

1 **Using Machine Learning Techniques for Data Quality**
2 **Monitoring at CMS Experiment**

3 by

4 Guillermo A. Fidalgo Rodríguez

5 A thesis presented for the degree of

6 BACHELLOR'S OF SCIENCE??

7 in

8 Physics

9 UNIVERSITY OF PUERTO RICO
10 MAYAGÜEZ CAMPUS

11 2018

12 Approved by:

13 _____
14 Sudhir Malik, Ph.D.

15 President, Graduate Committee
Date

16 _____
17 Héctor Méndez, Ph.D.

18 Member, Graduate Committee
Date

19 _____
20 Samuel Santana Colón, Ph.D.
21 Member, Graduate Committee

Date

22 _____
23 Rafael A. Ramos, Ph.D.
24 Chairperson of the Department

Date

²⁵ Abstract

²⁶ The Data Quality Monitoring (DQM) of CMS is a key asset to deliver high-quality
²⁷ data for physics analysis and it is used both in the online and offline environment. The cur-
²⁸ rent paradigm of the quality assessment is labor intensive and it is based on the scrutiny of
²⁹ a large number of histograms by detector experts comparing them with a reference. This
³⁰ project aims at applying recent progress in Machine Learning techniques to the automa-
³¹tion of the DQM scrutiny. In particular the use of convolutional neural networks to spot
³² problems in the acquired data is presented with particular attention to semi-supervised
³³ models (e.g. autoencoders) to define a classification strategy that doesn't assume previ-
³⁴ous knowledge of failure modes. Real data from the hadron calorimeter of CMS are used
³⁵ to demonstrate the effectiveness of the proposed approach.

³⁶ The Large Hadron Collider (LHC) is the world's largest and most powerful particle ac-
³⁷celerator designed to discover physics Beyond the Standard Model (BSM). The Standard
³⁸Model (SM) of fundamental particles and their interactions has been extremely successful
³⁹in describing phenomena in the atomic and subatomic realms in the last 100 years. After
⁴⁰the discovery of a boson with properties consistent with the SM Higgs boson at the LHC
⁴¹(by CMS and ATLAS Experiments in 2012), attention has shifted to searches for new
⁴²physics that can explain the fine-tuned cancellation of large quantum corrections required
⁴³to stabilize the Higgs boson mass at 125 GeV. Among BSM scenarios, Supersymmetry
⁴⁴(SUSY) is a new symmetry between fermions and bosons that provides an elegant mech-
⁴⁵anism to mitigate this hierarchy problem and overcome shortcomings of the Standard
⁴⁶Model. However, the data from the LHC has so far shown no signs of BSM, however,
⁴⁷stringent limits have been placed on its possible signatures. The future High Luminosity
⁴⁸run of the LHC (HL-LHC) from 2023-2035 will enable us to make precise measurements
⁴⁹of several of the SM parameters, in particular, the scalar sector and open the search for
⁵⁰new physics proposed by many BSM models and identify any new physics hint, including
⁵¹SUSY, that could also possibly be observed during the current run. This requires the CMS
⁵²collaboration to perform a series of upgrades that will ensure the capabilities of the CMS
⁵³detector to match the HL-LHC running conditions. One of these upgrades is to build an
⁵⁴entire new silicon Tracker that includes Pixel Detector and Outer Tracker. The work pre-
⁵⁵sented in this thesis is aligned with evolution of the CMS experiment. A simulation study
⁵⁶is performed to determine the effect of HL-LHC Pixel Detector on stubs (matching hit
⁵⁷pairs from the same particle track on two adjacent silicon layers of the modules in Outer
⁵⁸Tracker) which is key to future possible BSM discoveries and SUSY in particular. A
⁵⁹new method to measure the Z-invisible ($Z \rightarrow \nu\bar{\nu}$) background for the stop quark searches
⁶⁰in all-hadronic channel is presented. The method based on γ +jets provides an improve-
⁶¹ment over the method used for Z-invisible background in analysis presented here that
⁶²searches for SUSY in the all-hadronic channel with missing transverse momentum and a
⁶³customized top tagger. The data was collected by the CMS Detector from proton-proton

⁶⁴ collisions at the LHC at a center-of-mass energy of 13 TeV with an integrated luminosity
⁶⁵ of 35 fb^{-1} collected in 2016.

⁶⁶ Resumen

67 El Large Hadron Collider (LHC) constituye el acelerador de partículas más poderoso
68 y más grande del mundo, diseñado con el propósito de descubrir evidencia de fenómenos
69 físicos que trascienden el modelo estándar. El modelo estándar de partículas elemen-
70 tales, y las interacciones que describe, han sido extremadamente exitosos en explicar
71 fenómenos de la física al nivel atómico y sub-atómico durante el pasado siglo. Luego
72 del descubrimiento de un bosón con propiedades consistentes con el bosón de Higgs
73 (como se ha descrito por el modelo estándar), el enfoque se ha tornado hacia búsquedas de
74 fenómenos físicos que sean capaces de explicar la cancelación de correcciones cuánticas
75 grandes, necesarias para obtener un bosón de Higgs con una masa de 125 GeV. Dentro de
76 los modelos propuestos, la Supersimetría (SUSY) consiste de un tipo nuevo de simetría
77 entre fermiones y bosones que provee una solución elegante para este problema de jer-
78 arquía, entre otras deficiencias. Sin embargo, hasta el momento no se ha logrado observar
79 ninguna evidencia (de data obtenida del LHC) de fenómenos que van más allá del mod-
80 elo estándar, aunque si se han impuesto límites sobre las posibles señales de búsqueda.
81 La implementación del LHC de alta luminosidad (HL-LHC, por sus siglas en inglés) en-
82 tre 2023-2035, va a proveer la posibilidad de realizar medidas precisas de varios de los
83 parámetros del modelo estándar facilitando la búsqueda de varios modelos propuestos, in-
84 cluyendo SUSY. Esto está sujeto a que la colaboración de CMS actualice el detector para
85 que pueda funcionar bajo las nuevas condiciones de funcionamiento del LHC. Una de es-
86 tas actualizaciones incluye la construcción total de un nuevo sistema de rastreo de silicio,
87 el cual incluye un detector de pixeles y un rastreador externo. El trabajo presentado en
88 esta tesis está alineado con la evolución del detector CMS. Se realizó un estudio basado
89 en simulación con el propósito de determinar el efecto que va tener el nuevo detector de
90 pixeles sobre la formación de “stubs” (definido como la coincidencia de un par de señales
91 en capas de silicio adyacentes en los componentes del rastreador externo), los cuales for-
92 man la base para el descubrimiento de física nueva, en particular SUSY. Se presenta un
93 método nuevo para estimar las interacciones de fondo relacionadas el decaimiento de Z
94 en un par de neutrinos ($Z \rightarrow \nu\bar{\nu}$) para búsquedas del top squark en estados finales con cero
95 leptones. El método está basado en la realización que una muestra de eventos de $\gamma+jets$
96 debe exhibir mejores resultados que el método utilizado previamente para búsquedas de
97 SUSY en estados finales de cero leptones, con momentum transverso desaparecido y un
98 identificador de top-quarks. La data utilizada para el estudio fue obtenida de colisiones
99 de protones generadas por el LHC en el detector CMS con un centro de masa de 13 TeV
100 y una luminosidad integrada de 35 fb^{-1} durante el 2016.

¹⁰¹ **Acknowledgements**

¹⁰² I wish to thank United States State Department and University of Michigan for pro-
¹⁰³ viding the opportunity to work abroad at CERN during the 2018 Winter Semester. I also
¹⁰⁴ wish to thank CERN staff, CMS Experiment , Texas Tech University, and the University
¹⁰⁵ of Puerto Rico at Mayagüez, with special thanks to Dr. Federico de Guio for his local
¹⁰⁶ mentorship and Dr. Nural Akchurin, and Dr. Sudhir Malik for their guidance. A very
¹⁰⁷ special thanks to Dr. Jean Krisch for accepting me for this great research opportunity and
¹⁰⁸ Dr. Steven Goldfarb for being a wonderful host and overall local guidance at CERN.

¹⁰⁹ List of Figures

110	Figure 2.1:	The CERN Accelerator Complex [?].	4
111	Figure 2.2:	The CMS Detector Layout [?].	5
112	Figure 2.3:	Overview of the CMS Tracker Layout [?].	6
113	Figure 2.4:	Geometrical layout of the CMS Electromagnetic Calorimeter [?].	8
114	Figure 2.5:	Layout of the CMS ECAL illustrating its various components [?].	9
115	Figure 2.6:	Geometrical layout of the HCAL showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [?].	10
116	Figure 2.7:	Quarter-view of the CMS muon system (left) and a muon in the transverse plane leaves a curved trajectory across the four muon detector stations (right).	11
117	Figure 2.8:	Transverse slice of the CMS detector, showing the individual detector subsystems and particle signatures in each. The particle type can be inferred by combining the detector response in the different subdetectors [?].	13
118	Figure 2.9:	Phase-2 Pixel Detector Layout[ref].	14
119	Figure 3.1:	Schematic representation of the Standard Model particles.	16
120	Figure 3.2:	Schematic representation of the Standard Model particles with the added SUSY particles.	18
121	Figure 4.1:	Sketch of one quarter of the tracker layout in r-z view. In the Inner Tracker the green lines correspond to pixel modules made of two readout chips and the yellow lines to pixel modules with four readout chips. In the Outer Tracker the blue and red lines represent the two types of modules p_T modules - (2S and PS).	22
122	Figure 4.2:	Sketch of one quarter of the Outer Tracker in r-z view. Blue (red) lines represent PS (2S) modules. The three sub-detectors, named TBPS, TB2S, and TEDD, are indicated. All overlapping layers are shown separately.	22
123	Figure 4.3:	Sketch of four dees (left) forming a double-disc (centre), and drawing of a part of a TEDD double-disc (right), illustrating the overlap of modules in ϕ and z . The upper two dee support structures are removed in order to show all layers of modules. The modules ob the disc are arranged in 15 or 12 rings.	23
124	Figure 4.4:	Correlation of signals in closely-spaced sensors enables rejection of low- p_T particles; the channels shown in green represent the selection window to define an accepted stub.	24
125			

146	Figure 4.5:	Number of stubs per event for standard pixel geometry ($8s4l$)	25
147	Figure 4.6:	Comparison of total number of stubs per event between the standard layout and the other three considered. $8s4l$ vs $8s3l$ (left), $8s4l$ vs $7s4l$ (middle) and $8s4l$ vs $6s3l$ (right).	26
150	Figure 4.7:	Comparison for number of stubs per event between pixel geometry $8s4l$ with a dead disc (blue) and both pixel geometry $8s4l$ (right plot) and $7s4l$ (left plot) (blue). Both the standard geometry and $7s4l$ are shown in blue in their respective plots.	26
154	Figure 4.8:	Comparison of total number of stubs per η value between the standard layout and the other three layouts considered. All four different layouts exhibit the same distribution. $8s4l$ vs $8s3l$ (left), $8s4l$ vs $7s4l$ (middle) and $8s4l$ vs $6s3l$ (right).	27
158	Figure 4.9:	Comparison of total number of stubs per each of the 6 barrel layers. $8s4l$ vs $8s3l$ (left), $8s4l$ vs $7s4l$ (middle) and $8s4l$ vs $6s3l$ (right). Tracker Barrel layers shown below.	27
161	Figure 4.10:	Comparison of total number of stubs per each of the 5 Endcap Double Discs. $8s4l$ vs $8s3l$ (left), $8s4l$ vs $7s4l$ (middle) and $8s4l$ vs $6s3l$ (right). Tracker Endcap Discs shown below.	28
164	Figure 4.11:	Comparison of total number of stubs for all rings across the five endcap double-discs. $8s4l$ vs $8s3l$ (left), $8s4l$ vs $7s4l$ (middle) and $8s4l$ vs $6s3l$ (right). The double discs are shown below.	28
167	Figure 5.1:	Simplified model diagrams show of the various Supersymmetry signals considered in this analysis: the T2tt model (top left), the T1tttt model (top right), the T1ttbb model (middle left), the T5tttt (middle right), and the T5ttcc model (bottom).	31
171	Figure 5.2:	Efficiency of the top quark tagger as a function of generator-level top quark p_T for the monojet (red boxes), dijet (magenta triangles), and trijet (green upside-down triangles) categories and for their combination (blue circles), as determined using T2tt signal events with a top squark mass of 850 GeV and an LSP mass of 100 GeV. The vertical bars indicate the statistical uncertainties.	33
177	Figure 5.3:	Search region definitions in the kinematic variables.	35
178	Figure 5.4:	Comparison between the total SM background from simulation and CMS data for N_t (top left), N_b (top right), m_{T2} (bottom left) and p_T^{miss} (bottom right). Total SM backgrounds and signals are scaled to same data yield for a shape comparison.	36
182	Figure 5.5:	Distribution of p_T^{miss} for both the single-electron (left) and single-muon (right) SB data samples compared to predictions for SM processes. The hatched bands indicate the statistical uncertainties in the total SM prediction. Note that the data and the predictions for all backgrounds except that for $t\bar{t}$, single top quark, and W+jets events are identical between the left and right plots.	37

188	Figure 5.6:	The p_T^{miss} (left) and N_b (right) distributions of data and simulation in the loose dimuon CS after applying both the normalization and shape correction factors. The lower panels show the ratio between data and simulation. Error bars in the plot account only for statistical uncertainties. The values in parentheses indicate the integrated yields for each component.	38
194	Figure 5.7:	Observed events compared to the corrected SM background predictions for all 84 search regions in the full data of 35.9 fb^{-1} collected in 2016. The lower panel shows the ratio between the data and the total SM simulation. The gray bands show the total uncertainty related to the background prediction.	39
199	Figure 5.8:	The 95% CL upper limit on the production cross section of the T2tt simplified model as a function of the top squark and LSP masses. The solid black curves represent the observed exclusion contour with respect to NLO+NLL [?] signal cross sections and the change in this contour due to variation of these cross sections within their theoretical uncertainties. The dashed red curves indicate the mean expected exclusion contour and the region containing 68% of the distribution of expected exclusion limits under the background-only hypothesis.	40
208	Figure 5.9:	The 95% CL upper limit on the production cross section of the T1tttt (upper left), T1ttbb (upper right), T5tttt (bottom left), and T5ttcc (bottom right) simplified models as a function of the gluino and LSP masses.	41
212	Figure 6.1:	Leading-order Feynman diagrams for Z+ jets and $\gamma+$ jets processes.	43
213	Figure 6.2:	Shown are both the p_T^γ (left) and $p_T^{\gamma(miss)}$ (right) distributions before applying any corrections. $p_T^{\gamma(miss)}$ is obtained by adding the p_T^γ to the total p_T^{miss} in every event.	45
216	Figure 6.3:	Plots for Fake Rate (left) and Purity (right) as a function of the photon p_T are shown. The events are selected are required to have a $p_T > 200 \text{ GeV}$, be within the ECAL acceptance range, and pass the loose ID selection cuts. This selection was produced in order to verify the values given by the E/ γ POG. As can be seen, the efficiency (purity) is seen to agree with the values of the loose photon ID/isolation selection.	46
223	Figure 6.4:	Plots for Fake Rate (left) and Purity (right) as a function of the photon p_T are shown. These plots include photons with the full control region selection. Aside from exhibiting lower statistics, the plots seem to agree with the fake rate and purity before all the control region cuts are applied.	46
228	Figure 6.5:	Results of study of the Z+jets to $\gamma+$ jets cross-section ratio for both data and MadGraph simulation.	49
230	Figure 6.6:	Jet multiplicity and the associated S_γ scale factor in the loose photon control region before any corrections are applied.	50
232	Figure 6.7:	N_{jet} distribution in the tight $\mu\mu$ control region after S_γ corrections.	50

233	Figure 6.8: p_T^γ (left) and $p_T^{\gamma(miss)}$ (right) distributions after applying the $S_\gamma(N_j)$ scale factor. Comparing to Figure 6.2, an improvement in the agreement between data/MC can be observed.	51
234		
235		
236	Figure 6.9: N_b distribution before (left) and after (right) applying the $S_\gamma(N_j)$ scale factor.	51
237		
238	Figure 6.10: H_T and m_{T2} distributions applying the $S_\gamma(N_j)$ scale factor.	52
239	Figure 6.11: Shown are data/MC comparisons for the p_T^{miss} (left) and H_T (right) distributions after applying both the N_j -dependent shape corrections (S_γ) and the global normalization scale factor (R_{norm}).	52
240		
241		
242	Figure 6.12: Shown are data/MC comparisons for the N_j (left) and N_b (right) distributions after applying both the N_j -dependent shape corrections (S_γ) and the global normalization scale factor (R_{norm}).	53
243		
244		
245	Figure 6.13: Systematic uncertainty in the final prediction, as a function of the search bin, associated to the MC statistics.	54
246		
247	Figure 6.14: $Z \rightarrow \nu\bar{\nu}$ background prediction for all search bins, including the breakdown of the various uncertainties.	55
248		

Contents

250	List of Figures	iv
251	1 Introduction	1
252	2 Large Hadron Collider and the Compact Muon Solenoid	3
253	2.1 The LHC Machine	3
254	2.2 The CMS Detector	4
255	2.2.1 Silicon Tracking System	6
256	2.2.2 Electromagnetic Calorimeter	7
257	2.2.3 Hadron Calorimeter	8
258	2.2.4 Magnet	10
259	2.2.5 Muon Detector	10
260	2.2.6 Trigger and Data Acquisition	11
261	2.2.7 Event Reconstruction	12
262	2.2.8 Future Upgrade of Pixel Detector	13
263	3 Theoretical Motivation	16
264	3.1 The Standard Model	16
265	3.2 Indications of Physics Beyond the Standard Model	17
266	3.3 Supersymmetric extension of the Standard Model	18
267	3.3.1 Simplified Models Approach to Supersymmetry	19
268	4 Optimizing the Phase-2 Pixel Geometry for the Level-1 Trigger in the Outer Tracker	21
269	4.1 Upgraded Phase-2 Tracker	21
270	4.1.1 Phase-2 Tracker Layout	22
271	4.1.2 Tracker Input to the L1-Trigger	22
272	4.2 Tracker Stub Simulation	23
273	4.2.1 Results of Effect of the Phase 2 Pixel Detector on the L1 Stub Rate	25
274	4.3 Results	29
275	5 Search for SUSY in the 0-Lepton Final State	30
276	5.1 Introduction	30
277	5.2 Analysis Description	30
278	5.2.1 Signal Models	31
279	5.2.2 Top Tagger	32
280	5.2.3 Lepton and Track Veto	33
281	5.2.4 Baseline Event Selection	34

283	5.2.5	Search Bin Definition	34
284	5.3	Background Estimation	35
285	5.3.1	Background from $t\bar{t}$, Single Top Quark and W+jets Events	35
286	5.3.2	The $Z \rightarrow \nu\bar{\nu}$ Background	37
287	5.3.3	The QCD Multijet Background	38
288	5.3.4	Background From Rare and Other Processes	39
289	5.4	Results and Interpretation	39
290	6	Estimation of the $Z \rightarrow \nu\bar{\nu}$ Background	42
291	6.1	Introduction	42
292	6.2	The Irreducible $Z \rightarrow \nu\bar{\nu}$ Background	42
293	6.3	The Loose γ +jets Control Region	43
294	6.3.1	Photon ID and Isolation	43
295	6.3.2	Photon Selection	44
296	6.3.3	Photon Purity and Fake Rate	45
297	6.4	The $Z \rightarrow \mu^+\mu^-$ Control Region	47
298	6.4.1	Muon ID and Isolation	48
299	6.4.2	Muon Selection in the Tight Control Region	48
300	6.5	Analysis	49
301	6.5.1	Shape Correction Using the γ + jets Control Sample	49
302	6.5.2	Normalization Correction Using the tight $Z \rightarrow \mu^+\mu^-$ Control Sample	51
303	6.6	Results	53
304	6.6.1	Systematics	53
305	6.6.2	$Z \rightarrow \nu\bar{\nu}$ Estimation for the Search Bins	54
306	7	Conclusion	56
307	7.1	L1 Stub Counting for the HL-LHC CMS Tracker Upgrade	56
308	7.2	Search for SUSY in the All-Hadronic Channel	57
309	7.3	Estimation of the $Z \rightarrow \nu\bar{\nu}$ + jets Background with a Hybrid Method	57

310

Chapter 1

311

Introduction

312 The recently discovered Higgs boson at the LHC could be the final particle required
313 in the highly successful theory of the Standard Model. However, the SM is not with-
314 out its shortcomings, for instance one requires fine-tuned cancellations of large quantum
315 corrections for the Higgs boson to have a mass at the electroweak symmetry breaking
316 scale known as the hierarchy problem. The magnitude of this fine-tuning makes one sus-
317 pect that there is some dynamical mechanism which makes this fine-tuning “natural”.
318 SUSY, a popular extension of the SM offers a well-motivated explanation: the dominant
319 role in cancelling the quantum effects would come from scalar partners to the bottom/top
320 quarks (sbottoms/stops), due to the large 3rd generation Yukawa couplings in the SM,
321 and fermionic partners to the Higgs boson (higgsinos). The lightest SUSY particle, the
322 neutralino, could also be a candidate for Dark Matter. Alternatively, the naturalness of
323 electroweak symmetry breaking could be effectively explained by non-SUSY BSM mod-
324 els involving strong dynamics. Searching for sparticles in the all-hadronic mode is very
325 attractive because it makes up a large portion of branching fractions for typical signals.
326 The final state involves events with high jet multiplicity and missing transverse energy
327 and this search method could eventually decipher the mysteries surrounding the Higgs
328 boson.

329
330 The work presented in this thesis is about discovering physics beyond the Standard
331 Model and is three fold: (1) A simulation study is performed to determine the effect of
332 CMS Phase-2 Pixel Detector on stubs (matching hit pairs from the same particle track on
333 two adjacent layers of silicon modules) (2) An analysis involving a search for Supersym-
334 metry in the all-hadronic channel with missing transverse momentum and a customized
335 top tagger (3) An improved new method based on $\gamma + \text{jets}$ to determine $Z \rightarrow \nu\bar{\nu}$ background
336 in search for the SUSY signature.

337
338 This thesis work was carried out at the LPC (LHC Physics Center) at Fermilab. The
339 funding for this work was provided by the **NSF (National Science Foundation - NSF**
340 **Award ID 1506168) and LPC (Guest and Visitor) Fellowship.**

341
342 Chapter 2 presents a basic description of the Large Hadron Collider and CMS Detec-
343 tor (including trigger, data acquisition and Phase-2 upgrade plans), which are the basic
344 tools used to collect data in order to successfully complete the thesis work.

345

346 Chapter 3 gives a brief motivation to look for extensions for physics Beyond the Stan-
347 dard Model for which Supersymmetry (SUSY) is the most popular choice. It describes
348 SUSY briefly and in relation to Simplified Models few of which are used in the analysis
349 described in the thesis.

350
351 Chapter 4 is dedicated to a study optimizing the geometry of Phase-2 Pixel Detector
352 for the High Luminosity-LHC. A simulation is performed using the $t\bar{t}$ sample to estimate
353 the impacting of pixel detector on the formation of stubs used in the outer tracker to de-
354 termine particle tracks that are used in the Level-1 (L1) trigger. This work was done as
355 part of service work for the CMS experiment.

356
357 Chapter 5 presents an analysis on the search for SUSY in the 0-Lepton final state
358 where an inclusive search for events with final states that contain missing transverse mo-
359 mentum and reconstructed top quarks is performed. The signal models used in this study
360 include the production of three different types of SUSY particles. Two of which are the
361 top squark and the gluino, the supersymmetric partners of the SM top and gluon, respec-
362 tively. The third one is the neutralino that is considered to be the lightest SUSY particle
363 (LSP) under the Minimal Supersymmetric Standard Model (MSSM) and a possible can-
364 didate for Dark Matter. Background from $t\bar{t}$, single top quark, W+jets and $Z \rightarrow \nu\bar{\nu}$ events
365 was estimated.

366
367 Chapter 6 details an improved new method based on γ +jets to determine $Z \rightarrow \nu\bar{\nu}$
368 background in search for the SUSY signature. This estimation procedure builds upon the
369 $Z \rightarrow \nu\bar{\nu}$ method used in SUSY analysis described in Chapter 5 (also referred to as 2016
370 analysis) and aims to refine the overall background calculation as well as reduce the un-
371 certainties associated with the previous method. To accomplish this, an additional γ +jets
372 control sample (CS) is used in conjunction with the tight $Z \rightarrow \mu^+\mu^-$ control region used in
373 the 2016 analysis estimation. The new γ +jets CS provides a more data-driven estimation
374 procedure with the added benefit of a substantially larger production cross-section than
375 Z +jets process used before.

376
377 The conclusions and results of each chapter are presented in the corresponding chap-
378 ter.

379
380 This thesis work has been presented at several internal meetings of the CMS Experi-
381 ment and at the following international meetings and conferences:

- 382 1. **Andrés Abreu** gave a talk “*Estimation of the Z Invisible Background for Searches*
383 *for Supersymmetry in the All-Hadronic Channel*” at “APS April 2018: American
384 Physical Society April Meeting 2018, 14-17 Apr 2018”, Columbus, OH
- 385 2. **Andrés Abreu** gave a talk “*Phase-2 Pixel upgrade simulations*” at the “USLUA
386 Annual meeting: 2017 US LHC Users Association Meeting, 1-3 Nov 2017”, Fer-
387 milab, Batavia, IL
- 388 3. **Andrés Abreu** gave a talk, “*Direct production of top squark pairs in all-hadronic*
389 *channel*” presentation at “FNAL50: Fermilab 50th Anniversary Symposium and
390 Users Meeting and New Perspectives Workshop, Fermilab, Batavia, June 5-8, 2017”,
391 Batavia, IL

³⁹² Chapter 2

³⁹³ Large Hadron Collider and the ³⁹⁴ Compact Muon Solenoid

³⁹⁵ 2.1 The LHC Machine

³⁹⁶ The Large Hadron Collider (LHC) [?], located at the European Organization for Nu-
³⁹⁷ clear Research (CERN) complex in Geneva Switzerland, is the world's largest and most
³⁹⁸ powerful particle collider as well as the most complex experimental device ever assem-
³⁹⁹ bled. The main motivation for its construction was uncovering the nature behind elec-
⁴⁰⁰ troweak symmetry breaking due to the Higgs mechanism as well as to search for new
⁴⁰¹ physical phenomena Beyond the Standard Model (BSM) [?]. Since its completion in
⁴⁰² 2008, the LHC has helped researchers obtain a plethora of significant results in the field
⁴⁰³ of particle physics and it successfully achieved its purpose with the discovery of the Higgs
⁴⁰⁴ Boson in 2012.

