

**Search for supersymmetry in channels with two Higgs  
bosons and missing transverse momentum at CMS**

by

**Emily Fair**

B.S., Whitworth University, 2016

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Physics  
2022

Committee Members:

Keith Ulmer, Chair

Kevin Stenson

Oliver DeWolfe

Alysia Marino

Jeremy Darling

Fair, Emily (Ph.D., Physics)

Search for supersymmetry in channels with two Higgs bosons and missing transverse momentum  
at CMS

Thesis directed by Prof. Keith Ulmer

We present a search for physics beyond the Standard Model in channels with two Higgs bosons, each decaying via the process  $H \rightarrow b\bar{b}$ , and large missing transverse momentum. The search uses a data sample produced in proton-proton collisions with center-of-mass energy  $\sqrt{s} = 13$  TeV, collected by the CMS experiment at the CERN LHC, and corresponding to an integrated luminosity of  $137 \text{ fb}^{-1}$ . The Standard Model backgrounds are predicted using data-driven techniques and the prediction is compared to the signal region yields to test for the presence of new physics. The observed yields in the signal regions agree with the background estimation, enabling us to set limits on the production cross sections for pairs of the lightest neutralino  $\tilde{\chi}_1^0$  and pairs of gluinos using simplified supersymmetry models. For the electroweak production of nearly degenerate higgsinos,  $\tilde{\chi}_1^0$  masses in the range 175 to 1025 GeV are excluded. Gluinos as heavy as 2330 GeV are excluded. These represent the best limits to date for these simplified models, significantly extending the exclusions from previous results. This search is also used in tandem with five other analyses in an electroweakino combined search, extending the limits further than any individual analysis. In addition to these searches, this thesis describes the on-going and future upgrades to the LHC and the CMS detector and experiment. The upgrades are designed to substantially increase the instantaneous luminosity of the LHC, making it easier for future analyses to probe new physics.

## **Acknowledgements**

I would like to thank everyone who helped me get to this point, from my friends and family to my advisor. Without their help, support, and encouragement, I can guarantee this wouldn't have happened.

## Contents

### Chapter

|          |                                                     |          |
|----------|-----------------------------------------------------|----------|
| <b>1</b> | <b>Introduction</b>                                 | <b>1</b> |
| <b>2</b> | <b>Theoretical background</b>                       | <b>4</b> |
| 2.1      | The Standard Model of particle physics . . . . .    | 4        |
| 2.1.1    | Fermions . . . . .                                  | 8        |
| 2.1.2    | Bosons . . . . .                                    | 9        |
| 2.1.3    | The fundamental forces . . . . .                    | 9        |
| 2.1.4    | Lagrangian of the Standard Model . . . . .          | 10       |
| 2.1.5    | Quantum electrodynamics . . . . .                   | 11       |
| 2.1.6    | Weak interactions . . . . .                         | 12       |
| 2.1.7    | Quantum chromodynamics . . . . .                    | 15       |
| 2.1.8    | Electroweak theory . . . . .                        | 17       |
| 2.1.9    | The Higgs mechanism . . . . .                       | 19       |
| 2.1.10   | The parameters of the Standard Model . . . . .      | 21       |
| 2.2      | Supersymmetry . . . . .                             | 22       |
| 2.2.1    | The fine-tuning problem . . . . .                   | 24       |
| 2.2.2    | The Minimal Supersymmetric Standard Model . . . . . | 27       |
| 2.2.3    | SUSY breaking . . . . .                             | 28       |
| 2.2.4    | R-parity . . . . .                                  | 29       |
| 2.2.5    | SUSY searches . . . . .                             | 31       |



|          |                                          |           |
|----------|------------------------------------------|-----------|
| <b>3</b> | <b>The Large Hadron Collider</b>         | <b>35</b> |
| <b>4</b> | <b>The CMS Detector</b>                  | <b>39</b> |
| 4.1      | Relevant kinematic variables . . . . .   | 40        |
| 4.2      | Silicon tracker . . . . .                | 41        |
| 4.2.1    | Pixel detector . . . . .                 | 45        |
| 4.2.2    | Strip detector . . . . .                 | 46        |
| 4.3      | Electromagnetic calorimeter . . . . .    | 46        |
| 4.4      | Hadronic calorimeter . . . . .           | 50        |
| 4.4.1    | Hybrid photodiodes . . . . .             | 52        |
| 4.4.2    | Barrel and endcap calorimeters . . . . . | 52        |
| 4.4.3    | Outer calorimeter . . . . .              | 53        |
| 4.4.4    | Forward calorimeter . . . . .            | 53        |
| 4.5      | Superconducting magnet . . . . .         | 53        |
| 4.6      | Muon system . . . . .                    | 54        |
| 4.6.1    | Drift tubes . . . . .                    | 56        |
| 4.6.2    | Cathode strip chambers . . . . .         | 57        |
| 4.6.3    | Resistive plate chambers . . . . .       | 57        |
| 4.7      | Trigger system . . . . .                 | 58        |
| <b>5</b> | <b>Event reconstruction</b>              | <b>62</b> |
| 5.1      | Basic detector elements . . . . .        | 63        |
| 5.1.1    | Tracks . . . . .                         | 64        |
| 5.1.2    | ECAL & HCAL superclusters . . . . .      | 66        |
| 5.1.3    | Muon tracks . . . . .                    | 67        |
| 5.2      | Particle flow . . . . .                  | 67        |
| 5.3      | High-level objects . . . . .             | 69        |
| 5.3.1    | Jets . . . . .                           | 69        |

|          |                                                                                                   |           |
|----------|---------------------------------------------------------------------------------------------------|-----------|
| 5.3.2    | b-tagging jets . . . . .                                                                          | 69        |
| 5.3.3    | Missing energy . . . . .                                                                          | 72        |
| <b>6</b> | <b>Search for supersymmetry in channels with two Higgs bosons and missing transverse momentum</b> | <b>74</b> |
| 6.1      | Signal models . . . . .                                                                           | 75        |
| 6.2      | Data and simulation . . . . .                                                                     | 77        |
| 6.2.1    | Data and triggers . . . . .                                                                       | 78        |
| 6.2.2    | Simulation . . . . .                                                                              | 82        |
| 6.3      | Event selection . . . . .                                                                         | 86        |
| 6.3.1    | Baseline selection . . . . .                                                                      | 86        |
| 6.3.2    | Higgs candidate reconstruction . . . . .                                                          | 91        |
| 6.4      | Background estimation . . . . .                                                                   | 92        |
| 6.4.1    | Region definition . . . . .                                                                       | 93        |
| 6.4.2    | The ABCD method . . . . .                                                                         | 96        |
| 6.4.3    | The resolved signature . . . . .                                                                  | 98        |
| 6.4.4    | The boosted signature . . . . .                                                                   | 100       |
| 6.5      | Validation and systematic uncertainties . . . . .                                                 | 103       |
| 6.5.1    | Resolved background uncertainties . . . . .                                                       | 105       |
| 6.5.2    | Boosted background uncertainties . . . . .                                                        | 114       |
| 6.5.3    | Signal systematic uncertainties . . . . .                                                         | 120       |
| 6.6      | Results . . . . .                                                                                 | 124       |
| 6.6.1    | Resolved observed data yields . . . . .                                                           | 127       |
| 6.6.2    | Boosted observed data yields . . . . .                                                            | 127       |
| 6.6.3    | Statistical analysis . . . . .                                                                    | 131       |
| 6.6.4    | Simplified model interpretations . . . . .                                                        | 133       |

|          |                                                                    |            |
|----------|--------------------------------------------------------------------|------------|
| <b>7</b> | <b>Electroweak combined search</b>                                 | <b>139</b> |
| 7.1      | Individual searches . . . . .                                      | 141        |
| 7.1.1    | Search for one lepton and two b jets . . . . .                     | 141        |
| 7.1.2    | Search for an oppositely charged same-flavor lepton pair . . . . . | 142        |
| 7.1.3    | Search for two or three soft leptons . . . . .                     | 142        |
| 7.1.4    | Search for three or more leptons . . . . .                         | 143        |
| 7.1.5    | Search for multiple jets . . . . .                                 | 143        |
| 7.2      | Correlated systematics . . . . .                                   | 144        |
| 7.3      | Interpretations . . . . .                                          | 144        |
| <b>8</b> | <b>High-Luminosity LHC</b>                                         | <b>150</b> |
| 8.1      | LHC injector upgrade project . . . . .                             | 152        |
| 8.2      | CMS Phase 2 upgrade . . . . .                                      | 154        |
| 8.2.1    | Tracker . . . . .                                                  | 154        |
| 8.2.2    | Calorimeter endcaps . . . . .                                      | 155        |
| 8.2.3    | Muon endcaps . . . . .                                             | 156        |
| 8.3      | Phase 2 Level-1 Trigger . . . . .                                  | 156        |
| 8.4      | Track Finder . . . . .                                             | 157        |
| 8.4.1    | Tracking algorithm . . . . .                                       | 159        |
| 8.5      | Global Track Trigger . . . . .                                     | 161        |
| 8.6      | Track-based missing energy . . . . .                               | 163        |
| 8.6.1    | Track quality . . . . .                                            | 164        |
| 8.6.2    | Track-based missing energy performance . . . . .                   | 166        |
| 8.7      | Vertex finding . . . . .                                           | 167        |
| 8.8      | Track jets . . . . .                                               | 168        |
| 8.9      | Trigger demonstrator tests . . . . .                               | 169        |
| 8.10     | Data quality monitoring . . . . .                                  | 170        |

|          |                    |            |
|----------|--------------------|------------|
| <b>9</b> | <b>Conclusions</b> | <b>176</b> |
|----------|--------------------|------------|

|  |                     |            |
|--|---------------------|------------|
|  | <b>Bibliography</b> | <b>179</b> |
|--|---------------------|------------|

## **Appendix**

|          |                                                           |            |
|----------|-----------------------------------------------------------|------------|
| <b>A</b> | <b>Additional interpretations for the EWK Combination</b> | <b>189</b> |
|----------|-----------------------------------------------------------|------------|

|     |                             |     |
|-----|-----------------------------|-----|
| A.1 | Additional models . . . . . | 189 |
|-----|-----------------------------|-----|

|     |                                                                    |     |
|-----|--------------------------------------------------------------------|-----|
| A.2 | Search for an oppositely charged same-flavor lepton pair . . . . . | 190 |
|-----|--------------------------------------------------------------------|-----|

|     |                           |     |
|-----|---------------------------|-----|
| A.3 | Interpretations . . . . . | 190 |
|-----|---------------------------|-----|

## Tables

### Table

|      |                                                                          |     |
|------|--------------------------------------------------------------------------|-----|
| 2.1  | SM Fermions . . . . .                                                    | 7   |
| 2.2  | SM Bosons . . . . .                                                      | 7   |
| 2.3  | MSSM particles . . . . .                                                 | 30  |
| 6.1  | Data sets . . . . .                                                      | 78  |
| 6.2  | SM samples . . . . .                                                     | 84  |
| 6.3  | Rarer SM samples . . . . .                                               | 85  |
| 6.4  | Signal samples . . . . .                                                 | 86  |
| 6.5  | Boosted MC region yields . . . . .                                       | 104 |
| 6.6  | Resolved background uncertainties . . . . .                              | 106 |
| 6.7  | Data yields and associated uncertainties in the 0H+b region . . . . .    | 116 |
| 6.8  | $\beta_{\text{norm}}$ in the validation samples . . . . .                | 119 |
| 6.9  | Boosted background uncertainties . . . . .                               | 120 |
| 6.10 | Signal systematics summary . . . . .                                     | 125 |
| 6.11 | Resolved observed yields . . . . .                                       | 130 |
| 6.12 | Boosted observed yields . . . . .                                        | 131 |
| 7.1  | Topologies considered for the EWK Combination . . . . .                  | 141 |
| 7.2  | Summary of systematic uncertainties for EWK Combination . . . . .        | 145 |
| 8.1  | $\Delta z$ requirements for track $p_{\text{T}}^{\text{miss}}$ . . . . . | 163 |

|     |                                                                               |     |
|-----|-------------------------------------------------------------------------------|-----|
| 8.2 | Track quality criteria for track $p_T^{\text{miss}}$ and track jets . . . . . | 164 |
| 8.3 | FPGA resource usage for the GTT system . . . . .                              | 170 |
| A.1 | All topologies considered for the EWK Combination . . . . .                   | 191 |

## Figures

### Figure

|      |                                                             |    |
|------|-------------------------------------------------------------|----|
| 2.1  | SM particles . . . . .                                      | 5  |
| 2.2  | Example of Feynman diagram . . . . .                        | 11 |
| 2.3  | Beta minus decay . . . . .                                  | 13 |
| 2.4  | Meson and baryon multiplets . . . . .                       | 16 |
| 2.5  | CMS cross sections . . . . .                                | 23 |
| 2.6  | Top loop corrections . . . . .                              | 24 |
| 2.7  | Stop loop corrections . . . . .                             | 26 |
| 2.8  | SUSY cross sections . . . . .                               | 31 |
| 2.9  | Gluino simplified models . . . . .                          | 33 |
| 2.10 | Electroweakino simplified models . . . . .                  | 33 |
| 2.11 | CMS gluino mass limits . . . . .                            | 34 |
| 2.12 | CMS electroweakino mass limits . . . . .                    | 34 |
| 3.1  | CERN accelerator complex . . . . .                          | 36 |
| 3.2  | Track and vertex reconstruction of a single event . . . . . | 38 |
| 4.1  | Cutaway diagram of CMS . . . . .                            | 40 |
| 4.2  | Transverse slice of CMS . . . . .                           | 42 |
| 4.3  | Schematic of silicon ionization . . . . .                   | 43 |
| 4.4  | Phase 1 tracker . . . . .                                   | 44 |

|      |                                                                     |     |
|------|---------------------------------------------------------------------|-----|
| 4.5  | Track $z_0$ position and primary vertex resolution . . . . .        | 45  |
| 4.6  | Calorimeter showers in CMS . . . . .                                | 47  |
| 4.7  | Schematic of the electromagnetic calorimeter . . . . .              | 48  |
| 4.8  | Schematic of the layers in the preshower detector . . . . .         | 49  |
| 4.9  | Electromagnetic calorimeter mass resolution . . . . .               | 50  |
| 4.10 | Schematic of the HCAL . . . . .                                     | 51  |
| 4.11 | Improved jet momenta resolution from the HCAL upgrade . . . . .     | 52  |
| 4.12 | Magnetic field inside CMS . . . . .                                 | 54  |
| 4.13 | The muon system . . . . .                                           | 55  |
| 4.14 | Drift tubes . . . . .                                               | 56  |
| 4.15 | Cathode strip chambers . . . . .                                    | 57  |
| 4.16 | Resistive plate chambers . . . . .                                  | 59  |
| 4.17 | Phase 1 L1 Trigger . . . . .                                        | 60  |
| 5.1  | Secondary vertex for jets . . . . .                                 | 70  |
| 5.2  | Phase 1 $p_T^{\text{miss}}$ resolution . . . . .                    | 73  |
| 6.1  | Resolved and boosted topologies . . . . .                           | 76  |
| 6.2  | SUSY models . . . . .                                               | 77  |
| 6.3  | Integrated luminosities . . . . .                                   | 79  |
| 6.4  | Measured trigger efficiencies . . . . .                             | 81  |
| 6.5  | Signal vs background topology . . . . .                             | 87  |
| 6.6  | Analysis regions . . . . .                                          | 93  |
| 6.7  | Key variables for resolved event selection . . . . .                | 95  |
| 6.8  | Boosted mass regions . . . . .                                      | 96  |
| 6.9  | Key variables for the boosted event selection . . . . .             | 97  |
| 6.10 | $\langle m_{\text{bb}} \rangle$ shape in b-tag categories . . . . . | 99  |
| 6.11 | $\kappa$ corrections . . . . .                                      | 100 |



|      |                                                                                                                  |     |
|------|------------------------------------------------------------------------------------------------------------------|-----|
| 6.12 | Boosted background composition . . . . .                                                                         | 102 |
| 6.13 | Boosted $p_T^{\text{miss}}$ shape agreement . . . . .                                                            | 102 |
| 6.14 | Boosted signal to sideband yields ratio . . . . .                                                                | 104 |
| 6.15 | Boosted MC background prediction . . . . .                                                                       | 104 |
| 6.16 | $\langle m_{\text{bb}} \rangle$ and $\Delta R_{\text{max}}$ in the single-lepton validation sample . . . . .     | 108 |
| 6.17 | $\kappa$ values in the single-lepton validation sample . . . . .                                                 | 109 |
| 6.18 | $\langle m_{\text{bb}} \rangle$ and $\Delta R_{\text{max}}$ in the dilepton validation sample . . . . .          | 111 |
| 6.19 | $\kappa$ values in the dilepton validation sample . . . . .                                                      | 112 |
| 6.20 | $\langle m_{\text{bb}} \rangle$ and $\Delta R_{\text{max}}$ in the low- $\Delta\phi$ validation sample . . . . . | 113 |
| 6.21 | $\kappa$ values in the low- $\Delta\phi$ validation sample . . . . .                                             | 114 |
| 6.22 | Calculation of the $p_T^{\text{miss}}$ shape closure systematic . . . . .                                        | 116 |
| 6.23 | $f_{\text{bkg}}$ agreement in the validation samples . . . . .                                                   | 118 |
| 6.24 | Observed and predicted yields . . . . .                                                                          | 126 |
| 6.25 | Observed $\langle m_{\text{bb}} \rangle$ distributions in $\Delta R_{\text{max}}$ . . . . .                      | 128 |
| 6.26 | Observed $\langle m_{\text{bb}} \rangle$ distributions in $p_T^{\text{miss}}$ . . . . .                          | 129 |
| 6.27 | Observed $m_J$ distributions . . . . .                                                                           | 132 |
| 6.28 | TChiHH-G combined limits . . . . .                                                                               | 135 |
| 6.29 | TChiHH-G limits, resolved vs boosted . . . . .                                                                   | 136 |
| 6.30 | TChiHH limits, 2D mass . . . . .                                                                                 | 137 |
| 6.31 | T5HH limits . . . . .                                                                                            | 138 |
| 7.1  | GMSB signal models for the EWK Combination . . . . .                                                             | 140 |
| 7.2  | Higgsino-bino signal models for the EWK Combination . . . . .                                                    | 140 |
| 7.3  | Best analysis for neutralino-neutralino production to H . . . . .                                                | 147 |
| 7.4  | Neutralino-neutralino results for the EWK Combination . . . . .                                                  | 148 |
| 7.5  | Neutralino-neutralino results for H boson production . . . . .                                                   | 149 |
| 7.6  | Higgsino-bino exclusion limits . . . . .                                                                         | 149 |

|      |                                                                                       |     |
|------|---------------------------------------------------------------------------------------|-----|
| 8.1  | HL-LHC schedule . . . . .                                                             | 151 |
| 8.2  | Phase 2 tracker . . . . .                                                             | 155 |
| 8.3  | Phase 2 L1 trigger schematic . . . . .                                                | 158 |
| 8.4  | Phase 2 track $z_0$ resolution . . . . .                                              | 159 |
| 8.5  | A stub from a sensor module . . . . .                                                 | 160 |
| 8.6  | Phase 2 track-finding efficiency . . . . .                                            | 160 |
| 8.7  | Improved tracking efficiency for displaced signatures . . . . .                       | 162 |
| 8.8  | Bend $\chi^2$ as a track quality variable . . . . .                                   | 166 |
| 8.9  | L1 rate and trigger efficiency for track $p_T^{\text{miss}}$ . . . . .                | 167 |
| 8.10 | FastHisto vertexing algorithm . . . . .                                               | 173 |
| 8.11 | Primary vertex resolution . . . . .                                                   | 173 |
| 8.12 | Diagram of the two-layer jet clustering algorithm . . . . .                           | 174 |
| 8.13 | Track jet efficiency . . . . .                                                        | 174 |
| 8.14 | Track-based MHT trigger efficiency . . . . .                                          | 175 |
| A.1  | Slepton production . . . . .                                                          | 189 |
| A.2  | $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ signal models for the EWK Combination . . . . . | 190 |
| A.3  | Neutralino-chargino results for the EWK Combination . . . . .                         | 191 |
| A.4  | Best analysis for WZ and WH topologies . . . . .                                      | 192 |
| A.5  | Exclusion for the WZ and WH topologies . . . . .                                      | 193 |
| A.6  | Preliminary results for slepton pair production . . . . .                             | 194 |

## Chapter 1

### Introduction

The heart of particle physics is attempting to understand, study, classify, and predict the fundamental pieces of matter and their basic interactions. High energy particle colliders, like the Large Hadron Collider (LHC) [1], are a very effective way to produce subatomic particles and observe their interactions. In 2015, the LHC broke the world record for the highest energy collisions [2] and has continued to collide protons at a center-of-mass energy of 13 TeV, allowing for incredibly large data samples to be collected and studied. Along the LHC's 27-kilometer circumference sit four experimental detectors (CMS and ATLAS being the two general-purpose detectors), each of which collects and processes the collision decay products. Through decades of theory and experimentation, we have developed a thorough and consistent description of all observed subatomic particles and the strong and electroweak forces that form our Universe. This theory, the Standard Model (SM) of particle physics [3], was completed when the last remaining particle, the Higgs boson, was discovered jointly by the CMS and ATLAS collaborations in 2012 [4, 5].

There are, however, physical and theoretical reasons to believe that the Standard Model is not a complete description. For physical motivations, the Standard Model has no accurate description of gravity, a force we see interacting macroscopically. We are also aware of the presence of dark matter, or at least its effects in cosmological observations, which is believed to account for 85% of the total matter in the Universe, yet the Standard Model has no particle candidate for dark matter. For theoretical motivations, one indicator that the SM is incomplete is known as the fine-tuning problem. Since the Higgs boson is a scalar particle, its mass receives quantum corrections that

are strongly dependent on a large-scale cutoff, like the Planck scale. This implies the mass of the Higgs boson should be many orders of magnitude larger than it is experimentally measured. Only by severely fine-tuning the bare mass of the Higgs can we get the experimentally-determined Higgs mass; fine-tuning model parameters is usually an indicator of an incomplete theory, especially if there is no physical explanation or driving force for why fine-tuning is needed.

This thesis investigates a potential extension to the Standard Model that could help solve these and other problems the SM alone cannot fix. Supersymmetry (SUSY) introduces space-time symmetries that relate fermionic and bosonic degrees of freedom, giving each SM particle a “superpartner” that differs by  $\frac{1}{2}$ -unit of spin. The simplest of these SUSY models is called the Minimal Supersymmetric Standard Model (MSSM) and contains the smallest (“minimal”) number of new particle states and interactions [6]. An exact symmetry would require a SM particle and its corresponding superpartner to have identical mass; however, we have not observed any of these superpartners and thus it is assumed the symmetry is broken and the superpartners must have remarkably heavier masses. SUSY gives a natural solution to the fine-tuning problem and provides a candidate for dark matter: the lightest SUSY particle (LSP). The LSP is weakly-interacting and would escape detection from any particle detector, taking with it some energy from the collision. The key, then, to searches for SUSY is being able to successfully measure and reconstruct the missing energy in an event collision.

The search presented in this thesis is a search for physics beyond the SM. In Chapter 2, we examine the current theoretical understanding of particle physics, used as motivation for our search. In Section 2.1, we describe the Standard Model of particle physics, including its immensely successful description of three of the four fundamental forces, through the Lagrangian for each force and its associated particles. We also touch on some of the limitations of this theory. Section 2.2 describes the theory of supersymmetry (SUSY) and why it is such an attractive solution to the fine-tuning problem.

We then move on to the experimental apparatus. Chapter 3 presents a description of the Large Hadron Collider (LHC), the high-energy proton-proton collider that allows us to study particle

collisions and interactions, while Chapter 4 gives a thorough explanation of the CMS detector, which is responsible for particle detection and reconstruction from the collision data it collects. The event reconstruction, including particle identification, is included in Chapter 5.

The focus of this thesis, a search for new particles predicted by SUSY, is presented in Chapter 6. We begin with a motivation for the search, as well as a discussion of the SUSY signal models in which we interpret our results. A description of the data samples and the generation requirements for simulated samples is given in Section 6.2. The event criteria we impose to select the events in line with our analysis signature, including large amounts of missing energy, are listed in Section 6.3. The SM background estimation methods are described in Section 6.4, and the validation of these methods is given in Section 6.5. This section also includes details on the systematic uncertainties that are derived and assigned to both the background predictions and to the SUSY signal models used to interpret our results. Section 6.6 shows the results of the analysis, including the interpretations in the context of these signal models. After the description of this search, Chapter 7 describes a different, combined electroweak analysis that uses the results of Chapter 6 along with five other analyses for a broader range of interpretations.

Finally, a description of the Phase 2 upgrade of CMS for the High-Luminosity LHC (HL-LHC) data-taking period is given in Chapter 8, which includes work toward improving the performance of the data acquisition system for future runs. A summary of the two searches in Chapters 6 and 7, including how the Phase 2 upgrade will benefit future SUSY analyses, is given in Chapter 9.

## Chapter 2

### Theoretical background

In this Chapter, we examine the current theoretical understanding of particle physics. The Standard Model (SM) is the theory that describes three of the four fundamental forces and classifies the known constituents of matter. The SM was developed in stages throughout the second half of the 20th century, driven by theorists and experimentalists alike, and has been tremendously successful; however, there are a variety of missing pieces (such as gravity, the fourth fundamental force) that the SM cannot explain. This thesis investigates an extension to the SM called supersymmetry, so first we must have a good understanding of the SM itself and why it is an incomplete picture. In Section 2.1, we describe the Standard Model of particle physics, deriving its Lagrangian and highlighting deficiencies in the current theory. In Section 2.2, we examine the theory of supersymmetry and explain why it is an appealing extension.

#### 2.1 The Standard Model of particle physics

The Standard Model (SM) of particle physics describes three of the four fundamental forces and the subatomic particles that mediate and interact via these forces. The subatomic particles are broken down into fermionic particles and bosonic force carriers, and the SM describes how these particles interact with each other through quantum field theory. Figure 2.1 shows the elementary particles of the SM and their properties. The theories summarized in the rest of this section are adapted from Refs. [3, 7–10].

In the SM, elementary particles are categorized by half-integer (fermions) or integer (bosons)

## Standard Model of Elementary Particles

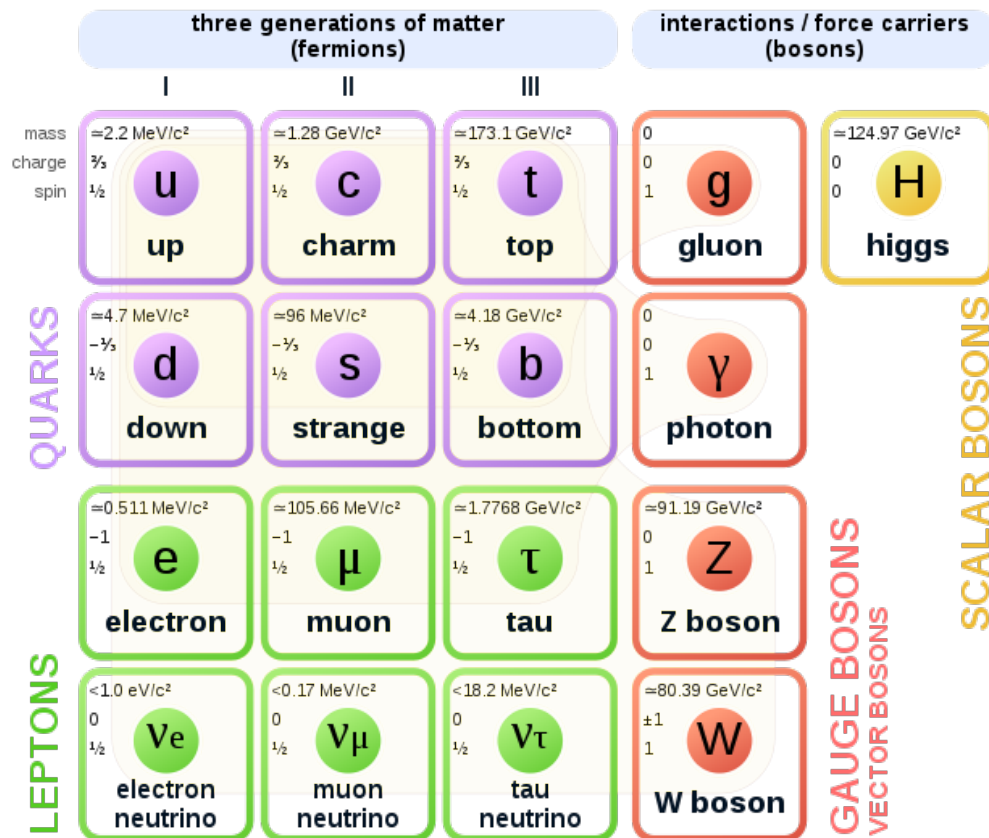


Figure 2.1: The Standard Model particles. [11]

spin. Fermions are categorized as quarks or leptons and make up the matter that we see. They are classified into three families or generations, as shown in Table 2.1. Bosons are separated into scalar bosons, such as the Higgs boson, and vector or gauge bosons, the force “carriers” or “mediators” responsible for force interactions. The gauge bosons only interact with particles that have the right kind of charge. The properties for all the integer-spin particles of the SM are listed in Table 2.2. Fermions and the bosons they interact with are described in more detail in Sections 2.1.1 and 2.1.2, respectively. The manifestation of these interactions, the fundamental forces, is described in Section 2.1.3.

The Lagrangian for the SM is briefly introduced in Section 2.1.4. We will go through its derivation in the sections that follow, through the symmetry group of the SM:  $SU(3) \times SU(2) \times U(1)$ . The unitary group  $U(1)$  describes electromagnetism, discussed in Section 2.1.5.  $SU(2)$  models the weak interaction between fermions and three gauge bosons, described in further detail in Section 2.1.6.  $SU(3)$  governs the strong interaction and is used in quantum chromodynamics, discussed in Section 2.1.7. Discussed further in Section 2.1.8,  $SU(2) \times U(1)$  generates the electroweak theory, describing both the electromagnetic and weak nuclear forces simultaneously, which breaks down to  $U(1)_{\text{EM}}$  when the electroweak symmetry is spontaneously broken. Lastly, the Higgs mechanism, the spontaneous symmetry breaking caused by interactions between the Higgs field and bosons, generates the masses of the particles and is discussed in Section 2.1.9. A list of the parameters of the SM is given in Section 2.1.10, all of which must be determined experimentally.



|         | Name              | Symbol     | Mass                       | Charge | Generation |
|---------|-------------------|------------|----------------------------|--------|------------|
| Leptons | electron          | e          | 0.511 MeV                  | -1     | 1          |
|         | electron neutrino | $\nu_e$    | <1.1 eV                    | 0      | 1          |
|         | muon              | $\mu$      | 106 MeV                    | -1     | 2          |
|         | muon neutrino     | $\nu_\mu$  | <0.19 MeV                  | 0      | 2          |
|         | tau               | $\tau$     | 1.78 GeV                   | -1     | 3          |
|         | tau neutrino      | $\nu_\tau$ | <18.2 MeV                  | 0      | 3          |
| Quarks  | up                | u          | $2.16^{+0.49}_{-0.26}$ MeV | +2/3   | 1          |
|         | down              | d          | $4.67^{+0.48}_{-0.17}$ MeV | -1/3   | 1          |
|         | charm             | c          | $1.27 \pm 0.02$ GeV        | +2/3   | 2          |
|         | strange           | s          | $93^{+11}_{-5}$ MeV        | -1/3   | 2          |
|         | top               | t          | $172.76 \pm 0.30$ GeV      | +2/3   | 3          |
|         | bottom            | b          | $4.18^{+0.03}_{-0.02}$ GeV | -1/3   | 3          |

Table 2.1: The spin- $\frac{1}{2}$  particles of the SM [12].

|               | Name        | Symbol   | Interaction     | Charge  | Mass [GeV]           |
|---------------|-------------|----------|-----------------|---------|----------------------|
| Vector bosons | Photon      | $\gamma$ | Electromagnetic | 0       | -                    |
|               | W bosons    | $W^\pm$  | Weak            | $\pm 1$ | $80.379 \pm 0.012$   |
|               | Z boson     | Z        | Weak            | 0       | $91.1876 \pm 0.0021$ |
|               | Gluon       | g        | Strong          | 0       | -                    |
| Scalar bosons | Higgs boson | H        | -               | 0       | $125.25 \pm 0.17$    |

Table 2.2: The integer-spin particles of the SM, and the force they mediate [12].

### 2.1.1 Fermions

Fermions are represented as solutions to the free-particle Dirac equation, with the Lagrangian and resulting equation of motion:

$$\begin{aligned}\mathcal{L} &= i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi, \text{ and} \\ i\gamma^\mu\partial_\mu\psi - m\psi &= 0.\end{aligned}\tag{2.1}$$

The SM is a chiral theory, meaning left-handed and right-handed fields transform differently. The left- and right-handed fields are eigenstates of the chiral projection operators  $P_L$  and  $P_R$ , and the charge conjugate of a right-handed field is left-handed:

$$\begin{aligned}P_L\psi_L &\equiv \frac{1-\gamma^5}{2}\psi_L, & P_L\bar{\psi}_R &= \bar{\psi}_R, & P_L\psi_R &= P_L\bar{\psi}_L = 0, \\ P_R\psi_R &\equiv \frac{1+\gamma^5}{2}\psi_R, & P_R\bar{\psi}_L &= \bar{\psi}_L, & P_R\psi_L &= P_R\bar{\psi}_R = 0,\end{aligned}\tag{2.2}$$

where  $\gamma^5$  is the fifth gamma matrix ( $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ ). The left-handed quarks and leptons are arranged into three generations and come in doublets of  $\pm\frac{1}{2}$  weak isospin,

$$\begin{aligned}Q^\alpha &= \begin{pmatrix} u^\alpha \\ d^\alpha \end{pmatrix}, \quad \begin{pmatrix} c^\alpha \\ s^\alpha \end{pmatrix}, \quad \begin{pmatrix} t^\alpha \\ b^\alpha \end{pmatrix}, \\ L &= \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix},\end{aligned}\tag{2.3}$$

where  $\alpha$  is the color index for the quarks. The quark and lepton doublets are contained in  $\Psi_L$ :

$$\Psi_L^\top = (Q_1, \bar{u}_1, \bar{e}_1, \bar{d}_1, L_1, \quad Q_2, \dots, L_2, \quad Q_3, \dots, L_3),\tag{2.4}$$

where the bar indicates the antiparticle (charge conjugate field) and the color-charge  $\alpha$  for the quarks is suppressed. The corresponding right-handed singlets  $\psi_R$  (with a weak isospin of 0) are:

$$\begin{aligned}u_\alpha &= \bar{u}, \bar{c}, \bar{t}, \\ d_\alpha &= \bar{d}, \bar{s}, \bar{b}, \\ e &= \bar{e}, \bar{\mu}, \bar{\tau}.\end{aligned}\tag{2.5}$$

### 2.1.2 Bosons

The SM has twelve gauge bosons (the photon, three weak bosons, and eight gluons) and twelve gauge fields:

$$\begin{aligned} G_\mu^{1\dots 8} : & \quad \text{the gluon fields,} \\ W_\mu^{1,2,3}, B_\mu^0 : & \quad \text{the electroweak W and B boson fields.} \end{aligned} \tag{2.6}$$

The electroweak boson fields mix to create physically-observable states: the neutral Z boson, the massless and neutral photon, and the massive charged W bosons.

The Higgs field  $\phi$ , a complex scalar field and a weak isospin doublet, couples together left- and right-handed fermions and is responsible for giving the gauge bosons their masses. Like all fundamental fields, the Higgs field has a particle that mediates it. This is the Higgs boson, whose properties are listed in Table 2.2. Similarly, the Yukawa interaction is responsible for fermionic masses.

### 2.1.3 The fundamental forces

There are four fundamental forces in our Universe: the gravitational force, the electromagnetic force, the weak force, and the strong force. Only three of these forces are included in the SM – there is no explanation for gravity. These forces govern all interactions that occur in the Universe and result from the exchange of force-carrier particles called gauge bosons. Each force has its own corresponding boson: the strong force is mediated by the gluon, the W and Z bosons mediate the weak force, the photon carries the electromagnetic force, and the (still undiscovered) graviton mediates gravity.

As the name suggests, the strong force is the most powerful of the four. The strength of the strong interaction is represented with a coupling constant,  $\alpha_s$ , which is a dimensionless indicator of interaction strength. The electromagnetic force is the next strongest, with a coupling constant of  $\alpha=1/137$  at lower energy scales. Despite its name, the weak force is many times stronger than gravity, but is the weakest of the three forces included in the SM. The weak coupling constant is

typically between  $10^{-7}$  and  $10^{-6}$  in the energy range around 1 GeV, since coupling constants are dependent on the interaction and energy scale. Gravity is the weakest of these four forces, with a coupling constant of  $10^{-39}$ , but is not included in the SM.

The effective range of the forces – the required distance between two particles for there to be a high probability of interaction – also varies. Gravity and electromagnetism both have infinite range. Gravity acts on all particles and objects in the Universe, no matter the distance, while an electromagnetic wave will travel indefinitely and undiminished unless it encounters a particle that absorbs it. The strong and weak forces do not manifest macroscopically because they have very short effective ranges. The strong interaction has an effective range of  $10^{-15}$  m, slightly more than the radius of a nucleon. The weak interaction has an even shorter effective range of  $10^{-17}$  to  $10^{-16}$  m, about 1/1000 the diameter of a proton, due to the heavy masses of the W and Z bosons.

#### 2.1.4 Lagrangian of the Standard Model

The Lagrangian of the SM contains terms that couple the particles and their relevant interactions. The gauge interactions act on the fermions through the covariant derivative  $D_\mu$  (a partial derivative) and a field  $A_\mu$ :

$$D_\mu \Psi_L = (\partial_\mu + gA_\mu)\Psi_L. \quad (2.7)$$

Using this covariant derivative, we can construct the field strength tensor,

$$F_{\mu\nu} = -\frac{i}{g}[D_\mu, D_\nu]. \quad (2.8)$$

We can then write the Lagrangian for the SM in terms of the fields described in Sections 2.1.1 and 2.1.2:

$$\begin{aligned} \mathcal{L}_{\text{SM}} = & -\frac{1}{4}F_{\mu\nu}^a F_a^{\mu\nu} + \bar{\Psi}_L i\gamma^\mu D_\mu \Psi_L + \bar{\psi}_R i\gamma^\mu D_\mu \psi_R \\ & + (D^\mu \phi)^\dagger (D_\mu \phi) + \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2 \\ & - y_{ij} \tilde{\phi} \bar{Q}_i u_j - y_{ij} \phi \bar{Q}_i d_j - y_{ij} \phi \bar{L}_i e_j + hc, \end{aligned} \quad (2.9)$$

where  $hc$  is the Hermitian conjugate. The first three terms correspond to the interactions between fermions and gauge bosons. The middle three terms correspond to the Higgs mechanism, while

the last line corresponds to the Yukawa interaction, with  $y$  being the Yukawa couplings. The more compact (and popular for merchandise) form of the SM Lagrangian is:

$$\begin{aligned}
\mathcal{L}_{SM} = & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\
& + i\bar{\psi}\not{D}\psi + hc \\
& + |D_{\mu}\phi|^2 - V(\phi) \\
& + \bar{\psi}_i y_{ij} \psi_j \phi + hc.
\end{aligned} \tag{2.10}$$

We will go through these terms through the corresponding field theories and interactions in the following sections.

### 2.1.5 Quantum electrodynamics

Quantum electrodynamics (QED) is a quantum field theory of the electromagnetic force that mathematically describes how light and matter interact. QED describes the force between two electrically-charged particles as coming from the interaction and exchange of photons and is visually represented by Feynman diagrams that show electrons' and photons' momenta as lines and the interactions – absorption or creation of photons – as vertices. Figure 2.2 shows one such diagram.

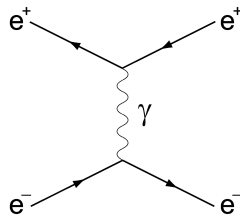


Figure 2.2: An example of a Feynman diagram, specifically electron-positron scattering [13].

The  $U(1)_{\text{EM}}$  symmetry group corresponds geometrically to a circle that is invariant to rotations and is associated with charge conservation. The Lagrangian of a free-particle Dirac fermion is invariant under a  $U(1)$  transformation,

$$\psi(x) \xrightarrow{U(1)} e^{iq\alpha} \psi(x), \tag{2.11}$$

where  $q$  is the charge and  $\alpha$  is a constant. If we allow the above U(1) transformation to be space-time dependent,  $\alpha(x)$ , we must introduce a new spin-1 vector field  $A^\mu$  for the Lagrangian to remain invariant and define our covariant derivative as follows:

$$\begin{aligned} D_\mu &:= \partial_\mu - ieA_\mu, \quad \text{and} \\ A_\mu &\xrightarrow{\text{U(1)}} A_\mu + \frac{1}{e}\partial_\mu\alpha. \end{aligned} \tag{2.12}$$

We can then define the whole QED Lagrangian:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.13}$$

with the electromagnetic field tensor  $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ . This spin-1 field  $A_\mu$  is the massless photon.

### 2.1.6 Weak interactions

Geometrically, SU(2) is the rotational symmetry of a sphere, and models the weak interactions between fermions and three bosons, the  $W^+$ ,  $W^-$ , and  $Z$  bosons. The properties of these bosons and fermions are described above in Table 2.2 and 2.1, respectively. The weak force conserves weak isospin  $T$ . More commonly  $T_3$ , the projection of the weak isospin along the quantization axis, is used. Left-handed fermions have a weak isospin of  $T_3 = \pm\frac{1}{2}$ , where the sign is the same as the electric charge of charged fermions<sup>1</sup>. Right-handed fermions have a weak isospin of  $T = T_3 = 0$ . The fermions involved in the weak force can be either elementary particles (leptons or quarks) or composite particles (e.g. protons), although all weak interactions are between elementary particles at the deepest level.

There are three basic types of weak interaction. Two of them involve charged, intermediate vector bosons and are called charged current interactions. The third type is called the neutral current interaction<sup>2</sup>. These interactions are:

<sup>1</sup> Neutrinos and anti-neutrinos are assigned the opposite sign of their corresponding charged lepton, since they are electrically neutral. Left-handed neutrinos have  $T_3 = +\frac{1}{2}$ , being paired with negatively charged left-handed leptons, while right-handed anti-neutrinos have  $T_3 = -\frac{1}{2}$ .

<sup>2</sup> It is interesting to note that these are called *charged* and *neutral* current interactions not because of the charge carried by the  $W$  and  $Z$  bosons, but because the weakly interacting fermions form a current with total electric charge

- (1) A charged lepton (for example, an electron or muon) can emit a  $W^-$  boson or absorb a  $W^+$  boson and convert into a corresponding neutrino, for example:  $\mu^- + W^+ \rightarrow \nu_\mu$ .
- (2) A down-type quark with charge  $-\frac{1}{3}$  can emit a  $W^-$  boson or absorb a  $W^+$  boson and convert it into a superposition of up-type quarks, for example:  $d \rightarrow u + W^-$ . The exact content of this superposition is given by the CKM matrix (see Ref. [14]).
- (3) Either a lepton or a quark can emit or absorb a Z boson, for example:  $e^- \rightarrow e^- + Z^0$ .

Unlike strong interactions, weak interactions allow quarks to change flavor, since the weak eigenstates are not exactly the mass eigenstates. This is a key component in the radioactive decay of atoms. In beta minus decays, shown in the Feynman diagram in Figure 2.3, the down quark inside a neutron changes into an up quark with the emission of an electron and an anti-electron neutrino, and the neutron becomes a proton.

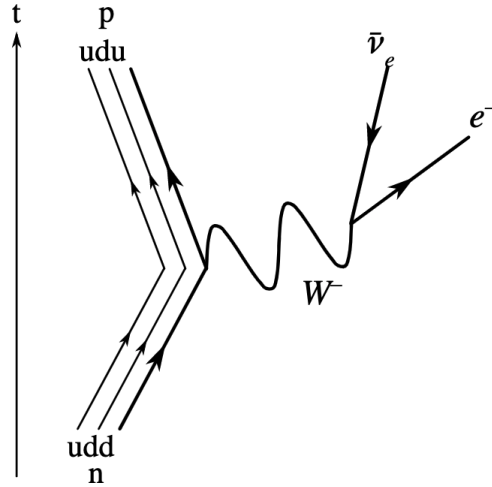


Figure 2.3: The Feynman diagram for a beta minus decay [15].

Let us postulate beta decays in the same manner Fermi did in 1934, suggesting that the interaction is due to a local current-current interaction:

$$\mathcal{L} = C \bar{\psi}_p \gamma^\mu \psi_n \bar{\psi}_e \gamma_\mu \psi_\nu + cc, \quad (2.14)$$

---

that is nonzero or zero. The naming convention predates the concept of a mediator boson. For example, in a charged current interaction, we can infer the presence of an electron neutrino if we identify an electron. In a neutral current interaction, we cannot directly identify the flavor of the neutrino.

with  $\psi_x$  for the interaction fields. After precise empirical measurements, the coefficients in Equation 2.14 have been well-determined:

$$\mathcal{L}_{\text{weak}} = -\frac{g_W}{\sqrt{2}} [\bar{\psi}_p \gamma^\lambda (1 - \frac{g_A}{g_v} \gamma_5) \psi_n [\bar{\psi}_e \gamma_\lambda (1 - \gamma_5) \psi_\nu] + cc. \quad (2.15)$$

In Equation 2.15, we see the vector and axial vector coupling constants,  $g_v$  and  $g_A$ , respectively. The vector coupling constant for quarks  $g_v$  is related to that for leptons:  $g_v = G_F V_{ud}$ , where  $G_F$  is the Fermi coupling constant and  $V_{ud}$  is one of the matrix elements in the CKM quark mixing matrix (see Ref. [14]). Historically,  $G_F$  was used to predict the masses of the W and Z bosons, since it provides a low energy value for the weak coupling. The ratio of the coupling constants is parameterized by  $\lambda = g_A/g_v$ , one of the free parameters of the SM listed in Section 2.1.10.

Focusing on the  $\bar{\psi}_e \psi_\nu$  interaction in Equation 2.15, let us subject the left-handed field  $(1 - \gamma_0) \psi_\nu$  to a parity transformation. This field possesses nonzero components only in the first two entries and is multiplied by  $\gamma^0$  for the transformation, which moves the nonzero components to the lower two entries. This produces the right-handed field  $(1 + \gamma_5) \gamma_0 \psi_\nu$ , which is not present in the equation. We see, then, that the weak interaction violates parity symmetry. Right-handed fermions can only participate in neutral current interactions, since they do not interact with W bosons.

Useful for particle identification, particle decays through the weak interaction tend to occur more slowly than other interactions. This is partially due to the large mass of the W bosons. The decay rate is proportional to

$$\frac{g_W^2}{Q^2 - m_W^2 c^2} \xrightarrow{Q \ll m_W c} \frac{g_W^2}{m_W^2 c^2}, \quad (2.16)$$

where  $g_W$  is the weak coupling constant and  $Q^2$  is the momentum transferred by the W boson. In the limit where  $Q \ll m_W c$ , the large W boson mass contributes to the longer lifetime. As an example, a neutral pion (which decays electromagnetically) has a lifetime of about  $10^{-16}$  seconds, whereas a charged pion has a lifetime of about  $10^{-8}$  seconds.



### 2.1.7 Quantum chromodynamics

Quantum chromodynamics (QCD) is the theory of interactions between quarks and gluons due to the strong force. QCD is a quantum field theory with symmetry group  $SU(3)$ , with spin-1 massless gluons as the force carrier. In the strong force, the three kinds of charge the quarks carry are referred to as “color” (categorized as red, green, and blue, and anti-red, anti-green, and anti-blue for the antiparticles), hence why we refer to the theory as *chromodynamics*. In addition to mediating the strong force, gluons themselves carry both color and anti-color and self-interact. There are six quarks categorized into three generations, whose properties are listed in Table 2.1. Each quark  $q$  has an antiquark  $\bar{q}$  with opposite electric and color charge.

$SU(3)$  symmetry is an exact symmetry under the different colors of quarks, mediated by gluons. According to  $SU(3)$ ,  $3 \times 3$  color combinations form an octet:

$$\begin{aligned} &|r\bar{g}\rangle, |r\bar{b}\rangle, |g\bar{b}\rangle, |g\bar{r}\rangle, |b\bar{r}\rangle, |b\bar{g}\rangle, \\ &\frac{1}{\sqrt{2}}(|r\bar{r}\rangle - |g\bar{g}\rangle), \frac{1}{\sqrt{6}}(|r\bar{r}\rangle + |g\bar{g}\rangle - 2|b\bar{b}\rangle), \end{aligned} \quad (2.17)$$

and a singlet:

$$\frac{1}{\sqrt{3}}(|r\bar{r}\rangle + |g\bar{g}\rangle + |b\bar{b}\rangle). \quad (2.18)$$

The color singlet is invariant under rotations in color space and therefore cannot be exchanged between color charges. It is, however, the color state of a  $q\bar{q}$  pair. Quarks are bound together by the strong force to create colorless hadrons, a requirement in QCD called color confinement. Mesons are quark-antiquark pairs and baryons are three-quark combinations. The color state of a baryon is  $\frac{1}{\sqrt{6}}(|rgb\rangle - |rbg\rangle + |gbr\rangle - |grb\rangle + |brg\rangle - |bgr\rangle)$ .

Let us look at just the three lightest quarks (u, d, s) to see how they combine into mesons and baryons<sup>3</sup>.  $q\bar{q}$  states can be combined into two nonets of spin-0 and spin-1 mesons, shown at the top of Figure 2.4. The spin-0 mesons include kaons and pions. The six mesons at the edges of each nonet are diagonally opposite their antiparticles, while the three in the center are their own antiparticle. Similarly, baryons are organized into a spin- $\frac{1}{2}$  octet, containing the proton and

---

<sup>3</sup> These quarks are an order of magnitude lighter than the other three, so this is a reasonable split.

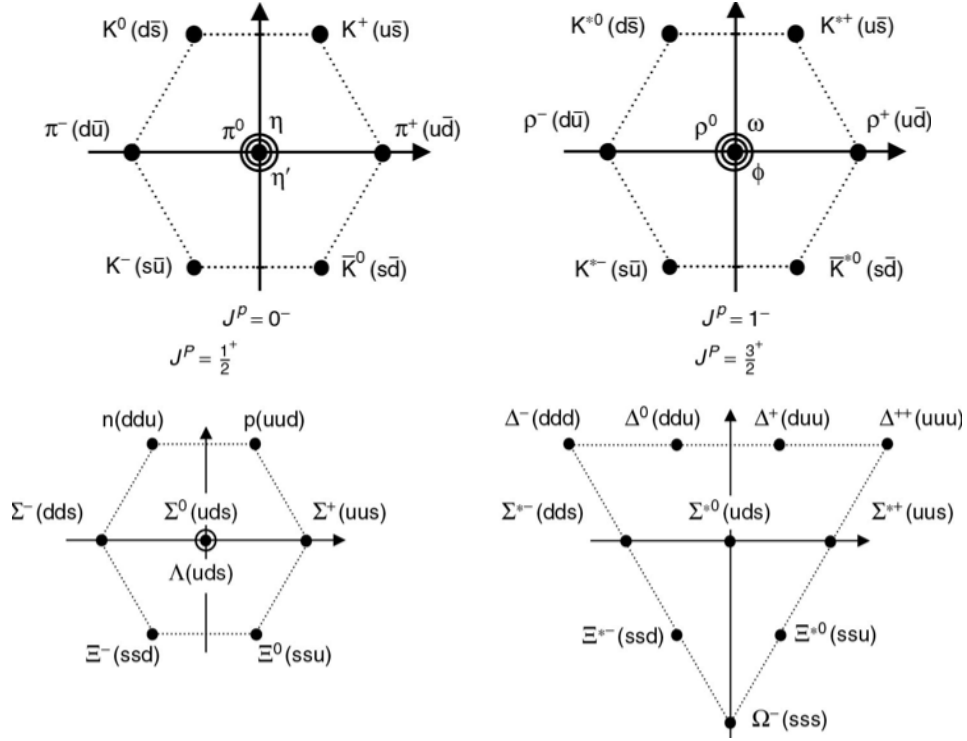


Figure 2.4: Meson (top row) and baryon (bottom row) multiplets allowed under flavor SU(3) symmetry. The  $q\bar{q}$  states are organized into nonets of spin-0 (top left) and spin-1 (top right) particles. The  $qqq$  states are organized into an octet of spin- $\frac{1}{2}$  particles (bottom left) and a decuplet of spin- $\frac{3}{2}$  particles (bottom right) [9].

neutron, and a spin- $\frac{3}{2}$  decuplet (bottom of Figure 2.4).

Due to color confinement, free quarks cannot be experimentally detected. Unlike in electromagnetism, the force between two color charges remains constant regardless of distance. As the charges are separated, the total energy of the system grows linearly with the separation distance until a  $q\bar{q}$  pair is formed from the vacuum, producing a new hadron. The process of forming new hadrons around a free gluon or quark is called hadronization and ensures only colorless states exist. When a quark or gluon is produced inside a particle accelerator, it causes new hadrons to form around it which travel in the same direction. This forms a narrow jet of particles. The jet is detected and measured to determine the properties of the original colored particle.

Let us consider the SU(3) symmetry for a quark  $q$ :

$$q \xrightarrow{\text{SU}(3)} e^{ig_s \xi_a \lambda_a / 2} q, \quad (2.19)$$

where  $g_s$  is the strong coupling constant,  $\xi$  is an 8-dimensional vector of arbitrary functions, and  $\lambda$  represents the eight  $3 \times 3$  Gell-Mann matrices [3]. The gauge covariant derivative, including the spin-1 vector fields  $G_\mu^{1 \cdots 8}$  to maintain SU(3) gauge invariance, is:

$$\begin{aligned} D_\mu &:= \partial_\mu - i \frac{g_s}{2} G_\mu^a \lambda_a, \quad \text{and} \\ G_\mu^c &\xrightarrow{\text{SU}(3)} G_\mu^c - \frac{1}{g_s} \partial_\mu \xi_c - f_{abc} \xi_a G_\mu^b, \end{aligned} \quad (2.20)$$

where  $f_{abc}$  are the structure constants of SU(3), satisfying  $[\lambda_a, \lambda_b] = 2i f_{abc} \lambda_c$ , and the indices  $a, b, c$  all range from 1 to 8. The complete Lagrangian for QCD can then be written as:

$$\mathcal{L}_{\text{QCD}} = \bar{q}(i\gamma^\mu D_\mu - m)q - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}, \quad (2.21)$$

with the field strength tensor for gluon field  $a$  defined as:  $G_{\mu\nu}^a \equiv \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c$ .

### 2.1.8 Electroweak theory

At first glance, the electromagnetic and weak forces appear quite different. The weak force has a very short effective range, at distances shorter than the diameter of a nucleus, while the electromagnetic force can extend indefinitely. Additionally, the weak force is roughly  $10^8$  times weaker than the electromagnetic force. Yet the electromagnetic force and the weak force can be understood as two facets of the same force, the electroweak force. The gauge group  $\text{SU}(2) \times \text{U}(1)$  provides a unified description of the electromagnetic and weak forces, the electroweak theory.

Historically, the electroweak theory arose from attempts to produce a self-consistent gauge theory, analogous to QED, for the weak force. It was discovered during the 1960s that this was indeed possible, but only if the theory also included the electromagnetic force. This theory required three massless bosons of weak isospin ( $W^1, W^2, W^3$ ) and a massless boson of weak hypercharge ( $B^0$ ). These are not yet physical fields, as we know the bosons that mediate the weak interaction have mass. One additional problem with this theory is that a fermionic mass term would couple left-handed and right-handed fermions, which belong to different weak representations.

Both problems can be solved by spontaneous symmetry breaking (SSB). The Higgs mechanism (discussed in the next section) describes how gauge bosons gain their masses through SSB

of a gauge theory. Additionally, an effective fermionic mass term can be written by coupling the operator that receives a vacuum expectation value (from SSB) to the fermions, if the operator has specific quantum numbers.

From observation, we know that the electroweak theory needs to break down to  $U(1)_{\text{EM}}$  after SSB and that the weak interaction can be accurately represented by  $SU(2)$ . However,  $SU(2) \times U(1)_{\text{EM}}$  cannot describe electroweak theory, since we know the  $W$  bosons are electrically charged, meaning the generator of  $U(1)_{\text{EM}}$  must not commute with some of the  $SU(2)$  generators (which it does by definition). This small problem is solved by introducing a different quantum number, weak hypercharge  $Y$ .

The symmetry breaking pattern  $SU(2) \times U(1)_Y \rightarrow U(1)_{\text{EM}}$ , then, is the minimal SSB pattern that accurately describes the Universe, and thus the gauge group of the electroweak theory. The generators of  $SU(2)$  and  $U(1)$  are the weak isospin ( $T$ ) and the weak hypercharge ( $Y$ ), respectively, and together give a relationship for the electric charge of a particle:

$$q = T_3 + \frac{1}{2}Y. \quad (2.22)$$

While some particles have a weak isospin of zero, all known spin- $\frac{1}{2}$  particles have a nonzero weak hypercharge. Neither weak isospin nor weak hypercharge are conserved when particles interact with the Higgs field; only the unique linear combination in Equation 2.22 is conserved. The strength of the particle interactions is determined by the vector and axial vector coupling constants  $g_v$  and  $g_A$ , defined in Section 2.1.6, whose values indicate the type of interaction:

Electromagnetic interactions:  $g_v = q, g_A = 0$ ,

Charged current weak interactions:  $g_v = -g_A = T$ , (2.23)

Neutral current weak interactions:  $g_v = T_3 - 2q \sin^2 \theta, g_A = -T_3$ .

Let us consider a  $U(1)$  transformation acting on fermion  $\psi$  and an  $SU(2)$  transformation on an isospin doublet  $\Psi$ :

$$\begin{aligned} \psi(x) &\xrightarrow{U(1)} e^{ig'Y\alpha(x)/2}\psi(x) \quad \text{and} \\ \Psi(x) &\xrightarrow{SU(2)} e^{ig_W\xi_i(x)\sigma_i/2}\Psi(x), \end{aligned} \quad (2.24)$$

where  $g'$  is the hypercharge coupling constant,  $Y$  is the hypercharge operator,  $g_W$  is the weak coupling constant,  $\alpha$  and  $\xi$  are arbitrary space-time dependent functions, and  $\sigma_i$  are the three Pauli spin matrices. We must introduce four new spin-1 vector fields for the Lagrangian to remain invariant,  $B_\mu$  and  $W_\mu^a$  ( $a=1,2,3$ ), and define our covariant derivative:

$$\begin{aligned} D_\mu &:= \partial_\mu - i\frac{g'}{2}YB_\mu - i\frac{g_W}{2}\sigma^a W_\mu^a, \\ B_\mu &\xrightarrow{\text{U}(1)} B_\mu - ig'\partial_\mu\alpha, \quad \text{and} \\ W_\mu^a &\xrightarrow{\text{SU}(2)} W_\mu^a - g_W\partial_\mu\xi^a - g_W\epsilon_{abc}\xi^b W_\mu^c, \end{aligned} \tag{2.25}$$

where  $\epsilon_{abc}$  are the structure constants of SU(2).

As discussed in Section 2.1.6, the weak interaction depends on particle chirality, as only left-handed particles and right-handed antiparticles interact with the W bosons [16]. Let us, then, account for this in the Lagrangian:

$$\mathcal{L}_{\text{EWK}} = \bar{\Psi}_L i\gamma^\mu D_\mu \Psi_L + \bar{\psi}_R i\gamma^\mu D'_\mu \psi_R - \frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}, \tag{2.26}$$

where  $\Psi_L$  are the left-handed quarks and leptons and  $\psi_R$  are the right-handed, as defined in Equations 2.3 and 2.5. The covariant derivative acting on the right-handed singlet ( $D'_\mu$ ) does not contain the third term in Equation 2.25, since right-handed fermions do not interact with the W bosons.

### 2.1.9 The Higgs mechanism

The SU(2) symmetry requires the W and Z bosons to be massless. Experimentally, we know this is not the case, and they, in fact, have relatively large masses ( $m_W = 80.4$  GeV,  $m_Z = 91.2$  GeV). The phrase ‘‘Higgs mechanism’’ refers to the W and Z bosons gaining mass through electroweak symmetry breaking. The mechanism adds a quantum field (the Higgs field) that permeates all Hilbert space and is described by an isospin doublet with both neutral and electrically-charged components:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{2.27}$$

Let us start by writing the Lagrangian for this field,

$$\mathcal{L}_\phi = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi), \quad (2.28)$$

where our covariant derivative  $D_\mu$  is the same as in Equation 2.25. The Higgs mechanism is accomplished by minimizing the Higgs potential

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad (2.29)$$

where the constants  $\mu^2$  and  $\lambda$  are positive. At the minimum, the ground state of the Higgs field is non-zero. We may choose the ground state such that:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix}, \quad (2.30)$$

where  $\nu$  is the acquired vacuum expectation value (VEV), satisfying  $\nu^2 = \mu^2/\lambda$ .

When the Higgs field develops a VEV, the symmetry is spontaneously broken, allowing the electroweak fields to mix and the Z and W bosons to acquire mass. If we substitute in this ground state and keep the resulting quadratic terms from the  $(D^\mu \phi)^\dagger (D_\mu \phi)$  term of Equation 2.28, we can see the effect on the gauge fields:

$$\frac{1}{8} \nu^2 [g_W^2 (W_\mu^1 W_1^\mu + W_\mu^2 W_2^\mu) + (g' B_\mu - g_W W_3^\mu)^2]. \quad (2.31)$$

The physical gauge bosons are excitations of linear combinations of these fields:

$$\begin{aligned} W_\mu^\pm &= \frac{W_\mu^1 \mp i W_\mu^2}{\sqrt{2}}, \\ Z_\mu &= \cos \theta_w W_\mu^3 - \sin \theta_w B_\mu, \\ A_\mu &= \sin \theta_w W_\mu^3 + \cos \theta_w B_\mu, \end{aligned} \quad (2.32)$$

where  $\theta_w$  is the Weinberg or weak mixing angle defined as  $\tan \theta_w = g'/g_W$ . The neutral gauge bosons  $W_\mu^3$  and  $B_\mu$  mix and have a  $2 \times 2$  non-diagonal mass matrix, giving us the mass eigenstates

for  $Z_\mu$  and  $A_\mu$ . We can then obtain the mass terms:

$$\begin{aligned} m_W &= \frac{\nu}{2} g_W, \\ m_Z &= \frac{m_W}{\cos \theta_w} = \frac{\nu}{2} \sqrt{g_W^2 + g'^2}, \\ m_A &= 0. \end{aligned} \tag{2.33}$$

We see that the photon ( $A_\mu$ ) remains massless after the symmetry is broken, agreeing with observations: the  $SU(2) \times U(1)$  symmetry is broken to  $U(1)_{\text{EM}}$ . The coupling constant for  $U(1)_{\text{EM}}$  ( $e$ ) can now be related to the couplings before symmetry breaking through the Weinberg angle,  $e = g_W \sin \theta_w = g' \cos \theta_w$ .

