

**Shining Light on Dark Matter,
One Photon at a Time**

by

Brandon Leigh Allen

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

Doctorate of Science in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Author
Department of Physics
May 18, 2019

Certified by
Christoph E.M. Paus
Professor
Thesis Supervisor

Accepted by
Arthur C. Smith
Chairman, Department Committee on Graduate Theses

Shining Light on Dark Matter, One Photon at a Time

by

Brandon Leigh Allen

Submitted to the Department of Physics
on May 18, 2019, in partial fulfillment of the
requirements for the degree of
Doctorate of Science in Physics

Abstract

A search is conducted for new physics in final states containing a photon and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. The data collected by the CMS experiment at the CERN LHC correspond to an integrated luminosity of 35.9 inverse femtobarns. No deviations from the predictions of the standard model are observed. The results are interpreted in the context of dark matter production and limits on new physics parameters are calculated at 95% confidence level. For the two simplified dark matter production models considered, the observed (expected) lower limits on the mediator masses are both 950 (1150) GeV for 1 GeV dark matter mass.

Thesis Supervisor: Christoph E.M. Paus

Title: Professor

Acknowledgments

This is the acknowledgements section. You should replace this with your own acknowledgements.

Contents

1	Introduction	15
2	Motivation	17
2.1	Astrophysical Evidence for Dark Matter	17
2.2	Simplified Models for the WIMP Paradigm	17
3	The CMS Detector	19
3.1	Detector Overview	19
3.1.1	Silicon Pixel Detector	19
3.1.2	Silicon Strip Tracker	19
3.1.3	Electromagnetic Calorimeter	19
3.1.4	Hadronic Calorimeter	19
3.1.5	Muon Detectors	20
4	Reconstruction	21
4.1	Reconstruction	21
4.1.1	Particle Flow	21
4.1.2	MET Reconstruction	21
4.2	Calibration	21
4.2.1	Trigger Efficiency	22
4.2.2	Photon Scale Factor	22
4.2.3	Pixel Veto Scale Factor	22
4.2.4	Lepton Scale Factors	22

4.2.5	Jet Energy Scale	22
4.3	Non-collision signatures	22
4.3.1	Beam halo	22
4.3.2	ECAL Spikes	22
4.3.3	Fake MET	22
5	The Monophoton Analysis	23
5.1	Event Selection	23
5.2	Irreducible backgrounds	23
5.3	Misidentified backgrounds	25
5.3.1	Electrons	25
5.3.2	Hadrons	25
5.4	Non-collision backgrounds	25
5.4.1	Spikes	25
5.4.2	Beam halo	25
5.5	Statistical Interpretation	26
5.6	Results	27
5.6.1	Pre-fit and post-fit distributions	27
5.6.2	Limits	31
6	Comparison with Other Results	33
6.1	Monophoton	33
6.2	Monojet / Mono- Z	33
6.3	Direct Detection	33
6.4	Indirect Detection	33
7	Conclusion	35

List of Figures

- 5-1 Transfer factors $R_{ee\gamma}^{Z\gamma}$ (left) and $R_{\mu\mu\gamma}^{Z\gamma}$ (right). The uncertainty bands in green (inner) and orange (outer) show the systematic uncertainty, and the combination of systematic and statistical uncertainty arising from limited MC sample size, respectively. The systematic uncertainties considered are the uncertainties in the data-to-simulation correction factors ρ for the lepton identification efficiencies. 23
- 5-2 Transfer factors $R_{e\gamma}^{W\gamma}$ (left) and $R_{\mu\gamma}^{W\gamma}$ (right). The uncertainty bands in green (inner) and orange (outer) show the systematic uncertainty, and the combination of systematic and statistical uncertainty arising from limited MC sample size, respectively. The systematic uncertainties considered are the uncertainties in the data-to-simulation correction factors ρ for the lepton identification efficiencies. 24
- 5-3 Transfer factor $f_{W\gamma}^{Z\gamma}$. The uncertainty bands in green (inner) and orange (outer) show the systematic uncertainty, and the combination of systematic and statistical uncertainty arising from limited MC sample size, respectively. The systematic uncertainties considered are the uncertainties from higher-order theoretical corrections. 25

5-4 Comparison between data and MC simulation in the four control regions: $e e \gamma$ (upper left), $\mu \mu \gamma$ (upper right), $e \gamma$ (lower left), $\mu \gamma$ (lower right) before and after performing the simultaneous fit across all the control samples and signal region, and assuming absence of any signal. The last bin of the distribution includes all events with $E_T^\gamma > 1000$ GeV. The ratios of data with the pre-fit background prediction (red dashed) and post-fit background prediction (blue solid) are shown in the lower panels. The bands in the lower panels show the post-fit uncertainty after combining all the systematic uncertainties. 28

