Title

The BEST Thesis: The Boosted Event Shape Tagger, A Search for Vector-like Quarks, and A Real GEM in CMS

Author

Regnery, Brendan James

Publication Date

2023

Peer reviewed|Thesis/dissertation

The BEST Thesis: *The Boosted Event Shape Tagger, A Search for Vector-like Quarks, and A Real GEM in CMS*

By

Brendan James Regnery
Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

Robin Erbacher, Chair

John Conway

Albert De Roeck

Committee in Charge

2023

To my parents and siblings,
who taught me how to live.

To Christopher Webster and Philip Martin,
who taught me to live.

To the Alps,
who taught me what it means to be alive.

Contents

List of Figures

List of Tables

Abstract

The BEST Thesis: *The Boosted Event Shape Tagger, A Search for Vector-like Quarks, and A Real GEM in CMS*

This thesis is a collection of three topics that take place at the Compact Muon Solenoid experiment—a particle detector which observes proton-proton collisions at CERN's Large Hadron Collider. These topics are the installation of GEMs into the CMS experiment, the development of the Boosted Event Shape Tagger, and a search for vector-like quarks.

The LHC is undergoing upgrades which will increase the instantaneous luminosity to 5×10^{34} cm^{-2}s^{-1}, a factor of 2.5 higher than the current maximum value. Therefore, the experiments are implementing upgrades to cope with the augmented particle rates. In the muon system of the Compact Muon Solenoid (CMS) experiment, Gas Electron Multipliers (GEMs) are being installed to complement the existing Cathode Strip Chambers (CSCs). This will provide a more precise measurement of the muon bending angle and thus improve the muon trigger capabilities. GEMs are micro-pattern gaseous detectors with high rate capabilities–ideal for the forward regions of the CMS muon system. In preparation for the LHC Run 3, 144 GEM chambers were installed in the first muon station and are now operational in Run 3. This thesis introduces the GEM technology and discusses the production, installation, commissioning, and operation of the new GEM muon detectors at CMS.

The first GEMs in CMS will improve the identification of muons, but proper identification of hadronic decays requires the development of new analysis tools. Jets from heavy particles (top, Higgs, W, Z) have characteristic patterns that can be identified by Lorentz boosting the jet to various hypothetical rest frames. The Boosted Event Shape Tagger (BEST) is a deep neural network that utilizes this technique to classify heavy particles from QCD background. A version of BEST was previously used for 2016 collision data. This version was improved on for the full Run 2 dataset. In the effort to improve BEST resulted in a method for creating images of jets in rest frames. These images were passed

to a convolutional neural network for classification. This thesis discusses this method and the other improvements to BEST in detail.

The improved version of BEST was used to search for a pair of vector-like quarks in an all hadronic final state in LHC Run 2 data. Vector-like quarks arise in extensions to the Standard Model which aim to solve the gauge hierarchy problem. This search uses BEST to classify collision events into 126 orthogonal regions. The H_T distributions are tested in each region for the presence of signal and exclusion limits are set for T' and B' masses. This search is currently being approved by the CMS experiment, so only expected limits are presented—the expected sensitivity of the search if no signal is present. The process for setting expected limits is completed using Monte Carlo simulated data and data driven estimates, so no collision data from the signal region are included.

Sommario

La tesi BEST: *il Boosted Event Shape Tagger, una ricerca di Vector-like Quarks e un vero GEM in CMS*

Questa tesi è una raccolta di tre argomenti nell'ambito dell'esperimento Compact Muon Solenoid, un rivelatore di particelle installato al Large Hadron Collider del CERN. Gli argomenti trattati sono l'installazione delle GEM nell'esperimento CMS, lo sviluppo del Boosted Event Shape Tagger e la ricerca di vector-like quark.

I miglioramenti apportati ad LHC aumenteranno la luminosità istantanea a 5×10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, un fattore di 2,5 superiore al massimo valore attuale. Pertanto, gli esperimenti stanno implementando upgrade per far fronte all'aumento di flusso di particelle. Nel sistema di muoni dell'esperimento Compact Muon Solenoid (CMS), si stanno installando dei rivelatori con tecnologia Gas Electron Multiplier (GEM) per integrare le Cathode Strip Chambers (CSC) esistenti. Ciò consentirà di misurare con maggiore precisione l'angolo di curvatura dei muoni e quindi di migliorare le capacità di trigger dei muoni. Le GEM sono rivelatori a gas veloci, ideali per le regioni ad elevata pseudo-rapidità del sistema a muoni di CMS. In preparazione al Run 3 di LHC, sono state installate 144 camere GEM nella prima stazione a muoni e sono ora operative sin dall'inizio di Run 3. Questa tesi introduce la tecnologia GEM e discute la produzione, l'installazione, il commissioning e il funzionamento dei nuovi rivelatori di muoni GEM nell'esperimento CMS.

L'argomento successivo trattato nella tesi riguarda lo sviluppo di nuovi strumenti di analisi, atti a una corretta identificazione dei decadimenti adronici. I jet di particelle pesanti (top, Higgs, W, Z) hanno schemi caratteristici che possono essere identificati mediante boosting di Lorentz del jet in vari ipotetici sistemi di riferimento a riposo. Il Boosted Event Shape Tagger (BEST) è una deep neural network che impiega questa tecnica per classificare le particelle pesanti, distinguendole dal fondo QCD. Una versione di BEST è stata utilizzata in precedenza per i dati delle collisioni del 2016. Quella versione è stata migliorata per l'intero set di dati di Run 2. Nel tentativo di migliorare BEST è stato creato un metodo per creare immagini di jet in sistemi di riferimento a riposo.

Queste immagini sono state passate a una rete neurale convoluzionale per la classificazione. Questa tesi illustra in dettaglio il metodo e gli altri miglioramenti apportati a BEST.

La versione migliorata di BEST è stata utilizzata per la ricerca di una coppia di vector-like quark in uno stato finale interamente adronico nei dati del Run 2 di LHC. I vector-like quark sono presenti nelle estensioni del Modello Standard che mirano a risolvere il problema della gerarchia di gauge. Questa ricerca utilizza BEST per classificare gli eventi di collisione in 126 regioni ortogonali. Le distribuzioni H_T sono state testate in ogni regione per verificare la presenza di segnale e sono stati fissati limiti di esclusione per le masse T' e B'. Questa ricerca è attualmente in fase di approvazione da parte della collaborazione CMS, pertanto vengono presentati solo i limiti attesi, ovvero la sensibilità prevista della ricerca in assenza di segnale. Il processo di definizione dei limiti attesi è completato utilizzando dati di simulazione Monte Carlo e stime basate sui dati, quindi non sono inclusi i dati di collisione della regione del segnale.

Acknowledgments

I would like to thank my parents (Jim and Marcia Regnery) and my siblings (Alexander and Kara Regnery). Since I was young, they encouraged my curiosity of the world and let me ramble on about insane ideas. Their love and support through this constant process made this entire PhD possible.

I am extremely grateful to my advisor, Robin Erbacher. She has been there to guide and support me through some of my most difficult days. Robin taught me as much about life as she did about physics. Along the way, she constantly let me explore ideas and discover my true potential as a physicist.

Grazie dal più profondo del cuore a Giovanni Mocellin. Giovanni mi ha insegnato tutto quello che so sulle GEM. Ha visto il potenziale e l'interesse in me quando nessun altro lo vedeva. Questa tesi non sarebbe stata possibile senza di lui. Ora sarà un amico per la vita.

Muchas gracias Johan Sebastian Bonilla Castro. Johan apoyó a todo nuestro grupo en el CERN durante el covid. Me enseñaron mucho sobre los reactores y sobre ser físico. Sin ellos no habría llegado al final de esta tesis.

Grazie dal più profondo del cuore a Michele Bianco. Michele è stato come un padre per il gruppo durante il COVID. Incoraggiava costantemente la mia passione per l'italiano e per il ciclismo, anche quando gli altri mi prendevano in giro. Ricorderò sempre il giorno in cui Michele mi disse: "Nel necrologio di Enrico Fermi c'era scritto 'l'americano nato in Italia', in quello di Brendan ci sarà scritto 'l'italiano nato in America'".

I could not have undertaken this adventure without John Conway. He always challenged me to think deeply about ideas, never settling for a surface-level response. Perhaps most importantly, he taught me that the BEST ideas come up in the most unlikely of ways like when riding a bicycle.

Words cannot express my gratitude to my undergraduate advisor, Darin Acosta. He encouraged me to study ecology and physics at the same time, then mentored me as I discovered my love for particle physics. Without Darin's constant guidance, I would not be at this point.

Muchas gracias Johny Jaramillo. Johny fomentó mi amor por el ciclismo, por Colombia y por el mundo. Amplió mis horizontes en la vida y me animó a recordar siempre a las personas que el mundo parece olvidar.

I am deeply indebted to Will Kalderon. Will invited me on the adventures of a lifetime, the adventures that I have always dreamed of having. Along the way (and often on top of a mountain), Will taught me about statistics, about particle reconstruction, and about how navigate my way in the field. These moments helped make me the person that I always wanted to be. Will also made the beautiful figures in the incredibly important Appendix A.

Grazie di cuore a Roberto Seidita e Caterina Aruta. Io, Cate e Rob abbiamo scritto le nostre tesi inseme. Tanti fine settimana, tante notte lunghe serate, tante pizze, tante conversatione di fisica, e finalemente abbiamo finito !

A special thanks to Samuel Meehan. Sam was one of the first friends I made at CERN. He always made sure to include me, encourage me, and mentor me. Sam helped me through some of the most difficult moments in my life. I will never forget discussing the physics of avalanches with Sam and a random French mountain man while we took a break from the TMB to eat wild blue berries.

Grazie mille a Margherita Boselli e Federica Baldassari. Margherita ha visto la mia passione per le attività di divulgazione e mi ha aiutato a partecipare all'insegnamento agli studenti delle scuole superiori di tutto il mondo. Si è anche assicurata di insegnarmi molto sulla cultura italiana lungo il percorso. Spero che pensi che io sia più figo di miocuggino. Federica mi ha aiutato a scrivere articoli di divulgazione per CMS. Insieme, mi hanno costretto a pensare e spiegare la fisica in modi completamente nuovi.

Thank you to Jared Sturdy. Jared introduced me to GEMs, mentored me during my first year at CERN, and taught me about git and programming with a team.

Thank you to Pieter Everaerts. Pieter mentored me through the last portion of graduate school and taught me incredibly important lessons on how to create, lead, and manage a group of people.

Grazie mille a Federica Maria Simone e Camilla Galloni per avermi sempre aiutato.

Camilla ha insegnato a destreggiarmi nella difficile burocrazia del CERN. Fede mi ha insegnato a conoscere le GEM e si è sempre assicurata che ricevessi un riconoscimento per il mio lavoro.

Merci à Gilles De Lentdecker, Laurent Petre et Daniel Teyssier. Ils m'ont aidé à grandir dans le groupe GEM et à trouver mon intérêt pour les détecteurs.

Thank you to Lucas Corcodilos for teaching me almost everything that I know about the Higgs Combine Tool and mentoring me in being the B2G statistics contact.

Grazie mille a Deborah Pinna. Deborah mi ha fatto da mentore per diventare il fisico che sono oggi.

Thank you to Grace Haza. Grace and I started as PhD students together at Davis, went to CERN together, and worked through the pandemic.

Thank you to Jackson Burzynski. I will never forget skiing throughout Grimentz-Zinal during the COVID-19 pandemic while Jackson taught me about Lie groups and symmetry breaking.

Grazie mille a Tiziana Puopolo. Tiziana è come una sorella maggiore per me; si è presa il tempo di incoraggiarmi a imparare l'italiano e mi ha aiutato a imparare a cooperare in un ambiente internazionale.

Thank you to my partner on analysis, Abhishek Das. Together, we conducted a search for a new particles. We both had long nights getting all of the pieces to work together.

Grazie a Antonello Pellecchia, Francesco Ivone, Nicole Rosi, Davide Fiorina, Vincenzo Mastrapasqua, Simone Calzaferri e Viviana Mele: studenti italini delle GEM e miei amici.

Thank you to Seulgi Kim, Yechan Kang, Dayoung Kung, Nimantha Perera, Kailasapathy Balashangar, Kalpanie Madara Liyanage, Manuel Rodriguez, Felipe Ramirez, Jeremie Merlin, Marcello Maggi, Louis Lemair, and Devin Aebi. This group spans the world and taught me so much about cultures.

Thank you to Peter Bond. Pete helped me more times than I can count: when my car broke down, when I had trouble at work, when I got injured, or when I just needed a friend. In particular, Pete's support during the COVID lockdown made sure that I had a family nearby, when mine was so far away.

Merci à Isabelle De Bruyn pour son aide en français et pour m'avoir inclus dans le groupe CSC. Mais à elle, je dois dire "Never. I'll never turn to the dark side. you've failed, your highness. I am a Jedi..."

Thank you to Samantha Abbott, Ethan Cannaert, Reyer Band, and Yao Yao. It was a pleasure to be a graduate student with this fun group of students, together we learned a lot.

From my time as an undergraduate at the University of Florida, I would like to thank Justin Hugon, Matt Carver, Dimitri Bourilkov, Alexander Madorsky, Vivan Nguyen, Zachary Montague, and Pierluigi Bortignon. A special thank you goes to Vivan who has been at CERN with me and helped support me throughout graduate school.

Vielen Dank an Clemens Lange, Anna Benecke, Andreas Hinzmann, Kinga Wozniak, Tomke Schröer, Silke Möbius, und Knut Zoch. Sie lehrten mich etwas über ihre Kultur und brachten mir gleichzeitig etwas über Statistik, Düsen und Physik bei. Ihre ausführlichen Berichte haben diese Arbeit erst möglich gemacht.

Thank you to Dan Guest, Claire Antel, Bingxuan Liu, Jennifer Roloff, and Matt LeBlanc for many lunches, dinners, and mountain adventures. Matt even had the idea for the name of this thesis.

Grazie mille a Enea Facoetti, Francesco Chiari, Lorenzo Ghilardi e Sarah Fornoni. Loro mi hanno insegnato che "tutto va bene quando sono in montagna con i miei amici."

Thank you to Christine McClean, Justin Pilot, Andrés Delannoy, Reyer Band, and Riju Dasgupta for always including me and for helping me find my place at the University of California, Davis.

Thank you to my extended family around the world. I wish that my grandparents, Richard and Martha Skerl, could see their first grandchild with a PhD. May they rest in peace.

Enfin, merci à Sébastien Paul Tubau pour m'avoir aidé à trouver mon centre, à gérer ma perte, et à franchir la ligne d'arrivée pour terminer cette thèse.

Preface

To all of the graduate students out there, I know that graduate school is difficult, but trust me when I say that you can do it! Graduate school has been a series of hurdles interspersed with the best moments of my life. When I started graduate school, I never expected the path to unfold the way it did.

During my first year of graduate school, one of my close friends died; this moment sent shock waves through my life. Then, in my second year of graduate school, another one of my close friends died just a few days before I was about to fly across the world to work at CERN, away from all of my family.

Chris and Phil were about the same age as me, we had previously lived together for almost 4 years and the experience had a huge impact on me. Chris was in medical school and Phil was in psychology graduate school. I spent a lot of time processing their loss by cycling, hiking, and skiing throughout the Alps. During this time, I felt connected to them and this connection has stayed with me. I think anyone experiencing a loss will understand exactly what I mean.

I will never understand why I am completing my PhD, but they are not. That is why I want this thesis to remember them! The two brilliant goofballs who started as freshmen with me at the University of Florida.

Then in my third year of graduate school, I was really starting to feel like things were turning around and my string of bad luck was about to end. However it was the year 2020 ...

In 2020, the world changed forever with the COVID-19 pandemic and the entire course of this thesis along with it. This terrible disaster affected lives around the world and will define the beginning of the "2020s" in world history textbooks.

I still remember the very beginning of this pandemic. I remember being on a bus with the CERN Ski club, with some of my new friends from across the world. We were returning from Grindelwald and had many outings planned for the rest of the ski season, but the organizers had begun to mention that the season might get cancelled. I thought

this was crazy, how could this possibly be?

The very next day, President Emmanuel Macron addressed all of us residing in France with a speech that resonated deep inside all of us

> *Nous sommes en guerre. Pas contre une autre nation, mais contre un ennemi invisible et insaisissable.*
>
> President Emmanuel Macron

It was at this exact moment that I begun to realize the gravity of the situation. The next day I sat with my friends outside of the main restaurant at CERN as the world shut around us and I did not know when I would see them again or what would befall us in the coming months. Thanks to the kindness of my group, I was given a simple choice, stay here in France or return home not knowing when I would come back.

The choice for me was simple, CERN had become my home and I wanted to stay. I cannot think of a simple choice that has had bigger impacts on my life. My entire PhD and future was shaped by this one moment, where I did not even give a second thought.

Going into graduate school, my main interest was in using machine learning to better understand jets at collider experiments. I had a side interest in detectors, but I had not realized how much this interest was about to grow.

As an American student at CERN, my PhD was supposed to be focused on analysis work with enough detector work to satisfy my service requirements. I was tasked to help with gas electron multipliers at CMS as my service task. But then COVID hit and many students disappeared to their home countries (some for years). Our group hit a labor shortage like was never predicted.

My job as a student who chose to stay during these difficult times was to help as much as possible (and as much as I felt was safe). What resulted was my chance to help in many crucial steps of the GE1/1 detector. I expanded all over the project and worked on production, installation, and commissioning. It was my chance to participate in more activities than typical graduate students ever get to help with. In fact, it made me realize my passion for detectors in a way that I never expected!

The group of us at CERN was small. I became very close my fellow UC Davis graduate

students and to my fellow graduate students in the lab from around the world. This taught me so much about languages, cultures, and humanity. I feel lucky to have had this experience.

In this way, I am grateful for my path as a graduate student. I remember being warned before graduate school "Physics graduate school is very tough". This was true of course and was made even harder by the loss of my friends and the pandemic. However, these situations allowed me to discover my passions and connect with humanity in a way that I never thought possible. And for that, I am incredibly thankful.

So...to anyone out there reading this, here is my story! I hope it can be helpful to you! And if you are a graduate student, just remember that you can make it! "Andiamo!" „Los geht's!" «On y va !»

Chapter 1

Introduction

> *Tutta la storia della vita sulla Terra ci insegna che la «diversità» è un valore fondamentale. La ricchezza della vita, infatti, è dovuta alla sua diversità: diversità di enzimi, di cellule, di piante, di organismi, di animali. Anche per la storia delle idee è stato così. La diversità delle culture, delle filosofie, dei modelli, delle strategie e delle invenzioni ha permesso la nascita e lo sviluppo delle varie civiltà.*
>
> Piero Angela, *Viaggi nella scienza: Il mondo di Quark*

Particle physics is about probabilities, endless possibilities, that make the universe the way it is today. Simply put, particle physics aims to answer one question: what makes up everything? This has plagued humans since the very beginning of recorded history. Answers have been proposed by philosophers, theologists, and scientists. It is a simple question, one that children commonly ask their parents as they develop an awareness of their surroundings. But the answer could not be more complex. It is only the foundation of everything.

In the late 1800s and early 1900s, scientists probed the structure of atoms. This began with J.J Thomson's discovery of the electron and evolved when Ernest Rutherford discovered the dense, positive nucleus at the center of the atom. These discoveries led to complicated theories describing the orbits of electrons and forces binding the nucleus together. Matter was previously described as either particles or waves. However, these new observations of complicated dynamics could only be explained using probabilities

causing the line between particles and waves to blur. Particle dynamics could only be explained by wave functions—this was the rise of quantum mechanics.

As more discoveries were made, physicists began unite all of the fundamental forces and elementary particles into one mathematical theory, the result is the Standard Model of particle physics. This is the most accurate scientific model to date, showing an unprecedented agreement to experimental data. The Standard Model states that matter in the universe consists of fermions (particles of matter) and bosons (force carrying particles). Fermions consist of quarks, which bind together to make hadrons (like protons and neutrons), and leptons, which consist of electron-like particles and neutrinos. The interactions between these particles are described by three fundamental forces: the electromagnetic, weak, and strong forces.

However, the Standard Model is far from a complete theory. One of the major fundamental forces, gravity, is still not described at the quantum level. Cosmological observations of galactic rotations point to the existence of more matter than can be observed. This Dark Matter most likely originates from currently undiscovered particles.

The Standard Model also has a number of theoretical curiosities. One of these is the gauge hierarchy problem which relates to divergent terms in the theory of the Higgs boson. This problem can be solved by building on the Standard Model to create a possible new theory which predicts the existence of new particles. A number of these new theories predict the existence of vector-like quarks, which are capable of interactions not possible for quarks in the Standard Model. These hypothetical particles can be created at high energy particle collisions and have a distinct experimental signature.

Probing the structure of the universe is done via cosmological observation, nuclear experiments, and observing particle collisions. Obtaining new insights into the Standard Model requires utilizing the cutting edge of particle detectors, hence particle physicists are constantly striving to improve accelerator and detector technology in the pursuit of new discoveries. Technological advances in particle physics directly led to the creation of medical imaging devices, cancer therapy, micro-electronics, and even the internet.

With each iteration of experiments, technology greatly improved, but the costs sky-

rocketed. The only way to afford these large experiments was by collaboration. One large scale collaboration began after World War II. In the late 1940s and 1950s, the United States entered a post-war economic boom and began to drive the world's scientific research. War-torn Europe could hardly keep up. Countries had to rebuild, there simply was no possibility for a single nation to keep pace with scientific research. Therefore, in 1954, 12 European countries—some previously at war with each other—united to create the Conseil Européen pour la Recherche Nucléaire (CERN).

Throughout the late twentieth century, CERN competed with the US by creating ever larger accelerators for discovering new particles. This evolved into machines like CERN's Super Proton Synchrotron collider or the the American Tevatron. At the end of the century, the large costs of high energy physics experiments became too much for the US, alone, to afford—the last proposed US super collider, the Super Conducting Super Collider[1], was cut by US congress in the 1990s due to its astronomical cost marking an end to the competition. This left one large particle collider to create the world's highest energy collisions, the Large Hadron Collider (LHC) at CERN.

Out of this grew a world wide collaboration. One in which no single country owns or drives particle physics. This prompted CERN to change its name to the European Organization for Nuclear Research in order to highlight its new place as a global scientific collaboration. Every day at CERN, students and scientists who come from countries around the world collaborate toward one common goal: a better understanding of the fundamental building blocks of the universe. Never before has something like this happened; CERN has become a model for international collaboration.

The current *capolavoro* (masterpiece) of CERN is the LHC, a 27 km long proton-proton collider located 100 m underground on the border of Switzerland and France. Around the LHC are four interaction points where protons collide. At each interaction point is a large experiment which observes the collisions. One of these experiments is the Compact Muon Solenoid (CMS) experiment. This experiment consists of many different types of particle detectors and acts like a giant camera—recording the results of collision events.

[1]The Super Conducting Super Collider would have been larger (87.1 km in circumference) and would have produced higher energy collisions (20 TeV) than the LHC.

In 2018, the LHC finished a three year long period of data-taking called "Run 2" and shut down for three years of upgrades (called Long Shutdown 2). These were the first in a series of upgrades to prepare for the High-Luminosity (HL) LHC. The HL-LHC will pack more protons into higher density bunches, the resulting bunch crossings (collision events) will have many more interactions leading to more data in a shorter amount of time. This will result in rates of detector occupancy never before seen in a particle experiment.

At CMS, one of the many subsystems being upgraded is the muon system—the gas based detector system responsible for detecting muons, heavy electron-like particles. A novel type of gaseous detector called the Gas Electron Multiplier (GEM) is being added to the muon system to cope with the high rates. During Long Shutdown 2, the first layer of GEM detectors was produced, installed, and commissioned for LHC Run 3.

This thesis can be thought of as a collection of stories focused around three main topics: The upgrade of the CMS muon system with GEM detectors, the use of advanced machine learning algorithms to create BEST—an algorithm for identifying hadronic decays of heavy particles, and a search for a new type of particle called vector-like quarks using CMS data taken during LHC Run 2. The thesis structure is outlined below.

Chapter 2 describes the Standard Model in mathematical detail using quantum field theory. This chapter focuses on Electroweak theory and describes the theory by explaining the mathematical implications as the universe cooled after the Big Bang. The resulting Standard Model has five types of physically manifesting bosons and twelve types of fermions. Then the chapter highlights some shortcomings of the Standard Model, in particular the gauge hierarchy problem which motivates vector-like quark theories. The particles and interactions described in this chapter form the foundation for all of the research in this thesis.

Chapter 3 presents a set of possible solutions to the gauge hierarchy problem which predict the existence of a new type of particle, vector-like quarks. The possible interactions and resulting experimental predictions are presented here.

Chapter 4 summarizes the CMS experiment. This cylindrical onion consists of many layers of different types of particle detectors. This chapter aims to summarize each layer.

There are two main objectives in this thesis: to improve this experimental apparatus by adding a new layer and to use this apparatus to search for vector-like quarks.

Chapter 5 is an overview of the physics inside of a gas electron multiplier, a new type of particle detector that is being added to the CMS experiment. Many of the physical processes described in this chapter are also relevant for the other layers of CMS described in chapter 3.

Chapter 6 describes in detail my involvement with the production, installation, commissioning, and operation of GE1/1—a new layer in the CMS experiment. GE1/1 is the result of collaboration with countries around the world and much of my work took place during the COVID-19 pandemic with a small and dedicated team.

Chapter 7 changes topic back to the CMS experiment and explains how signals in the detector are carefully constructed into particle candidates. This process of particle reconstruction creates the objects used to analyze collision events at the LHC forming the building blocks for the Vector-like Quark search.

Chapter 8 dives into the realm of machine learning and describes how these techniques are used to classify clusters of particles in the CMS detector with a neural network called BEST: the Boosted Event Shape Tagger. BEST utilizes the results of special relativity to boost into hypothetical rest frames and analyze the shapes in each frame. A great deal of my work went into attempting to improve BEST with cutting edge neural network techniques. BEST is the central tool used in the vector-like quark search.

Chapters 9, 10, and 11 summarize my work with a small team to search for vector-like quarks using BEST. This search utilizes a data driven background estimation that can be adapted for similar analyses. The results are analyzed using a statistical method developed by CERN physicists to test the new theory in experimental data. This method is described and then the results for the vector-like quark search are presented.

Chapter 2

The Standard Model of Particle Physics

Barrabás llegó a la familia por vía marítima, anotó la niña Clara con su delicada caligrafía.

Isabel Allende, *La Casa de los Espíritus*

Particle physics studies the fundamental interactions of the universe. Everything from the nuclear interactions occurring in the center of stars to simple electron interactions. The Standard Model is currently the best tool for achieving this goal. This model is the result of combined efforts from generations of physicists; it encapsulates centuries of experimental results and is the most accurate scientific model to date.

The Standard Model consists of three generations of quarks and leptons (fermions) along with 5 physically manifesting bosons. Leptons consist of electrons, muons, taus, and three neutrinos. The quarks are fractionally charged particles which are observed in bound states called hadrons. The bosons are mediators of interactions; a number of bosons are present in the Standard Model, 5 of which manifest at our energy scales.

Group theory is the language of the Standard Model and requires some mathematical introduction

2.1 Mathematical Introduction

Physics aims to describe our universe mathematically. Frequently, this requires describing how objects behave under transformations. The mathematical study of this behavior is called Group Theory. It is one of the greatest mathematical achievements—providing underlying structures used in physics, materials science, and chemistry. For particle physics, it is crucial to have a basic understanding of several important groups and definitions [50].

The Standard Model is built on the unitary group and its special unitary subgroup. A unitary group $U(n)$ is defined as:

Definition 2.1.1 *$U(n)$ is the group of all $n \times n$ matrices, T, such that $T^{-1} = T^{\dagger}$.*

In other words, $U(n)$ is the group of all unitary $n \times n$ matrices. It has an important subgroup called the special unitary group $SU(n)$.

Definition 2.1.2 *$SU(n)$ is the group of all $T \in U(n)$ such that $\det(T) = 1$.*

$U(n)$ and $SU(n)$ are both examples of Lie groups. Lie groups represent a large field of study in mathematics; many thorough texts exist on the subject and a complete definition requires a large mathematical background. However, the Standard Model is limited to matrix Lie groups.

Definition 2.1.3 *A matrix Lie group is a group of invertible $n \times n$ matrices on $\mathbb{R}$ or $\mathbb{C}$.*

This structure allows for a mathematical description of continuous motions and thus symmetries in physics [50].

Each Lie group has a corresponding Lie algebra

Definition 2.1.4 *Let G be a matrix Lie group. Then, the Lie algebra is the set of all matrices X such that $e^{tX} \in G$ for all real numbers t.*

The Lie algebra may be thought of as the elements that are infinitesimally close to the identity in a Lie group. In physics, these elements are called generators and describe infinitesimal transformations. Generators become particularly important when used in representation theory [50].

Representation theory formalizes the language of transformations in physics.

Definition 2.1.5 *Let G be a matrix Lie group, V be a vector space, and $g \in G$. Consider a map ϕ such that $\phi : g \to X$ where X is a linear operator on V. Together, ϕ and V are called a representation.*

A representation takes a group element and maps it to a linear operator on a vector space. When the representations of the generators are exponentiated, they describe finite transformations (i.e. the Lie group elements). Thus, objects in physics are vectors in a vector space, transformations of objects are linear operators that act on those vectors, and generators are used to construct those linear operators [50, 51]. This language of transformations makes it possible to mathematically describe particle interactions.

2.2 The Standard Model Gauge Group

Centuries of experiments have led to the observation of symmetries in nature. The observed symmetries of the universe are embedded in our physical models. Mathematically, this is accomplished with various gauge groups. Some of these symmetries are only present at a high energy scale and manifest differently at non-extreme energies—these are called "broken" symmetries. The Standard Model is described mathematically by the gauge group $SU(3) \times SU(2) \times U(1)$ and utilizes both broken and unbroken symmetries [52].

Quantum chromodynamics (QCD) is the theory of strong interactions and is described by $SU(3)$, an unbroken symmetry. This $SU(3)$ theory yields three types of charge called "colour" and describes how quarks interact and bind together into hadrons.

$SU(2) \times U(1)$ describes the electroweak interaction and unlike QCD is a broken symmetry. This broken symmetry is at the heart of the Standard Model, it describes the electroweak force at high values of energy and breaks into two separate forces at the low energy scales we observe in our daily lives. In fact, this clever theory happens to break in exactly the right way to explain all electroweak observations to date [53].

2.3 Electroweak Theory

In 1961, Glashow proposed the idea of using $SU(2) \times U(1)$ to describe electroweak interactions [54]. But there was a problem! This theory required that the bosons mediating

the electroweak interactions were massless, which disagreed with experimental observations of weak interactions. Then in 1964, Higgs, Brout, and Englert proposed a method for spontaneous symmetry breaking by introducing a scalar field with a non-zero vacuum expectation value. The field became known as the "Higgs field" and this method became known as the "Higgs mechanism" [55, 56]. This theoretical breakthrough finally provided an explanation for massive bosons. But the Higgs mechanism is even more encompassing than that, it provides an explanation for the origin of mass in the Standard Model.

In the late 1960s Weinberg and Salam used the Higgs mechanism to describe electroweak symmetry breaking. Thus, describing low energy electromagnetic interactions with the massless photon and weak interactions with the massive W and Z bosons [57, 58]. The resulting theory has incredible experimental agreement—encompassing centuries of results and successfully predicting many observations. Because of this success, Glashow, Salam, and Weinberg were awarded the Nobel Prize in 1979 for electroweak unification.

The Glashow-Salam-Weinberg (GSW) theory consists of three stages during the evolution of the universe:

- The high energy electroweak interaction (i.e. during the early universe).
- The spontaneous symmetry breaking that occurs via the Higgs mechanism as the universe cools.
- The resulting low energy electromagnetic and weak interactions (i.e. the present day).

2.3.1 The High Energy Electroweak Interaction

To start understanding this theory, consider the period of time immediately after the Big Bang. In GSW theory, the very early universe ($t < 10^{-12}$ s) existed at temperatures greater than 10^{16} K where there was one electroweak interaction. In this early universe, the Higgs field existed in an excited state so all of the fermions were massless. These electroweak interactions are described by the kinetic terms in Eq. 2.1 [59, 51].

$$\mathcal{L} = \bar{E}_L(i\not{D})E_L + \bar{e}_R(i\not{D})e_R + \bar{Q}_L(i\not{D})Q_L + \bar{u}_R(i\not{D})u_R + \bar{d}_R(i\not{D})d_R \tag{2.1}$$

In this Lagrangian, there are two important things to disentangle: the covariant derivative and the fermion fields. First, consider the covariant derivative for a $SU(2)$ fermion field

$$D_\mu = \partial_\mu - igA^a_\mu \tau^a - ig' Y B_\mu \,. \tag{2.2}$$

Y represents the $U(1)$ charge called weak hypercharge and τ^a are the generators of $SU(2)$. The τ^a are the normalized Pauli matrices, $\tau^a = \frac{1}{2}\sigma^a$, where the eigenvalue, T^3, of the τ^3 operator is called the weak isospin. A^a_μ and B_μ are the $SU(2)$ and $U(1)$ gauge bosons with coupling constants g and g'. These massless bosons are called Goldstone bosons.

Fermions are chiral particles, meaning that the left- and right-handed fermions are two different representations of the same $SU(2)$ gauge group. In Eq. 2.1, the L and R subscripts denote the left-handed and right-handed fermion fields. The left-handed fields are represented as doublets of $SU(2)$ and the right-handed fields are singlets of $SU(2)$ [51]:

$$E_L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \,, \quad Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L \,, \quad e_R \,, \quad u_R \,, \quad d_R \,. \tag{2.3}$$

This interaction of massless fermions with massless Goldstone bosons is the high energy electroweak interaction.

2.3.2 Spontaneous Symmetry Breaking via the Higgs Mechanism

As the universe cooled below 10^{16} K, the Higgs field settled to a ground state with a non-zero vacuum expectation value and broke $SU(2) \times U(1)$ symmetry [59]. This relaxation of the Higgs field is shown in Fig. 2.1. The unique shape of the Higgs potential is key to the mass inducing symmetry breaking.

The Higgs field is a scalar field, ϕ, that has a non-zero vacuum expectation value. One way of describing this field is by

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix} \,. \tag{2.4}$$

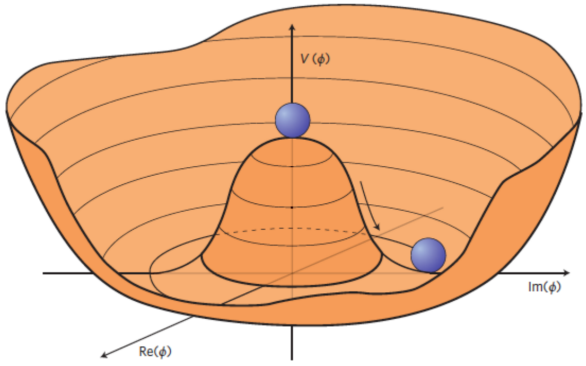

Figure 2.1: The Higgs potential with a non-zero vacuum expectation value. Figure from [1].

This scalar field undergoes the gauge transformation

$$\phi \to e^{i\alpha^a \tau^a} e^{i\beta/2} \phi \,, \tag{2.5}$$

with $\alpha^1 = \alpha^2 = 0$ and $\alpha^3 = \beta$. Thus, $\langle\phi\rangle$ is invariant. This causes the emergence of one massless boson (the photon) and three massive bosons. The relevant terms for determining the masses are

$$\Delta\mathcal{L} = (D_\mu \phi)^2 = \frac{1}{2}\begin{pmatrix} 0 & \nu \end{pmatrix} (g A^a_\mu \tau^a + \frac{1}{2} g' B_\mu)(g A^{b\mu} \tau^b + \frac{1}{2} g' B^\mu) \begin{pmatrix} 0 \\ \nu \end{pmatrix} , \tag{2.6}$$

which reduces to

$$\Delta\mathcal{L} = \frac{1}{2}\frac{\nu^2}{4}[g^2 (A^1_\mu)^2 + g^2 (A^2_\mu)^2 + (-g A^3_\mu + g' B_\mu)^2] \,. \tag{2.7}$$

This equation describes three massive vector bosons and the massless photon [51]. They are

$$\text{The W bosons} \quad W^\pm_\mu = \frac{1}{\sqrt{2}}(A^1_\mu \mp i A^2_\mu) \,, \qquad \text{with} \quad m_W = g\frac{\nu}{2} \,, \tag{2.8}$$

$$\text{The Z boson} \quad Z^0_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(g A^3_\mu - g' B_\mu) \,, \qquad \text{with} \quad m_Z = (\sqrt{g^2 + g'^2})\frac{\nu}{2} \,, \tag{2.9}$$

$$\text{The photon} \quad A_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(g' A^3_\mu + g B_\mu) \,. \tag{2.10}$$

The universe cooled from a hot state with one electroweak interaction to the universe we know today with two interactions: electromagnetic and weak [51]. But what about the weak hypercharge and weak isospin? How are these quantities relevant to our observable quantities?

To answer this question, let's return to Eq. 2.2, the covariant derivative for a Fermion field. Using the eigenstates of the broken $SU(2)\times U(1)$ symmetry, the covariant derivative can be rewritten as

$$D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}}(W_\mu^+\tau^+ + W_\mu^-T^-) - i\frac{1}{\sqrt{g^2+g'^2}}Z_\mu(g^2\tau^3 - g'^2Y) + i\frac{gg'}{\sqrt{g^2+g'^2}}A_\mu(\tau^3+Y)\,, \tag{2.11}$$

with $\tau^\pm = \tau^1 \pm i\tau^2$. The last term in this covariant derivative corresponds to the electromagnetic interaction. From this term, the electron charge and corresponding quantum number can be identified

$$e = \frac{gg'}{\sqrt{g^2+g'^2}}\,, \tag{2.12}$$

$$Q = T^3 + Y\,. \tag{2.13}$$

Thus, the electric charge quantum number is the sum of weak isospin and weak hypercharge. Also, a new measurable parameter can be introduced called the Weinberg angle θ_w. This angle is defined such that

$$cos\theta_w = \frac{g}{\sqrt{g^2+g'^2}}\,, \qquad\qquad sin\theta_w = \frac{g'}{\sqrt{g^2+g'^2}}\,. \tag{2.14}$$

This allows the change of basis relating Eq. 2.2 to Eq. 2.11 to be written as

$$\begin{pmatrix} Z^0 \\ A \end{pmatrix} = \begin{pmatrix} cos\theta_w & -sin\theta_w \\ sin\theta_w & cos\theta_w \end{pmatrix} \begin{pmatrix} A^3 \\ B \end{pmatrix}\,, \tag{2.15}$$

and as a result

$$g = \frac{e}{sin\theta_w}\,, \tag{2.16}$$

$$m_W = m_Z cos\theta_w\,. \tag{2.17}$$

This means that electroweak interactions can be written in terms of three measurable parameters: e, θ_w, and m_w [51]. The electron charge, e, was first experimentally measured by Millikan's oil drop experiment in 1909 [60]. The Weinberg angle, θ_w, was measured by many different experimental observations, such as: neutrino-proton scattering, neutrino-electron scattering, and the Z branching ratio [51]. The measured value is $\theta_w = 28.75^{\text{o}}$ [53]. The last missing piece is the W boson mass which was measured at CERN (along with the Z boson) in 1983 at $m_W = 82\,\text{GeV}$. GSW theory predicts then that the Z mass should be 93.5 GeV which is in good agreement with CERN's measured value of 91.2 GeV [53, 61, 62]. Now, the Tevatron collider and LHC have measured all of these parameters at even greater precision and the observed values continue to show good agreement with the theory.

Fermion Masses

The Higgs mechanism was motivated by the need to include massive bosons in the Standard Model. But fermion chirality adds in some complications for accounting for fermion masses. The left- and right-handed fermions belong to different representations of $SU(2)$ preventing the addition of simple mass terms [51]. Once again, the Higgs mechanism provides a solution to this dilemma.

To gain an intuition for the fermion mass terms, consider a full field $\psi = \psi_L + \psi_R$ and the Dirac Lagrangian $\mathcal{L} = \bar{\psi}(\not{p} - m)\psi$. This results in

$$\mathcal{L} = \bar{\psi}_L \not{p} \psi_L + \bar{\psi}_R \not{p} \psi_R + m(\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L)\,. \tag{2.18}$$

The intuition here is that a mass term will be of the form $m(\bar{\psi}_L \psi_R + h.c.)$ [59]. Such a term arises from the interaction of the Higgs field with the fermion fields. Consider the interaction of Higgs field and the first generation leptons. This results in

$$\Delta\mathcal{L}_e = -\lambda_e \bar{E}_L \cdot \phi e_R + h.c. \tag{2.19}$$

$$= -\frac{1}{\sqrt{2}} \lambda_e \nu \bar{e}_L e_R + h.c. + \dots\,, \tag{2.20}$$

where λ_e is a new dimensionless coupling constant (the Yukawa coupling) and ϕ was

replaced with the vacuum expectation value. Thus,

$$m_e = \frac{1}{\sqrt{2}}\lambda_e \nu \,. \tag{2.21}$$

The electron mass is scaled by λ_e allowing the electron mass to differ greatly from the vacuum expectation value. This same procedure applies to the other lepton generations, giving rise to the muon and tau masses [59]. However, it is important to note that in GSW theory (and the current Standard Model) the neutrinos are massless which is experimentally shown to be false by the observation of neutrino oscillations [63, 64]. Thus, neutrinos must obtain mass from something other than the Higgs field.

Quarks obtain mass in a similar way as electrons. Consider one generation of quarks and the interaction of these quark fields with the Higgs field

$$\Delta\mathcal{L} = \lambda_d \bar{Q}_L \cdot \phi d_R - \lambda_u \epsilon^{ab} \bar{Q}_{La} \phi^\dagger u_R + h.c. \tag{2.22}$$

$$= -\frac{1}{\sqrt{2}}\lambda_d \nu \bar{d}_L d_R - \frac{1}{\sqrt{2}}\lambda_u \nu \bar{u}_L u_R + h.c. + \dots \,. \tag{2.23}$$

Thus, the up and down quark masses are

$$m_d = \frac{1}{\sqrt{2}}\lambda_d \nu \,, \qquad\qquad m_u = \frac{1}{\sqrt{2}}\lambda_u \nu \,. \tag{2.24}$$

The other generations obtain mass in the same way. The quark masses, like the lepton masses, depend on a dimensionless coupling λ_i allowing for the large difference in masses observed experimentally. λ_i are the Yukawa couplings; it is important to note that they are arbitrary and non-minimal [51].

2.3.3 Low Energy Electroweak Currents and Interactions

The three generations of quarks allow for terms that mix generations. This is more easily seen by considering a basis for the quark fields that diagnolizes their Higgs couplings (i.e. the mass matrix). A change of basis is a unitary transformation, in this case

$$u_L^i = U_u^{ij} u_L^{'j} \,, \qquad\qquad d_L^i = U_d^{ij} d_L^{'j} \,, \tag{2.25}$$

where $u_L^i = (u_L,\ c_L,\ t_L)$ and $d_L^i = (d_L,\ s_L,\ b_L)$ are the quark fields in the original basis and $u_L^{'i}$ and $d_L^{'i}$ are the quark fields in the mass eigenstate. The W current can be derived

from the covariant derivative for a fermion field (Eq. 2.11) and expressed in the new mass eigenstate

$$J_W^{\mu+} = \frac{1}{\sqrt{2}}\bar{u}_L^i \gamma^\mu d_L^i \tag{2.26}$$

$$= \frac{1}{\sqrt{2}}\bar{u}_L^{'i} \gamma^\mu (U_u^\dagger U_d)_{ij} d_L^{'j} \tag{2.27}$$

$$= \frac{1}{\sqrt{2}}\bar{u}_L^{'i} \gamma^\mu V_{ij} d_L^{'j}\,. \tag{2.28}$$

The new unitary matrix V is the Cabibbo-Kobayashi-Maskawa (CKM) matrix [51].

The CKM matrix describes the flavour mixing of quarks mediated by the W boson, it is expressed as

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}\,. \tag{2.29}$$

The off-diagonal elements specify interactions that occur across generations [65].

In the case of leptons, there is no right-handed neutrino, therefore there are no tree level generation changing interactions.

The Z current can also be derived using Eq. 2.11 and is given by

$$\begin{aligned} J_Z^\mu = \frac{1}{cos\theta_w}[&\bar{\nu}_L\gamma^\mu(\frac{1}{2})\nu_L + \bar{e}_L\gamma^\mu(-\frac{1}{2} + sin^2\theta_w)e_L + \bar{e}_R\gamma^\mu(sin^2\theta_w)e_R \\ &+ \bar{u}_L\gamma^\mu(\frac{1}{2} - \frac{2}{3}sin^2\theta_w)u_L + \bar{u}_R\gamma^\mu(-\frac{2}{3}sin^2\theta_w)u_R \\ &+ \bar{d}_L\gamma^\mu(-\frac{1}{2} + \frac{1}{3}sin^2\theta_w)d_L + \bar{d}_R\gamma^\mu(\frac{1}{3}sin^2\theta_w)d_R]\,. \end{aligned}$$

In this equation, the Z boson has no flavour changing interactions [51].

Finally, the photon interactions are described by the standard electromagnetic current which can again be derived using Eq. 2.11

$$J_{EM}^\mu = \bar{e}\gamma^\mu(-1)e + \bar{u}\gamma^\mu(+\frac{2}{3})u + \bar{d}\gamma^\mu(-\frac{1}{3})d\,. \tag{2.30}$$

The photon, also does not allow for any flavour changing interactions [51]

Therefore, in the Standard Model, flavour changing neutral currents do not occur at tree level and are suppressed.

The emergence of mass terms for the W and Z bosons creates a potential that varies as $U(r) \propto e^{-Mr}$. Thus, the weak force is a short range force that falls off at distances greater than $1/M$. It is the magnitude of these masses that creates a "weaker" force than the electromagnetic force [59].

Electrically charged weak currents only affect left-handed fermions[1]. Whereas electromagnetic currents affect both left- and right-handed fermions [53].

Electromagnetic currents describe the interactions of electrically charged particles and are mediated by the neutral photon. These interactions must conserve fermion flavour and obey charge (exchanging a particle with its antiparticle) and parity (flipping left-handed particles into right-handed particles and vice versa) symmetry. Examples of electromagnetic interactions include: electron-positron annhilation and electron-electron scattering [53].

Neutral weak interactions are mediated by the Z boson. All interactions mediated by the photon can also be mediated by the Z boson. The neutral neutrinos have weak isospin and weak hypercharge, so the Z boson can mediate neutrino scattering and neutrino-antineutrino annhilation/production. Additionally, neutral weak interactions can violate parity symmetry, but conserve fermion flavour. The Z boson is its own anti-particle, so it has no weak isospin and no weak hypercharge [53].

Electrically charged weak interactions are mediated by the W boson and are the only flavour changing interactions in the Standard Model. Due to the lack of right-handed neutrinos [2], electrically charged weak vertices with leptons cannot occur across generations [53]. However, this is not true for quarks. The CKM matrix gives the allowed quark-W vertices and these interactions can be visualized nicely in Fig. 2.2.

In the interactions described by the CKM matrix the W boson carries weak isospin, but no weak hypercharge. Additionally, there is an important phase in the matrix. Before GSW theory, it was predicted that the combined symmetry of charge would be conserved. This can be simply thought of in a decay process by replacing the particles with their

[1] Often, the weak interaction is denoted as $SU_L(2)$.

[2] To date, no right-handed neutrinos have been observed. However, they have not been completely ruled out experimentally and may interact very rarely—suppressing lepton flavour changing interactions.

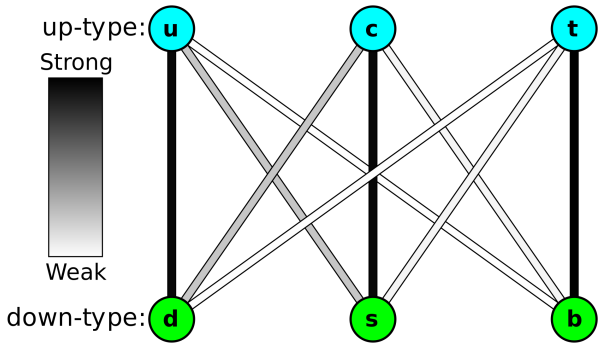

Figure 2.2: The allowed quark-W vertices. Figure from the public domain.

antiparticles and then swapping the chirality (handedness) of the particles. If charge-parity symmetry holds, then both processes should occur with the same frequency. The phase in the CKM matrix allows this to be violated. This was observed indirectly with the Fitch-Cronin experiment [66] in 1964 and then directly with the NA48 experiment at CERN [67] and KTeV experiment at Fermilab [68]. A larger symmetry, charge-parity-time, is still conserved. [53].

2.3.4 The Higgs Boson

The introduction of a scalar field with a non-zero vacuum expectation value has another consequence, it predicts the existence of a massive scalar boson, the Higgs boson. The mass of this new boson is given by

$$m_h = \sqrt{\frac{\lambda}{2}}\nu\,, \tag{2.31}$$

which relies on a dimensionless constant λ like the constants for the fermion masses [51].

Interactions with the Higgs conserve flavour and the interaction strength defines the mass of the interacting particles. The Higgs boson is the only mediator that has weak hypercharge [51].

The Higgs boson was discovered in 2012 at CERN by the ATLAS and CMS experiments with a mass of 125 GeV; it was the last fundamental Standard Model particle to be discovered [69, 70]. Further measurements confirmed that the observed boson was indeed a scalar particle that coupled as predicted by the Standard Model [71]. This further confirmed the predictions of the Standard Model.

2.4 Beyond the Standard Model

There are several notable issues with the Standard Model that motivate the need for a more complete theory. For example:

- In GSW theory the neutrinos are massless, which was proven false with the observation of neutrino oscillations at the SNO and Super-Kamiokande experiments [64, 63]. This means that neutrinos must obtain mass by some mechanism not explained in the Standard Model

- Cosmological observations point to the existence of dark matter [53, 72, 73, 74, 75]. Such matter is not explained by any particle in the Standard Model.

- One of the most common forces experienced in our daily lives, gravity, has no mediator in the Standard Model [51].

- Our universe is composed of matter rather than a mix of matter and anti-matter. While a small amount of charge-parity violation occurs in the Standard Model, it does not occur nearly at the scale needed to explain the matter/anti-matter asymmetry observed in our universe [51, 76].

These are some of the most commonly discussed short comings of the Standard Model. In addition to these observational issues, there are a number of theoretical problems; the most important for this thesis is the gauge hierarchy problem.

2.4.1 Gauge Hierarchy Problem

A hierarchy problem is created when divergent terms must be counteracted by introducing free parameters at a hierarchy of scales.

One such problem is caused by the the Higgs self-energy, E_{self}^{Higgs}, where every particle that couples to the Higgs contributes a radiative-correction term. The issue is most exacerbated in the top loop, as illustrated in Fig. 2.3, since its contribution to E_{self}^{Higgs} is

$$E_{self}^{Higgs} \propto \int^{\Lambda} d^4k \left(\frac{1}{k^2 - m_f^2} + \frac{2m_f^2}{(k^2 - m_f^2)^2} \right) . \tag{2.32}$$

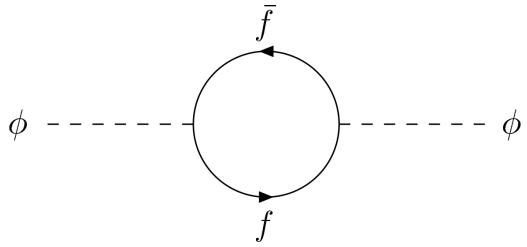

Figure 2.3: The fermion contribution to the Higgs self energy. Figure from [2].

Λ is the integral cutoff, typically the Standard Model cutoff[3] $\Lambda = 10^{16}$ GeV. The second term is quadratically divergent with respect to the arbitrary Λ value. The problem is readily apparent if we compare the scale of the measured mass of the Higgs 125 GeV to that of the bare mass needed to complement the arbitrary choice of $\Lambda \sim 10^{16}$ GeV,

$$m_h^2 = m_{h0}^2 + \delta m_h^2 \,,$$
$$m_{h0}^2 = (\Lambda \sim \mathcal{O}(10^{16}) \text{ GeV})^2 - (125\,\text{GeV})^2 \,.$$

We see that the tuning grows finer with the increasing scale of the integral cutoff (i.e. next lightest new physics scale). Were there to be no new physics, the cutoff scale (and thus the fine-tuning) is on the order of M_{Plank} [51, 2]

This is an aesthetic problem with the Standard Model in that it "feels" wrong. It "feels" wrong because there has been a problem like this before that lead to a more complete theory—namely, the electron self-energy problem. A similar situation can be created experimentally when studying ferromagnets. However, it should not "feel" like there is an experimenter playing with the parameters of the Standard Model!

The problem with the electron self-energy arose in the early twentieth century. When considering the diagram in Fig. 2.4 using Dirac's one-particle theory, a quadratic divergence is observed. This was an incomplete picture though, when Dirac's hole theory was introduced this divergence became a logarithmic divergence. This was addressed in quantum electrodynamics when Feynman used renormalization to resolve logarithmic divergences [77]. In this case, the quadratic divergence pointed to an incomplete theory and motivated QED. The same mechanism may be behind the Higgs self-coupling divergence.

[3]In renormalization, a constant Λ is introduced and the theory is only valid for values below this constant.

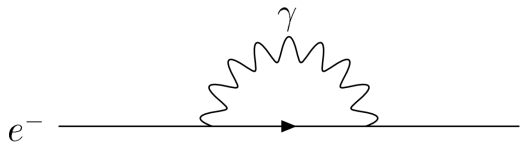

Figure 2.4: The electron self energy. Figure from [2].

A similar situation is observed in ferromagnets. At temperatures near 0 K, a ferromagnet has a spin expectation value on the order of the atomic parameters. An experimenter can finely adjust the temperature until the system reaches a point where the spin expectation value is much smaller then the value predicted by atomic parameters. In this situation, there is no problem because an experimenter was adjusting the parameters, but there is no experimenter "fine tuning"[4] the radiative corrections to the Higgs mass [51]. This is why the Higgs self-coupling divergence "feels" wrong and motivates the need for a more complete theory.

Many theories beyond the Standard Model address this problem, but in doing so predict new particles. One such theory is supersymmetry, which predicts that each Standard Model particle has a higher mass super partner. The minimal version of supersymmetry predicts particles that should have already been discovered at particle colliders prompting the rise of alternative theories. The next chapter discusses a type of particle that arises in some of these theories.

[4]Sometimes this problem is referred to as a fine tuning problem for this reason. It is also called a naturalness problem, because the principle of naturalness states that all parameters should be of the same order.

Chapter 3

Vector-like Quarks (VLQs)

The longer I live, the more uninformed I feel. Only the young have an explanation for everything.

Isabel Allende, *The City of the Beasts*

There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened.

Douglas Adams, *The Restaurant at the End of the Universe*

The gauge hierarchy problem (sec. 2.4.1) is caused by a quadratic divergence in the Higgs self-energy. One way to address this problem is to extend the Higgs sector. This can be achieved in several approaches, the two highlighted here are: compositeness of the Higgs boson and extending the Higgs sector with an extra dimension.

In composite Higgs models [78], the Higgs boson is a bound state and fermions have high mass partners. In early versions of this model, fermions were hypothesized to exist in a partially composite state where fermions become a linear combination of a light mass and heavy mass state predicting the existence of high mass resonances. Later versions focused on developing a mechanism around the top quark wherein the top has a heavier partner. In both cases, a fourth generation of non-chiral quarks arises [79].

In little Higgs models, the Higgs boson is a pseudo-Goldstone boson and a symmetry breaking at the TeV scale is added. This symmetry breaking gives rise to the light Higgs boson and eliminates the problematic radiative corrections. Like the composite Higgs models, these theories also predict the existence of a top partner [80].

In all of these theories, a new fourth generation of quarks must have left- and right-handed weak eigenstates with the same weak quantum numbers. These new quarks will couple to heavy quarks, have colour charge, and have fractional charge. It is these properties that have motivated the name vector-like quarks (VLQs).

VLQs are appealing in many ways. In addition to solving the gauge hierarchy problem, VLQs also provide new sources of CP violation. This will help to further explain the matter anti-matter asymmetry observed in our universe. VLQs are aesthetically nice providing a fourth generation in the Standard Model. A chiral fourth generation has already been ruled out experimentally, so VLQs are the simplest way of adding a fourth generation to the Standard Model [48].

Extensions of the Higgs sectors are not the only theories that predict the existence of vector-like fermions. Hypothetically speaking, vector-like fermions could be added to the Standard Model without addressing the gauge hierarchy problem. However, the VLQs in these theories arise from extending the Higgs sector yielding predictions that have unique experimental consequences.

3.1 Symmetry Breaking in the "Littlest Higgs"

VLQs do not obtain their mass from a Yukawa coupling to the Higgs field. Instead, the mass is the result of a broken symmetry. In all of the composite Higgs and little Higgs models, there is some higher symmetry that is broken. This section will focus on the method of symmetry breaking presented in the "littlest Higgs" model [80].

The gauge hierarchy problem is specifically concerned about the large difference in m_h^2/Λ^2 where Λ is the cutoff scale, which is 10^{16} GeV in the Standard Model. But, what if there is another broken symmetry which has a cutoff around $\Lambda = 10^4$ GeV? This is a much more manageable scale. In the littlest Higgs model, a higher $SU(5)$ symmetry with

this lower cutoff scale is proposed. This symmetry breaks $SU(5) \to SO(5)$ creating 14 Goldstone bosons. These bosons are "eaten" creating the electroweak and Higgs bosons that we observe [80].

This is achieved by considering the $SU(5)$ subgroup $G_1 \times G_2 = [SU(2) \times U(1)]^2$, which is broken to give rise to the electroweak gauge symmetry $SU(2) \times U(1)$. This model has the Lagrangian

$$\mathcal{L} = \mathcal{L}_K + \mathcal{L}_t + \mathcal{L}_\psi \,, \tag{3.1}$$

where $\mathcal{L}_K$ are the kinetic terms, $\mathcal{L}_t$ is the origin of the top Yukawa coupling, and $\mathcal{L}_\psi$ creates the other Yukawa couplings. The term $\mathcal{L}_t$ shows the origin of VLQs in this model

$$\mathcal{L}_t = \lambda_1(q_3 h + f\tilde{t})u_3^{'c} + \lambda_2 f\tilde{t}\tilde{t}^c + \ldots \tag{3.2}$$

This results in a heavy fermion $\tilde{t}$ that is multiplied by a linear combination of $u_3^{'c}$ and $\tilde{t}^c$. It also yields the top Yukawa coupling

$$\lambda_t q_3 h u_3^c \text{ where } \lambda_t = \frac{\lambda_1 \lambda_2}{\sqrt{\lambda_1^2 + \lambda_2^2}} \,. \tag{3.3}$$

The heavy fermion gives a negative contribution to the Higgs mass

$$-\frac{3\lambda_t^2}{8\pi^2} m'^2 \log\frac{\lambda^2}{m'^2} \,. \tag{3.4}$$

When this negative mass term dominates over the positive mass terms it triggers electroweak symmetry breaking [80].

In this theory, the fermions become $(5, 3)$ and $(5, \bar{3})$ multiplets under $SU(5) \times SU(3)_{\text{color}}$ The top and bottom quark become a mixture of a $(5, 3)$ multiplet and a new quark doublet field q, while the anti-top is a mixture of a $(5, \bar{3})$ multiplet and the $SU(5)$ singlet field t^c. The q and t^c fields induce the symmetry breaking and give rise to a Higgs boson with a physical mass consistent with the observed mass of 125 GeV. All of this, while avoiding quadratic divergences [80]. An example of the new VLQ loop contributions to the Higgs mass is shown in Fig. 3.1

In this theory, the VLQs arise as a doublet field $q = (T', B')$ and a singlet field t^c. This is merely one way in which VLQs may arise.

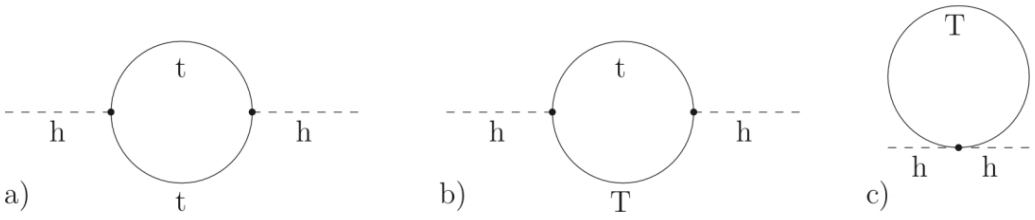

Figure 3.1: The top partner loop contributions to the Higgs mass in the littlest Higgs model. From [3].

3.2 Possible VLQ Multiplets

In composite Higgs and little Higgs models there are many possible multiplets of VLQs. If mixtures of VLQ multiplets are not considered, then there are seven possible multiplets. All seven have the property that the left- and right-handed states are the same and therefore both states interact weakly. All VLQs in these multiplets are spin-1/2 particles (hence, fermions) and the proposed VLQs X, T', B', Y have electric charge $Q = 5/3, 2/3, -1/3, -4/3$ respectively. The $SU(2)$ representation, weak hypercharge Y, and weak isospin T_3 for each multiplet is summarized in Table 3.1 [47, 48].

The multiplet that seems most aesthetically pleasing in the context of the Standard Model is the doublet (T', B'). The remaining sections in this chapter will focus on the consequences of this doublet, but very similar predictions arise in the other six multiplets. These predictions are summarized in [48].

3.3 VLQ Mixing

The Standard Model quarks are predicted to have components from the new VLQ doublet, but only the third generation is predicted to have a sizeable contribution from the VLQs. Therefore, VLQ-quark mixing is expected to be heavily dominated by the third generation. This mixing is described by two 2x2 unitary matrices $U^u_{L,R}$ and $U^d_{L,R}$

$$\begin{pmatrix} t_{L,R} \\ T_{L,R} \end{pmatrix} = U^u_{L,R} \begin{pmatrix} t^0_{L,R} \\ T^0_{L,R} \end{pmatrix} = \begin{pmatrix} cos\theta^u_{L,R} & -sin\theta^u_{L,R} e^{i\phi_u} \\ sin\theta^u_{L,R} e^{-i\phi_u} & cos\theta^u_{L,R} \end{pmatrix} \begin{pmatrix} t^0_{L,R} \\ T^0_{L,R} \end{pmatrix} , \tag{3.5}$$

$$\begin{pmatrix} b_{L,R} \\ B_{L,R} \end{pmatrix} = U^d_{L,R} \begin{pmatrix} b^0_{L,R} \\ B^0_{L,R} \end{pmatrix} = \begin{pmatrix} cos\theta^d_{L,R} & -sin\theta^d_{L,R} e^{i\phi_d} \\ sin\theta^d_{L,R} e^{-i\phi_d} & cos\theta^d_{L,R} \end{pmatrix} \begin{pmatrix} b^0_{L,R} \\ B^0_{L,R} \end{pmatrix} , \tag{3.6}$$

	SM quarks	Singlets	Doublets	Triplets
	$\begin{pmatrix} u \\ d \end{pmatrix}$ $\begin{pmatrix} c \\ s \end{pmatrix}$ $\begin{pmatrix} t \\ d \end{pmatrix}$	$\begin{pmatrix} T' \end{pmatrix}$ $\begin{pmatrix} B' \end{pmatrix}$	$\begin{pmatrix} X \\ T' \end{pmatrix}$ $\begin{pmatrix} T' \\ B' \end{pmatrix}$ $\begin{pmatrix} B' \\ Y \end{pmatrix}$	$\begin{pmatrix} X \\ T' \\ B' \end{pmatrix}$ $\begin{pmatrix} T' \\ B' \\ Y \end{pmatrix}$
$SU(2)$	$q_L = 2$ $q_R = 1$	1	2	3
Y	$q_L = 1/6$ $u_R = 2/3$ $d_R = -1/3$	$2/3 \quad -1/3$	$7/6 \quad 1/6 \quad -5/6$	$2/3 \quad -1/3$
T_3	$u_L = 1/2$ $d_L = -1/2$ $q_R = 0$	0	$X = 1/2 \quad T' = 1/2 \quad B' = 1/2$ $T' = -1/2 \quad B' = -1/2 \quad Y = -1/2$	$X = 1 \quad T' = 1$ $T' = 0 \quad B' = 0$ $B' = -1 \quad Y = -1$

Table 3.1: Seven possible multiplets of Vector-like Quarks. The row $SU(2)$ denotes how each interacts under the weak force, Y is weak hypercharge for each particle state, and T_3 is weak isospin for each particle state [47, 48].

where b, t, B, T are the mass eigenstates and b^0, t^0, B^0, T^0 are the weak eigenstates [48].

This mixing alters the Standard Model lagrangian by adding W, Z, and H terms, a portion of these new terms describes quark-VLQ interactions:

$$\begin{aligned}
\mathcal{L}_W =& -\frac{g}{\sqrt{2}}\bar{Q}\gamma^\mu(V^L_{Qq}P_L + V^R_{Qq}P_R)qW^+_\mu + h.c. \\
& -\frac{g}{\sqrt{2}}\bar{q}\gamma^\mu(V^L_{qQ}P_L + V^R_{qQ}P_R)QW^+_\mu + h.c.\,, \\
\mathcal{L}_Z =& -\frac{g}{2cos\theta_W}\bar{q}\gamma^\mu(\pm X^L_{qQ}P_L \pm X^R_{qQ}P_R)QZ_\mu + hc\,, \\
\mathcal{L}_H =& -\frac{gm_Q}{2M_W}\bar{q}(Y^L_{qQ}P_L + Y^R_{qQ}P_R)QH + h.c.\,,
\end{aligned}$$

where $q = t, b$ and $Q = T', B'$. In the case of the (T', B') doublet the couplings become:

VLQ-quark couplings to the W boson

$$V_{Tb}^{L} = sin\theta_L^u cos\theta_L^d e^{-i\phi_u} - cos\theta_L^u sin\theta_L^d e^{-i\phi_d}\,, \qquad V_{Tb}^{R} = -cos\theta_R^u sin\theta_R^d e^{-i\phi_d}\,,$$

$$V_{tB}^{L} = cos\theta_L^u sin\theta_L^d e^{i\phi_d} - sin\theta_L^u cos\theta_L^d e^{i\phi_u}\,, \qquad V_{tB}^{R} = -sin\theta_R^u cos\theta_R^d e^{i\phi_u}\,.$$

VLQ-quark couplings to the Z boson

$$X_{tT}^{R} = -sin\theta_R^u cos\theta_R^u e^{i\phi_u}\,, \qquad X_{bB}^{R} = -sin\theta_R^d cos\theta_R^d e^{i\phi_d}\,,$$

$$X_{tT}^{L} = 0\,, \qquad X_{bB}^{L} = 0\,.$$

VLQ-quark couplings to the Higgs boson

$$Y_{tT}^{L} = sin\theta_R^u cos\theta_R^u e^{i\phi_u}\,, \qquad Y_{tT}^{R} = \frac{m_t}{m_T} sin\theta_R^u cos\theta_R^u e^{i\phi_u}\,,$$

$$Y_{bB}^{L} = sin\theta_R^d cos\theta_R^d e^{i\phi_d}\,, \qquad Y_{bB}^{R} = \frac{m_b}{m_B} sin\theta_R^d cos\theta_R^d e^{i\phi_d}\,.$$

These terms are quite different from how generations interact. In the Standard Model, only the W boson can cause a generational change at tree level. However, the T', B' interact with t, b at tree level with the W, Z, and Higgs bosons [48].

