

FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

TARGETING THE MINIMAL SUPERSYMMETRIC STANDARD MODEL WITH THE
COMPACT MUON SOLENOID EXPERIMENT

By

SAMUEL LOUIS BEIN

A Dissertation submitted to the
Department of Physics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2016

Samuel Louis Bein defended this dissertation on June 23, 2016.
The members of the supervisory committee were:

Harrison B. Prosper
Professor Directing Dissertation

Michael Ruse
University Representative

Andrew Askew
Committee Member

Takemichi Okui
Committee Member

Jorge Piekarewicz
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.

ACKNOWLEDGMENTS

There are so many to thank for so much; this is my attempt. Thank you Harrison for one thousand good opportunities and a two thousand good ideas, and for letting me know that it doesn't matter where or how I train my multivariate discriminant, that even if Moses were to bring it down from the mountain top, I would be free to use it, provided I can reliably model its output in the data. Thank you Sabine for investing in me, for showing me what immunity to group think looks like, and for connecting me with so many who are now my colleagues. Thank you Sezen for persevering with me through the most challenging moments of the pMSSM paper (are they moments if they last for months and years?), and Lukas, Dipan, and Joe Bochenek for all the patience and generosity in the early years. Thank you Marc for making me quit using Mathematica and start using Root, and for having the world's biggest hard drive in your brain; I would not be finishing right now if it weren't for you. Thank you Keith and Bob for teaching me so much when you didn't have to, and for looking close enough to see value in my work. Thank you Hongxuan for having the most incredible sense of what the right research direction is; I realized that most of the things I did were things you encouraged early on. Thank you to Fermilab for hosting me so generously for two years; those years made all the difference; in that vein, thanks to Joe Pastika and Kevin Pedro for showing me so many ropes. Thank you Jory for reminding me to be a human being even though I'm a physicist. Thank you Jim and Jorge for getting me excited about nuclear physics, and for supporting me even though I did high energy. Thank you Take and Laura for caring so deeply about my education and always being willing to mentor. Thank you Andrew for putting an important fear of failure in me, and for reminding me to never use a jet veto, unless I'm totally desperate, and thank you Todd for magically having helpful, key information about every physics problem I ever discussed in your presence. Thank you Michael Ruse for agreeing to invest time in my work and managing to have an incredible amount of insight about subjects unrelated to your beat. Thank you Arka and Ajeeta for being my friends in addition to insightful work mates, and Daniel for being my most trusted

physics consultant and friend; I hope our paths stay intertwined in the coming years. Thank you Rishi for expeditiously helping with the limit plots. Lastly, thanks to my family, Gloria, Rob, Bud, and Joe for the unthinkable amount of love and support throughout my PhD career, not to mention my entire life.

TABLE OF CONTENTS

List of Tables	ix
List of Figures	xiii
Abstract	xxviii
1 Introduction	1
2 Quantum Field Theory	4
2.1 Amplitudes	4
2.2 Fields	6
2.3 Symmetry	6
2.4 Lagrangian density	7
2.5 Particles (quadratic term)	8
2.6 Feynman diagrams (interaction terms)	9
2.7 Conserved quantities	12
2.8 Symmetry breaking	12
2.9 Commutation relations	15
3 The Standard Model	16
3.1 SM fields and symmetries	16
3.2 SM Lagrangian	18
3.3 Electroweak symmetry breaking	18
3.4 SM parameters and visualization	21
3.5 Problems with the standard model	22
3.5.1 Observations in tension with the standard model	22
3.5.2 Fine tuning and unexplained patterns in the standard model	25
3.6 Problems with the standard model	26
3.6.1 Observations in tension with the standard model	27
3.6.2 Fine tuning and unexplained patterns in the standard model	29
3.7 This document	30
4 Supersymmetry	35
4.1 Supersymmetry in our universe	36
4.2 The MSSM	39
4.3 The cMSSM	40
4.4 The pMSSM	42
4.5 Dark matter candidate	43

5 The Large Hadron Collider	44
5.1 LHC key features	44
5.2 The injection sequence	48
5.3 The goals of the LHC	49
5.3.1 The objective of high energy	50
5.3.2 The objective of high luminosity	50
6 The Compact Muon Solenoid	51
6.1 Coordinates and observables	52
6.2 Silicon tracker	53
6.3 Electromagnetic calorimeter	54
6.3.1 ECAL calibration	57
6.4 Hadronic calorimeter	61
6.5 Superconducting solenoid	62
6.6 Muon system	62
6.7 Trigger and DAQ	63
6.8 Particle and event reconstruction	66
6.9 Phase II upgrade	67
6.9.1 Calorimeter research and development	68
7 The Viability of SUSY: The pMSSM Interpretation	75
7.1 Bayes' theorem	75
7.2 The pMSSM interpretation	76
7.3 Methods for the combination CMS analyses	77
7.3.1 Construction of the prior	77
7.3.2 Incorporation of the CMS data	81
7.4 Results	84
7.4.1 Global significance	85
7.4.2 Impact on parameters	85
7.4.3 Correlations among pMSSM parameters	90
7.5 Nonexcluded regions in the pMSSM parameter space	91
7.5.1 Characterization of signal events	93
8 CMS Searches for Supersymmetry at $\sqrt{s} = 13$ TeV	116
8.1 Trigger methods	117
8.1.1 Measuring the efficiency without bias	118
8.1.2 Multivariate classification	119
8.2 QCD background estimation	124
8.2.1 The origin of E_T^{miss} in QCD events	124
8.2.2 Model assumptions and likelihood	125
8.2.3 Rebalance and smear method	125
8.2.4 Closure	129
8.2.5 Applying the method in data	132

8.2.6	Uncertainties in the prediction	132
8.3	$Z \rightarrow \nu\bar{\nu}$ background estimation	142
8.3.1	The relationship between $Z \rightarrow \nu\bar{\nu}$ and $Z \rightarrow l^+l^-$	142
8.3.2	Procedure with 13 TeV data	143
8.3.3	Closure	147
8.3.4	Applying the method in data	148
8.3.5	Uncertainties in the prediction	149
8.4	Results of 2015 SUSY searches	158
8.4.1	Multi-jet + H_T^{miss} SUSY search: targeting gluino production at $\sqrt{s} = 13$ TeV	158
8.4.2	Search for SUSY in events with top-tagged jets at $\sqrt{s} = 13$ TeV	169
9	Targeting the Non-Excluded pMSSM with Multivariate Discriminants	178
9.1	Methods for exploring difficult pMSSM points	179
9.1.1	Low- H_T^{miss} discriminant	179
9.1.2	Low- H_T discriminant	180
9.2	Proof of principle analysis	182
9.2.1	Backgrounds	185
9.2.2	Results	188
10	Summary and Conclusion	192
Appendix		
A	Additional pMSSM Checks and Studies	198
A.1	Change of prior	198
A.2	Profile vs marginal likelihoods	200
B	Analysis Tables	213
C	Other Discriminating Variables	217
C.1	Observables probing supersymmetry	217
D	Public Database of LHC Searches	220
D.1	Toward a public analysis database for LHC new physics searches using MAD-ANALYSIS 5	220
D.2	Implemented analyses and their validation	222
D.2.1	CMS-SUS-13-012: search for new physics through jet multiplicity and missing energy	223
D.3	Limit setting	228
D.4	MadAnalysis 5 conclusions	229
D.5	Validation document: SUS-13-012	230
D.6	Validation document: SUS-14-002	249

Bibliography	258
Biographical Sketch	269

LIST OF TABLES

3.1	The fields and particles of the Standard Model and their quantum numbers. The gluon color can take on eight unique combinations of $c_i \bar{c}_j$ ($i, j = 1, 2, 3$) where $c_1 = r$ (red), $c_2 = g$ (green), and $c_3 = b$ (blue).	21
4.1	The supermultiplets of the MSSM.	39
5.1	The hardware at the interaction points of the LHC.	47
7.1	The Measurements that are the basis of the non-DCSprior $p^{\text{non-DCS}}(\theta)$ for the pMSSM parameters, their observed values and likelihoods. The observables are the decay branching fractions $\mathcal{B}(b \rightarrow s\gamma)$ and $\mathcal{B}(B_s \rightarrow \mu\mu)$, the SUSY to SM ratio for the branching fraction of the decay $B^+ \rightarrow \tau\nu$ ($R[B^+ \rightarrow \tau\nu]$), the difference in the muon anomalous magnetic moment from its SM prediction Δa_μ , the strong coupling constant at the Z boson mass $\alpha_s(m_Z)$, the top and bottom quark masses m_t and $m_b(m_b)$, the Higgs boson mass m_h and signal strength μ_h , and sparticle mass limits from LEP. All data except μ_h were used in the initial MCMC scan. See text for details.	80
7.2	The CMS analyses considered in this study. Each row gives the analysis description, the center-of-mass energy at which data were collected, the associated integrated luminosity, the likelihood used, and the reference to the analysis documentation.	82
8.1	The measured decay modes of the Z boson, as reported in [1].	142
9.1	Expected and observed counts and uncertainties, along with the prediction for the signal yield.	189
B.1	$Z \rightarrow \nu\nu$ background prediction and uncertainties, corresponding to 2.3 fb^{-1} for all search regions in the hadronic top squark search.	214
B.2	Observed event counts, and counts based on the background predictions in for the first 36 search bins from the multi-jet+ H_T^{miss} SUSY search.	215
B.3	Observed event counts, and counts based on the background predictions in from the multi-jet+ H_T^{miss} SUSY search.	216
D.1	Summary of yields for the baseline cuts for the T1qqqq and T1tttt topologies, as compared to the official CMS-SUS-13-012 results posted on the CMS twiki. The results are for the $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125) \text{ GeV}$ benchmark point.	225

D.2	Same as Table D.1 but for the T2qq and T5VV topologies. The benchmark points used are $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0}) = (700, 100)\text{GeV}$ for T2qq and $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)\text{GeV}$ for T5VV.	225
D.3	The cut flow for the baseline selection in CMS SUS-13-012 for the T1qqqq working point $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)\text{GeV}$. The second column is the official account and our own results are given in column 3. The official counts are normalized to luminosity $\mathcal{L} = 19.5/\text{fb}$ and cross section $\sigma = 10.17 \text{ pb}$, and our counts are normalized to match the official count after the first cut, MET Cleaning.	231
D.4	The signal region (SR) counts in CMS SUS-13-012 for the T1qqqq scenario after all selection has been applied. Column 2 is the official account obtained through generous correspondence with Christian Sanders, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.3.	232
D.5	The cut flow for the baseline selection in CMS SUS-13-012 for the T1tttt working point $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)\text{GeV}$. The second column is the official account and our own results are given in column 3. The official counts are normalized to luminosity $\mathcal{L} = 19.5/\text{fb}$ and cross section $\sigma = 10.17 \text{ pb}$, and our counts are normalized to match the official count after the first cut, MET Cleaning. .	235
D.6	The signal region (SR) counts in CMS SUS-13-012 for the T1tttt scenario after all selection has been applied. Column 2 is the official account obtained through generous correspondence with Christian Sanders, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.5.	236
D.7	The cut flow for the baseline selection in CMS SUS-13-012 for the T5VV working point $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)\text{GeV}$. The second column is the official account and our own results are given in column 3. The official counts are normalized to luminosity $\mathcal{L} = 19.5/\text{fb}$ and cross section $\sigma = 10.17 \text{ pb}$, and our counts are normalized to match the official count after the first cut, MET Cleaning. .	239
D.8	The signal region (SR) counts in CMS SUS-13-012 for the T5VV scenario after all selection has been applied. Column 2 is the official account obtained through generous correspondence with Christian Sanders, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.7.	240
D.9	The cut flow for the baseline selection in CMS SUS-13-012 for the T2qq working point $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0}) = (700, 100)\text{GeV}$. The second column is the official account and our own results are given in column 3. The official counts are normalized	

to luminosity $\mathcal{L} = 19.5/\text{fb}$ and cross section $\sigma = 63.4 \text{ pb}$, and our counts are normalized to match the official count after the first cut, MET Cleaning.	243
D.10 The signal region (SR) counts in CMS SUS-13-012 for the T2qq scenario after all selection has been applied. Column 2 is the official account obtained through generous correspondence with Christian Sanders, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.9.	244
D.11 The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-500-125 and the MA5 results are given in column 3.	250
D.12 The signal region (SR) counts in CMS CMS-SUS-14-001 for the signal model working point T2tt-500-125 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.11.	250
D.13 The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-650-25 and the MA5 results are given in column 3.	251
D.14 The signal region (SR) counts in CMS CMS-SUS-14-001 for the signal model working point T2tt-650-25 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.13.	251
D.15 The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-350-0 and the MA5 results are given in column 3.	252
D.16 The signal region (SR) counts in CMS CMS-SUS-14-001 for the signal model working point T2tt-350-0 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.15.	252
D.17 The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-500-100 and the MA5 results are given in column 3.	254
D.18 The signal region (SR) counts in CMS CMS-SUS-14-001 for the signal model working point T2tt-500-100 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.17.	254

D.19	The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-650-50 and the MA5 results are given in column 3.	256
D.20	The signal region (SR) counts in CMS CMS-SUS-14-001 for the signal model working point T2tt-650-50 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.19.	256

LIST OF FIGURES

2.1	An example of an initial state $ i\rangle$ with two incoming protons and a final state $ f\rangle$ with two outgoing protons and a photon γ . The blue object represents the interaction that takes place during the collision.	5
2.2	An example of a field. The dark blue arrows show the magnitude and direction of the gravity field between two massive parallel plates shown in light blue and green.	6
2.3	Diagrams a and b are two of the leading Feynman diagrams in the theory defined by Equation 2.3. Diagram c is forbidden by the theory, and so does not contribute to the amplitude.	10
2.4	The shape of the potential for three choices of parameter a: positive, zero, and negative. The field ϕ always centers on a local minimum of the potential, which is marked in the three cases by the red dot.	14
3.1	The fields of the standard model. From inside out, concentric annuli represent the spin-0, spin-1, and spin- $\frac{1}{2}$ fields.	16
3.2	A parallel coordinates plot of the 19 standard model parameters. Values on the black axes have dimensions of energy and are read from the left, while the values on the blue axes are read from the right.	32
3.3	Subfigure (a) [2] shows the orbital velocity of objects in galaxy NGC 3198 vs their distance from the galactic center. Subfigure (b) [3] shows an X-ray photograph taken by the Chandra X-ray observatory of the aftermath of the collision of two galaxy clusters, 1E0657-558, overlaid on a contour map of the gravitation lensing magnitude factor κ derived from a separate image taken by the Magellan 6.5 m telescope.	33
3.4	Diagrams for the one-loop corrections to the VEV from a generic fermion (a) and a generic scalar (b).	33
3.5	Subfigure (a) [2] shows the orbital velocity of objects in galaxy NGC 3198 vs their distance from the galactic center. Subfigure (b) [3] shows an X-ray photograph taken by the Chandra X-ray observatory of the aftermath of the collision of two galaxy clusters, 1E0657-558, overlaid on a contour map of the gravitation lensing magnitude factor κ derived from a separate image taken by the Magellan 6.5 m telescope.	34
3.6	Diagrams for the one-loop corrections to the VEV from a generic fermion (a) and a generic scalar (b).	34

4.1	The cancelation.	37
5.1	The layout of the Large Hadron Collider. The scale is approximate and the separation between the beams has been exaggerated to show the crossings of beam 1 (blue) and beam 2 (red) at interaction points 1, 2, 5, and 8.	45
5.2	Cross section of a cryodipole, the most common segment of the LHC [4].	46
6.1	The CMS detector, pulled apart during the long shutdown of 2014.	52
6.2	A schematic view of the CMS tracker. The direction of view is toward the center of the LHC.	54
6.3	An ECAL endcap PbWO ₄ crystal, length 220 mm (24.7 X ₀).	55
6.4	The layout of the ECAL barrel (a) in η - ϕ coordinates, and the forward (c) and rear (b) ECAL endcaps in cartesian coordinates. The shading shows the occupancy of a given run. OT stands for out-of-time, indicating that the occupancy shown in the color plot is of energy deposits that were detected out of time with the LHC bunch crossings.	56
6.5	The trigger primitives available from the calorimeters available to the CMS L1 trigger (a), and the observables derived from the primitives designed to discriminate between signal and background e/ γ candidates.	59
6.6	The signal efficiency in the barrel and endcap as a function of the p_T of the e/ γ candidates for various thresholds on the H/E and isolation.	60
6.7	The background efficiency vs the signal efficiency for cut-based selection and the selection of the BDT based whose inputs are the cut-based observables. The ideal point is indicated by the gold circle located at full signal efficiency and zero background efficiency. For a given background efficiency, the cut-based signal efficiency is within a few percent of the BDT efficiency.	61
6.8	The definition of the centrality (left), along with the distributions of the centrality for signal and background e/ γ candidates in simulation (right).	62
6.9	The background efficiency vs the signal efficiency for cut-based selection, the plane BDT selection, and the selection of the BDT trained with the addition of the centrality variable. The ideal point is indicated by the gold circle located at full signal efficiency and zero background efficiency. For a given signal efficiency, a suppression of the background on the order of 10-15% is possible.	63
6.10	A schematic view of the CMS HCAL. The direction of view is toward the center of the LHC.	64

6.11	An event display of the CMS muon system. Two muons are reconstructed in the muon system.	65
6.12	A shashlik calorimeter tower.	68
6.13	A scatter plot of the accumulated track positions weighted by the in-time energy deposited in the shashlik tower, for a set of muon runs recorded in July, 2014. The 4×4 tower array structure, the gaps between towers, and the holes bored for the WLS fibers can be seen as green features.	69
6.14	Top: heat maps of the ADC counts recorded at the upstream and downstream faces of the shashlik module. Bottom: Diagnostic displays of channel-by-channel information from the shashlik module. The "Chi2 Vs. Channel" display is effective at revealing noisy channels, that is, channels with a large ADC count generated by electronic noise rather than a genuine particle signal.	71
6.15	Top: The pedestal noise by channel. Bottom: A scatter plot of the positions of accumulated hits and tracks recorded by the wire chambers; quality hits are colored red, and refer to hits consistent with a particle track.	72
6.16	Three-dimensional views of reconstructed tracks and calorimetric energy deposits in events containing a 120 GeV proton. The view can be rotated by the user, allowing different viewing perspectives of the event, as illustrated in the top and bottom figures.	73
6.17	The jet response (left) and energy resolution (right) in the shashlik endcap as a function of simulated jet p_T in a sample of simulated $t\bar{t}$ events. The red distributions correspond to the scenario without pile-up, and pink to that with pile-up of 140 (an average of 140 collisions per bunch crossing), which is the expected amount of pile-up during the post-Phase II LHC operations.	74
7.1	The distribution of model points, weighted by the non-DCS prior density, of the Z -significance for the individual 8 TeV searches (a–c), and for 7 TeV combined and 7+8 TeV combined searches (d). The leftmost bins contain the underflow entries.	86
7.2	A summary of the impact of CMS searches on our knowledge of the gluino mass in the pMSSM parameter space. Plots (a)-(d) compare the non-DCSprior distribution of the gluino mass (blue filled histograms) to posterior distributions after data from various CMS searches (line histograms), where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the gluino mass for various combinations of CMS data and data from Higgs boson measurements, where the shaded grey band gives the the statistical uncertainty on the black histogram. Plot (f)	

shows the distribution of the gluino mass versus the Z -significance calculated from the combination of all searches. 97

- 7.3 A summary of the impact of CMS searches on our knowledge of the \tilde{u}_L mass (equivalently, the \tilde{c}_L mass) in the pMSSM parameter space. Plots (a)-(d) compare the non-DCSprior distribution of the \tilde{u}_L mass to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the \tilde{u}_L mass for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the \tilde{u}_L mass versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading. 98
- 7.4 A summary of the impact of CMS searches on our knowledge of the mass of the lightest colored SUSY particle (LCSP) in the pMSSM parameter space. Plots (a)-(d) compare the non-DCSprior distribution of the LCSP mass to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the LCSP mass for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the LCSP mass versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading. 99
- 7.5 A summary of the impact of CMS searches on our knowledge of the \tilde{t}_1 mass in the pMSSM parameter space. Plots (a)-(d) compare the non-DCSprior distribution of the \tilde{t}_1 mass to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the \tilde{t}_1 mass for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the \tilde{t}_1 mass versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading. 100
- 7.6 A summary of the impact of CMS searches on our knowledge of the $\tilde{\chi}_1^0$ mass in the pMSSM parameter space. Plots (a)-(d) compare the non-DCSprior distribution of the $\tilde{\chi}_1^0$ mass to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the $\tilde{\chi}_1^0$ mass for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the $\tilde{\chi}_1^0$ mass versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading. 101

7.7	A summary of the impact of CMS searches on our knowledge of the the mass of the lightest non-degenerate (LND) chargino in the pMSSM parameter space. Plots (a)-(d) compare the non-DCSprior distribution of the LND $\tilde{\chi}^\pm$ mass to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the LND $\tilde{\chi}^\pm$ mass for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the LND $\tilde{\chi}^\pm$ mass versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading.	102
7.8	A summary of the impact of CMS searches on our knowledge of the the logarithm of the cross section for inclusive sparticle production in 8 TeV pp collisions, $\log_{10}(\sigma_{\text{SUSY}}^{8 \text{ TeV}})$, in the pMSSM parameter space. Plots (a)-(d) compare the non-DCSprior distribution of the $\log_{10}(\sigma_{\text{SUSY}}^{8 \text{ TeV}})$ to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the $\log_{10}(\sigma_{\text{SUSY}}^{8 \text{ TeV}})$ for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the $\log_{10}(\sigma_{\text{SUSY}}^{8 \text{ TeV}})$ versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading. In plot (d), the apparent enhancement of the left tail of the posterior density with respect to the prior is due to the suppression of the right tail and an overall renormalization.	103
7.9	Comparison of prior and posterior distributions after several combinations of data from the CMS searches for the \tilde{u}_R, \tilde{c}_R mass, \tilde{b}_1 mass, $\tilde{e}_L, \tilde{\mu}_L$ mass, $\tilde{\tau}_1$ mass, $\tilde{\chi}_2^0$ mass, $\tilde{\chi}_1^\pm$ mass, the higgsino mass parameter μ , $\tan \beta$, and A mass.	104
7.10	Comparison of prior and posterior distributions after several combinations of data from the CMS searches for $\Omega_{\tilde{\chi}_1^0}$, $\xi \sigma^{\text{SD}}(p\tilde{\chi}_1^0)$, and $\xi \sigma^{\text{SI}}(p\tilde{\chi}_1^0)$	105
7.11	Marginalized non-DCS distributions (first column), compared with posterior distributions (second column) and survival probabilities (third column) after inclusion of the combined CMS searches, are shown for the $\tilde{\chi}_1^0$ mass versus gluino mass (first row), the top squark mass (second row), and the logarithm of the cross section for inclusive sparticle production at 8 TeV (third row).	106
7.12	The twelve most common principal processes in the pMSSM, listed in order of their frequency before the constraints of the CMS searches. Both on-shell and off-shell states are included. Indices of particle charge, flavor, and chirality are ignored in the construction, with the exception of the flavor of the third-generation squarks and quarks. Asterisks in the labels indicate where process names involving long decay chains have been abbreviated.	107

7.13	(a) shows the fraction of excluded (dark) and nonexcluded (light) points out of all considered points, by principal process. Color is assigned to the processes that are most common after the constraints of the CMS searches, which are selected for further study. The dominance of principal processes, as defined in Eq. 7.15, is given in (b) where the bands show the RMS range of the dominance.	108
7.14	A parallel coordinates plot showing several hundred selected nonexcluded model points for the six most common principal processes, with seven key properties. From the left, the selected properties are: the principal process, the 8 TeV signal production cross section (in \log_{10} scale), the average value of the E_T^{miss} , the average number of b-jets, leptons, and jets, and finally, the average p_T momentum of the leading jet. Color is assigned based on the principal process. Orange codes for process 1, blue for process 2, green for 3, red for 4, violet for 7, and cyan for 10. Lines arching toward higher vertical positions typically indicate more “discoverable” scenarios.	109
7.15	A parallel coordinates plot of the nonexcluded pMSSM points with the axes set as the principal process, the average E_T^{miss} (in GeV), and the fiducial cross section (in linear scale) for various thresholds on the E_T^{miss} . All nonexcluded points corresponding to processes 1, 2, 3, 4, 7, and 10 that have a fiducial cross section greater than 100 fb are shown. Color is assigned to values of the principal process in the same manner as in Fig. 7.14.	110
7.16	A parallel coordinates plot of the nonexcluded pMSSM points with the axes set as the principal process, the average H_T (in GeV), and the fiducial cross section (in linear scale) for various thresholds on the H_T . All nonexcluded points corresponding to processes 1, 2, 3, 4, 7, and 10 that have a fiducial cross section greater than 60 fb are shown. Color is assigned to values of the principal process in the same manner as in Fig. 7.14.	111
7.17	A parallel coordinates plot of the nonexcluded pMSSM points with the axes set as the principal process and the fiducial cross section (in linear scale) for various thresholds on the multiplicity of jets. All nonexcluded points corresponding to processes 1, 2, 3, 4, 7, and 10 that have a fiducial cross section greater than 200 fb are shown. Color is assigned to values of the principal process in the same manner as in Fig. 7.14.	112
7.18	A parallel coordinates plot of the nonexcluded pMSSM points with the axes set as the principal process and the fiducial cross section (in linear scale) for various thresholds on the sub-leading lepton p_T (in GeV). All nonexcluded points corresponding to processes 1, 2, 3, 4, 7, and 10 that have a fiducial cross section greater than 30 fb are shown. Color is assigned to values of the principal process in the same manner as in Fig. 7.14.	113

7.19	A parallel coordinates plot of the nonexcluded pMSSM points with the axes set as the principal process and the fiducial cross section (in linear scale) for various thresholds on the leading b-tagged jet p_T (in GeV). All nonexcluded points corresponding to processes 1, 2, 3, 4, 7, and 10 that have a fiducial cross section greater than 20 fb are shown. Color is assigned to values of the principal process in the same manner as in Fig. 7.14.	114
7.20	The 18 th and 34 th most frequent principal process in the pMSSM.	115
8.1	The ratio of the monojet trigger efficiency for events passing the single electron reference trigger to the trigger efficiency for all events in a simulated $t\bar{t}$ sample, as a function of the offline H_T and H_T^{miss} . The ratio is consistent with 1 throughout the $H_T - H_T^{\text{miss}}$ plane.	120
8.2	The monojet trigger efficiency, measured in real data, as a function of the offline H_T and H_T^{miss}	121
8.3	The trigger efficiency as a function of the offline H_T^{miss} for the monojet trigger, measured the traditional way (histograms) and by the NN method (blue function), for $N_{\text{jets}} = 4$, with varying values of the H_T . The green function lines show the 1-sigma uncertainty on the efficiency.	123
8.4	The likelihood function for the jet energy scale factor in two regions of the phase space of the parton-level jet four-vector. These distributions are equivalent to the smearing templates. The histograms (red) are smoothly interpolated using splines (black).	126
8.5	The parton-level and reconstruction-level H_T^{miss} and azimuthal separation between the H_T^{miss} and leading jet in QCD simulation. The red functions are taken as probability distributions making up the prior, $P[H_T^{\text{miss}}(\vec{J}_{\text{part}})]$ and $P[\Delta\phi_{H_T^{\text{miss}}, p_T^1}(\vec{J}_{\text{part}})]$	128
8.6	The jet energy response for the leading jet (left) and sub-leading jet (right) for reconstructed jets (green) and rebalanced jets (blue). Parton-level jets are required to have $p_T > 30$ GeV and $ \eta < 2.4$. The resolution of rebalanced jets is better (smaller) than that of reconstruction-level jets by bout 10%.	129
8.7	Comparisons of kinematic distributions between the direct simulation and the rebalance and smear method applied to simulation, after the baseline selection of the multi-jet SUSY search.	134
8.8	Comparisons of kinematic distributions between the direct simulation and the rebalance and smear method applied to simulation, after the baseline selection of the multi-jet SUSY search.	135

8.9	Comparisons of kinematic distributions between the direct simulation and the rebalance and smear method applied to simulation, in the inverted $\Delta\phi$ control region of the multi-jet SUSY search.	136
8.10	Comparisons of kinematic distributions between the direct simulation and the rebalance and smear method applied to simulation, in the inverted $\Delta\phi$ control region of the multi-jet SUSY search.	137
8.11	The ratio of scatter plots of key observables for hadronic SUSY searches between the direct simulation and the rebalance and smear method applied to simulation.	138
8.12	Comparisons of the H_T^{miss} distribution between prediction and expectation using the classical method (left) and the new method (right).	139
8.13	Signal contamination removal. The midnight blue (and yellow) histograms show the distribution of QCD (and signal) events taken directly from simulation. The black (and red) histograms show the distributions of QCD (and signal) and after being processed by the rebalance and smear procedure. The signal has been removed from the high- H_T^{miss} region, leaving an accurately modeled QCD contribution.	140
8.14	The distribution of the H_T in the H_T control sample for the unweighted offline H_T (red), the unweighted online H_T (pink), and the online H_T weighted by the pre-scale values.	141
8.15	The Z boson acceptance in simulated events with $Z \rightarrow e^+e^-$	145
8.16	The Z boson reconstruction efficiency in simulated events with $Z \rightarrow e^+e^-$ decays.	146
8.17	The Z boson isolation efficiency as a function of the ΔR between the two electrons and the geometric mean of the activities of the electrons in simulated events with $Z \rightarrow e^+e^-$ decays.	147
8.18	Comparison of $Z \rightarrow \nu\bar{\nu}$ prediction (blue) and expectation (red) for the H_T (1) and E_T^{miss} (2), the jet multiplicity (3) and the M_{T2} (4), after the baseline selection of the 2015 hadronic stop analysis. The ratio is of the prediction to the expectation. All events are from simulation.	151
8.19	Comparison of $Z \rightarrow \nu\bar{\nu}$ prediction (blue) and expectation (red) for the b-jet multiplicity (1) and top tag multiplicity (2), after the baseline selection of the 2015 hadronic stop analysis. The ratio is of the prediction to the expectation. All events are from simulation.	152
8.20	Comparison of $Z \rightarrow \nu\bar{\nu}$ prediction (blue) and expectation (red) for the H_T (1) and E_T^{miss} (2), the jet multiplicity (3) and the M_{T2} (4), after a loosened baseline	

selection for the 2015 hadronic stop analysis. The ratio is of the prediction to the expectation. All events are from simulation.	153
8.21 Comparison of $Z \rightarrow \nu\bar{\nu}$ prediction (blue) and expectation (red) for the b-jet multiplicity (1) and top tag multiplicity (2), after the loosened baseline selection for the 2015 hadronic stop analysis. The ratio is of the prediction to the expectation. All events are from simulation.	154
8.22 Comparison of data/simulation event weights for the $Z \rightarrow \nu\bar{\nu}$ prediction based on the electron sample (green) and the muon sample (magenta) for the b-jet multiplicity 0 (left) b-jet multiplicity greater than 0 (right), after the loosened baseline selection for the 2015 hadronic stop analysis. The denominator of the weights are the counts taken from on $Z \rightarrow l^+l^-$ simulation, and the numerator is the count in $Z \rightarrow l^+l^-$ data where the $t\bar{t}$ contribution has been subtracted using simulation.	155
8.23 The data-simulation normalization factor computed from the $\mu\mu$ sample (magenta), the ee sample (green), and the combined ee + $\mu\mu$ sample (red).	156
8.24 Comparisons between reweighted simulation and dilepton data.	157
8.25 The simplified models used for the optimization and interpretation of the multi-jet + H_T^{miss} search. They are T1qqqq (upper left), T1tttt (upper right), T1bbbb (lower left), and T5VV (lower right) scenarios.	159
8.26 The ratio of the trigger efficiency for events passing the single electron reference trigger to the trigger efficiency for all events in a simulated $t\bar{t}$ sample, as a function of the offline H_T and H_T^{miss} . The ratio is consistent with 1 within the region of the baseline selection of $H_T > 500$ GeV and $H_T^{\text{miss}} > 200$ GeV of the CMS hadronic searches.	160
8.27 The trigger efficiency for HLT_PFHT350_PFMET100* as a function of the search variables H_T and H_T^{miss} . The dashed (solid) blue lines show the distributions of the denominator (numerator) samples. These results were used in the CMS PAS on the commissioning of 13 TeV observables for SUSY searches [5].	161
8.28 The signal region boundaries in the planes of $H_T^{\text{miss}} - H_T$ (top) and $N_{\text{jets}} - N_{\text{b-jets}}$ (bottom) for the multi-jet + H_T^{miss} search. The pie charts represent the relative contributions to the Standard Model backgrounds in the region $N_{\text{jets}} = 4-6$, $N_{\text{b-jets}} = 0$ (top), and $H_T^{\text{miss}} = 200-500$ GeV, $H_T = 500-750$ GeV (bottom).	162
8.29 The QCD prediction based on the inverted $\Delta\phi$ extrapolation method compared with that of the rebalance and smear method. Each type of uncertainty in the rebalance and smear prediction appears as a color-filled rectangle centered on the predicted value, summed in quadrature with the inner uncertainties.	

Uncertainties in the inverted $\Delta\phi$ prediction are the statistical systematic uncertainties reported in [6], summed in quadrature.	164
8.30 Left: the distribution of the significance of the deviation between the predictions from the inverted $\Delta\phi$ and rebalance and smear methods. Right: the distribution of the fractional uncertainties in the counts predicted by the two methods. The upper overflow bin is populated by bins for which the prediction is identically 0.	165
8.31 Observed (black contour) and expected (red contour) limits on the masses of the gluino and LSP using the QCD estimate based on the inverted $\Delta\phi$ method (left) and the rebalance and smear method (right), in the context of the T1qqqq model. The dashed (grey) lines indicate the $\tilde{\chi}_1^0 = m_{\tilde{g}}$ diagonal.	167
8.32 Observed (black contour) and expected (red contour) limits on the masses of the gluino and LSP using the QCD estimate based on the inverted $\Delta\phi$ method (left) and the rebalance and smear method (right), in the context of the T1tttt model.	167
8.33 Observed (black contour) and expected (red contour) limits on the masses of the gluino and LSP using the QCD estimate based on the inverted $\Delta\phi$ method (left) and the rebalance and smear method (right), in the context of the T1bbbb model.	168
8.34 The observed (black contour) and expected (red contour) limits on the masses of the gluino and LSP using the QCD estimate based on the inverted $\Delta\phi$ method (left) and the rebalance and smear method (right), in the context of the T5VV model.	168
8.35 The simplified models used for the optimization and interpretation of the hadronic search for SUSY in events with top-tagged jets. They are T2tt (left) and T1tttt (right).	169
8.36 Search bin definitions after baseline selection cuts.	171
8.37 The predicted counts originating from standard model $Z \rightarrow \nu\bar{\nu}$ production, based on the $\mu\mu$ sample (top), the ee sample (center), and the combined $\mu\mu+ee$ sample (bottom). Each type of uncertainty, described in Section 8.3, appears as a color-filled rectangle centered on the predicted value, summed in quadrature with the inner uncertainties.	174
8.38 Data are shown as black points. The total predictions are shown in filled solid area. The bottom plot shows the ratio of data over total background prediction in each search bin. Only statistical uncertainties are propagated to the ratio. The numbering scheme is given in Ref. [7].	175

8.39	Observed (black contour) and expected (red contour) limits on the top squark and LSP masses in the context of the T2tt model. The top-left plot corresponds to the $\mu\mu$ -derived $Z \rightarrow \nu\bar{\nu}$ prediction, the top-right to the ee-derived prediction, and the bottom plot to the prediction of the combined $\mu\mu+ee$ sample.	176
8.40	Observed (black contour) and expected (red contour) limits on the gluino and LSP masses in the context of the T1tttt model. The left plot corresponds to the $\mu\mu$ -derived prediction for the $Z \rightarrow \nu\bar{\nu}$ counts, and the right plot to the muon+ee-derived prediction.	177
9.1	Top: the distribution of the output of the low- H_T^{miss} data-driven discriminant for a sample of simulated QCD events, compared with the distribution obtained by applying the rebalance and smear method to simulation. Also shown are the distributions of the output for a few signal points corresponding to various SMS scenarios. Bottom: the efficiency of selecting signal events as a function of the efficiency for selecting background events, based on a scan through all possible thresholds on the discriminant.	181
9.2	Top: The distribution of the output of the low- H_T muon data-driven multivariate discriminant for a sample of simulated $Z \rightarrow \nu\bar{\nu}$ events, compared with the distribution obtained by applying the data-driven dielectron prediction method to simulation. Also shown are the distributions of the output for a few signal points corresponding to various SMS scenarios. Bottom: the efficiency of selecting signal events vs that of selecting background events, based on a scan through all possible thresholds on the discriminant.	183
9.3	Kinematic distributions of selected low-mass gluino pMSSM points. The benchmark pMSSM point, referred to as survivor 1, is shown in blue.	184
9.4	Comparison of the normalized discriminant between simulated $t\bar{t}$ events with 2 b-tags and 0 b-tags, in the baseline region.	186
9.5	Comparison of the normalized discriminant between simulated $t\bar{t}$ events and W+jets events, in the signal baseline region.	187
9.6	Comparison of the output of the discriminant shapes in real data between control regions with various relative contributions from $t\bar{t}$ and W+jets events.	187
9.7	Predicted and observed counts for the low- H_T search using the fully data-driven prediction (top), as well as the prediction from pure simulation (bottom), shown for comparison. The yellow bands are the total uncertainties in the data-driven prediction (top), and the purely statistical uncertainties in the simulated counts (bottom).	190

A.1	Left: posterior/prior of $m_{\tilde{g}}$ (top) and m_{U_L} (bottom). Right: Ratio of the posterior to the prior of $m_{\tilde{g}}$ (top) and m_{U_L} (bottom).	199
A.2	Left: posterior/prior of m_{U_R} (top) and $m_{\tilde{t}_1}$ (bottom). Right: Ratio of the posterior to the prior of m_{U_R} (top) and $m_{\tilde{t}_1}$ (bottom).	202
A.3	Left: posterior/prior of $m_{\tilde{\chi}_1^0}$ (top) and $m_{\tilde{b}_1}$ (bottom). Right: Ratio of the posterior to the prior of $m_{\tilde{\chi}_1^0}$ (top) and $m_{\tilde{b}_1}$ (bottom).	203
A.4	Left: posterior/prior of $m_{\tilde{l}_1}$ (top) and $m_{\tilde{\tau}_1}$ (bottom). Right: Ratio of the posterior to the prior of $m_{\tilde{l}_1}$ (top) and $m_{\tilde{\tau}_1}$ (bottom).	204
A.5	Left: posterior/prior of $m_{\tilde{\chi}_1^\pm}$ (top) and m_A (pole) (bottom). Right: Ratio of the posterior to the prior of $m_{\tilde{\chi}_1^\pm}$ (top) and m_A (pole) (bottom).	205
A.6	The results of 10 of the 19 MLP regressions on the pMSSM parameters.	206
A.7	The results of the 9 of the 19 MLP regressions on the pMSSM parameters.	207
A.8	The output of the MLP used to derive the correction function. The red distribution is the original set of pMSSM points and the blue distribution is the set generated by a random sampling of the separable likelihood function.	208
A.9	Comparison of the frequentist profile likelihood (dashed blue line) to the non-DCS prior (solid black line) for 6 of the pMSSM parameters.	209
A.10	Comparison of the frequentist profile likelihood (dashed blue line) to the non-DCS prior (solid black line) for an additional 6 of the pMSSM parameters.	210
A.11	Comparison of the frequentist profile likelihood (dashed blue line) to the non-DCS prior (solid black line) for an additional 6 of the pMSSM parameters.	211
A.12	Comparison of the frequentist profile likelihood (dashed blue line) to the non-DCS prior (solid black line) for a pMSSM parameter.	212
D.1	Comparison between the official and MADANALYSIS 5 results for the H_T distribution after all baseline cuts, for the T2qq simplified model of CMS-SUS-13-012 with $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0}) = (700, 100)\text{GeV}$	226
D.2	Same as Fig. D.1 but for the H_T^{miss} distribution.	227
D.3	Same as Fig. D.1 but for the N_{jets} distribution.	227
D.4	The 95% CL exclusion limit (in red) in the $\tilde{\chi}_1^0$ versus \tilde{g} mass plane reproduced with the MADANALYSIS 5 implementation [8] of CMS-SUS-13-016. For com-	

parison, the full and dashed grey lines show the official CMS result with its $\pm 1\sigma$ uncertainty from Fig. 6 of [9]. The limit setting in the region where one of the tops from the gluino decay is off-shell, <i>i.e.</i> for $m_{\tilde{g}} \lesssim 800\text{GeV}$, is work in progress.	229
D.5 Diagram of the dominant SUSY production mechanism for the T1qqqq signal model.	231
D.6 Comparison of the distributions of H_T between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1qqqq signal model.	233
D.7 Comparison of the distributions of H_T^{miss} between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1qqqq signal model.	233
D.8 Comparison of the distributions of NJets between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1qqqq signal model.	234
D.9 Additional checks: comparison between ours and the official distributions of NJets after $\text{BL}+H_T>1250$ cuts (left), and H_T after $\text{BL}+\text{NJets}>7$ cuts (right), for the T1qqqq signal model.	234
D.10 Diagram of the dominant SUSY production mechanism for the T1tttt signal model.	235
D.11 Comparison of the distributions of H_T between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1tttt signal model.	237
D.12 Comparison of the distributions of H_T^{miss} between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1tttt signal model.	237
D.13 Comparison of the distributions of NJets between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1tttt signal model.	238
D.14 Additional checks: comparison between ours and the official distributions of NJets after $\text{BL}+H_T>1250$ cuts (left), and H_T after $\text{BL}+\text{NJets}>7$ cuts (right), for the T1tttt signal model.	238
D.15 Diagram of the dominant SUSY production mechanism for the T5VV signal model.	239

D.16	Comparison of the distributions of H_T between the official and our own samples after the “n-1” cut, Min $\Delta(\phi)$ (left), and after all baseline cuts (right), for the T5VV signal model.	241
D.17	Comparison of the distributions of H_T^{miss} between the official and our own samples after the “n-1” cut, Min $\Delta(\phi)$ (left), and after all baseline cuts (right), for the T5VV signal model.	241
D.18	Comparison of the distributions of NJets between the official and our own samples after the “n-1” cut, Min $\Delta(\phi)$ (left), and after all baseline cuts (right), for the T5VV signal model.	242
D.19	Additional checks: comparison between ours and the official distributions of NJets after $\text{BL}+H_T>1250$ cuts (left), and H_T after $\text{BL}+\text{NJets}>7$ cuts (right), for the T5VV signal model.	242
D.20	Diagram of the dominant SUSY production mechanism for the T2qq signal model.	243
D.21	Comparison of the distributions of H_T between the official and our own samples after the “n-1” cut, Min $\Delta(\phi)$ (left), and after all baseline cuts (right), for the T2qq signal model.	245
D.22	Comparison of the distributions of H_T^{miss} between the official and our own samples after the “n-1” cut, Min $\Delta(\phi)$ (left), and after all baseline cuts (right), for the T2qq signal model.	245
D.23	Comparison of the distributions of NJets between the official and our own samples after the “n-1” cut, Min $\Delta(\phi)$ (left), and after all baseline cuts (right), for the T2qq signal model.	246
D.24	Additional checks: comparison between ours and the official distributions of NJets after $\text{BL}+H_T>1250$ cuts (left), and H_T after $\text{BL}+\text{NJets}>7$ cuts (right), for the T2qq signal model.	246
D.25	The 95% CL exclusion limits (in red) reproduced with our MADANALYSIS 5 implementation compared to the official limits (in black). Top left: T1qqqq, top right: T1tttt, bottom left: T5VV, bottom right: T2qq simplified models.	248
D.26	Diagram for the T2tt SMS topology. Several mass combinations of the stop and LSP are used in the tables below as benchmark comparison scenarios.	249
D.27	MA5 and CMS unit-normalized kinematic distributions after the baseline selection for the T2tt working point (350,0).	253

D.28	MA5 and CMS unit-normalized kinematic distributions after the baseline selection for the T2tt working point (500,100).	255
D.29	MA5 and CMS unit-normalized kinematic distributions after the baseline selection for the T2tt working point (650,50).	257

ABSTRACT

An interpretation of CMS searches for evidence of supersymmetry in the context of the minimal supersymmetric Standard Model (MSSM) is given. It is found that supersymmetric particles with color charge are excluded in the mass range below about 400 GeV, but neutral and weakly-charged sparticles remain non-excluded in all mass ranges. Discussion of the non-excluded regions of the model parameter space is given, including details on the strengths and weaknesses of existing searches, and recommendations for future analysis strategies. Advancements in the modeling of events arising from quantum chromodynamics and electroweak boson production, which are major backgrounds in searches for new physics at the LHC, are also presented. These methods have been implemented as components of CMS searches for supersymmetry in proton-proton collisions resulting in purely hadronic events (i.e., events with no identified leptons) at a center of momentum energy of 13 TeV. These searches, interpreted in the context of simplified models, exclude supersymmetric gluons (gluinos) up to masses of 1400 to 1600 GeV, depending on the model considered, and exclude scalar top quarks with masses up to about 800 GeV, assuming a massless lightest supersymmetric particle. A search for non-excluded supersymmetry models is also presented, which uses multivariate discriminants to isolate potential signal candidate events. The search achieves sensitivity to new physics models in background-dominated kinematic regions not typically considered by analyses, and rules out supersymmetry models that survived 7 and 8 TeV searches performed by CMS.

CHAPTER 1

INTRODUCTION

A series of discoveries spanning three centuries has brought to light the existence of 25 elementary particles. These particles, along with the forces that govern their interactions, are described by a mathematical model many consider to be one of the great intellectual achievements of the 20th century. The model, called the Standard Model of particle physics (for a thorough review, see Ref. [10]), has enabled scientists to understand the process by which the sun radiates light, the forces that bind together solid matter, and details about the history of the universe extending back nearly to the beginning of time. The Standard Model has been tested by countless observations from hundreds of astronomical and collider experiments, and its predictions have been borne out almost without exception. But despite its extraordinary success, the Standard Model is not believed to give a complete picture of nature.

There are three reasons the Standard Model is thought to be incomplete. The first arises from a handful of laboratory and astronomical findings that are inconsistent with predictions. These findings include the discovery of neutrino masses [11], as well as the measured magnetic dipole moment of the muon [12]. The second is the observation of so-called emergent phenomena, behaviors of systems that become manifest as the number of particles making up the systems grows. *More is Different* [13] describes several examples of emergent phenomena, including the appearance of high-temperature superconductivity, as well as the quantum Hall effect, suggesting that the reductionist Standard Model is of limited predictive power in large systems. The third reason the Standard Model is believed to be incomplete is the presence of fine tuning; that is, patterns that cannot not be explained without the addition of some new physics beyond the Standard Model (BSM). These patterns, which are embedded in the Standard Model, perhaps hint at the existence of some new structure, like supersymmetry, or at the existence of many other universes. The latter explanation is

sometimes viewed as non-scientific because it doesn't appear to make testable predictions, but the former, as will be seen, is well-motivated, and can be tested by studying the collisions of highly energetic particles. To illuminate these arguments, it is necessary to first explain what exactly the Standard Model is.

The Standard Model is an example of a general class of theories called quantum field theories. The task of Chapter 2 of this dissertation is to define quantum field theory, providing an amount of detail needed to define and contextualize the nomenclature used throughout the dissertation, and to provide enough detail as is necessary to then describe the Standard Model in Chapter 3. A reader who is already familiar with QFT and the Standard Model may choose to skip these chapters. Following the description of the Standard Model is a discussion of supersymmetry (SUSY) that focuses on how SUSY might manifest itself in the context of the Standard Model. Then, sections describing two experimental facilities, the CERN Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) detector, describe how these facilities operate and why they are ideal for testing most variants of the hypothesis of supersymmetry. Embedded within the CMS chapter are subsections detailing work that has been carried out to examine, improve and maintain the performance of particle detectors.

The remainder of the dissertation consists of sections describing the results of work carried out in search of evidence of supersymmetric phenomena in the particle collisions observed and recorded by the CMS detector during the 2015 run of the LHC (Run 2). These sections include: a detailed analysis quantifying how searches for supersymmetry at CMS have constrained our knowledge of supersymmetry; a discussion of data analysis techniques that increase the sensitivity of BSM searches performed at CMS, including developments in the modeling of Standard Model background events arising from quantum chromodynamics (QCD) interactions as well as interactions that produce electroweak bosons; and a demonstration of new methods that may allow LHC searches to detect certain models of supersymmetry using subsets of data that are not traditionally considered for analysis. Finally, a summary is given on work that has been performed outside the CMS collaboration to create and support a public database of LHC BSM physics analyses. The database con-

tains publicly-available implementations of CMS analyses that can be used by independent theorists to constrain BSM models.

CHAPTER 2

QUANTUM FIELD THEORY

Quantum field theory (QFT) is the framework that fully describes a system of particles that are behaving according to the principles of quantum mechanics and special relativity. Quantum mechanics *alone* is known to accurately describe systems whose constituent objects are very small—much smaller than their wave functions [14]—but its accuracy breaks down as soon as those objects begin to travel at close to the speed of light, c . On the other hand, special relativity *alone* will accurately describe a system of objects traveling at any speed, including those near c , but it breaks down when the objects are very small. QFT is the theory that remains valid in the limit of both high object speed and small object size. Quantum field *theories* are specific manifestations of QFT that may describe real systems like the vacuum state of the universe (i.e., the standard model) or quantum many-body systems (i.e., a doped lattice), or may describe systems that have been dreamed up for the purpose of studying and understanding QFT more generally.

2.1 Amplitudes

From an experimental point of view, QFT is an algorithm that predicts amplitudes. An amplitude, M_{fi} , is a number that relates to the probability for a system to evolve from a specified initial state, labeled as $|i\rangle$, into a specified final state labeled as $|f\rangle$. For concreteness, the initial state could be two highly energetic protons (hydrogen nuclei) making their way toward a head-on collision, and the specified final state could be two protons and a photon (a particle of light) leaving the scene of the collision a moment later (See Fig. 2.1). Or, the initial state could be you reading this sentence at time $t=0$, and the final state could be you taking a sip of coffee a moment later. The probability that the system evolves from

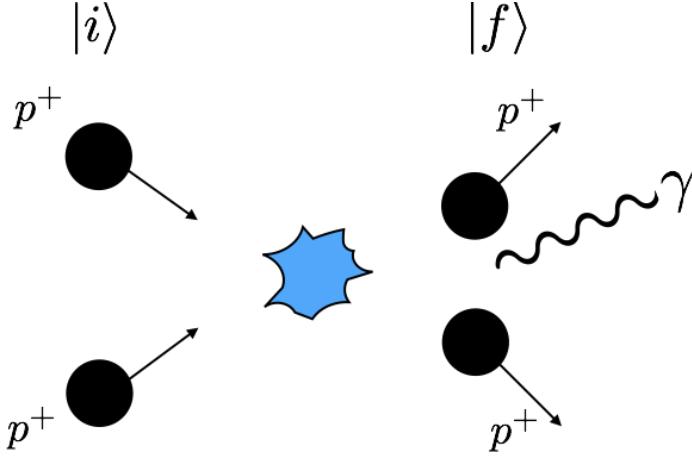


Figure 2.1: An example of an initial state $|i\rangle$ with two incoming protons and a final state $|f\rangle$ with two outgoing protons and a photon γ . The blue object represents the interaction that takes place during the collision.

state $|i\rangle$ into $|f\rangle$ is equal to the squared modulus of the amplitude,

$$P(i \rightarrow f) = |\mathcal{M}_{fi}|^2, \quad (2.1)$$

and is in fact the only observation about a theory that can be made in physics experiments. After an initial and a final state are chosen, $P(i \rightarrow f)$ can be computed for a given QFT hypothesis and compared to the $P(i \rightarrow f)$ observed in experiment. If the two answers are (not) consistent, evidence can begin to build for (against) the hypothesis. Commonly, $P(i \rightarrow f)$ is computed for two or more competing hypotheses, and the degree of consistency between each hypothesis and the experimental result are compared. In this way, a given hypothesis may be “favored by the data” relative to some other hypothesis, an idea that is further discussed in Chapter 7. Clearly, having a way to predict amplitudes is of utmost experimental importance. However, since most physicists are ultimately interested in understanding the physics that underlies the amplitude calculation—in other words, the physics that describes nature more deeply—a bit more detail is called for.

2.2 Fields

The primary degrees of freedom of a QFT are the fields. Mathematically, fields are single- or vector-valued functions of the spacetime coordinates, time and position in space. Physically, fields are interpreted as invisible objects that fill space, taking on a density that varies from point to point, and sometimes an associated direction that likewise varies from point to point (Fig. 2.2). A field can carry any number of properties, including energy, momentum, angular momentum, and charge density. When two or more fields interact, the relative local densities of the fields undergo changes, which can result in the exertion of forces or the creation of matter.

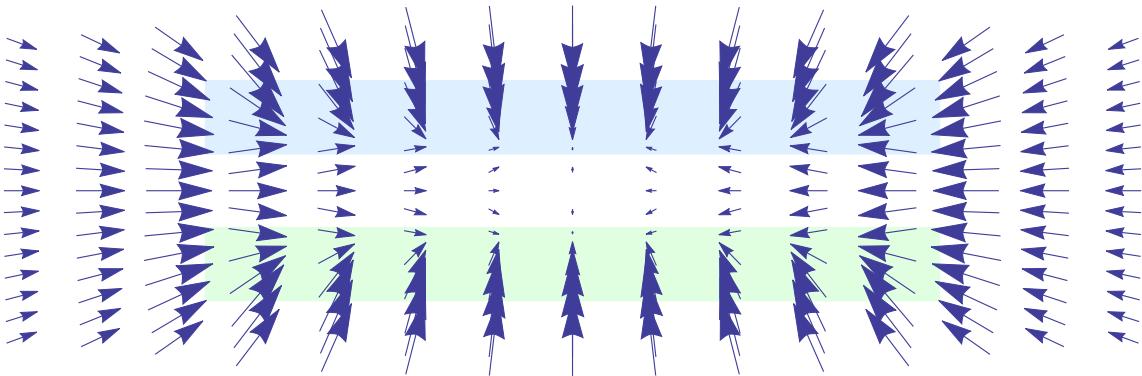


Figure 2.2: An example of a field. The dark blue arrows show the magnitude and direction of the gravity field between two massive parallel plates shown in light blue and green.

2.3 Symmetry

The way fields interact with one another depends on the set of symmetries that are present in the theory. Here, a symmetry is a property that ensures that some aspect of a system remains unchanged when the system is transformed in some way. Certain symmetries are inherent in every QFT, and these are the 10 non-spin related spacetime symmetries corresponding to the 10 generators of the Poincaré group [15]: 3 spatial translations and 1 time translation, 3 rotations, and 3 boosts. Here, boosts are the generators of the Lorentz

group, transforming a system from one inertial reference frame to another. All of these transformations correspond to invariant quantities, giving rise to the laws of the conservation of momentum, energy, angular momentum, and of the spacetime interval defined in special relativity.

Other symmetries are referred to as internal or “gauge” symmetries, and are free to be chosen during the construction of a QFT. An example of a gauge symmetry is the so-called unitary-1, or $U(1)$, symmetry, which is respected by objects that do not change when they undergo planar rotations. In an abstract geometrical sense, circles respect $U(1)$ symmetry because their shapes do not change when they are rotated about their centers by any angle; in the sense of a gauge symmetry, $U(1)$ is notably respected by the QFT known as quantum electrodynamics (QED), the first successful QFT, because QED contains a parameter which, if rotated about the complex unit circle, yields back the same theory.

The transformation of a field under a given symmetry can be represented by an operator \hat{T} , which is an object that changes (operates) on another object. The transformation on a field ϕ can be written as

$$\hat{T}[\phi] = \phi', \quad (2.2)$$

where ϕ is the original field and ϕ' is the transformed field. The criterion for a theory to be invariant under a transformation \hat{T} is that the Lagrangian density be unchanged after the replacement of the old fields by the transformed fields.

2.4 Lagrangian density

The mathematical object that encodes all the information in a perturbative (well-behaved) QFT is the Lagrangian density \mathcal{L} , sometimes just called the Lagrangian. The Lagrangian is a measure of the potential energy of a system subtracted from the kinetic energy, and is composed of terms proportional to the products of the field densities. The exact form of \mathcal{L} follows from the set of fields and symmetries that are chosen in the QFT, and the various terms contain information about different physical processes. As an example, we examine the Lagrangian for a QFT in which there is only one field ϕ , which is single valued (a scalar)

and real. The Lagrangian density for this theory is given by

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - a\phi^2 - b\phi^4, \quad (2.3)$$

where a and b are constants and ∂ is the partial derivative taken with respect to the spatial coordinates. It is apparent that this Lagrangian is an expansion in powers of the field, and this is a general feature of Lagrangians; they are in fact Taylor expansions in the fields.

The concept of symmetry can be illuminated in the context of this Lagrangian, because \mathcal{L} can be seen to respect a particular symmetry called \mathbb{Z}_2 symmetry. \mathbb{Z}_2 symmetry is the invariance of a system under the operation of reversing the sign of the fields. In other words, a Lagrangian that respects a \mathbb{Z}_2 symmetry is invariant under the operation of \hat{T} , where \hat{T} is defined by

$$\hat{T}[\phi] \equiv \phi' = -\phi. \quad (2.4)$$

All terms in this Lagrangian contain even powers of the field ϕ , and so all minus signs brought about by the operation of \hat{T} on \mathcal{L} cancel, giving back the original Lagrangian:

$$\begin{aligned} \hat{T}[\mathcal{L}] \equiv \mathcal{L}' &= -\frac{1}{2}\partial_\mu(-\phi)\partial^\mu(-\phi) - a(-\phi)^2 - b(\phi)^4 \\ &= -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - a\phi^2 - b\phi^4 \\ &= \mathcal{L}. \end{aligned} \quad (2.5)$$

If \mathcal{L} had contained a term proportional to ϕ^3 , the theory would not respect a \mathbb{Z}_2 symmetry, because $\hat{T}[\mathcal{L}]$ would not yield back \mathcal{L} .

As stated, each term in the Lagrangian carries specific physical meaning. The first term in Equation 2.3 accounts for the kinetic energy stored in the field, the second term accounts for particle mass, and the third term is an interaction term. These points are discussed now.

2.5 Particles (quadratic term)

A field typically exhibits a resonance at an energy value proportional to the coefficient of the quadratic term in the Lagrangian; in this case, twice the value of a . This resonance is referred to as a particle, and has a mass defined by the energy of the resonance:

$$m_\phi = 2a. \quad (2.6)$$

Particles are localized quanta, or pieces, of the fields, and they carry the same properties as the fields. These properties include energy E and momentum components p_i ($i = x, y, z$), which can be combined into a single object called a four-vector,

$$p^\mu = (E, p_x, p_y, p_z)^\mu, \quad (2.7)$$

whose magnitude is invariant under Lorentz transformations; particles can also carry charge, a property that is invariant under gauge transformations; and they have spin, where spin is the “intrinsic” angular momentum possessed by a particle in the absence any interaction taking place. The smallest natural increment of spin is equal to half of Planck’s constant \hbar , and thus all particles have a spin equal to an integer multiplied by this unit:

$$s_\phi = N \cdot \frac{\hbar}{2}. \quad (2.8)$$

Particles for which N is an even integer are called bosons, and those with an odd N are called fermions. The bosonic or fermionic nature of particles in a system determines many bulk properties of the system, most importantly whether or not the particles can be arranged to form rigid structures; fermions are able to do so, and bosons are not.

2.6 Feynman diagrams (interaction terms)

As mentioned in Section 2.1, the primary role of QFT from an experimental point of view is to predict amplitudes. To a first approximation, the amplitude for a system to evolve from state $|i\rangle$ to state $|f\rangle$ is given by the quantity

$$M_{fi} = \langle f | \hat{H} | i \rangle, \quad (2.9)$$

where the time evolution operator \hat{H} can be derived from the Lagrangian. There are various methods for computing this amplitude, but the most widely used is Feynman’s method. Feynman’s method amounts to doing a perturbative expansion over the possible ways (modes) for the state $|i\rangle$ to transform into state $|f\rangle$. Each possible mode is weighted by a number that can be computed from a diagram, and after the partial amplitudes for all

possible diagrams have been computed, they are added together to form the amplitude M_{fi} . By Equation 2.1 the square of the amplitude yields the probability for state $|i\rangle$ to evolve into state $|f\rangle$.

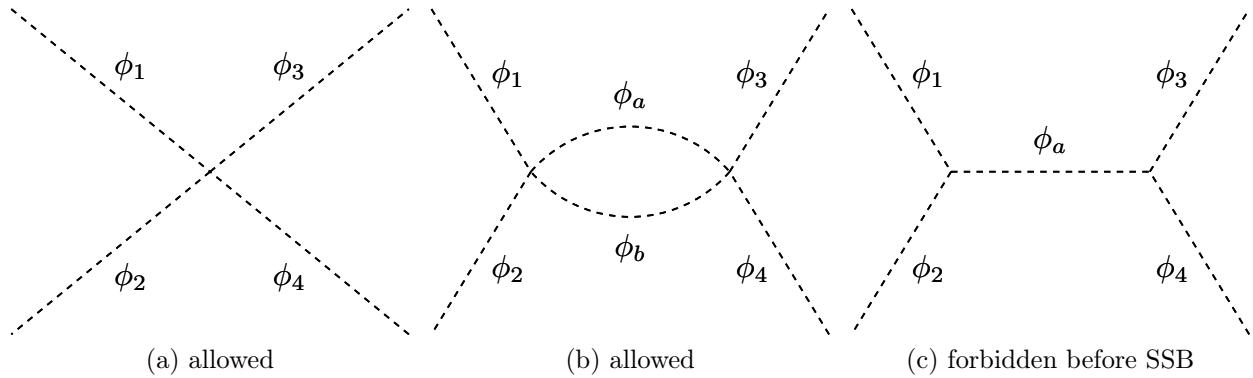


Figure 2.3: Diagrams a and b are two of the leading Feynman diagrams in the theory defined by Equation 2.3. Diagram c is forbidden by the theory, and so does not contribute to the amplitude.

Suppose, for example, we are working within the theory defined by the Lagrangian in Equation 2.3, that the initial state $|i\rangle$ consists of two identical incoming particles ϕ_1 and ϕ_2 , and that the final state $|f\rangle$ consists of two outgoing particles ϕ_3 and ϕ_4 . Figure 2.3 shows two of the leading Feynman diagrams (a and b), each representing a possible mode for the specified process, and one diagram c that is not allowed by the theory. Each allowed diagram contributes a partial amplitude, and the sum of all partial amplitudes is the total amplitude M_{fi} . The two essential features of these and any Feynman diagrams are:

- lines, which represent particles of the field ϕ , where dashes signify that in this case the particles have no spin, and
- vertices, which are the fundamental interactions between particles.

By convention, time flows horizontally, and so the initial state consists of the particles on the left side of a diagram, and the final state consists of the particles on the right side. Diagrams a and c are called tree level diagrams because they appear to have branches that do not close to form loops. Diagram b, by contrast, contains an internal loop of particles, and so is not

a tree-level diagram. Tree-level diagrams typically yield the largest partial amplitudes, and thus are often the most important.

The guiding principle for constructing diagrams is that any diagram not forbidden by the theory will, in principal, contribute to the total amplitude. Allowed diagrams are those for which every vertex within a diagram is associated with a term in the Lagrangian. For our \mathcal{L} , the only interaction term is the quartic term, $c\phi^4$, and this term gives rise to the four-particle vertex seen in the center of the diagram in Fig. 2.3 (a). This can be generalized, since a term in a Lagrangian proportional to n powers of a field gives rise to an n -particle vertex in the theory. The more complicated diagram in Fig. 2.3 (b) is built out of two copies of the four-particle vertex, and so is allowed. One could imagine a vertex of only three ϕ particles, shown in Fig. 2.3 c. However, the three-particle vertex does not correspond to any term in the Lagrangian. For such a vertex to exist, \mathcal{L} would have to contain a term proportional to ϕ^3 , and it was already established that such a term would violate the \mathbb{Z}_2 symmetry respected by the theory. Therefore, the three-particle vertex, and diagram c, are forbidden by symmetry, and so do not enter into the amplitude calculation.

In principle, one should construct and compute all possible diagrams and amplitudes, of which there are an infinite number, but this turns out to be unnecessary if the goal is to make reasonably accurate predictions. The simplest diagrams—the tree-level diagrams—typically give the largest contribution to the amplitude because they have fewer vertices than do more complicated diagrams. Each vertex contributes a factor of $4! \cdot b$ to the partial amplitude, and provided that $4! \cdot b$ is always less than 1, diagrams with more vertices yield smaller partial amplitudes. Taking an approximation where only the tree-level diagrams (diagram a) contribute, the amplitude and probability are given by¹:

$$M_{fi} = -(4!) b, \quad (2.10)$$

$$P(i \rightarrow f) = (-4!)^2 b^2 = 576 b^2. \quad (2.11)$$

The quantity $(4!)b$ is sometimes referred to as a coupling constant, λ , named so because its value dictates the strength of the amplitude associated with the interaction. A couple

¹For a complete description of the Feynman rules, see [15].

of observations: first, the amplitude contains a single power of the constant b , because the contributing diagram has only a single vertex. Second, a notable feature of the probability is that $P(i \rightarrow f)$ is constant with respect to all kinematic properties of the outgoing particles. This means that if many collisions were to occur in succession, the distribution of the final state particles would be isotropic. Of course, this is a conclusion reached by analyzing the leading order contributions. If diagram b or other higher-order diagrams had been included, the prediction of isotropic final states might not have been absolute, but any deviation from this behavior would be suppressed by additional factors of b .

2.7 Conserved quantities

Noether's theorem [16] states that every continuous symmetry respected by a theory is associated with some conserved quantity, and the same is often true in the case of discrete symmetries. It was established that the QFT at hand respects a \mathbb{Z}_2 symmetry, so what is the associated conserved quantity? The answer can be deduced through a bit of analysis of the interaction terms of the Lagrangian. In \mathcal{L} , terms with an even number of powers of the field are present, which means that vertices involving an even number of particles are allowed. However, terms with odd powers of the fields are not present, which means that vertices with an odd number of particles are not allowed. This fact means that while it is possible for a pair of particles to be created or destroyed, the same is not true for a single particle. If the initial number of particles is even, the final number will also be even; if the initial number is odd, the final number will be odd. The conserved quantity is:

$$\theta = n \pmod{2}, \quad (2.12)$$

where n is the number of particles in the system. In this theory, the value of θ for an isolated system can never change.

2.8 Symmetry breaking

The previous analysis of the scalar Lagrangian was valid but not unique. It was assumed that the parameters a and b were positive, but it happens that the dynamics of the theory

change considerably if the parameter a takes on a negative value. Since the Lagrangian is constructed as the difference between the kinetic and potential energy of the system, the second and third term define a potential energy V :

$$V(\phi) = a\phi^2 + b\phi^4. \quad (2.13)$$

For positive a , $V(\phi)$ exhibits a minimum at the value $\phi = 0$, as seen in Fig. 2.4 (a) and (b). But if a is negative, $V(\phi)$ exhibits two minima, located at

$$\phi_{min} \equiv \text{VEV} = \pm \sqrt{\frac{a}{2b}}, \quad (2.14)$$

where VEV is the vacuum expectation value, that is, the average value of the field in the vacuum.

Physical systems tend to evolve towards a state of lower energy, and so it is no surprise that the appearance of these energy minima leads to a qualitative change in the behavior of the system. The field density will naturally settle into one of these minima, located at an energy value equal to the VEV. When this happens, a visible change takes place at the level of the Lagrangian. The field ϕ splits into a static component equal to the VEV, and a dynamic part, which is the modified field φ :

$$\phi \rightarrow \text{VEV} + \varphi. \quad (2.15)$$

The Lagrangian in Equation 2.3 becomes

$$\begin{aligned} \mathcal{L} &\rightarrow -\frac{1}{2}(\partial_\mu[\text{VEV} + \varphi])^2 - a[\text{VEV} + \varphi]^2 - b[\text{VEV} + \varphi]^4 \\ &= \frac{1}{2}(\partial_\mu\varphi)^2 - (a\text{VEV}^2 + b\text{VEV}^4) - (2a\text{VEV} + 4b\text{VEV}^3)\varphi \\ &\quad - (a + 6b\text{VEV}^2)\varphi^2 - 4b\text{VEV}\varphi^3 - b\varphi^4 \equiv \mathcal{L}_{\text{broken}}. \end{aligned} \quad (2.16)$$

The terms without any powers of φ are non-physical and so can be ignored, and through a renaming of the constants, the Lagrangian can be written as

$$\mathcal{L}_{\text{broken}} = \frac{1}{2}(\partial_\mu\varphi)^2 - \alpha\varphi^2 - \gamma\varphi^3 - \beta\varphi^4. \quad (2.17)$$

Note that $\mathcal{L}_{\text{broken}}$ closely resembles the original Lagrangian \mathcal{L} , but differs in a few key ways. For one, the coefficients of the quadratic and quartic terms have changed from a and b to

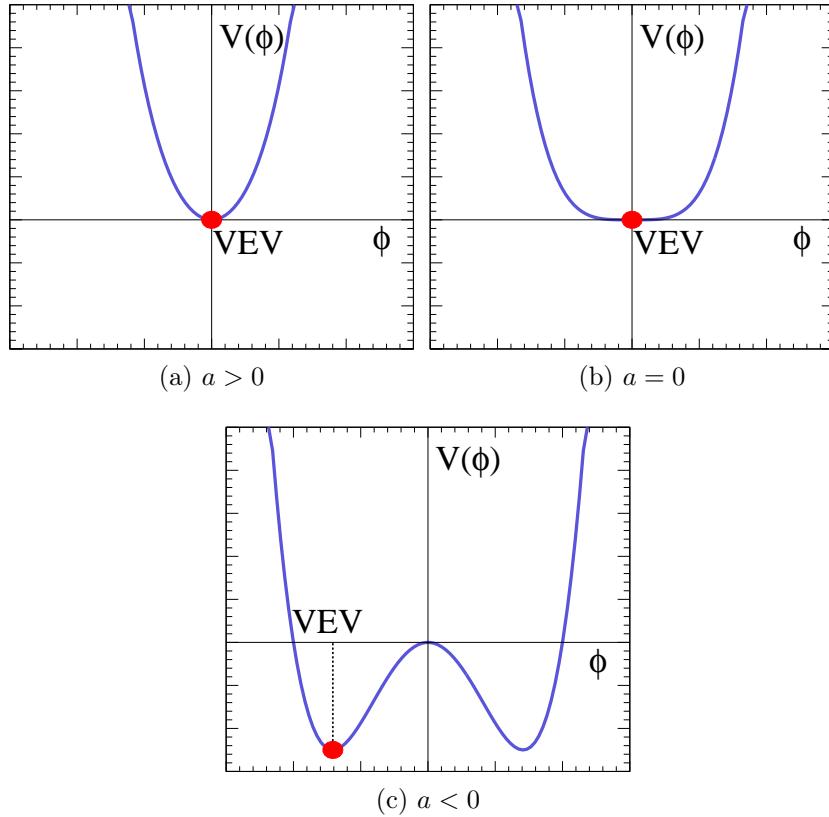


Figure 2.4: The shape of the potential for three choices of parameter a : positive, zero, and negative. The field ϕ always centers on a local minimum of the potential, which is marked in the three cases by the red dot.

α and β , where the latter depend on the magnitude of the symmetry breaking (the VEV). This means that the mass and coupling strength of the four-particle vertex have changed, which will ultimately result in a re-scaling of $P(i \rightarrow f)$.

Then, the fact that there is now a cubic term ($\gamma\varphi^3$) is particularly significant. In \mathcal{L} , the cubic term was explicitly forbidden by the \mathbb{Z}_2 symmetry, but upon the introduction of a non-zero VEV, \mathbb{Z}_2 is no longer respected. The symmetry is said to have been “spontaneously broken.” As a result of the emergent cubic term, the theory has a new allowed vertex of three particles, and so the interaction in the diagram of Fig. 2.3 c is no longer forbidden! This changes the prediction for the amplitude, which now has four contributing tree-level diagrams instead of just one. The probability now depends on the the kinematics of the

particles, and so the outgoing particles will no longer be distributed isotropically. Clearly, symmetry breaking has changed the physics considerably.

Perhaps the most striking change brought about by the breaking of the \mathbb{Z}_2 symmetry is that the previously conserved quantity θ (in Equation 2.12) is no longer conserved. Through processes involving the new three-particle vertex, the system can now start out with an even number of particles and end up with an odd number, or vice versa. Indeed, symmetry breaking is seen to have fundamentally changed the physics.

The concept of spontaneous symmetry breaking (SSB) is particularly relevant for the standard model because it is key to the mechanism by which all the particles in the standard model acquire mass, known as the Brout-Englert-Higgs mechanism [17]. Before proceeding to describe the standard model, there is one last concept that must be mentioned, and that is of commutation relations.

2.9 Commutation relations

Symmetries are defined by an invariance under the action of an operator, and operators are defined by a set of commutation relations. In other words, they are defined by how they transform the states they act on, and by how they behave when applied in sequence with other operators. In general, the commutator and anticommutator of two operators \hat{T}_1 and \hat{T}_2 are defined by

$$[\hat{T}_1, \hat{T}_2]\phi \equiv \hat{T}_1(\hat{T}_2\phi) - \hat{T}_2(\hat{T}_1\phi) \quad (2.18)$$

$$\{\hat{T}_1, \hat{T}_2\}\phi \equiv \hat{T}_1(\hat{T}_2\phi) + \hat{T}_2(\hat{T}_1\phi), \quad (2.19)$$

respectively. Broadly speaking, boson fields commute, meaning their commutators are equal to 0, and fermion fields anticommute, meaning their anticommutators are equal to 0. For more information on commutators and QFT, please see Ref. [15].

The reader has now received an introduction to the basic concepts and vocabulary to understand the following discussion about the standard model and the search for supersymmetry.

CHAPTER 3

THE STANDARD MODEL

The standard model (SM) is the quantum field theory that describes the fundamental fields and symmetries observed in nature thus far. These fields, artistically represented in Figure 3.1, and symmetries are discussed here.

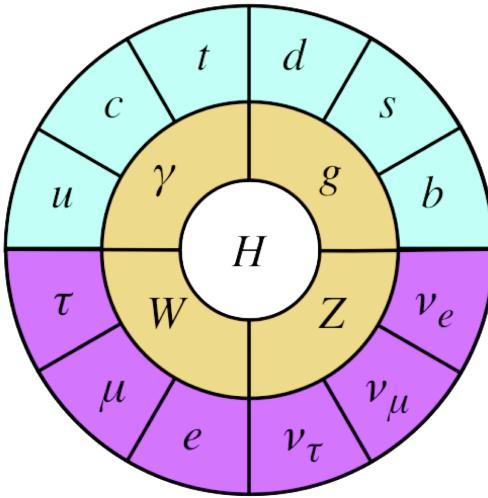


Figure 3.1: The fields of the standard model. From inside out, concentric annuli represent the spin-0, spin-1, and spin- $\frac{1}{2}$ fields.

3.1 SM fields and symmetries

Like all QFTs, the standard model is symmetric under transformations of the 10 non-spin related spacetime degrees of freedom, which map to the 10 generators of the Poincare group.

In addition to the usual spacetime symmetries, the standard model exhibits invariance under transformations of three gauge groups, each group being associated with a set of massless spin-1 (vector) fields:

- the circle group $U(1)_Y$, whose transformation can be represented by a unitary scalar operator (complex one) multiplied by a charge called the weak hypercharge Y . The associated vector field is called B_μ ;
- the $n = 2$ special unitary group $SU(2)_L$, whose three transformations can be represented by the 2×2 Pauli matrices σ_i ($i = 1, 2, 3$) multiplied by a charge called the weak isospin T_3 . The three associated vector fields are $W_\mu^{1,2,3}$;
- the $n = 3$ special unitary group $SU(3)_C$, whose eight operations can be represented by the 3×3 Gell-Mann matrices λ^j ($j = 1, \dots, 8$) multiplied by a charge called the color $C = (r, g, b)$. The eight associated vector fields are $G_\mu^{1,\dots,8}$.

In total, that makes 12 vector fields associated with three gauge symmetries, all of which can be summarized by the tensor product of groups that defines the gauge symmetry of the standard model:

$$U(1)_Y \times SU(2)_L \times SU(3)_C. \quad (3.1)$$

The standard model also has 12 spin-1/2 (fermion) fields ψ , sometimes referred to as “matter fields,” 6 of which are the lepton fields and 6 are quark fields. Of the lepton fields, there are three charged fields, electron (e^-), muon (μ^-), and tau (τ^-), and three neutral fields, electron neutrino (ν_e), muon neutrino (ν_μ), and tau neutrino (ν_τ). Of the quark fields, there are three “up-type” fields, up (u), charm (c), and top (t), and three “down-type” fields, down (d), strange (s), and bottom (b). Fermions contain a property called chirality, and they can either be left-chiral, or right-chiral, which are sometimes referred to as left- or right-handed. For mysterious but experimentally verified reasons, left-handed fermion fields transform differently under the $SU(2)_L$ gauge symmetry than do the right-handed fermion fields, which results in an asymmetry in the properties of fermions; specifically, there are left-handed and right-handed charged leptons, left-handed neutrinos, but no right-handed neutrinos.

The last field in the standard model is a complex scalar doublet field (ϕ), named the Higgs field after one of the theorists who predicted its existence in 1964 [17].

3.2 SM Lagrangian

The standard model Lagrangian [18] can be written in a somewhat simplified form as

$$\begin{aligned}\mathcal{L}_{\text{SM}} = & \frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{2}\mu^2\phi^2 - \frac{\lambda}{4}\phi^4 + \frac{1}{4}g_1^2(B_\mu)^2\phi^2 + \frac{1}{4}\sum_{i=1}^3 g_2^2(\sigma^i W_\mu^i)^2\phi^2 \\ & + \sum_{i,j=1}^9 \psi_i y_{ij} \psi_j \phi + \sum_{i=1}^8 \frac{1}{4} \left[\partial_\mu G_\nu^i - \partial_\nu G_\mu^i - \sum_{j,k=1}^8 g_s f_{ijk} G_\mu^j G_\nu^k \right]^2 + \mathcal{L}_{\text{gauge,kin}}.\end{aligned}\quad (3.2)$$

The first three terms are much like the ϕ^4 scalar Lagrangian introduced in Section 2.4, except for one trivial and one non-trivial difference. The trivial difference is that the names of the coefficients on ϕ^2 and ϕ^4 have been tweaked slightly to reflect their true meaning (the scalar mass and a coupling constant, respectively), and the non-trivial difference is that the field ϕ is now a complex scalar doublet rather than a real-valued number, and so $\phi^2 \equiv \phi^\dagger \phi$, where † denotes hermitian conjugation. The fourth and fifth term are determined from invariance under the $U(1)_Y \times SU(2)_L$ (electroweak) symmetry and are the allowed interactions of the W and B vector fields with the scalar field, where g_1 and g_2 are coupling constants. The sixth term features a sum over all pairs of charged fermion fields, and amounts to the set of allowed interaction terms between the fermion fields and the scalar; the constants y_{ij} are called Yukawa couplings. The seventh term defines the properties of the gluon fields, where g_s is the strong coupling constant and f_{ijk} are so-called structure constants determined by the $SU(3)$ symmetry group. The final term, $\mathcal{L}_{\text{gauge,kin}}$ includes terms for the kinetic energy of fermion and gauge boson fields and a few higher-order interaction terms, but is not essential for this discussion.

3.3 Electroweak symmetry breaking

One conspicuous feature of this Lagrangian is the absence of any mass terms apart from the scalar mass term. The symmetry in Expression 3.1 actually forbids quadratic terms for either the fermion or gauge boson fields, and so one might expect this model to describe a system which does not contain massive particles—certainly not our universe. However, as with the unbroken ϕ^4 theory described earlier, this Lagrangian does not explicitly reveal

all the dynamics. The coefficient of the quadratic term (in this case μ^2) is measured to be negative, and this causes the scalar potential $V(\phi)$ to exhibit a minimum value at

$$|\phi|_{min} = \text{VEV} = \sqrt{\frac{-\mu^2}{\lambda}}. \quad (3.3)$$

The universe is thought to have long ago slipped into the ground state with ϕ centered on the vacuum expectation value VEV. In this case, the Lagrangian can be re-expressed in terms of the fields through the replacement

$$\phi \rightarrow \text{VEV} + h,$$

where VEV acts as a constant and h is the new dynamic field, the Higgs field. The above replacement yields a Lagrangian that no longer respects the $U(1)_Y \times SU(2)_L$ symmetry:

$$\begin{aligned} \mathcal{L}_{\text{SM}}' &\rightarrow \mathcal{L}'_{\text{SM}} \\ \mathcal{L}'_{\text{SM}} &= \frac{1}{2}(\partial_\mu[\text{VEV} + h])^2 - \frac{1}{2}\mu^2[\text{VEV} + h]^2 - \frac{\lambda}{4}[\text{VEV} + h]^4 \\ &+ \frac{1}{4}g_1^2(B_\mu)^2[\text{VEV} + h]^2 + \frac{1}{4}\sum_{i=1}^3 g_2^2(\sigma^i W_\mu^i)^2[h + \text{VEV}]^2 \\ &+ \sum_{i,j=1}^9 \psi_i y_{ij} \psi_j [\text{VEV} + h] + \sum_{i=1}^8 \frac{1}{4} \left[\partial_\mu G_\nu^i - \partial_\nu G_\mu^i - \sum_{j,k=1}^8 g_s f_{ijk} G_\mu^j G_\nu^k \right]^2 \\ &+ \mathcal{L}_{\text{gauge,kin}}, \end{aligned} \quad (3.4)$$

$$\begin{aligned} &= \frac{1}{2}(\partial_\mu h)^2 + \frac{1}{2}\mu^2 h^2 - \lambda \text{VEV} h^3 - \frac{\lambda}{4} h^4 + \frac{1}{4}g_1^2 \text{VEV}^2 (B_\mu)^2 \\ &+ \frac{1}{2}g_1^2 \text{VEV} (B_\mu)^2 h + \frac{1}{4}g_1^2 (B_\mu)^2 h^2 + \frac{1}{4}g_2^2 \text{VEV}^2 \sum_{i=1}^3 (\sigma^i W_\mu^i)^2 \\ &+ \frac{1}{2}g_2^2 \text{VEV} \sum_{i=1}^3 (\sigma^i W_\mu^i)^2 h^2 + \frac{1}{4}g_2^2 \sum_{i=1}^3 (\sigma^i W_\mu^i)^2 h^4 + \sum_{i,j=1}^9 \psi_i y_{ij} \psi_j h \\ &+ (\text{VEV}) \sum_{i,j=1}^9 \psi_i y_{ij} \psi_j + \frac{1}{4} \sum_{i=1}^8 \left[\partial_\mu G_\nu^i - \partial_\nu G_\mu^i - \sum_{j,k=1}^8 g_s f_{ijk} G_\mu^j G_\nu^k \right]^2 \\ &+ \mathcal{L}_{\text{gauge,kin}}. \end{aligned} \quad (3.5)$$

The terms highlighted in blue are quadratic in the fields and indicate that several standard model particles have “acquired” mass through their interactions with the Higgs field. The third blue term indicates that the 9 charged standard model fermions each now have an associated mass equal to the VEV multiplied by their respective Yukawa coupling. Similarly, the first and second blue terms show that the gauge fields B and W are associated with masses as well; the electroweak symmetry has been spontaneously broken. Note that there are still no terms quadratic in the gluon fields G , and so the gluons remain massless; the $SU(3)_C$ symmetry remains unbroken. We are left with a model that more closely resembles nature, but there is one stone left unturned. It is known that nature retains one massless electroweak gauge boson, the photon γ , and so it must be that one of the electroweak field terms has a quadratic term equal to zero. This is achieved through a mixing of the B and W fields, wherein new fields are defined as various combinations of the old fields:

$$W_\mu^+ = \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2) \quad (3.6)$$

$$W_\mu^- = \frac{1}{\sqrt{2}}(W_\mu^1 + iW_\mu^2) \quad (3.7)$$

$$Z_\mu = -\sin\theta_W B_\mu + \cos\theta_W W_\mu^3 \quad (3.8)$$

$$A_\mu = +\cos\theta_W B_\mu + \sin\theta_W W_\mu^3, \quad (3.9)$$

where θ_W is the electroweak mixing angle (the Weinberg angle), defined as:

$$\theta_W = \tan^{-1}(g_1/g_2). \quad (3.10)$$

The new gauge fields are associated with masses equal to

$$m_W = \frac{1}{2}\text{VEV}g_2, \quad (3.11)$$

$$m_Z = \frac{1}{2}\text{VEV}\sqrt{g_1^2 + g_2^2}, \quad (3.12)$$

$$m_A = 0. \quad (3.13)$$

The massless field A_μ is identified as the electromagnetic field, or the photon field. This field respects a new $U(1)$ symmetry, the symmetry of electromagnetism, $U(1)_{EM}$, whose

generator of transformations is the well-known electric charge $Q = T_3 + \frac{1}{2}Y$. It is said that the electroweak symmetry breaks to the electromagnetic symmetry:

$$U(1)_Y \times SU(2)_L \times SU(3)_C \rightarrow U(1)_{EM} \times SU(3)_C \quad (3.14)$$

3.4 SM parameters and visualization

A summary of the standard model fields and particles is given in Table 3.1. In total, the

Table 3.1: The fields and particles of the Standard Model and their quantum numbers. The gluon color can take on eight unique combinations of $c_i\bar{c}_j$ ($i, j = 1, 2, 3$) where $c_1 = r$ (red), $c_2 = g$ (green), and $c_3 = b$ (blue).

name	fields	spin	Y	T_3	Q	C	mass states
fermion							
charged lepton	e_L, e_R	$\frac{1}{2}$	$-1, -2$	$-\frac{1}{2}, 0$	-1	0	e^-
	μ_L, μ_R						μ^-
	τ_L, τ_R						τ^-
neutrino	$\nu_{eL}, -$		$-1, -$	$+\frac{1}{2}, -$	$0, -$		ν_1, ν_2, ν_3
	$\nu_{\mu L}, -$						
	$\nu_{\tau L}, -$						
up-type quark	u_L, u_R		$+\frac{1}{3}, +\frac{4}{3}$	$+\frac{1}{2}, 0$	$+\frac{2}{3}$	r, g, b	hadrons (baryons & mesons)
	c_L, t_R						
	t_L, t_R						
down-type quark	d_L, d_R		$+\frac{1}{3}, -\frac{2}{3}$	$-\frac{1}{2}, 0$	$-\frac{1}{3}$	r, g, b	
	s_L, s_R						
	b_L, b_R						
boson							
Higgs	H	0	1	$-\frac{1}{2}$	0	0	h
B boson	B	1	0	0	0	0	γ, Z
W boson	W ³		0	0	0	0	
	W ²		0	$+1(W^+)$	$+1(W^+)$	0	W ⁺ , W ⁻
	W ¹		0	$-1(W^-)$	$-1(W^-)$	0	
gluon	G ^{1,...,8}		0	0	0	$c_i\bar{c}_j$	glueballs

standard model has 19 free parameters, all of which have been measured. They are

- 9 fermion mass parameters, one for each charged fermion;

- 1 Higgs mass parameter m_h (the Higgs boson was discovered and m_h measured in 2012 [19]);
- 1 vacuum expectation value, the VEV;
- 3 field strength constants, g_1 , g_2 , and g_3 ; and
- 5 additional mixing parameters appearing in $\mathcal{L}_{\text{gauge,kin}}$ not discussed herein: θ_{12} , θ_{23} , and θ_{13} , δ , and θ_{QCD} .

The values of these parameters can be represented using the technique of parallel coordinates visualization [20]. In a parallel coordinates plot, independent variables are represented by vertical parallel axes, and points in the hyperspace spanned by the independent variables are represented by horizontal lines traversing the plot from left to right. In Figure 3.2, the standard model is represented as the red line, intersecting each axis at the value of the parameter assumed by the theory, in this case the measured value.

The information in Equation 3.1, Table 3.1, and Figure 3.2 fully defines the standard model.

3.5 Problems with the standard model

As stated earlier, there are three reasons why the standard model is believed not to provide a full explanation of all natural phenomena. The first pertains to certain observations that seem to directly contradict predictions of the standard model. The second is the behavior seen in some systems that, while not directly contradicting the predictions of the standard model, seem to go beyond the standard model's ability to account for them. And finally, there is the puzzling observation that the standard model contains a high degree of fine tuning. While the second issue is of much importance, it is not directly relevant to this work, and so further discussion is left to the literature. However, the following is a discussion of the first and last issue.

3.5.1 Observations in tension with the standard model

The most striking shortcoming of the standard model is that it lacks a description for gravity. Gravity was the first force to be fully understood over large distances, but it will

likely be the last to be fully understood at very short distances. This is because the force of gravity is very weak, as it carries a coupling constant that is 10^{34} times smaller than the electromagnetic coupling α . Gravitational effects would not be observable in particle collisions below a center of mass energy close to the Planck scale (10^{19} GeV), mindbendingly larger than the energetic reach of modern particle colliders. Most theoretical models that could describe the standard model and gravity, sometimes called theories of everything or ultraviolet (UV) completion models, manifest new phenomena only above a large energy called Λ_{UV} , roughly in the vicinity of the Planck scale.

