, this time using the temperature sensors on the board directly and measuring the difference from the SiPM temperature and tray temperature as the TECs were turned on to full power. The data of this test was compared against another prototype board used for data acquisition testing and the results were found to be similar. Since the prototype components will not be used in the final detector, data was also taken for power values above the nominal operating current,

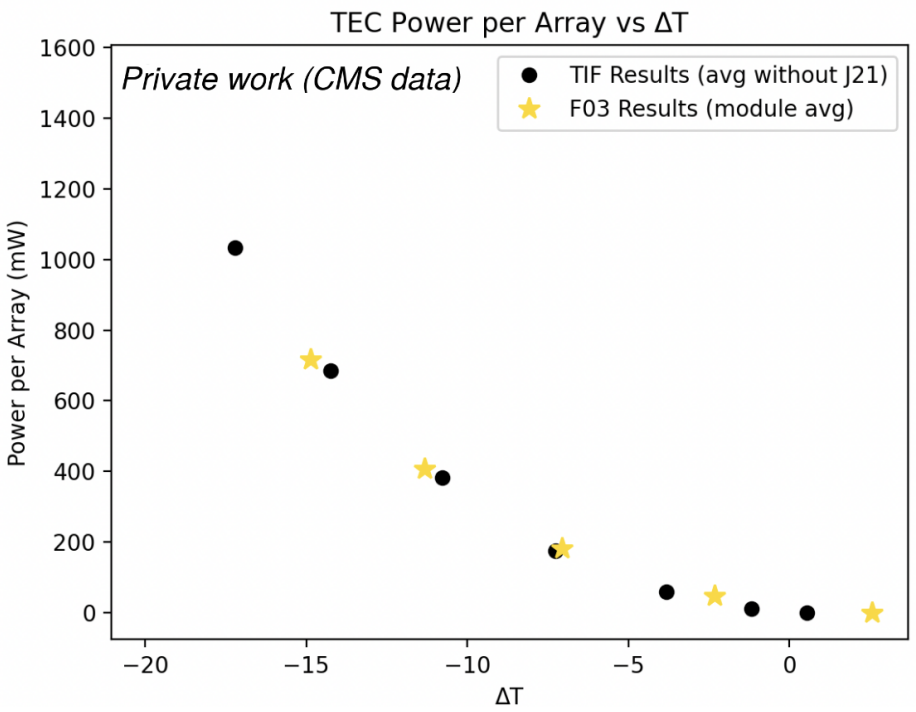


Figure 6.9: *Graph of temperature difference between a SiPM on the prototype board and the cold plate at different current values. The temperature values from module J21, one of the six modules on the board, were omitted from the TIF average because the RTD at that location was found to not be properly connected to the cold tray.*

which show achievable temperature reduction beyond $10°C$ with the TECs.

6.4 Conclusion

This thesis has demonstrated an overview of particle physics suitable to understand a segment of the physics program underway at the CMS on the LHC beamline, with emphasis on topics most interesting to the author. Notably the analysis of $H \to \tau\tau$ decays at CMS was not published or fully unblinded at the time of this writing, but expected results are already promising. Due to improvements in τ_h candidate identification at the online and offline level, the differential measurement of key variables of the Higgs boson are expected to improve by up to 40% in some categories. Though this is a strong improvement, this will not be the last word on the topic. Additional data from Run 3 will be available in the near future, final re-processing campaigns with updated reconstruction will be undertaken before

Run 4, jet tagging and τ_h identification can be further improved[142;143], embedding methods for $Z/\gamma + jets$ backgrounds will be re-implemented for use in Run 3 analyses, machine learning will replace the cut-based JetFakes estimation[144], and additional unknown advances in methodology will be made in the meantime.

Particle physics is an exciting field because almost the moment a measurement is made, it becomes out of date and could be improved with a new method, new dataset, or a new detector. The best one can hope to achieve is what I have documented here: a measurement that is thought out, contains no or minimal mistakes, and successfully taught the researcher something along the way.

Bibliography

[1] N. Bohr. On the Constitution of Atoms and Molecules. *Phil. Mag. Ser. 6*, 26, 1913. doi: 10.1080/14786441308634955.

