

Measurement of $H \rightarrow b\bar{b}$ in Associated Production with the CMS Detector

by

Daniel Robert Abercrombie

B.S., Pennsylvania State University (2014)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

© Massachusetts Institute of Technology 2020. All rights reserved.

Author
Department of Physics
July 22, 2020

Certified by
Christoph M. E. Paus
Professor of Physics
Thesis Supervisor

Accepted by
Nergis Mavalvala
Associate Department Head, Physics

Measurement of $H \rightarrow b\bar{b}$ in Associated Production with the CMS Detector

by

Daniel Robert Abercrombie

Submitted to the Department of Physics
on July 22, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

We measured $VH \rightarrow b\bar{b}$ with the CMS Detector.

Thesis Supervisor: Christoph M. E. Paus

Title: Professor of Physics

Acknowledgments

Thanks.

Contents

1	Introduction	11
1.1	Measurement of the Higgs Cross Section	12
1.2	Motivation for the Measurement	12
1.3	Historic Context	13
1.4	Using the CMS Detector	13
2	Theory	15
2.1	The Higgs Mechanism	16
2.2	Associated Production	18
2.2.1	Production Mechanisms of Vector Bosons	19
2.2.2	Decay Channels of Vector Bosons	24
2.3	Decay Channels of the Higgs	25
2.4	Other Relevant Standard Model Processes	27
3	The CMS Detector	29
3.1	Associated Production at the LHC	29
3.2	Detector Requirements	31
3.3	Detector Design	32
3.3.1	Solenoid Magnet	34
3.3.2	Silicon Pixel Detector	34
3.3.3	Electromagnetic Calorimeter	35
3.3.4	Hadronic Calorimeter	36
3.3.5	Muon Chambers	37

3.4	Event Reconstruction	37
3.4.1	Charged Particle Tracks	38
3.4.2	Calorimeters	39
3.4.3	Linking and Particle Identification	39
3.5	Trigger	41
3.6	Accessing Data	42
4	Simulation	43
4.1	Backgrounds to the Analysis	43
4.2	Event Generation	43
4.2.1	Tree Level Simulation	44
4.2.2	Parton Showers	44
4.3	Detector Simulation	44
4.4	Corrections to Simulation	44
4.4.1	Smearing	44
5	Event Selection	45
5.1	Object Definitions	45
5.1.1	Variable Definitions	46
5.1.2	Muons	47
5.1.3	Electrons	48
5.1.4	Jets	49
5.1.5	MET	50
5.1.6	Undesirable Particles	50
5.2	Removal of QCD	51
5.3	Categories of Vector Boson Decay	51
5.3.1	0 Leptons	51
5.3.2	1 Lepton	51
5.3.3	2 Leptons	51
5.4	Topology of Higgs Decay	51
5.4.1	Resolved Jets	51

5.4.2	Boosted Jet	51
5.5	Control Regions	52
5.5.1	Light Flavor Jets	52
5.5.2	Heavy Flavor Jets	52
5.5.3	$t\bar{t}$	52
6	Analysis Results	53
6.1	Systematic Uncertainties	53
6.2	Combination Fit Method	53
6.3	Results	53
7	Conclusions	55
A	Detector Projects	57
A.1	Dynamo Consistency	57
A.2	Workflow Web Tools	58
B	Physics Calculations	59
C	Data Format	61
D	Generator Parameters	63
E	Data Card	65

Chapter 1

Introduction

One of the most curious features of physics at small scales, which will likely frustrate students for the rest of time, is that certain events are not deterministic and only have a probability of happening. There is no guarantee that an electron and a positron approaching each other at high energies will annihilate and produce a muon and an anti-muon. However, this event might still occur at a later time at with the exact same initial conditions. Furthermore, the observation of resonances, where this is more likely to happen when the electron and positron approach each other at particular speeds, does not mean that a Z boson was present in a given interaction. It just means that the weak component of the electroweak force significantly increases the probability of the muonic final state, given the total energy of the initial state. The sum of probabilities from different possible field interactions with particular initial conditions is the only thing we can measure. This is also only possible when observing many events with the same initial conditions.

This point is difficult to convey concisely, so many laypeople, as well as some practicing physicists, are confused by the terminology adopted by the field. But this distinction is relevant to the topic of this work. This document presents a measurement of a cross section. Cross section is the name given to the probability of an interaction occurring. Reported cross sections can be split up to describe different contributions to final states, and they can be collated into what are called “production cross sections” which describe the probabilities of particular intermediate states

“occurring” (even though intermediate states never exist in reality).

The main point is that if there exists some interesting particle, and it interacts with other particles, you can see an increased probability of certain initial states resulting in certain final states. This can teach the observer about the role of the interesting particle, without ever directly seeing it.

1.1 Measurement of the Higgs Cross Section

The purpose of the following document is to present the methods and results of measuring the strength of the coupling between the Higgs boson and bottom quarks. In this context, the Higgs boson makes up one of the previously mentioned intermediate states that cannot be shown as present in a given event. The cross section measurement relies on a number of physics processes that will be accounted for in this document.

To measure this coupling, the Higgs boson must first be “produced” before measuring its coupling strength to bottom quarks. Since we are not technologically advanced enough to achieve this generation using bottom quarks directly, we use measured Higgs generation rates from normal constituents of protons. We constrain ourselves further by requiring that the Higgs is generated by associated production.

After the Higgs is generated, it can decay into a number of different particles. This work is only concerned with one kind of decay.

The math that this all relies on is presented in Chapter 2.

1.2 Motivation for the Measurement

This is Thomas Kuhn’s “normal science”.

Precision measurements are needed to be certain of what we think.

Precision measurements often lead to discrepancies that are explained by a fundamental shift in the model.

The Standard Model is a good one. It will not be fully replaced, but at worst

expanded upon. Just like Newton's Laws are still a reasonable approximation for General Relativity, The Standard Model is a good approximation for most things we have been able to interact with so far.

The only lingering questions are Dark Matter and Dark Energy, but there is no reason to assume that precise measurements of known phenomena will not lead to an explanation.

1.3 Historic Context

First, we have The Standard Model.

Parts were proven correct by the observation of the weak bosons.

The Higgs was observed in 2013.

The Higgs decaying to $b\bar{b}$ was observed in 2018. [1]

1.4 Using the CMS Detector

The CMS detector is a general purpose detector used to make many observations of conditions unattainable on Earth outside of the LHC.

It has many stationary parts, and a couple of moving ones too.

This device is described in detail in Chapter 3.

Chapter 2

Theory

Before diving into the description of the experimental apparatus, an explanation of why it is expected to work is needed. There are many textbooks that cover the Standard Model, as there are many students who study it. Much of what follows is taken from the book by Mark Thompson [2].

The Standard Model Lagrangian can be defined as a sum of Lagrangians that each describe the interactions between different fermions and bosons. Equations of motion can be extracted from a Lagrangian \mathcal{L} for a particle field ϕ_i using the Euler-Lagrange equations.

$$\delta_\mu \left(\frac{\delta \mathcal{L}}{\delta(\delta_\mu \phi_i)} \right) - \frac{\delta \mathcal{L}}{\delta \phi_i} = 0 \quad (2.1)$$

In the measurement of $H \rightarrow b\bar{b}$ in associated production, many components of the Standard Model are of interest. These will be introduced as needed. First, I will give a brief explanation of Higgs field's non-zero vacuum energy, a trait that makes the Higgs one of the central keystones to Standard Model. After that, the electroweak Lagrangian will be described since the cross section of associated production depends on the coupling of the Higgs boson to the W and Z vector bosons. The coupling of the electroweak force to fermions is also important to understand both the generation of these intermediate states and the resulting final state that the CMS detector records.

Another important factor for this work is a the decay of the Higgs boson itself into bottom quarks. This depends on the Higgs directly coupling to fermions. Finally, we will briefly consider the part of the Lagrangian describing the strong force. Since the LHC is a hadron collider, understanding of the strong force is required to extract data from LHC collisions.

2.1 The Higgs Mechanism

In both of these components of the Standard Model Lagrangian, the Higgs coupling actually gives the vector bosons and massive fermions their mass [3–5]. (In this work, neutrinos can be treated as massless.) The granting of mass happens for two reasons: the Higgs field has a non-zero vacuum expectation value, and the Higgs field couples to vector boson and massive fermion fields.

The Higgs can be described as two complex scalar fields in a weak isospin doublet with a quartic potential. The Lagrangian for a free Higgs is then

$$\mathcal{L} = (\delta_\mu \phi)^\dagger (\delta^\mu \phi) - (\mu^2 (\phi^\dagger \phi) + \lambda (\phi^\dagger \phi)^2) \quad (2.2)$$

Through the virial theorem, the potential has a minimum value when

$$\phi^\dagger \phi = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2} \quad (2.3)$$

This potential of the Higgs field breaks the $SU(2) \times U(1)$ symmetry of the Standard Model Lagrangian. Through this non-zero vacuum expectation value, the Higgs then has a constant influence in other parts of the Standard Model Lagrangian. For this measurement, three interactions that the Higgs makes with this influence need to be considered: the Higgs interacting with itself, the Higgs interacting with the electroweak vector bosons, and the Higgs interacting with quarks.

