Search for a Higgs-like boson decaying to two photons in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV

Guo Chen

Submitted to the graduate degree program in Department of Physics and Astronomy and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Committee members

Dr. Graham Wilson, Chairperson

Dr. Philip Baringer

Dr. Dave Besson

Dr. Kyoungchul Kong

Dr. Brian Laird

Date defended: March 1st, 2013

The Dissertation Committee for Guo Chen certifies
that this is the approved version of the following dissertation :

Search for a Higgs-like boson decaying to two photons in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV

Dr. Graham Wilson, Chairperson

Date approved: April 15th, 2013

Abstract

We present a search for the Standard Model (SM) Higgs boson decaying to two photons using the 9.6 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV collected by the D0 detector at the Tevatron from July 2002 to September 2011. The data corresponds to the complete data-set in the RunII period of the Tevatron. This thesis employs a multivariate technique, boosted decision trees, to optimize the discrimination of signal from background. No significant excess of data above theoretical prediction is observed, so limits on the ratio of the production cross section times the branching ratio to the SM value are set at 95% confidence level as a function of hypothetical Higgs boson mass. The expected and observed limits on this ratio are 8.7 and 12.8 respectively at 125 GeV. We also interpret the data in the fermiophobic Higgs model. The result has reached a sensitivity of 114 GeV, and we set a lower limit on the fermiophobic Higgs mass of $M_{H_f} > 113$ GeV.

The work described in this thesis directly results in the following papers:

Phys. Rev. Lett. **107**, 151801 (2011)

arXiv:1301.5358 [hep-ex], accepted by Phys. Rev. D (2013)

Acknowledgements

This thesis is the culmination of years of my dedication. In retrospect, there are so many great people to thank for. Because without them, I alone would not achieve this far and complete this thesis. This section is specially for those people.

First, I would like to thank my parents, Xinjian Chen and Linbin Zhou, and my brother Chuan Chen. Even though they cannot read English much and have almost no idea what I am working on, their virtues of sincerity, kindness and hard work have great influence on me and they have been my golden examples.They always encourage me to do what I am interested in, though I sometimes do well while sometimes do badly. It is quite easy to find flowers and applause when one is successful, but it is my parents who always stay alongside me when I am in valleys of life. This thesis is absolutely dedicated to them.

Next, I would like to thank cordially to Dr. Graham Wilson and Dr. Aurelio Juste. When I first stepped into the field of high energy physics, I found myself too ignorant to code out even “hello world”. Graham is quite kind, smart and knowledgeable and he has been guiding me patiently. I can never ever thank more to Graham for his guidance, kindness and continuous support over these years. And I wish he would never remember that he lent me the book *Introduction to Elementary Particles, David Griffiths* years ago but I never returned. Aurelio acts as my local advisor at Fermilab, even though he is not really local. At most of the time, we talk through video conference and exchange emails when he is in Spain. He is extraordinarily talented and always has many brilliant ideas to bet with me. (Un)fortunately in my impression he wins every time. He is also a very kind advisor. Without him, I believe I cannot reach this

far. It is my foremost honor to collaborate heavily with him, and hereby I would like thank him sincerely for his guidance as well. In the research on Higgs, I certainly owe a million thanks to Dr. Xuebing Bu and Dr. Zhenyu Yu. They are smart, hard working and they solve problems rapidly and effectively. I must admit that I have learned a lot from them, including working style. Heartiest thanks must also go to the editorial board 13 chaired by Dr. Elemer Nagy, as well as former and current Higgs conveners for their helpful suggestions and careful scrutiny.

Before I really started research on Higgs bosons, I worked heavily with Dr. Junjie Zhu, Dr. Mika Vesterinen and Dr. Terry Wyatt on Z boson transverse momentum measurement. A large thank you must fly to them for helping me to speed up data analysis and software ability.

Besides research, I have worked as a CFT expert and so many thanks to CFT and CTT experts for their kind education and listening to my studies. They are Dr. George Ginther, Dr. Jadwiga Warchol, Dr. Stefan Gruenendahl and Dr. Paul Rubinov.

Happy thanks go to those dear friends who make my life at Fermilab colorful: Dr. Yunhe Xie, Dr. Liang Li, Dr. Hang Yin, Dr. Weigang Geng, Jiaming Yu, Siqi Yang, Peng Jiang, Pengfei Ding, Dikai Li, and Vivek Parihar.

At last but not at the least, I am beyond grateful to the help and support from those in the HEP group of the university of Kansas, including Dr. Philip Baringer, Dr. Alice Bean, Dr. Dave Besson and Dr. Jadranka Sekaric. Also special thanks to my dear friends Dr. Guowei Xu, Dr. Justace Clutter, Dr. Mihailo Backovic, Dr. Nataly Ozak and Priyanka Saxena. They made my life in the university of Kansas exciting and unforgetful.

Though I cannot print out all the names, it is worthy noting that many great people helped me in different ways along this journey. Thank them all here.

Contents

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**

1.1 Standard Model 9

1.2 Higgs Mechanism 10

1.3 Higgs Boson Production and Decay Modes 12

1.4 Fermiophobic model 16

2 Accelerator and Detector **18**

2.1 Accelerator System 18

2.2 DØ Detector 20

2.2.1 Coordinates, Kinematic Quantities and Units 20

2.2.2 Central Tracking System 21

2.2.3 Preshower System 25

2.2.4 Calorimeter 28

2.2.5 Muon System 29

2.2.6 Luminosity Monitor 30

2.2.7 Trigger System 31

3 Object Reconstruction and Identification **33**

3.1 Tracks 34

3.2 Primary Vertices 35

3.3 Photons 35

3.3.1 Photon Reconstruction 35

3.3.2 Photon Identification 42

3.3.3 Photon Energy Scale and Resolution 44

3.4 Jets 45

3.5 Missing Transverse Energy 46

4 Samples **47**

4.1 Data Sample 47

4.2 Monte Carlo Simulation 48

5 Analysis **50**

5.1 Revertexing 51

5.1.1 Algorithm 53

5.1.2 Calibration 55

5.2 Event Selection 65

5.3 Background Modeling 68

5.3.1 Corrections 68

5.3.2 $Z/\gamma^* \to ee$ Drell-Yan (ZDY) Background 70

5.3.3 γ+jet and Di-jet Backgrounds 71

5.3.4 Direct Di-photon Production (DDP) 74

5.3.5 Diphoton Invariant Mass Distribution 76

5.4 Multivariate Analysis 79

5.4.1 Decision Tree and Boosting 79

5.4.2 Input Variables and Training 82

5.4.3 Final Discriminant 97

5.5 Systematic Uncertainties 100

6 Result **103**

6.1 Limit Calculation Method 103

6.2 SM Higgs Boson Search Result 108

6.3 Fermiophobic Interpretation 112

7 Conclusion **116**

References **117**

A Drell-Yan $Z/\gamma^* \to ee$ background **125**

B SM systematic uncertainties: ε_γ and ε_j **127**

C Diphoton Invariant Mass Method **128**

List of Figures

1.1 (a) The indirect and direct constraints on the Higgs boson mass are shown with dashed and solid 68% CL contour. (b) This figure "shows the $\Delta\chi^2$ curve derived from high-Q^2 precision electroweak measurements, performed at LEP and by SLD, CDF, and D0, as a function of the Higgs-boson mass, assuming the Standard Model to be the correct theory of nature. The preferred value for its mass, corresponding to the minimum of the curve, is at 94 GeV, with an experimental uncertainty of +29 and -24 GeV (at 68 percent confidence level derived from $\Delta\chi^2 = 1$ for the black line, thus not taking the theoretical uncertainty shown as the blue band into account). The precision electroweak measurements tell us that the mass of the Standard-Model Higgs boson is lower than about 152 GeV (one-sided 95 percent confidence level upper limit derived from $\Delta\chi^2 = 2.7$ for the blue band, thus including both the experimental and the theoretical uncertainty). This limit increases to 171 GeV when including the LEP-2 direct search limit of 114 GeV shown in yellow." [5] . 3

1.2 The p-value as a function of m_H under the background-only hypothesis. Also shown are the median expected values assuming a SM signal is present, evaluated separately at each m_H. The associated dark and light-shaded bands indicate the 1 s.d. and 2 s.d. fluctuations of possible experimental outcomes [7]. 4

1.3 Combined search results: (a) The observed (solid) 95% CL upper limit on the signal strength as a function of m_H and the expectation (dashed) under the background-only hypothesis. The dark and light shaded bands show the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties on the background-only expectation. (b) The observed (solid) local p_0 as a function of m_H and the expectation (dashed) for a SM Higgs boson signal hypothesis ($\mu = 1$) at the given mass. (c) The best-fit signal strength $\hat{\mu}$ as a function of m_H. The band indicates the approximate 68% CL interval around the fitted value [8]. 5

1.4 (a) The CL_S values for the SM Higgs boson hypothesis as a function of the Higgs boson mass in the range 110–145 GeV. The background-only expectations are represented by their median (dashed line) and by the 68% and 95% CL bands. (b) The observed local p-value for 7 TeV and 8 TeV data, and their combination as a function of the SM Higgs boson mass. The dashed line shows the expected local p-values for a SM Higgs boson with a mass m_H [9]. 6

1.5 The observed local p-value for the five decay modes and the overall combination as a function of the SM Higgs boson mass. The dashed line shows the expected local p-values for a SM Higgs boson with a mass m_H [9]. 7

1.6 Signal strength (σ/σ_{SM}) for the combination and individual decay modes in (a) ATLAS and (b) CMS [8, 9]. 8

1.7 The elementary particles of the Standard Model [12]. 10

1.8 The potential of the scalar field ϕ in the case of $\mu^2 < 0$, $\lambda > 0$ [13]. 11

1.9 Feynman diagrams of (a) gluon fusion, (b) vector boson fusion, (c) associated production with a weak boson and (d) associated production with a top quark pair [14]. 13

1.10 Higgs production cross sections at the Tevatron ($\sqrt{s} = 1.96$ TeV) and LHC ($\sqrt{s} = 7$ TeV) [15]. 14

1.11 SM Higgs boson decay modes and branching ratios [16]. 15

1.12 Feynman diagrams of the SM Higgs boson decaying into a photon pair through top quark and W boson loops [17]. 15

1.13 Branching ratios for the SM Higgs boson and the three scenarios in Ref. [10]. H_{SM} is the standard model Higgs boson. H_{u} is a Higgs boson with Yukawa couplings only with up-type fermions. H_{ew} is a Higgs boson that couples to all particles except the top and bottom quark. H_{f} is a Higgs boson with no tree-level coupling to fermions [24]. 16

2.1 Accelerator Overview [41]. 19

2.2 The upgraded DØ Detector [43]. 22

2.3 DØ Tracking System [43]. 23

2.4 DØ Silicon Detector [43]. 24

2.5 A curved ribbon consisting of two layers of scintillating fibers [43]. 25

2.6 A cross section view of the routing of the clear fiber waveguides on the south face of the central cryostat [43]. 25

2.7 An example fit to the LED spectrum from an axial CFT fiber. The solid histogram is the data; the smooth curve is the fit [43]. 26

2.8 Cross section and geometrical layout of the CPS and FPS scintillator strips [43]. 27

2.9 Complete ϕ-segment of a FPS module [43]. 27

2.10 DØ Uranium/Liquid-argon Calorimeter [49]. 29

2.11 Liquid argon gap and signal board unit cell for the calorimeter [43]. 29

2.12 DØ Uranium/Liquid-argon Calorimeter showing segmentation in η and depth [43]. 30

2.13 A schematic view of the location of the LM detectors [43]. 30

2.14 A schematic view of the trigger system of the DØ detector [43]. 31

3.1 A schematic diagram of typical particle interaction locations in the DØ detector [52]. . . . 33

3.2 A schematic view of the isolation definition [24]. 36

3.3 Normalized O_{NN} input variables for photons and jets [24]. 39

3.4 The three-body (di-lepton+γ) mass versus the two-body (di-lepton) mass from $Z \to \ell^+\ell^-\gamma$ ($\ell = e, \mu$) data events [60]. To pick up the pure final state radiation (FSR) γ candidates, the two-body mass is required to be less than 82 GeV while simultaneously the three-body mass is required to be within 82 - 102 GeV. 40
3.5 Normalized distributions of O_{NN} from real and fake photons. 41
3.6 Illustration of the Hits-on-the-road algorithm [61]. 41

4.1 The delivered and recorded integrated luminosity as a function of time at the DØ detector [73]. 48

5.1 The efficiencies for finding the true hard scatter vertex by using the default DØ reconstruction algorithm (d0reco) and max-track algorithm (mntrk) as a function of diphoton transverse momentum in the gluon fusion signal MC sample. They are shown as the black and red lines respectively in the plot. The blue line illustrates the frequency with which the two algorithms yield the same vertex or close enough. 52
5.2 A schematic illustration of the photon pointing in the $z-y$ plane (left) and $x-y$ plane (right) [91]. 53
5.3 Track multiplicity distribution of the hard-scatter vertex in p20 (upper) and p17 (lower) data (black), uncorrected MC (red) and corrected MC (magenta) using $Z/\gamma^* \to ee$ events. . 57
5.4 Track multiplicity distribution of the min-bias vertices in p20 (upper) and p17 (lower) data (black), uncorrected MC (red) and corrected MC (magenta) using $Z/\gamma^* \to ee$ events. . . . 58
5.5 The probability of finding the true vertex using calibrated max-track vertexing as a function of di-EM p_T for p20 (upper) p17 (lower) data (black) and MC (red) samples using $Z/\gamma^* \to e^+e^-$ events. 59
5.6 Photon pointing resolution fit in an example rapidity bin $|\eta| < 0.4$ for p20 data (upper) and MC (lower) samples using $Z/\gamma^* \to ee$ events. 60
5.7 Photon pointing resolution after calibration in p20 (upper) and p17 (lower) data (black), uncorrected MC (red) and corrected MC (magenta) samples using $Z/\gamma^* \to e^+e^-$ events. . . 61

5.8 Photon pointing significance fit in an example rapidity bin $|\eta| < 0.4$ for p20 data (upper) and MC (lower) samples using $Z/\gamma^* \to ee$ events. 62

5.9 Photon pointing significance after calibration in an example rapidity bin $|\eta| < 0.4$ for p20 data (upper) and MC (lower) samples using $Z/\gamma^* \to e^+e^-$ events. 63

5.10 The final calibrated algorithm's efficiency of finding the true vertex as a function of di-EM p_T for p20 (upper) p17 (lower) data (black) and MC (red) samples using $Z/\gamma^* \to e^+e^-$ events. 64

5.11 v15 and v16 calorimeter-only di-EM trigger efficiency as a function of di-EM invariant mass, where the two red dotted lines illustrate the assigned 0.1% uncertainty. 66

5.12 The photon O_{NN} distributions in different data taking epochs are compared in different η_{det} bins. 69

5.13 UET distribution after correction to MC simulations. 70

5.14 Validation of the shape of the $M_{\gamma\gamma}$ distribution in the CCCC region. The $\gamma\gamma$ background from the SHERPA MC and the γ+jet and jet+jet backgrounds from the orthogonal sample are compared to their corresponding 4x4 matrix method solutions. 73

5.15 Feynman diagrams that contribute to the direct diphoton production. Solid lines are quarks and antiquarks. Wavy and curly lines represent photons and gluons. (a)-(e) and (h)-(l) are considered as leading order and next-to-leading order processes respectively in Ref. [97]. The diagrams of (f) and (g) are single and double fragmentation. 74

5.16 An example of the measured differential cross section in direct diphoton production as a function of $M_{\gamma\gamma}$, $p_T^{\gamma\gamma}$, $\Delta\phi_{\gamma\gamma}$ and $|\cos\theta^*|$ ($\cos\theta^* = \tanh[(\eta_1 - \eta_2)/2]$). The data are compared to the theoretical predictions from RESBOS, DIPHOX, PYTHIA AND SHERPA. The ratio of the differential cross sections between data and SHERPA are displayed as black points with uncertainties in the bottom plots [98]. 75

5.17 Diphoton invariant mass distribution of signal, background and data. A hypothetical standard model signal at 125 GeV is shown. The signal yield is scaled by a factor of 100 for better visualization. 76

5.18 (a) Reconstructed diphoton invariant mass distribution for a SM Higgs boson signal MC sample with $M_H = 125$ GeV. The red line represents the fit to the distribution described in the text. (b) Mass resolution as a function of diphoton invariant mass from the fit to SM Higgs boson signal MC samples. 78

5.19 Collins-Soper frame illustration: P_1, P_2 are the proton and anti-proton momenta. l, l' are the photon momenta. $\hat{z}$ is a unit vector along the bisector of the proton beam momentum and the negative of the anti-proton beam momentum when they are boosted into the center-of-mass frame of the diphoton pair. The white plane is the diphoton plane; the gray plane is the proton-anti-proton plane. $\hat{h}$ is a unit vector in the proton-anti-proton plane but transverse to $\hat{z}$. θ and ϕ are the θ^* and ϕ^* in our notation. 84

5.20 $M_{\gamma\gamma}$ distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible. 86

5.21 p_T^1 distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible. 87

5.22 p_T^2 distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible. 88

5.23 $p_T^{\gamma\gamma}$ distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible. 89

5.24 $\Delta\phi_{\gamma\gamma}$ distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible. 90

5.25 $\cos\theta^*$ distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible. 91

5.26 ϕ^* distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible. 92

5.27 $\not{E}_T$ distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible. 93

5.28 O_{NN}^1 distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible. 94

5.29 O^2_{NN} distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible. 95

5.30 An illustration of the over-training check for a 125 GeV SM Higgs boson. The Kolmogorov-Smirnov test indicates the consistency of the distributions from the training sample and testing sample. 96

5.31 Final discriminants for 125 GeV signal with a linear scale. 98

5.32 Final discriminants for 125 GeV signal with a log scale. 99

6.1 Example distributions of log-likelihood ratio (LLR) under $S+B$ and B-only hypotheses. The LLR value from the observed data is shown as a vertical dashed line, separating the LLR distributions into two areas for the two hypotheses each. The red shaded area is CL_{S+B} and also the p-value of the test hypothesis. The blue shaded area is 1-CL_B, representing the p-value of the null hypothesis [108]. 106

6.2 Results for the SM Higgs search from using the joint BDT and sideband as the final discriminant. (a) Limits at 95% CL on the ratio of $\sigma \times \mathscr{B}(H \to \gamma\gamma)$ to the SM prediction as a function of Higgs mass. The observed limit is shown as a solid black line while the expected limit under the background-only hypothesis is shown as a dashed red line. The green and yellow areas correspond to 1 and 2 standard deviations (s.d.) around the expected limit for the background-only hypothesis. (b) Log-likelihood ratios (LLR) as a function of Higgs mass. 109

6.3 Data after subtraction of the post-fit background at $M_H = 125$ GeV in (a) the photon-enriched sample and (b) the jet-enriched sample. The expected SM Higgs signal is normalized to the corresponding observed limit in Table 6.1. The bands represent the 1 s.d. uncertainties from the background-only hypothesis fit. 110

6.4 Final discriminants for the fermiophobic signal at 125 GeV in linear scale. 113

6.5 Final discriminants for the fermiophobic signal at 125 GeV in log scale. 114

6.6 Results for the fermiophobic Higgs search. (a) Upper limits at 95% CL on the ratio of $\sigma \times \mathscr{B}(H_f \to \gamma\gamma)$ to the fermiophobic Higgs model prediction as a function of Higgs mass. The observed limit is shown as a solid black line while the expected limit under the background-only hypothesis is shown as a dashed red line. The green and yellow areas correspond to 1 and 2 standard deviations (s.d.) around the expected limits under the background-only hypothesis. (b) Upper limits at 95% CL on the value of $\mathscr{B}(H_f \to \gamma\gamma)$. The theoretical prediction from the fermiophobic Higgs model is shown as a blue dashed line. 115

A.1 The Z/γ^* invariant mass distribution by requiring that both electrons have tracks. 126

A.2 The Z/γ^* invariant mass distribution by requiring that one electron has a track and one electron does not. It is fit with an exponentially decaying function to model the smooth background, and with a convolution of a Gaussian function and a Breit-Wigner function to model the Z boson signal. In the mass window of $[85, 100]$ GeV, the fit yields 4005.7 Z bosons and the Monte Carlo yields 3755.1 Z bosons. 126

A.3 The Z/γ^* invariant mass distribution by requiring that both electrons have no track. 126

C.1 Results for the SM Higgs boson search from using the diphoton invariant mass as the final discriminant. (a) Log-likelihood ratio as a function of Higgs mass. The observed limit is shown as a solid black line while the expected limit under the background-only hypothesis is shown as a dashed red line. The green and yellow areas correspond to 1 and 2 standard deviations (s.d.) around the median LLR from the background-only hypothesis. (b) Limits at 95% CL on $\sigma \times \mathscr{B}(H \to \gamma\gamma)$ relative to the SM prediction as a function of Higgs mass. . 129

List of Tables

1.1 Cross sections (pb) and branching ratios in the standard model and fermiophobic model for the signal samples [25]. 17

3.1 p20 (p17) core cuts definitions for CC ($-1.1 < \eta_{det} < 1.1$) and EC ($1.5 < |\eta_{det}| < 2.5$) regions. For p20, eq1 = $7.3\eta_{det}^2$ - $35.9|\eta_{det}|$ + 45.7, and eq2 = $7.5\eta_{det}^2$ - $36.0|\eta_{det}|$ + 44.8. For p17, eq1 = $2.74\eta_{det}^2$ - $16.3|\eta_{det}|$ + 25.0, and eq2 = $5.96\eta_{det}^2$ - $30.6|\eta_{det}|$ + 40.7. 43

5.1 Resolution in z-position for the central preshower detector and the four electromagnetic calorimeter layers for RunII. 51

5.2 Event selection efficiencies from 130 GeV Higgs MC for p17 and p20. 67

5.3 CPS match efficiencies and scale factors measured with Zγ data and MC. 67

5.4 (a) shows the number of events in the data for the 4 categories. (b) shows the number of $\gamma\gamma$, γ+jet and dijet events in the data from the 4x4 matrix method. The number of DY events is estimated from MC. The quoted uncertainties are statistical only. 72

5.5 Systematic uncertainties for different sources. 102

6.1 Expected and observed limits on the ratio of $\sigma \times \mathscr{B}(H \to \gamma\gamma)$ to the SM prediction as a function of Higgs mass using the BDT. 108

6.2 Signal, backgrounds and data yields for the **photon-enriched** sample within the $M_H \pm 30$ GeV mass window, for M_H = 105 GeV to M_H = 145 GeV in 10 GeV intervals. The background yields are from a fit to the data. The uncertainties include both statistical and systematic contributions added in quadrature and take into account correlations among processes. The uncertainty on the total background is smaller than the sum in quadrature of the uncertainties in the individual background sources due to the anti-correlation resulting from the fit. 111

6.3 Signal, backgrounds and data yields for the **jet-enriched** sample within the $M_H \pm 30$ GeV mass window, for M_H = 105 GeV to M_H = 145 GeV in 10 GeV intervals. The background yields are from a fit to the data. The uncertainties include both statistical and systematic contributions added in quadrature and take into account correlations among processes. The uncertainty on the total background is smaller than the sum in quadrature of the uncertainties in the individual background sources due to the anti-correlation resulting from the fit. 111

6.4 Branching ratio comparison for a SM Higgs boson (H) and a fermiophobic Higgs boson (H_f) decaying into two photons. 112

6.5 Expected and observed upper limits at 95% CL on the branching ratio of a fermiophobic Higgs boson decaying into two photons $(H_f \to \gamma\gamma)$ as a function of the mass using the BDT method. 112

B.1 Yields change in % from the 4x4 matrix method in four scenarios of varying ε_γ and ε_j about their uncertainties. N_{jj} and N_{gj} are the yields of jet+jet and γ+jet respectively. 127

C.1 Expected and observed limits on the ratio of $\sigma \times \mathscr{B}(H \to \gamma\gamma)$ to the SM prediction as a function of Higgs mass using the BDT. 128

Chapter 1

Introduction

It is human nature to be curious about the fundamental composition of the universe, matter, and even ourselves-human beings. Particle physics tries to satisfy this curiosity by exploring what the fundamental particles are and explaining how they interact with one another.

Since the late 1970's, the Standard Model (SM) became the cornerstone theory to describe elementary particles and their interactions. The elementary particles include three generations of quarks and leptons, and the mediators of the interactions, the gauge bosons. The interactions are based on the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$, where $SU(3)_C$ is the QCD gauge theory [1] to explain interactions between quarks, and $SU(2)_L \times U(1)_Y$ is the Glashow-Weinberg-Salam electroweak theory [2] to explain electromagnetic and weak interactions of quarks and leptons. The last decades have witnessed the success of the Standard Model through the prediction and discovery of the $W^{\pm}$ and Z weak bosons, charm, bottom and top quarks, and numerous precision measurements that are consistent with SM predictions.

One of the most profound insights that the SM provides is the origin of mass of the fundamental particles. In the mid 1960's, the mechanism of spontaneous electroweak symmetry breaking (EWSB) was proposed by Higgs, Brout, Englert, Guralnik, Hagen and Kibble [3] to generate the three weak boson ($W^{\pm}$ and Z) masses. This led to the prediction of the existence of a massive particle-the Higgs boson. Great efforts have been taken to search for this particle at the CERN

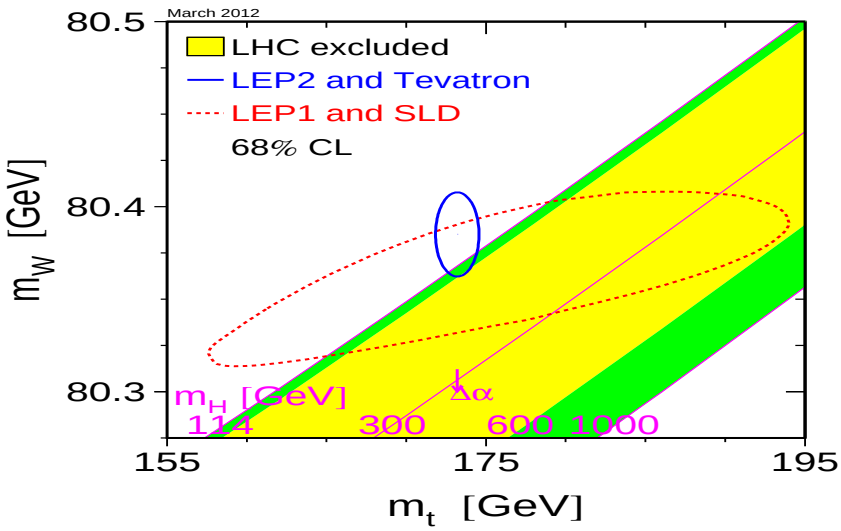
e^+e^- Collider (LEP), the Tevatron and the LHC over the decades, and to improve the constraints on the Higgs boson mass, M_H.

Direct searches at LEP [4] set a lower limit of $M_H > 114.4$ GeV at 95% confidence level (CL); indirect constraints from precision electroweak observables [5] set an upper limit of $M_H < 152$ GeV at 95% CL. The direct and indirect constraints are shown in Figure 1.1.

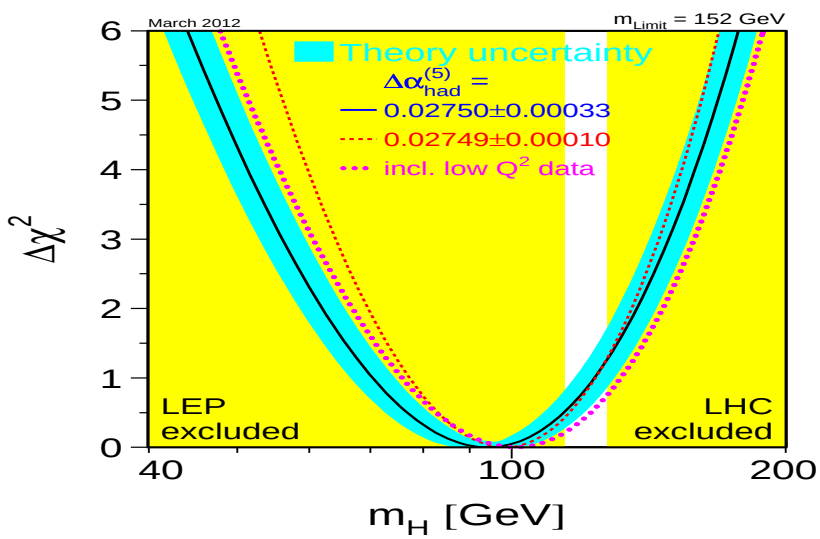
The search for the SM Higgs boson did not have a major breakthrough until the summer of 2012. The combination of searches at the Tevatron [6] excluded the mass ranges of $100 < M_H < 103$ GeV and $147 < M_H < 180$ GeV, and observed an excess with a maximum local significance of 3.1 standard deviations (s.d.) at $M_H = 125$ GeV from the $VH(H \to b\bar{b})$ searches [7], see Figure 1.2; more strikingly, the ATLAS and CMS Collaborations at the LHC observed significant excesses of events in data at $M_H \approx 125$ GeV with local significances of 5.9 s.d. (see Figure 1.3) and 5.0 s.d. (see Figure 1.4) respectively [8, 9].

The $H \to \gamma\gamma$ and $H \to ZZ^*$ decays play significant roles in the above combined searches at the LHC, as shown in Figure 1.5. In particular, $H \to \gamma\gamma$ is sensitive to new physics beyond the SM because of the loop-mediated production mode $gg \to H$ and decay mode $H \to \gamma\gamma$. For instance, alternative models of electroweak symmetry breaking [10] can suppress the coupling of Higgs bosons to fermions, in which a so-called benchmark fermiophobic Higgs boson has no tree-level couplings to fermions at all. This results in a greatly enhanced diphoton branching ratio, $\mathscr{B}(H \to \gamma\gamma)$. Interestingly, the best-fit cross sections to the excess in $H \to \gamma\gamma$ decays at the LHC showed a deviation of about 1.5 s.d. from the theory prediction (Figure 1.6), even though a more detailed global fit shows no significant deviations [11]. Thus, analyzing more data can be beneficial to a more definitive conclusion.

In this chapter, we first present a brief review of the Standard Model, followed by the Higgs mechanism, and then the production and decay modes of the Higgs boson, and finally the fermiophobic model.



(a)



(b)

Figure 1.1: (a) The indirect and direct constraints on the Higgs boson mass are shown with dashed and solid 68% CL contour. (b) This figure "shows the $\Delta\chi^2$ curve derived from high-Q^2 precision electroweak measurements, performed at LEP and by SLD, CDF, and D0, as a function of the Higgs-boson mass, assuming the Standard Model to be the correct theory of nature. The preferred value for its mass, corresponding to the minimum of the curve, is at 94 GeV, with an experimental uncertainty of +29 and -24 GeV (at 68 percent confidence level derived from $\Delta\chi^2$ = 1 for the black line, thus not taking the theoretical uncertainty shown as the blue band into account). The precision electroweak measurements tell us that the mass of the Standard-Model Higgs boson is lower than about 152 GeV (one-sided 95 percent confidence level upper limit derived from $\Delta\chi^2$ = 2.7 for the blue band, thus including both the experimental and the theoretical uncertainty). This limit increases to 171 GeV when including the LEP-2 direct search limit of 114 GeV shown in yellow." [5]

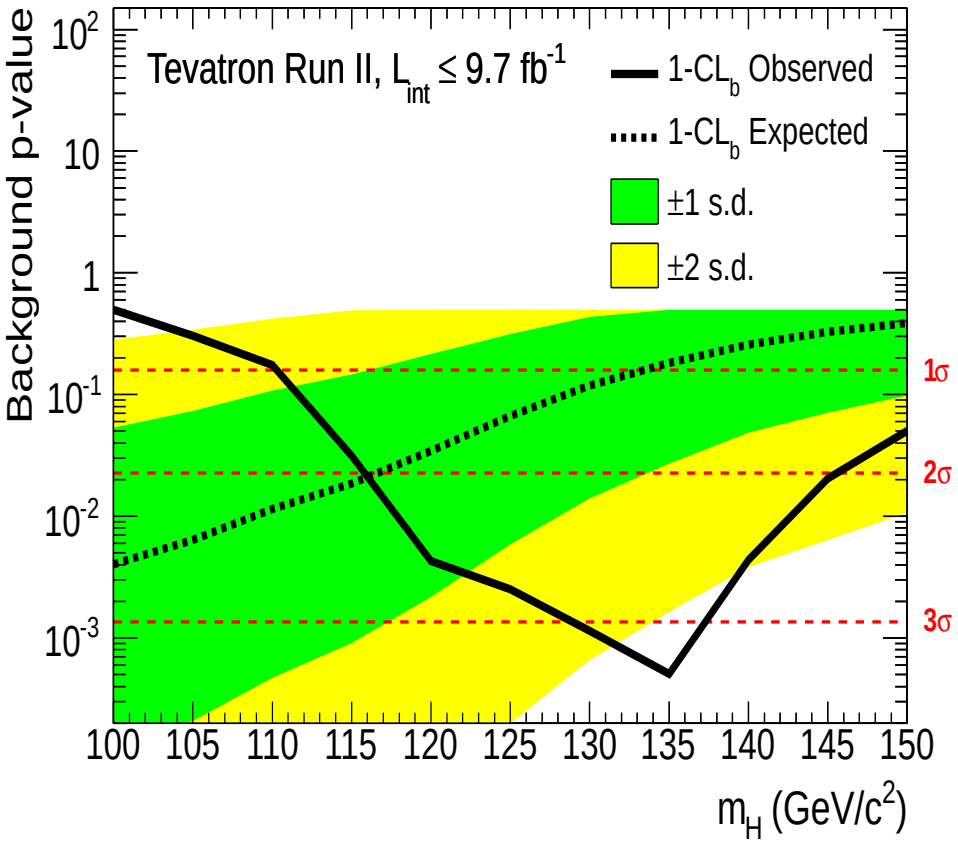


Figure 1.2: The p-value as a function of m_H under the background-only hypothesis. Also shown are the median expected values assuming a SM signal is present, evaluated separately at each m_H. The associated dark and light-shaded bands indicate the 1 s.d. and 2 s.d. fluctuations of possible experimental outcomes [7].

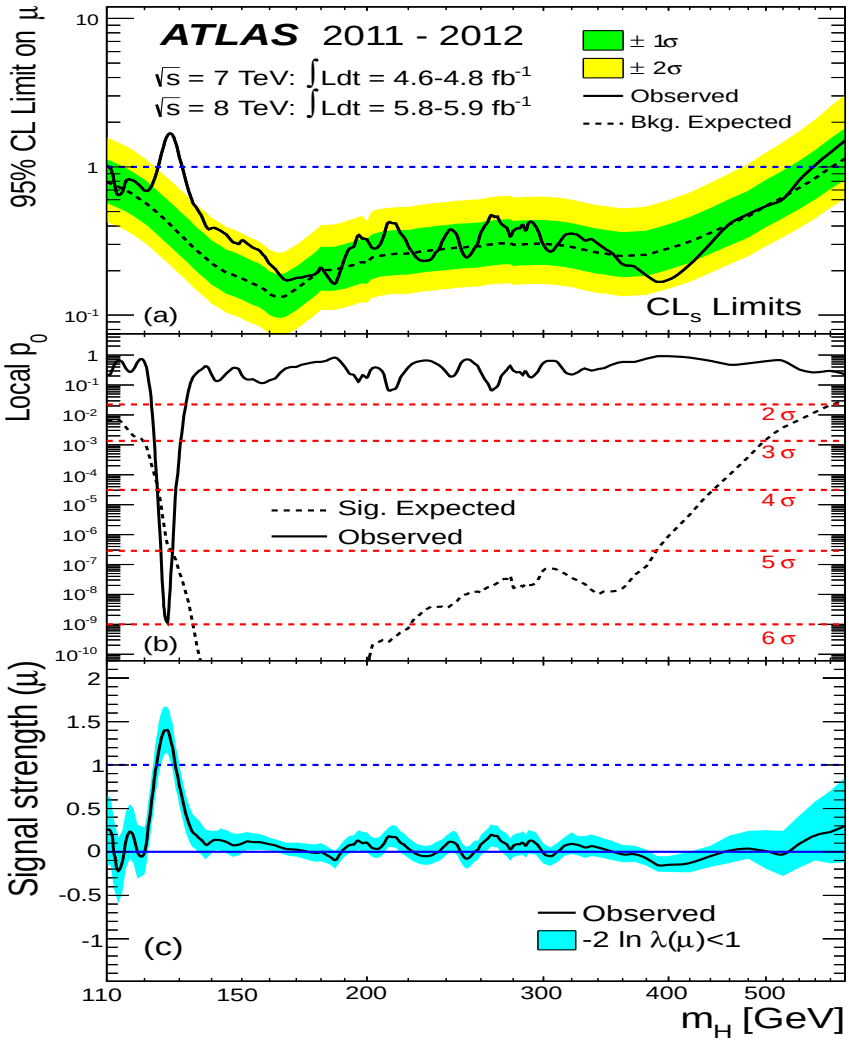
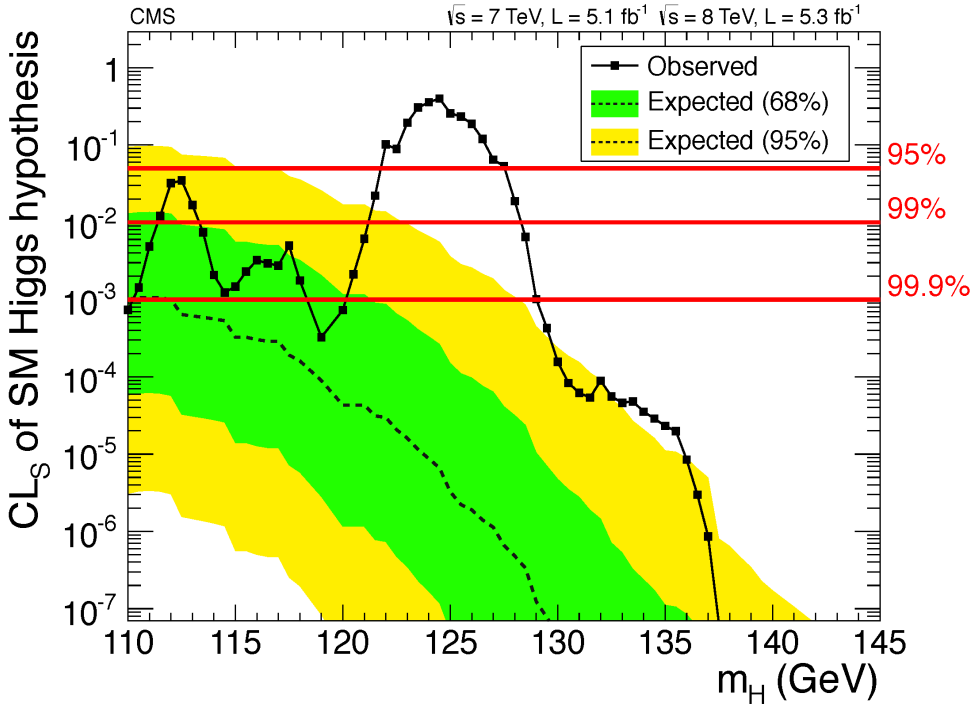
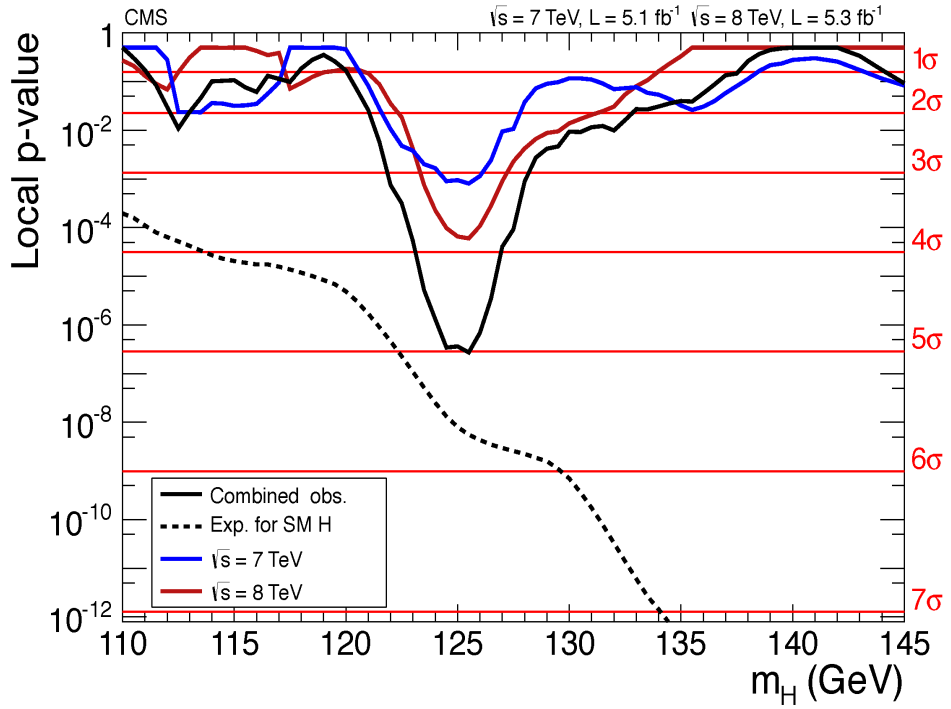


Figure 1.3: Combined search results: (a) The observed (solid) 95% CL upper limit on the signal strength as a function of m_H and the expectation (dashed) under the background-only hypothesis. The dark and light shaded bands show the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties on the background-only expectation. (b) The observed (solid) local p_0 as a function of m_H and the expectation (dashed) for a SM Higgs boson signal hypothesis ($\mu = 1$) at the given mass. (c) The best-fit signal strength $\hat{\mu}$ as a function of m_H. The band indicates the approximate 68% CL interval around the fitted value [8].



(a)



(b)

Figure 1.4: (a) The CL_S values for the SM Higgs boson hypothesis as a function of the Higgs boson mass in the range 110–145 GeV. The background-only expectations are represented by their median (dashed line) and by the 68% and 95% CL bands. (b) The observed local p-value for 7 TeV and 8 TeV data, and their combination as a function of the SM Higgs boson mass. The dashed line shows the expected local p-values for a SM Higgs boson with a mass m_H [9].