5-5 Observed E_T^γ distributions in the horizontal (left) and vertical (right) signal regions compared with the post-fit background expectations for various SM processes. The last bin of the distribution includes all events with $E_T^\gamma > 1000$ GeV. The expected background distributions are evaluated after performing a combined fit to the data in all the control samples and the signal region. The ratios of data with the pre-fit background prediction (red dashed) and post-fit background prediction (blue solid) are shown in the lower panels. The bands in the lower panels show the post-fit uncertainty after combining all the systematic uncertainties. The expected signal distribution from a 1 TeV vector mediator decaying to 1 GeV DM particles is overlaid. 29

5-6 Covariances between the predicted background yields in all the E_T^γ bins of the horizontal and vertical signal regions. The bin labels specify which signal region the bin belongs to and what number bin it is for that region. 30

5-7	The ratio of 95% CL upper cross section limits to the theoretical cross section (μ_{95}), for DM simplified models with vector (left) and axial-vector (right) mediators, assuming $g_q = 0.25$ and $g_{\text{DM}} = 1$. Expected $\mu_{95} = 1$ contours are overlaid in red. The region under the observed contour is excluded. For DM simplified model parameters in the region below the lower violet dot-dash contour, and also above the corresponding upper contour in the right hand plot, cosmological DM abundance exceeds the density observed by the Planck satellite experiment.	31
-----	--	----

List of Tables

- 5.1 Expected event yields in each E_T^γ bin for various background processes in the horizontal signal region. The background yields and the corresponding uncertainties are obtained after performing a combined fit to data in all the control samples, excluding data in the signal region. The observed event yields in the horizontal signal region are also reported. 27
- 5.2 Expected event yields in each E_T^γ bin for various background processes in the vertical signal region. The background yields and the corresponding uncertainties are obtained after performing a combined fit to data in all the control samples, excluding data in the signal regions. The observed event yields in the vertical signal region are also reported. 29

Chapter 1

The Monophoton Analysis

The main event.

1.1 Event Selection

1.2 Irreducible backgrounds

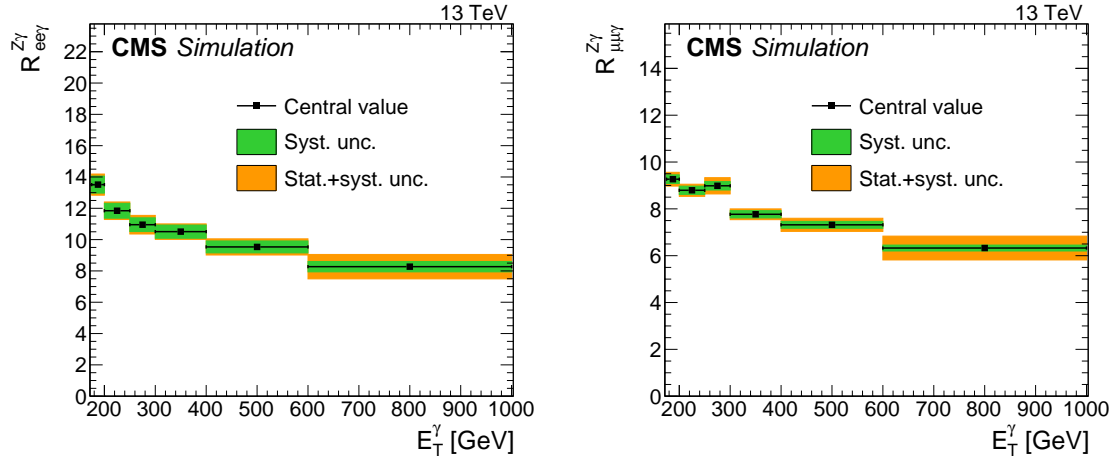


Figure 1-1: Transfer factors $R_{ee\gamma}^{Z\gamma}$ (left) and $R_{\mu\mu\gamma}^{Z\gamma}$ (right). The uncertainty bands in green (inner) and orange (outer) show the systematic uncertainty, and the combination of systematic and statistical uncertainty arising from limited MC sample size, respectively. The systematic uncertainties considered are the uncertainties in the data-to-simulation correction factors ρ for the lepton identification efficiencies.

