

**Search for Higgs Boson Production Beyond the
Standard Model Using the Razor Kinematic
Variables in pp Collisions at $\sqrt{s} = 8$ TeV
and
Optimization of Higgs Boson Identification Using
a Quantum Annealer**

Thesis by

Alexander Mott

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy



California Institute of Technology
Pasadena, California

2015

(Defended May 20, 2015)

© 2015

Alexander Mott

All Rights Reserved

Acknowledgments

I would like to thank my PhD advisor, Maria Spiropulu, for all of her support, advice, and guidance over the course of my graduate studies. She has been an incredible mentor and her passion and enthusiasm for experimental physics have made it a pleasure as well as a privilege to work with her. I would also like to thank Harvey Newman for his insightful and valuable help and insight.

I would like to thank Massimiliano Chiorboli, who helped me get started on my very first project on CMS; Adi Bornheim, who helped me learn about the Higgs to two photon analysis; and Maurizio Pierini, whose support and deep knowledge of particle physics have been extremely helpful in many different projects.

I would like to thank Daniel Lidar and Josh Job for their help in all stages of the quantum annealer project. Their support and insight made that entire project possible.

Finally, I would like to thank my parents, Anne and Geoffrey, and my brother Christian. Without their support and encouragement I never would have been able to complete my studies and this thesis.

Abstract

In the first part of this thesis we search for beyond the Standard Model physics through the search for anomalous production of the Higgs boson using the razor kinematic variables. We search for anomalous Higgs boson production using proton-proton collisions at center of mass energy $\sqrt{s} = 8$ TeV collected by the Compact Muon Solenoid experiment at the Large Hadron Collider corresponding to an integrated luminosity of 19.8 fb^{-1} .

In the second part we present a novel method for using a quantum annealer to train a classifier to recognize events containing a Higgs boson decaying to two photons. We train that classifier using simulated proton-proton collisions at $\sqrt{s} = 8$ TeV producing either a Standard Model Higgs boson decaying to two photons or a non-resonant Standard Model process that produces a two photon final state.

The production mechanisms of the Higgs boson are precisely predicted by the Standard Model based on its association with the mechanism of electroweak symmetry breaking. We measure the yield of Higgs bosons decaying to two photons in kinematic regions predicted to have very little contribution from a Standard Model Higgs boson and search for an excess of events, which would be evidence of either non-standard production or non-standard properties of the Higgs boson. We divide the events into disjoint categories based on kinematic properties and the presence of additional b-quarks produced in the collisions. In each of these disjoint categories, we use the razor kinematic variables to characterize events with topological configurations incompatible with typical configurations found from standard model production of the Higgs boson.

We observe an excess of events with di-photon invariant mass compatible with

the Higgs boson mass and localized in a small region of the razor plane. We observe 5 events with a predicted background of $0.54 \pm 0.28(\text{*syst.*})$, which observation has a p-value of 10^{-3} and a local significance of 3.35σ . This background prediction comes from 0.48 predicted non-resonant background events and 0.07 predicted SM higgs boson events. We proceed to investigate the properties of this excess, finding that it provides a very compelling peak in the di-photon invariant mass distribution and is physically separated in the razor plane from predicted background. Using another method of measuring the background and significance of the excess, we find a 2.5σ deviation from the Standard Model hypothesis over a broader range of the razor plane.

In the second part of the thesis we transform the problem of training a classifier to distinguish events with a Higgs boson decaying to two photons from events with other sources of photon pairs into the Hamiltonian of a spin system, the ground state of which is the best classifier. We then use a quantum annealer to find the ground state of this Hamiltonian and train the classifier. We find that we are able to do this successfully in less than 400 annealing runs for a problem of median difficulty at the largest problem size considered. The networks trained in this manner exhibit good classification performance, competitive with the more complicated machine learning techniques, and are highly resistant to overtraining. We also find that the nature of the training gives access to additional solutions that can be used to improve the classification performance by up to 1.2% in some regions.

Contents

Abstract	v
I Search for Higgs boson production beyond the Standard Model using the Razor kinematic variables in pp collisions at $\sqrt{s} = 8$ TeV	1
1 Introduction	3
2 Theory	7
2.1 The Standard Model	7
2.1.1 The Higgs Mechanism	10
2.1.2 SM Higgs Boson Production Modes	10
2.2 Supersymmetry	11
2.2.1 Simplified Models	14
3 Experimental Setup	19
3.1 The Large Hadron Collider	19
3.1.1 Collisions and Pileup	20
3.2 The Compact Muon Solenoid	21
3.2.1 The CMS Solenoid	22
3.2.2 The CMS Tracker	23
3.2.2.1 Track Reconstruction	26
3.2.2.2 Primary Vertex Reconstruction	29
3.2.3 The CMS Electromagnetic Calorimeter	29

3.2.3.1	ECAL Clusters	32
3.2.4	The CMS Hadron Calorimeter	35
3.2.5	The CMS Muon System	37
4	Object Reconstruction and Identification	41
4.1	Particle Flow	42
4.2	Photons	45
4.2.1	Photon Identification	46
4.2.2	Photon Energy Regression	53
4.3	Jets	61
4.3.1	Measuring Pileup Energy	62
4.3.2	Jet Corrections	62
4.3.3	Identifying bottom quark decays	64
4.4	Missing Energy	68
4.5	The CMS Trigger	71
5	Search for Supersymmetry in Events with a Higgs Boson Decaying to Two Photons	77
5.1	Blinding Procedure	78
5.2	Datasets and Triggers	78
5.2.1	SM Higgs Boson Monte Carlo Samples	79
5.3	Event Selection	80
5.3.1	Photon Selection	80
5.3.2	Jet and E_T^M Selection	83
5.3.3	Selection of B-tagged Jets	84
5.3.4	Event Boxes	85
5.4	The Razor Kinematic Variables	89
5.4.1	General Definition	89
5.4.2	Modification of the Razor Variables for $H \rightarrow \gamma\gamma$ topologies . .	92
5.5	Background Prediction	93
5.5.1	$m_{\gamma\gamma}$ Signal Region	97

5.5.2	Non-Resonant Background Prediction	98
5.5.3	Resonant Background	101
5.5.4	Sideband Extrapolation Validation	105
5.6	Signal Regions	109
5.6.1	Signal Region Algorithm	111
5.6.2	Signal Region Event Yields	112
5.7	Systematics	112
5.7.1	Background Statistics Systematic	117
5.7.2	Background Prediction Uncertainty	118
5.7.3	Fit Choice Systematic	121
5.8	Inverted Analysis Procedure	122
5.8.1	Inverse Analysis Fits	124
6	Results and Conclusions	129
6.1	Unblinded Results	129
6.1.1	Observed Limits in the Simplified Model Scan	134
6.2	Inverted Analysis Results	136
6.2.1	Significance of the Inverted Analysis Excess	136
6.3	Excess Events Characterization	138
6.3.1	Location within the Signal Bin	139
6.3.2	Event Topologies	140
6.3.3	Event Distributions	156
6.3.3.1	Global Event Variables	156
6.3.3.2	Photon Variables	157
6.3.3.3	Mega-Jets	158
6.3.3.4	Jets	159
6.4	Conclusion and Future Work	160

II Optimization of Higgs boson identification using a quan-

tum annealer	165
7 Introduction	167
8 Quantum Annealing	171
8.1 Simulated Annealing	171
8.2 Quantum Annealing	172
9 Experimental Setup	179
9.1 Encoding the Problem	179
9.2 Embedding the Problem on the Chimera Graph	183
9.3 Practical Implementation of the Embedding	187
9.4 The D-Wave 2 TM Machine	190
9.5 Training Sample	193
10 Results	199
10.1 Network Training Results	200
10.1.1 Mapping Physical Chains to Logical Nodes	200
10.1.2 Solution Accuracy	202
10.2 Performances of the Network in Classifying Signal Events	205
10.3 Improving Success Probability and Performance	209
10.4 Current Status and Outlook	212
Bibliography	215

Part I

**Search for Higgs boson production
beyond the Standard Model using
the Razor kinematic variables in
pp collisions at $\sqrt{s} = 8$ TeV**

Chapter 1

Introduction

In *A short history of nearly everything*, Bill Bryson said [1]:

Physics is really nothing more than a search for ultimate simplicity,
but so far all we have is a kind of elegant messiness.

In particle physics, this elegant messiness is known as the Standard Model (SM), which aims to understand the most fundamental building blocks of nature. The SM is one of the most successful theories in the history of science and has accurately and precisely described practically every experimental result obtained over the last few decades at the relevant scales [2]. It has also had success at predicting the presence of subsequently observed particles such as a wide variety of mesons, the top quark, and the Higgs boson.

Despite all this success, the SM doesn't conquer every experimental challenge. It doesn't explain the mass of neutrinos required by the observation of neutrino oscillations; it doesn't explain dark matter or dark energy; it doesn't reconcile with general relativity or explain why the gravitational and electroweak scales are so far apart; and it doesn't explain the matter-antimatter asymmetry observed in the universe. It also takes quite a lot of parameters as input and thus mostly can't explain why the particles have the masses they do (the Higgs mechanism does predict masses of some particles based on other observables). It also doesn't explain why (or if) there are exactly three generations of fermions.

This thesis attempts to address some of these challenges by looking for beyond the standard model production of the Higgs boson, and in doing so help provide experi-

mental evidence for the deeper theory of nature: Bryson’s “ultimate simplicity.” The production of the Higgs boson is very precisely predicted by the standard model based on its vacuum expectation value and the masses of the fermions, so any anomalous production is strong evidence for physics not predicted by the SM. We consider this from the point of view of the production of supersymmetric particles that decay into (among other things) Higgs bosons but we aim to keep this as general as possible to allow for other interpretations. We use the di-photon decays of the Higgs boson to perform our search, since this channel gives a very clean experimental signature and still has a sufficiently high branching ratio to allow a large sample of decays to this final state.

In Chapter 2 we give an overview of the portions of theory of the standard model and supersymmetry relevant to this study. Entire theses can and have been written on each of these theories so we restrict ourselves only to the key points relevant to our experimental search. We discuss the particle component of the SM and how the forces arise through preserved local symmetries of the theory. We discuss the recently discovered Higgs boson, its theoretical origins, its properties, and the problem of the divergence of the radiative corrections to its mass. We then discuss how this problem can be solved through the introduction of a new supersymmetry, mostly in the context of the minimal supersymmetric standard model. This model has a large number of degrees of freedom, so we discuss the simplifications made to make some of its predictions experimentally testable while retaining its key attributes.

In Chapter 3 we discuss the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) detector. This analysis uses 19.8 fb^{-1} of data from proton-proton collisions recorded by the CMS detector during the 2012 run of the LHC taken at $\sqrt{s} = 8 \text{ TeV}$. We describe the various subdetectors of CMS used to measure different attributes of the particles produced by the collisions and discuss how the raw measurements made by the detectors are built into detector-level quantities.

In Chapter 4 we discuss how the detector-level quantities are transformed into measurements of the energy and momentum of particles produced in the collisions. We discuss the algorithm used to reconstruct all particles in the events with a special

focus paid to the method of reconstructing and identifying photons and jets. We discuss the multivariate methods used to provide excellent estimation of the energy of photons used in the event, which is of vital importance when working in the $H \rightarrow \gamma\gamma$ channel. We also discuss the measurement of missing energy from the detector energy deposits, which can be used to gain some information about particles escaping from the detector.

In Chapter 5 we present the analysis used to search for anomalous production of the Higgs boson. We describe the blinding procedure that prevents us from looking in the signal region before freezing the analysis and the method of estimating the backgrounds using data-driven and Monte Carlo (MC) techniques. We discuss the kinematic variables, known as the *razor variables*, used to achieve excellent sensitivity to kinematic configurations indicative of beyond the standard model processes and used to cut away standard model background.

In Chapter 6 we unblind the analysis and show the results of the search. We present both the results of the main analysis and a cross-check analysis designed to measure background in a different way to ensure that we are not biased by our method. We observe an excess of events in a signal sensitive region and proceed to analyze and characterize this excess. While the excess has a moderate significance, it is a small absolute number of events, so characterization cannot be performed in a compelling way. We conclude by looking forward to the next run of the LHC, where the true nature of the observed deviation will be resolved.

Chapter 2

Theory

2.1 The Standard Model

The Standard Model (SM) of particle physics is our best current description of the interactions between all known elementary particles. It includes spin- $\frac{1}{2}$ fermions that make up matter and obey the Pauli exclusion principle, spin-1 bosons that mediate forces between these fermions, and the spin-0 Higgs boson that “gives mass” to elementary particles. The fermions are divided into leptons (Table 2.1) and quarks (Table 2.2). The bosons are summarized in Table 2.3.

Name	Symbol	Charge	Mass (GeV)
electron	e^-	-1	$5.11 \times 10^{-4} \pm 1.1 \times 10^{-11}$
muon	μ^-	-1	$0.11 \pm 3.5 \times 10^{-9}$
tau	τ^-	-1	$1.78 \pm 1.6 \times 10^{-4}$
electron neutrino	ν_e	0	$< 2 \times 10^{-9}$
muon neutrino	μ_μ	0	$< 2 \times 10^{-4}$
tau neutrino	μ_τ	0	< 0.02

Table 2.1: The properties of the leptons in the SM. (source: PDG [3])

Mathematically, the SM is a gauge theory defined by a Lagrangian \mathcal{L} involving fields. The Lagrangian of the SM is the most general renormalizable expression respecting a set of symmetries. These symmetries are given by a gauge group, the action of whose members leave the lagrangian invariant. The Lagrangian is required to respect these symmetries not only globally, but locally as well. The promotion of a global symmetry of the Lagrangian to a local symmetry requires the introduction

Name	Symbol	Charge	Mass (GeV)
up	u	$+\frac{2}{3}$	$2.3_{-0.5}^{0.7} \times 10^{-3}$
charm	c	$+\frac{2}{3}$	1.3 ± 0.025
top	t	$+\frac{2}{3}$	$173 \pm 0.51 \pm 0.71$
down	d	$-\frac{1}{3}$	$4.8_{-0.3}^{+0.5} \times 10^{-3}$
strange	s	$-\frac{1}{3}$	0.095 ± 0.005
bottom	b	$-\frac{1}{3}$	4.18 ± 0.03

Table 2.2: The properties of the quarks in the SM. (source: PDG [3])

Name	Symbol	Charge	Mass (GeV)	Spin
photon	γ	0	0	1
W-boson	W^\pm	± 1	80.4 ± 0.015	1
Z-boson	Z^0	0	91.2 ± 0.0021	1
gluon	g	0	0	1
Higgs boson	H	0	125.7 ± 0.4	0

Table 2.3: The properties of the bosons in the SM. Charge represents EM charge only and central values for the masses are quoted. (source: PDG [3])

of a bosonic field to form a gauge covariant derivative. This field can be interpreted as a force between different points in space-time and the quantization of this field is interpreted as the force carrier [4].

Each interaction in the SM is associated with an interaction strength α . A consequence of the renormalizability of the theory is that these strengths will change, or “run,” with the energy of the interaction. The way in which the coupling constants run is dependent on the group structure and form factors of the symmetry group governing the interaction and corresponding form factors.

The interactions of the standard model particles under the electromagnetic interaction can be described by quantum electrodynamics (QED), which is a gauge theory based on a $U(1)_{em}$ symmetry group [4]. Promoting this global symmetry to a local symmetry requires that the lagrangian be invariant under space-time dependent complex phase changes

$$\psi(x) \rightarrow e^{i\alpha(x)}\psi(x)$$

which requires the introduction of a spin-1 massless gauge boson, the photon. Technically the masslessness of the photon is a consequence of its interaction with the

Higgs field, which will be discussed further in section 2.1.1.

In addition to their electromagnetic interactions, the quarks also interact via the strong force. The gauge theory of the strong force, called quantum chromodynamics (QCD), is based on the $SU(3)_C$ symmetry group [5, 6]. The quarks carry a charge under this symmetry, known as color charge, which is selected from one of three charges (red, green, and blue). Promoting the $SU(3)$ symmetry to a local symmetry introduces a massless spin-1 boson that mediates this force, known as the gluon. Unlike in QED, where the photon is neutral, the gluon carries color charge with two separate colors allowing differently charged quarks to interact. There are 8 gluons with different combinations of color index corresponding to the the 8 generators of the adjoint representation of $SU(3)$.

The structure of the field strength tensor for the gluon gives rise to a phenomenon known as *asymptotic freedom*, where the strong force weakens between particles at high energies. Over short distances, colored particles behave like free particles, while as distances increase the strength of the strong force increases. If two colored particles get too far apart, there is eventually so much energy in the gluon field between them that it is energetically favorable to create a new quark-antiquark pair. If a colored particle is produced with a large momentum, as in LHC collisions, this process will be repeated many times, creating a large number of color neutral baryons travelling in the direction of the initial particle. This process is known as hadronization and forms what we call a jet.

The final component of the SM is the weak interaction, which is based on an $SU(2)_L \times U(1)_Y$ symmetry group. Gauging this group produces 4 massless spin-1 gauge bosons (in fact, the $U(1)_{em}$ group of the electromagnetic force is actually a subgroup of this interaction and the gauging of this full electroweak theory produces the photon), and the mixing of these produces the W^\pm and Z^0 bosons. These bosons are massless in the basic electroweak theory and acquire mass through the Higgs mechanism.

2.1.1 The Higgs Mechanism

The Higgs mechanism is a process of adding an additional $SU(2)$ doublet of spin-0 fields (ϕ) to the electroweak lagrangian, which produces electroweak symmetry breaking (EWSB) [7–11]. This field is assigned a quadratic potential

$$V(\phi) = -\mu^2\phi^\dagger\phi + \frac{\lambda}{2}(\phi^\dagger\phi)^2$$

with $\lambda, \mu^2 \in \mathbb{R}$ and $\mu^2 > 0$. This potential has a degenerate ground state where the Higgs field acquires a nonzero vacuum expectation value (VEV) $v = \sqrt{\mu^2/\lambda}$ due to the requirement that $\mu^2 > 0$. This VEV breaks the $SU(2)_L \times U(1)_Y$ electroweak symmetry.

One combination of generators of $SU(2)_L \times U(1)_Y$ leaves the VEV invariant, which leaves the boson massless. This combination produces $U(1)_{em}$ and the electromagnetic theory. The other combinations of generators acquire additional longitudinal degrees of freedom and gain mass, and produces the W^\pm and Z^0 bosons. One neutral scalar field from the Higgs doublet is left and is associated with the Higgs boson. This boson in turn gives mass to the fermions through Yukawa couplings. The Higgs boson acquires a mass from the VEV equal to $m_H = \sqrt{2\lambda}v$. The parameters μ^2 and λ are not set by the theory, but CMS and ATLAS have measured $m_H = 125.7 \pm 0.4$ GeV [3].

2.1.2 SM Higgs Boson Production Modes

At $\sqrt{s} = 8$ TeV, there are four major production modes to produce a Higgs boson in pp collisions, shown in Figure 2.1 [12]. For $m_H = 125$ GeV, the SM cross sections and expected number of total events to be produced in the 2012 CMS dataset can be seen in Table 2.4.

The different production modes are associated with different kinematics and final states for the events containing the Higgs boson. Additional objects may be produced in the interaction from initial state radiation of the partons from the proton or from

final state radiation of one of the produced objects. These processes are simulated using PYTHIA 6 [13] and POWHEG [14–18] and use GEANT 4 [19] to simulate interaction with the detector.

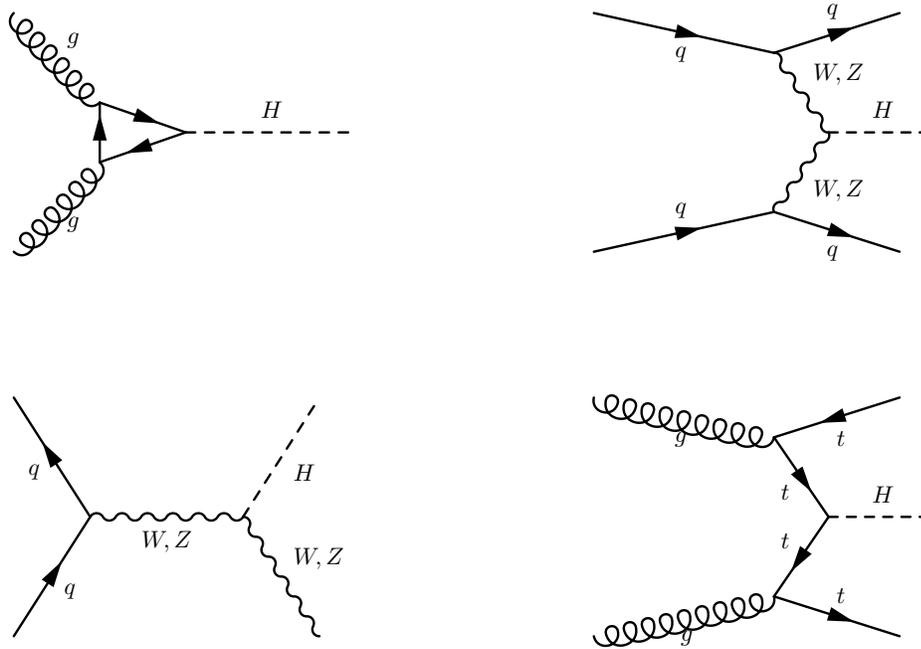


Figure 2.1: The four production modes of the SM Higgs boson accessible in the LHC with $\sqrt{s} = 8$ TeV and $\int \mathcal{L} dt = 19.8 \text{ fb}^{-1}$. They are called gluon fusion (top left), vector boson fusion (top right), vector boson associated production (bottom left) and top associated production (bottom right).

2.2 Supersymmetry

While the Higgs mechanism provides an elegant solution to the problem of electroweak symmetry breaking, it introduces at least one additional problem: the mass of the higgs boson receives enormous quantum corrections from every particle that couples to it (so, every massive particle). The Higgs boson coupling to a fermion f will receive

Process Name	Process	Cross-section (pb)	Expected Events in 19.8fb ⁻¹
Gluon Fusion	$pp \rightarrow H$	19.27	869.9
Vector Boson Fusion	$pp \rightarrow qqH$	1.578	71.2
Associated Production	$pp \rightarrow VH$	1.1199	50.6
Top Associated Production	$pp \rightarrow ttH$	0.1293	5.8
Total		22.1	997.6

Table 2.4: The SM Higgs boson production mechanisms considered as background to this analysis. The cross-sections are for 8 TeV pp collisions with $m_H = 125$ GeV. To get the expected events, we use $Br(H \rightarrow \gamma\gamma) = 2.28 \times 10^{-3}$. Cross-sections are taken from [12].

a one loop quantum correction to its mass of

$$\Delta m_H = -\frac{|\lambda_f|^2}{8\pi^2} \Lambda_{UV}^2 + \dots$$

where λ_f is the strength of the coupling ($|\lambda_{top}| \approx 1$), Λ_{UV} is the scale at which new physics enters and cuts off the loop integral, and the higher order terms go at most with $\log(\Lambda_{UV})$. If we consider the coupling to a scalar, we instead get contributions of the form

$$\Delta m_H = \frac{\lambda_S}{16\pi^2} [\Lambda_{UV}^2 + m_S^2 \log(\Lambda_{UV}/m_S) + \dots]$$

If we believe that the SM is the full theory, then we must take $\Lambda_{UV} \approx M_{Planck}$ where quantum gravity becomes important. This provides a value of m_H 30 orders of magnitude incompatible with the measured value [20–22].

One solution is to invent new physics at a lower scale so the Λ_{UV} is lower; however, this still raises issues if they couple even indirectly to the Higgs boson. A more robust solution is to cancel the divergences in Δm_H by a symmetry. We notice that the contributions from fermions and scalars have opposite sign, so if each fermion in the standard model is accompanied by two complex scalars with $\lambda_S = |\lambda_f|^2$ then the quadratic divergences can cancel. Since there are no candidates for this pairing within the standard model, a new symmetry between fermions and bosons called *supersymmetry* (SUSY) is posited [23].

There are many possible models that implement this possibility, but the simplest is the direct symmetrization of the fields of the SM, called the minimal supersymmetric

standard model (MSSM). We form chiral supermultiplets pairing each standard model fermion with its supersymmetric partner. Since the left-handed and right-handed components of the fermions have different gauge transformation properties in the SM, each must have its own complex scalar partner (hence we achieve the required 2 SUSY particles for each fermion). The scalar partners of the fermions are known as sfermions (squarks, sleptons, etc.) and are usually denoted with a tilde (\tilde{q} , \tilde{l} , etc.).

The vector bosons are associated with spin- $1/2$ fermionic superpartners, referred to as gauginos and placed in gauge supermultiplets. Before EWSB, the $SU(2)_L \times U(1)_Y$ symmetry of the EW gauge is associated with 4 vector bosons: W^\pm , W^0 , and B^0 , each of which has its own fermionic superpartner denoted with a $\tilde{}$ (e.g. \tilde{W}^\pm). After EWSB, just as the SM W^0 and B^0 mix to give the Z and γ , the \tilde{W}^0 and \tilde{B}^0 mix to give \tilde{Z} and $\tilde{\gamma}$. Taken together with the Higgsinos (also spin- $1/2$), these sparticles can mix to form 2 mass eigenstates which we call neutralinos $\chi_{i=1,2}^0$; the charged \tilde{W}^\pm states can also mix to form 2 charginos $\chi_{i=1,2}^\pm$. In many models, the lightest neutralino (written χ_1^0 or, more commonly, χ_0) is the lightest supersymmetric particle.

Supersymmetry does not independently preserve baryon and lepton number, which leaves it open to predicting proton decay and lifetimes incompatible with the observation of the lifetime of the universe and lower limits on the proton lifetime. One way to avoid this is to introduce a quantity known as R-parity

$$R \equiv (-1)^{2s+3(B-L)}$$

that is conserved by supersymmetry (s is the spin of the field). It can either be added in an ad-hoc way, or it can be added as part of a model that conserves R-parity automatically [24]. By construction, $R=+1$ for all SM particles and $R=-1$ for all SUSY particles. The introduction of this conserved quantity implies that there are always an even number of SUSY particles in any interaction, which has important phenomenological consequences:

- SUSY particles must be pair produced at the LHC
- a SUSY particle must always decay into an odd number of SUSY particles as

well as an arbitrary number of SM particles

- the lightest supersymmetric particle (LSP) is stable

The third point follows as consequence of the second (the lightest particle has no valid R-parity preserving decays). This implies that a weak scale LSP could potentially be a candidate for a particle explanation of dark matter. It also means the LSPs would escape the CMS detector undetected.

Since this whole discussion began by introducing SUSY to cancel divergences in the Higgs boson mass, it is no surprise that massive SUSY particles couple to the SM Higgs boson. This means that the decays $\tilde{q} \rightarrow H\chi_0$ and $\chi_2 \rightarrow H\chi_0$ should be allowed when there is enough mass splitting between the particles to allow it. So far, SUSY has eluded all attempts to find it, but these decays, coupled with the measurement of the mass of (a) Higgs boson, offer a promising new channel.

2.2.1 Simplified Models

If supersymmetry is realized in nature it must be a broken symmetry, since sparticle states with identical masses to their SM partners have been excluded by previous experiments. This implies that there must be supersymmetry violating terms in the SUSY Lagrangian; in order to avoid adding additional ultraviolet divergences, this breaking should be *soft*. With all the soft terms added, SUSY adds several hundred new masses, mixing angles and phases to the model.

Since a model with several hundred parameters is difficult to work with when presenting experimental results, we often interpret our results in the context of simplified models that make the assumption that only a few states have masses that are accessible at the LHC [25]. In particular, in this work, we focus on a model where only the neutralinos and charginos are accessible at the LHC, with the lightest neutralino $\tilde{\chi}_1^0$ being the LSP and the chargino $\tilde{\chi}_1^\pm$ being the NLSP. In models such as this, the key

decays to consider are

$$\begin{aligned} \tilde{\chi}_2^0 &\rightarrow H\tilde{\chi}_1^0 & \tilde{\chi}_2^0 &\rightarrow Z\tilde{\chi}_1^0 \\ \tilde{\chi}_2^0 &\rightarrow W^\pm\tilde{\chi}_1^\mp & \tilde{\chi}_1^\pm &\rightarrow W^\pm\tilde{\chi}_1^0 \end{aligned}$$

If we assume that the χ_2^0 and χ_1^\pm are nearly degenerate (we drop the $\tilde{\chi}$ notation for convenience), then the 3^{rd} decay is through a virtual W^* , which decays to low momentum particles. In an R-parity conserving model, the phenomenology of these decay channels is a set of final states with either two vector bosons, two Higgs bosons, or a vector boson and a Higgs boson, all with two additional neutralinos that escape the detector [26]. If we consider only models where the mass splitting between the NLSP and LSP is $\Delta m \gtrsim 130$ GeV, then all of these bosons would be on-shell. Of course there are other potential decay channels for the neutralinos and charginos that may or may not be allowed or dominant in different regions of parameter space, but we focus on cases where the decays listed are most important.

From an experimental point of view, final states including a Higgs boson have some attractive properties that can be exploited to increase the sensitivity to these models. The Higgs boson has a much smaller width than the Z or W (≈ 4 MeV versus 2.1 and 2.5 GeV respectively), so in a real experiment the observed width is just the resolution of the detector and the major production modes are straightforward to model. Its decay into photons gives it a very clean experimental signature, and its relatively small standard model cross section means it is quite easy to find kinematic regions with virtually no predicted Higgs boson events without going onto the tails of the distributions. In this study, therefore, we focus only on decays that produce a Higgs boson.

Figure 2.2 shows the diagrams of the simplified model that we study here. We require at least one leg to have the $\chi_2 \rightarrow H\chi_1$ decay and allow the other leg to decay as it will. We have already said that we take the χ_1^\pm and the χ_2 to be near degenerate, so we are left with only two more parameters to consider: the mass of the LSP m_{χ_0} and the mass of the NLSP $m_{\chi_1^\pm} \approx m_{\chi_2}$. In order to keep the Higgs boson

on shell, we require that $\Delta m = m_{\chi_2} - m_{\chi_0} > 130$ GeV, but otherwise we allow for different possible values for the two masses. We call these combinations of different mass parameters a *simplified model scan* (SMS), and we will use these scans over valid mass combinations to design the analysis and provide some interpretation for this type of model. Since these SMSs are not really full models, we use them mostly to help guide the analysis design and allow comparisons between different analyses. We keep the analyses general enough to allow re-interpretation with other models.

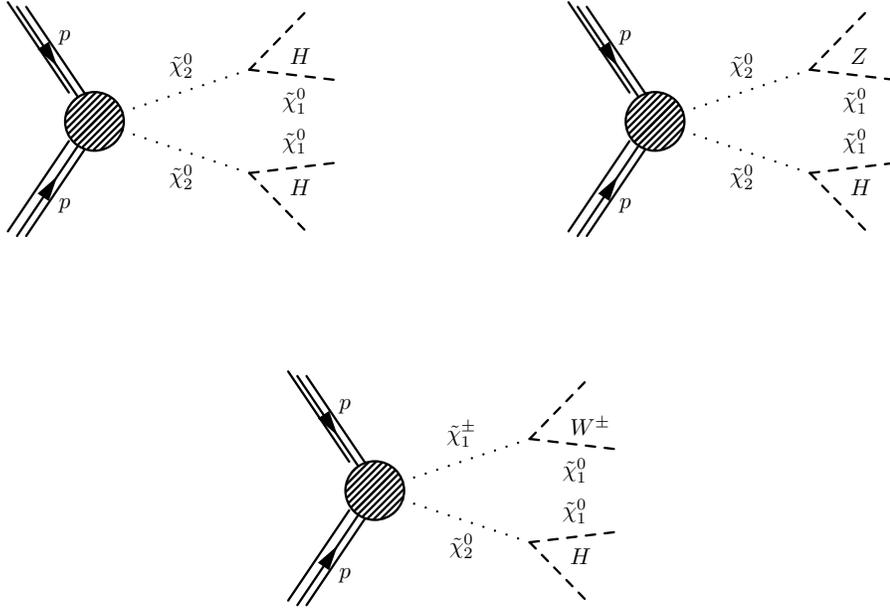


Figure 2.2: The diagrams under consideration in the simplified model studied here. The diagrams on the top both have production of two neutralinos (χ_2); the diagram on the top left has both of those neutralinos decaying $\chi_2 \rightarrow H\chi_0$, while the diagram on the right has one neutralino decaying $\chi_2 \rightarrow H\chi_0$ and the other decaying $\chi_2 \rightarrow Z\chi_0$. The diagram on the bottom has asymmetric $\chi_2\chi_1^\pm$ production with the neutralino decaying $\chi_2 \rightarrow H\chi_0$ and the chargino decaying $\chi_1^\pm \rightarrow W^\pm\chi_0$.

Figure 2.3 shows the expected cross-section for neutralino-neutralino and chargino-neutralino production as a function of the mass [27, 28]. We can see that the cross-section falls by two orders of magnitude between 130 and 500 GeV, so it is useful to calculate how many events we expect to see in the LHC data as a function of the mass. The LHC has collected 19.8 fb^{-1} of data at $\sqrt{s} = 8$ TeV, and we only want

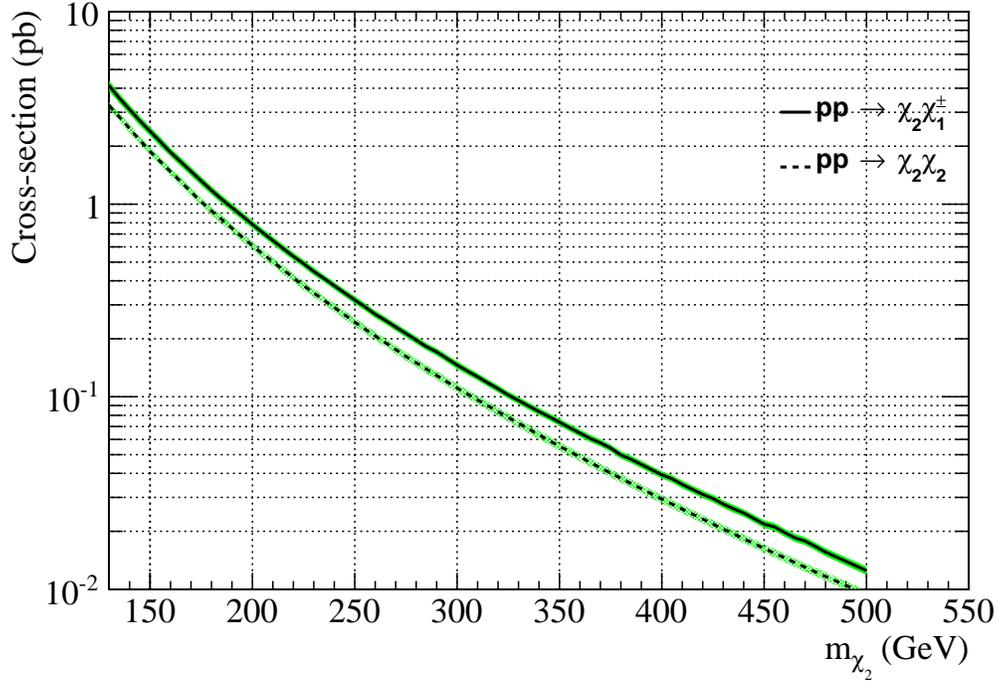


Figure 2.3: The production cross section for neutralino-neutralino and chargino-neutralino production as a function of the mass in the simplified model under consideration for pp collisions at $\sqrt{s} = 8$ TeV.

the decays of the Higgs boson to two photons, so we account for the branching ratio (0.0028 at $m_H = 125$ GeV) so we can calculate that we expect to see one event for a model with a cross-section of $\sigma = 0.04$ pb with an acceptance of 50%. Based on the scaling of the cross section, we only consider models with $m_{\chi_2} < 200$ GeV, which would produce at least a few events under realistic acceptance. We can therefore define the set of models that are guiding this analysis as those shown in Figure 2.2 with $m_{\chi_2} < 200$ GeV and $m_{\chi_1^0} < 70$ GeV.

Chapter 3

Experimental Setup

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a superconducting hadron collider located at CERN near Geneva, Switzerland [29]. It occupies a 26.7 km circumference tunnel that lies between 45 m and 170 m below the surface originally built for the LEP experiments. It is designed to provide collisions between two beams of protons traveling in opposite directions within the ring each with an energy of 7 TeV at a peak instantaneous luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Thus far, the machine has achieved 8 TeV collisions (4 TeV in each beam) and a peak luminosity nearing $7.7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ [30].

The beams are divided into clusters of protons, called bunches, within the machine. The LHC is designed to operate with a 25 ns spacing between the bunches and a total of 2,808 bunches, each containing 1.15×10^{11} protons, within the ring. Thus far only a spacing of 50 ns between the bunches has been achieved with a peak of 1380 bunches circulating. The bunch charges have been higher than design, with a peak of 1.6×10^{11} protons per bunch achieved.

The beam circulates in two parallel rings where the bunches are bent using superconducting dipole magnets. For cost and space reasons, the two counter-rotating beams are housed in adjacent beam tubes within the same cryostat. Since the beams have the same charge, they each require a magnetic field pointing in the opposite direction to stay within the ring. Dipole magnets are used to achieve this, producing two fields of nearly equal magnitude and opposite direction. This configuration is

shown in Figure 3.1.

LHC DIPOLE : STANDARD CROSS-SECTION

CERN AC/DI/MM - HE107 - 30 04 1999

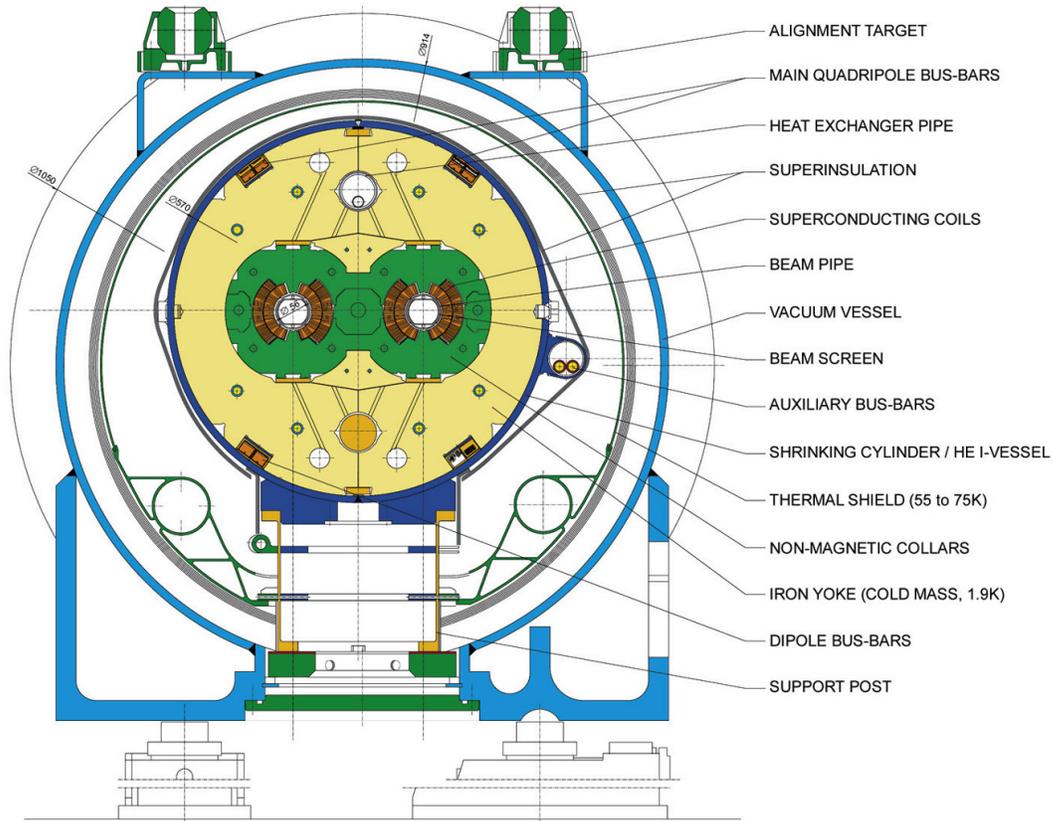


Figure 3.1: Schematic cross section of an LHC dipole magnet.

The dipole magnets are NbTi superconductors maintained at 2 K by a cryostat. They are capable of producing a field of up to 8 T, with the exact strength set depending on the energy of the protons within the machine. The LHC rings contain 1,232 of these magnets.

3.1.1 Collisions and Pileup

As mentioned before, the peak luminosity achieved in 2012 has been $L = 7.7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ divided into 1380 bunches separated by 50 ns. For most of the length of the ring, the two beams counterrotate in separate parallel beampipes. To produce collisions, the two beams are steered into a single beampipe and pass through each other at 8 interaction points (IPs) around the ring [29]. The collisions are then a stochastic process,

where some protons from one bunch collide with other protons from the other bunch. One can calculate the peak expected number of collisions per crossing from the peak luminosity ($\mathcal{L} = 7.7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ per bunch), the total proton-proton cross section at $\sqrt{s} = 8 \text{ TeV}$ ($\sigma_{pp} = 101.7 = 101.7 \times 10^{-27} \text{ cm}^2$, measured by TOTEM [31]) and the bunch spacing ($n = 5.0 \times 10^{-8} \text{ s}$):

$$\langle I \rangle = \mathcal{L} * \sigma_{pp} * n = 39.1 \quad (3.1)$$

So in each bunch crossing, we expect about 39 individual proton-proton interactions at peak luminosity. We are only interested in a very tiny fraction of the collisions that are produced, but each time one collision in the bunch crossing produces a particle of interest (a Higgs boson, for example), we will also record all the particles produced by the other collisions in the same bunch crossing. We call these interactions “pileup” and dealing with it is one of the largest challenges for any analysis at the LHC. We describe in Section 3.2.2.2 how we determine which vertex produced particles of interest (the primary vertex) and in Section 4.3.1 how we identify and remove the energy from particles produced by these pileup interactions from the reconstruction of the particles produced by the interaction at the primary vertex.

3.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is one of two general purpose detectors at the LHC [32] [33]. A schematic outline is shown in Figure 3.2, showing all the major components. CMS is designed to accurately measure all charged and neutral particles produced by a proton-proton collision using a series of different components or *sub-detectors*, each specialized at measuring specific attributes of products from the collision. At its core is the eponymous superconducting solenoid that produces a 3.8 T magnetic providing large bending power to enable precise determination of the momentum of charged particles. The detector is divided into 4 main subdetectors: the tracker, for measuring the path of charged particles in the magnetic field; the

electromagnetic calorimeter (ECAL), for measuring the shower from electrons and photons; the hadron calorimeter (HCAL), for measuring the shower from hadrons; and the muon systems, for measuring the energy and momentum of muons.

We define a coordinate system that will be used consistently when talking about positions of subdetectors and particles with the (0,0,0) point at the center of the detector. The \hat{x} direction points toward the center of the LHC ring from the center and the \hat{y} direction points upward toward the surface. The \hat{z} direction, then, points along the beamline to the left side of the detector (toward the Jura mountains) to make a right-handed coordinate system. Azimuthal angle (ϕ) is measured in the x-y plane from the x axis while polar angle is measured from the z axis. We will frequently use the variable pseudorapidity, which is defined as

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

We call a vector “transverse” to the beamline if $\vec{v} \cdot \hat{z} = 0$ (i.e., \vec{v} lies in the x-y plane). The detector is divided into three major regions: the barrel, covering the central (low η) part of the detector transverse to the beamline; the endcaps, covering the intermediate (middle η) regions; and the forward regions, covering the very forward (high η) region where particles from the interaction emerge nearly parallel to the beamline.

3.2.1 The CMS Solenoid

The CMS solenoid is shaped like a hollow cylinder centered around the beamline, with an internal diameter of 6 m and a length of 12.5 m [34] [35]. It is a superconducting magnet designed around a 4 layer NbTi coil to produce a 4 T magnetic field within the hollow bore; at full current, the solenoid has a stored energy of 2.6 GJ. The total cold mass of the solenoid is 220 T, giving a relatively high Energy/Mass ratio of 11.6 KJ/kg, causing a large (0.15%) mechanical deformation of the solenoid during energizing, significantly larger than the deformations found for previous detector magnets.

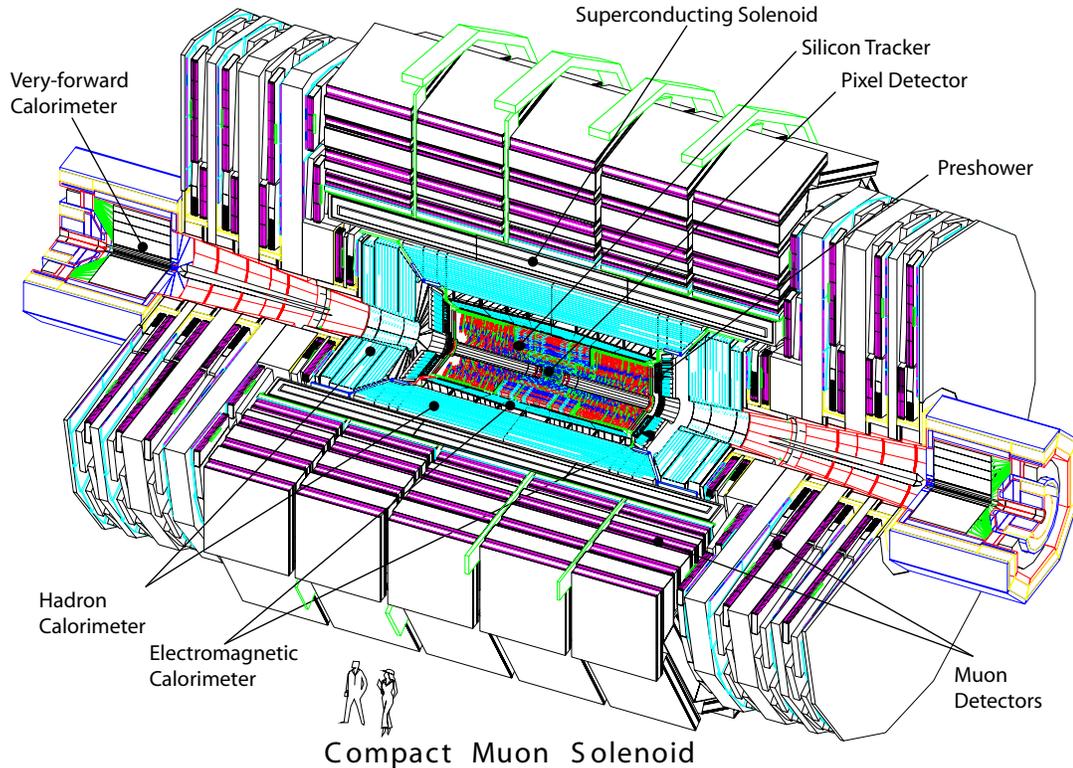


Figure 3.2: Schematic overview of the CMS detector.

The solenoid is situated within an 10,000 T iron return yoke consisting of 3 distinct layers divided into 7 regions (5 wheels in the central region and 2 endcap regions). A picture of the return yoke under construction can be seen in Figure 3.3. The solenoid produces a significant field outside the bore and the iron yoke guides the return field back along distinct channels (see Figure 3.4), allowing muon chambers to be placed outside the solenoid for muon identification and measurement.

3.2.2 The CMS Tracker

The CMS tracker is designed to measure the trajectory of individual charged particles produced in an interaction as they bend in the magnetic field produced by the solenoid. At 50 ns bunch spacing, each bunch crossing produces up to 40 interactions (see Section 3.1.1), which altogether produce around 2000 charged particles [33]. The CMS tracker is composed of a pixel detectors with three layers in the barrel region



Figure 3.3: The assembly of the yoke.

and 2 disks in the endcap region and a silicon strip tracker with 10 layers in the barrel region and 3+9 disks in the endcap region [36]; the configuration can be seen in Figure 3.5.

The three cylindrical layers of the pixel detector are situated at radii of 4.4, 7.3, and 10.2 cm from the beamline. The two disks extend from ≈ 6 to 15 cm in the radial direction and are located at $z = \pm 34.5$ cm and $z = \pm 46.5$ cm from the interaction point. Together, these cover the pseudorapidity range $-2.5 < \eta < 2.5$. The barrel pixels consist of 48 million channels while the endcap disks consist of 18 million channels each. Each channel has dimensions $100 \times 150 \mu\text{m}^2$, which allows the total occupancy of the pixels to be kept very low ($\approx 10^{-4}$) even with the very high particle multiplicities expected at design luminosity. The placement of the barrel and pixel layers allows almost the entire region between $-2.5 < \eta < 2.5$ to be covered by at least 3 layers of pixels (see Figure 3.6).

The strip tracker is located outside the pixel detector and is composed of silicon de-

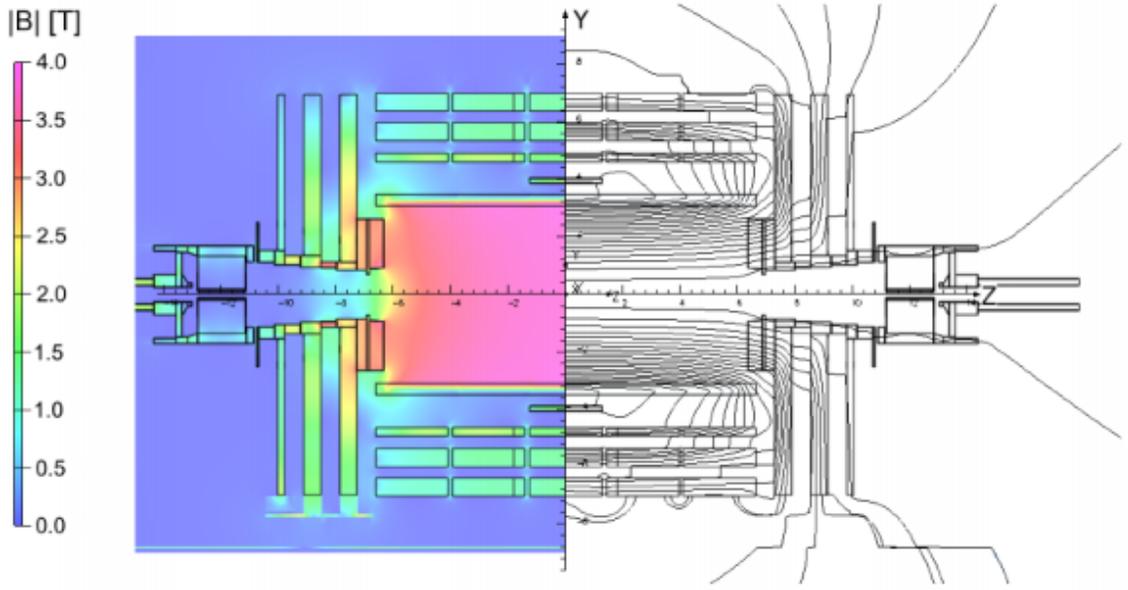


Figure 3.4: Schematic view of the CMS detector with the magnetic field strength shown (left half, color Z axis) and the magnetic field lines shown (right half). One sees the magnetic field is strong and uniform inside the solenoid, while outside the solenoid it is concentrated inside the iron return yokes, which guide the magnetic field lines.

ectors with a typical cell size of $10 \text{ cm} \times 80 \mu\text{m}$ in the inner strips (tracker inner barrel or TIB, and tracker inner disks or TID: $20 \text{ cm} < r < 55 \text{ cm}$) and $25 \text{ cm} \times 180 \mu\text{m}$ in the outer strips (tracker outer barrel or TOB, and tracker endcaps or TEC: $55 \text{ cm} < r < 110 \text{ cm}$). The increased distance from the interaction point keeps the occupancy manageable despite the larger size of the strip detectors (2-3% in the inner strips and $\approx 1\%$ in the outer strips). There are a total of 9.6 million strip detectors arranged in 15148 detector modules covering a total area of over 200 m^2 . The inner tracker (TIB/TID) provides up to 4 $r - \phi$ measurements on a given trajectory. The TOB provides another 6 $r - \phi$ while the TEC provides up to 9 ϕ measurements per trajectory. The individual measurements from the layers of the pixel and silicon trackers are used by the track reconstruction to reconstruct the trajectories of particles moving through the tracker.

The material budget for the tracker is shown in Figure 3.7. One sees that in the very central region $|\eta| < 1$, the tracker is very “thin” (less than one radiation length and around 0.2 interaction lengths), while in the more forward regions $1.5 < |\eta| < 3$

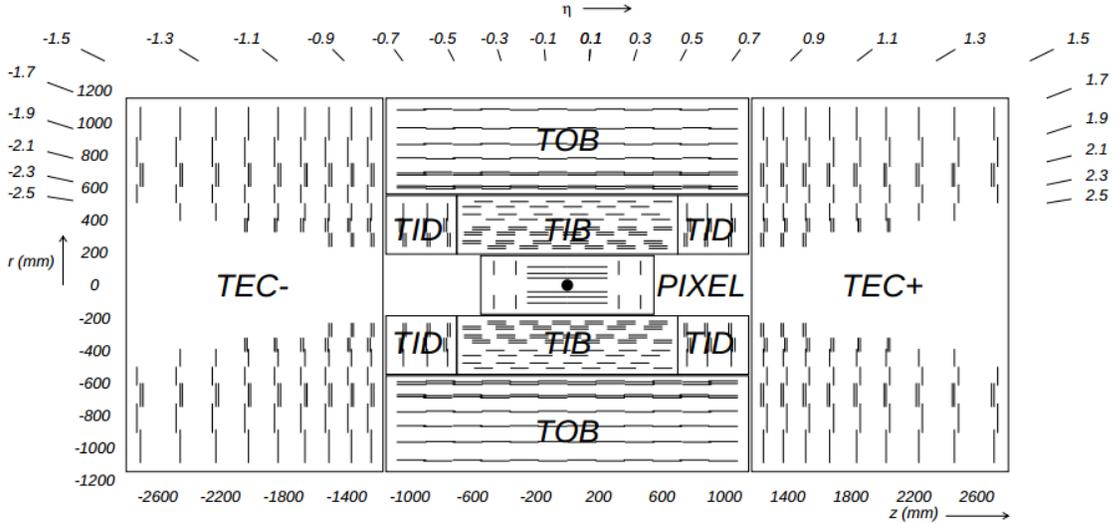


Figure 3.5: The layout of the subdetectors within the CMS tracker. The three barrel and two endcap layers of the pixel detector are shown in the box around the interaction region (the solid dot in the middle). The silicon tracker is divided into Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB) layers covering the barrel and Tracker Inner Disks (TID) and Tracker EndCaps (TEC \pm) covering the endcaps. From [32]

it is comparatively thick. This will play an important role in our reconstruction of photons and electrons, which is discussed in Section 4.2.

3.2.2.1 Track Reconstruction

We reconstruct the path of charged particles through the CMS tracker using the hits measured by the pixel and strip detectors and the equations of motion for charged particles moving in the homogeneous magnetic field produced by the CMS solenoid. From a given seed, the definition of which is discussed below, an initial trajectory is estimated from the best fit to those points in 3D space, varying the kinematics of the potential charged particle. In an iterative procedure, more hits are added to the trajectory, to further refine the estimation until all hits are accounted for.

The iterative algorithm used to find tracks is known as the Combinatorial Track Finder (CTF) [37]. The collection of reconstructed tracks is produced by multiple applications (iterations) of the CTF algorithm. The aim is to find the easiest tracks (e.g. high p_T tracks produced near the interaction region) first, remove the hits

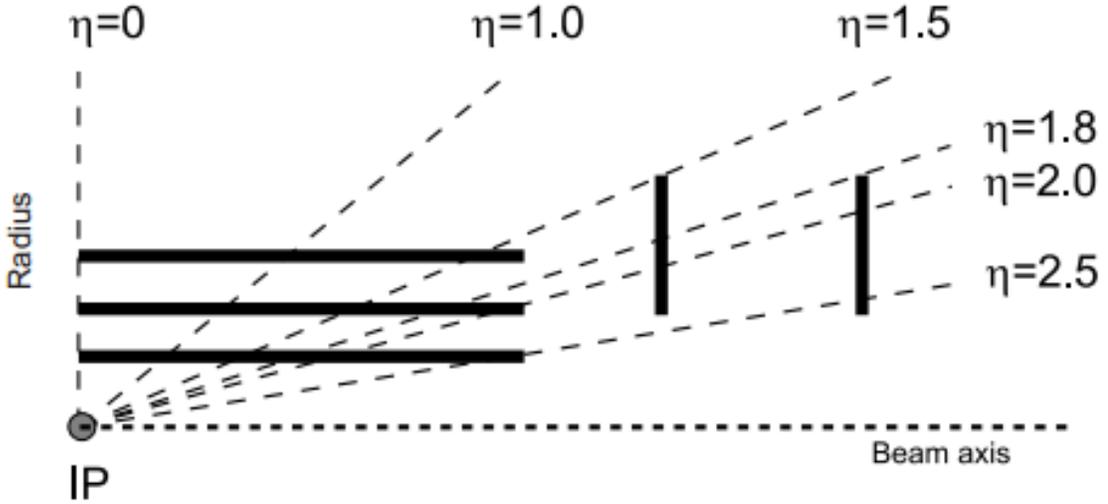


Figure 3.6: The coverage of the pixel detector. One can see that almost the entire region between $-2.5 < \eta < 2.5$ is covered by 3 layers of pixels, with a small region near the edge of the coverage only covered by the innermost barrel layer and the outermost endcap disk.

associated with these tracks, and then move on to progressively harder to find tracks (lower p_T or displaced from the interaction region) with the smaller hit collection reducing the computational challenge at each step. The algorithm uses 6 iterations: iterations 0-2 are designed to find prompt tracks with successively lower p_T or track quality, while iterations 3-5 are designed to find tracks originating outside of the main interaction region.

In addition to the hits recorded by the tracker, the track reconstruction algorithm uses information about the beam spot, which is a measurement of the average interaction location over many events within a given period of collisions. This serves as an initial estimate for the interaction location for each event. Additionally, initial per-event vertex information is reconstructed from triplets of hits in successive layers of the pixel detectors (*pixel tracks*), which are extrapolated back to the interaction region to form pixel vertices (in a manner analogous to that described in Section 3.2.2.2). This provides a fast estimation of the locations of the primary vertices of the event, which are then used to fit the full tracks.

The first step in the process is seed generation. The seeds are fit to define the initial

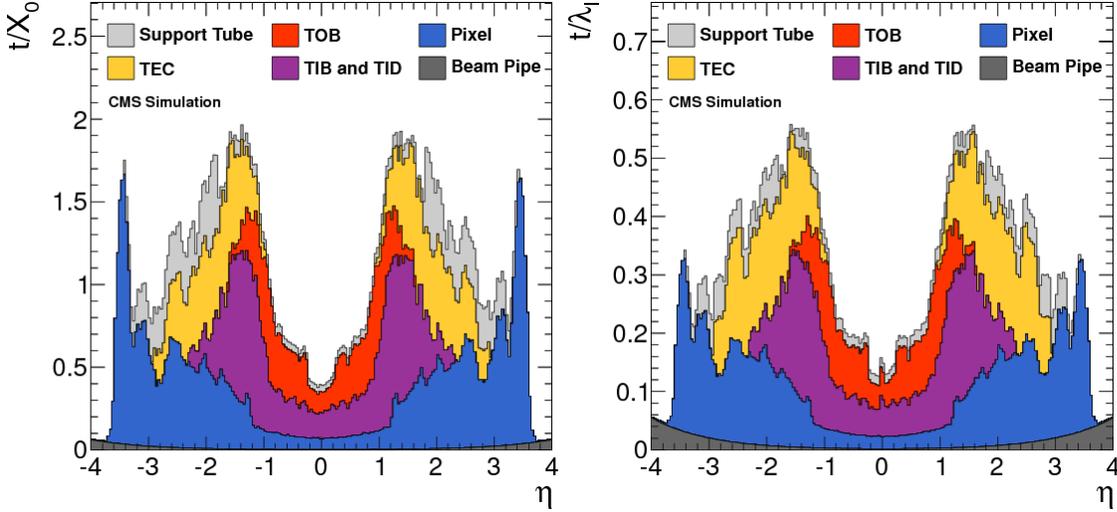


Figure 3.7: The tracker material budget as a function of η in units of the radiation length (left) and the hadronic interaction length (right). From [37]

track parameters and trajectory seed. The trajectory is extrapolated to subsequent layers of the detector using the Kalman filter method [38,39] with additional hits from the subsequent layers added and the track parameters added at each layer. Each layer provides information about the position and uncertainty of each hit and the amount of material crossed, which is used to account for multiple Coulomb scattering and energy loss. This procedure continues until the edge of the tracker is reached or there are no more valid hits. Once all the hits are identified, the entire track is refit to obtain the best estimate of the track parameters.

Iteration 0 of the CTF algorithm uses pixel triplets as seeds with a constraint on the fit requiring $p_T > 0.8$ GeV. This is designed to find high- p_T tracks originating from the interaction region (called prompt tracks). Iteration 1 requires two hits either in the pixels or in the two inner rings of the three inner TEC layers (used to increase coverage in the high η region) in addition to a pixel vertex. This vertex is required to pass quality criteria, the most important of which is that it be reconstructed from at least 4 pixel tracks. In this iteration $p_T > 0.6$ GeV is required for the track fit. Iteration 2 is searching for low- p_T prompt tracks, and so requires a pixel triplet seed but the requirement is $p_T > 0.075$ GeV. Iteration 3 requires mixed seeds in the pixels and strips, which allows for slightly displaced vertices. Iterations 4-5 require seeds in

the endcaps and barrels of the strips to find greatly displaced vertices.

Once all the 6 iterations have run, the track collection is filtered to remove likely fakes. Tracks are rejected if the reduced χ^2 of the track fit to the hits is too large, if there are too few hits, if the track is too incompatible with a pixel vertex, or if the track is too incompatible with the beamspot. The surviving *High Purity* tracks are considered measurements of the motion of charged particles through the tracker as well as to reconstruct the primary vertices in the event.

3.2.2.2 Primary Vertex Reconstruction

The locations of the multiple proton proton interactions within each bunch crossing (see Section 3.1.1) are determined by the primary vertex (PV) reconstruction. This determines the location and uncertainty of all interactions using the reconstructed tracks. High purity tracks that are consistent with the beamspot are clustered based on the z coordinate of the point of closest approach to the beamspot using a deterministic annealing algorithm [37,40]. Each track is assigned to exactly one candidate vertex, with the total number of candidates determined by the algorithm. The candidates with at least two tracks assigned to them are then fit with an adaptive vertex fitter [41] to determine the 3D position and other parameters of the vertex.

The performance of the PV reconstruction algorithm is illustrated in Figure 3.8. For high track multiplicities, the resolution is typically less than 20 μm in x and less than 25 μm in z for general inelastic collisions. The efficiency to reconstruct a vertex is also measured to be $> 99.75\%$ for vertices with more than 2 tracks (and $> 98\%$ for vertices with 2 tracks). This excellent resolution and efficiency allows most of the individual vertices in an event to be reconstructed distinctly even within the high pileup regime of the LHC.

3.2.3 The CMS Electromagnetic Calorimeter

The CMS Electromagnetic Calorimeter (ECAL) is designed to provide excellent energy resolution for electrons and photons produced by the collisions. It is a crystal

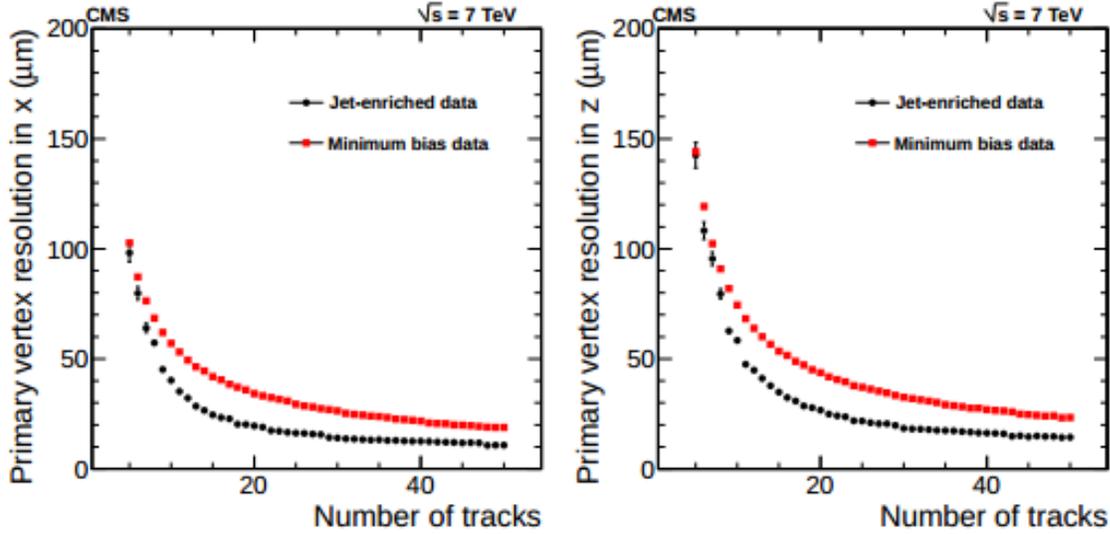


Figure 3.8: The x (left) and z (right) resolution of the CMS PV reconstruction algorithm as a function of the number of tracks. The resolution in y is nearly identical to the resolution in x. The minimum bias sample (red) estimates the resolution for general inelastic pp collisions by requiring only the presence of the a bunch crossing in the detector. The jet-enriched (black) sample estimates the resolution for collisions producing more transverse energy by requiring the presence of a 20 GeV jet (see Section 4.3). Figure taken from [37].

calorimeter made of 75,848 lead tungstate (PbWO_4) crystals split with 61200 crystals in the ECAL Barrel (EB) ($|\eta| < 1.48$) and 7324 crystals in each ECAL Endcap (EE) ($1.48 < |\eta| < 3.0$) [42]. A silicon lead pre-shower detector (ES) is installed in the region $1.65 < |\eta| < 2.6$ designed to improve the discrimination power between photons and π^0 particles. A schematic of the layout can be seen in Figure 3.9. The preshower consists of two lead radiators each followed by a layer of silicon strip detectors. The first is 2 radiation lengths thick and the second is 1 radiation length; these absorbers are designed to initiate showers from incoming electromagnetic particles, which are then detected by the silicon detectors [43].

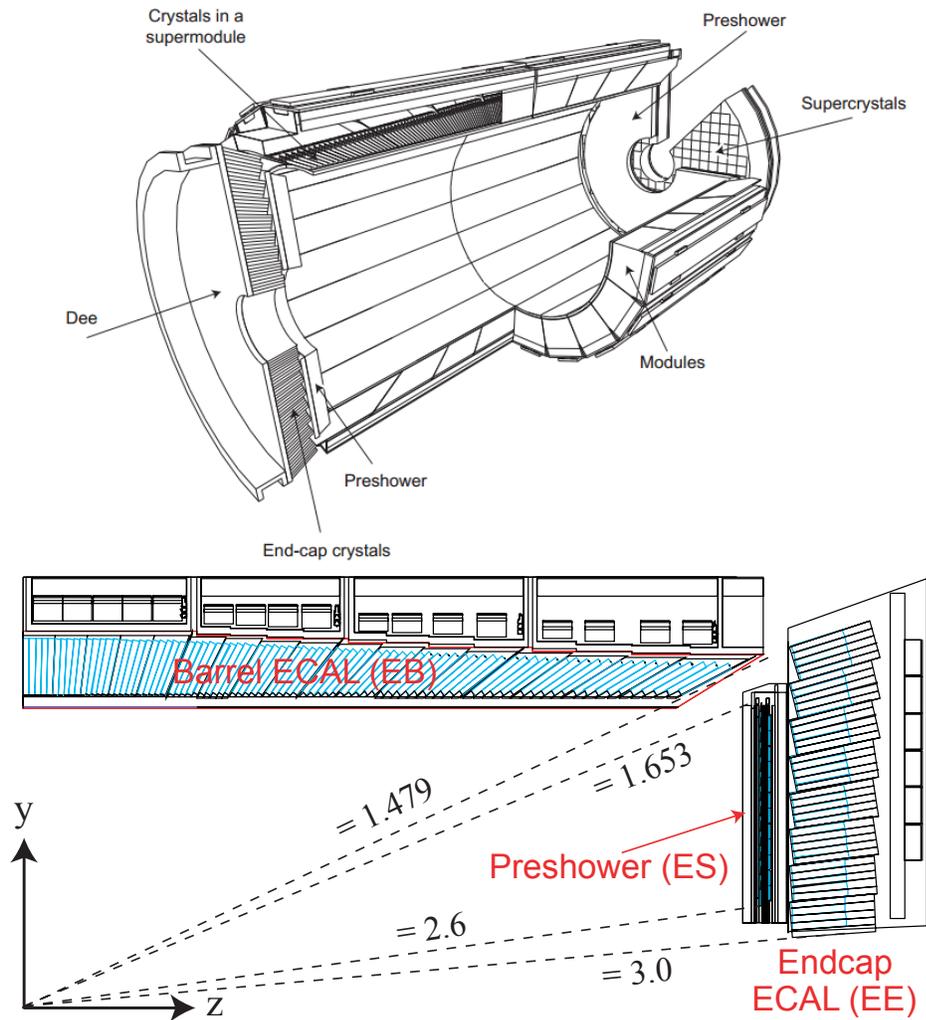


Figure 3.9: a schematic of the CMS ECAL layout. At top is a cut-away schematic of the whole detector showing the barrel, endcaps, and preshower subdetectors. A schematic showing the η coverage of the various subdetectors is shown at bottom.

The compact, high granularity design of the ECAL is made possible by the characteristics of the crystal material. Lead tungstate has a high density (8.28 g/cm³), short radiation length ($X_0 = 0.89$ cm), and small Molière radius (2.2 cm). These characteristics mean that the showers from incident high energy electrons and photons have small transverse spread and can be contained within relatively short crystals. In EB, the crystals have a truncated pyramid shape with a front face measuring 2.2 cm \times 2.2 cm and are 23 cm in length. In EE, the crystals are 22 cm long and their front faces measure 2.86 cm \times 2.86 cm. The crystals in the endcap are arranged in a projective geometry, pointing toward the center of the interaction region.