3.4 Experimental Signatures

The doublet T', B' primarily interacts at tree level via

$$T' \to W^+ b\,, \qquad T' \to Zt\,, \qquad T' \to Ht\,,$$

$$B' \to W^- t\,, \qquad B' \to Zb\,, \qquad B' \to Hb\,.$$

Experimentally, the VLQs can be produced singly or in pairs. This project is restricted to pair production of VLQs which is described in Fig. 3.2. This production mode is expected to be one of the most common modes at pp colliders[1]. The apparatus used for searching

[1]The proton is composed of quarks and gluons. Frequently, pp collisions result in gluon-gluon interactions making gluon-gluon fusion a prominent production mode.

for these VLQs and the subsequent search is discussed in the next chapters.

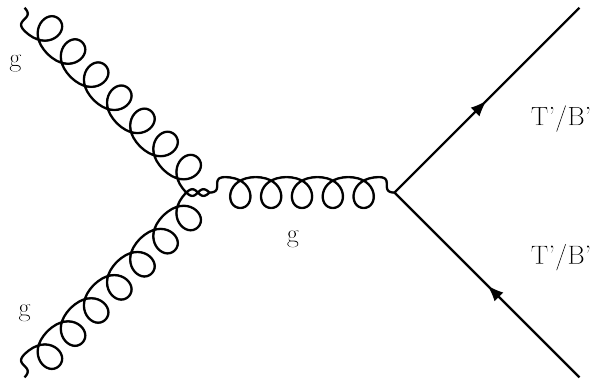

Figure 3.2: Pair production of VLQs.

Chapter 4

The Large Hadron Collider and Compact Muon Solenoid

All have their worth... and each contributes to the worth of the others.

J.R.R. Tolkien, *The Silmarillion*

The LHC is a 27 km long proton-proton collider located between 45 m and 100 m underground across the border of France and Switzerland. The apparatus is designed to for proton collisions every 25 ns with a 14 TeV center-of-mass energy and is capable of Pb 82^+ heavy ion collisions with 574 TeV per nucleus. Around the collider are four interaction points where collisions are observed by large experiments; the four main ones are the ATLAS experiment, the Compact Muon Solenoid (CMS), A Large Ion Collider Experiment (ALICE), and the LHC beauty (LHCb) experiment.

The primary mission of the LHC was to search for the Higgs boson and new processes beyond the Standard Model. The Higgs boson was discovered in 2012 by the ATLAS and CMS experiments [69, 70], but the search for new physics continues. This is done at four interaction points around the LHC. This chapter introduces the LHC and then describes the CMS experiment in detail. The CMS experiment is the apparatus is the subject of the rest of this thesis.

4.1 The Large Hadron Collider

In the 1980s, the world was experiencing the height of the Cold War between the western powers and the eastern soviet powers. The US and Soviet Union battled for superiority in nearly every field, even an ice hockey game, the 1980 Olympic semi-final (USA vs. USSR) in Lake Placid, was a battleground for demonstrating superiority. Scientific advancement was no exception; the space race progressed from satellites in orbit to the first man on the moon, all to prove scientific superiority. Europe seemed to be lagging behind. However, in particle physics a very different story was taking place, this time between Europe and the US in a field where superiority was often demonstrated by investing in the world's biggest colliders.

The discovery of the W and Z bosons in 1983 at large masses confirmed Electroweak unification, but required the existence of some larger theory to cause these large masses. The leading explanation was the symmetry breaking introduced in the Higgs mechanism (Section 2.3.2). Thus, the hunt for the Higgs boson was on. This story begins with a three way race for scientific superiority; the year was 1983 and the US, Europe, and the Soviet Union were battling for superiority by attempting to discover the Higgs boson. The US proposal was to build an 87 km long ring in Texas called the Superconducting Super Collider and the European proposal was a 27 km ring on at CERN the border of Switzerland and France called the Large Hadron Collider (LHC). In an effort to keep in the race, the Soviet Union proposed a 21 km long ring near Moscow called the UNK proton accelerator.

Then, the narrative took a completely unexpected turn; Mikhail Gorbachev began to open the Soviet Union to the world[1], leading to its downfall. As the bloc collapsed, it entered into economic difficulties and cancelled the UNK proton accelerator. This made the US Congress significantly less motivated to continue funding the evermore expensive Superconducting Collider, leading to its cancellation in 1993—leaving only the LHC at CERN.

CERN was born out of international collaboration after World War II and this tradi-

[1]The names of Gorbachev's famous reforms were Perestroika (restructuring) and Glasnost (openness).

tion continued with the LHC, where CERN grew by uniting American and former Soviet scientists. The astronomical price tag of the LHC was only made possible by this collaboration and this collider became a beacon of peace through international cooperation.

4.1.1 A Proton's Journey

The LHC complex consists of a chain of smaller accelerators that bring particles up to subsequently higher energies. At each stage, the particles reach precise energy levels before being injected into the next accelerator.

In the case of protons, acceleration begins at the Linear Accelerator 4 (LINAC2 before 2020) with a simple bottle of hydrogen. Here, hydrogen atoms with an extra electron (H^-) are pulsed into 100 µs bunches using radio frequency quadrupoles and then accelerated to 160 MeV. Once this energy is obtained, the negative hydrogen ions are stripped of the electrons and the protons are injected into the Proton Synchrotron Booster. At this stage, the protons are bent into a circle by dipole magnets and accelerated to 2 GeV before being injected into the Proton Synchrotron. This intermediate stage allows for more protons to be injected into the Proton Synchrotron than if the LINAC injected protons directly [81, 74].

The Proton Synchrotron was the flagship accelerator at CERN in 1959, and now serves to increase the protons to 26 GeV before injecting into the Super Proton Synchrotron. The Super Proton Synchrotron was also a former CERN collider that began operation in 1972, now it serves as the final stage in the LHC injection chain. Here, protons are accelerated to 450 GeV and are injected at two points into the LHC splitting the proton beam into two beams that travel opposite directions around the LHC [81, 74]. The full CERN accelerator complex, including the LHC injection chain, is shown in Fig. 4.1.

The injector accelerators utilize conventional (room-temperature) electromagnets, while the LHC utilizes superconducting magnets constructed from niobium-tin. These 8.3 T magnets are operated at 1.9 K using liquid helium cooling. The 27 km ring contains dipole magnets for bending the protons in a circular path, quadrupole and sextupole magnets for squeezing the protons into tight bunches, and radio-frequency cavities to accelerate the protons while maintaining bunches with 40 MHz frequency. The LHC is capable of accel-

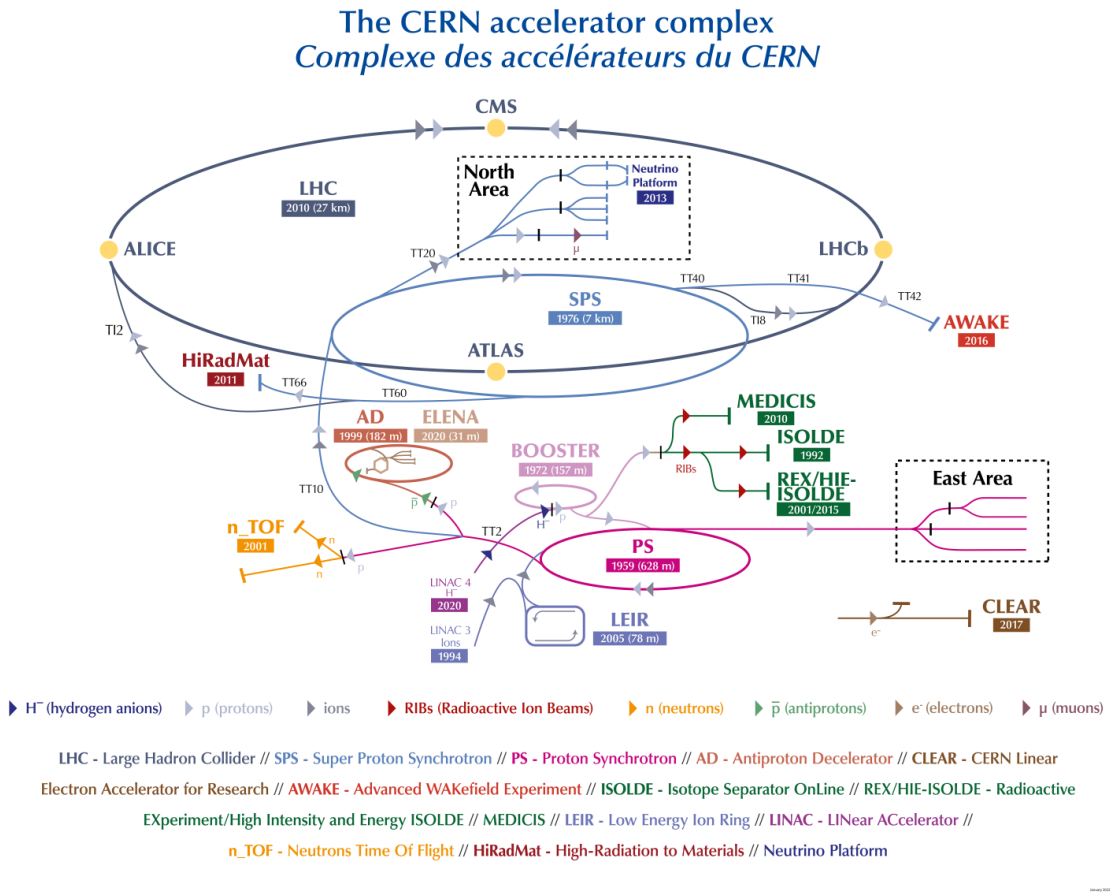

Figure 4.1: The CERN accelerator complex, including the LHC injection chain.

erating 2808 bunches of 10^{11} protons to create 14 TeV collisions with a design luminosity of $10^{34}\,\mathrm{cm}^{-1}\,\mathrm{s}^{2}$ [81, 74].

4.1.2 Proton Collisions

Physical processes arising from proton-proton interactions are determined from probabilistic quantum field theory calculations. Quantum field theory is used to calculate the cross section σ or the probability of a process happening. When a proton beam is created at the LHC, the protons are squeezed into tight bunches giving a probability of interaction called instantaneous luminosity $\mathcal{L}$. The integrated luminosity $\int \mathcal{L}dt$ can be roughly though of as the amount of data. Thus, the predicted number of events for a specific process are given by

$$N_{events} = \sigma \cdot \int \mathcal{L}dt\,. \tag{4.1}$$

These collisions can be split into two different categories: soft and hard scattering. In a soft scatter, the protons interact at large distances and only a small amount of radiation is transferred. This happens frequently when two proton bunches cross paths at interaction point. In a hard scatter, the protons interact at short distance exchanging a high amount of momentum. The center of mass energy $\sqrt{s}$ represents the highest possible momentum exchange [81, 74].

4.1.3 The LHC Timeline

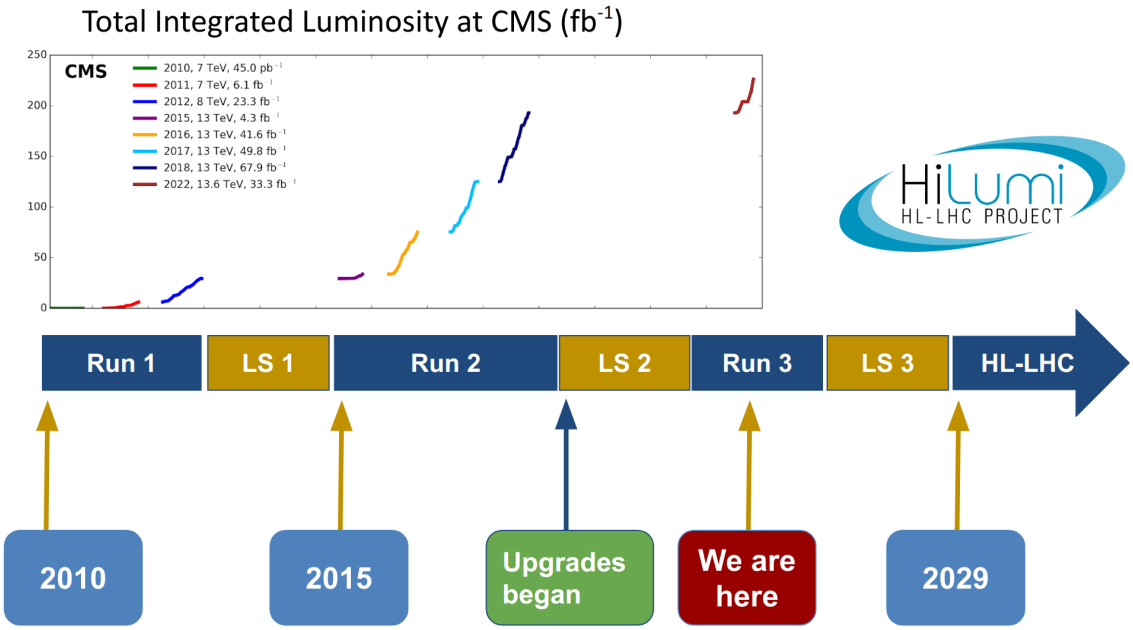

Figure 4.2: A brief timeline of LHC with the recorded integrated luminosity by CMS.

The LHC began taking collision data in 2010 with $\sqrt{s} = 7\,\text{TeV}$ and increased to $\sqrt{s} = 8\,\text{TeV}$ in 2012. This first period of data taking is called "LHC Run 1" and when it concluded, the LHC was closed for two years of upgrades called "Long Shutdown 1". This improved the luminosity and allowed the LHC to begin operations at $\sqrt{s} = 13\,\text{TeV}$. The next period of data taking, Run 2, took place from 2015 to 2018. The data collected by CMS for Run 2 is used in the search for vector-like quarks. After Run 2, the LHC began the first step of upgrades for a major luminosity increase, the High Luminosity or HL-LHC. This will begin operation in Run 4, but the extensive changes needed already began in Long Shutdown 2. This included the CMS muon system upgrade with the first

GEM muon detectors. Now, Long Shutdown 2 has finished and the LHC finished the first year of data taking for Run 3 with $\sqrt{s} = 13.6\,\text{GeV}$ collisions. A brief timeline with the integrated luminosity recorded by CMS is shown in Fig. 4.2.

4.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is one of two general purpose experiments at the LHC. The official letter of intent was submitted in 1992 and, together with ATLAS, CMS will allow for the LHC program to provide two independent measurements. This would allow for fast verification of new discoveries such as the (then hypothetical) Higgs boson or beyond Standard Model processes.

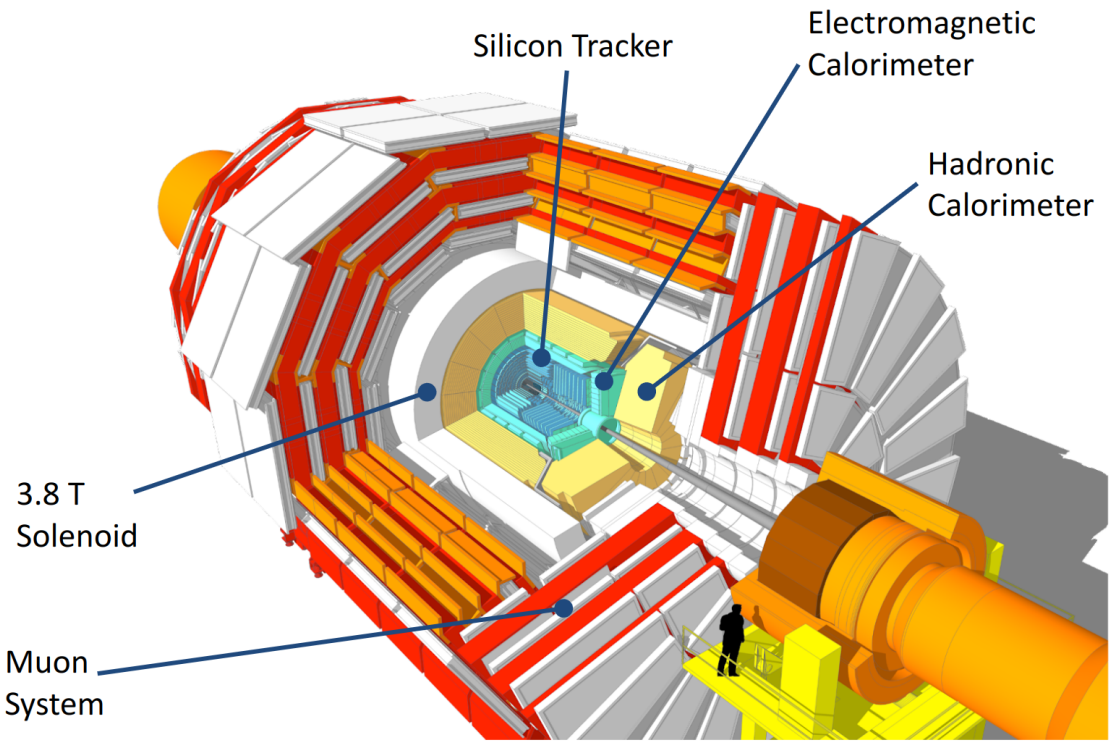

Figure 4.3: The CMS Experiment at the LHC.

The CMS experiment acts like a cylindrical onion with layers of different particle detectors, each serving a unique purpose. After collisions, positions of charged particles are precisely measured using the silicon tracking system. Then the calorimeters are used to measure the energy of all known particles except muons and neutrinos. These subsystems

are contained within a 3.8 T superconducting solenoid. This causes charged particles to curve in the detector. Outside of the solenoid is the muon system consisting of gaseous detectors. A full overview of the CMS experiment is shown in Fig. 4.3. CMS is 21.6 m in length, is 14.6 m in diameter, and weighs 12.500 t. Once all of the data is acquired and processed, this cylindrical onion acts like a giant camera taking snapshots of particle collisions [4, 82].

4.2.1 Coordinates

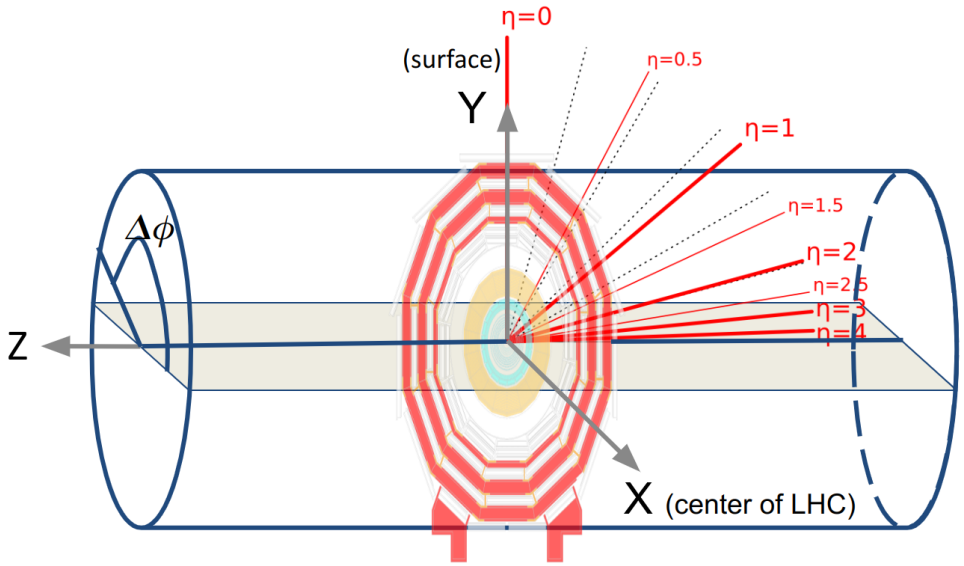

Figure 4.4: The coordinate system in the CMS experiment.

Collisions observed in the CMS experiment are described with a common coordinate system. In Cartesian coordinates, the z-axis points along the beamline, the x-axis points to the center of the LHC ring, and the y-axis points up to the surface. In spherical coordinates, r is the radial distance from the interaction point, ϕ is the azimuthal angle in the $x - y$ plane, and θ is the polar angle from the z-axis. The coordinates used to describe collisions utilize spherical coordinates with an adjusted θ defined as

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

This is called pseudorapidity and differences in η are Lorentz invariant along the z-axis (beamline). A diagram of the coordinate system in CMS is shown in Fig. 4.4. In general,

the detector subsystems at CMS are split into regions called the barrel, or central, region for $|\eta| < 1.5$ and endcap, or forward, region for $|\eta| > 1.5$ [83, 74, 81, 82].

4.2.2 Silicon Tracking System

Silicon semiconductors make excellent tracking detectors. When a charged particle passes through, it creates an electron-hole pair. A potential difference is applied to the top and bottom of a silicon sensor to cause the electrons and holes to drift apart, inducing a signal that can be readout. This process is very similar to the physics inside of a gaseous detector described in Chapter 5, but takes place on a much smaller scale. Silicon can be etched into small sensors that are connected with fine readout wires, resulting in a detector with significantly higher position resolution when compared to other technologies. Therefore, silicon sensors were chosen for the innermost tracking detectors of the CMS experiment [82].

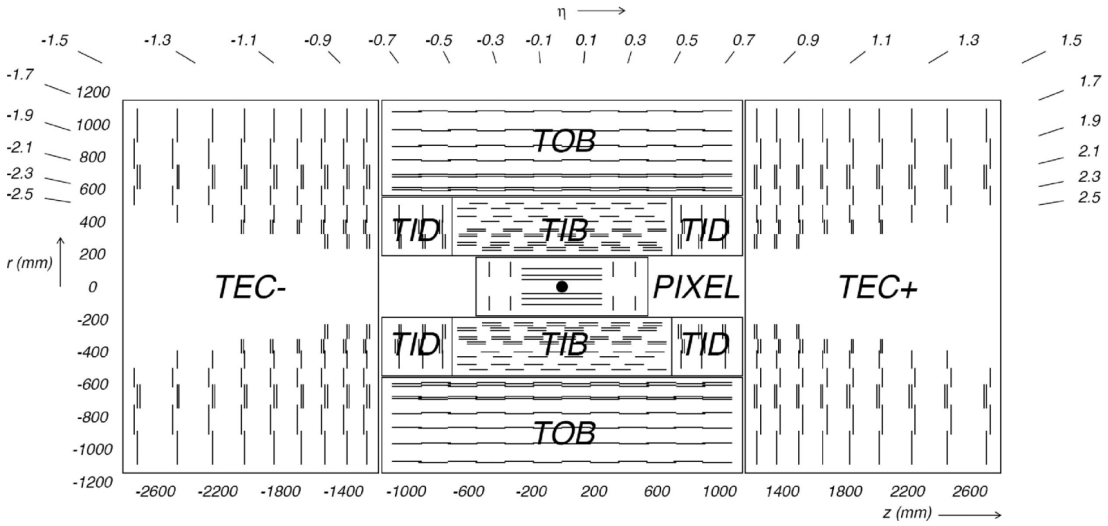

Figure 4.5: A diagram of the CMS inner tracker (in the (y, z) plane) made from silicon pixels and microstrips. Figure from [4].

The CMS tracking system consists of silicon pixels and microstrips. The highly precise pixels are placed nearest to the interaction point and then are surrounded by microstrips. The system has a total length of 5.8 m and a diameter of 2.6 m; A detailed overview of the CMS tracking system is shown in Fig. 4.5. Charged particles curve as they travel through the tracker due to the magnetic solenoid allowing a precise reconstruction of

the momentum. The CMS tracker has a momentum resolution of 1% for particles with $p_T < 20\,\text{GeV}$ and better for increasing p_T [74, 83, 82].

Silicon Pixels

The CMS pixel detector was replaced with an upgraded detector during the year end technical stop at the end of 2017. Thus, the first year of LHC run-2 utilized an older detector consisting of three layers of pixel rings in the barrel (at $r = 44$, 73, 103 mm) and two layers of pixel disks in the endcaps (at $z = \pm 345$ and ± 465 mm). The upgrade increased the number of pixel layers and brought the innermost layer closer to the beamline to better identify primary and secondary vertices. The barrel region contains four rings ($r = 29$, 68, 109, 160 mm) and the endcaps consist of three disks with inner and outer components (roughly $z = 30$, 40, 50 cm). A comparison of the upgraded (phase-1) and original pixel detector is shown in Fig. 4.6. The pixels have a spatial resolution of 10 µm in the (r, ϕ) plane and 20 µm along the z-axis [74, 83, 82, 5].

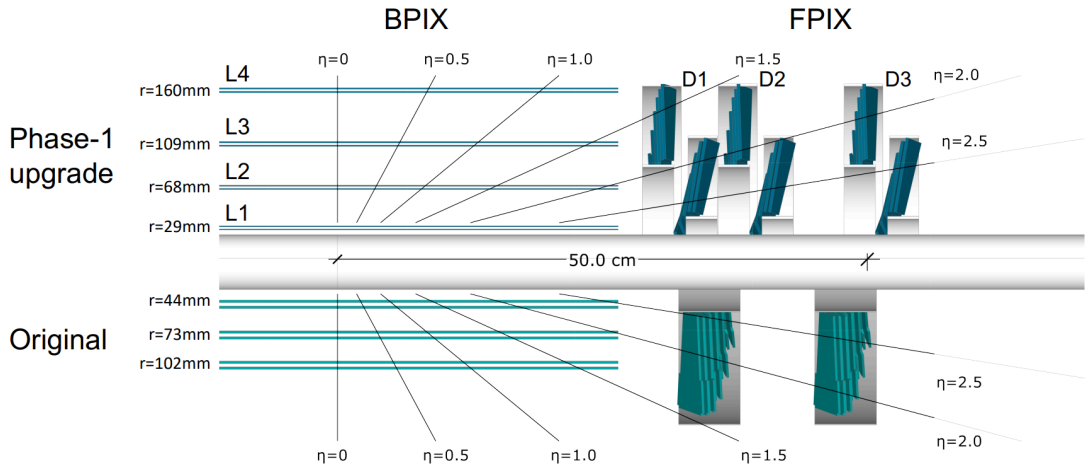

Figure 4.6: A comparison of the original and phase 1 CMS pixel detectors in the (y, z) plane. FPIX refers to the forward pixel detector and BPIX refers to the barrel pixel detector. Figure from [5].

Silicon Microstrips

The silicon microstrips operate with the same physics principles as the pixels but have a

much larger area and lower spatial resolution. The CMS microstrips consist of four smaller subsystems: the Tracker Inner Barrel, the Tracker Outer Barrel, the Tracker Outer Disks, and the Tracker End Cap. Each subsystem has a spatial resolution between 35 µm and 52 µm [74, 82].

4.2.3 Calorimeters

Calorimeters are detectors designed to create showers of subsequently lower energy particles in order to measure the energy of incident particles. Showers can be produced via the electromagnetic and strong interactions. Electromagnetic calorimeters consist of materials that create showers from photons and electrons. While hadronic calorimeters contain dense nuclei that induce showers via the strong force. These measurements can take place with homogeneous or sampling calorimeters. In a sampling calorimeter, the material inducing the shower is distinct from the material where the energy is deposited. In this case, some energy is undetected and must be estimated. While in a homogeneous calorimeter, the material inducing the shower is sensitive and the energy can be accurately measured. CMS contains a homogeneous electromagnetic calorimeter (ECAL) and sampling hadronic calorimeter (HCAL).

Electromagnetic Calorimeter

The CMS ECAL is constructed from lead tungstate ($PbWO_4$) crystals. $PbWO_4$ has a short radiation length ($X_0 = 0.89\,\text{cm}$), short Moliere radius (2.2 cm), and high radiation hardness (10 Mrad). The crystals also for the construction of a hermetic, homogeneous calorimeter—ideal for the size constraints caused by the magnetic solenoid. These crystals are capable of emitting 80% of the light in a shower within 25 ns, making for fast readout. The light is readout with silicon avalanche diodes (in the barrel) and vacuum phototriodes (in the endcaps). These technologies were carefully chosen for operation in the high magnetic field [82, 4].

The barrel section ($|\eta| < 1.479$) of the ECAL is equipped with 61,200 crystals for fine granularity. Each crystal has a cross sectional face of $22 \times 22\,\text{mm}^2$ and is 230 mm in length (25.8 X_0). The endcaps ($1.479 < |\eta| < 3.0$) are each equipped with 7,324 crystals that

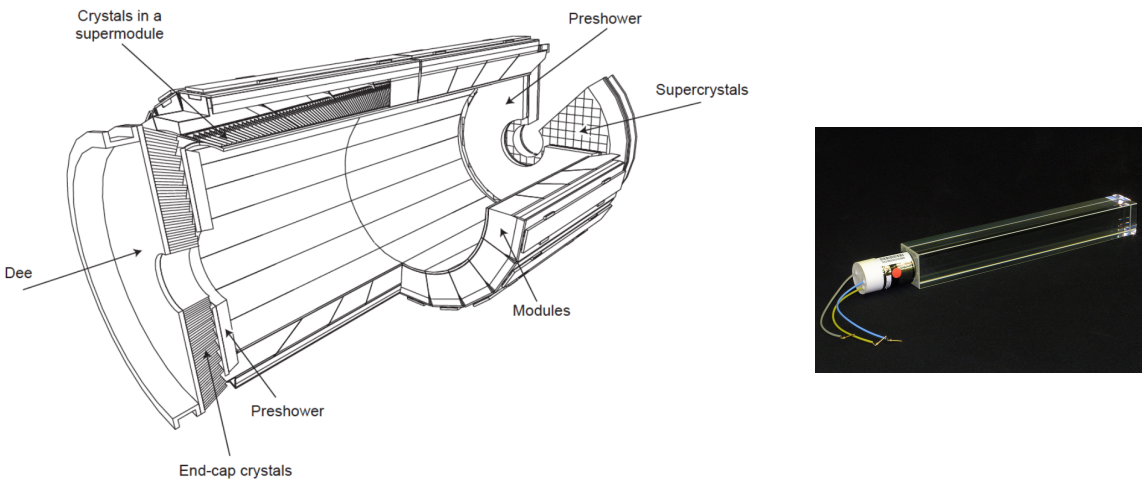

Figure 4.7: (Left) A diagram of the CMS ECAL. (Right) An individual ECAL crystal used in the endcap—each individual crystal took around two days to grow. Figures from [4].

have a cross sectional face of $28.6 \times 28.6\,\mathrm{mm}^2$ and are 220 mm in length (24.7 X_0). To improve the granularity in the forward regions, each endcap is equipped with a preshower at $1.653 < |\eta| < 2.6$. This is constructed from two silicon sensor planes and allows for the discrimination of a π^0 that creates two photon showers from an individual γ which creates only one shower. An overall diagram and an individual ECAL crystal are shown in Fig. 4.7 [25, 74, 82].

The energy resolution for the ECAL is given by

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E[\mathrm{GeV}]}} \oplus \frac{12\%}{E[\mathrm{GeV}]} \oplus 0.3\%\,, \tag{4.3}$$

in this case the first term describes stochastic effects during the showering process, the second term describes electronic noise, and the final term describes calibration errors and non-uniformity in the energy measurement. This amounts to an energy resolution between 0.5% and 1.5% for energies in the range 10 to 250 GeV [74, 83, 25, 82, 4].

Hadronic Calorimeter

The CMS HCAL is specially designed to measure the energy of hadronic decays and, after reconstruction, can be used to estimate the amount of missing transverse momentum. The HCAL is a sampling calorimeter consisting of four subsystems: hadron barrel (HB),

hadron outer (HO), hadron endcap (HE), and hadron forward (HF). An overview of the inner subsystems is shown in Fig. 4.8. Stringent requirements were placed on the HCAL due to the magnetic solenoid—the calorimeter had to be built from non-magnetic materials. For this reason, high purity brass (70% Cu, 30% Zn) was chosen for HB, HO, and HE. HE was constructed from over 1 million World War II artillery shell casings from the Russian navy. These former weapons were built with high purity brass and could be repurposed (when combined with copper from the US) to create the 600 tons of brass needed for HE; this became a scientific symbol of peace within the CMS experiment. Some of the old shell casings are shown in Fig. 4.8 [25, 82, 4].

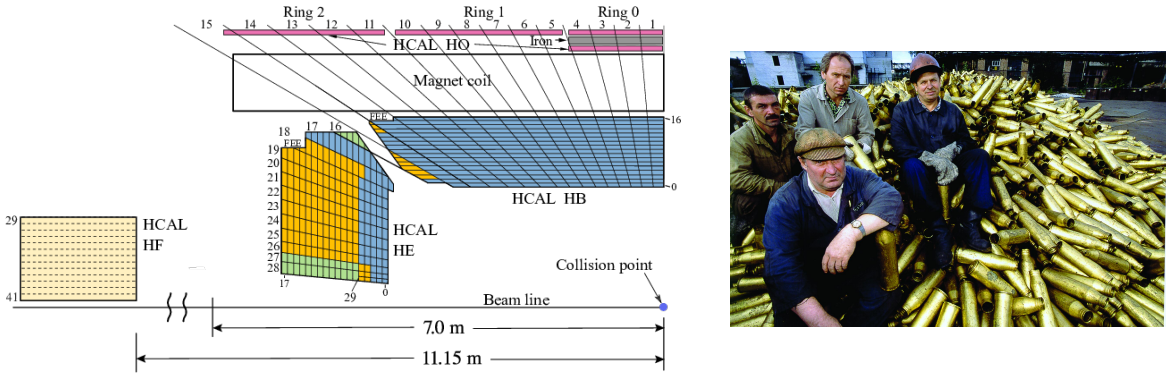

Figure 4.8: (Left) A diagram of HCAL in the (r, z) plane. Figure from [6] (Right) Russian navy shells that were melted down and combined with US copper to make HE—a symbol of peace within CMS. Figure from the CMS collaboration.

HB, HO, and HE all consist of alternating layers of brass and active plastic scintillator. The plastic scintillator is readout with hybrid photomultiplier tubes (replaced by SiPMs for LHC Run 3). HB extends within the magnetic solenoid to provide coverage for $|\eta| < 1.3$. HE provides coverage in the endcaps for $1.3 < |\eta| < 3.0$. In order to ensure that the calorimeter system captures 11.8 interaction lengths, HCAL was supplemented with HO outside of the magnetic solenoid covering $|\eta| < 1.26$. HF was placed in the very forward region of CMS providing coverage up to $|\eta| < 5.2$. This high energy, high radiation environment required that HF be constructed from more resistant materials, so HF is built from alternating steel absorbers and quartz scintillators [25, 4].

The overall energy resolution of HCAL is

$$\frac{\sigma_E}{E} = \frac{115.3\%}{\sqrt{E[\text{GeV}]}} \oplus 5.5\%\,, \tag{4.4}$$

where the first term is due to the stochastic nature of a shower and the second term is due to inefficiency during calibration [83, 4].

4.2.4 Magnetic Solenoid

At the time of construction, the CMS superconducting magnetic solenoid was the most powerful solenoid in world—capable of storing 2.6 GJ of energy. The 220 t magnet is comprised of a 4-layer winding of stabilized, reinforced NbTi for a total length of 12.5 m and a 6 m diameter. The superconducting volume is cooled with liquid helium to an operating temperature of 4.5 K. When operational, the magnet contains 19.14 kA flowing through 2168 turns to create an operational magnetic field of 3.8 T [25, 4]. The resulting magnetic field is shown in Fig. 4.9.

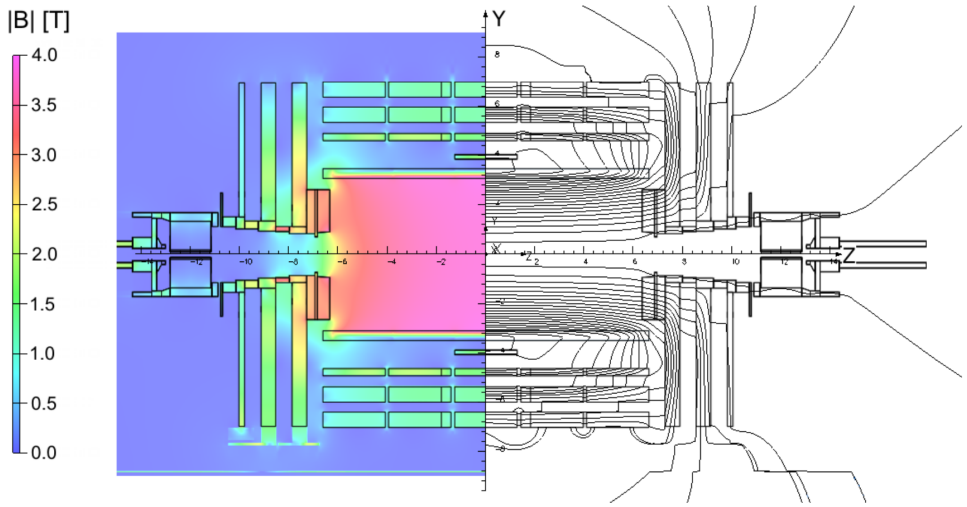

Figure 4.9: The magnetic field at the CMS experiment in the (r, z) plane. Figure from [7].

As charged particles travel through the magnetic field, they curve inside of the CMS tracker. This curvature allows for charge identification and transverse momentum reconstruction. CMS has less than 5% charge misidentification for muons with $p_T < 200\,\text{GeV}$ and less than 10% for muons with $p_T < 1\,\text{TeV}$. The transverse momentum (i.e. the

component perpendicular to the beam) is obtained after reconstructing tracks and follows the equation

$$p_T = 0.3zB\rho\,, \tag{4.5}$$

where z refers to particle charge as ze, B is the magnetic field, and ρ is the radius of curvature of the charged particle track. The transverse momentum resolution is given by

$$\frac{\sigma(p_T)}{p_T} = \frac{\sigma_s p_T}{0.3BL^2}\sqrt{\frac{720}{N+4}}\,, \tag{4.6}$$

where σ_s is the spatial resolution, L is the length of the solenoid, and N is the number of tracker hits used in reconstruction [12]. This amounts to a momentum resolution of 1% for charged particles with $p_T = 100\,\text{GeV}$ [74, 4].

The flux from the solenoid is returned through a 10.000 t iron yoke comprising five wheels and two endcaps. The muon system is interspersed through the return yoke providing a stable magnetic field for the operation of the muon detectors [4].

4.2.5 Muon System

The outermost detectors at CMS comprise the robust and redundant muon system. This system is capable of detecting ionizing particles that traverse the entire HCAL and ECAL such as muons with $p_T > 3\,\text{GeV}$ or hypothetical long lived particles that decay outside of the calorimeters. The muon system is comprised of three types of gaseous detectors and a fourth was added during Long Shutdown 2. The working principles of these gaseous detectors are described in detail in Chapter 5. Drift tubes (DTs) are used in the barrel region $|\eta| < 1.2$ which contains low occupancy and low, stable magnetic fields. Cathode strip chambers (CSCs) are used in the endcap regions ($0.8< |\eta| < 2.4$); this technology operates well with the high particle rates and non-uniform magnetic field in this region. Interspersed throughout the muon system are Resistive plate chambers (RPCs) to supplement DTs and CSCs providing fast response and high time resolution. RPCs were planed to provide coverage to $|\eta| < 2.4$, however, technological limitations in high radiation environments only allowed RPCs for $|\eta| < 2.1$. Gas electron multipliers (GEM) were added after LHC Run 2 to the high eta regions to improve the measurement of the muon

bending angle, which is critical to reducing the level-1 trigger rates. The GEM technology and CMS GEM upgrades are detailed in Chapters 5 and 6. Overall, the muon system provides muon identification and improves upon the tracker p_T resolution for muons (for reconstruction details, see Chapter 7) [25, 4].

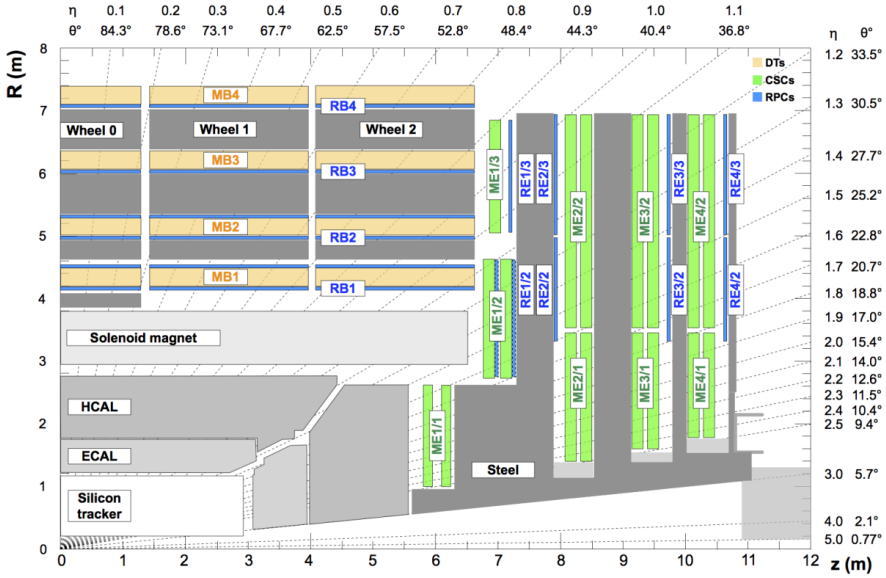

Figure 4.10: An (r, z) quadrant of CMS during LHC Run 2 (before upgrades). MB are drift tubes, ME are cathode strip chambers, and RE/RB are the resistive plate chambers. Figure from [8].

Drift Tubes

The DTs comprising the barrel region are divided into 5 wheels containing 12 azimuthal sectors. Each wheel contains 4 rings of DT chambers, where each chamber is 2.5 m in length and the widths vary from 1.9 m to 4.1 m depending on the position. The chambers consist of rectangular prism drift cells that are 42 mm by 13 mm and travel the length of the chamber. At the center of each cell is 50 µm gold-plated steel wire. The gas volume contains a mixture of Ar/CO_2 (85%/15%). Four staggered layers of drift cells comprise a superlayer. A diagram of a drift cell and DT chamber are shown in Fig. 4.11. Two superlayers are oriented with wires along the z axis to measure the (r, ϕ) coordinate and one superlayer is oriented orthogonally to measure the (r, z) coordinate. The DT chambers have a detection efficiency over 98% and spatial resolutions ranging from 200 µm (central

wheel) to 600 µm (external wheels) to provide an offline reconstruction resolution of 100 µm [25, 4].

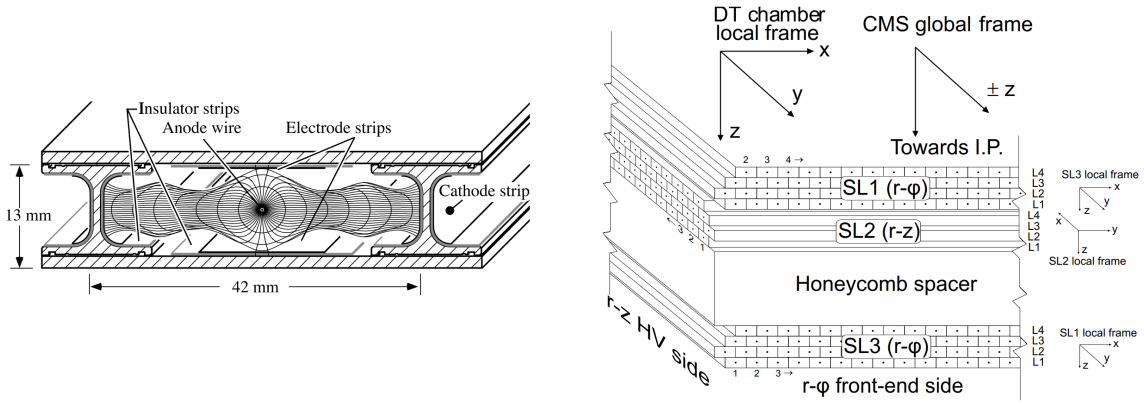

Figure 4.11: (Left) An individual DT cell. (Right) A diagram of a DT chamber in CMS. Figures from [9].

Cathode Strip Chambers

The CSCs are multi-wire proportional counters which provide short drift lengths for fast signal collection—ideal for the high rate region in the endcaps. The CSCs are divided into four disks per endcap with the first (inner most) disk being subdivided into three rings and the outer disks being sub divided into two rings. All rings, except the outer most on the first disk, are staggered to provide 10° overlap. All chambers, except for the ring nearest to the interaction point (called ME1/1), contain wires that are perpendicular to radial strips. The wires in ME1/1 are inclined by 29° to compensate for the Lorentz angle of electrons drifting in the high magnetic field. Radial strips are placed at ground and measure the polar angle. Wires are placed at 3.6 kV and measure the azimuthal angle. The gas volume contains a mixture of $Ar/CO_2/CF_4$ (40%/50%/10%). The position resolution in the range 70 to 150 µm and the θ position is determined with a precision of 0.5 cm. Fig. 4.12 shows an example of a CSC chamber and diagram of a CSC signal [25, 4].

Resistive Plate Chambers

The RPCs provide redundancy and high time resolution in the Muon system via six layers in the barrel and layers in all four endcap disks. CMS uses double-gap RPCs constructed

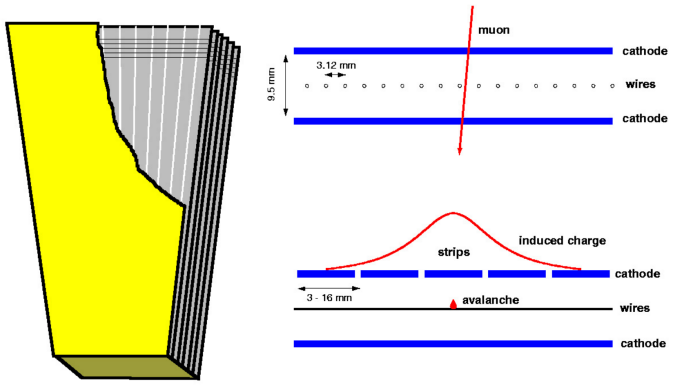

Figure 4.12: A diagram of a signal in a CMS CSC chamber. Figure from [10].

from four bakelite ([84]) planes forming two 2 mm gaps. The bakelite planes are coated with graphite to form electrodes and insulated aluminium strips to collect and read-out induced signals. The gas volume contains uses a freon-isobutane-sulfur hexafluoride (95.2%/4.5%/0.3%) mixture; Fig. 4.13 shows a diagram of the CMS RPCs. The resulting detector has a time resolution of 1.5 ns [25, 4].

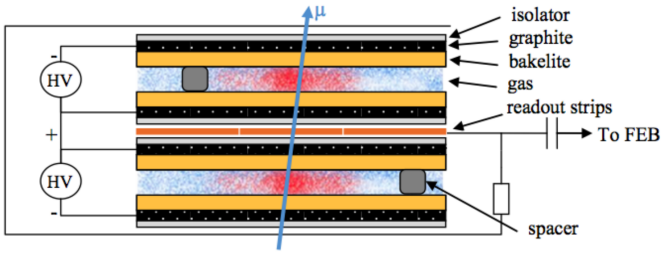

Figure 4.13: A diagram of a CMS RPC. Figure from [10].

4.2.6 Acquiring Data with CMS

At the CMS experiment, bunches of protons cross every 25 ns resulting in a collision rate of 40 MHz. It is not possible nor necessary to record each individual interaction so only events of interest are recorded. This is accomplished with a two step triggering process consisting of the level-1 (L1) trigger and high level trigger (HLT). The L1 trigger selects events and reduces the data rate to $\mathcal{O}(10^2\,\text{kHz})$. Then the data acquisition (DAQ) system collects data and passes it through the HLT. This process further selects events and reduces the data rate to $\mathcal{O}(10^2\,\text{Hz})$. Overall, the trigger reduces the rate by a factor of 10^6. Diagrams of the trigger and DAQ systems are shown in Fig. 4.14. The resulting data

is then sent to a world wide computing grid for full reconstruction and analysis [74, 83, 4].

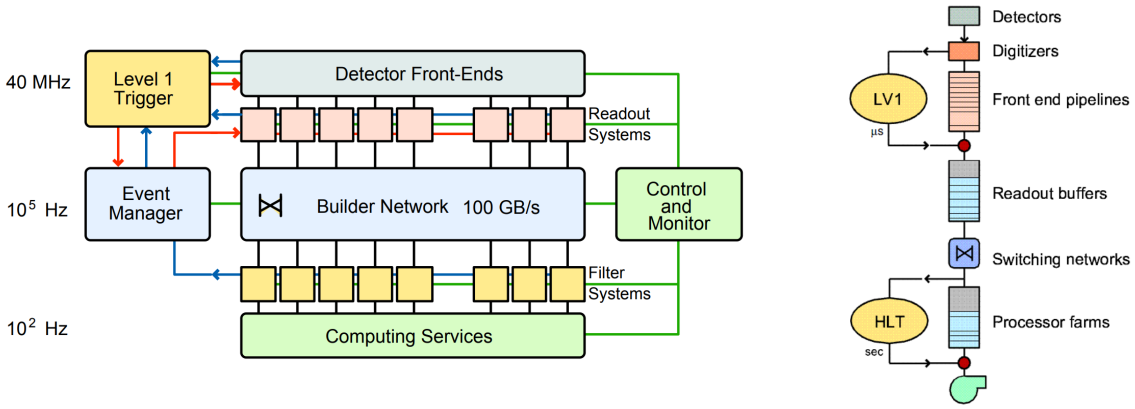

Figure 4.14: (Left) A diagram of the CMS Data Acquisition system (Right) A diagram of a CMS Trigger system. Figures from [4].

Level-1 Trigger

The L1 trigger analyzes every proton bunch crossing and provides a quick decision in 3.2 μs. In order to uphold the time constraints, only an easy to process subset of coarsely segmented data can be used to issue the decision; this is done with the calorimeters and muon system. With the calorimeters, the trigger issues a decision based on the energy in an event. With the muon system, the muon p_T can be quickly estimated and used to make a decision. If measurements in an event are above specified thresholds, then an L1 accept is issued and the event is passed onto the DAQ system. The L1 Trigger is implemented in programmable electronics, utilizing FPGAs whenever possible and ASICs with memory look up tables when necessary [74, 4].

Data Acquisition and High Level Trigger

Once an L1 accept is issued, the full granularity of the collision event is readout by the DAQ system. During this process, the data is passed through the HLT consisting of a filter farm of over 1,000 commercial processors. The HLT utilizes the full granularity of the event to make decisions on more complicated calculations. These decisions happen on the order of seconds and are more similar to selections in a physics analysis. Data passing these criteria are then stored at an $\mathcal{O}(10^2\,\text{Hz})$ rate and sent to a world wide computing

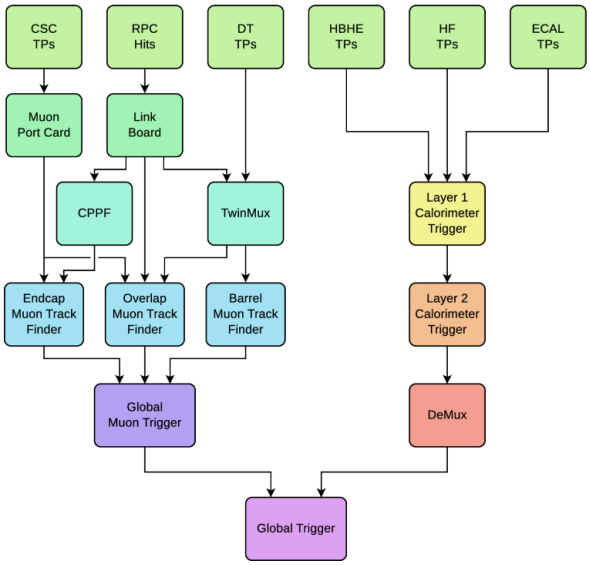

Figure 4.15: A diagram of the components in the CMS Level-1 Trigger than can lead to a L1 accept. Figure from [11].

grid [74, 4]. The reconstruction and analysis of data recorded by the CMS experiment are discussed starting in Chapter 7.

Chapter 5

Gas Electron Multipliers

> *Quando si è molto giovani capita di non sapere bene chi si è e che cosa si vuole dalla vita. Indubbiamente però noi tutti disponiamo di un misterioso filo conduttore che prima o poi finirà per farci scegliere ciò che per indole è già latente in noi, e servirà a costruire la nostra personalità.*
>
> Walter Bonatti, *Un Mondo Perduto*

The 20th century saw an explosion of detector technologies, but none have proved as timeless as the gaseous detector. This detector is based on the simple principle of creating electron-ion pairs that drift in an electric field. Now this detector technology can be seen virtually everywhere. It is often seen in films in the form of a "Geiger" counter.

Despite a theoretically simple idea, gaseous detectors evolved throughout the 20th century as electronic capabilities improved. This includes the rise of personal computers and associated techniques for printing micro-chips. This allowed for the development of micro pattern gas detectors (MPGDs), which can now be seen in every major experiment at the LHC and are frequently included as crucial components for proposed experiments at future colliders.

5.1 Interactions of Ionizing Radiation with Matter

The design of gaseous detectors is based around the interactions of ionizing radiation with matter, specifically the interactions of charged particles and photons with matter. One

of the key principles of a gaseous detector is the formation of an electron avalanche in a high electric field. This process will be described in the next section. The radiation basics described in this section are used to explain the fundamental processes inside of gaseous detectors. The equations summarizing these processes rely on a number of commonly used variables listed in Table 5.1 [85]. Throughout this section, the interactions with the two gases selected for use in the CMS GEM project are highlighted: argon and carbon dioxide.

Variable	**Meaning**
N_A	Avogadro's number
m_e	Electron mass
r_e	Classical electron radius
A	Atomic mass of the absorbing material
z	Charge number of the incident paricle
β, γ, c	Have the normal relativistic meanings

Table 5.1: Variables commonly used when discussing ionizing radiation.

5.1.1 Charged Particles

In each interaction of a charged particle in matter energy is transferred from the incident particle to the medium resulting in a loss of energy by the incident particle. This loss depends on the energy of the incident particle and is summarized in Fig. 5.1. As the energy of the incident particle increases, the mechanism for energy loss changes. For gaseous detectors, the two most relevant interactions are ionization and bremsstrahlung.

Ionization

One mechanism by which charged particles can lose energy in matter is by colliding with atomic electrons. The resulting collision transfers enough energy to the atomic electron to overcome the binding energy. This results in the creation of an electron-ion pair, illustrated in Fig. 5.2.

The average differential energy loss (or mass stopping power), due to ionization, for

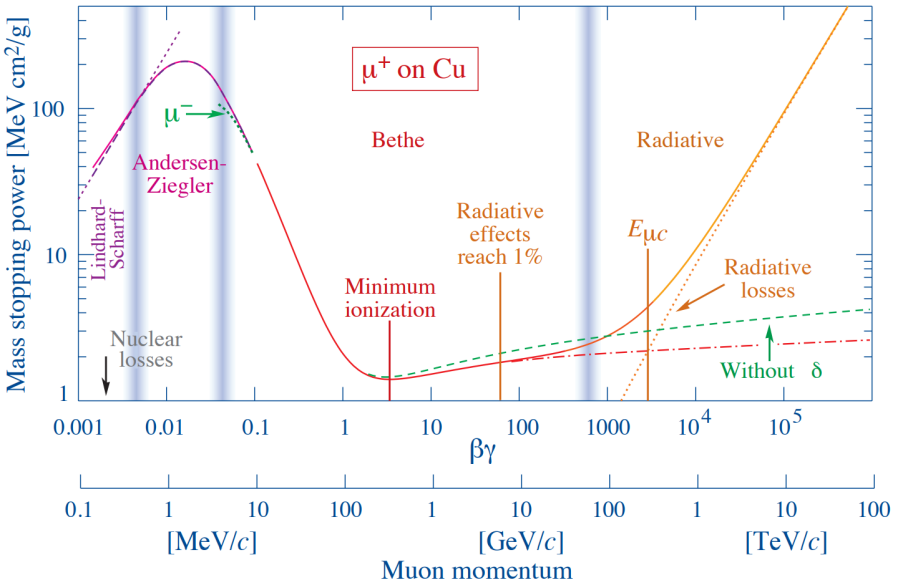

Figure 5.1: The energy loss by a muon traveling through Copper; The different regions highlight different mechanisms for energy loss. Figure from [12].

intermediate relativistic charged particles is described by the Bethe-Bloch equation

$$\langle \frac{dE}{dx} \rangle = \frac{-4\pi N_A r_e^2 m_e c^2 z^2}{A\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \tag{5.1}$$

where I is the mean excitation energy and W_{max} is the maximum energy transfer to an electron in a single collision [12].

This semi-classical expression describes the energy loss as a function of $\beta\gamma$. The energy loss initially decreases as the velocity of the particle increases due to the β^{-2} term. Then, minimum ionization occurs in the region near $\beta\gamma \approx 3$. This region of minimum ionization is often exhibited by muons that originate from high energy collisions like those at the LHC or from cosmic rays [85].

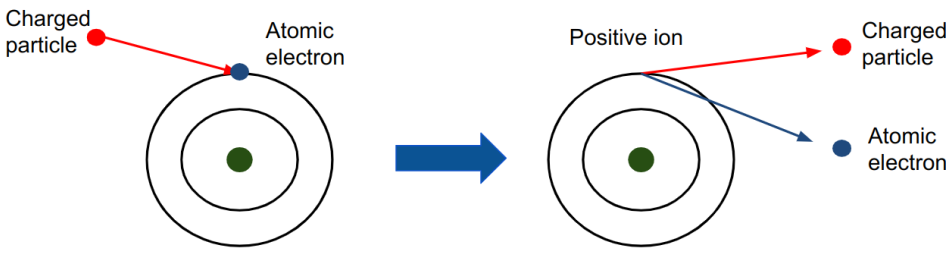

Figure 5.2: A charged particle collides with an atomic electron producing an electron-ion pair.

As the velocity increases further into the relativistic regime, the space in front of the relativistic particle becomes compact, creating a dense region with a higher probability of losing energy. This is reflected by the $\ln(\gamma^2)$ term which dominates at these energies. The increased energy loss is the "relativistic" rise. Note that the $\ln(\gamma^2)$ is corrected with $\delta(\beta\gamma)$ to reduce the rise to $\ln(\gamma)$. This correction is due to the density effect where atomic electrons cause a screening of the projectile's electric field [85, 86, 12, 25]. This behavior of the Bethe-Bloch equation is summarized by the red line in Fig. 5.1.

In gaseous detectors, incident charged particles give rise to electron-ion pairs with free electrons that have enough energy to generate secondary electron-ion pairs. This is the initial step for detecting a charged particle. The CMS GEM detectors are designed to detect clusters of electron-ion pairs that form when a charged muon in the minimum ionization region travels through the detector.

Bremsstrahlung

The process of a charged particle decelerating is called bremsstrahlung (bremsen: to brake, Strahlung: radiation). This "braking radiation" broadly means any type of charged particle deceleration. However, the term bremsstrahlung in English is associated with the specific case of charged particles decelerating in matter. In this process, the incident charged particle is deflected by the electromagnetic field of the nucleus and loses energy by emitting an x-ray photon [85, 12]. The cross section for emitting a photon is inversely proportional to the particle mass squared. Thus, electrons significantly lose energy via this process. Bremsstrahlung for an electron is illustrated in Fig. 5.3.

In gaseous detectors, bremsstrahlung is the cause of energy loss for high energy electrons and is one source of photons within the detector. This energy loss rises nearly linearly for high-energy electrons; the behavior is shown in the radiative region of Fig. 5.1. It can be approximated with

$$\left|\frac{dE}{dx}\right| \approx \frac{E}{X_0}\,, \tag{5.2}$$

where X_0 is the radiation length. The radiation length is the characteristic amount of matter traversed by a high energy electron and is defined as the mean distance where the

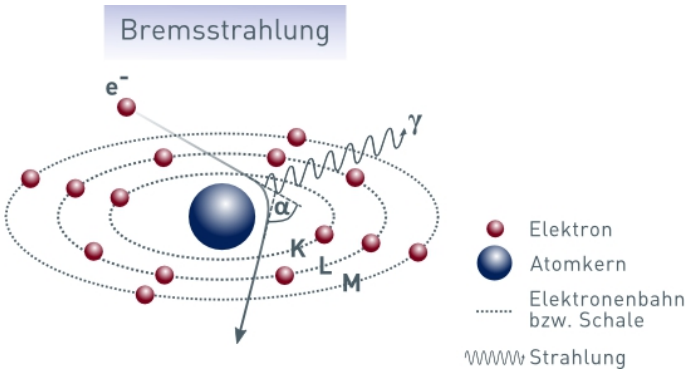

Figure 5.3: As an electron (Elektron) travels near the nucleus (Atomkern), the electron loses energy by radiating a photon (Strahlung: radiation). Figure from [13].

electron loses all but $1/e$ of its energy by bremsstrahlung. The radiation length is defined as

$$X_0 = \frac{A}{4\alpha r_e^2 N_A}\left\{Z^2[L_{rad} - f(Z)] + ZL'_{rad}\right\}^{-1}, \tag{5.3}$$

where $f(Z)$ is an infinite sum characteristic to the material and L_{rad}, L'_{rad} are properties of the material. Typical values of L_{rad}, L'_{rad} and an approximation of $f(Z)$ can be found in [12].

The radiation lengths for the CMS GEM gases are $19.55\,\mathrm{g/cm^2}$ for argon and $36.19\,\mathrm{g/cm^2}$ for carbon dioxide[1] [71].

5.1.2 Photons

Photons lose energy in matter by interacting via absorption and scattering. The reduction of intensity of a incident beam of intensity I_0 is a function of distance travelled in matter and is expressed as

$$I(x) = I_0 e^{-\mu x}, \tag{5.4}$$

where μ is the absorption coefficient[2] [87]. This coefficient depends on both the material used and the energy of the incident photon. It is summarized by the equation

$$\frac{\mu}{\rho}(E_\gamma) = \frac{N_A}{A}\sigma_{photoelectric}(E_\gamma) + Z\frac{N_A}{A}\sigma_{scattering}(E_\gamma) + \frac{N_A}{A}\sigma_{pair\ production}(E_\gamma), \tag{5.5}$$

[1] Often radiation lengths for gases is given in $\mathrm{g/cm^2}$.
[2] μ is sometimes called the linear attenuation coefficient

where σ is the cross section associated with the three major electromagnetic properties by which photons lose energy. These are:

1. Photoelectric effect,

2. Scattering (Compton, Rayleigh, and Thomson),

3. Pair production,

which are described below [86, 25, 85].

Photoelectric Effect

The dominant interaction with matter for low energy photons is the photoelectric effect. Photons with an energy $h\nu$ that exceeds the binding energy E_b of an atomic electron can be absorbed and cause the emission of an electron with kinetic energy T_{e^-}

$$T_{e^-} = h\nu - E_b \,. \tag{5.6}$$

In the case that the energy of the absorbed photon is near the binding energy, the photoelectron is emitted at nearly 90° with respect to the incident photon direction. As the energy of the photon increases, the photoelectron is emitted in a more forward direction [85].

The cross section of a photon being absorbed by an outer orbital electron is relatively small, so frequently photoelectrons originate from the K-shell of the atom. The hole left by the emitted electron is filled by high orbital electrons. The subsequent atomic electron transitions cause the emission of photons, which is commonly called 'fluorescence'. Fluorescence photons frequently cause additional photoelectrons. The excess energy of the transitioning electron can cause a second electron to be emitted; this process is called "Auger emission". The photoelectric effect and subsequent fluorescence is illustrated in Fig. 5.4 [16, 85].

Photons emitted by the bremsstrahlung process in gaseous detectors are typically in the x-ray regime with enough energy to be absorbed by atomic electrons and create photoelectrons which can contribute to an avalanche process. Unfortunately, these low

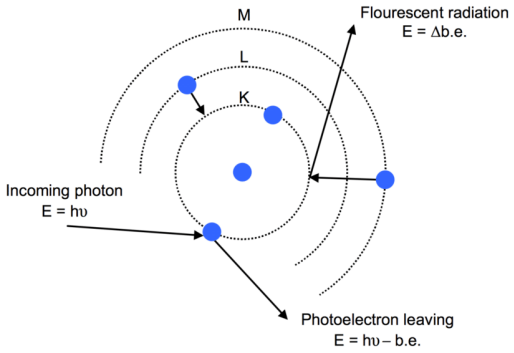

Figure 5.4: An atom emitting an electron via the photoelectric effect and the subsequent fluorescence caused by an electron transition. Figure from [14].

energy particles originating from secondary interactions can greatly increase the time length of a signal in a detector. Therefore, quickly absorbing low energy photons and photoelectrons is a crucial point of study in gaseous detectors.

Compton Scattering

The dominant photon scattering process relevant for this discussion of gaseous detectors is Compton scattering. This occurs when a photon with energy much higher than the binding energy of an atomic electron, $h\nu >> E_b$, scatters with an outer orbital (quasi-free) electron. The process is described graphically in Fig. 5.5.

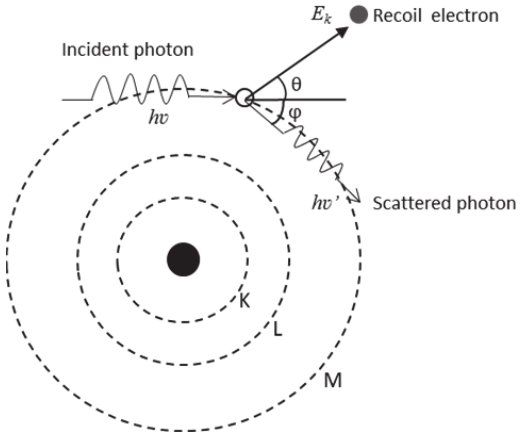

Figure 5.5: Compton scattering of a photon with an atomic electron. This process releases the electron from the atom. Figure from [15].

This occurs for photons with energies of 100 keV to $>$ 1 GeV and is described by the

equation

$$\lambda' - \lambda = \frac{h}{m_e c}(1 - \cos\theta)\,, \tag{5.7}$$

where λ' is wavelength of the scattered photon, λ is the wavelength of the incident photon, and θ is the angle between the incident photon trajectory and the scattered photon trajectory. This results in a scattered photon with energy

$$h\nu' = \frac{h\nu}{1 + \epsilon(1 - \cos\theta)}, \text{ where } \epsilon = \frac{h}{\lambda m_e c} \tag{5.8}$$

and a recoil electron with kinetic energy

$$T_{e^-} = h\nu - h\nu'\,. \tag{5.9}$$

The interaction probability increases with the electron density of the material. The differential cross section describing this probability is given by the Klein-Nishina formula

$$\frac{d\sigma}{d\Omega} = Z r_e^2 \left(\frac{1}{1 + \epsilon(1 - \cos\theta)}\right)^2 \left(\frac{1 + \cos^2\theta}{2}\right) \left(1 + \frac{\epsilon^2(1 - \cos\theta)^2}{(1 + \cos^2\theta)[1 + \epsilon(1 - \cos\theta)]}\right) \tag{5.10}$$

[16, 25, 85, 86].

High energy x-rays are frequently used to study signal formation (signal duration, shape, etc) in gaseous detectors. These x-rays generate electron-ion pairs via Compton scattering and cause signals just like the electron-ion pairs created by charged particles.