Using the transfer factor $R_{\ell\ell\gamma}^{Z\gamma}$, the total estimated event yield $T_{\ell\ell\gamma}$ in each dilepton control region in the i^{th} bin of the E_T^γ distribution can be expressed as

$$T_{\ell\ell\gamma,i} = \frac{N_i^{Z\gamma}}{R_{\ell\ell\gamma,i}^{Z\gamma}} + b_{\ell\ell\gamma,i}, \quad (1.1)$$

where $N^{Z\gamma}$ is the number of $Z(\rightarrow \nu\bar{\nu})+\gamma$ events in the combined signal regions and $b_{\ell\ell\gamma}$ is the predicted contribution from other background sources in the dilepton control region, namely $t\bar{t}\gamma$, $VV\gamma$, and misidentified hadrons. The subscript i indicates that the quantities are evaluated in bin i of the E_T^γ distribution.

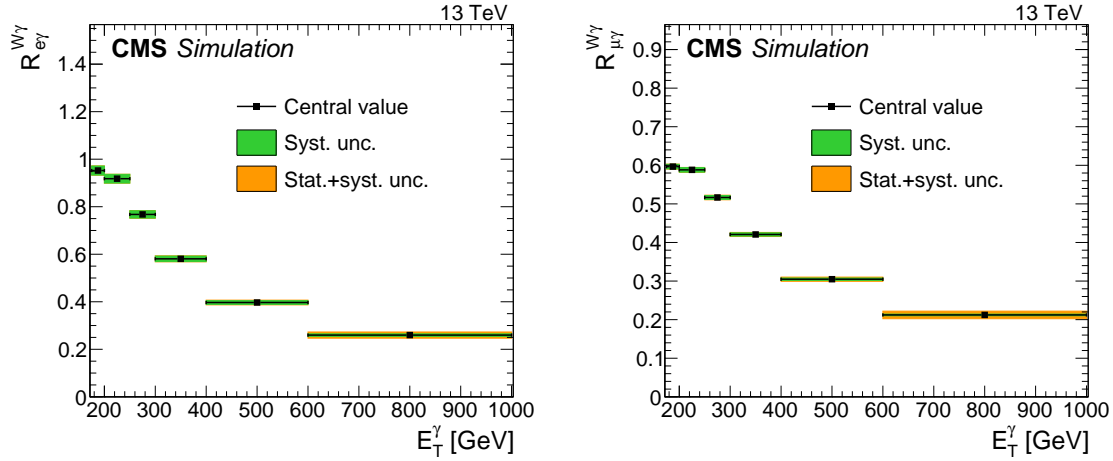


Figure 1-2: Transfer factors $R_{e\gamma}^{W\gamma}$ (left) and $R_{\mu\gamma}^{W\gamma}$ (right). The uncertainty bands in green (inner) and orange (outer) show the systematic uncertainty, and the combination of systematic and statistical uncertainty arising from limited MC sample size, respectively. The systematic uncertainties considered are the uncertainties in the data-to-simulation correction factors ρ for the lepton identification efficiencies.

Using $R_{\ell\gamma}^{W\gamma}$ and $f_{W\gamma}^{Z\gamma}$, the total estimated event yield $T_{\ell\gamma}$ in each single-lepton control region in the i^{th} bin of the E_T^γ distribution can be expressed as

$$T_{\ell\gamma,i} = \frac{N_i^{Z\gamma}}{R_{\ell\gamma,i}^{W\gamma} f_{W\gamma,i}^{Z\gamma}} + b_{\ell\gamma,i}, \quad (1.2)$$

where $b_{\ell\gamma}$ is the predicted contribution from other background sources in the single-lepton regions, namely misidentified electrons and hadrons and other minor SM processes.

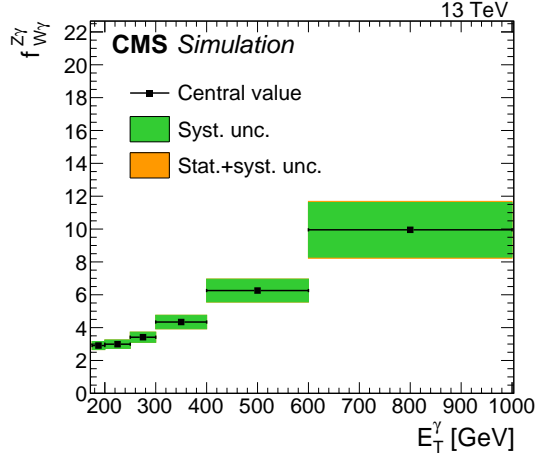


Figure 1-3: Transfer factor $f_{W\gamma}^{Z\gamma}$. The uncertainty bands in green (inner) and orange (outer) show the systematic uncertainty, and the combination of systematic and statistical uncertainty arising from limited MC sample size, respectively. The systematic uncertainties considered are the uncertainties from higher-order theoretical corrections.