Photons and electrons passing through the ECAL crystals lose energy to the crystal material, which produces scintillation light. PbWO₄ crystals emit scintillation light in a broad spectrum with a maximum between 420-430 nm. The scintillation light is measured by avalanche photodiodes (APDs) [44,45] and vacuum phototriodes (VPTs) [46] in the EB and EE respectively. The quantum efficiency and cross-section of the detectors are such that an incident particle depositing 1 MeV of energy in the crystal will produce an average of 4.5 detected photoelectrons in either EB or EE.

The energy resolution of the ECAL is measured in test-beams as a function of the energy of the incident particles [47]. In the barrel, this resolution is measured to be:

$$\frac{\sigma_E}{E} = \frac{2.83\%}{\sqrt{E/(1\text{GeV})}} \oplus \frac{12.4\%}{E/(1\text{GeV})} \oplus 0.26\% \quad (3.2)$$

The resolution *in situ* will depend on additional factors, including the material budget in front of the crystal.

3.2.3.1 ECAL Clusters

The energy deposited in each ECAL crystal during the collision is measured by the APD or VPT attached to the crystal. Since the width of the ECAL crystals is roughly one Molière radius of lead tungstate, we expect the deposits left by real electrons and photons passing through the ECAL to be several crystals wide. We start by forming basic clusters (BCs), which are 5 \times 5 groups of crystals, centered around local maxima

in the energy deposited in the ECAL [42]. In test beam measurements, it has been shown that a cluster of this size will contain 97% of the energy deposited by a high energy photon or electron [33, 47]. *In situ*, however, the picture is more complicated: the presence of the tracker in front of the ECAL and the strong magnetic field can cause electrons to bremsstrahlung, depositing energy over a large region in the ϕ direction. Similarly, it can cause photons to convert to electron-positron pairs, which can then deposit energy over much larger areas than photons that do not convert.

To accurately reconstruct the energy with which the particle was produced in the collision, the energy lost to bremsstrahlung must be clustered and recovered as well. This is done by clustering the basic clusters into superclusters (SCs), which are designed to contain all of the energy deposited in the ECAL by the particle in question. The process of forming these clusters in the EB is called the “hybrid” supercluster algorithm and is described extensively in [33, 48]. It starts by defining a seed crystal with high energy and then dynamically searches in the ϕ direction for more strips of high energy deposits that are more consistent with radiated energy than a separate prompt particle. The extended area is then reclustered in the ϕ direction into new basic clusters. A schematic view is shown in Figure 3.10. In the endcap the procedure is different due primarily to the presence of the preshower detector. The preshower is roughly 3 radiation lengths thick, which degrades the resolution of EE by a factor of roughly 2 compared to a similar particle incident on EB. A simpler algorithm consisting of clustering distinct basic clusters in the ϕ direction is used in EE that also merges in energy information from ES.

While the supercluster is an extended object within the ECAL, it is meant to encapsulate the energy deposited by a particle that had momentum in a specific direction. To estimate this direction, it is important to define the position of the supercluster as a distinct point. The position is defined as a weighted sum of the positions of the individual hits making up the SC. Let $\{\vec{x}_i | 1 \leq i \leq N\}$ be the 2D positions of the energy deposits making up the SC in the ECAL, with each point \vec{x}_i

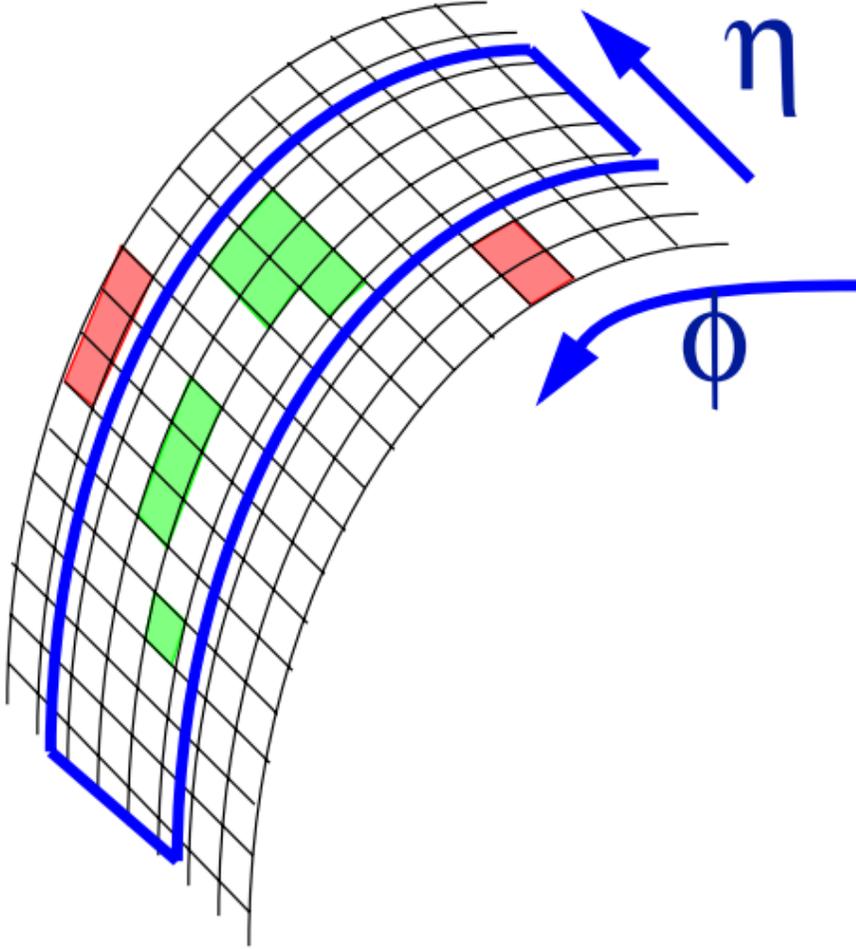


Figure 3.10: An illustration of the hybrid supercluster algorithm. Green squares indicate clustered energy while red squares indicate unclustered energy. Figure from [48]

having energy E_i , then the SC position is defined as

$$\vec{x}_{\text{SC}} = \frac{\sum_{i=1}^N w_i \vec{x}_i}{\sum_{i=1}^N w_i}$$

where the weights are defined

$$w_i = \max \left(0, 4.2 + \log \left(\frac{E_i}{\sum_{i=1}^N E_i} \right) \right)$$

The value of 4.2 in the above equation is found by optimization.

3.2.4 The CMS Hadron Calorimeter

The purpose of the CMS Hadron Calorimeter (HCAL) is to measure the energy of charged and neutral hadrons produced by the collision. When a quark or gluon is produced in the final state of a collision, it will quickly hadronize into jets of hadrons, the energies and directions of all of which must be measured to reconstruct the energy and direction of the initial particle. To measure as many of these particles as possible, the HCAL provides many interaction lengths of material to stop particles and coverage out to large $|\eta|$.

The HCAL is composed of three subdetectors: the HCAL barrel (HB) covering $|\eta| < 1.4$, the HCAL endcaps (HE) covering $1.3 < |\eta| < 3.0$, and the forward HCAL (HF) covering $3.0 < |\eta| < 5.2$ [33, 49]. The HB is situated between the ECAL barrel and the solenoid covering a radius of $1.78 \text{ m} < r < 2.88 \text{ m}$ from the centerline of the detector. The HE are placed on either side of the interaction region behind the ECAL endcaps. The HF is placed around the beampipe on both sides of the detector at $z = \pm 11.2 \text{ m}$ from the interaction point, outside the endcap return yokes. A schematic representation of the layout can be seen in Figure 3.11. There is also an outer hadron calorimeter (HO) placed outside the solenoid, but this was not used for data collected in 2012.

The HB is composed of 36 identical wedges, each covering 20° in ϕ . Each wedge is composed of 17 layers of plastic scintillators alternating with 16 layers of brass absorber plates (except the first and last plates, which are stainless steel for structural support). The first and last layers of scintillators are 9 mm thick and the other 15 layers are 3.7 mm. The inner stainless steel plate is 61 mm thick and the outer one is

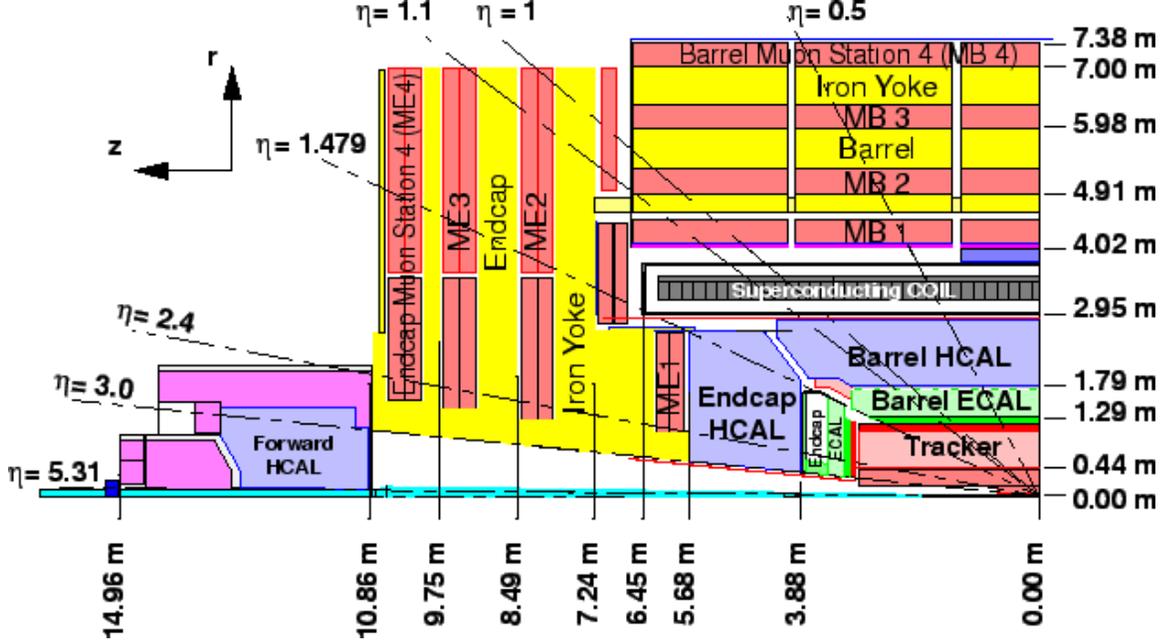


Figure 3.11: Schematic of the CMS detector showing the separation of subdetectors into barrel and endcap regions as well as the HF.

75 mm thick. The first 8 layers of brass plates are 50.5 mm thick and the remaining 6 are 56.5 mm thick. Each layer of scintillator is composed of multiple tiles that are machined to cover a region $\Delta\eta \times \Delta\phi = 0.087 \times 0.087(5^\circ)$, forming 32 HB “towers” in the η direction and 144 towers in the ϕ direction. The scintillators are read out by wavelength shifting fibers, each of which is read out by a multipixel hybrid photodiode (HPD) with a gain of ≈ 2000 .

The HE is tapered and overlaps with the last tower of the HB. The HE is also divided into 20° slices in ϕ matching the HB. There are 19 active plastic scintillators interleaved with absorbers, all of which are brass. All the absorbers are 78 mm thick and all the scintillators are 3.7 mm. The towers cover the same range in $\Delta\eta \times \Delta\phi$ as in HB up to $|\eta| = 1.74$, where the $\Delta\phi$ size is increased to 0.174 (10°) to accommodate the bending radius of the fibers and the $\Delta\eta$ size increases as shown in table 3.1. The scintillators in HE are also read out by HPDs with similar gain to those used in HB.

The forward calorimeter is meant to capture particles with large momentum parallel to the beam direction that end up in the very forward $3.0 < |\eta| < 5.0$ region. The extreme radiation flux coming from the LHC makes this a very challenging en-

$ \eta $ range	$\Delta\eta$ size
1.740 - 1.830	0.090
1.830 - 1.930	0.100
1.930 - 2.043	0.113
2.043 - 2.172	0.129
2.172 - 2.322	0.150
2.322 - 2.500	0.178
2.500 - 2.650	0.150
2.650 - 3.000	0.350

Table 3.1: The $\Delta\eta$ size of the plastic scintillators in HE as a function of η .

environment and requires the HF to be extremely radiation hard. At design energy, the inner edge of the HF (at $|\eta| = 5.0$) would expect to receive a radiation dose of $2 \times 10^5 \text{ Gy}/fb^{-1}$ compared to the expected dose in the inner edge of HE (at $|\eta| = 3.0$) of $20 \text{ Gy}/fb^{-1}$ [49]. This extreme difference in dose necessitates a very different design for the HF detector compared to HB and HE.

The HF is composed of radiation hard quartz fibers embedded in steel absorbers. The fibers are split between long (1.65 m) and short (1.43 m) fibers arranged alternating in the absorber with a separation of 5 mm creating effectively 2 longitudinal sampling points within the absorber. The signal is produced by Cherenkov light in the fibers. The fibers run parallel to the beamline and are bundled at the back of the detector and routed to phototubes for readout. The HF is split into 13 towers; 12 have $\Delta\eta \approx 0.175$ and one (the one at highest $|\eta|$) has $\Delta\eta = 0.3$. The towers have $\Delta\phi = 0.174$ up until $|\eta| > 4.716$ where the ϕ segmentation increases to $\Delta\phi = 0.348$.

3.2.5 The CMS Muon System

The CMS Muon system surrounds the solenoid and is responsible for the identification and accurate measurement of muons passing through the CMS detector [50]. Muons produced by a collision in CMS behave similarly to electrons, in that they lose energy as they pass through matter due to bremsstrahlung radiation, but being ≈ 200 times heavier than electrons, they lose much less energy (by a factor that goes as $\frac{1}{m^2}$). This means that, while a $\mathcal{O}(10 \text{ GeV})$ electron is fully stopped in the ECAL, a similar energy muon passes through the tracker, both calorimeters and the solenoid. Unlike

a neutrino, which also passes through all of these systems, a muon deposits small amounts of energy in each one, meaning that one can detect the presence of a muon by looking for small energy deposits in multiple subsystems traveling out of the detector. The muon system, then, measures the deposits of energy left by the muon after it travels out of the central region of the detector.

The muon system is composed of three different types of subdetectors occupying different regions in $|\eta|$. Since the magnetic field is no longer homogeneous in the region outside the solenoid (see Figure 3.4) each subdetector is optimized to deal with the different magnetic field regimes in which they are situated as well as the different radiation environments. In the central region $|\eta| < 1.2$ the muon system uses drift tube chambers (DTs), which provide excellent spatial resolution. Forward from the DTs from $0.9 < |\eta| < 2.4$, the muon system uses cathode strip chambers (CSCs), which sacrifice some spatial resolution for increased radiation resistance and less susceptibility to the more intense gradient of the magnetic field. Resistive plate chambers (RPCs) are used in front of the DTs and CSCs in the region $|\eta| < 1.6$ to help with triggering by providing good temporal resolution.

The DTs are composed of 250 individual drift chambers each of which is composed of drift cells with a ≈ 400 ns maximum drift time. These chambers are arranged in 4 concentric rings around the beamline at radii of approximately 4.0 m, 4.9 m, 6.0 m, and 7.0 m. They are segmented into 5 distinct slices in the z-direction, each approximately 2.7 m in length. In the ϕ direction, the chambers increase in size with increasing r to ensure that each one covers 30° in the ϕ direction. The chambers are staggered, as shown in the bottom right of Figure 3.12 to provide 360° ϕ coverage.

The CSCs consist of 234 chambers in each endcap region. The individual chambers are staggered to provide at least 3 layers in the entire range $0.9 < \eta < 2.4$, given the constraints of the positioning of the solenoid and other subdetectors. The layout can be seen exactly in Figure 3.12.

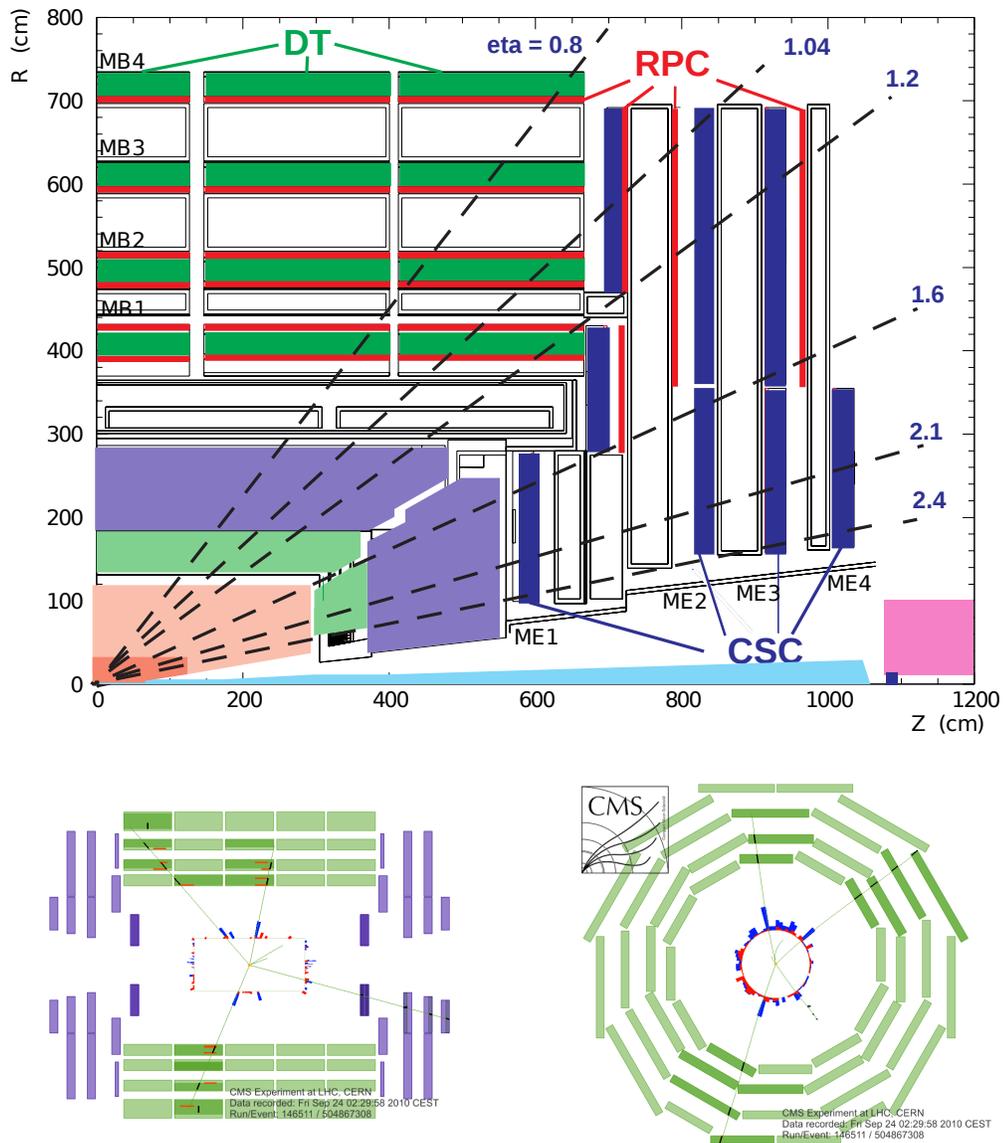


Figure 3.12: A schematic layout of the CMS muon system. The top plot shows r - z view of one quarter of the system with the DTs, CSCs, and RPCs labeled as well as lines corresponding to the $|\eta|$ within the detector. The bottom plots show the r - z (left) and r - ϕ (right) layout of the whole system as well as the readout data from a single $\sqrt{s} = 7$ TeV collision; the green colored channels are the DTs and the blue color are the CSCs (RPCs not shown). Both plots from [51].

The muon chambers measure the hit positions of muons as they are bent by the magnetic field channeled by the iron return yokes. The hits in the muon chambers can be fitted by a track (in a manner similar to the tracks in the tracker) and the momentum of the muon can be measured by the curvature of the track. The muon momentum resolution using only these hits can be see in the left side of Figure 3.13. One sees that the resolution becomes considerably worse as one goes to higher $|\eta|$ reflecting the lower sensitivity of the CSCs due to the harsher radiation environment. Since muons also create tracks in the tracker, one can do a combined fit to the hits in the tracker and those in the muon chamber to get a better estimate of the momentum of the muon. The resolution using this fit method (with two different implementations) is shown in the right side of Figure 3.13. One sees that, with the full fit, CMS obtains 1-2% resolution for muons in the central region < 6% resolution in the whole region $|\eta| < 2.1$.

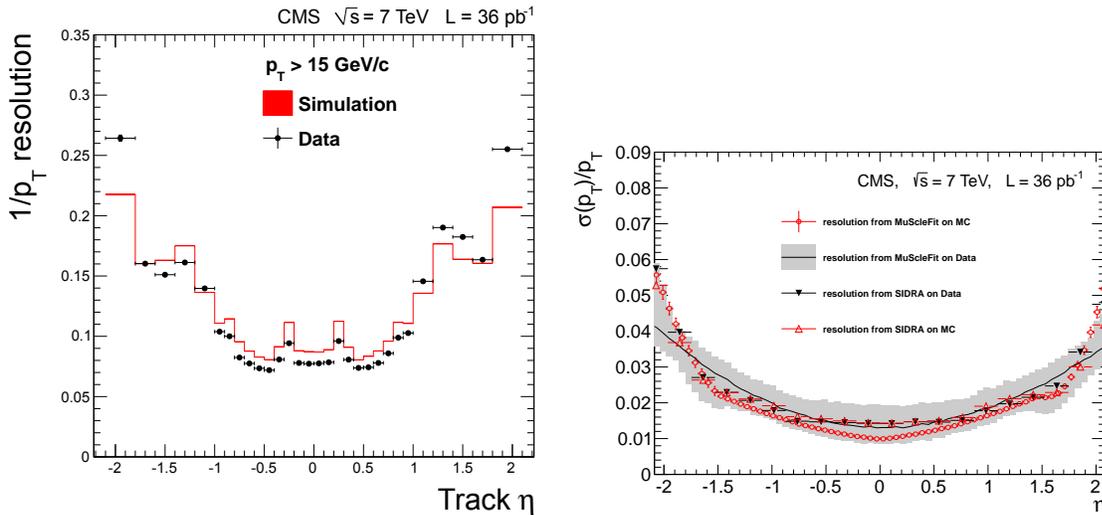


Figure 3.13: The resolution of the muon system using only the hits in the muon chambers (left) and using a combined fit to the hits in the muon system and the hits in the tracker (right) using two different fit methods. Plots taken from [51].

Chapter 4

Object Reconstruction and Identification

The goal of the analysis is to search for anomalous production of events with a $H \rightarrow \gamma\gamma$ and at least one other particle in the final state . For this we require a robust method for identifying photons and discriminating them from electrons. We also require a method for reconstructing other particles, which accurately determines the kinematics of the particle, but remains agnostic to the actual type of particle. Finally, we require a quantity that measures some of the kinematics of particles that escape the detector.

A particle produced in a collision in CMS will traverse multiple subdetectors and leave deposits in some or all of them. The shape of these deposits, which systems they are in, and their interaction with the magnetic field gives information about what type of particle was produced and what its energy and direction were. We have discussed in Section 3.2 how the raw data recorded by individual subsystems is reconstructed into local information about the particles movements through the system. We focus now on the approach of combining this information across subsystems to reconstruct the global information about the particle's kinematics.

While there are typically many objects produced in a collision, we will focus mostly on objects with large momentum perpendicular to the beamline. We measure this by looking at a particle's transverse energy, defined as:

$$E_T \equiv E/\cosh(\eta)$$

we will also be concerned with a particles transverse momentum, which is defined simply as the x-y component of the momentum:

$$\vec{p}_T = \vec{p} - \vec{p} \cdot \hat{z}$$

We care primarily about these quantities because energy transverse to the beamline is indicative of the energy produced only from the interaction process while energy parallel to the beamline has a contribution from the \sqrt{s} of the interacting partons.

4.1 Particle Flow

The goal of particle flow (pf) is to identify all stable particles produced by the event as they pass through the detector by combining information from multiple subdetectors [52, 53]. The goal is to end up with a list of candidate particles each with the best estimate of its energy, direction, and ID produced from the reconstructed tracks, calorimeter clusters and muon tracks in the event. The key then is to have a global way of linking reconstructed elements in the 4 systems as described here. Each link between objects in different subdetectors is assigned a score quantifying its quality (exact methods are discussed below). The algorithm will then produce blocks of linked objects that can be reconstructed.

It should be noted that calorimeter clusters for particle flow are formed in slightly different ways from those described in Section 3.2. The clusters are seeded by local maxima in the deposited energy and are then grown from these seeds by including any cell adjacent to the cluster with energy above a certain threshold (80 MeV in EB, 300 MeV in EE and 800 MeV in HCAL). These “topological” clusters are centered at their energy weighted geometric center in $\eta - \phi$ and are used in place of the traditional clustering methods described before. Energy lost to radiation is recovered in other ways, so there is no need to try to cluster all the radiated energy, which was the goal of the supercluster algorithm.

The first step is to link charged tracks with calorimeter clusters. This is done by

taking each reconstructed high-quality track consistent with a primary vertex and extrapolating it into the calorimeters to the following depths:

1. a depth in the ECAL corresponding to the expected maximum shower profile *or* the 2 layers of the ES
2. one interaction length into the HCAL

A track is then linked to any cluster in either (or both) of those detectors that overlaps with the extrapolated track. The link is assigned a score based on the $\eta - \phi$ distance between the center of the cluster and the extrapolated track. In order to capture any Bremsstrahlung photons produced by the charged particle, tangents along the curved track are extrapolated linearly to the ECAL. Any clusters overlapping these extrapolated tangents are linked with a link score determined by the distance from the center of the cluster to the extrapolated tangent.

Next calorimeter clusters are linked between the ES, ECAL, and HCAL. These calorimeters have different granularities, with ES being the most granular and HCAL being the least. A cluster in a more granular calorimeter is linked to a cluster in a less granular calorimeter if its center is within the cluster in the less granular detector. For instance, an ECAL cluster would link to an HCAL cluster if its center were within the HCAL tower. The link distance is defined as the $\eta - \phi$ distance between the centers of the clusters.

The last step is linking charged tracks from the inner tracker to muon tracks in the outer tracker. This is done when a global fit between an inner track and a muon track returns a reasonable χ^2 . Since the granularity of the inner tracker allows multiple charged tracks to fit one muon track, any ambiguity is resolved by selecting the track which produces the best χ^2 [32]. The χ^2 of the fit is the link quality.

Blocks are then formed so that all objects within a block are linked and there are no links between blocks. The largest possible block would be several tracks, several ES and ECAL clusters, one HCAL cluster and one global muon track, but most clusters will contain only a subset of these. Each block is then analyzed to check for candidate pf particles. The algorithm runs until the pool of blocks is exhausted.

If a block contains a muon track linked with a track from the inner tracker, it can form a pf muon. This happens when the momentum estimated by the inner track is within 3σ of the momentum from the combined fit. If this is the case, the inner track and muon track are removed from the block. If, after the deletion, the block contains only a single object or no object it is deleted, otherwise it is reinserted into the block pool. This helps address the case of a muon overlapping with another particle.

A block containing at least one linked track and an ECAL cluster may produce a pf electron [54]. Each track linked to a cluster goes through an identification and refit procedure described in [37] and [54] to check if it is compatible with the way in which an electron should lose energy in the tracker (essentially exploiting the entire tracker as a pre-shower detector). A multivariate approach is used to determine whether the re-fitted track and cluster are compatible with an electron. If they are determined to be compatible, a pf electron is formed, and the linked track and cluster are removed from the block while the block is re-entered into the block pool (if non-empty).

For the remainder of the algorithm, tracks with relative uncertainty on p_T larger than the (measured) expected uncertainty on calorimeter clusters are ignored. A block with a valid track linked to one or more calorimeter clusters, where the sum of the energy of the clusters is compatible with the track momentum within uncertainties, forms a pf charged hadron with the momentum determined by fitting the track momentum and the calorimeter deposits. Any remaining tracks in the block give rise to additional pf-charged hadrons with momentum and energy taken from the track, with the mass assumed to be the pion mass. Blocks with a track linked to calorimeter clusters, where the track momentum is significantly bigger than the calorimeter energy, are re-checked for muons using relaxed criteria and allowing tracker-only muons (muons without fitted tracks in the muon chambers).

Blocks with a track linked to a cluster with higher energy than the track are treated as either a neutral hadron or a pf photon. If the ΔE between the track and the sum of the clusters is greater than the ECAL energy, then a pf photon is created with the ECAL energy and a pf neutral hadron is created with the remaining $\Delta E - E_{ECAL}$, otherwise a pf photon is created with ΔE . In either case, the energy of these neutral

particles is removed from the clusters and the block is re-entered into the pool (by construction, the energies of the track and the cluster are now compatible, so it will give rise to a pf-charged candidate).

Blocks containing only linked ECAL HCAL clusters are reconstructed as neutral hadrons. Finally, unlinked ECAL and HCAL clusters are clustered into pf photons and pf neutral hadrons, respectively. The energies of the particles are determined solely by the calorimeter deposits.

The pf algorithm is very good at finding many of the particles in an event with lots of energy and tracks, where a traditional detector-by-detector algorithm might struggle. We have found that the algorithm has some trouble discriminating between photons and neutral hadrons. For most cases, this is not a source of problems, but for analysis specifically trying to tag $H \rightarrow \gamma\gamma$ within an event, it is. We therefore use a different reconstruction to find and reconstruct photons, based on a more ECAL-driven approach.

4.2 Photons

A photon is a neutral electromagnetic object, which should deposit virtually all of its energy in the ECAL. Since it has no charge, it will not bend in the magnetic field of the detector and will not leave a track in the tracker. Since the ECAL crystals are ≈ 26 radiation lengths in EB and ≈ 22 radiation lengths in EE, the photon should leave most or all of its energy in the ECAL, with very little energy reaching the HCAL. Our signature for a photon is, therefore, a cluster of energy in the ECAL which is not consistent with the trajectory of any track in the tracker and which has a small energy in the HCAL towers behind the ECAL clusters. Every ECAL supercluster is a photon candidate; further information about the shape of the cluster and energy deposits around the supercluster is used to discriminate between prompt photons produced by the collision and other objects or non-prompt photons.

4.2.1 Photon Identification

The key sources of superclusters not originating from real prompt photons are energy deposited directly by hadrons, clusters produced by hadrons decaying to photons (e.g. $\pi^0 \rightarrow \gamma\gamma$), and superclusters from electrons. In order to veto superclusters from these three effects we perform photon identification (vetoing clusters not consistent with SCs from photons), photon isolation (vetoing clusters surrounded by other energy), and electron veto (vetoing superclusters more consistent with electrons). We find that it is extremely difficult to reliably identify photons with low p_T , so we only consider clusters with energy consistent with $p_T > 20$ GeV [55, 56].

Photon identification is concerned with making sure that the supercluster originated from a real electromagnetic object. We check that the shape of the SC is consistent with the deposits usually left by EM objects using a variable $\sigma_{i\eta i\eta}$, which is an energy weighted width of the cluster in the η direction. While electrons or converted photons may bend in the magnetic field and deposit energy in a relatively broad range of ϕ , they usually produce a narrow cluster in η , so this variable is useful for both electrons, photons, and converted photons. The variable is defined as:

$$\sigma_{i\eta i\eta}^2 \equiv \frac{\sum_i i w_i (\eta_i - \eta_{seed})^2}{\sum_i w_i} \quad (4.1)$$

The index i runs over the ECAL crystals in a 5×5 matrix centered on the most energetic crystal. Here the weights are given by

$$w_i \equiv \max\left(0, 4.7 + \ln\left(\frac{E_i}{E_{5 \times 5}}\right)\right) \quad (4.2)$$

where η_i represents the pseudorapidity of the center of the i^{th} crystal (in units of crystals) while E_i represents its energy. η_{seed} represents the energy of the seeding crystal of the supercluster and $E_{5 \times 5}$ is the energy of a 5×5 crystal matrix around the seeding crystal. The value of 4.7 in the weight equation is set so the sum in equation 4.1 only considers crystals with energy $E_i/E_{5 \times 5} > 0.9\% = e^{-4.7}$. This helps protect $\sigma_{i\eta i\eta}$ from being sensitive to background energy. The logarithmic dependence

on the energy makes the variable more sensitive to relatively small energy deposits on the outside of the matrix, which improves its usefulness since photons tend to be very sharply peaked. One can see, then that clusters with large $\sigma_{i\eta i\eta}$ will tend to have wide distributions in η , while those with smaller values will have narrower distributions.

The distributions of $\sigma_{i\eta i\eta}$ for real photons and for clusters from other objects in MC are shown in Figure 4.1 for clusters in the barrel and endcap. We see that the real photons have characteristically wider showers (larger $\sigma_{i\eta i\eta}$) in EE than in EB. This behavior is expected, since the tracker material budget (figure 3.7) and the presence of the preshower detector mean there is considerably more material between the interaction point and EE than between the interaction point and EB (about 5 radiation lengths compared to 1). One sees that clusters originating from a prompt, high p_T ($p_T > 25$ GeV) photon have quite narrow distribution in $\sigma_{i\eta i\eta}$ compared to clusters originating from other effects. Here “fake” photons are composed of anything that is not a prompt photon, so this includes hadronic deposits, electrons, and photons from secondary sources. We can further identify clusters from e/γ compared

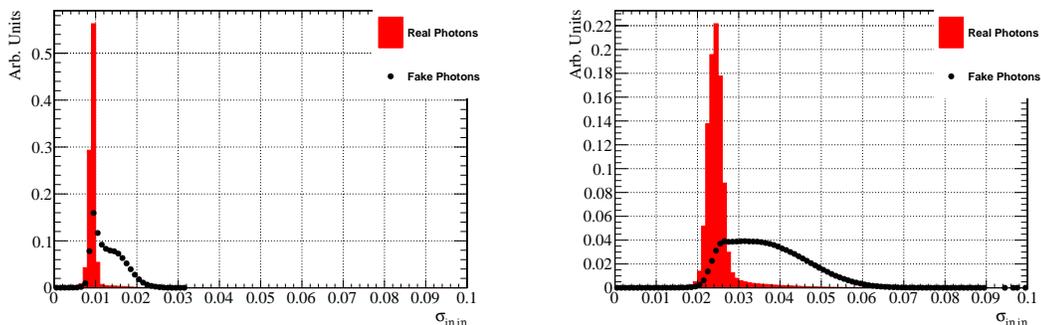


Figure 4.1: Distribution of $\sigma_{i\eta i\eta}$ in MC simulation for clusters matched to a generated photon (red) and those not matched to a generated photon (black). The distributions are for the EB (left) and EE (right). All distributions are normalized to unit area to see shape differences.

to clusters of hadronic origin by looking at the relative amount of energy left in the ECAL versus the portion of the HCAL directly behind the cluster. Since the ECAL crystals are 25-26 radiation lengths deep, we expect most of the energy to be deposited in the ECAL and very little to punch through and end up in the HCAL. To measure this, we use the ratio H/E , which is the energy in the HCAL tower at the

same η - ϕ position as the most energetic crystal of the SC divided by the energy of the supercluster. For real photons, we would expect this ratio to be small or zero.

The distributions for H/E for clusters from prompt photons and those not from prompt photons are shown in Figure 4.2. One can see that both distributions have a large peak at 0, where there is no energy in the HCAL tower behind the cluster. The dip in the next bin is due to zero suppression of the HCAL readout (measured energies below a certain value are treated as zero, so very small non-zero H/E values are suppressed), but for most of the distribution with $H/E > 0.05$ we have an order of magnitude more fake photons than real photons. These variables are effective at

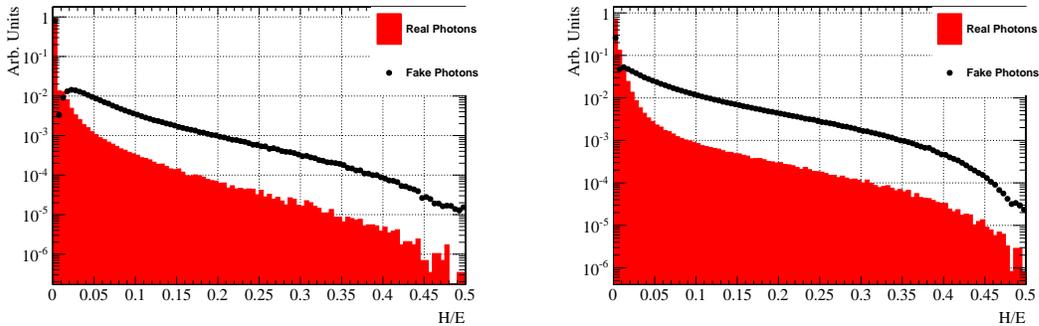


Figure 4.2: Distribution of H/E in MC simulation for clusters matched to a generated photon (red) and those not matched to a generated photon (black). The distributions are for the EB (left) and EE (right). All distributions are normalized to unit area to see shape differences.

rejecting deposits in the ECAL not due to real photons, but they can do very little if real photons are produced by a non-prompt (not originating from the primary interaction) process. The hadronization process of a quark or gluon produced in the event can occasionally produce a high energy π^0 or η^0 , which can decay to high-energy collimated photon pairs. The signal from these will look very similar or identical to a prompt photon in $\sigma_{i\eta i\eta}$ and H/E so these variables are not effective in rejecting these processes without rejecting substantial numbers of real photons as well. To overcome this, we exploit the fact that hadronization usually produces large numbers of energetic hadrons that interact with the tracker, ECAL and HCAL, so large amounts of energy deposited in the detector in an annulus around the cluster of interest would indicate that the cluster is from a non-prompt photon of this sort. We call clusters

that do not have large energy deposits around them *isolated*, while those that do are *non-isolated*.

We use the candidate objects reconstructed by the particle flow algorithms to calculate the isolation of the cluster. In all cases we consider only particles within the *isolation cone*: $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$, where $\Delta\eta \equiv |\eta_{cluster} - \eta_{pf\ candidate}|$ and $\Delta\phi \equiv |\phi_{cluster} - \phi_{pf\ candidate}|$. This is broken down into 3 distinct flavors of isolation:

Charged Hadron Isolation: We sum the energy of all pf charged candidates consistent with the primary vertex within the isolation cone and with $\Delta R > 0.02$. This ΔR cut is to prevent rejecting the event if the main cluster is actually an electron, which would leave a track in the tracker and thus be could be included as a pf charged candidate (we will reject electrons with a separate step). By “consistent with the primary vertex” we mean that the reconstructed position of the candidate’s vertex satisfies $|z_{cand} - z_{PV}| < 0.2$ cm and $|\sqrt{x_{cand}^2 - x_{PV}^2} + \sqrt{y_{cand}^2 - y_{PV}^2}| < 0.1$ cm. This requirement is to help reduce the effect of pileup, which is discussed further below.

Neutral Hadron Isolation: We sum the energy of all pf neutral candidates in the isolation cone.

pf Photon Isolation: We sum the energy of all pf photons within isolation cone and with $\Delta\eta > 0.015$ (for clusters in EB) or $\Delta R > 0.07$ (for clusters in EE). We ignore pf photons that are too close to our candidate photon to avoid the complications of matching our candidate photon to a pf photon, which can be technically challenging. Since this small central region adds no additional rejection power for secondary photons, it is more straightforward to ignore it.

These isolation variables perform well when the vast majority of energy comes from the primary interaction, but when a substantial amount of energy is deposited in the detector from particles produced by pileup vertices, their efficiency to select real electrons or photons decreases. This effect is most pronounced for the neutral hadron and photon isolations, which are made of particles that leave no tracks in the tracker, since we cannot associate the pf-candidates to a vertex. The effect is smaller for the charged hadron isolation, since we require the tracks to be consistent with the

primary vertex, but since multiple vertices can occur within small Δz regions along the beamline we still get an effect from particles produced at secondary vertices near the primary vertex. For charged hadrons, each additional interaction in the event typically adds about 1% to the total energy sum.

To combat the influence of pileup on the isolation variables, we measure the average energy from pileup in the detector in the manner described in Section 4.3.1 and in [57] and use this to correct the variables. Let ρ be the average energy from pileup in the detector, we then correct each isolation by subtracting $\rho \times A_{eff}$, where A_{eff} is an $|\eta|$ dependent effective area for each isolation. The values of A_{eff} are shown in Figure 4.3. They are computed by measuring the average value of the isolation sum as a function of ρ in bins of $|\eta|$ and setting A_{eff} equal to the measured slope in each bin.

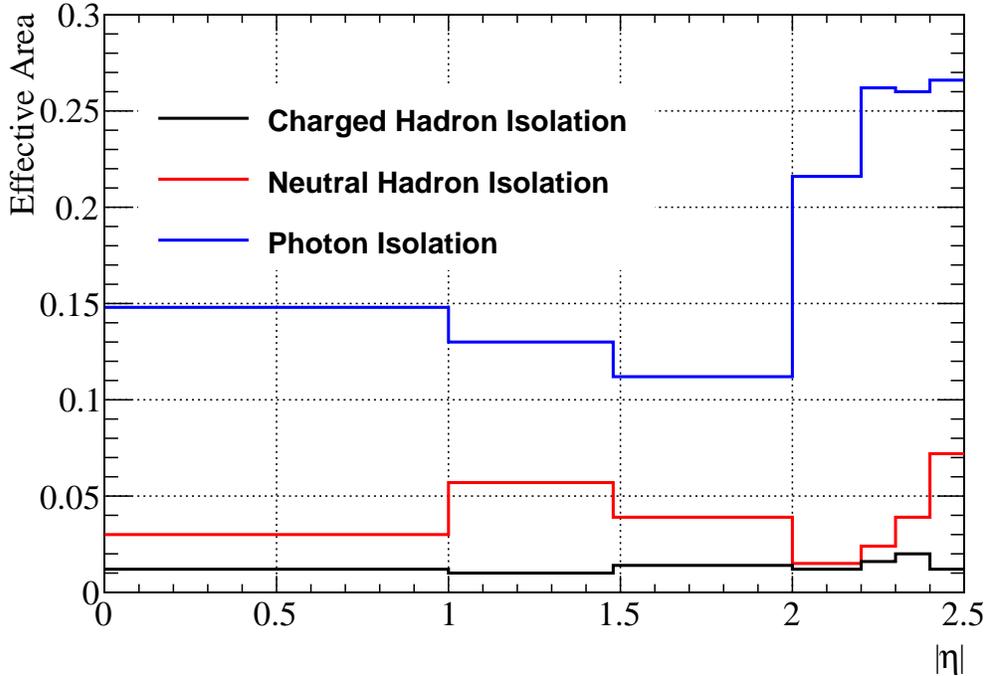


Figure 4.3: The effective area (A_{eff}) for the pileup correction to the isolation cones for the three different types of isolation as a function of $|\eta|$. These values come from the measured slope of the isolation sum in a sample of MC photons and jets.

The effect of the correction can be seen in Figure 4.4. We observe that the value of the isolation sums for the photon and neutral hadron isolation exhibit roughly linear

trends as a function of the number of vertices, while the charged hadron isolation exhibits a much smaller trend. After correction the trends are strongly suppressed, indicating we have successfully eliminated the contribution to the sum from particles produced by pileup vertices, while leaving the contribution from particles produced by the primary vertex.

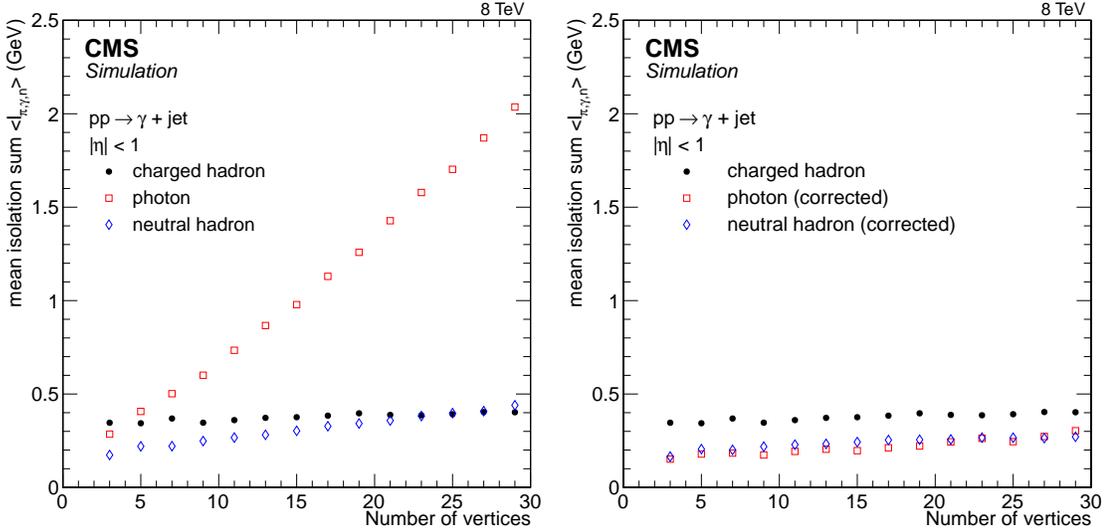


Figure 4.4: The mean isolation sum for photons with $p_T > 50$ GeV for $pp \rightarrow \gamma + j$ events as a function of the number of reconstructed vertices for the three different types of isolation. The left plot is without the ρ -correction and the right plot has the correction applied. Figure from [56].

Cutting on all of these variables together is effective at selecting real prompt photons and vetoing fakes. We define a series of working points at different photon selection efficiencies, maximizing the fake rejection efficiency at each point. Since some of the variables have different characteristics in the EB and EE, we define the cuts differently for the two regions. The values for the cuts are given in Tables 4.1 and 4.2 for the EB and EE, respectively.

The final element to the photon identification is to separate photons and electrons. This can be done by searching for a track consistent with an electron pointing to the cluster. There are two methods used to construct this veto: the *pixel veto* and the *conversion-safe electron veto*.

The pixel veto looks for any pixel track seed (see Section 3.2.2.1) that is consistent

ECAL Barrel	Loose (90%)	Medium (80%)	Tight (70%)
$\sigma_{i\eta i\eta}$	0.012	0.011	0.011
H/E	0.05	0.05	0.05
charged hadron isolation	2.6	1.5	0.7
neutral hadron isolation	$3.5 + 0.04p_T^\gamma$	$1.0 + 0.04p_T^\gamma$	$0.4 + 0.04p_T^\gamma$
pf photon isolation	$1.3 + 0.005p_T^\gamma$	$0.7 + 0.005p_T^\gamma$	$0.5 + 0.005p_T^\gamma$

Table 4.1: Variable cuts for three different working points for the photon ID and isolation in the Barrel. Variables are required to be less than the indicated value. The value in parenthesis is the target photon selection efficiency for the working point.

ECAL Endcap	Loose (85%)	Medium (75%)	Tight (65%)
$\sigma_{i\eta i\eta}$	0.034	0.033	0.031
H/E	0.05	0.05	0.05
charged hadron isolation	2.3	1.2	0.5
neutral hadron isolation	$2.9 + 0.04p_T^\gamma$	$1.5 + 0.04p_T^\gamma$	$1.5 + 0.04p_T^\gamma$
pf photon isolation	-	$1.0 + 0.005p_T^\gamma$	$1.0 + 0.005p_T^\gamma$

Table 4.2: Variable cuts for three different working points for the photon ID and isolation in the Endcap. Variables are required to be less than the indicated value. The value in parenthesis is the target photon selection efficiency for the working point.

with a charged track arriving within a small window around the supercluster position. This is a relatively harsh requirement, since it can veto real photons that happened to be consistent with a soft track from another particle as well as photons that convert to e^+e^- pairs within the pixel detector. This veto is found to be 94.4% (81.0%) efficient to select photons in the EB (EE) and 98.6% (95.7%) efficient to reject electrons in EB (EE).

The looser conversion-safe electron veto (CSEV) can be used to recover the photons rejected by the pixel veto at the expense of allowing more electrons into the sample. This requires that there be no reconstructed track with a hit in the inner layer of the pixel detector pointing to the ECAL cluster. By requiring the track have a hit in the first layer of the pixels, we reduce the chances of vetoing a converted photon, which would typically convert somewhere within the tracker and thus not leave hits in the early layers (hence the name “conversion-safe”). We would still reject a photon that converts upon interaction with the beampipe (before the first layer of the pixels), but there is little that can be done to recover these photons, since they

appear to be electrons at every point in the detector. The CSEV has higher efficiency to select photons: 99.1% (97.8%) in the EB (EE) but also a lower efficiency to reject electrons: 94.7% (80.4%) in the EB (EE).

4.2.2 Photon Energy Regression

One of the most important components for any search involving the Higgs boson decaying to two photons is a very accurate measurement of the photon energy. The Higgs boson has a very small intrinsic width (< 10 MeV), so the width of the resonance observed in data will be entirely dominated by the mass resolution of the resonance. When the correct identification of the primary vertex is not an issue, the mass resolution of a pair of photons is proportional to the sum in quadrature of the energy resolutions of the two photons. The CMS ECAL has excellent resolution on its own (this was one of the design goals of the detector) and we find that using additional techniques can substantially improve the performance.

To get the best possible photon energy measurement, we use a multivariate regression that uses multiple observables about the distribution of energy deposits within the cluster and global information about the event to estimate the energy of the photon [56, 58, 59]. We use a boosted decision tree (BDT) trained with gradient descent for regression implemented using the TMVA package to implement the MVA [60]. This is trained using a sample of simulated photons with $p_T > 25$ GeV, and the training target is to predict E_{true} , which is the energy of the photon at the generator level, given E_{raw} , the energy of the supercluster, using the other input variables.

There are 38 (31) input variables to the regression for photons in EB (EE) that encode a large amount of information about the cluster and event. The most physically interesting variables are listed in Table 4.3. In addition to these, there are 14 variables describing how the hits are layed out within the most energetic basic cluster within the SC (called the *seed cluster*) and 2 variables that describe the position of the center of the SC relative to the center of the seed cluster. In the barrel we also add information about the local position of the seed cluster, so that the regression can learn about local features of the EB. In the EE we add information about how

much of the energy was deposited in the preshower detector.

variable	description
E_{SC}	the energy of the supercluster
η, ϕ	the position of the cluster
R_9	the energy in a 3×3 crystal matrix around the most energetic crystal divided by the cluster energy
$\frac{E_{5 \times 5}}{E_{SC}}$	the energy in a 5×5 crystal matrix around the most energetic crystal divided by the cluster energy
η -width, ϕ -width	the width of the cluster in the η and ϕ directions
N_{BC}	number of basic clusters making up the supercluster
H/E	the energy in the HCAL behind the cluster (described in Section 4.2.1)
ρ, N_{PV}	the energy density (section 4.3.1) and number of vertices in the event
$\sigma_{i\eta i\eta}, \sigma_{i\eta i\phi}, \sigma_{i\phi i\phi}$	the widths of the clusters (equation 4.1 and the extension to the $\eta - \phi$ and $\phi - \phi$ directions)

Table 4.3: The input variables to the photon energy regression.

We measure the performance of the regression with the default energy reconstruction in samples orthogonal to the training sample, and observe that it substantially improves the energy response. Figure 4.5 shows the value of $\frac{E_{reco}}{E_{true}}$ for photons in simulated $H \rightarrow \gamma\gamma$ events produced through vector boson fusion. We see a large peak at 1 for both methods, but using the regression substantially reduces the fraction of mis-measured photons on the tails of the distribution and increases the fraction that are well reconstructed. Table 4.4 shows the fraction of “well measured photons” (within 2% of the true energy) and the fraction of “poorly measured” (more than 6% from the true energy) for the regression and non-regression energy in the EB and EE. We see that an additional 8.1% (11.7%) of photons become “well measured” in the EB (EE) by using the regression, which indicates that it is working well.

At its most fundamental level, a regression BDT works by finding small regions of parameter space where all the events need very similar correction and using the average in this region as its prediction. The accuracy of this depends on a number of things, including the size of the training sample and how well the true response can be estimated by a smooth function, but in the real world there is always some spread of the training targets within a single classification group. The width of the

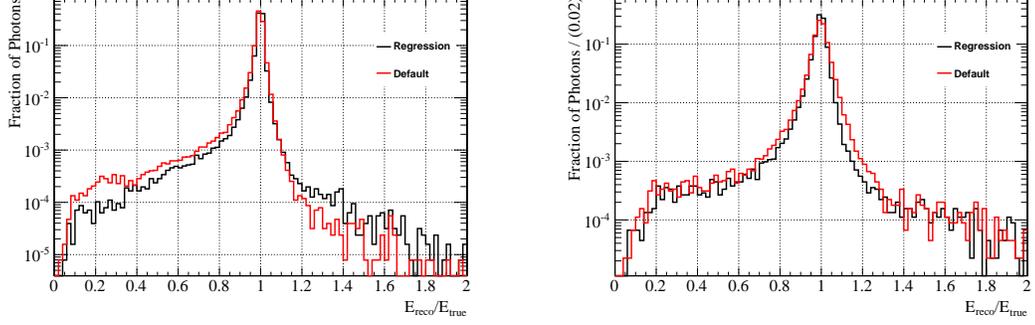


Figure 4.5: $\frac{E_{reco}}{E_{true}}$ for the regression energy (black) and the default energy (red) for photons in the EB (left) and EE (right) in a MC sample of $pp \rightarrow qqH$ events.

region	regression	non-regression
EB		
$0.98 < \frac{E_{reco}}{E_{true}} < 1.02$	82.5%	74.4%
$\frac{E_{reco}}{E_{true}} < 0.94$	4.0%	6.1%
$1.06 < \frac{E_{reco}}{E_{true}}$	0.9%	0.7%
EE		
$0.98 < \frac{E_{reco}}{E_{true}} < 1.02$	59.4%	47.7%
$\frac{E_{reco}}{E_{true}} < 0.94$	7.6%	9.8%
$1.06 < \frac{E_{reco}}{E_{true}}$	2.5%	6.0%

Table 4.4: The fraction of events with reconstructed energies close to the generator values (top lines) and far from the generator values (bottom lines) in EB (top table) and EE (bottom table).

spread of the correction values within a class of events can be taken as an estimate of the uncertainty on the predicted value of the photon energy. When the boosting algorithm is applied we take the prediction of the uncertainty on the energy as the weighted average of the uncertainties from each tree in the final set. This gives us a per-photon estimate of the uncertainty on the error σ_E .

We can investigate how well this σ_E variable corresponds to the true error of the prediction in MC by looking at the difference between the regression energy and the true energy in units of σ_E . Figure 4.6 shows the distribution of $\frac{E_{reco} - E_{true}}{\sigma_E}$ for photons in the EB and EE in the same VBF $H \rightarrow \gamma\gamma$ sample used in Figure 4.5. We also include a fit of the distribution with a convolution of a crystal ball function and a Gaussian to model the resolution. We find that there is an offset in the mean of the fit, indicating that our energy estimate is not perfect, but that the width of the

central part of the distribution is 0.99 ± 0.02 in EB and 1.12 ± 0.05 in EE, indicating that our estimate of the uncertainty is doing a good job of estimating the actual error.

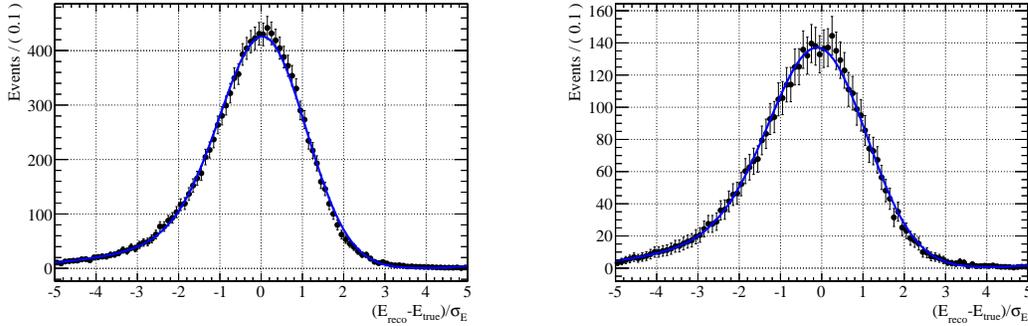


Figure 4.6: The error on the regression energy in units of σ_E in the EB (left) and EE (right) for photons in a MC sample of $pp \rightarrow qqH$ events. The fit is a Crystal Ball function (a Gaussian core with a power law low end tail [61]) convoluted with a Gaussian and the fit has a core width of 0.99 ± 0.02 for the EB fit and 1.12 ± 0.05 for the EE fit.

This estimate of the per-photon error is extremely useful for identifying high quality photons. One feature that we make use of in this analysis is to allow the selection of high resolution photons that will have better mass resolution for any resonance, such as the $H \rightarrow \gamma\gamma$. An example of this is shown in Figure 5.9, which shows the shape of the SM Higgs boson peak for $m_H = 125$ GeV in a category selecting photons with good resolution (small σ_E/E) and another category selecting photons with poor resolution (large σ_E/E). The details will be discussed in depth later, but the conclusion is that the width of the Higgs boson peak is much smaller with high resolution photons.

Another benefit of this resolution estimate σ_E/E is that it can help reject fake photons using additional information that isn't available to the cut-based photon ID. When the regression is trained, it only knows about real photons, so for every object it sees in the data, it tries to identify the region of parameter space for real photons that look most like the object. If the object is not a real prompt photon but a jet with a secondary photon or a neutral hadron faking a photon, it will generally not look like a real photon and be marginalized by the regression and placed in a class

with large variance and high σ_E . Figure 4.7 shows the σ_E/E distribution for MC events passing a basic $H \rightarrow \gamma\gamma$ selection with a pair of photons with invariant mass $110 < m_{\gamma\gamma} < 180$. The MC sample is made up of prompt-prompt, prompt-fake, and fake-fake components from a variety of different processes. The distributions shown are per-photon and are separated into “real” photons that match a generate level prompt photons, and “fake” photons that do not (photons matching a generator level electron are vetoed in this distribution).

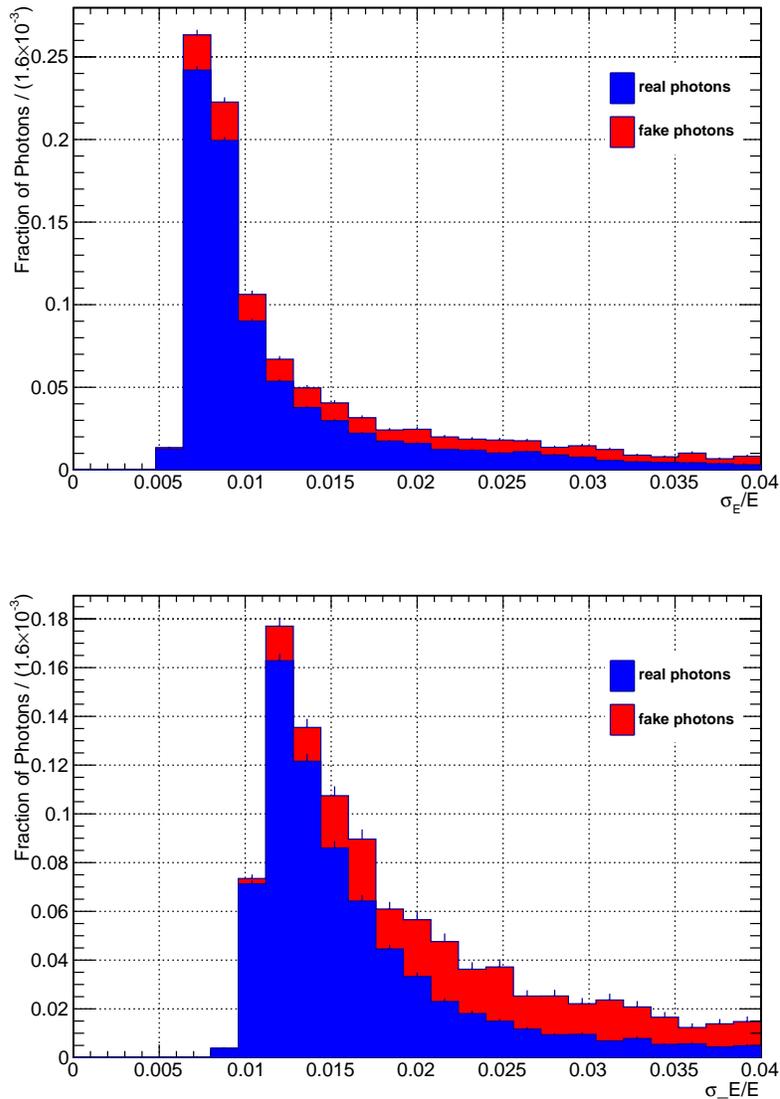


Figure 4.7: The distribution of σ_E/E in MC for photons in the EB (top) and EE (bottom) for events with two photons with mass in the range $110 < m_{\gamma\gamma} < 180$ GeV.

We can see from Figure 4.7 that real photons preferentially occupy small values of σ_E/E (with the scale being different in the EB and EE) while fake photons typically have a much more uniform distribution. Figure 4.8 shows the percentage of all reconstructed photons that are real prompt photons as a function of σ_E/E . We can see that the ratio starts near 100% in both EB and EE and falls off slowly as the value increases. What this suggests is that one can enhance the proportion of real photons even in a sample already passing ID and Isolation requirements by cutting in this variable.

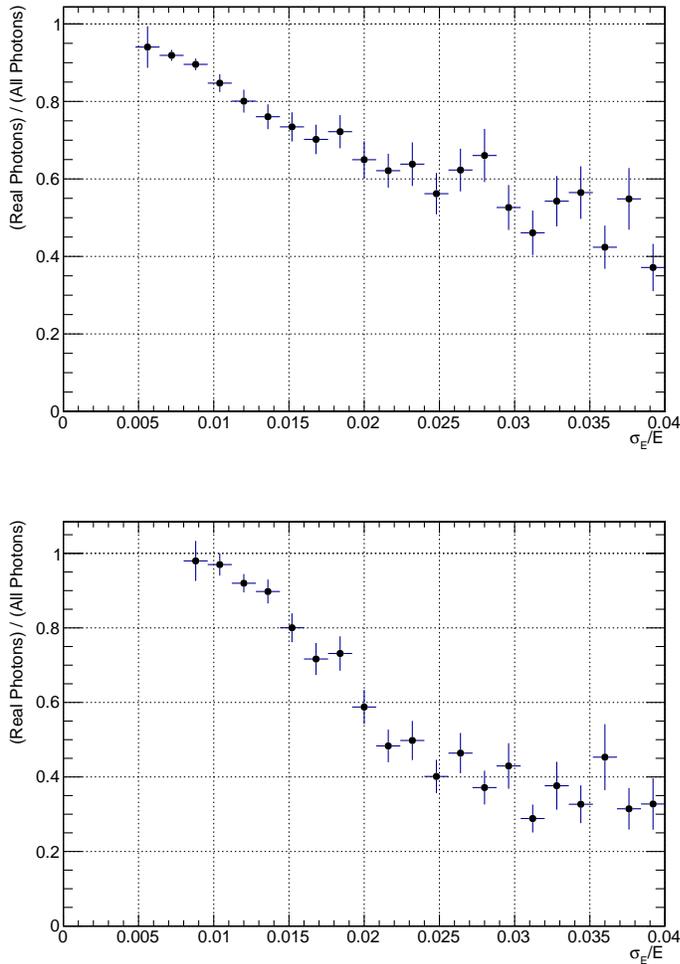


Figure 4.8: Fraction of real photons as a function of σ_E/E for photons in the EB (left) and EE (right) for events with two photons with a mass in the range $110 < m_{\gamma\gamma} < 180$ GeV.

In Figure 4.9 we can see the effect of cutting in the variable σ_E/E . The left plot shows the purity of the sample left after cutting on σ_E/E , which is defined as the fraction of the sample after the cut that are real photons. The right plot shows the efficiency of the cut, which is defined as the fraction of all real photons selected by cutting at the indicated value of σ_E/E . The value of the efficiency does not go to one in the plot because there are some photons with $\sigma_E/E > 0.04$ in both EB and EE.

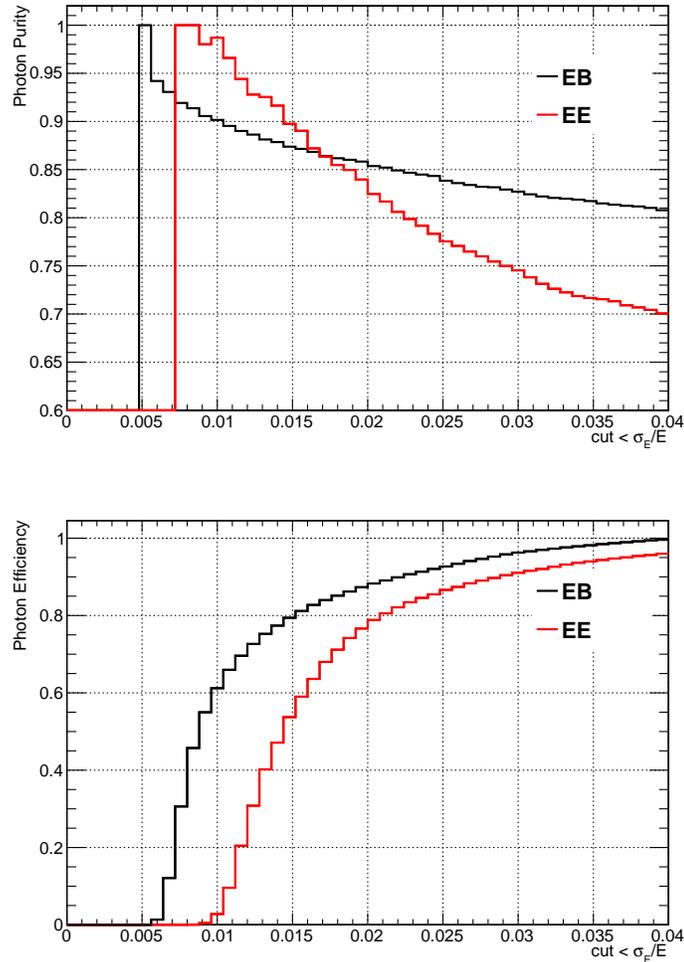


Figure 4.9: Purity (left) and Efficiency (right) of the photon sample selecting photons with σ_E/E below the value on the X-axis for photons in the EB (black) and EE (red). For the definition of purity and efficiency see text.

While this variable has a lot of nice properties in simulation, using it in practice is somewhat more complicated. Ideally one might want to use the shape from Figure 4.8 to assign a likelihood to each photon indicating whether it is real, which could be

used in a multi-dimensional fit to extract a Higgs boson signal. Unfortunately, more sophisticated techniques are stymied by the level of data/MC agreement for this variable. Figure 4.10 shows a comparison of the σ_E/E variable for photons with an analysis selection applied and an invariant mass requirement of $110 < m_{\gamma\gamma} < 150$. It is split into 4 regions in the $|\eta|$ of the photon. We see that the MC does not describe well the shape of the peak at very low σ_E/E in the barrel (top two plots), but does a reasonable job describing the tail of the distribution. In the endcaps (bottom plots) it generally performs better, though there are some regions where the shape is not described well.

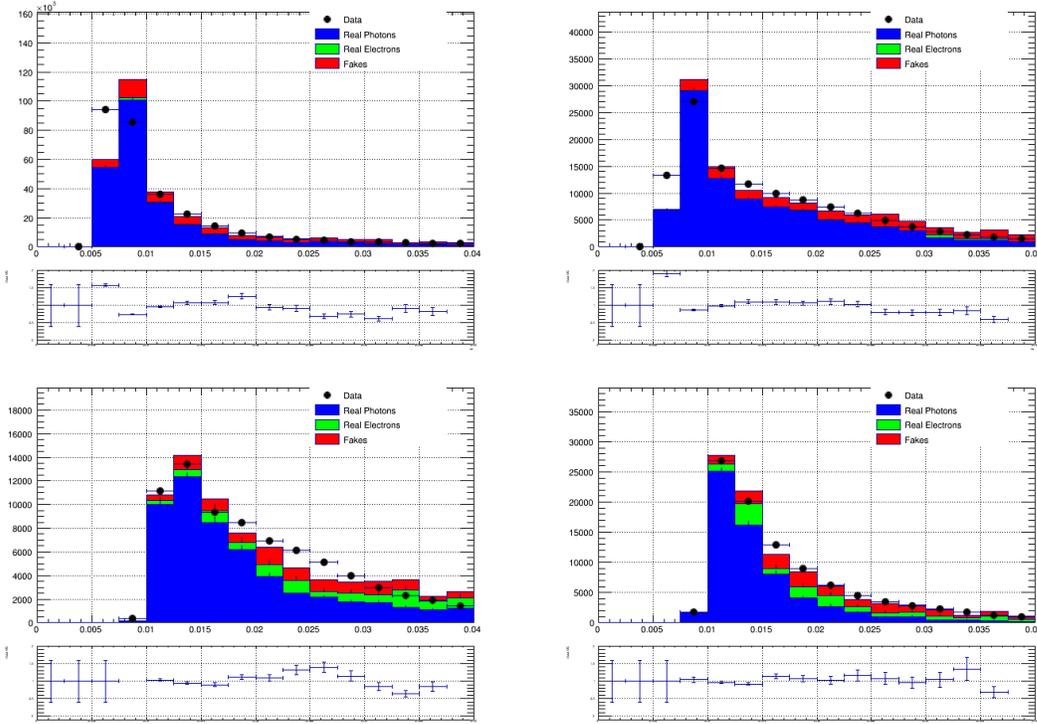


Figure 4.10: Comparison of σ_E/E in data (points) and MC (fill) for photons in $0.00 < |\eta| < 1.00$ (top left), $1.00 < |\eta| < 1.44$ (top right), $1.56 < |\eta| < 2.00$ (bottom left), and $2.00 < |\eta| < 2.50$ (bottom right) in events with two photons with $p_T > 25$ GeV passing the ID and Isolation with $110 < m_{\gamma\gamma} < 150$. The ratio plots are Data/MC in the range $[0, 2]$.

Because the level of agreement is poor on the edge of the distribution, we cut on the variable to select photons well away from the region of disagreement. If we choose the value of $\sigma_E/E < 0.015$ in EB and $\sigma_E/E < 0.20$, which has roughly 80% efficiency for

real photons in each region, we can compare the total efficiency for all reconstructed photon objects using this cut in data versus MC. Table 4.5 shows this comparison in the four regions. We can see that the efficiencies are relatively similar in all regions despite the sometimes substantial difference in the shapes. In the analysis we apply an $|\eta|$ -dependent scale factor to the MC to correct for this efficiency difference, and apply the uncertainties on the efficiency as a systematic error on the event yield in MC.

