

Sensitivity study of a search for a charged scalar particle in proton-proton collisions at a center of mass energy of 14 TeV

Kays Haddad

Department of Physics
McGill University, Montréal

August 2018



A thesis submitted to McGill University in partial fulfillment of the
requirements of the degree of Master of Science.

©Kays Haddad 2018

Contents

Abstract	i
Résumé	ii
Acknowledgements	iii
1 Introduction	1
2 The Standard Model and Beyond	4
2.1 The Standard Model Higgs Sector	6
2.2 Extensions of the Standard Model Higgs Sector	7
2.3 The Model: A Generic Charged Higgs Boson	9
3 Data Simulation	15
3.1 Simulation Chain	15
3.2 Signal	18
3.3 Background	19
3.3.1 Standard Model Backgrounds	19
3.3.2 Fake Signal Background	20
4 Detector Simulation	22
4.1 Geometry of the Detector	22
4.2 Four-Momentum Smearing	26
4.3 Object Reconstruction	26

4.3.1	Photon Reconstruction	27
4.3.2	Jet Clustering	28
4.4	Pileup Mitigation	29
5	Analysis	33
5.1	Preselection Cuts	34
5.1.1	Detector Acceptance Cuts	34
5.1.2	Background Discrimination Cuts	35
5.2	Kinematic Fitting	37
5.2.1	$m_{jj\gamma}$ Distribution	40
5.3	Limit Setting	46
5.4	Uncertainties	48
5.4.1	PDF and Scale Uncertainties	48
5.4.2	Integration Errors on Cross-section Estimates	51
5.4.3	Statistical Errors on Efficiencies	52
6	Results	54
6.1	Limits on Production Cross-Section	54
6.2	Discovery Potential in the GM Model	56
7	Summary and Outlook	58
	References	60
A	Measurement of DELPHES Detector Resolutions	65
A.1	Matching Truth and Reconstructed Objects	65
A.2	Energy Resolution	66
A.3	η and ϕ Resolutions	67

List of Figures

2.1	The Feynman diagram for the signal process.	11
2.2	Di-Higgs invariant mass with different virtual bosons.	12
3.1	Example background Feynman diagrams.	21
4.1	Schematic of the pseudorapidity variable.	24
4.2	The DELPHES detector.	25
5.1	Kinematic distributions motivating background discrimination cuts.	36
5.2	Separation between objects in the same grouping.	39
5.3	Effect of kinematic fit on signal.	41
5.4	Invariant mass distributions for test masses below 130 GeV.	42
5.5	Effect of kinematic fit on the QCD background.	43
5.6	Example of the CBG fit.	44
5.7	Reconstructed $m_{jj\gamma}$ resolutions.	45
5.8	Invariant mass distribution of signal plus background.	46
6.1	Upper limits on $\sigma(pp \rightarrow H^+H^-) \times BR(H^\pm \rightarrow W^\pm\gamma)^2$	55
6.2	Effects of sources of uncertainty on the upper limits.	55
A.1	Relative difference between smeared and truth energy of photons in one bin.	68
A.2	Photon energy detector resolution for $ \eta < 2.5$	68
A.3	Photon η and ϕ detector resolutions for $ \eta < 2.5$	69

List of Tables

2.1	The SM fermions.	5
3.1	List of standard model background processes.	20
4.1	The charged hadron tracking efficiency used in the DELPHES simulation. . . .	24
5.1	Values for the background discrimination cuts.	37
5.2	Separation between jets and photons allowed in a viable grouping.	39
5.3	Cut flow for dominant background and two signal test masses.	40
5.4	Final signal and background efficiencies.	45
5.5	Scale and PDF uncertainties on signal efficiencies.	51
5.6	Scale and PDF uncertainties on background efficiencies.	52
5.7	Statistical error on signal efficiencies.	52
5.8	Statistical error on background efficiencies.	53
6.1	GM model parameters and the resulting values of $m_{H_5^\pm}$	56
6.2	$\sigma(pp \rightarrow H_5^+ H_5^-)$ in the GM model.	57

Abstract

Many models that extend the Standard Model Higgs sector predict the existence of at least one charged Higgs boson, in addition to the neutral Higgs boson expected from the Standard Model. While many searches have been performed, a charged Higgs boson has not yet been observed. Using simulated data, the sensitivity of a search for a charged Higgs boson after the LHC Run 3 was studied. An analysis for a search in the channel $pp \rightarrow H^+H^- \rightarrow (W^+\gamma)(W^-\gamma) \rightarrow (jj\gamma)(jj\gamma)$ with a charged Higgs with a mass greater than 100 GeV was developed, and expected upper limits on the production cross-section times branching ratio of this process, $\sigma(pp \rightarrow H^+H^-) \times BR(H^\pm \rightarrow W^\pm\gamma)^2$, were calculated. Limits of $95.1 - 1.88$ fb were set for charged Higgs bosons with masses in the range $130 - 500$ GeV.

Résumé

Il existe de nombreux modèles théoriques qui prédisent l'existence non seulement du boson Higgs neutre du Modèle Standard, mais aussi d'au moins un boson Higgs électriquement chargé. Malgré les recherches déjà accomplies, des traces indiquant l'existence d'un tel boson Higgs électriquement chargé n'ont pas jusqu'ici été révélées. En utilisant des données simulées, les perspectives d'observer un boson Higgs chargé, en utilisant les données qui seront enregistrées durant la période LHC Run 3, ont été étudiées. Une analyse a été développée afin d'identifier des données compatibles avec la réaction $pp \rightarrow H^+H^- \rightarrow (W^+\gamma)(W^-\gamma) \rightarrow (jj\gamma)(jj\gamma)$ avec un Higgs chargé de masse plus grande que 100 GeV. De plus, des limites attendues ont été calculées sur la valeur de la section efficace fois le rapport de branchement de ce processus, $\sigma(pp \rightarrow H^+H^-) \times BR(H^\pm \rightarrow W^\pm\gamma)^2$. Des limites entre 95.1 – 1.88 fb ont été obtenues pour des masses du Higgs chargé entre 130 – 500 GeV.

Acknowledgements

I would like to thank first and foremost my M.Sc supervisor, Prof. Brigitte Vachon, for giving me the opportunity to work in a field I am passionate about, for her continuous support and guidance throughout the execution of this work, and for reviewing and providing feedback on this thesis.

I am very grateful to the ATLAS group at McGill for providing a stimulating environment with a wide range of expertise from which to learn and develop skills both related to, and separate from my own research. In particular I would like to thank some senior members of the group — Anabel Chuinard, Robert Keys, Sebastien Prince, Heather Russell, and Tamara Vasquez-Schroeder — for providing assistance when the path forward was not very clear.

I would also like to thank Rasa Rejali for her continued support in and beyond this project, making long and stressful work days more enjoyable, listening to and providing feedback on my ideas about my research, and for helping me plan my academic future.

Finally, I would like to thank my parents, Farid Haddad and Hind Kaddah, for their constant encouragement in my pursuit of my academic goals, and their support in all aspects of my life. Without them none of my accomplishments would have been possible.

Chapter 1

Introduction

The question of what matter is made of at the most fundamental scales has captivated scientists and philosophers for millennia. However all discoveries of subatomic particles have only occurred in the last 120 years, with particle collider experiments playing seminal roles in many of these discoveries. Such experiments have come a long way, from Rutherford’s gold foil experiment discovering the atomic nucleus [1], to more modern and powerful *synchrotrons* [2] — which use electromagnetic fields to accelerate and steer particles into high energy collisions — observing particles that only exist for fractions of a second. Currently, the largest and most powerful particle collider is the Large Hadron Collider (LHC) [3] at the European Organization for Nuclear Research (CERN), in Geneva, Switzerland.

In operation since 2008, the LHC collides individual protons at unprecedented energies, and hosts several detectors that record the results of these collisions. In 2012, two of these detectors, belonging to the ATLAS [4] and CMS [5] collaborations, were responsible for the discovery of the long anticipated Higgs boson [6, 7]. The existence of the Higgs boson was predicted by the Standard Model of particle physics (SM) [8], and its discovery solidified the SM as the most complete and successful theory of physics on subatomic scales — a distinction which the SM holds to this day. Indeed, before its discovery the Higgs boson was the only particle predicted by the SM that had not yet been discovered.

Despite the successes of the SM, however, it does not provide a complete account of all phenomena observed in the universe. For instance, one of the most well known shortcomings of the SM is that it does not predict or explain the nature of dark matter, a type of matter much more abundant than the "regular" matter described by the SM. Insights into the nature of dark matter – and into several other problems not addressed by the SM – requires looking beyond the SM. Thus, moving forward, the most pressing question in experimental particle physics is what, if anything, will be found in particle collider experiments henceforth. From a theoretical perspective, there are many models extending the SM that may hold the answer to this question.

A subset of the extensions of the SM addresses its problems by postulating the existence of more than one Higgs Boson. These extensions take advantage of the fact that the mechanism by which the Higgs boson emerges in the SM is not unique. The work presented herein focuses on one common prediction of models with several Higgs bosons: the existence of an electrically charged Higgs boson. The observation of this particle and measurements of its properties would have deep implications for the landscape of physics beyond the SM. Many searches have already been performed for a charged Higgs boson, but none have found any hints of the existence of such a particle (see Section 2.2 for some examples).

The objective of this work is to study the feasibility of observing a charged Higgs boson – should one exist – using the LHC multi-purpose detectors, after the LHC Run 3. This is done for a 4-jet-2-photon final state, produced by the decay of pair-produced charged Higgs bosons into one photon and one W boson each, and the subsequent hadronic decay of the W bosons. In addition to being a novel search channel, this final state is interesting to study because of the capability of detectors to reconstruct all objects produced in the reaction. To this end, this work develops a technique to search for this process and separate it from potential background reactions. The sensitivity of the analysis developed to search for the existence of a charged Higgs boson is quantified in terms of expected limits on the production cross-section times branching ratio, $\sigma(pp \rightarrow H^+H^-) \times BR(H^\pm \rightarrow W^\pm\gamma)^2$, for

different charged Higgs boson test masses. Such limits represent upper limits on this quantity in the event that the analysis described herein provides no evidence for a charged Higgs boson when applied to real data.

The layout of the thesis is as follows. Chapter 2 elaborates on how the Higgs boson arises in the SM, explores popular extensions to the SM Higgs sector, and presents the model under study in this work. Details of the generation of simulated signal and background events are explained in Chapter 3. Chapter 4 is dedicated to an explanation of the detector simulation and how the generated datasets are processed into the samples that are analyzed. The analysis technique used is detailed in Chapter 5, along with a discussion of the uncertainties considered. Results are presented and are interpreted in the context of the Georgi-Machacek model in Chapter 6. Finally, Chapter 7 presents a summary and outlook of the work.

Chapter 2

The Standard Model and Beyond

The SM [8] is a theory of the most fundamental constituents of the universe and their interactions. It describes seventeen fundamental particles and three of the fundamental forces through which they interact: the electromagnetic force, the weak nuclear force, and the strong force. These seventeen particles can be divided into three categories.

The first category is that of *vector bosons*, which comprises four of the particles in the SM. As *bosons*, their defining property is that they possess integer spin quantum number. They are termed "vector" bosons because they all have spin quantum number 1. They are responsible for mediating the fundamental forces. The vector bosons in the SM are:

- the photon (γ): perhaps the most well-known member of the SM, the photon is responsible for mediating the electromagnetic interaction.
- the neutral Z boson (Z^0): responsible for mediating the weak nuclear force.
- the charged W bosons ($W^{+/-}$): responsible for mediating the weak nuclear force.
- the gluon (g): responsible for mediating the strong force.

The second category of known elementary particles is that of *fermions*, comprising twelve particles in the SM. In contrast to bosons, fermions have half-integer spin quantum number

Generation	1	2	3
leptons	electrons, e^-	muons, μ^-	tauons, τ^-
	electron neutrinos, ν_e	muon neutrinos, ν_μ	tau neutrinos, ν_τ
quarks	up, u	charm, c	top, t
	down, d	strange, s	bottom, b

Table 2.1: Fermions of the SM. Fermions in the same row have the same electric charge, with electron-type leptons having charge (in units of the elementary charge) -1 , neutrinos being neutral, up-type quarks having charge $+2/3$, and down type quarks having charge $-1/3$.

— in fact all fermions in the SM have spin $1/2$ — and make up matter. Fermions in the SM are divided into three *generations* of four fermions each (Table 2.1). Each generation is further divided into two *leptons* — fermions that don’t interact via strong force — and two *quarks* — fermions that do. The masses of the fermions generally increase as one goes to the right in a row of Table 2.1. This is referred to as a *normal mass hierarchy*. It remains unknown whether the neutrinos obey this hierarchy.

The number of fundamental particles described by the SM nearly doubles if one counts *antimatter* particles. All electrically charged particles in the SM have an anti-particle partner that shares all the same interactions and properties as the ”normal” particle, but has opposite sign electric charge. It remains unknown whether neutrinos have antiparticle partners distinct from themselves, or are their own antiparticles [9].

The final category is reserved for the Higgs boson, h . Strictly speaking, the Higgs boson can be categorized with the vector bosons, as they are all bosons with integer spin. However the Higgs boson is differentiated from the vector bosons since it has spin quantum number 0 (it is a *scalar boson*). The distinction of the Higgs boson from all other particles in the SM manifests in another way: it is a boson, so it does not make up matter, yet it does not mediate any fundamental forces either.

Despite these facts, the existence and properties of the Higgs boson have profound consequences for the SM. In the SM it is responsible for giving mass to the W and Z bosons and all fermions except neutrinos, and simultaneously allowing for the unification of the electromagnetic and weak interactions above the electroweak scale [10]. Perhaps its most

important feature, however, is that its existence prevents the divergence of the WW scattering amplitude, that would otherwise violate unitarity at energies much higher than the W mass [11]. To do all of this there must exist at least one Higgs boson, but there is no reason that several other Higgs bosons cannot exist and contribute to these phenomena as well.

The rest of this section is dedicated to a discussion of the Higgs boson within (Section 2.1) and beyond (Section 2.2) the SM, and of the particular model of a charged Higgs boson studied in this work (Section 2.3).

2.1 The Standard Model Higgs Sector

The SM assumes the existence of one scalar *Higgs field* ϕ in nature [8]. The Higgs field transforms under a subset of the SM symmetries, represented by the product group $SU(2) \times U(1)$. This product group describes the unification of the weak nuclear force and the electromagnetic force at high energies. At these energies, the four aforementioned weak and electromagnetic vector bosons are replaced by four massless vector gauge bosons that correspond to this product group: W_i^μ with $i = 1, 2, 3$ for $SU(2)$, and B^μ for $U(1)$. It is these bosons that the Higgs field directly interacts with.

At low energies, the Higgs field exhibits a behaviour that is unique to it in the SM: it acquires a non-zero *vacuum expectation value* (vev), denoted by $\langle \phi \rangle = v$. This means that the numerical value of the field ϕ in the configuration that minimizes its potential energy is non-zero, and results in the breaking of the $SU(2) \times U(1)$ symmetry of the SM [10]. The breaking of this symmetry due to the non-zero vev of the Higgs field is called *spontaneous symmetry breaking*. Perturbations around the minimum energy configuration are identified with the Higgs boson, h .

The non-zero vev of the Higgs field causes the $SU(2) \times U(1)$ gauge bosons to be rearranged

into the massive W and Z bosons, and the massless photon [10]:

$$W_\mu^\pm = \frac{W_{1,\mu} \mp iW_{2,\mu}}{\sqrt{2}}, \quad (2.1)$$

$$Z_\mu = z_1 W_{3,\mu} - z_2 B_\mu, \quad (2.2)$$

$$\gamma_\mu = z_2 W_{3,\mu} + z_1 B_\mu, \quad (2.3)$$

where $z_{1,2}$ are constants. The generation of masses through spontaneous symmetry breaking is known as the *Higgs mechanism*. The masslessness of the photon indicates that the gauge group representing electromagnetism — $U(1)_{\text{em}}$ — is a symmetry of the SM after the Higgs field acquires a non-zero vev [8]. All interactions of the SM must therefore exhibit this $U(1)_{\text{em}}$ symmetry, which manifests itself in the conservation of electric charge before and after any reaction.

2.2 Extensions of the Standard Model Higgs Sector

Though the SM has proven to be a remarkably accurate theory, it is well known that it does not describe every subatomic phenomenon observed in nature. Extensions of the SM are developed with the intention of solving one or more of these shortcomings, by introducing new particles and allowing them to interact with SM particles. The subset of SM extensions that enlarges the SM Higgs sector does so by including new scalar fields that acquire non-zero vevs and thus partake in spontaneous symmetry breaking.

Two of the more studied extensions of the Higgs sector are the Georgi-Machacek (GM) model [12] and the Two-Higgs-Doublet Model (2HDM) [13]. The former is motivated by the fact that the spontaneous symmetry breaking mechanism in the SM is the simplest possible scenario. The GM model aims to provide an alternative that is consistent with observations, in order to better understand different possible spontaneous symmetry breaking mechanism. The 2HDM addresses the generation of neutrino masses while simultaneously predicting the

existence of a particle that could contribute to the make up of dark matter.

Both of these models add a second scalar field χ to the SM, and are different in the exact manner in which χ transforms under $SU(2)$. In the SM, ϕ can be thought of as having two components which transform into one another under an $SU(2)$ transformation — it is an $SU(2)$ *doublet*. The field χ of the 2HDM is also a $SU(2)$ doublet (hence the name of the model), whereas the field χ of the GM model has three components that transform into one another under an $SU(2)$ transformation — it is an $SU(2)$ *triplet*. They differ from one another even more in the interactions they allow between the field χ and the SM particles: the GM model allows χ to interact with all particles of the SM that ϕ interacts with, whereas different variations of the 2HDM allow different interactions between χ and SM particles.