1.3 Misidentified backgrounds

1.3.1 Electrons

1.3.2 Hadrons

1.4 Non-collision backgrounds

1.4.1 Spikes

1.4.2 Beam halo

The splitting of the signal region can be thought of as a two-bin fit. Collision processes occupy the relative fractions of phase space in the horizontal (H) and vertical (V) signal regions, $C_H = 1/\pi$ and $C_V = (\pi - 1)/\pi$, respectively. The corresponding fractions for beam halo events are determined by selecting a halo-enriched sample where the halo identification is inverted. Thus, a fit of the two signal regions provides an estimate of the overall normalization of the beam halo background, denoted h . The E_T^γ dependence of the halo background is encoded in $n_{K,i}^{\text{halo}}$, the unit-normalized

beam halo prediction in the i^{th} bin of the signal region $K \in \{H, V\}$. Using the notation introduced in Section 5.2, the total estimated background T_K in the two signal regions are

$$\begin{aligned} T_{K,i} &= C_K(N_i^{Z\gamma} + N_i^{W\gamma}) + hn_{K,i}^{\text{halo}} + C_K b_{K,i} \\ &= C_K(1 + f_{W\gamma,i}^{Z\gamma})^{-1} N_i^{Z\gamma} + hn_{K,i}^{\text{halo}} + C_K b_{K,i}, \end{aligned} \quad (1.3)$$

where $b_{K,i}$ is the total contribution to bin i of region K from electron and hadron misidentification, ECAL spikes, and other minor SM background processes.

1.5 Statistical Interpretation

Free parameters of the fit are the yield of $Z(\rightarrow \nu\bar{\nu})+\gamma$ background in each bin of the signal regions ($N_i^{Z\gamma}$) and the overall normalization of the beam halo background (h). Bin-by-bin yields of $W(\rightarrow \ell\nu)+\gamma$ and $Z(\rightarrow \ell\bar{\ell})+\gamma$ samples in all regions are related to the yield of $Z(\rightarrow \nu\bar{\nu})+\gamma$ through the MC prediction through the transfer factors defined in Section 5.2. The transfer factors are allowed to shift within the aforementioned theoretical and experimental uncertainties.

The background-only likelihood that is maximized in the fit is

$$\begin{aligned} \mathcal{L} &= \prod_i \{ \mathcal{L}_{\text{signal}} \times \mathcal{L}_{\text{single-lepton}} \times \mathcal{L}_{\text{dilepton}} \} \times \mathcal{L}_{\text{nuisances}} \\ &= \prod_i \left\{ \prod_{K=H,V} \mathcal{P}(d_{K,i} | T_{K,i}(\vec{\theta})) \times \prod_{\ell=e,\mu} \mathcal{P}(d_{\ell\gamma,i} | T_{\ell\gamma,i}(\vec{\theta})) \times \prod_{\ell=e,\mu} \mathcal{P}(d_{\ell\ell\gamma,i} | T_{\ell\ell\gamma,i}(\vec{\theta})) \right\} \times \prod_j \mathcal{N}(\theta_j) \\ &= \prod_i \left\{ \prod_{K=H,V} \mathcal{P}\left(d_{K,i} \left| \left(1 + f_{W\gamma,i}^{Z\gamma}\right)^{-1} C_K N_i^{Z\gamma} + hn_{K,i}^{\text{halo}}(\vec{\theta}) + C_K b_{K,i}(\vec{\theta}) \right.\right) \right. \\ &\quad \times \prod_{\ell=e,\mu} \mathcal{P}\left(d_{\ell\gamma,i} \left| \frac{N_i^{Z\gamma}}{R_{\ell\gamma,i}^{W\gamma}(\vec{\theta}) f_{W\gamma,i}^{Z\gamma}(\vec{\theta})} + b_{\ell\gamma,i}(\vec{\theta}) \right.\right) \\ &\quad \times \prod_{\ell=e,\mu} \mathcal{P}\left(d_{\ell\ell\gamma,i} \left| \frac{N_i^{Z\gamma}}{R_{\ell\ell\gamma,i}^{Z\gamma}(\vec{\theta})} + b_{\ell\ell\gamma,i}(\vec{\theta}) \right.\right) \left. \right\} \times \prod_j \mathcal{N}(\theta_j), \end{aligned} \quad (1.4)$$

following the notation introduced in Section 5.2, and where $\mathcal{P}(n|\lambda)$ is the Poisson

probability of n for mean λ , \mathcal{N} denotes the unit normal distribution, and $d_{X,i}$ is the observed number of events in bin i of region X . Systematic uncertainties are treated as nuisance parameters in the fit and are represented by $\vec{\theta}$. Each quantity Q_j with a nominal value \overline{Q}_j and a standard deviation of the systematic uncertainty σ_j appears in the likelihood function as $\overline{Q}_j \exp(\sigma_j \theta_j)$.