Like bosons, fermions also acquire mass through their interaction with the Higgs field [16], through Yukawa couplings which are proportional to the VEV  $\nu$ . Briefly, we define the Lagrangian as:

$$\mathcal{L}_{\text{Yukawa}} = -y_{ij} \tilde{\phi} \bar{Q}_i u_j - y_{ij} \phi \bar{Q}_i d_j - y_{ij} \phi \bar{L}_i e_j + hc, \tag{2.34}$$

where  $y$  are the different Yukawa couplings matrices,  $i, j$  are the family indices from 1 to 3, and  $\tilde{\phi} = i\tau_2 \phi^*$  where  $\tau_2$  is a Pauli matrix. After SSB and shifting the scalar field by  $\nu$ , the masses for the fermions are:

$$m_i = \frac{1}{\sqrt{2}} y_i \nu, \quad i = e, u, d. \tag{2.35}$$

### 2.1.10 The parameters of the Standard Model

There are 18 parameters of the Standard Model that must be determined experimentally:

- $m_u, m_d, m_s, m_c, m_b, m_t, m_e, m_\mu, m_\tau$ : the quark and lepton masses, as stated in Table 2.1,
- $\lambda, A, \rho, \eta$ : the parameters that describe the mixing between quark mass and flavor eigenstates, as in the CKM matrix in Ref. [14],
- $m_H$  and  $\nu$ : the mass and the vacuum expectation of the Higgs field, determined to be  $m_H = 125 \text{ GeV}$  and  $\nu = 246 \text{ GeV}$ , and

- $\alpha, \alpha_s, G_F$ : the coupling constants for the relative strengths of the gauge group.  $\alpha$  is dependent on the energy scale it is measured in, and so two values are given:

$$\begin{aligned}
\alpha &\equiv e^2/4\pi, & \alpha(q^2 \approx 0) &\approx 1/137.0, \\
\alpha(q^2 = m_Z^2 = (91 \text{ GeV})^2) &\approx 1/127.5, \\
\alpha_s &\equiv g_s^2/4\pi, & \alpha_s(q^2 = m_Z^2) &= 0.1184 \pm 0.0007, \\
G_F &\equiv \sqrt{2} g_W^2/8m_W^2, & G_F(q^2 \approx 0) &= 1.1663787 \times 10^{-5} \text{ GeV}^2,
\end{aligned}$$

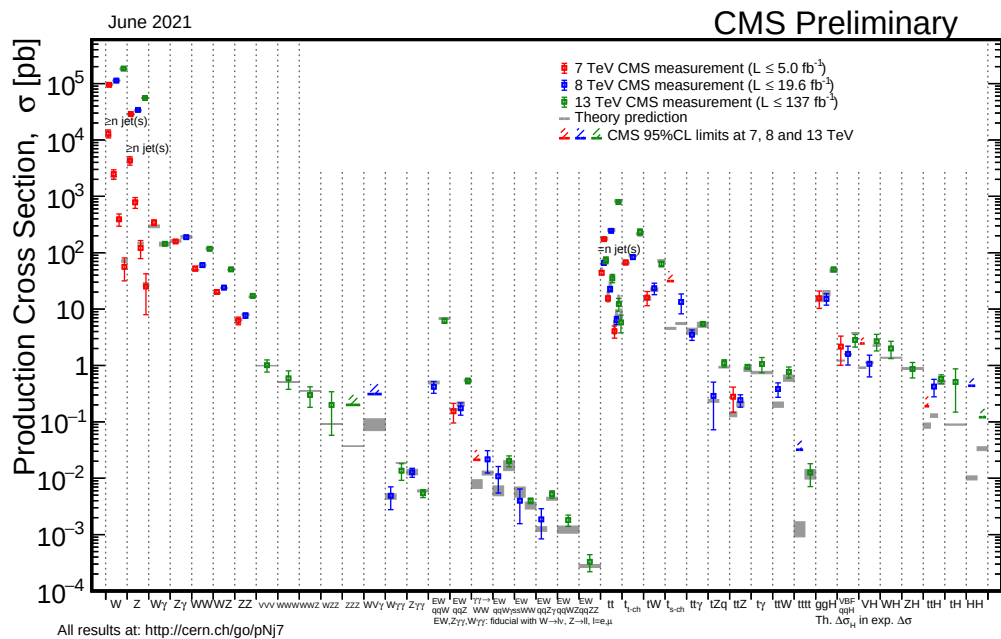
where  $m_W$  and  $m_Z$  are the masses of the W and Z bosons, respectively, and  $g_W$  and  $g_s$  are the weak and strong coupling constants.

## 2.2 Supersymmetry

The Standard Model has been widely successful, predicting the observed production rates of SM processes to high precision (as shown in Figure 2.5) and the existence of elementary particles before their empirical discovery, such as the top quark [17, 18], the tau neutrino [19], and the Higgs boson [4, 5]. Despite all of its success, the SM is an incomplete description of our Universe, the most glaringly obvious deficiency being that it is unable to incorporate gravity. Additionally, the SM does not account for the non-zero masses of neutrinos [20] nor neutrino oscillations [21], which we have experimentally detected and measured; it has no mechanism for the acceleration of our Universe, often attributed to dark energy [22, 23], and it contains no viable candidate for dark matter, which cosmological observations imply make up 85% of all matter [24]. Lastly, the quantum corrections to the Higgs mass are unstable and divergent, requiring a fine-tuning of the bare mass that is considered unnatural by most scientists [25].

These variances from nature, while a problem for the SM, give motivation and direction for new physics. Physics beyond the Standard Model (BSM) refers to theories that attempt to explain these deficiencies, and we consider the theory of supersymmetry (SUSY) in this thesis. A more complete description of SUSY is given in Refs. [27–29].





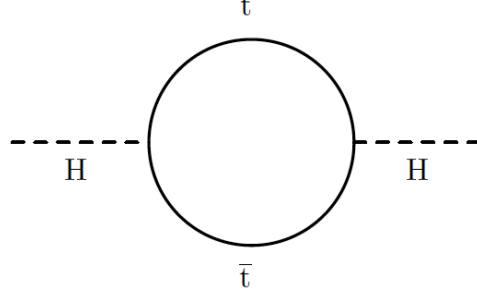


Figure 2.6: The loop correction to the Higgs mass from the top quark.

SUSY is a spacetime symmetry that uses a fermionic operator  $Q$  to relate bosons to fermions:

$$Q|Boson\rangle = |Fermion\rangle \quad \text{and} \quad Q|Fermion\rangle = |Boson\rangle. \quad (2.36)$$

Every SM particle has a superpartner related by Equation 2.36 that differs by  $\frac{1}{2}$ -unit of spin but is otherwise identical in properties (mass, hypercharge, weak isospin, color charge), assuming the symmetry is unbroken. Before describing SUSY further, we will discuss the fine-tuning problem in more detail, to see why such a symmetry would be beneficial.

### 2.2.1 The fine-tuning problem

Fine-tuning is the process of precisely adjusting parameters of a model to fit with observation and is regarded as problematic without a known mechanism or explanation as to why the parameters must have those exact values [30].

The physical mass of particles can be thought of as a bare mass  $m_0$  plus “loop” corrections  $\Delta m$  [31], where the  $m$  in Equations 2.13, 2.15, 2.21, and 2.26 is the bare mass. For the Higgs mass, the VEV is very unstable under quantum loop corrections due to the presence of virtual particles. These corrections are much larger than the physical mass of the Higgs, which means we must precisely fine-tune the bare mass. Let us consider only the top quark loop correction, the largest contribution, as shown in Figure 2.6. We can write the coupling of the Higgs to the top quark as:

$$\mathcal{L} = -\frac{y_t}{\sqrt{2}}\phi_L^\dagger t_R + hc, \quad (2.37)$$

where  $y_t$  is the top Yukawa coupling and  $\bar{t}_L t_R$  are the left- and right-handed components of the top quark. The mass of the top quark is  $m_t = y_t \nu / \sqrt{2}$ , with  $\nu$  as the VEV.

This coupling is valid from the electroweak scale – where the masses of the W, Z, and Higgs bosons are roughly the same order of magnitude – to some arbitrarily high scale<sup>4</sup>. We will take this high scale as the Planck scale, the highest scale at which quantum field theory holds; for any scale beyond that, gravity becomes too strong to ignore and must be included. Imposing, then, a momentum cutoff at the Planck scale  $\Lambda$ , we evaluate the top-quark loop contribution to the Higgs mass:

$$\Delta m_H^2 \Big|_{\text{top}} = -\frac{3|y_t|^2}{8\pi^2} \left[ \Lambda^2 - 3m_t^2 \ln \left( \frac{\Lambda^2 + m_t^2}{m_t^2} \right) \right], \quad (2.38)$$

where we have dropped terms that are finite as  $\Lambda \rightarrow \infty$  since we are only concerned with the asymptotic behavior of  $\Delta m_H^2$ . Let us divide Equation 2.38 into its quadratic  $\Delta m_H^2 \Big|_{\text{top}}^{\Lambda^2}$  and logarithmic  $\Delta m_H^2 \Big|_{\text{top}}^{\ln}$  terms:

$$\begin{aligned} \Delta m_H^2 \Big|_{\text{top}}^{\Lambda^2} &= -\frac{3|y_t|^2}{8\pi^2} \Lambda^2, \quad \text{and} \\ \Delta m_H^2 \Big|_{\text{top}}^{\ln} &= \frac{9m_t^2 |y_t|^2}{8\pi^2} \ln \left( \frac{\Lambda^2 + m_t^2}{m_t^2} \right). \end{aligned} \quad (2.39)$$

Since the corrections are quadratic in the Planck scale, they are roughly 30 orders of magnitude larger than the Higgs mass,  $m_h^2 \approx (100 \text{ GeV})^2$ . Specifically, this requires the Higgs bare mass to be finely-tuned in order to almost completely cancel out these quantum corrections: this is the fine-tuning problem.

Many new theories have been created in order to avoid this level of fine-tuning. One such theory is supersymmetry, which contains the same number of bosonic and fermionic degrees of freedom. Suppose we introduce a scalar partner of the top quark, denoted  $\tilde{t}$ , whose interactions

---

<sup>4</sup> Since the SM is incomplete, there must be new physics somewhere between the electroweak scale and the scale where gravity can no longer be ignored. This new physics would couple to the Higgs and so affect the Higgs mass. A cutoff is necessary and imposed at the point where these new particles would appear. Since we don't know where that cutoff should be, we will assume it is the highest scale at which QFT holds, the Planck scale.

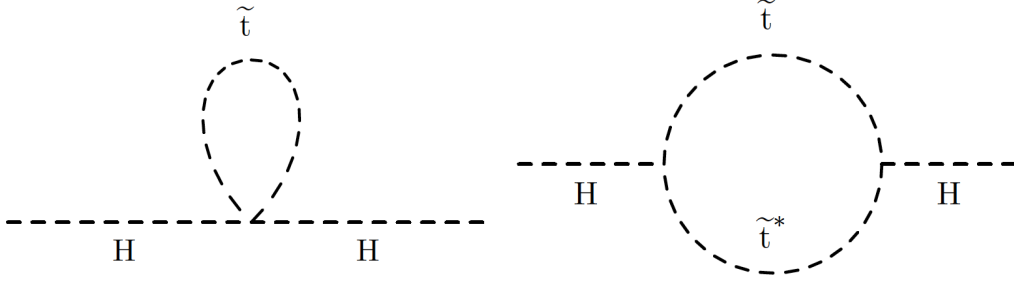


Figure 2.7: The loop correction to the Higgs mass from the top squark.

are described by:

$$\mathcal{L}_{\tilde{t}} = -\frac{\tilde{\lambda}(\phi)^2}{2}(|\tilde{t}_L|^2 + |\tilde{t}_R|^2) - \phi(\tilde{\mu}_L|\tilde{t}_L|^2 + \tilde{\mu}_R|\tilde{t}_R|^2) - m_{\tilde{t}_L}|\tilde{t}_L|^2 - m_{\tilde{t}_R}|\tilde{t}_R|^2. \quad (2.40)$$

Let us calculate the quadratic contribution to the Higgs mass, from the leading-order loops in Figure 2.7, and assume that  $\tilde{\lambda} = |y_t|^2$ :

$$\Delta m_{\text{H}}^2 \Big|_{\tilde{t}}^{\Lambda^2} = \frac{3|y_t|^2}{8\pi^2} \Lambda^2. \quad (2.41)$$

We see that Equation 2.41 exactly cancels out the quadratic contribution from the top quark in Equation 2.39. For the logarithmic contribution, we will assume  $\tilde{\mu}_L = \tilde{\mu}_R = 2\tilde{\lambda}m_t^2$ , and give a small perturbation to the left- and right-handed components of the  $\tilde{t}$  mass:  $m_{\tilde{t}_L}^2 = m_t^2 + m_{\Delta_L}^2$  and  $m_{\tilde{t}_R}^2 = m_t^2 + m_{\Delta_R}^2$ . The logarithmic contribution from the  $\tilde{t}$  corrections is then:

$$\begin{aligned} \Delta m_{\text{H}}^2 \Big|_{\tilde{t}}^{\ln} = & -\frac{9|y_t|^2}{16\pi^2} \left[ (m_t^2 + m_{\Delta_R}^2) \ln \left( \frac{\Lambda^2 + m_t^2 + m_{\Delta_R}^2}{m_t^2 + m_{\Delta_R}^2} \right) \right. \\ & \left. + (m_t^2 + m_{\Delta_L}^2) \ln \left( \frac{\Lambda^2 + m_t^2 + m_{\Delta_L}^2}{m_t^2 + m_{\Delta_L}^2} \right) \right]. \end{aligned} \quad (2.42)$$

As  $m_{\Delta_R}^2 \rightarrow 0$  and  $m_{\Delta_L}^2 \rightarrow 0$ , we see that  $\Delta m_{\text{H}}^2 \Big|_{\tilde{t}}^{\ln} \rightarrow \Delta m_{\text{H}}^2 \Big|_{\text{top}}^{\ln}$ . By introducing a scalar particle ( $\tilde{t}$ ) with identical couplings as the top quark ( $t$ ), provided the mass difference is small, the top quark loop corrections are cancelled out. Under supersymmetry,  $\tilde{t}$  is the superpartner of the top quark, and there is no longer any need for excessive fine-tuning of the Higgs mass [32].

### 2.2.2 The Minimal Supersymmetric Standard Model

The Minimal Supersymmetric Standard Model (MSSM) [28] is the most basic extension to the SM that incorporates supersymmetry and is based on the same gauge symmetry as the SM,  $SU(3) \times SU(2) \times U(1)$ . In the MSSM, SM fermions gain spin-0 superpartners that have an “s” in front of their fermion’s name for “scalar” (for example, the superpartner to the top quark is called a top squark or a stop) and denoted with a tilde ( $\tilde{t}$ ). Quarks and leptons and their corresponding superpartners are arranged in chiral supermultiplets, sometimes called matter supermultiplets, that consist of a two-component Weyl fermion and a complex scalar field. Before electroweak symmetry breaking, the gauge bosons ( $W^\pm$ ,  $W^0$ , and  $B^0$ ) receive spin- $\frac{1}{2}$  superpartners ( $\widetilde{W}^\pm$ ,  $\widetilde{W}^0$ , and  $\widetilde{B}^0$ ) called winos and binos. Gluons are partnered with gluinos,  $\tilde{g}$ . The SM gauge bosons and their fermionic superpartners are arranged in vector supermultiplets, sometimes called gauge supermultiplets. Both vector and chiral supermultiplets contain the same number of fermionic and bosonic degrees of freedom.

A single higgsino (the fermionic superpartner of the Higgs boson) would lead to a gauge anomaly<sup>5</sup> and would cause the theory to be inconsistent [28]. If two higgsinos are added, their contributions would exactly cancel out the gauge anomaly. Additionally, only a  $Y = \frac{1}{2}$  Higgs chiral supermultiplet has the necessary Yukawa couplings to give up-type quarks their masses (and  $Y = -\frac{1}{2}$  for down-type quarks and charged leptons). The MSSM requires a spin-0 Higgs isospin doublet, with opposite hypercharge, in addition to the SM Higgs doublet. Each of these spin-0 doublets has its own chiral superpartner. The scalar components for these doublet fields are

$$H_d = \begin{pmatrix} H_d^0 \\ H_d^- \end{pmatrix} \quad H_u = \begin{pmatrix} H_u^+ \\ H_u^0 \end{pmatrix}, \quad (2.43)$$

which give masses to the isospin  $+\frac{1}{2}$  and  $-\frac{1}{2}$  fermion, respectively.

---

<sup>5</sup> One of the conditions for the cancellation of gauge anomalies is  $\text{Tr}[T_3^2 Y] = \text{Tr}[Y^3] = 0$ . In the SM, this gauge anomaly is cancelled just by the existence of the quarks and leptons. A single superpartner to the Higgs must be a weak isospin doublet with weak hypercharge  $Y = \frac{1}{2}$  or  $Y = -\frac{1}{2}$ . Either of these alone would add a non-zero contribution to the trace and ruin the cancellation of the gauge anomaly.

### 2.2.3 SUSY breaking

Since we have not yet observed any supersymmetric particles, we can assume their masses must be different than their SM superpartners [33, 34]. The details of this symmetry breaking are not yet known, but we can express it as:

$$\mathcal{L}_{\text{MSSM}} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}}, \quad (2.44)$$

where  $\mathcal{L}_{\text{SUSY}}$  contains the gauge and Yukawa interactions and  $\mathcal{L}_{\text{soft}}$  contains our ignorance on how SUSY is specifically broken. Similar to electroweak symmetry breaking, the various components of the fields may mix. The W and Z bosons absorb three of the eight degrees of freedom from the two Higgs fields. The remaining five degrees of freedom become mass eigenstates: two scalar Higgs particles  $h^0$  and  $H^0$ , a pseudoscalar Higgs  $A^0$ , and two charged Higgs particles  $H^\pm$ . The neutral superpartners of the electroweak bosons, the wino and bino, and the two neutral higgsinos have the same quantum numbers, and so they mix to form four mass eigenstates. These are called neutralinos and are denoted  $\tilde{\chi}_i^0$ , where  $i = 1$  is the lightest. Similarly, the MSSM has two charginos  $\tilde{\chi}_i^\pm$  from the mixing of the charged winos and charged higgsinos.

The goldstino, a massless neutral fermion, is added as a consequence of SUSY breaking. Goldstone's theorem [35] states that when a symmetry is spontaneously broken, one or more massless particles must be present to couple to the vacuum. Under supersymmetry, the Higgs mechanism, which describes what occurs when a global symmetry is gauged<sup>6</sup>, is unaffected and continues to operate.

In models where SUSY breaking is mediated at a low scale [36, 37], the goldstino is nearly massless relative to the other particles. Accounting for gravity, the superpartner for the elusive and hypothesized graviton is the massless spin- $\frac{3}{2}$  gravitino. As SUSY is spontaneously broken, the gravitino acquires mass by absorbing the goldstino, the gauge field to which it couples. The gauge bosons associated with the broken generators of the symmetry group acquire mass proportional to the gauge couplings and the VEV, and the corresponding would-be Goldstone boson is eliminated.

---

<sup>6</sup> Gauging a global symmetry means modifying the theory so that it is invariant under spatially-varying transformations.

This process is called the super-Higgs mechanism. Both the gravitino and the goldstino are denoted by  $\tilde{G}$ . A list of particles introduced in the MSSM is given in Table 2.3.

#### 2.2.4 R-parity

Imposed in the MSSM is a discrete symmetry called R-parity, which enforces lepton and baryon number conservation [28]. R-parity is defined by

$$R = (-1)^{2s+3B+L}, \quad (2.45)$$

where  $s$  is the quantum spin number and  $B$  and  $L$  are the baryon and lepton numbers. The R-parity quantum numbers are then  $R = +1$  for SM particles and  $R = -1$  for their supersymmetric partners. If we assume R-parity conservation, as is done for the models considered in this thesis, we introduce significant consequences: (i) SUSY particles are always produced in pairs, (ii) there is always an odd number of SUSY particles in their decay products (usually one), and (iii) the lightest SUSY particle (LSP) is stable. While a colored or electrically-charged LSP is not explicitly required, a stable SUSY particle with either of those properties would have already been observed [38]. As a result of empirical exclusions and inconsistencies with our observed Universe [14, 28, 39], this leaves the lightest neutralino as the most favored option for the LSP. This gives an attractive side effect of SUSY: the LSP, a massive and weakly-interacting particle, is a prime dark matter candidate. The LSP would not be directly detected inside particle colliders but would instead leave a momentum imbalance in the event collision (more details in Section 5.3.3). Triggering on events with large momentum imbalance (discussed in Sections 4.7 for Phase 1 and Section 8.3 for Phase 2) gives analyses a good handle for selecting events consistent with the presence of LSPs. Although not discussed in this thesis, SUSY searches that use R-parity violating models are discussed in references [40, 41].

| Name                     | Gauge Eigenstates                                     | Mass Eigenstates                                                      | Spin | $P_R$ |
|--------------------------|-------------------------------------------------------|-----------------------------------------------------------------------|------|-------|
| Higgs boson              | $H_u^0 H_d^0 H_u^+ H_d^-$                             | $h^0 H^0 A^0 H^\pm$                                                   | 0    | +1    |
| Gluino                   | $\tilde{g}$                                           | same                                                                  | 1/2  | -1    |
| Squarks ( $\tilde{q}$ )  | $\tilde{u}_L \tilde{u}_R \tilde{d}_L \tilde{d}_R$     | same                                                                  | 0    | -1    |
|                          | $\tilde{s}_L \tilde{s}_R \tilde{c}_L \tilde{c}_R$     | same                                                                  | 0    | -1    |
|                          | $\tilde{t}_L \tilde{t}_R \tilde{b}_L \tilde{b}_R$     | $\tilde{t}_1 \tilde{t}_2 \tilde{b}_1 \tilde{b}_2$                     | 0    | -1    |
| Sleptons ( $\tilde{l}$ ) | $\tilde{e}_L \tilde{e}_R \tilde{\nu}_e$               | same                                                                  | 0    | -1    |
|                          | $\tilde{\mu}_L \tilde{\mu}_R \tilde{\nu}_\mu$         | same                                                                  | 0    | -1    |
|                          | $\tilde{\tau}_L \tilde{\tau}_R \tilde{\nu}_\tau$      | $\tilde{\tau}_1 \tilde{\tau}_2 \tilde{\nu}_\tau$                      | 0    | -1    |
| Neutralinos              | $\tilde{B}^0 \tilde{W}^0 \tilde{H}_u^0 \tilde{H}_d^0$ | $\tilde{\chi}_1^0 \tilde{\chi}_2^0 \tilde{\chi}_3^0 \tilde{\chi}_4^0$ | 1/2  | -1    |
| Charginos                | $\tilde{W}^\pm \tilde{H}_u^\pm \tilde{H}_d^\pm$       | $\tilde{\chi}_1^\pm \tilde{\chi}_2^\pm$                               | 1/2  | -1    |
| Goldstino                | $\tilde{G}$                                           | same                                                                  | 1/2  | -1    |
| Gravitino                |                                                       |                                                                       | 3/2  |       |

Table 2.3: The particles of the MSSM. The first- and second-generation sfermion gauge eigenstates are assumed not to mix due to experimental constraints on charge-parity symmetry violating phases and flavor changing neutral currents [28].



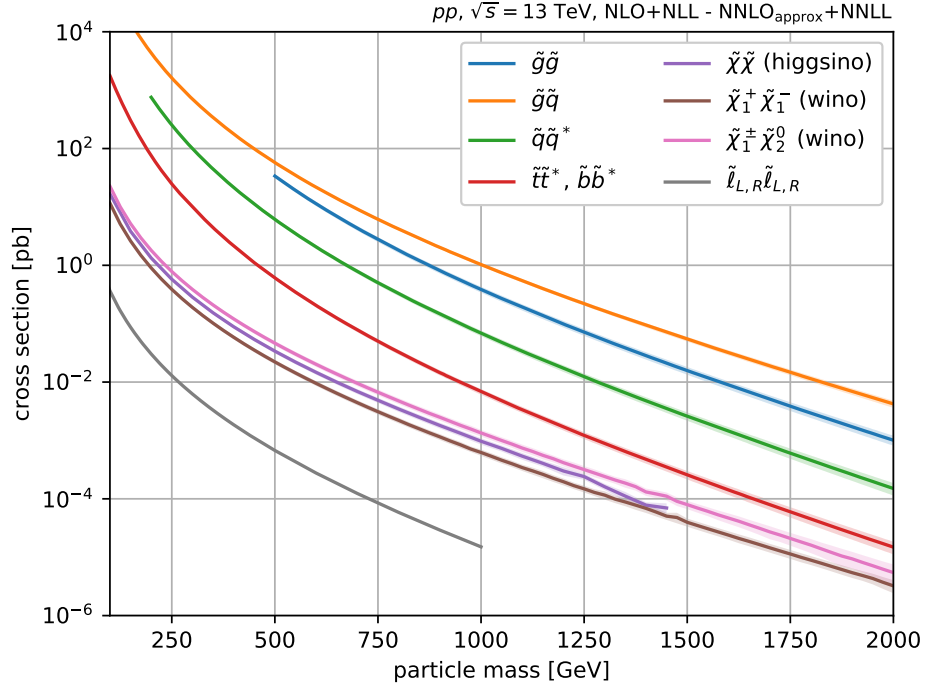


Figure 2.8: Theoretical production cross sections of SUSY processes [42].

### 2.2.5 SUSY searches

Figure 2.8 shows the production cross sections for different SUSY particles, demonstrating why many SUSY searches target gluino and squark production. After the squarks and gluinos decay, their decay chains will each contain one lightest supersymmetric particle (LSP). The lightest supersymmetric particle is stable and weakly-interacting, and therefore it will leave the detector without leaving any direct trace, carrying away some of the momentum from the proton-proton collision. These particles will produce a large missing energy signature, which many SUSY searches require in the final states.

Since gluons have the largest production cross section in proton-proton collisions, gluino pair production becomes an attractive target for SUSY analyses, as they can be produced significantly at the LHC [43–46]. In R-parity conserving SUSY models, gluinos eventually decay into an LSP, typically along with many quarks and the SM gauge bosons or the Higgs. The quarks register as jets in the detector, and the LSP leaves a momentum imbalance. Typical gluino signatures, then,

involve four jets, or the decay products of the SM gauge or Higgs bosons, and missing energy. Figure 2.9 shows some of the simplified gluino pair production models, where the filled-in circle represents all possible production contributions that lead to the resulting gluino pair.

To date, no searches have discovered supersymmetric particles. The size of the expected signals from different SUSY models can be translated into cross section upper limits at 95% confidence level. At the time of writing, the strongest limit set by CMS excludes gluino masses up to roughly 2.3 TeV [47].

Electroweakino production is much more difficult. Neutralinos and charginos only interact with the weak vector bosons and so are not directly produced at hadron colliders in copious amounts. The heavier electroweakinos eventually all decay into the lightest neutralino (the LSP) through the W bosons (for the charginos) or a Z boson (for the neutralinos), where the LSP leaves undetected. Typical electroweakino signatures involve final states with leptons, or SM gauge or Higgs bosons, and missing energy. Leptons are usually preferred, since they provide a cleaner signature in the detector. Until recently, the LHC did not produce enough collision data for a direct search to be possible; as the integrated luminosity increased, these searches became easier, but the limits were rather weak and restricted to lower masses. With the Run 2 (2016-2018) data, however, it is now possible to search for higher-mass electroweakinos, and the LHC has a number of dedicated searches [48–52]. Figure 2.10 shows some of the simplified electroweakino pair production models, where the filled-in circle represents all possible production contributions. The limits for electroweakino mass exclusion are highly dependent on the mass splitting between the electroweakinos. At the time of writing, the strongest limits set by CMS on the  $\tilde{\chi}_2^0$  mass (for a massless  $\tilde{\chi}_1^0$ ) excludes masses up to 1300 GeV, and excludes the mass of the  $\tilde{\chi}_1^0$  up to 800 GeV (for a  $\tilde{\chi}_2^0$  of 1300 GeV) [50].

A synopsis of the mass reach for different final states from searches at CMS is shown in Figure 2.11 for gluino pair production and in Figure 2.12 for electroweakino pair production.

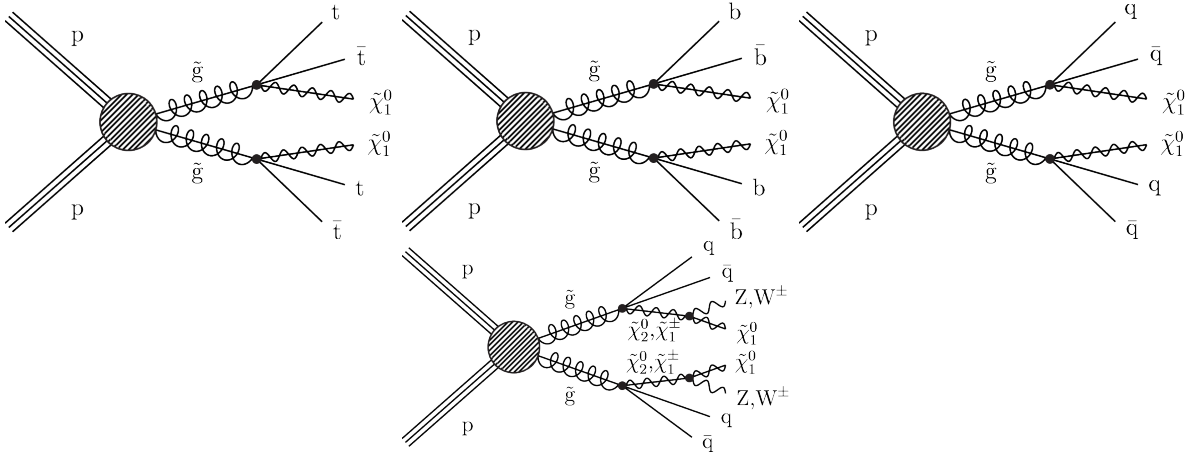


Figure 2.9: Feynman diagrams for gluino pair production with different final states.

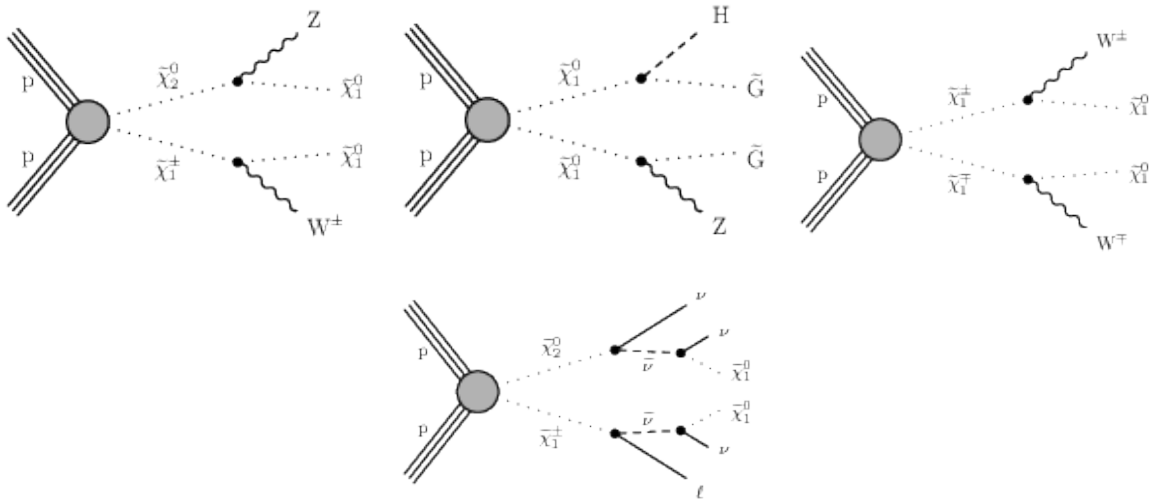


Figure 2.10: Feynman diagrams for electroweakino pair production with different final states.

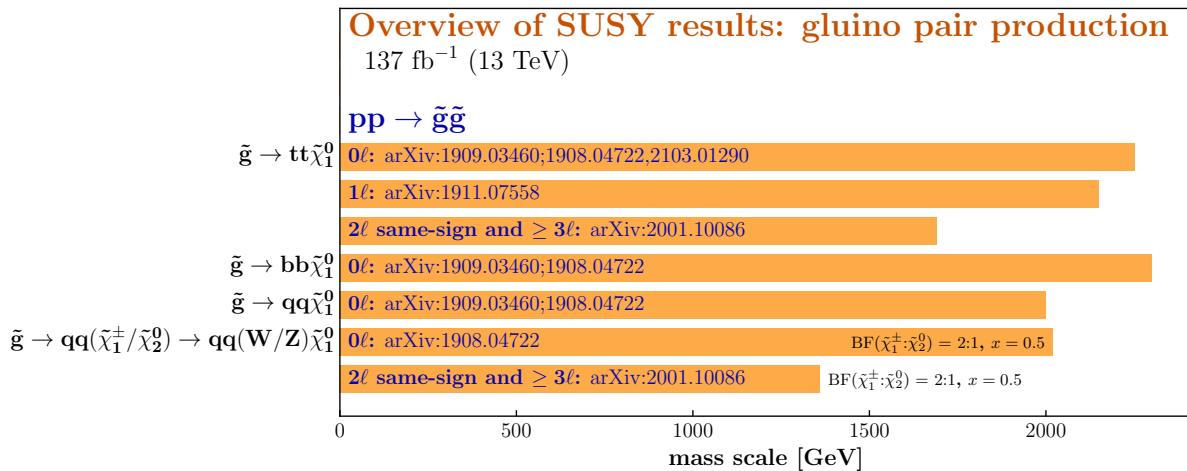


Figure 2.11: Synopsis of the mass reach (limits at 95% CL) obtained for different final states of gluino pair production, using simplified models. The arXiv and physics analysis summary numbers in each line indicate the analyses with relevant results [53].

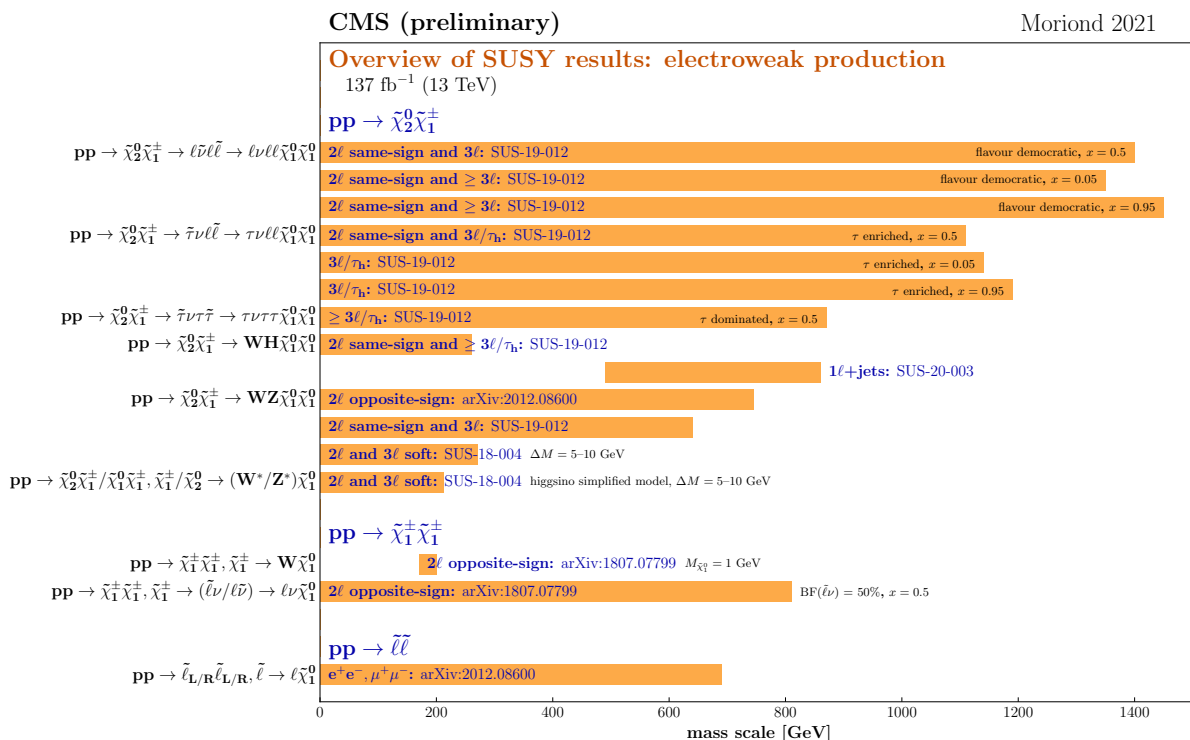


Figure 2.12: Synopsis of the mass reach (limits at 95% CL) obtained for different final states of electroweakino pair production, using simplified models. The arXiv and physics analysis summary numbers in each line indicate the analyses with relevant results [53].

## Chapter 3

### The Large Hadron Collider

The Large Hadron Collider (LHC) is currently the world's largest particle accelerator, housed within a tunnel  $\sim 100$  meters underground on the French-Swiss border, at CERN, Conseil Européen pour la Recherche Nucléaire (European Organization for Nuclear Research in English) [54]. Figure 3.1 shows the different stages of the CERN accelerator complex. Inside the 27 km circumference tunnel, two counter-rotating beam pipes contain protons moving near the speed of light. 1232 superconducting dipole magnets bend the protons along the circular trajectory of the LHC, while quadrupole magnets with alternating focusing and defocusing fields allow the beams to be stabilized and focused. Electric fields from 400 MHz radio frequency (RF) cavities accelerate the protons and separate the beam into bunches, each about 25 ns apart. The protons are successively accelerated until they reach the LHC, the last ring of the CERN accelerator complex.

To begin, a large electric field is used to ionize hydrogen gas, and the resulting protons are fed into a linear accelerator (Linac2), increasing their energy up to 50 MeV. The protons then move through three consecutive synchotrons, each increasing the protons' energy before moving them to the next stage: the Proton Synchrotron Booster (PSB or BOOSTER in Figure 3.1) increases the energy of the protons to 1.5 GeV, the Proton Synchrotron (PS) to 25 GeV, and the Super Proton Synchrotron (SPS) to 450 GeV. The final acceleration, up to 6.5 TeV, occurs in the two beam pipes of the LHC itself. The two beams are narrowed and collide at four intersection points, where the detectors of the experiments are located: CMS, LHCb, ATLAS, and ALICE. The CMS detector (described further in Chapter 4) and the ATLAS detector are considered general-purpose

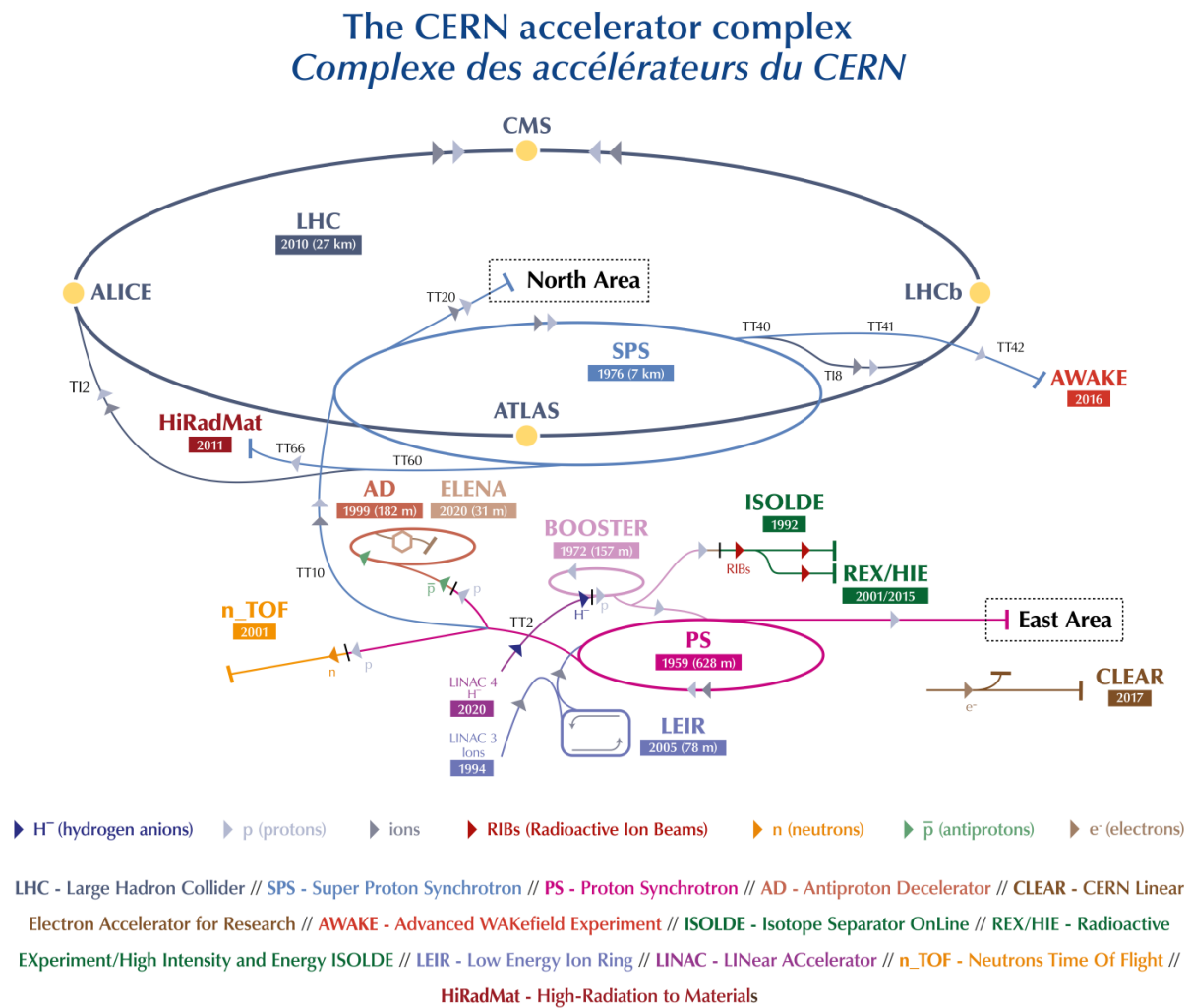


Figure 3.1: The accelerator complex at CERN [55].

detectors, since they more fully surround the collision area than the LHCb and ALICE detectors, allowing for a wide variety of particles and events to be reconstructed from the collected data. The LHCb detector specializes in events that contain b hadrons, while ALICE specializes in heavy-ion collisions. This thesis uses data collected by the CMS detector from 2016-2018, known as “Run 2”.

The expected rate of collisions for a given physical process is dependent on the production cross section ( $\sigma$ ) and the instantaneous luminosity ( $\mathcal{L}$ ):

$$\frac{dN}{dt} = \sigma \cdot \mathcal{L}, \quad (3.1)$$

where  $N$  is the number of expected collisions. Increasing the instantaneous luminosity, then, increases the event rate, giving us access to more collisions and thus more rare decays. In general, the instantaneous luminosity depends on the shape and density of the bunches and how frequently the beams interact, and is defined as:

$$\mathcal{L} = \frac{N_p^2 N_b f_{rev}}{4\pi\epsilon_n\beta^*} F, \quad (3.2)$$

where  $N_p$  is the number of particles per 25 ns bunch,  $N_b$  is the number of bunches per beam,  $f_{rev}$  is revolution frequency determined by the LHC circumference, and  $F$  is a reduction factor that accounts for the beam-crossing angle. The LHC injector limits  $\epsilon_n$ , the normalized transverse beam emittance, while the mechanical aperture limits  $\beta^*$ , the beta function at the interaction point. The number of collisions per bunch crossing varies with the instantaneous luminosity and averaged about 29 during Run 2.

Each collision creates a vertex in the detector, where particle trajectories estimate the interaction point of the two protons. The primary vertex is the vertex with the largest momentum, while secondary vertices are known as pileup vertices. Figure 3.2, which represents an event with 78 reconstructed vertices from the 2012 data-taking period, showcases how difficult it is to reconstruct tracks and vertices. Increasing the instantaneous luminosity increases the number of tracks and vertices and is one of the main challenges of the High-Luminosity LHC (HL-LHC), described in detail in Chapter 8. The LHC has more than doubled its design luminosity, with a recorded luminosity of over  $2 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$  in 2018.

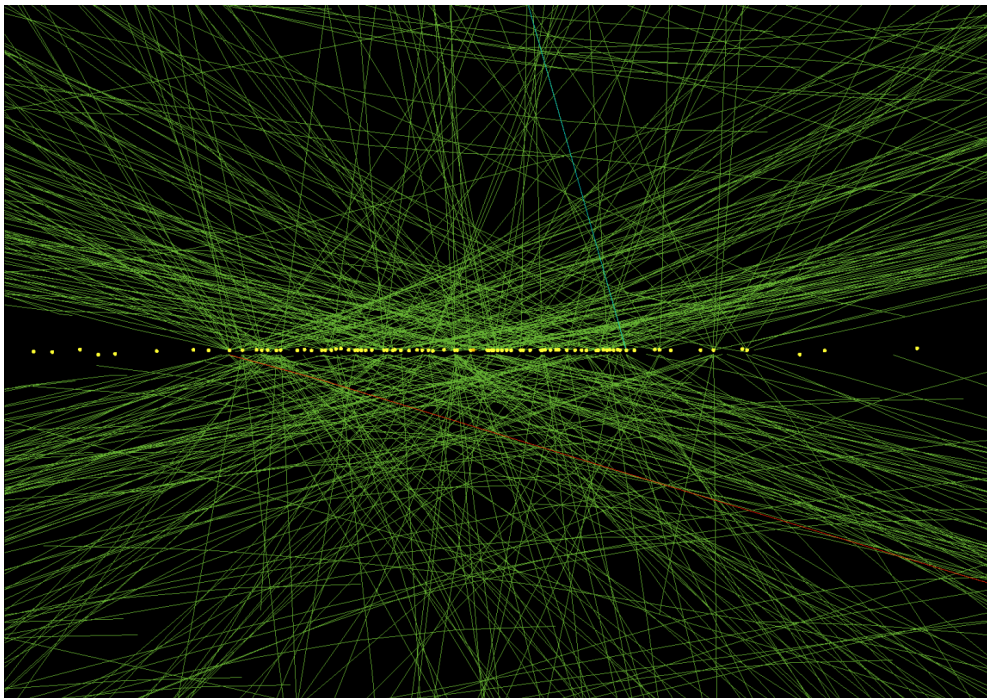


Figure 3.2: A simulation of the track and vertex reconstruction for one event with 78 vertices, collected by the CMS detector in 2012 [56].



## Chapter 4

### The CMS Detector

The Compact Muon Solenoid<sup>1</sup> (CMS) detector is a general purpose detector at the LHC and is composed of a variety of subdetectors designed to absorb and measure the out-going particles of a collision.

As shown in Figure 4.1, these subdetectors form successive, concentric layers centered around the interaction point and structured in a very specific way. The silicon tracker surrounds the beamline and allows for the reconstruction of charged particles. As the innermost layer, it enables precise vertex (the interaction point) measurements, and the choice of silicon allows particles to pass through without being destroyed. The electromagnetic and hadronic calorimeters (ECAL and HCAL), situated between the tracker and the solenoidal magnet, are separate subsystems and measure the energy of electrons and photons, and charged and neutral hadrons, respectively. The calorimeters measure the energy of incoming particles by forcing the particles to interact with certain materials and to deposit their energy in the form of a cascade (or shower) of particles. The solenoid is then placed after the calorimeters to absorb any remaining particles, except muons and neutrinos. Finally, the outermost layer is the extensive muon system.

Each subdetector records some amount of information per collision, briefly processes it into primitive objects, and passes on the relevant information to the trigger system (discussed in Section 4.7), which decides whether a particular event is saved. After the event is saved, it is fully

---

<sup>1</sup> The CMS detector is named appropriately, as it is designed to fit an incredible amount of electronics and hardware inside the 6-m-diameter bore of the solenoidal magnet, with the remaining 80% by volume of the detector designated for the muon systems.

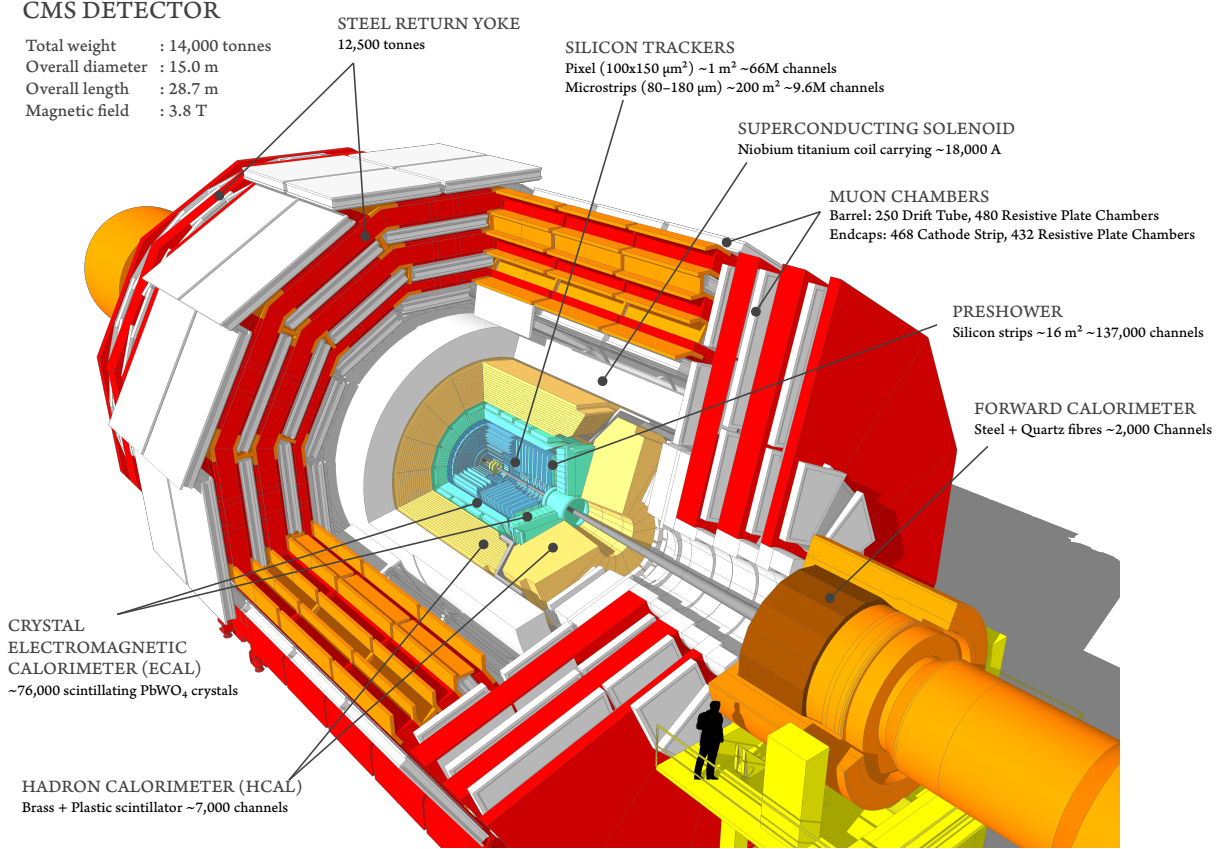


Figure 4.1: A cutaway diagram of the CMS detector, with the subsystems labeled [57].

reconstructed offline into particles and objects, such as jets and energy sums, which are used for analyses (see Chapter 5 for a full description of the event reconstruction).

#### 4.1 Relevant kinematic variables

The origin of the CMS coordinate system is the center of the detector. The  $x$ -axis points radially inward, toward the center of the LHC, while the  $y$ -axis points vertically upward. The  $z$ -axis is parallel to the beamline, following the direction of the counter-clockwise rotating beam (when viewed from above). Since the incoming protons are aligned with the  $z$ -axis, the total momentum in the transverse plane is expected to be zero due to momentum conservation. As a result, the transverse momentum  $p_T = \sqrt{p_x^2 + p_y^2}$  of the particles is often used in analyses. Momentum and energy sums of all particles per collision event are also typically in the transverse plane, such as

the missing transverse energy  $p_T^{\text{miss}}$ .

Due to the cylindrical shape of CMS, a cylindrical coordinate system is also used. The radial coordinate  $r$  is the distance from the beamline in the transverse ( $x$ - $y$ ) plane. The azimuthal angle  $\phi$  in the transverse plane and the polar angle  $\theta$ , measured from the  $z$ -axis, are used to describe a particle's direction with respect to the interaction point. More commonly the pseudorapidity  $\eta = -\ln[\tan(\theta/2)]$ , is used, where a vector perpendicular to the beamline corresponds to  $\eta = 0$ . CMS is generally divided into two regions: the central barrel ( $|\eta| < 1.0$ ) and the forward endcaps ( $|\eta| > 1.8$ ). The transition region ( $1.0 < |\eta| < 1.8$ ) between the barrel and the endcaps is more of an overlap region than a physical region of the detector. The tracking detectors cover the range  $|\eta| < 2.5$ , while the ECAL and HCAL cover  $|\eta| < 3$ . Forward calorimeters extend the coverage to  $|\eta| < 5.2$ . Muons are measured within  $|\eta| < 2.4$ . A transverse slice of the CMS detector is shown in Figure 4.2.

The solenoidal magnet, described in more detail in Section 4.5, surrounds the tracker and calorimeter volume and provides a magnetic field of 3.8 T. The magnetic field is crucial as it bends charged-particle trajectories, allowing for measurement of the particles' momenta.

## 4.2 Silicon tracker

The innermost subdetector of CMS is the silicon tracker. The tracker uses ionization deposits (“hits”) in the silicon layers to reconstruct the trajectories of all charged particles (kaons, pions, electrons, muons) called tracks. The tracks are used to identify the primary vertex of a collision event, crucial for mitigating the effects from pileup, and are associated with signatures in other parts of the CMS detector to make full physics objects (more in Section 5.2). The 3.8 T magnetic field bends the trajectories of charged particles, allowing for the charge and momentum of individual particles to be easily identified. The tracker consists of layers of modules, silicon sensors bonded to readout chips. The thin layers of silicon are set up as reverse-biased p-n junctions and are ionized when charged particles pass through the tracker, creating electron-hole pairs. Electric fields accelerate the ionized charge through the sensors to the readout chips, which digitize and store

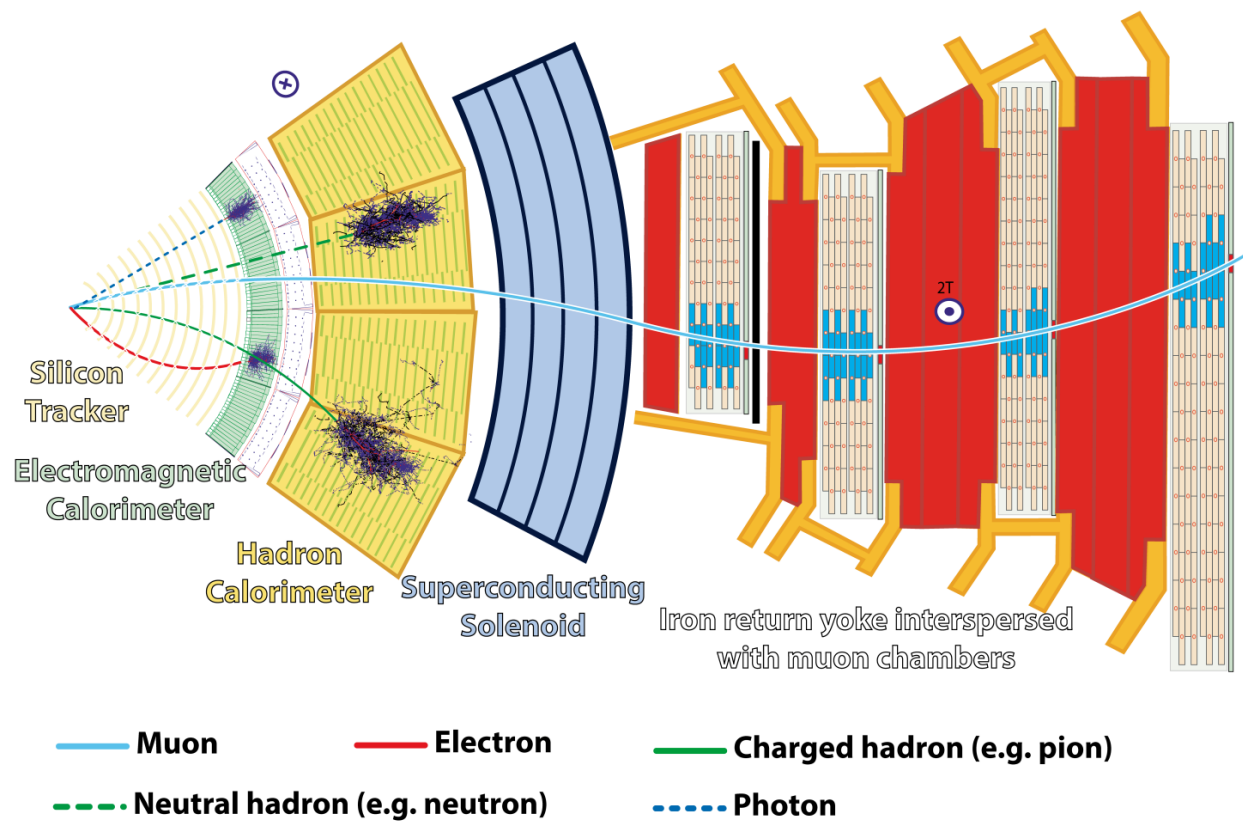


Figure 4.2: A transverse slice of the CMS detector, with the beam axis perpendicular to the page. Trajectories of SM particles are shown as they pass through each subsystem [58].

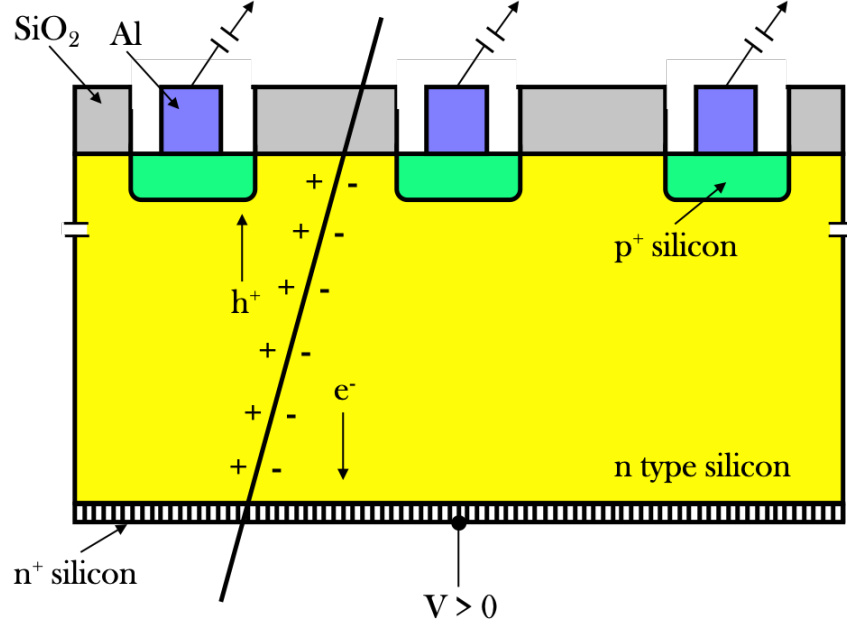


Figure 4.3: A schematic of the ionization of the silicon layers.

the hit information. Figure 4.3 shows a schematic of the ionization of the silicon. Due to the relative closeness to the beamline, the tracker requires differing levels of granularity and radiation hardness, and is thus divided into two components. The pixel detector, discussed in more detail in Section 4.2.1, has finer spatial segmentation and is responsible for the most precise vertexing. The strip detector (Section 4.2.2) lies concentrically outward from the pixel detector and is composed of micro-strip sensors, where most of the hits lie. We include a diagram of the geometry and layers of the tracker in Figure 4.4.

The momentum resolution in the barrel region ( $|\eta| < 1.6$ ) is  $\lesssim 2\%$  or less for a charged particle with  $p_T = 100 \text{ GeV}$ . The excellent track  $z_0$  resolution in the pixel detector (Figure 4.5, left), where the track  $z_0$  is the estimated vertex position of the charged particle along the beam line and the resolution is measured with respect to simulated truth tracks, gives rise to a primary vertex resolution of down to  $15 \mu\text{m}$  (Figure 4.5, right). A resolution such as this allows the position of the pileup vertices to be identified and tracks associated with those vertices to be cleanly removed. Tracks from pileup vertices can overwhelm energy sums, such as missing energy, and make identifying potential sources of real momentum imbalance difficult. Further details on

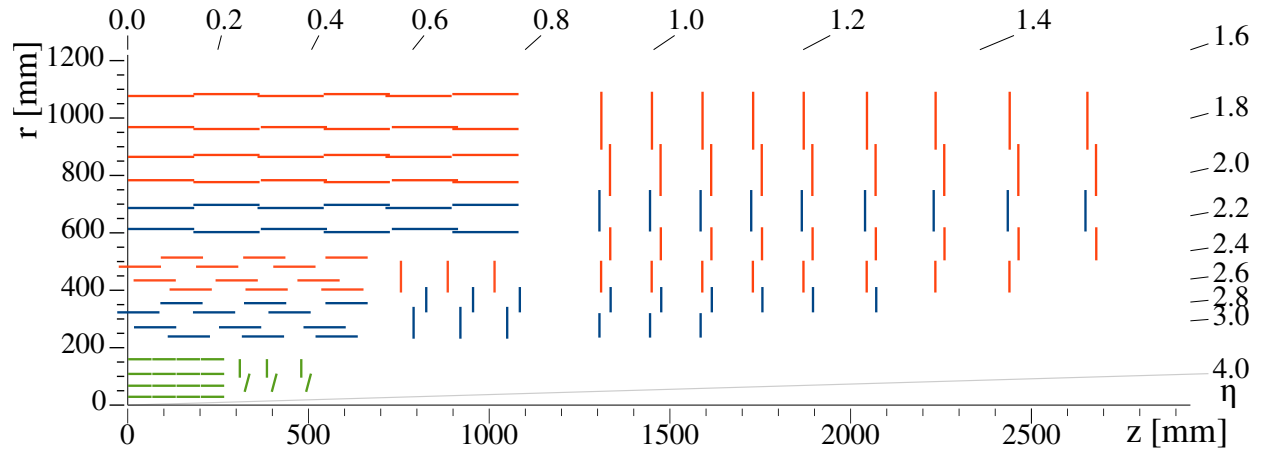


Figure 4.4: A quadrant of the CMS tracker detector, the innermost subdetector. The pixel detector is the green lines closest to the beamline. The red and blue lines are the strip detector, with single-sided and double-sided strip modules, respectively [59].