[2] Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. W. H. Freeman, San Francisco, 1973. ISBN 978-0-7167-0344-0, 978-0-69117779-3.

[3] Peter J. Mohr, David B. Newell, Barry N. Taylor, and Eite Tiesinga. CODATA recommended values of the fundamental physical constants: 2022*. *Rev. Mod. Phys.*, 97 (2), 2025. doi: 10.1103/RevModPhys.97.025002.

[4] V. Mukhanov. *Physical Foundations of Cosmology*. Cambridge University Press, Oxford, 2005. ISBN 978-0-521-56398-7. doi: 10.1017/CBO9780511790553.

[5] Franz Mandl and Graham Shaw. *Quantum Field Theory*. Wiley, 1985.

[6] David Griffiths. *Introduction to elementary particles*. Cambridge University Press, 2008. ISBN 978-3-527-40601-2.

[7] CMS Collaboration. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716, 2012. doi: 10.1016/j.physletb.2012.08.021.

[8] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716, 2012. doi: 10.1016/j.physletb.2012.08.020.

[9] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3), 2024. doi: 10.1103/PhysRevD.110.030001.

[10] Wikipedia contributors. Standard Model Diagram. https://en.wikipedia.org/w/index.php?title=Standard_Model&oldid=1367543769, 2026. [Online; accessed 4-August-2026].

[11] Andrei V. Gritsan, Raoul Rontsch, Markus Schulze, and Meng Xiao. Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques. *Phys. Rev. D*, 94(5), 2016. doi: 10.1103/PhysRevD.94.055023.

[12] CMS Collaboration. High-level hadronic tau lepton triggers of the CMS experiment in proton-proton collisions at $\sqrt{s} = 13.6$ TeV. *JINST*, 21(04), 2026. doi: 10.1088/1748-0221/21/04/P04002.

[13] SNO Collaboration. Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89, 2002. doi: 10.1103/PhysRevLett.89.011301.

[14] James Joyce. *Finnegans Wake*. Faber and Faber, 1939.

[15] Nora Brambilla, Simon Eidelman, Christoph Hanhart, Alexey Nefediev, Cheng-Ping Shen, Christopher E. Thomas, Antonio Vairo, and Chang-Zheng Yuan. The XYZ states: experimental and theoretical status and perspectives. *Phys. Rept.*, 873, 2020. doi: 10.1016/j.physrep.2020.05.001.

[16] Wikipedia contributors and Hugo Spinelli. Graphic of Key Terms in Particle Physics. https://simple.wikipedia.org/wiki/Hadron#/media/File:Bosons-Hadrons-Fermions-RGB-png2.png, 2026. [Online; accessed 13-August-2026].

[17] Michael E. Peskin. What is the Hierarchy Problem? *Nucl. Phys. B*, 1018, 2025. doi: 10.1016/j.nuclphysb.2025.116971.

[18] ALPHA Collaboration. Observation of the effect of gravity on the motion of antimatter. *Nature*, 621(7980), 2023. doi: 10.1038/s41586-023-06527-1.

[19] Peter Ade et al. The Simons Observatory: Science goals and forecasts. *JCAP*, 02, 2019. doi: 10.1088/1475-7516/2019/02/056.

[20] Csaba Csaki. The Minimal supersymmetric standard model (MSSM). *Mod. Phys. Lett. A*, 11, 1996. doi: 10.1142/S021773239600062X.

[21] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83, 1999. doi: 10.1103/PhysRevLett.83.3370.

[22] C. E. Swartz. *The Fundamental Particles*. Addison-Wesley Publishing Company, 1965.

[23] Luigi Di Lella and Carlo Rubbia. 60 Years of CERN Experiments and Discoveries: The Discovery of the W and Z Particles. https://cds.cern.ch/record/2103277/files/9789814644150_0006.pdf, 2015. [Online; accessed 4-August-2026].

[24] Rudiger Schmidt. *Challenges and Goals for Accelerators in the XXI Century*, chapter 10. World Scientific, doi:10.1142/9789814436403_0010, 2021.

[25] P. Lebrun S. Myers R. Ostojic J. Poole P. Proudlock O. Bruning, P. Collier, editor. *LHC Design Report - Volume I: The LHC Main Ring*. CERN, 2004. doi: 10.5170/CERN-2004-003-V-1.