Another observation not predicted by the standard model is that of neutrino mass. The Higgs field does not interact with neutrinos because neutrinos are left-chiral and have no right-chiral counterparts (see Figure 3.1); neutrinos in the standard model are massless. However, neutrinos were recently observed to oscillate between flavors [11], a behavior that is only possible if the neutrinos have mass. This seemingly innocuous observation is in fact quite profound, because it implies that some still unknown theory of particles, with masses of higher energy than the standard model, exists. A number of theories have emerged that could potentially explain neutrino mass, including the types I, II, and III seesaw mechanisms [21], all of which predict heavy right-handed neutrinos whose mass is inversely proportional to the corresponding left-handed neutrino mass. If a lower bound on the masses of the standard model neutrinos were to be measured, this would specify the mass scale of the heavy neutrinos. Unfortunately, no lower bound has been established as yet, and so there are few clues about the nature of the true theory, or at what energy scale the phenomena of the new theory might be observable.

Another piece of evidence that suggests the standard model falls short of providing a full description of nature is the observation of dark matter in galaxies and in the intergalactic regions of the universe [22][23]. Measurements of the baryonic mass of galaxies (the mass due to standard model particles) based on observed galactic luminosities are inconsistent with estimates of the total mass of galaxies based on rates of galactic rotation. The baryonic mass is, on average, a factor 6 lower than the total inferred mass, so it seems that there is a great deal of extra mass in galaxies not accounted for by the known particles. The best-motivated

explanation for this perceived excess is that some yet-undiscovered type of matter exists in abundance within galaxies, but doesn't emit or absorb light (and hence the moniker, "dark matter"), and therefore has, at most, weak interactions.

This inferred dark matter may come in the form of new particles, particles not described by the standard model, which may interact with the standard model particles at very high energy scales. This hypothesis is supported by observations of the speed of objects in galaxies as a function of their distance from the galactic center (see Fig. 3.5 a). The expected speed of objects moving in the nearby galaxy NGC 3198 is given as a function of the orbital radius for the baryonic matter-only hypothesis (solid line peaking at 5 kpc), as well as for the baryonic matter+dark matter hypothesis (top solid line). The data (black points) show excellent agreement with the non-interacting dark matter hypothesis. Furthermore, it is claimed that direct evidence of dark matter is evident in the overlay in Figure 3.5 (b), in which the aftermath of the collision of two galaxy clusters has been observed. The points of maximum baryonic matter density, represented in black, are displaced from the points of maximum total mass density as determined by the magnitude of gravitational lensing, indicated by the red contours. The interpretation is that the collision of the clusters caused the baryonic matter to collide and slow down, while the dark matter passed through the collision without slowing, resulting in the separation of the two populations of matter.

Dark matter may be discovered by direct or indirect detection experiments (see Ref. [24] for a review) or collider experiments like ATLAS [25] and CMS [26]. A number of theories have been proposed as possible explanations of cosmological dark matter, including certain supersymmetric extensions of the standard model (discussed in Chapter 4).

Other observations requiring some explanation beyond the standard model include the fact that there is an abundance of matter in the cosmos but almost no anti-matter (baryon asymmetry), and the existence of the dark energy that accounts for the accelerating expansion of the universe [27].

3.5.2 Fine tuning and unexplained patterns in the standard model

In a scientific context, tuning is the notion that the parameters of a model can either take on seemingly random values, or they can appear to follow some pattern. An un-tuned (or “natural”) model refers to the former, and a highly-tuned (or fine-tuned) model refers to the latter. When an apparent pattern has been observed, centuries of science practice and common sense suggests that there should be a reason for the pattern. When a model seems to be fine-tuned, one of two possible, scientifically acceptable explanations must be the case:

- Either there are enough randomized versions (near copies) of the model that the chance emergence of such a pattern is a statistical inevitability; simultaneously, some kind of selection bias has lead the observer to examine the particular instance of the model that exhibited the pattern; or
- there exists some mechanism that generates the observed pattern.

The standard model has a number of examples of unexplained fine tuning. Notably, it features a pattern of masses that is quite unusual. As discussed, the origin of particle mass originates from the VEV acquiring a non-zero value through spontaneous symmetry breaking. However, although the VEV is measured to be greater than 200 GeV, it is astonishingly small compared with the natural expectation, and the reason is the following. The VEV, while normally viewed as a constant in the Lagrangian, actually carries the quantum numbers of the Higgs scalar field, and so interacts with all massive particles. These interactions lead to Feynman diagrams that contain internal loops of the particles. Such diagrams, as in Fig. 3.6, lead to corrections to the value of the VEV given by

$$\text{fermion: } (\Delta \text{VEV})^2 = -\frac{\lambda_f^2}{8\pi^2} \Lambda_{\text{UV}}^2, \quad (3.15)$$

$$\text{scalar: } (\Delta \text{VEV})^2 = \frac{\lambda_S^2}{16\pi^2} [\Lambda_{\text{UV}}^2 - 2m_S^2 \ln(\Lambda_{\text{UV}}/m_s)], \quad (3.16)$$

where λ_f and λ_S are the Yukawa coupling and scalar coupling constants, and Λ_{UV} is the cutoff energy of the ultraviolet theory. Every massive particle in the standard model, and in principle any other particle that may occupy the electroweak vacuum, contributes such a correction term. Because Λ_{UV} is near the Planck scale, the VEV receives large corrections

that are to 10^{15} times larger than the VEV itself. This series of huge numbers is somehow summing to a small number, in a striking example of fine tuning. There is not as yet a scientifically acceptable explanation, since there is no identified mechanism, and no copies of the standard model have been found.

The unexplained cancelation of the 1-loop contributions to the VEV is known as the hierarchy problem, and it impacts nearly every aspect of the standard model. Were the VEV to assume a natural value, the entirety of the particle spectrum of the standard model would be pushed up to the GUT scale. This mystery is especially relevant in the era of TeV-scale particle colliders like the LHC, because a mechanism that could explain the pattern of cancelations is expected to emerge at the electroweak energy scale. If there is no such mechanism, then we are all but forced to accept the premise that there somehow exists a multitude of versions of the standard model “somewhere” in order to make the emergence of such a pattern statistical inevitable. But even in such a case, it is not clear where these copies are, or what selection bias might be responsible for our observing the pattern that is seen.

3.6 Problems with the standard model

As stated earlier, there are three reasons why the standard model is believed not to provide a full explanation of all natural phenomena. The first pertains to certain observations that seem to directly contradict predictions of the standard model. The second is the behavior seen in some systems that, while not directly contradicting the predictions of the standard model, seem to go beyond the standard model’s ability to account for them. And finally, there is the puzzling observation that the standard model contains a high degree of fine tuning. While the second issue is of much importance, it is not directly relevant to this work, and so further discussion is left to the literature. However, the following is a discussion of the first and last issue.

3.6.1 Observations in tension with the standard model

The most striking shortcoming of the standard model is that it lacks a description for gravity. Gravity was the first force to be fully understood over large distances, but it will likely be the last to be fully understood at very short distances. This is because the force of gravity is very weak, as it carries a coupling constant that is 10^{34} times smaller than the electromagnetic coupling α . Gravitational effects would not be observable in particle collisions below a center of mass energy close to the Planck scale (10^{19} GeV), mindbendingly larger than the energetic reach of modern particle colliders. Most theoretical models that could describe the standard model and gravity, sometimes called theories of everything or ultraviolet (UV) completion models, manifest new phenomena only above a large energy called Λ_{UV} , roughly in the vicinity of the Planck scale.

Another observation not predicted by the standard model is that of neutrino mass. The Higgs field does not interact with neutrinos because neutrinos are left-chiral and have no right-chiral counterparts (see Figure 3.1); neutrinos in the standard model are massless. However, neutrinos were recently observed to oscillate between flavors [11], a behavior that is only possible if the neutrinos have mass. This seemingly innocuous observation is in fact quite profound, because it implies that some still unknown theory of particles, with masses of higher energy than the standard model, exists. A number of theories have emerged that could potentially explain neutrino mass, including the types I, II, and III seesaw mechanisms [21], all of which predict heavy right-handed neutrinos whose mass is inversely proportional to the corresponding left-handed neutrino mass. If a lower bound on the masses of the standard model neutrinos were to be measured, this would specify the mass scale of the heavy neutrinos. Unfortunately, no lower bound has been established as yet, and so there are few clues about the nature of the true theory, or at what energy scale the phenomena of the new theory might be observable.

Another piece of evidence that suggests the standard model falls short of providing a full description of nature is the observation of dark matter in galaxies and in the intergalactic regions of the universe [22][23]. Measurements of the baryonic mass of galaxies (the mass due to standard model particles) based on observed galactic luminosities are inconsistent with

estimates of the total mass of galaxies based on rates of galactic rotation. The baryonic mass is, on average, a factor 6 lower than the total inferred mass, so it seems that there is a great deal of extra mass in galaxies not accounted for by the known particles. The best-motivated explanation for this perceived excess is that some yet-undiscovered type of matter exists in abundance within galaxies, but doesn't emit or absorb light (and hence the moniker, "dark matter"), and therefore has, at most, weak interactions.

This inferred dark matter may come in the form of new particles, particles not described by the standard model, which may interact with the standard model particles at very high energy scales. This hypothesis is supported by observations of the speed of objects in galaxies as a function of their distance from the galactic center (see Fig. 3.5 a). The expected speed of objects moving in the nearby galaxy NGC 3198 is given as a function of the orbital radius for the baryonic matter-only hypothesis (solid line peaking at 5 kpc), as well as for the baryonic matter+dark matter hypothesis (top solid line). The data (black points) show excellent agreement with the non-interacting dark matter hypothesis. Furthermore, it is claimed that direct evidence of dark matter is evident in the overlay in Figure 3.5 (b), in which the aftermath of the collision of two galaxy clusters has been observed. The points of maximum baryonic matter density, represented in black, are displaced from the points of maximum total mass density as determined by the magnitude of gravitational lensing, indicated by the red contours. The interpretation is that the collision of the clusters caused the baryonic matter to collide and slow down, while the dark matter passed through the collision without slowing, resulting in the separation of the two populations of matter.

Dark matter may be discovered by direct or indirect detection experiments (see Ref. [24] for a review) or collider experiments like ATLAS [25] and CMS [26]. A number of theories have been proposed as possible explanations of cosmological dark matter, including certain supersymmetric extensions of the standard model (discussed in Chapter 4).

Other observations requiring some explanation beyond the standard model include the fact that there is an abundance of matter in the cosmos but almost no anti-matter (baryon asymmetry), and the existence of the dark energy that accounts for the accelerating expansion of the universe [27].

3.6.2 Fine tuning and unexplained patterns in the standard model

In a scientific context, tuning is the notion that the parameters of a model can either take on seemingly random values, or they can appear to follow some pattern. An un-tuned (or “natural”) model refers to the former, and a highly-tuned (or fine-tuned) model refers to the latter. When an apparent pattern has been observed, centuries of science practice and common sense suggests that there should be a reason for the pattern. When a model seems to be fine-tuned, one of two possible, scientifically acceptable explanations must be the case:

- Either there are enough randomized versions (near copies) of the model that the chance emergence of such a pattern is a statistical inevitability; simultaneously, some kind of selection bias has lead the observer to examine the particular instance of the model that exhibited the pattern; or
- there exists some mechanism that generates the observed pattern.

The standard model has a number of examples of unexplained fine tuning. Notably, it features a pattern of masses that is quite unusual. As discussed, the origin of particle mass originates from the VEV acquiring a non-zero value through spontaneous symmetry breaking. However, although the VEV is measured to be greater than 200 GeV, it is astonishingly small compared with the natural expectation, and the reason is the following. The VEV, while normally viewed as a constant in the Lagrangian, actually carries the quantum numbers of the Higgs scalar field, and so interacts with all massive particles. These interactions lead to Feynman diagrams that contain internal loops of the particles. Such diagrams, as in Fig. 3.6, lead to corrections to the value of the VEV given by

$$\text{fermion: } (\Delta \text{VEV})^2 = -\frac{\lambda_f^2}{8\pi^2} \Lambda_{\text{UV}}^2, \quad (3.17)$$

$$\text{scalar: } (\Delta \text{VEV})^2 = \frac{\lambda_S^2}{16\pi^2} [\Lambda_{\text{UV}}^2 - 2m_S^2 \ln(\Lambda_{\text{UV}}/m_s)], \quad (3.18)$$

where λ_f and λ_S are the Yukawa coupling and scalar coupling constants, and Λ_{UV} is the cutoff energy of the ultraviolet theory. Every massive particle in the standard model, and in principle any other particle that may occupy the electroweak vacuum, contributes such a correction term. Because Λ_{UV} is near the Planck scale, the VEV receives large corrections

that are to 10^{15} times larger than the VEV itself. This series of huge numbers is somehow summing to a small number, in a striking example of fine tuning. There is not as yet a scientifically acceptable explanation, since there is no identified mechanism, and no copies of the standard model have been found.

The unexplained cancelation of the 1-loop contributions to the VEV is known as the hierarchy problem, and it impacts nearly every aspect of the standard model. Were the VEV to assume a natural value, the entirety of the particle spectrum of the standard model would be pushed up to the GUT scale. This mystery is especially relevant in the era of TeV-scale particle colliders like the LHC, because a mechanism that could explain the pattern of cancelations is expected to emerge at the electroweak energy scale. If there is no such mechanism, then we are all but forced to accept the premise that there somehow exists a multitude of versions of the standard model “somewhere” in order to make the emergence of such a pattern statistical inevitable. But even in such a case, it is not clear where these copies are, or what selection bias might be responsible for our observing the pattern that is seen.

3.7 This document

The subject of this dissertation is a symmetry, that, if realized in the context of the Standard Model, may resolve a number of the above mentioned riddles, including the hierarchy problem. The symmetry, known as supersymmetry [28], is the only remaining spacetime symmetry not yet found in nature. Supersymmetry offers a potentially viable explanation for the existence of dark matter, and if ever discovered, could motivate certain models of grand unification, theories of everything, and string theories.

This document summarizes work that has been carried out to shed light on the question of whether a supersymmetric extension of the Standard Model is possible given the observations made so far by the Compact Muon Solenoid (CMS), and if not, whether a definitive answer can be expected from analysis of current and future experimental data. Additionally, analysis techniques are discussed that have been applied to the Run 2 (2015) dataset. I have authored the text in a style suitable for a graduate student interested in contributing to the effort of

globally understanding the experimental viability of supersymmetry, and in developing new techniques to search for supersymmetry at the Large Hadron Collider (LHC).

The body of this document is organized as follows. I proceed with a brief description of supersymmetry and how it may manifest itself at the LHC. Chapter 5 gives a description of the LHC, and Chapter 6 provides a description the Compact Muon Solenoid (CMS). During the discussion of the CMS detector, I make short forays into studies and techniques I have developed while contributing to a number of technical upgrades within CMS.

In Chapter 7, I describe an analysis [29] I helped perform within CMS to understand how the hypothesis of supersymmetry is constrained by the data that has been collected and analyzed by CMS and other experiments. The analysis is a global interpretation of a representative set of CMS SUSY searches in the context of a general model of supersymmetry that is a proxy for the 120-parameter model called the minimal supersymmetric standard model. The submodel is called the phenomenological MSSM (pMSSM). The goal of the chapter is to summarize what lessons have been learned about the viability of supersymmetric hypotheses based on LHC results.

In Chapter 8, I proceed to describe various analysis techniques that I have developed to improve the sensitivity of CMS searches to supersymmetry models that have not been excluded by the CMS Run 1 (2011-2012) analyses. This includes novel methods for modeling of Standard Model QCD and $Z \rightarrow \nu\bar{\nu}$ processes, which are dominant sources of background in searches that look for evidence of supersymmetry, as well as novel techniques for modeling triggers (triggers are defined in Chapter 6) useful for analyses that are brave enough to hunt for supersymmetry in kinematic regions where the background is dominant. I also give key details about two 2015 and 2016 analyses that I contributed to while developing the mentioned techniques [6][7]. The chapter ends with a proof of principal demonstration of a bold strategy for searching for evidence of supersymmetry, namely, the feasibility of employing multivariate discriminants to maximize sensitivity to models of supersymmetry.

Chapter D summarizes contributions I made to a project [30] outside the CMS collaboration that has the goal of making the results of CMS and ATLAS SUSY searches more accessible to the theoretical physics community.



Figure 3.2: A parallel coordinates plot of the 19 standard model parameters. Values on the black axes have dimensions of energy and are read from the left, while the values on the blue axes are read from the right.

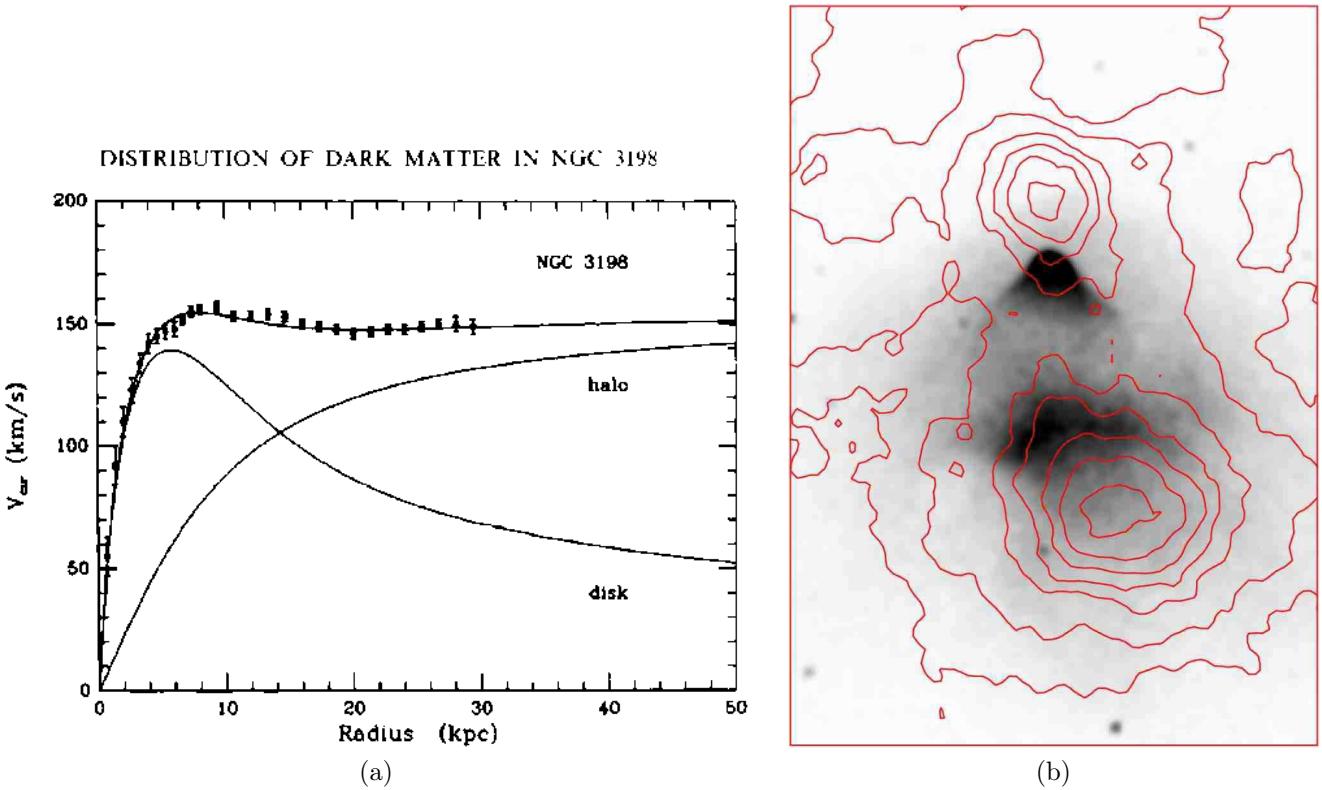


Figure 3.3: Subfigure (a) [2] shows the orbital velocity of objects in galaxy NGC 3198 vs their distance from the galactic center. Subfigure (b) [3] shows an X-ray photograph taken by the Chandra X-ray observatory of the aftermath of the collision of two galaxy clusters, 1E0657-558, overlaid on a contour map of the gravitation lensing magnitude factor κ derived from a separate image taken by the Magellan 6.5 m telescope.

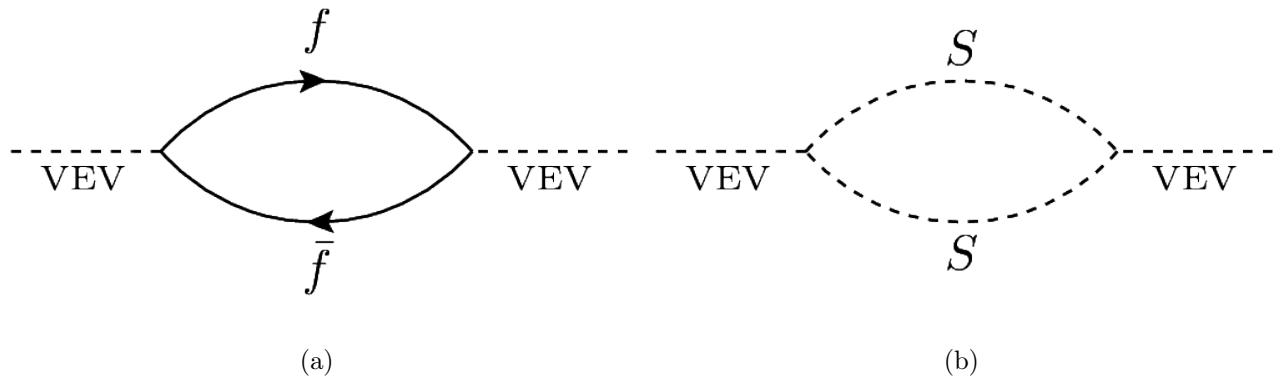


Figure 3.4: Diagrams for the one-loop corrections to the VEV from a generic fermion (a) and a generic scalar (b).

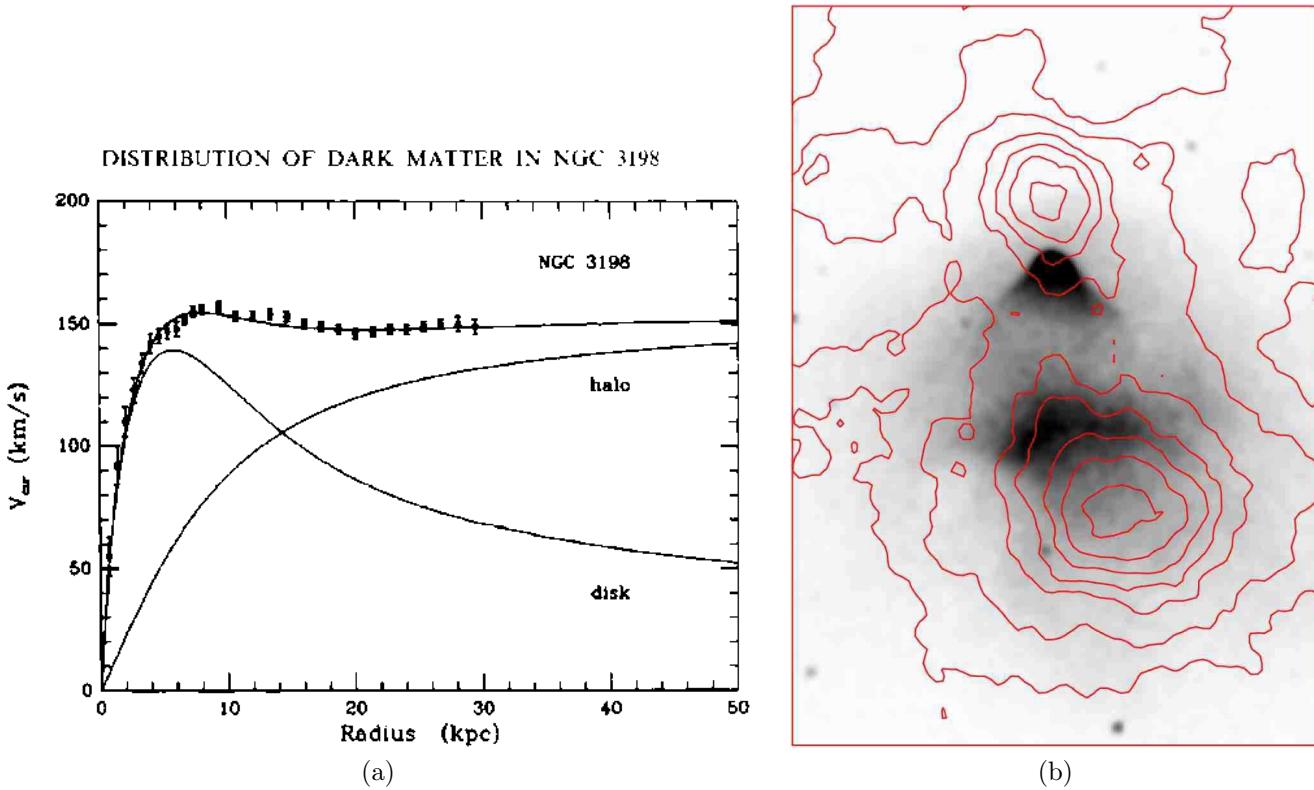


Figure 3.5: Subfigure (a) [2] shows the orbital velocity of objects in galaxy NGC 3198 vs their distance from the galactic center. Subfigure (b) [3] shows an X-ray photograph taken by the Chandra X-ray observatory of the aftermath of the collision of two galaxy clusters, 1E0657-558, overlaid on a contour map of the gravitation lensing magnitude factor κ derived from a separate image taken by the Magellan 6.5 m telescope.

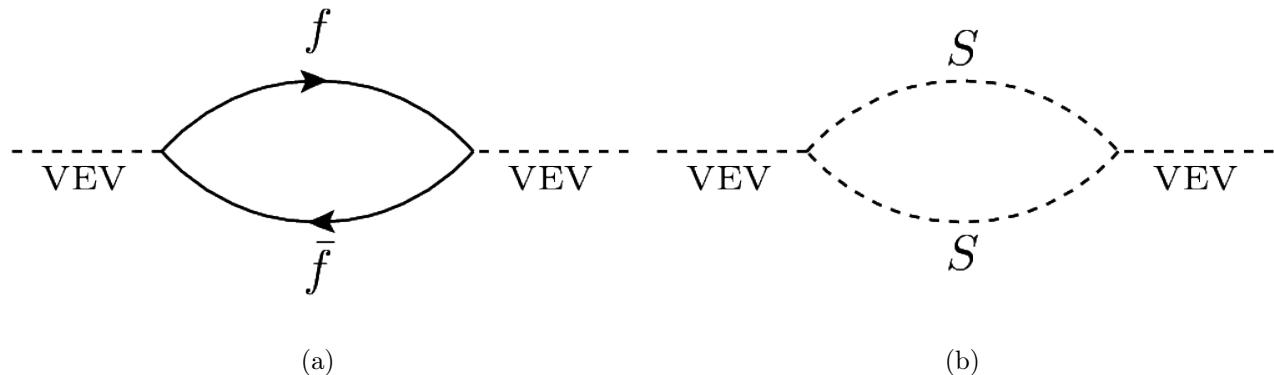


Figure 3.6: Diagrams for the one-loop corrections to the VEV from a generic fermion (a) and a generic scalar (b).

CHAPTER 4

SUPERSYMMETRY

Supersymmetry was originally developed out of an attempt to understand why all particle interactions seem to obey Poincaré symmetries, but not all interactions allowed by the Poincaré group are manifest in nature [31, 32]. The idea that a larger symmetry may exist, one that governs which interactions are allowed and which are not, was a prevailing source of curiosity at the time, and remains so today.

Because the ten generators of the Poincaré group have been mapped to the non-spin related spacetime symmetries (four translations, three rotations, and three boosts), it is natural to suppose that any larger symmetry containing the Poincaré group should involve the only remaining degrees of freedom, namely spin. Such a symmetry, which is called a supersymmetry, would generate transformations that shift quantum states by a fundamental unit of spin, $\hbar/2$, modeled as the action of an operator \hat{Q} such that

$$\hat{Q} |\text{boson}\rangle = |\text{fermion}\rangle, \quad \hat{Q}^\dagger |\text{fermion}\rangle = |\text{boson}\rangle. \quad (4.1)$$

Since the $|\text{boson}\rangle$ and $|\text{fermion}\rangle$ states differ by a spin of $\hbar/2$, conservation of angular momentum requires that the operator \hat{Q} itself carry spin $\hbar/2$, and so \hat{Q} is identified as a fermionic operator obeying the fermion anti-commutation relations

$$\left\{ \hat{Q}_\alpha, \hat{Q}_\beta^\dagger \right\} = -2\sigma_{\alpha\beta}^\mu \hat{P}_\mu, \quad (4.2)$$

$$\left\{ \hat{Q}_\alpha, \hat{Q}_\beta \right\} = 0, \quad \left\{ \hat{Q}_\alpha^\dagger, \hat{Q}_\beta^\dagger \right\} = 0. \quad (4.3)$$

The pair of states in (4.1), $|\text{boson}\rangle$ and $|\text{fermion}\rangle$, form a single object called a supermultiplet, where the fermion state is considered to be a field and the boson state is the corresponding “superfield”, or vice versa. As with all QFTs, the fields are quantized, and each field

¹Here, \hat{P}_μ is the generator of momentum translations, σ^μ is a four-vector of SU(2) rotation generators (Pauli matrices), and α and β denote spinor indexes.

begets a particle that carries the quantum numbers of the field. Thus, in a supersymmetric framework, there are always two particles, dubbed “superpartners”, that are identical copies of each other—except for their spin, of course, which differs by a fundamental increment. In particular, since \hat{Q} commutes with the gauge generators of the theory as well as the momentum generator,

$$[\hat{Q}_\alpha, \hat{T}^\mu] = 0, \quad [\hat{Q}_\alpha, \hat{P}^\mu] = 0, \quad (4.4)$$

the particle and superpartner possess the same charges under all gauge groups and also have the same mass.

In a Lagrangian obeying supersymmetry, the terms involving the superfields only carry model parameters that appear in the non-supersymmetric portion of the Lagrangian. This means that all couplings, mixing angles, and masses associated with the superfields are either already set by the gauge symmetries, or can be measured once and for all through experiments that probe the non-supersymmetric sector of the theory. This is true, however, only if supersymmetry is an unbroken symmetry. As we’ll learn shortly, if the laws of physics do exhibit any supersymmetry at all, it must be a broken supersymmetry, at least at low energies close the vacuum state.

4.1 Supersymmetry in our universe

It was not until some time after its conception that supersymmetry began to be viewed as an idea to be taken seriously. The idea became popular when it was learned that supersymmetry could potentially solve the hierarchy problem. The unnaturally small Standard Model VEV receives large corrections from each boson and fermion that interacts with the vacuum (Section 3.6.2). But there is some relief, or possibly a clue, in the fact that the bosons and fermions contribute with the opposite sign (compare 3.17 and 3.18). If each Standard Model fermion were to be paired with an equivalent boson and vice versa, the quadratic VEV corrections would all cancel, and the smallness of the VEV would be a consequence of symmetry.

Indeed, a supersymmetric framework would yield a set of particles that automatically induces this cancelation. Where each Standard Model field generates a huge correction to

the VEV, the corresponding superfield would generate the opposite huge correction, resulting in no net change to the electroweak breaking scale (see Fig. 4.1). Supersymmetry could be the mechanism to explain nature's unusual mass pattern.

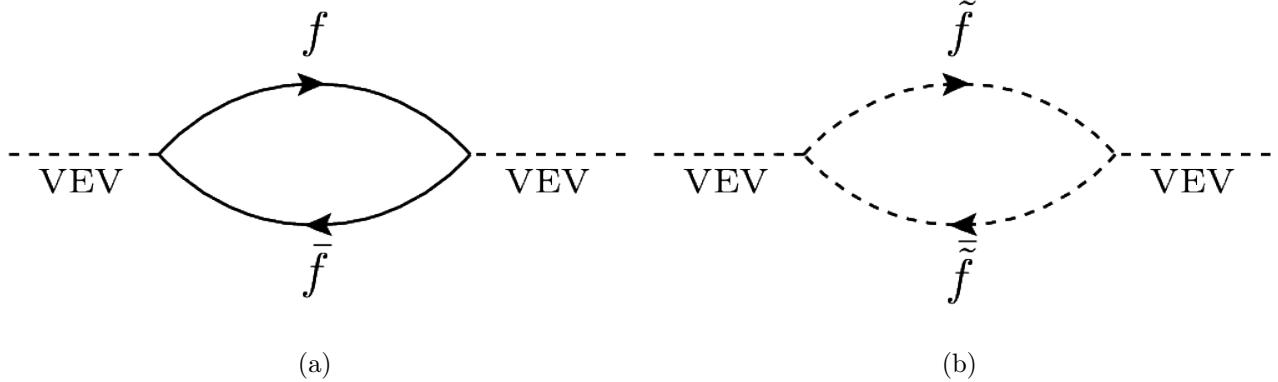


Figure 4.1: The cancelation.

But before supersymmetry can be seriously considered as a potential candidate, we must deal with the fact that the viability of a supersymmetric hypothesis in our universe faces two serious obstacles. They are:

1. A supersymmetric Standard Model could result in the proton being unstable.
2. Supersymmetry has not been observed in any experiment before.

Perhaps it is best to deal with these one by one.

1. If supersymmetric interactions existed, protons would quickly decay because supersymmetry introduces interactions that break the accidental Standard Model symmetries that conserve baryon and lepton numbers. Two quarks inside the proton could annihilate to form a lepton superpartner (slepton), and the proton would fly to pieces after a period of a few seconds to a year. However, experiments [33] have determined that the proton lifetime is greater than 10^{32} years, so clearly there is a spectacular discrepancy to be resolved. The situation can most easily be remedied by introducing an additional \mathbb{Z}_2 symmetry called *R*-parity [34], which, as shown in Section 2.6, requires that superpartners always be produced or annihilate in pairs. This forbids the vertices involved in the proton's decay, thereby preserving proton

stability. It should be noted that R -parity is not simply an ad hoc symmetry, but arises naturally out of many grand unification theories like $\text{SO}(10)$, as well as out of continuous gauge symmetries that conserve $B-L$.

2. If the Standard Model exhibited an unbroken supersymmetry, the masses of the superpartners would be identical to the masses of Standard Model particles, and the superpartner of the electron would have been discovered perhaps a century ago. Supersymmetry has not been observed in any experiment so far; therefore, in order to be a viable hypothesis, it must be that the superpartners are much heavier than their Standard Model counterparts, somehow. This can be arranged, so long as supersymmetry is a so-called “softly” broken symmetry, meaning that a certain set of terms can be added to the Lagrangian that violate supersymmetry by giving extra mass to the superpartners.

The fact that supersymmetry must be broken should not be taken as a strike against the hypothesis; as discussed in Section 3, the Standard Model is already host to a very important symmetry that happens to be broken, namely, the electroweak $U(1) \times SU(2)$ symmetry.

As stated, even a broken supersymmetry will cure the Standard Model’s hierarchy problem, as long as the supersymmetry is only broken by a certain class of terms. Generically, these terms are of the form

$$\mathcal{L}_{SOFT} = -\frac{1}{2}M_a\tilde{\lambda}_a\tilde{\lambda}_a - \frac{1}{6}b_{ijk}\tilde{f}_i\tilde{f}_j\phi_k - \frac{1}{2}b_{ij}\tilde{f}_i\tilde{f}_j - m_{ij}^2\tilde{f}_j^*\tilde{f}_i. \quad (4.5)$$

The $\tilde{\lambda}$ ’s are the superpartner fields of gauge bosons (gauginos), the \tilde{f} ’s are the superpartner fields of fermions (sfermions), and the ϕ ’s are scalar fields; repeated indices within a term imply a summation over species. The red and blue terms are responsible for endowing the superpartners with mass, and are needed to explain why supersymmetry so far has never been observed, solving problem 2. The green terms indicate the scalar trilinear interactions, and the orange are called the non-holomorphic terms [15], which don’t play a large role in this investigation but are included for completeness. I’ll now explain how supersymmetry, along with this generic set of soft breaking terms, could be realized in the context of the Standard Model.

Table 4.1: The supermultiplets of the MSSM.

Supermultiplets	spin 0	spin 1/2	spin 1	number of fields/2
Higgs, Higgsino	H_u	\tilde{H}_u		2
	H_d	\tilde{H}_d		2
quark, squark	\tilde{Q}_L	Q_L		6
	\tilde{u}_R^*	u_R^\dagger		3
	\tilde{d}_R^*	d_R^\dagger		3
lepton, slepton	\tilde{L}_L	L_L		6
	\tilde{e}_R^*	e_R^\dagger		3
B boson, bino		\tilde{B}	B	1
W boson, wino		\tilde{W}	W	3
gluon, gluino		\tilde{g}	g	8

4.2 The MSSM

Introducing supersymmetry into the Standard Model yields the set of supermultiplets listed in Table 4.1. You'll notice that there are two Higgs supermultiplets. The reason, in short, is that supersymmetry requires there to be a second Higgs doublet to prevent an electroweak gauge anomaly [28]. Represented here are the fields that make up the minimal supersymmetric Standard Model (MSSM), which predicts the existence of 41 extra fields in addition the 33 fields of the Standard Model. That is 33 superfields matching to the known Standard Model fields, and eight new fields arising from the second Higgs doublet. The supersymmetric Lagrangian containing these fields is given by

$$W_{MSSM} = \bar{u}\mathbf{y}_d Q H_u - \bar{d}\mathbf{y}_d Q H_d - \bar{e}y_e L H_d + \mu H_u H_d. \quad (4.6)$$

Table 4.1 summarizes the fields of the MSSM before symmetry breaking. As discussed, an unbroken supersymmetric Standard Model doesn't exist. Therefore, to complete the description of the MSSM, the generic soft-breaking Lagrangian in Equation 4.5 must be expressed in terms of the specific superfields listed in Table 4.1. Suppressing the chirality

labels, one obtains

$$\begin{aligned}
\mathcal{L}_{\text{SOFT, MSSM}} = & -1/2(M_1 \tilde{B}\tilde{B} + M_2 \tilde{W}\tilde{W} + M_3 \tilde{g}\tilde{g}) \\
& -\tilde{\bar{u}}\hat{\mathbf{a}}_u \tilde{Q}H_u - \tilde{\bar{d}}\hat{\mathbf{a}}_d \tilde{Q}H_d - \tilde{\bar{e}}\hat{\mathbf{a}}_e \tilde{Q}H_d \\
& -\tilde{Q}^\dagger \hat{\mathbf{m}}_Q^2 \tilde{Q} - \tilde{L}^\dagger \hat{\mathbf{m}}_L^2 \tilde{L} - \tilde{\bar{u}}\hat{\mathbf{m}}_u^2 \tilde{\bar{u}}^\dagger - \tilde{\bar{d}}\hat{\mathbf{m}}_d^2 \tilde{\bar{d}}^\dagger \\
& -\tilde{\bar{e}}\hat{\mathbf{m}}_e^2 \tilde{\bar{e}}^\dagger - bH_u H_d \\
& -m_{H_u}^2 H_u^* H_u - m_{H_d}^2 H_d^* H_d,
\end{aligned} \tag{4.7}$$

where the bold hatted symbols are complex 3×3 matrices in family space, and colors show the correspondence between similar terms in Equations 4.5 and 4.7. In total, Equation 4.7 contains 107 unknown parameters [35], which along with the four parameters from the second Higgs doublet, span the 111 unknown free parameters of the MSSM. That is 111 new parameters that must in principle be measured in future experiments, if the MSSM is indeed a model that describes nature.

Unfortunately, the large number of free parameters renders the MSSM parameter space impossible to explore exhaustively, at least for now. But, the vast set of experimental observations collected at particle physics experiments over decades, as well as theoretical motivations, can be used to constrain the model space, and make studying the MSSM a possibility. A number of constrained submodels of the MSSM have been developed and studied, two of which are discussed now.

4.3 The cMSSM

The constrained MSSM, or cMSSM [36], is a well-named submodel of the MSSM that makes many assumptions about the MSSM. This reduces enormously the dimensionality of the model. Sufficient assumptions are made to reduce the number of supersymmetry breaking parameters from 107 to just five. Central to these defining assumptions is the idea that at some energy near the GUT scale (10^{16} GeV), the superpartners become mass degenerate, a feature motivated by certain grand unification scenarios. This allows for all the sfermion masses to be set to a single value at the selected high energy, and the same for the gaugino

and Higgsino masses, that is,

$$\textcolor{red}{M}_3 = \textcolor{red}{M}_2 = \textcolor{red}{M}_1 \equiv m_{1/2}, \quad (4.8)$$

$$\hat{\mathbf{m}}_Q^2 = \hat{\mathbf{m}}_L^2 = \hat{\mathbf{m}}_u^2 = \hat{\mathbf{m}}_d^2 = \textcolor{brown}{m}_{H_u}^2 \mathbb{1} = \textcolor{brown}{m}_{H_d}^2 \mathbb{1} \equiv m_0^2 \mathbb{1}, \quad (4.9)$$

$$\hat{\mathbf{a}}_{u,d,e} \equiv A_0 \cdot \textcolor{cyan}{y}_{t,b,\tau} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (4.10)$$

$$\textcolor{blue}{b} \equiv B_0 \cdot \text{sign}(\mu). \quad (4.11)$$

The five black symbols on the right hand side, $m_{1/2}$, m_0 , A_0 , B_0 , and $\text{sign}(\mu)$ are the free parameters of the cMSSM, and the cyan $y_{t,b,\tau}$ are the known Standard Model values of weak hypercharge for the top quark, bottom quark, and tau lepton. Because most of these are matrix equations, there are many more constraints than equations; 102 constraints to be specific. As mentioned, the cMSSM has very low dimensionality, and as a result, the model can be readily used to interpret experimental data. However, most of the constraints responsible for the reduction in dimensionality are not particularly well motivated, as they amount to assumptions about the laws of physics at scales that have never been observed by humans. Furthermore, any conclusions made about the cMSSM based on data collected in particle experiments are highly contingent, since only a tiny fraction of the phenomenological possibilities within the MSSM are realized in the cMSSM.

This is not to say that models like the cMSSM have not played an important role in the evolution of our understanding of supersymmetry. They have provided benchmarks against which to compare the past and present sensitivity of experiments that probe supersymmetric hypotheses. They may play an increasingly important role in the future, if a significant part of the superpartner spectrum is discovered, and hints of the true supersymmetry breaking mechanism begins to appear. Such a breaking scheme could be quite elegant, described by only a few independent parameters. That said, very constrained models that make many untested assumptions are less useful when, as yet, there is absolutely no hint of supersymmetry.

4.4 The pMSSM

The phenomenological MSSM (pMSSM) is another submodel of the MSSM, one based on far fewer assumptions than the cMSSM. In particular, the pMSSM makes no assumption about how the supersymmetry breaking parameters are related at the GUT scale. The pMSSM [37] was developed with the aim of making as few assumptions about the MSSM as possible, while reducing the size of the model's parameter space to a level that is practicable and workable. The assumptions defining the pMSSM are:

- there are no new flavor changing neutral currents;
- there are no new sources of CP violation at tree level;
- the first and second generation fermions masses are degenerate, and
- the lightest of the superpartners is the neutralino.

The first two assumptions are motivated by the observation that CP-violating processes are extremely rare, even at loop level (higher order than tree-level), and flavor changing neutral currents are not observed at all. The third assumption is borne of the expectation that a model with two independent fermion generations will be as phenomenologically complete as a model with three generations, particularly in the early phases of discovery when only part of the full superpartner spectrum has been detected. The final assumption ensures that the lightest superpartner is electrically neutral, which is strongly motivated by cosmological observations. This leaves the following 19 parameters free:

- three gaugino mass parameters M_1 , M_2 , and M_3 ;
- one higgsino mass parameter μ and one pseudo-scalar Higgs mass m_A ;
- ten sfermion mass parameters $m_{\tilde{f}}$, where $\tilde{f} = \tilde{Q}_1, \tilde{U}_1, \tilde{D}_1, \tilde{L}_1, \tilde{Q}_3, \tilde{U}_3, \tilde{D}_3, \tilde{L}_1$, and \tilde{E}_3 (imposing $m_{\tilde{Q}1} = m_{\tilde{Q}2} = m_{\tilde{L}1} = m_{\tilde{L}2}$);
- three trilinear couplings A_t , A_b , and A_τ , and
- the ratio of the two Higgs vacuum expectation values $\tan(\beta) = \text{VEV}_A/\text{VEV}_h$.

With the particle masses free to vary independently throughout the model space, the pMSSM captures most of the possible, experimentally relevant phenomenology of the MSSM. Despite the still rather large dimensionality of 19 free parameters, the model has been successfully used for the interpretation of experimental results, as demonstrated in “The pMSSM interpretation of CMS 7 and 8 TeV results” [29], as well as in [38, 39]. A detailed account is given in Chapter 7.

4.5 Dark matter candidate

If supersymmetry is realized in Nature in the context of the Standard Model, and if R -parity is conserved, there exists a new heavy stable particle, called the LSP (lightest supersymmetric particle). If this particle is electrically neutral and colorless, it matches the criteria necessary to be dark matter. There are several such candidates, including the superpartners of Higgs boson, the Z boson, and the photon. This possibility serves as an additional motivation for searching for supersymmetry in particle colliders.

CHAPTER 5

THE LARGE HADRON COLLIDER

The world's foremost tool for observing physics at the electroweak scale is a particle accelerator and collider facility called the Large Hadron Collider (LHC) [40]. Located at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland, the LHC is the largest and perhaps most complex machine ever constructed, and the source of the world's most energetic laboratory-prepared particles. This machine offers our best chance of observing evidence of new structure that may describe our universe at the smallest scales, including perhaps the particles and fields of supersymmetry.

5.1 LHC key features

The primary feature of the LHC is a meter-wide, 26.7 kilometer-long cryogenic vacuum cylinder forming a nearly circular path below ground in France and Switzerland. The LHC ring occupies a 4 meter-wide tunnel whose depth ranges between 45 and 170 meters beneath the surface. During normal operation, the LHC ring houses two counter-circulating beams of energetic protons, which are steered by superconducting magnets and brought to intersect at various points around the ring. There are four such crossing points, located at the centers of the four LHC experiments, ATLAS, ALICE, CMS, and LHC-b, where high energy protons collide. The layout of the LHC is shown in Figure 5.1.

The LHC ring does not form a perfect circle, but rather is made of alternating curved and straight sections. The curved sections are 2500 meters long in arc length and are instrumented with superconducting dipole magnets. The straight sections are 530 meters long, and their centers (labeled as points 1-8 in Figure 5.1) have either been instrumented with hardware that serves necessary functions for the operation of the LHC, or with of the experimental facilities.

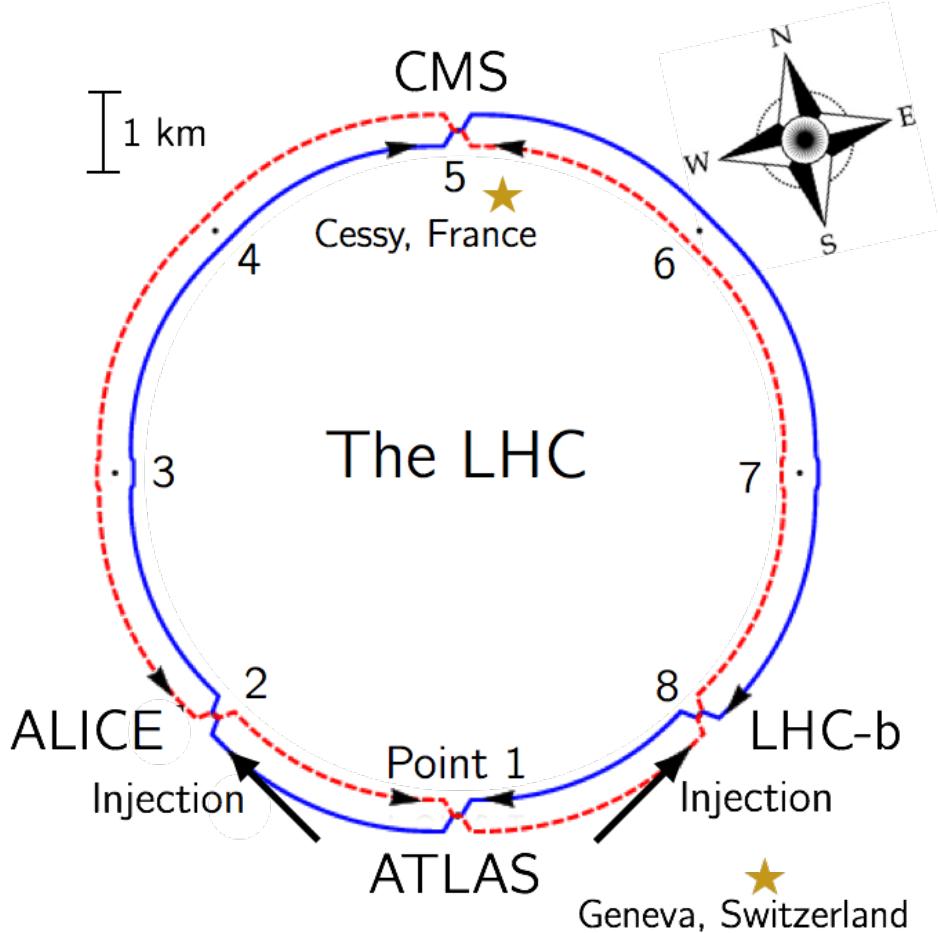


Figure 5.1: The layout of the Large Hadron Collider. The scale is approximate and the separation between the beams has been exaggerated to show the crossings of beam 1 (blue) and beam 2 (red) at interaction points 1, 2, 5, and 8.

The hardware and functionality of each of the LHC’s eight points is summarized in Table 5.1. The primary experimental facilities, mentioned above, observe high energy proton and heavy ion collisions and perform both standard model measurements and searches for BSM physics. In addition to the four primary experiments, Table 5.1 lists in parentheses a few additional experimental facilities that share interaction points with the primary four. TOTEM shares the experimental cavern with CMS on the LHC ring at point 5 and measures, among other observables, proton-proton elastic scattering cross sections; MOEDAL shares the interaction point 8 with LHC-b and performs searches for particles with magnetic monopole moments, and LHC-f is a future experiment that will share interaction point 1

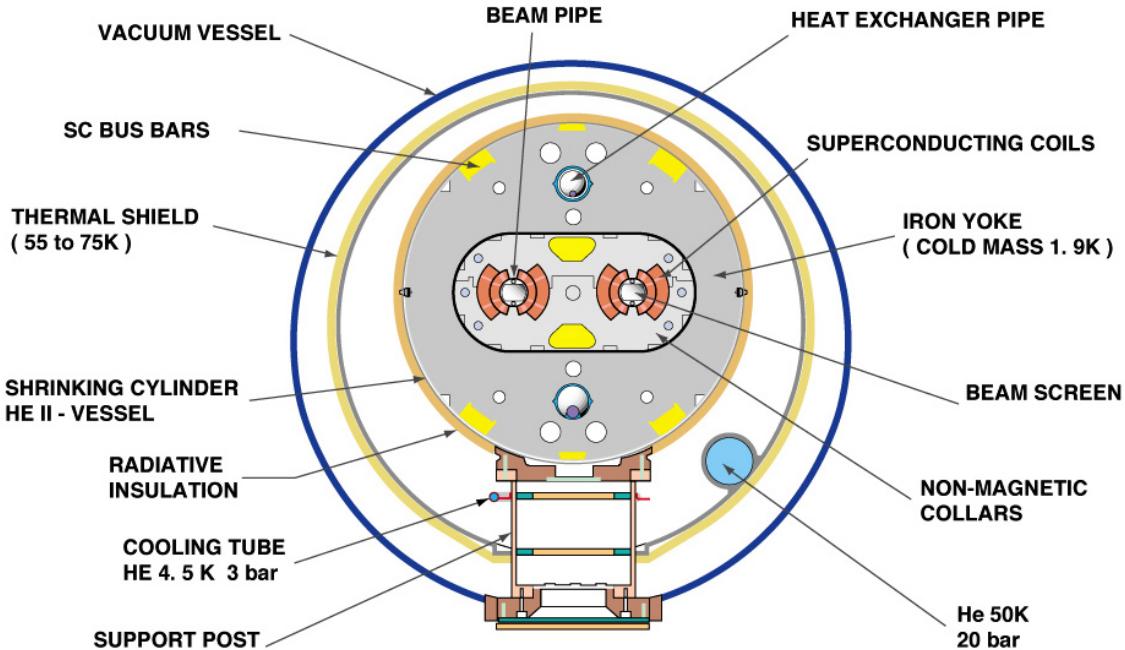


Figure 5.2: Cross section of a cryodipole, the most common segment of the LHC [4].

with ATLAS and study particles that travel nearly parallel with the beam direction (the forward direction).

Figure 5.2 shows a typical cross section of the LHC ring. Within the outer steel casing are not only the two proton beams, but also the magnets and other instrumentation required to control and manipulate the beams. A continuous ultra high vacuum is maintained in the outer volume in order to thermally insulate the inner volume. The inner volume is filled with an iron yoke and held at a temperature of 1.9 K by supercooled liquid helium, which flows through the heat exchanger pipe. Inside the yoke and in thermal equilibrium with the iron is a twin bore assembly of niobium titanium superconducting coils that produce magnetic fields with strengths of up to 9 Tesla (T). Within the focus of the coils are the two parallel evacuated beam pipes, highways for subatomic particles.

Table 5.1: The hardware at the interaction points of the LHC.

point	hardware	functionality
1	ATLAS experiment [41]	<ul style="list-style-type: none"> • full coverage particle detector • proton-proton collisions • standard model measurements and searches for evidence of new particles
	LHC-f experiment [42]	<ul style="list-style-type: none"> • forward particle detector
2	ALICE experiment [43]	<ul style="list-style-type: none"> • partial coverage particle detector • heavy ion collisions
3	RF cavities	<ul style="list-style-type: none"> • momentum cleaning
4	sextupole magnets	<ul style="list-style-type: none"> • acceleration
5	CMS experiment [26]	<ul style="list-style-type: none"> • full coverage particle detector • proton-proton collisions • standard model measurements and searches for new particles
	TOTEM experiment [44]	<ul style="list-style-type: none"> • forward particle detector
6	quadrupole magnets	<ul style="list-style-type: none"> • beam focusing
7	sextupole magnets	<ul style="list-style-type: none"> • betatron cleaning
8	LHC-b experiment [45]	<ul style="list-style-type: none"> • partial coverage particle detector • heavy ion collisions
	MOEDAL experiment [46]	<ul style="list-style-type: none"> • forward physics

The LHC’s superconducting coils produce various magnetic field geometries to serve different functions. About 1200 cryodipole magnets are positioned in the curved LHC segments and produce dipole fields that bend the protons along their arched trajectories; about 400 magnets located in the LHC’s short straight sections produce quadrupole fields that are used to focus the beams, and thousands of additional magnets produce sextupole, octopole and decapole fields that purify and adjust the beams in various ways.

Additional systems are responsible for the refrigeration of liquid helium, the powering and protection of the superconducting coils, and the maintenance and dumping of the particle beams.

5.2 The injection sequence

Long before they reach the LHC beam pipes, protons are accelerated and focused by a multi-staged process. Initially the protons are stored as the nuclei of gaseous hydrogen atoms, kept in a sealed bottle at the CERN accelerator complex. When the LHC operators begin the procedure of filling the machine with protons, the hydrogen in the bottle is injected into a duoplasmatron [40], which subjects the hydrogen to an electric field and a bombardment of electrons. The electrons ionize the hydrogen atoms, and the electric field carries the liberated hydrogen nuclei (protons) through the duoplasmatron cavity, providing them with 100 keV of kinetic energy. Upon leaving the duoplasmatron, the protons are gathered by a quadrupole magnet and guided into the aperture of a linear accelerator called LINAC2, where they are accelerated by radio frequency (RF) cavities up to an energy of 50 MeV. LINAC2 discharges the protons into a cyclotron called the Proton Synchrotron Booster (PSB), which accelerates the protons repeatedly as they cycle around a 20-meter circular path. After reaching an energy of 1.4 GeV within the PSB, the protons are injected into a larger cyclotron called the Proton Synchrotron (PS). The PS elevates the protons’ kinetic energy to 25 GeV before passing them to the world’s second largest particle accelerator, the Super Proton Synchrotron (SPS), which takes the protons up to an energy of 450 GeV. The energy is now high enough for the protons to be injected into the LHC.

Protons are injected into the LHC at 450 GeV and accelerated up to the highest energy ever achieved at a laboratory, 6.5 TeV. The beams are collided head-on, yielding a collision energy of 13 TeV in the proton-proton center of mass frame. To understand why it is necessary to push the energy frontier in this way, it is necessary to explain the goals of the LHC.

5.3 The goals of the LHC

One of the goals of the LHC is to produce and observe new, heavy particles in the collisions of protons. This of course assumes that such particles actually exist. Assuming they exist, and for concreteness, that the new particles are the particles of supersymmetry, there are two main ingredients required to produce SUSY particles, both of which can be understood by examining the expression that governs how many proton-proton collisions are expected to result in the creation of SUSY particles:

$$N_{\text{SUSY}} = \mathcal{L} \cdot \sigma_{\text{SUSY}}. \quad (5.1)$$

Here, N_{SUSY} is the number of events in which SUSY particles are present, where “event” is defined as the collection of particles emanating from a single collision. The quantity σ_{SUSY} is the interaction cross section, a measure of the probability for a supersymmetric interactions to occur during a collision. The symbol \mathcal{L} stands for the integrated luminosity, and is a measure of the total number of protons collided:

$$\mathcal{L} = \frac{N_{\text{collisions}}}{\sigma_{\text{pp}}}. \quad (5.2)$$

Here, $N_{\text{collisions}}$ is the total number of collisions achieved at the LHC and σ_{pp} is the cross section for protons to interact. Clearly, since the goal is to maximize N_{SUSY} , it is important to maximize both σ_{SUSY} and \mathcal{L} . The maximization of σ_{SUSY} is achieved by imparting as much kinetic energy as possible to the protons. The maximization of \mathcal{L} is achieved by making the beam intensities, i.e., the number of protons per unit volume and time as possible. A few details on these two objectives are now given.

5.3.1 The objective of high energy

Owing to energy conservation, the probability of producing a supersymmetric particle gets lower as the hypothetical mass of the particle increases. However, this limitation can be overcome by increasing the energy of the colliding protons. In other words, σ_{SUSY} , and thus N_{SUSY} , increase as a function of the proton energy. This is why the primary objective of the LHC is to achieve high energy collisions.

5.3.2 The objective of high luminosity

While high energy is the primary goal of the LHC, the other essential goal is to provide as large a beam intensity as possible. This is achieved by maximizing $N_{\text{collisions}}$, by colliding as many protons as possible within a given period of time. To this end, the proton beams are not continuous beams, but are segmented into bunches, which are groups of a few billion protons, where the bunches are spaced apart in time intervals of 50 ns (25 ns) in Run 1 (Run 2). Maximizing the number of observed collisions is important because new physics processes are expected to be very rare, and as few as one in 10 billion collisions might contain the kind of new physics that is being looked for. The design luminosity of the LHC is $\mathcal{L} = 10^{34}$ cm $^{-2}$ s $^{-1}$.

CHAPTER 6

THE COMPACT MUON SOLENOID

Centered at collision point 5 of the LHC, and located 100 meters below the town of Cessy, France, is an underground cavern that houses a particle detector called the Compact Muon Solenoid (CMS) [40] [26]. Weighing 12,500 tonnes and occupying a volume of 3600 cubic meters, CMS (shown in Fig. 6.1) is the most massive particle detector to be constructed at a collider experiment. The experimental collaboration that built, operates and uses the data collected by the detector is called the CMS collaboration, and comprises over 3000 participants from 199 institutions in 43 countries.

The CMS detector was designed to study, among other things, particles produced through rare electroweak processes. At the LHC, such processes occur amid an overwhelming number of QCD dijet events (events with two hadronic showers that recoil back-to-back) in an environment of high instantaneous luminosity. To reconstruct and identify particles amid the large background, the various sub-detectors of CMS measure various properties of the particles emerging from the collisions, and this information is used to form particle candidates. The properties that can be measured include the electric charge, energy, momentum, direction of motion, particle identity (ID), and the decay lifetime of particles. The ability to use information from the various sub-detectors to identify and characterize individual particles is made possible by CMS's fine granularity, meaning the sub-detectors are composed of small, closely-spaced units, or cells. The fine granularity is key to accurately measuring the position of particles passing through the detector, and minimizing the chances that the same cell will be activated by two particles simultaneously, which can complicate the reconstruction. This complication is enhanced by the fact that the high instantaneous luminosity of the LHC leads to, in most cases, the occurrence of multiple simultaneous collisions per bunch crossing rather than just a single collision, an effect called pile-up. A precise determination of the momentum of particles is made possible by a high magnetic field that fills the space of

the inner sub-detectors, the tracker and electromagnetic calorimeter, described below. The global analysis of each event and its constituent particles is performed by an algorithm called particle flow [47], some details of which are provided in the sections below.

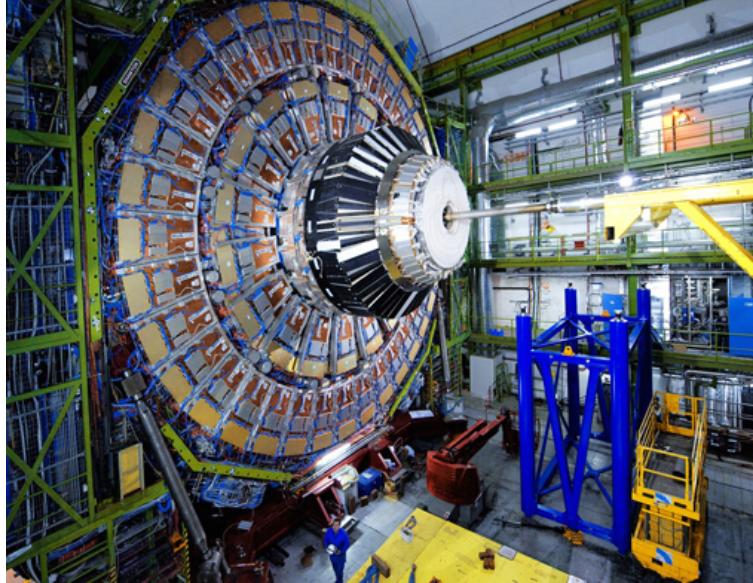


Figure 6.1: The CMS detector, pulled apart during the long shutdown of 2014.

A photograph of the CMS detector is given in Fig. 6.1. The sub-detectors are nested cylindrically around the LHC beams, centered on the interaction point. A description of the coordinate system used by CMS, as well as a set of key observables that CMS is capable of reconstructing, is given in the following section. Then, the main systems are described, starting from the inside (closest to the beam lines) and moving outward.

6.1 Coordinates and observables

The CMS detector makes use of both a right-handed cartesian coordinate system as well as an approximately Lorentz invariant spherical coordinate system (HEP coordinate system). In the cartesian system, \hat{x} points toward the center of the LHC, \hat{y} points up, and \hat{z} points in the direction of the counterclockwise rotation of the LHC beam pipe as viewed from above. The HEP coordinate system is spanned by the basis vectors p_T (transverse momentum),

η (pseudorapidity), and ϕ (azimuthal angle), which translate to the right-handed cartesian coordinate system through the relations:

$$p_T = \sqrt{(p_x)^2 + (p_y)^2} \quad (6.1)$$

$$\eta = -\ln[\tan(\theta/2)] \quad (6.2)$$

$$\phi = \tan^{-1}(p_y/p_x), \quad (6.3)$$

where θ is the polar angle from the positive \hat{z} axis.

The missing transverse momentum \vec{p}_T^{miss} is defined as the negative of the vector sum of the momenta of all visible particles in an event, projected on the plane transverse to the beam:

$$\vec{p}_T^{\text{miss}} \equiv - \sum_i^n (\vec{p}_T)_i. \quad (6.4)$$

The \vec{p}_T^{miss} has magnitude $|\vec{p}_T^{\text{miss}}|$, or just E_T^{miss} , and direction ϕ^{miss} , and is an estimate of the transverse momentum of the invisible particles. Since the true mass of an invisible particle is not observable, such masses are conventionally assumed to be zero, and the magnitude of the \vec{p}_T^{miss} vector is thus equated with the missing transverse energy E_T^{miss} . The \vec{p}_T^{miss} can be used to construct the four-vector

$$(p_T^{\text{miss}})^\mu \equiv (E_T^{\text{miss}}, \vec{p}_T^{\text{miss}})^\mu. \quad (6.5)$$

Another key observable, H_T , is a measure of the total hadronic energy in an event, and is defined as the scalar sum of the transverse momenta of jets in an event, or

$$H_T \equiv - \sum_i^n (p_T)_i. \quad (6.6)$$

6.2 Silicon tracker

The innermost system of CMS, immediately encountered by particles emerging from collisions, is the tracker [48]. Shown in Fig. 6.3, the tracker consists of two subsystems: a 600 cm-long inner pixel detector extending out to a radius of 11 cm, and a 5.8 m-long outer strips tracker extending out to a radius of 1.2 m. The pixel detector is made of 66

million silicon pixels, and the strips detector contains 10 million silicon strips. The density of readout channels is highest within the innermost detector elements because the intensity of particle flux is highest in this region. Tracks left by muons (hadrons) are reconstructed with a very high efficiency of greater than 99% (90%) with a fake rate as low as 1%, for transverse momenta as low as 150 MeV. At the design luminosity of the LHC of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$,

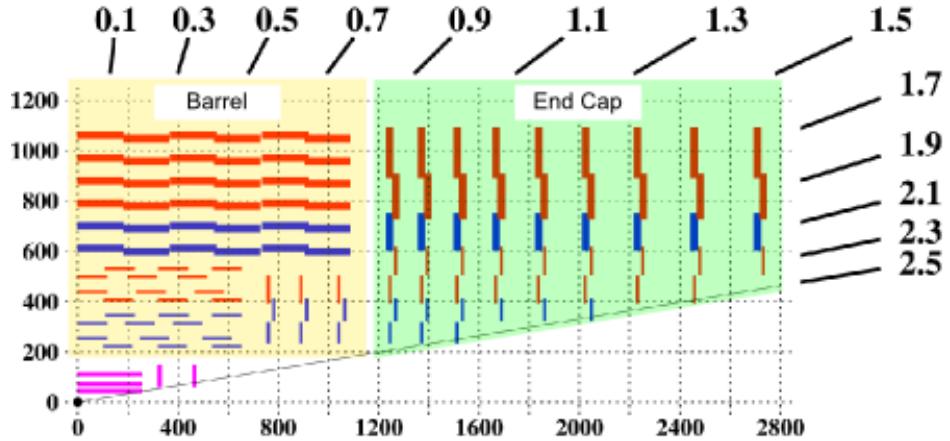


Figure 6.2: A schematic view of the CMS tracker. The direction of view is toward the center of the LHC.

approximately 1,000 particles are created each 25 ns bunch crossing. This means that during the reconstruction, the innermost pixel layer is responsible for disentangling the tracks of about one million particles per mm^2 each second. The pixel detector is also responsible for resolving each particle's impact parameter, which is the distance of closest approach of the particle's trajectory to the point of the proton-proton collision, which is called the primary vertex. This task will be particularly challenging when the LHC enters its high luminosity phase, since the occurrence of multiple collisions within the same bunch crossing, will become much more prevalent.

6.3 Electromagnetic calorimeter

Upon reaching the outer edge of the CMS tracker, particles enter the electromagnetic calorimeter (ECAL). The ECAL is made of 72,000 lead tungstate (PbWO_4) crystals, like



Figure 6.3: An ECAL endcap PbWO_4 crystal, length 220 mm ($24.7 X_0$).

the one shown in Fig. 6.3. The material was chosen for its high density and high optical transparency, which, respectively, guarantee a short radiation length ($X_0 = 0.89 \text{ cm}$) and efficient transmission of light. Highly-relativistic charged particles emit bremsstrahlung radiation as they traverse the crystals, resulting in electromagnetic showers, and the light is collected by photodetectors mounted on the backs of the ECAL crystals. The amount of radiation produced is proportional to the energy of traversing particles, which key to determining the energy of particles during the event reconstruction. PbWO_4 is also chosen for its small Molière radius of 2.2 cm, which results in an improved spatial resolution of particles over that possible with other materials.

In the barrel, crystals are organized into rectangular modules composed of 400 or 500 crystals each, and modules are arranged in groups of four to form supermodules. There are 36 supermodules in total, which combine to cover a large rectangular area in the space of η and ϕ , as shown in Fig. 6.4 (a). This large rectangle covers all values of ϕ and reaches an η of ± 1.48 . In the endcaps, crystals are arranged in groups whose number varies depending on the location. The modules combine to form sectors, of which there are a total of 18, shown in Fig. 6.4 (b) and (c). Sectors, as well as supermodules, are each read out by a front end digitizer (FED).

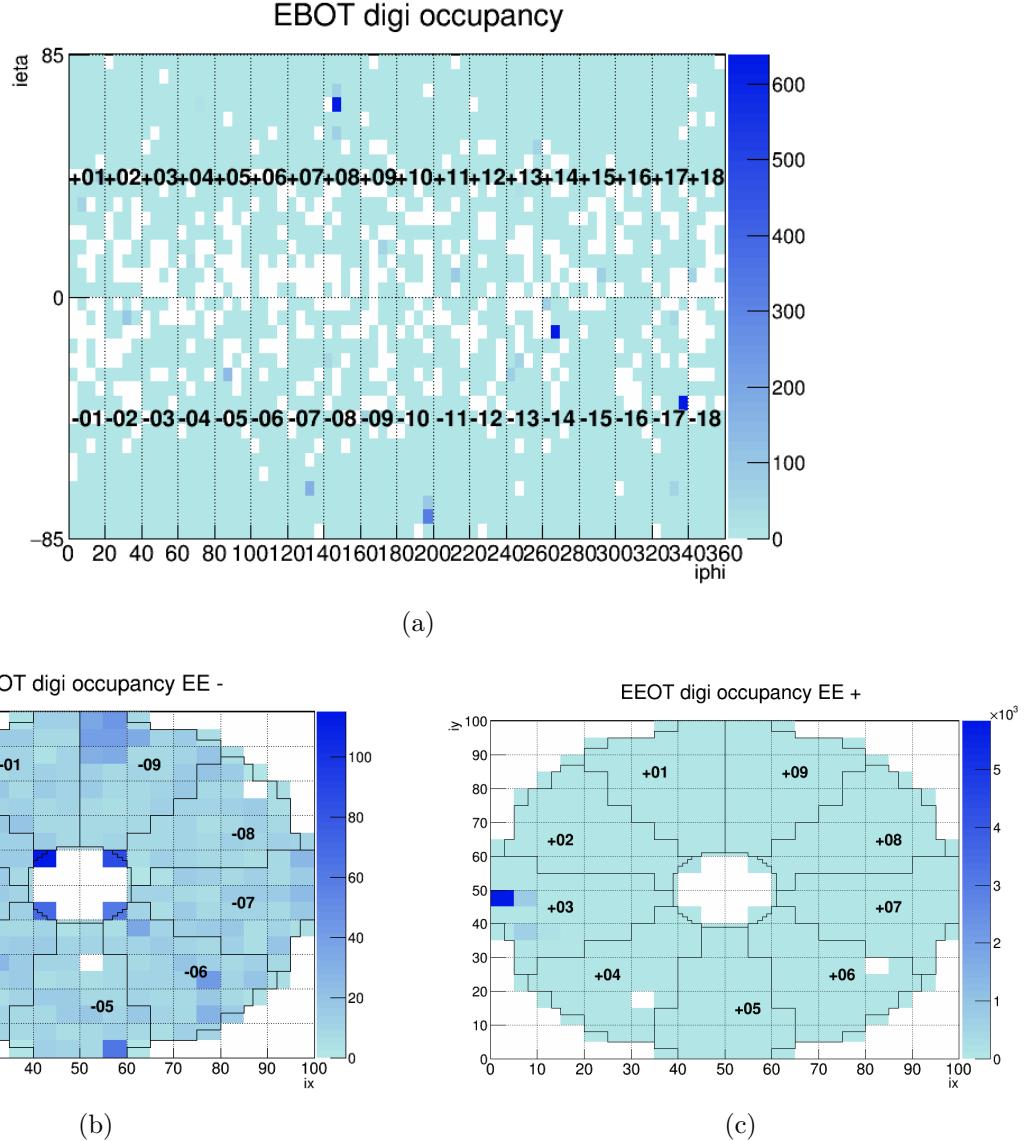


Figure 6.4: The layout of the ECAL barrel (a) in η - ϕ coordinates, and the forward (c) and rear (b) ECAL endcaps in cartesian coordinates. The shading shows the occupancy of a given run. OT stands for out-of-time, indicating that the occupancy shown in the color plot is of energy deposits that were detected out of time with the LHC bunch crossings.

Deposits in the calorimeters are analyzed in the reconstruction to yield calorimeter clusters. A calorimeter cluster is composed of cells corresponding to a relative maximum in the amount of energy deposited, and the cluster energy is equal to the energy deposited in the central cell summed with the energy deposited in neighboring cells, provided that a neigh-

boring cell has an energy deposit exceeding 1 GeV. The ECAL is capable of determining the energy of photons and electrons with a resolution of 1–5%, depending on the pseudorapidity, and charged hadrons with a resolution of about 10%.

6.3.1 ECAL calibration

Ensuring an equivalent and accurate response among the various crystals (channels) requires a detailed calibration procedure. The ECAL is calibrated *in situ*, or while the detector is operating, and a number of methods are employed. The first method exploits the principle of the ϕ -invariance of collisions. Since the primary proton beams are unpolarized, outgoing particles are distributed isotropically in the azimuthal angle ϕ , and this allows channels with different ϕ locations but within a small η region to be checked against one another. The second method, which is the most important, relies on the tracker to give an accurate estimate of the momentum of isolated electrons. The momentum of the tracks can be used to scale the response of the ECAL signal to the appropriate value. This method makes use of electrons from events in which a W boson is produced and decays to $e + \nu_e$, in which the outgoing electrons exhibit a broad p_T spectrum that peaks around 40 GeV. The third method employs tag and probe methods [47] in which neutral mesons decay to two γ particles.

The second method relies on the clean measurement of the momenta of electrons from the tracker, but as will be explained in Section 6.8, information from the tracker is not available to the CMS level 1 (L1) trigger (Section 6.8 gives a description of the L1 trigger). Therefore, the collection (triggering) of events to be used for calibration must rely on information from the ECAL itself, as well as the hadronic calorimeter (HCAL), the system described in the next section. It is crucial to ensure that enough events with low-energy electrons are triggered in order to calibrate the ECAL, while preventing the large background of charged pions from overwhelming the calibration sample. For this, a selection is applied to events containing e/γ candidates, which is a collection of four-vectors that contain nearly all electrons and charged pions. The goal is to discriminate between the pion background and the electrons to be used for calibration, and ultimately reject as many candidates believed to be pions

as possible, while reconstructing enough electrons to calibrate the ECAL. The fundamental triggering information used in the selection are the called trigger primitives, and are:

- et2x1: the energy of the e/γ candidate, equal to the summed energy deposited in the two adjacent ECAL trigger towers, which are 5×5 arrays of crystals, closest to the crystal with maximum energy deposit.
- et12x12: the energy of a jet candidate, equal to the sum of all ECAL and HCAL energy in the 12×12 array of trigger towers surrounding the e/γ candidate, multiplied by 4.

The simulated distributions of these primitives for signal and background e/γ candidates is shown in Fig. 6.5 (a), where a proximity matching has been applied between the e/γ candidates and the parton-level, often referred to as “truth”-level, particles. These primitives are the building blocks of observables designed to give good signal/background discrimination when used in a cut-based, that is, threshold-based, selection, which are:

- hadronic energy fraction (H/E): The ratio of the energy deposited in the of the 3×3 array of HCAL towers centered on the e/γ seed to that deposited in the corresponding 3×3 array of ECAL towers:

$$H/E = \sum_{3 \times 3} \text{HCAL} / \sum_{3 \times 3} \text{ECAL}; \quad (6.7)$$

- isolation: The sum of the energy deposits surrounding the ECAL seed divided by the energy of the seed:

$$\text{isolation} = (0.25 \times \text{et12x12} - \text{et2x1}) / \text{et2x1}. \quad (6.8)$$

Simulated distributions of these observables are shown for signal and background e/γ candidates in Fig. 6.5 (b).

The efficiency, or fraction of e/γ candidates matched to true electrons that are selected by the baseline selection of

- isolation < 2.0
- $H/E < 0.5$,

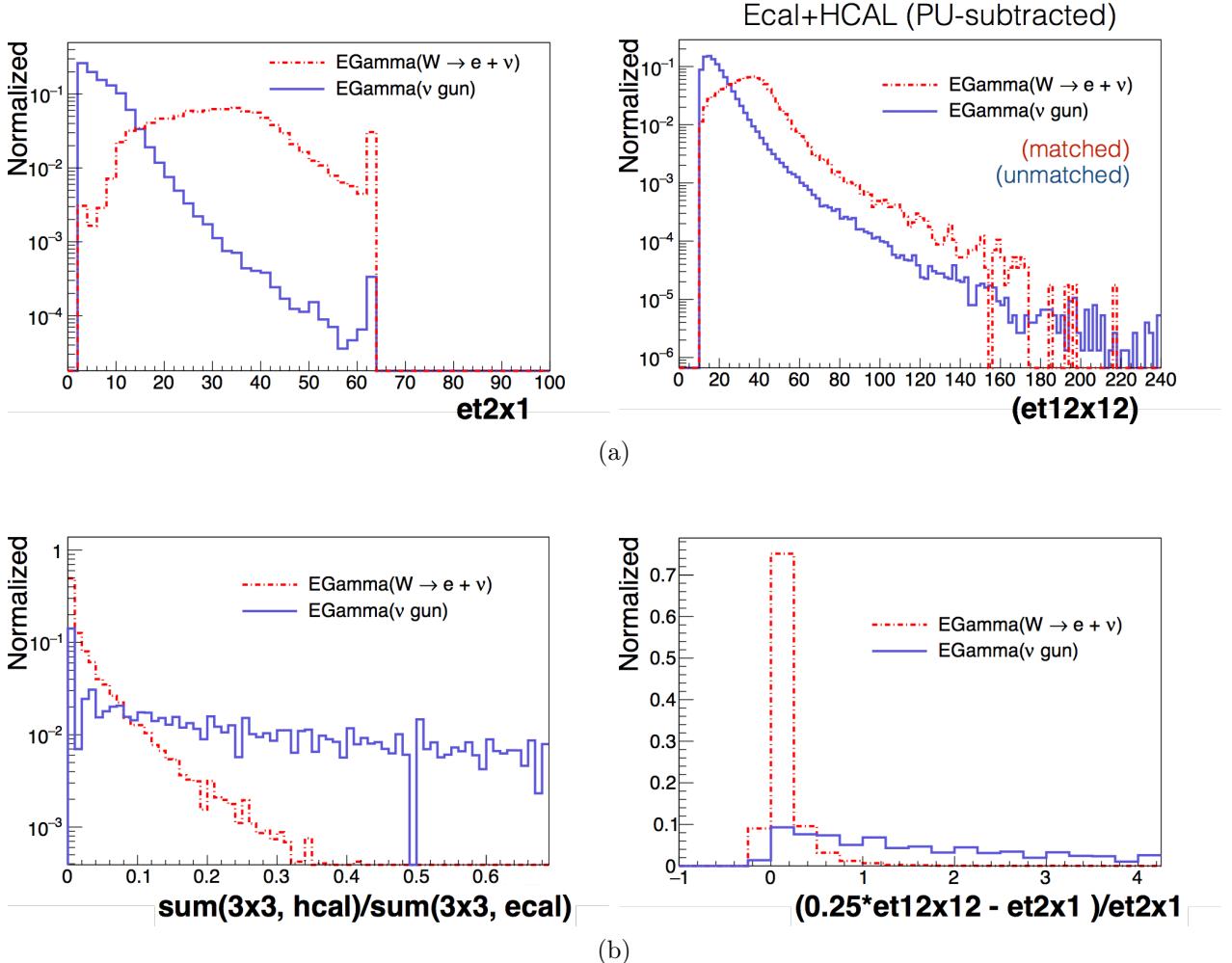


Figure 6.5: The trigger primitives available from the calorimeters available to the CMS L1 trigger (a), and the observables derived from the primitives designed to discriminate between signal and background e/γ candidates.

is shown in Fig. 6.6 for various thresholds, shown in the figure.

The performance of the currently applied cut-based selection was investigated by comparing the existing performance with that obtained through the use of an idealized discriminator. The idealized discriminator chosen was a boosted decision tree (BDT) [49] whose input variables are the observables H/E and isolation. The BDT was trained using half of the available simulated signal candidates and half of the background candidates, while half of both collections was set aside as an independent test sample. Figure 6.7 characterizes the performance

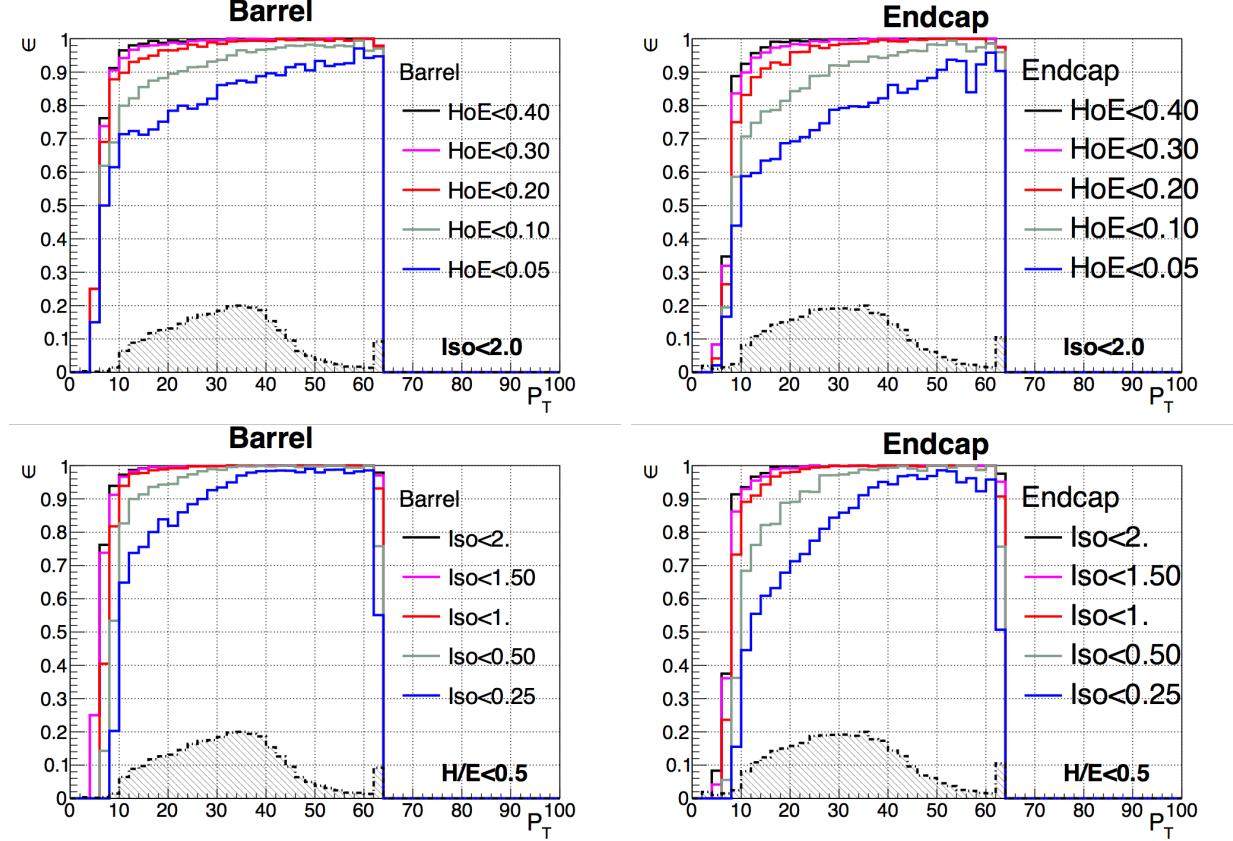


Figure 6.6: The signal efficiency in the barrel and endcap as a function of the p_T of the e/γ candidates for various thresholds on the H/E and isolation.

of the cut-based and BDT selection, showing the trade-off between the signal efficiency and background contamination in the test sample. The performance of the idealized discriminant is comparable to the performance of the selected working points of the cut-based approach, indicating that the values of the currently-employed selection is well-optimized.

The addition of new observables in the selection was also investigated. For example, the centrality, explained in called the centrality, explained in Fig. 6.8 The change in performance of the BDT with the centrality is shown in Fig. 6.9. Including the centrality improves the performance by about 10-15%, which implies a 10-15% reduction in the background rate for a given signal efficiency. This result has been presented to the ECAL DPG working group and is available if a 10-15% reduction in the trigger rate is required for calibrating in subsequent runs. A cut-based selection on this observable would need to be applied, since

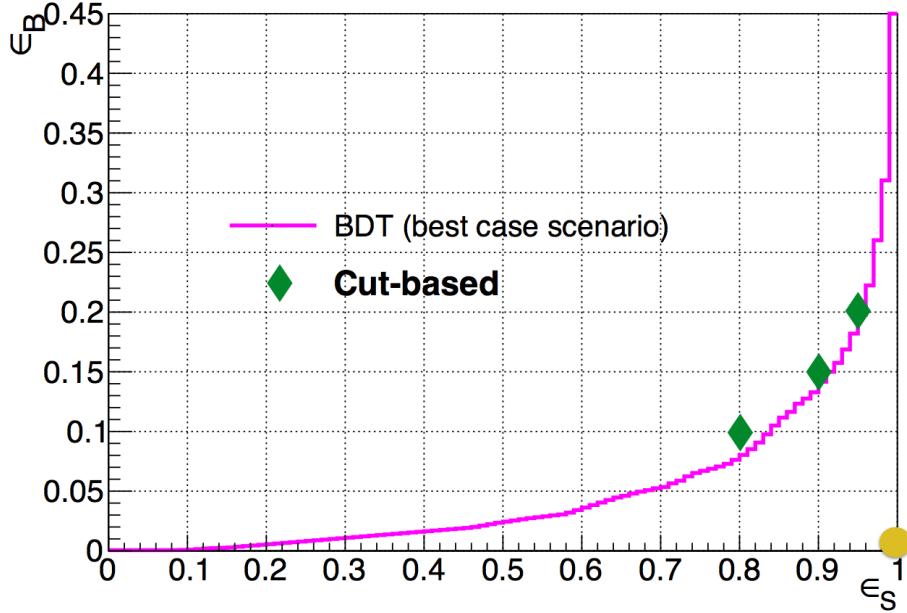


Figure 6.7: The background efficiency vs the signal efficiency for cut-based selection and the selection of the BDT based whose inputs are the cut-based observables. The ideal point is indicated by the gold circle located at full signal efficiency and zero background efficiency. For a given background efficiency, the cut-based signal efficiency is within a few percent of the BDT efficiency.

the use of multivariate discriminants is not possible, at present, in the L1 trigger.

6.4 Hadronic calorimeter

Immediately surrounding the CMS ECAL and inside the CMS solenoid magnet is a sampling hadronic calorimeter (HCAL), shown in Fig. 6.10. Constructed from plates that alternate between steel, brass, and plastic scintillating material, the HCAL comprises a barrel with 36 azimuthal wedges covering the range $|\eta| < 1.3$, and two endcaps covering the range $1.3 < |\eta| < 3$. The number of interaction lengths (λ_0) in the barrel ranges between 5 to 10 λ_0 , and about 10 λ_0 in the endcaps. The geometry of the 36 barrel wedges closely matches the geometry of the 36 ECAL supermodules. An individual HCAL cell subtends the solid angle of a 5×5 array of ECAL crystals. The geometry of the endcaps differs from that of the ECAL.

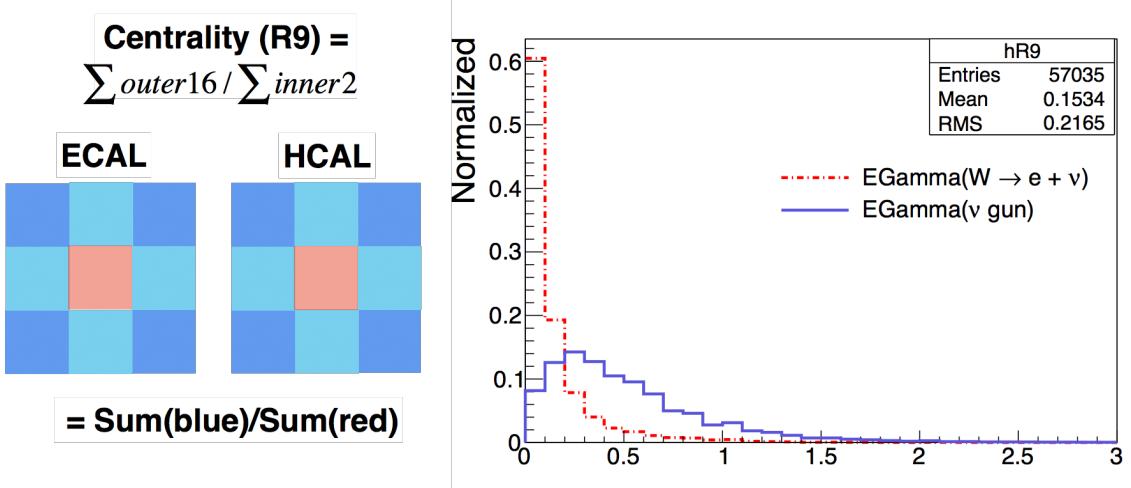


Figure 6.8: The definition of the centrality (left), along with the distributions of the centrality for signal and background e/γ candidates in simulation (right).