1.6 Results

1.6.1 Pre-fit and post-fit distributions

Figure 5-4 shows the observed E_T^γ distributions in the four control regions compared with the results from simulations before and after performing the simultaneous fit across all the control samples and signal region, and assuming absence of any signal. Figure 5-5 shows the observed E_T^γ distributions in the horizontal and vertical signal regions compared with the results from simulations before and after performing a combined fit to the data in all the control samples and the signal region. The observed distributions are in agreement with the prediction from SM and noncollision backgrounds.

Table 1.1: Expected event yields in each E_T^γ bin for various background processes in the horizontal signal region. The background yields and the corresponding uncertainties are obtained after performing a combined fit to data in all the control samples, excluding data in the signal region. The observed event yields in the horizontal signal region are also reported.

E_T^γ [GeV]	[175, 200]	[200, 250]	[250, 300]	[300, 400]	[400, 600]	[600, 1000]
$Z\gamma$	81.2 ± 8.0	88.2 ± 8.4	38.8 ± 4.8	26.8 ± 3.7	8.8 ± 1.9	1.4 ± 0.7
$W\gamma$	27.9 ± 3.7	29.9 ± 3.9	11.4 ± 1.7	6.3 ± 1.2	1.4 ± 0.4	0.1 ± 0.1
Misid. electrons	22.5 ± 2.7	25.7 ± 2.7	10.5 ± 1.0	8.2 ± 0.7	2.7 ± 0.2	0.5 ± 0.0
Misid. hadrons	5.2 ± 2.2	9.3 ± 1.8	3.1 ± 0.7	1.0 ± 0.3	0.4 ± 0.1	0.0 ± 0.0
Other SM	13.6 ± 2.0	19.6 ± 1.3	13.9 ± 0.4	4.2 ± 0.2	0.8 ± 0.0	0.1 ± 0.0
ECAL spikes	4.3 ± 1.3	2.7 ± 0.8	0.5 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Total prediction	154.6 ± 8.3	175.4 ± 8.8	78.2 ± 5.3	46.6 ± 4.0	14.1 ± 2.1	2.1 ± 0.8
Observed	150 ± 12	166 ± 13	76.0 ± 8.7	44.0 ± 6.6	19.0 ± 4.4	4.0 ± 2.0

The expected yields in each bin of E_T^γ for all backgrounds in the horizontal and