[26] Lyndon Evans, editor. *The Large Hadron Collider: A Marvel of Technology*. EPFL Press, 2009. ISBN 978-1-4398-0401-8.

[27] CERN. Our History. https://home.cern/about/who-we-are/our-history/, 2026. [Online; accessed 4-August-2026].

[28] Rudiger Schmidt. *Challenges and Goals for Accelerators in the XXI Century*, chapter 4. World Scientific, doi:10.1142/9789814436403_0004, 2021.

[29] Stephen Holmes, Ronald S. Moore, and Vladimir Shiltsev. Overview of the Tevatron Collider Complex: Goals, Operations and Performance. *JINST*, 6, 2011. doi: 10.1088/1748-0221/6/08/T08001.

[30] CERN. The CERN Accelerator Complex. https://cds.cern.ch/record/2800984/files/CCC-v2022.png, 2022. [Online; accessed 4-August-2026].

[31] M. G. Holloway and C. P. Baker. How the Barn was Born. https://physicstoday.aip.org/letters/how-the-barn-was-born, 1972. [Online; accessed 5-August-2026].

[32] CMS Collaboration. proton-proton collisions since 2010 (Run 1 + Run 2 + Run 3). https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, 2026. [Online; accessed 6-August-2026].

[33] Werner Herr and Bruno Muratori. The Concept of Luminosity. https://cds.cern.ch/record/941318/files/p361.pdf, 2003. [Online; accessed 5-August-2026].

[34] CMS Collaboration. Output rates and bandwidth of the High-Level Trigger during 2025 proton-proton runs. 2026.

[35] Chris Quigg. Long Live the Tevatron. https://cerncourier.com/a/long-live-the-tevatron/, 2011.

[36] O. Bruning and L. Rossi. *The High Luminosity Large Hadron Collider – HL-LHC*, chapter Chapter 1. World Scientific, 2024. doi: 10.1142/9789811278952_0001.

[37] ALICE Collaboration. ALICE: Technical proposal for a large ion collider experiment at the CERN LHC. 1995.

[38] LHCb Collaboration. LHCb technical proposal: A Large Hadron Collider Beauty Experiment for Precision Measurements of CP Violation and Rare Decays. 1998.

[39] ATLAS Collaboration. ATLAS: Technical proposal for a general-purpose p p experiment at the Large Hadron Collider at CERN. 1994. doi: 10.17181/CERN.NR4P.BG9K.

[40] CMS Collaboration. CMS, the Compact Muon Solenoid: Technical proposal. 1994.

[41] CMS Collaboration. Extracting the speed of sound in quark–gluon plasma with ultrarelativistic lead–lead collisions at the LHC. *Rept. Prog. Phys.*, 87(7), 2024. doi: 10.1088/1361-6633/ad4b9b.

[42] CMS Collaboration. Observation of $\gamma\gamma \to \tau\tau$ in proton-proton collisions and limits on the anomalous electromagnetic moments of the τ lepton. *Rept. Prog. Phys.*, 87(10), 2024. doi: 10.1088/1361-6633/ad6fcb.

[43] Tai Sakuma. Cutaway diagrams of CMS detector. https://cds.cern.ch/record/2665537/files/, 2019. CMS-OUTREACH-2019-001.

[44] K. Eggert and W. Hoogland. The LHC Experimental Programme. *Europhys. News*, 23, 1992.

[45] Alayar Kangarlu. Physics of high field mri and applications to brain tumor imaging. In Herbert B. Newton and Ferenc A. Jolesz, editors, *Handbook of Neuro-Oncology NeuroImaging*. Academic Press, San Diego, 2008. ISBN 978-0-12-370863-2. doi: https://doi.org/10.1016/B978-012370863-2.50023-3.

[46] D. W. McRobbie. Occupational exposure in MRI, 2012.

[47] Izaak Neutelings. CMS coordinate system - Graphics with TikZ in LaTeX. https://tikz.net/axis3d_cms/, 2021.

[48] J.D. Jackson. Particle Data Group: Reviews, Tables, and Plots - Kinematics. https://pdg.lbl.gov/2026/reviews/rpp2026-rev-kinematics.pdf, 2000.