⁴⁰⁵
⁴⁰⁶ The LHC consists of several superconducting magnets and accelerating structures ar-
⁴⁰⁷ ranged along a 27-kilometer circumference ring which serve to boost and bend the proton
⁴⁰⁸ (or other heavy ion) beams along its path. In order for the particle beams to maintain the
⁴⁰⁹ orbit while they are accelerated, a magnetic field of 8.3 T is produced by the supercon-
⁴¹⁰ ducting magnets along the circular tunnel, 100 meters below the ground. The magnets are
⁴¹¹ maintained at a temperature of 1.9 K by using liquid helium. The two beams are made
⁴¹² to travel in opposite directions and in separate beam pipes which are kept at ultrahigh
⁴¹³ vacuum and are accelerated to near the speed of light. They are then collided head-on at
⁴¹⁴ any of the four interaction points located around the LHC ring. Approximately 1.5×10^{11}
⁴¹⁵ protons are generated every 25 ns with an energy of 26 GeV at the Proton Synchrotron
⁴¹⁶ (PS). The proton beam is then accelerated to an energy of 450 GeV at the Super Proton
⁴¹⁷ Synchrotron (SPS) [?] before being delivered to the LHC. At design level, each bunch is
⁴¹⁸ accelerated to an energy of 7 TeV and are made to collide at a centre-of-mass energy of
⁴¹⁹ up to 14 TeV with a frequency of 40 MHz. The current proton beams achieve an instan-
⁴²⁰ taneous luminosity on the order of $10^{32} \text{ cm}^{-2} \cdot \text{s}^{-1}$ and a total integrated luminosity of 37
⁴²¹ fb^{-1} was reached during 2016 [?].

⁴²²
⁴²³ The four main detectors comprising the LHC machine are CMS, ATLAS [?], LHCb
⁴²⁴ [?] and ALICE [?]. Both CMS and ATLAS are general purpose detectors whose initial
⁴²⁵ designs had the detection of the SM Higgs boson, with its wide range of decay modes,

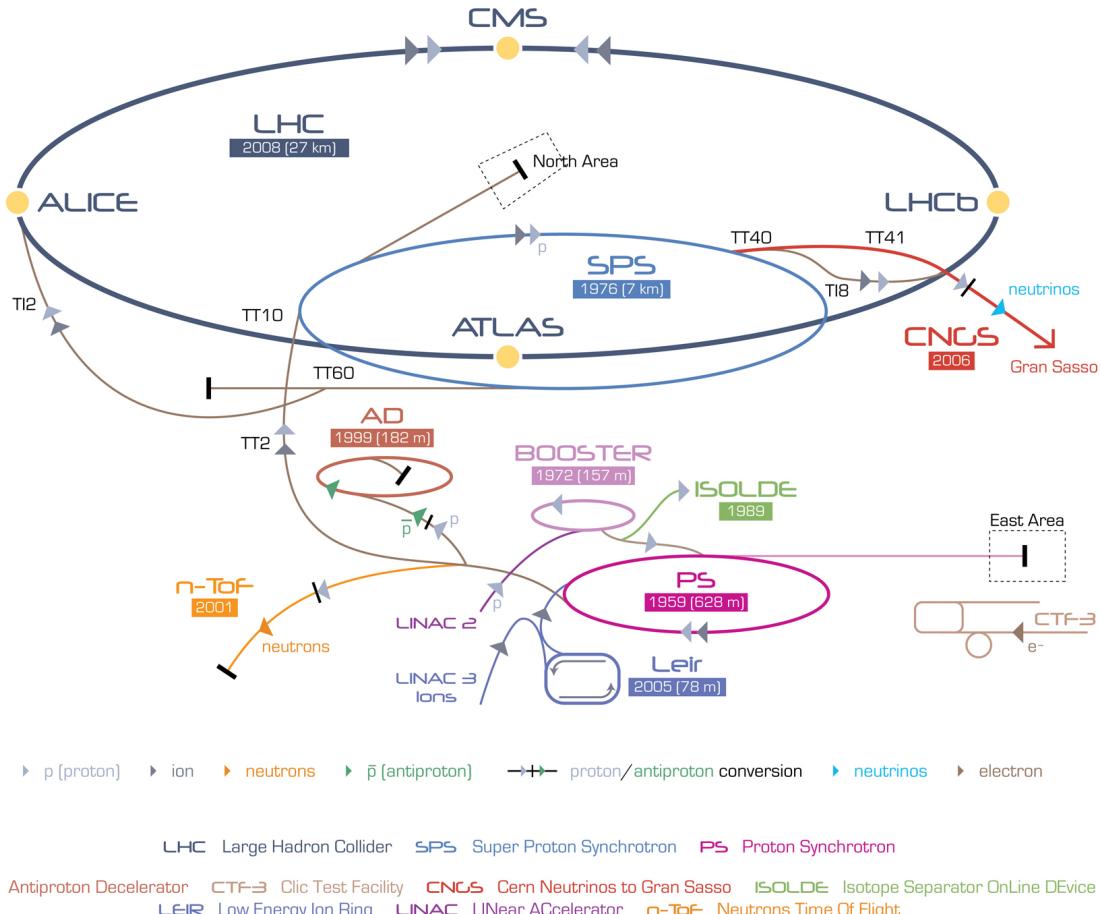


Figure 2.1: The CERN Accelerator Complex [?].

in mind. Both detectors managed to accomplish this goal when a 126 GeV scalar boson consistent with the SM Higgs was independently verified by both experiments in July of 2012. Furthermore, the designs for CMS and ATLAS allow for the search of many other additional phenomena in BSM physics such as Supersymmetry, Dark Matter [?], Dark Sector [?], etc. On the other hand, the LHCb and ALICE detectors focus on more particular kinds of searches. The main motivation for the LHCb experiment, where the b stands for beauty, concerns itself with the measurement of CP violation parameters in b-hadron interactions and studies cover a wide range of aspects of Heavy Flavor Electroweak and QCD physics. Meanwhile, the ALICE experiment focuses on the study of heavy ion ($Pb-Pb$) nuclei collisions at a centre-of-mass energy of 2.76 TeV in order to better understand the physics behind strongly interacting matter at extreme energy densities.

2.2 The CMS Detector

The Compact Muon Solenoid (CMS) detector is a general purpose particle detector designed to investigate various physical phenomena concerning the SM and beyond it, such as Supersymmetry, Extra Dimensions and Dark Matter. As its name implies, the detector is a solenoid which is constructed around a superconducting magnet capable of producing a magnetic field of 3.8 T. The magnetic coil is 13m long with an inner diameter

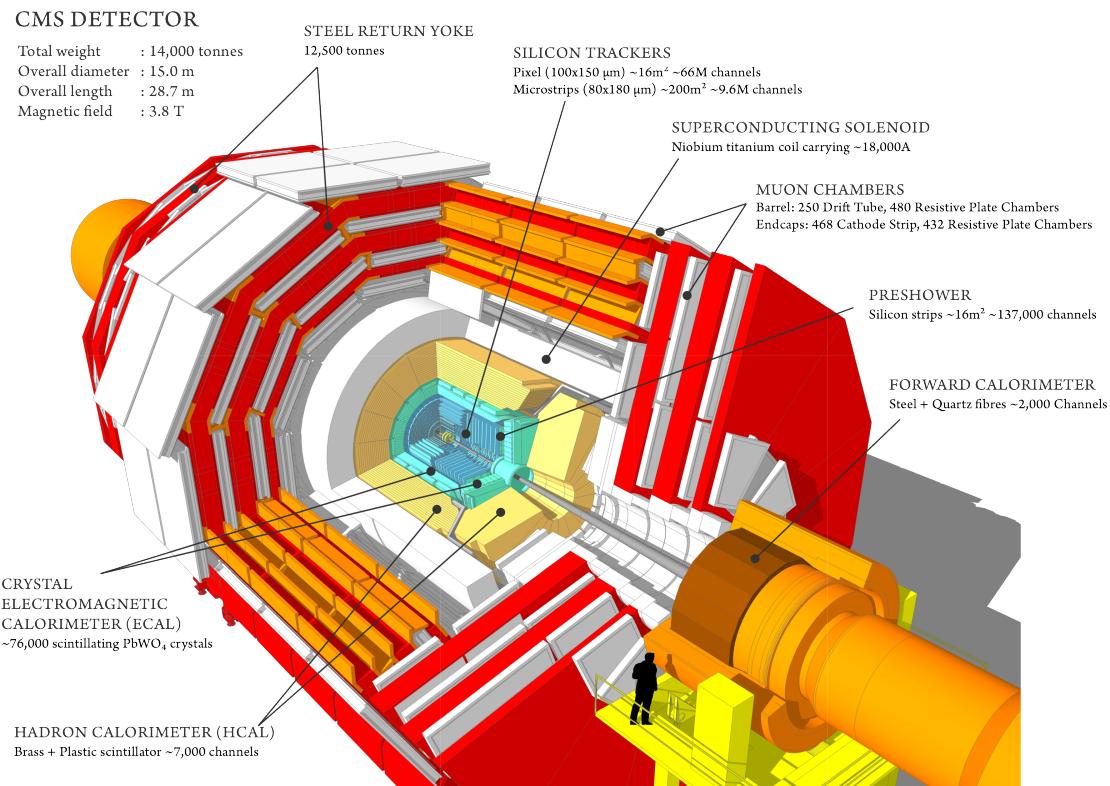


Figure 2.2: The CMS Detector Layout [?].

of 6m, making it the largest superconducting magnet ever constructed. The CMS detector itself is 21m long with a diameter of 15m and it has a weight of approximately 14,000 tonnes. The CMS experiment is one of the largest scientific collaborations in the history of mankind with over 4,000 participants from 42 countries and 182 institutions.

In order to meet the many needs of the SM and BSM searches, and the goals of the LHC physics program, the CMS detector was designed with the following features:

- A magnet with large bending power and high performance muon detector for good muon identification and momentum resolution over a wide range of momenta and angles.
- An inner tracking system capable of high reconstruction efficiency and momentum resolution requiring pixel detectors close to the interaction region.
- An electromagnetic calorimeter able to provide good electromagnetic energy resolution and a high isolation efficiency for photons and leptons.
- A hadron calorimeter capable of providing precise missing-transverse-energy (p_T^{miss}) and dijet-mass resolution.

A general layout of the CMS detector and all its constituent sub-detectors can be seen in Figure 2.2. The configuration of the CMS sub-detectors follow a cylindrical layer pattern that is symmetrical about the interaction region and consists of a central barrel with endcaps on both ends.

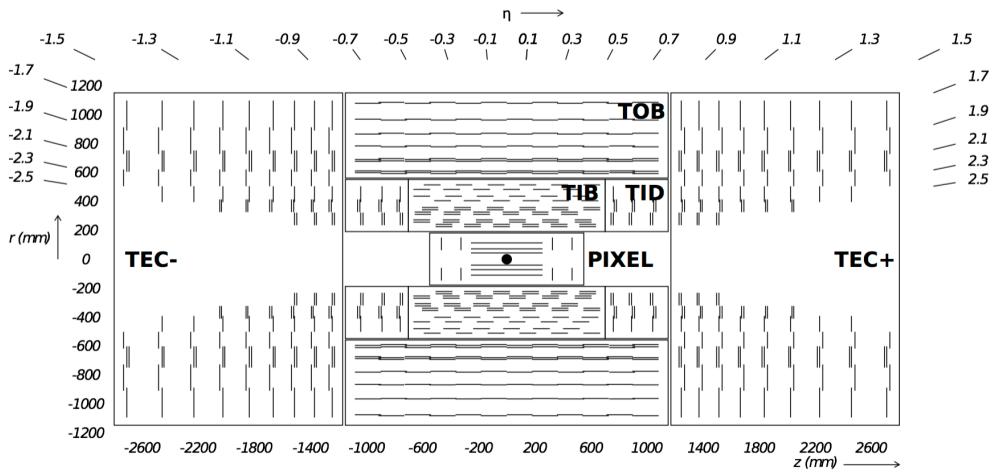


Figure 2.3: Overview of the CMS Tracker Layout [?].

468 The coordinate system for the CMS detector design uses a right-hand rule convention
 469 centered around the ideal interaction point to describe the positions of objects in the experiment.
 470 The z-axis is defined along the direction of the LHC beam, with the x-axis
 471 pointing towards the center of the LHC ring. In terms of polar coordinates then, r is the
 472 radial distance from the center of the pipe, the polar angle θ is measured against the z-axis
 473 and the azimuthal angle ϕ is measured with respect to the x-axis. However, the pseudo-
 474 rapidity η is generally preferred over the polar angle θ . The pseudorapidity is defined
 475 as:

$$476 \quad \eta = -\ln \tan \frac{\theta}{2}.$$

477 2.2.1 Silicon Tracking System

478 The CMS tracking system was designed with the goal of obtaining precise and effi-
 479 cient measurements for the trajectories of charged particles resulting from proton-proton
 480 collisions at the LHC. In addition, it allows for the precise measurement of secondary ver-
 481 tices and impact parameters needed to efficiently identify the heavy flavours produced in
 482 many interesting physics channels. Due to the LHC's design Luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$,
 483 the currently installed CMS phase-1 tracker is expected to handle an average of 1000 par-
 484 ticles from over 20 overlapping proton-proton interactions per bunch crossing, every 25
 485 ns. This required a detector technology capable of achieving a high granularity and fast
 486 response as well as being tolerant to the radiation produced from the intense particle flux.
 487 These considerations lead to a tracker design composed entirely of silicon detector tech-
 488 nology which features an active silicon area of about 200 m^2 , making it the largest silicon
 489 tracker ever built [?].

490
 491 The CMS tracker is built in a cylindrical manner around the interaction point and has
 492 a diameter of 2.5 m and a length of 5.8 m. It is comprised of a pixel detector with three
 493 barrel layers, positioned at a distance between 4.4 cm and 10.2 cm from the interaction
 494 region, and a silicon strip tracker with 10 barrel detection layers extending to a radius of
 495 1.1 m. Each system is made complete by endcaps at opposite sides of the barrel, consist-
 496 ing of 2 discs for the pixel detector and 3 plus 9 discs for the strip tracker, extending the

497 acceptance region of the tracker up to a pseudorapidity $|\eta| < 2.5$.

498

499 The pixel detector is the part of the CMS tracking system closest to the interaction
500 region and consists of 3 barrel layers (BPix) and 2 endcap discs (FPix). It is responsible
501 for providing precise tracking points in $r\phi$ and z , a feature that is required for the small
502 impact parameter resolution, needed for good secondary vertex reconstruction. The de-
503 tector contains 1440 modules covering an area of approximately 1 m^2 making up a total of
504 66 million pixels. The sensors were designed using an n-on-n concept from $320 \mu\text{m}$
505 thick silicon wafers and are fabricated with read-out chips (ROCs) that are bump-bonded
506 to the sensor in standard $0.25 \mu\text{m}$ CMOS technology. Each of the pixels has a pitch size
507 of $100 \times 150 \mu\text{m}^2$, which corresponds to an occupancy of about 10^{-4} per bunch crossing.

508

509 The silicon strip tracker is built surrounding the pixel detector and consists of three
510 large subsystems. The Tracker Inner Barrel and Disks (TIB/TID) extend to a radius of
511 55 cm and are composed of four barrel layers, completed by three disks at each end. The
512 TIB/TID employs the use of $320 \mu\text{m}$ thick silicon micro-strip sensors in order to deliver
513 up to 4 $r\phi$ measurements on a trajectory. Surrounding the TIB/TID is the Tracker Outer
514 Barrel (TOB), which consists of 6 barrel layers and has an outer radius of 116 cm. The
515 TOB extends symmetrically in z between $\pm 118 \text{ cm}$ and provides an additional 6 $r\phi$
516 measurements for a trajectory. Beyond the TOB's z range lie the Tracker Endcaps (TEC+
517 and TEC-, where the sign indicates their position respect to z). Each TEC consists of 9
518 discs with up to 7 rings of silicon micro-strip detectors, providing up to 9 additional ϕ
519 measurements per trajectory. The CMS silicon strip tracker has a total active silicon area
520 of 198 m^2 and is composed of 15,148 sensor modules.

521 2.2.2 Electromagnetic Calorimeter

522 The CMS Electromagnetic Calorimeter (ECAL) is a hermetic homogeneous calorime-
523 ter whose function is to measure the energy of particles that interact via the electromag-
524 netic force. With the use of 75,848 scintillator crystals, it is capable of providing good
525 energy resolution within the requirements of the ambitious LHC program. In particular,
526 the ECAL's design was optimized to search for diphoton events resulting from Higgs bo-
527 son decays ($H \rightarrow \gamma\gamma$).

528

529 The ECAL is composed of two main sub-systems – the barrel calorimeter (EB) and the
530 endcap calorimeter (EE) – and is completed by a preshower calorimeter (ES), as shown
531 in [Figure 2.5](#). It covers a solid angle up to a pseudorapidity of $|\eta| < 3$, with the EB ex-
532 tending in the range of $|\eta| < 1.479$ and the EE covering a range of $1.479 < |\eta| < 3$. Both
533 subsystems are composed of lead tungstate (PbWO_4) scintillator crystals which provide
534 a fast response time and high radiation tolerance, both crucial requirements for optimal
535 performance at LHC operating conditions. In addition, the properties of the PbWO_4 crys-
536 tals (high density, short radiation length and small Moli  re radius) led to the design of a compact
537 calorimeter with fine granularity.

538

539 The barrel part of the ECAL (EB) is composed of specially designed avalanche pho-
540 todiodes (APD). It consists of 61,200 crystals, forming a total volume of 8.14 m^3 and
541 weighing about 67.4 t. The crystals that form the EB are 230 mm long with a cross-

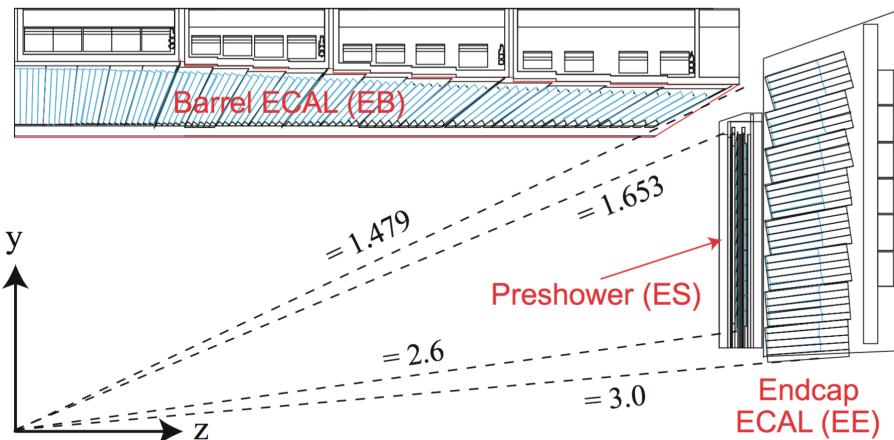


Figure 2.4: Geometrical layout of the CMS Electromagnetic Calorimeter [?].

542 section of $22 \times 22 \text{ mm}^2$ at the front face and $26 \times 26 \text{ mm}^2$ at the back. They are organized
 543 in pairs within thin-walled alveolar structures called submodules. The submodules are
 544 assembled into different types of modules that differ by their location with respect to η
 545 and contain 400 or 500 crystals each. These modules are then arranged in sets of four
 546 modules, called supermodules, and contain 1700 crystals each. Thus, the EB is composed
 547 of two half-barrels, each consisting of 18 supermodules.

548

549 In contrast, the photodetectors used in the endcap section of the ECAL are vacuum
 550 phototriodes (VPT) [?]. Each of the endcaps contain 7,324 crystals which, in total, oc-
 551 cupy a volume of 2.90 m^3 and weigh about 24.0 t. The crystals in the EE are 220 mm in
 552 length with a cross-section of $28.62 \times 28.62 \text{ mm}^2$ for the front face and $30 \times 30 \text{ mm}^2$ in the
 553 rear. They are all identical in shape and are arranged in mechanical units of 5×5 crystals,
 554 called supercrystals (SC), which consist of carbon-fibre alveola structures. Each of the
 555 endcaps are divided into two semi-circular structures, called *Dees*, which hold a total of
 556 3,662 crystals.

557

558 The ES preshower is located before the EE detector and spans a pseudorapidity range
 559 of $1.653 < |\eta| < 2.6$. Its main purpose is to identify neutral pions in the endcaps as
 560 well as to improve the determination of electrons and photons with high granularity. The
 561 preshower consists of two layers and has a total thickness of 20 cm. The first layer is
 562 conformed by lead radiators which initiate electromagnetic showers from incoming pho-
 563 tons and electrons. Meanwhile, the second layer is composed of silicon strips which are
 564 capable of measuring the deposited energy and the transverse shower profiles.

565 2.2.3 Hadron Calorimeter

566 The CMS Hadron calorimeter (HCAL) conforms the next layer of the CMS detector.
 567 It is a sampling calorimeter that consists of alternating layers of massive absorbing brass
 568 plates and plastic scintillator tiles and is of particular importance for the measurement of
 569 hadron jet energy and p_T^{miss} . The HCAL detector is located in between the outer extent of
 570 the ECAL ($R = 1.77 \text{ m}$) and the inner extent of the magnet coil ($R = 2.95 \text{ m}$). Similar to
 571 the other CMS subsystems, it's composed of a barrel part (HB) and an endcap part (HE).

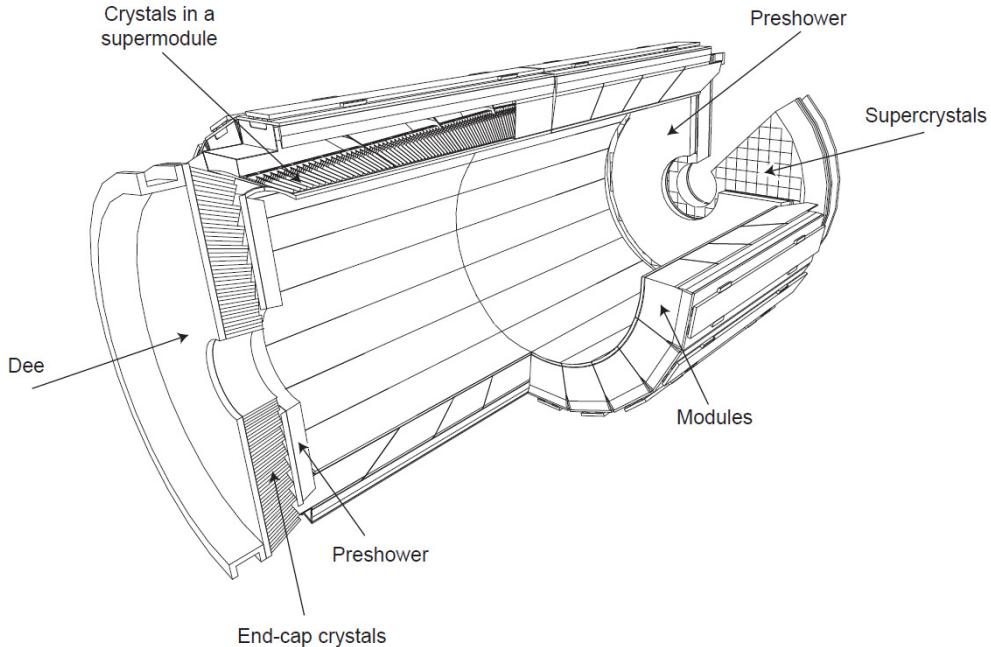


Figure 2.5: Layout of the CMS ECAL illustrating its various components [?].

572 In addition, it features a tail-catching outer calorimeter (HO), located outside the magnet,
 573 and a forward calorimeter (HF) in the very forward region near the beam line. A layout
 574 of the HCAL system can be seen in [Figure 2.6](#).

575

576 The barrel component of the HCAL is a sampling calorimeter which covers the pseudo-
 577 rapidity range $|\eta| < 1.3$. It consists of both the HB and the HO detectors. The reason
 578 behind separating the barrel detector into the HB and HO is due to the limited amount
 579 of space available for the barrel detector. The HB is located within the superconducting
 580 magnet coil and is supplemented by the HO in between the outer solenoid coil and the
 581 muon chambers. Therefore, the HO acts as a tail-catcher in order to improve the jet en-
 582 ergy measurements and p_T^{miss} , with the solenoid in between acting as absorber material.
 583 The HB consists of two half-barrel sections, identified as HB+ and HB- due to their ge-
 584ometrical location, which are composed of 36 identical azimuthal wedges. The wedges,
 585 which are constructed out of flat brass absorber plates, are aligned parallel to the beam
 586 axis and are segmented into four azimuthal angle (ϕ) sections.

587

588 The HE covers a significant amount of the pseudorapidity in the range of $1.3 < |\eta| <$
 589 3, a region containing about 34% of the particles produced in the final state. Due to
 590 the high luminosity of the LHC, the HE is required to have a high radiation tolerance at
 591 $|\eta| \simeq 3$, as well as being capable of handling high counting rates. Similar to the HB, the
 592 HE is also composed of brass absorber plates and scintillator plates which are read out
 593 by wavelength shifting fibers. The light captured by the scintillators merges within the
 594 wavelength shifting fibers and then it's read out by hybrid photo-diodes. The scintillators
 595 are partitioned in towers with an area of $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$.

596

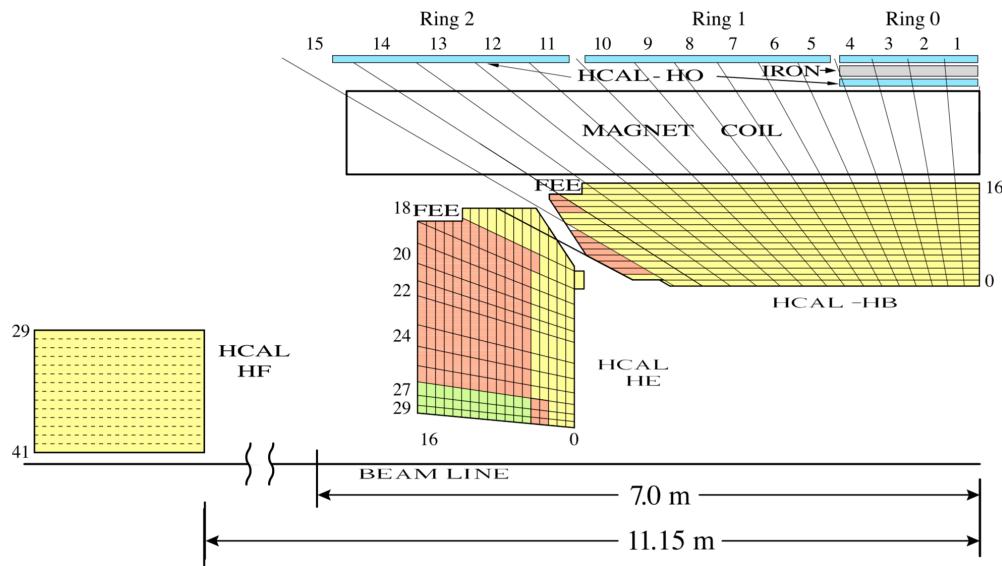


Figure 2.6: Geometrical layout of the HCAL showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [?].

597 2.2.4 Magnet

598 The CMS superconducting solenoid is one of the driving features of the detector de-
 599 sign. It is capable of providing a magnetic field with a 3.8T magnitude, which allows
 600 for the large bending power needed for precise particle transverse momentum (p_T) mea-
 601 surements. The magnet is made of four layers of stabilized reinforced Niobium-Titanium
 602 (NbTi) and has a cold mass of 220 t. The solenoid consists of a 13 m long coil with an
 603 internal diameter of 6 m, which houses both the tracking and calorimetric system. This
 604 design allows for particles to be measured prior to crossing the magnetic coil which sig-
 605 nificantly improves the energy resolution.

606 2.2.5 Muon Detector

607 As implied by the detector's name, precise and robust muon measurements have been
 608 a central theme of the CMS experiment since the early stages of its design. The detector
 609 design takes into account that muons behave as minimum ionizing particles (MIPs) [?]
 610 and can therefore manage to traverse the tracker and calorimeters with minimal energy
 611 loss. Furthermore, due to their relatively long lifetime they can be efficiently identified
 612 by a dedicated system at the outer region of the detector. Consequently, the CMS muon
 613 systems comprise the outermost layer of the detector, which are integrated into the mag-
 614 net return yoke that surrounds the solenoid.