The first two interactions manifest in the Lagrangian when we force the $SU(2) \times U(1)$ symmetry on the Lagrangian in Equation (2.2). The derivatives must be

replaced.

$$\delta_\mu \rightarrow D_\mu = \delta_\mu + i\frac{g_W}{2}\boldsymbol{\sigma} \cdot \mathbf{W}_\mu + ig'\frac{Y}{2}B_\mu \quad (2.4)$$

ϕ can be rewritten to satisfy the vacuum expectation value in the gauge that will give us the massless neutral boson known as a photon.

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (2.5)$$

This leads to the following expansion for the kinetic term of the Lagrangian.

$$\begin{aligned} (D_\mu \phi)^\dagger (D^\mu \phi) &= \frac{1}{2}(\delta_\mu h)(\delta^\mu h) + \frac{1}{8}g_W^2(W_\mu^{(1)} + iW_\mu^{(2)})(W^{(1)\mu} - iW^{(2)\mu})(v + h)^2 \\ &\quad + \frac{1}{8}(g_W W_\mu^{(3)} - g' B_\mu)(g_W W^{(3)\mu} - g' B^\mu)(v + h)^2 \end{aligned} \quad (2.6)$$

Terms that are quadratic in terms of the gauge boson fields reveal the mass of the fields. Taking $h(x) \rightarrow 0$, the terms for $W^{(1)}$ and $W^{(2)}$ are the just

$$\frac{1}{4}g_W^2 v^2 W_\mu^{(1)} W^{(1)\mu} \quad \text{and} \quad \frac{1}{4}g_W^2 v^2 W_\mu^{(2)} W^{(2)\mu},$$

giving the mass.

$$m_W = \frac{1}{2}g_W v \quad (2.7)$$

The quadratic terms for $W^{(3)}$ and B mix to give a non-diagonal mass matrix \mathbf{M} .

$$\frac{v^2}{8} \begin{pmatrix} W_\mu^{(3)} & B_\mu \end{pmatrix} \mathbf{M} \begin{pmatrix} W^{(3)\mu} \\ B^\mu \end{pmatrix} = \frac{v^2}{8} \begin{pmatrix} W_\mu^{(3)} & B_\mu \end{pmatrix} \begin{pmatrix} g_W^2 & -g_W g' \\ -g_W g' & g'^2 \end{pmatrix} \begin{pmatrix} W^{(3)\mu} \\ B^\mu \end{pmatrix} \quad (2.8)$$

The non-diagonal matrix allow $W^{(3)}$ and B to mix. Physical states must be represented by a diagonal Hamiltonian. Diagonalizing the term above gives masses of the

physical states.

$$\frac{1}{8}v^2 \begin{pmatrix} A_\mu & Z_\mu \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & g_W^2 + g'^2 \end{pmatrix} \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} = \frac{1}{2} \begin{pmatrix} A_\mu & Z_\mu \end{pmatrix} \begin{pmatrix} m_A^2 & 0 \\ 0 & m_Z^2 \end{pmatrix} \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} \quad (2.9)$$

This gives us the masses of the neutral gauge bosons.

$$m_A = 0 \quad \text{and} \quad m_Z = \frac{1}{2}v\sqrt{g_W^2 + g'^2} \quad (2.10)$$

From the simple act of requiring $\text{SU}(2) \times \text{U}(1)$ symmetry on the Lagrangian of a scalar doublet with non-zero vacuum expectation value, the masses of all the electroweak gauge bosons have been produced.

2.2 Associated Production

The next thing to consider is the couplings also produced by this process. The couplings will allow us to determine more precisely the parameters above by measuring cross sections.

The physical states of W^+ and W^- bosons can be written as the raising and lowering operators for isospin.

$$W^\pm = \frac{1}{\sqrt{2}} (W^{(1)} \mp iW^{(2)}) \quad (2.11)$$

The second term of Equation (2.6) can be further expanded.

$$\frac{1}{4}g_W^2 W_\mu^- W^{+\mu} (v + h)^2 = \frac{1}{4}g_W^2 v^2 W_\mu^- W^{+\mu} + \frac{1}{2}g_W^2 v W_\mu^- W^{+\mu} h + \frac{1}{4}g_W^2 W_\mu^- W^{+\mu} h^2 \quad (2.12)$$

The second term on the right hand side of Equation (2.12) gives us the coupling strength of a vertex with a Higgs and two W bosons.

$$g_{HWW} = \frac{1}{2}g_W^2 v = g_W m_W \quad (2.13)$$

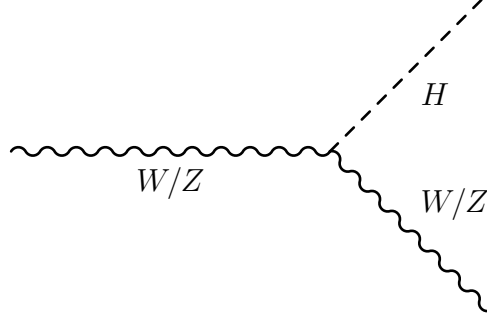


Figure 2-1: Above is the Feynman diagram for associated production. The W or Z boson radiates a Higgs boson. Both bosons later decay into particles detected by CMS.

The coupling to the Z boson can also be found from Equation (2.9) by substituting $(v + h)^2$ back in for v^2 and extracting the terms proportional to $hZ_\mu Z^\mu$.

$$g_{HZZ} = \frac{1}{2} (g_W^2 + g'^2) v = \sqrt{g_W^2 + g'^2} m_Z \quad (2.14)$$

When arranged in a way that the W or Z boson radiates the Higgs, as opposed to a Higgs decaying into a pair of W or Z bosons, the process is called associated production or *Higgstrahlung*. The vertex showing associated production is pictured in Figure 2-1.

2.2.1 Production Mechanisms of Vector Bosons

The W and Z bosons are themselves intermediate states, never existing in a directly observable manner. They must be produced through interacts with stable fermions. Since the LHC is a hadron collider, considering the vector bosons' couplings with quarks would be most relevant.

Quarks are fermions that couple to each other through the strong force, resulting from a $SU(3)$ symmetry. There are three generations of quarks each consisting of a pair of quark types. Their mass eigenstates are denoted as down-type or up-type. Table 2.1 displays some of the characteristics of these quarks. A feature of quarks is that their mass eigenstates do not match their weak eigenstates. There is a mixing among the down-type quarks that is parametrized by the Cabibbo-Kobayashi-

Table 2.1: The quarks and some stuff about them.

	1st gen.	2nd gen.	3rd gen.	Q	$I_W^{(3)}$
down-type	d	s	b	$-\frac{1}{3}$	$-\frac{1}{2}$
up-type	u	c	t	$+\frac{2}{3}$	$+\frac{1}{2}$

Maskawa (CKM) matrix.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2.15)$$

The mass eigenstates are denoted as d, s , and b , while d', s' , and b' are the weak eigenstates. This mixing allows quarks to change generations through interaction with W^\pm bosons, which raise or lower the weak isospin. The following is the charge current vertex interaction.

$$-i \frac{g_W}{\sqrt{2}} \begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix} \gamma^\mu \frac{1}{2} (1 - \gamma^5) \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

The vertices for this interaction is shown in Figure 2-2 arranged in a way to show the processes of generating a W^+ or W^- boson from annihilating quarks. The γ matrices in the interaction are present because the $SU(2)$ component of the Standard Model only interacts with left-handed fermions and right-handed anti-fermions. For this reason, the $SU(2)$ component is more accurately labelled $SU(2)_L$. From Equation (2.11), the W^\pm bosons are completely made up of the $W^{(1)}$ and $W^{(2)}$ components of the $SU(2)_L$, so they also only interact with left-handed fermions and right-handed anti-fermions.

Both the photon and the Z boson mix the $SU(2)_L$ and $U(1)$ components of the Standard Model. Production of the Z needs to be directly understood for this measurement, but it is more straightforward to determine the strength of the Z couplings

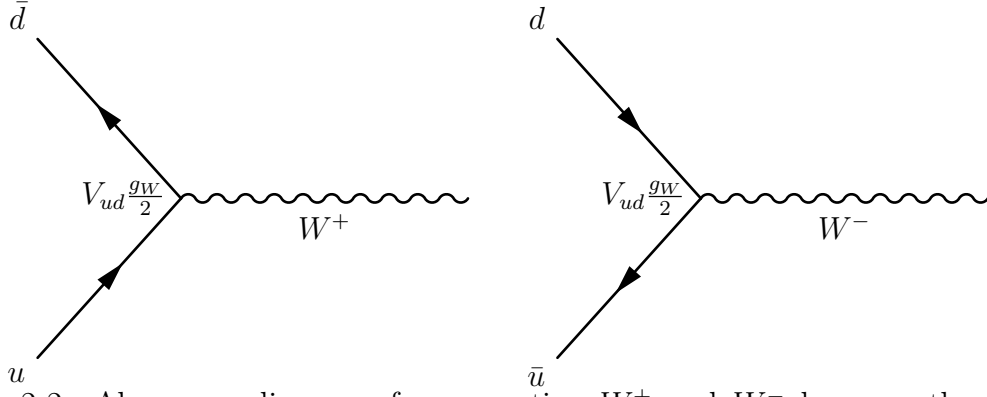


Figure 2-2: Above are diagrams for generating W^+ and W^- bosons. the u and d quarks in the diagram can be replaced with any up-type or down-type quark, respectively. The CKM matrix element would in the vertex element would be changed accordingly.