EB	110 < $m_{\gamma\gamma}$ < 150 GeV		150 < $m_{\gamma\gamma}$ < 180 GeV	
	Data	MC	Data	MC
0.00 < $ \eta $ < 1.00	81.3% \pm 0.2%	79.7% \pm 0.2%	84.6% \pm 0.3%	82.7% \pm 0.4%
1.00 < $ \eta $ < 1.44	57.5% \pm 0.2%	54.8% \pm 0.4%	66.3% \pm 0.5%	62.5% \pm 0.4%
EE	110 < $m_{\gamma\gamma}$ < 150 GeV		150 < $m_{\gamma\gamma}$ < 180 GeV	
	Data	MC	Data	MC
1.56 < $ \eta $ < 2.00	58.2% \pm 0.3%	58.9% \pm 0.4%	67.1% \pm 0.6%	64.0% \pm 0.8%
2.00 < $ \eta $ < 2.50	75.3% \pm 0.3%	75.9% \pm 0.4%	79.5% \pm 0.6%	78.1% \pm 0.8%

Table 4.5: Fraction of events with $\sigma_E/E < 0.015$ ($\sigma_E/E < 0.020$) in EB (EE) in data and MC for events with two photons with $p_T > 25$ GeV passing the ID and Isolation in two different invariant mass regions.

4.3 Jets

The goal of the jet reconstruction is to capture all of the particles produced by the hadronization of quarks and gluons from the interaction. Hadrons produced in this way may not have momenta perfectly parallel to the momentum of the initially produced parton, so we must sum over a cone surrounding the initial direction to capture all of the produced particles. We use the collection of all pf particles produced by the particle flow algorithm (see Section 4.1). We cluster these particles using the anti- k_T algorithm [62] with a radius $R=0.5$ in $\Delta\eta$ - $\Delta\phi$. We use the implementation of the anti- k_T algorithm provided in the `FastJet` package [63, 64]. We label each set of clustered particles a “jet” with a 4-vector equal to the sum of the 4-vectors of all the constituent (clustered) pf-candidates (all treated as massless for the summation).

4.3.1 Measuring Pileup Energy

A variation of the jet reconstruction procedure also provides a method to estimate the average energy density (ρ) coming from particles produced by pileup vertices in the event. To accomplish this, we cluster jets using the k_T clustering algorithm [65, 66], which has been shown to have a tendency to organize a uniform background of soft particles into structures with area πR^2 , where R is the radius parameter given to the clustering algorithm [57]. If the algorithm clusters a jet over an area which contains only a uniform background of soft particles from pileup, then we can recover the average energy density of these soft particles as $\rho = p_T^{jet}/(\pi R^2)$.

Since we cannot easily check for a given jet whether it is composed of only soft particles from pileup, we use the k_T algorithm to cluster all jets in the event with $R = 0.6$ and compute the median value of *rho*:

$$\rho = \text{median} \left(\frac{p_T^{jet i}}{\pi(0.6)^2} \right) \quad (4.3)$$

The relatively large value of $R = 0.6$ is chosen to ensure that the jet areas encompass multiple soft particles so that the calculation is not biased by the large amount of empty area within the jet. The median in equation 4.3 is a good estimator of ρ , since we find that the number in all cases that the number of jets clustered by this algorithm from pileup and detector noise is much greater than the number of jets from the hard collision.

4.3.2 Jet Corrections

The anti- k_T algorithm run on pf-candidate particles produces a good estimate of the energy of the hadronized parton, and we can improve upon this estimate by applying corrections [67–69]. These corrections are grouped into three types applied in sequence: offset correction, MC corrections, relative correction, and absolute correction. These correction are a function of the jet p_T and η as well as global information about the event.

The offset correction is applied first and is designed to correct for the inclusion of particles produced by pileup vertices within the clustered jet area. In Section 4.3.1 we detailed how the average energy from pileup is computed to get the energy density ρ ; the other ingredient is to measure the jet area A . Using the algorithm radius $R = 0.5$ to get the jet area tends to overestimate the size of the jet, since the algorithm need not cluster every piece of energy within the cone. To get a better estimate of the real size of the jet, a large number of extremely soft 4-vectors are added into the event (so soft that they do not change any of the jet properties) and the jet clustering algorithm is re-run. The jet area is taken to be the $\eta - \phi$ extent of of the soft 4-vectors that are clustered with the jet. The corrected jet p_T is then computed:

$$p_T^{L1\ corr.} = p_T^{raw} - \rho\beta(\eta)A \quad (4.4)$$

where $\beta(\eta)$ is a pseudorapidity dependent correction for non-uniformity of the energy response. The size of the correction as a function of η and the number of vertices is shown in Figure 4.11.

The MC calibration is applied next, which is derived in simulation to correct jets to have the same p_T as the matched generated jet. The corrections are derived in simulated QCD events produced with PYTHIA6.4.22 Tune Z2 [13] and simulated in the CMS detector with GEANT4 [19]. Jets are reconstructed in the simulated events, have the pileup correction applied, and are then spatially matched with $\Delta R < 0.25$ to a MC particle jet. The average response $\langle f \rangle = \left\langle \frac{p_T^{reco}}{p_T^{gen}} \right\rangle$ is computed in bins of the reconstructed jet η and p_T . The MC calibration correction applied in data is then simply $\frac{1}{\langle f \rangle}$ for the $\eta - p_T$ bin corresponding to the jet's kinematics. The correction factor derived in this way is shown in Figure 4.12. The correction factor is typically > 1 , indicating that the reconstructed jets tend to underestimate the true energy (as is expected if particles are missed by the reconstruction or clustering).

Finally, “relative” and “absolute” corrections are applied to ensure that the jet energy scale and response is the same throughout the whole detector and at all jet p_T scales. The “relative” correction ensures that the response of jets in the forward

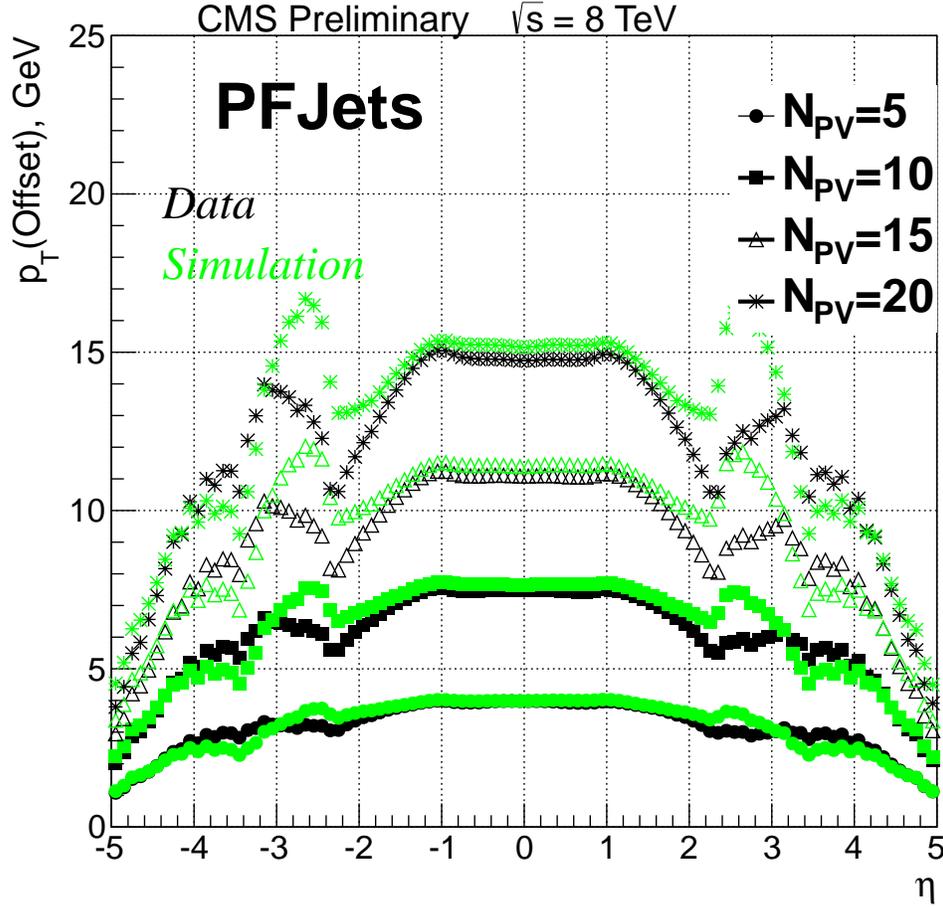


Figure 4.11: The average magnitude of the pileup correction $p_T(\text{Offset}) = \rho A$ as a function of η for simulation (green) and data (black). 4 different scenarios of the number of primary vertices (N_{PV}) in the event are shown, to illustrate the scaling with pileup. Plot from [70]

region is the same as that in the central region by applying an η -dependent correction. The “absolute” energy scale ensures that the low p_T response of the jets is the same as the high p_T response by applying a p_T -dependent correction.

4.3.3 Identifying bottom quark decays

The identification of jets resulting from the decays of bottom quarks is quite important for the experiment, since many physics processes have large coupling to b-quarks, including the top quark, the Higgs boson, and many beyond-the-standard-model pro-

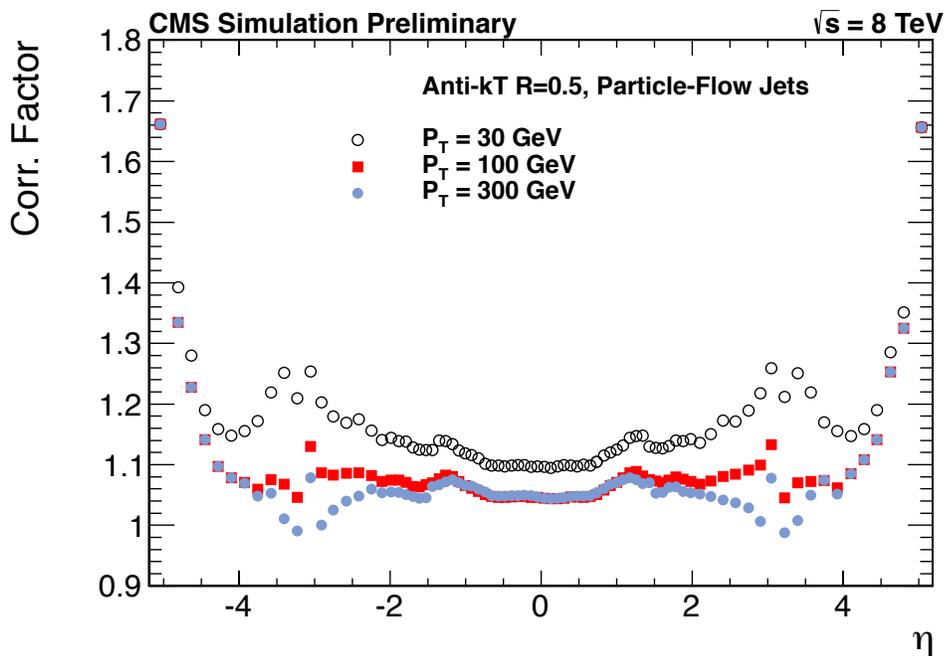


Figure 4.12: The average value of the MC calibration correction as a function of η for 3 different jet p_T values. One sees that the correction factor gets smaller (the initial p_T estimate is better) for higher p_T jets. Plot from [70]

cesses. While jets from b quarks look similar to jets from lighter quarks and gluons in many respects, there are certain properties of the particle and the detector that make it possible to discriminate with reasonable efficiency. The bottom has a relatively long lifetime, which allows it to move several millimeters from the interaction before decaying, mostly through $b \rightarrow cW^*$, which results in a final state with either three quarks or one quark, one charged lepton, and a neutrino.

The relatively light bottom mass compared to a typical jet p_T measured by CMS means that the visible decay products of the b are mostly collimated and reconstructed as a single jet by the pf algorithm. The b -jet identification uses the displaced decay of the b into 2-3 charged particles, looking within each reconstructed jet for an intersection of tracks displaced from the primary vertex by an amount consistent with the lifetime of a b quark with the energy of the reconstructed jet. In practice the algorithm uses a number of variables related to the quality of the tracks and their level of consistency with a secondary (displaced) vertex within the jet to create a likelihood ratio. The full details are given in [71], but the result is called the combined

secondary vertex (CSV) discriminator, which is a score between 0 (not-b-like) and 1 (b-like) for each jet.

Figure 4.13 shows the distribution of CSV for a sample of jets in data and MC. One sees that the fraction of b-quark jets is highly enriched at large values of the discriminator [72]. The discrimination is, however, not perfect, so we define a series of working points based on the mistag rate (the probability for a non-b-jet to have a CSV score higher than the discriminator). These are labeled CSVL (10% mistag rate), CSVM(1% mistag rate), and CSVT(0.1% mistag rate). Table 4.6 shows the cut values along for these working points as well as the measured efficiency to tag a b-quark for that working point (e.g. 18% of real b-quarks will be missed by the CSVL working point).

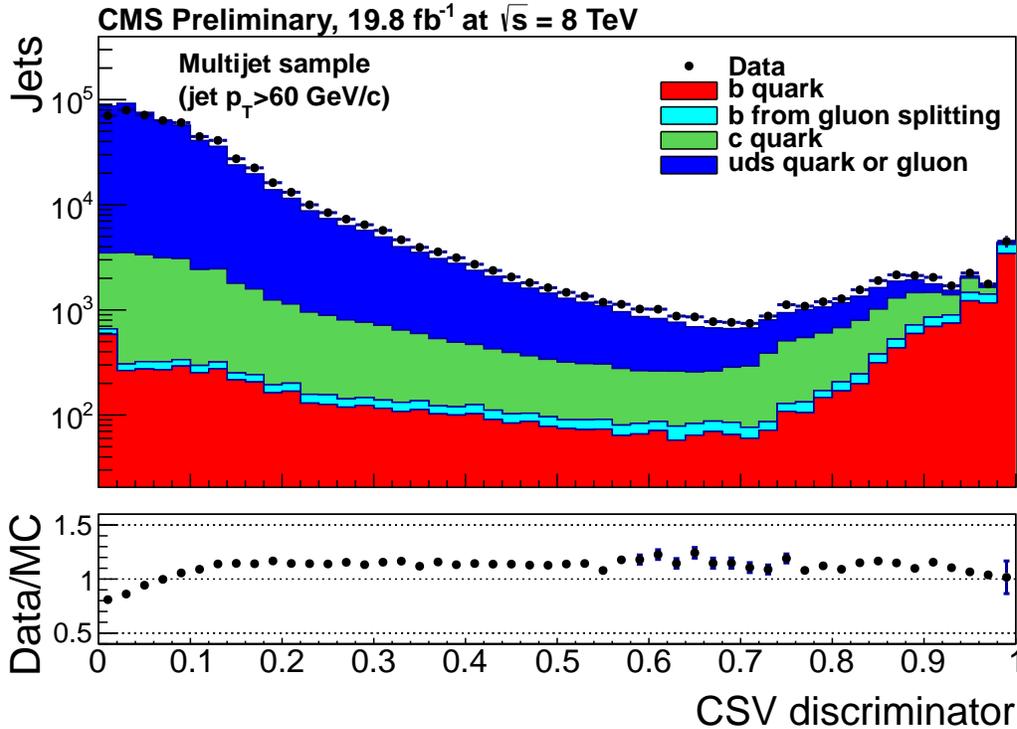


Figure 4.13: The distribution of the CSV discriminator for bottom quark jets. The data is selected in a sample with many high p_T jets and the score for each jet is plotted. The stack histogram is the distribution from MC simulation for the same selection split by the type of jet determined by MC truth. Plot from [72].

At higher p_T , it becomes more difficult to identify b-quarks. The decay products

Working Point	Score	Average Mistag Rate	Average b-jet efficiency
CSVL	0.244	9.9%	82%
CSVM	0.679	1.1%	68%
CSVT	0.898	0.2%	52%

Table 4.6: The cut values

become increasingly collinear, making it difficult to separate the hit patterns into two or three distinct tracks and the tracks become straighter, making it more difficult to accurately measure p_T and constrain the track parameters. Figure 4.14 shows the mistag probability of the medium working point as a function of the jet p_T . One sees that it is almost 4 times as likely to identify a light jet as a b-quark at 900 GeV than at 50 GeV. The difference in mistag rate between simulation and data reflects the slight differences between the tracks in simulation and in data.

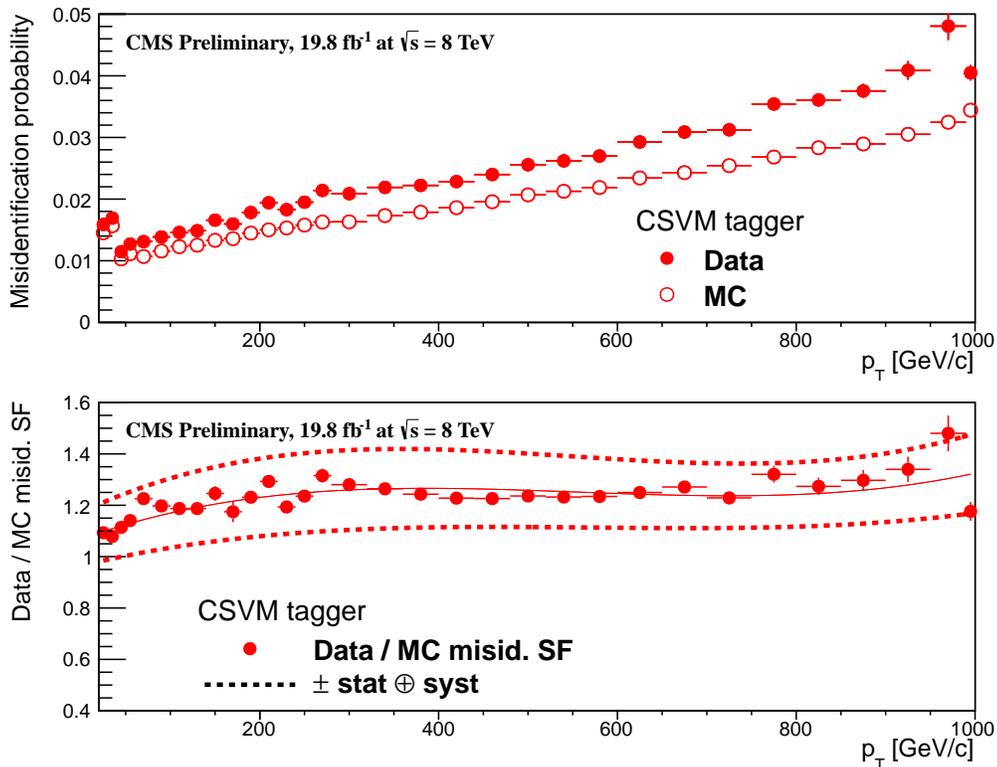


Figure 4.14: The mistag rate (light quark identified as b-quark) as a function of the jet p_T in data (solid points) and MC (hollow points). The ratio of the mistag rate is shown on the bottom. This data/MC difference is corrected for in the analysis. Plot from [72].

4.4 Missing Energy

If we could perfectly detect and measure all of the decay products from only the main interaction, we would expect to find that the vectoral sum of their momenta in the plane transverse to the beamline would be 0, since the incident protons have momentum only in the \hat{z} direction. If we continue the paradigm of perfect measurement, but take into account the existence of particles that don't interact in our detector (e.g. neutrinos), we expect to see some events where the sum of the transverse momentum is non-zero indicating the presence of momentum carried away by the invisible particle. If we imagine an event where a neutrino is produced and escapes the detector and where \vec{p}_T^i are the transverse momenta of the visible objects in the event, then we expect

$$\vec{p}_T^{neutrino} = - \sum_i \vec{p}_T^i \quad (4.5)$$

We call this the missing transverse energy (MET, also denoted \vec{E}_T^M or $E_T^M \equiv |\vec{E}_T^M|$).

Reconstructing the MET in a real event follows equation 4.5. While we don't have access to the \vec{p}_T of every particle produced, we can get a good estimate by using the pf-candidates described in Section 4.1. We then compute

$$\vec{E}_T^M = - \sum_i \vec{p}_T^i \quad (4.6)$$

where the index i runs over all pf-candidates [73]. To the extent that the pf algorithm has reconstructed all particles in the event, this reproduces equation 4.5.

Just as jets are corrected to provide better performance, so too are corrections applied to the E_T^M [74]. The first step is to correct for pileup. There should be little true E_T^M (neutrinos) from pileup collisions, but imperfect energy measurement of the pf-candidates can create fake E_T^M . We find that most of the mis-measurement occurs for neutral particles, since they get no help from the tracker, so we first measure the imbalance of charged pf candidates associated with vertices other than the primary vertex $\left(\vec{E}_T^M\right)_{charged} = - \sum_{i \in ch. pf-cand.} \vec{p}_T^i$. We take this as a good estimate of the true

charged component of E_T^M from pileup and, since the pileup should have little total E_T^M , we set $\left(\vec{E}_T^M\right)_{neutral}^{true} = -\left(\vec{E}_T^M\right)_{charged}$. We measure in Monte Carlo as scale factor $R((E_T^M)_{charged})$ such that $\left\langle\left(\vec{E}_T^M\right)_{neutral}^{true}\right\rangle = R((E_T^M)_{charged})\left\langle\left(\vec{E}_T^M\right)_{neutral}^{measured}\right\rangle$. From this, we derive the PU-corrected E_T^M :

$$(E_T^M)^{PU-corr} = E_T^M + (1 - R((E_T^M)_{charged})) \sum_{i \in ch. \text{ pf-cand.}} \vec{p}_T^i \quad (4.7)$$

The second correction takes advantage of the derived jet energy corrections (section 4.3.2). We partition the pf-candidate collection based on whether the candidate was clustered into the jet collection ($\{clustered\}$) or not ($\{unclustered\}$). We can then write

$$E_T^M = - \sum_{i \in \text{pf-cand.}} \vec{p}_T^i = - \sum_{i \in \{clustered\}} \vec{p}_T^i - \sum_{i \in \{unclustered\}} \vec{p}_T^i$$

We can then replace $\sum_{i \in \{clustered\}} \vec{p}_T^i$ with $\sum_{jets} \vec{p}_T^{jet}$, where \vec{p}_T^{jet} has all the jet corrections applied. We can then write the fully corrected E_T^M variable as

$$(E_T^M)^{corr} = - \sum_{i \in \{unclustered\}} \vec{p}_T^i - \sum_{jets} \vec{p}_T^{jet} + (1 - R((E_T^M)_{charged})) \sum_{i \in ch. \text{ pf-cand.}} \vec{p}_T^i \quad (4.8)$$

The performance of the E_T^M is measured in events with a Z boson decaying in the channels $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ and in events with a high p_T photon and a jet. Figure 4.15 shows good agreement for the E_T^M value between data and simulation over a large range and a variety of processes. The resolution is measured by treating the transverse momentum (\vec{q}_T) of the dilepton pair or the photon as perfectly measured (which is a fair approximation given the relative resolutions between jets, E_T^M and $e/\mu/\gamma$) and measuring the vectorial sum p_T of the jets (\vec{u}) and the E_T^M in the direction

parallel and perpendicular to the Z or γ .

$$u_{\parallel} + (E_T^M)_{\parallel} = q_T + \Delta_{\parallel}$$

$$u_{\perp} + (E_T^M)_{\perp} = \Delta_{\perp}$$

From this one can derive an event-by-event uncertainty on the E_T^M . The average value of this uncertainty as a function of the total energy of the event (the sum of the E_T of all pf-candidates) is shown in Figure 4.16. One sees that the resolution is consistent in all the MC samples and consistent with the value measured in data. The somewhat poor agreement between in the E_T^M variable between data and MC in the photon plus jet sample reflects the difficulty of effectively simulating all of the QCD processes that produce prompt, secondary, or fake photons or in their final states. This is one of the reasons that simulated QCD diphotons events are not used for the background prediction in the SM Higgs to two photon searches, or in this analysis.

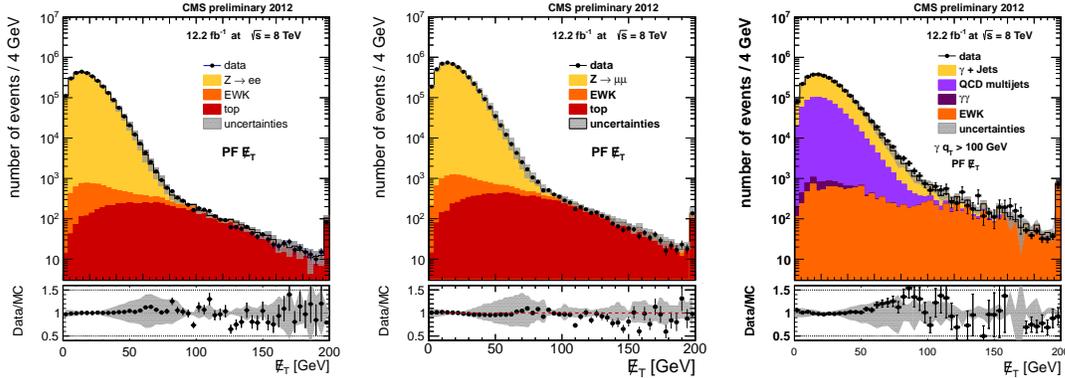


Figure 4.15: The distribution of E_T^M for events with a high p_T e^+e^- pair (left) or a high p_T $\mu^+\mu^-$ pair (middle) consistent with the Z mass ($60 < m_{ll} < 120$ GeV) and events with a high p_T photon in the EB (right). The colored stack plots are the distributions of E_T^M in the simulation of the various background processes that populate the selection criteria. Plot from [74].

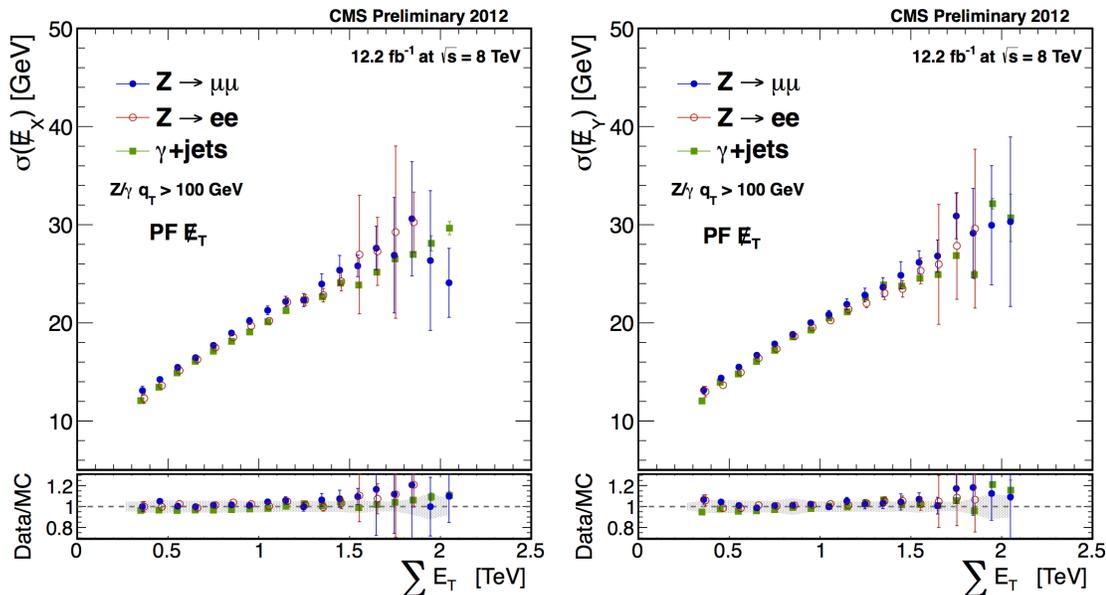


Figure 4.16: The resolution of the E_T^M in the x (left) and y (right) directions as a function of the sum of the energy of all pf candidates in the event. The ratios at the bottom show that the measured resolution is consistent in simulation and data. Plot from [74].

4.5 The CMS Trigger

The collisions at the LHC predominantly (by several orders of magnitude) produce QCD interactions, which are not the main focus of the CMS physics program. In order to avoid saving the data from all of these collisions, which would amount to > 1 terabyte/s of data, a trigger system is used to select physically interesting events in real time based on their contents [75]. The CMS trigger system is divided into two components: the Level 1 (L1) trigger and the high level trigger (HLT). The L1 is a hardware based system that processes every collisions observed in the detector and uses coarse criteria to select potentially interesting events. The HLT is a software based trigger running on 13000 CPU cores that runs on each event selected by the L1 and is able to do a more sophisticated reconstruction.

When a collision occurs in CMS, the output from the detectors is held in a buffer and sent to the L1 trigger system. This system must decide within $4 \mu\text{s}$ whether to reject the event, in which case the data is flushed from the buffer, or whether

to accept the event, in which case the data is sent to the HLT [76]. The L1 uses only information from the calorimeters and muon chambers (not the tracker) and is divided into two components: the global calorimeter trigger (GCT) and the global muon trigger (GMT). The GCT takes information from the ECAL and HCAL and reconstructs jets and e/γ objects (it cannot distinguish between electrons and photons because of the lack of track information) as well as global information about the event (the total energy, the scalar sum of the jet momenta, and the missing energy). The GMT takes the information from the muon chambers and looks for muon candidates (again using only muon chamber information).

An L1 trigger is a request for a certain number of L1 objects or L1 global sums to be above a given threshold. For instance, a L1 trigger could request the presence of two e/γ objects with $p_T > 10$ GeV, or be asymmetric and require one above 13 GeV and another above 7 GeV. Since the L1 has limited time and information with which to reconstruct objects, the energy of the object reconstructed at L1 will only be an estimate of the true energy of the object, and there will be some resolution that characterizes the accuracy of this reconstruction. Since we cut on the L1 object p_T and accept any object with p_T greater than the threshold, this resolution creates a *turn-on curve*, which says how likely an object of a given true p_T is to pass the selection.

Figure 4.17 shows the turn-on curve for selecting an e/γ object at L1 with $p_T > 20$ GeV. One can see that an object with a true p_T of 20 GeV has only $\approx 30\%$ chance (efficiency) of being selected by this trigger, indicating that there is an inherent underestimation of the energy of e/γ at L1. One also sees that there is a small $< 20\%$ efficiency to select objects that had a true $p_T < 20$ GeV and a 5% (12.5%) inefficiency to reject objects in EB (EE) that had $p_T = 25$ GeV. When we discuss triggers, we often consider the plateau efficiency, which is the region where the efficiency is no longer changing quickly as a function of p_T and is approaching some constant maximum value (in the case of Figure 4.17, this is at about 30 GeV and is approaching 1). In this case we say that the trigger is nearly fully efficient to select electrons with $p_T > 30$ GeV and the inefficiency is included as a scale factor and systematic on the

MC yields.

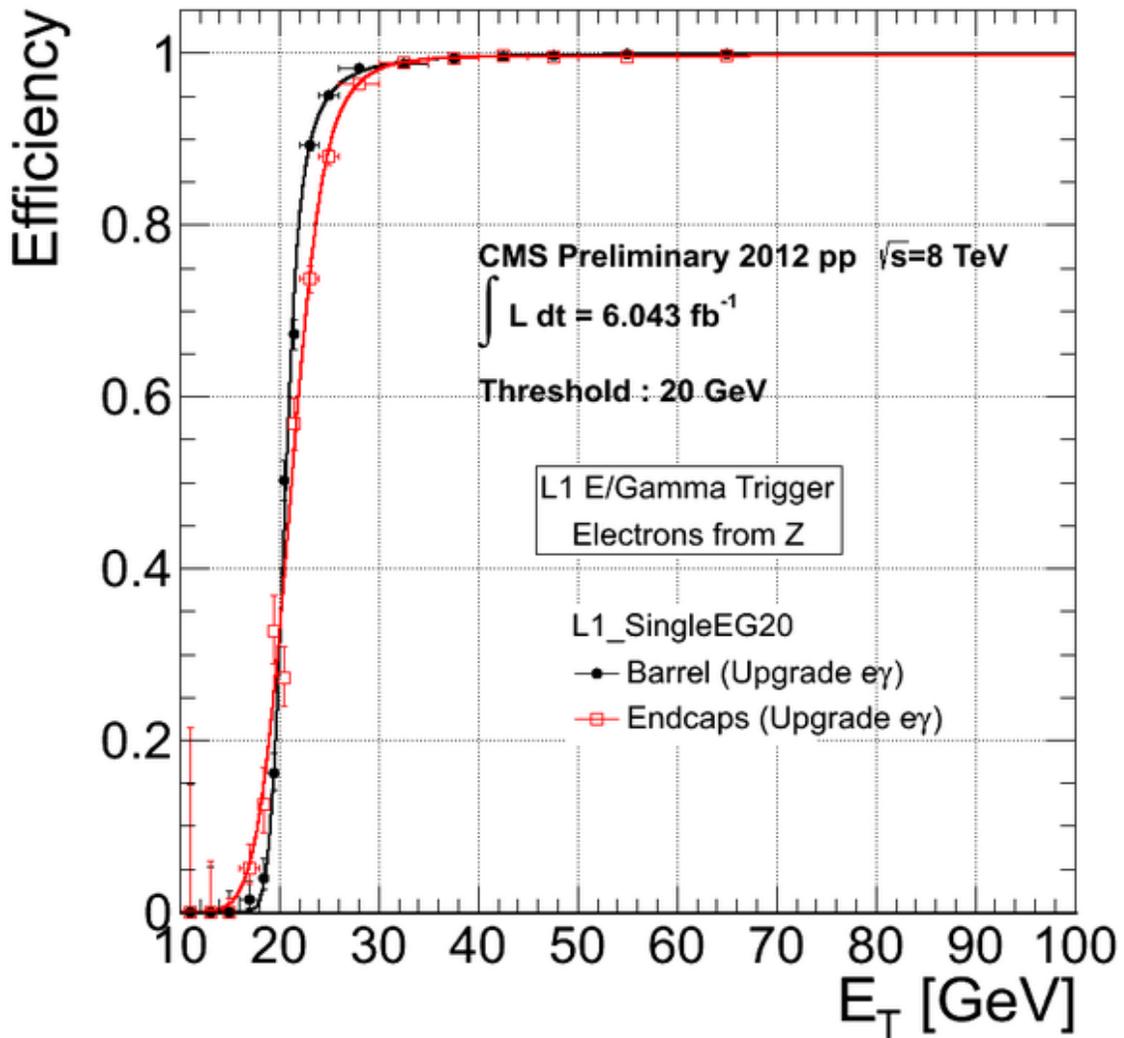


Figure 4.17: The efficiency of a single e/γ object with $p_T > 20$ GeV with the L1 trigger as a function of the true p_T of electron. Plot from [77].

The L1 trigger takes in the full collision rate of 20 MHz of events and is configured to select the most interesting 100 kHz of events. These 100 kHz of events are passed to the HLT, which is able to do a more sophisticated reconstruction and analysis of the events. Given the sometimes large resolution between the L1 and the true energy of the particle, the goal of the L1 is usually to provide the loosest possible triggers that fit within the 100 kHz bandwidth allocation (this limit is set by technical limitations of the system).

Since our final state contains a $H \rightarrow \gamma\gamma$ decay, we are interested in final states containing two photons. We ask for any event that passed the L1 with at least one e/γ candidate with $p_T > 22$ GeV (this is the loosest requirement available in the L1). The HLT then performs a reconstruction similar to the version described above, but with two caveats. The first is that the HLT must decide whether to accept an event within ≈ 200 ms on average and the second is that it does not have access to all of the calibration information about the run. The first caveat means that certain very CPU intensive activities (like the particle flow reconstruction or certain aspects of the tracking) cannot be run on most events. The second means that there will still be resolution effects between the HLT and the true object energies, despite the much more accurate reconstruction possible at the HLT.

For photons, the effects of these caveats are not extreme. The energy reconstruction of the photon does not need the particle flow information, but it is missing some corrections that are derived offline, and it cannot use the regression. The lack of particle flow reconstruction does, however, mean that the isolation must be done quite differently than described in Section 4.2.1. Instead of the pf isolation, the HLT uses *detector-based* isolation, which looks at energy deposited in the calorimeters and tracker around the direction of the photon. The full details of the HLT-based isolation can be found in [76, 78], but the essence is that the energy in the ECAL and HCAL in an annulus around the photon plus the energy of all tracks in a cone around the photon are summed, and the photon is rejected if this sum is above some threshold. This is conceptually similar to the pf-based isolation, and indeed was used as the offline isolation criteria in many early analyses from CMS [79].

Figure 4.18 shows the turn-on curve for a single photon with $p_T > 36$ GeV at HLT. The resolution coming from the inaccuracy in the energy reconstruction can be seen in the width of the turn-on region, and we see that the trigger represented by the black curve is $> 95\%$ efficient for photons with offline $p_T > 40$ GeV. The different colors represent three different photon identification schemes. The red represents the isolation discussed above, where the isolation sum is required to be below some threshold. We can see that this scheme has an inconsistent plateau efficiency, which

reflects the mismatch of this detector-based scheme when compared to the pf-based isolation used offline. The blue curve represents an alternate scheme that has no cut on the ID or isolation variables, but only requires the photon to have $R_9 > 0.85$. This produces a smooth plateau, but does not reach 100% efficiency since not all real offline photons have $R_9 > 0.85$. The black curve represents the logical OR of the two requirements, accepting photons if they either pass the ID and isolation requirements or pass the R_9 cut. This produces the black turn-on curve, which has a consistent plateau with an efficiency near 100%. This is the trigger that is used for this analysis.

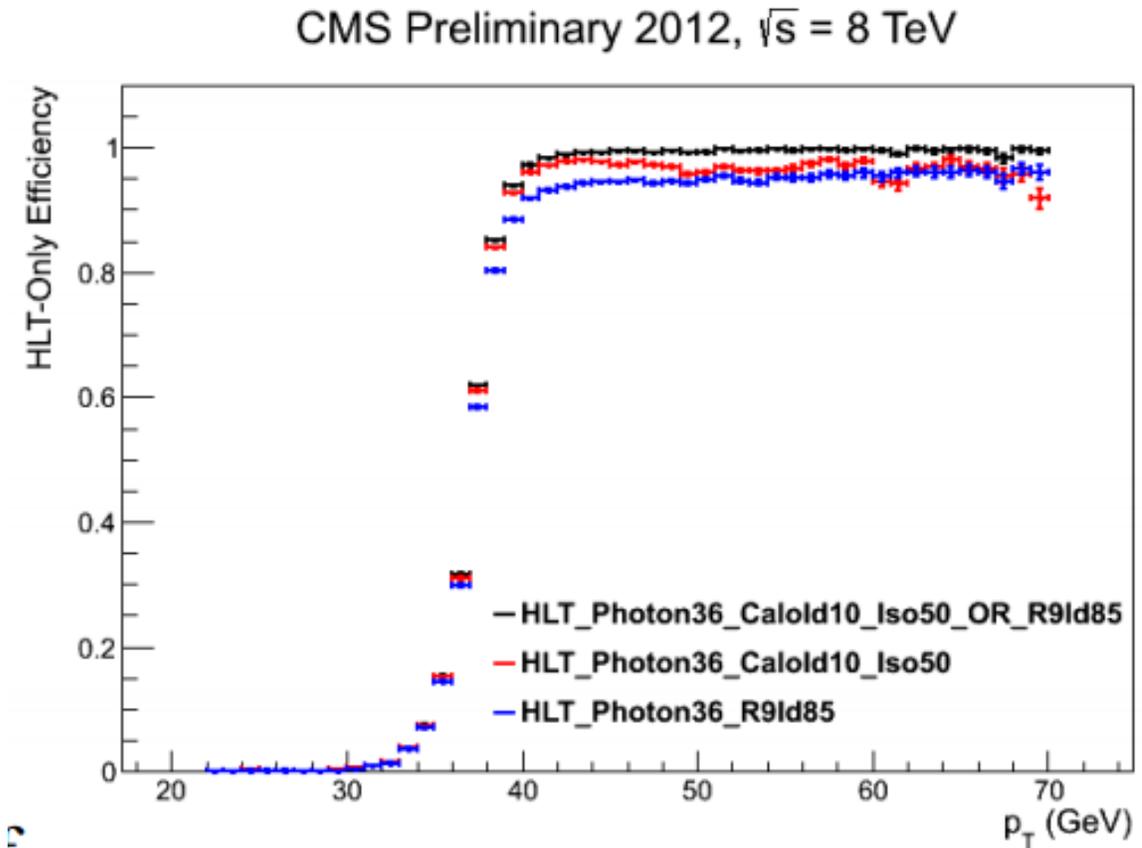


Figure 4.18: The turn-on curve for an HLT photon with $p_T > 36$ GeV as a function of the offline reconstructed p_T of the photon. The three different colors represent three different ID schemes that are described in the text. Plot from [78]

Chapter 5

Search for Supersymmetry in Events with a Higgs Boson Decaying to Two Photons

The discovery of the Higgs boson by the CMS and ATLAS experiments [80,81] provides new channels to search for beyond the standard model (BSM) physics at the LHC. The production modes of the Higgs boson are fully predicted in the standard model, so any evidence of anomalous production is *prima facie* evidence of BSM physics. In particular, some models of supersymmetry predict that the Higgs boson should be produced in the decay chain of pair produced supersymmetric particles. To search for phenomena of this type, we use the razor variables (described in section 5.4) to define kinematic regions with small expected contributions from the SM Higgs boson and from SM processes that produce two photons, and search for excess of events compared to predicted yields.

We look for events with a resonant di-photon pair consistent with $H \rightarrow \gamma\gamma$. This decay process has a small branching ratio (0.23% for $m_H = 125$ GeV), making it rare (we expect only ≈ 1000 such events in 19.8 fb^{-1} of data taken at $\sqrt{s} = 8$ TeV) and has non-trivial background from continuum di-photon production, making finding these events challenging. This challenge is compensated by the excellent photon resolution achieved by CMS thanks to the ECAL, ensuring that the $H \rightarrow \gamma\gamma$ decay shows up as a narrow resonance in the $m_{\gamma\gamma}$ spectrum on top of smoothly falling background [59]. This allows the background to be determined by a fit to the

$m_{\gamma\gamma}$ distribution, obviating the need for modeling of the complicated SM processes that comprise the background. These features are why $H \rightarrow \gamma\gamma$ was one of the first channels to independently achieve 5σ significance for the presence of the Higgs boson in 2012. We use all of these characteristics to our advantage in this search for BSM production of the Higgs boson.

5.1 Blinding Procedure

The analysis is performed blindly, i.e. with all selection requirements determined without using the experimental data in regions where the search is looking for signals of new physics. Specifically this is achieved by not looking at any data with $m_{\gamma\gamma}$ consistent with the mass of the Higgs boson ($120 < m_{\gamma\gamma} < 131$ GeV) until all background studies and estimations were completed and fully documented and reviewed. The analysis is tuned using data in the $m_{\gamma\gamma}$ sidebands ($m_{\gamma\gamma} \notin [120, 131]$), MC simulation of the SM Higgs boson and MC simulation of certain benchmark SUSY signals that produce Higgs bosons as part of the decay chain of a supersymmetric particle. The blinded analysis is presented in chapter 5 and the unblinded results are discussed in chapter 6.

5.2 Datasets and Triggers

The analysis uses data collected by the CMS experiment in 2012. The data collected during that run is split into 4 eras based on computing needs and LHC machine schedules: Run2012A, Run2012B, Run2012C, and Run2012D. The LHC evolved over the course of 2012, providing periods of higher luminosity and pileup as the year progressed. As a consequence, the era toward the end of 2012 (Run2012D) contained periods of much higher pileup and luminosity than did the era at the beginning of the year (Run2012A). These difference are summarized in Table 5.1; one can see that Run2012D contains substantially more data and higher pileup than does Run2012A.

Era	$\int \mathcal{L} \text{ (fb}^{-1}\text{)}$	Average pileup
Run2012A	0.5	17
Run2012B	4.8	19
Run2012C	7.1	20
Run2012D	7.3	21

Table 5.1: Datasets used in analysis

We collect the data using a trigger that selects two photons. Both photons are required to have $H/E < 0.10$ and either $R_9 > 0.85$ or all of:

- $\sigma_{i\eta i\eta} < 0.014(0.035)$ EB (EE)
- ECAL isolation < 5 GeV
- HCAL isolation < 5 GeV

H/E is the energy in the HCAL behind the ECAL cluster divided by the energy in the ECAL; R_9 is the energy in a 3×3 matrix around the highest energy crystal divided by the energy of the supercluster, and $\sigma_{i\eta i\eta}$ is the log-energy weighted width in the η direction of the cluster (all discussed in Section 4.2.1). The ECAL and HCAL isolation are detector level-isolation variables described in Section 4.5. The highest p_T photon is required to have $p_T > 36$ GeV and the next highest p_T photon must have $p_T > 22$ GeV. This trigger was used for the discovery of the standard model Higgs boson in 2012 [59] and is designed to be extremely loose: selecting many objects that are not photons in order to have a high efficiency to select events with two real photons.

5.2.1 SM Higgs Boson Monte Carlo Samples

The analysis relies on using data to form background predictions wherever possible, but the production of the Standard Model Higgs boson cannot be estimated in this way. Since this search looks for BSM production of the Higgs bosons, its standard model production is a background that must be accounted for. To this end, we use MC simulation as discussed in Section 2.1.2 and summarized in Table 2.4.

5.3 Event Selection

The selection requires the presence of two reconstructed high p_T photons, since the final state should contain a Higgs boson decaying to two photons. The p_T thresholds are the lowest possible values that keep us in the plateau of the trigger efficiency. The razor variables used for this search require the existence of at least one additional object in the event, so we will require the presence of at least one jet, which can be hadrons from the production of a quark or gluon, a lepton, or a photon as described in Section 4.3.

5.3.1 Photon Selection

We preselect events with at least two photons above 25 GeV and with at least one above 40 GeV (vetoing any photon that falls in the pseudorapidity range $1.44 < |\eta| < 1.56$, to avoid mismeasurement from the EB-EE transition region) passing the loose photon ID ($\sigma_{i\eta i\eta}$ and H/E requirements in tables 4.1 and 4.2). For each event in the preselected sample, we select the pair with the highest $\sum p_T$ of the photons with invariant mass $100 < m_{\gamma\gamma}$, satisfying the 40/25 p_T requirements, the $|\eta|$ veto and the loose photon ID. In $\approx 99\%$ of selected events, there are only two photons satisfying the ID with sufficiently high p_T , so this is simply a selection, but in the cases where there are 3 or more photons in the event, it provides a procedure to select one pair of photons. We call the selected pair the *candidate pair*.

For each preselected event, we veto the event if either photon in the event fails the loose photon isolation (pf-isolation requirements in tables 4.1 and 4.2) or if either photon falls in the EE ($|\eta| > 1.5$). This method for selecting the photon-pairs is driven by the design of our analysis. In Section 5.4 we will discuss how the objects in the event are divided into hemispheres for the construction of the razor variables; in this procedure we will treat the two photons selected as a single massive four vector representing the di-photon system. In the case where these photons are coming from the decay of a single massive particle, the decay products should be forced into a single

hemisphere to get the correct peaking behavior from the M_R variable. In the case where the two photons come from continuum processes, this treatment will continue to produce the smoothly falling background we expect from processes without a heavy scale in the event.

To take advantage of this key difference between peaking background and continuum background, it is very important that in events with a Higgs boson we correctly identify the photons. We prefer to reject an event where the wrong photons are selected rather than keep it with incorrectly determined kinematics. In $\approx 29\%$ of standard model $H \rightarrow \gamma\gamma$ events, one of the two photons falls into EE (see Figure 5.1). If we didn't consider the endcaps, we could easily have events in which one photon from the Higgs boson ends up in EE and we associate an unrelated real or fake photon with the photon from the Higgs boson that went in EB. This would create an odd situation where the kinematics of the di-photon system looks like background, but the kinematics of the rest of the event is recoiling against the massive Higgs boson.

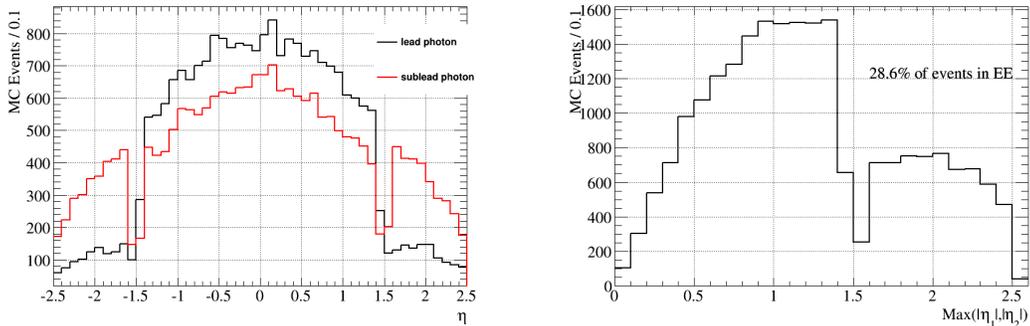


Figure 5.1: The distribution of the η (left) of the highest p_T (black) and second-highest p_T (red) photons in the event and the distribution of $\max(\eta_{\text{highest}}, \eta_{\text{second-highest}})$ (right) for an SMS with $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0 \rightarrow bb\gamma\gamma\chi_0\chi_0$ with $m_{\chi_2} = 130$ GeV and $m_{\chi_0} = 1$ GeV. From the right plot, one calculates that $\approx 29\%$ of events have at least one photon falling in the EE.

We have an analogous concern for the photon isolation: we specifically look for events with photons in high jet multiplicity environments in this analysis, so there are inevitably cases where a photon and a jet are in near vicinity or overlap. If we applied the photon isolation before looking for photons, we would end up in exactly the same situation as we considered for the inclusion of the endcap, the isolation would reject

the photon because of the overlap with the jet, and we could end up with confused kinematics.

We illustrate this in MC simulation of $pp \rightarrow t\bar{t}H \rightarrow \gamma\gamma$ and $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0 \rightarrow bb\gamma\gamma\chi_0\chi_0$ events. We select pairs of photons in two ways: first, the method described above; and second selecting the highest $\sum p_T$ photons falling in EB passing loose ID and isolation. We find that these two methods have very similar selection efficiencies for $t\bar{t}H$ (31.5% for our method and 33.1% for the other) and the SUSY model (56.7% and 57.7%, respectively). We then compare the invariant mass distributions for the photon pairs selected by these methods in Figure 5.2. There are substantially fewer events with incorrect $m_{\gamma\gamma}$ (not near 125 GeV) with the selection used in the analysis. The overall efficiency is very similar between the two methods, since in the large majority of events there are only two photons, so there is no net difference between the events.

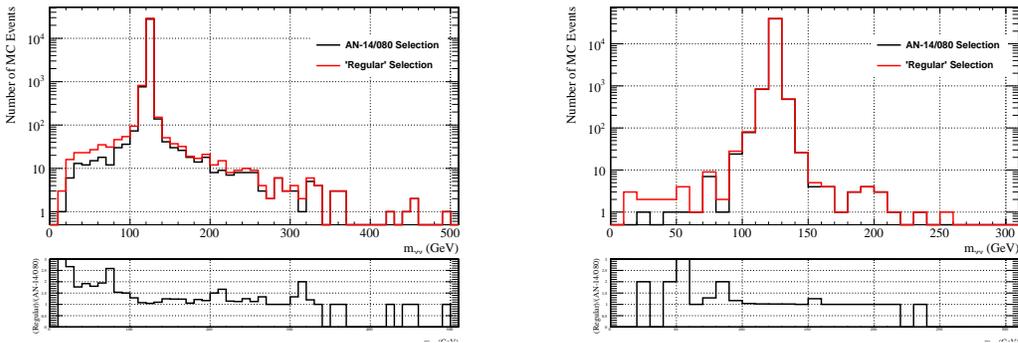


Figure 5.2: The invariant mass of the photon pairs selected using the method used in this analysis (black) and an alternate selection where the ID, isolation, and EB cut occur before pair selection (red) for $pp \rightarrow t\bar{t}H$ MC (left) and $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0$ MC (right). The ratios are the efficiency for the “regular” selection divided by the efficiency of the selection used in this analysis.

The photon energy is corrected using the energy regression described in Section 4.2.2. As described, this regression gives not only an excellent estimation of the true energy of the photon, but also a per photon uncertainty on that energy: σ_E . We will use this value to group events based on the relative resolution σ_E/E of the photons in the event, as described in Section 5.3.4.

Finally, the event is vetoed if the p_T of the di-photon system (denoted $p_T^{\gamma\gamma}$) is

below 20 GeV. We select this because we find that our signal usually produces a Higgs boson with a moderate transverse boost, while the continuum background does not. This distribution is shown in Figure 5.3; we see that below 20 GeV there is very little contribution from our signal hypothesis but a substantial contribution from the continuum backgrounds.

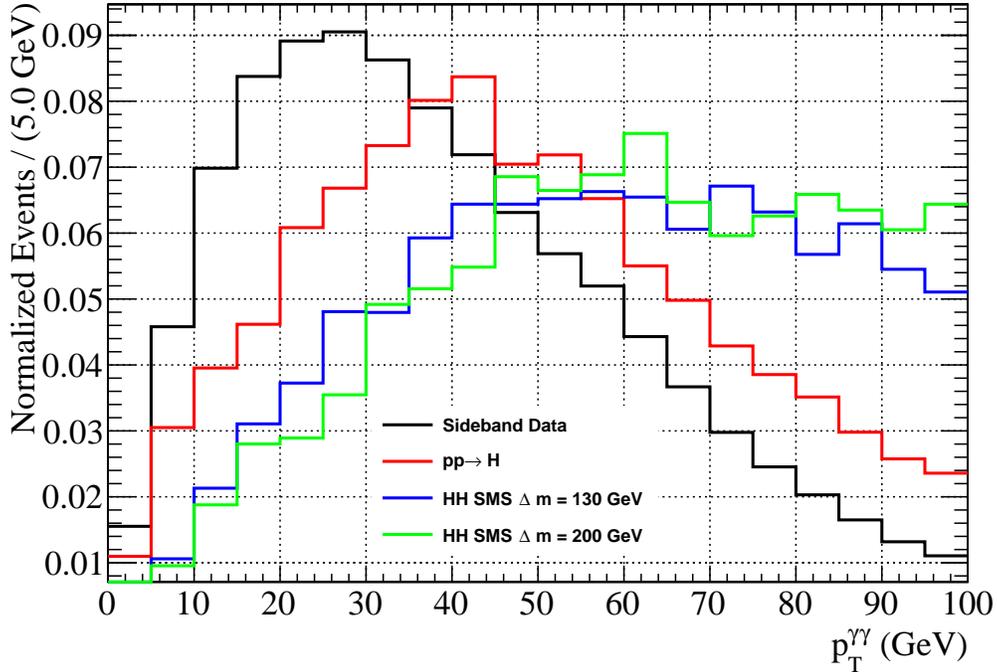


Figure 5.3: The distribution of the p_T of the di-photon system for data in the $m_{\gamma\gamma}$ sideband (black), SM Higgs boson MC (red) and the SMS with $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0$ with $m_{\chi_0} = 1$ GeV and $m_{\chi_2} = 130$ GeV (blue) or 200 GeV (green). The distribution is truncated $p_T^{\gamma\gamma} = 100$ GeV to show the low p_T behavior of the distributions for optimizing the selection. The selected cut is at $p_T^{\gamma\gamma} > 20$ GeV, where there is virtually no contribution from SUSY-like signals.

5.3.2 Jet and E_T^M Selection

We use anti- k_T clustered particle flow jets with jet corrections applied as described in Section 4.3. We consider all such jets in the event and select those meeting the following criteria:

- $p_T \geq 30$ GeV

- $|\eta| < 3.0$
- $\Delta R(jet, \gamma) > 0.5$ for both selected photons

The p_T requirement on the jet is motivated by the difficulty of accurately reconstructing low p_T jets as well as the observation that the jets produced in the signal typically have a relatively hard spectrum. We require the jets to be in the HBHE region since the resolution for jets falling in HF is worse than that in the HBHE and it is much more difficult to remove contributions from pileup in the absence of the tracker and ECAL. Finally, we require that the jets be far away from each photon. Because the jets are constructed from all pf-candidates, the photons in the event are usually reconstructed also as jets. Rather than go through the pf-candidate collection and remove the photons (which is error prone, since the pf-Photon reconstruction differs from the regular photon reconstruction) it is simpler and more effective to simply veto any reconstructed jet that would contain either photon within its cone.

We require at least one jet in the event to pass our jet selection, which is required since we need at least 2 objects to compute the razor variables (described in Section 5.4). Since our signal typically produces a Higgs boson and at least one other object, this is a relatively mild requirement.

We use E_T^M reconstructed from pf-candidates and corrected as described in Section 4.4. We apply a full suite of filters to the events, vetoing any event which has a possible contribution from detector noise. We also require at least 1 valid primary vertex to be reconstructed in the event.

5.3.3 Selection of B-tagged Jets

In all selected events, each jet is analysed with the CSV algorithm to test its probability of originating from a b-quark as described in Section 4.3.3. The goal is to identify the presence of either $H \rightarrow b\bar{b}$ or $Z \rightarrow b\bar{b}$ in the event in addition to the photon pair. In each event we compute two values $m_{bb,H}$ and $m_{bb,Z}$, which will be used to test compatibility with these hypotheses.

To compute the values of these variables, we use the following algorithm:

1. define $m_{bb,H} = 0$ and $m_{bb,Z} = 0$
2. let $J = \{j_b\}$ be the set of all jets in the event passing the CSVL b-tag
3. if $J = \emptyset$ **exit**
4. let $P = \{(j_b^\alpha, j_b^\beta) : j_b^\alpha \in J, j_b^\beta \in J, j_b^\alpha \neq j_b^\beta\}$ be the set of all pairs of jets in J
5. if $P = \emptyset$ **exit**
6. for all $(j_b^\alpha, j_b^\beta) \in P$:
 - (a) if j_b^α does not pass CSVM and j_b^β does not pass CSVM goto 6
 - (b) compute the invariant mass of the b-jets: m_{bb}
 - (c) if $|m_{bb,H} - 125| > |m_{bb} - 125|$ then set $m_{bb,H} = m_{bb}$
 - (d) if $|m_{bb,Z} - 91.2| > |m_{bb} - 91.2|$ then set $m_{bb,Z} = m_{bb}$

This algorithm finds the pair of b-tag jets, one passing CSVL and one passing CSVM, that have invariant mass closest to the mass of the Higgs boson and Z, and sets $m_{bb,H}$ and $m_{bb,Z}$ to 0 if these cannot be found. These variables will be used to assign events to different categories in Section 5.3.4, in order to find an enriched sample of events consistent with $HH+X \rightarrow \gamma\gamma bb+X$ or $HZ+X \rightarrow \gamma\gamma bb+X$. Events without b-tagged jets, with only 1 b-tagged jet, or with two that don't pass the CSV requirements are *not* rejected from the rest of the analysis, they will simply have $m_{bb,H} = m_{bb,Z} = 0$ and be classified into one of the three non-btagged boxes.

5.3.4 Event Boxes

With the events and objects selected, we divide the events into disjoint event categories (called boxes), based on their kinematics and objects. The boxes assignment is hierarchical and exclusive: an event is assigned only to the first box it satisfies regardless of its other properties. The ordering of the boxes is based upon the capacity to provide improved S/B for the simplified models targetted in this analysis. We design the final two boxes to be largely model-independent and also to capture

any event that is distinguished from background by the razor variables. These events do not have kinematics consistent with one of our models, but are very important to maintain sensitivity to other potential models that this analysis can capture. We use several variables to form these boxes:

- $p_T^{\gamma\gamma}$: the p_T of the selected di-photon system
- $m_{bb,H}$ and $m_{bb,Z}$: The masses of any b-quark pairs, defined in Section 5.3.3
- σ_E/E : the resolutions of each of the photons, measured from the regression

The boxes are:

1. **HighPt Box:** $p_T^{\gamma\gamma} > 110$ GeV
2. **Hbb Box:** $110 \leq m_{bb,H} \leq 140$
3. **Zbb Box:** $76 \leq m_{bb,Z} \leq 106$
4. **HighRes Box:** Both photons in the di-photon system have $\sigma_E/E < 0.015$
5. **LowRes Box:** All other events

The heirarchy of the boxes is the order in which they are enumerated. A schematic is also shown in Figure 5.4.

As noted before, a given event will be placed in only in the highest box whose criteria it meets, so an event in the Zbb Box would have: $p_T^{\gamma\gamma} \leq 110$ GeV, $m_{bb,H} \notin [110, 140]$, and $m_{bb,Z} \in [76, 106]$. An event in the Zbb box may have more than one pair of b-tag jets, but no pair will have mass $110 < m_{bb} < 140$ (or else it would fall in the Hbb box). Likewise, an event in the HighRes box may have any number of b-tag jets (including 0), but no pair, with at least one CSVN jet, may have $110 < m_{bb} < 140$ or $76 < m_{bb} < 106$ (or it would have fallen into one of the Hbb or Zbb boxes). This implies that the number of jets and number of b-tag distributions differ between the Hbb/Zbb boxes and the others.

The different kinematic cuts in each box and the resultant difference in the object distribution is the primary reason that no extrapolation or prediction used in one

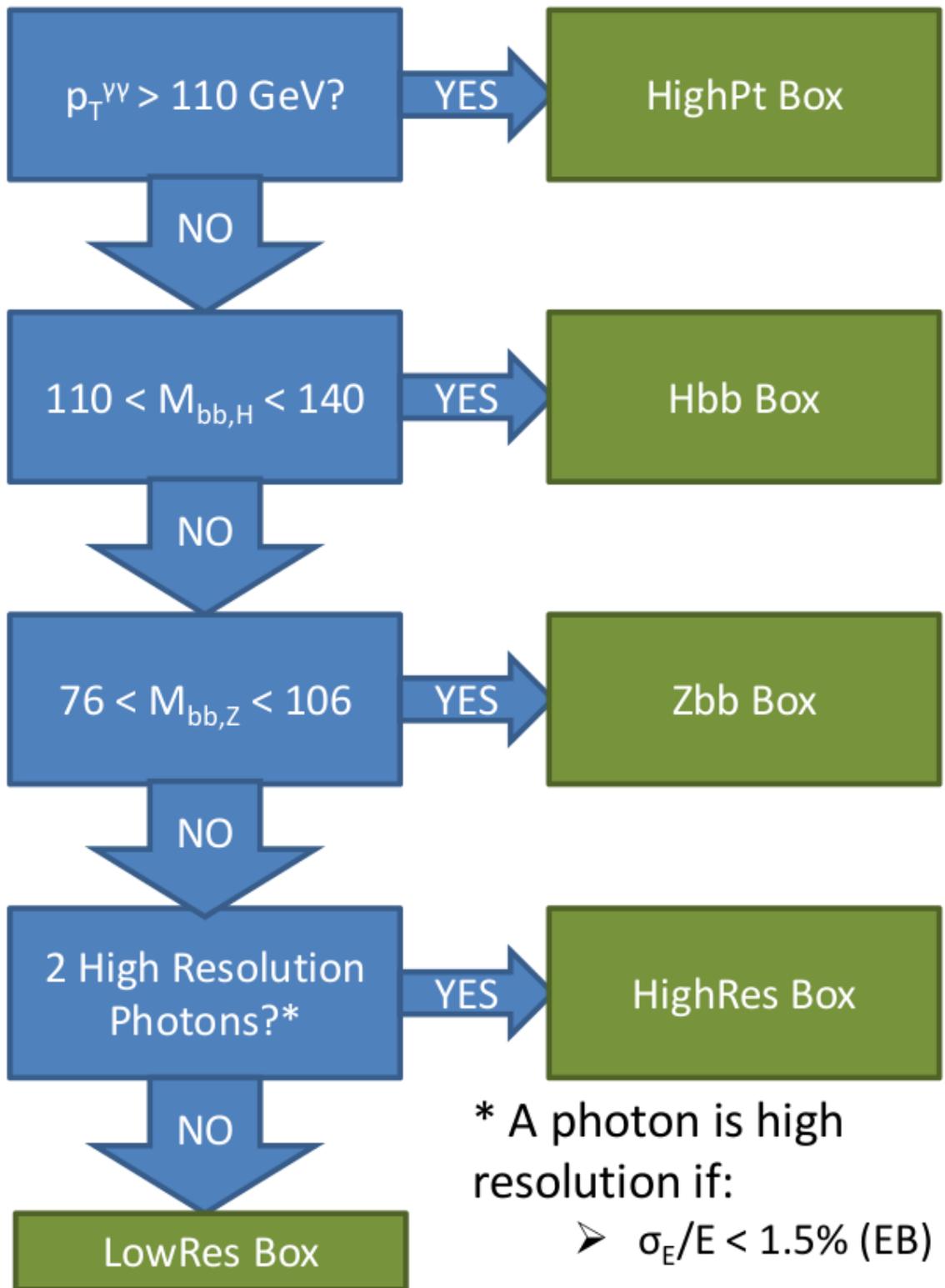


Figure 5.4: Hierarchy of event boxes.

box is ever applied in any other box. The shape of the $m_{\gamma\gamma}$ distribution, for example, will be very different for events with boosted di-photon systems (HighPt box) and for events with a high mass pair of b-tagged jets. Similarly, we expect that the resolution of a resonance will be much sharper in events with two high resolution photons than in events with at least one poor resolution photon. Therefore, while we can extrapolate in the $m_{\gamma\gamma}$ and $R^2 - M_R$ planes within a box (with suitable systematic errors), we never extrapolate between boxes, since the kinematics are fundamentally different.

The motivation for the three boxes comes from a study of the kinematics of the simplified models we are interested in. For events where there is a large mass splitting between the pair-produced particle and the LSP, we expect the Higgs boson produced in the decay chain to be boosted, and hence preferentially populate the **HighPt box**. We expect this box to have low background, as non-resonant processes are less likely to produce a di-photon pair that is both high mass and highly boosted. For events where there is a second Higgs boson produced in, for example, a second similar decay chain, we would expect the majority of these events to decay to $b\bar{b}$ pairs resonant at the Higgs boson mass. Similarly, if there is a Z-boson produced in the event, we would expect to see some population of events with a $b\bar{b}$ near the Z-resonance. One can get a sense of how the box selection relates to the kinematics by examining Figure 5.5.

Tables 5.2 and 5.3 shows the expected yield of background, SM Higgs boson, and SMS events in the five analysis boxes. We can see that both SMSs should produce most of their events in the HighPt box, consistent with our argument that the Higgs boson system should be boosted in the models we consider. In contrast, the SM production of the Higgs boson is predominantly in the HighRes box. Furthermore, we can see that the highest S/B (and S/\sqrt{B}) for the $\chi_2\chi_2$ SMS occurs in the Hbb and Zbb boxes, consistent with the rarity of final states with resonant $b\bar{b}$ and $\gamma\gamma$ pairs. Finally we can see that the HighRes box is significantly more powerful than the LowRes box in terms of signal significance, consistent with the selection of high resolution photons.

The final two categories are used to capture signal events in a less model-dependent way. By selecting high resolution photons in the HighRes box we reject fake photons,

Box	Combinatoric background	Peaking background	Total background
HighPt	$348.9 \pm +40.6$	$22.9^{+2.2}_{-2.1}$	$371.9^{+40.6}_{-40.6}$
Hbb	$3.3 \pm +2.5$	$0.1^{+0.0}_{-0.0}$	$3.5^{+2.5}_{-2.5}$
Zbb	$6.1 \pm +3.6$	$0.2^{+0.0}_{-0.0}$	$6.3^{+3.6}_{-3.6}$
HighRes	$829.0 \pm +74.4$	$34.9^{+3.1}_{-2.9}$	$863.9^{+74.5}_{-74.5}$
LowRes	$1635.8 \pm +89.0$	$14.5^{+1.4}_{-1.3}$	$1650.3^{+89.1}_{-89.1}$

Table 5.2: The expected number of events in the signal region of each box for background processes. The events are required to have $M_R > 150$ GeV and no selection is applied on R^2 . Monte Carlo yields are set to the expected number of events in $19.8 fb^{-1}$ of $\sqrt{\hat{s}} = 8\text{TeV}$ collisions with the Higgs boson cross-section is set to $\mu = 1.00 \times \mu_{SM}$.

Box	$\tilde{\chi}_2^0 \tilde{\chi}_2^0$ production				$\tilde{\chi}_2^0 \tilde{\chi}_1^\pm$ production			
	$m_{\tilde{\chi}_2^0} = 130$ GeV		$m_{\tilde{\chi}_2^0} = 175$ GeV		$m_{\tilde{\chi}_2^0} = 130$ GeV		$m_{\tilde{\chi}_2^0} = 175$ GeV	
	Efficiency	Yield	Efficiency	Yield	Efficiency	Yield	Efficiency	Yield
HighPt	$8.6^{+0.5\%}_{-0.5\%}$	$12.6^{+0.7}_{-0.7}$	$11.1^{+0.6\%}_{-0.6\%}$	$5.1^{+0.3}_{-0.3}$	$8.9^{+0.6\%}_{-0.6\%}$	$16.6^{+1.0}_{-1.0}$	$10.7^{+0.5\%}_{-0.5\%}$	$6.4^{+0.3}_{-0.3}$
Hbb	$1.8^{+0.1\%}_{-0.1\%}$	$2.6^{+0.2}_{-0.2}$	$1.8^{+0.1\%}_{-0.1\%}$	$0.8^{+0.1}_{-0.1}$	$0.0^{+0.0\%}_{-0.0\%}$	$0.0^{+0.0}_{-0.0}$	$0.0^{+0.0\%}_{-0.0\%}$	$0.0^{+0.0}_{-0.0}$
Zbb	$0.6^{+0.0\%}_{-0.1\%}$	$0.9^{+0.1}_{-0.1}$	$0.7^{+0.1\%}_{-0.1\%}$	$0.3^{+0.0}_{-0.0}$	$0.0^{+0.0\%}_{-0.0\%}$	$0.0^{+0.0}_{-0.0}$	$0.0^{+0.0\%}_{-0.0\%}$	$0.0^{+0.0}_{-0.0}$
HighRes	$4.7^{+0.2\%}_{-0.2\%}$	$6.9^{+0.3}_{-0.4}$	$5.3^{+0.3\%}_{-0.3\%}$	$2.5^{+0.1}_{-0.1}$	$5.2^{+0.3\%}_{-0.3\%}$	$9.7^{+0.6}_{-0.6}$	$6.5^{+0.3\%}_{-0.3\%}$	$3.9^{+0.2}_{-0.2}$
LowRes	$1.9^{+0.1\%}_{-0.1\%}$	$2.8^{+0.1}_{-0.2}$	$2.2^{+0.1\%}_{-0.1\%}$	$1.0^{+0.1}_{-0.1}$	$2.2^{+0.1\%}_{-0.1\%}$	$4.1^{+0.3}_{-0.3}$	$2.6^{+0.1\%}_{-0.1\%}$	$1.6^{+0.1}_{-0.1}$

Table 5.3: The expected number of events in the signal region of each box several signal benchmark points. The events are required to have $M_R > 150$ GeV and no selection is applied on R^2 . Monte Carlo yields are set to the expected number of events in $19.8 fb^{-1}$ of $\sqrt{\hat{s}} = 8\text{TeV}$ collisions with the chargino and neutralino cross-sections are set to their NLO values.

which typically have wide clusters and thus worse resolutions, and we select from among the real photons those that were better measured and more likely to produce a narrow resonance.