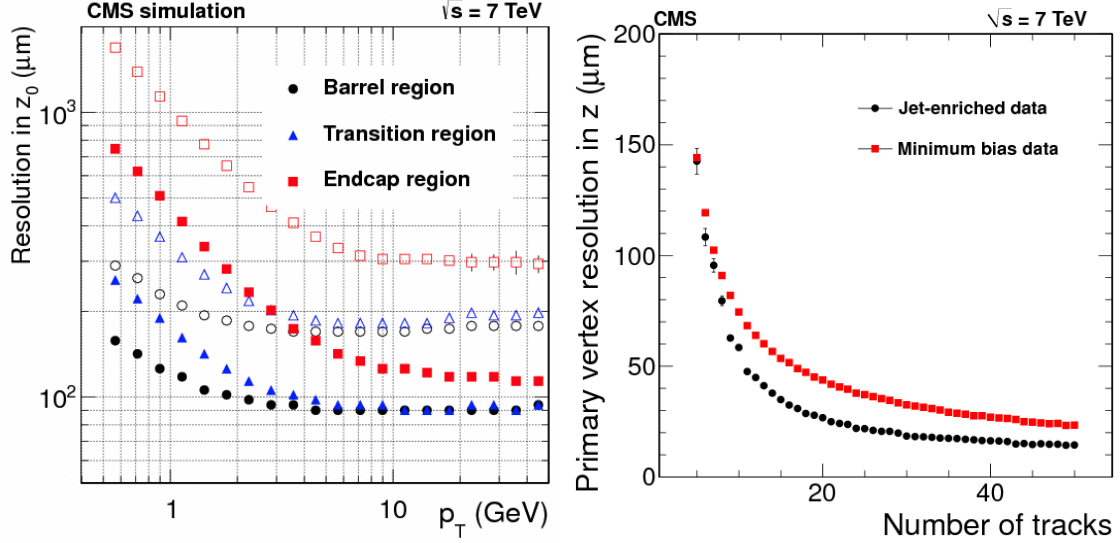


Figure 4.5: Left: Track resolution in  $z$  as a function of  $p_T$  for pixel tracks in simulated  $t\bar{t}$  events in the barrel, transition, and endcap regions. For each bin in  $p_T$ , the solid (open) symbol corresponds to the half-width of the 68% (90%) interval centered on the most probable value of the residuals distributions.

Right: Primary vertex resolution in  $z$  as a function of the number of tracks at the fitted vertex, for two kinds of common events, using simulated  $t\bar{t}$  collisions. [60]

the track and primary vertices reconstruction, including track-finding performance, are given in Section 5.1.1.

#### 4.2.1 Pixel detector

The pixel detector, the innermost subdetector of CMS, consists of 1,440 silicon modular detector units (“modules”), with up to  $100 \times 150$  μm pixels per module. These modules provide the needed spatial granularity for precise track measurements. The barrel consists of three concentric cylinders (“layers”), with radii of 4.4, 7.3, and 10.2 cm from the beamline, while the endcaps consist of two discs each, at  $z = \pm 35.5$  cm and  $z = \pm 48.5$  cm, with radii from 6 to 15 cm. In total, there are 65,894,400 pixels, covering a readout area of  $|\eta| < 2.5$ .

In 2016, the pixel detector was replaced to account for the increasing luminosity of the LHC. As the luminosity is increased, more proton-proton interactions occur each bunch crossing, resulting in more particles and more hits inside the pixel detector. The new detector includes four layers in

the barrel and three discs in each endcap (compared to the original three layers and two discs); the readout electronics were also upgraded to accommodate the increased hit rate.

#### 4.2.2 Strip detector

The strip detector surrounds the pixel detector and is made up of 9.3 million micro-strip sensors. The modules are oriented parallel to the beamline in roughly 10 cm long strips, with the strip pitch varying from 80 to 180  $\mu\text{m}$ . The barrel is made up of ten layers, from  $r = 20$  cm to  $r = 116$  cm; the endcaps on either side of the barrel each have twelve discs, between  $z = \pm 58$  cm and  $z = \pm 282$  cm. To provide a  $z$  measurement in the barrel and an  $r$  measurement in the endcaps, some layers have two modules that are slightly rotated (100 mrad) relative to each other, the blue layers in Figure 4.4. Like the pixel detector, the strip detector covers an area of  $|\eta| < 2.5$ .

### 4.3 Electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL) is a hermetic, homogeneous detector, designed to absorb and measure the energy from photons and electrons. When high-energy electrons enter the ECAL, they lose energy predominantly through photon radiation (bremsstrahlung) as they interact with the lead tungstate crystals that make up the ECAL. Similarly, photons create  $e^+e^-$  pairs via material interactions, losing energy. Both electrons and photons create an electromagnetic shower, generating light in the scintillators of the detector. A diagram of various showers in CMS is shown in Figure 4.6. The collected light is used to measure the energy of the particle.

The ECAL is composed of 75,848  $\text{PbWO}_4$  crystals, ideal for their high density (8.28  $\text{g}/\text{cm}^3$ ) and short radiation length (0.89 cm), which leads to small transverse radii for the induced electromagnetic showers and therefore a compact calorimeter. The ECAL consists of a barrel (EB) and two endcap (EE) regions. Additionally, a preshower detector is installed in front of each endcap. Over 99% of the electromagnetic energy within the shower is captured by the ECAL. On-detector electronics calculate the total transverse energy deposited and record the electromagnetic shower



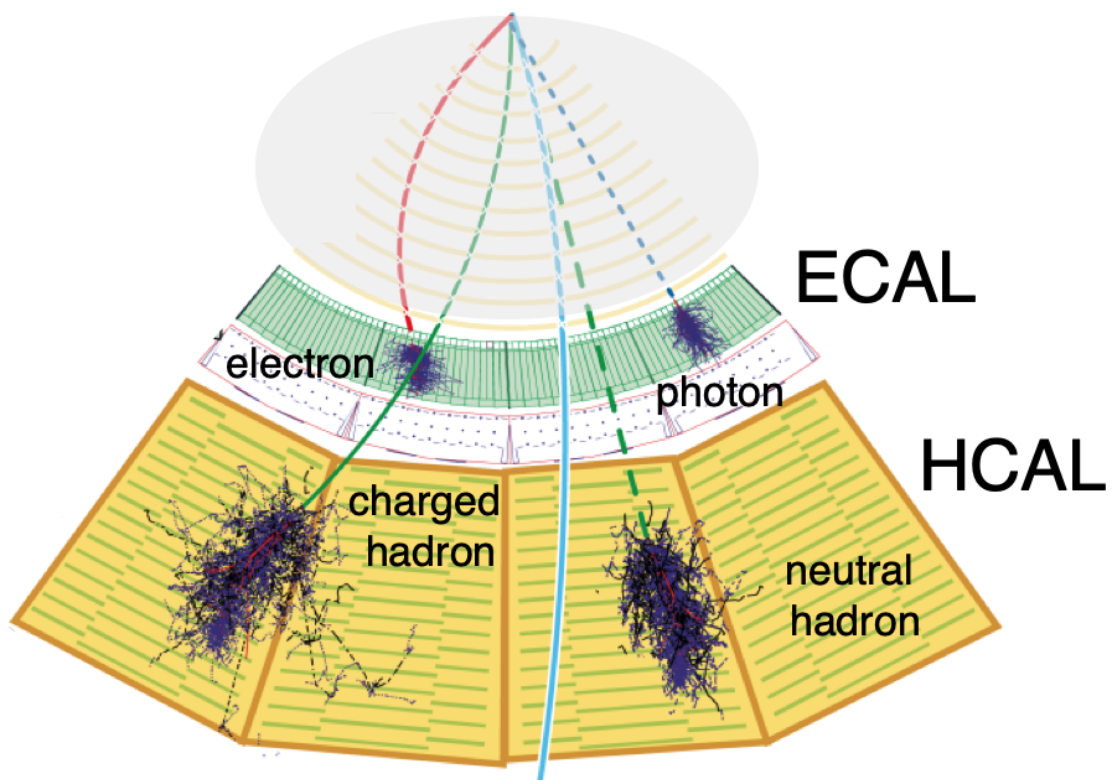


Figure 4.6: A diagram showing the different calorimeter shower types inside CMS. This diagram is modified from Ref. [58].

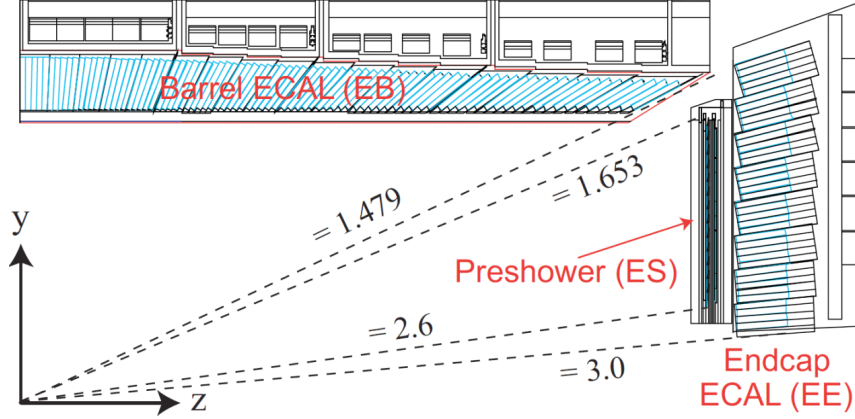


Figure 4.7: A schematic of the CMS electromagnetic calorimeter, emphasizing the crystal orientation. The small gap in  $\eta$  between the EB and the preshower detector is highlighted [61].

shape profile and send both to the L1 trigger. A schematic for the ECAL is shown in Figure 4.7.

The EB covers a range of  $|\eta| < 1.479$ , and the EE covers  $1.479 < |\eta| < 3.0$ . In the barrel, the crystals measure  $2.2 \times 2.2 \times 23 \text{ cm}^3$  and are oriented radially outward from the interaction point. Each crystal is aligned with a constant value of  $\eta$  to minimize energy loss between the crystals. In the endcaps, the crystals measure  $3 \times 3 \times 22 \text{ cm}^3$ . Photodiodes and phototriodes collect the light from each crystal in the EB and EE, respectively, which is further amplified and digitized.

The preshower detector in front of each endcap (shown in Figure 4.8) is a two-layer sampling calorimeter, consisting of a lead absorber, silicon-strip sensors, a second lead absorber, and an orthogonal plane of strip sensors. The 2 mm-wide strip sensors provide finer granularity, when compared to the 3 cm-wide crystals in the EB and EE, and the orthogonal planes of sensors allow for precise shower shape measurements. Most notably, the preshower detector prevents the misidentification of photons originating from neutral pion decays. Such decays often produce two closely-spaced low-energy photons that the ECAL cannot resolve. The energy deposits within the preshower detector, in tandem with those in the rest of the ECAL and an extrapolated path between the two, are used to deduce if seemingly high-energy photons are, indeed, high-energy photons or photon pairs.

The high ECAL granularity, which allows for a mass resolution of around 1%, enabled the identification of the Higgs boson in 2012 in the  $H \rightarrow \gamma\gamma$  channel [4, 5]. The mass resolution and

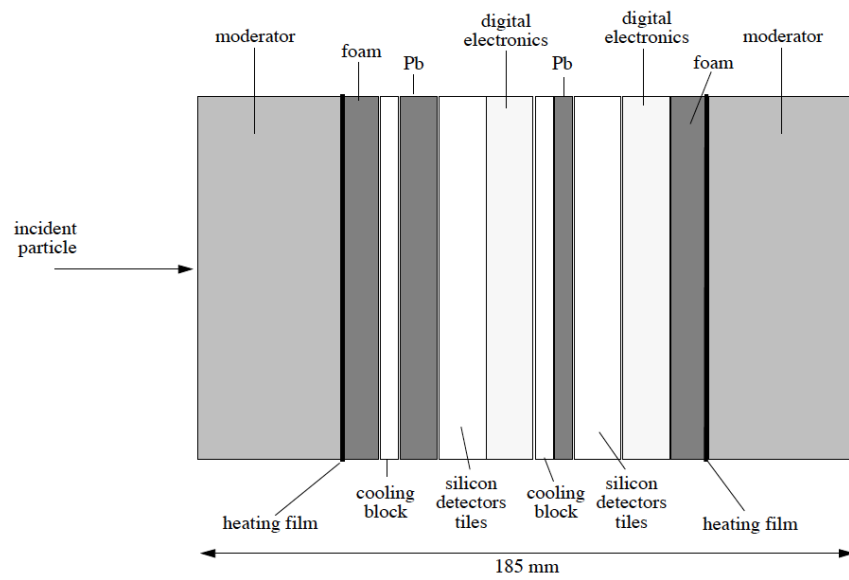


Figure 4.8: Schematic of the layers in the preshower detector [62].

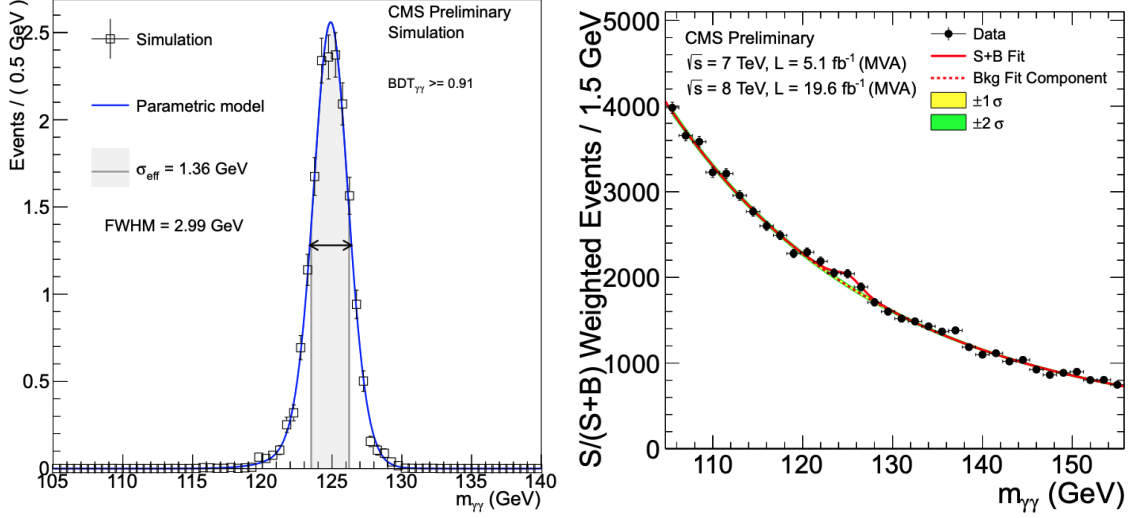


Figure 4.9: The invariant mass distribution for a Higgs boson of  $m_H = 125$  GeV decaying into two photons (left) shows the mass resolution of around 1%, made possible by the granularity of the ECAL. On the right, the background and signal+background parametric models fitted to the data, with a clear bump at 125 GeV. [63]

resulting  $S/(S+B)$  with a clear bump at the Higgs mass of 125 GeV can be seen in Figure 4.9.

Details on the electron- and photon identification efficiency are given in Section 5.2.

#### 4.4 Hadronic calorimeter

The hadronic calorimeter (HCAL) is composed of layers of brass and scintillator plates and is used to measure the position, energy, and arrival time of hadrons. When a particle (or secondary particle) passes through the plastic scintillation layers, it produces a rapid light pulse. A particle's energy is measured by summing over the amount of light collected in a given region of layers, called an HCAL tower, which roughly follows the particle's path through the HCAL. The pulses in each tower are summed together and then converted into electronic signals by hybrid photodiodes (HPDs). The transverse energy of each HCAL tower is computed and sent to the L1 trigger.

The HCAL is divided into four regions: the barrel (HB) from  $|\eta| < 1.4$ , the endcaps (HE) from  $1.3 < |\eta| < 3$ , the forward calorimeters (HF) from  $3 < |\eta| < 5.2$ , and the outer barrel (HO) from  $|\eta| < 1.2$ . There are 36 barrel wedges and 36 endcap wedges. These regions are more clearly shown in Figure 4.10.



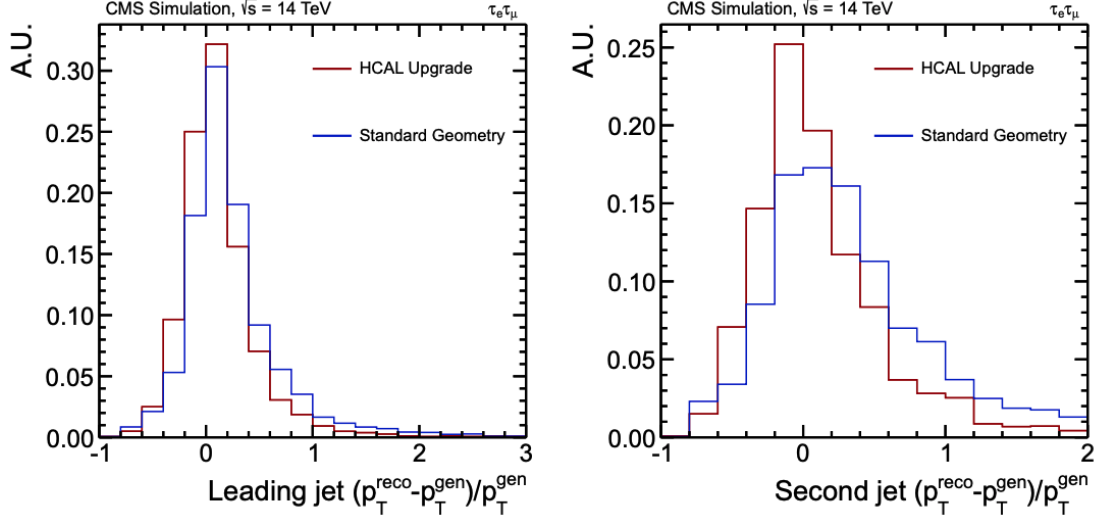


Figure 4.11: Jet  $p_T$  resolution for the  $p_T$ -leading (left) and  $p_T$ -subleading (right) jets in an event. The resolution (full width at half-maximum) for the upgrade and standard geometry are 0.40 and 0.47 respectively for the leading jet and 0.47 and 0.73 respectively for the sub-leading jet. [65].

After the Phase 1 upgrades, the HCAL granularity significantly improved the signal-to-background ratio of energy measurements. In particular, the transverse momentum resolution of particle-flow jets (discussed in Section 5.3.1) improved due to better separation of hadronic clusters, as shown in Figure 4.11, which in turn improved the missing energy resolution.

#### 4.4.1 Hybrid photodiodes

Hybrid photodiodes (HPDs) are photosensors designed to operate within the detector's high magnetic field. Light pulses from the different HCAL regions are sent by fiber-optic waveguides to the HPDs, which amplify the response approximately 2000 times. The surface of an HPD converts light into electrons via the photoelectric effect and accelerates the electrons across a few-mm gap to the surface of a silicon diode. In total, there are 420 HPDs in CMS.

#### 4.4.2 Barrel and endcap calorimeters

The HB and HE each consist of 17 alternating layers of 5 cm-thick brass or steel<sup>2</sup> and 1 cm-thick scintillating plastic. Incoming particles interact within the brass absorber, inducing hadronic

<sup>2</sup> The absorber layers nearest and farthest from the interaction point are made from steel, the rest from brass.

showers. Flashes of blue-violet light are produced by the scintillating molecules embedded in the plastic and are absorbed by very small (less than 1 mm) wavelength-shifting fibers, which shift the light to green, since the photodetectors have low sensitivity in the ultraviolet. The green light is then carried by optic cables to readout boxes located within the HCAL.

#### 4.4.3 Outer calorimeter

The HO is located outside the magnet and ensures no undetected hadronic energy extends beyond the HB. It consists of two scintillating layers that straddle the innermost layer of the yoke and uses the magnet itself as an absorber. The HO is divided along roughly the same lines as the HB and uses HPDs to collect the light.

#### 4.4.4 Forward calorimeter

The HF is installed outside of the CMS endcaps, at  $|z| = \pm 11.2$  m from the interaction point. Since the HF is situated at such high values of  $\eta$ , it is exposed to a high rate of energetic particles, inspiring a more radiation-hard approach than the other regions of the HCAL. 165 cm-thick steel plates are used as absorbers and quartz fibers as the active material to collect light. Photons and electrons can be identified separately from hadrons by the shower depth, since hadrons produce more uniform energy deposits. Instead of using HPDs, the light is read out by photomultiplier tubes, with quartz or glass windows, which can withstand higher radiation.

### 4.5 Superconducting magnet

The solenoidal magnet is a crucial component of CMS, providing the necessary magnetic field to deflect charged particles and precisely measure their momenta. The magnet is 12.5 m in length, weighing 220 metric tons, with a 6 m diameter bore that houses the tracker, the ECAL, and the HCAL subdetectors. The superconducting niobium-titanium coils produce a 3.8 T field throughout the tracker volume. The magnetic field lines are returned by an iron yoke, which is interlaced with the muon system. The iron yoke is a 12-sided structure that surrounds the coils, 13 m long and 14

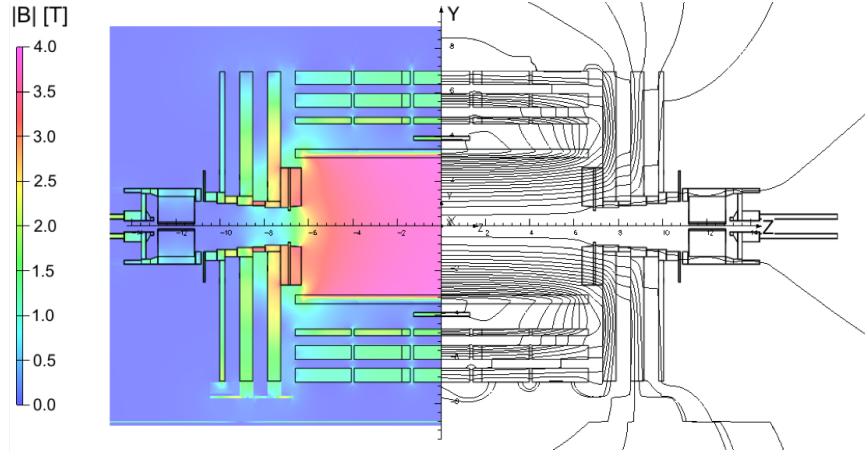


Figure 4.12: Simulation of the magnetic field inside the CMS detector [66].

m in diameter, and weighs 10,000 metric tons. It is composed of five wheels in the barrel and six discs in the endcap, each with three layers, and filters out most particles, allowing only muons and weakly interacting particles to pass through. Figure 4.12 shows a depiction of the magnetic field throughout the detector.

## 4.6 Muon system

The strategic placement of the calorimeters and the solenoid magnet ensure that any particle that reaches the outer muon systems is most likely a muon. The muon system is responsible for the identification, measurement, and triggering of muons. Since muons are too massive to induce showers in the calorimeter, the muon system uses a dedicated tracker. The muon system consists of three subsystems, each of which relies on incoming muons to ionize gas, and measures the resulting charge. The drift tubes (DT) located in the barrel ( $|\eta| < 1.3$ ) and the cathode strip chambers (CSC) in the endcaps ( $0.9 < |\eta| < 2.4$ ) provide good position and momentum resolution, while the resistive plate chambers (RPC) are spread throughout the barrel and endcaps and provide excellent timing resolution. These three subdetectors are shown in the colored-portions of Figure 4.13 and are discussed in further detail in the following sections.

The combination of information from all of the muons systems gives a muon reconstruction efficiency of over 99%, with momenta resolution of 1% in the barrel and 3% in the endcaps for a



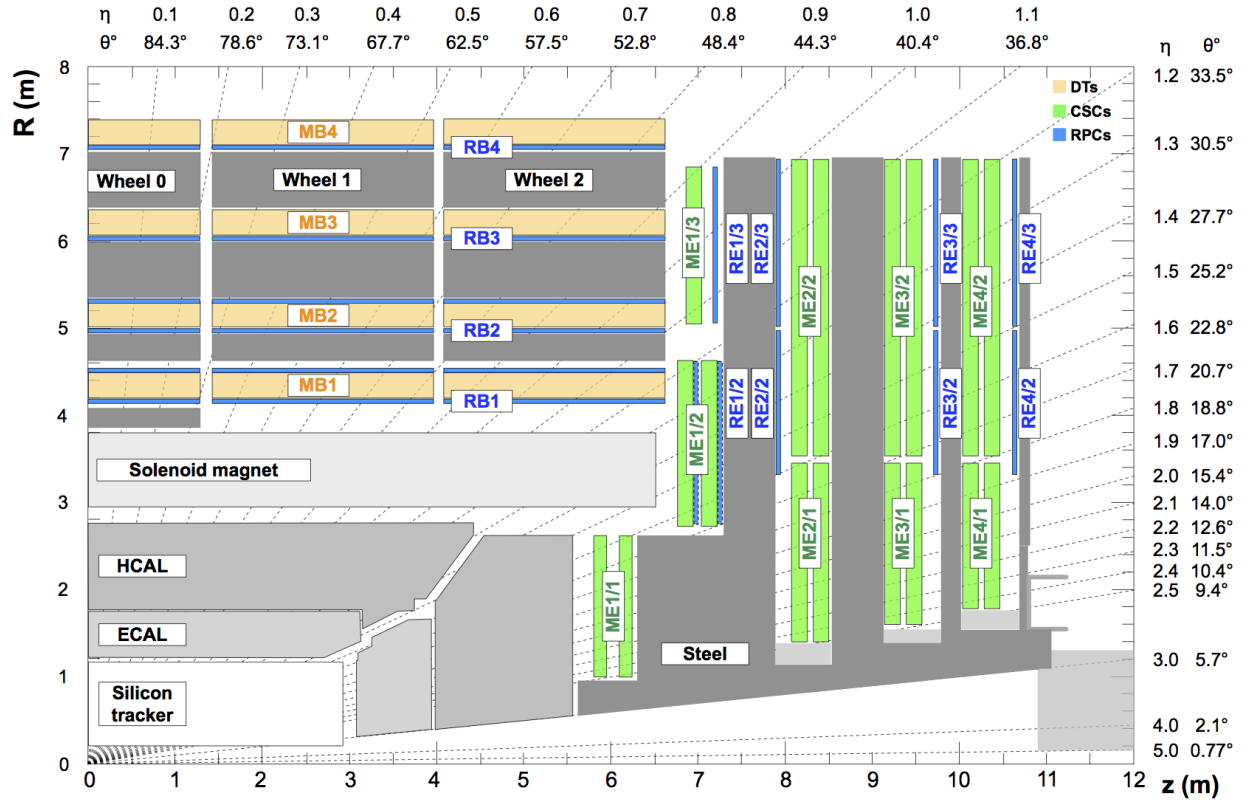


Figure 4.13: One quadrant of the CMS detector, with the different subdetectors of the muon system in color [67].

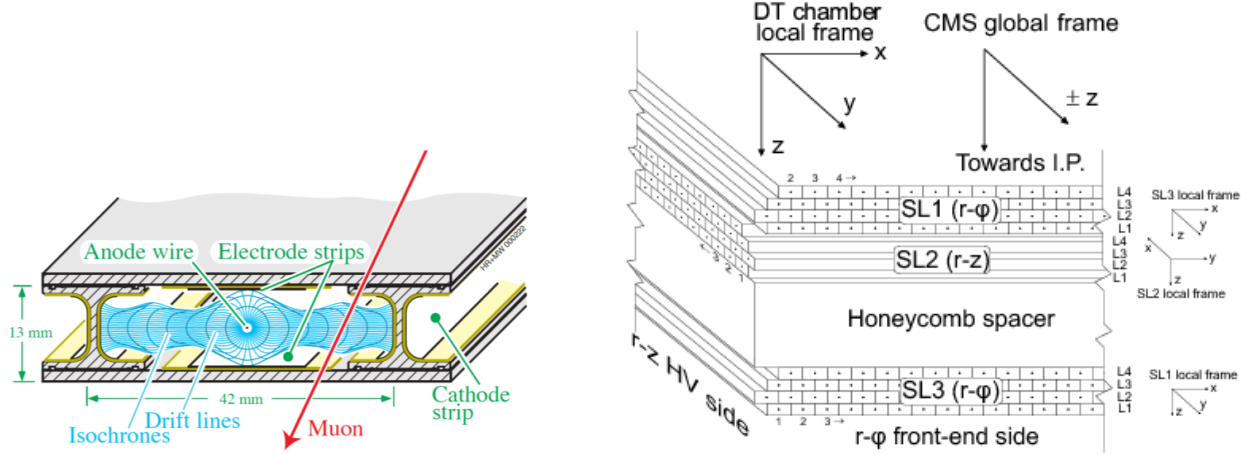


Figure 4.14: Left: Cross section of a drift tube cell [69]. Right: Schematic of a drift tube chamber, showing the orientation of the three superlayers [70].

muon with  $p_T = 100 \text{ GeV}$  [68]. During the Run 2 data-taking period, the spatial resolution was  $80\text{--}120 \text{ }\mu\text{m}$  in the DTs,  $40\text{--}150 \text{ }\mu\text{m}$  in the CSCs, and  $0\text{--}1.2 \text{ cm}$  in the RPCs. The time resolution achievable was 3 ns or better, for all three systems, and allowed particles from different bunch crossings to be identified.

#### 4.6.1 Drift tubes

The drift tubes (DT) of the muon system are composed of drift cells, each 4 cm wide and 2–4 m long, and are predominantly used to measure muon positions in the barrel. When an incident muon passes through the cell, it ionizes the gaseous mixture within (85% Ar, 15%  $\text{CO}_2$ ), and a high voltage wire strung within each drift cell collects the electrons. The DTs give two position coordinates for the muons, recording where the electrons hit the wire and calculating the muon's original distance from the wire using the electrons' speed. Four layers of drift cells, staggered by one-half width, form a superlayer, each with 60 cells. Three superlayers form a chamber, averaging  $2 \times 2.5 \text{ m}$  in size. Per chamber, the two outer superlayers are oriented for  $r-\phi$  measurements while the middle is orthogonal to allow for measurements in  $z$ . A cross-sectional view of a drift cell and a schematic of a DT chamber can be seen in Figure 4.14. The DT system consists of four DT chambers, at differing radii inside the return yoke of the magnet.

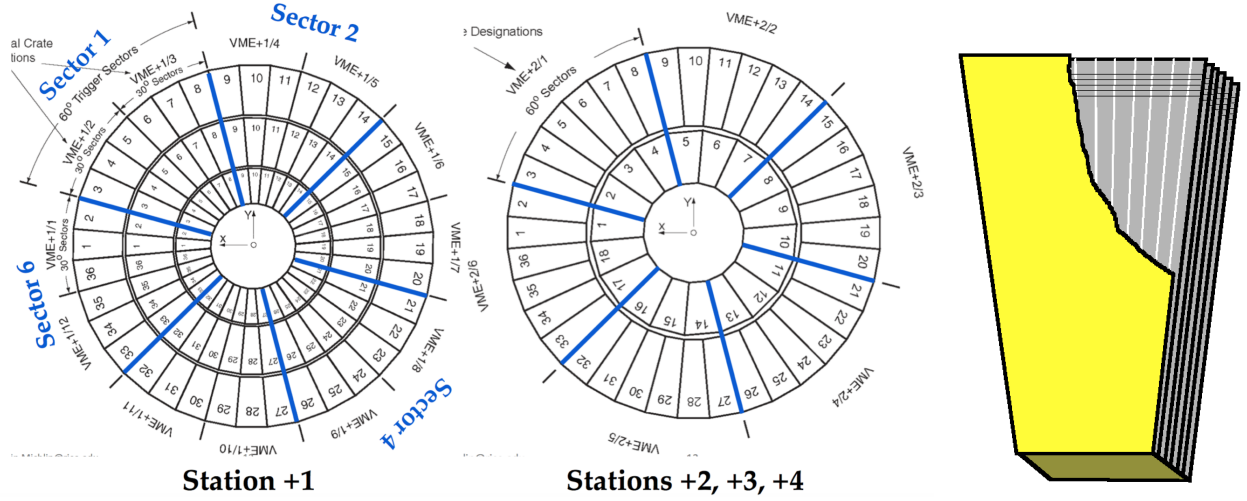


Figure 4.15: Left: Layout of the CSCs [71]. Right: Cross-sectional view of one chamber. The strips are oriented vertically while the wires are horizontal.

#### 4.6.2 Cathode strip chambers

The cathode strip chambers (CSC) are used to measure the radial position of muons in the endcap discs, where particle flux is high, and the magnetic field becomes stronger and less isotropic. Each CSC module is composed of six layers of copper cathode strips crossed with six layers of anode wires, separated by a layer of gaseous Freon and isobutane. Incident muons ionize the gas, causing electrons to move toward the wires and positive ions to move toward the cathode strips. The CSCs provide precise measurements of the muon hit's radial coordinate and  $\phi$  since the wires and strips are perpendicular. There are 468 chambers, arranged in two or three rings per disc, with four discs located in each endcap. The geometry of the system and a diagram of a cathode strip chamber are shown in Figure 4.15.

#### 4.6.3 Resistive plate chambers

Resistive plate chambers (RPC) are interspersed between the CSC and DT systems and have very good timing resolution, providing additional trigger information for the muon system. The RPCs consist of two parallel plates coated with conductive graphite paint, one positively charged and one negatively charged, separated by a gas mixture of 95.2% Freon ( $C_2H_2F_4$ ), 4.5% isobutane

( $\text{iC}_4\text{H}_{10}$ ), and 0.3% hexafluoride ( $\text{SF}_6$ ). Incoming muons ionize the gas and the strong electric field between the plates induces an electron avalanche. External aluminum strips that are separated and insulated from the graphite coating record the readout. There are four stations of six layers of chambers in the barrel and four stations with three layers of chambers in each endcap. The hit pattern across the strips allows for quick measurements of the muon's momentum, which is used for triggering. The RPCs have a time resolution of 1 ns, allowing for clear identification of muons from different bunch crossings. Figure 4.16 shows a cross section schematic of an RPC and the layout of the RPCs in the barrel.

## 4.7 Trigger system

The collision rate inside the LHC reaches up to 40 MHz, providing  $\sim 60$  TB/s of information if all events were kept. Since CMS cannot process nor store this much data, a two-level trigger system is used to quickly decide which events are kept based on the raw information collected by the subdetectors.

The first level (L1) trigger is implemented in the detector hardware (ASICs and programmable logic chips, FPGAs) and reduces the rate from 40 MHz to  $\sim 100$  kHz, the readout limit of the electronics. The L1 trigger receives information from only the calorimeters and the muon systems (not the tracker) and uses up to 512 distinct algorithms to determine which detector signals are consistent with candidate objects (such as energy clusters indicative of an electron) and should therefore be kept. A schematic of the Phase 1 L1 trigger system is shown in Figure 4.17, demonstrating which subdetectors and trigger systems send information to the Global Trigger. The Global Trigger must make a final event decision, to keep or not to keep, a fixed number of bunch crossings after the event collision. This amount of time is called *latency*. The L1 trigger has a design latency of 128 bunch crossings, so it must make the final decision within  $3.8 \mu\text{s}$ . The buffer pipelines of the readout electronics have limited storage capacity and must store all the information needed for the final decision, and so they are the limiting factor for the latency. This latency allows for the partial reconstruction and sorting of some objects.

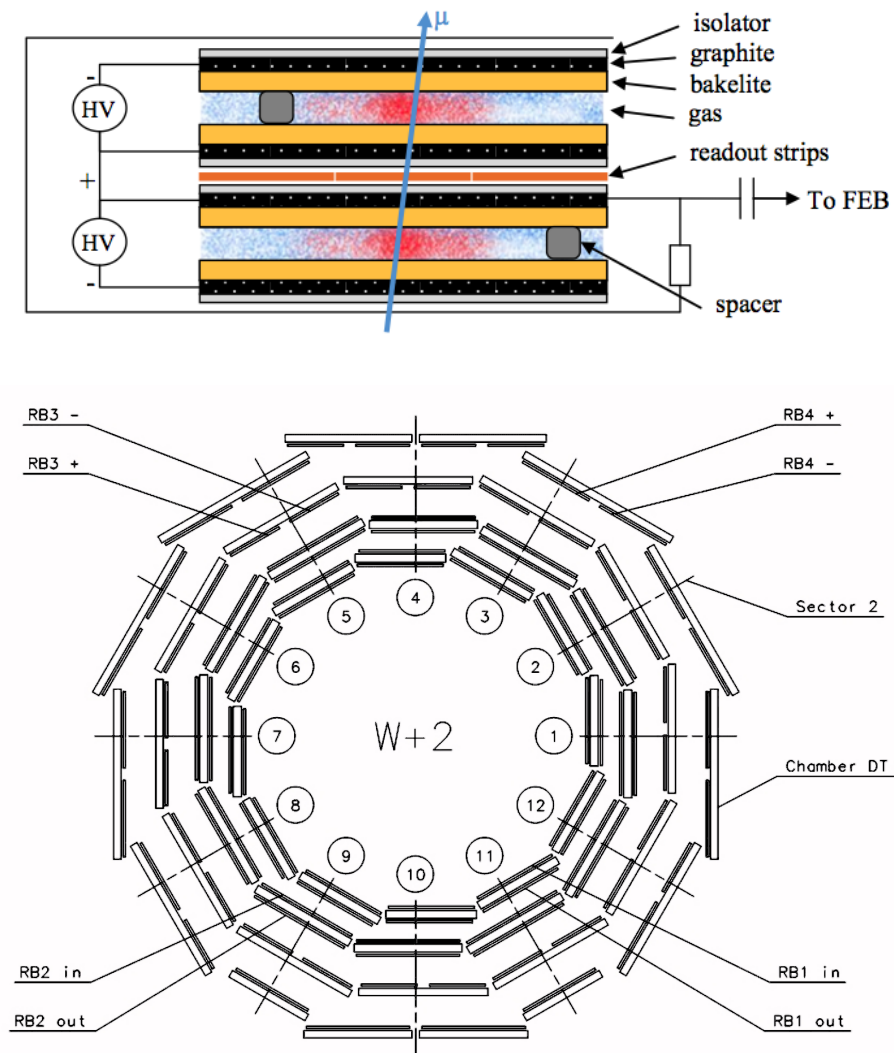


Figure 4.16: Top: Schematic of a resistive plate chamber [72]. Bottom: The barrel RPC system [73].

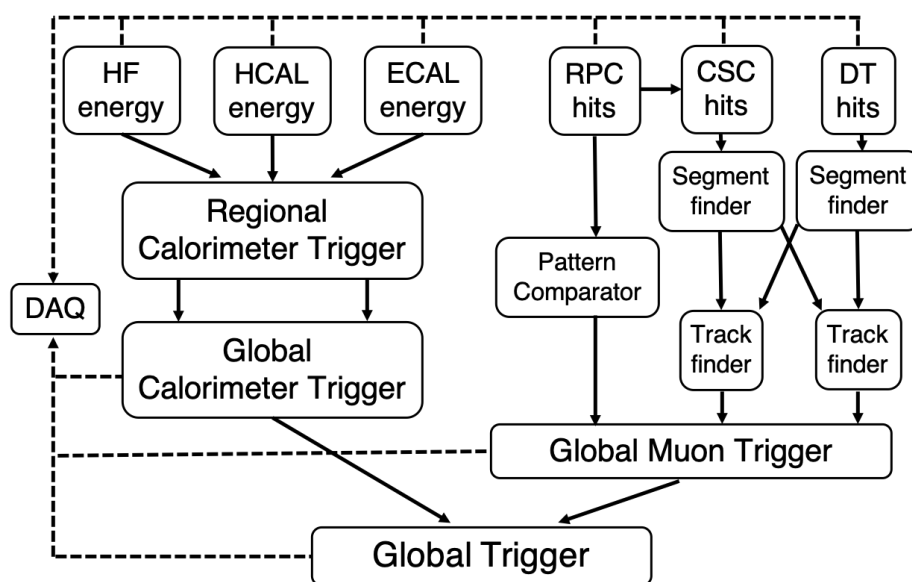


Figure 4.17: Overview of the CMS L1 trigger system for Phase 1. Trigger primitives (top row) are passed to regional and then global triggers (middle rows). The Global Trigger makes the final L1 trigger decision. The data acquisition system (DAQ) reads data from various triggers and subsystems for offline storage.

Even without tracker information, reconstruction of all particle types is still necessary in order to make a trigger menu that covers the full range of interesting physics; therefore, the events accepted at L1 are sent to a second level, known as the High Level Trigger (HLT). The HLT is a quicker version of the full offline reconstruction software, and, unlike the L1 trigger, it receives information from all subdetectors, including the tracker. Tracks are reconstructed using a similar (but simpler) iterative algorithm to what is used in offline reconstruction (further details in Section 5.1.1) and are then sent to be used in the Particle Flow (PF) algorithm, along with the information from the L1 trigger. PF is a reconstruction technique that identifies particles using information from the full detector and clusters them into more complex objects. More details on PF can be found in Section 5.2.

HLT decision paths, more commonly just called “triggers,” are developed and optimized to target specific event topologies. These triggers are sequences of reconstruction and selection algorithms, reproducing the offline requirements for a specific physics object (i.e., muons, electrons, jets, missing transverse energy) or for combinations of objects. Triggers for missing transverse energy are of particular note. Large amounts of missing energy can be an indicator of physics beyond the SM, as any weakly interacting unknown particles (such as dark matter or LSP candidates) would escape detection. Events are discarded throughout the reconstruction if they do not meet the trigger requirements, saving time and resources. For the Run 2 data-taking period used in this thesis, the HLT menu was composed of over 400 different triggers. The HLT further reduces the rate to an average of 400 Hz for storage.

## Chapter 5

### Event reconstruction

The data collected by the CMS detector are in energy deposits and detector hits, which need to be processed before they can be used in analyses. In this chapter, we discuss how particles and objects, such as leptons and jets, are identified and reconstructed. These objects are used to define the event selection criteria discussed in Section 6.3.

Figure 4.1 shows the trajectory of certain SM particles as they pass through the different subdetectors of CMS, leaving traces in multiple layers. Charged particle trajectories (tracks) and origin points (vertices) can be determined from the hits in the tracker layers, the innermost part of CMS [60]. These trajectories are bent by the magnetic field, allowing for measurements of the particle momentum and charge. After the tracker, electrons and photons are absorbed in the ECAL [74], leaving clusters of energy in adjacent cells that are used to determine the energy and direction of the particles themselves. Similar to the ECAL, the HCAL [65] absorbs charged and neutral hadrons (that do occasionally initiate a hadronic shower in the ECAL), whose energies and directions can be estimated using the corresponding clusters. Lastly, muons can be reconstructed from the hits they leave in the outermost subdetectors, the muon subsystems [68]. Neutrinos are the only known SM particles that do not leave any trace in the detector.

Local reconstruction occurs in individual subdetector systems, forming the most primitive of objects. Further reconstruction, rather at trigger-level or offline, such as the tracks from the tracker in Section 5.1.1, proceeds after these initial primitive objects are created:

- **Tracker clusters:** Local algorithms look for pixels or strips with a signal above a specific



threshold. These pixels/strips are then used as seeds for clusters, which are built by adding together neighboring pixels/strips.

- **ECAL clusters:** Localized energy deposits are used to determine the position, time of arrival, and energy of ECAL clusters.
- **HCAL clusters:** Similarly, localized energy deposits are used in local reconstruction to determine the position, time of arrival, and energy of HCAL clusters.
- **Track segments in the DTs:** Local reconstruction gives the position of a muon hit in individual cells. Muon track segments are built from hits within a given superlayer.
- **Track segments in the CSCs:** Two-dimensional muon hits are obtained in each layer. These hits are then combined to create three-dimensional track segments within each chamber. The local reconstruction provides time of arrival and the position of the muon hit from the charge induced on the strips.
- **Track segments in the RPCs:** The position of a muon hit is reconstructed from the clusters of hits on the strips.

## 5.1 Basic detector elements

The primitive objects listed above are then used in reconstruction algorithms to create the four basic elements: tracks from the tracker (Section 5.1.1), ECAL and HCAL superclusters (Section 5.1.2), and muon tracks (Section 5.1.3). After initial detection and local reconstruction, the four elements are combined via the Particle Flow (PF) algorithm, which takes in information from all subdetectors to fully identify and reconstruct the particles used for analyses. A more detailed explanation of PF is given in Section 5.2. Finally, in Section 5.3 we describe some of the higher-level objects that provide particularly helpful handles for BSM analyses.

### 5.1.1 Tracks

As discussed in Chapter 3, the LHC is designed to run with a peak instantaneous luminosity of  $\mathcal{L} = 10^{34} \text{cm}^{-2} \text{s}^{-1}$  with the proton bunch crossings occurring every 25 ns. The average number of proton-proton collisions per bunch crossing, called pileup, varies with the instantaneous luminosity, averaging about  $\langle n_{\text{vertex}} \rangle = 29$  for the full Run 2 data set. This pileup, as well as the strict time requirement and the granularity of the silicon trackers, makes reconstructing tracks and vertices challenging.

Tracks are reconstructed by first forming track seeds or clusters from a small number of hits in the pixel and strip detectors. The average efficiency for hit reconstruction is  $>99\%$  when known defective pixel modules are excluded, and roughly  $99.8\%$  in the strip detector. This pattern recognition stage begins by looping through a subset of possible pairs of hits in individual layers, beginning with the innermost layers. When pairs of hits are identified, a helix is constructed through the hits and the beam axis. The track is then propagated backwards toward the beamline to the next internal layer, and the algorithm looks for hits near the estimated point of intersection. The propagation finishes when a minimum number of hits (two or three<sup>1</sup>) is found. If there are too few layers hit, or too many missed layers, the track is abandoned. Since the reconstruction algorithm begins with all detector hits, it often finds the same track seeds many times, and so duplicate track seeds are removed at the end of the pattern recognition step.

The reconstruction proceeds using a Kalman filter (KF) algorithm, which begins with the helix computed in the pattern recognition stage, projects the track seeds onto successive detector layers to look for additional hits, and updates the track information after each hit is found. If there are several likely hit candidates for a track, each candidate is added in turn to the fit, and the hit candidate that produces the the best track  $\chi^2$  value is kept. The tracks are accepted and saved if they have a minimum number of hits, not too many consecutive missing layers, and an acceptable  $\chi^2$  value. After the KF, duplicate tracks are removed; the longest track is kept if it has an acceptable  $\chi^2$  value, otherwise the track with the smallest  $\chi^2$  is kept.

---

<sup>1</sup> The minimum number of hits is dependent on which layers or discs were hit.

The final step of the tracking algorithm performs a complete track fit, again using a Kalman filter, but starting from the innermost hit. This refitting mitigates potential bias from constraints, such as a beamspot constraint if put in place. The trajectory estimate is first obtained by using the KF fit on the first four innermost hits. The fit then continues outward through the list of track hits. A smoothing stage follows this complete fit, where a second KF is initialized using the result from the first, and iteratively fits the track from the outermost layer toward the beam spot.

For prompt muons, the tracking efficiency throughout Run 2 is essentially 100% for the entire muon pseudorapidity acceptance [75,76]. In the barrel region ( $|\eta| < 1.4$ ), muons with  $p_T = 100$  GeV have resolutions of 2.8% in transverse momentum and  $10\ \mu\text{m}$  and  $30\ \mu\text{m}$  in transverse and longitudinal impact parameters. Because electrons lose much of their energy through bremsstrahlung radiation before reaching the outer layers of the tracker, electron tracks are reconstructed using a modified Gaussian sum filter [77], which account for the energy loss. In the barrel, the efficiency for electrons exceeds 90% for  $p_T > 0.4$  GeV.

The primary vertex of the event is reconstructed from the available tracks and distinguishes the vertex of the hard interaction from the pileup vertices. After applying quality criteria to the tracks, such as the number of hits in the tracker, the tracks are clustered using a deterministic annealing (DA) algorithm [78], which finds a global minimum by iteratively considering all possible solutions, calculating a probability that each solution is “good,” and progressively moving toward the optimum solution. Additionally, the DA algorithm is extended with a rejection term for outliers originating from secondary vertices or mismeasured tracks. After the DA clustering, the candidate vertices with at least two tracks are re-fitted using an adaptive vertex fitter [79]. For this analysis, we require at least one good primary vertex, classified as being consistent with the beam spot and having a minimum of four tracks. The primary vertex of the event is the one with the largest sum of the squared track momenta ( $\sum p_T^2$ ); other vertices are classified as pileup vertices. Every event will have a primary vertex reconstructed by definition and analyses place additional requirements for selecting events with suitable primary vertices. Typically, the primary vertex resolution is  $10\text{-}12\ \mu\text{m}$  in all three dimensions, leading to excellent removal of tracks from pileup vertices when required.

### 5.1.2 ECAL & HCAL superclusters

The clustering procedure is run separately for the ECAL and the HCAL, but the same structure of algorithm is used. In the calorimeters, clusters are formed by adding all of the hits in the crystals that neighbor the crystal with the largest energy deposit. This proceeds until all clusters have been reconstructed and there are no remaining calorimeter hits. Cells are clustered in the EB, EE, HB, HE, and the first two layers of the preshower detector. Each cell in the HF is considered a cluster. The cluster begins with a single crystal, with an energy deposit above a certain threshold. This is called a cluster seed and is reconstructed locally in detector modules. The cells next to each cluster seed are added together to create topological clusters, (a  $5 \times 5$  array for the ECAL, and roughly  $0.09 \times 0.09$  in  $\eta - \phi$  for the HCAL) with the original cluster seed at the center. These can contain multiple cluster seeds.

The final clusters are determined by iteratively dividing the energy from each cell across the cluster seeds for every topological cluster.  $N$  clusters are identified assuming Gaussian distributions of  $N$  energy deposits. Each cluster has energy  $E$ , a position in the  $\eta - \phi$  plane, and a width scale  $\sigma$  that is set by the size of the crystal. The cluster location is recalculated at every iteration, which continues until the change in position is less than some small fraction of the position resolution.

The ECAL energy resolution for electrons with  $E_T \approx 45$  GeV is better than 2% in the barrel regions, and 2–5% elsewhere, with the energy resolution for photons between 1–5%. The ECAL was specifically designed to provide excellent invariant mass resolution and fine transverse granularity to aid in photon identification, in order to enhance the sensitivity to the  $H \rightarrow \gamma\gamma$  decay, the decay channel the Higgs boson was first discovered in. The high-resolution ECAL changed the primary function of the HCAL, shifting it to focus more on particle isolation and identification rather than energy resolution (with the exception of neutral hadron energy). For a 300 GeV pion beam, the energy resolution in the HCAL is under 10%. When information from the ECAL is combined with the HCAL, jets are measured with a resolution of  $\Delta E/E \approx 100\% \sqrt{E[\text{GeV}]} \oplus 5\%$ .

### 5.1.3 Muon tracks

The hits in the DT, CSC, and RPC subdetectors are used to reconstruct muon tracks. Reconstruction of local track segments is done quickly to allow for triggering and bunch crossing identification. The track segments are then passed to a track-finding algorithm and used as the seeds. The track reconstruction then proceeds using the Kalman filter (KF) method, in a very similar fashion as is used for reconstructing tracker tracks. Section 5.1.1 provides more detail on the Kalman filter method. Briefly, the KF begins with a seed (a track segment) and iteratively adds track segments from the DTs and hits from the CSCs and RPCs, running first outward, and then inward. If a hit were to increase the track  $\chi^2$  above a certain threshold, it is rejected. Muon tracks are then matched to tracks from the tracker, with a transverse momentum resolution of 1.3-2.0% for muons with  $p_T$  [20,100] GeV in the barrel, and better than 6% in the endcaps.

## 5.2 Particle flow

The particle flow (PF) algorithm [80] uses information from all of the subdetectors, specifically in Section 5.1, for final object reconstruction. The PF algorithm identifies the candidates as charged or neutral hadrons, photons, electrons, or muons, and feeds these candidates into higher level algorithms.

Iterative clustering and track reconstruction methods are used in PF in order to reconstruct objects with high efficiency and low fake rates. The Kalman filter method is used to determine single particle tracks (see Section 5.1.1). Once complete, incoming elements are grouped together in the  $\eta - \phi$  plane if they could come from the same particle. After these “blocks” are formed, the actual particle reconstruction proceeds in order of ambiguity for each reconstructed object, first muons, then electrons, and then photons and neutral particle candidates. Once used, blocks are discarded.

For muons, the muon tracks iteratively add the tracker tracks to form global muons, which are then re-fit. Like the reconstruction of both tracker tracks and muon tracks, the tracker tracks

are accepted only if their addition keeps the resulting global muon fit below a certain  $\chi^2$  threshold. The global muon is kept (and identified as a global muon by PF) if the global muon momentum is compatible with the tracker-only momentum, within three standard deviations.

Electron candidates are formed from linking tracker tracks compatible with an electron to ECAL clusters. The reconstruction process is more complicated than it is for muons, because electrons interact with the tracker material, emitting bremsstrahlung radiation that must be properly accounted for. A Gaussian sum filter algorithm [77] is used to associate the radiated energy with the electron candidate, recovering the lost energy. All ECAL clusters in a PF block are associated with the electron candidate, if very strict requirements on the track-to-cluster compatibility are satisfied. This ensures optimal energy containment. The reconstruction efficiency for electrons is typically greater than 95%, and better at higher  $p_T$  values. Photons are identified by ECAL clusters that cannot be associated with a track and are assumed to have 100% reconstruction efficiency. Physics analyses apply specific electron- and photon-identification criteria, a trade-off between retaining high efficiency and removing candidates that were misidentified as electrons or photons.

The charged and neutral hadrons are reconstructed from the remaining blocks, after the electron and muon candidates are removed. The charged hadron candidates are formed from the remaining tracks, and require compatibility with HCAL clusters, while neutral hadrons typically leave small energy deposits in the ECAL.

After all blocks have been processed, PF sends off its list of particle candidates to higher-level algorithms. These include collections for electrons, muons, and photons. For the remaining sections and chapters of this thesis, all event observables (such as missing energy and isolated muons) are created from the PF candidates, unless otherwise indicated. PF candidates are also now used almost exclusively for the HLT trigger paths discussed in Section 4.7.

### 5.3 High-level objects

#### 5.3.1 Jets

Since gluons and quarks are colored particles, as they pass through the detector, they cause  $q\bar{q}$  pairs to pop-up around them, in a process called hadronization. These new particles typically travel in the same direction as the original quark or gluon, creating a collimated spray that is reconstructed as a jet. Jets are clustered from all PF candidates using the anti- $k_T$  algorithm [81], which starts with clustering two of the closest candidates together and continues to reconstruct the jet in an enclosed area until a termination distance  $d_{i,j}$  is reached:

$$d_{i,j} = \min(k_{t,i}^{-2}, k_{t,j}^{-2}) \frac{\Delta_{i,j}^2}{R^2} \quad \text{and} \quad d_{i,B} = k_{t,i}^{-2}, \quad (5.1)$$

where  $d_{i,j}$  is the distance between entities (particles, pseudojets)  $i$  and  $j$ , and  $d_{i,B}$  is the distance between the entity  $i$  and the beam.  $R$  is a measure of the radius of the jet in the  $\eta - \phi$  plane, and is defined as  $R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ .  $\Delta_{i,j}^2$  is the spatial distance in the  $y - \phi$  plane and  $k_{t,i}$  is the transverse momentum. The  $d_{i,j}$  is calculated iteratively until it reaches  $d_{i,B}$ , at which point the entity  $i$  becomes a jet. For this analysis, the clustering is performed with distance parameter  $R = 0.4$  (AK4 jets) for jets containing the fragmentation products of a single parton, and with distance parameter  $R = 0.8$  (AK8 jets) for wider jets that contain multiple partons.

If the energy fraction of one of the jet components (neutral hadron fraction, neutral EM fraction, or charged EM fraction) exceeds 99% of the total jet energy, this indicates the jet is from detector noise and is rejected [82]. Additionally, the energy of the jets is corrected for the non-uniform detector response, and the jet  $p_T$  is recalculated after removing the estimated contribution from neutral PF candidates coming from pileup.

#### 5.3.2 b-tagging jets

Jets that originate from  $b$  quarks are useful for reconstructing the Higgs bosons used in this analysis. As mentioned in Section 2.1.7, individual  $b$  quarks will hadronize, forming a  $b$  hadron.

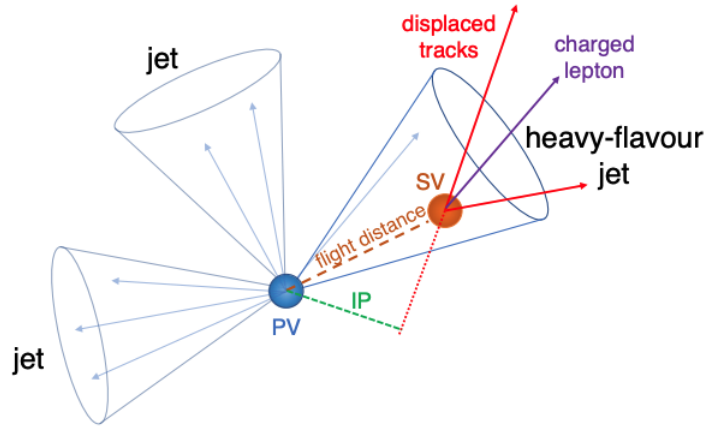


Figure 5.1: Diagram of a jet with a secondary vertex (SV) from the decay of a b (or c) hadron, resulting in charged-particle tracks that are displaced with respect to the primary vertex (PV), and hence have a large impact parameter (IP) value.

The relatively long lifetime of a b hadron (1.5 ps), in comparison to light-flavored hadrons, is such that it will travel hundreds of microns before decaying, creating a secondary vertex that can be reconstructed. A secondary vertex (SV), shown in Figure 5.1, is one that is spatially distanced from the primary vertex, and provides a very useful handle for identifying (“tagging”) jets that come from b quarks, called b jets.

Several algorithms for b jet identification, called b-taggers, are supported by the CMS experiment. For AK4 jets, we use a version of the combined secondary vertex algorithm based on deep neural networks (DEEPCSV [83]), which prescribes a score based on how likely a jet is to contain a particle from a SV. The SV is crucial to this identification process, as are the tracks associated with this SV. The following parameters are input into the training of the deep neural network:

- SV 2D flight distance significance
- Number of SV
- Number of tracks from SV
- Corrected SV mass
- SV energy ratio
- 3D IP significance of the first four tracks



- Summed track  $E_T$  ratio
- $\Delta R(\text{summed tracks, jet})$
- Track  $\eta_{rel}$
- Track  $p_{T,rel}$
- Track distance
- Track decay length
- First track 2D IP significance above c threshold
- Number of selected tracks
- $\Delta R(\text{track, jet})$
- Jet  $p_T$
- Jet  $\eta$

A b-jet tagging efficiency of 85% corresponds to a misidentification rate of about 10% for a light-parton jet and 40% for a charm jet. This is the loose (L) working point (WP) for the DEEPCSV algorithm. The medium (M) and tight (T) WPs have b-jet tagging efficiencies of 70 and 50%, with corresponding light-parton jet mistag rates of 1 and 0.1%, and charm jet mistag rates of 12 and 2.5%, respectively. All three WPs are used in the analysis of this thesis; further analysis specific requirements on b jets are described in Section 6.3.2.

In a similar way, wider jets can be identified that contain two b quarks. For this analysis, we use a deep-learning-based double-b tagging algorithm [84] (mass-decorrelated) that assigns a score on how likely an AK8 jet is to contain two b quarks, called bb jets in this thesis. The algorithm relies on a variable called  $N$ -subjettiness [85], which uses the shape of the jet and the internal jet structure to essentially count the number of subjets inside a given jet. Twenty-seven properties of jets (more in Ref. [83]) are input as parameters to the training, as well as eight properties of the particle-flow jet constituents and two properties of the SVs associated with the jet. The score is mass-decorrelated by adding a penalty term to the loss function minimized during the training, which penalizes the difference between the mass distribution shapes for tagged and untagged jets. The bb-tagging criteria for analyses typically require an AK8 jet to pass one of the centrally-defined

working points. For the analysis in Chapter 6, this is the loose WP, corresponding to an efficiency of 90% and a misidentification rate of 5% for light partons.

Generally, the efficiency for tagging a b jet is not the same in simulation as it is in data. To correct for this difference, we compute event-by-event weights that use generator-level parton flavor,  $p_T$ , and  $\eta$  to determine the probability that a given event will be above the DEEPCSV working points. To determine these weights [86], we compare multijet,  $Z \rightarrow b\bar{b}$ , and  $t\bar{t}$  events in data and simulation, and apply them as corrections (called scale factors) to both signal and SM background simulation. Similarly, for bb jets, we derive true tagging scale factors (SFs) from comparisons of  $g \rightarrow b\bar{b}$  and multijet events, and mistag scale factors from mostly  $t\bar{t}$ ,  $Z$ +jets, and  $W$ +jets events. The SFs are roughly on the order of a few percent to 10% corrections.