6.5 Superconducting solenoid

Dividing the inner detector from the outer is a 12-m long, 6-m-diameter superconducting solenoid magnet, the system most critical for the functionality of the CMS detector. Niobium Titanium coils, kept at a temperature of 3.2 K and carrying a stored energy of 2.6GJ, are responsible for filling the inner detector with a uniform 3.8 T magnetic field. The superconducting coils are cooled by liquid helium that is supplied and cooled via a multi-stage cryogenic system that is managed by the External Cryogenics Group of the LHC. Oil contamination in this system resulted in sub-optimal performance of the CMS solenoid in 2015, and led to a 20—30% loss in the data collected.

6.6 Muon system

The outermost subdetector of CMS is the muon system. The rest of the CMS detector, which is nested within the muon systems, amounts to 16 interaction lengths of material, and so the only particles that are able to interact with the muon system are muons. Like the other subdetectors, the muon system coheres to the azimuthal symmetry of the CMS solenoid. Composed of a barrel and two endcaps, the muon system provides complete pseu-

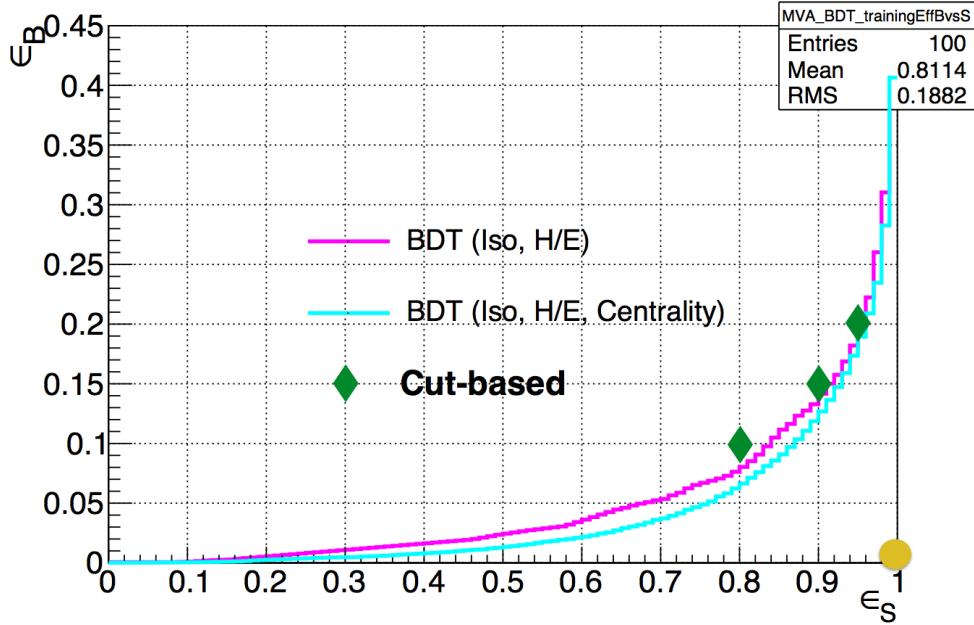


Figure 6.9: The background efficiency vs the signal efficiency for cut-based selection, the plane BDT selection, and the selection of the BDT trained with the addition of the centrality variable. The ideal point is indicated by the gold circle located at full signal efficiency and zero background efficiency. For a given signal efficiency, a suppression of the background on the order of 10-15% is possible.

dorapidity coverage in the range $|\eta| < 2.4$. The muon detector, shown in Fig. 6.11, is made of 4 concentric, cylindrical stations, and each station is composed of 12 ionizing drift chambers [40]. On its own, the muon system is capable of determining the momentum of muons with a p_T of up to 1 TeV with a resolution of 15% in the barrel and 40% in the endcaps. When utilizing information from both the muon system and the inner tracker, muon momentum can generally be determined to within 5%.

6.7 Trigger and DAQ

The two most critical components for the CMS detector to take data are the trigger and the data acquisition system (DAQ). The trigger's purpose is to analyze one billion proton-proton collisions per second and quickly decide which events are to be written to disk, and which are discarded. Because of the finite computing speed and storage capacity, a maximum

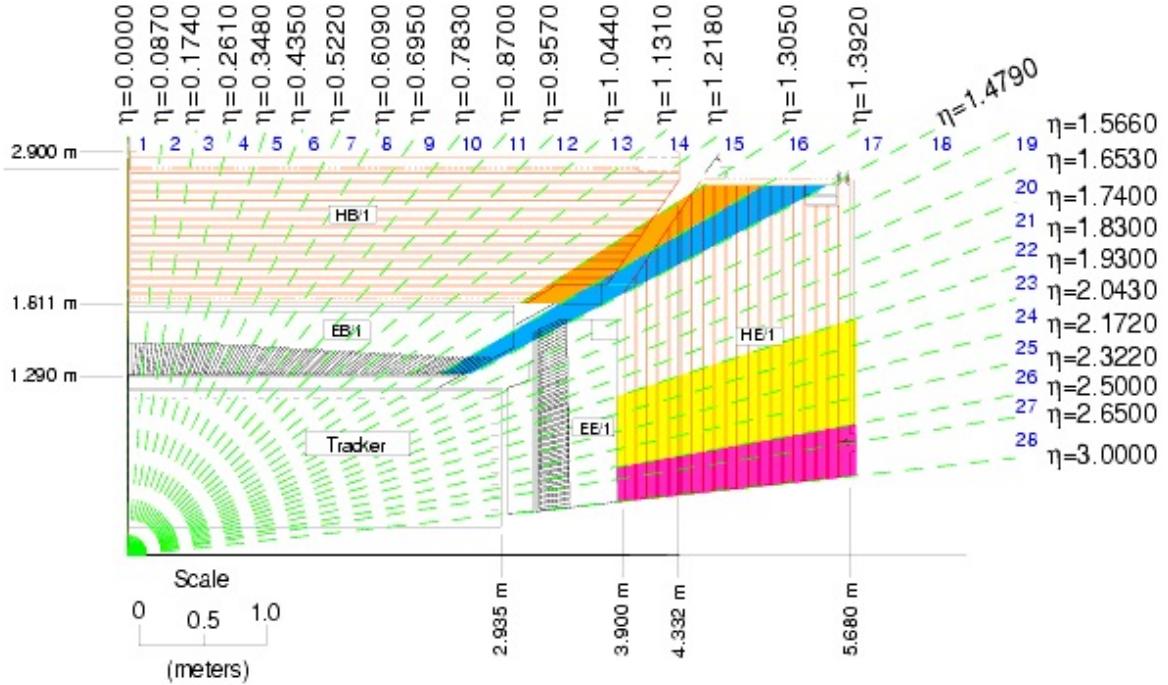


Figure 6.10: A schematic view of the CMS HCAL. The direction of view is toward the center of the LHC.

of about 1000 collision events per second can be recorded. Therefore, the vast majority of collisions are never reconstructed.

In order to reduce the number of events by 6 orders of magnitude, CMS employs a two-level trigger system. The two logically sequential trigger stages are: the level-1 (L1) trigger, and the high level trigger (HLT), each of which lowers the rate of events by a factor of a thousand. The L1 trigger is based on custom-manufactured programmable electronics and is responsible for analyzing each collision immediately after every bunch crossing, and discarding 99.9% of the events. This trigger has access to coarse information from the calorimeters and muon system, namely, energy deposits in the calorimeters and hits in the muon chambers, collectively called trigger primitives. The L1 trigger uses these primitives to compute estimates of the four-vectors of μ candidates and e/γ candidates, as well as estimates of the H_T and E_T^{miss} . The L1 trigger is allowed about $3 \mu\text{s}$ to analyze each event, and if it accepted, the event is passed to the HLT.

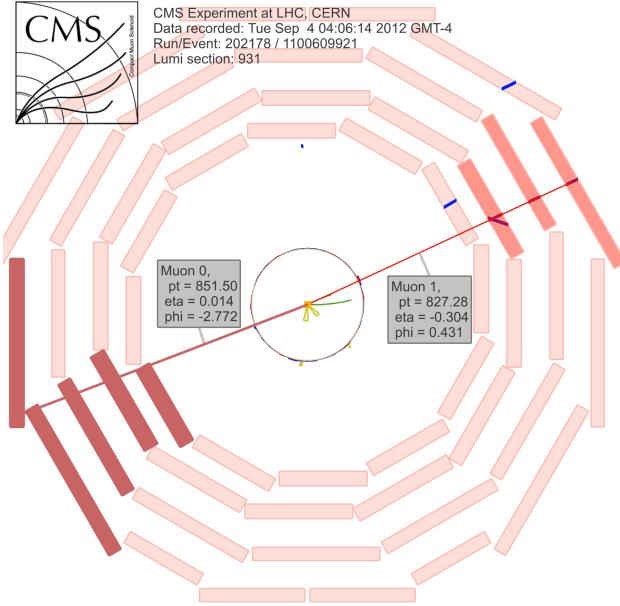


Figure 6.11: An event display of the CMS muon system. Two muons are reconstructed in the muon system.

The HLT, which is based in software, is capable of analyzing events with a much greater degree of sophistication than the L1. The information available to the high level trigger includes information from the CMS tracker, finely-granulated calorimetric depositions, and global objects that are constructed from the combination of all detectors. The determination of the objects such as the H_T and E_T^{miss} is much more precise in the HLT than in the L1, but less precise than the in the offline reconstruction. Thus, at the analysis level, the trigger efficiency must be measured using the offline information.

The DAQ is the computer system responsible for storing information about events that have been accepted by the trigger. It is capable of processing 100 Gigabytes, or 100 events per second on average. The DAQ integrates and analyzes the signals from about 55 million channels within the CMS detector, which are sent from about 600 independent front-end readout drivers. After processing, the DAQ sends the event information into mass storage on the CMS Tier-0 computing enter, hosted by the CERN Data Centre.

6.8 Particle and event reconstruction

The reconstruction of particles and event-level observables is performed by the particle flow algorithm [47]. Particle flow analyzes the basic elements from each detector system (tracks in the tracker, calorimeter clusters, and muon tracks), and associates the elements to form single particle candidates. The final result is a global event description that attempts to characterize every particle in each event. Jets and event-level observables such as the missing transverse energy E_T^{miss} are constructed from these particles.

Muons are identified based on information from the muon system and the inner tracker. Specifically, a muon is identified as a track in the muon system that is consistent in energy and direction with a track in the inner tracker. A selection is made on the χ^2 of a fit of the muon track to the combined hits in the tracker and muon system, and the momentum estimated by the muon system is required to fall within three standard deviations of the estimate made by the tracker.

Electrons are identified as tracks in the inner tracker that are spatially and energetically consistent with energy clusters in the ECAL. Because electrons can emit photons when traversing the tracker, their trajectories can exhibit abrupt changes. Therefore, a special tracking algorithm called a Gaussian Sum Filter [47] is used to model the trajectories of electron tracks in a realistic way. The algorithm that reconstructs energy clusters in the ECAL accounts for photon emission by identifying energy deposits consistent with possible trajectories of emitted photons, and adding them to the total cluster energy. After the track is associated with an ECAL cluster, information from the ECAL and tracker is used to identify electrons, including the quality of the fit associated with the reconstructed track, properties related to the distribution of the energy deposited in the ECAL, the ratio of the energy deposited in the HCAL behind the relevant ECAL cluster to that deposited in the ECAL, and the individual momentum estimates from the tracker and ECAL. Charged hadrons are identified as the set of track candidates matched to ECAL clusters that did not pass the electron identification criteria.

Photons are identified in a similar manner as electrons, but it is required that no track point in the direction of the electromagnetic energy cluster, and little or no energy deposited in the HCAL. Remaining energy deposits in the HCAL are identified as neutral hadrons.

The result of the above algorithm is a set of particles, called particle flow candidates, representing all identified particles in the event. Jets are reconstructed by applying the anti- k_T algorithm [50], with a distance parameter of 0.4 or 0.5, to the particle flow candidates. Estimates of the jet energy are corrected for effects of pile-up and for non-uniformities in the detector. The jet energy resolution ranges from 5–20%, depending on the jet momentum and direction. The E_T^{miss} is computed as the magnitude of the vectorial sum of the four-momenta of all particle flow candidates and jets in the event, after the energy corrections have been applied to the jets.

6.9 Phase II upgrade

In the year 2023, the high luminosity LHC (HL-LHC) will begin operations of what is referred to as Phase II of the LHC. At that stage, the instantaneous luminosity will increase by a factor of 10 from its Run 1 design value, up to $\mathcal{L} \approx 10^{35} \text{ cm}^{-2}\text{s}^{-1}$. A primary motivation for this transformation is to make it possible to study in great detail whatever phenomena may be discovered in the current phase of operation. This LHC upgrade will be accompanied by an upgrade in the hardware of the CMS detector, including improvements to various subsystems, and in some cases, the replacement of the old subsystems with new detectors. The tracker will likely be extended to cover the pseudorapidity range up to $|\eta| < 4$, the muon system will likewise be extended into the forward region, the level 1 trigger system will be modified so that it considers information from the tracker in addition to the calorimeters and muon system, and the ECAL and HCAL endcaps will be replaced by more sophisticated detectors.

Work has been carried out in preparation for the last item in this list, the calorimeter endcap upgrade. This section describes a portion of this work that I carried out. Some of the results were shown at the 2014 CMS upgrade Jamboree, and were included as part of

the information that informed the decision of the down-select, in which one endcap design was selected for construction.

6.9.1 Calorimeter research and development

Two options were considered for the set of detectors that will replace the CMS endcaps during the Phase II upgrade [51]: a sampling “shashlik” calorimeter, and a high granularity calorimeter (HGCAL). Ultimately, the HGCAL design was chosen for construction. The shashlik calorimeter, in design, is composed of towers 114 mm in length with square cross sections of 14×14 mm. Each tower (Figure 6.12) is composed of 28 layers of Tungsten

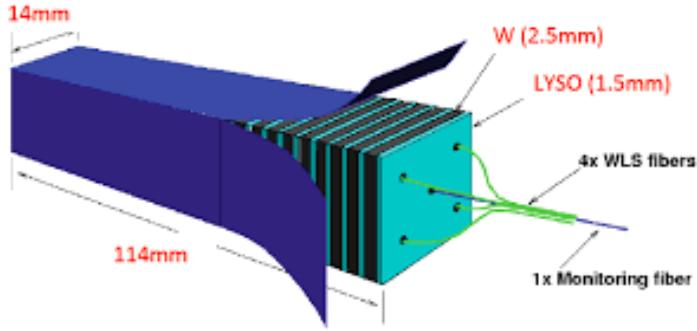


Figure 6.12: A shashlik calorimeter tower.

alternating with 29 layers of LYSO, where LYSO ($\text{LuY}_3\text{SiO}_5(\text{Ce})$) is a high density (7.1 g/cm³) scintillating material largely developed for use in particle detectors by Caltech [52]. The HGCAL is a sampling calorimeter that will be built from alternating lead and copper sampling layers, with signal readout from silicon photosensors. The HGCAL design is notable for its ability to provide longitudinal as well as lateral information about the distribution of energy deposited by high-energy particles.

In April-May of 2014 and July-August, 2014, test beam experiments were held at Fermilab at the M-Test facility [53] to study and characterize a prototype module of the shashlik calorimeter (Figure 6.13 illustrates the prototype structure). The testbeam facility provided access to beams of particles with an energy per particle ranging from 1 to 120 GeV. Several characteristics of the beam could be chosen in real time, such as the species of particles

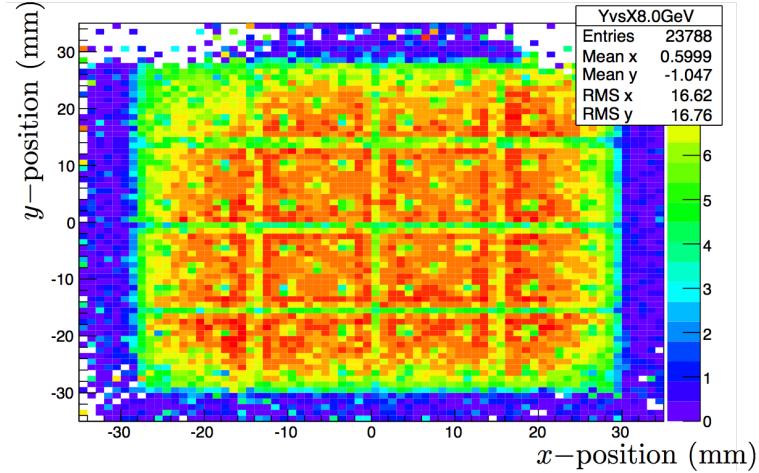


Figure 6.13: A scatter plot of the accumulated track positions weighted by the in-time energy deposited in the shashlik tower, for a set of muon runs recorded in July, 2014. The 4×4 tower array structure, the gaps between towers, and the holes bored for the WLS fibers can be seen as green features.

populating the beam, as well as the energy per particle. The facility also provided the use of two wire chambers, which were used to determine the time and position of incident particles. The shashlik prototype consisted of a 4×4 array of towers, and each tower (Figure 6.12) was read out by four wavelength-shifting (WLS) fibers fed longitudinally through the towers and connected to silicon photomultiplier sensors (SiPMs).

While working help set up and manage the testbeams, I authored an interactive event display that was used to view data as it was being collected. The event display is a flexible, python-based program with a graphical user interface (GUI) that can be used to view both live as well as stored data. The GUI is largely based on the TEve class of the Root software package [54], which makes use of a slots and signals framework to enable communication between objects defined in the program. Signals carry information related to one or several user-defined objects, can be produced as a result of various actions taken by the user or from the data stream, while slots are a class of objects that are capable of receiving this information and initiating changes to their associated user-defined objects. For a concrete example, after processing the tracking information for a given event, the program produces a signal that is received by the module that projects a 3-dimensional event view, which is

updated to display a set of tracks.

The event display contains several view panels, including

- a heat map (color plot) of the maximum analog-to-digital count (ADC) recorded by each channel during an event in real space, as shown in Fig. 6.14 (top);
- a display of the readout and χ^2 of the samples by channel, and a display of the pedestal noise, as shown in Fig. 6.14, (bottom);
- a display of the pedestal noise by channel, where the pedestal is the expected number of ADC counts in the absence of any signal, as shown in Fig. 6.15;
- a view of the in time hits in the wire chambers, where “in time” refers to the criterion that the recorded time of the signal be consistent with the time of the recorded energy deposited in the calorimeter, as shown in Fig. 6.15, and
- a 3-dimensional view of the hits, tracks, wire chambers, and the shashlik calorimeter with color corresponding to the estimated energy deposited in each tower, as shown in Fig. 6.16.

The event display has since been adopted for use at testbeam experiments dedicated to studying the HGCAL, where a prototype of the future calorimeter is being characterized. The software for the event display can be obtained at the url in Ref. [55].

Jet/ E_T^{miss} analysis. After testbeam operations, I and my colleague Arka Santra performed analysis of the data collected, as well as of simulated events. The energy resolution of jets in the CMS endcaps for various phases of the CMS detector are shown in Fig. 6.17. At large p_T , the energy resolution in PU 140 events is seen to drop below 10%, indicating an excellent level of performance that surpasses that described in Ref. [51].

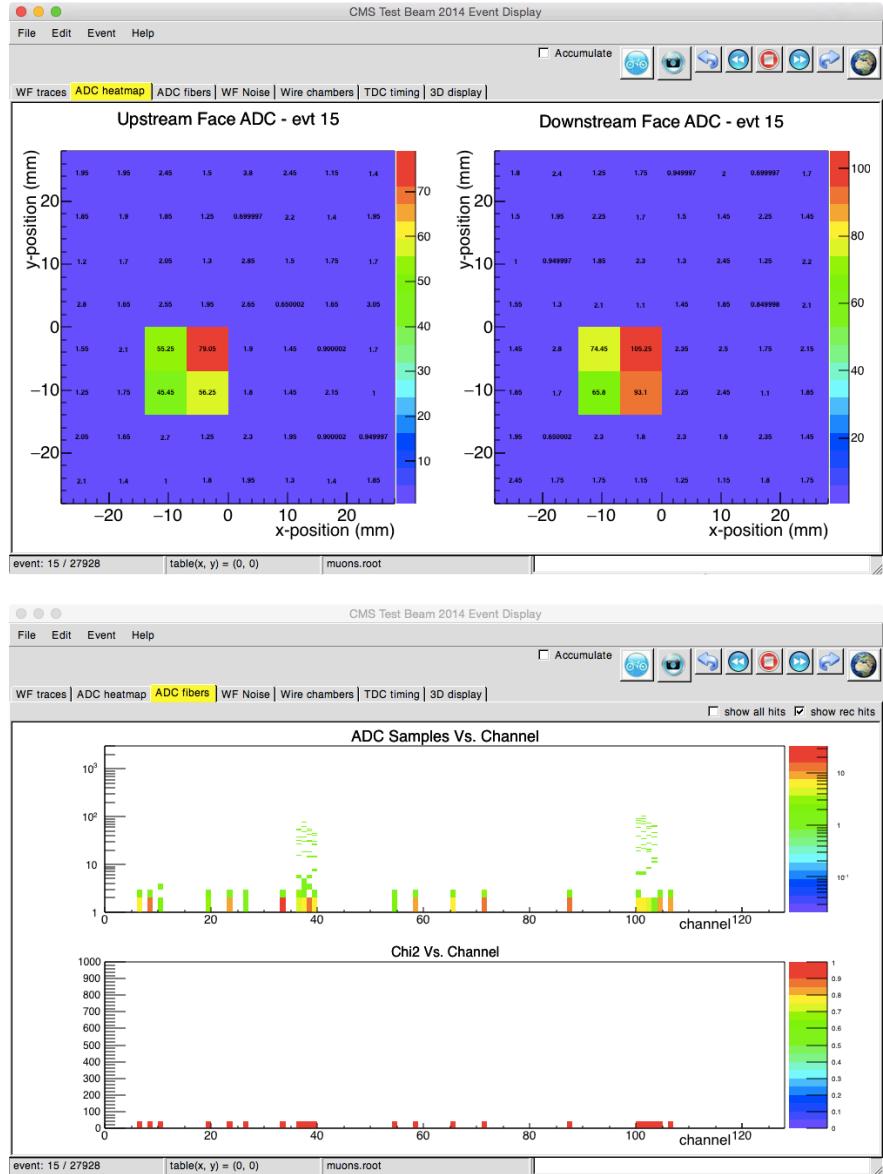


Figure 6.14: Top: heat maps of the ADC counts recorded at the upstream and downstream faces of the shashlik module. Bottom: Diagnostic displays of channel-by-channel information from the shashlik module. The "Chi2 Vs. Channel" display is effective at revealing noisy channels, that is, channels with a large ADC count generated by electronic noise rather than a genuine particle signal.

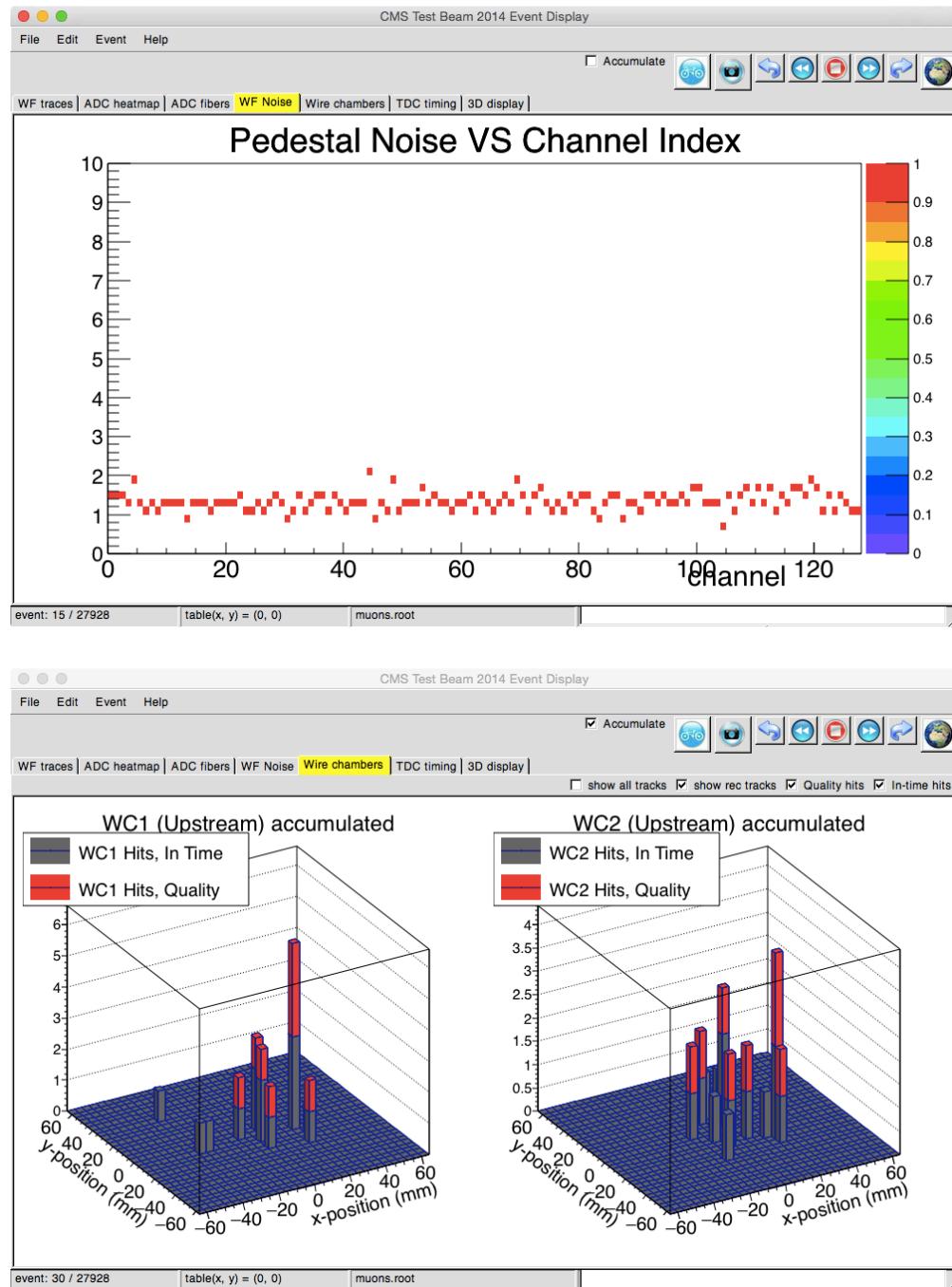


Figure 6.15: Top: The pedestal noise by channel. Bottom: A scatter plot of the positions of accumulated hits and tracks recorded by the wire chambers; quality hits are colored red, and refer to hits consistent with a particle track.

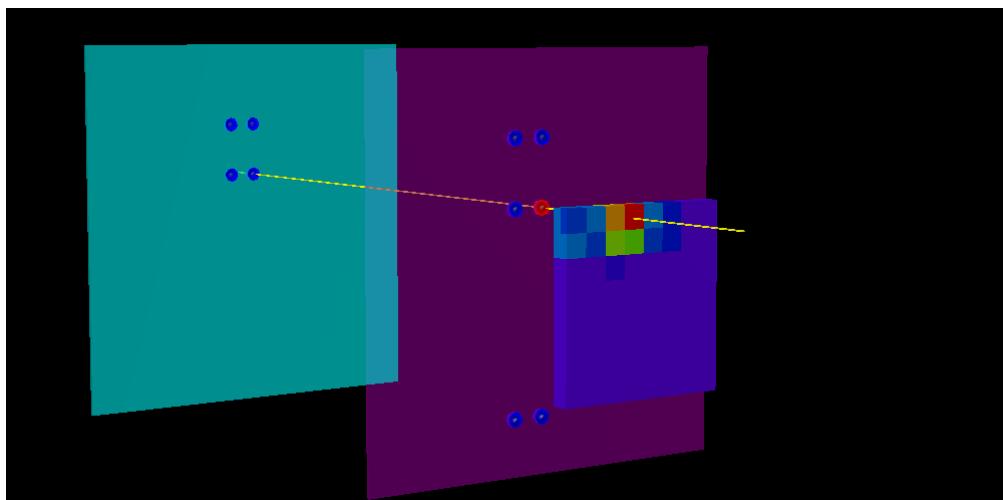
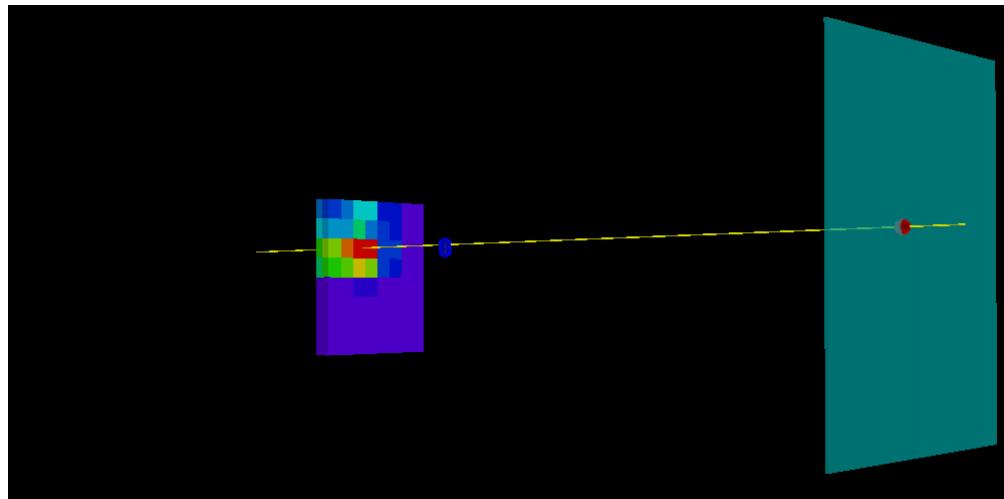


Figure 6.16: Three-dimensional views of reconstructed tracks and calorimetric energy deposits in events containing a 120 GeV proton. The view can be rotated by the user, allowing different viewing perspectives of the event, as illustrated in the top and bottom figures.

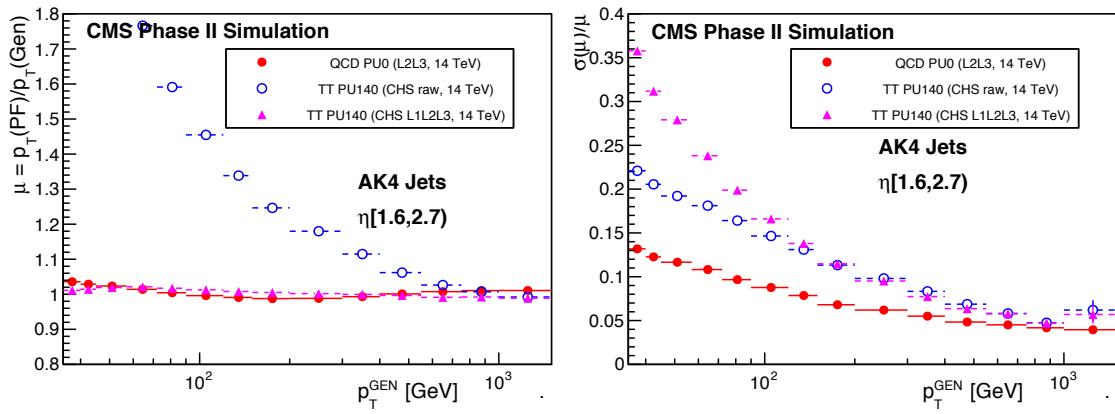


Figure 6.17: The jet response (left) and energy resolution (right) in the shashlik endcap as a function of simulated jet p_T in a sample of simulated $t\bar{t}$ events. The red distributions correspond to the scenario without pile-up, and pink to that with pile-up of 140 (an average of 140 collisions per bunch crossing), which is the expected amount of pile-up during the post-Phase II LHC operations.

CHAPTER 7

THE VIABILITY OF SUSY: THE PMSSM INTERPRETATION

CMS has performed many searches for evidence of SUSY at the LHC. To assess the impact of a representative set of these searches on the MSSM, an interpretation has been performed in the framework of the phenomenological MSSM (pMSSM), taken as a proxy for the MSSM (see Section 4.4 for an introduction to the pMSSM). A global Bayesian analysis is performed, incorporating results from a broad range of CMS searches, as well as constraints from other experiments. Because the pMSSM incorporates several well-motivated assumptions that reduce the 120 parameters of the MSSM to just 19 parameters defined at the electroweak scale, it is possible to assess the how the MSSM is constrained in a relatively straightforward way.

This chapter begins with a brief introduction to Bayes’ theorem, and follows with modified excerpts from the paper [29]. These sections are augmented by additional key details not discussed in the paper, including a discussion of the robustness of the results with respect to the choice of prior (prior is defined in the next section), and an expanded discussion of the non-excluded regions. But first, a short introduction to Bayes’ theorem.

7.1 Bayes’ theorem

Bayes’ theorem, or “the Pythagorean theorem [of] probability,” according to eminent statistician Harold Jeffreys [56], is a fundamental statement in probability theory relating conditional and absolute probabilities. If A and B represent two potentially true statements,

- $P(A)$ and $P(B)$ are the absolute probabilities of A and B being true, respectively, and
- $P(A|B)$ is the conditional probability of A being true given that B is true,

then Bayes' theorem is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (7.1)$$

Bayes' theorem is applicable in many physics problems as a means of inference, namely, as a way of inferring knowledge about a theory given data:

$$P(\text{theory}|\text{data}) = \frac{P(\text{data}|\text{theory})P(\text{theory})}{P(\text{data})} = P(\text{data}|\text{theory})P(\text{theory}), \quad (7.2)$$

where $P(\text{data})$ has been set to 1. In this expression, the factor $P(\text{theory})$ is called the prior probability density, and incorporates knowledge about a given theory prior to the consideration of the data at hand; $P(\text{data}|\text{theory})$ is referred to as the likelihood, and incorporates the data into our knowledge about the theory; $P(\text{theory}|\text{data})$ is called the posterior probability density, and represents our state of knowledge about the theory given all information available. If the posterior density for a given parameter differs significantly from its prior density, then the data have provided useful information about the parameter. To make meaningful conclusions, it is necessary to start with a prior that encodes as much relevant information as possible. Bayesian inference is used in the following sections, as well as in the Chapter 8 in the context of QCD background estimation.

7.2 The pMSSM interpretation

The purpose of this work is to assess how the current data constrain the MSSM using the more tractable pMSSM as a proxy. The conclusions are based on the observations of a representative subset of CMS search results that analyzed data corresponding to integrated luminosities of 5.0 fb^{-1} at 7 TeV and 19.5 fb^{-1} at 8 TeV. The considered searches include hadronic searches, both general searches and searches targeting top squark production; also included are searches with leptonic final states, both general and EW-targeted, as well as monojet searches. For a selected set of pMSSM parameter points, event samples were simulated using the CMS fast detector simulation [57] and analyzed. The 7 and 8 TeV data are treated consistently; in particular, the same set of points in the pMSSM model phase space are used in the analysis of all searches. This approach greatly facilitates the combination of the results from the 7 and 8 TeV (Run 1) data. The points were chosen randomly

from a larger set of points that are consistent with pre-LHC experimental results and basic theoretical constraints.

The statistical analysis is based largely on the Bayesian approach of Refs. [58, 59]. The work is an extension of Ref. [60], which interpreted three independent CMS analyses based on an integrated luminosity of about 1 fb^{-1} of data [61, 62, 63] in terms of the pMSSM, confirming that the approach is both feasible and more successful in yielding general conclusions about SUSY than those based on constrained SUSY models. Furthermore, the diversity of phenomena covered by the pMSSM is also helpful in suggesting new approaches to searches for SUSY at the LHC. A similar study has been performed by the ATLAS experiment [39].

7.3 Methods for the combination CMS analyses

The results from the selected CMS analyses are used to construct posterior densities of model parameters, masses, and observables. The posterior density of the model parameters, which are denoted by θ , is given by

$$p(\theta|D^{\text{CMS}}) \propto L(D^{\text{CMS}}|\theta) p^{\text{non-DCS}}(\theta), \quad (7.3)$$

where D^{CMS} denotes the data analyzed by the direct CMS SUSY searches, $L(D^{\text{CMS}}|\theta)$ is the associated CMS likelihood that incorporates the impact of these direct CMS searches, and $p^{\text{non-DCS}}(\theta)$ is the prior density constructed from results not based on direct CMS SUSY searches (non-DCS results). The posterior density for a single observable λ is obtained by integrating the full posterior density over the parameters,

$$p(\lambda|D^{\text{CMS}}) = \int \delta[\lambda - \lambda'(\theta)] p(\theta|D^{\text{CMS}}) d\theta, \quad (7.4)$$

where $\lambda'(\theta)$ is the value of the observable as predicted by model point θ (θ identifies the model point). Equation 7.4 is approximated using Monte Carlo (MC) integration.

7.3.1 Construction of the prior

The prior $p^{\text{non-DCS}}(\theta)$ encodes the boundaries of the considered parameter space, the constraints from non-DCS data, and several theoretical conditions. It is formulated as a

product of four factors,

$$p^{\text{non-DCS}}(\theta) \propto \left[\prod_j L(D_j^{\text{non-DCS}} | \lambda_j(\theta)) \right] \cdot p(c\tau(\tilde{\chi}^\pm) < 10 \text{ mm} | \theta) \cdot p(\text{theory} | \theta) \cdot p_0(\theta), \quad (7.5)$$

which I will explain one by one.

The initial prior $p_0(\theta)$ is taken to be uniform in the pMSSM sub-space (see Section 4.4 for definitions of the parameters),

$$\begin{aligned} -3 &\leq M_1, M_2 \leq 3 \text{ TeV}, \\ 0 &\leq M_3 \leq 3 \text{ TeV}, \\ -3 &\leq \mu \leq 3 \text{ TeV}, \\ 0 &\leq m_A \leq 3 \text{ TeV}, \\ 2 &\leq \tan \beta \leq 60, \\ 0 &\leq m_{\tilde{Q}_{1,2}}, m_{\tilde{U}_{1,2}}, m_{\tilde{D}_{1,2}}, m_{\tilde{L}_{1,2}}, m_{\tilde{E}_{1,2}}, m_{\tilde{Q}_3}, m_{\tilde{U}_3}, m_{\tilde{D}_3}, m_{\tilde{L}_3}, m_{\tilde{E}_3} \leq 3 \text{ TeV}, \\ -7 &\leq A_t, A_b, A_\tau \leq 7 \text{ TeV}, \end{aligned} \quad (7.6)$$

and the formally unbounded SM subspace defined by m_t , $m_b(m_b)$, and $\alpha_s(m_Z)$; the non-DCS measurements, which are listed in Table 7.1, constrain these SM parameters to narrow ranges. The sub-space defined in Eqs. (7.6) covers the phenomenologically viable parameter space for the LHC and is large enough to cover sparticle masses to which the LHC might conceivably be ultimately sensitive. The term $p(\text{theory} | \theta)$ imposes the theoretical constraints listed at the end of Section 4.4. In this study, signal events were simulated using a fast detector simulation program [57] that does not accurately model the detector response to massive long-lived charged particles traversing the calorimeters. Therefore, the parameter space considered has been restricted to the set of model points for which the chargino decays quickly, having a mean lifetime of $c\tau(\tilde{\chi}^\pm) < 10 \text{ mm}$. The factor $p(c\tau(\tilde{\chi}^\pm) < 10 \text{ mm} | \theta)$ imposes this requirement. Both $p(\text{theory} | \theta)$ and $p(c\tau(\tilde{\chi}^\pm) < 10 \text{ mm} | \theta)$ are unity if the inequalities are satisfied and zero otherwise.

The product of likelihoods $L(D_j^{\text{non-DCS}} | \lambda_j(\theta))$ in Equation (7.5) over measurements j is associated with non-DCS data, $D^{\text{non-DCS}}$, which imposes constraints from precision measurements and a selection of pre-LHC searches for new physics. The measurements used and

their associated likelihoods are listed in Table 7.1. Data from DM experiments has not been included in the prior in order to avoid bias from cosmological assumptions (e.g., DM density and distribution, assumption of one thermal relic, no late entropy production, etc.).

Since the explicit functional dependence of the prior $p^{\text{non-DCS}}(\theta)$ on θ is not available *a priori*, but the predictions $\lambda(\theta)$ are available point by point, it is natural to represent the prior as a set of points sampled from it. Owing to the complexity of the parameter space, the sampling is performed using a Markov chain Monte Carlo (MCMC) method [64, 65, 66, 67, 59].

All data in Table 7.1 except the Higgs boson signal strengths $\mu_h = \sigma/\sigma_{\text{SM}}$ were used in the original MCMC scan. The μ_h measurements were incorporated into the prior post-MCMC. A number of measurements, marked “reweight” in the last column, were updated during the course of this study as new results became available. The weights, applied to the subset of scan points which were selected for simulation, were computed as the ratio of the likelihoods of the new measurements shown in Table 7.1 to the previous measurements.

For a given point θ , the predictions $\lambda(\theta)$ — including those needed to calculate the likelihoods $L(D^{\text{non-DCS}}|\lambda(\theta))$ — are obtained as follows. The physical masses and interactions are calculated using the SUSY spectrum generator SOFTSUSY 3.3.1 [78], with the input parameters θ defined at $M_{\text{SUSY}} = \sqrt{m_{\tilde{t}_1} m_{\tilde{t}_2}}$. This calculation includes 1-loop corrections for sparticle masses and mixings, as well as 2-loop corrections for the small Higgs boson mass. Low-energy constraints are calculated with SUPERISO v3.3 [79]. MICROMEGAs 2.4.5 [75, 76, 77] is used to check the compatibility of pMSSM points with sparticle mass limits from LEP and other pre-LHC experiments. MICROMEGAs is also used to compute the DM relic density, and the spin-dependent and spin-independent DM-nucleon scattering cross sections; these observables are not used in the construction of the prior, but it is shown how they are affected by the CMS searches. The program SDECAY 1.3 [80] is used to generate sparticle decay tables and HDECAY 5.11 [81] to generate Higgs boson decay tables. The program LILITH 1.01 [72, 73] is used for evaluating the Higgs boson signal likelihood based on ATLAS [82] and CMS [83] measurements, following the approach explained in Section 2.3 of Ref. [84]. The experimental results used in LILITH are the signal strengths of the Higgs boson

Table 7.1: The Measurements that are the basis of the non-DCSprior $p^{\text{non-DCS}}(\theta)$ for the pMSSM parameters, their observed values and likelihoods. The observables are the decay branching fractions $\mathcal{B}(b \rightarrow s\gamma)$ and $\mathcal{B}(B_s \rightarrow \mu\mu)$, the SUSY to SM ratio for the branching fraction of the decay $B^+ \rightarrow \tau\nu$ ($R[B^+ \rightarrow \tau\nu]$), the difference in the muon anomalous magnetic moment from its SM prediction Δa_μ , the strong coupling constant at the Z boson mass $\alpha_s(m_Z)$, the top and bottom quark masses m_t and $m_b(m_b)$, the Higgs boson mass m_h and signal strength μ_h , and sparticle mass limits from LEP. All data except μ_h were used in the initial MCMC scan. See text for details.

i	Observable $\mu_i(\theta)$	Constraint $D_i^{\text{non-DCS}}$	Likelihood function $L[D_i^{\text{non-DCS}} \mu_i(\theta)]$	Comment
1	$\mathcal{B}(b \rightarrow s\gamma)$ [68]	$(3.43 \pm 0.21^{\text{stat}} \pm 0.24^{\text{th}} \pm 0.07^{\text{sys}}) \times 10^{-4}$	Gaussian	reweight
2	$\mathcal{B}(B_s \rightarrow \mu\mu)$ [69]	$(2.9 \pm 0.7 \pm 0.29^{\text{th}}) \times 10^{-9}$	Gaussian	reweight
3	$R(B^+ \rightarrow \tau\nu)$ [68]	1.04 ± 0.34	Gaussian	reweight
4	Δa_μ [70]	$(26.1 \pm 6.3^{\text{exp}} \pm 4.9^{\text{SM}} \pm 10.0^{\text{SUSY}}) \times 10^{-10}$	Gaussian	
5	$\alpha_s(m_Z)$ [1]	0.1184 ± 0.0007	Gaussian	
6	m_t [71]	$173.20 \pm 0.87^{\text{stat}} \pm 1.3^{\text{sys}}$ GeV	Gaussian	reweight
7	$m_b(m_b)$ [1]	$4.19_{-0.06}^{+0.18}$ GeV	Two-sided Gaussian	
8	m_h	LHC: $m_h^{\text{low}} = 120$ GeV, $m_h^{\text{high}} = 130$ GeV	$1 \text{ if } m_h^{\text{low}} \leq m_h \leq m_h^{\text{high}}$ $0 \text{ if } m_h < m_h^{\text{low}} \text{ or } m_h > m_h^{\text{high}}$	reweight
9	μ_h	CMS and ATLAS in LHC Run 1, Tevatron	LILITH 1.01 [72, 73]	post-MCMC
10	Sparticle masses	LEP [74] (via MICROMEGA [75, 76, 77])	1 if allowed 0 if excluded	

decay modes $Y = (\gamma\gamma, WW^*, ZZ^*, b\bar{b}, \tau\bar{\tau})$ in terms of the primary Higgs boson production modes gluon-gluon fusion (ggF), vector boson fusion (VBF), associated production with a W or Z boson (Wh and Zh, commonly denoted as Vh), and associated production with a top-quark pair (tth) as published by ATLAS, CMS, and Tevatron experiments. When these signal strengths are given as 2-dimensional (2D) confidence level (CL) contours in, e.g., the $\mu_{\text{ggF+tth}}(Y)$ versus $\mu_{\text{VBF+Vh}}(Y)$ plane, the likelihood is approximated by fitting a 2D Gaussian function to the 68% CL contour provided by the experiments. For each experiment, a χ^2 is computed using $-2 \log L_Y = \chi_Y^2$ for each decay mode Y , and the combined χ^2 is then obtained by summing over all the individual χ_Y^2 values. Additional information on signal strengths (and invisible decays) in one dimension is included analogously, using the published likelihood function when available or else the Gaussian approximation.

The uncertainty in the anomalous magnetic moment of the muon includes a component that accounts for theoretical uncertainties in the SUSY calculations.

The large window on the Higgs boson mass of 120–130 GeV covers the theoretical uncertainty in the Higgs boson mass calculation in the MSSM. All tools use the SUSY Les Houches accord [85] for data entry and output. Approximately 20 million points are sampled from $p^{\text{non-DCS}}(\theta)$ using multiple MCMC chains, but omitting the prompt chargino requirement. When that requirement is imposed, the number of sampled points is reduced by 30%, and the fraction of bino-like LSPs (see Chapter 4 Section 4.2) is enhanced from about 40 to 50%. A random subsample of 7200 points is selected for simulation studies. Given the large dimensionality of the model, this is a rather sparse scan. Nevertheless, the scan density is sufficient to learn much about the viability of the pMSSM model space. Distributions of model parameters in this subsample were compared with distributions from independent subsamples of similar size, as well as distributions from the original large sample, and consistency between distributions was observed within statistical uncertainties.

7.3.2 Incorporation of the CMS data

The analyses considered in this work are those listed in Table 7.2, and explore final-states characterized by a variety of event-level observables: the scalar sum of the transverse momenta of jets (H_T); the magnitude of the vector sum of the transverse momenta of final-state particles (E_T^{miss} or H_T^{miss}); a measure of the transverse mass [86] (see Appendix C for more details) in events in which two particles each decay to one invisible and one reconstructed particle (M_{T2}); the multiplicity of jets identified as originating from a b quark (b-jets); and a range of lepton multiplicities, including opposite-sign (OS) and like-sign (LS) lepton pairs. Other analyses that were not included in this study but which may impose additional constraints on the model space include searches for SUSY in the single lepton channel with one or multiple b-jets [87] and searches for top squark production [88] in the single lepton channel. The searches considered comprise hundreds of signal regions and address a large diversity of possible signal topologies.

The CMS likelihoods $L(D^{\text{CMS}}|\theta)$ are calculated for each of these analyses (or combinations of analyses), using different forms of likelihood depending on the nature of the results that are available. The first form of likelihood (*counts*) uses observed counts, N , and as-

Table 7.2: The CMS analyses considered in this study. Each row gives the analysis description, the center-of-mass energy at which data were collected, the associated integrated luminosity, the likelihood used, and the reference to the analysis documentation.

Analysis	\sqrt{s} [TeV]	\mathcal{L} [fb^{-1}]	Likelihood
Hadronic $H_T + H_T^{\text{miss}}$ search [89]	7	4.98	counts
Hadronic $H_T + E_T^{\text{miss}} + \text{b-jets}$ search [90]	7	4.98	counts
Leptonic search for EW prod. of $\tilde{\chi}^0, \tilde{\chi}^\pm, \tilde{l}$ [91]	7	4.98	counts
Hadronic $H_T + H_T^{\text{miss}}$ search [92]	8	19.5	counts
Hadronic M_{T2} search [93]	8	19.5	counts
Hadronic $H_T + E_T^{\text{miss}} + \text{b-jets}$ search [94]	8	19.4	χ^2
Monojet searches [95]	8	19.7	binary
Hadronic third generation squark search [96]	8	19.4	counts
OS dilepton (OS ll) search [97] (counting experiment only)	8	19.4	counts
LS dilepton (LS ll) search [98] (only channels w/o third lepton veto)	8	19.5	counts
Leptonic search for EW prod. of $\tilde{\chi}^0, \tilde{\chi}^\pm, \tilde{l}$ [99] (only LS, 3 lepton, and 4 lepton channels)	8	19.5	counts
Combination of 7 TeV searches	7	-	binary
Combination of 7 and 8 TeV searches	7,8	-	binary

sociated background estimates, $B \pm \delta B$; the second (χ^2) uses profile likelihoods (Section A.2 gives a discussion on profile likelihoods), $T(\mu, \theta)$, where $\mu = \sigma/\sigma^{\text{SUSY}}(\theta)$ is the signal strength modifier and σ and $\sigma^{\text{SUSY}}(\theta)$ are the observed and predicted SUSY cross sections, respectively, while the third (*binary*) joins either of the first two kinds of result together with a signal significance measure Z , and is used for combining results from overlapping search regions. The three forms of the likelihood used and the signal significance measure Z are described in the following.

Counts likelihood For a single-count analysis, the likelihood is given by

$$L(D^{\text{CMS}}|\theta) = \int \text{Poisson}(N|s(\theta) + b) p(b|B, \delta B) db, \quad (7.7)$$

where N is the observed count, $s(\theta)$ and b are the expected number of signal and background counts, respectively, and $B \pm \delta B$ is the estimated number of background event

counts and its uncertainty. The prior density for b , $p(b|B, \delta B)$, is modeled as a gamma density, $\text{gamma}(x; \alpha, \beta) = \beta \exp(-\beta x)(\beta x)^{\alpha-1}/\Gamma(\alpha)$, with α and β defined such that the mode and variance of the gamma density are B and $(\delta B)^2$, respectively. For analyses that yield multiple independent counts, the likelihood is the product of the likelihoods of the individual counts. For analyses with multiple counts, the background predictions for the different search regions are treated as uncorrelated. Systematic effects on the signal counts are taken into account by varying the signal yield by multiplying it with a signal strength modifier μ with values $1 - \delta\mu, 1, 1 + \delta\mu$, where $\delta\mu$ is the fractional value of the systematic uncertainty.

χ^2 likelihood This likelihood is used for CMS searches that provide profile likelihoods, $T(\mu, \theta) \equiv L(D^{\text{CMS}}|\mu, \theta, \hat{\nu}(\mu, \theta))$, for the signal strength modifier μ , where ν represents the nuisance parameters and $\hat{\nu}(\mu, \theta)$ their conditional maximum likelihood estimates. Taking $\hat{\mu}$ to be the signal strength modifier that maximizes $T(\mu, \theta)$, it can be shown that the quantity $t = -2 \ln [T(1, \theta)/T(\hat{\mu}, \theta)]$ follows a χ^2 density with one degree of freedom in the asymptotic limit [100],

$$L(D^{\text{CMS}}|\theta) \approx \exp(-t/2)/\sqrt{2\pi t}, \quad (7.8)$$

which is adopted as the CMS likelihood in this case. The systematic uncertainties in the signal yield can again be incorporated by varying the values of μ .

Z -significance This study uses a signal significance measure defined by

$$Z(\theta) = \text{sign}[\ln B_{10}(D, \theta)] \sqrt{2 |\ln B_{10}(D, \theta)|}, \quad (7.9)$$

where

$$B_{10}(D, \theta) = \frac{L(D|\theta, H_1)}{L(D|H_0)} \quad (7.10)$$

is the local Bayes factor for data D , at point θ , and $L(D|\theta, H_1)$ and $L(D|H_0)$ are the likelihoods for the signal plus background (H_1) and background only (H_0) hypotheses, respectively. The function $Z(\theta)$ is a signed Bayesian analog of the frequentist “ n -sigma”. The case $Z \gg 0$ would indicate the presence of a signal at a significance of Z standard deviations, while the case $Z \ll 0$ would indicate the absence of signal, i.e., an *exclusion* at a significance of Z standard deviations. The Z -significance is the basis of the binary likelihood.

Binary likelihood This likelihood is used for combining results from search regions in which data may not be independent, for example, multiple counts from overlapping search regions. The data are first divided into subsets for which either a count or χ^2 likelihood can be calculated. For each subset j , with data D_j , $Z_j(\theta)$ is computed using Equation (7.9). An overall significance measure that includes all subsets under consideration is defined by

$$Z(\theta) \equiv Z_{j_{\max}}(\theta), \quad (7.11)$$

where j_{\max} is the index of the maximum element in the set $\{|Z_j(\theta)|\}$. This quantity is used to define the binary likelihood as follows,

$$L(D^{\text{CMS}}|\theta) = \begin{cases} 1 & \text{if } Z(\theta) > -1.64, \\ 0 & \text{if } Z(\theta) \leq -1.64, \end{cases} \quad (7.12)$$

where $Z(\theta) = -1.64$ corresponds to the frequentist threshold for exclusion at the 95% CL. Systematic uncertainties are incorporated by computing each $Z_j(\theta)$ with varying values of μ , and using these recalculated $Z_j(\theta)$ to compute the binary likelihood. Although use of the binary likelihood entails a loss of information, it is a convenient approach in cases of non-disjoint data, where a proper likelihood calculation is not feasible without more information. In this study, binary likelihoods are used for the monojet searches, which have overlapping search regions, and for combining the 7 TeV, and 7+8 TeV results, where the analyses use non-disjoint data.

To compute likelihoods and Z -significances, expected signal counts for the search regions of each analysis are computed for the 7200 pMSSM points. The simulated events for each model point, which were generated using PYTHIA 6.4 [101] and processed with the CMS fast detector simulation program [102], are passed through the analysis procedures in order to determine the counts. For each pMSSM point, 10,000 events have been simulated.

7.4 Results

The results of the study are presented in terms of three types of comparisons. The first type of comparison is of the distribution of the Z -significance for different combinations of

analyses. The second comparison is between the prior and posterior densities of the pMSSM parameters. The third comparison is between the total number of points in the pMSSM parameter space, and the number of points that survive the CMS analyses considered. For the last comparison, the survival probability is shown as a function of model parameters, where the survival probability in a region Θ of the pMSSM parameter space, is defined by

$$\frac{\int_{\Theta} p^{\text{non-DCS}}(\theta) H(Z(\theta) + 1.64) d\theta}{\int_{\Theta} p^{\text{non-DCS}}(\theta) d\theta}, \quad (7.13)$$

where H is the Heaviside step function with a threshold value $Z = -1.64$, which again is the threshold for exclusion at the 95% CL.

7.4.1 Global significance

Distributions of Z -significance are shown in Fig. 7.1 for all the CMS searches included in this study: 8 TeV searches, combinations of 7 TeV searches, and combinations of 7+8 TeV searches. The farther a Z distribution is from zero, the greater the impact of the analysis on the pMSSM parameter space. As noted in Section 7.3, negative and positive values indicate a preference for the background only (H_0) and the signal plus background (H_1) hypotheses, respectively.

All 8 TeV searches lead to distributions with negative tails, indicating that each disfavors some region of the pMSSM parameter space. The searches making the greatest impact are the $H_T + H_T^{\text{miss}}$ and M_{T2} searches, which disfavor a significant portion of the parameter space. The M_{T2} , $H_T + E_T^{\text{miss}} + b\text{-jets}$, EW, and OS dilepton searches, which yield modest excesses over the SM predictions, have Z -significances up to 4.

As expected, the combined 7+8 TeV result has a greater impact than any individual analysis. Overall, the impact of the 7 TeV combined result is relatively small as indicated by the high peak around zero.

7.4.2 Impact on parameters

Figure 7.2 shows the impact of the CMS searches on our knowledge of the gluino mass. Figures 7.2 (a)-(d) show marginalized distributions of the gluino mass. Posterior distributions obtained using three signal strength modifier values $\mu = 0.5, 1.0, 1.5$ illustrate the

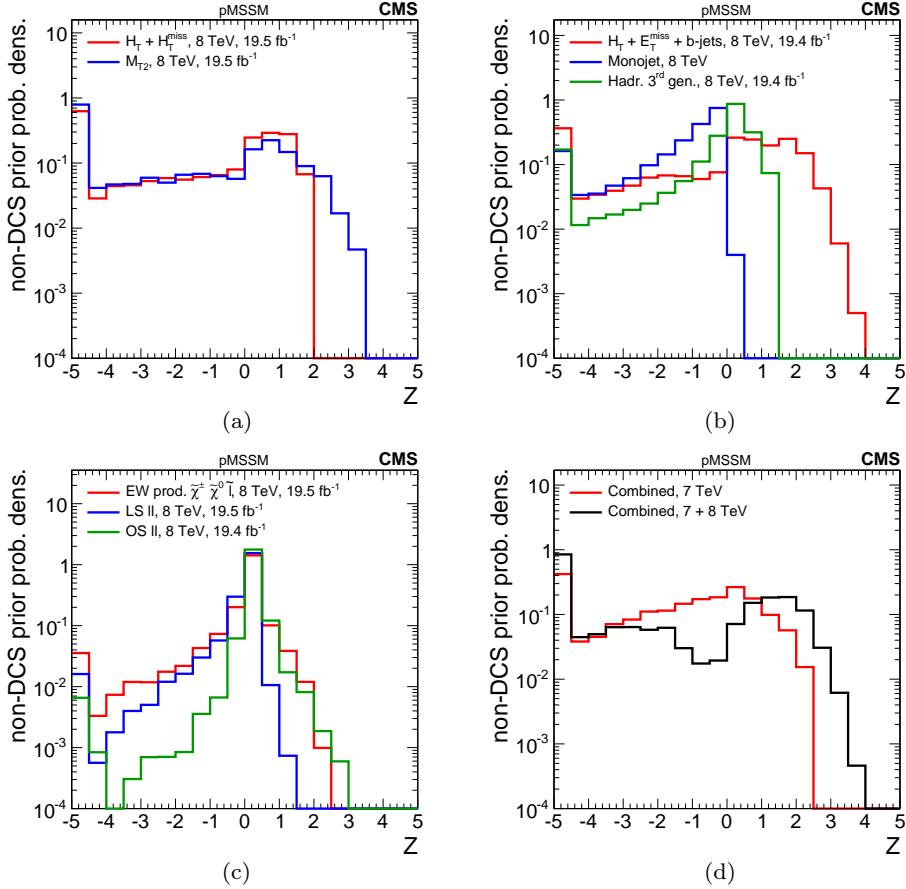


Figure 7.1: The distribution of model points, weighted by the non-DCS prior density, of the Z -significance for the individual 8 TeV searches (a–c), and for 7 TeV combined and 7+8 TeV combined searches (d). The leftmost bins contain the underflow entries.

effect of a $\pm 50\%$ systematic uncertainty in the predicted SUSY signal yields. Since the uncertainty in the signal efficiency typically varies between 10 and 25%, and the uncertainty in the signal cross section ranges between 30 and 50%, this prescription is considered to be conservative. Figure 7.2 (a) shows the strong impact of the hadronic analyses on the gluino mass distribution. The $H_T + H_T^{\text{miss}}$ search strongly disfavors the region below 1200 GeV, while the M_{T2} search leads to a distribution with two regions of peaking probability, one at relatively low mass, around 600 to 1000 GeV, and one above 1200 GeV. In Fig. 7.2 (b), the other hadronic analyses also disfavor the low-mass region, though to a lesser degree, and two of these analyses (the $H_T + E_T^{\text{miss}} + b$ -jets and the hadronic third generation) also exhibit

secondary preferred regions around 1100 GeV, while Fig. 7.2 (c) shows that the EW, OS dilepton, and LS dilepton searches have little impact on the gluino mass distribution. Figure 7.2 (d) compares the prior distribution to posterior distributions after inclusion of the combined 7 TeV and combined 7+8 TeV data. The 7 TeV data already have sufficient sensitivity to exclude much of the low-mass gluino model space, and the 8 TeV data further strengthen this result. The enhancements induced by the hadronic searches in the 800–1300 GeV range disappear in the combination since the observed excesses driving the enhancements are not consistent with a single model point or group of model points.

Figure 7.2 (e) shows the survival probability (Equation 7.13) as a function of gluino mass for the combined 7 TeV, and 7+8 TeV results. The CMS searches exclude all the pMSSM points with a gluino mass below 500 GeV, and can probe scenarios up to the highest masses covered in the scan. As may be expected, masses of order 3 TeV are not probed directly but rather through the production of lighter particles in the model. Finally, Fig. 7.2 (f) shows the Z -significance versus gluino mass. A slight negative correlation for positive Z values and gluino masses is observed below 1200 GeV; Z declines slightly as mass increases, which indicates that some small excesses of events observed by the various searches are consistent with models with light gluinos.

Figures 7.3 and 7.4 similarly summarize the impact of searches on the first- and second-generation left-handed up squark mass and the mass of the lightest colored SUSY particle (LCSP), respectively. The picture is similar to that for the gluino mass. For both \tilde{u}_L and the LCSP, the M_{T2} search shows a preference for masses from 500 to 1100 GeV. The overall impact of the searches on \tilde{u}_L is less than the impact on the gluino mass owing to the more diverse gluino decay structure that can be accessed by a greater number of searches. For the LCSP, the overall impact is the least because the LCSP has the fewest decay channels; nevertheless CMS searches exclude about 98% of the model points with an LCSP mass below 300 GeV; in the surviving 2% of these model points (6 points), the LCSP is always the \tilde{d}_R . The searches can be sensitive to scenarios with LCSP masses up to \sim 1500 GeV. Again, the Higgs boson results make a negligible contribution. In each case a negative correlation is

observed between the Z -significance and the sparticle mass for positive Z values and masses below 1200 GeV; this is most pronounced for the LCSP.

Figure 7.5 illustrates what information the searches provide about the mass of the lightest top squark \tilde{t}_1 . The difference between the prior and posterior distributions is minor. The reason is that the low-energy measurements such as the $b \rightarrow s \gamma$ branching fraction (see Table 7.1) impose much stronger constraints on the mass of the \tilde{t}_1 than do the analyses. This is not to say the CMS analyses are insensitive to top squark masses. The posterior distribution for the M_{T2} search exhibits an enhancement at $m_{\tilde{t}_1} < 1$ TeV relative to the non-DCS distribution. This enhancement does not appear in the combined posterior density because it is suppressed by observations of other more sensitive searches, which is a good illustration of the importance of doing global analyses. In the distribution of $m_{\tilde{t}_1}$ versus Z , the positive (negative) Z values have a slight negative (positive) correlation with the \tilde{t}_1 mass below 1 TeV, indicating that the CMS analyses have some direct sensitivity to top squarks with masses up to 1 TeV. The overall conclusion is that light top squarks with masses of the order of 500 GeV cannot be excluded.

Turning now to the EW sector, Fig. 7.6 shows the effect of the CMS data on our knowledge of the mass of the lightest neutralino $\tilde{\chi}_1^0$. The hadronic searches disfavor low $\tilde{\chi}_1^0$ masses; the hadronic searches targeting specific topologies also have an effect, although smaller, and the leptonic searches have a marginal impact. The 7+8 TeV combined distribution is very similar to the M_{T2} distribution, especially in the lower mass region, indicating that this search is the most sensitive to the $\tilde{\chi}_1^0$ mass. The main constraint on the $\tilde{\chi}_1^0$ mass arises indirectly through correlations with other sparticle masses. Since $\tilde{\chi}_1^0$ is the LSP, its mass is constrained by the masses of the heavier sparticles. As CMS searches push the probability distributions for the colored particles to higher values, more phase space opens for $\tilde{\chi}_1^0$ and the $\tilde{\chi}_1^0$ distributions shift to higher values. The survival probability distribution shows that no $\tilde{\chi}_1^0$ mass is totally excluded at the 95% CL by CMS. In general, the non-excluded points with light $\tilde{\chi}_1^0$ are those with heavy colored sparticles. The fact that the survival probability decreases below a $\tilde{\chi}_1^0$ mass of ~ 700 GeV shows that CMS searches are sensitive up to this mass value. The Higgs boson data disfavor neutralino masses below about 60 GeV, that is,

the mass range in which invisible decays $h \rightarrow \tilde{\chi}_1^0 \tilde{\chi}_1^0$ could occur; this is visible in the first bin in Fig. 7.6 (d) (see Ref. [72]).

In the MSSM, the lightest chargino becomes degenerate with the lightest neutralino when $|M_1| \geq \min(|M_2|, |\mu|)$. Therefore, it is informative to examine the impact of searches on the lightest non-degenerate (LND) chargino, defined as

$$\text{LND } \chi^\pm = \begin{cases} \tilde{\chi}_1^\pm & \text{if } |M_1| < \min(|M_2|, |\mu|) \\ \tilde{\chi}_2^\pm & \text{if } |M_1| > \min(|M_2|, |\mu|). \end{cases} \quad (7.14)$$

Figure 7.7 summarizes what information has been gained about the mass of the LND chargino. Again, the impact of the CMS searches is found to be rather limited and no chargino mass can be reliably excluded. It is worth noticing the impact of the leptonic searches. In Fig. 7.7 (c), the distributions differ from the non-DCS distribution, while these searches have negligible impact on most of the other SUSY observables and parameters. The survival probability is lowest in the first bin where the LND mass is between 0 and 200 GeV, but a small percentage of points still survive.

A more generic view is possible by looking at the overall CMS impact on the inclusive SUSY production cross section for 8 TeV, which is shown in Fig. 7.8. The most probable total sparticle cross section in the non-DCS prior is approximately 100 fb. The effect of the CMS SUSY searches is to reduce this value by an order of magnitude. The $H_T + H_T^{\text{miss}}$ search has the largest individual contribution to this because of its ability to address a great diversity of final states comprising different sparticle compositions. The survival probability distribution confirms that CMS is sensitive to SUSY scenarios with total cross sections as low as 1 fb.

In Fig. 7.9, the non-DCS and post-CMS distributions are compared after 7 and 7+8 TeV data for several other observables. The impact of the CMS data on the first and second generation right-handed up squarks is lower than on the corresponding left-handed up squarks. This is because left-handed up squarks in the MSSM form doublets with mass-degenerate left-handed down squarks, while the right-handed up and down squarks are singlets and their masses are unrelated. Therefore, the sensitivity to left-handed up squarks of a given mass is increased by the left-handed down squarks of the same mass. A mild impact on

the bottom squark mass is observed, where CMS disfavors masses below 400 GeV. The searches also have some sensitivity to the selectron and stau masses, which comes from the leptonic searches. The impact on $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ masses is relatively larger, mostly due to the dedicated EW analyses. The searches have little or no impact on the masses of the light and heavy pseudoscalar Higgs bosons. The preference of the Higgs data for negative values of the higgsino mass parameter μ comes primarily from the fact that the measured signal strength normalized to its SM value for $Vh \rightarrow b\bar{b}$ (where V is a W or a Z boson) is currently slightly below one. Models with negative values of μ give rise to radiative corrections that reduce the bottom Yukawa coupling, thereby creating a preference for $\mu < 0$ [84]. The $\tan\beta$ distribution is largely unaffected by both the SUSY searches and the current Higgs boson data evaluated via LILITH 1.01.

The impact of CMS searches on some observables related to dark matter are also investigated. Figure 7.10 shows distributions of the dark matter relic density, the spin-dependent (SD) direct detection cross section, and spin-independent (SI) direct detection cross section. In Fig. 7.10 (a), the relic density is seen to take on a bimodal probability density. The lower peak corresponds primarily to model points with bino-like LSPs, and the upper peak is mainly due to points with wino- and higgsino-like LSPs. The combined CMS searches lead to a noticeable enhancement of the lower peak. In Fig. 7.10 (b) and (c), minor differences are seen between the prior and posterior densities for the direct detection cross section.

7.4.3 Correlations among pMSSM parameters

A virtue of high-dimensional models like the pMSSM is that they enable the examination of correlations among parameters not possible in the context of more constrained models.

Figure 7.11 compares marginalized distributions in two dimensions of non-DCS (left) and post-CMS distributions (middle), and also shows the post-CMS to non-DCS survival probability (right) for several observable pairs. The first row of distributions show that the CMS impact on our knowledge of the $\tilde{\chi}_1^0$ mass is strongly correlated with the gluino. Since $\tilde{\chi}_1^0$ is the LSP, light colored particles imply a light $\tilde{\chi}_1^0$. Consequently, the disfavoring of light colored sparticles implies the disfavoring of a light $\tilde{\chi}_1^0$. In the last row, it is seen

that the $\tilde{\chi}_1^0$ mass is correlated most strongly with the cross section and that light $\tilde{\chi}_1^0$ LSPs are indeed disfavored for the reason just given. However, scenarios with $\tilde{\chi}_1^0$ masses around 100 GeV can still survive even though they have cross sections above 1 pb! These and other high cross section model points are discussed in Section 7.5. The probability distributions and survival probability for $\tilde{\chi}_1^0$ versus \tilde{t}_1 mass are shown in the third row. Although the post-CMS probabilities shift towards higher values, the survival probabilities never really go to zero. Although current simplified model scenarios exclude large parts of the \tilde{t}_1 - $\tilde{\chi}_1^0$ plane, pMSSM scenarios with relatively low \tilde{t}_1 masses (500 GeV) are not significantly disfavored by the CMS searches.

As is true of any Bayesian analysis, the results depend both on the likelihood and the prior. Studies were therefore performed to assess how the conclusions would change if a different choice of initial prior had been made. A log-uniform prior ($p_0(\theta)$) in Equation 7.5) is found to yield posterior densities very similar to those using the uniform prior. The most significant exception are the densities for the masses of the $\tilde{\chi}_1^0$ and $\tilde{\chi}_1^\pm$, which are shifted 10–20% toward higher values with respect to the densities derived from the uniform prior. It is found that the marginalized likelihood distributions are consistent with the profile likelihoods, suggesting that a frequentist analysis based on the profile likelihoods would yield similar conclusions. More details are given in the next section. Details of these studies are provided in Appendix A.

7.5 Nonexcluded regions in the pMSSM parameter space

Of the 7200 pMSSM points considered in this study, about 3700 cannot be excluded by CMS analyses based on their Z -significance (Fig. 7.1 (d)), although more than half of these nonexcluded points have a total cross section greater than 10 fb at $\sqrt{s} = 8$ TeV. Given this unexpected result, it is of interest to characterize this nonexcluded subspace in order to shed light on why the CMS analyses are not sensitive to these points, which can help guide the design of future analyses. To this end, the nonexcluded subspace is decomposed

into the dominant physical processes, and an idealized analysis of final state observables is performed.

Decomposition into processes. For the decomposition, signal events are analyzed at the generator level for each model point, and the pair of SUSY particles most frequently produced directly from the proton-proton interaction is taken as the production mode for that model point. Then the principal (dominant) process for the point is built as a tree diagram starting from the pair of SUSY mother particles and following the decay modes with the highest branching fractions until endpoints consisting of only SM particles and LSPs are reached. Indices of particle charge, flavor, and chirality are ignored in the construction, with the exception of the flavor of the third-generation squarks and quarks. Over 100 distinct principal processes are found among the total 7200 studied points, of which the first twelve are listed in Fig. 7.12. Many of the principal processes are seen to correspond to common SMS scenarios, while others depict more unusual scenarios with long decay chains.

The distribution of principal processes for excluded and nonexcluded points is given in Fig. 7.13 (a). It is seen that processes involving direct gluino production (5 and 8) are excluded with a much higher frequency than they survive, and those with EW gaugino production (2, 3, and 10) survive with a higher frequency than they are excluded. Processes with first-generation squark production (1 and 7) survive and are excluded at similar rates, and processes with slepton production (12) have exceptionally high survival rates. These trends are likely attributable to the difference in the production cross section between colored and noncolored particles for a given SUSY mass scale. The overflow bin (other), which contains many principal processes, including modes of colored and noncolored particle production, indicates a survival rate approximately equal to the exclusion rate. The dominance is defined for each model point as the ratio of the cross section of the principal process to the total SUSY production cross section at 8 TeV,

$$\text{dominance} \equiv \sigma_{\text{principal}}^{8 \text{ TeV}} / \sigma_{\text{tot}}^{8 \text{ TeV}}, \quad (7.15)$$

and is shown in Fig. 7.13 (b). Most values of the dominance are in the range 0.05–0.60. The excluded and nonexcluded values for the dominance are seen to agree within the RMS of

the distributions, indicating that the presence of multiple event types within a single model hypothesis does not significantly impact our ability to exclude such a model point.

Dedicated searches exist that correspond to some of the most frequent principal processes, indicating areas where the SMS approach is likely well optimized. For example, points with principal processes 1, $\tilde{q}\tilde{q}(\tilde{q} \rightarrow q\tilde{\chi}_1^0)$, and 4, $\tilde{b}\tilde{b}(\tilde{b} \rightarrow b\tilde{\chi}_1^0)$, enjoy searches that target these processes explicitly. A few principal processes have not been explicitly targeted by the host of CMS SUSY searches, including processes 2, $\tilde{\chi}_1^\pm\tilde{\chi}_1^0(\tilde{\chi}_1^\pm \rightarrow W^\pm\tilde{\chi}_1^0)$, and 3, $\tilde{\chi}_1^\pm\tilde{\chi}_2^0(\tilde{\chi} \rightarrow V/h\tilde{\chi}_1^0)$, the asymmetric EW gaugino production modes. New searches that target these or the other overlooked processes may serve to broaden the overall sensitivity to the pMSSM.

7.5.1 Characterization of signal events

Next, the nonexcluded model space is characterized by the predicted final states in order to shed light on what signatures may serve to target the nonexcluded points in Run 2. A loose set of baseline physics objects and event variables are defined at the generator level, as follows:

- Leptons: electrons, muons, or taus having a transverse momentum p_T greater than 5 GeV and an isolation less than 0.2. Here, isolation = $[(\sum_i p_{T,i}) - p_T]/\sum_i p_{T,i}$, where the sums run over all detector-visible particles i within a ΔR cone of 0.5 around the object, with $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$;
- Jets: particles clustered with the anti-kT jet algorithm [50] with distance parameter 0.5. The jets are required to have a p_T greater than 20 GeV;
- b-jets: jets matched to a b hadron within a ΔR of 0.5;
- E_T^{miss} : the missing transverse energy, calculated as the magnitude of the vector sum of the transverse momenta of visible particles with $p_T > 5$ GeV;
- H_T : the scalar sum of the p_T of the jets with a $p_T > 50$ GeV.

A parallel coordinates visualization technique is used that enables the display of multiple dimensions (as introduced in Chapter 3). Figure 7.14, shows nonexcluded points corresponding to the six selected principal processes (those denoted by color in Fig. 7.14). Vertical axes

are chosen to represent meaningful properties of the model points, and each model point is represented as a curved line traversing the plot from left to right, intersecting each axis at the parameter value taken by the model point. The curvature of the lines is added to help distinguish between similar pMSSM points, but the trajectories of the lines between the axes do not carry physical information. A number of distinct scenarios are seen to have survived the CMS analyses. A minimum threshold of 20 fb has been applied to the 8 TeV signal cross sections to limit the scope to those points that could potentially still be probed with the Run 1 data set using an expanded set of analyses and techniques.

The nonexcluded points associated with principal processes 1, $\tilde{q}\tilde{q}(\tilde{q}\rightarrow q\tilde{\chi}_1^0)$, and 4, $\tilde{b}\tilde{b}(\tilde{b}\rightarrow b\tilde{\chi}_1^0)$, are seen to give rise to large average E_T^{miss} , jet multiplicities between 2 and 4, and moderate to low cross sections due to the large masses of the squarks. Given the higher cross sections in Run 2, these high E_T^{miss} scenarios will become increasingly more accessible.

Model points with principal processes 2, $\tilde{\chi}_1^\pm\tilde{\chi}^0(\tilde{\chi}_1^\pm\rightarrow W^\pm\tilde{\chi}_1^0)$, and 3, $\tilde{\chi}_1^\pm\tilde{\chi}_2^0(\tilde{\chi}\rightarrow V/h\tilde{\chi}_1^0)$, typically predict large cross sections, in the range 100 fb pb, but a limited number of physical observables, primarily due to compression in the mass spectrum between the LSP and the other EW gauginos. These points peak low in the average multiplicity of jets, leptons, and in average E_T^{miss} . They could potentially be probed with searches that involve events with initial state radiation and soft boson decay products that are aligned with the E_T^{miss} .

Points with principal processes 3, $\tilde{\chi}_1^\pm\tilde{\chi}_2^0(\tilde{\chi}\rightarrow V/h\tilde{\chi}_1^0)$, and 10, $\tilde{\chi}_1^0\tilde{\chi}_1^0$, tend to follow the trend profiled by process 2, $\tilde{\chi}_1^\pm\tilde{\chi}^0(\tilde{\chi}_1^\pm\rightarrow W^\pm\tilde{\chi}_1^0)$, differing primarily in the lepton multiplicity and, in the case of at least one lepton, in the average p_T of the highest- p_T lepton (leading lepton). The close resemblance of processes 10 and 2 is mostly due to the fact that the mass difference between the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_1^0$ is frequently very small (less than 1 GeV), causing the ensuing off-shell W boson of process 2 to produce undetectably soft particles.

Points with principal processes 5, $\tilde{g}\tilde{g}(\tilde{g}\rightarrow qq\tilde{\chi}_1^0)$, and 6, $\tilde{g}\tilde{q}(\tilde{g}\rightarrow \tilde{q}q\tilde{\chi}_1^0)$, the most frequent modes involving gluinos, are not highlighted in Fig. 7.14, since their frequency among nonexcluded points is relatively small. Several of the nonexcluded models with very light gluino masses (less than 700 GeV) correspond to principal process 6, with mass differences between the \tilde{g} and LSP that range around 100 GeV. Sensitivity to these model points may be possible

by considering final states with three or fewer jets and E_T^{miss} thresholds that are lower than typically applied.

Points with principal process 7, $\tilde{q}\tilde{q}(\tilde{q} \rightarrow q\tilde{\chi}_1^\pm)^*$, do not display distinct trends in the properties selected, which is partly due to these points having a low dominance of around 0.1. Such model points have a diverse set of secondary processes, which are not directly examined here.

A general observation about the model points in Fig. 7.14 is the significant anticorrelation of observables, which manifests as the criss-crossing of lines between the axes. For example, model points with very high average E_T^{miss} tend to have very low cross sections, and vice versa. This is a consequence of the fact that, having observed no significant excess of events in data, the surviving model points are those with very few experimentally accessible observables, or they would have been excluded.

Signal fiducial cross sections. With over 50% of all nonexcluded points corresponding to cross sections of greater than 10 fb, it is critical to further examine why these points were not accessed in Run 1. To attempt to gain an understanding, the signal is further characterized by evaluating fiducial cross sections corresponding to a range of final-state observables. The fiducial cross section σ_f of a final-state is defined for each model point as

$$\sigma_f = \sigma_{\text{tot}}^{8 \text{ TeV}} A, \quad (7.16)$$

where A is the acceptance times signal efficiency computed as the fraction of simulated signal events passing a set of event-level criteria. A set of final-state observables are examined that loosely correspond to trigger thresholds or signal regions of the examined searches. Figures 7.15-7.19 show the impact of adjusting various thresholds on the fiducial cross sections of nonexcluded points.

Some principal processes can be associated with large fiducial cross sections, depending on the final state considered. For example, points with mostly first-generation squark production give rise to large fiducial cross sections for events with high H_T , resulting in Fig. 7.16 showing mostly orange-colored points; and points with production involving EW gauginos give rise to substantial fiducial cross sections for events with a high multiplicity of

soft leptons, which explains the unaccompanied blue and green lines in Fig. 7.18. Somewhat striking is the behavior of the E_T^{miss} fiducial cross section (Fig. 7.15), which can increase rapidly (by up to a factor of ten) as the threshold is relaxed from 200 to 100 GeV. It is apparent that many of the nonexcluded regions are not accessible with thresholds of 200 GeV, a common criterion applied offline to achieve full efficiency with the triggers. The fiducial cross section decreases noticeably as the threshold is further increased from 200 to 300 GeV. Similar behavior is seen for the H_T fiducial cross section (Fig. 7.16). Fiducial cross sections are quite large for these final states when a threshold of 300 GeV is applied, but fall off substantially for higher thresholds.

Of course, a loosening of the object thresholds would increase the background yield as well as signal yield. Therefore, a thorough survey of analysis techniques and specific backgrounds will be necessary to select optimal values for kinematic thresholds and other analysis techniques to probe the most difficult points. However, the lesson that nonexcluded pMSSM models have large cross sections in background-rich kinematic regions is an open invitation for the development of new techniques that improve signal to background discrimination and background modeling. If the MSSM is realized in one of these difficult regions, the hunt for SUSY will force us to either abandon the LHC, or become more creative.

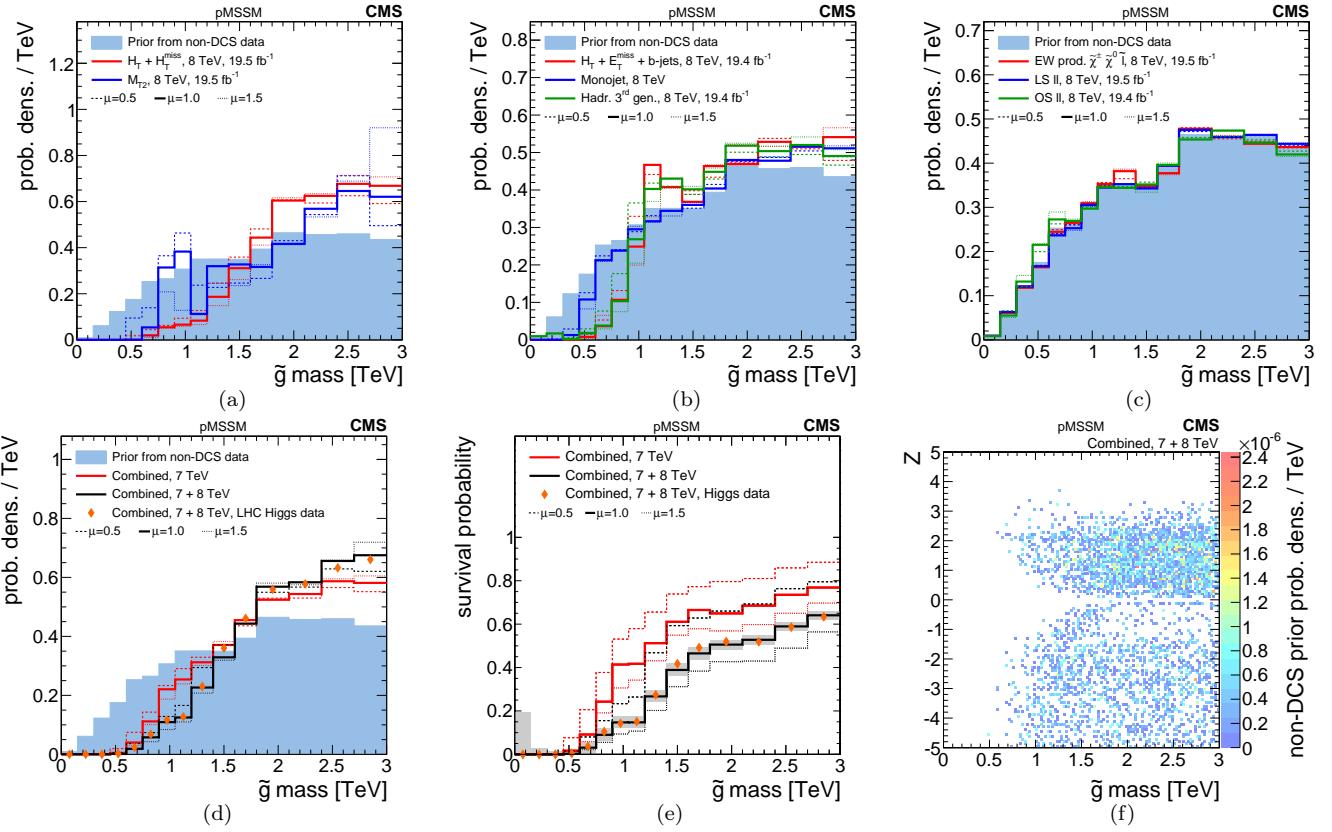


Figure 7.2: A summary of the impact of CMS searches on our knowledge of the gluino mass in the pMSSM parameter space. Plots (a)-(d) compare the non-DCS prior distribution of the gluino mass (blue filled histograms) to posterior distributions after data from various CMS searches (line histograms), where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the gluino mass for various combinations of CMS data and data from Higgs boson measurements, where the shaded grey band gives the the statistical uncertainty on the black histogram. Plot (f) shows the distribution of the gluino mass versus the Z-significance calculated from the combination of all searches.

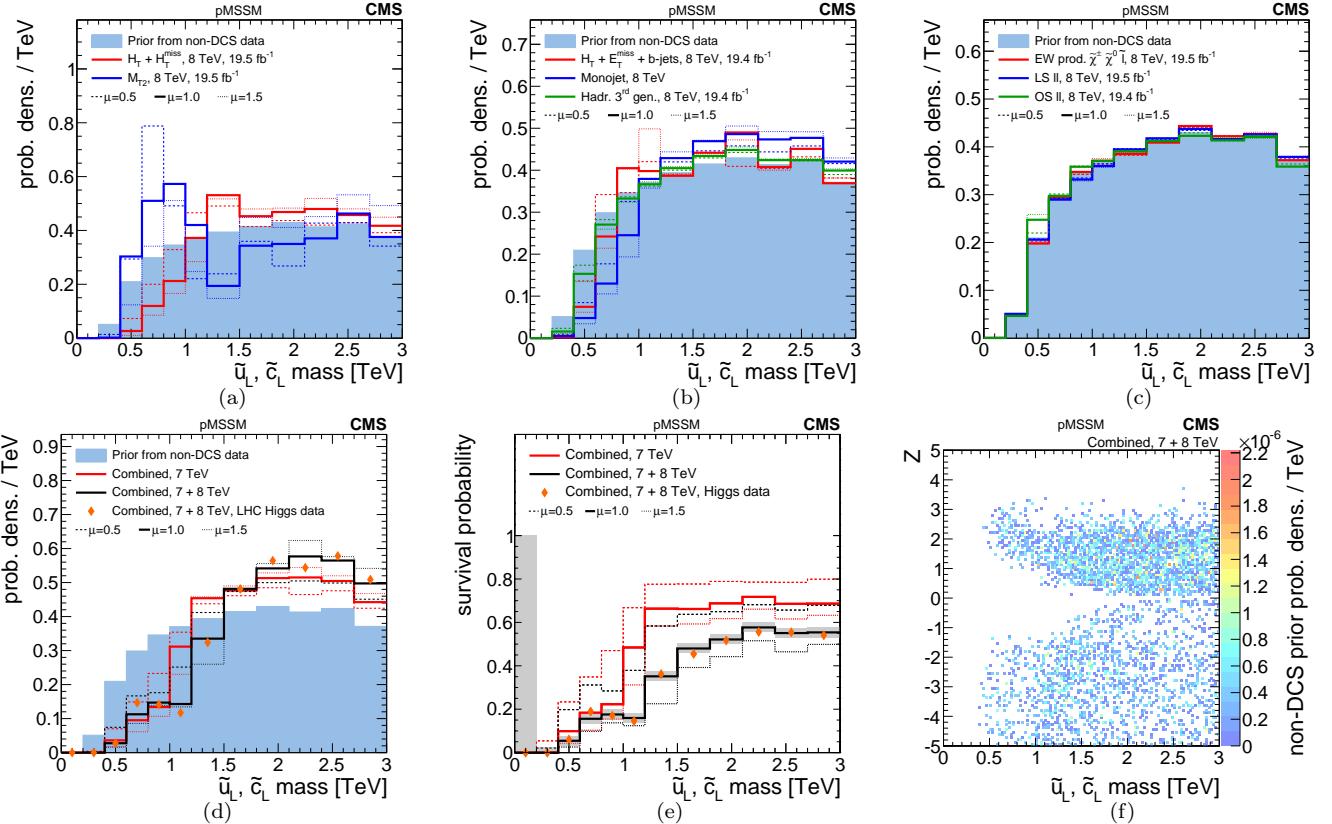


Figure 7.3: A summary of the impact of CMS searches on our knowledge of the \tilde{u}_L mass (equivalently, the \tilde{c}_L mass) in the pMSSM parameter space. Plots (a)-(d) compare the non-DCS prior distribution of the \tilde{u}_L mass to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the \tilde{u}_L mass for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the \tilde{u}_L mass versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading.