5.4 The Razor Kinematic Variables

5.4.1 General Definition

The razor variables are designed based on a generic process of pair production of heavy particles, each decaying to a visible particle and an undetected particle. The variables characterize, event-by-event, whether an observed event is compatible with this sort of process [82]. These variables have been used previously by the CMS [83, 84] and

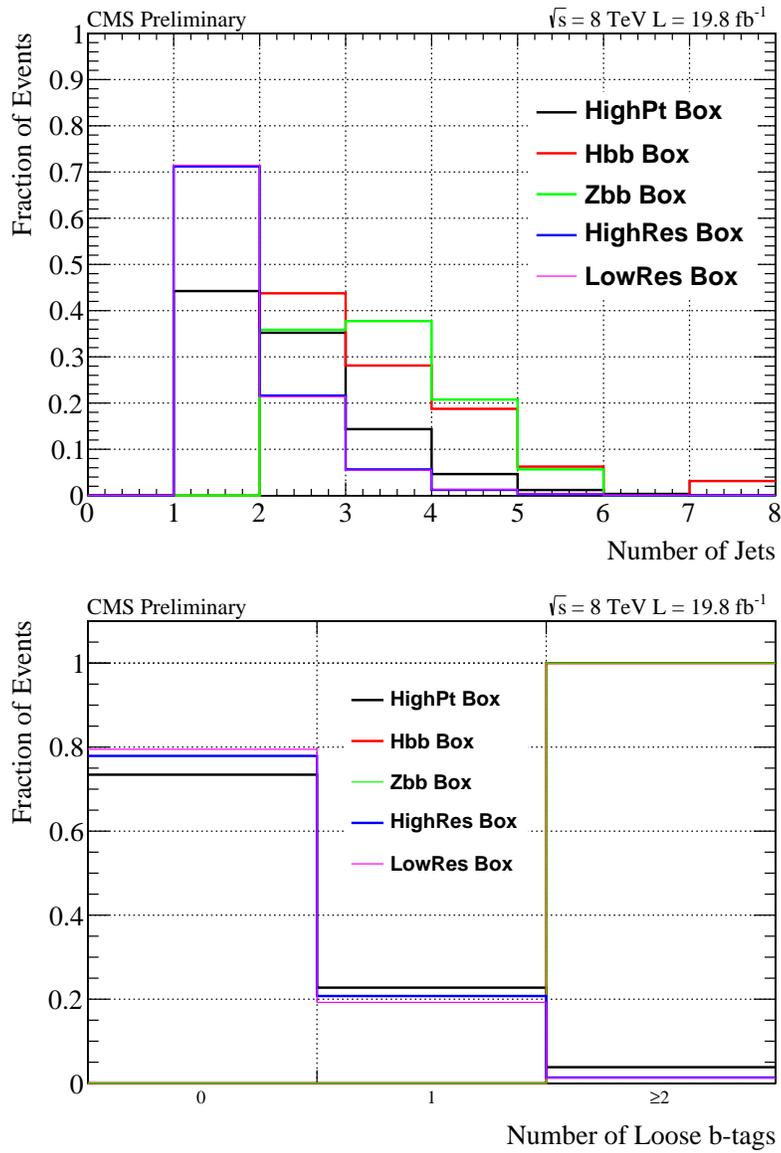


Figure 5.5: The number of jets (top) and number of CSVM b-tag jets (bottom) for the 5 analysis boxes.

ATLAS [85] collaborations to search for processes where a pair of heavy squarks are produced, with each squark decaying $\tilde{q} \rightarrow q\chi_0$, where χ_0 is the LSP and q represents a quark. If we let \vec{p}^{j1} \vec{p}^{j2} be the momenta of the two visible jets, and $p_z^{ji} \equiv \vec{p}^{ji} \cdot \hat{z}$ be the z-component of the momentum, then we define the variable M_R :

$$M_R \equiv \sqrt{(|\vec{p}^{j1}| + |\vec{p}^{j2}|)^2 - (p_z^{j1} + p_z^{j2})^2} \quad (5.1)$$

This variable is, by construction, invariant under longitudinal boosts. In the limit where the initial state has small transverse boost, we find that

$$M_R \approx \gamma_\Delta M_\Delta = \gamma_\Delta \frac{M_{\tilde{q}}^2 - M_{\chi_0}^2}{M_{\tilde{q}}} \quad (5.2)$$

where γ_Δ is the boost factor from the center-of-mass frame to the di-squark rest frame, $M_{\tilde{q}}$ is the mass of the heavy pair-produced particle (squark), and M_{χ_0} is the mass of the undetected particle (LSP). In the limit where the mass of the LSP is small compared to the squark, we then expect $M_R \approx M_{\tilde{q}}$, so M_R estimates the mass of the squark. When we relax our conditions on the initial state transverse boost, we find that the value of M_R will peak near $\gamma_\Delta M_\Delta$.

A second variable M_T^R that uses only transverse information is defined:

$$M_T^R \equiv \frac{1}{\sqrt{2}} \sqrt{E_T^M (p_T^{j1} + p_T^{j2}) - \vec{E}_T^M \cdot (\vec{p}_T^{j1} + \vec{p}_T^{j2})} \quad (5.3)$$

where \vec{p}_T^{ji} are the transverse momenta of the two jets and $p_T^{ji} \equiv |\vec{p}_T^{ji}|$. For perfectly measured jet p_T and E_T^M , M_T^R is designed to have a kinematic endpoint M_Δ . We then define the razor dimensionless ratio:

$$R \equiv \frac{M_T^R}{M_R} \quad (5.4)$$

Other analyses have found that background has a simple exponential shape in the variable R^2 , so we will use this for consistency with the published literature even though we do not explicitly rely on that shape in this analysis [86].

Thus far we have considered only the case where each heavy particle decays to exactly one visible object and one invisible object. To be more general, we must treat the case where each heavy particle decays to many objects, and the possibility that the decays are not symmetric. We do this by taking all visible objects in the event and assigning them to one of two mega-jets, which are constructed as the sum of the four-vectors of their components. The assignment is done by finding the configuration that minimizes $M_{hem\ 1}^2 + M_{hem\ 2}^2$. With these mega-jets, we have now forced an arbitrary event configuration into a di-jet-like topology. We then compute M_R and M_T^R exactly as in equations 5.1 and 5.3 with the four-vectors of the two mega-jets used as j_1 and j_2 .

5.4.2 Modification of the Razor Variables for $H \rightarrow \gamma\gamma$ topologies

The variables defined in Section 5.4.1 are designed to be useful for any generic signal topology. For our specific case, where we select events that have a high invariant mass $\gamma\gamma$ pair, we modify the definition slightly to incorporate our topology. Specifically, we modify the mega-jet assignment algorithm to force the two photons to always be placed in the same mega-jet, which is done by giving the mega-jet algorithm a collection consisting of the the four vectors of all the selected jets in the event and the four vector of the rest frame of the di-photon system (remember that the photons are specifically excluded from the selected jet collection by the ΔR cuts). The algorithm minimizes the sum in quadrature of the invariant mass of the mega-jets as usual.

With these ingredients, we can investigate the phenomenology of the razor variables for the signal topology of interest. Figure 5.6 shows the M_R , M_T^R and R^2 distributions for the model $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0 \rightarrow b\bar{b}\gamma\gamma\chi_0\chi_0$ at two mass splitting hypotheses. We see the behavior predicted for the variable in Section 5.4.1: the M_R distribution is peaking near the mass splitting between the heavy particle (χ_2) and the invisible particle (χ_0). The events have $M_T^R \lesssim M_\Delta$, where the inequality isn't an exact edge because of the imperfect reconstruction of jets and E_T^M . Figure 5.7 shows the distri-

bution of events in the $R^2 - M_R$ plane for signal-like events and for standard model $pp \rightarrow \gamma\gamma + X$ background (where we require $100 < m_{\gamma\gamma}$).

The name “razor” comes from the ability of the R^2 variable to cut away SM background while leaving selecting signals with heavy resonances. This is illustrated in Figure 5.8, which shows the M_R distribution for background and several signal hypotheses for various R^2 cuts. One sees that the background is heavily suppressed by the R^2 cut, without compromising signal efficiency. Not only is the overall yield of background suppressed by the razor cut, but the exponential slope is also markedly changed, which further improves the sensitivity to models with moderate to high M_R , since a mild R^2 cut increases the potential significance of a signal.

5.5 Background Prediction

There are two major categories of background for this analysis: background from standard model production of the Higgs boson, and background from other processes that produce photons (or objects we mis-identify as photons). We will call these categories “resonant” and “non-resonant,” respectively, and will treat them differently. The non-resonant background is a mixture of many different processes, some of which are quite difficult to model. There are, for instance, contributions from rare cases where two jets are produced through QCD processes and both jets are mis-identified as photons, which occurs in roughly 1 in 10^5 QCD interactions that produce a jet with $p_T > 40$ GeV. Creating a sufficiently high statistics sample of such events requires production of an intractably large number of simulated events (10 million produced events was equivalent to only 189 pb^{-1} of data). Because of these difficulties, we use data to model the background from non-resonant processes. For the resonant processes, we use MC simulation of SM $H \rightarrow \gamma\gamma$ events. We combine simulated events produced through the 4 major production channels (listed in Table 2.4) weighted by their relative production cross section.

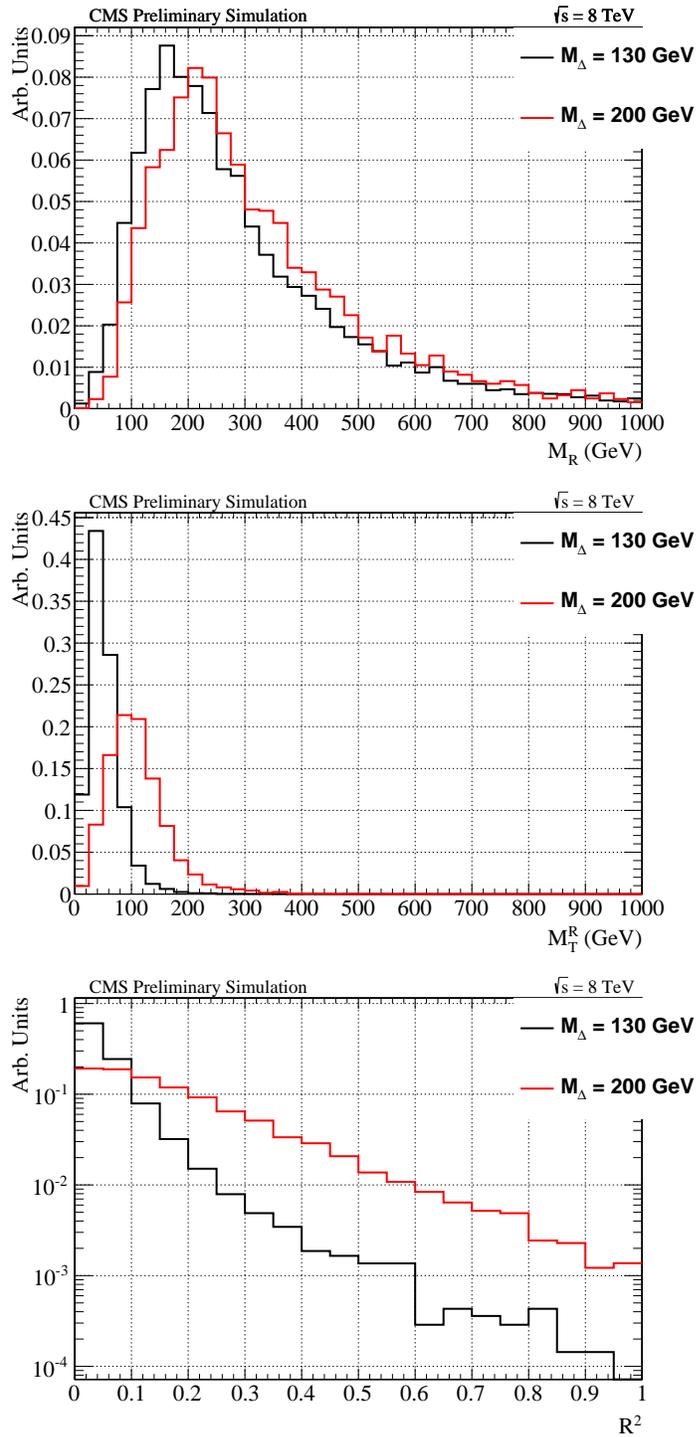


Figure 5.6: The distributions of M_R (top), M_T^R (middle), and R^2 (bottom) for the model $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0 \rightarrow bb\gamma\gamma\chi_0\chi_0$ with $m_{\chi_0} = 1 \text{ GeV}$ and $m_{\chi_2} = 130 \text{ GeV}$ (black) or $m_{\chi_2} = 200 \text{ GeV}$ (red).

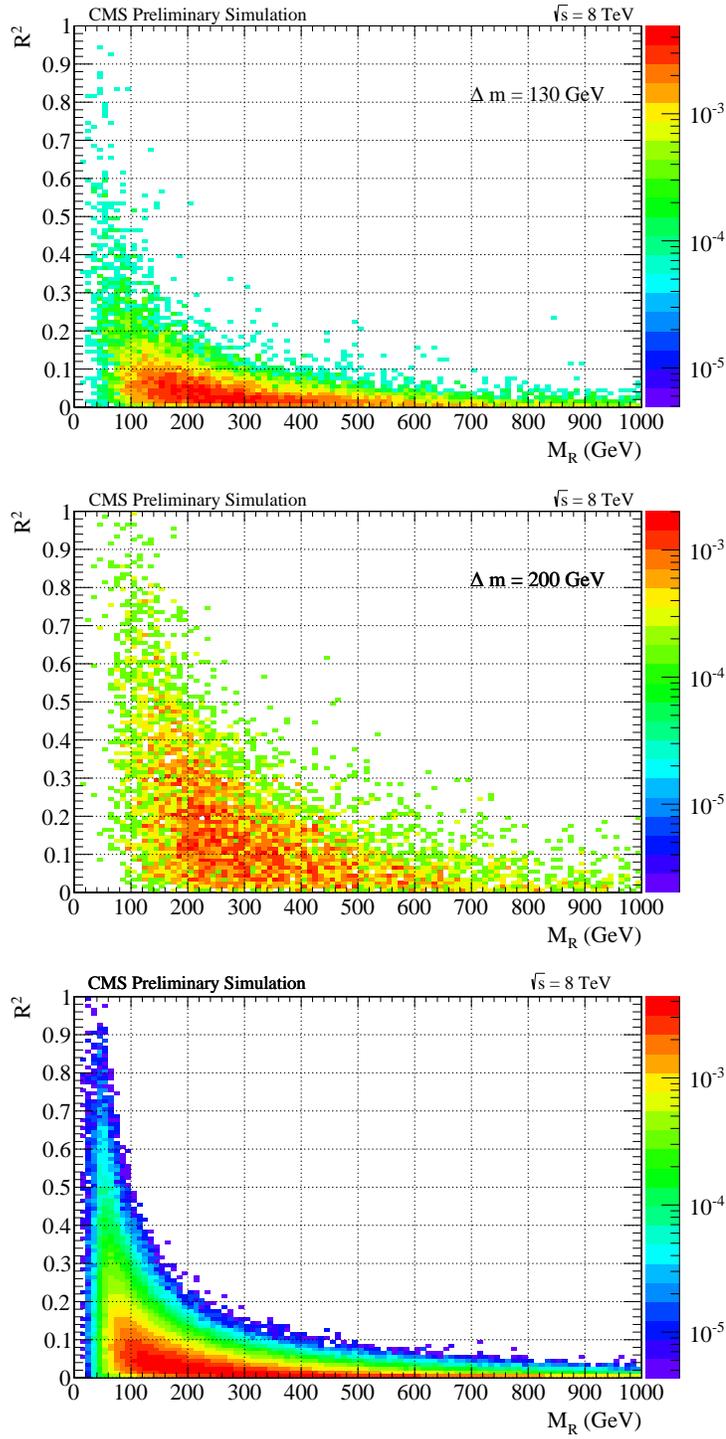


Figure 5.7: The distributions of R^2 versus M_R for the model $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0 \rightarrow bb\gamma\gamma\chi_0\chi_0$ with $m_{\chi_0} = 1 \text{ GeV}$ and $m_{\chi_2} = 130 \text{ GeV}$ (top) or $m_{\chi_2} = 200 \text{ GeV}$ (middle), along with the distribution for SM $pp \rightarrow \gamma\gamma + \text{jets}$ (bottom)

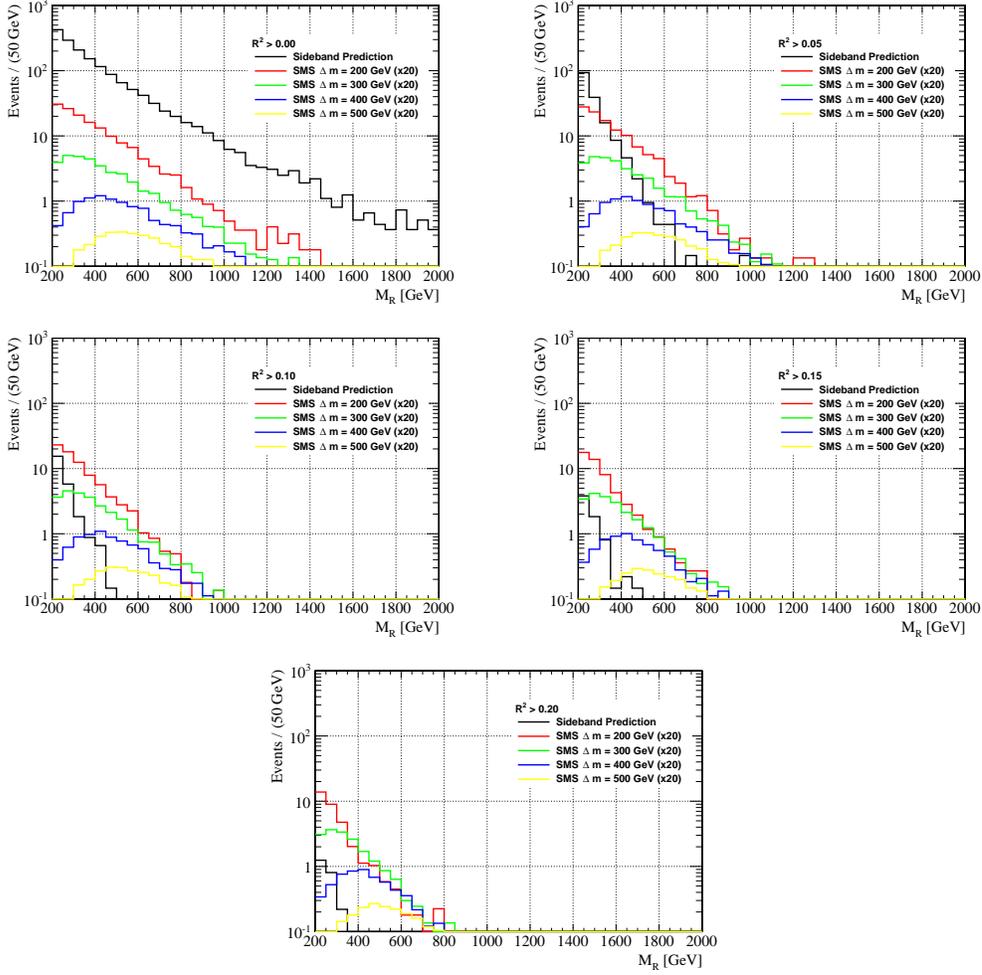


Figure 5.8: M_R distributions for $m_{\gamma\gamma}$ -sideband data (black) and $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0$ SMS events with $M_{\chi_0} = 1$ GeV and $\Delta M \approx M_{\chi_2}$ as indicated in the legend (expected yields scaled by factor 20), for events with R^2 cuts ranging from $R^2 > 0$ to $R^2 > 0.20$ as indicated in the legend. The normalization is the expected number of background events in data (see Section 5.5) for the data distribution and $20\times$ the expected number of signal events for each mass point.

5.5.1 $m_{\gamma\gamma}$ Signal Region

The first step is to define regions in the $m_{\gamma\gamma}$ spectrum where we expect both the photons from a Higgs boson decay (either SM or BSM production) to peak. While the Higgs boson has a very narrow width, the resolution of the photon reconstruction means that reconstructed $m_{\gamma\gamma}$ for the photon pair follows some (roughly Gaussian) distribution centered at m_H but with a few GeV width. The exact width is dependent on the kinematics of the photons, so we measure it in each box. We use MC produced with $m_H = 125$ GeV, select events and photon pairs in the manner described in Section 5.3, and then examine the distribution of $m_{\gamma\gamma}$. Figure 5.9 shows the distribution of $m_{\gamma\gamma}$ for SM Higgs bosons in the HighPt, HighRes, and LowRes boxes; there are not enough SM Higgs boson events in the Hbb and Zbb boxes to make a reliable distribution. For the three boxes with enough statistics, we compute in each box a value σ_{eff} , which is defined such that the interval $[125 - \sigma_{eff}, 125 + \sigma_{eff}]$ contains 68.2% of the expected SM Higgs boson events in the box. Since the distributions are not perfectly Gaussian, σ_{eff} is not quite the standard deviation of the sample, but it is an estimate of the Gaussian core of the distribution.

With σ_{eff} determined from the MC (we use $\sigma_{eff} = 2$ GeV for the Hbb and Zbb box where there are not enough MC statistics), we then set the signal region in each box to be $[125 - 2 \times \sigma_{eff}, 126 + 2 \times \sigma_{eff}]$. Since the signal is not perfectly Gaussian, $\pm 2\sigma_{eff}$ does *not* necessarily contain 95.5% of the events, but it is not too far away. The discrepancy between 125 for the lower bound and 126 for the upper bound is to take into account the current uncertainty on the mass of the Higgs boson (the PDG currently estimates the mass of the Higgs boson to be 125.7 ± 0.4 [3]). Table 5.4 shows the measured σ_{eff} and resulting signal regions in each box. As expected, our box with high resolution photons (HighRes box) has the smallest σ_{eff} , while the box explicitly containing poor resolution photons has the largest.

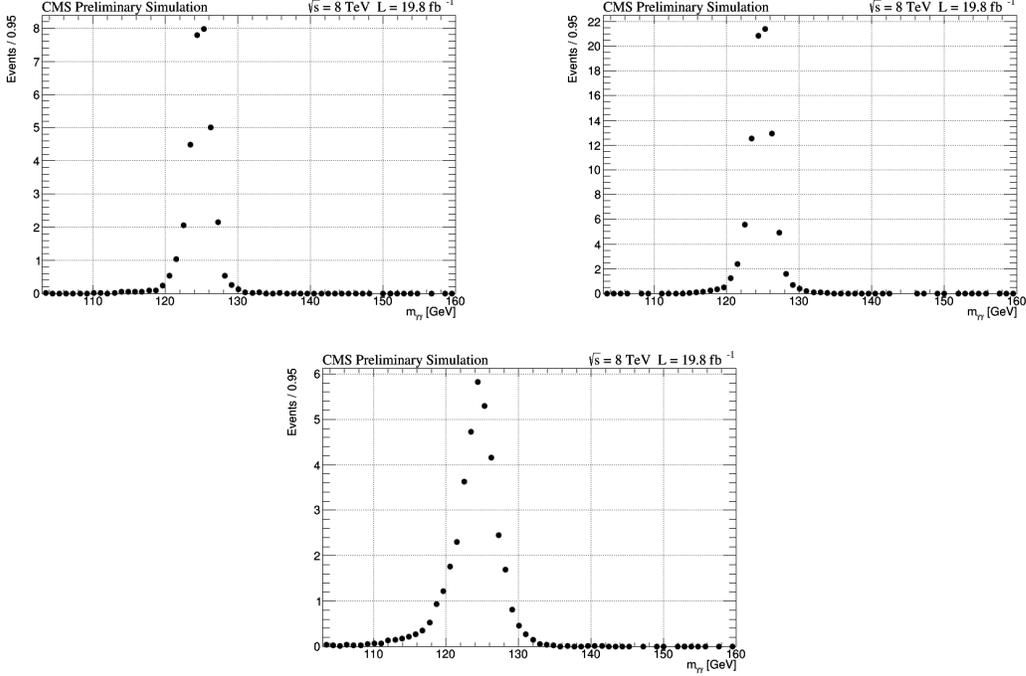


Figure 5.9: $m_{\gamma\gamma}$ distributions for selected standard model Higgs boson events. The events are divided into the HighPt Box (top left), HighRes Box (top right), and LowRes Box (bottom). The distributions are normalized to the expected number of events in the indicated integrated luminosity.

5.5.2 Non-Resonant Background Prediction

We model the shape of the non-resonant background in the $R^2 - M_R$ plane by taking the shape in $m_{\gamma\gamma}$ sidebands around the signal regions in each box. The ranges of the sideband regions are the same in the each box and are given in Table 5.5. The data from the upper and lower sideband is summed and the resulting shape in the $R^2 - M_R$ plane can be seen in Figure 5.10. One can see several important features of the razor variables from these distributions:

1. In all boxes except the HighPt box, the data is clustered at low M_R and R^2 . The kinematic requirements placed on the objects (30 GeV on the jet and 20 GeV on the di-photon system) prevent events from having $M_R \lesssim 50$ GeV, but after this turn on region, we see that the distribution is steeply falling in M_R .
2. In the HighPt box, the requirement on the di-photon ($p_T^{\gamma\gamma} > 110$ GeV) pushes

Box	σ_{eff}	Signal Region Range
HighPt	1.52 GeV	[121.96,129.04]
Hbb	2 GeV*	[121,130]
Zbb	2 GeV*	[121,130]
HighRes	1.48 GeV	[122.04,128.96]
LowRes	2.50 GeV	[120,131]

Table 5.4: The σ_{eff} and resulting signal region ranges for the different boxes. The σ_{eff} is computed from SM Higgs boson Monte Carlo and the Signal Region is defined as $[125 - 2 \times \sigma_{eff}, 126 + 2 \times \sigma_{eff}]$. *: For the Hbb and Zbb Categories, the MC statistics are insufficient to compute a reliable σ_{eff} , so it is set equal to 2 GeV.

the turn-on region up to higher M_R . Since these background events will mostly have the two photons recoiling against one or more jets of equal magnitude, this implies the existent of a jet system with $p_T > 110$ GeV. These two systems together make $M_R \lesssim 220$ GeV kinematically unfavorable. Once this turn-on region is passed, then the exponential fall begins.

3. The Hbb and Zbb boxes have very few events. This is by construction, since the final state $b\bar{b}\gamma\gamma$ is relatively rare and even rarer when the $b\bar{b}$ pair is required to have invariant mass near m_H or m_Z .

Lower Sideband	$103 < m_{\gamma\gamma} < 120$
Upper Sideband	$131 < m_{\gamma\gamma} < 160$

Table 5.5: The definitions of the Lower and Upper Sidebands.

With the shape of the non-resonant background determined, we determine the normalization of the prediction. This is done by doing a fit to the $m_{\gamma\gamma}$ distribution in each box in the disjoint region $m_{\gamma\gamma} \in [103, 120] \cup [131, 160]$ and continuing the function in the whole region $103 < m_{\gamma\gamma} < 160$. The fit function chosen is a double exponential defined as

$$p(m_{\gamma\gamma}; f, \alpha_1, \alpha_2) = fe^{-\alpha_1 m_{\gamma\gamma}} + (1 - f)e^{-\alpha_2 m_{\gamma\gamma}} \quad (5.5)$$

where the parameters are constrained $0 < f < 1$ and $\alpha_i > 0$. We consider several other possible functions and find the the double exponential is a good choice for

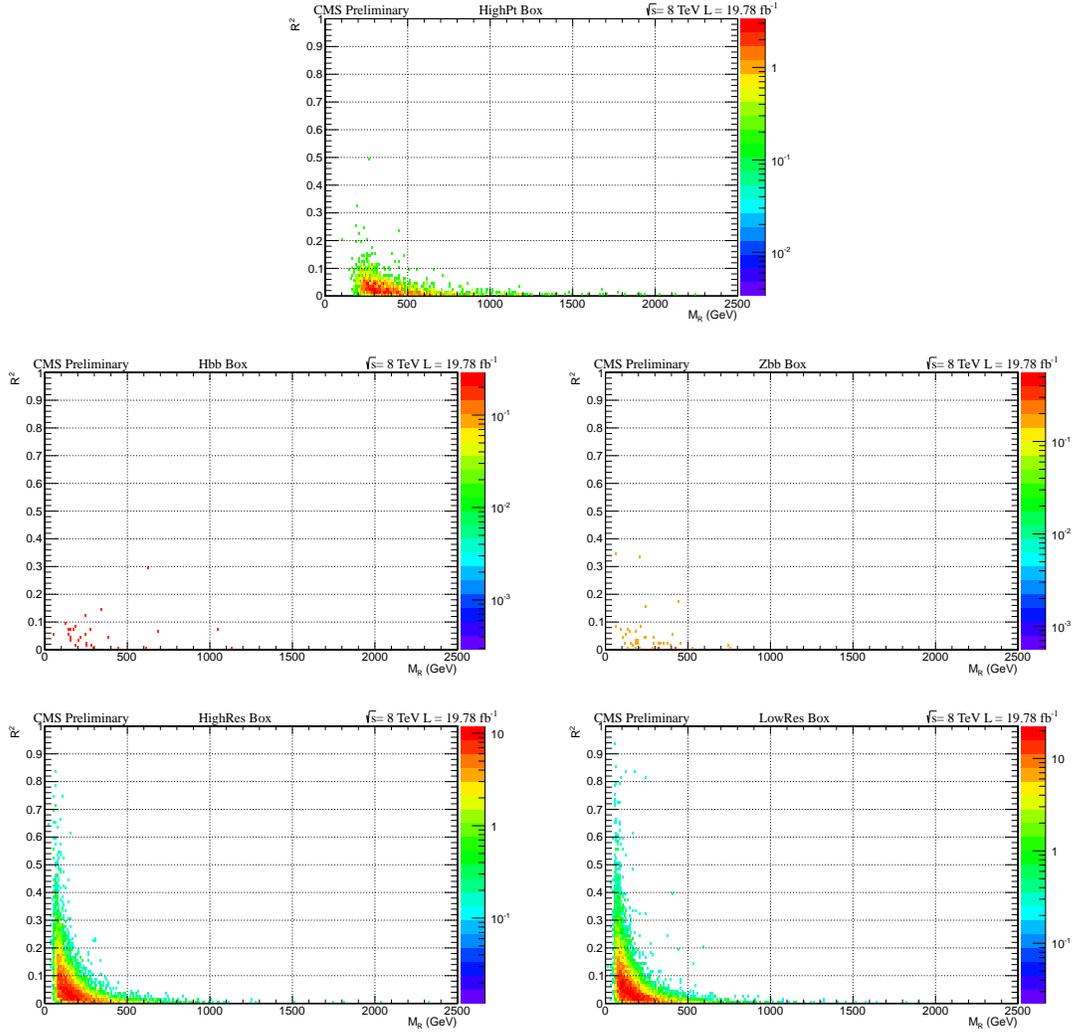


Figure 5.10: Distribution of events in the $R^2 - M_R$ plane taken from the data sideband regions. The plots correspond to the 5 data boxes: HighPt (top), Hbb (middle left), Zbb (middle right), HighRes (bottom left), and LowRes (bottom right).

our particular data, and use the very slight difference in the predicted shape as a systematic error on the final measurement; this is discussed further in Section 5.7.3. The fits with the $m_{\gamma\gamma}$ distribution blinded are shown in Figure 5.11. One sees that the selected shape is in good agreement with the data in all five boxes.

With the fit performed, the integral of the fitted shape in the signal region is taken as the predicted number of non-resonant background events in the $m_{\gamma\gamma}$ signal region. From there a scale factor is computed for each box, equal to the number of predicted events in the signal region divided by the actual number of events in the sideband regions; the scale factor gives the amount the distributions in Figure 5.10 must be scaled by to get the non-resonant background prediction. The derived scale factors are shown in Table 5.6. The errors here come from the uncertainty on the integral of the fitted function and the statistical error on the number of events in the sideband (though the former is in all cases much larger). The scaled M_R - R^2 distributions are shown in Figure 5.12. Here the binning is also coarser on the x and y axes to more accurately represent the resolution we have on the M_R and R^2 variables.

Box	Background Prediction Scale Factor
HightPt	0.160 ± 0.0046
Hbb	0.156 ± 0.045
Zbb	0.185 ± 0.053
HighRes	0.165 ± 0.0033
LowRes	0.260 ± 0.0030

Table 5.6: The scale factors derived from the fits to the $m_{\gamma\gamma}$ background. The errors come from Poisson sideband errors and the errors on the integral of the fit in the signal region.

5.5.3 Resonant Background

The resonant background prediction is taken as the shape in the R^2 - M_R plane of the SM Higgs boson MC within the signal regions defined in Table 5.4. The normalization is set by scaling the MC to the number of SM Higgs boson events expected in 19.8 fb^{-1} . The normalized distributions are shown in Figure 5.13, with the same binning as in Figure 5.12.

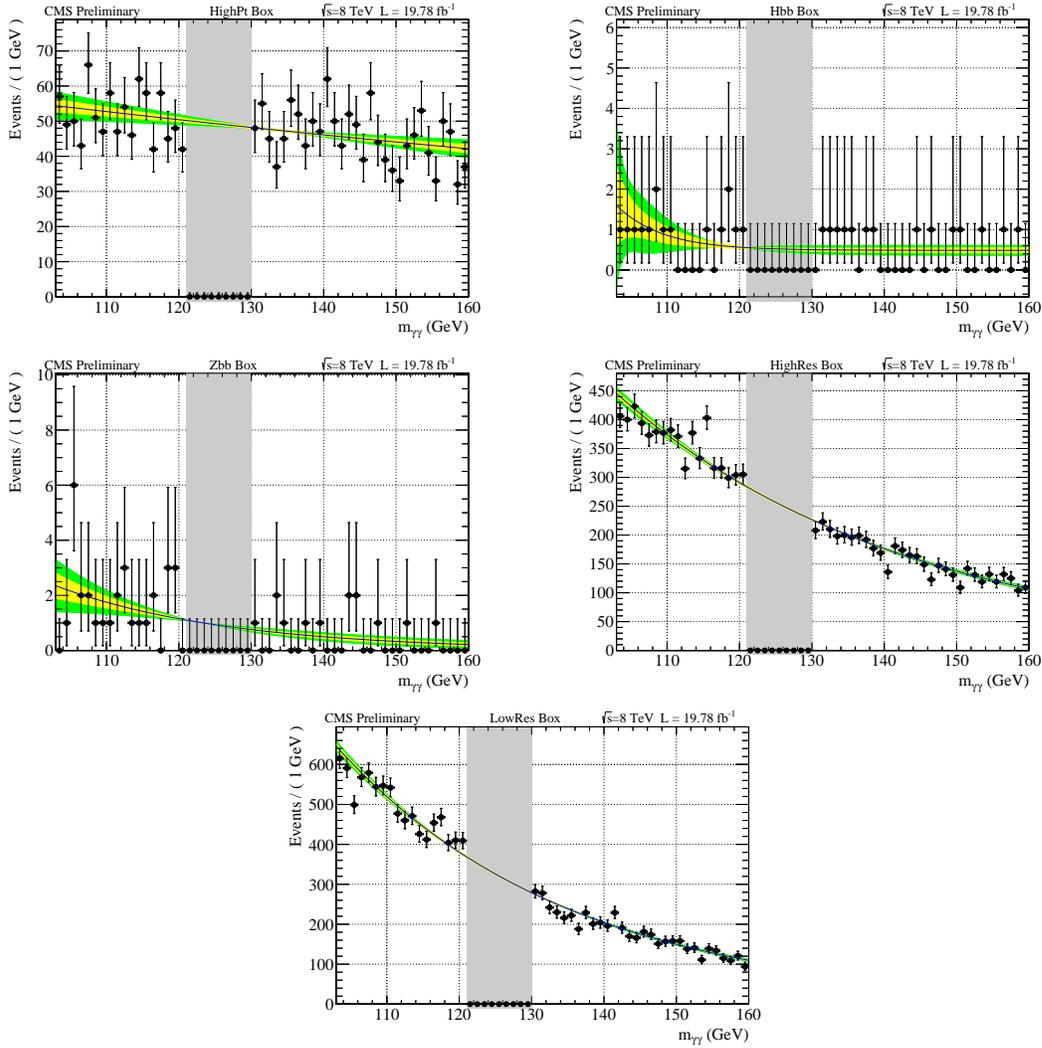


Figure 5.11: The blinded fits to the $m_{\gamma\gamma}$ distribution in the HighPt (top left), Hbb (top right), Zbb (middle left), HighRes (middle right), and LowRes (bottom) categories. Fits are shown with their $\pm 1\sigma$ (yellow) and $\pm 2\sigma$ (green) error bands. The gray region is the blinding window.

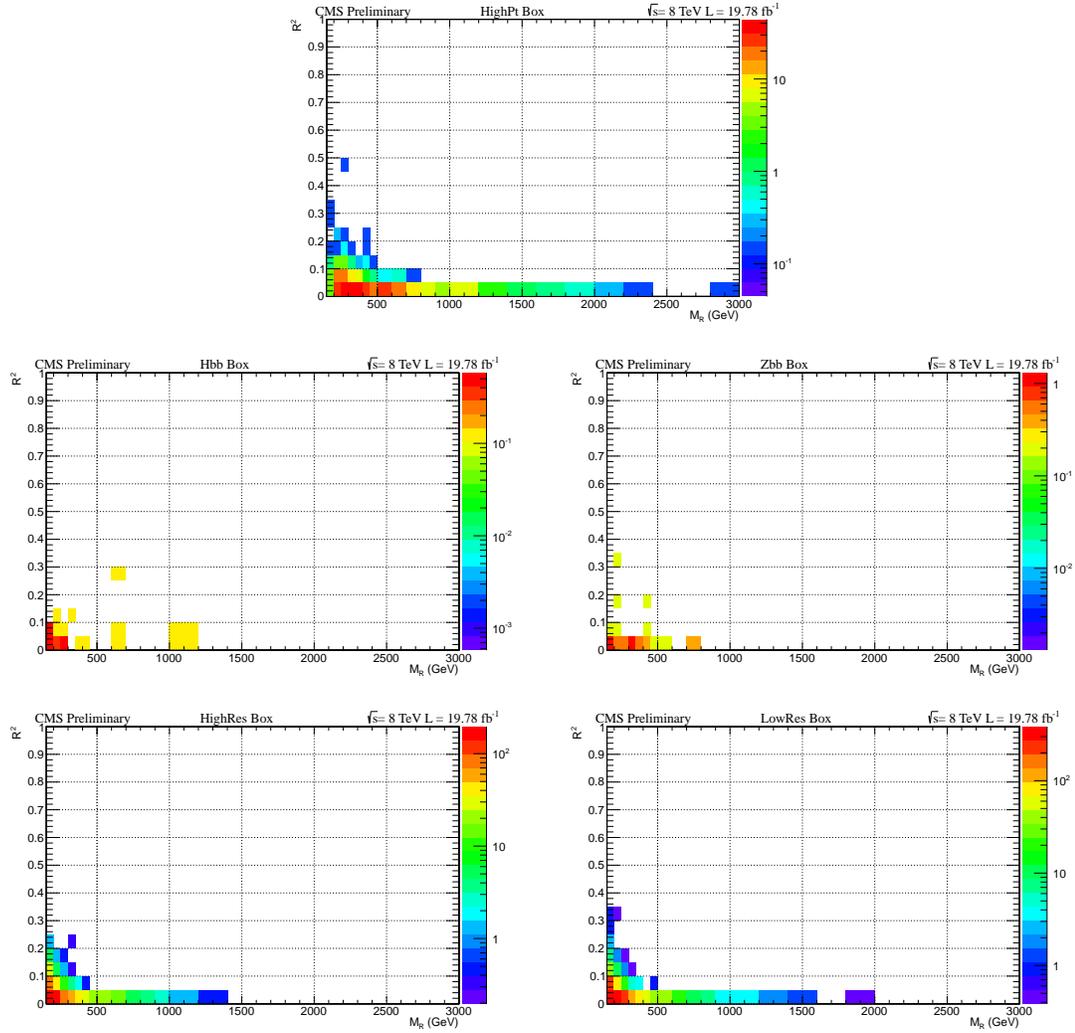


Figure 5.12: Distribution of events in the $R^2 - M_R$ plane taken from the data sideband regions. The plots correspond to the 5 data boxes: HighPt (top), Hbb (middle left), Zbb (middle right), HighRes (bottom left), and LowRes (bottom right). This plot contains the same events as Figure 5.10 but with coarser binning, corresponding more accurately to the actual resolution of the M_R and R^2 variables, and with the scale factors applied to show the actual background prediction.

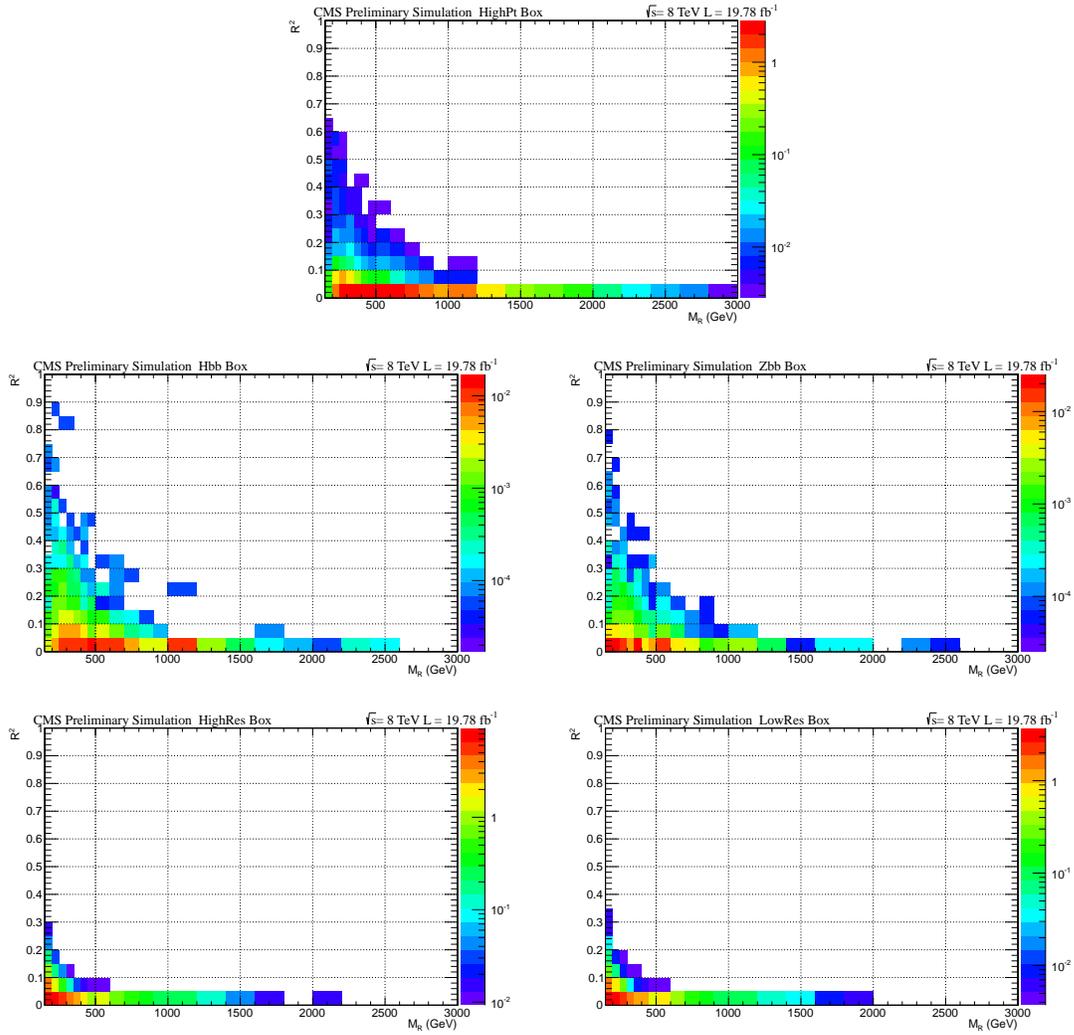


Figure 5.13: Distribution of events in the $R^2 - M_R$ plane taken from the SM Higgs boson MC signal regions. The plots correspond to the 5 data boxes: HighPt (top), Hbb (middle left), Zbb (middle right), HighRes (bottom left), and LowRes (bottom right). Note that the Z-axis scale is different in each box.

5.5.4 Sideband Extrapolation Validation

The extrapolation of the $R^2 - M_R$ shape from the $m_{\gamma\gamma}$ sideband region into the $m_{\gamma\gamma}$ signal region is extremely important for this analysis to function properly, so we perform several tests to make sure that it is valid. We begin by testing the extrapolation on events where we invert the photon isolation, which means we require at least one of the photons to *fail* the isolation. Since the isolation is designed to be highly efficient for selecting prompt photons, the inverse selects very few real photons. As we are looking for Higgs bosons decaying to two real photons, this inverted isolation sample has very few real Higgs boson events.

We predict ≈ 14 real $H \rightarrow \gamma\gamma$ events compared to ≈ 5100 background events in the signal region, which is a signal contamination of $\approx 0.27\%$ distributed over the entire razor plane. Looking at the plots in Figure 5.13 we can see that these events should largely occupy the low R^2 low M_R regions of the HighPt, HighRes, and LowRes boxes, where the background statistics are quite high. We therefore expect this region to be totally dominated by non-resonant background, and so we can perform the sideband extrapolation on this sample and compare it to the observed data in the signal region.

This does not violate our blinding policy, since we expect such a small contribution from SM (and BSM) Higgs boson events and since these events are in an orthogonal sample to our final selection. The spirit of the blinding policy is to avoid tuning cuts in a way that might enhance features of the observed data. This helps avoid a type I error where the analysis is tuned to enhance a fluctuation in the data. Since this data sample with the inverted isolation cut is explicitly orthogonal to our final selection, we are not sensitive to this sort of problem by looking at this data. While there is some contribution from SM Higgs in this sample, it is so small compared to the non-resonant background that it is virtually impossible to detect. We also compare only large regions in R^2 and M_R to make it harder to notice small clusters in the plane that could be evidence of a deviation.

The results of the test are shown in Figure 5.14. The plot on the bottom shows the observed yield in the inverted isolation signal region (defined as $121 < m_{\gamma\gamma} < 130$),

while the plot on the top shows the number of standard deviations of the observed yield from the background prediction, defined as

$$z = \frac{\text{observed} - \text{expected}}{\sqrt{\text{expected} \times (1 + f_s)}} \quad (5.6)$$

where f_s is the scale factor derived for the sample. The denominator is designed to be the sum in quadrature of the statistical error and the systematic error coming from the limited sideband statistics (described in detail in Section 5.7.1). This procedure tends to over-estimate the deviation in low statistics regions, where this Gaussian approximation breaks down, but even so we see excellent agreement between the observed yield in the signal region and the prediction derived from the sideband.

Since the inversion of the isolation changes the makeup of the background with respect to the real selection, we perform a second validation with the isolation applied normally, but with the regions we consider shifted. We cannot use data from $m_{\gamma\gamma} < 103$ GeV, because we risk significant contributions from $Z \rightarrow e^+e^-$, where both electrons are identified as photons. The Drell-Yan cross section is high enough that this can be significant in mass ranges compatible with the Z. This is partly a consequence of our choice of the conversion safe electron veto (Section 4.2.1), where we accept the lower electron rejection efficiency in order to gain the higher photon efficiency. This minor inconvenience for one cross-check is less important for the analysis than the higher global photon efficiency, since we have little Drell-Yan contamination in the main analysis. For this test we define a shifted sideband region of $m_{\gamma\gamma} \in [130, 135] \cup [150, 155]$ and a shifted signal region of $m_{\gamma\gamma} \in [140, 145]$, where we expect no peaking background. Figure 5.15 shows the results of the test and the available statistics. The statistical precision is worse than in the inverted isolation case, but we still see no deviation of the observed yield from the predicted yield.

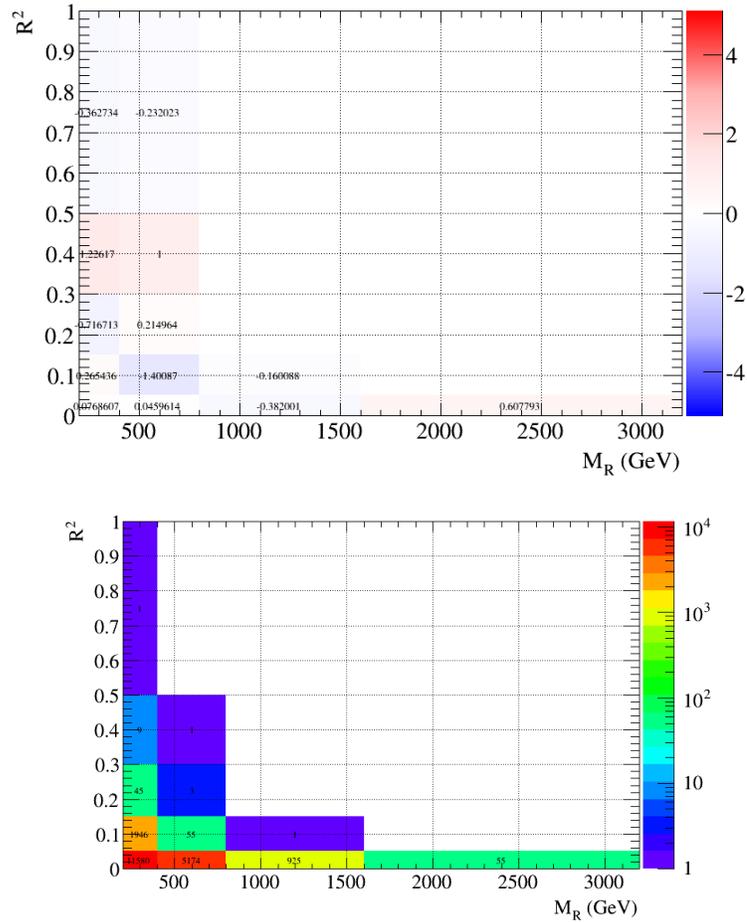


Figure 5.14: Validation of the $R^2 - M_R$ extrapolation of the sideband to the signal region with the selection altered so that at least one photon fails the isolation. (top) The actual yield in the signal region minus the prediction from the sidebands divided by the sum in quadrature of the statistical error of the signal region and the systematic uncertainty of the sideband prediction (from the sideband statistics). (bottom) The number of events in the signal region. The signal region is $m_{\gamma\gamma} \in [121, 130]$ and the sideband regions are $m_{\gamma\gamma} \in [103, 120]$ and $m_{\gamma\gamma} \in [131, 160]$.

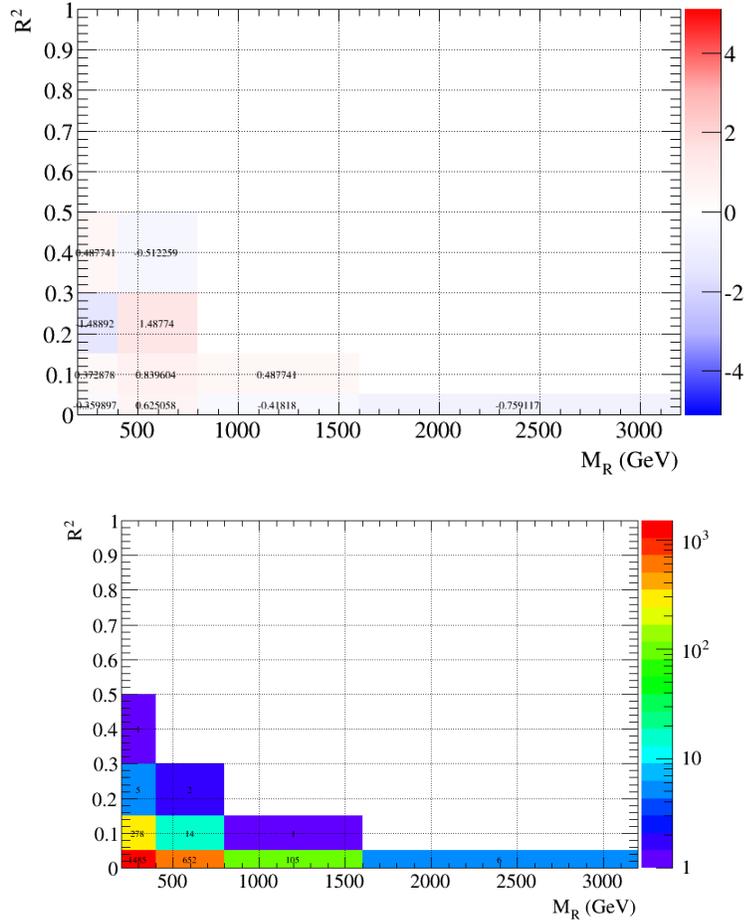


Figure 5.15: Validation of the $R^2 - M_R$ extrapolation of the sideband to the signal region with shifted $m_{\gamma\gamma}$ regions in data. (top) The actual yield in the signal region minus the prediction from the sidebands divided by the sum in quadrature of the statistical error of the signal region and the systematic uncertainty of the sideband prediction (from the sideband statistics). (middle) The percent difference of the signal region from the sideband prediction (computed as $[(\text{signal region}) - (\text{sideband region})]/(\text{signal region})$). (bottom) The number of events in the signal region. The signal region is $m_{\gamma\gamma} \in [140, 145]$ and the sideband regions are $m_{\gamma\gamma} \in [130, 135]$ and $m_{\gamma\gamma} \in [150, 155]$.

5.6 Signal Regions

One can see in Figure 5.12 that the limited statistics of the sideband sample requires coarser binning to provide meaningful background predictions. In order to take into account the falling exponential nature of the backgrounds and the expectation that the signal should have larger R^2 values than typical backgrounds, we create signal regions that are large at high values of M_R and R^2 and smaller at low values of M_R and R^2 , on the core of the background distribution.

The exact algorithm followed is described in Section 5.6.1; here we will explain qualitatively what is being done to develop the intuition for the algorithm. A SUSY-like signal can peak at any M_R value (dependent on the mass-splitting between the SUSY particles), and tends to occupy intermediate to high values of R^2 . A standard model process will follow an exponentially falling distribution in both R^2 and M_R , and so will tend to cluster in the lower left corner of the plane. With this in mind, the goal is to create signal regions that encompass narrow regions (approximately the resolution) in M_R and an R^2 range that cuts out most of the SM background. Looking at the plots in Figure 5.10, one can see that the natural set of signal bins would have a step-like shape, looking only at relatively high R^2 values for small values of M_R and looking at almost the entire R^2 range at high M_R .

One constraint we add to the problem is that we don't want signal regions with very small background predictions. Since we take our background prediction from the sidebands, there will be a granularity to the background we can predict in any given region of the plane equal to one sideband event \times the scale factor in the box we consider, e.g. the smallest non-zero background we can predict in the HighRes box is 0.16 events, since that is the scale factor in that box. Furthermore, having a very small background prediction means the systematic error becomes very large compared to the background prediction (see Section 5.7.1). Having bins with 0 predicted background creates a challenge, since it makes a single observed event extremely significant and needlessly complicates the statistical analysis. We therefore want to choose a set of signal regions that are predicted to never have fewer than one background event.

These considerations determine our procedure for setting the signal regions. We start from very high M_R and look at events in the whole $0 \leq R^2 < 1$ region. We integrate downward in M_R until the region we are looking at contains one predicted background event (sideband+SM Higgs boson). Once we reach that, we call this a signal bin. For the next bin, we don't want to look at the whole $0 \leq R^2 < 1$ region, because we know the low R^2 region will contain most of the background but little signal (again, this can be seen in Figure 5.10). We therefore step up by 0.05 in R^2 and start from the previous M_R edge we found considering the region from $0.05 \leq R^2 < 1$, again integrating to lower values in M_R until we find another region that contains one event. We call that region another signal region, move up again by 0.05 in R^2 and repeat until we reach $M_R = 150$, the lower M_R edge of our plane. In this way, we create a set of regions starting at high R^2 for low M_R and low R^2 for high M_R each going up to $R^2 = 1$. This procedure gives us a set of edges in M_R (the places where we moved up in R^2) and R^2 (the 0.05 steps); we bin the lower left corner of the plane by tracing those down to the M_R and R^2 axis, respectively.

Since the non-resonant background prediction comes from data and has limited statistics on the tails of the distribution, we perform this procedure on SM di-photon MC plus SM Higgs boson MC. This prevents our choice of signal regions from being unduly influenced by fluctuations on the tail of the sideband prediction. While the MC has much higher statistics, we also don't want to be too sensitive to any local fluctuations in the MC. In order to be agnostic to this, when we move downward in M_R to find the one event background, we do so in relatively large steps in M_R . These steps depend on the exact value of M_R we are considering and are designed to be roughly 10%-20% of the M_R value, which is close to the resolution of the M_R variable, as is discussed in the next section. The choice of the $R^2 = 0.05$ step is based on the observation that $0.00 \leq R^2 < 0.05$ contains most of the background events with $M_R > 500$ GeV.

5.6.1 Signal Region Algorithm

To create the signal regions, we follow a standard algorithm that was chosen and frozen before studying sideband data and, of course, before looking at signal region data. We choose the bins by examining background Monte Carlo and SM Higgs boson Monte Carlo. We form the full MC background distribution in each box by adding the SM background MC to the SM Higgs boson MC, and based on this distribution we define signal regions using the following algorithm. In the following, we use $p(M_R, R^2)$ to mean the full MC bkg distribution:

1. let $R_{cut}^2 = 0$ and $M_R^{cut} = 3000$
2. let $M_R^{edges} = \{3000\}$
3. define $I(M_R, R^2) \equiv \int_{M_R}^{\min(M_R^{edges})} dm \int_{R^2}^1 dr p(m, r)$
4. if $I(M_R^{cut}, R_{cut}^2) \geq 1$ goto 8
5. if $M_R^{cut} \leq 150$ goto 11
6. $M_R^{cut} = M_R^{cut} - \begin{cases} 50 & M_R^{cut} < 500 \\ 100 & 500 \leq M_R^{cut} < 1000 \\ 200 & 1000 \leq M_R^{cut} \end{cases}$
7. goto 4
8. $M_R^{edges} = M_R^{edges} \cup \{M_R^{cut}\}$
9. $R_{cut}^2 = R_{cut}^2 + 0.05$
10. goto 5
11. $M_R^{edges} = M_R^{edges} \cup \{M_R^{cut}\}$

This gives us a list of edges in the M_R plane, listed in inverse order from 3000 to 150. If we let $M_R^{edges} = \{M_{R,1}, M_{R,2}, \dots, M_{R,N} | M_{R,i} > M_{R,i+1}\}$, then divide

the plane into bins: $\{[M_{R,2}, M_{R,1}) \otimes [0, 1), [M_{R,3}, M_{R,2}) \otimes [0, 0.05), [M_{R,3}, M_{R,2}) \otimes [0, 0.10), [M_{R,4}, M_{R,3}) \otimes [0, 0.05), [M_{R,4}, M_{R,3}) \otimes [0.05, 0.10), [M_{R,4}, M_{R,3}) \otimes [0.10, 1.00) \dots\}$.

For example, if our algorithm returns $M_R^{edges} = \{3000, 1400, 400, 250, 150\}$, then we would divide the plane into the following bins:

- $[1400, 3000) \otimes [0, 1)$
- $[400, 1400) \otimes [0.00, 0.05), [400, 1400) \otimes [0.05, 1.00)$
- $[250, 400) \otimes [0.00, 0.05), [250, 400) \otimes [0.05, 0.10), [250, 400) \otimes [0, 10, 1.00)$
- $[150, 250) \otimes [0.00, 0.05), [150, 250) \otimes [0.05, 0.10), [150, 250) \otimes [0.10, 0.15), [150, 250) \otimes [0.15, 1.00)$

The sideband prediction binned according to the output of the algorithm is shown in Figure 5.16. One can see that the HighPt and LowRes boxes each have 15 bins, the HighRes box has 10 bins, and the Hbb and Zbb boxes each have 3 bins. One sees that none of the bins have 0 predicted sideband events, which will simplify the statistical analysis.

5.6.2 Signal Region Event Yields

The signal regions for each box are listed in tables 5.7- 5.11. We see in the tables the predicted background from resonant (“SM Higgs boson”) and non-resonant (“Sideband”) processes, as well as the expected yield from a signal of the two types of SMS under consideration at two different mass points for the neutralino. The errors on the background prediction and MC come from the statistics of the sideband data and MC sample respectively, which is discussed further in Section 5.7.1.

5.7 Systematics

Our systematic errors fall into three broad categories: those affecting the normalization of the signal and background MC, those affecting the objects within the MC, and those describing the uncertainty on the background prediction. For each systematic

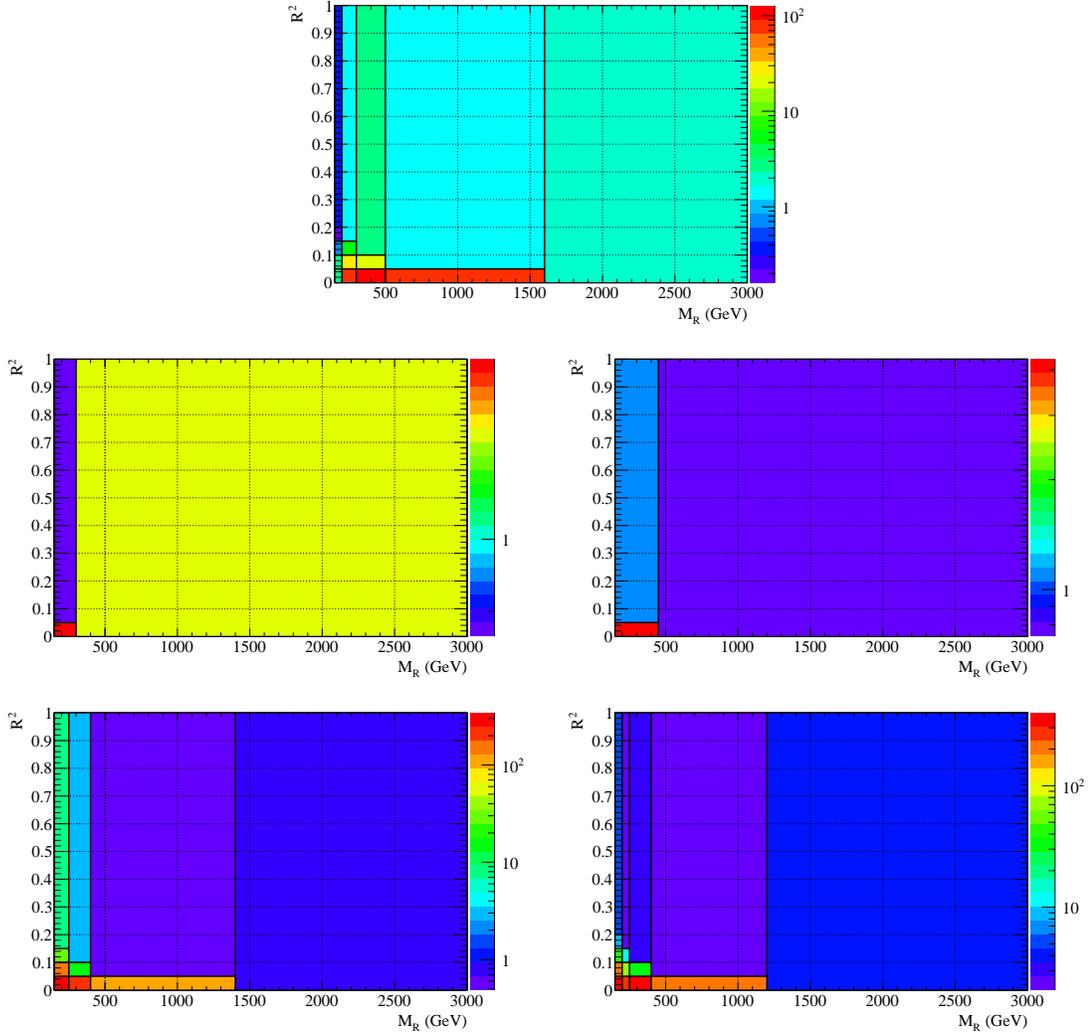


Figure 5.16: The prediction of the non-resonant background from the $m_{\gamma\gamma}$ sideband binned according to the regions selected by the signal region binning algorithm.

error, we evaluate its size and its degree of correlation within the R^2 - M_R plane and between the boxes.

The first category of systematics are those affecting only normalization of the Monte Carlo. These systematics are listed in Table 5.12. The luminosity systematic is a measurement of the uncertainty of the delivered luminosity, which gives a fully correlated error on the MC normalization. The trigger efficiency systematic is a measure of the uncertainty in the plateau of the trigger efficiency. We have measured that the trigger efficiency for our selection is $81 \pm 5\%$, where the uncertainty comes from the statistics of the sample from which we measure the turn-on curve. The MC

HighPt Box				$pp \rightarrow \chi_1^\pm \chi_2 \rightarrow W^\pm H \chi_0 \chi_0$		$pp \rightarrow \chi_2 \chi_2 \rightarrow HH \chi_0 \chi_0$	
M_R Range	R^2 Range	Sideband	SM Higgs boson	130 GeV	200 GeV	130 GeV	200 GeV
150 - 200	0.00 - 0.05	3.04 ± 0.697	0.12 ± 0.017	0.00 ± 0.003	0.00 ± 0.000	0.01 ± 0.006	0.00 ± 0.000
150 - 200	0.05 - 0.10	3.20 ± 0.715	0.11 ± 0.017	0.00 ± 0.004	0.00 ± 0.004	0.01 ± 0.006	0.01 ± 0.004
150 - 200	0.10 - 0.15	0.80 ± 0.358	0.02 ± 0.006	0.01 ± 0.009	0.00 ± 0.000	0.01 ± 0.005	0.01 ± 0.006
150 - 200	0.15 - 0.20	0.16 ± 0.160	0.01 ± 0.006	0.02 ± 0.014	0.01 ± 0.006	0.00 ± 0.002	0.01 ± 0.005
150 - 200	0.20 - 1.00	0.32 ± 0.226	0.05 ± 0.011	0.21 ± 0.041	0.28 ± 0.038	0.00 ± 0.002	0.11 ± 0.018
200 - 300	0.00 - 0.05	66.52 ± 3.262	2.94 ± 0.087	0.13 ± 0.032	0.03 ± 0.013	0.49 ± 0.043	0.09 ± 0.017
200 - 300	0.05 - 0.10	31.66 ± 2.251	1.08 ± 0.052	0.22 ± 0.042	0.09 ± 0.022	0.30 ± 0.034	0.19 ± 0.024
200 - 300	0.10 - 0.15	6.24 ± 0.999	0.14 ± 0.019	0.21 ± 0.041	0.11 ± 0.024	0.07 ± 0.017	0.21 ± 0.025
200 - 300	0.15 - 1.00	1.28 ± 0.452	0.14 ± 0.019	0.66 ± 0.073	1.04 ± 0.073	0.02 ± 0.009	0.66 ± 0.045
300 - 500	0.00 - 0.05	128.09 ± 4.527	7.76 ± 0.141	0.49 ± 0.063	0.28 ± 0.038	1.38 ± 0.072	0.46 ± 0.038
300 - 500	0.05 - 0.10	19.19 ± 1.752	0.88 ± 0.047	0.41 ± 0.058	0.47 ± 0.049	0.31 ± 0.034	0.64 ± 0.044
300 - 500	0.10 - 1.00	2.40 ± 0.619	0.23 ± 0.024	0.67 ± 0.073	1.60 ± 0.091	0.05 ± 0.013	1.25 ± 0.062
500 - 1600	0.00 - 0.05	82.67 ± 3.637	8.83 ± 0.150	0.53 ± 0.065	0.53 ± 0.052	1.13 ± 0.065	0.65 ± 0.045
500 - 1600	0.05 - 1.00	1.28 ± 0.452	0.25 ± 0.025	0.43 ± 0.059	1.04 ± 0.073	0.08 ± 0.017	0.93 ± 0.053
1600 - 3000	0.00 - 1.00	2.08 ± 0.577	0.39 ± 0.032	0.01 ± 0.010	0.01 ± 0.007	0.02 ± 0.008	0.02 ± 0.007

Table 5.7: HighPt Box. Event yields for SM Higgs boson are normalized to the SM cross section. Event yields for the SMS models are for NLO theory cross sections. The column label refers to m_{χ_2} ; in all columns $m_{\chi_1} = 1$ GeV. The errors come only from the statistics of the sideband or MC sample.

Hbb Box				$pp \rightarrow \chi_1^\pm \chi_2 \rightarrow W^\pm H \chi_0 \chi_0$		$pp \rightarrow \chi_2 \chi_2 \rightarrow HH \chi_0 \chi_0$	
M_R Range	R^2 Range	Sideband	SM Higgs boson	130 GeV	200 GeV	130 GeV	200 GeV
150 - 300	0.00 - 0.05	1.30 ± 0.451	0.01 ± 0.006	0.00 ± 0.002	0.00 ± 0.000	0.36 ± 0.037	0.07 ± 0.014
150 - 300	0.05 - 1.00	0.87 ± 0.368	0.02 ± 0.007	0.00 ± 0.005	0.00 ± 0.004	0.27 ± 0.032	0.50 ± 0.039
300 - 3000	0.00 - 1.00	1.16 ± 0.425	0.10 ± 0.016	0.00 ± 0.005	0.00 ± 0.004	0.17 ± 0.025	0.28 ± 0.029

Table 5.8: Hbb Box. Event yields for SM Higgs boson are normalized to the SM cross section. Event yields for the SMS models are for NLO theory cross sections. The column label refers to m_{χ_2} ; in all columns $m_{\chi_1} = 1$ GeV. The errors come only from the statistics of the sideband or MC sample.

is normalized to take into account the total trigger efficiency and so the uncertainty is added as a systematic error on this normalization.