Photons emitted via bremsstrahlung can interact within gaseous detectors via Compton scattering contributing to the development of an avalanche. As mentioned earlier, unabsorbed photons and additional electrons can lengthen the signal duration worsening the time resolution.

Electron-Positron Pair Production

The most important interaction for high energy photons is the creation of an electron-positron pair. This most often occurs when a photon with energy greater than $2m_e(\approx 1.022\,\text{MeV})$ is absorbed by the intense electric field near the atomic nucleus. The energy required is near the mass of two electrons because the nucleus satisfies the conservation of momentum, but gains little recoil energy so nearly all energy goes into producing

the electron-positron pair. The conservation of momentum can also be satisfied by the interaction with an atomic electron, however the recoil energy is much greater and in this case the energy of the incident photon must be above $4m_e$ [16, 25, 85].

Pair production is not one of the dominant processes in the CMS GEM detectors, but is mentioned here for a complete overview of the photon interactions within gaseous detectors. A summary of the photon cross section as a function of energy is shown in Fig. 5.6. This takes into account the photoelectric effect, Compton scattering, and pair production to summarize the cross sections for the gases used in the CMS GEM detectors.

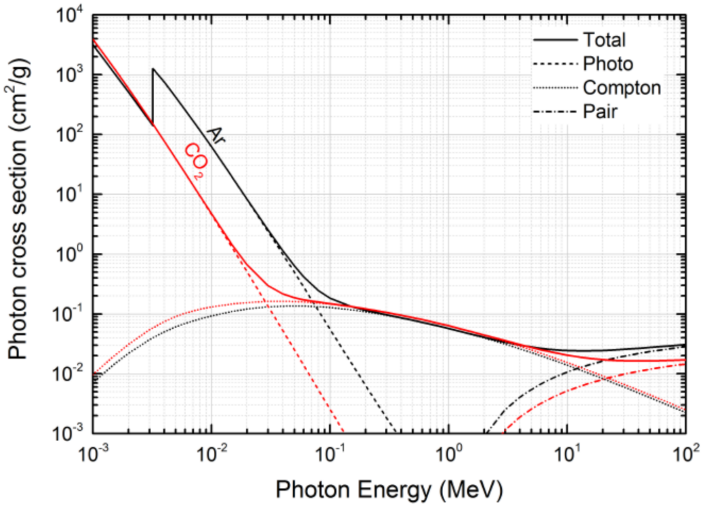

Figure 5.6: The photon cross section as a function of photon energy for argon and carbon dioxide. Figure from [16] using [17].

5.2 Principles of Gaseous Detectors

As an ionizing particle enters a gaseous detector, it sets off a cascade that eventually results in a signal. All of the interactions described in the previous section are used to develop a signal and detect an ionizing particle.

5.2.1 Ionization

This process begins when a charged particle travels through the gas volume and collides with an atomic electron creating an electron-ion pair (primary ionization). The now free electron often has enough energy to undergo a second collision creating additional

electron-ion pairs (secondary ionization), resulting in an electron-ion cluster. The total number of electron-ion pairs from an ionization event is the sum of primary and secondary pairs $n_T = n_P + n_S$. The average number of total pairs becomes

$$\langle n_T \rangle = \frac{\langle \frac{dE}{dx} \rangle L}{W}\,, \tag{5.11}$$

where $\langle dE/dx \rangle$ is given by the Bethe-Bloch equation, L is the path length, and W is the average energy required to create an electron-ion pair [25, 85].

The number of primary ionization pairs follows a Poisson distribution

$$P(k, \langle n_P \rangle) = \frac{\langle n_P \rangle^k e^{-\langle n_P \rangle}}{k!}\,, \tag{5.12}$$

where k is the number of clusters per unit length. However, the total number of pairs is heavily influenced by the production of photons which create subsequent pairs via the photoelectric effect. Thus, the total number of pairs follows a Landau distribution with an average that can be approximated as

$$\langle n_T \rangle \approx \frac{E_\gamma}{W}\,, \tag{5.13}$$

which utilizes the approximation that $E_\gamma >> E_{photoelectric}$ [25, 85].

The probability of a charged particle traveling through the gas volume without creating an electron-ion pair is $P(0, \langle n_P \rangle)$. This can be used to calculate the theoretical maximum efficiency for a gaseous detector

$$\epsilon = 1 - P(0, \langle n_P \rangle) = 1 - e^{-\langle n_P \rangle}\,. \tag{5.14}$$

The CMS GEM detectors use an argon/carbon dioxide mixture (70/30%) and have an ionization region that is 3mm in thickness. This results in a maximum theoretical efficiency of $\epsilon = 99.98\%$. Average values for commonly used gases are listed in Table 5.2.

5.2.2 Motion in Gases

Inside of a gaseous detector is a strong electric field. As the ions and electrons move throughout the volume, the particles exhibit two types of motion: thermal expansion (i.e. diffusion) and drift due to the electric field.

Gas	$\langle n_P \rangle$ [cm^{-1}]	$\langle n_T \rangle$ [cm^{-1}]	$\left\langle \frac{dE}{dx} \right\rangle$ [keV cm^{-1}]
Ar	25	100	2.4
CO_2	35	100	3.4
i-C_4H_{10}	90	220	4.5
CF_4	63	120	6.4

Table 5.2: The average number of primary and secondary ion pairs, and the average energy loss for a MIP with charge $q = e$ for common gases used in gaseous detectors. These values are true at $T = 20\,^\circ$°C and $p = 1$ atm. Adapted from data in [25, 49, 16].

Drift

After the cluster of electron-ion pairs forms, the charged particles drift in the electric fields present in the detector. The ions and electrons have a thermal velocity, but also drift parallel to the electric field. The velocity is limited by collisions with other atoms. This interaction can be thought of as a friction-like force which reaches an equilibrium with the forces from the electric and magnetic fields. Thus, the electrons and ions quickly reach a terminal velocity. The positively charged ions have a larger size and heavier mass resulting in more frequent collisions and a lower terminal velocity. In comparison, the electrons collide less often and obtain a higher terminal velocity. Electrons are also more significantly affected by the presence of a magnetic field. The electron motion with constant drift velocity is described as

$$\left\langle m\frac{d\vec{v_D}}{dt} \right\rangle = e(\vec{E} + \vec{v_D} \times \vec{B}) - \frac{m}{\tau}\vec{v_D} \approx 0\,, \tag{5.15}$$

where τ is the characteristic time between collisions with the gas atoms, $\vec{E}$ is the electric field, and $\vec{B}$ is the magnetic field [25, 85].

An important characteristic is μ_-, the electron mobility, and is defined as

$$\mu_- = \frac{e\tau}{m}\,. \tag{5.16}$$

Using the electron mobility and the Lamor frequency ($\omega = eB/m$), equation 5.15 can be solved for the electron drift velocity

$$\vec{v_D} = \frac{\mu_- E}{1 + \omega^2\tau^2}\left[\hat{E} + \omega\tau\hat{E} \times \hat{B} + \omega^2\tau^2(\hat{E} \cdot \hat{B})\hat{B}\right]. \tag{5.17}$$

This equation describes the electron drift inside of the detector and can be further understood by considering two categories of gases used in detectors: 'hot' and 'cold' gases [25].

'Hot' gases contain atoms (or molecules) with few low lying energy levels. In 'hot' gases, τ is not constant as the electric field increases. An example of a 'hot' gas mixture is argon gas with methane [25].

'Cold' gases contain many low lying energy levels. This results in a constant τ and therefore constant μ_-. Thus, in absence of a magnetic field, the drift velocity of an electron in a 'cold' gas is

$$v_D = \mu_- E \,. \tag{5.18}$$

It is important to note that the electron mobility is greatly affected by the pressure and temperature of the gas. The gas mixture used by the CMS GEM project (Ar/CO_2) a cold gas mixture [25, 85].

As the electrons travel in the gas mixture, their typical thermal velocity is $\approx 10^7$cm/s with a net drift in the direction of the electric field. Whereas the positive ions have a significantly slower velocity due to their larger size and heavier mass, the typical ion thermal velocity is $\approx 10^5$cm/s. In the case of the CMS GEM gas mixture (Ar/CO_2 at 70/30%) the electron velocity is $7{\cdot}10^6$cm/s in an electric field of $E = 2.6\,\mathrm{kV/cm}$ [25, 85].

Diffusion

The newly created electrons and ions will diffuse away from areas of high concentration to low concentration. This effect is more significant for electrons which have a higher velocity and longer mean free path. The amount of diffusion is affected by the forces from the electric and magnetic fields inside of the detector. High electric fields limit diffusion in the transverse directions as the electrons drift parallel to the field lines. Electrons spiral around the magnetic field lines, again limiting the amount of movement transverse to the magnetic field. Containing this diffusion to a small area during the formation of a signal is critical for high spatial and time resolution [85].

5.2.3 Avalanche

In areas of high electric field, freed electrons can gain enough energy to cause secondary ionization. When this happens the electrons multiply forming an avalanche[3]. Typical electron avalanches have a maximum size of $\approx 10^8$ electrons due to charge repulsion preventing further growth [85].

The number of electron-ion pairs produced by a single electron traveling 1 cm along the electric field is called Townsend's first ionization coefficient α. This coefficient is exactly the inverse of the mean free path for ionization $\alpha = \lambda^{-1}$ and can be approximated with

$$\frac{\alpha}{P} = A \exp\left(\frac{-BP}{E}\right), \tag{5.19}$$

where P is pressure, E is the electric field, and A and B are properties of the gas [85].

Let the direction along the electric field be the y-axis. The change in the number of electron-ion pairs between y and $y + dy$ is

$$dN = \alpha N dy\,. \tag{5.20}$$

Inside a uniform electric field, this results in a total electron multiplication over a distance Δy given by

$$M(\Delta y) = e^{\alpha \Delta y}\,. \tag{5.21}$$

When the electric field is non uniform, as is often the case, α becomes a function of position and the total multiplication becomes

$$M(x) = \exp\left[\int_x^0 \alpha(y) dy\right]. \tag{5.22}$$

The resulting avalanche has a tear-drop shape and is shown in Fig. 5.7.

The electrons will diffuse during the avalanche process. As a result, the electrons will distribute over the anode. Choosing gases and electric fields which limit this diffusion and improve the resolution of the detector are important aspects of study.

As mentioned in the section on ionizing radiation (Section 5.1), electrons can ionize atoms with electrons in high orbitals, causing the release of photons when the atomic

[3]The avalanche is sometimes called a "Townsend" avalanche or "Townsend" discharge after John Sealy Townsend who discovered the process in 1897.

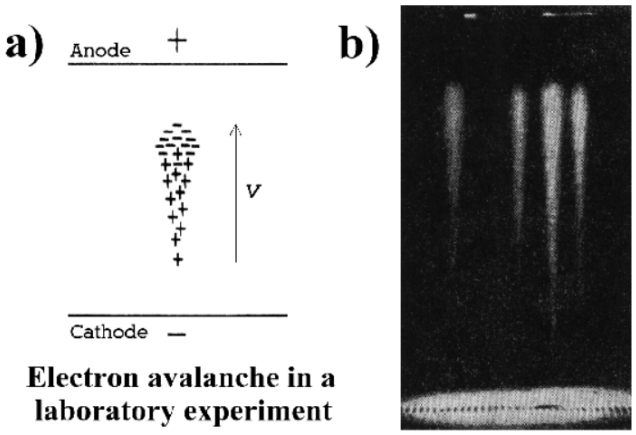

Figure 5.7: The characteristic teardrop shape of an electron avalanche inside of a gaseous detector. Figure from [18].

electrons transition. Additionally, the electrons can radiate photons via bremsstrahlung. The photons interact via the photoelectric effect and Compton scattering to produce additional electron-ion pairs (and more photons). As a result, another term is needed to describe the contribution from photons, this is Townsend's second ionization coefficient γ [49].

This second coefficient depends on the amplification, A, of the first avalanche. When A increases but $A\gamma << 1$, then $A\gamma$ gives the probability of producing an after-discharge. If the photons travel far, then the size of the avalanche increases which decreases the overall resolution. "Quenching" gases like carbon dioxide are present in gas mixtures to absorb these extra photons [49].

Up to a certain amplification, the signal is proportional to the number of primary electrons. This is called the "proportional" regime. If the gas mixture poorly absorbs photons, then an ionization event will cause the avalanche to spread through the entire gas volume as shown in Fig. 5.8. This is the "Geiger" regime. In this case, the detector has poor resolution, but becomes a highly effective counter requiring little signal amplification. When the amplification is further increased and $A\gamma > 1$, the gas undergoes dielectric breakdown [49, 25]. These operational regimes are summarized in Fig. 5.9.

The CMS GEMs need to have spatial resolution of 0.8 mm in the azimuthal direction and temporal resolution of less than 10 ns [26]. To meet these demands, the detector is

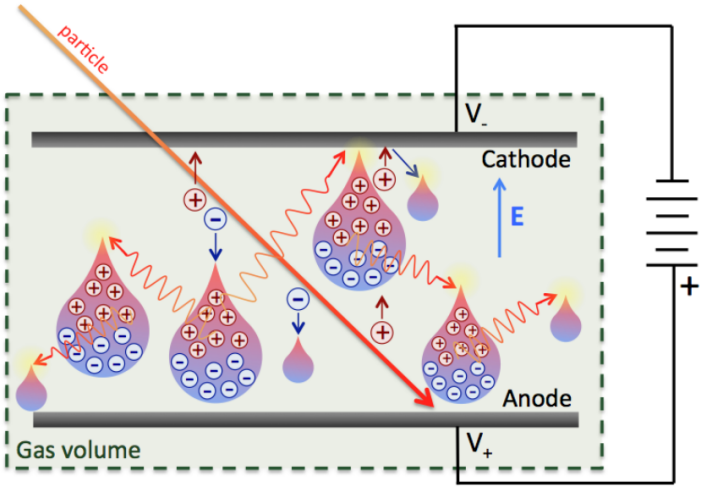

Figure 5.8: The "Geiger" regime where emitted photons travel long distances and create avalanches throughout the detector volume. Figure from [19].

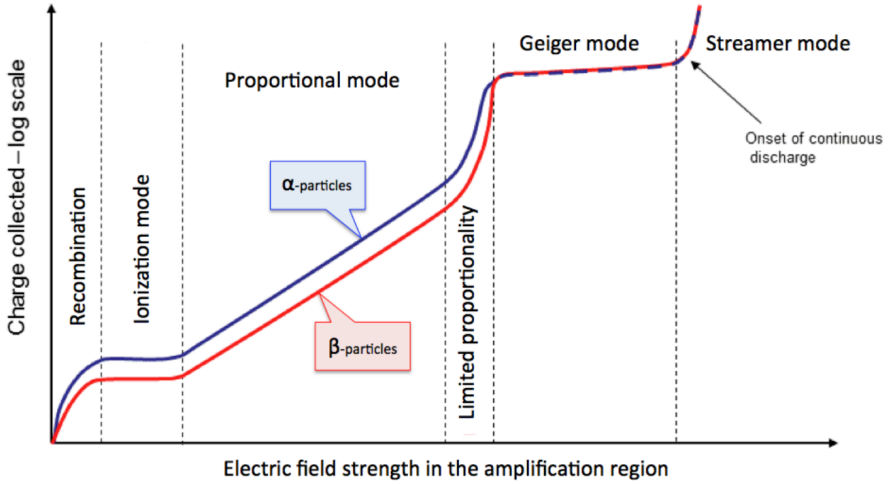

Figure 5.9: As the electric field increases, the amount of charge in the gas generated during an ionization event increases until the dielectric breakdown of the gas. Figure from [19].

designed to operate in the "proportional" regime with a gas mixture containing carbon dioxide for quenching.

5.2.4 Signal Generation

The resulting electron avalanche drifts toward a read-out anode, inducing a signal. The induced signal takes on the order of 1 to 10 ns to form after an ionization event. Some gaseous detectors (like cathode strip chambers) also have an additional signal component

due to the ions; ions are 10^3 times slower and as a result produce a much longer signal on the cathode. Novel gaseous detectors (like GEMs) have faster signals by using only the anode signal [16, 25].

5.3 Evolution of Micro Pattern Gas Detectors

Gaseous detectors underwent a boom in the 1940s and to this day they are a fundamental tool for detecting particles. When compared to silicon detectors, gaseous detectors can cover much wider areas for significantly cheaper costs with slight decreases in position and time resolution. This makes gaseous detectors ideal for detecting ionization over large areas as is often required in the muon system at a collider experiment. Through the decades, gaseous detectors evolved significantly. In each iteration the detectors improved in spatial resolution, reduced in recovery time, and experienced less aging effects [16, 86].

The gas detector boom started in the early 1900s with the single wire proportional counter (SWPC). This was first designed by Rutherford and Geiger in 1908 and was first constructed by Müller in 1928. A SWPC consists of a metal wire anode surrounded by a coaxial metal cylinder cathode, as shown in Fig. 5.10.

This chamber is filled with an inert gas that produces a high amount of electron-ion pairs and a quenching gas to absorb excess photons inhibiting a self sustaining discharge (i.e. dielectric breakdown). Early versions of the SWPC were operated in the "Geiger" regime and are referred to as "Geiger counters". When an ionization event occurs it releases a large amount of electrons which induce a signal on the wire anode. All that can be determined by a Geiger counter is that an ionizing particle passed somewhere in the detector volume. As electronics improved, the SWPC was operated in the "proportional" regime where the amount of collected charge is proportional to the energy deposited by the incident particle [16, 25, 86].

The main disadvantage of the SWPC is an inability to precisely locate where the ionizing particle passed. All that can be concluded is that the particle passed somewhere in the detector. This issue was addressed in the 1960s by G. Charpak with the multi wire proportional counter (MWPC)[4]. A MWPC is operated in the "proportional" regime

[4]G. Charpak was awarded the Nobel Prize for the MWPC in 1992. This was the last time (at least as

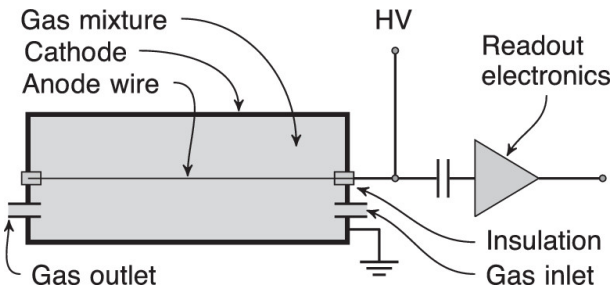

Figure 5.10: The Single Wire Proportional counter consists of an anode wire surrounded by a coaxial metal cylinder cathode and filled with a mixutre of inert and quenching gases. Figure from [20].

and consists of parallel layers of anode wires arranged in a grid alternating with cathode strips. The anode wires create a dense electric field. When an ionizing particle travels through the chamber, clusters of electron-ion pairs form. The electrons drift into regions of high electric field near the anode wires forming an avalanche. The 2D position of the passing particle can be precisely reconstructed by comparing the signal strength in the grid of anode wires and cathode strips, as shown in Fig. 5.11 [16, 86].

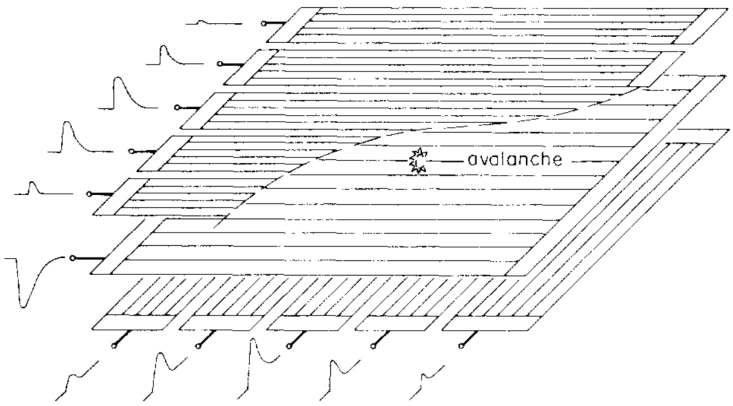

Figure 5.11: A grid of anode wires and cathode strips used to precisely detect the position of an avalanche. Figure from [21].

MWPCs are unable to be used in high radiation environments. A high flux of ionizing particles causes a high charge density within the chamber. This distorts the electric field lines inside of the chamber—limiting the gain resulting in efficiency loss. This problem motivated the need for a new detector design and the development of photolithography

of 2022) that the physics Nobel Prize was awarded to a single person and one of the only prizes awarded for detector instrumentation.

in the 1980s enabled a new generation of gaseous detectors [16, 86].

Photolithography is a class of light based techniques used for printing patterns from thin films (of suitable material) onto a substrate. This allowed for the development of micro-electronics and the rise of personal computers. The new ability to print conducting strips on an insulating surface at 200 µm distances resulted in a new generation of gas detectors called micro-pattern gas detectors (MPGDs) [16, 86].

The first of these new MPGDs was the micro-strip gas counter (MSGC). The traditional anode wires from the MWPC became narrow strips. These narrow anode strips are printed on an insulating board next to wide cathode strips. The strip pattern and resulting electric field are shown in Fig. 5.12. These narrow strips allow for high position resolution. The resulting electric field lines create a high electron collection efficiency and prevent the back-flow of positive ions into the drift region of the detector. By preventing the back-flow of positive ions, the rate capabilities of MGSCs are vastly superior to those of MWPCs [16, 86].

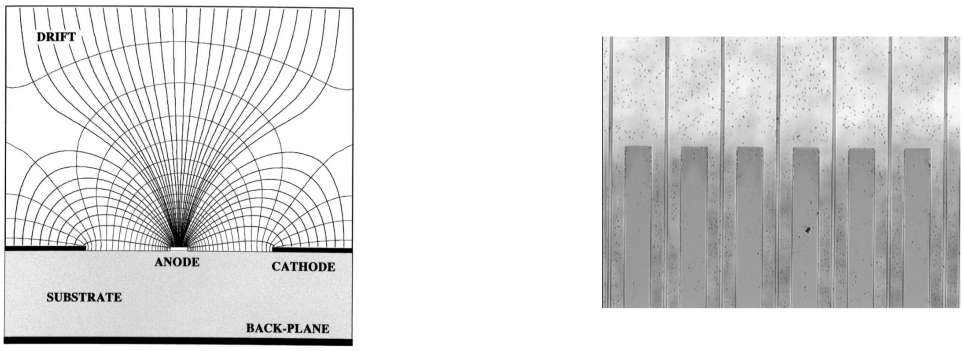

Figure 5.12: (Left) The electric field inside a MSGC. (Right) The individual strips. The narrow strips are anodes and the wide ones are cathodes. Figures from [22].

However, there was a major problem with the MSGC. Events with ionization largely exceeding the average value can occur at collider experiments. When the MSGCs are operated at the high gains necessary to detect minimum ionization, the large ionization events create damaging discharges which fuse the strips resulting in detector "dead spots" as shown in Fig. 5.13 [86].

One solution to this issue is to divide the amplification and readout stages. The discharged probability can be reduced even further by operating multiple amplification

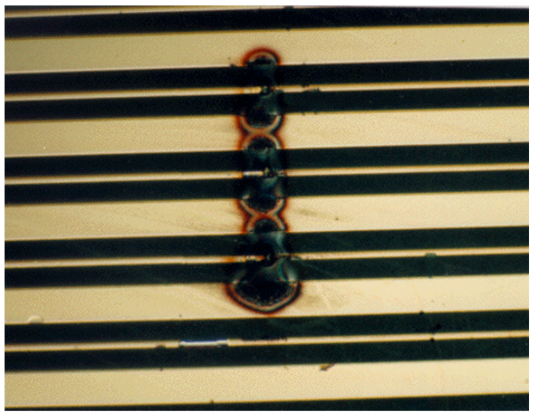

Figure 5.13: A microscopic view of the strips in a MSGC after a destructive discharge. Figure from [19].

stages at reduced gains. In the 1990s, two new types of MPGDs were developed in order to expand on this observation: the micro-mesh gaseous structure (MICROMEGAS) and the gas electron multiplier (GEM) [86, 88].

This provided much needed relief for existing MSGC detectors with discharge issues. The inner tracker for the HERA-B experiment[5] was built from MSGCs with discharge issues. By adding a GEM pre-amplification stage, the tracker was able to be safely operated for ten years [86].

Like the HERA-B experiment, the original design of the CMS tracker included an outer tracker made from MSGCs. This detector was designed to be a 3 m long, 3 m diameter cylinder able to record an average of 6 hits per track (Fig. 5.14) [23]. The design was later abandoned in favor of silicon strips due to the potential for damaging discharges during operation. But the interest in MPGDs at CERN did not end here.

At the same time that CMS was abandoning the MSGC tracker, the GEM technology was being further developed and intensely studied. Progress with GEM and MICROMEGAS led to the creation of the Research and Development group 51 (RD51) at CERN. RD51 is a cross-experiment collaboration created in 2008 to study MPGDs. The progress made in this group directly led to the development of GEM projects in the ALICE, LHCb, and CMS experiments as well as the MICROMEGAS project in the ATLAS

[5]HERA-B was a B meson factory experiment that used beam halo from DESY's electron-ion collider to collide with a thin wire and produce B mesons for observation.

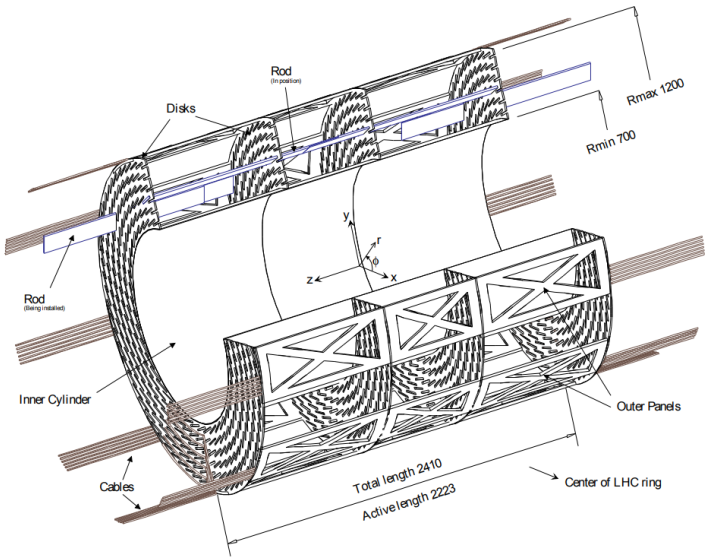

Figure 5.14: The CMS MSGC tracker design which was later replaced with silicon strips, from [23].

experiment.

At the CMS experiment, the GEM technology was chosen for upgrades to the muon spectrometer and this technology will be the focus for the rest of this chapter and the subsequent chapter.

5.4 GEM

Gas electron multipliers (GEMs) are a subset of MPGDs named after a characteristic foil used inside of the chamber. The GEM foil consists of a layer of insulating polymer surrounded on the top and bottom by a conducting layer. Microscopic holes ($\approx 50\,\mu$m in diameter) are etched throughout the foil in high density hexagonal patterns (50-100 holes/mm). The GEM holes have a characteristic hour-glass shape; these holes and the surface of a GEM foil are shown in Fig. 5.15 [16, 25, 27].

Each GEM hole works like a proportional counter with a potential difference applied across the two conducting layers to make an electric dipole with a sharp electric field. This electric field operates in the proportional regime where typical values of ΔV are 350-500 V [27].

A GEM chamber consists of three separate regions: drift, amplification, and induction.

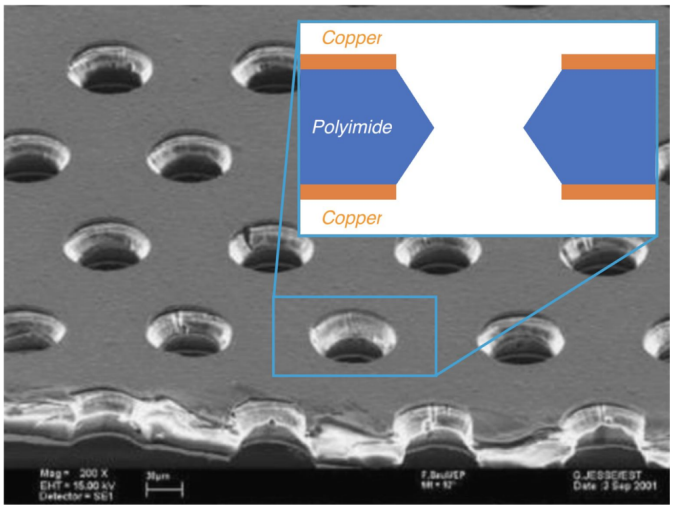

Figure 5.15: A GEM foil shown with a scanning electron microscope (from [24]) and a diagram highlighting the GEM foil (from [25]).

The drift region is optimized based on the gas mixture for the ionizing particles to have a high probability of creating at least one electron-ion pair. The resulting electrons drift to the holes in the GEM foils [27].

The GEM foils compose the amplification region; the sharp electric field within the holes creates an electron avalanche. This amplification process is commonly called multiplication and is shown in Fig. 5.16. The maximum amplification of a single GEM foil is on the order of 10^3. In the case of a detector with multiple GEM foils, the region between foils is called a transfer region [19, 27].

As the electrons enter the induction region, a charge is induced on the readout strips, creating a signal. It is important to note that some electrons drift backwards to the surface the foil. In general, this effect is reduced by increasing the induction field. In other words, as the induction field increases, the electron collection efficiency increases [27].

Many different patterns of gap spacing and numbers of foils have been studied and used. One of the standard foil configurations is three layers of GEM foils in the amplification region—the triple-GEM design.

The spacing between each of the foils is a delicate balance between improving time resolution and decreasing damaging discharge probability. One of the most commonly

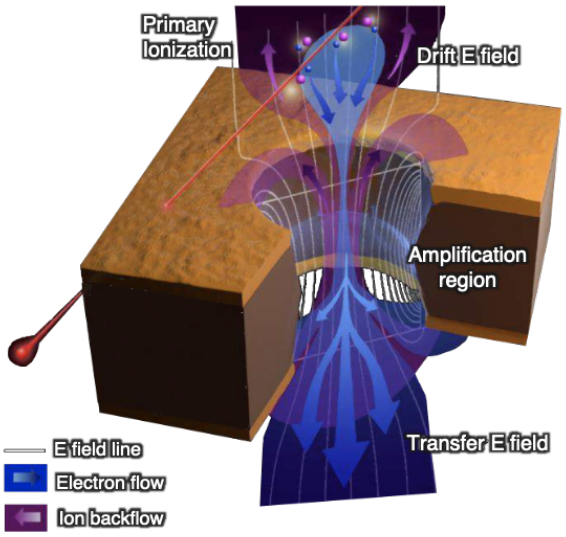

Figure 5.16: The electron multiplication due to the sharp electric field within a GEM hole. Figure from [25] where it was adapted from [26].

used schemes is "3/2/2/2" where there is a 3 mm gap between the drift cathode and first GEM foil (drift region), 2 mm gaps between the foils (i.e. the transfer regions), and a 2 mm gap between final GEM foil and readout (induction region). This scheme is shown in Fig. 5.17.

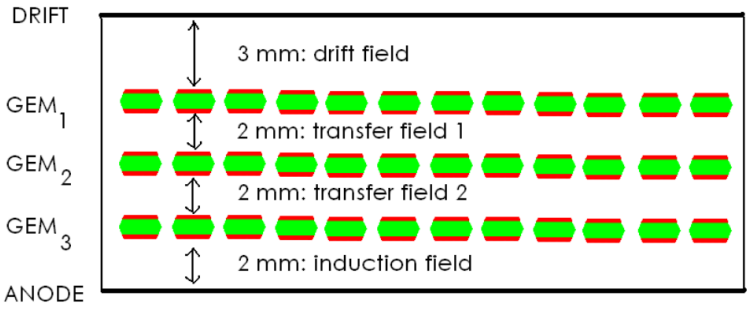

Figure 5.17: The 3/2/2/2 scheme for a triple-GEM detector. Figure from [27].

The 3 mm drift region provides a high probability for ionizing radiation to produce at least one electron ion-pair while keeping the distance as small as possible for better time resolution and rate capabilities. The other 2 mm gaps decrease the probability of a damaging discharge while trying to be as small as possible. The average electric fields for this scheme are shown in Table 5.3.

The main advantage of GEM detectors over MSGCs is the separation of the ampli-

Region	**Gap [mm]**	**Electric field [kV/cm]**
Drift	3	2.4
Transfer 1	2	3.6
Transfer 2	2	3.6
Induction	2	3.6
Region	**Voltage [V]**	**Average Electric field [kV/cm]**
ΔGEM1	400	80
ΔGEM2	360	72
ΔGEM3	325	65

Table 5.3: Standard electric field values for a GEM with a 3/2/2/2 scheme. Table from [19].

fication and readout regions which significantly reduces the risk of damaging discharges. GEM detectors exhibit higher rate capabilities when compared to traditional wire chambers [89]. For these reasons the GEM technology is being widely used at the CERN LHC experiments. For example:

- The ALICE experiment is using a 4 foil GEM detector as part of its time projection chamber.
- The TOTEM experiment is using a triple GEM detector as part of the "T2 telescope".
- The LHCb Experiment is using a triple GEM detector as part of the muon system.

The design chosen for CMS is the triple GEM and the specific details about the chamber design, electronics, and tests are described in the next chapter.

Chapter 6

New GEMs in CMS

Per un corridore il momento più esaltante non è quando si taglia il traguardo da vincitori. E' invece quello della decisione, di quando si decide di scattare, di quando si decide di andare avanti e continuare anche se il traguardo è lontano.

Fausto Coppi

The High Luminosity LHC (HL-LHC) is currently scheduled to begin operation in 2029. The upgraded LHC will have an instantaneous luminosity of $5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$— more than five times the luminosity of the current LHC, resulting in significantly increased background rates. The current CMS experiment is unable to cope with these enormous background rates. Therefore, the CMS experiment will be significantly upgraded over the next years to prepare for Run 4 of the LHC. One component of these upgrades involves improving the triggering and tracking capabilities of the CMS muon system by installing gas electron multipliers (GEMs)

GEMs are novel gas chamber detectors capable of dealing with the high rates expected in the HL-LHC. The GEM upgrade consists of three rings of GEM detectors installed in three different forward regions in each end of the CMS experiment. The LHC Long Shutdown 3, scheduled for 2025, will not provide enough time for the full installation of the GEM system. Therefore, the first ring of GEM detectors (called GE1/1) was installed in the LHC Long Shutdown 2.

This chapter described the production, installation, commissioning, and operation of

the new GE1/1 station in the CMS experiment. During production, I participated in the seventh step of quality control. Then, I used that experience to motivate, design, and lead the "trolley test" a critical electronics test administered directly before installation. In 2021, I served as a co-coordinator for the commissioning of GE1/1 and worked with a team of experts to investigate the noise and VTRx issues. I performed laboratory studies of the VTRx cooling prototypes and participated in the operation of the detector as an on-call expert at the start of data taking.

6.1 The GE1/1 Detector

GE1/1 consists of two layers of GEM chambers paired together into a "super chamber". These two layers are placed next to the first station[1] of cathode strip chambers (CSCs). This region contains $1.6 < |\eta| < 2.2$ and the location is shown in Fig. 6.1.

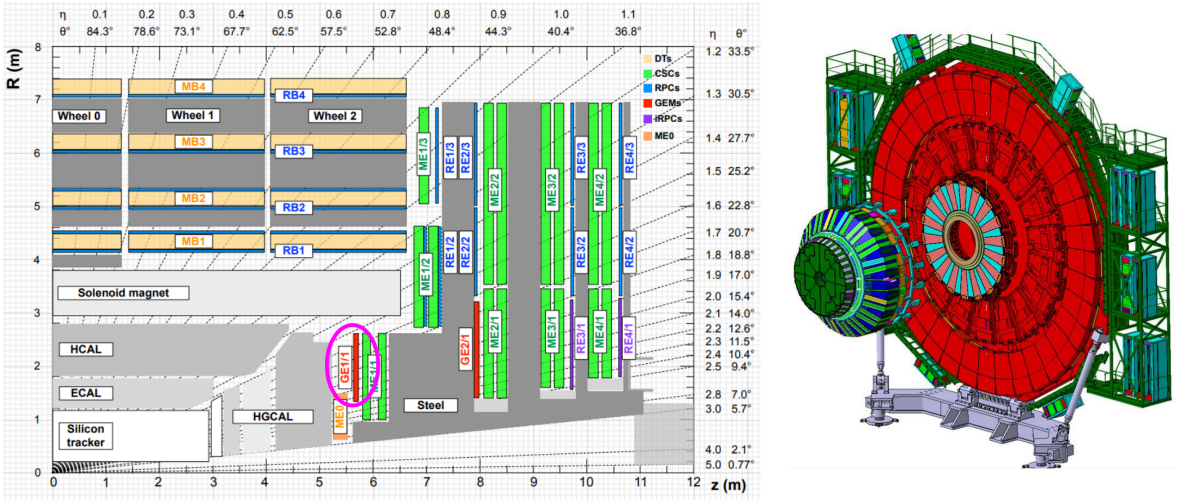

Figure 6.1: (Left) A cross sectional cutout of one quarter of the CMS experiment where the LHC beam travels horizontally across R=0 and collisions occur at R=0 and Z=0. The purple circle highlights the location of GE1/1. Figure from [26]. (Right) The GE1/1 ring inside the CMS endcap. Figure from [26].

GE1/1 provides two muon hits in addition to the six muon hits provided by the CSCs. These additional hits provide a more precise measurement of the muon bending angle in the high magnetic field and thus a more precise measurement of the muon p_T. This measurement reduces the number of "fake" muon triggers, significantly reducing the

[1]First disk of the endcap going out from the interaction region.

L1 trigger rate. Fig. 6.2 shows how the two additional hits improve the bending angle measurement and the predicted trigger rate reduction. This is a crucial step for creating a muon system which can handle the high HL-LHC rates.

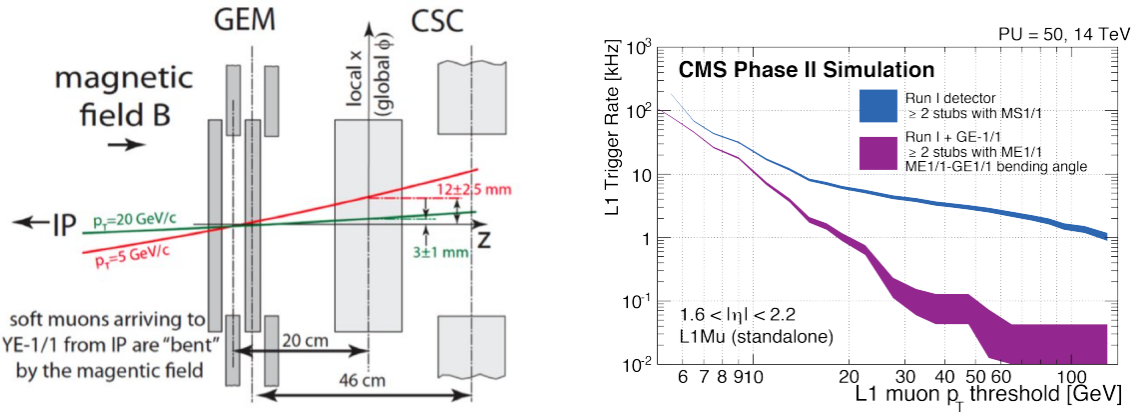

Figure 6.2: (Left) An example of a more precise bending angle measurement due to the addition of two GEM layers. Figure from [26]. (Right) The expected trigger rate due to the addition of two GEM hits based on simulations with Run I rates. Figure from [26].

The specialized GEM chamber designed for GE1/1 consists of a trapezoidal shape. Due to the support braces on the nose of the CMS endcaps, two different types of chambers had to be designed to ensure maximum available coverage. "Short" chambers fit underneath the supports and "long" chambers fit in between the supports. The dimensions of the GE1/1 chambers are shown in Table 6.1.

Parameter	Short Chamber	Long Chamber
Length (along the center)	106.1 cm	120.9 cm
Width (short end)	23.1 cm	23.1 cm
Active volume thickness	0.7 cm	0.7 cm
Total chamber thickness	3.5 cm	3.5 cm
Active readout area	0.345 m^2	0.409 m^2
Active chamber volume	2.6 L	3 L

Table 6.1: Dimensions of the Long and Short GE1/1 chambers, adapted from [26].

The GE1/1 chambers are triple GEM detectors, so the gas volume is divided into four regions as shown in Fig. 6.3. The drift region is the largest with a gap of 3 mm; this is to

ensure a high probability (99.98%) of producing at least one electron-ion pair, but is still kept as small as possible to reduce the time it takes the clusters to reach the first GEM foil. Then, Transfer 1, Transfer 2, and the Induction gap have sizes of 1 mm, 2 mm, 1 mm respectively. Ideally, all three gaps would have a size of 1 mm, but transfer 2 has a gap of 2 mm in order to reduce the probability of a discharge of propagating through the entire stack of foils [19, 26].

A number of gas mixtures were considered for use in the GE1/1 detector, including $Ar/CF_4/C_4H_{10}$, $Ar/CO_2/CF_4$, and Ar/CO_2. The gas mixture selected for GE1/1 is Ar/CO_2 at 70/30%. Other gaseous detectors require the presence of CF_4 to reduce the effects of aging[2]. However, GEM detectors do not use wire anode readouts so they do not exhibit classical aging. This allows for the elimination of CF_4—an increasingly expensive gas with harmful environmental effects that can cause ozone layer depletion and that has a high global warming potential [3] [26].

The voltage scheme in Fig. 6.3 is a typical scheme for optimal gain in GE1/1 with an Ar/CO_2 gas mixture. The values used in operation vary proportionally depending on the observed optimal working point for each detector. Note that the readout is set to 0 V so that the front-end electronics can be nicely placed on the readout board.

Historically, GEM chambers were assembled by gluing the foils to the support frame. But this had a number of drawbacks. For example if the tension was improperly set, there was no way to adjust the tension after the glue dried. The GEM project at CMS uses an innovative technique to adjust the foil tension and this technique has now become an industry standard.

The full GE1/1 chamber design is shown in Fig. 6.4 and the electronics will be discussed in detail in the next section.

[2]Classical aging is a process that occurs in wire chambers where debris collect on the wires reducing the effectiveness of the amplification.

[3]GWP or Global Warming Potential is the amount of heat absorbed by a gas in the atmosphere relative to the amount of heat absorbed by an equivalent mass of CO_2.

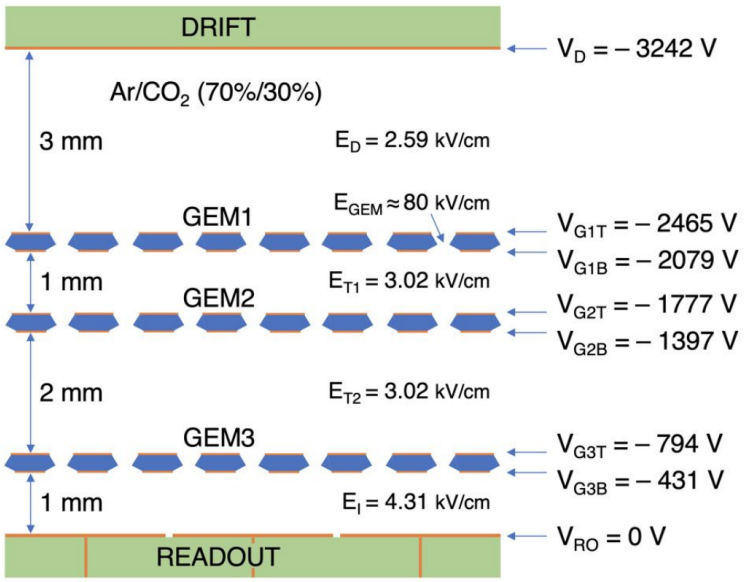

Figure 6.3: The three GEM foils, voltage settings, gaps, and typical values of electric field used in the GE1/1 chambers. The electric field strength used during operation varies proportionally based on the observed optimal working point. Figure from [25].

6.2 GE1/1 Electronics

The GE1/1 detectors have on-chamber front-end electronics that convert analog to digital signals and send data outside of the cavern and to the Cathode Strip Chambers (CSCs) through fiber optic cables. Outside of the main experimental cavern at CMS is a service cavern containing the GE1/1 back-end electronics which pack the data from the GEM inputs.

The electronics described in this section were used extensively for this thesis work.

6.2.1 Front-end Electronics

The GE1/1 front-end electronics consist of four main components: the GEM electronics board, the VFAT3 front-end chip, the OptoHybrid, and the FEAST. The location of the front-end electronics is shown in Fig. 6.4. The FEASTs are radiation hard step-down DC-DC converters that deliver precise output voltages. The other electronic components are discussed in greater detail below.

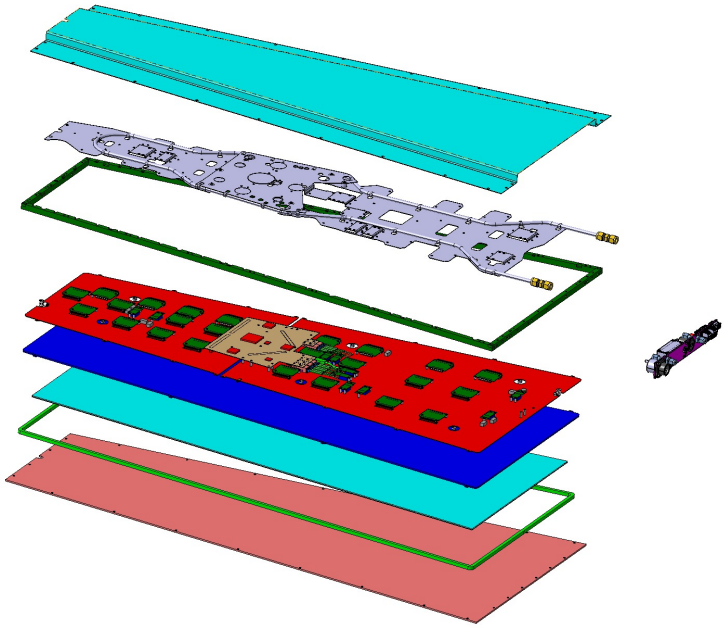

Figure 6.4: An exploded view of a GE1/1 chamber. The red bottom is the drift cathode, the inner light blue area represents the stack of GEM foils, the green outlines are the external frames, the dark blue block is the readout board, the red block is the GEM electronics board equipped with front-end electronics, the silver plate is the copper cooling plate, the light blue cover is a protective aluminium chimney, and the external green block is the HV filter.

6.2.1.1 GEB: GEM Electronics Board

The GEM Electronics Board (GEB) is a multi-layer PCB that sits on top of the GE1/1 readout board. This thin board aids in keeping the GE1/1 detector compact and has three main functions:

- Electrically shielding the electronics from the detector
- Distributing power and ground for all of the front-end electronics
- Providing a link for electric signals between the VFATs and the OptoHybrid

The GEB is split into two parts, one for the narrow side of the detector and one for the wide side of the detector. The two sides of the GEB are visible as the red surface in Fig. 6.4. Each side covers four η-partitions and they are connected via LV cables. The GEB is secured using screws to the external frame and grounded via copper tongues

attached to the external frame. Throughout the GEB are 12-pin connectors for the FEASTs, 100-pin Panasonic connectors for the VFATs, and SAMTEC connectors for the OptoHybrid. Two different GEBs had to be developed for the Long and Short chambers [25, 26].

6.2.1.2 VFAT3

The VFAT3 is the front-end chip for the GE1/1 detector. This chip evolved from the specialized VFAT2 chip used for GEM detectors inside the TOTEM experiment, with improved time resolution. The VFAT3 has 128 readout channels; each channel corresponds to a strip in the GE1/1 detector and has an independent data path (i.e. full granularity). The resulting circuit is shown in Fig. 6.5 [90, 28].

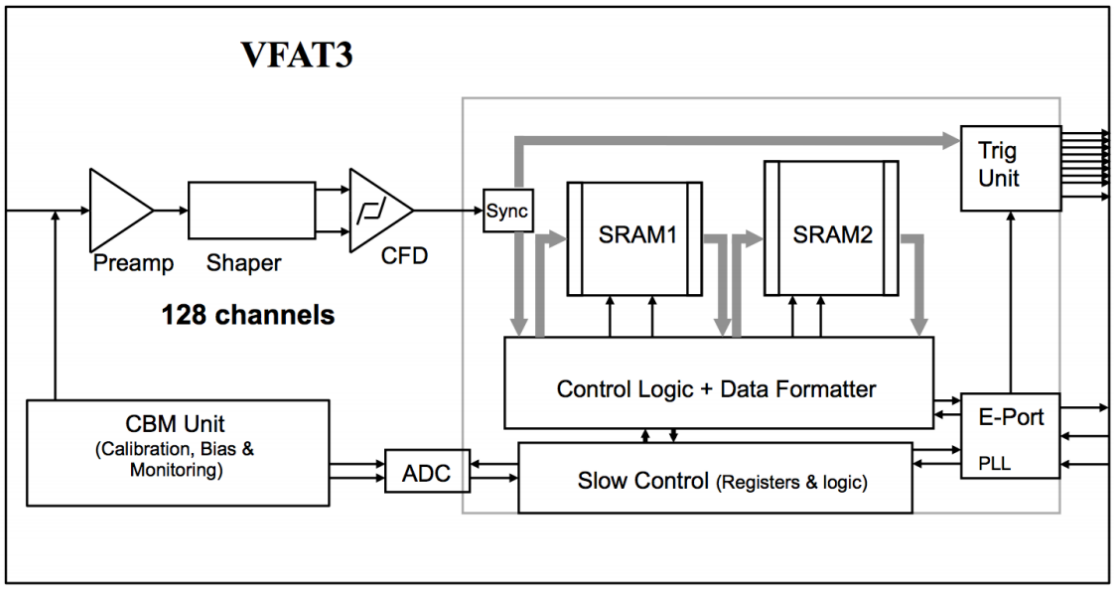

Figure 6.5: The VFAT3 Circuit consisting of a full granularity data path with 128 readout channels. Figure from [26].

The VFAT3 chip is placed on a specialized board called a VFAT3 hybrid to connect to the GE1/1 readout board (the readout strips) and the GEB. A 130-pin Panasonic connector connects the hybrid to the readout board, this consists of two pins for grounding and one pin per readout strip (128 strips per VFAT). A 100-pin Panasonic connector connects the hybrid to the GEB. These connections are shown in Fig. 6.6.

Signals from the strips are transmitted to the VFAT3 chip; following the pre-amp

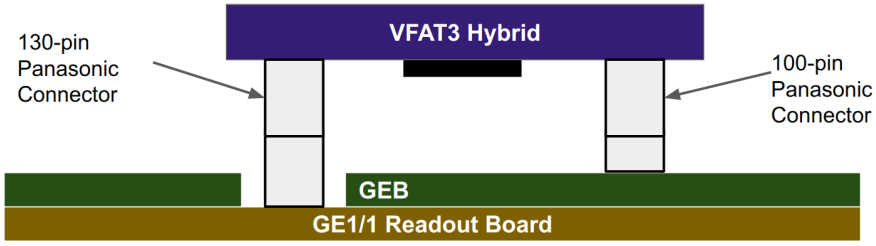

Figure 6.6: The connection of the VFAT3 hybrid to the GEB and GE1/1 readout board.

and shaper of each readout channel is a constant fraction discriminator (CFD). This is a circuit that uses the CFD technique to improve time resolution by reducing time-walk. Previously, the time of a signal was defined as the time that the leading edge of a signal charge pulse passed a threshold value. Since the signals vary in amplitude, the time a signal crosses the threshold can be delayed. This phenomenon is known as "time-walk" and the CFD technique reduces the effects of time-walk by breaking a signal into two components, inverting one signal and delaying the other, then combining them together to create a zero crossing. Using this method the time of a signal is determined based on a *constant fraction* of the pulse height. The resulting improvement is shown in Fig. 6.7 and the VFAT3 implementation of the technique is shown in Fig. 6.8 [28, 90].

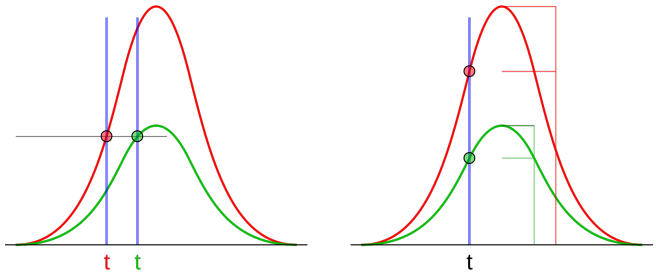

Figure 6.7: (Left) Two signals that occur at the same time pass the threshold at different times resulting in a measured time discrepancy. This phenomenon is known as time-walk. (Right) A constant fraction of the peaks occurs at the same time. By applying the CFD technique the zero crossing will occur at the same time (i.e. this constant fraction shown) for these two signals. Figure from the public domain.

The resulting data path inside the VFAT3 is shown in more detail in Fig. 6.9. Each readout channel is connected to a circuit that is split into two parts. One side consists of a shaper and a zero crossing comparator (ZCC) to create the CFD function. The other side

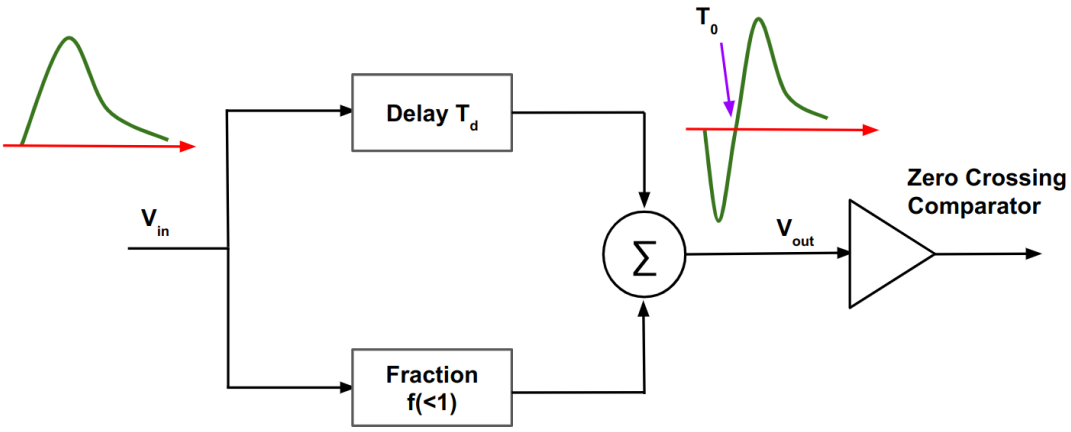

Figure 6.8: The CFD circuit in the VFAT3. Adapted from [28].

consists of an arming comparator (ARM) with a fixed threshold. The results of the two paths are combined. The separation of the CFD function and the ARM comparator also allows the VFAT3 to be operated in a "leading edge" mode where the circuit is configured to determine time based on when the leading edge of a signal passes a fixed threshold.

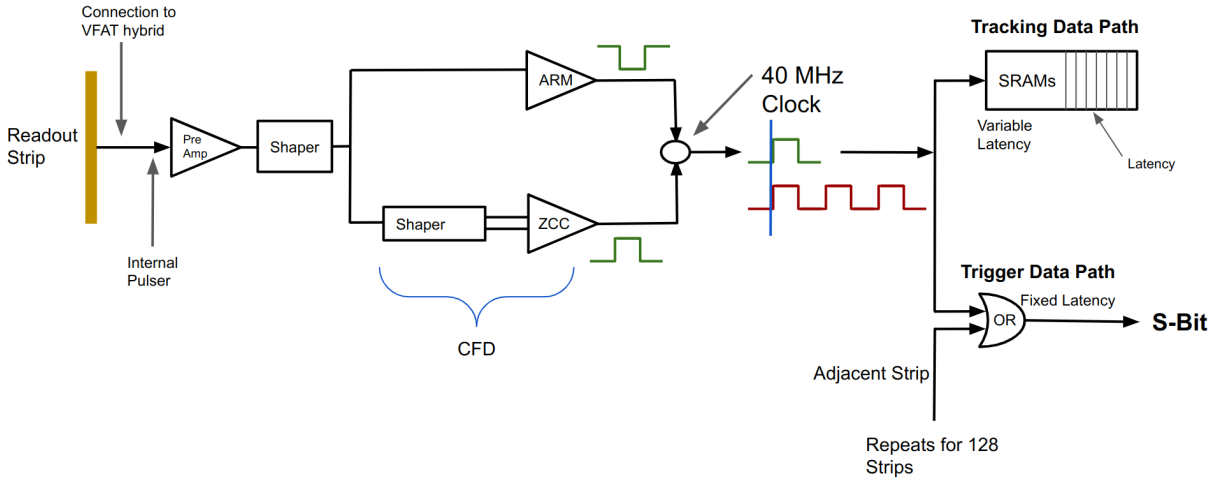

Figure 6.9: The data path inside of the VFAT3 for a single readout strip. There are 128 of these paths, one for each data channel in the VFAT.

After the two components are combined, the signal becomes a square pulse. This is sent to a 40 MHz clock[4] where the signal is aligned relative to the clock. The circuit is then split again. This time into a "tracking" data path and a "trigger" data path.

[4]Collisions at the LHC occur every 25 ns, i.e. bunch crossings at 40 MHz.

The trigger data path has fixed latency and is capable of operating in two modes: single data rate (SDR) for half granularity[5] and double data rate (DDR) for full granularity. The standard operational mode for GE1/1 is SDR. In this case, the trigger path of one strip passes through an OR gate with the trigger path from an adjacent strip. The result is called an "S-bit". If one (or both) of these channels have a signal that passed the threshold, then the S-bit becomes 1. 8 S-bits are combined into a byte and sent off of the VFAT via a differential pair called an S-bit line. Each VFAT has 8 parallel S-bit lines[6]. Figure 6.10 shows a detailed view of the GE1/1 VFAT3 hybrid with a zoomed in region highlighting the differential pairs which compose the S-bit lines.

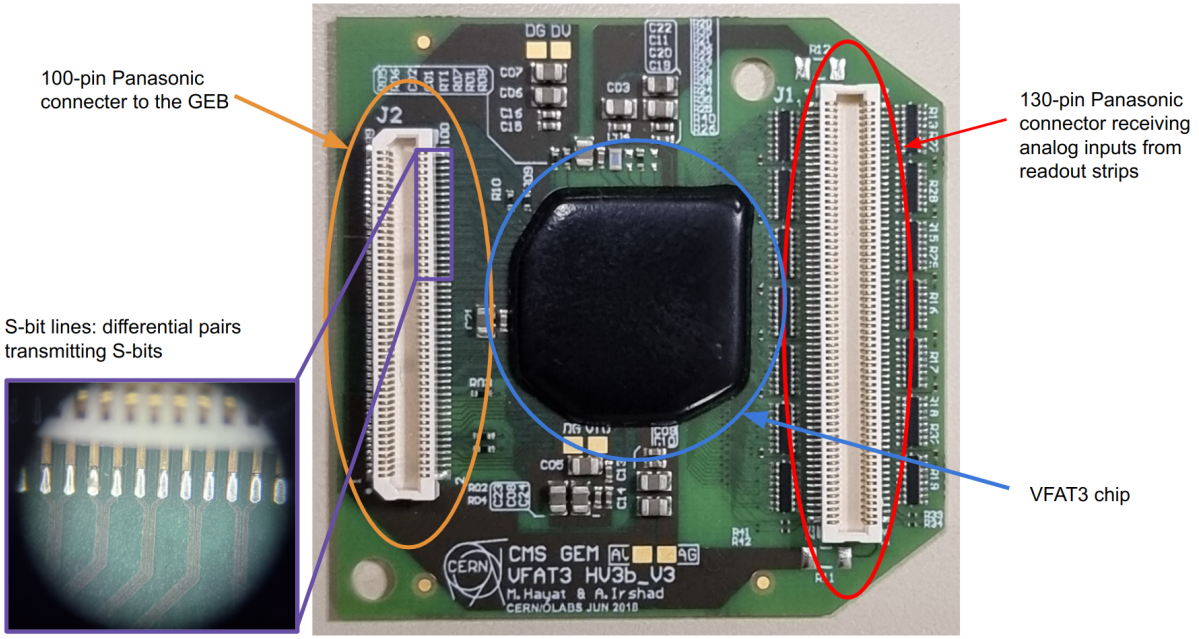

Figure 6.10: The GE1/1 VFAT3 hybrid highlighting the Panasonic connectors and a zoomed in region showing the differential pairs that compose the S-bit lines.

The tracking data path has full granularity and variable latency. Data in the tracking path is sent to a buffer called an SRAM (Static Random Access Memory). If the VFAT receives a Level 1 accept, then tracking information is located in the buffer based on the recorded latency and sent from the VFAT to the OptoHybrid.

[5]Half granularity meaning half of the number of readout channels.

[6]8 S-bit lines, 8 S-bits per S-bit line, 2 data channels per S-bit. Thus, the results of 16 strips are sent per S-bit line.

The components in the VFAT data path (pre-amp, shapers, comparators) have configurable values that can be adjusted for optimal performance.

6.2.1.3 OptoHybrid

The OptoHybrid sits at the center of each GE1/1 chamber and is responsible for communication between the chamber and the outside world.

This is accomplished in the data path with three gigabit transceivers (GBTs), each taking care of data from a region of the detector. Each GBT is connected to a versatile link (VTRx) capable of sending and receiving optical signals. The first GBT receives data from six VFATs while the other two GBTs are receive data from nine VFATs. This enables slow control commands to be sent along the first GBT in order to configure all of the detector [25, 26].

Each OptoHybrid is equipped with a Xilinx Vertex 6 FPGA which processes trigger hits and controls two transmitter versatile links (VTTx) to send trigger information from GE1/1 to the nearby CSC chambers and to the GEM back-end electronics [25, 26].

The OptoHybrid receives data from the tracking and trigger paths of the VFATs. The GBTs send optical signals via the VTRxs to transmit the tracking data to the back-end electronics in the service cavern. The trigger data is elaborated in the FPGA and sent through the VTTx. From the back-end electronics, the OptoHybrid receives triggers, clock, and slow control commands.

6.2.2 Back-end Electronics

The goal of the back-end electronic system is to interface the GEM detectors with CMS DAQ (Data-Acquisition), TTC (Trigger Timing and Control), and the trigger. This is accomplished using multi-Gbit/s links that adhere to a standard approach from the telecommunication industry called μTCA. This allows for high throughput (2 Tbit/s) and high availability (the probability of being interrupted is $\approx 10^{-5}$). μTCA is the standard for all CMS phase I upgrades [25, 26].

The μTCA crate houses two slots for μTCA carrier hub (MCH) modules and 12 slots for advanced mezzanine cards (AMCs). The first MCH slot is occupied by a commercial MCH module which provides gigabit Ethernet communication and control. The second

MCH slot houses the AMC13 module developed by Boston University. It is the CMS standard module for providing an interface between the μTCA crate to CMS DAQ and TTC. The AMC used by GE1/1 is the calorimeter trigger processor card (CTP7) developed by the University of Wisconsin. Each μTCA create contains 6 CTP7s each one capable of interfacing with 12 GE1/1 chambers [25, 26].

Together, the chambers, front-end electronics, and back-end electronics compose the GE1/1 system. The electronics undergo heavy usage in the experiment, but before being installed in CMS all of these components are carefully tested with a full CMS-like readout train during production.

6.3 GE1/1 Production

The GE1/1 detectors were tested in 8 steps of quality controls (QC) while being assembled. Each step of QC was designed to ensure optimal construction. The process of building a GEM chamber starts with sending the materials to production sites and the full process is outlined below [26, 91].

Figure 6.11: The author (left) participating in the assembly of a GE1/1 chamber in a clean room.

This process starts with chamber materials sent to production sites around the world.

- **QC 1:** *Foil and Components Inspection* - This consists of a visual inspection of the GEM foils, careful measurements of the frames, and preliminary tests of the readout PCB to ensure no damage before assembly.

- **QC 2 'Fast':** *First HV test on GEM foils* - A potential difference of 550 V is applied across the foils and the output resistance is carefully monitored to ensure no defects or contaminants.

- **Assembly**

- **QC 2 'Long':** *Second HV test on GEM foils* - The GEM foils are placed in a nitrogen[7] (N_2) chamber where a potential difference of 600 V is applied across the foils and the discharge rates are carefully monitored until the foils have a discharge rate of 3 discharges per 5 hours.

- **QC 3:** *Gas leak test* - After assembly, the chamber is tested at high pressures and the leak rate is carefully monitored to ensure that the leak rate will remain less that 1% of the incoming flow rate.

- **QC 4:** *HV stability (divider)* - A voltage divider is connected to the electrodes for the top and bottom of each GEM foil to provide HV at constant proportion to each electrode. Then the chamber is tested to ensure that a high voltage can be continuously maintained. This is done by increasing the voltage to the divider in steps of 100 V while the chamber is flushed with CO_2.

- **QC 5:** *Gain and response uniformity* - The chamber uses a gas mixture of Ar/CO_2 70/30% and is placed inside of a copper box with an x-ray gun in order to measure the gain and ensure that the response is uniform throughout the detector.

After assembly, the completed chambers are sent back to CERN.

- **QC 6:** *HV stability (multichannel)* - The chamber is again flushed with CO_2 and tested at HV, this time with a multichannel system capable of applying different voltages to each foil. This is done to further test and improve the chamber stability.

[7]Nitrogen is a non-amplifying gas with a large breakdown voltage, allowing for high voltages to be applied without concern of damaging the foils.

- **QC 7:** *On-chamber electronics test* - The chamber is equipped with front-end electronics and detailed measurements are taken to ensure proper operation of the electronics.

- **QC 8:** *Cosmic stand measurements* - The single layer GE1/1 detectors are coupled into a GE1/1 super chamber, and placed on a cosmic ray stand. Here the efficiency of each chamber is measured with cosmic rays. Each chamber must have an efficiency above 95% before being fully qualified for use in the CMS experiment.

In terms of production, this thesis work is primarily focused on one step, the QC 7 on-chamber electronics test. The work on QC 7 was used as a basis for early GE1/1 commissioning and to design the "trolley" test. For details on QC steps 1 to 5 see reference [91] and for details on QC 8 see reference [25].

6.3.1 QC 7: On-chamber Electronics Test

The chamber is equipped with front-end electronics after passing QC 6. In QC 7, the front-end electronics are extensively tested. My work during production focused on performing the QC 7 test as part of a dedicated team.

The QC 7 stand consists of a full CMS-like readout train with a μTCA crate equipped with an AMC13, an MCH, and a CTP7. The tests in QC 7 can be summarized as follows:

1. Test communication between back-end and front-end electronics.

2. Calibrate the electronics and identify optimal working points for the VFAT circuits (pre-amp, shapers, and comparators).

3. Identify bad connections to the VFAT (digital and analog sides) while measuring the intrinsic noise in the electronics.