[49] D. Kovalskyi, M. Tadel, A. Mrak-Tadel, B. Bellenot, V. Kuznetsov, C. D. Jones, L. Bauerdick, M. Case, J. Mulmenstadt, and A. Yagil. Fireworks: A Physics Event Display for CMS. *J. Phys. Conf. Ser.*, 219, 2010. doi: 10.1088/1742-6596/219/3/032014.

[50] Thomas McCauley. Dimuon event information derived from the Run2010B public Mu dataset. https://opendata.cern.ch/docs/about, 2014. CERN Open Data Portal.

[51] CMS Collaboration. CMS Collaboration Publications. https://cms-results.web.cern.ch/cms-results/public-results/publications/index.html, 2026.

[52] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3, 2008. doi: 10.1088/1748-0221/3/08/S08004.

[53] CMS Collaboration. Development of the CMS detector for the CERN LHC Run 3. *JINST*, 19, 2024. doi: 10.1088/1748-0221/19/05/P05064.

[54] CMS Collaboration. Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 15, 2020. doi: 10.1088/1748-0221/15/10/P10017.

[55] CMS Collaboration. The CMS trigger system. *JINST*, 12, 2017. doi: 10.1088/1748-0221/12/01/P01020.

[56] CMS Collaboration. Performance of the CMS high-level trigger during LHC Run 2. *JINST*, 19, 2024. doi: 10.1088/1748-0221/19/11/P11021.

[57] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *JINST*, 16, 2021. doi: 10.1088/1748-0221/16/05/P05014.

[58] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$= 13 TeV. *JINST*, 13, 2018. doi: 10.1088/1748-0221/13/06/P06015.

[59] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9, 2014. doi: 10.1088/1748-0221/9/10/P10009.

[60] Siona Ruth Davis. Interactive Slice of the CMS detector. https://cds.cern.ch/record/2205172/files/, 2016. CMS-OUTREACH-2016-027.

[61] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12, 2017. doi: 10.1088/1748-0221/12/10/P10003.

[62] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_{T} jet clustering algorithm. *JHEP*, 04, 2008. doi: 10.1088/1126-6708/2008/04/063.

[63] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet user manual. *Eur. Phys. J. C*, 72, 2012. doi: 10.1140/epjc/s10052-012-1896-2.

[64] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *JINST*, 15, 2020. doi: 10.1088/1748-0221/15/09/p09018.

[65] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup Per Particle Identification. *JHEP*, 10, 2014. doi: 10.1007/JHEP10(2014)059.

[66] CMS Collaboration. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13, 2018. doi: 10.1088/1748-0221/13/10/P10005.

[67] Izaak Neutelings. Hadronic tau decay - Graphics with TikZ in LaTeX. https://tikz.net/tau_decay/, 2017.

[68] CMS Collaboration. Identification of hadronic tau lepton decays using a deep neural network. *JINST*, 17, 2022. doi: 10.1088/1748-0221/17/07/P07023.

[69] Huilin Qu and Loukas Gouskos. Jet tagging via particle clouds. *Phys. Rev. D*, 101, 2020. doi: 10.1103/PhysRevD.101.056019.

[70] Huilin Qu, Congqiao Li, and Sitian Qian. Particle Transformer for Jet Tagging. 2022.

[71] CMS Collaboration. Comparison of the performance of tau reconstruction and identification algorithms in Run 3. Technical report, 2025.

[72] Marzena Lapka and Ewa Lopienska. Facts about the CMS Experiments. https://cds.cern.ch/record/2687235?ln=en, 2019.

[73] Freya Blekman. Same LHC, same CMS, more physics. https://cms.cern/news/same-lhc-same-cms-more-physics, 2024.

[74] J. Troska, G. Cervelli, F. Faccio, K. Gill, R. Grabit, R. Macias Jareno, A. M. Sandvik, and F. Vasey. Optical readout and control systems for the CMS tracker. *IEEE Trans. Nucl. Sci.*, 50, 2003. doi: 10.1109/TNS.2003.815124.

[75] CMS Collaboration. CMS: The computing project. Technical design report. 2005.

[76] Worldwide LHC Computing Grid Collaboration. Tiers of WLCG. https://wlcg-public.web.cern.ch/tiers, 2026.