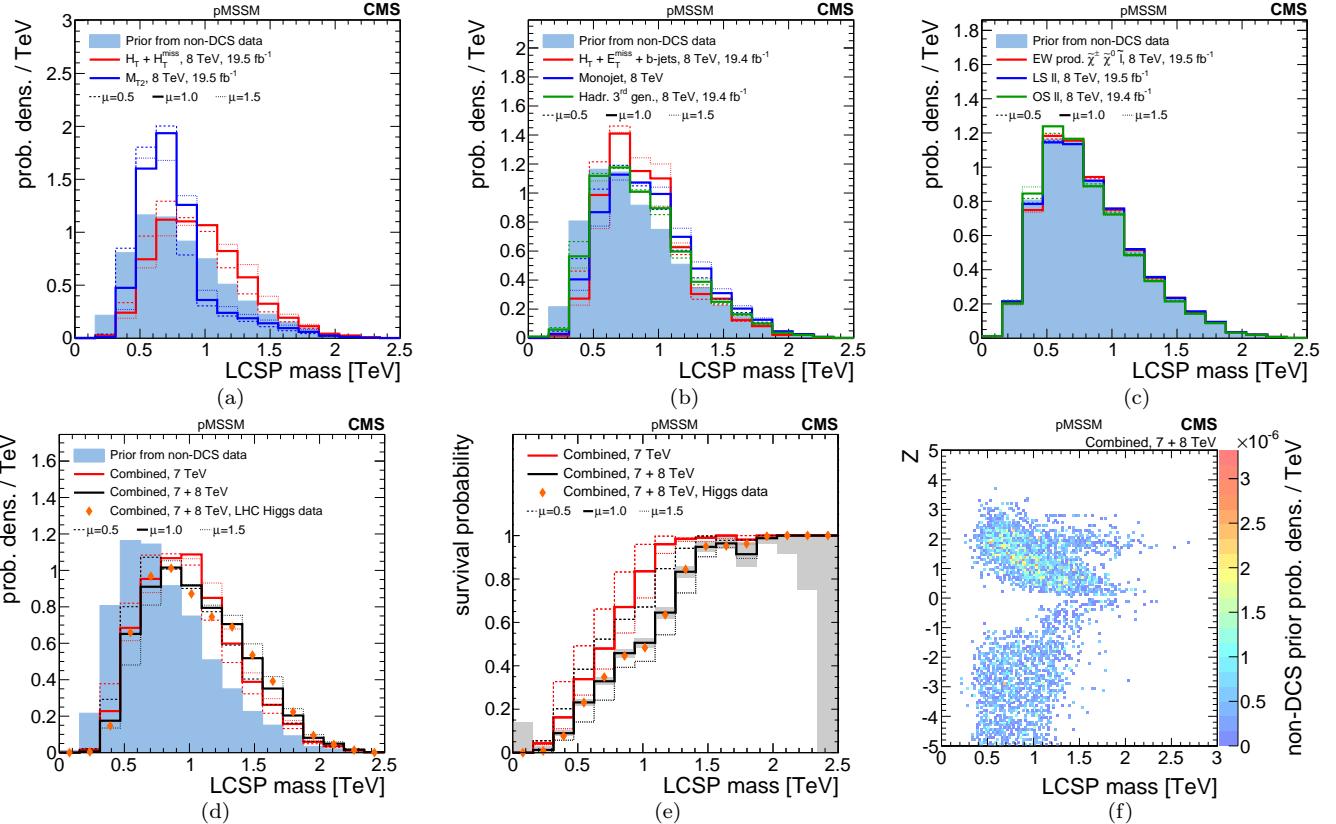


Figure 7.4: A summary of the impact of CMS searches on our knowledge of the mass of the lightest colored SUSY particle (LCSP) in the pMSSM parameter space. Plots (a)-(d) compare the non-DCSprior distribution of the LCSP mass to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the LCSP mass for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the LCSP mass versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading.

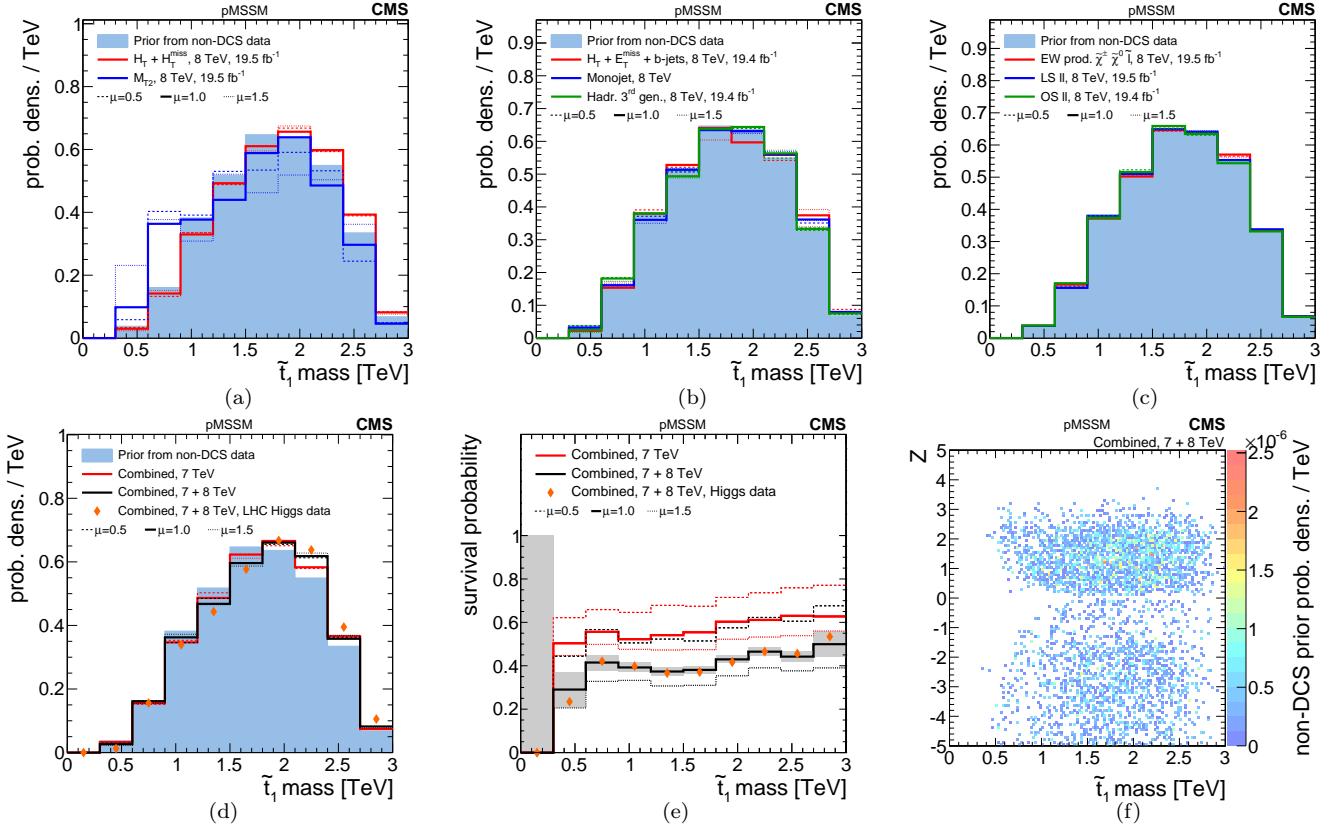


Figure 7.5: A summary of the impact of CMS searches on our knowledge of the \tilde{t}_1 mass in the pMSSM parameter space. Plots (a)-(d) compare the non-DCS prior distribution of the \tilde{t}_1 mass to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the \tilde{t}_1 mass for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the \tilde{t}_1 mass versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading.

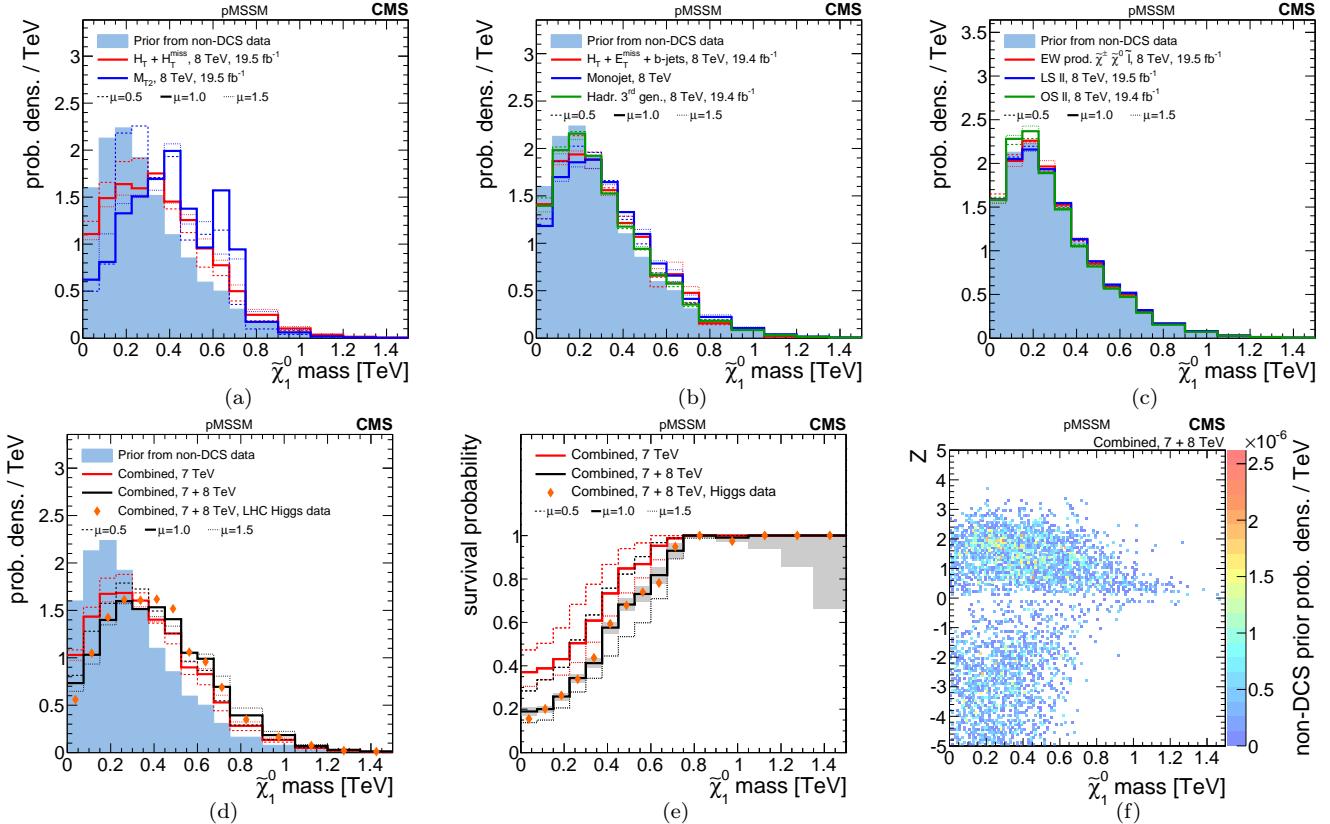


Figure 7.6: A summary of the impact of CMS searches on our knowledge of the $\tilde{\chi}_1^0$ mass in the pMSSM parameter space. Plots (a)-(d) compare the non-DCS prior distribution of the $\tilde{\chi}_1^0$ mass to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the $\tilde{\chi}_1^0$ mass for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the $\tilde{\chi}_1^0$ mass versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading.

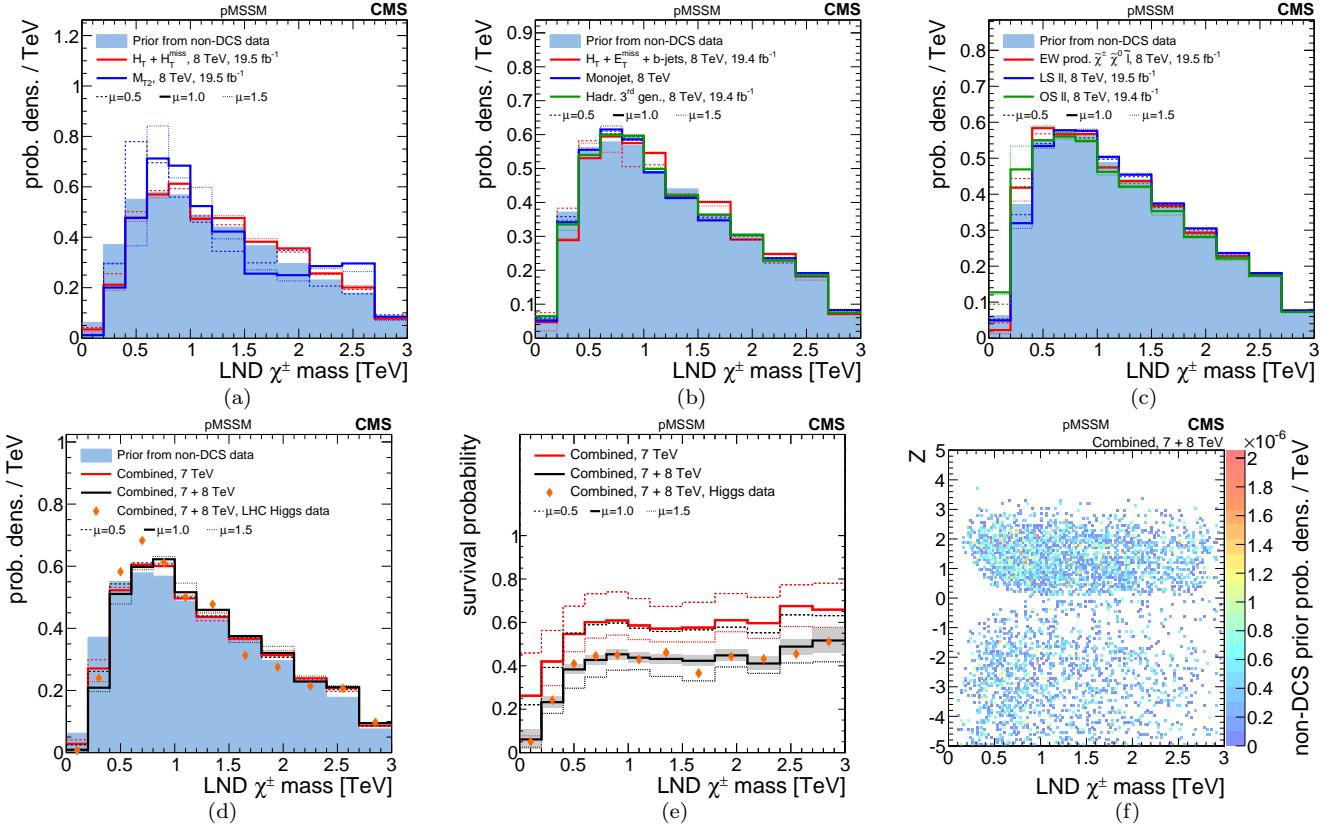


Figure 7.7: A summary of the impact of CMS searches on our knowledge of the the mass of the lightest non-degenerate (LND) chargino in the pMSSM parameter space. Plots (a)-(d) compare the non-DCSprior distribution of the LND $\tilde{\chi}^\pm$ mass to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the LND $\tilde{\chi}^\pm$ mass for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the LND $\tilde{\chi}^\pm$ mass versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading.

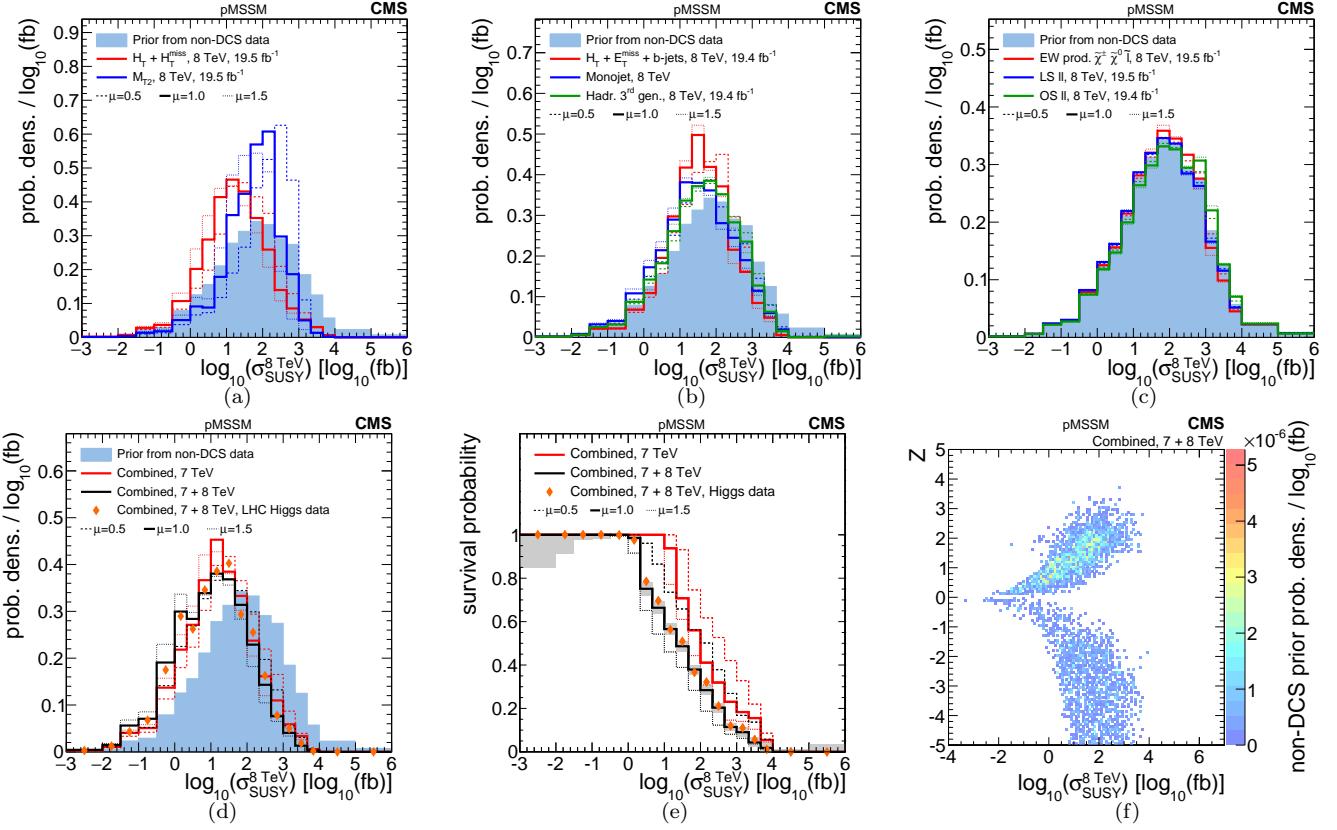


Figure 7.8: A summary of the impact of CMS searches on our knowledge of the the logarithm of the cross section for inclusive sparticle production in 8 TeV pp collisions, $\log_{10}(\sigma_{\text{SUSY}}^{\text{8 TeV}})$, in the pMSSM parameter space. Plots (a)-(d) compare the non-DCSprior distribution of the $\log_{10}(\sigma_{\text{SUSY}}^{\text{8 TeV}})$ to posterior distributions after data from various CMS searches, where (d) shows the combined effect of CMS searches and the Higgs boson results. Plot (e) shows survival probabilities as a function of the $\log_{10}(\sigma_{\text{SUSY}}^{\text{8 TeV}})$ for various combinations of CMS data and data from Higgs boson measurements. Plot (f) shows the distribution of the $\log_{10}(\sigma_{\text{SUSY}}^{\text{8 TeV}})$ versus the Z -significance calculated from the combination of all searches. See Fig. 7.2 for a description of the shading. In plot (d), the apparent enhancement of the left tail of the posterior density with respect to the prior is due to the suppression of the right tail and an overall renormalization.

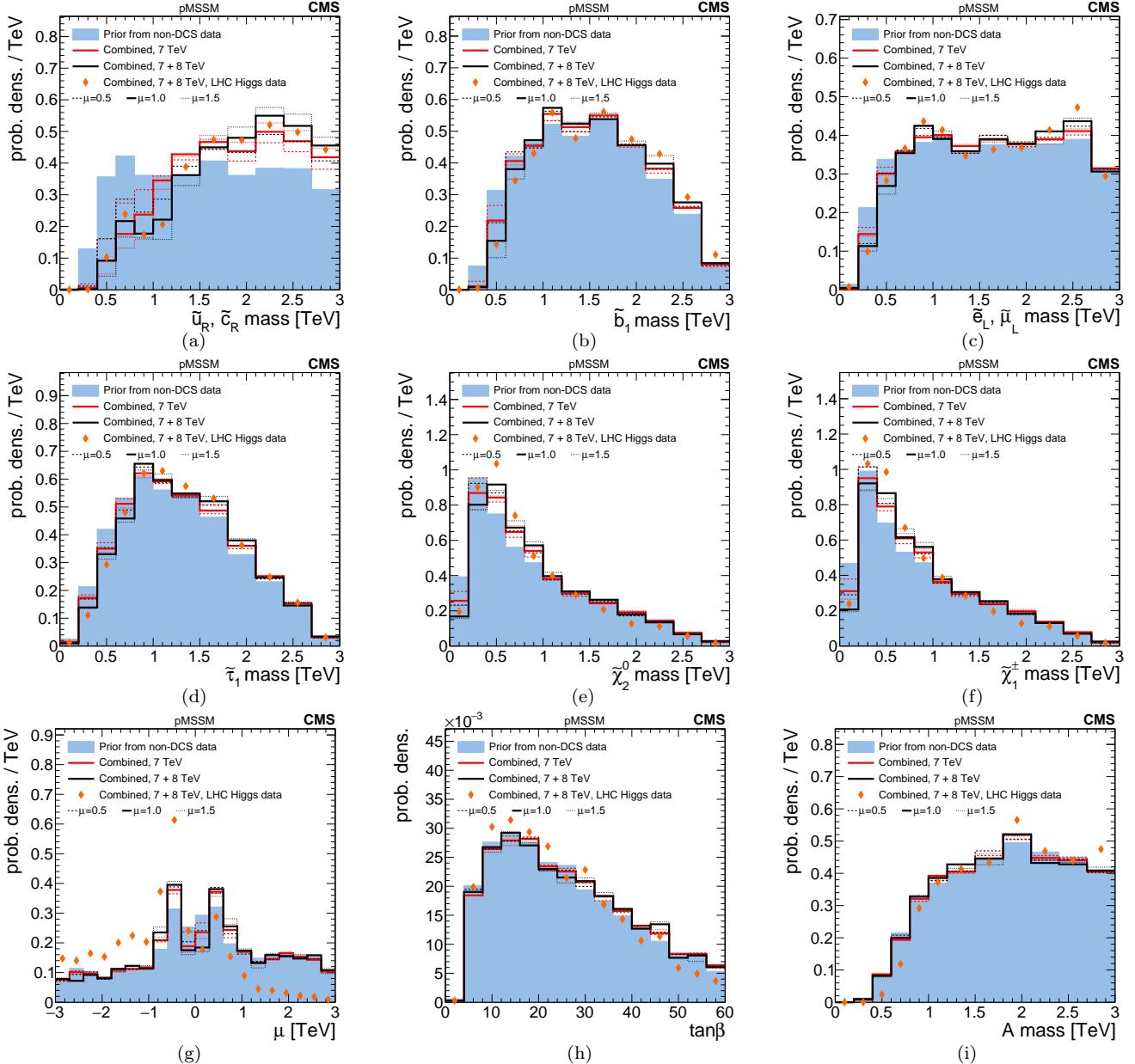


Figure 7.9: Comparison of prior and posterior distributions after several combinations of data from the CMS searches for the \tilde{u}_R , \tilde{c}_R mass, \tilde{b}_1 mass, \tilde{e}_L , $\tilde{\mu}_L$ mass, $\tilde{\tau}_1$ mass, $\tilde{\chi}_2^0$ mass, $\tilde{\chi}_1^\pm$ mass, the higgsino mass parameter μ , $\tan\beta$, and A mass.

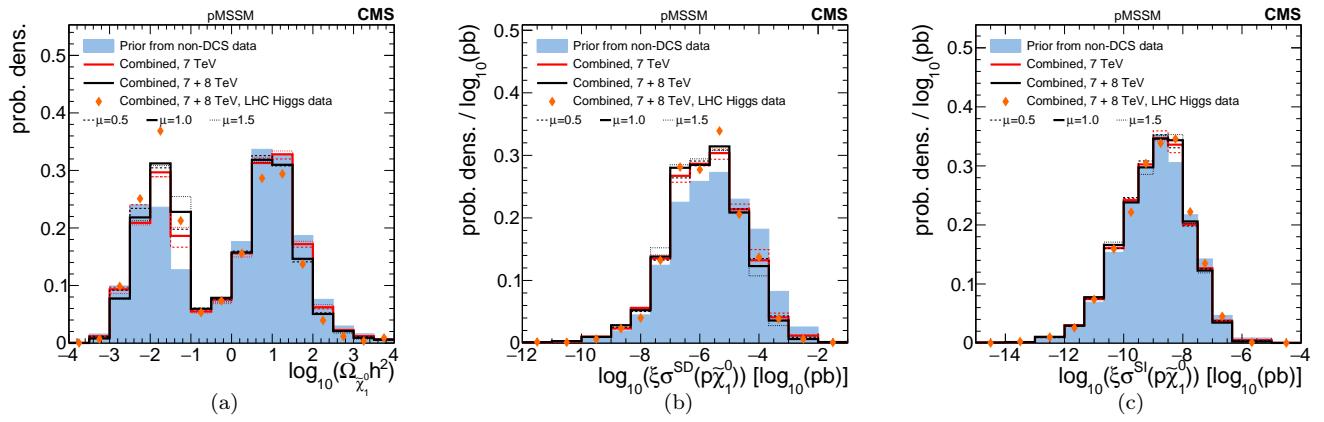


Figure 7.10: Comparison of prior and posterior distributions after several combinations of data from the CMS searches for $\Omega_{\tilde{\chi}_1^0}$, $\xi\sigma^{\text{SD}}(p\tilde{\chi}_1^0)$, and $\xi\sigma^{\text{SI}}(p\tilde{\chi}_1^0)$.

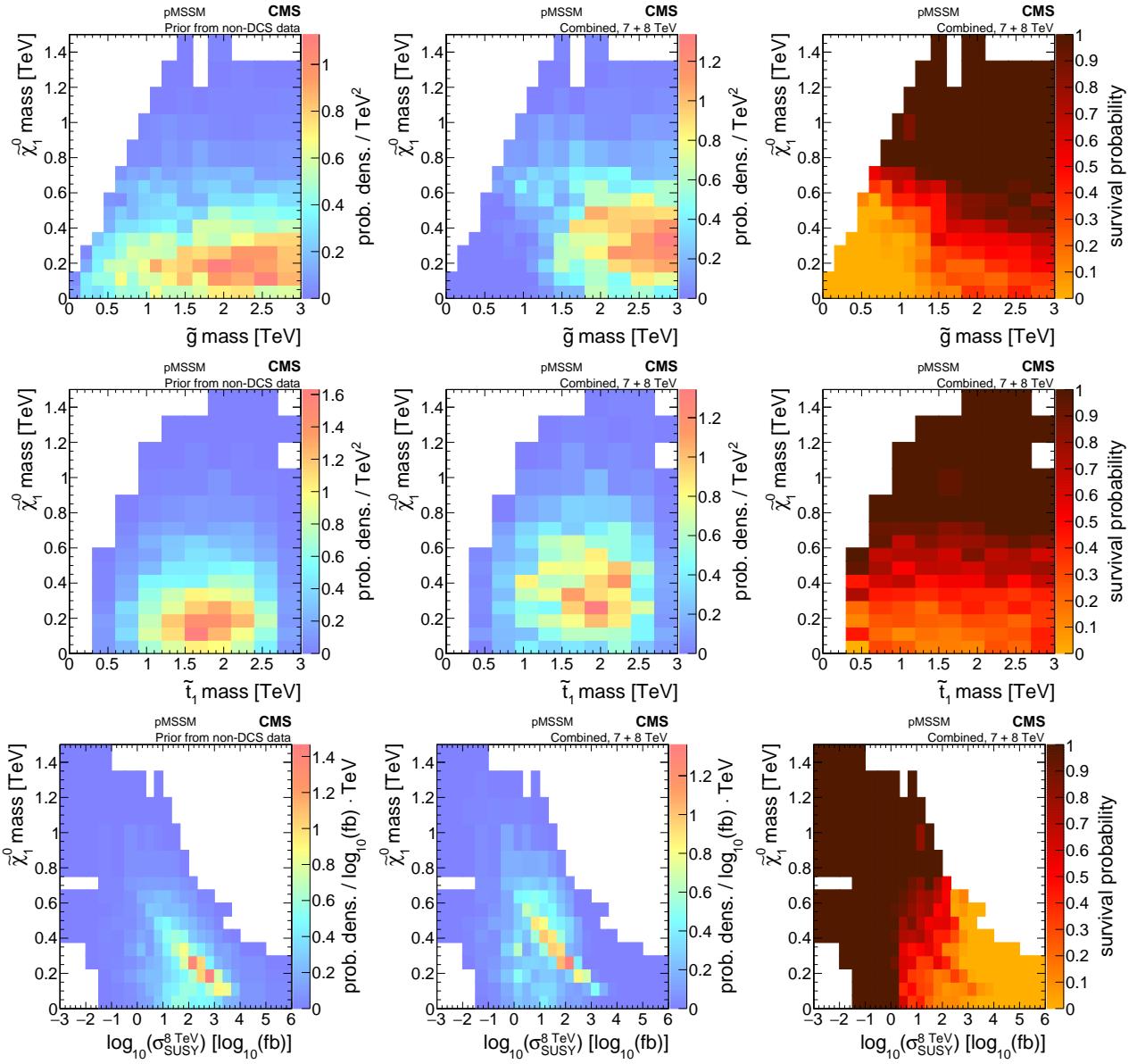


Figure 7.11: Marginalized non-DCS distributions (first column), compared with posterior distributions (second column) and survival probabilities (third column) after inclusion of the combined CMS searches, are shown for the $\tilde{\chi}_1^0$ mass versus gluino mass (first row), the top squark mass (second row), and the logarithm of the cross section for inclusive sparticle production at 8 TeV (third row).

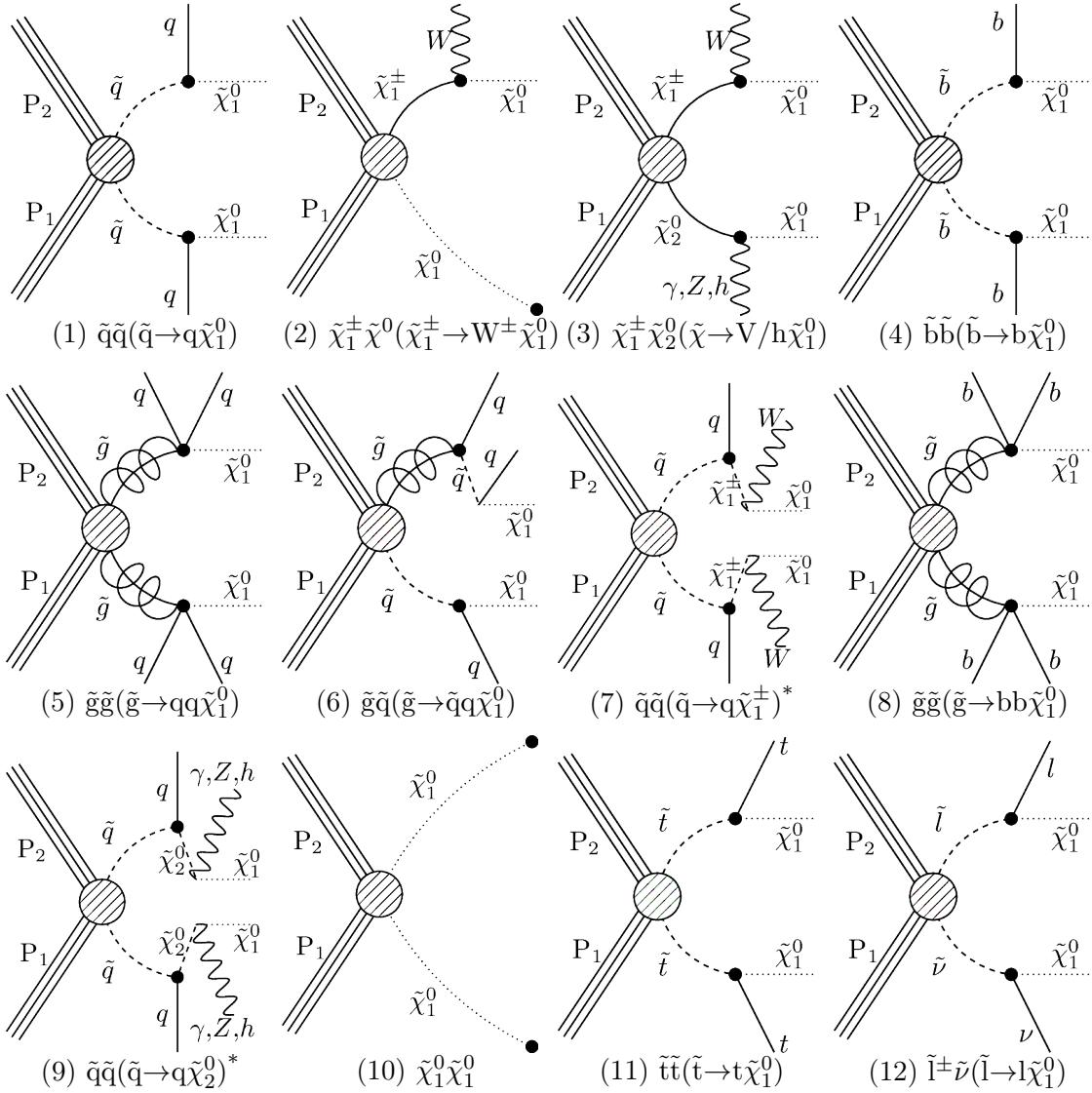


Figure 7.12: The twelve most common principal processes in the pMSSM, listed in order of their frequency before the constraints of the CMS searches. Both on-shell and off-shell states are included. Indices of particle charge, flavor, and chirality are ignored in the construction, with the exception of the flavor of the third-generation squarks and quarks. Asterisks in the labels indicate where process names involving long decay chains have been abbreviated.

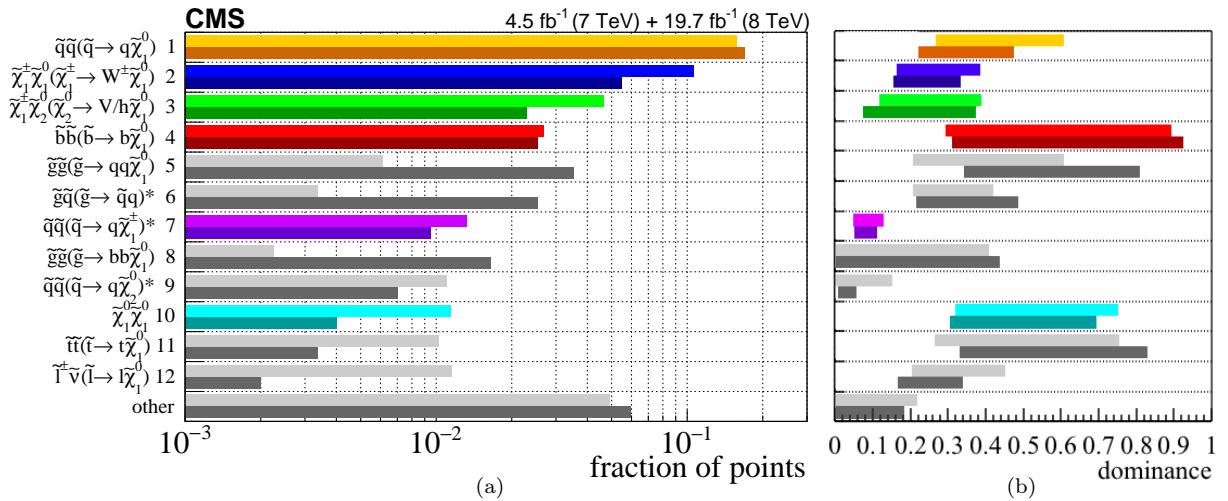


Figure 7.13: (a) shows the fraction of excluded (dark) and nonexcluded (light) points out of all considered points, by principal process. Color is assigned to the processes that are most common after the constraints of the CMS searches, which are selected for further study. The dominance of principal processes, as defined in Eq. 7.15, is given in (b) where the bands show the RMS range of the dominance.

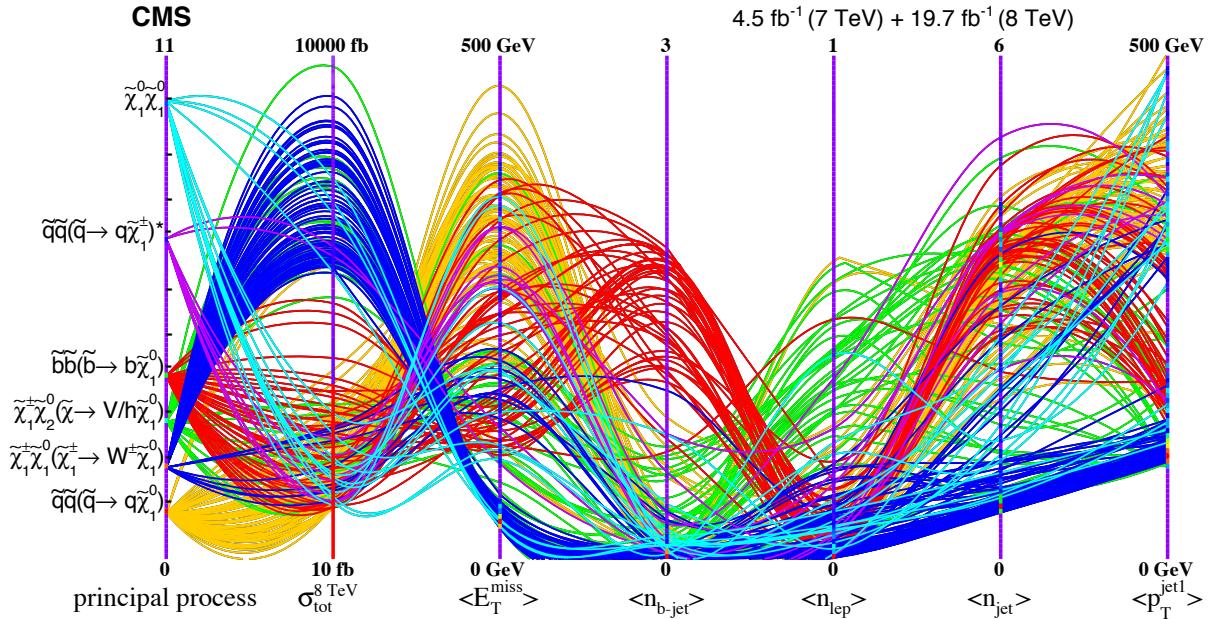


Figure 7.14: A parallel coordinates plot showing several hundred selected nonexcluded model points for the six most common principal processes, with seven key properties. From the left, the selected properties are: the principal process, the 8 TeV signal production cross section (in \log_{10} scale), the average value of the E_T^{miss} , the average number of b-jets, leptons, and jets, and finally, the average p_T momentum of the leading jet. Color is assigned based on the principal process. Orange codes for process 1, blue for process 2, green for 3, red for 4, violet for 7, and cyan for 10. Lines arching toward higher vertical positions typically indicate more “discoverable” scenarios.

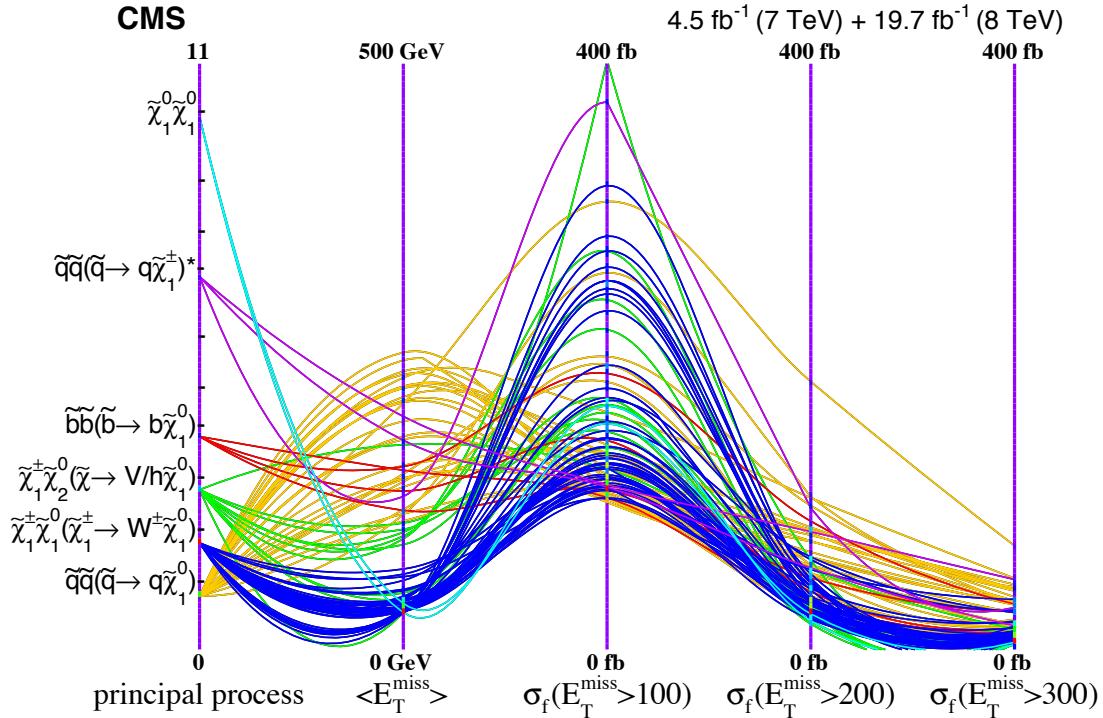


Figure 7.15: A parallel coordinates plot of the nonexcluded pMSSM points with the axes set as the principal process, the average E_T^{miss} (in GeV), and the fiducial cross section (in linear scale) for various thresholds on the E_T^{miss} . All nonexcluded points corresponding to processes 1, 2, 3, 4, 7, and 10 that have a fiducial cross section greater than 100 fb are shown. Color is assigned to values of the principal process in the same manner as in Fig. 7.14.

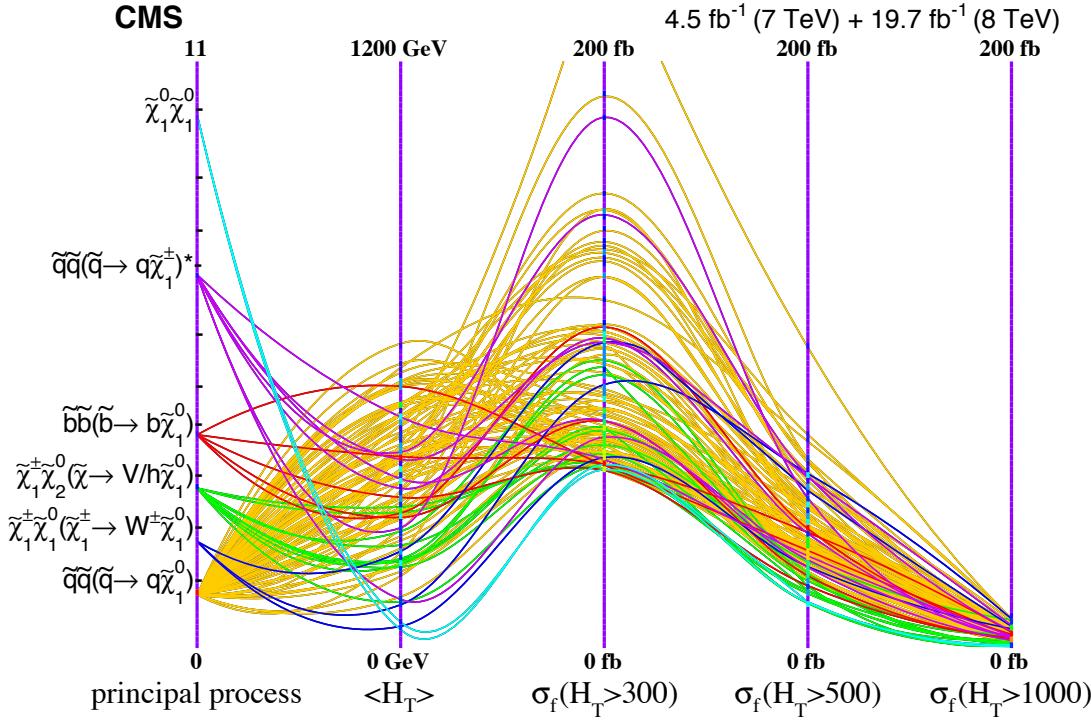


Figure 7.16: A parallel coordinates plot of the nonexcluded pMSSM points with the axes set as the principal process, the average H_T (in GeV), and the fiducial cross section (in linear scale) for various thresholds on the H_T . All nonexcluded points corresponding to processes 1, 2, 3, 4, 7, and 10 that have a fiducial cross section greater than 60 fb are shown. Color is assigned to values of the principal process in the same manner as in Fig. 7.14.

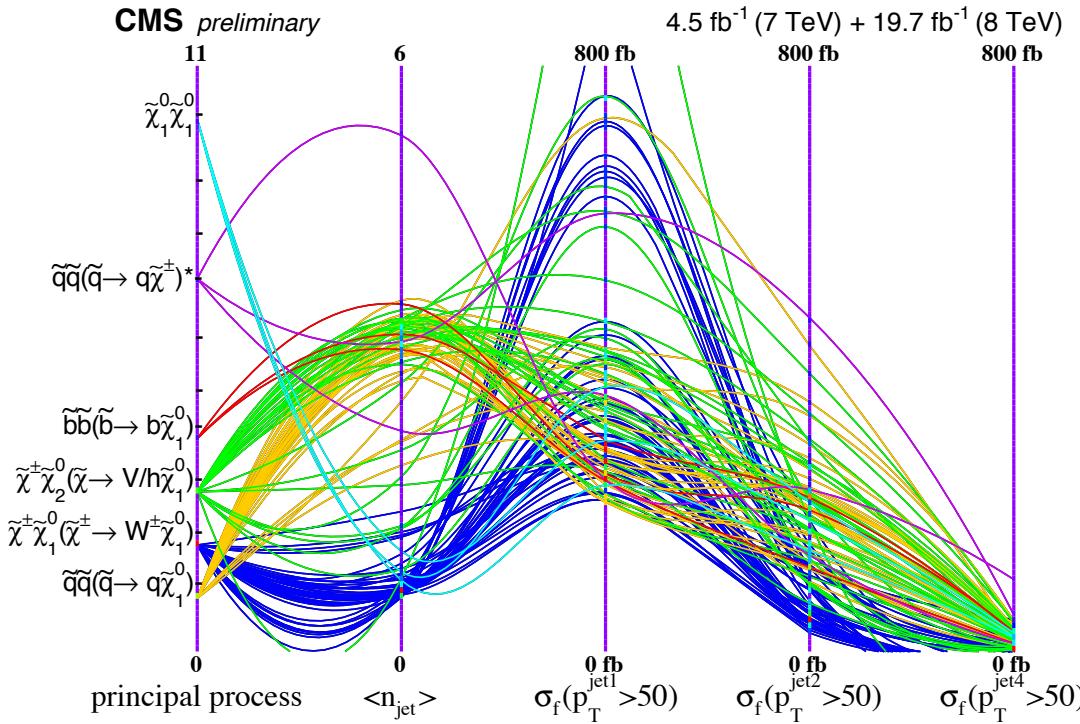


Figure 7.17: A parallel coordinates plot of the nonexcluded pMSSM points with the axes set as the principal process and the fiducial cross section (in linear scale) for various thresholds on the multiplicity of jets. All nonexcluded points corresponding to processes 1, 2, 3, 4, 7, and 10 that have a fiducial cross section greater than 200 fb are shown. Color is assigned to values of the principal process in the same manner as in Fig. 7.14.

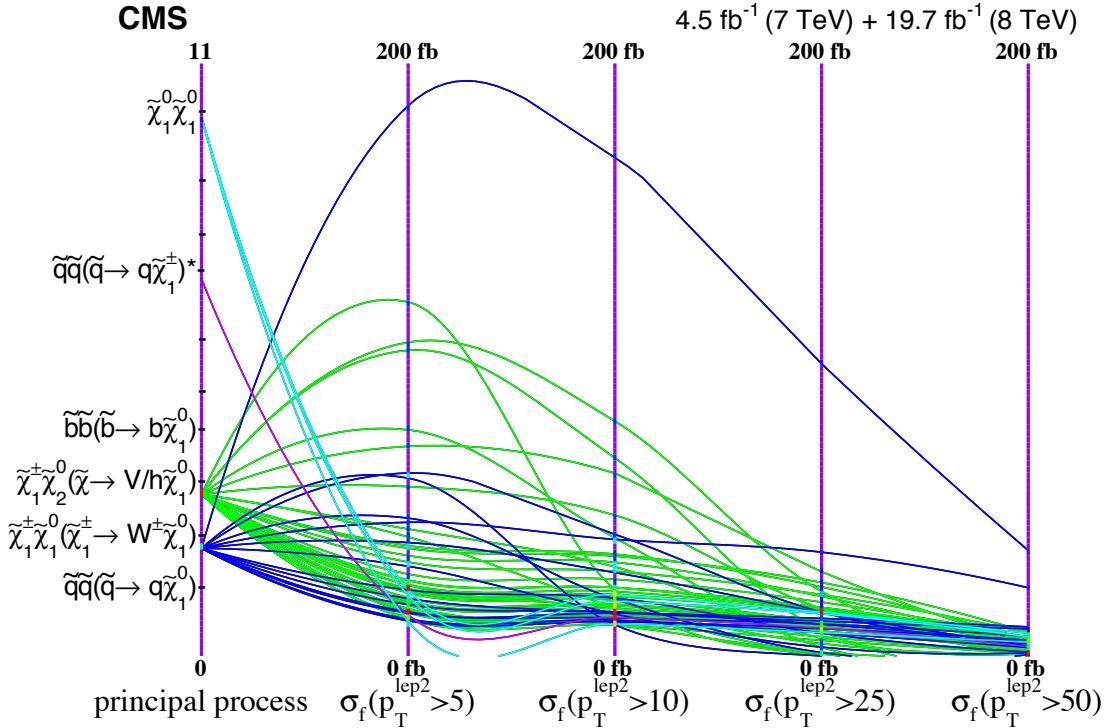


Figure 7.18: A parallel coordinates plot of the nonexcluded pMSSM points with the axes set as the principal process and the fiducial cross section (in linear scale) for various thresholds on the sub-leading lepton p_T (in GeV). All nonexcluded points corresponding to processes 1, 2, 3, 4, 7, and 10 that have a fiducial cross section greater than 30 fb are shown. Color is assigned to values of the principal process in the same manner as in Fig. 7.14.

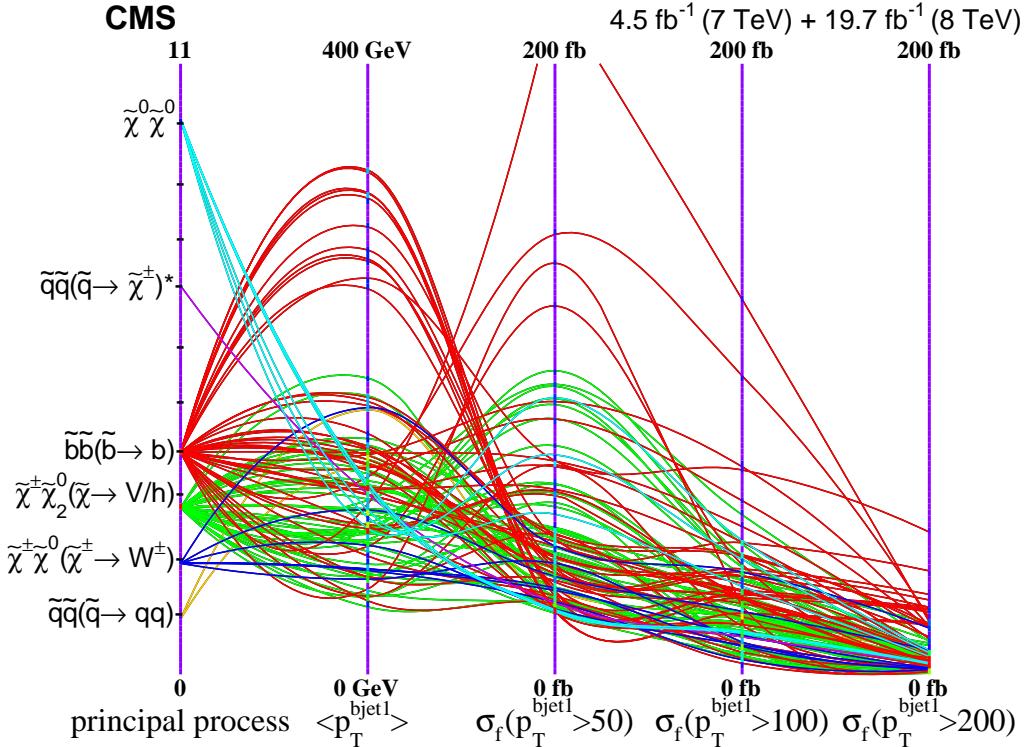


Figure 7.19: A parallel coordinates plot of the nonexcluded pMSSM points with the axes set as the principal process and the fiducial cross section (in linear scale) for various thresholds on the leading b-tagged jet p_T (in GeV). All nonexcluded points corresponding to processes 1, 2, 3, 4, 7, and 10 that have a fiducial cross section greater than 20 fb are shown. Color is assigned to values of the principal process in the same manner as in Fig. 7.14.

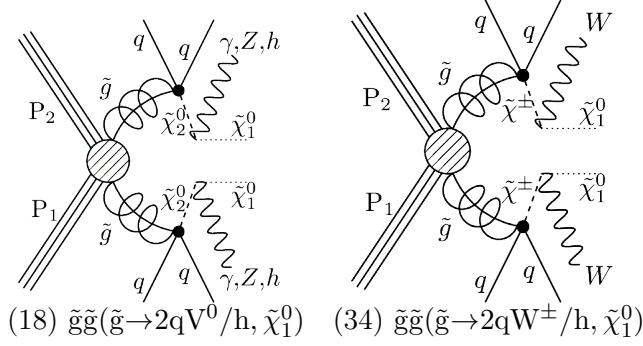


Figure 7.20: The 18th and 34th most frequent principal process in the pMSSM.

A note on low-mass gluinos. Model points with gluino masses as low as 600 GeV survive CMS analyses, but these scenarios do not correspond to the top 12 processes listed in Fig. 7.12. Such model points are characterized primarily by the processes shown in Fig. 7.20. The intermediate EWkinos in the gluino decay chains serve to lower the p_T of jets, and therefore H_T , in signal events. These scenarios tend to have mass splittings between the gluino and LSP of less than a few hundred GeV, which leaves little kinetic energy available for the intermediate decay bosons, which are typically off shell. These features lead to a lowered signal efficiency in searches that require high jet multiplicity and large H_T . These scenarios may be probed by searches with relatively low or moderate thresholds on the H_T and H_T^{miss} , and by requiring initial state radiation to boost the SUSY particles into one hemisphere. Then, looking for soft particles, which may potentially originate from the gluino decays, that travel in the direction of the E_T^{miss} may serve to separate signal events from the number of QCD and $Z \rightarrow \nu\bar{\nu}$ background events. In the case of Fig. 7.20 (b), requiring an identified track that disappears in the tracker, a possible signature of the long-lived chargino, may greatly increase the sensitivity of searches in the case that the LSP and $\tilde{\chi}^\pm$ are sufficiently degenerate in mass.

CHAPTER 8

CMS SEARCHES FOR SUPERSYMMETRY AT $\sqrt{S} = 13$ TEV

One of the findings of Chapter 7, and seen in Fig. 7.1, is that searches for SUSY events with purely hadronic final states have a particularly significant impact on our knowledge of the MSSM. Furthermore, it was found that SUSY models characterized by large fiducial cross sections corresponding to low thresholds on the E_T^{miss} , H_T , and N_{jets} survived 7 and 8 TeV CMS searches.

After making these findings, I contributed to CMS hadronic analyses [6] and [7], where I developed methods that increase the sensitivity of the searches to SUSY models that predict large numbers of events produced in the low- E_T^{miss} and low- H_T regions. The primary backgrounds in these regions are multi-jet events arising from QCD interactions, and events arising from the electroweak production of a Z boson that decays into neutrinos, $Z \rightarrow \nu\bar{\nu}$. I developed a data-driven QCD event simulator, based on the rebalance and smear method, that allows the prediction of QCD background counts to be made with uncertainties that are lower than those of alternative methods by up to an order of magnitude. I developed and performed a data-driven $Z \rightarrow \nu\bar{\nu}$ background estimation method that uses events in which a Z boson is produced and decays into two electrons or two muons as a proxy for $Z \rightarrow \nu\bar{\nu}$ background events. I also built upon a technique that maximizes the usefulness of triggers in kinematic regions in which the trigger is not fully turned on.

The above techniques are discussed generally in the following sections, that is, they are discussed as stand-alone methods, independent of a particular analysis, with only occasional reference made to the analyses in which the methods were applied. The aim of this approach is to accentuate the universality of the techniques. Moreover, special effort was paid to ensure that the methods remain valid in kinematics regions with lower E_T^{miss} and H_T than the baseline regions of the hadronic analyses, and the methods are used in the next chapter

in a new context. Section 8.4 discusses how these methods have been applied specifically in the context of two high- E_T^{miss} hadronic searches: the multi-jet + H_T^{miss} search [6], and the search for SUSY with top squarks [7]. Section 8.4 gives a brief introduction to the analyses [6] and [7], a detailing of the baseline selection for each, a discussion of the background estimates and uncertainties obtained using the above methods, and the final results of the searches, namely, limits on sparticle masses set in the context of simplified models. Then, in the chapter that follows, the use of multivariate techniques in background-dominated kinematic regions like the low- E_T^{miss} region and the low- H_T region is explored as a way to probe nonexcluded pMSSM points.

8.1 Trigger methods

Searches for SUSY at the LHC typically suppress the overwhelming QCD background by requiring events to have a large E_T^{miss} as a criterion for firing the trigger. Additionally, for searches in the hadronic channel, a minimum amount of H_T can also be required, or some combination of E_T^{miss} and H_T . Hadronic searches typically collect data using triggers that accept events (fire) if these observables exceed some threshold. However, because online calculations must be performed rapidly, only rough approximations of the H_T and E_T^{miss} can be obtained in real time. Therefore, substantial discrepancies can arise between the online objects and the more precise offline objects, since the latter benefit from the availability of greater computing resources. In order to guarantee that a trigger fire on an event with a desired offline E_T^{miss} , the online E_T^{miss} threshold must be sufficiently lower than the offline E_T^{miss} threshold. This introduces the concept of a trigger turn-on curve, which is the nickname for the trigger efficiency, that is, the probability of the trigger firing, as a function of an offline variable. E_T^{miss} trigger turn-on curves have notoriously gradual slopes, because the calculation of the E_T^{miss} is particularly sensitive to jet energy resolutions, which differ substantially in the offline and online event reconstruction.

Two conditions must be met in order for a trigger to be useful. First, the trigger must have a sufficiently high efficiency for selecting signal events. And second, it must be possible to make an unbiased measurement of the trigger efficiency. The former is typically a matter

of choosing an appropriate trigger. For the latter, it is best to measure the trigger efficiency using real data, but in order to do so without bias, one needs a sample of events, referred to as a reference sample, that is independent of the events that pass the trigger.

8.1.1 Measuring the efficiency without bias

Suppose the goal is to make a measurement of the efficiency of trigger A. One can make an unbiased measurement by obtaining a sample of events collected by trigger B, so long as the absolute probability of firing trigger A, $P(A)$, is equal to the conditional probability of trigger A firing given that trigger B has also fired, $P(A|B)$:

$$P(A) = P(A|B). \quad (8.1)$$

In other words, trigger B should be chosen such that its triggering criteria are independent of those of trigger A. Whether or not Equation 8.1 holds can be by simulating the triggers. Finally, the unbiased estimate of the efficiency of trigger A is the fraction of events that fire trigger B that also fire trigger A:

$$\hat{\epsilon}_A = \frac{N_{\text{fire A and B}}}{N_{\text{fire B}}}. \quad (8.2)$$

This is an estimate of $P(A|B)$ with an expectation value equal to $P(A)$ if, and only if, $P(AB) = P(A)P(B)$, that is triggers A and B are independent.

One might choose to collect data using the trigger named

- `HLT_PFMETNoMu90_PFMHTNoMu90`,

referred to as the monojet trigger. Here, PFMET and PFMHT refer to the online E_T^{miss} and H_T^{miss} computed using the particle flow (PF) algorithm (Section 6.8), NoMu refers to the fact that muons are neglected in the online computation of these observables, and the number 90 refers to the triggering thresholds in units of GeV. The trigger efficiency must be estimated as a function of the offline H_T and H_T^{miss} . In this example, the monojet trigger is treated as trigger A above.

Trigger B, or the reference trigger, is chosen to be the trigger named

- `HLT_Ele27_eta2p1_WP Loose_Gsf_v*`,

which selects events that contain an electron reconstructed by the GSF algorithm (Section 6.8) with a p_T greater than 27 GeV and pseudorapidity less than than 2.1, and take the base sample to be the subset of these events that have $N_{\text{jets}} > 3$ and a single offline reconstructed electron.

A sample of simulated $t\bar{t}$ events is used to establish the validity of Equation 8.1. First, the conditional probability of firing the primary trigger, given that the reference trigger fired, is computed as a function of the offline H_T and H_T^{miss} . Then, the absolute probability of firing the primary trigger is computed by replacing the reference sample with the entire sample of simulated events, as a function of the same variables. The ratio of the conditional probability (“method”) to the absolute probability (“truth”) is shown in Fig. 8.26.

The primary and reference triggers have been shown to satisfy Equation 8.1, so there is now license to measure the trigger efficiency in real data. Figure 8.2 shows the efficiency, computed using Equation 8.2. The efficiency is nearly 1 for kinematic regions with $H_T^{\text{miss}} > 200$ GeV, making the monojet trigger an ideal choice for a hadronic searches for physics in the low- H_T region.

If the condition in 8.1 is found not to hold for a particular reference trigger, the efficiency can still be measured, but a systematic uncertainty must be applied to the estimate of the efficiency to account for the bias. These systematic uncertainties can typically be reduced or eliminated by more intelligently choosing a set of observables on which the efficiency estimation depends, in this case the H_T and H_T^{miss} . This is discussed below under the subsection heading “Multivariate trigger techniques,” in the context of more advanced techniques for determining the trigger efficiency.

8.1.2 Multivariate classification

It is possible for the trigger efficiency to vary as a complicated function of the observables in events. Often, a trigger decision involves a long chain of HLT (Chapter 6 Section 6.8) modules, or filters, that can potentially induce unforeseen dependencies of the probability of firing the trigger on non-trivial correlations among the observables. One way to account

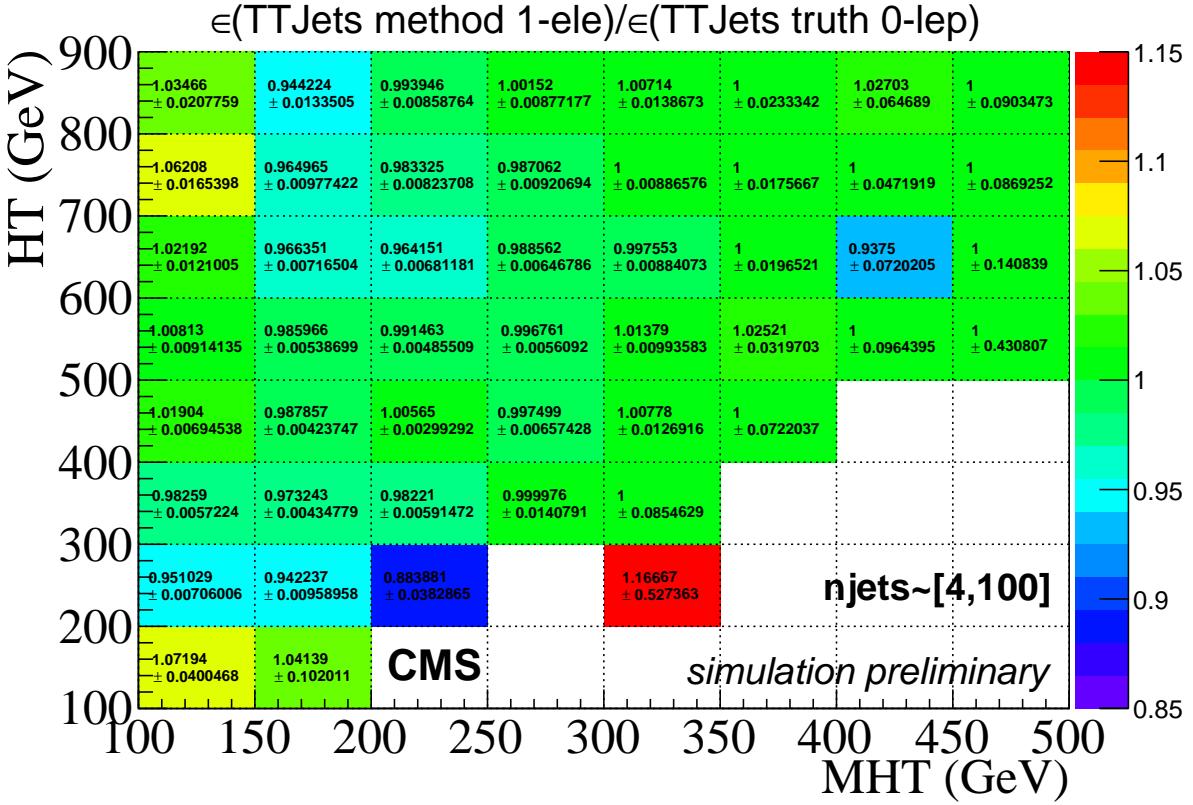


Figure 8.1: The ratio of the monojet trigger efficiency for events passing the single electron reference trigger to the trigger efficiency for all events in a simulated $t\bar{t}$ sample, as a function of the offline H_T and H_T^{miss} . The ratio is consistent with 1 throughout the $H_T - H_T^{\text{miss}}$ plane.

for these correlations is to increase the dimensionality of the parametrization of the trigger efficiency estimation. For example, the jet multiplicity could be added as a third parameter to accompany the previously described H_T and H_T^{miss} parameterization if it were believed that N_{jets} were a determinant of the efficiency. However, adding additional dimensions requires the division of the event sample into a larger number of bins, which reduces the counts in each bin. This approach is ultimately plagued by the curse of dimensionality.

It is possible to reduce the dimensionality of the parametrization to just 1 dimension by constructing an appropriate single-valued function of the relevant observables. Such a function can be obtained through the use of a Bayesian neural network (NN) classifier, as previously demonstrated in the context of b-tagging triggers [103].

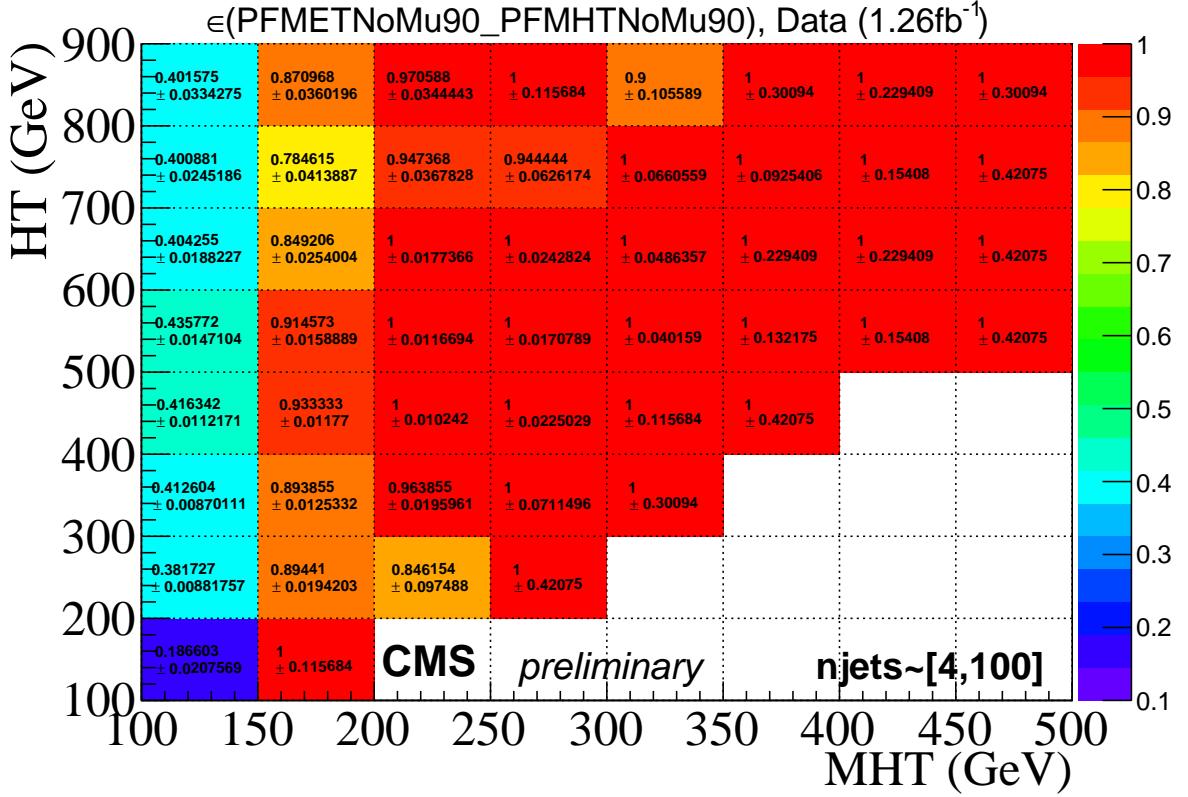


Figure 8.2: The monojet trigger efficiency, measured in real data, as a function of the offline H_T and H_T^{miss} .

As stated in Equation A.3, the output of such a classifier can be interpreted as

$$\text{NN} = P(s|\vec{\theta}) = \frac{p(\vec{\theta}|s)}{p(\vec{\theta}|s) + p(\vec{\theta}|b)}, \quad (8.3)$$

where $\vec{\theta}$ is a vector of observed quantities, also called the data, and s and b are, respectively, the true and false outcomes—in this case, the decision made by the trigger to fire or not. Since a NN is capable of training with a vector θ of arbitrary length, it is worth including any observable that may be relevant for the determination of the efficiency. To train the NN for the monojet trigger efficiency, the single electron reference sample is divided into events that fire the monojet trigger, which serve as the signal training events, and events that fail the monojet trigger, which serve as the background training events. There are several internal NN parameters, which govern aspects of the output such as the responsiveness of

the output to small fluctuations in the data. These parameters are assigned gaussian prior probability densities that are broad distributions relative to the likelihood functions, which are highly corrugated, a combination that results in posterior densities for the NN parameters being dominated by the likelihoods. To obtain an uncertainty in the efficiency estimate, the posterior densities of the NN parameters are randomly sampled, and an ensemble of efficiency estimates are obtained. The mode of the envelope of this ensemble defines the central efficiency estimate, and the uncertainty band taken as the interval that is centered on the mode and contains 68% of the estimated efficiency values.

Figure 8.3 shows the dependence of the monojet trigger efficiency on a few kinematic variables, both for the NN-derived efficiency and the efficiency computed in the traditional way. As values of the H_T are scanned over, the NN efficiency reveals a dependence on correlations between the H_T and H_T^{miss} . Importantly, this manifests as the loss of efficiency for very high H_T around $H_T^{\text{miss}} \approx 200 - 300$ GeV. Traditional, binned efficiency measurements may easily hide such inefficiencies, particularly if regions with low efficiency are combined with regions of maximum efficiency. The NN method provides is a natural way to avoid bias associated with binned efficiency estimates.

This method is particularly well-suited for analyses that consider events in regions of kinematic phase space in which the trigger efficiency is varying as a function of the observables. Typically, analyses avoid such regions, since traditional trigger efficiency estimation techniques can hide potentially detrimental effects in these regions, such as the efficiency loss at high H_T revealed in Fig. 8.3. Such kinematic regions include the low to moderate H_T and H_T^{miss} regions, which were identified as potentially signal rich in Chapter 7.

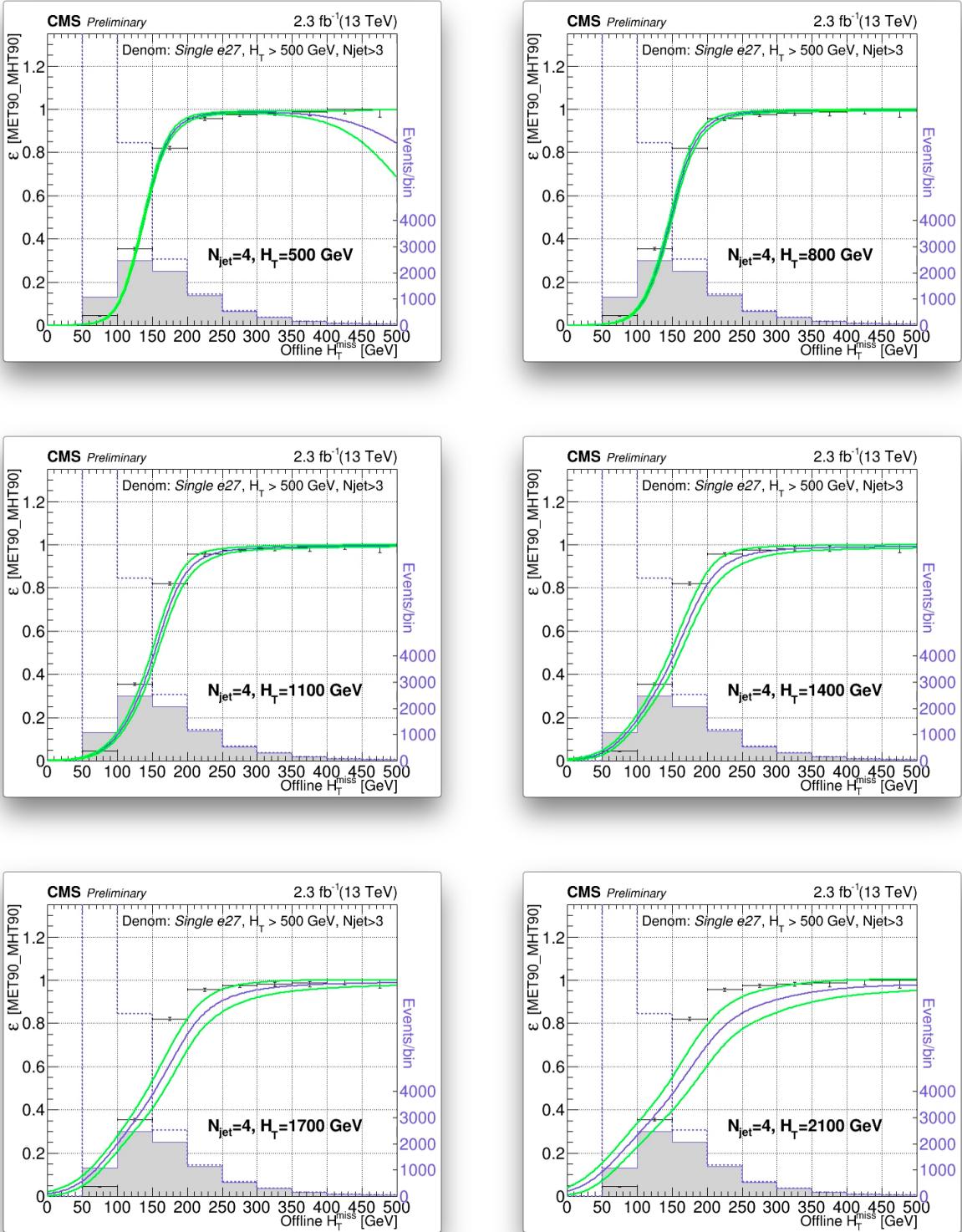


Figure 8.3: The trigger efficiency as a function of the offline H_T^{miss} for the monojet trigger, measured the traditional way (histograms) and by the NN method (blue function), for $N_{\text{jets}} = 4$, with varying values of the H_T . The green function lines show the 1-sigma uncertainty on the efficiency.

8.2 QCD background estimation

Purely hadronic events originating from QCD interactions in proton-proton collisions account for the vast majority of events observed at the LHC. Not surprisingly, these events are a major background to new physics signals that may manifest in the hadronic channel, particularly in regions of low E_T^{miss} . Several features of multi-jet QCD events are poorly modeled, including the production cross section, jet multiplicity, and relationships between the directions of jets. This motivates the development of data-driven approaches to QCD estimation. Generally, one takes advantage of the particularly well-modeled aspects of event simulation, but relies on the real data to model as many features as possible. As with any data-driven approach, the method must be robust against possible signal contamination that may enter into control regions.

8.2.1 The origin of E_T^{miss} in QCD events

In order to estimate the QCD background contribution in signal regions with large E_T^{miss} , it is relevant to note that the only stable particles in the standard model that are invisible to the CMS detector are neutrinos. Apart from rare final states containing heavy-flavor jets, QCD events are free of neutrinos, and therefore exhibit little or no E_T^{miss} at the parton level. For this reason, the E_T^{miss} typically serves as an excellent discriminating variable between the large QCD background and events with significant E_T^{miss} , such as models of R -parity conserving supersymmetry. However, QCD events *only* exhibit zero E_T^{miss} at the parton level (E_T^{miss})_{part}. The final-state particles are detected by a tracker and calorimeters with finite momentum and energy resolution. Mis-measurements of the jet momenta propagate into the measured missing transverse energy, which can result in the measurement of large missing transverse energy, (E_T^{miss})_{meas}, and so QCD events can occupy a similar kinematic region as events predicted by models of new physics. These statements can be summarized as follows:

$$(\vec{E}_T^{\text{miss}})_{\text{part}} \equiv - \sum_{i=1}^n (\vec{p}_T)_{i,\text{part}} = 0 \quad (8.4)$$

$$(\vec{E}_T^{\text{miss}})_{\text{meas}} \equiv - \sum_{i=1}^n (\vec{p}_T)_{i,\text{meas}} \neq 0, \quad (8.5)$$

where i is the particle index and n is the number of particles in the event.

8.2.2 Model assumptions and likelihood

In general, the magnitudes of the four-vectors of jets at the parton and reconstruction levels differ. However, assuming their directions are identical, the following expression holds for the collection of n jets of a given event,

$$\vec{J}_{\text{meas}} = \hat{C} \vec{J}_{\text{part}}, \quad (8.6)$$

where \vec{J}_{meas} and \vec{J}_{part} are $n \times 1$ vectors of the reconstructed jet four-vectors and the parton-level jet four-vectors, and \hat{C} is diagonal $n \times n$ matrix whose elements are the jet energy scale factors (c_1, c_2, \dots, c_n). The likelihood for a scale factor $c_i \equiv p_{i,\text{meas}}^\mu / p_{i,\text{part}}^\mu$, given by

$$L_i \equiv P(p_{i,\text{meas}}^\mu | p_{i,\text{part}}^\mu) = P(c_i | p_{i,\text{part}}^\mu), \quad (8.7)$$

is derived from simulation as the distribution of the ratio of reconstruction-level jet momentum to the associated parton-level jet momentum. The association of parton- and reconstruction-level jets is made by the matching criterion,

$$\Delta R(p_{\text{meas}}^\mu, p_{\text{parton}}^\mu) < 0.4. \quad (8.8)$$

Additionally, an isolation criterion,

$$p_T / \sum_{i=1}^{N_{\text{jets}}} [(p_T)_i \cdot \Theta(0.5 - \Delta R(p_i^\mu, p^\mu))] < 1.01, \quad (8.9)$$

is applied both to the parton-level jets and to the reconstruction-level jets that contribute to the likelihood. The jet p_T likelihoods (jet response functions), as shown in Fig. 8.4, are derived in simulation. The likelihoods are measured in 800 finely-spaced bins of parton-level jet p_T and η .

8.2.3 Rebalance and smear method

The rebalance and smear method, originally developed by the authors of the Dissertations [104][105][106] and the paper [107], exploits the relationships given in Equations 8.5,

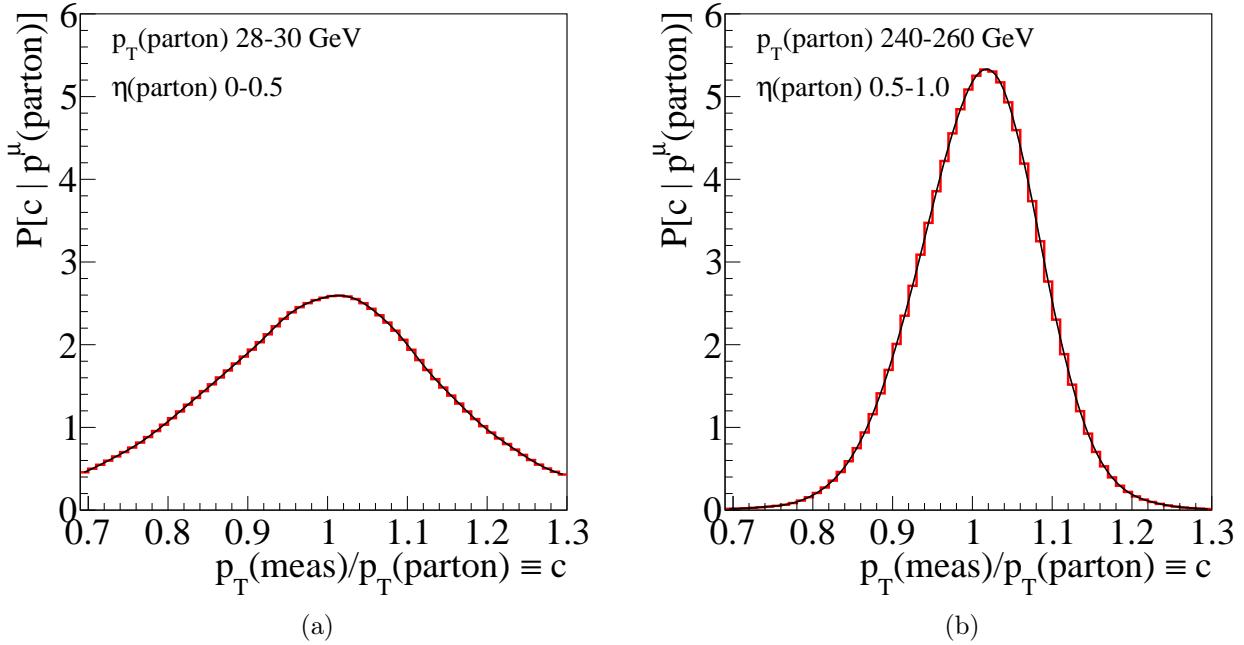


Figure 8.4: The likelihood function for the jet energy scale factor in two regions of the phase space of the parton-level jet four-vector. These distributions are equivalent to the smearing templates. The histograms (red) are smoothly interpolated using splines (black).

along with the jet energy resolution, to form a transfer function between the measured and parton-level jets. The first step of the method is to identify the optimal matrix of scale factors \hat{C} that transforms the collection of reconstruction-level jets into a set that resembles a parton-level jet collection, and (historically) yields an event with $H_T^{\text{miss}} = 0$; jets in the resulting collection are referred to as the rebalanced jets. In the second step, the four-vectors of the rebalanced jets are smeared according to the likelihood for the scale factor given in Equation 8.7; events can be smeared as many times as is necessary to make predictions in all signal regions. The above procedure, applied to all events in a QCD-enriched data control sample, yields an event sample that is the basis of the QCD background prediction. Predictions for the QCD background in the signal regions are derived from cuts applied to this sample, which we may refer to as the prediction event sample. I have adapted key aspects of the methodology, which I will now describe step-by-step.

Rebalance procedure: a Bayesian approach. The objective is to rebalance the collection of measured jets so that it resembles a parton-level collection. In the re-envisioned approach, it is possible to systematically incorporate prior knowledge about the true H_T^{miss} distribution and jet response functions to constrain the jets. An inversion of the likelihood function for the jet energy scale factors is performed using Bayes' theorem (see Chapter 7 for a brief introduction). The probability density for the parton-level jet collection can be written as

$$P(\vec{J}_{\text{part}} | \vec{J}_{\text{meas}}) \sim P(\vec{J}_{\text{meas}} | \vec{J}_{\text{part}}) \cdot \pi(\vec{J}_{\text{part}}), \quad (8.10)$$

where $\pi(\vec{J}_{\text{part}})$ is the n -dimensional prior probability density for the parton-level jet collection. Treating the jets as mutually independent allows the likelihood to be factorized as

$$\begin{aligned} P(\vec{J}_{\text{meas}} | \vec{J}_{\text{part}}) &= \prod_{i=1}^{N_{\text{jets}}} L_i = \prod_{i=1}^{N_{\text{jets}}} P(p_{i,\text{meas}}^\mu | p_{i,\text{part}}^\mu) \\ &= \prod_{i=1}^{N_{\text{jets}}} P(c_i | p_{i,\text{part}}^\mu). \end{aligned} \quad (8.11)$$

The prior allows our knowledge about the parton-level missing transverse energy outlined in Equation 8.5 to constrain of the jets. However, while equations 8.5 are true for the E_T^{miss} , high pileup conditions necessitate the use of the so-called missing transverse hadronic energy H_T^{miss} , defined as

$$\vec{H}_T^{\text{miss}} \equiv - \sum_{i=1}^{N_{\text{jet}}} (\vec{p}_T)_i \cdot \Theta(30 \text{ GeV} - (p_T)_i). \quad (8.12)$$

The threshold on the p_T applied via the Heaviside function to remove jets originating from pileup interactions spoils the equality in Equation 8.5. Jets with p_T less than 30 GeV can recoil off harder jets, and this produces non-zero true H_T^{miss} . However, the parton-level H_T^{miss} is still small compared to the reconstruction-level H_T^{miss} , as seen in the distribution of parton- and reconstruction-level H_T^{miss} in Fig. 8.5, implying that the underlying H_T^{miss} distribution may provide a meaningful constraint on our knowledge of the parton-level system. A difference is also seen in the angular distribution of the H_T^{miss} , in the polar coordinate system with the z-axis defined along the direction of the leading jet, and this information is likewise incorporated.