615 The muon system is capable of three main functions: muon identification, muon p_T
 616 measurement and triggering. It is composed of three different types of detectors, all of
 617 which make use of gaseous chamber technology. This choice of detector provides a cost
 618 efficient way of covering most of the full solid angle, featuring a total of 25,000 m² of de-
 619 tection plates. As a consequence of the shape of the solenoid magnet, the muon detector
 620 was designed to have a cylindrical barrel section as well as two planar endcap regions.
 621

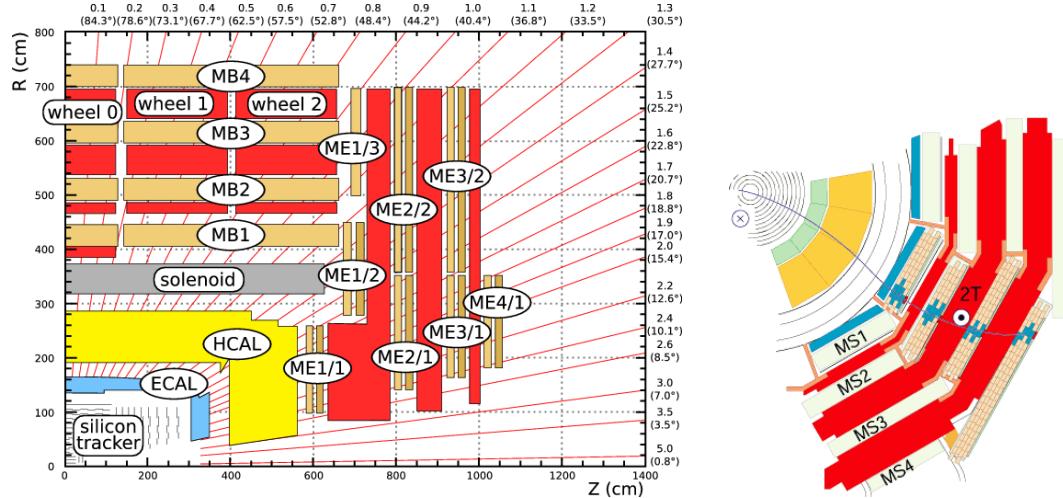


Figure 2.7: The left diagram shows a quarter-view of CMS with both the muon barrel (MB) and endcap (ME) stations [?]. The right diagram shows a muon in the transverse plane leaves a curved trajectory across the four muon detector stations [?].

For the barrel region of the muon detector, four layers of drift tube (DT) modules are used, covering a pseudorapidity of up to $|\eta| < 1.2$. These four layers, called “stations”, are arranged in cylindrical concentric layers around the beam line, where the first three layers have 60 DTs each and the outer cylinder has 70. Each of the DT stations contain 12 individual gas filled tubes, all of which have a 4cm diameter and a center electrode. The use of DTs as tracking detectors for the barrel muon system is possible because of the low expected muon rate and the relatively low strength of the local magnetic field.

The endcap regions of the muon system are subject to a higher muon rate, and cover the range $0.9 < |\eta| < 2.4$ where the magnetic field is stronger and less homogeneous. Considering these conditions, cathode strip chambers (CSCs) are employed, which feature high granularity, fast response time and adequate radiation hardness. CSCs consist of arrays of positively-charged “anode” wires crossed with negatively-charged copper “cathode” strips within a gas volume. They are trapezoidal in shape and can cover either 10° or 20° in ϕ . Furthermore, CSCs have the advantage of featuring both precision muon measurement and muon trigger in a single device.

The third type of detector used in the CMS muon system are called resistive plate chambers (RPCs) and can be found in both the barrel and endcap regions. They consist of gaseous parallel-plate detectors capable of providing precise timing information and adequate spatial resolution. Because of their excellent time resolution, RPCs provide the capability of tagging the time of an ionizing event between 2 consecutive LHC bunch crossings (BX) in a much shorter time (~ 1 ns) than the interval between the BXs (25 ns). For this reason, an RPC-based dedicated muon trigger device can be implemented to unambiguously identify the relevant BX to which a muon track is associated with, despite the high rate of events and background expected at the LHC.

2.2.6 Trigger and Data Acquisition

Due to the vast volume of data originating from the proton-proton collisions (delivered by the LHC at a rate of 40 MHz), a method of eliminating the majority of the un-

interesting/unwanted events was a requisite for the CMS detector design. This event rate reduction is achieved by the implementation of the so-called trigger system, which manages to select the potentially interesting interactions and reduce the rate from a staggering 40 TBs^{-1} to a manageable value of just a few hundred Hz.

656

The CMS trigger system is implemented using a two-stage rate reduction, which combines both a hardware and software phase. The combination of both of these triggers is designed to reduce the rate by a factor of $\sim 10^6$. The first stage used in the rate reduction is purely hardware based and it is called the Level 1 (L1) Trigger [?], which consists of both Field Programmable Gate Array (FPGA) and Application Specific Integrated Circuit (ASIC) technology. During this initial stage, the rate is reduced to about 100 kHz with a latency of $3.2 \mu\text{s}$. This time interval constrains the trigger decision, allowing only for data from the calorimeters and muon system to be processed. Trigger primitives (TP) from these subsystems are processed through a series of steps before the combined event information is evaluated by the global trigger (GT) where the final decision, whether or not to accept the event, is made.

668

The second stage, which implements offline-quality reconstruction algorithms in its event selection, is referred to as the High-Level Trigger (HLT) [?]. The event selection process for the HLT requires that physics objects for each event, such as electrons, muons and jets, are reconstructed and undergo predetermined identification criteria.

673

2.2.7 Event Reconstruction

In order to reconstruct the events the particle flow (PF) algorithm [?] is used. This algorithm gathers information from all the CMS sub-detectors to reconstruct charged and neutral hadrons, photons, muons, and electrons. It relies on an efficient and pure track reconstruction, a clustering algorithm able to distinguish overlapping tracks originated from different vertices, and an efficient link procedure to associate each particle deposit in the sub-detectors. Once all the deposits of a particle are associated, it can be correctly identified and its four-momentum optimally determined from the combined information of the sub-detectors. The resulting list of particles are then used to reconstruct higher level objects such as jets, taus, missing transverse energy, and to compute charged lepton and photon isolation, etc [?]. The CMS experiment is provided with millions of collisions per second, which need to be triggered, detected, stored and analyzed in a collaboration of several thousand physicists. This huge amount of data and the complexity of the detector require a flexible data model that serves all the needs of the collaboration. The data format is optimized for performance and flexibility of the reconstruction for the end user's analysis.

689

Event information from each step in the simulation and reconstruction chain is logically grouped into what is called a data tier [?]. From the physicist's point of view the most important data tiers are RECO, which contains all reconstructed objects and hits, and AOD (a subset of RECO). The AOD will contain a copy of all the high-level physics objects (such as muons, electrons, taus, etc.) and enough information about the event to support all the typical usage patterns of a physics analysis. It also contains a summary of the RECO information sufficient to support typical analysis actions such as track re-

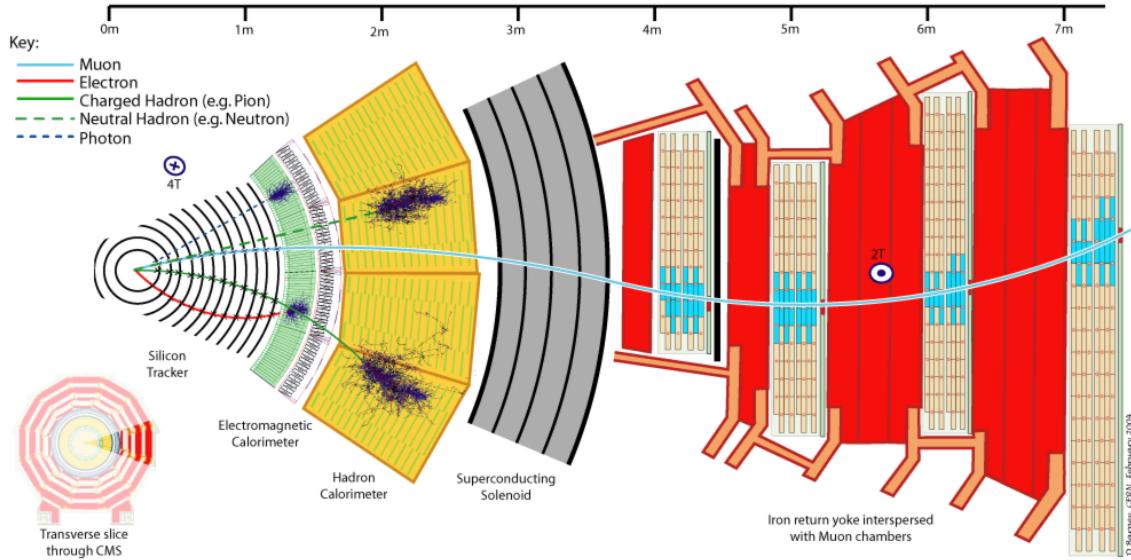


Figure 2.8: Transverse slice of the CMS detector, showing the individual detector subsystems and particle signatures in each. The particle type can be inferred by combining the detector response in the different subdetectors [?].

fitting with improved alignment or kinematic constraints, re-evaluation of energy and/or position of ECAL clusters based on analysis-specific corrections. The format of each data tier is ROOT [?]. ROOT is a framework for data processing developed at CERN with the sole purpose of aiding high energy physics research. The various AOD datasets are stored worldwide at various data tier centers. From the AOD's the analysis groups create data structures called NTuples containing only the high-level physics objects needed for their particular analysis.

2.2.8 Future Upgrade of Pixel Detector

After the first LHC shutdown called LS1 (2013-2014), and the installation of the Phase-1 Pixel Detector [?] in early 2017, among other things, the LHC is planning another series of upgrades during two major shutdowns, called LS2 and LS3, currently planned for 2019-2020 and ~ 2024 , respectively. LS2 would result in a further increase of the luminosity beyond the original design value, to over $2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. With the LS3 upgrade of the LHC (called HL-LHC [?]) the luminosity is expected to reach up to $7.5 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. Correspondingly, the CMS collaboration has planned a series of further upgrades [?, ?] that will ensure the capabilities of the CMS detector to match to the HL-LHC running conditions, while taking the opportunity to improve the performance and repair any problems uncovered during the data-taking periods. The UPRM group will continue its involvement in the Phase-1 Pixel Detector operations and Phase-2 Pixel Upgrade design.

The HL-LHC conditions of instantaneous peak luminosities of up to $7.5 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and an integrated luminosity of the order of 300 fb^{-1} would result in 1 MeV neutron equivalent fluence of $2.3 \cdot 10^{16} \text{ neq/cm}^2$ and a total ionizing dose (TID) of 12MGy (1.2 Grad) at the center of CMS, where its innermost component, the Phase-2 Pixel Detector will be installed. The detector should be able to withstand the above radiation dose, han-

723 dle projected hit rates of $3\text{GHz}/\text{cm}^2$ at the lowest radius, be able to separate and identify
 724 particles in extremely dense collision debris, deal with a pileup of 140-200 collisions per
 725 bunch crossing and have a high impact parameter resolution. This translates into requiring
 726 a detector design that is highly granular, has thinner sensors and smaller pixels, and faster,
 727 radiation hard electronics compared to its Phase-1 counterpart. The selection of interesting
 728 physics events at the Level-1 (L1) trigger and inefficiency of selection algorithms in
 729 high pileup conditions further require the Tracker to be included in this trigger stage, helping
 730 reduce the event rate from 40 MHz rate to 7.5 kHz. The physics goals also require
 731 an increase in Pixel Detector coverage to $|\eta| = 4.0$ which improves the p_T^{miss} resolution
 732 and particle-flow event reconstruction by providing p_T measurements and trajectories for
 733 charged particles entering the calorimeters. The p_T^{miss} resolution is an essential perfor-
 734 mance parameter for many BSM physics searches including SUSY and extra dimension
 735 models where particles escape undetected from the detector space. The smaller pixel size
 736 will further improve b-tagging as well as hadronic reconstruction and track reconstruction
 737 efficiencies within boosted jets, which can be produced from new heavy objects decaying
 738 to Higgs, Z bosons, or top quarks – all heavy probes that can be exploited for new physics
 739 searches. Improving the b-tagging capabilities directly affects our analysis due to the im-
 740 portance of properly identifying top quark decays.

741

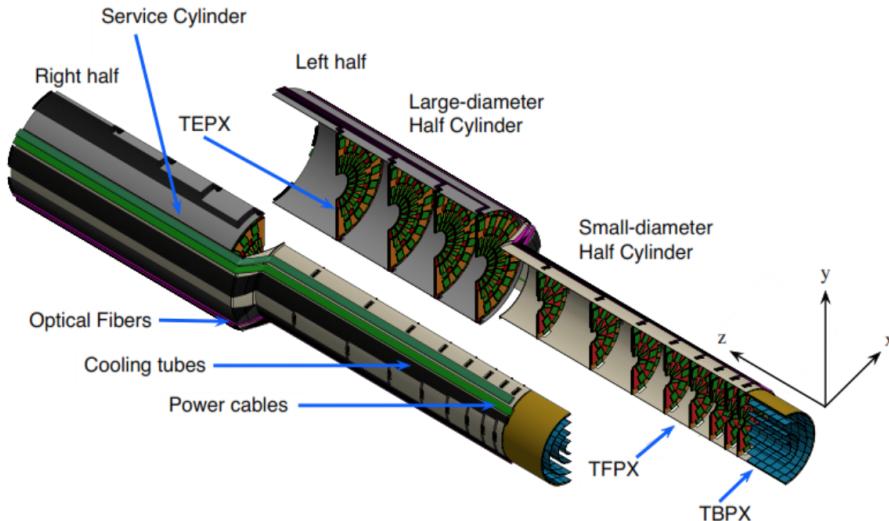


Figure 2.9: Phase-2 Pixel Detector Layout[ref].

742 The Phase-2 Pixel Detector [?,?] baseline design comprises a barrel part with 4 layers
 743 of Tracker Barrel Pixel Detector (TBPX), 8 small double-discs per side of Tracker For-
 744 ward Pixel Detector (TFPX) and 4 large double-discs per side of Tracker Endcap Pixel
 745 Detector (TEPX). This forward and end part is referred to as 8l4s (8 TFPX and 4 TEPX).
 746 In the TBPX the pixel modules are arranged in “ladders”. In each layer, neighboring ladders
 747 are mounted staggered in radius, so that $r\text{-}\phi$ overlap between the ladders is achieved.
 748 The modules on a ladder do not overlap in z . A projective gap at $\eta = 0$ is avoided by
 749 mounting an odd number of modules along z , and by splitting the barrel mechanics in z
 750 into slightly asymmetric halves. In TFPX and TEPX the modules are arranged in concen-
 751 tric rings. Each double-disc is physically made of two discs, which facilitates to mount
 752 modules onto four planes, with overlaps in r as well as $r\text{-}\phi$. Each disc is split into two

753 halves, and these D-shaped structures are referred to as “dees”. The TEPX will provide
754 the required luminosity measurement capability, by an appropriate implementation of the
755 readout architecture. In total, the pixel detector will have an active surface of approxi-
756 mately 4.9 m^2 . [Figure 2.9](#) shows the layout of the Phase-2 Pixel Detector.

757 Chapter 3

758 Theoretical Motivation

759 3.1 The Standard Model

760 Quantum Mechanics and the Theory of Special Relativity, along with the plethora of
761 new particles discovered over the past century or so, has led to the formulation of the
762 Standard Model of Particle Physics (SM) [?]. It is a theory concerning the electromagnetic,
763 weak [?], and strong nuclear interactions [?], as well as classifying all the subatomic
764 particles known. It provides our best current understanding of the universe, what matter
765 is made of and how it is held together. It rests on two basic ideas: all matter is made of
766 particles, and these particles interact with each other by exchanging other particles associated
767 with the fundamental forces. The basic components of matter are fermions and the
768 force carriers are bosons and fermions have half-integer values of spin whereas bosons
769 have integer values.

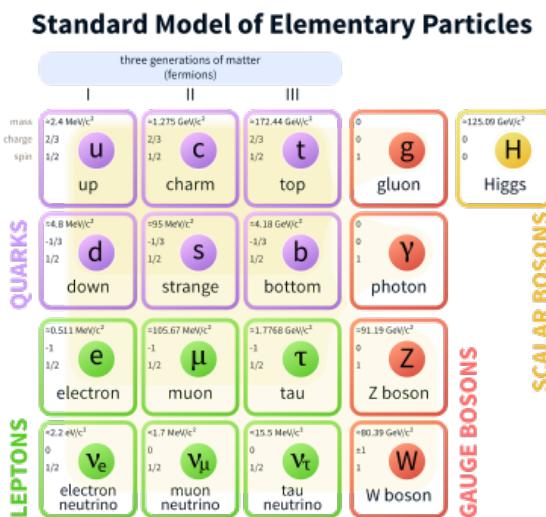


Figure 3.1: Schematic representation of the Standard Model particles. Shown are the three generations of matter (formed by fermions), the gauge bosons and the Higgs boson.

770 Fermions come in two families as shown in Figure 3.1 - the leptons and quarks. The
771 lepton family has six members, meanwhile the quark family contains six quarks. The
772 up and down quarks are found inside protons and neutrons. The twelve fermions are the
773 building blocks of matter and each one has a spin value of $\frac{1}{2}$. These particles interact

774 with each other through fundamental forces. Each force comes with one or more force
775 carriers. The nuclear force comes with the gluon and binds the quarks within the proton
776 and neutrons. The photon is associated with the electromagnetic force and is responsible
777 for mediating many of the forces humans encounter on a daily basis. The weak interaction
778 is responsible for radioactivity and is mediated by the Z and W bosons. The gluons,
779 photons and the Z and W bosons all have a spin of 1. The Standard Model is both remark-
780 ably simple and very powerful with nearly every measured quantity in particle physics
781 laboratories over the past five decades falling right on the predicted value (within exper-
782 imental error margins) [?, ?]. Until recently, the only missing piece of the SM was the
783 Higgs boson [?]. The Higgs boson is a particle corresponding to the Higgs field which,
784 according to the SM, gives mass to all the fundamental particles. It was finally discovered
785 at the LHC in 2012 [?, ?].

786 3.2 Indications of Physics Beyond the Standard Model

787 The Standard Model gives an incomplete picture of the makeup of the Universe as
788 listed below:

- 789 1. The Standard Model does not include gravity and does not explain why gravity is
790 so much weaker than the electromagnetic or nuclear forces [?]. The weak force
791 is 10^{24} times as strong as gravity. It is also incompatible with the most successful
792 theory of gravity to date, general relativity [?, ?].
- 793 2. There is a wide spectrum of masses among the building blocks of matter [?]. Why?
794 The electron is about 200 times lighter than the muon and 3,500 times lighter than
795 the τ -lepton, meanwhile the top quark is 75,000 times heavier than the up quark.
796 Why is there such a wide spectrum of masses among the building blocks of matter?
- 797 3. The SM introduces particle masses through a process known as spontaneous sym-
798 metry breaking [?, ?, ?] caused by the Higgs field. Within the standard model, the
799 mass of the Higgs gets some very large quantum corrections due to the presence of
800 virtual particles (mostly virtual top quarks). These corrections are much larger than
801 the actual mass of the Higgs. This means that the bare mass parameter of the Higgs
802 in the standard model must be fine-tuned in such a way that almost completely
803 cancels the quantum corrections. This level of fine-tuning is deemed unnatural.
- 804 4. It explains only about 5% of matter present in the universe. However, 26% exists
805 as dark matter [?] that behaves just like other matter (in terms of gravity) but which
806 only interacts weakly (if at all) with the Standard Model fields. Yet, the Standard
807 Model does not supply any fundamental particles that are good dark matter candi-
808 dates. The rest (69%) is dark energy [?], a constant energy density for the vacuum.
809 Attempts to explain dark energy in terms of the standard model have failed.
- 810 5. Neutrinos are predicted to be massless particles by the Standard Model [?]. How-
811 ever, neutrino oscillation experiments have shown that neutrinos do have mass [?, ?].
812 Mass terms for the neutrinos can be added to the standard model by hand, but these
813 lead to new theoretical problems as these mass terms need to be extraordinarily
814 small and it is doubtful if the neutrino masses would arise in the same way than the
815 masses of other fundamental particles do in the Standard Model.

- 816 6. The universe is made out of mostly matter but the standard model predicts that matter
 817 and antimatter should have been created in almost equal amounts if the initial
 818 conditions of the universe did not involve disproportionate matter relative to anti-
 819 matter [?]. But no mechanism, sufficient to explain matter–antimatter asymmetry,
 820 exists in the Standard Model.

821 Thus, as long as we observe various phenomena at low energy the Standard Model
 822 behaves properly, however, it lacks robustness at higher energy. To overcome its short-
 823 comings several theories exist beyond the Standard Model that include various exten-
 824 sions of the standard model through supersymmetry, such as the Minimal Supersym-
 825 metric Standard Model (MSSM) [?, ?] and Next-to-Minimal Supersymmetric Standard
 826 Model (NMSSM) [?], or entirely new ones like such as string theory [?] and extra di-
 827 mensions [?]. As these theories tend to reproduce the entirety of current phenomena, the
 828 question of which theory is the right one can only be determined by experiments. With the
 829 advent of even more powerful accelerators like the LHC, we are within reach of energy
 830 levels which existed only shortly after the Big Bang where the Standard Model has issues.

831 3.3 Supersymmetric extension of the Standard Model

832 The supersymmetry theories are based on a symmetry between fermions and bosons.
 833 It is similar to solving the ‘electron mass hierarchy problem’ in quantum mechanics,
 834 where the number of particles is doubled: in addition to the electron, there is also a
 835 positron. The virtual electron–positron contributions solved the problem of electrons
 836 having a small mass by smearing out the electric charge. Supersymmetry is an analo-
 837 gous theory where once again the set of particles is doubled, and in doing so the loop
 838 contributions of one particle to the Higgs are cancelled by the loop contributions of its
 839 super-partner. It extends space-time symmetry since it relates matter particles to force
 840 particles. It relates particles with different spins but the same gauge charges.

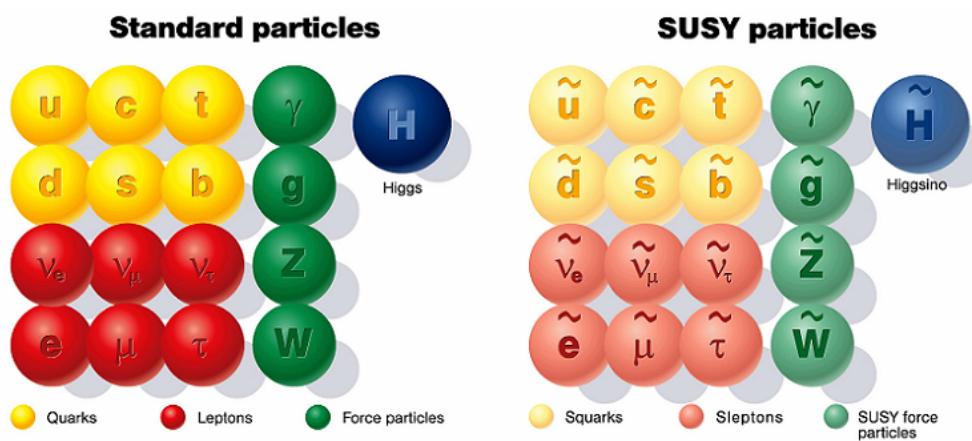


Figure 3.2: Schematic representation of the Standard Model particles with the added SUSY particles.

841 The gluon (QCD) has a fermionic partner, the gluino. Similarly, the spin 1 gauge
 842 bosons W^+ , W^0 , W^- and B^0 have their counterparts, called winos and binos, that, af-
 843 ter electroweak symmetry breaking, mix to give the mass eigenstate Z^0 and γ , and the

844 zino and photino respectively. For every lepton/quark there is a bosonic partner (i.e. a
845 scalar slepton/squark). There are three families for each of the quark and lepton super-
846 multiplets. The left-handed and right-handed pieces of the squarks and sleptons are two
847 separate components, as in the corresponding SM sector. The Higgs doublet is extended
848 by another doublet leading to other Higgs boson particles (Higgsinos).

849

850 Even though supersymmetry solves many problems in particle physics, it also poses
851 new problems. If these new particles had the same mass as their counterparts in the stan-
852 dard model, we would have already observed them. Since none of these particles have
853 been found yet, SUSY must be a broken symmetry. What makes superpartners heavier
854 than ordinary particles? Why are superpartners so well hidden in rare phenomena? This
855 arbitrary mass spectrum for the superpartners would have effects that are far too large
856 in rare processes that change the flavor of particles. There must be some special reason
857 why such effects are well hidden. How do we extract information on the mechanism of
858 supersymmetry breaking? How does supersymmetry impact cosmology? Is the lightest
859 supersymmetric partner what really composes Dark Matter? Different mechanisms have
860 been suggested on how this symmetry is broken and have led to different phenomenologi-
861 cal scenarios like the minimal supergravity model (mSUGRA) [?, ?], the gauge-mediated
862 supersymmetry breaking (GMSB) [?, ?, ?] model or R-parity [?, ?]. A discussion of
863 mSUGRA and GMSB is beyond the scope of this dissertation. R-parity is a new sym-
864 metry that has been added to the minimal supersymmetry scenario which prevents the
865 violation of the leptonic and baryonic numbers. All SM fields have even R-parity, while
866 SUSY particles have odd R-parity. The most obvious experimental constraint comes from
867 the non-observation of proton decay [?], which would violate both baryonic and leptonic
868 number conservation. Due to R-parity every interaction vertex in the theory involves an
869 even number of sparticles, meaning that they must be pair-produced. The decay chains
870 of the produced particles are characterized by the presence of one stable particle, generi-
871 cally called the Lightest Supersymmetric Particle (LSP), that may account for all the dark
872 matter in the Universe. However, in other models, like the R-parity violating SUSY this
873 may not apply. In other scenarios, those LSP can be regarded as vanishing mass and may
874 not be enough to satisfy the dark matter in the Universe.

875 3.3.1 Simplified Models Approach to Supersymmetry

876 Simplified models [?] are a new approach for characterizing LHC supersymmetry re-
877 sults. In a simplified model only a few new particles and a single decay topology are
878 introduced, and hence these models are easy to constrain. Traditionally, testing a particle
879 physics theory against experimental results requires calculating what the theory predicts
880 in an experimental situation using a particle accelerator detector. This translates into
881 computing: the masses of new particles, their decay widths, their branching ratios and
882 production cross-sections. This information is subsequently used in Monte Carlo simula-
883 tions of passage of proton-proton collisions products and their decays through the CMS
884 detector. This is done under the hypothesis that the theory is true. The simulated data
885 is analyzed just like the actual experimental data resulting from proton-proton collisions.
886 The theoretical and the experimental results can then be compared. Though conceptually
887 uncomplicated it requires access to the experimental data, and the step of simulating the
888 detector response requires intimate knowledge of the CMS detector by theorist commu-

nity. It is also restrictive being model/theory dependent. The CMS collaboration, like ATLAS, does not share experimental details but only the final results via publications. The Simplified Model approach circumvents this issue by presenting results on very simple phenomenological models instead of a specific model, such as the constrained MSSM (cMSSM) [?]. Each simplified models encompasses a specific phenomenological feature that is common to many different theories. A limited set of hypothetical particles and decay chains are introduced to produce a given topological signature. The amplitudes describing the production and decays of these particles are parametrized in terms of the particle masses and their branching ratios to daughter particles. This makes the analysis of simplified models less model dependent. The Simplified Models assume that a particular decay signature can be realized without specifying the exact mechanism, offering the possibility to overcome small branching ratios. Some Simplified Models with their corresponding production and decay modes are shown in the [Table 3.1](#). We have used T2tt, T1ttt, T1ttbb, T5ttt and T5ttcc models in analysis in this thesis. Some of these topologies are further discussed in more detail in [chapter 5](#), since they correspond to some of the signals for SUSY in the treated analysis.