The theory uncertainties on the Higgs boson are per-process uncertainties on the production cross section taken from [12]. These uncertainties come from uncalculated higher-order EW and QCD radiative corrections and from uncertainties on the PDF. While these uncertainties are quite important for the search for the SM Higgs boson, they are relatively small contributions to our uncertainty. The signal theory uncertainties affect the normalization of our signal models and are described in [27] and

Zbb Box				$pp \rightarrow \chi_1^\pm \chi_2 \rightarrow W^\pm H \chi_0 \chi_0$		$pp \rightarrow \chi_2 \chi_2 \rightarrow HH \chi_0 \chi_0$	
M_R Range	R^2 Range	Sideband	SM Higgs boson	130 GeV	200 GeV	130 GeV	200 GeV
150 - 450	0.00 - 0.05	4.28 ± 0.890	0.08 ± 0.014	0.00 ± 0.005	0.00 ± 0.004	0.14 ± 0.023	0.05 ± 0.012
150 - 450	0.05 - 1.00	1.12 ± 0.454	0.04 ± 0.009	0.00 ± 0.005	0.01 ± 0.006	0.11 ± 0.020	0.27 ± 0.029
450 - 3000	0.00 - 1.00	0.74 ± 0.371	0.04 ± 0.010	0.00 ± 0.000	0.00 ± 0.000	0.02 ± 0.010	0.02 ± 0.008

Table 5.9: Zbb Box. Event yields for SM Higgs boson are normalized to the SM cross section. Event yields for the SMS models are for NLO theory cross sections. The column label refers to m_{χ_2} ; in all columns $m_{\chi_1} = 1$ GeV. The errors come only from the statistics of the sideband or MC sample.

HighRes Box				$pp \rightarrow \chi_1^\pm \chi_2 \rightarrow W^\pm H \chi_0 \chi_0$		$pp \rightarrow \chi_2 \chi_2 \rightarrow HH \chi_0 \chi_0$	
M_R Range	R^2 Range	Sideband	SM Higgs boson	130 GeV	200 GeV	130 GeV	200 GeV
150 - 250	0.00 - 0.05	343.53 ± 7.529	13.56 ± 0.186	0.29 ± 0.048	0.09 ± 0.022	0.75 ± 0.053	0.11 ± 0.018
150 - 250	0.05 - 0.10	134.47 ± 4.710	4.94 ± 0.112	0.41 ± 0.057	0.16 ± 0.028	0.47 ± 0.042	0.16 ± 0.022
150 - 250	0.10 - 0.15	31.64 ± 2.285	0.90 ± 0.048	0.37 ± 0.054	0.25 ± 0.036	0.11 ± 0.021	0.22 ± 0.026
150 - 250	0.15 - 1.00	7.75 ± 1.131	0.27 ± 0.026	0.52 ± 0.065	1.24 ± 0.080	0.05 ± 0.013	0.83 ± 0.050
250 - 400	0.00 - 0.05	198.08 ± 5.717	9.21 ± 0.153	0.26 ± 0.046	0.14 ± 0.027	0.41 ± 0.039	0.13 ± 0.020
250 - 400	0.05 - 0.10	14.21 ± 1.531	0.47 ± 0.035	0.17 ± 0.037	0.16 ± 0.029	0.08 ± 0.017	0.16 ± 0.022
250 - 400	0.10 - 1.00	2.58 ± 0.653	0.09 ± 0.015	0.10 ± 0.029	0.75 ± 0.062	0.01 ± 0.007	0.41 ± 0.035
400 - 1400	0.00 - 0.05	95.41 ± 3.968	5.36 ± 0.117	0.19 ± 0.039	0.20 ± 0.032	0.24 ± 0.030	0.14 ± 0.021
400 - 1400	0.05 - 1.00	0.48 ± 0.283	0.06 ± 0.012	0.05 ± 0.019	0.21 ± 0.033	0.00 ± 0.004	0.15 ± 0.022
1400 - 3000	0.00 - 1.00	0.81 ± 0.365	0.07 ± 0.014	0.00 ± 0.004	0.00 ± 0.004	0.00 ± 0.003	0.00 ± 0.003

Table 5.10: HighRes Box. Event yields for SM Higgs boson are normalized to the SM cross section. Event yields for the SMS models are for NLO theory cross sections. The column label refers to m_{χ_2} ; in all columns $m_{\chi_1} = 1$ GeV. The errors come only from the statistics of the sideband or MC sample.

LowRes Box				$pp \rightarrow \chi_1^\pm \chi_2 \rightarrow W^\pm H \chi_0 \chi_0$		$pp \rightarrow \chi_2 \chi_2 \rightarrow HH \chi_0 \chi_0$	
M_R Range	R^2 Range	Sideband	SM Higgs boson	130 GeV	200 GeV	130 GeV	200 GeV
150 - 200	0.00 - 0.05	384.91 ± 10.583	3.01 ± 0.088	0.05 ± 0.020	0.03 ± 0.012	0.11 ± 0.020	0.02 ± 0.007
150 - 200	0.05 - 0.10	194.92 ± 7.531	1.33 ± 0.058	0.08 ± 0.026	0.03 ± 0.012	0.13 ± 0.022	0.02 ± 0.008
150 - 200	0.10 - 0.15	47.82 ± 3.730	0.26 ± 0.026	0.08 ± 0.026	0.06 ± 0.017	0.04 ± 0.012	0.04 ± 0.011
150 - 200	0.15 - 0.20	8.58 ± 1.580	0.05 ± 0.012	0.06 ± 0.023	0.04 ± 0.015	0.02 ± 0.009	0.01 ± 0.006
150 - 200	0.20 - 1.00	5.20 ± 1.230	0.05 ± 0.011	0.13 ± 0.033	0.27 ± 0.037	0.00 ± 0.004	0.19 ± 0.024
200 - 250	0.00 - 0.05	290.83 ± 9.199	2.45 ± 0.079	0.05 ± 0.021	0.01 ± 0.007	0.15 ± 0.024	0.01 ± 0.006
200 - 250	0.05 - 0.10	71.47 ± 4.561	0.54 ± 0.037	0.08 ± 0.026	0.04 ± 0.014	0.07 ± 0.017	0.05 ± 0.012
200 - 250	0.10 - 0.15	10.14 ± 1.717	0.07 ± 0.013	0.05 ± 0.021	0.04 ± 0.014	0.02 ± 0.008	0.05 ± 0.012
200 - 250	0.15 - 1.00	2.60 ± 0.870	0.01 ± 0.006	0.05 ± 0.020	0.15 ± 0.028	0.00 ± 0.003	0.20 ± 0.025
250 - 400	0.00 - 0.05	397.64 ± 10.757	3.81 ± 0.098	0.10 ± 0.028	0.07 ± 0.019	0.17 ± 0.025	0.06 ± 0.014
250 - 400	0.05 - 0.10	29.11 ± 2.910	0.25 ± 0.025	0.07 ± 0.023	0.06 ± 0.017	0.03 ± 0.010	0.05 ± 0.012
250 - 400	0.10 - 1.00	3.38 ± 0.992	0.03 ± 0.009	0.06 ± 0.022	0.29 ± 0.039	0.00 ± 0.004	0.20 ± 0.025
400 - 1200	0.00 - 0.05	183.23 ± 7.302	2.52 ± 0.080	0.09 ± 0.027	0.08 ± 0.020	0.10 ± 0.020	0.07 ± 0.015
400 - 1200	0.05 - 1.00	2.08 ± 0.778	0.02 ± 0.008	0.02 ± 0.012	0.12 ± 0.025	0.00 ± 0.004	0.05 ± 0.012
1200 - 3000	0.00 - 1.00	3.90 ± 1.065	0.08 ± 0.014	0.00 ± 0.005	0.00 ± 0.000	0.00 ± 0.000	0.00 ± 0.000

Table 5.11: LowRes Box. Event yields for SM Higgs boson are normalized to the SM cross section. Event yields for the SMS models are for NLO theory cross sections. The column label refers to m_{χ_2} ; in all columns $m_{\chi_1} = 1$ GeV. The errors come only from the statistics of the sideband or MC sample.

reflect similar sorts of normalization uncertainties. These uncertainties are relatively important for excluding certain signal hypotheses, but can also be avoided by setting cross section upper limits.

The next category of uncertainties are those that affect the Monte Carlo as uncertainties on the properties of objects, which uncertainties are summarized in Table 5.13. These uncertainties are treated by varying the quantity up and down by the size of

Source	value	target
luminosity	2.5%	Signal Models, SM Higgs boson MC
trigger efficiency	5%	Signal Models, SM Higgs boson MC
Higgs boson theory	2%-8%	SM Higgs boson MC
signal theory x-sec uncertainty	≈ 13%	

Table 5.12: Systematic uncertainties considered in the analysis affecting only the global normalization of MC samples.

Source	value	target
jet energy scale	shape(3%)	Signal Models, SM Higgs boson MC
photon energy and resolution	shape (1%)	Signal Models, SM Higgs boson MC
b-tagging ID	shape (0.1%)	Signal Models, SM Higgs boson MC
σ_E/E uncertainty	shape	Signal Models, SM Higgs boson MC

Table 5.13: Object-level systematic uncertainties considered in the analysis affecting the normalization and shape of MC samples.

uncertainty and re-running the entire analysis chain on the events with the variations. The new MC background predictions from these up/down variations are taken as the $\pm 1\sigma$ size of the resultant uncertainty. Apart from the b-tagging uncertainty (discussed below), these do not affect the weight of the events (as the normalization uncertainties do), but they can cause events to have different values of M_R and/or R^2 , migrate between the boxes, or be rejected by the selection.

For an example of this, consider the first uncertainty in the table, which is the uncertainty on the jet energy scale. To measure the effect of this, we run the analysis taking all jets in each event to have energy 1σ less than their mean measured energy. This could cause, for example, an event which has only a single jet of 30.1 GeV to be rejected in the -1σ version of the analysis. Similarly, since the energies of all the jets and E_T^M in the events are less, the values of M_R and R^2 will be different in the $\pm 1\sigma$ analyses, so the shapes in the R^2 - M_R plane will be different. We consider in this way the uncertainties on the jet energy scale, which effects the jets and E_T^M , the photon energy and scale, and separately the uncertainty on σ_E/E . This uncertainty is taken from the size of the discrepancy between σ_E/E in data and MC and causes migration of events between the HighRes and LowRes categories.

The b-tagging systematic is treated a bit differently, since it changes the normalization of the MC. There is a measured difference in the efficiency to tag b-quarks in data versus that in MC that is corrected for in the analysis. This correction is done by weighting events based on the number and p_T of the b-tagged versus non-btagged jets in the event, and the uncertainty on this correction is taken as a systematic. The size of the uncertainty varies as a function of the number and p_T of the jets in the event, and so is computed event by event and varies the reweighting up and down.

Source	value	target
background prediction uncertainty	1% – 50%	background shape
sideband statistics	1 – 100%	statistics in the data sidebands
fit choice	$\approx 1\%$	background normalization
MC statistics	varies	statistics in SM Higgs boson and SMS MC

Table 5.14: Systematic uncertainties arising from limited statistics considered in the analysis affecting the normalization and shape of the MC and sideband prediction.

The treatment of these uncertainties somewhat overestimates their effect on the analysis. Since we vary all objects up and down at the same time, we are getting the envelope of worst possible cases where the uncertainties are fully correlated across all objects. This is done in this way since the correlations are not fully known—and not easy to measure—and because these systematics are also small compared to the next set of systematics which represent bin-to-bin fluctuations.

The final set of systematics are those that come from statistical uncertainties and are summarized in Table 5.14. These are by far the biggest systematics and the only ones that affect the sideband prediction taken from data as well as the MC. They are discussed individually below.

5.7.1 Background Statistics Systematic

As described earlier, our background prediction is composed of two pieces (the sideband prediction and SM Higgs boson MC), each of which has finite statistics. The amount of statistics create an uncertainty on the background prediction in each signal bin equal to the statistical uncertainty on the number of events in the sideband of that bin times the scale factor for the box. For example, the first bin of the HighRes box ($150 < M_R < 250$ $0.00 < R^2 < 0.05$) has 2082 events in the sideband and the box has a scale factor of 0.165, so the predicted non-resonant background in that region is 343.5 events. Using the Gaussian approximation, the error on 2082 is $\sqrt{2082} = 45.6$ and so we take the systematic uncertainty on the non-resonant background prediction to be $45.6 \times 0.165 = 7.5$. We perform an exactly analogous procedure for the MC, where the statistics of the MC sample in each signal region replace the statistics of the sideband region.

We find that this procedure is effective even in regions with sufficiently small statistics that the Gaussian approximation of the error on the number of sideband events does not hold. We show this with a toy study: imagine a region in the HighRes box (scale factor=0.165) with 2 events in the $m_{\gamma\gamma}$ sideband and 0 contribution from the SM Higgs boson. This gives us a predicted background of 0.33 background events. If we observe N events in this region, the correct procedure to evaluate the significance of the observation is to throw toys, with a toy consisting of:

- sample $N_{sideband}^{toy} Pois(2)$
- sample $N_{bkg}^{toy} Pois(N_{sideband}^{toy} \times 0.165)$

One then sets a p-value based on these toys. We use the statistical tools developed for the CMS SM Higgs boson search [80], the implementation of which prefers Gaussian systematics, so we approximate our true Poisson uncertainties by:

- sample $\mu_{toy} Gaus(2 * 0.165, \sqrt{2} * 0.165)$
- sample $N_{bkg}^{toy,approx} Pois(\mu_{toy})$

Figure 5.17 shows the distribution of the number of background events found in 10,000 toy events for the full procedure and the Gaussian approximation. One sees excellent agreement between the two across the full range of possible N_{bkg}^{toy} . We therefore use this Gaussian approximation for as needed systematics.

5.7.2 Background Prediction Uncertainty

Figures 5.14 and 5.15 show that the sideband prediction procedure closes to within the statistical error, but cannot exclude the possibility that there are systematic shifts significantly smaller than the statistical error. Specifically we consider the possibility that the $R^2 - M_R$ distribution shifts significantly as a function of $m_{\gamma\gamma}$, and so the upper and lower sidebands have slightly different shapes. To check this, we compute the non-resonant background prediction using only the lower sideband and compute the prediction using only the upper sideband and compare them. Figure 5.18 shows

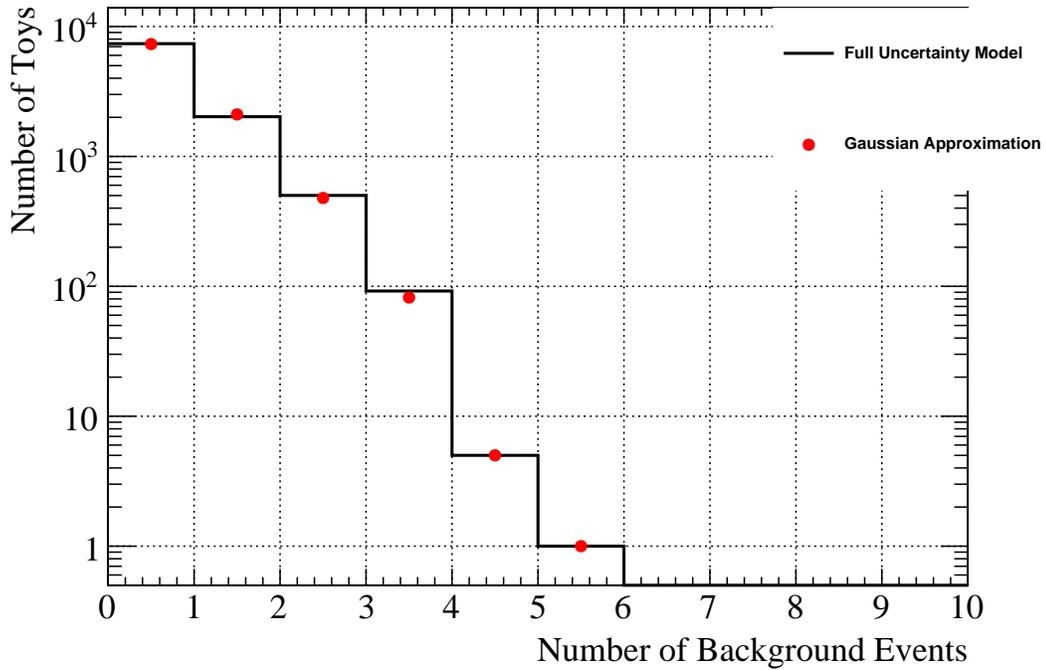


Figure 5.17: Distribution of the number of background events in 10,000 toy events for the full uncertainty calculation (black) and the Gaussian approximation (red), showing excellent agreement.

the comparison in each signal bin. All the deviations we see are within the statistical uncertainty.

To cover any residual difference between the upper and lower sideband predictions, we assign a systematic to the background yield in each bin to capture the difference between the upper and lower predictions. We do this in the following way: in each box, we derive two additional scale factors s_{upper} and s_{lower} to predict the signal region background from the upper and lower sidebands, respectively. Then, for each signal bin, we let N_{upper} be the number of events in the upper sideband, N_{lower} be the number of events in the lower sideband, and s be the nominal scale factor derived for the box the bin is in. The nominal background prediction is just $s \times (N_{lower} + N_{upper})$. We then have two additional predictions for the background: $s_{lower} \times N_{lower}$ and $s_{upper} \times N_{upper}$.

In the case where $|s_{lower}N_{lower} - s_{upper}N_{upper}| > s\sqrt{N_{upper} + N_{lower}}$, we apply $\max(s_{lower}N_{lower}, s_{upper}N_{upper}) - s \times (N_{lower} + N_{upper})$ as a $+1\sigma$ systematic on the background prediction and $s \times (N_{lower} + N_{upper}) - \min(s_{lower}N_{lower}, s_{upper}N_{upper})$ as a

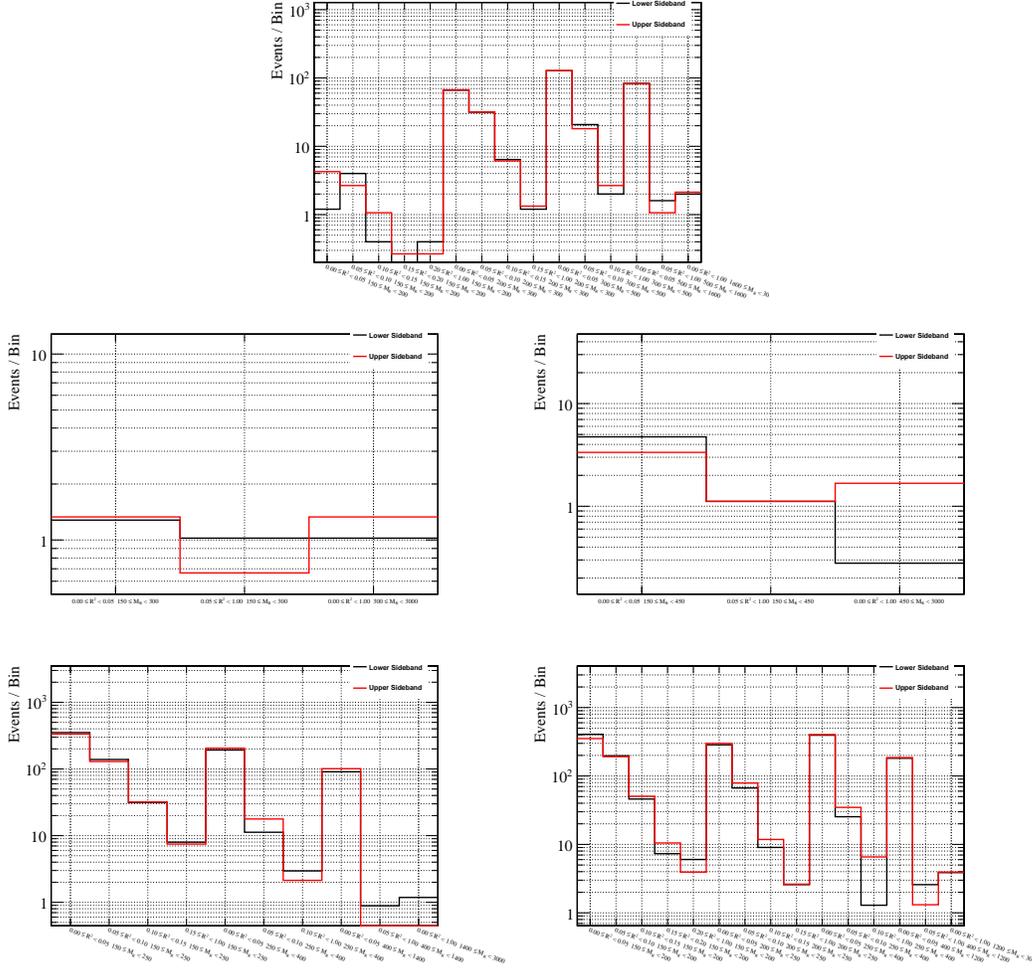


Figure 5.18: Comparison of the predictions derived from the lower sideband only (black) and upper sideband only (red). The comparison is broken down by box in the HighPt box (top), Hbb box (middle left), Zbb box (middle right), HighRes box (bottom left), and LowRes box (bottom right).

-1σ systematic. In the case where $|s_{lower}N_{lower} - s_{upper}N_{upper}| \leq s\sqrt{N_{upper} + N_{lower}}$, we apply $s\sqrt{N_{upper} + N_{lower}}$ as the $\pm 1\sigma$ systematic. This reflects the fact that an observed difference smaller than the statistical uncertainty on the background cannot be trusted as an accurate estimate of the systematic. Note that in this case this systematic has the exact same magnitude as the sideband statistics systematic. This systematic is treated as uncorrelated in all signal bins to be conservative in the treatment.

5.7.3 Fit Choice Systematic

This systematic deals with the possibility that the double exponential was not the best choice to fit our non-resonant background. While this function choice is well motivated physically, it is not the only well motivated function choice. To measure any potential impact that this could have on the background prediction, we repeat the background fit in the 5 boxes using different choices of fit function. Table 5.15 summarizes the results (the exact definition of the functions is in Table 5.16). The table contains the integral of the fit function in the signal region as extracted from the fit function. We introduce a systematic on the scale factor equal to the percent difference between the integral of the Double Exponential and the average of the integrals of the functions.

One sees that the maximum deviation is 1.7% in the Hbb category, while all other deviations are $\leq 1\%$. For low statistics categories, the size of this error is negligible compared to the systematics from statistics, but in the high statistics categories it can become a larger effect.

Fit Function	HighPt	Hbb	Zbb	HighRes	LowRes
Double Exp	0.127	0.161	0.163	0.131	0.206
Modified Exp	0.128	0.166	0.166	0.134	0.208
Single Exp	0.127	0.169	0.163	0.128	0.206
Single Power Law	0.127	0.166	0.157	0.128	0.200
Double Power Law	0.124	0.158	0.158	0.132	0.202
Average	0.127	0.164	0.161	0.131	0.204
Difference	0.4%	1.7%	1.0%	0.6%	0.6%

Table 5.15: The integral in the signal region of the fit of the specified function to the data in the corresponding box (the integral is normalized to 1 over the whole range $103 < m_{\gamma\gamma} < 160$). The second to last line is the average of the integrals and the final line is the percentage difference between this average and the value for the double exponential (which is the nominal fit choice for the analysis).

fit type	formula
modified exponential	$e^{\alpha m^\lambda}$
single exponential	$e^{\alpha m}$
double exponential	$f \times e^{\alpha_1 m} + (1 - f) \times e^{\alpha_2 m}$
single power law	m^λ
double power law	$f \times m^{\lambda_1} + (1 - f) \times m^{\lambda_2}$

Table 5.16: definition of alternate fit functions

5.8 Inverted Analysis Procedure

The key technique used in the main analysis is the prediction of the non-resonant background using the distribution of events in the $m_{\gamma\gamma}$ sideband in the R^2 - M_R plane. We have shown that this technique is very robust in both data control regions and in Monte Carlo, but there are still systematics associated with the statistical uncertainty on the closure tests. Since these are the largest systematics we have in this analysis, we also perform a parallel analysis that predicts the background in a different way to have a relatively independent cross-check of the analysis; we call this the *inverse* or *inverted* analysis.

In the main analysis we fit the $m_{\gamma\gamma}$ distribution and then look at the shape in the R^2 - M_R plane; in this analysis we invert this ordering by cutting on R^2 and M_R to remove SM Higgs boson background and then fit the $m_{\gamma\gamma}$ distribution. The goal is to find a region of the R^2 - M_R plane with virtually no SM Higgs boson background but large acceptance for BSM Higgs boson production and look for a resonance at the Higgs boson mass. This is analogous to how searches for less common production modes of the Higgs boson are performed, by going to kinematic regions with little gluon fusion production and looking for a resonance. This method also has the benefit of a substantially reduced look-elsewhere-effect due to the much smaller number of signal region.

We define the cuts on M_R and R^2 on a box-by-box basis based on the following considerations:

- **small contribution from the SM Higgs boson:** The idea is to fit $m_{\gamma\gamma}$ in a $R^2 - M_R$ region where we expect virtually no SM Higgs boson, so if we see a

peak at the Higgs boson mass it suggests non-standard Higgs boson production;

- **small total background production:** We expect any possible SUSY signals to be relatively small, so a small background is desirable to enhance significance;
- **mass-splitting independence:** SUSY models with different mass splittings occupy different regions of the $R^2 - M_R$ planes. In the main analysis we capture this by binning in the plane, but we do not have that luxury here. We therefore maximize significance over a wide range of mass splittings.

With these considerations in mind, we define the criteria for selecting the cuts in R^2 and M_R for each box. We require that there be less than one remaining SM Higgs boson event after the cut, to satisfy the first criterion. From the possible cuts that satisfy this, we choose the one that maximizes the expected S/\sqrt{B} for all the signal points of our SMS with $m_{LSP} = 1\text{GeV}$. Table 5.17 shows the result of using this optimization procedure in the five analysis boxes. For all but the HighPt box, we end up with cuts that are fairly loose and inclusive in M_R and R^2 . For the HighPt box, we would be forced to cut out a large portion of the plane by using a single cut, so we use two disjoint regions to cover the plane.

Box	minimum M_R	minimum R^2
Highp _T region 1	450	0.05
Highp _T region 2 ¹	150	0.10
Hbb	150	0.00
Zbb	150	0.00
HighRes	250	0.05
LowRes	250	0.05

Table 5.17: Threshold on M_R and R^2 defining the search region in the inverse-logic analysis. Unlike the case of the default analysis, the razor variables are used to reduce the peaking background to a negligible level while the $m_{\gamma\gamma}$ distribution is used to characterize a possible signal induced by a new mechanism of Higgs boson production. (¹: This region will end at $M_R = 450\text{ GeV}$ to avoid overlap with region 1)

5.8.1 Inverse Analysis Fits

We do a S+B fit for the inverted analysis to extract the significance of any peak at or near the Higgs boson mass. In order to cover a wide variety of possible background shapes, we consider a number of potential background fit functions. The list of functions considered is shown in Table 5.18. We perform the same procedure of fitting in the sidebands of $m_{\gamma\gamma}$ as in the main analysis to define the shape of the background function. In order to correctly weight all of these potential background functions based on how well they fit the sideband data, we use the Akaike Information Criterion to form a composite background model [87–89].

fit type	formula
modified exponential	$e^{\alpha m^\lambda}$
single exponential	$e^{\alpha m}$
double exponential	$f \times e^{\alpha_1 m} + (1 - f) \times e^{\alpha_2 m}$
single power law	m^λ
double power law	$f \times m^{\lambda_1} + (1 - f) \times m^{\lambda_2}$
2 nd order bernstein polynomial	$\sum_{i=1}^{N=2} c_i b_{i,N=2}(m)$
3 rd order bernstein polynomial	$\sum_{i=1}^{N=3} c_i b_{i,N=3}(m)$

Table 5.18: definition of background fit functions considered for the inverted analysis

To do this, we fit every function to the data and compute the *AIC* value defined as:

$$AIC = -2\log(\mathcal{L}) + 2k + \frac{2k(k+1)}{N-k-1} \quad (5.7)$$

where k is the number of parameters of the function and N is the number of fitted data points. The *AIC* can be thought of as a goodness of fit corrected for the number of degrees of freedom that the function has and encodes information about whether adding more parameters to a fit is improving its quality sufficiently.

Given this set $\{f_i\}$ of fit functions, each with AIC value $\{a_i\}$, we can compute the minimum $a_{min} = \min_i a_i$, which can be considered the best fit function found. One could simply use this function to fit the background, but in the case that there are many functions with similar AIC values, this may not be the optimal strategy.

Instead we can assign each function an AIC weight

$$w_i = \frac{e^{-\frac{1}{2}(a_i - a_{min})}}{\sum_j e^{-\frac{1}{2}(a_j - a_{min})}}$$

We can then form a composite background model $f_{AIC} = \sum_i w_i f_i$, which contains information about how well all the shapes considered describe the data [89].

Table 5.19 shows the ΔAIC values for the various fit choices in the six regions of the inverse analysis. One can see that typically a single exponential is the best choice, indicated by $\Delta a_i = 0$. In all of these cases, the double exponential and triple exponential reduce to the same curve as the single exponential and fit no better. In two cases the single power law is the best fit (For the Zbb this is tied with the single exponential and in the case of HighPt1 it only beats it by a little bit). Finally, the second order polynomial is the best fit for the HighPt2 case.

function	Parameters	$a_i - a_{min}$					
		HighPt1	HighPt2	Hbb	Zbb	HighRes	LowRes
double exponential	3	4.88	4.92	7.52	15.00	4.18	4.15
single exponential	1	0.11	2.04	0.00	0.00	0.00	0.00
triple exponential	5	10.52	9.17	15.50	47.00	8.33	8.27
modified exponential	2	2.66	4.06	2.82	5.00	1.68	1.89
single power	1	0.00	1.93	0.35	0.00	0.44	0.30
double power	3	5.39	5.13	7.85	15.00	4.62	4.45
second order poly	3	4.92	0.00	7.18	11.79	3.82	3.77
third order poly	4	8.34	2.21	12.76	40.56	2.86	5.84

Table 5.19: ΔAIC values of the various fit choices in the different boxes.

We use the weights in Table 5.19 to build a composite background model

$$f_{AIC}(m_{\gamma\gamma}) = \sum_i w_i f_i(m_{\gamma\gamma})$$

Figure 5.19 shows the composite AIC background model for the 6 analysis region of the inverted analysis. The weights are computed from the ΔAIC values from Table 5.19 and the parameters of the individual functions are taken from each of their fits to the sideband data.

This composite model will form the background only (B) hypothesis in each analysis region. To test for the presence of signal, we also create a signal plus background (S+B) model by adding a single Gaussian as the signal model. Since we are looking for the SM Higgs boson, we float the mean of the signal model with a Gaussian penalty centered at the $m_{\gamma\gamma} = 125.7$ GeV with a width of 0.4 GeV (the PDG mass an uncertainty of the SM Higgs boson). We fix the width of the signal model in each analysis box to be equal to the σ_{eff} of the SM Higgs boson from Table 5.4. We then float the number of signal events $0 \leq N_{sig} < \infty$.

The systematic errors for this version of the analysis are much easier to treat. There is no Monte Carlo, so there are not efficiency-based systematics to consider. The key systematic comes from the uncertainty of the fit and the choice of the background function. The uncertainties of the fit can be taken into account by computing the profile likelihood of the number of signal events to measure the significance of any peak. We do not enforce any relationship between the number of signal events in each box, in order to be model independent.

The uncertainty coming from the choice of the background function can be estimated by measuring the uncertainties on the AIC weights from Table 5.19. We do this using *bootstrapping*, which is sampling the dataset with replacement. If we let D be the dataset with N_D entries, then we form a bootstrap dataset B_i by randomly choosing an element $x \in D$ N_D times where x is *not* removed from D after being selected, so the same element may end up in B multiple times.

We compute 10^5 of these bootstrap datasets and perform the AIC procedure on each one of them independently. This gives a 10^5 measurements of each of the individual w_i for each model. Since the values of w_i are highly correlated, we measure the covariance matrix of the weights from these bootstrap datasets as a measure of the allowed variance of the weights. When measuring the statistical significance of any potential excess, we do the AIC background fit with the AIC weights floated with a fully-correlated Gaussian penalty term given by the covariance matrix. This gives us a term in the profile likelihood that allows the relative AIC weights to be adjusted to account for uncertainty in the background model selection.

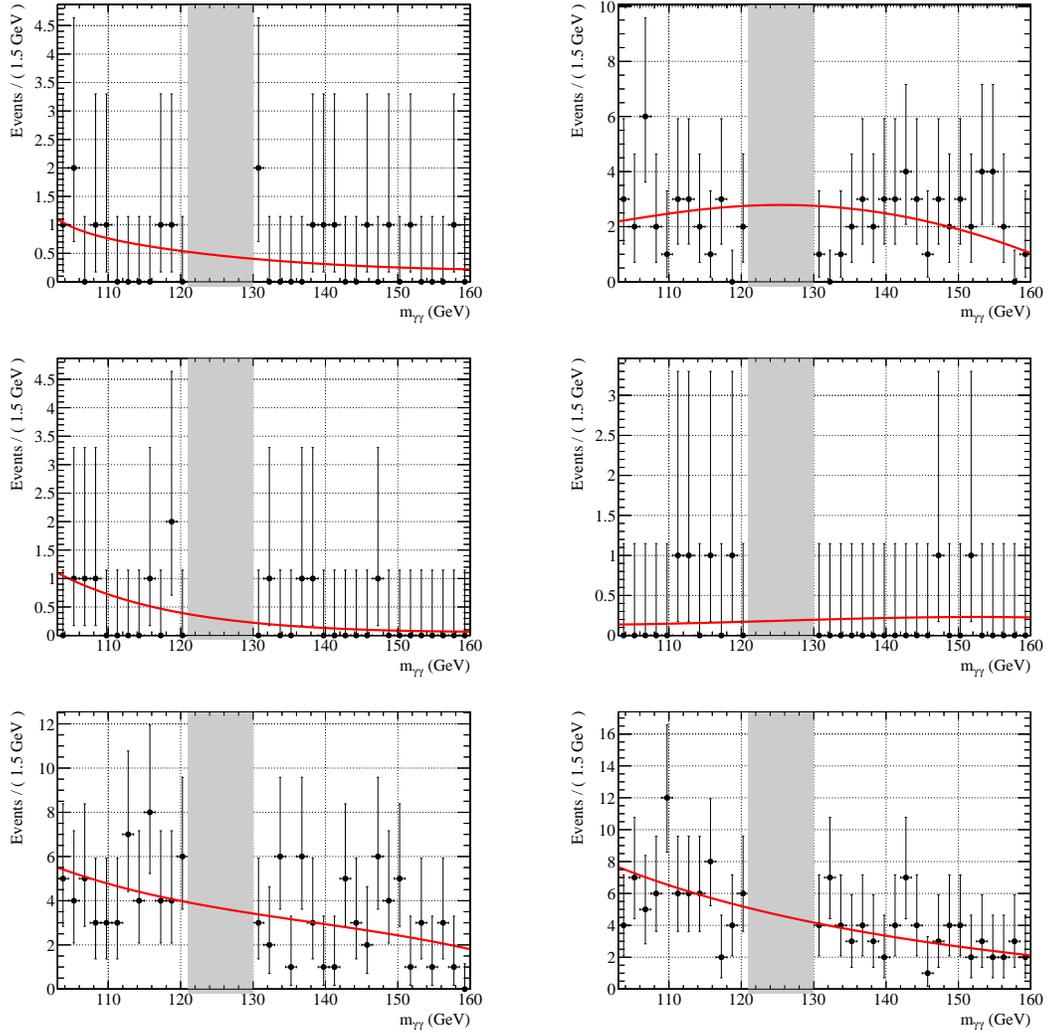


Figure 5.19: The AIC background fits for the HighPt1 (top left), HighPt2 (top right), Hbb (middle left), Zbb (middle right), HighRes (bottom left), and LowRes (bottom right) inverted analysis regions. The grey area with $121 < m_{\gamma\gamma} < 130$ GeV is the blinding region.

Chapter 6

Results and Conclusions

Having established the analysis selection path and background estimation in Chapter 5 we proceed to examine the signal region in the data, i.e. unblind the analysis and obtain the results. In Section 6.1 we examine the data in the signal $m_{\gamma\gamma}$ region and compare to the background prediction formed from the sum of continuum and SM Higgs background. We observe an excess of events in one region of the R^2 - M_R plane, so we proceed to attempt to characterize this excess to the extent the limited number of events allows us. At this stage we will also establish limits on the SMS models we used to establish the analysis design. In Section 5.8 we use a difference method of setting the background prediction that is not subject to some of the systematics of the main analysis; we call this analysis the *inverted analysis*. In Section 6.3 we look at some of the kinematic and other properties of the events to begin to understand the excess.

6.1 Unblinded Results

The analysis is designed to be sensitive to beyond the standard model production of a Higgs boson and at least one other object over a wide range of masses of the BSM particles. It does this by prioritizing configurations most typical of the expected signature of BSM physics, with a boosted Higgs or di-Higgs production by breaking the events down into five boxes as described in Section 5.3.4. The R^2 - M_R plane is then binned to create regions with small expected background from SM signatures

but potentially large signal from BSM processes.

Figure 6.1 shows the unblinded distributions of $m_{\gamma\gamma}$ in the five boxes. These are the fits from which the normalization of the continuum background is derived as described in Section 5.5. We see no major deviation of the data from the fits in the $m_{\gamma\gamma}$ signal region in any other 5 boxes, which is expected since these distributions are integrated over the entire R^2 - M_R plane, so background dominates over both SM and potential BSM production of the Higgs.

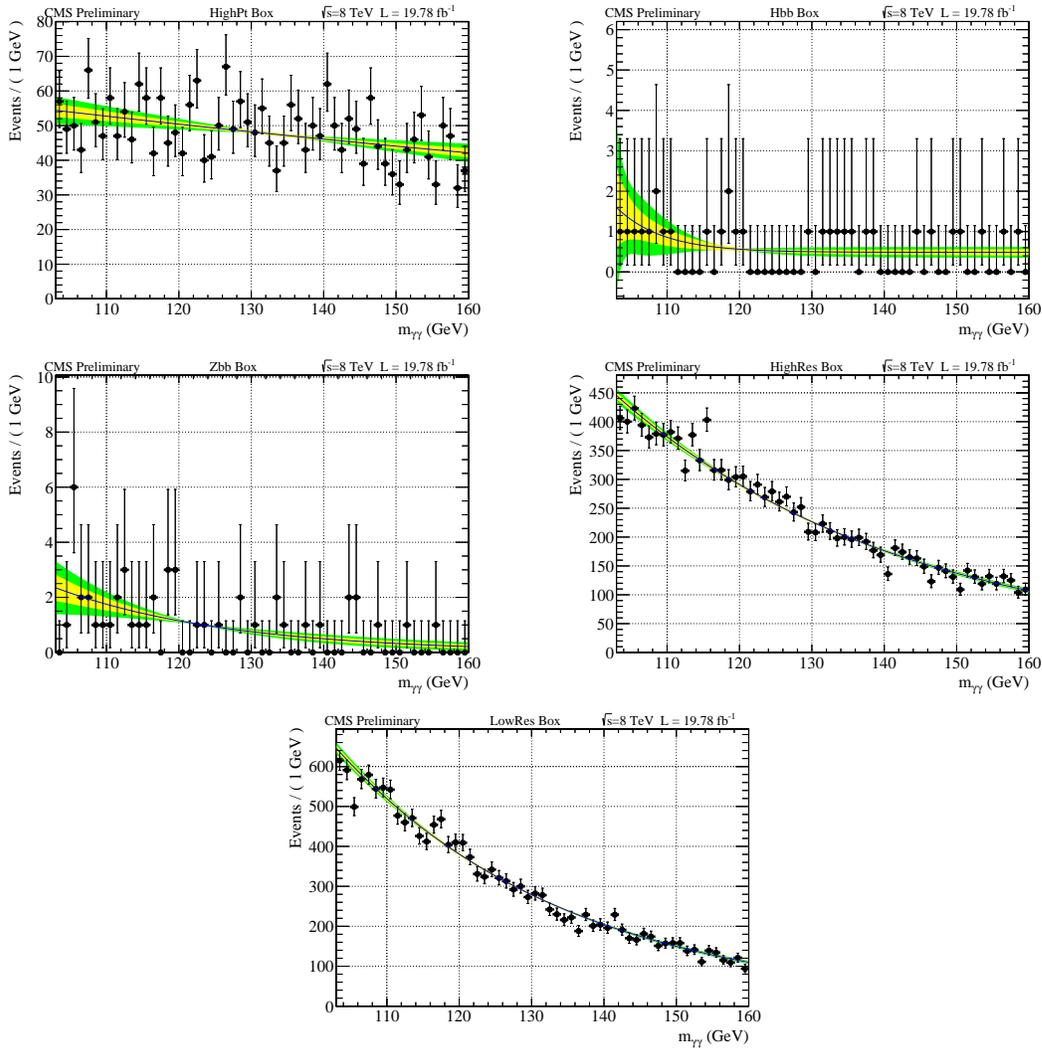


Figure 6.1: The unblinded fits to the $m_{\gamma\gamma}$ distribution in the HighPt (top left), Hbb (top right), Zbb (middle left), HighRes (middle right), and LowRes (bottom) categories. Fits are shown with their $\pm 1\sigma$ (yellow) and $\pm 2\sigma$ (green) error bands. The corresponding blinded distributions are in Figure 5.11

Tables 6.1-6.5 show the observed event yield in each of the signal regions of the analysis. The observed yield is compared against the expected yield from the sum of sideband prediction and SM Higgs MC, where the errors are the sum in quadrature of the systematic errors on the yield in each bin. Finally, the double-sided p-value of the observation compared to the expectation and the associated significance are quoted. The p-value is computed using pseudo-experiments (toys) by sampling the predicted background for each toy from the sum of independent Gaussians for each systematic and then throwing a Poisson around the toy mean and recording whether the observed deviation is greater than or equal to the observed deviation in data. One can see that the agreement between observation and expectation is excellent in the HighPt, Hbb, Zbb, and LowRes boxes and in all but one bin in the HighRes box. In the bin $400 < M_R < 1400$ $0.05 < R^2 < 1.00$ in the HighRes box, we see 5 events in the $m_{\gamma\gamma}$ signal region, but predict only $0.54^{+0.278}_{-0.278}$, which corresponds to a p-value of 0.001 (3.4σ).

M_R region	R^2 region	observed events	expected background	p-value	significance (σ)
150 - 200	0.00 - 0.05	4	$3.16^{+0.995}_{-0.995}$ (syst.)	0.782000	0.28
150 - 200	0.05 - 0.10	2	$3.31^{+1.022}_{-1.022}$ (syst.)	0.656000	0.45
150 - 200	0.10 - 0.15	1	$0.82^{+0.507}_{-0.507}$ (syst.)	1.000000	-0.00
150 - 200	0.15 - 0.20	0	$0.17^{+0.226}_{-0.226}$ (syst.)	1.000000	-0.00
150 - 200	0.20 - 1.00	0	$0.37^{+0.453}_{-0.452}$ (syst.)	1.000000	-0.00
200 - 300	0.00 - 0.05	87	$69.47^{+7.112}_{-7.112}$ (syst.)	0.115000	1.58
200 - 300	0.05 - 0.10	42	$32.74^{+4.726}_{-4.726}$ (syst.)	0.211000	1.25
200 - 300	0.10 - 0.15	8	$6.37^{+2.017}_{-2.017}$ (syst.)	0.655000	0.45
200 - 300	0.15 - 1.00	1	$1.42^{+1.391}_{-1.391}$ (syst.)	1.000000	-0.00
300 - 500	0.00 - 0.05	117	$135.85^{+13.964}_{-13.964}$ (syst.)	0.306000	1.02
300 - 500	0.05 - 0.10	21	$20.06^{+4.624}_{-4.624}$ (syst.)	0.958000	0.05
300 - 500	0.10 - 1.00	3	$2.63^{+2.172}_{-2.172}$ (syst.)	1.000000	-0.00
500 - 1600	0.00 - 0.05	82	$91.50^{+13.657}_{-13.657}$ (syst.)	0.575000	0.56
500 - 1600	0.05 - 1.00	0	$1.53^{+1.072}_{-1.072}$ (syst.)	0.436000	0.78
1600 - 3000	0.00 - 1.00	2	$2.47^{+1.734}_{-1.734}$ (syst.)	1.000000	-0.00

Table 6.1: Number of Events observed in the signal region compared to expected background in the **HighPt** box.

We note by comparing tables 5.10 and 5.11 that one expects roughly 43% as much SM Higgs in the LowRes box between $400 < M_R < 1200$ $0.05 < R^2 < 1.00$ compared to the HighRes box with $400 < M_R < 1400$ $0.05 < R^2 < 1.00$ (0.03 events versus 0.07

M_R region	R^2 region	observed events	expected background	p-value	significance (σ)
150 - 300	0.00 - 0.05	0	$1.32^{+1.073}_{-1.001}$ (syst.)	0.591000	0.54
150 - 300	0.05 - 1.00	0	$0.89^{+0.909}_{-0.900}$ (syst.)	0.748000	0.32
300 - 3000	0.00 - 1.00	1	$1.26^{+1.545}_{-1.507}$ (syst.)	1.000000	-0.00

Table 6.2: Number of Events observed in the signal region compared to expected background in the **Hbb** box.

M_R region	R^2 region	observed events	expected background	p-value	significance (σ)
150 - 450	0.00 - 0.05	1	$1.89^{+1.073}_{-1.073}$ (syst.)	0.794000	0.26
150 - 450	0.05 - 1.00	1	$0.76^{+1.267}_{-0.942}$ (syst.)	1.000000	-0.00
450 - 3000	0.00 - 1.00	1	$3.63^{+3.523}_{-2.988}$ (syst.)	0.512000	0.66

Table 6.3: Number of Events observed in the signal region compared to expected background in the **Zbb** box.

events). Coupled with the observation that the non-resonant background is about 4 times larger in the LowRes box, we would expect a small signal producing a Higgs to show up in the HighRes box but not the LowRes box.

M_R region	R^2 region	observed events	expected background	p-value	significance (σ)
150 - 250	0.00 - 0.05	357	$357.09^{+29.548}_{-29.552}$ (syst.)	1.000000	-0.00
150 - 250	0.05 - 0.10	147	$139.41^{+12.551}_{-12.550}$ (syst.)	0.650000	0.45
150 - 250	0.10 - 0.15	35	$32.54^{+3.880}_{-3.880}$ (syst.)	0.767000	0.30
150 - 250	0.15 - 1.00	7	$8.02^{+2.370}_{-2.302}$ (syst.)	0.788000	0.27
250 - 400	0.00 - 0.05	213	$207.29^{+18.566}_{-18.565}$ (syst.)	0.804000	0.25
250 - 400	0.05 - 0.10	20	$14.68^{+3.762}_{-2.692}$ (syst.)	0.317000	1.00
250 - 400	0.10 - 1.00	3	$2.67^{+1.692}_{-1.477}$ (syst.)	1.000000	-0.00
400 - 1400	0.00 - 0.05	109	$100.77^{+13.843}_{-12.783}$ (syst.)	0.631000	0.48
400 - 1400	0.05 - 1.00	5	$0.54^{+0.278}_{-0.278}$ (syst.)	0.001030	3.35
1400 - 3000	0.00 - 1.00	0	$0.88^{+0.832}_{-0.797}$ (syst.)	0.744000	0.33

Table 6.4: Number of Events observed in the signal region compared to expected background in the **HighRes** box.

M_R region	R^2 region	observed events	expected background	p-value	significance (σ)
150 - 200	0.00 - 0.05	407	$387.92^{+18.491}_{-18.482}$ (syst.)	0.499000	0.68
150 - 200	0.05 - 0.10	201	$196.25^{+10.538}_{-10.540}$ (syst.)	0.790000	0.27
150 - 200	0.10 - 0.15	54	$48.08^{+3.932}_{-3.932}$ (syst.)	0.487000	0.70
150 - 200	0.15 - 0.20	13	$8.63^{+1.486}_{-1.486}$ (syst.)	0.205000	1.27
150 - 200	0.20 - 1.00	5	$5.25^{+2.527}_{-2.468}$ (syst.)	1.000000	-0.00
200 - 250	0.00 - 0.05	297	$293.27^{+14.574}_{-14.576}$ (syst.)	0.876000	0.16
200 - 250	0.05 - 0.10	78	$72.01^{+5.093}_{-5.093}$ (syst.)	0.619000	0.50
200 - 250	0.10 - 0.15	7	$10.20^{+1.624}_{-1.624}$ (syst.)	0.413000	0.82
200 - 250	0.15 - 1.00	1	$2.61^{+1.537}_{-1.478}$ (syst.)	0.537000	0.62
250 - 400	0.00 - 0.05	386	$401.45^{+23.360}_{-23.360}$ (syst.)	0.616000	0.50
250 - 400	0.05 - 0.10	21	$29.36^{+4.466}_{-4.466}$ (syst.)	0.243000	1.17
250 - 400	0.10 - 1.00	4	$3.41^{+2.464}_{-1.884}$ (syst.)	0.881000	0.15
400 - 1200	0.00 - 0.05	159	$185.75^{+18.646}_{-18.646}$ (syst.)	0.218000	1.23
400 - 1200	0.05 - 1.00	4	$2.10^{+1.754}_{-1.695}$ (syst.)	0.524000	0.64
1200 - 3000	0.00 - 1.00	3	$3.98^{+1.900}_{-1.841}$ (syst.)	0.860000	0.18

Table 6.5: Number of Events observed in the signal region compared to expected background in the **LowRes** box.

6.1.1 Observed Limits in the Simplified Model Scan

Using the yields and uncertainties shown in tables 6.1-6.5, we can set limits on the production cross-sections of the simplified models targeted by the analysis. Since the distributions in the R^2 - M_R plane for the models look different depending on the mass of the SUSY particles, we set these limits as a function of the mass of the produced neutralino (we set the mass of the chargino in the asymmetric case to be equal to the mass of the neutralino). Based on the level of sensitivity we achieve, we focus on the case where the mass of the LSP is 1 GeV the limits for which can be seen in Figure 6.2.

For these limits we assume that the branching ratios are $Br(\chi_2 \rightarrow H\chi_0) = 100\%$ and $Br(\chi_1^\pm \rightarrow W^\pm\chi_0) = 100\%$. We can see that in the case of $\chi_2\chi_1^\pm$ production (left plot) we exclude production cross sections between 3 and 1.5 pb in the range $130 \leq m_{\chi_2} = m_{\chi_1^\pm} \leq 200$ GeV. For the case of symmetric $\chi_2\chi_2$ production in the HH final state, the exclusion limit is slightly worse, excluding cross-sections between 4 and 1.5 pb in the range $130 \leq m_{\chi_2} \leq 200$ GeV. It is also worth noting that the observed limit agrees very well (left plot) or produces a stronger limit (right plot) than expected in each case. This reflects the fact that there is excellent agreement between observed data and predicted background in most regions and a slight deficit of events in the Hbb box (Table 6.2), which is the most important box for the $\chi_2\chi_2$ case.

Figure 6.3 shows the observed and expected limit divided by the theoretical cross section at each mass. From this, we can easily read what masses are excluded in this model. We can exclude models with $\chi_2\chi_1^\pm \rightarrow HW^\pm\chi_0\chi_0$ and $m_{\chi_2} = m_{\chi_1^\pm} < 156$ GeV and models with $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0$ with $m_{\chi_2} < 136$ GeV. For both models we stay within twice the theory cross-section for the entire mass range $130 \leq m_{\chi_2} = m_{\chi_1^\pm} \leq 200$ GeV.

The large deviation observed in the HighRes box is not reflected in the limits for two important reasons. The first is that the M_R value of the excess is higher than would be typically produced by these models in the mass range of interest. From

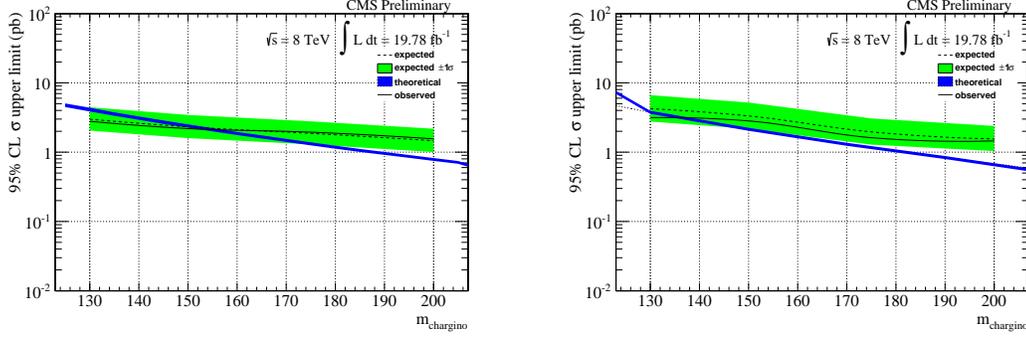


Figure 6.2: The observed cross-section upper limit for the $\chi_2\chi_1^\pm \rightarrow HW^\pm\chi_0\chi_0$ (left) and $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0$ (right) SMSs as a function of the mass of the neutralino (set equal to the mass of the chargino in the left plot) with the mass of the LSP set $m_{\chi_0} = 1$ GeV. The expected limit, computed from the expected background, is shown as a dashed line, while the observed limit from the observed data is shown as the solid line. The expected theoretical cross section is shown as the dotted line with the theory error in blue.

Figure 5.8 one can see that a model with $m_{\chi_2} = 200$ GeV should typically produce three times as many events in the region $200 \leq M_R < 400$ GeV than in the region $400 \leq M_R < 600$ GeV and many of these should be at high R^2 . Since we observe no excess in that region, the observed excess does not resemble the expected distribution for a low mass excess.

We can see from Figure 5.8 that a model with $400 < m_{\chi_2} < 500$ would peak in the right place for the observed excess, but this model is disfavored for other reasons. The first is that the theoretical cross-section, which can be seen in Figure 2.3 is $\leq 4 \times 10^{-3}$ pb for masses of this sort, so we only expect about 1 event in the best case in the data collected so far. The second issue with this model is that we would expect to see an excess in one of the HighPt, Hbb and/or Zbb boxes associated with the excess in HighRes. Recalling the discussion in Section 5.3.4, these boxes are designed to specifically target the signatures of the SMS models under consideration as shown in Table 5.2. The absence of an excess in any of these boxes suggests that the observed deviation is not coming from a signal that looks like the models we consider.

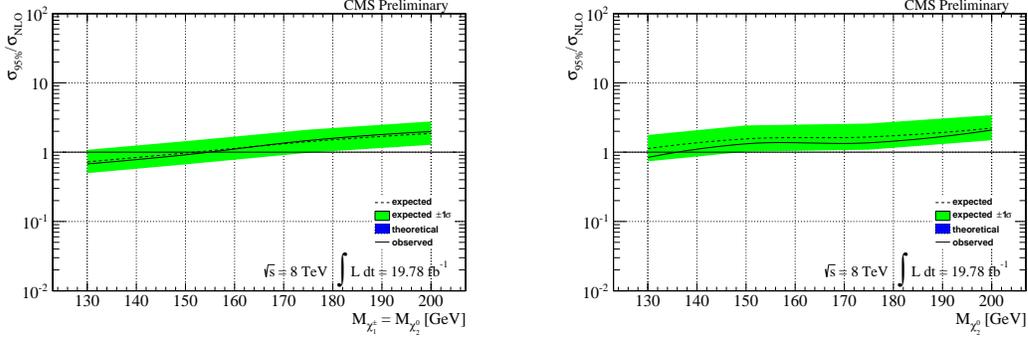


Figure 6.3: The observed cross-section upper limit for the $\chi_2\chi_1^\pm \rightarrow HW^\pm\chi_0\chi_0$ (left) and $\chi_2\chi_2 \rightarrow HH\chi_0\chi_0$ (right) SMSs as a function of the mass of the neutralino (set equal to the mass of the chargino in the left plot) with the mass of the LSP set $m_{\chi_0} = 1$ GeV divided by the theory cross section. The theory error is added in quadrature to the experimental error, though the experimental error dominates everywhere.

6.2 Inverted Analysis Results

Using the procedure detailed in Section 5.8.1 we perform the B and S+B fits to the data in each of the inverse analysis regions. Figure 6.4 shows the distribution of events in $m_{\gamma\gamma}$ in each of the region as well as the two fits. We observe no excess of events in the HighPt2, Hbb, Zbb, and LowRes regions. We observe an excess in the HighPt1 region and in the HighRes region.

6.2.1 Significance of the Inverted Analysis Excess

We measure the significance of the excess in the inverted analysis by looking at the profile likelihood of the number of signal events from the fitted S+B model. We float the AIC weights of the potential background models as described in Section 5.8.1 using the bootstrap covariance matrix. Figure 6.5 shows the profile log likelihood for the two regions that show excesses in the inverted analysis.

The HighPt1 box has 1.4 fitted signal events in the peak. The profile likelihood in Figure 6.5 shows that the $-2\Delta\log(\mathcal{L}(N_{signal} = 0)) = 0.62$, which is a significance of 0.8σ . We expect 0.4 SM Higgs events in this region, so the observed deviation from the SM Higgs hypothesis is $-2\Delta\log(\mathcal{L}(N_{signal} = 0.4)) = 0.30$, which is a significance

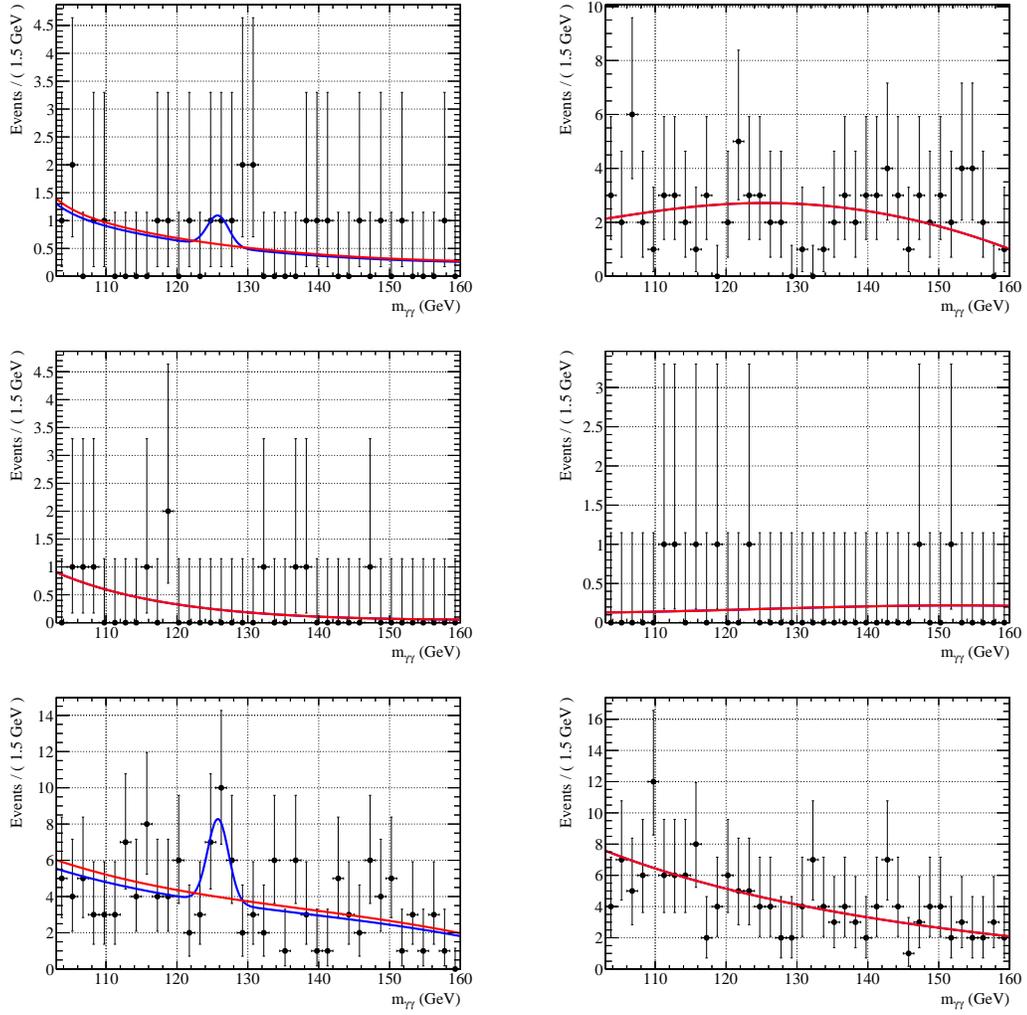


Figure 6.4: The distribution of events in $m_{\gamma\gamma}$ for the inverse analysis for the HighPt1 (top left), HighPt2 (top right), Hbb (middle left), Zbb (middle right), HighRes (bottom left), and LowRes (bottom right) regions. The fits are background only (red) and signal plus background (blue). The background model is the AIC-weighted composite model and the signal model is a single Gaussian with mean constrained by the measured Higgs boson mass.

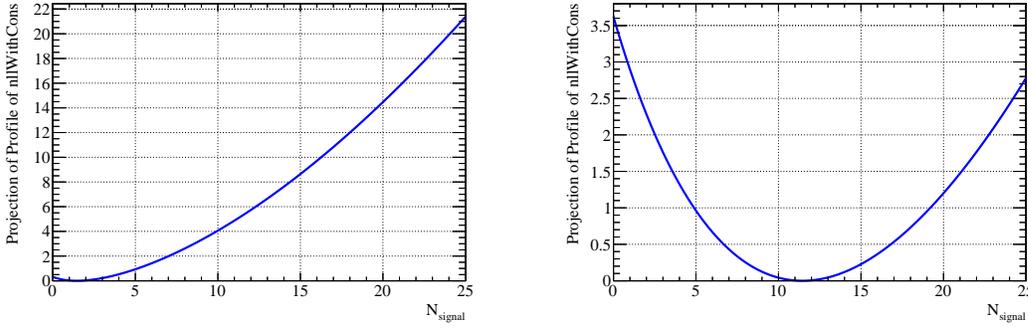


Figure 6.5: The profile log likelihood of the number of signal events for the HighPt1 box (left) and HighRes box (right) for the fit to data with the constrained floated AIC weights background model and the Gaussian signal model.

of 0.6σ .

The HighRes box has 11.5 fitted signal events in the peak. The profile likelihood in Figure 6.5 shows that the $-2\Delta\log(\mathcal{L}(N_{\text{signal}} = 0)) = 7.2$, which is a significance of 2.7σ . We expect 0.8 SM Higgs events in this region, so the observed deviation from the SM Higgs hypothesis is $-2\Delta\log(\mathcal{L}(N_{\text{signal}} = 0.8)) = 6.1$, which is a significance of 2.5σ .

Since the mean of the signal model is constrained by the measured value of the SM Higgs mass and there are only six orthogonal analysis region, the look elsewhere effect for this inverse analysis is very small. Since the inverse analysis is designed as a cross-check analysis, we simply observe that the observed significance in the HighRes box is inline with the look-elsewhere corrected significance in the main analysis. The observed excess in the HighPt1 box has very small significance, and is compatible with the level of agreement observed in the main analysis.

6.3 Excess Events Characterization

The excess of events in the HighRes box is significant enough to justify further characterization of the events to check for any potential problems and to attempt to determine if their properties substantially differ from events outside the signal $m_{\gamma\gamma}$ region and from predicted shapes of Higgs produced through the SM channels. One impor-

tant point from Table 5.10 is that we expect only 0.07 ± 0.02 SM Higgs events in this region (0.06 from $pp \rightarrow ttH$ and 0.01 from $pp \rightarrow VH$), so observing 5 would be extremely unlikely.

6.3.1 Location within the Signal Bin

We begin by zooming in on the region of the HighRes box with the observed excess. Figure 6.6 shows the distribution of events in the bin with the excess for the $m_{\gamma\gamma}$ signal region and for the background predicted from the $m_{\gamma\gamma}$ sidebands. One sees that the background prediction clusters at low M_R and R^2 , which is expected for an exponentially falling distribution, while 4 of the 5 events in the signal region are at relatively large values of M_R and R^2 , and are separated from the background within the plane.

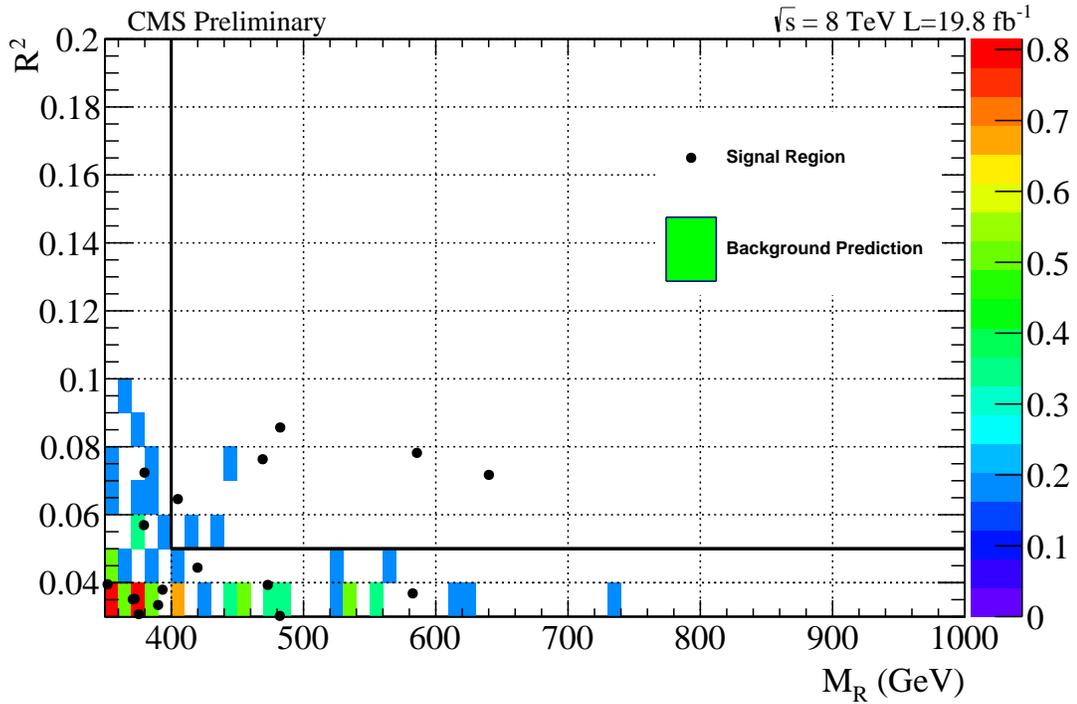


Figure 6.6: The distribution of events in the $m_{\gamma\gamma}$ signal region (black points) compared with the predicted background distribution from sideband (colored regions). The black lines indicate the edge of the bin with the observed excess.

Figure 6.7 shows the R^2 distribution in the HighRes box after various M_R cuts.

One can see the deviation corresponding to the excess in the middle plot. Furthermore, one sees the physical separation of the signal events from the background in the bottom plot with $M_R > 450$ GeV. There we see 4 events in the $m_{\gamma\gamma}$ signal region, and no events in the $m_{\gamma\gamma}$ sideband (there are 0.036 predicted SM Higgs events there). Figure 6.8 shows the corresponding distribution in M_R for two R^2 cuts. The inclusive distribution agrees extremely well, while the distribution for $R^2 > 0.05$ has a clear tail of events in the signal region that are not predicted in the sideband region.

Figure 6.9 shows the distribution of $m_{\gamma\gamma}$ in the region corresponding to the excess ($M_R > 400$ and $R^2 > 0.05$ in the HighRes box) together with a S+B fit. One sees the 5 events clustered in the higgs mass window together with the three events at lower $m_{\gamma\gamma}$ that makeup the sideband prediction. A background only (double exponential) fit is also performed on this distribution, and we find

$$\sqrt{-2(\log(\mathcal{L}_{S+B}) - \log(\mathcal{L}_B))} = 4.02$$

corresponding to a significance of 4.05σ . Figure 6.10 shows the same distribution and fit but with the R^2 cut shifted up by 0.01 in the top plot and the M_R cut shifted up by 50 GeV in the bottom plot. One sees that shifting up in R^2 cuts away two of the events in the $m_{\gamma\gamma}$ sideband, while leaving all 5 in the signal region, while shifting up 50 GeV in M_R gives no sideband events and 4 events in the higgs mass window.

6.3.2 Event Topologies

One can gain some insight into the structure of the events and whether they could be coming from some unlucky conspiracy of detector noise by looking at the layout of objects and the readout of the detector. We want to observe how the Higgs system is interacting with the jet system(s) and the \vec{E}_T^M , as well as the presence of any b-tagged jets in the event. We see in these events no high p_T isolated leptons; there is a single 30 GeV muon in one event, but it is non-isolated and overlaps with a highly b-tagged jet (CSV=0.998) and we conclude that this muon is coming from the secondary decay from the b quark. There are further soft muons, but none with p_T exceeding 7.5 GeV.