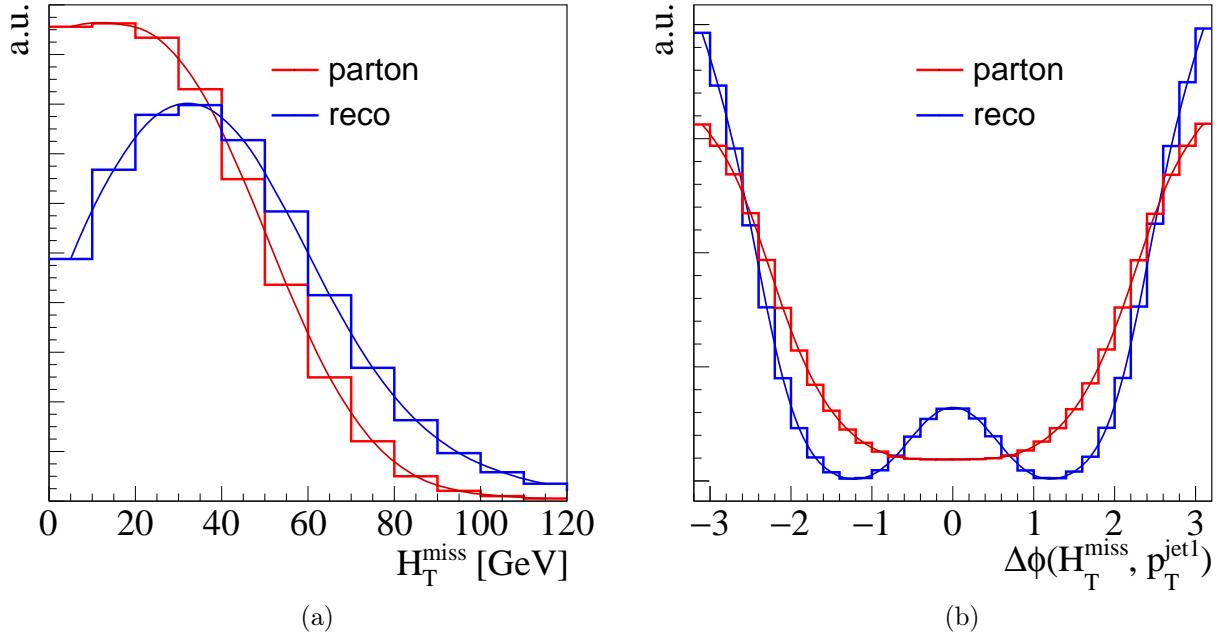


Figure 8.5: The parton-level and reconstruction-level H_T^{miss} and azimuthal separation between the H_T^{miss} and leading jet in QCD simulation. The red functions are taken as probability distributions making up the prior, $P[H_T^{\text{miss}}(\vec{J}_{\text{part}})]$ and $P[\Delta\phi_{H_T^{\text{miss}}, p_T^1}(\vec{J}_{\text{part}})]$

The expression relating the parton-level H_T^{miss} to the jet energy scale factors is used, in conjunction with the red probability densities, to form factors in the prior. The templates for the $\Delta\phi$ between the H_T^{miss} and the leading jet are derived separately in bins of the b-jet multiplicity, for the bins of 0, 1, and ≥ 2 b-jets. This choice and its implications are discussed later. For the case of one or more b-jets, the $\Delta\phi$ between the H_T^{miss} and the leading b-jet is used. Combining Eqs. 8.10 and 8.11 gives

$$P(\vec{J}_{\text{part}} | \vec{J}_{\text{meas}}) \sim \prod_{i=1}^{N_{\text{jets}}} P(p_{i,\text{meas}}^\mu | p_{i,\text{part}}^\mu) \cdot P[H_T^{\text{miss}}(\vec{J}_{\text{part}})] \cdot P[\Delta\phi_{H_T^{\text{miss}}, p_T^1}(\vec{J}_{\text{part}})] \cdot \pi_0(\vec{J}_{\text{part}}), \quad (8.13)$$

where $\pi_0(\vec{J}_{\text{part}})$ is the initial prior on the parton jet four-vectors, taken to be uniform. Having constructed a posterior density, the parton-level jets are inferred by integrating with respect to the reconstruction-level jet four-vectors, or alternatively, by performing a likelihood maximization. The second approach is taken.

Posterior density of jet momenta. Bayes' theorem suggests that incorporating new relevant information will lead to improved knowledge about a system. In the case of the rebalance and smear method, a rebalanced event is expected to more closely resemble the parton-level event than does the reconstruction-level event. Figure 8.6 shows that the resolution of rebalanced jets is better than that of reconstructed jets by approximately 10%. We also observe the peak of the response to be slightly lower for rebalanced jets than for

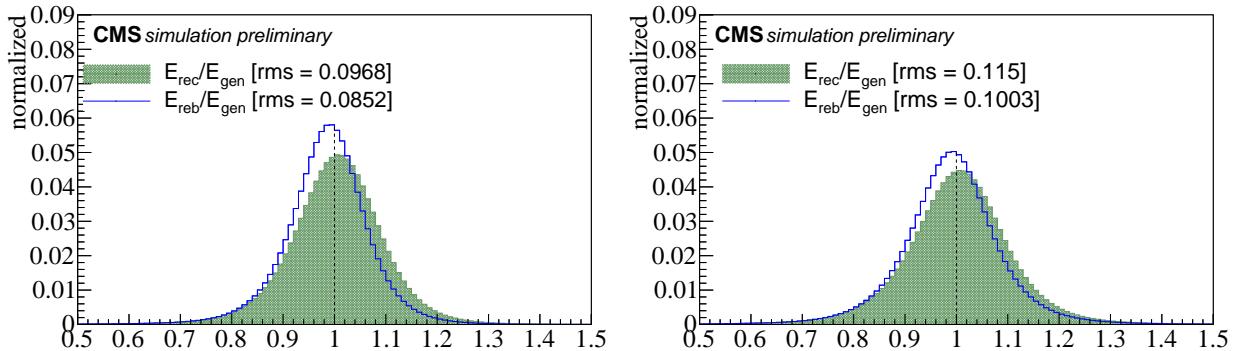


Figure 8.6: The jet energy response for the leading jet (left) and sub-leading jet (right) for reconstructed jets (green) and rebalanced jets (blue). Parton-level jets are required to have $p_T > 30$ GeV and $|\eta| < 2.4$. The resolution of rebalanced jets is better (smaller) than that of reconstruction-level jets by about 10%.

reconstructed jets. This is consistent with the peak of the jet p_T likelihood functions being centered at values slightly higher than 1, as seen in Fig. 8.4.

8.2.4 Closure

The rebalance and smear prediction is applied to simulated events and compared with the result obtained directly from the simulation. A set of baseline selection, similar to that applied in the search [6], is applied. Events are required to have:

- no reconstructed, isolated particle track with a $p_T > 10$ GeV and $|\eta| < 2.4$;
- $H_T > 500$ GeV;
- $H_T^{\text{miss}} > 150$ GeV;

- $N_{\text{jets}} \geq 4$, where jets are required to have a $p_T > 30$ GeV and $|\eta| < 2.4$, and pass a loose set of quality selection criteria.
- an azimuthal separation between the H_T^{miss} and the leading four jets $\Delta\phi(H_T^{\text{miss}}, \text{jet}_{1,2,3,4}) > 0.5, 0.5, 0.3, 0.3$.

Figures 8.7 through 8.8 show this comparison for a number of observables after the baseline selection of the CMS multi-jet SUSY search, which is described in Section 8.4. The plots show “n-1” distributions, meaning all baseline selection have been applied except that of the x-axis variable. Then, Figs. 8.9 through 8.10 show the comparison in an inverted $\Delta\phi$ region that has a selection equivalent to the baseline but with the selection on the $\Delta\phi$ between the H_T^{miss} and the jets inverted. Figure 8.11 shows the comparison in two dimensions for selected pairs of observables.

The degree of consistency between the prediction and expectation is excellent in all considered kinematic regions, with the exception of large $N_{\text{b-jets}}$ region. Not only are the relevant 1-dimensional distributions accurately modeled, but all pairwise correlations of the kinematic observables examined show good closure. This includes correlations between the directions and magnitudes of the momenta of jets within an event, as well as relationships between the directions of the H_T^{miss} and the jets, and the magnitudes of the H_T^{miss} and H_T . The most significant exception to these statements is with the distribution of the multiplicity of b-tagged jets, where non-closure is evident for events with ≥ 3 b-jets on the order $\approx 50\%$. A possible source of this discrepancy is the choice of binning for the parton-level H_T^{miss} templates that constitute the prior, namely, $N_{\text{b-jets}} = 0, 1$, and ≥ 2 . A modification of this choice to include templates for the individual bins of $N_{\text{b-jets}} = 2, 3$, and ≥ 4 may lead to improved results in the region of large b-jet multiplicity.

The method is ready to be applied in the data, but I first give a summary of the key developments, a note on the strengths of the method, and a breakdown of uncertainties associated with the method.

What were the key developments? It appears to be feasible to use correlations among jets in QCD events as a means of achieving signal-background discrimination in

kinematic regions dominated by QCD, such as the regions of low- H_T^{miss} and low H_T . The possibility of applying the method in the low- H_T^{miss} region is for the first time feasible. It was not possible before because the classical method does not accurately model the QCD kinematics in this region (see Fig. 8.12). The primary reasons for these improvements with respect to the classical method are that a realistic parton-level H_T^{miss} constraint is used in the rebalance procedure, whereas, in the classical method, every event is rebalanced to an H_T^{miss} of 0 or a constant value chosen by the user, amounting to a delta function constraint. Second, the full jet response has been used in the likelihood maximization, rather than gaussian approximations. Gaussian approximations can lead to mis-modeling of the jet p_T or H_T^{miss} spectrum in events containing jets with small p_T , which exhibit a highly non-Gaussian response.

The rebalance and smear method is robust against contamination from signal events in the prediction sample. The reason is that the rebalance procedure displaces signal-like, high- E_T^{miss} events from the high- E_T^{miss} region, to regions of low- E_T^{miss} ; after smearing, the proportion of events in the high- E_T^{miss} region truly originating from QCD processes is nearly 100% (see Fig. 8.13). The robustness of the prediction to possible signal contamination in the control region allows for real events to be used directly for the prediction. Another strength is that the rebalanced events can be smeared an indefinite number of times, which allows for the accumulation of an indefinitely large prediction sample, computer resources permitting. Therefore, estimates of the expected QCD event count and uncertainty are made in extreme tails, even in regions for which no simulated QCD events are available, or for which the expected count is much less than 1. Alternatively, the rebalanced events can be smeared once to generate a training sample of background events for a multivariate discriminant, and subsequent smearing can generate a statistically-independent prediction sample. This is further explored in Chapter 9. Essentially, this approach establishes a data-driven QCD event generator.

8.2.5 Applying the method in data

To derive the rebalance and smear prediction using real data, samples are collected with a series of pre-scaled H_T triggers,

- HLT_PFHT200_*,
- HLT_PFHT300_*,
- HLT_PFHT350_*,
- HLT_PFHT600_*,

and the un-prescaled trigger, HLT_PFHT800_*. Events collected by these triggers are pooled into a single sample, here referred to as the H_T control sample. Events are rejected if the trigger with the highest possible H_T threshold that could have fired did not fire. This criterion ensures that the probability of an event being selected is equal to the inverse of the pre-scale, since events with large H_T that happen to trigger one of the lower H_T triggers would otherwise inflate the probability slightly. Distributions of the weighted and un-weighted online H_T , and un-weighted offline H_T are shown in Fig. 8.14. The online pre-scale weighted H_T is smooth, as expected. The jet p_T likelihoods are modified by jet resolution scale factors, derived by the jet/ E_T^{miss} POG [108] to ensure the compatibility of the smearing with the true jet resolution. Rebalanced events are smeared a number of times proportional to the event’s pre-scale value, which serves to apply the pre-scale weight, and ensures that all prediction events have equal weight.

8.2.6 Uncertainties in the prediction

Uncertainties in the prediction include the direct statistical uncertainty in the weighted number of counts, statistical uncertainty associated with the H_T control sample, uncertainty in the jet energy resolution, uncertainty in the parton-level H_T^{miss} prior density, and uncertainty corresponding to any non-closure.

The statistical uncertainty in the prediction is the poisson uncertainty in the weighted number of counts in each bin. This uncertainty can, in principle, be reduced to arbitrarily

small values by increasing the number of smearing trials. Given independent search regions, statistical uncertainties are uncorrelated across bins.

The statistical uncertainty in the control sample is evaluated by making the prediction once using the entire data set, and once having randomly discarded half the control sample events. Scaling the half-discarded prediction up by a factor of two and taking the discrepancy with the original prediction yields a bin-by-bin systematic uncertainty. Uncertainties are once again uncorrelated across all bins.

The uncertainty in the jet energy resolution is evaluated by taking the difference between the nominal prediction and the prediction obtained by modifying the jet p_T likelihoods according to the jet energy resolution uncertainties provided by [108]. Rather than evaluating the bin-by-bin discrepancy, the discrepancy in the inclusive H_T^{miss} distribution integrated over the bounds of each signal region between the nominal and modified predictions is taken as a percent uncertainty. This methods ensures that statistical uncertainties are not duplicated based on the limited statistics of the control sample. Uncertainties are taken to be fully correlated across all bins.

Uncertainty in the H_T^{miss} prior is evaluated in an analogous way as the previous uncertainty, except the difference in predictions is that between the nominal prediction and the prediction made with a modified H_T^{miss} prior density. The modified prior density is obtained by reweighting the nominal H_T^{miss} prior by the ratio of the real reconstructed H_T^{miss} to the simulated reconstructed H_T^{miss} between the values of $H_T^{\text{miss}} = 0$ and 200 GeV in a QCD-enriched kinematic region. The weights take values between 0.6 and 1.0, with negligible uncertainties. Uncertainties are treated as correlated across bins.

An uncertainty of 50% is assigned to predicted counts in search bins with greater than 2 b-tags to accommodate the non-closure in simulation. In the future, this uncertainty may be eliminated by modifying the binning of the prior template as prescribe in the previous section. Non-closure uncertainties are treated as uncorrelated across bins.

The values of the uncertainties, along with the bin-by-bin prediction, for the multi-jet+ H_T^{miss} SUSY search are given in Section 8.4.

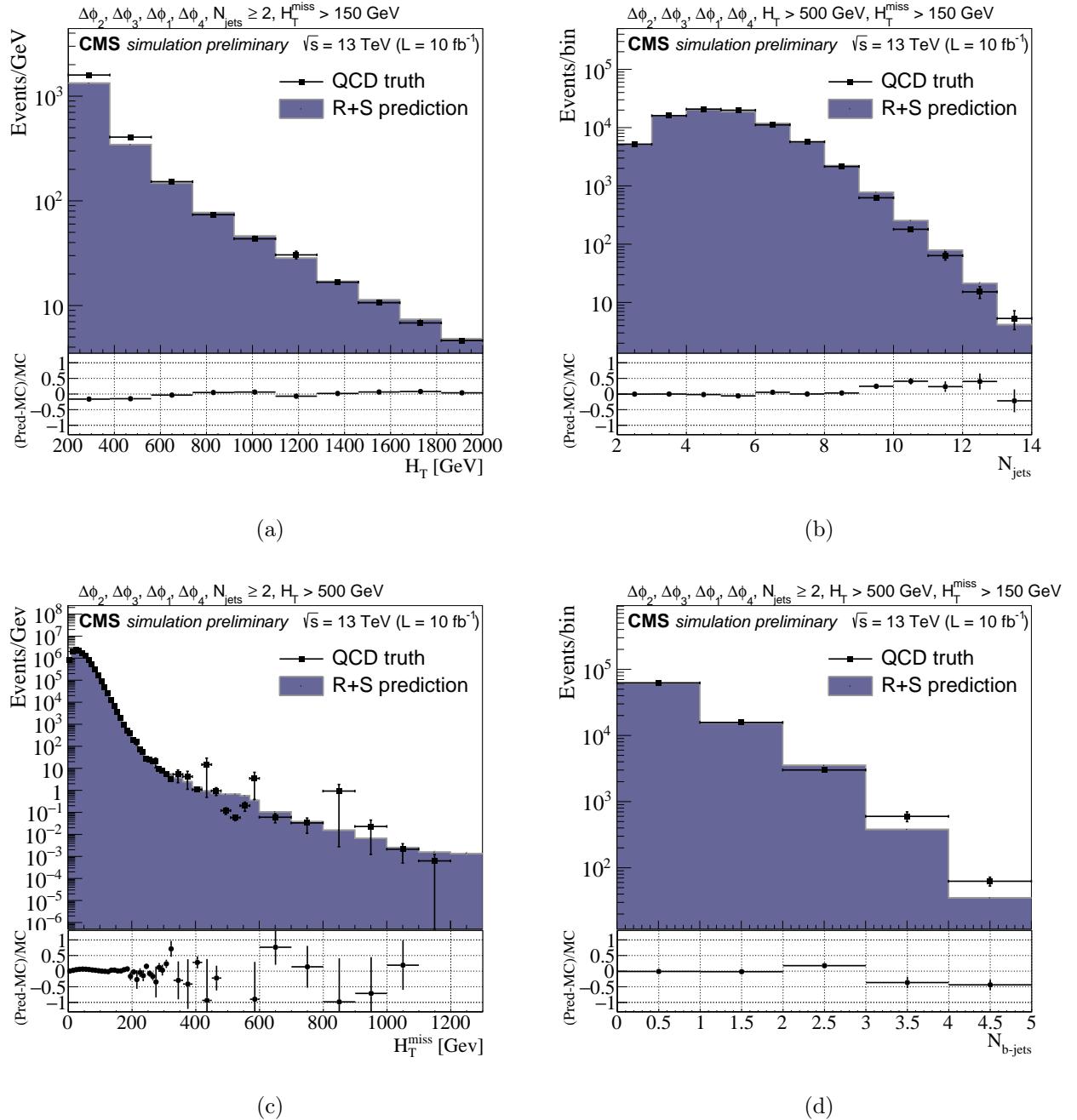


Figure 8.7: Comparisons of kinematic distributions between the direct simulation and the rebalance and smear method applied to simulation, after the baseline selection of the multi-jet SUSY search.

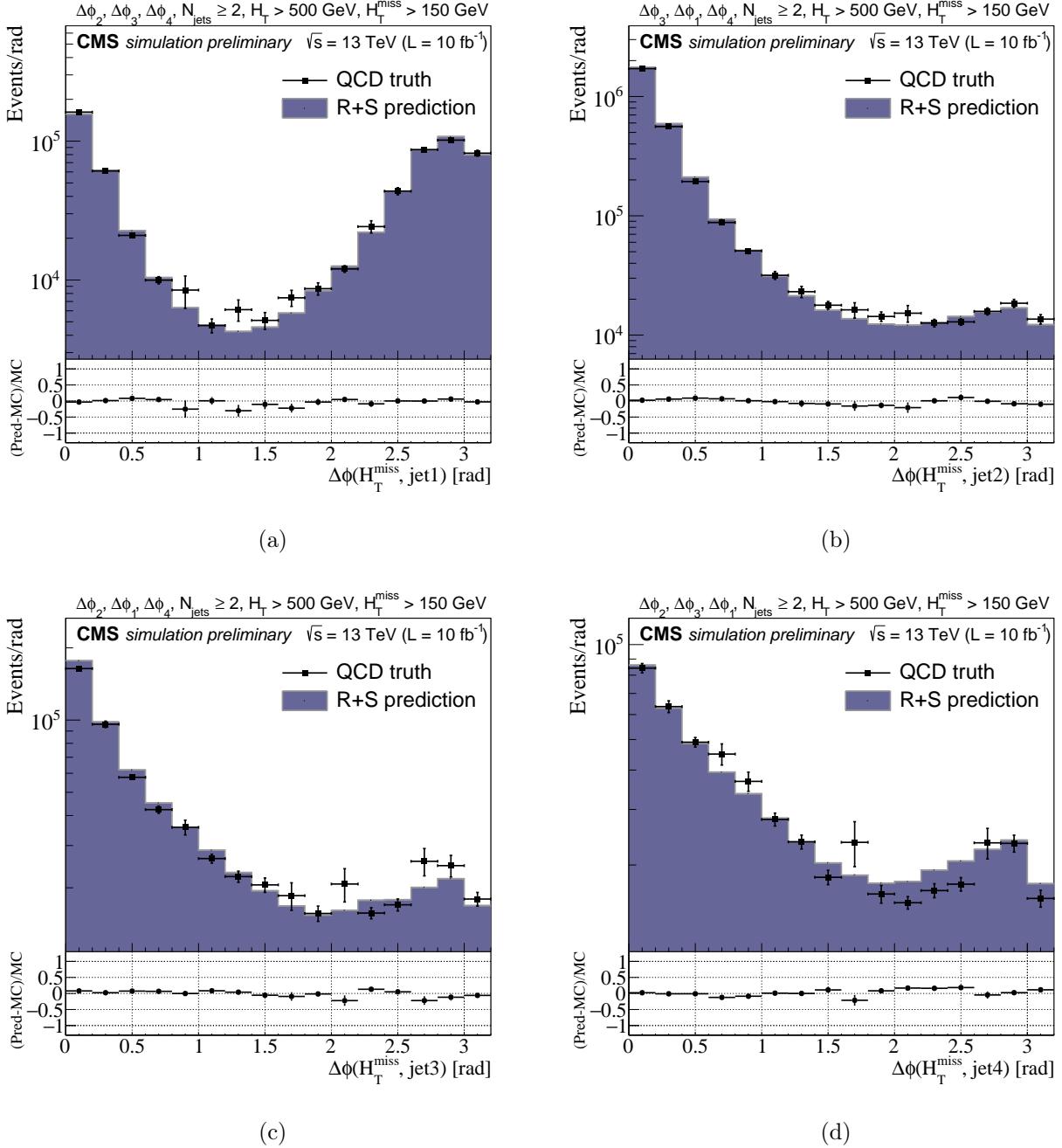


Figure 8.8: Comparisons of kinematic distributions between the direct simulation and the rebalance and smear method applied to simulation, after the baseline selection of the multi-jet SUSY search.

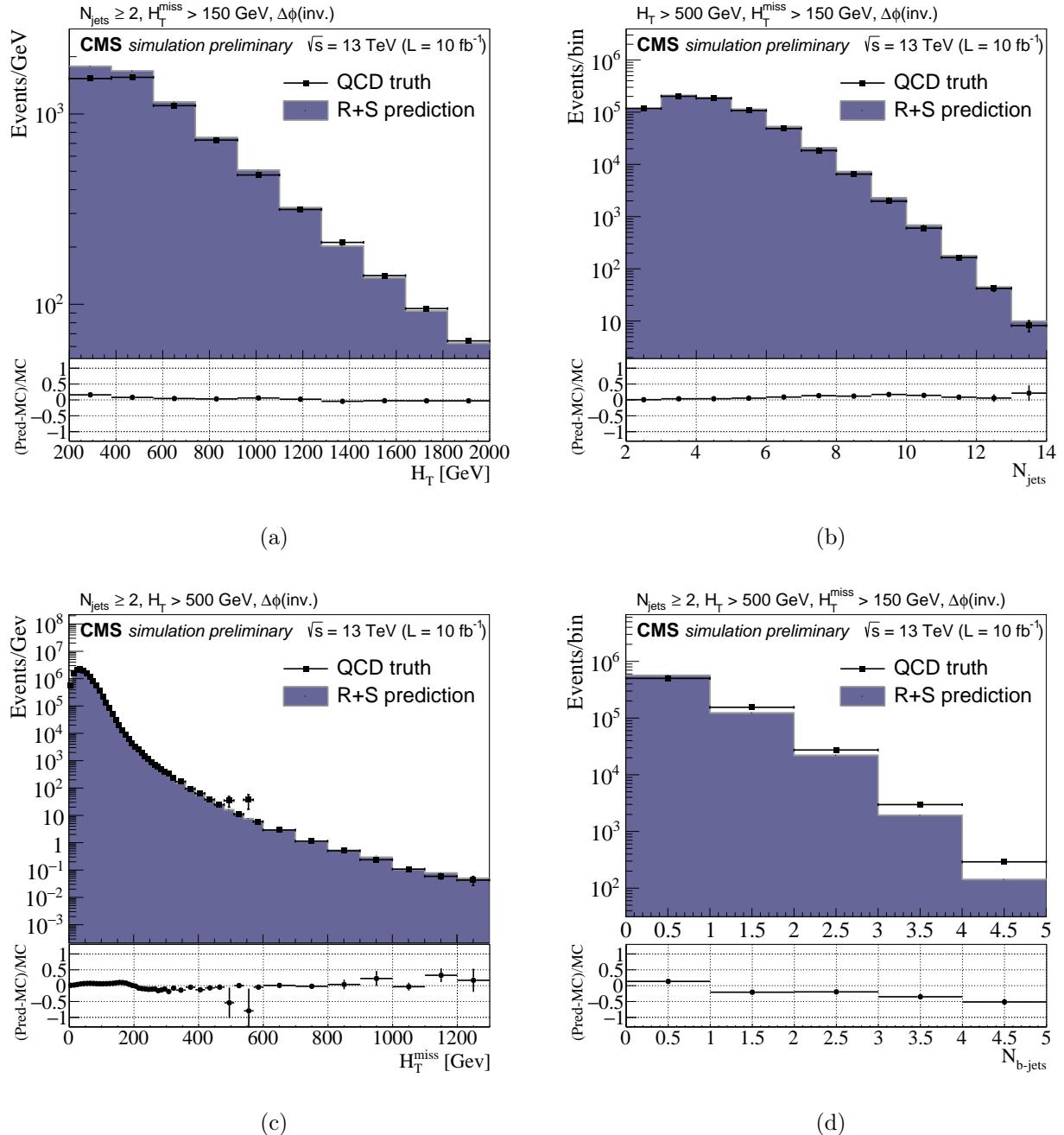


Figure 8.9: Comparisons of kinematic distributions between the direct simulation and the rebalance and smear method applied to simulation, in the inverted $\Delta\phi$ control region of the multi-jet SUSY search.

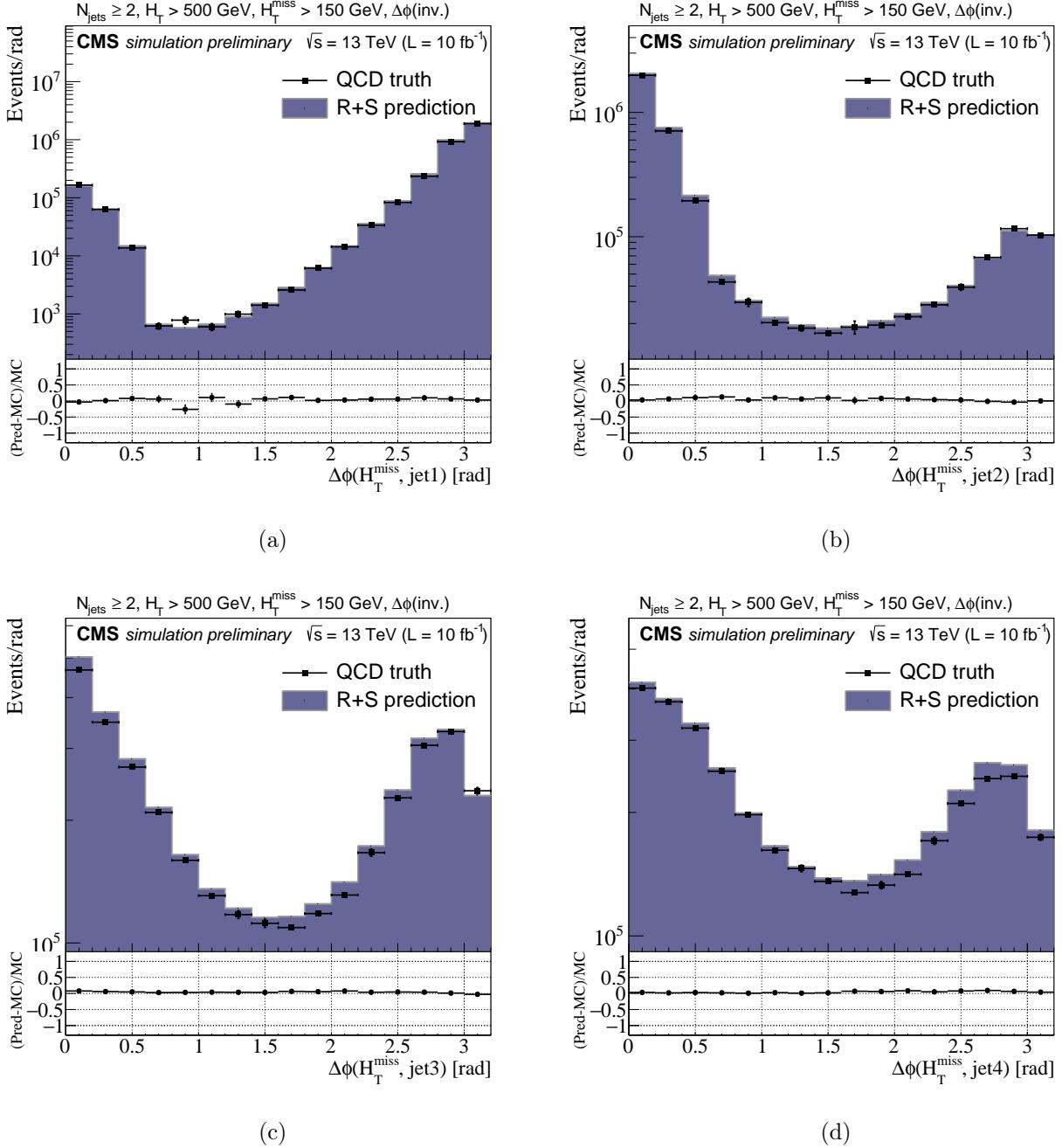


Figure 8.10: Comparisons of kinematic distributions between the direct simulation and the rebalance and smear method applied to simulation, in the inverted $\Delta\phi$ control region of the multi-jet SUSY search.

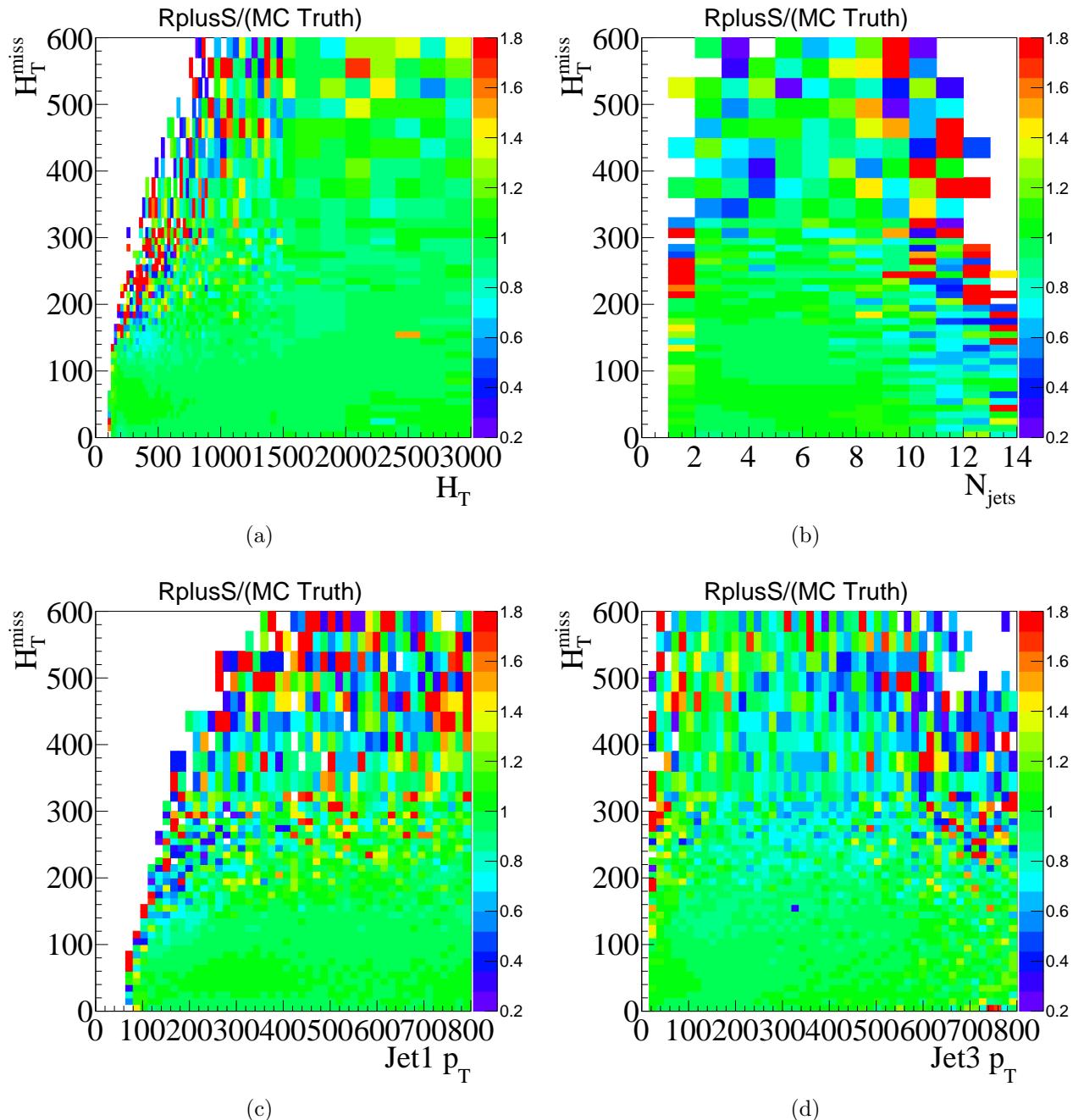


Figure 8.11: The ratio of scatter plots of key observables for hadronic SUSY searches between the direct simulation and the rebalance and smear method applied to simulation.

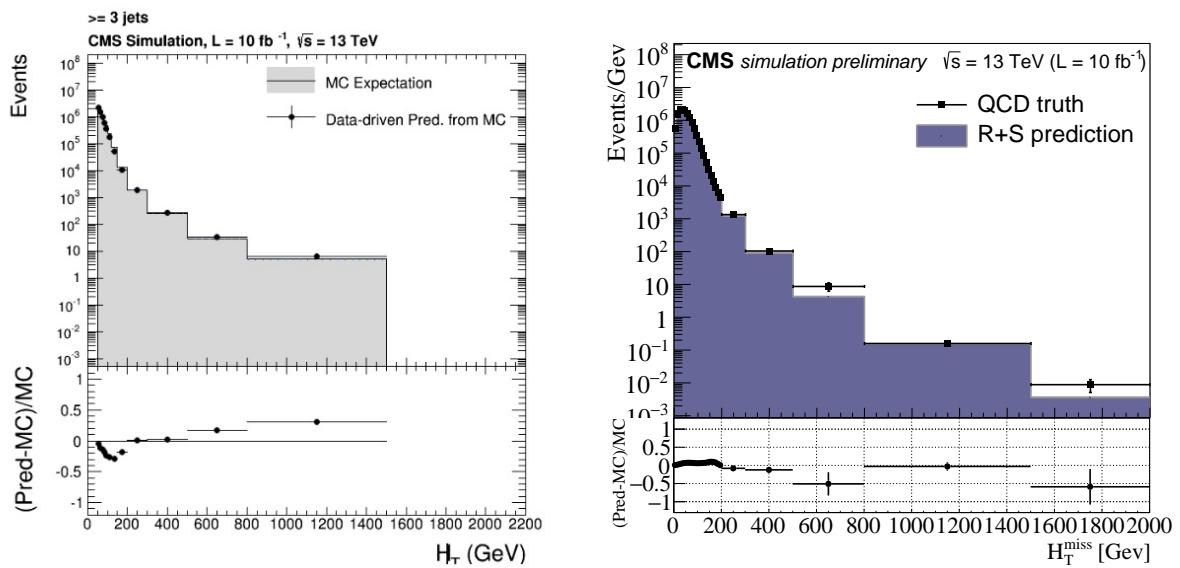


Figure 8.12: Comparisons of the H_T^{miss} distribution between prediction and expectation using the classical method (left) and the new method (right).

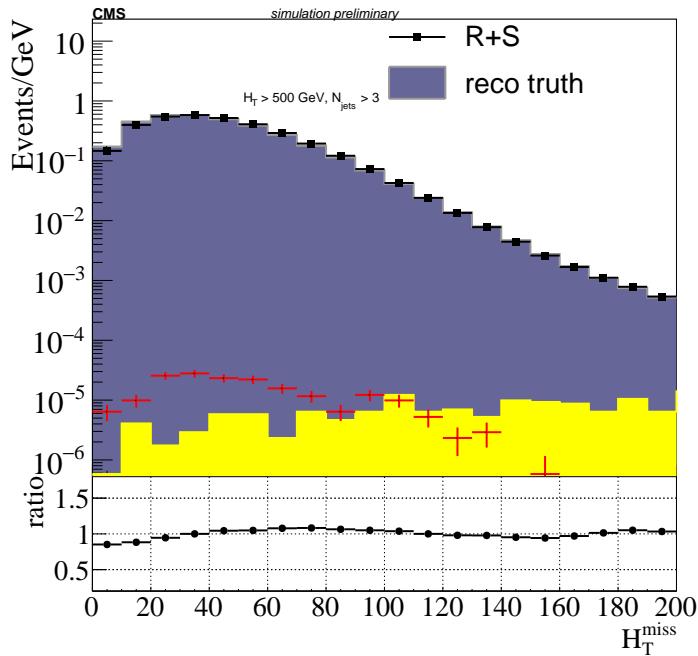


Figure 8.13: Signal contamination removal. The midnight blue (and yellow) histograms show the distribution of QCD (and signal) events taken directly from simulation. The black (and red) histograms show the distributions of QCD (and signal) and after being processed by the rebalance and smear procedure. The signal has been removed from the high- H_T^{miss} region, leaving an accurately modeled QCD contribution.

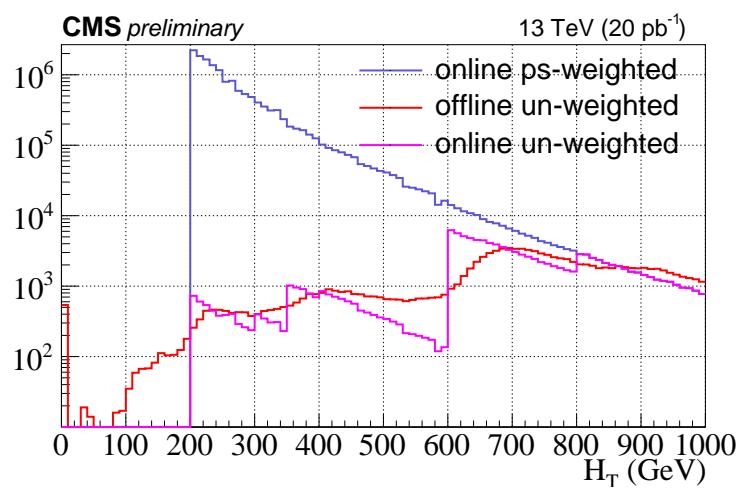


Figure 8.14: The distribution of the H_T in the H_T control sample for the unweighted offline H_T (red), the unweighted online H_T (pink), and the online H_T weighted by the pre-scale values.

8.3 $Z \rightarrow \nu\bar{\nu}$ background estimation

Events in which Z bosons are produced in association with jets account for a large fraction of the high- E_T^{miss} all-hadronic events at the LHC. Not surprisingly, these events are a major background to new physics that may manifest signals in the hadronic channel. Since the jet multiplicity beyond 4 jets, as well as the relationships between the directions of jets, are not simulated accurately, it is generally advisable to employ data-driven methods for $Z \rightarrow \nu\bar{\nu}$ background estimation, especially when the signal region is populated by events with several jets. Typically, data-driven approaches make use of similarities between the kinematics of $Z \rightarrow \nu\bar{\nu}$ events and $Z \rightarrow X\bar{X}$ events, where X is one of the other Z boson decay products, or of the similarities between events with Z bosons and events with photons. The former approach is adopted in the following.

8.3.1 The relationship between $Z \rightarrow \nu\bar{\nu}$ and $Z \rightarrow l^+l^-$

The kinematics of Z bosons are independent of the decay mode of the Z boson, as is the case for any particle that can undergo a variety of decays. The decay modes and branching fractions \mathcal{B} for the Z boson are listed in Table 8.1.

Table 8.1: The measured decay modes of the Z boson, as reported in [1].

decay mode	$\mathcal{B}(\%)$
e^+e^-	3.363 ± 0.004
$\mu^+\mu^-$	3.366 ± 0.007
$\tau^+\tau^-$	3.370 ± 0.008
$\nu\bar{\nu}$	20.00 ± 0.06
$q\bar{q}$	69.91 ± 0.06

Events featuring the $Z \rightarrow \nu\bar{\nu}$ decay mode are largely indistinguishable from other events with E_T^{miss} and no leptons, and therefore obtaining a reasonably pure data sample of $Z \rightarrow \nu\bar{\nu}$ events is not feasible. However, events with $Z \rightarrow l^+l^-$ decays can be selected with high purity by requiring two opposite-sign, same-flavor, well-isolated leptons, whose invariant mass falls within a range similar to $|m_{ll} - m_Z| < 20$ GeV. For events in such a sample, the Z boson can

be reconstructed by summing the four-vectors of the two leptons, and if the Z is removed from the event, a proxy sample for real $Z \rightarrow \nu\bar{\nu}$ events is obtained. The process of removing the tagged Z boson is referred to as event cleaning.

The resulting “cleaned” event sample is expected to exhibit *all* the characteristics of the real $Z \rightarrow \nu\bar{\nu}$ events, barring two exceptions: first, distributions of extensive properties derived from the $l\bar{l}$ sample will be too low by an overall normalization factor of $20/3.4 \approx 5.9$ because of the difference in branching fractions of the $Z \rightarrow \nu\bar{\nu}$ and $Z \rightarrow l^+l^-$ decay modes; and second, the shapes of distributions based on the $Z \rightarrow l^+l^-$ sample will be altered because the acceptance, reconstruction efficiency, and isolation efficiency of the lepton selection. To correct for the first difference, $Z \rightarrow l^+l^-$ -derived distributions are scaled up by the normalization factor 5.9. To correct for the second difference, events in the $Z \rightarrow l^+l^-$ sample are weighted by the inverse of the lepton efficiencies mentioned. The prediction for the $Z \rightarrow \nu\bar{\nu}$ count N_i in a given signal region i is therefore given by

$$N_i = N_i(Z \rightarrow l^+l^-) \cdot \frac{1}{\epsilon_{\text{acc}}^Z} \cdot \frac{1}{\epsilon_{\text{rec}}^Z} \cdot \frac{1}{\epsilon_{\text{iso}}^Z} \cdot \frac{\mathcal{B}(Z \rightarrow \nu\bar{\nu})}{\mathcal{B}(Z \rightarrow l^+l^-)}, \quad (8.14)$$

where $N_i(Z \rightarrow l^+l^-)$ is the observed count in signal region i from the cleaned dilepton sample, and the ϵ^Z 's are the efficiencies of acceptance, reconstruction, and isolation.

8.3.2 Procedure with 13 TeV data

A procedure was developed previously [7] to estimate the $Z \rightarrow \nu\bar{\nu}$ counts using events with two opposite-sign muons, using simulated events as well as real events collected in 2015. Muon events were chosen rather than electron events, since the muon efficiency is roughly 10% larger than the electron efficiency, a choice that yields the larger of the two dilepton samples. However, the number of real events containing $Z \rightarrow \mu^+\mu^-$ decays are limited because of the integrated luminosity recorded in 2015 is relatively small, and so an estimate of the $Z \rightarrow \nu\bar{\nu}$ count could not be made in all signal regions. Thus, a data-simulation hybrid approach was taken, detailed in Ref. [7], in which simulated $Z \rightarrow \nu\bar{\nu}$ events are reweighted based on the ratio of counts between data and simulation in bins of jet multiplicity and b-jet multiplicity, and are subsequently taken as the prediction. Systematic uncertainties, as

well as the weights, are derived in control regions believed to be signal-poor; the systematic effects include those associated with the normalization uncertainty, shape uncertainties, and uncertainties associated with the reweighting.

I developed the method to incorporate a data event sample with two opposite-sign electrons to complement, and further constrain, the $Z \rightarrow \nu\bar{\nu}$ background prediction. On its own, the ee-based prediction is not expected to yield a better prediction, namely, a prediction with smaller uncertainties, than the $\mu\mu$ -based prediction; however, given that electron and muon efficiencies are comparable within about 10%, the combination of $\mu\mu$ and ee data should result in a reduction in control region statistical uncertainties by $\approx 30\%$.

Reconstructed electrons are selected with p_T greater than 33 (36) GeV, a threshold chosen to ensure the double electron trigger,

- `HLT_DoubleEle33_CaloIdL_GsfTrkIdVL_MW_v`,

has a nearly 100% probability of selecting electrons passing the kinematic selection. Electrons are also required to be isolated. An isolated electron has an isolation value less than 0.1, where isolation is defined as the ratio of the p_T sum of reconstructed particles within a ΔR cone around the lepton (excluding the lepton) to the p_T of the lepton, where the cone size R varies with the the lepton p_T :

$$\text{isolation} = \sum_i (p_T)_i \cdot \Theta[R^* - \Delta R_i]/(p_T)_{\text{lepton}}, \quad (8.15)$$

$$R^* = \begin{cases} 0.2 & : p_T \leq 50 \text{ GeV} \\ (10 \text{ GeV})/p_T & : 50 < p_T \leq 200 \text{ GeV} \\ 0.05 & : p_T > 200 \text{ GeV}. \end{cases}$$

An essential objective is finding a suitable parametrization of the efficiencies, ϵ_{acc}^Z , ϵ_{rec}^Z , and ϵ_{iso}^Z , that will result in the shapes of distributions derived from the cleaned ee sample agreeing with those of the $Z \rightarrow \nu\bar{\nu}$ events. Since it is not the Z bosons themselves that are reconstructed, but their decay products, a good parametrization must incorporate information about the leptons, but characterize the Z boson itself. This approach is distinct from the approach taken in the $\mu\mu$ method where the efficiencies are parametrized in terms

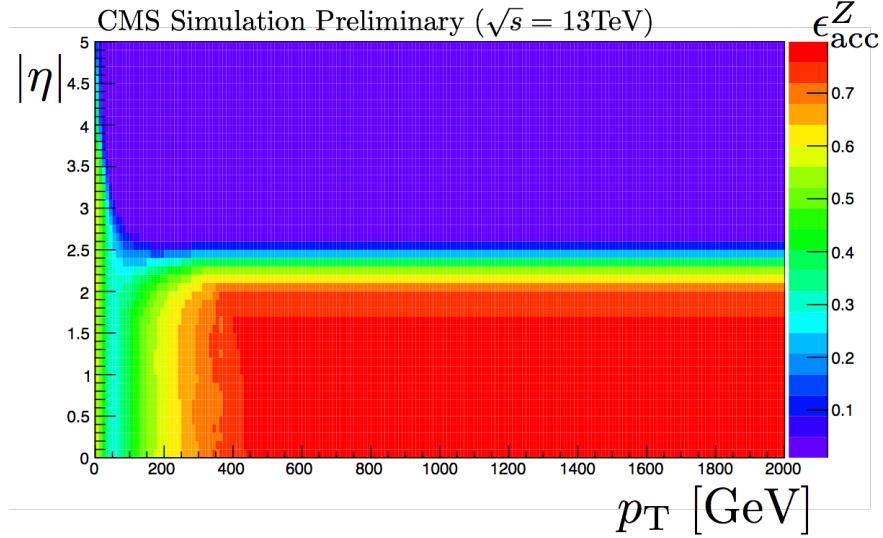


Figure 8.15: The Z boson acceptance in simulated events with $Z \rightarrow e^+e^-$.

of properties of individual leptons, and the efficiencies for the two leptons in each event are multiplied to yield an event-level efficiency. The efficiency maps are now presented, based on the parametrization that was ultimately chosen.

Acceptance ϵ_{acc}^Z . The acceptance is defined as the fraction of $Z \rightarrow e^+e^-$ events where both electrons satisfy the kinematic criteria

- $p_T > 33$ GeV
- $|\eta| < 2.5$.
- $|m_{ll} - m_Z| < 20$ GeV

Since the Z boson p_T and η are correlated with but not equal to the p_T and η of both electrons, these observables are a well-suited option for the acceptance parametrization. Figure 8.15 shows the acceptance as a function of the inferred Z boson η and p_T .

Reconstruction efficiency ϵ_{rec}^Z . The reconstruction efficiency is defined as the fraction of $Z \rightarrow e^+e^-$ events passing the acceptance criteria detailed in the previous section for which there are two reconstructed electrons in the event. The electron reconstruction is performed with the particle flow algorithm [47] introduced in Chapter 6. A set of selection criteria based

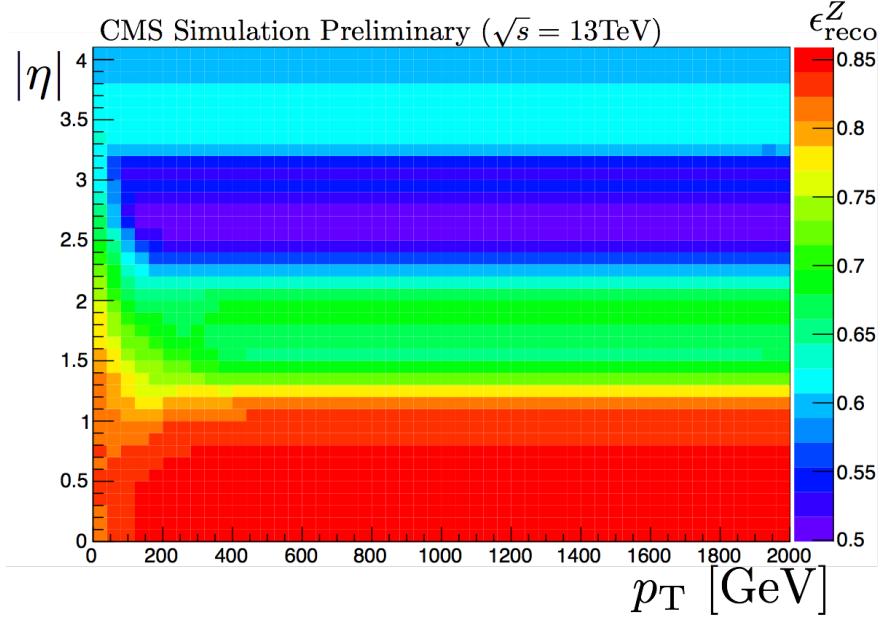


Figure 8.16: The Z boson reconstruction efficiency in simulated events with $Z \rightarrow e^+e^-$ decays.

on a recommendation by the Physics Object Group (POG) called the “Cut Based VETO” selection [109] is used, which features a very high overall electron efficiency of around 95%.

Since the reconstruction efficiency is expected to vary with the lepton p_{T} and η , the Z boson p_{T} and η are again selected as the observables used in the efficiency parametrization. Figure 8.16 shows the reconstruction efficiency as a function of the inferred Z boson η and p_{T} .

Isolation efficiency ϵ_{iso}^Z . The isolation efficiency is the fraction of $Z \rightarrow e^+e^-$ events passing the acceptance and reconstruction criteria detailed in the previous sections for which there are two isolated electrons in the event.

Unlike for the acceptance and the reconstruction efficiency, the isolation of the electrons is not strongly correlated with the p_{T} or η of the parent Z boson, at least for Z p_{T} less than 300 GeV. Observables that describe the Z boson, but which are correlated with the electron isolation, must be constructed. There are two main effects that can cause an event to fail the isolation: first, one or both of the electrons may happen to travel in a direction collinear with jets or other particles in the event. Second, in the case of a highly energetic Z boson,

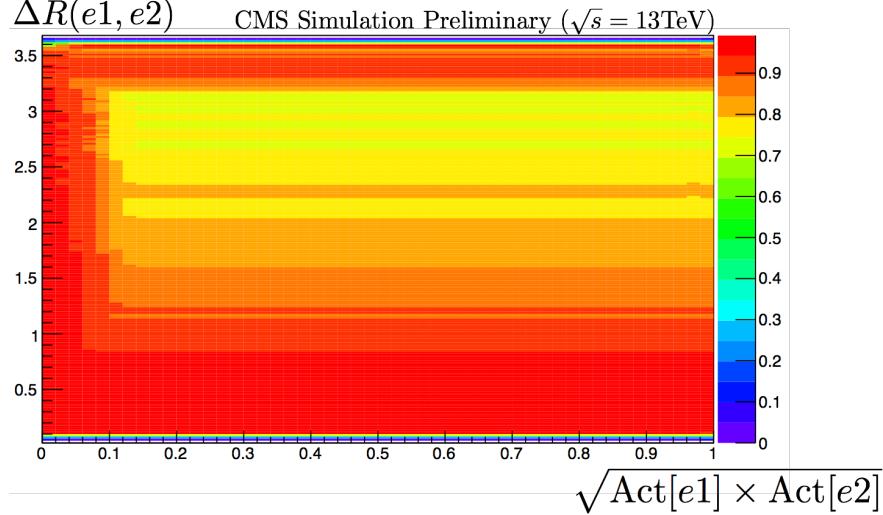


Figure 8.17: The Z boson isolation efficiency as a function of the ΔR between the two electrons and the geometric mean of the activities of the electrons in simulated events with $Z \rightarrow e^+e^-$ decays.

the electrons can be highly collinear with each other. It makes sense to parametrize the efficiency as a function of one variable that captures the effects of the first scenario, and one variable that captures those of the second. The variables chosen are the $\Delta R(e_1, e_2)$ between the two electrons, and the geometric mean of the activity of the two electrons, where the activity is defined as the p_T sum of reconstructed particles in an annulus around the lepton in the $\eta - \phi$ plane with an inner radius of the isolation cone radius and an outer radius of 0.4, divided by the lepton p_T :

$$\text{activity} = \sum_i (p_T)_i \cdot \Theta[0.4 - \Delta R_i] \cdot \Theta[\Delta R_i - R^*]/(p_T)_{\text{lepton}}, \quad (8.16)$$

where R^* is defined above. Figure 8.17 shows the isolation efficiency map using this parametrization.

8.3.3 Closure

The prediction is applied to simulated $Z \rightarrow l^+l^-$ events and compared with the result obtained directly from $Z \rightarrow \nu\bar{\nu}$ simulation. The baseline selection of [6],

- no reconstructed, isolated lepton with a $p_T > 10$ GeV and $|\eta| < 2.4$;

- no reconstructed, isolated particle track with a $p_T > 10$ GeV and $|\eta| < 2.4$;
- $H_T > 500$ GeV;
- $E_T^{\text{miss}} > 200$ GeV;
- $\Delta\phi(H_T^{\text{miss}}, \text{jet}_{1,2,3}) > 0.5, 0.5, 0.3$;
- $N_t \geq 1$, where N_t is the multiplicity of jets identified as originating from a top quark;
- $N_{\text{b-jets}} \geq 1$, and
- $M_{T2} > 200$ GeV (see Appendix C for more details).

described in greater detail in the Section 8.4, is applied. Figures 8.18 and 8.19 show the distributions for the $Z \rightarrow \nu\bar{\nu}$ prediction using Equation 8.14, the efficiencies presented in the previous section, the branching fractions quoted in Table 8.1, compared with the expectation from $Z \rightarrow \nu\bar{\nu}$ simulation. Figures 8.20 and 8.21 show the comparisons in the loosened baseline control region (also described in Section 8.4) in which the E_T^{miss} and H_T selection has been relaxed.

The prediction and expectation agree to within a few percent across all distributions, indicating the soundness of the method; in other words, the method closes. I note that the closure holds in the region of low to moderate E_T^{miss} , where it was previously determined that large nonexcluded SUSY signals could be manifest (see Chapter 7), and which will be considered in Chapter 9. The method is now at a stage where it can be applied in a fully data-driven prediction.

8.3.4 Applying the method in data

The final $Z \rightarrow \nu\bar{\nu}$ prediction used in [7] relies on a reweighting of simulated $Z \rightarrow \nu\bar{\nu}$ events to account for differences between data and simulation. Weights are derived in bins of N_{jets} and $N_{\text{b-jets}}$ by dividing the counts observed in a real dilepton data sample, which has been treated with the data-driven background estimation (Equation 8.14), by the counts predicted directly by $Z \rightarrow \nu\bar{\nu}$ simulation, after a loosened baseline selection has been applied. The loose selection is defined similarly to the baseline selection, but with

- $H_T > 200$ GeV;
- $E_T^{\text{miss}} > 0$ GeV;
- $M_{T2} > 0$ GeV;

The weights derived from the ee sample are compared with those based on the $\mu\mu$ sample in Fig. 8.22. For events with no b-tagged jets, the weights from the ee and $\mu\mu$ sample agree within the statistical uncertainties. For events with one or more b-tagged jet, at least one bin of jet multiplicity shows a noticeable departure between the two methods. It remains to be determined if this effect is statistical in nature, or if it represents a systematic difference between the electron and muon methods. Examination of the data collected in 2016 will play an important role in answering this question.

An overall simulation-to-data normalization factor is derived in a control region with 0 b-tagged jets and applied to the $Z \rightarrow \nu\bar{\nu}$ simulation as part of the final prediction. A comparison of the normalization factor derived from the ee sample with that derived from the $\mu\mu$ sample is shown in Fig. 8.23. The electron- and muon-derived normalization factors agree within statistical uncertainties, indicating that a trivial combination of the electron and muon data is justifiable in the derivation of the normalization factor and its uncertainty. After combining the electron and muon prediction samples, the systematic uncertainty in the $Z \rightarrow \nu\bar{\nu}$ prediction associated with the normalization is reduced by approximately 30%. Uncertainties are discussed now.

8.3.5 Uncertainties in the prediction

Uncertainties in the $Z \rightarrow \nu\bar{\nu}$ prediction include the statistical uncertainty of the simulated samples, uncertainty in the normalization, uncertainty associated with differences between real and simulated data, the statistical uncertainty in the $N_{\text{jets}} - N_{\text{b-jets}}$ weights, and other small uncertainties related to jet energy, E_T^{miss} , factorization scale, and PDFs.

The statistical uncertainty in the prediction is the bin-by-bin poisson uncertainty in the weighted simulated $Z \rightarrow \nu\bar{\nu}$ counts. These uncertainties are negligible for the analysis [7].

The statistical uncertainty in the normalization, as well as uncertainties in the $N_{\text{jets}} - N_{\text{b-jets}}$ weights, are propagated to the signal region, and treated as correlated across all search regions.

Uncertainties associated with the differences in the shapes of simulated and real distributions are computed bin-by-bin in the so-called loose N_{top} region, defined analogously to the baseline selection but with no criterion placed on the multiplicity of top-tagged jets. For each search bin, four comparisons are made between the data and reweighted simulated counts, one for each of the four observables, $E_{\text{T}}^{\text{miss}}$, $M_{\text{T}2}$, $N_{\text{b-jets}}$, and N_{top} . All but the observable considered are integrated over their full ranges, and the observable considered is integrated over the values set by the search bin for that observable. The maximum discrepancy among the four comparisons is taken as a percent uncertainty in the predicted $Z \rightarrow \nu\bar{\nu}$ count in the bin. Comparisons of dilepton data, treated with the data-driven method, and reweighted simulation are shown in Fig. 8.24. Uncertainties are taken to be uncorrelated across bins.

The values of the uncertainties, along with the bin-by-bin prediction, for the search for SUSY in events with top-tagged jets, are given in Section 8.4.

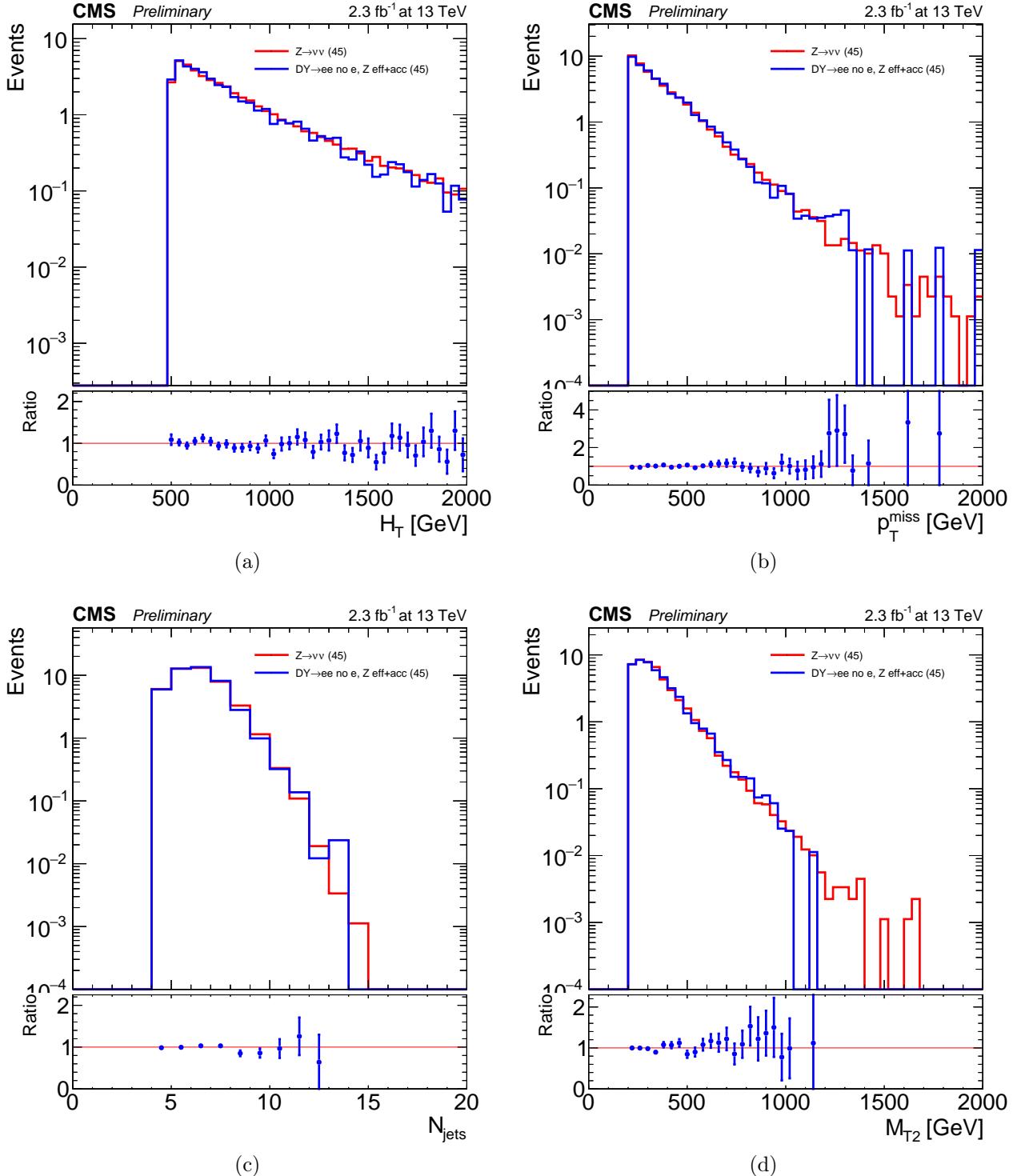


Figure 8.18: Comparison of $Z \rightarrow \nu\bar{\nu}$ prediction (blue) and expectation (red) for the H_T (1) and E_T^{miss} (2), the jet multiplicity (3) and the M_{T2} (4), after the baseline selection of the 2015 hadronic stop analysis. The ratio is of the prediction to the expectation. All events are from simulation.

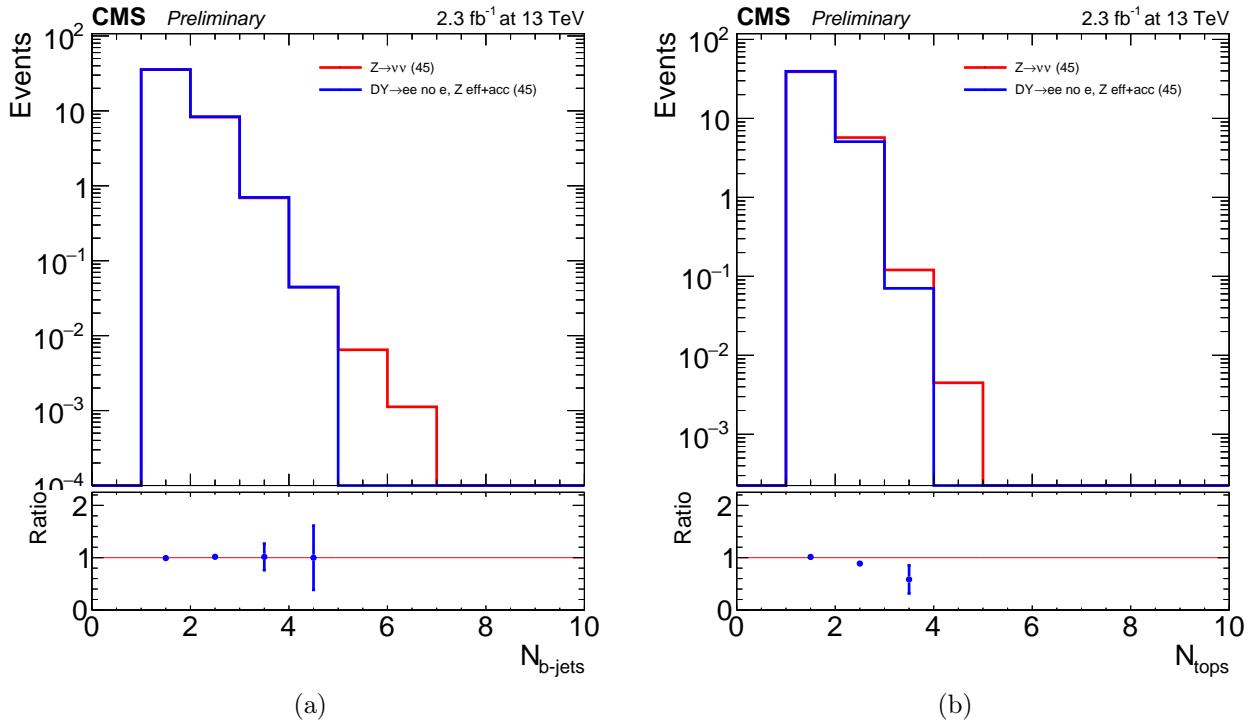


Figure 8.19: Comparison of $Z \rightarrow \nu\bar{\nu}$ prediction (blue) and expectation (red) for the b-jet multiplicity (1) and top tag multiplicity (2), after the baseline selection of the 2015 hadronic stop analysis. The ratio is of the prediction to the expectation. All events are from simulation.

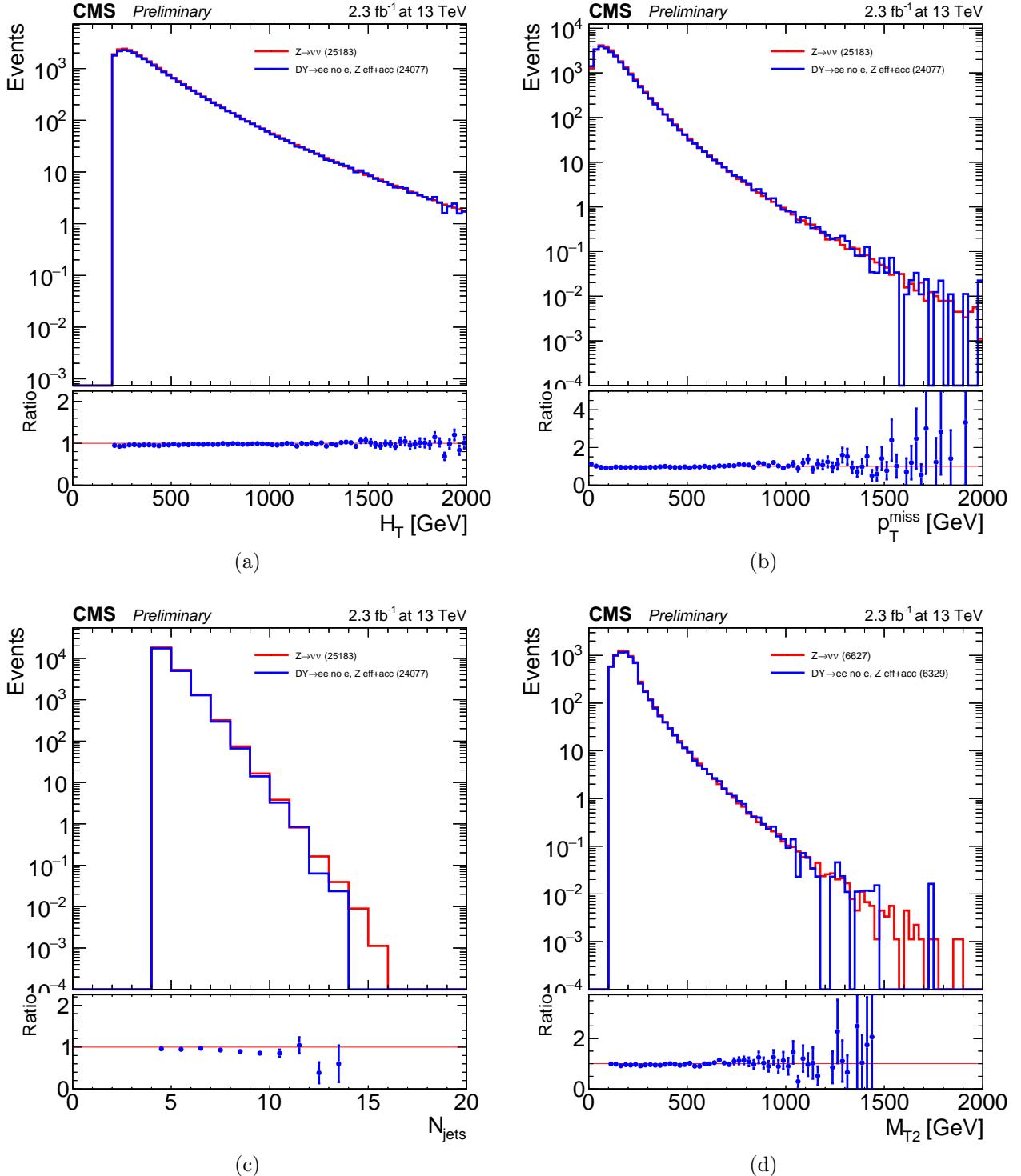


Figure 8.20: Comparison of $Z \rightarrow \nu\bar{\nu}$ prediction (blue) and expectation (red) for the H_T (1) and E_T^{miss} (2), the jet multiplicity (3) and the M_{T2} (4), after a loosened baseline selection for the 2015 hadronic stop analysis. The ratio is of the prediction to the expectation. All events are from simulation.

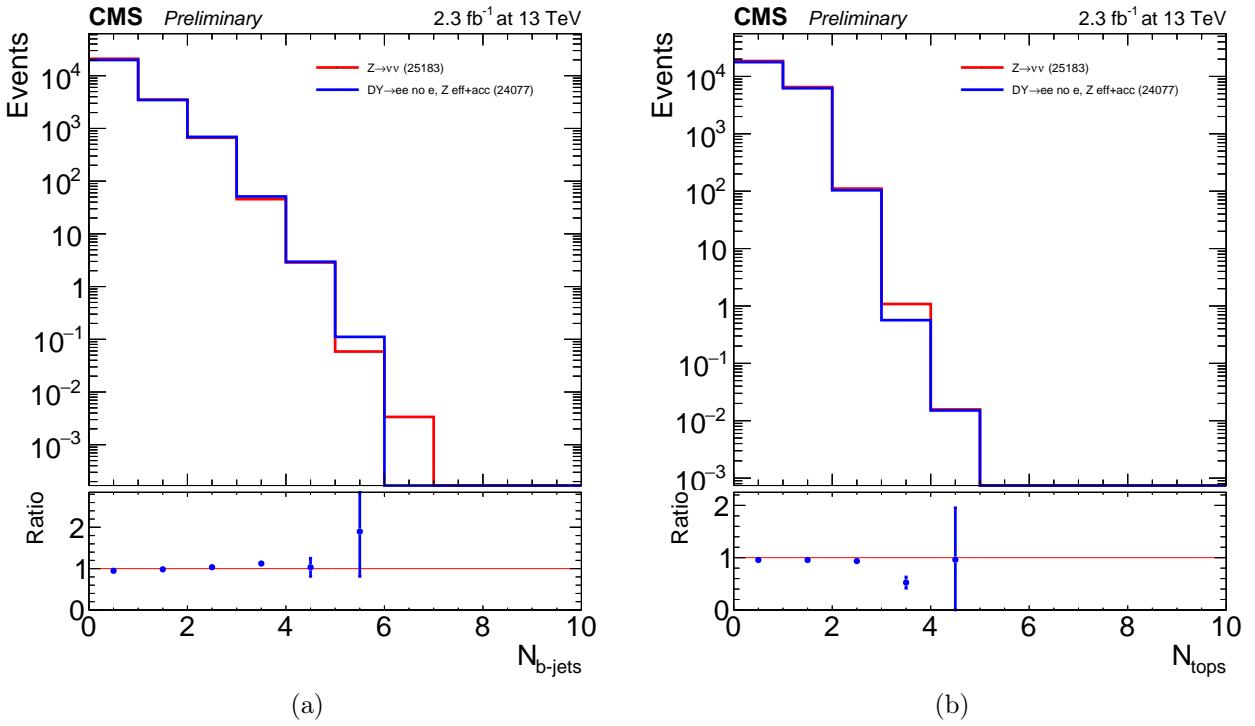


Figure 8.21: Comparison of $Z \rightarrow \nu\bar{\nu}$ prediction (blue) and expectation (red) for the b-jet multiplicity (1) and top tag multiplicity (2), after the loosened baseline selection for the 2015 hadronic stop analysis. The ratio is of the prediction to the expectation. All events are from simulation.

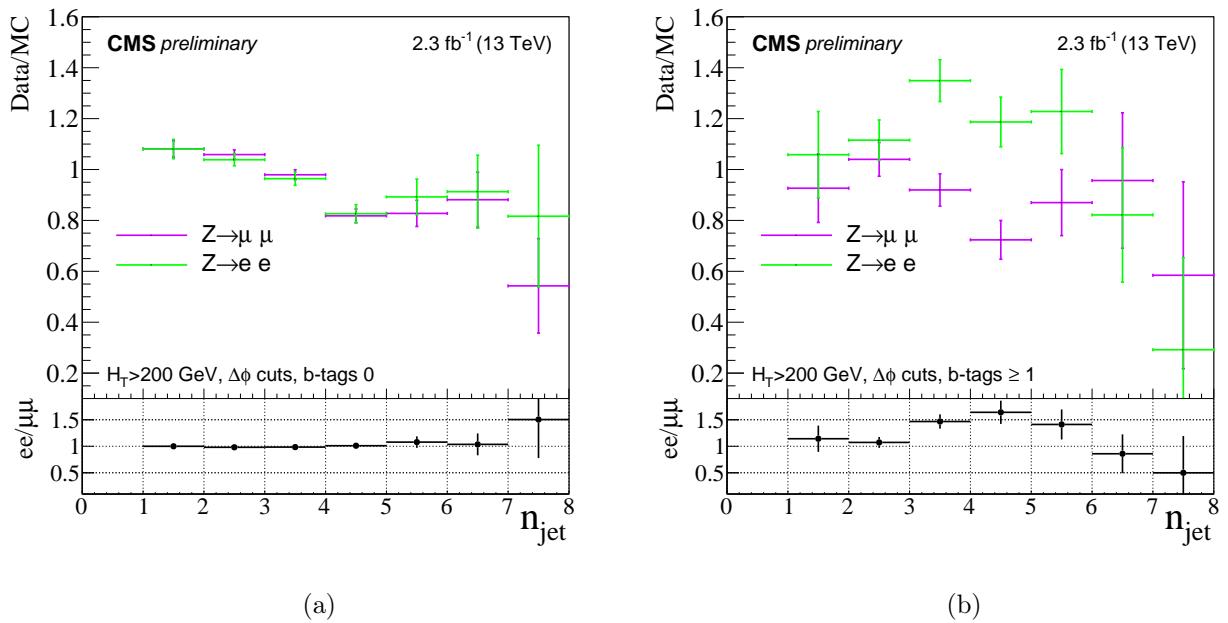


Figure 8.22: Comparison of data/simulation event weights for the $Z \rightarrow \nu \bar{\nu}$ prediction based on the electron sample (green) and the muon sample (magenta) for the b-jet multiplicity 0 (left) b-jet multiplicity greater than 0 (right), after the loosened baseline selection for the 2015 hadronic stop analysis. The denominator of the weights are the counts taken from $Z \rightarrow l^+l^-$ simulation, and the numerator is the count in $Z \rightarrow l^+l^-$ data where the $t\bar{t}$ contribution has been subtracted using simulation.

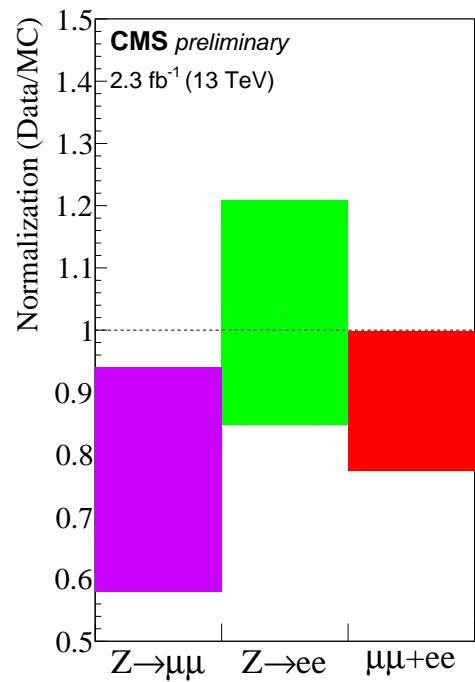


Figure 8.23: The data-simulation normalization factor computed from the $\mu\mu$ sample (magenta), the ee sample (green), and the combined ee + $\mu\mu$ sample (red).

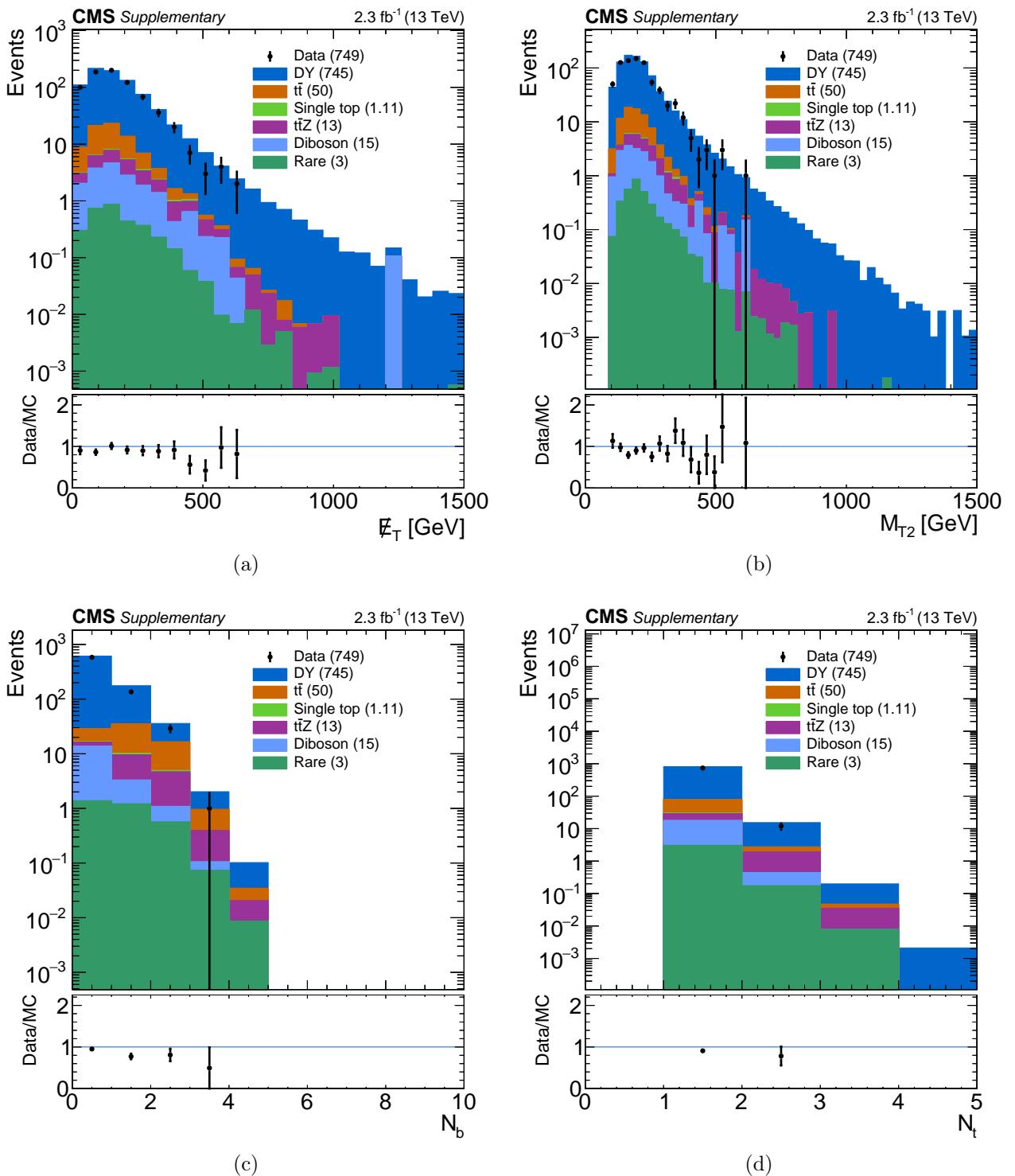


Figure 8.24: Comparisons between reweighted simulation and dilepton data.

8.4 Results of 2015 SUSY searches

The QCD background estimation, $Z \rightarrow \nu\bar{\nu}$ background estimation, and trigger techniques just described were carried out in the context of two CMS SUSY searches. This section briefly describes these analyses, their baseline selection and search region definitions, the primary backgrounds, and an interpretation of their findings in terms of various simplified models.

This section also gives the predictions and systematic uncertainties of the background estimation methods in the search bins of the two hadronic analyses, and interpretation of the results in the context of various simplified SUSY models. Upper limits are placed on the mass of the gluino between 1400 and 1600 GeV, depending on the model, and assuming a massless LSP. Limits on the top squark mass are set around 800 GeV, assuming a massless LSP. I note that the following results are the product not only of myself, but also the work of dozens of analysis team members. I provide a high degree of detail when referring to work that was specifically carried out by me. The searches are based on a sample of proton-proton collision data collected at $\sqrt{s} = 13$ TeV with the CMS detector at the CERN LHC in 2015, corresponding to an integrated luminosity of 2.3 fb^{-1} .

8.4.1 Multi-jet + H_T^{miss} SUSY search: targeting gluino production at $\sqrt{s} = 13$ TeV

This search [6] looks for an excess of events with four or more jets, no identified and isolated electron, muon, or isolated charged track, large H_T , and large H_T^{miss} . The principal standard model backgrounds are events with top quarks, W bosons and jets, Z bosons and jets, and QCD multi-jet production, and are evaluated using control samples in the data as well as information based on simulated events.

The search targets simplified model scenarios corresponding to gluino pair production. In the various models, the gluino decays 100% of the time into: an LSP and a light flavor q \bar{q} pair (T1qqqq model), an LSP and a t \bar{t} pair (T1tttt model), an LSP and a b \bar{b} pair (T1bbbb model). Also considered is a model in which the gluino decays into a q \bar{q} pair and a next-to-lightest EWkino, $\tilde{\chi}_2^0$ or χ_1^\pm (T5VV model), where the masses of the intermediate $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ states are taken to be the mean of $m_{\tilde{\chi}_1^0}$ and $m_{\tilde{g}}$. These four models are shown in Fig. 8.25.

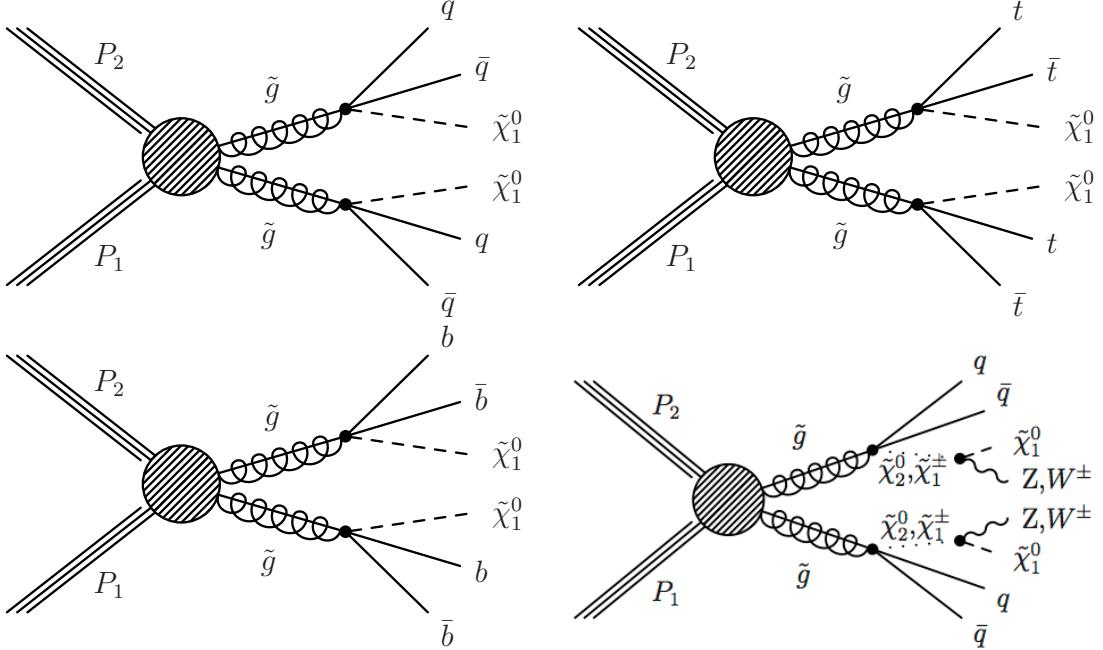


Figure 8.25: The simplified models used for the optimization and interpretation of the multi-jet + H_T^{miss} search. They are T1qqqq (upper left), T1tttt (upper right), T1bbbb (lower left), and T5VV (lower right) scenarios.

Events are collected using the hadronic trigger

- `HLT_PFHT350_PFMET100_*`,

where PFHT and PFMET refer to the online H_T and E_T^{miss} computed using the particle flow (PF) algorithm (Section 6.8), and the numbers 350 and 100 refer to the respective triggering thresholds in units of GeV.

At level 1 of the trigger system, events are triggered if they have a calorimeter-based H_T of 175 GeV. If the event is accepted, the HLT trigger then requires a calorimeter-based $H_T > 280$ GeV and a calorimeter-based $E_T^{\text{miss}} > 70$ GeV. Finally, the HLT trigger applies a lower threshold of 350 GeV on the H_T in coincidence with a threshold on the E_T^{miss} above 100 GeV computed using all particles reconstructed from information from the tracker and calorimeters.

It is seen that the choice of reference trigger, namely the single electron trigger,

- `HLT_Ele27_eta2p1_WP Loose_Gsf_v*`,

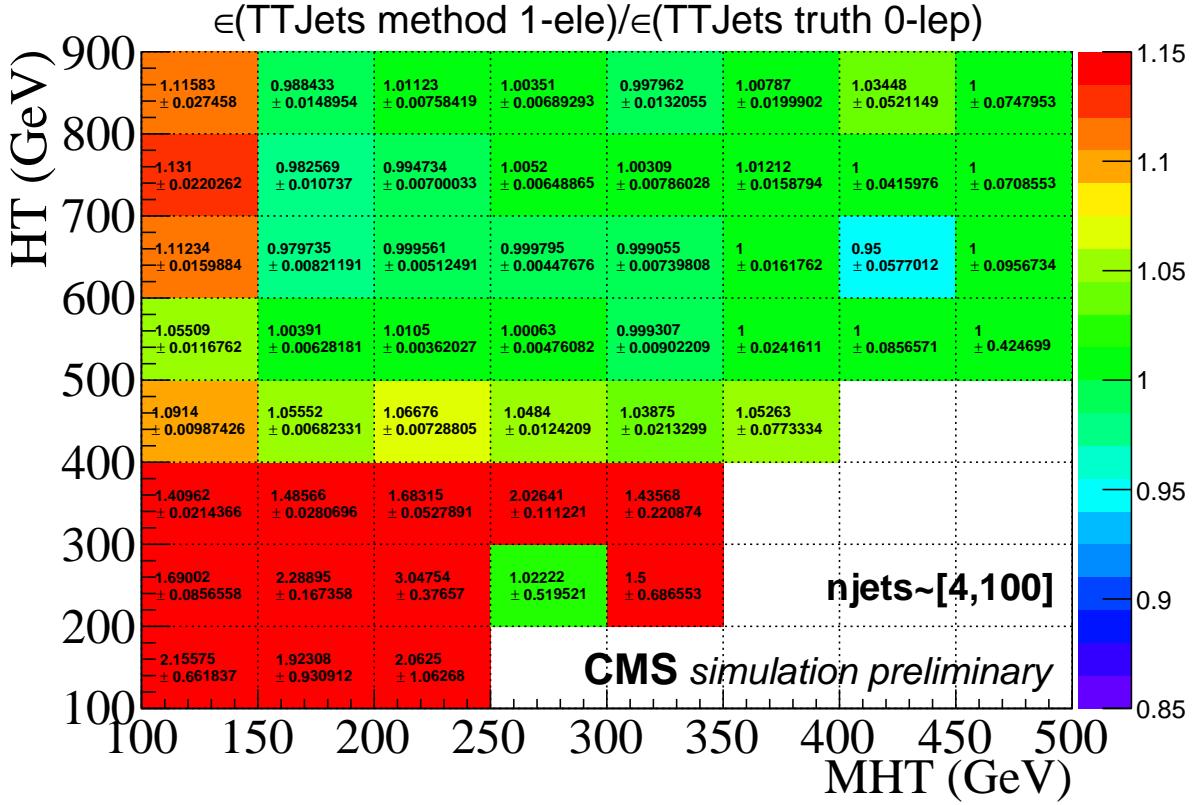


Figure 8.26: The ratio of the trigger efficiency for events passing the single electron reference trigger to the trigger efficiency for all events in a simulated $t\bar{t}$ sample, as a function of the offline H_T and H_T^{miss} . The ratio is consistent with 1 within the region of the baseline selection of $H_T > 500$ GeV and $H_T^{\text{miss}} > 200$ GeV of the CMS hadronic searches.

allows for an unbiased estimate of the trigger efficiency in the region $H_T^{\text{miss}} > 150$ GeV and $H_T > 500$. The region in which the efficiency can be measured without bias includes the baseline selection, which imposes thresholds on the offline H_T^{miss} and H_T of 200 and 500 GeV.