### 5.3.3 Missing energy

Weakly interacting particles, such as neutrinos, leave no energy deposits in the detector and as such cannot be directly detected. Instead, we infer their presence from the imbalance in transverse momentum. Missing transverse energy (also called missing transverse momentum,  $\vec{p}_T^{\text{miss}}$ ) [87] is calculated from the negative vector sum of all PF candidates,

$$\vec{p}_T^{\text{miss}} = - \sum_i \vec{p}_{T,i}, \quad (5.2)$$

the visible particles in the event. While the net momentum of the protons in the collisions is zero, the fraction of momenta in individual partons within the proton along the  $z$ -axis cannot be known. Because of this, it is common to sum the momenta of all particles in the transverse plane since that is conserved and known. Large momentum imbalance indicates the presence of an undetected particle, such as a neutrino or particles not accounted for in the Standard Model. This is of particular importance to beyond the SM analyses since it can be a direct indication of a new particle (such as the lightest supersymmetric particle or a dark matter candidate) escaping detection. Unfortunately, due to detector resolution and imperfect reconstruction, we cannot perfectly measure the energy of every particle, which can lead to anomalous  $p_T^{\text{miss}}$ . Dedicated studies of the  $p_T^{\text{miss}}$  resolution have

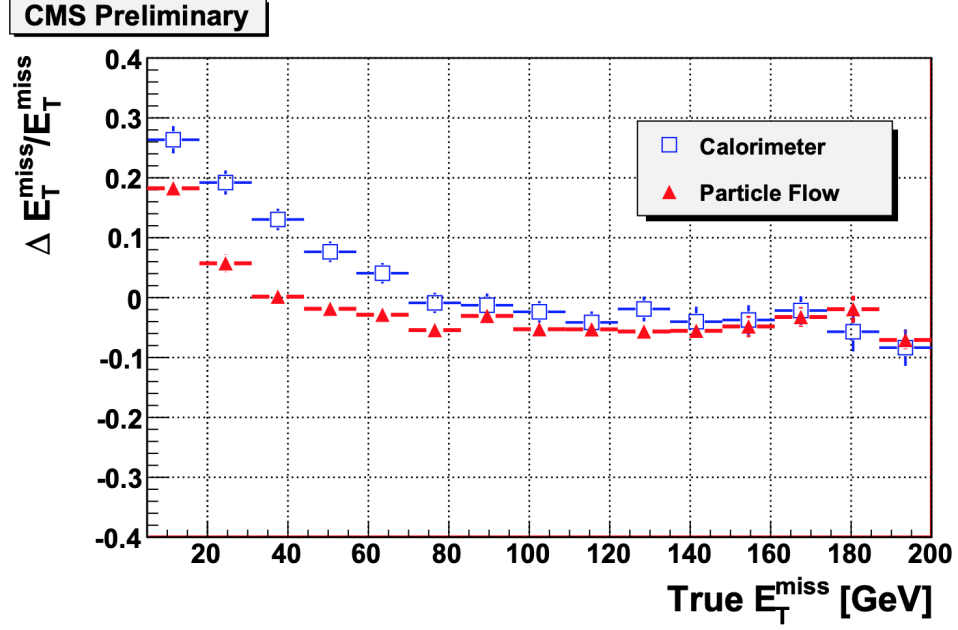


Figure 5.2: Distribution of  $(p_{T,\text{reco}}^{\text{miss}} - p_{T,\text{true}}^{\text{miss}})/p_{T,\text{true}}^{\text{miss}}$  as a function of  $p_{T,\text{true}}^{\text{miss}}$  in simulated  $t\bar{t}$  events for PF  $p_T^{\text{miss}}$  (red triangles) and calorimeter  $p_T^{\text{miss}}$  (blue squares).  $p_T^{\text{miss}}$  is occasionally written as  $E_T^{\text{miss}}$ , where  $E$  stands for energy.

determined filters and corrections (discussed in Section 6.3.1) to correct the reconstructed  $p_T^{\text{miss}}$  and to remove events with anomalous  $p_T^{\text{miss}}$ , to increase our chances of selecting events with true  $p_T^{\text{miss}}$ . Figure 5.2 shows the  $p_T^{\text{miss}}$  resolution as a function of the true  $p_T^{\text{miss}}$  from simulation, comparing PF  $p_T^{\text{miss}}$  to the  $p_T^{\text{miss}}$  from the calorimeters only. For events with true  $p_T^{\text{miss}} > 20$  GeV, the resolution for PF  $p_T^{\text{miss}}$  is under 10%.

## Chapter 6

### Search for supersymmetry in channels with two Higgs bosons and missing transverse momentum

We will now describe the main premise of this thesis, an analysis searching for physics beyond the SM in events with two Higgs bosons and large missing transverse energy. We target the Higgs bosons in the decay  $H \rightarrow b\bar{b}$ , which has the largest branching fraction for the Higgs, 58% [88]. As discussed previously, large missing transverse energy is indicative of a particle leaving the detector unnoticed (such as the LSP), and therefore gives a nice signature for SUSY searches. Preliminary results from this search have been made public [89]. Detailed documentation appears in an internal note, Ref. [90].

The general approach of this analysis is to first identify kinematic requirements that have high efficiency for selecting signal events and low efficiency for selecting SM backgrounds, specifically  $t\bar{t}$ ,  $W$ +jets,  $Z$ +jets, and to a small degree QCD multijet events. As we will discuss in Section 6.3, we impose selection criteria to reduce our main SM backgrounds. In particular, semi-leptonic  $t\bar{t}$  events are our largest background, since they produce  $b$  quarks and true  $p_T^{\text{miss}}$  from the neutrino of the  $W$  decay. We substantially reduce the contribution from this background by requiring an all-hadronic final state. Any events with charged leptons are removed.  $Z \rightarrow \nu\nu$  events are also a predominant background, since the neutrinos provide large missing energy. These events are removed by tight  $b$ -tagging requirements.

To test for the presence of BSM contributions in the data, we must develop a way to measure the remaining SM background contribution, so that we may compare the SM background prediction

to the data. The background prediction method is described in Section 6.4, while the validation of the method, including the background and signal systematic uncertainties, is detailed in Section 6.5. The data yields and analysis results are described in Section 6.6, including the interpretation of the results using simplified SUSY models.

In this analysis, we look for an all-hadronic final state containing four b quarks and large missing energy. For the Higgs reconstruction, then, we expect two b-tagged jets per Higgs decay, four b jets per event. An interesting topology occurs, however, when a Higgs boson is produced with large enough momentum: its decay products are emitted at smaller angles, collimating enough to be reconstructed as a single jet. This boosted topology can arise when the parent particles of the Higgs (the neutralinos in our case) are massive enough, and thus impart higher momentum to the final state particles. This analysis focuses partially on the boosted topology, which is increasingly sensitive to higher mass SUSY particles; previous SUSY searches have already excluded lower mass neutralinos and gluinos (see Section 2.2.5). We therefore consider two types of topologies: events in which the b-tagged jets are resolved in distinct AK4 jets, and events in which the b jets are close enough together to be captured in a single wide-cone AK8 jet. These topologies can be seen in Figure 6.1. We refer to the separated and merged possibilities as the resolved signature and the boosted signature, respectively.

## 6.1 Signal models

Targeting a full SUSY model, like the MSSM, in a search for new physics is difficult because the multitude of parameters add complexity to signal generation, and interpretations can be too model-specific. Simplified models are more often used [91, 92], and designed to involve only a few SUSY particles, with the remaining particle masses set intentionally to be out of the LHC’s reach. These models exhibit final states that are representative of a number of SUSY models with a wide range of parameters and are particularly helpful for comparing different supersymmetry-inspired analyses. For the search for physics beyond the SM covered in this thesis, three simplified SUSY models are considered, as shown in Figure 6.2: two different electroweak pair production scenarios

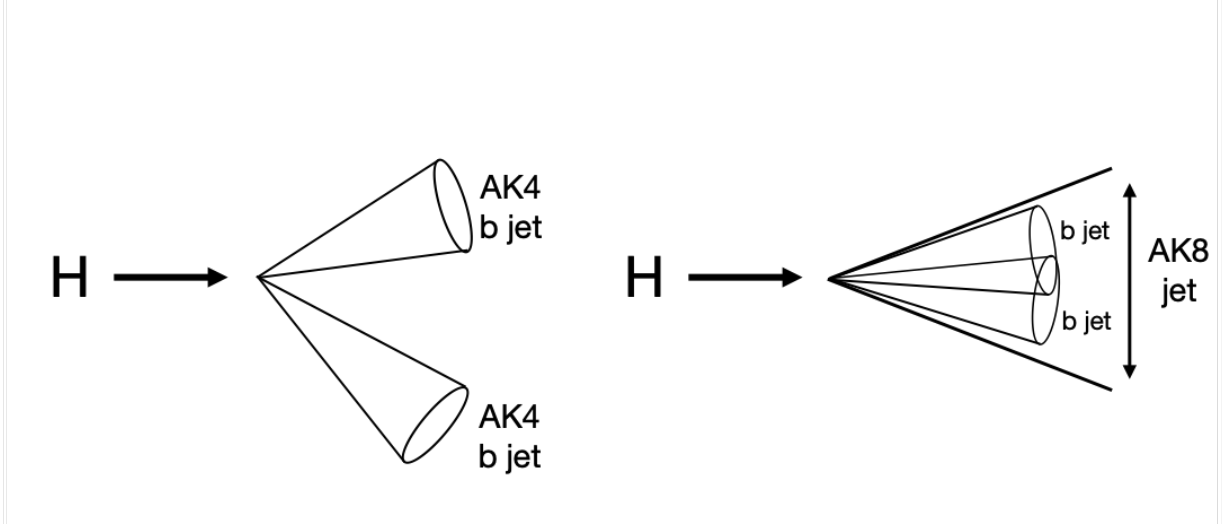


Figure 6.1: Diagrams of the resolved (left) and boosted (right) topologies targeted in this analysis. Each Higgs decay produces two b quarks. At high enough momentum ( $p_T > 300$  GeV), the b quarks are collimated enough to be captured in a single jet.

(left and center) and gluino pair production (right).

For the first electroweakino pair production model (left in Figure 6.2), the TChiHH-G model, we take the massless goldstino as the LSP and the higgsino-like lightest neutralino  $\tilde{\chi}_1^0$  as the next-to-lightest SUSY particle (NLSP) [93]. The NLSPs are produced in the cascade decays of different combinations of neutralinos and charginos. When the lighter neutralinos  $\tilde{\chi}_{1,2}^0$  and charginos  $\tilde{\chi}_1^\pm$  are nearly mass degenerate, all of their cascade decays can lead to the production of the NLSP (and other, low-momentum particles). Integrating over the possible contributions of the produced charginos and neutralinos ( $\tilde{\chi}_1^0 \tilde{\chi}_2^0$ ,  $\tilde{\chi}_1^0 \tilde{\chi}_1^\pm$ ,  $\tilde{\chi}_2^0 \tilde{\chi}_1^\pm$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ ) leads to an effective rate for  $\tilde{\chi}_1^0 \tilde{\chi}_1^0$  production [94, 95] that is significantly larger than that of any of the individual pairs. The larger production rate boosts experimental sensitivity. The higgsino-like NLSPs then decay via  $\tilde{\chi}_1^0 \rightarrow \gamma \tilde{G}$ ,  $\tilde{\chi}_1^0 \rightarrow H \tilde{G}$ , or  $\tilde{\chi}_1^0 \rightarrow Z \tilde{G}$ , where the undetected  $\tilde{G}$  can generate large  $p_T^{\text{miss}}$ . The TChiHH-G model considered in this thesis assumes a 100% branching ratio to the  $H \tilde{G}$  final state.

The more generic model, TChiHH, is the electroweak production of two higgsino-like neutralinos  $\tilde{\chi}_2^0$  and  $\tilde{\chi}_3^0$ . This model contains a bino-like LSP ( $\tilde{\chi}_1^0$ ) and higgsino-like NLSPs with  $m(\tilde{\chi}_2^0) = m(\tilde{\chi}_3^0)$ . Unlike in TChiHH-G described above, we consider only the higgsino production cross section for  $\tilde{\chi}_2^0 \tilde{\chi}_3^0$ , assuming we are not sensitive to the other combinations ( $\tilde{\chi}_2^0 \tilde{\chi}_1^\pm$ ,  $\tilde{\chi}_3^0 \tilde{\chi}_1^\pm$ ,

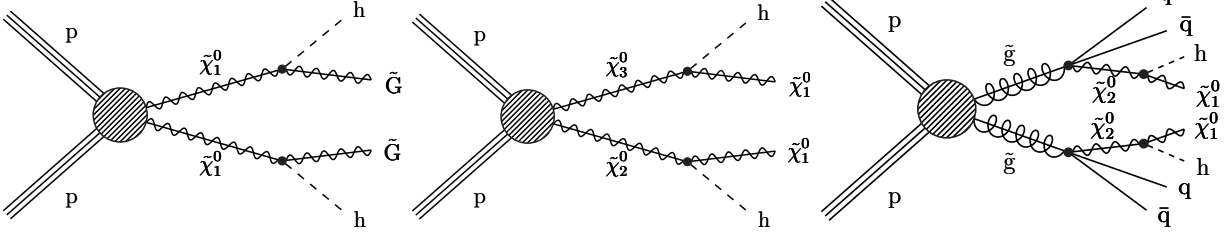


Figure 6.2: Diagrams for the SUSY models we interpret our results in, TChiHH-G (left), TChiHH (center), and T5HH (right). In each diagram, the filled circle represents the sum of processes that can lead to the SUSY particles shown.

$\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ ), since  $\tilde{\chi}_1^\pm$  cannot decay to a Higgs boson.

For the T5HH model, the proton-proton interaction produces a pair of gluinos  $\tilde{g}$ , each of which decays to the second-lightest neutralino  $\tilde{\chi}_2^0$ , along with the emission of SM quarks. For simulation purposes, we set the mass of the  $\tilde{\chi}_2^0$  to be 50 GeV below  $m(\tilde{g})$ , so that the quarks are soft and do not normally register as jets. The  $\tilde{\chi}_2^0$  further decays into a SM Higgs boson and the lightest neutralino  $\tilde{\chi}_1^0$ , which is taken to be the LSP. Many models predict that the production rates of high-momentum Higgs bosons are greater in gluino cascade decays than in direct Higgsino production [96, 97]. This is driven by the significantly larger strong production cross section of gluinos vs electroweakinos.

For all models, we impose requirements for large momentum imbalance, indicative of the presence and escaped-detection of the LSP. While dependent on the LSP, a dark matter candidate, these SUSY signal models were not built with dark matter candidates or detection in mind, and the search itself makes no assumptions related to dark matter. Simulation details for these signal samples, including the mass ranges considered, can be found in Section 6.2.2.2.

## 6.2 Data and simulation

The SM background estimation methods described in Section 6.4 are based primarily on observed event yields in data. A data-driven approach allows us to use the data directly, without relying on our ability to completely and perfectly model complex interactions and the detector itself. In Section 6.2.1, we describe the data samples used in this analysis, as well as the HLT

decision paths used for data collection. Monte Carlo (MC) simulation is used to optimize the event selection criteria, validate the background estimation methods (including any necessary correction factors), and interpret the observed data in the context of a signal model. The details on sample generation, for both SM backgrounds and simplified SUSY models, are given in Section 6.2.2.

### 6.2.1 Data and triggers

| Year  | Era | $\int \mathcal{L}$ [pb <sup>-1</sup> ] | Year  | Era | $\int \mathcal{L}$ [pb <sup>-1</sup> ] | Year     | Era | $\int \mathcal{L}$ [pb <sup>-1</sup> ] |
|-------|-----|----------------------------------------|-------|-----|----------------------------------------|----------|-----|----------------------------------------|
| 2016  | A   | -                                      | 2017  | A   | -                                      | 2018     | A   | 14024.2                                |
|       | B   | 5746.4                                 |       | B   | 4793.3                                 |          | B   | 6907.9                                 |
|       | C   | 2572.9                                 |       | C   | 9632.7                                 |          | C   | 6894.8                                 |
|       | D   | 4242.3                                 |       | D   | 4247.7                                 |          | D   | 31719.5                                |
|       | E   | 3924.3                                 |       | E   | 9314.0                                 |          | E   | -                                      |
|       | F   | 3104.5                                 |       | F   | 13498.4                                |          | F   | -                                      |
|       | G   | 7575.8                                 |       | G   | -                                      |          | G   | -                                      |
|       | H   | 8649.0                                 |       | H   | -                                      |          | H   | -                                      |
| Total |     | 35815.2                                | Total |     | 41486.1                                | Total    |     | 59546.4                                |
| Total |     |                                        |       |     |                                        | 136847.7 |     |                                        |

Table 6.1: The integrated luminosities by run era for the MET primary data set.

The search in this thesis uses  $137 \text{ fb}^{-1}$  of data collected by the CMS experiment from 2016 to 2018, nearly 40 times larger than the collision data delivered in 2015, and increasing each year, as seen in Figure 6.3. The LHC did not run for three years straight, but instead was broken up into run eras: seven run eras in 2016, five in 2017, and four in 2018. The integrated luminosity for each run era is given in Table 6.1. The data are examined by physics object groups to ensure they meet quality standards: checks include the consistency of run conditions, detector hardware statuses, and validating calibration results. If a portion of the run data (called a “block”) does not meet the necessary quality standards, it is removed. This is why CMS recorded a different integrated luminosity in Figure 6.3 than the LHC delivered, and also why not all of the data recorded is used for this analysis. The certified data is measured using the BRIL Work Suite [98], and corresponds to  $35.9 \text{ fb}^{-1}$  in 2016,  $41.5 \text{ fb}^{-1}$  in 2017, and  $59.5 \text{ fb}^{-1}$  in 2018.

Primary data sets are blocks of data with similar event selection requirements, and are col-



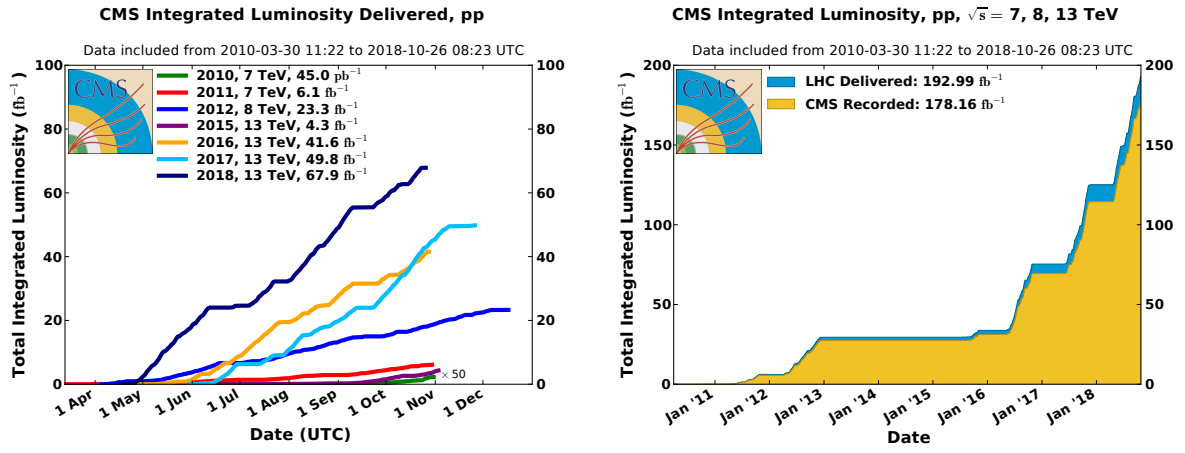


Figure 6.3: Integrated luminosity delivered by year (left) and cumulatively recorded by CMS (right) [99].

lected requiring certain sets of HLT paths. The primary data sets include the list of PF candidates, particle tracks, ECAL and HCAL clusters, vertices, and higher-level physics objects, as described in Chapter 5. For this analysis, we use the MET, SINGLEMUON, SINGLEELECTRON, and SINGLEPHOTON primary data sets. For 2018, we use the EGAMMA data set, which is a combination of the SINGLEELECTRON and SINGLEPHOTON data sets. The MET data set is used for the search region of this analysis; all others are used for validation purposes. As such, the integrated luminosities listed in Table 6.1 are for the MET primary data set; the integrated luminosities for the other primary data sets are typically within 1% of those listed.

The data for the analysis region was obtained using triggers that require selected events to have at least 90 GeV of  $p_T^{\text{miss}}$  calculated at the HLT. The threshold is tightened as the instantaneous luminosity increases during the different run periods. Further information on the trigger system can be found in Section 4.7.

The  $p_T^{\text{miss}}$  calculated at the HLT has poorer resolution than the  $p_T^{\text{miss}}$  reconstructed offline. Trigger efficiencies, defined as the fraction of events that pass the triggers after certain criteria are applied, are measured in data and applied as weights to simulated events to account for the difference in  $p_T^{\text{miss}}$  resolution. A single-electron reference trigger, chosen to be independent of our search region, is used to measure the trigger efficiencies. This reference trigger requires exactly one electron with  $p_T < 35$  GeV, at least three AK4 jets, and high  $\Delta\phi$  between the jets and  $\vec{p}_T^{\text{miss}}$ . The trigger efficiency shown in Figure 6.4 is different per data-taking year. As such, the trigger efficiencies are derived per year and applied accordingly to the simulated samples. The trigger efficiency reaches a plateau efficiency of over 99% for  $p_T^{\text{miss}} > 260$  GeV, meaning the resolved signature operates partly on the turn-on section of the efficiency curve, since it requires  $p_T^{\text{miss}} > 150$  GeV.

Although the reference trigger is unbiased and the effect is largely negligible to the analysis, the electron presence in the trigger measurement region changes the measured trigger efficiencies, in comparison to the all-hadronic case of our analysis. A systematic uncertainty is assigned to account for this trigger inefficiency. Additional trigger effects, including the statistical uncertainty

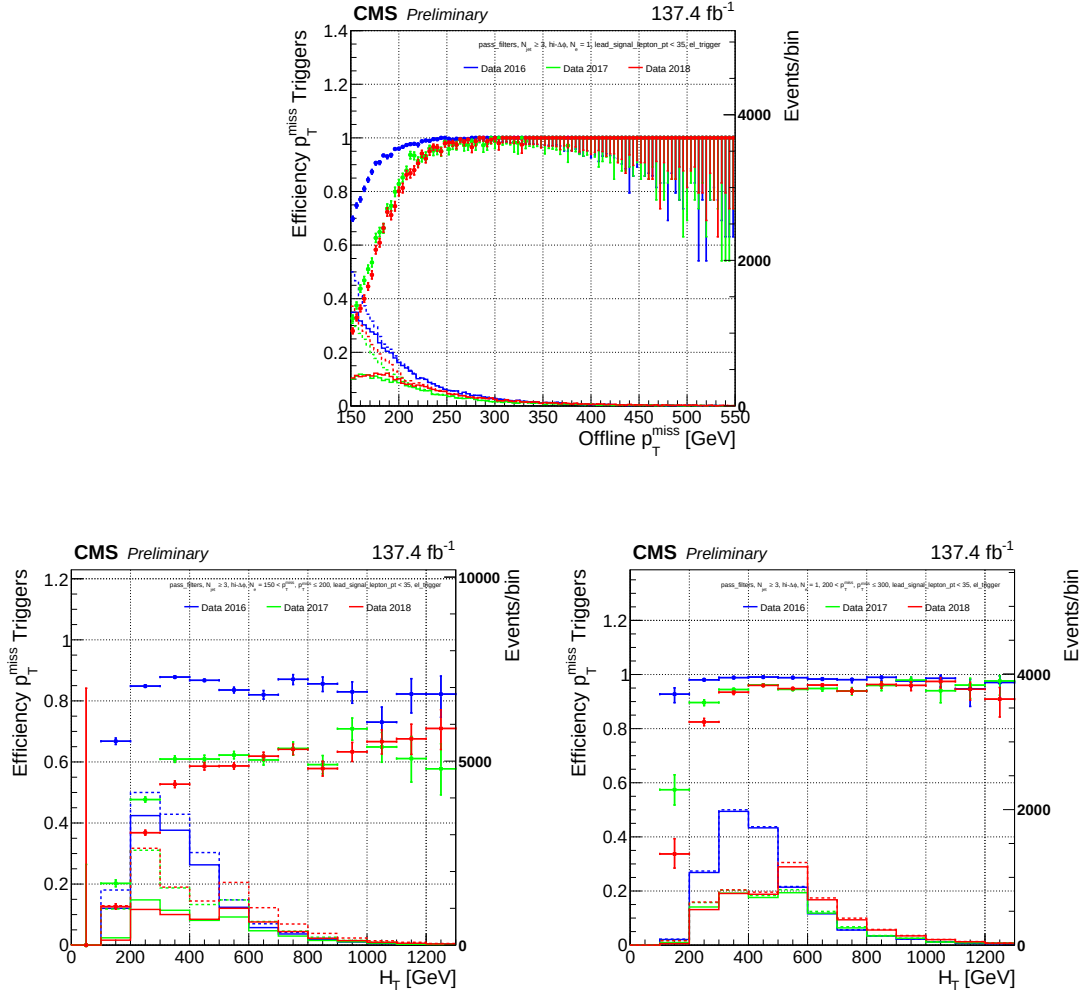


Figure 6.4: Trigger efficiencies for the search region as a function of  $p_T^{\text{miss}}$  (top),  $H_T$  for  $150 < p_T^{\text{miss}} < 200$  GeV (bottom left), and  $H_T$  for  $200 < p_T^{\text{miss}} < 300$  GeV (bottom right), for 2016 (blue), 2017 (green), and 2018 data (red). Events for the trigger efficiency measurement are selected with a single-electron trigger, with the additional requirements of high  $\Delta\phi$  and at least three AK4 jets.

and the kinematic differences between the reference and signal triggers, are included and added in quadrature.

### 6.2.2 Simulation

To complement our data-driven approach, we use simulated MC samples to optimize and to validate our background prediction methods, and to generate any necessary correction factors or systematic uncertainties of those methods. Additionally, we generate signal models, described in the Section 6.1, to interpret the observed data.

For both SM and signal samples, we use the MADGRAPH [100] or POWHEG [101] event generators to simulate the hard interaction and PYTHIA [102] for parton showering and fragmentation. GEANT [103] is used to model the full detector response for the SM background samples and the gluino T5HH signal sample. Because the detector simulation can be computationally taxing, we use a simplified description of the detector, called FastSim [104], to simulate the electroweak models.

To model the presence of pileup, samples are generated with a reasonable number of interactions per bunch crossing, depending on the instantaneous luminosity. The number of pp interactions is then reweighted at analysis-level to match the pileup distribution observed in data. All events are generated using the NNPDF 3 [105] set of parton distribution functions (PDF), which are used to model the quark and gluon momentum distributions within the proton.

#### 6.2.2.1 Standard Model simulation

SM processes are simulated to optimize our event selection criteria, create and validate our data-driven background estimation methods, and to develop any corrections or scale factors (SF) needed to match the data. These samples are generated for 2016, 2017, and 2018 separately, since the detector conditions are different per year. For quantities measured in MC to not be limited by the statistical precision relative to the statistical uncertainty in data, we aim to generate samples with much higher event counts than the data. These samples are then weighted to the appropriate luminosity, depending on the year. The size of the simulated samples is limited by the

computational time and the storage capacity for the samples after generation.

The production of  $t\bar{t}$ , W, Z, Drell-Yan,  $\gamma$ , and QCD processes listed in Table 6.2 are simulated at leading order (LO), with the exception of inclusive Drell-Yan in 2017. These samples make up the majority of the expected SM background or are used to validate the background methods. The rightmost column shows the integrated luminosity of the samples, which is dependent on the production cross section  $\sigma$  and the number of events generated. The W, Z, Drell-Yan,  $\gamma$ , and QCD samples are generated in ranges of  $H_T$ , defined as the scalar sum of  $p_T$  for all AK4 jets in an event. For  $t\bar{t}$ +jets, W+jets, and Drell-Yan processes, we also use samples with larger event counts for high amounts of generator-level  $p_T^{\text{miss}}$  or  $H_T$ . The generator-level  $p_T^{\text{miss}}$  selection for  $t\bar{t}$  events is listed in the second column of Table 6.2, with the lower value  $p_T^{\text{miss}} > 80$  GeV used for the 2018 samples, and the higher value ( $p_T^{\text{miss}} > 150$  GeV) used for 2016 and 2017. Events are removed from the inclusive samples in order to avoid double counting. Single-top quark events and rare SM backgrounds, such as  $t\bar{t}$  production in association with bosons, diboson processes, and  $t\bar{t}t\bar{t}$ , shown in Table 6.3, are produced at next-to-leading order (NLO).

#### 6.2.2.2 Signal simulation

Simplified signal models are useful for interpreting the data observed in analyses. In this analysis, we target two types of SUSY production:

- electroweak pair production, and
- gluino pair production.

Signal events for the simplified models are generated in the same manner as the SM backgrounds, with the MADGRAPH generator in LO mode, using the NNPDF 3 PDF set, and followed with the PYTHIA generator for showering and fragmentation. We use a simplified description of the detector called FastSim [104] to simulate the electroweak models. We generate scale factors to account for any differences with respect to the full simulation used for the backgrounds. Additionally, the electroweak models are generated assuming the H bosons only decay to  $b\bar{b}$ , the largest branching

| Process                                                   | Generator Selection          | $\sigma$ [pb] |          |          | $\int \mathcal{L}$ [fb <sup>-1</sup> ] |           |          |
|-----------------------------------------------------------|------------------------------|---------------|----------|----------|----------------------------------------|-----------|----------|
|                                                           |                              | 2016          | 2017     | 2018     | 2016                                   | 2017      | 2018     |
| $t\bar{t}, t \rightarrow \ell\nu, \bar{t} \rightarrow 2q$ | -                            | 182.16        | 182.37   | 182.34   | 338.27                                 | 338.66    | 314.03   |
| $t\bar{t}, \bar{t} \rightarrow \ell\nu, t \rightarrow 2q$ | -                            | 182.16        | 182.36   | 182.34   | 331.26                                 | 310.95    | 328.67   |
| $t\bar{t}, 2\ell$                                         | -                            | 87.34         | 87.43    | 87.42    | 353.06                                 | 324.59    | 328.31   |
| $t\bar{t}, t \rightarrow \ell\nu, \bar{t} \rightarrow 2q$ | $p_T^{\text{miss}} > 80/150$ | 9.68          | 10.43    | 54.11    | 1792.22                                | 1377.87   | 1555.97  |
| $t\bar{t}, \bar{t} \rightarrow \ell\nu, t \rightarrow 2q$ | $p_T^{\text{miss}} > 80/150$ | 9.68          | 10.40    | 53.36    | 1755.90                                | 791.73    | 1489.24  |
| $t\bar{t}, 2\ell$                                         | $p_T^{\text{miss}} > 80/150$ | 5.919         | 6.154    | 37.71    | 1647.82                                | 1415.12   | 1549.83  |
| $W \rightarrow \ell\nu_\ell$                              | -                            | 61527         | 61527    | 61527    | 1.41                                   | 1.26      | 1.15     |
|                                                           | $70 \leq H_T \leq 100$       | 1656.07       | 1501.30  | 1503.89  | 6.05                                   | 14.80     | 18.71    |
|                                                           | $100 \leq H_T \leq 200$      | 1647.50       | 1612.26  | 1611.90  | 6.04                                   | 22.24     | 18.31    |
|                                                           | $200 \leq H_T \leq 400$      | 440.76        | 476.23   | 478.94   | 88.45                                  | 44.62     | 53.18    |
|                                                           | $400 \leq H_T \leq 600$      | 59.73         | 67.69    | 68.08    | 129.91                                 | 211.45    | 87.15    |
|                                                           | $600 \leq H_T \leq 800$      | 14.77         | 16.85    | 16.93    | 1264.92                                | 1288.53   | 1167.77  |
|                                                           | $800 \leq H_T \leq 1200$     | 6.73          | 7.55     | 7.66     | 1163.82                                | 2704.93   | 1096.45  |
|                                                           | $1200 \leq H_T \leq 2500$    | 1.627         | 1.685    | 1.696    | 4224.79                                | 12026.50  | 4500.90  |
|                                                           | $H_T \geq 2500$              | 0.0393        | 0.0407   | 0.0404   | 67157.45                               | 528312.40 | 80997.72 |
| $Z \rightarrow \nu\bar{\nu}$                              | $100 \leq H_T \leq 200$      | 343.2885      | 342.8630 | 345.4666 | 1414.29                                | 1507.93   | 1457.49  |
|                                                           | $200 \leq H_T \leq 400$      | 95.11         | 103.59   | 102.89   | 260.35                                 | 209.26    | 226.22   |
|                                                           | $400 \leq H_T \leq 600$      | 13.14         | 14.71    | 14.67    | 750.66                                 | 620.98    | 648.42   |
|                                                           | $600 \leq H_T \leq 800$      | 3.134         | 3.633    | 3.627    | 1840.22                                | 1568.33   | 1585.19  |
|                                                           | $800 \leq H_T \leq 1200$     | 1.444         | 1.565    | 1.574    | 1502.43                                | 1315.27   | 1313.04  |
|                                                           | $1200 \leq H_T \leq 2500$    | 0.3531        | 0.3469   | 0.3472   | 1454.34                                | 976.97    | 988.59   |
|                                                           | $H_T \geq 2500$              | 0.0085        | 0.0078   | 0.0079   | 47627.32                               | 863.09    | 45676.51 |
| $Z/\gamma^* \rightarrow \ell^+\ell^-$                     | -                            | 6020.85       | 6225.42  | 6225.42  | 8.26                                   | 33.38     | 16.09    |
|                                                           | $70 \leq H_T \leq 100$       | 215.62        | 171.22   | 171.22   | 44.95                                  | 54.57     | 58.52    |
|                                                           | $100 \leq H_T \leq 200$      | 171.46        | 188.06   | 188.06   | 64.25                                  | 59.54     | 61.31    |
|                                                           | $200 \leq H_T \leq 400$      | 52.58         | 56.51    | 56.51    | 182.74                                 | 189.99    | 192.33   |
|                                                           | $400 \leq H_T \leq 600$      | 6.761         | 8.046    | 8.046    | 1438.43                                | 1270.20   | 1205.27  |
|                                                           | $600 \leq H_T \leq 800$      | 1.676         | 1.967    | 1.967    | 4946.62                                | 4444.08   | 4504.29  |
|                                                           | $800 \leq H_T \leq 1200$     | 0.8314        | 0.8556   | 0.8556   | 3215.29                                | 3640.82   | 3667.88  |
|                                                           | $1200 \leq H_T \leq 2500$    | 0.1427        | 0.1926   | 0.1926   | 4177.73                                | 3247.92   | 2785.27  |
|                                                           | $H_T \geq 2500$              | 0.0032        | 0.0027   | 0.0027   | 125305                                 | 153127    | 155954   |
| $\gamma$ +jets                                            | $100 \leq H_T \leq 200$      | 5363.00       | 5045.08  | 5030.57  | 2.76                                   | 3.16      | 3.07     |
|                                                           | $200 \leq H_T \leq 400$      | 1178.00       | 1130.82  | 1125.44  | 42.45                                  | 44.13     | 43.94    |
|                                                           | $400 \leq H_T \leq 600$      | 131.80        | 124.81   | 124.81   | 88.79                                  | 107.28    | 109.91   |
|                                                           | $H_T \geq 600$               | 44.27         | 40.80    | 40.54    | 263.99                                 | 205.69    | 307.26   |
| QCD                                                       | $100 \leq H_T \leq 200$      | 28060000      | 23700000 | 23700000 | 0.0029                                 | 0.0039    | 0.0040   |
|                                                           | $200 \leq H_T \leq 300$      | 1710000       | 1548751  | 1558084  | 0.03                                   | 0.04      | 0.03     |
|                                                           | $300 \leq H_T \leq 500$      | 347500        | 323196   | 323761   | 0.16                                   | 0.07      | 0.17     |
|                                                           | $500 \leq H_T \leq 700$      | 32060         | 30069    | 30193    | 1.95                                   | 1.87      | 1.83     |
|                                                           | $700 \leq H_T \leq 1000$     | 6829.00       | 6361.49  | 6325.76  | 2.29                                   | 3.15      | 7.61     |
|                                                           | $1000 \leq H_T \leq 1500$    | 1207.00       | 1095.27  | 1098.15  | 12.60                                  | 15.41     | 14.08    |
|                                                           | $1500 \leq H_T \leq 2000$    | 120.00        | 100.19   | 99.99    | 98.66                                  | 116.12    | 109.56   |
|                                                           | $H_T \geq 2000$              | 25.25         | 20.63    | 20.43    | 238.40                                 | 287.97    | 268.05   |

Table 6.2: Standard model simulated samples. The sample cross section  $\sigma$  and an equivalent sample integrated luminosity,  $\int \mathcal{L}$ , is given for each year. For the three  $t\bar{t}$  processes with a  $p_T^{\text{miss}}$  generator selection, the first corresponds to the 2018 samples, while the second corresponds to the selection imposed in both 2016 and 2017.

| Process                 | Generator Selection                                           | $\sigma$ [pb] |          |          | $\int \mathcal{L}$ [fb <sup>-1</sup> ] |         |         |
|-------------------------|---------------------------------------------------------------|---------------|----------|----------|----------------------------------------|---------|---------|
|                         |                                                               | 2016          | 2017     | 2018     | 2016                                   | 2017    | 2018    |
| t                       | s-channel                                                     | 5.401         | 5.996    | 5.993    | 185.14                                 | 1652.35 | 3331.19 |
| t                       | t-channel                                                     | 136.02        | 113.30   | 123.47   | 493.35                                 | 51.77   | 1249.74 |
| $\bar{t}$               | t-channel                                                     | 80.95         | 67.91    | 73.62    | 479.44                                 | 58.02   | 1074.35 |
| t                       | tW-channel                                                    | 20.22         | 18.67    | 18.61    | 161.08                                 | 251.87  | 58.34   |
| $\bar{t}$               | tW-channel                                                    | 20.20         | 18.71    | 18.65    | 161.20                                 | 288.12  | 58.27   |
| t $\bar{t}$ Z           | $Z \rightarrow \ell^+ \ell^-$ , $Z \rightarrow \nu \bar{\nu}$ | 0.2529        | 0.2432   | 0.2432   | 54142.5                                | 32617.8 | 54605.3 |
|                         | $Z \rightarrow q \bar{q}$                                     | 0.5297        | 0.5104   | 0.5104   | 1414.76                                | 18985.1 | 18753.9 |
| t $\bar{t}$ W           | $W \rightarrow \ell \nu_\ell$                                 | 0.2043        | 0.2043   | 0.2043   | 25847.1                                | 24447.1 | 24042.8 |
|                         | $W \rightarrow q \bar{q}$                                     | 0.4062        | 0.4064   | 0.4062   | 2051.45                                | 1997.31 | 2056.37 |
| t $\bar{t}$ H           | $H \rightarrow b \bar{b}$                                     | 0.2934        | 0.2934   | 0.2934   | 33556.9                                | 27266.5 | 32324.5 |
| t $\bar{t}$ $\gamma$    | -                                                             | 3.697         | 3.697    | 3.697    | 3989.41                                | 1987.86 | 1269.11 |
| t $\bar{t}$ t $\bar{t}$ | -                                                             | 0.009103      | 0.008213 | 0.008213 | 27463.5                                | 121758  | 282662  |
| WH                      | $W \rightarrow \ell \nu_\ell$ , $H \rightarrow b \bar{b}$     | 0.2616        | 0.2616   | 0.2616   | 8332.14                                | 3732.42 | 3719.42 |
| WW                      | $WW \rightarrow \ell \nu_\ell q \bar{q}$                      | 49.997        | 49.997   | 49.997   | 39.99                                  | 175.66  | 384.01  |
|                         | $WW \rightarrow \ell \nu_\ell \ell \nu_\ell$                  | 12.18         | 12.18    | 12.18    | 164.15                                 | 164.23  | 637.12  |
| WZ                      | $W \rightarrow \ell \nu_\ell$ , $Z \rightarrow q \bar{q}$     | 10.73         | 11.66    | 11.76    | 5899.91                                | 1629.92 | 4919.00 |
|                         | $W \rightarrow \ell \nu_\ell$ , $Z \rightarrow \nu \bar{\nu}$ | 3.054         | 3.294    | 3.322    | 557.88                                 | 1509.79 | 510.90  |
|                         | $W \rightarrow \ell^+ \ell^-$ , $Z \rightarrow q \bar{q}$     | 5.595         | 5.606    | 5.606    | 4726.68                                | 4920.11 | 5029.19 |
|                         | $W \rightarrow \ell \nu_\ell$ , $Z \rightarrow \ell^+ \ell^-$ | 4.430         | 4.430    | 4.430    | 449.97                                 | 2480.48 | 446.22  |
| ZH                      | $Z \rightarrow \nu \bar{\nu}$ , $H \rightarrow b \bar{b}$     | 0.1029        | 0.1029   | 0.1029   | 28421.6                                | 25033   | 44375.4 |
| ZZ                      | -                                                             | 16.523        | 16.523   | 16.523   | 120.323                                | 118.003 | 119.772 |

Table 6.3: Single-top, t $\bar{t}$  +X, and diboson simulated samples. The sample cross section  $\sigma$  and an equivalent sample integrated luminosity,  $\int \mathcal{L}$ , is given for each year.

| Decay Mode                                                                                                                              | SMS name | Generator | Mass Scan [GeV]                                                                                                                               |
|-----------------------------------------------------------------------------------------------------------------------------------------|----------|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| $\tilde{\chi}_2^0 \tilde{\chi}_3^0 \rightarrow \text{HH} \tilde{\chi}_1^0 \tilde{\chi}_1^0$                                             | TChiHH   | FASTSIM   | $m_{\tilde{\chi}_3^0} = m_{\tilde{\chi}_2^0},$<br>$m_{\tilde{\chi}_2^0} = 127 - 1200,$<br>$m_{\tilde{\chi}_1^0} = m_{\tilde{\chi}_2^0} - 125$ |
| $\tilde{\chi}_i^{0,\pm} \tilde{\chi}_j^{0,\mp} \rightarrow \tilde{\chi}_1^0 \tilde{\chi}_1^0 \rightarrow \text{HH} \tilde{G} \tilde{G}$ | TChiHH-G | FASTSIM   | $m_{\tilde{\chi}_1^0} = 127 - 1500,$<br>$m_{\tilde{G}} = 1$                                                                                   |
| $\tilde{g}\tilde{g}, \tilde{g} \rightarrow q\bar{q}\tilde{\chi}_2^0, \tilde{\chi}_2^0 \rightarrow \text{H}\tilde{\chi}_1^0$             | T5HH     | FULLSIM   | $m_{\tilde{g}} = 1000 - 2500,$<br>$m_{\tilde{\chi}_2^0} = m_{\tilde{g}} - 50,$<br>$m_{\tilde{\chi}_1^0} = 1$                                  |

Table 6.4: Signal samples used in the analysis. All decay modes begin with a proton-proton interaction. The gluino pair production model, T5HH, is generated so that the quarks produced are soft and do not normally register.

fraction (58%) [88]. The gluino model is generated using the full detector simulation and includes all H decays. Further details on the signal samples are discussed in Section 6.1, and summarized in Table 6.4.

### 6.3 Event selection

The simplified SUSY models used in this thesis produce b jets and large transverse momentum imbalance. Since our models have fully-hadronic final states, we remove events that have isolated electrons or muons. Because of this, and technical considerations such as filters, corrections, and trigger requirements, we optimize our selection criteria to find these events.

#### 6.3.1 Baseline selection

The baseline selection is the set of search criteria imposed on our events, in order to pick out events compatible with our SUSY models. The baseline selection is slightly different between the boosted and resolved signatures, since they target slightly different topologies (but the same final state). For both signatures, we reconstruct two potential Higgs candidates, each containing two b quarks. Figure 6.5 shows a signal event (left) and a  $t\bar{t}$  single-lepton background event (right), the dominant background in this analysis. In signal events, four b jets are produced from the two Higgs



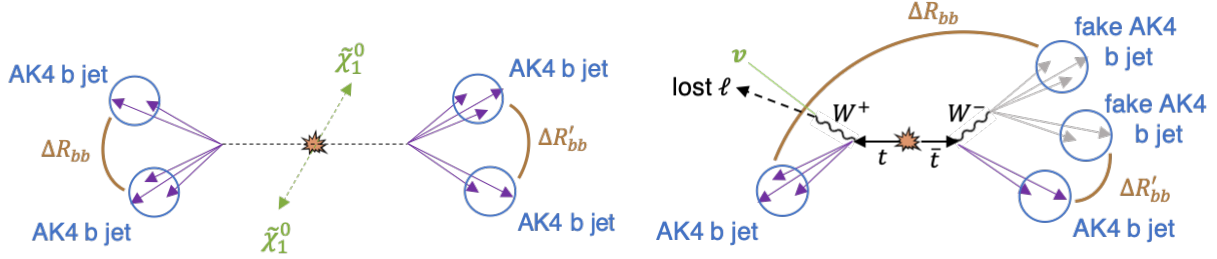


Figure 6.5: Configuration of signal (left) and  $t\bar{t}$  (right) events, our dominant background, shown here for how extra b jets could arise in the resolved signature.

decays, while the top quark decay in  $t\bar{t}$  events only produces two b jets. Additional b jets found in background events can arise from mistagging or from gluon radiation, either initial-state radiation (ISR) or final-state radiation (FSR), followed by  $g \rightarrow b\bar{b}$ . The neutrino produced in leptonic decays provides the needed  $\vec{p}_T^{\text{miss}}$ . Additional SM backgrounds, such as  $Z \rightarrow \nu\nu$  (in which the neutrinos result in true  $p_T^{\text{miss}}$ ) and QCD multijet events (in which the  $p_T$  of a jet is mismeasured, creating a source of fake  $p_T^{\text{miss}}$ ) can also arise.

To mitigate these backgrounds, we establish our baseline selection as follows:

(1) **Electron and muon vetoes**

Events with identified and isolated electrons or muons with  $p_T > 10$  GeV are rejected, where the leptons are PF candidates described in Section 5.2. Electrons are identified using the “cut-based veto” working point of the electron identification algorithm [106], corresponding to roughly 95% reconstruction efficiency. Muons are required to pass the medium identification working point [107]. Isolation [108] requires the  $p_T$  sum of the particles within a cone of  $\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$  to be less than 10% (20%) of the electron (muon)  $p_T$ . The radius of the cone depends on the lepton  $p_T$ :

$$\Delta R = \begin{cases} 0.2 & \text{for } p_T < 50 \text{ GeV} \\ 10 \text{ GeV}/p_T & \text{for } 50 \leq p_T \leq 200 \text{ GeV} \\ 0.05 & \text{for } p_T > 200 \text{ GeV} \end{cases} \quad (6.1)$$

The isolation requirement is put in place to avoid removing events with muons from b

hadron decays, and to avoid rejecting events with electrons only from bremsstrahlung radiation.

## (2) Angular requirement

To suppress QCD background, events are rejected if any of the four highest- $p_T$  AK4 jets are approximately aligned with  $\vec{p}_T^{\text{miss}}$ , indicating the observed  $\vec{p}_T^{\text{miss}}$  may be from undermeasuring the  $p_T$  of the jet, and therefore is not genuine. Specifically, we require  $\Delta\phi_{1,2,3,4} > 0.5, 0.5, 0.3, 0.3$ , where  $\Delta\phi_i \equiv \Delta\phi(p_T^{\text{miss}}, \text{AK4jet}_i)$ . A relatively loose selection on  $\Delta\phi$  for high momentum jets provides excellent background rejection, while preserving high signal efficiency. For lower- $p_T$  jets, the discrimination power begins to fade. We collectively refer to this set of cuts as the  $\Delta\phi$  requirement.

## (3) Missing transverse energy

We require PF  $p_T^{\text{miss}}$  (Section 5.3.3)  $> 150$  GeV, to improve signal efficiency while rejecting QCD background. The boosted signature has a higher  $p_T^{\text{miss}}$  requirement,  $p_T^{\text{miss}} > 300$  GeV, to ensure a boosted topology where the Higgs decay products are captured in a single AK8 jet.

## (4) Isolated track vetoes

To suppress backgrounds with a  $\tau$  lepton or an electron or muon that is not identified or fails the lepton veto criteria, we reject events with one or more isolated charged tracks. Leptonic tracks must satisfy  $p_T > 5$  GeV and have  $<20\%$  isolation within a cone of radius  $\Delta R = 0.3$ , where isolation is the scalar  $p_T$  sum of all charged-hadron, neutral-hadron, and photon PF candidates in the cone divided by the track  $p_T$ . Hadronic tracks must satisfy  $p_T > 10$  GeV and have  $<10\%$  isolation. To retain more signal, events are only vetoed if the isolated track is consistent with originating from a W boson. We require  $m_T(\text{trk}, p_T^{\text{miss}}) < 100$  GeV, where

$$m_T(\text{trk}, p_T^{\text{miss}}) \equiv \sqrt{2p_T^{\text{trk}} p_T^{\text{miss}} (1 - \cos\Delta\phi)}, \quad (6.2)$$

where  $\Delta\phi$  is the angular separation between the track and  $\vec{p}_T^{\text{miss}}$ .

(5) **Event cleaning:**

The following recommended filters are applied to reject events with anomalous sources of  $p_T^{\text{miss}}$ . Unless otherwise indicated, the filters are applied only to the full simulation samples and the data.

- **BEAM HALO:** Particles produced upstream form a beam halo, that travel through the detector along the beam axis. If any CSC hits or ECAL clusters are consistent with certain halo shapes, studied from out-of-time hits, the event is discarded.
- **HBHE NOISE and HBHEISO NOISE:** In the HB and HE, scintillator tiles are prone to noise that may register as energy deposits. Geometric patterns, channel multiplicities, and the pulse shape are used to determine which events should be removed.
- **ECAL DEAD CELLS:** ECAL trigger primitives are bounded and can saturate, leading to dead cells. Events in which the trigger primitives are at or near threshold are flagged, since the dead cells can lead to significant bias in measuring particle  $p_T$ .
- **MUON JETS:** Failures in muon reconstruction can create exaggerated momentum imbalance. We reject events if any jet has  $p_T > 200$  GeV, muon energy fraction  $> 0.5$ , and  $\Delta\phi(\text{jet}, \vec{p}_T^{\text{miss}}) > \pi - 0.4$ .
- **LOW NEUTRAL JETS:** We remove events with anomalously low neutral EM energy fraction, specifically if the leading- $p_T$  AK4 jet has neutral EM energy fraction  $< 0.03$  and  $\Delta\phi(\text{jet}, \vec{p}_T^{\text{miss}}) > \pi - 0.4$ .
- **ONE GOOD VERTEX:** Events without a well-defined vertex often have anomalous  $p_T^{\text{miss}}$ . We require one reconstructed vertex that satisfies certain position and quality criteria, as described in Section 5.1.1. This filter is also applied to the signal samples that are generated using the fast simulation.
- **EE BAD SUPERCLUSTER NOISE:** We remove events that contain noise from bad crystals in the endcaps of the ECAL. This filter is only applied to data.

- $p_T^{\text{miss}}$  RATIO: To protect against particle flow failures, which can result in a large discrepancy between  $p_T^{\text{miss}}$  calculated with particle flow (PFMET) and  $p_T^{\text{miss}}$  calculated with only calorimeter information (CaloMET), we require:  $\text{PFMET}/\text{CaloMET} < 2$ . Additionally, we reject events where the PF  $p_T^{\text{miss}}$  is more than two times larger than the  $H_T^{\text{miss}}$ , where  $H_T^{\text{miss}}$  is the negative vector sum of all jets in an event. This filter is also applied to the signal samples that are generated using the fast simulation.
- PF JET ID: We remove AK4 and AK8 jets from an event if they have  $|\eta| < 2.5$ , charged-hadron fraction less than 0.1, and  $p_T > 20$  GeV (or  $p_T > 170$  GeV for AK8 jets). Jets such as these are usually from detector noise and can effect our  $\Delta\phi$  selection, and the  $H_T$  and  $H_T^{\text{miss}}$  of an event.
- ECAL NOISE IN JETS: In the 2017 data, increased calorimeter noise in the endcaps of the detector resulted in artificial  $p_T^{\text{miss}}$ . Even after correcting the  $p_T^{\text{miss}}$ , some events with noisy jets and induced  $p_T^{\text{miss}}$  remain. We remove events if either of the two leading- $p_T$  jets with  $2.4 < |\eta| < 5$  have  $p_T > 250$  GeV and  $\Delta\phi(\text{jet}, \vec{p}_T^{\text{miss}}) > 2.6$  or  $< 0.1$ .
- $H_T$ -RATIO: Events with noisy energetic jets in the HF are observed between  $3.0 < \eta < 3.1$ . We reject events if they have  $H_T^5/H_T > 1.2$  or  $\Delta\phi \geq 5.3 H_T^5/H_T - 4.78$ , where  $H_T^5$  is the  $H_T$  calculated including jets up to  $|\eta| = 5$ , while  $H_T$  uses only jets up to  $|\eta| = 2.4$ .  $\Delta\phi$  is between the leading- $p_T$  AK4 jet and  $\vec{p}_T^{\text{miss}}$ .
- HEM VETO: The HEM issue in some runs of 2018 may generate spurious jets, electrons, and  $p_T^{\text{miss}}$ . Specifically, an HCAL submodule broke, and the region of the detector it covered ( $-3.2 \leq \eta \leq -1.3$  and  $-1.57 \leq \eta \leq -1.3$ ) was entirely depleted of energy deposits. This led to a significant number of reconstructed isolated electrons (HCAL information is used to discriminate real isolated electrons that leave clusters in the ECAL from jets that leave clusters in the HCAL), that were triggered on. Additionally, the reconstruction of jet energy relies heavily on HCAL clusters. Events are vetoed if any isolated electrons are in the region  $-3.0 < \eta < -1.4$  and  $-1.57 < \phi < -0.87$ ,

or if any AK4 jets are in the region  $-3.2 < \eta < -1.2$  or  $-1.77 < \eta < -0.67$ . Since this effect is not modeled in our simulated samples, we remove a proportionate fraction of events in MC.

The resolved and boosted methods each have further requirements, as discussed below, for the reconstruction of the Higgs boson candidates.

### 6.3.2 Higgs candidate reconstruction

For the resolved signature, we require 4 or 5 AK4 jets in the event, each with  $p_T > 30$  GeV. The possibility of a fifth jet increases signal sensitivity without greatly increasing the combinatorics for the Higgs candidate selection. Out of these, we choose the four AK4 jets with the highest b-tagging scores, the four most b-like jets. We require at least two of these jets to pass the tight working-point (WP) of the DEEPCSV algorithm (see Section 5.3.2). These four jets are grouped into three possible pairings of Higgs boson candidates (bb-bb), and we select the pairing with the smallest mass difference  $\Delta m_{bb}$  between the two Higgs candidates  $m_{(bb)_1}, m_{(bb)_2}$ :

$$\Delta m_{bb} = |m_{(bb)_1} - m_{(bb)_2}|. \quad (6.3)$$

This method uses the fact that signal events contain two particles of identical mass, without using the value of the mass itself. Methods that use the known mass to select the best candidate tend to sculpt an artificial peak in the background. Only events where the masses of the two Higgs boson candidates are similar,  $\Delta m_{bb} < 40$  GeV, are kept. We then calculate the average mass as:

$$\langle m_{bb} \rangle = \frac{1}{2} (m_{(bb)_1} + m_{(bb)_2}), \quad (6.4)$$

and require  $\langle m_{bb} \rangle < 200$  GeV. Lastly, we require the angular separation between the two b-jets of each Higgs candidate to be small,  $\Delta R_{\max} < 2.2$ . This criteria removes a large portion of  $t\bar{t}$  events, where the  $\Delta R$  for at least one bb-pairing is typically much larger than in signal events, as shown in Figure 6.5.

For the boosted signature, we require  $p_T^{\text{miss}} > 300$  GeV and two AK8 jets, each with  $p_T > 300$  GeV. The  $p_T$  requirement ensures that bosons with mass  $\sim 125$  GeV will have both decay products captured within a single AK8 jet:  $p_T \approx 2m_H/R$ , where  $R$  is the radius of the jet (see Section 5.3.1). We select the two AK8 jets with the highest  $p_T$  as our Higgs candidates, and require them to have a jet mass (derived using the soft-drop mass algorithm [109, 110]) between 60 and 260 GeV. By targeting boosted topologies, we significantly reduce SM background rates. In particular, since most SM processes do not produce hadronically decaying bosons, the jet masses will typically be below our baseline selection of 60 GeV, even if they have large  $p_T$ .

After the baseline selection is applied and the Higgs candidates are reconstructed, each of the two signatures further bins the events into mass signal regions (SR), where we expect the most signal, and mass sideband regions (SB), which are background-dominated and used to predict the background of the signal regions. These bins are further divided into different b- or bb-tagging categories and other kinematic variables, as described in Section 6.4.1, and are used to estimate the Standard Model background of the analysis. The combined selection retains roughly 10% of signal events in the signal bins.

## 6.4 Background estimation

The Standard Model (SM) background yields are determined using a data-driven approach that uses two discriminating and uncorrelated variables to estimate the yield in a signal region. The benefit of a data-driven approach is that we can be confident the results we see are real, and not an effect of poor or under-performing simulation.

For this analysis, these variables are the Higgs boson candidate mass  $m_H$  and the b-tagging discriminators, which are different between the boosted and resolved methods. This method, termed the “ABCD” method, estimates the background in a given signal region ( $A$ ) using the event yields observed in three neighboring regions: the mass sideband region of the same tagged category as the signal region ( $B$ ), and the control regions where events fall within the Higgs mass window ( $C$ ) and the mass sidebands ( $D$ ) but fail the tagging requirement. The control regions  $C$  and  $D$  are expected

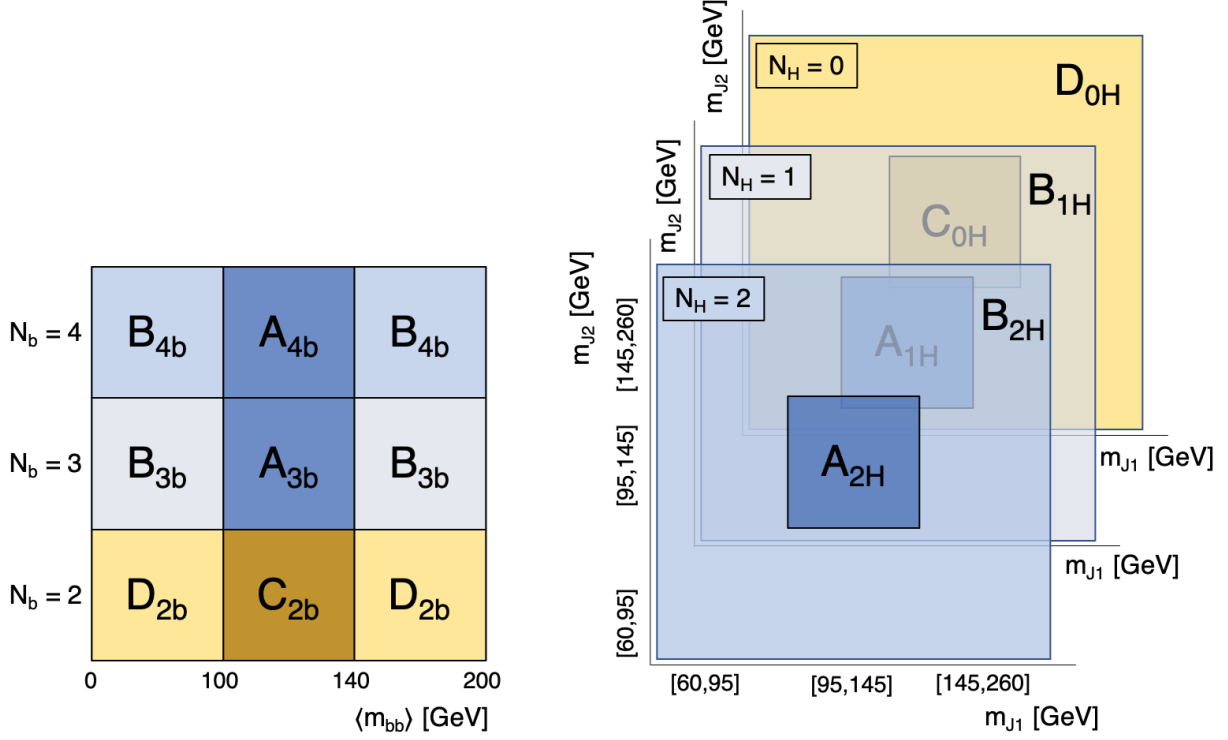


Figure 6.6: The Higgs mass window and mass sideband regions used for the resolved (left) and boosted (right) background predictions, in different b- and bb-tagging categories. Further binning in other kinematic variables is applied.

to have very few signal events. These regions differ between the resolved and boosted background estimation methods, and are described further in Section 6.4.1, and shown in Figure 6.6.

#### 6.4.1 Region definition

For the resolved signature, six regions are defined in the number of b-tagged jets  $N_b$  and the mean mass of the two Higgs boson candidates  $\langle m_{bb} \rangle$ . We define the Higgs boson mass window as  $100 < \langle m_{bb} \rangle \leq 140$  GeV, and the regions on either side of this window,  $\langle m_{bb} \rangle < 100$  GeV and  $140 < \langle m_{bb} \rangle \leq 200$  GeV, as mass sidebands. The following mutually exclusive b-tagging categories are used to classify events for the resolved method:

- $N_b = 2$ , “2b”: two tight b-tags,
- $N_b = 3$ , “3b”: two tight b-tag, one medium b-tag,

- $N_b = 4$ , “4b”: two tight b-tags, one medium b-tag, one loose b-tag.

The b-tagging WPs are inclusive, meaning a jet that passes the tight WP also passes the medium and loose WPs. The  $N_b = 2$  category is used as a control region to determine the kinematic shape of the background, while the  $N_b = 3$  and 4 categories are enriched in signal events. The six regions, then, are the three  $N_b$  categories, split into the Higgs mass window and the mass sidebands. To further enhance the sensitivity of the background prediction, we subdivide these regions into four  $p_T^{\text{miss}}$  bins and two  $\Delta R_{\text{max}}$  bins, for a total of eight planes of the six bins:  $p_T^{\text{miss}}$  [150–200, 200–300, 300–400, and >400 GeV] and  $\Delta R_{\text{max}}$  [0–1.1 and 1.1–2.2]. Each  $p_T^{\text{miss}}$ - $\Delta R_{\text{max}}$  plane has two signal bins, the 3b mass SR ( $A_{3b}$ ) and the 4b mass SR ( $A_{4b}$ ). The key variables for these bins are shown in Figure 6.7, with representative signal samples overlaid to motivate region boundaries.

For the boosted signature, six regions are defined in the number of bb-tagged jets  $N_H$  and the masses of both the leading- and subleading- $p_T$  jets ( $m_{J1}, m_{J2}$ ). This two-dimensional mass plane, shown in Figure 6.8, defines the Higgs mass window (SR) and sideband regions (SB). An event falls into the mass SR if both AK8 jets have a mass within the Higgs mass window,  $95 < m_J \leq 145$  GeV. If one or both jets has a mass outside of this window, the event falls into the mass SB. Similar to the resolved method, we define the following mutually exclusive bb-tagging categories:

- $N_H = 2$ , “2H”: both Higgs candidates pass the bb-tagging criteria,
- $N_H = 1$ , “1H”: only one Higgs candidate passes the bb-tagging criteria,
- $N_H = 0$ , “0H”: neither Higgs candidate passes the bb-tagging criteria.