As a result of spontaneous symmetry breaking, these models predict the existence of a scalar particle consistent with the Higgs boson of the SM and several other scalar particles. Most notably, both models produce charged scalar particles — also referred to as charged Higgs bosons. Other models that produce scalar particles in addition to the SM Higgs boson, such as supersymmetric extensions to the SM [14], also predict the existence of charged Higgs bosons. Results from searches for a charged Higgs boson can therefore be reinterpreted in the context of several models, and can be used to constrain the parameter space of many of these models. There have been several such searches, and in the absence of evidence for the existence of a charged Higgs boson, both model-dependent and model-independent limits have been set on its mass and cross-section times branching ratio for various processes. Some existing experimental limits include

- the lower bound of 78.6 GeV on the mass of charged Higgs boson consistent with the 2HDM by LEP [15],
- a model independent limit on $\sigma_{\text{VBF}}(H^\pm) \times BR(H^\pm \rightarrow W^\pm Z)$ of $573 - 36$ fb in the mass range $200 - 2000$ GeV by CMS [16],
- a model independent limit on $\sigma(pp \rightarrow [b]\bar{t}H^\pm) \times BR(H^\pm \rightarrow \tau\nu)$ of 1.9 pb to 15 fb in

the mass range $200 - 2000$ GeV by ATLAS [17].

With much of the focus in particle physics on phenomena beyond the SM, and the fact that an extended Higgs sector offers plausible descriptions for such phenomena, it is worthwhile to develop analyses sensitive to an extended Higgs sector. Furthermore a charged Higgs boson could provide evidence for — and have the potential to discriminate between — some of the most prominent extensions of the SM Higgs sector. Understanding what kind of signatures are characteristic of a charged Higgs boson and the sensitivity of current detectors to these signatures can therefore help motivate specific searches for new physics, as well as lay the ground work for analyses used by these searches.

2.3 The Model: A Generic Charged Higgs Boson

Experiments are agnostic to the specifics of spontaneous symmetry breaking; all that can be observed in the laboratory are the decays of any Higgs boson, regardless of how they emerge. This fact allows for phenomenological studies — such as this one — to be performed without reference to a complete model of spontaneous symmetry breaking, and as a result be applicable to many models extending the SM Higgs sector. In this section the effective theory that enables the study presented herein is introduced, considering only experimentally observable quantities without making reference to any underlying theory of spontaneous symmetry breaking.

The decay of the charged Higgs that is under study — and therefore that needs to be modelled — is described symbolically as

$$H^\pm \rightarrow W^\pm \gamma, \tag{2.4}$$

where the charged scalar has been represented by a complex field H^+ ($H^{+*} = H^-$). The

effective Lagrangian describing this process is

$$\begin{aligned}\mathcal{L}_{H^+} &= D^\mu H^+ D_\mu^* H^- - m_{H^\pm}^2 H^+ H^- - \mathcal{L}_{\text{int}}, \\ \mathcal{L}_{\text{int}} &= g_{HW\gamma} H^+ W_{\mu\nu}^- F^{\mu\nu} + \text{h.c.}, \quad D^\mu = \partial^\mu + ieA^\mu + ig_Z Z^\mu,\end{aligned}\tag{2.5}$$

where $m_{H^\pm}$ is the mass of the charged scalar, Z_μ is the Z -boson field, $g_{HW\gamma}$ is an effective coupling constant, $W_{\mu\nu}^-$ is the W^- -boson field strength, $F_{\mu\nu}$ is the electromagnetic field strength, A_μ is the photon field, e is the fundamental charge, and g_Z is a coupling constant. The field strengths are given by $W_{\mu\nu}^- \equiv D_\mu^A W_\nu^- - D_\nu^A W_\mu^-$ and $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$, where $D_\mu^A \equiv \partial_\mu - ieA_\mu$.

This Lagrangian is largely model-independent. The couplings of the charged Higgs to the photon must be present for a charged scalar to preserve the $U(1)_{\text{em}}$ electromagnetic gauge symmetry of the SM — if they weren't present the scalar could not be charged. Moreover any pair production involving a photon has an identical pair production mode where the Z -boson replaces the photon due to the definitions of the Z -boson and the photons in terms of the $SU(2) \times U(1)$ gauge bosons (Equations 2.2 and 2.3). Furthermore, the coupling of the charged Higgs to a photon and a W boson is also generic: even in a model (such as the GM model [18]) where there are no tree level couplings between a charged Higgs and these two bosons, such an interaction can be generated at loop level and the model presented herein would represent the effective coupling of such a theory. Figure 2.1 shows the Feynman diagram that describes the signal process.

The only model-dependence in this Lagrangian arises from the values of the couplings g_Z and $g_{HW\gamma}$; however, since the sensitivity of the search presented in this manuscript is quantified in terms of limits on the production cross-section times branching fraction of the process of interest, the precise values of these couplings are unimportant.

Since the $\gamma H^+ H^-$ coupling is fixed by gauge invariance, the value of g_Z is the only free parameter, affecting the fraction of $H^+ H^-$ pairs that would be produced by an s-

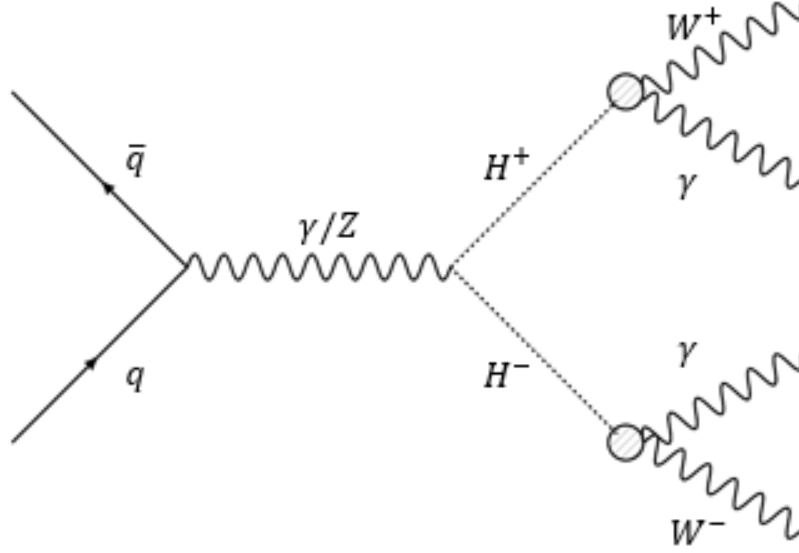


Figure 2.1: Feynman diagram for the process $pp \rightarrow H^+H^- \rightarrow W^+\gamma W^-\gamma$. The blobs at the $H^\pm W^\mp \gamma$ vertices indicate that the decay of the charged Higgs is described by an effective Lagrangian. In a full model the blob may be replaced by a vertex or a loop.

channel photon to those produced by an s-channel Z -boson. In the VH^+H^- interaction ($V = \gamma, Z$) the Z -boson and the photon differ from one another only by their masses; the Z -boson is massive whereas the photon is not. In the computation of cross-sections for $pp \rightarrow V^* \rightarrow H^+H^-$, where $*$ means V is virtual, this difference manifests itself in the propagator P_V of V , as shown schematically in Equation 2.6.

$$P_V \sim \frac{1}{p^2 - m_V^2}. \quad (2.6)$$

For a mass $m_{H^\pm} \geq m_Z$, energy conservation at the VH^+H^- vertex requires $p^2 \geq 4m_{H^\pm}^2$. The width of the Z -boson is only approximately 1/36th of its mass [19], so requiring $p^2 \geq 4m_{H^\pm}^2 \gtrsim 4m_Z^2$ ensures that p^2 is very far away from the m_Z^2 pole. It is therefore expected that the mediator should have little to no effect on decay kinematics.

By this reasoning, the pair-production kinematics of the charged Higgs will be influenced by the mass of the virtual boson mediating the pair-production more strongly for lower $m_{H^\pm}$. Such an influence would manifest itself in an excess of events in the di-Higgs invariant

mass distribution close to the mass of the virtual boson. To verify the dependence of the pair-production kinematics on the type of virtual boson exchange, the di-Higgs invariant mass $m_{H^+H^-}$ is examined for decays involving only a photon and decays involving only a Z boson as the virtual boson. This is done for the lightest mass analyzed, which, as will be justified in Section 5, is $m_{H^\pm} = 130$ GeV. Charged Higgs four-vectors are pair-produced using MadGraph [20] as described in Section 3.1. The di-Higgs invariant mass distributions are shown in Figure 2.2(a) for $m_{H^\pm} = 130$ GeV, with Figure 2.2(b) showing the distributions obtained assuming a higher charged Higgs mass to demonstrate any dependence on $m_{H^\pm}$.

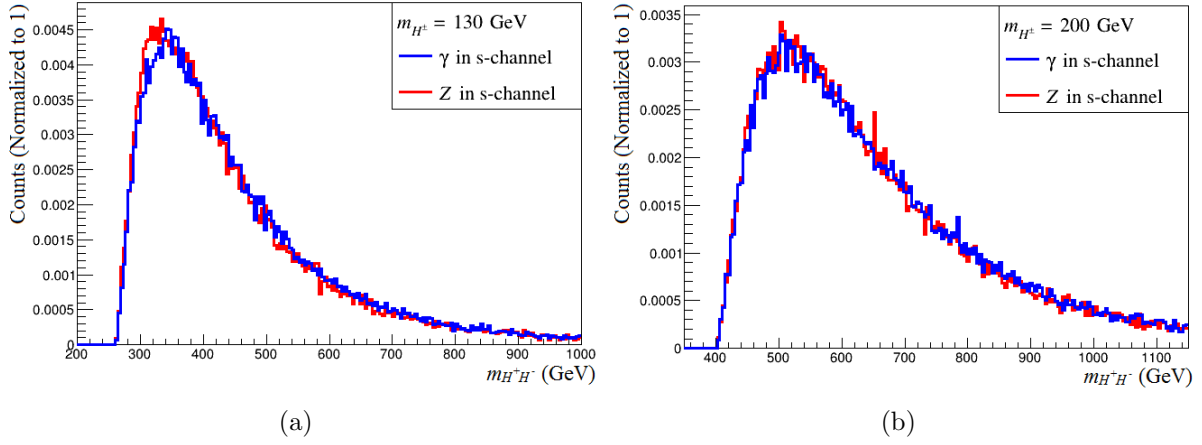


Figure 2.2: The di-Higgs invariant mass for (a) $m_{H^\pm} = 130$ GeV and (b) $m_{H^\pm} = 200$ GeV with a photon and, separately, a Z boson mediating charged Higgs boson pair production.

This figure shows a slight difference in the distributions at low masses, where the charged Higgs pair production mediated by Z bosons exhibit an excess over the production mediated by photons. In the case where both photons and Z bosons are allowed to mediate the production, the distribution interpolates between the two shown, with the exact distribution depending on the relative values of e and g_Z . Since the difference in kinematics in Figure 2.2(a) is small and disappears with increasing $m_{H^\pm}$, the exact value of g_Z is interpreted as being unimportant for the purpose of this work.

In the case of the coupling constant $g_{\gamma HW}$, its value only affects the charged Higgs decay rate by scaling its value, but the kinematic distributions of the decay products are fixed by

the kinematics of the charged Higgs. In order to set limits on the production cross-section times branching ratio only the kinematic distributions are important, so the value of this coupling constant is also irrelevant to the analysis.

It must be verified that the effective $H^\pm W^\mp \gamma$ coupling has the relevant symmetries of the SM. The full symmetry group of the SM is $SU(3) \times SU(2) \times U(1)$ [8] before symmetry breaking, with $SU(3)$ representing the strong force. As described above, spontaneous symmetry breaking reduces this group to $SU(3) \times U(1)_{\text{em}}$. Therefore this effective coupling must be invariant under $SU(3)$ and $U(1)_{\text{em}}$. Furthermore, a quantum field theory must incorporate special relativity — done by requiring Lorentz invariance of the theory — and must exhibit Hermiticity — in order for probabilities derived from the theory to add to 1 [8].

- $SU(3)$ gauge invariance: The electromagnetic field strength tensor $F^{\mu\nu}$ and the W boson field strength tensor $W^{\mu\nu}$ are invariant under $SU(3)$ since the electromagnetic and weak vector bosons do not interact via the strong force. Similarly, since the charged Higgs is part of a model of spontaneous symmetry breaking, it can only interact via the electromagnetic and weak forces. As all fields in the $H^\pm W^\mp \gamma$ coupling are invariant under $SU(3)$, the coupling itself is invariant as well.
- $U(1)_{\text{em}}$ gauge invariance: The electromagnetic field strength tensor $F^{\mu\nu}$ is gauge invariant by definition. Furthermore, since H^+ and W^- have opposite charges under $U(1)_{\text{em}}$, they transform oppositely under a $U(1)_{\text{em}}$ gauge transformation: if $H^+ \rightarrow e^{i\alpha(x)}$ then $W_\mu^- \rightarrow e^{-i\alpha(x)}$. The only subtlety arises from the W^- -boson field strength. Under a local $U(1)_{\text{em}}$ gauge transformation, $W_{\mu\nu}^-$ transforms as

$$W_{\mu\nu}^- = D_\mu^A W_\nu^- - D_\nu^A W_\mu^- \quad (2.7)$$

$$\rightarrow [\partial_\mu - ieA_\mu + i\partial_\mu\alpha(x)] (e^{-i\alpha(x)} W_\nu^-) \quad (2.8)$$

$$\begin{aligned} & - [\partial_\nu - ieA_\nu + i\partial_\nu\alpha(x)] (e^{-i\alpha(x)} W_\mu^-) \\ & = e^{-i\alpha(x)} \partial_\mu W_\nu^- - ie^{-i\alpha(x)} W_\nu \partial_\mu \alpha(x) - ieA_\mu e^{-i\alpha(x)} W_\nu^- + i\partial_\mu \alpha(x) e^{-i\alpha(x)} W_\nu^- \end{aligned} \quad (2.9)$$

$$\begin{aligned}
& -e^{-i\alpha(x)}\partial_\nu W_\mu^- + ie^{-i\alpha(x)}W_\mu\partial_\nu\alpha(x) + ieA_\nu e^{-i\alpha(x)}W_\mu^- - i\partial_\nu\alpha(x)e^{-i\alpha(x)}W_\mu^- \\
& = e^{-i\alpha(x)}W_{\mu\nu}^-,
\end{aligned} \tag{2.10}$$

so the combination $H^+W_{\mu\nu}^-$ is also gauge invariant. Therefore, the effective coupling is gauge invariant under the electromagnetic gauge group.

- Lorentz invariance: All Lorentz indices appearing in the $H^\pm W^\mp \gamma$ interaction term are contracted, so the interaction is Lorentz invariant.
- Hermiticity: Every term in the Lagrangian 2.5 is either Hermitian or has its Hermitian conjugate explicitly included. The Lagrangian is therefore Hermitian.

The effective field theory described by Equation 2.5 does not fully describe the charged Higgs under study: in order to encode the total decay width of the charged Higgs, $\Gamma_{H^\pm}$, all vertices involving the Higgs must be present in the Lagrangian. Doing so entails writing a complete model that includes the charged Higgs and is not the aim of the work. Thus the value of $\Gamma_{H^\pm}$ required for the simulation of data is treated as an external input parameter, and is not calculated based on all possible charged Higgs interactions according to a particular model.

There are two experimentally relevant regions to consider; either $\Gamma_{H^\pm}$ is larger or smaller than the detector resolution. Each region requires a specifically tailored analysis technique. This work focuses on the case where $\Gamma_{H^\pm}$ is smaller than the detector resolution. In this region the value of $\Gamma_{H^\pm}$ does not affect the signal in the detector, thus allowing the results obtained to be extended to any model in which a charged Higgs has a total decay width smaller than the detector resolution. The detector energy resolutions, as quantified in Section A.2, are dominated by the jet energy resolution. For energies around the charged Higgs masses analyzed, this resolution is on the order of 10's of GeV. Therefore, for the simulation of the process of interest, the value of $\Gamma_{H^\pm}$ is chosen to be 1 GeV.

Chapter 3

Data Simulation

The goal of this work is to lay the foundation for an analysis strategy to search for a charged Higgs boson in proton-proton collisions, assuming an integrated data luminosity of 300 fb^{-1} [21]. In the context of the LHC, this quantity of data is foreseen to be available after the LHC Run 3, when protons will be collided at a center of mass energy of 14 TeV. The work presented herein is based on simulated data samples of signal and background events, with simulated signal events corresponding to the reaction described in Section 3.2.

This section provides a description of the software tools and specifications used to generate signal and background event samples. It comprises an explanation of the event simulation chain in Section 3.1, followed by a definition of the signal under study in Section 3.2, and concluded with a discussion of background considerations in Section 3.3.

3.1 Simulation Chain

Five programs are used in the generation of signal and background samples. The first of these is **FeynRules** [22], used to define the extension to the SM described above. **FeynRules** outputs the model in a Universal FeynRules Output (UFO) [23] format, that is read by matrix-element (ME) generators for the simulation of events. As all of the background

processes occur in the SM and UFO files for the SM are readily available, it is only necessary to generate UFO files for the signal.

The aforementioned UFO file is then used by **MadGraph** [20] to generate kinematics for signal events and all expected dominant SM background processes, with the exception of the *QCD background* (defined in the caption of Table 3.1), which is generated using **Sherpa** [24]. **MadGraph** and **Sherpa** are ME generators, and events generated by them are referred to as *truth-level* events. At this level, all decays and scatterings conserve momentum and energy exactly.

Due to the fact that protons are composite particles, the collision of two protons is actually a collision of its constituents: quarks and gluons. Of utmost importance in the simulation of proton-proton collisions is the ability to describe which particles within the protons will collide and with what fractions of the proton’s total energy. This information cannot be calculated using perturbation theory, and is instead encoded in the empirically determined *parton distribution function* (PDF). ME generators must make use of a PDF set to produce physically plausible results. In this work, signal and background events are generated from proton-proton collisions with the NNPDF23_lo_as_0130_qed [25, 26] PDF.