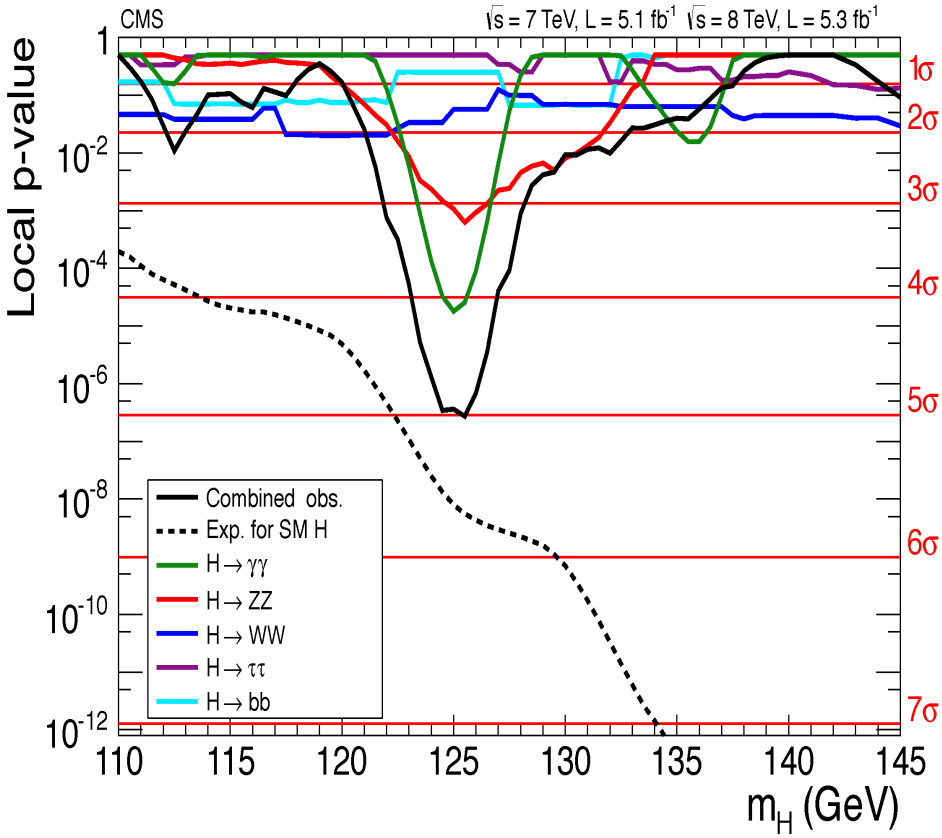
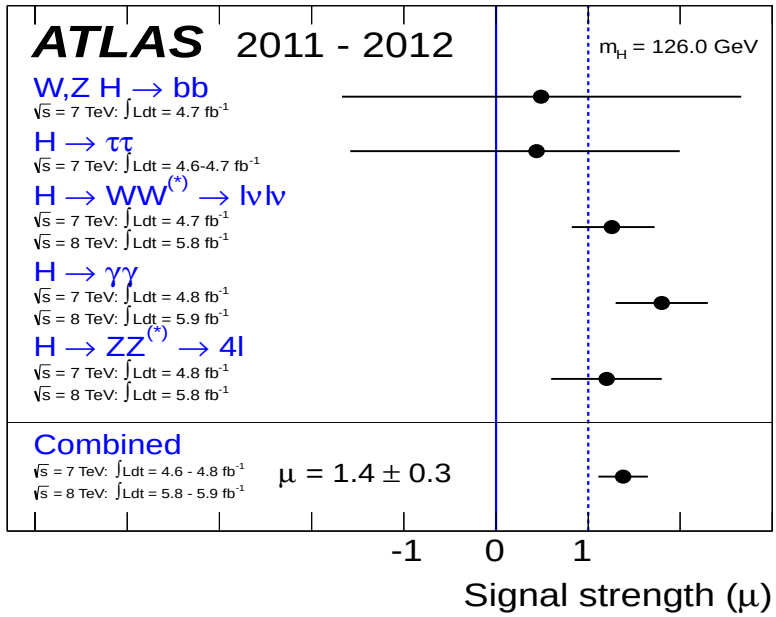
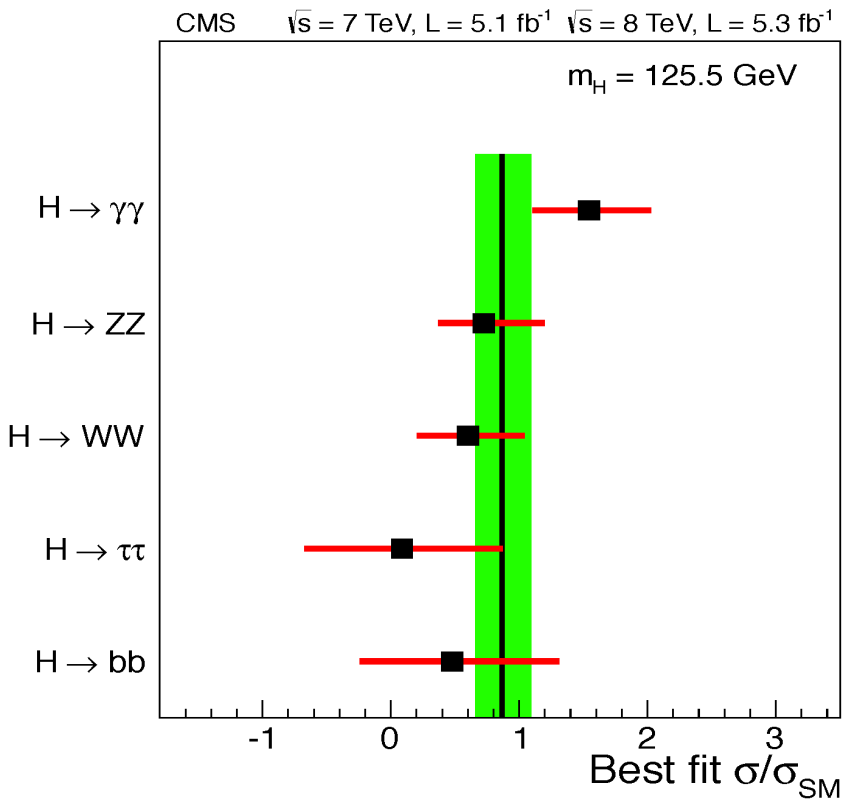


Figure 1.5: The observed local p-value for the five decay modes and the overall combination as a function of the SM Higgs boson mass. The dashed line shows the expected local p-values for a SM Higgs boson with a mass m_H [9].



(a)



(b)

Figure 1.6: Signal strength (σ/σ_{SM}) for the combination and individual decay modes in (a) ATLAS and (b) CMS [8, 9].

1.1 Standard Model

In general, there are two groups of particles: bosons and fermions. Bosons are the integer-spin particles that are characterized by Bose-Einstein statistics. Fermions are the half-integer spin particles that are described by Fermi-Dirac statistics and follow the Pauli exclusion principle.

For the elementary particles in the Standard Model, the fermions consists of 6 quarks and 6 leptons. The six quarks are named as up (u), down (d), charm (c), strange (s), top (t) and bottom (b). The six leptons are the electron (e), muon (μ), tau (τ), electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ). The quarks and leptons are grouped into 3 generations with mass increasing with each generation, as shown in Figure 1.7 [12]. Quarks have fractional electric charge. For instance, the up-type quarks have an electric charge of +2/3 and the down-type quarks possess charge of -1/3. Besides the electric charge, quarks also carry color charge: red, green and blue. In the SM, each elementary particle has a counterpart whose electric charge is opposite; it is called an anti-particle. So taking into account anti-particle and color charge, there are 36 quarks and 12 leptons. Quarks are not allowed to exist alone in the SM due to color confinement. But multiple quarks can constitute a particle, called a hadron. A hadron is called a meson if it is comprised of a quark and an anti-quark. If a hadron is made up of three quarks (anti-quarks), it is called a baryon (anti-baryon).

Quarks and leptons are fundamental particles to construct matter. Their interactions are mediated by the four bosons: gluon, photon, W and Z boson. The strong interaction between quarks is mediated by gluons. The photon, W and Z boson are the "force carriers" in electroweak interactions.

The strong and electroweak interactions can be described by the SM Lagrangian. This Lagrangian keeps local gauge symmetry, *ie*, is invariant under local $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge transformations for fermions and gauge bosons. However, the fermion field and gauge field in this Lagrangian are massless. The mass term cannot be added without violating $SU(2) \times U(1)$ gauge symmetry, which is the most important and fundamental concept in the SM. A solution was proposed in the mid 1960's to solve this problem. This solution is the Higgs-Brout-Englert-Guralnik-

Hagen-Kibble mechanism of spontaneous symmetry breaking, a.k.a the Higgs mechanism.

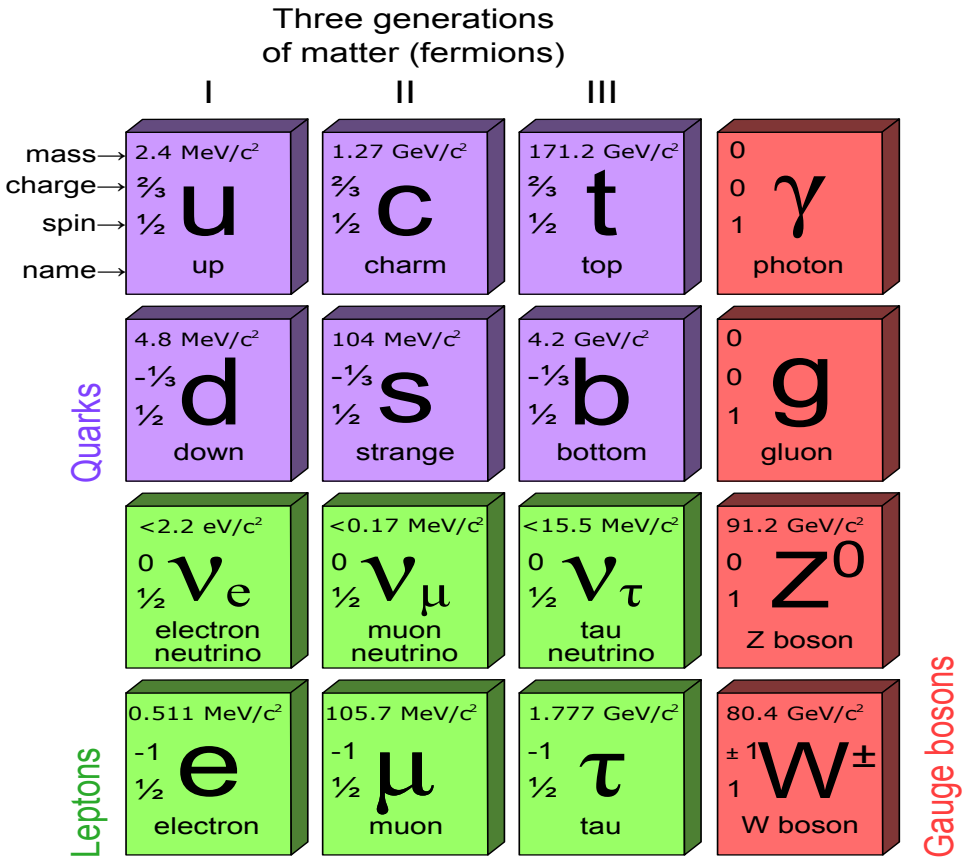


Figure 1.7: The elementary particles of the Standard Model [12].

1.2 Higgs Mechanism

The Higgs mechanism introduces a complex doublet scalar field ϕ.

$$\phi = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{1.1}$$

This is the simplest choice to keep the Lagrangian to be gauge invariant and also generate the masses of the three gauge bosons ($W^\pm$ and Z boson) because at least three degrees of freedom are required. The Lagrangian of the scalar field is given by

$$\mathscr{L} = (D^\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \tag{1.2}$$

where $V(\phi)$ is the potential

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2. \tag{1.3}$$

The sign of μ^2 and λ determine the shape of this potential $V(\phi)$. λ can only be positive to make sure the potential has a lower bound. When μ^2 is positive, the scalar field ϕ has a zero vacuum expectation value (*vev*) at the potential minimum.The first term $-\mu^2\phi^\dagger\phi$ in the Lagrangian is the mass term and it simply represents a particle with a mass of μ^2. When μ^2 is negative, the potential has a "sombrero" shape as shown in Figure 1.8 [13]. It is worth noting that the ground state or vaccum expectation value (*vev*) is not at zero now but at v with a value given by $< 0|\phi|0 >$

$$vev =< 0|\phi|0 >= \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \tag{1.4}$$

$$v = \sqrt{-\frac{\mu^2}{\lambda}}. \tag{1.5}$$

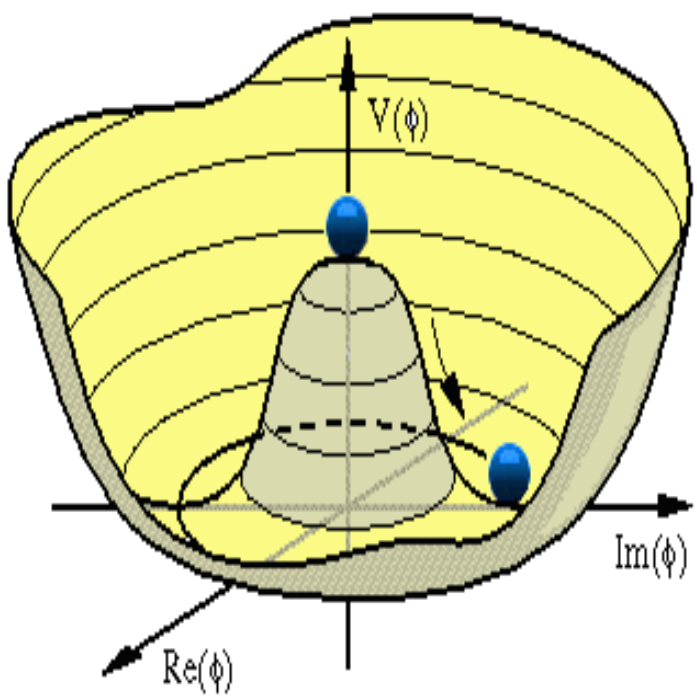


Figure 1.8: The potential of the scalar field ϕ in the case of $\mu^2 < 0$, $\lambda > 0$ [13].

Now ϕ can be written as an expansion around the *vev* after removing the phase by a gauge transformation.

$$\phi = \begin{pmatrix} 0 \\ \frac{v+H}{\sqrt{2}} \end{pmatrix}. \tag{1.6}$$

After substitution of ϕ into the Lagrangian and collecting bilinear terms (mass terms) after redefining the gauge fields, the masses of the gauge bosons including the Higgs boson are given by

$$M_W = \frac{1}{2} v g_2, \tag{1.7}$$

$$M_Z = \frac{1}{2} v \sqrt{g_2^2 + g_1^2}, \tag{1.8}$$

$$M_A = 0, \tag{1.9}$$

$$M_H = \sqrt{2\lambda v^2}. \tag{1.10}$$

where $g_{1,2}$ are gauge coupling constants.

The vacuum expectation value v can be determined from measurements of the W and Z mass but this can only constrain one of the two parameters in the expression for the mass of the Higgs boson, which leads to the Higgs mass being a free parameter in the Standard Model.

However, the Higgs mass can be constrained from the measurement of the W and top mass because the Higgs boson couples to the W and top quark through loop diagrams. In the past decades, the electroweak precision measurements have constrained the Higgs mass. The goodness of the fit provides an indirect indication of the most probable Higgs mass as shown in Figure 1.1.

1.3 Higgs Boson Production and Decay Modes

At hadron colliders such as the Tevatron, the dominant production mechanisms for a light SM Higgs boson are gluon fusion (GF) ($gg \to H$), associated production with a W or Z boson ($q\bar{q}' \to VH$, $V = W, Z$), and vector boson fusion (VBF) ($VV \to H$). Since the LHC is a proton-proton collider operating at a higher energy ($\sqrt{s} = 7(8)$ TeV), besides the three production mechanisms

enumerated above, the associated production with a top quark pair ($q\bar{q} \to t\bar{t}H$) can have a sizeable contribution. The Feynman diagrams of the four production processes are shown in Figure 1.9 [14] and the production cross sections are shown in Figure 1.10 [15].

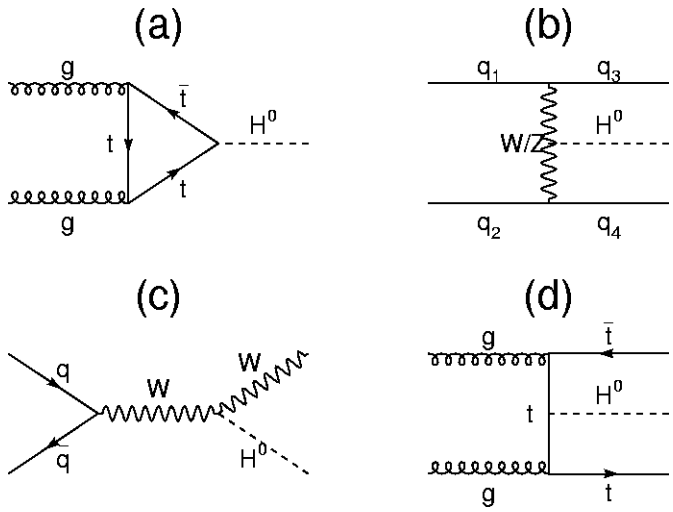


Figure 1.9: Feynman diagrams of (a) gluon fusion, (b) vector boson fusion, (c) associated production with a weak boson and (d) associated production with a top quark pair [14].

Figure 1.11 [16] shows the decay modes and branching ratios as a function of the Higgs mass. At the Tevatron the most sensitive SM Higgs boson searches rely on the $VH(H \to b\bar{b})$ process for $M_H < 125$ GeV. While at the LHC, the $H \to \gamma\gamma$ decay mode becomes one of the most promising discovery channels at low M_H (< 140 GeV), although the branching ratio of $\mathscr{B}(H \to \gamma\gamma)$ is relatively small ($\approx 0.2\%$). Since the Higgs boson does not couple to massless photons directly, the Feynman diagram for the standard model Higgs boson decaying into a photon pair is through top quark and W boson loops, as shown in Figure 1.12 [17].

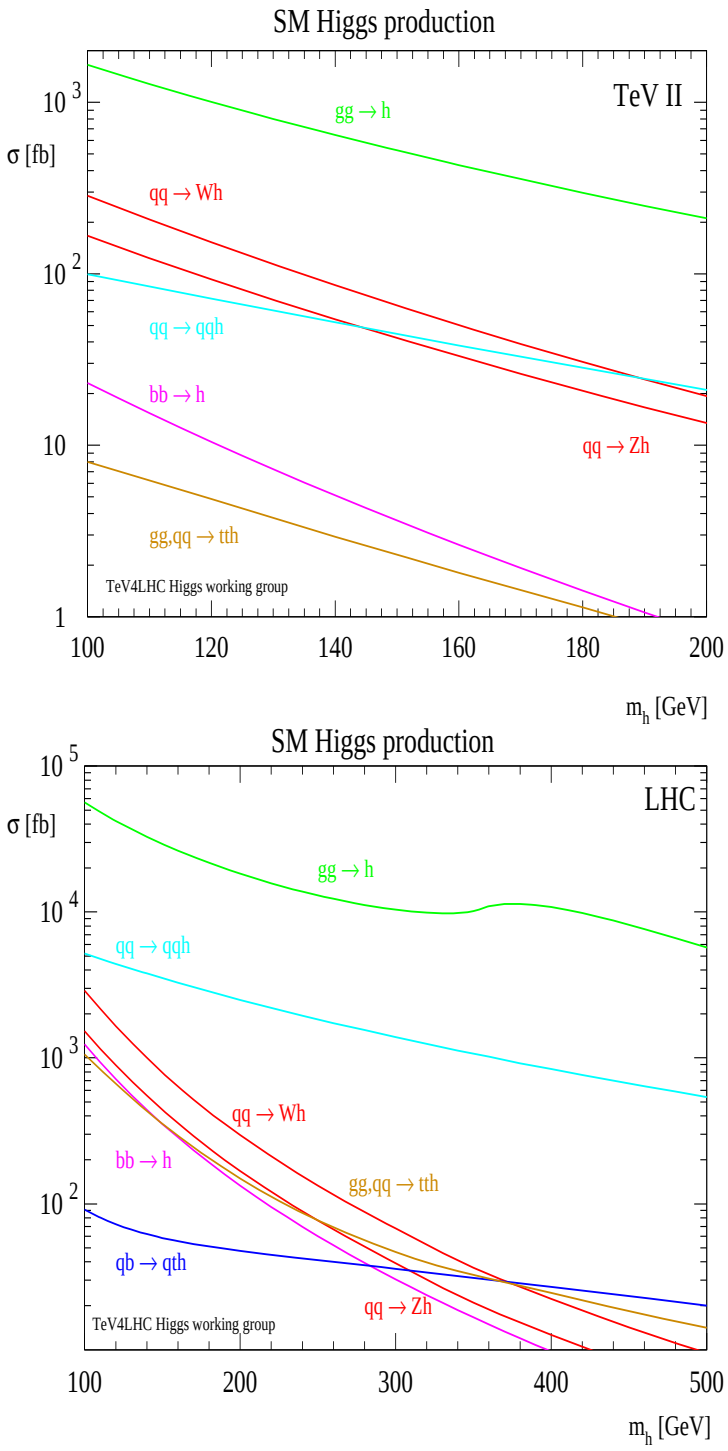


Figure 1.10: Higgs production cross sections at the Tevatron ($\sqrt{s} = 1.96$ TeV) and LHC ($\sqrt{s} = 7$ TeV) [15].

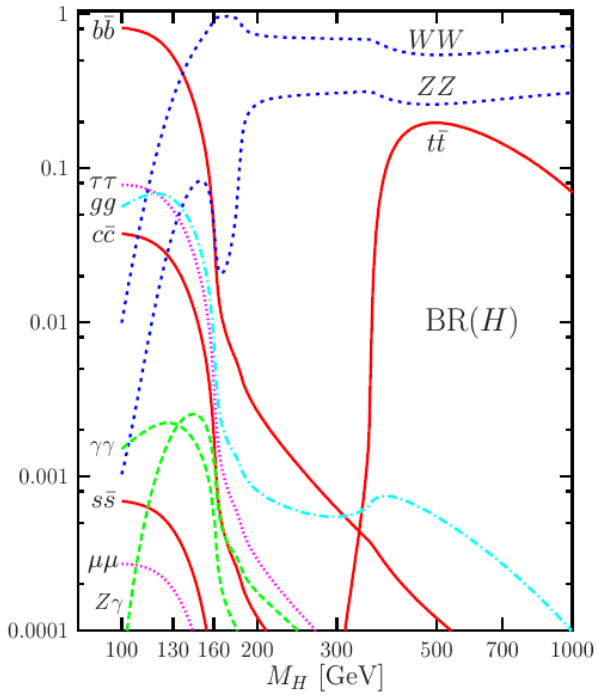


Figure 1.11: SM Higgs boson decay modes and branching ratios [16].

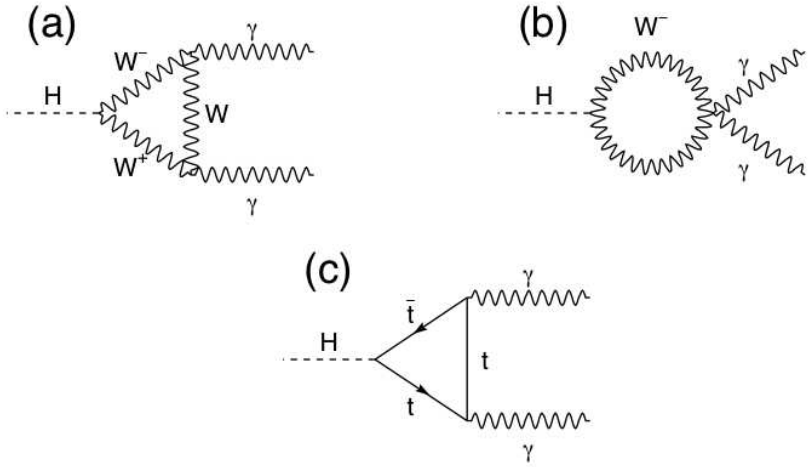


Figure 1.12: Feynman diagrams of the SM Higgs boson decaying into a photon pair through top quark and W boson loops [17].

1.4 Fermiophobic model

As discussed in the previous sections, the search for the SM Higgs boson in the $H \rightarrow \gamma\gamma$ channel at the Tevatron is difficult due to the fairly small branching ratio. However, the loop mediated decay through top quarks and weak bosons are sensitive to new physics. In some models, as extensions to the standard model [18, 19, 20, 21, 22], the Higgs boson decaying into a photon pair can be enhanced through suppressing the coupling to fermions. The greatly enhanced branching ratio provides sensitivity to direct searches for the so-called *fermiophobic* Higgs boson in the diphoton channel at the Tevatron [10, 23]. Particularly in Ref. [10], three scenarios suppressing the coupling of a Higgs boson to fermions are considered: (i) all fermions are not coupled to the Higgs boson; (ii) only down-type fermions are not coupled; (iii) only top and bottom quarks are not coupled to the Higgs boson. The resulting $\gamma\gamma$ branching ratios are all larger than those in the SM, which is shown in Figure 1.13 [24].

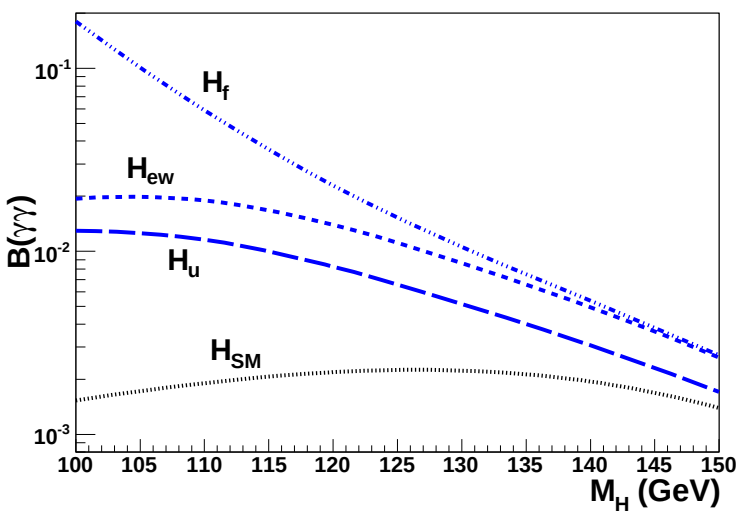


Figure 1.13: Branching ratios for the SM Higgs boson and the three scenarios in Ref. [10]. H_{SM} is the standard model Higgs boson. H_{u} is a Higgs boson with Yukawa couplings only with up-type fermions. H_{ew} is a Higgs boson that couples to all particles except the top and bottom quark. H_{f} is a Higgs boson with no tree-level coupling to fermions [24].

Among the various fermionic coupling suppression models, we consider the benchmark fermiophobic Higgs boson[1] that assumes zero tree-level couplings to all fermions, while the coupling

[1] A benchmark fermiophobic Higgs boson is called a fermiophobic Higgs boson in the following text.

strengths to bosons are retained. Thus the production mechanism through gluon fusion in Figure 1.9 (a)(d) and the decay mode in Figure 1.12 (c) are forbidden.

This Higgs boson is the H_{f} in Figure 1.13. Table 1.1 [25] lists the cross sections and the branching ratios.

mass (GeV)	ggH(NNLO)	WH(NNLO)	ZH(NNLO)	VBF(NLO)	$\mathscr{B}$(SM)	$\mathscr{B}$(fermiophobic)
100	1.82	0.281	0.163	0.100	0.00159	0.185
105	1.58	0.239	0.140	0.0924	0.00178	0.104
110	1.39	0.204	0.120	0.0852	0.00197	0.0603
115	1.22	0.175	0.104	0.0787	0.00213	0.0366
120	1.07	0.150	0.0902	0.0727	0.00225	0.0233
125	0.949	0.130	0.0785	0.0672	0.00230	0.0156
130	0.843	0.112	0.0685	0.0622	0.00226	0.0107
135	0.751	0.0972	0.060	0.0576	0.00214	0.00759
140	0.671	0.0846	0.0527	0.0533	0.00194	0.00544
145	0.601	0.0737	0.0463	0.0494	0.00168	0.00390
150	0.539	0.0644	0.0408	0.0458	0.00137	0.00273

Table 1.1: Cross sections (pb) and branching ratios in the standard model and fermiophobic model for the signal samples [25].

Searches for the fermiophobic Higgs boson have been performed at the CERN e^+e^- Collider (LEP) [26, 27, 28, 29] and the combined result [30] from the four experiments at LEP excluded a fermiophobic Higgs boson with a mass below 109.7 GeV at 95% confidence level. Searches have been done at the Tevatron by CDF [31, 32, 33] and DØ [34, 35, 36], and at the LHC by ATLAS [37] and CMS [38]. The most restrictive mass range for the fermiophobic Higgs boson comes from the combination of $H \to \gamma\gamma$, $H \to W^+W^-$ and $H \to ZZ$ searches by CMS, excluding $110 < M_{H_f} < 194$ GeV.

Chapter 2

Accelerator and Detector

The data used in this study are collected with the DØ detector at the Tevatron. The Tevatron is part of the Fermi National Accelerator Laboratory located in Batavia, Illinois, USA. The accelerator system generates proton and anti-proton beams and collides them at a center of mass energy of $\sqrt{s} = 1.96$ TeV. Before the LHC started smooth running on March 30th 2010 at $\sqrt{s} = 7$ TeV, it was the hadron collider with the highest energy in the world [39, 40]. In this Chapter, we first sketch the production process of the proton and anti-proton beams in Section 2.1 on the accelerator system and then detail how particles are detected after collisions in Section 2.2 on the DØ detector.

2.1 Accelerator System

The accelerator system is mainly composed of four parts in terms of function: the proton source, Main Injector, antiproton source, and Tevatron. The whole system is shown in Figure 2.1 [41].

The proton source system includes the Pre-accelerator (Preacc), the Linear Accelerator (Linac) and the Booster. The linac is a Cockcroft-Walton accelerator that provides a source of negatively charged hydrogen ions (H^-). The H^- gas gains an energy of 750 keV through a column from the charged dome (-750 kV) to the grounded wall. The linac further accelerates the H^- ions from the Preacc to an energy of 400 MeV through two sections: the low energy drift tube Linac (DTL) and the high energy side coupled cavity Linac (SCL). After the Linac, the 400 MeV negative hydrogen

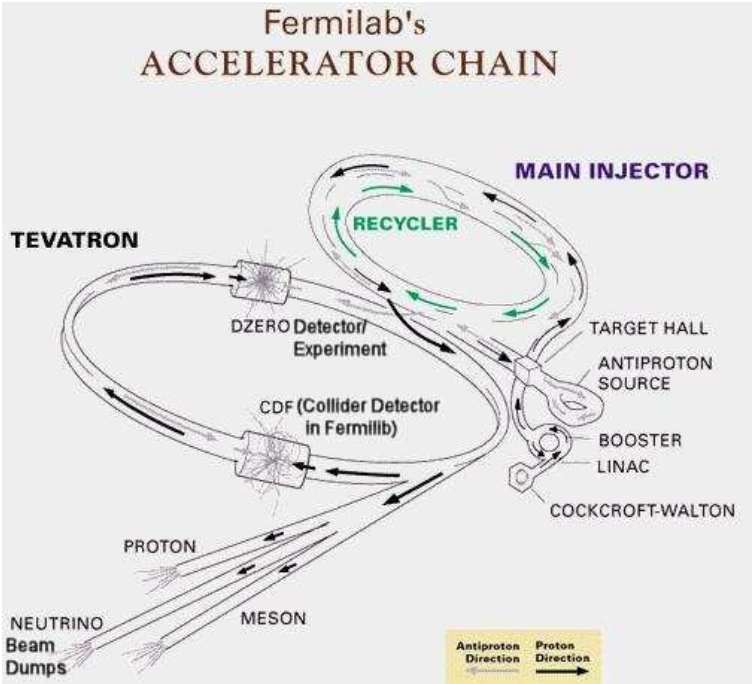


Figure 2.1: Accelerator Overview [41].

ions move into a circular accelerator, the Booster. In the Booster, H^- ions are stripped of their two electrons by a thin carbon film and become protons. The protons are then accelerated to 8 GeV and prepared to pass into the Main Injector (MI).

The MI is a synchrotron accelerator seven times the circumference of the Booster. It can accelerate the proton beam from 8 GeV to either 120 GeV when it is used to generate anti-protons, or to 150 GeV when preparing the proton beam for the Tevatron.

When the 120 GeV proton beam enters the antiproton source that is composed of the Debuncher and Accumulator, it strikes a nickel target and produces secondary particles from which 8 GeV anti-protons are collected and sent to the Debuncher. In the Debuncher, high momentum spread anti-protons are captured and sent to the Accumulator for storage.

The final stage of proton and anti-proton beam preparation happens in the Tevatron. The Tevatron is the largest synchrotron accelerator at Fermilab with a circumference of 6.28 km. The electric fields produced in a series of radio frequency cavities provide the energy to accelerate the proton and anti-proton to the final energy of 980 GeV. The magnets made from superconducting

niobium/titanium alloy are kept in $\sim$4 K liquid helium and are used to steer and focus the beams. The proton and anti-proton beams are not continuous but are split into 36 individual bunches and spread out in the ring accordingly. The interval between every two bunch crossings is 396 ns.

Collisions of proton and anti-proton bunches happen at two places around the Tevatron ring where the two detectors are placed: DØ and CDF.

2.2 DØ Detector

The DØ experiment was proposed in 1983 and its life can be divided into two major periods. Between 1992 and 1996, the DØ detector was used to study high p_T phenomena at $\sqrt{s} = 1.8$ TeV. This period is called Run I. During the 1996-2001 period, along with the Tevatron upgrades [42], the DØ detector was significantly upgraded. After upgrading [43], it consists of a central tracking system, a preshower system, a calorimeter system and a muon system. Then the DØ experiment began the so-called Run II operation from March 2001 until the final shut down in September 2011. The data we analyzed in this study is from the RunII period and so the upgraded RunII DØ detector will be described in detail. Figure 2.2 shows the overview of the upgraded DØ detector since 2001.

2.2.1 Coordinates, Kinematic Quantities and Units

Before further describing the DØ detector in detail, it is necessary to define the coordinates and a few simple kinematic quantities.

A Cartesian coordinate system is used with its origin at the detector center. The z-axis is along the direction of the proton beam and the y-axis is upward. The x-axis is then naturally defined by the vector product of the y-axis and z-axis, pointing out of the Tevatron ring and parallel to the horizon.

$$\hat{x} = \hat{y} \times \hat{z}. \tag{2.1}$$

The $x - y$ plane, or the transverse plane, is of special importance because the kinematics along

the z-axis cannot be reconstructed due to the beam pipe. For example, the transverse momentum p_T is an important kinematic variable, defined in Equation 2.2.

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{2.2}$$

It is also sometimes useful to convert the Cartesian coordinate system (x, y, z) into a spherical coordinate system (r, θ, ϕ). For instance, the rapidity y, shown in Equation 2.3, is a boost parameter and can be used to describe the direction of a particle relative to the beam line.

$$y = \frac{1}{2}\ln(\frac{E + p_z}{E - p_z}), \tag{2.3}$$

where E is the energy of the particle and p_z is the momentum along the z axis.

When the energy is much larger than the mass, which is usually satisfied in high energy experiments, the pseudorapidity, η[1], is more convenient to use as an approximation of rapidity. It is defined in Equation 2.4

$$\eta = -\ln(\tan\frac{\theta}{2}). \tag{2.4}$$

Now a particle's 4-momentum vector $(E, \vec{p})$ can be represented as (E, p_T, η, ϕ).

The natural unit system with $\hbar = c = 1$ is used, where the mass, momentum and energy are in electron-volts (eV), 1 eV$= 1.6 \times 10^{-19}$ J in SI units.

2.2.2 Central Tracking System

The central tracking system consists of the silicon microstrip tracker (SMT) and the central fiber tracker (CFT) surrounded by a 2T solenoidal magnet as shown in Figure 2.3, covering up to $|\eta| \lesssim 1.7$. The main functions of the tracker include: 1) locate the primary interaction vertex with a resolution of $\sim 35\mu m$; 2) provide high precision momentum measurement of charged particles in the central region. It is necessary for studies of the top quark, electroweak physics, B physics and

[1]Physics η represents the particle position with respect to the primary interaction vertex, often denoted as η_{phy}. If the origin is set to be the center of the detector, it is called detector η or denoted as η_{det}

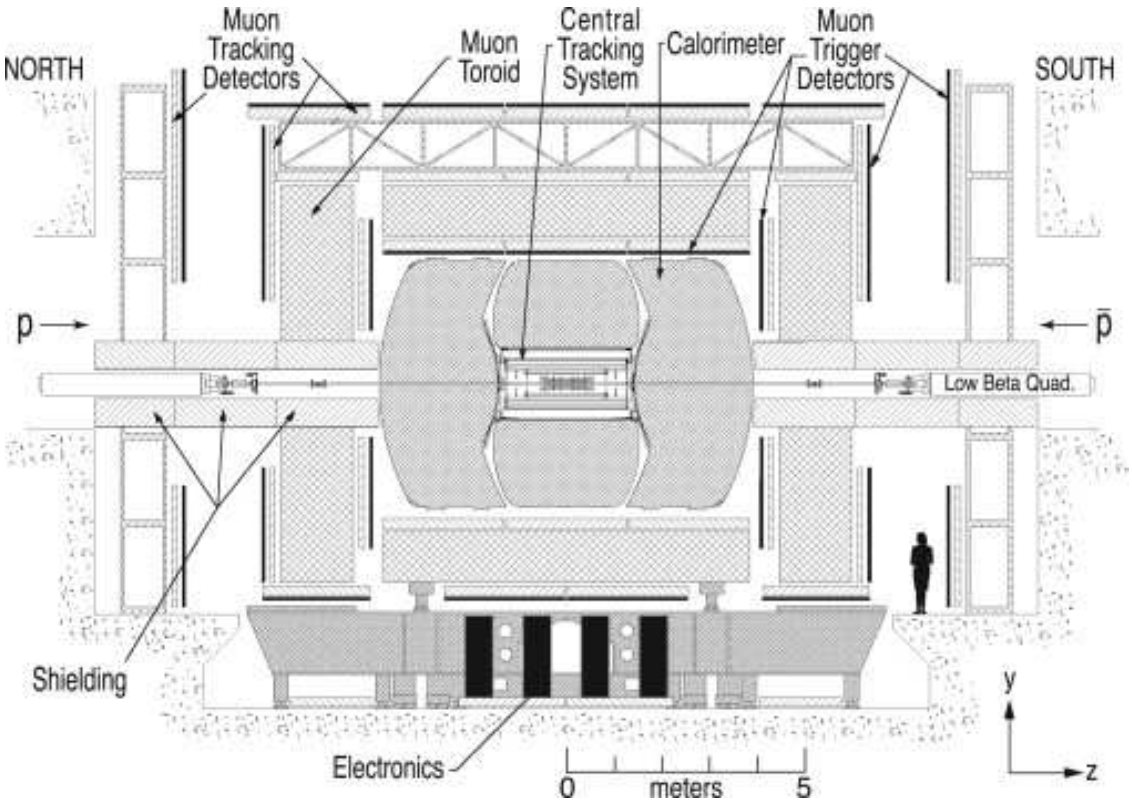


Figure 2.2: The upgraded DØ Detector [43].

searches for new phenomena and the Higgs boson.

Silicon Microstrip Tracker (SMT)

The SMT [44] is used for both tracking and vertexing over nearly the full coverage of the calorimeter and muon systems. The large coverage is achieved by constructing the geometry as in Figure 2.4.

There are six barrels in the central region, with four concentric silicon readout layers for each barrel. On each layer, there is an array of sensors called ladders. When charged particles pass through the silicon material, electron-hole pairs are produced and signals are collected by sensors via bias voltage. The first two layers have 12 ladders each and the outer two layers have 24 ladders each, for a total of 432 ladders. After the barrels are the F-disks. An F-disk has 12 double-sided wedge detectors at high $|z|$. In the most forward regions, two large-diameter disks called H-disks are placed to help tracking of charged particles with large $|\eta|$.

Central Fiber Tracker (CFT)

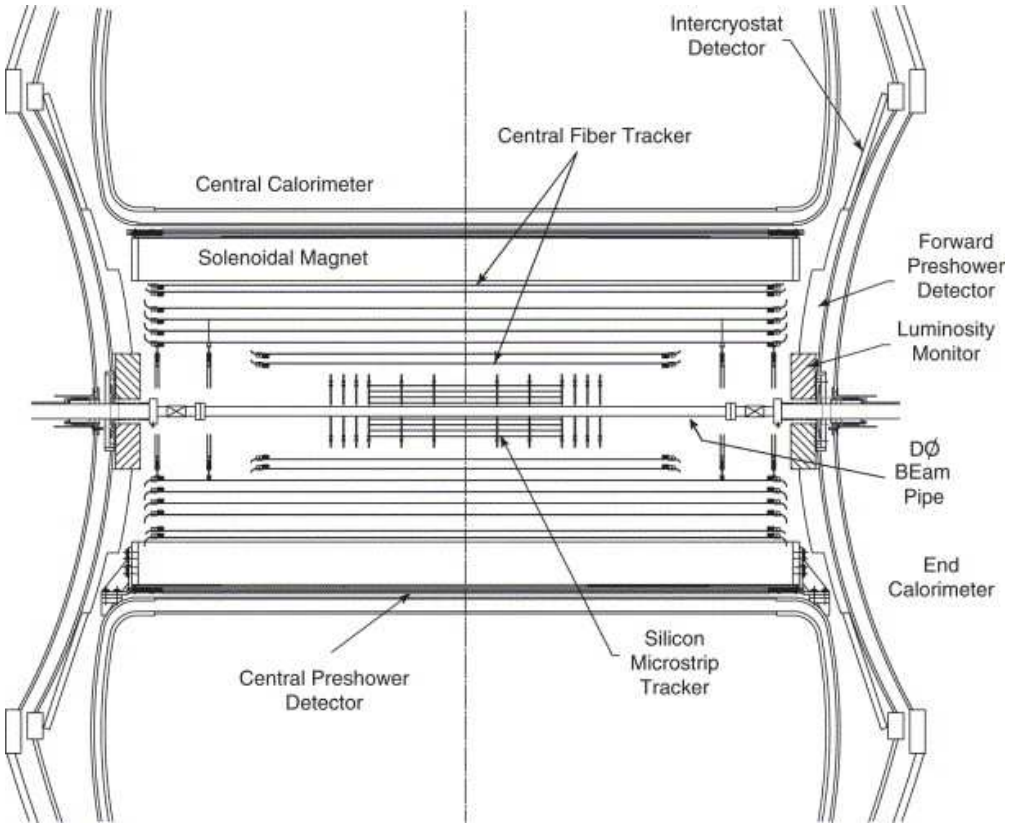


Figure 2.3: DØ Tracking System [43].