The efficiency, estimated without bias using Equation 8.2, is shown as a function of the offline H_T and H_T^{miss} in Fig. 8.27, where the entire 2015 dataset has been used. The statistical uncertainties in the plot correspond to the 68% CL Clopper-Pearson interval [110]. Additionally, a systematic uncertainty is assigned to the efficiency equal to the difference between the efficiency obtained from applying the method described to a sample of simulated $t\bar{t}$ events and that derived from a set of simulated signal events. Such discrepancies may arise due to any number of subtle differences in the content of the events in $t\bar{t}$ and signal

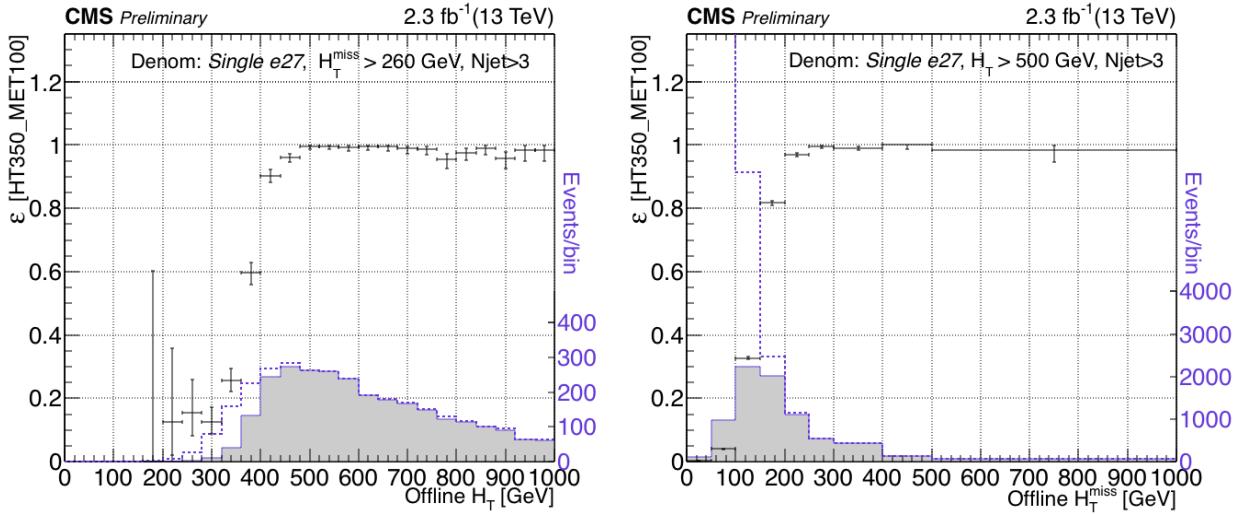


Figure 8.27: The trigger efficiency for HLT_PFHT350_PFMET100* as a function of the search variables H_T and H_T^{miss} . The dashed (solid) blue lines show the distributions of the denominator (numerator) samples. These results were used in the CMS PAS on the commissioning of 13 TeV observables for SUSY searches [5].

samples.

The probability for the trigger to fire on an event that passes the baseline selection is greater than 97%. The baseline event selection can be summarized as follows. Events are accepted if they have

- no reconstructed, isolated lepton with a $p_T > 10$ GeV and $|\eta| < 2.4$, where the isolation is as defined in Equation 8.15;
- no reconstructed, isolated particle track with a $p_T > 10$ GeV and $|\eta| < 2.4$;
- $H_T > 500$ GeV;
- $H_T^{\text{miss}} > 200$ GeV;
- $N_{\text{jets}} \geq 4$, where jets are required to have a $p_T > 30$ GeV and $|\eta| < 2.4$, and
- $\Delta\phi(H_T^{\text{miss}}, \text{jet}_{1,2,3,4}) > 0.5, 0.5, 0.3, 0.3$.

After this selection, events are further subdivided into 72 exclusive bins in a four-dimensional array of H_T^{miss} , the number of jets, the number of tagged bottom quark jets, and H_T . Figure

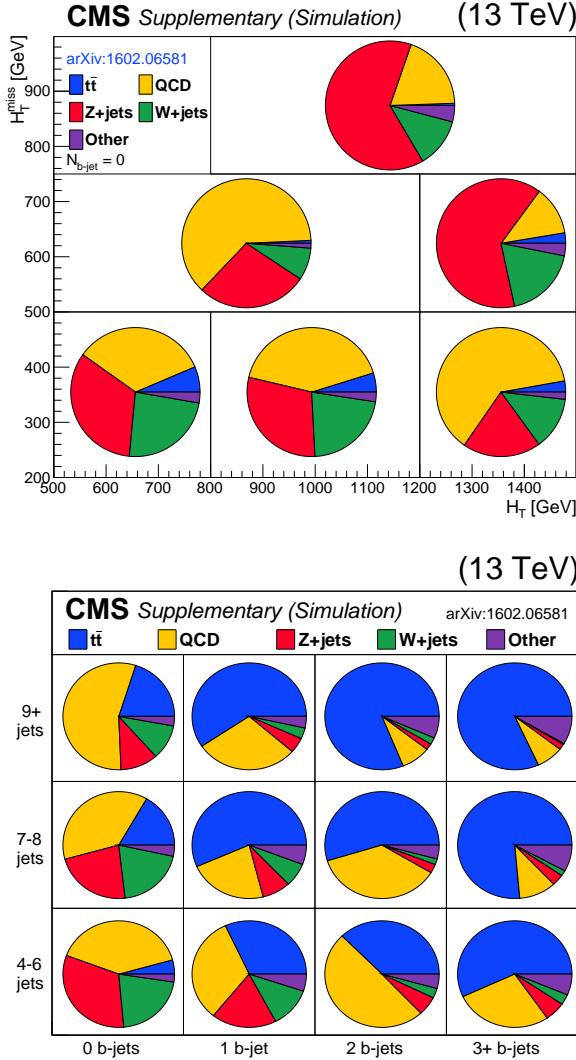


Figure 8.28: The signal region boundaries in the planes of $H_T^{\text{miss}} - H_T$ (top) and $N_{\text{jets}} - N_{\text{b-jets}}$ (bottom) for the multi-jet + H_T^{miss} search. The pie charts represent the relative contributions to the Standard Model backgrounds in the region $N_{\text{jets}} = 4-6$, $N_{\text{b-jets}} = 0$ (top), and $H_T^{\text{miss}} = 200-500$ GeV, $H_T = 500-750$ GeV (bottom).

8.28 shows the boundaries of 6 bins in the $H_T - H_T^{\text{miss}}$ plane, and 12 bins in the plane of $N_{\text{jets}} - N_{\text{b-jets}}$. The 72 bins included in this search are defined by each unique combination of a box in the $H_T^{\text{miss}} - H_T$ plane and a box in the $N_{\text{jets}} - N_{\text{b-jets}}$ plane. The bin numbering scheme is given in Appendix B.

Background estimation. The relative contribution of each standard model background in selected regions is also shown in Fig. 8.28. The $t\bar{t}$, $W+jets$, and $Z+jets$ backgrounds are estimated by data-driven methods detailed in [6], but are not discussed thoroughly here. The QCD background is largest in regions of low H_T^{miss} and high H_T , and its bin-by-bin estimation is detailed here.

The QCD background is estimated using two independent data-driven methods, one making use of a control region defined analogously to the baseline selection but with the $\Delta\phi$ requirement inverted, and by the rebalance and smear method described in Section 8.2. The two methods are considered independent because the predictions are derived from nearly disjoint data control samples, and because they use different methodology.

In the $\Delta\phi$ method, the counts in the inverted $\Delta\phi$ control region are related to the counts in the signal region by factors derived using a combination of input from real and simulated data. The information from the real data comes from events in the least sensitive search bins so as to minimize potential signal contamination. In this method, other standard model backgrounds are subtracted using estimates derived by the data-driven methods employed in the search. The largest sources of uncertainty come from the estimates of the standard model backgrounds that are subtracted. Further details about the inverted $\Delta\phi$ method are given in the literature.

The bin-by-bin prediction based on the rebalance and smear method, along with the systematic uncertainties described in Section 8.2, are shown in Fig. 8.29, with the predictions and uncertainties based on the inverted $\Delta\phi$ method superimposed. The uncertainties in the rebalance and smear counts are derived based on the methodology described in Section 8.2. For rebalance and smear, the largest uncertainty in bins with a large count is the uncertainty in the jet energy resolution, ranging from 10–30%. In bins with exceptionally small predicted counts, the statistical uncertainty is dominant, ranging from 20–40%. In bins with three or more b-tagged jets, the non-closure uncertainty is the largest, at 50%.

No evidence of inconsistency is seen between the two predictions. The distribution of the fractional differences in the predictions of the two methods, divided by the uncertainty in the fraction (significance), is shown in Fig. 8.30 (a). After fitting the significance distribution

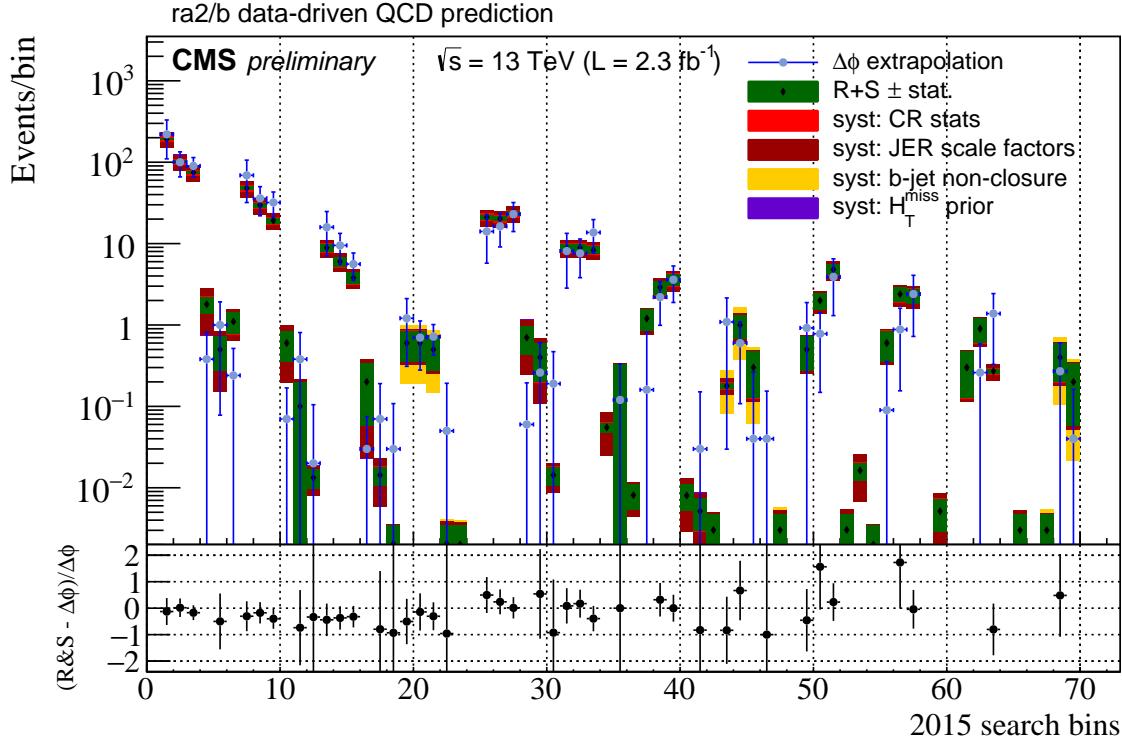


Figure 8.29: The QCD prediction based on the inverted $\Delta\phi$ extrapolation method compared with that of the rebalance and smear method. Each type of uncertainty in the rebalance and smear prediction appears as a color-filled rectangle centered on the predicted value, summed in quadrature with the inner uncertainties. Uncertainties in the inverted $\Delta\phi$ prediction are the statistical systematic uncertainties reported in [6], summed in quadrature.

with a Gaussian function, as shown in the Figure, the best fit values for the mean and standard deviation indicate consistency between the two predictions. Fig. 8.30 (b) shows the distribution of fractional uncertainties in the inverted $\Delta\phi$ and rebalance and smear predictions. It is seen that the inverted $\Delta\phi$ method has a prediction of zero for 22 of the search bins, whereas the rebalance and smear method gives a non-zero prediction for all but four bins. The fractional uncertainties in the rebalance and smear predictions are, overall, much smaller than those of the inverted $\Delta\phi$ predictions. This is the case in all 72 bins, except for bin 21, where the 50% non-closure uncertainty in the rebalance and smear prediction renders an overall larger uncertainty. The bi-modal shape of the rebalance and smear distribution in Fig. 8.30 (b) is due to the b-tag non-closure systematic uncertainty;

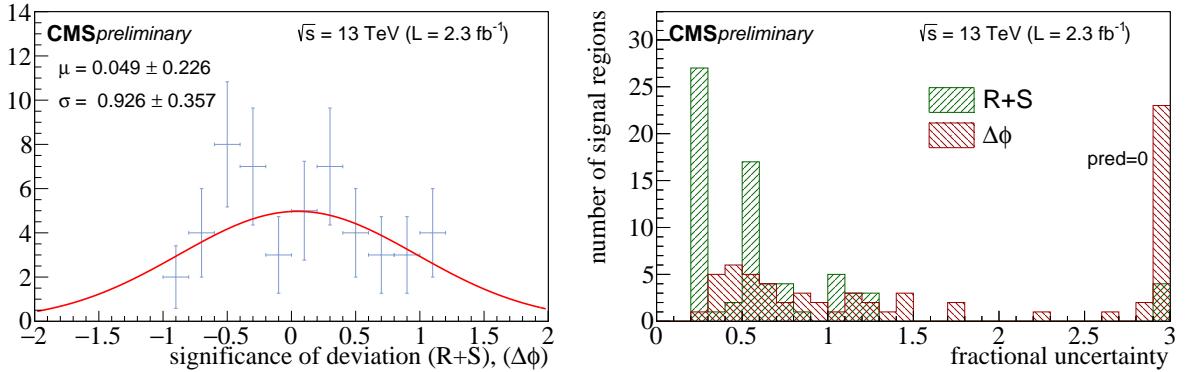


Figure 8.30: Left: the distribution of the significance of the deviation between the predictions from the inverted $\Delta\phi$ and rebalance and smear methods. Right: the distribution of the fractional uncertainties in the counts predicted by the two methods. The upper overflow bin is populated by bins for which the prediction is identically 0.

this uncertainty is responsible for the peak around 0.5. Bin 35 was found to have an error in its kinematic definition, and so the prediction from the inverted $\Delta\phi$ method has been used in its stead.

Results and interpretation. The observed numbers of events in the 72 search bins are shown in the Table in Appendix B, along with the summed predictions for the SM backgrounds. The predicted background is observed to be compatible with the data in all 72 regions, within uncertainties. Therefore, I do not observe evidence for new physics. These results are interpreted in the context of the simplified models shown in Fig. 8.25.

These results are interpreted in terms of limits on the signal cross section as a function of the gluino and LSP mass. A likelihood function \mathcal{L} is constructed as the product of Poisson probability functions, one for each search bin, with various nuisance parameters accounting for the yields of the four background classes and the signal strength modifier. \mathcal{L} also contains log-normal density functions that constrain the background yields by their estimated uncertainties. Uncertainties in the signal yield are associated with the renormalization and factorization scales, ISR, the jet energy scale, the b-tagging efficiency, and statistical uncertainties based on the simulated sample sizes, and are determined as functions of $m_{\tilde{\chi}_1^0}$ and $m_{\tilde{g}}$.

The test statistic is $q_\mu = -2 \ln (\mathcal{L}_\mu / \mathcal{L}_{\max})$, where \mathcal{L}_{\max} is the maximum likelihood determined by a best fit to all parameters, including the signal strength modifier, and \mathcal{L}_μ is the maximum likelihood for a fixed signal strength. Asymptotic formulae [111] and the CL_s [112, 113] are used to set limits.

Upper limits on the sparticle masses are identified as the set of points for which the 95% confidence level upper limit on the cross section is equal to the SUSY signal cross section computed at NLO+NLL. Expected limits are derived by evaluating the test statistic assuming Poisson fluctuations around the predicted numbers of background events.

The limits are shown in Figs. 8.31–8.34. For a massless LSP, gluinos are excluded with masses below 1460, 1550, 1600, and 1470 GeV, for the T1qqqq, T1tttt, T1bbbb, and T5VV scenarios. The bands on the observed exclusion curves show the effect of varying the signal cross section by changing the renormalization and factorization scales by a factor of 2 and using the PDF4LHC recommendation [114] for the PDF uncertainty. This shows the sensitivity of the limits to uncertainties in the signal cross section.

The expected limits based on the rebalance and smear prediction are comparable to those based on the inverted $\Delta\phi$ extrapolation. Observed limits on the gluino and LSP masses in the T1qqqq, T1tttt, and T5VV models are stronger for rebalance and smear by 20–50 GeV, in the compressed region. In the uncompressed region, the T1qqqq and T5VV limits are stronger by 10 and 50 GeV with the rebalance and smear prediction. In all other comparisons, the observed mass limits are nearly indistinguishable. The difference in color between left and right plots shows the comparison in the upper limit on the signal cross section between the two methods, throughout the $m_{\tilde{g}} - m_{\tilde{\chi}_1^0}$ plane. In the compressed region of the T1tttt model, the limits using input from the rebalance and smear prediction represent the strongest limits of any published CMS result, at the time of this writing. The same is true for the T5VV model, throughout the $m_{\tilde{g}} - m_{\tilde{\chi}_1^0}$ plane.

These limits significantly surpass those that were set at $\sqrt{s} = 8$ TeV, for which the corresponding limits are around 1150 GeV [115, 107] for the three T1 models and 1280 GeV [107] for the T5VV model.

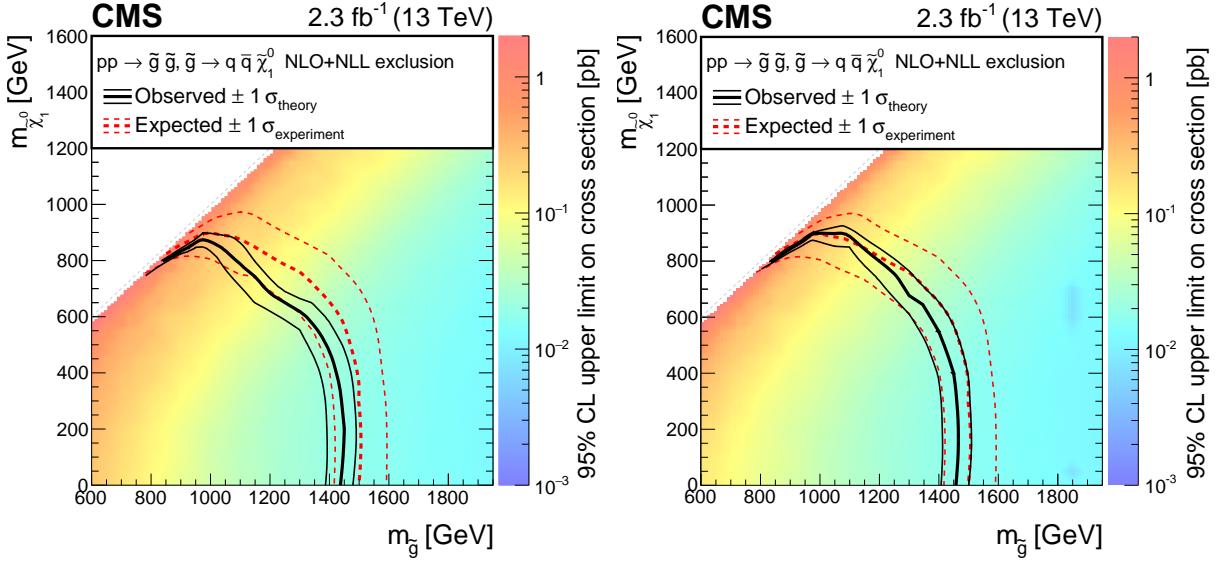


Figure 8.31: Observed (black contour) and expected (red contour) limits on the masses of the gluino and LSP using the QCD estimate based on the inverted $\Delta\phi$ method (left) and the rebalance and smear method (right), in the context of the T1qqqq model. The dashed (grey) lines indicate the $\tilde{\chi}_1^0 = m_{\tilde{g}}$ diagonal.

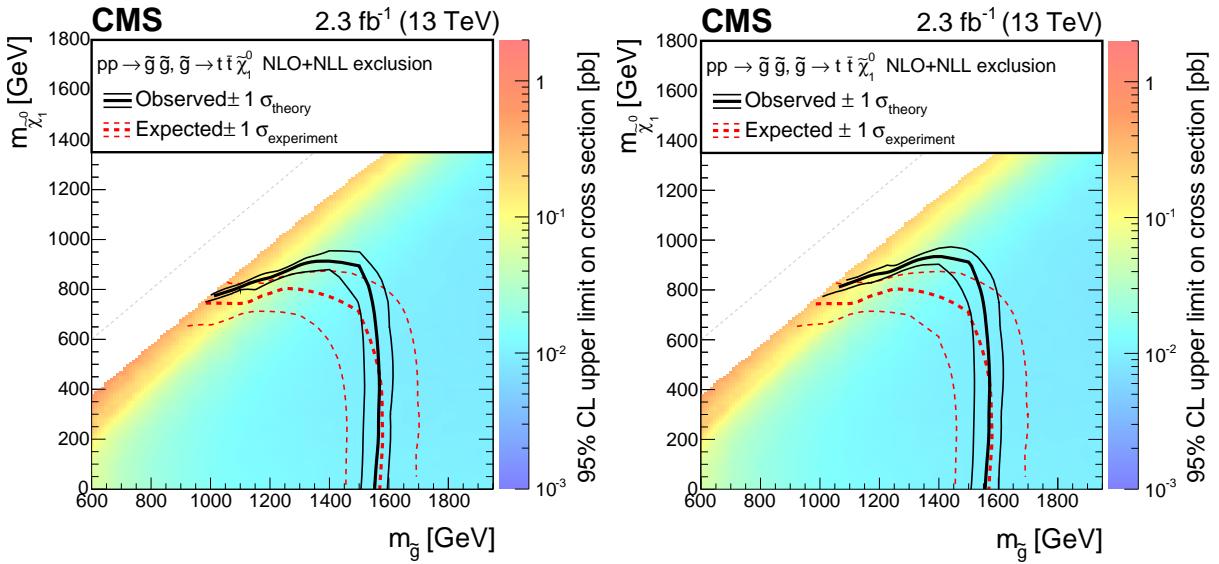


Figure 8.32: Observed (black contour) and expected (red contour) limits on the masses of the gluino and LSP using the QCD estimate based on the inverted $\Delta\phi$ method (left) and the rebalance and smear method (right), in the context of the T1tttt model.

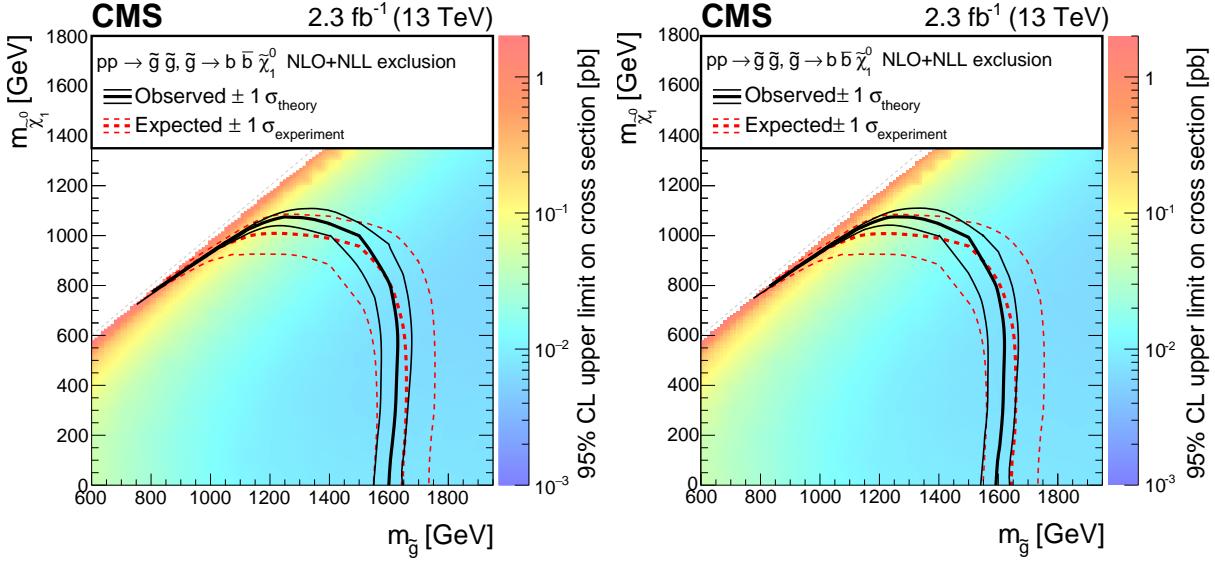


Figure 8.33: Observed (black contour) and expected (red contour) limits on the masses of the gluino and LSP using the QCD estimate based on the inverted $\Delta\phi$ method (left) and the rebalance and smear method (right), in the context of the T1bbbb model.

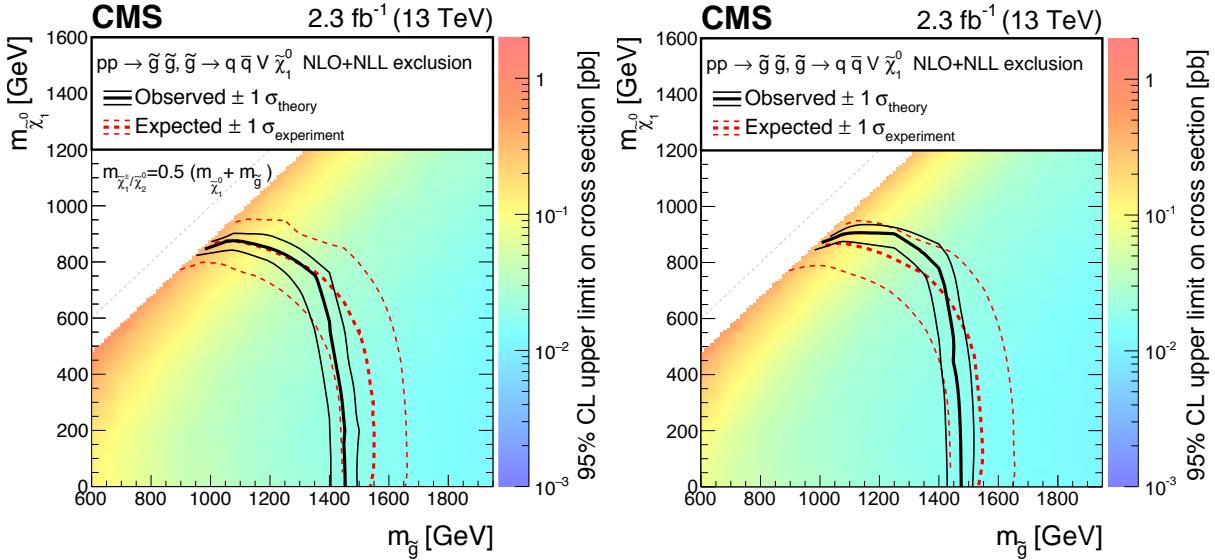


Figure 8.34: The observed (black contour) and expected (red contour) limits on the masses of the gluino and LSP using the QCD estimate based on the inverted $\Delta\phi$ method (left) and the rebalance and smear method (right), in the context of the T5VV model.

8.4.2 Search for SUSY in events with top-tagged jets at $\sqrt{s} = 13$ TeV

This search [7] looks for evidence of SUSY in events with jets identified as top quarks. An excess of events containing top-tagged jets (top tags) over the standard model background may be evidence of the production of top squark pairs, whose existence has been motivated as a solution to the hierarchy problem (Section 3.6). The possibility of top squarks with masses below about 450 GeV has largely been ruled out by results from other experiments (Chapter 7). However, considerations of naturalness implore us to look for evidence of top squarks nonetheless. The analysis is optimized to be sensitive to signatures resembling the simplified models shown in Fig. 8.35.

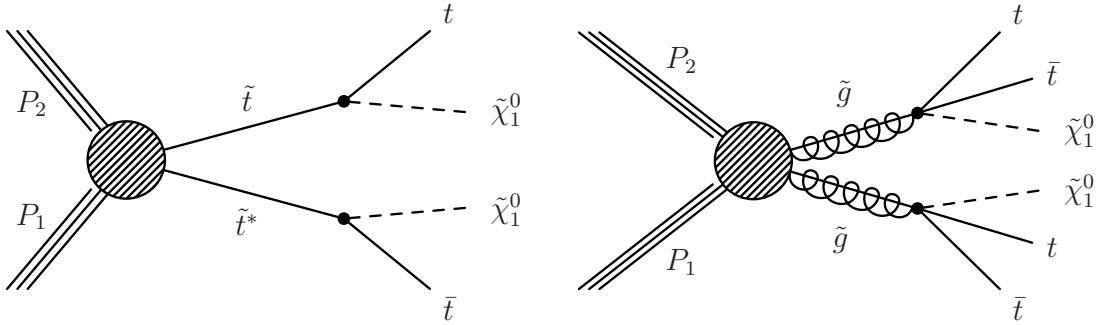


Figure 8.35: The simplified models used for the optimization and interpretation of the hadronic search for SUSY in events with top-tagged jets. They are T2tt (left) and T1tttt (right).

In addition to the multiplicity of top tags, other observables are used for signal/background discrimination, including the H_T , E_T^{miss} , N_{jets} , $N_{\text{b-jets}}$, and the transverse mass M_{T2} [86]. Top tags are constructed by considering all three-jet combinations in an event, and selecting a combination with an invariant mass consistent with the top quark mass. Additional topological requirements are placed on the top-tagged jet to reject background from QCD jet combinations. After identifying a top-tagged jet, the remaining jets in the event are analyzed and checked for consistency with a second top quark in the event, inferred by the presence of a b-tagged jet. In the event that both systems (the top-tagged system and the

remnant b-tagged jet) are reconstructed, the $M_{\text{T}2}$ is computed taking the four-vectors of the two top candidates as input, along with the $E_{\text{T}}^{\text{miss}}$ of the event.

Events are collected using the same hadronic trigger used in the work of the previous section,

- HLT_PFHT350_PFMET100_*

The probability for the trigger to accept an event that passes the baseline selection is greater than 98%. For the baseline selection, events are selected if they have

- no reconstructed, isolated muon with a $p_{\text{T}} > 10 \text{ GeV}$, $|\eta| < 2.4$, and isolation < 0.2 ;
- no reconstructed, isolated electron with a $p_{\text{T}} > 10 \text{ GeV}$, $|\eta| < 2.5$, and isolation < 0.1 ;
- no reconstructed, isolated lepton (hadron) track with a $p_{\text{T}} > 5$ (10) GeV , $|\eta| < 2.5$, and isolation 0.2;
- $H_{\text{T}} > 500 \text{ GeV}$;
- $E_{\text{T}}^{\text{miss}} > 200 \text{ GeV}$;
- $\Delta\phi(H_{\text{T}}^{\text{miss}}, \text{jet}_{1,2,3}) > 0.5, 0.5, 0.3$;
- $N_t \geq 1$, where N_t is multiplicity of top tags;
- $N_{\text{b-jets}} \geq 1$, and
- $M_{\text{T}2} > 200 \text{ GeV}$.

After the baseline selection, events are further subdivided into 37 search bins defined by rectangular boundaries in the dimensions of $E_{\text{T}}^{\text{miss}}$, $M_{\text{T}2}$, $N_{\text{b-jets}}$, and N_t , with boundaries and bin numbers shown in Fig. 8.36.

The most significant background comes from the SM $t\bar{t}$ production or W-boson production in association with jets, where the W decays into leptons that are either not accepted, not reconstructed, or not isolated. The next largest contribution comes from $Z \rightarrow \nu\bar{\nu}$ production in association with jets, including b-jets, in which the neutrino pair gives rise to large $E_{\text{T}}^{\text{miss}}$ and the top quark conditions are satisfied by an accidental combination of the jets. The QCD multi-jet contribution and the contribution from other rare SM processes are subdominant across all search bins.

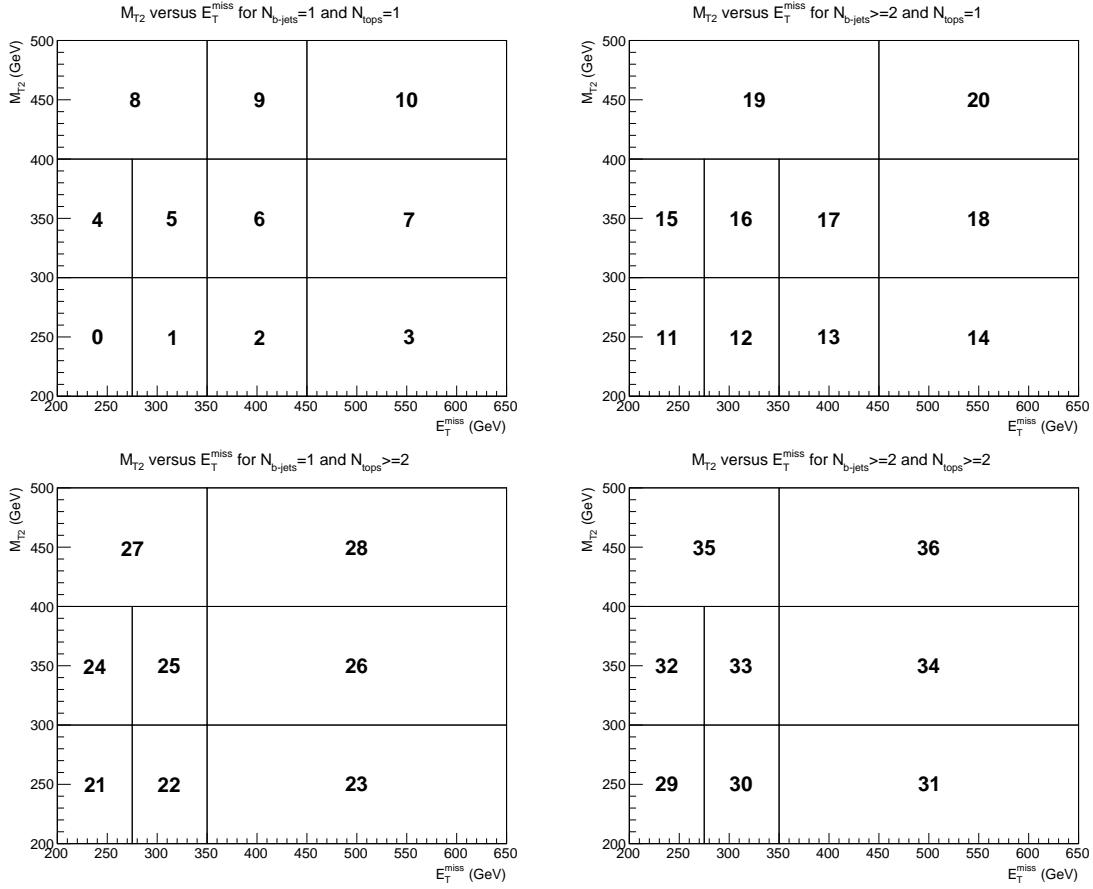


Figure 8.36: Search bin definitions after baseline selection cuts.

Background estimation. The $t\bar{t}$, $W+jets$, and QCD backgrounds are estimated by data-driven methods detailed in [7], but are not discussed thoroughly here. The $Z \rightarrow \nu\bar{\nu}$ background is most dominant in regions of small H_T and large E_T^{miss} , and its bin-by-bin estimation is detailed here.

The $Z \rightarrow \nu\bar{\nu}$ background is estimated using the ee and $\mu\mu$ samples as a proxy for the true sample in which the Z decays into neutrinos. The method used to derive the prediction, as well as uncertainties, is that described in Section 8.3. The values and uncertainties in the bin-by-bin prediction are shown in Fig. 8.37 for the $\mu\mu$ -derived prediction, the ee-derived prediction, and the prediction based on the combined $\mu\mu+ee$ sample. The $\mu\mu$ -derived and ee-derived predictions are compatible within the uncertainties, but the ee-derived prediction

is systematically larger than that of the $\mu\mu$ -derived prediction by at least 40%. This is due to the difference in the normalization factor-derived in the two methods, as well as differences in the data-simulation weights.

Results and interpretation. The predicted number of SM background events and the number of events observed in data for each of the search bins are shown in Figure 8.38 and Tabulated in Appendix B. The predicted background counts are observed to be compatible with the data in all 37 regions, within uncertainties. Therefore, no evidence of new physics is observed. These results are interpreted in the context of the simplified models shown in Fig. 8.25.

The results are interpreted in terms of the simplified model scenarios shown in Fig. 8.35 using an analogous limit setting procedure discussed in Section 8.4.1. Uncertainties in the signal yields are similar to those discussed in Section 8.4.1, with additional uncertainties associated with differences in the top-quark tagging efficiency and false positive rate between the CMS fast and full simulations, on the order of a few percent.

Upper limits on the observed cross section for the signal models T2tt and T1tttt are shown in Figs. 8.39 and 8.40, along with contours corresponding to the expected (red) and expected (black) upper limits on the SUSY masses, assuming the NLO+NLL theoretical SUSY cross sections. In the case of T2tt, limits are shown assuming the predicted $Z \rightarrow \nu\bar{\nu}$ counts based on the $\mu\mu$, ee, and $\mu\mu+ee$ samples. For T1tttt, just the limits on the $\mu\mu$ and $\mu\mu+ee$ samples are shown.

In the case of T2tt, the ee-derived expected limit is somewhat less constraining than that of the expected muon-derived limit, as was expected from the outset. The ee-derived observed limit on the top squark mass is actually stronger than the $\mu\mu$ -derived observed limit by about 20 GeV, for small values of the LSP mass. This difference is due to the considerably larger ee-derived $Z \rightarrow \nu\bar{\nu}$ prediction. The combined $\mu\mu+ee$ -derived expected limit is comparable to the $\mu\mu$ -derived expected limit. The observed limit based on the combined result is slightly more constraining on the top squark mass than the $\mu\mu$ result, by about 10 GeV, for small LSP masses. The largest observed upper limit on $m_{\tilde{t}}$ for low LSP mass comes from the ee estimate, and is approximately 780 GeV, coinciding with the

expected limit. In the compressed region, the strongest observed limit comes from the ee prediction, where LSPs are excluded with masses below about 180 GeV, which is nearly in agreement with the expected limit.

In the T1tttt model, only negligible differences appear between the $\mu\mu$ and the combined $\mu\mu+ee$ predictions. Given the limited constraining power of the ee control sample, this sample be better used in more creative ways, for example, as a training sample for a multivariate discriminant. This is explored in the next chapter.

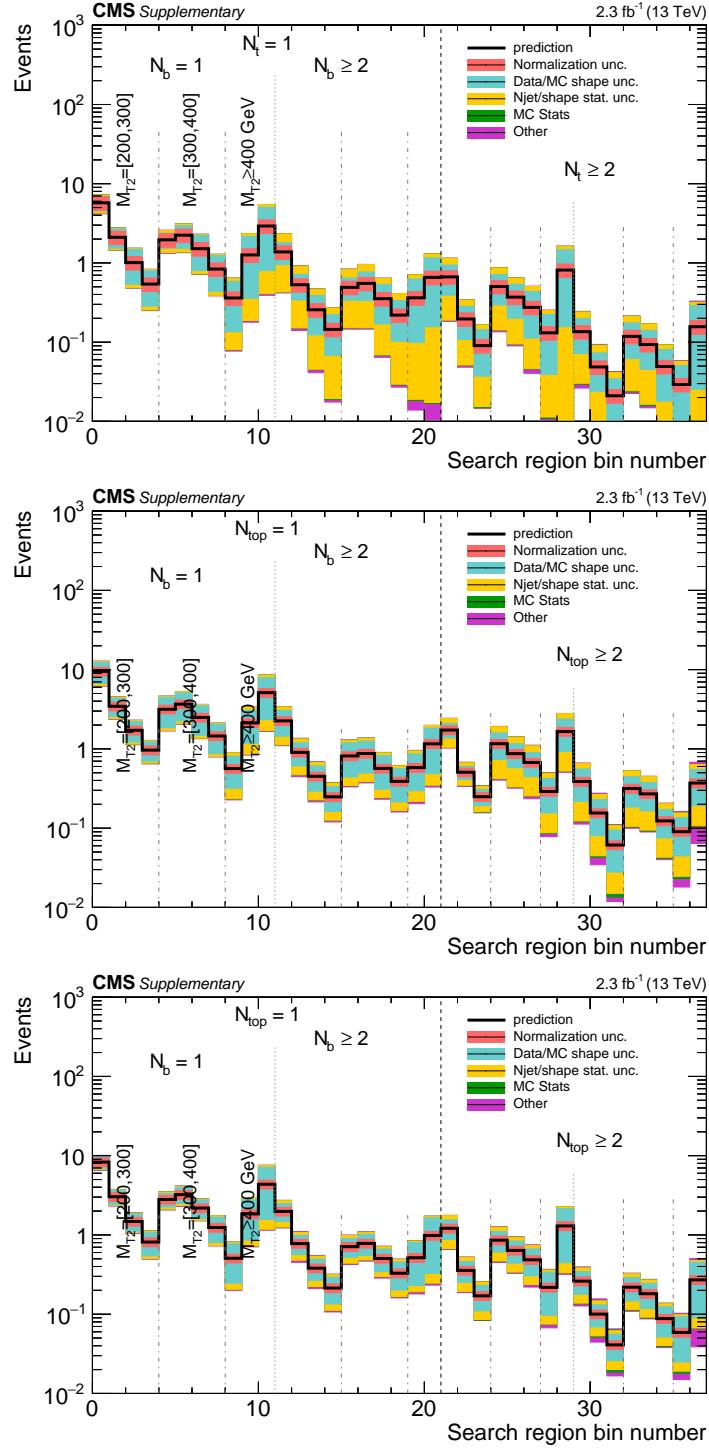


Figure 8.37: The predicted counts originating from standard model $Z \rightarrow \nu\bar{\nu}$ production, based on the $\mu\mu$ sample (top), the ee sample (center), and the combined $\mu\mu+ee$ sample (bottom). Each type of uncertainty, described in Section 8.3, appears as a color-filled rectangle centered on the predicted value, summed in quadrature with the inner uncertainties.

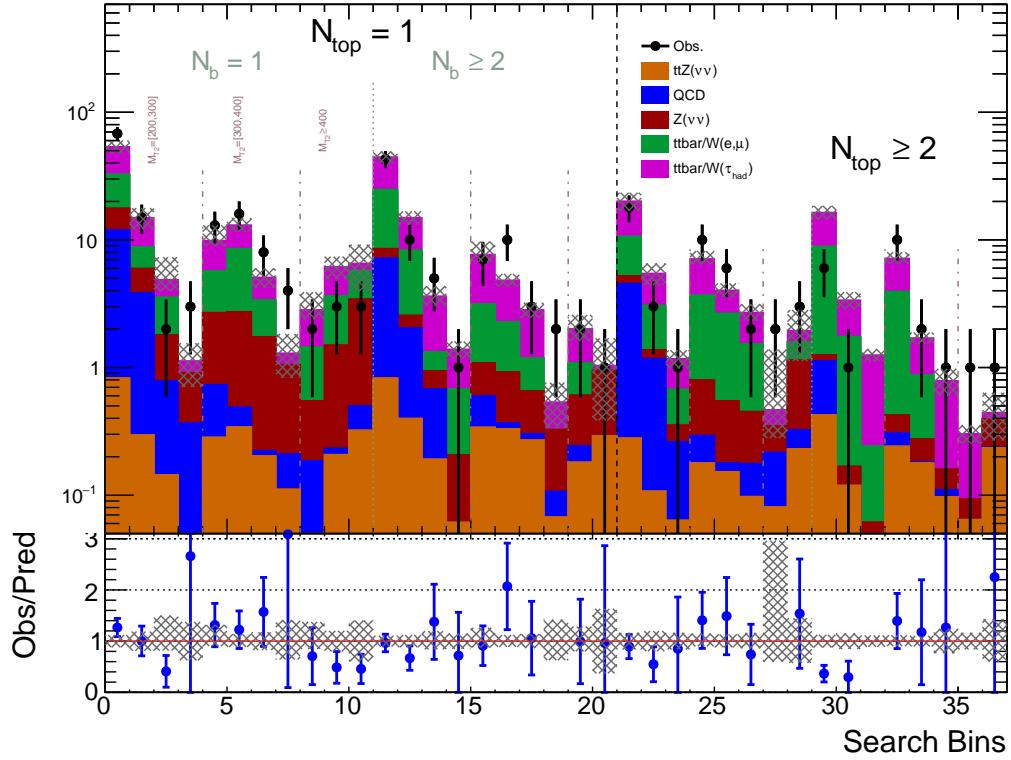


Figure 8.38: Data are shown as black points. The total predictions are shown in filled solid area. The bottom plot shows the ratio of data over total background prediction in each search bin. Only statistical uncertainties are propagated to the ratio. The numbering scheme is given in Ref. [7].

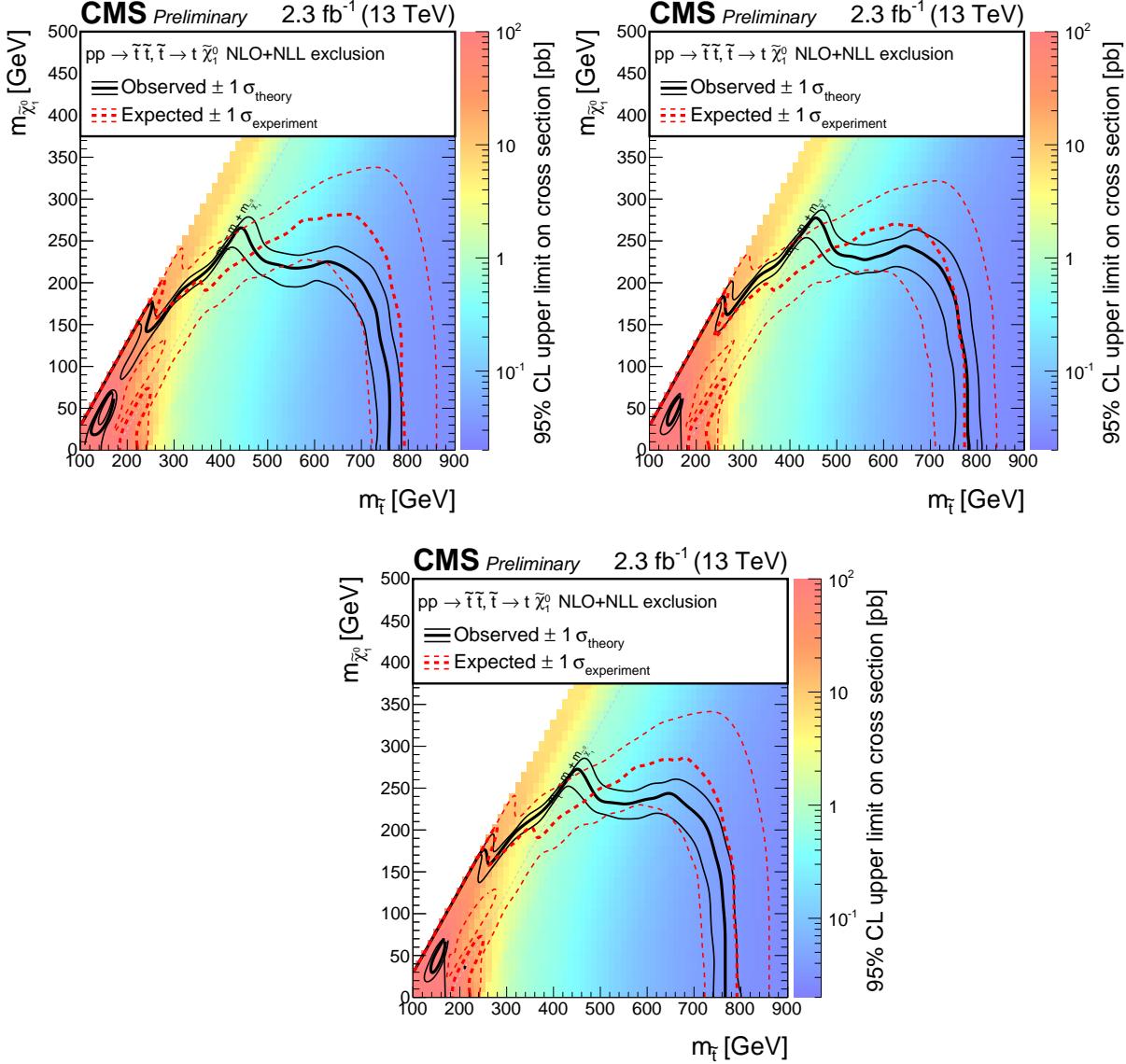


Figure 8.39: Observed (black contour) and expected (red contour) limits on the top squark and LSP masses in the context of the T2tt model. The top-left plot corresponds to the $\mu\mu$ -derived $Z \rightarrow \nu\bar{\nu}$ prediction, the top-right to the ee-derived prediction, and the bottom plot to the prediction of the combined $\mu\mu + \text{ee}$ sample.

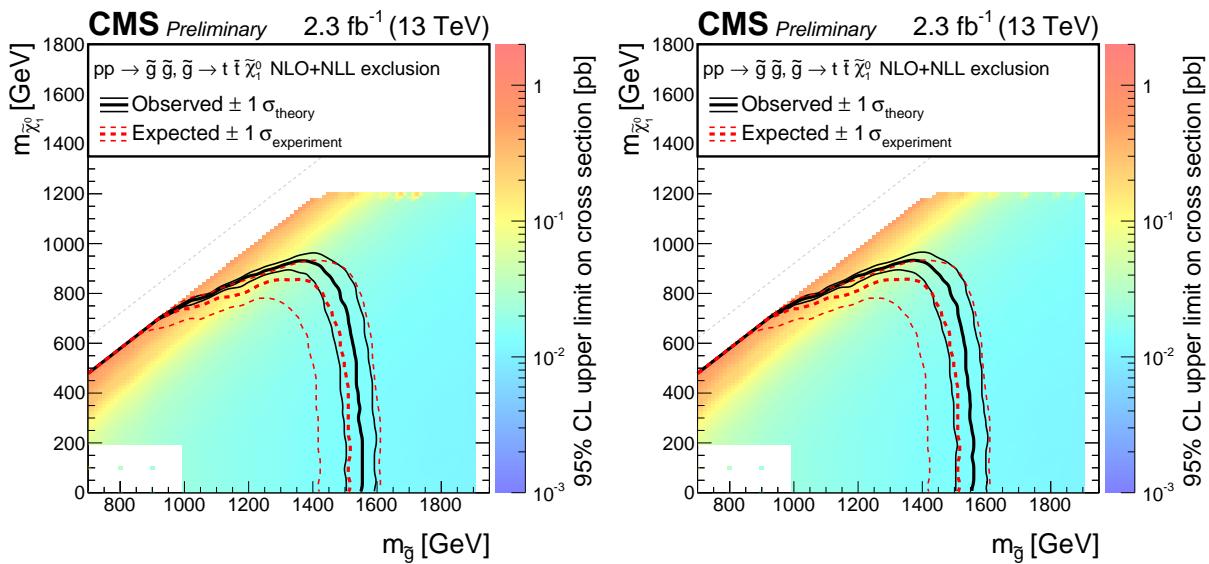


Figure 8.40: Observed (black contour) and expected (red contour) limits on the gluino and LSP masses in the context of the T1tttt model. The left plot corresponds to the $\mu\mu$ -derived prediction for the $Z \rightarrow \nu\bar{\nu}$ counts, and the right plot to the muon+ee-derived prediction.

CHAPTER 9

TARGETING THE NON-EXCLUDED PMSSM WITH MULTIVARIATE DISCRIMINANTS

Designing analyses to target the non-excluded regions of the pMSSM is challenging for a number of reasons. The fact that the signal events in this subset contain soft objects and low E_T^{miss} means that the observables typically used to distinguish between Standard Model backgrounds and the signal no longer function as excellent discriminators.

However, the situation is not completely dire. The H_T , E_T^{miss} , and the prevalence of high p_T objects are popular largely because they exploit well-understood differences between typical new physics signals and background. There are presumably other event characteristics that differ between events arising from supersymmetry and those from standard model processes, such as the relative angles between particles, the presence and characteristics of leptons and heavy-flavor quarks in the events, as well as various combinations of the four-vector components of particles. Information not considered in most searches, but to which we have ready access, include the p_T , η , and ϕ of particles in the event. Correlations among these observables may serve to distinguish between standard model events and BSM events.

In the next section, the methods detailed in this document are combined into an outline for a set of possible strategies for targeting non-excluded SUSY models that predict large cross sections in the most challenging all-hadronic kinematic regions: those of low E_T^{miss} , where QCD is the dominant background, and low H_T , where $Z \rightarrow \nu\bar{\nu}$ events are the dominant background. Multivariate methods are used to build discriminants that increase the sensitivity to supersymmetry models in these regions. Then, Section 9.2 presents a preliminary analysis that has been carried out following one such strategy.

9.1 Methods for exploring difficult pMSSM points

Can our sensitivity to the pMSSM be increased through the use of multivariate discriminants? in order to proceed, answering the following questions is key:

1. Does the set of particle four-vectors in an event contain additional information that distinguishes signal from background events?
2. Can this information be accurately modeled in real and simulated events?
3. Is any such discrimination power generalizable to a large class of interesting models, or is it limited to a narrowly-defined subset of models or even a single model?

I set out to answer these questions by considering events in two background-dominated kinematic signal regions, corresponding to modified versions of the baseline selection of the CMS multi-jet + H_T^{miss} search [6]. The first signal region is taken to be the low- H_T^{miss} sideband of the baseline selection [6] where the range $H_T^{\text{miss}} = 150\text{--}200 \text{ GeV}$ is considered, and the second is the taken to be the low- H_T sideband, such that events in the range $H_T = 400\text{--}500 \text{ GeV}$ are considered. In each signal region, a separate multivariate discriminant is trained and tested, making use of the data-driven background estimation methods in two ways: first, to obtain real events for the background training sample, and second, to obtain an independent set of events for testing.

9.1.1 Low- H_T^{miss} discriminant

The kinematic region defined by the range $H_T^{\text{miss}} = 150\text{--}200 \text{ GeV}$ is largely dominated by QCD multi-jet events. Therefore, I focus on rejecting QCD events while accepting hypothetical SUSY using the following strategy.

- Create a pure sample of QCD events using the data-driven QCD generator based on the rebalance and smear method (Section 8.2);
- obtain a few samples of SUSY signal events corresponding to models that are nearly excluded by the existing CMS searches (I consider SUSY models within the SMS framework, T1qqqq with $m_{\tilde{g}} = 1000 \text{ GeV}$ and $m_{\tilde{\chi}_1^0} = 800 \text{ GeV}$, T1bbbb with $m_{\tilde{g}} = 1000 \text{ GeV}$ and $m_{\tilde{\chi}_1^0} = 900 \text{ GeV}$, and T1tttt with $m_{\tilde{g}} = 1200 \text{ GeV}$ and $m_{\tilde{\chi}_1^0} = 800 \text{ GeV}$);

- train a discriminant using rebalanced and smeared events for the background training, and pooled events from the three SUSY models for the signal training events;
- observe the accuracy of the modeling of the discriminant output in a sample of simulated QCD events, and
- observe potential sensitivity to a range of signal models.

A boosted decision tree (BDT) [49] is trained using the signal and background samples, taking the p_T , η , and ϕ of each of the four leading jets, as well as the H_T^{miss} , H_T , and N_{jets} as inputs. The output of the discriminant for signal and background is shown in Fig. 9.1. The modeling of the QCD background in simulation by the rebalance and smear method is shown in Fig. 9.1, overlaid with the distribution pertaining to a few signal model points. Agreement to within a few percent is observed between the prediction and expectation, evidence that the improved rebalance and smear method can accurately model the correlations of up to 15 observables. Also shown are the distributions of the discriminant for the three chosen signal models, as well as a graph showing the tradeoff between the signal efficiency and background rejection (ROC curve). This discriminant is seen to yield significant increases in sensitivity to a broad class of models, and its output is shown to be successfully modeled using the rebalance and smear method described in this document.

9.1.2 Low- H_T discriminant

A similar procedure is applied in the low- H_T signal region. In this region, the dominant standard model background is from the production of Z bosons in association with jets, where the Z decays into neutrinos. Therefore, I focus on rejecting $Z \rightarrow \nu\bar{\nu}$ events while accepting SUSY signal events with high efficiency. Using the data-driven $Z \rightarrow \nu\bar{\nu}$ background estimation methods described in Section 8.3, and continuing the approach of using real data samples for the background training events for multivariate discriminants, the following strategy is used.

- Obtain a $Z \rightarrow \nu\bar{\nu}$ prediction sample based on a real sample of $Z \rightarrow \mu^+\mu^-$ events that have been cleaned of muons using the procedure described in Section 8.3;

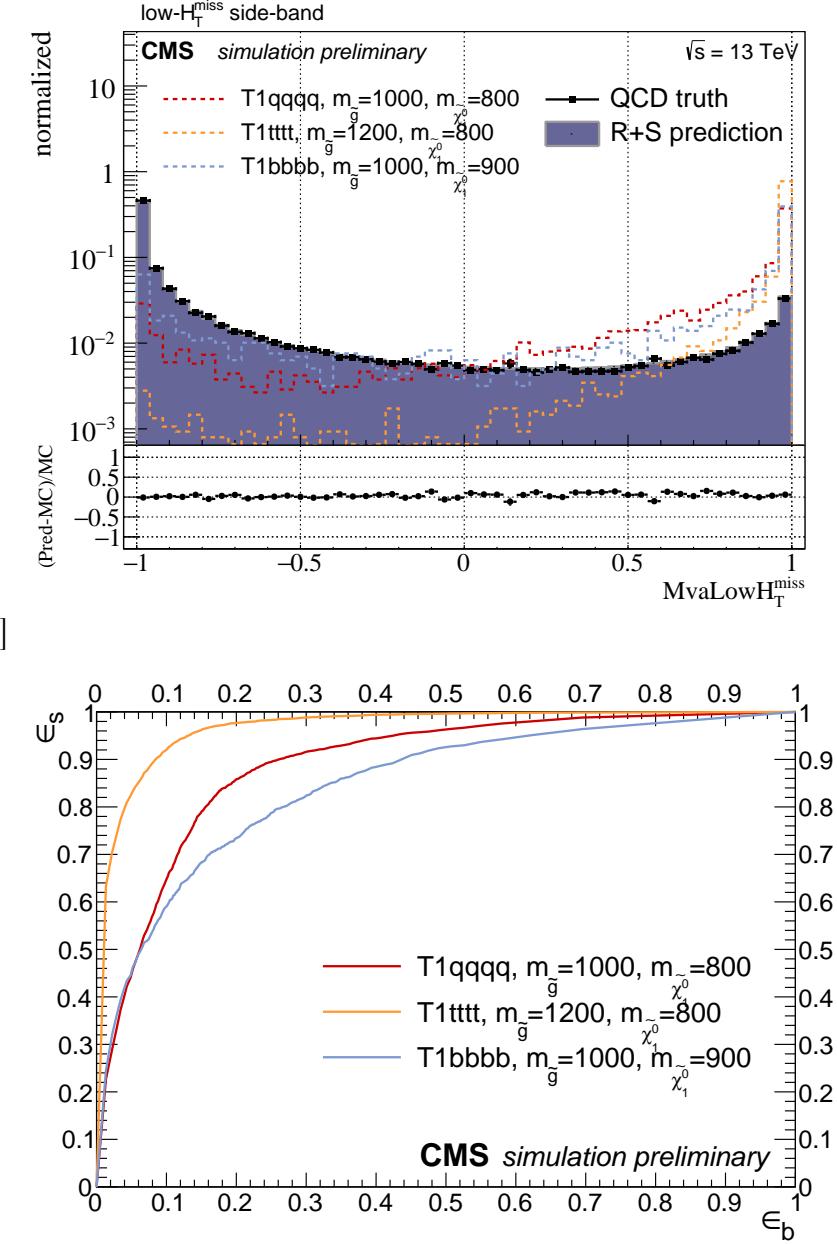


Figure 9.1: Top: the distribution of the output of the low- H_T^{miss} data-driven discriminant for a sample of simulated QCD events, compared with the distribution obtained by applying the rebalance and smear method to simulation. Also shown are the distributions of the output for a few signal points corresponding to various SMS scenarios. Bottom: the efficiency of selecting signal events as a function of the efficiency for selecting background events, based on a scan through all possible thresholds on the discriminant.

- use the same samples of SUSY signal events considered for the low- H_T^{miss} exercise;
- train a BDT using the $Z \rightarrow \nu\bar{\nu}$ prediction sample for the background events, and the pooled events from the SUSY model points for the signal events;
- observe the accuracy of the modeling of the discriminant output by comparing the data-driven prediction applied in simulation using dielectron sample with the output obtained directly from $Z \rightarrow \nu\bar{\nu}$ simulation, and
- observe potential sensitivity to the signal models.

The modeling of the discriminant distribution for a simulated sample of $Z \rightarrow \nu\bar{\nu}$ events is shown in Fig. 9.2, along with the distribution for the sample obtained by applying the dielectron prediction method to $Z \rightarrow e^+e^-$ simulation. Overlaid are distributions of the chosen signal model points. Once again, significant sensitivity gains are achieved across a range of SUSY models, and the discriminant is accurately modeled by the data-driven method developed herein.

9.2 Proof of principle analysis

As a proof of principle, I now present an analysis that follows the second approach described in Section 9.1.2. Rather than targeting signal events corresponding to SMS models, however, the non-excluded regions of the pMSSM are used to design the search. A benchmark pMSSM point is chosen where the gluino and LSP masses are 644 and 400 GeV, respectively. The principal physics process corresponding to this point is that shown in Fig. 7.20 (b), in which gluinos are pair produced, each gluino decays into a $q\bar{q}$ pair and a 400 GeV chargino, and the chargino decays into a W boson and an LSP. This model point survived the run 1 CMS searches, and is referred to as survivor 1 in the figures. Fig. 9.3 shows distributions of key observables predicted by the benchmark signal scenario, normalized to a luminosity of 2.3 fb^{-1} . The H_T distribution, which peaks at 350 GeV, illuminates why this model was not excluded in run 1. The 7 and 8 TeV hadronic analyses imposed a threshold of 500 GeV on the H_T as part of the baseline selection. I note that the same selection is applied in the 13 TeV hadronic searches, which likely have also not excluded this point.

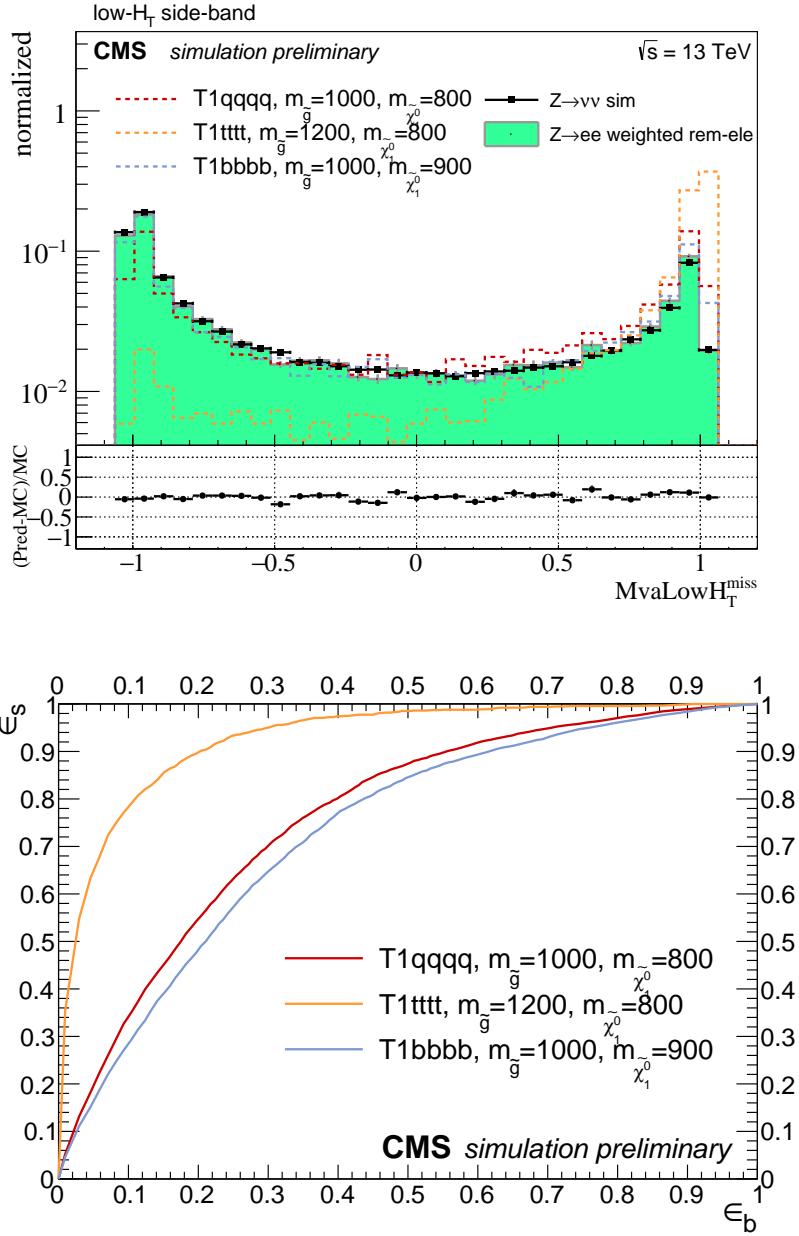


Figure 9.2: Top: The distribution of the output of the low- H_T muon data-driven multivariate discriminant for a sample of simulated $Z \rightarrow \nu\bar{\nu}$ events, compared with the distribution obtained by applying the data-driven dielectron prediction method to simulation. Also shown are the distributions of the output for a few signal points corresponding to various SMS scenarios. Bottom: the efficiency of selecting signal events vs that of selecting background events, based on a scan through all possible thresholds on the discriminant.

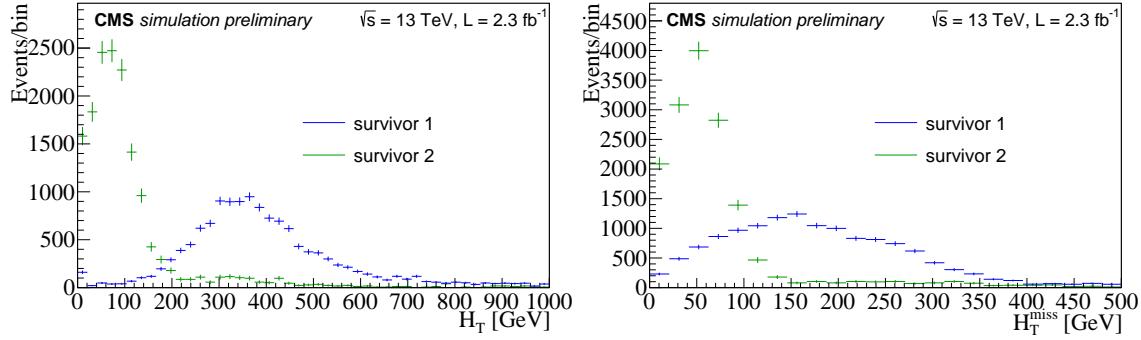


Figure 9.3: Kinematic distributions of selected low-mass gluino pMSSM points. The benchmark pMSSM point, referred to as survivor 1, is shown in blue.

Figure 9.3 suggests that a good choice of signal region is a low-\$H_T\$ sideband of the CMS multi-jet + H_T^{miss} search [6]. I choose a baseline selection:

- $400 < H_T < 500 \text{ GeV}$;
- $H_T^{\text{miss}} > 200 \text{ GeV}$;
- $N_{\text{jets}} \geq 4$, where jets are required to have a $p_T > 30 \text{ GeV}$ and $|\eta| < 2.4$, and
- $\Delta\phi(H_T^{\text{miss}}, \text{jet}_{1,2,3,4}) > 0.5, 0.5, 0.3, 0.3$,

which applies to all control regions and the signal baseline region. Events in the signal baseline region are further required to satisfy

- $N_{\text{lep}} = 0$, with lepton $p_T > 10 \text{ GeV}$ and $|\eta| < 2.4$, and isolation defined analogously with [6];
- $N_{\text{b-jets}} = 0$.

Events are collected using the monojet trigger featured in Section 8.1. Events are required to have both an online H_T^{miss} and E_T^{miss} in excess of 90 GeV. Because there is no online threshold on the H_T , events satisfying the baseline selection are accepted by the monojet trigger with nearly 100% probability.

A BDT is trained using events from excluded pMSSM points for the signal training sample, and a set of $Z \rightarrow \nu\bar{\nu}$ proxy events for the background training sample, where the

proxy events are those used to train the BDT in the previous section. Approximately 27,000 signal events and 6,000 background events are used in the training, choices dictated by the availability of events. The 15 inputs to the BDT are the H_T^{miss} , ϕ^{miss} , H_T , N_{jets} , and the p_T , η , and ϕ of the four leading jets in an event, and the output is a single-valued discriminant. The discriminant ranges from -1 to 1 , where -1 corresponds to more background-like events, and $+1$ to more signal-like events. The range is divided evenly into 10 bins. The left-most bin is seen to receive a negligible contribution from the signal, and so is used as a normalization control region. The left 9 bins are designated as signal regions labeled, from left to right, “SR1”,..., “SR9”.

The principal sources of background events are those of the multi-jet + H_T^{miss} search [6]. These include events in which a Z boson is produced and decays into neutrinos ($Z \rightarrow \nu\bar{\nu}$), events in which jets are produced in association with a W boson that decays into leptons that are not identified (W+jets), events in which a top quark pair is produced in association with jets and final-state leptons are not identified ($t\bar{t}$), and QCD multi-jets events. Given that the BDT is sensitive to the correlations among the 15 input observables, it is advantageous to carry out a fully data driven background prediction. The estimation procedure is described below.

9.2.1 Backgrounds

Predictions for the $Z \rightarrow \nu\bar{\nu}$ and QCD counts in the signal regions are readily carried out using the rebalance and smear and dilepton cleaning methods, respectively. Uncertainties in the QCD estimates are obtained using the procedure described in Section 8.2. Uncertainties in the $Z \rightarrow \nu\bar{\nu}$ prediction are statistical uncertainties associated with the two-electron control region, the residuals of the simulated closure test, and a 5% systematic uncertainty associated with the efficiency maps, which are derived in simulation. The backgrounds of W+jets and $t\bar{t}$ are estimated using the following method.

Prediction for $t\bar{t}$ and W+jets. Events arising from the signal benchmark scenario do not contain heavy flavor jets, that is, jets originating from bottom or top quarks, nor events with leptons. Therefore, a sample of real events having one or more b-tagged jet or

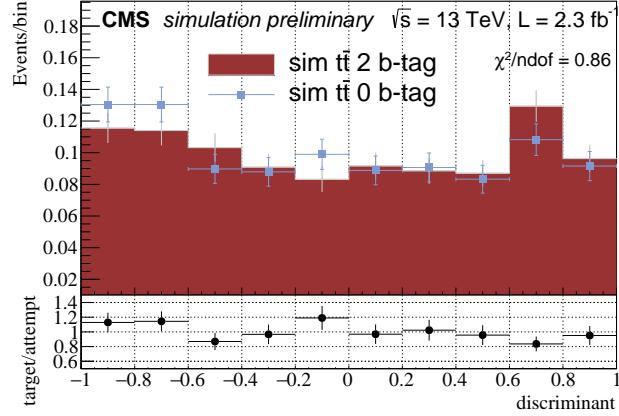


Figure 9.4: Comparison of the normalized discriminant between simulated $t\bar{t}$ events with 2 b-tags and 0 b-tags, in the baseline region.

one or more lepton is expected to comprise primarily $t\bar{t}$ and W+jets events, and no signal events. Because the only difference between $t\bar{t}$ events with 0 b-tags and those with 2 b-tags is whether or not the b-jets in the event are tagged as such, one expects the shape of the discriminant distribution of the 0 and 2 b-tag event sets to be similar. This is confirmed in simulation, as seen in Fig. 9.4, where the discriminant distribution for $t\bar{t}$ events with 0 b-tags is compared with the same for events with 2 b-tags; it is found that the shapes agree within statistical uncertainties. This similarity motivates the use of a real 2 b-tag data sample to define the $t\bar{t}$ template.

Simulation also reveals that W+jets events yield a discriminant distribution that is indistinguishable from that of $t\bar{t}$ events (Figure 9.5). Furthermore, simulation indicates that the shapes of the output of the discriminant in the baseline region are indistinguishable across a range of final states, among them final states with 0, 1, and 2 b-tagged jets, as well as final states with 0 and 1 lepton, for both $t\bar{t}$ events and W+jets events. Therefore, real data samples corresponding to these final states are used to derive shape templates for the combined contribution from $t\bar{t}$ events and W+jets backgrounds. The similarity of the shapes for the two backgrounds is checked in real data by comparing samples with various relative contributions from $t\bar{t}$ and W+jets. These regions include the 2 b-tag 0-lepton region, the

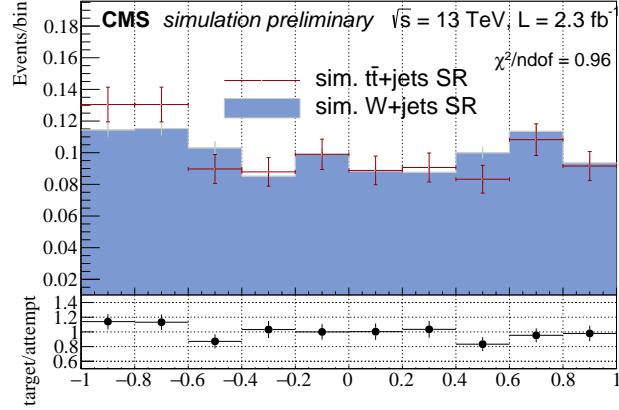


Figure 9.5: Comparison of the normalized discriminant between simulated $t\bar{t}$ events and $W+jets$ events, in the signal baseline region.

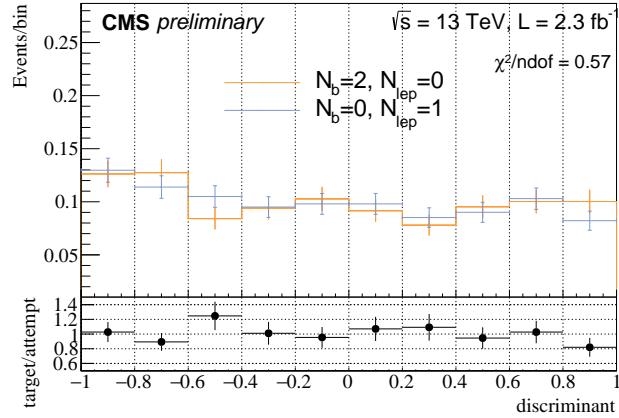


Figure 9.6: Comparison of the output of the discriminant shapes in real data between control regions with various relative contributions from $t\bar{t}$ and $W+jets$ events.

0 b-tag 1-lepton region, and regions with various numbers of b-tags and leptons. Shapes of distributions defined with the above sets of selection are seen to be compatible in Fig. 9.6.

The $t\bar{t}/W+jets$ shape template and associated uncertainty is derived from the pooled b-tag and lepton control regions discussed above. The normalization of the $t\bar{t}/W+jets$ template is fixed by the observed count in the normalization control region (the leftmost bin of the discriminant), less the predicted count from the other backgrounds in the normalization control region, namely, the leftmost bin of the discriminant. Given this choice of normalization, the

predicted $t\bar{t}$ and $W+jets$ event counts in the signal regions are anti-correlated with the prediction for the other backgrounds in the normalization control region. This anti-correlation is taken into account in the final prediction by generating ensembles of background predictions, where each prediction in the ensemble receives the full treatment described above, but has a unique value for the normalization obtained from a random sampling of a gaussian whose mean and standard deviation is the predicted count and uncertainty from non- $t\bar{t}$ and $W+jets$ in the normalization region. All uncertainties are propagated through the ensemble.

Other backgrounds from rare standard model processes include single-top quark production, two and three boson production, and make only a minor contribution to the total background. These backgrounds are simulated, and uncertainties of 100% are assumed for the counts in all signal regions. The signal counts are estimated from a sample of simulated events generated by PYTHIA 8 [101] and processed with the CMS fast detector simulation program [102]. Uncertainties reflect the uncertainty in the signal cross section of 13% and statistical uncertainties in the event sample on the order of 10%. An additional uncertainty, guessed to be 20%, is assigned to account for PDF, scale, and ISR uncertainties. These uncertainties have not been computed directly due to time constraints, and this treatment should be improved in future iterations of the analysis.

9.2.2 Results

Estimates of the total background counts and uncertainties in the signal regions, along with the observed counts, are given in Table 9.1. The full results are also shown in Fig. 9.7, with the distribution of the benchmark pMSSM point included in the background stack. The results based on pure simulation are also shown for comparison, although are not used in the interpretation. The results are interpreted using a “counts” likelihood, similar to that described in Section 7.3.2, given by

$$L(D^{\text{CMS}}|\theta) = \prod_{i=1}^9 \int \text{Poisson}(N_i|s_i + b_i) p(b_i|B_i, \delta B_i) p(s_i|S_i, \delta S_i) db_i ds_i, \quad (9.1)$$

where the product is over the 9 signal regions. For each search bin i , N_i is the observed count, s_i and b_i are the expected number of signal and background counts, respectively, $B_i \pm \delta B_i$

Table 9.1: Expected and observed counts and uncertainties, along with the prediction for the signal yield.

signal region	expected bkg. count	observed count	signal count
1	229 ± 48	172	9 ± 5
2	159 ± 39	192	18 ± 8
3	157 ± 35	161	36 ± 13
4	128 ± 31	163	41 ± 14
5	137 ± 35	137	68 ± 21
6	167 ± 40	157	57 ± 18
7	162 ± 39	164	104 ± 29
8	167 ± 36	194	118 ± 33
9	127 ± 30	191	61 ± 19

is the estimate of the background count and its uncertainty, and $S_i \pm \delta S_i$ is the estimate of the signal count and its uncertainty. The prior density for b_i , $p(b_i|B_i, \delta B_i)$, is modeled as a gamma density, $\text{gamma}(x; \alpha, \beta) = \beta \exp(-\beta x)(\beta x)^{\alpha-1}/\Gamma(\alpha)$, with α and β defined such that the mode and variance of the gamma density are B_i and $(\delta B_i)^2$. An analogous prescription is used for the signal prior. The signed significance Z described in Section 7.3.2 is also computed, where $Z < -1.64$ indicates exclusion at the 95% CL, and $Z \gg 0$ indicates evidence for the signal hypothesis with a significance equal to Z .

The expected value of Z under the background-only hypothesis is -5.16 , indicating an expected exclusion of the benchmark model with a significance of approximately 5 standard deviations. The expected Z under the background + signal hypothesis is 4.69 , indicating that, if the signal hypothesis were true, I would expect to observe evidence for the signal with a significance of approximately 4 standard deviations. The observed Z -significance is -3.58 , indicating that the benchmark model is excluded at a significance of about 3 standard deviations. The observed exclusion is weaker than the expected exclusion because of observed excesses over the predicted counts in the most sensitive signal regions.

The benchmark pMSSM point, which survived previous CMS analyses, has been excluded. This preliminary analysis stands as a demonstration that, given a willingness to adopt new techniques, we can look for supersymmetry in the most background-dominated

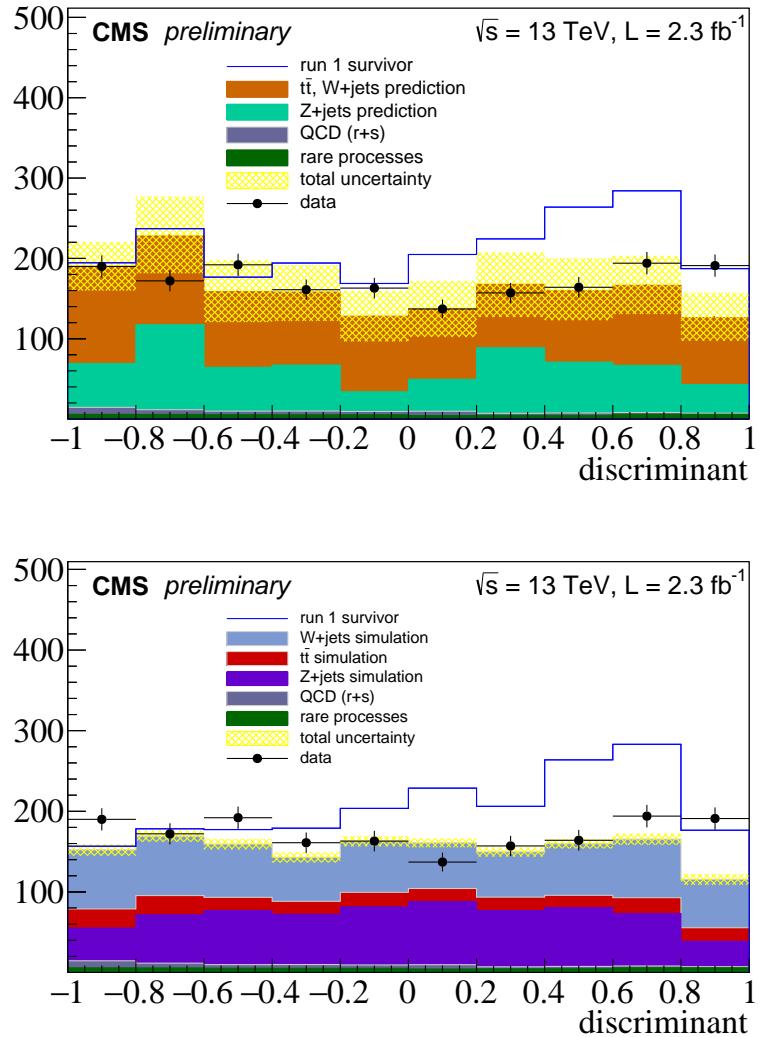


Figure 9.7: Predicted and observed counts for the low- H_T search using the fully data-driven prediction (top), as well as the prediction from pure simulation (bottom), shown for comparison. The yellow bands are the total uncertainties in the data-driven prediction (top), and the purely statistical uncertainties in the simulated counts (bottom).

kinematic regions, where evidence of new physics may be manifest. If supersymmetry is not discovered at high- H_T and high H_T^{miss} , these regions and techniques may be key to ruling out, or providing evidence for, weak-scale supersymmetry at the LHC.

A note on the discrimination power. It is possible that providing more sophisticated variables as input to a multivariate discriminant could result in better performance

than that achieved by the discriminants discussed here. Appendix C gives a brief introduction to a few such observables, some of which have been central to CMS and/or ATLAS searches for supersymmetry. Studies that evaluate the effectiveness with which these observables probe the pMSSM, taken alone and in combinations with each other, may prove informative. Also, training with a larger sample of background events may increase the separation power of the discriminant.

CHAPTER 10

SUMMARY AND CONCLUSION

Our understanding of the nature of electroweak symmetry breaking may be nearing a turning point. If no evidence of new phenomena at energy scales at or around 1 TeV comes to light at the LHC, there appears to be fine tuning in the standard model. This would be an extraordinary finding, since fine tuning is typically thought to arise only in systems of which multiple similar copies of the system exist somewhere. Considering that the system in question is the universe, then the most acceptable explanation would seem to be that there exist other copies of the universe “somewhere.” Supersymmetry is a proposed symmetry, which if realized in the context of the standard model, could explain the apparent fine tuning in the standard model as a natural, and unavoidable, consequence of the laws of physics. The hypothesis of supersymmetry that is consistent with our current understanding of the universe predicts the existence of at least 30 new fundamental fields, many of which would manifest particles that could potentially be detected in proton collisions at the LHC.

One of the most daunting challenges in making definitive statements about supersymmetry (SUSY) is that there are a multitude of ways in which SUSY could manifest itself. The values of the masses of the new particles could be ordered in any one of an inconceivably large number of permutations. This means that the evidence for SUSY could come a large variety of ways, and it is difficult to be sure if, in the absence of evidence, there is evidence of absence. Searches for supersymmetry at the LHC are typically guided by simplified models, which hypothesize that only two or three of the new particles are present in the collisions of the LHC, and the rest are ignored, or assumed to take on very large masses. So far, after a large number of searches performed by the CMS and ATLAS experiments and interpreted in terms of the simplified models, there is absolutely no hint of supersymmetry. Many so-called natural supersymmetry models have been ruled out. However, given the limited scope of the models, it can be difficult to argue for or against the weak-scale SUSY hypothesis.

To begin to address fundamental limitations of interpreting results in the context of simplified models, I contributed to an analysis that interpreted the results of 7 and 8 TeV CMS searches in terms of a SUSY model that more compressively covers the set of potential particle spectra allowed by the MSSM, called the phenomenological minimal supersymmetric standard model (pMSSM) [29]. A number of reasonable assumptions about the MSSM are made which, along with the incorporation of results from past experiments and non-CMS data, constrain the MSSM to a model with just 19 free parameters. The pMSSM is thus an excellent proxy for the MSSM that captures most of the MSSM's phenomenological features that are relevant at the LHC.

Several important lessons were learned about the viability of the supersymmetric hypothesis given the constraints of CMS 7 and 8 TeV analyses. For example, it was found that gluinos, which, in the simplified model paradigm are found to be excluded below a mass of about 1.2 TeV, are viable in the pMSSM with masses as low as 600 GeV. Scenarios with these very light gluinos tend to feature lightest SUSY particles, or LSPs, with masses within a few hundred GeV of the gluino mass, which causes the decay products of the gluino to be of low energy (soft), and thus either are not reconstructed in the detector or do not satisfy the event selection of the searches. Additionally, these scenarios typically feature long, sometimes exotic, decay chains, in which the gluinos would decay to LSPs through one or several intermediate particles, typically squarks, neutralinos, or charginos; the production of a gluino in association with a squark also accounts for many nonexcluded scenarios with low-mass gluinos. Such scenarios would tend to predict more particles in the events, but with small transverse momenta than those in scenarios that are excluded; there is a trade-off between the average energy of particles and the multiplicity of particles.

Similar conclusions were made with regard to first- and second-generation squarks. It was found that our knowledge of the third generation squarks (top and bottom squarks) has not been significantly impacted by the CMS searches because results from previous experiments, such as measurements of b quark properties, excluded most of the low-mass top and bottom squark scenarios before CMS analyses arrived on the scene. The lightest colored particle among the surviving pMSSM points has a mass of around 400 GeV. It was found that no

mass range of the LSP can be excluded, even in the range below 100 GeV. Scenarios with very light LSPs predict enormous cross sections of SUSY events, but the events have few or no properties to which our current analyses are sensitive.

After a decomposition of the nonexcluded model space into principal processes, it was found that the second most common process in the pMSSM is a process for which there is no simplified model targeted by any of the CMS or ATLAS searches, namely, the direct production of one LSP and one chargino. In many cases, the chargino and LSP are nearly mass degenerate, which can lead to either soft, off-shell W bosons in the decay of the chargino, or in the case of extreme mass degeneracy, charginos with lifetimes as long as 10 mm. In either case, soft W bosons or long-lived charginos could provide the observables necessary to probe this difficult class of model points. Specifically, a set of event selection that includes initial-state radiation in conjunction with a reconstructed soft lepton or disappearing track traveling in a similar azimuthal direction as the missing transverse momentum, could define the starting point for such an analysis.

Moreover, two general findings were made. The first is that searches for SUSY in the hadronic channel were found to be the most effective at probing the MSSM, by a significant margin. Second, many pMSSM points with very high cross sections survive the 7 and 8 TeV CMS searches because they predict signal events that arise mostly in background-dominated kinematic regions, such as regions of low- H_T , low- H_T^{miss} , and low jet multiplicity. Because these kinematic regions are background-dominated, new discriminating variables may be required to allow for sensitivity to the nonexcluded signals predicted to appear in these regions. After making these findings, I spent much of my remaining PhD career developing and implementing methods that increase the sensitivity of hadronic searches in regions of low- H_T , H_T^{miss} , and N_{jets} .

I cast the rebalance and smear method, a procedure for data-driven QCD estimation, as a problem of Bayesian inference. A prior probability distribution of the parton-level H_T^{miss} in QCD events, combined with a likelihood function built out of the full jet response functions, was used to construct a posterior density for the set of parton-level jets in an event that proved to model QCD events with good accuracy. The result is a data-driven QCD

event generator that more accurately models standard model multi-jet events than previous estimation methods, even reproducing the correct correlations between the magnitudes and directions of the jet four-vectors in the events. The method yields a good prediction even in the case that there is signal contamination in the data control samples from which the prediction is derived, which means that the method has been applied using real data events, thereby preserving the true correlations of jets and the jet multiplicity exhibited by nature, quantities that are notoriously poorly-modeled in Monte Carlo event simulation. Moreover, the method has been applied repeatedly, which serves to drive the statistical uncertainty in the QCD prediction to small values in all search regions. This method has been developed in the context of the CMS multi-jet + H_T^{miss} search for supersymmetry at $\sqrt{s} = 13$ TeV [6], in which limits were set on the masses of the gluino in the range 1400–1600 GeV, assuming a massless LSP, in the context of simplified models.