Model name	Production mode	Decay	Visibility
T1	$\tilde{g}\tilde{g}$	$\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_{\text{LSP}}$	All-Hadronic
T2	$\tilde{q}\tilde{q}^*$	$\tilde{q} \rightarrow q\tilde{\chi}_{\text{LSP}}$	All-Hadronic
T5zz	$\tilde{g}\tilde{g}$	$\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_2^0, \tilde{\chi}_2^0 \rightarrow Z\tilde{\chi}_{\text{LSP}}$	All-Hadronic
			Opposite-Sign Dileptons Multileptons
T3w	$\tilde{g}\tilde{g}$	$\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_{\text{LSP}}$ $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_1^\pm, \tilde{\chi}_1^\pm \rightarrow W^\pm\tilde{\chi}_{\text{LSP}}$	Single Lepton + Jets
T5lnu	$\tilde{g}\tilde{g}$	$\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_1^\pm, \tilde{\chi}_1^\pm \rightarrow \ell\nu\tilde{\chi}_{\text{LSP}}$	Same-Sign Dileptons
T3lh	$\tilde{g}\tilde{g}$	$\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_{\text{LSP}}$ $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_2^0, \tilde{\chi}_2^0 \rightarrow \ell^+\ell^-\tilde{\chi}_{\text{LSP}}$	Opposite-Sign Dileptons
T1bbbb	$\tilde{g}\tilde{g}$	$\tilde{g} \rightarrow b\bar{b}\tilde{\chi}_{\text{LSP}}$	All-Hadronic (b)
T1ttt	$\tilde{g}\tilde{g}$	$\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_{\text{LSP}}$	All-Hadronic (b) Single Lepton + Jets (b) Same-Sign Dileptons (b) Inclusive (b)
T2bb	$\tilde{b}\tilde{b}^*$	$\tilde{b} \rightarrow b\tilde{\chi}_{\text{LSP}}$	All-Hadronic (b)
T6ttww	$\tilde{b}\tilde{b}^*$	$\tilde{b} \rightarrow t\tilde{\chi}^-, \tilde{\chi}^- \rightarrow W^-\tilde{\chi}_{\text{LSP}}$	Same-Sign Dileptons (b)
T2tt	$\tilde{t}\tilde{t}^*$	$\tilde{t} \rightarrow t\tilde{\chi}_{\text{LSP}}$	All-Hadronic (b)

Table 3.1: Some Simplified Models with their corresponding production and decay modes.

905 **Chapter 4**

906 **Optimizing the Phase-2 Pixel Geometry
907 for the Level-1 Trigger in the Outer
908 Tracker**

909 **4.1 Upgraded Phase-2 Tracker**

910 This chapter is dedicated to a study that seeks to optimize the geometry of the Phase-2
911 Pixel Detector (for the HL-LHC) in terms of how the upgraded Inner Tracker (IT) affects
912 the overall stub production (the concept of a stub is defined in [subsection 4.1.2](#)). The
913 goal of this study was to determine how the removal of discs from the Phase-2 Forward
914 Pixel detector would affect the production of L1 stubs in the upgraded CMS Outer Tracker
915 (OT). More information of the Phase-2 Tracker can be found at [\[?\]](#). This study has a direct
916 impact on the future of SUSY analyses due to the importance of the tracker in searches
917 for BSM physics.

918

919 In the HL-LHC era (2023-2035) an integrated luminosity of $3,000 \text{ fb}^{-1}$ is expected to
920 be reached at a center-of-mass energy of 14 TeV. This data will enable us to make precise
921 measurements of several SM parameters (particularly the scalar sector). This will also
922 lay the groundwork for new physics searches proposed by many of the BSM models. In
923 addition, the HL-LHC project is essential to identify any hints of new physics, including
924 SUSY, that could possibly be observed during the LHC Run 2 or 3, or beyond. The
925 bigger, and highly granular pixel and silicon tracker, and its extended η coverage, will
926 improve jet sub-structure, boosted object identification and heavy object tagging to hunt
927 for BSM physics, specifically Supersymmetry. A discovery could happen with various
928 few sigma excesses and then interpretation of the new signal will require a large amount
929 of data. At the HL-LHC, much gain is expected for weakly produced SUSY particles
930 because they are characterized by small production cross-sections. Considering the case
931 that gluinos and squarks are heavy, search for electroweak production channels will offer
932 the best option to discover SUSY. At the HL-LHC, charginos/neutralinos are expected to
933 be discovered (significance of 5σ) up to masses of about 950 GeV. In the case of strong
934 production, the mass reach for gluinos goes up to 2.2 TeV. This study is a contribution in
935 the potential future discovery of SUSY and other BSM physics.

4.1.1 Phase-2 Tracker Layout

The Phase-2 tracker design features an Inner Tracker based on silicon pixel modules and an Outer Tracker made from silicon modules with strip and macro-pixel sensors. The IT is composed of a Tracker Barrel Pixel Detector (TBPX) with four barrel layers, and 8 small Tracker Endcap Pixel Detector (TEPX) discs plus 4 Tracker Forward Pixel Detector (TFPX) large discs in the forward direction.

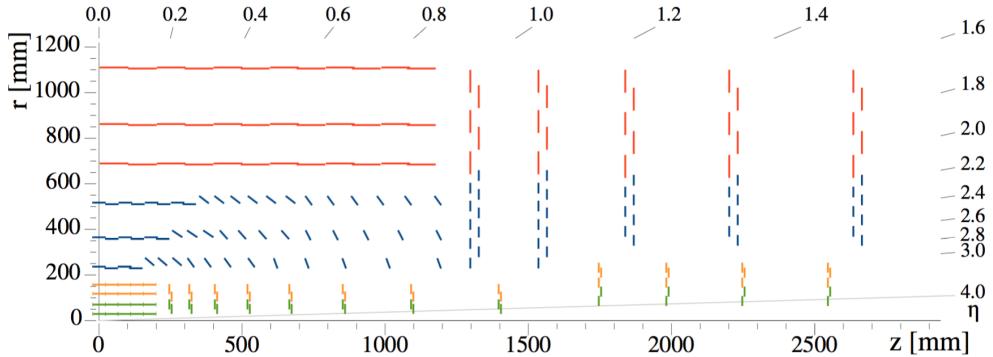


Figure 4.1: Sketch of one quarter of the tracker layout in r-z view. In the Inner Tracker the green lines correspond to pixel modules made of two readout chips and the yellow lines to pixel modules with four readout chips. In the Outer Tracker the blue and red lines represent the two types of modules p_T modules - (2S and PS).

The OT will feature so-called “ p_T modules”, which are capable of providing tracking information for the L1 trigger. The OT consists of six barrel layers (TBPS and TB2S) and five endcap double-discs (TEDD) which are composed of two types of p_T modules, called 2S and PS modules. The first two endcap double-discs contain a total of 15 rings and the remaining three feature 12 rings each. This is shown in [Figure 4.2](#).

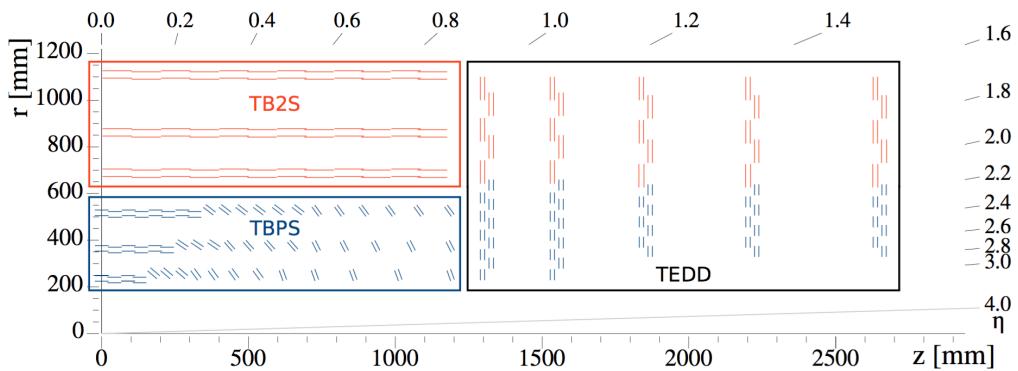


Figure 4.2: Sketch of one quarter of the Outer Tracker in r-z view. Blue (red) lines represent PS (2S) modules. The three sub-detectors, named TBPS, TB2S, and TEDD, are indicated. All overlapping layers are shown separately.

4.1.2 Tracker Input to the L1-Trigger

The enhancement of the trigger performance involves both a higher output rate of interesting events and an improved discriminating power of the event selection, which is

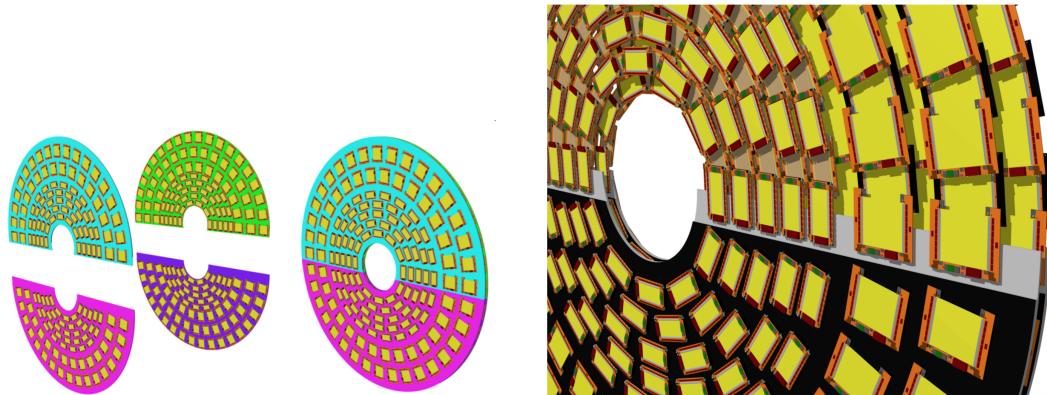


Figure 4.3: Sketch of four dees (left) forming a double-disc (centre), and drawing of a part of a TEDD double-disc (right), illustrating the overlap of modules in ϕ and z . The upper two dee support structures are removed in order to show all layers of modules. The modules ob the disc are arranged in 15 or 12 rings.

more challenging in a high-pileup environment. Improved discriminating power will be achieved by using more information in the trigger decision, with a longer latency available for its processing. The use of tracking information in the L1 trigger will improve the p_T resolution of various objects at L1 (e.g. jets), it will allow the exploitation of information on track isolation, and will contribute to the mitigation of pileup of 140-200 collisions per bunch crossing. CMS plans to enhance the first level trigger rate from presently 100 kHz to 750 kHz and to increase the latency from the present value of $3.2 \mu s$ to $12.5 \mu s$. The front-end electronics and the L1 trigger track reconstruction need to comply with these new requirements. The necessity of providing tracking information to the L1 trigger is a main driver for the design of the OT, including its module concept. The use of tracking information in the L1 trigger implies that the tracker has to send out self-selected information at every bunch crossing. Such functionality relies upon local data reduction in the front-end electronics, in order to limit the volume of data that has to be sent out at 40 MHz. This is achieved with modules that are capable of rejecting signals from particles below a certain p_T threshold, referred to as “ p_T modules” [?]. Tracks from charged particles are bent in the transverse plane by the 3.8 T field of the CMS magnet, with the bending angle depending on the p_T of the particle. The modules are composed of two single-sided closely-spaced sensors read out by a common set of front-end ASICs that correlate the signals in the two sensors and select the hit pairs (referred to as “stubs”) compatible with particles above the chosen p_T threshold (as shown in Figure 4.4). Stubs are defined as matching pairs of hits on the adjacent silicon layers of the p_T modules. Whether or not a stub is accepted depends on the p_T of the particle for a certain p_T threshold value.

4.2 Tracker Stub Simulation

The baseline detector design comprises a barrel part with 4 layers of TBPX, 8 small double-discs per side of TFPX and 4 large double-discs per side of TEPEX. This forward and end part is referred to as 8l4s (8 TFPX and 4 TEPEX). Any changes made to the detector design or its geometry, due to any circumstance, directly impacts the physics variables, such as tracking efficiency, fake rate, p_T resolution, b-tagging and primary track reconstruction and needs to be simulated to understand its impact. The removal of discs from the Phase-2 Forward Pixel detector (TFPX/TEPEX) would affect the production of L1 stubs

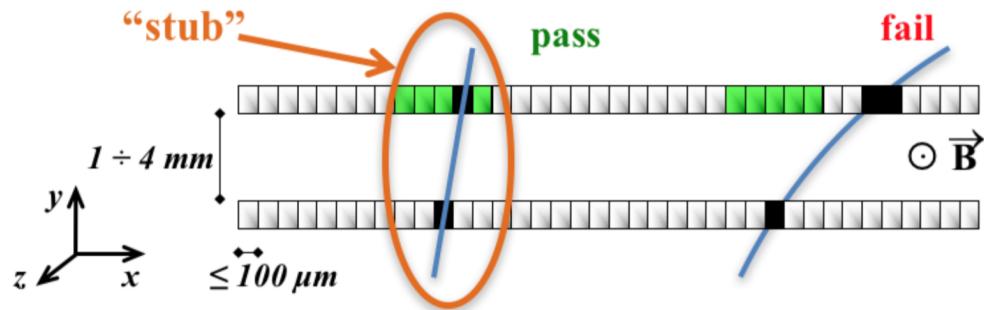


Figure 4.4: Correlation of signals in closely-spaced sensors enables rejection of low- p_T particles; the channels shown in green represent the selection window to define an accepted stub.

in the upgraded CMS OT. A decrease in the total number of stubs is expected when less material is present due to secondary interactions.

982

983 To perform this study we considered the following four different geometries of the
984 Pixel Detector:

- 985 • 8 small discs and 4 large discs ($8s4l$) – designated as the default geometry
- 986 • 8 small discs and 3 large discs ($8s3l$) - one large disc removed.
- 987 • 7 small discs and 4 large discs ($7s4l$) - one small disc removed.
- 988 • 6 small discs and 3 large discs ($6s3l$) - one large disc and two small discs removed.
- 989 • A version of the default pixel geometry with a dead disc (used to verify our find-
990 ings).

991 The response of the detector due to design modifications is simulated using the offi-
992 cial CMS software package CMSSW. The detector geometry is generated with a design
993 tool for innovative silicon tracking detectors called tkLayout [?] and is exported into the
994 standard CMSSW format along with the materials like sensing silicon elements, elec-
995 tronics and inactive material. The exported tracker geometry serves as input for the full
996 simulation and reconstruction performance studies with CMSSW. The design of the CMS
997 Phase-2 Tracker includes Tracker information for L1 triggering. The study was made
998 from $t\bar{t}$ samples with a center-of-mass energy of 14 TeV and 200 pileup. The samples
999 were created with the CMS software version referred to as CMSSW_9_3_2. The stub in-
1000 formation is used as input for the track fitting algorithm and allows tracks used for L1
1001 trigger to be reconstructed within approximately $5 \mu s$ (this includes a $1 \mu s$ estimate for
1002 the time required to transmit the stub data from the detector to the counting room) to
1003 remain within the overall L1 trigger latency budget of $12.5 \mu s$. Stubs are formed from
1004 two closely spaced sensors in Tracker measurement layers and provide linked position
1005 measurements for each particle passing through them. [Table 4.1](#) shows the stub window
1006 p_T threshold values used for the study.

TEDD1	TEDD2	TEDD3	TEDD4	TEDD5	
5.5	5.5	6.0	6.5	6.5	R1
4.5	5.0	5.5	6.0	6.0	R2
3.5	4.0	5.0	6.0	6.0	R3
3.0	3.0	2.5	3.0	3.0	R4
2.5	2.0	3.5	3.0	3.0	R5
4.0	3.0	2.5	2.0	2.0	R6
3.5	3.0	2.5	2.0	2.0	R7
3.0	3.0	2.5	2.0	2.0	R8
3.0	2.5	2.0	2.0	1.5	R9
2.5	2.0	2.0	1.5	1.5	R10
2.0	2.0	1.5	1.5	1.5	R11
2.0	2.0	1.5	1.0	1.0	R12
1.5	1.5				R13
1.5	1.5				R14
1.0	1.0				R15

TB2S3	6.5											
TB2S2	5.5											
TB2S1	4.5											
TBPS3	3.5	4.5	4.5	4.0	4.0	4.0	4.0	3.5	3.5	3.5	3.0	3.0
TBPS2	2.0	3.0	3.0	3.0	3.0	2.5	2.5	2.5	2.5	2.5	2.5	2.0
TBPS1	2.0	2.5	2.5	2.5	2.5	2.5	2.5	2.0	2.0	1.5	1.5	1.0
FLAT	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12

Table 4.1: Stub window tunings for CMSSW_9_3_2. Shown are the SW values for TEDD (top table) and for the OT barrel (below). Obtained from [this github repository](#).

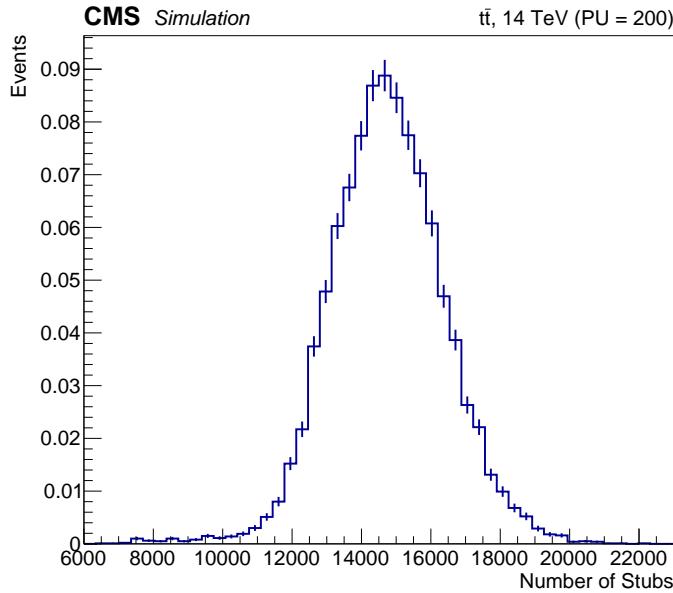


Figure 4.5: Number of stubs per event for standard pixel geometry (8s4l).

1007 4.2.1 Results of Effect of the Phase 2 Pixel Detector on the L1 Stub 1008 Rate

- 1009 (a) **Number of Stubs:** With the current (above table) stub window tuning, the total
1010 number of stubs are on average $\sim 15,000$ stubs as shown in [Figure 4.5](#).

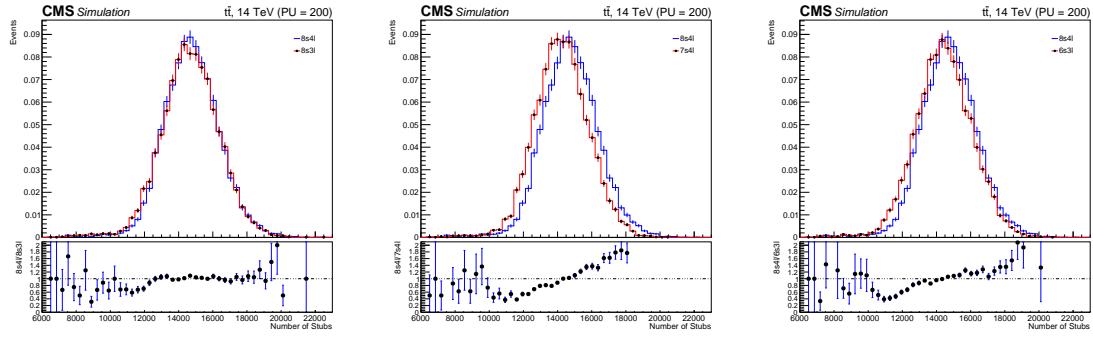


Figure 4.6: Comparison of total number of stubs per event between the standard layout and the other three considered. 8s4l vs 8s3l (left), 8s4l vs 7s4l (middle) and 8s4l vs 6s3l (right).

1011 **(b) Comparison of Stubs between default Pixel Geometry (8s4l) to disc removed:**

1012 A comparison of the total number of stubs per event between the standard geometry to
 1013 the others that have a disc or more removed (8s3l, 7s4l, 6s3l) was performed ([Figure 4.6](#)).
 1014 Out of the three comparisons shown, the largest shift in the stub distribution can be seen
 1015 from the 7s4l geometry. It was found that the removal of a smaller pixel disc has a more
 1016 pronounced effect on stub production in the OT than a larger pixel disc.
 1017

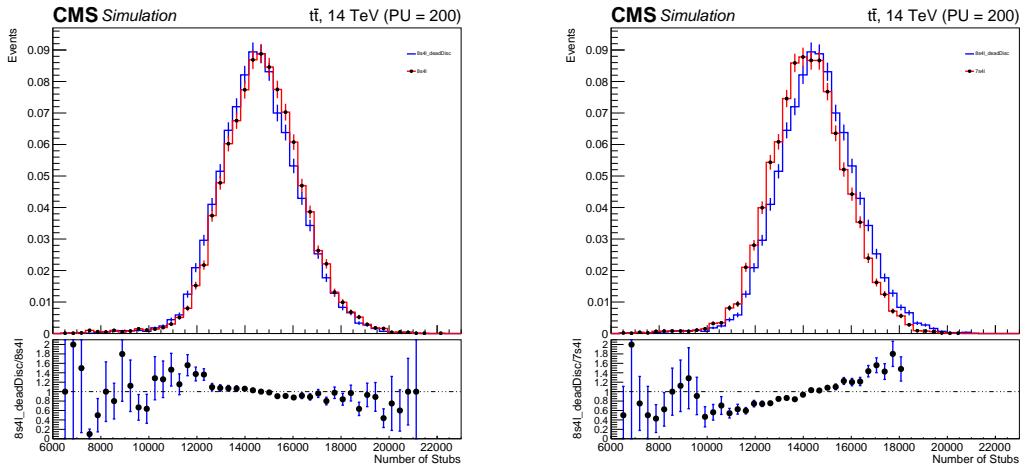


Figure 4.7: Comparison for number of stubs per event between pixel geometry 8s4l with a dead disc (blue) and both pixel geometry 8s4l (right plot) and 7s4l (left plot) (blue). Both the standard geometry and 7s4l are shown in blue in their respective plots.

1018 **(c) Comparison with a disc dead in default geometry:**

1019 An additional study, using pixel geometry 8s4l with a dead disc, was made to verify
 1020 the effect observed on the L1 stub counting. [Figure 4.7](#) (right) shows a comparison be-
 1021 tween 2 geometries that are (in principle) identical but one detects no hits on the second
 1022 disc in the positive side. This Figure shows a slight deviation from the original 8s4l ge-
 1023 ometry.
 1024

Comparing the number of stubs for the geometry with the dead disc (blue) to the one with the removed small disc (red), the same behavior (Figure 4.7, left) is observed as before. As was expected, the stub distribution for pixel geometry $8s4l$ has a higher mean value than $7s4l$, despite the dead disc. The removal of discs in the pixel detector causes the total number of stubs to decrease due to secondary interactions with the material. We also studied the effect of the pixel geometries on the stubs with respect to the pseudorapidity (η) as shown in Figure 4.8.

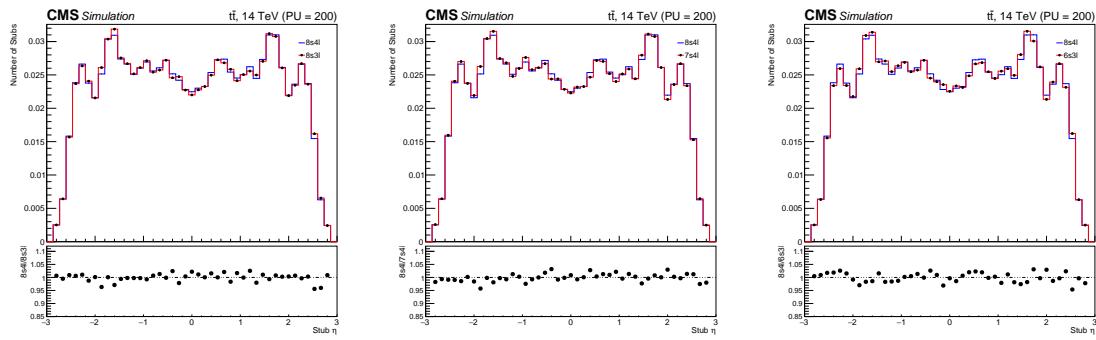


Figure 4.8: Comparison of total number of stubs per η value between the standard layout and the other three layouts considered. All four different layouts exhibit the same distribution. $8s4l$ vs $8s3l$ (left), $8s4l$ vs $7s4l$ (middle) and $8s4l$ vs $6s3l$ (right).

We further studied the effect of the pixel geometries on the stubs per barrel (Figure 4.9), end discs (Figure 4.10) of the OT and per ring of the endcap discs (Figure 4.11).

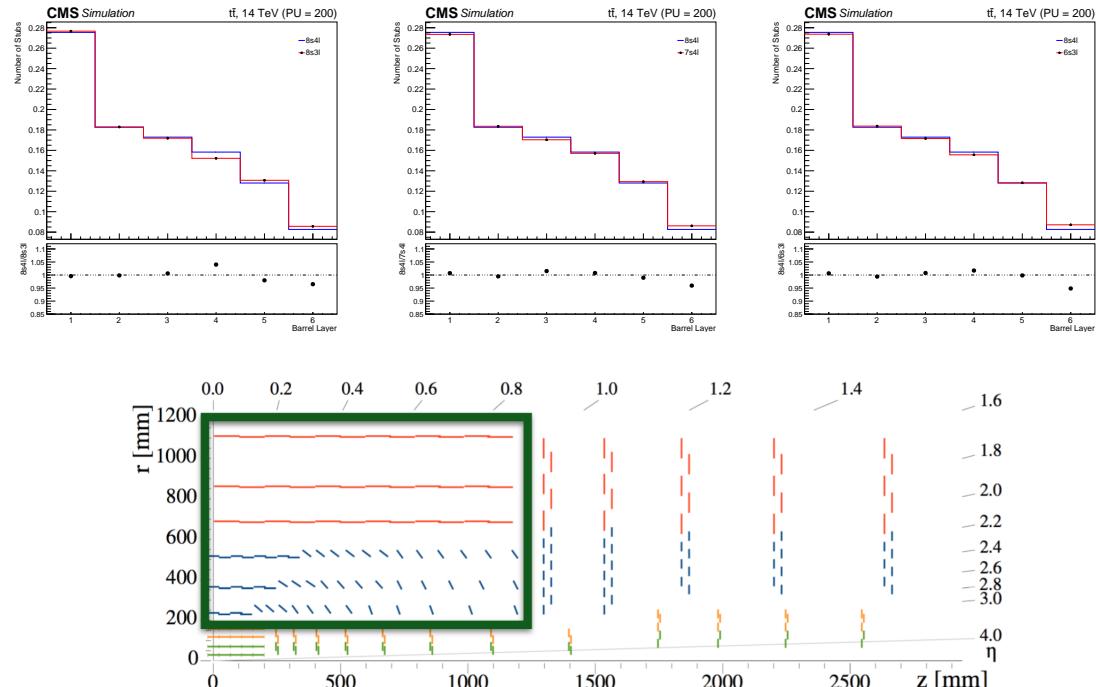


Figure 4.9: Comparison of total number of stubs per each of the 6 barrel layers. $8s4l$ vs $8s3l$ (left), $8s4l$ vs $7s4l$ (middle) and $8s4l$ vs $6s3l$ (right). Tracker Barrel layers shown below.