To bb-tag the AK8 jets, we use the loose WP ( $>0.7$ ) of a deep-learning-based double-b tagging algorithm (more in Section 5.3.2), corresponding to an efficiency of 90% for  $H \rightarrow b\bar{b}$  [84]. Most  $H \rightarrow b\bar{b}$  signal events have  $N_H = 2$ , while the  $N_H = 1$  category is enriched with other Higgs decay modes. The  $N_H = 0$  category is used as a control region to determine the kinematic shape of the background. The six regions, then, are the three  $N_H$  categories, split into the Higgs mass window and the mass sidebands. For each of the six regions, three  $p_T^{\text{miss}}$  bins are used to further separate



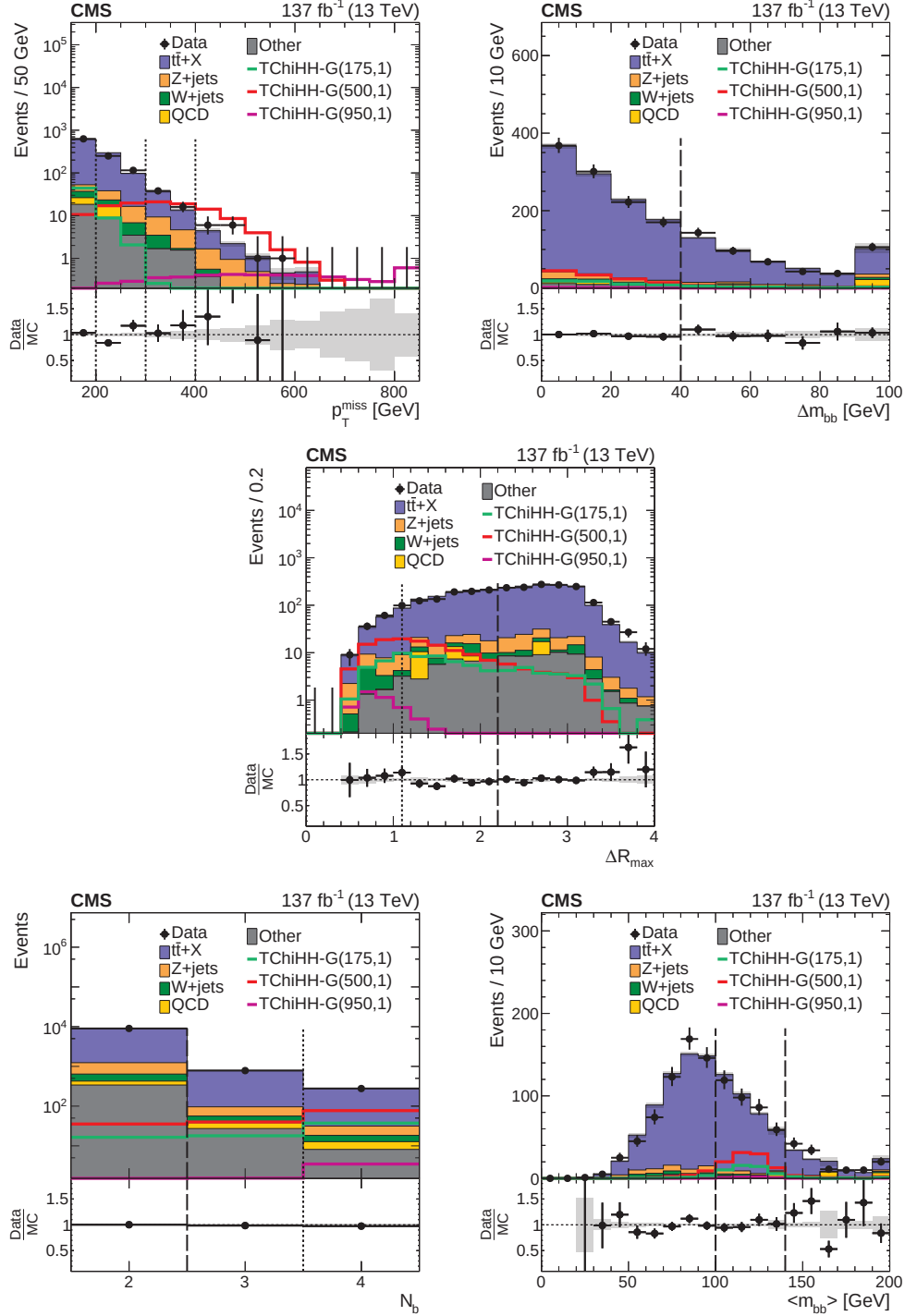


Figure 6.7: Distributions in key analysis variables of the resolved signature for events satisfying the baseline requirements and  $N_b = 3$  or 4, except for those on the variable plotted: (upper left)  $N_b$ , (upper right)  $p_T^{\text{miss}}$ , (middle left)  $\Delta m_{bb}$ , (middle right)  $\langle m_{bb} \rangle$ , and (lower)  $\Delta R_{\text{max}}$ . The SM background are scaled by factors of 0.93–0.97 to match the total data yield in each plot, and facilitate comparisons by-eye. Three signal mass points are shown to motivate region boundaries. The lower panel in each plot shows the ratio of observed data to (scaled) MC yields.

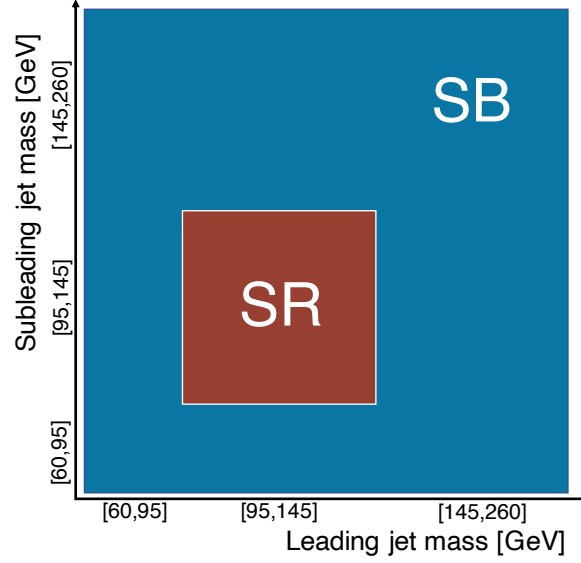


Figure 6.8: The two-dimensional mass plane that defines the Higgs mass window (red) and mass sideband (blue) regions.

signal from background,  $300 < p_T^{\text{miss}} < 500$  GeV,  $500 < p_T^{\text{miss}} < 700$  GeV, and  $p_T^{\text{miss}} > 700$  GeV. The key variables for the boosted signature, including overlaid signal, are shown in Figure 6.9.

#### 6.4.2 The ABCD method

As mentioned in the beginning of this section, the ABCD method is a data-driven approach that estimates the background yield of a signal region  $A$  by the yields in neighboring regions,  $B$ ,  $C$ , and  $D$ . The specific regions for the resolved and boosted background methods are given in Section 6.4.1.

In a standard ABCD method, the background rate in the signal regions can be estimated with the yields in the corresponding mass sideband and control regions:

$$A = B * (C/D) \quad (6.5)$$

This background prediction is sound if the two variables that define the ABCD plane are uncorrelated. The shape of the distribution of variable 1 is unchanged whether or not variable 2 is applied, and vice versa. This means the ratio of events when variable 1 is or is not applied ( $A/B$ ) should be equal to to the ratio of events when variable 2 is or is not applied ( $C/D$ ).

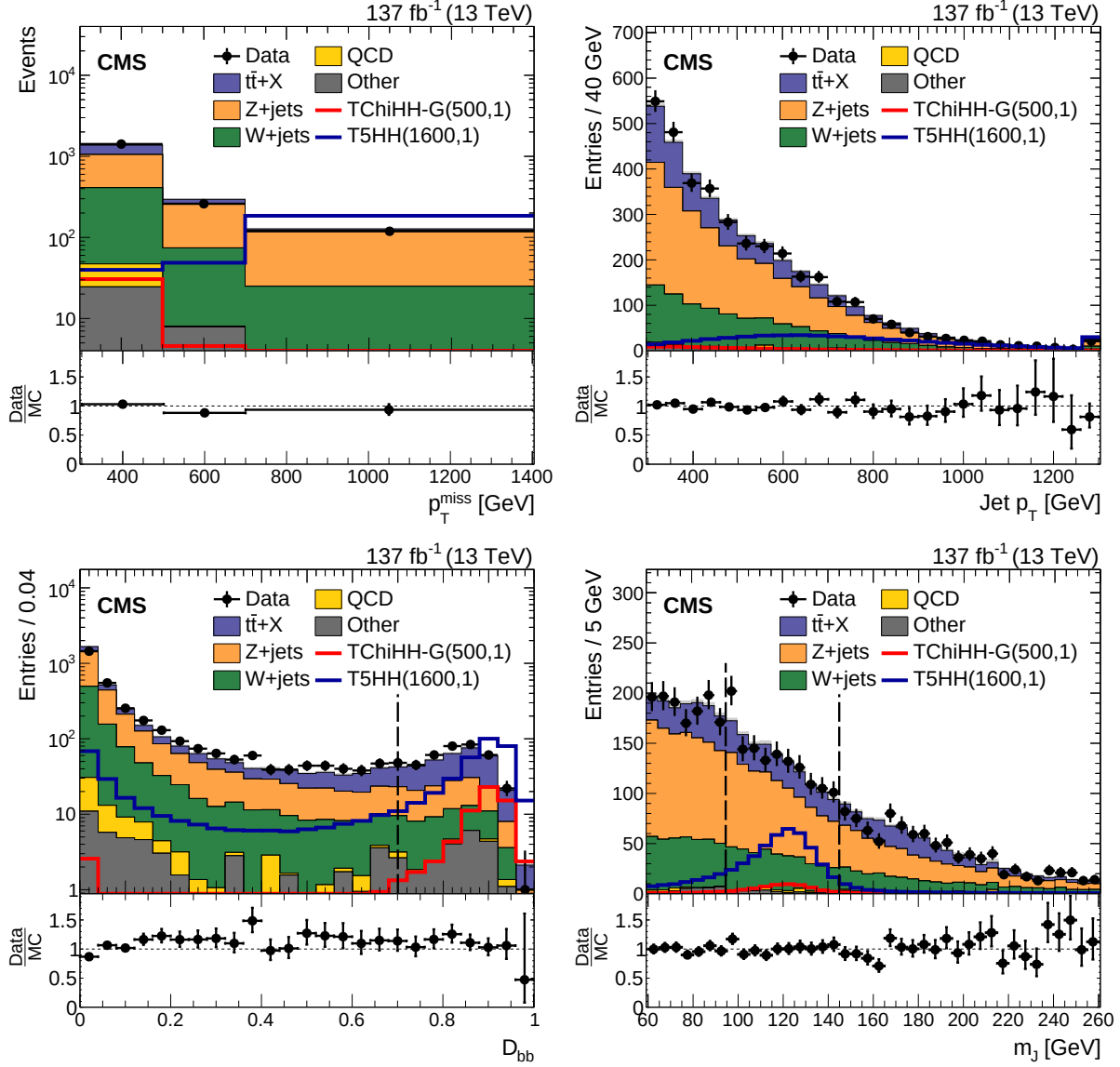


Figure 6.9: Distributions in key analysis variables after the baseline requirements for the boosted signature, except for those on the variable plotted: (upper left)  $p_T^{\text{miss}}$ , (upper right) jet  $p_T$ , (lower left)  $D_{bb}$ , and (lower right)  $m_J$ . Except for the  $p_T^{\text{miss}}$  distribution, each plot contains two entries per event, for each of the two  $p_T$ -leading AK8 jets. The SM background are scaled by 86% to match the total data yield in each plot. Two signal models are shown to motivate region boundaries. The lower panel in each plot shows the ratio of data to (scaled) simulated yields.

As previously mentioned, the ABCD method is data-driven: the yields for the  $A$ ,  $B$ ,  $C$ , and  $D$  regions of Equation 6.5 are taken from data. When the two discriminating variables are partially correlated, the equation does not stand since our assumed ratio equality  $A/B = C/D$  is incorrect. The effect of any such correlation can be corrected with a double ratio  $\kappa$ , which is a measure of how far off  $A/B \neq C/D$ :

$$\kappa_n = \left( \frac{N_A}{N_B} \right)_n / \left( \frac{N_C}{N_D} \right), \quad (6.6)$$

where  $n$  is the signal region of interest.  $\kappa$  is measured and obtained from simulated events, and then applied to the corresponding signal region yields in data:

$$A = B * (C/D) * \kappa \quad (6.7)$$

As expected, when the two ABCD variables are uncorrelated, the  $\kappa$  factors are close to unity.

Further details on the background estimation methods specific to each signature are discussed in the following sections.

### 6.4.3 The resolved signature

The background estimation method for the resolved phase space splits the ABCD regions by the Higgs candidate mass  $\langle m_{bb} \rangle$  (100–140 GeV for the Higgs mass window, 0–100 and 140–200 GeV for the mass sidebands) and the number of b-tagged jets  $N_b$  (3b and 4b are used for the signal regions, 2b is used for the control region), discussed in Section 6.4.1. The ABCD method is operated in every plane of  $p_T^{\text{miss}}\text{--}\Delta R_{\text{max}}$ , and for the  $N_b = 3$  and  $N_b = 4$  signal regions accordingly.

As shown in Figure 6.10, the  $\langle m_{bb} \rangle$  distributions for the three b-tag categories agree within the available statistics, meaning  $\langle m_{bb} \rangle$  and  $N_b$  are roughly uncorrelated. The background in all three b-tag categories is dominated by events that contain only two b quarks, and so it makes sense the  $\langle m_{bb} \rangle$  distributions would like the same. The additional b-tagged jets in the  $N_b = 3$  and  $N_b = 4$  categories come from mistagged light-quark or gluon jets. Any correlation between  $\langle m_{bb} \rangle$  and  $N_b$  is captured in the double-ratio  $\kappa$  correction, derived in simulated samples, and discussed in more detail in Section 6.4.3.1.

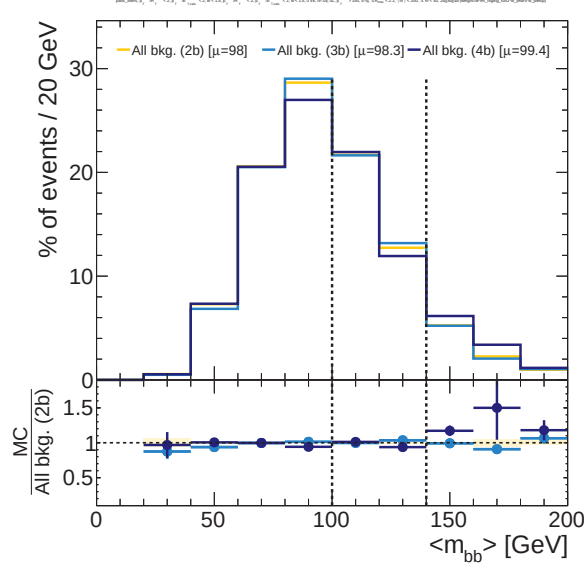


Figure 6.10: Comparison of  $\langle m_{bb} \rangle$  between the three b-tag categories.

The final background estimation for each signal region is:

$$N_{A_{ib}}^{\text{predicted}} = \left( N_{B_{ib}} \frac{N_{C_{2b}}}{N_{D_{2b}}} \right)^{\text{observed}} \kappa^{\text{MC}}, \quad (6.8)$$

where  $i$  indicates the  $N_b = 3$  or 4 tag category.

#### 6.4.3.1 $\kappa$ correction

The performance of the background estimation method within simulation can be determined by dividing the true yield of the signal region with the prediction. This ratio is  $\kappa$ . If the two discriminating variables of the ABCD method are perfectly uncorrelated, the prediction would exactly match the yield of the signal bin, and  $\kappa$  would be 1.  $\kappa$  is obtained by dividing the yields for the signal region by the prediction from the ABCD method:

$$\kappa \equiv N_A^{\text{MC}} / \left( N_B^{\text{MC}} \frac{N_C^{\text{MC}}}{N_D^{\text{MC}}} \right). \quad (6.9)$$

For the resolved method, the  $\kappa$  values for high- $\Delta R_{\text{max}}$  bins are consistent with 1; however, those for the low- $\Delta R_{\text{max}}$  bins are not, ranging from 0.72 to 2.35, indicating a correlation between  $N_b$  and  $\langle m_{bb} \rangle$ . This correlation is due to the difference in the number of ISR jets for each b-tag

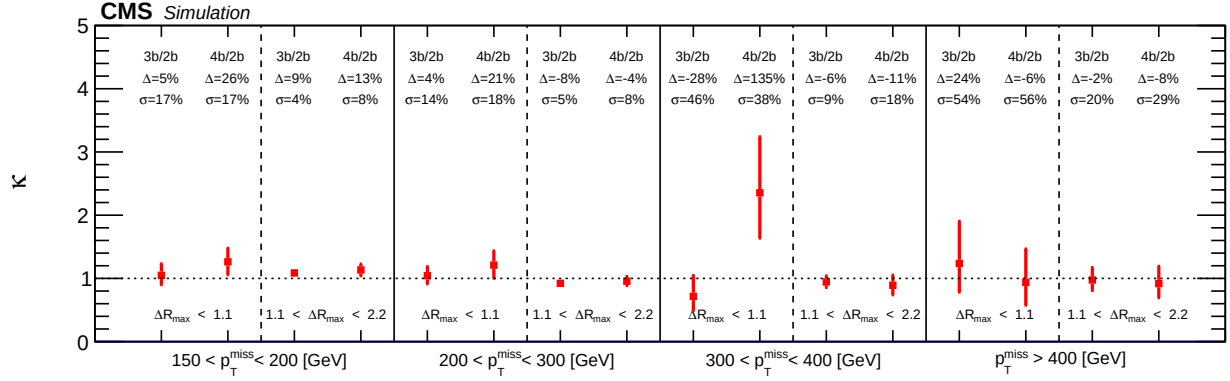


Figure 6.11: The  $\kappa$  values derived in MC as a measure of how well the ABCD prediction matches the true yield for each signal bin.  $\kappa$  is applied as a correction to the background estimation method in the data.

category. Each Higgs candidate is reconstructed from four jets, while the main SM background,  $t\bar{t}$ , only has two b quarks. The  $N_b = 3$  and 4 categories, then, have a larger probability of using ISR jets in the Higgs reconstruction than the  $N_b = 2$  category. The ISR jets are typically more energetic, and so the resulting Higgs candidate tends to have a slightly harder  $\langle m_{bb} \rangle$  distribution. The ISR jet distribution for  $t\bar{t}$  and the other SM backgrounds is well modeled, and so we can derive  $\kappa$  corrections using simulated MC samples and apply them to the observed data.

The calculation of the  $\kappa$  corrections begins with a mixture of the main SM backgrounds, weighted by the appropriate production cross sections. We apply the resolved baseline selection, to match the expected kinematics of our signal bins in data. For each of the 16 signal bins, we calculate a  $\kappa$  correction, shown in Figure 6.11. Additionally,  $\kappa$  values are measured in dedicated validation samples, comparing data to MC, and used to quantify systematic uncertainties on the resolved background method. These are discussed in Section 6.5.1.

#### 6.4.4 The boosted signature

For the boosted method, the SM background is estimated for each signal region ( $A_{1,2}$ ), consisting of two parts: a  $p_T^{\text{miss}}$  shape obtained from a designated control region, and a  $p_T^{\text{miss}}$ -integrated background from an ABCD method that is used to normalize the  $p_T^{\text{miss}}$  shape. In comparison to a pure ABCD method that operates in every bin, this method has the added benefit

that it is not limited by the statistics of the mass sidebands. As mentioned in Section 6.4.1, the ABCD regions are split by the two-dimensional mass plane of the Higgs candidates (95–145 GeV for the Higgs mass window, 60–95 GeV and 145–260 GeV for the mass sidebands) and the number of bb-tagged jets (1 and 2 bb-tagged jets for the signal region, no bb-tagged jets as a control region). There are two signal bins in the boosted method: the 1H mass SR ( $A_{1H}$ ) and the 2H mass SR ( $A_{2H}$ ). We further divide the six ABCD regions into three  $p_T^{\text{miss}}$  bins, [300–500, 500–700, >700] GeV, to improve signal sensitivity.

The  $p_T^{\text{miss}}$  shape  $f_{\text{bkg}}$  is obtained in data from a subset of the  $N_H = 0$  control region ( $C + D$ ), with the additional requirement of at least one tight b-tagged AK4 jet  $N_{b,T} > 0$ . This region, called “0H+b,” requires the additional b-tagged jet in order to enrich it in  $t\bar{t}$  events, more closely approximating the SM content of the signal regions, as can be seen using MC in Figure 6.12. Figure 6.13 shows the  $p_T^{\text{miss}}$  shape agreement between the two signal regions and the 0H+b control region in simulation as proof of concept: the 0H+b region can be used to estimate the  $p_T^{\text{miss}}$  shape of both the  $N_H = 1$  and  $N_H = 2$  signal regions. The  $f_{\text{bkg}}$  that is used for the background estimation method comes from observed data.

We compute a fraction  $f(p_T^{\text{miss}})_{\text{bkg},i} = (p_T^{\text{miss}})_i / (p_T^{\text{miss}})_{\text{tot}}$  in the control region for each of the three  $p_T^{\text{miss}}$  search ranges, where  $(p_T^{\text{miss}})_i$  are the yields in each  $p_T^{\text{miss}}$  bin and  $(p_T^{\text{miss}})_{\text{tot}}$  is the total integral. The fraction  $f(p_T^{\text{miss}})_{\text{bkg},i}$  is then the fraction of total background events expected in a given  $p_T^{\text{miss}}$  search bin.

The background normalization  $\beta_{\text{norm}}$  is the second half of the total prediction, and normalizes the  $p_T^{\text{miss}}$  shape.  $\beta_{\text{norm}}$  is calculated separately for  $N_H = 1, 2$  and is given by

$$\beta_{\text{norm},iH} = N_{B_{iH}} \cdot \left( \frac{N_{C_{0H}}}{N_{D_{0H}}} \right), \quad (6.10)$$

where  $i = 1, 2$  indicates the two signal regions,  $A_{1H}$  and  $A_{2H}$ . Each ABCD region in Equation 6.10 is first integrated in  $p_T^{\text{miss}}$ . This background normalization is unbiased if the ratio of mass window to mass sideband events (SR/SB) remains consistent across  $N_H$ . Any residual correlation would be corrected with the double-ratio  $\kappa$ , the ratio between the true yield of the signal region in

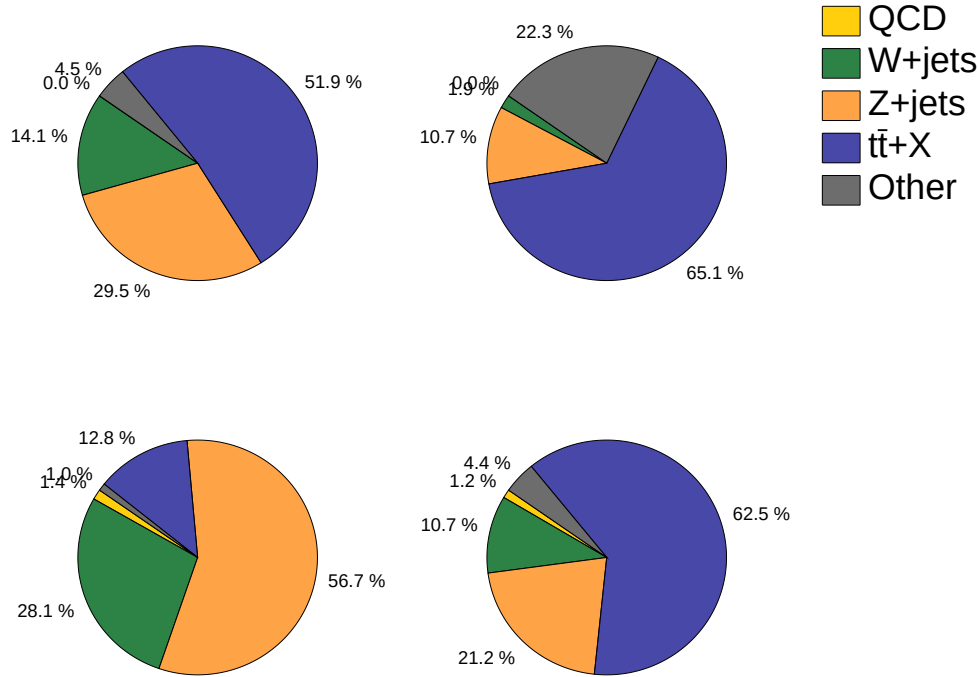


Figure 6.12: The background composition for the  $A_{1H}$  (top left) and  $A_{2H}$  (top right) signal regions, and the control region used to derive the  $p_T^{\text{miss}}$  shape, 0H+b (bottom right), in MC. The background composition for the  $N_H = 0$  region is also shown (bottom left) to demonstrate the effect of the  $N_{b,T} > 0$  requirement.

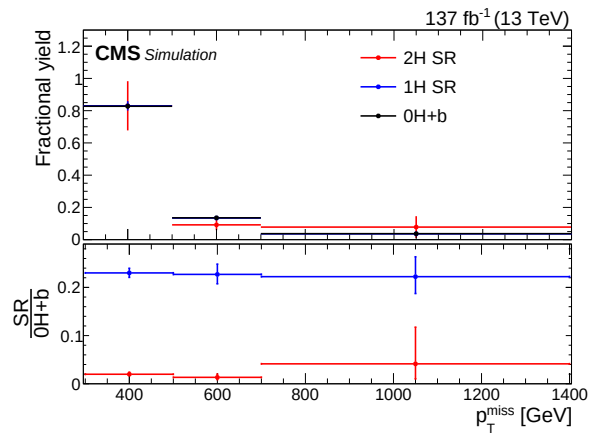


Figure 6.13: Using simulated MC, the  $p_T^{\text{miss}}$  shape agreement between the signal regions and the 0H+b control region that is used to model both. The top panel shows the normalized  $p_T^{\text{miss}}$  shape comparison, while the bottom panel shows the (unnormalized) ratio of the signal region to control region  $p_T^{\text{miss}}$  shape.



MC and the prediction  $\beta_{\text{norm},iH}$ . Table 6.5 shows the prediction  $\beta_{\text{norm},iH}$  and the corresponding agreement to the true signal yields,  $\kappa$ , in MC. As Figure 6.14 demonstrates, SR/SB is the same regardless of bb-tagging category, and so the corresponding  $\kappa$  values are consistent with unity. We therefore do not apply a  $\kappa$  correction, but account for its statistical precision as a systematic of the background method. Other systematic uncertainties assigned to the boosted background prediction are discussed in Section 6.5.2.

The background prediction can then be summarized as:

$$N_{A_{iH}}^{\text{predicted}}(p_T^{\text{miss}}) = (f_{\text{bkg}}(p_T^{\text{miss}}) \cdot \beta_{\text{norm}})^{\text{observed}}. \quad (6.11)$$

Figure 6.15 shows a comparison in MC of the full background prediction  $N_{A_{iH}}(p_T^{\text{miss}})$  to the true yields of each signal region. Using MC for this closure test shows how well our prediction  $\beta_{\text{norm},iH} \cdot f_{\text{bkg}}$  models the actual yields  $A_{iH}$ , and is useful for both proof-of-concept and confirmation that  $\kappa$  corrections are not needed. In the case of the actual background prediction method in data, we expect the prediction to be rather similar in behaviour to the MC prediction, but a SUSY signature would increase the yields in the  $A_1$  and  $A_2$  signal bins, leading to poor agreement.

## 6.5 Validation and systematic uncertainties

In this section, we discuss the validation of the background estimation methods, and how the associated systematic uncertainties are defined and derived. We use dedicated validation samples in both data and MC to derive the  $\kappa$  corrections, validate the background predictions, and quantify systematic uncertainties. The validation samples are beneficial because they allow us to look at real data (not used in our analysis region) and compare it with generated MC samples of the same SM backgrounds. This gives us confidence in our methodology and in how well the MC models the backgrounds. A validation sample is used to study each of the primary SM backgrounds ( $Z \rightarrow \nu\nu$ ,  $t\bar{t} \rightarrow \ell\nu qq'$ ,  $W \rightarrow \ell\nu$ , QCD), which are rather enhanced directly in the SM background process or a representative topology:

- Single-lepton: used to validate the  $t\bar{t}$  and  $W$ +jets processes. The topology of events from

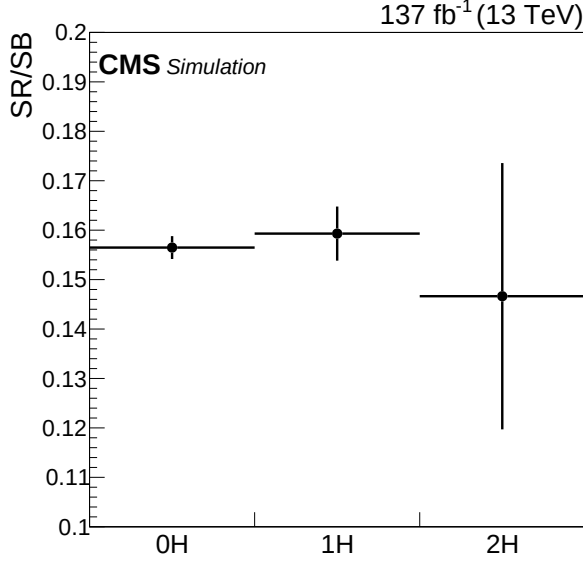


Figure 6.14: The ratio of the mass signal to mass sideband yields in MC across  $N_H$ . A consistent ratio shows the Higgs candidate mass and the bb-tagger are uncorrelated.

| Region                   | MC Yields       |
|--------------------------|-----------------|
| $A_{2H}$                 | $3.9 \pm 0.7$   |
| $B_{2H}$                 | $26.9 \pm 1.9$  |
| $A_{1H}$                 | $45.9 \pm 1.4$  |
| $B_{1H}$                 | $288 \pm 4$     |
| $C_{0H}$                 | $234 \pm 3$     |
| $D_{0H}$                 | $1495 \pm 9$    |
| $\beta_{\text{norm},2H}$ | $4.2 \pm 0.3$   |
| $\beta_{\text{norm},1H}$ | $45.1 \pm 0.9$  |
| $\kappa_{2H}$            | $0.94 \pm 0.17$ |
| $\kappa_{1H}$            | $1.02 \pm 0.04$ |

Table 6.5: Yields in the  $A_{iH}$  bins for the boosted signature using simulated data. The background normalization  $\beta_{\text{norm},iH}$  is given for both bins. The ratio  $\kappa$  of the yields in the signal bins to the corresponding  $\beta_{\text{norm}}$  is found to be consistent with 1.0 in MC.

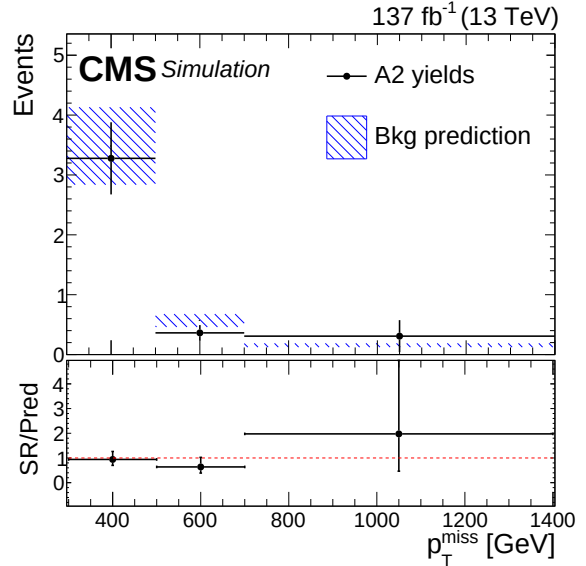
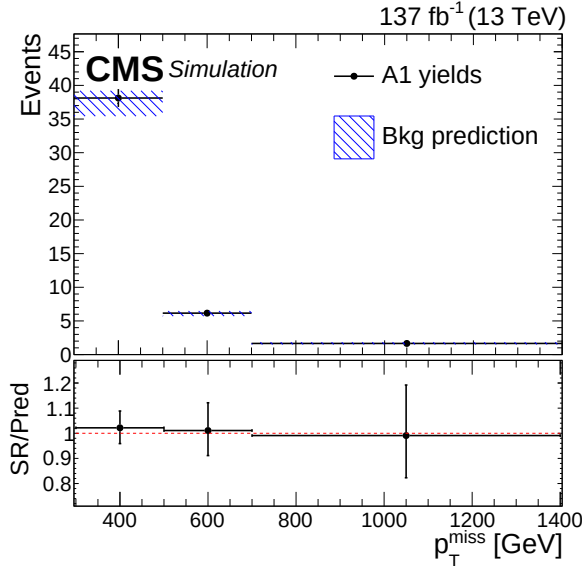


Figure 6.15: The signal region yields (black) compared to the background prediction (blue) for the 1H (left) and 2H (right) signal regions using MC events. The bands of uncertainty are based on the statistical uncertainty of  $\beta_{\text{norm}}$  and  $f_{\text{bkg}}$  only. The bottom panel is the ratio of the signal region yields to the prediction, which is consistent with 1.0 across  $p_T^{\text{miss}}$  for both signal regions. This unitary ratio means our background method is sound at predicting the yields of the signal regions, in a SM-background-only hypothesis.

direct W or top quark production is the same, regardless of whether the lepton (electron or muon) is identified or lost.

- Low- $\Delta\phi$ : used to validate the QCD multijet background for the resolved signature. The  $\Delta\phi$  requirement is reversed to explicitly select events in which the AK4 jet momentum is likely under-measured, resulting in close alignment with  $\vec{p}_T^{\text{miss}}$ , giving spuriously large momentum imbalance.
- Dilepton: used to validate the Z+jets background for the resolved signature. Specifically, the  $Z \rightarrow \ell\ell$  decay shares many topological similarities to  $Z \rightarrow \nu\nu$ , and provides a direct handle on the physics of Z production. The leptons from this decay are not involved in the reconstruction of the Higgs candidate mass, nor are they correlated with the presence of additional b jets.
- Single-photon: used to validate the Z+jets background for the boosted signature. When a Z boson has a high enough  $p_T$ , such that its mass is negligible, it is expected to have similar kinematics to the photon (also massless, neutral, electroweak gauge boson). The photon is removed from event reconstruction, giving a similar amount of momentum imbalance as  $Z \rightarrow \nu\nu$ .

The uncertainties in the background predictions for the resolved and boosted phase space regions are described in Sections 6.5.1 and 6.5.2, respectively. Signal systematics are discussed in Section 6.5.3.

### 6.5.1 Resolved background uncertainties

The background estimation method for the resolved signature (described in Section 6.4.3) relies on  $\langle m_{bb} \rangle$  and  $N_b$  being roughly uncorrelated, and on the Monte Carlo simulation to correctly model any residual dependence. For the nominal ABCD plane (six regions), we construct a corresponding ABCD plane using the validation samples, using a single-lepton sample to check  $t\bar{t}$  and W+jets events, a dilepton sample to validate Z+jets events, and a low- $\Delta\phi$  sample that is enriched in QCD events to verify the QCD background. We then compute  $\kappa$  factors in both data and MC,

| Source                                 | Uncertainty [%]                    |         |    |         |    |         |    |      |    |
|----------------------------------------|------------------------------------|---------|----|---------|----|---------|----|------|----|
|                                        | $p_{\text{T}}^{\text{miss}}$ [GeV] | 150–200 |    | 200–300 |    | 300–400 |    | >400 |    |
|                                        | $N_{\text{b}}$                     | 3       | 4  | 3       | 4  | 3       | 4  | 3    | 4  |
| $1.1 < \Delta R_{\text{max}} < 2.2$    |                                    |         |    |         |    |         |    |      |    |
| $\kappa$ stat. unc.                    |                                    | 4       | 8  | 5       | 8  | 9       | 18 | 20   | 29 |
|                                        | $t\bar{t}$                         | 2       | 7  | 2       | 7  | 1       | 6  | 1    | 7  |
| $\Delta\kappa(\text{data, MC})/\kappa$ | V+jets                             | <1      | <1 | 1       | <1 | 1       | 1  | 2    | 2  |
|                                        | QCD                                | <1      | <1 | <1      | 1  | <1      | <1 | <1   | 1  |
| $\kappa$ total syst. unc.              |                                    | 2       | 7  | 2       | 7  | 1       | 6  | 2    | 7  |
| $\Delta R_{\text{max}} < 1.1$          |                                    |         |    |         |    |         |    |      |    |
| $\kappa$ stat. unc.                    |                                    | 17      | 17 | 14      | 18 | 46      | 38 | 54   | 56 |
|                                        | $t\bar{t}$                         | 11      | 18 | 10      | 17 | 6       | 13 | 6    | 9  |
| $\Delta\kappa(\text{data, MC})/\kappa$ | V+jets                             | 1       | 1  | 4       | 5  | 10      | 6  | 6    | 10 |
|                                        | QCD                                | 1       | <1 | <1      | 2  | <1      | <1 | <1   | <1 |
| $\kappa$ total syst. unc.              |                                    | 11      | 18 | 11      | 18 | 12      | 14 | 8    | 13 |

Table 6.6: Summary of the background uncertainties for the resolved signature. These percentages are fractional uncertainties for a given background process. Scaling these percentages by the relative background composition and adding them in quadrature gives the total systematic uncertainty on  $\kappa$ . The systematics propagated from the validation samples are small compared to the data statistical uncertainty of the ABCD method, the predominant uncertainty of the resolved signature.

comparing the ABCD prediction to the yields of the signal bins  $A_{3b}$  and  $A_{4b}$ . This comparison is called “closure.” We assign systematic uncertainties on the  $\kappa$  values, accounting for any correlation between  $\langle m_{bb} \rangle$  and  $N_b$ . Additionally, any difference between the MC and data is captured inside these systematics. For each validation region, we define  $\Delta_\kappa$  as the difference between  $\kappa^{\text{data}}$  and 1.0, and  $\sigma_{\text{stat}}$  as the statistical uncertainty of  $\kappa^{\text{data}}$  relative to unity. The systematic uncertainty is assigned as the larger of these values. These uncertainties are then scaled by the relative background composition of our SM backgrounds in the analysis region, and propagated to the signal bins.

Table 6.6 shows the total systematic uncertainties for each signal bin based on the uncertainties derived in the validation samples. These systematic uncertainties are small compared to the statistical uncertainty of the ABCD method, which is the statistical uncertainty in data in the  $B$ ,  $C$ , and  $D$  regions. Further details on the validation sample requirements can be found in Sections 6.5.1.1, 6.5.1.2, and 6.5.1.3.

### 6.5.1.1 Single-lepton validation sample

The single-lepton validation sample, used to model the  $t\bar{t}$  background, has the same baseline requirements as the analysis region samples, excluding the lepton and isolated track vetoes. This gives the same topology as the  $t\bar{t}$  background, which is predominantly single-lepton ( $t\bar{t} \rightarrow \ell\nu qq'$ ) where the lepton is rather not identified or is outside of our acceptance. The  $\Delta\phi$  cut is also removed, since the single-lepton sample is free from QCD contamination. We instead require exactly one lepton with  $p_T > 30$  GeV and impose the following requirement:

$$m_T(\ell, p_T^{\text{miss}}) \equiv \sqrt{2p_T^\ell p_T^{\text{miss}}(1 - \cos \Delta\phi)} < 100 \text{ GeV}, \quad (6.12)$$

where  $\Delta\phi$  is the angular separation between the lepton and  $\vec{p}_T^{\text{miss}}$ . This  $m_T$  requirement helps pick out events consistent with a  $W$  decay: if the neutrino in the  $W$  decay is the only source of true  $p_T^{\text{miss}}$  in the event,  $m_T$  will peak at or below the  $W$  mass of 80.4 GeV. The lepton presence allows us to cleanly trigger on events with lower  $p_T^{\text{miss}}$ , and so we add two additional  $p_T^{\text{miss}}$  bins,  $p_T^{\text{miss}} < 75$  GeV and  $75 < p_T^{\text{miss}} \leq 150$  GeV, to study the  $p_T^{\text{miss}}$  dependence of  $\kappa$  in a wider range.

The simulation modeling of the  $\Delta R_{\text{max}}$  and  $\langle m_{bb} \rangle$  distributions is assessed for this validation sample in Fig. 6.16, which shows good data-to-simulation agreement. The  $\langle m_{bb} \rangle$  shape is also similar to the search sample, validating the assumption that this validation sample can be used to model and assess the  $t\bar{t}$  and  $W$ +jets backgrounds of the search sample.

We assess the validity of the background estimation method in  $t\bar{t}$ -like events by comparing the  $\kappa$  values in data and simulation. Figure 6.17 shows the closure of the method, the ratio of the true yields to the prediction, between the simulation and data. We observe good closure within statistical uncertainty across all  $p_T^{\text{miss}}$  bins and no apparent dependency on  $p_T^{\text{miss}}$ , and thus integrate in  $p_T^{\text{miss}}$  (the last panel of Figure 6.17) to increase the statistical precision. As previously mentioned, the systematic uncertainty is assigned as the larger of the non-closure  $\Delta_\kappa$  and the statistical precision of  $\kappa$   $\sigma_{\text{stat}}$ . The results in Figure 6.17 correspond to an uncertainty of 13% (18%) in the  $N_b = 3$  (4) bins for low  $\Delta R_{\text{max}}$ , and of 2% (10%) in the  $N_b = 3$  (4) bins for high  $\Delta R_{\text{max}}$ . These percentages are fractional uncertainties for the  $t\bar{t}$  background. Scaling these percentages by the relative background

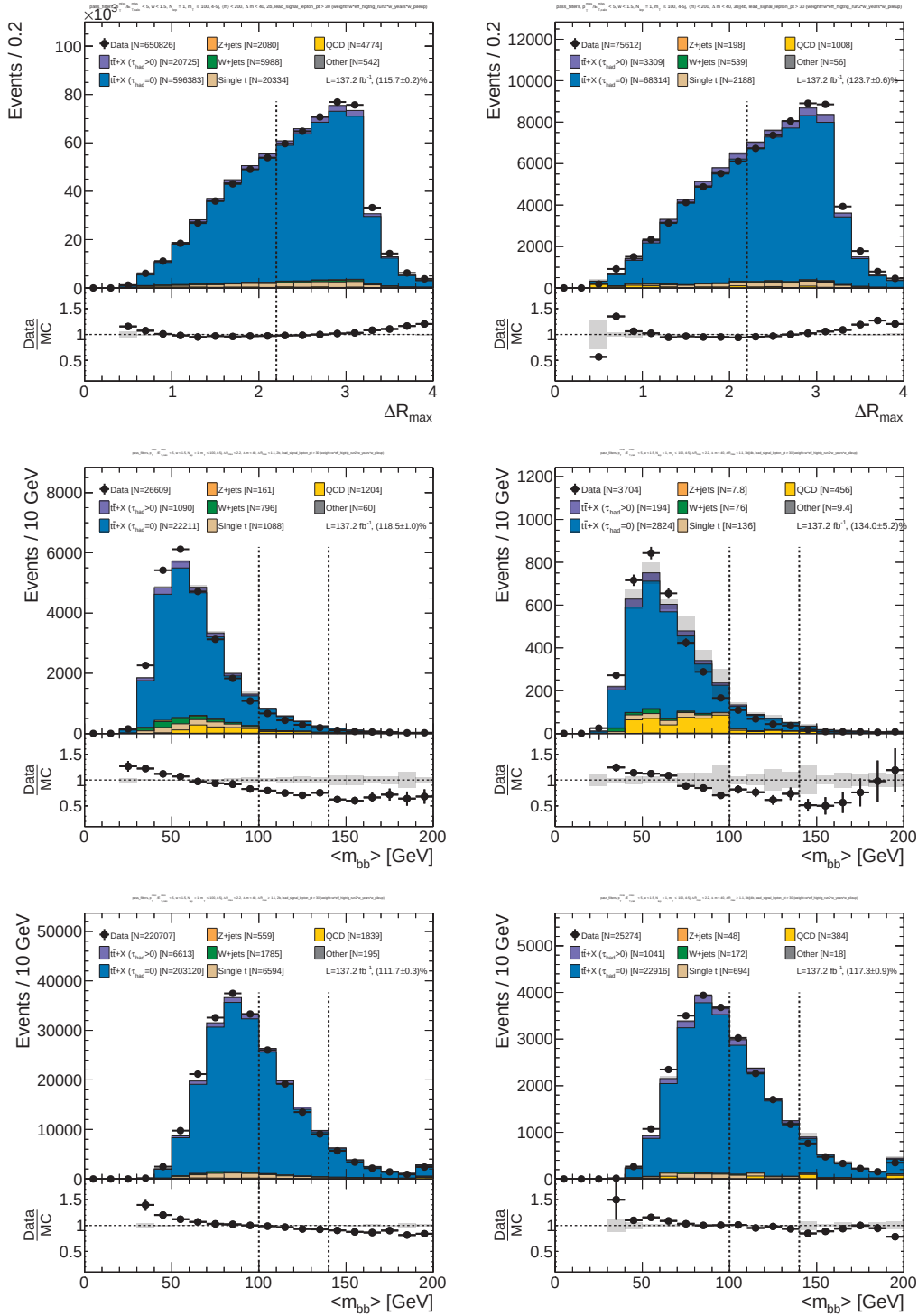


Figure 6.16: Comparison of data to simulation in the single-lepton validation for  $N_b = 2$  (left column) and  $N_b = 3$  or  $4$  (right column) in  $\Delta R_{\max}$  (top row),  $\langle m_{bb} \rangle$  with  $\Delta R_{\max} < 1.1$  (middle row), and  $\langle m_{bb} \rangle$  with  $1.1 \leq \Delta R_{\max} < 2.2$  (bottom row). The simulation is normalized to the data yields, with the scaling factor indicated under the legend.

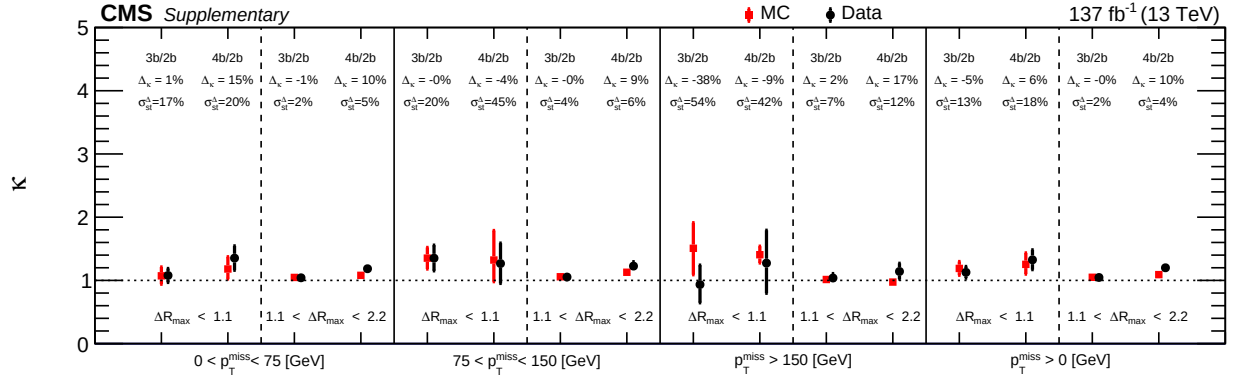


Figure 6.17:  $\kappa$  values measured in data and simulation in the single-lepton validation sample, for the 2b-3b and 2b-4b ABCD planes in each  $p_T^{\text{miss}} - \Delta R_{\text{max}}$  bin.  $\Delta_\kappa$  is the difference between the value of  $\kappa$  in data and unity, and  $\sigma_{\text{stat}}$  is the statistical uncertainty of  $\kappa$  measured in data relative to unity.

composition and adding them in quadrature gives the total systematic uncertainty on  $\kappa$  in Table 6.6.

### 6.5.1.2 Dilepton validation sample

A validation sample enriched in  $Z \rightarrow \ell\ell$  decays is used to assess the background prediction method for Z+jets events. This sample provides a direct handle on the physics of Z production, with near-identical topology, since the leptons are not involved in the Higgs candidate reconstruction, nor are they correlated with the presence of b-tagged jets. The dilepton validation sample has the same baseline requirements as the analysis region sample, excluding the lepton and isolated track vetoes. Additionally, we remove the  $\Delta\phi$  requirement, since requiring two leptons makes the contamination from QCD events negligible. We instead select events with two electrons or muons of opposite sign, at least one of which has  $p_T > 30$  GeV, a dilepton mass within 15 GeV of the Z mass, and  $p_T^{\text{miss}} < 50$  GeV. We compute the  $p_T$  of the dilepton system,  $p_T(Z)$ , and use this as a proxy for the true  $p_T^{\text{miss}}$  that would be measured in a  $Z \rightarrow \nu\nu$  decay. Given the small branching fraction of  $Z \rightarrow \ell\ell$  decays and the large  $t\bar{t}$  contamination associated with a high  $N_b$  selection, we lower the  $N_b$  requirements for regions A and B (C and D) to  $N_b = 1$  (0), using the medium WP of the b-tagger.

We divide the  $Z \rightarrow \ell\ell$  sample into bins of  $p_T(Z)$ , adding two additional bins of  $p_T(Z) < 75$  GeV

and  $75 < p_T(Z) \leq 150$  GeV, like the single-lepton validation sample. As shown in Figure 6.18, there is excellent agreement between data and simulation in  $\Delta R_{\max}$  and  $\langle m_{bb} \rangle$  for both  $N_b = 0$  and  $N_b = 1$ .

We assess the validity of the background prediction method of Z+jets events by comparing the  $\kappa$  factors between data and simulation. Figure 6.19 shows the closure of the method, the ratio of the true yields to the prediction, between the simulation and data. We observe good closure within statistical uncertainty across all  $p_T(Z)$  bins and no apparent dependency on  $p_T(Z)$ , and thus integrate in  $p_T(Z)$  to increase the statistical precision. As previously mentioned, the systematic uncertainty is assigned as the larger of the non-closure  $\Delta_\kappa$  and the statistical precision of  $\kappa$   $\sigma_{\text{stat}}$ . The results in Figure 6.19 correspond to uncertainties of 16% and 5% for low and high  $\Delta R_{\max}$ , respectively. These percentages are fractional uncertainties for the V+jets background. Scaling these percentages by the relative background composition and adding them in quadrature gives the total systematic uncertainty on  $\kappa$  in Table 6.6.

### 6.5.1.3 Low- $\Delta\phi$ validation sample

To obtain a QCD-enriched validation sample, we invert the  $\Delta\phi$  requirement of the search sample. This selects events in which the AK4 jet momentum is likely under-measured, resulting in close alignment with  $\vec{p}_T^{\text{miss}}$ . Like the dilepton validation sample, we lower the  $N_b$  requirements for regions *A* and *B* (*C* and *D*) to  $N_b = 1$  (0). As shown in Figure 6.20, there is reasonable agreement between data and simulation for both  $N_b = 0$  and  $N_b = 1$ .

We assess the validity of the background prediction method for QCD events by comparing the  $\kappa$  factors in data and simulation. Figure 6.21 shows the closure of the method, the ratio of the true yields to the prediction, between the simulation and data. We observe good closure within statistical uncertainty across all  $p_T^{\text{miss}}$  bins and no apparent dependency on  $p_T^{\text{miss}}$ , and thus integrate in  $p_T^{\text{miss}}$  to increase the statistical precision. As previously mentioned, the systematic uncertainty is assigned as the larger of the non-closure  $\Delta_\kappa$  and the statistical precision of  $\kappa$   $\sigma_{\text{stat}}$ , which is 7% and 9% in the low- and high- $\Delta R_{\max}$  bins, respectively. These percentages are fractional uncertainties



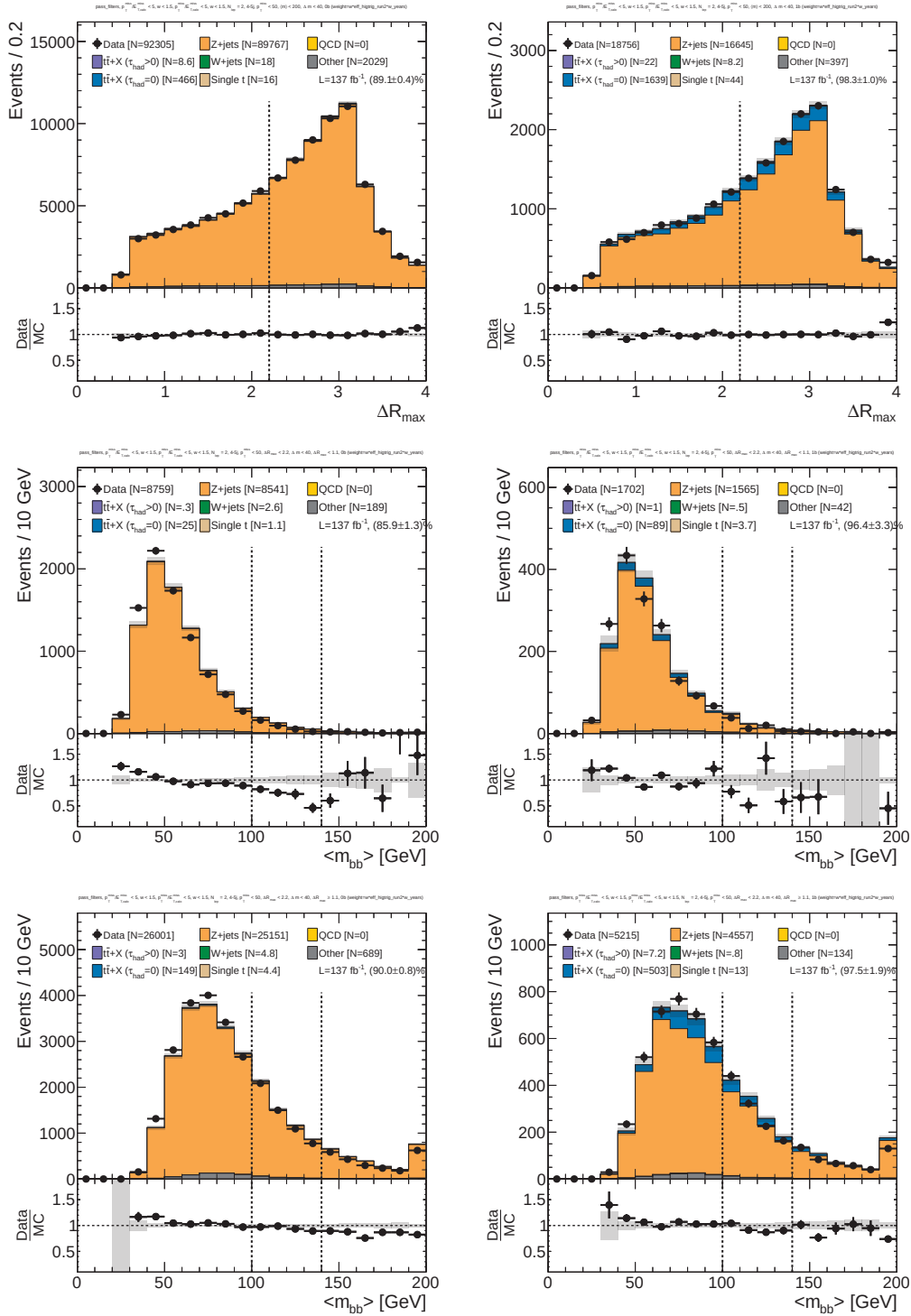


Figure 6.18: Comparison of data to simulation in the dilepton validation sample for  $N_b = 0$  (left column) and  $N_b = 1$  (right column) in  $\Delta R_{\max}$  (top row),  $\langle m_{bb} \rangle$  with  $\Delta R_{\max} < 1.1$  (middle row), and  $\langle m_{bb} \rangle$  with  $1.1 \leq \Delta R_{\max} < 2.2$  (bottom row). The simulation is normalized to the data yields, with the scaling factor indicated under the legend.

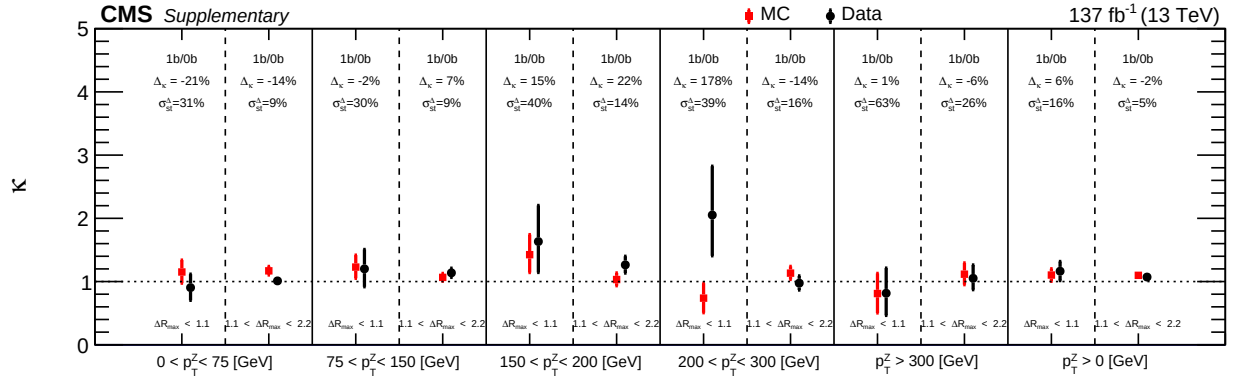


Figure 6.19:  $\kappa$  values measured in data and simulation in the dilepton validation sample, for  $N_b = 0$  and  $N_b = 1$ .  $\Delta_\kappa$  is the difference between the value of  $\kappa$  in data and unity.  $\sigma_{\text{stat}}$  is the statistical uncertainty of  $\kappa$  measured in data relative to unity.

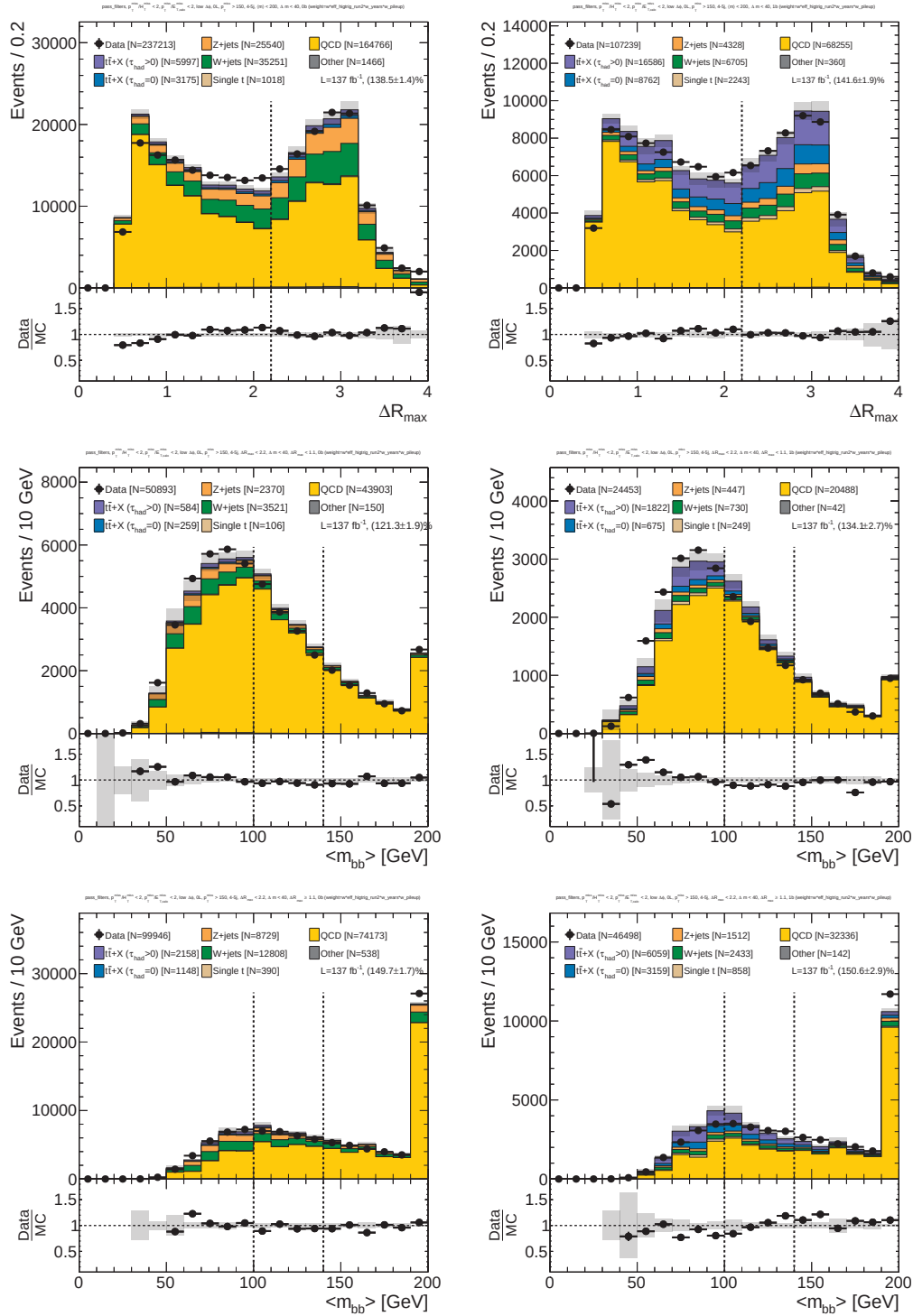


Figure 6.20: Comparison of data to simulation in the low- $\Delta\phi$  validation sample for  $N_b = 0$  (left column) and  $N_b = 1$  (right column) in  $\Delta R_{\max}$  (top row),  $\langle m_{bb} \rangle$  with  $\Delta R_{\max} < 1.1$  (middle row), and  $\langle m_{bb} \rangle$  with  $1.1 \leq \Delta R_{\max} < 2.2$  (bottom row). The simulation is normalized to the data yields, with the scaling factor indicated under the legend.