Surrounding the SMT is the Central Fiber Tracker (CFT) [45]. We first sketch the geometry of the CFT and then describe the working principle of how charged particles are detected with the CFT.

The CFT consists of scintillating fibers mounted on eight concentric support cylinders with radii of 20-52 cm. The length of the innermost two and outer six cylinders are 1.66 m and 2.52 m respectively. The length difference allows the placement of the SMT H-disks. The outer cylinders cover up to $\eta \lesssim 1.7$. Each cylinder has a doublet layer of fibers with one layer (axial layer) oriented along the beam direction (z) and a second layer (stereo) at a stereo angle in ϕ of $+3^{\circ}$ (u) or -3° (v). From the center moving outward, the doublet layers are placed in the order of $zu-zv-zu-zv-zu-zv-zu-zv$. All scintillating fibers are connected to clear fiber waveguides that propagate the scintillation light to visible light photon counters (VLPCs).

The journey of a charged particle begins with interaction with the scintillating fibers. The scintillating fibers are 835 μm in diameter and 1.66 or 2.52 m in length. The scintillating fibers

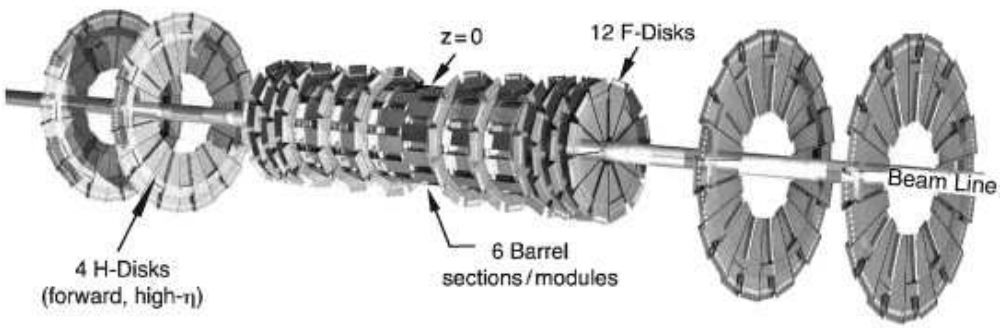


Figure 2.4: DØ Silicon Detector [43].

are assembled into ribbons consisting of two layers of 128 fibers each, as seen in Figure 2.5. The curved ribbons are then installed on the support cylinders. The scintillating fibers are made of polystyrene (PS) doped with an organic fluorescent dye paraterphenyl (pT), a wave-shifter dye 3-hydroxyflavone (3HF) and two claddings. Photons at a wavelength of 340 nm are produced by the interaction of charged particles with the PS and pT, and then absorbed by 3HF but re-emitted at 530 nm. The re-emitted photons from scintillating fibers propagate to clear fiber waveguides. The clear fiber is structurally and chemically similar to the scintillating fiber, but without fluorescent dyes. The attenuation lengths are ~ 5 m for the scintillating fibers and ~ 8 m for clear fibers. The length of the waveguides range from 7.8 m to 11.9 m. Figure 2.6 illustrates the routing of the waveguides. The other ends of waveguides point to VLPCs where the photons are converted to electronic signals. VLPCs are impurity-band silicon avalanche photodetectors that operate at 9 K. They provide rapid response, good quantum efficiency ($\geq 75\%$), high gain (22000 to 65000), low gain dispersion and the capability of functioning in a high background environment. 8 VLPCs are mounted on a VLPC chip and 128 chips are installed on a VLPC cassette. So each cassette provides 1024 VLPC pixels, or channels, of light-sensitive detector. There are, in total, 76800 channels in the CFT. The electronic signals from the VLPCs are amplified by the analog front-end boards (AFEs) that are also mounted on the cassette body. In addition to amplifying the signals, the AFEs also provide trigger discriminator signals, temperature control, and bias-voltage control.

After the passage of scintillating light through all the fibers to VLPCs, the generated signals are usually small, ~ 6 photoelectrons (pe). In order to keep acceptable efficiency for triggers while

controlling noise rates, the individual channel thresholds are set at around 1.0 pe. In order to test the connectivity, monitor the stability of the VLPC readout and calibrate the response of VLPCs channel-by-channel, fast Nichia blue-emitting LED pulsers are used to generate LED spectra as in Figure 2.7

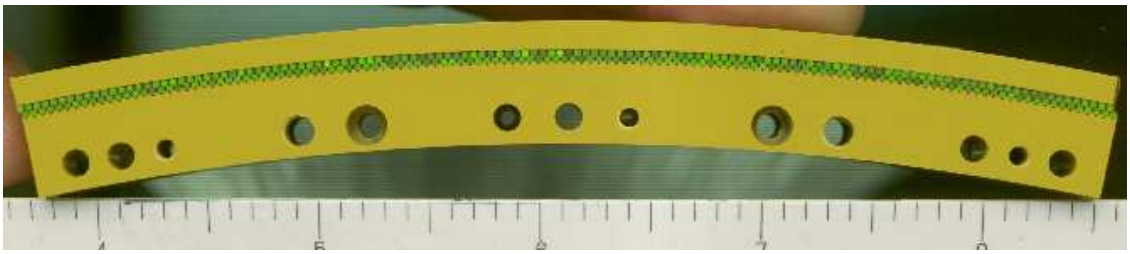


Figure 2.5: A curved ribbon consisting of two layers of scintillating fibers [43].

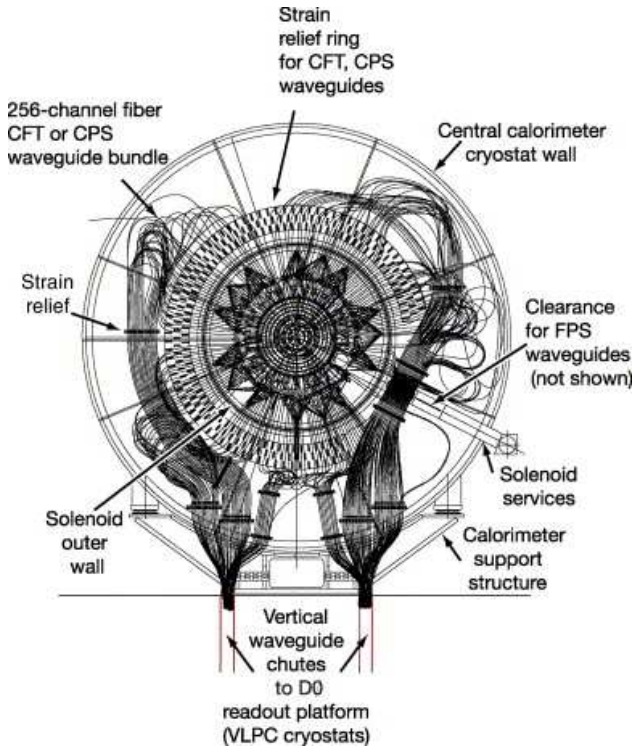


Figure 2.6: A cross section view of the routing of the clear fiber waveguides on the south face of the central cryostat [43].

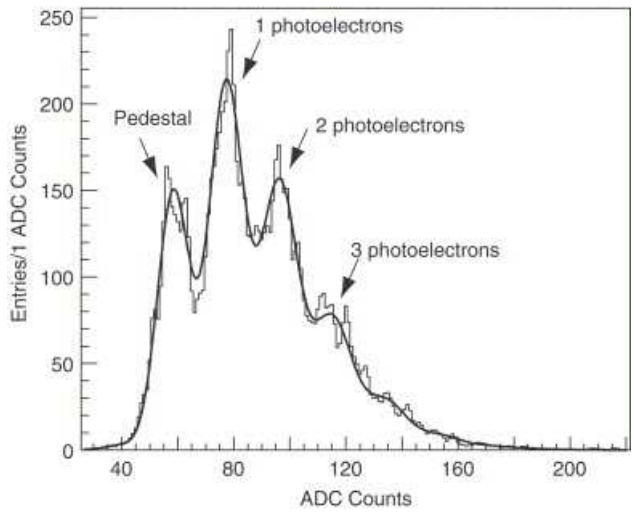


Figure 2.7: An example fit to the LED spectrum from an axial CFT fiber. The solid histogram is the data; the smooth curve is the fit [43].

2.2.3 Preshower System

The preshower system is a hybrid of a tracking system and a calorimeter. It not only enhances the spatial matching between tracks and calorimeter showers, but also can be used offline to correct the electromagnetic energy measurement of the central and end calorimeters for losses in the solenoid and upstream material. Therefore, the preshower system is helpful for photon and electron identification. By location and structure, it consists of two sub-components. The central preshower detector (CPS) [46] covers the region $|\eta| < 1.3$ and the two forward preshower detectors (FPS) [47] cover $1.5 < |\eta| < 2.5$. Both the CPS and FPS are made from triangular strips of scintillators, as shown in Figure 2.8. Embedded at the center of each strip is a wavelength shifting (WLS) fiber that collects and transfers the emitted light to the read-out. Since the triangles are interleaved, there is no dead space between strips. Most tracks traverse more than one strip, allowing for strip-to-strip interpolations and improving the position measurement, which aids the vertex identification for photons.

Central Preshower (CPS)

The CPS is comprised of three concentric cylindrical layers of triangular scintillator strips. The three layers of scintillator are arranged in an axial-*u*-*v* geometry, with a *u* stereo angle of 23.774°

and a v stereo angle of 24.016°. Each layer has 1280 strips and is made from eight octant modules. The modules consist of two 1/32” stainless steel skins with the scintillator strips sandwiched in between.

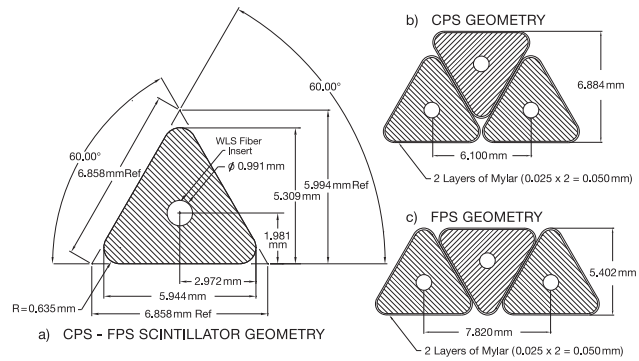


Figure 2.8: Cross section and geometrical layout of the CPS and FPS scintillator strips [43].

Forward Preshower (FPS)

The two FPS detectors are mounted on the spherical heads of the end calorimeter cryostats. Each detector is made from two layers of two planes of scintillator strips. The two layers are separated by a 2 radiation length (X_0) thick lead-stainless-steel absorber, as shown in Figure 2.9. The upstream layers are known as the minimum ionizing particle (MIP) layers; the downstream layers behind the absorber are called the shower layers. Charged particles passing through the detector will register minimum ionizing signals in the MIP layer, allowing measurement of the location (in η, ϕ and z) of the track. However, photons generally do not interact with the MIP layer, but do produce a shower signal in the shower layer.

2.2.4 Calorimeter

The energy of particles, either charged or neutral, are measured with the calorimeter system, primarily a uranium/liquid-argon sampling calorimeter system [48]. Since the momentum of a photon cannot be determined with the central tracker, the calorimeter system is the most important individual sub-detector for this study.

The whole system consists of a central calorimeter (CC) covering $|\eta| < 1.1$ and two (north and

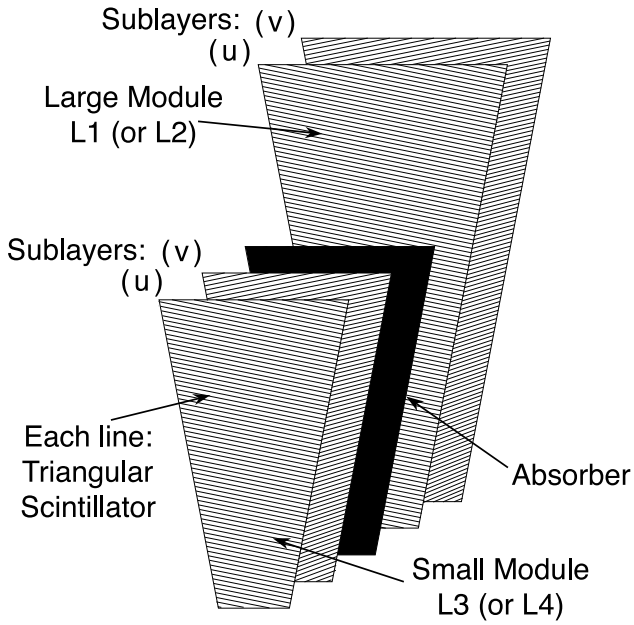


Figure 2.9: Complete ϕ-segment of a FPS module [43].

south) end calorimeters (EC) covering the forward region, $1.5 < |\eta| < 4.2$. The calorimeters contain an electromagnetic (EM) section followed outward by fine hadronic (FH) and coarse hadronic (CH) sections. An isometric view of the system geometry is shown in Figure 2.10 [49].

Each calorimeter is made from a number of cells. A typical cell structure is shown in Figure 2.11. It has a thin absorption plate made of almost pure depleted uranium for the EM sections, or uranium niobium (2%) alloy for the FH sections. The CH sections use relatively thick copper (in CC) or stainless steel (in EC) plates. The gaps between the absorption plates are filled with liquid argon as the active medium. Liquid argon is abundant in nature and relatively cheap. It does not trap charges and allows the ionization produced in electromagnetic or hadronic showers to be collected by the signal boards without amplification. It also provides the relative simplicity of calibration, the flexibility provided in segmenting the calorimeter into transverse and longitudinal cells, good radiation hardness, and relatively low cost per channel for readout electronics [50]. Particles traverse and ionize the liquid argon, generating a number of liberated electrons. The liberated electrons drift to the signal boards for readout under a 2.0 kV potential. The electron drift time across the 2.3 mm liquid argon gap is approximately 450 ns. The gap thickness was chosen to be large enough to observe minimum ionizing particle signals and to avoid fabrication difficulties.

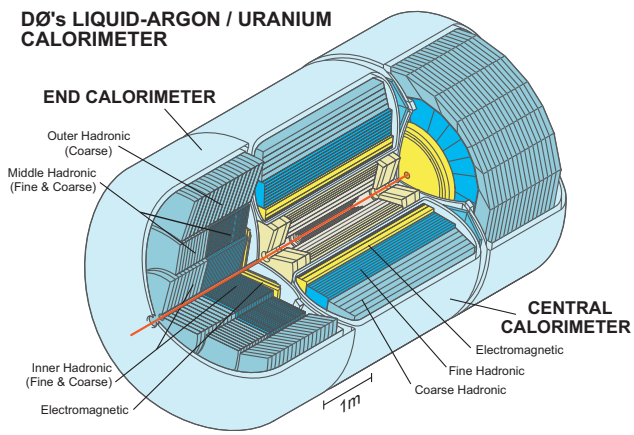


Figure 2.10: DØ Uranium/Liquid-argon Calorimeter [49].

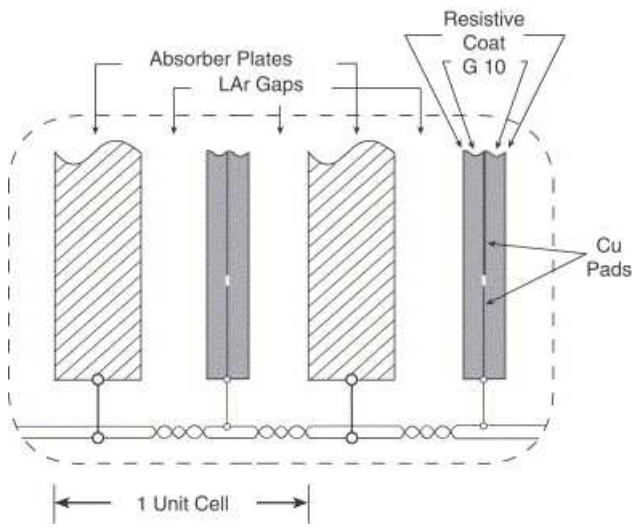


Figure 2.11: Liquid argon gap and signal board unit cell for the calorimeter [43].

As shown in Figure 2.12, the readout cells form pseudo-projective towers with a size of $\Delta\eta \times \Delta\phi \approx 0.1 \times 0.1$. The EM section is segmented into 4 layers of 1.4, 2.0, 6.8, and 9.8 X_0 thickness. The third layer, where an electromagnetic shower usually reaches its maximum, is further segmented twice into cells covering 0.05×0.05 in the $\eta \times \phi$ plane. In the ϕ direction, the CC-EM section has 32 modules. The detector responses are different when particles are incident on the modules and intermodule crack (ϕ-crack). A region is called ϕ-fiducial in CC, if a EM cluster position in ϕ at the third layer is > 0.02 from the crack.

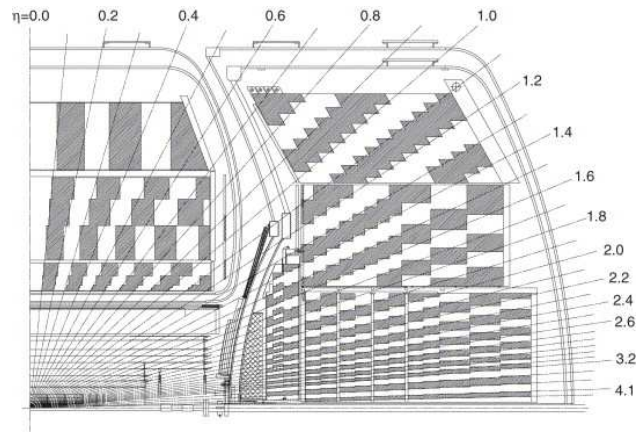


Figure 2.12: DØ Uranium/Liquid-argon Calorimeter showing segmentation in η and depth [43].

2.2.5 Muon System

Muons traverse further than electrons, photons and jets. They are detected by the muon system located at the outer most part of the DØ detector. The upgraded DØ detector adds a brand-new forward muon system in addition to the original central muon system and toroidal magnet, extending muon detection from $|\eta| \leq 1.0$ to $|\eta| \approx 2.0$.

The central muon system contains proportional drift tubes (PDTs) in three layers (A, B and C). Layers A and B have trigger scintillation counters. Instead of PDTs, the new forward muon system uses mini drift tubes (MDTs), and also trigger scintillation counters and beam pipe shielding.

The toroidal iron magnet generates a 1.8 T magnetic field outside the calorimeter, providing a second measurement of the muon momentum in addition to the central tracker.

2.2.6 Luminosity Monitor

The luminosity monitor (LM) serves to determine the luminosity at the DØ interaction region by detecting inelastic $p\bar{p}$ collisions. It is placed in front of the end calorimeters at $z = \pm 140$ cm, covering $2.7 < |\eta| < 4.4$, as shown in Figure 2.13.

The luminosity $\mathscr{L}$ is determined from the average number of inelastic collisions per beam crossing ($\bar{N}_{LM}$) measured by the LM:

$$\mathscr{L} = \frac{f\bar{N}_{LM}}{\sigma_{LM}}, \tag{2.5}$$

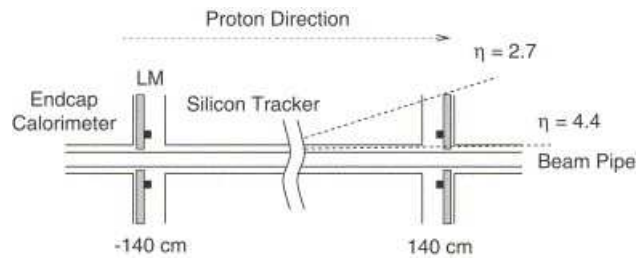


Figure 2.13: A schematic view of the location of the LM detectors [43].

where f is the beam frequency and σ_{LM} is the effective cross section that includes the acceptance and efficiency of the LM detector [51].

It is important to remove beam halo backgrounds when counting $p\bar{p}$ interactions. It is done by constraining the z coordinate of the interaction vertex to be $|z_v| < 100$ cm. z_v is calculated by Equation 2.6

$$z_v = \frac{c}{2}(t_- - t_+), \tag{2.6}$$

where $t_\pm$ are the times-of-flight for particles hitting the LM placed at ± 140 cm. Beam halo particles usually have $|z_v| \approx 140$ cm, and so are usually eliminated by the requirement of $|z_v| < 100$ cm.

2.2.7 Trigger System

The Tevatron accelerator system provides bunch crossings every 396 ns, which is $1/396 \approx 2.5$ MHz. To record every single event becomes unrealistic because of dead time of the trigger system, data recording rate and disk space. The DØ detector uses the trigger system with three distinct successive levels, called level 1 (L1) level 2 (L2) and level 3 (L3), to only select those interesting physics events to be recorded. Each succeeding level examines fewer events but in greater detail and more complexity. This is shown in Figure 2.14.

L1 examines every event for interesting properties in hardware. For example, the L1 calorimeter trigger (L1Cal) searches for energy deposition patterns exceeding programmed limits on transverse energy deposits; the central track trigger (L1CTT) examines track candidates to see if the p_T exceeds a preset threshold. After L1 selection, the trigger accept rate is reduced to 2 kHz.

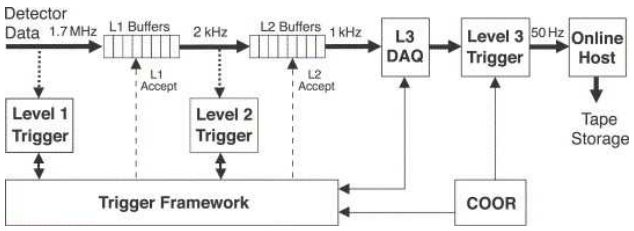


Figure 2.14: A schematic view of the trigger system of the DØ detector [43].

The second stage, L2, collects data from the L1 trigger system to form physics objects and test for correlations in these physics objects across detector sub-systems. The L2 trigger system further selects events for L3 to process and decide, reducing the trigger rate from 2 kHz to 1 kHz.

Candidate events passing the L1 and L2 selection, come to L3 for the decision whether or not to be recorded on tape. L3 is a fully programmable software trigger system and it decides on complete physics objects and their correlations. Additionally, L3 performs a limited reconstruction of events. The output rate of L3 is capped at around 300 Hz limited by data recording speed, media costs and reconstruction capacities.

Chapter 3

Object Reconstruction and Identification

When particles traverse the materials of the DØ detector, the interaction happens in different sub-detectors, as shown in Figure 3.1 [52] (neutrinos usually do not interact with the materials at all and are not detected directly). Moreover, the interaction with materials, shower width and depth for example, are different between photons, electrons and jets. In this Chapter, we describe the methods and variables for particle reconstruction and identification relevant to this study, mainly for photons since this is a study based on the di-photon final state.

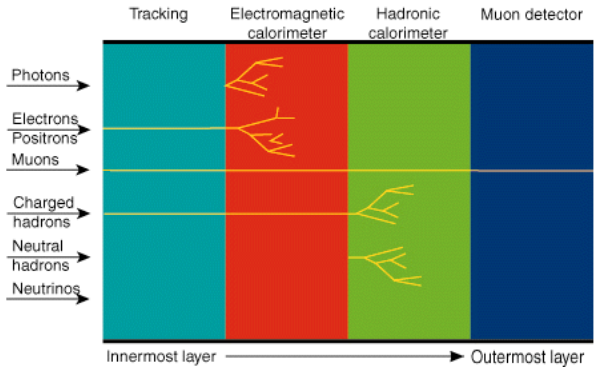


Figure 3.1: A schematic diagram of typical particle interaction locations in the DØ detector [52].

3.1 Tracks

When a charged particle (with charge q) traverses the detector in the uniform magnetic field (B), it results in a cluster of hits in adjacent SMT strips and CFT fibers in the tracker. The connection of these hits form a track showing the trajectory of the charged particle in the magnetic field. The curvature (ρ) of a track can be used to measure the transverse momentum (p_T) of a physics object (electron for example), since $\rho = q|B|/p_T$. In this study, tracks are used to reconstruct the primary vertex and distinguish photons from electrons.

In reality, there are a lot of hits observed in the tracker and a number of combinations for possible tracks. DØ adopts two different track reconstruction algorithms to find tracks and then remove duplicate tracks.

The first is the Histogramming Track Finder (HTF) [53] that determines track parameters by finding the intersection of clusters of hits for possible tracks in the parameter space. The hit positions in (x,y) coordinates are transformed into lines in the (ρ,ϕ) parameter space through the Hough transform.[1] Because clusters of hits belonging to the same track have the same (ρ,ϕ), the intersection of lines represents the parameters of the track. This is implemented for each cluster of hits and form a 2D histogram. The peaks of the histogram represent the parameters of the track candidates. Then a 2D Kalman filter [54] is used to extract the track parameters more accurately through a fit and remove bad tracks from the candidate pool.

The second is called Alternative Algorithm (AA) [55]. It first searches for a "track seed" that consists of three SMT hits and the hits must satisfy several conditions. Then the track seed grows outward through the remaining layers of the tracker, and incorporates the hits within a small spatial separation along the extrapolation using the Kalman filter. The algorithm stops if encountering three consecutive misses.

[1] The Hough transform is a concept extensively used in image analysis. A very simple case is to transform a straight line in (x,y) space into a point in the parameter space (r,θ), or polar coordinate. $r = \sqrt{x^2+y^2}, \theta = \tan^{-1}(\frac{y}{x})$.

3.2 Primary Vertices

A vertex is a common point from which tracks originate. It is reconstructed from a fit to a group of tracks consistent with originating from the same point. Each vertex must have at least two tracks with $p_T > 0.5$ GeV associated with it and the tracks must have at least two SMT hits.

For a single bunch crossing, there can be multiple primary vertices close to multiple $p\bar{p}$ interactions. These vertices form a vertex list. One of them on the list, denoted the hard scatter vertex , should be from the hard scattering that produces the physics objects with high transverse momenta that we are interested in; the others are from soft parton interactions, called minimum bias (min-bias) vertices. Tracks from the hard scatter vertex tend to have higher p_T. DØ uses this fact to construct a probability to identify the hard scatter vertex in the list.

However, the default algorithm in DØ to identify the hard scatter vertex from the vertex list yields low efficiency to find the true hard scatter vertex in our study of diphoton final states. In Section 5.1 of Chapter 5, we will discuss the algorithm for *revertexing* diphoton events in detail.

3.3 Photons

3.3.1 Photon Reconstruction

Photons, similar to electrons, develop showers and deposit energies in the cells of the EM section of the calorimeter. A group of such cells is called an EM cluster. Photons are, in the perspective of detection, EM clusters that are reconstructed by a Simple Cone algorithm [56]. The cluster energy is calculated as the sum of the energies in all the EM and FH1 cells in a cone of size $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.2$, centered on the tower with the highest fraction of the photon energy.

To be an EM cluster that looks like a photon/electron candidate, the transverse energy of the cluster should satisfy $E_T > 1.5$ GeV, deposit most of its energy in the EM section of the calorimeter, be well isolated, *etc*. In the following, we define and explain these variables individually.

EM fraction (f_{EM})

f_{EM} is the fraction of a cluster's energy deposited in the EM section of the calorimeter with respect to the total energy deposited in the EM and hadronic sections.

$$f_{EM} = \frac{E_{EM}}{E_{EM} + E_{Had}}, \tag{3.1}$$

where E_{EM} and E_{Had} are the energy measured in the EM and hadronic sections of the calorimeter in a cone of radius $\Delta R = 0.2$, respectively.

Isolation (f_{iso})

f_{iso} is the isolation defined in Equation 3.2.

$$f_{iso} = \frac{E_{tot}(\Delta R < 0.4) - E_{EM}(\Delta R < 0.2)}{E_{EM}(\Delta R < 0.2)}, \tag{3.2}$$

where $E_{tot}(\Delta R < 0.4)$ is the total energy in a cone of radius $\Delta R = 0.4$ around the direction of the cluster, summed over the entire cluster layers of the calorimeter and $E_{EM}(\Delta R < 0.2)$ is the energy in a cone of $\Delta R = 0.2$, summed over the cells of only the EM layers. Figure 3.2 illustrates the calculation of f_{iso}.

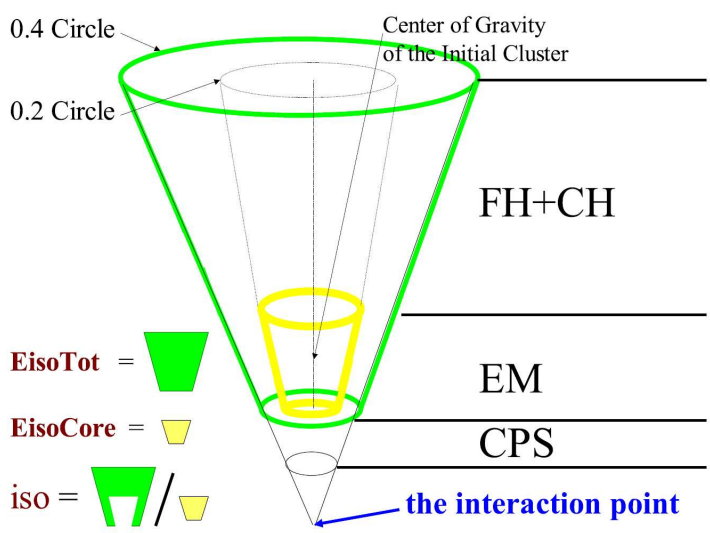


Figure 3.2: A schematic view of the isolation definition [24].

It is worthy to note that photons, as well as electrons, deposit most of their energies in the

EM section of the calorimeter and they are well isolated physics objects. Contrary to photons and electrons, hadronic jets deposit most of their energy in the hadronic section and are composed of a number of particles resulting in poor isolation. Therefore, the EM clusters found by the DØ reconstruction algorithm (EMReco) are required to have $f_{EM} > 0.9$ and $f_{iso} < 0.15$.

Shower width (*sigphi*)

A shower shape variable for a CC photon (a photon in the central region with $|\eta| < 1.1$) is defined using the width of the cluster in the $r-\phi$ plane, *sigphi*. It equals the energy weighted distance between all fired EM3 cells and the centroid position, defined below.

$$\begin{aligned} sigphi &= \sum log(\frac{E^i_{cell}}{E_{EM3}}) \times (\Delta r_i)^2 \\ &= \sum log(\frac{E^i_{cell}}{E_{EM3}}) \times (-sin\phi_{EM} \cdot x^i_{cell} + cos\phi_{EM} \cdot y^i_{cell})^2, \end{aligned} \tag{3.3}$$

where E^i_{cell} is the energy of the cell i, E_{EM3} is the total energy in the EM3 layer, and Δr_i is the transverse distance between cell i and the centroid position.

H-Matrix (*HMx*7(8))

HMx7(8) is a variable based on a $7(8) \times 7(8)$ covariance matrix (M) of 7(8) longitudinal and transverse shower variables. It is mainly used to discriminate electrons from hadronic jets. But since photons showers share similarities with electrons, this variable (*HMx*8) is used in the EC region to identify photon candidates.

The 8 observables include the four energy fractions in each of the four EM layers, the total EM cluster energy, the primary vertex z-position, and transverse shower widths in ϕ and z. HMx7 uses the same variables except the transverse shower width in z.

The covariance matrix M is built using Monte Carlo electrons, with the matrix elements given by Equation 3.4

$$M_{ij} = \frac{1}{N} \sum_{n=1} N(x^n_i - \overline{x_i})(x^n_j - \overline{x_j}), \tag{3.4}$$

where the sum is performed over N electrons; x^n_i is the value of variable i for the nth electron.

The consistency of the cluster shape with an EM cluster is given by a χ^2 computed with Equa-

tion 3.5

$$\chi^2 = \sum_{i,j}(x_i^{'} - \bar{x_i})H_{ij}(x_j^{'} - \bar{x_j}), \tag{3.5}$$

where $H \equiv M$. An EM cluster will have a low χ^2.

Artificial neural network (*ANN*)

To further suppress jets misidentified as photons, an artificial neural network (ANN) [57] is trained to combine the information from a set of variables sensitive to the differences between photons and jets in the tracker and the energy distributions in the calorimeter and CPS. The ANN input contains the following variables[2].

- *trkiso*, *i.e.* the scalar sum of the p_T of all $p_T > 0.5$ GeV tracks ($p_{T\mathrm{trk}}^{\mathrm{sum}}$) originating from the hard scatter vertex in an annulus of $0.05 < \Delta R < 0.4$ around the EM cluster,

- the number of cells above the transverse energy (E_T) dependent threshold ($0.004 \times E_T + 0.25$ GeV) in the first EM calorimeter layer within $\Delta R < 0.2$ and $0.2 < \Delta R < 0.4$ of the EM cluster,

- the number of CPS clusters within $\Delta R < 0.1$ of the EM cluster,

- the squared-energy-weighted width of the energy deposit in the CPS [59]: $\frac{\sum_i E_i^2 \times (\phi_{EM} - \phi_i)^2}{\sum_i E_i^2}$, where E_i and ϕ_i are the energy and azimuthal angle of the ith strip, and ϕ_{EM} is the azimuthal angle of the EM cluster at the EM3 layer.

Distributions of these variables are shown in Figure 3.3.

[2]These 5 variables are for CC photons, so the ANN is also called ANN5. For EC photons, there are 4 input variables and the ANN is also called ANN4. Please refer to [58] for details.

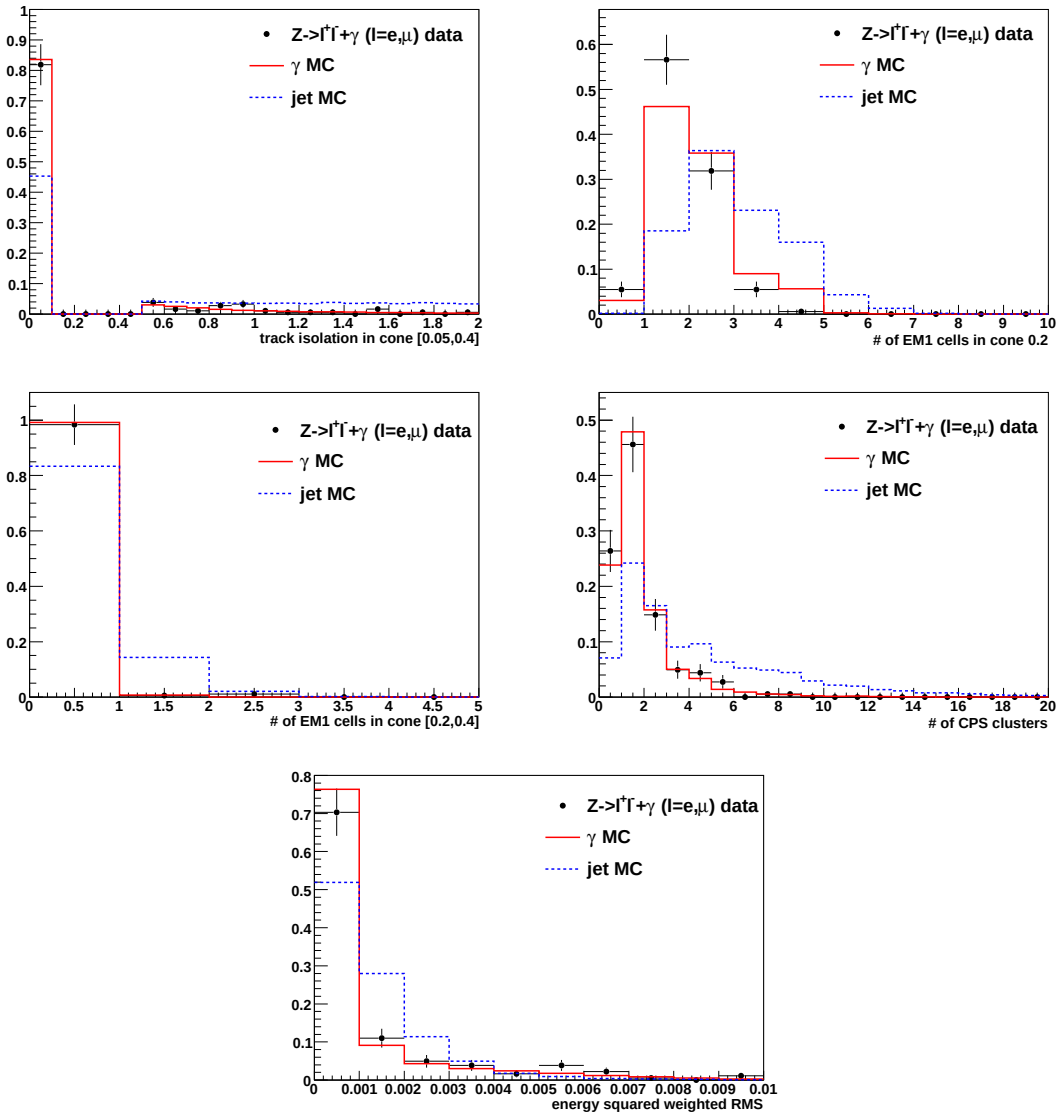


Figure 3.3: Normalized O_{NN} input variables for photons and jets [24].

The ANN is trained using diphoton and dijet Monte Carlo (MC) samples and its performance is verified using a data sample of $Z \to \ell^+\ell^-\gamma$ ($\ell = e, \mu$) events (see Figure 3.4). Figure 3.5 compares the ANN output (O_{NN}) spectrum for photons and jets. When $O_{NN} > 0.1$ is required for a photon candidate, it is almost fully efficient for real photons while rejecting $\sim 40\%$ of misidentified jets.

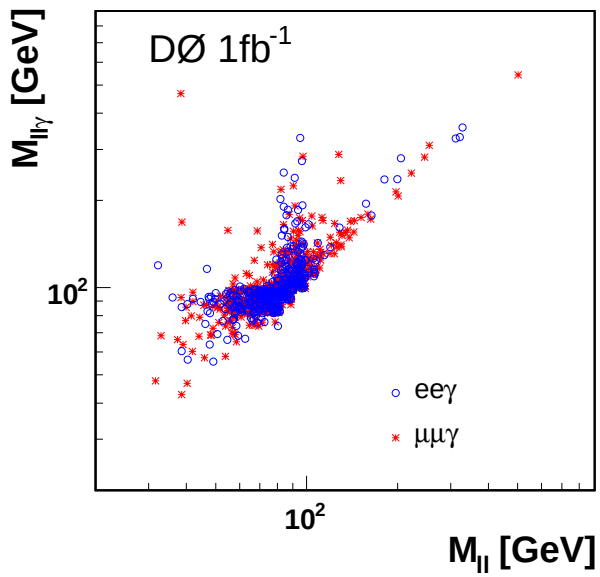


Figure 3.4: The three-body (di-lepton+γ) mass versus the two-body (di-lepton) mass from $Z \to \ell^+\ell^-\gamma$ ($\ell = e, \mu$) data events [60]. To pick up the pure final state radiation (FSR) γ candidates, the two-body mass is required to be less than 82 GeV while simultaneously the three-body mass is required to be within 82 - 102 GeV.

Spatial track match probability (P_{trk})

Photons are similar to electrons in shower shape and development but they still have distinctions, one of which is that photons usually do not have associated tracks. P_{trk} is a probability calculated from the χ^2_{trk} of the spatial separation significance between the EM cluster position in the third layer of the EM calorimeter (EM3) and the position from extrapolated tracks.

$$\chi^2_{trk} = (\frac{\Delta\eta}{\sigma_\eta})^2 + (\frac{\Delta\phi}{\sigma_\phi})^2, \tag{3.6}$$

where $\Delta\eta$ and $\Delta\phi$ are the differences between the extrapolated track position and the EM cluster position in the EM calorimeter. σ_η and σ_ϕ are the extrapolated track uncertainties in η and ϕ, respectively.

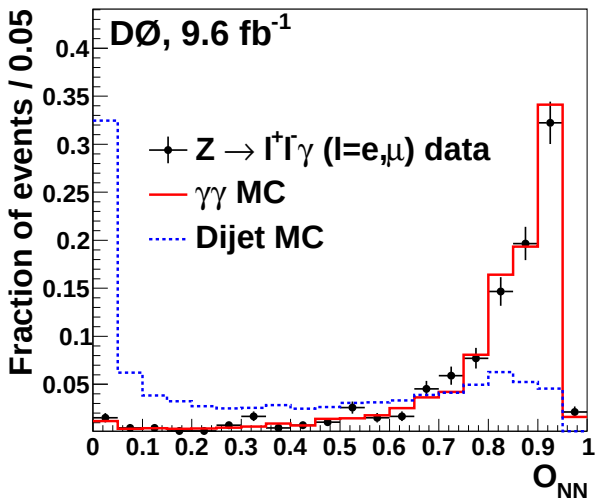


Figure 3.5: Normalized distributions of O_{NN} from real and fake photons.

Hits-on-the-road discriminant (D_{HOR})

Sometimes an electron leads to an EM cluster but the associated track is neither reconstructed nor matched to the EM cluster due to tracking inefficiency, leaving only a series of hits in the tracker. The tracking inefficiency also becomes worse in events with a high number of $p\bar{p}$ interactions. A “hits-on-the-road" discriminant (D_{HOR}) [61] is used to include the hits information and so further separate photons from electrons.

For an EM object, a road is defined from the hard scatter vertex of the event to either the EM cluster in EM3, or the CPS cluster if the EM cluster matches with a CPS cluster. Two roads (left and right) are considered as shown in Figure 3.6 to allow for the electron and positron hypotheses.

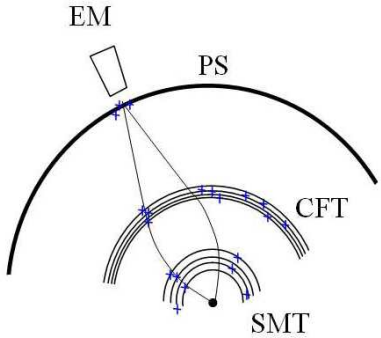


Figure 3.6: Illustration of the Hits-on-the-road algorithm [61].

The number of fired CFT fibers and SMT hits are counted within 4σ from the road. D_{HOR} is calculated by Equation 3.7

$$
\begin{aligned}
D_{HOR} &= \frac{P_e(N_{hits})}{P_e(N_{hits}) + P_\gamma(N_{hits})}, \\
P_e(N_{hits}) &= \frac{\sum_{i=0}^{N_{hits}}(N^e_{hits}(i))}{\sum_{i=0}^{24}(N^e_{hits}(i))}, \\
P_\gamma(N_{hits}) &= \frac{\sum_{i=N_{hits}}^{24}(N^\gamma_{hits}(i))}{\sum_{i=0}^{24}(N^\gamma_{hits}(i))},
\end{aligned}
\tag{3.7}
$$

where P_e and P_γ are the probabilities for an EM object to be an electron or photon; N^e_{hits} and N^γ_{hits} are the distributions for the total number of hits for electrons and photons. The maximum number of hits is 24, dictated by the geometry of of CFT (8 doublet layers, $8 \times 2 = 16$) and the SMT (8 measurement planes).