[77] CMS Collaboration. CMSSW: CMS Offline Software. https://cms-sw.github.io/, 2026.

[78] CMS Collaboration. CMS event displays of Higgs to two photon candidate. https://cds.cern.ch/record/2736135, 2020.

[79] 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, 2014. doi: 10.1007/JHEP07(2014)079.

[80] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11, 2004. doi: 10.1088/1126-6708/2004/11/040.

[81] 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, 2010. doi: 10.1103/PhysRevD.81.075022.

[82] Christian Bierlich et al. A comprehensive guide to the physics and usage of PYTHIA 8.3. *SciPost Phys. Codeb.*, 2022, 2022. doi: 10.21468/SciPostPhysCodeb.8.

[83] Stefan Gieseke, P. Stephens, and Bryan Webber. New formalism for QCD parton showers. *JHEP*, 12, 2003. doi: 10.1088/1126-6708/2003/12/045.

[84] Enrico Bothmann et al. Event generation with Sherpa 3. *JHEP*, 12, 2024. doi: 10.1007/JHEP12(2024)156.

[85] S. Agostinelli et al. GEANT4—A simulation toolkit. *Nucl. Instrum. Meth. A*, 506, 2003. doi: 10.1016/S0168-9002(03)01368-8.

[86] Mark Thomson. *Modern particle physics*. Cambridge University Press, New York, 2013. ISBN 978-1-107-03426-6, 978-1-139-52536-7. doi: 10.1017/CBO9781139525367.

[87] CMS Collaboration. A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature*, 607(7917), 2022. doi: 10.1038/s41586-022-04892-x. [Erratum: Nature 623, (2023)].

[88] Bruce Mellado Garcia, Pasquale Musella, Massimiliano Grazzini, and Robert Harlander. CERN Report 4: Part I Standard Model Predictions. 2016. csv file at https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR.

[89] CMS Collaboration. Combined measurements and interpretations of Higgs boson production and decay in proton-proton collisions at $\sqrt{s} = 13$ TeV. 2026.

[90] LHC Higgs WG1. Predictions for Production Cross Sections of the Higgs Boson at the LHC and HL-LHC. Technical report, 2026.

[91] Nicolas Berger et al. Simplified template cross section — Stage 1.1 and 1.2. 2019. doi: 10.21468/SciPostPhysCommRep.15.

[92] CMS Collaboration. Search for Neutral Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 713, 2012. doi: 10.1016/j.physletb.2012.05.028.

[93] CMS Collaboration. Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector. *Phys. Lett. B*, 779, 2018. doi: 10.1016/j.physletb.2018.02.004.

[94] CMS Collaboration. Constraints on anomalous $H \to VV$ couplings from the production of Higgs bosons decaying to τ lepton pairs. *Phys. Rev. D*, 100(11), 2019. doi: 10.1103/PhysRevD.100.112002.

[95] CMS Collaboration. Measurements of Higgs boson production in the decay channel with a pair of τ leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 83(7), 2023. doi: 10.1140/epjc/s10052-023-11452-8.

[96] CMS Collaboration. Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 06, 2022. doi: 10.1007/JHEP06(2022)012.

[97] CMS Collaboration. Measurement of the production cross section of a Higgs boson with large transverse momentum in its decays to a pair of τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 857, 2024. doi: 10.1016/j.physletb.2024.138964.

[98] CMS Collaboration. Analysis of the CP structure of the Yukawa coupling between the Higgs boson and tau leptons in proton-proton collisions at $\sqrt{s} = 13.6$ TeV. 2026.

[99] CMS Collaboration. Measurement of the inclusive and differential Higgs boson production cross sections in the decay mode to a pair of τ leptons in pp collisions at $\sqrt{s}$ = 13 TeV. *Phys. Rev. Lett.*, 128(8), 2022. doi: 10.1103/PhysRevLett.128.081805.

[100] ATLAS Collaboration. Search for the Standard Model Higgs boson in the H to $\tau^+\tau^-$ decay mode in $\sqrt{s} = 7$ TeV pp collisions with ATLAS. *JHEP*, 09, 2012. doi: 10.1007/JHEP09(2012)070.

[101] ATLAS Collaboration. Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector. *JHEP*, 04, 2015. doi: 10.1007/JHEP04(2015)117.