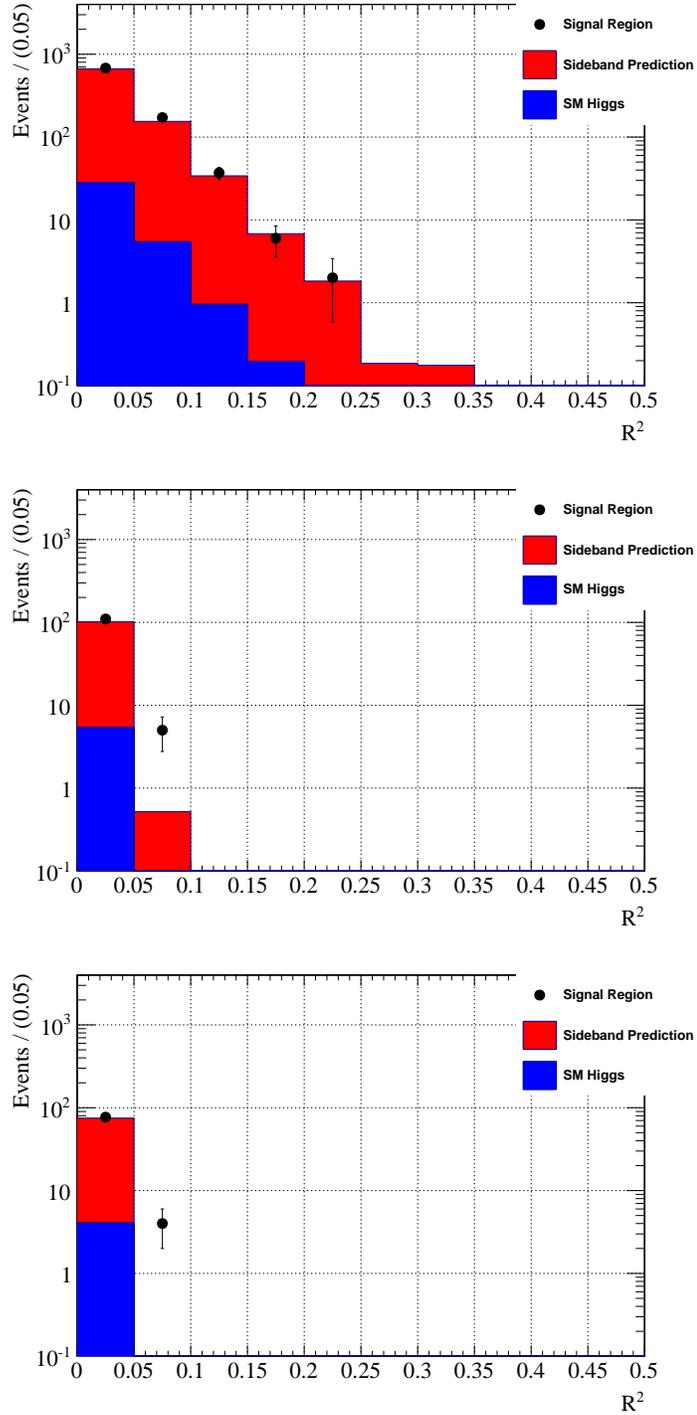


Figure 6.7: The R^2 distribution in the HighRes box for $M_R > 150$ GeV (top), $M_R > 400$ GeV (middle), and $M_R > 450$ GeV (bottom) for data in the signal region (points) and for background prediction (red and blue fill). One sees excellent agreement between the observed and expected distributions for the inclusive sample. One can similarly see the deviation that occurs in the $0.05 < R^2 < 0.10$ bin of the middle distribution corresponding to the observed excess.

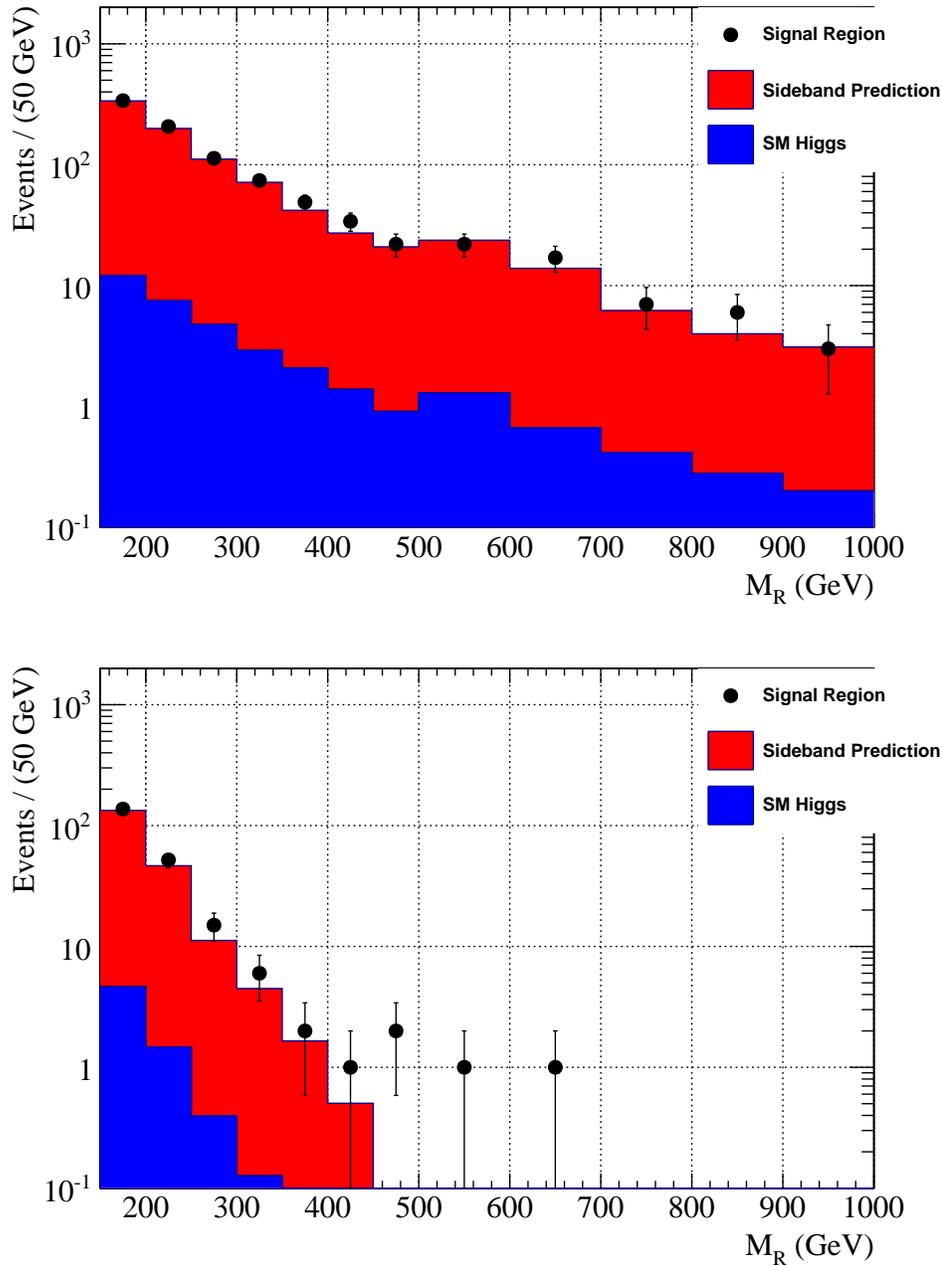


Figure 6.8: The M_R distribution in the HighRes box for $R^2 > 0.00$ (top) and $R^2 > 0.05$ (bottom) for data in the signal region (points) and for background prediction (red and blue fill). One sees excellent agreement between the observed and expected distributions for the inclusive sample. One can similarly see the tail of events in M_R corresponding to the observed excess.

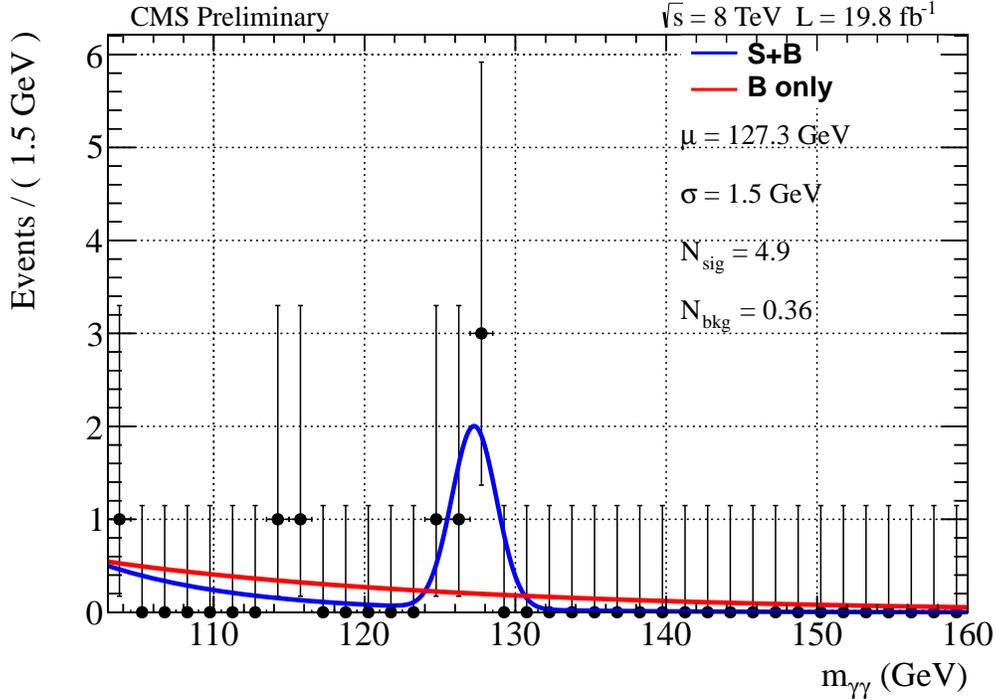


Figure 6.9: The distribution of $m_{\gamma\gamma}$ for events in the HighRes box with $M_R > 400$ and $R^2 > 0.05$ with a signal+background fit consisting of a Gaussian with width fixed to $\sigma = 1.48$ GeV and a double exponential and a background only fit of a double exponential.

Two of the photons in the Higgs system are also identified as electrons, but they fail the CSEV (Section 4.2.1) because they do not have hits in the inner layer of the pixel detector. The tracks are well within the acceptance of the pixel detector, so we conclude that these electrons are from a secondary conversion of the photon rather than a real produced electron.

Figures 6.11-6.19 show the layout of the objects in the five events in the $p_x - p_y$ plane and in the $p_{\parallel} - p_{\perp}$ planes. The Higgs system is shown in red, the jets are shown in black, the mega-jets are shown in blue and the MET is shown in green. Note that when a jet is alone in a mega-jet, the jet line overlaps with the mega-jet line. Distributions of variables including $\Delta\phi$ distributions will be shown in Section 6.3.3, but from these schematics we see no obvious issues with the events, nor suspicious repeated features. One thing we do notice is that all 5 events have their di-photon system pointing in the negative direction in η . While this is unusual, it is not necessarily significant

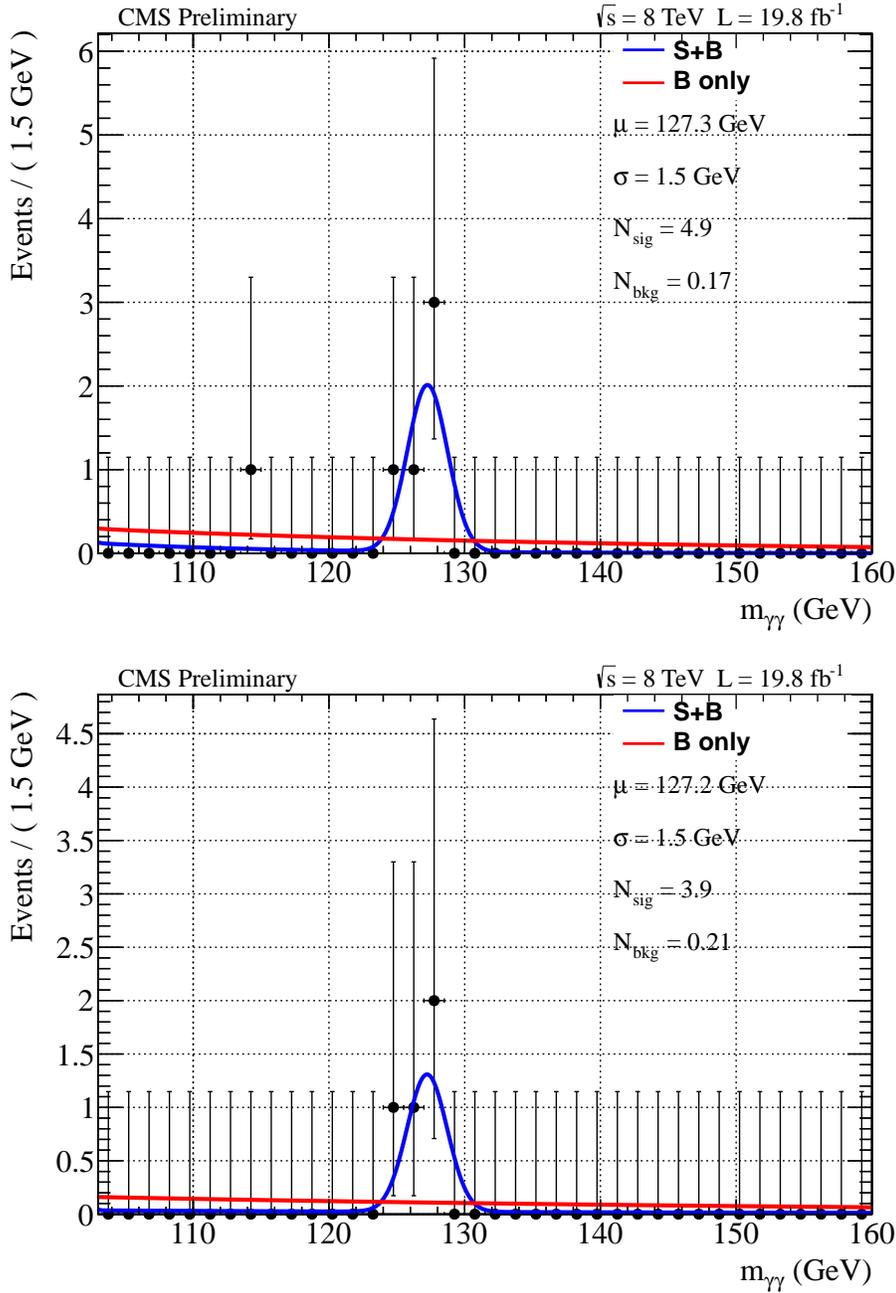


Figure 6.10: The distribution of $m_{\gamma\gamma}$ for events in the HighRes box with $M_R > 400$ and $R^2 > 0.06$ (top) and $M_R > 450$ and $R^2 > 0.05$ (bottom) with a signal+background fit consisting of a Gaussian with width fixed to $\sigma = 1.48 \text{ GeV}$ and a double exponential. The background only fit is a double exponential in the top plot and a single exponential in the bottom (the lack of sideband events makes the double exponential unstable).

(the odds of observing all 5 events with di-photon system on one side of the detector is 6.25%, assuming a symmetric parent distribution). We also note that 2 of the three sideband events have di-photon systems pointing in the positive η direction, and observing 6/8 events in one direction has a probability of 11% (p-value 0.14).

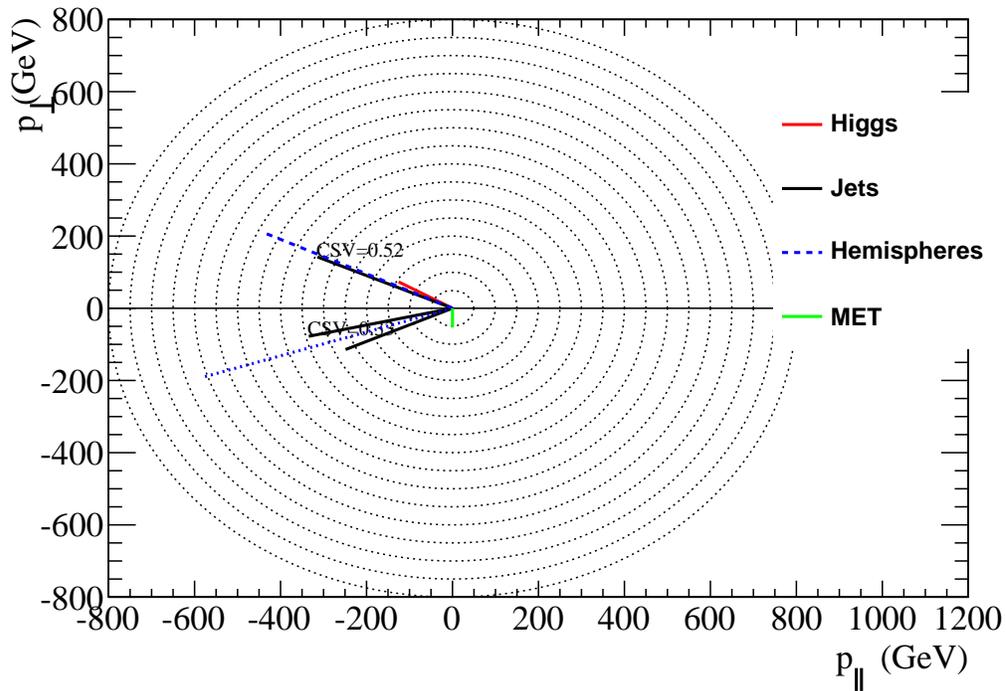
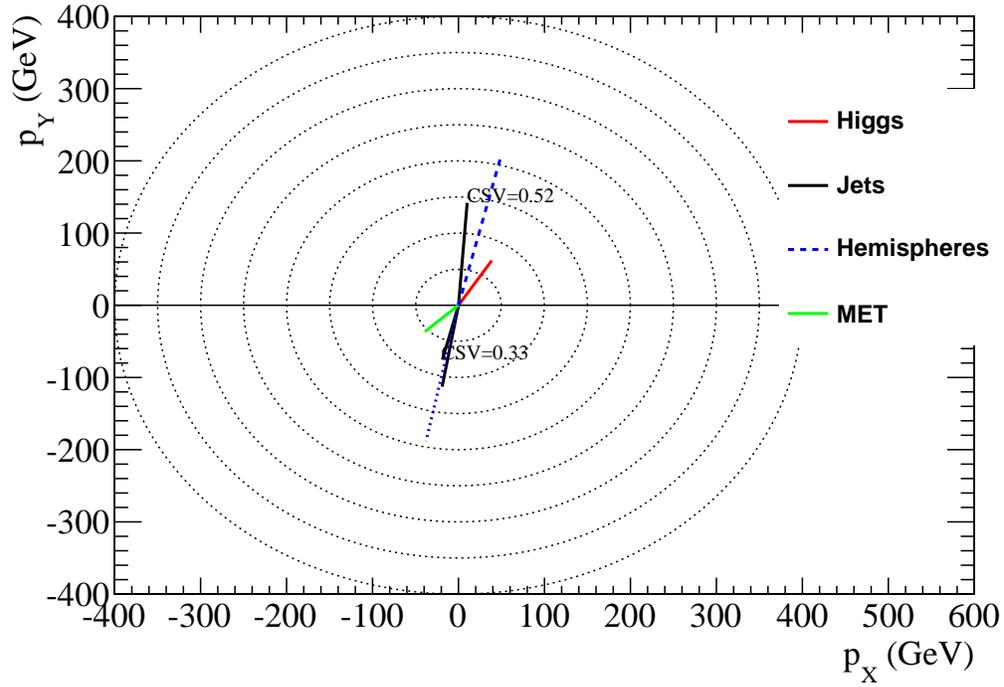


Figure 6.11: Event 0. The layout of the four vectors of objects in the event. The CSV score for any jet passing CSVL is also shown.

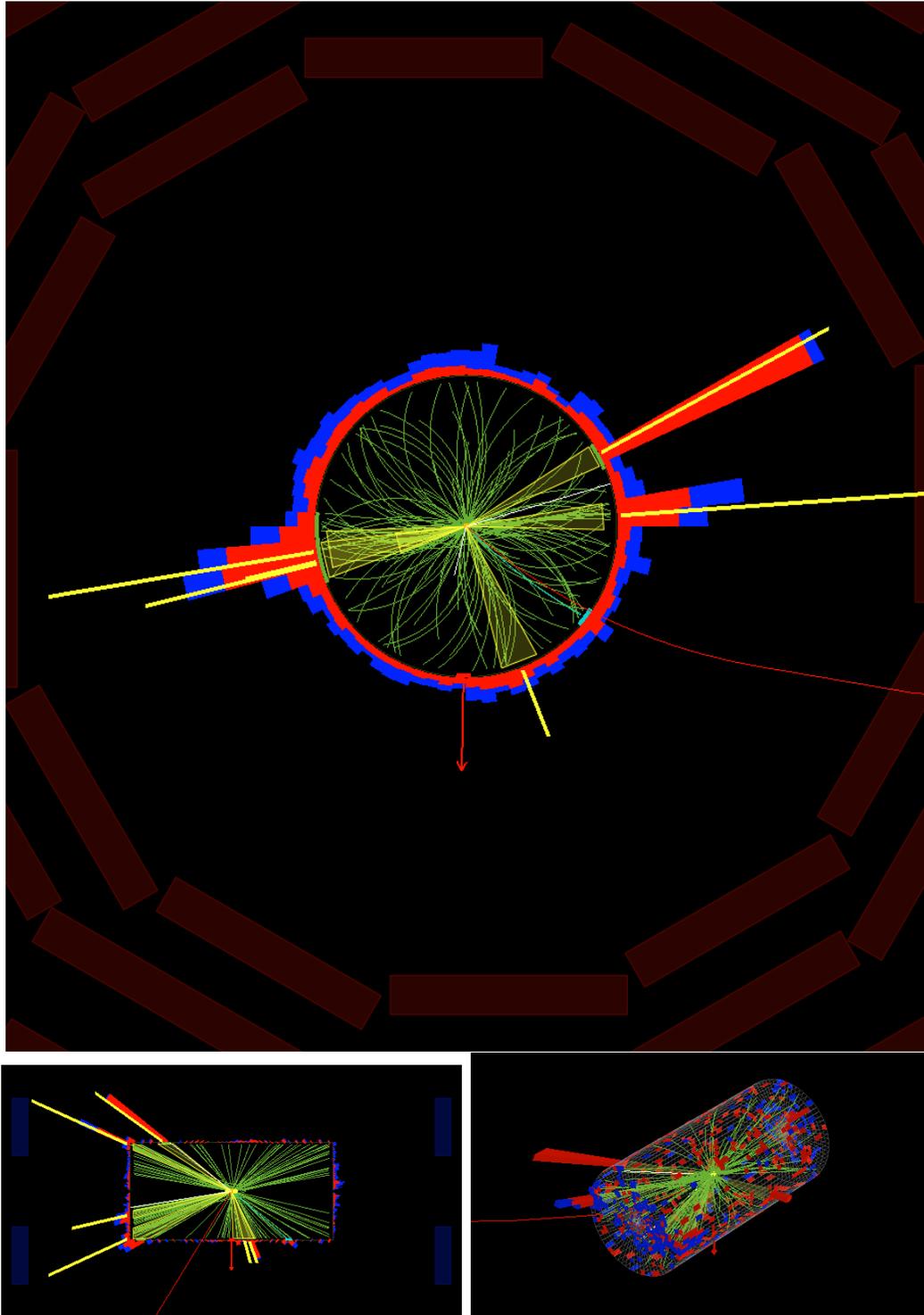


Figure 6.12: Event Display for Event 0

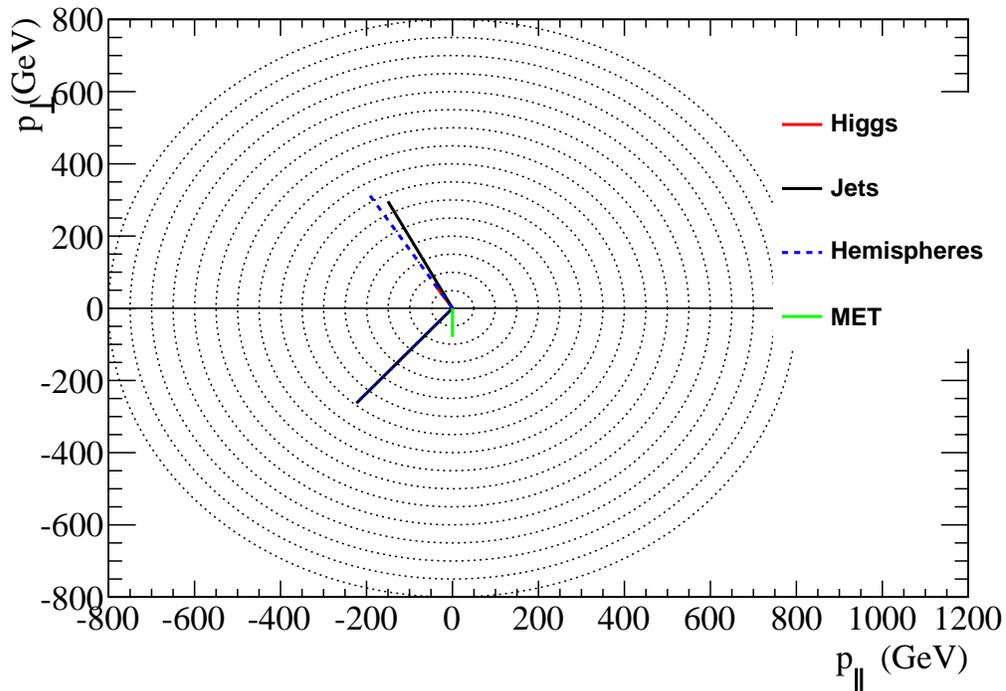
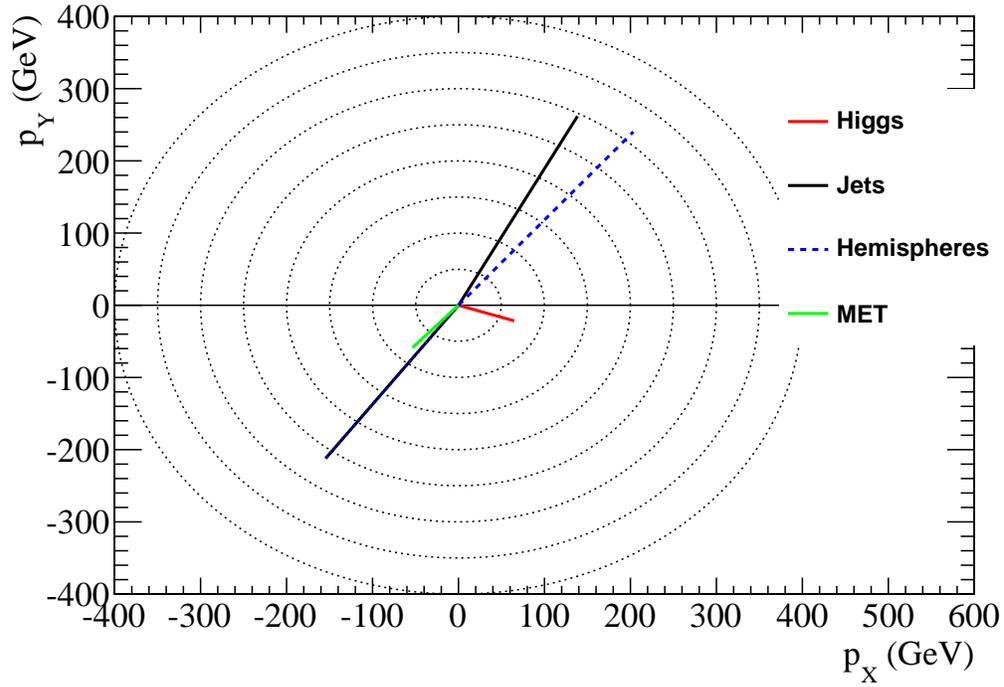


Figure 6.13: Event 1. The layout of the four vectors of objects in the event. The CSV score for any jet passing CSVL is also shown.

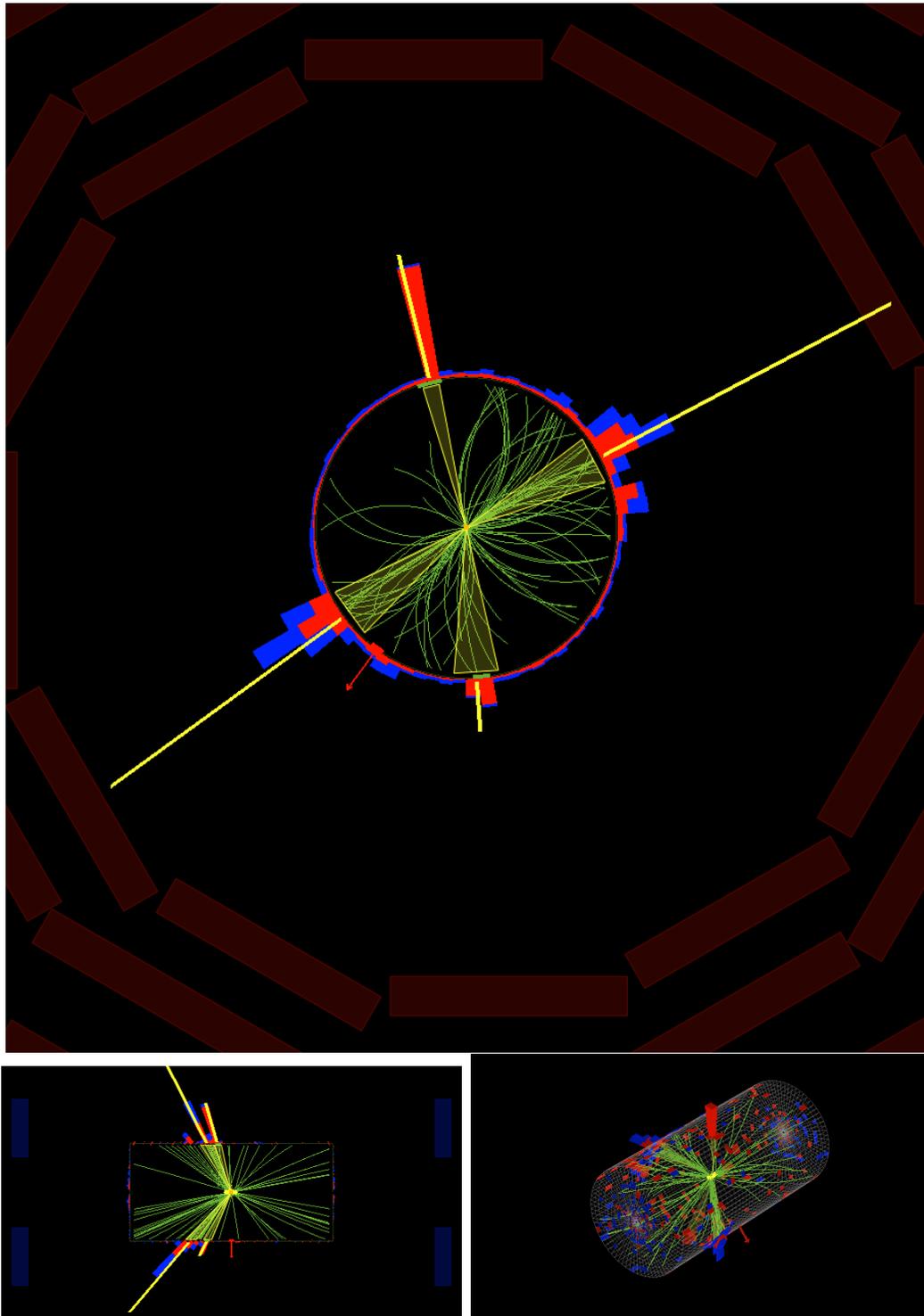


Figure 6.14: Event Display for Event 1

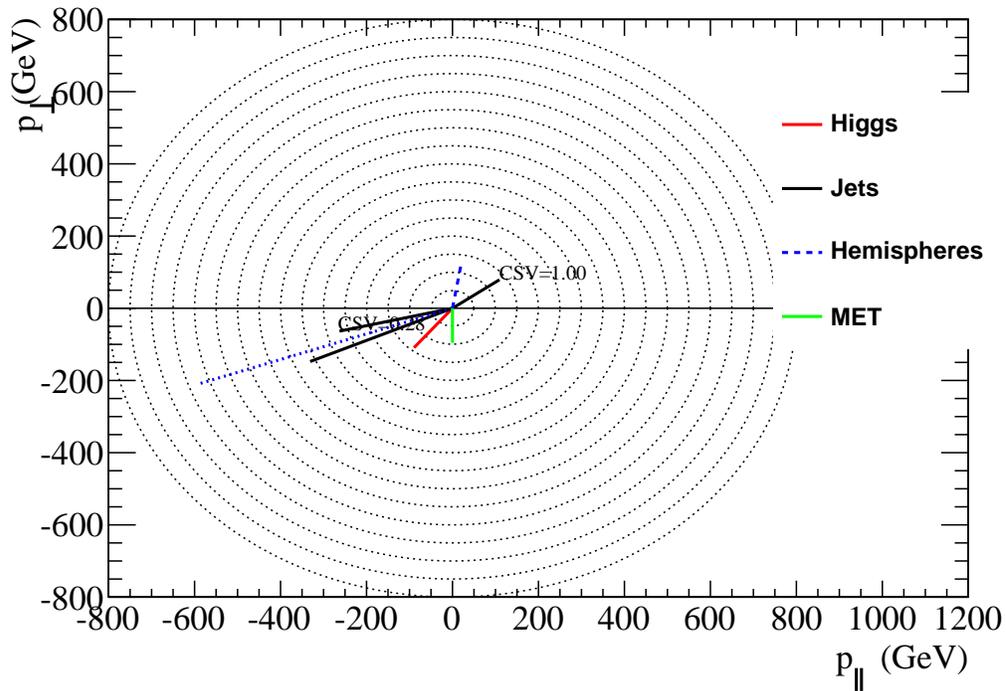
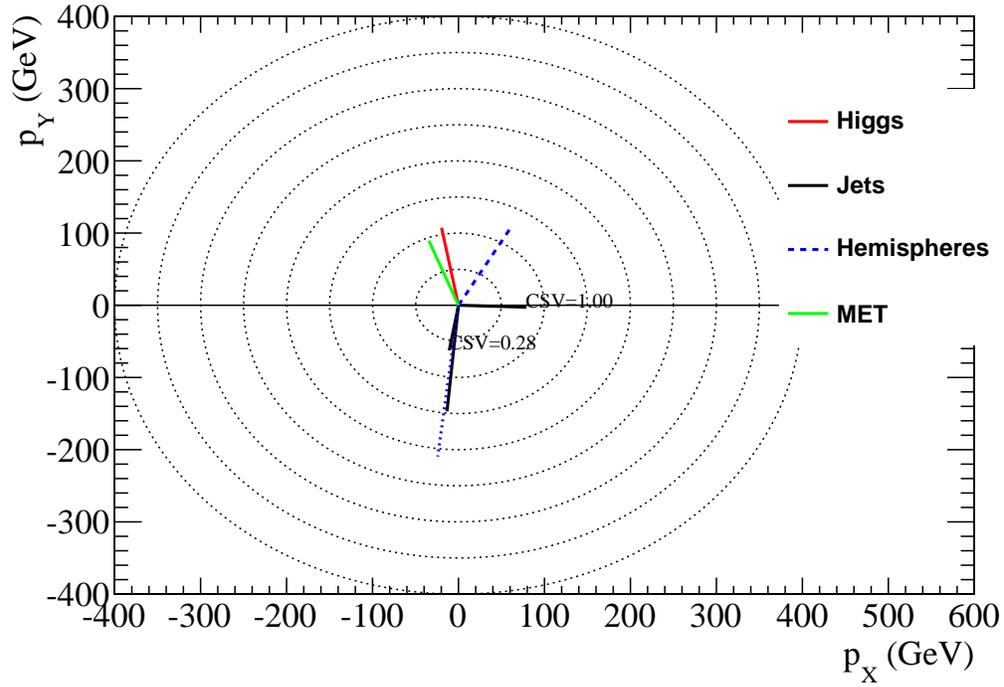


Figure 6.15: Event 2. The layout of the four vectors of objects in the event. The CSV score for any jet passing CSVL is also shown.

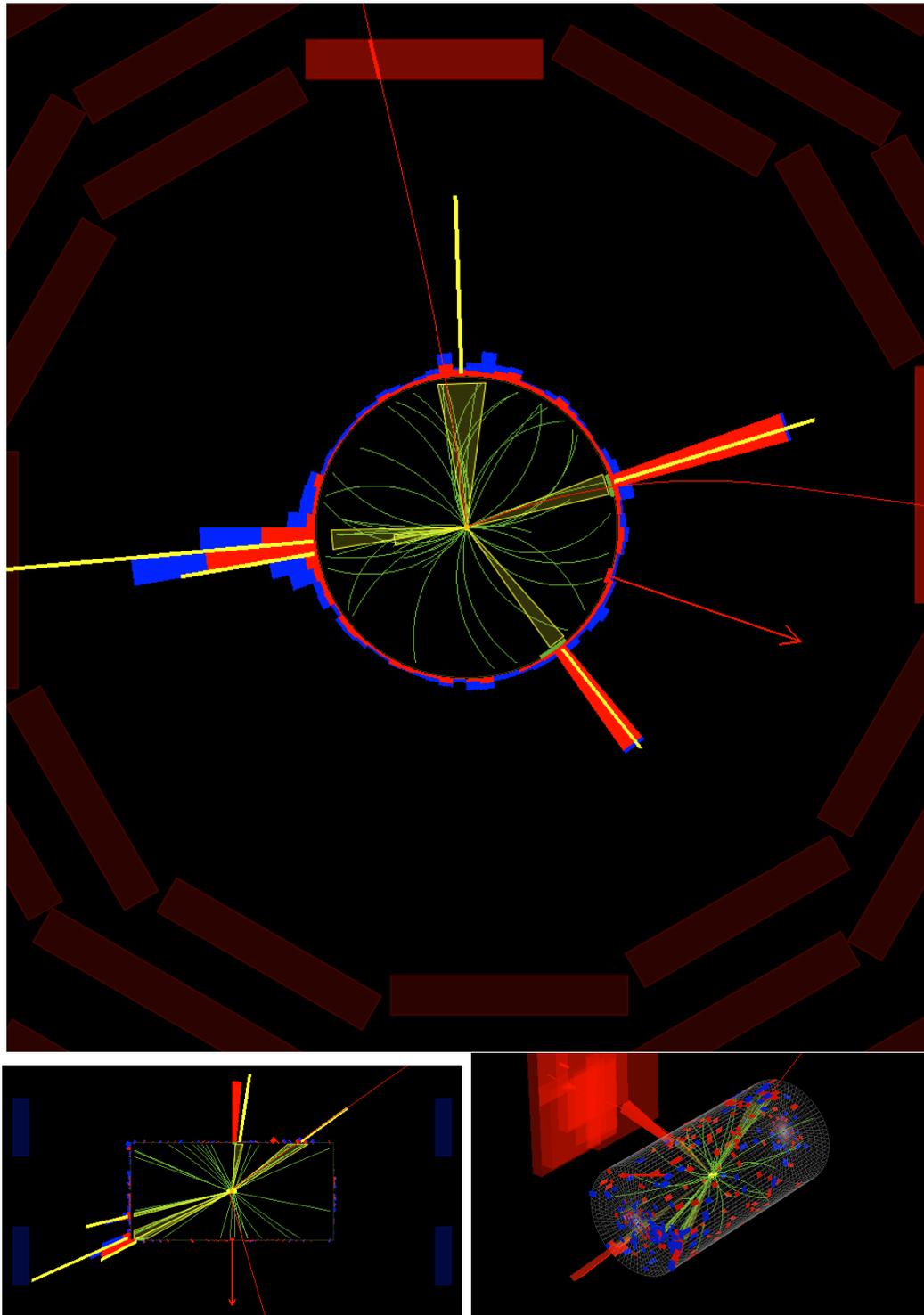


Figure 6.16: Event Display for Event 2

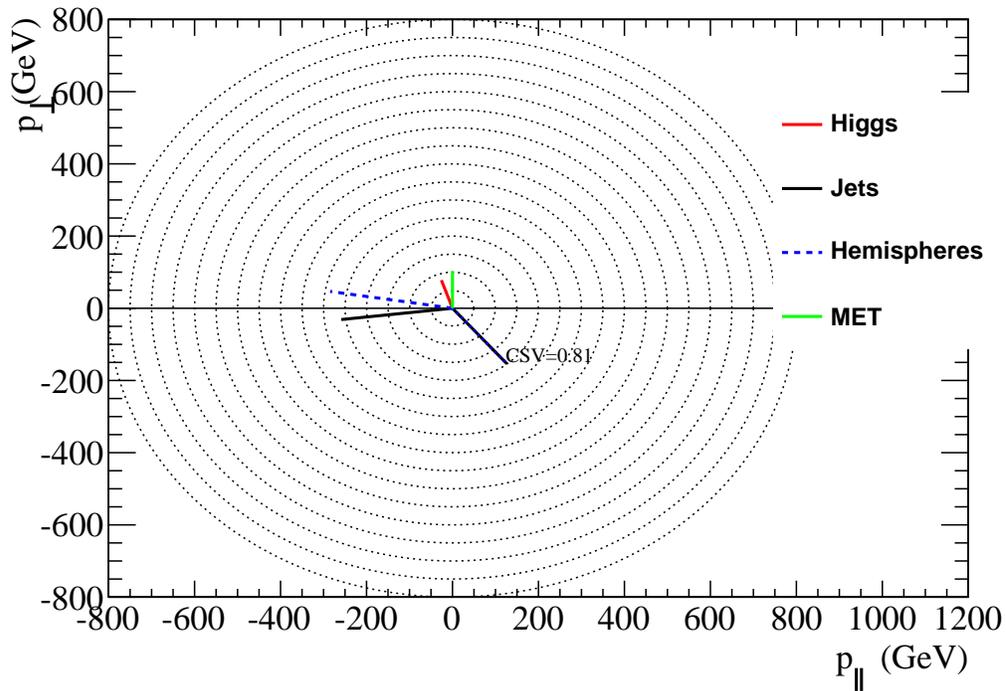
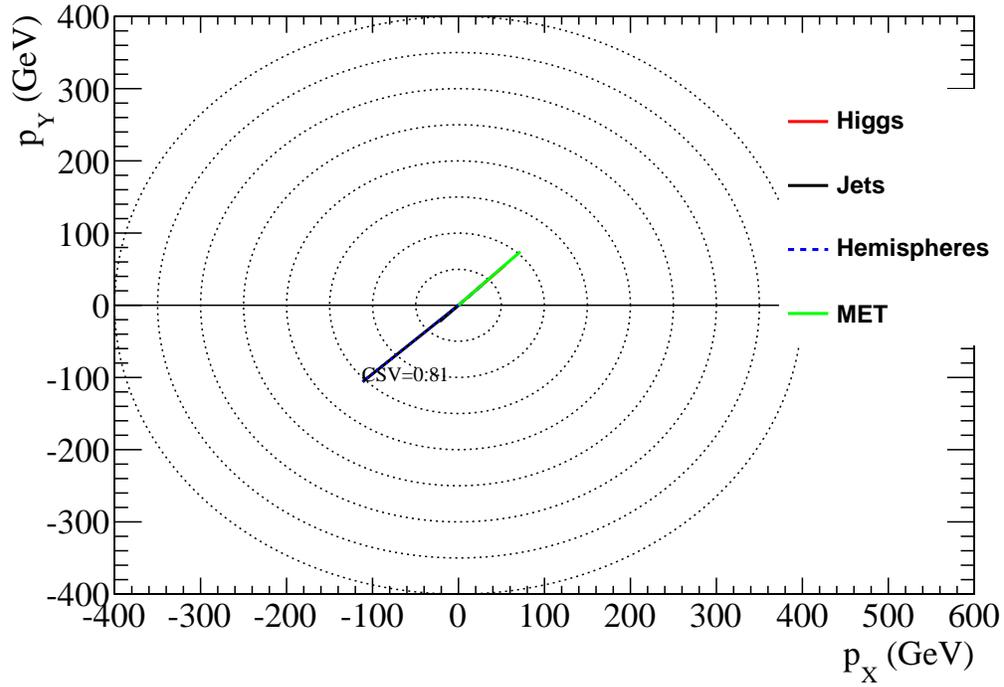


Figure 6.17: Event 3. The layout of the four vectors of objects in the event. The CSV score for any jet passing CSVL is also shown.

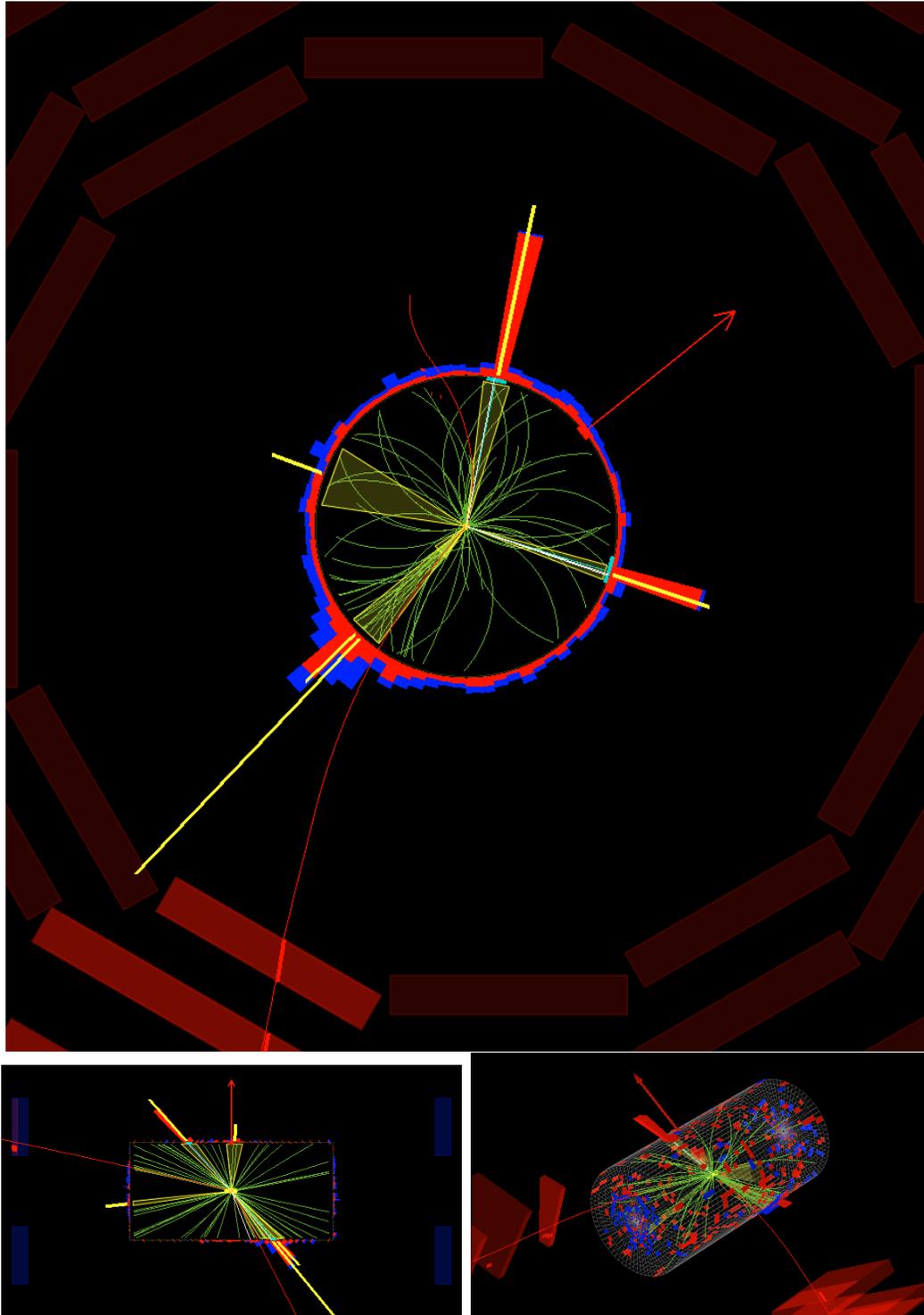


Figure 6.18: Event Display for Event 3

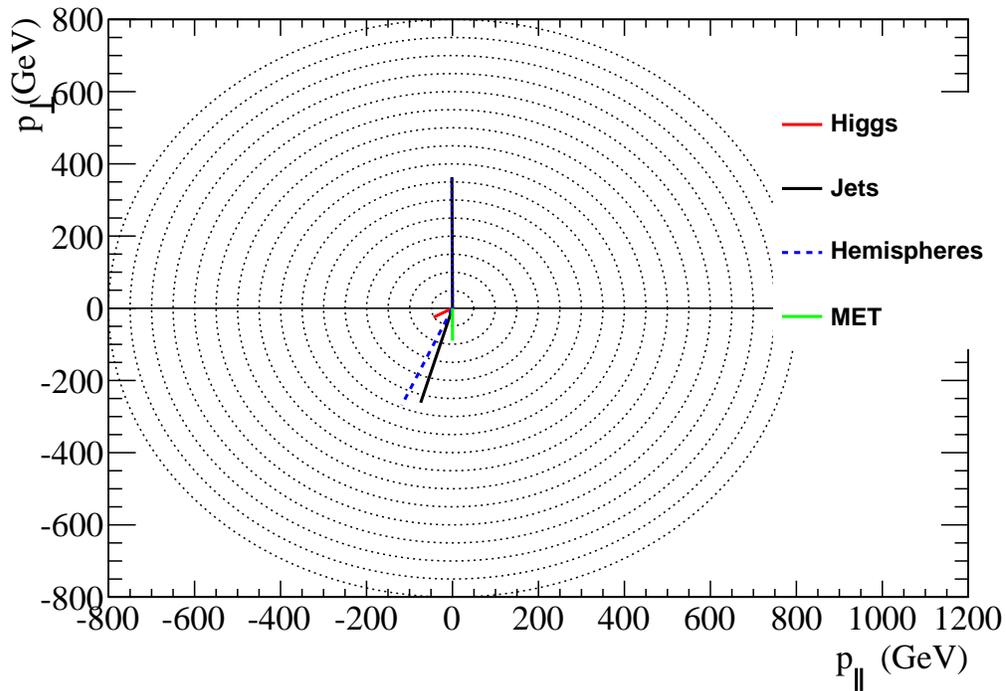
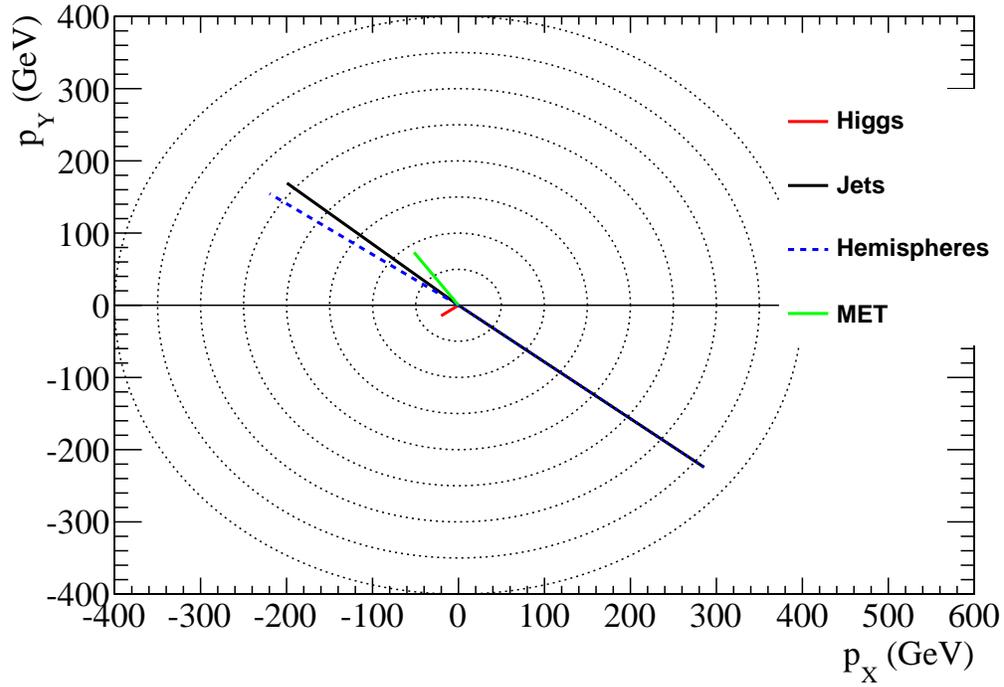


Figure 6.19: Event 4. The layout of the four vectors of objects in the event. The CSV score for any jet passing CSVL is also shown.

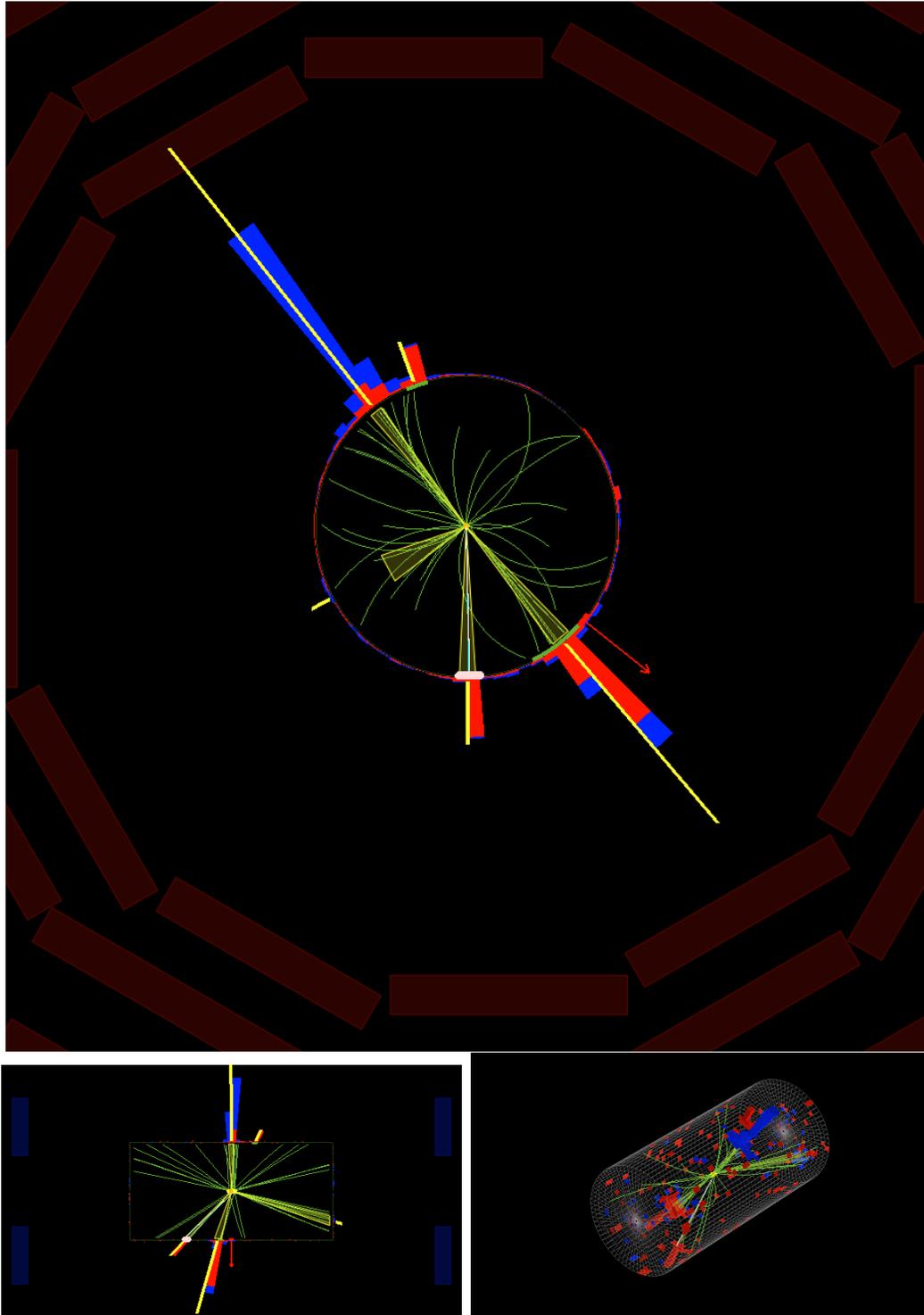


Figure 6.20: Event Display for Event 4

6.3.3 Event Distributions

We proceed to look at the 5 events in several variables to check for any indications of pathological features, and to guide the effort toward classifying these events. To do this effectively, we also compare to various backgrounds:

- **$m_{\gamma\gamma}$ sideband:** Look at events in the HighRes box in the $m_{\gamma\gamma}$ sideband with $M_R > 350$ $R^2 > 0.03$
- **$pp \rightarrow ttH$:** Look at events in the HighRes box in the $m_{\gamma\gamma}$ signal region with $M_R > 400$ $R^2 > 0.05$
- **$pp \rightarrow VH$:** Look at events in the HighRes box in the $m_{\gamma\gamma}$ signal region with $M_R > 400$ $R^2 > 0.05$

We loosen the M_R and R^2 cuts on the $m_{\gamma\gamma}$ sideband in order to get more statistics (there are only three sideband events with $M_R > 400$ and $R^2 > 0.05$). We choose the two SM Higgs processes because they are the only two that have any contribution in this $R^2 - M_R$ region (even though they shouldn't produce this many events).

While the observed number of events is not compatible with ttH or VH based on cross-section, one can use these comparisons to judge whether the deviation could be coming from incorrect kinematics in the MC or from a BSM process that resembles ttH or VH in certain kinematic quantities.

6.3.3.1 Global Event Variables

We begin by looking at the H_T and E_T^M of the events in Figure 6.21. One sees that the H_T of the 5 events is not very consistent with the $m_{\gamma\gamma}$ sideband data, but could conceivably be sampled from either of the Higgs distributions. We see a similar trend in the E_T^M distribution: the events do not look like continuum background, but are more consistent with the Higgs processes.

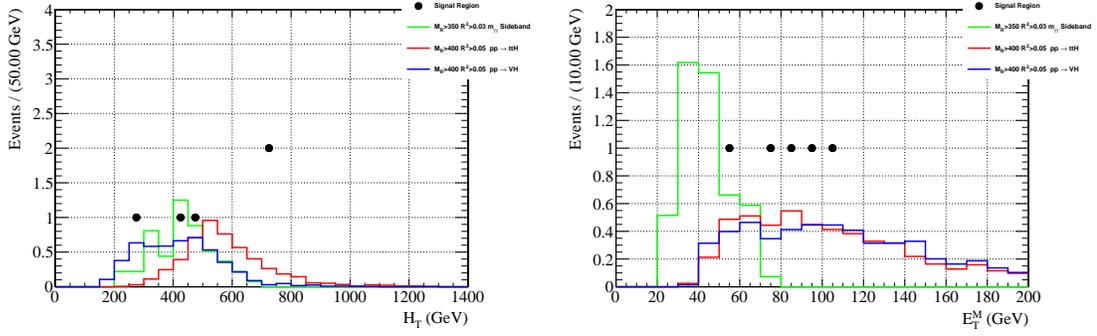


Figure 6.21: The H_T (left) and E_T^M (right) for the 5 events (black points) and 3 backgrounds (green, red, blue).

6.3.3.2 Photon Variables

We next look at the variables coming from the photons in the events. Figure 6.22 shows the $p_T^{\gamma\gamma}$ of the events. We see that it is relatively uniform and consistent with the backgrounds and SM Higgs. This tells us that we don't have a clustering of events near either the analysis selection cutoff (20 GeV) or the HighPt box cutoff (110 GeV).

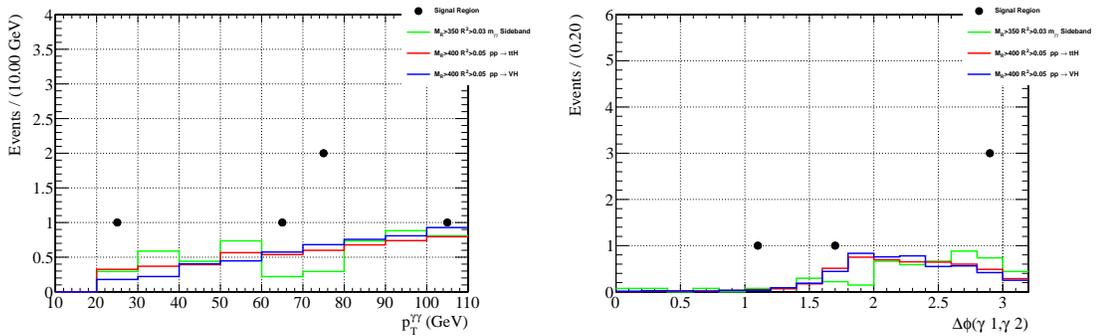


Figure 6.22: The $p_T^{\gamma\gamma}$ (left) and the $\Delta\phi$ between the photons (right) for the 5 events (black points) and 3 backgrounds (green, red, blue).

In figures 6.23 and 6.24 we see the p_T and η distributions of the photons in the events. These show no obvious pathologies and the backgrounds, SM Higgs and data are all consistent. The already observed tendency of the di-photon system is manifest here as well, but adds no additional information from the probabilities already observed.

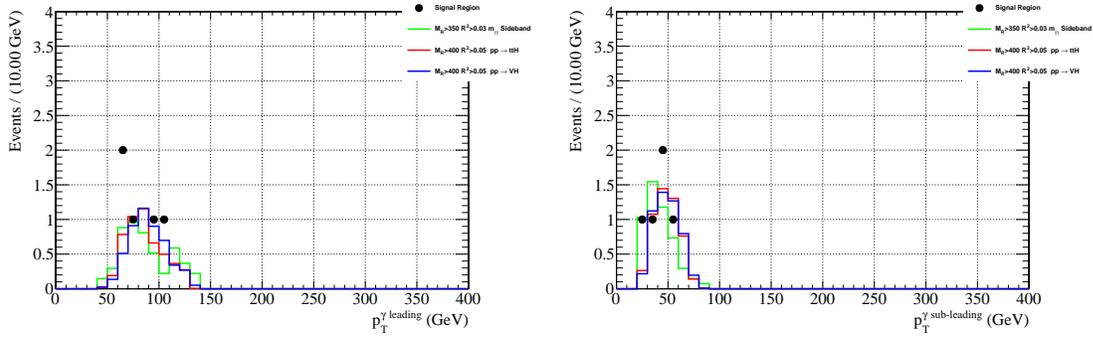


Figure 6.23: The p_T for the leading (left) and subleading (right) photons for the 5 events (black points) and 3 backgrounds (green, red, blue).

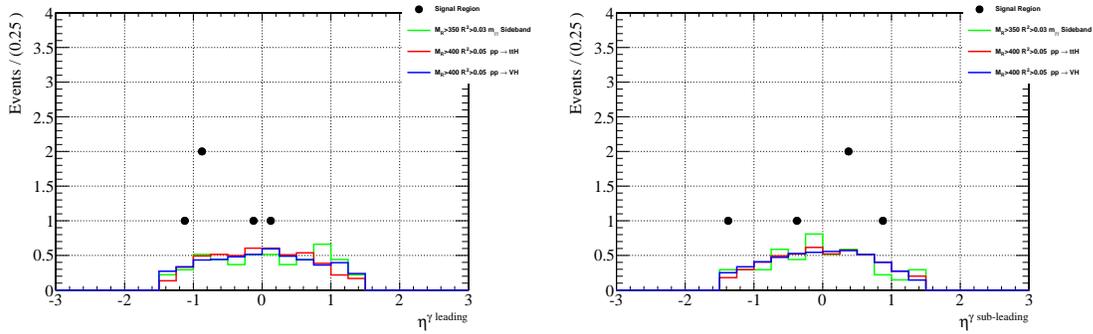


Figure 6.24: The η for the leading (left) and subleading (right) photons for the 5 events (black points) and 3 backgrounds (green, red, blue).

6.3.3.3 Mega-Jets

We next look at the kinematics of the mega-jets. Figure 6.25 shows the p_T of the mega-jets of the events, where we observe that the distribution of the 5 events is somewhat broader than any one process would predict, but not incompatible. Figure 6.27 shows the reconstructed invariant mass of the mega-jets. Note that if a mega-jet contains only one jet it will have a mass of zero, while if it contains only the diphoton system it will have an invariant mass equal to $m_{\gamma\gamma}$. One sees several interesting trends in this figure that illustrate the physics going into the mega-jets. The VH MC has a peak at 125 GeV, which implies that a substantial number of events have the higgs alone in a mega-jet, which one would expect from the final state configuration of a higgs recoiling against a vector boson. The ttH has a peak at 300 GeV, indicating that the Higgs is often paired with one of the two top quarks in the event. The background

here has a peak at the top mass, indicating that there are a number of tops in the non-resonant background.

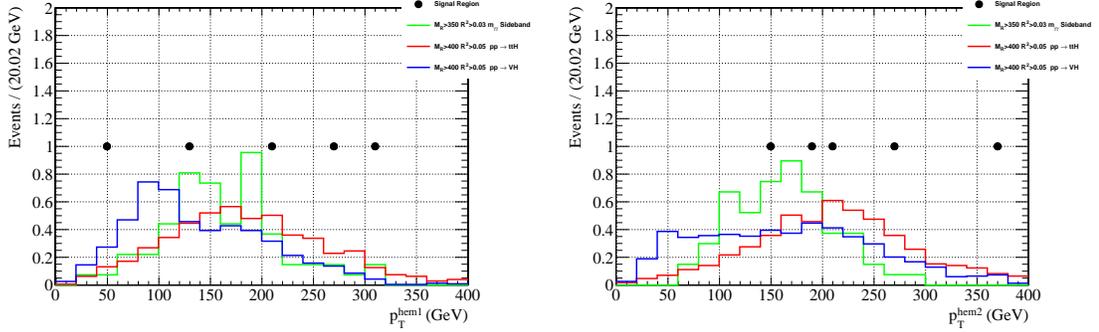


Figure 6.25: The p_T for the mega-jet containing (left) and not-containing (right) the di-photon system for the 5 events (black points) and 3 backgrounds (green, red, blue).

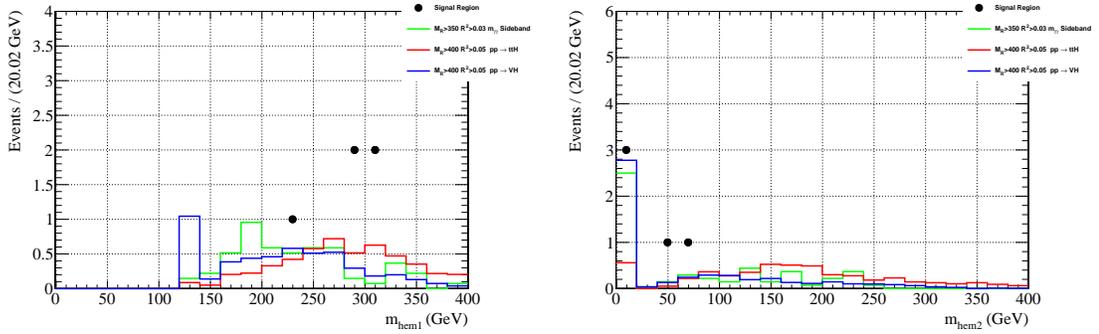


Figure 6.26: The mass of the mega-jet containing (left) and not-containing (right) the di-photon system for the 5 events (black points) and 3 backgrounds (green, red, blue).

Figure 6.28 shows the $\Delta\phi$ between the mega-jets and between the mega-jets and the MET vector. We see that the configurations observed in the 5 events are consistent with both the $m_{\gamma\gamma}$ sideband data and the MC.

6.3.3.4 Jets

Figure 6.29 shows the p_T of the jets in the events. We see that several of the events have jets that are somewhat harder than would be predicted by any of the MC or the sideband data. Since the backgrounds are all in the same kinematic region, it

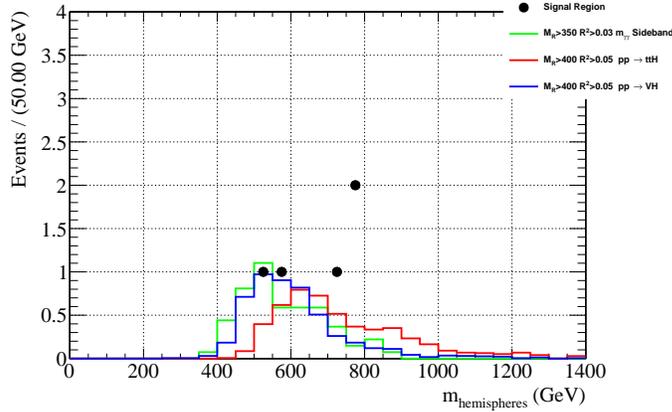


Figure 6.27: The invariant mass of the two mega-jets for the 5 events (black points) and 3 backgrounds (green, red, blue).

suggests that the R^2 and M_R cuts alone are not sufficient to select such unusually high p_T jets.

Figure 6.30 shows several other features of the jets. The top plot, the number of jets, suggests that these events are not consistent with ttH configurations, which would tend to produce many more jets than we observe. Indeed, this plot is strong evidence that the observed excess is not coming from an excess of ttH events. The other two plots are the CSV score of the highest and second highest scored jets in the events. We see that ttH, since it produces two real b jets, tends to have at least 1 jet with a very high CSV score, while VH tends to have a very low score, since it does not usually produce b-jets.

6.4 Conclusion and Future Work

We have performed an analysis looking for non-standard production of the Higgs boson using the razor kinematic variables. We have used the $H \rightarrow \gamma\gamma$ decay mode to give access to a narrow resonance on top of smoothly falling background to allow a fully data-driven prediction for the standard model background. We have used the R^2 and M_R variables to select regions of phase space with small expected resonant and non-resonant standard model background and observed an excess with a local significance of 3.4σ .

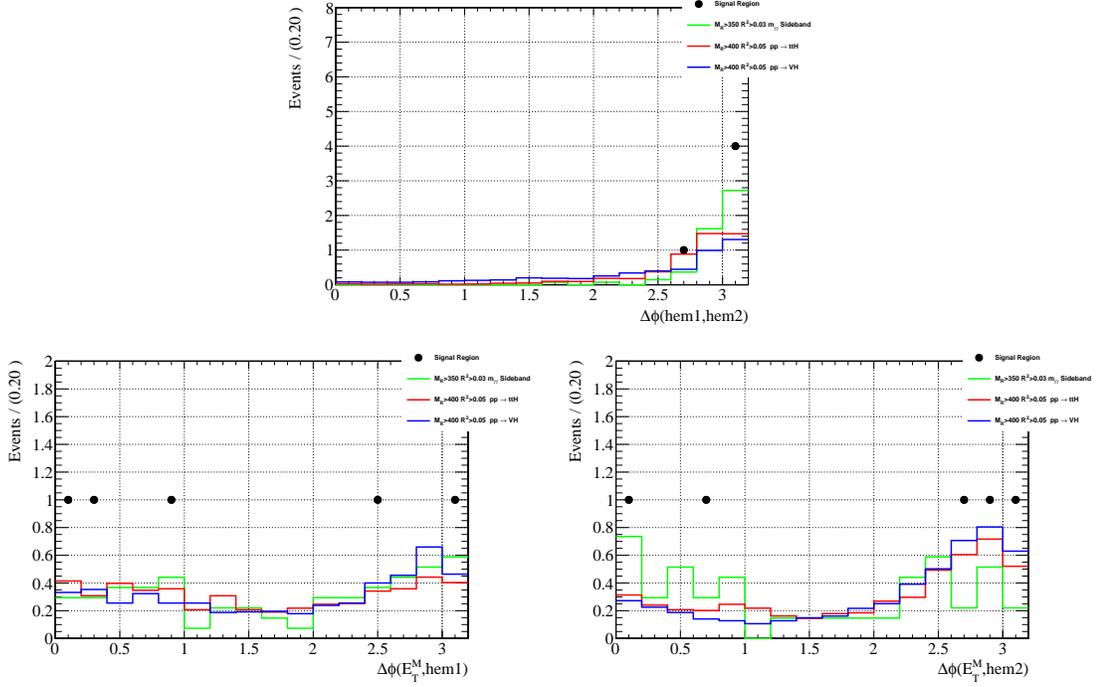


Figure 6.28: The $\Delta\phi$ between the two mega-jets (top), the mega-jet containing the higgs and the \vec{E}_T^M (left) and the mega-jet not-containing the higgs and the \vec{E}_T^M (right) for the 5 events (black points) and 3 backgrounds (green, red, blue).