4.2. Tracker Stub Simulation

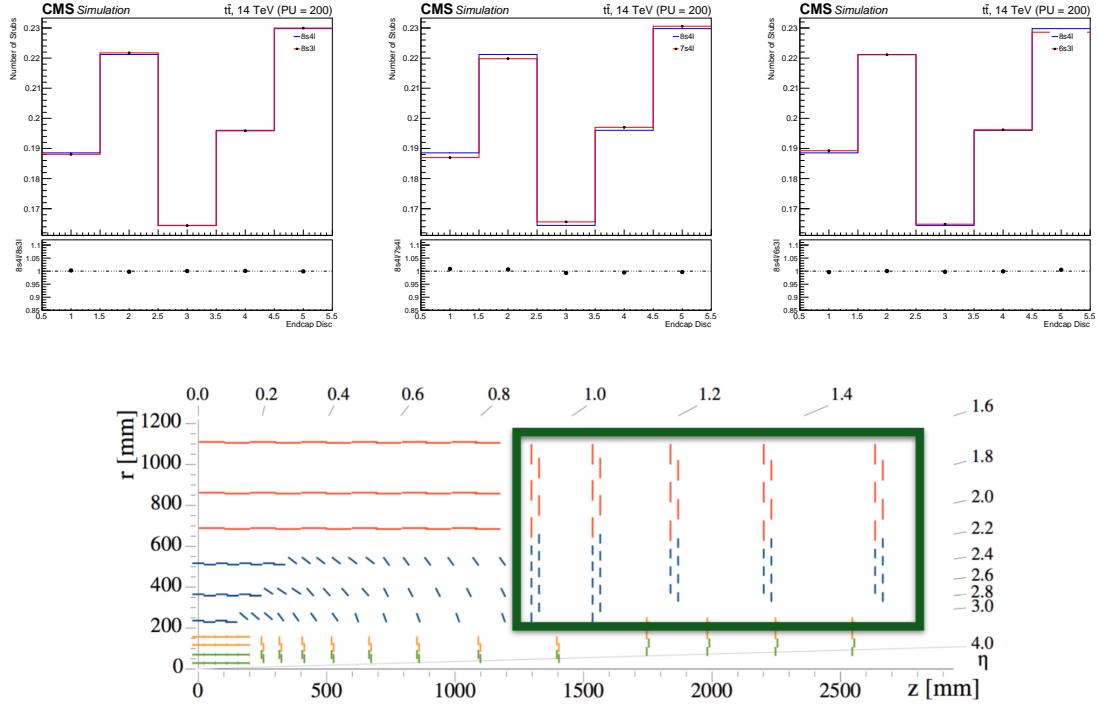


Figure 4.10: Comparison of total number of stubs per each of the 5 Endcap Double Discs. 8s4l vs 8s3l (left), 8s4l vs 7s4l (middle) and 8s4l vs 6s3l (right). Tracker Endcap Discs shown below.

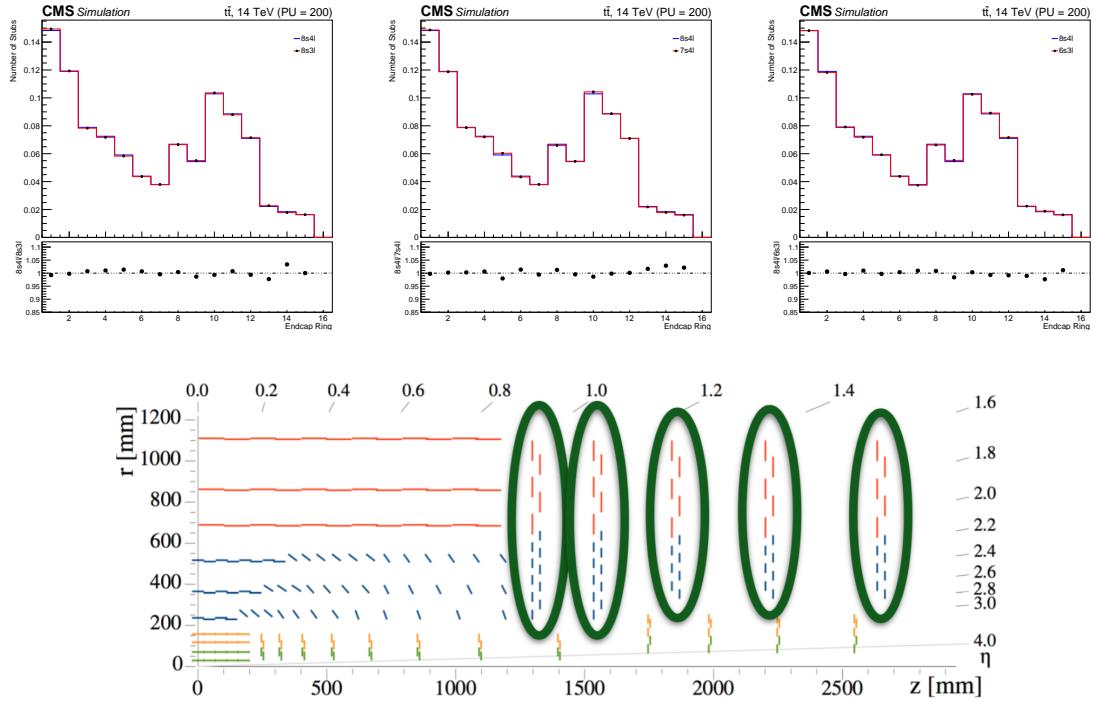


Figure 4.11: Comparison of total number of stubs for all rings across the five endcap double-discs. 8s4l vs 8s3l (left), 8s4l vs 7s4l (middle) and 8s4l vs 6s3l (right). The double discs are shown below.

1034 **4.3 Results**

1035 The above study shows that the removal of discs in the pixel detector affects the num-
1036 ber of stubs formed in the OT. A reduction in the amount of material (i.e. removal of
1037 discs) translates into a reduced amount of secondary particle interactions and hence, a de-
1038 crease in the net amount of stubs is observed. In other words more material (more discs)
1039 in the pixel layer leads to more secondaries, increasing the total amount of stubs in a given
1040 sample. However, the effect is small and a detailed study with respect to various η regions
1041 is planned in the future, as stub formation is also sensitive to η windows. Knowing the
1042 origin of particles causing stubs is also an important area to look into.

1043 **Chapter 5**

1044 **Search for SUSY in the 0-Lepton Final
1045 State**

1046 **5.1 Introduction**

1047 An overview of a search for Supersymmetry in the all-hadronic channel with missing
1048 transverse momentum and a customized top tagger is presented. The data used for the
1049 analysis was collected in 2016 by the CMS experiment at the LHC, and corresponds to
1050 an integrated luminosity of 35.9 fb^{-1} from proton-proton collisions at a center-of-mass
1051 energy (\sqrt{s}) of 13 TeV. The results and procedures presented are the product of an analysis
1052 conducted in collaboration with the hadronic SUSY analysis team based at Fermilab in
1053 Batavia, IL, USA. The published results can be obtained from [?].

1054 **5.2 Analysis Description**

1055 The study consists of an inclusive search for events with final states that contain miss-
1056 ing transverse momentum (p_T^{miss}) and reconstructed top quarks. The signal models used
1057 in this study include the production of three different types of SUSY particles. Two of
1058 which are the top squark (\tilde{t}) and the gluino (\tilde{g}), the supersymmetric partners of the SM top
1059 quark (t) and gluon (g). The third one is the neutralino ($\tilde{\chi}_1^0$), considered to be the lightest
1060 SUSY particle (LSP) under the Minimal Supersymmetric Standard Model (MSSM), and
1061 a possible candidate for Dark Matter.

1062 The signal data is selected by requiring events to have a minimum amount of jets (N_j)
1063 and b-jets (N_b) as well as a large p_T^{miss} . Other factors, such as the number of top jets
1064 (N_t), the scalar sum of the jet transverse momentum (H_T), and the so-called stransverse
1065 mass (m_{T2})¹ are also considered in order to select events of interest. The search region is
1066 then defined in exclusive bins of N_t , N_b , H_T , p_T^{miss} and m_{T2} . Additional discriminatory
1067 variables used in the analysis include the azimuthal angle between p_T^{miss} and the leading
1068 jets of an event ($\Delta\phi$), the pseudorapidity (η) and the cone radius size (defined as $\Delta R =$

¹ m_{T2} is a minimization of two transverse masses (m_T) with a constraint that the sum of the transverse momenta of both $\tilde{\chi}_1^0$'s is equal to the transverse momentum of the event, i.e., $\vec{p}_T = \vec{q}_T^{(1)} + \vec{q}_T^{(2)}$. The mathematical definition is given by $m_{T2} \equiv \min_{\vec{q}_T^{(1)} + \vec{q}_T^{(2)} = \vec{p}_T} \left[\max \left\{ m_T^2(\vec{p}_T^{(1)}, \vec{q}_T^{(1)}; m_{\tilde{\chi}_1^0}), (\vec{p}_T^{(2)}, \vec{q}_T^{(2)}; m_{\tilde{\chi}_1^0}) \right\} \right]$.

¹⁰⁷⁰ $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ used in the isolation of various physics objects.

¹⁰⁷¹ 5.2.1 Signal Models

¹⁰⁷² The signal models used in this analysis are based on two different scenarios: direct
¹⁰⁷³ top squark production and gluino mediated top squark production. In the former sce-
¹⁰⁷⁴ nario we consider one decay within the Simplified Model Spectra (SMS) [?] framework
¹⁰⁷⁵ called ‘T2tt’ (Figure 5.1, top left), where a top squark pair is produced directly from the
¹⁰⁷⁶ proton-proton collision and then decays into a pair of top quarks and a pair of neutralinos
¹⁰⁷⁷ ($\tilde{t} \rightarrow t\tilde{\chi}_1^0$).

¹⁰⁷⁸

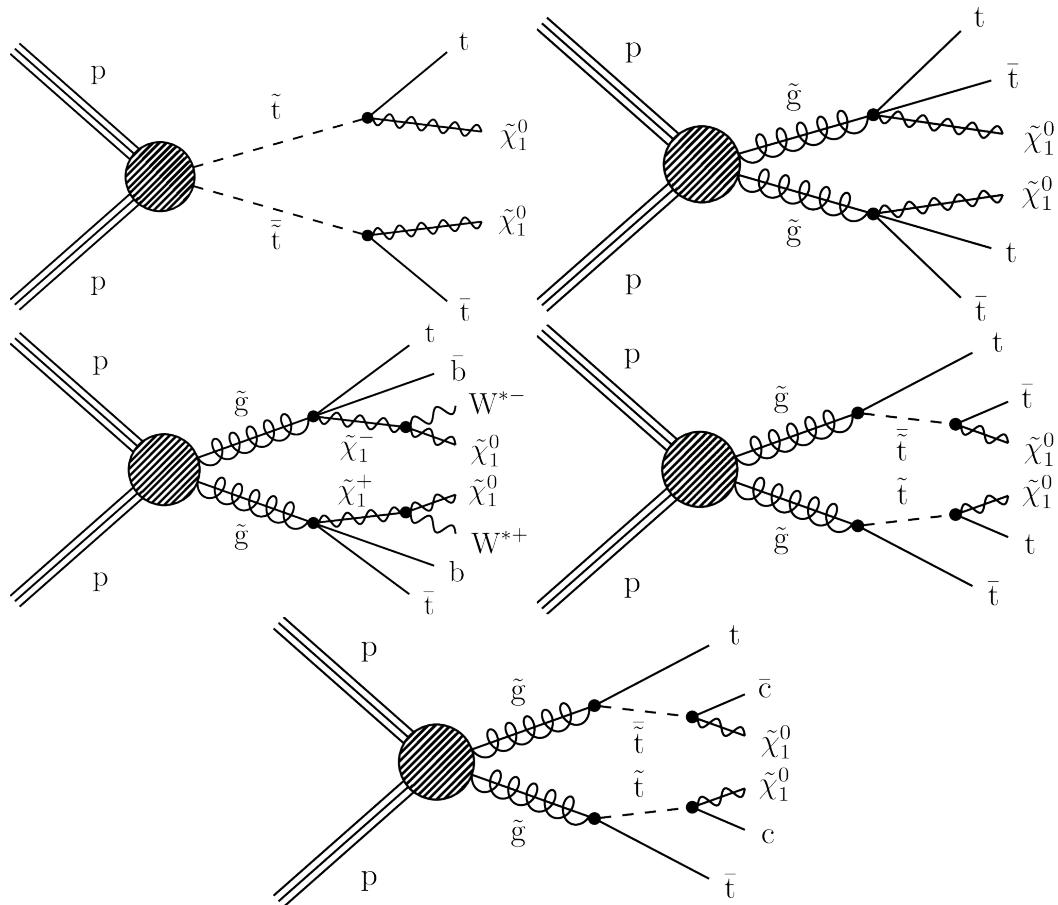


Figure 5.1: Simplified model diagrams show of the various Supersymmetry signals considered in this analysis: the T2tt model (top left), the T1tttt model (top right), the T1ttbb model (middle left), the T5tttt (middle right), and the T5ttcc model (bottom).

¹⁰⁷⁹ In the case of gluino mediated top squark production there are several different de-
¹⁰⁸⁰ cay modes considered. The T1tttt model (Figure 5.1, top right) consists of two gluinos
¹⁰⁸¹ decaying into a pair of top quarks and a neutralino ($\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$), accounting for situations
¹⁰⁸² in which the top squark is too heavy to be produced directly but the gluino is not. In the
¹⁰⁸³ T1ttbb model (Figure 5.1, middle left), a pair of gluinos to either a top or bottom squark
¹⁰⁸⁴ as $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$ (25%), $\tilde{g} \rightarrow b\bar{b}\tilde{\chi}_1^0$ (25%), or $\tilde{g} \rightarrow t\bar{b}\tilde{\chi}_1^\pm$ (50%), where $\tilde{\chi}_1^\pm$ denotes the lightest
¹⁰⁸⁵ positive ($\tilde{\chi}_1^+$) or negative ($\tilde{\chi}_1^-$) chargino. Due to the small difference in mass between the
¹⁰⁸⁶ $\tilde{\chi}_1^+$ (or its conjugate) and the LSP ($\Delta m(\tilde{\chi}_1^+, \tilde{\chi}_1^0) = 5$ GeV), the two particles are taken

1087 to be nearly mass degenerate. The T1ttbb model provides sensitivity to situations where
 1088 there are mixed states of top and bottom squarks.

1089

1090 The T5tttt model consists of a pair of gluinos, where each decays into a top quark and
 1091 an on-shell top squark. The top squark then decays into a top quark and a neutralino. For
 1092 this model, a mass difference of $\Delta m(\tilde{t}, \tilde{\chi}_1^0) = 175$ GeV between the top squark and the
 1093 LSP is expected. This model provides sensitivity to a region of T2tt where the $t\bar{t}$ back-
 1094 ground is very similar to the signal and $\Delta m(\tilde{t}, \tilde{\chi}_1^0)$ is very close to the top squark mass.
 1095 In the T5ttcc model (Figure 5.1, bottom) each gluino decays into a final state consisting
 1096 of a top quark, a charm quark and a neutralino ($\tilde{g} \rightarrow \bar{t}c\tilde{\chi}_1^0$). This model is very similar
 1097 to the T5tttt model with the difference that $\Delta m(\tilde{t}, \tilde{\chi}_1^0) = 20$ GeV and the on-shell top
 1098 squark decays into a charm quark and the LSP. The T5ttcc model provides sensitivity to
 1099 situations where the on-shell top squark is kinematically forbidden from decaying into a
 1100 top quark.

1101

1102 All five signal models described share similar final states containing two neutralinos
 1103 and up to four top quarks. Since the neutralino is considered stable and only interacts
 1104 weakly with matter, it cannot be picked up by the detector. Therefore, missing transverse
 1105 momentum (p_T^{miss}) is one of the most important variables used to compare event yields
 1106 between the signal and the SM background.

1107 5.2.2 Top Tagger

1108 The distinguishing feature of this analysis is its powerful top-tagging algorithm. It
 1109 is designed to provide high reconstruction efficiency over the full range of the top quark
 1110 p_T for the considered SUSY signal models. This top-tagger combines the use of sev-
 1111 eral different jet clustering algorithms in order to identify three different categories of
 1112 top quark jets, designated as “monojet”, “dijet” and “trijet”. The tagger makes use of
 1113 identified jets that were reconstructed with the anti- k_T [?] clustering algorithms combined
 1114 with additional corrections and selection criteria provided by the Cambridge–Aachen [?]
 1115 and soft-drop de-clustering [?] algorithms. In addition, multivariate analysis (MVA) tech-
 1116 niques, such as the random forest decision tree algorithm [?], were applied in order to
 1117 further decrease the amount of reconstructed fake tops.

1118

1119 There are three top-quark jet categories that take into account that top quark decay
 1120 products get closer together as the top quark p_T becomes higher. For specific p_T values,
 1121 the decay products of a hadronic process will be reconstructed either as one (“monojet”)
 1122 or two (“dijet”) jets rather than three (“trijet”). The p_T values for which this is observed
 1123 depends on the size of the jets that are being used. In the case of a highly boosted jet,
 1124 the anti- k_T algorithm is used within a cone size $\Delta R \sim 0.8$ (called AK8), which captures
 1125 all decay products of the top quark and is reconstructed as a single jet. This technique
 1126 requires that the top quark starts with a p_T of at least 400 GeV to have the decay products
 1127 fully captured in the 0.8 jet cone. For a $p_T < 400$ GeV, resolved top-tagging techniques
 1128 are more efficient, which require three jets independently clustered within a cone size of
 1129 $\Delta R \sim 0.4$ (AK4). Both types of algorithms are used in order to obtain a high efficiency
 1130 over a wide range of top quark p_T .

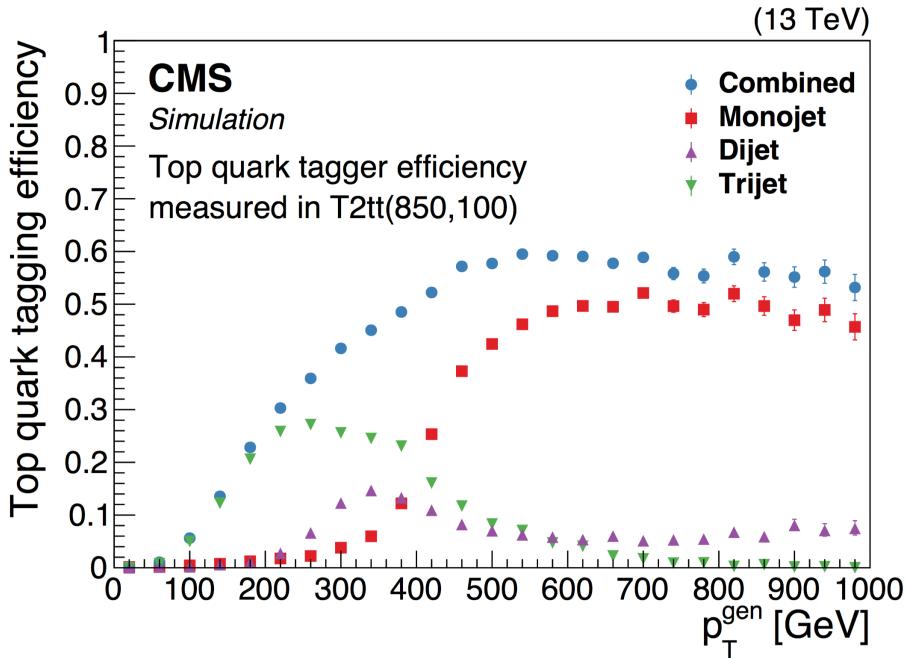


Figure 5.2: Efficiency of the top quark tagger as a function of generator-level top quark p_T for the monojet (red boxes), dijet (magenta triangles), and trijet (green upside-down triangles) categories and for their combination (blue circles), as determined using T2tt signal events with a top squark mass of 850 GeV and an LSP mass of 100 GeV. The vertical bars indicate the statistical uncertainties.

In the case of the dijet top quark decay category, an AK8 jet with a $p_T > 200$ GeV is combined with a loose AK4 jet to form a top quark candidate. Both jets are required to be within a cone of radius $\Delta R = 1$. For the trijet category, three AK4 jets are combined, with all three jets required to be within a radius of $\Delta R = 1.5$. [Figure 5.2](#) shows the efficiency of the algorithm for each of the three categories of top jets. The efficiency was determined for the T2tt signal model considering a top squark mass of 850 GeV and an LSP mass of 100 GeV, and is calculated as:

$$\text{Efficiency} = \frac{\text{Number of generator-level top quarks matched to a tagged top}}{\text{Number of generator-level top quarks in an event}}.$$

The top quark candidates are required to be within $|\eta| < 2.0$. Furthermore, a cone size of $\Delta R < 0.4$ is used to match the reconstructed top quark to the generated-level top quark. The average misidentification rate as a function of the p_T^{miss} is found to be around 20% for simulated $Z(\nu\bar{\nu})+jets$ events with a selection criteria similar to the one applied to data: $N_j \geq 4$, $N_b \geq 1$, $p_T^{miss} > 250$ GeV and no isolated electron or muon with $p_T > 10$ GeV.

5.2.3 Lepton and Track Veto

Since the search only involves models with purely hadronic final states, events which contain electrons or muons are vetoed. This requires the isolation of the muon/electron candidates to a cone size of $\Delta R < 0.2$ for $p_T \leq 50$ GeV and 0.05 for ≥ 200 GeV. This ΔR requirement decreases inversely to the lepton p_T in the range $50 < p_T < 200$ GeV in order to account for the collimation of a heavy object's decay products as its Lorentz boost increases. The electron and muon objects are considered to be isolated when their

relative isolation² is less than 0.1 and 0.2, respectively. Due to contributions from pileup, the isolation sum is corrected using an estimate of the amount of pileup energy in the cone.

The events that manage to pass the lepton veto are subjected to undergo an isolated charged-particle track veto. The purpose of this veto is to suppress events which contain τ leptons that decay hadronically or misidentified electrons and muons. The track requirements for this veto are $p_T > 5 \text{ GeV}$, $|\eta| < 2.5$ and a relative track isolation less than 0.2. In addition, the isolated track veto is only applied when the transverse mass of an isolated track- \bar{p}_T^{miss} is $m_T < 100 \text{ GeV}$, consistent with W boson decays. This veto successfully reduces the background of leptonically decaying W bosons by about 40%.

5.2.4 Baseline Event Selection

The data selection process begins with the triggers and follows with a pre-selection and the definition of the search bins. The selection criteria applied is optimized for high trigger efficiency as well as being sensitive to a variety of new-physics scenarios. The events selected from data meet the following baseline conditions:

1167

- 1168 • Satisfy the filters designed to remove detector and beam-related noise.
- 1169 • Undergo the lepton, isolated-track, and charged-hadron vetoes.
- 1170 • Have a final state with $N_j \geq 4$, $N_b \geq 1$, $N_t \geq 1$, $p_T^{\text{miss}} > 250 \text{ GeV}$ and $H_T > 300$ GeV.
- 1172 • $m_{T2} > 200 \text{ GeV}$, in order to reduce background contributions from $t\bar{t}$ events.
- 1173 • To reduce the background arising from the QCD multijet background the azimuthal angle between the p_T^{miss} and the three leading jets of an event is required to be $\Delta\phi(p_T^{\text{miss}}, j_{1,2,3}) > 0.5, 0.5, 0.3$, where j_1 , j_2 , and j_3 indicate the three leading jets in order of decreasing p_T .

5.2.5 Search Bin Definition

The search described in this chapter was performed by defining 84 non-overlapping search regions (Figure 5.3). Regions that contain events with $N_b \leq 2$ and $N_t \leq 2$ use N_b , N_t , p_T^{miss} and m_{T2} . In contrast, if the events contain $N_b \geq 3$ and $N_t \geq 3$, then N_b , N_t , p_T^{miss} and H_T are used for that region. The reason H_T is used for these regions instead of m_{T2} has to do with the fact that for events that contain many jets, the jets arising from the decay of specially heavy objects are not always correctly associated with said object, giving the m_{T2} variable a relatively broad and flat distribution. Therefore, in regions with $N_b \geq 3$ and $N_t \geq 3$, H_T is found to provide better discrimination between signal and background.

²The ratio between the isolation sum and the candidate p_T .

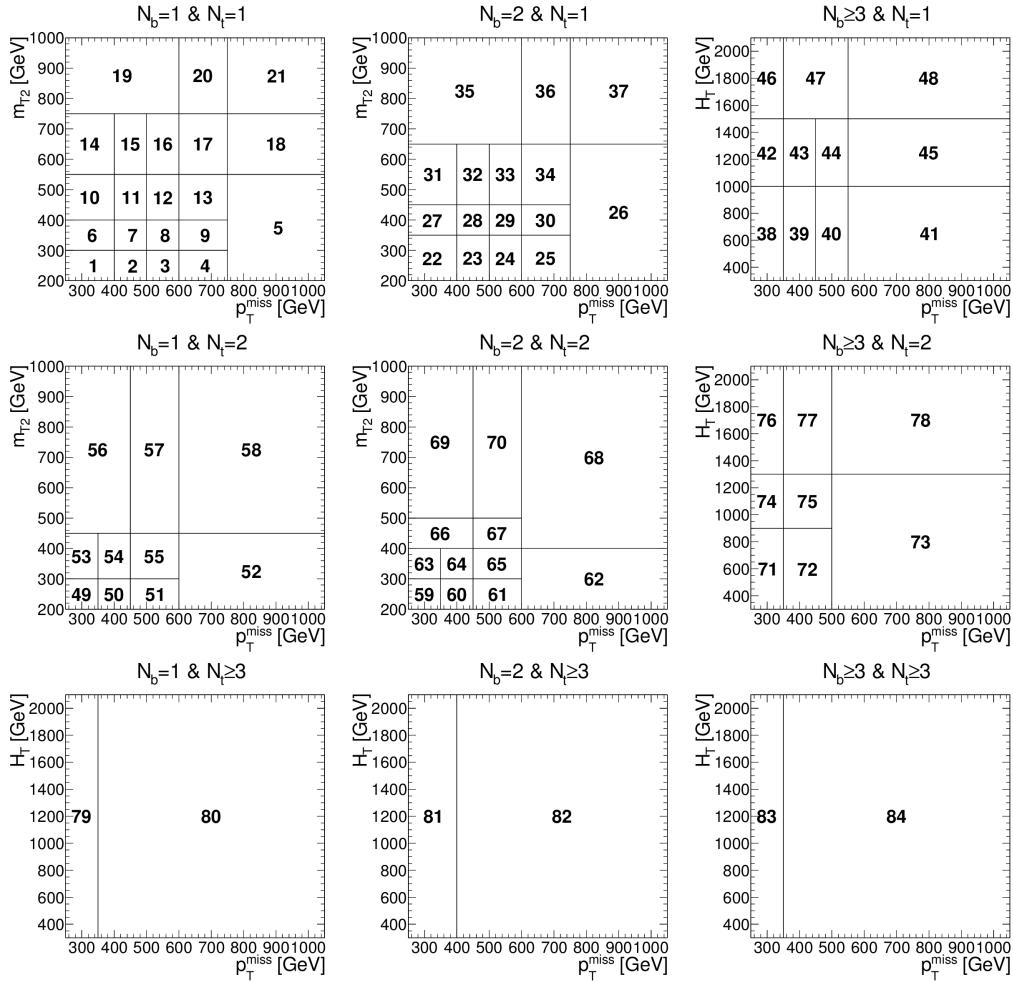


Figure 5.3: Search region definitions in the kinematic variables.