to left- and right-handed fermions by exploiting the symmetry of photon interactions. That is, the photon interacts the same with left and right handed charged fermions, and not at all with neutral fermions. This is shown directly with experiments with leptons. The charged leptons, electrons, muons, and taus, interact with photons, while the respective neutrinos do not. From the mixing in Equation (2.8), the photon and Z fields can be expressed as the following.

$$A_\mu = B_\mu \cos \theta_W + W_\mu^{(3)} \sin \theta_W \quad (2.16)$$

$$Z_\mu = -B_\mu \sin \theta_W + W_\mu^{(3)} \cos \theta_W \quad (2.17)$$

θ_W is known as the weak mixing angle. The relative strengths of the B and $W^{(3)}$ couplings are determined directly through lepton electro-magnetic characteristics, keeping in mind that $W^{(3)}$ only interacts with left handed particles. The following are the electro-magnetic interaction strengths of left- and right-handed electrons and neutri-

nos.

$$e_L : \quad Qe = \frac{1}{2}g'Y_{e_L} \cos \theta_W - \frac{1}{2}g_W \sin \theta_W \quad (2.18)$$

$$\nu_L : \quad 0 = \frac{1}{2}g'Y_{\nu_L} \cos \theta_W - \frac{1}{2}g_W \sin \theta_W \quad (2.19)$$

$$e_R : \quad Qe = \frac{1}{2}g'Y_{e_R} \cos \theta_W \quad (2.20)$$

$$\nu_R : \quad 0 = \frac{1}{2}g'Y_{\nu_R} \cos \theta_W \quad (2.21)$$

Y_{e_L} and Y_{ν_L} must be equal to maintain $SU(2)_L$ symmetry. To satisfy these constraints, the follow definition of Y is needed.

$$Y = 2 \left(Q - I_W^{(3)} \right) \quad (2.22)$$

The following relationship also arises from these experimental constraints.

$$e = g_W \sin \theta_W = g' \cos \theta_W \quad (2.23)$$

Returning to the Z boson, from Equation (2.17), and defining

$$g_Z = \frac{e}{\sin \theta_W \cos \theta_W}, \quad (2.24)$$

we have the following couplings to left- and right-handed fermions.

$$\begin{aligned} -\frac{1}{2}g' \sin \theta_W (Y_{f_L} \bar{u}_L \gamma^\mu u_L + Y_{f_R} \bar{u}_R \gamma^\mu u_R) + I_W^{(3)} g_W \cos \theta_W (\bar{u}_L \gamma^\mu u_L) = \\ g_Z \left((I^{(3)} - Q \sin^2 \theta_W) \bar{u}_L \gamma^\mu u_L - Q \sin^2 \theta_W \bar{u}_R \gamma^\mu u_R \right) \end{aligned} \quad (2.25)$$

Now the coupling of the Z to left- and right-handed quarks can be calculated from Table 2.1, remembering that $I_W^{(3)}$ for right-handed fermions is 0. Diagrams showing the interaction strengths of fermion- Z vertices are shown in Figure 2-3.

Thus vector bosons couple to quarks, the constituents of hadrons, which means they can be produced at the LHC. As mentioned earlier in this section, quarks interact

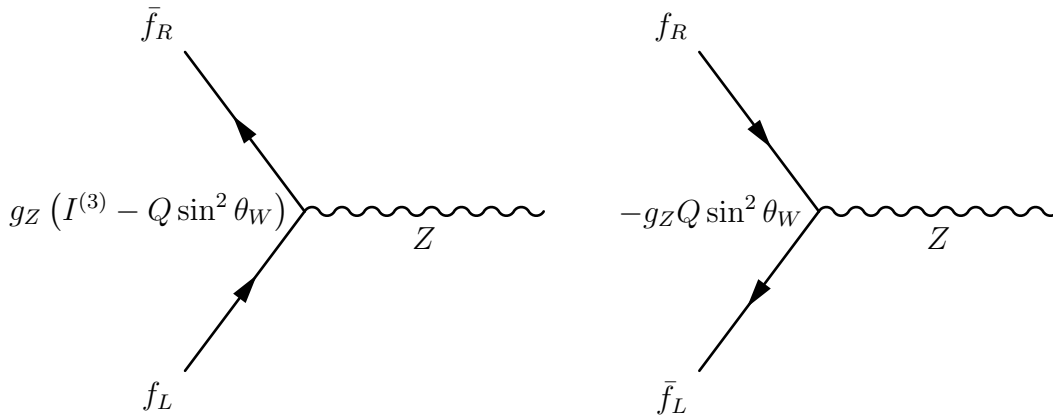


Figure 2-3: Above are diagrams for generating Z bosons. Left- and right-handed fermions are both coupled to, but with different coupling strengths.

through an $SU(3)$ symmetry that results in the strong force. The three states that this symmetry supports are known as color states, and they are labelled red, green, and blue, or r , g , and b . The resulting gauge bosons are known as gluons, and they carry the following color states.

$$r\bar{g}, g\bar{r}, r\bar{b}, b\bar{r}, g\bar{b}, b\bar{g}, \frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g}) \text{ and } \frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} - 2b\bar{b})$$

At low energy, the coupling constant for the strong force is on the order of unity. This leads to color confinement, so that quarks an appreciable distance apart do not interact with each other. To achieve this, all observable hadronic states are color singlets. The most common hadronic states are mesons, made of a quark/anti-quark pair with the color singlet state

$$\psi(q\bar{q}) = \frac{1}{\sqrt{3}}(r\bar{r} + g\bar{g} + b\bar{b}), \quad (2.26)$$

and baryons, made of three quarks with the following color singlet state.

$$\psi(qqq) = \frac{1}{\sqrt{6}}(rgb - rbg + gbrgrb + brg - bgr) \quad (2.27)$$

Baryons can also be composed of three anti-quarks, which has a state corresponding to Equation (2.27), but with anti-color.

For this measurement, protons are collided at the LHC. The proton consists of

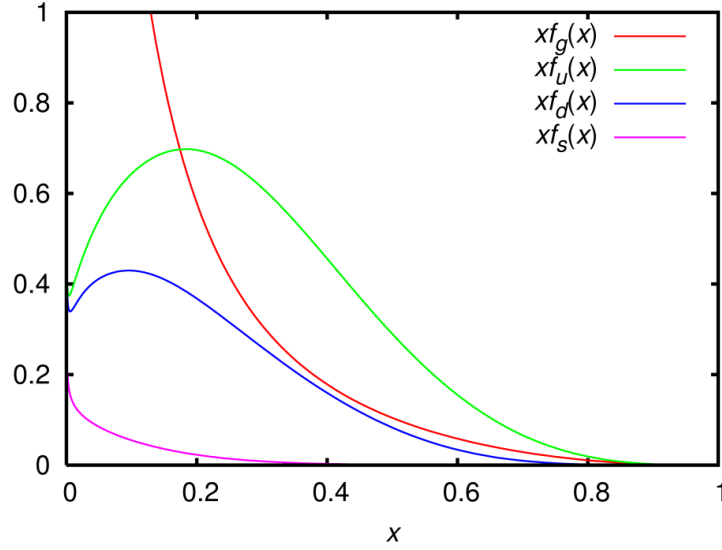


Figure 2-4: The P.D.F.

two u quarks, and one d quark. Since the three quarks inside the proton interact strongly, there are also many virtual gluons and quark/anti-quark pairs present at all times. The quantity and energies of all these partons are not able to be calculated since QCD is non-perturbative. They can be measured in deep inelastic scattering experiments though. In these, electrons are scattered off of protons, and parton distribution functions (PDFs) can be measured. The PDFs for protons are shown in Figure 2-4.

From these things, we can predict the cross section of generating W and Z bosons at the LHC. These need to be very massive though, since they're going to radiate a Higgs. This complicates the calculation a bit.

2.2.2 Decay Channels of Vector Bosons

Due to the couplings described in Section 2.2.1, the vector bosons decay into quarks. However, in the hadronic environment produced at the LHC these are not the best indicators of a vector boson intermediate state. This measurement uses leptonic decays in the final state since the contributions of the vector boson intermediate states to leptonic final states of appropriate kinematics are larger compared to other contributions to this final state.

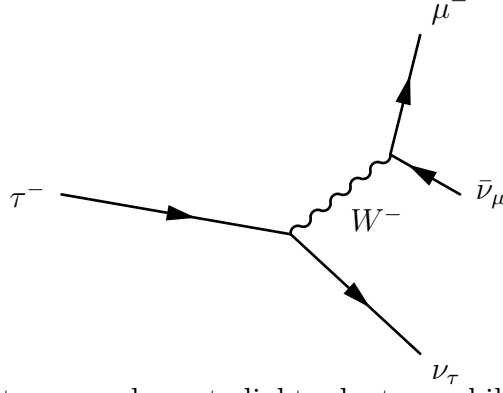


Figure 2-5: Heavier leptons can decay to lighter leptons while emitting two neutrinos. Above is an example of a decay of $\tau \rightarrow \nu_\tau \mu \bar{\nu}_\mu$. The neutrinos cannot be measured at CMS, so it is better to avoid such decays in the analysis.