3.3.2 Photon Identification

The photon identification (ID) is constructed by certain requirements on the variables described in Section 3.3.1.

The following photon ID definitions, named "core cut", are chosen and optimized for DØ RunII data [62]. The core cut efficiencies are different in data and MC, therefore the ratio of the efficiency in data and MC, also denoted as scale factors, are measured and implemented in MC as corrections. The scale factors for photon ID are a product of two factors that are measured individually. The first factor relates to the requirements on shower shapes. This factor is measured via the tag-and-probe method implemented on $Z/\gamma^* \to e^+e^-$ data and MC, which takes advantage of the fact that an electron's shower profile is similar to that of a photon and that $Z/\gamma^* \to e^+e^-$ events are statistically abundant in both data and MC. The second factor relates to the track veto for photons. A genuine photon could be rejected by the track veto in two cases: 1) the photon converts to an e^+e^- pair in the upstream material and one or more tracks are reconstructed, with a probability of $(5.9 \pm 0.2)\%$; 2) a random track from underlying events is matched to the photon, with a probability of $\sim 0.05\%$ [63]. This factor is measured using radiated photons from the

$Z/\gamma^* \rightarrow l^+l^-\gamma\ (l = e, \mu)$ process.

Taking into account the two factors stated above, the measured scale factors are presented in Ref. [62] (RunIIa) and Ref. [64] (RunIIb).

Variables	CC core0	CC core1	CC core2	EC core0	EC core1	EC core2
$f_{iso} <$	0.15	0.10	0.07	0.15	0.10	0.07
$f_{EM} >$	0.90	0.95	0.97	0.90	0.95	0.97
$trkiso <$	2.0	2.0	1.5	1.5	1.5	1.5
$HMx8 <$	-	-	-	-	30	30
$sigphi <$	18(14)	18(14)	18(14)	eq1	eq1	eq1
$sigz <$	-	-	-	eq2	eq2	eq2
$ANN5 >$	0.1	0.1	0.3	-	-	-
$ANN4 >$	-	-	-	0.05	0.1	0.3
$P_{trk} <$	$0.0(10^{-4})$	$0.0(10^{-4})$	$0.0(10^{-4})$	0.001	$0.0(10^{-4})$	$0.0(10^{-4})$
$D_{hor} <$	0.8(0.5)	0.9(0.5)	0.9(0.5)	-	-	-

Table 3.1: p20 (p17) core cuts definitions for CC ($-1.1 < \eta_{det} < 1.1$) and EC ($1.5 < |\eta_{det}| < 2.5$) regions. For p20, eq1 = $7.3\eta_{det}^2$ - $35.9|\eta_{det}|$ + 45.7, and eq2 = $7.5\eta_{det}^2$ - $36.0|\eta_{det}|$ + 44.8. For p17, eq1 = $2.74\eta_{det}^2$ - $16.3|\eta_{det}|$ + 25.0, and eq2 = $5.96\eta_{det}^2$ - $30.6|\eta_{det}|$ + 40.7.

3.3.3 Photon Energy Scale and Resolution

At the Tevatron, there are not statistically large amounts of pure photon data coming from a resonance to study the photon energy scale and resolution, which is still true even if the Higgs boson is found. However, because electrons share similar shower profiles with photons and they are both identified as EM clusters in the calorimeter, $Z/\gamma^* \to e^+e^-$ events are used for photon energy calibration. There is a shortcoming of this method for the energy scale calibration because an electron loses more energy in the material in front of the calorimeter especially at larger rapidities, resulting an "over-correction" for photons. In the previous studies [36, 65], this over-correction is covered by a large systematic uncertainty (0.59%). In this study, we apply an η-dependent correction ($\sim$1-5%) based on GEANT simulation of the DØ detector response to compensate for the difference.

Photon energy scale

The photon energy scale re-calibration [66] is done on top of the existing corrections for ϕ_{mod} [67] and instantaneous luminosity [68], rather than derived from scratch. Single photon MC events are generated at energy points of $E = 15, 20, 25, ..., 70, 80, 90$ and 100 GeV, and then used to measure the ratio of the photon transverse momentum at generator level (p_T^{true}) to the average reconstructed photon transverse momentum $\bar{p}_T^{reco}$ as a function of $\bar{p}_T^{reco}$. The shape of the function $f(\bar{p}_T^{reco})$ is dependent on the physics η, and so the events are examined in 7 rapidity bins.

$$\frac{p_T^{true}}{\bar{p}_T^{reco}} = f(\bar{p}_T^{reco}). \tag{3.8}$$

Photon energy resolution

The energy resolution of the calorimeter can be written as Equation 3.9 [69].

$$\frac{\sigma(E)}{E} = \sqrt{C^2 + \frac{S^2}{E} + \frac{N^2}{E^2}}. \tag{3.9}$$

The constant term C comes from the non-uniformities in the gain calibration. The sampling term $\frac{S}{E^{\frac{1}{2}}}$ relates to the fluctuations of the shower development, such as fluctuations of the sampling

fraction (the fraction of the energy deposited in the active medium). The noise term $\frac{N}{E}$ is due to the electronic noise and the low activity of the uranium. For high energy photons, the energy resolution is approximately equal to the constant term C in the leading order. C is measured to be $\sim 2\%$ [70].

Worse energy resolution in data is observed than in MC events given by the GEANT simulation of the DØ detector. Thus the MC events are further smeared to match data through a "over-smearing" process [67].

3.4 Jets

The hard-scattering of $p\bar{p}$ collisions produces high momentum quarks and gluons (quarks and gluons are collectively called partons). Partons cannot exist in free form due to color confinement in QCD, and so hadronize on the fly into a narrow cone of particles such as pions and kaons. Such a collection of particles is called a jet. In this sense, a jet is the observable of a parton.

A jet is reconstructed from the clusters in the calorimeter within a cone of a certain radius[3] using the Ban iterative midpoint cone algorithm [71]. The basic idea of this algorithm is:1) find proto-jets from preclusters that are composed of reconstructed calorimeter towers; 2) use midpoints of proto-jets pairs as seeds to find more proto-jets so as to suppress jets from soft radiation;3) update the increased proto-jets lists via a split-and-merge process to ensure that each tower belongs to one jet only.

Since a jet usually contains charged particles, there are some reconstructed tracks inside a jet cone. By using tracks, a jet is said to be "vertex confirmed" if there are at least two tracks in the jet cone associated with the primary vertex. The vertex confirmation is used to improve the counting of jet multiplicity since the resolution in the z axis is much better for tracks than jets. Another way to use track information is b-tagging [72].

The measured jet energy is usually different from the parton's energy for various reasons. To compensate for the difference, the energy scale is corrected by a jet energy scale correction (JES),

[3]The algorithm "JCCA(JCCB)" uses a cone of radius $\mathscr{R} = 0.7(0.5)$.

and the jet resolution in MC is corrected by a method of jet shifting, smearing, and removal (JSSR) [71].

3.5 Missing Transverse Energy

In $p\bar{p}$ collisions at the Tevatron, the transverse momenta of the initial partons are close to zero, which implies that the sum of the transverse momenta of all final state physics particles should almost vanish by momentum conservation. However, neutrinos or any new physics particles that escape from direct detection by the DØ detector, as well as instrumental reasons (thermal noise in the calorimeter, *etc*), may result in an imbalance of the total transverse momentum, *i.e.*, the missing transverse energy $\not{E}_T$.

The $\not{E}_T$ calculation is performed in two stages. Firstly, it is computed as the negative vector sum of the energies from all cells of the EM and FH calorimeters. The cells from the CH calorimeter are excluded due to a high-level noise.

In the second stage where physics objects are already reconstructed and identified, the $\not{E}_T$ is re-computed by incorporating corrections to the physics objects. For instance, an EM cluster that could be an electron or photon need be corrected for its energy scale; jet energies are corrected for JES, JSSR, in-cone muons, and leakage of hadronic showers in the CH calorimeter, and so on. After these corrections to the physics objects, a re-computation of $\not{E}_T$ is performed for more accuracy. Since the physics objects such as electrons, photons and jets are identified as clusters of energies in the calorimeter, the vector sum of the unclustered energies (UE) can be calculated to characterize the underlying activity of the calorimeter in an event.

Chapter 4

Samples

In this chapter, we describe the data sample and Monte Carlo (MC) simulations used in this study.

4.1 Data Sample

For the RunII period, the accelerator division delivered 11.9 fb^{-1} of $p\bar{p}$ collisions, with 10.7 fb^{-1} recorded by the D0 detector from April 2002 to September 2012, see Figure 4.1 [73]. The recorded data is filtered to remove bad luminosity blocks and events by the Data Quality Group [74], resulting in a data sample corresponding to an integrated luminosity of 9.6 fb^{-1} that is used for this study.

In 2006, the DØ detector was upgraded with the "layer 0" installed in the SMT as the innermost layer. Since then, the reconstruction software was updated to incorporate the hardware configuration change. So the data sample is divided into two subsets, often called RunIIa which corresponds to the 1.1 fb^{-1} data before the upgrade in 2006, and RunIIb[1] for the rest of the data. The reconstruction software version begins with p17 and p20 for RunIIa and RunIIb data respectively. So RunIIa (b) data are often referred as p17 (p20) data, as well as Monte Carlo simulations.

For efficient usage of computer resource, the data sample is reduced to various *skims* according

[1]There are mainly four updates on the physics object reconstruction after the year of 2006, and so the RunIIb data can be further subdivided into RunIIb1, 2, 3,and 4 datasets.

to the final states by the Common Sample Group [75]. In this study, we use the $2EMhighpt$ skim that requires two EM clusters with $p_T > 12$ GeV be present in each event.

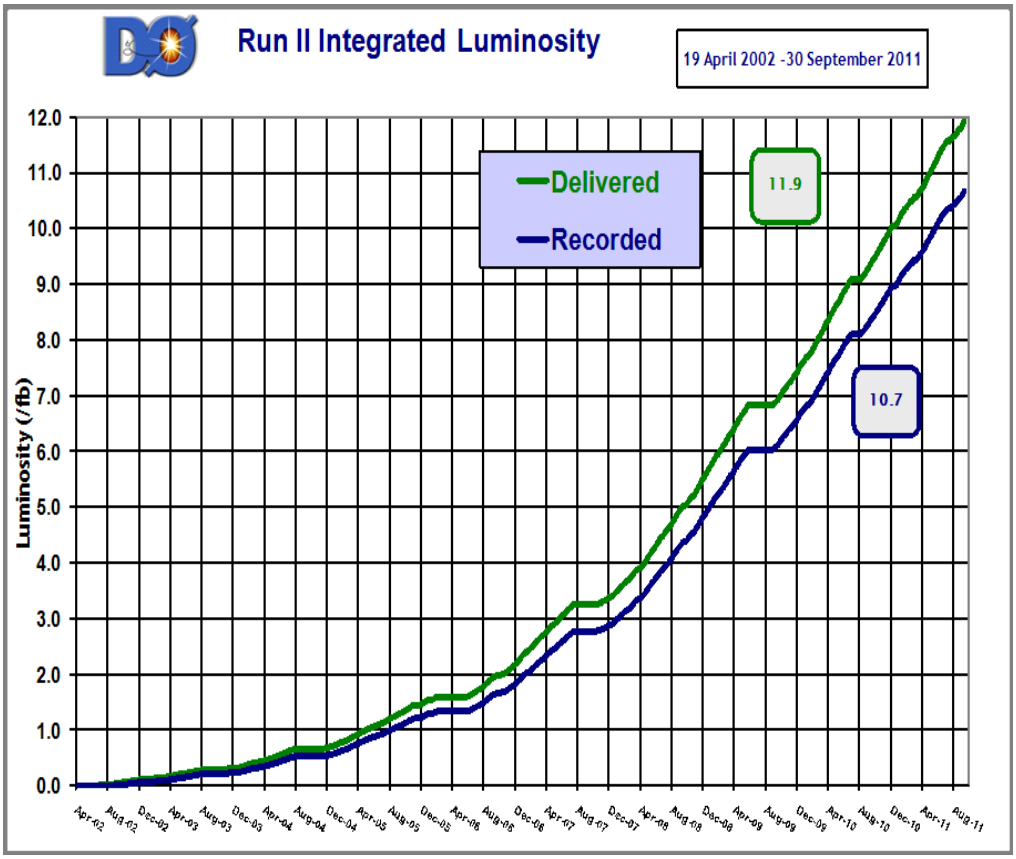


Figure 4.1: The delivered and recorded integrated luminosity as a function of time at the DØ detector [73].

4.2 Monte Carlo Simulation

All the signal and background processes are modelled using Monte Carlo simulations, except for the $\gamma+$ jet and dijet backgrounds.

Signal samples

The signal Monte Carlo samples for the SM Higgs boson are generated using PYTHIA [76] with the CTEQ6L1 [77] parton distribution functions (PDFs), for 100 GeV$\leq M_H \leq$150 GeV in intervals of 5 GeV. The samples corresponding to each of the three dominant SM Higgs boson production mechanisms at the Tevatron are normalized using the next-to-next-to-leading order (NNLO) plus

next-to-next-to-leading-logarithm (NNLL) theoretical cross sections for GF [78] and NNLO cross sections for VH and VBF processes [79, 80], computed with the MSTW 2008 PDF set [81].

The branching ratio predictions are from HDECAY [82]. Moreover, the p_T of the Higgs boson in the GF process is corrected to match the calculations by HQT at NNLO and NNLL accuracy [83].

In the fermiophobic model where the GF process is absent, we assume the VH and VBF signals are produced at the same rate as the SM. The branching ratios are also calculated with HDECAY but with the option "fermiophobic" turned on in the software.

Background samples

Backgrounds of instrumental and physical origin are considered in this search. The major instrumental backgrounds include γ + jet (γj), dijet (jj) and $Z/\gamma^* \to l^+l^-(l = e, \tau)$ (ZDY) production, with jets or electrons misidentified as photons. The major physical background is from direct $\gamma\gamma$ production (DDP) where two isolated photons with high transverse momenta are produced.

$\gamma j + jj$ events are obtained from independent data control samples as discussed later in Section 5.3. The sample for the ZDY process is simulated using ALPGEN [84], with showering and hadronization from PYTHIA, and with the NNLO cross section [85] for normalization. Additionally, the Z boson p_T spectrum is corrected to match data [86]. The DDP sample is generated using SHERPA [87].

All the MC samples are processed through a detailed GEANT-based [88] simulation of the D0 detector. In order to accurately model the effects of multiple $p\bar{p}$ interactions and detector noise, data events from random $p\bar{p}$ crossings that have an instantaneous luminosity spectrum similar to the events in this analysis are overlaid on the MC samples. These MC events are then processed using the same reconstruction algorithm as data.

Chapter 5

Analysis

In this Chapter, we elaborate in detail the data analysis. The search for the SM Higgs boson in the channel of $H \rightarrow \gamma\gamma$ is conducted by using the samples described in Chapter 4.

An innovation implemented in this study is the *revertexing*, it significantly enhanced the efficiency to find the true vertex. In Section 5.1, the algorithm and calibration are described.

Section 5.2 and Section 5.3 illustrate the event selection and background modeling, which provides the foundation of another important innovation, the multivariate technique.

It is known that the search for the $H \rightarrow \gamma\gamma$ channel at the Tevatron is challenging due to the small branching ratio (maximum of $\approx 0.2\%$ near 125 GeV). Nevertheless, the experimental signature is very clean thanks to a narrow resonance on top of a smoothly-falling background in the diphoton mass spectrum. So the analyses at DØ [65] and at CDF [89] in the past focused only on the peak-searching by examining the diphoton invariant mass distribution. In this study, we implement the multivariate technique (Section 5.4) for the first time and achieve the best sensitivity in this channel at the Tevatron. In the multivariate analysis, the diphoton mass spectrum still provides the majority of the discriminating power between signal and background. So we also scan the diphoton mass spectrum for a resonance (Section 5.3.5) and compare it with the result from the multivariate analysis.

The last Section 5.5 of this Chapter deals with systematic uncertainties.

5.1 Revertexing

From Section 3.2 in Chapter 3, we know that the reconstruction of the primary vertex relies on the associated tracks and the track p_T. It performs well for hard scatter (high p_T) events with electrons because the electrons themselves usually result in high p_T associated tracks. However, photons are charge-neutral particles that usually do not have tracks. So the default DØ vertex reconstruction (reco) algorithm yields a low efficiency to find the true hard-scatter vertex, for the diphoton final states. The previous search for $H \to \gamma\gamma$ [65] and the diphoton differential cross section measurement [90] replace the default DØ vertex reco algorithm with a *max-track* algorithm. The algorithm identifies the hard scatter vertex in a diphoton event as the vertex with the maximum number of associated tracks in the vertex list, regardless of the track p_T. We use the signal MC sample (gluon fusion) to demonstrate the performance difference of the default DØ reco and max-track algorithms, as shown in Figure 5.1 The true primary vertex is said to be found or matched if the chosen vertex is within 1 cm of the true vertex in the z direction, *i.e.*, $|dz| < 1$ cm. Within 1 cm, even if the chosen vertex is not *the* true primary vertex, the resulting kinematic difference is negligible.

In this study, we add another element called "photon pointing" to increase the chance of finding the true primary vertex. It is best explained together with the concept of "CPS match".

CPS match and photon pointing

When a photon deposits energy in the CPS stereo layers as well as the four layers of the EM calorimeter, the spatial information that these five layers provide can significantly improve the knowledge of the z-position of the photon point-of-origin. Because of the fine segmentation, the spatial resolution of the CPS is excellent. Table 5.1 shows the resolutions in the central region for the CPS and EM layers.

Quantity	CPS	EM1	EM2	EM3	EM4
σ_z	3.7 mm	15 mm	16 mm	6.8 mm	19 mm

Table 5.1: Resolution in z-position for the central preshower detector and the four electromagnetic calorimeter layers for RunII.

For the CPS cluster associated with an EM object, the centroid positions of the CPS cluster and

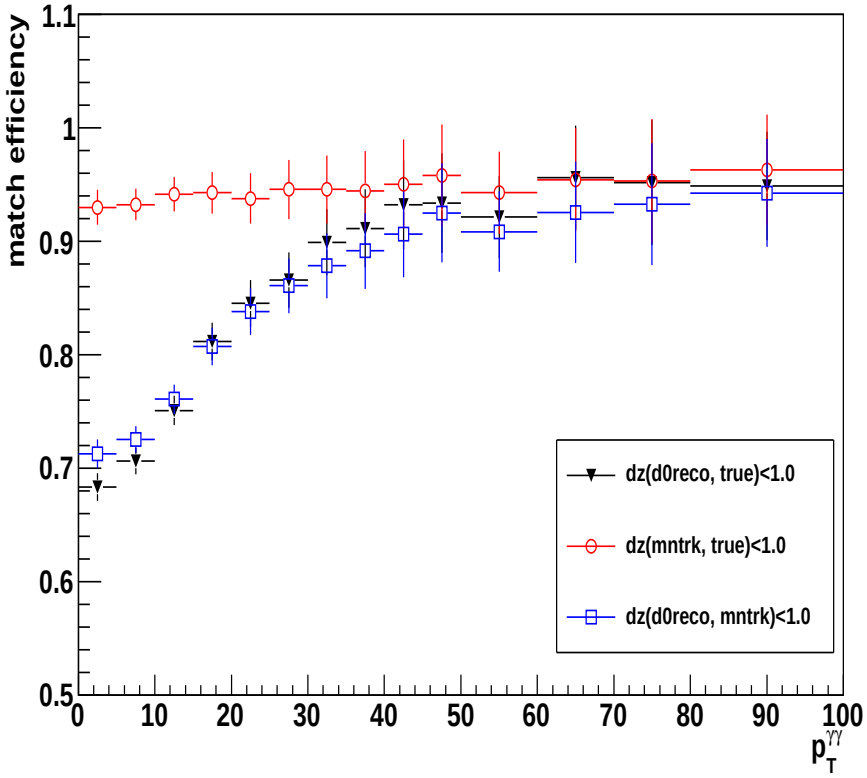


Figure 5.1: The efficiencies for finding the true hard scatter vertex by using the default DØ reconstruction algorithm (d0reco) and max-track algorithm (mntrk) as a function of diphoton transverse momentum in the gluon fusion signal MC sample. They are shown as the black and red lines respectively in the plot. The blue line illustrates the frequency with which the two algorithms yield the same vertex or close enough.

the four EM clusters in the four layers are calculated and then fit to a straight line. If the χ^2 of the fit is less than 25, the EM object is said to have a "CPS match" [91]. The fitted line is extrapolated to the origin to extract the z-coordinate of the EM object's origin, as well as impact parameter. This procedure is called "photon pointing" [92], illustrated in Figure 5.2. We use photon pointing when the photon candidate has a CPS match as discussed below. It is worth mentioning that "photon pointing" merely uses the information from the central preshower detector and the calorimeter while no track information is included. So it can serve as an independent method to locate the hard-scatter vertex of the diphoton event.

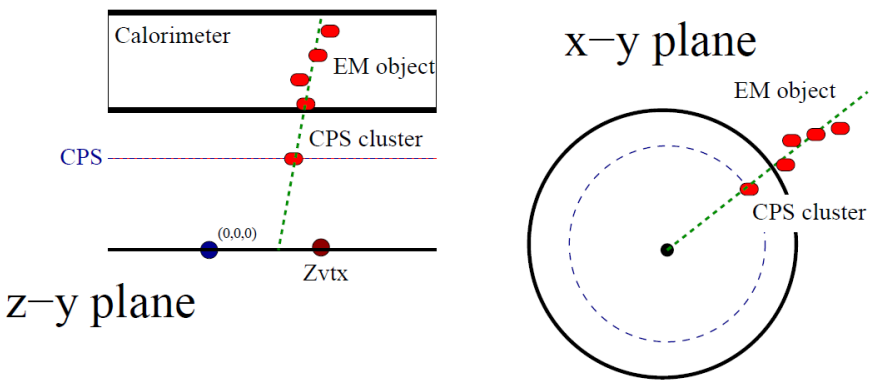


Figure 5.2: A schematic illustration of the photon pointing in the $z-y$ plane (left) and $x-y$ plane (right) [91].

5.1.1 Algorithm

As mentioned above, the max-track vertexing means to identify the vertex with the maximum associated track multiplicity as the hard-scatter vertex. The photon pointing is used in such a way that the vertex from the primary vertex list closest to the pointed z-coordinate is identified as the hard-scatter vertex. The two vertexing methods have strengths in different cases. For example, when no vertex can be found near the photon pointed position, it is better to use max-track vertexing.

The algorithm combines the usage of max-track vertexing and photon pointing as illustrated in the following:

- If neither photon candidate has a CPS match, the max-track vertexing is used.

- If only one photon candidate has a CPS match and the pointing significance[1] is less than 3, the photon pointing from the matched photon is used. Otherwise, the max-track vertexing is used.

- If both photon candidates have a CPS match with the following two conditions satisfied: i) the 2-photon pointing significance[2] is less than 3 and ii) the pointing difference significance[3] is less than 3, then the photon pointing from both photons is used ($< z_p >$ is used to select the hard-scatter vertex from the vertex list). If the two conditions are not satisfied simultaneously, we examine whether one photon pointing significance is larger than 3 and the other's is less than 3. If this is the case, the photon pointing from the photon with less pointing significance is used. Otherwise, we use the max-track vertexing.

5.1.2 Calibration

To ensure the same performance of the algorithm in data and MC, we calibrate the two ingredients of the algorithm, the max-track vertexing and photon pointing, by using $Z/\gamma^* \to e^+e^-$ data and MC samples.

[1]The 1-photon pointing significance S_1 is defined as

$$S_1 = |z_p - z_{vtx}|/\sigma_p,$$

where z_p and σ_p are the pointed z-coordinate and error; z_{vtx} is the z-coordinate of the vertex closest to the pointed position from the photon with CPS match.

[2]The 2-photon pointing significance S_2 is defined as

$$\begin{aligned} S_2 &= | < z_p > - < z_{vtx} > |/ < \sigma_p >, \\ < z_p > &= (z_{1p}\sigma_{2p}^2 + z_{2p}\sigma_{1p}^2)/(\sigma_{1p}^2 + \sigma_{2p}^2), \\ < \sigma_p > &= (\sigma_{1p}\sigma_{2p})/\sqrt{\sigma_{1p}^2 + \sigma_{2p}^2}, \end{aligned}$$

where $z_{1(2)p}$ and $\sigma_{1(2)p}$ are pointed z-coordinates and errors from photon 1 and 2; $< z_{vtx} >$ is the z-coordinate of the vertex closest to the error-weighted pointed position $< z_p >$ from the two photons.

[3]The 2-photon pointing difference significance S_D is defined as

$$S_D = |z_{1p} - z_{2p}|/\sqrt{\sigma_{1p}^2 + \sigma_{2p}^2}.$$

Max-track vertexing calibration

For max-track vertexing, we find that the MC is different from data in three aspects:

- The number of vertices: MC has fewer vertices than data, so for this aspect we reweight the distribution of the number of vertices of MC to that of data.

- The track multiplicity of the hard-scatter vertex: the MC has higher track multiplicity due to higher tracking efficiency. So we apply "track-loose" tracking scale factors from Muon ID [93] to the hard-scatter vertex in MC events to correct its track multiplicity, as shown in Figure 5.3.

- The track multiplicity of the min-bias vertices: it is known that the cluster error of hits of the overlaid events are mistakenly from MC (should be directly from data). So we randomly generate track multiplicities for min-bias vertices by sampling from the track multiplicity distribution in data, as shown in Figure 5.4.

After the three corrections, we evaluate the efficiencies for max-track vertexing to find the correct vertex as a function of diphoton p_T in $Z/\gamma^* \to e^+e^-$ data and MC samples. The true vertex in the $Z/\gamma^* \to ee$ data/MC is defined as the default DØ reco vertex verified by both of the two electron tracks. But when counting the number of tracks for the true vertex, two are subtracted to emulate two photons. The average efficiencies are about the same for data and MC. For p20, we measure $64.6 \pm 0.1\%$ for data and $65.8 \pm 0.1\%$ for MC. For p17, we measure $80.7 \pm 0.3\%$ for data and $80.9 \pm 0.2\%$ for MC. The di-EM p_T dependent efficiencies are shown in Figure 5.5.

Photon pointing calibration

For CPS pointing, we mainly calibrate

- the pointing resolution ($|z_p - z_{vtx}|$), the difference between the z-coordinate from photon pointing and the true vertex

- the pointing significance ($|z_p - z_{vtx}|/\sigma_p$), the pointing resolution divided by pointing error.

To calibrate pointing resolution, we fit the $|z_p - z_{vtx}|$ distribution in $Z/\gamma^* \to e^+e^-$ data and MC samples to measure the resolutions, σ_{data} and σ_{MC}. This is done in four rapidity bins: $|\eta| < 0.4$, $0.4 < |\eta| < 0.6$, $0.6 < |\eta| < 0.8$ and $|\eta| > 0.8$. Figure 5.6 shows the fit in an example rapidity bin, $|\eta| < 0.4$. Then z_p in the MC sample is smeared according to Equation 5.1. Figure 5.7 shows the pointing resolution in MC agrees well with data after the smearing.

$$Z_{\text{new pointing}} = Z_{\text{old pointing}} + RND \cdot \sqrt{\sigma_{data}^2 - \sigma_{MC}^2}, \tag{5.1}$$

where RND is a random number generated from a standard normal distribution N(0,1).

After calibration of the pointing resolution, we continue to examine the pointing significance for both data and MC samples, as shown in Figure 5.8. Because the vertexing algorithm decides when to use the photon pointing based upon the pointing significance, it makes more sense statistically to calibrate the significance distribution to be a standard normal distribution N(0,1). This is done by multiplying the pointing error by the fitted standard deviation from Figure 5.8. Then both data and MC have a standard normal distribution for the pointing significance, shown in Figure 5.9.

After the above calibrations of each element in the new algorithm, the average efficiencies for the new algorithm to find the true vertex for p20 data and MC samples are $94.0 \pm 0.1\%$ and $95.7 \pm 0.1\%$, and that for p17 data and MC samples are $97.6 \pm 0.1\%$ and $98.7 \pm 0.1\%$ respectively. Figure 5.10 shows the efficiency as a function of diphoton p_T.

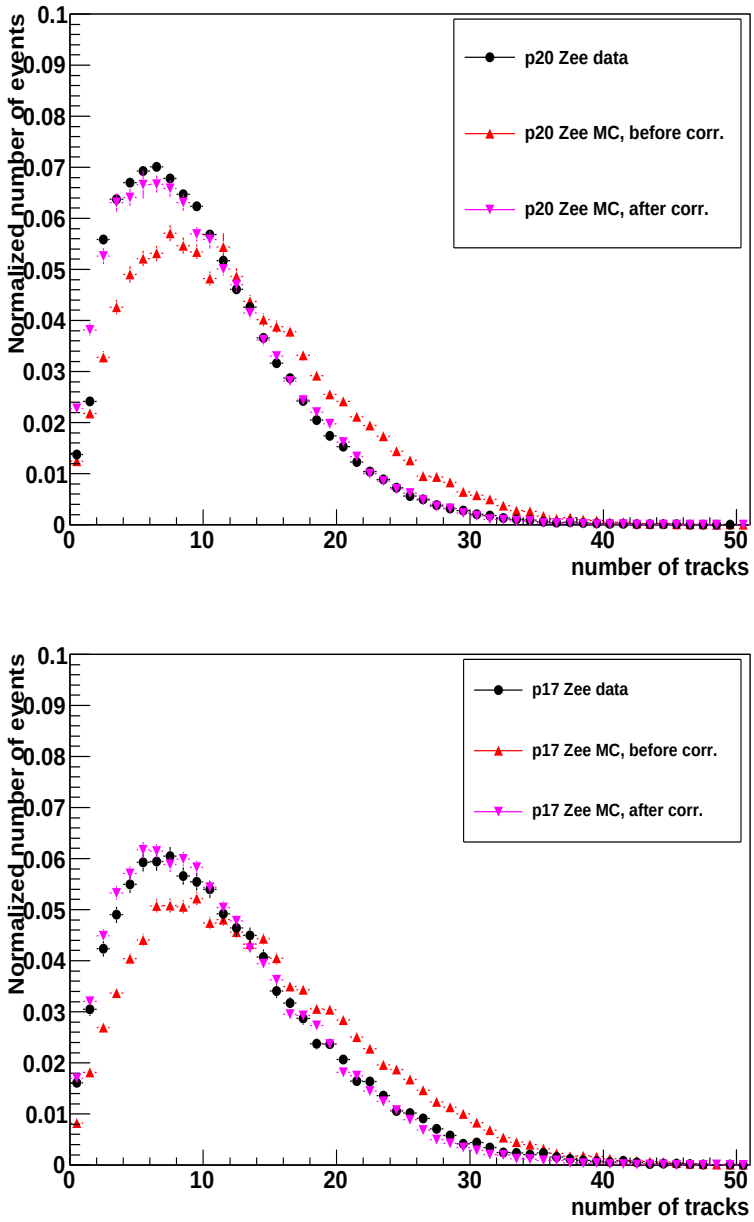


Figure 5.3: Track multiplicity distribution of the hard-scatter vertex in p20 (upper) and p17 (lower) data (black), uncorrected MC (red) and corrected MC (magenta) using $Z/\gamma^* \rightarrow ee$ events.

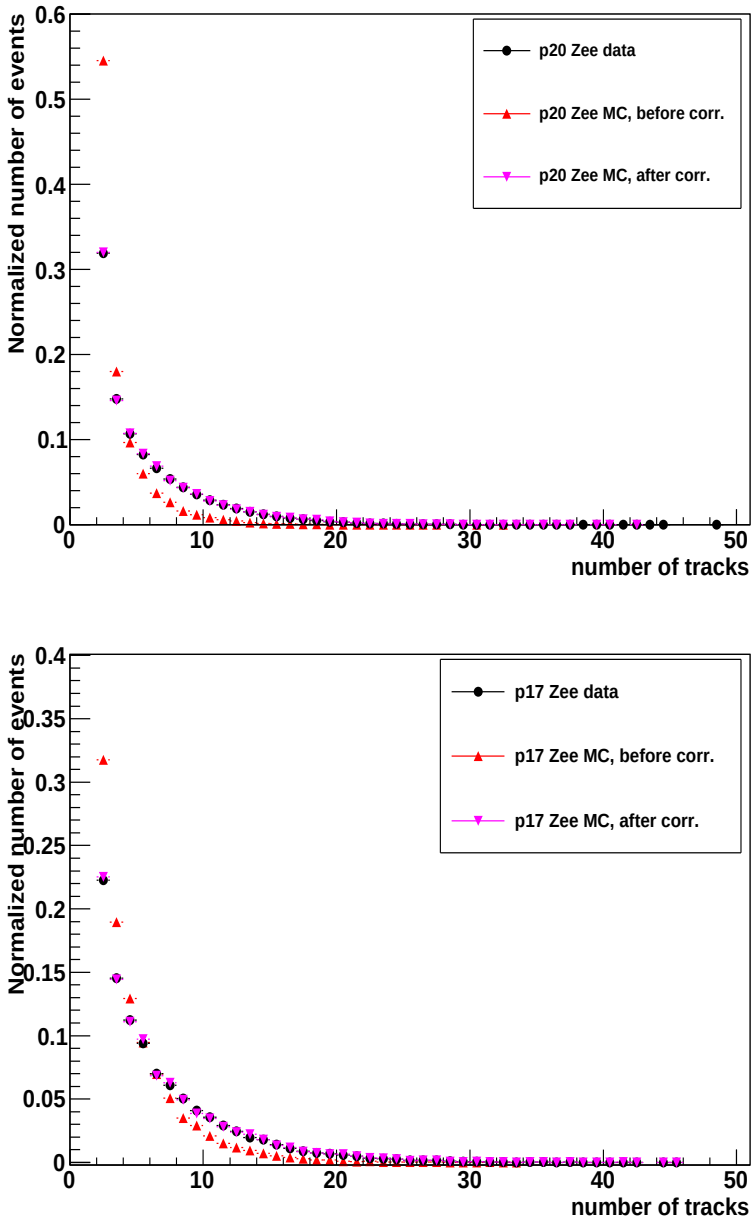


Figure 5.4: Track multiplicity distribution of the min-bias vertices in p20 (upper) and p17 (lower) data (black), uncorrected MC (red) and corrected MC (magenta) using $Z/\gamma^* \rightarrow ee$ events.

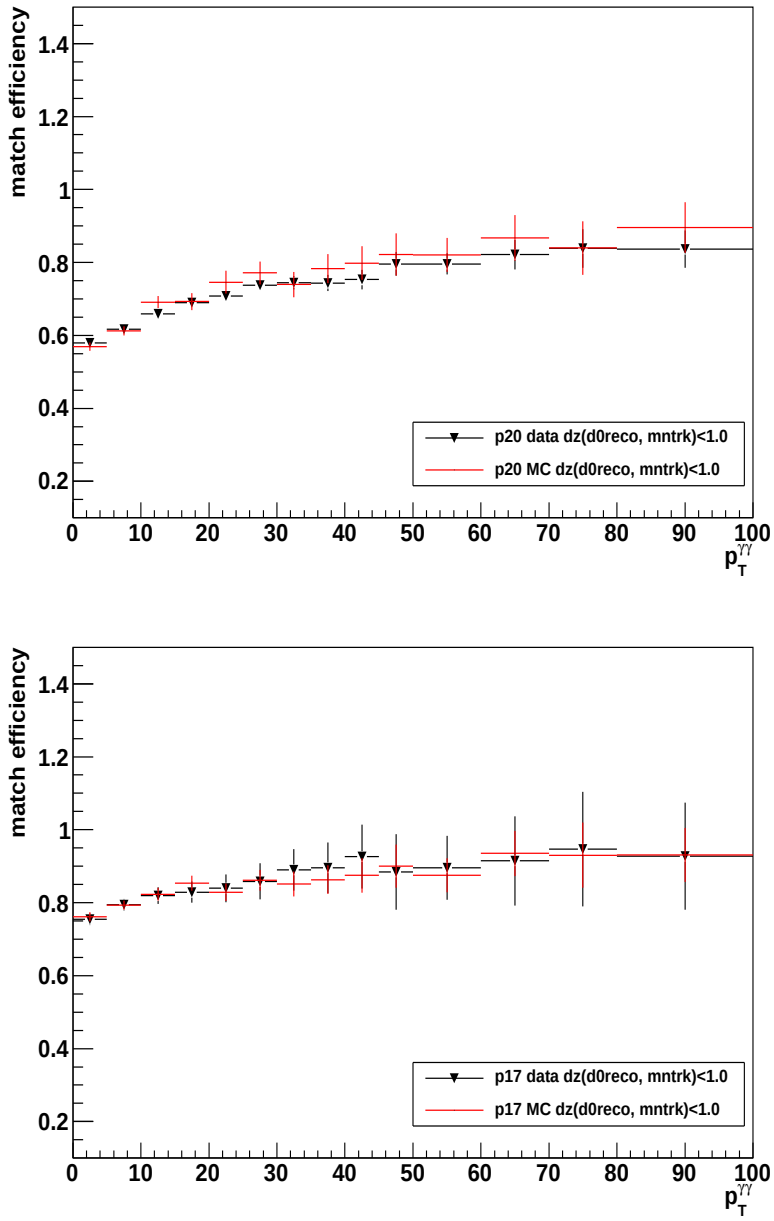


Figure 5.5: The probability of finding the true vertex using calibrated max-track vertexing as a function of di-EM p_T for p20 (upper) p17 (lower) data (black) and MC (red) samples using $Z/\gamma^* \rightarrow e^+e^-$ events.

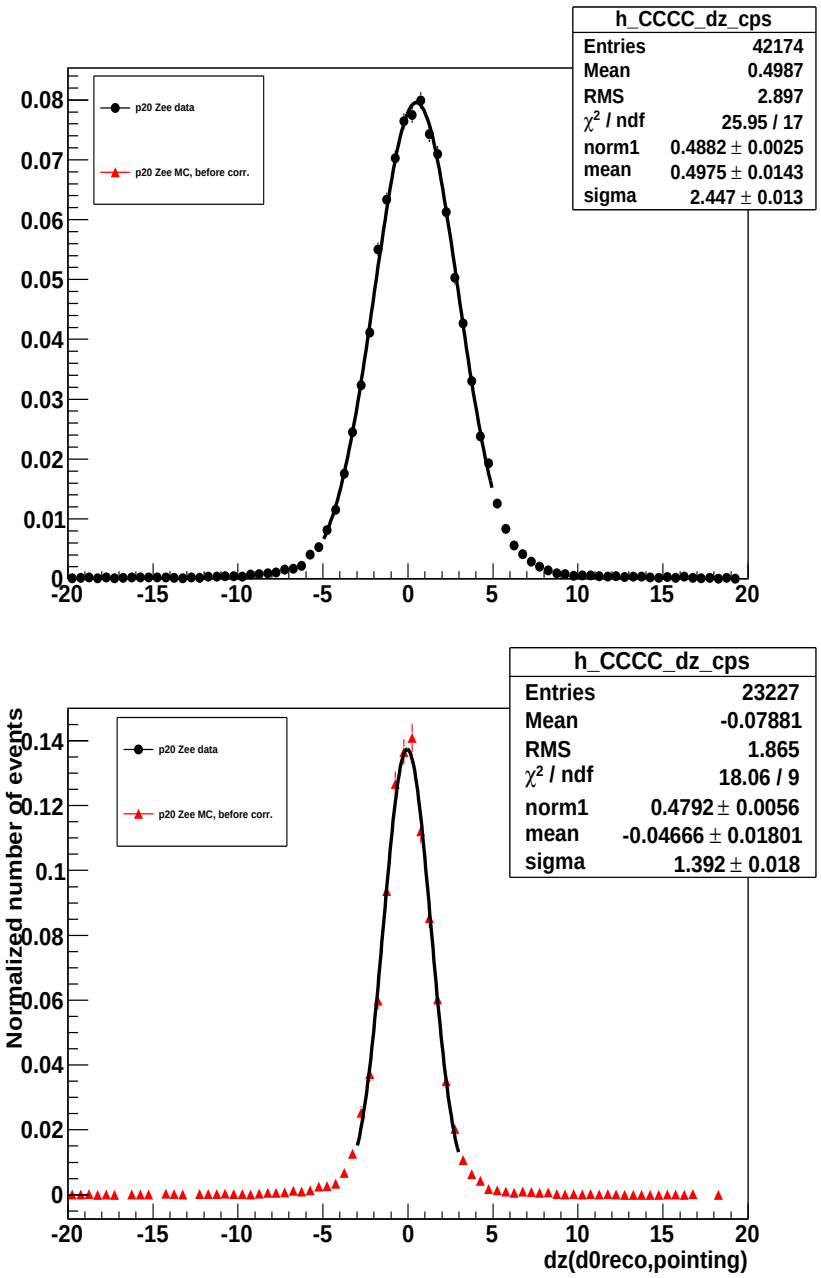


Figure 5.6: Photon pointing resolution fit in an example rapidity bin $|\eta| < 0.4$ for p20 data (upper) and MC (lower) samples using $Z/\gamma^* \rightarrow ee$ events.