We have performed a statistical analysis of the results and found that there is no excess consistent with the SUSY simplified models targeted by the analysis. We therefore set limits on the production cross-section of these simplified models at the several pb level. The location of the observed excess makes it incompatible with the expected shape of an excess for either of the models under consideration, so it has no effect on the limits.

We have cross-checked this result using a logically inverted analysis that uses the razor variables to reject background and then makes the $m_{\gamma\gamma}$ fit the signal sensitive part of the analysis. We have seen that this method also shows an excess with a slightly smaller significance, but less potential correction from the look elsewhere effect. The number of excess events is larger in the inverted analysis, but the significance is smaller because of the larger background.

We looked at the 5 event excess in a variety of different ways, including its location in the 3D $m_{\gamma\gamma}$ - R^2 - M_R plane, which is consistent with a Higgs boson being produced

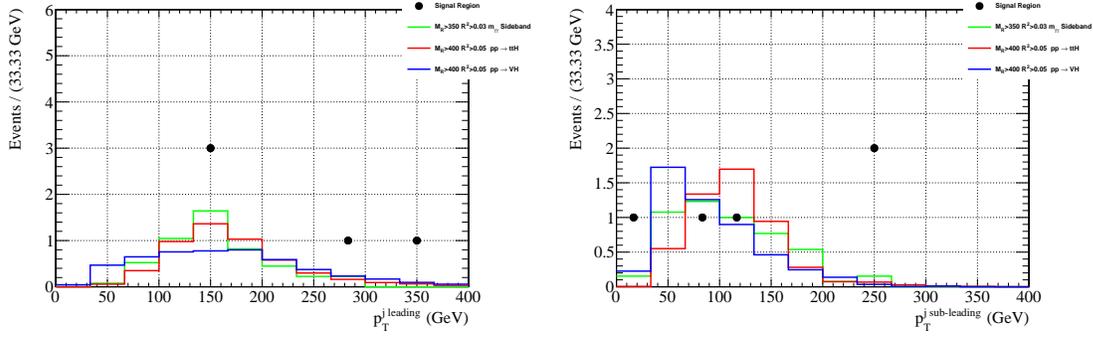


Figure 6.29: The p_T of the leading (left) and subleading (right) jets for the 5 events (black points) and 3 backgrounds (green, red, blue).

in a non-standard way. The locations of the events in the R^2 - M_R plane are quite separate from the predicted background, and the requirement of a $m_{\gamma\gamma}$ pair near the Higgs boson mass makes it very unlikely that this excess could be produced by some detector malfunction, which is also confirmed by looking at the detector readouts and object configurations. Indeed the fact that these events are peaking in both the $m_{\gamma\gamma}$ plane and in the R^2 - M_R plane, which are largely uncorrelated, makes it less likely that this excess is coming from some mundane source.

Based on the predicted differential cross-section of the SM Higgs boson in the R^2 - M_R region where the excess occurs, we can largely discount the possibility that this is simply a fluctuation in the number of SM Higgs boson events. This is especially true when one considers that most of the SM Higgs boson events in this region should be produced by $pp \rightarrow t\bar{t}H$ production, and the number of jets observed in our excess is not compatible with the $t\bar{t}H$ hypothesis (Figure 6.30). Without $t\bar{t}H$, we are left with only 0.1 predicted events coming from vector-boson associated production, and several other distributions from Section 6.3.3 make this prospect less likely.

It is possible, therefore, that this is an early stage of a detection of beyond the standard model production of the Higgs boson. While the significance of the excess is too low to make a definitive statement, several pieces of evidence suggest that this excess does not look very much like a fluctuation of the background or a known production mechanism of the Higgs boson. With only five events, it is very hard to say much more about what this could be, a large variety of models could potentially

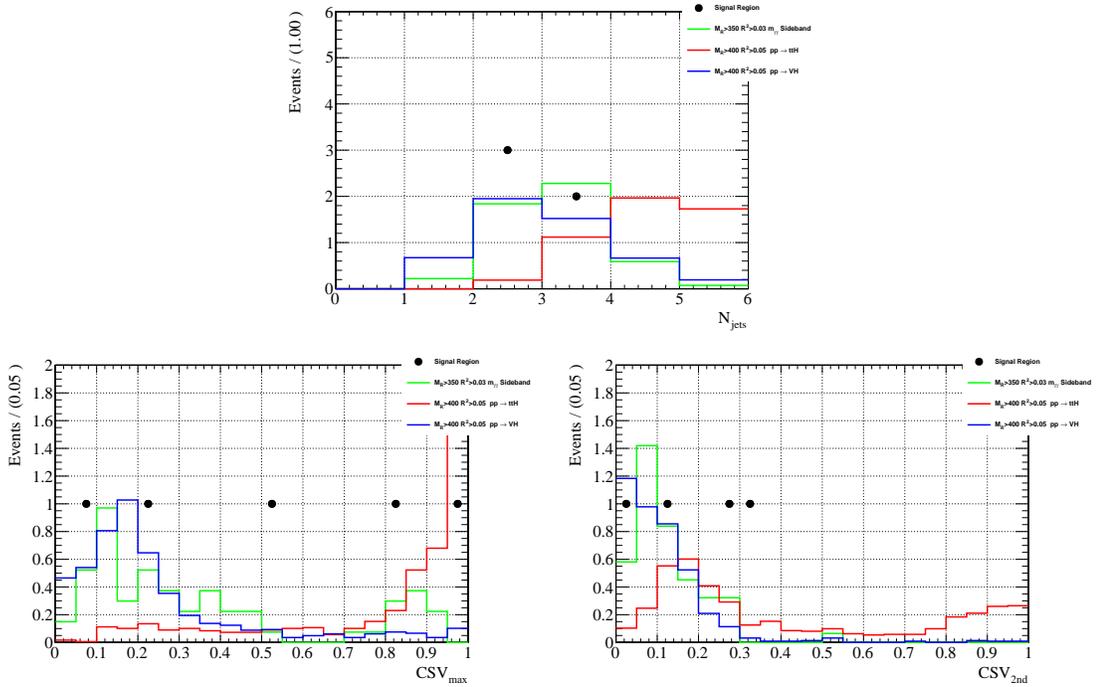


Figure 6.30: The number of selected jets in the event (top) and the CSV score of the highest CSV jet (bottom left) and the second highest CSV jet (bottom right) for the 5 events (black points) and 3 backgrounds (green, red, blue). One event has its second jet falling out of tracker acceptance, and so its CSV score is -10 and it is not included on the plot.

fit this observation.

The design of the analysis and the lack of an excess in other boxes can inform model building to some extent. The lack of an excess in either of the Hbb or Zbb boxes suggest that this is single Higgs production. Certainly any model that predicts additional H or (to a lesser extent) Z would have to explain the null observation in these boxes. The lack of excess in the HighPt box suggests the Higgs is not highly boosted, although the larger background at high M_R in the HighPt box could cover this to some extent. Little can be said from the lack of excess in the LowRes box since, by design, this box should be much less sensitive to events with real photons.

Some work is already being done to evaluate models that could produce this sort of resonance. One curious observation is that the events are all at similar values of R^2 , typically with $R^2 \approx 0.07 - 0.08$. Most SUSY models have broad distributions in R^2 (see, for instance, Figure 5.7) and finding one compatible with this clustering

in R^2 is challenging. This is a place where more data is vital, since it would help explain whether this clustering is a real phenomenon or a statistical effect. Similarly the observed clustering of the mass of the mega-jet containing the Higgs boson near 300 GeV could be a real phenomenon, which might give some insight about what the production mechanism is.

Because of these considerations, this analysis must be repeated with more data and at higher energy during Run 2 of the LHC. This will help resolve the questions about the observed clustering in some distributions. The relative yields at different center-of-mass energies will provide some insight into the production mode if this excess is confirmed, and higher statistics will allow finer binning in the R^2 - M_R plane and more analysis boxes to isolate different effects. The potential for lower pileup due to a 25 ns run of the LHC will also be a welcome addition to this analysis, since identifying photons at high pileup is especially challenging and our efficiencies suffer as a result. Finally, additional signal hypotheses will be helpful in guiding the design of an evolved analysis. The shape of the excess observed in this analysis suggests other types of signals might be more appropriate, so using other hypotheses that fit the observed yield better will be helpful to better tune the Run 2 analysis.

Part II

Optimization of Higgs boson identification using a quantum annealer

Chapter 7

Introduction

In part one we looked for the Higgs boson in non-standard kinematic configurations to look for beyond the standard model physics producing standard model Higgs boson, but it hasn't yet been established that the particle observed by CMS and ATLAS is the standard model Higgs boson. To show that this is the SM Higgs boson and to understand the true form of the EW symmetry breaking Lagrangian requires all of its properties to be measured to extremely high precision. To do this at a particle collider, one needs to produce a lot of Higgs bosons and have high efficiency to select them and reject background. We focus here on how to solve this second problem specifically in the context of the Higgs boson decay to two photons.

The search for the SM Higgs boson in the two photon decay channel at both ATLAS and CMS use a variety of multivariate analysis (MVA) techniques to enhance signal sensitivity [59,80,81,90]. Typically these use a variety of boosted decision trees and multilayer perceptrons to exploit deep correlations that differ between the signal and background samples and so select events that don't on their surface resemble their true classifications. This is an extremely powerful technique used in a variety of areas, but there are some limitations when it comes to the application to particle physics. The first is that the training samples are taken from MC simulation of the process in question, which must be fed through a simulation of the detector to the expected signal in the detectors. This simulation is quite mature, but still has some trouble especially with deep non-linear correlation between observables, which is specifically what is used in these MVAs.

The second issue is the sheer number of different processes that form the background to most Higgs boson searches. Without a specific model of the alternate hypothesis, an MVA anomaly detector needs to be used, which essentially looks in the data for configurations not predicted by the data. Unfortunately, unless the anomaly detector knows about every single possible background process, it can easily promote some rare background to attention as an anomalous signal. There are techniques to help deal with this, but it remains an extremely difficult problem when using these sorts of MVAs.

We propose to use a simpler type of classifier, still formally an MVA but with far less ability to get stuck on deep non-linear correlations between variables. This classifier relies on explicitly linearized correlations to perform the classification and protects itself from overtraining by using binary connections. This configuration makes it much easier to control, and if need be correct, correlations between variables in the MC and the binary nature of the connection makes it much more robust against getting stuck in the possibly highly non-convex optimization space. In essence this classifier is not dissimilar to optimizing a cut based analysis for signal sensitivity, but does so using many more potential variables.

These advantages come at a price, training this sort of classifier is a provably NP-hard problem, meaning that the training time increases exponentially in the number of input variables. This sort of classifier typically needs large numbers of highly correlated input variables to attain good performance, so this has traditionally been a serious problem. We use the newly viable technique of *quantum annealing* to train the classifier. This technique relies on the quantum adiabatic theorem to negotiate its way through the non-convex solution landscape to find the optimal network configuration in, ideally, polynomial time.

We use an implementation of a quantum annealer produced by D-Wave, Inc. a company based out of Burnaby, BC, Canada. The device we use is owned by the Lockheed Martin Corporation and jointly operated by Lockheed Martin and the University of Southern California. It has not been proven conclusively that the machine is indeed a true quantum annealer, but it has been shown that it can solve problems

correctly.

In Chapter 8 we discuss the theory of annealing and its application to solving problems. We begin with simulated annealing, which simulates on a computer the different but related algorithm of thermal annealing. This algorithm solves a problem by converting it into a energy landscape and simulating a state influenced by thermal fluctuations in that landscape. The thermal excitations of the state allow it to escape local minima in the energy landscape and as the temperature is gradually lowered gets progressively frozen into the solution. We then discuss the theory of quantum annealing, which has a conceptually similar evolution except it has no thermal fluctuations and uses quantum tunneling to escape from the valleys of local minima in the ground state.

In Chapter 9 we discuss the specific implementation of the classifier we use and how it is trained. We discuss how we cast the training problem into a form that can be solved by quantum annealing and the details of the machine that we use to perform the annealing. The machine has some particular engineering constraints that complicate our problem somewhat, and we discuss the techniques used to overcome these challenges. Finally we discuss the training sample and variables we propose to use to train the classifier.

In Chapter 10 we show the results of the training of the classifier on the quantum annealer. We investigate the performance of the resulting classifier and look at how long the annealer takes to solve the problem. We show that the quantum nature of the training is providing additional benefits in the form of added performance from excited state solutions to our Hamiltonian that encode additional classifiers. Finally we look toward future applications of quantum annealing to HEP and future studies to improve the performance of our classifier.

Chapter 8

Quantum Annealing

Quantum Annealing is a process for finding the ground state solution of a complicated Hamiltonian using adiabatic evolution. The idea is to construct a Hamiltonian whose ground state can be easily constructed and slowly evolve the system from the Hamiltonian of the easy to construct ground state to the Hamiltonian of the problem that we wish to solve. We will first discuss the related concept of simulated annealing, which simulates thermal annealing to develop intuition for how this problem can solve complicated systems in a fully classical regime. We will then discuss the theory behind quantum annealing in Section 8.2 and the practical details of implementing it in chapter 9.

8.1 Simulated Annealing

The problem of finding minima and maxima of various classes of functions is extremely important in a wide variety of disciplines. Finding the lowest energy state of a system is of fundamental importance in quantum physics and chemistry, and maximizing global return or utility is a hallmark principle in finance and economics. There are many different techniques for approaching this problem, from deterministic functional methods to statistical sampling methods, and there is often more than one correct method to solve a problem. One method that has been used with general success for very complicated non-convex optimization problems is simulated annealing.

Simulated annealing is especially suited to problems that have a large number

of solutions which are local minima of the function being optimized [91, 92]. Most iterative techniques have difficulty in topologies like this, since it is very hard to judge whether a minimum found is a global minimum or just a local one. Certain tricks can be used to escape from local minima, such as randomizing starting parameters or even the problem parameters, but these can dramatically increase the solution time of the problem. The common feature that these techniques share is a reliance on gradient descent for the optimization, which is very effective for convex or near convex problems, but not optimal for highly non-convex problems. Simulated annealing solves this problem by allowing locally non-optimal steps meeting certain criteria.

Simulated annealing is an iterative technique that does not rely on gradient descent, but instead uses the Metropolis algorithm to escape local minima [93]. The calculation is started by assigning a virtual temperature $T = T_0$, which is gradually decreased until it reaches $T = T_f = 0$. Let $f(\vec{\alpha})$ be the function we are seeking to minimize over its parameters $\vec{\alpha}$, we start our function in some initial state $\vec{\alpha}_0$ with energy $f(\vec{\alpha}_0)$ at temperature T_0 . We then randomly choose a small offset in the parameter of function $\Delta\vec{\alpha}$ and compute the energy of the function at this offset $f(\vec{\alpha}_0 + \Delta\vec{\alpha})$. If $f(\vec{\alpha}_0 + \Delta\vec{\alpha}) < f(\vec{\alpha}_0)$ then we accept the update $\Delta\vec{\alpha}$. If, however, $f(\vec{\alpha}_0 + \Delta\vec{\alpha}) \geq f(\vec{\alpha}_0)$ then we accept the update with probability $e^{-\Delta E_0/kT_0}$ where $\Delta E_0 = f(\vec{\alpha}_0 + \Delta\vec{\alpha}) - f(\vec{\alpha}_0)$.

This process is then repeated on the state $f(\vec{\alpha}_1)$, which is either $f(\vec{\alpha}_0)$ or $f(\vec{\alpha}_0 + \Delta\vec{\alpha})$ depending on whether the update was accepted. This algorithm is then repeated as we lower the temperature, where at each step we accept the change with probability $\max(1, e^{-\Delta E_i/kT_i})$. These thermal fluctuations allow the solution to “jump” over small barriers in the energy landscape and escape from local minima.

8.2 Quantum Annealing

Quantum annealing solves similar problems to simulated annealing using different methods. If it can be realized fully, it should be able to solve certain classically hard problems much more quickly than classical algorithms, which would make it extremely

useful in a large number of important situations. Using quantum annealing requires a collection of qubits that can be arranged easily into a fully entangled uniform superposition of their 0 and 1 states, and a system for constructing a Hamiltonian whose ground state necessarily encodes the solution to the problem. The details of the machine used to realize this system are discussed in Section 9.4 and the procedure for constructing a problem Hamiltonian that can encode the problem of training a learner to recognize the Higg boson are given in Section 9.1; here we shall assume that the hardware exists and the problem Hamiltonian can be constructed.

We have already required that it be possible to construct the time-independent Hamiltonian H_p , so we turn to constructing a second time-independent Hamiltonian: the initial state Hamiltonian H_0 . Let n be the number of qubits in the system, and let us denote the state of our system by a vector in a 2^n dimensional Hilbert space with basis vectors

$$|z_1\rangle |z_2\rangle \cdots |z_n\rangle \quad (8.1)$$

We can then define a state which is a uniform superposition of all unit vectors as

$$|\psi_0\rangle = \prod_{i=1}^N \frac{|0\rangle + |1\rangle}{\sqrt{2}} = \frac{1}{2^{n/2}} \sum_{\text{basis vectors}} |z_1\rangle |z_2\rangle \cdots |z_n\rangle \quad (8.2)$$

where the sum is over all basis vectors in the hilbert space [94]. We said as a requirement of our system that this state be easy to construct, so it must also be easy to construct the Hamiltonian that has this as a ground state, which we will call H_0 . If, for instance, we are conceptualizing our qubits as spins, then this state is achieved by applying a large traverse magnetic field.

We now construct a new time-dependent Hamiltonian that starts equal to the initial state H_0 at $t = 0$ and ends as the problem state Hamiltonian H_p at $t = t_f$. This is easy to construct as

$$H(t) = A(t)H_0 + B(t)H_p \quad (8.3)$$

subject to the constraints

$$A(0) > 0 \qquad A(t_f) = 0 \qquad (8.4)$$

$$B(0) = 0 \qquad B(t_f) > 0 \qquad (8.5)$$

We will place some additional requirements on A and B shortly. Figure 8.1 shows the actual instantiations of $A(t)$ and $B(t)$ that are used for our study. $H(t)$ now has the property that at $t = 0$ its ground state is the state given in equation 8.2 and at $t = t_f$ its ground state encodes the solution to the problem encoded by H_p .

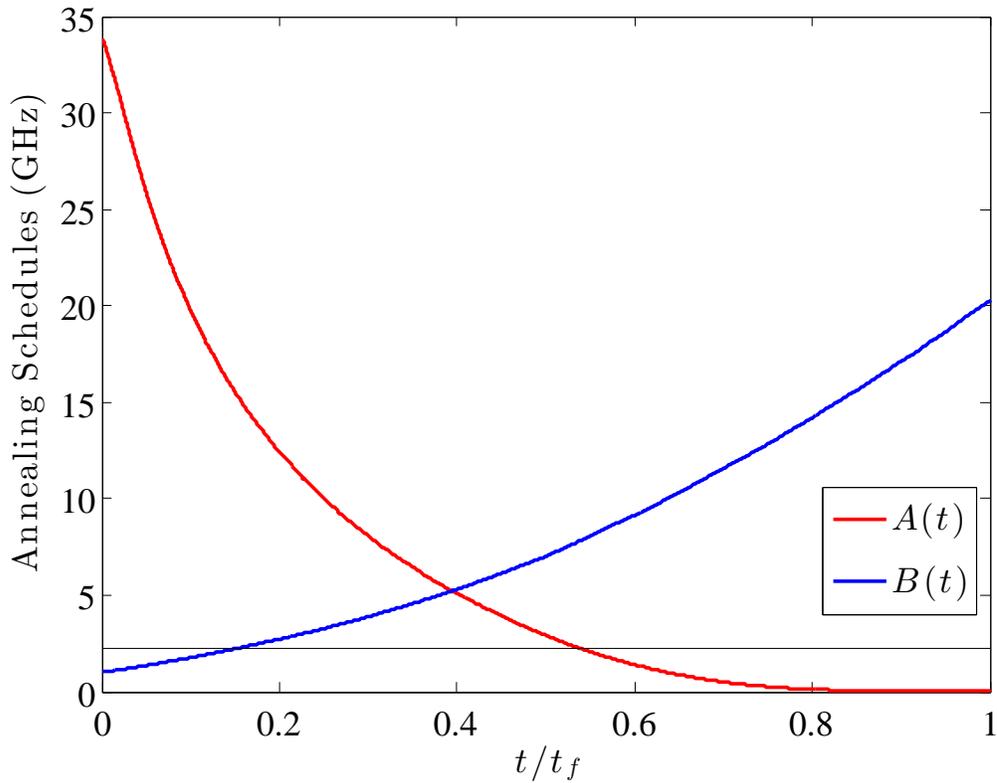


Figure 8.1: The Annealing Schedule ($A(t)$ and $B(t)$) for in the quantum annealing implementation used in our study.

We now call on the quantum adiabatic theorem, which states that if the Hamiltonian of a system transforms slowly from H^i to H^f , a particle beginning in the n^{th} eigenstate of H^i will end up in the n^{th} eigenstate of H^f [95, 96]. This means that, as long as our functions $A(t)$ and $B(t)$ qualify as “slowly” for the purpose of the

adiabatic theorem, our system that starts in the ground state of H_0 will end in the ground state of H_p and thus solve the problem we have posed.

To define what “slowly” means in the context of the adiabatic theorem, let us define $s \equiv \frac{t}{t_f}$ and write our time-dependent Hamiltonian as $H(s) = A(s)H^i + B(s)H^f$ [97]. We define $E_i(s)$ as the eigenvalues of $H(s)$ with

$$E_i(s) \leq E_j(s) \mid i < j$$

so that $E_0(s)$ is the energy of the ground state of H at time s and $E_1(s)$ is the energy of the first excited state. According to the adiabatic theorem, we can say that

$$\lim_{t_f \rightarrow \infty} |\langle \psi_0 | \psi(t_f) \rangle| = 1 \quad (8.6)$$

Following the derivation in [97] and [96], we can then define the minimum gap as

$$g_{min} = \min_{0 \leq s \leq 1} (E_1(s) - E_0(s)) \quad (8.7)$$

and we can then say that the condition under which equation 8.6 holds is that

$$t_f \gg \frac{1}{g_{min}^2} \max_{0 \leq s \leq 1} \left| \left\langle \psi_1(s) \left| \frac{dH(s)}{ds} \right| \psi_0(s) \right\rangle \right| \quad (8.8)$$

where $|\psi_1(s)\rangle$ is the first excited state of $H(s)$. This tells us that the required time to anneal the problem is governed by g_{min}^{-2} . This minimum energy difference also tells us how our solution time will scale with the size of the problem; a “hard” problem will have an exponential scaling of g_{min} with the problem size, and hence an exponential scaling of the required running time, while an “easy” problem would scale polynomially.

So far this discussion has focused on a perfectly isolated system operating at 0 temperature, so that the effects of thermal fluctuations and interaction between the system and the environment causing decoherence could be ignored. These effects can be thought of as adding additional terms to the Hamiltonian to characterize coupling

to a thermal bath and to the environment $H_{tot} = H_{anneal} + H_{bath} + H_{int}$, where H_{anneal} is the (time-dependent) annealing hamiltonian, H_{bath} is the coupling of the system to the thermal bath, and H_{int} is the coupling of the system to the environment [98–100]. The dynamics of this total Hamiltonian can be analyzed using Markovian master equations that describe the time evolution of an adiabatic system weakly coupled to a thermal bath [101]. One key insight from this analysis is that the decoherence drives the annealer to the Gibbs state of H_{anneal} at the equilibrium temperature of the system [100].

In this Gibbs state regime, the annealing system becomes inherently probabilistic: the annealer will find the ground state with some probability related to the distribution of the Gibbs state of the problem Hamiltonian. If $\rho(t)$ is the time-dependent density matrix of the thermally coupled system, then we can say that the probability of ending up in the ground state will be close to

$$\langle \psi_0 | \rho(t_f) | \psi_0 \rangle \xrightarrow{t_f \rightarrow \infty} \langle \psi_0 | \frac{e^{-\beta H_P}}{\text{Tr } e^{-\beta H_P}} | \psi_0 \rangle \quad (8.9)$$

in the infinite time limit [100]. In the non-infinite time limit, the success probability depends on the nature of the spectral density of the bath and the form of the coupling to the environment [98]. For example, if the spectral density of the bath is Ohmic ($S(\omega) \propto \omega \frac{e^{-\omega/\omega_c}}{1-e^{-\beta\omega}}$, where ω_c is some cutoff frequency) then we find that the probability to end up in the ground state goes like

$$p_{GS} \propto 1 - e^{-\varepsilon/\alpha} \quad (8.10)$$

$$\varepsilon = \left| \left\langle \psi_1(s) \left| \frac{dH(s)}{ds} \right| \psi_0(s) \right\rangle \right| \quad (8.11)$$

where α is a constant describing the energy scale of the interaction. For small problem sizes, the master equations can be solved numerically to estimate the ground state probability as a function of t_f (see [100]), but for larger problems, this swiftly becomes intractable. It is typically easier to *measure* the solution probability as a function of t_f and/or bath temperature and use that to gain insight about g_{min} and the dynamics

of the decoherence of the system [102].

Chapter 9

Experimental Setup

Our goal is to create a classifier that can be trained using a quantum annealer to classify events with a real Higgs Boson and separate them from background. The classifier takes in information about the event and returns a score reflecting how likely it is that the object is signal. There are three key components to setting up this system:

1. how to cast the problem of training this classifier in a way that can be encoded on a quantum computer
2. how the quantum annealer is realized in hardware
3. how to setup the training sample in such a way that it can be used effectively

We will go through the steps of the problem in this order and then come in chapter 10 to results of the study.

9.1 Encoding the Problem

We first address the problem of training a classifier to solve a generic machine learning problem where we have a training sample containing information about each event and whether that event is signal or background. We will do this by turning the problem of training such a classifier into an optimization problem, which is set up such that finding the minimum energy solution is equivalent to finding the best classifier;

such problems can then easily be mapped to Hamiltonians and solved by quantum annealing [103].

Let $\mathcal{T} = \{(\vec{x}_i, y_i)\}$ be the training data, where \vec{x}_i is some set of variables about event i and y_i is the indicator variable

$$y_i = \begin{cases} +1 & \text{if event } i \text{ is signal} \\ -1 & \text{if event } i \text{ is background} \end{cases}$$

we define a *weak classifier* $z(\vec{x}_i)$ as any function that has the property that $z(\vec{x}) > 0$ increases the probability that \vec{x} is a signal event or, more formally, $p(y_i = +1 | z(\vec{x}_i) > 0) > p(y_i = +1)$ and $-1 \leq z(\vec{x}) \leq 1 \forall \vec{x}$. If the probability of an event being signal or background are equal before any cuts, than the weak classifier condition reduces to saying that and event \vec{x} is more likely to be signal than background if $z(\vec{x}) > 0$. So a weak classifier is any function which separates signal signal from background, even if the separation is extremely small.

We can find many functions over the training data that satisfy the definition of a weak classifiers, so we can define a set of weak classifiers $\{z_i(\vec{x})\}$ for $1 \leq i \leq N$. In a typical problem, we might be able to identify a huge number of weak classifiers, so we would like a systematic way to select the best ones to form a stronger classifier. Selecting the subset of weak classifiers the produce the best strong classifier is a difficult problem, and is it this problem that we will turn into an optimization problem and solve. We form a new classifier $O(\vec{x})$ as the weighted sum of weak classifiers:

$$O(\vec{x}; \vec{w}) = \frac{1}{N} \sum_{i=1}^N w_i z_i(\vec{x}) \quad (9.1)$$

where $w_i \in \{0, 1\}$ and $\vec{w} \equiv (w_0, w_1, \dots, w_N) \in \{0, 1\}^N$ are the to-be-determined weights. Since the weights are binary, any instance of this classifier is selecting some subset of the weak classifiers to evaluate. In principle we could say that $O(\vec{x}; \vec{w}) > 0$ indicates signal and $O(\vec{x}; \vec{w}) < 0$ indicates background, but in practice it will be useful to think of $O(\vec{x}; \vec{w})$ as a continuous output with more positive values being

more signal like and more negative values being more background-like.

We begin by computing the error from the output of O , which is asking whether $O(\vec{x}; \vec{w})$ classifies the event \vec{x} in the training sample correctly. We can then define a function that indicates whether the classification of event i was successful:

$$L(\vec{x}_i; \vec{w}) = \frac{1}{2} (y_i - \text{sign}[O(\vec{x}_i)])$$

This function is interesting, but as mentioned before we prefer the continuous classifier, since it is more useful and easier to train. We therefore define a per-event error as

$$\delta(\vec{x}_i; \vec{w}) = (y_i - O(\vec{x}_i))^2 \quad (9.2)$$

This function is continuous, so it is useful for optimization, but it has a slight technical problem in that it prefers large numbers of weights being 1, since $O(\vec{x}) \in [-\frac{\|\vec{w}\|}{N}, \frac{\|\vec{w}\|}{N}]$. This means that even a perfect classifier would have $\delta(\vec{x}; \vec{w}) = (1 - \frac{\|\vec{w}\|}{N})^2$. For a typical problem, we expect $\|\vec{w}\| \ll N$, so we can approximate the error as $(1 - 2\frac{\|\vec{w}\|}{N})$ and add this as a correction to the error function. We will proceed without this correction for the time being and add it at the final step in a slightly rearranged form.

We can now compute the total error:

$$\delta(\vec{w}) = \sum_{i=1}^{N_{\mathcal{T}}} \delta(\vec{x}_i; \vec{w}) = \sum_{i=1}^{N_{\mathcal{T}}} (y_i - O(\vec{x}_i))^2 \quad (9.3)$$

$$= \sum_{i=1}^{N_{\mathcal{T}}} y_i^2 + \sum_{i=1}^{N_{\mathcal{T}}} O(\vec{x}_i)^2 - 2 \sum_{i=1}^{N_{\mathcal{T}}} O(\vec{x}_i) y_i \quad (9.4)$$

$$= \sum_{i=1}^{N_{\mathcal{T}}} y_i^2 + \sum_{i=1}^{N_{\mathcal{T}}} \sum_{j=1}^N \sum_{k=1}^N w_j w_k z_j(\vec{x}_i) z_k(\vec{x}_i) - 2 \sum_{i=1}^{N_{\mathcal{T}}} \sum_{j=1}^N y_i w_j z_j(\vec{x}_i) \quad (9.5)$$

where $N_{\mathcal{T}}$ is the size of the training sample, and we have substituted the expression for $O(\vec{x})$ in the third line. The first term is simply equal to the size of the training sample and may be ignored. We rewrite the remaining terms in a more suggestive

fashion

$$\delta(\vec{w}) \propto \sum_{j=1}^N \sum_{k=1}^N w_j w_k \left(\sum_{i=1}^{N_{\mathcal{T}}} c_j(\vec{x}_i) c_k(\vec{x}_i) \right) - 2 \sum_{j=1}^N w_j \left(\sum_{i=1}^{N_{\mathcal{T}}} y_i c_j(\vec{x}_i) \right)$$

The terms in parentheses are not dependent on the weights, so we can then define

$$C_{jk} \equiv \sum_{i=1}^{N_{\mathcal{T}}} c_j(\vec{x}_i) c_k(\vec{x}_i) \quad (9.6)$$

$$C_j^y \equiv \sum_{i=1}^{N_{\mathcal{T}}} y_i c_j(\vec{x}_i) \quad (9.7)$$

$$(9.8)$$

and we are left with

$$\delta(\vec{w}) \propto \sum_{j=1}^N \sum_{k=1}^N C_{jk} w_j w_k - 2 \sum_{j=1}^N C_j^y w_j \quad (9.9)$$

To equation 9.9, we add the correction from equation 9.2. However, it is not always the case that one wants to exactly cancel the error, since minimizing over the corrected equation could still cause overtraining, where the network picks up statistical features of the training sample. This tends to happen when the network turns on classifiers with only small effect on the total efficiency of the network in the training sample. To combat this, we introduce a term $\lambda \|\vec{w}\| = \lambda \sum_{j=1}^N w_j$, where λ is a tunable parameter. With this term added, our problem of training the network is now equivalent to finding the \vec{w} that minimizes

$$\sum_{j=1}^N \sum_{k=1}^N C_{jk} w_j w_k + \sum_{j=1}^N (\lambda - 2C_j^y) w_j \quad (9.10)$$

For encoding on the quantum annealer, it is more convenient to write this equation in terms of variables $s_i \in \{-1, 1\}$ instead of $w_i \in \{0, 1\}$. This is straightforward by

assigning $w_i = \frac{1}{2}(s_i + 1)$; plugging this into equation 9.10 gives

$$\frac{1}{4} \sum_{j=1}^N \sum_{k=1}^N C_{jk}(s_j + 1)(s_k + 1) + \frac{1}{2} \sum_{j=1}^N (\lambda - 2C_j^y)(s_j + 1) \quad (9.11)$$

If we assign

$$C'_{jk} = \frac{1}{4} C_{jk} \quad (9.12)$$

$$C_j^{y'} = C_j^y - \frac{1}{2} \sum_{k=1}^N C_{jk} \quad (9.13)$$

and drop the constant term $\frac{1}{4} \sum_{j=1}^N \sum_{k=1}^N C_{jk}$ and set $\lambda' = \frac{1}{2}\lambda$, then equation 9.14 becomes

$$\sum_{j=1}^N \sum_{k=1}^N C'_{jk} s_j s_k + \sum_{j=1}^N (\lambda' - C_j^{y'}) s_j \quad (9.14)$$

9.2 Embedding the Problem on the Chimera Graph

To understand how this problem can be embedded on a quantum annealer, it is useful to think of equation 9.14 as describing the parameters of a fully connected graph with N binary nodes (K_N) (shown for $N=6$ in figure 9.1) where each node n_j has some local field B_j and the coupling between nodes i and j has strength B_{jk} . One can then map the value of the weight parameter s_j to the node n_j , the entries of $C_j^{y'}$ can be thought of as local fields B_j and the entries of C'_{jk} map to B_{jk} . If one defines the energy of the node as $E_j = s_j * B_j + \sum_{k=1}^N B_{jk} s_j s_k$ and the energy of the graph as $E_{graph} = \sum_{j=1}^N s_j$, then the problem of minimizing equation 9.14 is identical to finding the lowest energy configuration of the graph for $n_j \in \{-1, 1\}$.

Casting the problem in this manner is useful, because it allows us to more easily visualize what happens when we encode the problem onto the D-Wave. Because of engineering constraints, the D-Wave quantum annealer does not provide couplings between every pair of qubits; instead it implements connections between a qubit and

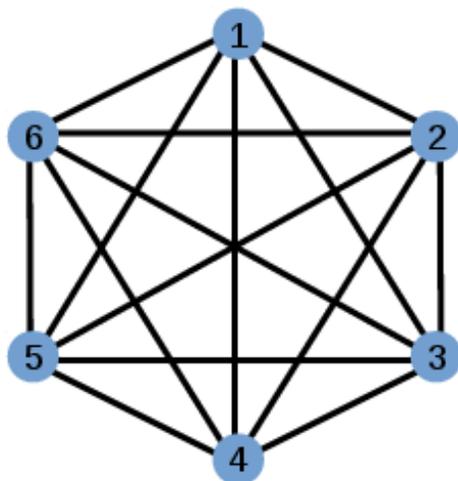


Figure 9.1: A fully connected graph with 6 nodes

at most six nearest neighbors [103–108]. One can interpret the allowed connects between qubits as a graph, where each qubit is a node and each edge is a valid connection between a qubit and a neighbor. In this interpretation the hardware implementation of the D-Wave, known as the *chimera graph*, can be visualized as the graph shown in figure 9.2. It consists of 8 qubit unit cells ($K_{4,4}$) each glued together by joining qubits on the right hand side of the $K_{4,4}$ horizontally and those on the left hand side vertically. In all, the hardware graph consists of an 8×8 array of these unit cells, totalling 512 qubits.

Clearly, our problem encoded in equation 9.14 that maps to the K_N graph in figure 9.1 will not map directly to the chimera graph unless most of the entries of C'_{jk} are 0. Since this will not generally be the case, we use a method of embedding the fully connected graph within a larger sparsely connected graph in such a way that they encode the same problem. In the following we will refer to the graph K_N as the *logical graph* and its embedding into the chimera graph as the *physical graph*. Similarly, a node in the logical graph is a *logical node* while a node in the physical graph is a *physical node*.

The key element in embedding the logical graph into the physical graph is the mapping of logical nodes to chains of physical nodes. A chain of nodes, for this

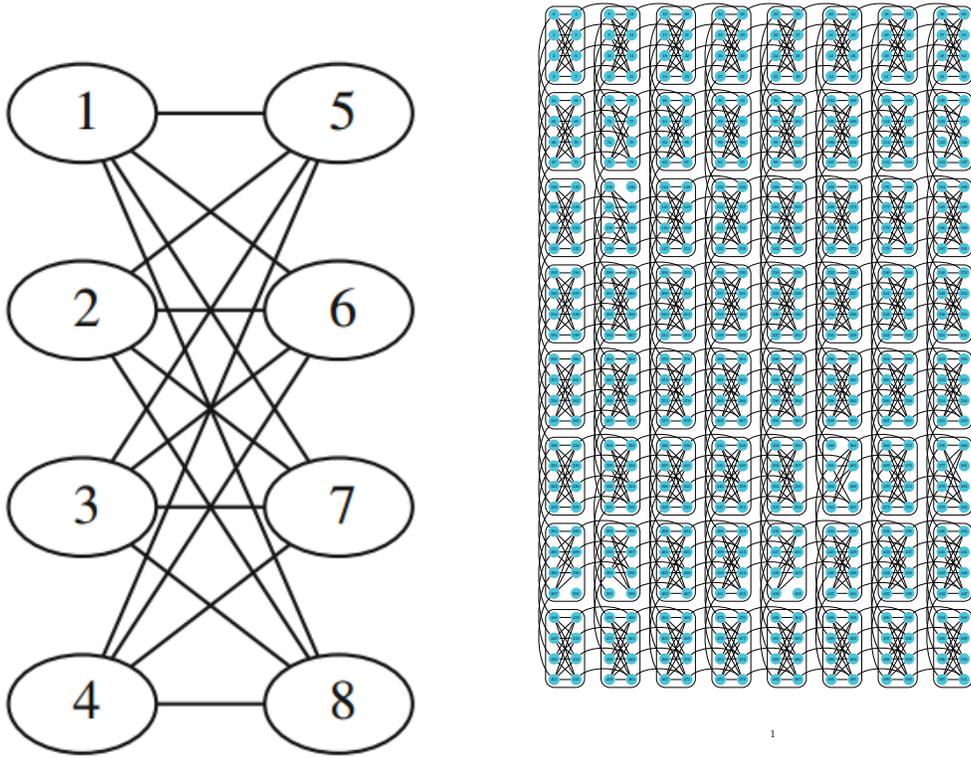


Figure 9.2: The unit cell ($K_4, 4$) of the chimera graph (left) and the arrangement of the unit cells to create the full hardware graph (right).

purpose, is a set of nodes such that there is a path through the graph that touches each node in the chain and no nodes outside the chain. More formally, a set of nodes within a graph \mathcal{G} is a chain if it forms a connected subgraph of \mathcal{G} . Let $\mathcal{C}_p = \{n_i^p | 1 \leq i \leq M\}$ be a chain in the physical graph and $\{e_{i,i+1}^p | 1 \leq i \leq M-1\}$ be a set of edges such that $e_{i,i+1}^p$ connects nodes n_i^p and n_{i+1}^p . If we then set the coupling strength of the edge $e_{i,i+1}^p$ to $E_{i,i+1}^p = -|E_{chain}|$, we introduce terms in the network energy that look like $\sum_{i=1}^{N-1} E_{i,i+1}^p n_i^p n_{i+1}^p = -|E_{chain}| \sum_{i=1}^{N-1} n_i^p n_{i+1}^p$, which is obviously minimized if $n_i^p = n_{i+1}^p$ (i.e. if all elements in the chain point in the same direction). If we take E_{chain} much larger than any other field in the problem, then this term would outweigh any other possible term in the energy so the solution would always keep the bits in the chain aligned.

Using this concept, we map each logical node in the problem to a chain of nodes

in the physical problem. If we solve the problem in the physical graph, then we can readout the solution in the logical graph by taking the value of the nodes within each chain and setting the the corresponding logical node to that value. We have two criteria for the chain mapping

- all chains must be disjoint
- For each pair of logical nodes n_α^l and n_β^l ($\alpha \neq \beta$), mapped to physical chains $\{n_i^{p,\alpha} | 1 \leq i \leq N_\alpha\}$ and $\{n_i^{p,\beta} | 1 \leq i \leq N_\beta\}$ respectively, there must be an edge in the chimera graph connecting $n_i^{p,\alpha}$ to $n_j^{p,\beta}$ for some $1 \leq i \leq N_\alpha$ and $1 \leq j \leq N_\beta$.

The first point is just saying that chains cannot cross each other. The second point says that each chain must be adjacent to every other chain in at least one place (two chains \mathcal{C}_i and \mathcal{C}_j are adjacent if there exists at least one edge of the physical graph connect connecting nodes n and m with $n \in \mathcal{C}_i$ and $m \in \mathcal{C}_j$).

With these definitions we can write the exact definition of the implementation of the problem on the physical problem given a logical problem and a mapping from each logical node to a chain of physical nodes (called an embedding). Let $\{n_i^l | 1 \leq i \leq N_{log}\}$ be the list of logical nodes and $\{n_j^p | 1 \leq j \leq N_{phys}\}$ be the list of physical nodes. For each n_i^l define $\mathcal{C}_i \subset \{n_j^p\}$ such that $\{n^p | n^p \in \mathcal{C}_i\}$ is the chain corresponding to logical node i . We can then set the coupling strength on each node n_j^p in the physical problem to

$$E_j^p = \begin{cases} \frac{E_i^l}{|\mathcal{C}_i|} & \text{if } \exists i \mid n_j^p \in \mathcal{C}_i \\ 0 & \text{if } \nexists i \mid n_j^p \in \mathcal{C}_i \end{cases}$$

The disjointness of the chains implies that if i exists, it is unique. This is essentially saying that the energies from the logical problem are evenly divided amongst the nodes in the corresponding logical chain.

We have already described what happens to edges within a given chain, so all that remains is to set the edges between chains. Let $\mathcal{E}^l = \{(i, j)^l\}$ be the set of edges in the logical problem (i.e. the edges of the fully connected graph) and let $\mathcal{E}^p = \{(l, m)^p\}$ be the set of edges in the physical graph (i.e. the chimera graph). Let $E_{i,j}^l$ be the strength

of the $(i, j)^l$ coupling, and let $\mathcal{E}_{i,j} = \{(l, m)^p \mid (l, m)^p \in \mathcal{E}^p \ \& \ n_l^p \in \mathcal{C}_i \ \& \ n_m^p \in \mathcal{C}_j\}$ we can then set

$$E_{l,m}^p = \begin{cases} -E_{chain}^2 & \text{if } \exists i \mid n_l^p, n_m^p \in \mathcal{C}_i \\ \frac{E_{i,j}^l}{|\mathcal{E}_{i,j}|} & \text{if } \exists i, j \mid (l, m)^p \in \mathcal{E}_{i,j} \\ 0 & \text{else} \end{cases}$$

The first case occurs when the nodes are within a chain, the second case divides the coupling between logical nodes over all the valid couplings between the chains, and the third term sets all the rest of the couplings to 0. Again, the disjointness of the chains assures that the cases are mutually exclusive and that if $\exists i, j$, they are unique.

We will require many more physical nodes than logical nodes to embed the problem. We measured the median number of required physical nodes as a function of the problem size in figure 9.3 with the median taken over 200 random problems of each logical size. The scaling with logical problem size is found to be roughly quadratic. This comes from the linear scaling of the mean chain length shown on the right plot. This scaling predicts that, on 512 physical nodes, we should be able to embed a logical problem of up to 36 variables. A problem of this size, however, can only reliably be embedded around half the time (since its a median), so to get a true upper limit on the logical problem size one should look at the maximum physical problem size over the 200 random problems, which is shown in figure 9.4. From this we conclude that we can reliably embed problems with up to 32 variables. It has been shown by Choi that the tree width of the 512 node chimera graph is 33, meaning that it can admit an embedding of up to a theoretical maximum of 33 logical nodes, so this result is in line with the prediction [105, 109].

9.3 Practical Implementation of the Embedding

For consistency with the literature, we will adopt the traditional notation convention for the fields being applied to the physical problem. We will use h_i to mean the local field on the i^{th} node of the physical graph and J_{ij} to refer to the coupling between

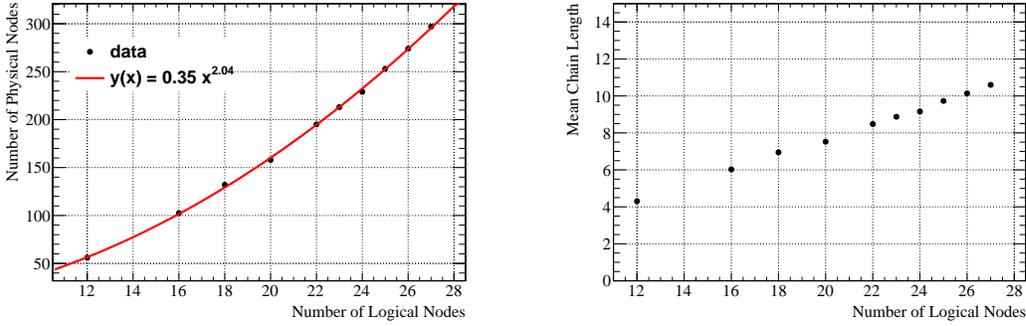


Figure 9.3: The median number of physical nodes required to embed the logical network (left) and the average length of the chains as a function of the chains (right) as a function of the number of nodes in the logical problem.

nodes i and j of the physical problem.

In the actual implementation of the hardware, there are some additional complications to consider related to the physical limitations of the system. The first is noise on the chip which leads to an uncertainty on the value of the coupling actually programmed onto any edge or node, compared with the intended value. Each individual qubit has very slightly different physical properties, which are compensated for by the design of the qubits (discussed in section 9.4), but this leads to slightly different dynamic ranges for the magnetic fields that implement the biases and couplings (typically on order ≈ 10 picohenries). The architecture is designed so that the dynamic range for the local biases (h_i) is twice that for the inter-qubit couplers. For ease of discussion, we will use units in which $-1 \leq J_{ij} \leq 1$ and $-2 \leq h_i \leq 2$. In these units, it has been found that the actual value of the magnetic field implemented on the chip has an uncertainty of ± 0.05 . Since our problem is invariant under uniform scaling of the Hamiltonian, this noise model suggests that the optimal procedure is to scale all the couplers up to the maximum value allowed by the dynamic range (i.e. so that the largest coupler in the physical problem has $\max(\frac{|h_i|}{2}, |J_{ij}|) = 1$ in these units).

A second complication arises from the possibility that a chain is *broken*, which is when the qubits within a chain are not all aligned. This creates a difficulty in converting the physical problem back into the logical problem after optimization. There are ways to convert broken chains, which will be discussed in the next chapter,

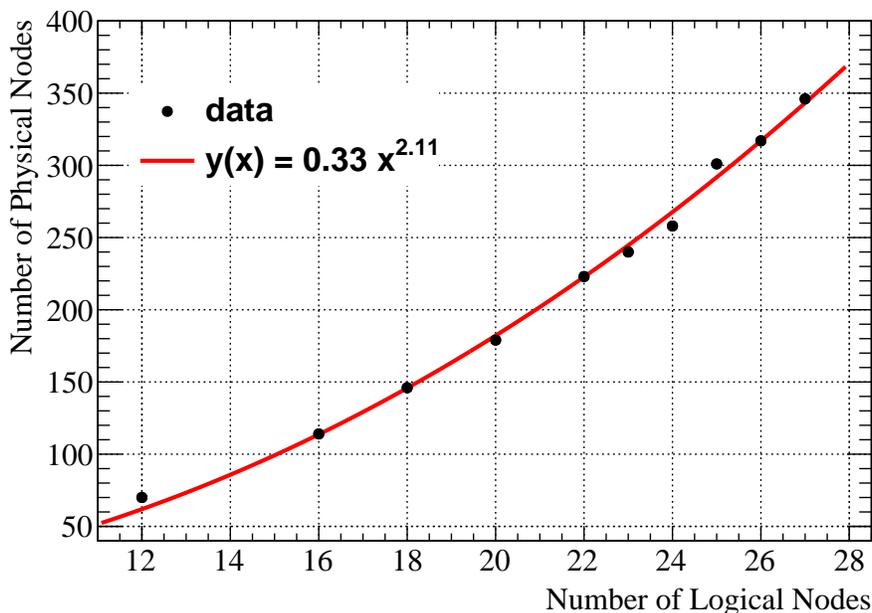


Figure 9.4: The max number of physical nodes required to embed the logical network (left) and the average length of the chains as a function of the chains (right) as a function of the number of nodes in the logical problem.

but the best way to deal with them would be to prevent them from happening.

We said in the last section that the ideal way to prevent chains from breaking would be to set the intra-chain coupling constant larger than any energy scale in the problem to create huge energy penalties for breaking. The actual dynamics are somewhat more subtle than this, because we still need the chains to be able to flip when the correlation terms make it energetically favorable. Since correlation terms touch only one or two qubit(s) in the chain, as the anneal progresses only a subset of the qubits will see the energy penalty for (mis)alignment with a neighbor. What needs to happen, then, is that if that energy penalty is sufficiently strong, it should flip the qubits directly coupled, which should then induce the entire chain to flip. If there energy coupling within the chain completely overwhelms all of the couplings between adjacent chains, then this cannot happen and the correlation terms will have no effect on the final solution.

If is difficult *a-priori* to solve this problem, so we instead set all intra-chain couplings to -1 and tune the rest of the problem. Since the intra-chain Hamiltonian is not

a part of the logical problem, we can arbitrarily rescale it with respect to the rest of the problem. We have done studies measuring the optimal strength of the couplers with respect to the problem and found that best performance is typically achieved when the problem scale is slightly less than the intra-chain scale $\max(|J_{i,j}^{inter-chain}|) \lesssim 1$. This gives allows the flipping behavior required for the annealing to be useful, at the expense of chains occasionally *breaking* or not all pointing in the same direction. A method of dealing with broken chains is discussed in section 10.1.1.

9.4 The D-Wave 2TM Machine

The D-Wave 2TM is a machine built by D-Wave Systems Inc. that is designed to implement a 512 qubit quantum annealer connected like the chimera graph of Figure 9.2. There is substantial ongoing research into whether the machine is a true quantum annealer and whether it is able to use quantum correlations in a meaningful way to solve problems [102, 108, 110–118]. One goal of our study is to see whether this machine can solve large, physically motivated problems; we do not directly seek to answer the question of whether this is a true quantum annealer, but we do look at some metric of its performance versus other algorithms.

The qubits in the D-Wave 2TM are built from superconducting Josephson Junctions (JJ), which are two superconducting materials separated by a thin layer of insulator or non-superconducting material [119, 120]. The key insight for this arrangement is that the supercurrent flowing across a JJ will be proportional to the sine of the phase difference across the junction. The basic building block to the D-Wave 2TM qubits are loops of superconducting wire with one Josephson junction and an inductor called *flux qubits*. This creates an arrangement where an external applied magnetic field will produce a persistent current around the loop. The flux through this loop is quantized and, if an external magnetic field with a magnitude of half the flux quantum is applied, it creates a Hamiltonian with two symmetric degenerate minima separated by a tunable energy barrier.

These degenerate minima correspond to macrostates with all current circulating

clockwise or counter-clockwise around the flux qubit. Figure 9.5 shows the shape of the potential and the configurations of the ground and first excited states. The energy levels of the ground state and the first excited state form the energy eigenbasis for the qubit. An alternative basis representing the current or flux in the loop can be formed by taking symmetric and antisymmetric combinations of the ground state and the first excited states. The basis states in this representation correspond to clockwise and counter-clockwise currents. Having these two alternate basis representations is the feature that makes the flux qubit useful: the direction of current flow can be mapped to 0 and 1 and can be easily read out at the end of a computation and, the qubit can be placed in a uniform superposition of 0 and 1 by going to the ground state in the energy basis.

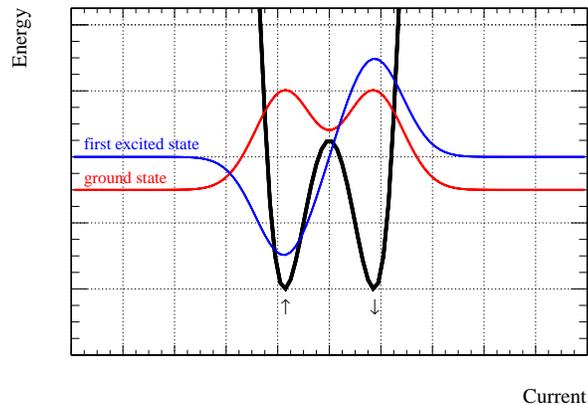


Figure 9.5: The configuration of the potential (black line) when the applied flux is set to one half flux quantum. The red and blue lines show the symmetric ground state and the antisymmetric first excited state.

The full configuration used in the D-Wave 2TM machine is described thoroughly in [121]. It adds more loops of JJs to each flux qubit to give additional tunable parameters for calibration and to allow the application of two distinct fields: one to set up in the energy basis and the other to read out in the current basis. Coupling between the qubits is achieved by adding additional loops that allow for mutual inductance between adjacent qubits.

The qubits are fabricated on a silicon chip with Nb wires forming the loops and JJs. The chip is cooled and operated at a temperature of 20 mK to reduce thermal

noise (far below the critical temperature of Nb, which is 9 K). The annealing schedule is fixed by the hardware design, with the only tunable parameter being the anneal time ($T_A > 20\mu s$). The system provides an API for programming field strengths onto the chip, so the exact details are abstracted from the user. One sets a magnetic field of a coupler or local field using the units described in section 9.3 where $J_{ij} \in [-1, 1]$ and $h_i \in [-2, 2]$ and the API takes care of converting this into the actual bias field that needs to be applied to any given loop. The API also reads the current values off the chip at the end of the anneal time, returning a list of bits $q_i \in \{-1, 1\}$ representing the measured current direction of each qubit.

A variety of sources of noise can spoil the evolution of the machine, the most important of which is thermal noise. Given the size of the typical inductances (pH) and current (μA) encoded on the device, typical energy level separations will be $\approx 10^{-5}$ eV, while 20 mK gives a thermal energy scale of 2×10^{-6} eV. This means it is relatively easy for thermal fluctuations to cause transitions between energy levels, even when the system evolution is slow enough that the adiabatic theorem applies. Furthermore, we have little control over what happens to the separation of the energy levels during the annealing process, and as discussed in section 8.2 the difficulty of the problem is controlled by g_{min}^{-2} , where g_{min} is the minimum gap between the ground state and the first excited state at any point in the annealing process.

Since the D-Wave 2TM machine operates at finite temperature, we will often be in the regime where kT is of order g_{min} , and we will have some decoherence from interactions of the system with its environment, we will invariably end up in a probabilistic regime. As discussed, this means that the final state of the annealer will be sampled from some probability distribution that will depend on the dynamics of the interaction between the system and the bath (e.g. equation 8.9). In particular, it will end up in the ground state with some probability related to the dynamics of the annealing problem (e.g. equation 8.10 if the noise distribution is ohmic and the anneal time fulfils the conditions of equation 8.8). To ensure that the ground state is found, we can run the annealer multiple times on the same problem and take the lowest energy solution we find. If each run of the annealer has a probability p_g to end

up in the ground state, then the probability that we find the ground state at least once after R runs is

$$P(R; p_g) = 1 - (1 - p_g)^R$$

Using this we can ensure that we find the ground state at least once with confidence C

$$C \leq 1 - (1 - p_g)^R \tag{9.15}$$

$$R \log(1 - p_g) \leq \log(1 - C) \tag{9.16}$$

$$R \geq \frac{\log(1 - P(R; p_g))}{\log(1 - p_g)} \tag{9.17}$$

We can, therefore, find the ground state with probability C by running the annealer at least $\frac{\log(1 - P(R; p_g))}{\log(1 - p_g)}$ times. Since the value of p_g is usually extremely difficult to calculate, we will usually measure p_g on small problems where the ground state can be found exactly by other means and then extrapolate to an estimate of p_g for larger problems.

9.5 Training Sample

We seek to train the network to recognize events which have a $H \rightarrow \gamma\gamma$ decay and distinguish it from events which have a di-photon system with similar invariant mass, but arising from standard model processes. We generate events at $\sqrt{s} = 8$ TeV using PYTHIA 6.4 [13] for the signal $H \rightarrow \gamma\gamma$ events with $m_H = 125$ GeV produced through gluon fusion and SHERPA [122] for the SM background events. We select events that have two photons with $|\eta| < 2.5$ with one photon having $p_T > 30$ GeV and the other having $p_T > 22$ GeV to simulate realistic detector acceptance and trigger requirements. We also require that all events have di-photon invariant mass $122.5 < m_{\gamma\gamma} < 127.5$, which ensures that we are looking only at the background directly underneath the Higgs boson mass peak, and that the classifier cannot learn to cut based on mass information (since this information is typically used to perform a

fit). The major diagrams contributing to the training sample are shown in Figure 9.6.

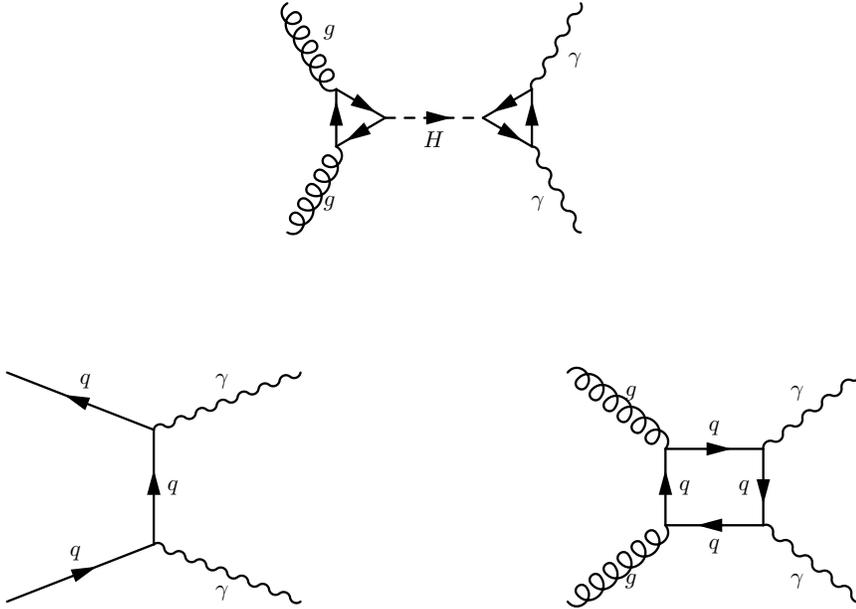


Figure 9.6: Diagrams generated for the training sample. The signal is gluon-fusion higgs to two photons (top) and the background is standard model production of two real photons (bottom).

We use the generator information about the photon kinematics as the variables for the classifier. We identify 8 variables of interest about the problem, which are listed in Table 9.1. These variables encode different information about the events that we expect to differ slightly if there is a heavy particle producing the diphoton pair versus production from other processes. Figure 9.7 shows the distributions of the variables for the signal and background samples. One can see that all have slightly different shapes in the different samples, though some are quite similar. These variables were selected because they represent a large variety of different information about the momenta and direction of the photons. The correlations of the photons from the decay of the Higgs boson compared to the combinatorial background will typically produce differences in variables where differences or separations of the photons are considered.

variable	description
$p_T^1/m_{\gamma\gamma}$	transverse momentum of the highest p_T photon divided by the invariant mass of the diphoton pair
$p_T^2/m_{\gamma\gamma}$	transverse momentum of the second-highest p_T photon divided by the invariant mass of the diphoton pair
$(p_T^1 + p_T^2)/m_{\gamma\gamma}$	sum of the transverse momentum of the two photons divided by their invariant mass
$(p_T^1 - p_T^2)/m_{\gamma\gamma}$	difference of the transverse momentum of the two photons divided by their invariant mass
$p_T^{\gamma\gamma}/m_{\gamma\gamma}$	transverse momentum of the diphoton system divided by the invariant mass
$\Delta\eta$	separation in the η direction of the two photons
ΔR	sum in quadrature of the separation in η and ϕ directions of the two photons ($\sqrt{\Delta\eta^2 + \Delta\phi^2}$)
$ \eta^{\gamma\gamma} $	the η value of the diphoton system

Table 9.1: The kinematic variables identified for use as inputs to the learner.

In order to use the treatment of the classifier discussed in section 9.1, we must transform our kinematic variables into weak classifiers. A weak classifier should have the property that $p(\vec{x} \in S | z(\vec{x}) > 0) > p(\vec{x} \in S)$ and that $z \in [-1, 1]$, which none of our variables satisfy in their out-of-the-box form. Fortunately, it is easy to transform a variable v into a weak classifier h using a standard procedure. Intuitively, what we want to do is find a cut in the distribution where the S/B on one side of the cut is higher than the S/B for the entire distribution. We then shift the distribution so that this cut point is at 0 and then reflect about 0 if needed so that the higher S/B region occurs for values greater than 0. Finally we rescale the distribution separately above and below 0 so that it falls in the range $[-1, 1]$. Actually, we find that it is better to truncate the distribution on either side so that the ranges are not driven by the (low statistics) tails of the distributions, so we identify some maximum point in the shifted variable and assign all values greater than this point to the classifier value +1 and some minimum point assigned to -1.

To make this discussion more precise, we detail the exact algorithm used to convert a variable v into a weak classifier z . Using the notation that $S(v)$ is the binned pdf of the signal events over the variable v and $B(v)$ is the binned background pdf, we define several values and functions for each variable:

1. Find the value v_0^{low} such that $\int_{-\infty}^{v_0^{low}} S(v')dv' = 0.7$ and v_0^{up} such that $\int_{v_0^{up}}^{\infty} S(v')dv' = 0.7$. v_0^{low} is now the point in the distribution such that 70% of signal events have $v < v_0^{low}$, while 70% of signal events have $v > v_0^{up}$.
2. Compute $b^{low} = \int_{-\infty}^{v_0^{low}} B(v')dv'$ and $b^{up} = \int_{v_0^{up}}^{\infty} B(v')dv'$. b^{low} now tells us how many background events have $v < v_0^{low}$ and b^{up} tells us how many background events have $v > v_0^{up}$.
3. If $b^{low} < 0.7$ let $v^{shift}(v) = v_0^{low} - v$, otherwise if $b^{up} < 0.7$ let $v^{shift}(v) = v - v_0^{up}$, otherwise reject the classifier. If $b^{low} < 0.7$, it tells us that S/B improves by cutting on $v < v_0^{low}$. If it doesn't, then we check if $b^{high} < 0.7$, in which case cutting at $v > v_0^{up}$ improves S/B. If neither of these is the case, then the distribution doesn't lend itself to being cast as a weak classifier in a simple way. For this study we reject these variables, but one could imagine going in and defining these by hand.
4. Find v_{+1} such that $\int_{v_{+1}}^{\infty} S[(v^{shift})^{-1}(x)] dx = 0.1$ and v_{-1} such that $\int_{-\infty}^{v_{-1}} B[(v^{shift})^{-1}(x)] dx = 0.1$. $v_{\pm 1}$ are the truncation points for the weak classifier, if $v^{shift}(v) \notin [v_{-1}, v_{+1}]$, then we will assign the output to +1 or -1 depending on whether it is above or below the range.

With these definitions we define the classifier:

$$z(v) = \begin{cases} +1 & \text{if } v_{+1} < v^{shift}(v) \\ \frac{v^{shift}(v)}{v_{+1}} & \text{if } 0 < v^{shift}(v) \leq v_{+1} \\ \frac{v^{shift}(v)}{|v_{-1}|} & \text{if } v_{-1} < v^{shift}(v) \leq 0 \\ -1 & \text{if } v^{shift}(v) < v_{-1} \end{cases} \quad (9.18)$$

This has, by construction, the properties we require from a weak classifier. This can be seen by noting that $\int_0^{\infty} S[(v^{shift})^{-1}(v')] dv' = 0.7$ and $\int_0^{\infty} B[(v^{shift})^{-1}(v')] dv' < 0.7$, so we can see that the probability of an event being a signal event given that

$v^{shift}(v) > 0 \implies z(v) > 0$ is

$$p(\vec{x} \in S \mid v^{shift}(v_x) > 0) = \frac{N_S \left(\int_0^\infty S [(v^{shift})^{-1}(v')] dv' \right)}{N_S \left(\int_0^\infty S [(v^{shift})^{-1}(v')] dv' \right) + N_B \left(\int_0^\infty B [(v^{shift})^{-1}(v')] dv' \right)}$$

$$> \frac{0.7N_S}{0.7(N_S + N_B)} = \frac{N_S}{N_S + N_B} = p(\vec{x} \in S)$$

Where N_S is the number of signal events in the training sample and N_B is the number of background events.

Since our classifier in section 9.1 is relatively simple and our procedure for creating the classifiers removes some information about the tails of the variables, we add more variables to our problem that can capture some additional information about the correlations between variables. There are several ways this can be incorporated, but for this study we simply add the product terms between variables. Specifically, for the 8 variables v_i listed in Table 9.1 we compute the function

$$p(v_i, v_j) = \begin{cases} v_i \times v_j & \text{if } b_i^{low} > 0.7 \ \& \ b_j^{low} > 0.7 \\ v_i \times \frac{1}{v_j} & \text{if } b_i^{low} > 0.7 \ \& \ b_j^{low} < 0.7 \\ \frac{1}{v_i} \times v_j & \text{if } b_i^{low} < 0.7 \ \& \ b_j^{low} > 0.7 \\ \frac{1}{v_i} \times \frac{1}{v_j} & \text{if } b_i^{low} < 0.7 \ \& \ b_j^{low} < 0.7 \end{cases} \quad (9.19)$$

where b_i^{low} is the value computed in step 2 of the weak classifier procedure. Using this prevents the case where we would multiply a variable which favors signal at small values with one that favors signal at large values, which would washout the discriminating power of their product. Using these product terms, the 8 original variables of turn into 36 potential weak classifiers. 4 of these are rejected in the weak classifier transformation, so we are left with 32 weak classifiers. These will be used to train the classifier as described in this section.

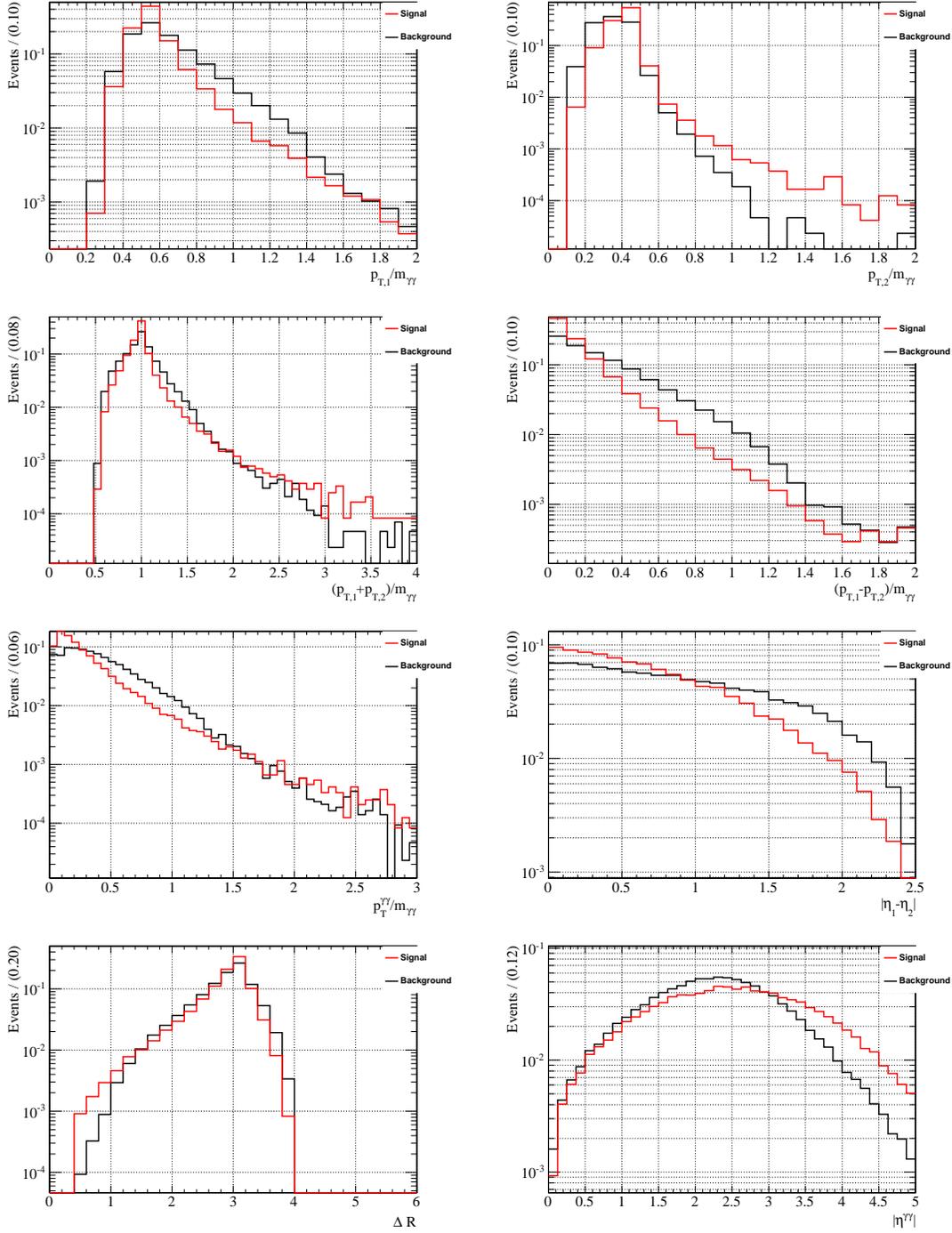


Figure 9.7: The distributions of the kinematic variables used in the learner. The distributions are $p_T^1/m_{\gamma\gamma}$ (first row, left), $p_T^2/m_{\gamma\gamma}$ (first row, right), $(p_T^1 + p_T^2)/m_{\gamma\gamma}$ (second row, left), $(p_T^1 - p_T^2)/m_{\gamma\gamma}$ (second row, right), $p_T^{\gamma\gamma}/m_{\gamma\gamma}$ (third row, left), $\Delta\eta$ (third row, right), ΔR (fourth row, left), $|\eta^{\gamma\gamma}|$ (fourth row, right). In all plots, the distribution for the signal is in red and the distribution for the background is in black.