It is possible, and is in fact often the case, that the outgoing particles of a particular SM process are unstable. This instability can mean one of two things, or both together: among the outgoing particles there may be one or more particles that decay spontaneously into lighter SM particles, or one or more strongly interacting particles. In the first case, such particles will decay until there are no longer any spontaneously decaying particles in the final state.

Instability in the second case is slightly more complicated. The strong force is *confining*, meaning that particles that interact through it cannot exist independently, but must instead always appear in bound states that are uncharged under the $SU(3)$ group that represents the strong force — in other words, they must appear in *colorless* bound states. At the LHC, however, the energies are too high for stable bound states (called *hadrons*) to form immediately

after collisions. Instead, quarks and gluons produced in a hard-scattering interaction between two protons undergo so-called *parton showering* and hadronization processes, through which large numbers of additional quarks and gluons are produced along the trajectories of the original quarks and gluons [27]. This continues until there is sufficiently low kinetic energy among the quarks and gluons to allow for hadrons to form. Phenomenologically, this results in collimated *jets* of hadrons, which is what is observed in detectors.

MadGraph only generates event kinematics and estimates cross-sections and decay rates, but it does not decay or hadronize outgoing truth-level particles. Unstable particle decay, parton showers, and hadronization for samples generated with **MadGraph** are performed using **Pythia8** [28, 29]. **Sherpa** is able to decay outgoing particles and simulate parton showers and hadronization processes as well as generate truth-level events, so simulated events contributing to the QCD background include outgoing particle decay, parton showers, and hadronization using **Sherpa**.

The final step in the simulation of event samples is the detector simulation. **DELPHES** [30] provides fast detector simulations intended for phenomenological studies, and is used to estimate the detector response to signal and background processes. Events produced by the detector simulation are referred to as *detector-level* or *reconstructed* events. In contrast to truth-level events, detector-level kinematics only approximately conserve energy and momentum. This is due to uncertainties in the reconstructed objects' four-momenta as a result of resolution effects of a real detector (see Section 4.2).

A final phenomenon that must be considered in the simulation of data samples is the so-called *pileup*. At the LHC, proton bunches cross one another in the detectors millions of times per second, with several tens of low-energy (*soft*) interactions occurring per proton bunch crossing that are not head-on (*hard*) collisions. The presence of these simultaneous soft interactions results in additional low-energy particles being detected. These extra particles that do not originate from a hard-scatter event are what is referred to as pileup. At the LHC, the amount of pileup is quantified by the average number of interactions per bunch

crossing $\langle\mu\rangle$. Pileup must be incorporated in the generation of simulated data samples for an accurate simulation of proton-proton collisions at the LHC.

There are two steps necessary for the inclusion of pileup effects in the data simulation chain. The first is to generate the physical events that cause the pileup. These are called *minimum bias* events, referring to the fact that they only pass *triggers* — sets of conditions for the recording of an event, determined by a combination of the sensitivity of the detectors, and kinematic requirements enforced by reconstruction software — with the least stringent criteria for selection, and hence triggers that produce the least biased sample of the physics processes occurring in a bunch crossing. This first step is done using `Pythia8`. Next, these minimum bias events must be mixed with the simulated signal and background event samples. This was done by `DELPHES` as part of the detector response simulation.

Pileup can significantly alter the kinematic distributions of processes under study, especially when the average number of interactions per bunch crossing is as high as that expected at the LHC Run 3. Therefore it is extremely important to apply event reconstruction techniques in the analysis of the data that attempt to subtract the contribution of pileup to a data sample. A discussion of the techniques employed to this end is saved until Chapter 4, when the detector simulation is discussed in greater detail.

A center-of-mass energy of 14 TeV, with the energy evenly divided between both protons, is assumed for the generation of all signal and background samples, with an average number of interactions per bunch crossing of $\langle\mu\rangle = 140$ — these will be the conditions at the LHC Run 3.

3.2 Signal

The signal process considered in this work is the pair-production of charged Higgs bosons, decaying as in Equation 2.4, with the W bosons forced to decay hadronically. Therefore, the

full process as generated at truth-level is

$$pp \rightarrow H^+ H^- \rightarrow (W^+ \gamma)(W^- \gamma) \rightarrow (jj\gamma)(jj\gamma). \quad (3.1)$$

This reaction is simulated for charged Higgs masses from 100 GeV to 250 GeV in increments of 10 GeV, and from 300 GeV to 500 GeV in increments of 50 GeV. A total of 1.2 million events are generated and analyzed for each signal test mass.

3.3 Background

Two categories of backgrounds are simulated. The first is the *Standard Model background* category, comprising SM processes that can reproduce the final state of the signal process of interest. Second, the *fake signal background*, in which physics objects are misidentified upon reconstruction, must be considered.

3.3.1 Standard Model Backgrounds

Several SM processes resulting in the same final state particles as those from the process of interest are considered as SM background processes. The full list of processes considered, along with their cross-sections, is given in Table 3.1. The cross-sections listed in this table are the estimates calculated by the ME generator used to simulate the events. Figure 3.1 shows an example Feynman diagram for each of the two types of background processes with the largest cross-sections. While only two strongly interacting particles are depicted in these diagrams, additional quarks and gluons can be produced at the particle shower stage. This leads to the possible reconstruction of four or more jets in the event, making the final state indistinguishable from that of the signal process.

The dominant background process will be the one that is observed most frequently in an analysis of proton-proton collisions. While the cross-sections determine which process

Process	Cross-section (fb)	Process	Cross-section (fb)
$pp \rightarrow \geq 2j + 2\gamma$	1.61×10^5	$pp \rightarrow W^\pm Z \gamma \gamma$	1.94×10^{-1}
$pp \rightarrow ggh$	8.68	$pp \rightarrow ZZ \gamma \gamma$	8.55×10^{-2}
$pp \rightarrow b\bar{b}h$	1.51	$pp \rightarrow W^+W^-h$	5.93×10^{-3}
$pp \rightarrow W^\pm h$	1.19	$pp \rightarrow hh \rightarrow b\bar{b}h$	5.24×10^{-3}
$pp \rightarrow W^+W^- \gamma \gamma$	9.31×10^{-1}	$pp \rightarrow W^\pm Zh$	2.42×10^{-3}
$pp \rightarrow Zh$	6.47×10^{-1}	$pp \rightarrow ZZh$	1.73×10^{-3}
$pp \rightarrow t\bar{t}h \rightarrow W^+bW^-\bar{b}h$	2.87×10^{-1}		

Table 3.1: All SM background processes considered in order of decreasing cross-section. All weak gauge bosons are forced to decay hadronically, and, unless explicitly stated, all Higgs bosons are forced to decay to two photons. The cross-section for the first process is calculated for matrix-element processes involving at least 2, and up to 5 jets (j). All further jets are generated at the particle shower and hadronization stage of the event generation. Events generated in the first process make up to so-called *QCD background* since all jets are produced from QCD processes.

would occur most frequently purely from a theoretical standpoint, identifying the dominant background requires a parameterization of the experimental likelihood that each background would be observed when it occurs. Combining the theoretical likelihood for the occurrence of a background process with the experimental *efficiency* of observation, ϵ , the dominant background will be the one with the largest *effective cross-section* $\sigma_{\text{eff}} = \sigma\epsilon$.

As σ is fixed by the SM, mitigating this type of background entails minimizing ϵ . The methods applied to this end are discussed in Section 5.1.2.

A total of 1.2 million events are generated and analyzed for each background process listed in Table 3.1.

3.3.2 Fake Signal Background

Particles produced in proton-proton collisions can be misidentified as the result of several effects, such as detector imperfections, the granular nature of the information collected by a typical detector, and limitations of the algorithms used to identify particles by interpreting energy depositions. Of particular relevance to this work is the probability that a jet is misidentified as a photon. It is assumed that this probability is 10^{-3} — a conservative

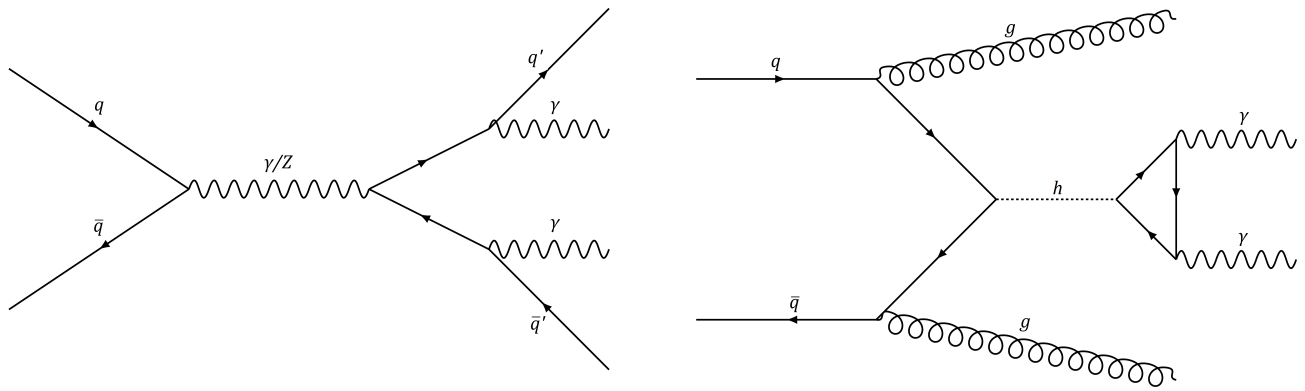


Figure 3.1: Example Feynman diagrams for the two types of background processes with the largest cross-section, the QCD background (left) and $pp \rightarrow ggh \rightarrow gg\gamma\gamma$ (right).

estimate approximating the fake rate observed by ATLAS with loose photon identification requirements [31]. DELPHES is thus configured to misidentify 1 jet of every 1000 as a photon. This effect is incorporated in the generation of all signal and background samples.

Chapter 4

Detector Simulation

In this chapter, the detector simulation performed by DELPHES is discussed in detail. In Section 4.1, the coordinates used to map the detector and to parametrize particle four-momenta are introduced, along with a breakdown of the different components of the detector. The remainder of the chapter is dedicated to a description of how events are reconstructed: Section 4.2 touches on the simulation of detector resolution effects, Section 4.3 outlines the algorithms used to reconstruct jets and photons from energy depositions in the detector, and finally Section 4.4 details the methods through which the pileup contamination of the simulated samples is estimated, and how this contamination is mitigated.

4.1 Geometry of the Detector

As particle detectors used in collider experiments typically have a cylindrical geometry, it is useful to parametrize four-vectors of physics object in terms of cylindrical coordinates. Suppose the proton beam lies along the z -axis. The following variables are used to describe the trajectory of a physics object in the detector:

- p_T , the transverse momentum: The component of the momentum orthogonal to the

beam axis,

$$p_T^2 = p_x^2 + p_y^2. \quad (4.1)$$

- η , the pseudorapidity: The angular separation of a physics object from the direction orthogonal to the beam axis. It is defined as

$$\eta = \tanh^{-1} \left(\frac{p_z}{|\vec{p}|} \right). \quad (4.2)$$

The value $\eta = 0$ corresponds to a physics object with momentum orthogonal to the beam axis, whereas $\eta \rightarrow \infty$ corresponds to one parallel to the beam axis (Figure 4.1). The pseudorapidity is used instead of the polar angle θ because, in the limit where the momentum of a particle is much larger than its mass — an approximation which is reasonable at LHC energies — the pseudorapidity transforms very simply under Lorentz boosts: it changes by an additive constant [32]. This implies that differences in η are Lorentz invariant, a useful property since the boosts along the beam axis of the partons undergoing the hard-scatter are unknown.

- ϕ , the azimuthal angle. ϕ lies in the range $[0, 2\pi)$ to cover the entire cylinder.

DELPHES simulates the response of an ideal detector — one with no gaps in its coverage, and no dead material. It comprises four subdetectors intended to measure the four-momenta of different types of particles. Starting with the subdetector closest to the beam and moving outward, these are (Figure 4.2)

- the *tracker*, where charged particles deposit energy in hits along their trajectory. From the hits in the tracker, *tracks* are reconstructed which map the trajectory of the charged particle. The tracker is simulated to cover the pseudorapidity range $|\eta| < 2.5$. The DELPHES simulation allows the user to define a track reconstruction efficiency for different species of charged particles. The only charged particles relevant to the signal

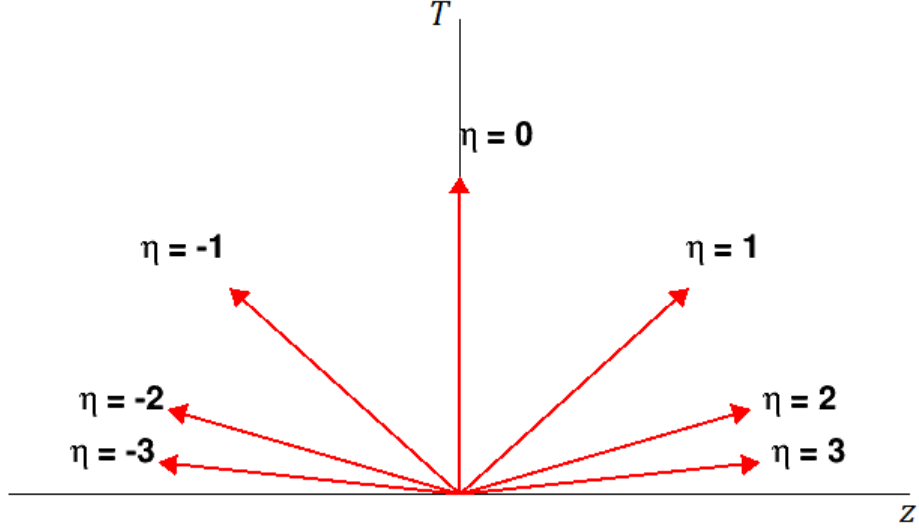


Figure 4.1: The direction of motion of an object starting at the origin with the indicated value of η . Here, the T axis represents any direction transverse to the beam.

	$ \eta \leq 1.5$	$1.5 < \eta \leq 2.5$	$ \eta > 2.5$
$p_T < 0.1$ GeV	0	0	0
$0.1 \text{ GeV} < p_T < 1$ GeV	0.7	0.6	0
$p_T > 1$ GeV	0.95	0.85	0

Table 4.1: The charged hadron tracking efficiency used in the DELPHES simulation, as a function of both η and p_T .

process are charged hadrons, whose track reconstruction efficiency is the DELPHES default, described in Table 4.1.

- the *barrel calorimeter*, which is further divided into an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). In the DELPHES idealization of a detector, electromagnetic objects such as photons, electrons and positrons are simulated to deposit all of their energy in the ECAL, and hadrons deposit all of their energy in the HCAL. In a real detector, there is some energy deposition by electromagnetic (hadronic) objects in the HCAL (ECAL). Both the ECAL and HCAL in the DELPHES simulation are assumed to be segmented into $\eta \times \phi$ *towers* of size 0.1×0.175 , covering the pseudorapidity range $|\eta| < 2.5$.

- the *end-cap calorimeter*, where particles produced at large $|\eta|$ ($2.5 < |\eta| < 5$) deposit their energy. The end-cap calorimeter is distinguished from the barrel calorimeter in its $\eta \times \phi$ segmentation of 0.2×0.35 .
- the *muon system*, where the trajectory of muons is measured. It covers the pseudorapidity range $|\eta| < 2.5$.

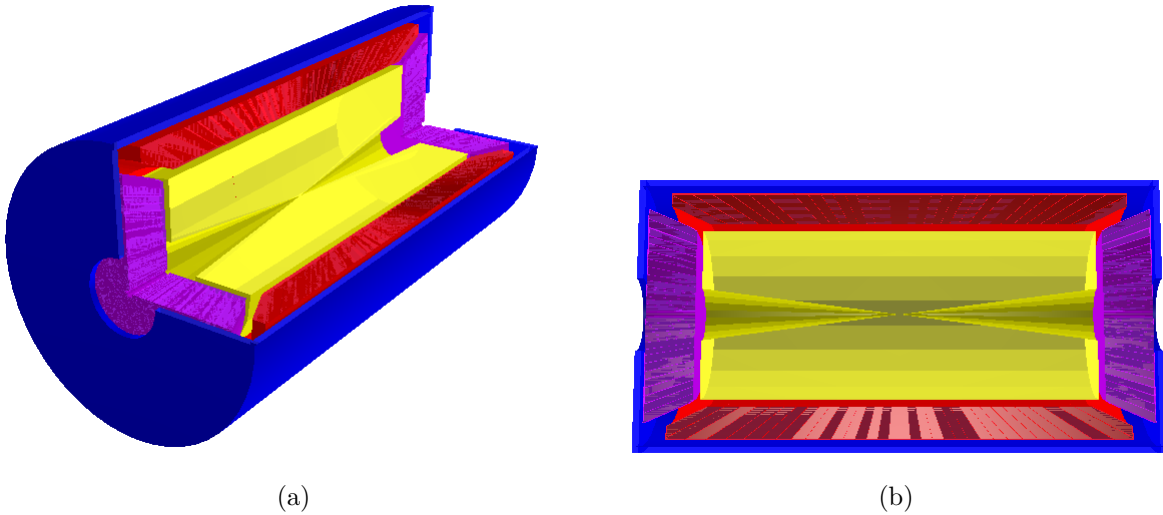


Figure 4.2: The DELPHES detector, visualized by the DELPHES event display. The tracker is shown in yellow, the calorimeter (including both the barrel and end-cap calorimeters) in red and purple, and the muon system in blue. The conical cut-out from the tracker depicts the $|\eta| = 2.5$ boundary of the tracker.

As neither the process of interest nor the backgrounds involved muons, only the tracker and the calorimeters are relevant to this work.

The DELPHES simulation reconstructs events by first generating measurements in each of the different detector subsystems based on the truth-level event information. The information recorded in the different subdetectors is then processed as described in the following sections.