These tests are administered first using fan cooling with heat sinks placed on the sensitive electronics and then administered again after installation of the cooper cooling plate and aluminium chimney. A chamber on a QC 7 stand with fan cooling is shown in Fig. 6.12

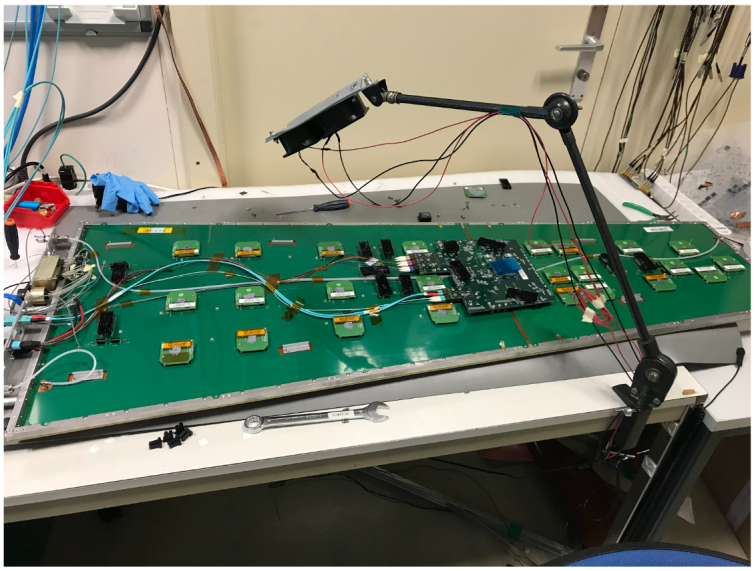

Figure 6.12: A GE1/1 Long chamber on a QC 7 stand with fan cooling.

First, communication is established with the VFATs by determining the optimal phase for the GBTs. Then, a number of calibration scans take place in the VFATs to ensure optimal performance. After properly calibrating the electronics, the other tests can proceed. Errors that arise during either of these steps correspond to a hardware issue and the affected VFAT(s) or OptoHybrid are replaced. QC 7 then continues with testing the analog and digital sides of the VFAT.

VFAT Analog Side Test

As detailed above, the VFAT3 is equipped with a full granularity data path with 128 readout channels. This means that each readout channel corresponds to one strip inside of the GE1/1 chamber as shown in Fig. 6.9. In order to test the health of each channel, the VFAT is capable of delivering an internal calibration pulse. The test pulse is delivered at the beginning of the circuit—mimicking a signal from the readout strip. To test the analog side of the VFAT, each channel is pulsed with 100 pulses at a fixed charge while the charge is increased and the number of recorded pulses are measured. This creates an "S" shaped curve for each of the 128 channels and is shown in Fig. 6.13.

The width of this "S" shaped curve corresponds to intrinsic noise[8]. During QC 7,

[8]The noise is due to the electronics and the detector.

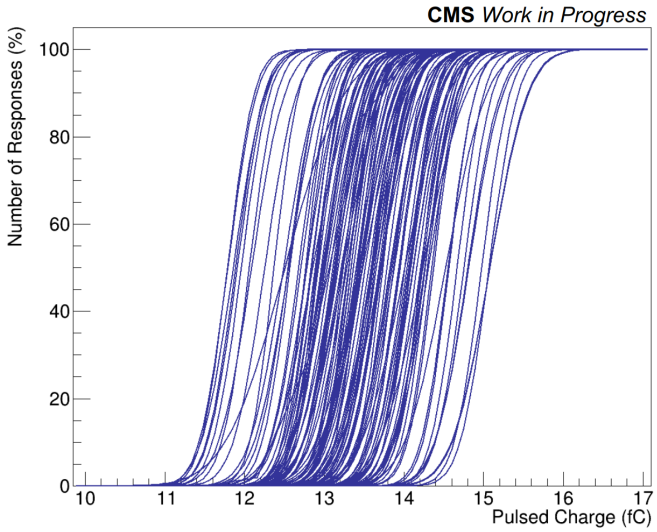

Figure 6.13: For each of the 128 data channels 100 pulses are administered to the VFAT while the pulse charge is increased. The number of pulses recorded is measured, creating an "S" shaped curve for each channel.

this noise is monitored and can be significant if there was an issue during electronics installation. The most common source of noise during QC 7 is the cooling plate. The cooling plate is secured to the frame of the detector where the grounding is located. The aluminium chimney above the chamber creates a Faraday cage and the cooling plate acts like an antenna. If the grounding is electrically connected to the cooling plate, then the cooling plate will bring in outside noise. To mitigate this, the cooling plate must be carefully insulated from the frame using Kapton tape and tested to ensure that it is not electrically connected to the chamber grounding. The resulting Kapton protection is shown in Fig. 6.14.

The channels are monitored in greater detail by measuring the intrinsic noise without delivering an internal pulse. The VFAT3 is equipped with an arming comparator (Fig. 6.9) that converts analog to digital signals by creating a square pulse when a signal exceeds a threshold charge. The user can increase the threshold charge value to discriminate data from electronic noise. During this test the response due to noise is sampled 100 times for each channel as the threshold is increased. An example of this test in shown in Fig. 6.15. In this image, a number of channels show activity due to noise. This is noise due to capacitance of the detector propagating through the connection of the data channels to

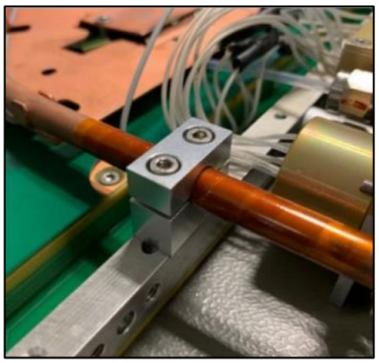

Figure 6.14: Kapton tape wrapped around a cooling pipe to insulate the copper cooling plate from the GE1/1 grounding.

the readout. Channels that show no noise have become disconnected from the readout [25]. In this case, the analog of the VFAT has become disconnected and must be carefully plugged in again as shown in Fig. 6.16.

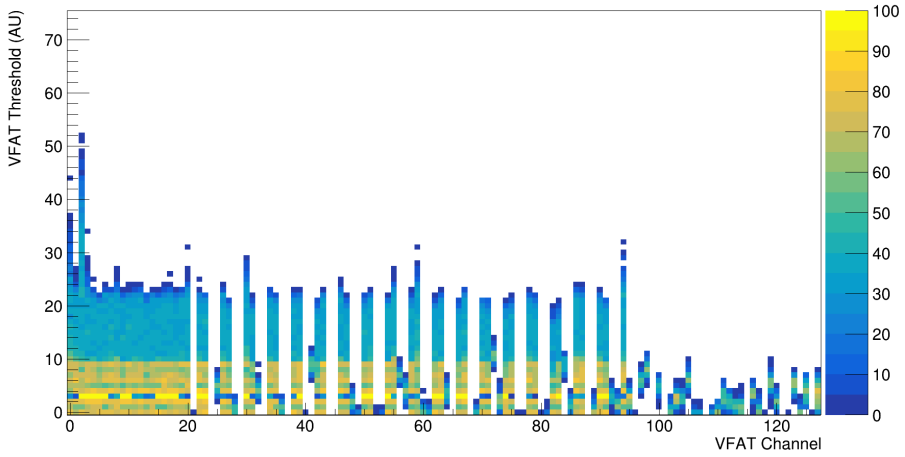

Figure 6.15: The noise on the left side is the intrinsic noise of the detector (capacitance noise) and electronics, on the right hand side only the electronic noise is visible. This indicates that the analog side of this VFAT has become disconnected.

VFAT Digital Side Test

An important component of the digital side of the VFAT are the 8 parallel S-bit lines where trigger data are transmitted. Intrinsic noise in the electronics will cause a high trigger hit rate and again the effects of noise are mitigated by applying a threshold charge in the arming comparator. During the trigger test the threshold value is increased while

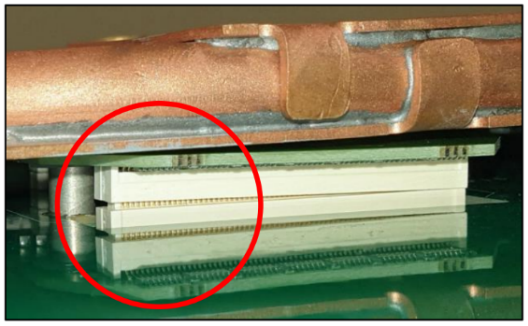

Figure 6.16: The analog side of the VFAT has become disconnected from the readout board.

measuring the trigger hit rate to measure the health of the trigger path and calibrate the arming comparator against electronic noise. This test is shown in Fig. 6.17, where the blue line shows the expected behavior of a healthy trigger path and the flat red line indicates an issue in the trigger path.

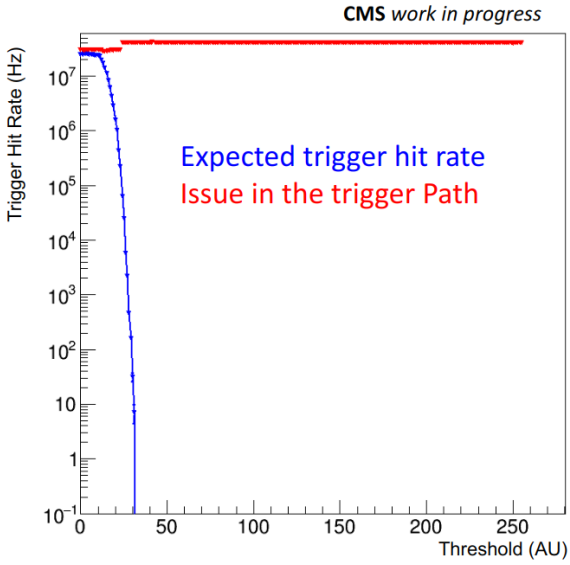

Figure 6.17: The trigger hit rate is measured in Hz as the threshold value is increased. A healthy response is shown on the blue curve, whereas the flat red line indicates there is an issue in the trigger path.

The red line in Fig. 6.17 shows that the noise never decreases as the threshold is increased; at low threshold values the observed line is the intrinsic electronic noise and at higher threshold values becomes a saturated flat line due to the clock (which injects low charge at 40 MHz. On the VFAT3 hybrid (Fig. 6.10) the S-bit lines are on the side of

Panasonic connector away from a securing pillar which fixes the VFAT connection to the GEB. Thus, the observed noise indicates a disconnection in the Trigger path, often caused by an improperly connected VFAT or, less frequently, a damaged pin on the OptoHybrid (Fig. 6.18). Trigger path issues frequently emerge after installation of the cooling plate. This rigid plate can apply pressure in undesirable locations causing bad connections. The experience in QC 7 prompted design changes for the VFAT3 hybrid and cooling plates that will be used for GE2/1 and ME0, where both the hybrid and cooling will have enhanced flexibility.

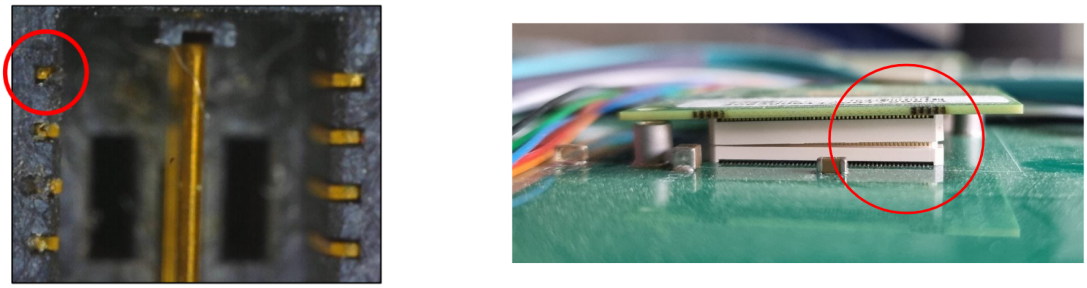

Figure 6.18: (Left) a damaged pin on the SAMTEC connector of the OptoHybrid or (Right) a disconnection of the digital side of the VFAT are frequent causes of issues in the trigger path.

After successfully passing these tests before and after installation of the cooling plate, two GE1/1 chambers are coupled into a super chamber and installed into the QC 8 cosmic stand for efficiency evaluation. Once all 8 steps of quality controls are passed, then the GE1/1 chambers are loaded into a mobile trolley, ready for installation into the CMS experiment.

6.4 GE1/1 Installation

After the GE1/1 chambers have passed all the stages of quality control, they are transported to the CMS experiment for installation. This means transporting the chambers from CERN Prevessin site to Point 5; a 10 km journey across French farm roads (Fig. 6.19). Then, the chambers are lowered 100 m underground. The transportation is done around 3 am local time so that the chambers can be transported at low speeds across these bumpy roads.

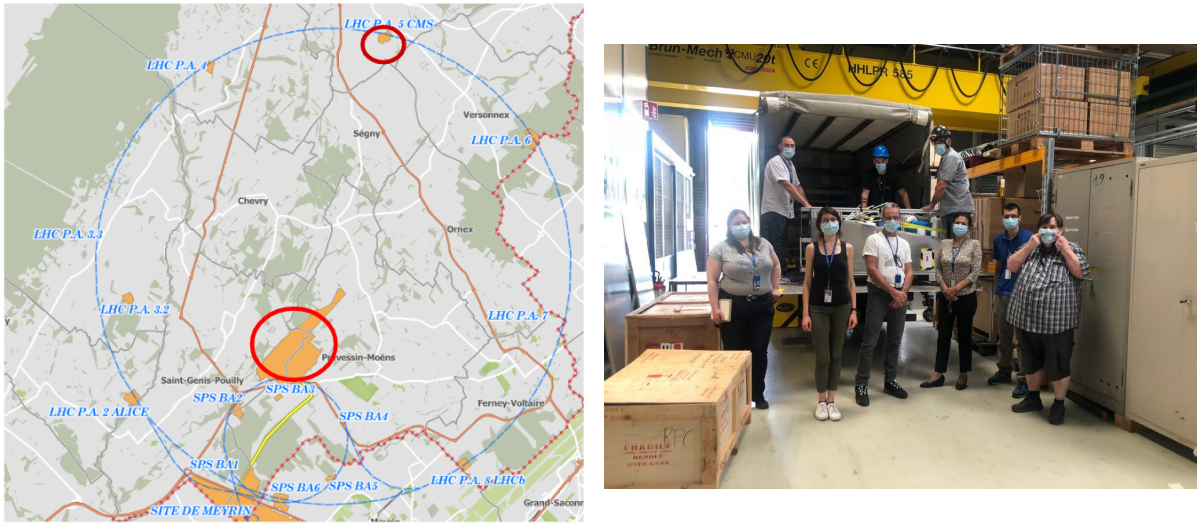

Figure 6.19: (Left) The blue ring is the outline of the LHC, the small red ring is the location of the CMS experiment, and the larger red ring is the location of the Prevessin cite. The GE1/1 super chambers are transported from Prevessin to the CMS experiment by truck. (Right) The GEM team loading the truck with a trolley of chambers the afternoon before an early morning departure.

6.4.1 The "Trolley Test"

During the installation in the first endcap, several chambers that were installed had broken trigger lines. These were due to disconnections on the digital side of the VFATs, which most likely arose during the transportation of the chambers to CMS. Therefore, it was clear to me that a new test was needed. I designed this test to take place on the cavern floor after the chambers had been transported.

The cavern floor is not an ideal location to perform a sensitive electronics test like QC 7, so a new one was designed for this unique environment. This test required a connection to the back-end electronics and LV power. A μTCA crate cannot be installed in the CMS experimental cavern, so a "flying" fiber optic cable was routed from a patch panel to connect the chambers to the GE1/1 back-end located in a separate service cavern. The fiber optic cables were handed carefully down to the cavern floor, see Fig. 6.20. The LV was supplied via a bench top power supply. In the experimental cavern, there is no access to cooling, so this test needed to take place quickly ($\approx$ 20 minutes).

Given the time constraints, a faster version of the trigger path test was needed, so a low

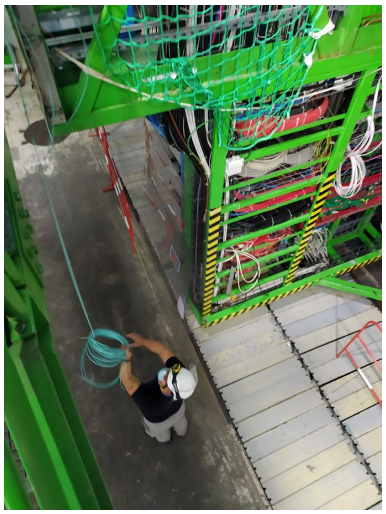

Figure 6.20: Setting up a "flying" fiber optic cable in the CMS cavern.

level tool was developed to quickly scan the trigger path for any possible disconnections. This short time also meant that a proper scan for optimal VFAT calibration values for the electronics cannot be performed. So a set of calibration values was created based on the recommendations by the VFAT3 developers [28, 90]. These DAC values are able to bring the electronics into an operational regime, but not able to calibrate them for optimal use. The resulting "trolley" test is shown in Fig. 6.21 and was performed as follows:

1. Establish a connection to the chamber and scan the GBT phases.

2. Configure the detector by setting the VFAT electronic calibration values to the recommended values.

3. Scan the noise in the tracking path to look for any disconnections from the readout. This test is the same as the one used in QC 7 and an example can be seen in the QC 7 section in Fig. 6.15.

4. Configure the detector.

5. Perform a low level S-bit test to look for any disconnections in the trigger path. This test had to be redesigned from QC 7 in order to be fast and the new test is described in detail below.

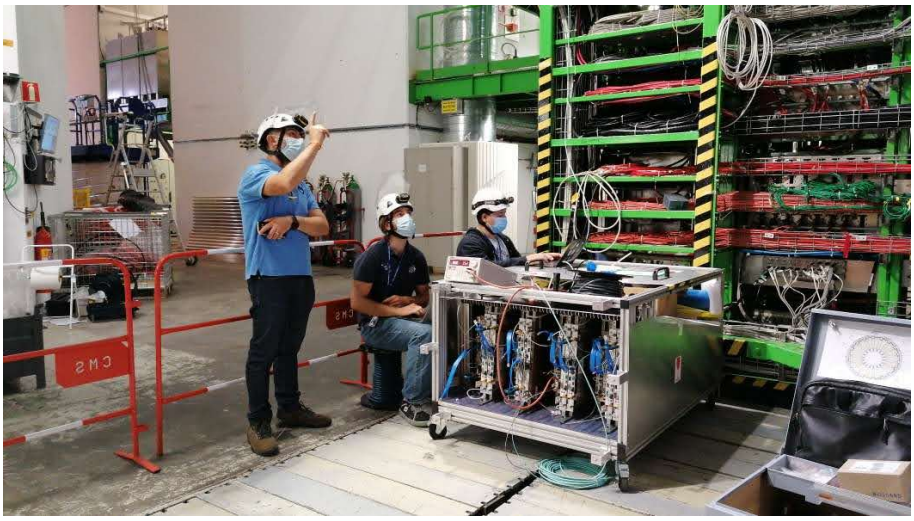

Figure 6.21: The setup of the "Trolley" Test inside the CMS cavern. The trolley contains the GE1/1 chambers; on top of the trolley is a bench top power supply to provide LV for the test. In the photo, Michele Bianco (left) is pointing out the GEM installation to the author (center) and Johny Jaramillo (right) while they trolley test is being performed.

The new S-bit test was designed to be fast. Rather than scanning over different threshold values, the global threshold is set to a high value (100 ADC units) to ensure that noise will not contaminate the trigger path. Then one data channel for each S-bit is pulsed and monitored. When one channel is pulsed, the expected result is one observed S-bit. Each S-bit is an OR of two data channels and 8 S-bits travel on an S-bit line. Thus, if one S-bit line is disconnected then 9 S-bits are observed—the correct S-bit and the 8 S-bits corresponding to the broken S-bit line. A visual example of this test is shown in Fig. 6.22.

This test proved to be crucial for the successful installation of GE1/1. The test identified 5 chambers with issues. These chambers were able to be transported back to the GEM lab and repaired before being installed in the experiment.

Now 144 GE1/1 chambers have been installed in the CMS experiment and will be operated during run 3 (and beyond) of the LHC.

6.5 GE1/1 Commissioning

Installation of the GE1/1 chambers into the CMS experiment was a major milestone for the GEM project, but there is a lot that must me done before GE1/1 can be operated

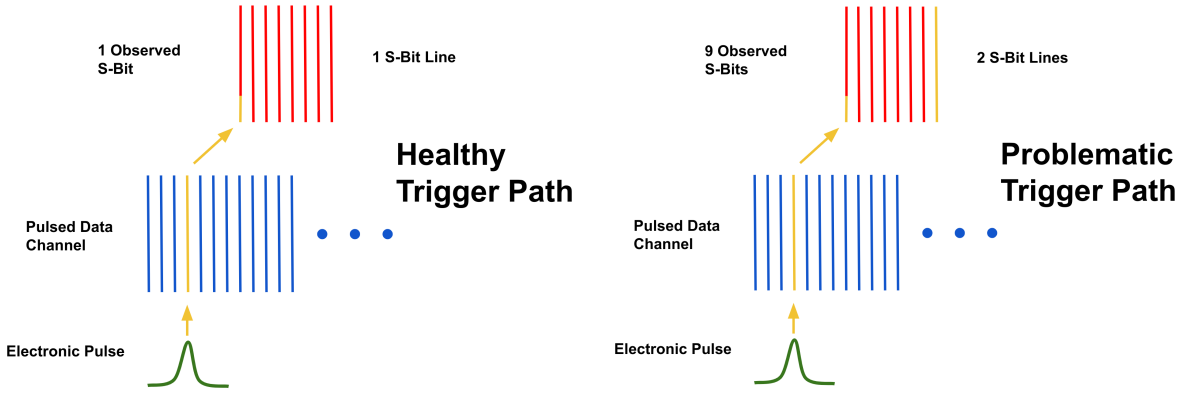

Figure 6.22: (Left) A healthy trigger path with one S-bit for one pulsed channel. (Right) A problematic trigger path with 9 S-bits for one pulsed channel (each S-bit line provides a path for 8 S-bits); this indicates a disconnection in the trigger path.

in the CMS experiment. The process of preparing the GEM chambers for collisions after the installation is called "commissioning". In 2021, I served as GE1/1 commissioning co-coordinator and worked with teams of experts to prepare GE1/1 for operation in LHC run 3.

Commissioning begins with a hybrid stage where the services (optical fibers, cooling, LV and HV cables) are attached and tested. Then this progresses to an early stage of chamber commissioning which tests the connections, electronics, and mappings for individual chambers. This is the first time the chambers have been operated in the CMS experiment, so they need to be studied in this new environment.

After completing these early steps, the chambers need to be aligned in 4D (space and time) with the other detectors at CMS. Then, GE1/1 needs to be added to the CMS level 1 trigger via the Endcap Muon Trackfinder. Finally, the efficiency of detecting muons needs to be measured. These steps are done using cosmic ray muons.

6.5.1 Services Commissioning

While the chambers were being installed into the CMS experiment, all of the services (optical, gas, cooling, LV, HV) had to be properly attached and tested. One of the crucial services are the fiber optic cables attached to each chamber, which must be carefully connected to a patch panel on the side of the experiment. Here, the cables must be carefully coiled (according to the fiber bending radius) and properly plugged in. The

patch panel reduces the number of cables that travel outside of the experimental cavern. The fiber optic cables then travel to the service cavern where they are connected to the back-end electronics. An example of coiling and connecting fibers in the experimental cavern is shown in Fig. 6.23. After installing and powering the GE1/1 chambers, the optical power received by the GEM back-end is carefully monitored after plugging in cables to each chamber. The observed power should be between 300 and 500 µW for each VTRx and VTTx. This ensures that no optical cables were damaged during installation. The proper connection and routing of all the services is then tested in greater detail during the chamber commissioning.

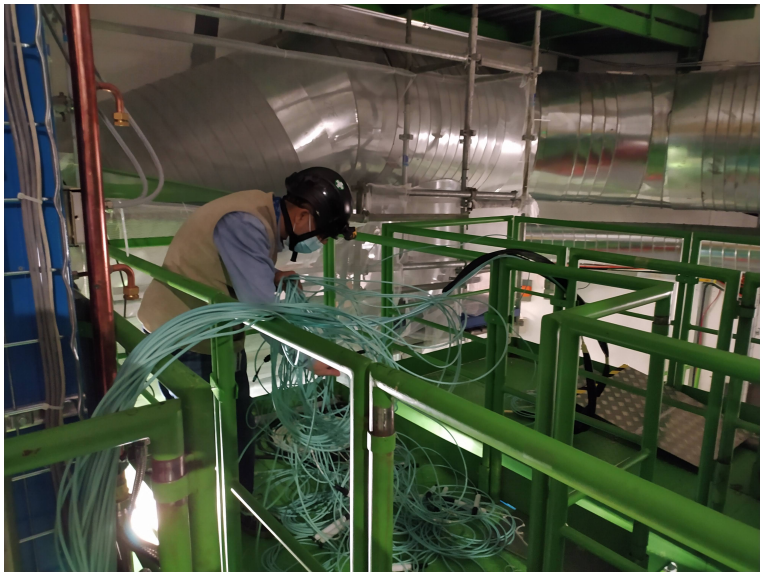

Figure 6.23: The green cables contain optical fibers coming from the GE1/1 detectors. At this point, the fibers are on top of the CMS experiment and here they are being carefully organized and coiled before being attached to a patch panel.

6.5.2 Chamber Commissioning

The next commissioning phase began as soon as the chambers were installed in CMS. The first step was to power the chamber LV using the DCS and make sure that the LV on the DCS correctly maps to the chamber on the experiment and in the DAQ system. This confirms that the fiber optic and LV cables have been correctly installed.

The next step was to redo the steps of QC 7 on every chamber. This is done to check for any problems that may have arisen after installation. This is also the first measurement

of noise in the CMS cavern and is used to calibrate the front-end electronics.

After performing these tests and resolving any mapping or electronics issues, the HV system was operated. The GEM foils had to be turned on in steps to carefully eliminate any pollutants that may have entered into the chambers. This is known as "HV training". The GEM HV training was carefully studied in QC 8 and is outlined completely in [25]. The HV training has to be adapted based on time allowed by CMS operations and based on the gas mixture[9]. The normal steps for HV training in Ar/CO_2 (70/30%) are listed below:

1. Increase the voltage of each foil to $\Delta V = 420\,\text{V}$ and maintain the voltage for 4 hours.

2. Power the electrodes in the following way and maintain the voltage for 4 hours: Drift 810 V, G1Top 403 V, G1Bottom 315 V, G2Top 396 V, G2Bottom 630 V, G3Top OFF, G3Bottom OFF.

3. Power the electrodes in the following way and maintain the voltage for 4 hours: Drift 799 V, G1Top 398 V, G1Bottom 311 V, G2Top 391 V, G2Bottom 621 V, G3Top 100 V, G3Bottom OFF.

4. Maintain the voltage settings and increase G3Top to 186 V and maintain for 4 hours.

5. Maintain the voltage settings and increase G3Top to 373 V and maintain for 4 hours.

6. Power the chamber to the nominal value for the desired working point.

Once the HV can be operated, the HV mapping must be tested. This was done by setting the global threshold to a common high value and checking for occupancy in the S-bit registers in the DAQ system before and after powering the HV. This was done one chamber at a time. The chamber in the DAQ system should match the chamber powered in HV, if not, there is a mapping issue that needs to be resolved.

[9]Sometimes HV training is done in pure CO_2 in order to operate the foils at higher voltages for more effective cleaning of pollutants.

During the early commissioning of GE1/1, two major issues were discovered. The first was high noise in the CMS experiment. The second was a communication instability.

6.5.3 Noise Mitigation at CMS

As GE1/1 was being installed and tested, it was discovered that the detectors were experiencing significantly higher noise in the LV electronics compared to the laboratory. This prompted a detailed noise investigation into the possible cause. During this process, it was observed that the noise was reduced by operating the chambers with a bench top power supply.

The noise occurs at high frequencies and, while the source of noise is still not fully understood, the solution was to develop a low pass filter. Many iterations of chamber side low pass filters were designed and tested in the CMS cavern and the finalized design is shown in Fig. 6.24.

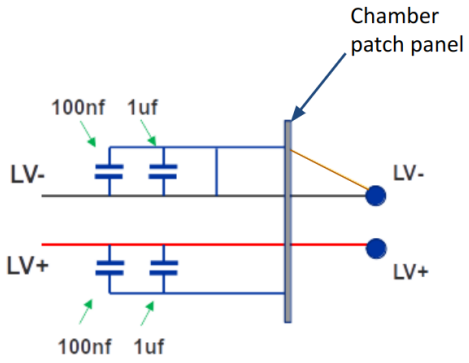

Figure 6.24: The chamber-side low pass filter designed to mitigate noise on the GE1/1 chambers.

After finalizing the design, a low pass filter needed to be installed to each of the 144 chambers. This endeavor required modified screw drivers to install the filters into tight locations made difficult to access due to support beams. Once installed, the noise in each chamber was remeasured. The resulting noise reduction in one of the endcaps is shown in Fig. 6.25. An operational amount of noise in the LV system is approximately $< 1\,\mathrm{fC}$. This endeavor brought almost all of the GE1/1 chambers into the operational noise region.

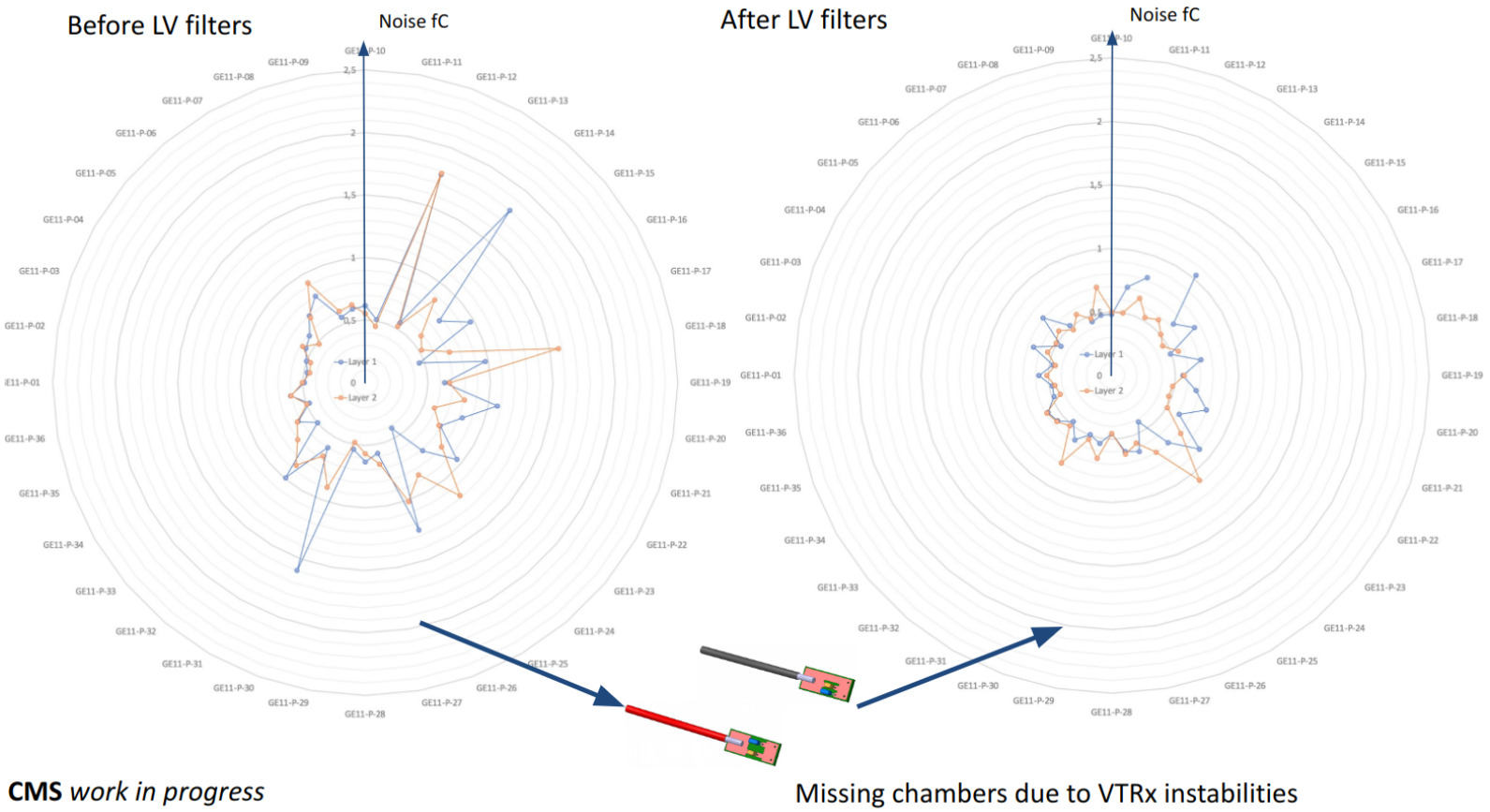

Figure 6.25: The noise (in fC) for each chamber in one of the CMS endcaps. The charge value increases radially outward and each label around the circle represents the position of a GE1/1 super chamber. The blue line refers to layer 1 (the chamber nearest the interaction point) and the orange line refers to layer 2 (the chamber further away from the interaction point).

6.5.4 VTRx Issue and Solution

As detailed earlier in the chapter, a key component of the GE1/1 electronics is the versatile link (VTRx) for transmitting and receiving optical signals from the chamber to the back-end electronics. The VTRx receives optical signals via a photodiode and transmits signals via a laser (Fig. 6.26) [29].

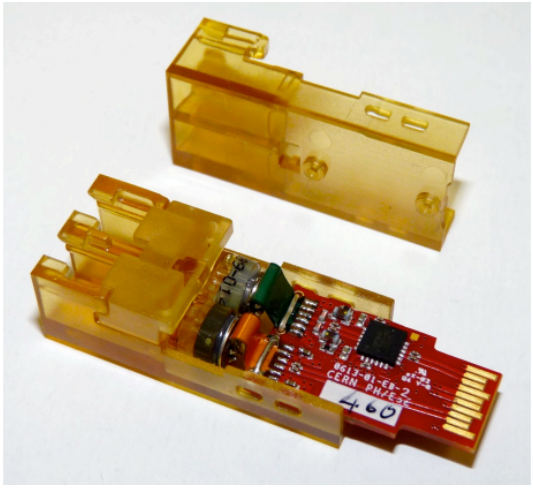

Figure 6.26: The Versatile Link (VTRx) developed by the CERN electronics group; the green flexwire is connected to the laser and the orange flexwire is connected to the photodiode. Figure from [29].

During the commissioning phase of GE1/1, there were disruptions in communication between the back-end electronics and the chamber-side electronics. At the same time, a large investigation was taking place by the CERN electronics group into reported issues with the VTRx. They discovered that epoxy glue placed on the photodiode "outgasses" releasing particulates that can deposit on the optical fiber and interrupt communication (Fig. 6.27). The VTRx is a component used by many detector subsystems at CERN, creating an urgent need for a solution to this problem.

The CERN electronics developed a procedure for baking the VTRxs at 80 °C for 1000 hours. Simultaneously, the HCAL subsystem at CMS discovered that the build up of particulates can be prevented with a proper temperature gradient between the optical fiber and the photodiode. Alas, the detrimental build up of particulates can be prevented by cooling the photodiode. This discovery allowed HCAL and other subsystems to make last minute modifications to the cooling before Run 3. However, at this point GE1/1 was fully

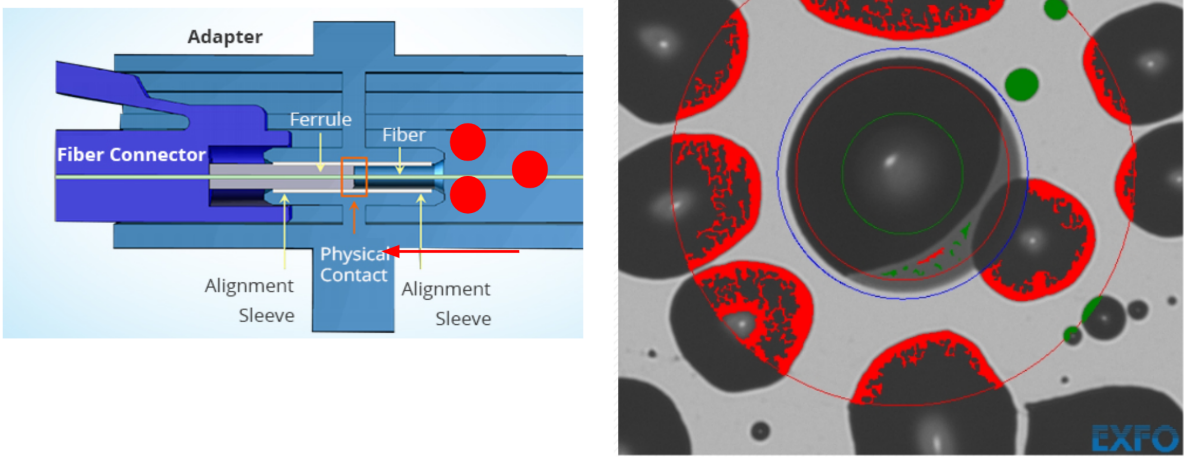

Figure 6.27: (Left) shows the process of particulates building up on the tip of the optical fiber inside of a connector (fiber diagram adapted from [30]) and (Right) shows how the tip of a GE1/1 fiber is affected by outgassing under a microscope. The black circles are the particulates which accumulate on the fiber causing communication interruptions.

installed, so for Run 3 the GE1/1 system will have to deal with periodic communication interruptions which seem to affect 5% of the VTRxs.

There are 3 VTRxs per chamber each responsible for communication with a Gigabit Transceiver (GBTx), so different regions of the GE1/1 chambers are affected during a communication loss.

The GE1/1 detectors will remain inaccessible for the remainder of LHC Run 3, so a solution to the this problem will be implemented during the Long Shutdown 3. The first step of the solution will be to replace all of the VTRxs with VTRxs that have been baked by the CERN electronics group. The second step is to apply cooling to the VTRx.

A prototype of this new VTRx cooling was designed and tested in the GEM lab at CERN. I installed and tested this prototype while helping to finalize the installation procedure. The cooling is placed above VFAT #12 and extends underneath the optohybrid. A prototype is shown in Fig. 6.28. The cooling is placed in thermal contact with the VFATs using thermal pads recommended by the HCAL group.

In an early version of the prototype, foam padding was placed underneath the VTRx

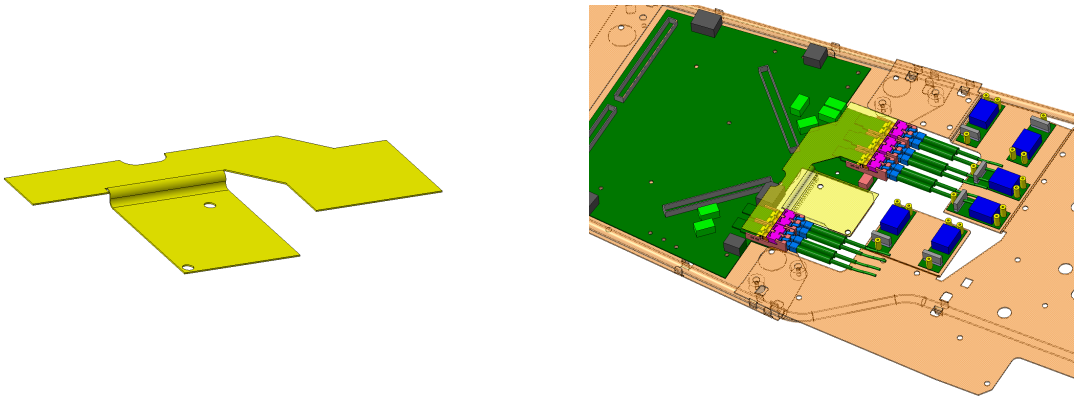

Figure 6.28: (Left) The design of the prototype GE1/1 VTRx Cooling plate. (Right) The VTRx Cooling plate fit with the GE1/1 optohybrid and VTRxs.

cooling to apply extra upward pressure as shown in Fig. 6.29. The hope was that this pressure would improve the thermal contact. However, during the testing of new cooling the pressure proved to be too much; the excess pressure caused disturbances in communication. Therefore, the foam padding was removed in the next version and thermal pads were very precisely cut to avoid any pressure on the flexwire connectors as demonstrated in Fig. 6.29.

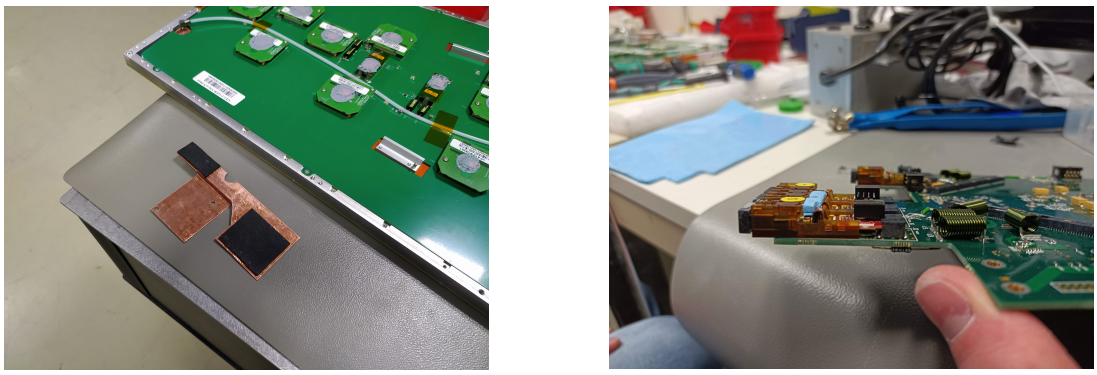

Figure 6.29: (Left) The foam padding that caused the pressure issue. This padding was removed. (Right) The precisely cut thermal pads.

After resolving the issue with excess pressure, the VTRx cooling prototype became a success. The temperature gradient between the photodiode and fiber tip was measured using a modified VTRx. During testing, the chamber was operated for over 190 hours and had a stable temperature difference of 3 °C, Fig. 6.30 . Additionally, the RSSI current (current from the photodiode) for all three VTRxs was monitored and remained at ex-

pected high values of $> 300\,\mu A$; the observed currents for this test are shown in Fig. 6.30 . Previously the temperature difference between fiber tip and photodiode was 15 °C, this was much higher than the ideal gradient determined by HCAL, but the new cooling has brought GE1/1 into this region. The GEM result and comparison to HCAL are shown in Fig. 6.31 .

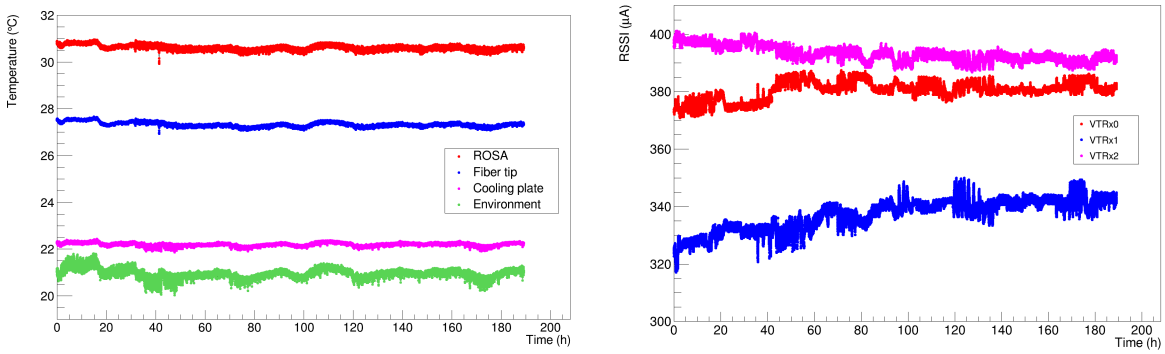

Figure 6.30: (Left) The temperature data of the monitored VTRx; the difference between the fiber tip (blue) and photodiode (red) was approximately 3 °C for 190 hours. (Right) The RSSI current of all three VTRxs on the optohybrid.

This is an effective solution to this problem, however it will take a massive endeavor to implement it. All 144 chambers will have to be extracted from the CMS experiment. Then, the VTRxs on each chamber will be replaced and the new VTRx cooling installed. This is a significant disturbance to the front-end electronics, so a full QC 7 electronics test is needed to validate the chamber before re-insertion into the experiment. This requires a dedicated team and months of testing during Long Shutdown 3.

6.5.5 Commissioning with Cosmic Rays

Muons originating from Cosmic Rays were used to prepare the detectors in the CMS experiment for LHC operations. This was done in two ways, by taking data with the experiment with the magnet turned off and with the magnet at 3.8 T. These data taking runs at CMS are referred to as "global runs". At first, these runs were used to test GE1/1 in the L1 trigger and to properly align GE1/1 in time and space with the other subsystems at CMS.

After completing these initial steps, the global runs were used to study the efficiency

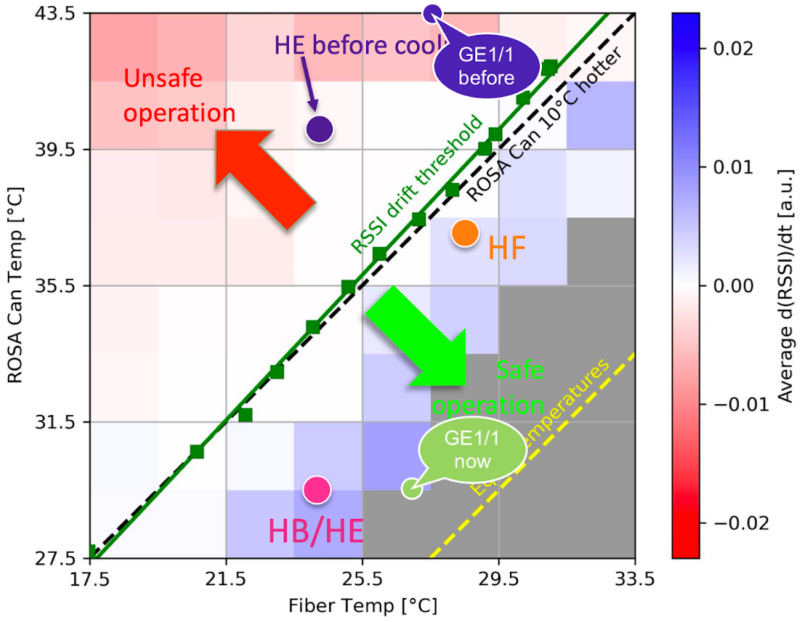

Figure 6.31: This plot was made by the CMS HCAL group. The y-axis gives the photodiode temperature and the x-axis gives the fiber tip temperature. The color is a measure of how much the RSSI current fluctuates. Below the diagonal line is the "safe" operational region and the addition of cooling brings GE1/1 below this line. Figure from [31].

of the GE1/1 detectors. This is done by measuring a muon track in the Cathode Strip Chamber system and then *propagating* the track to the GE1/1 system. Then, a corresponding GE1/1 hit is searched for in the vicinity of the propagated hit. This process is shown in Fig. 6.32 and the corresponding efficiency is defined as

$$\text{Efficiency} = \frac{\text{Matched Hits}}{\text{Total Propagated Hits}}\,. \tag{6.1}$$

The theoretical efficiency of the GE1/1 detectors is greater than 99%. However, electronic and detector imperfections can lower the efficiency of a GEM chamber. Therefore, during the commissioning process the objective is to reach an efficiency of greater than 90%. An example of an ideal chamber efficiency increasing as the HV settings are increased is shown in Fig. 6.33. A CMS event display showing a Cosmic Ray muon traveling through the CMS experiment (including GE1/1) is shown in Fig. 6.34.

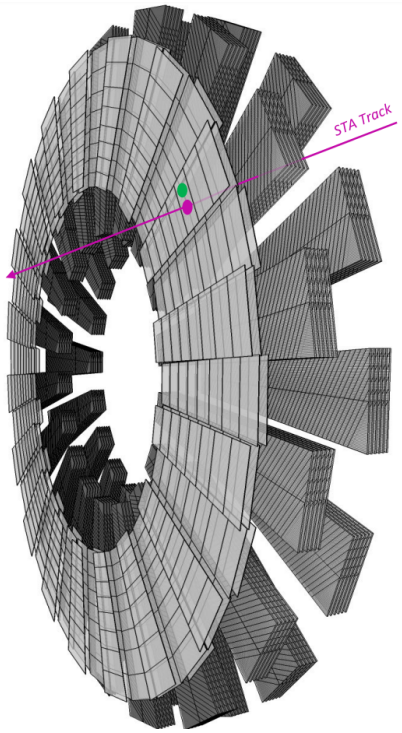

Figure 6.32: The thin light grey chambers are the GE1/1 system and the thick dark grey chambers are Cathode Strip Chambers. The purple line shows a muon track recorded by the Cathode Strip Chambers that is propagated to the GE1/1 system and the green dot shows the corresponding hit in GE1/1. Figure from [26] and modified by Francesco Ivone.

6.6 The future of GEMs at CMS

The GE1/1 detector is now operational in LHC Run 3 and will continue to operate in the HL-LHC. An event display of one of the first Run 3 events with GE1/1, and a very happy GEM team, is shown in Fig. 6.35. During the LHC Long Shutdown 3, the GE1/1 detectors will have to be extracted from the experiment to allow for the installation of the High Granularity Calorimeter. At which point the GE1/1 detectors will be equipped with VTRx cooling and re-tested before being installed in the experiment. The cooling will need to be carefully applied to GE1/1 and hopefully the details recorded in this chapter can prove useful to future GEM teams.

GE1/1 is only the first pair of GEM rings to be installed in the CMS experiment, two more are planned to be installed. The next GEM station will be GE2/1 and is scheduled for installation during an LHC extended technical stop at the beginning of 2024. GE2/1 will cover $1.6 < |\eta| < 2.4$, slightly more than GE1/1, but further away from

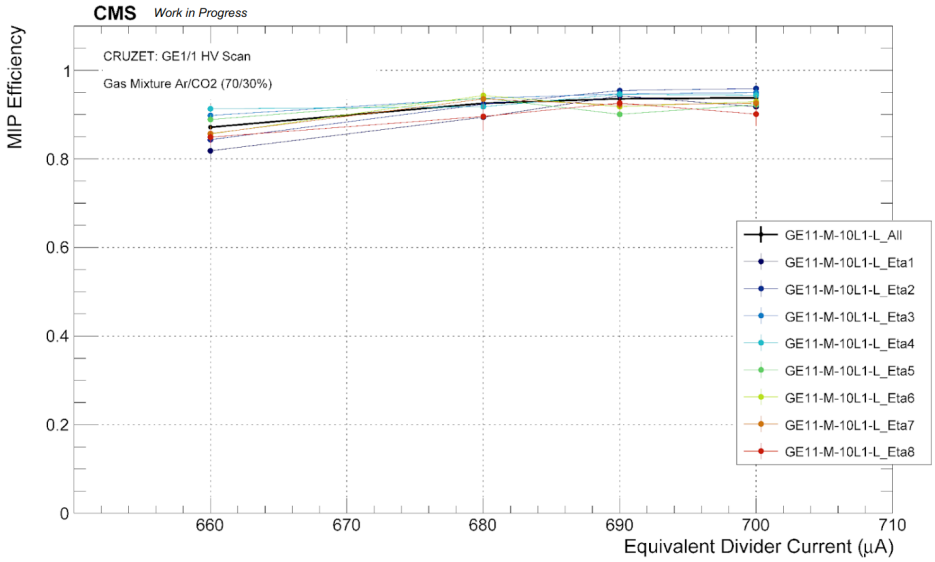

Figure 6.33: The efficiency of a GE1/1 chamber as the HV is increased. The equivalent divider current is a measure of HV in the chamber, as this is increased, the efficiency of the chamber improves in each η-partition.

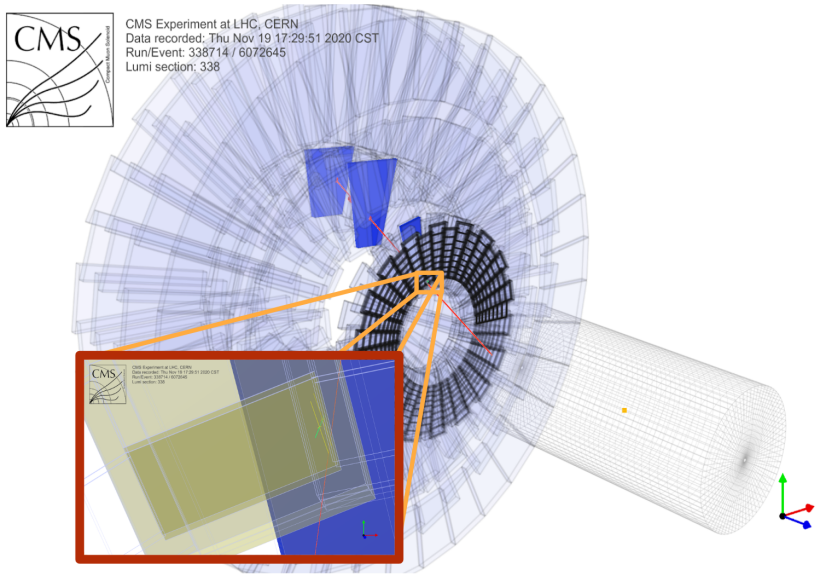

Figure 6.34: A Cosmic Ray muon travels through the CMS experiment during a Cosmic run without magnetic field. The zoomed in region shows the muon hits in GE1/1.

the interaction point. Therefore, it needs to cover a large surface area. To allow for this, the GE2/1 detectors are split into four separate GEM modules. This will be one of the largest area GEM detectors ever constructed.

ME0 is the last GEM station scheduled for installation during LHC Long Shutdown

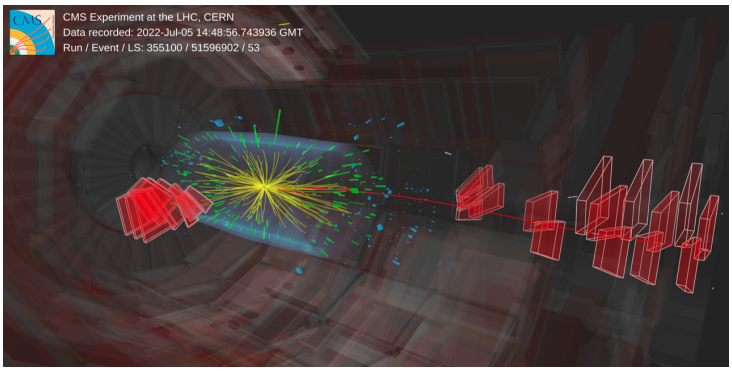

Figure 6.35: (Left) One of the first collision events from LHC run 3 showing a muon traveling through the muon endcap. The inner most muon detectors are GE1/1. (Right) The GEM team, with the author on the right, in the CMS control room during the beginning of LHC run 3.

3. This new detector will consist of 6 layers of GEM chambers and will occupy η-region $2.4 < |\eta| < 2.8$. This will be the first muon detector in this η-region, but it comes with a significant number of challenges due to the expected high rates. A number of novel techniques are being employed for the first time to make a detector capable of handling these rates. These include radial electrode segmentation [92], and random hole sectorization [93].

The techniques being utilized at CMS have far reaching results that allow for other experiments (and industries) to create large area GEM detectors with high rate capabilities. Simply put, the future of GEMs *shines* brightly.

Chapter 7

Particle Reconstruction in CMS

≫Und bei unseren unbekannten Abenteuern gibt es ja auch widersprüchliche Fassungen.≪

≫Du meinst deine und die Wahrheit?≪, fragt das Känguru.

≫Genau≪, sage ich. ≫Deine und die Wahrheit≪

Marc-Uwe Kling *Die Känguru-Apokryphen*

The CMS experiment is designed to record a wide variety of measurements, from an array of different particle detector technologies. The results of an individual proton-proton bunch crossing looks like a scattered record of signals, energy measurements, and noise. In order to be used in an analysis, these measurements need to be assembled into a more useful object. At CMS, these measurements are used to create six types of particle candidates: muons, photons, electrons, charged hadrons, and neutral hadrons.

Hadronic particles are clustered together to form "Jets". Jets come with a series of important definitions, features, and areas of research. Together with leptons and photons, these objects form the building blocks of an analysis at CMS.

The previous chapters described the outermost subsystem at the CMS experiment, the muon system, and they detailed it's upgrade with the GEM detectors. This chapter briefly describes how the detector subsystems at CMS reconstruct particles and then moves into the very heart of the CMS experiment in order to utilize the high resolution trackers and calorimeters which capture the decays of hadronic particles.

7.1 Reconstruction

A collision event at CMS leaves a lot of information scattered throughout the detector. This includes signals in the tracking and muon systems (hits) and deposits of energy in the calorimeters. This information needs to be carefully combined together to display the paths of the particles in the detector in a process called event reconstruction.

An understanding of particles and their paths through the experiment is necessary in order to properly reconstruct them. The tracker is designed to be as thin as possible to minimize the probability of a neutral particle (neutral hadron or photon) from interacting with the detector while still maintaining a high probability to have a signal due to a charged particle. The ECAL is designed to cause and capture the showers of electrons and photons while the HCAL causes and captures the showers of hadrons. The muon system tracks the charged particles which do not shower within the detector.

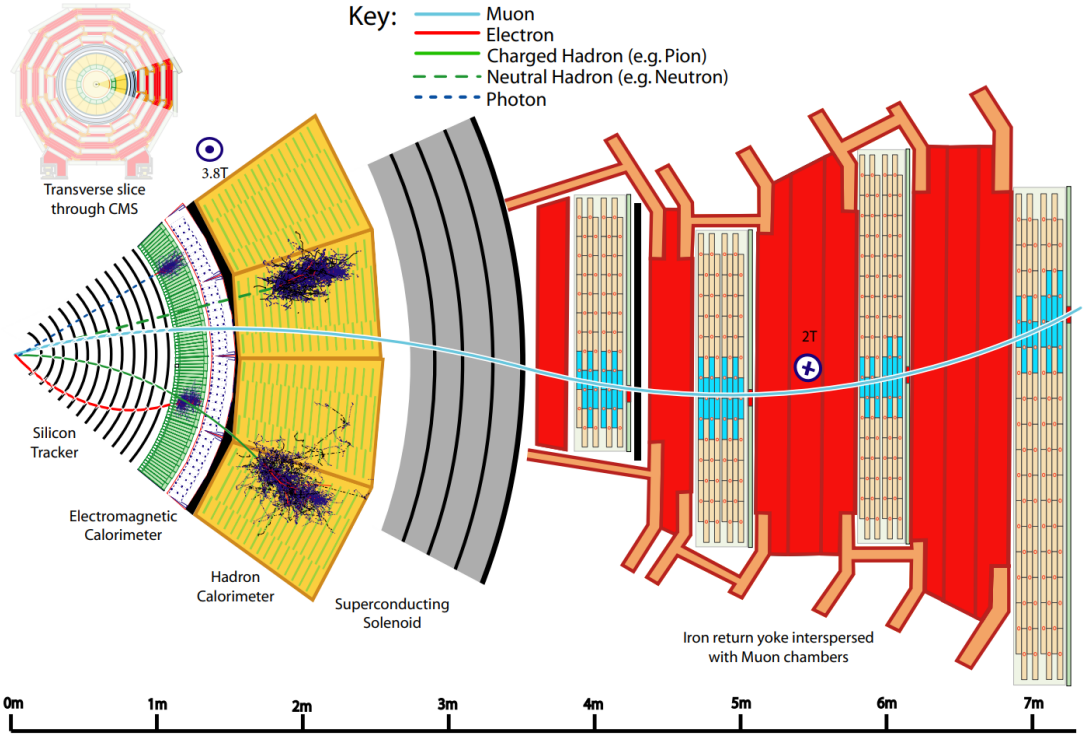

Figure 7.1: A transverse slice of barrel region of CMS showing the paths of different particles inside of the detectors. Figure from [32].

This resulting particle paths are summarized in Fig. 7.1. Electrons leave hits in the tracker as they travel on a curved path in the magnetic field, then deposit energy in the

ECAL. Photons will usually travel through the tracker without hits and then deposit energy in the ECAL. Charged hadrons travel similarly to electrons leaving a curved path of hits in the tracker, but deposit most of their energy in the HCAL. Neutral hadrons travel similarly to photons while depositing most of their energy in the HCAL. Muons will leave hits in the tracker, but travel through the HCAL and ECAL without showering, then pass through the muon system leaving hits in a path that curves in the opposite direction as the magnetic field switches direction outside of the solenoid.

7.1.1 Particle Flow

The CMS algorithm used to reconstruct particles is called "particle flow". The central idea of this algorithm is to link the signals in different subsystems together to reconstruct particles. The particle flow (PF) algorithm can be described in three steps:

1. Use hits to reconstruct tracks and energy deposits to reconstruct calorimeter clusters
2. Establish links between the detector subsystems by creating PF blocks
3. Identify PF blocks as particles to create PF candidates

Each of these steps are made possible at CMS due to the fine-grained tracking, large uniform magnetic field, highly segmented calorimeters, and high efficiency muon system. These steps are summarized below and are described in detail in [32]. The concept of creating particle candidates from a detector measurement with Particle Flow is summarized in Fig. 7.2.

Step 1: Tracks and Clusters

The tracking portion of this step fits a series of hits in the tracking system using an iterative approach called combinatorial track finding based on Kalman Filtering [94]. This approach works in three stages:

1. Seeds are generated from hits compatible with the path of a charged particle
2. A trajectory is built by gathering hits from all tracker layers compatible with the path of a charged particle originating from the seed

3. A final fit determines the charged particle properties: origin, transverse momentum, and direction

The first iteration of reconstructed tracks are required to be seeded with at least two hits in consecutive layers, to have at least eight hits with at most one missing layer, to originate from a cylinder around the interaction region within a few mm radius of the beam, and to have $p_T \geq 0.9$. The track finding algorithm starts with these strict requirements on the track quality. As tracks are identified, the corresponding hits are removed before the next iteration of the algorithm. With each iteration, the quality criteria are loosened until no more tracks can be reconstructed. This approach increases efficiency while keeping the misidentification rate as low as possible [83, 32].

In the calorimetry portion of this step, the energy deposits in the HCAL and ECAL are collected into bunches. For each region of the calorimeter, there is a corresponding noise threshold. The clustering process begins by searching cells with energy above the noise threshold and identifying the highest energy cell relative to neighboring cells. This cell becomes the seed for topological clustering. Then cells around the seed with energy at least twice as high as the threshold are added to the cluster. The applied noise thresholds can cause an observed energy less than true energy; this is especially true for low p_T particles [83, 32].

In the muon system, the hits are combined to create muon tracks. At this point, the muon tracks, tracker tracks, ECAL clusters, and HCAL clusters are all separate objects.

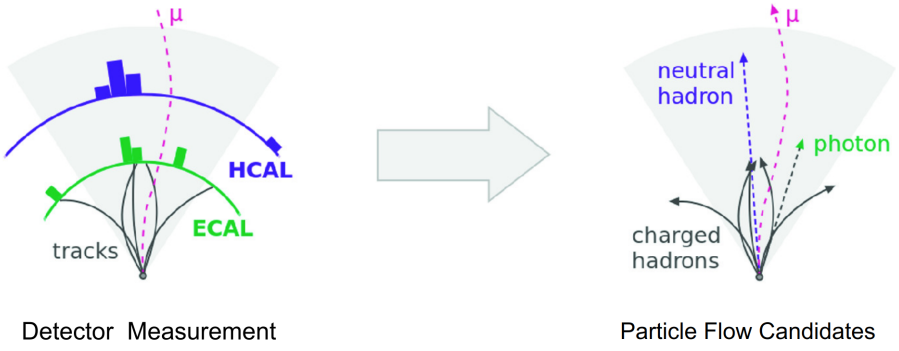

Figure 7.2: (Left) Detector measurement showing tracks from tracker hits and towers from calorimeter measurements. These measurements are constructed into particle candidates (right) with the Particle Flow algorithm. Figure from [33].

Step 2: Links between subsystems

Now the tracks and clusters are linked across subsystems into PF blocks. This process starts with linking the muon system to the tracker system. First, the path from the muon system is extrapolated into the tracker. Then, the path from the tracker is extrapolated into the muon system. If the two extrapolations agree with the reconstructed paths, then the two are linked into a PF block and the tracks are removed from consideration for the next PF objects [83, 32].

After all of the muon tracks are used, the algorithm continues. If an ECAL cluster is within the envelope of an HCAL cluster, then the two are linked [83, 32].

After linking ECAL and HCAL clusters, the algorithm links the tracker to the calorimeters. First, the tracks are extrapolated into the calorimeters; if there is a cluster, then the tracks are linked to the cluster. The quality of this link is measured by the distance between the extrapolation and clusters in (η, ϕ) space (η is pseudorapidity and ϕ is the azimuthal angle). As these PF blocks are created, they are removed from consideration during the next iteration. [83, 32].

Step 3: Formation of a PF Candidate

The identification of PF candidates starts with muons. After a PF block linking the tracker and muon system is established it is labeled as a muon candidate with a corresponding quality rating. Electron candidates are the next to be identified as the PF blocks which link a track to energy in the ECAL. Then, ECAL clusters without a corresponding track are identified as photon candidates. Next, HCAL clusters without a track are labeled as neutral hadrons [83, 32].

After this initial step of identification, the ECAL and HCAL clusters linked to tracks are considered. The clusters are calibrated and then the momenta of the charged particles is estimated from the tracks. If the difference between the calibrated energy and momenta estimate is small, then the particles are identified as charged hadron candidates. If the difference is large, then the cluster also consists of neutral particles (neutral hadrons and photons), which can be divided into two cases. If the difference is smaller than the energy

deposited inside of the ECAL and is $> 500\,\text{MeV}$, then the energy is labeled as a photon candidate. If the difference is larger than the energy deposited inside of the ECAL, then the energy in the ECAL is labeled as a photon and subtracted from the difference. The rest of the difference is labeled as neutral hadron [32].

Finally, the vector sum of the transverse momenta in the event is calculated and its negative becomes the missing transverse momentum of the event [95, 32].

7.1.2 Vertex Reconstruction

A vertex is a point of intersecting tracks due to a particle interaction. The proton-proton interactions from a bunch crossing are called primary vertices; these include the "signal" interaction and secondary interactions called "pile-up". Additional vertices displaced from the interaction point are called "secondary vertices". Secondary vertices arise from particles that travel a small distance into the detector before decaying, like flavour hadrons and taus. A primary vertex and nearby secondary vertex are shown in Fig. 7.3. The vertex reconstruction process is summarized below and a complete overview can be found in [96].

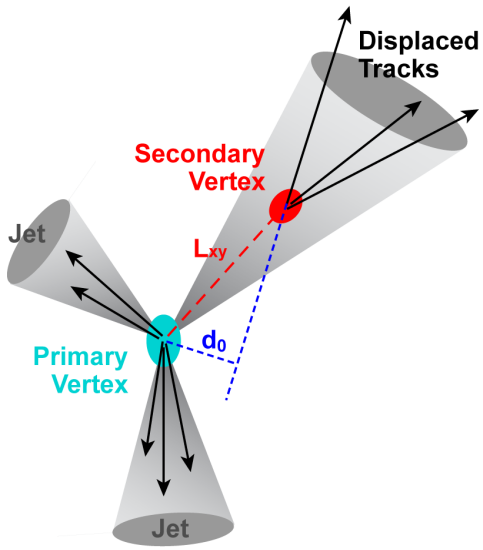

Figure 7.3: A primary vertex and displaced secondary vertex. The distance d_0 is the transverse impact parameter which can be used as a measure of displacement from the beam. Figure from [34].