I extended the method for the estimation of the $Z \rightarrow \nu\bar{\nu}$ background in an analysis searching for evidence of top squarks in the CMS data [7], so that the final prediction is constrained by real events containing two electrons. Previously, the estimation was constrained by samples of events with two muons, so this method provides both a check for and an improvement in the $Z \rightarrow \nu\bar{\nu}$ background prediction. The main hurdles were accurately modeling the acceptance, reconstruction efficiency, and isolation efficiency of Z bosons decaying into pairs of electrons, and combining the electron and muon data in a coherent way. The result is a parametrization of the efficiency that can be applied in data as a way to predict the standard model $Z \rightarrow \nu\bar{\nu}$ distributions, including for observables that rely on complicated relations between properties of the jets, such as multivariate discriminants. Furthermore, the efficiency parameterization may be applied to the muon samples in future versions of the hadronic top squark search, since the level of consistency between the prediction and the expectation is high, including in regions of low- H_T^{miss} and H_T . Limits were set on the masses of the top squark and the LSP.

I made use of, and built upon, methods of measuring the efficiency of triggers in a way that either avoids or properly accounts for bias that may be introduced choice of the control sample used to derived the efficiency. The result is an estimate of the trigger efficiency that

is a smoothly varying function than can depend on multiple final state observables. The benefits of this technique include an easing of the difficulty of conducting an analysis in kinematic regions in which the trigger is not fully efficient. Often these regions are avoided in analysis, sometimes out of a concern that unknown bias affect the traditional, binned efficiency measurements, and sometimes out of the conventional notion that such regions are not populated by a large number of signal events, and so aren't worth analyzing. However, SUSY may exist and manifest itself in these difficult regions, and so tools like these may be key if we decide to rise to the challenge.

I established that multivariate techniques can be employed as a means of significantly increasing the sensitivity of searches in regions of low H_T^{miss} and H_T . Multivariate discriminants that take jet properties as input can provide good signal/background discrimination for the main backgrounds in these regions, which are primarily standard model QCD and $Z \rightarrow \nu\bar{\nu}$ events. Moreover, the data-driven methods of background estimation presented in this document and employed high- E_T^{miss} regions, have been shown to accurately model the background distributions of such discriminants.

I contributed to detector research and development, helping to conduct test beam activities and analysis of data for the CMS Phase II upgrade. To this end, I authored an event display that allowed the data to be viewed during and after their collection. The event display has been adopted for use in other testbeam experiments, and I look forward to employing it in future work.

Finally, I worked on a project [30] to help to make the results of published CMS searches more accessible to theorists. This project established a public framework in which anyone can write code that calculates the signal efficiency \times acceptance of a CMS or ATLAS analysis. Once an analysis has been re-implemented and validated, theorists, or anyone interested in how a given model is constrained by the CMS data, can quickly improve their understanding of their model. The details of the project, along with the corresponding published paper [30], are given in Appendix D.

Much work has been carried out to learn the true nature of electroweak symmetry breaking. Clearly, a great deal more work remains to be done before we will understand this

curious aspect of nature. But there is good reason to believe that with ever more effective and rigorous analyses, with increasingly bold ideas and careful reasoning, we may be able to state conclusively, by the end of the LHC era, that either we have discovered supersymmetry or that the weak-scale SUSY hypothesis has been ruled out.

APPENDIX A

ADDITIONAL PMSSM CHECKS AND STUDIES

A.1 Change of prior

For any choice of prior, there will be some parameterization in which the prior density is constant (that is, uniform or flat). The difficult question to answer is: in which parameterization ought the prior density be constant? In the pMSSM paper, the ur-prior (coined by Glen Cowan from the German prefix *ur* meaning original or primitive), is taken to be uniform in a physically motivated parameterization. However, *a priori* it is not clear that this is the parameterization in which the prior should be flat. Foregoing a discussion of how best to choose a parametrization, it is examined how the results are affected when the choice of ur-prior is changed.

Figures A.1-A.2 show the prior and posterior density projections onto a suite of parameters presented in the paper [29], given the choice of a flat prior (red) and a prior proportional to the product of the log of each parameter (blue), referred to here as the log prior. The main conclusions of the paper stand with respect to this change, namely that

- pMSSM points with gluino masses below 500 GeV are excluded, but a small number points with 550 GeV gluinos survive (Fig. 7.2);
- the probability density of the light top squark mass is not significantly shifted by the CMS analyses (Fig. 7.5);
- no LSP mass or $\tilde{\chi}^\pm$ mass can be excluded (Fig. 7.6), and
- the most probable value (MPV) of the signal cross section has been pushed down by a factor of 10 (Fig. 7.8).

The posterior MPVs for most parameters agree within 5% between the flat and log priors. The most significant difference is in the impact on the LSP mass, where the log prior prefers

slightly larger values of the LSP mass (Fig. 7.6). This is because the LSP in general exhibits a high density in the low mass region, and this region is suppressed due to the log Jacobian. The result is that the bulk of the prior and posterior is shifted down.

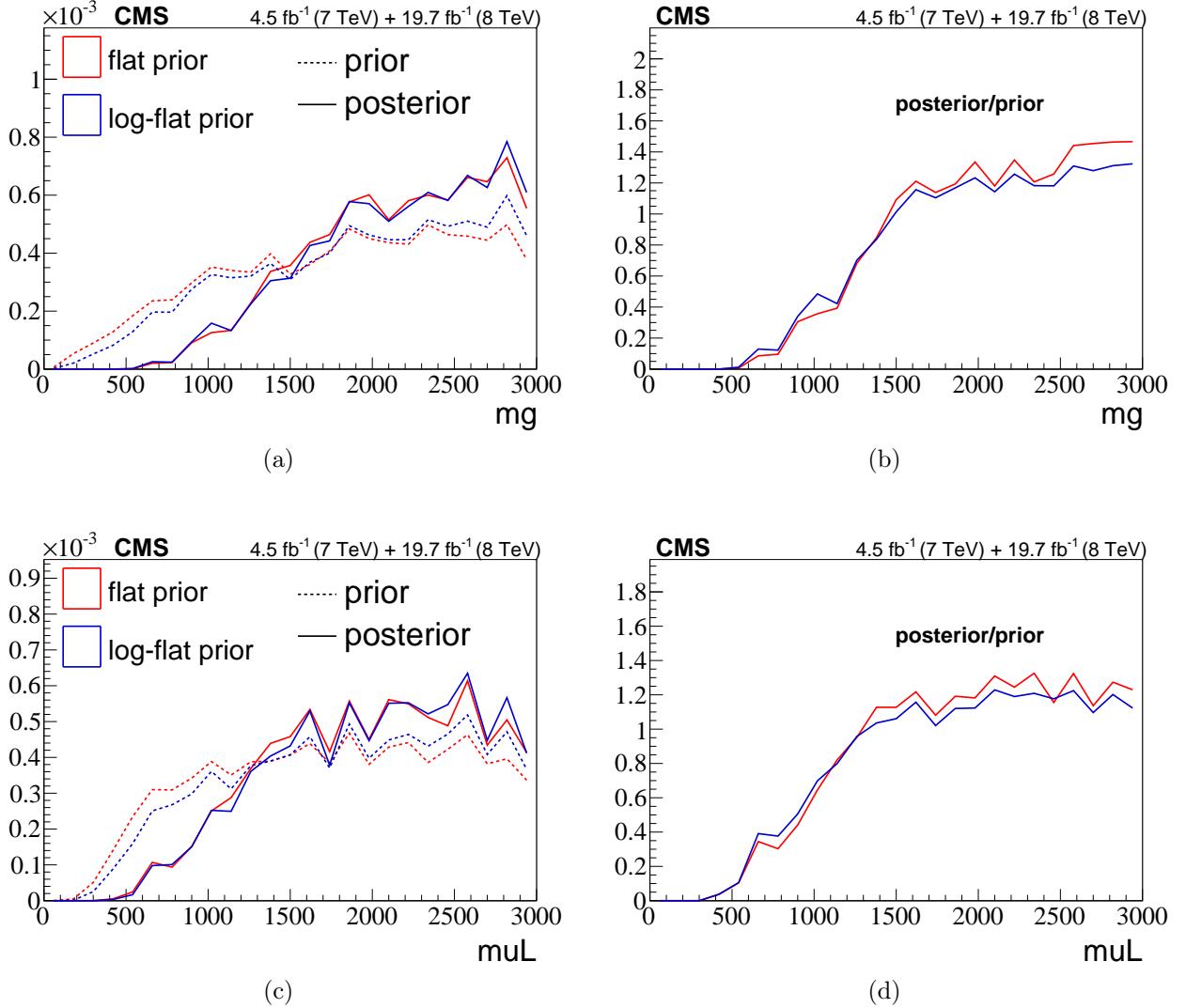


Figure A.1: Left: posterior/prior of $m_{\tilde{g}}$ (top) and μ_L (bottom). Right: Ratio of the posterior to the prior of $m_{\tilde{g}}$ (top) and μ_L (bottom).

A.2 Profile vs marginal likelihoods

Experience suggests that when profile and marginal likelihoods are in approximate agreement, the conclusions obtained via frequentist and Bayesian methods approximately agree. Therefore, it is instructive to examine the profile likelihoods. To do so, a smooth analytic expression for the 19-D non-DCS likelihood is constructed. The non-DCS likelihood can be factorized as a separable component equal to the product of 19 1-D functions, and a correction function that accounts for the statistical dependencies among parameters, which, for simplicity, are referred to using the slightly imprecise term “correlation”:

$$P(D^{\text{non-DCS}}|\vec{\theta}) = C(\vec{\theta}) \cdot \prod_i^{19} f(\theta_i), \quad (\text{A.1})$$

where $\vec{\theta}$ is a vector of the 19 pMSSM parameters of which θ_i is the i th element, the $f(\theta_i)$ functions are the 1-D non-DCS marginal likelihoods,

$$f(\theta_i) = \int_1 d\theta_1 \cdots \int_j d\theta_j \cdots P(D^{\text{non-DCS}}|\vec{\theta}) \pi_F(\vec{\theta}) \quad j \neq i, \quad (\text{A.2})$$

using a flat ur-prior, $\pi_F(\vec{\theta})$, and $C(\vec{\theta})$ encodes the correlations between the pMSSM parameters induced by the non-DCS measurements and constraints.

The 1-D f functions are estimated using TMVA regression methods [49]. For each regression, the swarm of points that constitute a discretized approximation to the non-DCS likelihood is projected onto a single pMSSM parameter, forming a 1-D histogram. The heights of the histogram bins are taken as the target, and the centers of the bins are taken as the input variable. A multi-layer perceptron (MLP)[49] is trained on a portion of the pMSSM points. The resulting $f(\theta_i) \equiv \text{MLP}(\theta_i)$ functions are shown in Figures A.6, along with the target histograms. The MLP approximations are seen to be reasonable.

The correlation function $C(\theta)$ is estimated using a multivariate classifier whose training sample consists of signal events corresponding to the original scan points, and background events consisting of pseudo-data generated from a sampling of the separable regression product. The classifier is an MLP whose output (see Figure A.8) can be interpreted generally as

$$D = \frac{p(\vec{\theta}|s)}{p(\vec{\theta}|s) + p(\vec{\theta}|b)}, \quad (\text{A.3})$$

given that the training was performed with equal numbers of signal (s) and background (b) events. In this case, the output is

$$D = \frac{p(D^{\text{non-DCS}}|\vec{\theta})}{p(D^{\text{non-DCS}}|\vec{\theta}) + \prod_i^{19} f(\theta_i)}. \quad (\text{A.4})$$

Rearranging Equation(A.4) gives

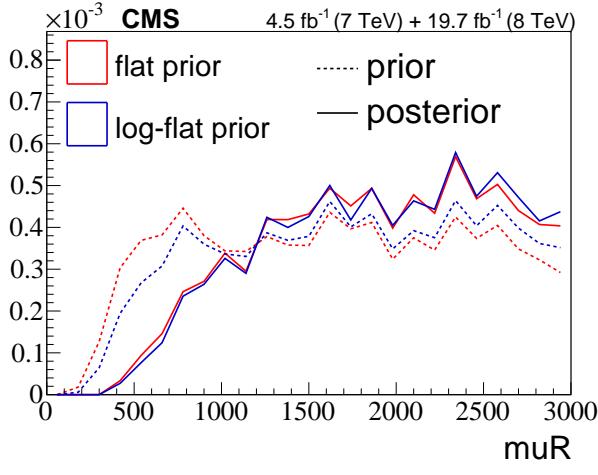
$$P(D^{\text{non-DCS}}|\vec{\theta}) = \frac{D}{1 - D} \prod_i^{19} f(\theta_i), \quad (\text{A.5})$$

and the correlation function (Equation(A.1)) is identified as

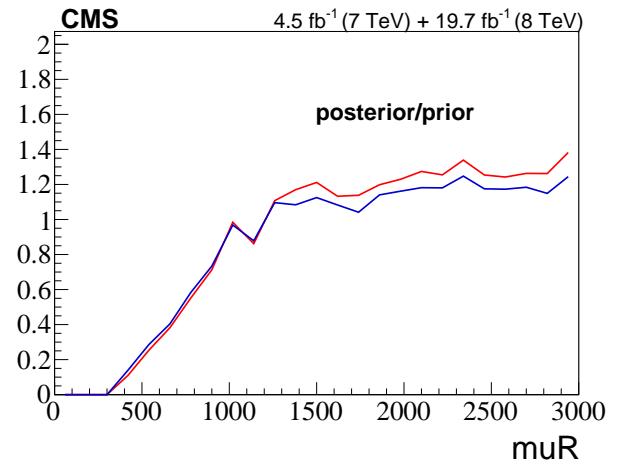
$$C(\vec{\theta}) = \frac{D}{1 - D}. \quad (\text{A.6})$$

The 1-D profile likelihoods were derived by maximizing Equation A.5 with respect to the $n - 1 = 18$ parameters.

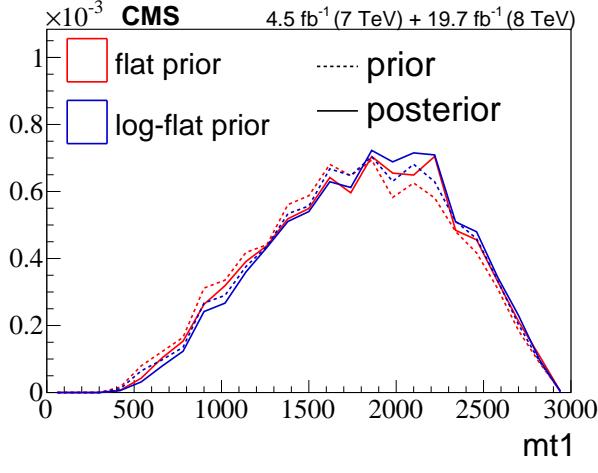
Comparisons between the profile likelihoods and the 1-D non-DCS marginal likelihoods are given in Figures A.9-A.12. The comparison shows general agreement to within $\sim 10\%$, with a couple of exceptions. The first notable difference is in the bino mass Lagrangian parameter M_1 , where the profile likelihood is relatively enhanced at large absolute values of M_1 . If the LSP is bino-like, confidence levels (CLs) based on the posterior densities would tend to under-exclude LSP masses; however, no LSP mass is excluded by the Z-significance, and so this does not alter the statement about LSPs made in the paper ???. Finally, the trilinear coupling A_t appears to have a roughly 30% enhancement in the negative region for the likelihood with respect to the marginal likelihood. However, the paper makes no explicit statement about the impact of CMS analysis on the trilinear couplings. The otherwise general similarity of the two sets suggests that a frequentist analysis based on the profile likelihoods would yield conclusions consistent with those in the CMS pMSSM paper. The Z-significance is a mapping of the Bayes factor that roughly maps to a frequentist n-sigma, the only difference being that rather than maximizing the likelihood to eliminate nuisance parameters, the likelihood is marginalized.



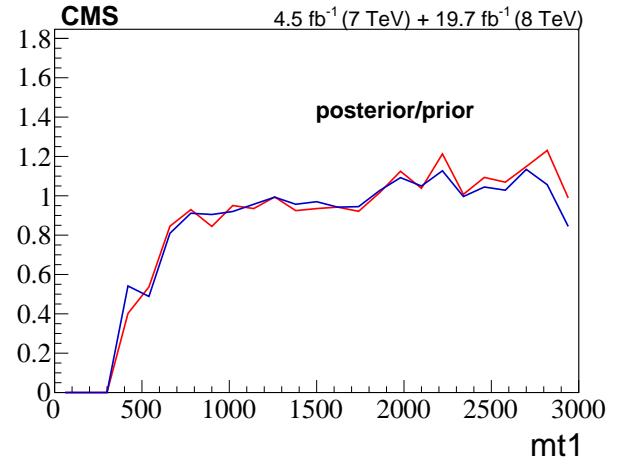
(a)



(b)

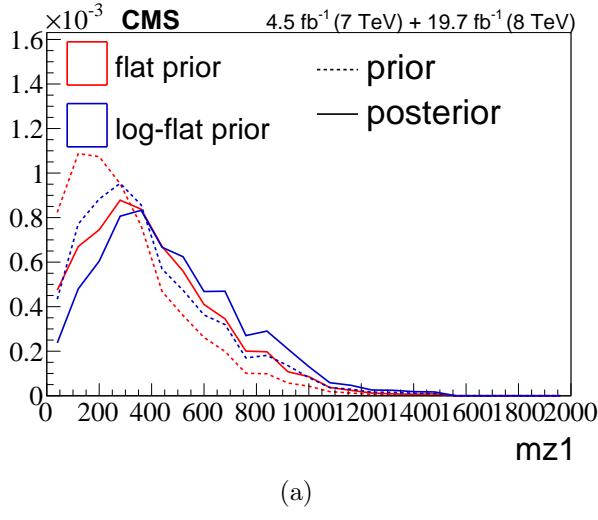


(c)

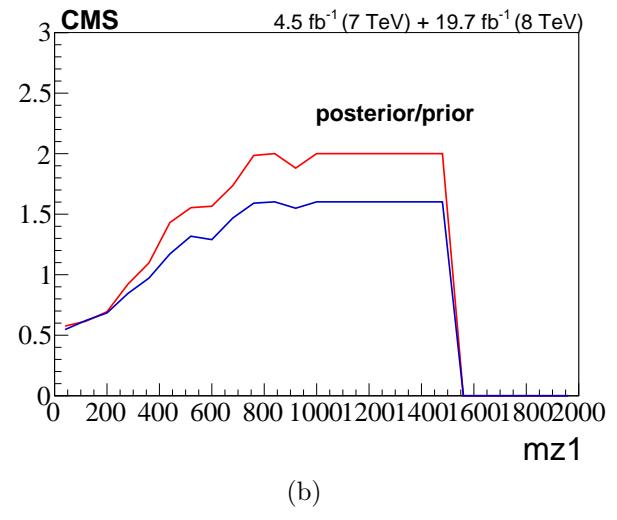


(d)

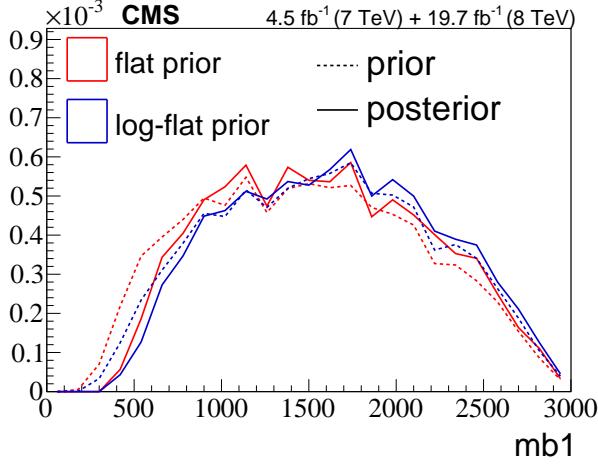
Figure A.2: Left: posterior/prior of m_{U_R} (top) and $m_{\tilde{t}_1}$ (bottom). Right: Ratio of the posterior to the prior of m_{U_R} (top) and $m_{\tilde{t}_1}$ (bottom).



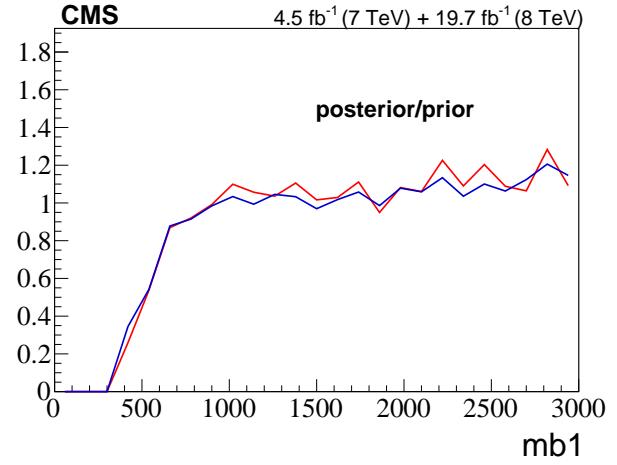
(a)



(b)

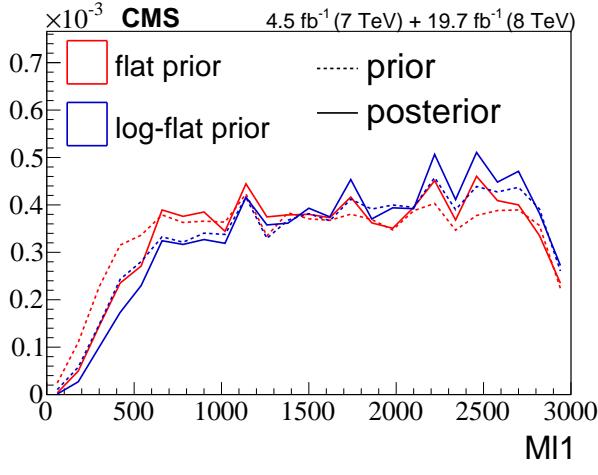


(c)

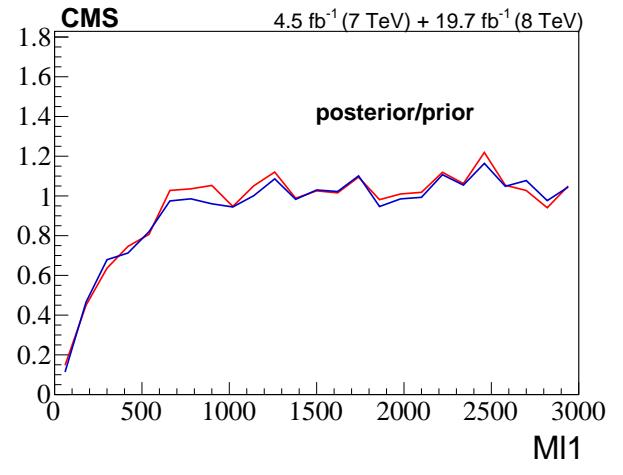


(d)

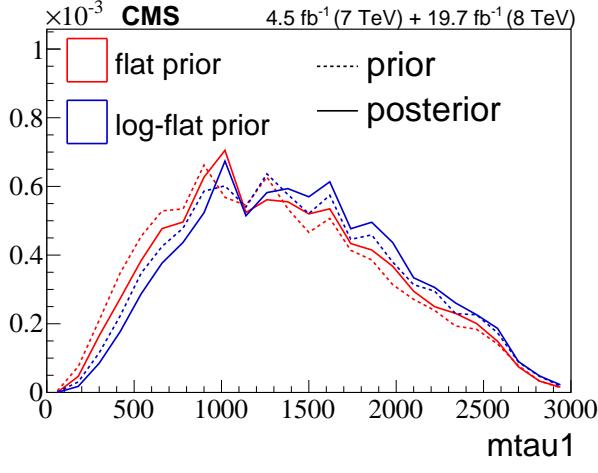
Figure A.3: Left: posterior/prior of $m_{\tilde{\chi}_1^0}$ (top) and $m_{\tilde{b}_1}$ (bottom). Right: Ratio of the posterior to the prior of $m_{\tilde{\chi}_1^0}$ (top) and $m_{\tilde{b}_1}$ (bottom).



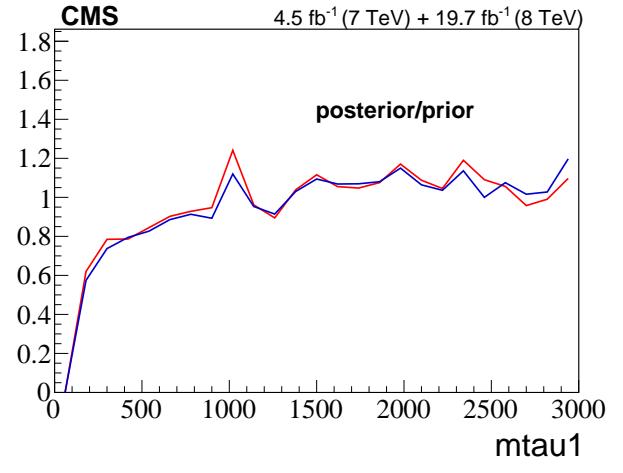
(a)



(b)

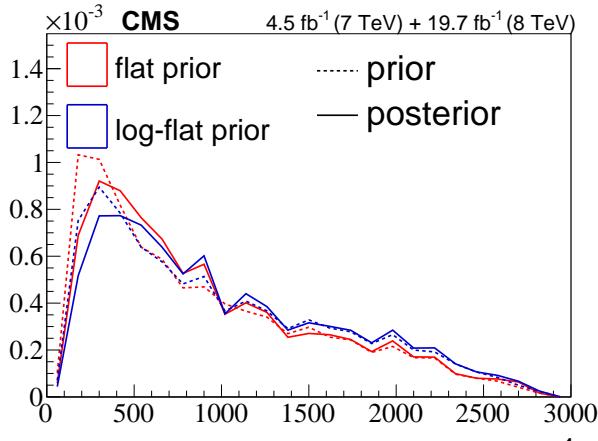


(c)

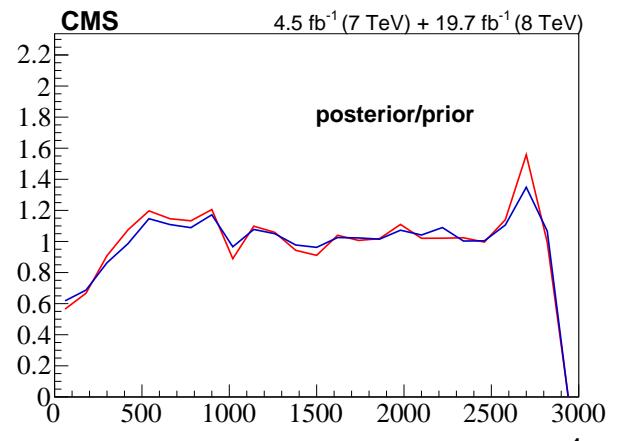


(d)

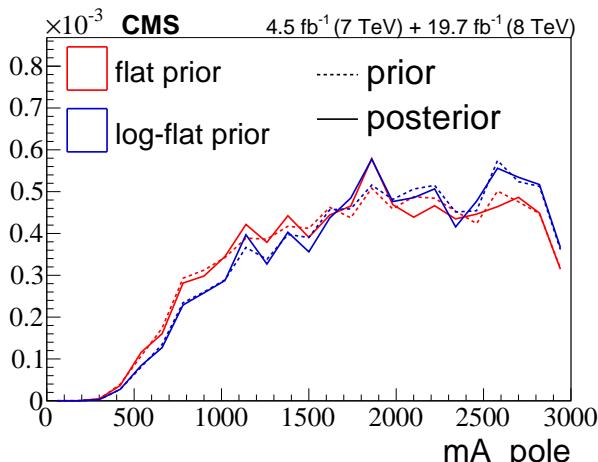
Figure A.4: Left: posterior/prior of $m_{\tilde{l}_1}$ (top) and $m_{\tilde{\tau}_1}$ (bottom). Right: Ratio of the posterior to the prior of $m_{\tilde{l}_1}$ (top) and $m_{\tilde{\tau}_1}$ (bottom).



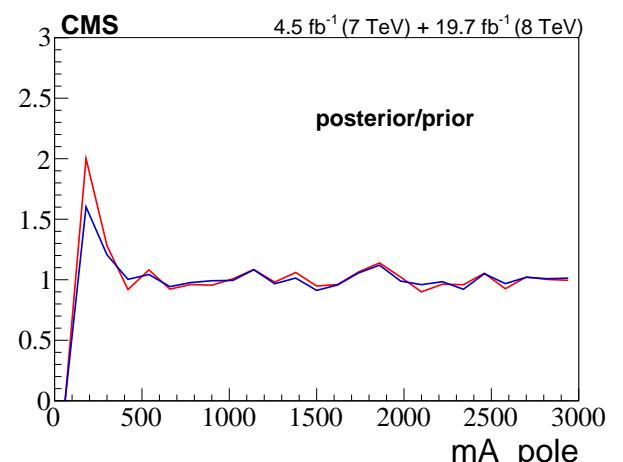
(a)



(b)



(c)



(d)

Figure A.5: Left: posterior/prior of $m_{\tilde{\chi}_1^\pm}$ (top) and $m_A(\text{pole})$ (bottom). Right: Ratio of the posterior to the prior of $m_{\tilde{\chi}_1^\pm}$ (top) and $m_A(\text{pole})$ (bottom).

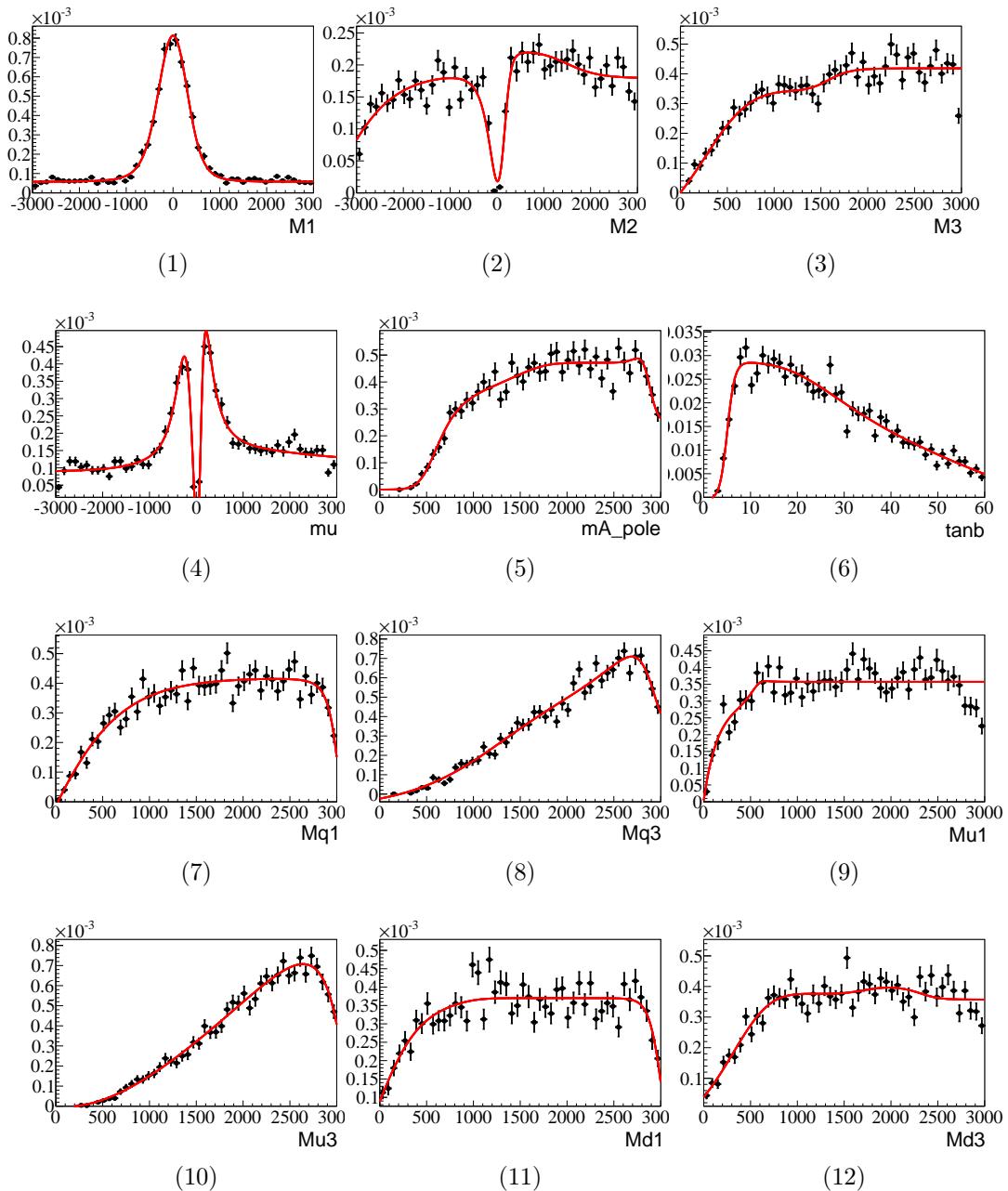


Figure A.6: The results of 10 of the 19 MLP regressions on the pMSSM parameters.

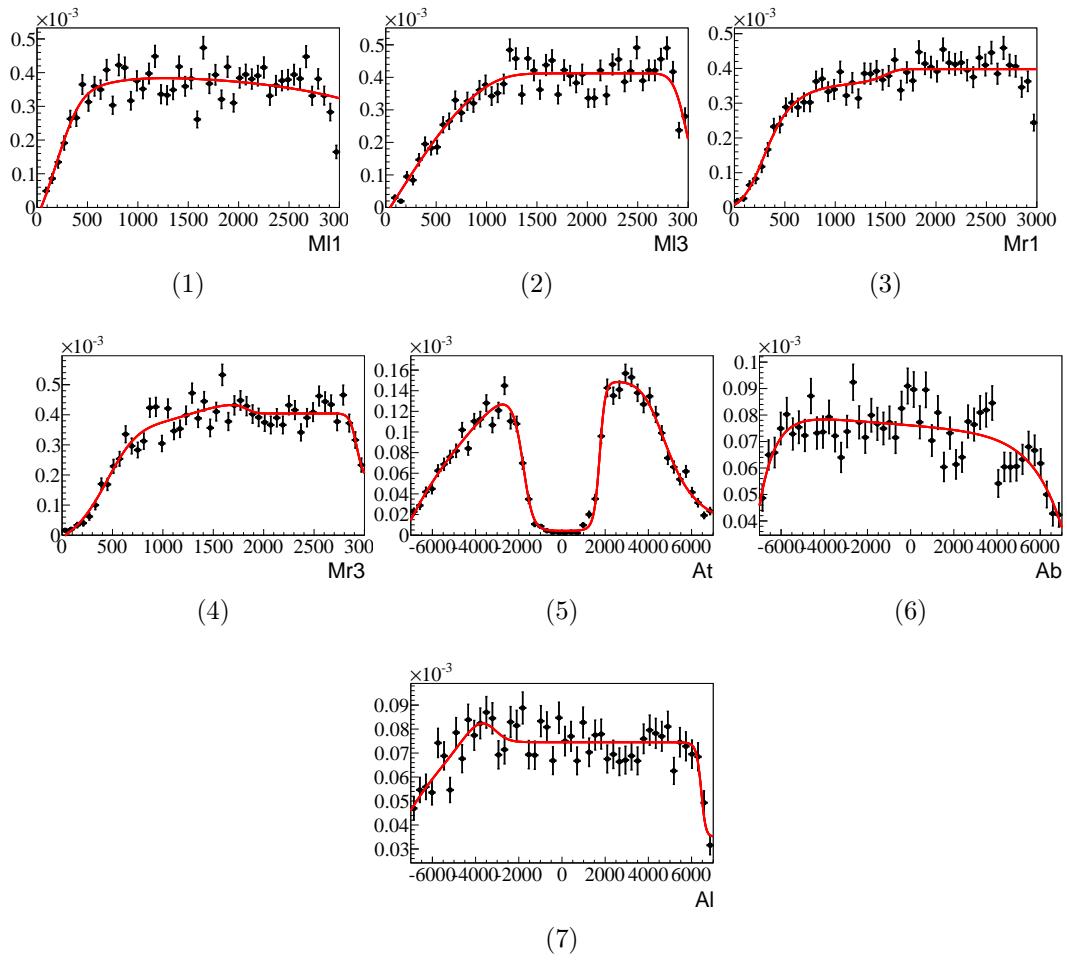


Figure A.7: The results of the 9 of the 19 MLP regressions on the pMSSM parameters.

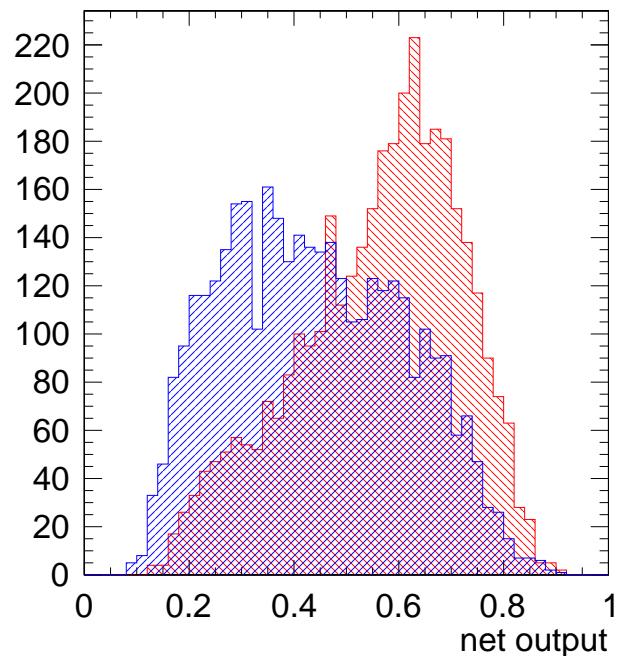


Figure A.8: The output of the MLP used to derive the correction function. The red distribution is the original set of pMSSM points and the blue distribution is the set generated by a random sampling of the separable likelihood function.

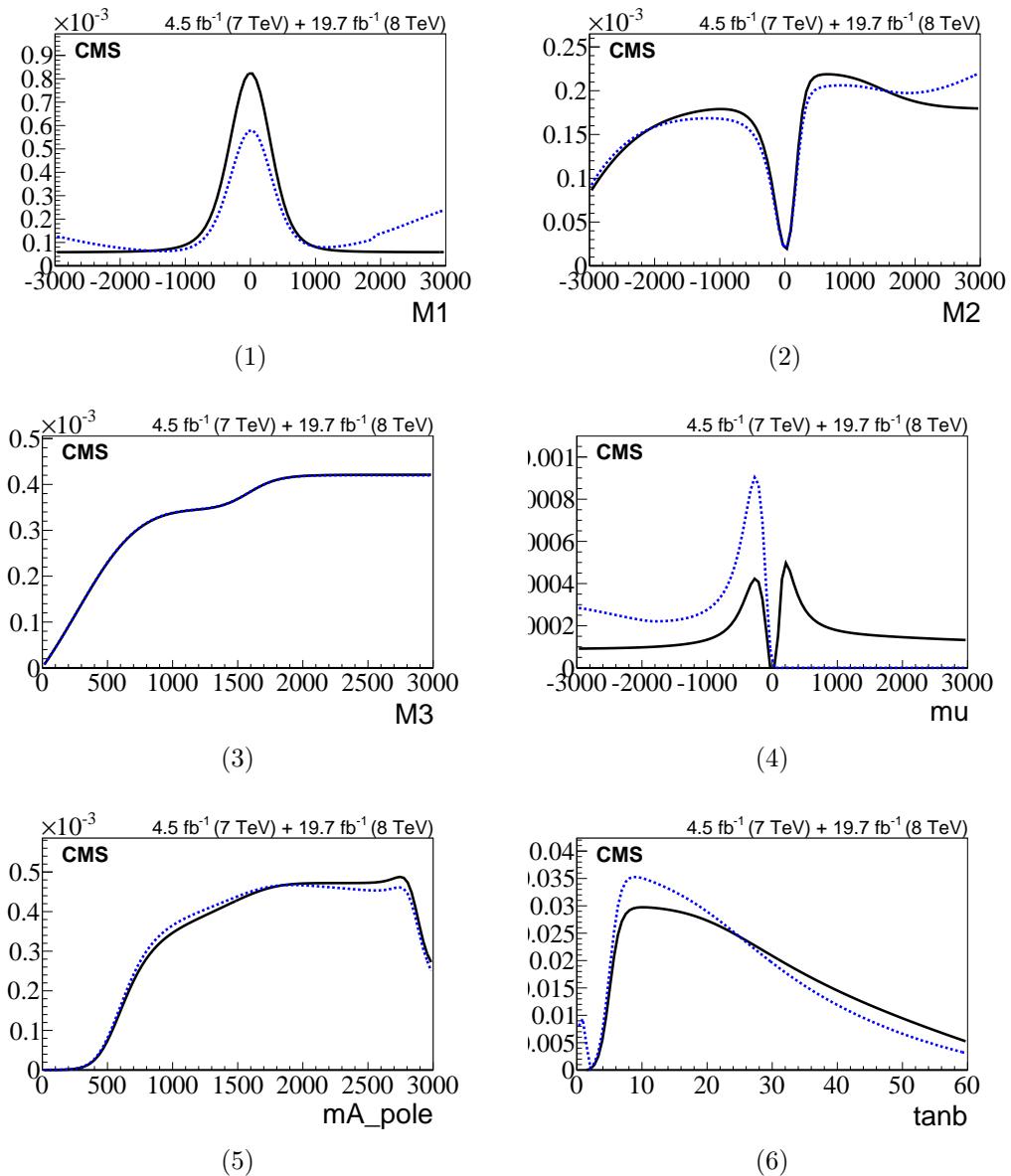


Figure A.9: Comparison of the frequentist profile likelihood (dashed blue line) to the non-DCS prior (solid black line) for 6 of the pMSSM parameters.

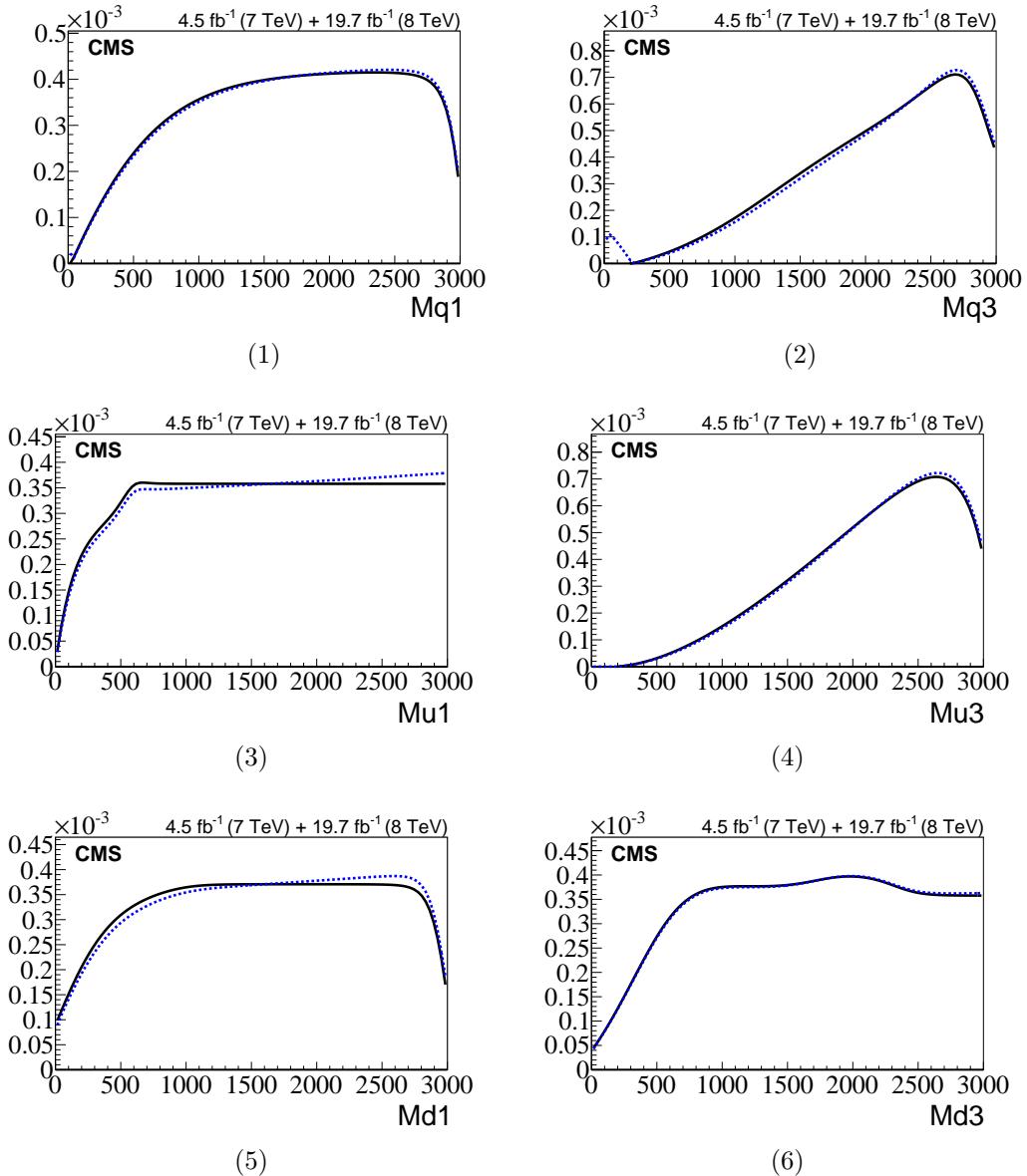


Figure A.10: Comparison of the frequentist profile likelihood (dashed blue line) to the non-DCS prior (solid black line) for an additional 6 of the pMSSM parameters.

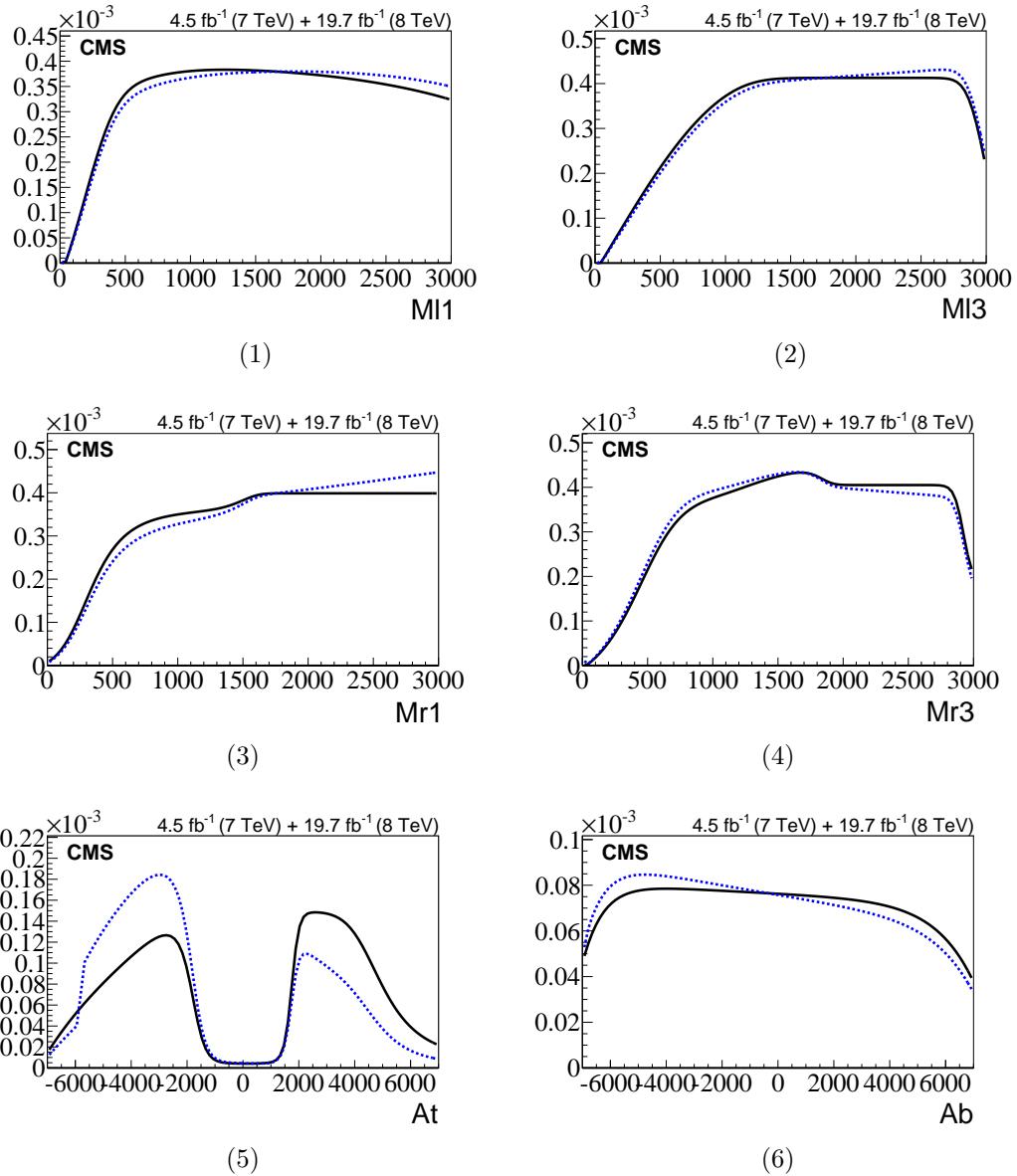


Figure A.11: Comparison of the frequentist profile likelihood (dashed blue line) to the non-DCS prior (solid black line) for an additional 6 of the pMSSM parameters.

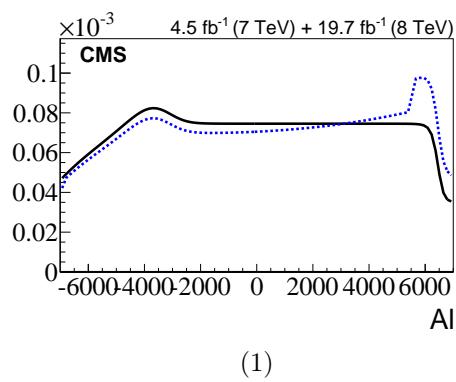


Figure A.12: Comparison of the frequentist profile likelihood (dashed blue line) to the non-DCS prior (solid black line) for a pMSSM parameter.

APPENDIX B

ANALYSIS TABLES

The following are tables of expected and observed counts for the multi-jet H_T^{miss} search [6] and the hadronic top squark search [7].

Table B.1: $Z \rightarrow \nu\nu$ background prediction and uncertainties, corresponding to 2.3 fb^{-1} for all search regions in the hadronic top squark search.

Search Bin	N_{tops}	$N_{\text{b-tags}}$	M_{T2} [GeV]	$E_{\text{T}}^{\text{miss}}$ [GeV]	$Z \rightarrow \nu\nu$ pred ($\mu\mu$)	$Z \rightarrow \nu\nu$ pred ($\mu\mu$)
0	1	1	200-300	200-275	6.52 ± 1.70	7.79 ± 4.27
1	1	1	200-300	275-350	2.39 ± 0.77	2.81 ± 1.17
2	1	1	200-300	350-450	1.13 ± 0.58	1.41 ± 0.58
3	1	1	200-300	450+	0.60 ± 0.32	0.78 ± 0.28
4	1	1	300-400	200-275	2.22 ± 0.67	2.59 ± 1.68
5	1	1	300-400	275-350	2.52 ± 0.98	3.00 ± 1.54
6	1	1	300-400	350-450	1.70 ± 0.87	2.03 ± 1.08
7	1	1	300-400	450+	0.94 ± 0.51	1.19 ± 0.59
8	1	1	400+	200-350	0.42 ± 0.34	0.46 ± 0.27
9	1	1	400+	350-450	1.39 ± 1.21	1.74 ± 1.06
10	1	1	400+	450+	3.19 ± 2.83	4.13 ± 2.73
11	1	2+	200-300	200-275	1.67 ± 1.49	1.85 ± 1.53
12	1	2+	200-300	275-350	0.63 ± 0.59	0.75 ± 0.56
13	1	2+	200-300	350-450	0.30 ± 0.30	0.38 ± 0.29
14	1	2+	200-300	450+	0.17 ± 0.18	0.20 ± 0.15
15	1	2+	300-400	200-275	0.60 ± 0.54	0.68 ± 0.60
16	1	2+	300-400	275-350	0.67 ± 0.61	0.71 ± 0.58
17	1	2+	300-400	350-450	0.43 ± 0.42	0.46 ± 0.38
18	1	2+	300-400	450+	0.26 ± 0.28	0.32 ± 0.26
19	1	2+	400+	200-450	0.44 ± 0.49	0.47 ± 0.39
20	1	2+	400+	450+	0.75 ± 0.85	0.94 ± 0.82
21	2+	1	200-300	200-275	0.79 ± 0.60	1.63 ± 0.97
22	2+	1	200-300	275-350	0.23 ± 0.18	0.48 ± 0.25
23	2+	1	200-300	350+	0.10 ± 0.09	0.24 ± 0.1
24	2+	1	300-400	200-275	0.61 ± 0.47	1.08 ± 0.79
25	2+	1	300-400	275-350	0.44 ± 0.35	0.81 ± 0.51
26	2+	1	300-400	350+	0.32 ± 0.28	0.62 ± 0.41
27	2+	1	400+	200-350	0.16 ± 0.16	0.27 ± 0.22
28	2+	1	400+	350+	0.95 ± 0.99	1.47 ± 1.05
29	2+	2+	200-300	200-275	0.17 ± 0.17	0.37 ± 0.27
30	2+	2+	200-300	275-350	0.06 ± 0.06	0.15 ± 0.12
31	2+	2+	200-300	350+	0.03 ± 0.03	0.06 ± 0.04
32	2+	2+	300-400	200-275	0.15 ± 0.15	0.30 ± 0.25
33	2+	2+	300-400	275-350	0.11 ± 0.12	0.26 ± 0.19
34	2+	2+	300-400	350+	0.06 ± 0.07	0.12 ± 0.09
35	2+	2+	400+	200-350	0.04 ± 0.04	0.09 ± 0.08
36	2+	2+	400+	350+	0.18 ± 0.23	0.34 ± 0.31

APPENDIX C

OTHER DISCRIMINATING VARIABLES

A number of observables were discussed in this document without their full definitions having been given, due to the amount of detail required. These observables have been largely embraced by the particle physics community because they are particularly effective at probing physics models that predict the pair production of new heavy particles that decay semi-invisibly—that is, each heavy particle decays into one detectable particle and one invisible particle. This signature is the hallmark of a number of BSM models, such as generic models of dark matter, 4th generation leptons, and mostly relevantly, R-parity conserving supersymmetry. The goal of this appendix is to provide a more in depth description of these observables.

C.1 Observables probing supersymmetry

In the following, $q_1^\mu = (E_1, \vec{q}_1)^\mu$ and $q_2^\mu = (E_2, \vec{q}_2)^\mu$ are taken to be the lab frame four-vectors of the visible daughters emanating from heavy particles 1 and 2. $\nu_1^\mu = (E_{\nu 1}, \vec{\nu}_1)^\mu$ and $\nu_2^\mu = (E_{\nu 2}, \vec{\nu}_2)^\mu$ are the unknown four-vectors vectors of the invisible daughters of heavy particles 1 and 2.

It is not required that the new heavy particles decay to exactly two daughters. To accommodate potentially complex decay chains that would give rise to a large number of final state particles, the events are partitioned into two “mega-jets”[41] or hemispheres, where a mega-jet is the four-vector sum of all final state particles within a hemisphere. q_1^μ and q_2^μ are then taken to be the mega-jet four-vectors.

Razor variables. The Razor variables [117] M_R and R^2 have been shown [118] to be a probe of R-parity conserving supersymmetry models. They exploit the fact that should new heavy particles be created as result of the high energy reach of the LHC, they will likely

be produced on threshold, and have nearly zero kinetic energy in the center of momentum frame of the colliding partons. In this case, a single Lorentz boost in the $\hat{\eta}$ direction suffices to translate from the lab frame to the center of momentum frame of the hard interaction, where the conservation of momentum assures that all four daughter systems have the same momentum magnitude. The value of this momentum, written in terms of lab frame observables is referred to as M_R , and is equal to

$$M_R = 2 \sqrt{\frac{(q_{2z}E_1 - q_{1z}E_2)^2}{(q_{1z} - q_{2z})^2 - (E_1 - E_2)^2}}. \quad (\text{C.1})$$

The distribution of M_R in a sample of signal events is a peak somewhere in the vicinity of the mass of the mother particle. In the M_R range of immediate interest, a large contribution from Standard Model processes is expected, so an additional variable R is constructed to suppress these backgrounds, defined as

$$R \equiv M_T^R / M_R. \quad (\text{C.2})$$

where

$$M_T^R = \sqrt{\frac{E_T^{miss}}{2}(q_{1T} + q_{2T}) - \vec{E}_T^{miss} \cdot (\vec{q}_1 + \vec{q}_2)}. \quad (\text{C.3})$$

Events arising from Standard Model processes give rise to a smoothly falling distribution in the R^2 - M_R plane, and signal-like events give rise to a 2-dimensional peak far from the origin. In the case of an observed excess over the Standard Model prediction, an analysis of the extracted signal can provide information about the underlying physics process, including the difference in mass between the new mother and daughter particle.

Stransverse mass M_{T2} . The transverse mass m_T has been used to study events in which a single heavy particle is produced and undergoes a semi-invisible decay—for example, when a W boson decays to a charged lepton and a neutrino. Assuming the detectable daughter has four-momentum $q^\mu = (E, \vec{q})^\mu$ and the invisible daughter's transverse momentum is inferred from the presence of missing transverse energy, the transverse mass is written as

$$m_T(q^\mu, p_T^{miss}) = \sqrt{2q_T E_T^{miss} (1 - \cos(\phi^{miss} - \phi_q))}. \quad (\text{C.4})$$

When m_T is computed for a set of signal events with one invisible particle in the final state, a distribution with a kinematic endpoint is yielded, and the mass of the mother particle can be inferred as the value of the endpoint. This technique has been used to perform some of the most precise measurements of the W boson mass to date [119].

The stransverse mass M_{T2} [86] is an extension of m_T to systems with two invisible particles in the final state instead of one, and is defined as

$$M_{T2} = \min_{\vec{\nu}_1 + \vec{\nu}_2 = \vec{p}_T^{\text{miss}}} [\max\{m_T(q_1^\mu, \nu_1^\mu), m_T(q_2^\mu, \nu_2^\mu)\}]. \quad (\text{C.5})$$

M_{T2} has been shown [120] to not only provide discrimination between Standard Model events and supersymmetry-like events, but to exhibit once again a kinematic endpoint whose value indicates the mass of the heavy mother particle.

Other observables include α_T , [121], the jet mass, and the contransverse mass m_{CT} , and may be worth including in a multivariate discriminant.

APPENDIX D

PUBLIC DATABASE OF LHC SEARCHES

In the last few years, I have contributed to a project to help make results published by CMS and ATLAS more accessible to theorists working outside the collaborations. The LPSC in Grenoble, the institution where the project is based, generously hosted me for periods of a few weeks to a month during the summers of 2013, 2014, and 2015. The following is a selection from the paper [30].

D.1 Toward a public analysis database for LHC new physics searches using MADANALYSIS 5

It is a challenge and in cases impossible for the broader community of physicists, namely theorists, to work out the implications of the LHC results in the context of all these different models, to derive the relevant limits, point out possible loopholes in the current searches, and help design the analyses for the next phase of LHC running at higher energy.

To this end, many groups have been developing private code for the interpretation of the LHC results. Moreover, recently some public tools became available, which serve the whole community. For the interpretation in the context of Simplified Models, there are SMODELS [122] and FASTLIM [123]. SMODELS takes the spectrum of any BSM scenario, decomposes it into SMS topologies, and compares it to the cross section upper limits from more than 50 ATLAS and CMS SMS results. FASTLIM reconstructs the visible cross sections from pre-calculated efficiency and cross section tables for simplified event topologies, currently taking into account 11 ATLAS analyses which mainly focus on searches for supersymmetric partners of the top and bottom quarks (top and bottom squarks, respectively). For confronting simulated events of any model to LHC results, there is CHECKMATE [124]. This program currently has 8 ATLAS and 1 CMS SUSY analyses implemented, which it re-interprets based on fast simulation. Another tool, XQCAT [125], is designed specifically

for testing scenarios with heavy extra quarks, based on a CMS search for top partners as well as two SUSY searches.

In this paper, we follow a complementary approach. We present the implementation of several ATLAS and CMS supersymmetry (SUSY) analyses in MADANALYSIS 5 [126, 127], with simulation of detector effects based on DELPHES 3 [128], and propose to create a public analysis database (PAD) within this framework. MADANALYSIS 5 offers a user-friendly platform for collider phenomenology, and the PAD we propose will be easily accessible to and extendible by the whole community.

In BSM searches, sets of selection criteria are designed in order to maximize the sensitivity to expected signals of new physics. These define so-called signal and control regions, described in the experimental publications. For interpreting a search in the context of a given new physics model, one has to implement these selection criteria together with a description of the detector performance (emulating the various object definitions and efficiencies) in a simulation tool. Based on simulated event samples for the model being tested, the expected number of signal events in the various signal regions (SRs) can then be computed and compared to the number of observed events and the number of expected SM background events, which are reported in the experimental publication.

Non-collaboration members however do not have access to the experimental data, nor the Monte Carlo (MC) event set simulated with an official collaboration detector simulation. This renders the implementation and validation of ATLAS and CMS analyses for re-interpretation of the experimental results in general contexts a tedious task, even more so as the information given in the experimental papers is often incomplete (we will comment more on this in Section 3). We therefore think that a common platform for collecting object definitions, cuts, and all other information necessary to reproduce or use the results of the analyses will be of great value to the high-energy physics community. Moreover, as our project follows an Open Access and Open Data approach, we hope that it will benefit the scientific communication and in particular motivate ATLAS and CMS to provide more information on their analyses, in line with the Les Houches Recommendations [129].

D.2 Implemented analyses and their validation

To start the analysis database, we have implemented and validated the following ATLAS and CMS SUSY searches at $\sqrt{s} = 8$ TeV and an integrated luminosity of about 20 fb^{-1} :

ATLAS:

- Search for top squarks and sbottoms in final states with no lepton and two b -jets [130]:
ATLAS-SUSY-2013-05;
- Search for charginos, neutralinos and sleptons in final states with two leptons [131]:
ATLAS-SUSY-2013-11;

CMS:

- Search for top squarks in the single-lepton final state [88]: CMS-SUS-13-011;
- Search for gluinos and squarks in events with three or more jets and E_T^{miss} [107]: CMS-SUS-13-012;
- Search for gluinos in opposite-sign dilepton events, large number of jets, b -jets and E_T^{miss} [9]: CMS-SUS-13-016.

Several more analyses are currently being implemented and validated.

A list of all available analyses, which is evolving, instructions on how to use them, as well as more detailed validation notes can be found on the MADANALYSIS 5 wiki page [132]. The recast scripts themselves are published via INSPIRE [133], in order to make them citable (INSPIRE assigns each submission a DOI [134]) and to ensure that changes can be traced reliably through a rigorous versioning system.

Before proceeding, some general comments are in order. Generally, we cannot reproduce cleaning cuts (for, *e.g.*, cosmic rays and beam effects). Moreover, some basic jet quality

criteria must be skipped as we do not have vertex information. This is, however, expected to have a small impact on signal events. In addition, event weights are typically applied by ATLAS and CMS to correct simulated events with respect to data. We take such event weights into account whenever they are available. Otherwise they are neglected and contribute to the overall uncertainty of the procedure.

Finally, while the selection criteria that define the various SRs are usually clear and well documented, information on the preselection cuts is often missing. In particular, trigger efficiencies, information about isolation, efficiencies for leptons, and the order in which preselection cuts are applied is crucial for reliably reproducing an analysis, but this information is often incomplete in the experimental publications. We hope that this will improve over time and the necessary information will be given systematically either in the physics paper or in a performance note, as also advertised in [129].

Below we give some details on an example of a CMS analysis validated in the framework of MADANALYSIS 5, namely, the CMS multi-jet + H_T^{miss} search permed at $\sqrt{s} = 8$ TeV. the level of documentation by the experimental collaboration, and the validation of our MADANALYSIS 5 implementations.

D.2.1 CMS-SUS-13-012: search for new physics through jet multiplicity and missing energy

This CMS search for new physics in the hadronic activity in events with no leptons [107] targets a number of different signal topologies, in particular:

- gluino-pair production with $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_1^0$, denoted as T1qqqq topology in the following;
- gluino-pair production with $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$, denoted as T1tttt;
- gluino-pair production with $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_2^0/\tilde{\chi}_1^\pm$, followed by $\tilde{\chi}_2^0, \tilde{\chi}_1^\pm \rightarrow Z/W\tilde{\chi}_1^0$, generically denoted as T5VV; and
- squark-pair production with $\tilde{q} \rightarrow q\tilde{\chi}_1^0$, denoted as T2qq,

following the CMS simplified models naming scheme [135].

The analysis comprises 36 non-overlapping signal regions, each one defined as a rectangular box volume in the space spanned by the variables N_{jets} , H_T , and H_T^{miss} . Here N_{jets} is the jet multiplicity of the event, H_T is the scalar sum of the jet transverse momenta, and H_T^{miss} is the magnitude of the vector sum of the jets transverse momenta. Explicitly,

$$H_T = \sum_{\text{jets}} p_T, \quad H_T^{\text{miss}} = |\vec{H}_T^{\text{miss}}| = \left| \sum_{\text{jets}} \mathbf{p}_T \right|. \quad (\text{D.1})$$

The event selection was primarily determined from the documentation in [107]. This document describes six baseline selection criteria on the events, named MET Cleaning, No Lepton, $N_{\text{jets}} > 2$, $H_T > 500\text{GeV}$, $H_T^{\text{miss}} > 200\text{GeV}$, and Min $\Delta\phi(\text{jets}, \vec{H}_T^{\text{miss}})$. We note that the MET Cleaning cut involves a detailed consideration of spurious signals in the CMS detector, which we cannot simulate with DELPHES. Instead, we simply multiply our event count by the efficiency given by CMS. (We stress again that such efficiencies being publicly available is extremely helpful.)

We validated the recast code against cut-flow tables and distributions of the kinematic variables provided by the CMS analysis team as per our request. The benchmark scenarios used are $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)\text{GeV}$ for the T1qqqq, T1tttt and T5VV topologies, and $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0}) = (700, 100)\text{GeV}$ for the T2qq topology, with production cross sections of 10.2 fb and 63.4 fb, respectively [136]. For the T5VV topology, one also needs the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ masses; they are set to 612.5GeV for the $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)$ GeV benchmark point.

The complete validation material from CMS is available in form of the PDF documents `ma5_validation_CMS-SUS-13-012.pdf` in the “Attachments” section on the analysis’ wiki page [132], and also featured in Section D.5. These files correspond to the simplified SUSY models of the same names. For each of the four simplified-model scenarios, the CMS collaboration provided us with 10^5 events in LHE format along with cut-flow tables and distributions in the variables N_{jets} , H_T , and H_T^{miss} after each cut. As before, we passed these LHE files to PYTHIA 6.4 [101] for showering and hadronization and finally to DELPHES for detector simulation. The merging of the partonic events that exhibit different jet multiplicities was performed according to the setup read from the LHE files provided by CMS.

Table D.1: Summary of yields for the baseline cuts for the T1qqqq and T1tttt topologies, as compared to the official CMS-SUS-13-012 results posted on the CMS twiki. The results are for the $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)$ GeV benchmark point.

cut	T1qqqq		T1tttt	
	CMS result	MA 5 result	CMS result	MA5 result
MET Cleaning	190.6	190.6	190.5	190.5
No Lepton	190.3	190.6	95.9	101.0
+ $N_{\text{jets}} > 2$	188.1	188.5	95.8	100.9
+ $H_T > 500$ GeV	187.6	188.1	95.1	100.0
+ $H_T^{\text{miss}} > 200$ GeV	158.7	159.7	75.4	81.2
+ Min $\Delta(\phi)$	130.8	131.1	62.3	66.9

Table D.2: Same as Table D.1 but for the T2qq and T5VV topologies. The benchmark points used are $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0}) = (700, 100)$ GeV for T2qq and $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)$ GeV for T5VV.

cut	T2qq		T5VV	
	CMS result	MA 5 result	CMS result	MA5 result
MET Cleaning	1215.2	1215.2	189.9	189.9
No Lepton	1212.8	1215.2	136.2	142.1
+ $N_{\text{jets}} > 2$	675.9	691.5	135.9	141.7
+ $H_T > 500$ GeV	619.5	638.4	135.5	141.3
+ $H_T^{\text{miss}} > 200$ GeV	524.0	539.6	108.8	115.2
+ Min $\Delta(\phi)$	460.7	476.1	89.6	95.2

A detail that required additional correspondence with the CMS analysis team were the pseudorapidity (η) cuts on the electrons and muons used for the lepton veto. We learned that the only requirement on these leptons is that $|\eta| < 2.4$, and they are allowed to reside in the overlap region between the electromagnetic calorimeter barrel and the endcap. We also checked the dependence on the jet energy scale (JES) correction, which is set in the CMS DELPHES card, to have good agreement in the N_{jets} , H_T and H_T^{miss} distributions, and found JES=1.0 to be optimal.

The results of our cut-flow counts for the various simplified models are shown alongside the official counts in Tables D.1 and D.2. The results were obtained by normalizing with the cross section for each of the benchmark points and for an integrated luminosity of 19.3 fb^{-1} . Moreover, some distributions after the baseline cuts for the case of the T2qq topology are shown in Figs. D.1–D.3. The distributions are normalized to unity and overlaid on the official plots obtained from the collaboration.

The agreement between the official and MADANALYSIS 5 results is better than 10% throughout the baseline cut flows. The largest discrepancy arises from the lepton veto cut, which leads to a difference of up to about 5% in the cut flow. The shapes of the distributions qualitatively match very well, and the peaking bins are in accordance with the official results. (This also holds for the other distributions not shown here for space considerations). The MADANALYSIS 5 implementation is available as [137], and a detailed validation note comparing the recast results to those of CMS can be found at [132].

The full validation document, along with the validation document for SUS-14-002, are given in Sections D.5 and D.6.

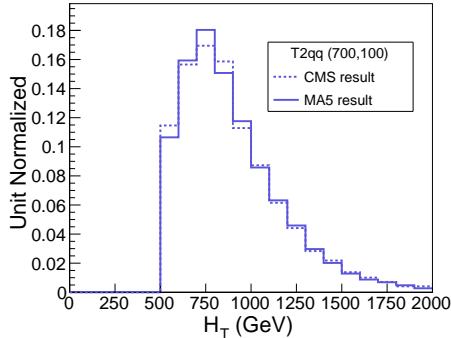


Figure D.1: Comparison between the official and MADANALYSIS 5 results for the H_T distribution after all baseline cuts, for the T2qq simplified model of CMS-SUS-13-012 with $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0}) = (700, 100)\text{GeV}$.

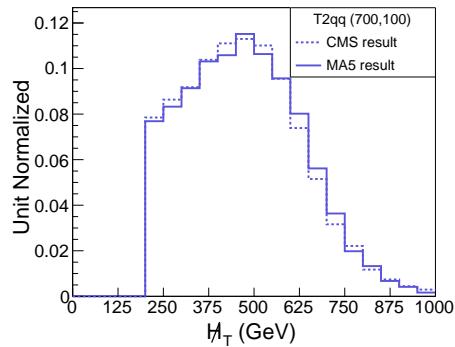


Figure D.2: Same as Fig. D.1 but for the H_T^{miss} distribution.

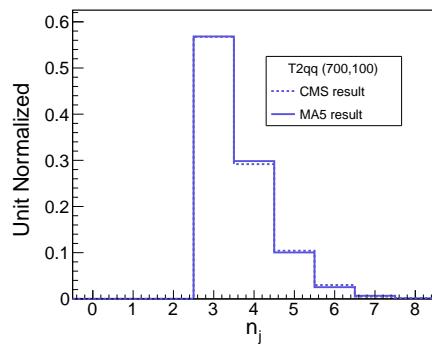


Figure D.3: Same as Fig. D.1 but for the N_{jets} distribution.

D.3 Limit setting

For the statistical interpretation of the results, we provide on [132] a PYTHON script, `exclusion_CLs.py`, for computing exclusion limits using the CL_s prescription [113].¹ This code can also be installed on a user system by typing in, from the MADANALYSIS 5 interpreter, the command

```
install RecastingTools
```

which results in the file `exclusion_CLs.py` being present at the root of any working directory created in the expert mode of MADANALYSIS 5. We refer to [127, 132] for details on the creation of MADANALYSIS 5 working directories.

The core of the calculation works as follows. First, the number of signal events (n_s) is obtained as the product of the luminosity, signal cross section and acceptance \times efficiency for the SR of interest. Then, the number of observed events (n_{obs}) and the nominal number of background events (\hat{n}_b) and its uncertainty (Δn_b), are retrieved from an information card stored in the Analyzer directory, provided by the software package. A large number of toy MC experiments (10^5 by default) are then generated from the Poisson distribution $\text{poiss}(n_{\text{obs}}|n_{\text{expected}})$, corresponding to the distribution of the total number of events in the SR under the background-only hypothesis on the one hand ($n_{\text{expected}} = n_b$), and under the signal + background hypothesis ($n_{\text{expected}} = n_s + n_b$) on the other hand. We assume that the uncertainty on the number of background events is modeled as $\text{gauss}(\hat{n}_b, \Delta n_b)$, and for each toy MC the number of background events n_b is randomly generated from this normal distribution. Under the two different hypotheses, p -values are then calculated using the number of events actually observed at the LHC, and finally used to compute the CL_s value.

We have tested the limit-setting code on the analyses presented in this paper and generally found good agreement with the official exclusions from ATLAS and CMS. As an illustrative example, Fig. D.4 shows the 95% CL exclusion limit in the neutralino versus gluino mass plane reproduced with the MADANALYSIS 5 implementation [8] of CMS-SUS-13-016.

¹The PYTHON code requires SciPy libraries to be installed.

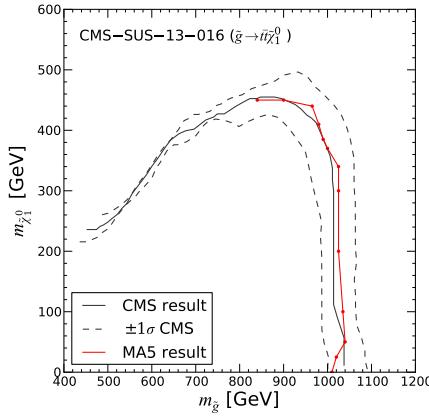


Figure D.4: The 95% CL exclusion limit (in red) in the $\tilde{\chi}_1^0$ versus \tilde{g} mass plane reproduced with the MADANALYSIS 5 implementation [8] of CMS-SUS-13-016. For comparison, the full and dashed grey lines show the official CMS result with its $\pm 1\sigma$ uncertainty from Fig. 6 of [9]. The limit setting in the region where one of the tops from the gluino decay is off-shell, *i.e.* for $m_{\tilde{g}} \lesssim 800$ GeV, is work in progress.

D.4 MadAnalysis 5 conclusions

We have presented a new scheme for developing and deploying implementations of LHC analyses based on fast simulation within the MADANALYSIS 5 framework. This can serve to create a public analysis database, which may be used and developed further by the whole community. The code for the five analyses [138, 137, 8, 130, 131] that we published together with this paper are intended as a starting point for this database and may conveniently be used as templates for other analyses.

We propose that the C++ code of new implementations within this scheme be published via INSPIRE [133], as done here, best together with the physics paper they have been developed for. This way, each analysis implementation is assigned a Digital Object Identifier (DOI) [134], ensuring that it is uniquely identifiable, searchable and citable. In addition it is very useful if a detailed validation note is made available on the MADANALYSIS 5 wiki page [132].

The ease with which an experimental analysis can be implemented and validated may serve as a useful check for the experimental collaborations for the quality of their docu-

mentation. Note, finally, that the platform we are proposing might also be used by the experimental collaborations to directly provide implementations of their analyses for fast simulation, thereby assuring the maximum usability of their results, as for example envisaged in level 1 of the CMS statement on “data preservation, re-use and open access policy” [139].

The author of a re-implementation could be a new member of CMS or ATLAS who wishes to learn about an analysis, or a who has recently carried out an analysis and wants the analysis to have a larger impact. Such a volunteer must read the relevant CMS paper and implement the object and event selection in a relatively easy-to-use framework based on c++, and validate their implementation by comparing their results for cut flows and exclusion limit curves with those that CMS makes public. Sample validation material has been given above, and examples of complete validation documents are provided in Sections D.5 and D.6.

It is important for the legacy of the LHC that its experimental results can be used by the whole high-energy physics community. We hope that our project contributes to this aim.

D.5 Validation document: SUS-13-012

The following describes the work documented in [137], in association with the paper [137].

Bein, Samuel (Florida State U.); Sengupta, Dipan (LPSC, Grenoble)

We present the results of the synchronization of the MA5 implementation of the SUS-13-012 multi-jet + H_T^{miss} SUSY search for SUSY in data collected by the CMS experiment at $\sqrt{s} = 8$ TeV. The performance of the implementation is evaluated by comparing the MA5-derived results with a set of cut flow tables and kinematic distributions provided by CMS for this purpose. The simplified models T1qqqq (Figure D.5), T1tttt (Figure D.10), T5VV (Figure D.15), and T2qq (Figure D.20), are used as common benchmarks, with values of the masses of the gluino and neutralino taking on a range of values. Two types of comparisons are given in Figs. D.3–D.24 give cut flow tables and normalized kinematic distributions. In some cases the cut flow tables give the number of events normalized to 100%; in other cases the tables are normalized to the cross section times the integrated luminosity. The

normalization convention used by CMS was followed. Dashes hold the place of values that were not provided in the CMS cut flow tables.

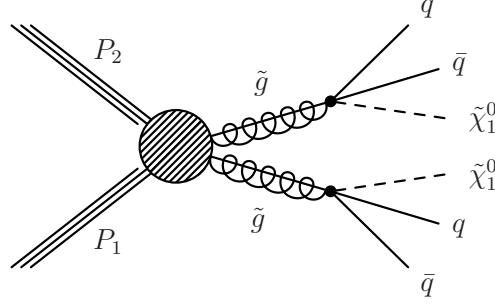


Figure D.5: Diagram of the dominant SUSY production mechanism for the T1qqqq signal model.

Cut Name	Official Count (Eff)	MA5 Count (Eff)
MET Cleaning	190.6 (xxx)	190.6 (xxx)
No Lepton	190.3 (99%)	190.6 (100%)
NJets>2	188.1 (98%)	188.49 (98%)
$H_T > 500$	187.6 (99%)	188.07 (99%)
$H_T^{\text{miss}} > 200$	158.7 (84%)	159.72 (84%)
Min $\Delta(\phi)$	130.8 (82%)	131.11 (82%)

Table D.3: The cut flow for the baseline selection in CMS SUS-13-012 for the T1qqqq working point $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)\text{GeV}$. The second column is the official account and our own results are given in column 3. The official counts are normalized to luminosity $\mathcal{L} = 19.5/\text{fb}$ and cross section $\sigma = 10.17 \text{ pb}$, and our counts are normalized to match the official count after the first cut, MET Cleaning.

T1qqqq simplified model.

Signal Region Name	Official	MA5
NJets3-5, H_T 500-800, H_T^{miss} 200-300	1.4	1.21
NJets3-5, H_T 500-800, H_T^{miss} 300-450	2.4	2.08
NJets3-5, H_T 500-800, H_T^{miss} 450-600	1.7	1.36
NJets3-5, H_T 500-800, $H_T^{\text{miss}}>600$	0.6	0.60
NJets3-5, H_T 800-1000, H_T^{miss} 200-300	2.1	1.81
NJets3-5, H_T 800-1000, H_T^{miss} 300-450	2.9	3.75
NJets3-5, H_T 800-1000, H_T^{miss} 450-600	4.2	3.74
NJets3-5, H_T 800-1000, $H_T^{\text{miss}}>600$	4.1	4.04
NJets3-5, H_T 1000-1250, H_T^{miss} 200-300	4.2	3.70
NJets3-5, H_T 1000-1250, H_T^{miss} 300-450	8.1	6.93
NJets3-5, H_T 1000-1250, H_T^{miss} 450-600	7.6	7.18
NJets3-5, H_T 1000-1250, $H_T^{\text{miss}}>600$	10.6	10.63
NJets3-5, H_T 1250-1500, H_T^{miss} 200-300	3.9	3.64
NJets3-5, H_T 1250-1500, H_T^{miss} 300-450	7.3	6.74
NJets3-5, H_T 1250-1500, $H_T^{\text{miss}}>450$	15.6	16.52
NJets3-5, H_T >1500, H_T^{miss} 200-300	4.5	4.41
NJets3-5, H_T >1500, $H_T^{\text{miss}}>300$	17.9	18.80
NJets6-7, H_T 500-800, H_T^{miss} 200-300	0.1	0.08
NJets6-7, H_T 500-800, H_T^{miss} 300-450	0.1	0.05
NJets6-7, H_T 500-800, $H_T^{\text{miss}}>450$	0.1	0.04
NJets6-7, H_T 800-1000, H_T^{miss} 200-300	0.3	0.24
NJets6-7, H_T 800-1000, H_T^{miss} 300-450	0.6	0.51
NJets6-7, H_T 800-1000, $H_T^{\text{miss}}>450$	0.8	0.71
NJets6-7, H_T 1000-1250, H_T^{miss} 200-300	0.9	0.91
NJets6-7, H_T 1000-1250, H_T^{miss} 300-450	1.8	1.74
NJets6-7, H_T 1000-1250, $H_T^{\text{miss}}>450$	2.8	2.94
NJets6-7, H_T 1250-1500, H_T^{miss} 200-300	1.2	1.16
NJets6-7, H_T 1250-1500, H_T^{miss} 300-450	2.4	2.46
NJets6-7, H_T 1250-1500, $H_T^{\text{miss}}>450$	4.1	5.16
NJets6-7, H_T >1500, H_T^{miss} 200-300	2.3	2.56
NJets6-7, H_T >1500, $H_T^{\text{miss}}>300$	9.8	11.50
NJets>7, H_T 500-800, $H_T^{\text{miss}}>200$	0.0	0.0
NJets>7, H_T 800-1000, $H_T^{\text{miss}}>200$	0.0	0.01
NJets>7, H_T 1000-1250, $H_T^{\text{miss}}>200$	0.2	0.28
NJets>7, H_T 1250-1500, $H_T^{\text{miss}}>200$	0.5	0.75
NJets>7, H_T >1500, $H_T^{\text{miss}}>200$	2.2	2.69

Table D.4: The signal region (SR) counts in CMS SUS-13-012 for the T1qqqq scenario after all selection has been applied. Column 2 is the official account obtained through generous correspondence with Christian Sanders, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.3.

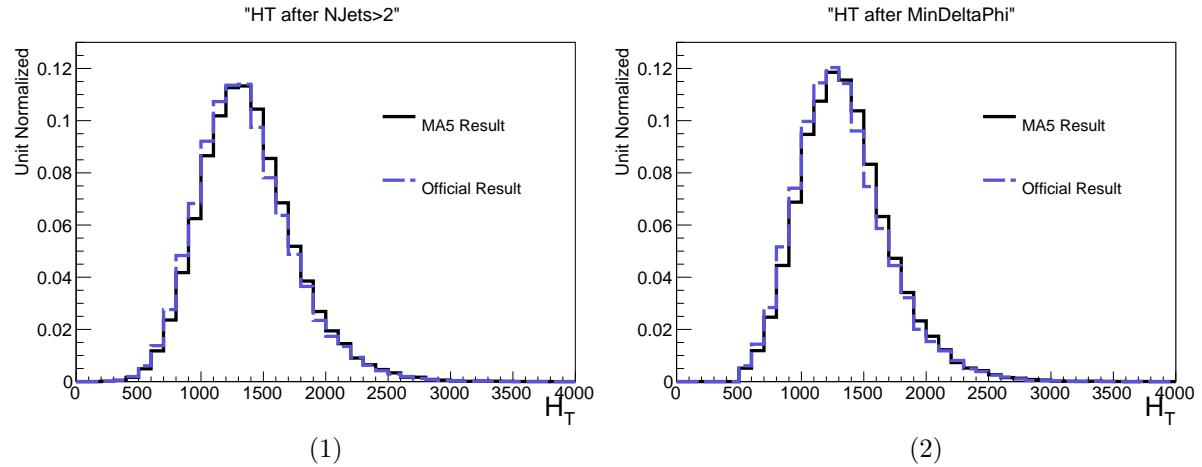


Figure D.6: Comparison of the distributions of H_T between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1qqqq signal model.

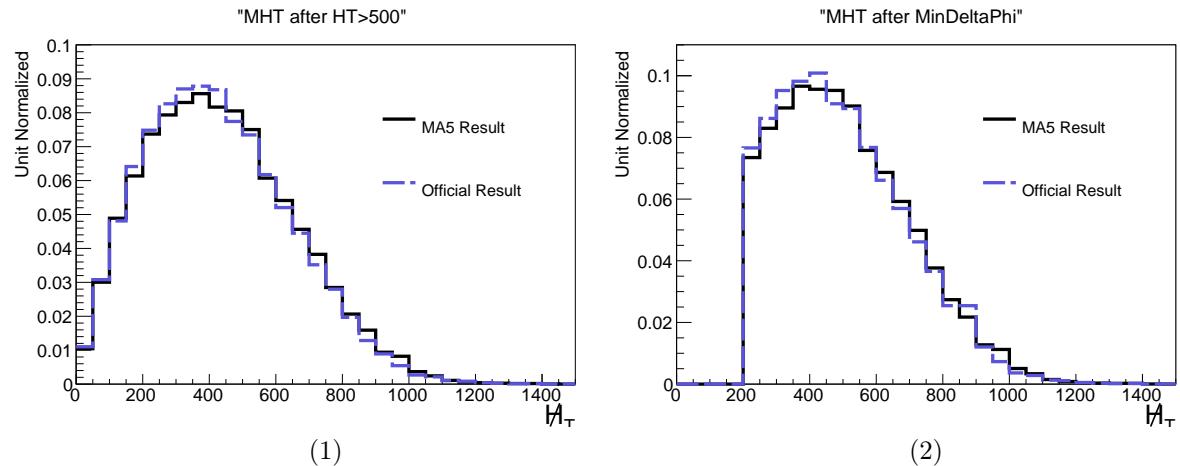


Figure D.7: Comparison of the distributions of H_T^{miss} between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1qqqq signal model.

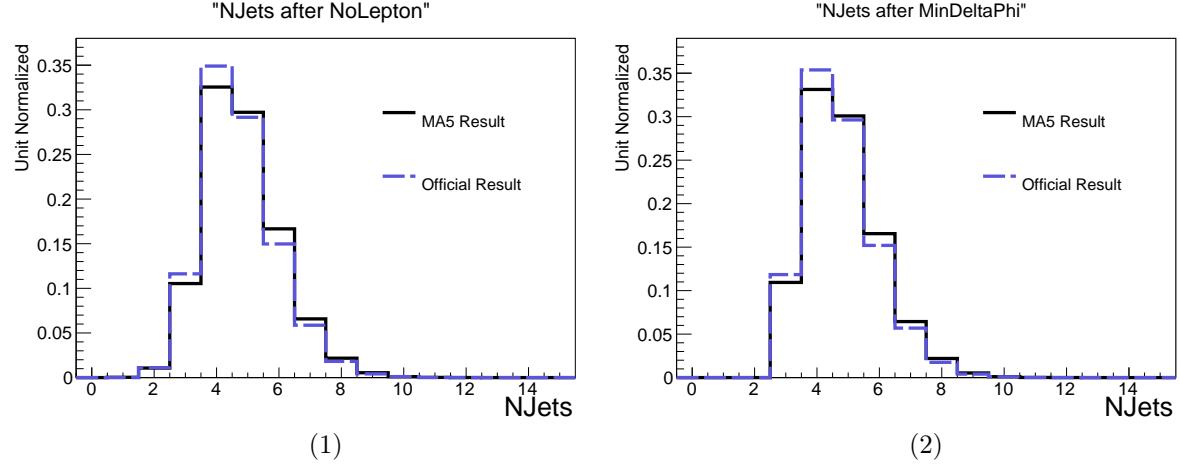


Figure D.8: Comparison of the distributions of NJets between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1qqqq signal model.