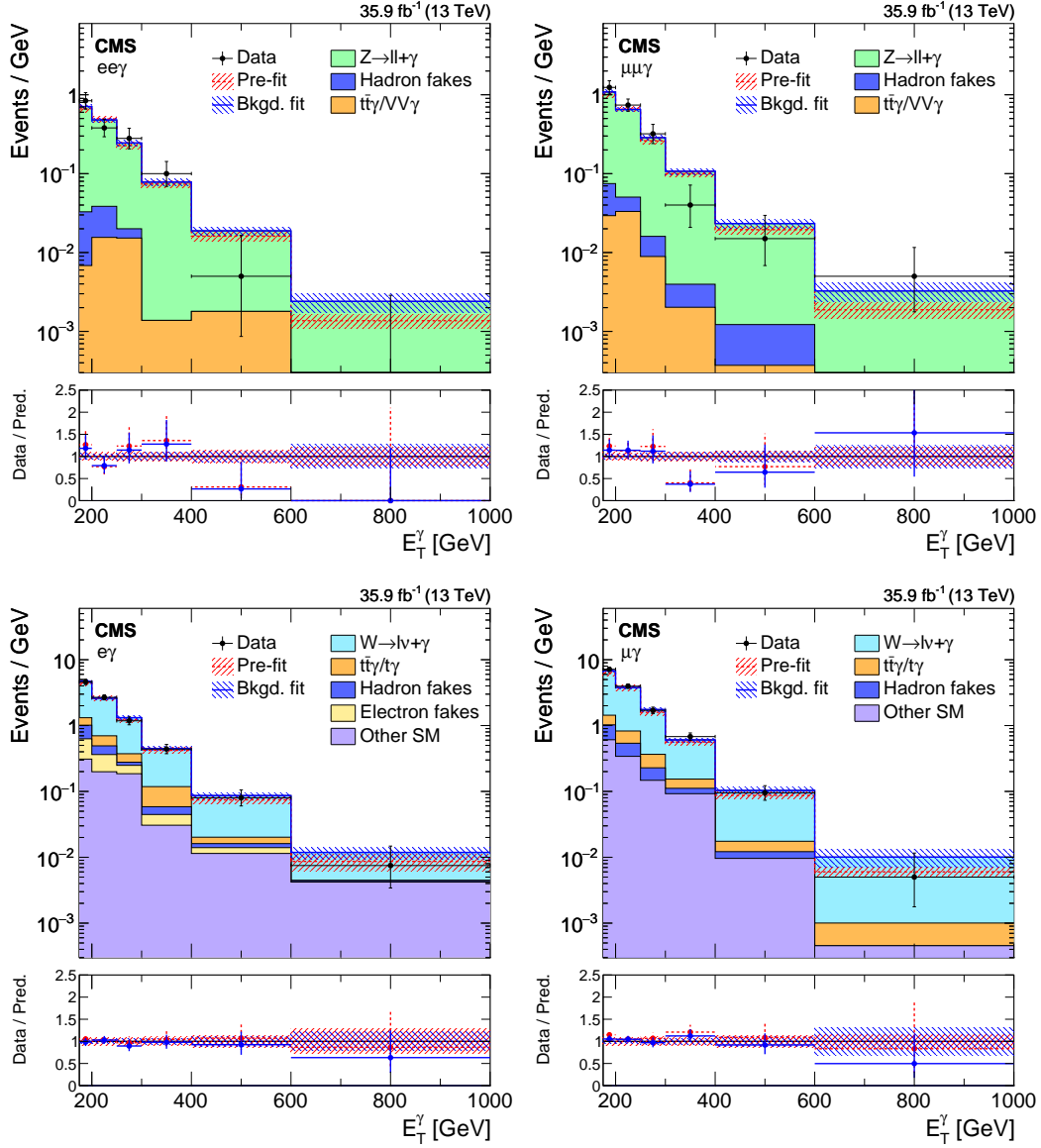


Figure 1-4: Comparison between data and MC simulation in the four control regions: $e e \gamma$ (upper left), $\mu \mu \gamma$ (upper right), $e \gamma$ (lower left), $\mu \gamma$ (lower right) before and after performing the simultaneous fit across all the control samples and signal region, and assuming absence of any signal. The last bin of the distribution includes all events with $E_T^\gamma > 1000$ GeV. The ratios of data with the pre-fit background prediction (red dashed) and post-fit background prediction (blue solid) are shown in the lower panels. The bands in the lower panels show the post-fit uncertainty after combining all the systematic uncertainties.

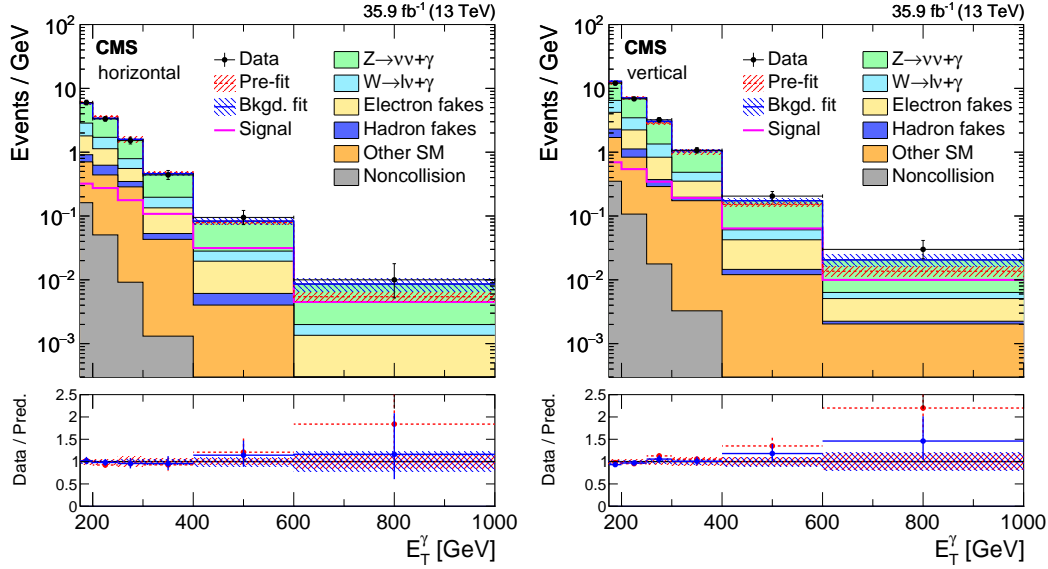


Figure 1-5: Observed E_T^γ distributions in the horizontal (left) and vertical (right) signal regions compared with the post-fit background expectations for various SM processes. The last bin of the distribution includes all events with $E_T^\gamma > 1000$ GeV. The expected background distributions are evaluated after performing a combined fit to the data in all the control samples and the signal region. The ratios of data with the pre-fit background prediction (red dashed) and post-fit background prediction (blue solid) are shown in the lower panels. The bands in the lower panels show the post-fit uncertainty after combining all the systematic uncertainties. The expected signal distribution from a 1 TeV vector mediator decaying to 1 GeV DM particles is overlaid.