4.2 Four-Momentum Smearing

One of the first steps in the DELPHES detector simulation is the slight alteration of truth-level four-vectors – referred to as *smearing* – to simulate uncertainties in object reconstruction due to detector resolutions. The amount of truth-level four-vector smearing is chosen to closely simulate the spatial and energy resolutions achievable in a typical collider experiment, and depends on the subdetectors in which a truth-level particle deposits energy. In DELPHES a kinematic quantity x is smeared according to

$$x_{\text{smeared}} = x_{\text{truth}} + \delta x, \quad (4.3)$$

where δx is a Gaussian distributed random variable with mean of 0 and width of σ_x . The width σ_x is referred to as the *resolution* of x , and is the parameter of interest when quantifying the smearing.

The knowledge of the experimental resolutions is necessary in order to perform the kinematic fit described in Section 5.2. A detailed account of the quantification of the experimental resolutions is given in Appendix A.

4.3 Object Reconstruction

DELPHES uses a simplified *particle flow* algorithm, which combines tracker and calorimeter information, to reconstruct the four-momenta of physics objects. The algorithm outputs *particle flow tracks* and *particle flow towers* which are determined as follows.

First, every track that is reconstructed is also stored as a particle flow track, originating from a charged particle. The particle flow tracks are then used to define the particle flow towers.

Consider a given calorimeter tower. Let E_e and E_h be the total energy measured in the ECAL and HCAL, respectively, for this particular calorimeter tower. Define $E_{e,\text{track}}$ to be

the sum of the energies of all charged particles that deposited energy in the ECAL at the given tower, and that had their tracks reconstructed. Define $E_{h,\text{track}}$ to be the HCAL analog. Finally, define

$$E_{\text{tower}} = \max(0, \Delta E_e) + \max(0, \Delta E_h), \quad (4.4)$$

$$\Delta E_{e/h} = E_{e/h} - E_{e/h,\text{track}}. \quad (4.5)$$

This quantity is calculated for each calorimeter tower, with $E_{e/h,\text{track}} = 0$ if there are no tracks pointing to a particular tower. A particle flow tower with energy E_{tower} is created whenever $E_{\text{tower}} > 0$. The collection of particle flow towers and tracks is used to reconstruct all physics objects.

4.3.1 Photon Reconstruction

Photons are reconstructed from electromagnetic particle flow towers that have no associated tracks pointing to them, provided they satisfied an isolation criterion:

For a photon γ_i the isolation variable is defined as

$$I(\gamma_i) = \frac{1}{p_T(\gamma_i)} \sum_{\substack{i \neq j \\ \Delta R(i,j) < \Delta R_{\text{max}} \\ p_T(j) > p_{T,\text{min}}}} p_T(\gamma_j), \quad (4.6)$$

where

$$\Delta R(i, j) = \sqrt{\Delta \eta^2(i, j) + \Delta \phi^2(i, j)}, \quad (4.7)$$

and ΔR_{max} and $p_{T,\text{min}}$ are taken to be the DELPHES defaults of 0.5 and 0.5 GeV respectively. Photons with $I(\gamma_i) > 0.12$, which is also the DELPHES default, are rejected.

This isolation requirement ensures that reconstructed photons do not originate from a hadronization process, in which case they would be expected to have nearby activity in the

detector.

4.3.2 Jet Clustering

Because of the composite nature of jets, their signature in a detector is much less compact than, say, a photon's. Furthermore, when a detector measures the deposition of several jets in close proximity, it can be difficult to tell where the signature of one jet ends and where another begins. This task, and the reconstruction of jets in general, is left to so-called *jet algorithms*, which cluster track and tower information into jets.

The anti- k_t [33] jet algorithm is used to cluster the particle flow tracks and towers into jets. This algorithm falls into a category of jet algorithms known as *sequential clustering* algorithms, which define jets by iterated reference to a metric, quantifying the distance between particles or *pseudojets* — objects composed of multiple particles, created before the end of the algorithm. The algorithm is the same among all sequential clustering algorithms, with the only difference between them being the metric used. Given a specific metric d , the general sequential clustering algorithm is as follows [33]:

1. Beginning with input object i in a set, calculate the distance $d(i, j)$ between i and every other input object j in the set. Calculate furthermore the distance $d(i, B)$ between input object i and the beam line.
2. Determine $\min[d(i, B), \min_j d(i, j)]$.
3. If $\min[d(i, B), \min_j d(i, j)] = d(i, B)$, promote object i to be a jet and remove it from the set of input objects. If $\min[d(i, B), \min_j d(i, j)] = \min_j d(i, j) = d(i, k)$ for some k , combine input objects i and k into a pseudojets by adding their four-momenta, and replace objects i and k with the pseudojets in the set of input objects.
4. Repeat until there are no input objects or pseudojets remaining in the set.

In the anti- k_t algorithm the distance metric used is

$$d(i, j) \equiv \min \left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2} \right) \frac{\Delta R^2(i, j)}{R^2}, \quad d_{iB} \equiv \frac{1}{p_{T,i}^2}, \quad (4.8)$$

where R is called the radius parameter which sets the scale for the size of the jets in $\eta - \phi$ space. At the end of the algorithm, jets are cones with radii R centered at the sum of the four-momenta contained in the jet. The value $R = 0.4$ is used for this work, chosen to minimize overlap between distinct objects while simultaneously maximizing the number of particles belonging to a jet that are included in the reconstructed jet.

The anti- k_T algorithm exhibits the important quality of *infrared safety* [33]. During hadronization it is common that quarks emit low energy (~ 1 GeV) gluons. Scattering amplitudes that describe this process have low-energy, or infrared (IR), divergences that make it a non-perturbative process; it is therefore desirable to work with observables that are insensitive to such a common yet difficult process to describe [34]. As high-momentum objects produce smaller values for the metric in Equation 4.8, these objects are clustered together first. Therefore, the summing of object four-vectors by this jet algorithm ensures that low energy objects do not have a significant effect on the reconstructed jet four-vectors.

4.4 Pileup Mitigation

Large amounts of pileup during LHC runs distort kinematic distributions and renders more difficult the measurements of the four-momenta of particles from the hard-scatter, due to the production and detection of many secondary low-energy particles that are not of interest. To mitigate the impact of pileup on the final reconstructed event, special algorithms are employed on an event-by-event basis. In the DELPHES detector simulation and event reconstruction, the pileup mitigation algorithm consists of two techniques.

The first technique reduces pileup due to charged particles for which a track has been

reconstructed. In the DELPHES simulation the hard-scatter always occurs at the geometrical center of the detector (the location of a scatter is referred to as a *vertex*). By contrast, as is the case in a real experiment, interactions producing pileup occur over a spatial extent around the hard-scatter event. This information can be used to identify tracks originating from pileup. When a track is reconstructed, the vertex from which the track originated can also be determined. If this vertex is further from the location of the hard-scatter than the resolution with which the hard-scatter vertex can be determined, the track is assumed to originate from pileup and is removed from the list of tracks considered for further analysis. The resolution used for this study is 100 microns, and is typical of current collider experiments [35].

The second technique mitigates the pileup contribution from neutral particles, as well as that from charged particles with no reconstructed tracks. This component of the pileup causes an overall increase in the energy of reconstructed jets because it gets clustered into the jets. To account for this, the component of a jet's energy that originates from pileup is estimated and subtracted from the jet energy. This is done by first calculating the *jet area*, which parameterizes the susceptibility of a jet to pileup contamination.

Once particle flow towers and tracks are clustered into jets, the jet area is computed for each jet using the *active area algorithm* [36]. After jets have been clustered, this algorithm adds a large number of infinitesimal p_T objects, called *ghosts*, to the collection of stable particles before jet clustering, randomly and densely in the $\eta - \phi$ plane. Jets are then clustered once more with the new collection of stable particles plus the set of ghosts, $\{g_i\}$. Thanks to the IR safety of the anti- k_t algorithm the addition of infinitely soft objects does not significantly alter the four-momenta of the reconstructed jets (indeed this area algorithm is only compatible with IR safe jet algorithms). Given the number density of ghosts in the $\eta - \phi$ plane, ν_g , the area of a jet j according to the specific distribution of ghosts, $\{g_i\}$, is

$$A(j|\{g_i\}) \equiv \frac{N_g(j)}{\nu_g}. \quad (4.9)$$

where $N_g(j)$ is the number of ghosts clustered into j .

This definition of the jet area depends on the specific distribution of ghosts. Since the distribution is generated randomly, this area need not be the same if the procedure is carried out twice. To make this area more robust under the selection of a new set of ghosts, this area is averaged over a large variety of ghost distributions, and the limit of large ghost number density is taken:

$$A(j) \equiv \lim_{\nu_g \rightarrow \infty} \langle A(j|\{g_i\}) \rangle_g. \quad (4.10)$$

This is the *scalar active area*. It is extended to the four-vector active area $A_\mu(j)$ by defining

$$A_\mu(j|\{g_i\}) \equiv \frac{1}{\nu_g \langle g_T \rangle} \sum_i g_{i,\mu}, \quad (4.11)$$

where the sum is over all ghosts within j and $\langle g_T \rangle = \sum_{i \in j} g_{i,T} / N_g(j)$. Note that the transverse component of this area corresponds to the scalar active area.

Given a measure of the jet area, one can define the second parameter required to estimate the pileup contribution to a jet's energy: an energy density of jets, ρ [37], given by

$$\rho \equiv \text{Md}_i \left(\frac{p_{i,T}}{A_{i,T}} \right), \quad (4.12)$$

where Md represents a median over all reconstructed jets i in the event. This definition is motivated by the fact that, in the presence of high pileup, there will be a large number of reconstructed jets in each event that originate from pileup. All such jets have energies much lower than jets originating from the hard-scatter. This separation between the energies of the pileup jets and the hard-scatter jets, along with the large number of pileup jets, ensures that ρ is pulled towards the energy density of the pileup jets. As such, it provides an estimate for the energy density of the pileup jets, without accounting for the energy of the hard-scatter jets. This is exactly the quantity that needs to be subtracted from the reconstructed jet

energy.

The four-momentum of each jet in the event is then corrected by subtracting the portion of pileup contamination estimated to be within that jet's area:

$$p'_{\text{jet}}{}^\mu = p_{\text{jet}}{}^\mu - \rho A_{\text{jet}}{}^\mu. \quad (4.13)$$

Any jet with $p_T < \rho A_T$, or with transverse momentum below a specified threshold (20 GeV is used in this work) is removed from the event and therefore not considered further in the analysis.

Jets remaining after the mitigation of pileup are converted to photons at the rate mentioned in Section 3.3. Converted jets are removed from the list of reconstructed jets, and added to the list of reconstructed photons passing the isolation criterion. These are the jets and photons that are studied in the remainder of the analysis.

With a thorough understanding of how the detector-level samples are produced, the analysis of these simulated data sets can now be discussed.

Chapter 5

Analysis

The discovery of a new particle at a collider experiment requires direct observation of the new particle. This entails a measurement of the mass of the new particle, by measuring the *invariant mass* of its decay products. Specifically, this means that if a previously unobserved particle χ decays to n daughter particles with four-momenta p_i^μ , the invariant mass

$$m_{\text{inv}} \equiv \sqrt{\left(\sum_{i=1}^n p_i^\mu\right) \left(\sum_{j=1}^n p_{j,\mu}\right)} \quad (5.1)$$

is computed for each event, and the distribution of this quantity should have a statistically significant peak at $m_{\text{inv}} = m_\chi$.

With this in mind, the analysis of the generated samples comprises three main steps, with the goal of accentuating a peak in the invariant mass distribution of two jets plus one photon, $m_{jj\gamma}$:

1. apply cuts to the signal and background samples to select events that can be reconstructed by a detector and to discriminate against background,
2. apply a kinematic fit to properly group the final state objects into two two-jet-one-photon groups, to further discriminate against background, and to improve the resolu-

tion of the $m_{jj\gamma}$ distribution. The resulting $m_{jj\gamma}$ resolutions are used to calculate the signal and background efficiencies in a specified mass window,

3. assuming no evidence for a charged Higgs boson is found, limits on $\sigma(pp \rightarrow H^+ H^-) \times BR(H^\pm \rightarrow W^\pm \gamma)^2$ are calculated as a function of the charged Higgs mass.

5.1 Preselection Cuts

The purpose of the preselection cuts is twofold: first they remove all objects that can not be reconstructed in a real experiment for a variety of reasons, and second they are chosen to remove as many background events as possible from the region of interest. Every cut applied is motivated by at least one of these two objectives. Cuts motivated by the first objective are called *detector acceptance cuts*, and those satisfying the second are called *background discrimination cuts*.

5.1.1 Detector Acceptance Cuts

An event must possess kinematics that allow for it to be triggered on — so that the detector knows to record this specific event — and once triggered on the objects in the final state must be compatible with the process of interest. The triggering criterion is satisfied by requiring the presence of two photons with transverse momentum of at least 20 GeV, falling within the fiducial volume of the detector, defined by the pseudorapidity range $|\eta| < 2.5$. All events with less than two photons with these kinematic properties are rejected.

After triggering on an event, the event must be compatible with the process of interest, in the sense that it must have the correct final state. This requires further that there are at least four jets in the final state that can be reconstructed. To be reconstructible, a jet must have sufficiently high energy, and must fall within the fiducial volume of the detector. These conditions are satisfied by requiring that jets have at least 20 GeV of transverse

momentum, and that they fall within $|\eta| < 4.9$. All events with less than four jets satisfying these conditions are rejected.

If an event contains more than four jets with $p_T > 20$ GeV and $|\eta| < 4.9$, the four highest p_T jets are assumed to originate from the charged Higgs bosons. Similarly, if more than two photons possesses $|\eta| < 2.5$ and $p_T > 20$ GeV the two photons with the highest p_T are assumed to originate from the charged Higgs bosons. Only these four jets and two photons are considered for the rest of the analysis.

5.1.2 Background Discrimination Cuts

After selecting events compatible with the signal, cuts are applied to discriminate against the SM background processes. Because of the extremely large cross-section of the QCD background compared to all other backgrounds, cuts are chosen specifically with the goal of eliminating as many QCD background events as possible. Two cuts are found to be particularly successful in eliminating a significant fraction of the QCD background.

Objects in the QCD background have on average lower transverse momenta than those in the signal. This is most pronounced in the transverse momenta of the leading photons, and in the scalar sum of the transverse momenta of all objects in the event, $H_T \equiv \sum |p_T|$, so all events with values of these quantities lower than a specified cut are excluded from the remainder of the analysis. The cut values at each charged Higgs test mass are optimized to produce the most stringent limits on the production cross-section times branching ratio at the end of the analysis.

As the test mass increases the distributions of the signal leading photon p_T and H_T shift toward higher values. This allows the cuts on these quantities to also be increased, further reducing the background contamination of the higher mass samples without compromising much on the signal efficiency. Figure 5.1 shows the distributions of both of these variables for the QCD background, and for four charged Higgs test masses. The trend of the signal distributions to higher values of both variables with increasing charged Higgs mass is clear

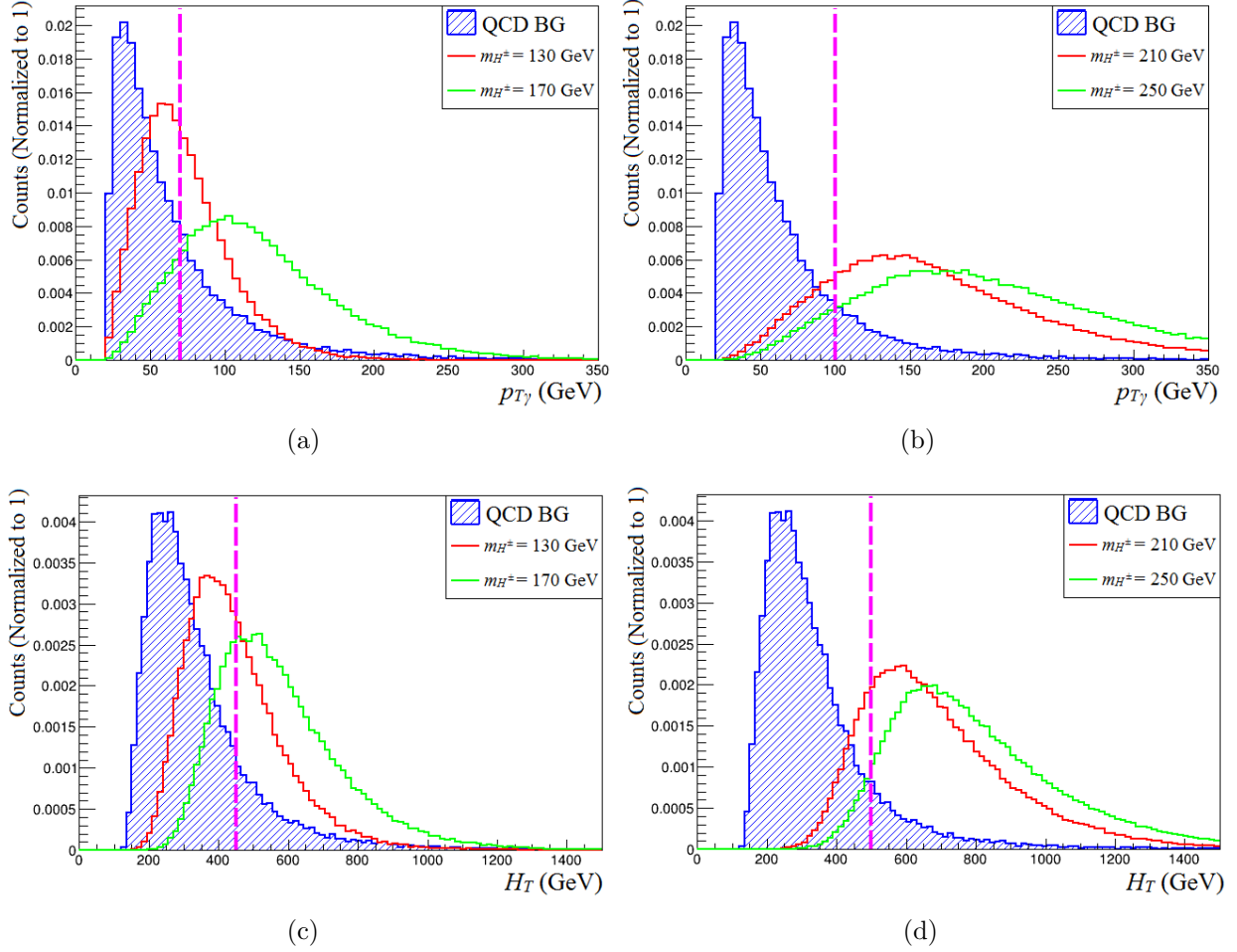


Figure 5.1: The distributions of the leading photon transverse momentum for the QCD background and signal events with masses (a) $m_{H^\pm} = 130$ GeV and 170 GeV, and (b) $m_{H^\pm} = 210$ GeV and 250 GeV. The distributions of the scalar sum of transverse momenta for the QCD background and signal events with masses of (c) $m_{H^\pm} = 130$ GeV and 170 GeV and (d) $m_{H^\pm} = 210$ GeV and 250 GeV. The dashed line on each plot marks the location of the cut used to suppress background.

from this figure. The cuts used for the different charged Higgs test masses are tabulated in Table 5.1.