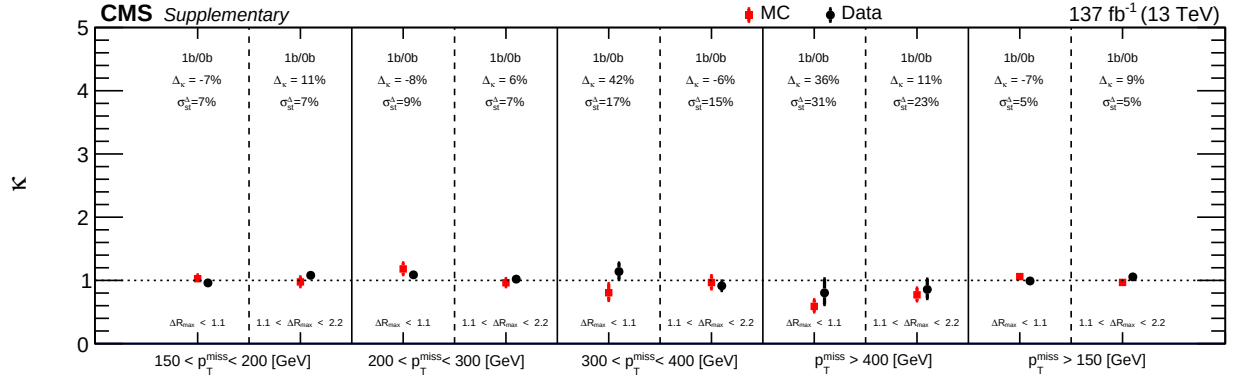


Figure 6.21:  $\kappa$  values measured in data and simulation in the low- $\Delta\phi$  validation sample, for  $N_b = 0$  and  $N_b = 1$ .  $\Delta_\kappa$  is the difference between the value of  $\kappa$  in data and unity.  $\sigma_{\text{stat}}$  is the statistical uncertainty of  $\kappa$  measured in data relative to unity.

for the QCD background, and scaling these uncertainties by the relative background composition and adding them in quadrature gives the total systematic uncertainty on  $\kappa$ , which we include in Table 6.6.

### 6.5.2 Boosted background uncertainties

Similar to the resolved method, validation samples are used to confirm the validity of the boosted background estimation method (described in Section 6.4.4) and to check the consistency of data and MC. There are two main validation regions, one for each of the main backgrounds. The  $t\bar{t}$  and  $W$ +jets background processes are studied with the single-lepton validation sample described in the resolved systematics above (Section 6.5.1.1). Briefly, we require exactly one lepton with  $p_T > 30$  GeV and  $m_T < 100$  GeV and remove the  $\Delta\phi$  requirement since QCD contamination is no longer a concern.

To validate the background from  $Z$ +jets, a single-photon validation sample is used with photon  $p_T > 100$  GeV. The photon, used as a proxy for the  $Z$  boson in  $Z \rightarrow \nu\nu$  events, is removed from the event; the missing photon  $p_T$  serves as a source of  $p_T^{\text{miss}}$ . The  $p_T^{\text{miss}}$  requirement is loosened to  $p_T^{\text{miss}} > 200$  GeV to increase statistical precision, only affecting the  $\kappa$  calculation and validation (the  $p_T^{\text{miss}}$  shape is only considered for  $p_T^{\text{miss}} > 300$  GeV). All other event selection requirements for

the two validation regions are the same as for the analysis regions.

The background estimation method for the boosted signature is split into two parts: the  $p_T^{\text{miss}}$  shape  $f_{\text{bkg}}$  (described in Section 6.5.2.1) and the background normalization  $\beta_{\text{norm}}$  (described in Section 6.5.2.2). Each portion has four systematics:

- The statistical uncertainty in data, considered the dominant uncertainty
- The statistical uncertainty in MC closure
- The agreement between data and MC in the validation samples
- The effect of known MC mismodeling

This last systematic uncertainty covers the effect of known MC mismodeling on both  $\beta_{\text{norm}}$  and  $f_{\text{bkg}}$  from the following sources: b and double-b tagging, jet energy scale, jet energy resolution, ISR corrections, and renormalization ( $\mu_R$ ) and factorization ( $\mu_F$ ) scales. The individual sources and their uncertainties, as applied to the signal samples, are described in Section 6.5.3. The effect of each of these contributions ranges from 0 to 5%, and the total systematic uncertainty is calculated by taking the largest uncertainty for each source (typically, there is an up and a down variation), and adding those in quadrature.

#### 6.5.2.1 Systematic uncertainties in $f_{\text{bkg}}$

As stated above, the  $p_T^{\text{miss}}$  shape  $f_{\text{bkg}}$  has four sources of uncertainty. The first is the statistical uncertainty in data in the control region “0H+b”, where  $f_{\text{bkg}}$  is taken from. This statistical uncertainty is in each individual  $p_T^{\text{miss}}$  bin, for 300–500 GeV, 500–700 GeV, and >700 GeV. The yields and associated percentage of background per  $p_T^{\text{miss}}$  bin are in Table 6.7. These yields result in systematic uncertainties of 8%, 18%, and 37%, for each of the bins.

The second source of uncertainty comes from the  $p_T^{\text{miss}}$  shape closure in MC, how well the  $p_T^{\text{miss}}$  shape in the 0H+b region models that of the signal bins,  $A_{1\text{H}}$  and  $A_{2\text{H}}$ . The shape uncertainty is derived from the shape difference of two fits, one for a constant and the other with a floating

| $p_T^{\text{miss}}$ [GeV] | $N_{\text{obs}}$ | $f_{\text{bkg}}$  |
|---------------------------|------------------|-------------------|
| 300 – 500                 | 142              | $0.789 \pm 0.030$ |
| 500 – 700]                | 31               | $0.172 \pm 0.028$ |
| > 700                     | 7                | $0.039 \pm 0.014$ |

Table 6.7: Data yields in the 0H+b control region. The percentage of the background that lies in each  $p_T^{\text{miss}}$  bin,  $f_{\text{bkg}}$ , is also provided. The statistical uncertainty on  $f_{\text{bkg}}$  per bin is 8%, 18%, and 38%, in increasing  $p_T^{\text{miss}}$ .

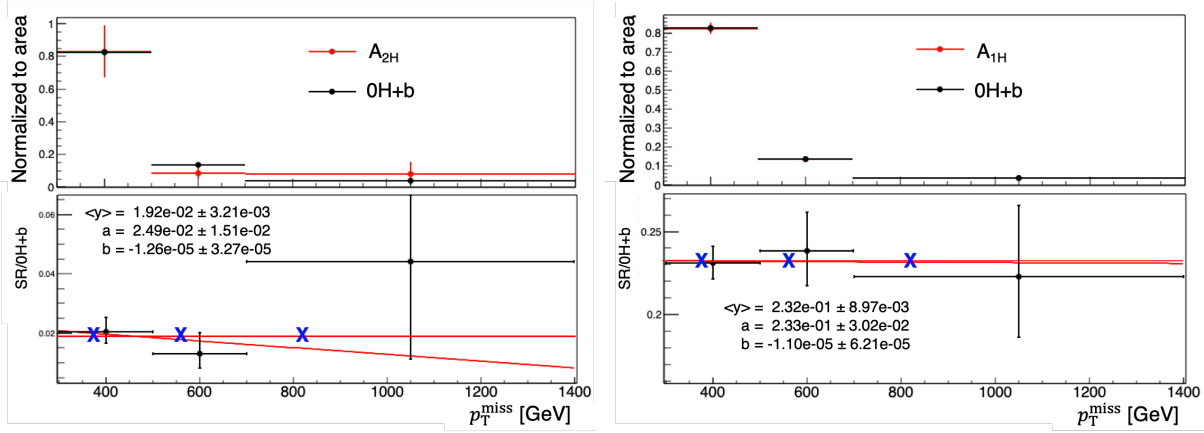


Figure 6.22: A visual of how the  $p_T^{\text{miss}}$  shape closure systematic is derived for  $A_{2H}$  (left) and  $A_{1H}$  (right) in MC. The top panels show the normalized distributions of  $A$  compared to the 0H+b region. The bottom panel shows the ratio of  $A$  to the 0H+b region. Each bottom panel is fit with a flat constant,  $\langle y \rangle$ , and a first-degree polynomial,  $a + b \cdot x$ . The fit parameters are listed. The blue x's indicate the mean average of each  $p_T^{\text{miss}}$  bin, at 370, 565, and 830 GeV.

slope. These fits and associated parameters can be seen in Figure 6.22. The agreement between these fits is converted to a shape uncertainty and assigned as a systematic uncertainty per  $p_T^{\text{miss}}$  bin: 13%, 19%, and 63% (2%, 3%, and 8%) for the  $A_{2H}$  ( $A_{1H}$ ) signal bin, for the  $p_T^{\text{miss}}$  bins in increasing value.

The third systematic uncertainty accounts for potential background composition mismodeling in the MC compared to the data, a crucial component for  $f_{\text{bkg}}$  to model the  $p_T^{\text{miss}}$  shape of the signal bins  $A_{iH}$ . Figure 6.23 shows the  $p_T^{\text{miss}}$  shape comparison in the validation samples for the  $A_{1H}$  and  $A_{2H}$  selection compared to the 0H+b selection. While the agreement in the  $p_T^{\text{miss}}$  shape is acceptable, the available statistical precision is limited in both the single-lepton and single-photon regions. Instead of assigning a systematic based on this agreement, which would be more of a measure of available statistics, we vary the Z+jets cross section by a factor of  $\frac{1}{2}$  and by a factor of

2, and assign a systematic based on the change of the  $p_T^{\text{miss}}$  shape.  $t\bar{t}$  has a softer  $p_T^{\text{miss}}$  shape than Z+jets, and so  $f_{\text{bkg}}$  is highly dependent on the MC correctly modeling the background composition. We assign uncertainties of 3%, 7%, and 32%, in increasing  $p_T^{\text{miss}}$  bins, for both signal bins.

Lastly, we check the effect of known MC modeling uncertainties on the MC closure. The uncertainties on  $f_{\text{bkg}}$  are found to be +0.6%, -2.3%, and -4.8% in the three  $p_T^{\text{miss}}$  bins.

### 6.5.2.2 Systematic uncertainties in $\beta_{\text{norm}}$

Like the uncertainties on  $f_{\text{bkg}}$ , the background normalization has four sources of uncertainty. The first is the statistical uncertainty of  $\beta_{\text{norm},iH}$  in data, from the  $B_{iH}$ ,  $C_{0H}$ , and  $D_{0H}$  regions, 10% and 19% for  $N_H = 1$  and  $N_H = 2$ , respectively.

The second source of uncertainty comes from the MC closure,  $\kappa$ , the agreement between  $\beta_{\text{norm},iH}$  and the true yields of  $A_{iH}$  in MC. As can be seen in Table 6.5,  $\beta_{\text{norm},iH}$  is consistent with the true  $A_{iH}$  yields for both  $N_H = 1$  and  $N_H = 2$ , meaning  $\kappa$  is consistent with unity. Because of this agreement, we do not apply a  $\kappa$  correction, but account for the statistical uncertainty on  $\kappa$ , 4% for  $N_H = 1$  and 17% for  $N_H = 2$ , as a systematic uncertainty.

The third source of uncertainty in  $\beta_{\text{norm}}$  is determined by the agreement of  $\kappa$  between simulation and data in the validation samples. Table 6.8 lists the signal bin yields,  $\beta_{\text{norm},iH}$ , and  $\kappa_{iH}$  in the single-lepton and single-photon validation samples. We compare the background normalization and the closure,  $\kappa$ , between simulation and data. Although there is a scale factor between the simulation and data, the ratio ( $\kappa$ ) between the  $A_{iH}$  yield and  $\beta_{\text{norm},iH}$  is consistent between MC and data. We assign a systematic uncertainty of the larger of the discrepancy in  $\kappa$  values between data and MC or the combined statistical uncertainty of the data and MC, in the single-lepton validation sample, resulting in 13% for  $N_H = 2$  and 9% for  $N_H = 1$ . Even with the less strict  $p_T^{\text{miss}}$  requirement, we find that the statistics are too low in the single-photon sample. Thus, we assign no systematic based on this validation region, but instead take this as a cross-check validation.

Lastly, we check the effect of known MC modeling uncertainties on  $\beta_{\text{norm},iH}$  and find uncertainties of 4.0% and 5.7% for  $N_H = 1$  and  $N_H = 2$ , respectively.

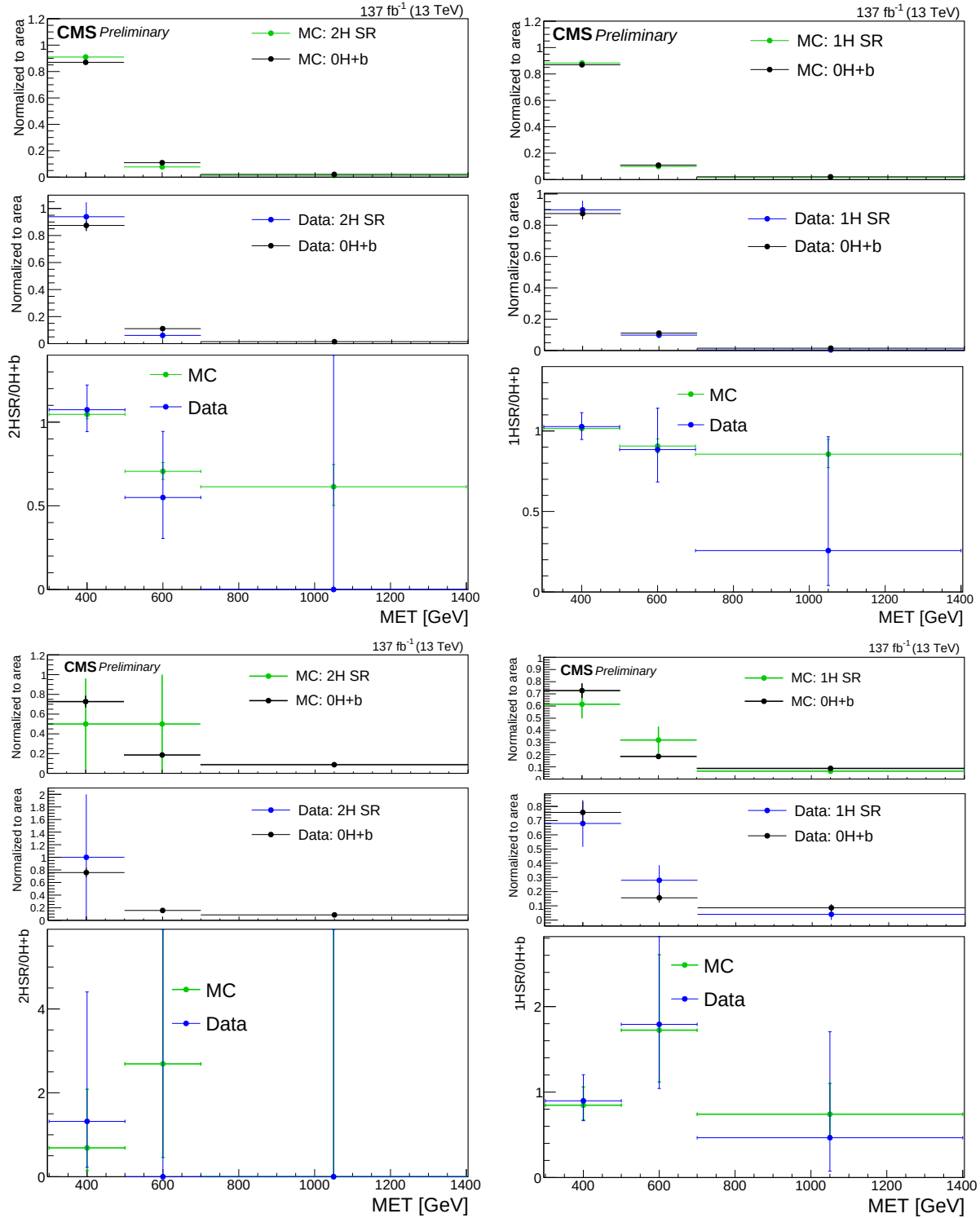


Figure 6.23:  $f_{\text{bkg}}$  comparison between data and MC in the  $A_{2H}$  (left) and  $A_{1H}$  (right) signal bins in the single-lepton (top row) and single-photon (bottom row) validation samples. The top panel of each plot shows the normalized  $p_T^{\text{miss}}$  distributions of the signal bins and the 0H+b region in MC, while the middle panel shows the comparison in data. The bottom panel shows the (unnormalized) ratio of the signal bin to the 0H+b region for both data and MC. Due to the limited statistics of the validation sample, we use this as a cross-check validation only.



| Region                   | MC Yields       | Region                   | Data Yields     |
|--------------------------|-----------------|--------------------------|-----------------|
| $A_{2H}$                 | $130 \pm 3$     | $A_{2H}$                 | $82 \pm 9$      |
| $A_{1H}$                 | $426 \pm 5$     | $A_{1H}$                 | $265 \pm 16$    |
| $\beta_{\text{norm},2H}$ | $125.4 \pm 1.6$ | $\beta_{\text{norm},2H}$ | $77 \pm 6$      |
| $\beta_{\text{norm},1H}$ | $440 \pm 5$     | $\beta_{\text{norm},1H}$ | $287 \pm 18$    |
| $\kappa_{2H}$            | $1.03 \pm 0.03$ | $\kappa_{2H}$            | $1.07 \pm 0.14$ |
| $\kappa_{1H}$            | $0.97 \pm 0.01$ | $\kappa_{1H}$            | $0.92 \pm 0.08$ |

| Region                   | MC Yields       | Region                   | Data Yields     |
|--------------------------|-----------------|--------------------------|-----------------|
| $A_{2H}$                 | $4 \pm 3$       | $A_{2H}$                 | $3 \pm 1.7$     |
| $A_{1H}$                 | $72 \pm 15$     | $A_{1H}$                 | $56 \pm 7$      |
| $\beta_{\text{norm},2H}$ | $1.11 \pm 0.18$ | $\beta_{\text{norm},2H}$ | $3.9 \pm 0.8$   |
| $\beta_{\text{norm},1H}$ | $53 \pm 4$      | $\beta_{\text{norm},1H}$ | $68.3 \pm 4.7$  |
| $\kappa_{2H}$            | $3.73 \pm 2.48$ | $\kappa_{2H}$            | $0.77 \pm 0.48$ |
| $\kappa_{1H}$            | $1.35 \pm 0.30$ | $\kappa_{1H}$            | $0.82 \pm 0.12$ |

Table 6.8: Region yields in MC (left) and data (right), in the single-lepton (top row) and single-photon (bottom row) validation samples. For the single-photon sample, the  $p_T^{\text{miss}}$  requirement is loosened to  $p_T^{\text{miss}} > 200$  GeV to increase the statistical precision.  $\kappa$  is the ratio of the true  $A_{iH}$  yields to the region's corresponding  $\beta_{\text{norm},iH}$ . The statistics are too low in the single-photon sample for a meaningful validation, so we use this as a cross-check only, and no systematic is assigned. All uncertainties are statistical only.

| Source                            | $p_T^{\text{miss}}$ (GeV) | Uncertainty (%) |    |
|-----------------------------------|---------------------------|-----------------|----|
|                                   |                           | 1H              | 2H |
| Sideband statistics               | > 300                     | 10              | 19 |
|                                   | [300, 500]                | 8               |    |
|                                   | [500, 700]                | 18              |    |
|                                   | > 700                     | 38              |    |
| $\kappa$ stat. unc.               | > 300                     | 4               | 17 |
|                                   | [300, 500]                | 2               | 13 |
| $p_T^{\text{miss}}$ shape closure | [500, 700]                | 3               | 19 |
|                                   | > 700                     | 9               | 63 |
|                                   | > 300                     | 9               | 13 |
| Bkg. composition                  | [300, 500]                | 3               |    |
|                                   | [500, 700]                | 7               |    |
|                                   | > 700                     | 32              |    |
| MC mismodeling                    | > 300                     | 4               | 6  |
|                                   | [300, 500]                | 1               |    |
|                                   | [500, 700]                | 2               |    |
|                                   | > 700                     | 5               |    |

Table 6.9: Summary of the systematic uncertainties in the background estimation for the boosted signature and comparison with sideband statistical uncertainties. The systematic uncertainties pertaining to  $\beta_{\text{norm}}$  appear in the rows labeled “>300 GeV” in the  $p_T^{\text{miss}}$  column. The row labeled  $\Delta(\text{data}, \text{MC})$  gives the contribution derived from the comparison of simulation with data in the one-lepton validation sample, and those labeled “MC mismodeling” refer to the corrections to simulation discussed in Section 6.5.3.

### 6.5.2.3 Summary of background systematics for the boosted method

Table 6.9 shows the total systematic uncertainties for each signal bin derived using the validation samples. The largest systematic uncertainties are found to be the data statistics of the analysis regions (ranging from 8-38%), the background composition for the highest  $p_T^{\text{miss}}$  bin (32%), and the  $p_T^{\text{miss}}$  shape closure (ranging from 2-63%), the highest  $p_T^{\text{miss}}$  bin again being the most affected. All other systematics fall within 1-17%.

### 6.5.3 Signal systematic uncertainties

We consider a variety of systematic uncertainties on the signal efficiency and distribution, as calculated from the signal MC.

- Luminosity: 2.3%-2.5%.** The integrated luminosity is estimated based on data from five subdetectors [111]. The subdetectors used are the pixel detector, the muon DT, the HF calorimeter, and two additional sensors, the Pixel Detector Telescope [112] and the Fast Beam Conditions Monitor diamond sensors [113]. Each subdetector has its own associated algorithm [111–114] that estimates the integrated luminosity to which our signal samples are scaled to. We must account for the uncertainty of the luminosity estimate. The calculated uncertainties are 2.5% (2016) [111], 2.3% (2017) [115], and 2.5% (2018) [116], respectively, for the different data-taking years.
- PF Jet ID: 1%.** The jet ID criteria defined in Section 6.3.1 determine how likely the jets are to be from noise. If a jet fails the jet ID requirement, it is removed from the event. The  $H_T$  and  $\Delta\phi$  variables are recalculated if necessary. In full simulation, we measure the jet ID efficiency to be  $\geq 99\%$ , and so we take a flat 1% as the systematic uncertainty.
- b-tag efficiency: <1%-7%.** As noted in Section 5, differences in b-tag efficiency between data and MC are corrected for by using scale factors, binned in jet  $p_T$  and  $\eta$ . For the systematic, the scale factors are varied by their uncertainties and these variations are propagated as migrations between the different region bins. The b-tagging and charm-mistagging are varied together, while the light flavor mistagging scale factors are varied separately. The effect of varying the b/c tagging scale factors ranges from 2-7%, while the effect of varying the light mistagging scale factors varies from 0-3%.
- b-tag FastSIM corrections: <1%.** Signal samples generated using the fast simulation package must be corrected to match the full simulation performance, including the b-tagging and mistagging performance. Separate correction factors are derived for b/charm-jets and light flavor jets, as functions of the jet  $p_T$  and  $\eta$  [117–119], and applied to the fast simulation samples. As with the scale factors above, the correction factors for b/c jets and light flavor jets are varied independently by their uncertainties, and these variations are propagated as migrations among the different signal bins.

- **bb-tag efficiency: 7%-15%.** The bb-tagging and mistagging scale factors account for differences in data compared to simulation and are computed as functions of the AK8 jet  $p_T$ . These scale factors are applied only in the boosted method. The tagging scale factors are provided centrally by a dedicated Physics Object Group at CMS [120–122]. The mistagging scale factors were computed as described in Section 5.3.2. For the signal systematic, the scale factors are varied by their uncertainty.
- **bb-tag FastSIM corrections: <1%.** The bb-tagging performance in the fast simulation must be corrected to match the full simulation. These correction factors and associated uncertainties were computed by comparing the tagging efficiencies for truth-matched  $H \rightarrow b\bar{b}$  jets in simulated signal samples and applied only in the boosted method. The signal systematic uncertainties are derived by varying the scale factors by their uncertainties.
- **AK8 jet mass FastSIM correction: 2%-4%.** The AK8 jet mass performance in the fast simulation must be corrected to match the full simulation. These correction factors and associated uncertainties were computed by comparing the efficiency of the Higgs mass window selection (95-145 GeV) between simulated signal samples, using both full and fast simulation. The signal systematic uncertainties are derived by varying the scale factors by their uncertainties. Both the scale factors and their associated signal systematic uncertainties are applied to the boosted method only.
- **$p_T^{\text{miss}}$  mismeasurement: <1%-13%.** A correction is applied to the  $p_T^{\text{miss}}$  variable to account for a discrepancy observed in the fast simulation compared to the full simulation. Reconstructed  $p_T^{\text{miss}}$  distributions are found to be too wide compared to the full simulation. To correct for this, the region yields using the reconstructed and generator-level  $p_T^{\text{miss}}$  values are averaged and used. The systematic uncertainty on this correction is assigned to be half the difference in the region yields for when the reconstructed  $p_T^{\text{miss}}$  and the generator-level  $p_T^{\text{miss}}$  are used.

- **Signal sample statistics: <1%-18%.** After event selection, the absolute number of simulated events generated is treated as an independent systematic uncertainty for each signal bin.
- **Trigger efficiency: <1%-13%.** The trigger efficiencies are measured in the data as a function of  $p_T^{\text{miss}}$  and  $H_T$  (see Section 6.2.1). The nominal efficiencies are varied up and down by the associated bin uncertainty for the signal systematic. The trigger efficiency uncertainty is treated as fully correlated across the search bins.
- **Pileup reweighting: <1%-9%.** We reweight the pileup distributions in simulation to more closely match those in data. A truth pileup distribution is generated by combining luminosity measurements with the best measured value for the minimum bias cross section. The minimum bias cross section is then varied within uncertainties, and these variations are propagated as migrations between the different signal bins.
- **Prefiring correction: <1%.** Due to an ECAL timing error which propagated to the L1 trigger primitives, events with significant ECAL energy in the region  $2 < \eta < 3$  self-vetoed in 2016 and 2017 data. A recipe was centrally-produced to create event weights that model this issue in the corresponding MC. This correction and its associated uncertainty are propagated to the signal yields, and only applied to the 2016 and 2017 samples.
- **ISR: <1%-18%.** We correct our signal samples for the distribution of ISR jets. The weights are normalized to maintain the total cross sections of the signal samples. For the strong production signal samples, the correction is binned in the number of ISR jets: 1.00, 0.920, 0.821, 0.715, 0.662, 0.561, and 0.511 for  $N_{\text{ISR}} = 0, 1, 2, 3, 4, 5, 6+$ . We take 50% of the deviation from unity as the systematic uncertainty for the strong production. For the electroweak signal samples, we reweight the  $p_T^{\text{miss}}$  of the system of SUSY particles,  $p_T^{\text{ISR}}$ . The reweighting factors range between 1.18 for  $p_T^{\text{ISR}}=125$  GeV and 0.78 for  $p_T^{\text{ISR}} > 600$  GeV. We take the deviation from 1.0 as the systematic uncertainty. The corrections are

applied to the simulated signal events as event weights and are varied according to their uncertainties for the signal systematic.

- **Theory scales: <1%-7%.** The parton distribution function of the proton depends on two theory scales, the renormalization scale  $\mu_R$  and the factorization scale  $\mu_F$ . The uncertainty is calculated using the envelope of the weights from varying these scales by a factor of 2 [123,124].
- **Jet energy corrections: <1%-12%.** We apply jet energy corrections (JECs) to scale the jet response to that seen in data. These corrections, typically <1%-20% in size, are dependent on the jet  $p_T$  and  $\eta$ . We propagate the JEC variations and recalculate any jet-dependent variables, including  $H_T$ ,  $p_T^{\text{miss}}$ , and  $\Delta\phi$ .
- **Jet energy resolution: <1%-7%.** The jet momenta in MC samples are smeared to match the jet energy resolution (JER) in data, to correct the simulated detector response. The smearing factors are varied according to the uncertainties on the jet energy resolution measurements. These variations are propagated into the various jet-dependent variables, including  $H_T$ ,  $p_T^{\text{miss}}$ , and  $\Delta\phi$ .
- **AK8 jet mass resolution: <1%-9%.** The AK8 jet mass differs slightly between simulation and data. We measured the mean and width of the W boson peak in data and in MC. The width of the peak is consistent between data and simulation. For the signal systematic, the mass value for each AK8 jet is shifted up by 2.6 GeV, which accounts for the difference in the mean between simulation and data, scaled to the Higgs mass.

## 6.6 Results

We perform the data-driven background estimations described in Section 6.4 first with the signal bins blinded, so as to not bias our results. After the background prediction is calculated, we unblind the signal bins and compare to the SM prediction. In this section, we include distributions

| Source                                   | Relative uncertainty (%) |         |
|------------------------------------------|--------------------------|---------|
|                                          | Resolved                 | Boosted |
| Signal sample statistics                 | 0–18                     | 1–15    |
| ISR modeling                             | 0–2                      | 0–18    |
| Theory scales $\mu_R$ and $\mu_F$        | 0–2                      | 0–7     |
| Pileup reweighting                       | 0–2                      | 0–9     |
| Integrated luminosity                    |                          | 2.4     |
| Jet energy scale                         | 0–7                      | 0–12    |
| Jet energy resolution                    | 0–7                      | 0–7     |
| Trigger efficiency                       | 1–13                     | 0–4     |
| Prefiring correction                     | 0–1                      |         |
| AK8 jet mass resolution                  | —                        | 0–9     |
| btagging efficiency                      | 2–7                      | —       |
| bmistagging                              | 1–3                      | —       |
| bb tagging efficiency                    | —                        | 6–15    |
| Uncertainties due to the fast simulation |                          |         |
| Jet ID                                   |                          | 1       |
| $p_T^{\text{miss}}$ modeling             | 0–13                     | 0–12    |
| AK8 jet mass resolution                  | —                        | 2–4     |
| btagging efficiency                      | 0–1                      | —       |
| bmistagging                              | 0–1                      | —       |
| bb tagging efficiency                    | —                        | < 1     |

Table 6.10: List of sources of systematic uncertainties and their typical impact on the signal. The range is reported as the median 68%, accounting for all signal regions for every signal mass point considered. Entries reported as 0 correspond to values smaller than 0.5%. Low statistics bins dominate the tails.

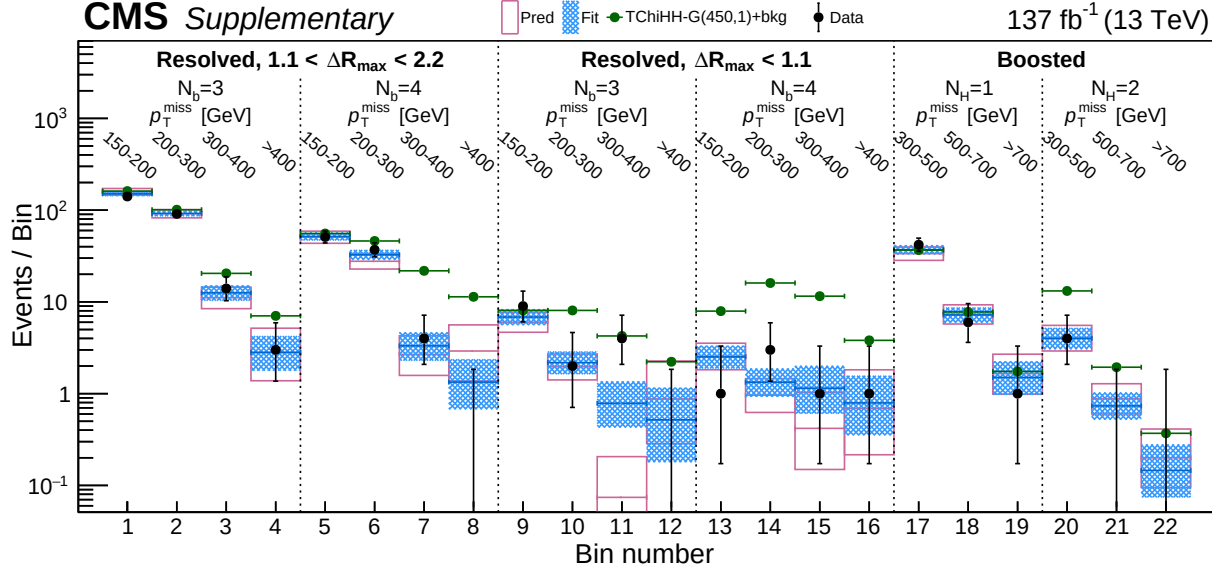


Figure 6.24: Observed and predicted yields in the search regions. The points with error bars represent the observed yields, the magenta outline bands the predicted background yields (“Pred”) with their total uncertainties derived as described in Sections 6.4 and 6.5, and the blue shaded bands the values determined by the background-only fit (“Fit”). The green lines are the sum of the yields from the data background prediction and the yields of the TChiHH-G(450,1) model with signal strength 1, motivating what we would have expected to see had there been a signal.

of several key analysis variables in data and compare them with the corresponding distributions from simulated event samples. It is important to note, however, that while the agreement is generally good, the analysis does not require good data to MC agreement. As shown in Figure 6.24, which summarizes the observed and predicted SM yields, no significant deviation in 21 of the 22 signal bins with the SM expectation is observed. One of the bins of the resolved signature shows an excess, but after careful examination we find no exceptional features. Further details, as well as the observed and predicted yields, are discussed in Sections 6.6.1 and 6.6.2 for the resolved and boosted signatures, respectively. Section 6.6.3 describes the statistical treatment of the analysis, and how we quantify the compatibility of the data with a signal hypothesis. We then interpret our data using three simplified SUSY models, and set limits on the production cross sections, as discussed in Section 6.6.4.



### 6.6.1 Resolved observed data yields

The observed distributions in  $\langle m_{bb} \rangle$  for the resolved signature are shown in Figures 6.25 and 6.26, in the  $N_b = 3$  and  $N_b = 4$  regions (SR+SB). The corresponding distributions in the  $N_b = 2$  region are overlaid, reweighted event-by-event with  $\kappa$  values and normalized to the observed event yields of the signal bin. Overall, we observe fairly good agreement between the  $N_b = 3, 4$   $\langle m_{bb} \rangle$  distributions and the  $N_b = 2$  distribution, and we do not observe a large excess in the Higgs mass window.

The observed yields and predicted SM background are given in Table 6.11. For 15 of the 16 bins of the resolved signature, deviations with respect to SM predictions are consistent with statistical fluctuations. In the bin corresponding to  $\Delta R_{\max} < 1.1$ ,  $N_b = 3$ ,  $300 < p_T^{\text{miss}} < 400$  GeV, we observe 4 events in data while the prediction is  $0.07^{+0.13}_{-0.05}$ . Small event counts in the  $N_b = 3$  SB and  $N_b = 2$  SR lead to very asymmetric confidence bands and are the primary driver for a lower-than-anticipated prediction. We have examined the four observed events in detail and find no exceptional features. We have also generated and studied several pseudo-experiments to determine the probability of observing 4 or more events given this background estimate. Lastly, the models we consider for potential contributions beyond the SM predict that events would be distributed over multiple bins, contrary to our observations here.

### 6.6.2 Boosted observed data yields

The observed distributions in  $m_J$  and in  $(m_{J1}, m_{J2})$  for the boosted signature are shown in Figure 6.27, with simulated SM backgrounds overlaid, for the  $N_H = 0, 1$ , and 2 regions, integrated in  $p_T^{\text{miss}}$ . The MC is scaled to the integral of the observed data, to facilitate by-eye shape comparisons. Agreement between the data and MC is not necessary, since our background estimation methods are data-driven. Included in the figures are representative distributions for potential SUSY signals, with their peaks well within the H boson mass window; we do not observe an excess in the data at the Higgs mass. These figures show statistical uncertainties only, as systematic uncertainties are

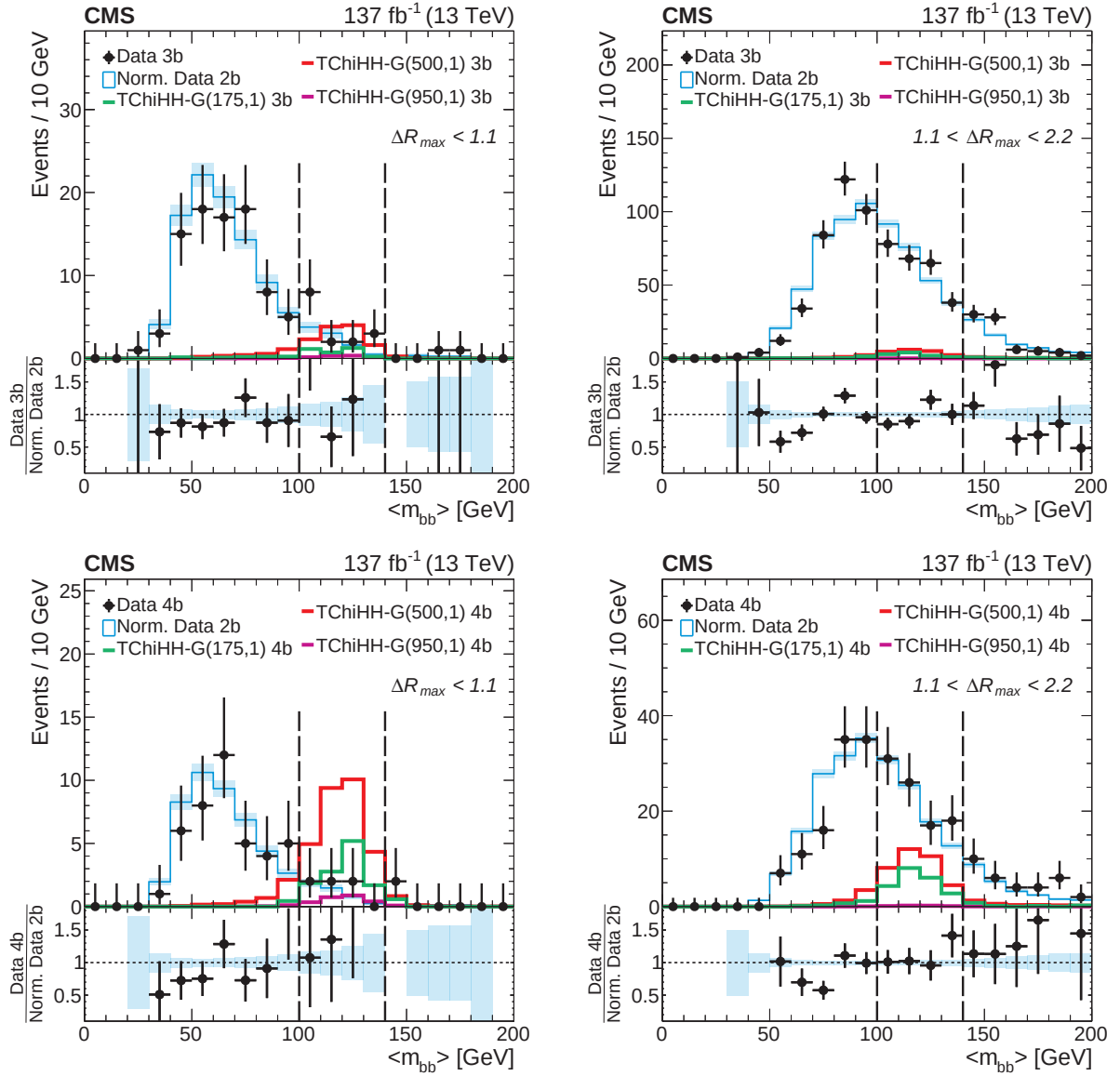
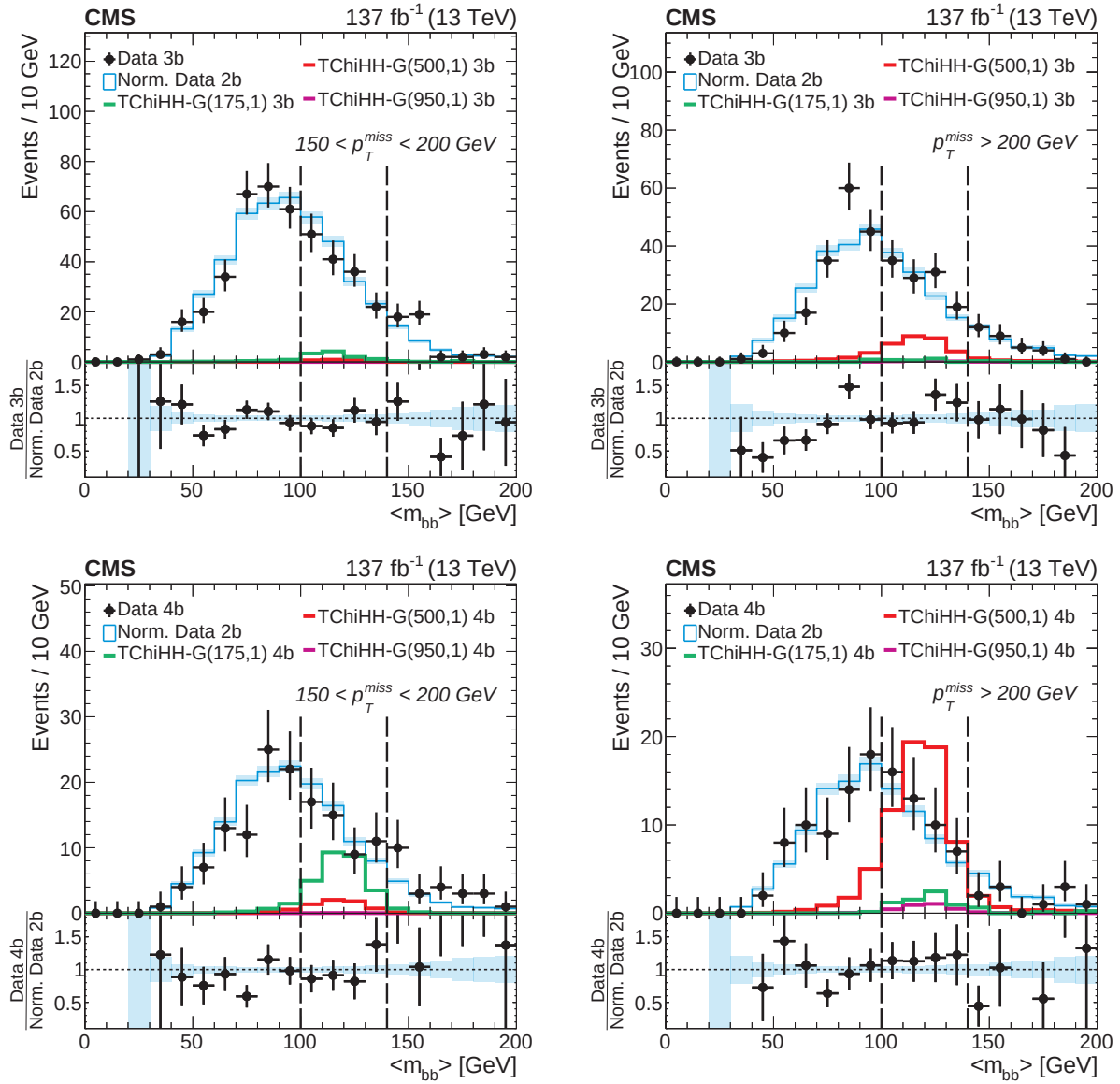


Figure 6.25: Distributions in  $\langle m_{bb} \rangle$  for the (upper row)  $N_b = 3$  and (lower row)  $N_b = 4$  data, denoted by black markers, with error bars indicating the statistical uncertainties. The overlaid cyan histograms show the corresponding distributions of the  $N_b = 2$  data, reweighted with the appropriate  $\kappa$  values and scaled in area to the  $N_b = 3$  or  $N_b = 4$  distribution and with statistical uncertainties indicated by the cyan shading. The ratio of these distributions appears in the lower panel. The red, green, and violet histograms show simulations of representative signals, denoted TChiHH-G( $m(\tilde{\chi}_1^0), m(\tilde{G})$  [GeV]) in the legends. The left-hand set of plots corresponds to  $\Delta R_{\max} < 1.1$ , while the right-hand set corresponds to  $1.1 < \Delta R_{\max} < 2.2$ , in both cases integrated over  $p_T^{\text{miss}}$ .



| $\Delta R_{\text{max}}$ | $N_b$ | $p_T^{\text{miss}}$ [GeV] | $\kappa$                        | $N_{\text{SR}}^{\text{pred}}$ | $N_{\text{SR}}^{\text{fit}}$ | $N_{\text{SR}}^{\text{obs}}$ |
|-------------------------|-------|---------------------------|---------------------------------|-------------------------------|------------------------------|------------------------------|
| 1.1–2.2                 | 3     | 150–200                   | $1.07 \pm 0.04 \pm 0.02$        | $159 \pm 13$                  | $150.5^{+8.8}_{-8.5}$        | 141                          |
|                         |       | 200–300                   | $0.92 \pm 0.04 \pm 0.02$        | $91.3^{+9.8}_{-9.0}$          | $92.1^{+6.9}_{-6.5}$         | 91                           |
|                         |       | 300–400                   | $0.94^{+0.09}_{-0.08} \pm 0.01$ | $11.1^{+3.3}_{-2.7}$          | $12.5^{+2.6}_{-2.2}$         | 14                           |
|                         |       | > 400                     | $0.99^{+0.20}_{-0.17} \pm 0.02$ | $2.8^{+2.4}_{-1.4}$           | $2.8^{+1.4}_{-1.0}$          | 3                            |
|                         | 4     | 150–200                   | $1.07 \pm 0.06 \pm 0.10$        | $50.7^{+8.2}_{-7.2}$          | $51.7^{+5.4}_{-5.0}$         | 51                           |
|                         |       | 200–300                   | $0.93 \pm 0.07 \pm 0.08$        | $27.7^{+5.7}_{-4.9}$          | $32.6^{+4.3}_{-3.9}$         | 37                           |
|                         |       | 300–400                   | $0.91^{+0.16}_{-0.15} \pm 0.07$ | $2.7^{+1.6}_{-1.1}$           | $3.3^{+1.3}_{-1.0}$          | 4                            |
|                         |       | > 400                     | $1.03^{+0.32}_{-0.26} \pm 0.09$ | $2.9^{+2.7}_{-1.5}$           | $1.34^{+1.03}_{-0.66}$       | 0                            |
| < 1.1                   | 3     | 150–200                   | $1.25 \pm 0.13 \pm 0.15$        | $6.0^{+1.7}_{-1.3}$           | $6.9^{+1.5}_{-1.3}$          | 9                            |
|                         |       | 200–300                   | $1.08^{+0.14}_{-0.13} \pm 0.12$ | $1.97^{+0.74}_{-0.56}$        | $2.18^{+0.70}_{-0.55}$       | 2                            |
|                         |       | 300–400                   | $0.85^{+0.36}_{-0.25} \pm 0.10$ | $0.07^{+0.13}_{-0.05}$        | $0.78^{+0.58}_{-0.36}$       | 4                            |
|                         |       | > 400                     | $1.23^{+0.64}_{-0.44} \pm 0.10$ | $0.87^{+1.30}_{-0.58}$        | $0.50^{+0.62}_{-0.33}$       | 0                            |
|                         | 4     | 150–200                   | $1.23^{+0.20}_{-0.19} \pm 0.21$ | $2.55^{+0.98}_{-0.74}$        | $2.52^{+0.81}_{-0.64}$       | 1                            |
|                         |       | 200–300                   | $1.10^{+0.21}_{-0.19} \pm 0.19$ | $0.96^{+0.49}_{-0.34}$        | $1.33^{+0.55}_{-0.41}$       | 3                            |
|                         |       | 300–400                   | $2.40^{+0.91}_{-0.75} \pm 0.32$ | $0.42^{+0.61}_{-0.27}$        | $1.15^{+0.86}_{-0.54}$       | 1                            |
|                         |       | > 400                     | $0.96^{+0.55}_{-0.37} \pm 0.12$ | $0.65^{+2.29}_{-0.44}$        | $0.65^{+1.18}_{-0.31}$       | 1                            |

Table 6.11: For each SR of the resolved signature, the MC correction factor  $\kappa$ , predicted background yield  $N_{\text{SR}}^{\text{pred}}$ , yield from the background-only ( $\mu = 0$ ) fit  $N_{\text{SR}}^{\text{fit}}$ , and observed yield  $N_{\text{SR}}^{\text{obs}}$ . The first and second uncertainties in the  $\kappa$  factors are statistical and systematic, respectively. The uncertainties in  $N_{\text{SR}}^{\text{pred}}$  and  $N_{\text{SR}}^{\text{fit}}$ , extracted from the maximum likelihood fit, include both statistical and systematic contributions. The interpretation of results in the bin corresponding to  $\Delta R_{\text{max}} < 1.1$ ,  $N_b = 3$ ,  $300 < p_T^{\text{miss}} < 400$  GeV is discussed in the text.

| $N_H$ | $p_T^{\text{miss}}$ [GeV] | $N_{\text{bkg}}^{\text{tot}}$ | $f_{\text{bkg}}$  | $N_{\text{SR}}^{\text{pred}}$ | $N_{\text{SR}}^{\text{fit}}$ | $N_{\text{SR}}^{\text{obs}}$ |
|-------|---------------------------|-------------------------------|-------------------|-------------------------------|------------------------------|------------------------------|
| 1     | 300–500                   | $42.6 \pm 4.2$                | $0.789 \pm 0.030$ | $33.6^{+6.1}_{-5.2}$          | $37.0^{+4.2}_{-4.0}$         | 42                           |
|       | 500–700                   |                               | $0.172 \pm 0.028$ | $7.3^{+2.0}_{-1.6}$           | $7.2^{+1.5}_{-1.3}$          | 6                            |
|       | > 700                     |                               | $0.039 \pm 0.014$ | $1.65^{+1.04}_{-0.66}$        | $1.50^{+0.75}_{-0.53}$       | 1                            |
| 2     | 300–500                   | $5.1 \pm 1.0$                 | $0.789 \pm 0.030$ | $4.0^{+1.5}_{-1.1}$           | $4.0^{+1.2}_{-1.0}$          | 4                            |
|       | 500–700                   |                               | $0.172 \pm 0.028$ | $0.88^{+0.40}_{-0.28}$        | $0.74^{+0.29}_{-0.21}$       | 0                            |
|       | > 700                     |                               | $0.039 \pm 0.014$ | $0.20^{+0.21}_{-0.10}$        | $0.14^{+0.13}_{-0.07}$       | 0                            |

Table 6.12: For each  $N_H$  SR of the boosted signature, the total predicted background yield  $N_{\text{bkg}}^{\text{tot}}$ , and for each  $p_T^{\text{miss}}$  bin the fraction  $f_{\text{bkg}}$ , both with their statistical uncertainties, the predicted background yield  $N_{\text{SR}}^{\text{pred}}$ , the yield from the background-only ( $\mu = 0$ ) fit  $N_{\text{SR}}^{\text{fit}}$ , and the observed yield  $N_{\text{SR}}^{\text{obs}}$ . The values of  $N_{\text{SR}}^{\text{pred}}$  and  $N_{\text{SR}}^{\text{fit}}$  are extracted from the maximum likelihood fit, with uncertainties that include both statistical and systematic contributions.

computed only for the total yields in the analysis regions.

The observed yields and predicted SM background are given in Table 6.12. The deviations of the observations with respect to SM predictions are consistent with statistical fluctuations for all 6 signal bins of the boosted signature.

### 6.6.3 Statistical analysis

Interpreting our results in the context of our simplified models, the absence of signal allows us to place upper limits on the production cross sections. We express how incompatible our data with the background plus signal (s+b) hypothesis relative to the background-only (b) hypothesis as a 95% confidence level (CL). For setting limits, we use the LHC-style  $\text{CL}_s$  approach, which is a modified frequentist method [125–128]. We construct the likelihood function in each bin, a product of Poisson probabilities to observe  $N$  events. The test statistic  $q_\mu$  is used to determine how likely our s+b hypothesis is, with

$$q_\mu = -2 \ln \mathcal{L}_\mu / \mathcal{L}_{\text{max}}, \quad (6.13)$$

for a given signal strength  $\mu$ .  $\mathcal{L}_\mu$  is the likelihood for signal strength  $\mu$ , and  $\mathcal{L}_{\text{max}}$  is the maximum likelihood, if all parameters are allowed to vary. For setting limits, a signal point is excluded when the 95% CL upper limit on  $\mu$  reaches one.

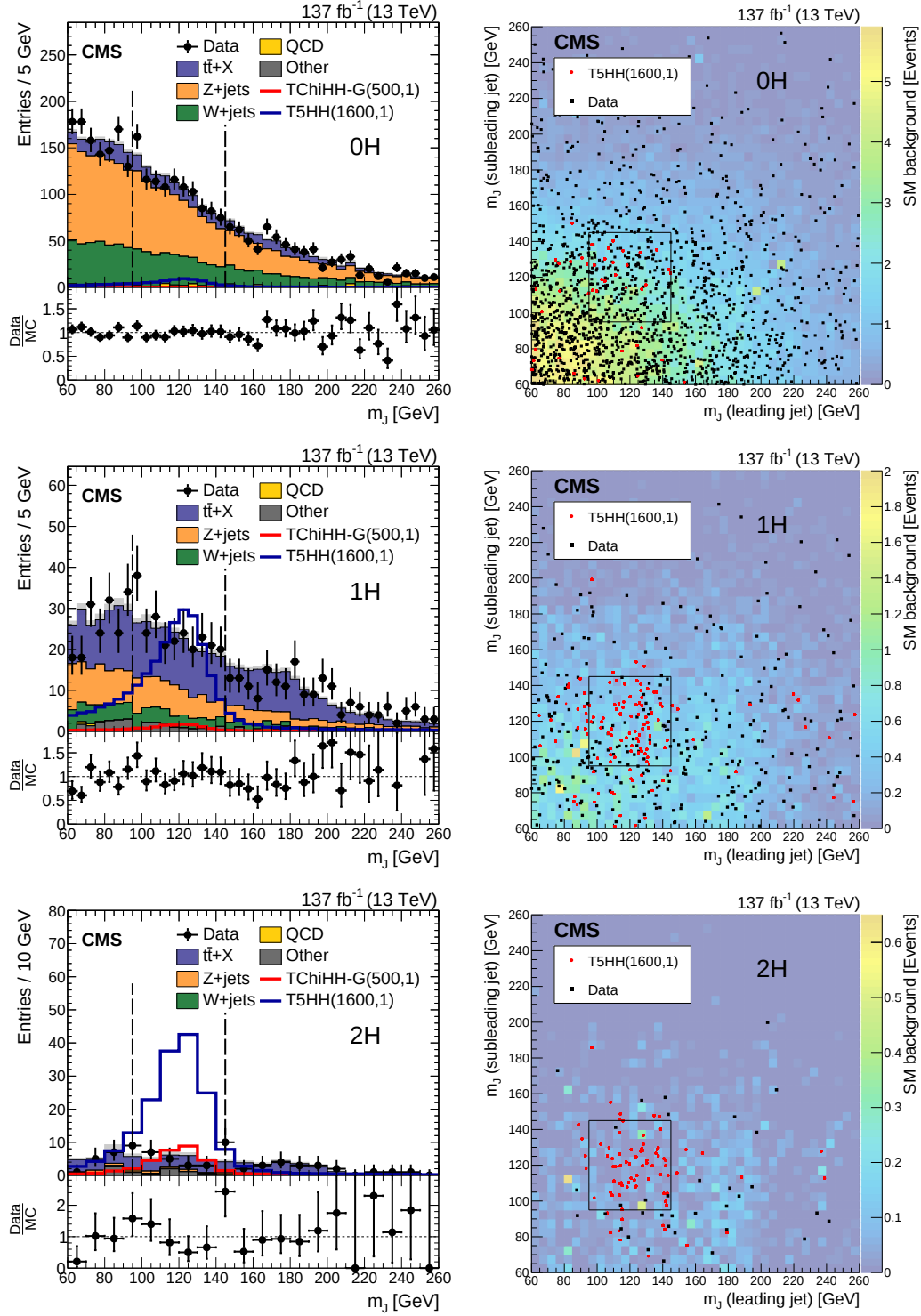


Figure 6.27: Distributions in  $m_J$  for the boosted signature, integrated in  $p_T^{\text{miss}}$ . The left column contains two entries per event. The SM yields are scaled to the data integral, by factors of 0.84, 0.92, and 1.13 in the 0H (upper), 1H (middle), and 2H (lower) plots, respectively. The data to simulation ratio appears in each lower panel. Two simulated signal samples are also shown. In the correlation plots (right column) the color scale represents the SM background, black dots the data, and red dots the expected signal.

We use the Higgs Combination tool [129] to simultaneously fit all signal and sideband regions. The yields in each region are encoded as Poisson observables, with appropriate constraints (Equations 6.5 and 6.6) among them to implement the background estimation methods. Signal contamination in the control regions is accounted for by directly including the relevant signal yields for all analysis regions. All systematic uncertainties, for both the backgrounds methods and signal samples, enter the statistical interpretation as log-normal nuisance parameters. Additionally, the  $\kappa$  factors are treated as log-normal nuisance parameters. The signal uncertainties are modeled with a single value for all signal points and are correlated across all signal bins.

We can use pseudo-datasets with observation from a Poisson probability as our background, and a given signal strength. If we wish to exclude the signal at a 95% CL, we can require the fraction of these pseudo-datasets that are less than  $q_\mu$  to be smaller than 5%, that is:

$$\frac{N(q_\mu^{\text{toy}} < q_\mu)}{N_{\text{toys}}} = \text{CL}_{s+b}, \quad \text{CL}_{s+b} < 0.05. \quad (6.14)$$

For the  $\text{CL}_s$  method, we are also interested in the fraction of these pseudo-datasets with no signal strength ( $\mu=0$ ):

$$\frac{N(q_{\mu=0}^{\text{toy}} > q_{\mu=0})}{N_{\text{toys}}} = \text{CL}_b. \quad (6.15)$$

The CLs ratio is the ratio of these probability values,

$$\text{CL}_s = \frac{\text{CL}_{s+b}}{\text{CL}_b}, \quad (6.16)$$

where  $\text{CL}_b$  ( $\text{CL}_{s+b}$ ) is the p-value for the data fit to the background only (signal plus background) hypothesis. The asymptotic approximation is used for limit setting.

#### 6.6.4 Simplified model interpretations

We set upper limits on the simplified model production cross sections, which are evaluated at 95% confidence level (CL). This is a measure of how incompatible our data with the s+b hypothesis relative to the b-only hypothesis is. These limits are shown in Figures 6.28-6.31.

Figure 6.28 shows the upper limits across different  $\tilde{\chi}_1^0$  masses, with the lightest SUSY particle (in this model, the  $\tilde{G}$ ) fixed at  $m_{\tilde{G}} = 1$  GeV, for the combined approach. This is for the scenario

of all possible higgsino productions. The observed limits at lower masses are generally weaker than expected due to the excess of observed events over the predicted background in the resolved signature, while the observed limits are generally stronger than expected at higher masses due to the deficit of observed events in the boosted signature. In all cases, the observed limits are within  $2\sigma$  of the expectation. The data exclude  $\tilde{\chi}_1^0$  masses from 175 to 1025 GeV. This represents the best limit to date on this model, extending previous CMS results [130], which exclude 230–770 GeV, and previous ATLAS results [131], which exclude mass ranges 130–230 GeV and 290–880 GeV.

The resolved and boosted signatures are shown separately in the bottom of Figure 6.29, to clearly indicate the sensitivity of each signature. As expected, the resolved method provides most of the sensitivity at lower NLSP masses, where the energy available in the  $\tilde{\chi}_1^0 \rightarrow \tilde{G}H$  decay is limited. At higher NLSP masses, the Higgs bosons receive a larger Lorentz boost, and the b quarks from the decay tend to merge; the crossover where the boosted signature is more sensitive is about 800 GeV for the expected limits, and 600 GeV for the observed limits since the boosted signal bins show no excess in the data. This also shows how complimentary the two signatures are, extending the exclusion reach further than either alone could have accomplished.

Figure 6.30 shows the upper limits in the two-dimensional mass plane of the higgsino  $\tilde{\chi}_2^0$  and the lightest SUSY particle  $\tilde{\chi}_1^0$ , for the combined approach, where only the  $\tilde{\chi}_2^0\tilde{\chi}_3^0$  production is included. The resolved signature fully drives the exclusion power, since the boosted method has no sensitivity at low masses. While the expected limit excludes a substantial region, the observed exclusion is limited to a small region where  $m(\tilde{\chi}_1^0)$  is less than 15 GeV, again because of the single-bin excess. For  $m(\tilde{\chi}_1^0) = 1$  GeV the excluded range of  $m(\tilde{\chi}_2^0)$  is 265–305 GeV.

The 95% CL upper limit across different  $\tilde{g}$  masses is shown in Figure 6.31. For this model, gluino pair production with a  $\tilde{\chi}_2^0$  NLSP slightly less massive than the gluino and a light LSP, masses  $m(\tilde{g}) < 2330$  GeV are excluded. This is the strongest mass limit for this model to date, extending a previous CMS result [132], which excluded  $m(\tilde{g}) < 2010$  GeV. The choice of large NLSP mass, which leads to energetic H bosons, lends most of the exclusion power to the boosted signature.



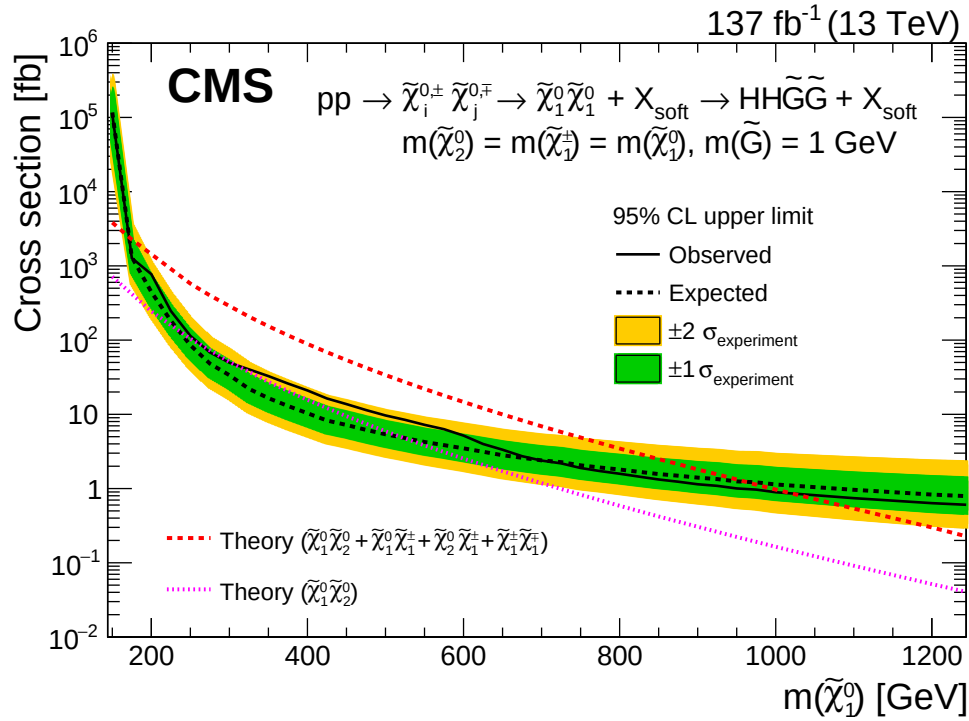


Figure 6.28: Observed and expected upper limits at 95% CL for TChiHH-G. The dashed black line and the green and yellow bands show the expected limit with its 1- and 2-standard deviation uncertainties, while the solid black line shows the observed limit. The theoretical cross section is indicated by the dashed red line for the degenerate set of all charginos and neutralinos, and by the dotted magenta line under the assumption that only the combination  $\tilde{\chi}_1^0 \tilde{\chi}_2^0$  contributes.