Table 1.2: Expected event yields in each E_T^γ bin for various background processes in the vertical signal region. The background yields and the corresponding uncertainties are obtained after performing a combined fit to data in all the control samples, excluding data in the signal regions. The observed event yields in the vertical signal region are also reported.

E_T^γ [GeV]	[175, 200]	[200, 250]	[250, 300]	[300, 400]	[400, 600]	[600, 1000]
$Z\gamma$	172 ± 17	190 ± 18	83 ± 10	58.6 ± 7.9	18.0 ± 3.9	3.1 ± 1.6
$W\gamma$	59.9 ± 7.8	63.6 ± 7.8	24.6 ± 3.5	13.4 ± 2.4	3.0 ± 0.8	0.3 ± 0.2
Misid. electrons	48.4 ± 5.6	56.2 ± 5.1	23.4 ± 1.8	15.7 ± 1.4	5.6 ± 0.4	1.2 ± 0.1
Misid. hadrons	15.1 ± 4.4	14.5 ± 3.1	4.2 ± 0.8	2.3 ± 0.8	0.5 ± 0.1	0.1 ± 0.1
Other SM	33.8 ± 4.1	36.6 ± 2.7	13.6 ± 0.5	17.1 ± 0.6	2.4 ± 0.1	0.8 ± 0.0
ECAL spikes	9.3 ± 2.8	5.7 ± 1.7	0.9 ± 0.3	0.3 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
Total prediction	339 ± 18	366 ± 19	150 ± 11	107.5 ± 8.7	29.6 ± 4.3	5.4 ± 1.7
Observed	301 ± 17	342 ± 19	161 ± 13	107 ± 10	41.0 ± 6.4	12.0 ± 3.5

vertical signal regions after performing a combined fit to data in all the control samples, excluding data in the signal regions, are given in Tables 5.1 and 5.2, respectively. The covariances between the predicted background yields across all the E_T^γ bins in the two signal regions are shown in Fig. 5-6. The expected yields together with the covariances can be used with the simplified likelihood approach detailed in Ref. [?] to reinterpret the results for models not studied in this thesis