Once cuts are applied to suppress the background processes, a method was then required to determine how the four jets and two photons being analyzed should be grouped into two-jet-one-photon groups to reconstruct the charged Higgs bosons. Towards this end, a kinematic fit was applied to events passing the preselection cuts.

	$m_{H^\pm}$ range (GeV)				
	[130, 200]	[210, 250]	[300, 350]	[400, 450]	500
Leading photon p_T (GeV)	70	100	150	250	300
H_T (GeV)	450	500	700	850	1100

Table 5.1: Values for the background discrimination cuts on the leading photon p_T , and H_T , for all test masses.

5.2 Kinematic Fitting

As mentioned previously, a measurement of energy or momentum before and after any scattering process will often result in only approximate conservation of these quantities because of uncertainties in the measurements due to finite detector resolutions. Measured quantities can be altered within these uncertainties, with the aim of imposing conservation laws, by applying a *kinematic fit* on the data. It is expected that, by forcing the satisfaction of conservation laws taking into account experimental resolutions, measured quantities become more accurate.

In practice, a kinematic fit involves defining a χ^2 function that is minimized subject to the imposed physical constraints, through the method of Lagrange multipliers [38]. For the purpose of this work, the χ^2 function minimized is

$$\chi^2 = \sum_i \left[\frac{(p_{T,i} - p_{T,i}^{\text{meas}})^2}{\sigma_{p_T,i}^2} + \frac{(\eta_i - \eta_i^{\text{meas}})^2}{\sigma_{\eta,i}^2} + \frac{(\phi_i - \phi_i^{\text{meas}})^2}{\sigma_{\phi,i}^2} \right], \quad (5.2)$$

subject to the constraints

$$\begin{aligned} 0 &= \sqrt{(j_1 + j_2)_\mu (j_1 + j_2)^\mu} - m_W, \\ 0 &= \sqrt{(j_3 + j_4)_\mu (j_3 + j_4)^\mu} - m_W, \\ 0 &= \sqrt{(j_1 + j_2 + \gamma_1)_\mu (j_1 + j_2 + \gamma_1)^\mu} - \sqrt{(j_3 + j_4 + \gamma_2)_\mu (j_3 + j_4 + \gamma_2)^\mu}, \end{aligned} \quad (5.3)$$

where the sum in Equation 5.2 is over the 4 jets and 2 photons considered in the analysis, x_i^{meas} are the measured kinematic quantities, x_i are the variables that are modified to min-

imize the χ^2 , and $\sigma_{x,i}$ are their detector resolutions. The detector resolutions are measured by comparing the truth-level and reconstructed object four-momenta, and are found to depend on both energy and pseudorapidity. Details of how they are obtained are provided in Appendix A. The first two constraints in Equation 5.3 reflect the expectation that the jets come from the decay of two W bosons, while the third requires that both charged scalars that are pair-produced have the same invariant mass.

If it is possible to find four-vectors that minimize Equation 5.2 while satisfying Equation 5.3, then the minimization converges. Intuitively, a smaller χ^2 suggests a better fit since the measured quantities did not have to be altered much to satisfy the constraints. Formally, the quality of the convergent solution can be quantified using the *p-value* of the χ^2 , defined as [39]

$$p(\chi^2; n_{\text{dof}}) = \int_{\chi^2}^1 f_{\chi^2}(x; n_{\text{dof}}) dx, \quad (5.4)$$

where $f_{\chi^2}(x; n_{\text{dof}})$ is the χ^2 probability distribution function given n_{dof} degrees of freedom. The function minimized has 3 degrees of freedom: one per Lagrange multiplier. The *p-value* represents the probability that the minimized χ^2 value would be above the measured value purely by statistical fluctuations. A large *p-value* therefore suggests a better fit.

A priori it is unknown which photons/jets came from which charged Higgs. In other words, the assignment of jets j_i and photons γ_k , for $i = 1, 2, 3, 4$ and $k = 1, 2$, is unknown. Therefore the fit is performed with all viable assignments of four jets and two photons into two 2-jet-1-photon groups; there are at most six viable assignments per event. The viability of a grouping is assessed based on the proximity of the objects that are to be grouped together: if two jets come from the same W boson then they should be close to one another in ΔR . Similarly, if two jets and one photon come from the same charged Higgs they should be close according to the same metric.

The maximum allowed ΔR between jets and photons in a viable grouping depends on

	$m_{H^\pm}$ range (GeV)			
	[130, 200]	[210, 350]	[400, 450]	500
$\Delta R(jj)$	2.8	1.5	1.4	1.2
$\Delta R(j\gamma)$	1.95	2.5	2.7	3

Table 5.2: Maximum values of ΔR allows between jets and photon in a viable grouping, for all test masses.

the test mass. As the charged Higgs get heavier they are produced with less kinetic energy. Consequently the photons and W bosons are less boosted, and therefore the jets get further from the photons in $\eta - \phi$ space with increasing $m_{H^\pm}$. Furthermore, a larger $m_{H^\pm}$ means there is more kinetic energy provided to the photon and the W boson. The increased kinetic energy of the W boson produces more collimated jets, and hence the ΔR between the jets decreases. Figure 5.2 shows the evolution of the ΔR distributions with $m_{H^\pm}$. The maximum object ΔR separations allowed for a viable grouping for the different charged Higgs masses are tabulated in Table 5.2. These values are optimized to produce the most stringent limits on the production cross-section times branching ratio at the end of the analysis.

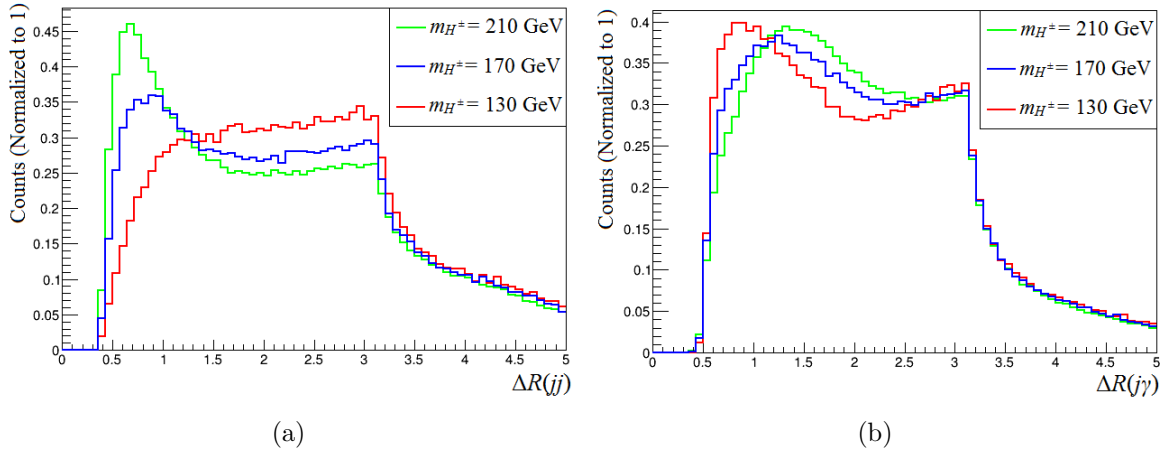


Figure 5.2: The ΔR separation between jets and photons in the same grouping, after choosing the grouping using the kinematic fit with the highest p -value. As explained in the text, (a) shows that jets from the same W boson are produced with decreasing ΔR as $m_{H^\pm}$ increases, whereas (b) shows that jets and photons from the same charged Higgs are produced with increasing ΔR as $m_{H^\pm}$ increases.

It is possible that no groupings in an event produce a convergent fit, in which case it is

Cut	QCD Background	Signal	
		$m_{H^\pm} = 130 \text{ GeV}$	$m_{H^\pm} = 170 \text{ GeV}$
Detector Acceptance	6.17(3)%	12.8(1)%	27.0(1)%
Background Discrimination	0.77(1)%	3.47(5)%	17.8(1)%
Kinematic Fit	0.199(6)%	2.56(4)%	12.5(1)%

Cut	QCD Background	Signal	
		$m_{H^\pm} = 300 \text{ GeV}$	$m_{H^\pm} = 350 \text{ GeV}$
Detector Acceptance	6.17(3)%	54.34(7)%	58.04(7)%
Background Discrimination	0.193(4)%	38.72(7)%	43.15(7)%
Kinematic Fit	0.039(2)%	27.53(6)%	29.93(7)%

Table 5.3: The effects of the cuts used for $m_{H^\pm} \leq 200 \text{ GeV}$ (top) and for $300 \text{ GeV} \leq m_{H^\pm} < 400 \text{ GeV}$ (bottom) on the absolute background and signal efficiencies. The errors are statistical and are calculated using equation 5.14.

assumed that the charged scalars did not pass the preselection cuts and the event is rejected. If more than one grouping produces a convergent fit, it is assumed that the grouping with the largest p -value is the correct one. By truth-matching (see Appendix A.1) the jets and photons used in the kinematic fit, it is found that the correct grouping was selected by this method 50% of the time. A two-jet-one-photon invariant mass distribution is constructed from the groupings that maximize the p -value.

The effects of all the cuts and the kinematic fit on the efficiencies of the QCD background and some signal samples are tabulated in Table 5.3. The background efficiencies for the QCD background in Table 5.3 can be used to find that the effective cross-section of the QCD background remains up to almost 40 times larger than the total cross-section of the subdominant background (Table 3.1). Therefore all SM backgrounds are negligible in comparison to the QCD background, and are ignored.

5.2.1 $m_{jj\gamma}$ Distribution

The $m_{jj\gamma}$ distribution serves two purposes in this analysis. The first is to validate that the cuts and the kinematic fit does not affect the signal in a manner that obscures the peak at

the test mass; if the analysis is successful and compatible with the process of interest, there should be a clearly defined peak at the test mass.

By examining this distribution, it is determined that the analysis developed is effective in reconstructing a peak in the $m_{jj\gamma}$ distribution at the charged Higgs mass for test masses at or above 130 GeV. Figure 5.3 shows an example of the result of the cuts and the kinematic fit on the $m_{H^\pm} = 200$ GeV sample.

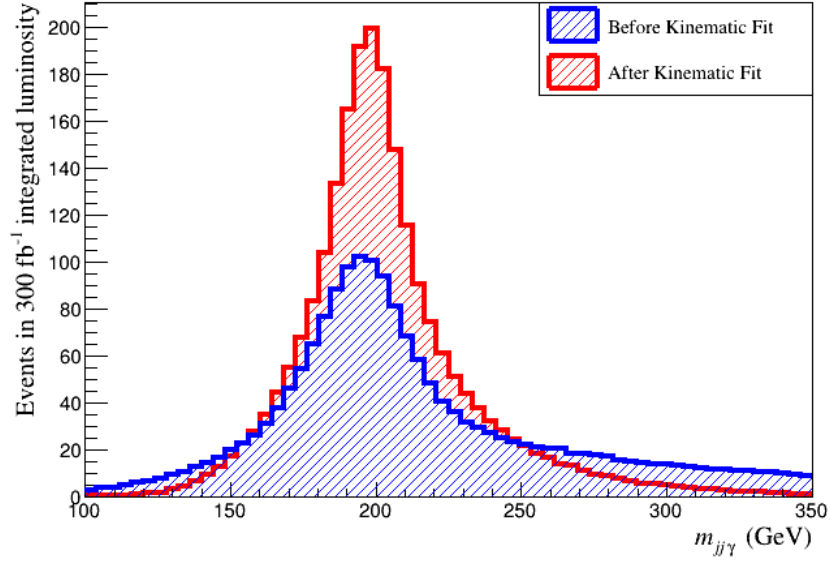


Figure 5.3: The effect of the kinematic fit on the $m_{jj\gamma}$ distribution for the $m_{H^\pm} = 200$ GeV sample. The distribution is scaled to the upper limit on the production cross-section times branching ratio set for this test mass (see figure 6.1).

Below $m_{H^\pm} = 130$ GeV, however, it is found that this technique could not produce an accurate peak at the test mass value. The reason for this is a bias produced by the combination of the fit and the detector acceptance cuts, observed as few charged Higgs candidates having an invariant mass below 100 GeV. This observed cutoff in the $m_{jj\gamma}$ distributions is understood as follows: the detector acceptance cuts require photons with at least 20 GeV of transverse momentum, while the kinematic fit forces the selected pairings of jets to have an invariant mass of 80 GeV. Both of these requirements set lower limits on the energy of the final state, so to output this amount of energy a decaying charged Higgs must have at least 100 GeV of energy. The only way to guarantee this amount of energy is

available for the decay is for the charged Higgs to possess this energy in its invariant mass.

This sharp drop of the $m_{jj\gamma}$ distribution around 100 GeV resulted in incorrect locations of the peaks of the distributions for $m_{H^\pm} < 130$. This effect is portrayed in Figure 5.4. To avoid this reconstruction bias, the analysis is restricted to test masses $m_{H^\pm} \geq 130$ GeV.

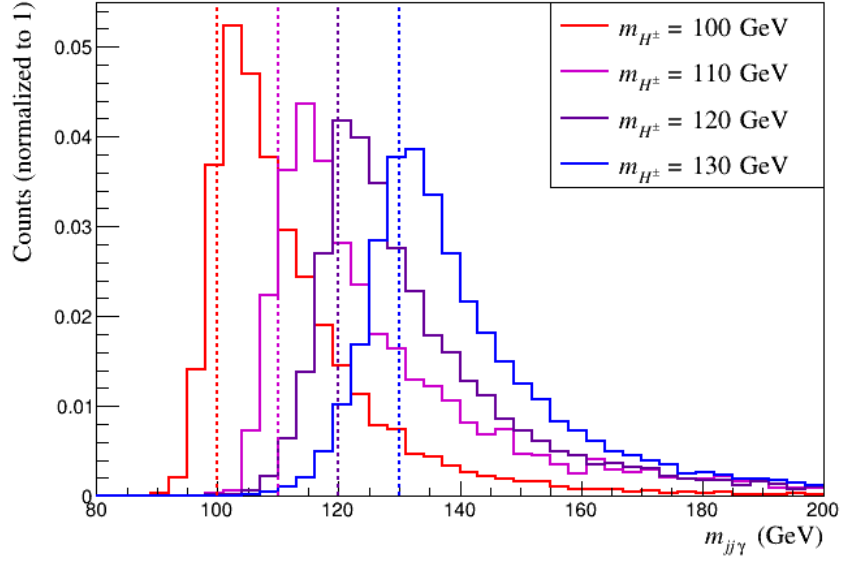


Figure 5.4: The invariant mass distributions for signal samples with test mass $m_{H^\pm} \leq 130$ GeV. The dashed lines show the locations of the test masses, and are to be compared to the locations of the reconstructed peaks of the distributions in the same color.

The second purpose served by the $m_{jj\gamma}$ distribution is to define one final cut that aided in setting more stringent limits on $\sigma \times BR$. As will be shown in Section 5.3, the limit depends on both the signal and background efficiencies.

To obtain the most stringent value for the limit, it is key to minimize the background efficiency while maintaining as high of a signal efficiency as possible. As shown in Figure 5.3, it is clear that the reconstructed signal events are concentrated around the test mass. Figure 5.5 shows that the background remains widely distributed, even after the kinematic fit. One should not be inclined to count background events towards the final efficiency if they lie past a certain distance from the peak, since they do not make the reconstructed signal peak more difficult to discern. The $m_{jj\gamma}$ distribution allowed the quantification of this distance.

This distance varied with the test mass, and is given by the resolution of the $m_{jj\gamma}$ peak.

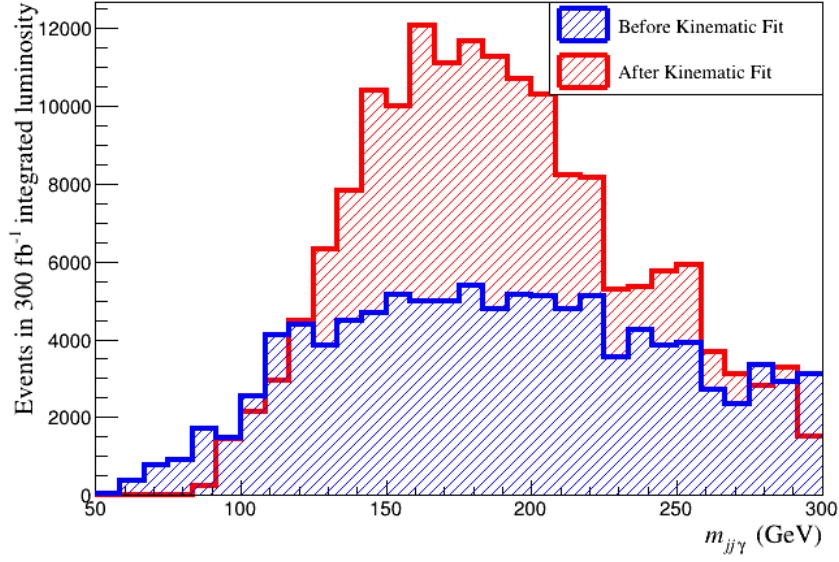


Figure 5.5: The effect of the kinematic fit on the $m_{jj\gamma}$ distribution for the QCD background, applying the set of cuts for test masses below 200 GeV.