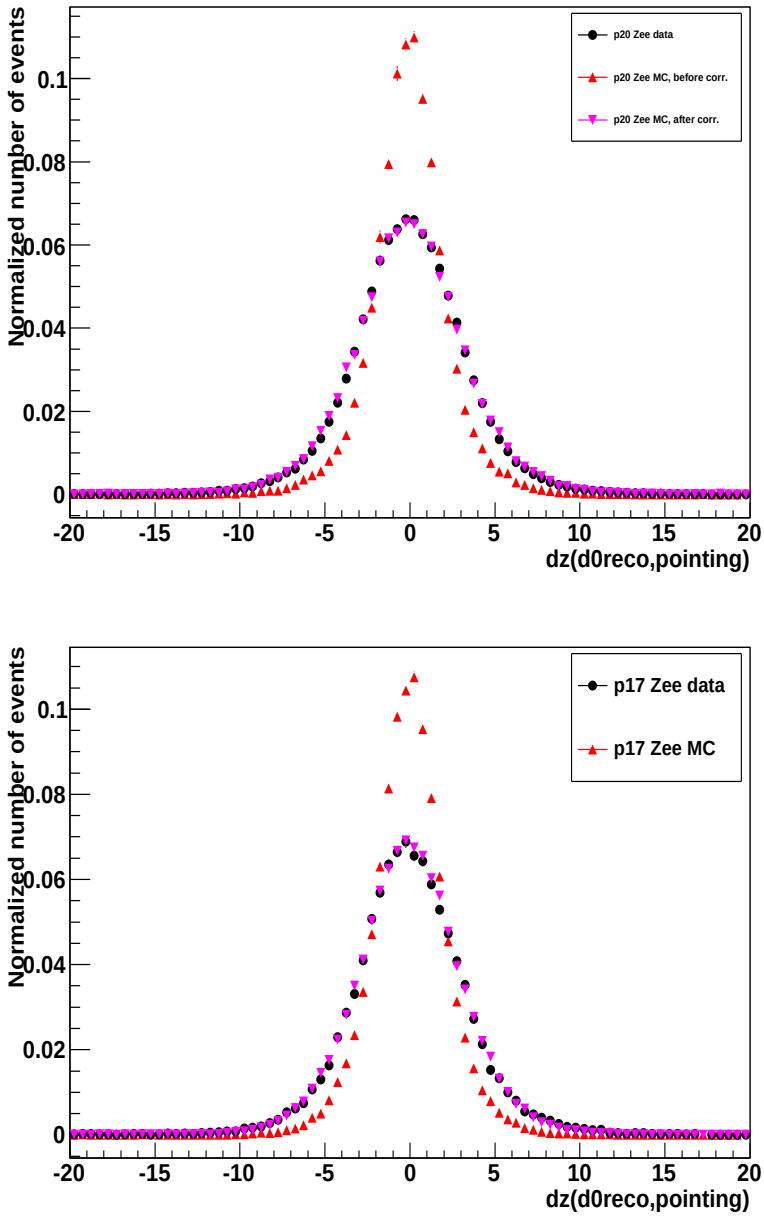


Figure 5.7: Photon pointing resolution after calibration in p20 (upper) and p17 (lower) data (black), uncorrected MC (red) and corrected MC (magenta) samples using $Z/\gamma^* \to e^+e^-$ events.

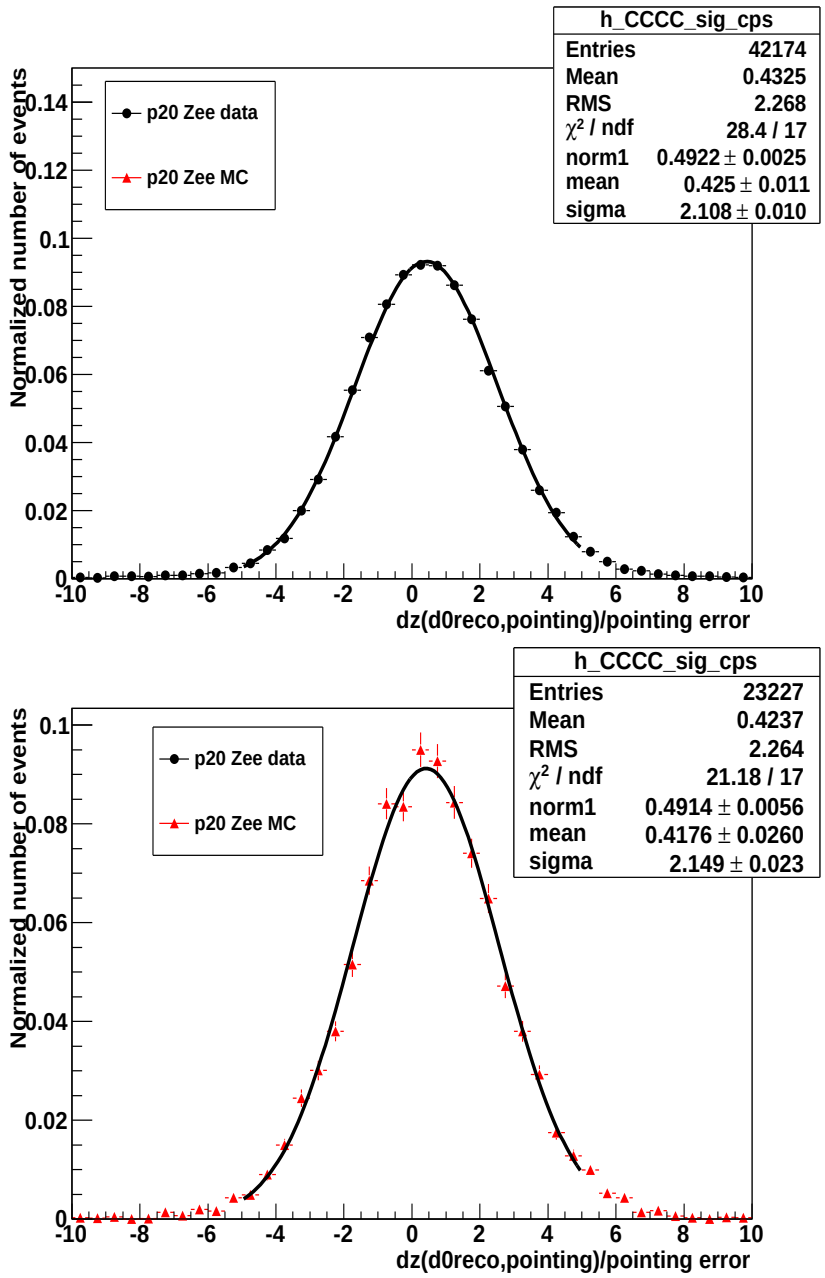


Figure 5.8: Photon pointing significance fit in an example rapidity bin $|\eta| < 0.4$ for p20 data (upper) and MC (lower) samples using $Z/\gamma^* \rightarrow ee$ events.

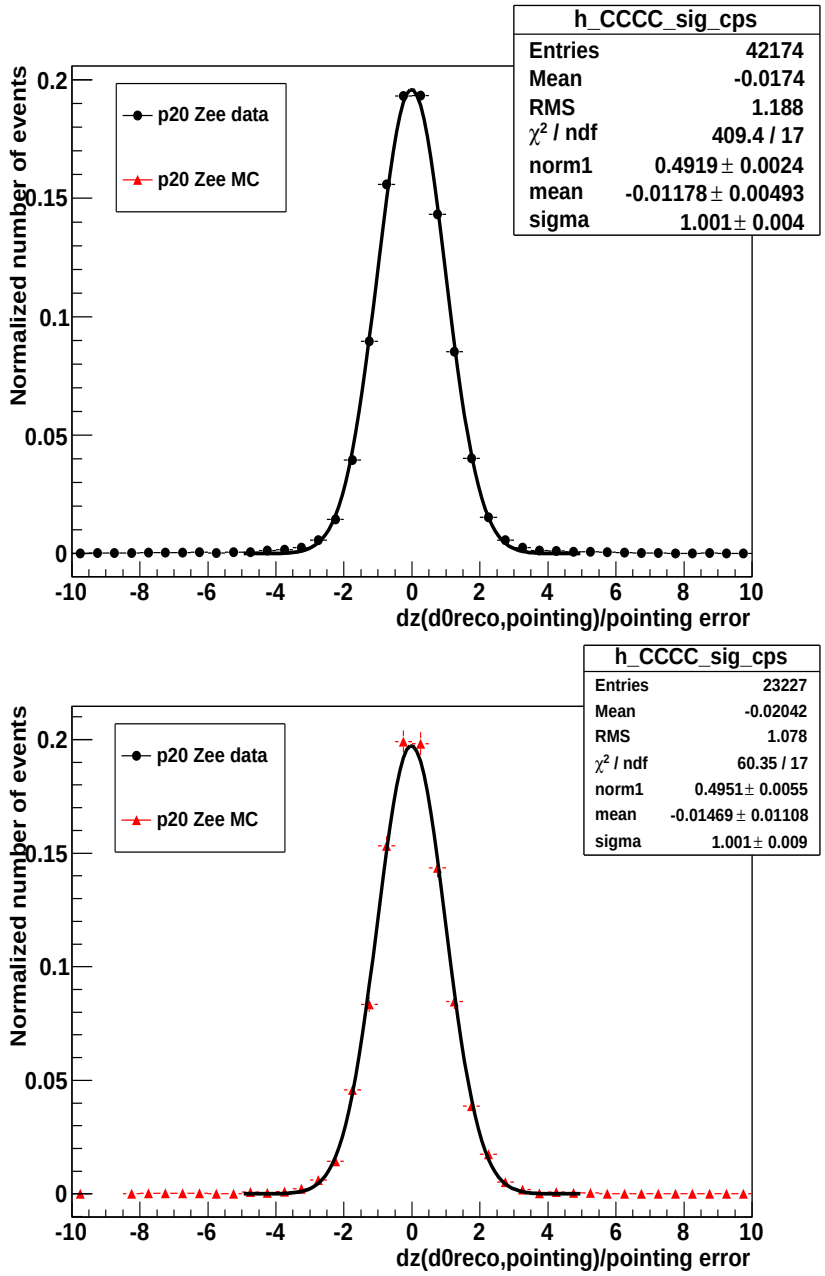


Figure 5.9: Photon pointing significance after calibration in an example rapidity bin $|\eta| < 0.4$ for p20 data (upper) and MC (lower) samples using $Z/\gamma^* \to e^+e^-$ events.

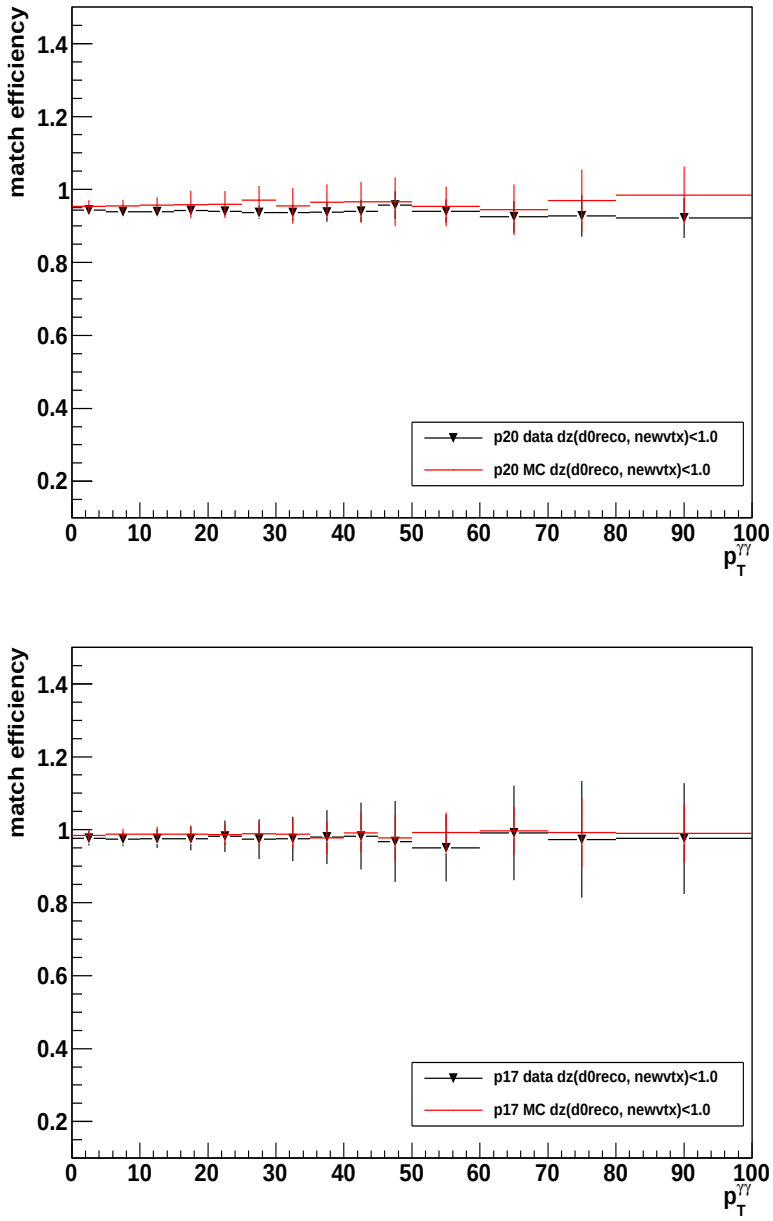


Figure 5.10: The final calibrated algorithm's efficiency of finding the true vertex as a function of di-EM p_T for p20 (upper) p17 (lower) data (black) and MC (red) samples using $Z/\gamma^* \rightarrow e^+e^-$ events.

5.2 Event Selection

To extract interesting diphoton events from the data and MC samples, a set of requirements should be satisfied besides the good quality of data. It includes requirements on the trigger, the hard-scatter vertex position, photon quality and CPS match.

Trigger requirement

Each event is required to fire the di-EM trigger. There are 6 di-EM trigger lists in different data taking epochs: v8-11, v12, v13, v14, v15 and v16. Each trigger list is a combination of the triggers in the three-level trigger system, requiring that there are at least two EM clusters with loose shower shapes and p_T above a threshold varying from 15 GeV to 25 GeV.

The trigger efficiency for v8-v14 is estimated to be 100% at a precision of 0.1% with $M_{ee} > 50$ GeV using $Z/\gamma^* \rightarrow e^+e^-$ events [94]. For v15 and v16 the trigger efficiency is found to be greater than 97% for a di-EM invariant mass of 50 GeV and increases to about 100% at 90 GeV. So in this study, the invariant mass of the two photon candidates is required to be greater than 60 GeV to ensure very high trigger efficiency. The dependence of the trigger efficiency on di-EM invariant mass for v15 and v16 is shown in Figure 5.11 and parametrized empirically as

$$\varepsilon_{\text{trigger}}^{\text{v15,16}}(M_{\text{di-EM}}) = \frac{1}{2}p_2(1.0 + \text{erf}(\frac{M_{\text{di-EM}} - p_0}{\sqrt{2}p_1})) \tag{5.2}$$

$$p_0 = -16.39, \;\; p_1 = 34.55, \;\; p_2 = 0.9995; \tag{5.3}$$

Primary vertex position requirement

The *revertexing* is discussed in Section 5.1 at the beginning of this Chapter. The z-coordinate of the vertex is required to be within 60 cm of the geometrical center of the detector. After the *revertexing* procedure, the photon kinematics and related photon ID variables are updated.

Photon quality requirement

The photon candidates are selected in the central region, with $p_T > 25$ GeV and satisfy the "core1" quality requirement discussed in Section 3.3.2 of Chapter 3.

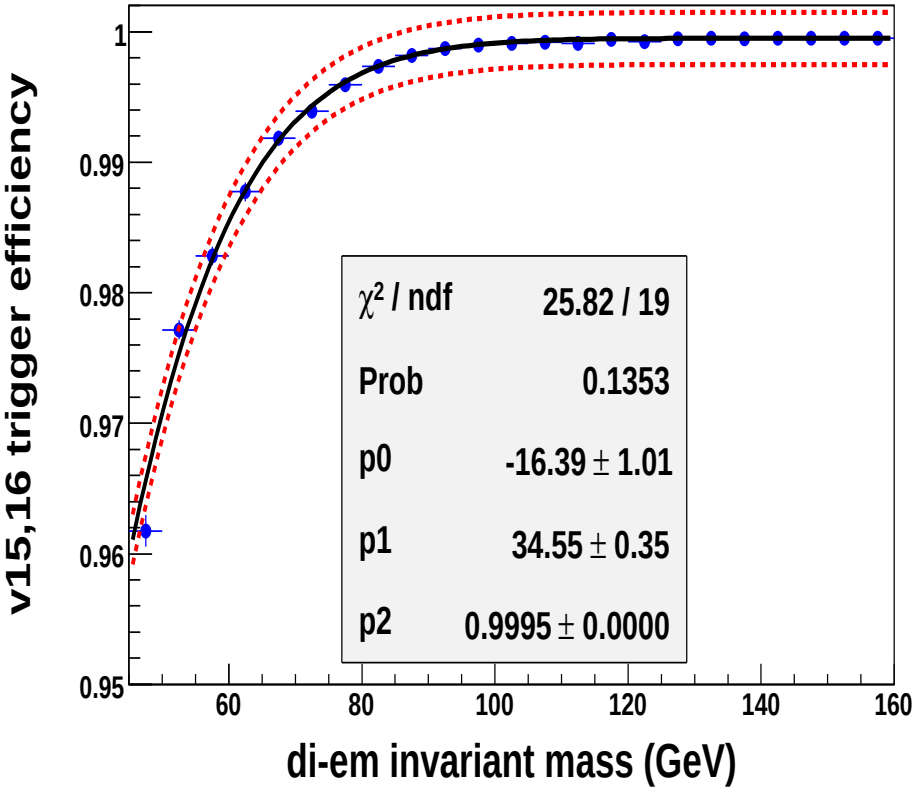


Figure 5.11: v15 and v16 calorimeter-only di-EM trigger efficiency as a function of di-EM invariant mass, where the two red dotted lines illustrate the assigned 0.1% uncertainty.

Tables 5.2 shows the selection efficiencies for the gluon fusion, associated and vector boson fusion production of 130 GeV Higgs MC. The average event selection efficiency for the signal MC events is about 21%.

sample	gluon fusion	associated production	vector boson fusion
p17	0.222 ± 0.001	0.208 ± 0.001	0.233 ± 0.001
p20	0.219 ± 0.001	0.206 ± 0.001	0.234 ± 0.001

Table 5.2: Event selection efficiencies from 130 GeV Higgs MC for p17 and p20.

CPS match requirement

Additionally, we require at least one of the two photon candidates has a CPS match. The CPS matching efficiencies are measured using the method documented in Ref. [62]. Table 5.3 shows the CPS match efficiencies for different core cuts in the central region.

core cut	CC core0	CC core1	CC core2
average efficiency(data)	0.725	0.726	0.730
average efficiency(MC)	0.745	0.748	0.752
scale factor(data/MC)	0.983 ± 0.014	0.980 ± 0.015	0.980 ± 0.016

Table 5.3: CPS match efficiencies and scale factors measured with Zγ data and MC.

5.3 Background Modeling

Three major backgrounds are considered in this study: (i) $Z/\gamma^* \rightarrow l^+l^-(l = e, \tau)$ (ZDY) production, where both electrons are misidentified as photons due to tracking inefficiency; (ii) γ+jet and dijet events, where the jet(s) are misidentified as photon(s); (iii) direct $\gamma\gamma$ production (DDP), which is the irreducible physics background. In this Section, we first describe some important corrections to the Monte Carlo simulations and then describe how each background is modeled.

5.3.1 Corrections

The Monte Carlo simulations have general corrections (luminosity profile correction, and beam z-reweighting), EM cluster pre-selection efficiency [95], photon ID efficiency and CPS match efficiency corrections applied. Here we describe another two corrections that are unique to this study: the ANN output (O_{NN}) shape correction and the unclustered energy correction.

$\mathbf{O_{NN}}$ Shape Correction

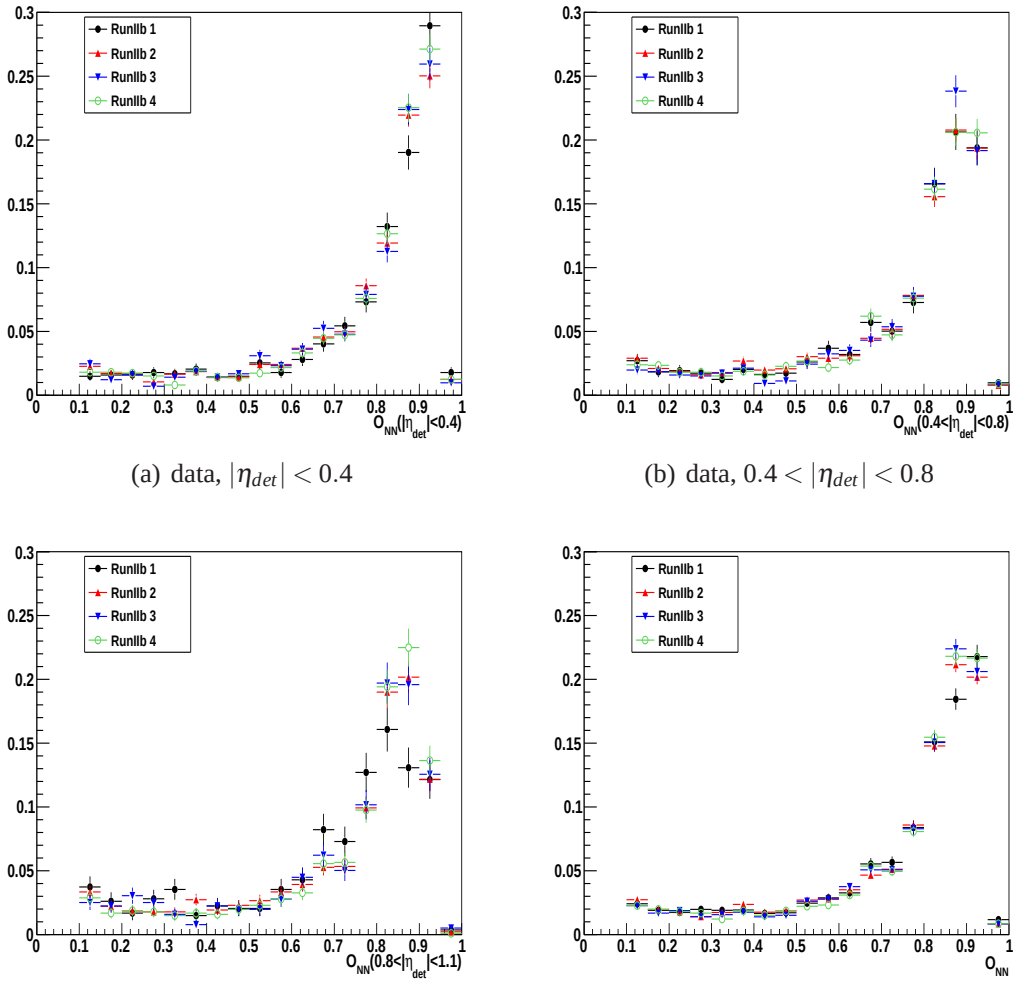
O_{NN} is an essential variable in photon identification and is used extensively in this study, so it is important to calibrate its shape. As mentioned in Section 4.1 in Chapter 4, the RunIIb data epoch can be divided into finer subepochs, namely RunIIb1, 2, 3, and 4. In principle, it is ideal to model each data taking epoch with Monte Carlo simulation of the corresponding reco version. For example, RunIIb1 MC is used to model RunIIb1 data. However, the ideal sometimes cannot be achieved with limited resources. For instance, we only have RunIIb1 MC to model the DDP background.

We observed that the O_{NN} shape varies in different data epochs, as shown in Figure 5.12. More specifically, the O_{NN} shapes in RunIIb2,3,4 data are very similar while they are different from that in RunIIb1 data, which means that RunIIb1 MC with a specific correction can simulate adequately the O_{NN} shapes of the full RunIIb dataset. The correction is shown in Equation 5.4 and

implemented in three regions: $|\eta_{det}| < 0.4$, $0.4 < |\eta_{det}| < 0.8$ and $0.8 < |\eta_{det}| < 1.1$.

$$correction = \frac{\text{RunIIb1 } O_{NN} \text{ shape} \times \mathscr{L}_{\text{RunIIb1}} + \text{RunIIb3 } O_{NN} \text{ shape} \times \mathscr{L}_{\text{RunIIb234}}}{\text{RunIIb1 } O_{NN} \text{ shape} \times \mathscr{L}_{\text{RunIIb1234}}}, \tag{5.4}$$

where $\mathscr{L}_X$ is the integrated luminosity for epoch X.



(a) data, $|\eta_{det}| < 0.4$

(b) data, $0.4 < |\eta_{det}| < 0.8$

(c) data, $0.8 < |\eta_{det}| < 1.1$

(d) data, inclusive

Figure 5.12: The photon O_{NN} distributions in different data taking epochs are compared in different η_{det} bins.

Unclustered Energy Correction

The total unclustered energy in the transverse plane (UET) is the total scalar p_T of calorimeter cells not associated by the d0correct algorithm to a clustered object. UET is used in the calculation of $\not\!\!E_T$, so a calibration of UET results in a better modeling of $\not\!\!E_T$. The correction is derived in this way: first subtract the UET distribution of data from that of the data-based background, and then divide it by the MC-based background after normalizing the MC-based background to the same integral. The final ratio is the correction that should be applied to MC samples. After correction, the UET distribution in background agrees well with data as shown in Figure 5.13.

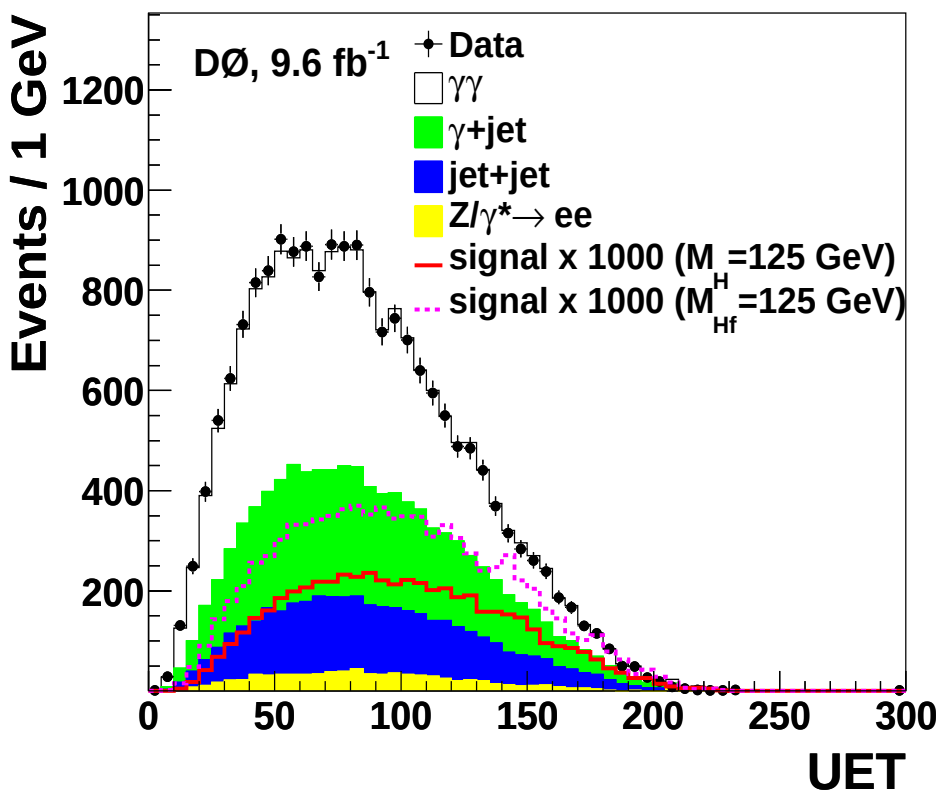


Figure 5.13: UET distribution after correction to MC simulations.

5.3.2 $Z/\gamma^* \to ee$ Drell-Yan (ZDY) Background

The $Z/\gamma^* \to ee$ background is estimated from Monte Carlo simulation using ALPGEN [84], with showering and hadronization from PYTHIA, and using the NNLO cross section [85] for normaliza-

tion. Besides the general corrections, the MC events are also corrected for the "electron no-track" efficiency. This efficiency measured the probability of an electron that fails the track veto to fake a photon. We find that the probability is about 1-3%, depending on azimuthal angle ϕ [64].

5.3.3 γ+jet and Di-jet Backgrounds

The γ+jet and dijet backgrounds are estimated with data. The estimate is performed in two steps. In the first step, a so-called "4×4 matrix method" is used to estimate the yields of this background. It basically decomposes the data events into components belonging to each of the three backgrounds. One drawback of this approach is the limited statistics of data. Therefore in the second step, we select an orthogonal data sample, or control data sample, by reversing certain requirements in the event selection. In this way, we can significantly boost the statistics for this background, and more importantly, the shape of this background can be estimated with higher statistical precision.

4x4 matrix method

The idea of the "4×4 matrix method" is to use the observed classified data events to calculate the background composition of the data through an efficiency matrix [96].

Following the event selection, a tightened O_{NN} requirement[4] ($O_{NN} > 0.75$) is used to classify the events into four classes: (i) both photons, (ii) only the photon with the highest p_T (leading photon), (iii) only the photon with the second highest p_T (trailing photon), or (iv) neither of the two photons, pass this requirement.

The corresponding numbers of events, after subtracting the ZDY contribution, are denoted as (i) N_{pp}, (ii) N_{pf}, (iii) N_{fp} and (iv) N_{ff}. The different efficiencies of the $O_{NN} > 0.75$ requirement for photons (ε_γ) and jets (ε_{jet}) allow us to estimate the sample composition by solving a system of linear equations 5.5:

$$(N_{pp}, N_{pf}, N_{fp}, N_{ff})^T = \mathscr{E} \times (N_{\gamma\gamma}, N_{\gamma j}, N_{j\gamma}, N_{jj})^T, \tag{5.5}$$

[4]Note that this requirement is not an event selection requirement. It can be viewed as a boundary to classify events.

where $N_{\gamma\gamma}$ (N_{jj}) is the number of $\gamma\gamma$ (jj) events and $N_{\gamma j}$ ($N_{j\gamma}$) is the number of γj events with the leading (trailing) cluster as the photon. The 4×4 matrix $\mathscr{E}$ is constructed with the efficiency terms ε_γ and ε_{jet} as in Equation 5.6, parametrized as a function of $|\eta|$ for each photon candidate and estimated in photon and jet MC samples. We validate ε_γ with data of radiated photons from charged leptons in Z boson decays ($Z \to l^+l^-\gamma, l = e, \mu$) and ε_{jet} with jet data [90].

$$\begin{pmatrix} (1-\varepsilon_{j1})(1-\varepsilon_{j2}) & (1-\varepsilon_{j1})(1-\varepsilon_{\gamma 2}) & (1-\varepsilon_{\gamma 1})(1-\varepsilon_{j2}) & (1-\varepsilon_{\gamma 1})(1-\varepsilon_{\gamma 2}) \\ (1-\varepsilon_{j1})\varepsilon_{j2} & (1-\varepsilon_{j1})\varepsilon_{\gamma 2} & (1-\varepsilon_{\gamma 1})\varepsilon_{j2} & (1-\varepsilon_{\gamma 1})\varepsilon_{\gamma 2} \\ \varepsilon_{j1}(1-\varepsilon_{j2}) & \varepsilon_{j1}(1-\varepsilon_{\gamma 2}) & \varepsilon_{\gamma 1}(1-\varepsilon_{j2}) & \varepsilon_{\gamma 1}(1-\varepsilon_{\gamma 2}) \\ \varepsilon_{j1}\varepsilon_{j2} & \varepsilon_{j1}\varepsilon_{\gamma 2} & \varepsilon_{\gamma 1}\varepsilon_{j2} & \varepsilon_{\gamma 1}\varepsilon_{\gamma 2} \end{pmatrix} \tag{5.6}$$

The method is implemented on an event-by-event basis, resulting in weights corresponding to γ+jet, dijet and $\gamma\gamma$. The sum of the weights for γ+jet and dijet over all events are the estimated yields for the γ+jet and dijet background respectively. Table 5.3.3 shows the classification and results of the 4×4 matrix method.

(a)

N_{ff}	2719
N_{fp}	3600
N_{pf}	4916
N_{pp}	7796
Data	19031

(b)

N_{DY}	795 ± 17
N_{jj}	3330 ± 222
$N_{\gamma j} + N_{j\gamma}$	5738 ± 406
$N_{\gamma\gamma}$	9168 ± 262
Data	19031

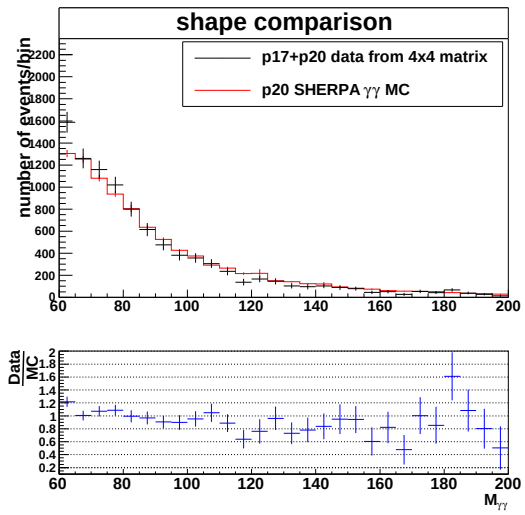
Table 5.4: (a) shows the number of events in the data for the 4 categories. (b) shows the number of $\gamma\gamma$, γ+jet and dijet events in the data from the 4x4 matrix method. The number of DY events is estimated from MC. The quoted uncertainties are statistical only.

The 4×4 matrix method not only provides the estimation for background yields but also a validation of shape modeling for each of the $\gamma\gamma$, γ+jet and dijet backgrounds, even though statistically limited . Figure 5.14 shows the comparison of backgrounds from the 4×4 matrix method with those from the more statistics-enriched samples discussed in this Section.

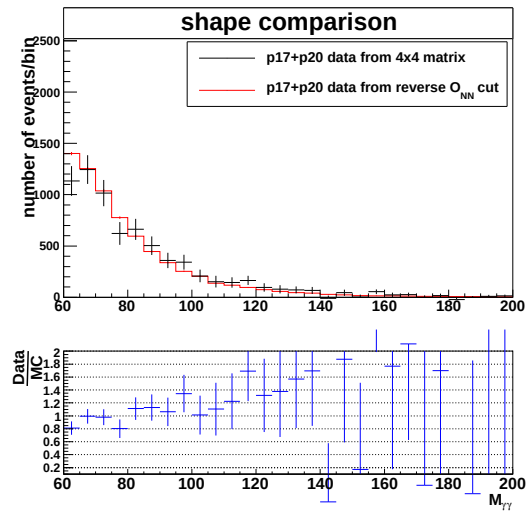
Orthogonal sample

Due to the limited statistics of the matrix method, especially in the high mass region, we obtain the

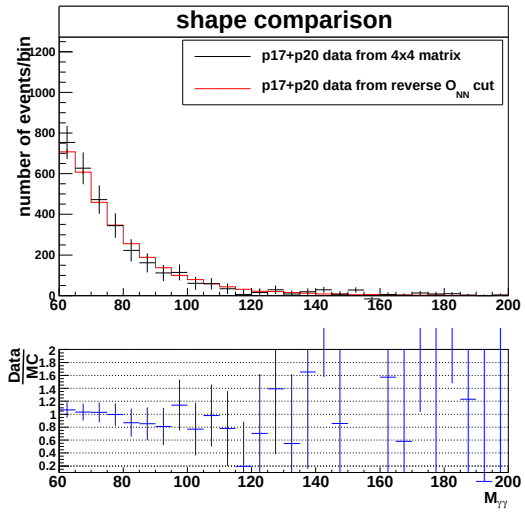
shapes of the γ+jet and dijet backgrounds from an orthogonal data sample by reversing the O_{NN} requirement in the event selection (*i.e.*, $O_{NN} < 0.1$), dropping the *sigphi* requirement, and relaxing the *trkiso* requirement to 5 GeV. The shapes of kinematic distributions for the γ+jet background are obtained by requiring one of the two photon candidates to satisfy $O_{NN} < 0.1$. And the shapes of kinematic distributions for dijet background are obtained by requiring both of the two photon candidates to satisfy $O_{NN} < 0.1$.



(a) $\gamma\gamma$ vs SHERPA $\gamma\gamma$ MC



(b) γ+jet vs orthogonal sample



(c) dijet vs orthogonal sample

Figure 5.14: Validation of the shape of the $M_{\gamma\gamma}$ distribution in the CCCC region. The $\gamma\gamma$ background from the SHERPA MC and the γ+jet and jet+jet backgrounds from the orthogonal sample are compared to their corresponding 4x4 matrix method solutions.

5.3.4 Direct Di-photon Production (DDP)

The hadronic SM processes are a copious source of diphoton events, in which high momentum photon pairs can be produced and constitute the physics background in the form of a steeply-falling continuum mass spectrum. They are produced from scatterings of quark-antiquark, (anti)quark-gluon, and gluon-gluon. Figure 5.15 [97] shows the possible Feynman diagrams for direct diphoton production (DDP). Figure 5.15 (a) shows the lowest order process where a photon pair is produced from $q\bar{q}$ annihilation. The process involves radiative corrections in Figures 5.15 (b)-(e), (h), (i)-(l) of $\mathscr{O}(\alpha_s)$, $\mathscr{O}(\alpha_s^2)$ and $\mathscr{O}(\alpha_s^3)$ in the strong coupling strength, respectively. Figures 5.15 (f) and (g) represent the single-photon fragmentation scenarios.

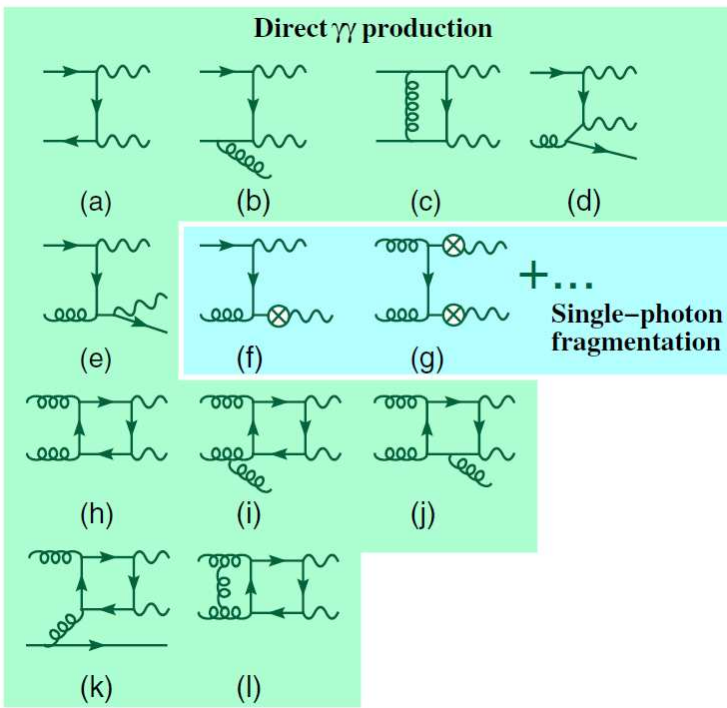


Figure 5.15: Feynman diagrams that contribute to the direct diphoton production. Solid lines are quarks and antiquarks. Wavy and curly lines represent photons and gluons. (a)-(e) and (h)-(l) are considered as leading order and next-to-leading order processes respectively in Ref. [97]. The diagrams of (f) and (g) are single and double fragmentation.

In practice, we choose SHERPA [87] to model DDP because the DDP differential cross section measurement [90, 98] shows that this process is best modeled by the Simulation of High-Energy Reactions of Particles, namely SHERPA. The photon fragmentation function in SHERPA is modeled

by an interleaved QCD+QED parton shower including higher-order real-emission matrix elements.

Figure 5.16 shows good agreement between data and SHERPA predictions.

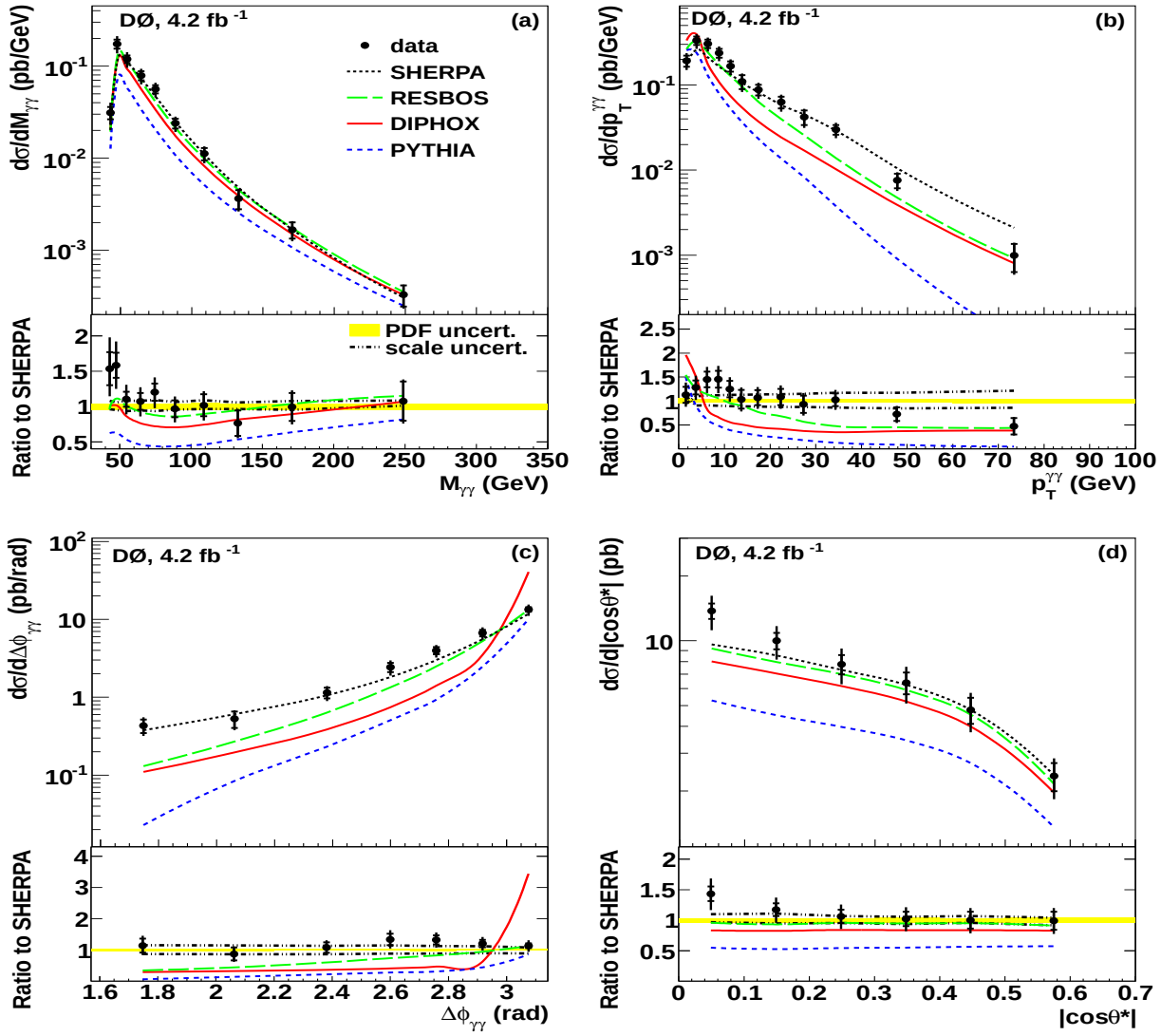


Figure 5.16: An example of the measured differential cross section in direct diphoton production as a function of $M_{\gamma\gamma}$, $p_T^{\gamma\gamma}$, $\Delta\phi_{\gamma\gamma}$ and $|\cos\theta^*|$ ($\cos\theta^* = \tanh[(\eta_1 - \eta_2)/2]$). The data are compared to the theoretical predictions from RESBOS, DIPHOX, PYTHIA AND SHERPA. The ratio of the differential cross sections between data and SHERPA are displayed as black points with uncertainties in the bottom plots [98].