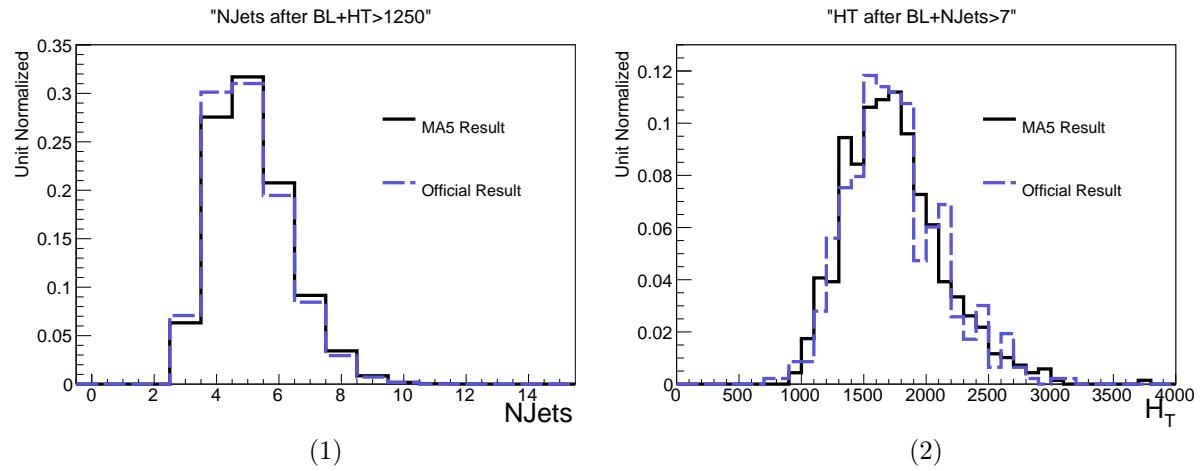


Figure D.9: Additional checks: comparison between ours and the official distributions of NJets after $\text{BL} + H_T > 1250$ cuts (left), and H_T after $\text{BL} + \text{NJets} > 7$ cuts (right), for the T1qqqq signal model.

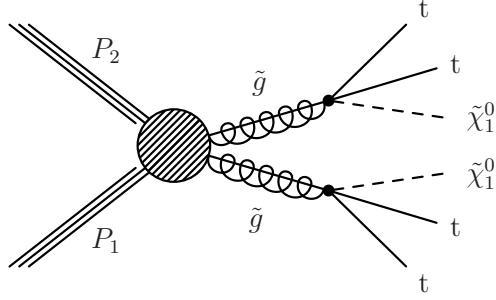


Figure D.10: Diagram of the dominant SUSY production mechanism for the T1tttt signal model.

Cut Name	Official Count (Eff)	MA5 Count (Eff)
MET Cleaning	190.5 (xxx)	190.5 (xxx)
No Lepton	95.9 (50%)	101.04 (53%)
NJets>2	95.8 (99%)	100.87 (99%)
$H_T > 500$	95.1 (99%)	100.01 (99%)
$H_T^{\text{miss}} > 200$	75.4 (79%)	81.23 (81%)
Min $\Delta(\phi)$	62.3 (82%)	66.92 (82%)

Table D.5: The cut flow for the baseline selection in CMS SUS-13-012 for the T1tttt working point $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)\text{GeV}$. The second column is the official account and our own results are given in column 3. The official counts are normalized to luminosity $\mathcal{L} = 19.5/\text{fb}$ and cross section $\sigma = 10.17 \text{ pb}$, and our counts are normalized to match the official count after the first cut, MET Cleaning.

T1tttt simplified model.

Signal Region Name	Official	MA5
NJets3-5, H_T 500-800, H_T^{miss} 200-300	0.8	0.85
NJets3-5, H_T 500-800, H_T^{miss} 300-450	1.4	1.22
NJets3-5, H_T 500-800, H_T^{miss} 450-600	0.8	0.85
NJets3-5, H_T 500-800, $H_T^{\text{miss}} > 600$	0.2	0.31
NJets3-5, H_T 800-1000, H_T^{miss} 200-300	0.5	0.45
NJets3-5, H_T 800-1000, H_T^{miss} 300-450	0.7	1.00
NJets3-5, H_T 800-1000, H_T^{miss} 450-600	1.0	1.03
NJets3-5, H_T 800-1000, $H_T^{\text{miss}} > 600$	0.8	0.79
NJets3-5, H_T 1000-1250, H_T^{miss} 200-300	0.5	0.53
NJets3-5, H_T 1000-1250, H_T^{miss} 300-450	1.0	0.83
NJets3-5, H_T 1000-1250, H_T^{miss} 450-600	0.8	0.87
NJets3-5, H_T 1000-1250, $H_T^{\text{miss}} > 600$	0.9	1.01
NJets3-5, H_T 1250-1500, H_T^{miss} 200-300	0.4	0.40
NJets3-5, H_T 1250-1500, H_T^{miss} 300-450	0.5	0.58
NJets3-5, H_T 1250-1500, $H_T^{\text{miss}} > 450$	0.8	0.81
NJets3-5, $H_T > 1500$, H_T^{miss} 200-300	0.3	0.34
NJets3-5, $H_T > 1500$, $H_T^{\text{miss}} > 300$	0.9	1.01
NJets6-7, H_T 500-800, H_T^{miss} 200-300	0.9	0.81
NJets6-7, H_T 500-800, H_T^{miss} 300-450	1.2	0.85
NJets6-7, H_T 500-800, $H_T^{\text{miss}} > 450$	0.6	0.44
NJets6-7, H_T 800-1000, H_T^{miss} 200-300	1.5	1.16
NJets6-7, H_T 800-1000, H_T^{miss} 300-450	2.5	2.35
NJets6-7, H_T 800-1000, $H_T^{\text{miss}} > 450$	2.5	2.59
NJets6-7, H_T 1000-1250, H_T^{miss} 200-300	1.8	1.71
NJets6-7, H_T 1000-1250, H_T^{miss} 300-450	3.4	3.37
NJets6-7, H_T 1000-1250, $H_T^{\text{miss}} > 450$	4.5	5.21
NJets6-7, H_T 1250-1500, H_T^{miss} 200-300	1.4	1.46
NJets6-7, H_T 1250-1500, H_T^{miss} 300-450	2.2	2.43
NJets6-7, H_T 1250-1500, $H_T^{\text{miss}} > 450$	2.8	3.34
NJets6-7, $H_T > 1500$, H_T^{miss} 200-300	1.1	1.16
NJets6-7, $H_T > 1500$, $H_T^{\text{miss}} > 300$	3.4	3.99
NJets>7, H_T 500-800, $H_T^{\text{miss}} > 200$	0.2	0.15
NJets>7, H_T 800-1000, $H_T^{\text{miss}} > 200$	1.9	1.69
NJets>7, H_T 1000-1250, $H_T^{\text{miss}} > 200$	5.7	6.37
NJets>7, H_T 1250-1500, $H_T^{\text{miss}} > 200$	5.9	7.28
NJets>7, $H_T > 1500$, $H_T^{\text{miss}} > 200$	6.0	7.53

Table D.6: The signal region (SR) counts in CMS SUS-13-012 for the T1tttt scenario after all selection has been applied. Column 2 is the official account obtained through generous correspondence with Christian Sanders, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.5.

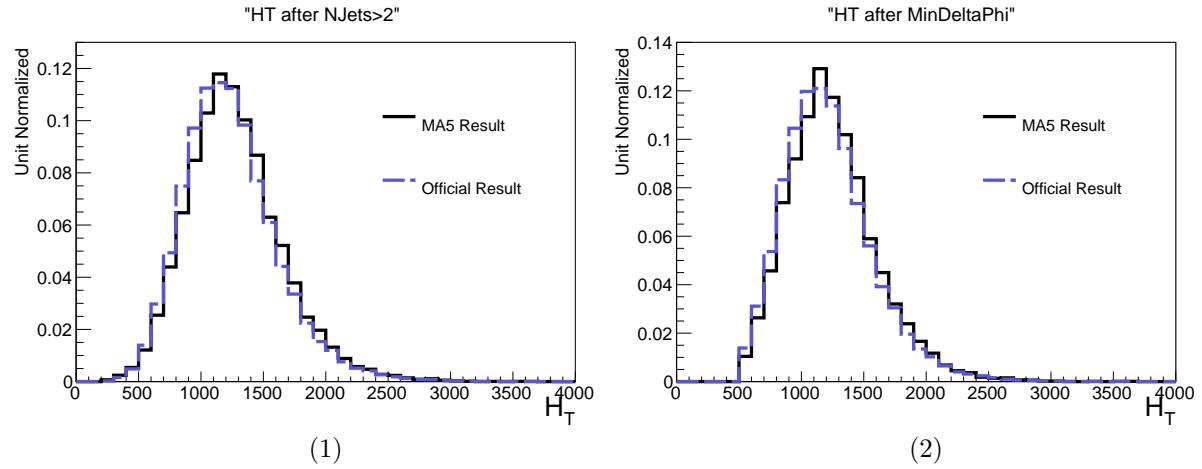


Figure D.11: Comparison of the distributions of H_T between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1tttt signal model.

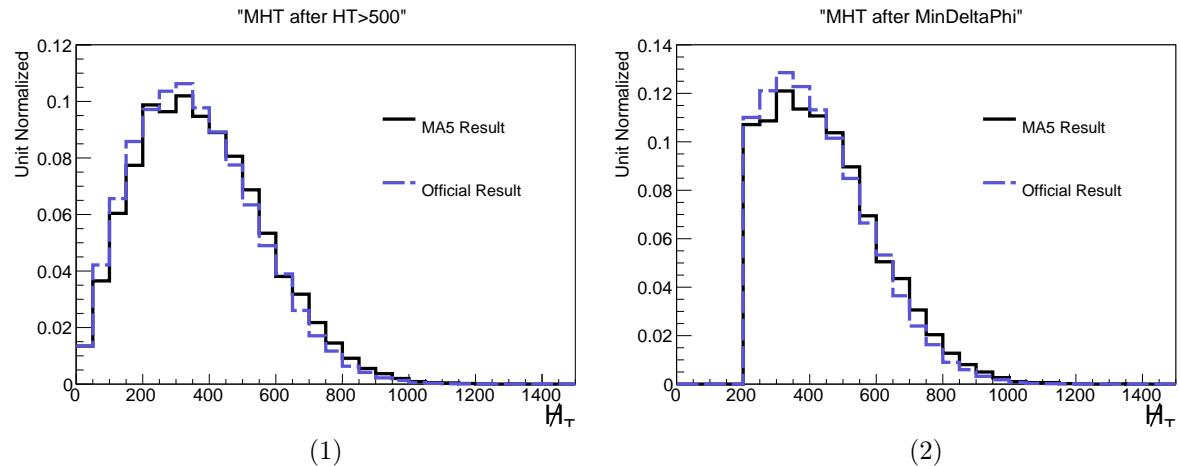


Figure D.12: Comparison of the distributions of H_T^{miss} between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1tttt signal model.

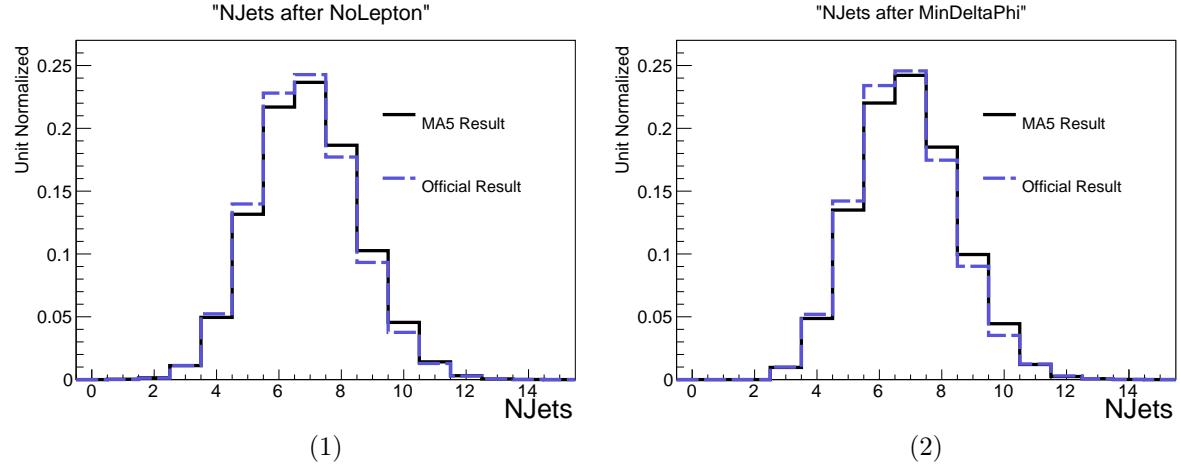


Figure D.13: Comparison of the distributions of NJets between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T1tttt signal model.

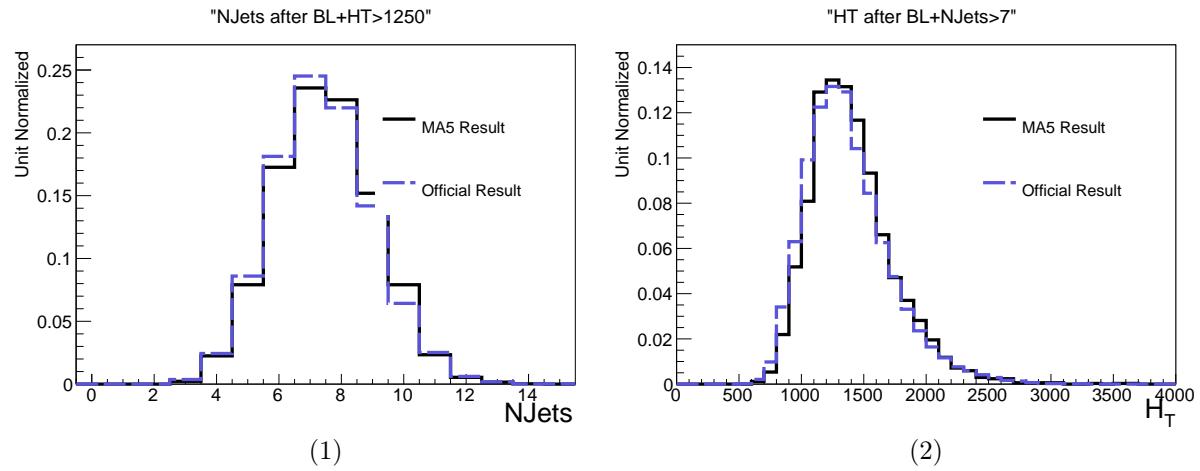


Figure D.14: Additional checks: comparison between ours and the official distributions of NJets after $BL+H_T > 1250$ cuts (left), and H_T after $BL+NJets > 7$ cuts (right), for the T1tttt signal model.

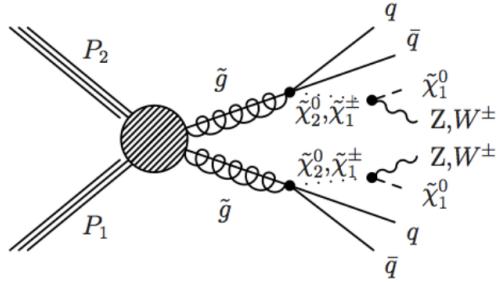


Figure D.15: Diagram of the dominant SUSY production mechanism for the T5VV signal model.

Cut Name	Official Count (Eff)	MA5 Count (Eff)
MET Cleaning	189.9 (xxx)	189.9 (xxx)
No Lepton	136.2 (71%)	142.07 (74%)
NJets>2	135.9 (99%)	141.69 (99%)
$H_T > 500$	135.5 (99%)	141.26 (99%)
$H_T^{\text{miss}} > 200$	108.8 (80%)	115.23 (81%)
Min $\Delta(\phi)$	89.6 (82%)	95.22 (82%)

Table D.7: The cut flow for the baseline selection in CMS SUS-13-012 for the T5VV working point $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0}) = (1100, 125)\text{GeV}$. The second column is the official account and our own results are given in column 3. The official counts are normalized to luminosity $\mathcal{L} = 19.5/\text{fb}$ and cross section $\sigma = 10.17 \text{ pb}$, and our counts are normalized to match the official count after the first cut, MET Cleaning.

T5VV simplified model.

Signal Region Name	Official	MA5
NJets3-5, H_T 500-800, H_T^{miss} 200-300	1.0	1.18
NJets3-5, H_T 500-800, H_T^{miss} 300-450	1.8	1.77
NJets3-5, H_T 500-800, H_T^{miss} 450-600	1.1	1.09
NJets3-5, H_T 500-800, $H_T^{\text{miss}} > 600$	0.3	0.31
NJets3-5, H_T 800-1000, H_T^{miss} 200-300	1.5	1.08
NJets3-5, H_T 800-1000, H_T^{miss} 300-450	1.7	2.40
NJets3-5, H_T 800-1000, H_T^{miss} 450-600	2.1	2.12
NJets3-5, H_T 800-1000, $H_T^{\text{miss}} > 600$	1.2	1.43
NJets3-5, H_T 1000-1250, H_T^{miss} 200-300	1.9	1.84
NJets3-5, H_T 1000-1250, H_T^{miss} 300-450	3.1	3.23
NJets3-5, H_T 1000-1250, H_T^{miss} 450-600	2.8	2.66
NJets3-5, H_T 1000-1250, $H_T^{\text{miss}} > 600$	2.1	2.41
NJets3-5, H_T 1250-1500, H_T^{miss} 200-300	1.3	1.35
NJets3-5, H_T 1250-1500, H_T^{miss} 300-450	2.3	2.03
NJets3-5, H_T 1250-1500, $H_T^{\text{miss}} > 450$	3.2	3.67
NJets3-5, H_T >1500, H_T^{miss} 200-300	1.1	1.06
NJets3-5, H_T >1500, $H_T^{\text{miss}} > 300$	3.7	3.77
NJets6-7, H_T 500-800, H_T^{miss} 200-300	0.4	0.29
NJets6-7, H_T 500-800, H_T^{miss} 300-450	0.4	0.32
NJets6-7, H_T 500-800, $H_T^{\text{miss}} > 450$	0.2	0.15
NJets6-7, H_T 800-1000, H_T^{miss} 200-300	1.2	1.06
NJets6-7, H_T 800-1000, H_T^{miss} 300-450	1.9	1.73
NJets6-7, H_T 800-1000, $H_T^{\text{miss}} > 450$	1.7	1.65
NJets6-7, H_T 1000-1250, H_T^{miss} 200-300	3.1	2.66
NJets6-7, H_T 1000-1250, H_T^{miss} 300-450	4.6	4.72
NJets6-7, H_T 1000-1250, $H_T^{\text{miss}} > 450$	5.9	5.77
NJets6-7, H_T 1250-1500, H_T^{miss} 200-300	2.7	2.89
NJets6-7, H_T 1250-1500, H_T^{miss} 300-450	4.4	4.72
NJets6-7, H_T 1250-1500, $H_T^{\text{miss}} > 450$	5.8	6.57
NJets6-7, H_T >1500, H_T^{miss} 200-300	2.7	3.01
NJets6-7, H_T >1500, $H_T^{\text{miss}} > 300$	9.2	10.94
NJets>7, H_T 500-800, $H_T^{\text{miss}} > 200$	0.0	0.01
NJets>7, H_T 800-1000, $H_T^{\text{miss}} > 200$	0.4	0.33
NJets>7, H_T 1000-1250, $H_T^{\text{miss}} > 200$	2.3	2.50
NJets>7, H_T 1250-1500, $H_T^{\text{miss}} > 200$	3.8	4.48
NJets>7, H_T >1500, $H_T^{\text{miss}} > 200$	6.0	7.84

Table D.8: The signal region (SR) counts in CMS SUS-13-012 for the T5VV scenario after all selection has been applied. Column 2 is the official account obtained through generous correspondence with Christian Sanders, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.7.

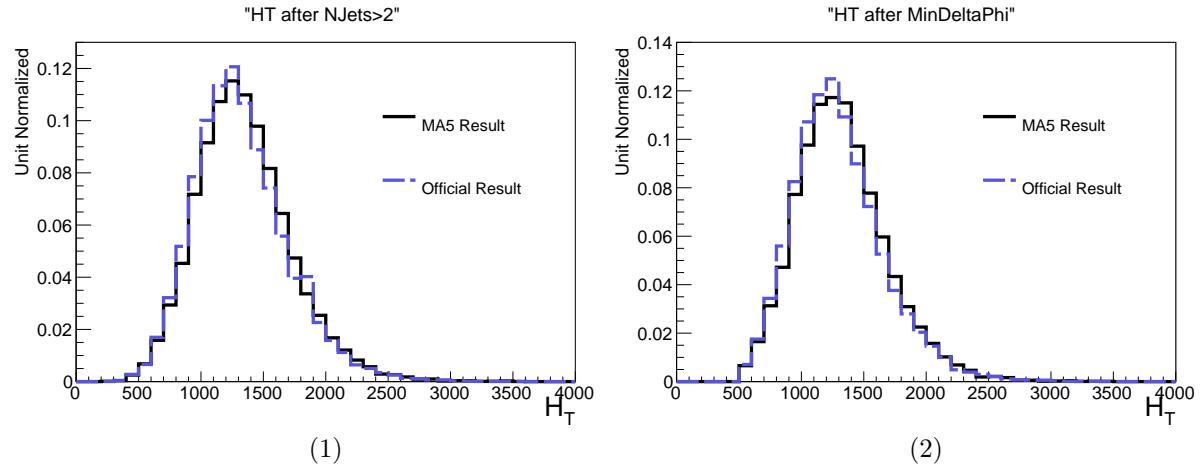


Figure D.16: Comparison of the distributions of H_T between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T5VV signal model.

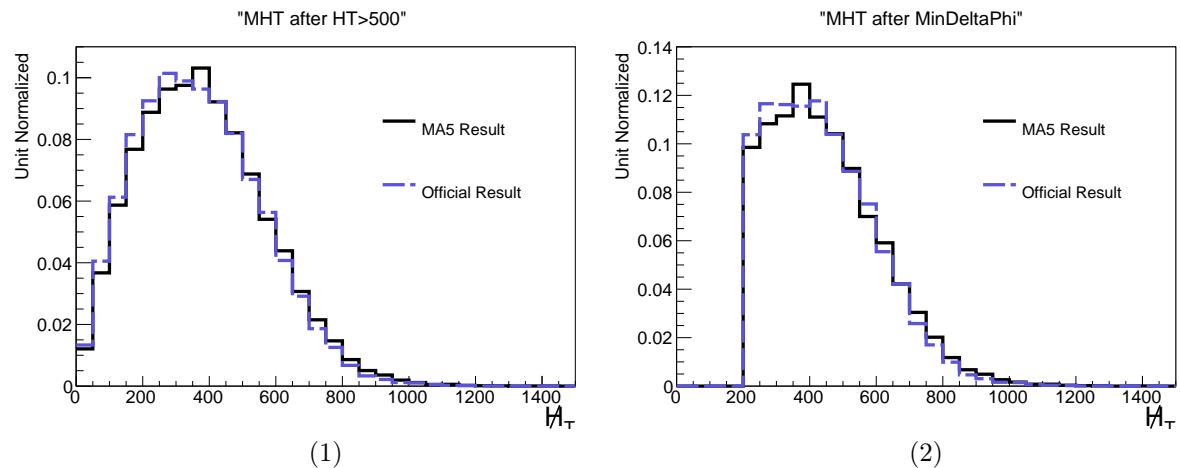


Figure D.17: Comparison of the distributions of H_T^{miss} between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T5VV signal model.

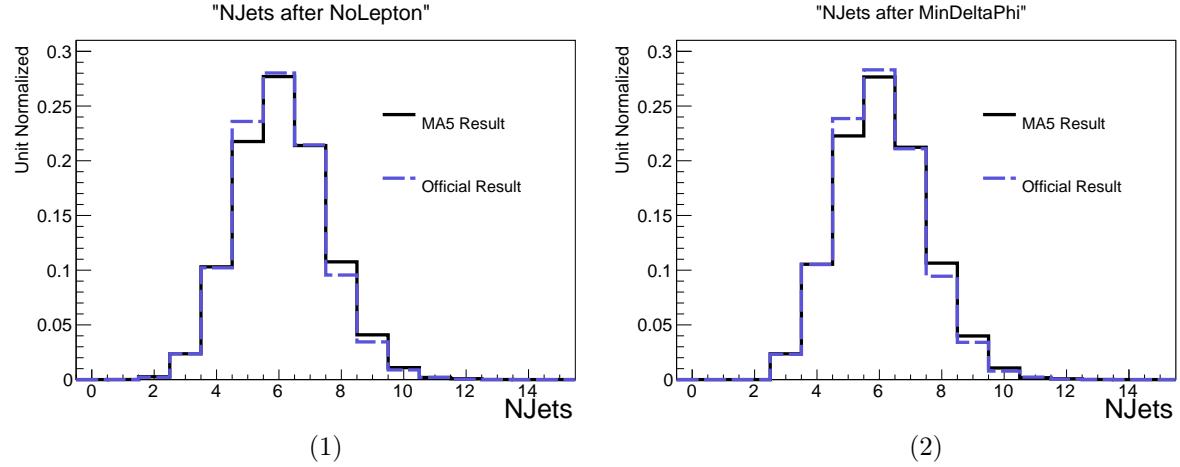


Figure D.18: Comparison of the distributions of NJets between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T5VV signal model.

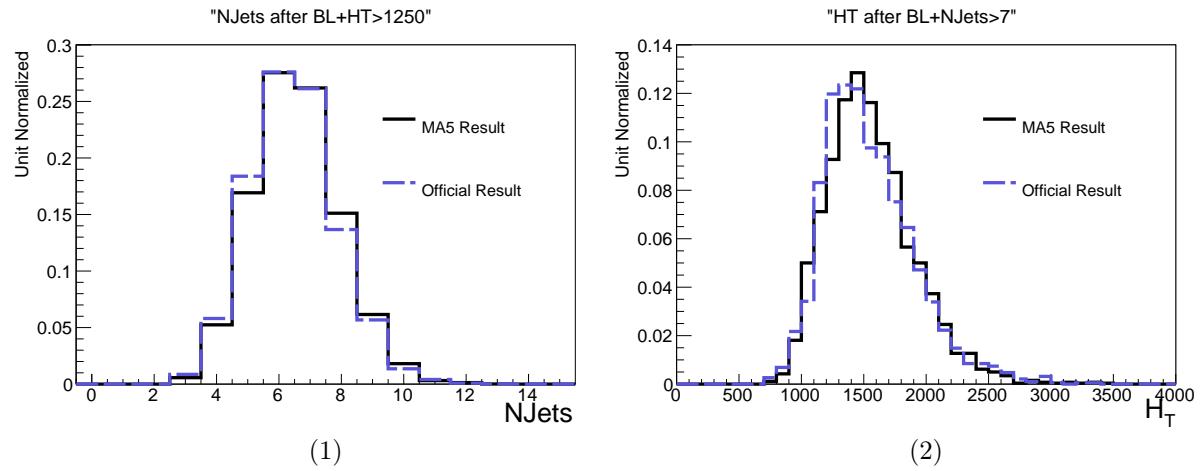


Figure D.19: Additional checks: comparison between ours and the official distributions of NJets after $BL+H_T > 1250$ cuts (left), and H_T after $BL+N\text{Jets} > 7$ cuts (right), for the T5VV signal model.

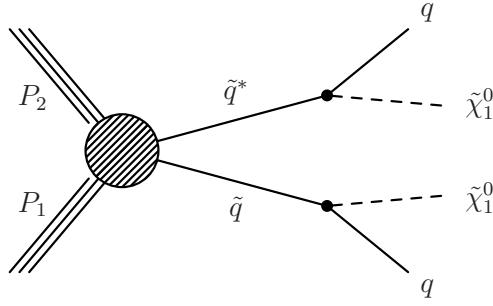


Figure D.20: Diagram of the dominant SUSY production mechanism for the T2qq signal model.

Cut Name	Official Count (Eff)	MA5 Count (Eff)
MET Cleaning	1215.2 (xxx)	1215.2 (xxx)
No Lepton	1212.8 (99%)	1215.2 (100%)
NJets>2	675.9 (55%)	691.54 (56%)
$H_T > 500$	619.5 (91%)	638.41 (92%)
$H_T^{\text{miss}} > 200$	524.0 (84%)	539.59 (84%)
Min $\Delta(\phi)$	460.7 (87%)	476.12 (88%)

Table D.9: The cut flow for the baseline selection in CMS SUS-13-012 for the T2qq working point $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0}) = (700, 100)\text{GeV}$. The second column is the official account and our own results are given in column 3. The official counts are normalized to luminosity $\mathcal{L} = 19.5/\text{fb}$ and cross section $\sigma = 63.4 \text{ pb}$, and our counts are normalized to match the official count after the first cut, MET Cleaning.

T2qq simplified model.

Signal Region Name	Official	MA5
NJets3-5, H_T 500-800, H_T^{miss} 200-300	35.3	35.10
NJets3-5, H_T 500-800, H_T^{miss} 300-450	70.4	73.44
NJets3-5, H_T 500-800, H_T^{miss} 450-600	71.5	73.82
NJets3-5, H_T 500-800, $H_T^{\text{miss}} > 600$	23.6	28.78
NJets3-5, H_T 800-1000, H_T^{miss} 200-300	18.1	17.20
NJets3-5, H_T 800-1000, H_T^{miss} 300-450	21.9	32.19
NJets3-5, H_T 800-1000, H_T^{miss} 450-600	38.1	38.14
NJets3-5, H_T 800-1000, $H_T^{\text{miss}} > 600$	35.2	36.74
NJets3-5, H_T 1000-1250, H_T^{miss} 200-300	10.9	12.15
NJets3-5, H_T 1000-1250, H_T^{miss} 300-450	21.7	20.31
NJets3-5, H_T 1000-1250, H_T^{miss} 450-600	20.7	21.54
NJets3-5, H_T 1000-1250, $H_T^{\text{miss}} > 600$	21.8	23.59
NJets3-5, H_T 1250-1500, H_T^{miss} 200-300	4.3	5.53
NJets3-5, H_T 1250-1500, H_T^{miss} 300-450	8.1	7.85
NJets3-5, H_T 1250-1500, $H_T^{\text{miss}} > 450$	16.1	16.86
NJets3-5, H_T >1500, H_T^{miss} 200-300	3.7	3.68
NJets3-5, H_T >1500, $H_T^{\text{miss}} > 300$	13.	13.45
NJets6-7, H_T 500-800, H_T^{miss} 200-300	0.8	0.40
NJets6-7, H_T 500-800, H_T^{miss} 300-450	1.0	0.44
NJets6-7, H_T 500-800, $H_T^{\text{miss}} > 450$	0.4	0.44
NJets6-7, H_T 800-1000, H_T^{miss} 200-300	0.5	0.58
NJets6-7, H_T 800-1000, H_T^{miss} 300-450	1.1	1.26
NJets6-7, H_T 800-1000, $H_T^{\text{miss}} > 450$	1.5	1.63
NJets6-7, H_T 1000-1250, H_T^{miss} 200-300	1.0	0.61
NJets6-7, H_T 1000-1250, H_T^{miss} 300-450	1.2	1.33
NJets6-7, H_T 1000-1250, $H_T^{\text{miss}} > 450$	2.5	3.24
NJets6-7, H_T 1250-1500, H_T^{miss} 200-300	0.6	0.61
NJets6-7, H_T 1250-1500, H_T^{miss} 300-450	1.2	0.61
NJets6-7, H_T 1250-1500, $H_T^{\text{miss}} > 450$	1.4	1.84
NJets6-7, H_T >1500, H_T^{miss} 200-300	0.6	0.30
NJets6-7, H_T >1500, $H_T^{\text{miss}} > 300$	2.3	1.80
NJets>7, H_T 500-800, $H_T^{\text{miss}} > 200$	0.0	0.0
NJets>7, H_T 800-1000, $H_T^{\text{miss}} > 200$	0.0	0.0
NJets>7, H_T 1000-1250, $H_T^{\text{miss}} > 200$	0.2	0.27
NJets>7, H_T 1250-1500, $H_T^{\text{miss}} > 200$	0.3	0.10
NJets>7, H_T >1500, $H_T^{\text{miss}} > 200$	0.3	0.13

Table D.10: The signal region (SR) counts in CMS SUS-13-012 for the T2qq scenario after all selection has been applied. Column 2 is the official account obtained through generous correspondence with Christian Sanders, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.9.

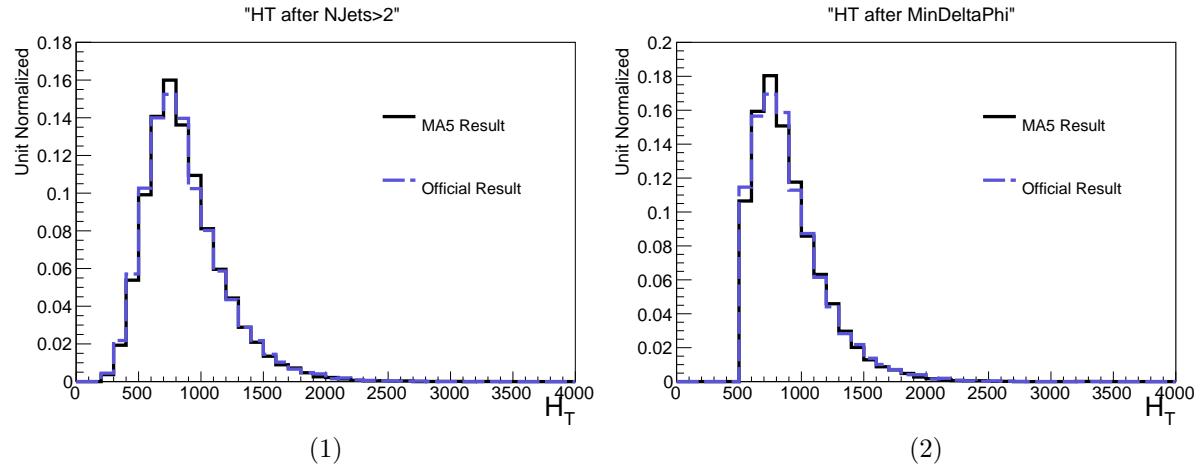


Figure D.21: Comparison of the distributions of H_T between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T2qq signal model.

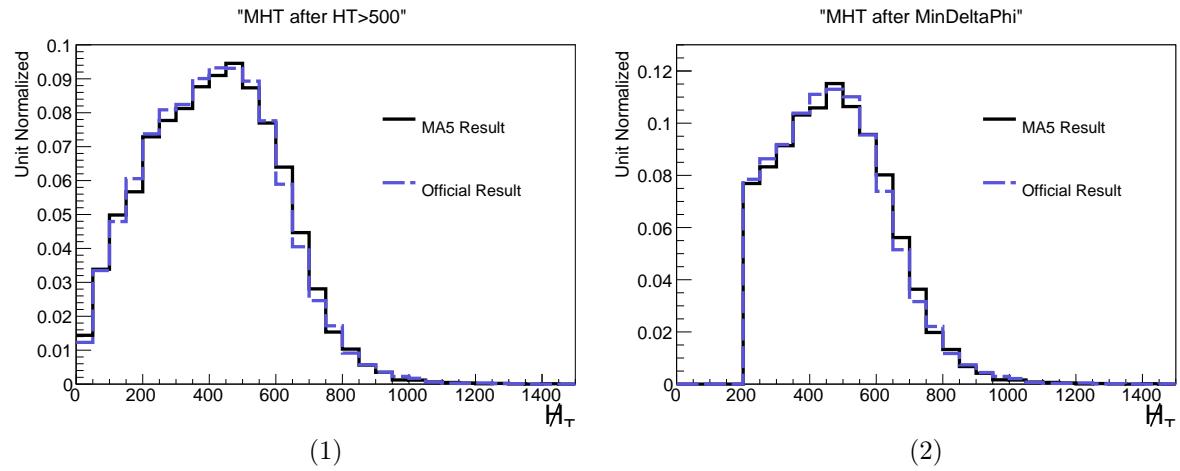


Figure D.22: Comparison of the distributions of $H_{T\text{miss}}$ between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T2qq signal model.

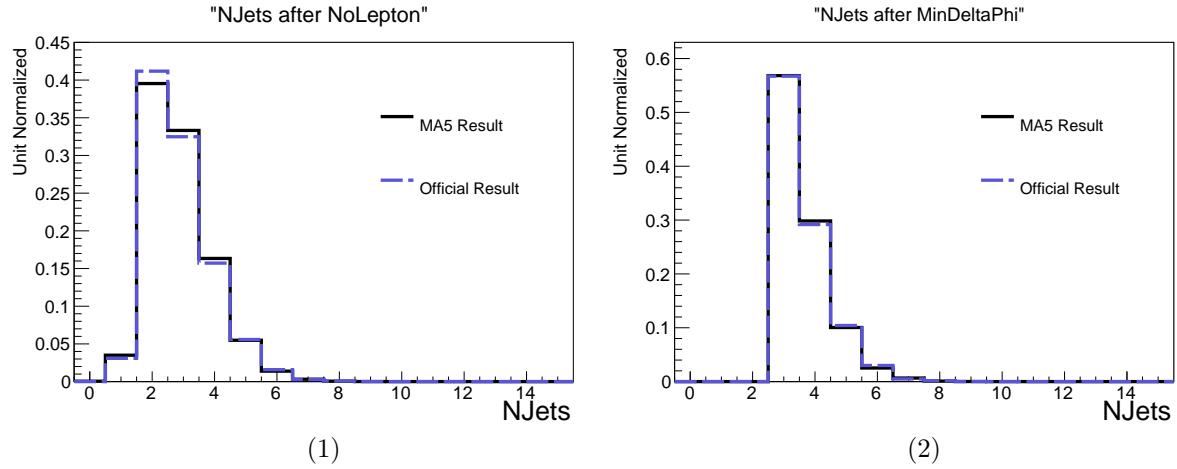


Figure D.23: Comparison of the distributions of NJets between the official and our own samples after the “n-1” cut, $\text{Min } \Delta(\phi)$ (left), and after all baseline cuts (right), for the T2qq signal model.

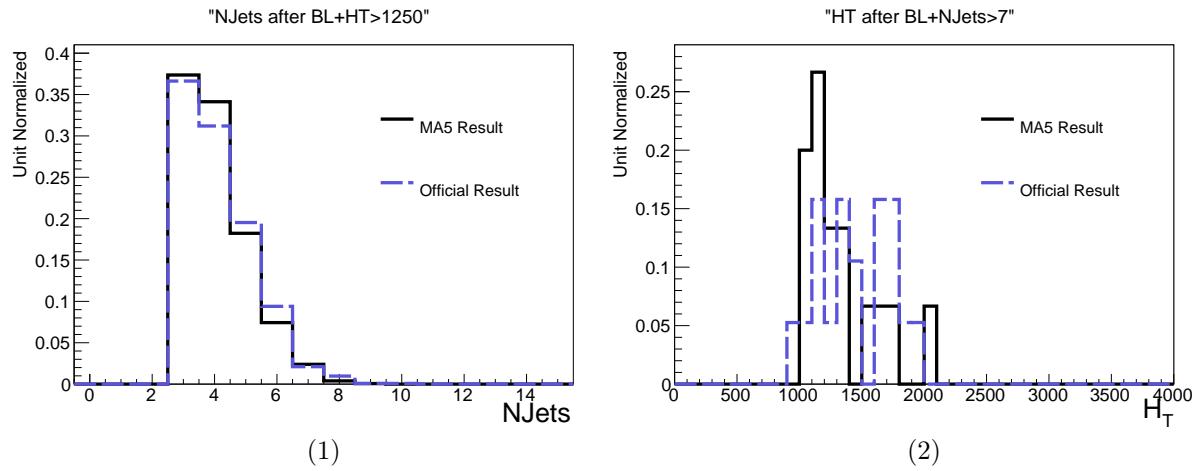


Figure D.24: Additional checks: comparison between ours and the official distributions of NJets after $\text{BL} + H_T > 1250$ cuts (left), and H_T after $\text{BL} + \text{NJets} > 7$ cuts (right), for the T2qq signal model.

Exclusion Limits. It is also instructive to reproduce the 95% CL exclusion lines in the $(m_{\tilde{g}}, m_{\tilde{\chi}_1^0})$ and $(m_{\tilde{q}}, m_{\tilde{\chi}_1^0})$ mass planes. Figure D.25 shows the limit curves (in red) obtained with our MADANALYSIS 5 implementation and using the `exclusion_CLs.py` code superimposed on the official CMS exclusion with its $\pm 1\sigma$ theoretical uncertainty (solid and dashed black lines). For the T1qqqq ($\tilde{g} \rightarrow q\bar{q}\chi_1^0$) and T5VV ($\tilde{g} \rightarrow q\bar{q}V\chi_1^0$) simplified models, the limits are reproduced very well. Also for T1tttt ($\tilde{g} \rightarrow t\bar{t}\chi_1^0$) the agreement is reasonably fine. For T2qq ($\tilde{q} \rightarrow q\chi_1^0$), however, we encounter a rather erratic behavior for LSP masses above about 200–250GeV.

It should be noted that our limit setting procedure differs from that used by CMS, and so should be considered a rough estimate. Our procedure is as follows. For any given point on the mass plane, the production cross section is taken from the LHC SUSY cross sections twiki and the signal acceptance is computed with our MADANALYSIS 5 recast code for the analysis. Then, the most sensitive signal region (SR), out of the 36 total SRs, is determined based on the number of expected signal and background counts. Finally, the CLs exclusion value is determined from the expected signal, expected background, expected uncertainty on the background, and observed counts in the most sensitive SR.

The primary difference between our method and that used by CMS is that we consider only the most sensitive SR in the exclusion calculation, whereas CMS uses all signal regions simultaneously and considers correlations in the uncertainty between signal regions. This difference introduces a certain volatility in our exclusion limits, which can however be mitigated by demanding that jumps in exclusion between two neighbouring points close in mass be not too large. As can be seen in Fig.!D.25, we obtain excellent results for the T1qqqq and T5VV scenarios; the exclusion curve for T1tttt shows more fluctuations but none the less matches the official result reasonably well. The only problematic case is the T2qq topology with LSP masses above 200–250GeV: here our procedure clearly does not well reproduce the official limit curve. To improve the situation, we would need the statistical model from CMS for combining the 36 SRs. Unfortunately, this is currently not available.

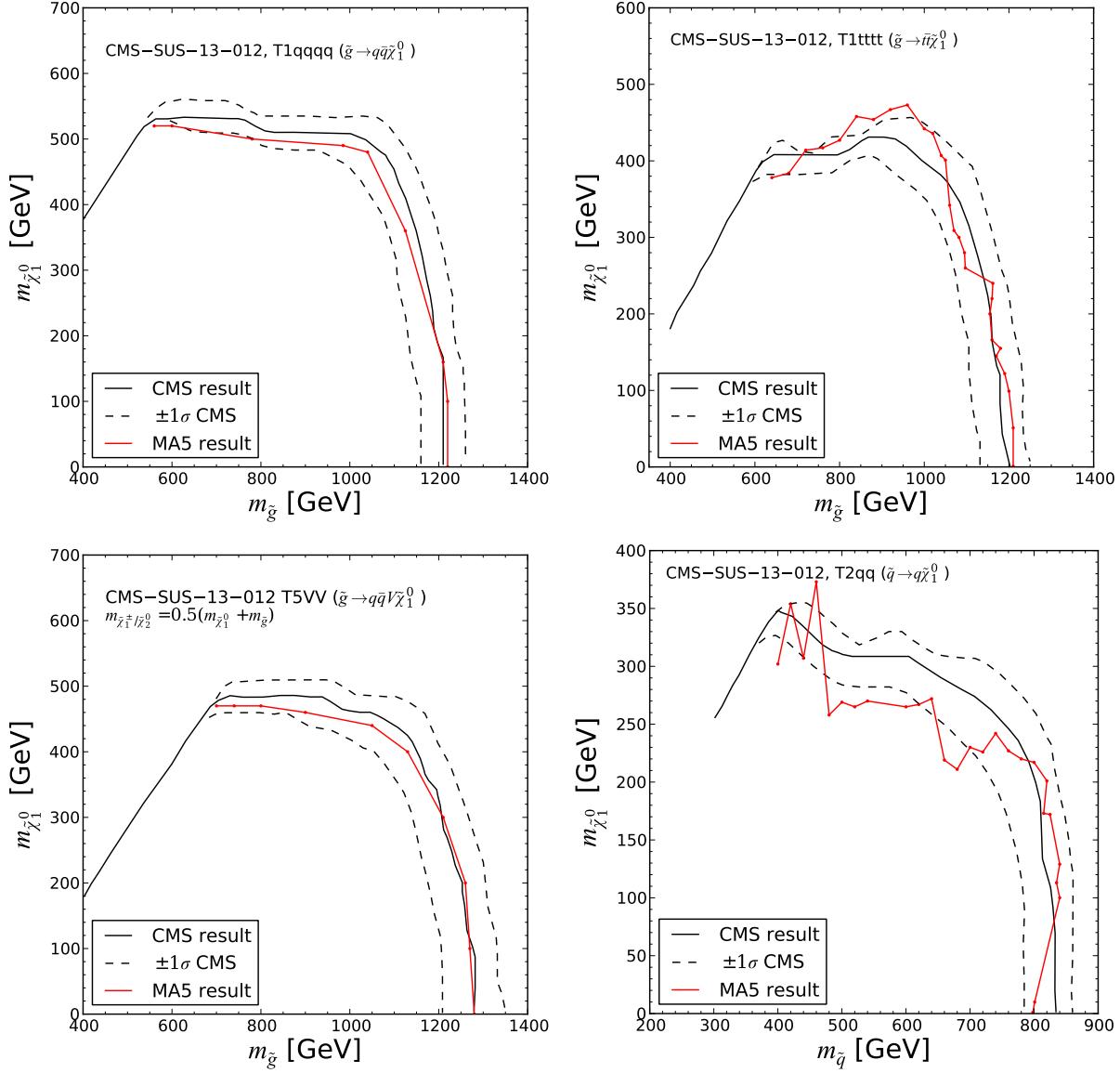


Figure D.25: The 95% CL exclusion limits (in red) reproduced with our MADANALYSIS 5 implementation compared to the official limits (in black). Top left: T1qqqq, top right: T1tttt, bottom left: T5VV, bottom right: T2qq simplified models.

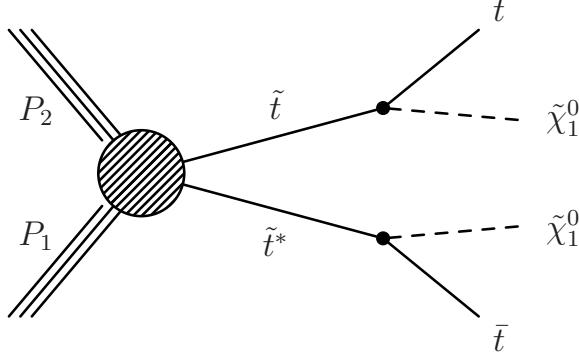


Figure D.26: Diagram for the T2tt SMS topology. Several mass combinations of the stop and LSP are used in the tables below as benchmark comparison scenarios.

D.6 Validation document: SUS-14-002

The following describes the work documented in [140], in association with the paper [140].

Atmasiddha, Prachi (IISER); Bein, Samuel (Florida State U.)

We present the results of the synchronization of the MA5 implementation of the SUS-14-001 “top tagging” SUSY search for top squarks. The performance of the implementation is evaluated by comparing the MA5-derived results with a set of cut flow tables and kinematic distributions provided by CMS for this purpose. The simplified model T2tt (Fig. D.26) is used as a common benchmark, with values of the masses of the stop and neutralino taking on a range of values. Two types of comparisons are given in Figs. D.11-D.19: cut flow tables and normalized kinematic distributions. In some cases the cut flow tables give the number of events normalized to 100%; in other cases the tables are normalized to the cross section times the integrated luminosity. The normalization convention used by CMS was followed. Question marks hold the place of values that were not provided in the CMS cut flow tables.

Cut Name	CMS Count(Eff)	MA5 Count(Eff)
Event Cleaning	98.13 (xxx)	98.13 (xxx)
No Mu	72.16 (73%)	72.21 (73%)
No Ele	55.41 (76%)	55.50 (76%)
Njet70>1	49.55 (89%)	50.07 (90%)
Njet50>3	31.16 (62%)	32.14 (64%)
Njet30>4	26.25 (84%)	27.10 (84%)
Min $\Delta(\phi)$	22.46 (85%)	23.15 (85%)
Nbjets>0	19.63 (87%)	19.69 (85%)
MET>200	12.21 (62%)	12.95 (65%)
Top Reco	- (-)	5.79 (44%)
MTsum>500	4.87 (39%)	4.95 (85%)

Table D.11: The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-500-125 and the MA5 results are given in column 3.

Signal Region Name	CMS	MA5
MET200-350, Nbjets=1	1.19	1.25
MET>350, Nbjets=1	0.93	1.12
MET200-350, Nbjets>1	1.64	1.38
MET>350, Nbjets>1	1.11	1.19

Table D.12: The signal region (SR) counts in CMS CMS-SUS-14-001 for the signal model working point T2tt-500-125 after all selection has been applied. Column 2 is the CMS ac-count, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.11.

Cut Name	CMS Count(Eff)	MA5 Count(Eff)
Event Cleaning	97.44 (xxx)	97.44 (xxx)
No Mu	72.5 (74%)	71.81 (73%)
No Ele	55.55 (76%)	54.98 (76%)
Njet70>1	52.72 (94%)	51.86 (94%)
Njet50>3	34.55 (65%)	34.66 (66%)
Njet30>4	28.49 (82%)	28.86 (83%)
Min $\Delta(\phi)$	24.98 (87%)	25.22 (87%)
Nbjets>0	21.81 (87%)	21.67 (85%)
MET>200	17.6 (80%)	17.73 (81%)
Top Reco	- (-)	9.10 (51%)
MTsum>500	8.37 (47%)	8.52 (93%)

Table D.13: The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-650-25 and the MA5 results are given in column 3.

Signal Region Name	CMS	MA5
MET200-350, Nbjets=1	1.06	0.91
MET>350, Nbjets=1	2.49	2.93
MET200-350, Nbjets>1	1.34	1.24
MET>350, Nbjets>1	3.48	3.43

Table D.14: The signal region (SR) counts in CMS CMS-SUS-14-001 for the signal model working point T2tt-650-25 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.13.

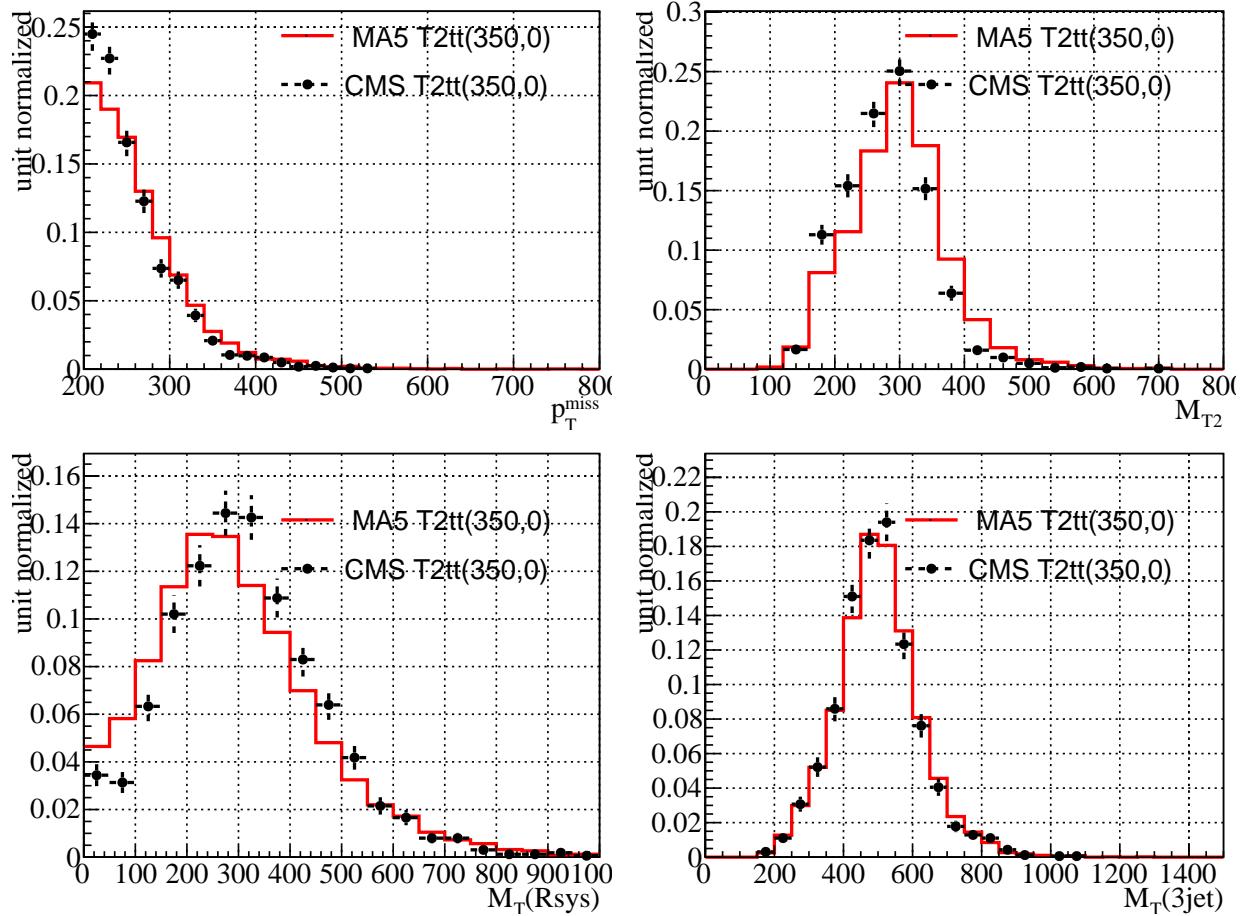
Cut Name	CMS Count(Eff)	MA5 Count(Eff)
Event Cleaning	15662.0 (xxx)	15662.0 (xxx)
No Mu	- (-)	11568.97 (73%)
No Ele	8802.0 (56%)	8927.82 (77%)
Njet70>1	- (-)	7380.74 (82%)
Njet50>3	- (-)	4350.13 (58%)
Njet30>4	3113.0 (35%)	3653.80 (83%)
Min $\Delta(\phi)$	2205.0 (70%)	2972.08 (81%)
Nbjets>0	2200.0 (99%)	2539.64 (85%)
MET>200	- (-)	1010.90 (39%)
Top Reco	- (-)	314.46 (31%)
MTsum>500	182.9 (8%)	213.18 (67%)

Table D.15: The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-350-0 and the MA5 results are given in column 3.

Signal Region Name	CMS	MA5
MET200-350, Nbjets=1	-	95.08
MET>350, Nbjets=1	-	13.66
MET200-350, Nbjets>1	-	90.88
MET>350, Nbjets>1	7.5	13.55

Table D.16: The signal region (SR) counts in CMS CMS-SUS-14-001 for the signal model working point T2tt-350-0 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.15.

Figure D.27: MA5 and CMS unit-normalized kinematic distributions after the baseline selection for the T2tt working point (350,0).



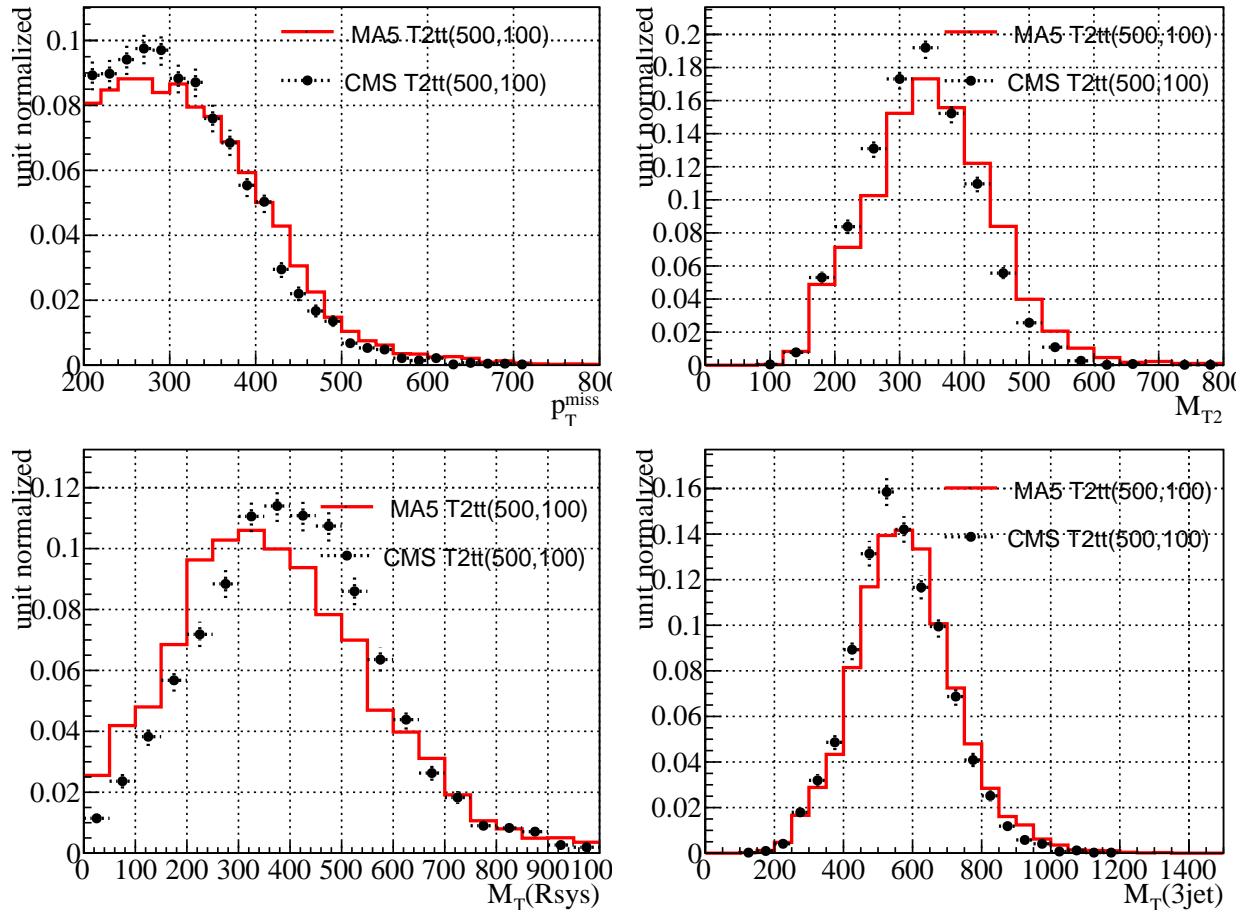
Cut Name	CMS Count(Eff)	MA5 Count(Eff)
Event Cleaning	1660.0 (xxx)	1660.0 (xxx)
No Mu	- (-)	1229.14 (74%)
No Ele	927.0 (55%)	942.80 (76%)
Njet70>1	- (-)	856.80 (90%)
Njet50>3	- (-)	551.71 (64%)
Njet30>4	419.0 (45%)	468.27 (84%)
Min $\Delta(\phi)$	360.0 (85%)	400.29 (85%)
Nbjets>0	314.0 (87%)	341.67 (85%)
MET>200	- (-)	229.78 (67%)
Top Reco	- (-)	105.02 (45%)
MTsum>500	85.9 (27%)	90.82 (86%)

Table D.17: The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-500-100 and the MA5 results are given in column 3.

Signal Region Name	CMS	MA5
MET200-350, Nbjets=1	-	21.48
MET>350, Nbjets=1	-	20.52
MET200-350, Nbjets>1	-	25.31
MET>350, Nbjets>1	19.8	23.50

Table D.18: The signal region (SR) counts in CMS CMS-SUS-14-001 for the signal model working point T2tt-500-100 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.17.

Figure D.28: MA5 and CMS unit-normalized kinematic distributions after the baseline selection for the T2tt working point (500,100).



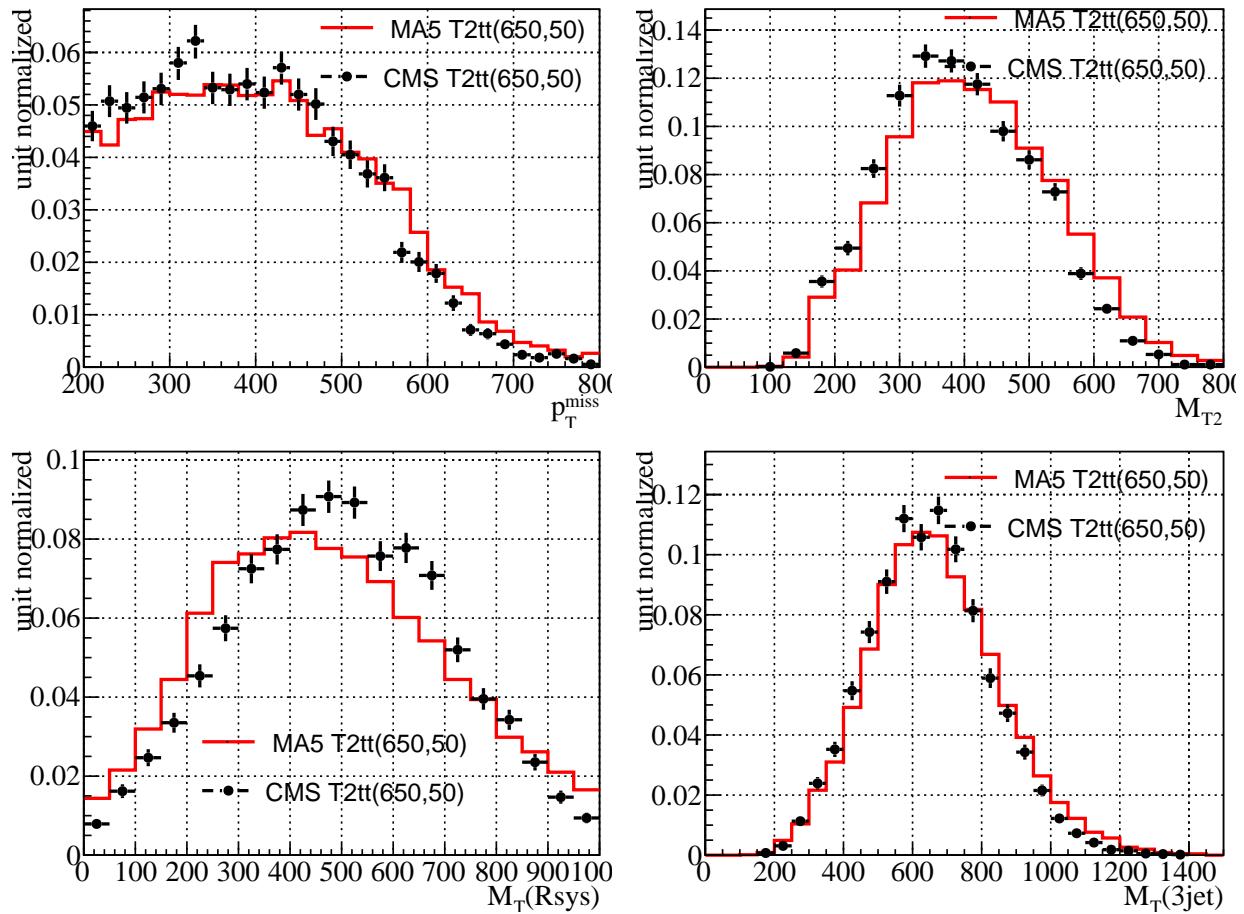
Cut Name	CMS Count(Eff)	MA5 Count(Eff)
Event Cleaning	270.8 (xxx)	270.8 (xxx)
No Mu	- (-)	199.25 (73%)
No Ele	152.0 (56%)	152.26 (76%)
Njet70>1	- (-)	143.65 (94%)
Njet50>3	- (-)	95.87 (66%)
Njet30>4	75.0 (49%)	79.95 (83%)
Min $\Delta(\phi)$	66.0 (88%)	69.84 (87%)
Nbjets>0	58.0 (87%)	59.99 (85%)
MET>200	- (-)	48.58 (80%)
Top Reco	- (-)	25.36 (52%)
MTsum>500	22.7 (39%)	23.71 (93%)

Table D.19: The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-650-50 and the MA5 results are given in column 3.

Signal Region Name	CMS	MA5
MET200-350, Nbjets=1	-	2.77
MET>350, Nbjets=1	-	8.08
MET200-350, Nbjets>1	-	3.35
MET>350, Nbjets>1	9.3	9.49

Table D.20: The signal region (SR) counts in CMS CMS-SUS-14-001 for the signal model working point T2tt-650-50 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table D.19.

Figure D.29: MA5 and CMS unit-normalized kinematic distributions after the baseline selection for the T2tt working point (650,50).



BIBLIOGRAPHY

- [1] Particle Data Group, “Review of Particle Physics,” *Chin. Phys. C*, vol. 38, p. 090001, 2014.
- [2] T. S. van Albada, J. N. Bahcall, K. Begeman, and R. Sancisi, “The Distribution of Dark Matter in the Spiral Galaxy NGC-3198,” *Astrophys. J.*, vol. 295, pp. 305–313, 1985.
- [3] D. Clowe, S. W. Randall, and M. Markevitch, “Catching a bullet: Direct evidence for the existence of dark matter,” *Nucl. Phys. Proc. Suppl.*, vol. 173, pp. 28–31, 2007.
- [4] J.-L. Caron, “Cross section of LHC dipole.. Dipole LHC: coupe transversale..” AC Collection. Legacy of AC. Pictures from 1992 to 2002., May 1998.
- [5] “Commissioning the performance of key observables used in SUSY searches with the first 13 TeV data,” Aug 2015.
- [6] V. Khachatryan *et al.*, “Search for supersymmetry in the multijet and missing transverse momentum final state in pp collisions at 13 TeV,” *Phys. Lett. B*, 2016.
- [7] C. Collaboration, “Search for direct production of top squark pairs decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV,” 2016.
- [8] D. Sengupta and S. Kulkarni, “MadAnalysis 5 implementation of CMS-SUS-13-016,”
- [9] C. Collaboration, “Search for supersymmetry in pp collisions at $\sqrt{s} = 8$ TeV in events with two opposite sign leptons, large number of jets, b-tagged jets, and large missing transverse energy.,” 2013.
- [10] J. L. Rosner, “The Standard model in 2001,” in *Heavy flavor physics: Theory and experimental results in heavy quark physics and CP violation. Proceedings, 55th Scottish Universities Summer School in Physics, SUSSP 2001, St. Andrews, UK, August 7-23, 2001*, pp. 1–56, 2001.
- [11] Y. Fukuda *et al.*, “Evidence for oscillation of atmospheric neutrinos,” *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, 1998.
- [12] G. W. Bennett *et al.*, “Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL,” *Phys. Rev.*, vol. D73, p. 072003, 2006.

- [13] R. A. Broglia, “More is different: 50 years of nuclear BCS,” 2012.
- [14] J. J. Sakurai and J. Napolitano, *Modern quantum mechanics*. Addison-Wesley, 2011.
- [15] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. 1995.
- [16] E. Noether, “Invariant Variation Problems,” *Gott. Nachr.*, vol. 1918, pp. 235–257, 1918. [Transp. Theory Statist. Phys.1,186(1971)].
- [17] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964.
- [18] P. Langacker, “Structure of the standard model,” *Adv. Ser. Direct. High Energy Phys.*, vol. 14, pp. 15–36, 1995.
- [19] S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett.*, vol. B716, pp. 30–61, 2012.
- [20] M. d’Ocagne, *Coordonnées parallèles et axiales: Méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles*. 1885.
- [21] T. Yanagida, “Horizontal Symmetry and Masses of Neutrinos,” *Prog. Theor. Phys.*, vol. 64, p. 1103, 1980.
- [22] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, “The Evolution of Large Scale Structure in a Universe Dominated by Cold Dark Matter,” *Astrophys. J.*, vol. 292, pp. 371–394, 1985.
- [23] E. Komatsu *et al.*, “Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation,” *Astrophys. J. Suppl.*, vol. 180, pp. 330–376, 2009.
- [24] T. M. Undagoitia and L. Rauch, “Dark matter direct-detection experiments,” *J. Phys.*, vol. G43, no. 1, p. 013001, 2016.
- [25] G. Aad *et al.*, “The ATLAS Experiment at the CERN Large Hadron Collider,” *JINST*, vol. 3, p. S08003, 2008.
- [26] S. Chatrchyan *et al.*, “The CMS experiment at the CERN LHC,” *JINST*, vol. 3, p. S08004, 2008.

- [27] G. Goldhaber, “The Acceleration of the Expansion of the Universe: A Brief Early History of the Supernova Cosmology Project (SCP),” *AIP Conf. Proc.*, vol. 1166, pp. 53–72, 2009.
- [28] S. P. Martin, “A Supersymmetry primer,” 1997. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [29] V. Khachatryan *et al.*, “Phenomenological MSSM interpretation of CMS searches in pp collisions at $\text{sqrt}(s) = 7$ and 8 TeV,” 2016.
- [30] B. Dumont, B. Fuks, S. Kraml, S. Bein, G. Chalons, E. Conte, S. Kulkarni, D. Sengupta, and C. Wymant, “Toward a public analysis database for LHC new physics searches using MADANALYSIS 5,” *Eur. Phys. J.*, vol. C75, no. 2, p. 56, 2015.
- [31] Yu. A. Golfand and E. P. Likhtman, “Extension of the Algebra of Poincare Group Generators and Violation of p Invariance,” *JETP Lett.*, vol. 13, pp. 323–326, 1971. [Pisma Zh. Eksp. Teor. Fiz.13,452(1971)].
- [32] J. Wess and B. Zumino, “Supergauge Transformations in Four-Dimensions,” *Nucl. Phys.*, vol. B70, pp. 39–50, 1974.
- [33] M. Miura, “Search for the proton decay in Super-Kamiokande,” in *Proceedings, 44th Rencontres de Moriond on Electroweak Interactions and Unified Theories*, pp. 525–530, 2009.
- [34] G. R. Farrar and S. Weinberg, “Supersymmetry at Ordinary Energies. 2. R Invariance, Goldstone Bosons, and Gauge Fermion Masses,” *Phys. Rev.*, vol. D27, p. 2732, 1983.
- [35] S. Dimopoulos and D. W. Sutter, “The Supersymmetric flavor problem,” *Nucl. Phys.*, vol. B452, pp. 496–512, 1995.
- [36] J. Wess, “SUPERSYMMETRY AND SUPERGRAVITY,” in *22nd International Conference on High Energy Physics. Vol. 1*, p. II.84, 1984.
- [37] C. F. Berger, J. S. Gainer, J. L. Hewett, and T. G. Rizzo, “Supersymmetry Without Prejudice,” *JHEP*, vol. 02, p. 023, 2009.
- [38] M. Carena, J. Lykken, S. Sekmen, N. R. Shah, and C. E. M. Wagner, “The pMSSM Interpretation of LHC Results Using Rernormalization Group Invariants,” *Phys. Rev.*, vol. D86, p. 075025, 2012.

- [39] G. Aad *et al.*, “Summary of the ATLAS experiment’s sensitivity to supersymmetry after LHC Run 1 interpreted in the phenomenological MSSM,” *JHEP*, vol. 10, p. 134, 2015.
- [40] A. Breskin and R. Voss, *The CERN Large Hadron Collider: Accelerator and Experiments*. Geneva: CERN, 2009.
- [41] G. Aad *et al.*, “Multi-channel search for squarks and gluinos in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS detector,” *Eur. Phys. J.*, vol. C73, no. 3, p. 2362, 2013.
- [42] O. Adriani *et al.*, “The LHCf detector at the CERN Large Hadron Collider,” *JINST*, vol. 3, p. S08006, 2008.
- [43] L. Leistam, “The ALICE experiment: a large ION collider experiment for LHC,” 2003.
- [44] G. Anelli *et al.*, “The TOTEM experiment at the CERN Large Hadron Collider,” *JINST*, vol. 3, p. S08007, 2008.
- [45] A. A. Alves, Jr. *et al.*, “The LHCb Detector at the LHC,” *JINST*, vol. 3, p. S08005, 2008.
- [46] J. L. Pinfold, “The MoEDAL experiment - a new light on LHC physics,” *EPJ Web Conf.*, vol. 95, p. 03030, 2015.
- [47] F. Beaudette, “The CMS Particle Flow Algorithm,” in *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013)*, pp. 295–304, 2013.
- [48] V. Veszpremi, “Operation and performance of the CMS tracker,” *JINST*, vol. 9, p. C03005, 2014.
- [49] A. Hocker *et al.*, “TMVA - Toolkit for Multivariate Data Analysis,” *PoS*, vol. ACAT, p. 040, 2007.
- [50] M. Cacciari, G. P. Salam, and G. Soyez, “The Anti- k_t jet clustering algorithm,” *JHEP*, vol. 04, p. 063, 2008.
- [51] B. Bilki, “CMS Forward Calorimeters Phase II Upgrade,” *J. Phys. Conf. Ser.*, vol. 587, no. 1, p. 012014, 2015.
- [52] J. Chen, R. Mao, L. Zhang, and Y. Ren-Zhu, “Large size LSO and LYSO crystals for future high energy physics experiments,” *IEEE Trans. Nucl. Sci.*, vol. 54, pp. 718–724, 2007.

- [53] “Fermilab m-test facility.” <http://www-ppd.fnal.gov/FTBF/>.
- [54] R. Brun and F. Rademakers, “ROOT: An object oriented data analysis framework,” *Nucl. Instrum. Meth.*, vol. A389, pp. 81–86, 1997.
- [55] “Particle physics event display.” <https://github.com/sbein/EventDisplay>.
- [56] H. Jeffreys, “Scientific Inference, Third edition,” *Cambridge University Press; 3rd edition*, 1974.
- [57] S. Abdullin, P. Azzi, F. Beaudette, P. Janot, and A. Perrotta, “The fast simulation of the CMS detector at LHC,” *J. Phys. Conf. Ser.*, vol. 331, p. 032049, 2011.
- [58] C. P. Robert, *The Bayesian Choice: from Decision-Theoretic Foundations to Computational Implementation*. Springer Verlag, New York, 2nd ed., 2007.
- [59] A. O’Hagan, *Bayesian Inference*, vol. 2B of *Kendall’s Advanced Theory of Statistics*. Edward Arnold, London, 1994.
- [60] S. Sekmen, S. Kraml, J. Lykken, F. Moortgat, S. Padhi, *et al.*, “Interpreting LHC SUSY searches in the phenomenological MSSM,” *JHEP*, vol. 02, p. 075, 2012.
- [61] S. Chatrchyan *et al.*, “Search for Supersymmetry at the LHC in Events with Jets and Missing Transverse Energy,” *Phys. Rev. Lett.*, vol. 107, p. 221804, 2011.
- [62] S. Chatrchyan *et al.*, “Search for new physics with same-sign isolated dilepton events with jets and missing transverse energy,” *Phys. Rev. Lett.*, vol. 109, p. 071803, 2012.
- [63] S. Chatrchyan *et al.*, “Search for new physics in events with opposite-sign leptons, jets, and missing transverse energy in pp collisions at $\sqrt{s} = 7$ TeV,” *Phys. Lett. B*, vol. 718, p. 815, 2013.
- [64] A. A. Markov, *Extension of the limit theorems of probability theory to a sum of variables connected in a chain*. reprinted in Appendix B of: R. Howard, *Dynamic Probabilistic Systems, volume 1: Markov Chains*, John Wiley and Sons, 1971.
- [65] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines,” *J. Chem. Phys*, vol. 21, p. 1087, 1953.
- [66] W. K. Hastings, “Monte carlo sampling methods using markov chains and their applications,” *Biometrika*, vol. 57, p. 97, 1970.

- [67] B. A. Berg, *Markov chain monte carlo simulations and their statistical analysis*. World Scientific, Singapore, 2004.
- [68] Y. Amhis *et al.*, “Averages of b-hadron, c-hadron, and τ -lepton properties as of summer 2014,” 2014.
- [69] V. Khachatryan *et al.*, “Observation of the rare $B_s^0 \rightarrow \mu^+ \mu^-$ decay from the combined analysis of CMS and LHCb data,” *Nature*, vol. 522, p. 68, 2015.
- [70] K. Hagiwara, R. Liao, A. D. Martin, D. Nomura, and T. Teubner, “ $(g-2)_\mu$ and $\alpha(M_Z^2)$ re-evaluated using new precise data,” *J.Phys.*, vol. G38, p. 085003, 2011.
- [71] Tevatron Electroweak Working Group, CDF and D0 Collaborations, “Combination of CDF and D0 results on the mass of the top quark using up to 8.7 fb^{-1} at the Tevatron,” 2013.
- [72] J. Bernon, B. Dumont, and S. Kraml, “Status of Higgs couplings after run 1 of the LHC,” *Phys. Rev. D*, vol. 90, p. 071301, 2014.
- [73] J. Bernon and B. Dumont, “Lilith: a tool for constraining new physics from Higgs measurements,” *Eur. Phys. J. C*, vol. 75, no. 9, p. 440, 2015.
- [74] Joint LEP2 SUSY Working Group, the Aleph, Delphi, L3 and Opal Collaborations.
- [75] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, “MicrOMEGAs: A program for calculating the relic density in the MSSM,” *Comput. Phys. Commun.*, vol. 149, p. 103, 2002.
- [76] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, “micrOMEGAs: Version 1.3,” *Comput. Phys. Commun.*, vol. 174, p. 577, 2006.
- [77] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, “Dark matter direct detection rate in a generic model with micrOMEGAs 2.2,” *Comput. Phys. Commun.*, vol. 180, p. 747, 2009.
- [78] B. Allanach, “SOFTSUSY: a program for calculating supersymmetric spectra,” *Comput. Phys. Commun.*, vol. 143, p. 305, 2002.
- [79] F. Mahmoudi, “SuperIso v2.3: A program for calculating flavor physics observables in supersymmetry,” *Comput. Phys. Commun.*, vol. 180, p. 1579, 2009.

- [80] M. Muhlleitner, A. Djouadi, and Y. Mambrini, “SDECAY: A Fortran code for the decays of the supersymmetric particles in the MSSM,” *Comput. Phys. Commun.*, vol. 168, p. 46, 2005.
- [81] A. Djouadi, J. Kalinowski, and M. Spira, “HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension,” *Comput. Phys. Commun.*, vol. 108, p. 56, 1998.
- [82] “Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data,” Tech. Rep. ATLAS-CONF-2014-009, CERN, Geneva, Mar 2014.
- [83] V. Khachatryan *et al.*, “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV,” *Eur. Phys. J. C*, vol. 75, p. 212, 2015.
- [84] B. Dumont, J. F. Gunion, and S. Kraml, “The phenomenological MSSM in view of the 125 GeV Higgs data,” *Phys. Rev. D*, vol. 89, p. 055018, 2014.
- [85] P. Z. Skands, B. Allanach, H. Baer, C. Balazs, G. Belanger, *et al.*, “SUSY Les Houches accord: Interfacing SUSY spectrum calculators, decay packages, and event generators,” *JHEP*, vol. 07, p. 036, 2004.
- [86] C. G. Lester and D. J. Summers, “Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders,” *Phys. Lett.*, vol. B463, pp. 99–103, 1999.
- [87] S. Chatrchyan *et al.*, “Search for supersymmetry in pp collisions at $\sqrt{s}=8 \text{ TeV}$ in events with a single lepton, large jet multiplicity, and multiple b jets,” *Phys. Lett. B*, vol. 733, pp. 328–353, 2014.
- [88] S. Chatrchyan *et al.*, “Search for top-squark pair production in the single-lepton final state in pp collisions at $\sqrt{s} = 8 \text{ TeV}$,” *Eur. Phys. J. C*, vol. 73, no. 12, p. 2677, 2013.
- [89] S. Chatrchyan *et al.*, “Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 7 \text{ TeV}$,” *Phys. Rev. Lett.*, vol. 109, p. 171803, 2012.
- [90] S. Chatrchyan *et al.*, “Search for supersymmetry in events with b-quark jets and missing transverse energy in pp collisions at 7 TeV,” *Phys. Rev. D*, vol. 86, p. 072010, 2012.
- [91] S. Chatrchyan *et al.*, “Search for electroweak production of charginos and neutralinos using leptonic final states in pp collisions at $\sqrt{s} = 7 \text{ TeV}$,” *JHEP*, vol. 11, p. 147, 2012.

- [92] S. Chatrchyan *et al.*, “Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 8$ TeV,” *JHEP*, vol. 06, p. 055, 2014.
- [93] S. Chatrchyan *et al.*, “Search for supersymmetry in hadronic final states using MT2 in pp collisions at $\sqrt{s} = 7$ TeV,” *JHEP*, vol. 10, p. 018, 2012.
- [94] S. Chatrchyan *et al.*, “Search for gluino mediated bottom- and top-squark production in multijet final states in pp collisions at 8 TeV,” *Phys. Lett. B*, vol. 725, p. 243, 2013.
- [95] V. Khachatryan *et al.*, “Search for dark matter, extra dimensions, and unparticles in monojet events in proton-proton collisions at $\sqrt{s} = 8$ TeV,” *Eur. Phys. J. C*, vol. 75, p. 235, 2015.
- [96] V. Khachatryan *et al.*, “Searches for third-generation squark production in fully hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV,” *JHEP*, vol. 06, p. 116, 2015.
- [97] V. Khachatryan *et al.*, “Search for physics beyond the standard model in events with two leptons, jets, and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV,” *JHEP*, vol. 04, p. 124, 2015.
- [98] S. Chatrchyan *et al.*, “Search for new physics in events with same-sign dileptons and jets in pp collisions at $\sqrt{s} = 8$ TeV,” *JHEP*, vol. 01, p. 163, 2014.
- [99] V. Khachatryan *et al.*, “Searches for electroweak production of charginos, neutralinos, and sleptons decaying to leptons and W, Z, and Higgs bosons in pp collisions at 8 TeV,” *Eur. Phys. J. C*, vol. 74, p. 3036, 2014.
- [100] S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses,” *Annals Math. Statist.*, vol. 9, p. 60, 1938.
- [101] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual,” *JHEP*, vol. 05, p. 026, 2006.
- [102] C. Collaboration, “The fast simulation of the CMS detector at LHC,” *J.Phys.Conf.Ser.*, vol. 331, p. 032049, 2011.
- [103] W. D. P. H. Sekmen, Sezen; Teo, “Modeling Trigger Efficiencies Without Prejudice,” 2011.
- [104] S. A. Koay, *A Search for Dark Matter Production with Jets and Missing Momentum Signature in Proton-Proton Collisions at 7 TeV*. PhD thesis, UC, Santa Barbara, 2011.

- [105] M. Schröder, *Quality of jet measurements and impact on a search for new physics at CMS*. PhD thesis, Hamburg U., 2012.
- [106] K. Goebel, *Probing supersymmetry based on precise jet measurements at the CMS experiment*. PhD thesis, U. Hamburg, Dept. Phys., 2015.
- [107] S. Chatrchyan *et al.*, “Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 8$ TeV,” *JHEP*, vol. 06, p. 055, 2014.
- [108] “Jet resolution scale factors.” <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>. CMS twiki.
- [109] “Egamma pog recommendations.” <https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2>.
- [110] C. J. Clopper and E. S. Pearson, “The Use of Confidence of Fiducial Limits Illustrated in teh Case of the Binomial,” *Biometrika*, vol. 26, no. 4, pp. 404–413, 1934.
- [111] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J. C*, vol. 71, p. 1554, 2011.
- [112] T. Junk, “Confidence level computation for combining searches with small statistics,” *Nucl. Instr. and Meth. A*, vol. 434, p. 435, 1999.
- [113] A. L. Read, “Presentation of search results: the CL_s technique,” *J. Phys. G*, vol. 28, p. 2693, 2002.
- [114] M. Botje, J. Butterworth, A. Cooper-Sarkar, A. de Roeck, J. Feltesse, S. Forte, A. Glazov, J. Huston, R. McNulty, and T. Sjöstrand, “The PDF4LHC working group interim recommendations.” 2011.
- [115] S. Chatrchyan *et al.*, “Search for gluino mediated bottom- and top-squark production in multijet final states in pp collisions at 8 TeV,” *Phys. Lett. B*, vol. 725, p. 243, 2013.
- [116] C. M. Bishop and ???, “Pattern recognition and machine learning, 2006,” ????, vol. 60, no. 1, pp. 78–78, 2012.
- [117] C. Rogan, “Kinematical variables towards new dynamics at the LHC,” 2010.
- [118] S. Chatrchyan *et al.*, “Inclusive search for supersymmetry using the razor variables in pp collisions at $\sqrt{s} = 7$ TeV,” *Phys. Rev. Lett.*, vol. 111, no. 8, p. 081802, 2013.

- [119] B. Abbott *et al.*, “A measurement of the W boson mass,” *Phys. Rev.*, vol. D58, p. 092003, 1998.
- [120] V. Khachatryan *et al.*, “Searches for Supersymmetry using the M_{T2} Variable in Hadronic Events Produced in pp Collisions at 8 TeV,” *JHEP*, vol. 05, p. 078, 2015.
- [121] L. Randall and D. Tucker-Smith, “Dijet Searches for Supersymmetry at the LHC,” *Phys. Rev. Lett.*, vol. 101, p. 221803, 2008.
- [122] S. Kraml, S. Kulkarni, U. Laa, A. Lessa, V. Magerl, W. Magerl, D. Proschofsky-Spindler, M. Traub, and W. Waltenberger, “SModelS v1.0: a short user guide,” 2014.
- [123] M. Papucci, K. Sakurai, A. Weiler, and L. Zeune, “Fastlim: a fast LHC limit calculator,” *Eur. Phys. J.*, vol. C74, no. 11, p. 3163, 2014.
- [124] M. Drees, H. Dreiner, D. Schmeier, J. Tattersall, and J. S. Kim, “CheckMATE: Confronting your Favourite New Physics Model with LHC Data,” *Comput. Phys. Commun.*, vol. 187, pp. 227–265, 2015.
- [125] D. Barducci, A. Belyaev, M. Buchkremer, J. Marrouche, S. Moretti, and L. Panizzi, “XQCAT: eXtra Quark Combined Analysis Tool,” *Comput. Phys. Commun.*, vol. 197, pp. 263–275, 2015.
- [126] E. Conte, B. Fuks, and G. Serret, “MadAnalysis 5, A User-Friendly Framework for Collider Phenomenology,” *Comput. Phys. Commun.*, vol. 184, pp. 222–256, 2013.
- [127] E. Conte, B. Dumont, B. Fuks, and C. Wymant, “Designing and recasting LHC analyses with MadAnalysis 5,” *Eur. Phys. J.*, vol. C74, no. 10, p. 3103, 2014.
- [128] J. de Favereau, C. Delaere, P. Demin, A. Giannanco, V. Lemaître, A. Mertens, and M. Selvaggi, “DELPHES 3, A modular framework for fast simulation of a generic collider experiment,” *JHEP*, vol. 02, p. 057, 2014.
- [129] S. Kraml *et al.*, “Searches for New Physics: Les Houches Recommendations for the Presentation of LHC Results,” *Eur. Phys. J.*, vol. C72, p. 1976, 2012.
- [130] G. Chalons, “MadAnalysis 5 implementation of ATLAS-SUSY-2013-05,”
- [131] B. Dumont, “MadAnalysis 5 implementation of ATLAS-SUSY-2013-11: di-leptons plus MET,”
- [132] “MADANALYSIS 55 analysis package.” <http://madanalysis.irmp.ucl.ac.be/wiki/PhysicsAnalysisDatabase#no1>.

- [133] “Inspire hep.” <http://inspirehep.net/help/data-submission>.
- [134] “Doi citation framework.” <http://www.datacite.org>.
- [135] S. Chatrchyan *et al.*, “Interpretation of Searches for Supersymmetry with simplified Models,” *Phys. Rev. D*, vol. 88, p. 052017, 2013.
- [136] M. Krämer, A. Kulesza, R. van der Leeuw, M. Mangano, S. Padhi, T. Plehn, and X. Portell, “Supersymmetry production cross sections in pp collisions at $\sqrt{s} = 7$ TeV.” 2012.
- [137] S. Bein and D. Sengupta, “MadAnalysis 5 implementation of CMS-SUS-13-012,”
- [138] B. Dumont, B. Fuks, and C. Wymant, “MadAnalysis 5 implementation of CMS-SUS-13-011: search for stops in the single lepton final state at 8 TeV,”
- [139] “CMS data preservation, re-use and open access policy.” <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB>ShowDocument?docid=6032>.
- [140] P. Atmasiddha, S. Bein, and S. Sharma, “MadAnalysis 5 implementation of CMS-SUS-14-001-TopTag,” 2015.

BIOGRAPHICAL SKETCH

I attended the University of Colorado from 2007 to 2011, where I studied under the mentorship of theoretical nuclear physicist James R. Shepard. I conducted research on two- and three-nucleon interactions for my undergraduate honors thesis, and graduated with a bachelor's degree in physics and economics. From 2011 to 2016, I carried out research for my Ph.D. with Florida State University under the supervision of experimental particle physicist Harrison B. Prosper, and was based at the Fermi National Accelerator Laboratory for the final two years. This research included

- performing searches for supersymmetry in data collected by CMS;
- performing interpretations of CMS searches based on the minimal supersymmetric standard model;
- developing techniques for the modeling of standard model backgrounds;
- developing hardware and software for test beam experiments to study and characterize components of future versions of the CMS detector;
- contributing to the creation of a database of public implementations of CMS and ATLAS searches for use by physicists outside the collaborations.

I defended my dissertation in June, 2016. In the fall of 2016, I will begin a postdoctoral fellowship with the University of Hamburg, where I will help to carry out searches for evidence of new physics with the CMS experiment and develop hardware for the CMS silicon tracker.