1187 5.3 Background Estimation

1188 Since there are many SM processes which exhibit the same characteristics as those of
 1189 the signal models (e.g. multiple jets, missing energy, etc.) it is of the utmost importance to
 1190 be able to properly account for their contributions. These interactions, which closely re-
 1191 semble the signal that's being searched for, are called the SM background. However, they
 1192 cannot be determined solely from the detector's reconstructed data itself, which is why
 1193 MC simulation is used to recreate them using theoretical predictions from the SM. These
 1194 simulations give us a good idea of what the signal region should look like if there were no
 1195 signal events, and therefore any significant excess when comparing the MC background to
 1196 data could signify a potential discovery. The relevant backgrounds for this particular anal-
 1197 ysis are: the lost-lepton background, the hadronic τ background, the $Z \rightarrow \nu\bar{\nu}$ background
 1198 and the QCD multijet background.

1199 5.3.1 Background from $t\bar{t}$, Single Top Quark and W+jets Events

1200 The vast majority of the expected SM background (around 70%) is due to $t\bar{t}$ and
 1201 W+jets events where the W decays into a lepton and a neutrino. Since the expected signal
 1202 is purely hadronic, a portion of the leptonic and semi-leptonic decays are vetoed and do

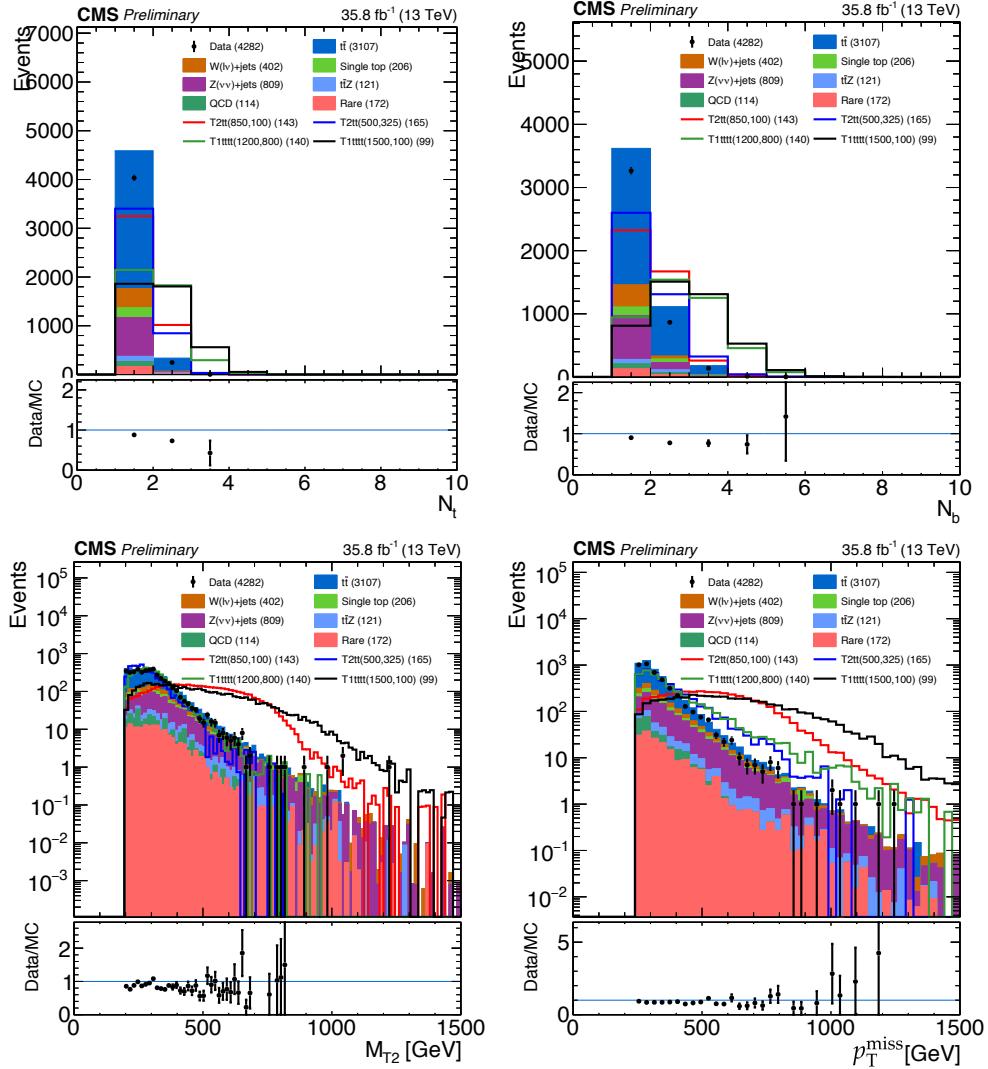


Figure 5.4: Comparison between the total SM background from simulation and CMS data for N_t (top left), N_b (top right), m_{T2} (bottom left) and p_T^{miss} (bottom right). Total SM backgrounds and signals are scaled to same data yield for a shape comparison.

not form part of the total SM background. However, there are two different scenarios in which the lepton vetoes can be satisfied. For instance, when the W decays into a τ which decays hadronically (τ_h), the τ gets reconstructed as a jet and passes the veto. If, on the other hand, the W decays into an electron or a muon, the veto is satisfied when the corresponding lepton is said to be “lost”. In other words, the lepton veto fails to reject events when light leptons (electrons and muons) are not isolated, not identified/reconstructed, or are out of the acceptance region. Both scenarios are evaluated together with a single-lepton data control sample (CS), subjected to the same trigger used for signal events.

1211

1212 The total predicted amount of lost-lepton and τ_h events in any given search region is
 1213 defined by the net sum of single-electron and single-muon events in the respective CS,
 1214 corrected by a translation factor that is determined from simulation. Due to differences
 1215 in how they are detected, the single-electron and single-muon samples are determined
 1216 separately. The translation factor used to correct the search bins is defined as the ratio
 1217 of simulated τ_h and lost-lepton events in the search region over the number of simulated

1218 single-electron or single-muon events in the corresponding CS region.

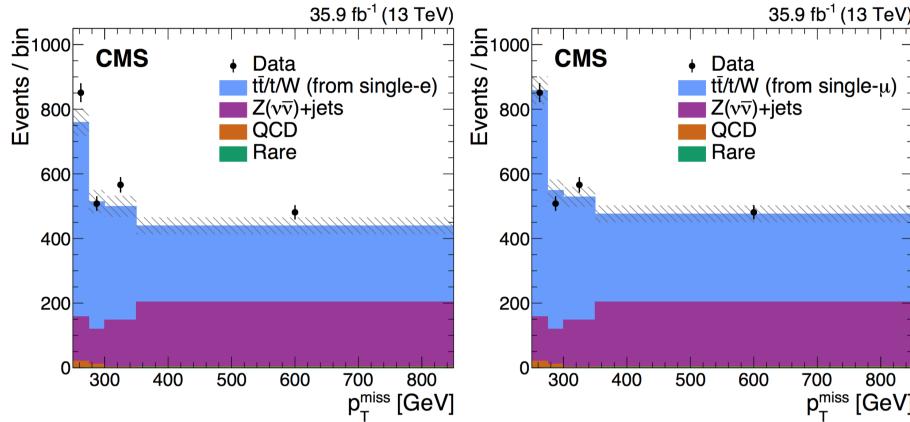


Figure 5.5: Distribution of p_T^{miss} for both the single-electron (left) and single-muon (right) SB data samples compared to predictions for SM processes. The hatched bands indicate the statistical uncertainties in the total SM prediction. Note that the data and the predictions for all backgrounds except that for $t\bar{t}$, single top quark, and $W+jets$ events are identical between the left and right plots.

1219 In order to test this method, a sideband (SB) region is selected which is orthogonal
 1220 to the search region. This region is defined by the same selection criteria applied to data
 1221 except for $N_t = 0$, $N_b \geq 2$, and $\Delta\phi(p_T^{miss}, j_{1,2,3,4}) > 0.5$, where the last two require-
 1222 ments are applied in order to suppress contributions from $Z(\nu\bar{\nu})+jets$ and QCD multijet
 1223 processes. The SB is divided into four intervals of p_T^{miss} , where the contribution of lost-
 1224 lepton and τ_h to the corresponding intervals are determined by multiplying the appropriate
 1225 single -electron and single-muon CS (Figure 5.5) by a translation factor from simulation.
 1226 The contribution to the SB from other backgrounds, such as $Z+jets$, QCD multijet and
 1227 rare events, is estimated directly from simulation. The total SM background prediction is
 1228 found to agree with the data within uncertainty, confirming the validity of the translation
 1229 factor procedure.

1230

1231 Systematic uncertainties stemming from the prediction of $t\bar{t}$, single top quark and
 1232 $W+jets$ background events are determined from various sources: The statistical uncer-
 1233 tainty associated to the translation factors determined for each search region (1–40%),
 1234 the lepton reconstruction and isolation efficiency (7–43%), the jet and p_T^{miss} energy scale
 1235 and resolution (maximum of 64%), the ISR modeling (maximum of 13%), the PDF's
 1236 (maximum of 32%) and the b-jet tagging efficiency (1%).

1237 5.3.2 The $Z \rightarrow \nu\bar{\nu}$ Background

1238 The $Z \rightarrow \nu\bar{\nu}$ background is derived using simulated $Z \rightarrow \nu\bar{\nu}$ events that have been
 1239 corrected for observed differences between data and simulation. In order to correctly esti-
 1240 mate this background two scale factors are used to weigh the simulated events: R_{norm} and
 1241 $S_{DY}(N_j)$ which correct the normalization of the simulation and the shape of the simulated
 1242 N_j distribution, respectively. These scale factors are calculated from the dimuon control
 1243 region which includes events with two muons and no muon or isolated track veto. In the
 1244 region where $81 < m < 101$ GeV, the two muons are treated as if they were neutrinos.

1245

The first scale factor, R_{norm} , is computed by comparing the expected event yield in the dimuon control region from the Drell-Yan (DY) simulation with the observed event yield in data after subtraction of the other SM processes. The second scale factor, S_{DY} , depends on the number of jets (N_j) in the event and is also computed from the dimuon control region in which the signal region requirements on p_T^{miss} , N_t and m_{T2} are removed, and the H_T requirement is relaxed to $H_T > 200$ GeV. For each N_j bin this scale factor is calculated from the ratio between data and the DY simulation. Its value ranges between 0.6 and 1.1, depending on the N_j bin.

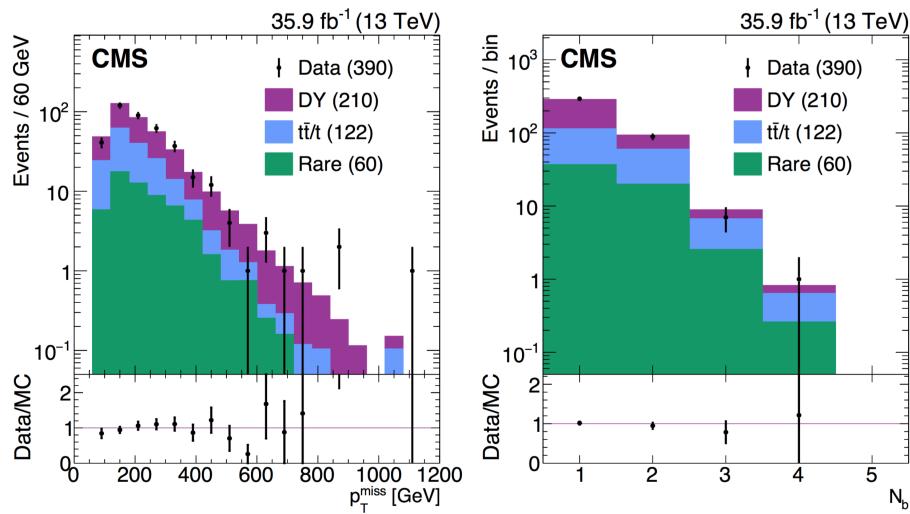


Figure 5.6: The p_T^{miss} (left) and N_b (right) distributions of data and simulation in the loose dimuon CS after applying both the normalization and shape correction factors. The lower panels show the ratio between data and simulation. Error bars in the plot account only for statistical uncertainties. The values in parentheses indicate the integrated yields for each component.

The systematic uncertainties for the $Z(\nu\bar{\nu})+j$ ets background are obtained from the shape differences between data and simulation in the loose dimuon CS in terms of the N_b , N_t , p_T^{miss} , m_{T2} and H_T after the normalization factor (R_{norm}) has been applied. These include the statistical uncertainty in the N_j shape correction (1–46%) and in the overall normalization correction (7.6%). Additional systematic uncertainties account for the jet and p_T^{miss} energy scales (1–71%), the b tagging efficiency (1–23%), the PDFs and the renormalization and factorization scales (1–48%), the statistical uncertainty in the simulation (1–81%, with some search regions as high as 100%), and the trigger (up to 14%). An additional uncertainty is defined from the shift in the central value between the data and simulation in the distributions (14–44% depending on the search region).

5.3.3 The QCD Multijet Background

To estimate the QCD Multijet background a signal-depleted, QCD multijet-rich data control sample is used. This control sample is defined by inverting the preselection requirements on $\Delta\phi(p_T^{miss}, j_{1,2,3})$ and subtracting contributions of other SM backgrounds, such as $t\bar{t}$, $W+jets$, and $Z+jets$. For $t\bar{t}$ and $W+Jets$ the same methods (lost-lepton and hadronic τ) are used to estimate the contributions for this QCD multijet-enriched control region. In the case of the $Z \rightarrow \nu\bar{\nu}$ background, simulation is used for its estimation due to its small number. Following that, a translation factor, partly determined by data and partly

by simulation, is used to convert the number of QCD multijet events measured in the data control region into a QCD multijet prediction for each search region bin. This translation factor, called T_{QCD} , is computed as the simulated ratio between the signal region and the inverted- $\Delta\phi$ control region, in bins of p_T^{miss} and m_{T2} where the bins are bounded in the same way as the signal bins.

1277

The systematic uncertainty associated with the QCD multijet background for each search region is obtained as the difference between the event yield from simulation of QCD multijet processes and the prediction obtained by applying the background prediction procedure to simulated QCD multijet samples (30–500%). Other sources include the statistical uncertainty in the translation factors (30–300%) and the subtraction of the non-QCD-multijet SM contributions to the QCD control sample (2–50%).

1284

5.3.4 Background From Rare and Other Processes

1285

The background contribution from rare processes forms only a tiny fraction of the overall SM background and has therefore a small effect on the final result. The estimation for these processes are determined directly from simulation, with the biggest contribution coming from $t\bar{t}Z$. With the exception of the $t\bar{t}Z$ process, all other remaining backgrounds are combined. The comparison between simulation and the obtained data were found to be in agreement within a statistical uncertainty of 30%, which is attributed to the systematic uncertainty in the estimation of the $t\bar{t}Z$ background.

1292

5.4 Results and Interpretation

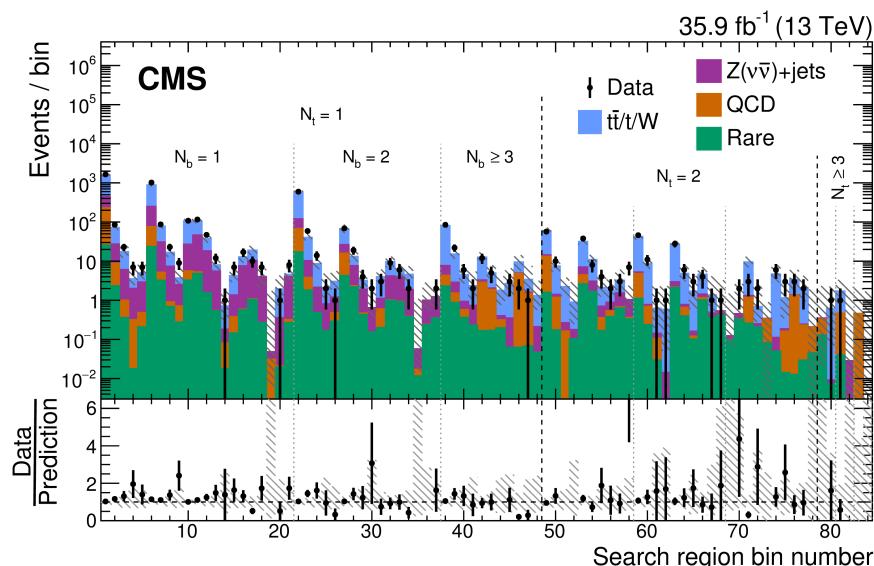


Figure 5.7: Observed events compared to the corrected SM background predictions for all 84 search regions in the full data of 35.9 fb^{-1} collected in 2016. The lower panel shows the ratio between the data and the total SM simulation. The gray bands show the total uncertainty related to the background prediction.

1293

Figure 5.7 shows the summary of all observed events compared to the corrected SM background simulation for all of the 84 search bins. The data obtained shows no statisti-

1294

1295 cally significant deviation from the predicted SM background. The biggest contribution
 1296 to the background is attributed to the $t\bar{t}$ and $W+jets$ processes, followed by $Z(\nu\bar{\nu})+jets$,
 1297 which could be dominant in regions that have a high p_T threshold. The contributions arising
 1298 from the QCD multijet and rare backgrounds are found to be nearly negligible in all
 1299 of the search regions.

1300

1301 Exclusion limits are calculated for each of the signal models discussed in this chapter
 1302 by applying a binned likelihood fit on the data. The likelihood function is obtained for
 1303 each of the 84 search regions as well as for each of the background data control samples
 1304 (single-electron, single-muon, and QCD) from the product of the Poisson probability den-
 1305 sity function. Exclusion limits were placed on the top squark, gluino and LSP production
 1306 cross-sections with a 95% confidence level (CL), calculated using a modified frequentist
 1307 approach with the CL_s criterion [?, ?] and asymptotic results for the test statistic [?].

1308

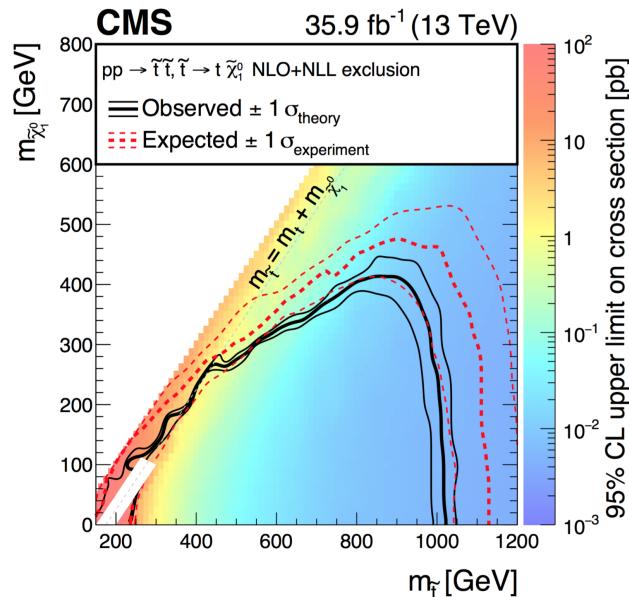


Figure 5.8: The 95% CL upper limit on the production cross section of the T2tt simplified model as a function of the top squark and LSP masses. The solid black curves represent the observed exclusion contour with respect to NLO+NLL [?] signal cross sections and the change in this contour due to variation of these cross sections within their theoretical uncertainties. The dashed red curves indicate the mean expected exclusion contour and the region containing 68% of the distribution of expected exclusion limits under the background-only hypothesis.

1309

1310 The uncertainties from the search region modeling are taken into account for each of
 1311 the 84 bins and arise from several different sources: the statistical uncertainty in the sim-
 1312 ulated event samples, the integrated luminosity (2.5% [?]), the lepton and isolated-track
 1313 veto efficiencies (up to 6.8%), the b tagging efficiency (up to 21%), the trigger efficiency
 1314 (up to 2.6%), the renormalization and factorization scales (up to 3.5%), the ISR modeling
 1315 (up to 46%), the jet energy scale corrections (up to 34%), the top quark reconstruction
 1316 efficiency (up to 14%), and the modeling of the fast simulation compared with the full
 1317 simulation for top quark reconstruction and mis-tagging (up to 24%). All of the uncer-
 1318 tainties, with the exception of the statistical precision of the simulation, are taken to be
 fully correlated between the search regions. Contributions from signal contamination in

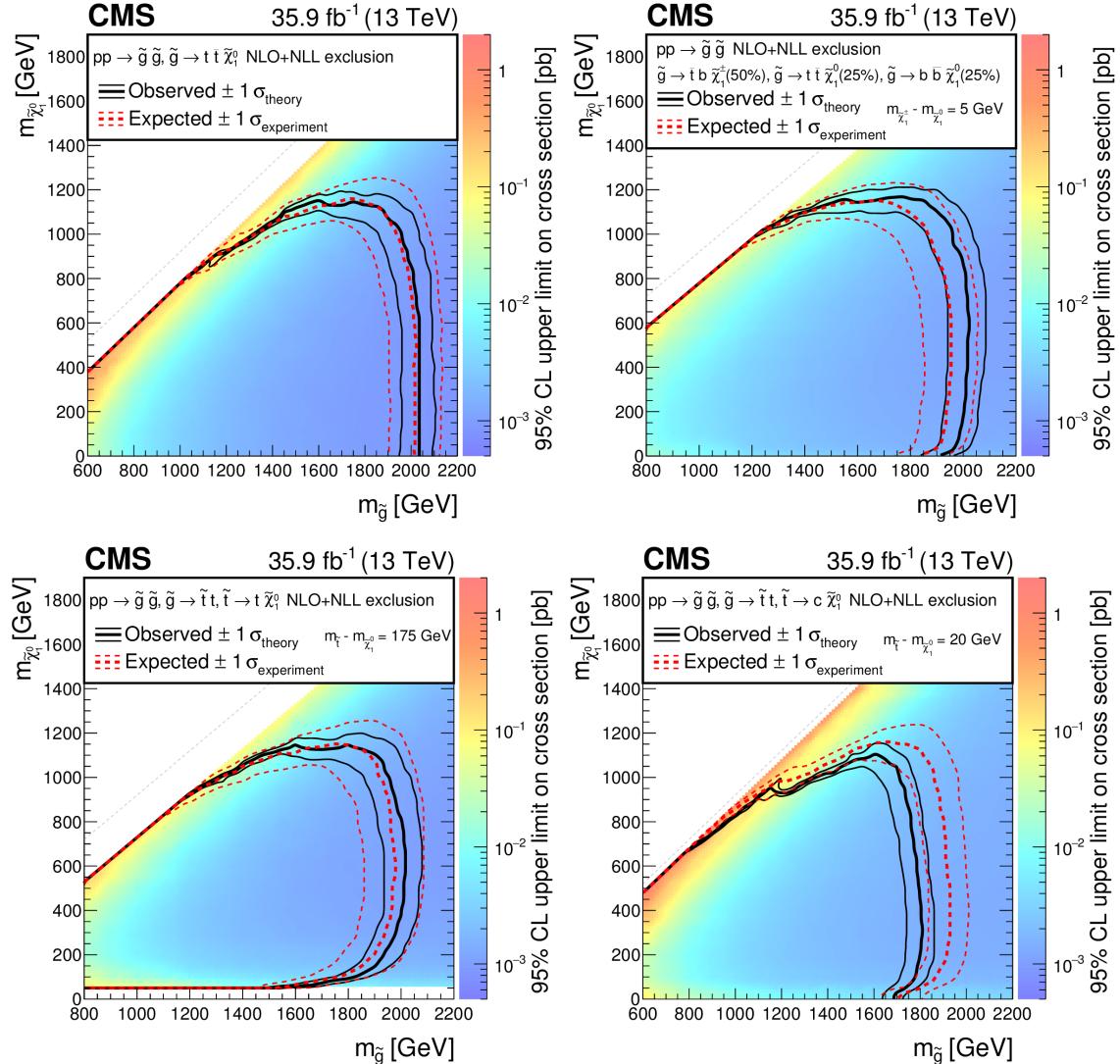


Figure 5.9: The 95% CL upper limit on the production cross section of the T1tttt (upper left), T1ttbb (upper right), T5tttt (bottom left), and T5ttcc (bottom right) simplified models as a function of the gluino and LSP masses.

the signal modeling are found to be only significant from the single-lepton control samples and negligible for the rest.

1321

1322 The 95% CL exclusion limits obtained for the T2tt model (Figure 5.8), which consists
 1323 of direct top squark production, excludes top squark masses up to 1020 GeV and LSP
 1324 masses up to 430 GeV. Meanwhile, Figure 5.9 shows the results for the gluino pair pro-
 1325 duction models: T1tttt, T1ttbb, T5tttt, and T5ttcc. For the T1tttt model, gluino masses of
 1326 up to 2040 GeV and LSP masses up to 1150 GeV are excluded, with corresponding limits
 1327 of 2020 and 1150 GeV for the T1ttbb model, 2020 and 1150 GeV for the T5tttt model,
 1328 and 1810 and 1100 GeV for the T5ttcc model.

1329 **Chapter 6**

1330 **Estimation of the $Z \rightarrow \nu\bar{\nu}$ Background**

1331 **6.1 Introduction**

1332 A detailed explanation for the estimation of the $Z \rightarrow \nu\bar{\nu}$ +jets background is presented
1333 in this chapter. The following estimation procedure builds upon the 2016 SUSY analysis
1334 summarized in [chapter 5](#) and aims to refine the overall background calculation as well
1335 as reduce the uncertainties associated with the previous method. To accomplish this, an
1336 additional γ +jets CS is used in conjunction with the tight $Z \rightarrow \mu^+\mu^-$ control region used in
1337 the 2016 analysis estimation. The new γ +jets CS provides a more data-driven estimation
1338 procedure with the added benefit of a substantially larger production cross-section than
1339 Z +jets processes [?].

1340 **6.2 The Irreducible $Z \rightarrow \nu\bar{\nu}$ Background**

1341 An important source of background in searches for SUSY in the 0-lepton final state
1342 comes from events in which a Z boson, accompanied by jets, decays into a pair of neu-
1343 trinos ($Z \rightarrow \nu\bar{\nu}$ +jets). The resulting final state is comprised of a large p_T^{miss} (from the
1344 neutrino pair) and multiple hadron jets, closely mimicking the SUSY signal. For similar
1345 searches, the $Z \rightarrow \nu\bar{\nu}$ contribution can make up a large portion of the background in many
1346 of the search bin regions (higher than 50% in some regions) [?, ?], and can compose up
1347 to about a third of the total SM background. For this particular analysis, the $Z \rightarrow \nu\bar{\nu}$ ac-
1348 counts for about 17% of the total SM background and owes its low value to the dedicated
1349 top-tagging algorithm outlined in [subsection 5.2.2](#).