Primary Vertices

The primary vertices are reconstructed in three steps:

1. Select high quality tracks using strict criteria as inputs for clustering
2. Cluster tracks that seem to originate from the same point
3. Fit the tracks to determine the location of the primary vertex

The tracks are selected based on the number of tracker hits and the transverse impact parameter (d_0) with respect to the beamline. The tracker hits ensure high quality tracks and the cut on the impact parameter prevents secondary vertices from being included. These selected tracks are clustered by a deterministic annealing algorithm [97]. After the primary vertices are calculated, the vertex with the largest $\sum p_T^2$ is labelled the "leading vertex". In simulation, the number of reconstructed primary vertices is, on average, 30% smaller than the total number of simulated interactions. This reconstruction efficiency is affected by pile-up mitigation technique. These techniques will be discussed in detail in the jet section [83, 96].

Secondary Vertices

Secondary vertices are displaced from the beam. On average, two thirds of secondary particles are charged. This justifies the use of charged hadron candidates to construct the secondary vertices. These vertices are reconstructed with an inclusive vertex fitter that uses all tracks with $p_T > 0.8$ GeV and $d_0 < 0.3$ cm where d_0 is the transverse impact parameter with respect to the beamline as shown in Fig. 7.3 [98, 99]. Nuclear interactions can create kinks in particle paths that resemble secondary vertices. In order to prevent this, during reconstruction it is required that a secondary vertex consists of at least three tracks where one must be an incoming track from a primary vertex and that the invariant mass formed by the incoming and outgoing tracks must be greater that 0.2 GeV [32]. The resulting secondary vertices are key for identifying flavour hadrons and taus.

7.1.3 Pile-up Mitigation

There are many proton-proton interactions during a bunch crossing in the CMS experiment. The vast majority of these interactions are due to peripheral interactions where the protons scatter off of each other without directly colliding. When an event at CMS is triggered on, the signal vertex for hard interactions is most likely the leading vertex. The leading vertex is defined as the primary vertex with the sum of tracks resulting in the highest value of $\sum p_T^2$. All other primary vertices coming from peripheral interactions are called "pile-up". The amount of pile-up per bunch crossing in the CMS experiment during LHC Run 2 is shown in Fig. 7.4.

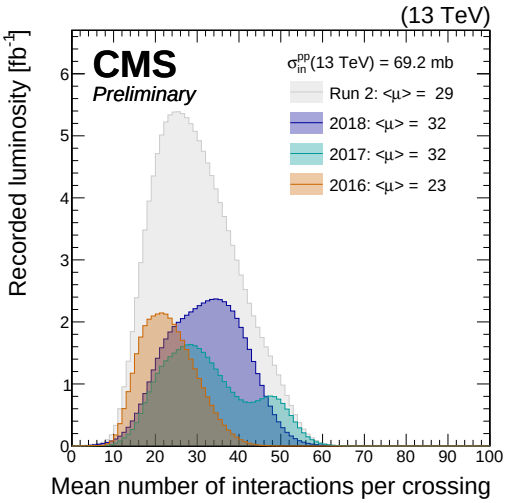

Figure 7.4: The number of interactions per proton bunch crossing in the CMS experiment during LHC Run 2. The vast majority of these interactions are pile-up. Figure from [35].

Algorithms to mitigate the effects of pile-up are a large area of study. Traditional mitigation techniques make use of three key pieces of information:

- The event wide pile-up density
- Vertex information from charged tracks
- Local distributions of pile-up with respect to particles from the leading vertex

The traditional techniques are summarized in [36]. In summary, each traditional algorithm alone is unlikely to optimally remove pile-up. Instead, the technique used in this thesis is

a relatively new algorithm that is quickly becoming standard at CMS called pile-up per particle identification (PUPPI) [36].

PUPPI is a flexible algorithm that combines global event information with local particle candidate information. It aims to remove pile-up rather than correct jet quantities by applying a weight to each particle candidate to rescale each particle's four-momentum. Ideally, particles from pile-up should receive a weight of 0 and particles from a hard scattering event should receive a weight of 1. The steps of the PUPPI algorithm are summarized here; a complete overview of the PUPPI algorithm is available at [36] and the CMS implementation at [35]

In order to construct a weight, a quantity α_i for each particle i must be defined where the distribution of α is different for pile-up particles and particles from hard scattering. For every particle i

$$\alpha_i = \log \sum_{j \in event} \frac{p_{Tj}}{\Delta R_{ij}} H \begin{cases} H = 1, & R_{\min} \leq \Delta R_{ij} \leq R_0 \\ H = 0, & \Delta R_{ij} < R_{\min} \text{ or } \Delta R > R_0 \end{cases}, \tag{7.1}$$

where ΔR_{ij} is the distance between particles i and j in (η, ϕ) space, R_0 defines the cone of nearby particles, and $R_{\min}$ is a regulator for collinear splitting. Collinear splitting of the parton shower causes particle i from a hard process to be closer to other particles from the same process. Whereas pile-up has no shower-like structure, so pile-up particles have a smaller value of α_i [36].

The central region of CMS ($|\eta| < 2.5$) has high resolution tracking that allows for distinguishing charged particles originating from the leading vertex. So in the central region of CMS particle j can belong to one of three different collections: the charged particles coming from pile-up (ch, PU), the charged particles coming from the leading vertex (ch, LV), or the neutral particles $(neutral)$ which have no tracking information. Therefore, α can be split into two regions: central (with tracking) and forward (without tracking).

$$\text{Central: } \alpha_i^C \text{ where the sum is over } j \in \ ch, LV\,,$$
$$\text{Forward: } \alpha_i^F \text{ where the sum is over } j \in \ event\,.$$

In other words, without tracking information, all particles are assumed to come from the leading vertex [36]. Distributions of these values from Monte Carlo simulated data are shown in Fig. 7.5.

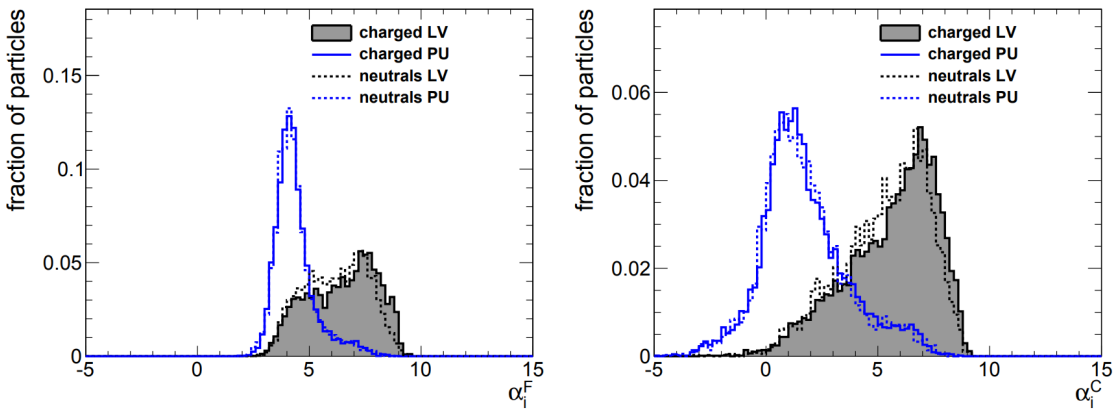

Figure 7.5: Distributions of α_i^F (left) and α_i^C (right) in Monte Carlo simulated data. Figure from [36].

The next step in creating a weight is to find α values that are characteristic of pile-up. In the central region

$$\bar{\alpha}_{PU}^C = \text{median}\left\{\alpha_{i\in ch,PU}^C\right\}, \qquad \sigma_{PU}^C = \text{RMS}\left\{\alpha_{i\in ch,PU}^C\right\}. \tag{7.2}$$

However, there are no tracks in the forward region, so the pile-up values must be measured in the central region and extrapolated to the forward region. At CMS, this is done using transfer factors derived from simulation

$$\bar{\alpha}_{PU}^F = \text{TF}_{\bar{\alpha}PU}\ \bar{\alpha}_{PU}^C, \qquad \sigma_{PU}^F = \text{TF}_{\sigma PU}\ \sigma_{PU}^C. \tag{7.3}$$

These values are then used to calculate χ_i^2. The CMS PUPPI specific definition is

$$\text{signed } \chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})|\alpha_i - \alpha_{PU}|}{\sigma_{PU}^2}. \tag{7.4}$$

α_i below the median is considered to be pile-up and above the mean is considered to be uncharacteristic of pile-up. When possible the values form the central region are used. The distribution of α for pile-up is roughly Gaussian and χ_i^2 resembles $\chi_{NDF=1}^2$. These values are then used to compute the weights

$$w_i = F_{\chi_{NDF=1}^2}(\chi_i^2), \tag{7.5}$$

where F is the cumulative distribution function for $\chi^2_{NDF=1}$ [36, 35].

The PUPPI algorithm uses these quantities and proceeds in seven steps:

1. α_i^C and α_i^F are computed for all charged pile-up in the event. Then these values are used to calculate the medians, RMS values, and χ_i^2 distribution.

2. All charged particles that are confirmed to be pile-up from vertex reconstruction are assigned $w_i = 0$ and all charged particles from the leading vertex are assigned $w_i = 1$.

3. The weights of all other particles in the event are calculated with Eq. 7.5.

4. Particles with $w_i < 0.01$ are rejected as noise and neutral particles must satisfy the condition:

$$w_i \cdot p_{T_i} > (A + B \cdot n_{vertices}) \text{ GeV} . \tag{7.6}$$

5. The four momentum of each particle is rescaled by its weight $p_i^\mu \to w_i p_i^\mu$.

6. The final set of particles is the PUPPI pile-up corrected event.

The result is the identification of charged particles from pile-up vertices that are removed from the event and the weighting of neutral (or untracked) particles [36, 35]. This is summarized in Fig. 7.6.

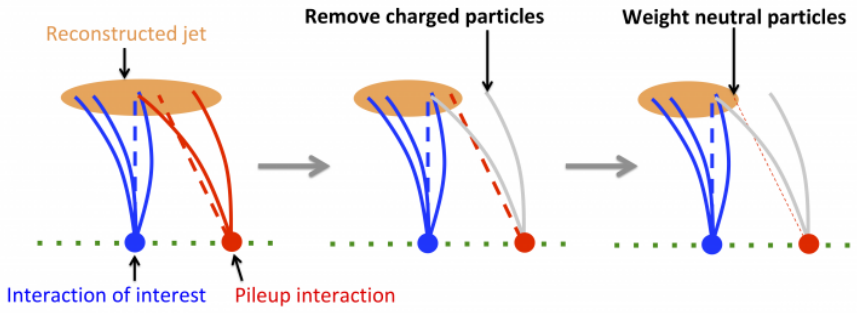

Figure 7.6: The pile-up corrections from PUPPI remove charged particles that originate from pile-up vertices and weight neutral (or untracked) particles. Figure from [37].

At CMS, the tunable PUPPI parameters are optimized in three different pseudorapidity regions. In the first two regions, $0 \leq |\eta| < 2.5$ and $2.5 \leq |\eta| < 3.0$, the parameters

optimized to ensure a near unity jet response as a function of pile-up. In the third region, $3.0 \leq |\eta| \leq 5.0$, the parameters are tuned to optimize the missing transverse momentum resolution. The tuned parameters for CMS during LHC run 2 are in Table 7.1 [35].

particle $\|\eta\|$	A [GeV]	B [GeV]	$\mathrm{TF}_{\alpha PU}$	$\mathrm{TF}_{\sigma PU}$
$[0, 2.5]$	0.2	0.015	1	1
$[2.5, 3]$	2.0	0.13	0.9	1.2
$[3, 5]$	2.0	0.13	0.75	0.95

Table 7.1: The tunable PUPPI parameters at CMS in each pseudorapidity region. Table adapted from [35].

One important note is that CMS currently uses a simple version of χ_i^2 which can be extended further. This can allow for other experimental inputs, like vertex uncertainties. It can also be extended to include information from future subsystems at CMS, like the MIP timing detector (MTD) which will be installed during LHC Long Shutdown 3 [36].

7.2 Constructing a Jet

Hadronic particles are clustered together into collections of particles with moments that lie in similar directions to create objects called "jets". At CMS, this process starts with reconstructing tracker and calorimeter data into particle candidates. Once successfully reconstructed, these candidates can be combined together into jets.

7.2.1 Jet Clustering

The process of combining particle candidates into jets is called "jet clustering". Experimentally speaking, the ideal jet is a conical object that forms a perfect circle in (η, ϕ) space. A fixed cone makes it easier to properly calibrate jets when compared to a variable size jet. The category of jet clustering algorithms that make conical objects with a fixed size are called "cone algorithms" [38].

However, theoretically there is a problem with this approach. For cone algorithms, the fixed boundaries can split a particle decay into two jets. If the cone is moved to fully capture the decay, then this will alter the number and contents of jets in the event. Pre-

venting this process from occurring for high momentum (hard) particles is called "collinear safety" and preventing it for low momentum (soft) particles is called "infrared safety". A jet clustering algorithm obeying both of these requirements is deemed infrared and collinear (IRC) safe [99].

The theoretically motivated category of jet clustering algorithms that sequentially cluster particle candidates into jets are called "sequential recombination algorithms". These algorithms are IRC safe, but often difficult to use in an experiment because they don't have fixed boundaries in the presence of soft radiation making calibrations difficult. Algorithms that have boundaries which are flexible around soft particles are called "soft-adaptable" and algorithms that maintain a fixed boundary in the presence of soft particles are called "soft-resilient" [38, 99, 83].

Sequential recombination algorithm starts by considering the distance d_{ij} between entities i and j and the distance d_{iB} between entity i and the beam. These entities can be particle candidates or pseudo-jets. Then clustering proceeds as follows:

1. Calculate d_{ij} and d_{iB} (the definitions vary depending on the algorithm)
2. Pick the smallest value d_{ij} or d_{iB}
    - If d_{ij} is the smallest, then combine entities i and j into one entity (pseudo-jet)
    - If d_{iB} is the smallest, then label entity i a jet and remove it from the list
3. Start at (1) using an updated list of entities

The distances for the most popular sequential recombination algorithms can be summarized with different powers n in the following equations

$$\begin{aligned} d_{ij} &= \min(p_{Ti}^{2n}, p_{Tj}^{2n}) \frac{\Delta_{ij}^2}{R^2}\,, \\ d_{iB} &= p_{Ti}^{2n}\,, \end{aligned}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is the distance in rapidity-azimuthal (y, ϕ) space, p_{Ti} is the transverse momentum of entity i, R is the desired jet radius in (y, ϕ) space, and n governs the power of the momentum versus geometrical (Δ_{ij}) scales [38].

The case $n = 1$ is the k_T algorithm where jets are clustered based on a p_T weighted (y, ϕ) distance. When $n > 1$, the particle ordering is maintained for finite Δ hence exhibiting a k_T like behavior.

The special case $n = 0$ is the Cambridge/Aachen algorithm which prioritizes only on Δ_{ij} proximity. The Cambridge/Aachen algorithm is not soft-resilient, but it is used at the CMS experiment for jet grooming (section 7.3.1). In this clustering algorithm, the nearest entities in (y, ϕ) space are combined together into pseudo-jets until a jet of radius R is formed [38]. This process is shown in Fig. 7.7

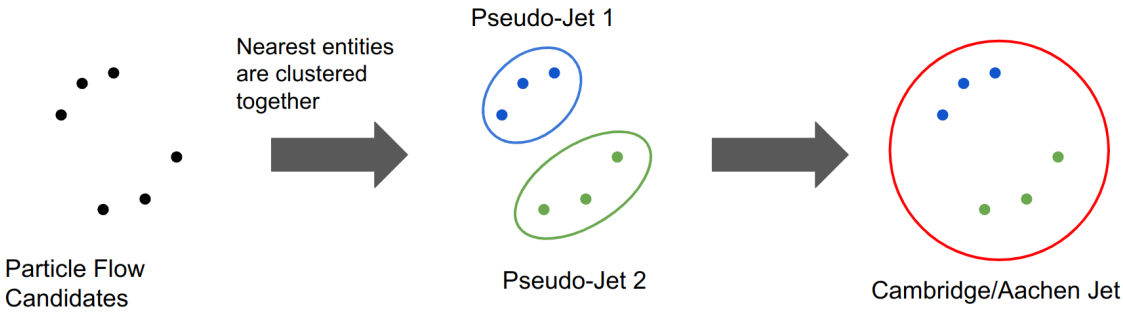

Figure 7.7: In the Cambridge/Aachen clustering algorithm, the nearest entities in (y, ϕ) space are clustered together into pseudo-jets until a jet is created.

The behavior for soft-radiation is the same for $p < 0$ and the specific case of $p = -1$ is the "anti-k_T" algorithm. In 2008, the anti-k_T algorithm was created to fulfill the need of a soft-resilient, IRC safe clustering algorithm. The clustering in this algorithm is a bit more difficult to describe graphically. To understand the clustering behavior consider a few well separated hard particles with $p_{T1}, p_{T2}, \ldots$ and many soft particles with $p_{Ti}, p_{Tj}, \ldots$

For hard particle 1 and soft particle i,

$$d_{1i} = \min\left(\frac{1}{p_{T1}^2}, \frac{1}{p_{Ti}^2}\right)\frac{\Delta_{1i}^2}{R^2} \tag{7.7}$$

is determined by the momentum of the hard particle. Thus, soft particles cluster with hard particles before each other.

- If a hard particle has no other hard particle within $2R$, it will accumulate all soft particles within R and become a perfect cone in (y, ϕ) space

- If another hard particle (particle 2) is present in $R < \Delta_{12} < R$, then both jets cannot be perfectly conical
  - If $p_{T1} >> p_{T2}$ then jet 1 will be conical and jet 2 will be partially conical (missing the overlap region)
  - If $p_{T1} = p_{T2}$ then neither jet will be a perfectly conical and will be divided down the middle
  - If $p_{T1} \approx p_{T2}$ then both cones will be clipped with boundary b between them such that $\Delta_{1b}/p_{T1} = \Delta_{2b}/p_{T2}$
- If hard particle 2 is present with $\Delta_{12} < R$ then 1 and 2 will cluster to form a single jet
  - If $p_{T1} >> p_{T2}$ then the cone will be centered around particle 1
  - if $p_{T1} \approx p_{T2}$ then the cone will include the union of cones (radius$< R$) around each hard particle plus a cone of radius R around the final jet.

Thus, soft radiation does not modify the jet shape (soft-resilient) and hard particles do [38].

Figure 7.8 shows a parton-level event plus $\approx 10^4$ random soft particles clustered with four standard algorithms: the k_T, Cambridge/Aachen, SISCone[1], and anti-k_T algorithms. Note that Cambridge/Aachen and k_T algorithms are jagged due to soft radiation, while SISCone has complicated shapes for composite jets. In anti-k_T, hard jets are circular with radius R and only soft jets suffer from complicated shapes. This makes the Anti-k_T algorithm IRC safe and easy to experimentally calibrate [38].

At the CMS experiment Particle Flow candidates are clustered with the anti-k_T algorithm with $R = 0.4, 0.8$ to create AK4 and AK8 jets (respectively). These are the standard reconstructed jets used in CMS analyses [99, 83].

[1]SISCone is an IRC safe cone algorithm.

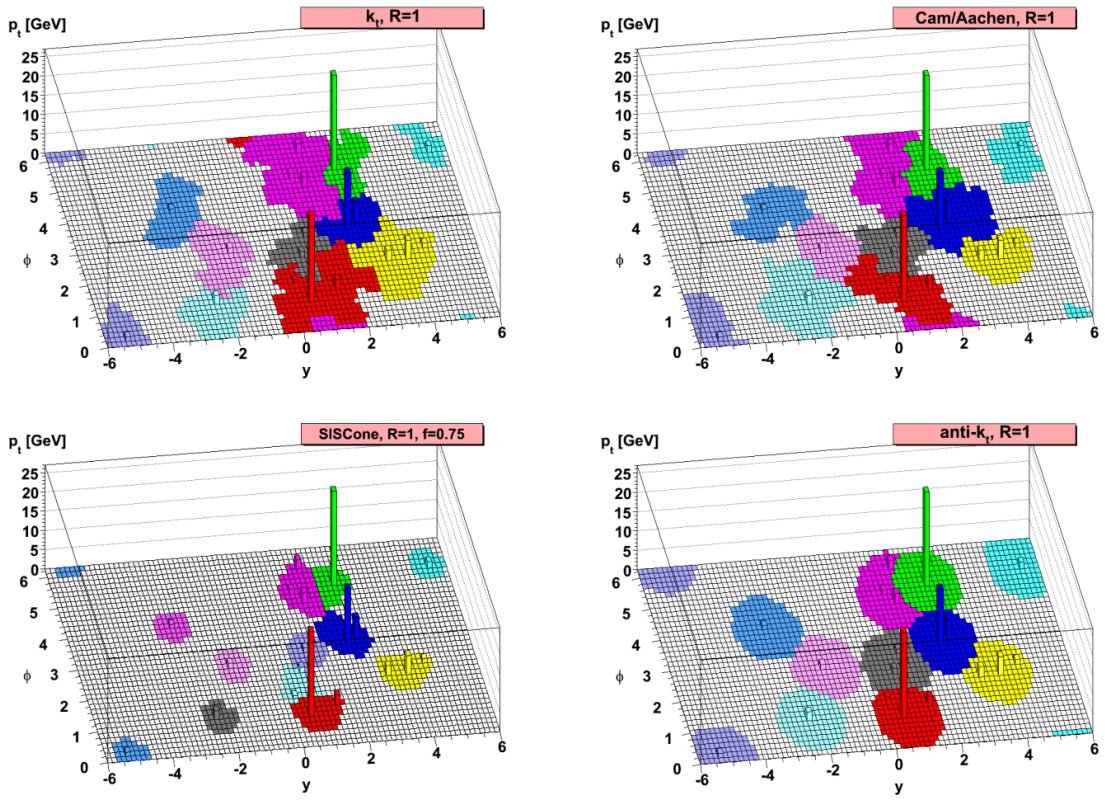

Figure 7.8: An event is clustered with four standard jet clustering algorithms with $R = 1$. The event consists of a parton-level event plus $\approx 10^4$ random soft particles. Figure from [38].

7.2.2 Jet Calibration

Reconstructed jets are subject to a number of effects that shift the reconstructed jet measurements from the true values. These include contributions from pile-up and detector effects (like calorimeter noise). Jet calibration is the process of correcting for these effects to ensure that Monte Carlo simulated data agrees with collision data. This consists of two components: corrections to the jet energy scale (with JEC) and corrections to the jet p_T resolution (with JER).

The CMS process for correcting the jet energy depends on the pile-up mitigation technique that is used. The process for jets mitigated with PUPPI is shown in Fig. 7.9.

The first step is to compare the particle-level jet (from simulated particles) to the reconstructed jet in Monte Carlo simulated data. The detector response used in CMS Monte Carlo simulation uses GEANT4 [100] and includes many detector effects. In this

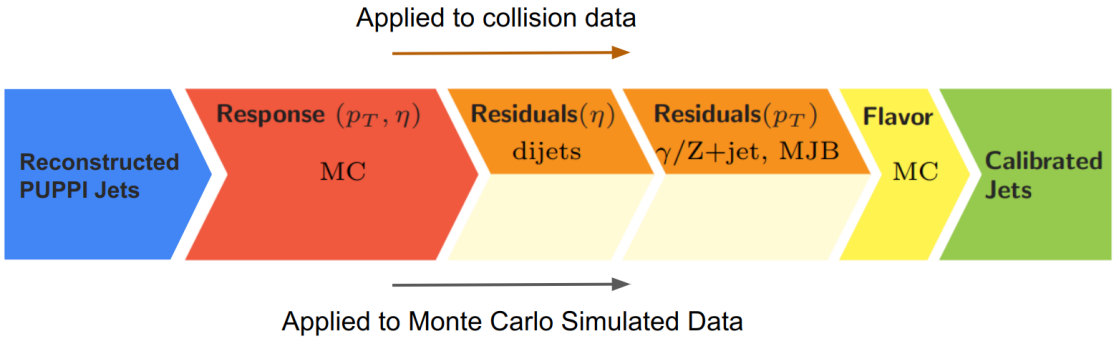

Figure 7.9: The CMS jet energy correction process for Jets mitigated with the PUPPI algorithm.

step the distribution of the detector response,

$$R = \frac{p_T(\text{reconstructed jet}) - p_T(\text{particle-level jet})}{p_T(\text{particle-level jet})} \tag{7.8}$$

is corrected to be on average zero. This corrects for known detector effects like non-linearity in p_T and η and electronic noise. The resulting correction is applied to collision data and Monte Carlo simulated data [83, 99, 101].

The next step is only applied to collision data. This step accounts for any residual differences between collision data and Monte Carlo simulated data as a function of η and p_T. These corrections are derived using collision data in well studied topologies. First, a dijet sample in data is used where one central jet $|\eta| < 1.3$ is used as a "tag" and a second in the opposite direction in ϕ is studied as a "probe". This corrects the jet response as a function of η. Then, Drell-Yan+jet (γ, $Z \to \mu\mu$, $Z \to ee$) and multi-jet samples are used to correct the jet response as a function of p_T [83, 99, 101]. Next, optional jet flavour corrections can be applied, these corrections are described in [101]. Together, these steps derive the JEC at CMS.

The JECs focus on adjusting the mean of the jet p_T response for simulation to agree with collision data. The next process is to correct the width of the jet p_T response (i.e. the resolution) with JER. These correction factors are derived in a similar method so that the resolution in simulation will agree with data as a function of η. The details of this process are described in [101]. After applying JEC and JER the jets are calibrated and ready for further analysis.

7.3 Jets for Jet Substructure

Boosted heavy particles decay into large-radius jets. Heavy particle decays often create characteristic features within the jet cones. For example, a boosted $H \to b\bar{b}$ decay contains a displaced vertex due to the lifetime of the b quark and two regions of energy in the calorimeter from the two b jets. The study of these unique features in large radius jets is called "jet substructure" and this field came to fruition around 2008. This was something completely unexpected during the planning of the LHC and now this field will shape the design of future detectors.

The radii of jets used at collider experiments is motivated by attempting to contain all of the unique substructure features from a heavy decay in a single jet. The two-body decay of a particle with mass m is roughly contained within a jet with radius

$$\Delta R \approx \frac{2m}{p_T}, \tag{7.9}$$

where p_T is the transverse momentum of the decaying particle [81]. The standard large radius, reconstructed jets at CMS are AK8 jets. This is intended to contain the decays of W, Z, H above 300 GeV.

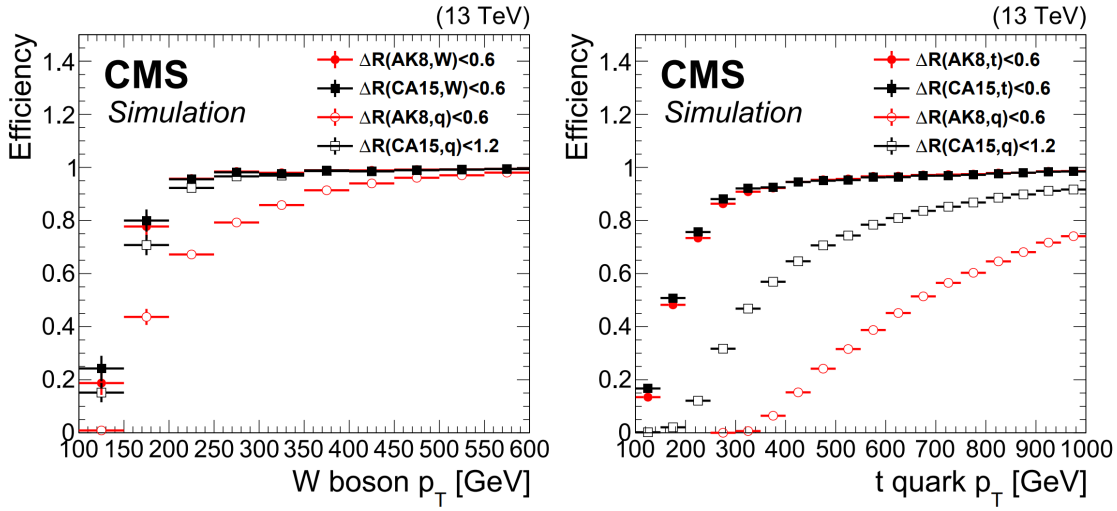

Figure 7.10: (Left) The containment of a simulated W boson (W) and its decay products (q) for Cambridge Aachen (CA) jets with R=1.5 and AK8 jets. (Right) The containment of a simulated top quark (t) and its decay products (q) for CA15 jets and AK8 jets. In this case, "well contained" is defined as within R=0.6 of the jet axis. Figure from [39].

However, top quark decays are much more difficult to constrain. In the Standard Model, $t \to Wb$ where the W immediately decays. This results in essentially a three-body decay which no longer obeys the rule of thumb. Figure 7.10 shows in CMS simulation that a W boson decay is well contained within the AK8 jet above 300 GeV, but a top quark decay is not well contained even at $p_T > 1000$ GeV. Thus, jet radii are optimized to contain the decays of all heavy boosted Standard Model particles except top quarks.

Large radius jets contain large amounts of particles not originating from the hard scatter. Thus, in order to extract unique substructure features from these jets, they must first undergo additional layers of processing.

7.3.1 Jet Grooming

Large-radius jets are particularly sensitive to contributions from pile-up, initial state radiation, and underlying events. These contributions can mask important substructure features. Grooming algorithms are employed to help remove this information. These algorithms recluster the jet constituents while removing soft and wide angle radiation that is most likely not due to the event of interest. The grooming algorithm employed by CMS during jet reconstruction is called "soft drop" [99].

The soft drop algorithm begins by reclustering a jet using the Cambridge/Aachen method. Then declustering along the clustering tree. The clustering tree is an diagram which shows the order in which entities (particle candidates or pseudo-jets) were clustered together while forming a jet. During this process the following condition is checked

$$\text{soft drop Condition: } \frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{cut} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}, \tag{7.10}$$

where p_{Ti} is the transverse momentum of particle i, ΔR_{12} is the distance between entities 1 and 2 in the rapidity-azimuth (y, ϕ) plane, R_0 is the jet radius, z_{cut} is the soft drop threshold, and β is an adjustable angular exponent. The soft drop condition is controlled by the parameters z_{cut} and β. When $\beta \to$ inf then the jet becomes an ungroomed jet [102].

The soft drop algorithm then proceeds as follows:

1. Undo the last step of the Cambridge/Aachen clustering algorithm by breaking the

jet j into pseudo-jets j_1 and j_2

2. Test the soft drop condition

   - If the condition is satisfied, label j the groomed jet and j_1, j_2 the soft drop subjets
   - If the condition is not satisfied, label the subjet with the highest p_T as j
     - If j is an event containing only one particle candidate then there are two options: either remove the entry j from the collection ("tagging mode") or label j as a jet ("grooming mode")
     - If j contains more than one clustered entity, then return to step (1)

It is important to note that tagging mode is only IRC safe for $\beta \leq 0$ and grooming mode is only IRC safe for $\beta > 0$ [102].

The soft drop algorithm moving along the cluster tree for a jet originating from a simulated boosted Z boson is shown in Fig. 7.11. However, visualizing the steps of the soft drop algorithm along a clustering tree is quite abstract. The steps can be much more easily visualized in Fig. 7.12 where the steps of the algorithm are shown in (y, ϕ) space [102].

The soft drop algorithm becomes the "Semi-classical" algorithm with $z_{cut} = 0.5$ and $\beta = 1.0$. It also becomes the "modified Mass Drop" algorithm [103] when $\beta = 0.0$. The ATLAS experiment recently began using $z_{cut} = 0.1$ and $\beta = 1.0$ in grooming mode, which is IRC safe [104]. However, the CMS experiment uses $z_{cut} = 0.1$ and $\beta = 0.0$ in grooming mode. This usage of the "mass drop" version is purely aimed at reducing soft radiation while maintaining high mass resolution [83, 99, 104]. However, it is important to note that the CMS condition is not IRC safe [102].

Subjets

At CMS, the subjets of an AK8 jet are defined as the subjets determined by the soft drop algorithm. These subjets belong to the cluster tree of the Cambridge/Aachen algorithm, so they do not have a fixed size and have very flexible boundaries. In other words, one

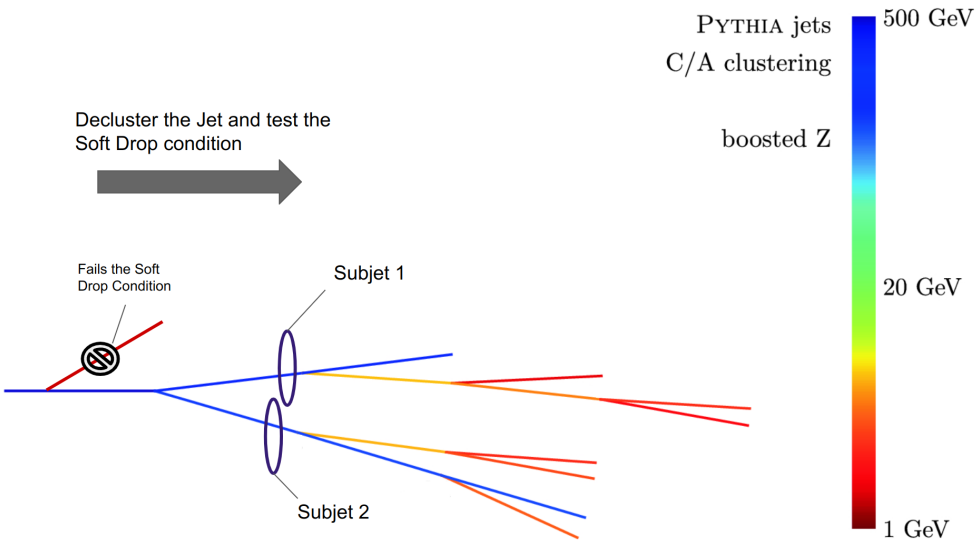

Figure 7.11: This is a cluster tree for a simulated boosted Z boson. The opening angle represents the (y, ϕ) distance and the color represents the energy of the entity. Proceeding from the left declusters the jet. Initially, some soft radiation is removed by soft drop, then two hard pseudo-jets passing the requirements are found. These two pseudo-jets become the soft drop subjets. Figure modified from the original format in [40].

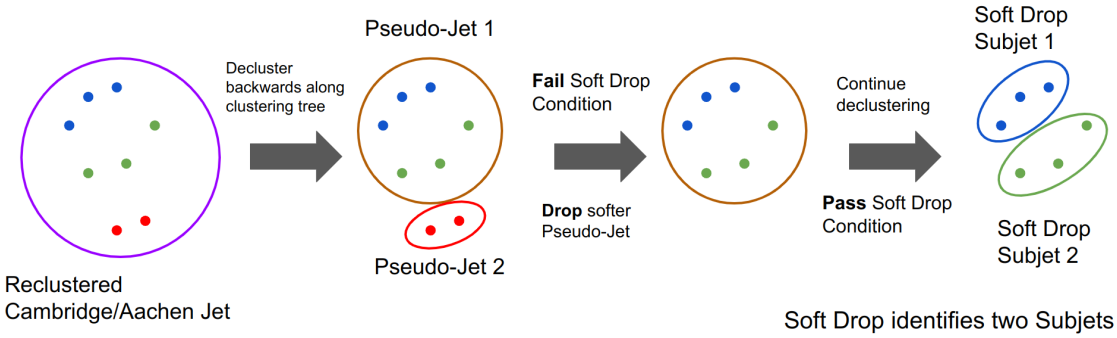

Figure 7.12: The soft drop algorithm declustering a jet in (y, ϕ) space. Initially a soft pseudo-jet is removed, then two hard pseudo-jets that pass the soft drop condition are identified. These two pseudo-jets become the soft drop subjets.

subjet could be radius 0.5 and the other could be 0.05. An important quantity to note is the soft drop mass, which is defined as the invariant mass of the jet after performing jet grooming with soft drop.

AK8 jets groomed with the soft drop algorithm ($z_{cut} = 0.1$ and $\beta = 0.0$) are the standard jets used to study substructure at CMS. These jets will be used for the neural network studies and analysis described in the subsequent chapters.

7.3.2 Jets with b Quarks

Hadrons containing b quarks travel a short distance from the interaction point before decaying in the detector. This creates a characteristic secondary vertex that makes a b hadron decay distinguishable from other particle decays. There are a number of algorithms at CMS for selecting jets containing a b quark; the CMS recommendation at the end of run 2 is the deepJet algorithm (also called deepFlavour).

DeepJet is a deep neural-network that trains low-level features of jet constituents. This consists of 16 properties of each charged jet constituent, 12 properties of each neutral jet constituent, and 12 properties of each secondary vertex. The output of this network gives a value from 0.0 to 1.0 as a measure of how likely the jet is to contain a b hadron. DeepJet is trained on and optimized for AK4 jets. However, it can also be applied to the soft drop subjets of an AK8 jet [105, 83].

7.4 Jet Creation Summary

The overall process for creating the jets discussed in this thesis is summarized in Fig. 7.13. After the Particle Flow candidates are weighted with PUPPI, candidates with weights equal to zero are passed to the FastJet algorithm [106] as ghosts[2] during clustering. The soft drop algorithm is used to create subjets and calculate a modified mass. The four vectors of all Particle-Flow candidates (even those from pile-up) and PUPPI weights are still accessible in the AK4 Jet and AK8 Jet collections. This allows teams at CMS to determine their own analysis specific criteria. The next chapter focuses on utilizing these jet objects to identify jets originating from boosted heavy particles.

[2]Ghost particles are included in a jet when contained with in the appropriate (η, ϕ) area, but are not used when clustering.

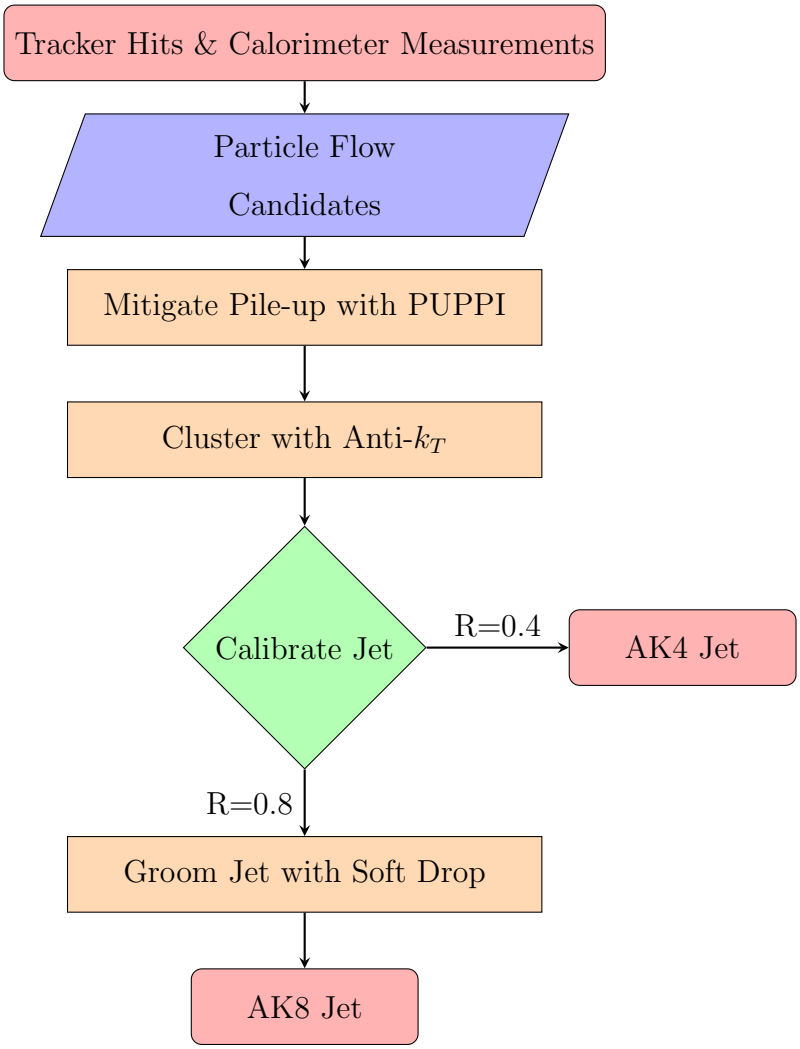

Figure 7.13: The CMS process for creating the jet collections used in this thesis.

Chapter 8

The BEST Story (Ever)

> *A composer's music should express the country of his birth, his love affairs, his religion, the books that have influenced him, the pictures he loves...*
>
> Sergei Rachmaninoff

In 1943 at the Cornell aeronautical laboratory, neurophysiologist Warren McCulloch and mathematician Walter Pitts postulated an electrical circuit to explain how neurons in the brain might work [107]. This idea led to the very first circuit-based artificial neuron called a perceptron—a tool that many scientists hoped would be able to solve complicated problems based on pattern recognition in much the same way as the human brain. However interesting, this first perceptron was underwhelming, only able to solve very simple problems.

Strange things can happen in science–sometimes an idea just needs time to mature. Over the subsequent decades, the perceptron was implemented as an algorithm [108] and then began to be connected into multiple layers. These networks were able to solve more complicated problems; the multilayer perceptrons are artificial neural networks and the field studying these networks is called deep learning. The first multilayer network was created in 1975, but suffered many complications during training [109].

The process of training a network involves minimizing a high-dimensional function and these difficulties were not overcome until the process of backpropagation came to widespread attention in the field in 1986 [110]. At this point, interest in neural networks

experienced a renewal that expanded the field to what it is today—a vast field of networks able to solve complicated problems from voice and picture recognition to self-driving cars. Naturally, these same techniques caused a shift in physics. A few applications at the CMS experiment include: triggering, particle tagging, and data analysis. Every day, physicists and computer scientists create novel approaches to interpret the vast amount of data produced by LHC.

In the case of the search for vector-like quarks, neural networks are used for jet tagging. The all-hadronic final state consists of 4 jets that are classified in six different categories using the Boosted Event Shape Tagger (BEST): top, Higgs, W, Z, bottom, or QCD. This chapter describes BEST and the effort to improve it. For the effort, I developed a method to make images of jets that had been boosted into hypothetical rest frames. Then, I designed and trained a network that uses these images to classify jets. Finally, I worked in a team to create an improved BEST that was ready to be used in the search for vector-like quarks.

8.1 Deep Learning for Jets

In 2008 an unexpected subfield arose at the LHC to study characteristic features of jets called jet substructure. This was mostly attributed to improvements in phenomenology, where observables were defined that could characterize jets. For example, n-subjettiness (τ defined in section 8.3) [111]. This is a measure of the number of separable regions of energy in a large radius jet.

This subfield continued to expand with the rise of deep neural networks. At first these new substructure variables were provided as high level input features into neural networks and used to train simple binary classifiers. For example: top versus QCD. Then, advances in deep learning and computing made it possible to utilize low-level input features, such as direct information about the jet constituents, to train powerful multiclassifiers.

In 2015, a method for creating images of jets made it possible to apply the field of computer vision to jet tagging. After the first conceptual papers [112, 113, 114], an image tagger quickly came to CMS called imageTop [39]. This tagger used two dimensional

convolutional layers to discriminate between top quarks and QCD.

The idea of using convolutional layers for jet taggers continued to evolve. In the same year, a group at CMS created a multiclassifier with one dimensional convolutional layers called deepAK8. Rather than using images, this tagger aimed to exploit low level features by comparing four vectors of jet constituents. The convolutional networks can use the four vectors to create geometric correlations. The hope is that this leads to a better tagger and that the network can be studied to find new jet substructure observables. This tagger became the standard at CMS for early full run-2 analyses using boosted jets, but has not yet lead to a better understanding of jet substructure [39].

One key component of any deep learning approach is data representation; in order to be represented for convolutional networks, jet constituent information is simplified to observe one dimensional correlations or represented as sparse images. In the former, the dimensionality of the problem is significantly reduced and in the later, the sparse images are very difficult to train a network with. Before this issue could even be studied in detail, the fast paced field of deep learning had made significant progress with graph networks. By 2018, graph networks became widespread and began to replace convolutional networks as the standard technique applied in jet physics at CMS. Graph networks more naturally describe jets by representing the jet constituents as elements of a discrete set of data points in n dimensional space; this set is known as a point cloud. At CMS, the successor to deepAK8 is called ParticleNet [115] and utilizes graph networks to classify jets.

At the same time that all of these developments were taking place, a new tagger was on the rise trying to exploit all that clever physics observables have to offer. This story begins in 2016 with a single paper [116] where the idea is to Lorentz boost along the axis of a jet into various hypothetical rest frames based on the known masses of Standard Model particles. In each of these frames, a number of event shape variables are calculated (defined in section 8.3).

These event shape variables were originally defined in the PEP/Petra era and used in the first run of the Tevatron. However, limitations in Monte Carlo techniques prevented

these variables from widespread use[1]. Thus, event shape variables were traditionally used in the lab frame and under-exploited. The rise of jet substructure along with the improved tracking, reconstruction, and clustering at today's experiments made it possible to realize the full potential of these variables.

The original paper showed that these event shape variables had discriminating power in these hypothetical rest frames. It also coined the term "boosted event shapes". This idea was used in CMS to create a tagger called the "boosted event shape tagger" or BEST. BEST is a simple dense neural network that uses the Boosted Event Shapes from four frames (top, Higgs, Z, and W) as input features and classifies jets as: top, Higgs, Z, W, bottom, or QCD.

BEST was used to search for vector-like quarks in 2016 data at CMS [43]. However, the BEST tagger underperformed when compared to deepAK8. This chapter—and a great deal of the work that went into this thesis—is about the effort to improve BEST and make it a competitive tagger at the CMS experiment.

8.2 Deep Learning Basics

BEST, deepAK8, imageTop, and ParticleNet are all deep neural networks trained to learn characteristic features of jets. Each of these networks use different architectures aimed at exploiting low or high level features. The theoretically most simple network is BEST, which consists of a fully connected dense neural network aimed at exploiting high level input features. The convolutional and graph neural networks used in the other taggers exploit low level input features.

In order to improve BEST to incorporate both low an high level features a number of techniques were used. The basics of deep learning and these techniques are summarized below. The field of machine learning also comes with its own set of metrics to evaluate training and compare network performances. These are briefly defined before being used throughout the rest of this thesis.

[1]Modelling background shapes with multiple jets was difficult until the rise of matrix Monte Carlo methods.

8.2.1 Neural Network Basics

The mathematical building block of a neural network is the artificial neuron which is loosely based on a biological neuron. A neuron takes in a series of inputs $\mathbf{x}$ these inputs are tensors that can be scalars, vectors, matrices, etc. The inputs undergo a tensor contraction with the kernel $\mathbf{W}$ and are summed with a bias $\boldsymbol{\theta}$. The result is passed to an activation function f which yields the neuron response

$$\boldsymbol{a}_\alpha = f\left(\mathbf{W}_{\alpha i}\mathbf{x}^i + \boldsymbol{\theta}_\alpha\right). \tag{8.1}$$

The kernel $\mathbf{W}$ and bias $\boldsymbol{\theta}$ are tunable parameters called weights. In order to introduce this concept, the rest of this subsection will focus on a simple perceptron where $\mathbf{W}$ is a matrix, θ is a vector where each element is a scalar bias, and $\mathbf{x}$ is a vector where each entry is a scalar input feature. In this case, the tensor contraction in the artificial neuron becomes a scalar product

$$a_j = f\left(\sum_{i=1}^{N} W_{ji}x_i + \theta_j\right). \tag{8.2}$$

A graphical description of an artificial neuron is shown in Fig. 8.1. In the rest of this chapter, "neuron" will refer to an artificial neuron.

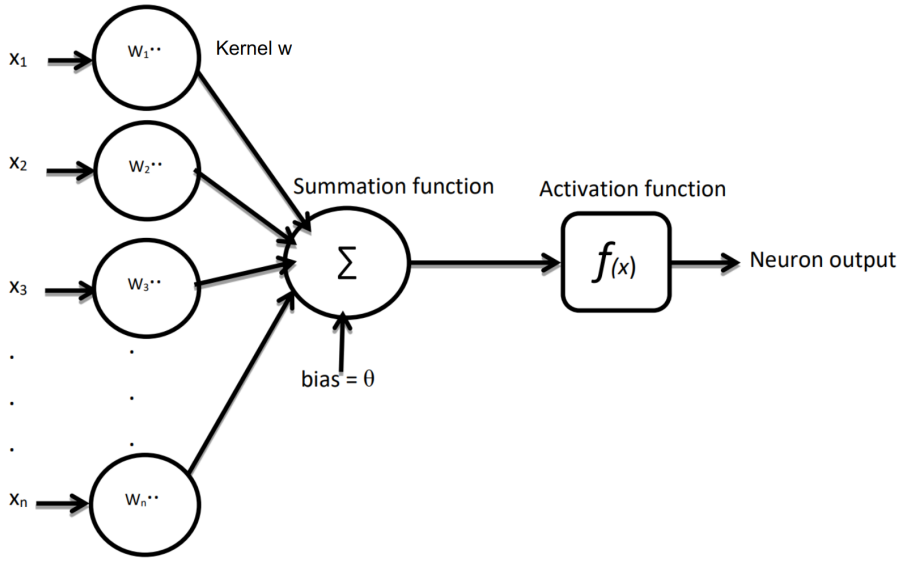

Figure 8.1: An artificial neuron with inputs $\mathbf{x}_i$ being contracted together to determine an output value. Figure from [41].

In a neural network, artificial neurons are linked together and trained on data. This thesis is limited to classification and in this case networks are trained on input features $\mathbf{x}$ belonging to different categories. The true categories of the categories are the targets $\mathbf{y}$ and the output of a neural network is a predicted category $\mathbf{y}_{pred}$. The training of a neural network proceeds in four steps:

1. Draw batch of the training input features $\mathbf{x}$ and targets (truth information) $\mathbf{y}$
2. Run the network on $\mathbf{x}$ and obtain $\mathbf{y}_{pred}$
3. Compute the mismatch between $\mathbf{y}_{pred}$ and $\mathbf{y}$ called the loss
4. Update the weights to reduce the loss on the batch

The final step of this process is called "backpropagation" and was made computationally possible to a technique called stochastic gradient descent [42].

The process proceeds as follows. First the loss $L_k(w)$ is computed as a function of each scalar weight w in the kernel $\mathbf{W}$ for an entry k in batch (subset) of the dataset. Then the weights are adjusted

$$w \to w - \frac{\eta}{N} \sum_{k=1}^{N} \nabla L_k(w) \,, \tag{8.3}$$

where η is the learning rate (or step size). When the batch size is just one entry from the dataset, this is the true stochastic gradient descent. When the batch size is larger than one, this is called mini-batch stochastic gradient descent. The process of stochastic gradient descent is shown if Fig. 8.2 [117, 42].

Before utilizing stochastic gradient descent, the loss was evaluated on the entire dataset. This was computationally very expensive; by training on batches of the dataset, the training time is significantly improved. However, there is still one common problem with this technique. During training, the neural network can get stuck in a local minimum of the loss function. Several variations on stochastic gradient descent are used for overcoming this problem to train network. One common method is to add an intermediate

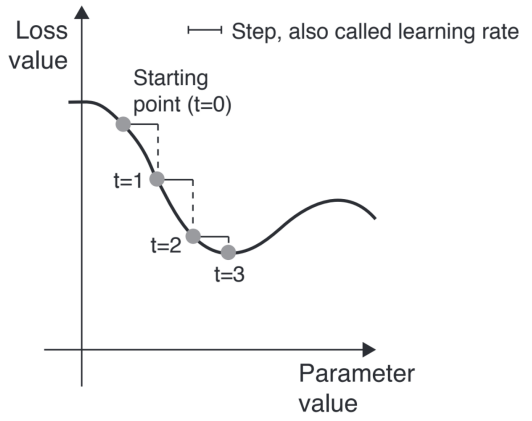

Figure 8.2: A visualization of stochastic gradient descent in one dimension. The weights are adjusted in each step as the loss is computed on a batch of the dataset. Figure from [42].

step while adjusting the weights

$$v_t = \gamma v_{t-1} + \frac{\eta}{N} \sum_{k=1}^{N} \nabla L_k(w) \,,$$

$$w \to w - v_t \,,$$

where v_t is the adjustment to the weights at step t and γ is user input scalar parameter. This concept (and the parameter γ) is called "momentum". This can be thought of as giving a "push" to an object so that it travel over the "hill" near a local minimum [117].

Stochastic gradient descent and these modified methods of updating the weights in a neural network are called "optimizers". Adam (adaptive momentum estimation) [118] is the optimizer used for training networks throughout this thesis. The Adam optimizer utilizes a modified version of momentum in stochastic gradient descent, the details of this process are described in [118]. This optimizer is commonly used in the machine learning community and currently recommended at CMS.

Like the various optimizers, there are different loss functions that can be minimized. Each loss function specializes in a different type of problem (binary classification, multi-classification, or regression). The neural networks used in this thesis are multiclassfiers that use the categorical cross entropy loss function which is defined for N categories as

$$\text{Loss} = -\sum_{i=1}^{N} \mathbf{y}_i \dot{\log}(\mathbf{y}_{pred_i}) \,, \tag{8.4}$$

where $\mathbf{y}$ is the vector with 1.0 for the true category and 0.0 for the others while $\mathbf{y}_{pred}$ is output vector of the neural network with an output value in each category. Categorical cross entropy is a tool to compare two probability distributions. Thus, $\mathbf{y}_{pred}$ must be normalized so that the elements sum to one.

The overall process of training a neural network through backpropagation is summarized in Fig. 8.3. In this figure each layer consists of neurons with weights that are updated by the optimizer using the loss score.

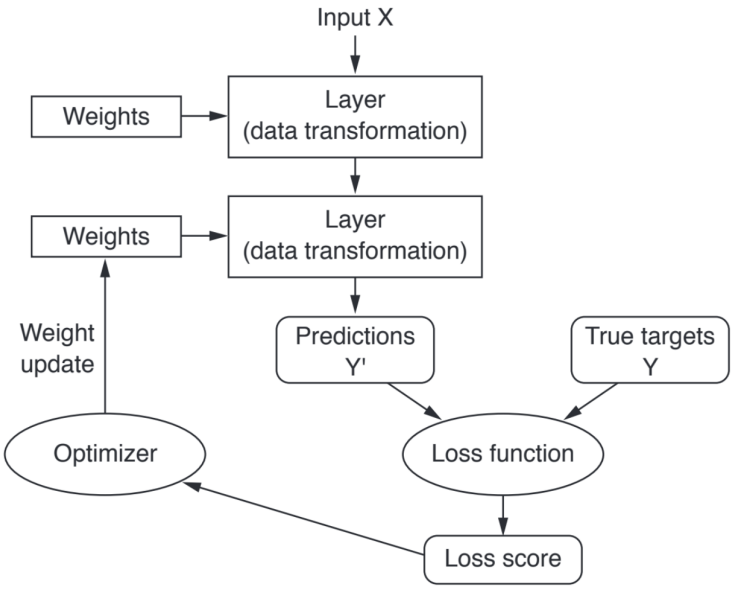

Figure 8.3: The process of training a neural network. Figure from [42].

The keen observer may have noticed that many of these terms used in deep learning are just a rebranding of the terms used in statistics. For example, "backpropagation" is just a type of minimization and the "loss function" is just a very high dimensional likelihood function. The heart of neural networks is just statistics.

8.2.2 Neural Network Architectures

Artificial neurons can be linked together in different ways to create a network. Each network architecture specializes in identifying different types of patterns in the input dataset. These have been well developed and studied for many industrial applications resulting in a wide range of knowledge on network architecture that can be adapted to the specific high energy physics case of classifying the origin of large radius jets.

Dense Neural Network

A dense (also called feedforward or fully connected) neural network consists of layers of artificial neurons. Each neuron in a layer is connected to every neuron in the next layer. The first and last layers are called the input and output layers respectively. Every layer in between is called a hidden layer. The structure of a simple dense neural network is shown in Fig. 8.4.

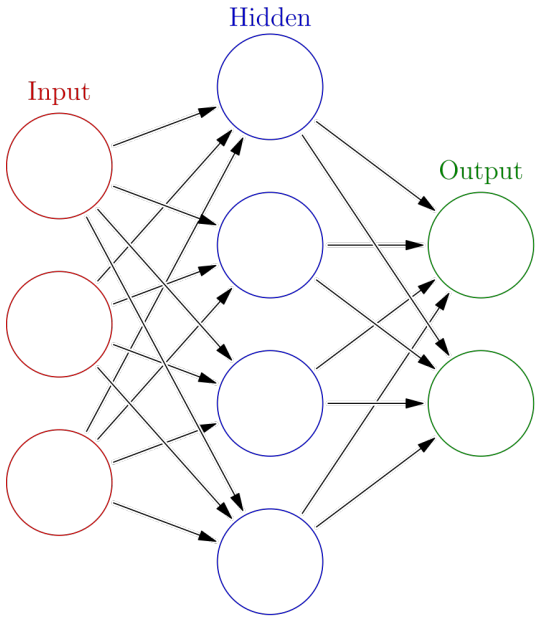

Figure 8.4: A dense neural network where the left hand side shows the input layer and each circle represents a artificial neuron. Then each neuron is connected to all of the neurons in the next layer through the hidden layer to the output layer. Figure publicly available.

A wide array of activation functions can be used in an artificial neuron. Often, different functions are chosen depending on the location of the nodes in the network. Hidden layers throughout this thesis use the Rectified Linear Unit (ReLU). This function is an element-wise operation

$$ReLU = \max(0, x)\,, \tag{8.5}$$

where x are the elements of the input tensor. With this activation function, the output of a neuron becomes the value of weighted sum when the element is non-zero. The ReLU activation function is widely used in the machine learning community [42].

When classifying an input dataset, it is desirable to have the output nodes normalized such that they can be treated like probabilities. For the output neurons in a binary classifier this is accomplished with the sigmoid activation function. For multiclassifiers, the sigmoid is modified to create the softmax activation function. This is defined as

$$\sigma_i(\mathbf{x}) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \text{ for } i = 1, \ldots, N \text{ and } \mathbf{x} = (x_1, \ldots, x_N)\,. \tag{8.6}$$

The resulting output sums to one across the categories and the value for each category can be treated like a probability [42].

The resulting dense networks handle scalar input features like the individual kinematics of a jet. These networks are advantageous over traditional machine learning techniques (like boosted decision trees) when there are a large number of input features and large training datasets as is often the case with CMS data.

Convolutional Neural Networks

Dense neural networks learn global patterns form the input feature space, but identifying jets often relies on more local features of the individual jet constituents. One tool for achieving this are convolutional neural networks (CNN). CNNs are commonly used in computer vision to categorize images. For example, a CNN can be trained on an input feature space consisting of pictures of elephants, lions, cats, and dogs. Then, it can categorize pictures into these four categories. The patterns in the images learned by the CNN are translationally invariant and subsequent convolutional layers allow CNNs to learn a spatial hierarchy of patterns. This is illustrated in Fig. 8.5 where the first layers identify local, translationally invariant features of the cat, such as the individual whiskers or shape of the ears. Then, the higher convolutional layers identify more global patterns like eyes, ears, or nose. These are used to identify the image as a cat [42].

Convolutions are operations on tensors called "feature maps". Commonly, 3D tensors are used where the dimensions are (height, width, depth). Depth is the number of input channels; in the case of image classification, often RGB (red green blue) images are used with each color representing a different input channel for a depth of three. A convolution creates an output feature map where each entry is a "filter" with the (height, width)

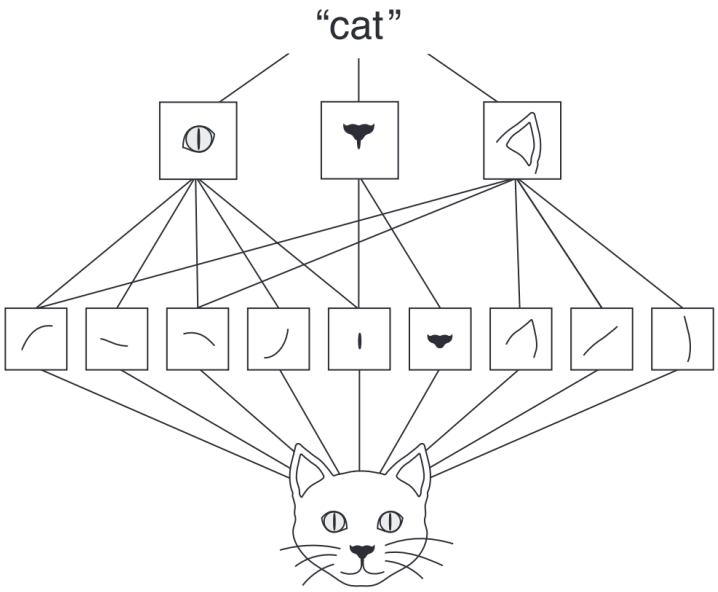

Figure 8.5: A convolutional neural network identifying an image of a cat starting with local features and then identifying more global patterns. Figure from [42].

from the input feature map the total number of filters (or depth) is set by the user when defining the convolutional layer. After training, each filter contains a specific aspect of the image. For example, in the case of the cat, the low level filters contain whiskers or outlines, then the high level filters contain ears, eyes, or nose. In addition to controlling the number of output filters, the user also controls the size of the patches (window height, window width) extracted from the input [42].

The convolution proceeds by sliding at windows according to the patch size. At each possible location a patch of surrounding features is extracted with (window height, window width, input depth). A tensor product is then performed between each patch and a tensor of learned weights (called a convolutional kernel). This creates an 1D vector with length equal to the output depth (i.e. number of filters). Each output vector is spatially reassembled into the 3D output feature map of (height, width, output depth). Thus, the spatial locations in the output feature map correspond to the same spatial locations in the input feature map [42]. This process is shown graphically in Fig. 8.6.

Convolutional layers typically use patch sizes of (3,3) or greater. Without intervention, these patches are unable to center on values at the edge of the tensor. A common method of dealing with this effect is to surround the edges with the value “zero” so that the

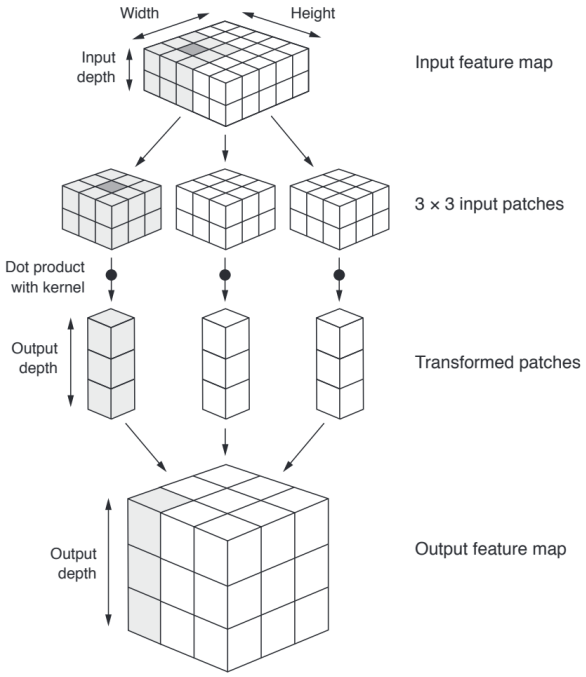

Figure 8.6: A graphical representation of a convolution. Figure from [42].

patches can create a window around every location. This method of "zero padding" is the one employed throughout the rest of the chapter [42].

After preforming convolutions it is essential to "down sample" the in the input maps. Down sampling reduces the amount of feature map coefficients and induces spatial filter hierarchies. Down sampling can be achieved by altering the stride length (see [42]) or by pooling. The pooling method with the most reliable results used throughout this feature is "max-pooling". Max-pooling aggressively down samples by extracting windows from input feature maps and sends the max value to an output feature map [42].

Like data for a dense neural network, feature maps should be normalized before a convolution is applied. Normalization makes the ranges of input samples more similar to each other creating a smoother loss surface. The input to the neural network is often normalized by the user. However, the values between each intermediate convolution are not normalized. In 2015, the technique "batch normalization" [119] was created to adaptively normalize data between convolutions even as the mean and variance change as training progresses [42].

Convolutional layers, max-pooling, and batch normalization make the basic compo-

nents of a CNN. After applying a sequence of these layers, the output tensor is flattened and sent to dense neural network for classification.

Recurrent Neural Networks

Artificial neurons in dense networks are only connected in one direction forward (i.e. on to the next layer), but neurons can be connected in any direction. When a artificial neuron is connected to itself (or other previous neurons), a feedback loop is created (Fig. 8.7). Neural networks containing feedback loops are called recurrent neural networks (RNNs). These feedback loops allow for the concept of memory to be introduced to a neural network.

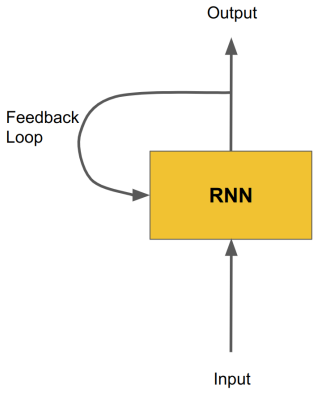

Figure 8.7: A feedback loop in a recurrent neural network (RNN).

RNNs take in a sequence of data like a string of text or images in a video. This sequence is treated as a single input. However, the input is no longer processed in a single step. Instead, the network internally loops over sequence elements while maintaining a "state" of information. A sequence consists of timesteps. The RNN loops over the sequence and at each timestep considers the current state at time t and the input, then combines them at time t. This process simulates memory [42].

In early RNNs, old signals gradually vanish during processing, meaning that the RNN is unable to retain information at time t that was observed many timesteps earlier. This is known as the "vanishing gradient problem" and it 1997 it was solved with the introduction of long short term memory (LSTM) [120]. An LSTM consists of a cell with neurons and feedback loops. Each cell has an input, output, and forget gate. This regulates the

flow in and out of (with the input and output gates) while also allowing the cell to discard irrelevant information (via the forget gate). Thus, creating a cell that can store information for an arbitrary amount of time [42].

LSTMs are used in handwriting recognition, voice recognition, and video processing. The LSTM was one of the most widely cited machine learning tools, in 2021 LSTM was cited over 16,000 times on Google Scholar. In terms of jets, LSTMs can identify patterns in sequences of jet constituents. For example, this is done in the deepJet and deepAK8 networks [105, 39].

Graph Neural Networks

The next generation of boosted jet taggers represent jet constituents as point clouds. A point cloud is simply a 3D representation of a data point. In this case the jet candidates can easily be represented as points. In the example of jet tagging with point cloud transformers [121], jet constituents are represented as points in (η, ϕ) space. Point clouds are a theoretically more accurate representation of jets than images. Networks that take point clouds as inputs are called "graph neural networks" and are rapidly gaining attention due to success in solving scientific problems. In July 2022, the GNN based AlphaFold by the Google DeepMind project correctly folded nearly all known biological proteins [122]. This result cemented GNNs as a highly valuable scientific tool.

GNNs were used for comparison, but not studied in this thesis work. Thus, a detailed discussion of GNNs goes far beyond the scope of this thesis. For an in-depth description of the highly popular point cloud based jet tagger called ParticleNet see [115] and for the point cloud transform see [121].