5.3.5 Diphoton Invariant Mass Distribution

The signature of the Higgs boson in $H \rightarrow \gamma\gamma$ is a narrow resonance on top of a smoothly falling background in the diphoton invariant mass spectrum. Therefore, examining the mass spectrum for a narrow resonance is an important and robust approach. We use this approach as a cross check with the multivariate technique for each assumed Higgs mass. The data and modeled background are shown in Figure 5.17. The SM signal yield is scaled up by a factor of 100 to be visible in the figure. The modeled background agrees well with the data.

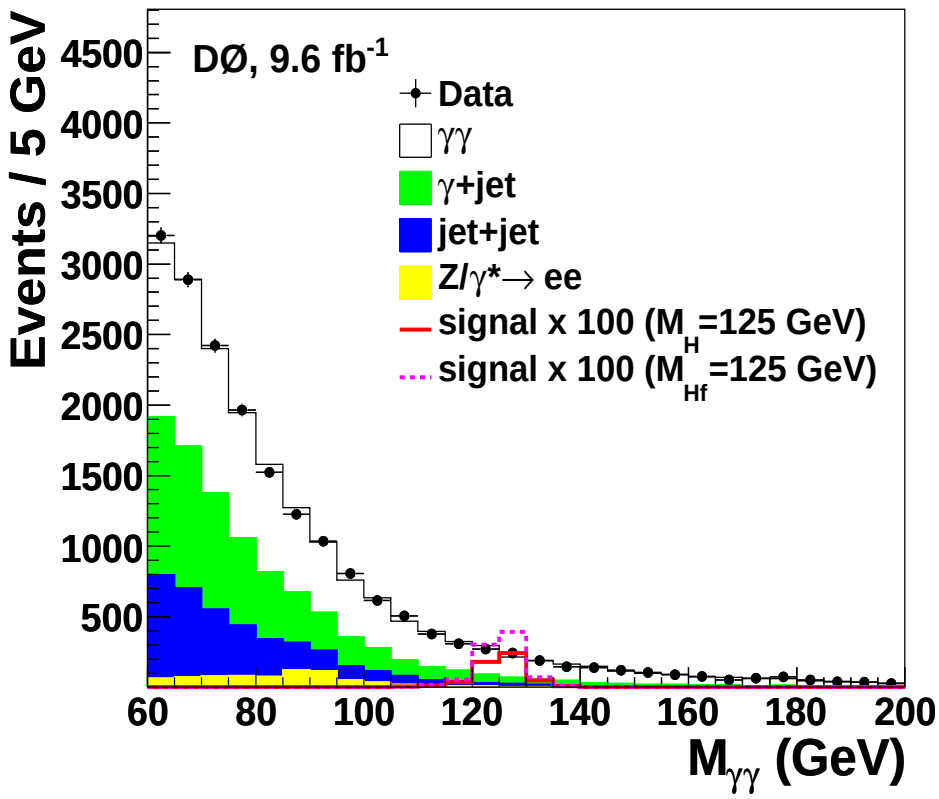


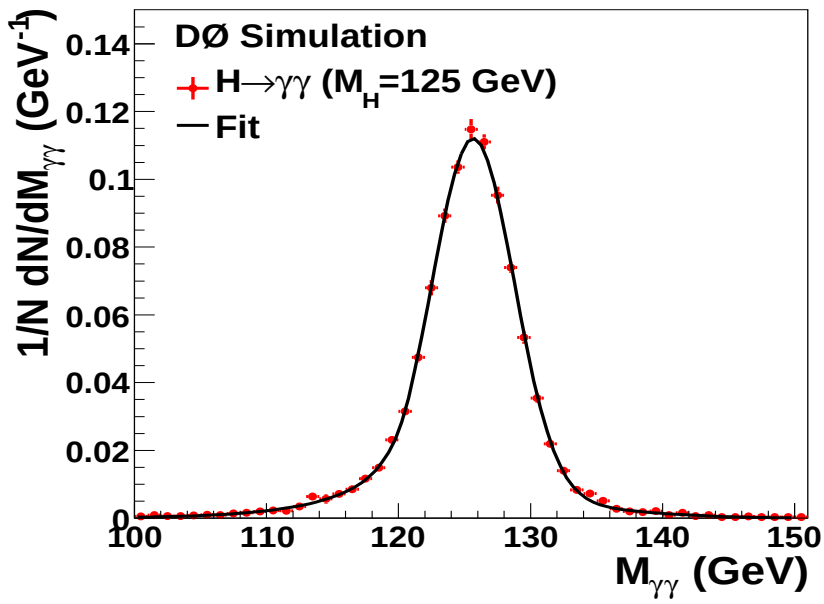
Figure 5.17: Diphoton invariant mass distribution of signal, background and data. A hypothetical standard model signal at 125 GeV is shown. The signal yield is scaled by a factor of 100 for better visualization.

The diphoton mass resolution is ~ 3 GeV. It is determined from a fit to the invariant mass spectrum. The fit function is a sum of a Crystal Ball function [99] as shown in Equation 5.7 and a Gaussian function. The former is used to model the narrow resonance and tails toward the lower mass region, and the latter with a wider standard deviation to model the right tail and outliers.

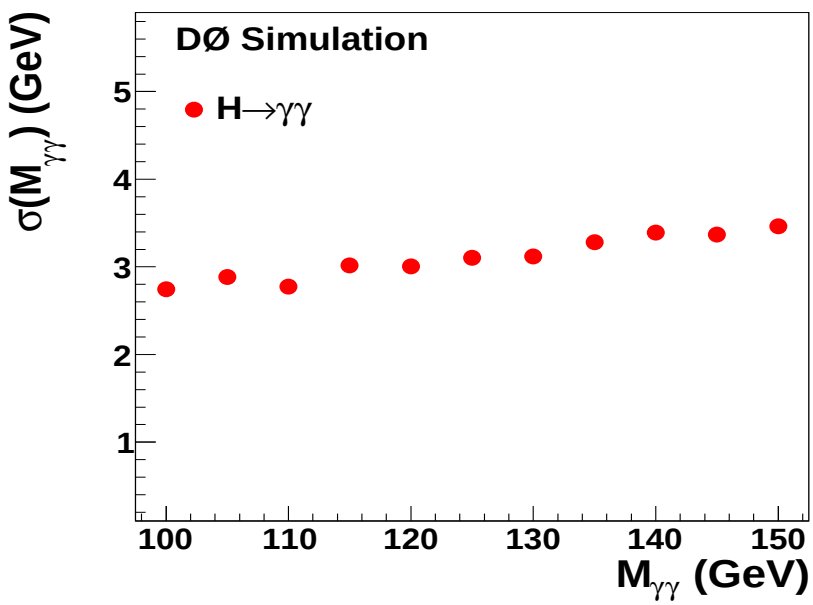
Figure 5.18 (a) shows an example fit for the 125 GeV signal. The mass resolution is extracted from the core component of the fit function, namely the width of the Crystal Ball function. The mass resolution as a function of assumed Higgs mass is shown in Figure 5.18 (b).

$$
\begin{aligned}
t &= (m_{\gamma\gamma} - \mu_{CB})/\sigma_{CB}, \\
f &= \begin{cases} N \cdot e^{-t^2/2} & t > -\alpha_{CB} \\ (\frac{n_{CB}}{\alpha_{CB}})^{n_{CB}} \cdot (\frac{n_{CB}}{\alpha_{CB}} - \alpha_{CB} - t)^{-n_{CB}} \cdot e^{-\alpha_{CB}^2/2} & \text{otherwise.} \end{cases}
\end{aligned}
\tag{5.7}
$$

where μ_{CB}, σ_{CB} are the mean and width of the Crystal Ball function; α_{CB}, n_{CB} dictate the location and shapes of the non-Gaussian tail; N is the normalization parameter.



(a)



(b)

Figure 5.18: (a) Reconstructed diphoton invariant mass distribution for a SM Higgs boson signal MC sample with $M_H = 125$ GeV. The red line represents the fit to the distribution described in the text. (b) Mass resolution as a function of diphoton invariant mass from the fit to SM Higgs boson signal MC samples.

5.4 Multivariate Analysis

Traditional cut-based techniques rely on a series of cuts on event selection variables to discriminate signal from background. For example, the searches for $H \to \gamma\gamma$ in Ref. [65] and Ref. [89] only examine the diphoton invariant mass for a narrow resonance. However, the separation power of the traditional technique is not optimized because the signal can be different from the background in several variables. In this study, we adopt a multivariate technique combining information from a list of variables into a final discriminant to achieve better signal significance $S/\sqrt{B}$.

There are several multivariate techniques such as a neural network. The one we used is a boosted decision tree (BDT), implemented with the Toolkit for Multivariate Analysis (TMVA) [100]. The BDT technique is extensively used in High Energy Physics, such as in the single top quark measurement [101]. In this Section, we first describe the BDT technique, and then show the selection of variables used in the BDT and the final discriminant.

5.4.1 Decision Tree and Boosting

Decision Tree

A decision tree is a supervised machine learning technique to extend a simple cut-based approach to a multivariate technique by recursively separating events based on a binary decision and continuing to analyze events that fail a particular criterion.

The usage can be split into two phases. Firstly, the decision tree needs to learn the difference between signal and background, and build a tree-structured classifier. This phase is called *training*. Training is performed with known signal and background samples. After training, the decision tree is implemented on statistically independent, signal and background samples. More importantly, the decision tree is also applied on data to separate data into signal-like and background-like events. After the second phase, the final discriminant distribution is populated and examined for a signal excess.

A *decision tree* is a binary tree. Consider the training phase, all the known signal and back-

ground events come to the first binary separation point, the *root node*. Given a list of n variables of interest $x_1, x_2, \ldots x_n$, for each variable x_i, the splitting value that gives the "best separation" of the events is found. Then the variable that gives the best separation is selected and the splitting value is the one found in the previous step. The root node is now split into two collections of events, or *child nodes*, one with mostly signal events and the other with mostly background events. This procedure is applied recursively to the two child nodes until "some condition" is met to stop the splitting. The terminal node is called a $leaf$.

There are two concepts above worthy of more explanation: the definition of best separation and the conditions that terminate the splitting.

At each node, the signal *purity* p is defined as the sum of signal weights divided by the sum of signal and background weights. When the node is composed of all signal or background, p gives two different values, 1 or 0. However, we need a measure of the *impurity* such that it reaches its maximum when signal and background are equally mixed, and it falls to its minimum when there are only signal events, or background events by symmetry. One of the popular choices is called the *Gini* Index, or Gini for short, as defined in Equation 5.8. As splitting moves forward, the impurity or Gini should decrease to achieve the goal of separation of signal and background. A desired splitting method, or best separation, should maximize the decrease of impurity. This is done by maximizing the difference of a parent node Gini and purity-weighted child nodes Gini, as shown in Equation 5.9.

$$Gini = 1 - p^2 - (1-p)^2 = 2p(1-p). \tag{5.8}$$

$$\Delta(impurity) = Gini(parent) - [p_A \cdot Gini(child\ A) + p_B \cdot Gini(child\ B)]. \tag{5.9}$$

The splitting of nodes would continue if it is not terminated, until each node contains either one signal event or one background event. This certainly results in 100% classification of signal and background events in the training phase but it loses its generality of classification on the unknown events, which is called *over-training*. So a minimum number of events in a node is required to minimize over-training. Additionally, it is not useful to further split nodes if the improvement on

the signal purity is under a certain threshold. If either of the two conditions above is met, the splitting is terminated and the training of a decision tree is finished. The output for a certain event from a decision tree is given by the purity of the terminal node it finally falls in.

Boosting

The terminal nodes can be categorized as signal leaves or background leaves according to the purity. The non-zero signal purity in a background leaf indicates misclassification of signal events, which affects the performance of a decision tree. This is discussed and solved by the introduction of a technique called *boosting* in the 90's [102]. The idea behind boosting is to assign a larger weight to the misclassified events, and train a new decision tree with the modified event weight, which allows the new tree to "learn" harder on the misidentified events. The procedure is performed iteratively, resulting in a collection of trees. The output of the decision trees is the weighted output of single decision trees.

A popular boosting method is called AdaBoost. After a tree T_n is trained, its associated error err_n is calculated as the fraction of the sum of misclassified event weights, as in Equation 5.10.

$$err_n = (\sum_i w_i \times I_n) / \sum_i w_i, \tag{5.10}$$

where w_i is the event weight and I_n is an indicator function for tree T_n. $I_n = 1$ for misclassified events, otherwise it is zero.

Then the tree weight α_n is defined from err_n as in Equation 5.11.

$$\alpha_n = \beta \ln(\frac{1 - err_n}{err_n}), \tag{5.11}$$

where β is a parameter to provide users with the option of adjusting the boosting strength.

For each misclassified event, its weight is multiplied by a factor of e^{α_n}. The sum of the total weights of the entire sample should be renormalized so that for each tree the number of events remains constant. Finally, the output of an event from the N boosted decision trees (T) is diluted

by the weighted outputs of single trees as shown in Equation 5.12.

$$T = \sum_{n=1}^{N} \alpha_n T_n. \tag{5.12}$$

The mathematical meaning of the boosting procedure is to minimize the deviation of the model response ($F(\mathbf{x})$), namely the T in the AdaBoost example above from the true value, y, obtained from the training sample. The measure of the deviation is called the *loss function* $L(F,y)$. It can be shown that the boosting procedure is dictated by the loss function. For example, the AdaBoost method is derived from the loss function $L(F,y) = e^{F(\mathbf{x})y}$, which leads to the reweighting procedure described in Equations 5.10-5.11. The exponential loss function has drawbacks of lack of robustness in the presence of outliers or mislabeled events. To overcome this weakness, the TMVA toolkit implements a technique called *GradientBoost* with the loss function of Equation 5.13. The boosting procedure dictated by this loss function cannot be obtained as straightforwardly as that for AdaBoost, but has to use a "steepest descent" step for the minimization, in which the gradient of the loss function needs to be calculated.

$$L(F,y) = \ln(1 + e^{-2F(\mathbf{x})y}). \tag{5.13}$$

In this study, we find that the decision trees with gradient boosting provide the best performance for rejecting background.

5.4.2 Input Variables and Training

Input Variables

Ten variables including the diphoton invariant mass are used as input variables to train decision trees:

- diphoton invariant mass, $M_{\gamma\gamma}$
- leading photon transverse momentum, p_T^1

- trailing photon transverse momentum, p_T^2
- diphoton transverse momentum, $p_T^{\gamma\gamma}$
- azimuthal angle between the two photon candidates, $\Delta\phi_{\gamma\gamma}$
- $\cos\theta^*$, in the Collins-Soper frame [103]
- ϕ^*, in the Collins-Soper frame
- missing transverse energy, $\not{E}_T$
- leading photon ANN output, O_{NN}^1
- trailing photon ANN output, O_{NN}^2

Complementary to $M_{\gamma\gamma}$, the kinematic variables of p_T^1, p_T^2, $p_T^{\gamma\gamma}$, $\Delta\phi_{\gamma\gamma}$ separate signal from the non-resonant background in momentum and opening angle distributions.

To minimize the uncertain impact of transverse momenta from initial-state quarks, the Collins-Soper frame [103] is used to explore more (di)photon angular distributions such as $\cos\theta^*$ and ϕ^*. In this frame, the z axis is defined as the bisector of the proton beam momentum and the negative of the anti-proton beam momentum when they are boosted into the center-of-mass frame of the diphoton pair. The variable θ^* is defined as the angle between the photon momentum and the z axis. The variable ϕ^* is defined as the angle between the diphoton plane and the proton-anti-proton plane. A schematic view of the frame and definitions of $\cos\theta^*$ and ϕ^* are illustrated in Figure 5.19.

When the Higgs boson is produced in association with a vector boson (W boson or Z boson), a large $\not{E}_T$ can be observed when the final states from W or Z boson decays contain neutrinos. It is more dramatic in the fermiophobic model in which VH is the dominant signal production mechanism. Moreover, photon identification variables such as O_{NN}^1 and O_{NN}^2 are included to further reject the residual background from misidentified photons.

The distributions of the ten input variables for signals, backgrounds and data in the inclusive mass range [60, 200] GeV are shown in Figures 5.20-5.29. The modeled background agrees well with data.

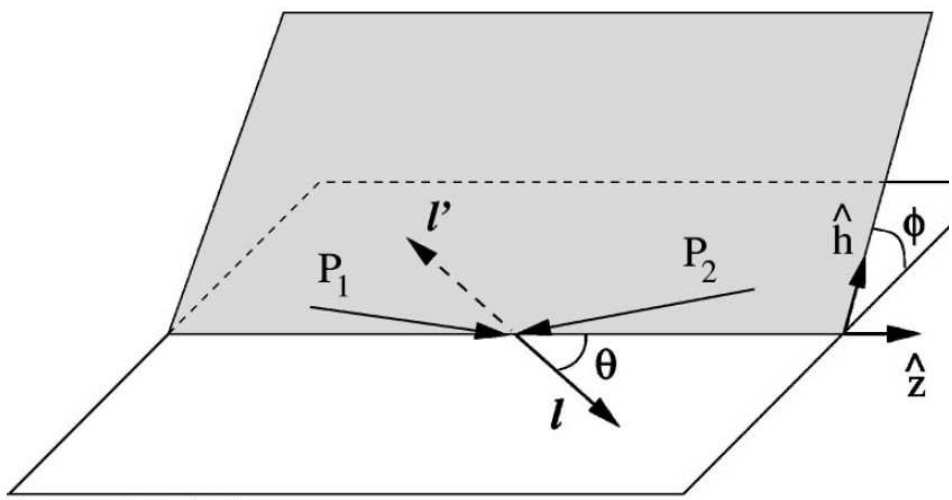
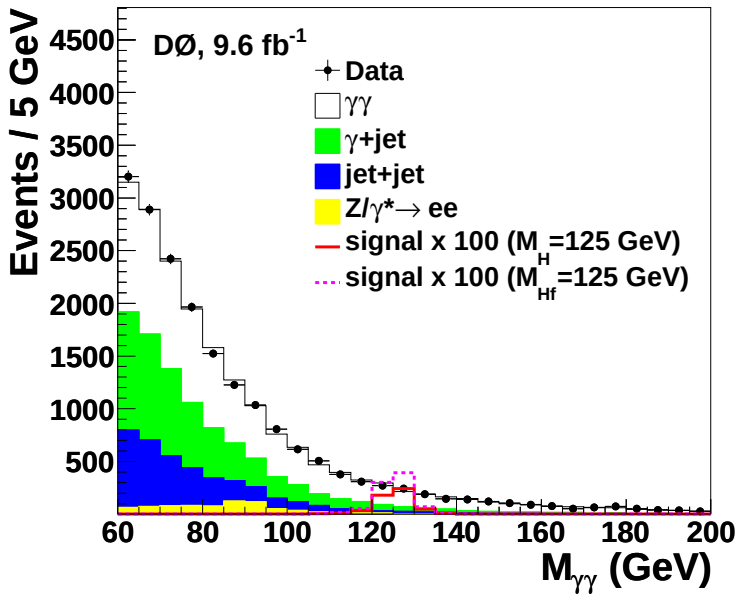


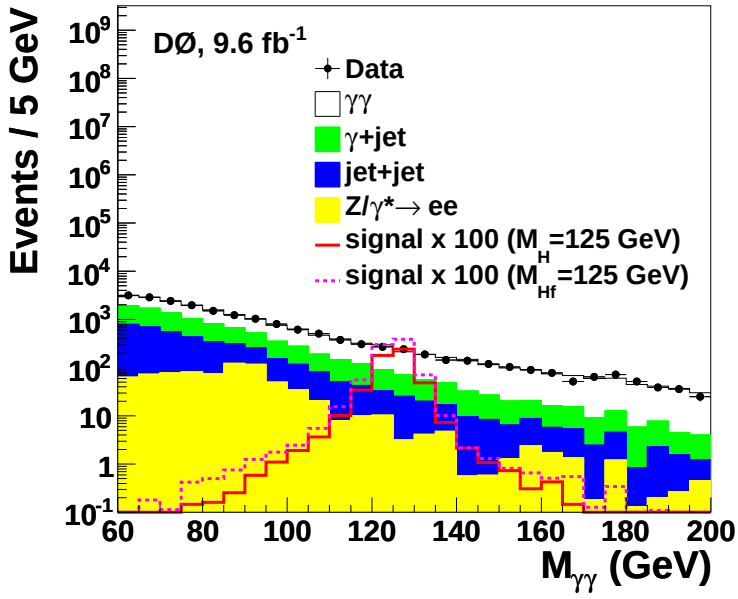
Figure 5.19: Collins-Soper frame illustration: P_1, P_2 are the proton and anti-proton momenta. l, l' are the photon momenta. $\hat{z}$ is a unit vector along the bisector of the proton beam momentum and the negative of the anti-proton beam momentum when they are boosted into the center-of-mass frame of the diphoton pair. The white plane is the diphoton plane; the gray plane is the proton-anti-proton plane. $\hat{h}$ is a unit vector in the proton-anti-proton plane but transverse to $\hat{z}$. θ and ϕ are the θ^* and ϕ^* in our notation.

Training

All the signal and background samples are separated into two independent samples by the parity of each event number. One is used for training; the other is used to populate the final discriminant distributions. In order to concentrate on the area near the Higgs boson resonance, a narrow mass window, for example ± 10 GeV, should be applied for an assumed Higgs boson mass. But considering the statistics for training, the mass window is enlarged to ± 30 GeV. After the training, the BDT response is tested on an independent sample (obtained by the TMVA framework during training) to check for over-training. This procedure is calling *testing*. The reasonable values of the Kolmogorov-Smirnov test in Figure 5.30 indicate that the training is robust and no severe over-training is observed.

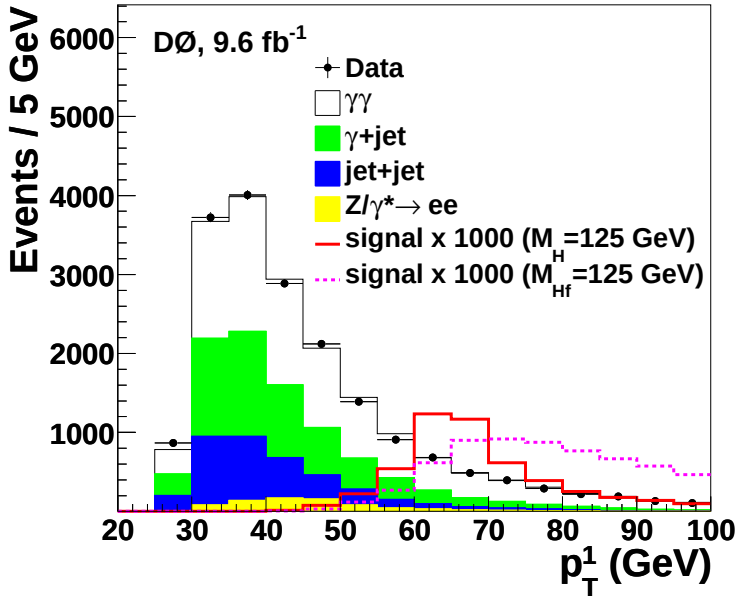


(a) $M_{\gamma\gamma}$ in linear scale

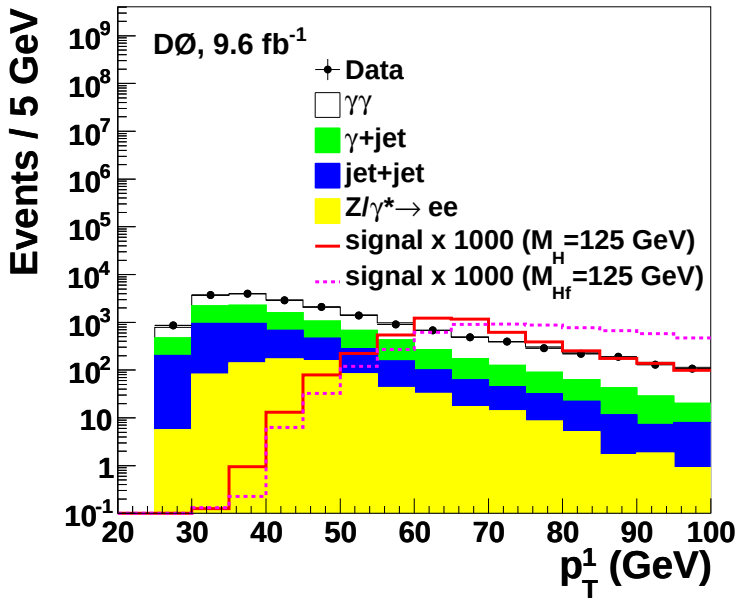


(b) $M_{\gamma\gamma}$ in log scale

Figure 5.20: $M_{\gamma\gamma}$ distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible.

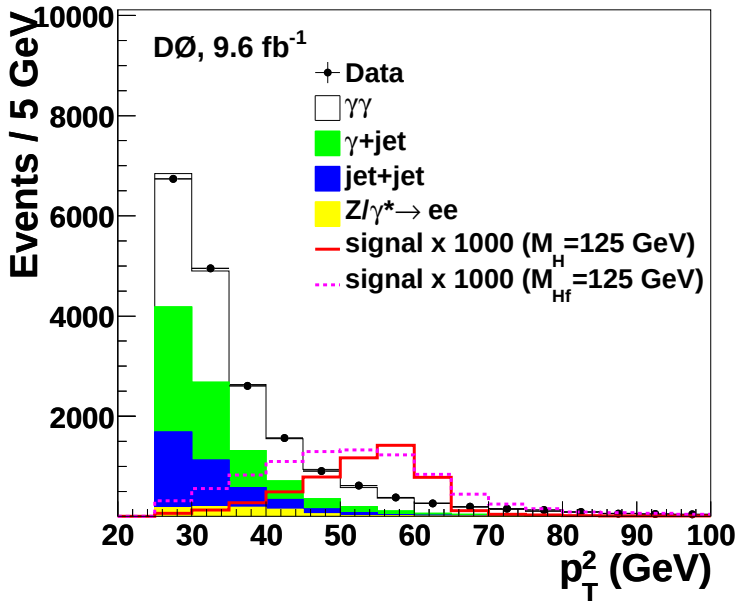


(a) E_T^1 in linear scale

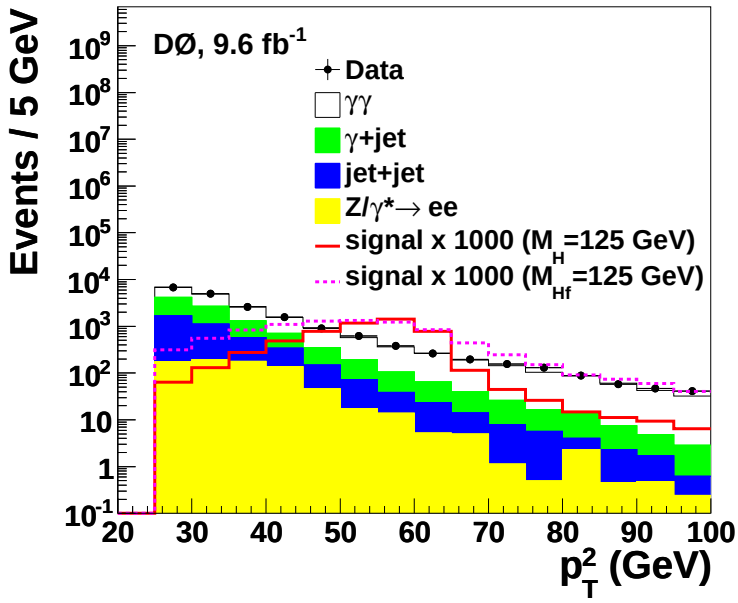


(b) E_T^1 in log scale

Figure 5.21: p_T^1 distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible.

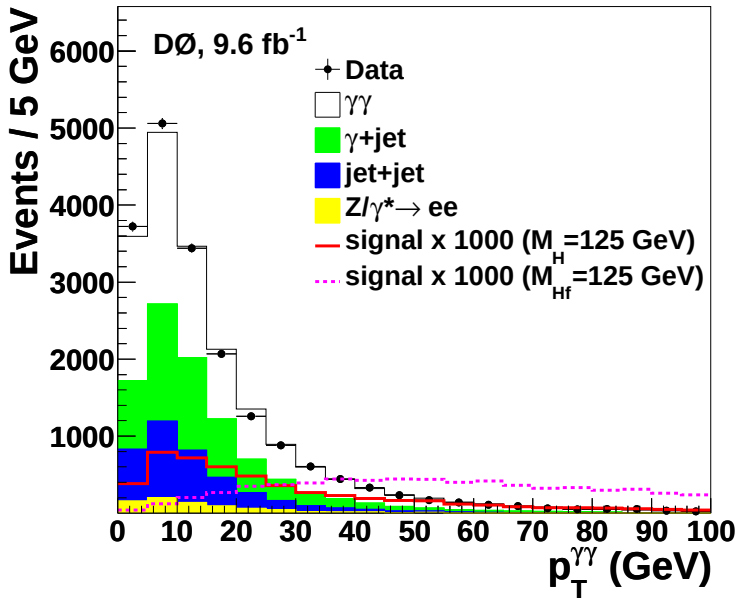


(a) E_T^2 in linear scale

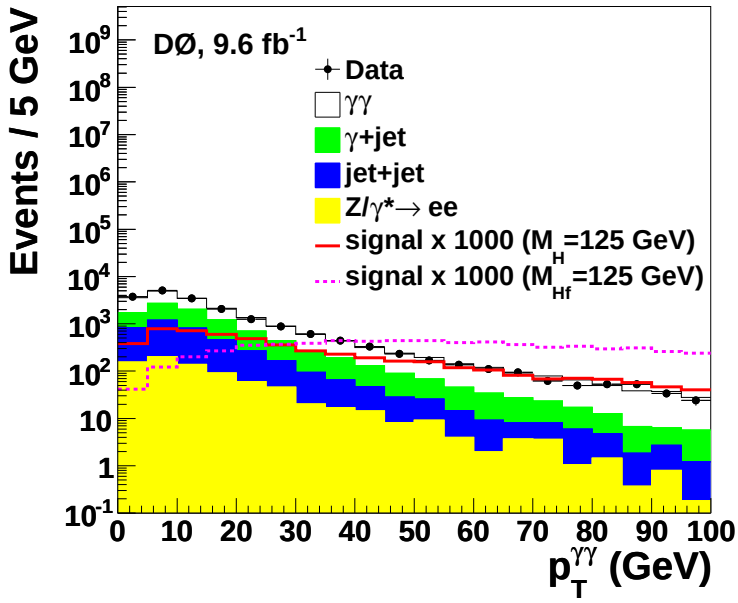


(b) E_T^2 in log scale

Figure 5.22: p_T^2 distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible.

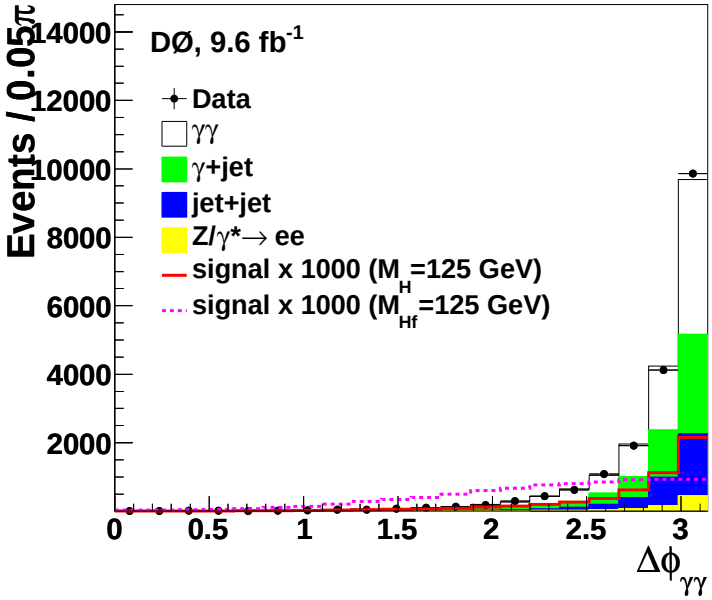


(a) $p_T^{\gamma\gamma}$ in linear scale

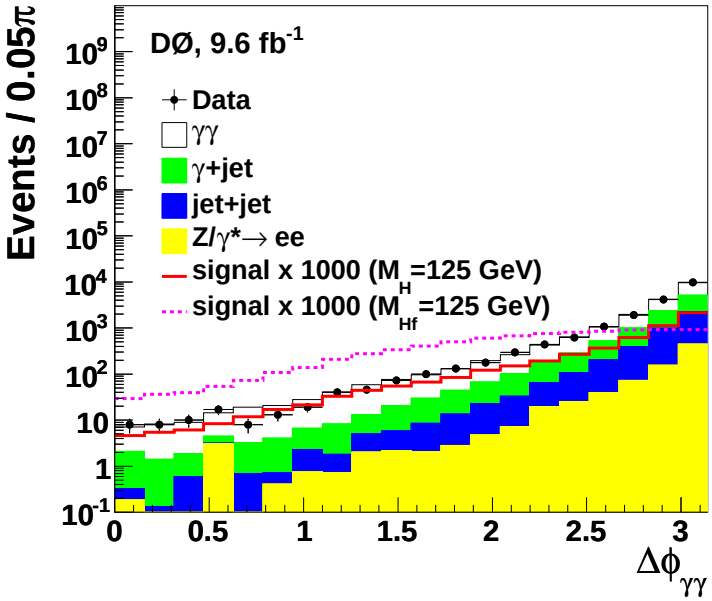


(b) $p_T^{\gamma\gamma}$ in log scale

Figure 5.23: $p_T^{\gamma\gamma}$ distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible.

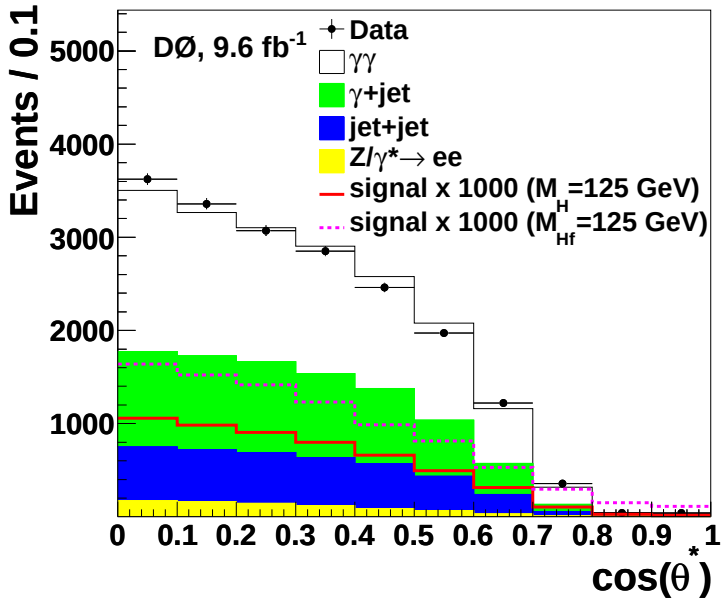


(a) $\Delta\phi^{\gamma\gamma}$ in linear scale

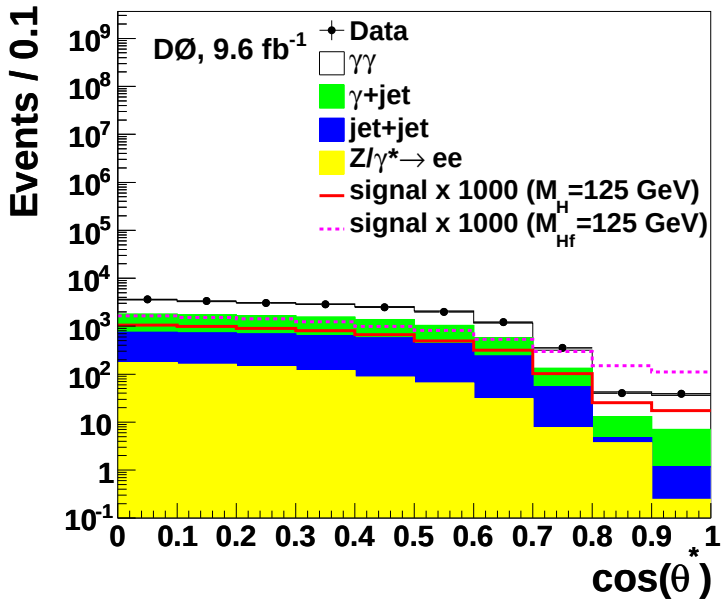


(b) $\Delta\phi^{\gamma\gamma}$ in log scale

Figure 5.24: $\Delta\phi_{\gamma\gamma}$ distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible.

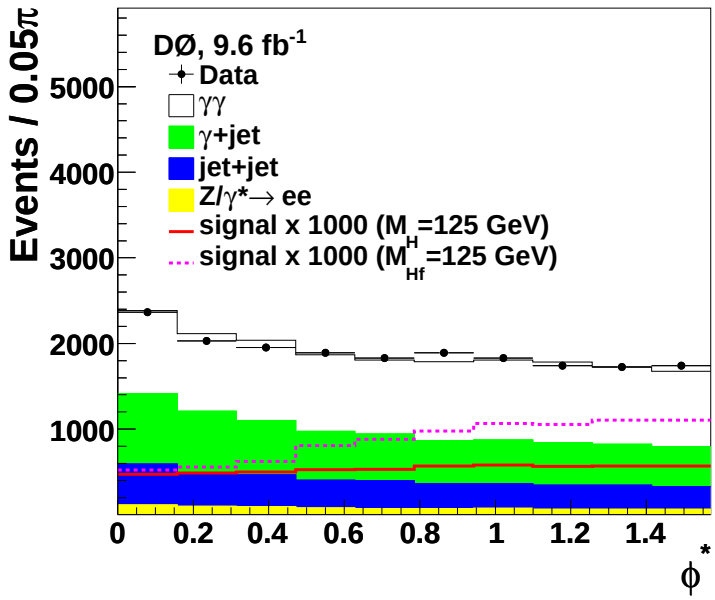


(a) $\cos\theta^*$ in linear scale

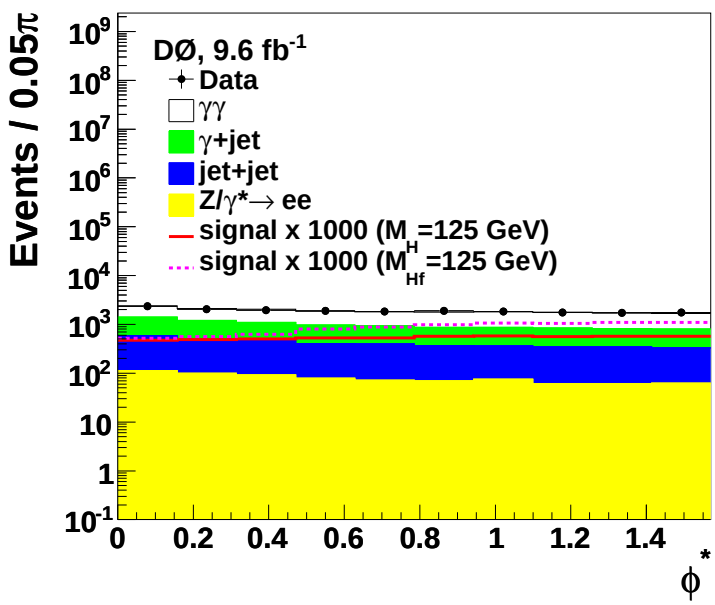


(b) $\cos\theta^*$ in log scale

Figure 5.25: $\cos\theta^*$ distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible.

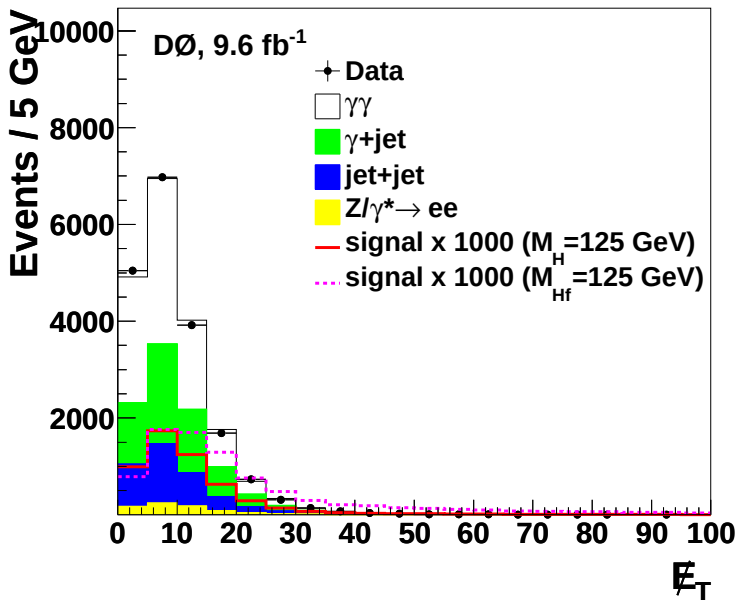


(a) ϕ^* in linear scale

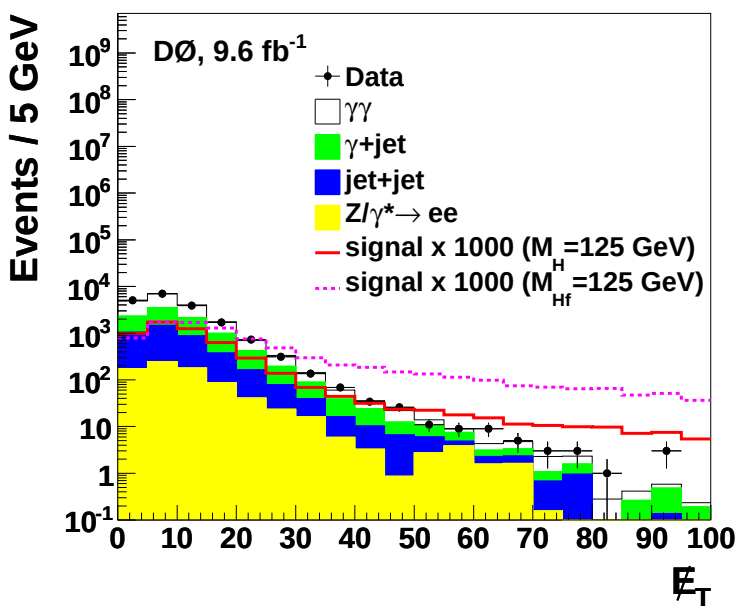


(b) ϕ^* in log scale

Figure 5.26: ϕ^* distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible.