With the goal of determining this resolution the reconstructed $m_{jj\gamma}$ distribution is fit with the sum of a crystal ball and Gaussian (CBG). The functional form of the crystal ball function is [40]

$$C(x; \beta, n, \mu, \sigma) = N \begin{cases} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right], & \text{for } \frac{x-\mu}{\sigma} > -\beta \\ \left(\frac{n}{|\beta|} \right)^n \exp \left[-\frac{|\beta|^2}{2} \right] \left(\frac{n}{|\beta|} - |\beta| - \frac{x-\mu}{\sigma} \right)^{-n}, & \text{for } \frac{x-\mu}{\sigma} \leq -\beta, \end{cases} \quad (5.5)$$

where N is a normalization constant, μ is the mean of the Gaussian core, σ is the standard deviation of the Gaussian core, β denotes the location relative to the mean where the Gaussian ends and the power-law tail begins (in units of σ), and n is the order of the power-law tail. The sign of β indicates the side of the Gaussian core on which the exponential power-law tail is stitched: positive β means the tail is on the left of the Gaussian, and negative β means it is on the right. An example of this fit for the $m_{H^\pm} = 200$ GeV sample is shown in Figure 5.6.

Both the CB and the standalone Gaussian in the fit provided an independent estimate for the width of the mass peak. The resolution is taken to be the weighted average of the

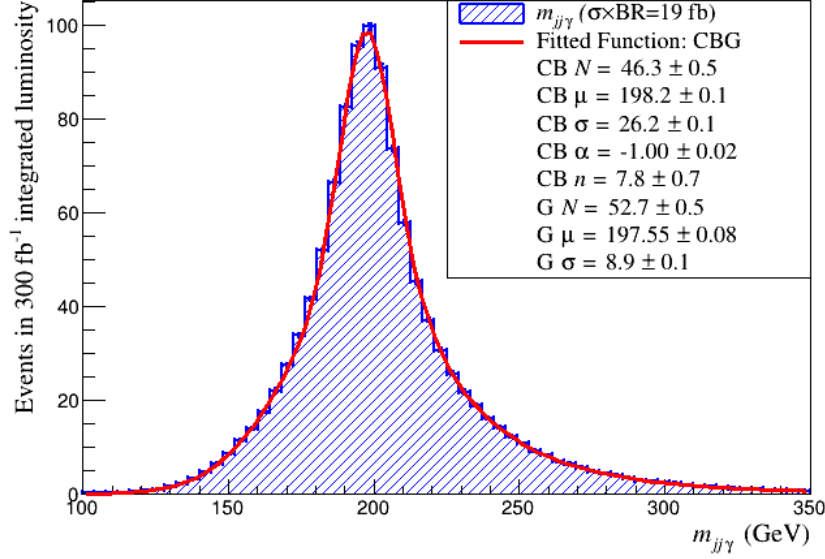


Figure 5.6: The $m_{H^\pm} = 200$ GeV test mass $m_{jj\gamma}$ distribution fit with the CBG function. The distribution is scaled to the upper limit on the production cross-section times branching ratio set for this test mass (see Figure 6.1).

widths, with the weight for each width being the fraction of the area under the Gaussian peak covered by the corresponding fit. Explicitly,

$$\sigma_{m_{jj\gamma}} = \frac{A_{CB}\sigma_{CB} + A_G\sigma_G}{A_{CB} + A_G}, \quad (5.6)$$

where σ_{CB} is the standard deviations of the Gaussian core of the CB function, σ_G is the standard deviation of the standalone Gaussian, A_{CB} is the area of the Gaussian core of the CB function, and A_G is the area of the standalone Gaussian that is within the Gaussian core of the CB function. The signal and background efficiencies used in the limit calculations for a specific test mass correspond to the fraction of generated events within the mass window $m_{H^\pm} - \sigma_{m_{jj\gamma}} < m_{jj\gamma} < m_{H^\pm} + \sigma_{m_{jj\gamma}}$. The reconstructed invariant mass resolutions of signal samples are plotted as function of test mass in Figure 5.7, and the final signal and background efficiencies in the mass windows associated to each charged Higgs test mass are tabulated in Table 5.4. The trend of increasing invariant mass resolutions with the charged Higgs test mass in Figure 5.7 is to be expected, since the absolute resolutions on the jet and photon

energies increase with increasing energy (see Section A.2).

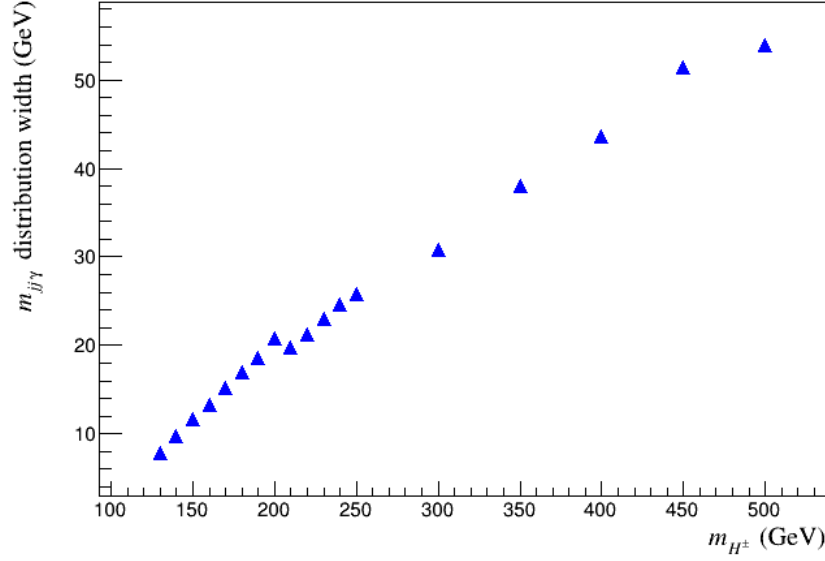


Figure 5.7: Reconstructed $m_{jj\gamma}$ resolutions as a function of the charged Higgs test mass. The errors on the widths are all less than 0.2 GeV, and are not visible on this plot. Discontinuities in the trend observed can be attributed to the changing values of background discrimination and kinematic fitting ΔR cuts.

	$m_{H^\pm}$ (GeV)								
	130	140	150	160	170	180	190	200	210
ϵ_s (%)	1.17(1)	2.42(1)	4.00(2)	5.72(2)	7.54(2)	9.27(3)	10.89(3)	12.55(3)	11.44(3)
$100\epsilon_b$ (%)	1.19(9)	2.0(1)	2.9(2)	3.8(2)	4.5(2)	4.9(2)	5.1(2)	5.2(2)	4.3(2)

	$m_{H^\pm}$ (GeV)								
	220	230	240	250	300	350	400	450	500
	13.27(3)	15.18(3)	16.68(3)	17.98(4)	17.33(3)	22.23(4)	18.30(4)	21.72(4)	17.43(3)
	4.1(2)	3.8(2)	3.6(2)	3.2(2)	1.6(1)	0.85(8)	0.46(6)	0.32(5)	0.17(3)

Table 5.4: Signal and background efficiencies within the mass windows of each test mass. These efficiencies are used to compute the limits on $\sigma(pp \rightarrow H^+H^-) \times BR(H^\pm \rightarrow W^\pm\gamma)^2$. The errors shown are statistical and are calculated using equation 5.14.

Figure 5.8 shows the sum of the $m_{jj\gamma}$ distributions for the QCD background and the $m_{H^\pm} = 200$ GeV test mass, as well as the sum of the $m_{jj\gamma}$ distributions for the QCD background and the $m_{H^\pm} = 300$ GeV test mass, in 300 fb^{-1} of integrated luminosity. This is what would be observed in a collider experiment in the presence of a 200 GeV or 300 GeV

charged Higgs boson.

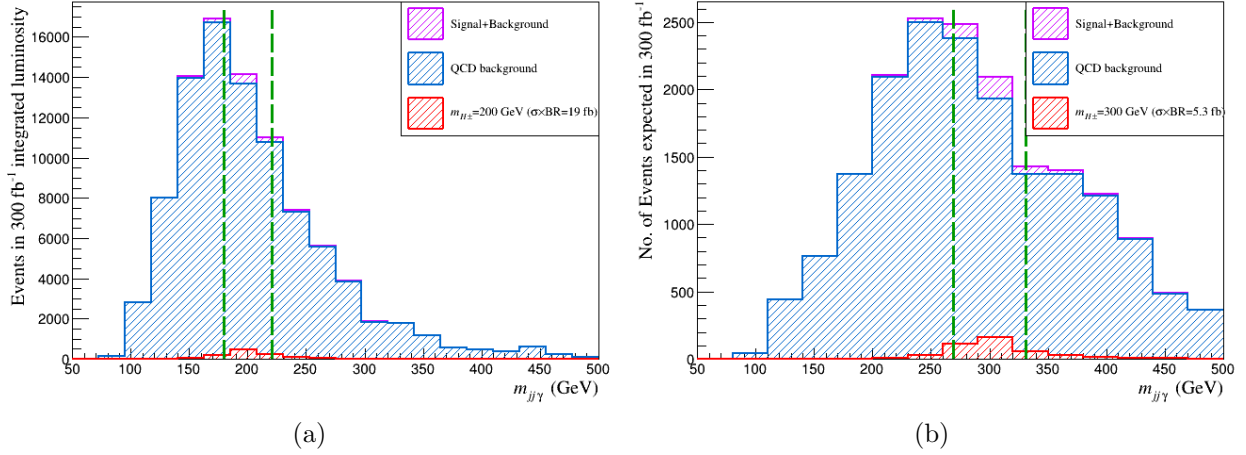


Figure 5.8: The $m_{jj\gamma}$ distribution for the sum of the QCD background and the (a) $m_{H^\pm} = 200$ GeV and (b) $m_{H^\pm} = 300$ GeV test mass in 300 fb^{-1} of integrated luminosity. The signal distributions are scaled to the upper limits on the production cross-section times branching ratio set for these test masses (see Figure 6.1). The green lines delimit the mass window for each test mass. Note that the difference in the QCD background shapes is due to the fact that different cut values are used for each mass depicted.

5.3 Limit Setting

While this method is effective in reconstructing charged Higgs bosons with masses $m_{H^\pm} \geq 130$ GeV, it is not guaranteed that applying this analysis to data would produce an observation of a charged Higgs boson, even if one exists that is consistent with the one studied herein. It is possible that such a particle is produced so rarely that an excess in the $m_{jj\gamma}$ distribution above the QCD background at the test mass is not observable. The question then arises of exactly how rare this process must be to go undetected. This rarity is expressed in terms of an upper limit on $\sigma(pp \rightarrow H^+H^-) \times BR(H^\pm \rightarrow W^\pm\gamma)^2$, representing the largest value possible for this quantity that would result in non-detection.

Setting limits on $\sigma(pp \rightarrow H^+H^-) \times BR(H \rightarrow W\gamma)^2$ is done using a Bayesian method. Following the derivation in [41], the probability of observing n events consistent with the signal process given an integrated luminosity $\mathcal{L}$, a signal efficiency ϵ_s , a signal cross-section

σ_{sig} , and a number b of background events obeys a Poisson distribution:

$$P(n|\sigma_{\text{sig}}, \epsilon_s, b) = \frac{\mu^n e^{-\mu}}{n!}, \quad \mu = \mathcal{L}\sigma_{\text{sig}}\epsilon_s + b. \quad (5.7)$$

From Bayes' rule [39], then, the probability that the true cross-section is σ_{sig} given n , ϵ , and b is

$$P(\sigma_{\text{sig}}|n, \epsilon_s, b) = \frac{P(n|\sigma_{\text{sig}}, \epsilon_s, b)P(\sigma_{\text{sig}})}{\int_0^\infty P(n|\sigma'_{\text{sig}}, \epsilon_s, b)P(\sigma'_{\text{sig}})d\sigma'_{\text{sig}}}, \quad (5.8)$$

where $P(\sigma_{\text{sig}})$ is the *prior* probability for the cross-section to take value σ_{sig} and is taken to be uniform up to some finite cross-section that is certainly above the bound to be set, and 0 above this cross-section. The prior used is

$$P(\sigma_{\text{sig}}) = \begin{cases} 10^{-3} \text{ fb}^{-1}, & \sigma_{\text{sig}} < 10^3 \text{ fb} \\ 0, & \sigma_{\text{sig}} > 10^3 \text{ fb}. \end{cases}$$

In the context of calculating an expected limit, it must be assumed that all events observed that are consistent with signal are in fact background. This implies that $n = b$. The α confidence level upper bound is then the value of the cross section σ_α that satisfies

$$\int_0^{\sigma_\alpha} P(\sigma_{\text{sig}}|\epsilon_s, b)d\sigma_{\text{sig}} = \alpha. \quad (5.9)$$

Note that since the process under study has the additional constraint that the W bosons decay hadronically, σ_{sig} is given by

$$\sigma_{\text{sig}} = \sigma(pp \rightarrow H^+H^-) \times BR(H \rightarrow W\gamma)^2 \times BR(W \rightarrow \text{hadrons})^2. \quad (5.10)$$

Therefore the α confidence level upper limit on $\sigma(pp \rightarrow H^+H^-) \times BR(H \rightarrow W\gamma)^2$ is given by $\sigma_\alpha/BR(W \rightarrow \text{hadrons})^2$.

Systematic and statistical uncertainties on the parameters are included in this limit calculation by convoluting equation 5.7 with the probability density function of the parameters. For example, suppose the parameter ϵ_s is distributed according to some probability density function $f(\epsilon'_s; \epsilon_s, \vec{\theta})$, where ϵ_s is the measured value and $\vec{\theta}$ represents all other parameters of the probability density function. Then $P(n|\sigma_{\text{sig}}, \epsilon_s, b)$ in equation 5.8 is replaced by

$$P(n|\sigma_{\text{sig}}, \epsilon_s, b) \rightarrow \int_0^1 P(n|\sigma_{\text{sig}}, \epsilon'_s, b) f(\epsilon'_s; \epsilon_s, \vec{\theta}) d\epsilon'_s. \quad (5.11)$$

In practice, b is measured through the general relation $b = \mathcal{L} \sum_i \sigma_{\text{bg},i} \epsilon_{\text{bg},i}$ where the sum is over all SM backgrounds considered. As all SM backgrounds are negligible compared to the QCD background, in this work $i = 1$. Then, $b = \mathcal{L} \sigma_{\text{bg}} \epsilon_{\text{bg}}$ where σ_{bg} and ϵ_{bg} are the cross-section and efficiency of the QCD background respectively. The three parameters that are input to the limit calculation are therefore σ_{bg} , ϵ_s , and ϵ_{bg} .

5.4 Uncertainties

Uncertainties on the three parameters that enter the limit calculation due to several sources are estimated: PDF and scale uncertainties on efficiencies and cross-sections, integration uncertainty in the calculation of the background cross-section, and statistical uncertainty on the efficiencies.

5.4.1 PDF and Scale Uncertainties

PDF and scale uncertainties arise due to an inability to fully describe the interactions under study at particle colliders. To address this shortcoming scales are introduced that are meant to parameterize the energy regime where perturbation theory provides a sufficient description, and where empirical results are needed to fill in the gaps that are inaccessible to the theory.

The fact that PDF sets themselves have errors on their parameters results in an uncertainty, called the *PDF uncertainty*, on the observables affected by the PDF. In NNPDF the uncertainties in the parameters of the PDF are encoded in 100 different *member sets*, each of which has slightly different values for these parameters. The *nominal* PDF set is the average of all member sets. It is this nominal PDF set that is used to simulate samples and estimate cross-sections.

By contrast to PDF uncertainties which arise from experimental uncertainties, *scale uncertainties* are meant to estimate the shortcomings of calculations performed in perturbation theory. Often when one calculates scattering amplitudes from Feynman diagrams within the SM at loop level, high energy (ultraviolet, UV) divergences appear. A common method used to make these divergences finite is to introduce a cutoff, Λ on the highest allowed energy in the theory. This cutoff represents the energy scale where the SM no longer provides an accurate description of the scattering, perhaps because new degrees of freedom become manifest. Instead of working directly with this cutoff, it is customary to introduce the *renormalization scale*, μ_R , and absorb the dependence of scattering amplitudes on this scale into the coupling constant of the theory [8].

A second scale, the *factorization scale* μ_F , similarly delimits an energy scale where perturbation theory is inadequate, only it is introduced to make finite IR divergences. Examples of such divergences occur in the aforementioned emission of low energy gluons. Instead of absorbing dependence on this scale into the coupling constant, it is absorbed into the PDF.

Both of these scales are entirely unphysical: if it was possible to compute scattering amplitudes at all orders in perturbation theory, or to simply compute the entire amplitude, all dependence on these scales would vanish. Since this is not possible, however, values must be chosen for these scales. Due to the fact that the selection of these scales is arbitrary, one must understand how sensitive simulated results are to these scales. Conventionally, this sensitivity is assessed by performing all calculations with nominal scales (this work used $\mu_R = \mu_F = m_Z$) and repeating the calculations while varying each scale independently up

and down by factors of two. This produces eight variations from the nominal scale choice: $\{2\mu_R, \mu_F\}$, $\{\mu_R/2, \mu_F\}$, $\{2\mu_R, 2\mu_F\}$, $\{\mu_R/2, 2\mu_F\}$, $\{2\mu_R, \mu_F/2\}$, $\{\mu_R/2, \mu_F/2\}$, $\{\mu_R, 2\mu_F\}$, $\{\mu_R, \mu_F/2\}$. The variation of simulated or calculated quantities with the choice of scale is what is called the scale uncertainty.