8.2.3 Training and Performance Metrics

The advanced mathematical structures of a neural network make it difficult to monitor the neurons directly during training and evaluation. Therefore, a number of simple metrics were created to accomplish these tasks. The training metrics enable the user to monitor the health of the network during training and provide a valuable tool for finding improvements. The evaluation metrics allow the user to see the success of the network

and compare the network to others.

The two most common metrics monitored during the training of a neural network are the loss score (output of the loss function for all entries in a dataset) and the accuracy. In the case of multiclassifiers, like the taggers discussed here, the categorical accuracy is used. This is defined as

$$acc = \frac{\text{Total Correctly Categorized Entries}}{\text{Total Entries in the dataset}} . \tag{8.7}$$

Before training a neural network the dataset is split into three separate subsets: training, validation, and test. Backpropagation is performed on batches of the training set. After the entire training set is run through (called an epoch) the loss and accuracy are computed on the validation set. Either metric evaluated on the training or validation set maybe used to define a stopping condition. It is common practice to use the accuracy on the validation set to stop training. The network may continue to show improvements in the accuracy and loss in the training set while saturating, or even increasing, in validation set. These are the signs of overtraining, when the model begins to train on the statistical noise of the training sample and is unable to generalize to new data. By defining a stopping criteria on the validation set, this ensures that the network no longer continues to train when the validation accuracy saturates. A third test set is used in order to tune the hyperparameters (such as number of layers, neurons, or other user input parameters). The test set is particularly important when utilizing hyperparameter optimization tools such as Bayesian optimization [42].

After training a network, the confusion matrix and receiver operator characteristic (ROC) are used to evaluate the performance of the network. In order to understand these, correctly and incorrectly data are split into four cases defined in Fig. 8.8.

A confusion matrix is a tool to visualize these quantities. The rows in a confusion matrix represent the true category Y and the columns represent the predicted category Y_{pred}. The bins of the confusion matrix contain the number of entries in Y that were predicted as X_{pred}. Thus, the diagonal are the true positives while the off diagonal elements are the incorrectly identified entries. Often, the confusion matrix is normalized by the number of entries in each row [123].

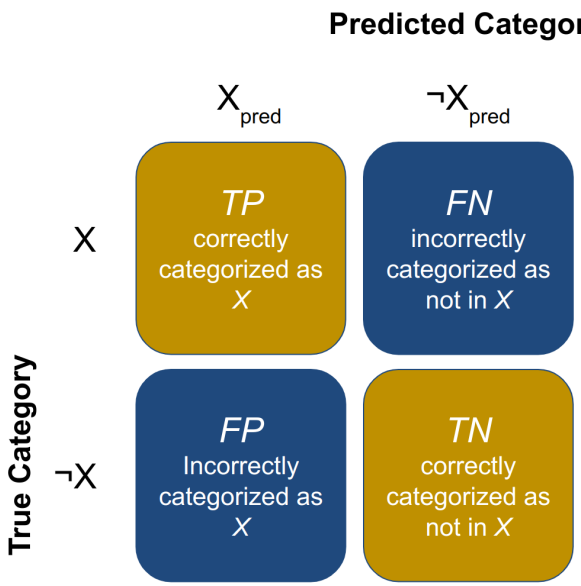

Figure 8.8: The definitions of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) for a category X.

A ROC plot shows the true positive rate (TPR) as a function of the false positive rate (FPR) where the rates are defined as

$$TPR = \frac{TP}{TP + FN},$$
$$FPR = \frac{FP}{FP + TN}.$$

This creates a curve shape where a perfect network would have a triangle in the upper left corner. An example is shown in Fig. 8.9. The area under the ROC curve (AUC) is a measure of how well the network can distinguish class X at some efficiency in Y and is a common metric used for comparing network performance.

In a multiclassifier, a ROC curve can be created for each individual category. However, often it is desirable to display the overall performance of a network on a single plot. Throughout this thesis this is accomplished with macro-averaging, where, for N categories

$$\overline{TPR} = \frac{\sum_{i=1}^{N} TPR_i}{N},$$
$$\overline{FPR} = \frac{\sum_{i=1}^{N} FPR_i}{N}.$$

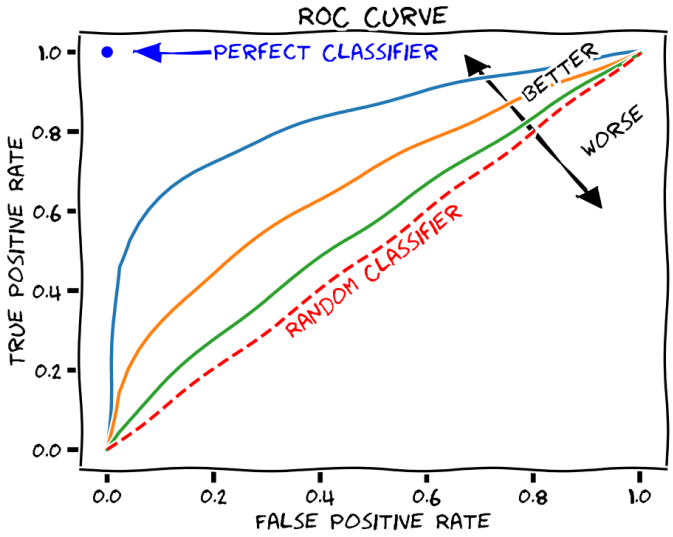

Figure 8.9: An example of a ROC curve where a perfect network would create a triangle in the upper left corner. Figure from the public domain.

8.3 BEST: Boosted Event Shape Tagger

The Boosted Event Shape Tagger (BEST) [116] is a multiclassifier which classifies AK8 jets as originating from a top quark, Higgs boson, W boson, Z boson, bottom hadron, or QCD. The central idea of this method is that the jet constituents are boosted into a hypothetical parent particle's rest frame, where the daughter particles are expected to be most isotropic. A graphical representation of this idea is shown in Fig. 8.10.

In each frame, the boosted particle flow candidates are reclustered into new AK4 jets. Then, the particle flow candidates and reclustered jets are used to calculate a number of observable quantities which become the input features to BEST. The original concept of BEST was created in 2016 [116] and used in a CMS analysis for the first time in 2018 [43]. This section details the original BEST then future sections discuss studies done to improve upon this tagger and make it competitive with other boosted jet taggers at CMS.

8.3.1 Method

Preparation of the input features begins with boosting the jet constituents. First, the reconstructed jet axis is taken to be the boost axis. This boost axis is described with a four vector $p_\mu = (\vec{p}, E)$ and to perform the boost, the negative of the momentum vector

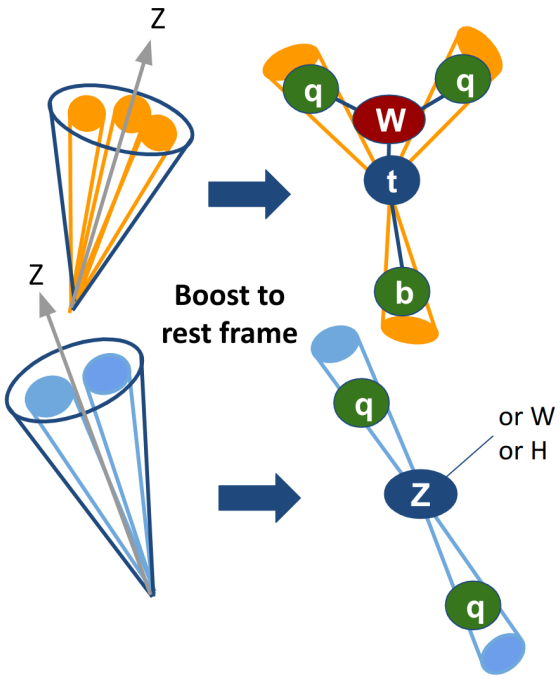

Figure 8.10: The left side shows the original AK8 jet and the right side shows the expected shape in the rest frame. The boost results in two rest frame jets for two-prong decays (like the Higgs, W, and Z bosons) and three rest frame jets for three-prong decays (like the top quark).

is used:

$$-\vec{p} = (-p_x, -p_y, -p_z)\,. \tag{8.8}$$

Then, to transform to the appropriate frame, a new energy is calculated using the jet momentum and the mass point corresponding to the desired rest frame

$$p_\mu = (-\vec{p}, m^2 + p^2)\,. \tag{8.9}$$

Four different rest frames are used based on hypothetical parent particles: Top frame ($m_t = 173.1\,\text{GeV}$), Higgs frame ($m_H = 125.1\,\text{GeV}$), Z frame ($m_Z = 91.2\,\text{GeV}$), and W frame ($m_W = 80.4\,\text{GeV}$). The next step is to define a new coordinate system after the boost

$$\begin{aligned}
\hat{z}' &= \hat{p}_{jet}\,, \\
\hat{x}' &= \hat{z}_{lab} \times \hat{z}'\,, \\
\hat{y}' &= -(\hat{x}' \times \hat{z}')\,,
\end{aligned}$$

where $\hat{z}_{lab}$ is the z axis along the beam line in the lab frame and $\hat{p}_{jet}$ is the direction of the jet axis which is boosted along. The standard azimuthal (ϕ) and polar (θ) angles are used with respect to the new axes. Finally, the particle flow jet constituents are reclustered using the anti-k_T algorithm (Section 7.2.1) with radius parameter R=0.4 for each boost.

In this method of creating the boost vector, the momentum is unaltered and the energy is recalculated based on the chosen mass. An alternative approach would be to leave the energy constant and alter the momentum based on the chosen mass. This alternative approach has not yet been investigated.

8.3.2 Input Features

The algorithm calculates Boosted Event Shape (BES) variables in each of the boosted frames. In total there are 59 variables, with 13 variables per each of the four boosted frames (Table 8.1) and 7 frame invariant variables (Table 8.2). Each input feature is explained in detail below.

Frame Dependent Input Features

Sphericity	Fox-Wolfram Moment H_1/H_0	m_{12}
Aplanarity	Fox-Wolfram Moment H_2/H_0	m_{13}
Thrust	Fox-Wolfram Moment H_3/H_0	m_{23}
Longitudinal Asymmetry	Fox-Wolfram Moment H_4/H_0	m_{1234}
Isotropy		

Table 8.1: For each frame, the inputs consist of shape measurements from the particle distribution and mass combinations of the four leading energy reclustered jets.

Sphericity and Aplanarity

The sphericity and aplanarity variables are calculated from the eigenvalues of the sphericity tensor

$$S_{\alpha\beta} = \frac{\Sigma_i p_i^\alpha p_i^\beta}{\Sigma_i |\vec{p}|^2}\,, \tag{8.10}$$

where i is a jet constituent, and α, β are the spatial coordinates x, y, z. The sphericity tensor has three eigenvalues which can be sorted as $\gamma_1 \leq \gamma_2 \leq \gamma_3$. Then, the sphericity is

$\frac{3}{2}(\gamma_2 + \gamma_3)$ and the aplanarity is simply $\frac{3}{2}\gamma_1$. Sphericity is a measure of the uniformity of the jet constituents. Sphericity is equivalent to $\sum p_T^2$ with respect to the sphericity axis. Aplanarity is a measure of how "dijet" like the event is, meaning how consistent the event is with having two regions of energy that travel in opposite directions.

Thrust

Thrust is the maximum value of

$$\max_{\vec{n}} \frac{\Sigma_i |\vec{p}_T^{\,i} \cdot \vec{n}|}{\Sigma_i p_T^i}\,, \tag{8.11}$$

where i is a jet constituent and $\vec{n}$ is the axis which maximizes the equation. Thrust is a measure of directionality of the jet; it has a value of 1.0 when all constituent momenta lie along or opposite to the thrust axis.

Isotropy

Isotropy measures how uniformly distributed a set of particles is. A value near 1 means that the set is uniformly distributed whereas a value near 0 means the set is linearly distributed.

$$\frac{1}{2\pi\Sigma_i p_T^i} \int_0^{2\pi} \sum_i |cos(\phi)p_x^i + sin(\phi)p_y^i| d\phi\,, \tag{8.12}$$

where i is a jet constituent.

Longitudinal Asymmetry

Longitudinal asymmetry uses the reclustered jets and is defined as

$$A_L = \frac{\Sigma_i p_z^i}{\Sigma_i p_T^i}\,, \tag{8.13}$$

where i is a reclustered jet. This measures the momentum in balance in the $\hat{z}'_{boost}$ direction. Ideally, the correct choice of boost will result in good momentum balance and yield a value near 0.

Fox-Wolfram Moments

Fox-Wolfram moments are rotationally invariant quantities based on spherical harmonics [124]. These moments describe the angular distribution of the jet constituents

$$H_l = \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j|}{S} P_l(cos(\phi_{i,j}))\,, \tag{8.14}$$

where i, j are jet constituents. The moments used as input features consist of H_1 to H_4 normalized by H_0. If the jet constituents were boosted such that their center of mass is perfectly stationary, then H_1 would be exactly zero. On the other hand, if all of the jet constituents were parallel, then H_1/H_0 would go to 1 [124].

Reclustered Jet Constituents

The four-vectors of the four highest p_T reclustered jets[2] are used to calculate four new four-vectors

$$J^\mu_{1i} = J^\mu_1 + J^\mu_i\,, \tag{8.15}$$

$$J^\mu_{1234} = J^\mu_1 + J^\mu_2 + J^\mu_3 + J^\mu_4\,, \tag{8.16}$$

where J^μ_i is the reclustered jet four-vector and $i = 2, 3, 4$ are the reclustered jets ranked in decreasing order of p_T. Then, the invariant masses of these four four-vectors are used as input features. When boosted to the correct frame, the new invariant masses should reconstruct to the mass of the heavy particle causing the jet.

Lab Frame Features

Jet SoftDrop Mass	Subjet b Discriminant	τ_{21}
Jet Charge	Subjet 1 b Discriminant	τ_{32}
	Maximum Subjet b Discriminant	

Table 8.2: These are lab frame input features that come from commonly utilized jet substructure techniques.

[2] p_T with respect to the new coordinate system.

N-Subjettiness

The n-subjettiness values provide a measure of the number of prongs. In other words, the number of separable regions of energy that likely come from a hard scatter (i.e. subjet-like). These subjet-like directions $\hat{n}_j$ are defined within the AK8 jet and are chosen to align with the dominate radiation direction using the k_T clustering algorithm. The n-subjettiness is defined as

$$\tau_N = \sum_i p_{Ti} \min\left\{\Delta R_{i,1}, \Delta R_{i,2}, \ldots, \Delta R_{i,N}\right\}^{\beta}, \tag{8.17}$$

where N refers to the number of light jet-like directions, i is a jet constituent, p_{Ti} is the transverse momentum of jet constituent i, $\Delta R_{i,N}$ is the distance in (η, ϕ) between the jet constituent i and the center of N, and β is an angular term.

The standard way to obtain discriminating power from n-subjettiness is to use the ratios $\tau_{21} = \tau_2 \ / \ \tau_1$ and $\tau_{32} = \tau_3 \ / \ \tau_2$. τ_{32} is closer to zero for top jets and closer to one for QCD jets while τ_{21} is closer to zero for two-prong jets (W, Z, H) and closer to one for QCD jets [125]. These are standard variables in jet tagging and were added to BEST for completeness.

Jet Properties

Two jet properties are used as input features: the soft drop mass and jet charge. The soft drop mass is the invariant mass of the jet after grooming with the soft drop algorithm (Section 7.3.1). The jet charge is a sum of p_T-weighted charges for all constituents i in a jet

$$\frac{1}{\Sigma_i [p_T(i)]^{0.6}} \times \sum_i q(i) [p_T(i)]^{0.6}\,. \tag{8.18}$$

This definition of jet charge is commonly used at CMS, alternative versions are discussed in detail in [126]. Ideally, the soft drop mass should be close to the mass of a heavy particle created in a hard scattering event. The jet charge helps in discriminating jets that originate from charged objects (W boson, top quarks, QCD) and jets that originate from neutral objects (Higgs and Z bosons).

b Discriminators

This version of BEST uses b-discriminator values from the combined secondary vertex (CSV) algorithm [127]. This algorithm is applied to the soft drop subjets and outputs a score from 0 to 1 on how likely it is that the subjet contains a secondary vertex from a b hadron.

8.3.3 Architecture, Training, and Performance

This version of BEST is a relatively simply dense neural network consisting of 59 input features, three hidden layers of 40 neurons, and 6 output categories. The hidden layers use the ReLU activation function and the output layers use the softmax activation function. This network was trained using the categorical cross entropy loss function so that each output category gives probability like results. All of this is done in Python using the "scikit-learn" package with the "MLPClassifier" module [128].

The network was trained on Monte Carlo simulated data of heavy resonances decaying to t, H, Z, W, b and of background QCD. The network was trained with 83,333 jets per category in the p_T range 500–1500 GeV. The resulting normalized confusion matrix is shown in Fig. 8.11. For more details on this version of BEST see [43, 39].

8.3.4 Limitations

The first and most notable limitation of the original BEST is the low amount of training statistics. The second is the use of an old version of "scikit-learn"; this is a powerful Python tool, but other application programming interfaces (APIs) such as Keras and PyTorch are more commonly used and better documented. The third limitation is that BEST is trained on high-level features and does not have access to the low-level particle flow features that deepAK8 and ParticleNet do.

The Jet Reclustering Problem

The final limitation is a bit subtle, it is about reclustering the jet constituents in the boosted frame (the jet reclustering problem). As seen in Section 7.8, jet clustering methods are based around points in rapidity-azimuth space. But in the boosted frame, rapidity

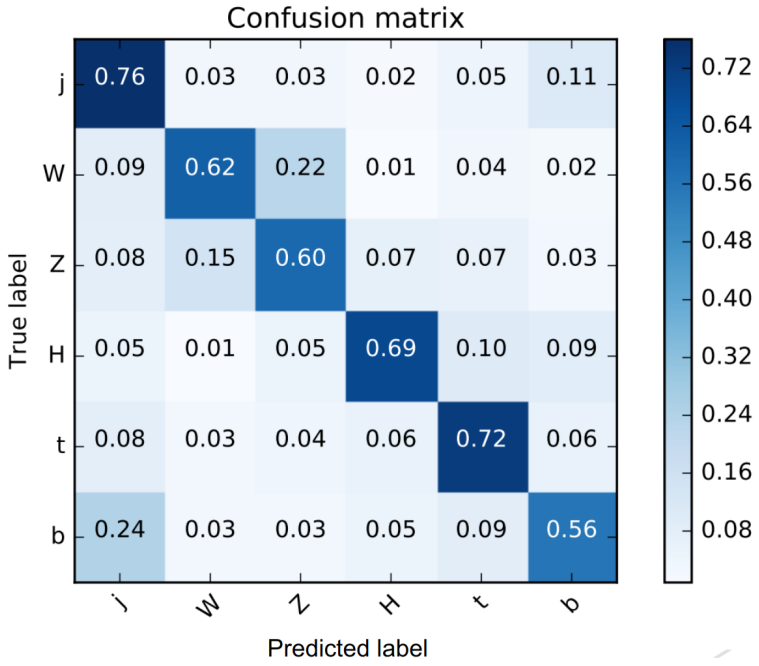

Figure 8.11: The confusion matrix for the version of BEST used in the analysis on 2016 data [43, 39]. This was evaluated on 2016 Monte Carlo simulated data. Here "j" stands for light jets (i.e. QCD jets).

is calculated with respect to the boost axis. The uninteresting candidates are near the z axis and the main candidates of interest are near the (x, y) plane. Thus, the uninteresting candidates will be clustered into multiple reclustered jets while the interesting ones will be clustered into two or three. Therefore it seems quite difficult to visualize this complicated boosted space with simply reclustered jets and event shape variables.

8.3.5 The Effort to Improve BEST

The first step of investigating improvements to BEST was to switch to using the Keras API with a TensorFlow backend [129, 130]. This provides access for a large library of deep learning tools. The next step was to increase the number of events for training. This was done by increasing the number of samples used in training. In order to avoid overlap in the training with the VLQ search, different Monte Carlo simulation samples were used in training than in the search presented in the subsequent chapters. Here, simply highly boosted jets were needed for each category. These were obtained by using hypothetical high mass resonances decaying to $\rightarrow HH$ (restricted to $H \rightarrow b\bar{b}$), $\rightarrow WW$, $\rightarrow ZZ$, $\rightarrow t\bar{t}$,

and $\rightarrow b\bar{b}$. For each of these samples, 21 resonant mass points were used (in GeV): 500, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000. For the QCD jets, p_T binned QCD Monte Carlo samples were used with $p_T > 470\,\text{GeV}$. Additionally, the AK8 jet p_T range was expanded to 500-2000 GeV.

A number of small improvements were made to the preprocessing of the data. Each category of samples was forced to have the same jet p_T spectrum. This was done by binning each sample into 20 GeV intervals, then assigning a probability to each jet. The number of selected jets per sample in each bin is equal to the sample with the minimum number of jets in that bin.

Then, each of the input features are independently standardized to aid in convergence of the network. In general, input features at small values (order 1) or features that are normally distributed at similar scales allow the network to converge faster. The resulting "scalar model" is created by a combination of the StandardScaler, MinMaxScaler, and MaxAbsScaler functions from the "scikit-learn" library for python [128].

The StandardScaler function moves the mean to zero and scales the feature distribution to unit variance. This is done for a feature x with

$$z = \frac{x - \mu}{\sigma}\,, \tag{8.19}$$

where μ is the mean of the samples and σ is the standard deviation of the samples.

The MinMaxScaler function translates and scales the input feature distribution to a given scale, in this case zero to one. This is done for a feature x by

$$x_{scale} = \frac{x - x_{min}}{x_{max} - x_{min}}\,, \tag{8.20}$$

where x_{max} and x_{min} are the maximum and minimum of the feature distribution respectively.

The MaxAbsScaler function translates and scales the input feature so that the maximum value of the absolute value of the feature distribution is one.

Only the training set is used to fit the scalar model. This is done to avoid "data leakage" from the validation and test sets. For this reason, data are preprocessed in the

following order:

1. Split into train, validation, and test sets
2. Match the p_T spectrum for each set
3. Fit the scalar model to the training set
4. Standardize all sets using the scalar model

If, instead, the scalar model was fit to data prior to splitting, then the sets would no longer be independent from one another. This alternative would allow the training set to learn subtle information from the validation and test sets through the scalar parameters. If a different scalar model was fit to each dataset, then the differences between the sets would cause inaccurate validation of the training and inaccurate evaluation of the network. Thus, it is crucial to ensure that the same scalar model from the training set is also applied to data when the network is used in a physics analysis.

Then a number of additional input features were added to try and increase the performance of the traditional BEST approach.

New Frame-Dependent Features

Improvements to the frame-dependent features focused on extracting more information from the reclustered jets. The first new input features are each element of the four-vector for the four highest p_T reclustered jets: $p_{xi}, p_{yi}, p_{zi}, E_i$. In Eq. 8.15 and 8.16, four four-vectors of reclustered jet combinations are created, the next input features are the cos of angle of each combination vector with respect to the new z. The last new frame dependent input features are $\cos(\Delta\theta_{ij})$ for the three highest p_T reclustered jets $i, j = 1, 2, 3$.

New Lab Frame Features

The lab frame features were also updated to extract as much information as possible. This included the addition of the first four N-jettiness variables τ_i, lab frame AK8 jet kinematics m, p_T, η, ϕ, and the number of secondary vertices within the AK8 jet cone.

Additionally, the b-tagging was updated to match the CMS recommendations first with deepCSV [131] and then with deepJet [105].

The reclustering problem, however, was much more difficult to address. This led to the creation of an entirely new idea called imageBEST that utilized a wide array of methods available from the deep learning field.

8.4 ImageBEST: An Image Based Version of BEST

Images are a powerful tool for visualizing complicated spaces and, with the developments in computer vision, images have become a powerful tool in deep learning. Jets can be represented as images where the pixel value corresponds to energy. Visually, this shows how energy is distributed in different types of jets. Such images also have discriminating power when used to train 2D convolutional networks. In the case of BEST, jet images are a way to get around the jet reclustering problem because each image shows the energy deposits in the space rather than reclustering candidates in an attempt to access the same information.

This section introduces the original method used at CMS to create jet images and then describes how it inspired and motivated me to create a method tailored to boosted frames that can be used with BEST. I wanted to incorporate these images into BEST to determine whether they would improve the tagging performance. The new method requires utilizing advanced CNN and RNN techniques, so I created an entirely new tagger called imageBEST. The last part of the section presents the results, comparisons, and conclusions from years of my work with imageBEST.

8.4.1 Classic Jet Images

In 2015, progress in the field of computer vision prompted the creation of jet images to be analyzed by CNNs [113]. The original techniques were refined to the technique in [114], which became the basis for imageTop at CMS [39]. This technique became the standard method for making jet images. Therefore, it is briefly summarized below.

First, jet constituent particle flow candidates are split into a 2D grid in (η, ϕ) space. The value of each bin in this 2D grid becomes the sum of the transverse energy of the

constituents within that bin, defined as

$$E_{Ti} = \frac{E_i}{\cosh(\eta_i)}\,, \tag{8.21}$$

for each constituent i. Then, the image is translated so that the leading Soft Drop subjet is at $(\eta, \phi) = (0, 0)$. The translations in ϕ are equivalent to rotations around the z axis and the translations in η are equivalent to a Lorentz boost along the z axis. The quantity E_T is invariant under Lorentz boosts (i.e. η translations), thus none of these translations affect the pixel intensity (bin value) [114].

Next, the image is rotated about $(\eta, \phi) = (0, 0)$ so that the second (subleading) Soft Drop subjet is at $-\pi/2$ (aligns with the vertical axis). If there is only one Soft Drop subjet, then the first principle component of pixel intensity is used instead. After the rotation, the bin values are recalculated since the original grid has been rotated. Then, the image is reflected so that the right side of the image has the highest sum of E_T. Finally, the bins are normalized by the squared sum $\sum I^2$ with I being the value in each bin [114]. An example of averages of images used in the imageTop tagger at CMS is shown in Fig. 8.12.

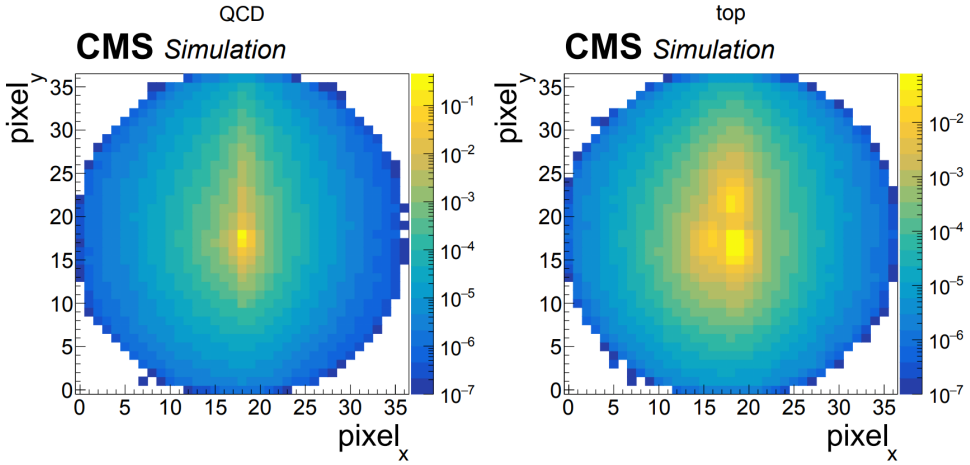

Figure 8.12: (Left) An averaged image of simulated QCD jets and (right) an averaged image of simulated Top Jets using the imageTop method [39]. Figure from [39].

8.4.2 The BEST Jet Image Method

The method of (η, ϕ) images works well in the lab frame, but not in the boosted frame. Most of the jet constituents of interest are perpendicular to the boost axis and as a result,

causing them to be smeared together in (η, ϕ) images. Therefore, a new image making technique was necessary.

The CNN identifies patterns in images, so it is crucial that these images be presented in a way that makes these patterns easily identifiable [113, 114, 112]. This is accomplished with a series of rotations in the rest frame. First, the leading candidate (highest energy jet constituent) is identified and the coordinate system is rotated twice so that the candidate is located at the z-axis. Next, a subleading candidate is identified. Theoretically there should be two separate areas of energy due to the two body decays. Thus, the subleading candidate is defined as the jet constituent with the highest energy in the region $\Delta\phi > 0.3$ away from the leading candidate. Then, a third rotation is performed about the new z-axis so that the subleading candidate is in the (x, z) plane. Each step in this process is shown in Fig. 8.13. After completing these rotations, the other jet constituents will be located throughout a sphere in this new coordinate system.

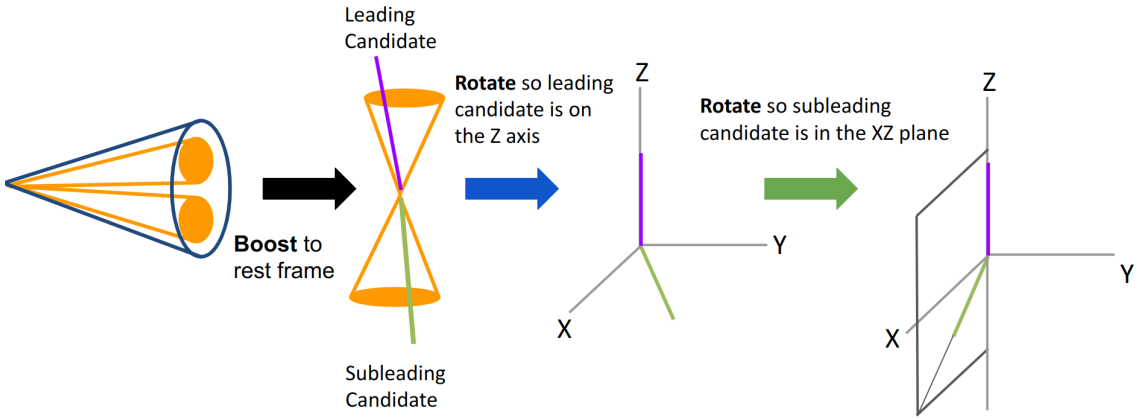

Figure 8.13: A graphical description of Lorentz boost and subsequent rotations necessary for the first attempt at making jet images for BEST.

The new sphere of jet constituents then needs to be projected onto a 2D rectangular grid to input properly into a 2D convolutional neural network. This is done by creating a two dimensional image with $\cos(\theta)$ (polar angle) on the y-axis, ϕ (azimuthal angle) on the x-axis, and color given by the energy contained within that portion of solid angle. The advantage of this approach is preserving solid angle between bins.

This method works well for two-prong jets by providing a consistent location in the

image for each region of energy. However, three-prong jets (e.g. top jets) have a third region of energy that will be smeared throughout the image. To prevent this, a series of reflections are implemented. First, the image is divided into top and bottom halves at $\cos(\theta) = 0$; the total energy in the top and bottom halves is calculated. If the energy in the bottom half is greater than the top, then the image is reflected about $\cos(\theta) = 0$. Next, the image is split into right and left halves at $\phi = 0$; the total energy in each half is again calculated. If the energy in the left half is greater than the energy in the right, then the image is reflected about $\phi = 0$.

Patterns in these images are visible when a large number of jets are averaged over. Averaged images of $\approx 10,000$ simulated top-jets are shown in each frame in Fig. 8.14. Notice that in the top-frame the top-jet has three distinct regions of energy. Whereas, in the other frames, the energy is smeared out in the upper right quadrant.

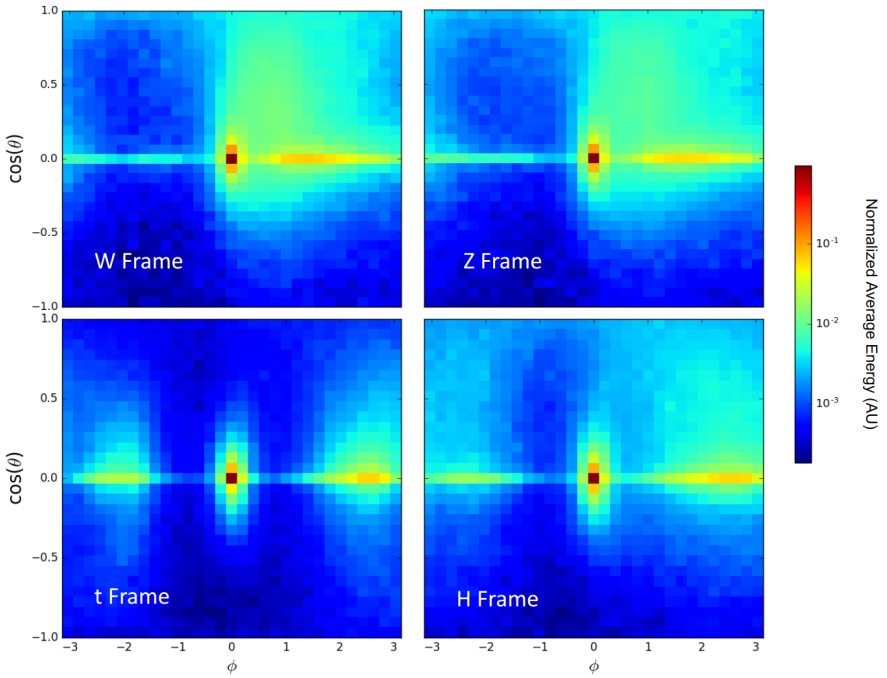

Figure 8.14: The first attempt of making images for BEST. These are averaged images of top-jets in each frame used for BEST. These images are the averages of over 10,000 jets originating from top quarks in CMS Monte Carlo simulated data.

However, there was a big problem with this method. Figure 8.15 shows the averaged images of $\approx 10,000$ simulated QCD-jets. These images show similar patterns to those of

the top-jet in the top frame making it difficult to train a network. This method forces symmetry into the images, even when it does not previously exist. Thus, a new image method was needed.

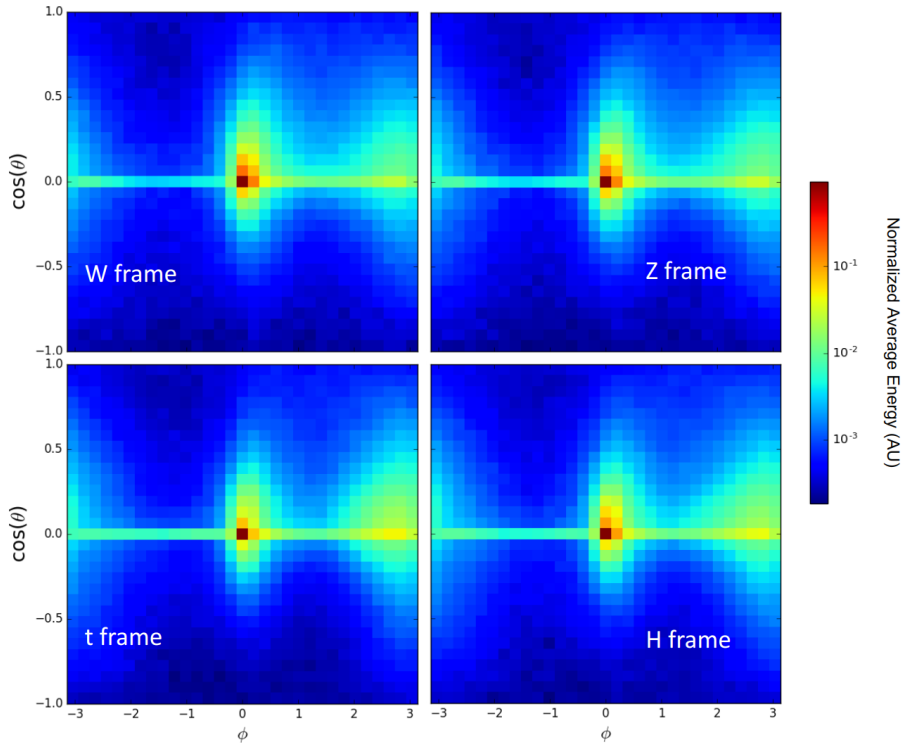

Figure 8.15: The first attempt of making images for BEST. These are averaged images of QCD-jets in each frame used for BEST. These images are the averages of over 10,000 jets originating from QCD in CMS Monte Carlo simulated data.

The central idea of the original BEST paper [116] is to exploit the symmetries that naturally arise in the correct rest frame of the jet—in other words, minimizing interventions to the frame. The rotations and reflections were abandoned to align the image method with the spirit of BEST. In this new method, the z-axis is the axis that the jet was boosted along (the jet axis in the lab frame). Then, no rotations are performed. The resulting sphere is again mapped to a 2D grid using $\cos(\theta)$ and ϕ. This new method is outlined in Fig. 8.16.

Again, there is a problem with this approach. When a jet is boosted to the wrong frame, the jet constituents should cluster near the positive and negative z-axis (the "poles" on a map). In the correct frame, jet constituents due to pile-up and the under-lying event

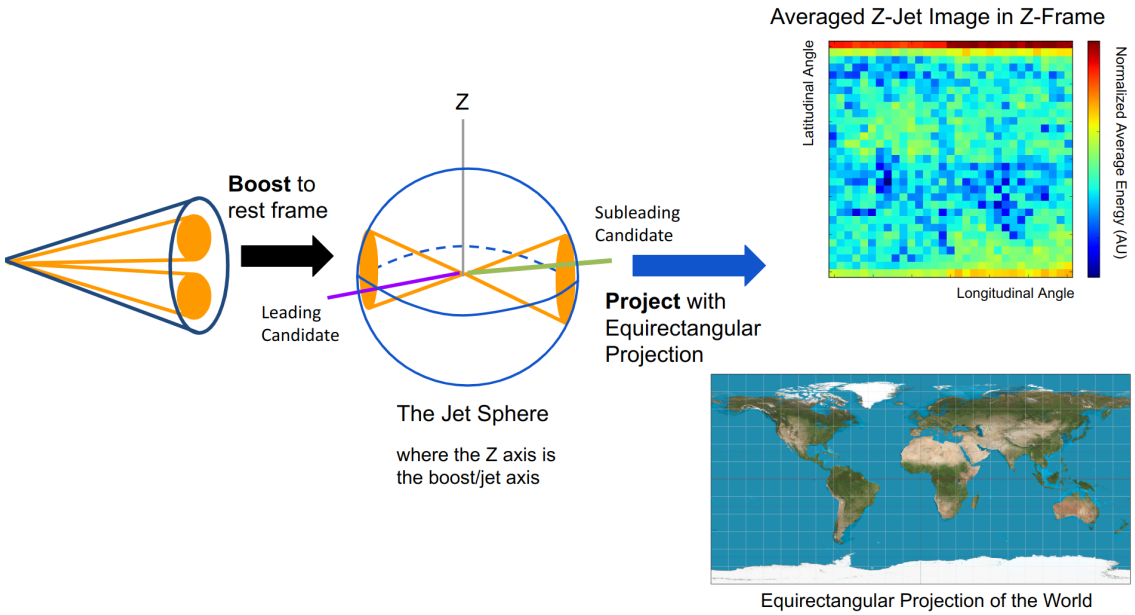

Figure 8.16: A graphical description of the modified image method inspired by the "BEST" approach with the boost axis as the z-axis. The image mapping is called the "equirectangular projection" in Cartography.

also cluster near the pole, leaving the characteristic constituents from the main decay process in the (x, y) plane (the "equator"). This method distorts the region near the poles causing the unimportant candidates to spread over a large area.

The mapping method in Fig. 8.16 is called the "equirectangular" projection in cartography. This is quite similar to the most commonly printed projection of the world, the "Mercator" projection[3]. The Mercator projection, like the equirectangular projection, distorts the relative size and shape of objects near the pole. The importance of this problem is nicely summarized in an episode of the American TV show "The West Wing" when a group of socially conscious cartographers advocate for a map projection which preserves relative size.

[3]The Mercator projection was designed by Flemish geographer Gerardus Mercator to make navigation by sea easier.

"It [the Mercator projection] distorts the relative size of nations and continents... What do maps have to do with social equality, you ask?"

"...In our society we unconsciously equate size with importance, and even power."

The West Wing

Much like for society, the relative size matters for training convolutional neural networks. By using the equirectangular projection, the network then assigns additional weight to these background candidates at the poles.

Thankfully, mapping a sphere onto rectangle is a problem that cartographers have been studying for centuries. In the case of BEST, a projection is needed that will confine the background candidates at the poles to a handful of bins while emphasizing the features at the equator. The projection chosen for this is the Cassini projection. Mathematically the projection is defined as

$$x = \arcsin(\cos(\theta - \frac{\pi}{2})\sin(\phi))\,,$$
$$y = \arctan 2(\tan(\theta - \frac{\pi}{2}), \cos(\phi))\,,$$

where $\arctan 2$ is the "2-argument arctangent" that returns a value β with range $-\pi < \beta \leq \pi$. The resulting method for creating images is illustrated in Fig. 8.17.

In the Cassini projection, the meridian becomes the equator. This causes a smearing of features along the equator, but leaves the poles constrained. This was crucial for allowing the CNN to pick up on features near the equator. The characteristic lobs of energy are shown in Fig. 8.18. In this image, the averaged Z-jet image has two lobs of energy near the equator which do not appear in the averaged QCD-jet image.

PUPPI weights are not applied to the Particle Flow candidates that make up the jet constituents. This idea was studied, however this creates very sparse images. When the CNN trains on these images, it is no longer resilient to the effects of pile-up or other backgrounds. Thus, when applied to data, the CNN will be unable to properly classify images where pile-up has bled through PUPPI and into the image.

CNNs train most efficiently on small pixel values; often images are normalized such that the pixel values are between zero and one. Initially, imageBEST used un-normalized

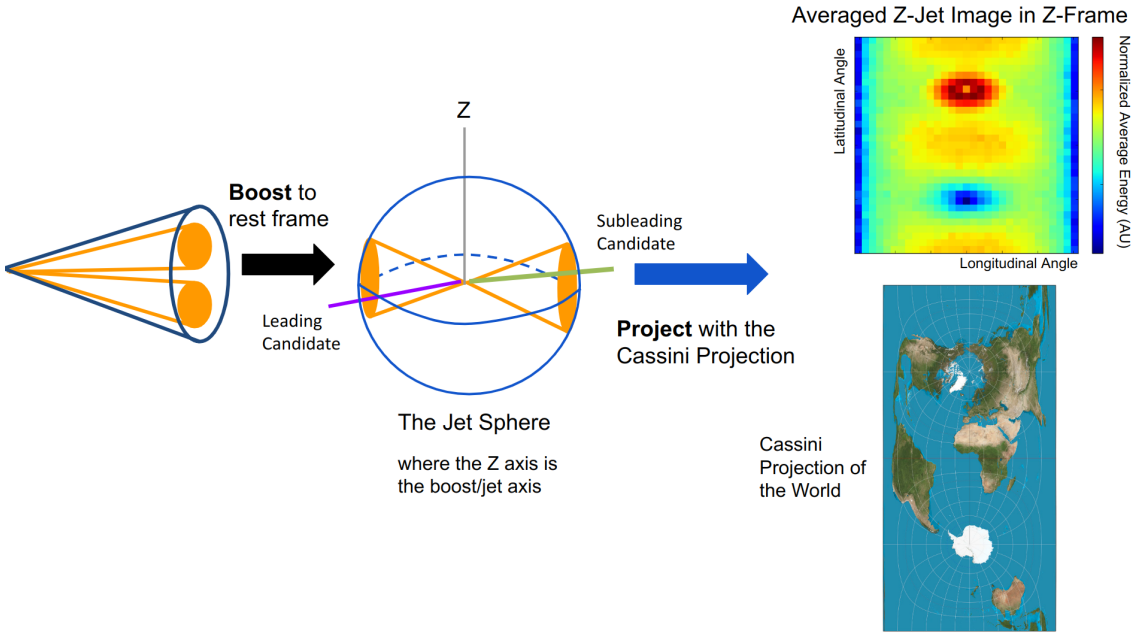

Figure 8.17: A graphical description of the BEST image making procedure using the Cassini projection.

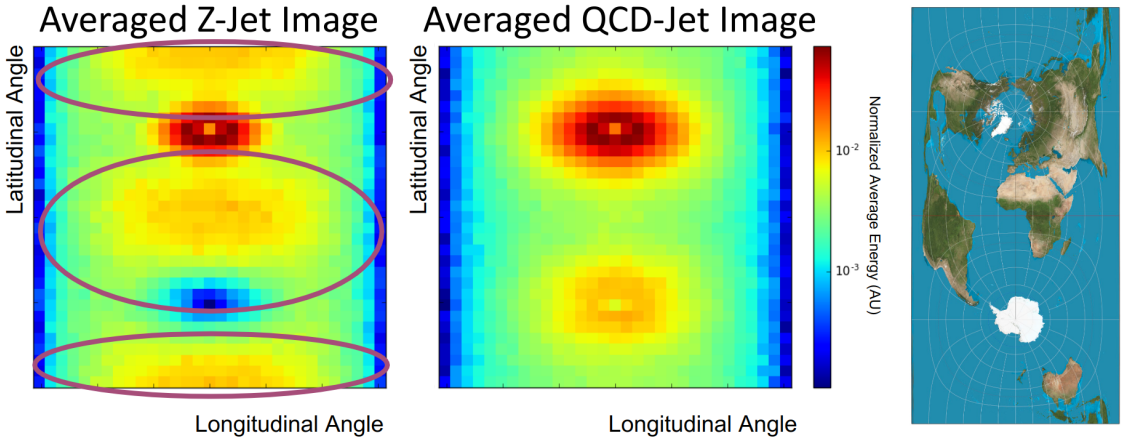

Figure 8.18: (Left) An averaged Z-jet image in the Z-frame with the characteristic lobs of energy circled. (Center) An averaged QCD-jet image in the Z-frame; note how it lacks the lobs of energy characteristic of a Z decay. (Right) A map of the world with the Cassini projection for reference.

pixel energy values. Then, the values were normalized by the total energy within the image. This yielded faster training and better accuracy. However, normalizing in this manner caused a majority of pixel values to be very close to zero. In an attempt to improve this, pixels were normalized by the highest energy particle flow candidate. This

showed a little improvement and next the images were normalized by the energy in the highest energy pixel. This yielded the best performance, but still a majority of pixel values were near zero. So one final study was made to scale the pixels using

$$\frac{2}{1+e^{-7x}} - 1\,, \tag{8.22}$$

where x is the pixel value normalized by the leading pixel energy. This did not yield any accuracy increase and the final version of the image method used normalization by the highest energy pixel.

An example of resulting images used to train "imageBEST" are shown in Fig. 8.19. This figure shows an average of over 100,000 simulated W-jets in each of the four BEST frames. A complete set of images for all BEST categories is included in Appendix B.

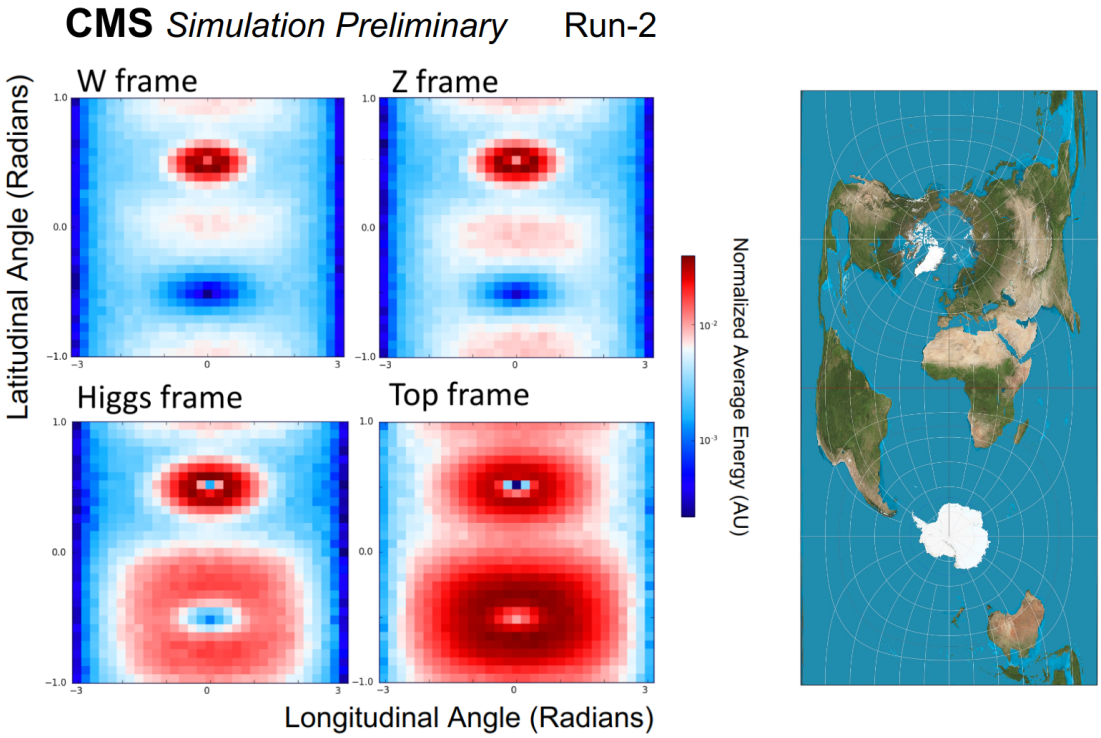

Figure 8.19: Averaged images of over 100,000 simulated W-jets created with the BEST image method. This figure contains an averaged image for each BEST rest frame.

8.4.3 ImageBEST Architecture

Before creating a full network to account for all the input features of BEST, smaller networks were created and tested on the images corresponding to a single rest frame.

This procedure allowed for the concentrated development of a single CNN structure that would be repeated for all frames. The original network consisted of three layers of 2D convolutional layers with batch normalization and max-pooling between each layer. The first layer utilized a patch size of (11,11). This large patch size is unusual in the field of computer vision, but [114] found that this large filter connects features in the sparse jet images. The other layers used a patch size of (2,2) before being flattened and output to a dense neural network for classification. The overall network structure is shown in Fig. 8.20.

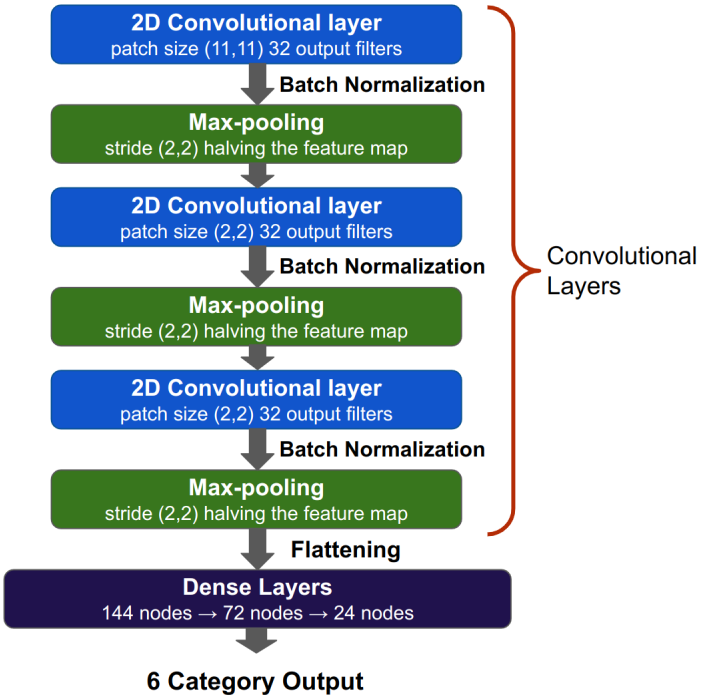

Figure 8.20: The first attempt for a CNN for BEST that was tested with a single rest frame image. This original network utilized 2D convolutional layers that caused an explosion of trainable parameters.

The usage of the 2D convolutional layers caused an explosion of trainable parameters in the network. This original structure proved to be difficult to train because of the large number of parameters. A common technique in computer vision to reduce the number of trainable parameters while still yielding high levels of accuracy is the depthwise-separable convolution [132].

A traditional 2D convolution utilizes a kernel (tensor of learned weights) with the

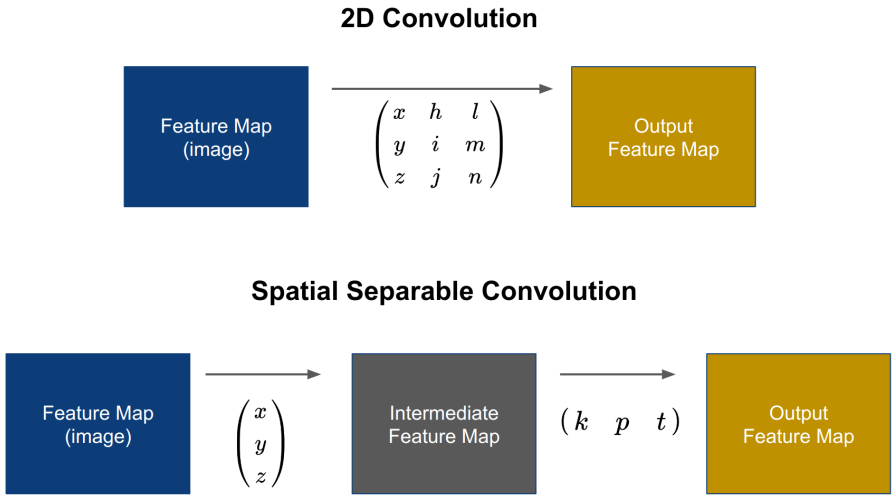

Figure 8.21: (Top) A 2D convolution with a single (3,3) kernel. (Bottom) A spatial separable convolution with a (3,1) kernel that creates an intermediate feature map and a (1,3) kernel that creates the output feature map. Inspired by [44].

height and width corresponding to the patch size. For a single filter, this kernel can be separated along the height and width into two kernels. Then two convolutions are applied (one with each kernel) creating an intermediate feature map. For a convolution with a single filter created by a (3,3) kernel, the kernel is separated into a (3,1) and a (1,3) kernel. First, the (3,1) kernel is applied to the input feature map, then the (1,3) kernel is applied to the intermediate feature map [42]. This operation is called a spatial separable convolution and is outlined in Fig. 8.21.

A depthwise-separable convolution performs an independent spatial separable convolution on each input channel. Then combines the channels with a (1,1) convolution over all channels; i.e. the depth is equal to the number of channels. Thus, separating the learning process of spatial and channel-wise input features. The result is 2D convolution with less operations and less trainable parameters. Depthwise-separable convolutions are faster to train, faster to use, and train more accurately on smaller datasets [42].

Replacing the 2D convolutions with depthwise-separable convolutions allowed for more complicated architectures in BEST. A number of different CNNs were tried during the process, including a modified version of AlexNet. However, the most successful architecture was again based off of the observations in [114]. The resulting architecture is shown

in Fig. 8.22. This architecture uses several convolutional layers with large patch sizes to scan the sparse jet images. The large number of convolutional layers was made possible by utilizing the depthwise-separable convolutions.

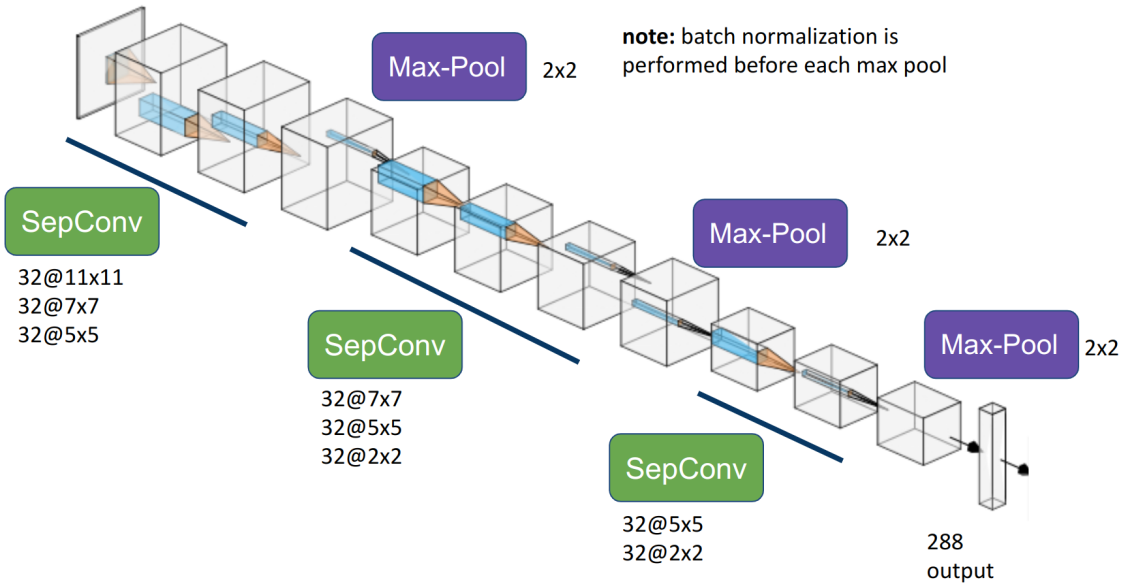

Figure 8.22: The CNN architecture used to process the images for a single frame. An output depth of 32 is used for each convolutional layer and the patch sizes in each layer are shown next to the output layer. This results in a flattened output with 288 features.

The next challenge was to combine the frames together into a single large network. One early attempt at this was to input the four frames together in one input feature map with (height=31, width=31, channels=4). However, this network performed similarly to a CNN trained with only one rest frame. In the process of optimizing the network, the four rest frames were split into four different CNNs each with the architecture shown in Fig. 8.22. Then, the outputs of each network were concatenated together. This yielded better performance.

The output of this new image network was combined together with the input features from BEST to create imageBEST. The resulting network is shown in Fig. 8.23. Unfortunately, this network had the same performance as BEST without the images. In this structure of imageBEST, there is no information shared between the networks even though these pictures are highly correlated. Each image is related by a Lorentz boost

creating a sequence, almost a small movie.

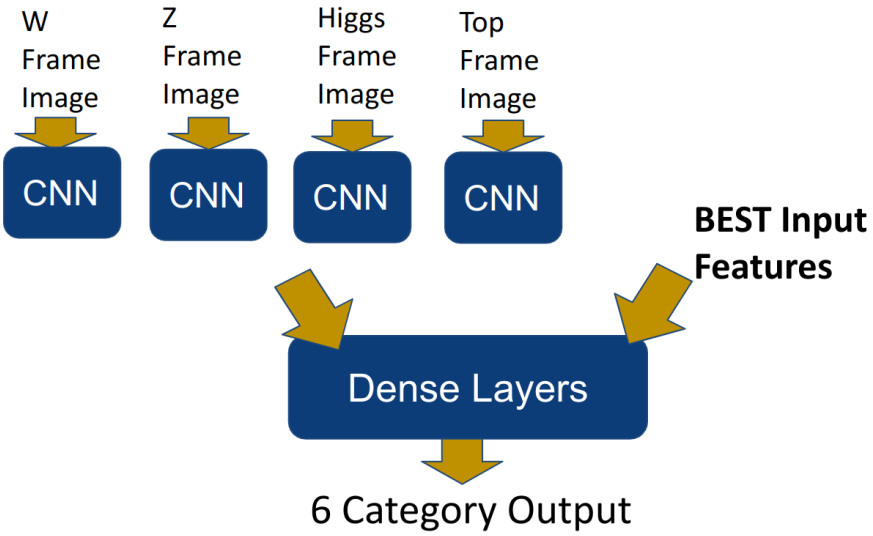

Figure 8.23: The first attempt at making an imageBEST. Each CNN consists of the network shown in Fig. 8.22. This network has the same performance as BEST without images.

In the subfield of video processing, separate frames of videos are treated like a sequence of images. Thus, a new architecture for the network was created inspired by this field. Here each CNN layer is merged in a time distributed layer and the output of all four is sent to an LSTM for sequence recognition. The time distributed layers force the weights to be shared between the four CNNs (rather than training separate weights for each network). This results in the CNN discriminating features that move between frames. These moving features create a sequence that is passed to an LSTM creating distinguishing features that are input into dense layers a long with the normal BEST input features. The resulting network structure is shown in Fig. 8.24. This network structure proved to be the most successful.

8.4.4 ImageBEST Training and Performance

The imageBEST network was trained on 300,000 jets per each category (top, Higgs, W, Z, bottom, QCD) using only 2017 Monte Carlo simulated data. The data were preprocessed and split using the method outlined in Section 8.3.5. During the training, the loss and accuracy are recorded for the training and validation sets. When the validation

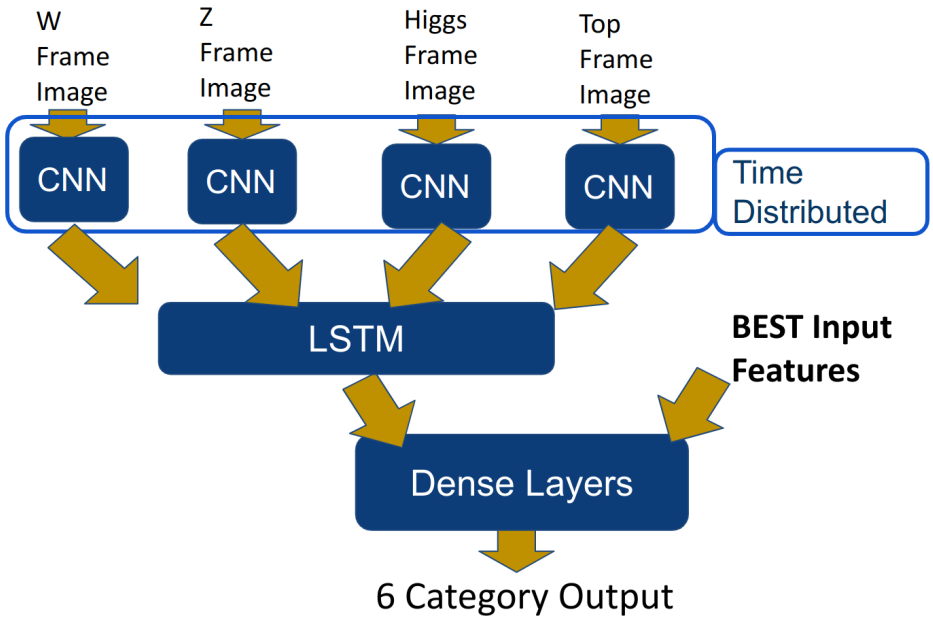

Figure 8.24: The imageBEST network. Each CNN consists of the network shown in Fig. 8.22 where the weights are shared among the four CNNs via time distributed layers.

accuracy shows no improvement for 10 epochs[4] then the training stops and the weights corresponding to the highest validation accuracy are restored. The metrics for the training of imageBEST are shown in Fig. 8.25.

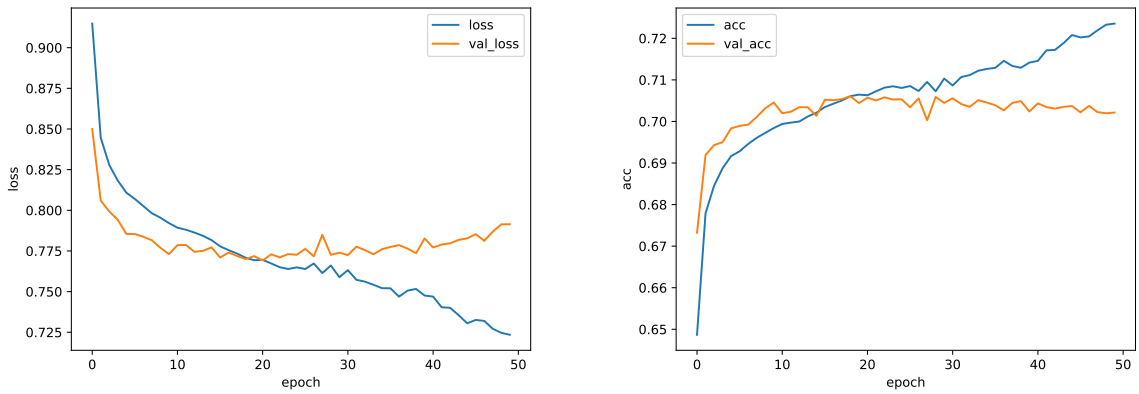

Figure 8.25: (Left) The loss evaluated on the training and validation sets and (right) the accuracy evaluated on the training and validation sets during the training of imageBEST.

[4]An epoch is when the entire training set has been used for backpropagation.

The resulting imageBEST network was evaluated on the testing set. The confusion matrix is shown in Fig. 8.26 and the ROC curves are used for comparisons in the next subsection.

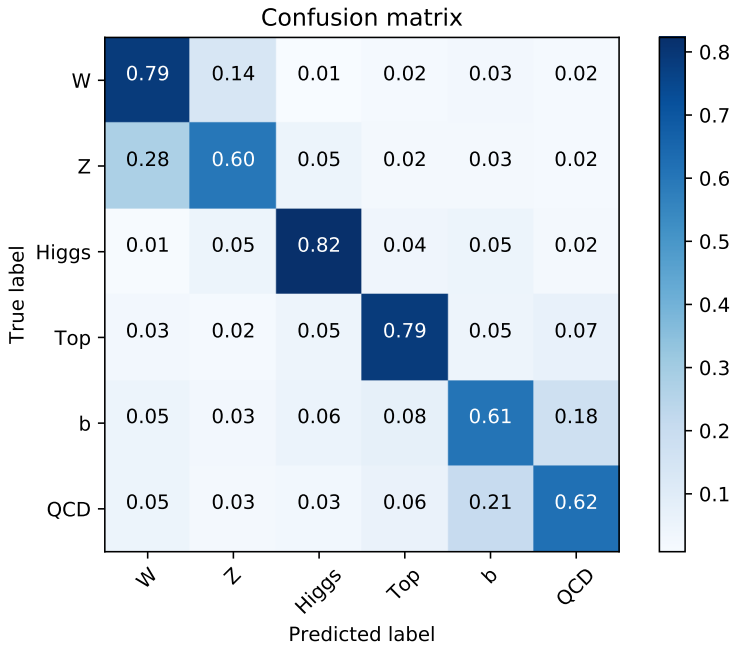

Figure 8.26: Confusion matrix for imageBEST showing the "true" particle on the y-axis and the "predicted" particle on the x-axis.

8.4.5 ImageBEST Comparisons and Conclusions

A series of networks were trained and evaluated to monitor and compare the versions of BEST. Unfortunately, it was no longer possible to directly use the version of BEST used in [43] due to a number of incompatibilities, so a new "2016-like" version was trained on the network architecture and features described in Section 8.3 with the updated p_T spectrum matching and standardization described in Section 8.3.5. A second, improved version of BEST was trained using the additional features described in Section 8.3.5. This network had four hidden layers, the first three with 256 neurons and the fourth with 144 neurons. The resulting network is simply labeled "BEST". A third network was trained only on the jet images. This "Image Only" network consisted of the same network architecture as imageBEST, but without the BEST input features. A fourth network was constructed from the point cloud transformer on the same datasets using the approach described in

[121] to train on particle flow candidates in the lab frame. This "PCT" was included to evaluate how BEST competes with graph based neural networks that utilize low level features.