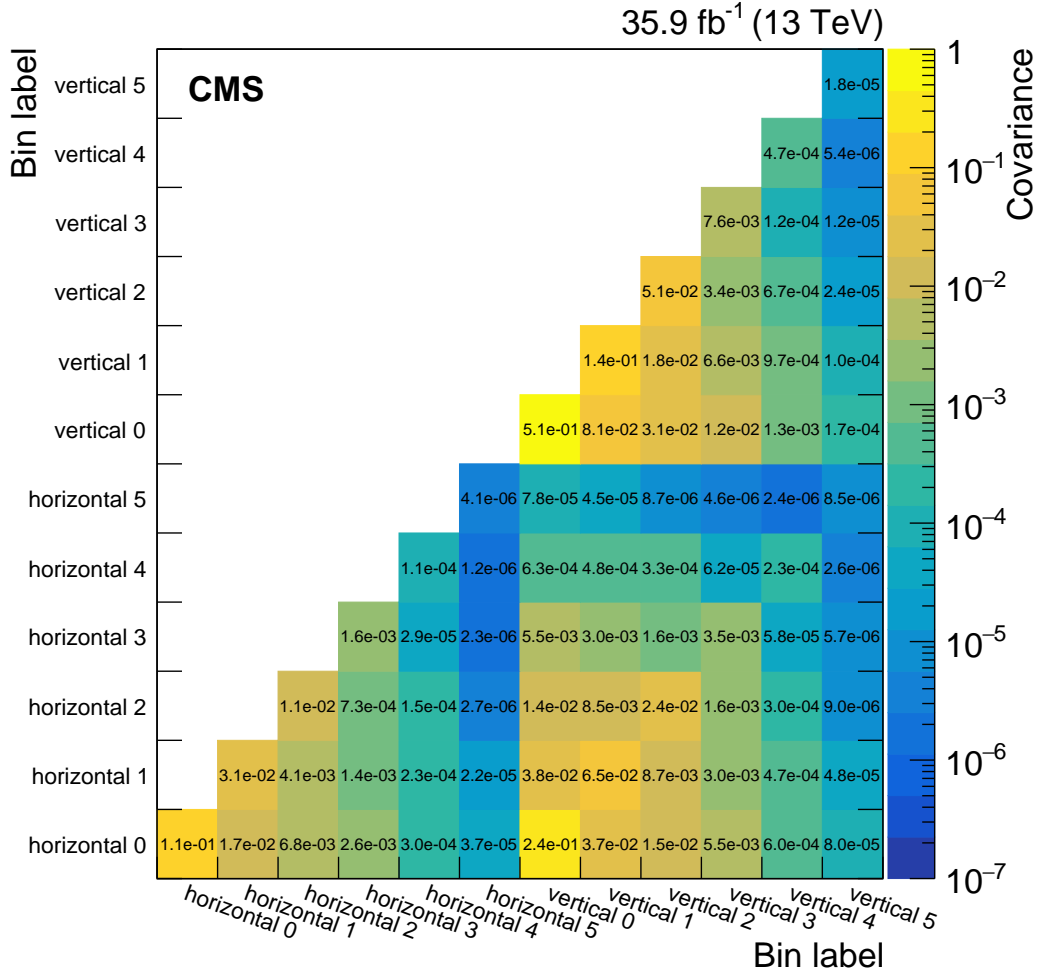


Figure 1-6: Covariances between the predicted background yields in all the E_T^γ bins of the horizontal and vertical signal regions. The bin labels specify which signal region the bin belongs to and what number bin it is for that region.

1.6.2 Limits

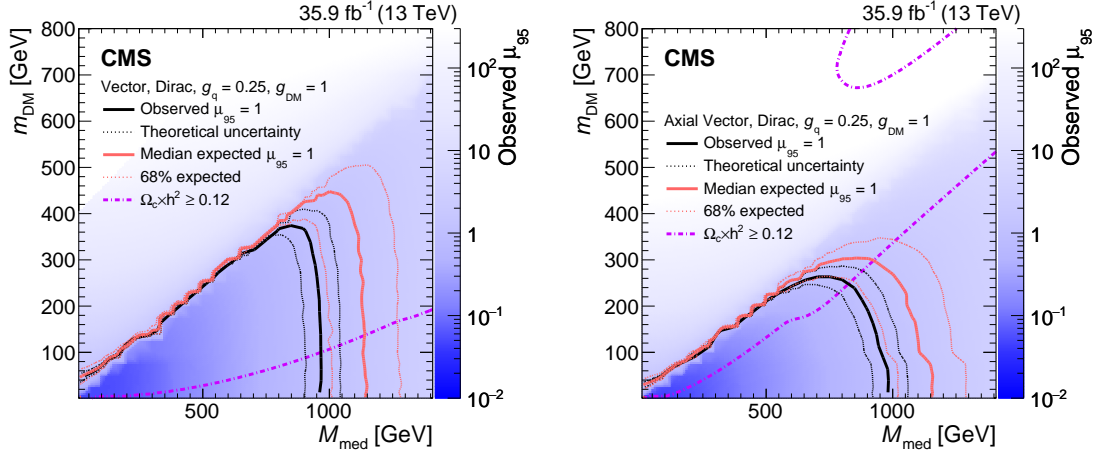


Figure 1-7: The ratio of 95% CL upper cross section limits to the theoretical cross section (μ_{95}), for DM simplified models with vector (left) and axial-vector (right) mediators, assuming $g_q = 0.25$ and $g_{\text{DM}} = 1$. Expected $\mu_{95} = 1$ contours are overlaid in red. The region under the observed contour is excluded. For DM simplified model parameters in the region below the lower violet dot-dash contour, and also above the corresponding upper contour in the right hand plot, cosmological DM abundance exceeds the density observed by the Planck satellite experiment.

Figure 5-7 shows the 95% CL upper cross section limits with respect to the corresponding theoretical cross section ($\mu_{95} = \sigma_{95\%}/\sigma_{\text{theory}}$) for the vector and axial-vector mediator scenarios, in the $M_{\text{med}}-m_{\text{DM}}$ plane. The solid black (dashed red) curves are the observed (expected) contours of $\mu_{95} = 1$. The σ_{theory} hypothesis is excluded at 95% CL or above in the region with $\mu_{95} < 1$. The uncertainty in the expected upper limit includes the experimental uncertainties. For the simplified DM LO models considered, mediator masses up to 950 GeV are excluded for values of m_{DM} less than 1 GeV.