Both of these uncertainties affected the background cross-section estimate, and the signal and background efficiencies.

To estimate the effects of PDF and scale uncertainties on the QCD background cross-section estimate, the cross-section is calculated using **Sherpa** once with each PDF member set with the nominal scales, and with the eight scale variations using the nominal PDF set. Then the standard deviation of the cross-section is calculated using

$$\sigma_{\sigma_b, \text{PDF}} = \sqrt{\frac{\sum_{i=1}^{N_{\text{PDF}}} (\sigma_{b,i} - \sigma_{b,0})^2}{N_{\text{PDF}}}}, \quad \sigma_{\sigma_b, \text{scale}} = \sqrt{\frac{\sum_{i=1}^{N_{\text{scale}}} (\sigma_{b,i} - \sigma_{b,0})^2}{N_{\text{scale}}}}, \quad (5.12)$$

where $N_{\text{PDF}} = 100$, $N_{\text{scale}} = 8$, and $\sigma_{b,0}$ is the nominal cross-section. It is found that $\sigma_{\sigma_b, \text{PDF}} = 4$ pb and $\sigma_{\sigma_b, \text{scale}} = 50$ pb, representing relative errors on the QCD background cross-section of 2.5% and 31% respectively.

This large difference between the scale and PDF uncertainties on the QCD background cross-section is expected, because of the fact that this background is produced via the strong force. As mentioned above, when one chooses values for the renormalization and factorization scales, one is indirectly choosing energy cutoffs as well, which curtails the Feynman diagram expansion. However, the coupling constant of the strong force is of order unity [8], and consequently QCD scattering amplitudes receive significant corrections from Feynman diagrams at high orders in perturbation theory. Cutting the Feynman diagram expansion short eliminates these non-negligible higher order corrections, resulting in large uncertainties on the QCD cross-section. The PDF, by contrast, does not suffer from the same lack of information as a reduced Feynman diagram expansion. The uncertainties on its parameters are experimentally determined, and the differences between variations of a PDF set are thus

small relative to the scale uncertainty.

To determine the effects of the PDF and scale variations on the signal and background efficiencies, the analysis is carried through to the end and the distributions of $m_{jj\gamma}$ are made for each of the $N_{\text{PDF}} + N_{\text{scale}}$ variations. Efficiencies are then calculated as the ratio of the sum of the weights of all events within the mass window to the sum of the weights of all events corresponding to a specific variation. Uncertainties are estimated using the standard deviations of the efficiency:

$$\sigma_{\epsilon, \text{PDF}} = \sqrt{\frac{\sum_i^{N_{\text{PDF}}} (\epsilon_i - \epsilon_0)^2}{N_{\text{PDF}}}}, \quad \sigma_{\epsilon, \text{scale}} = \sqrt{\frac{\sum_i^{N_{\text{scale}}} (\epsilon_i - \epsilon_0)^2}{N_{\text{scale}}}}, \quad (5.13)$$

where ϵ_0 is the nominal efficiency. Tables 5.5 and 5.6 tabulate the relative scale and PDF uncertainties on the signal and background efficiencies respectively.

		$m_{H^\pm}$ (GeV)							
		130	140	150	160	170	180	190	200
σ_ϵ/ϵ	scale	0.004	0.003	0.002	0.002	0.001	0.0009	0.0008	0.001
	PDF	0.006	0.005	0.005	0.005	0.005	0.005	0.004	0.005

$m_{H^\pm}$ (GeV)									
210	220	230	240	250	300	350	400	450	500
0.001	0.0005	0.0003	0.0002	0.0003	0.002	0.0007	0.002	0.0009	0.002
0.005	0.004	0.004	0.004	0.004	0.005	0.004	0.005	0.004	0.006

Table 5.5: The relative scale and PDF uncertainties on the signal efficiencies.

5.4.2 Integration Errors on Cross-section Estimates

The integration error on the QCD background estimate is calculated directly by **Sherpa**. It depends largely on the number of events generated, so a large enough sample is created to reduce the relative error to 1%.

		$m_{H^\pm}$ (GeV)							
		130	140	150	160	170	180	190	200
σ_ϵ/ϵ	scale	0.1	0.1	0.07	0.04	0.03	0.03	0.03	0.04
	PDF	0.09	0.1	0.05	0.02	0.02	0.02	0.02	0.03

$m_{H^\pm}$ (GeV)									
210	220	230	240	250	300	350	400	450	500
0.04	0.04	0.04	0.05	0.03	0.05	0.05	0.07	0.2	0.06
0.02	0.03	0.03	0.05	0.02	0.04	0.03	0.05	0.2	0.04

Table 5.6: The relative scale and PDF uncertainties on the QCD background efficiencies obtained for analysis cuts corresponding to the different test masses.

		$m_{H^\pm}$ (GeV)							
		130	140	150	160	170	180	190	200
σ_ϵ/ϵ		0.009	0.006	0.004	0.004	0.003	0.003	0.003	0.002

$m_{H^\pm}$ (GeV)									
210	220	230	240	250	300	350	400	450	500
0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002

Table 5.7: The relative statistical errors on the signal efficiencies.

5.4.3 Statistical Errors on Efficiencies

When applying a set of cuts to an event, one of two things can happen: the event will either pass the set of cuts, or it won't. The probability with which the event passes the set of cuts is the efficiency ϵ . Given a set of N events, then, the probability that k of them pass a set of cuts is binomially distributed with probability ϵ . Assuming that k is sufficiently far from both 0 and N (so that the 68% confidence interval does not include unphysical values of k), the error on k can be used to determine the error on ϵ [42]:

$$\sigma_\epsilon = \sqrt{\frac{\epsilon(1-\epsilon)}{N}}. \quad (5.14)$$

The relative statistical uncertainties on the efficiencies according to this formula are tabulated in Table 5.7 for the signal samples, and Table 5.8 for the QCD background.

		$m_{H^\pm}$ (GeV)							
		130	140	150	160	170	180	190	200
σ_ϵ/ϵ		0.09	0.06	0.05	0.05	0.04	0.04	0.05	0.05

$m_{H^\pm}$ (GeV)										
210	220	230	240	250	300	350	400	450	500	
0.04	0.04	0.05	0.05	0.05	0.07	0.1	0.1	0.2	0.2	

Table 5.8: The relative statistical errors on the QCD background efficiencies obtained for analysis cuts corresponding to the different test masses.

The scale uncertainty on the background cross-section, and the scale, PDF, and statistical uncertainties on the background efficiencies are considered in the calculation of the limits. The parameters are assumed to vary in a Gaussian manner with respect to all sources of uncertainty. Thus, incorporating these uncertainties, the 95% confidence limits are calculated by substituting

$$\begin{aligned}
P(n|\sigma_{\text{sig}}, \epsilon_s, b) &\rightarrow \int_0^1 \int_0^\infty d\sigma_b d\epsilon_b P(n|\sigma_{\text{sig}}, \epsilon_s, \mathcal{L}\sigma_b \epsilon_b) \\
&\quad \times G(\sigma_b; \sigma_{b,0}, \sigma_{\text{scale}}) G(\epsilon_b; \epsilon_{b,0}, \sigma_{\text{PDF}}) G(\epsilon_b; \epsilon_{b,0}, \sigma_{\text{scale}}) G(\epsilon_b; \epsilon_{b,0}, \sigma_{\text{stat}})
\end{aligned}
\tag{5.15}$$

in Equation 5.8 (where $G(x; x_0, \sigma_x)$ represents a Gaussian probability density function with mean x_0 and variance σ_x) and solving equation 5.9 for σ_α with $\alpha = 0.95$.

Chapter 6

Results

The effects of the analysis detailed herein on signal and background efficiencies can be interpreted both in model-independent (Section 6.1) and dependent (Section 6.2) ways to quantify the sensitivity of a search for the signal process.

6.1 Limits on Production Cross-Section

Using the method described above, expected 95% confidence limits are calculated for σ_{sig} for 18 scalar masses in the range 130 – 500 GeV. Using $BR(W \rightarrow qq') = 0.6741 \pm 0.0027$ [19] this is converted to an upper limit on $\sigma(pp \rightarrow H^+H^-) \times BR(H \rightarrow W\gamma)^2$. The results are shown in Figure 6.1.

To demonstrate the effects of the different sources of uncertainty on the limits, Figure 6.2 shows the difference between the limits with only the i th uncertainty included in the calculation (μ_i) and those with no uncertainties included in the calculation (μ_0). The scale uncertainty on the background cross-section is found to have the most significant effect on the calculation of the limits. Indeed, the inclusion of all other sources had a negligible effect on the limits once the cross-section scale uncertainty is accounted for. The trends in each of the three lines near the bottom of the plot roughly follow the values of the relative

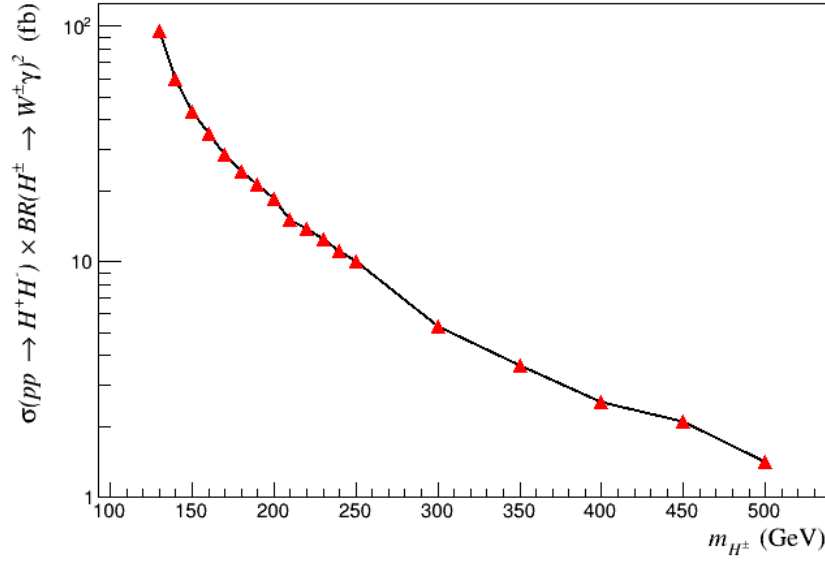


Figure 6.1: The expected 95% confidence upper limits on production cross-section times branching-ratio squared as a function of the mass of the charged scalar. The region above the markers is excluded with 95% confidence.

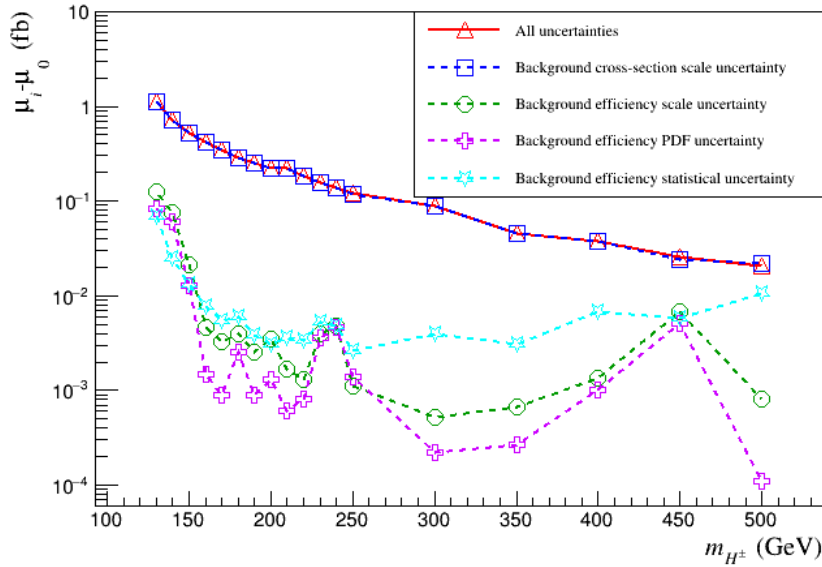


Figure 6.2: The combined and individual effects of each source of uncertainty on the calculation of the upper limit.

uncertainties they represent.

These limits are an order of magnitude lower than the model independent limits set in [16] in a similar channel, and are two orders of magnitude lower than the limits set in [17]. This competitiveness with previously set limits suggests that this method may be able to

set some of the most stringent limits on a charged Higgs boson production and decay mode to date.

6.2 Discovery Potential in the GM Model

To quantify the potential of this analysis to discover a charged Higgs boson, one must specify a complete model. This is due to the need for a cross-section estimate for the signal process, in order to make a prediction about the number of signal events to expect in real data. To this end, predictions about the value of $\sigma(pp \rightarrow H^+ H^-) \times BR(H^\pm \rightarrow W^\pm \gamma)^2$ are extracted from the GM model, for the singly charged Higgs in the scalar quintuplet, $H_5^\pm$.

First, **GM CALC** [43] is used to find experimentally and theoretically allowed values of the GM model parameters that are consistent with the test masses studied; a consistent set of parameters is one that produces a mass of the $H_5^\pm$ within 0.5 GeV of one of the test masses. This set of parameters is then used to make a UFO file so that $\sigma(pp \rightarrow H_5^+ H_5^-)$ could be computed using **MadGraph**. Table 6.1 shows the model parameter values that are manually chosen to produce each mass point, using the parameter symbols defined in [43]. The values of all other parameters can be found using the relations in [43].

	$m_{H^\pm}$ (GeV)							
	130	140	150	160	170	180	190	200
λ_2	1.27	-0.114	0.774	1.243	0.08608	0.175	1.26	0.82
λ_3	-0.498	-1.431	-0.524	-0.15	-0.796	-1.434	-0.519	-0.25
λ_4	0.638	1.464	0.639	0.403	1.01	1.449	0.652	0.6
λ_5	-4.796	2.207	-1.962	-4.999	-0.796	-0.685	-4.412	-2.27
M_1 (GeV)	1416.12	733.398	899.545	1899.771	383.29	274.233	1578.641	590.6
M_2 (GeV)	26.11	-228.83	-2.586	40.141	-167.61	-74.405	71.025	43.3
$\tan \theta$	0.651	0.895	0.845	1.062	0.386	0.362	0.94	0.485
$m_{H_5^\pm}$ (GeV)	130.29	140.50	149.99	159.93	170.22	180.04	189.99	200.24

Table 6.1: GM model parameters used to produce masses for the $H_5^\pm$ consistent with charged Higgs test masses studied. The last row shows the actual value of the mass of the $H_5^\pm$ that is produced by the parameters in the column above it.

Values for the branching ratio $BR(H_5^\pm \rightarrow W^\pm \gamma)$ are available in the GM model for

charged Higgs masses up to 200 GeV [44]. As the vertex $H_5^\pm W^\mp \gamma$ does not exist at tree level in the GM model, the calculation of this branching ratio is non-trivial, even at leading order, so the pair production cross section is only calculated for scalar masses below 200 GeV. The results are tabulated in Table 6.2. The values calculated for $m_{H_5^\pm} = 130$ and 200 GeV are lower than the expected upper limits set in the previous section, so the products of the production cross-sections and the branching ratios are also below the expected limits. All other masses below 200 GeV have pair production cross-sections that are slightly above the expected limits, but the branching ratios in [44] are on the order of 0.1 and below. Therefore the product $\sigma(pp \rightarrow H_5^+ H_5^-) \times BR(H_5^\pm \rightarrow W^\pm \gamma)^2$ is again below the expected limits. Consequently this particular charged Higgs boson with these masses can not be observed with this method in 300 fb^{-1} of integrated luminosity of $\sqrt{s} = 14 \text{ TeV}$ proton-proton collisions.

$m_{H_5^\pm} \text{ (GeV)}$	$\sigma(pp \rightarrow H_5^+ H_5^-) \text{ (fb)}$
130	86.6(2)
140	73.1(2)
150	52.0(1)
160	41.1(1)
170	33.75(8)
180	26.91(7)
190	21.79(4)
200	17.53(3)

Table 6.2: The $H_5^\pm$ pair production cross sections in the GM model for parameters consistent with different test masses. The quoted errors in the cross-sections are the integration errors estimated by **MadGraph**.

Chapter 7

Summary and Outlook

An analysis method has been developed for a search for a charged scalar particle with mass $m_{H^\pm} \geq 130$ GeV. The technique detailed herein has been tailored to a charged scalar particle that is pair produced in proton-proton collisions at $\sqrt{s} = 14$ GeV, with both pair produced particles then decaying hadronically through a $W^\pm\gamma$ intermediate state. Some of the main backgrounds that can be expected at the LHC Run 3 — high levels of pileup and the QCD jet background — were examined, and techniques were applied to mitigate their effects on the signal. A kinematic fit was used with the goals of further suppressing the background, while also accentuating the signal peak in the $m_{jj\gamma}$ distribution.

In the event that this method does not provide evidence for the existence of such a particle, it predicts upper limits on $\sigma(pp \rightarrow H^+H^-) \times BR(H^\pm \rightarrow W^\pm\gamma)^2$ of $95.1 - 1.88$ fb for test masses in the range $130 - 500$ GeV. These values are competitive with limits set by experiments in similar channels. It was found that the scale uncertainty on the background cross-section estimate was the dominant source of uncertainty in the calculation of these limits.

This analysis and these limits would apply to a large class of singly charged scalar particles consistent with the one studied herein: the most stringent assumptions made were on the total width and mass of the charged scalar, though this method can be extended to other test

mass values, within or above the studied range. The limits set for low $m_{H^\pm}$ would require slight adjustment to apply to specific models due to the small dependence on the identity of the vector boson mediating the pair production, but it is not expected that they would change much with such a correction.