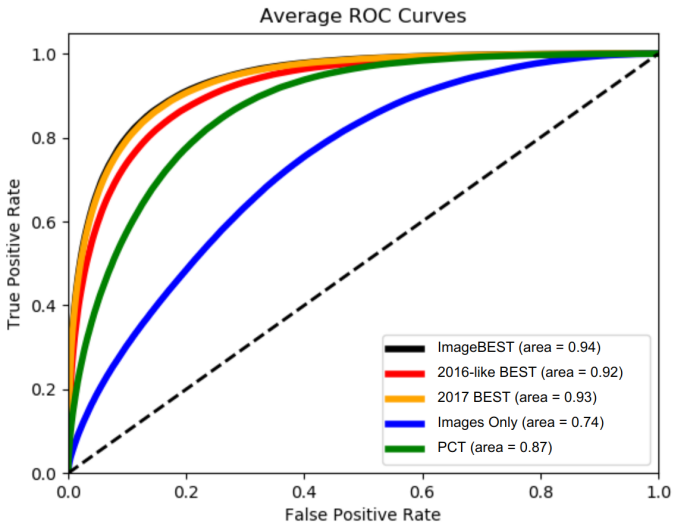

Figure 8.27: Macro averaged ROC curves imageBEST (black) is the best performing, but only very slightly better than BEST (orange) with a difference in area < 0.01.

The resulting macro average ROC curve comparison of all five networks is shown in Fig. 8.27. In this case the imageBEST network outperforms the others, but only very slightly with the difference in area < 0.01 between BEST and imageBEST. The two best performing categories for imageBEST are shown in Fig. 8.28.

In the end the performance increase from BEST to imageBEST was small. ImageBEST was the result of years of hard work with clever approaches to creating images and designing networks. ImageBEST combines the ability of CNNs to extract low level features with the high level features in BEST and leads to a profound conclusion—almost all of the information is captured by the high level features, the CNN is not adding much. Never underestimate the power of a simple network with physics-inspired inputs.

8.5 The New BEST

The underwhelming performance of imageBEST motivated a simple architecture for BEST consisting of high level input features with three layers of dense nodes. A series of addi-

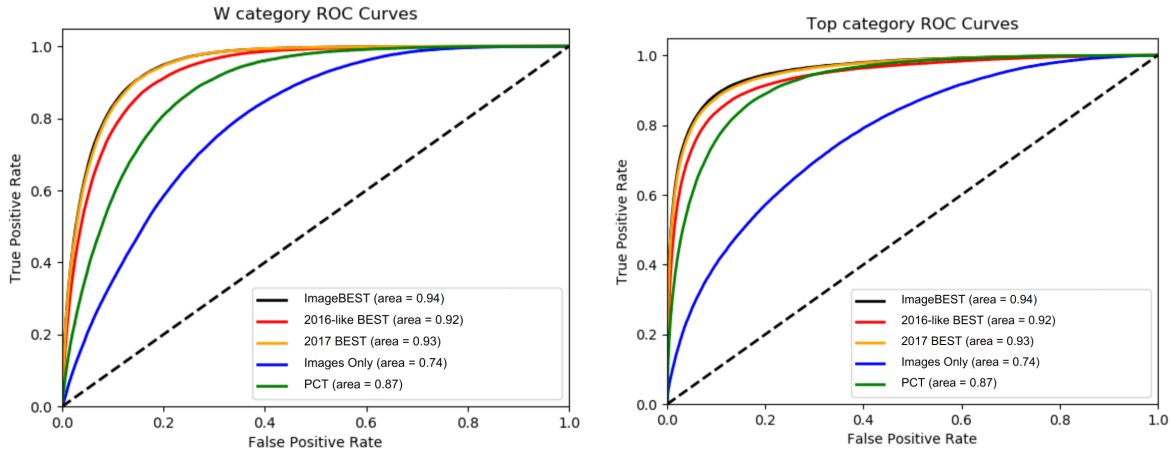

Figure 8.28: The two best performing categories for imageBEST (left) the W category and (right) the top category. In both cases imageBEST (black) is only slightly better than BEST (orange).

tional studies were done to confirm that the optimal rest frames were used[5].

This resulted in a change of frames; the new version of BEST utilizes seven different rest frames. The first three are based on hypothetical parent particles: Top frame (173.1 GeV), Higgs frame (125.1 GeV), and W frame (80.4 GeV). The other four were selected due to high discriminating power: 300 GeV frame, 400 GeV frame, Jet mass frame (lab frame jet mass with no corrections), and soft-drop mass frame (lab frame jet mass after soft-drop is applied). The input features described in Section 8.3.5 were used in each frame and a complete list of the frame dependent features is summarized in Table 8.3.

The lab frame input features include the new features and b-tagging updates described in Section 8.3.5. Additionally, Isotropy is frame invariant when the z-axis is the jet axis. Thus, only the top-frame Isotropy was used. Together, these encompass the single frame features of BEST and are summarized in Table 8.4.

8.5.1 Training Procedure

These boosted event shape variables are fed into a feedforward, fully connected set of dense layers. The architecture consists of: 271 input features x 140 node hidden layer x 140 node hidden layer x 140 node hidden layer x 6 node output layer. These 6 output nodes correspond to the 6 categories (top, Higgs, W, Z, bottom, QCD). The network

[5]These studies will be described in detail in Samantha Abbott's thesis.

Frame Dependent Input Features			
Sphericity	Aplanarity	Thrust	Longitudinal Asymmetry
FWM H_1/H_0	FWM H_2/H_0	FWM H_3/H_0	FWM H_4/H_0
p_{x_1}	p_{x_2}	p_{x_3}	p_{x_4}
p_{y_1}	p_{y_2}	p_{y_3}	p_{y_4}
p_{z_1}	p_{z_2}	p_{z_3}	p_{z_4}
E_1	E_2	E_3	E_4
m_{12}	m_{13}	m_{23}	m_{1234}
$cos_{12}(\theta)$	$cos_{13}(\theta)$	$cos_{23}(\theta)$	$cos_{1234}(\theta)$
$cos_{12}(\Delta\theta)$	$cos_{13}(\Delta\theta)$	$cos_{23}(\Delta\theta)$	

Table 8.3: List of frame dependent input features for training the Boosted Event Shape Tagger. For each frame, features such as the normalized Fox-Wolfram Moment (FWM) are examined, along with the four leading energy reclustered jets.

Single Frame Input Features		
Jet τ_1	Jet τ_2	Jet τ_{21}
Jet τ_3	Jet τ_4	Jet τ_{32}
Jet η	Jet ϕ	Jet p_T
Jet Mass	Jet Soft-Drop Mass	Jet Charge
Subjet 1 b Discriminant	Subjet 1 b-tag	Subjet 1 double b-tag
Subjet 2 b Discriminant	Subjet 2 b-tag	Subjet 2 double b-tag
Isotropy$_{top}$	Number of Secondary Vertices	

Table 8.4: List of single frame input features for training the Boosted Event Shape Tagger. All variables except for the top-frame Isotropy come from the lab frame. b tagging values provided by deepJet.

structure is shown in Fig. 8.29. Each tagging score is constrained to be between 0 and 1; a jet is classified as the category that has the largest score.

BEST is trained on 3×10^6 jets from each of the 6 categories, for 18×10^6 total training jets. The training is monitored by comparing performance to a statistically independent validation sample set of 375,000 jets per category, for 2.25×10^6 total validation jets. The

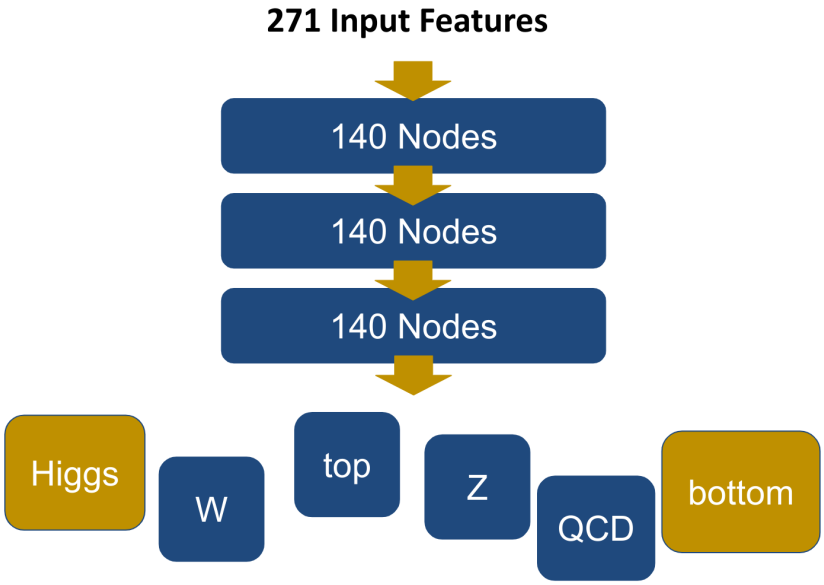

Figure 8.29: The structure used for the new version of BEST.

evaluation of the tagger after training is done on another data set, statistically independent from both the training and validation sets, but of equal size to the validation set. Each year of study (2016APV, 2016, 2017, and 2018) has a unique set of training, validation, and evaluation data. All data sets are split and standardized according to the procedure outlined in Section 8.3.5.

8.5.2 Performance

This improved version of BEST utilizes more input features, larger training samples, a better method of p_T shape matching, and a larger network. The resulting performance is shown in Fig. 8.30 with slight differences between each year. Note that the 2016 run is split into two data sets: 2016 and 2016 APV. 2016 APV refers to the LHC runs before the CMS tracker changed bias voltage values (increasing efficiency) in 2016. This version of BEST was used in the full Run 2 Vector-like quark search described in subsequent chapters.

8.6 Conclusion

At the CMS experiment, many advanced Machine Learning techniques are used to tag jets including convolutional, recurrent, and graph neural networks. BEST utilizes high level

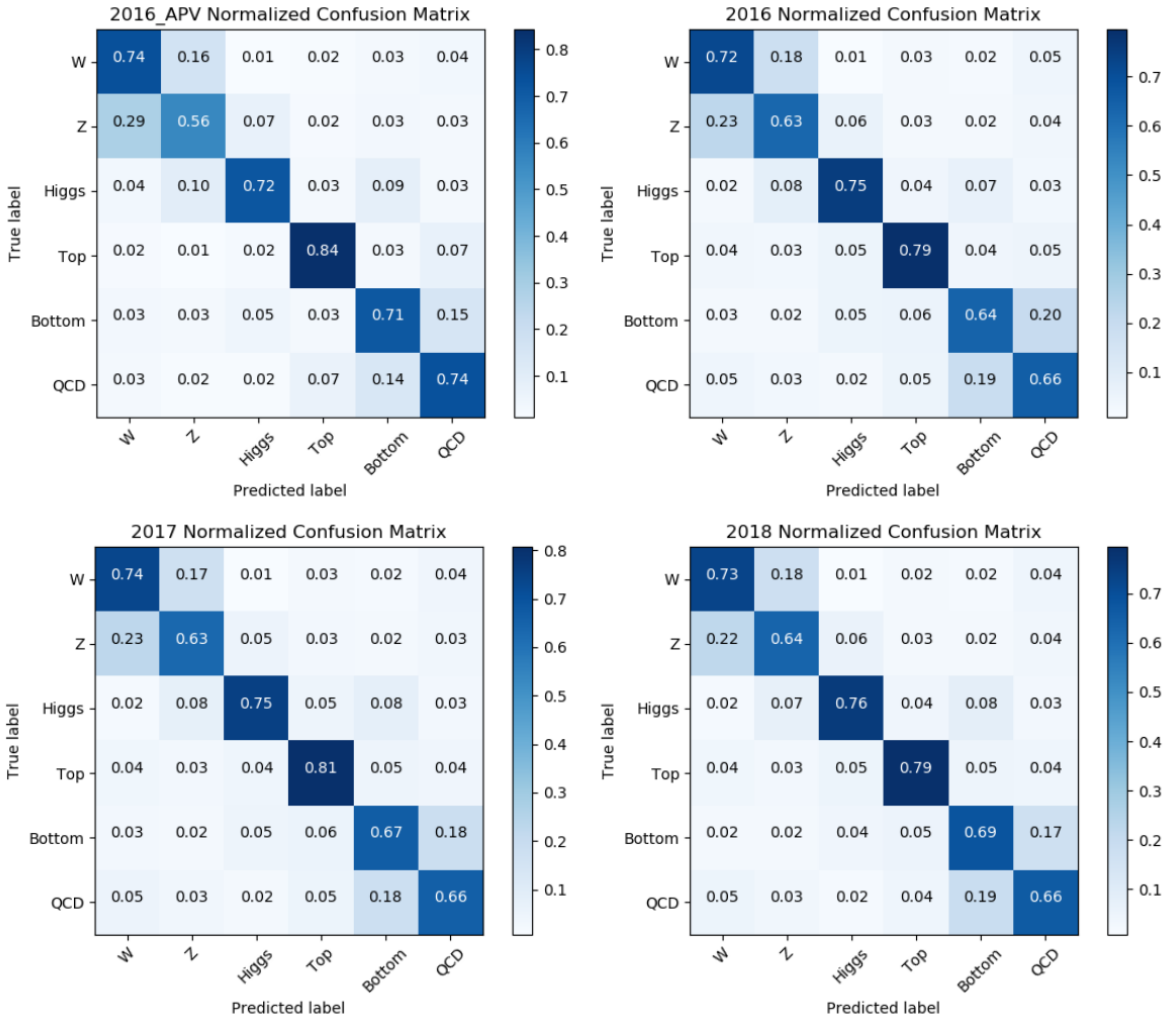

Figure 8.30: Confusion Matrices for BEST trained on simulation for each CMS data set during Run 2.

input features while many taggers aim to utilize low level features. Therefore, imageBEST was created to try and use both low and high level features. Surprisingly, imageBEST showed little improvement and the idea was abandoned. Ultimately, BEST is a simple neural network that outperformed many advanced techniques, highlighting the importance of clever physics-inspired input features.[6]

[6]This sentence was written on the top of Passo Spluga (Splügen Pass).

Chapter 9

Constructing a Search for Vector-like Quarks

≪Che cos'è per lei la neve?≫
Mi dice come sto, le cose su cui devo lavorare, la vivo e la uso per trovare me stessa. Un giorno le ho persino scritto una lettera. Quando non la sento sotto i piedi vuol dire che non sto sentendo nemmeno me stessa. La neve mi dà risposte, quelle che cerco e, soprattutto, quelle che non voglio sentire.

Sofia Goggia

The goal of any search for new physics is to design a set of selections to remove as many background processes as possible while retaining signal processes. In this case, the "signal" refers to the pair-production of vector-like quarks in the hadronic final state. This analysis relies heavily on the results of BEST in order to do this. Before applying BEST, a number of selections are used to ensure that events have high data quality, are orthogonal to other searches, and are free of common background processes.

After applying BEST, the events are categorized based on the output scores into 126 orthogonal regions (combinatorical result of 4 tags with 6 classes). This categorization is different from more traditional analyses, which focus on applying cuts to optimize signal over background. In each category, the figure of merit will be the total scalar sum of the jet transverse momenta (H_T). The selection and subsequent categorization and background

estimation are the topics of this chapter. I worked with a small, dedicated team on each of these topics.

9.1 Analysis Selection Criteria

The high mass of the VLQ produces significantly boosted third generation quarks and bosons, generating highly collimated jets with large p_T. For each VLQ we expect at least two large radius ($\Delta R = 0.8$) jets resulting in at least four massive objects for either $T\bar{T}$ or $B\bar{B}$ decays. Therefore, the signal regions require at least four AK8 jets and a three AK8 jet region is used for data driven background estimation. For events with more than four AK8 jets, only the four jets with the highest p_T are considered in this analysis.

9.1.1 Triggers

LHC Run 2 delivered proton collisions every 25 ns, so CMS had to develop specialized triggers for these experimental conditions. These can be broadly categorized into muon triggers and calorimeter triggers. Analyses using hadronic final states, like this one, use calorimeter triggers.

Identification of events with large amounts energy deposited in the calorimeter cells can be done using transverse momentum p_T or the total some of transverse energy H_T. Interactions due to soft scattering carry large amounts in the direction of the beamline, depositing very little energy in the transverse direction. Thus, events with large amounts of energy in the transverse direction are likely the result of a hard scatter. Therefore, triggering on the energy transverse to the beamline is a common choice. p_T triggers aim to identify events containing a jet likely originating from a boosted heavy object (such as W, Z, Higgs, or top). Whereas, H_T triggers aim to identify hard scattering events with large amounts of energy.

All of the triggers used in this analysis were developed for AK8 jets reconstructed with the Particle Flow algorithm. During proton collisions, it is possible for one trigger to have short periods a high rate preventing all of the events from being recorded in that path. This inefficiency can be overcome by combining multiple triggers together. In the 2016 run, three triggers are combined to maximize the efficiency. This includes two H_T

triggers (at 800 GeV and 900 GeV) and one p_T trigger (at 400 GeV). The pile-up at the LHC increased after 2016, resulting in higher trigger rates. The trigger menu changed to reflect this for the 2017 and 2018 runs. As a result, two new triggers were chosen to maintain similar levels of efficiency. In this case, one H_T trigger (at 1050 GeV) and one p_T trigger (at 400 GeV). These triggers are summarized in Table 9.1.

Year	**Trigger**
2016	HLT_AK8PFHT800
	or HLT_AK8PFHT900
	or HLT_AK8PFJet400
2017, 2018	HLT_AK8PFHT1050
	or HLT_AK8PFJet500

Table 9.1: The High Level Triggers used in each year of LHC Run 2.

The efficiency of each trigger was studied to ensure optimal performance. This is done on an orthogonal trigger data set; in this case, the single muon data set filled with events containing an isolated muon with $p_T > 27$ GeV (IsoMu27). The trigger efficiency is defined as

$$\epsilon = \frac{\text{\# events passed (trigger + selection)}}{\text{\# events passed (selection)}}, \tag{9.1}$$

where the selection is kept as minimal as possible. In this case, the selection only requires that there four or more AK8 jets with $p_T > 50$ GeV, $|\eta| < 2.4$, and that each jet has two subjets identified with the soft drop algorithm. The performance of the triggers is shown for each year by H_T in Fig. 9.1. The jet selections described in the next section require a minimum H_T of 1200 GeV, which is well within the fully efficient region.

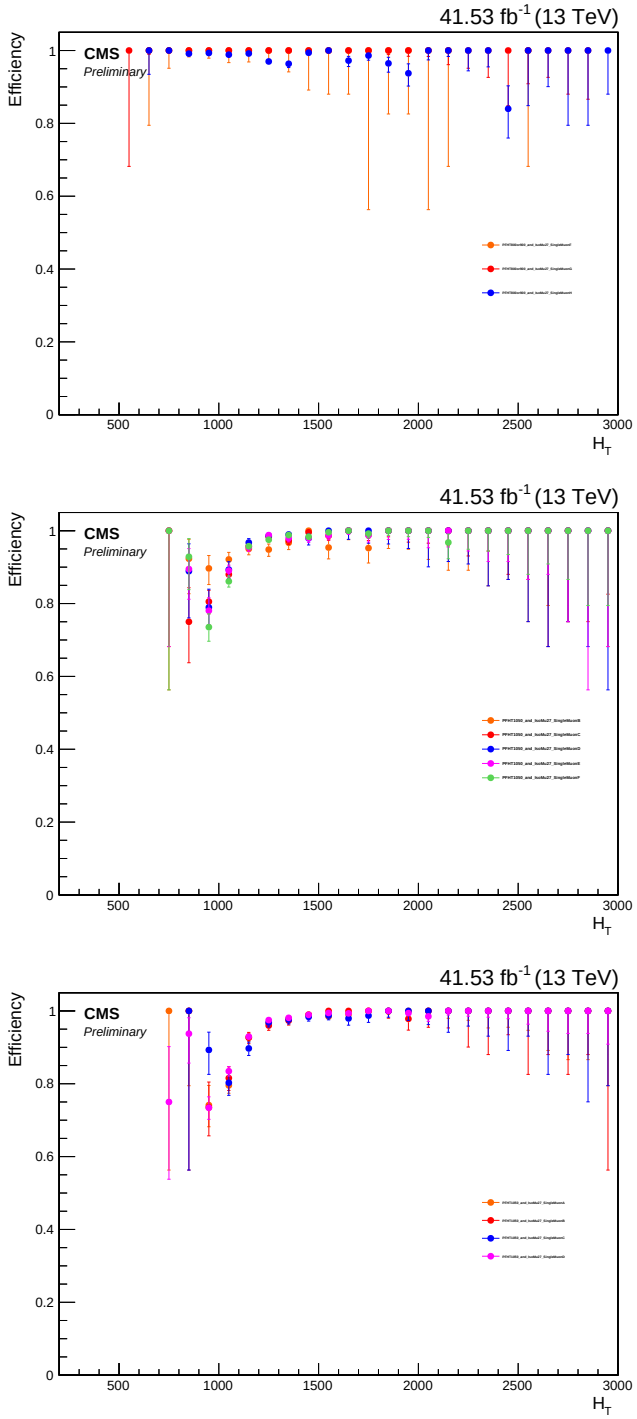

Figure 9.1: Trigger efficiencies for the jet triggers evaluated on the single muon data set. One efficiency curve per run year: 2016 (top), 2017 (center), 2018 (bottom).

9.1.2 Jet Selection

The selection of this analysis focuses on choosing jets that can be used with BEST, which requires several steps. The first set of requirements is to ensure that jets are well reconstructed. For this, jets are required to be inside the CMS tracker to fully utilize tracking information. This is enforced with a pseudorapidity cut of $|\eta| < 2.4$.

Then, the jets must pass ID requirements established by the CMS Jet MET group. For this, each jet must be clustered from particle flow candidates with the PUPPI algorithm applied. The clustering must be done using the anti-k_T algorithm with radius 0.8 and the soft drop algorithm must be applied to the resulting jet. The fraction of energy from various types of constituents must come largely from charged hadrons. This is enforced with a neutral hadron fraction < 0.90, charged hadron fraction > 0.0, charged Electromagnetic (electrons) fraction < 0.8, neutral electromagnetic (photons) fraction < 0.9, and muon fraction < 0.8. Each jet must consist of more than one constituent and have a charge multiplicity greater than one. A jet passing these requirements is considered a jet which passes the "ID" requirements.

Several requirements are specific to BEST. BEST requires two soft drop subjets and at least four reclustered jets (to sample the invariant mass of the combinations). The lightest expected AK8 jet originates from a b hadron decay, therefore a soft drop mass cut of $m_{SD} > 0.25\,\text{GeV}$ is applied. This soft drop mass cut is lower than many other analyses, but prevents a drop in signal efficiency.

All jets considered in this analysis have passed these cuts. Then, additional event level selections are applied to ensure that this search can fit into a broader CMS framework.

9.1.3 Additional selections

The CMS experiment aims to combine all searches for pair-production of vector-like quarks together to set broader exclusion limits. In order to do this, each search must occupy an orthogonal region of phase space. For the all hadronic final state, this is done by applying a lepton veto to ensure the search is orthogonal to leptonic final states.

Orthogonality to the single lepton final states is done by ensuring that events do not contain a tight muon with $p_T > 55.0\,\text{GeV}$ nor a tight electron $p_T > 55.0\,\text{GeV}$. A tight

electron is defined as a electron passing "mvaEleID-Fall17-noIso-V2-wp90" and a tight muon is defined by the CMS muon POG using the results in [133].

The multi-lepton final states require the presence of at least two loose leptons (electrons or muons) where both have $|\eta| < 2.4$, the highest energy lepton has $p_T > 40.0\,\text{GeV}$, and the second highest energy lepton has $p_T > 30.0\,\text{GeV}$. Loose electrons are defined as electrons passing a specialized election filter called "mvaEleID-Fall17-noIso-V2-wpLoose" and loose muons are defined by the CMS muon group using the results in [133]. If two such leptons are identified, then the event is excluded.

9.1.4 Analysis Selection Summary

Overall, events with three passing AK8 jets are used in the control region and events with four or more passing AK8 jets are used in the signal regions. In the signal region, each of the four highest p_T jets can either be classified as Higgs, top, bottom, W, Z, or QCD. That is four jets with six possibilities leading to 126 orthogonal regions. For example: WWWZ, tHWZ, bHWZ, etc. six of these regions are signal deficient and used for validation (the 3QCD+1 categories), the other 120 are used as signal regions. A complete summary of the analysis selections is shown in Table 9.2.

9.2 Background Estimation

The signal regions of this search are dominated by QCD background which is estimated with a Data Driven approach. The other backgrounds are estimated with Monte Carlo simulated data.

9.2.1 Monte Carlo Background Estimate

The number of expected Non-QCD background events is estimated as the number of Monte Carlo simulated events that fall in the signal region. The Non-QCD background processes are listed in table 9.3 along with their corresponding MC generators. (CITE MC Generators)

	Category	Requirement
Jets	Jet ID	neutral hadron fraction < 0.9,
		charged hadron fraction > 0.0,
		neutral EM fraction < 0.9,
		charged EM fraction < 0.8,
		muon fraction < 0.8,
		jet constituents > 1,
		charge multiplicity > 0.0
	Jet	$p_T > 400\,\text{GeV}$, $\lvert\eta\rvert < 2.4$, $m_{SD} > 0.25\,\text{GeV}$
	Selections	# SoftDrop Subjets = 2, # reclustered jets ≥ 4
Analysis	Single	No "tight" electrons with $p_T > 55.0\,\text{GeV}$
	Lepton Veto	"tight" electrons pass mvaEleID-Fall17-noIso-V2-wp90
		No "tight" muons with $p_T > 55.0\,\text{GeV}$
	Multi	Does not contain two "loose" leptons where
	Lepton Veto	one has $p_T > 40.0\,\text{GeV}$ and the other $p_T > 30.0\,\text{GeV}$
		"loose" electrons pass mvaEleID-Fall17-noIso-V2-wpLoose
	Event	Number of jets ≥ 4 AK8 Jets
	Selections	(=3 AK8 Jets for control region)

Table 9.2: Summary of the selections used in the search for pair production of vector-like quarks in an all hadronic final state.

9.2.2 Data Driven Background Estimation

QCD is not well modeled in Monte Carlo simulated data, therefore the QCD background is estimated with a data driven approach. This is done by selecting a region dominated by QCD events but depleted of signal events. The four AK8 jet signal is avoided in a QCD rich three AK8 jet control region. In this QCD control region, BEST classifies each jet as Higgs, W, Z, top, bottom, or QCD. Since this a QCD dominated region, QCD

$t\bar{t}$	Powheg + Pythia 8
W + jets	MadGraph aMC@NLO + Pythia 8
Z + jets	MadGraph aMC@NLO + Pythia 8
WW	Pythia 8
ZZ	Pythia 8
WZ	Pythia 8
$t\bar{t}$W	MadGraph aMC@NLO + Pythia 8
$t\bar{t}Z$	MadGraph aMC@NLO + Pythia 8
$t\bar{t}t\bar{t}$	MadGraph aMC@NLO + Pythia 8
$H \to b\bar{b}$	MadGraph aMC@NLO + Pythia 8
$H \to b\bar{b}$	Powheg + Pythia 8

Table 9.3: The non-QCD background processes included in the background estimate and the correspond Monte Carlo generators.

tags should dominate. However, the tagger is not fully efficient, so not all of the jets are QCD tagged. Thus, in this region, a new quantity is measured, the classification fraction defined as a function of p_T

$$\epsilon_X(p_T) = \frac{N_X}{N}\,, \tag{9.2}$$

where N_X is the number of jets classified as X (Higgs, W, Z, top, bottom, QCD) and N is the total number of jets. The classification fraction shows the fraction of jets correctly tagged as QCD and the fraction mistagged as Higgs, W, Z, top, or bottom. The classification fraction in the three jet control region is shown in Fig. 9.2.

The next step of estimating the background in the four AK8 jet signal region is to multiply the classification fractions of the four AK8 jets together. Each signal region has multiple possible permutations; for example in WWWZ, the first jet might be Z or the second jet might be Z and so on creating four possible permutations (ZWWW, WZWW, WWZW, WWWZ) which all belong to the same signal region. Thus, the product is summed over the possible permutations to create an event weight. Finally, the event weight is summed over the number of events to create a background estimate. Mathe-

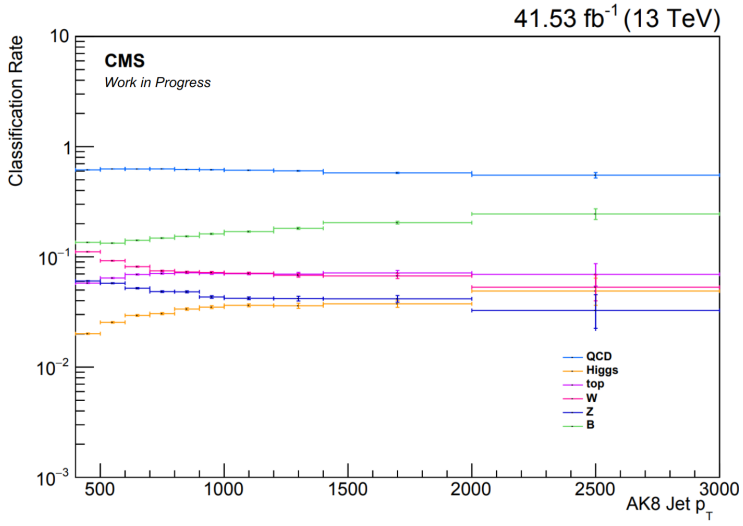

Figure 9.2: The classification fraction with respect to jet p_T in the three jet control region. In this region, it is assumed that all of the jets are QCD. The fraction tagged as QCD is the fraction of jets correctly tagged and the other fractions show the fraction mistagged as Higgs, W, Z, top, or bottom.

matically, this is expressed as

$$r = \sum_{events} \{ \sum_{permutations} [\prod_{i=1}^{4} \epsilon_{X_i}(p_T(i))] \} . \tag{9.3}$$

The estimated background in each of the 120 signal regions is shown in Fig. 9.3. During the background estimation, it is assumed that the control region contains only QCD events—creating an overestimation in the signal regions. However, this overestimate is adjusted using six nuisance parameters to account for non-QCD contributions in the control region. This process is detailed in the data driven background uncertainty description in section 10.2.

The H_T distributions of observed data, background estimate, and signal hypotheses are measured in each of the 120 regions. Then, these distributions are used for setting exclusion limits. The statistical methods and resulting expected exclusion limits are the subject of the next chapter.

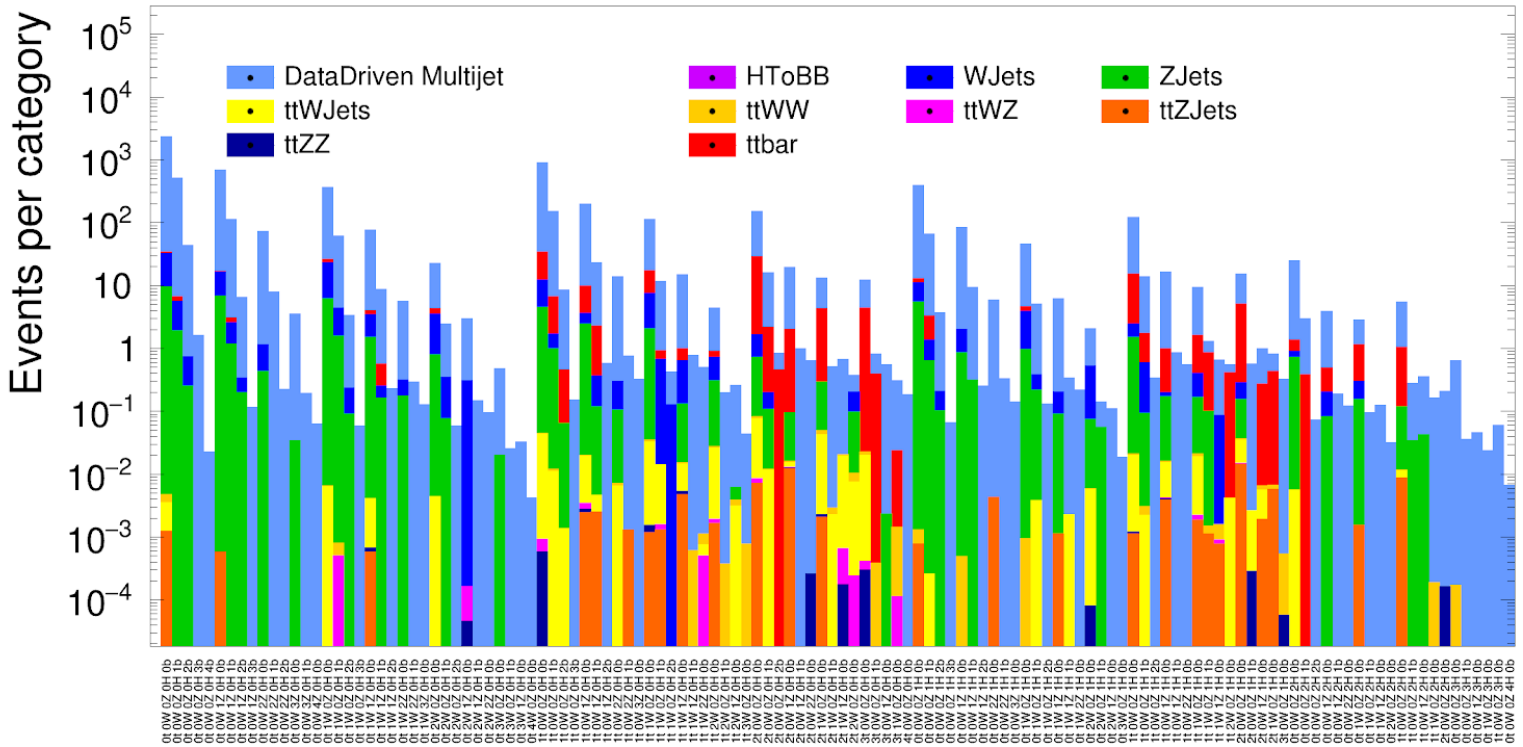

Figure 9.3: The estimated background in each of the 120 signal regions. Each bin represents one signal region labeled by the classification of the four jets (ex. WWWZ, WWHZ, etc.).

Chapter 10

Statistical Analysis

L'homme a toujours eu besoin de se confronter à des choses qui le dépassent. C'est en sortant de sa zone de confort qu'on apprend.

Thomas Pesquet

The final step in a search for new physics is to apply methods from the vast field of statistics to the results from the search. These methods test the agreement of collision data with the Standard Model given an alternative hypothesis (in this case vector-like quarks). If the data largely deviate from the Standard Model prediction, then the null hypothesis (the Standard Model only) can be rejected in support of an alternative hypothesis (vector-like quarks). If no deviation is observed, then exclusion limits are set on the alternative hypothesis.

This chapter describes the process of setting exclusion limits using the H_T bin results for all 120 signal regions. To accomplished this, I worked with a small team to perform a maximum likelihood fit where all sources of systematic uncertainty are described by carefully constructed nuisance parameters. Together, we set exclusion limits with an LHC statistical strategy called the CLs method. Before unblinding the signal regions, we performed a number of statistical tests in validation regions to ensure the health of the fit model. In addition to the CLs method, this chapter highlights six nuisance parameters that I motivated and constructed to account for overestimation in the data driven background estimate.

10.1 Limit Setting Method

In 2011, the ATLAS and CMS experiments at the LHC agreed on a common statistical strategy to search for the Higgs boson. The CLs method[1], chosen by the LHC Higgs combination group, allows for the combination of individual search results making it a valuable tool that continues to be used in searches for new physics [135, 136]. The process presented in this section is the result of decades of statistical results and begins with the construction of a likelihood function.

10.1.1 Maximum Likelihood

Maximum likelihood is a parameter estimation of the signal strength and background normalization. This technique utilizes a set x of measured quantities that depend on parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$. The maximum likelihood estimator is the $\boldsymbol{\theta}$ that maximizes the likelihood function $L(\boldsymbol{\theta})$. Algorithmically, it is much easier to minimize a function than maximize. Therefore, the maximum likelihood estimator is obtained by minimizing the negative log of the likelihood $-\ln L(\boldsymbol{\theta})$ which yields the same results as maximizing the likelihood [74, 12].

The likelihood function is defined as

$$L(\boldsymbol{\theta}) = \prod_i f(x_i; \boldsymbol{\theta}) \,, \tag{10.1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_m)$ is a set of statistically independent measurements with probability density function $f(x_i; \boldsymbol{\theta})$. In probability theory, as the number of trails N becomes large, discrete data become Poisson-distributed [137]. Thus, in high energy physics the data is assumed to follow a Poisson distribution $\mathcal{P}$ with mean λ_i so the likelihood becomes

$$L(\boldsymbol{\theta}) = \prod_i \mathcal{P}(n_i|\lambda_i) = \prod_i \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i} \,. \tag{10.2}$$

λ can parameterized as

$$\lambda = \mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta}) \,, \tag{10.3}$$

[1]The CLs method was originally developed by Alex Read as part of the LEP2 Higgs working group [134].

where $s(\boldsymbol{\theta})$ is signal yield, $b(\boldsymbol{\theta})$ is the background yield, and μ is an additional fit parameter. In a search for physics beyond the Standard Model, the background distribution is the Standard Model cross section σ_{sm} times the integrated luminosity L_{int} and the signal distribution is the cross section of the new theoretical process σ_{th} times L_{int}

$$\begin{aligned} b(\boldsymbol{\theta}) =& \sigma_{sm}\mathrm{L}_{int}\,, \\ s(\boldsymbol{\theta}) =& \sigma_{th}\mathrm{L}_{int}\,. \end{aligned}$$

For this reason, μ is defined as ratio of the measured cross section σ and the theoretical cross section σ_{th}

$$\mu = \frac{\sigma}{\sigma_{th}}\,. \tag{10.4}$$

μ is called the signal strength.

The likelihood used in the CLs method is

$$L(\mu, \boldsymbol{\theta}) = \prod_j \mathcal{P}(data_j|\mu \cdot s_j(\boldsymbol{\theta}) + b_j(\boldsymbol{\theta})) \cdot \prod_i \rho(\tilde{\theta}_i|\theta_i)\,, \tag{10.5}$$

where $\tilde{\theta}_i$ is the nominal value of the parameter and $\rho(\tilde{\theta}_i|\theta_i)$ is the probability density function of the parameter [74, 135, 134]. The distribution ρ is a way of applying constraints to the parameters θ_i.

Nuisance parameters

The values θ_i are often called "nuisance parameters"; these are quantities which enter the analysis but are of no intrinsic interest. The constraint $\rho(\tilde{\theta}_i|\theta_i)$ is applied because often the desired value of a nuisance parameter $\tilde{\theta}_i$ is known within an uncertainty and applying this function will help avoid false minima due to nonphysical values during the fit.

In terms of searches for new physics, these nuisance parameters represent systematic uncertainties. At CMS, there are two types of systematic uncertainties: shape and normalization. Shape uncertainties are constrained with a Gaussian probability density function

$$\rho(\tilde{\theta}_i|\theta_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(\theta_i - \tilde{\theta}_i)^2}{2\sigma^2} \right). \tag{10.6}$$

Normalization uncertainties are constrained with a log-normal distribution

$$\rho(\tilde{\theta}_i|\theta_i) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(- \frac{(\ln(\theta_i/\tilde{\theta}_i))^2}{2(\ln\kappa)} \right) \frac{1}{\theta_i} \,. \tag{10.7}$$

10.1.2 Hypothesis Testing with the Profile Likelihood

In this search for vector-like quarks, the signal strength μ is the parameter of interest and can be used to construct a hypothesis test. In this case the null hypothesis H_0 is that the signal strength is 0; for this reason, it is often referred to the background only hypothesis. This hypothesis is tested by utilizing the profile likelihood as a test statistic [74, 135].

Profile Likelihood

Let $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ be the values of the parameters which maximize L on a data set, where the data is taken from experiment or pseudo-experiment. Then, let $\hat{\boldsymbol{\theta}}_\mu$ be the values of the nuisance parameters which maximize L for a fixed signal strength μ. The process of fitting $\hat{\boldsymbol{\theta}}$ for given values of μ is called "profiling" and the profile likelihood is defined as

$$q_\mu = -2\ln \frac{L(data|\mu, \hat{\boldsymbol{\theta}}_\mu)}{L(data|\hat{\mu}, \hat{\boldsymbol{\theta}})} \;\; \hat{\mu} \leq \mu \,, \tag{10.8}$$

where the constraint $\hat{\mu} \leq \mu$ prevents the exclusion of any μ lower than the best fit $\hat{\mu}$ ensuring a one-side confidence interval. The profile likelihood can be used to test a hypothesis by constructing a p-value.

The p-value and Significance

The profile likelihood yields a probability density function $g(t_\mu|H)$ from the hypothesis H which is being tested. This can be used to quantify the agreement between data and the hypothesis with the p-value

$$p = \int_{q_\mu,obs}^{\infty} g(q_\mu|H)dq_\mu = \int_{q_\mu,obs}^{\infty} g(q_\mu|\mu, \hat{\boldsymbol{\theta}}_\mu)dq_\mu \,, \tag{10.9}$$

where $q_{\mu,obs}$ is the value of the test statistic observed in data. The profile likelihood is defined such that lower values of p correspond to worse agreement with the hypothesis. In the case of the null hypothesis, $\mu = 0$, if the p-value is sufficiently low then the null

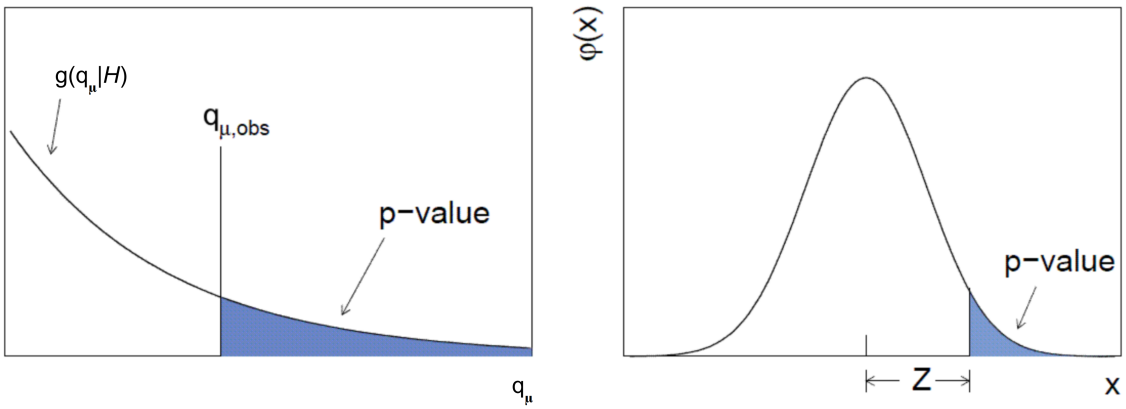

Figure 10.1: (Left) shows the relationship between the p-value and the probability density function of the profile likelihood for the hypothesis being tested. (Right) shows the relationship of the p-value to an area on a Gaussian tail resulting in the significance Z. Figure adapted from [45].

hypothesis can be rejected. When the null hypothesis is Standard Model background, then rejecting the null hypothesis supports physics beyond the Standard Model [74, 135].

Properly quantifying small p-values is quite difficult. Thus, in high energy physics, the p-value is expressed in terms of significance Z. Z is defined such that p is equal to the area under a standard Gaussian distribution from Z to ∞

$$Z = \Phi^{-1}(1-p)\,, \tag{10.10}$$

where Φ^{-1} is the inverse of the cumulative distribution of the single sided Gaussian. Z is just the number of standard deviations away the observed data is from the mean of a Gaussian distribution. This subtle definition is shown in Fig. 10.1. At collider experiments, the common practice is to set a significance of 5σ (p-value of 2.87×10^{-7}) as the requirement for discovery. When a discovery is not made, the p-value can be used to set limits on the signal model [74, 12].

Confidence Limits

If the calculated p-value is unable to exclude the null hypothesis then the upper limit on μ_s can be set using the CL_s ratio

$$CL_s = \frac{p_\mu}{1-p_0}\,. \tag{10.11}$$

The tested value of μ is said to be excluded at a confidence level α, where $\alpha = 1 - CL_s$. At CMS, 95% is chosen for upper limits. This is commonly referred to as the modified frequentist approach [74, 135, 134, 45]

Previously, the value $CL_{s+b} = p_\mu$ was used for confidence levels. However, this approach may lead to the exclusion of low signal strengths where the analysis is not sensitive. The CLs method was created to protect against this. The quantity CL_s can be determined by generating data from pseudo-experiments. A pseudo-experiment (also called pseudo-data or toy) is a Monte Carlo simulated distribution created from the model where the nuisance parameter values vary from the optimal values within the errors. Alternatively, CL_s can be approximated.

Asymptotic Limit

The probability density function $g(q_\mu|H)$ follows a known formula (described in [45]) in the limit of a large data sample. This known formula is used to create the asymptotic limit approximation and it can be used to remove the need to generate and fit toy data from pseudo-experiments. This approximation requires a representative dataset where the observed rates exactly match model predictions when the nuisance parameters are set to the mean value. This is the Asimov dataset[2]—the properties of this dataset are used to derive the expected limit and uncertainty bands [135, 45].

10.2 Systematic Uncertainties

In this search for the vector-like quarks, the nuisance parameters are systematic uncertainties. These are used to construct a model which is fit to H_T distributions in each of the 126 regions presented in the previous chapter. In this model, the parameter of interest is the signal strength μ as defined in Eq. 10.4. The systematic uncertainties are described in detail below.

- **Process Cross Sections**: The Monte Carlo simulated backgrounds have associated cross section uncertainties. These uncertainties in the cross sections used to

[2]The "Asimov" dataset name is inspired by the short story *Franchise* by Isaac Asimov where a computer selects a single representative person to vote on issues eliminating the need for large elections.

normalize simulated background processes are included. For the W+jets and Z+jets backgrounds, uncertainties of 15% are applied. For the subdominant diboson, rare top process (ttV, $4t$), and Higgs contributions, an uncertainty in the cross section value of 50% is applied.

- **Integrated Luminosity** A 2.5% uncertainty is applied over the integrated luminosity recorded by CMS during the 2016, 2017 and 2018 runs. This uncertainty modifies all simulated signal and background samples as a normalization uncertainties.

- **Pile-up reweighting** The simulated samples used in the analysis are reweighted to account for the discrepancy in the distribution of the number of pileup interactions per event in data/MC. This pile-up distribution is obtained using a proton-proton inelastic (minimum bias) cross section value of 69.2 mb. A systematic uncertainty in the distribution is obtained by varying the value by $\pm 4.6\%$, resulting in an uncertainty with both a normalization and shape component.

- **Jet Energy Scale and Resolution** Uncertainties in the corrections applied to jets are propagated to the final discriminating distributions by reconstructing events with the jet-level corrections shifted within their corresponding uncertainties, which have dependence on the jet p_T and η.

- **Parton Distribution Functions** The Monte Carlo simulated data utilize parton distribution functions from the NNPDF3.1 set [138]. An overall uncertainty for these functions is calculated using the method outlined in [139].

- **Q^2 Scale Uncertainties** Monte Carlo simulation that describes QCD processes utilize parton densities which have a factorization and renormalization scale dependence. This results in two uncertainties (μ_F, μ_R). These uncertainties affect our $t\bar{t}$ and $T'\bar{T}'$ signal Monte Carlo simulated samples. To appropriately account for this uncertainty we vary μ_F and μR up and down by factors of two to create a combined uncertainty envelope resulting in one "scale uncertainty" nuisance parameter [89].

Source	Prior	Applies to Samples
Integrated luminosity	2.5%	All
Jet Energy Scale	$\pm 1\sigma(p_T, \eta)$	All simulation
Jet Energy Resolution	$\pm 1\sigma(\eta)$	All simulation
Pile-up reweighting	$\pm 1\sigma$	All simulation
Parton distribution functions	$\pm 1\sigma$	$t\bar{t}$, VLQ
Diboson cross section	50%	WW, ZZ, WZ
Rare top process cross section	50%	ttW, ttZ, $ttWW$, $ttWZ$, $ttZZ$
W+jets cross section	15%	W+jets
Z+jets cross section	15%	W+jets
Q^2 Scale	$\pm 1\sigma$	$t\bar{t}$, VLQ

Table 10.1: General systematic uncertainties used in this search.

- **BEST Scale Factor Uncertainty** Jet classifiers have different performances in Monte Carlo simulated data and collision data. Scale factors are used to correct the performance in simulation to match that of the data. The scale factors are given an input value of one and adjusted with nuisance parameters. For top, Higgs, W, Z, and bottom, there are five uncertainties for the efficiency and five uncertainties for the mistag rate. Then, for QCD, there is one normalization uncertainty for categorization efficiency. This amounts to 11 uncertainties where the prior for each is assumed to be Gaussian-distributed with 5% uncertainty.

- **Data Driven Background Uncertainties** The data driven background estimate described in section 9.2.2 assumes that the control region contains only QCD events, but this creates an overestimate since there can be other contributions. To adjust this, the background estimate has six systematic uncertainties ($r^X_{DDSystematic}$), one calculated for each jet category where X = Higgs, top, bottom, W, Z, or QCD. The background uncertainties are calculated using the individual uncertainties (δ_X) on each ϵ_X. To determine $r^X_{DDSystematic}$, r is recalculated by varying ϵ_X for one category

Source	Prior	Applies to simulation
QCD-tag scale factor	5%	One per QCD-tagged jet
top-tag scale factor	5%	One per top-tagged jet
Higgs-tag scale factor	5%	One per Higgs-tagged jet
Z-tag scale factor	5%	One per Z-tagged jet
W-tag scale factor	5%	One per W-tagged jet
bottom-tag scale factor	5%	One per bottom-tagged jet
top-mistag scale factor	5%	One per top-tagged jet
Higgs-mistag scale factor	5%	One per Higgs-tagged jet
Z-mistag scale factor	5%	One per Z-tagged jet
W-mistag scale factor	5%	One per W-tagged jet
bottom-mistag scale factor	5%	One per bottom-tagged jet

Table 10.2: The systematic uncertainties on the BEST Scale factors which apply to all samples of Monte Carlo simulated data.

while the other ϵ_Y remain constant. This is represented mathematically as

$$r^X_{DDSystematicUp} = \sum_{events} \left\{ \sum_{permutations} \left[\prod_{i=1}^{4} [\epsilon_{X_i}(p_T(i)) + \delta^{up}_{X_i}(p_T(i))] \right] \right\}, \quad (10.12)$$

$$r^X_{DDSystematicDown} = \sum_{events} \left\{ \sum_{permutations} \left[\prod_{i=1}^{4} [\epsilon_{X_i}(p_T(i)) - \delta^{down}_{X_i}(p_T(i))] \right] \right\}. \quad (10.13)$$

These parameters are first fit to the six validation regions, then the measured values are used as inputs during the fit to the signal regions. These six nuisance parameters allow for the background estimate to account for non-QCD contributions to the control region and are an improvement over the analysis presented in [43].

The general analysis uncertainties are summarized in Table 10.1. The BEST scale factor uncertainties are applied to Monte Carlo simulated data; these are summarized in Table 10.2. The uncertainties related to the data driven background estimate are summarized in Table 10.3. Together, these are the nuisance parameters used in maximum likelihood fit for setting exclusion limits on the search for vector-like quarks.

Source	Prior	Applied to data driven estimate
Data driven estimate top	$\pm 1\sigma$	One per t-tagged jet
Data driven estimate Higgs	$\pm 1\sigma$	One per H-tagged jet
Data driven estimate Z	$\pm 1\sigma$	One per Z-tagged jet
Data driven estimate W	$\pm 1\sigma$	One per W-tagged jet
Data driven estimate bottom	$\pm 1\sigma$	One per b-tagged jet
Data driven estimate QCD	$\pm 1\sigma$	One per QCD-tagged jet

Table 10.3: The systematic uncertainties on the data driven background estimate used in this analysis.

10.3 Statistical Tests

The model is tested in background dominated validation regions before being applied to the signal regions. This process consists of four tests aimed at identifying potential problems. These tests are described in detail below.

10.3.1 Nuisance Parameter Pulls

The nuisance parameter values are compared to the assumption by measuring pulls. This is done by performing the maximum likelihood fit on Monte Carlo pseudo-experiments and then calculating the pull

$$\text{pull} = \frac{\theta_{fit} - \tilde{\theta}}{\sigma(\theta_{fit})}\,, \tag{10.14}$$

where θ_{fit} is fit of the nuisance parameter, $\tilde{\theta}$ is the nominal value, and $\sigma(\theta_{fit})$ is the variance. A pull with mean close to 0 was well chosen. When the pull width is smaller than one, then the nuisance variance is over-constrained in the fit and when it is greater than one, the variance is under-constrained in the fit [74]. The pull plot for an Asimov dataset is shown in Fig. 10.2. This dataset was constructed from the six validation regions

10.3.2 Nuisance Parameter Impacts

This test shows the "impact" of each nuisance parameter. This is the amount that the parameter-of-interest (in this case, signal strength) changes when a nuisance parameter is varied up and down. During this test, one nuisance parameter is fixed and the others

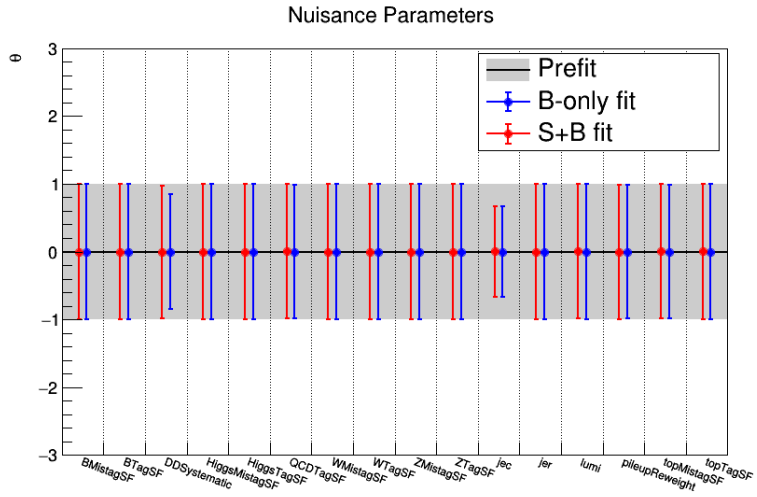

Figure 10.2: The nuisance parameter pulls for an Asimov dataset constructed from the six validation regions.

are profiled during a fit. The impact in the parameter-of-interest is measured when the nuisance parameter is fixed to $\pm\sigma$ from the nominal. Figure 10.3 shows the nuisance parameter impacts for an Asimov dataset constructed from the six validation regions. This example represents the ideal impacts for the analysis.

10.3.3 Goodness of Fit

This test shows the value of a test statistic measured in data and compares it to a distribution of test statistics measured in a series of pseudo-experiments (assuming $\mu = 0$). The chosen test statistic is the "saturated test statistic" described in detail in [140]. The measured test statistic in data is compared to the distribution of pseudo-experiments by creating a p-value. A value close to 1.0 or 0.0 indicates that the background-only model does not describe the data. Such a value indicates a poor model construction since this test is applied to background dominated validation regions. Figure 10.4 shows the Goodness of Fit in a combination of all six validation regions.

10.3.4 Signal Injection

Perhaps the most important of all these statistical tests is the signal injection test. This shows the ability of the model to respond to signal ensuring that the model does not mask signal and that it is not biased to find signal where none exists.

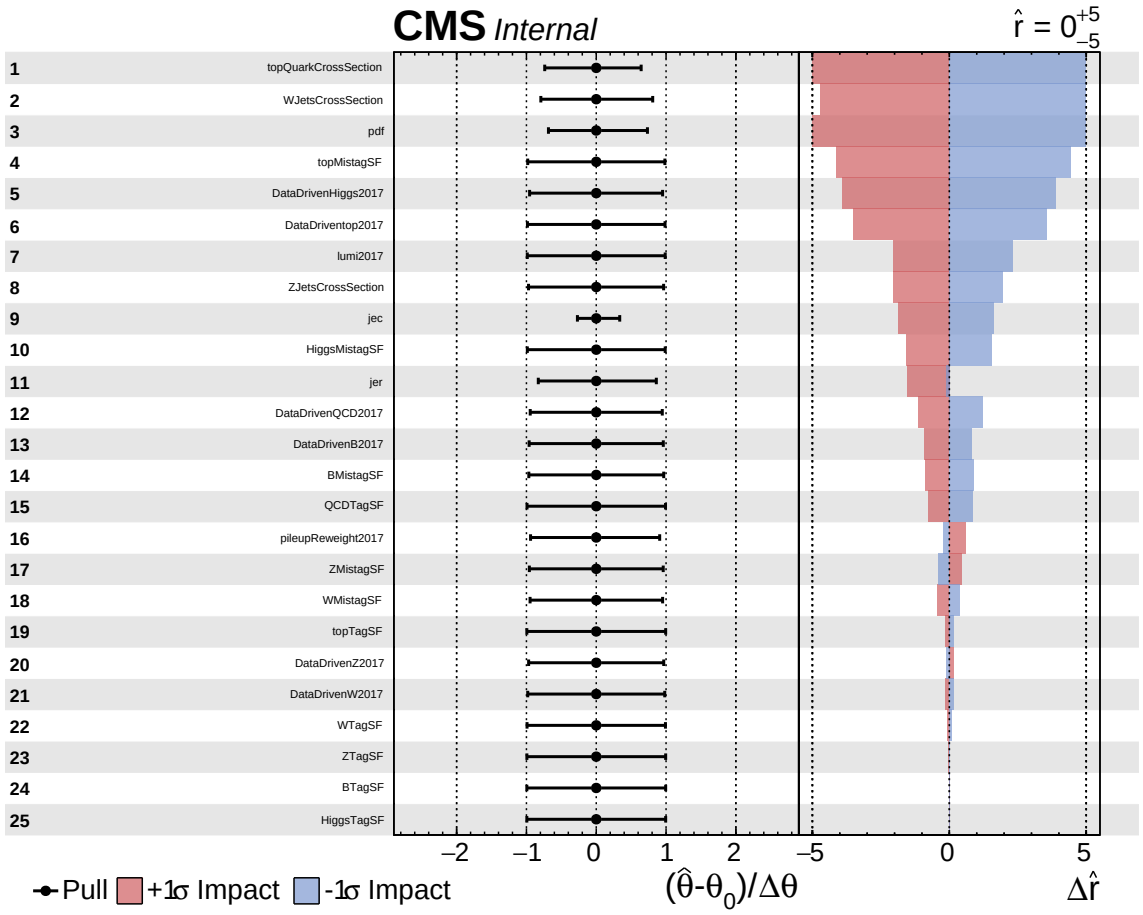

Figure 10.3: Nuisance parameter impacts for a fit to an Asimov dataset constructed by from the six validation regions. This fit uses the signal hypothesis of a T' with a mass of 1000 GeV.

First, a number of pseudo-experiments are created for different fixed values of μ. These pseudo-experiments use the nuisance parameters measured in the validation regions, but are extrapolated to signal region where the injected signal is added. Then, these models are fit with the statistical model. The results are then observed in two steps. First, in the distributions of $\mu - \mu_{inj}$ (where μ_{inj} is the injected signal) for pseudo-experiments created at different μ_{inj} values. Then, in the signal injection pull distributions which are normalized by $(\mu - \mu_{inj})/\mu_{err}$ where μ_{err} is the error in that measurement given by the

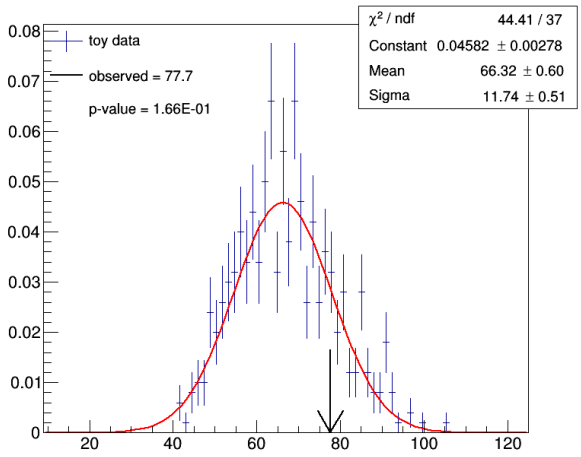

Figure 10.4: Goodness of Fit test performed on the six validation regions.

equation

$$\mu_{err} = \begin{cases} \mu_{HiErr} \text{ if } (\mu - \mu_{inj}) < 0 \\ \mu_{LoErr} \text{ if } (\mu - \mu_{inj}) > 0 \end{cases}, \tag{10.15}$$

with μ_{HiErr} is the upper error on μ and μ_{LoErr} is the lower error. In both cases, the distributions are fitted to Gaussian distributions. In the normalized case, a perfect model would show a Gaussian centered at 0 with $\sigma = 1$. An example of the signal injection test is shown in Fig. 10.5. In this example, a signal strength of one for B' with mass 1000 GeV is injected and the resulting toys are fit to create the Gaussian like distribution.

10.4 Limits

Exclusion limits are set by comparing observed and expected limits to theoretical predictions. In this search, limits are set on the mass of top-like and bottom-like vector like quarks for various branching rations. Figure 10.6 shows the expected limit for top-like vector like quarks (T') in a combination of 2017 and 2018 collision data. At the moment, these limits are blinded, meaning there is no observed limit from collision data. The signal hypothesis assumes a branching ratio of $T' \to tZ : T' \to tH : T' \to bW = 1 : 1 : 1$. The previous search on 2016 collision ($35.9\,\text{fb}^{-1}$) data had an expected exclusion limit near $m(T') \approx 1000\,\text{GeV}$ for this branching ratio. The new analysis improved the expected limit

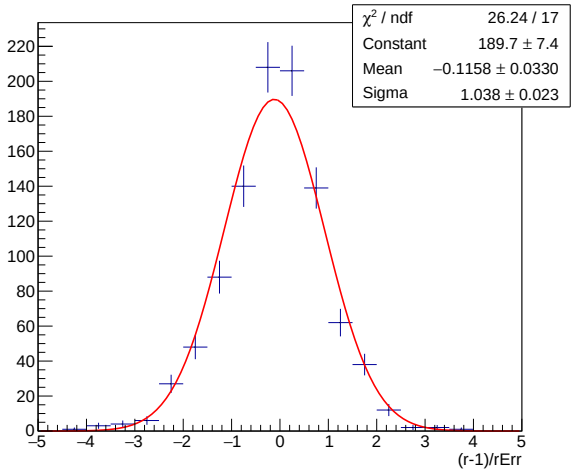

Figure 10.5: A signal strength of one injected for B' with mass 1000 GeV. The resulting distribution of fits to toy datasets is centered at zero with a Gaussian-like shape; this shows that there is no bias in the signal + background model.

to greater than 1300 GeV for a similar amount of data ($41.53\,\text{fb}^{-1}$), this limit becomes greater than 1400 GeV when combining 2017 and 2018 collision data. These limits will be further expanded to all years in LHC Run 2 for both T' and B' at various branching ratio hypotheses. Once approved by CMS, the search will be unblinded to set observed exclusion limits.

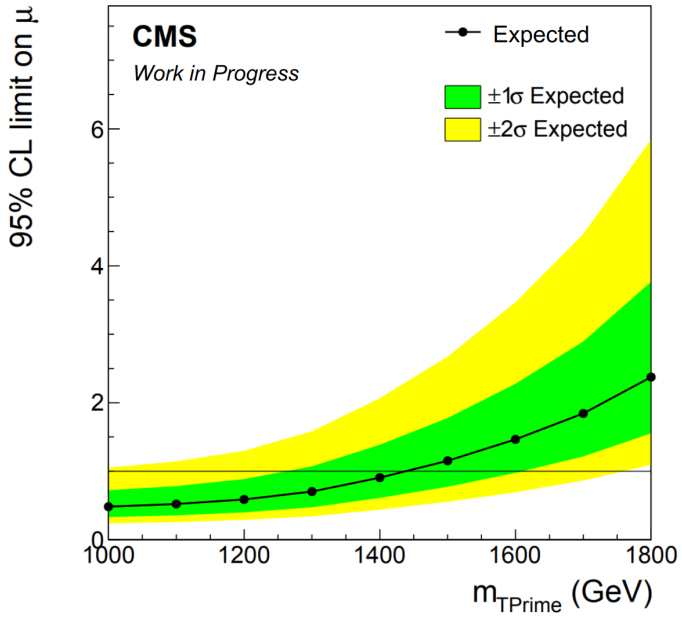

Figure 10.6: The expected limits for top-like vector-like quarks on a combination of 2017 and 2018 collision data. The analysis is sensitive until the point at which the expected limit crosses the signal strength $\mu = 1$ line.

Chapter 11

Conclusions

There is no real ending. It's just the place where you stop the story.

Frank Herbert

The Compact Muon Solenoid observes high energy proton collisions at CERN's Large Hadron Collider. The LHC has already completed two periods of data taking called Run 1 and Run 2. During the course of this thesis, the LHC was shutdown for upgrades and then began a third period of data taking, Run 3. However, the upgrade period was primarily focused on preparing for a High Luminosity version of the LHC which is currently scheduled to begin in 2029.

During Long Shutdown 2 of the LHC, the CMS experiment began a series of upgrades to prepare for operations in this high rate environment. This included adding the first set of Gas Electron Multipliers to the muon system. This thesis work helped produce, install, and commission the first CMS GEMs which are now operational during LHC Run 3.

GEMs are micro-pattern gas detectors with high rate capabilities. The technology arose while investigating improvements to micro strip gas counters and now GEMs are widely used in particle physics. At CMS, the technology was chosen to provide two new layers for level-1 trigger to use for measuring the muon bending angle. This will allow for lower muon p_T trigger thresholds during the high rates in HL-LHC.

After transporting the GEM chambers to the cavern floor and before installation, the electronics are quickly checked for disconnections. This fast electronics test can provide a

template for other detectors to use before installation into particle physics experiments. During commissioning it was discovered that an electronic component, the VTRx, needs to be carefully cooled. This component is used in many detector upgrades for the HL-LHC and these results can be used by other detector groups to design cooling. While the GEMs were being produced, installed, and commissioned, the data from the LHC was still being used physics searches and measurements.

The data recorded during LHC Run 2 was used to conduct a search for a pair of vector-like quarks decaying to four large radius jets. Vector-like quarks are an extension to the Standard Model that arose in several theories aiming to prevent divergent terms in the Higgs bare mass. These theories provide an alternative to supersymmetry and result from relatively simple extensions to the Standard Model.

This search for vector-like quarks used the Boosted Event Shape Tagger to classify jets creating 126 orthogonal regions. The H_T distributions in each region were measured to perform the search. Once approved by the CMS collaboration, exclusion limits on T' and B' masses will be placed using the CLs method.

A previous version of BEST was used to search for vector-like quarks in 2016 collision data at CMS. BEST is a neural network based jet tagger that classifies jets into six categories based on heavy object decays. The 2016 version was improved before being applied to the full Run 2 dataset. During the process of improving BEST, an image based method of representing jets in various rest frames was created. This method was used to train a convolutional neural network in hopes that this would be better able to identify patterns in complicated rest frames. This method showed some improvements, but not enough to warrant the significant increase in model complexity. BEST is a simple neural network that competes with many advanced techniques, highlighting the importance of clever physics-inspired input features. Although the image technique did not yield large improvements, this method of data representation could be adapted for future work with jet classification.