1350 There are several different methods that have been developed to estimate the $Z \rightarrow$
1351 $\nu\bar{\nu}$ +jets background [?]. Two of the commonly used methods involve the use of a control
1352 region dominated by $Z \rightarrow ll$ +jets, where the l stands for lepton (either a muon or an
1353 electron, in this case) or γ +jets events. The $Z \rightarrow ll$ channel has the advantage of having
1354 very similar kinematics to the $Z \rightarrow \nu\bar{\nu}$ region but suffers from low statistics (due in part
1355 to its small branching ratio), specially in the tight search regions used in typical SUSY
1356 searches. On the other hand, the γ +jets region has a much higher production cross-section
1357 but involves a completely different process. The hybrid method described in this chapter
1358 makes use of both control regions in order to estimate the $Z \rightarrow \nu\bar{\nu}$ background corrections,
1359 and aims to improve on the results of the 2016 method described in [subsection 5.3.2](#).



Figure 6.1: Leading-order Feynman diagrams for $Z + \text{jets}$ and $\gamma + \text{jets}$ processes. The ‘V’ in the figure can represent either Z or γ .

1361 6.3 The Loose $\gamma + \text{jets}$ Control Region

1362 The loose $\gamma + \text{jets}$ control region, as well as the photon ID/isolation selection, is de-
 1363 scribed in this section. The $\gamma + \text{jets}$ control region is chosen with the purpose of substitut-
 1364 ing the loose muon control region used in the 2016 version of this analysis for the calcu-
 1365 lation of the shape correction scale factors applied to the final estimation of the $Z \rightarrow \nu\bar{\nu}$
 1366 background. The $\gamma + \text{jets}$ control region is presumed to be better suited for this estima-
 1367 tion due to having a much higher cross-section and therefore, an expected reduction in
 1368 the statistical and systematic uncertainties associated to the shape correction factors are
 1369 expected.

1370 6.3.1 Photon ID and Isolation

1371 Three different working points are provided by the CMS EGM physics object group
 1372 (POG) for simple cut-based photon identification [?]. The three working points, called
 1373 loose, medium and tight, are chosen according to the requirements of the particular analy-
 1374 sis and differ on the amount of background rejection they offer, as well as on their average
 1375 photon selection efficiency. The higher the efficiency of a given ID, the lower the amount
 1376 of background that is rejected, and vice-versa. Table 6.1 shows the cut values that are
 1377 applied to photons that are found within both the ECAL barrel and endcap range. The
 1378 associated values to the efficiency and the background rejection rate are shown for each
 1379 of the three different photon ID selections.

1380 In order to obtain the high efficiency and background rejection rates shown, a robust
 1381 set of identification and isolation criteria are selected. A total of five parameters are
 1382 used for this simple cut based method. For photon identification, the H/E and the $\sigma I\eta I\eta$
 1383 variables are found to provide the best results. The H/E parameter is defined as the ratio
 1384 of the HCAL tower energy over the ECAL seed cluster energy. A threshold value is
 1385 selected on H/E to remove background from electrons that are detected in both the ECAL
 1386 and HCAL but have no reconstructed track [?]. The $\sigma I\eta I\eta$ variable is known as the
 1387 photon shower shape variable and is defined in the ECAL as the energy weighted standard
 1388 deviation of a single crystal within the 5×5 crystal η range, centered around the crystal
 1389 with maximum energy [?]. This variable is a key component in the identification of both
 1390 electrons and photons since it provides a measure of the shower width where most of the
 1391 energy has been deposited in a given ECAL crystal.

Table 6.1: Identification and isolation cut values for photons provided by the CMS EGM POG. Values are provided for the three working points described (loose, medium and tight) for both the ECAL barrel and endcaps. The photon selection efficiency of the three working points, as well as their associated background rejection rate, are provided.

Barrel	Loose (90.06%)	Medium (80.19%)	Tight (70.01%)
Background Rejection	Loose (85.73%)	Medium (88.87%)	Tight (90.66%)
H/E	0.105	0.035	0.020
$\sigma I\eta I\eta$	0.0103	0.0103	0.0103
ρ -corrected PF charged hadron isolation	2.839	1.416	1.158
ρ -corrected PF neutral hadron isolation	$9.188 + 0.0126 \cdot p_T^\gamma + 0.000026 \cdot p_T^{\gamma^2}$	$2.491 + 0.0126 \cdot p_T^\gamma + 0.000026 \cdot p_T^{\gamma^2}$	$1.267 + 0.0126 \cdot p_T^\gamma + 0.000026 \cdot p_T^{\gamma^2}$
ρ -corrected PF photon isolation	$2.956 + 0.0035 \cdot p_T^\gamma$	$2.952 + 0.0040 \cdot p_T^\gamma$	$2.065 + 0.0035 \cdot p_T^\gamma$

End Cap	Loose (90.81%)	Medium (80.06%)	Tight (70.11%)
Background Rejection	Loose (76.90%)	Medium (81.50%)	Tight (84.34%)
H/E	0.029	0.027	0.025
$\sigma I\eta I\eta$	0.0276	0.0271	0.0271
ρ -corrected PF charged hadron isolation	2.150	1.012	0.575
ρ -corrected PF neutral hadron isolation	$10.471 + 0.0119 \cdot p_T^\gamma + 0.000025 \cdot p_T^{\gamma^2}$	$9.131 + 0.0119 \cdot p_T^\gamma + 0.000025 \cdot p_T^{\gamma^2}$	$8.916 + 0.0119 \cdot p_T^\gamma + 0.000025 \cdot p_T^{\gamma^2}$
ρ -corrected PF photon isolation	$4.895 + 0.0040 \cdot p_T^\gamma$	$4.095 + 0.0040 \cdot p_T^\gamma$	$3.272 + 0.0040 \cdot p_T^\gamma$

1393 The other three parameters considered, comprise the isolation portion of the photon
 1394 selection cuts. These are the ρ -corrected particle flow (PF) charged hadron, neutral hadron
 1395 and photon isolation parameters. As can be seen from Table 6.1, two of these parameters
 1396 (the neutral hadron and photon isolation) have a dependence on p_T^γ . These cuts are used
 1397 to ensure that the identified photon is well isolated within its own cone and by rejecting
 1398 photons that are identified within close proximity to either a charged or a neutral hadron
 1399 [?]. The value ρ included in the name of each of these parameters refers to the total
 1400 pileup density [?]. Therefore, the term “ ρ -corrected” implies that these values, which are
 1401 sensitive to the residual contamination that arises from pile-up, have been corrected to
 1402 include these contributions.

1403 6.3.2 Photon Selection

1404 The event selection process for the γ +jets control region starts with photon candidates
 1405 that have a $p_T > 200$ GeV and are within the acceptance range of the CMS ECAL (given
 1406 by $|\eta| < 1.4442$ for the barrel and $1.566 < |\eta| < 2.5$ for the endcaps). The photons
 1407 are subjected to pass the loose ID/isolation cuts described in subsection 6.3.1 in order to
 1408 remove $\sim 85\%$ of the background processes and obtain a prompt photon sample that is
 1409 $\sim 90\%$ pure, on average. Additional restrictions include some of the same requirements
 1410 imposed on the signal baseline selection discussed in subsection 5.2.4. These include
 1411 $N_j \geq 4$, an $H_T > 300$ GeV, the $\Delta\phi$ requirements for leading jets and the lepton vetoes
 1412 described in subsection 5.2.3. The lepton veto, in particular, greatly improves the prompt
 1413 photon selection by removing many of the events in the simulated samples where a lepton
 1414 gets misidentified as a photon.

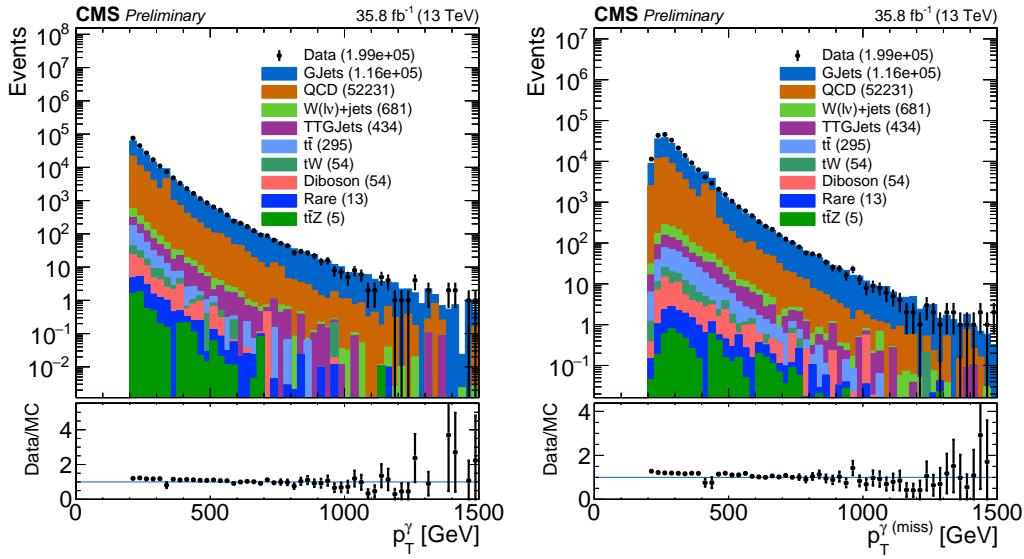


Figure 6.2: Shown are both the p_T^γ (left) and $p_T^{\gamma(\text{miss})}$ (right) distributions before applying any corrections. $p_T^{\gamma(\text{miss})}$ is obtained by adding the p_T^γ to the total p_T^{miss} in every event.

To further emulate the $Z \rightarrow \nu\bar{\nu} + \text{jets}$ background, a variable in which the photons are treated as p_T^{miss} is defined. We call this variable $p_T^{\gamma(\text{miss})}$ and we obtain it by adding the p_T^γ for every event to the total p_T^{miss} in the event. Both the p_T^γ and the resulting $p_T^{\gamma(\text{miss})}$ distributions are shown in Figure 6.2 as data/MC comparison plots, where the simulated backgrounds are stacked in order of ascending contribution.

The main contributions from simulation arise from the $\gamma + \text{jets}$, QCD and to a lesser extent, $t\bar{t}\gamma$. Other non-dominant backgrounds in the control region include contributions from $W(l\nu) + \text{jets}$, $t\bar{t}$, Diboson, tW , $t\bar{t}Z$ and rare processes. Most of these lesser backgrounds are nearly negligible (several orders of magnitude lower than the dominant backgrounds) and are considered to be mostly composed of fake photons. In addition to the cuts described, all of the simulation samples are subjected to weights that apply corrections to pileup as well as the b-tagging efficiency. Data, on the other hand, is obtained from a sample that contains events with at least one identified photon. Photons in this sample are also subjected to the high-level trigger HLT_Photon175, which restricts the selection to photons that have a $p_T > 175$ GeV. Both simulation and data are subjected to the same selection criteria established in this section.

6.3.3 Photon Purity and Fake Rate

Three different types of photons make up the $\gamma + \text{jets}$ CS: prompt photons, produced either directly or through fragmentation, and fake photons. Prompt photons are defined as photons which are formed shortly after the proton-proton collision (i.e. before the produced quarks and gluons have had enough time to form hadrons). Two types of photons fit in this category. The first type, which we designate as direct photons, are photons that are produced directly from the proton-proton interaction [?]. A secondary type of prompt photon, that is virtually indistinguishable from the direct photons at the detector level, originates from the decay of π^0 mesons and are called fragmentation photons. The final type of photon found in the CS corresponds to fake (or non-prompt) photons. The fake

¹⁴⁴² photon contribution typically arises from leptons (mostly electrons) whose tracks are not
¹⁴⁴³ properly reconstructed, yet leave energy measurements in the ECAL.

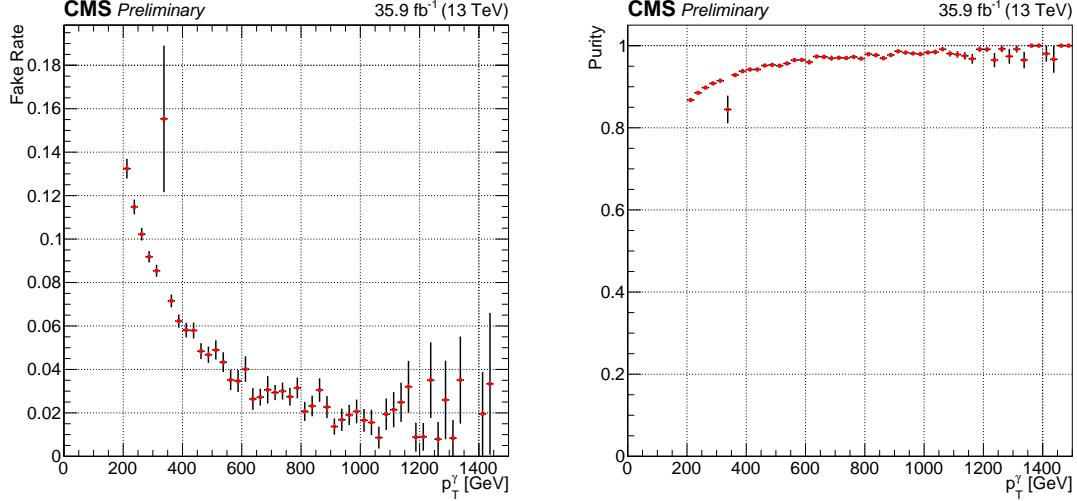


Figure 6.3: Plots for Fake Rate (left) and Purity (right) as a function of the photon p_T are shown. The events are selected are required to have a $p_T > 200$ GeV, be within the ECAL acceptance range, and pass the loose ID selection cuts. This selection was produced in order to verify the values given by the E/ γ POG. As can be seen, the efficiency (purity) is seen to agree with the values of the loose photon ID/isolation selection.

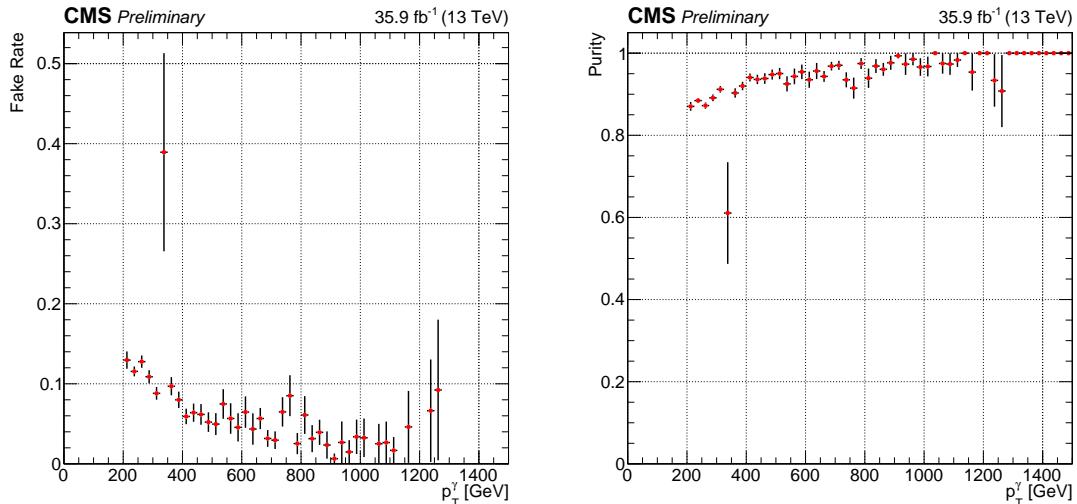


Figure 6.4: Plots for Fake Rate (left) and Purity (right) as a function of the photon p_T are shown. These plots include photons with the full control region selection. Aside from exhibiting lower statistics, the plots seem to agree with the fake rate and purity before all the control region cuts are applied.

¹⁴⁴⁴ In order to identify prompt photons, reconstructed photons from the $\gamma+jets$ and QCD
¹⁴⁴⁵ samples are matched to generator-level photons in space and momentum by requiring
¹⁴⁴⁶ $\Delta R(\gamma_{\text{gen}}, \gamma_{\text{reco}}) < 0.4$ and $0.5 < p_T^{\text{gen}}/p_T^{\text{reco}} < 2.0$, respectively. Any reconstructed photon
¹⁴⁴⁷ which fails to get matched to a generator level photon is labeled as a fake/non-prompt
¹⁴⁴⁸ photon. Direct photons are identified by further requiring that the reconstructed photons
¹⁴⁴⁹ be matched to a parton (a gluon or quark) in space as $\Delta R(\gamma, \text{parton}) > 0.4$. This require-
¹⁴⁵⁰ ment is intended to distinguish the reconstructed photons from highly boosted π^0 's, which

compose a large portion of the experimentally indistinguishable fragmentation photons. Finally, fragmentation photons are obtained exclusively from QCD simulation and are required to have $\Delta R(\gamma, \text{parton}) < 0.4$ in order to avoid double counting photons from the $\gamma+\text{jets}$ sample.

With all three types of photons defined, a study can be carried out from simulation to estimate their respective contributions to the defined control region. The study takes into account that any reconstructed photon in the $\gamma+\text{jets}$ or QCD samples can only be categorized as prompt (through direct production or fragmentation) or non-prompt (fake). The purity and fake rate can then be defined in terms of the relative proportions of prompt or non-prompt photons with respect to the sum of the contributions from all three types of photons. Identified direct photons are taken from the $\gamma+\text{jets}$ sample exclusively. Meanwhile, the fragmentation and fake photon contributions are taken from the QCD sample. The three quantities are then added together and their respective contributions are determined in terms of the photon p_T (Figure 6.3).

The photon purity (Figure 6.3 and Figure 6.4, right) is defined in terms of the prompt and non-prompt photons as:

1469

$$1470 \quad p_\gamma = \frac{\text{prompt}}{\text{prompt} + \text{fake}},$$

where the prompt photon portion comes from the sum of the direct photons (extracted from the $\gamma+\text{jets}$ sample) and the fragmentation photons (extracted from the QCD sample). The remaining non-prompt (or fake) photons all come from photons in the QCD sample that were not matched to truth-level photons in space and momentum with the specified required conditions. Meanwhile, the photon fake rate (Figure 6.3 and Figure 6.4, left) is defined from this same combination of samples as:

1477

$$1478 \quad f = \frac{\text{fake}}{\text{prompt} + \text{fake}},$$

Figure 6.3 shows the purity and fakerate for photons that pass the loose ID/selection, have a $p_T > 200$ GeV and are within the ECAL acceptance range. A sample is obtained in which 77% of the photons are direct, 12% are fragmentation and 11% are fakes. This implies an average purity of $\sim 89\%$ for this sample, well within the value that is expected. Figure 6.4 shows the same ratios for the loose $\gamma+\text{jets}$ control region described in subsection 6.3.2. Although the amount of statistics has decreased due to the additional cuts, a similar trend can be observed.

1486 6.4 The $Z \rightarrow \mu^+ \mu^-$ Control Region

The $Z \rightarrow \mu^+ \mu^-$ control region defined in this section is in every respect identical to the one applied in the 2016 analysis (subsection 5.3.2). The only difference between the 2016 method and the one discussed in this chapter is that the Drell-Yan (DY) sample is

only used for the normalization correction of the $Z \rightarrow \nu \bar{\nu}$ background. Therefore, the loose $\mu\mu$ control region is not used or applied in the calculation of the scale factors. In the following subsections only the tight $\mu\mu$ control region, and its usage to obtain the normalization scale factor R_{norm} , is discussed.

6.4.1 Muon ID and Isolation

The muons are selected using the “medium muon” selection [?], per the recommendation of the muon POG. The muon candidates in this selection satisfy $p_T > 10$ GeV and $|\eta| < 2.4$. Other additional cuts are applied to aid in the muon candidate selection, such as an impact parameter cut. Muons are also subjected to a PF relative-isolation (also referred to as mini-isolation) in which the cone size is inversely proportional to the muon p_T . This requirement enforces the p_T within the isolation cone to be at most 20% of the muon p_T in order to eliminate events with an isolated muon. Details of the medium photon selection are included in Table 6.2 and Table 6.3, while details of the impact parameter cut are summarized in Table 6.4.

Muon Medium ID	
Loose muon ID	Yes
Fraction of valid tracker hits >	0.80
Good Global muon OR Tight segment compatibility >	Yes OR 0.451

Table 6.2: Muon Medium ID 2016 HIP Safe

Good Global muon	
Global muon	Yes
Normalized global-track $\chi^2 <$	3
Tracker-Standalone position match <	12
Kick finder <	20
Segment compatibility >	0.303

Table 6.3: Muon Medium ID HIP Safe Good Global Muon

Muon Impact Parameter	
d0 <	0.2
dz <	0.5

Table 6.4: Additional Impact Parameter cut on Muons

6.4.2 Muon Selection in the Tight Control Region

Events are selected from data samples that contain exactly two oppositely charged muons ($\mu^+ \mu^-$), which fall within the invariant mass $81 < m_{ll} < 101$ GeV window for the Z boson. Additional cuts for the tight muon selection include baseline requirements such as an $H_T > 300$ GeV, $N_j \geq 4$, the $\Delta\phi$ baseline cut on leading jets, a $p_T^{miss} > 250$ GeV, an $m_{T2} > 200$ GeV and at least 1 top-tagged jet $N_t \geq 1$. In addition, the p_T of the two muons are required to be $p_T > 50$ GeV for the leading muon and $p_T > 20$ GeV for the sub-leading one. The only difference, when compared to the signal region is the missing lepton veto, in addition to the dimuon events being treated as p_T^{miss} . This makes for a region that exhibits very similar kinematics to the $Z \rightarrow \nu \bar{\nu}$ signal region, yet suffers from a lack of statistics.

1515 6.5 Analysis

1516 In this section a detailed explanation of the calculation of the scale factors for both
 1517 shape and normalization corrections is provided. The following methods make use of
 1518 the loose γ +jets and the tight $\mu\mu$ control regions defined in the previous sections. The
 1519 procedure involves extracting the shape corrections S_γ from the γ +jets control region and
 1520 afterwards obtain a single normalization correction factor R_{norm} from the tight $\mu\mu$ control
 1521 region. Both factors will then be applied to the final prediction of the $Z \rightarrow \nu\bar{\nu}$ background
 1522 in each of the required search bins.

1523 6.5.1 Shape Correction Using the γ + jets Control Sample

1524 In this section the validation of the γ +jets simulation is discussed in terms of the
 1525 shape of the loose photon control region. As it was shown in subsection 6.3.3, this con-
 1526 trol region has high purity for γ +jets events, particularly in regions of high p_T ($\gtrsim 300$).
 1527 In order to apply this correction factor it is assumed that the shape differences between
 1528 data and simulation are similar between $Z \rightarrow \nu\bar{\nu}$ and γ +jets events. This assumption is
 1529 validated in studies which compare the cross-section ratio of Z +jets to γ +jets events [?].
 1530 Figure 6.5 shows the results of this study, conducted in 2014, for both data and Mad-
 1531 Graph simulation with an integrated luminosity of 19.7fb^{-1} and a center-of-mass energy
 1532 of 8 TeV. It can be seen that for values of $p_T^{Z/\gamma} \gtrsim 300$ GeV, the ratio of the cross-section of
 1533 both processes becomes nearly constant. It is then a matter of applying a factor to account
 1534 for the difference in the amount of events between the Z +jets and γ +jets events in order
 1535 to obtain the total amount of $Z \rightarrow \nu\bar{\nu}$ +jets events. This factor is obtained from the tight
 1536 $\mu\mu$ control region, as shown in subsection 6.5.2.

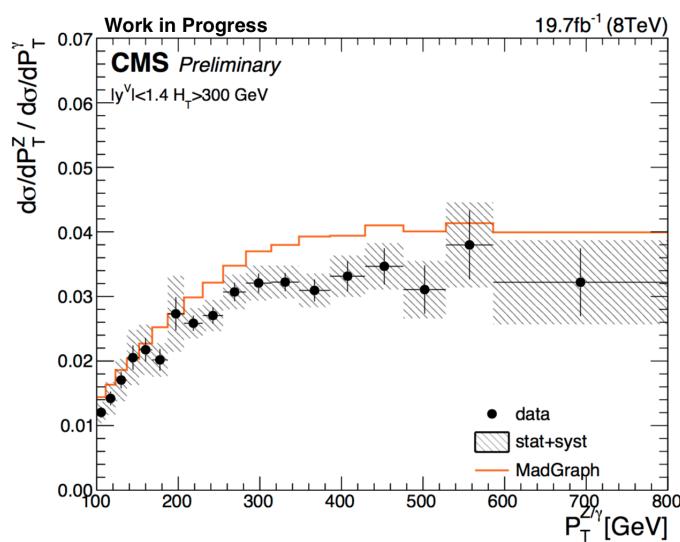


Figure 6.5: Results of study of the Z +jets to γ +jets cross-section ratio for both data and MadGraph sim-
 ulation. For high values of the vector boson transverse momentum, the ratio between these processes is
 observed to be nearly constant.

1537 In order to obtain the shape corrections, the ratio between data and simulation of the
 1538 jet multiplicity distribution is used (Figure 6.6). This is due to it exhibiting the highest
 1539 difference between the observed data and MC. The re-weight for the γ +jets simulation

sample is then accomplished by applying the N_{jet} dependent factor $S_\gamma(N_j)$. This scale factor is determined by taking the ratio of the data and simulation, after subtracting all other MC samples from data events:

$$S_\gamma^i = \frac{\text{Data}^i - \text{MC}_{\text{other}}^i}{\text{MC}_{\gamma+\text{jets}}^i},$$

where i denotes any given bin in the N_j distribution. The shape correction factors S_γ^i are displayed graphically in Figure 6.6 (right) for each N_j bin. These factors correct for differences in the jet multiplicity shape, while the overall normalization is estimated from the tight $\mu\mu$ control region. Figure 6.7 shows the N_j distribution in the tight $\mu\mu$ control region after the calculated scale factors have been applied. The S_γ correction will be applied to the $Z \rightarrow \nu\bar{\nu}$ simulation final prediction for each of the analysis search bins. The uncertainty associated with the scale factor is estimated from the event yields in the loose photon control region. This uncertainty will form part of the total systematic uncertainty in the final prediction.

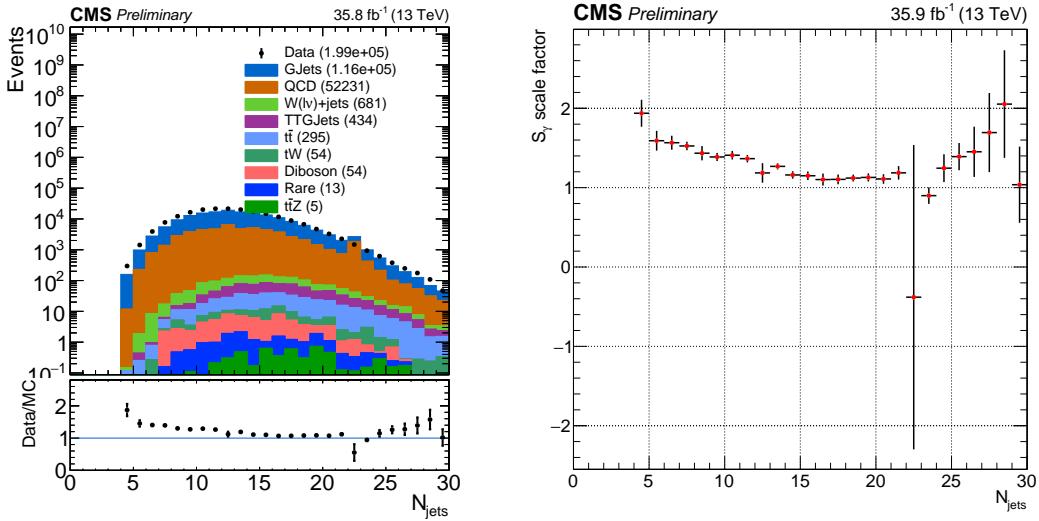


Figure 6.6: Jet multiplicity and the associated S_γ scale factor in the loose photon control region before any corrections are applied.