[102] ATLAS Collaboration. Search for the Standard Model Higgs boson produced in association with a vector boson and decaying into a tau pair in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev. D*, 93(9), 2016. doi: 10.1103/PhysRevD.93.092005.

[103] ATLAS Collaboration. Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 99, 2019. doi: 10.1103/PhysRevD.99.072001.

[104] ATLAS Collaboration. Measurements of Higgs boson production cross-sections in the

$H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 08, 2022. doi: 10.1007/JHEP08(2022)175.

[105] ATLAS Collaboration. Differential cross-section measurements of Higgs boson production in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 03, 2025. doi: 10.1007/JHEP03(2025)010.

[106] ATLAS Collaboration. Probing the Higgs boson CP properties in vector-boson fusion production in the $H \to \tau^+\tau^-$ channel with the ATLAS detector. *JHEP*, 10, 2025. doi: 10.1007/JHEP10(2025)092.

[107] CMS Collaboration. Identification of tau leptons using a convolutional neural network with domain adaptation. *JINST*, 20(12), 2025. doi: 10.1088/1748-0221/20/12/P12032.

[108] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *JINST*, 15(09), 2020. doi: 10.1088/1748-0221/15/09/P09018.

[109] Artur Kalinowski and Wiktor Matyszkiewicz. Efficient tau-pair invariant mass reconstruction with simplified matrix element techniques. *Nucl. Instrum. Meth. A*, 1086, 2026. doi: 10.1016/j.nima.2026.171318.

[110] CMS Collaboration. Luminosity measurement in proton-proton collisions at 13.6 TeV in 2022 at CMS. CMS Physics Analysis Summary CMS-PAS-LUM-22-001, 2024. URL https://cds.cern.ch/record/2890833.

[111] CMS Collaboration. Measurement of the offline integrated luminosity for the CMS proton-proton collision dataset recorded in 2023. Technical Report CMS-DP-2024-068, 2024. URL https://cds.cern.ch/record/2904808. CMS Detector Performance Note.

[112] CMS Collaboration. Luminosity determination using Z boson production at the CMS experiment. *Eur. Phys. J. C*, 84, 2024. doi: 10.1140/epjc/s10052-023-12268-2.

[113] CMS Collaboration. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS. *Eur. Phys. J. C*, 81, 2021. doi: 10.1140/epjc/s10052-021-09538-2.

[114] Marco Peruzzi, Giovanni Petrucciani, and Andrea Rizzi. The NanoAOD event data format in CMS. *J. Phys. : Conf. Ser.*, 1525(1), 2020. doi: 10.1088/1742-6596/1525/1/012038.

[115] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11, 2007. doi: 10.1088/1126-6708/2007/11/070.

[116] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06, 2010. doi: 10.1007/JHEP06(2010)043.

[117] John M. Campbell, R. Keith Ellis, Paolo Nason, and Emanuele Re. Top-Pair Production and Decay at NLO Matched with Parton Showers. *JHEP*, 04, 2015. doi: 10.1007/JHEP04(2015)114.

[118] Adam Kardos, Paolo Nason, and Carlo Oleari. Three-jet production in POWHEG. *JHEP*, 04, 2014. doi: 10.1007/JHEP04(2014)043.

[119] Christian Bierlich et al. A comprehensive guide to the physics and usage of PYTHIA 8.3. *SciPost Phys. Codeb.*, 2022, 2022. doi: 10.21468/SciPostPhysCodeb.8.

[120] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191, 2015. doi: 10.1016/j.cpc.2015.01.024.

[121] John Allison et al. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53, 2006. doi: 10.1109/TNS.2006.869826.

[122] J. Allison et al. Recent developments in Geant4. *Nucl. Instrum. Meth. A*, 835, 2016. doi: 10.1016/j.nima.2016.06.125.

[123] ROOT Programming Language. ROOT v6-32 - Reference Guide - LorentzVector Class Template. https://root.cern/doc/v632/classROOT_1_1Math_1_1LorentzVector.html#abef996904d310e997da4160c7ffa509c, 2026.

[124] CMS Data Preservation and Open Access group. CMS Open Data Workshop lessons. https://cms-opendata-workshop.github.io/workshop2024-lesson-physics-objects/03-muons.html, 2026.