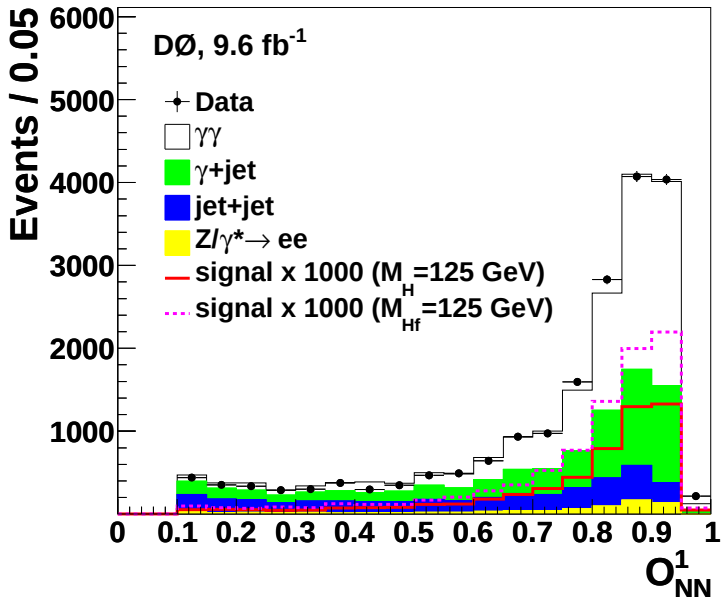


(a) $\not{E}_T$ in linear scale

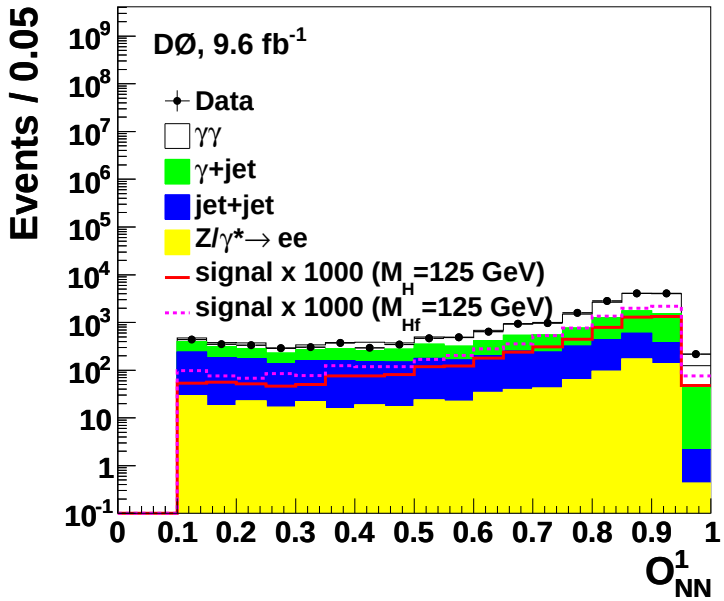


(b) $\not{E}_T$ in log scale

Figure 5.27: $\not{E}_T$ distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible.

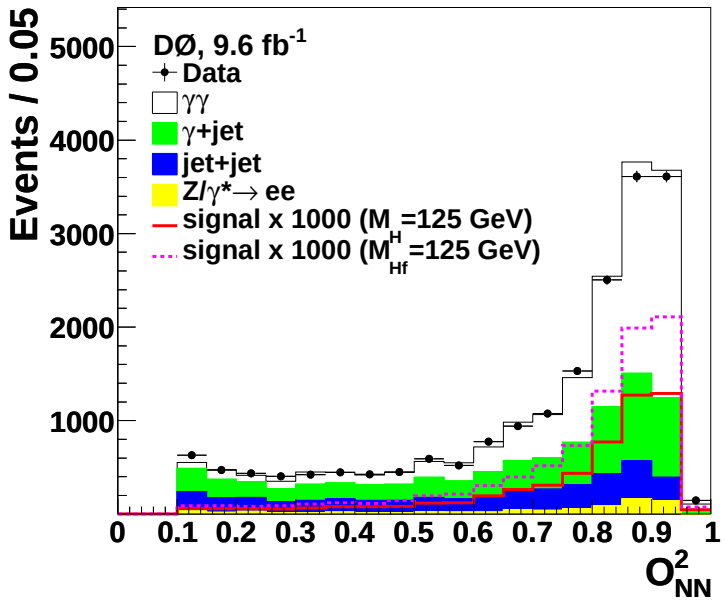


(a) O^1_{NN} in linear scale

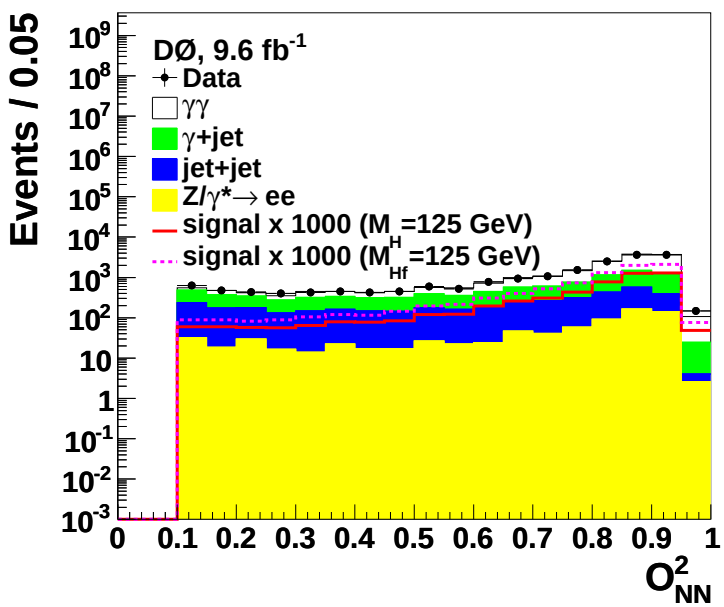


(b) O^1_{NN} in log scale

Figure 5.28: O^1_{NN} distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible.



(a) O^2_{NN} in linear scale



(b) O^2_{NN} in log scale

Figure 5.29: O^2_{NN} distributions for signals, backgrounds and data in the mass range [60, 200] GeV. The standard model signal and fermiophobic signal are represented by the red solid line and the magenta dotted line respectively. Signals with a mass of 125 GeV are assumed and multiplied by a factor of 1000 to be visible.

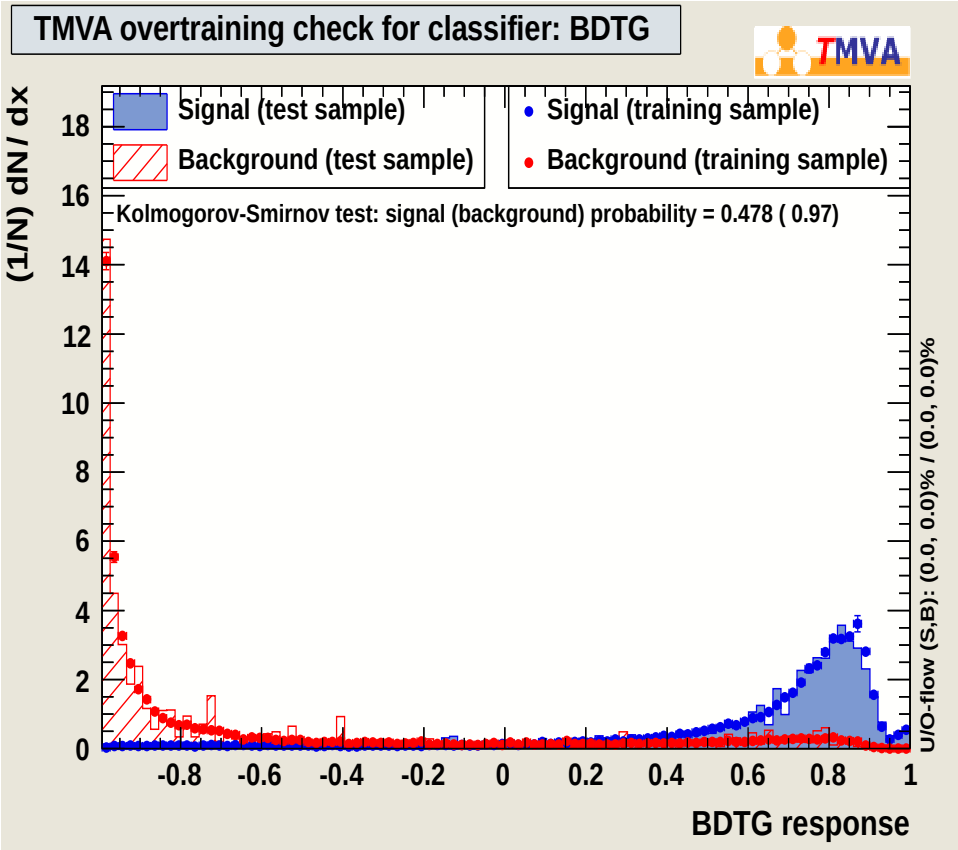
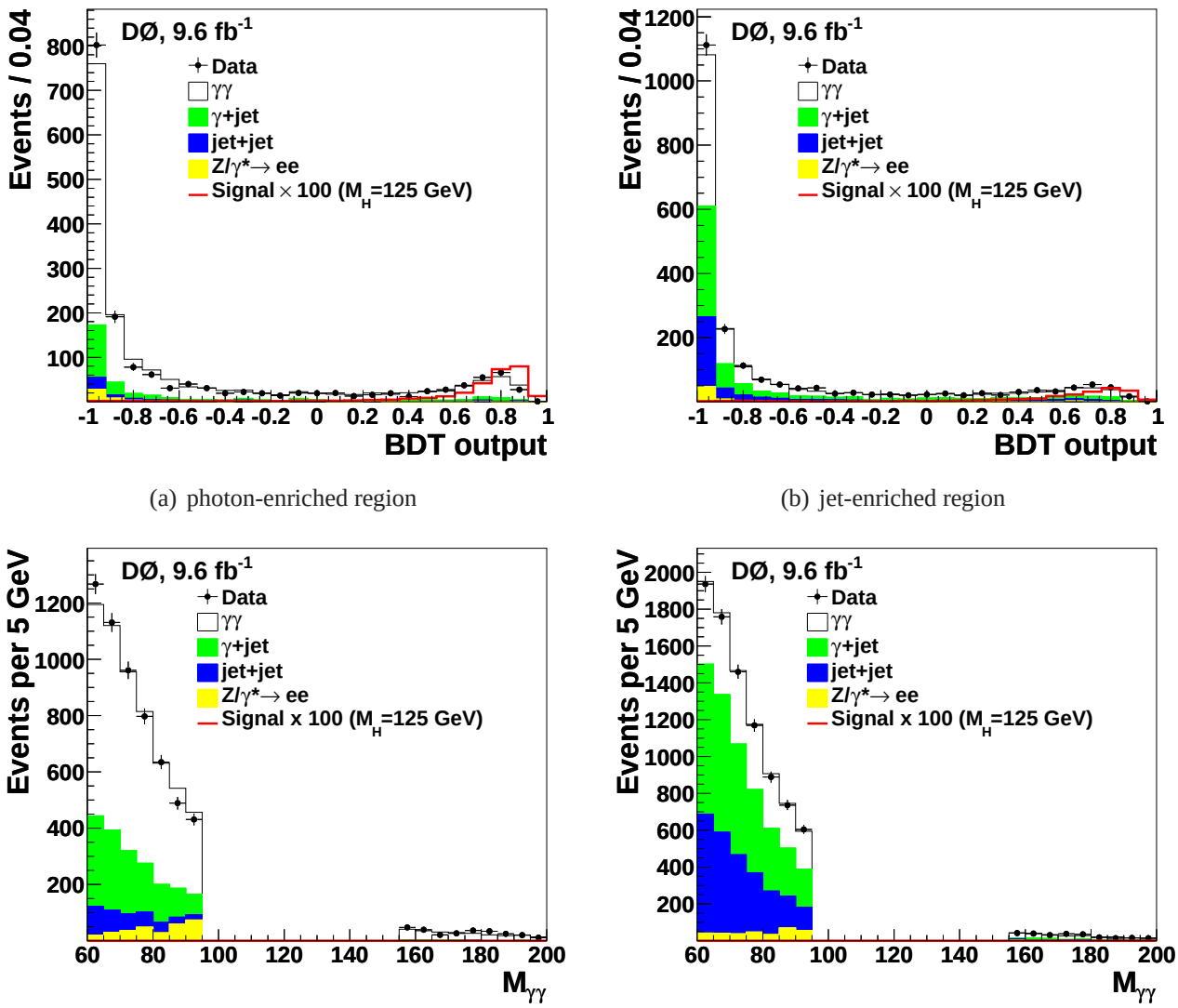


Figure 5.30: An illustration of the over-training check for a 125 GeV SM Higgs boson. The Kolmogorov-Smirnov test indicates the consistency of the distributions from the training sample and testing sample.

5.4.3 Final Discriminant

The BDT output distributions are categorized into two classes: the photon-enriched region (both photon candidates with $O_{NN} > 0.75$) and the jet-enriched region (at least one photon candidate with $O_{NN} < 0.75$). As mentioned, the BDT output distributions are obtained from events within $M_H \pm 30$ GeV for each assumed Higgs mass. In this study, sideband regions of the invariant mass spectrum outside the $M_H \pm 30$ interval are also included as final discriminants. Therefore for each M_H hypothesis, the four distributions (two BDT output distributions in the photon/jet-enriched regions plus two corresponding sidebands) are treated as independent sub-channels. As an illustration, the four final discriminant distributions for M_H = 125 GeV are shown in Figure 5.31 and Figure 5.32 with linear and logarithmic scales respectively.



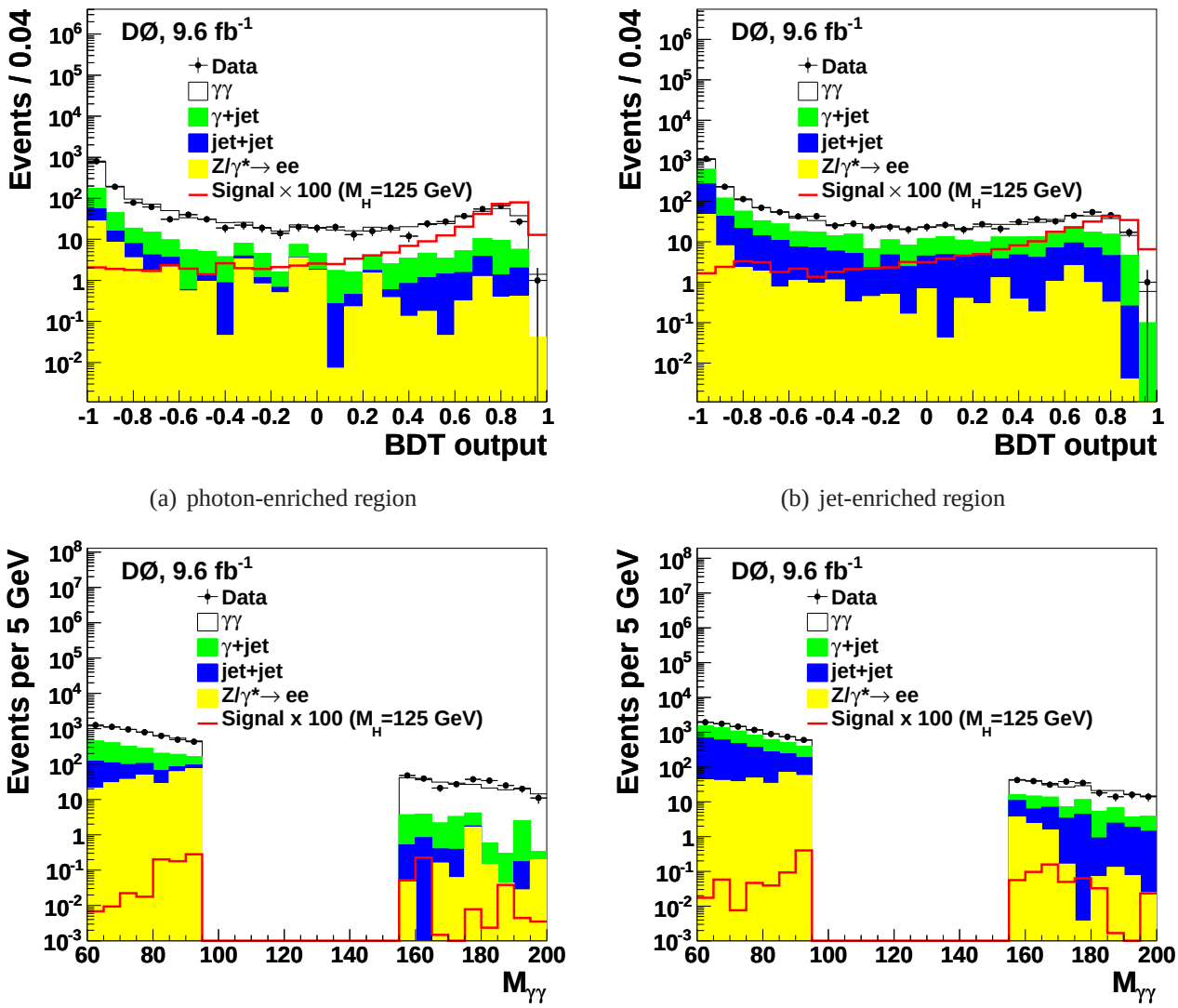
(a) photon-enriched region

(b) jet-enriched region

(c) photon-enriched region sideband

(d) jet-enriched region sideband

Figure 5.31: Final discriminants for 125 GeV signal with a linear scale.



(a) photon-enriched region

(b) jet-enriched region

(c) photon-enriched region sideband

(d) jet-enriched region sideband

Figure 5.32: Final discriminants for 125 GeV signal with a log scale.

5.5 Systematic Uncertainties

Systematic uncertainties involved in this analysis are:

- The total integrated luminosity uncertainty: 6.1%.
- The parton distribution functions (PDFs): the effect of the PDF uncertainty [104] on the signal acceptance is $1.7-2.2\%$ as a function of the assumed Higgs mass.
- Higgs p_T re-weighting: we re-weight the gluon-gluon fusion Higgs p_T spectrum from PYTHIA MC to the spectrum from the HQT NNLO event generator [83] for each assumed Higgs mass. In order to account for the systematic uncertainty from the soft gluon resummation procedure, we derive the shape systematic uncertainty by estimating the fractional change of the Higgs p_T distribution from RESBOS after varying the scale up ($2M_H$) and down ($0.5M_H$).
- Signal cross section uncertainties: 6.96% and 12.24% for the gluon fusion process from scale and PDF uncertainties; 6.18% for associated production and 4.91% for the vector boson fusion process [105].
- Efficiency scale factor: for Monte Carlo based signals and backgrounds, there are systematic uncertainties from the scale factors of the trigger efficiency, the EM cluster pre-selection efficiency, the photon ID efficiency and the track veto efficiency. Except for the Drell-Yan background, the track veto scale factor uncertainty is combined into the photon ID scale factor uncertainty. The track veto scale factor uncertainty for a single electron is estimated to be 9.0% (see Appendix A). The per event uncertainty is determined by simply multiplying the per object uncertainty by $\sqrt{2}$.
- The theoretical cross section uncertainty on the Drell-Yan $Z/\gamma^* \to ee$ normalization is 3.9% [85].

- ε_γ and ε_j: the uncertainties on photons and EM-like jets passing the $O_{NN} > 0.75$ criterion (ε_γ,ε_j) propagate into the uncertainty on the γ+jet and dijet background normalizations. The uncertainties on ε_γ and ε_j are 1.5% and 10% according to Ref [90]. See Appendix B for more detail.

- The scale uncertainty of the the QCD $\gamma\gamma$ Monte Carlo: the MC prediction of the shape of QCD $\gamma\gamma$ kinematics relies on how the scale is chosen. We have 3D ($M_{\gamma\gamma} - \Delta\phi_{\gamma\gamma} - p_T^{\gamma\gamma}$) weights at the generator level from PYTHIA to SHERPA in nominal scale, scale up and scale down, where the nominal scale choice is where the renormalization and factorization factor are equal to $M_{\gamma\gamma}$. For the scale up we double the renormalization and factorization factors. For the scale down we halve the renormalization and factorization factors. We weight the SHERPA MC by $\frac{scaleup}{nominal}$ to obtain a $+1\sigma$ variation and by $\frac{scaledown}{nominal}$ to obtain a -1σ variation.

- non-$\gamma\gamma$ (γ+jet, jet+jet) shape systematics: the shape of systematic uncertainties of the non-$\gamma\gamma$ background is estimated by comparing the mass spectrum from the orthogonal sample with the 4×4 matrix method solution. The latter suffers a lot from large statistical uncertainties. So we use a first-order polynomial function to fit the ratio of the shapes and symmetrize the fit function around the constant function $y = 1$.

The systematic uncertainties for the signal and each background component are summarized in Table 5.5.

source	signal	Drell-Yan	γ+jet	jet+jet	QCD $\gamma\gamma$
luminosity	6.1%	6.1%	-	-	-
PDF for $H \rightarrow \gamma\gamma$ acceptance	1.7% -2.2%	-	-	-	-
H p_T re-weighting	shape	-	-	-	-
trigger	0.1%	0.1%	-	-	-
pre-selection scale factor	0.5%	0.5%	-	-	-
ID scale factor	2.7%	2.7%	-	-	-
track veto scale factor	-	9.0%	-	-	-
cross section	4.91%-12.24%	3.9%	-	-	-
ε_γ	-	-	7.3%	5.4%	-
ε_j	-	-	0.8%	17%	-
QCD $\gamma\gamma$ MC scale	-	-	-	-	shape
non-$\gamma\gamma$ shape:	-	-	15%	10%	-

Table 5.5: Systematic uncertainties for different sources.

Chapter 6

Result

By examining the final discriminant distributions for various hypothetical Higgs boson masses, we observed no obvious excess of signal-like events above the background prediction in the data sample. Therefore upper limits on the product of the cross section and the branching ratio ($\sigma \times \mathscr{B}(H \to \gamma\gamma)$) are derived as a function of the assumed Higgs boson masses. In this Chapter, we first describe the statistical method for setting limits, and then present the results on the searches for the SM Higgs boson and the fermiophobic Higgs boson.

6.1 Limit Calculation Method

The limits are calculated at the 95% confidence level with the $\mathrm{CL_S}$ modified frequentist approach [106], implemented by the Confidence Level Limit Evaluator (COLLIE) [107]. We discuss the calculation method and treatment of statistical and systematic uncertainties in COLLIE below.

Generally speaking, limit calculation is about hypothesis testing. Two hypotheses are tested here. One is the null hypothesis H_0, which represents a model where only background (B) is present and there are no signal events from a Higgs boson; it is also called the B-only hypothesis. The other is the test hypothesis H_1, which represents a model where both background and signal ($S+B$) exist; it can be called the $S+B$ hypothesis.

A test statistic should be constructed to discriminate signal-like and background-like events,

providing a test of the two hypotheses. A quantity called the *likelihood ratio* Q satisfies this requirement and is defined as in Equation 6.1. From its name, Q is the ratio of the conditional probabilities of data given the two hypotheses. It measures the extent to which the data is more consistent with the null hypothesis or not, and *vice versa*.

$$Q = \frac{p(data|H_1)}{p(data|H_0)}. \tag{6.1}$$

For High Energy physics experiments, the number of events or the number of events in a bin of a histogram follows Poisson statistics. So the likelihood ratio Q can be explicitly expressed in a Poisson likelihood ratio as in Equation 6.2.

$$\begin{aligned} Q(s,b,d) &= \frac{p(data|H_1)}{p(data|H_0)} \\ &= \frac{p(data|S+B)}{p(data|B)} \\ &= \frac{e^{-(s+b)}(s+b)^d/d!}{e^{-b}b^d/d!} \\ &= e^{-s}(1+\frac{s}{d})^d \\ &= \prod_{i=1}^{N_{bins}} e^{-s_i}(1+\frac{s_i}{d_i})^{d_i}, \end{aligned} \tag{6.2}$$

where N_{bins} is the number of bins for the final discriminant distributions; s_i, b_i, d_i are the number of events in the i-th bin for signal, background and data.

For numerical reasons, it is better to transform the Poisson likelihood ratio to the negative log-likelihood ratio, namely LLR, as shown in Equation 6.3.

$$\begin{aligned} \text{LLR} &= -2\ln(Q(s,b,d)) \\ &= -2\ln(\prod_{i=1}^{N_{bins}} e^{-s_i}(1+\frac{s_i}{d_i})^{d_i}) \\ &= 2\sum_{i=1}^{N_{bins}}(s_i - d_i\ln(1+\frac{s_i}{b_i})). \end{aligned} \tag{6.3}$$

With the definition of LLR, we need to know the distributions of LLR for the $S+B$ and B-only hypotheses. It is achieved by generation of pseudo-data. Consider a particular bin in the final discriminant histogram, the pseudo-data model assumes that its number of data events is stochastically sampled from a Poisson distribution. The mean of the Poisson distribution is the sum of the physics processes under a certain hypothesis. For the B-only hypothesis, the mean is the number of events for background; for the $S+B$ hypothesis, the mean is the sum of number of events for signal and background. A sampling process to generate a pseudo-data event is called a $pseudo-experiment$. Thus the LLR distributions for the two hypotheses are populated by generating pseudo-experiments in a number of trials, as in Equation 6.4.

$$\text{LLR distribution} = 2\sum_{i=1}^{N_{bins}}(s_i - \hat{d}_i \ln(1+\frac{s_i}{b_i})), \tag{6.4}$$

where $\hat{d}_i$ is a pseudo-data event sampled from the distribution of Poisson(s_i+b_i) (Poisson(b_i)) for the $S+B$ (B-only) hypothesis. It must be mentioned that there is an important assumption in Equation 6.4. Each bin of the pseudo-data is independent and so there are N_{bins} of random variables that each follows its own Poisson statistics with its own mean.

By using the LLR test statistic, we can calculate a confidence level (CL) for signal exclusion. The method used here is a modified frequentist approach[1], CL_S. It requires calculation of two quantities called CL_{S+B} and CL_B. CL_{S+B} (CL_B) is the probability for the $S+B$ (B-only) hypothesis to produce an outcome more background-like than that observed in data. So CL_{S+B} is the p-value and CL_B is 1-p-value for their corresponding hypothesis. The calculation is illustrated in Equation 6.5 and Figure 6.1. In Figure 6.1, the separation of the medians for the $S+B$ and B-only hypotheses indicates the sensitivity of the search.

$$\text{CL}_S = \frac{\text{CL}_{S+B}}{\text{CL}_B} = \frac{p\text{-value}_{S+B}}{1-p\text{-value}_B}. \tag{6.5}$$

[1]A traditional frequentist approach uses CL_{S+B}, but is likely to produce false exclusion when data have an downward fluctuation significantly below background prediction [107].

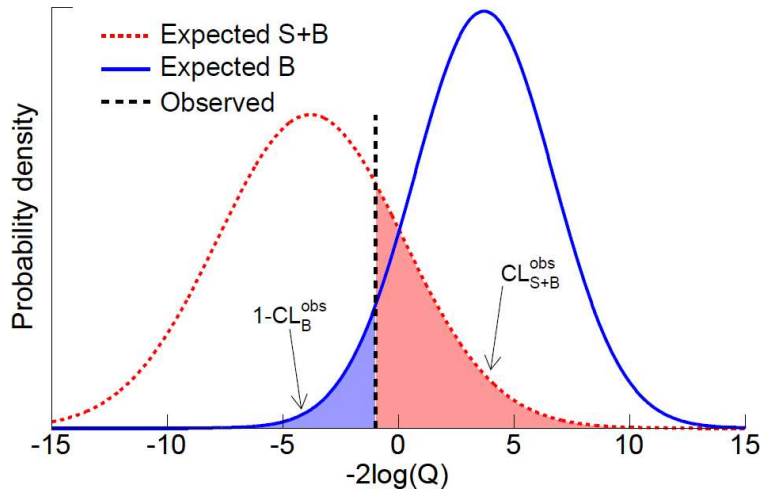


Figure 6.1: Example distributions of log-likelihood ratio (LLR) under $S+B$ and B-only hypotheses. The LLR value from the observed data is shown as a vertical dashed line, separating the LLR distributions into two areas for the two hypotheses each. The red shaded area is CL_{S+B} and also the p-value of the test hypothesis. The blue shaded area is 1-CL_B, representing the p-value of the null hypothesis [108].

With the definition of CL_S, we can exclude signal at a confidence level of $1-\alpha$ such that $\mathrm{CL}_s < \alpha$. The upper limits are usually expressed as ratios of the cross section upper limits to the theoretical predictions. Such ratios are the multipliers of the signal until the CL_S is less than a threshold, for example, 5%. Then we can say that the upper limits are set at 95% CL.

Until now, we have not included statistical and systematic uncertainties in the limit calculation. The limits at this stage are called *CLFast* limits in the COLLIE framework.

Still consider a final discriminant histogram, the statistical uncertainties are considered by an uncorrelated Gaussian smearing for each bin content. The width of the Gaussian is the per-bin content statistical uncertainty.

The treatment of systematic uncertainties is much more complicated and it involves a concept called *nuisance parameters*. A nuisance parameter is a parameter that is not specified but of no immediate interest to the hypothesis test. For example, the integrated luminosity is a parameter of no immediate interest in the context of setting upper limits on cross sections. Systematic uncertainties, accounting for our limited knowledge on nuisance parameters, are modeled with a prior probability distribution function (PDF) with the widths specified by the ± 1 standard deviation. In the pseudo-experiment, the nuisance parameter value is stochastically sampled from the prior PDF.

The PDF can be a Gaussian or log-normal distribution. In this study, we use Gaussian distributions as the prior PDFs. The correlations between systematic uncertainties are maintained during sampling.

The incorporation of systematic uncertainties usually degrades the sensitivity of a search. The impact can be significantly reduced by making a modification to the test statistic from the Poisson log-likelihood ratio defined by Equation 6.3. The modification requires a minimization of a Poisson χ^2 function to determine the best fit of background to data. χ^2 is a function of the nuisance parameters and their uncertainties (systematic uncertainties), as shown in Equation 6.6.

$$\chi^2(H) = -2\ln P(data|H,\vec{\theta}), \tag{6.6}$$

where the hypothesis H can be $S+B$ or B-only; and $\vec{\theta}$ is a list of nuisance parameters.

The χ^2 is minimized individually for the $S+B$ and the B-only hypotheses[2]. So Equation 6.3 is changed to the so called "profile likelihood ratio".

$$\begin{aligned} \text{profile LLR} &= -2\ln(Q) \\ &= -2\ln\left(\frac{p(data|S+B,\vec{\theta}_{s+b})}{p(data|B,\vec{\theta}_b)}\right), \end{aligned} \tag{6.7}$$

where $\vec{\theta}_{s+b}$ ($\vec{\theta}_b$) is the list of nuisance parameters from minimization of $\chi^2(S+B)$ ($\chi^2(B)$).

This approach is called *CLFit*2 in COLLIE and it is what we used for this study to set limits. A more detailed discussion about this method and others can be found in Ref. [107].

[2]The minimization can certainly be done for just the B-only hypothesis to reduce computational load, namely the *CLFit* method in COLLIE, but it does not yield the optimal sensitivity.

Higgs mass (GeV)	100	105	110	115	120	125	130	135	140	145	150
Expected limit	12.2	10.2	9.3	9.1	8.9	8.7	9.0	10.0	11.2	13.3	16.8
Observed limit	11.9	16.6	10.5	8.3	9.1	12.8	12.3	9.9	13.2	19.2	25.4

Table 6.1: Expected and observed limits on the ratio of $\sigma \times \mathscr{B}(H \to \gamma\gamma)$ to the SM prediction as a function of Higgs mass using the BDT.

6.2 SM Higgs Boson Search Result

The upper limits on the cross section times branching ratio ($\sigma \times \mathscr{B}(H \to \gamma\gamma)$) relative to the SM prediction are set at 95% CL using the CLFit2 method with COLLIE version V00-04-12.

Our most stringest expected limits come from the joint usage of the BDT in the ± 30 GeV mass window and the sidebands outside. The results of the limits and log-likelihood ratios are shown in Figure 6.2 and Table 6.1. The best expected limit is 8.5 at 125 GeV, where there is a slight excess over 1σ but less than 2σ. We also calculated the limits based solely on diphoton invariant mass in Appendix C as a cross check. Apparently, the two methods give consistent trends on the data and the BDT method provides significantly more stringent limits than the diphoton invariant mass method.

To examine the fit process discussed in the Section 6.1, we show the event rates after the fit process (post-fit) for signal, background and data in Tables 6.2-6.3 for the inclusive, photon-enriched and jet-enriched channels respectively. To better visualize the post-fit results and examine them for any potential excess, we also plot the data events after subtraction of background. An example plot is shown in Figure 6.3 at 125 GeV.

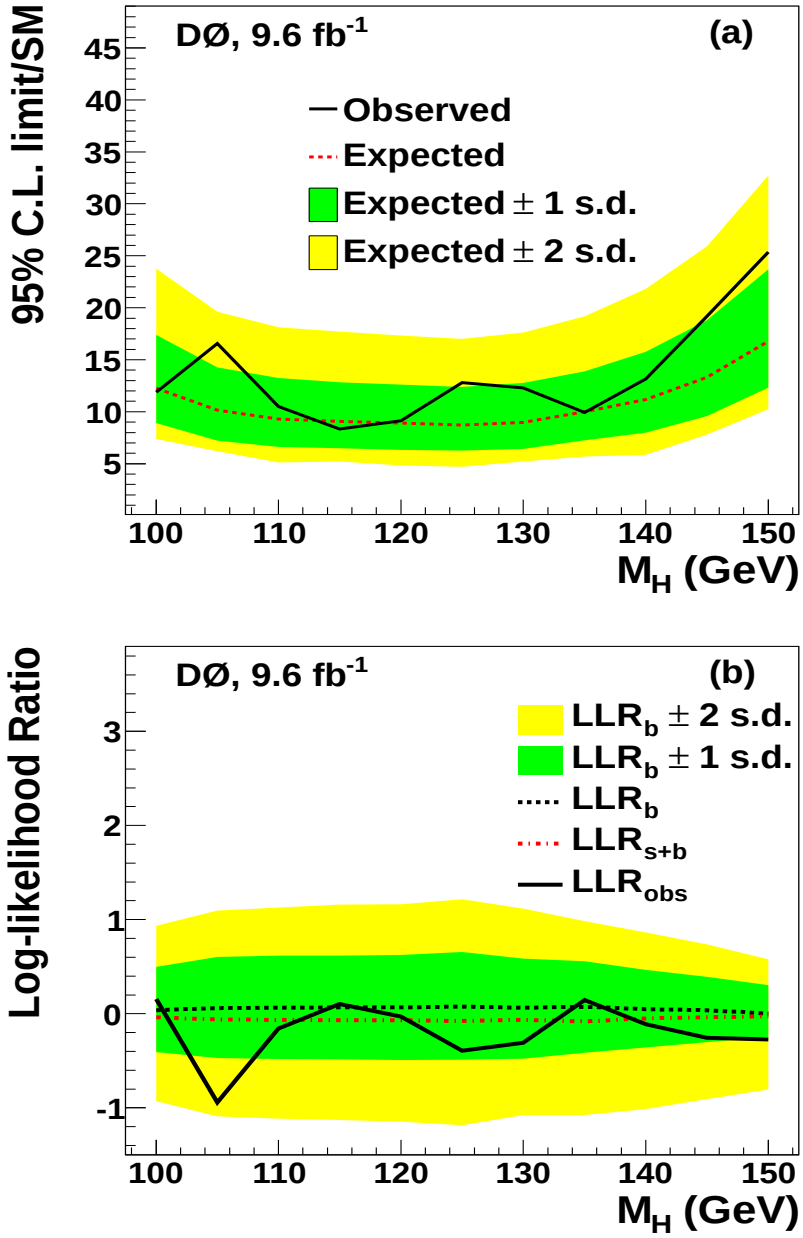


Figure 6.2: Results for the SM Higgs search from using the joint BDT and sideband as the final discriminant. (a) Limits at 95% CL on the ratio of $\sigma \times \mathscr{B}(H \to \gamma\gamma)$ to the SM prediction as a function of Higgs mass. The observed limit is shown as a solid black line while the expected limit under the background-only hypothesis is shown as a dashed red line. The green and yellow areas correspond to 1 and 2 standard deviations (s.d.) around the expected limit for the background-only hypothesis. (b) Log-likelihood ratios (LLR) as a function of Higgs mass.

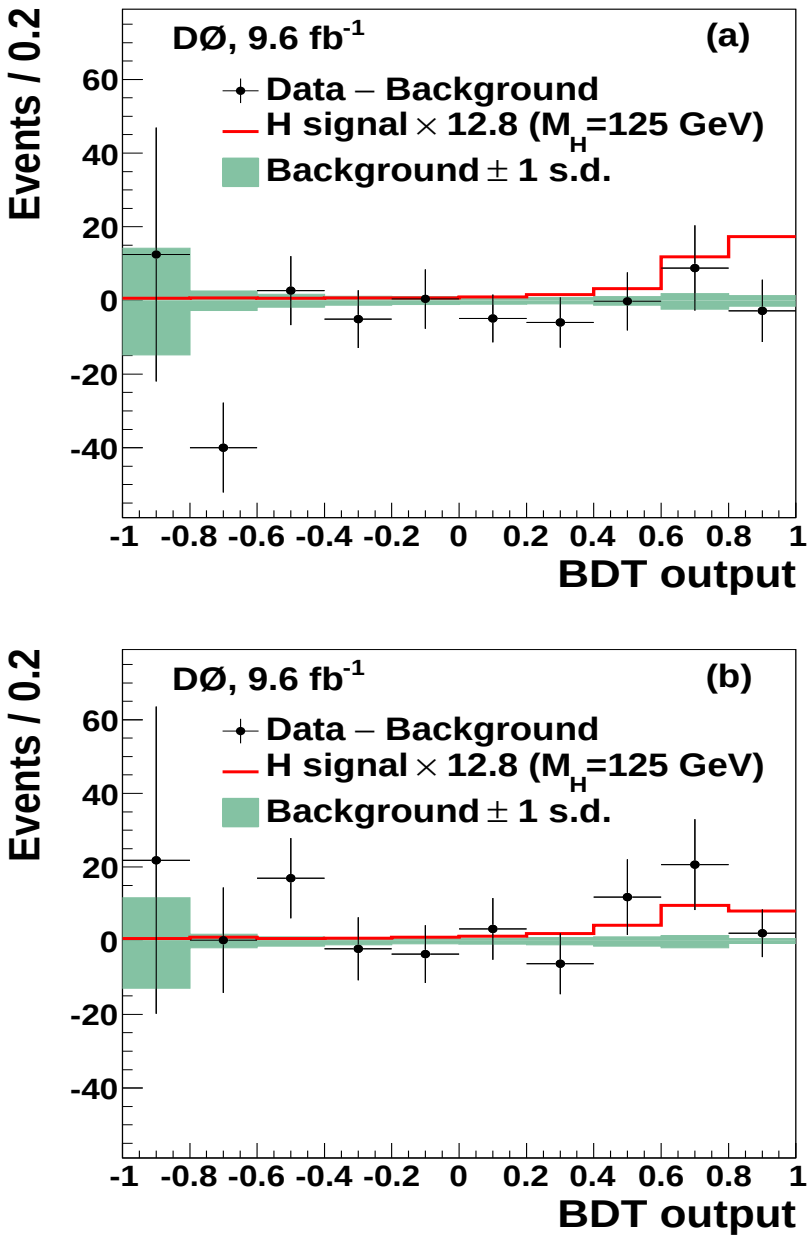


Figure 6.3: Data after subtraction of the post-fit background at $M_H = 125$ GeV in (a) the photon-enriched sample and (b) the jet-enriched sample. The expected SM Higgs signal is normalized to the corresponding observed limit in Table 6.1. The bands represent the 1 s.d. uncertainties from the background-only hypothesis fit.

M_H (GeV)	105	115	125	135	145
$\gamma\gamma$ (DDP)	2777 ± 65	1928 ± 44	1355 ± 31	980 ± 22	721 ± 17
$\gamma + jet$	704 ± 40	407 ± 24	238 ± 14	144 ± 9	88 ± 6
$jet + jet$	183 ± 16	93 ± 9	54 ± 6	34 ± 4	19 ± 2
$Z/\gamma^* \to e^+e^-$	219 ± 40	149 ± 30	51 ± 11	22 ± 5	11 ± 3
Total background	3883 ± 61	2577 ± 45	1698 ± 30	1180 ± 21	839 ± 16
Data	3777	2475	1664	1147	813
H signal	3.6 ± 0.4	3.5 ± 0.4	3.0 ± 0.4	2.2 ± 0.3	1.4 ± 0.2
H_f signal	49.8 ± 1.1	14.0 ± 0.3	4.8 ± 0.1	1.9 ± 0.1	0.79 ± 0.03

Table 6.2: Signal, backgrounds and data yields for the **photon-enriched** sample within the $M_H \pm 30$ GeV mass window, for M_H = 105 GeV to M_H = 145 GeV in 10 GeV intervals. The background yields are from a fit to the data. The uncertainties include both statistical and systematic contributions added in quadrature and take into account correlations among processes. The uncertainty on the total background is smaller than the sum in quadrature of the uncertainties in the individual background sources due to the anti-correlation resulting from the fit.

M_H (GeV)	105	115	125	135	145
$\gamma\gamma$ (DDP)	1969 ± 47	1406 ± 33	1012 ± 24	734 ± 17	545 ± 13
$\gamma + jet$	1852 ± 100	1101 ± 60	653 ± 36	391 ± 22	251 ± 15
$jet + jet$	1188 ± 94	647 ± 54	365 ± 31	219 ± 19	135 ± 12
$Z/\gamma^* \to e^+e^-$	227 ± 39	152 ± 28	61 ± 11	30 ± 7	20 ± 5
Total background	5236 ± 67	3307 ± 45	2091 ± 29	1374 ± 21	951 ± 17
Data	5287	3384	2156	1422	989
H signal	2.7 ± 0.3	2.6 ± 0.3	2.2 ± 0.3	1.7 ± 0.2	1.1 ± 0.1
H_f signal	34.8 ± 0.8	9.8 ± 0.3	3.4 ± 0.1	1.34 ± 0.04	0.56 ± 0.02

Table 6.3: Signal, backgrounds and data yields for the **jet-enriched** sample within the $M_H \pm 30$ GeV mass window, for M_H = 105 GeV to M_H = 145 GeV in 10 GeV intervals. The background yields are from a fit to the data. The uncertainties include both statistical and systematic contributions added in quadrature and take into account correlations among processes. The uncertainty on the total background is smaller than the sum in quadrature of the uncertainties in the individual background sources due to the anti-correlation resulting from the fit.