In J.K. Rowling's Harry Potter series, Albus Dumbledore famously said

Help will always be given at Hogwarts to those who ask for it.

J.K. Rowling

This thesis is the culmination of collaboration with scientists around the world, it would have not have been possible without the unique atmosphere created at CERN. For any young graduate student, perhaps the most important conclusion from this thesis is a rephrasing of Dumbledore's famous quote.

There is always help at CERN for those who ask for it.

Brendan Regnery

Bibliography

[1] J. Ellis, *Higgs physics*, 2013.

[2] M. Drees, *An Introduction to Supersymmetry*, Tech. Rep. hep-ph/9611409. APCTP-96-05. KEK-TH-501, Seoul Univ. Asia-Pacific Cent. Theor. Phys., Seoul, S Korea (Nov, 1996).

[3] J. Berger, J. Hubisz and M. Perelstein, *A fermionic top partner: naturalness and the lhc*, *Journal of High Energy Physics* **2012** (2012) .

[4] CMS collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[5] CMS Tracker Group collaboration, *The CMS Phase-1 Pixel Detector Upgrade*, *JINST* **16** (2021) P02027 [2012.14304].

[6] CMS collaboration, *Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV*, *JINST* **15** (2020) P05002 [1910.00079].

[7] CMS collaboration, *Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays*, *JINST* **5** (2010) T03021 [0910.5530].

[8] CMS collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JINST* **13** (2018) P06015 [1804.04528].

[9] C. Collaboration, *Performance of the CMS drift tube chambers with cosmic rays*, *Journal of Instrumentation* **5** (2010) T03015.

[10] CMS collaboration, *The Phase-2 Upgrade of the CMS Muon Detectors*, Tech. Rep. CERN-LHCC-2017-012, CMS-TDR-016, CERN, Geneva (2017).

[11] CMS collaboration, *Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JINST* **15** (2020) P10017 [2006.10165].

[12] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01.

[13] THM-Webredaktion, "Elektromagnetische strahlung."

[14] S. Chen, *A Study on Properties of Novel Metallic Foam for Nuclear Applications*, Ph.D. thesis, 08, 2015. 10.13140/RG.2.1.2700.6804.

[15] A.-L. Manninen, *CLINICAL APPLICATIONS OF RADIOPHOTOLUMINESCENCE (RPL) DOSIMETRY IN EVALUATION OF PATIENT RADIATION EXPOSURE IN RADIOLOGY Determination of absorbed and effective dose*, Ph.D. thesis, 11, 2014.

[16] F. Simone, *Development and characterization of the Fast Timing Micro-Pattern Gaseous Detector*, Ph.D. thesis, Università degli Studi di Bari Aldo Moro, 2018.

[17] M. Berger, J. Hubbell, S. Seltzer, J. Coursey and D. Zucker, *Xcom: Photon cross section database (version 1.2)*, 1999-01-01, 1999.

[18] H. Raether, *Electric Avalanches and Breakdown in Gases*, Butterworths Advanced Physics Series, Washington, Butterworths (1964).

[19] J.A. Merlin, *Study of long-term sustained operation of gaseous detectors for the high rate environment in CMS*, Ph.D. thesis, l'Université de Strasbourg, Apr, 2016.

[20] A. Winkler, A. Karadzhinova, T. Hilden, F. Garcia, G. Fedi, F. Devoto et al., *A gaseous proportional counter built from a conventional aluminum beverage can*, *American Journal of Physics* **83** (2015) .

[21] A. Breskin, G. Charpak, C. Demierre, S. Majewski, A. Policarpo, F. Sauli et al., *High-accuracy, bidimensional read-out of proportional chambers with short resolution times*, *Nuclear Instruments and Methods* **143** (1977) 29.

[22] A. Oed, *Position Sensitive Detector with Microstrip Anode for electron Multiplication with Gases*, *Nucl. Instrum. Meth. A* **263** (1988) 351.

[23] CMS collaboration, *The CMS tracker system project: Technical Design Report*, Tech. Rep. CERN-LHCC-98-006, CMS-TDR-5, CERN-LHCC-98-006, CMS-TDR-5, CERN: European Organization for Nuclear Research (1997).

[24] F. Sauli, *The gas electron multiplier (GEM): Operating principles and applications*, *Nucl. Instrum. Meth. A* **805** (2016) 2.

[25] G. Mocellin, *Performance of the GE1/1 detectors for the upgrade of the CMS muon forward system*, dissertation, RWTH Aachen University, Aachen, 2021. 10.18154/RWTH-2021-07066.

[26] CMS collaboration, *CMS Technical Design Report for the Muon Endcap GEM Upgrade*, Tech. Rep. CERN-LHCC-2015-012, CMS-TDR-013, CERN: European Organization for Nuclear Research (6, 2015).

[27] P. Everaerts, *Rate capability and ion feedback in GEM detectors*, Master's thesis, Gent U., 2006.

[28] P. Aspell et al., *VFAT3: A Trigger and Tracking Front-end ASIC for the Binary Readout of Gaseous and Silicon Sensors*, in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference*, p. 8824655, 2018, DOI.

[29] J. Troska, *Versatile link technical specification, part 2.1*, Tech. Rep. CERN: European Organization for Nuclear Research (2017).

[30] Irving, "Fiber optic adapter/coupler tutorial: Fs community."

[31] J. Dittmann and J. Wilson, *Vtrx failure investigation: Recent cms temeprature studies*, 2021.

[32] CMS collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *JINST* **12** (2017) P10003 [1706.04965].

[33] M.M. Defranchis, *Event reconstruction in cms*, in *First Measurement of the Running of the Top Quark Mass*, (Cham), pp. 47–65, Springer International Publishing (2021), DOI.

[34] D0 collaboration, *Observation of Single Top Quark Production*, *Phys. Rev. Lett.* **103** (2009) 092001 [0903.0850].

[35] CMS Collaboration collaboration, *Pileup mitigation at CMS in 13 TeV data*, Tech. Rep. CMS-PAS-JME-18-001, CERN, Geneva (2019).

[36] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup Per Particle Identification*, *JHEP* **10** (2014) 059 [1407.6013].

[37] ATLAS and CMS Collaborations collaboration, *Jet substructure measurements with CMS and ATLAS*, .

[38] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [0802.1189].

[39] CMS collaboration, *Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques*, *JINST* **15** (2020) P06005 [2004.08262].

[40] A. Andreassen, I. Feige, C. Frye and M.D. Schwartz, *JUNIPR: a framework for unsupervised machine learning in particle physics*, *The European Physical Journal C* **79** (2019) .

[41] J. Yacim and D. Boshoff, *Impact of artificial neural networks training algorithms on accurate prediction of property values*, *Journal of Real Estate Research* **40** (2018) 375.

[42] F. Chollet, *Deep Learning with Python*, Manning (Nov., 2017).

[43] CMS collaboration, *Search for pair production of vectorlike quarks in the fully hadronic final state*, *Phys. Rev. D* **100** (2019) 072001 [1906.11903].

[44] C.-F. Wang, "A basic introduction to separable convolutions."

[45] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [1007.1727].

[46] *Grisons: Les rhb ont battu le record du train le plus long du monde*, Oct, 2022.

[47] L. Panizzi, *Vector-like quarks: t' and partners*, *Nuovo Cim. C* **037** (2014) 69.

[48] J.A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer and M. Pérez-Victoria, *Handbook of vectorlike quarks: Mixing and single production*, *Physical Review D* **88** (2013) .

[49] C.W. Fabjan and H. Schopper, eds., *Elementary Particles. Detectors for Particles and Radiation.: Part 1: Principles and Methods*, vol. 21B1 of *Landolt-Boernstein - Group I Elementary Particles, Nuclei and Atoms*, Springer (2011), 10.1007/978-3-642-03606-4.

[50] N. Jeevanjee, *An introduction to tensors and group theory for physicists; 1st ed.*, Birkhäuser, New York, NY (2011), 10.1007/978-0-8176-4715-5.

[51] M.E. Peskin and D.V. Schroeder, *An introduction to quantum field theory*, Westview, Boulder, CO (1995).

[52] P. Langacker, *The Standard Model and Beyond*, Series in High Energy Physics, Cosmology and Gravitation, Taylor and Francis, Boca Raton, FL (2010).

[53] D.J. Griffiths, *Introduction to elementary particles; 2nd rev. version*, Physics textbook, Wiley, New York, NY (2008).

[54] S. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961) 579.

[55] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* **13** (1964) 321.

[56] P.W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* **13** (1964) 508.

[57] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264.

[58] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C* **680519** (1968) 367.

[59] T. Lancaster and S.J. Blundell, *Quantum field theory for the gifted amateur*, Oxford University Press, Oxford (Apr, 2014), 0199699321.

[60] R.A. Millikan., *On the elementary electrical charge and the avogadro constant*, *Phys. Rev.* **2** (1913) 109.

[61] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV*, *Phys. Lett. B* **122** (1983) 103.

[62] UA1 collaboration, *Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider*, *Phys. Lett. B* **126** (1983) 398.

[63] Y. Fukuda, T. Hayakawa, E. Ichihara, K. Inoue, K. Ishihara, H. Ishino et al., *Evidence for oscillation of atmospheric neutrinos*, *Physical Review Letters* **81** (1998) 1562–1567.

[64] SNO collaboration, *Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory*, *Phys. Rev. Lett.* **87** (2001) 071301 [nucl-ex/0106015].

[65] R. Barbieri, *Ten lectures on the electroweak interactions*, 2007.

[66] J.H. Christenson, J.W. Cronin, V.L. Fitch and R. Turlay, *Evidence for the 2π decay of the k_2^0 meson*, *Phys. Rev. Lett.* **13** (1964) 138.

[67] NA48 collaboration, *A New measurement of direct CP violation in two pion decays of the neutral kaon*, *Phys. Lett. B* **465** (1999) 335 [hep-ex/9909022].

[68] KTeV collaboration, *Observation of direct CP violation in $K_{S,L} \to \pi\pi$ decays*, *Phys. Rev. Lett.* **83** (1999) 22 [hep-ex/9905060].

[69] S. Chatrchyan, V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo et al., *Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc*, *Physics Letters B* **716** (2012) 30–61.

[70] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim et al., *Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc*, *Physics Letters B* **716** (2012) 1–29.

[71] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2020** (2020) 083C01.

[72] D. Clowe, A. Gonzalez and M. Markevitch, *Weak-lensing mass reconstruction of the interacting cluster 1e 0657-558: Direct evidence for the existence of dark matter*, *The Astrophysical Journal* **604** (2004) 596–603.

[73] E. Corbelli and P. Salucci, *The extended rotation curve and the dark matter halo of m33*, *Monthly Notices of the Royal Astronomical Society* **311** (2000) 441–447.

[74] D. Pinna, *Search for Dark Matter in Association with Top Quarks with the CMS Detector*, Ph.D. thesis, University of Zurich, 2017.

[75] B. Moran, W. Hoover and S. Bestiale, *Diffusion in a periodic lorentz gas*, *Journal of Statistical Physics* **48** (1987) 709.

[76] L. Canetti, M. Drewes and M. Shaposhnikov, *Matter and antimatter in the universe*, *New Journal of Physics* **14** (2012) 095012.

[77] J. Bovy, *The Self-energy of the electron: A Quintessential problem in the development of QED*, physics/0608108.

[78] M.J. Dugan, H. Georgi and D.B. Kaplan, *Anatomy of a Composite Higgs Model*, *Nucl. Phys. B* **254** (1985) 299.

[79] G. von Gersdorff, E. Ponton and R. Rosenfeld, *The dynamical composite higgs*, 2015.

[80] N. Arkani-Hamed, A.G. Cohen, E. Katz and A.E. Nelson, *The littlest higgs*, *Journal of High Energy Physics* **2002** (2002) 034–034.

[81] J.S. Bonilla Castro, *Reconstructing the Top Quark in a Search for a Pair-Produced Supersymmetric Partner in the All-Hadronic plus Missing Energy Final State Using* 139 fb^{-1} *of* $\sqrt{s} = 13$ *TeV Proton-Proton Collisions Delivered by the Large Hadron Collider and Collected by the ATLAS Detector*, Ph.D. thesis, University of Oregon, 2019.

[82] CMS collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*, Tech. Rep. CERN-LHCC-2006-001, CMS-TDR-8-1, CERN-LHCC-2006-001, CMS-TDR-8-1, CERN: European Organization for Nuclear Research (2006).

[83] A. Benecke, *Searches for new heavy bosons and vector-like quarks with the CMS experiment at* $\sqrt{s}$*= 13 TeV and novel pileup mitigation techniques*, Ph.D. thesis, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky, 2020.

[84] L.H. Baekeland, *The synthesis, constitution, and uses of bakelite.*, *Journal of Industrial & Engineering Chemistry* **1** (1909) 149 [https://doi.org/10.1021/ie50003a004].

[85] R.C. Fernow, *Introduction to Experimental Particle Physics*, Cambridge University Press (3, 1983).

[86] F. Sauli, *Gaseous Radiation Detectors: Fundamentals and Applications*, Cambridge University Press, Cambridge (2014).

[87] W.R. Leo, *Techniques for Nuclear and Particle Physics Experiments: A How to Approach*, Springer (1987).

[88] R. Bouclier, M. Capeans, W. Dominik, M. Hoch, J.C. Labbe, G. Million et al., *The Gas electron multiplier (GEM)*, *ICFA Instrum. Bull.* **1996** (1996) F53.

[89] P. Everaerts, *W Cross Section Measurement in the Electron Channel in pp Collisions at sqrt(s)= 7 TeV*, Ph.D. thesis, Massachusetts Institute of Technology, 2011.

[90] F. Licciulli et al., *Calibration, bias and monitoring system for the VFAT3 ASIC of the CMS GEM detector*, in *7th IEEE International Workshop on Advances in Sensors and Interfaces*, pp. 81–84, 2017, DOI.

[91] M. Abbas et al., *Quality control of mass-produced GEM detectors for the CMS GE1/1 muon upgrade*, *Nucl. Instrum. Meth. A* **1034** (2022) 166716 [2203.12037].

[92] M. Bianco, F. Fallavollita, D. Fiorina, A. Pellecchia, L.F.R. Garcia, N. Rosi et al., *Rate capability of large-area triple-GEM detectors and new foil design for the innermost station, ME0, of the CMS endcap muon system*, 2201.09021.

[93] A. Pellecchia, P. Verwilligen and A. Stamerra, *Performance of triple-GEM detectors for the CMS Phase-2 upgrade measured in test beam*, 2207.09906.

[94] W. Adam, B. Mangano, T. Speer and T. Todorov, *Track Reconstruction in the CMS tracker*, Tech. Rep. CMS-NOTE-2006-041, CERN, Geneva (2006).

[95] C.A. McLean, *Search for High-Mass Top Quark Pair Resonances with the CMS Experiment*, Ph.D. thesis, UC, Davis, 2018. 10.2172/1488582.

[96] T.C. Collaboration, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, *Journal of Instrumentation* **9** (2014) P10009.

[97] K. Rose, *Deterministic annealing for clustering, compression, classification, regression, and related optimization problems*, *Proceedings of the IEEE* **86** (1998) 2210.

[98] CMS collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *JINST* **13** (2018) P05011 [1712.07158].

[99] C.A. Mantilla Suarez, *Probing new physics using initial state radiation jets at the Large Hadron Collider*, Ph.D. thesis, Johns Hopkins University, 2020.

[100] S. Agostinelli, J. Allison et al., *Geant4—a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003) 250.

[101] CMS collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, *JINST* **12** (2017) P02014 [1607.03663].

[102] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146 [1402.2657].

[103] M. Dasgupta, A. Fregoso, S. Marzani and G.P. Salam, *Towards an understanding of jet substructure*, *Journal of High Energy Physics* **2013** (2013) .

[104] ATLAS collaboration, *Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions*, *Eur. Phys. J. C* **81** (2021) 334 [2009.04986].

[105] CMS Collaboration collaboration, *Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector*, .

[106] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896 [1111.6097].

[107] W. Mcculloch and W. Pitts, *A logical calculus of ideas immanent in nervous activity*, *Bulletin of Mathematical Biophysics* **5** (1943) 127.

[108] F. Rosenblatt, *The perceptron - a perceiving and recognizing automaton*, Tech. Rep. 85-460-1, Cornell Aeronautical Laboratory, Ithaca, New York (January, 1957).

[109] K. Fukushima, *Cognitron: A self-organizing multilayered neural network*, *Biological Cybernetics* **20** (2004) 121.

[110] D.E. Rumelhart, G.E. Hinton and R.J. Williams, *Learning Representations by Back-propagating Errors*, *Nature* **323** (1986) 533.

[111] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015 [1011.2268].

[112] J. Cogan, M. Kagan, E. Strauss and A. Schwartzman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **02** (2015) 118 [1407.5675].

[113] L.G. Almeida, M. Backović, M. Cliche, S.J. Lee and M. Perelstein, *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition*, *JHEP* **07** (2015) 086 [1501.05968].

[114] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, *JHEP* **07** (2016) 069 [1511.05190].

[115] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020) 056019 [1902.08570].

[116] J.S. Conway, R. Bhaskar, R.D. Erbacher and J. Pilot, *Identification of High-Momentum Top Quarks, Higgs Bosons, and W and Z Bosons Using Boosted Event Shapes*, *Phys. Rev. D* **94** (2016) 094027 [1606.06859].

[117] S. Ruder, *An overview of gradient descent optimization algorithms*, 2016. 10.48550/ARXIV.1609.04747.

[118] D.P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014. 10.48550/ARXIV.1412.6980.

[119] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, 2015. 10.48550/ARXIV.1502.03167.

[120] S. Hochreiter
and J. Schmidhuber, *Long Short-Term Memory*, *Neural Computation* **9** (1997) 1735 [https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf].

[121] V. Mikuni and F. Canelli, *Point cloud transformers applied to collider physics*, *Machine Learning: Science and Technology* **2** (2021) 035027.

[122] J.M. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger et al., *Highly accurate protein structure prediction with alphafold*, *Nature* **596** (2021) 583 .

[123] M. Grandini, E. Bagli and G. Visani, *Metrics for multi-class classification: an overview*, 2020. 10.48550/ARXIV.2008.05756.

[124] G.C. Fox and S. Wolfram, *Observables for the analysis of event shapes in e^+e^- annihilation and other processes*, *Phys. Rev. Lett.* **41** (1978) 1581.

[125] J. Thaler and K.V. Tilburg, *Maximizing boosted top identification by minimizing n-subjettiness*, *Journal of High Energy Physics* **2012** (2012) .

[126] CMS collaboration, *Measurements of jet charge with dijet events in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **10** (2017) 131 [1706.05868].

[127] C. Weiser, *A Combined Secondary Vertex Based B-Tagging Algorithm in CMS*, Tech. Rep. CMS-NOTE-2006-014, CERN, Geneva (2006).

[128] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al., *Scikit-learn: Machine learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825.

[129] F. Chollet et al., *Keras*, 2015.

[130] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015.

[131] CMS collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *JINST* **13** (2018) P05011 [1712.07158].

[132] F. Chollet, *Xception: Deep learning with depthwise separable convolutions*, 2016. 10.48550/ARXIV.1610.02357.

[133] T.C. collaboration, *Performance of CMS muon reconstruction in pp collision events at sqrt(s) = 7 TeV*, *Journal of Instrumentation* **7** (2012) P10002.

[134] A.L. Read, *Presentation of search results: the cls technique*, *Journal of Physics G: Nuclear and Particle Physics* **28** (2002) 2693.

[135] A.J. Gilbert, *Searches for neutral Higgs bosons decaying to tau pairs and measurement of the Z+b-jet cross section with the CMS detector*, Ph.D. thesis, Imperial College London, 2014.

[136] TheATLAS, TheCMS, TheLHCHiggsCombinationGroup collaboration, *Procedure for the LHC Higgs boson search combination in Summer 2011*, Tech. Rep. CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, CERN, Geneva (2011).

[137] G. Cowan, *Statistical data analysis*, Oxford University Press, USA (1998).

[138] R.D. Ball, V. Bertone, S. Carrazza, L.D. Debbio, S. Forte, P. Groth-Merrild et al., *Parton distributions from high-precision collider data*, *The European Physical Journal C* **77** (2017) .

[139] J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A.D. Roeck, J. Feltesse, S. Forte et al., *PDF4lhc recommendations for LHC run II*, *Journal of Physics G: Nuclear and Particle Physics* **43** (2016) 023001.

[140] R.D. Cousins, *Generalization of chisquare goodness-of-fit test for binned data using saturated models, with application to histograms*, 2013.

Appendices

Appendix A

The Search for the Best Cycling Road in the Alps

Roulez autant ou aussi peu, ou aussi longtemps ou aussi court que vous vous sentez. Mais roulez.

Eddie Merckyx

Much like searching for new particles, the search for the best[1] cycling road in the alps is one that continues over the years. This is a search full of excesses. Every year, new roads are discovered that outshine the current "best" road. Some question the importance of this endeavor: Why search for the best road when there might be a new road next year?

But the search for the best cycling road is extremely important for a pair of physicists at CERN: Will Kalderon and Brendan Regnery. And this search is of high importance for the CERN community at large! Which road should the CERN director general, Fabiola Gianotti, cycle over? Thus, this pair of physicists has prepared a list—at least as much of a list as one can make in the limited duration available during a PhD. So, without further ado, the best cycling roads in the alps.

[1] Not BEST like the BEST network.

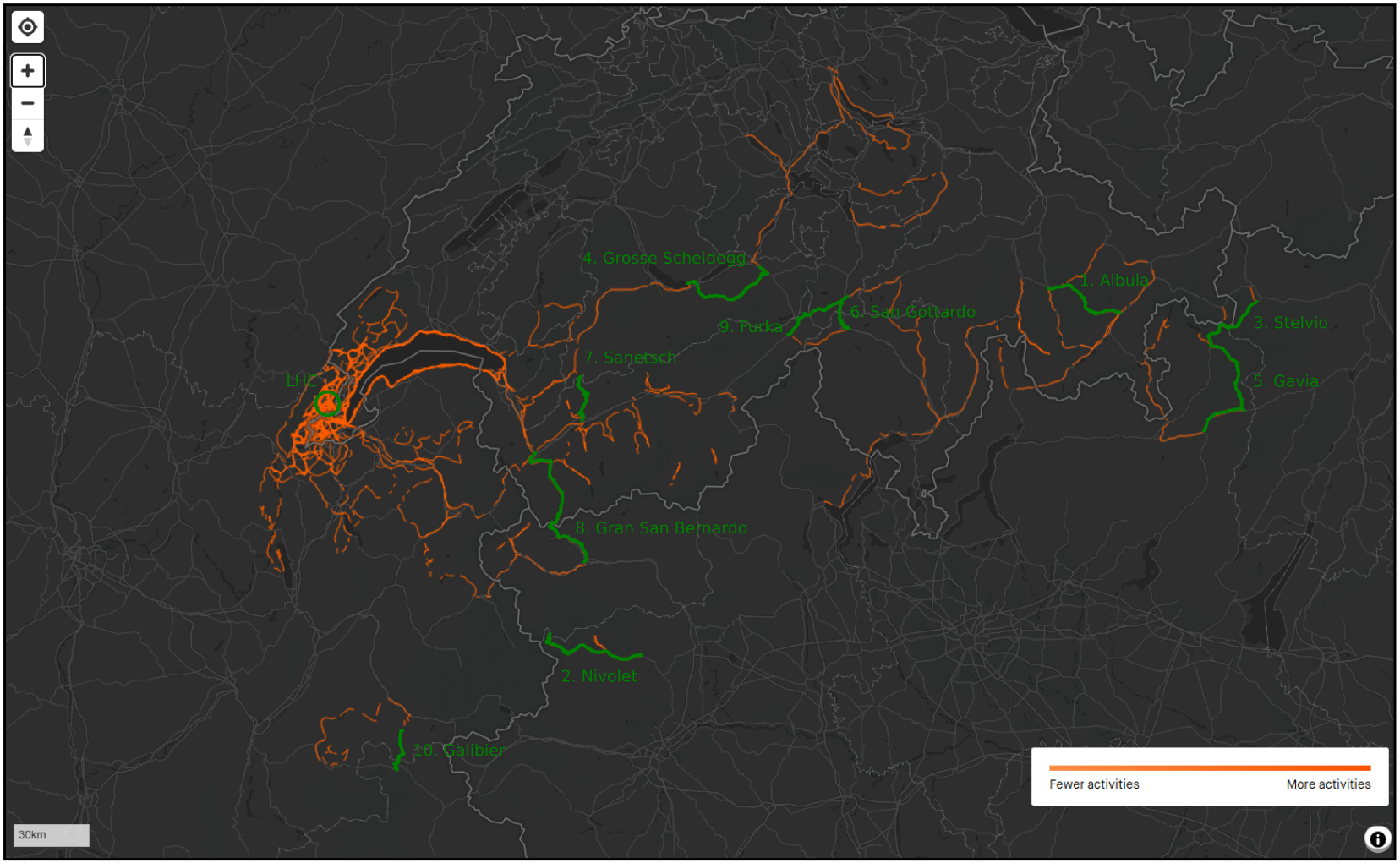

Figure A.1: A heat map from Strava (modified by Will Kalderon) of all the locations in France, Switzerland, and Italy that Brendan cycled during his PhD at CERN. In total, he cycled over 19.000 km.

10. Col du Galibier (2,642m)

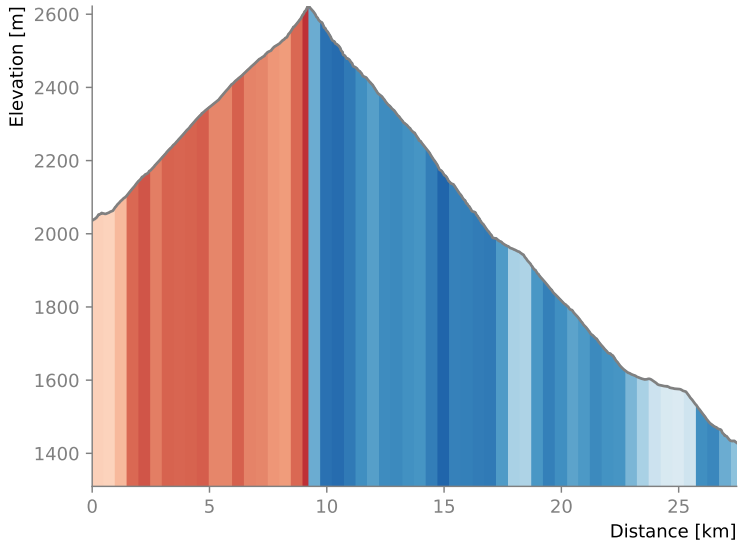

Figure A.2: A profile of the Col du Galibier as cycled by Will and Brendan. Figure from Will Kalderon.

The Col du Galibier is shrouded in history (and fog) as the center point of many decisive stages in Le Tour de France. This pass is high above the treeline and is located near the legendary French ski town, La Grave[2]. The main road to the Col starts from the Col du Lautaret and spirals upward through many high points of cycling history, including a memorial to the first organizer of Le Tour de France, Henri Desgrange. This is the same spot where, in 1926, Spanish cyclist Francisco Cepeda died after falling during a descent. This same climb has seen legendary battles between Bernard Hinault, Laurent Fignon, and Greg LeMond.

Physicists Will and Brendan described the ride up as a historical and exciting experience. However, their ride was surrounded by fog, which dulled the experience. If the physicists returned to this road, would it become higher on the list? Perhaps, but for now the sheer legend of this climb makes it deserve a spot in the top 10 cycling roads in the Alps.

[2]La Grave was the winter home of American freeskier, Doug Coombs. It has no pisted runs and is renown for extreme skiing.

Figure A.3: The legendary turns of the Col du Galibier during a short break in the fog.

9. Furka Pass (2,429m)

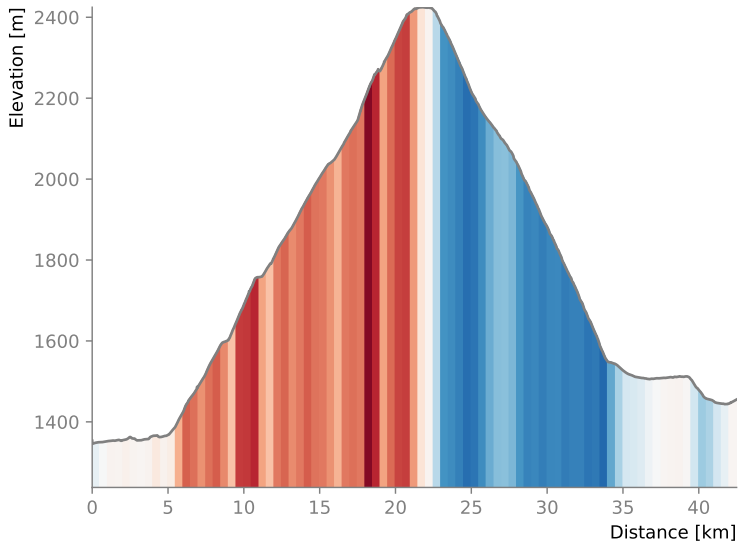

Figure A.4: A profile of the Furka pass as cycled by Will and Brendan. Figure from Will Kalderon.

The Furka pass is perhaps most famous for being featured in the iconic James Bond movie "Goldfinger". This pass transports brave cyclists from the Swiss canton Valais to the canton Uri. The Swiss giant spends very little time in the forest and offers almost

immediate alpine landscapes. Along the way, there is a must see vantage point of the Rhone glacier, the source of the large river which makes its way to the Mediterranean. And of course, the famous hotel from the James Bond movie "Goldfinger".

Figure A.5: Physicist Brendan Regnery (not James Bond, but don't tell him that!), preparing to take a call from 'M' while riding the new...Eco Edition...Aston Martin up the Furka pass.

8. Passo Gran San Bernardo (2,469m)

The Grand Saint Bernard pass is the origin of the large Saint Bernard dogs which were historically used for avalanche search and rescue by monks at the Hospice du Saint Bernard. This Hospice was originally created by Napoleon to aid travelers who needed to transport important supplies during the winter. Now the hospice and monks are led by Father Raphael—a very energetic and fun loving priest. However, when asked "what question do you get most tired of hearing" he replied "ugh...'where are the dogs?"".

This pass crosses between Italy and Switzerland. The Italian side features a grueling ascent from Aosta, while wiggling between Italian villages (often with French names). The landscape starts with forests, but often busy roads. Eventually the traffic enters a tunnel and cyclists are left on a much quieter road to the top with spectacular alpine views.

Physicists Brendan and Will have traveled to this pass in all four seasons of the year—

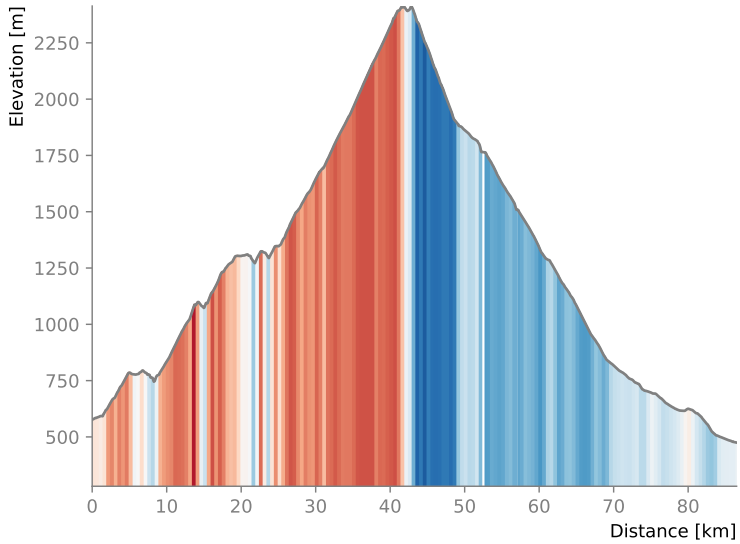

Figure A.6: A profile of the Grand Saint Bernard pass (Passo Gran San Bernardo) as cycled by Will and Brendan. Figure from Will Kalderon.

on bikes, skis, foot, and even a brief paddle board border crossing[3]. This pass comes highly recommended. Make sure to make the most of the daylight with the incredible landscapes and then enjoy a warm cup of tea in the cozy company of father Raphael.

Figure A.7: (Left) The Italian wiggles as the road rises above the Val d'Aosta. (Right) The cozy home of father Raphael at sunset.

[3]Along with Jackson Burzynski, Samuel Meehan, and Ethan Canneart.

7. Col du Sanetsch (2,242m)

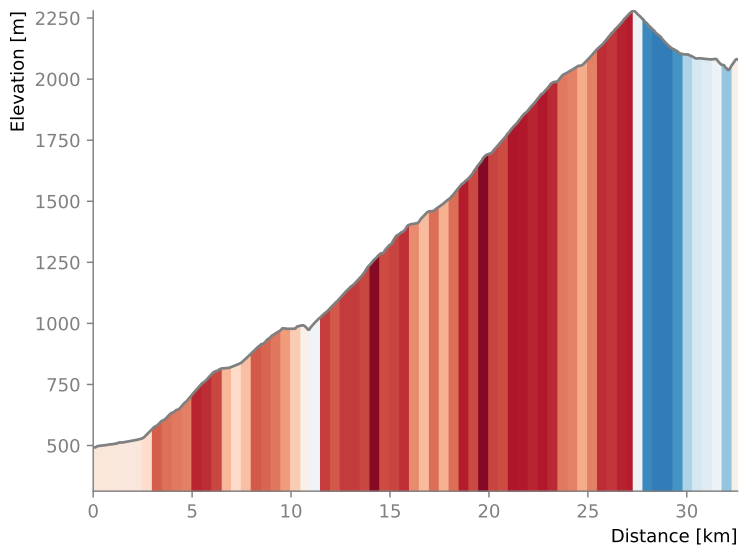

Figure A.8: A profile of the Col du Sanetsch as cycled by Will and Brendan. Figure from Will Kalderon.

The Col du Sanetsch is considered a "local secret" and many Valais residents consider it the best cycling road in Switzerland. Perhaps the most impressive part about this pass is the varied landscapes. It starts in the vineyards near Sion and then rounds into the forest under the steep cliffs characteristic of the Valais Alps. The road has very few cars making it the perfect playground for cyclists.

Figure A.9: The varied landscapes up the Col du Sanetsch, including a bit of "bicycle skiing".

As the road wiggles above the treeline, it offers impressive views of the nearby Alpine giants. After reaching the top, the road continues to a small dam where it ends. The

other side of the climb has a small cable car capable of carrying bicycles, allowing for long journeys into Vaud or Bern for the adventurous cyclist.

The col is high enough to receive early season snow; when the physicists visited the road from the top to the dam was covered in snow. Naturally, they tested out a new sport—bicycle skiing. This sport had limited success and may improve in future iterations.

6. Passo San Gottardo (2,106m)

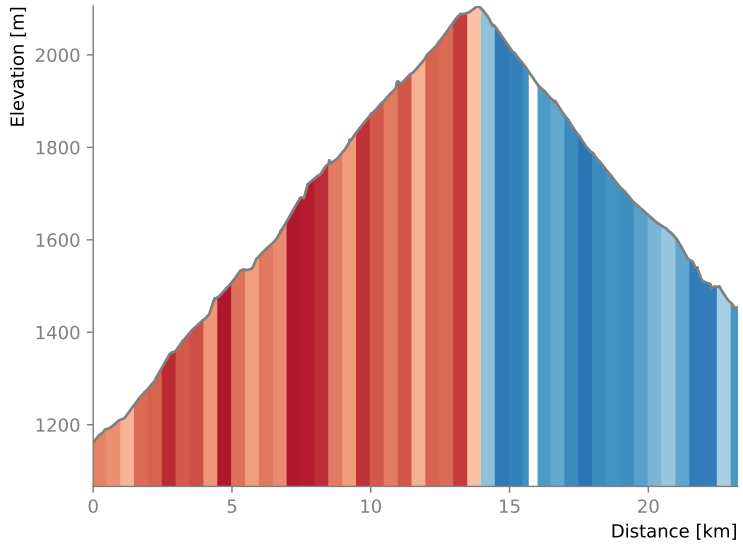

Figure A.10: A profile of the Passo San Gottardo (Gotthard pass) as cycled by Will and Brendan. Figure from Will Kalderon.

Passo San Gottardo gets the award for the prettiest hairpins in the Alps. However beauty comes with a price, these cobblestone bends are quite tiring on the arms during the ascent. This historic pass goes from the Italian speaking Swiss canton Ticino to the German speaking canton Uri. Underneath the pass is one of the world's longest railway tunnels (the Gotthard base tunnel) at 55 km long[4].

This climb starts in a beautiful Ticino valley with villages of pastel colored houses which characterize the region. Then, the road climbs through a forest in and out of cobblestone road. At multiple points the road splits, so it is important to stay on the

[4]The Gotthard base tunnel is an excellent way to travel to the Ticino canton for late season cycling adventures.

older, scenic route. Once above the treeline, the beautiful hairpins dominate the horizon as shown in Fig. A.11. The reviewers managed to cycle this road while the pass was closed to traffic, making for an epic car-free ascent.

Figure A.11: The cobblestone hairpins of Passo San Gottardo with reviewer Brendan Regnery (Left). This pass wins the prize for the prettiest hairpins in the Alps.

5. Passo Gavia (2,621m)

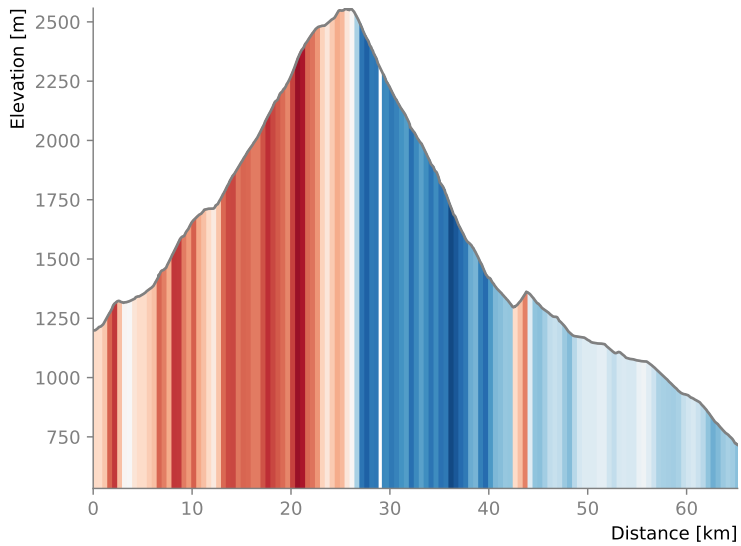

Figure A.12: A profile of the Passo Gavia as cycled by Brendan, Will, and Pete. Figure from Will Kalderon.

This wild Italian legend was made famous in the Giro d'Italia and is a brutal, rewarding climb. The initial climb winds through beautiful Italian towns characteristic of Lombardia, but then takes a turn into the remote mountains. The Passo Gavia gives a remote

mountain pass feel on a small road that winds through steep meadows of rocks near cliffs with a large amount of time spent above the treeline.

The top portion of this climb ends on a slowly rising plateau and is slightly anti-climatic. Figure A.13 shows one of the final rises on the climb. Then the road down to the valley is a steep narrow path, creating a very technical, exposed descent into a gorgeous village straddling a mountain river. The state of the roads was less than ideal, particularly on the descent. For this reason, the Passo Gavia is rated as number five on the list.

Figure A.13: One of the final rises in the Passo Gavia.

4. Grosse Scheidegg (1,962m)

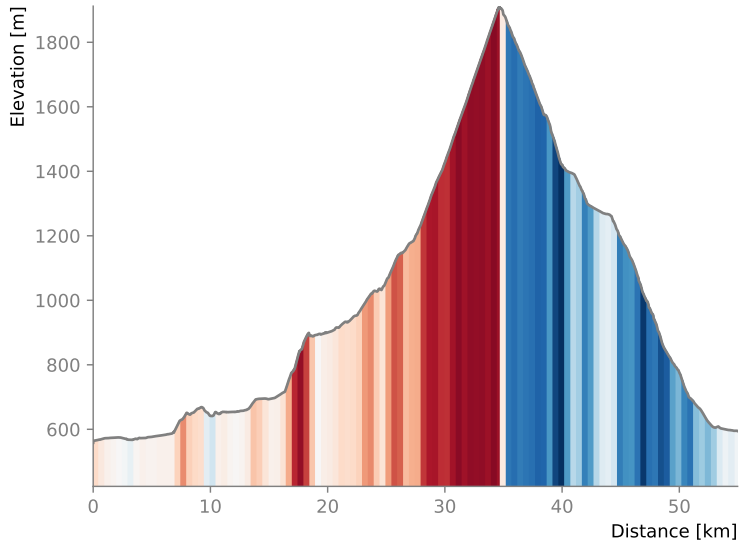

Figure A.14: A profile of the Grosse Scheidegg as cycled by Brendan. Figure from Will Kalderon.

This short, but steep, car-free road has some of the most phenomenal views of the Eiger, Mönch, and Jungfrau. The Bernese pass connects the famous ski town of Grindelwald to the village of Meiringen[5]. The pass wiggles through some alpine forests and lovely old Swiss villages, while maintaining a constant presence in the shadow of the Swiss giants. The steepness adds to the challenge while the impressive north face of the Eiger serves as a constant reminder that bigger challenges are not far away, as shown in Fig. A.15.

Physicist Brendan cycled over this climb on a two day bike packing trip from Geneva to Zurich. It is possible to cycle between these cities without cycling over giant climbs in the Alps, but why would anyone do that? After all, the Grosse Scheidegg is only a short detour and is the highlight of the route. While cycling this route, Brendan had a difficult time finding dinner due to Swiss restaurants closing at 9 pm. Thankfully, he was saved by a very kind Italian restaurant at 9:15 pm. This highlights another important life lesson: He who learns Italian shall never go hungry.

[5]Meiringen is famous for the nearby Reichenbach Falls where Sir Arthur Conan Doyle's Sherlock Holmes dramatically faces off with the infamous Moriarty.

Figure A.15: Physicist Brendan Regnery stopping by the Grosse Scheidegg while cycling from Geneva to Zurich. Off to the left of his shoulder is the Eiger.

3. Passo dello Stelvio (2,757m)

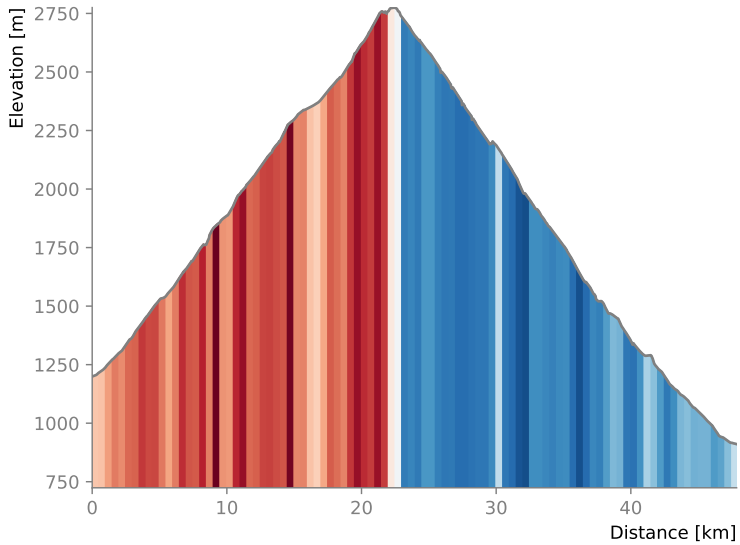

Figure A.16: A profile of the Passo dello Stelvio as cycled by Brendan, Will, and Pete. Figure from Will Kalderon.

Passo dello Stelvio is the mythical climb from the Giro d'Italia and the second highest paved pass in the Alps. This climb has been the scene of many Giro successes and controversies, making it one of the most legendary climbs in the world. The pass begins in the beautiful town of Bormio, where it wiggles through many different types of terrain: steep cliffs, long fields, beautiful water falls, and alpine vistas. The climb ends at a pass filled with souvenir stands with merchandise rippling in the wind—creating a true sense of being on top of the world. . . or at least the cycling world.

On north side of the pass is the beautiful, German speaking village of Stilfs in the Sud Tyrol region of Italy. The third, western side of the pass comes from Switzerland via the Umbrail pass. This is another cycling must when spending time in the region and it also marks a "Dreisprachenspitze"[6] between German (in Südtirol, Italy), Italian (in Lombardia, Italy), and Romansch (in Graubünden, Switzerland).

The physicists were accompanied on this pass by Peter Bond. Together, they battled their way to the top during a fantastic, car-free day. After a successful day, they feasted

[6]Three language peak.

on a new found local favorite—pizzocheri. When asked about this delicious food, Brendan said simply "it is the best[7]."

Figure A.17: Sua Eccellenza (His majesty) il Passo dello Stelvio! And some guy spoiling the view.

[7]Once again, this does not refer to the BEST network.

2. Colle di Nivolet (2,641m)

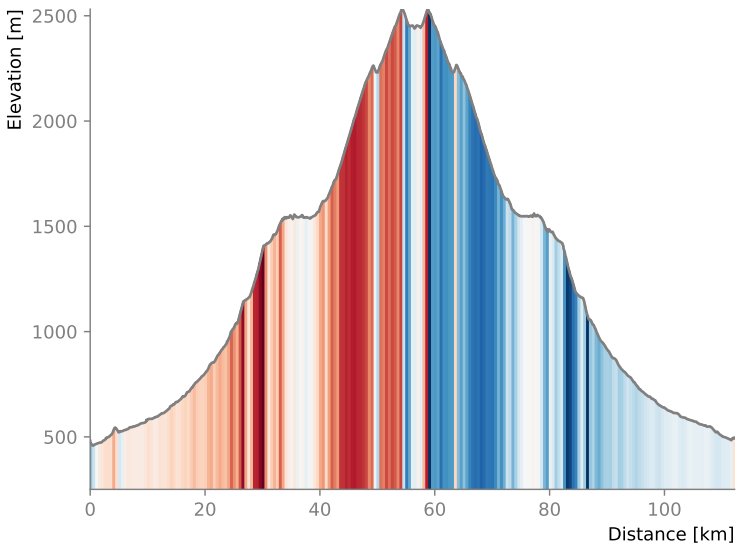

Figure A.18: A profile of the Colle di Nivolet as cycled by Will and Brendan. Figure from Will Kalderon.

Perhaps the "prettiest climb" on the list. Colle di Nivolet is a long climb, from start to finish it is over 2300 meters in elevation gain. Located in Piemonte, Italy this climb wiggles past a series of dams that generate power for the nearby city of Torino while traveling through a number of towns frequently visited by the Giro d'Italia.

The road starts in the hot, flat entrance to the alps. There is a crucial small side road that avoids an annoying tunnel while climbing up a beautiful valley. Once at the top, the climbing really starts. The climb goes by numerous dams that supply electricity to the city of Torino and each dam makes a pretty alpine lake. The final stretch of road is frequently closed to cars in the summer (make sure to check the times).

With every hair pin is a better and better view. The very top has a view of the Gran Paradiso and just beyond the top is a nice, wind sheltered lake where tired cyclists can take a refreshing swim before descending down. The long descent has many fantastic restaurants! Perfect for refueling after the long ride.

Figure A.19: The “prettiest climb” in the Alps: the Colle di Nivolet.

1. The Albula Pass (2,312m)

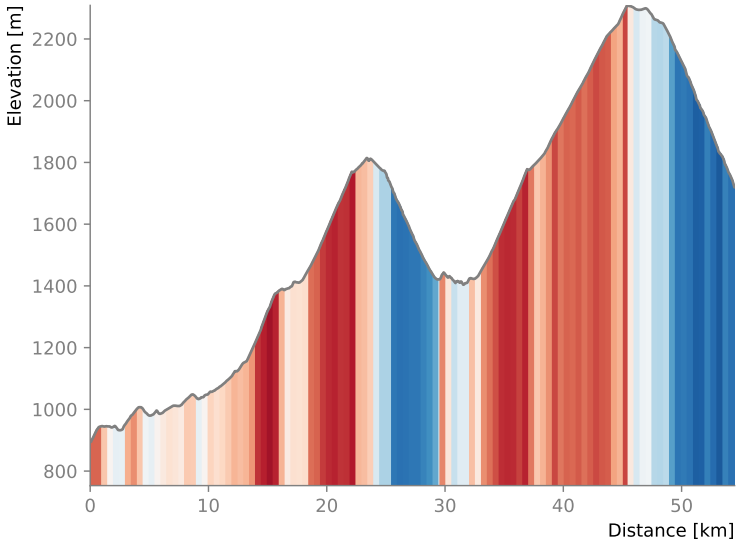

Figure A.20: A profile of the Albula pass as cycled by Will and Brendan. Including a scenic descent to watch the world's longest passenger train. Figure from Will Kalderon.

This Swiss pass in the Romansch speaking canton of Graubünden simply must be done from Filisur to La Punt. The ascent starts like many alpine ascents, through a lovely valley and then up beside a gorge to the town of Bergün. Here, the ascent has a special treat for train-lovers. This whole pass is filled with a long train track that spirals through tunnels and viaducts to support the Rhaetian Railway, making this a UNESCO world heritage site.

As the climb twists upward, the flora and fauna dramatically change, so the rider can enjoy many different tree species before climbing above the treeline. Once above the treeline, the scenery becomes true alpine terrain and hidden among the rocks are small bunkers for Switzerland to defend their country in case of an invasion. The climb finishes with a lovely view of Engadin mountains with a short descent into La Punt. As an added bonus, the Engadin is home to one of the world's best deserts: nüßtorte. Each family tightly guards their nüßtorte recipes, so each one is unique and worth trying.

Physicists Will and Brendan may have been slightly biased as they choose another car-free day to do this pass, but not just any car-free day. This was the day that the

Rhaetian railway decided to make the world's longest passenger train. This 1.91km train traveled slowly down the pass while the reviewers raced beside it from the Albula tunnel all the way to Bergün. Often, the train would occupy multiple viaducts and spiral tunnels at the same time. Figure A.21 shows the reviewers featured in a picture from "lematin.ch" news. Perhaps an even longer train can be arranged when CERN director general Fabiola Gianotti comes here to cycle the pass!

Figure A.21: Physicists Will Kalderon (blue) and Brendan Regnery (orange) featured in "lematin.ch" while racing down the Albula pass beside the world's longest passenger train. Figure from [46].

Conclusion

Cycling is a way to connect cultures, relieve stress, think about physics in new ways, and connect with one's inner self. Once, professor John Conway gave advice to a young student:

"I have solved the most difficult physics problems while simply cycling through the orchards near Davis, CA. Taking the time to let your mind wander while your legs spin will let you solve problems in ways you never imagined."

Appendix B

The BEST Images

ImageBEST was created to extract low and high level features from Jets at CMS. The method for creating these images is defined in Section 8.4 and averaged images for each category are shown below. Unfortunately, this method did not result in significant performance increase.

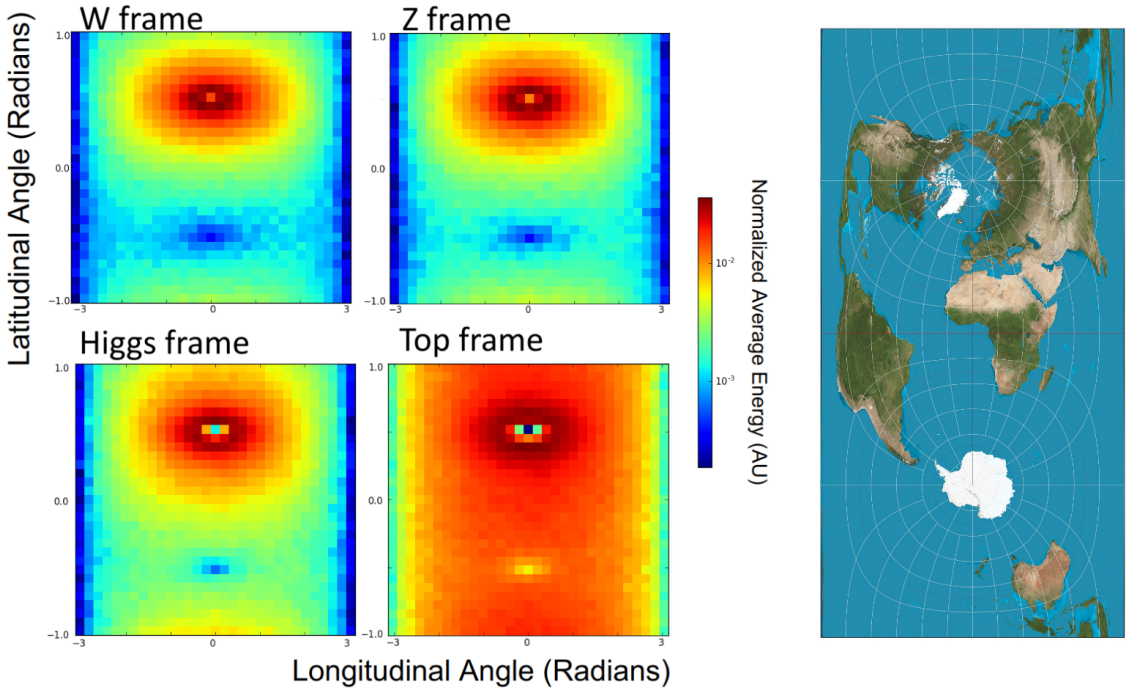

Figure B.1: Averaged images of over 100,000 simulated top jets created with the BEST image method. There is one image per BEST rest frame.

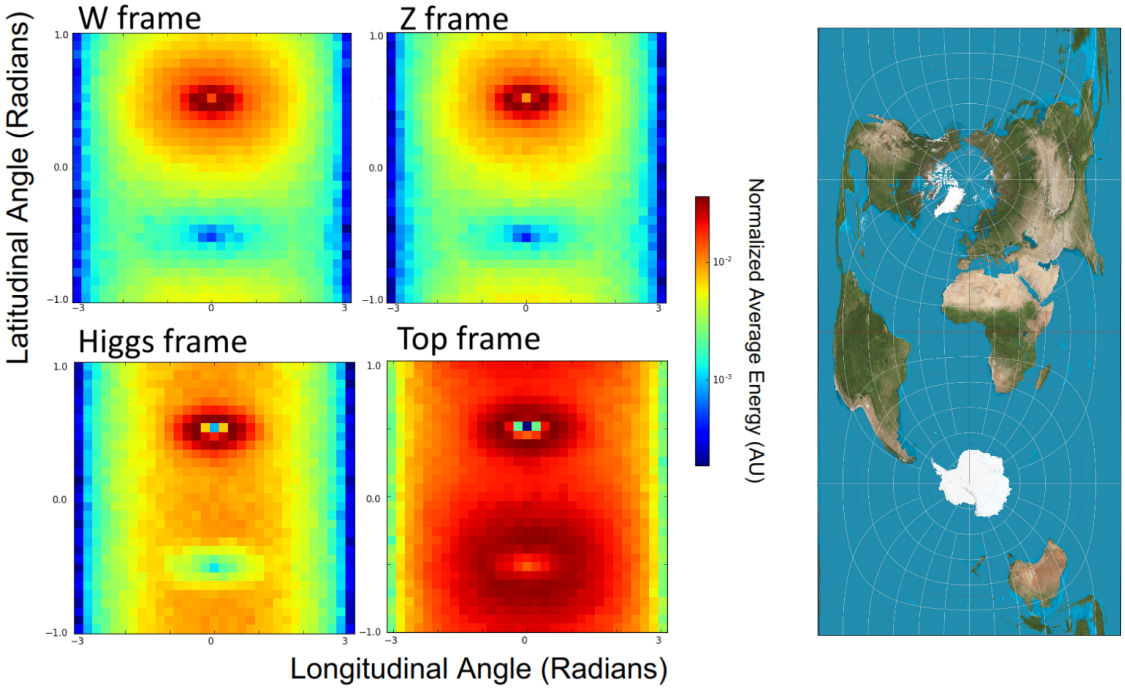

Figure B.2: Averaged images of over 100,000 simulated Higgs jets created with the BEST image method. There is one image per BEST rest frame.

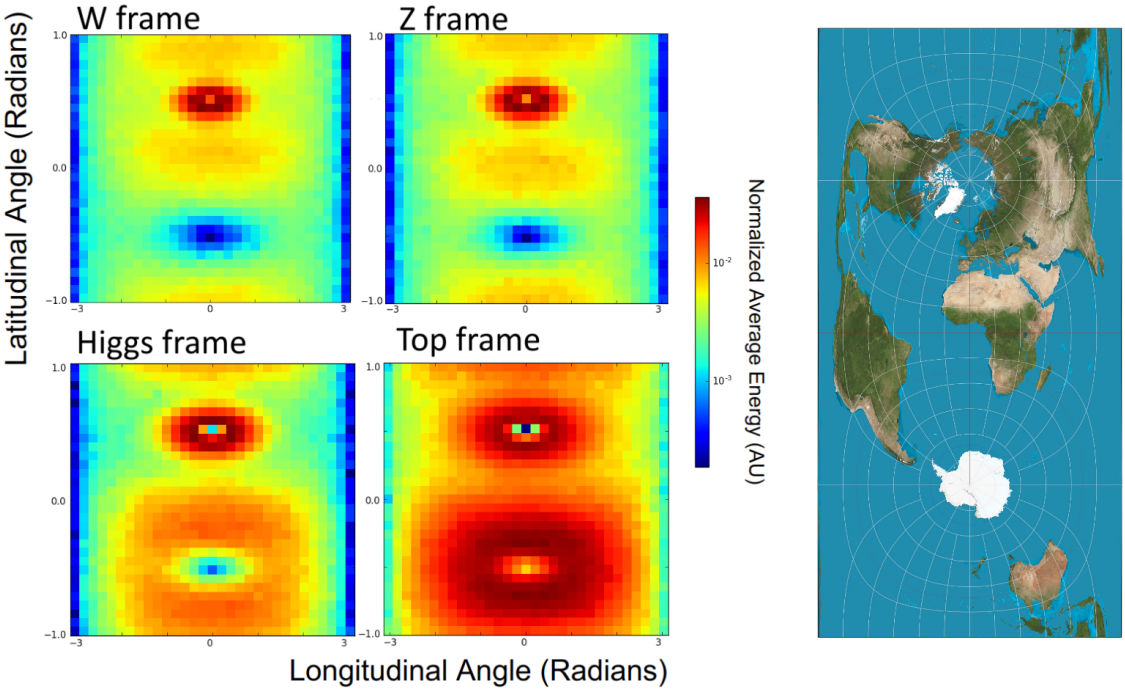

Figure B.3: Averaged images of over 100,000 simulated Z jets created with the BEST image method. There is one image per BEST rest frame.

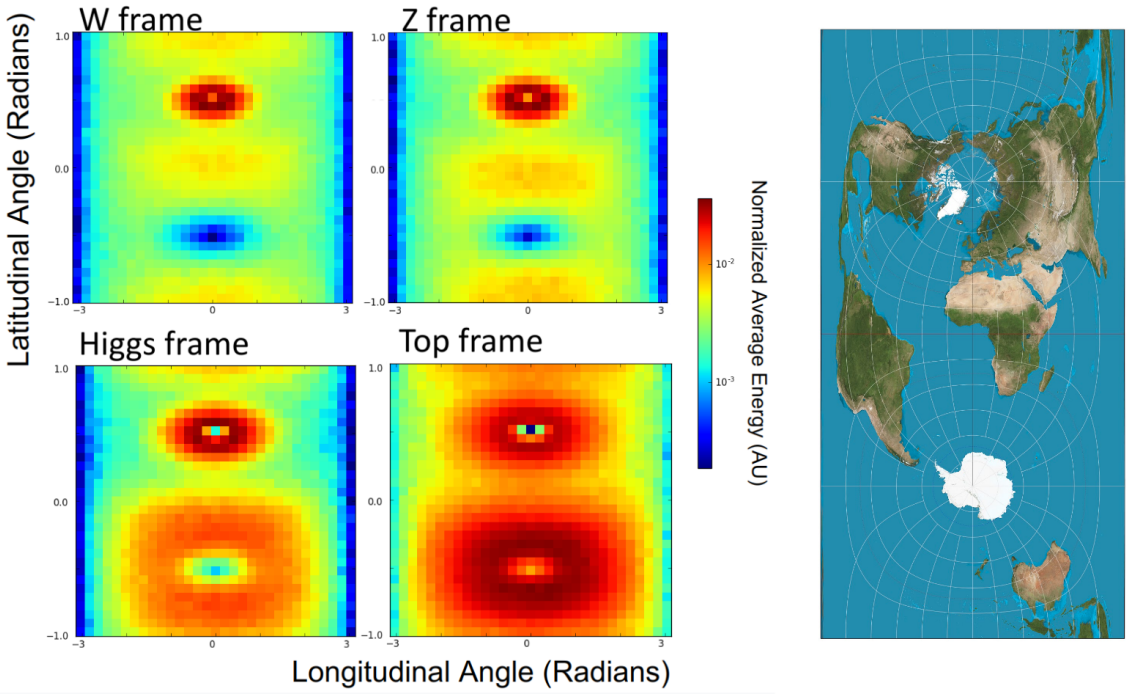

Figure B.4: Averaged images of over 100,000 simulated W jets created with the BEST image method. There is one image per BEST rest frame.

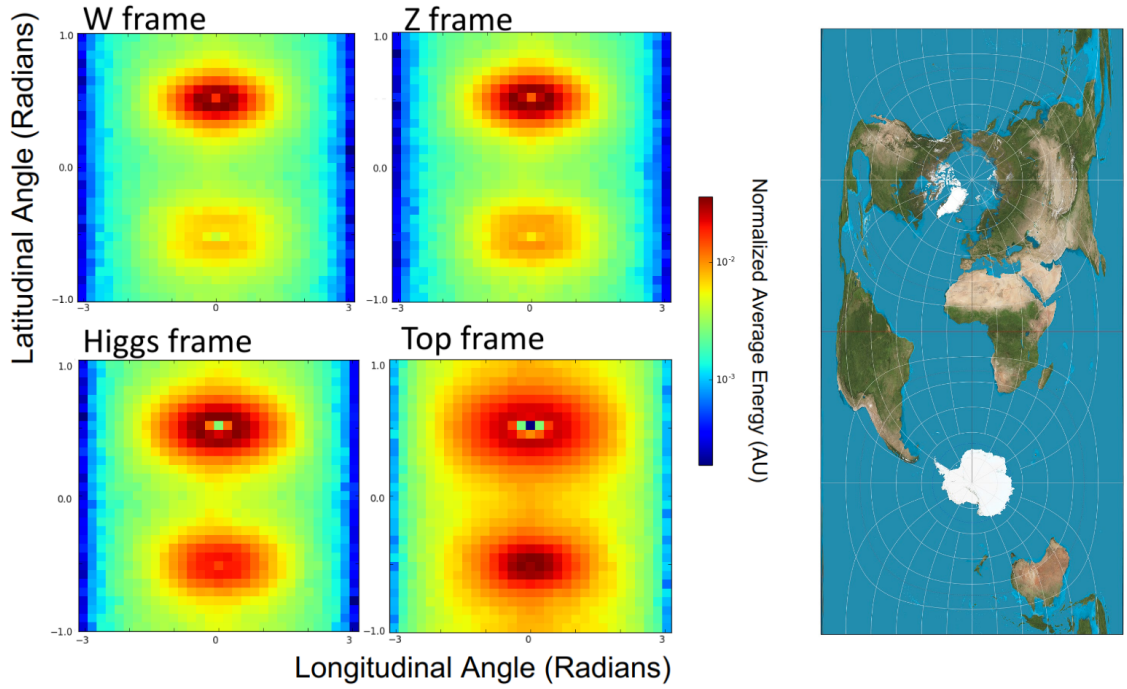

Figure B.5: Averaged images of over 100,000 simulated bottom jets created with the BEST image method. There is one image per BEST rest frame.

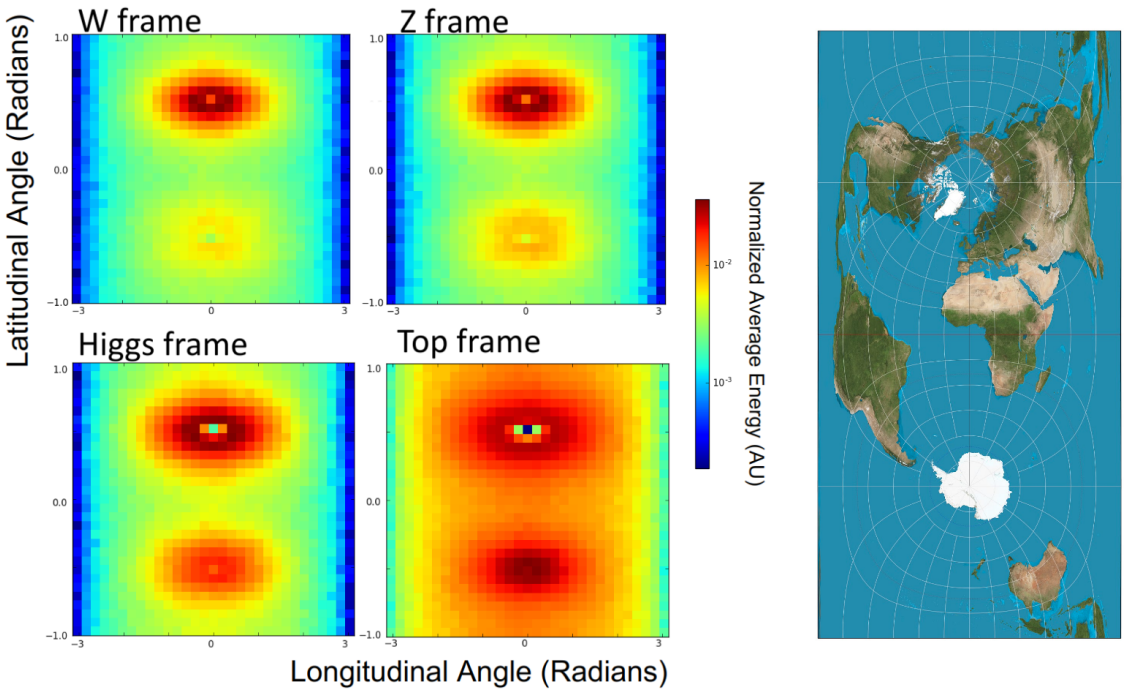

Figure B.6: Averaged images of over 100,000 simulated QCD jets created with the BEST image method. There is one image per BEST rest frame.

Appendix C

About the Author

Brendan is a skier, cyclist, hiker, and (oh yeah...) physicist, who is fervently trying to learn Italian, French, and German (but can't decide which to prioritize). He spends his time working with an experiment 100 meters underground or thousands of meters above sea level in the high Alps. While he was based at CERN for his PhD research, he cycled over 19,000 km. As a result he is often introduced as

> *"Hey, do you know Brendan?"*
> *"Do you mean the cyclist?"*
> *"Yeah, that guy"*

To the president of the CERN Ski Club, alpine section, he is known as

> *"that guy who signs up for every outing way too fast"*

And to the Italians at CERN he is known as

> *"that guy that keeps forcing me to listen to him practice his Italian while he rambles on about Pizzocheri and Taleggio... Ma che... I am not even from Lombardia... I prefer real pasta..."*

Figure C.1: Brendan Regnery