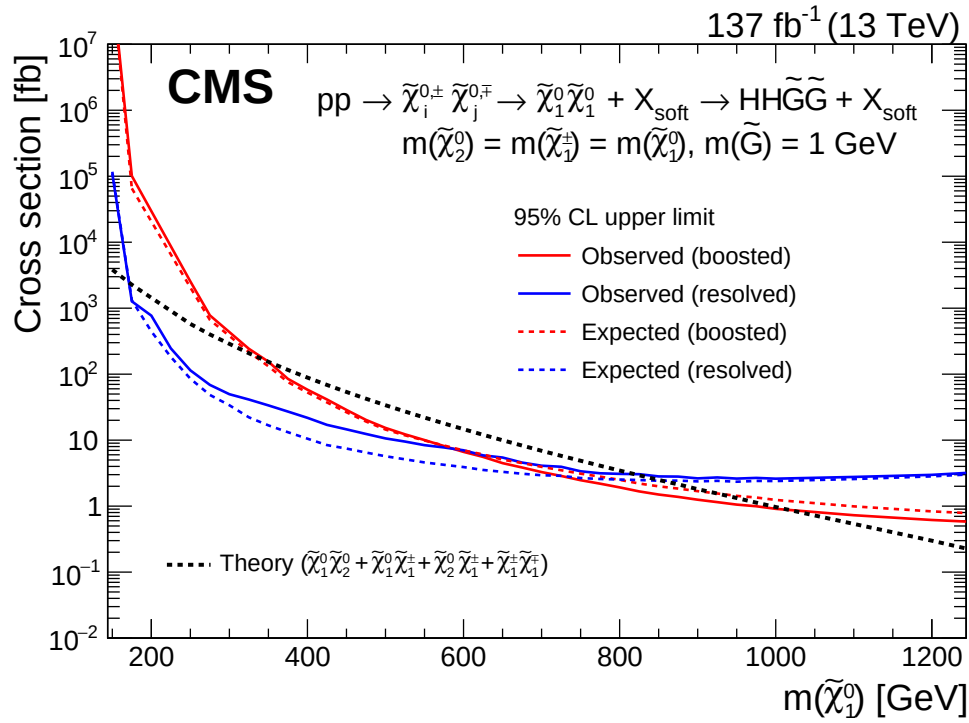


Figure 6.29: Observed (solid lines) and expected (dashed lines) upper limits at 95% CL for TChiHH-G for the resolved (red) and boosted (blue) signatures separately, to indicate the sensitivity of each method. The dashed black line shows the theoretical cross section computed under the assumption that the decay chains leading to the  $\tilde{\chi}_1^0 \tilde{\chi}_1^0$  intermediate state include a degenerate set of all charginos and neutralinos. No overlapping events were removed between the two methods.

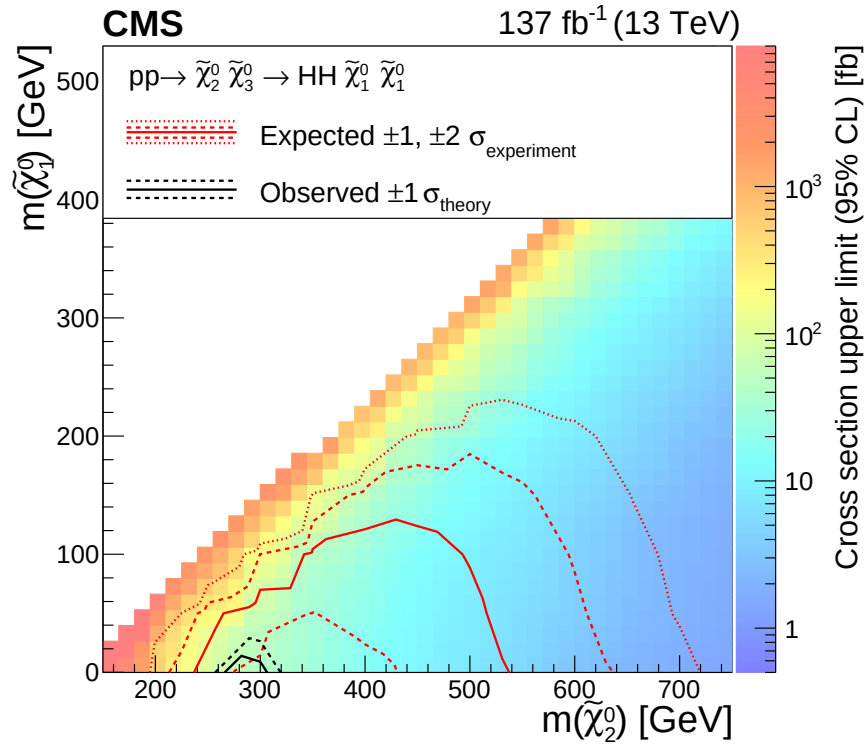


Figure 6.30: Limits at 95% CL on the cross section for the TChiHH signal model in which only the production of the intermediate state  $\tilde{\chi}_2^0 \tilde{\chi}_3^0$  is considered. The intersection of the expected upper limit (red) and observed upper limit (black) are drawn as exclusion curves, with their 1- (2-)standard deviation uncertainties shown in dashed (dotted) contours.

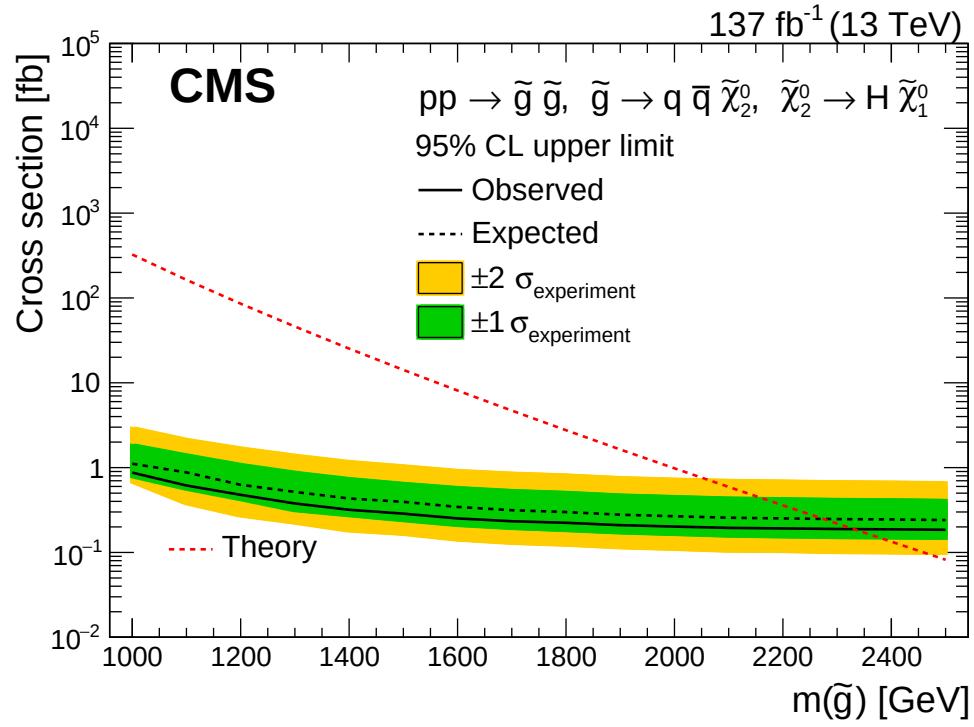


Figure 6.31: Observed (solid black) and expected (dashed black line and green and yellow bands) upper limits at 95% CL for T5HH. The dashed red line indicates the theoretical cross section.

## Chapter 7

### Electroweak combined search

Due to different production and decay modes, a SUSY signal could result in different final states, motivating a combined search with increased exclusion sensitivity. In this chapter, we describe a statistical combination of several CMS searches for electroweakino (EWK) production of SUSY particles [48–52, 89], including the search described above in Chapter 6 (hereafter “4b”). The EWK Combination is still an active area of work and has not yet been published. The results of this combined search are interpreted using simplified SUSY models, and are still preliminary at the time of writing. For these simplified models, we assume all sparticles other than those directly involved in the specified process have too large masses to be produced or are too large to affect the decays of the particles of interest.

The EWK Combination targets the pair production of the lightest neutralino ( $\tilde{\chi}_1^0$ ), each decaying into a Z or H boson and an effectively massless gravitino  $\tilde{G}$  (considered the LSP), as demonstrated in Figure 7.1. We consider three topologies, where each topology is denoted by the specific bosons produced  $\tilde{\chi}_1^0 \rightarrow Z\tilde{G}$  (ZZ topology),  $\tilde{\chi}_1^0 \rightarrow H\tilde{G}$  (HH), and a mixed topology of 50% of each of these decays (HZ). The bosons are expected to have SM-like properties and decay according to their SM branching fractions. Similar to the TChiHH-G model described in Section 6.1, these processes are described by a specific gauge-mediated SUSY breaking (GMSB) model with quasidegenerate higgsinos as the next-to-lightest SUSY particles. The production cross sections for the GMSB scenario are computed in the limit of mass-degenerate higgsino states  $\tilde{\chi}_1^\pm$ ,  $\tilde{\chi}_2^0$ , and  $\tilde{\chi}_1^0$ . The statistical combinations for some other relevant electroweak SUSY production and

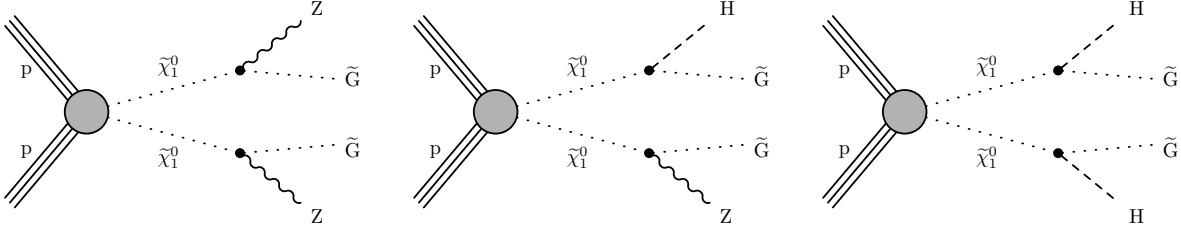


Figure 7.1: Pair production of  $\tilde{\chi}_1^0 \tilde{\chi}_1^0$  assuming the GMSB model. The two  $\tilde{\chi}_1^0$  particles decay into the  $\tilde{G}$  and (left) both to Z bosons, (center) a Z and a H boson, or (right) both to H bosons.

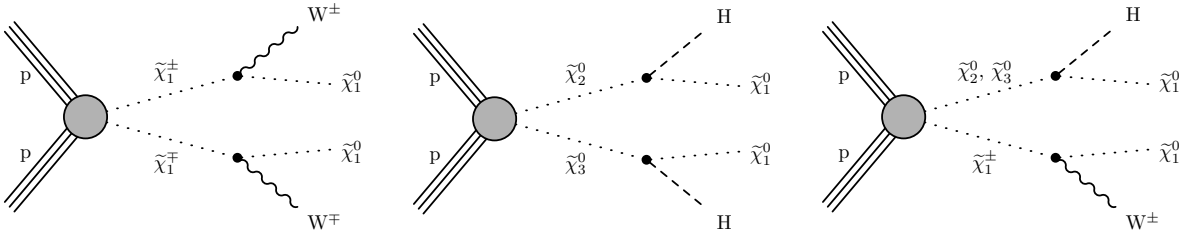


Figure 7.2: Production and decays modes used for the higgsino-bino interpretations, showing (left) the production of a pair of charginos followed by decays to W bosons and the LSP, (middle) the production of a pair of neutralinos followed by decays to H bosons and the LSP and (right) the production of chargino-neutralinos pairs followed by charginos (neutralinos) decaying to a W (H) boson and the LSP.

decay channels are given in Appendix A.

Additionally, the Combination targets a second model that assumes a bino-like LSP and the existence of a mass-degenerate higgsino triplet, considering the following production mechanisms:  $\tilde{\chi}_1^\pm \tilde{\chi}_1^0$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_1^0$ , and  $\tilde{\chi}_3^0 \tilde{\chi}_2^0$ . All electroweakinos involved in the production mechanisms are assumed to have higgsino-like couplings. The charginos decay to a W boson and the LSP with 100% branching fraction, while the neutralinos are assumed to decay to an H boson and the LSP with 100% branching fraction. The production mechanisms and decays modes for the higgsino-bino model are shown in Figure 7.2. This interpretation is the same simplified model as TChiHH described in Section 6.1, and is particularly nice since the other five analyses are sensitive to the higgsino production modes ( $\tilde{\chi}_2^0 \tilde{\chi}_1^\pm$ ,  $\tilde{\chi}_3^0 \tilde{\chi}_1^\pm$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ ) not considered by the “4b” analysis described in Chapter 6.

## 7.1 Individual searches

Table 7.1 summarizes which searches are used to place exclusion limits for each of the HH/HZ/ZZ topologies for the GMSB scenario and the WW/WH/HH topologies for the higgsino-bino interpretations. The event selection and the Monte Carlo (MC) simulated samples used are detailed in the corresponding references [48–52, 89]. The selections for all searches were checked to be mutually exclusive to ensure no events are double-counted. No significant deviations from the SM predictions were observed in any of the searches.

| Search        | GMSB scenario |    |    | Higgsino-bino |    |    |
|---------------|---------------|----|----|---------------|----|----|
|               | ZZ            | ZH | HH | WW            | WH | HH |
| 1 $\ell$ 2b   |               |    |    |               | ✓  |    |
| 4b            |               |    | ✓  |               |    | ✓  |
| 2 $\ell$ on-Z | ✓             | ✓  |    |               |    |    |
| > 3 $\ell$    | ✓             | ✓  | ✓  |               | ✓  |    |
| Hadr. WX      |               |    |    | ✓             | ✓  |    |

Table 7.1: Summary of the experimental searches considered in the combination and the signal topologies for which each search is used in the combined results. Table A.1 in Appendix A shows additional interpreted models.

### 7.1.1 Search for one lepton and two b jets

This search (“1 $\ell$ 2b”), described in Ref. [49], targets the WH topology. The analysis focuses on events with one charged lepton produced in a W boson decay, two b-tagged jets from a Higgs boson decay, and significant missing energy from the two LSPs and the neutrino. The search benefits from the large branching fraction for  $H \rightarrow b\bar{b}$  (58%) and the simpler background afforded by the leptonic final state. The two b-tagged jets are required to have an invariant mass consistent with the Higgs boson (90–150 GeV), as well a cotransverse mass incompatible with that of a pair of b jets resulting from top quark pair production. The transverse mass of the lepton and  $\vec{p}_T^{\text{miss}}$  is required to be larger than 150 GeV to remove events where the lepton and the  $p_T^{\text{miss}}$  both come from a single leptonic W decay. Boosted Higgs boson decays collimated into a single wide AK8 jet are identified using the DEEPAK8 algorithm. To enhance the sensitivity, the search regions are

categorized based on the presence of a bb-tagged AK8 jet and the number of AK4 jets. The signal bins are further binned in  $p_T^{\text{miss}}$ .

Additional interpretations involving this analysis are shown in Appendix A.

### 7.1.2 Search for an oppositely charged same-flavor lepton pair

This search [51] (“2 $\ell$ ”) selects events with two oppositely-charged (or opposite-sign, OS), same-flavor (SF) leptons ( $e^+e^-$  or  $\mu^+\mu^-$ ) and large  $p_T^{\text{miss}}$ . Events are required to have exactly two reconstructed OSSF leptons, which are required to be well identified and isolated, with  $p_T > 25$  (20) GeV for the leading (trailing) lepton. The invariant mass of the dilepton system is required to be consistent with the Z boson,  $86 < m_{\ell\ell} < 96$  GeV. Events with additional leptons or isolated tracks are vetoed.

A total of three signal regions are considered, targeting the WZ, HZ, and ZZ topologies. Events in the resolved WW/ZZ region are required to have at least two AK4 jets ( $p_T > 35$  GeV) and no b-tagged jets. The boosted topology is constructed by selecting events with fewer than 2 AK4 jets and at least one PV candidate. The HZ signal region requires two b-tagged jets consistent with a  $H \rightarrow b\bar{b}$  decay. In boosted regimes, AK8 jets are required to have  $p_T > 200$  GeV and to have their soft-drop mass and the ratio of the 2- to the 1-subjettiness be compatible with the 2-prong structure expected in W and Z decays. The three signal regions are then binned in  $p_T^{\text{miss}}$ .

### 7.1.3 Search for two or three soft leptons

This search [52] (“2 $\ell$  soft”) selects events with two or three low- $p_T$  leptons (requiring  $e^+e^-$  or  $\mu^+\mu^-$  and one additional lepton in the relevant selections), jets, and large  $p_T^{\text{miss}}$ . It targets the WZ topology where the mass difference between  $\tilde{\chi}_1^0$  and  $\tilde{\chi}_2^0$  is small, so the W and Z bosons are off-shell, and the observable decay products have low momentum.

The leptons are required to satisfy  $p_T < 30$  GeV and either  $p_T > 3.5$  or  $> 5$  GeV, depending on the signal region. The invariant mass is constrained to the range  $1 < m_{\ell\ell} < 50$  GeV or  $4 < m_{\ell\ell} < 50$  GeV, which strongly suppresses the SM backgrounds while retaining high enough acceptance for



compressed signal scenarios. Additional kinematic requirements are applied to further reduce the SM backgrounds, and the signal regions are binned in  $m_{\ell\ell}$  and  $p_T^{\text{miss}}$ .

The interpretations involving this analysis are shown in Appendix A.

#### 7.1.4 Search for three or more leptons

This search [50] (“ $> 3\ell$ ”) focuses on final states with either two light leptons (electrons and muons) with the same charge or three or more leptons, including hadronically decaying taus ( $\tau_h$ ), and  $p_T^{\text{miss}}$ . It targets topologies involving WZ, WH, ZZ, HZ, and HH production, although it is most sensitive to the WZ and ZZ models. Its sensitivity to models involving a Higgs boson is reduced by its relatively small leptonic branching fractions. The search is organized in a series of thirteen exclusive categories that target the different decay modes of the W, Z, and H bosons, depending on the sign and flavor of the lepton pair, and the presence of a  $\tau_h$ .

All  $\tau_h$  candidates are required to have at least  $p_T > 30$  GeV. The  $p_T$  requirements for the other lepton flavors varies between 10 and 25 GeV, depending on category, and were designed to ensure full trigger efficiency after the selection. Additional requirements on  $p_T^{\text{miss}}$ , varying from  $p_T^{\text{miss}} > 50$  GeV (categories with three or more leptons) to  $> 60$  GeV (category with a same charge lepton pair) are applied to reduce the significant background from nonprompt leptons. The presence of b jets is vetoed in all categories, reducing top-enriched backgrounds.

#### 7.1.5 Search for multiple jets

This search [48] (“hadr. WX”) focuses on final states with a pair of hadronically decaying bosons and large  $p_T^{\text{miss}}$ , targeting the WW, WZ, and WH signal topologies. Taking advantage of the larger hadronic branching fraction of H and Z bosons, and also the branching ratios to a pair of b quarks, the search is classified into two regions, a b-veto region (0 b-tagged jets) and a b-tag region (at least 1 b-tagged jet).

This search requires at least two AK8 jets,  $p_T^{\text{miss}} > 200$  GeV,  $HT > 300$  GeV, and two to six AK4 jets. The AK8 jets are tagged using different versions of the DEEPAK8 tagger, depending on

the targeted boson. The b-veto search region targets the WW and WZ signal models and requires the AK8 jet mass to be between 65–105 GeV. The b-tag search regions targets the WZ and WH ( $Z/H \rightarrow b\bar{b}$ ) signal models. The AK8 jets from the W boson decays are required to have a mass between 65–105 GeV, while the AK8 jets from the Z/H boson decays are required to have a mass between 75–140 GeV.

Additional interpretations involving this analysis are shown in Appendix A.

## 7.2 Correlated systematics

While the searches were all developed individually, each search was aware of a prospective combination in the early stages of development. Because of this, we strove for mutually exclusive signal regions wherever possible, typically through lepton count. The hadronic searches are naturally separated from the searches with leptonic final states, as long as the same lepton definition ( $p_T$ , identification, and isolation requirements) is used. The  $2\ell$  searches are separated from the  $> 3\ell$  by removing events with three or more leptons from the  $2\ell$  search regions. This avoids major correlations between the analyses that would be difficult to statistically handle.

Table 7.2 summarizes the sources of uncertainty among the analyses and whether or not they are correlated. The dominant uncertainties in the background predictions are not correlated among analyses, as they tend to be either statistical in nature or uncertainties in the methods themselves. There are many similar signal uncertainties that should be treated as correlated when combining the searches. For the relevant analyses, both the lepton identification and isolation efficiency are correlated. Additionally, the uncertainty in the modeling of the trigger efficiency is correlated between individual analyses that share primary HLT trigger paths, rather  $p_T^{\text{miss}}$ -dependent or single- or double-lepton triggers. See Section 6.5.3 for a description of the remaining sources.

## 7.3 Interpretations

The results of the searches described in Section 7.1 are interpreted using the simplified models described at the beginning of this chapter (see Figures 7.1 and 7.2). Cross section limits as a

| Source                              | Correlated? |
|-------------------------------------|-------------|
| General                             |             |
| MC sample size                      | No          |
| SM background normalization         | No*         |
| Luminosity                          | Yes         |
| Trigger efficiency                  | Partially   |
| Pileup                              | Yes         |
| Prefiring                           | Yes         |
| Objects and signal modeling         |             |
| Lepton efficiency                   | Yes         |
| Jet energy resolution               | Yes         |
| Jet energy scale                    | Yes         |
| b (mis)tagging efficiency           | Yes         |
| AK8 bb tagging efficiency           | Yes         |
| AK8 jet mass resolution             | Yes         |
| ISR modeling                        | Yes         |
| Theory scales $\mu_R$ and $\mu_F$   | Partially   |
| Attributable to the fast simulation |             |
| $p_T^{\text{miss}}$ modeling        | Yes         |
| b (mis)tagging                      | Yes         |
| AK8 bb tagging                      | Yes         |
| AK8 bb mass                         | Yes         |

Table 7.2: Sources of systematic uncertainties and the level of correlation between analyses. The (\*) indicates an exception - the WZ background normalization is correlated between the “ $> 3\ell$ ” [50] and the “ $2\ell$  soft” [52] searches.

function of the SUSY particle masses are set using a modified frequentist approach, employing the  $\text{CL}_s$  criterion and an asymptotic formulation [125–128], the same as described in Section 6.6.3. The observed data yields in control regions are typically incorporated either by a simultaneous maximum likelihood fit of the signal and control regions or through parameterization using gamma functions. Other nuisances, such as the signal systematic uncertainties described in Section 7.2, are implemented using lognormal functions or as alternate shapes of the relevant distributions.

The analysis with the best exclusion limit at each point in the plane for the scan of different branching fractions to the Higgs boson is shown in Figure 7.3, motivating the combination of these individual analyses. Each individual search is sensitive to a different area and combined have a better reach than any search alone. The expected and observed limits for the neutralino-neutralino production for the ZZ topology, the HH topology, and the 50% mixed topology H/Z are shown in Figure 7.4.  $\tilde{\chi}_1^0$  masses less than 840 GeV are excluded for ZZ,  $< 1000$  GeV for the HH topology, and  $< 760$  GeV for the 50% mixed topology. The ZZ and mixed topologies show better-than-expected exclusion, whereas the HH topology shows the one-bin excess from the “4b” analysis as a 1-2 sigma effect in the region where  $m(\tilde{\chi}_1^0) < 700$  GeV. Figure 7.5 shows the results for the scan of different branching fractions to the H boson. The expected mass exclusion limit varies between  $\sim 550 - 950$  GeV and is least stringent around  $B(\tilde{\chi}_1^0 \rightarrow H\tilde{G}) = 0.4$ . The observed limit ranges between  $\sim 750 - 975$  GeV, allowing us to exclude masses below 750 GeV independent of the branching fraction.

The results for the higgsino-bino interpretation, assuming mass degenerate higgsino-like  $\tilde{\chi}_2^0$ ,  $\tilde{\chi}_3^0$ , and  $\tilde{\chi}_1^\pm$  that decay to a bino-like  $\tilde{\chi}_1^0$  and a SM boson, are shown in Figure 7.6, excluding  $m(\tilde{\chi}_1^\pm) = m(\tilde{\chi}_2^0)$  between 225 and 800 GeV for a  $\tilde{\chi}_1^0$  with a mass of 50 GeV.

In general for the models considered in this combination, chargino masses are excluded up to 990 GeV, while neutralino masses are excluded up to 1000 GeV. The improvement is between 100–510 GeV with respect to the limits when using only 2016 data [133].

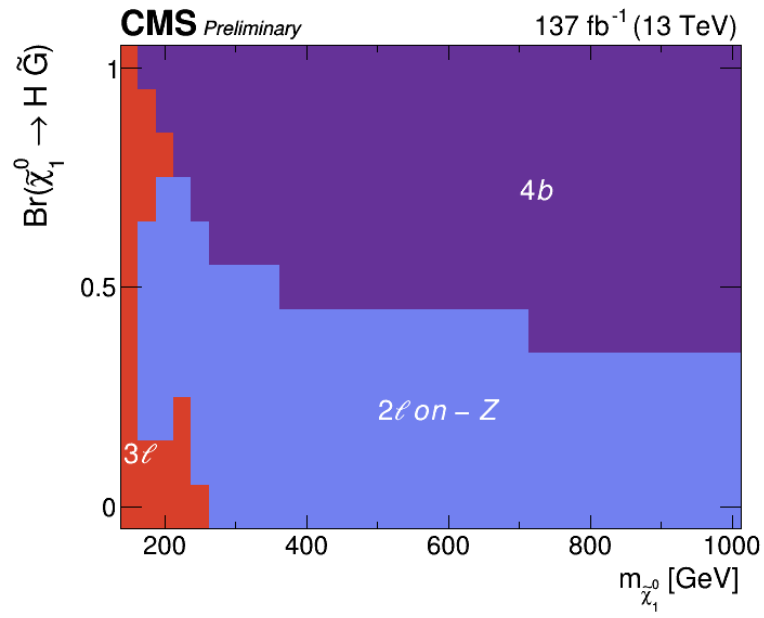


Figure 7.3: The analysis with the best exclusion limit at each point in the plane for the neutralino-neutralino production for different branching fractions to H boson.

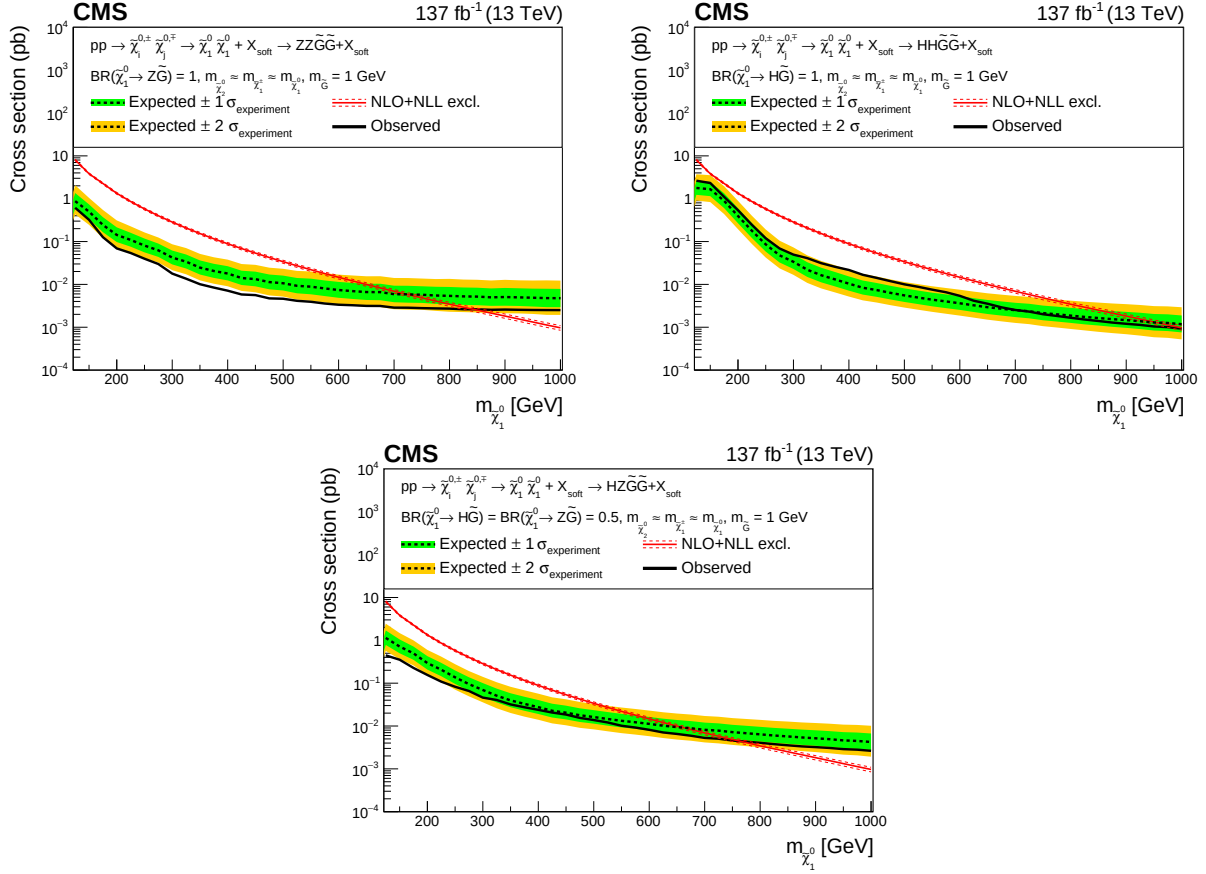


Figure 7.4: Expected and observed exclusion limits for the neutralino-neutralino production for the ZZ topology (upper left), the HH topology (upper right), and the mixed topology with 50% branching fraction to H and Z (bottom). The one-bin excess from the “4b” analysis can be seen in the HH topology as a 1-2 sigma effect from  $m(\tilde{\chi}_1^0) < 700$  GeV.

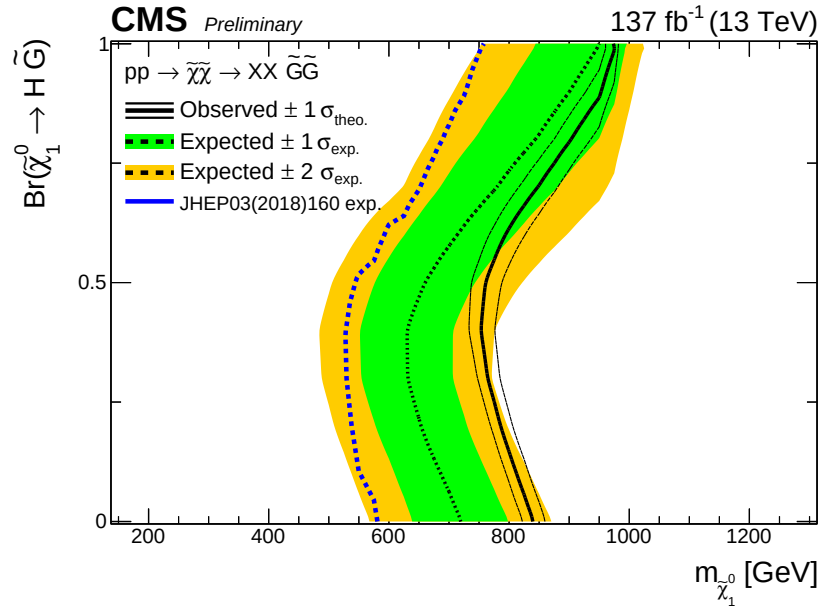


Figure 7.5: Results for the neutralino-neutralino production for different branching fractions to an H boson.

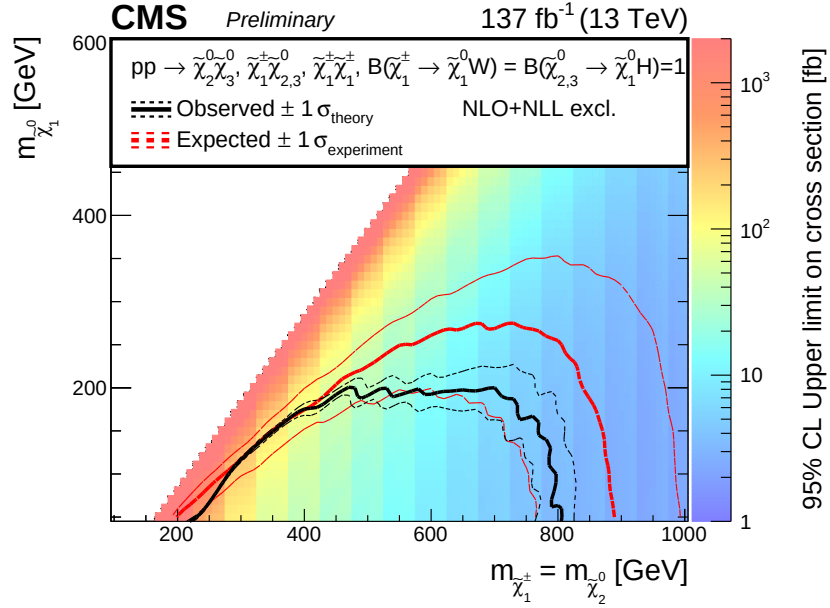


Figure 7.6: The expected and observed exclusion limits for the higgsino-bino interpretation.

## Chapter 8

### High-Luminosity LHC

The data used for the analysis of this thesis were collected from 2016-2018, in what is known as “Run 2” of the LHC. Since the end of Run 2, CMS has worked to improve the detector for the next data-taking period, “Run 3.” Run 3 was scheduled to begin in 2021 but was unfortunately delayed due to the global coronavirus pandemic. It is now expected to begin early in 2022 with 13.5 TeV center-of-mass energy and an integrated luminosity of up to  $500 \text{ fb}^{-1}$ . After that, the center-of-mass energy is constrained to the fixed ring size of the LHC, and so the LHC will proceed toward increasing the intensity. The High-Luminosity LHC (HL-LHC) is planned to begin collisions in 2027, starting “Phase 2” of the LHC running (“Phase 1” is Runs 1, 2, and 3). The HL-LHC will reach a luminosity of up to  $\mathcal{L} = 1.5 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ , compared to Phase 1’s peak luminosity of  $\mathcal{L} = 2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , and an increased average number of proton-proton collisions per bunch crossing (pileup) of around 200. The integrated luminosity is expected to be  $4000 \text{ fb}^{-1}$ , almost  $30\times$  that of Run 2 ( $137 \text{ fb}^{-1}$ ). The upgrades to the accelerator chain that are necessary to achieve such a high luminosity are detailed in Section 8.1. Section 8.2 describes the upgrades CMS is undergoing to accommodate this substantial increase in luminosity, requiring most of the subdetectors to be improved or replaced with more radiation-hard material and electronics. The LHC baseline plan, which includes the energy of the collisions and luminosity, can be seen in Figure 8.1.

The increased pileup poses a significant challenge for the L1 trigger, but the upgrade will enhance the physics selectivity while maintaining the necessary performance throughout the 10 year-long HL-LHC program. The results shown in the following sections are documented in the



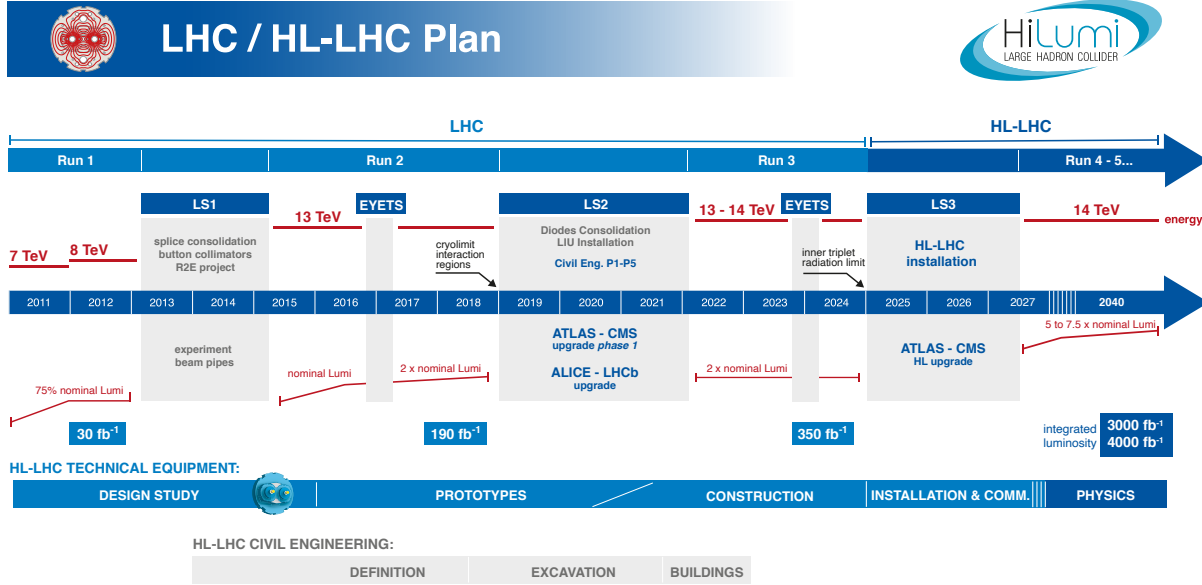


Figure 8.1: The LHC baseline plan, showing the energy of the collisions (upper red line) and luminosity (lower red lines). The second long shutdown (LS2), 2019–2022, is on-going at the time of writing, and involves upgrades to the LHC injectors, among other things. After LS3, 2025–2027, the machine will be in the High Luminosity configuration (HL-LHC), and Phase 2 running will begin.

Level-1 Trigger Technical Design Report [134] (TDR). The TDR was compiled to show the proposed scope for CMS for the Phase 2 L1 trigger upgrade project. Based on this TDR and subsequent review, the plan was accepted and is now part of the baseline upgrade project for CMS. Section 8.3 documents the upgrades to the L1 trigger, the largest of which is the inclusion of information from the tracker at L1. Sections 8.4 and 8.5 describe the newest editions to the L1 trigger, the Track Finder (TF) and the Global Track Trigger (GTT), respectively. The remaining sections are descriptions and simulated performance of the track-based algorithms that can now be included for Phase 2 running: track  $p_T^{\text{miss}}$  (Section 8.6), primary vertex identification (Section 8.7), and track jets (Section 8.8). Lastly, we describe the Data Quality Monitoring (DQM) system for the Phase 2 upgrade in Section 8.10, with particular focus on the trigger primitives from the Track Finder.

## 8.1 LHC injector upgrade project

The LHC Injector Upgrade (LIU) project [135–140] aims to increase the intensity and the brightness of the LHC injectors to match the HL-LHC requirements ( $\mathcal{L} = 1.5 \times 10^{35} \text{cm}^{-2}\text{s}^{-1}$ ) while ensuring the injector complex can function reliably through the entire HL-LHC running era. The project comprises a new  $\text{H}^-$  Linac (Linac4) as well as extensive upgrades to the PS Booster, PS and SPS synchrotrons. For an overview of the current injector complex, including descriptions of the different accelerators and synchrotrons, see Chapter 3.

A new injector called Linac4 began construction in 2008 in its own tunnel, parallel to the current 40 year-old linear accelerator (Linac2), so that it could be built without disrupting the LHC runs. Using radio frequency (RF) cavities at 352 and 704 MHz, Linac4 has replaced the 50 MeV proton Linac2 and delivers a 160 MeV, 40 mA  $\text{H}^-$  beam. The ions are then injected into the Proton Synchrotron Booster (PSB), where both electrons are stripped. The use of ions instead of protons reduces the beam loss at injection and allows more particles to accumulate in the synchrotron.

The PSB is the first synchrotron in the accelerator chain and contains four rings with a radius of 25 meters. The PSB was upgraded during the Long Shutdown 2 (LS2) to receive the injection of 160 MeV  $\text{H}^-$  ions from Linac4 and to increase the energy of the protons to 2 GeV (compared to the previous 1.4 GeV) before being injected into the Proton Synchrotron (PS). These were accomplished by upgrades to the injection and extraction magnets (the kicker and septa magnets); dipole, quadrupole, and corrector magnets; the transfer line between Linac4 and the PSB; and the power supply systems. The RF acceleration system was also completely replaced, based on cavities built using a composite magnetic material called FineMet, allowing for a larger bandwidth of frequencies, from 0.6 to 18 MHz.

The PS has 277 room-temperature electromagnets, including 100 dipoles that bend the beams around the ring. The accelerator operates at up to 26 GeV. New injection vacuum chambers have been installed that remove the aperture limitations from the PSB. The upgrades for the HL-LHC were completed during the LS2 and include a new beam-loss monitoring system, improved amplifier

units for the RF cavities, reduction of the impedance<sup>1</sup> of the 10 MHz RF system, and a longitudinal feedback system to mitigate longitudinal instabilities.

The Super Proton Synchrotron (SPS) operates at up to 450 GeV, and has 1317 room temperature electromagnets, including 744 dipoles. All of the RF cavities were dismantled and remounted during the LS2, and a new 200 MHz power plant was added to increase the RF power. New power amplifiers, a state-of-the-art beam-control system, and a new beam-dump system were also installed to handle the increased beam intensity and brightness. The upgrades greatly mitigate the electron cloud, one of the main limitations of the SPS up through Run 2.

Additionally, many systems in the injector chain will be upgraded to cope with the increased bunch intensity and beam brightness, including:

- New wide-aperture focusing quadrupole magnets made from superconducting Nb<sub>3</sub>Sn that generate a 12 T magnetic field (compared to the current 8 T) will be installed close to the interaction points of CMS and ATLAS.
- New collimators, made from a material that produces less electromagnetic interference on the beam, will be installed, providing enhanced beam cleaning. Roughly 60 of the 118 existing collimators will be replaced, and an additional 20 will be added.
- Superconducting crab cavities will be installed at either side of CMS and ATLAS, which are designed to give particle bunches a transverse momentum before collisions occur. This increases the overlap area of the colliding bunches and thus the probability of collisions.
- A full cryogenics upgrade will be installed to keep the existing and new magnets at the required 1.8 K.
- Innovative superconducting power lines will be installed, connecting the power converters to the accelerator. These cables are made of superconducting magnesium diboride and can carry currents up to 100,000 amps.

---

<sup>1</sup> The upgrade reduces the cavity impedance by roughly a factor of two.

## 8.2 CMS Phase 2 upgrade

The primary goal of the Phase 2 upgrade program is to maintain the excellent performance of Phase 1 with the increased challenging conditions of the HL-LHC, namely the increased pileup (PU). The beam brightness and the new focusing and crossing scheme at the interaction points will allow an instantaneous luminosity of up to  $\mathcal{L} = 1.5 \times 10^{35} \text{cm}^{-2}\text{s}^{-1}$ . Studies intended to understand the effect of radiation damage highlighted the need to replace both the tracker and the endcap calorimeters before Phase 2 running. Particle-flow reconstruction greatly mitigates PU, and so the tracker granularity will be increased to enable more precise determination of the primary vertex. The new endcap calorimeters will provide improved energy resolution, mostly for jets. The upgrades to these systems are detailed in Sections 8.2.1 and 8.2.2, respectively, while updates to the muon detector, including the addition of a new muon subsystem, are detailed in Section 8.2.3.

### 8.2.1 Tracker

By the end of Run 3, the tracker will have suffered significant radiation damage and will need to be entirely replaced. This enables us to improve the tracker design to allow use of the information from the tracker in the L1 trigger, one of the key features of the Phase 2 upgrade. This is made possible due to a unique design of the Phase 2 outer tracker (OT), with the addition of “ $p_T$  modules”. These modules perform hit correlations between two closely-spaced silicon sensors that filter out particles with  $p_T > 2 \text{ GeV}$ , allowing for a substantial rate reduction. The pairs of correlated hits, called “stubs,” are sent to the backend L1 track-finding system. The geometry for the Phase 2 tracker can be seen in Figure 8.2, with the red and blue lines corresponding to the OT (strip detector) and the green and orange lines to the inner tracker (pixel detector). The granularity of both the OT and the pixel systems will be increased by  $\sim 4\times$ , improving  $p_T$ -resolution in the OT, and overall b-tagging and track reconstruction efficiency within boosted jets. Up to 10 additional pixel discs will be added in each endcap, extending the tracker coverage close to  $|\eta| = 4$ , like the calorimeter range. With this extension, total and missing energy measurements will be greatly

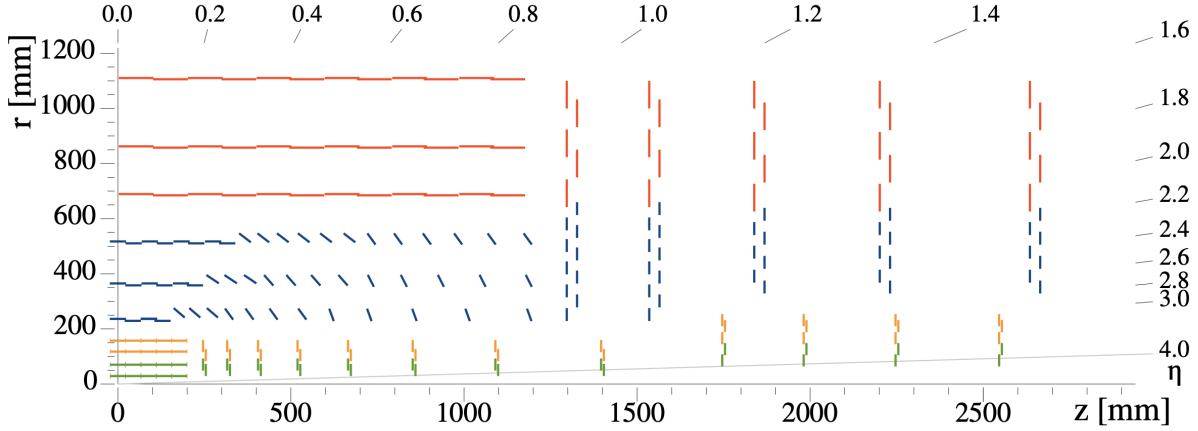


Figure 8.2: One quadrant of the Phase 2 tracker. The red and blue lines denote the Outer Tracker (used for triggering at L1), while the orange and green lines denote the Inner Tracker (pixel detector).

improved, necessary for future BSM analyses. At L1, only the OT is used for track finding, and so the resulting L1 tracks can only be found up to  $|\eta| < 2.4$ . The OT is composed of six barrel layers and five endcap discs. The tilted geometry of the barrel modules of the OT allows for better coverage in the transition region of roughly  $1.0 < |\eta| < 1.8$  and has the added benefit of a smaller material budget.

The removal of stubs incompatible with tracks with  $p_T > 2 \text{ GeV}$  reduces the stub rate by over 90%, greatly reducing the combinatorics for the L1 track reconstruction. The average  $t\bar{t}$  collision event with 200 pileup vertices has 180 tracks with  $p_T > 2 \text{ GeV}$ , each with 4–6 stubs, making reconstruction challenging, but not impossible. This reduction in stub rate is the underlying change that will allow tracks to be reconstructed and used at L1 for Phase 2.

### 8.2.2 Calorimeter endcaps

Just like the tracker, the ECAL and HCAL endcaps will have suffered significant radiation damage and will be replaced by the High Granularity Calorimeter (HGC). The HGC will provide detailed three-dimensional images of showers due to excellent transverse and longitudinal segmentation. The ECAL endcap will consist of roughly 30 tungsten and copper plates interwoven with silicon sensors for the active material. The hadronic system has a front section of twelve brass and

copper plates interwoven with silicon sensors, for a depth of  $3.5\lambda$ , covering the maximum hadronic shower measurement. A “backing hadron calorimeter” similar in design to the current HE is placed after the front section. Brass plates interleaved with scintillating tiles, whose light is read out by a wavelength-shifting fibers, make up the backing hadron calorimeter, and provide an overall depth of  $\approx 10\lambda$  for the full HCAL.

### 8.2.3 Muon endcaps

Currently, the muon system in the region  $1.5 \leq |\eta| \leq 2.4$  consists of four stations of CSCs and is the only region of the muon detector that lacks coverage from multiple systems, despite it being a challenging region for muon momentum resolution. In order to maintain sufficient L1 muon trigger acceptance in this region, the muon system will be enhanced with additional chambers built for higher rate capability. The first two stations are situated where the magnetic field is still relatively high, and so they will be fitted with Gas Electron Multiplier (GEM) chambers for optimal position resolution. This will improve the momentum resolution for standalone muon triggers, which will additionally improve the matching to muon tracks in the Global Muon Trigger. The last two stations will make use of lower-granularity RPCs that will provide good timing resolution to mitigate background effects. The addition of these four stations will increase the coverage of the muon systems to  $|\eta| \approx 3$ .

## 8.3 Phase 2 Level-1 Trigger

The Phase 2 L1 trigger upgrades consist of improvements to both hardware and algorithms. An overview of the current Phase 1 L1 trigger, including a general description of the two-tier trigger system employed by CMS, can be found in Section 4.7. For the first time ever, information from the tracker will be made available at L1. The upgraded hardware components include new FPGAs (Xilinx UltraScale+ class) and improved high-speed optical links, up to 28 Gb/s. The total latency for the trigger has been increased from  $4 \mu\text{s}$  to  $12.5 \mu\text{s}$ , which enables the use of tracker and high-granularity calorimeter information, and also the potential to use particle-flow techniques [80] or

a greater use of machine-learning based approaches. The L1 trigger has an input rate of 40 MHz and reduces it to 750 kHz. The HLT further reduces the output rate to 7.5 kHz.

The L1 trigger consists of four independent data-processing paths and a final global trigger that receives all output objects and issues the final L1 trigger decision. The schematic for the Phase 2 L1 trigger is shown in Figure 8.3. The calorimeter trigger builds electron candidates, hadronically decaying taus, jets, and associated energy sums. The muon trigger The track trigger reconstructs tracks (not available in Phase 1), primary vertices, and tracker-only based objects, such as jets and energy sums. Lastly, the correlator trigger produces a global event reconstruction. It creates particle-flow candidates and hosts a version of the Pileup Per Particle Identification (PUPPI) algorithm [141], which removes pileup.

As previously mentioned, information from the tracker was not available at L1 in Phase 1 running, and so we will focus on the Track Trigger. The Track Trigger consists of the Track Finder (TF) described in Section 8.4 and the Global Track Trigger (GTT) detailed in Section 8.5. The TF is an active area of work and focuses on reconstructing tracks from the hits in the tracker, within a small amount of time and with limited computational resources. The GTT receives the tracks from the TF, builds track-based objects, and sends the objects downstream to the Global Trigger. The GTT computes the primary vertex position, enabling PUPPI pileup mitigation, as well as standalone track-only objects such as missing energy and jets. Work is also in progress to extend the GTT to reconstructed displaced objects to trigger on new long-lived physics.

## 8.4 Track Finder

The Track Finder is responsible for reconstructing tracks from the hits in the tracker within  $5 \mu\text{s}$  per event. The track-finding algorithm, further described in Section 8.4.1, is implemented in firmware, since that is the only way to meet the latency requirements. The tracks themselves are then sent to down-stream users, other systems of the L1 trigger that build track-based objects for further trigger paths.

Since the tracks are reconstructed with a certain level of efficiency, we impose certain track

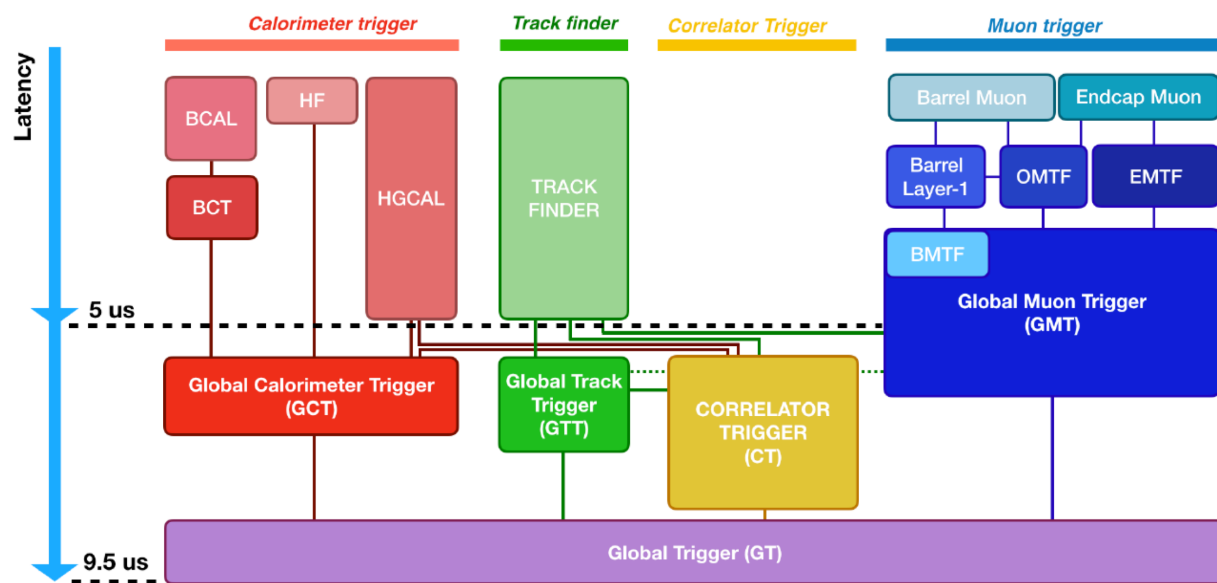


Figure 8.3: A schematic of the L1 trigger for Phase 2, showing the four independent data processing paths.



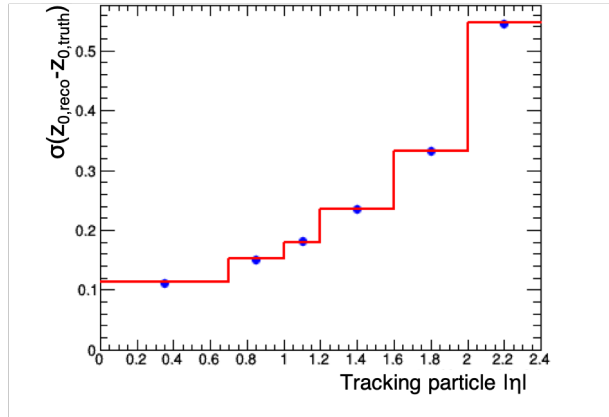


Figure 8.4: The  $z_0$  resolution for L1 trigger tracks for a  $t\bar{t}$  with 200 PU sample. The resolution is between the reconstructed track  $z_0$  and the  $z_0$  from the truth-level track matched to the reconstructed track, and is binned in the truth-level track (called a tracking particle)  $\eta$ . The  $z_0$  resolution worsens in the forward regions of the tracker.

selection criteria designed to rather limit the fraction of non-genuine tracks or to reduce the L1 rate for track-based objects. The excellent  $z_0$  track resolution, shown in Figure 8.4, allows for PU tracks to be easily identified from the hard scatter. Removing the PU tracks greatly reduces the L1 rate for the track-based objects described in more detail in Section 8.5.

#### 8.4.1 Tracking algorithm

The tracking algorithm developed for the L1 trigger is rather similar in structure to what is used in the offline reconstruction during Phase 1 (Section 5.1.1); it first forms seeds (called “tracklets”) from stubs in two adjacent layers combined with a beamspot constraint, projects those seeds onto nearby layers to look for additional compatible stubs, and then fits the final track candidates with a Kalman filter to identify the best stub candidates and calculate the track parameters. A schematic for a stub, including a brief description of stub  $p_T$ , is shown in Figure 8.5.

The tracking algorithm used is called the “hybrid” algorithm, which combines parts from two previous tracking algorithms: the seeds from the Tracklet approach and the Kalman filter fitting from the Time Multiplex Track Trigger (TMTT) approach. Only tracklets consistent with a particle with  $p_T > 2$  GeV,  $|\eta| < 2.4$ , and  $|z_0| < 15$  cm are kept. These requirements are from the constraints of the tracker geometry and the computational resources available at L1. The

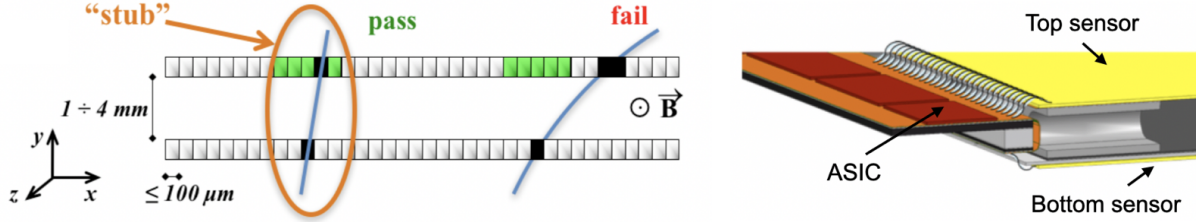


Figure 8.5: A sketch of a stub (left) and a  $p_T$  module (right). The stub is defined as the two hits in each layer of the module. From this, we obtain the stub bend (the lateral difference between the two hits) and a crude measurement of the  $p_T$  based on how much the trajectory bends. A low-momentum track ( $< 2$  GeV) falls outside of the acceptance window, and no stub is saved.

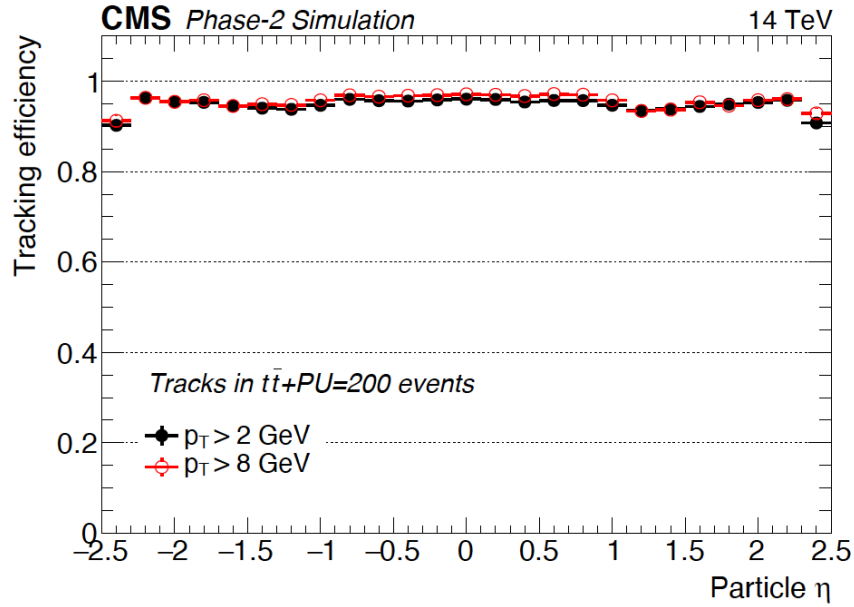


Figure 8.6: L1 track-finding efficiency as a function of  $\eta$  for tracks above 2 GeV (filled black markers) or 8 GeV (open red markers) in  $t\bar{t}$  events with 200 pileup vertices.

track-finding efficiency with these constraints is shown for a  $t\bar{t}$  sample with 200 pileup vertices in Figure 8.6, averaging 95% efficiency.

Using multiple parallel seeding combinations of two adjacent layers (L1+L2, L2+L3, L3+L4, L5+L6, L1+D1, L2+D1, D1+D2, D3+D4, where “L” is the barrel layer numbered radially outward and “D” is for the endcap discs) allows for high track-finding efficiency without overwhelming combinatorics. The tracklets are then projected both inward and outward onto nearby layers and discs, in search of stubs that are compatible with the projection in a narrow stub-matching window.

We require at least four stubs to form an L1 track; the list of stub candidates per track are sent to the Kalman filter step. Any track candidates with shared stubs are merged before fitting. The Kalman filter identifies the best stub candidates through a mixture of the  $\chi^2$  fit and a quick consistency check of the stub bend (the horizontal distance between the two hits in the module that make up a stub) with the expected  $p_T$  of the track that left it. Any stub candidates that do not meet these criteria are removed from consideration, and the stub that gives the resulting track the best  $\chi^2$  value (the “best-fit”) is chosen. Finally, the Kalman filter fits the final track to determine its parameters:  $p_T$ ,  $\eta$ ,  $\phi$ ,  $z_0$ , and  $d_0$ . The impact parameter  $d_0$  is an optional parameter of the fit, depending if the tracklet and Kalman filter assume the track originated from the beamline. Allowing a 5-parameter fit means we can also create displaced tracks, originating from secondary vertices. This extended L1 tracking includes “triplet” seeds, a combination of three layers/disc with no beamspot constraint, in addition to the regular seed combinations. The improvement in tracking performance for displaced signatures can be seen in Figure 8.7. Currently, the extended tracking is not in the baseline plan for the L1 trigger, but, due to its physics potential, we are working on improving and including it.

## 8.5 Global Track Trigger

The Global Track Trigger (GTT) receives the tracks from the TF and reconstructs track-based objects. These objects are sent to the Correlator Trigger, where particle flow (PF) is run, and to the Global Trigger, which makes the final event decision for L1. While many objects can be made from the L1 tracks, we will focus on the main three: track-based  $p_T^{\text{miss}}$  (discussed in Section 8.6), primary and pileup vertices (discussed in Section 8.7), and track-based jets (discussed in Section 8.8). Of these objects, the vertices are passed to the Correlator Trigger. These algorithms are implemented in firmware, since that is the only way to meet the strict latency requirement. Details on each algorithm, including track quality criteria and the effect on the trigger efficiency and rate, are given in the following sections.