There are three generations of leptons. Each generation consists of a charged lepton, and a neutral neutrino. The left-handed charged lepton and neutrino of each generation form an electroweak $SU(2)$ doublet. In order of increasing mass, the three generations are called electron, muon, and tau. Heavier charged leptons decay into lighter leptons via the weak force. Two neutrinos result from this decay, as shown in Figure 2-5, making the characteristics of the parent lepton's parent difficult to reconstruct. The tau lepton has enough mass to consistently decay before reaching the CMS detector. The tau lepton is also massive enough to also decay into quarks, making its measurement even more complicated. Muons have an average lifetime long enough to penetrate the entire detector, and electrons are stable particles. As a result, only muons and electrons are considered in this analysis. The Feynman diagrams for the decay channels of interest are shown in Figure 2-6.

2.3 Decay Channels of the Higgs

What we are ultimately interested in measuring is the contribution of the Higgs intermediate state to the final state of $b\bar{b}$. Since the Higgs is a $SU(2)_L$ doublet of scalar fields, the term $-g_f(\bar{L}\phi R + \bar{R}\phi^\dagger L)$ in the Standard Model Lagrangian is invariant under $SU(2)_L \times SU(1)_Y$, where L is a left-handed fermion doublet, and R is a right-handed singlet. If the Higgs doublet is expanded around the vacuum

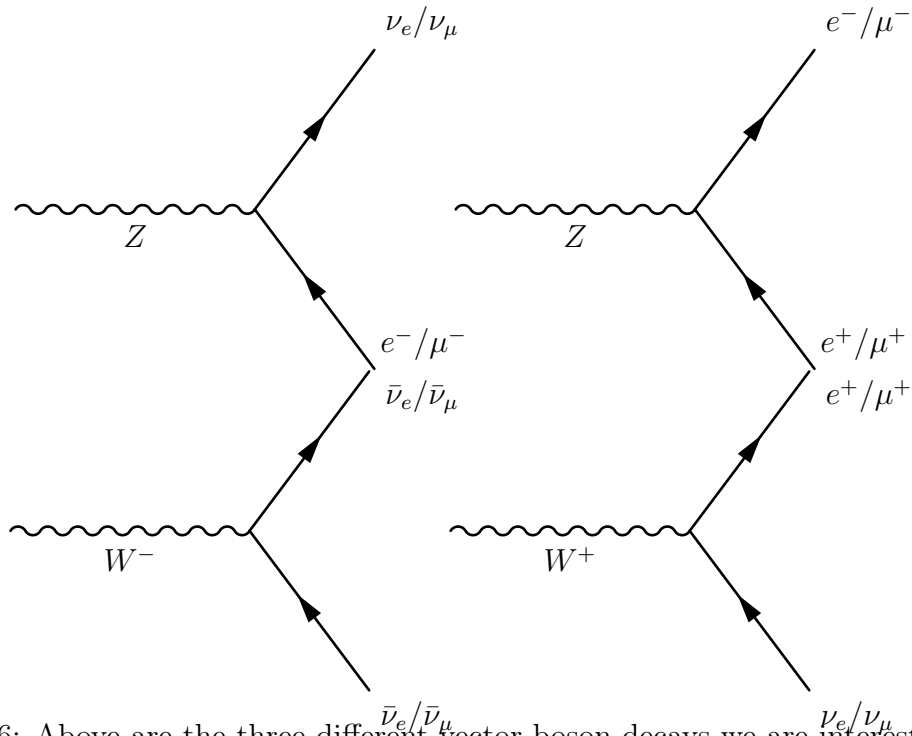


Figure 2-6: Above are the three different vector boson decays we are interested in.

expectation value, as Equation (2.5), the Lagrangian term becomes the following.

$$\mathcal{L}_f = -\frac{g_f}{\sqrt{2}}v(\bar{f}_L f_R + \bar{f}_R f_L) - \frac{g_f}{\sqrt{2}}h(\bar{f}_L f_R + \bar{f}_R f_L) \quad (2.28)$$

In Equation (2.28), f refers to the lower field of the fermion's $SU(2)_L$ doublet. The Lagrangian also includes terms for the upper field since the conjugate of ϕ has the same symmetries as ϕ .

The Lagrangian showing fermion-Higgs interactions in Equation (2.28) consists of two terms. Since v is constant, the first term would be consistent with a fermion's mass, assuming an appropriate coupling constant.

$$g_f = \sqrt{2}\frac{m_f}{v} \quad (2.29)$$

The second term is the coupling of the fermion to the Higgs field with the same coupling constant. This is the mechanism by which the Higgs give fermions their masses, and also why the Higgs couples more strongly to massive particles. The Feynman rule for the interaction vertex between the Higgs and fermions is proportional to the

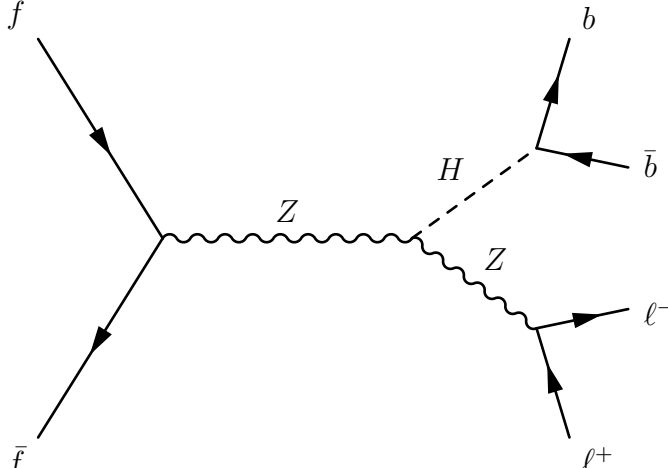


Figure 2-7: Above is the full Feynman diagram for $ZH \rightarrow \ell^+ \ell^- b \bar{b}$.

fermion's mass. Of the quarks, the b quark is the second most massive. The most massive t quark is too massive to be the final decay product of an on-shell Higgs. In fact, the predicted branching ratio of $H \rightarrow b\bar{b}$ is 57.8%. Therefore, measuring $H \rightarrow b\bar{b}$ is the most direct measurement to confirm this theory of quark masses. The diagram for this decay can be combined with the Feynman diagrams in Figure 2-1, Figure 2-2 or 2-3, and one of the decays in Figure 2-6 in order to generate the full Feynman diagrams for the processes being measured in this analysis. One such full diagram is shown in Figure 2-7.

2.4 Other Relevant Standard Model Processes

There are other processes that must be acknowledged in order to explain how the CMS detector works. I'll fill those in as needed?

Chapter 3

The CMS Detector

The Compact Muon Solenoid (CMS) detector, located at the LHC, consists of multiple sub-detectors. Ideally, time could be saved for the dedicated reader by only detailing the parts of the detector that are relevant for the analysis presented in this work. Unfortunately, a characteristic of hadron collider measurements is that almost all sub-systems of the detector are used to make the final measurement.

A brief overview of all the detector subsystems are therefore presented in this chapter. More can be learned about the design and motivations for the detector in the TDR [6]. Information presented on the physical CMS design parameters are taken directly from that document unless otherwise noted.

3.1 Associated Production at the LHC

The CMS detector only observes events. Before describing the devices that are used to observe and record events, the method of generating interesting events must be described. The CMS detector is located at the Large Hadron Collider (LHC). Described in detail in multiple publications [7], a brief description is given here.

The LHC, with a circumference of 26.7 km, is large enough to be considered located in multiple towns and countries, but it will suffice to say it is near Geneva, Switzerland at the European Organization for Nuclear Research (CERN), the main campus of which is addressed in Meyrin, Switzerland. This campus itself also spans

the border between Switzerland and France. This large circumference is needed since charged particles traveling in a circular path with radius r emit synchrotron radiation at the following rate.

$$P = \frac{q^2 p^4}{6\pi\epsilon_0 m^4 c^5 r^2} \quad (3.1)$$

The amount of power lost by the particles decreases quadratically with the size of the collider. In addition, the energy lost decreases with the mass of the accelerated particles to the fourth power. The LHC was built in the same tunnels that were used for LEP, which was a collider for electrons and positrons that took much of its data at $\sqrt{s} = 98 \text{ GeV}$. The resulting LHC is designed to collide protons at energies of $\sqrt{s} = 14 \text{ TeV}$, with the data for this analysis taken at 13 TeV . Though Figure 2-4 showed that not all of the energy from each proton goes into the interaction, there still is adequate phase space available to generate the massive off-shell vector bosons that are needed for *Higgstrahlung*, via the mechanisms described in Section 2.2.1.

The removal of a quark from one proton and an anti-quark from the other leads to two non-color-singlet states in close proximity to each other. The resulting spray of hadronic particles generated from the vacuum to restore color singlets are called jets. The detectors are designed to distinguish these jets from more interesting decay products in the interaction.