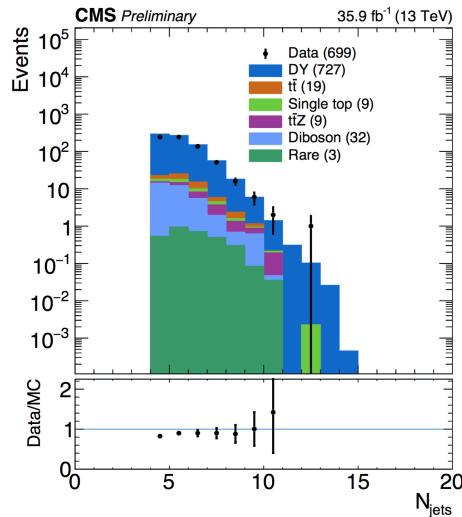


Figure 6.7: N_{jet} distribution in the tight $\mu\mu$ control region after S_γ corrections.

1553 The effect of the $S_\gamma(N_j)$ scale factor is shown for various distributions. These results
 1554 show that the overall agreement between data and simulation improves after applying the
 1555 corresponding shape corrections.

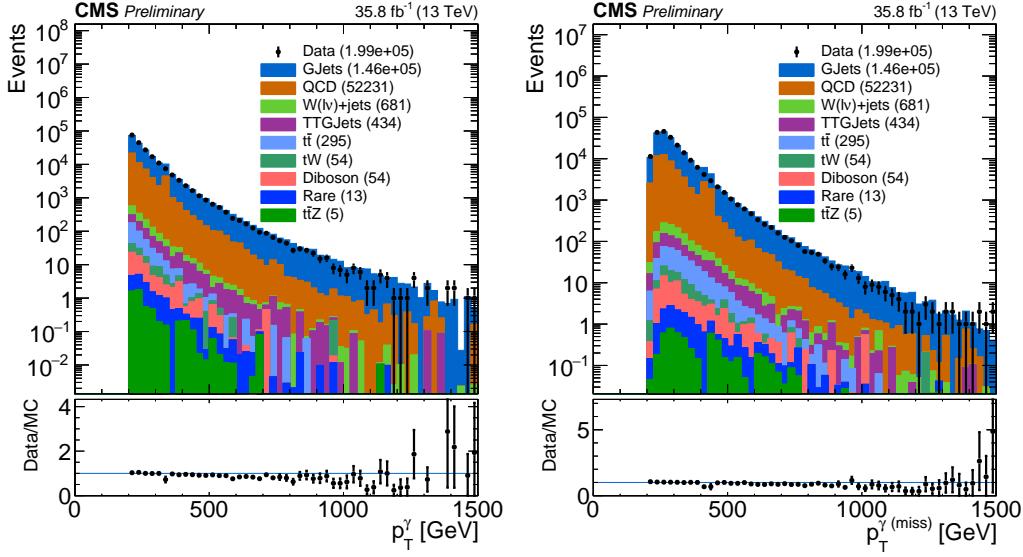


Figure 6.8: p_T^γ (left) and $p_T^{\gamma(\text{miss})}$ (right) distributions after applying the $S_\gamma(N_j)$ scale factor. Comparing to Figure 6.2, an improvement in the agreement between data/MC can be observed.

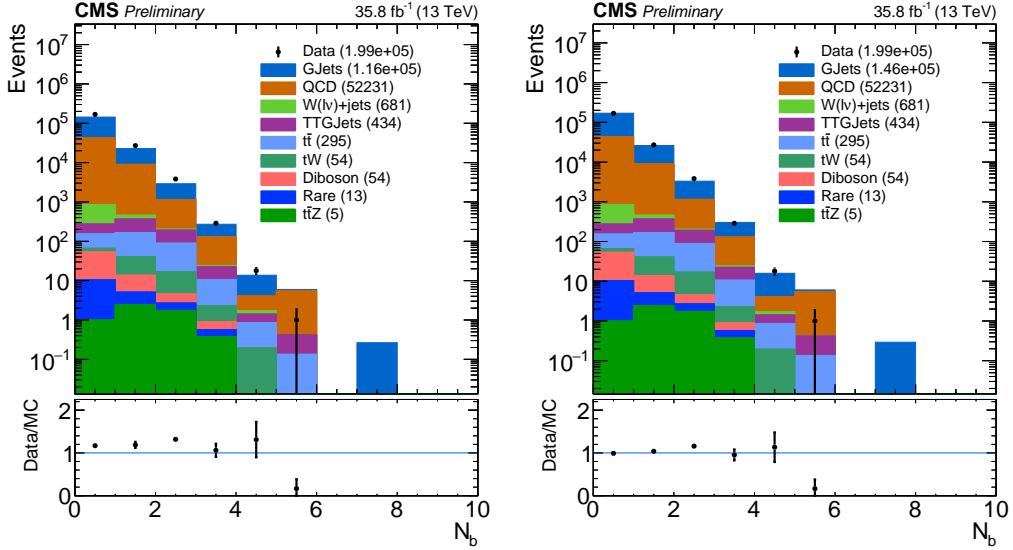


Figure 6.9: N_b distribution before (left) and after (right) applying the $S_\gamma(N_j)$ scale factor.

1556 6.5.2 Normalization Correction Using the tight $Z \rightarrow \mu^+\mu^-$ Control 1557 Sample

1558 In order to constrain the normalization of the $Z \rightarrow \nu\bar{\nu}$ simulation sample, a normaliza-
 1559 tion correction factor R_{norm} is calculated from the tight $\mu\mu$ control region defined in
 1560 subsection 6.5.2. Two categories are considered: the zero b-tagged jet category ($N_b = 0$),
 1561 and the ≥ 1 b-tagged jet category ($N_b \geq 1$). Both of these categories are statistically

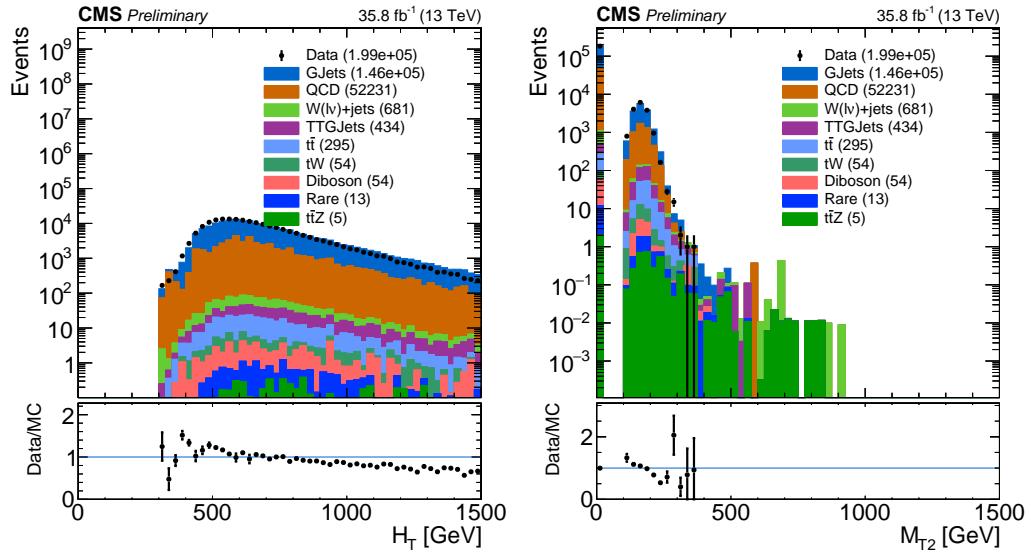


Figure 6.10: H_T and m_{T2} distributions applying the $S_\gamma(N_j)$ scale factor.

consistent with each other but the inclusive region ($N_b \geq 0$) has a lower overall uncertainty. The method used to calculate the normalization scale factor requires that the N_j -dependent shape correction factors already be applied. Then, the R_{norm} factor can be extracted from the ratio of the total event yield in data to that in the simulation. This factor is found to be:

$$R_{norm} = 1.070 \pm 0.085,$$

where the uncertainty includes only the associated statistical uncertainties on data and simulation. This uncertainty is found to be propagated to the final background prediction, see subsection 6.6.1.

1571

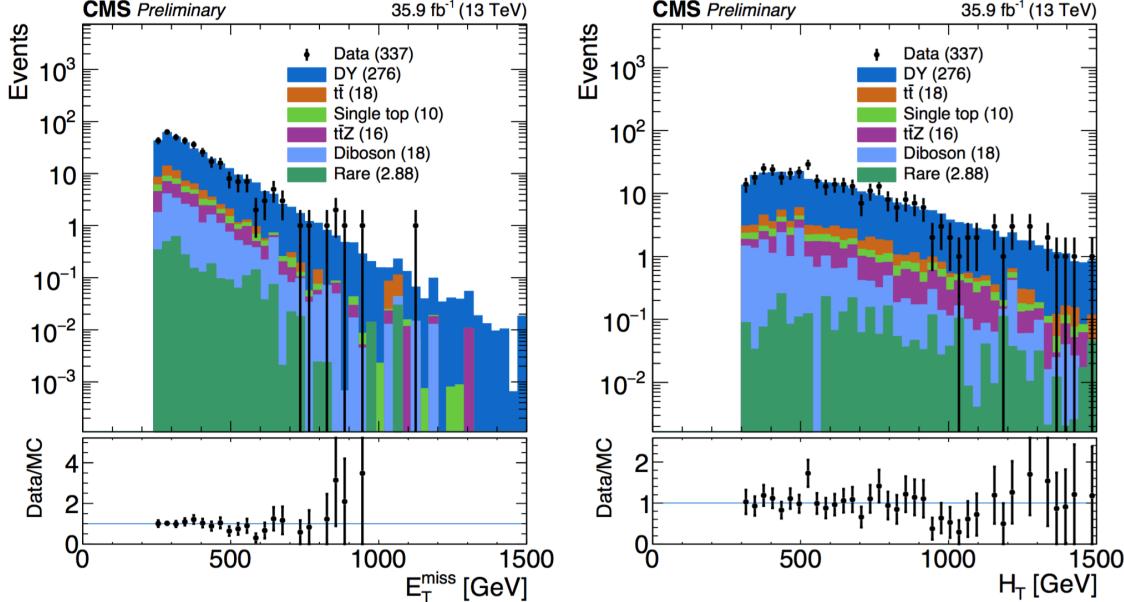


Figure 6.11: Shown are data/MC comparisons for the p_T^{miss} (left) and H_T (right) distributions after applying both the N_j -dependent shape corrections (S_γ) and the global normalization scale factor (R_{norm}).

1572 Data/MC comparisons are shown in [Figure 6.11](#) and [Figure 6.12](#) after applying R_{norm}
 1573 for several distributions in the study. With this final global scale factor all the required
 1574 ingredients for the central value of the $Z \rightarrow \nu\bar{\nu}$ background prediction are obtained.

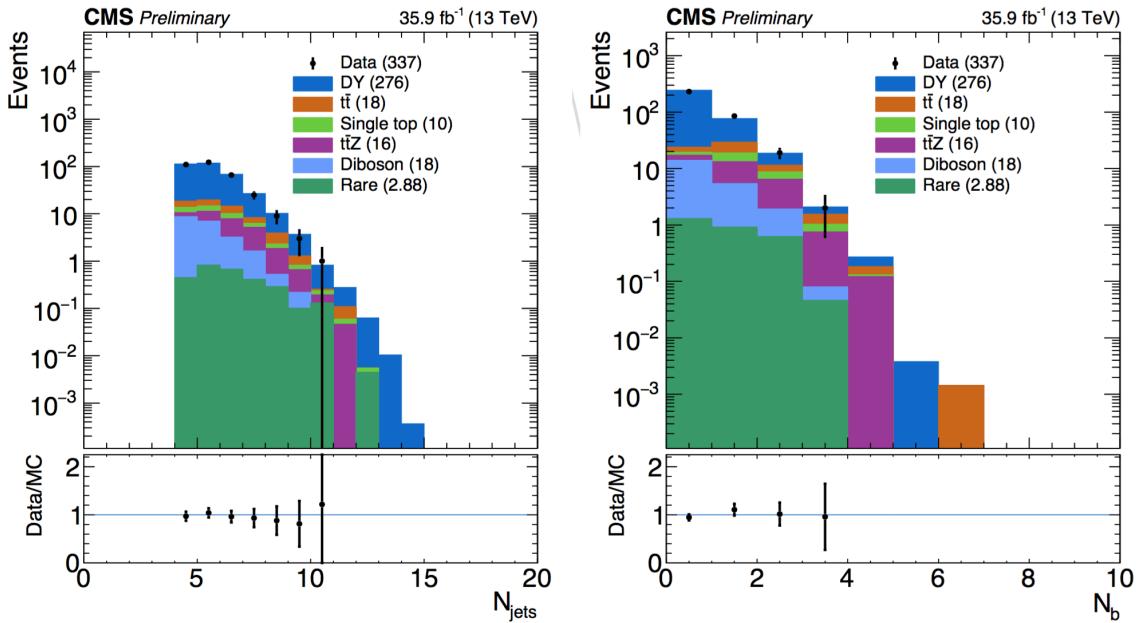


Figure 6.12: Shown are data/MC comparisons for the N_j (left) and N_b (right) distributions after applying both the N_j -dependent shape corrections (S_γ) and the global normalization scale factor (R_{norm}).

1575 6.6 Results

1576 In this section the results for the final estimation of the $Z \rightarrow \nu\bar{\nu}$ are presented.
 1577 The current study includes preliminary results using only data obtained at the CMS de-
 1578 tector during 2016. The results for this study are intended to confirm the assumption that
 1579 the additional $\gamma + \text{jets}$ control region introduced in this analysis reduce the overall uncer-
 1580 tainties obtained in the 2016 analyses (described in [chapter 5](#)). Furthermore, this study is
 1581 intended as a benchmark for future analyses of the SUSY stop group based in Fermilab
 1582 and will be the method used for the 2017 CMS data.

1583 6.6.1 Systematics

1584 Two categories of uncertainties for the $Z \rightarrow \nu\bar{\nu}$ prediction are considered: uncertain-
 1585 ties that are associated to the use of MC simulation and the uncertainties specifically
 1586 associated to the background prediction method. Several sources are acknowledged in the
 1587 first category mentioned such as PDF and renormalization/factorization scale choices, jet
 1588 and p_T^{miss} energy scale uncertainties b-tag scale factor uncertainties, and trigger efficiency
 1589 uncertainties. Given that the simulation sample is normalized to data in the tight control
 1590 region, uncertainties associated with the luminosity and cross-section are excluded. In
 1591 addition, the overall $Z \rightarrow \nu\bar{\nu}$ statistical uncertainty from MC simulation is also taken into
 1592 account.

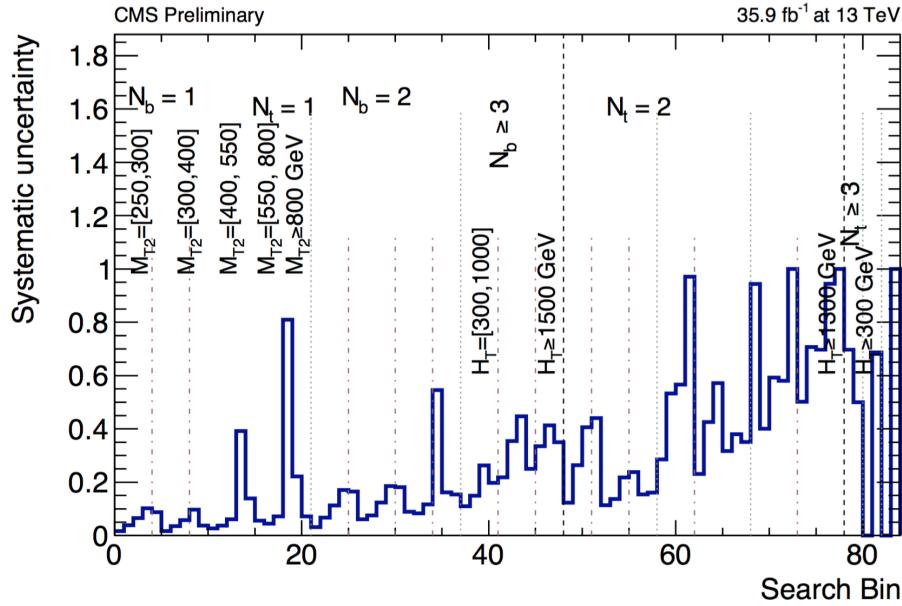


Figure 6.13: Systematic uncertainty in the final prediction, as a function of the search bin, associated to the MC statistics.

1594 The statistical uncertainty associated with each bin in the MC is propagated as a sys-
 1595 tematic uncertainty. The relative uncertainty per bin can be see in Figure 6.13. It shows
 1596 that the uncertainties for the MC vary from as low as 1% up to 81% and even 100% in
 1597 some regions. Since the final estimation is scaled using the global normalization factor
 1598 from the tight $\mu\mu$ control region (R_{norm}), the total uncertainty, due to limited amounts of
 1599 events in data, is propagated in the final prediction. This is also true for the $S_\gamma(N_j)$ scale
 1600 factor, in which the residual differences in search variables other than N_j are evaluated in
 1601 the loose photon control region. Both the uncertainty arising from the N_j re-weighting
 1602 as well as the residual differences are evaluated together. The uncertainty from R_{norm} is
 1603 propagated as a flat value of 7.9% uncertainty per each search bin.

1604 6.6.2 $Z \rightarrow \nu\bar{\nu}$ Estimation for the Search Bins

1605 The final estimation for the $Z \rightarrow \nu\bar{\nu}$ background calculated for all 84 search bins is
 1606 shown in Figure 6.14. The statistical uncertainty in bins that have zero events is treated
 1607 as the average weight (the sum of the weights squared over the weight) times the poisson
 1608 error on 0 which is 1.8. This average weight is calculated on the basis of a relaxed cut in
 1609 which $N_b \geq 2$ is required. For comparison, a cut in which $N_t > 2$ where two tops are
 1610 fake for the $Z \rightarrow \nu\bar{\nu}$ is used.

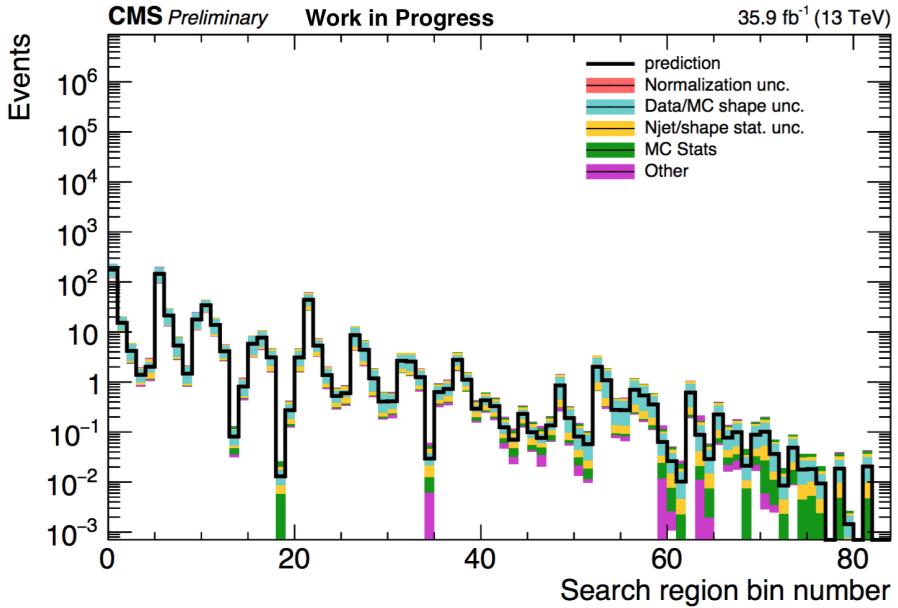


Figure 6.14: $Z \rightarrow \nu\bar{\nu}$ background prediction for all search bins, including the breakdown of the various uncertainties.

1611 **Chapter 7**

1612 **Conclusion**

1613 There are three major topics of research that were discussed in this dissertation: The
1614 simulation studies involving the counting of L1-stubs for the HL-LHC CMS Inner Tracker
1615 upgrade ([chapter 4](#)), the overall 2016 search for SUSY in the all-hadronic channel us-
1616 ing a customized top-tagger ([chapter 5](#)) and the improvements made for the estimation
1617 of the $Z \rightarrow \nu\bar{\nu} +$ jets background using an additional control region from $\gamma +$ jets events
1618 ([chapter 6](#)). These studies were explained in detail in their respective chapters and their
1619 individual results are provided. A summary of the most important results from each study
1620 is provided in this chapter.

1621 **7.1 L1 Stub Counting for the HL-LHC CMS Tracker Up-
1622 grade**

1623 Results from this study (detailed in [chapter 4](#)) reflect the overall effects that were ex-
1624 pected beforehand. The removal of discs from the standard pixel geometry (consisting
1625 of 8 small and 4 large discs) results in a noticeable reduction of stub hits in the upgraded
1626 CMS Outer Tracker. This effect is specially apparent if the disc that is removed is closer to
1627 the interaction point, due to the much larger volume of particles that are present in this re-
1628 gion. Therefore, the reduction in stubs is more pronounced when a small disc is removed
1629 (as in the case of the $7s4l$ geometry) than if a large disc is removed (as in the $8s3l$ pixel
1630 geometry). The reason for this effect stems from the fact that as particles travel through
1631 the various layers of the Inner Tracker material, some of them are bound to interact with
1632 it, producing particles that did not originate from the initial proton-proton collision. The
1633 stubs produced via such processes are considered to be “fake” stubs. To confirm these
1634 findings, an additional study was conducted using a sample that was virtually indistin-
1635 guishable from the standard pixel geometry, but with the second disc on the positive side
1636 “turned off” or “dead”. The results from this study confirm the initial findings and shows
1637 that there is indeed a correlation between the average number of stubs detected in the
1638 Outer Tracker and the total amount of material present in the upgraded Inner Tracker. An
1639 important factor that needs to be taken into account when interpreting these results is the
1640 re-optimization of the disc positions after removing a disc in the different pixel geome-
1641 tries considered. This feature could provide a possible explanation as to why the $6s3l$
1642 geometry, which has two less small discs than the standard geometry (and one less large
1643 one), was found to have less of an effect on the average number of stubs than the $7s4l$

1644 geometry.

1645 7.2 Search for SUSY in the All-Hadronic Channel

1646 The analysis presented in [chapter 5](#) shows the results of a search for SUSY in the
 1647 0-lepton final state using a customized top-tagger. The data was obtained from proton-
 1648 proton collisions at the CMS detector during 2016 with a total integrated luminosity of
 1649 35.9 fb^{-1} at a center-of-mass energy of 13 TeV. The search was conducted by speci-
 1650 fying 84 non-overlapping regions of phase space with varying requirements on the N_b ,
 1651 N_t , p_T^{miss} , H_T and m_{T2} variables ([subsection 5.2.5](#)). Several dominant and non-dominant
 1652 backgrounds were identified and estimated to account for all the majority of the processes
 1653 that were seen in the collected data. The estimation procedures and their respective sys-
 1654 tematic and statistical uncertainties are discussed in [section 5.3](#). The total background
 1655 prediction vs. data for all 84 search bins ([Figure 5.7](#)) shows no statistically significant de-
 1656 viation from the predicted SM background. The biggest sources background were shown
 1657 to be the $t\bar{t}$ and $W+jets$ processes, followed by $Z(\nu\bar{\nu})+jets$, which were seen to be dom-
 1658 inant in regions with a high p_T threshold. Meanwhile, the contributions from the QCD
 1659 multijet and rare backgrounds are found to be nearly negligible in all of the 84 search bins.
 1660 Exclusion limits were calculated from these results for each of the signal models used, by
 1661 applying a binned likelihood fit on the data. The likelihood function was obtained for
 1662 each of the 84 search regions as well as for each of the background data control sam-
 1663 ples from the product of the Poisson probability density function. Exclusion limits were
 1664 placed on the top squark, gluino and LSP production cross-sections with a 95% confi-
 1665 dence level (CL), calculated using a modified frequentist approach with the CL_s criterion
 1666 and asymptotic results for the test statistic. The 95% CL exclusion limits obtained for the
 1667 $T2tt$ model, which consists of direct top squark production, excludes top squark masses
 1668 up to 1020 GeV and LSP masses up to 430 GeV. For the $T1tttt$ model, gluino masses of
 1669 up to 2040 GeV and LSP masses up to 1150 GeV are excluded, with corresponding limits
 1670 of 2020 and 1150 GeV for the $T1ttbb$ model, 2020 and 1150 GeV for the $T5tttt$ model,
 1671 and 1810 and 1100 GeV for the $T5ttcc$ model.

1672 7.3 Estimation of the $Z \rightarrow \nu\bar{\nu} + jets$ Background with a 1673 Hybrid Method

1674 [Chapter 6](#) presents a different method (than the one briefly described in [chapter 5](#))
 1675 for estimating the total amount of $Z \rightarrow \nu\bar{\nu} + jets$ events per search bin. The new method
 1676 makes use of an additional ‘loose $\gamma + jets$ ’ control sample for the estimation of the $S_\gamma(N_j)$
 1677 shape correction factor, in addition to the ‘tight $Z \rightarrow \mu\mu + jets$ ’ sample (used in the 2016
 1678 analysis) to calculate the R_{norm} correction factor ([subsection 5.3.2](#)). The $\gamma + jets$ sample
 1679 was chosen to substitute the previously used ‘loose $Z \rightarrow \mu\mu + jets$ ’ control region, due to
 1680 its much higher cross-section and kinematic similarity to the $Z \rightarrow \nu\bar{\nu} + jets$ process at high
 1681 p_T . The implementation of this method sought to refine the results obtained in 2016 by
 1682 reducing the statistical uncertainties that stem from the low branching fraction of $Z \rightarrow \mu\mu +$
 1683 jets events. The $\gamma + jets$ control region was found to have a high purity and a low fake-rate
 1684 in the regime of high p_T that was being studied. The N_j -dependent S_γ^i was obtained for

1685 each bin i in N_j from a comparison between data and the $\gamma +$ jets MC, after subtracting
1686 the other backgrounds from data and normalizing both samples to 1. This resulted in the
1687 S_γ scale factor plot depicted in [Figure 6.6](#), on the right. As can be seen from this plot, the
1688 statistical uncertainty of the various correction factors is small due to the large number
1689 of events available. Using these values, the R_{norm} normalization correction factor was
1690 obtained from the tight $\mu\mu$ control sample as $R_{norm} = 1.070 \pm 0.085$. Both of these
1691 scale factors were then applied to the final estimation of the $Z \rightarrow \nu\bar{\nu} +$ jets background
1692 ([Figure 6.14](#)).