6.3 Fermiophobic Interpretation

As described in Section 1.4 of Chapter 1, the branching ratio in the fermiophobic model can be enhanced by an order of magnitude as can be seen in Table 6.4. In this section, we directly interpret the SM results in the context of the fermiophobic model. We use the same 9.6 fb^{-1} data samples and analysis technique as the search for the SM Higgs boson, except we retrain the BDT using fermiophobic signals. Figure 6.4 and Figure 6.5 show the four final discriminants for an example signal mass point (115 GeV) in linear and log scale respectively.

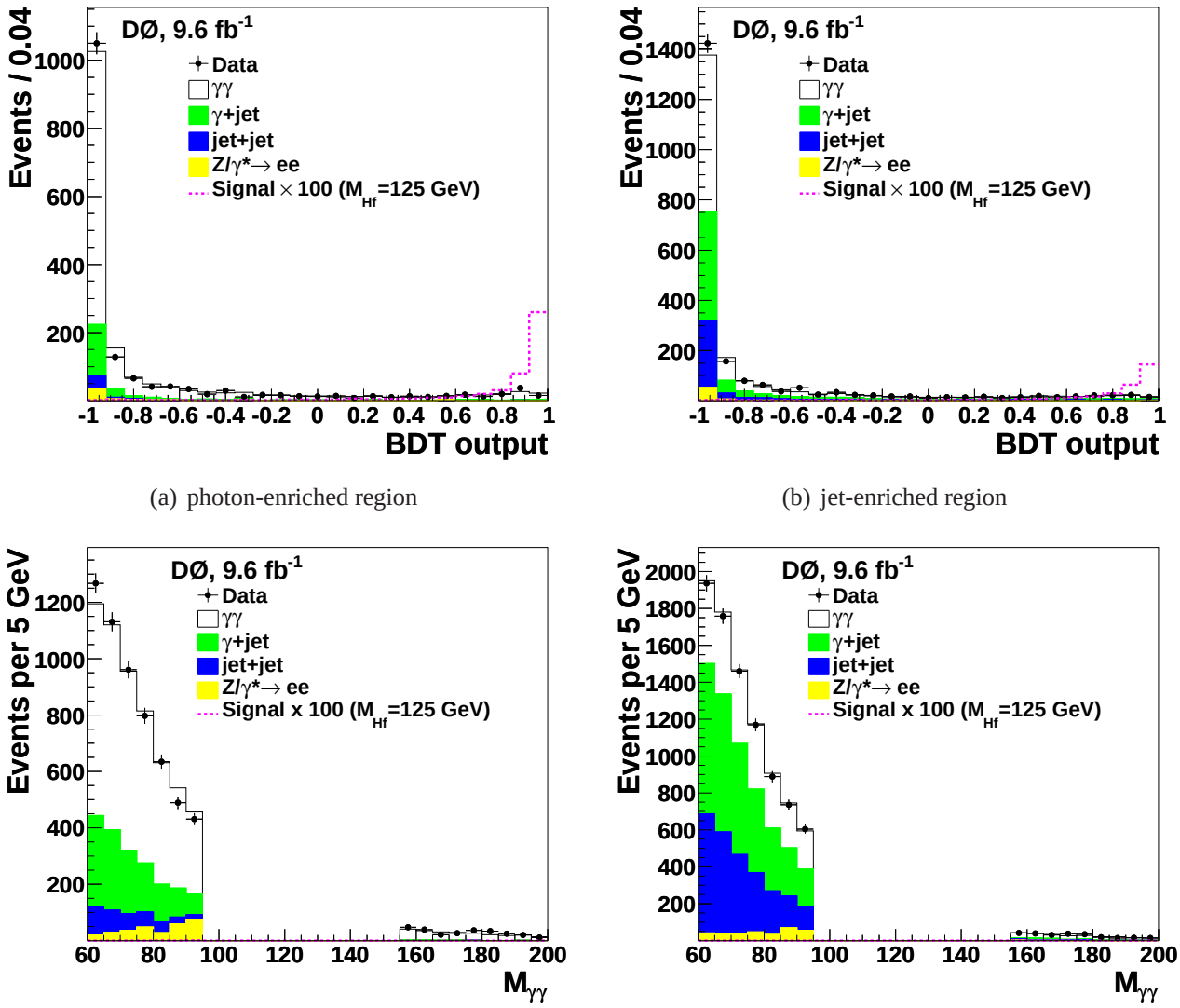
M_{H_f} (GeV)	100	110	120	130	140	150
$BR(H \to \gamma\gamma)$	0.0015	0.0019	0.0022	0.0022	0.0019	0.0014
$BR(H_f \to \gamma\gamma)$	0.185	0.060	0.023	0.011	0.005	0.0030
$BR(H_f \to \gamma\gamma)/BR(H \to \gamma\gamma)$	123	32	10	5	3	2

Table 6.4: Branching ratio comparison for a SM Higgs boson (H) and a fermiophobic Higgs boson (H_f) decaying into two photons.

We assume that in the fermiophobic model, the signal production processes for weak boson associated production and vector boson fusion have the same cross section as in the Standard Model. Therefore the limits are explicitly expressed as branching ratios in Table 6.5. They are compared with the existing LEP limits [30] and also the theoretical prediction for signal in Figure 6.6. The search region for the fermiophobic Higgs boson extends to a mass of 150 GeV, and the expected exclusion reaches 115 GeV, better than the combined 4 LEP experiments ($M_{H_f} > 109.7$ GeV). From the intersection of the observed limit with the theoretical prediction, we set a lower limit on the fermiophobic Higgs mass of $M_{H_f} > 114$ GeV at 95% CL.

Higgs mass (GeV)	100	105	110	115	120	125	130	135	140	145	150
Expected limit (%)	3.8	3.9	3.9	3.8	4.3	4.5	4.4	4.7	5.0	5.1	5.4
Observed limit (%)	5.8	4.7	4.0	4.6	4.4	5.5	5.1	7.0	5.3	5.4	4.2

Table 6.5: Expected and observed upper limits at 95% CL on the branching ratio of a fermiophobic Higgs boson decaying into two photons $(H_f \to \gamma\gamma)$ as a function of the mass using the BDT method.



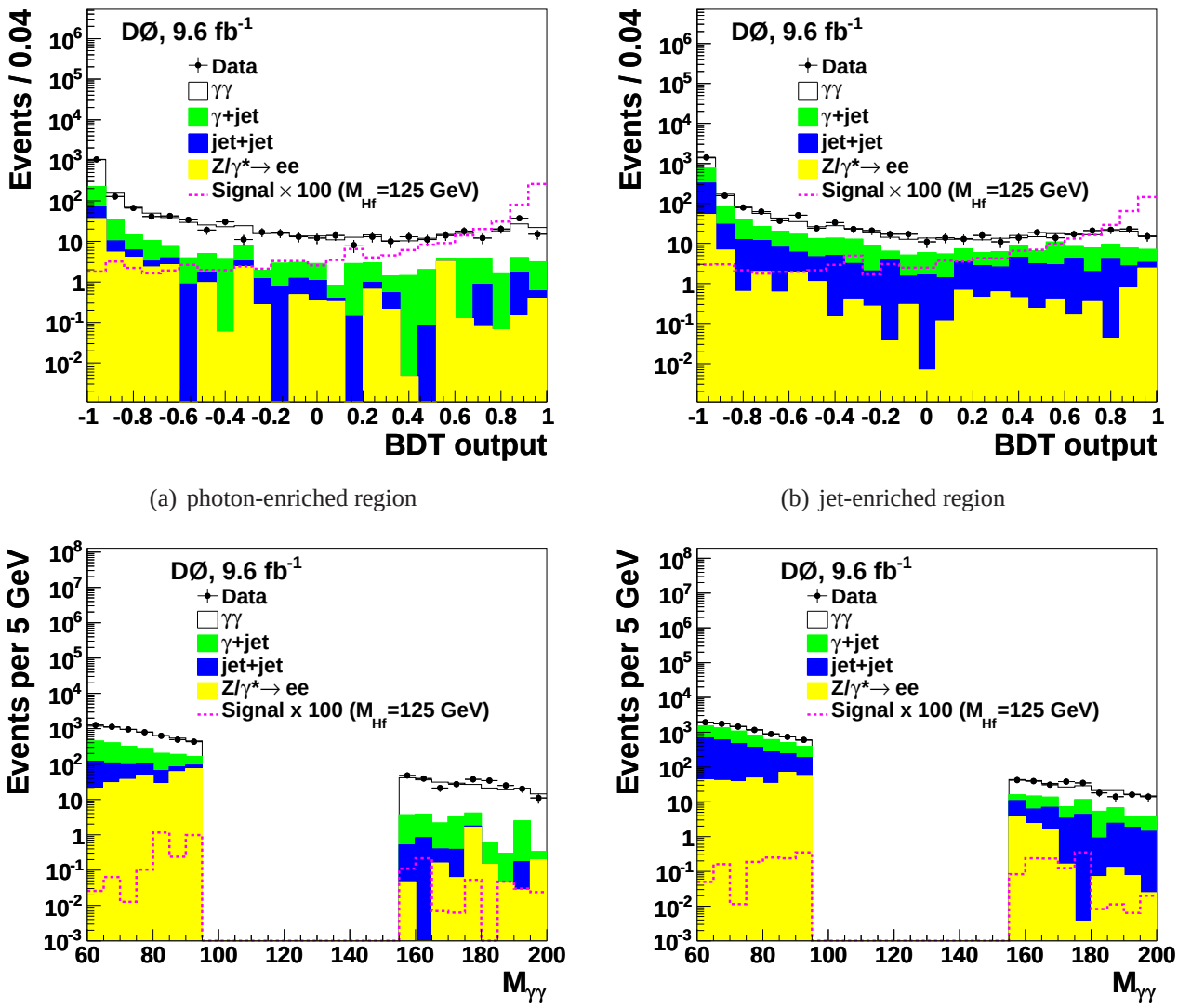
(a) photon-enriched region

(b) jet-enriched region

(c) photon-enriched region sideband

(d) jet-enriched region sideband

Figure 6.4: Final discriminants for the fermiophobic signal at 125 GeV in linear scale.



(a) photon-enriched region

(b) jet-enriched region

(c) photon-enriched region sideband

(d) jet-enriched region sideband

Figure 6.5: Final discriminants for the fermiophobic signal at 125 GeV in log scale.

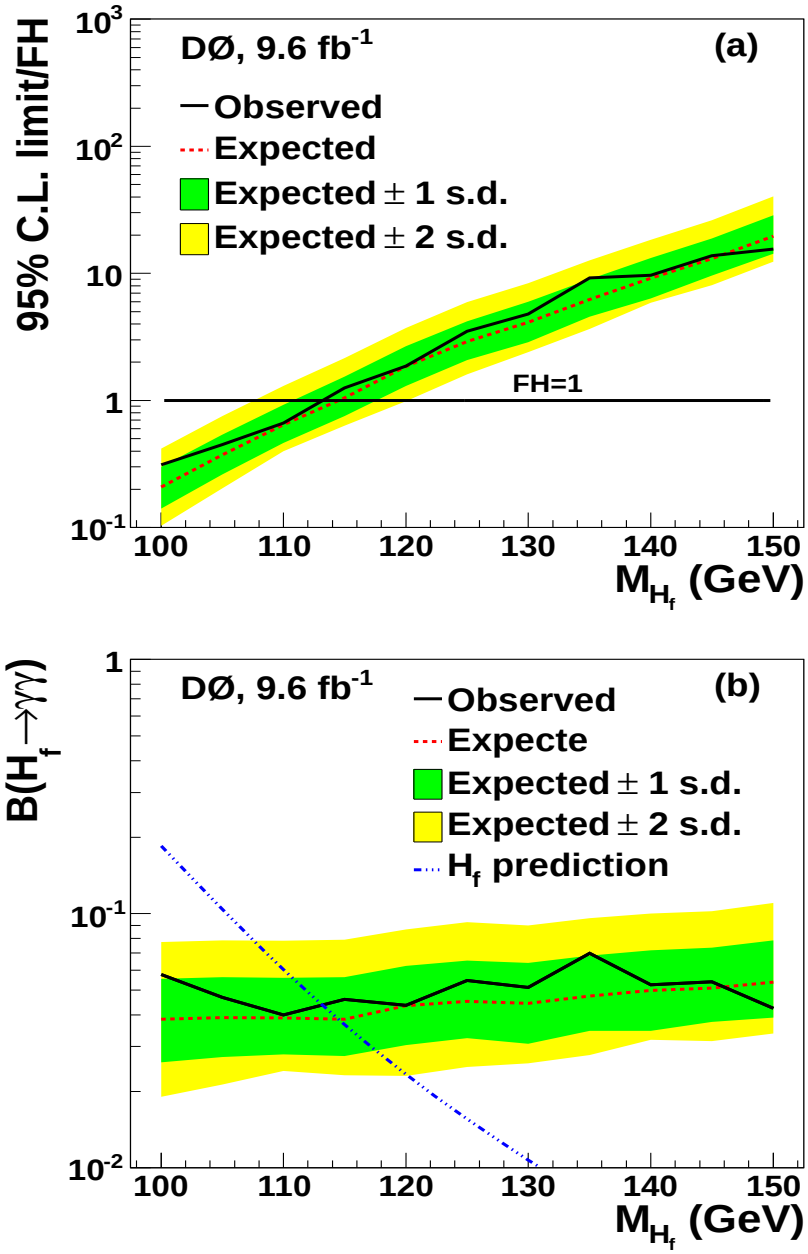


Figure 6.6: Results for the fermiophobic Higgs search. (a) Upper limits at 95% CL on the ratio of $\sigma \times \mathscr{B}(H_f \to \gamma\gamma)$ to the fermiophobic Higgs model prediction as a function of Higgs mass. The observed limit is shown as a solid black line while the expected limit under the background-only hypothesis is shown as a dashed red line. The green and yellow areas correspond to 1 and 2 standard deviations (s.d.) around the expected limits under the background-only hypothesis. (b) Upper limits at 95% CL on the value of $\mathscr{B}(H_f \to \gamma\gamma)$. The theoretical prediction from the fermiophobic Higgs model is shown as a blue dashed line.

Chapter 7

Conclusion

In this study, we presented a search for the SM Higgs boson in the diphoton final state using the complete data collected with the DØ detector at the Tevatron during the RunII period, April 2002-September 2011. The data corresponds to an integrated luminosity of 9.6 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. No significant excess of data above prediction is observed and so upper limits are set on the ratio of $\sigma \times \mathscr{B}(H \to \gamma\gamma)$ to the SM prediction at 95% CL, as shown in Figure 6.2. The expected limit reaches its minimum of 8.5 at 125 GeV, representing the best sensitivity search in the $H \to \gamma\gamma$ channel at the Tevatron.

Considering the SM-like Higgs boson observed at the LHC, the DØ data is consistent with both the background-only and the S+B hypothesis at 125 GeV.

This search is also extended to the fermiophobic interpretation, in which the gluon fusion signal is absent. The same analysis technique as that in the SM Higgs boson search is applied. No excess of data is observed and the upper limits are set on the branching ratio $\mathscr{B}(H_f \to \gamma\gamma)$ and compared to the theoretical predictions. The mass region below 113 GeV is excluded at 95% CL.

In the SM case, this search is combined with other searches at the Tevatron including $H \to b\bar{b}$, $H \to W^+W^-$, $H \to ZZ$ and $H \to \tau^+\tau^-$, to contribute to the overall Tevatron SM Higgs result. In the fermiophobic case, this search is combined with $H \to W^+W^-$ to provide the Tevatron's answer on the fermiophobic Higgs search.

References

[1] M. Gell-Mann, Phys. Lett. 8 (1964) 214;
G. Zweig, CERN-Report 8182/TH401 (1964);
H. Fritzsch, M. Gell-Mann and H. Leutwyler, Phys. Lett. B 47 (1973) 365;
D. Gross and F. Wilczek, Phys. Rev. Lett. 30 (1973) 1343;
H.D. Politzer, Phys. Rev. Lett. 30 (1973) 1346;
G.'t Hooft, Marseille Conference on Yang-Mills fields (1972).

[2] S. Glashow, Nuclear Phys. 22 (1961) 579;
S. Weinberg, Phys. Rev. Lett 19 (1967) 1264;
A. Salam, in: N. Svartholm (Ed.), Elementary Particle Theory, Almqvist and Wiksells, Stockholm, 1969, p.367.

[3] P. W. Higgs, Phys. Rev. Lett. 13 (1964) 508;
P. W. Higgs, Phys. Rev. 145 (1966) 1156;
F. Englert, R. Brout, Phys. Rev. Lett. 13 (1964) 321;
G. S. Guralnik, C.R. Hagen, T. Kibble, Phys. Rev. Lett. 13 (1965) 585;
T. Kibble, Phys. Rev. 155 (1967) 1554.

[4] ALEPH, DELPHI, L3, and OPAL Collaborations, The LEP Working Group for Higgs Boson Searches, Phys. Lett. B **565**, 61 (2003).

[5] The LEP Electroweak Working Group, "Status of March 2012"
http://lepewwg.web.cern.ch/LEPEWWG/

[6] Tevatron New Physics and Higgs Working Group and CDF and D0 Collaborations, arXiv:1207.0449 [hep-ex]

[7] T. Aaltonen *et al.* (CDF and D0 Collaboration), Phys. Rev. Lett. **109**, 071804 (2012);
T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **108**, 151803 (2012);
V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **108**, 151804 (2012).

[8] G. Aad *et al.* (ATLAS Collaboration), Phys. Lett. B. **716**, 1 (2012).

[9] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Lett. B. **710**, 26 (2012);
S. Chatrchyan *et al.* (CMS Collaboration), Phys. Lett. B. **716**, 30 (2012).

[10] S. Mrenna and J.D. Wells, Phys. Rev. D **63**, 015006 (2000).

[11] T. Plehn and M. Rauch, Europhys. Lett. **100**, 11002 (2012).

[12] http://en.wikipedia.org/wiki/Standard_Model

[13] http://www.hep.ph.ic.ac.uk/cms/physics/higgs.html

[14] http://www.hep.lu.se/atlas/thesis/egede/thesis-node10.html

[15] T. Hahn, S. Heinemeyer, F. Maltoni, G. Weiglein, S. Willenbrock, arXiv:hep-ph/0607308v2 (2006).

[16] A. Djouadi, Phys. Rept. **457** 1-216 (2008).

[17] http://www.physikblog.eu/wp-content/uploads/2011/04/higgs-gammagamma_egede.png

[18] H.E. Haber, G.L. Kane, and T. Sterling, Nucl. Phys. B **161**, 493 (1979).

[19] J.F. Gunion, R. Vega, and J. Wudka, Phys. Rev. D **42** 1673 (1990).

[20] J.L. Basdevant *et al.*, Phys. Lett. B **313**, 402 (1993).

[21] V. Barger *et al.*, arXiv:hep-ph/9211234.

[22] A.G. Akeroyd, Phys. Lett. B **368**, 89 (1996).

[23] G.L. Landsberg and K.T. Matchev, Phys. Rev. D **62**, 035004 (2000).

[24] X. Bu, "$H \to \gamma\gamma$ Search and Direct Photon Pair Production Differential Cross Section Measurement", PhD thesis, University of Science and Technology of China, (2010).

[25] The Tevatron New Phenomena and Higgs Working Group, "Cross Section and Branching Ratio Recommendations for Tevatron Higgs Searches", (2012).

[26] A. Heister *et al.* (ALEPH Collaboration), Phys. Lett. B **544**, 16 (2002).

[27] G. Abbiendi *et al.* (OPAL Collaboration), Phys. Lett. B **544**, 44 (2002).

[28] P. Achard *et al.* (L3 Collaboration), Phys. Lett. B **568**, 191 (2003).

[29] P. Abreu *et al.* (DELPHI Collaboration), Eur. Phys. J. C **35**, 313, (2004).

[30] A. Rosca (LEP Collaborations), arXiv:hep-ex/0212038.

[31] T. Affolder *et al.* (CDF Collaboration), Phys. Rev. D **64**, 092002 (2001).

[32] T. Affolder *et al.* (CDF Collaboration), Phys. Rev. Lett **103**, 061803 (2009).

[33] T. Affolder *et al.* (CDF Collaboration), arXiv:1207.6386.

[34] B. Abbott *et al.*, (D0 Collaboration), Phys. Rev. Lett. **82**, 2244 (1999).

[35] V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **101**, 051801 (2008).

[36] V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **107**, 151801 (2011).

[37] ATLAS Collaboration, Eur. Phys. J. C **72**, 2157 (2012).

[38] CMS Collaboration, J. High Energy Phys. **1209**, 111 (2012).

[39] M. Syphers, Fermilab Accelerator Division, Accelerator Concepts, Version 3.0 (2002).

[40] Fermilab Accelerator Division Document Databases, beamdocs.fnal.gov/AD/DocDB//Static/Lists//FullList.html

[41] http://www.phy.duke.edu/research/hep/images/tevatron.jpg

[42] V. Shiltsev, Status of Tevatron Collider Run II and Novel Technologies for Luminosity Upgrades, in: Proc. 2004 European Particle Accelerator Conference, Lucern, Switzerland, Vol. 1, 2004, p. 239

[43] V. M. Abazov *et al.*, Nucl. Instrum. Meth. A **565**, 463 (2006);
M. Abolins *et al.*, Nucl. Instrum. Methods Phys. Res. A **584**, 75 (2008);
R. Angstadt *et al.*, Nucl. Instrum. Methods Phys. Res. A **622**, 298 (2010).

[44] DØ Upgrade Collaboration, "DØ Silicon Tracker Technical Design Report", http://www-d0.fnal.gov/trigger/stt/smt/smt_tdr.ps.

[45] D. Adams et al. (DØ Collaboration), The D0 Upgrade: Central Fiber Tracker, Technical Design Report, DØ note 4164 (1999).

[46] M. Adams et al. (DØ Collaboration), Central Fiber Tracker Technical Design Report, DØ note 3014 (1996).

[47] DØ Collaboration, The DØ Upgrade: Forward Preshower, Muon System and Level 2 Trigger, DØ note 2894 (1996).

[48] DØ Collaboration, Calorimeter Electronics Upgrade for Run 2, (1998).

[49] http://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_dzero_detector.html

[50] S. Abachi et al. (DØ Collaboration), The DØ Detector, Nucl. Instr. and Methods, A338, 185 (1994).

[51] T. Edwards, *et al.*, FERMILAB-TM-2278-E (2004).

[52] http://www.physicsmasterclasses.org/exercises/keyhole/sp/detectors/part_det_interaction.gif

[53] A. Khanov, “HTF: histogramming method for finding tracks. The algorithm description.”, DØ note 3778, (2000).

[54] H. Greenlee, “The DØ Kalman Track Fit”, DØ note 4304, (2004).

[55] D. Adams, “Finding Tracks”, DØ note 2958, (1996);
G. Borissov, “AA track finding algorithm” Presentation given at the DØ plenary meeting on Feb 28th, (2003).

[56] F. Fleuret, “The DØ Electron/Photon Analysis Package EMAnalyze”, DØ note 3888, (2001).

[57] C. Peterson, T. Rognvaldsson and L. Lonnblad, “JETNET 3.0 A versatile Artificial Neural Nertwork Package”, Lund University Preprint LU-TP 93-29. Version 3.5 is used.

[58] D. Bandurin *et al.*, “Artificial neural network for forward photon selection”, DØ note 5895, (2009).

[59] A. Askew and D. Duggan, “CPS Variables for Photon Identification”, DØ note 4949, (2005).

[60] V.M. Abazov *et al.* (D0 Collaboration), Phys. Lett. B **653**, 378 (2007).

[61] O. Atramentov and Y. Maravin, “Utilizing CFT and SMT hits count for photon and electron reconstruction”, DØ Note 4444, (2004).

[62] D. Bandurin *et al.*, “Photon Identification for Run II data”, DØ note 6004, (2009).

[63] X. Bu *et al.*, “Search for Sneutrino resonance in the $e+\mu$ final states in RPV SUSY at DØ$'$, DØ note 5299, (2007).

[64] X. Bu *et al.*, “Photon Identification for RunIIb data”, DØ note 6057, (2011).

[65] V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **102**, 231801 (2009).

[66] X. Bu *et al.*, “η-dependent photon energy scale correction”, DØ note 6316, (2012).

[67] M. Vesterinen, “An Improved Energy Calibration and Monte Carlo Over-smearing for Electromagnetic Clusters”, DØ note 6040, (2010).

[68] H. Yin, “Measurement of the Forward-Backward Charge Asymmetry (A_{FB}) in $p\bar{p} \to Z/\gamma^* \to e^+e^-$ events at $\sqrt{s} = 1.96$ TeV”, DØ note 5867.

[69] C. Fabjan and F. Gianotti, Calorimetry for particle physics, Rev. Mod. Phys., Vol. 75 (2003).

[70] A. Melnitchouk and M. Wetstein, “Determination of the Constant term of the Electron Energy Resolution in the Central Calorimeter”, DØ note 5879, (2009).

[71] G. C. Blazey *et al.*, arXiv:hep-ex/0005012v2.

[72] V.M. Abazov *et al.* (D0 Collaboration), Nucl. Instrum. Methods A **620**, 490 (2010).

[73] http://d0server1.fnal.gov/Projects/Operations/D0RunII_DataTaking_files/image005.png

[74] DØ Collaboration, Data Quality Group, http://www-d0.fnal.gov/computing/data_quality.

[75] DØ Collaboration, Common Sample Group, http://www-d0.fnal.gov/Run2Physics/cs/.

[76] T. Sjöstrand *et al.*, J. High Energy Phys. **05**, 026 (2006). Version 6.409 is used.

[77] J. Pumplin *et al.*, J. High Energy Phys. **07**, 012 (2002);
D. Stump *et al.*, J. High Energy Phys. **10**, 046 (2003).

[78] C. Anastasiou, R. Boughezal, and F. Petriello, J. High Energy Phys. **04**, 003 (2009);
D. Florian and M. Grazzini, Phys. Lett. B **674**, 291 (2009).

[79] J. Baglio and A. Djouadi, J. High Energy Phys. **10**, 064 (2010).

[80] P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro, Phys. Rev. Lett. **105**, 011801 (2010).

[81] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, Eur. Phys. J. C **63**, 189 (2009).

[82] A. Djouadi, J. Kalinowski, and M. Spira, Comput. Phys. Commun. **108**, 56 (1998). Version 3.70 is used.

[83] G. Bozzi, S. Catani, D. de Florian, and M. Grazzini, Phys. Lett. B **564**, 65 (2003); Nucl. Phys. **B737**, 73 (2006).

[84] M. L. Mangano *et al.*, J. High Energy Phys. **07**, 001 (2003). Version 2.11 is used.

[85] R. Hamberg, W. L. van Neerven and T. Matsuura, Nucl. Phys. B **359**, 343 (1991) [Erratum-ibid. B **644**, 403 (2002)].

[86] V.M. Abazov *et al.*, (D0 Collaboration), Phys. Rev. Lett. **100**, 102002 (2008).

[87] T. Gleisberg *et al.*, J. High Energy Phys. **02**, 007 (2009). Version 1.2.2 is used.

[88] R. Brun and F. Carminati, CERN Program Library Long Writeup W5013, (1993).

[89] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **108**, 011801 (2012).

[90] V.M. Abazov *et al.*, (D0 Collaboration), Phys. Lett. B **690**, 108 (2010).

[91] Y. Maravin, (D0 Collaboration), "Optimization of central track matching to electromagnetic objects by utilizing central preshower stereo clusters", DØ note 4411, (2004).

[92] S. Kesisoglou, Brown University, Ph.D. Thesis, FERMILAB-THESIS-2004-44, UMI-31-74625, (2004).

[93] DØ Collaboration, Muon ID group, http://www-d0.fnal.gov/computing/algorithms/muon/muon_algo.html.

[94] L. Han *et al.*, "Measurement of the Forward-Backward Charge Asymmetry (A_{FB}) and Extraction of $\sin^2\theta_W^{eff}$ in $p\bar{p} \to Z/\gamma^* \to e^+e^-$ events at $\sqrt{s} = 1.96$ TeV", DØ note 5603, (2008).

[95] G. Chen, *et al.*, "Electron Preselection Efficiency study for RunII data", DØ note 6009, (2010).

[96] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. **95**, 022003 (2005).

[97] C. Balázs, E.L. Berger, P. Nadolsky, and C.-P. Yuan, Phys. Rev. D **76**, 013009, (2007).

[98] D0 Collaboration, Updated figures including predictions from SHERPA, http://www-d0.fnal.gov/Run2Physics/WWW/results/final/QCD/Q10B/

[99] J.E. Gaiser, PhD Thesis, SLAC-R-255 (1982).

[100] A. Hoecker *et al.*, arXiv:physics/0703039 [physics.data-an]; A. Hoecker *et al.*, PoS (ACAT) **040** (2007).

[101] E. Aguiló *et al.*, "Using Boosted Decision Trees to Search for Single Top Quarks in 1 fb^{-1} of Data", DØ note 5286, (2007).

[102] Y. Freund and R.E. Schapire, "Experiments with a New Boosting Algorithm" in Machine Learning: Proceedings of the Thirteenth International Confernece", pp. 148-156 (1996).

[103] J.C. Collins and D.E. Soper, Phys. Rev. D **16**, 2219 (1977).

[104] S. Alekhin, S. Alioli, R. D. Ball *et al.*, arXiv:1101.0536 [hep-ph]; M. Botje, J. Butterworth, A. Cooper-Sarkar *et al.*, arXiv:1101.0538 [hep-ph]

[105] I. W. Stewart and F. J. Tackmann, Phys. Rev. D **85**, 034011 (2012).

[106] T. Junk, Nucl. Instrum. Methods A **434**, 435 (1999); A. Read, J. Phys G **28**, 2693 (2002).

[107] W. Fisher, FERMILAB-TM-2386-E (2006).

[108] D. Gerbaudo, "Search for a Standard Model Higgs Boson with a Dilepton and Missing Energy Signature", PhD thesis, Princeton University, (2011).

Appendix A

Drell-Yan $Z/\gamma^* \rightarrow ee$ background

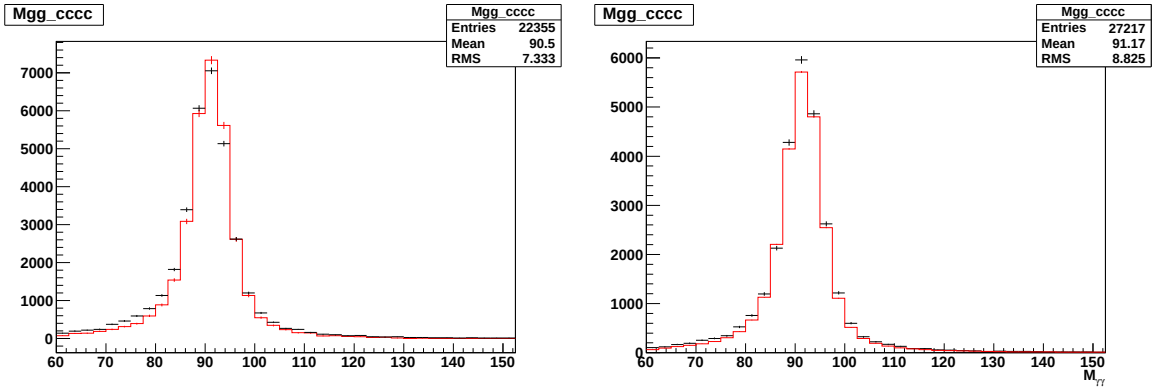
To validate the Z/γ^* background estimated from MC, especially the track-veto scale factors from reference [62] we use the RunIIb1 dataset and MC to test Z/γ^* yields in the following categories:

- both electrons are track-matched
- one electron is track-matched and the other electron is track-vetoed
- both electrons are track-vetoed

According to the photon ID algorithm [62], we define:
"track-matched" $= track_match_spatialchi2prob > 0.0$
"track-vetoed" $= track_match_spatialchi2prob < 0.0$ and $emhits_e_f_discriminant() < 0.9$(hits on the road)

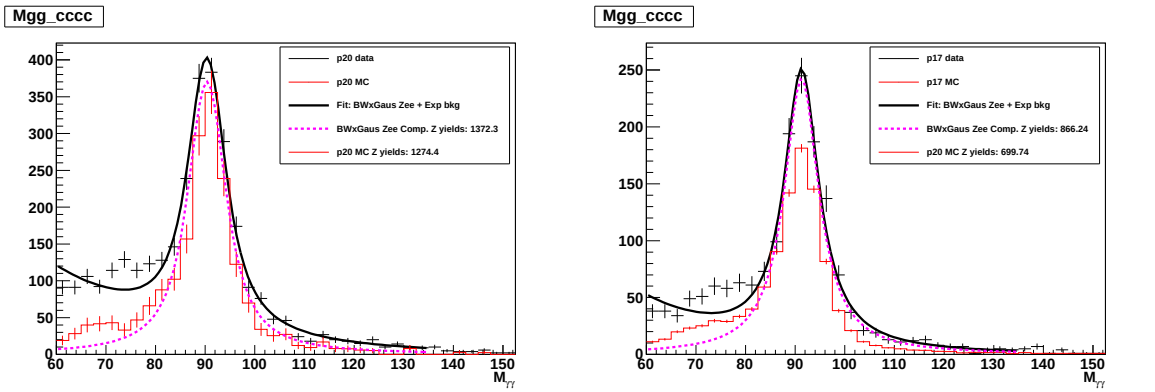
The two-track-matched Z/γ^* distribution is shown in Figure A.1. The one-track-matched-one-track-vetoed Z/γ^* distribution is shown in Figure A.2. The two-track-vetoed Z/γ^* distribution is shown in Figure A.3.

In the two-track-vetoed case, the Z/γ^* contribution is already very small.



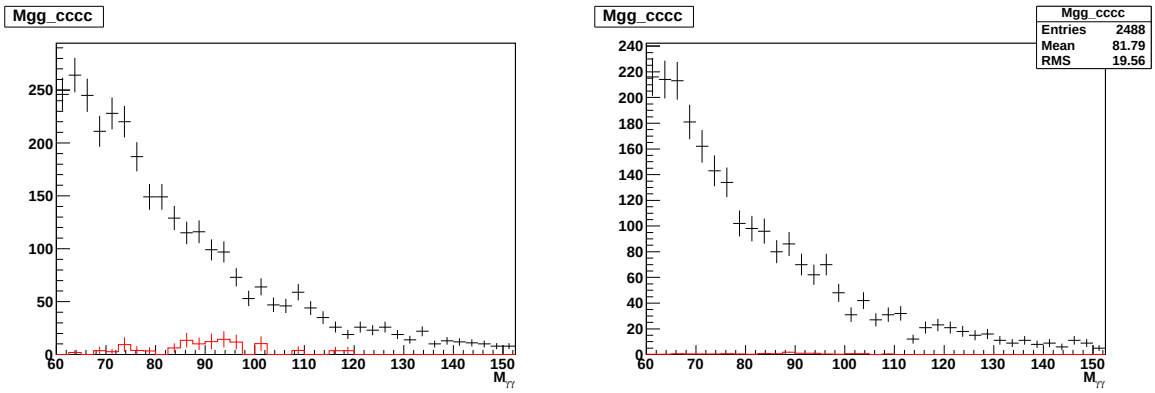
(a) p20 CCCC (b) p17 CCCC

Figure A.1: The Z/γ^* invariant mass distribution by requiring that both electrons have tracks.



(a) p20 CCCC (b) p17 CCCC

Figure A.2: The Z/γ^* invariant mass distribution by requiring that one electron has a track and one electron does not. It is fit with an exponentially decaying function to model the smooth background, and with a convolution of a Gaussian function and a Breit-Wigner function to model the Z boson signal. In the mass window of $[85, 100]$ GeV, the fit yields 4005.7 Z bosons and the Monte Carlo yields 3755.1 Z bosons.



(a) p20 CCCC (b) p17 CCCC

Figure A.3: The Z/γ^* invariant mass distribution by requiring that both electrons have no track.

Appendix B

SM systematic uncertainties: ε_γ and ε_j

For the 4x4 matrix method, the efficiency matrix needs the efficiencies for photons and EM-like jets passing the $O_{NN} > 0.75$ cut: ε_γ and ε_j. They are estimated from Monte Carlo so that the statistical uncertainties ($\sim 0.4\%$) are negligible compared to their systematic uncertainties ($1.5\%, 10\%$). Varying ε_γ and ε_j about their systematic uncertainties affects the normalization of the γ+jet and jet+jet contributions, thus further affecting the normalization and shape of the sum of γ+jet and jet+jet components (also called non-$\gamma\gamma$).

We vary ε_γ and ε_j individually and compare the γ+jet and jet+jet yields to the original solution from the 4x4 matrix method as shown in Table B.1.

Variation scenarios	$\delta N_{jj}/N_{jj}$	$\delta N_{gj}/N_{gj}$
$\varepsilon_\gamma + 1.5\%$, ε_j	5.4%	6.5%
$\varepsilon_\gamma - 1.5\%$, ε_j	-5.2%	-7.3%
ε_γ , $\varepsilon_j + 10.0\%$	17%	-0.3%
ε_γ , $\varepsilon_j - 10.0\%$	-13.6%	-0.8%

Table B.1: Yields change in % from the 4x4 matrix method in four scenarios of varying ε_γ and ε_j about their uncertainties. N_{jj} and N_{gj} are the yields of jet+jet and γ+jet respectively.

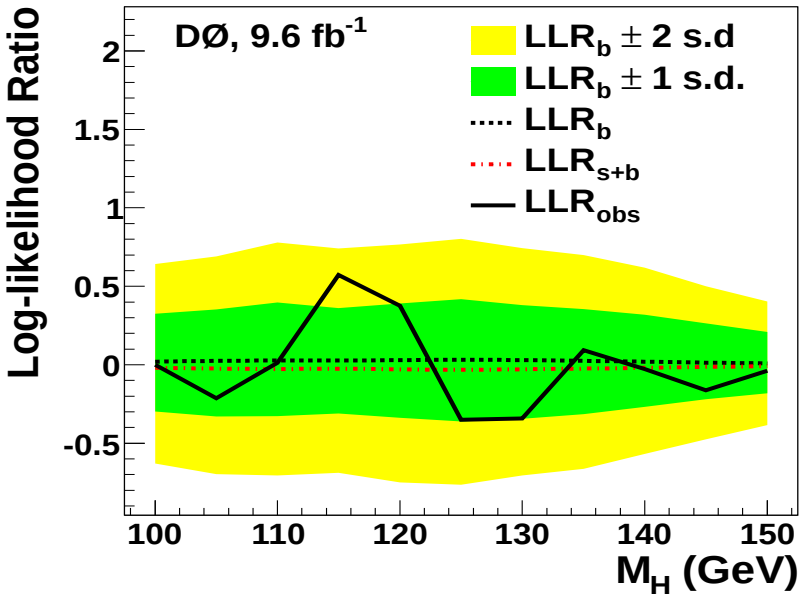
Appendix C

Diphoton Invariant Mass Method

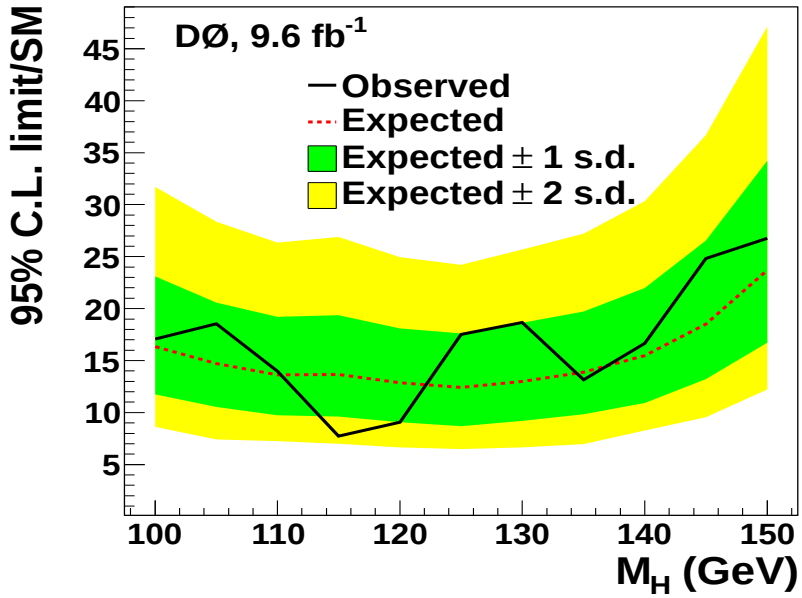
Using only the diphoton invariant mass as the final discriminant, we have the results of the log-likelihood ratio and limits shown in Table C.1 and Figure C.1. The shape of the median expected limits (dashed red line in Figure C.1 (b)) has a minimum, reaching its lowest value around 125 GeV. This shape is consistent with the expected $\sigma \times \mathscr{B}$ reaching its maximum at $\sim$ 125 GeV. The best sensitivity given by this mass method is 12.4. Almost all of the observed limits are consistent with the limits from the background-only hypothesis within 1σ, except for a deficit at 115 GeV and an excess at 125 GeV but within 2σ.

Higgs mass (GeV)	100	105	110	115	120	125	130	135	140	145	150
Expected limit	16.3	14.7	13.6	13.7	12.9	12.4	13.0	13.9	15.5	18.5	23.7
Observed limit	17.1	18.5	14.0	7.7	9.1	17.5	18.7	13.2	16.6	24.8	26.8

Table C.1: Expected and observed limits on the ratio of $\sigma \times \mathscr{B}(H \to \gamma\gamma)$ to the SM prediction as a function of Higgs mass using the BDT.



(a) LLR



(b) Limits

Figure C.1: Results for the SM Higgs boson search from using the diphoton invariant mass as the final discriminant. (a) Log-likelihood ratio as a function of Higgs mass. The observed limit is shown as a solid black line while the expected limit under the background-only hypothesis is shown as a dashed red line. The green and yellow areas correspond to 1 and 2 standard deviations (s.d.) around the median LLR from the background-only hypothesis. (b) Limits at 95% CL on $\sigma \times \mathscr{B}(H \to \gamma\gamma)$ relative to the SM prediction as a function of Higgs mass.