The luminosity of the LHC is given by the following formula.

$$\mathcal{L} = \frac{N_B^2 f_{\text{rev}} k_B}{4\pi B^* \epsilon_{xy}} \times F \quad (3.2)$$

N_B is the number of protons per bunch, f_{rev} is the frequency of beam revolutions, k_B is the number of bunches per beam, B^* and ϵ_{xy} describe the goodness of the beam, and F is a geometric collision factor.

$$F = \frac{1}{\sqrt{1 + \frac{(\sigma_s \tan \phi)^2}{\epsilon_{xy} \beta^*}}} \quad (3.3)$$

σ_s is the length of each bunch, and ϕ is the crossing angle. In order to generate a

large amount of data needed for measurements like associated production, the LHC operates at a high frequency of collisions, and generates many simultaneous collisions. For Run 2, there is a proton bunch crossing every 25 ns. The CMS detector must be able to read out and process data on that timescale. Each proton bunch includes over 100 billion protons [8].

3.2 Detector Requirements

One configuration of possible final state particles is shown in Figure 2-7. There, two oppositely charged leptons and two b quarks are the end decay products. The b quarks also hadronize form color singlets well before reaching the detector, but the resulting jets can actually be distinguished well from the jets resulting from the fragmenting protons.

Hadrons containing b quarks decay through the weak force since they require a flavor change. As mentioned before, the CKM matrix in Equation 2.15 quantifies the mixing between the different quark flavors. The value of V_{tb} is close to unity, and since the CKM matrix is unitary, V_{cb} and V_{ub} are small. This means the matrix element weak decays of the b hadrons is small. This is the only decay channel available to the lightest b hadrons, so their lifetimes are relatively long. The delayed decay results in a jet with a secondary vertex where many of its particles are generated from the vacuum at a distance from the initial collision point.

Alternate signatures of interest can be seen by substituting other vector boson final states from Figure 2-6. In these, there may be one or zero charged leptons, with one or two neutral leptons, respectively. Neutral particles are difficult to detect, with neutral leptons being capable of passing through the entire Earth without being part of a detectable interaction. The CMS detector therefore ignores the neutrinos, but their presence can still be inferred. Even with the variation in momentum along the beam direction, all partons in each proton have approximately zero momentum in the transverse direction. Therefore, the sum of the transverse momenta of all final state particles must also be zero. Many events in CMS have an overall imbalance in

the transverse plane. This imbalance is labelled Missing Transverse Energy, E_T^{miss} , or MET. Large MET in an event is often a sign of high energy neutrinos that the detector cannot detect.

We need to identify all of these interesting particles, as well as be able to reconstruct missing transverse momentum. In addition, the additional hadronic activity in the event, called pileup, must be mitigated. The energy of the decay products have energies on the scale of the masses of the parent particles. The detector must be capable of measuring jets and leptons with energies on the order of 10s or 100s of GeV. Better energy resolution for each of these decay products allows better separation of our signal process from background processes that generate very similar final states.

3.3 Detector Design

The CMS detector as a whole has cylindrical symmetry around the beam access. It is 21 meters long and 15 meters in diameter. There are gaps at either end to allow the beam, but otherwise tries to cover the full solid angle around the collision point. The azimuthal angle of a particle relative to the beam axis is described by pseudorapidity, η .

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (3.4)$$

The barrel portion of CMS can detect particles with $|\eta| < 2.4$, while the forward caps of the detector can reach $|\eta| < 5.0$.

Different technologies are better for measuring the energy or other kinematics variables of different particles. As a result, the CMS detector is made up of different sub-detector systems, arranged in cylindrical layers. Each layer consists of a “barrel” portion and two endcaps on either side.

The innermost layer is designed to extrapolate the tracks of charged particles back to their point of origin. This is called the Silicon Pixel Detector. The next layer is designed to measure the energies of photons and electrons. The third layer

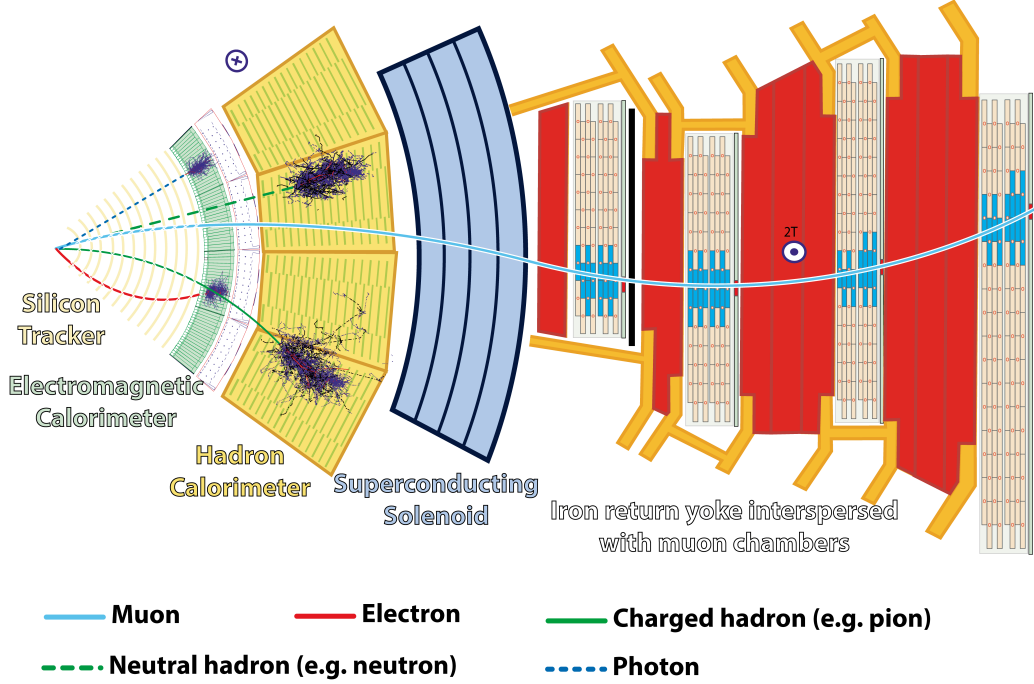


Figure 3-1: A slice of the CMS detector is shown above [9]. The four detector layers are labelled and show the penetration depths of various particles stable enough to travel a measureable distance.

measures the energies of both charged and neutral hadrons. Outside of these three layers is a superconducting solenoid, which generates a magnetic field for the entire detector. On the very outside of the detector are gas chambers designed to detect muons interspersed with the iron return yoke for the solenoid. A slice of the CMS detector showing the relative positions of each layer is shown in Figure 3-1.

The magnet is described first since the magnetic field it produces is a key part of most of the rest of the detector. After that, the sub-detectors are summarized in the order of closest to farthest from the beamline, since this is the order that particles would interact with the layers. Each sub-detector section also describes the measured performance during Run 2. Note that this measurement is an iterative process that depends on the event reconstruction described in Section 3.4, which in turn depends on the performance of the entire detector. The performance numbers are presented with each sub-detector design though so that it is immediately clear how effective each design has been.

3.3.1 Solenoid Magnet

Part of the CMS acronym acknowledges the role of the solenoid magnet. The presence of a magnetic field is necessary for accurate measurements of charged particles passing through the silicon pixel detector and muon chambers.

The magnetic field generated is designed to cause the path of a muon with 1 TeV of energy to bend enough to have a momentum resolution of 10%. Inside the solenoid, the magnetic field reaches 4 T, with a return field that is large enough to cause muon tracks to curve throughout the muon chambers outside the magnet.

A super-conducting solenoid enables the creation of a magnetic field with the required strength. A current of 19.5 kA is sent through 2168 turns over 12.9 m. The magnetic field stores 2.7 GJ of energy. In order to hold this, the structural components holding the magnet and the detector in place are strong enough to withstand 64 atm of hoop pressure.

3.3.2 Silicon Pixel Detector

The layer closest to the beamline is designed to obtain a precise track pointing to the origin of particles passing into the detector. It is made up of layers of many small pixels to do this. As distance from the interaction point increases, the pixel size also increases.

The innermost three layers, with the closest layer being a distance of $r = 4$ cm from the interaction point, are made of hybrid pixel detectors. Each pixel has dimensions of $100 \times 150 \mu\text{m}$. At this size, only one out of every ten thousand inner layer pixels is triggered in a typical LHC bunch crossing. Outside of the pixel detector layers, silicon strip detectors are used. These are placed in the region that is $20 < r < 55$ cm from the beamline. Strip dimensions give a cell size of approximately $10 \text{ cm} \times 80 \mu\text{m}$. 2–3% of cells are activated during a typical bunch crossing. The outermost layers are made of larger strips with cell sizes of $25 \text{ cm} \times 180 \mu\text{m}$. About 1% of these pixels are triggered each bunch crossing.

The active material of the silicon pixel detector is semi-conducting silicon. When

charged particles pass through, electron-hole pairs are generated and drift apart due to a bias voltage. The voltage change when these pairs reach their respective electrodes indicates a charged particle passed through. Because of this, the silicon pixel detector cannot detect any neutral particles, but it gives a point of origin for charged particles that is accurate enough to identify pileup.

In the beginning of 2017, the pixel detector was upgraded to handle the higher radiation environment of Run 2 [10]. A layer was added to both the barrel and endcap sections of the pixel detector. Firmware was also upgraded to keep the pixel detector operating at a frequency higher than the Run 2 collision frequency. With this upgrade, the detector operated with a 97% hit efficiency for all layers at the highest instantaneous luminosity. Layers beyond the first performed with greater than 99% hit efficiency [11].