Chapter 10

Results

We encode the problem of training the classifier on the D-Wave 2TM machine, run the annealing with a $20\mu s$ anneal time and analyze the results. In Figure 9.2 we presented the ideal chimera graph the machine seeks to implement, onto which we can reliably encode 32 variables, which is based on the fit to the most challenging embedding problem we found over 200 random instances. When the system is implemented in practice, however, the actual graph is slightly different because not all the qubits are activated (usually because they cannot be calibrated *in situ* to bring their performance to acceptable levels about the noise threshold).

The graph with the dead qubits indicated is shown in Figure 10.1. There are 8 dead nodes indicated on the graph, taking the total number of qubits down from 512 to 504. While the number of lost qubits may be small, their placement within the graph is significant because they interrupt the connectivity of the graph and make the placement of chains harder. Recalling our requirements for the chains from Section 9.2, we need all chains to be disjoint and to be adjacent to each other chain in at least one place in order to embed the fully connected problem. Losing one node, which means losing 5-6 connections in the graph, complicates the embedding. The net effect is that, rather than being able to embed up to 32 variables, we find we can reliably embed up to 30 variables.

For all the tests done here, we define 200 training samples from our signal and background Monte Carlo as described in Section 9.5. These samples are formed by randomly selecting 10,000 events from each of the signal and background samples

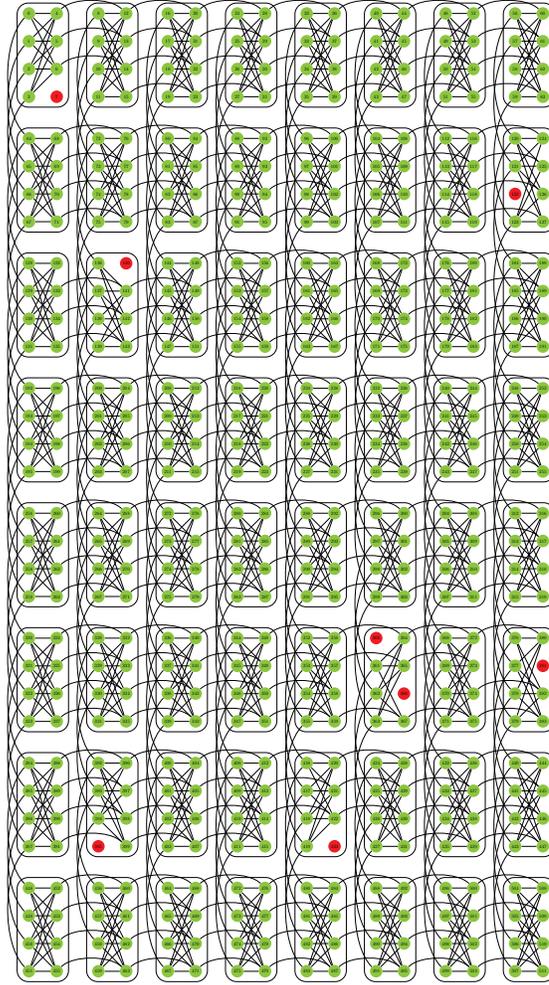


Figure 10.1: The chimera graph as it existed on the D-Wave 2TM machine used for these tests. The green nodes are valid nodes that can be programmed, while the 8 red nodes are “bad” nodes that are disabled.

(without replacement). This gives us 200 training sets that have statistical differences from each other, but represent the same processes.

10.1 Network Training Results

10.1.1 Mapping Physical Chains to Logical Nodes

The first stage of reading the output from the D-Wave 2TM is to understand how to decode the chains in the output of the physical problem. The setup of the problem expects that all the qubits in a single chain should point in the same direction, so that

it can be mapped to a single qubit in the logical problem, but this does not always occur in practice. When this is not the case, we say that the chain is *broken*. To get a sense of how often this occurs, we encode a 27 variable problem onto the D-Wave 2TM and run it 10,000 times for each of the 200 training samples. We then look within each chain in the physical problem and check whether it is broken. Figure 10.2 shows the fraction of chains that are *not* broken for each of the 200 training samples. One can see that the fraction varies quite a bit between problems, with a few having up to half of their chains intact but most having very few or virtually no unbroken chains.

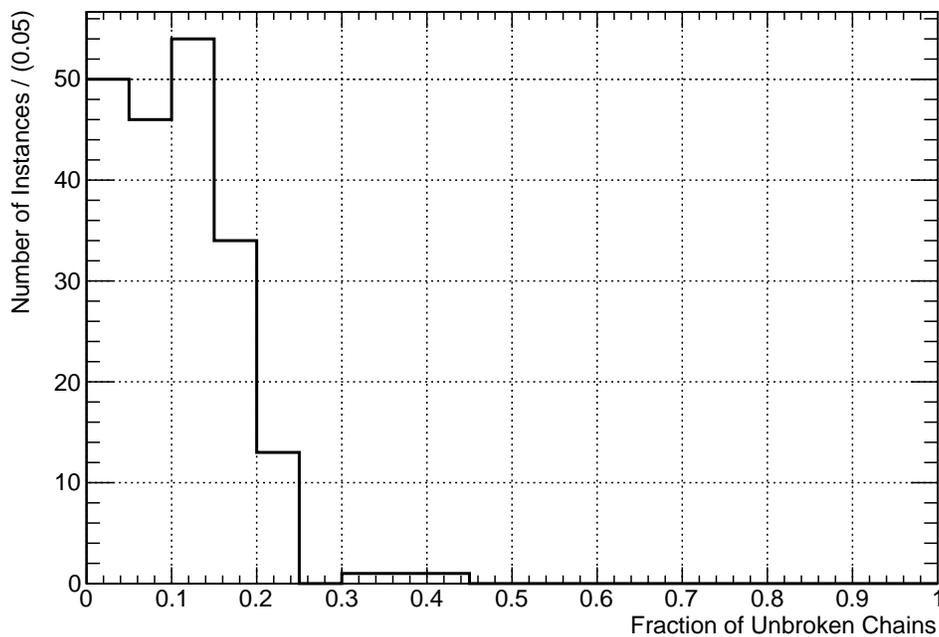


Figure 10.2: The fraction of non-broken chains in each of the 200 problems. This is found by looking at every chain found in the 10,000 runs of the machine.

Figure 10.2 tells us that we will need a method to fix broken chains, rather than discarding results when they happen. The simplest way to do this is to decide by taking the majority vote of all qubits within a chain as the decision for the logical qubit mapped to this chain. If there is a tie (equal number of up and down qubits), we decide the value of the logical node by a coin flip (effectively, the machine is giving us no information about this node). This method has the advantage of being very

quick to compute (requiring only a single pass over the physical network) and being quite effective.

We evaluated two other potential algorithms to decode these problems, which showed no substantial improvement over simple majority vote. This first was a weighted majority vote, where the votes in the direction of the h_i field were weighted more highly than votes in the opposite direction. This showed a small improvement in the performance for some instances, but is quite dangerous in the case that correlations are more important than local fields and so is not used. The second was an algorithm that iteratively used majority vote to fix one chain and then would look at the nearest neighbors to each bit changed by the vote and re-evaluate the neighbors based on the new energy landscape. This was found to slightly improve the performance at the cost of a very long post-processing time, which depended on the embedded graph size and the number of anneals. Since the goal is to speedup the solution of these problems, this additional scaling is not desirable.

10.1.2 Solution Accuracy

Having defined our procedure for correcting broken chains, we now turn to the question of whether the solutions returned solve our problem and, if so, how often. We run the annealer 10,000 times for each of the 200 training samples, which produces a set of 10,000 results for each sample (there can be duplicates within that set). We decode these results by majority vote, as described above, to get a set of results for our logical problem. We then compute the energies of each of the returned results and take the state with the lowest energy to be the best solution found by the annealer.

To check whether the best solution found by the annealer is the true ground state of the system, we use a CPU to exhaustively evaluate the energy of all 2^N possible configurations of the logical problem. The lowest energy state found in this method must be the true ground state of the system. If the lowest energy state from the annealer is the same as the true ground state, then we count the number of times (M) that the state was found out of the 10,000 annealing runs and say that the probability to solve the problem on the annealer is $\approx M/10^5$. If the lowest energy

state from the annealer is not the same as the true ground state, then the annealer did not solve the problem and the solution probability is $< 10^{-5}$.

Figure 10.3 shows the observed solution probability over the 200 training samples for a problem where we select 21 variables (top plot) and 26 variables (bottom plot). In all cases, fewer than 10 of the training samples were not solved correctly within 10,000 anneals, so we are not significantly biased by setting these few to 0. One sees that the distribution has a shape that may not be well represented by the mean, with most events being solved correctly in only a few percent of cases, so we also look at the median problem as well. One can see that the 26 variable problem is solved correctly less often, both on average (red lines) and in the 50th percentile case (blue line). This also indicates that, despite the large number of broken chains seen in Figure 10.2, we can still solve the problem effectively. The value of the mean and the median as a function of the number of variables in the logical problem is shown in Figure 10.4. This gives an indication of how the size of the problem influences the success probability.

We can turn the fraction success probability shown in Figure 10.4 into an estimate of the number of anneals required to produce at least one correct solution. We use the Chernoff bound discussed in Section 9.4 (equation 9.17) to turn the plot of the mean and median success probability as a function of the number of variables into a plot of the mean and median number of anneals required to solve the problem, as shown in Figure 10.5.

Figure 10.5 shows that with only a few hundred anneals on the machine, we can solve a problem of average difficulty. This is a significant result, since it is one of the largest problems encodable on the D-Wave 2TM machine (up to 350 qubits) and one of the first applications of the machine to solving a large, real world problem.

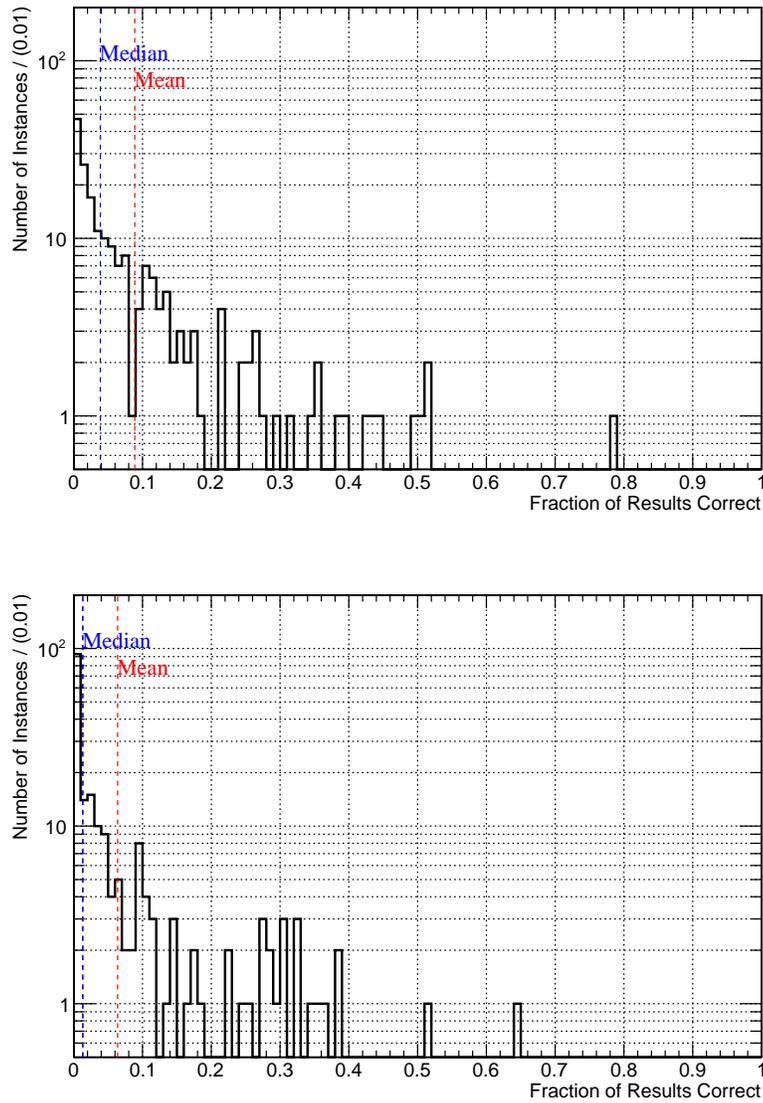


Figure 10.3: The distribution from the 200 training samples of the fraction of correct results from the 10,000 annealing runs for 21 variables (top) and 26 variables (bottom). The blue line shows the median of the distribution while the red line shows the mean.

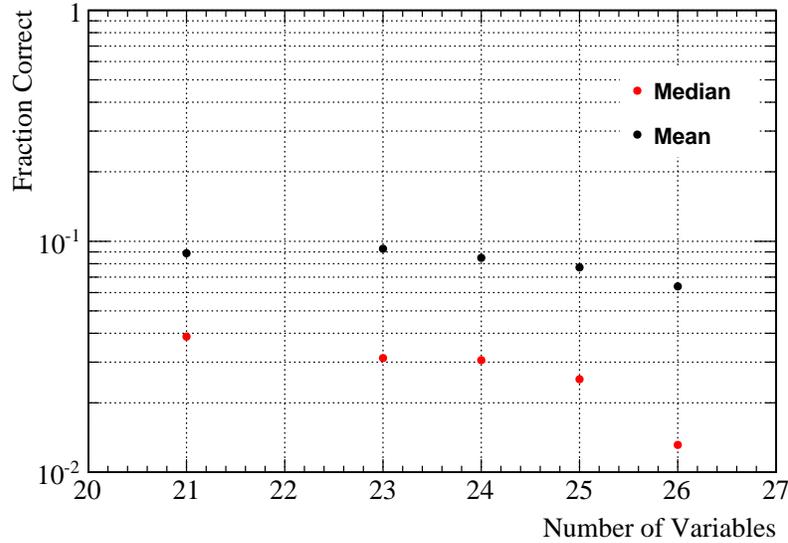


Figure 10.4: The fraction of the 10,000 annealing runs that returned the correct results taken as the mean (black) and median (red) over the 200 training samples.

10.2 Performances of the Network in Classifying Signal Events

Having shown that we can efficiently find the ground state of the network, we now evaluate how well it selects Higgs boson events and rejects continuum background. To evaluate the classifier trained by the D-Wave 2TM machine, we will evaluate its receiver operating characteristic (ROC) curve, which measures the level of signal efficiency and background rejection achieved by cutting on the output of the classifier. For every value v of the output of the classifier $O(x)$ we compute the signal efficiency by looking at the signal training sample (\mathcal{S}) and evaluating

$$\epsilon_{\mathcal{S}}(v) = \frac{1}{|\mathcal{S}|} \sum_{x \in \mathcal{S}} I_v(x) \quad (10.1)$$

where

$$I_v(x) = \begin{cases} 1 & \text{if } x \geq v \\ 0 & \text{else} \end{cases}$$

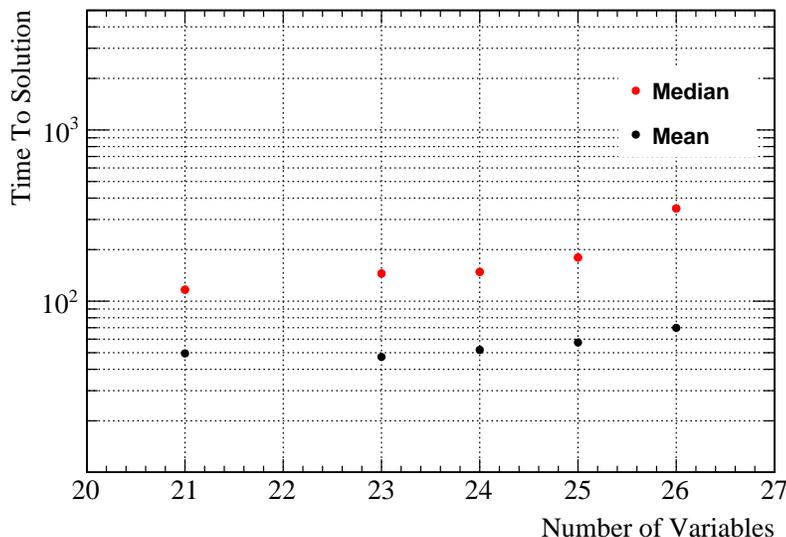


Figure 10.5: The time to solution with 99% accuracy in units of the number of annealing runs as a function of the number of variables in the logical problem for the mean solution (black) and the 50th percentile solution (red).

is the indicator function, so $\epsilon_S(v)$ is just the fraction of signal events with $O(x) > v$. We also evaluate the background efficiency of the cut on the background sample (\mathcal{B}):

$$\epsilon_B = \frac{1}{|\mathcal{B}|} \sum_{x \in \mathcal{B}} I_v(x) \quad (10.2)$$

and then define the background rejection as $r_B = 1 - \epsilon_B$.

The second component we need to benchmark our result is another classifier trained in a different way. We train a multilayer perceptron (MLP) [123, 124] using the TMVA package for ROOT [60] using the 8 input variables from Table 9.1. It has been shown that MLPs can approximate any smooth function with a single hidden layer [125], so we use a single hidden layer with 13 nodes in our network. We have found that the performance of our network is not very sensitive to small changes in the number of nodes or the initial seed values of our network (determined by training many times with randomized start values and training samples and observing the results). The MLP gives us a classifier with a continuous output estimating how signal-like or background-like an event is, which we can analyze using ROC curves in

an identical method to our linear classifier.

Figure 10.6 shows the ROC curve for the linear classifier trained on the D-Wave 2TM and the MLP trained using the TMVA package. One can see that they have quite different shapes and in some regions the linear classifier gives better performance while in some regions the MLP gives better performance. Here we use 26 variables to train the network and set the value of the penalty term to $\lambda = 0.01 \times \max_i(|h_i|)$. This level of performance makes the linear classifier already useful on its own, though we will discuss ways to improve it in Section 10.4.

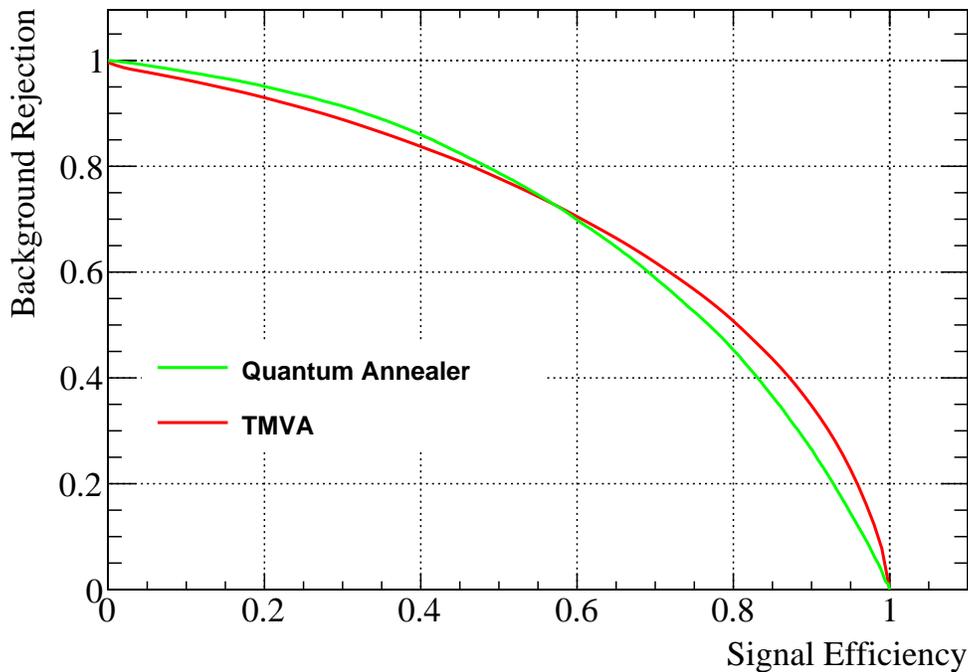


Figure 10.6: The ROC curve for the linear classifier with 26 variables and $\lambda = 0.01 \times \max_i(|h_i|)$ trained on the quantum annealer (green) and the MLP trained by TMVA. The curves represent the average value of the classifier taken over the 200 training samples, but the spread is smaller than the line width for both the linear classifier and the MLP.

Figure 10.6 showed the ROC with $\lambda = 0.01 \times \max_i(|h_i|)$, but we can also investigate how the performance is changed by increasing the value of λ . Figure 10.7 shows the ROC for $\lambda = 0.05 \times \max_i(|h_i|)$ and $\lambda = 0.80 \times \max_i(|h_i|)$. One sees that the ROC curve does not change substantially over a wide range of λ values (though it

does become slightly worse at large values of the signal efficiency). This indicates that the performance over most of the range is being driven by a few key variables and the rest are being used to drive performance in the high signal efficiency regime. This also gives us some insight into what the shape difference in Figure 10.6 is telling us; the MLP has access to more degrees of freedom (since it can weight and combine variables), which is most useful when you have very high signal efficiency (taking most of the data), but the few variables that drive the linear classifier are most important over the rest of the range.

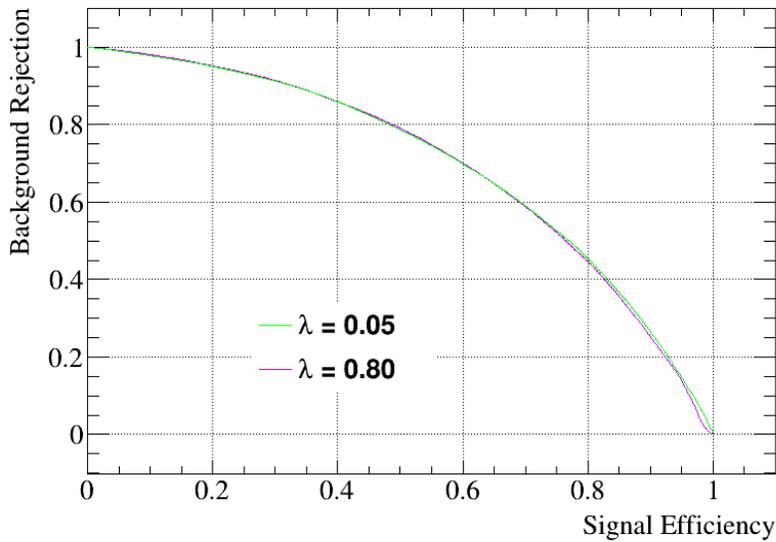


Figure 10.7: The ROC curve for the linear classifier with 26 variables trained on the D-Wave 2TM with $\lambda = 0.05 \times \max_i(|h_i|)$ (green) and $\lambda = 0.80 \times \max_i(|h_i|)$ (purple). One sees that the performance of the classifier is not very sensitive over a wide range of values of λ .

The remaining thing to check is whether our weak classifier is being over-trained on the quantum annealer. Over-training occurs when the classifiers learn statistical features of the training sample and reject real signal (or accept real background) based on the presence of those features. Fortunately, it is easy to check whether over-training has occurred, by looking at the distribution of the output of the classifier on the training sample and on an orthogonal validation sample where the categories are known. We have just such a sample for this study, since we sampled 10,000 signal and 10,000 background events out of total samples that contained (22,000 and 42,000)

events respectively. Figure 10.8 shows the distributions of the outputs for signal and background samples for the training and validation sample, as well as the result of the χ^2/NDF test on the histograms. We see that there are no systematic deviations in any of the samples tested and the χ^2 test indicates good agreement in all cases. This indicates that there is no over-training in any of these samples.

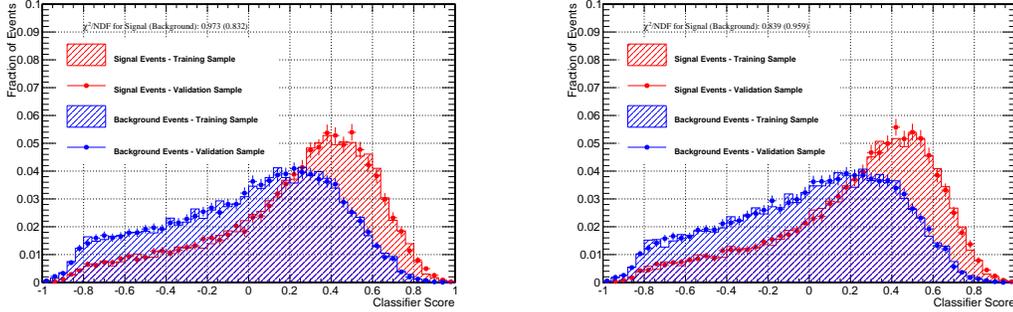


Figure 10.8: Comparison of the output of the network for signal samples (red) and background samples (blue) for the training sample (fill) and validation sample (points) for $\lambda = 0.01 \times \max_i(|h_i|)$ (left) and $\lambda = 0.05 \times \max_i(|h_i|)$ (right). The χ^2/NDF for the histogram is reported on the plots as well.

10.3 Improving Success Probability and Performance

So far, we have looked at the ground state solution to our linear classifier problem and evaluated its characteristics, we now investigate what happens if we include some excited state solutions in our analysis. An excited state solution is a network that does not minimize the energy of the logical Hamiltonian, which, if we go back to our discussion in Section 9.1, is one that does not minimize the error of the classifier. If we consider excited states within a few % in energy from the ground state, then these encode solutions that *almost* minimize the error, so they should be valid classifiers with slightly lower performance. Indeed, since our definition of the error is based on how far away from +1 a signal event is placed and how far away from -1 a background event is placed, which does not map directly to having the best ROC curve at every point in the plane, it is entirely possible that some of the excited state ROC curves outperform the ground state ROC at certain points.

Let us consider some set of excited states $\{e_i : 1 \leq i \leq N\}$; for each one we build a ROC curve $R_i(s)$ which gives the level of background rejection as a function of the level of signal efficiency s . Letting $R_0(s)$ be the ROC for the ground state, we can then define a new ROC curve

$$R_N^{sup}(s) = \max_{0 \leq i \leq N} R_i(s) \quad (10.3)$$

at every point of signal efficiency s , this picks the best network from all the ones we consider. Intuitively, the purpose of the ROC curve is to help choose a working point for the signal efficiency based on its level of background rejection (or vice-versa); this R_N^{sup} can be used in the same way except when we choose the working point we also choose the network that gives us the best performance at that working point.

With this definition, we can now sensibly define the ROC of the top N of solutions. We look at all solutions returned by the annealer and choose the N_P solutions that have energy within P% of the ground state energy and compute the ROC using Equation 10.3. We evaluate this for $P = 1\%, 2\% \dots, 10\%$ in Figure 10.9, where we can see the ROC curves for the excited states in the whole range in the top plot while the bottom plot shows the same curve zoomed in on the x and y axes to show the structure. We can see the gain from adding the excited states in Figure 10.10, which shows the difference between the background rejection for the top N% excited states and the ground state. We see that adding the excited states gets up to a 1% improvement in background rejection at a given signal efficiency. This effect is more important than this numerical value indicates: an additional 0.5% background rejection in a place where we already reject 92% of the background is a 6.3% reduction in remaining background.

We have seen that adding excited states gives better performances for the classifier, but this doesn't mean that we could run the annealer just once and take whatever excited state it returns. The annealer samples the probability distribution of the evolving system, so to ensure the level of performance seen here, we need to make sure that the results at least sample this distribution well. Since Figure 10.9 takes

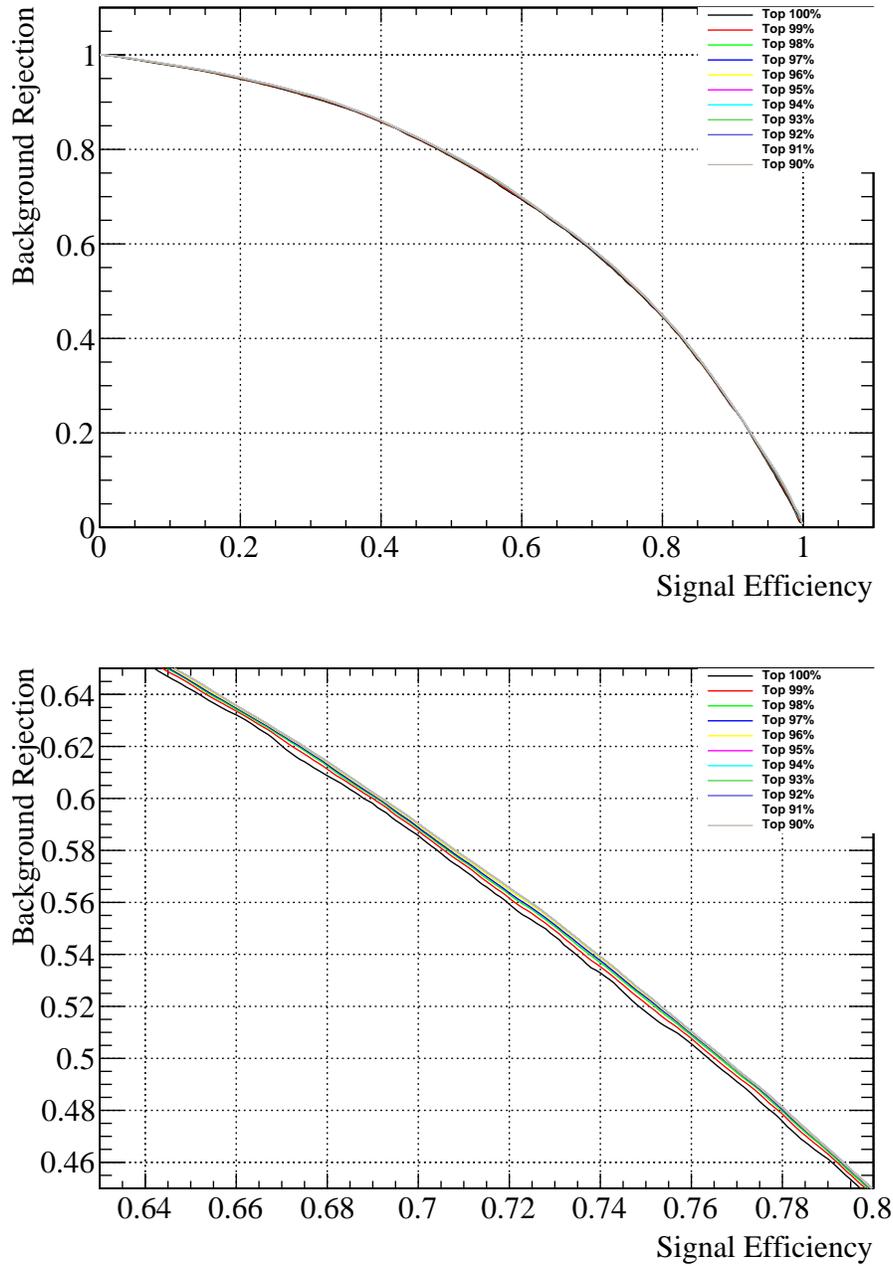


Figure 10.9: The ROC curves for the classifier with 26 variables trained on the D-Wave 2TM using the top n% of excited states to make the ROC curve. The top plot shows the entire range of the curve, while the bottom plot zooms in on a region of interest around the center to show the separation.

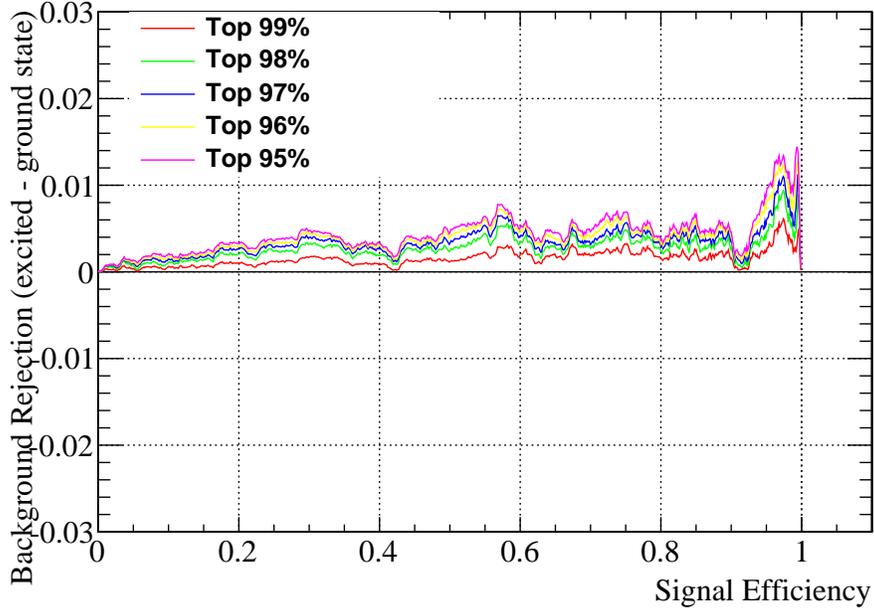


Figure 10.10: The difference of ROC curves with 26 variables for the top $n\%$ of excited states to the ROC curve for the ground state only. At each point of signal efficiency, this difference is defined as $R_N^{sup} - R_0$, so it reflects the additional background rejection gained from including excited states at each point of signal efficiency.

the supremum over many different excited states, we need a large number of these to achieve this level of performance, any one of the excited states may outperform the ground state at a single point of signal efficiency, but is extremely unlikely to outperform it everywhere (it is possible for this to happen, since the λ penalty means we technically don't choose the absolute minimum error network). The excited states should be treated as a nice benefit to the performance of the classifier, but the goal should still be to find the ground state at least once with high probability.

10.4 Current Status and Outlook

The goal of this study was to use the D-Wave 2TM machine to solve a real problem in high energy physics, benchmark the performance of the classifier and investigate the dynamics of the machine output. This has been successfully achieved and we have produced a classifier that is competitive with classifiers found using more mature

classical technologies. There are a wide variety of additional studies that have been considered during this project but left for future study in favor of producing a proof-of-concept result.

The first is to optimize the variable selection that goes into this problem and see if there are better ways to select the variables and convert them into weak classifiers. Our technique works well for the problem under consideration, but it is deliberately sacrificing information on the tails of the distribution that might be more important for other problems. We have found that we can rank input variables based on the value of λ at which they are always included in the classifier, which suggests the possibility of dynamically pruning useless variables and including other.

We also notice that it may be possible to embed substantially more variables into the network by cleverly manipulating the C_{ij} matrix. We operated from the point of view that the C_{ij} matrix defines a fully connected graph, which it technically does. With the noise on the couplers, however, we would be tempted to say that any value of $C_{ij} < 0.05$ is essentially mapped to a random value when applied to the physical problem. Using this information, it would be possible to dynamically sparsify the C_{ij} matrix making the initial graph non-fully connected and allowing a larger initial graph to be embedded into the same physical size. Future versions of the D-Wave machine may improve the noise model, so it is a bit dangerous to rely on this going forward, but it could have interesting application in the short term.

Finally, we see the simple linear classifiers we have built as an ideal candidate for boosting. Boosting is an algorithm where one trains a classifier, evaluates it on all members of the training sample, weights each event in the sample based on whether the event was correctly classified and then trains a new classifier. This procedure is done iteratively and at the end one has a large collection of classifiers, each of which will produce a different output on a single event. This has been shown to dramatically increase the power of simple classifiers, such as binary decision trees. One of the key criteria for doing the boosting effectively is that each of the individual classifiers should be hard to over-train, which our linear classifier is.

There are some technical challenges that would need to be addressed for boosting

to be usable with this training method. To boost, one needs to be able to apply weights to the training sample, which would essentially mean recalculating the C_{ij} and C_{iy} values after each iteration in a weighted way. This would require a procedure where the network is trained on the annealer, the training is read off and the network computed, then the C_{ij} and C_{iy} are recomputed and the new problem is written onto the annealer. This is certainly possible, but would be relatively slow at least in the current instantiation of the API. Furthermore, one would need to study how a case where the ground state is not found at one of the steps would influence the training; my guess is that it would not be a problem, but it would need careful study.

Quantum machine learning is a rich field with many potential applications to HEP and other fields. We have performed one of the first studies that has demonstrated a real application of a classifier trained on a quantum annealer. It is also the first study of which we are aware that applies a learner trained on a quantum annealer to a problem in HEP. We have demonstrated that this training can be done successfully and the success probability is sufficiently high to be usable. The performance of the classifier is comparable to the performance of classifiers trained using classical methods, and we have shown that the nature of our training has tangible benefits in the form of the excited states.

Bibliography

- [1] Bill Bryson. *A short history of nearly everything*. Broadway Books, 2003.
- [2] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Kennedy, R. Kogler, K. Mnig, M. Schott, and J. Stelzer. The electroweak fit of the standard model after the discovery of a new boson at the lhc. *The European Physical Journal C*, 72(11), 2012. URL: <http://dx.doi.org/10.1140/epjc/s10052-012-2205-9>, doi:10.1140/epjc/s10052-012-2205-9.
- [3] K.A. Olive et al. Review of Particle Physics. *Chin.Phys.*, C38:090001, 2014. doi:10.1088/1674-1137/38/9/090001.
- [4] Michael Peskin and Daniel Schroeder. *An introduction to quantum field theory*. 1995.
- [5] H. David Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, Jun 1973. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.30.1346>, doi:10.1103/PhysRevLett.30.1346.
- [6] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.30.1343>, doi:10.1103/PhysRevLett.30.1343.
- [7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964. URL:

- <http://link.aps.org/doi/10.1103/PhysRevLett.13.585>,
[doi:10.1103/PhysRevLett.13.585](https://doi.org/10.1103/PhysRevLett.13.585).
- [8] T. W. B. Kibble. Symmetry breaking in non-abelian gauge theories. *Phys. Rev.*, 155:1554–1561, Mar 1967. URL:
<http://link.aps.org/doi/10.1103/PhysRev.155.1554>,
[doi:10.1103/PhysRev.155.1554](https://doi.org/10.1103/PhysRev.155.1554).
- [9] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. URL:
<http://link.aps.org/doi/10.1103/PhysRevLett.13.508>,
[doi:10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [10] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964. URL:
<http://link.aps.org/doi/10.1103/PhysRevLett.13.321>,
[doi:10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
- [11] Peter W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156–1163, May 1966. URL:
<http://link.aps.org/doi/10.1103/PhysRev.145.1156>,
[doi:10.1103/PhysRev.145.1156](https://doi.org/10.1103/PhysRev.145.1156).
- [12] S. Dittmaier et al. *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*. CERN, Geneva, 2011. Comments: 153 pages, 43 figures, to be submitted to CERN Report. Working Group web page:
<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>.
- [13] Torbjrn Sjstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006. URL:
<http://stacks.iop.org/1126-6708/2006/i=05/a=026>.
- [14] P Nason. *JHEP*, 0411:040, 2004. [arXiv:hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146).
- [15] S. Frixione, P. Nason, and C. Oleari. *JHEP*, 0711:070, 2007. [arXiv:0709.2092](https://arxiv.org/abs/0709.2092).

- [16] S. Alioli, P. Nason, C. Oleari, and E. Re. *JHEP*, 1006:043, 2010.
[arXiv:1002.2581](#).
- [17] S. Alioli, P. Nason, C. Oleari, and E. Re. Higgs boson production in gluon fusion. *JHEP*, 0904, 2009.
- [18] P. Nason and C. Oleari. Higgs boson production in vector boson fusion. *JHEP*, 1002:037, 2010. [arXiv:0911.5299](#).
- [19] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytrcek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell'Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J.J. Gomez Cadenas, I. Gonzalez, G. Gracia Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F.W. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampn, V. Lara, V. Lefebvre, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. Mora de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O'Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M.G. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. Di Salvo, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, M. Takahata, S. Tanaka, E. Tcherniaev, E. Safai Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wander, H. Weber, J.P. Wellisch, T. Wenaus, D.C. Williams, D. Wright, T. Yamada, H. Yoshida, and D. Zschesche. Geant4a simulation toolkit. *Nuclear Instruments and Methods*

- in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. URL: <http://www.sciencedirect.com/science/article/pii/S0168900203013688>, doi:[http://dx.doi.org/10.1016/S0168-9002\(03\)01368-8](http://dx.doi.org/10.1016/S0168-9002(03)01368-8).
- [20] P. Fayet. Supersymmetry and weak, electromagnetic and strong interactions. *Physics Letters B*, 64(2):159 – 162, 1976. URL: <http://www.sciencedirect.com/science/article/pii/0370269376903191>, doi:[http://dx.doi.org/10.1016/0370-2693\(76\)90319-1](http://dx.doi.org/10.1016/0370-2693(76)90319-1).
- [21] P. Fayet and S. Ferrara. Supersymmetry. *Physics Reports*, 32(5):249 – 334, 1977. URL: <http://www.sciencedirect.com/science/article/pii/0370157377900667>, doi:[http://dx.doi.org/10.1016/0370-1573\(77\)90066-7](http://dx.doi.org/10.1016/0370-1573(77)90066-7).
- [22] Steven Weinberg. Implications of Dynamical Symmetry Breaking. *Phys.Rev.*, D13:974–996, 1976. doi:10.1103/PhysRevD.13.974.
- [23] Glennys R. Farrar and Pierre Fayet. Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry. *Physics Letters B*, 76(5):575 – 579, 1978. URL: <http://www.sciencedirect.com/science/article/pii/0370269378908584>, doi:[http://dx.doi.org/10.1016/0370-2693\(78\)90858-4](http://dx.doi.org/10.1016/0370-2693(78)90858-4).
- [24] V. Barger and Ernest Ma. Natural conservation of R parity in supersymmetry. *Phys. Rev. D*, 51:1332–1336, Feb 1995. URL: <http://link.aps.org/doi/10.1103/PhysRevD.51.1332>, doi:10.1103/PhysRevD.51.1332.
- [25] Daniele Alves, Nima Arkani-Hamed, Sanjay Arora, Yang Bai, Matthew Baumgart, Joshua Berger, Matthew Buckley, Bart Butler, Spencer Chang, Hsin-Chia Cheng, Clifford Cheung, R Sekhar Chivukula, Won Sang Cho, Randy Cotta, Mariarosaria DAlfonso, Sonia El Hedri, Rouven Essig (Editor),

Jared A Evans, Liam Fitzpatrick, Patrick Fox, Roberto Franceschini, Ayres Freitas, James S Gainer, Yuri Gershtein, Richard Gray, Thomas Gregoire, Ben Gripaios, Jack Gunion, Tao Han, Andy Haas, Per Hansson, JoAnne Hewett, Dmitry Hits, Jay Hubisz, Eder Izaguirre, Jared Kaplan, Emanuel Katz, Can Kilic, Hyung-Do Kim, Ryuichiro Kitano, Sue Ann Koay, Pyungwon Ko, David Krohn, Eric Kuflik, Ian Lewis, Mariangela Lisanti (Editor), Tao Liu, Zhen Liu, Ran Lu, Markus Luty, Patrick Meade, David Morrissey, Stephen Mrenna, Mihoko Nojiri, Takemichi Okui, Sanjay Padhi, Michele Papucci, Michael Park, Myeonghun Park, Maxim Perelstein, Michael Peskin, Daniel Phalen, Keith Rehermann, Vikram Rentala, Tuhin Roy, Joshua T Ruderman, Veronica Sanz, Martin Schmaltz, Stephen Schnetzer, Philip Schuster (Editor), Pedro Schwaller, Matthew D Schwartz, Ariel Schwartzman, Jing Shao, Jessie Shelton, David Shih, Jing Shu, Daniel Silverstein, Elizabeth Simmons, Sunil Somalwar, Michael Spannowsky, Christian Spethmann, Matthew Strassler, Shufang Su, Tim Tait (Editor), Brooks Thomas, Scott Thomas, Natalia Toro (Editor), Tomer Volansky, Jay Wacker (Editor), Wolfgang Waltenberger, Itay Yavin, Felix Yu, Yue Zhao, Kathryn Zurek, and LHC New Physics Working Group. Simplified models for lhc new physics searches. *Journal of Physics G: Nuclear and Particle Physics*, 39(10):105005, 2012. URL: <http://stacks.iop.org/0954-3899/39/i=10/a=105005>.

- [26] K Nakamura and Particle Data Group. Review of particle physics. *Journal of Physics G: Nuclear and Particle Physics*, 37(7A):075021, 2010. URL: <http://stacks.iop.org/0954-3899/37/i=7A/a=075021>.
- [27] Michael Kramer, Anna Kulesza, Robin van der Leeuw, Michelangelo Mangano, Sanjay Padhi, et al. Supersymmetry production cross sections in pp collisions at $\sqrt{s} = 7$ TeV. 2012. [arXiv:1206.2892](https://arxiv.org/abs/1206.2892).
- [28] Christoph Borschensky, Zoltan Gecse, Michael Kraemer, Robin van der Leeuw, Anna Kulesza, Michelangelo Mangano, Sanjay Padhi, Tilman Plehn,

Xavier Portell, and Sezen Sekmen.

<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections>.

- [29] L. Evans and P. Bryant. Lhc machine. *J. Instrum*, 3:S08001, 2008.
doi:10.1088/1748-0221/3/08/S08001.
- [30] M. Lamont. Status of the lhc. *J. Phys. Conf. Ser.*, 455(1):012001, 2013. URL:
<http://stacks.iop.org/1742-6596/455/i=1/a=012001>.
- [31] The TOTEM Collaboration. Luminosity-independent measurement of the proton-proton total cross section at $\sqrt{s} = 8$ TeV. *Phys. Rev. Lett.*, 111:012001, Jul 2013. doi:10.1103/PhysRevLett.111.012001.
- [32] The CMS Collaboration. The cms experiment at the cern lhc. *J. Instrum*, 3:S08004, 2008. doi:10.1088/1748-0221/3/08/S08004.
- [33] The CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006.
- [34] A. Herve, G. Acquistapace, D. Campi, P. Cannarsa, P. Fabbriatore, et al. Status of the CMS magnet. *IEEE Trans.Appl.Supercond.*, 12:385–398, 2002.
doi:10.1109/TASC.2002.1018425.
- [35] V.I. Klyukhin, A. Ball, F. Bergsma, D. Campi, B. Cure, et al. Measurement of the CMS Magnetic Field. *IEEE Trans.Appl.Supercond.*, 18:395–398, 2008.
arXiv:1110.0306, doi:10.1109/TASC.2008.921242.
- [36] The CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.
- [37] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the cms tracker. *Journal of Instrumentation*, 9(10):P10009, 2014. URL:
<http://stacks.iop.org/1748-0221/9/i=10/a=P10009>.

- [38] P. Billoir. Progressive track recognition with a kalman-like fitting procedure. *Comp. Phys. Comm*, 57(1-3):390–394, 1989.
- [39] P. Billoir and S. Qian. Simultaneous pattern recognition and track fitting by the kalman filtering method. *Nucl. Instrum. Meth. A*, 262:219, 1990.
- [40] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, Nov 1998. doi:10.1109/5.726788.
- [41] R Frhwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Adaptive Vertex Fitting. Technical Report CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.
- [42] The CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.
- [43] Chia Ming, Kuo, and the CMS ECAL Group. The commissioning and first results on the performance of the cms preshower detector. *Journal of Physics: Conference Series*, 293(1):012058, 2011. URL: <http://stacks.iop.org/1742-6596/293/i=1/a=012058>.
- [44] S. Baccaro, J.E. Bateman, F. Cavallari, V. Da Ponte, K. Deiters, P. Denes, M. Diemoz, Th. Kirn, A.L. Lintern, E. Longo, M. Montecchi, Y. Musienko, J.P. Pansart, D. Renker, S. Reucroft, G. Rosi, R. Rusack, D. Ruuska, R. Stephenson, and M.J. Torbet. Radiation damage effect on avalanche photodiodes. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 426(1):206 – 211, 1999. URL: <http://www.sciencedirect.com/science/article/pii/S0168900298014934>, doi:[http://dx.doi.org/10.1016/S0168-9002\(98\)01493-4](http://dx.doi.org/10.1016/S0168-9002(98)01493-4).
- [45] Z. Antunovic, I. Britvitch, K. Deiters, N. Godinovic, Q. Ingram, A. Kuznetsov, Y. Musienko, I. Puljak, D. Renker, S. Reucroft, R. Rusack,

- T. Sakhelashvili, A. Singovski, I. Soric, and J. Swain. Radiation hard avalanche photodiodes for the {CMS} detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 537(12):379 – 382, 2005. Proceedings of the 7th International Conference on Inorganic Scintillators and their Use in Scientific and Industrial Applications. URL: <http://www.sciencedirect.com/science/article/pii/S016890020401856X>, doi:<http://dx.doi.org/10.1016/j.nima.2004.08.047>.
- [46] K.W. Bell, R.M. Brown, D.J.A. Cockerill, P S Flower, P.R. Hobson, B.W. Kennedy, A L Lintern, C.S. Selby, O. Sharif, M. Sproston, and J.H. Williams. Vacuum phototriodes for the cms electromagnetic calorimeter endcap. *Nuclear Science, IEEE Transactions on*, 51(5):2284–2287, Oct 2004. doi:10.1109/TNS.2004.836053.
- [47] P Adzic. Energy resolution of the barrel of the cms electromagnetic calorimeter. *Journal of Instrumentation*, 2(04):P04004, 2007. URL: <http://stacks.iop.org/1748-0221/2/i=04/a=P04004>.
- [48] Emilio Meschi, T Monteiro, Christopher Seez, and Pratibha Vikas. Electron Reconstruction in the CMS Electromagnetic Calorimeter. Technical Report CMS-NOTE-2001-034, CERN, Geneva, Jun 2001.
- [49] CMS: The hadron calorimeter technical design report. 1997.
- [50] *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. URL: <http://cds.cern.ch/record/343814>.
- [51] The CMS collaboration. Performance of cms muon reconstruction in pp collision events at $\sqrt{s} = 7$ tev. *Journal of Instrumentation*, 7(10):P10002, 2012. URL: <http://stacks.iop.org/1748-0221/7/i=10/a=P10002>.

- [52] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009.
- [53] Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector. Technical Report CMS-PAS-PFT-10-001, 2010.
- [54] Vardan Khachatryan et al. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}=8$ TeV. 2015. [arXiv:1502.02701](https://arxiv.org/abs/1502.02701).
- [55] Photon reconstruction and identification at $\sqrt{s} = 7$ TeV. Technical Report CMS-PAS-EGM-10-005, CERN, 2010. Geneva, 2010.
- [56] Vardan Khachatryan et al. Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV . Submitted to *JINST*, 2015. [arXiv:1502.02702](https://arxiv.org/abs/1502.02702).
- [57] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Physics Letters B*, 659(12):119 – 126, 2008. URL: <http://www.sciencedirect.com/science/article/pii/S0370269307011094>, doi:<http://dx.doi.org/10.1016/j.physletb.2007.09.077>.
- [58] Yong Yang. *Search for a standard model Higgs boson decaying to two photons in CMS at the LHC*. PhD thesis, California Institute of Technology, 2013.
- [59] The CMS Collaboration. Observation of the diphoton decay of the higgs boson and measurement of its properties. *Eur. Phys. J. C*, 74:3076, 2014.
- [60] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. Tmva: Toolkit for multivariate data analysis. *PoS*, ACAT:040, 2007. [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).

- [61] Tomasz Skwarnicki. A STUDY OF THE RADIATIVE CASCADE TRANSITIONS BETWEEN THE UPSILON-PRIME AND UPSILON RESONANCES. DESY-F31-86-02.
- [62] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 0804:063, 2008. [arXiv:0802.1189](https://arxiv.org/abs/0802.1189), [doi:10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063).
- [63] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Catchment Area of Jets. *JHEP*, 0804:005, 2008. [arXiv:0802.1188](https://arxiv.org/abs/0802.1188), [doi:10.1088/1126-6708/2008/04/005](https://doi.org/10.1088/1126-6708/2008/04/005).
- [64] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur.Phys.J.*, C72:1896, 2012. [arXiv:1111.6097](https://arxiv.org/abs/1111.6097), [doi:10.1140/epjc/s10052-012-1896-2](https://doi.org/10.1140/epjc/s10052-012-1896-2).
- [65] S. Catani, Yu.L. Dokshitzer, and B.R. Webber. The k_T -clustering algorithm for jets in deep inelastic scattering and hadron collisions. *Physics Letters B*, 285(3):291 – 299, 1992. URL: <http://www.sciencedirect.com/science/article/pii/037026939291467N>, [doi:http://dx.doi.org/10.1016/0370-2693\(92\)91467-N](http://dx.doi.org/10.1016/0370-2693(92)91467-N).
- [66] S. Catani, Yu.L. Dokshitzer, M.H. Seymour, and B.R. Webber. Longitudinally-invariant k_T -clustering algorithms for hadron-hadron collisions. *Nuclear Physics B*, 406(12):187 – 224, 1993. URL: <http://www.sciencedirect.com/science/article/pii/055032139390166M>, [doi:http://dx.doi.org/10.1016/0550-3213\(93\)90166-M](http://dx.doi.org/10.1016/0550-3213(93)90166-M).
- [67] Jet Energy Resolution in CMS at $\sqrt{s}=7$ TeV. Technical Report CMS-PAS-JME-10-014, CERN, Geneva, 2011.
- [68] Jet energy scale uncertainty correlations between ATLAS and CMS. Technical Report CMS-PAS-JME-14-003, CERN, Geneva, 2014.

- [69] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002, 2011. URL: <http://stacks.iop.org/1748-0221/6/i=11/a=P11002>.
- [70] Status of the 8 TeV Jet Energy Corrections and Uncertainties based on 11 fb⁻¹ of data in CMS. May 2013. URL: <https://cds.cern.ch/record/1545350>.
- [71] The CMS collaboration. Identification of b-quark jets with the cms experiment. *Journal of Instrumentation*, 8(04):P04013, 2013. URL: <http://stacks.iop.org/1748-0221/8/i=04/a=P04013>.
- [72] Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events. Technical Report CMS-PAS-BTV-13-001, CERN, Geneva, 2013.
- [73] The CMS collaboration. Missing transverse energy performance of the cms detector. *Journal of Instrumentation*, 6(09):P09001, 2011. URL: <http://stacks.iop.org/1748-0221/6/i=09/a=P09001>.
- [74] MET performance in 8 TeV data. Technical Report CMS-PAS-JME-12-002, CERN, Geneva, 2013.
- [75] Valentina Gori. The cms high level trigger. *International Journal of Modern Physics: Conference Series*, 31:1460297, 2014. URL: <http://www.worldscientific.com/doi/abs/10.1142/S201019451460297X>, arXiv:<http://www.worldscientific.com/doi/pdf/10.1142/S201019451460297X>, doi:10.1142/S201019451460297X.
- [76] A Tapper and Darin Acosta. CMS Technical Design Report for the Level-1 Trigger Upgrade. Technical Report CERN-LHCC-2013-011. CMS-TDR-12, CERN, Geneva, Jun 2013. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I,

didier.claude.contardo@cern.ch. URL:
<https://cds.cern.ch/record/1556311>.

- [77] L1 calorimeter trigger upgrade: electron and photon performance. Jan 2015.
URL: <http://cds.cern.ch/record/1987288>.
- [78] The cms trigger. Technical Report CMS-TRG-12-001, CERN, 2015.
- [79] Search for a Higgs boson decaying into two photons in the CMS detector.
Technical Report CMS-PAS-HIG-11-021, CERN, Geneva, 2011. URL:
<https://cds.cern.ch/record/1376642>.
- [80] The CMS Collaboration. Observation of a new boson at a mass of 125 gev
with the cms experiment at the lhc. *Phys. Lett. B*, 716:30–61, 2012.
- [81] The ATLAS Collaboration. Observation of a new particle in the search for the
standard model higgs boson with the atlas detector at the lhc. *Phys. Lett. B*,
716:1–29, 2012.
- [82] C. Rogan. Kinematical variables towards new dynamics at the lhc. 2010.
[arXiv:1006.2727](https://arxiv.org/abs/1006.2727).
- [83] The CMS collaboration. Search for supersymmetry using razor variables in
events with b-tagged jets in pp collisions at $\sqrt{s} = 8 \text{ tev}$. 2015. Submitted to
Phys. Rev. D. [arXiv:1502.00300](https://arxiv.org/abs/1502.00300).
- [84] The CMS collaboration. Inclusive search for supersymmetry using razor
variables in pp collisions at $\sqrt{s} = 7 \text{ TeV}$. *Phys. Rev. Lett.*, 111:081802, Aug
2013. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.111.081802>,
[doi:10.1103/PhysRevLett.111.081802](https://doi.org/10.1103/PhysRevLett.111.081802).
- [85] The ATLAS collaboration. Multi-channel search for squarks and gluinos in
 $\sqrt{s} = 7 \text{ TeV}$ pp collisions with the atlas detector at the lhc. *The European
Physical Journal C*, 73(3), 2013. URL:

<http://dx.doi.org/10.1140/epjc/s10052-013-2362-5>,
doi:10.1140/epjc/s10052-013-2362-5.

- [86] Inclusive search for squarks and gluinos in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. D*, 85:012004, Jan 2012. URL:
<http://link.aps.org/doi/10.1103/PhysRevD.85.012004>,
doi:10.1103/PhysRevD.85.012004.
- [87] Ken Aho, DeWayne Derryberry, and Teri Peterson. Model selection for ecologists: the worldviews of aic and bic. *Ecology*, 95(3):631–636, Mar 2014. URL: <http://dx.doi.org/10.1890/13-1452.1>, doi:10.1890/13-1452.1.
- [88] H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, Dec 1974. doi:10.1109/TAC.1974.1100705.
- [89] Kenneth P. Burnham and David R. Anderson. Multimodel inference: Understanding aic and bic in model selection. *Sociological Methods & Research*, 33(2):261–304, 2004. URL:
<http://smr.sagepub.com/content/33/2/261.abstract>,
arXiv:<http://smr.sagepub.com/content/33/2/261.full.pdf+html>,
doi:10.1177/0049124104268644.
- [90] Georges Aad et al. Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector. *Phys.Rev.*, D90(11):112015, 2014. arXiv:1408.7084, doi:10.1103/PhysRevD.90.112015.
- [91] M. Vecchi S. Kirkpatrick, C. Gelatt. Optimization by simulated annealing. *Science*, 220(4598). doi:DOI:10.1126/science.220.4598.671.
- [92] Scott Kirkpatrick. Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics*, 34(5-6):975–986, 1984. URL:
<http://dx.doi.org/10.1007/BF01009452>, doi:10.1007/BF01009452.

- [93] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1953. doi:<http://dx.doi.org/10.1063/1.1699114>.
- [94] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, Joshua Lapan, Andrew Lundgren, and Daniel Preda. A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. *Science*, 292(5516):472–475, 2001. URL: <http://www.sciencemag.org/content/292/5516/472.abstract>, arXiv:<http://www.sciencemag.org/content/292/5516/472.full.pdf>, doi:10.1126/science.1057726.
- [95] David Griffiths. *Introduction to Quantum Mechanics*. Prentice Hall, 1994.
- [96] Satoshi Morita and Hidetoshi Nishimori. Mathematical foundation of quantum annealing. *Journal of Mathematical Physics*, 49(12):–, 2008. URL: <http://scitation.aip.org/content/aip/journal/jmp/49/12/10.1063/1.2995837>, doi:<http://dx.doi.org/10.1063/1.2995837>.
- [97] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution. 2000. arXiv:quant-ph/001106.
- [98] M. H. S. Amin, Peter J. Love, and C. J. S. Truncik. Thermally assisted adiabatic quantum computation. *Phys. Rev. Lett.*, 100:060503, Feb 2008. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.100.060503>, doi:10.1103/PhysRevLett.100.060503.
- [99] M. H. S. Amin, Dmitri V. Averin, and James A. Nesteroff. Decoherence in adiabatic quantum computation. *Phys. Rev. A*, 79:022107, Feb 2009. URL: <http://link.aps.org/doi/10.1103/PhysRevA.79.022107>, doi:10.1103/PhysRevA.79.022107.

- [100] Andrew M. Childs, Edward Farhi, and John Preskill. Robustness of adiabatic quantum computation. *Phys. Rev. A*, 65:012322, Dec 2001. URL: <http://link.aps.org/doi/10.1103/PhysRevA.65.012322>, doi:10.1103/PhysRevA.65.012322.
- [101] Tameem Albash, Sergio Boixo, Daniel A Lidar, and Paolo Zanardi. Quantum adiabatic markovian master equations. *New Journal of Physics*, 14(12):123016, 2012. URL: <http://stacks.iop.org/1367-2630/14/i=12/a=123016>.
- [102] N. G. Dickson, M. W. Johnson, M. H. Amin, R. Harris, F. Altomare, A. J. Berkley, P. Bunyk, J. Cai, E. M. Chapple, P. Chavez, F. Cioata, T. Cirip, P. deBuen, M. Drew-Brook, C. Enderud, S. Gildert, F. Hamze, J. P. Hilton, E. Hoskinson, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Lanting, T. Mahon, R. Neufeld, T. Oh, I. Perminov, C. Petroff, A. Przybysz, C. Rich, P. Spear, A. Tcaciuc, M. C. Thom, E. Tolkacheva, S. Uchaikin, J. Wang, A. B. Wilson, Z. Merali, and G. Rose. Thermally assisted quantum annealing of a 16-qubit problem. *Nat Commun*, 4:1903, May 2013. Article. URL: <http://dx.doi.org/10.1038/ncomms2920>.
- [103] Kristen L. Pudenz and Daniel A. Lidar. Quantum adiabatic machine learning. *Quantum Information Processing*, 12(5):2027–2070, 2013. URL: <http://dx.doi.org/10.1007/s11128-012-0506-4>, doi:10.1007/s11128-012-0506-4.
- [104] Inc. DWave Systems. <http://dwavesys.com>.
- [105] Vicky Choi. Minor-embedding in adiabatic quantum computation: I. the parameter setting problem. *Quantum Information Processing*, 7(5):193–209, October 2008. URL: <http://dx.doi.org/10.1007/s11128-008-0082-9>, doi:10.1007/s11128-008-0082-9.
- [106] A. Perdomo-Ortiz, N. Dickson, M. Drew-Brook, G. Rose, and A. Aspuru-Guzik. Finding low-energy conformations of lattice protein models

- by quantum annealing. *Sci. Rep.*, 2, 2012.
doi:<http://dx.doi.org/10.1038/srep00571>.
- [107] Christine Klymko, Blair D. Sullivan, and Travis S. Humble. Adiabatic quantum programming: minor embedding with hard faults. *Quantum Information Processing*, 13(3):709–729, 2014. URL:
<http://dx.doi.org/10.1007/s11128-013-0683-9>,
doi:10.1007/s11128-013-0683-9.
- [108] R. Harris, M. W. Johnson, T. Lanting, A. J. Berkley, J. Johansson, P. Bunyk, E. Tolkacheva, E. Ladizinsky, N. Ladizinsky, T. Oh, F. Cioata, I. Perminov, P. Spear, C. Enderud, C. Rich, S. Uchaikin, M. C. Thom, E. M. Chapple, J. Wang, B. Wilson, M. H. S. Amin, N. Dickson, K. Karimi, B. Macready, C. J. S. Truncik, and G. Rose. Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor. *Phys. Rev. B*, 82:024511, Jul 2010. URL:
<http://link.aps.org/doi/10.1103/PhysRevB.82.024511>,
doi:10.1103/PhysRevB.82.024511.
- [109] Vicky Choi. Minor-embedding in adiabatic quantum computation: II. minor universal graph design. *Quantum Information Processing*, 10:343–353, 2011.
- [110] T. Lanting, J. Przybysz, A. Yu. Smirnov, A. M. Spedalieri, F. H. Amin, M. J. Berkley, A. R. Harris, F. Altomare, S. Boixo, P. Bunyk, N. Dickson, C. Enderud, P. Hilton, J. E. Hoskinson, W. Johnson, M. E. Ladizinsky, N. Ladizinsky, R. Neufeld, T. Oh, I. Perminov, C. Rich, C. Thom, M. E. Tolkacheva, S. Uchaikin, B. Wilson, A. and G. Rose. Entanglement in a quantum annealing processor. *Phys. Rev. X*, 4:021041, May 2014. URL:
<http://link.aps.org/doi/10.1103/PhysRevX.4.021041>,
doi:10.1103/PhysRevX.4.021041.
- [111] Kamran Karimi, Neil G. Dickson, Firas Hamze, M.H.S. Amin, Marshall Drew-Brook and Fabian A. Chudak, Paul I. Bunyk, William G. Macready, and

Geordie Rose. Investigating the performance of an adiabatic quantum optimization processor. *Quant. Inf. Proc.*, 11(1):77, 2012.

- [112] Otfried Gühne and Gza Tth. Entanglement detection. *Physics Reports*, 474(16):1 – 75, 2009. URL: <http://www.sciencedirect.com/science/article/pii/S0370157309000623>, doi:<http://dx.doi.org/10.1016/j.physrep.2009.02.004>.
- [113] Sergio Boixo, Tameem Albash, Federico M. Spedalieri, Nicholas Chancellor, and Daniel A. Lidar. Experimental signature of programmable quantum annealing. *Nat Commun*, 4, Jun 2013. Article. URL: <http://dx.doi.org/10.1038/ncomms3067>.
- [114] Sergio Boixo, Troels F. Ronnow, Sergei V. Isakov, Zhihui Wang, David Wecker, Daniel A. Lidar, John M. Martinis, and Matthias Troyer. Evidence for quantum annealing with more than one hundred qubits. *Nat Phys*, 10(3):218–224, Mar 2014. Article. URL: <http://dx.doi.org/10.1038/nphys2900>.
- [115] Kristen L. Pudenz, Tameem Albash, and Daniel A. Lidar. Error-corrected quantum annealing with hundreds of qubits. *Nat Commun*, 5, Feb 2014. Article. URL: <http://dx.doi.org/10.1038/ncomms4243>.
- [116] Seung W. Shin, Graeme Smith, John A. Smolin, and Umesh Vazirani. How quantum is the d-wave machine. 2014. [arXiv:quant-ph/1401.7087](https://arxiv.org/abs/quant-ph/1401.7087).
- [117] Tameem Albash, Walter Vinci, Anurag Mishra, Paul A. Warburton, and Daniel A. Lidar. Consistency tests of classical and quantum models for a quantum annealer. *Phys. Rev. A*, 91:042314, Apr 2015. URL: <http://link.aps.org/doi/10.1103/PhysRevA.91.042314>, doi:10.1103/PhysRevA.91.042314.
- [118] Troels F. Rnnow, Zhihui Wang, Joshua Job, Sergio Boixo, Sergei V. Isakov, David Wecker, John M. Martinis, Daniel A. Lidar, and Matthias Troyer.

- Defining and detecting quantum speedup. *Science*, 345(6195):420–424, 2014. URL: <http://www.sciencemag.org/content/345/6195/420.abstract>, arXiv:<http://www.sciencemag.org/content/345/6195/420.full.pdf>, doi:10.1126/science.1252319.
- [119] B.D. Josephson. Possible new effects in superconductive tunnelling. *Physics Letters*, 1(7):251 – 253, 1962. URL: <http://www.sciencedirect.com/science/article/pii/0031916362913690>, doi:[http://dx.doi.org/10.1016/0031-9163\(62\)91369-0](http://dx.doi.org/10.1016/0031-9163(62)91369-0).
- [120] B. D. Josephson. The discovery of tunnelling supercurrents. *Rev. Mod. Phys.*, 46:251–254, Apr 1974. URL: <http://link.aps.org/doi/10.1103/RevModPhys.46.251>, doi:10.1103/RevModPhys.46.251.
- [121] R. Harris, J. Johansson, A. J. Berkley, M. W. Johnson, T. Lanting, Siyuan Han, P. Bunyk, E. Ladizinsky, T. Oh, I. Perminov, E. Tolkacheva, S. Uchaikin, E. M. Chapple, C. Enderud, C. Rich, M. Thom, J. Wang, B. Wilson, and G. Rose. Experimental demonstration of a robust and scalable flux qubit. *Phys. Rev. B*, 81:134510, Apr 2010. URL: <http://link.aps.org/doi/10.1103/PhysRevB.81.134510>, doi:10.1103/PhysRevB.81.134510.
- [122] Tanju Gleisberg, Stefan Hche, Frank Krauss, Andreas Schlicke, Steffen Schumann, and Jan-Christopher Winter. Sherpa 1., a proof-of-concept version. *Journal of High Energy Physics*, 2004(02):056, 2004. URL: <http://stacks.iop.org/1126-6708/2004/i=02/a=056>.
- [123] WarrenS. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943. URL: <http://dx.doi.org/10.1007/BF02478259>, doi:10.1007/BF02478259.

- [124] Eric B Baum. On the capabilities of multilayer perceptrons. *Journal of Complexity*, 4(3):193 – 215, 1988. URL: <http://www.sciencedirect.com/science/article/pii/0885064X88900209>, doi:[http://dx.doi.org/10.1016/0885-064X\(88\)90020-9](http://dx.doi.org/10.1016/0885-064X(88)90020-9).
- [125] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359 – 366, 1989. URL: <http://www.sciencedirect.com/science/article/pii/0893608089900208>, doi:[http://dx.doi.org/10.1016/0893-6080\(89\)90020-8](http://dx.doi.org/10.1016/0893-6080(89)90020-8).