[125] CMS Collaboration. An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data. *JINST*, 14(06), 2019. doi: 10.1088/1748-0221/14/06/P06032.

[126] E. Alvarez, L. Da Rold, M. Szewc, A. Szynkman, S. A. Tanco, and T. Tarutina. Improvement and generalization of ABCD method with Bayesian inference. *SciPost Phys. Core*, 7, 2024. doi: 10.21468/SciPostPhysCore.7.3.043.

[127] CMS Collaboration. Measurement of the $Z\gamma^* \to \tau\tau$ cross section in pp collisions at $\sqrt{s} = 13$ TeV and validation of τ lepton analysis techniques. *Eur. Phys. J. C*, 78(9), 2018. doi: 10.1140/epjc/s10052-018-6146-9.

[128] CMS Collaboration. The CMS Statistical Analysis and Combination Tool: Combine. *Comput. Softw. Big Sci.*, 8, 2024. doi: 10.1007/s41781-024-00121-4.

[129] W. Verkerke and D. Kirkby. The RooFit toolkit for data modeling. In *Proc. 13th International Conference on Computing in High Energy and Nuclear Physics (CHEP 2003): La Jolla CA, United States, March 24–28, 2003*, 2003. [eConf C0303241 (2003) MOLT007].

[130] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf. The RooStats project. In *Proc. 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT*

2010): Jaipur, India, February 22–27, 2010, 2010. doi: 10.22323/1.093.0057. [PoS (ACAT2010) 057].

[131] Albert M Sirunyan et al. Measurement of the inclusive and differential Higgs boson production cross sections in the leptonic WW decay mode at $\sqrt{s} = 13$ TeV. *JHEP*, 2021. doi: 10.1007/JHEP03(2021)003.

[132] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 2011. doi: 10.1140/epjc/s10052-011-1554-0. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[133] CMS Collaboration. CMS diversity and inclusion statistics from 2016 to 2023. 2024.

[134] CMS Collaboration. Performance of Hadronic Tau Reconstruction at the High Level Trigger using 2024 Data from the CMS Experiment at CERN. 2026.

[135] LHC Collaboration. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN, 2020. doi: https://doi.org/10.23731/CYRM-2020-0010. https://e-publishing.cern.ch/index.php/CYRM/issue/view/127.

[136] High Luminosity LHC project group. HL LHC Project Schedule. https://project-hl-lhc-industry.web.cern.ch/content/project-schedule, 2026.

[137] F. Hartmann, B. Hippolyte, F. Lanni, T. Nayak, C. Parkes, and P. Rumerio. *Upgrade of the Experimental Detectors for High Luminosity LHC*, chapter Chapter 3. World Scientific, 2024. doi: 10.1142/9789811278952_0003.

[138] Joel N. Butler and Tommaso Tabarelli de Fatis. A MIP Timing Detector for the CMS Phase-2 Upgrade. 2019.

[139] Groom D.E. Bichsel, H. and S.R. Klein. Particle Data Group: Reviews, Tables, and Plots - Passage of Particles Through Matter. https://pdg.lbl.gov/2026/reviews/rpp2026-rev-kinematics.pdf, 2005.

[140] A. Bornheim et al. Integration of thermo-electric coolers into the CMS MTD SiPM arrays for operation under high neutron fluence. *JINST*, 18(08), 2023. doi: 10.1088/1748-0221/18/08/P08020.

[141] B. Allmond. Testing the Cooling and Readout Chain of the Barrel MTD Detector for the HL-LHC Upgrade. https://meetings-archive.aps.org/apr/2023/f12/5/, 2023.

[142] Sarah Koller and Humberto Reyes-Gonzalez. The fundamental limit of jet tagging: Beyond top jets. In *23rd International Workshop on Advanced Computing and Analysis Techniques in Physics Research: Transforming the Scientific Process with AI*, 2026.

[143] ATLAS Collaboration. Transforming jet flavour tagging at ATLAS. *Nature Commun.*, 17(1), 2026. doi: 10.1038/s41467-025-65059-6.

[144] Arzunik Gevorgyan et al. A machine-learning method for the estimation of the background from jets misidentified as hadronic tau lepton decays. 10 2026.