3.3.3 Electromagnetic Calorimeter

The next layer of the detector is called the Electromagnetic Calorimeter or ECAL. This layer is designed to fully capture and measure the energy of photons and electrons. The ECAL is made of crystals of the scintillating material Lead Tungstate (PbWO_4). Each crystal is placed in the detector so that its smallest face is facing the collision point. These small faces have dimensions of 22×22 mm. The length of each crystal is 230 mm, and the far face is slightly larger at 26×26 mm. PbWO_4 has a radiation length of $\chi_0 = 8.9$ mm and a Moliere radius of 21 mm. This means each crystal is 25.8 radiation lengths, containing the full shower within the ECAL, and each shower is also localized to within one crystal from the initial ionization.

The scintillating properties of PbWO_4 are also desirable for observing LHC collisions. 4.5 photons for every MeV of deposited energy are ultimately detected by the photodiodes at the far end of the crystals. This is a low number for most experiments, but the only photons and electrons of interest in this measurement deposit at least 10s of GeV of energy. This gives the ECAL energy resolutions in the range of 5 – 10%. More importantly, the scintillation is very fast. 80% of the light from an interaction is emitted within the 25 ns between bunch crossings, making it easy to associate the

readouts with the appropriate collision.

During the operation of the CMS, the ECAL crystals are damaged by radiation. Some recovery occurs as the crystal structure falls back into the ground state, but over time, the performance of the crystals degrade. That degradation happens at different rates in different areas of the detector. Lasers are used to calibrate the ECAL online during the gaps between beams [12]. Resolution is measured by looking at $Z \rightarrow ee$ events. For Run 2, the barrel region of the ECAL performed with 1.6% resolution, and the other regions had a 5% resolution [13].

3.3.4 Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) has the same goal as the ECAL, where it contains particles and measures the energy emitted by them. However, it tries to do this for hadrons, such as protons, neutrons, and stable mesons. Since hadrons are much more massive than electrons, the ionizing collisions in a typical scintillator does not slow them down enough to contain them. Instead, they must interact via nuclear collisions to be attenuated. CMS uses brass for its HCAL due to its relatively short interaction length, the fact that it is non-magnetic, and its affordability.

The barrel of the HCAL is jacketed in stainless steel for structural support. This layer is 61 mm on the layer immediately next to the ECAL and 75 mm thick on the outer edge. The inside of the HCAL consists of brass absorber plates interspersed with plastic scintillator tiles. The layers closer to the beamline alternate 50.5 mm brass plates with 3.7 mm scintillator plates. Farther away, the brass plates are instead 56.5 mm thick. Wavelength shifting fibers are run through the scintillator tiles to allow photons to travel to the outside of the HCAL where they are detected by photodiodes.

The HCAL performance also degrades as it is exposed to radiation, like the ECAL in Section 3.3.3. The calibration for HCAL is performed using an embedded radioactive source, lasers and LEDs, and an *in situ* calibration using assumed symmetry in ϕ . With these methods, a response within 3.4% was maintained in the HCAL barrel and within 2.6% in the HCAL endcap [14].

3.3.5 Muon Chambers

Muons are the most penetrative particles that CMS attempts to detect. They are heavier than electrons, so they are not stopped in the ECAL. They do not interact via the strong nuclear force, so the high density of the HCAL also does not cause measurable interactions. Instead of stopping and measuring muons in calorimeters, CMS tracks their trajectory with both the pixel detectors on the inside of the detector and the muon chambers that make up the outer layer of the detector.

This is the only sub-detector system outside of the solenoid, but the returning magnetic field is still present, allowing the momentum of the muons to be extracted from the curvature of their trajectory. Layers of muon chambers act much like the pixel detectors, but at a larger and more distant scale. The muon chambers in the barrel region of $|\eta| < 1.2$ consist of drift tube chambers. In the endcaps, cathode strip chambers are used. The difference is to account for higher neutron backgrounds in the endcap, as well as a greater magnetic field. In both regions, resistive plate chambers are spaced between the layers of the other muon chambers.

The muon detectors were triggered above 90% of the time a muon passed through. The overall efficiency of the muon trigger, which relies heavily on the muon system and is described in more detail in Section 3.5, increases as a function of muon p_T and plateaus around 90% efficiency. The timing of the muon system assigns 1% of muons to the wrong event [15].

3.4 Event Reconstruction

Event reconstruction is done using a combination of tracks and calorimeter deposits. These are then linked across the different detector components to identify particles. This overview is taken largely from reference [16]. An interested reader can find more details there.

3.4.1 Charged Particle Tracks

Both the Silicon Pixel Detector and the Muon Chambers are designed for charged particles to leave tracks. In both sub-detectors, the basic steps for track reconstruction are the same. First, a track must be seeded. Usually, this is done by finding hits in consecutive layers that are consistent with a particle coming from the beamline. A Kalman Filter is then used to find hits in the other layers of the appropriate sub-detector that are consistent with the initial seed. Once more hits have been found, a fit is performed for the precise trajectory of the track.

Tracks are kept or discarded based on the number of layers that are missing hits and on the momentum of a charged track in the magnetic field of the solenoid. Making these parameters looser results in better recovery of tracks, but the high activity within the detector results in a combinatorial background. This background increases exponentially when the momentum cut is reduced, for example. To help reduce this background, tracks with missing hits use an iterative fit. Different seeds are found for each track to make sure that the resulting collections of hits remain the same.

Additional complications arise for each the pixel and muon detectors. The Silicon Pixel Detector is the only tracking detector that deals with electrons. Because of their small mass, electrons are likely to radiate energy while travelling through the magnetic field of CMS. This leads to complications in the calorimeters described in Section 3.4.2. It also means that the radius of curvature of an electron track can decrease appreciably within the pixel detector. This can lead to the Kalman Filter approach missing tracks entirely, depending on the number of hits required. Tracks with a large χ^2 and a certain number of hits are fit again using a Gaussian-Sum Filter (GSF). The GSF allows for fitting tracks that have significant energy loss, recovering electron and positron tracks.

A change of trajectory may also happen in the muon chambers, but this is due to multiple scattering in the return yoke. No specialized tracking algorithm is used to account for it. The muon tracking performs best when the tracks in the muon

chambers are successfully linked to a track in the pixel detector. The most common backgrounds in the muon chambers is caused when hadrons manage to punch through the HCAL. This can often be mitigated by considering the amount of energy deposited along the particle track in the other sub-detector systems.

3.4.2 Calorimeters

Clusters are identified in calorimeters, also using a seeding algorithm. First clusters with a large energy deposit are identified, and then nearby crystals are checked against noise thresholds. The energy deposition is assumed to have a Gaussian profile, and so a fit is performed to show that there are not multiple deposits in the same region of the calorimeter.

As mentioned in Section 3.3.3, the dimensions of the ECAL crystals are comparable to the Moliere radius, keeping clusters localized. Although a complication arises due to the bremsstrahlung from electrons that also leads to the need for using GSF to find tracks. Superclusters that are linked to electrons have a larger allowed range over ϕ to account for this. Photons are still well behaved within the ECAL, and the ECAL is therefore calibrated for photon deposits.

The calibration of HCAL is complicated by the fact that particles reaching it have to first pass through the ECAL. Initial calibration was done with a 50 GeV pion test beam, but the actual response is non-linear in energy as well as different for charged and neutral particles. Reasons for this difficulty include particles losing energy in the region between the ECAL and the HCAL, in addition to the energy lost in the ECAL. Therefore, there are calibration coefficients that are used depending on if the energy deposits are all in the HCAL, or in preceding sub-detectors as well.

3.4.3 Linking and Particle Identification

An important step in making sense of the various sub-detector readouts is linking tracks to calorimeter clusters. The general procedure is to extrapolate tracks from the inner tracker out to each calorimeter. A shower that originates within one radiation

or interaction length in the calorimeter along that track is linked with the track.

The bremsstrahlung from GSF electrons is linked to the track by looking along track tangents. A dedicated conversion finder is used to identify pair production within the pixel detector caused by either bremsstrahlung or prompt photons in order to not mistakenly link a charged particle track with what should otherwise be measured as a photon. This step, in addition to some ECAL clusters that do not have a track make it possible to identify isolated photons. On the other hand, it is still difficult to determine whether an electron track is well isolated or not. The large number of variables that go into identifying an electron leads to the training of a Boosted Decision Tree (BDT) to identify an electron. Separate BDTs are needed for the barrel and endcap regions of the detector.

For accurate HCAL readings, the linking algorithm also ECAL clusters and HCAL clusters along a path. These may not always be along a charged particle track. Multiple calorimeter links of this nature may be found, but only a single link is kept based on a distance assigned to each link. HCAL hits without a track are identified as neutral hadrons. HCAL hits with a linked track are likely charged hadrons. Though the ECAL clusters must be linked in order to determine the energy coefficient to calibrate the HCAL, ECAL hits without tracks are still identified as photons since photons carry some of the energy of jets.

Of particular interest to this analysis is also the secondary vertex step of linking. Charged particle tracks that do not go back to the interaction vertex are linked together if they share a common secondary vertex. These tracks must have an invariant mass of greater than 0.2 GeV to be kept. There must also be a track from the secondary vertex to the primary vertex, which would belong to a long-lived hadron. As mentioned in Section 3.2, this is the signature of a b jet. It is possible, however, that the secondary vertex is generated by a nuclear scattering within the pixel detector, so additional analysis is needed for each secondary vertex.