Results of this analysis were interpreted in the context of the GM model. It was found that this process would be unobservable for a charged Higgs consistent with the $H_5^\pm$ of the GM model, with mass $m_{H_5^\pm} \leq 200$ GeV. This is, however, only one of the many charged Higgs bosons that have been predicted as part of an extended Higgs sector.

Given a specific model that predicts values of $\sigma(pp \rightarrow H^+H^-) \times BR(H^\pm \rightarrow W^\pm\gamma)^2$ larger than the expected limits, and the signal and background efficiencies for the test masses studied, model-specific signal significances can be computed. These would provide estimates for the statistical significance of a signal above the background achievable with this analysis, thus quantifying the ability of this analysis to lead to a discovery if a charged Higgs boson exists. Alternatively, the expected limits set herein can be used to constrain model-specific parameter spaces.

Performing a search for a particle as rare as a charged Higgs boson at the LHC will require a very sophisticated analysis. It is the hope of the author that this work can provide inspiration for a future search, and insights into how such a search can proceed.



Bibliography

- [1] E. Rutherford. The scattering of α and β particles by matter and the structure of the atom. *Philosophical Magazine*, 21:669, 1911.
- [2] A. W. Chao and M. Tigner. *Handbook of accelerator physics and engineering*. Singapore: World Scientific, 1999.
- [3] L. Evans and P. Bryant. LHC Machine. *JINST*, 3:S08001, 2008.
- [4] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [5] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.
- [6] ATLAS Collaboration. A particle consistent with the higgs boson observed with the atlas detector at the large hadron collider. *Science*, 338:1576, 2012.
- [7] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012, arXiv: 1207.7235 [hep-ex].
- [8] C. P. Burgess and G. D. Moore. *The standard model: A primer*. Cambridge University Press, 2006.
- [9] B. Kayser. Are neutrinos their own antiparticles? *J. Phys. Conf. Ser.*, 173:012013, 2009, arXiv: 0903.0899 [hep-ph].
- [10] S. Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264, 1967.

- [11] M. Szleper. The Higgs boson and the physics of WW scattering before and after Higgs discovery. 2014, arXiv: 1412.8367 [hep-ph].
- [12] H. Georgi and M. Machacek. Doubly charged higgs bosons. *Nucl. Phys. B*, 262:463, 1985.
- [13] T. D. Lee. A theory of spontaneous t violation. *Phys. Rev. D*, 8:1226, 1973.
- [14] S.K. Vempati. Introduction to MSSM. 2012, arXiv: 1201.0334 [hep-ph].
- [15] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, and the LEP Higgs Working Group. Search for Charged Higgs bosons: Preliminary Combined Results Using LEP data Collected at Energies up to 209 GeV. 2001, arXiv:hep-ex/0107031.
- [16] CMS Collaboration. Search for charged higgs bosons produced via vector boson fusion and decaying into a pair of W and Z bosons using pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 119:141802, 2017.
- [17] ATLAS Collaboration. Search for charged higgs bosons produced in association with a top quark and decaying via $h^\pm \rightarrow \tau\nu$ using pp collision data recorded at $\sqrt{s} = 13$ tev by the atlas detector. *Phys. Lett. B*, 2016, 1603.09203.
- [18] K. Kumar K. Hartling and H.E. Logan. The decoupling limit in the Georgi-Machacek model. *Phys. Rev.*, D90(1):015007, 2014, arXiv: 1404.2640 [hep-ph].
- [19] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.
- [20] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 7:79, 2014, arXiv: 1405.0301 [hep-ph].

- [21] B. Schmidt. The high-luminosity upgrade of the lhc: Physics and technology challenges for the accelerator and the experiments. *Journal of Physics: Conference Series*, 706(2):022002, 2016.
- [22] A. Alloul et al. FEYNRULES 2.0 - A complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.*, 185:2250, 2014, arXiv: 1310.1921 [hep-ph].
- [23] C. Degrande et al. UFO - The Universal FEYNRULES Output. *Comput. Phys. Commun.*, 183:1201, 2012, arXiv: 1108.2040 [hep-ph].
- [24] T. Gleisberg et al. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009, arXiv: 0811.4622 [hep-ph].
- [25] NNPDF Collaboration. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244, 2013, arXiv: 1207.1303 [hep-ph].
- [26] NNPDF Collaboration. Parton distributions with QED corrections. *Nucl. Phys. B*, 877:290, 2013, arXiv: 1308.0598 [hep-ph].
- [27] D. Griffiths. *Introduction to elementary particles*. Wiley-VCH, 2008.
- [28] T. Sjöstrand et al. PYTHIA 6.4 physics and manual. *JHEP*, 5:026, 2006, hep-ph/0603175.
- [29] S. Mrenna T. Sjöstrand and P. Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852, 2008, arXiv: 0710.3820 [hep-ph].
- [30] J. de Favereau et al. DELPHES 3: a modular framework for fast simulation of a generic collider experiment. *JHEP*, 2:57, 2014, arXiv:1307.6346.
- [31] ATLAS Collaboration. Expected photon performance in the ATLAS experiment. Technical Report ATL-PHYS-PUB-2011-007, CERN, Geneva, 2011.

- [32] C. Y. Wong. *Introduction to high-energy heavy ion collisions*. Singapore: World Scientific, 1994.
- [33] G.P. Salam M. Cacciari and G. Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 4:063, 2008, arXiv: 0802.1189 [hep-ph].
- [34] G.P. Salam. Towards jetography. volume 67, chapter 2.1.4, page 637. 2010, arXiv: 0906.1833 [hep-ph].
- [35] M. Aaboud et al. Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC. *Eur. Phys. J.*, C77(5):332, 2017, arXiv: 1611.10235 [hep-ex].
- [36] G.P. Salam M. Cacciari and G. Soyez. The catchment area of jets. *JHEP*, 4:005, 2008, arXiv: 0802.1188 [hep-ph].
- [37] M. Cacciari and G.P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119, 2008, arXiv: 0707.1378 [hep-ph].
- [38] L. Lyons. *Statistics for nuclear and particle physicists*. Cambridge University Press, 1986.
- [39] V. Rohatgi and A.K. Md. E. Saleh. *An Introduction to Probability and Statistics*. Wiley, 1976.
- [40] T. Skwarnicki. *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*. PhD thesis, Cracow, INP, 1986.
- [41] J. Conway. *CDF/PUB/STATISTICS/PUBLIC/6428*, 2006.
- [42] M. Paterno. Calculating efficiencies and their uncertainties. *FERMILAB-TM-2286-CD*, 2004.

- [43] K. Kumar K. Hartling and H.E. Logan. GMCALC: a calculator for the Georgi-Machacek model. 2014, arXiv: 1412.7387 [hep-ph].
- [44] K. Hartling C. Degrande and H.E. Logan. Scalar decays to $\gamma\gamma$, $Z\gamma$, and $W\gamma$ in the Georgi-Machacek model. *Phys. Rev. D*, 96:075013, 2017, arXiv: 1708.08753 [hep-ph].

Appendix A

Measurement of DELPHES Detector Resolutions

As shown in Equation 4.3, in order to quantify the DELPHES detector resolutions it is necessary to compare the truth-level four-momentum of an object to its reconstructed counterpart. The process by which truth-level and reconstructed objects were matched is described in Section A.1. Sections A.2 and A.3 present the techniques used to measure the energy, and η and ϕ resolutions, respectively. Smearing was measured using a signal sample with no pileup overlaid, and no jet-to-photon fake rate.

A.1 Matching Truth and Reconstructed Objects

Smearing is performed on an object-by-object basis. As such, the distribution describing the smearing of a specific parameter must be built from a large sample of one type of object. From equation 4.3 it's clear that one needs to know the smeared and truth values of the parameter of interest, suggesting that reconstructed objects need to be paired with the truth-level objects that produced them, a process referred to as *truth-matching*.

Since smearing results in different four-vectors than are input to the detector simulation,

truth-matching is not as simple as searching for identical four-momenta. However, it is not expected that a truth-level object is drastically altered by smearing, and so a reconstructed object should remain close to its truth level counterpart in $\eta - \phi$ space. This closeness is quantified by the variable ΔR .

The truth-level counterpart to a reconstructed object was found by measuring its ΔR with all truth objects of the same identity (i.e. for a reconstructed jet, only truth jets were considered). Truth objects were considered viable candidates for the reconstructed object's partner if they satisfied $\Delta R < 0.2$ with the reconstructed object. Then, of all truth objects passing this cut, the one minimizing ΔR was hypothesized to be the truth-level counterpart. Once a truth-level object matched with a reconstructed object, it was removed from the list of candidates for all other reconstructed objects.

It was also possible that no truth-level object was close enough to the reconstructed object to be able to truth-match the reconstructed object. Such a scenario would arise in the case of extreme smearing where the reconstructed object ends up very distant from its truth-level counterpart, or if the truth-level counterpart was incorrectly paired with another reconstructed object. In this case, the smearing of the reconstructed object could not be measured.

A.2 Energy Resolution

The DELPHES simulation smears the energy of reconstructed objects differently depending on their η values. Hadronic calorimeter energy smearing had a different functional form in each of three η bins: $|\eta| < 1.7$, $1.7 < |\eta| < 3.2$, and $3.2 < |\eta| < 4.9$. The electromagnetic calorimeter had two η bins, but due to the coverage of electromagnetic calorimeters at experiments at the LHC, only one bin was relevant: $|\eta| < 2.5$.

After binning with respect to η , the reconstructed objects were further binned into 20 GeV energy bins from 10 – 190 GeV. In each bin the quantity $(E_{\text{smeared}} - E_{\text{truth}})/E_{\text{truth}}$ was

plotted and a Gaussian was fitted to each energy bin. An example of this histogram and fit is shown in figure A.1 for photons in the energy range 170 – 190 GeV, and the pseudorapidity range $|\eta| < 2.5$. The fitted standard deviation in each energy bin was taken to be the resolution. The resolutions at all energies were plotted (at the center of the corresponding energy bin) and the result was fitted with the functional form of the resolution implemented in DELPHES. Figure A.2 shows good agreement between the energy resolution measured using truth-matching and the resolution implemented in DELPHES for photons.

After validating this technique by comparing the measured energy resolutions with those hard-coded into DELPHES, the functional forms used for the energy resolutions in different $|\eta|$ regions were taken from DELPHES. For jets these were

$$\begin{aligned}
|\eta| < 1.7 : \sigma_E &= \sqrt{0.0302^2 E^2 + 0.5205^2 E + 1.59^2}, \\
1.7 < |\eta| < 3.2 : \sigma_E &= \sqrt{0.05^2 E^2 + 0.706^2 E}, \\
3.2 < |\eta| < 4.9 : \sigma_E &= \sqrt{0.0942^2 E^2 + E},
\end{aligned} \tag{A.1}$$

and for photons it was

$$|\eta| < 2.5 : \sigma_E = \sqrt{0.0017^2 E^2 + 0.101^2 E}. \tag{A.2}$$

A.3 η and ϕ Resolutions

The $x = \eta, \phi$ resolutions were treated very similarly to the energy resolutions, with the primary difference being the way the binning was performed. The DELPHES detector was segmented differently in the regions $|\eta| < 2.5$ and $2.5 < |\eta| < 5$ — the latter region only relevant for jets — so the resolution measurements were performed separately in these two regions.

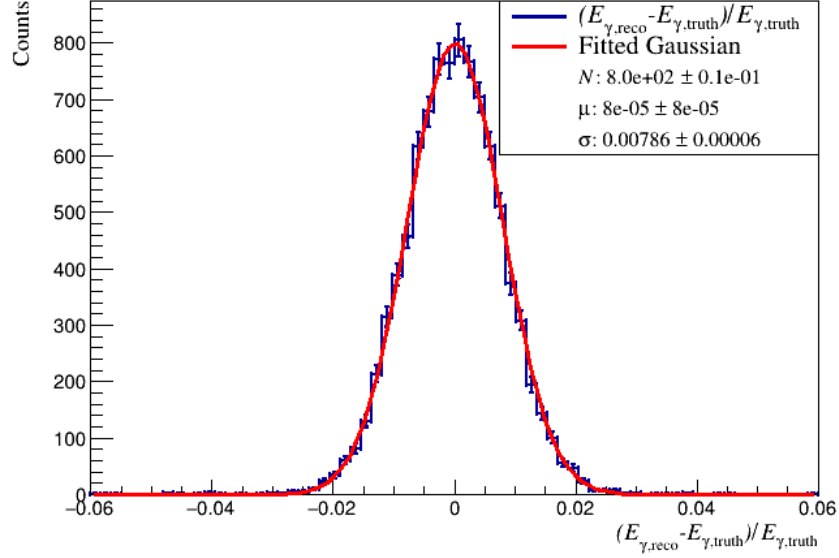


Figure A.1: The relative difference between the smeared and truth photon energy in the energy range 170 – 190 GeV and pseudorapidity range $|\eta| < 2.5$.

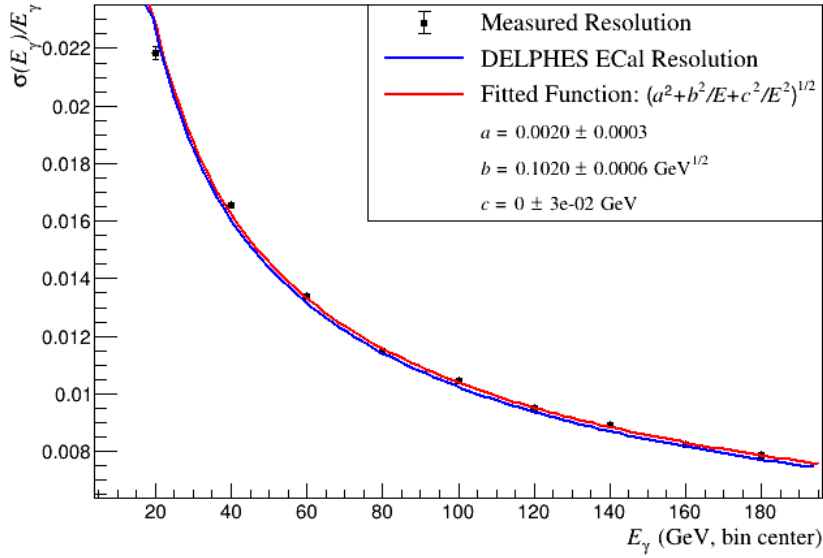


Figure A.2: Comparison between the measured energy resolution and the resolution implemented in DELPHES for the electromagnetic calorimeter for $|\eta| < 2.5$. Errors on the points from fitting for the standard deviation of the Gaussian in each energy bin are included, but are too small to be visible.

Within these two regions the quantity $x_{\text{reco}} - x_{\text{truth}}$ was binned first in $|\eta_{\text{reco}}| = 0.1$ bins, then separately in $E_{\text{reco}} = 20$ GeV bins. Gaussians were fit to each bin and the resolution in a bin was taken to be the fitted standard deviation. Binning with respect to $|\eta_{\text{reco}}|$ or E_{reco}

was found to have no effect on the resolutions in either $|\eta|$ region of the detector. Figure A.3 shows the resolution of $|\eta|$ and ϕ in the $|\eta| < 2.5$ region of the detector, binned with respect to $|\eta|$, measured using reconstructed photons.

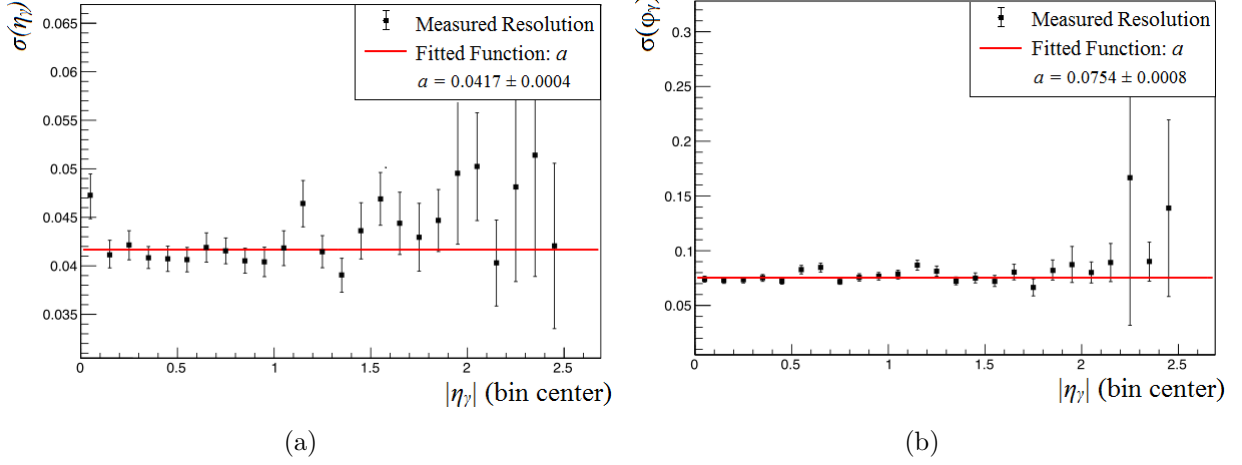


Figure A.3: (a) η and (b) ϕ resolution as a function of $|\eta|$ for photons with $|\eta| < 2.5$. Low statistics are responsible for the large errors and worsening agreement at high $|\eta|$.

The values of the jet $|\eta|$ and ϕ resolutions used in different $|\eta|$ regions were

$$\begin{aligned}
 |\eta| < 2.5 : \sigma_\eta &= 0.033, \sigma_\phi = 0.056, \\
 2.5 < |\eta| < 4.9 : \sigma_\eta &= 0.060, \sigma_\phi = 0.090,
 \end{aligned}
 \tag{A.3}$$

and for photons:

$$|\eta| < 2.5 : \sigma_\eta = 0.042, \sigma_\phi = 0.075.
 \tag{A.4}$$