The final link is made between tracks in the muon chambers and tracks in the inner tracker. Muons are identified as tracker muons if they only leave tracks in the inner tracker. This can often happen with low energy muons. They are called standalone

muons when only the track in the muon chamber is identified. When tracks are successfully linked in both sub-detectors, the resulting reconstructed particle is called a global muon. When a global muon is not well-isolated from other energy deposits, it must have tighter requirements on how it behaves in the muon chambers. This is to prevent energy from a jet from being attributed to a muon or vice versa. This is important for b jets since the decay that happens at a displaced vertex is a decay through the weak nuclear force.

3.5 Trigger

Collisions happen every 25 ns, corresponding to a frequency of 400 MHz. The amount of data that the detector generates at each collision is too large to store all of it at this frequency. Luckily, most collisions result in uninteresting data. A trigger system is used to identify interesting events and reduce the frequency of event writing to 1 kHz. This is done using two stages. The Level-1 (L1) trigger passes events with a frequency of 100 kHz, and the High Level Trigger (HLT) picks from the remaining events with a frequency of 1 kHz [17].

The L1 trigger is implemented in hardware. It was upgraded for Run 2 of the LHC to run on FPGAs on an Advanced Mezzanine Card. There are two main components of the L1 trigger. One considers calorimeter deposits, and the other examines the muon chambers. The overall L1 trigger fires when there are high energy, resolved calorimeter hits or if a possible muon is reconstructed. Due to the flexibility of FPGAs, the exact conditions of the firing are configurable [18].

After the L1 is fired, the data is sent to the High-Level Trigger (HLT), which is a computing farm that makes a final decision on whether or not to save the data using a rough event reconstruction. The use of 30,000 cores in the HLT allows for buffering data so the HLT has plenty of time to make this decision [19].

For this analysis, only a few of the possible HLT paths are of interest. The exact trigger names are given in Chapter 5, but for the most part, they only depend on three different identifiable objects. Figures 2-6 and 2-7 show the different final states

of interest. b jets are difficult to identify quickly since we must rely on the pixel detector's reconstruction of the secondary vertex, but the decay mode of the vector boson can be used for the trigger. In that case, only events with one of the following are worth saving and examining for $VH \rightarrow b\bar{b}$.

- ECAL deposits consistent with a high p_T electron
- muon chamber hits consistent with a high p_T muon
- an overall energy imbalance consistent with MET from neutrinos

For the specific decay channel in Figure 2-7, the HLT also includes paths where two electrons or two muons are identified.

3.6 Accessing Data

The final important piece of the CMS detector is its offline computing resources.

Tier-1 sites typically provide 30,000 CPU cores, while Tier-2 sites provide another 60,000 [20].

Chapter 4

Simulation

4.1 Backgrounds to the Analysis

In order to effectively measure Higgs production, we need to be able to accurately estimate other events that end up in our selection. To get started, a loose understanding of

4.2 Event Generation

We use different generators.

Details in Appendix D.

4.2.1 Tree Level Simulation

4.2.2 Parton Showers

4.3 Detector Simulation

4.4 Corrections to Simulation

4.4.1 Smearing

Muons

Electrons

Jets

Chapter 5

Event Selection

Since Chapter 2, a picture of the overall event topology of interest has been depicted. Until now, the description of the events have been vague. This chapter gives the specific cuts used to identify each particle and to reject or otherwise classify events. First, objects will be defined in terms of reconstructed variables. Then common cuts used to reject events are given. After that, cuts used to classify events into different decay channels are specified. This chapter ends with the differentiation between events where the Higgs decay products can be resolved as separate jets and where they are merged into a single massive jet.

5.1 Object Definitions

Section 3.4 describes how detector responses are linked to possible physical particles. Most of the particle ID techniques described so far also give false positives. What follows are tighter selections used in order to reduce these backgrounds. Once objects are defined, they can be counted for event classification.

Each type of particle generally has a method of pre-selection and additional cuts for a selection. For these objects, pre-selection often means that the object is defined well enough to veto events in categories that don't want the object to exist. The additional cuts are required for an event to classify as a category that wants the object to exist with reduced false positives.

5.1.1 Variable Definitions

Many of the object definitions use variables that have not been defined yet. They can be understood in terms of the reconstruction described in Section 3.4.

Lepton isolation is quantified using the following formula.

$$I = \frac{1}{p_T^\ell} \left(\sum p_T^{\text{charged}} + \max \left[0, \sum p_T^{\text{neutral}} + \sum p_T^\gamma - p_T^{\text{PU}} \right] \right) \quad (5.1)$$

The sums are over charged hadrons originating from the primary vertex and all neutral hadrons and photons within a distance of $\Delta R < 0.4$ from the lepton if it is a muon, or $\Delta R < 0.3$ from an electron. ΔR is a distance used in the (η, ϕ) plane.

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (5.2)$$

The term p_T^{PU} is defined as the following for muons.

$$p_T^{\text{PU}} = 0.5 \times \sum_i p_T^{\text{PU},i} \quad (5.3)$$

i refers to charged hadrons that do not originate from the primary vertex. Electrons use the following definition.

$$p_T^{\text{PU}} = \rho \times A_{\text{eff}} \quad (5.4)$$

A_{eff} is the area of the isolation cone, and ρ is the median of the p_T density of neutral particles in that area.

Particles can also be defined as coming from the primary vertex of an event or from pileup. Vertices are defined through deterministic annealing [21], using the closest approach of tracks to the beamline [22]. The primary vertex is the vertex with the greatest sum of E_T of the charged particles originating from it. After identification of the primary vertex, charged particles are classified as originating from the primary vertex or as pileup using their extrapolated track's distance in the transverse plane, d_{xy} and distance along the beamline, d_z .

5.1.2 Muons

An isolated muon gives one of the cleanest signatures in CMS, with only perhaps the exception of an isolated photon that does not undergo pair production in the pixel tracker. Muons can also show up in jets from weakly decaying hadrons, in which case they are not isolated. Since weakly decaying b jets are central to this analysis, events with non-isolated leptons are not rejected, but the distinction is important.

Pre-selected muons must meet the following requirements.

- The muon must have a relatively high energy of $p_T > 5 \text{ GeV}$.
- The muon should be located in the barrel within $|\eta| < 2.4$.
- The muon originates from the primary vertex, satisfying both $d_{xy} < 0.5 \text{ cm}$ and $d_z < 1.0 \text{ cm}$.
- The muon must pass a loose isolation cut of $I < 0.4$.
- The muon must be a PF muon.
- The muon is either a global muon or a tracker muon.

Fully selected muons have some additional cuts they must pass.

- They must have a higher transverse momentum at $p_T > 25 \text{ GeV}$. In events with two muons, such as caused by $Z \rightarrow \mu\mu$, the second muon can pass the slightly looser cut of $p_T > 15 \text{ GeV}$.
- The muon must be a global muon, leaving tracks in both the central tracker and the muon chambers.
- There must be more than five hits in the inner tracker with one hit on a pixel.
- The fit for the global muon track must be good with $\chi^2/ndof < 10$.
- The muon must be well isolated with $I < 0.06$

These definitions are accepted by all members of the CMS collaboration as loose and tight working points, respectively. This allows the analysis to use efficiency measurements created for wider use.

5.1.3 Electrons

The kinematic variables associated with an electron are extracted from the GSF fit. Pre-selected electrons must meet the following requirements.

- They must have a high transverse momentum with $p_T > 7 \text{ GeV}$
- They should be centered in the detector with $|\eta| < 2.4$.
- They originate from the primary vertex with $d_{xy} < 0.05 \text{ cm}$ and $d_Z < 0.2 \text{ cm}$.
- They pass a loose isolation cut of $I < 0.4$.

To reduce backgrounds, electrons are identified with the aid of an MVA [23]. Fully selected electrons pass the tight working point used by the CMS collaboration. In order to also match the samples of simulated electrons used in the training sample, the selected electrons also must pass the following cuts.

- The electron must have higher energy with $p_T > 15 \text{ GeV}$.
- The deposit of HCAL energy must be less than 9% of the ECAL energy deposit along the electron track.
- The track sum p_T component of the isolation must be less than 18% of the electron p_T .
- The electron must either have $|\eta| < 1.4442$ or $|\eta| > 1.5660$
- For electrons with $|\eta| < 1.4442$:
 - $\sigma_{i\eta i\eta} < 0.012$
 - Isolation in the ECAL cluster must be less than 0.4, and isolation in the HCAL must be less than 0.25.

- The difference between super cluster and track location of the electron must be small with $\Delta\eta < 0.0095$ and $\Delta\phi < 0.065$.
- For electrons with $|\eta| > 1.5660$:
 - $\sigma_{i\eta i\eta} < 0.033$
 - Isolation in the ECAL cluster must be less than 0.45, and isolation in the HCAL must be less than 0.28.

There is a gap in the η direction that accounts for a gap in the CMS detector. This gap in the active detecting volume is needed to accomodate various electronics and structural components.

5.1.4 Jets

As described in Section 3.1, hadrons produced at the LHC are often accompanied by sprays of particles called jets. Conservation of energy and momentum means that the sum of jet constituents give the kinematics of the initial parton that produced them. Jets are constructed by clustering all particle-flow candidates with the anti- k_T algorithm [24] using $R = 0.4$. Due to factors like pileup and detector response, the energy of the reconstructed jets are corrected [25].

Loose jet cuts, based on the constituents, are applied to remove jets constructed from detector noise. Jets that get a significant fraction of their energy from pileup are also removed. Preselected electrons and muons are also often reconstructed as jets. Any jet within $\Delta R < 0.4$ from a preselected lepton is removed.

To be considered for the study of decay products of the Higgs boson, jets must be within the barrel of the detector with $|\eta| < 2.5$. The jets must satisfy $p_T > 25$ GeV for the zero and one lepton signatures. The two lepton signature from $Z(\ell\ell)H$ is cleaner and can apply the looser criteria of $p_T > 20$ GeV to the jets.

Table 5.1: The cuts for each DeepCSV working point are defined for each year of Run 2 of the LHC. The working points are defined by their mistag rates.

Working Point	Mistag Rate	2016	2017	2018
Loose	10%	0.2219	0.1522	0.1241
Medium	1%	0.6324	0.4941	0.4184
Tight	0.1%	0.8958	0.8001	0.7527

Identification of b Jets

As mentioned in Section 3.2 while motivating the design of CMS, jets containing b hadrons have a distinct signature. This includes secondary vertices displaced from the beamline, as well as non-isolated leptons from weak decays. All of these features are considered in a deep neural network (DNN) called Deep Combined Secondary Vertex (DeepCSV) [26]. The output of DeepCSV has three working points that are defined based on the amount of false positives that can be expected in a collection of jets passing the cut. The specific values are different for each year of operation. They are given in Table 5.1.

5.1.5 MET

MET is corrected.

5.2 Removal of QCD

We have some cuts across the board on our objects in order to remove events that are just QCD.

5.3 Categories of Vector Boson Decay

Now that we are ready to count, we can count leptons in order to characterise potential vector bosons.

5.3.1 0 Leptons

5.3.2 1 Lepton

5.3.3 2 Leptons

5.4 Topology of Higgs Decay

5.4.1 Resolved Jets

We reconstruct two b jets.

5.4.2 Boosted Jet

When the Higgs has very high p_T , the jet clustering algorithms can find both daughter particles as being part of a single jet.

5.5 Control Regions

5.5.1 Light Flavor Jets

5.5.2 Heavy Flavor Jets

5.5.3 $t\bar{t}$

Chapter 6

Analysis Results

6.1 Systematic Uncertainties

6.2 Combination Fit Method

6.3 Results

Chapter 7

Conclusions

Appendix A

Detector Projects

Each collaborator must contribute to the operation of the CMS detector before his or her name is added to the author list. The operation of the detector is distinct from analyzing the data generated by the detector, so all collaborators must adopt some role outside of being a physicist.

This appendix details projects I completed in order to contribute to the operation of the CMS detector. The first project presented is the Dynamo Consistency project. It is a plugin for the dynamic data management system Dynamo [27] that compares the inventory of files Dynamo expects at a site with the files that are actually at a site. The other project described is known as Workflow Web Tools. This is a dynamic web server that displays errors reported by the CMS computing infrastructure to operators, and allows those operators to perform corrective actions through the web page. Workflow Web Tools also tracks operator actions for future use in training various machine learning models. Both projects produced software packages written in Python [28, 29] and available through the Python Package Index (PyPI) as `dynamo-consistency` and `workflowwebtools`.

A.1 Dynamo Consistency

[30] [31]

A.2 Workflow Web Tools

Appendix B

Physics Calculations

Appendix C

Data Format

Appendix D

Generator Parameters

Appendix E

Data Card

Bibliography

- [1] T. C. Collaboration, “Observation of higgs boson decay to bottom quarks,” *Physical Review Letters*, vol. 121, Sep 2018.
- [2] M. Thomson, *Modern particle physics*. Cambridge: Cambridge University Press, 2013.
- [3] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964.
- [4] P. W. Higgs, “Broken symmetries and the masses of gauge bosons,” *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.
- [5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global conservation laws and massless particles,” *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964.
- [6] T. C. Collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS, Geneva: CERN, 2006.
- [7] L. Evans and P. Bryant, “LHC machine,” *Journal of Instrumentation*, vol. 3, pp. S08001–S08001, aug 2008.
- [8] R. Bruce, G. Arduini, H. Bartosik, R. De Maria, M. Giovannozzi, G. Iadarola, J. Jowett, K. S. B. Li, M. Lamont, A. Lechner, E. Metral, D. Mirarchi, T. Pieloni, S. Redaelli, G. Rumolo, B. Salvant, R. Tomas Garcia, and J. Wenninger, “LHC Run 2: Results and challenges,” p. MOAM5P50. 7 p, Jul 2016.
- [9] D. Barney, “CMS Detector Slice.” CMS Collection., Jan 2016.
- [10] A. Dominguez, D. Abbaneo, K. Arndt, N. Bacchetta, A. Ball, E. Bartz, W. Bertl, G. M. Bilei, G. Bolla, H. W. K. Cheung, M. Chertok, S. Costa, N. Demaria, D. D. Vazquez, K. Ecklund, W. Erdmann, K. Gill, G. Hall, K. Harder, F. Hartmann, R. Horisberger, W. Johns, H. C. Kaestli, K. Klein, D. Kotlinski, S. Kwan, M. Pesaresi, H. Postema, T. Rohe, C. Schäfer, A. Starodumov, S. Streuli, A. Tricomi, P. Tropea, J. Troska, F. Vasey, and W. Zeuner, “CMS Technical Design Report for the Pixel Detector Upgrade,” Tech. Rep. CERN-LHCC-2012-016. CMS-TDR-11, Sep 2012. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.

- [11] A. Modak, “CMS Phase-1 Pixel Detector: Operational Experience, Performance and Lessons Learned,” Tech. Rep. CMS-CR-2019-283, CERN, Geneva, Nov 2019.
- [12] F. Monti, “CMS ECAL monitoring and calibration in LHC Run 2,” Tech. Rep. CMS-CR-2018-171, CERN, Geneva, Aug 2018.
- [13] N. Bartosik, “Performance of the CMS Electromagnetic Calorimeter in LHC Run2,” *PoS*, vol. LeptonPhoton2019, p. 126. 4 p, 2019.
- [14] M. Chadeeva and N. Lychkovskaya, “Calibration of the CMS hadron calorimeter in run 2,” *Journal of Instrumentation*, vol. 13, pp. C03025–C03025, mar 2018.
- [15] N. Pozzobon, “The CMS muon system: performance during the LHC run-2,” *Journal of Instrumentation*, vol. 14, pp. C11031–C11031, nov 2019.
- [16] A. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, M. Flechl, and et al., “Particle-flow reconstruction and global event description with the cms detector,” *Journal of Instrumentation*, vol. 12, p. P10003–P10003, Oct 2017.
- [17] M. Tosi, “The CMS trigger in Run 2,” Tech. Rep. CMS-CR-2017-340, CERN, Geneva, Oct 2017.
- [18] L. Cadamuro, “The CMS level-1 trigger system for LHC run II,” *Journal of Instrumentation*, vol. 12, pp. C03021–C03021, mar 2017.
- [19] H. Sert, “CMS High Level Trigger Performance in Run 2,” Tech. Rep. CMS-CR-2019-278, CERN, Geneva, Nov 2019.
- [20] K. Bloom, “Cms software and computing for lhe run 2,” *Proceedings of 38th International Conference on High Energy Physics — PoS(ICHEP2016)*, Feb 2017.
- [21] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2210–2239, 1998.
- [22] T. C. Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” *Journal of Instrumentation*, vol. 9, pp. P10009–P10009, oct 2014.
- [23] J. R. and, “CMS electron and photon performance at 13 TeV,” *Journal of Physics: Conference Series*, vol. 1162, p. 012008, jan 2019.
- [24] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-ktjet clustering algorithm,” *Journal of High Energy Physics*, vol. 2008, p. 063–063, Apr 2008.
- [25] V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, and et al., “Jet energy scale and resolution in the cms experiment in pp collisions at 8 tev,” *Journal of Instrumentation*, vol. 12, p. P02014–P02014, Feb 2017.

- [26] A. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, and et al., “Identification of heavy-flavour jets with the cms detector in pp collisions at 13 tev,” *Journal of Instrumentation*, vol. 13, p. P05011–P05011, May 2018.
- [27] Y. Iiyama, B. Maier, D. Abercrombie, M. Goncharov, and C. Paus, “Dynamo – handling scientific data across sites and storage media,” 2020.
- [28] G. Van Rossum and F. L. Drake Jr, *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [29] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.
- [30] A. Dorigo, P. Elmer, F. Furano, and A. Hanushevsky, “Xrootd-a highly scalable architecture for data access,” *WSEAS Transactions on Computers*, vol. 1, no. 4.3, pp. 348–353, 2005.
- [31] E. Laure, A. Edlund, F. Pacini, P. Buncic, M. Barroso, A. Di Meglio, F. Prelz, A. Frohner, O. Mulmo, A. Krenek, *et al.*, “Programming the grid with glite,” tech. rep., CERN, Geneva, Switzerland, 2006.