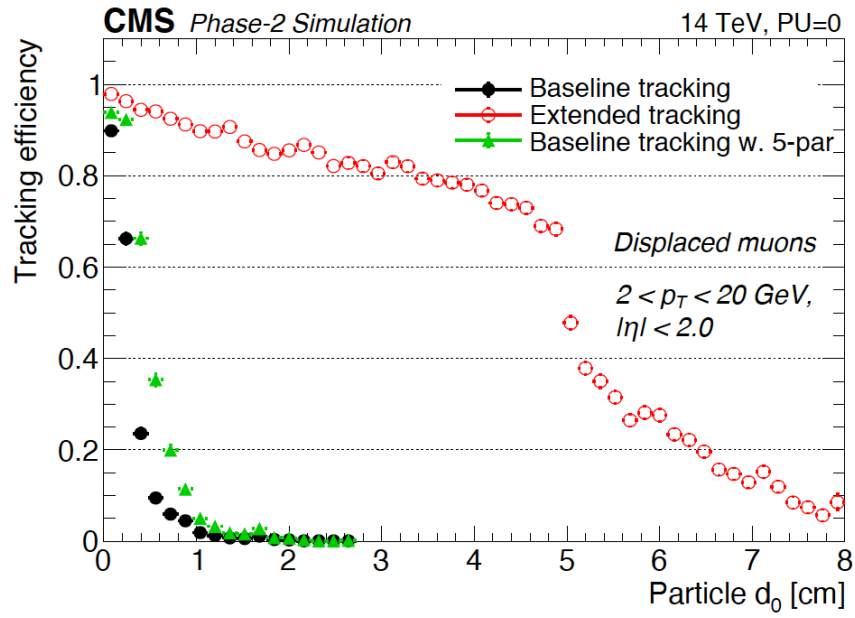


Figure 8.7: Preliminary tracking efficiency as a function of the impact parameter  $d_0$  for simulated displaced muons. The efficiency is defined with respect to truth-level particles, shown for the baseline tracking with an implied beamspot constraint ( $d_0 = 0$ ) in black, the baseline tracking that allows for a non-zero  $d_0$  in green, and the full extended tracking algorithm that uses the triplet seeds and a non-zero  $d_0$  in red.

## 8.6 Track-based missing energy

The vector sum of the transverse momenta ( $p_T^{\text{miss}}$ ) of all particles produced in the primary interaction is a key input for triggering on BSM signatures at L1 and is similar in structure to the  $p_T^{\text{miss}}$  described in Section 5.3.3. For the track-based algorithm described in this section, one of the main challenges is excluding fake tracks that have high transverse momentum which lead to large non-genuine  $p_T^{\text{miss}}$ . These fake tracks come from spurious combinations of stubs. Events containing these high- $p_T$  fake tracks would dominate a L1 track  $p_T^{\text{miss}}$  trigger. We have performed a number of optimization studies in simulation to reduce the number (or at least the effect) of fake tracks. The track quality criteria optimized from these studies are given in Section 8.6.1.

Like the other track-based objects, the track  $p_T^{\text{miss}}$  can easily remove tracks from PU interactions, which could overwhelm the  $p_T^{\text{miss}}$  from the PV. We impose  $\Delta z(z_{\text{PV}}, z_{\text{trk}})$  requirements to remove these PU tracks. The primary vertex described in Section 8.7 is input to the L1 track-based  $p_T^{\text{miss}}$  algorithm. A window of  $\Delta z(z_{\text{PV}}, z_{\text{trk}})$  is required for track  $\eta$  regions within the tracking acceptance, shown in Table 8.1. These specific values are based on the  $z_0$  resolution from the track trigger, which varies with track  $\eta$  (see Figure 8.4). Central tracks allow for a tight window of 0.36 cm while more forward tracks need a wider  $z$ -window (as large as 1.60 cm) to include tracks within  $3\sigma$  of the Gaussian  $z_{\text{trk}}$ -resolution. These selection criteria have been re-optimized since the TDR was published.

| Track $\eta$            | $ \Delta z $ [cm] |
|-------------------------|-------------------|
| $0.0 \leq  \eta  < 0.7$ | 0.37              |
| $0.7 \leq  \eta  < 1.0$ | 0.50              |
| $1.0 \leq  \eta  < 1.2$ | 0.60              |
| $1.2 \leq  \eta  < 1.6$ | 0.75              |
| $1.6 \leq  \eta  < 2.0$ | 1.00              |
| $2.0 \leq  \eta  < 2.4$ | 1.60              |

Table 8.1: Maximum  $\Delta z(z_{\text{PV}}, z_{\text{trk}})$  requirements between the primary vertex and  $z_{\text{trk}}$  in each  $\eta$  region for the track selection. These values are based on the  $z_0$  resolution from the track trigger, which varies with track  $\eta$ .

It is more important to reduce the amount of high- $p_T$  fake tracks than it is to maintain

high track efficiency, and so we impose track selection criteria in addition to the  $\Delta z$  requirements. The track selection shown in Table 8.2 is based mainly on the confines of the detector, the track reconstruction algorithm, and the available track fit quality parameters. The track quality requirements optimized specifically to remove fake tracks are the track  $\chi^2$  fit and a new variable which we call  $\chi_{\text{bend}}^2$ , which is the consistency between the measured stub bend from the hardware and the expected bend from the  $p_T$  of the track. Further details on  $\chi_{\text{bend}}^2$  are provided in Section 8.6.1.1. The addition of the  $\chi_{\text{bend}}^2$  requirement allows us to loosen the regular  $\chi^2$  criteria, improving our track efficiency while maintaining our reduction in fake tracks. Since the publishing of the L1 TDR, the  $\chi^2$  fit variable has been split into its two plane-components,  $r-z$  and  $r-\phi$ , and each has been normalized to the degrees of freedom of the track. As such, Table 8.2 shows the two components as separate quality variables.

| Track variable                            | Selection            | Purpose                          |
|-------------------------------------------|----------------------|----------------------------------|
| $p_T$                                     | $\geq 2 \text{ GeV}$ | Track reconstruction limitations |
| $ \eta $                                  | $\leq 2.4$           | Detector limitations             |
| $ z_0 $                                   | $\leq 15 \text{ cm}$ | Detector limitations             |
| $N_{\text{stubs}}$                        | $\geq 4$             | Track reconstruction limitations |
| $\Delta z(z_{\text{PV}}, z_{\text{trk}})$ | varied               | Track quality - PU reduction     |
| $\chi_{r-z}^2/dof$                        | $\leq 5.0$           | Track quality - fake reduction   |
| $\chi_{r-\phi}^2/dof$                     | $\leq 20.0$          | Track quality - fake reduction   |
| $\chi_{\text{bend}}^2$                    | $\leq 2.25$          | Track quality - fake reduction   |

Table 8.2: L1 track quality criteria for tracks input to both the L1 track-based jet clustering and the L1  $p_T^{\text{miss}}$  computation. Cuts are optimized to preserve a sharp trigger turn-on in track-based  $H_T$ ,  $H_T^{\text{miss}}$ , and  $p_T^{\text{miss}}$ , primarily from rejecting the bulk of fake tracks.

### 8.6.1 Track quality

The extended tracking described in Section 8.4.1 has looser  $\chi^2$  and stub bend criteria, which increases the number of non-genuine tracks. These tracks, termed “fake” tracks, are made from spurious combinations of stubs and also exist in the prompt tracking. We therefore impose quality criteria on the tracks, in hopes of removing these fake tracks while maintaining high track-finding

efficiency for the genuine tracks. Typical quality criteria include the number of stubs in a track, the track  $\chi^2$  fit, and a new variable called  $\chi_{\text{bend}}^2$  (further details in Section 8.6.1.1).

Additionally, the tracker has the unique ability (compared to other subdetectors) to identify tracks left from PU. In addition to the quality criteria on the input L1 tracks, we select tracks that are consistent with the reconstructed primary vertex of the event, typically placing a ceiling on  $\Delta z(z_{\text{PV}}, z_{\text{trk}})$ . This preserves a low trigger threshold while maintaining low data-taking rates. The quality criteria are the same for the track-based  $p_{\text{T}}^{\text{miss}}$  and the tracker jets, described in Sections 8.6 and 8.8, respectively.

#### 8.6.1.1 Stub bend consistency $\chi_{\text{bend}}^2$

To begin, we developed a new metric to distinguish genuine from fake tracks, called  $\chi_{\text{bend}}^2$ . The distribution of  $\chi_{\text{bend}}^2$  and the performance at retaining good tracks vs rejecting fake tracks can be seen in Figure 8.8.  $\chi_{\text{bend}}^2$  is a measure of how consistent the stub bend  $b_s$  and the track bend  $b_t$  are:

$$\chi_{\text{bend}}^2 = \frac{1}{N_{\text{stub}}} \sum \frac{(b_t - b_s)^2}{(\sigma_b)^2}, \quad (8.1)$$

where  $N_{\text{stub}}$  is the number of stubs on the track and the sum implicitly sums over each stub of the track, to make  $\chi_{\text{bend}}^2$  essentially a per-degree-of-freedom variable. The bend resolution  $\sigma_b$  is obtained from the difference in  $b_t$  and  $b_s$  for genuine truth-matched tracks. The matching done is a tight matching, meaning it requires every stub of the track to be truth-matched to stubs produced from generator-level, hard-interaction particles.  $b_s$  is measured directly from the hardware modules and is the lateral difference between the two hits in each stub. The track bend  $b_t$  is defined as:

$$b_t = -\frac{a r B c}{\Gamma \epsilon p_{\text{T}}}, \quad (8.2)$$

where  $a$  is the sensor spacing of the modules (1.6 or 4.0 mm),  $r$  is the radial position of the stub,  $B$  is the magnetic field inside the tracker (3.8112 T),  $c$  is the speed of light in a vacuum,  $\Gamma$  is the sensor pitch (90  $\mu\text{m}$ ),  $p_{\text{T}}$  is the transverse momentum of the track from the fit, and  $\epsilon$  is a correction

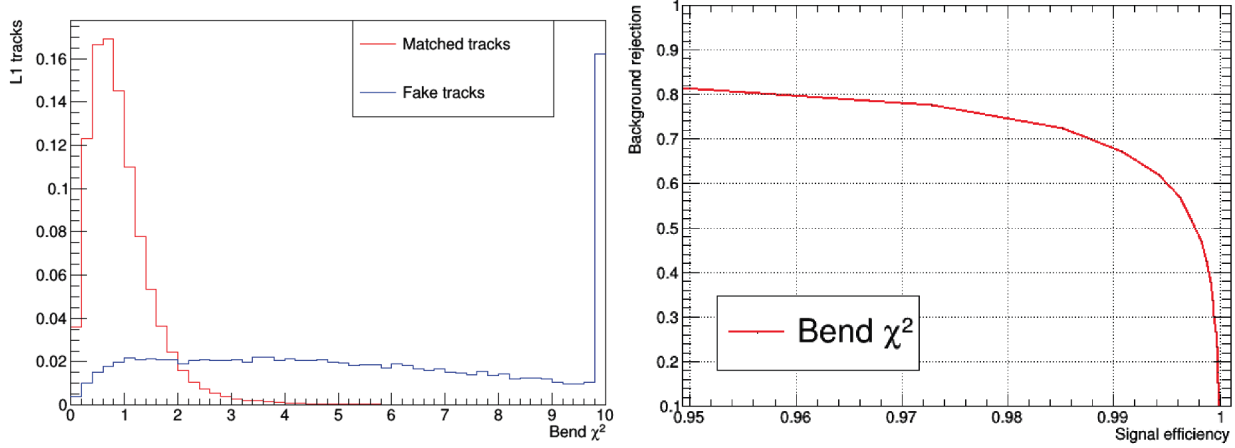


Figure 8.8: Distributions of  $\chi^2_{\text{bend}}$  for truth-matched and fake tracks (left) and a ROC curve (right). For this, a truth-matched track requires tight-stub matching, so every stub must be matched to a stub left by a truth-level, hard-interaction particle. Fake tracks cannot be matched to any truth-level particle and typically arise from spurious combinations of stubs. Each point along the ROC curve demonstrates the percentage of fake tracks removed on the  $y$ -axis (background rejection) and the percentage of genuine tracks retained (signal efficiency) on the  $x$ -axis for a given cut value of  $\chi^2_{\text{bend}}$ . These studies were performed with an older version of the tracking algorithm.

for the tilted modules in the barrel. The stub bend and a by-eye estimate of the track bend can be seen in Figure 8.5.

$\chi^2_{\text{bend}}$  is uncorrelated with the track fit  $\chi^2$ , since the track fit  $\chi^2$  only uses information from one of the two hits in a given stub. Applying both the track  $\chi^2$  and  $\chi^2_{\text{bend}}$  as quality criteria greatly reduces the fake rate while maintaining a high efficiency. Using an older tracking algorithm, we can reduce the track fake rate by a factor of three, with only a loss of 1.5% in efficiency.

### 8.6.2 Track-based missing energy performance

Applying the full track selection in Table 8.2, the trigger threshold is lowered to 70 GeV, the same threshold as when only simulated charged particle trajectories are used for the  $p_T^{\text{miss}}$  computation, “simulated track  $p_T^{\text{miss}}$ .” This allows a low trigger threshold for offline- $p_T^{\text{miss}}$  for BSM models like a BSM stop signal. Figure 8.9 shows the reduction in L1 rate and the trigger efficiency after the full track selection is applied, compared also to the simulated track  $p_T^{\text{miss}}$ . The L1 rate is the amount of data that would be kept for the corresponding threshold. For track  $p_T^{\text{miss}}$ , a reasonable rate of 25 kHz would allow us to keep events with  $p_T^{\text{miss}} > 200$  GeV with only the  $\Delta z$  selection and



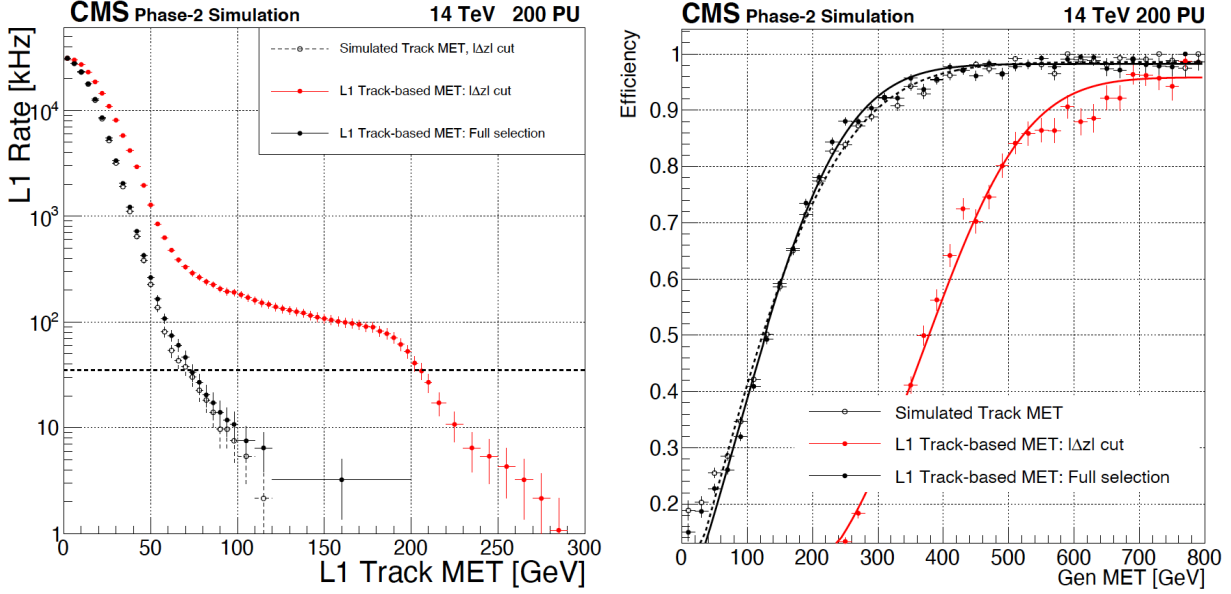


Figure 8.9: The track  $p_T^{\text{miss}}$  rate from single neutrino MC events with 200 PU interactions. The full track quality criteria reduce the threshold from 200 GeV (red markers) to 70 GeV (solid black markers), as seen in the L1 rate (left figure). This is roughly the threshold from simulated tracks (open black circles). The corresponding trigger efficiency (right) for SUSY stop MC with 200 PU, with  $m(\tilde{t})=1000$  GeV and  $m(\text{LSP})=775$  GeV, reaches 95% at 650 GeV and 350 GeV, for the  $\Delta z$ -only selection and the full selection, respectively.

$p_T^{\text{miss}} > 70$  GeV after the full selection. The corresponding trigger efficiency is a measure of how efficient those thresholds are, binned in generator-level  $p_T^{\text{miss}}$ . Compared to PF  $p_T^{\text{miss}}$ , track  $p_T^{\text{miss}}$  will always have a slower turn-on curve because the tracker lacks information from neutral particles. The compatibility of track-based  $p_T^{\text{miss}}$  and PF  $p_T^{\text{miss}}$ , however, could help provide a cleaner SUSY signature.

## 8.7 Vertex finding

The identification and reconstruction of a primary vertex (PV) from the tracks is crucial for mitigating pileup and highly beneficial for reducing data rates of trigger objects that target the hard scatter. The reconstructed vertex with the largest sum of track  $p_{T^2}$  is taken as the PV, similar to offline reconstruction. Since the publication of the TDR, the vertex-finding algorithms have been updated to include more than one vertex per event, with the PV as the leading index. We will describe here only one vertexing algorithm, called “FastHisto”. Other viable vertexing algorithms

are described in detail in Ref. [134].

The FastHisto algorithm is one of the simplest vertex reconstruction techniques and is based on three key principles:

- (1) The tracks associated to an individual vertex will have a small separation along the  $z$ -axis
- (2) Individual vertices will have a comparatively large separation
- (3) The primary vertex will be the one with the most activity, i.e. with the highest scalar sum of track  $p_T$

Ideally, the tracks will first undergo some selection process in order to remove fake tracks, tracks that result from spurious combinations of stubs and can often have very large  $p_T$ . However, this has not yet been fully studied, and the current lack-of-selection gives high vertex-finding efficiency. The input tracks are sorted into bins based on their  $z$ -position, essentially forming a histogram. Each entry is weighted by the track  $p_T^2$ , and the bin with the highest sum of the track  $p_T^2$  is assigned as the PV position. Figure 8.10 shows a visual of how the FastHisto algorithm works, using 200 PU interactions. The density of the tracks from the PV and their relatively high  $p_T$  make it easy to identify the PV amidst the PU tracks. The PV resolution is under 1 mm, as shown in Figure 8.11, which compares the performance in simulation to that in firmware-emulating software.

## 8.8 Track jets

A quick firmware-implemented track clustering algorithm allows for the creation of track-based jets, or track jets, from the L1 tracks. The track jets are then used in higher-level objects, such as track  $H_T$  and track  $H_T^{\text{miss}}$ , where  $H_T$  and  $H_T^{\text{miss}}$  are defined in Section 6.3.1. The same set of track quality criteria used for track  $p_T^{\text{miss}}$  (Table 8.2) is applied to the incoming tracks, as they were determined to be sufficient at keeping the track  $H_T$  and  $H_T^{\text{miss}}$  rates low while maintaining high trigger efficiency. These criteria make the track jets more resilient to PU and fake tracks. The track jets have a different  $\Delta z$  requirement than the track  $p_T^{\text{miss}}$ ; the clustering is done in parallel for ranges of  $z_{\text{trk}}$  and includes a  $\Delta z < 1$  cm requirement.

The clustering occurs in two steps in the  $\eta - \phi$  plane and is performed using a nearest-neighbor approach. The tracks are split into 24 regions in  $\eta$  and 27 regions in  $\phi$ . In the first layer of clustering, the sum of the track  $p_T$  in cell  $i$  ( $\sum p_T^{trk}$ ) <sub>$i$</sub>  is compared to the two neighboring cells in  $\phi$ , and the  $p_T$  is summed into the cell with the local maximum. The next step is clustering these  $\phi$ -clusters in  $\eta$ , also by nearest-neighbor. The cluster with the largest  $\sum p_T$  becomes the local maximum, and its neighbors are merged to create the final track jets. A schematic of the clustering steps is shown in Figure 8.12.

A comparison of the performance of this two-layer clustering to that of the anti- $k_T$  algorithm is shown in Figure 8.13, which shows high jet-clustering efficiency in both jet  $\eta$  and  $p_T$  for a SUSY stop sample with 200 PU. For the L1 trigger, triggers using the track jets will most likely use composite objects, like track  $H_T$  (the scalar sum of all  $p_T$ ) or track  $H_T^{\text{miss}}$  (the negative vectorial sum of the jet  $p_T$ ). The improved trigger efficiency for track  $H_T^{\text{miss}}$  at an L1 rate of 35 kHz after the full track selection requirements are applied is shown in Figure 8.14.

The option for extending tracking (Section 8.4.1) allows for the possibility of triggering on displaced track jets, which could be key to BSM searches. The displaced jets are still under study but are promising, as seen in Figure 8.7. Algorithmically, the displaced jets are created the same as the prompt jets but with looser track selection criteria. Additional criteria on the impact parameter  $d_0$  are put in place to ensure the jet is displaced, and we are currently performing optimization studies for the number of displaced tracks required within each jet for it to be tagged as displaced.

## 8.9 Trigger demonstrator tests

Ultimately, the L1 trigger is a hardware trigger, and the track-based algorithms described above will be implemented in firmware for the Phase 2 upgrade. In general, the baseline architecture for the GTT consists of 12 cards, each hosting a single Xilinx VU9P FPGA. The cards are arranged as a time-multiplexed system with a period of six, with two cards processing each event. Each card receives the full set of tracks from the track finder and sends the output track-based objects to the Global Trigger.

The track  $p_T^{\text{miss}}$  algorithm described in Section 8.6 has been implemented in firmware and tested on the FPGAs. We demonstrated these hardware tests, and the initial optimization studies, for the L1 TDR [134], which enabled the GTT to be included in the L1 trigger upgrade plan.

Since the primary vertex is required for the algorithm to remove PU tracks, the tracks are stored in 18 FIFOs while the vertex reconstruction is performed. When available, the primary vertex consistency and the track selection criteria listed in Table 8.2 are applied, and the remaining tracks are passed to the  $p_T^{\text{miss}}$  algorithm, which runs simultaneously for the 18 streams of tracks.

The FPGA resource usage for the track  $p_T^{\text{miss}}$  algorithm is shown in Table 8.3 and is within 3% for all categories on the VU9P. The algorithm was run at clock speeds up to 360 MHz and finishes 146 clock cycles after the primary vertex is available. This corresponds to a total of 171 (267) clock cycles from the arrival of the last (first) track each event which is within the required arrival time at the Global Trigger for the final event decision. Table 8.3 also shows the FPGA resource usage for the other GTT algorithms, the primary vertex (Section 8.7) and the track jets (Section 8.8).

| VU9P                | $n_{\text{FF}}$ | $n_{\text{LUT}}$ | $n_{\text{BRAM}}$ | $n_{\text{DSP}}$ | $f_{\text{clk,max}}$<br>[MHz] | Latency<br>(clocks) |
|---------------------|-----------------|------------------|-------------------|------------------|-------------------------------|---------------------|
| Available           | 2364480         | 1182240          | 2160              | 6840             | —                             | —                   |
| $p_T^{\text{miss}}$ | 31921 (1%)      | 35752 (3%)       | 2 (0%)            | 74 (1%)          | 360                           | 146                 |
| PV                  | 29137 (1%)      | 83741 (7%)       | 1 (0%)            | 1 (0%)           | 360                           | 121                 |
| Track jets          | 153890 (7%)     | 135117 (11%)     | 325 (15%)         | 0 (0.0%)         | 360                           | 175                 |

Table 8.3: Resource utilization of the VU9P FPGA for the GTT demonstrator system, at the time of the TDR.

## 8.10 Data quality monitoring

Data Quality Monitoring (DQM) is crucial for the health of the CMS detector and for providing a reliable certification of the data it collects [142, 143]. In general, the goal of the DQM system is to quickly discover and pinpoint errors or problems, rather in the detector itself or in the event reconstruction. The DQM tools allow for:

- creating, filling, and saving histograms with important detector or algorithm-specific vari-

ables

- automated quality and validity tests
- systems for online monitoring of the detector, offline monitoring of the reconstruction, and validation for new software releases and simulated samples
- visualization of the monitoring results
- workflows for the certification of data sets used for physics analyses

For online monitoring, the DQM system receives information from each subdetector, such as the trigger primitives (track segments, calorimeter clusters, and muon segments). Millions of histograms are filled for each daily run (about 17 hours) that are checked by automated systems and, when the automated tests indicate it is needed, by experts. Once the full run has been filled, the data are analyzed by the offline DQM system, which include more detailed checks than the online system. Each subdetector has a flag in the DQM framework that describes the final quality result. Lastly, the third component of the DQM system is release validation (RelVal). The RelVal DQM is run when the entire CMS software package is moved to a new release or to test the effects of any software changes or updates. The tests are run over a large sampling of physics processes, and the resulting histograms are compared bin-by-bin to corresponding histograms from an older release to identify any changes, intended or not.

The online and offline DQM tools for the Phase 2 upgrade of the LHC have not yet been written, since the HL-LHC will not begin running until 2027, but the RelVal DQM tools are currently useful for algorithm development and debugging. The DQM tools for the Track Trigger include histograms of basic track qualities (such as the  $p_T$  and the number of stubs per track) and variables for the stubs and clusters that make up each track. Additionally, the tools include truth-matched tracks, allowing for first-order studies of track resolution and efficiency. So far, these tools have been invaluable and have caught a number of unintended changes and bugs during the developmental stages of the L1 tracks, and were particularly helpful during the creation of the

hybrid tracking algorithm (see Section 8.4.1), which was created after the TDR and is still being studied and optimized.

The DQM tools for the other L1 trigger primitives (the calorimeter clusters and the muon hits and track segments) have also been written. While these have not been as immediately helpful, since these primitives currently exist during Phase 1 running, they will set the foundation and the framework for the online and offline DQM tools that will be developed in the future for Phase 2 running.

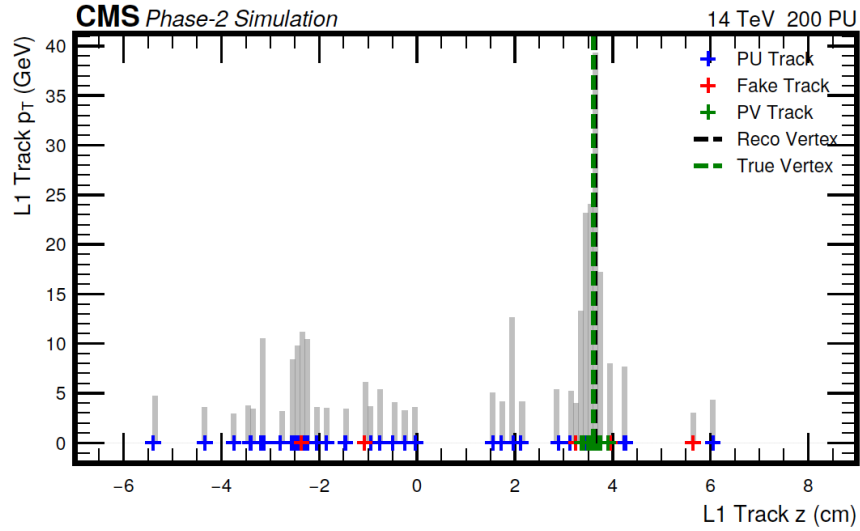


Figure 8.10: For a single event, the  $p_T$ -weighted distribution of the tracks'  $z$ -position. At the time of this study, the tracks were weighted by their  $p_T$ , not their  $p_T^2$ . The gray markers show the scalar sum of the track  $p_T$  per bin. The colored markers indicate the actual track position, with the source of the tracks shown using the different colors. The dashed lines indicate the  $z$ -position of the primary vertex, with the reconstructed vertex from the FastHisto algorithm in black, and the true generator-level vertex in green.

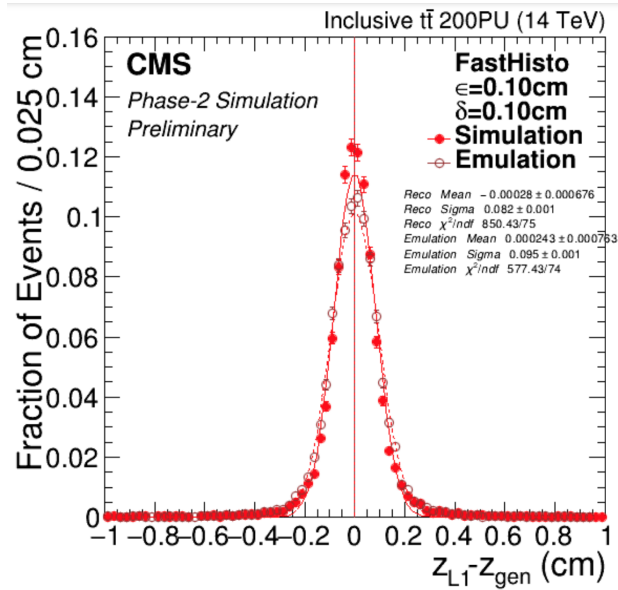


Figure 8.11: Comparison of the primary vertex resolution between simulation (filled circles) and emulation (open circles) of the FastHisto algorithm. For each, the reconstructed vertex is compared to the generator-level position of the interaction point.

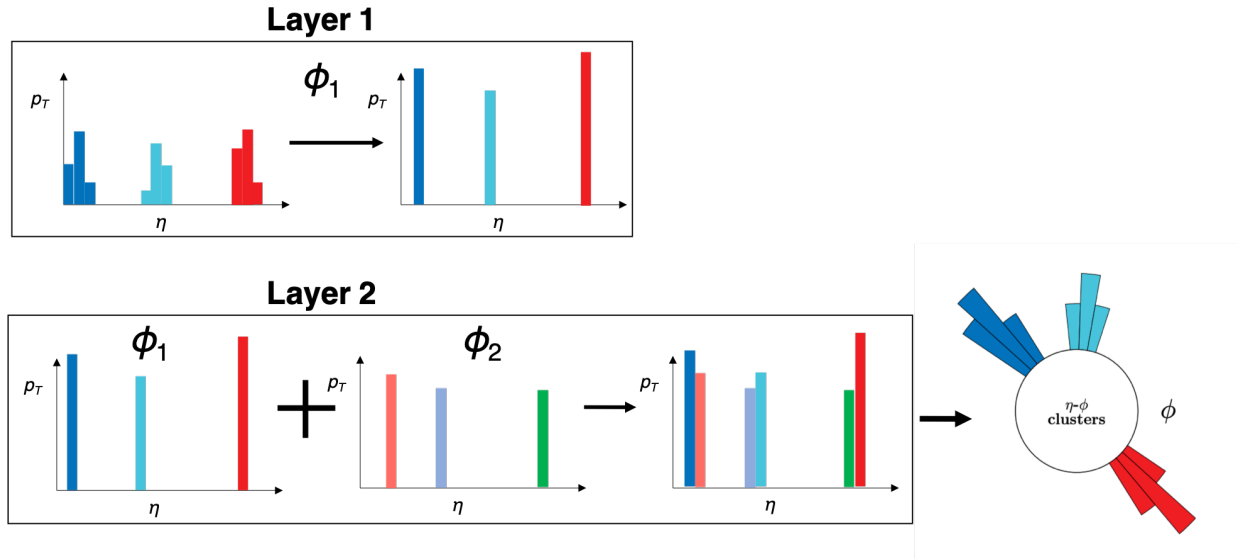


Figure 8.12: A diagram of the track-based two-layer jet clustering algorithm. In the first layer in each of the 27  $\phi$  bins, the nearest neighbors in  $\eta$  are clustered around a local  $p_T^{\max}$ . In the second layer, the  $\eta$ -clusters in each  $\phi$  bin are clustered around a local  $p_T^{\max}$ .

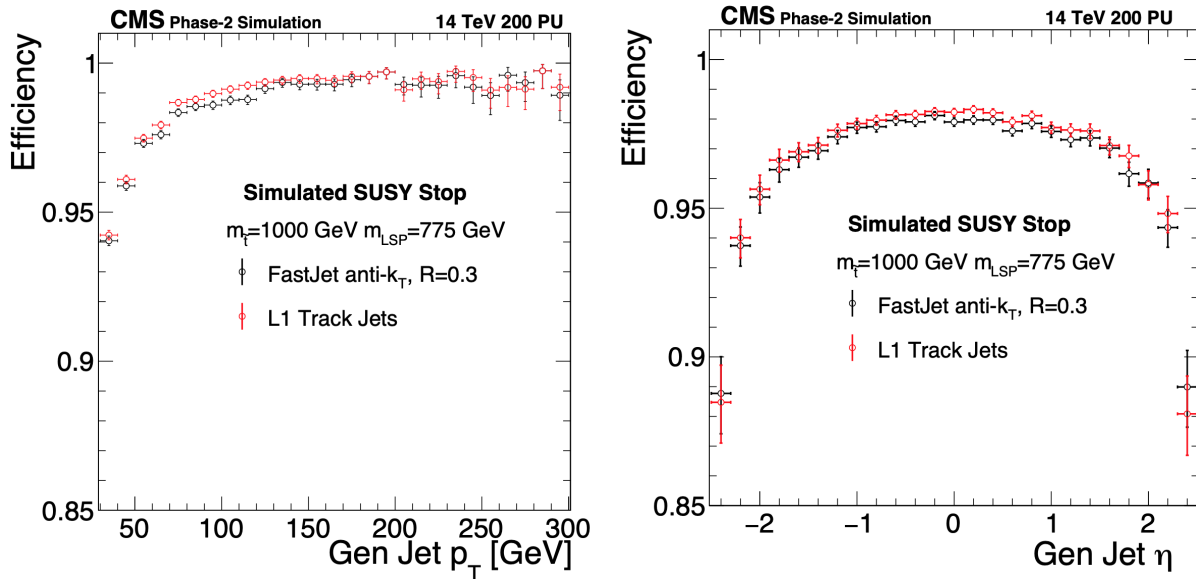


Figure 8.13: Comparison of the L1 track jet clustering and anti- $k_T$  clustering with  $R = 0.3$  using the same input collection of L1 reconstructed tracks from a SUSY stop sample with 200 PU. The efficiency for jet  $p_T$  (left) and  $\eta$  (right) is computed by matching the reconstructed jet to a generator-level jet within  $\Delta R < 0.4$ .



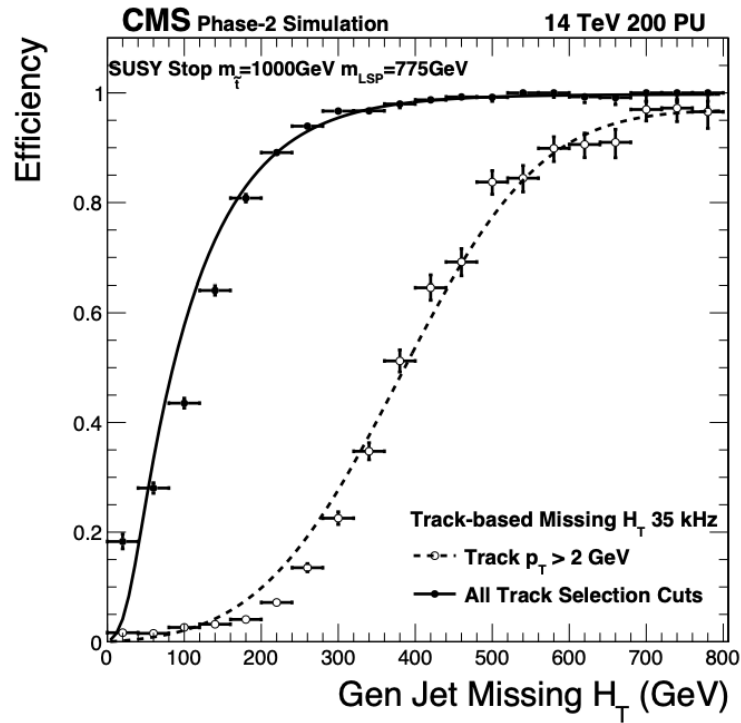


Figure 8.14: The trigger turn-on efficiency measured as the efficiency for events with generator-level  $H_T^{\text{miss}}$  to pass a given L1 trigger threshold. The generator-level jets used to compute these quantities have  $p_T > 30$  GeV and  $|\eta| < 2.4$  to emulate offline jet selection. A 35 kHz L1 threshold is chosen for the track  $H_T^{\text{miss}}$ . The 95% efficiency turn-on for the BSM stop signal is greatly lowered by the track selection requirements.

## Chapter 9

### Conclusions

The Standard Model of particle physics, a theory which describes the electroweak and strong interactions, continues to accurately describe our experimental results. Despite its success, the fine-tuning problem of the Higgs bare mass concerns many theorists, as it is usually an indication of an incomplete theory. Moreover, the SM can provide no explanation or description of dark matter. Supersymmetry is a simplistic and therefore appealing solution to these problems, providing a weakly-interacting candidate for dark matter and a set of superpartners that can cancel out the divergent quantum loop corrections to the Higgs mass.

This thesis presents a search for evidence of physics beyond the SM, targeting events with two Higgs bosons, each decaying to  $b\bar{b}$ , and  $p_T^{\text{miss}}$ , potentially indicative of an unknown particle escaping detection. The analysis is divided into two complementary regions of phase space, one where the two  $b$  quarks of each Higgs decay are contained in two distinct jets (resolved method) and one where the two  $b$  quarks are contained in a single wide-angle jet (boosted method). Careful treatment of the event classification is done to ensure no events are double-counted. The data-driven background estimation methods make use of orthogonal analysis regions in the Higgs candidate mass and  $b$ - or  $b\bar{b}$ -tagging requirements, and are further binned in  $\Delta R_{\text{max}}$  (for the resolved signature only) and  $p_T^{\text{miss}}$  (for both) to improve signal sensitivity. Standard Model backgrounds arise from  $Z$ +jets,  $W$ +jets,  $t\bar{t}$  (in which the leptons are not identified), and, to a smaller degree, QCD multijet production. Validation regions, where individual SM background processes are dominant, are defined to study and verify these processes and, in combination with simulation, are used to derive

corrections and uncertainties to the background prediction.

In all but one signal bin, the observed yields are statistically consistent with the background expectation. An excess is observed in one bin of the resolved signature, for which the global significance is  $1.9\sigma$  when all 16 bins are considered. After careful examination of the excess events, no exceptional features were found, and we determined the excess was not kinematically-aligned with our signal models.

These results are used to set limits on the production cross sections of certain supersymmetric particles, where we consider the electroweak production of nearly-degenerate higgsinos and the strong production of gluino pairs. Comparisons of our limits to the theoretical cross sections allow us to exclude  $\tilde{\chi}_1^0$  masses from 175-1025 GeV for higgsinos who decay to a  $\tilde{\chi}_1^0$  that in turn decays through a Higgs boson to the lightest SUSY particle, a massless goldstino. Considering only the  $\tilde{\chi}_2^0\tilde{\chi}_3^0$  production and a bino LSP, a small region where  $m(\tilde{\chi}_1^0)$  is less than 15 GeV is excluded; for  $m(\tilde{\chi}_1^0) = 1$  GeV,  $\tilde{\chi}_2^0$  masses between 265 and 305 GeV are excluded. For the gluino-pair production, we exclude gluinos with masses below 2330 GeV, where the  $\tilde{\chi}_2^0$  mass is set to be 50 GeV below the gluino mass, and the LSP has a mass of 1 GeV. These exclusions place significant constraints on natural SUSY, suggesting that SUSY alone might be insufficient to solve the problem of fine-tuning and that additional theory might be necessary.

Additionally, the results of this search were used to extend the reach of other SUSY models in an electroweak combination, extending electroweakino production limits further than any individual analysis. The current limits are rather stringent, expecting to exclude neutralino masses up to 980 GeV for the HH topology, up to 730 GeV for the ZZ topology, and up to 700 GeV for the mixed HZ topology. This search is still an active area of work, and the results will change once the signal regions are unblinded.

While newer and more innovative techniques could help unlock SUSY, the key to finding SUSY is more collision data. The LHC will continue running after this thesis is completed, defended, and archived, with plans for Run 3 to begin in 2022 and the HL-LHC (Phase 2) to begin running in 2027. Run 3 will reach 13.5 TeV center-of-mass energy and an integrated luminosity of up to

500 fb<sup>-1</sup>. The increased total data has a greater potential for new discoveries in physics.

Since the center-of-mass energy is restricted by the ring size of the LHC to 14 TeV, Phase 2 will instead focus on increasing the instantaneous luminosity by reducing the beam size at the collision point. Most importantly, the HL-LHC will have new 12 T quadrupole magnets (compared to 8 T of Phase 1), which have already been built and tested. Additionally, new and stronger (11 T) dipole magnets will be inserted to more powerfully bend the trajectories. The entire accelerator chain complex will be upgraded or replaced entirely to accommodate the luminosity required by the HL-LHC. The tracker and parts of the calorimeter and muon systems will also be replaced with more radiation-hard materials. The HL-LHC is projected to collect an enormous 4000 fb<sup>-1</sup> of data, which will improve the likelihood of detecting rare processes and improve statistically marginal measurements.

Extensive effort has gone into reconstructing the information from the tracker that will be available at Level 1 during Phase 2. The tracker will enable us to trigger on track-based objects, which are extra resilient against pileup, a necessary feature considering the 200 pileup interactions expected during Phase 2 runs. The inclusion of displaced tracking at L1, while not yet in the baseline plan, could greatly extend our physics potential for displaced signatures and would allow us to create higher-level objects (such as displaced vertices and displaced track jets) that provide a unique handle to trigger on. The reconstructed primary vertex will allow for track-based  $p_T^{\text{miss}}$ ; the analysis in this thesis highlighted the importance of triggering on  $p_T^{\text{miss}}$ , and track-based  $p_T^{\text{miss}}$  has the benefit of mitigating the effects of PU.  $p_T^{\text{miss}}$  could be the key to, and potentially only indicator of, new physics in future analyses, and the HL-LHC has a chance to unlock it.

## Bibliography

- [1] Lyndon Evans and Philip Bryant. LHC machine. Journal of Instrumentation, 3(08):S08001–S08001, AUG 2008.
- [2] Sarah Charley. LHC breaks energy record. Symmetry magazine, Apr 2015.
- [3] D. J. Griffiths. Introduction to Elementary Particles. Wiley-VCH, 2 edition, 2008.
- [4] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B, 716:30–61, 2012.
- [5] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. B, 716:1–29, 2012.
- [6] Howard Baer and Xerxes Tata. The Minimal Supersymmetric Standard Model, pages 127–189. Cambridge University Press, 2006.
- [7] F. Halzen and A. D. Martin. Quarks & Leptons. John Wiley & Sons, 1984.
- [8] C. Burgess and G. Moore. The Standard Model: A Primer. Cambridge University Press, 2007.
- [9] M. Thomson. Modern Particle Physics. Cambridge, 6 edition, 2016. Figures accessed online on 20 November 2021. <https://www.hep.phy.cam.ac.uk/~thomson/MPP/ModernParticlePhysics.html>.
- [10] Rosier-Lees Djouadi et al. The Minimal Supersymmetric Standard Model: Group Summary Report. Group Report hep-ph/9901246, PM-98-45, CERN, 1998.
- [11] Wikipedia contributors. Standard Model – Wikipedia, the free encyclopedia. Online; [https://en.wikipedia.org/wiki/Standard\\_Model](https://en.wikipedia.org/wiki/Standard_Model), accessed 20 November 2021.
- [12] P.A. Zyla et al. Review of Particle Physics. PTEP, 2020(8):083C01, 2020.
- [13] ChasEpstn. Bhabha T-channel. Online; <https://commons.wikimedia.org/w/index.php?curid=35015622>, accessed 29 December 2021.
- [14] Particle Data Group. Review of Particle Physics. Progress of Theoretical and Experimental Physics, 2020(8), 08 2020. 083C01.
- [15] Joel Holdsworth. Beta Negative Decay. Online; <https://commons.wikimedia.org/w/index.php?curid=1754301>, accessed 20 November 2021.

- [16] Steven Weinberg. A model of leptons. Physical Review Letters, 19(21):1264–1266, Nov 1967.
- [17] F. Abe, H. Akimoto, A. Akopian, M. G. Albrow, S. R. Amendolia, D. Amidei, J. Antos, C. Anway-Wiese, S. Aota, G. Apollinari, and et al. Observation of top quark production in  $p\bar{p}$  collisions with the collider detector at fermilab. Physical Review Letters, 74(14):2626–2631, Apr 1995.
- [18] S. Abachi, B. Abbott, M. Abolins, B. S. Acharya, I. Adam, D. L. Adams, M. Adams, S. Ahn, H. Aihara, J. Alitti, and et al. Observation of the top quark. Physical Review Letters, 74(14):2632–2637, Apr 1995.
- [19] K. Kodama, N. Ushida, C. Andreopoulos, N. Saoulidou, G. Tzanakos, P. Yager, B. Baller, D. Boehnlein, W. Freeman, B. Lundberg, and et al. Observation of tau neutrino interactions. Physics Letters B, 504(3):218–224, Apr 2001.
- [20] Susanne Mertens. Direct Neutrino Mass Experiments. Journal of Physics: Conference Series, 718:022013, May 2016.
- [21] Maria Concepcion Gonzalez-Garcia and Michele Maltoni. Phenomenology with Massive Neutrinos. Physics Reports, 460:1–129, 2007.
- [22] P. J. E. Peebles and Bharat Ratra. The cosmological constant and dark energy. Reviews of Modern Physics, 75(2):559–606, Apr 2003.
- [23] Anto. I. Lonappan, Sumit Kumar, Ruchika, Bikash R. Dinda, and Anjan A. Sen. Bayesian evidences for dark energy models in light of current observational data. Physical Review D, 97(4), Feb 2018.
- [24] Virginia Trimble. Existence and Nature of Dark Matter in the Universe. Annual Review of Astronomy and Astrophysics, 25(1):425–472, 1987.
- [25] Leonard Susskind. The gauge hierarchy problem, technicolor, supersymmetry, and all that. Phys. Rept., 104:181–193, 1984.
- [26] CMSPublic Web. 2018 bbttag Data/Monte Carlo Correction Factors. [https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation102X#DoubleB\\_and\\_DeepDoubleBvL\\_2018\\_S](https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation102X#DoubleB_and_DeepDoubleBvL_2018_S), 2021.
- [27] S. Dimopoulos, S. Raby, and Frank Wilczek. Supersymmetry and the scale of unification. Phys. Rev. D, 24:1681, 1981.
- [28] Stephen P. Martin. A Supersymmetry Primer. Adv. Ser. Direct. High Energy Phys., 18:1, 1998.
- [29] Pierre Fayet and S. Ferrara. Supersymmetry. Phys. Rept., 32:249, 1977.
- [30] Madeleine O’Keefe. Fine-tuning versus naturalness, Jan 2021.
- [31] Steven Weinberg. The Quantum Theory of Fields, Volume 1: Foundations. Cambridge University Press, 2005.
- [32] Jonathan L. Feng. Naturalness and the Status of Supersymmetry. Ann. Rev. Nucl. Part. Sci., 63:351, 2013.

- [33] Edward Witten. Dynamical Breaking of Supersymmetry. Nucl. Phys. B, 188:513, 1981.
- [34] Savas Dimopoulos and Howard Georgi. Softly Broken Supersymmetry and SU(5). Nucl. Phys. B, 193:150, 1981.
- [35] Abdus Salam Jeffrey Goldstone and Steven Weinberg. Broken Symmetries. Phys. Rev., 127:965, 1962.
- [36] Savas Dimopoulos, Michael Dine, Stuart Raby, and Scott D. Thomas. Experimental signatures of low-energy gauge mediated supersymmetry breaking. Phys. Rev. Lett., 76:3494–3497, 1996.
- [37] Konstantin T. Matchev and Scott D. Thomas. Higgs and Z boson signatures of supersymmetry. Phys. Rev., D62:077702, 2000.
- [38] P. Binétruy. Supersymmetry: Theory, Experiment, and Cosmology. Oxford Graduate Texts, 2006.
- [39] Toby Falk, Keith A. Olive, and Mark Srednicki. Heavy sneutrinos as dark matter. Physics Letters B, 339(3):248–251, Nov 1994.
- [40] Andreas Redelbach. Searches for Prompt  $R$ -Parity-Violating Supersymmetry at the LHC. Advances in High Energy Physics, 2015.
- [41] P.S. Bhupal Dev, Dilip Kumar Ghosh, and Werner Rodejohann.  $R$ -parity violating supersymmetry at IceCube. Physics Letters B, 762:116–123, Nov 2016.
- [42] CMSPublic Web. LHC SUSY Cross Section Working Group. [https://twiki.cern.ch/twiki/pub/LHCPhysics/SUSYCrossSections/SUSY\\_xsecs\\_13TeV\\_overview.pdf](https://twiki.cern.ch/twiki/pub/LHCPhysics/SUSYCrossSections/SUSY_xsecs_13TeV_overview.pdf), 2019.
- [43] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogio, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. Escalante Del Valle, and et al. Searches for physics beyond the standard model with the  $M_{T2}$  variable in hadronic final states with and without disappearing tracks in proton–proton collisions at  $\sqrt{s} = 13$  TeV. The European Physical Journal C, 80(1), Jan 2020.
- [44] A. M. Sirunyan, A. Tumasyan, W. Adam, J. W. Andrejkovic, T. Bergauer, S. Chatterjee, M. Dragicevic, A. Escalante Del Valle, R. Frühwirth, M. Jeitler, and et al. Search for top squark production in fully hadronic final states in proton-proton collisions at  $\sqrt{s} = 13$  TeV. Physical Review D, 104(5), Sep 2021.
- [45] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogio, T. Bergauer, M. Dragicevic, J. Erö, A. Escalante Del Valle, M. Flechl, R. Frühwirth, and et al. Search for supersymmetry in pp collisions at  $\sqrt{s} = 13$  TeV with  $137\text{ fb}^{-1}$  in final states with a single lepton using the sum of masses of large-radius jets. Physical Review D, 101(5), Mar 2020.
- [46] Albert M Sirunyan et al. Search for physics beyond the standard model in events with jets and two same-sign or at least three charged leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV. Eur. Phys. J. C, 80(8):752, 2020.

- [47] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambroggi, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. Escalante Del Valle, and et al. Search for supersymmetry in proton-proton collisions at 13 TeV in final states with jets and missing transverse momentum. Journal of High Energy Physics, 2019(10), Oct 2019.
- [48] CMS Collaboration. Search for electroweak production of supersymmetric particles in final states containing hadronic decays of WW, WZ, or WH and missing transverse momentum. Physics Analysis Summary CMS-PAS-SUS-21-002, CERN, Geneva, 2021.
- [49] Armen Tumasyan et al. Search for chargino-neutralino production in events with Higgs and W bosons using  $137 \text{ fb}^{-1}$  of proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$ . JHEP, 10:045, 2021.
- [50] Armen Tumasyan et al. Search for electroweak production of charginos and neutralinos in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$ . JHEP, 6 2021.
- [51] Albert M Sirunyan et al. Search for supersymmetry in final states with two oppositely charged same-flavor leptons and missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$ . JHEP, 04:123, 2021.
- [52] Armen Tumasyan et al. Search for supersymmetry in final states with two or three soft leptons and missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$ . JHEP, 11 2021.
- [53] CMS Collaboration. CMS Supersymmetry Physics Results. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>, 2021.
- [54] Oliver Sim Bruning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. LHC Design Report. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.
- [55] Esma Mobs. The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019. General Photo, Jul 2019.
- [56] CMS Collaboration and A Holzner. 78 reconstructed vertices in event from high-pileup run 198609. CMS Collection., Sep 2012.
- [57] Tai Sakuma. Cutaway diagrams of CMS detector. <http://cds.cern.ch/record/2665537>, May 2019.
- [58] David Barney. CMS Detector Slice. CMS Collection., Jan 2016.
- [59] Detector Performance Group. CMS Tracker Detector Performance Results. [https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK#Sketches\\_of\\_the\\_CMS\\_Tracker\\_Dete](https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK#Sketches_of_the_CMS_Tracker_Dete), 2020.
- [60] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. JINST, 9:P10009. 80 p, May 2014. Comments: Replaced with published version. Added journal reference and DOI.
- [61] Andrea Benaglia. The CMS ECAL performance with examples. JINST, 9:C02008, 2014.
- [62] CMS Collaboration. The CMS electromagnetic calorimeter project: Technical Design Report. CMS Technical Design Report CERN-LHCC-97-033, CMS-TDR-4, Dec 1997.



- [63] Andrea Benaglia. The CMS ECAL performance with examples. Conference Report CMS-CR-2013-430, CERN, Geneva, Nov 2013.
- [64] A.M. Sirunyan, A. Tumasyan, W. Adam, F. Ambroggi, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. Escalante Del Valle, M. Flechl, and et al. Calibration of the CMS hadron calorimeters using proton-proton collision data at  $\sqrt{s} = 13$  TeV. Journal of Instrumentation, 15(05):P05002–P05002, May 2020.
- [65] J Mans, J Anderson, B Dahmes, P de Barbaro, J Freeman, T Grassi, E Hazen, J Mans, R Ruchti, I Schimdt, T Shaw, C Tully, J Whitmore, and T Yetkin. CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter. Technical Design Report CERN-LHCC-2012-015, CMS-TDR-10, CERN, Sep 2012.
- [66] CMS collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. JINST, 5:T03021. 35 p, Oct 2009.
- [67] CMS Collaboration. Performance of the CMS Muon Detectors in 2016 collision runs. Detector Performance Note CMS-DP-2016-046, CERN, Jul 2016.
- [68] A M Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13$  TeV. JINST, 13:P06015, 2018.
- [69] Barbara Alvarez Gonzalez. CMS Muon Drift Tubes HL-LHC Slice Test. Oct 2021. arXiv: 2110.15052.
- [70] CMS Collaboration. Performance of the CMS Drift Tube Chambers with Cosmic Rays. Performance of the CMS Drift Tube Chambers with Cosmic Rays. JINST, 5:T03015 . 45 p, Nov 2009.
- [71] CMS EMTF DQM group. Cathode strip chambers. <https://twiki.cern.ch/twiki/bin/view/Main/NEUNewCoopGuide>, 2017.
- [72] Andres Cabrera. Endcap HV Cable Reshuffling. <https://twiki.cern.ch/twiki/bin/view/Main/EndcapReshufflingRpc>, 2015.
- [73] Roumyana Hadjiiska. RPC Performance Results for 2016, approved for EPS Conference in Venice. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/RPCDPGResultsforVenice>, 2017.
- [74] CMS Collaboration. Performance and Operation of the CMS Electromagnetic Calorimeter. JINST, 5:T03010. 39 p, Oct 2009.
- [75] Walaa Elmetenawee. CMS track reconstruction performance during Run 2 and developments for Run 3. PoS, ICHEP2020:733. 8 p, Dec 2020.
- [76] Silvia Goy Lopez. CMS Detector Performance during LHC Run 1 and projections for Run 2. Nucl. Part. Phys. Proc., 273-275:1048–1054. 7 p, 2016.
- [77] Wolfgang Adam, R Frühwirth, Are Strandlie, and T Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. CMS Note CMS-NOTE-2005-001, CERN, Geneva, Jan 2005.

- [78] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. Proceedings of the IEEE, 86(11):2210–2239, 1998.
- [79] R Fruhwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Adaptive Vertex Fitting. CMS Note CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.
- [80] A.M. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, M. Flechl, and et al. Particle-flow reconstruction and global event description with the cms detector. Journal of Instrumentation, 12(10):P10003–P10003, Oct 2017.
- [81] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- $k_t$  jet clustering algorithm. JHEP, 04:063, 2008.
- [82] CMS Collaboration. Jet performance in pp collisions at  $\sqrt{s} = 7$  TeV. CMS Physics Analysis Summary CMS-PAS-JME-10-003, CERN, 2010.
- [83] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. JINST, 13:P05011, 2018.
- [84] CMS Collaboration. Performance of deep tagging algorithms for boosted double quark jet topology in proton-proton collisions at 13 tev with the phase-0 cms detector. CMS Detector Performance Note CMS-DP-2018-046, CERN, 2018.
- [85] Jesse Thaler and Ken Van Tilburg. Identifying boosted objects with n-subjettiness. Journal of High Energy Physics, 2011(3), Mar 2011.
- [86] CMS B-tag POG. Methods to apply b-tagging efficiency scale factors. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMETHODS>, 2015.
- [87] CMS Collaboration. Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector. Physics Analysis Summary CMS-PAS-JME-16-004, CERN, Geneva, 2016.
- [88] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector. CERN, 2016.
- [89] CMS Collaboration. Search for higgsinos in channels with two Higgs bosons and missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13$  TeV. Physics Analysis Summary CMS-PAS-SUS-20-004, CERN, Geneva, 2021.
- [90] J. B. Kim, M. Oshiro, A. Ovcharova, J. Richman, W. T. Ford, E. MacDonald, R. Patel, A. Perloff, K. Stenson, K. A. Ulmer, and N. Zipper. Search for new physics in channels with two Higgs bosons and missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13$  TeV. CMS AN-19-112. Available on the CMS information server., 2022.
- [91] Daniele Alves, Nima Arkani-Hamed, Sanjay Arora, Yang Bai, Matthew Baumgart, Joshua Berger, Matthew Buckley, Bart Butler, Spencer Chang, Hsin-Chia Cheng, and et al. Simplified models for LHC new physics searches. Journal of Physics G: Nuclear and Particle Physics, 39(10):105005, Sep 2012.

- [92] CMS Collaboration. Interpretation of searches for supersymmetry with simplified models. Phys. Rev. D, 88:052017. 24 p, Jan 2013. Comments: Submitted to Phys. Rev. D.
- [93] Joshua T. Ruderman and David Shih. General Neutralino NLSPs at the Early LHC. JHEP, 08:159, 2012.
- [94] Benjamin Fuks, Michael Klasen, David R. Lamprea, and Marcel Rothering. Gaugino production in proton-proton collisions at a center-of-mass energy of 8 TeV. JHEP, 10:081, 2012.
- [95] Benjamin Fuks, Michael Klasen, David R. Lamprea, and Marcel Rothering. Precision predictions for electroweak superpartner production at hadron colliders with RESUMMINO. Eur. Phys. J. C, 73:2480, 2013.
- [96] Graham D. Kribs, Adam Martin, Tuhin S. Roy, and Michael Spannowsky. Discovering Higgs Bosons of the MSSM using Jet Substructure. Phys. Rev. D, 82:095012, 2010.
- [97] Stefania Gori, Pedro Schwaller, and Carlos E.M. Wagner. Search for Higgs Bosons in SUSY Cascade Decays and Neutralino Dark Matter. Phys. Rev. D, 83:115022, 2011.
- [98] BRIL Group. Bril work suite. <http://cms-service-lumi.web.cern.ch/cms-service-lumi/brilwsdoc.html>, 2017.
- [99] CMS Collaboration. Cms luminosity - public results. [https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Multi\\_year\\_plots](https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Multi_year_plots), 2020.
- [100] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. JHEP, 07:079, 2014.
- [101] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. JHEP, 06:043, 2010.
- [102] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. Comput. Phys. Commun., 191:159, 2015.
- [103] S. Agostinelli et al. GEANT4 — a simulation toolkit. Nucl. Instrum. Meth. A, 506:250, 2003.
- [104] S. Abdullin, P. Azzi, F. Beaudette, P. Janot, and A. Perrotta. The fast simulation of the CMS detector at LHC. J. Phys. Conf. Ser., 331:032049, 2011.
- [105] Richard D. Ball et al. Parton distributions for the LHC Run II. JHEP, 04:040, 2015.
- [106] A.M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, A. Escalante Del Valle, R. Frühwirth, M. Jeitler, N. Krammer, L. Lechner, and et al. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. Journal of Instrumentation, 16(05):P05014, May 2021.

- [107] A.M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogio, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, and et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s}=13$  TeV. Journal of Instrumentation, 13(06):P06015–P06015, Jun 2018.
- [108] Keith Rehermann and Brock Tweedie. Efficient Identification of Boosted Semileptonic Top Quarks at the LHC. JHEP, 03:059, 2011.
- [109] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft Drop. JHEP, 05:146, 2014.
- [110] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. Towards an understanding of jet substructure. JHEP, 09:029, 2013.
- [111] CMS Collaboration. CMS Luminosity Measurements for the 2016 Data Taking Period. Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, Geneva, 2017.
- [112] Andreas Kornmayer. The CMS Pixel Luminosity Telescope. Conference Report CMS-CR-2015-121, CERN, Geneva, Jun 2015.
- [113] A. Bell, E. Castro, R. Hall-Wilton, W. Lange, W. Lohmann, A. Macpherson, M. Ohlerich, N. Rodriguez, V. Ryjov, R.S. Schmidt, and et al. Fast beam conditions monitor BCM1F for the CMS experiment. Nuclear Instruments and Methods in Physics Research, 614(3):433–438, Mar 2010.
- [114] CMS Collaboration. CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. Physics Analysis Summary CMS-PAS-LUM-13-001, CERN, Geneva, 2013.
- [115] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at  $\sqrt{s} = 13$  TeV. Physics Analysis Summary CMS-PAS-LUM-17-004, CERN, Geneva, 2018.
- [116] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at  $\sqrt{s} = 13$  TeV. Physics Analysis Summary CMS-PAS-LUM-18-002, CERN, Geneva, 2019.
- [117] CMS B-tag POG. 2016 Legacy FastSim Correction Factors. [https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation2016Legacy#FastSim\\_correction\\_factors](https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation2016Legacy#FastSim_correction_factors), 2016.
- [118] CMS B-tag POG. 2017 FastSim Correction Factors. [https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation94X#FastSim\\_correction\\_factors](https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation94X#FastSim_correction_factors), 2017.
- [119] CMS B-tag POG. 2018 FastSim Correction Factors. [https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation102X#FastSim\\_correction\\_factors](https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation102X#FastSim_correction_factors), 2018.
- [120] CMS B-tag POG. 2016 bbttag Data/Monte Carlo Correction Factors. [https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation2016Legacy#DoubleB\\_2016\\_Scale\\_Factors\\_in\\_QC](https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation2016Legacy#DoubleB_2016_Scale_Factors_in_QC), 2016.
- [121] CMS B-tag POG. 2017 bbttag Data/Monte Carlo Correction Factors. [https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation94X#DDBvL\\_2017\\_Scale\\_Factors\\_in\\_QCD](https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation94X#DDBvL_2017_Scale_Factors_in_QCD), 2017.

- [122] CMS B-tag POG. Summaries of CMS cross section measurements. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined>, 2018.
- [123] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi. The  $t\bar{t}$  cross-section at 1.8 TeV and 1.96 TeV: A Study of the systematics due to parton densities and scale dependence. *JHEP*, 04:068, 2004.
- [124] Stefano Catani, Daniel de Florian, Massimiliano Grazzini, and Paolo Nason. Soft gluon resummation for Higgs boson production at hadron colliders. *JHEP*, 07:028, 2003.
- [125] Alexander L. Read. Presentation of search results: The CL<sub>s</sub> technique. *J. Phys. G*, 28:2693, 2002.
- [126] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999.
- [127] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: 10.1140/epjc/s10052-013-2501-z].
- [128] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, CERN, 2011.
- [129] CMS Higgs Combination Group. Documentation of the roostats-based statistics tools for higgs pag. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit>, 2015.
- [130] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambroggi, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, and et al. Search for higgsino pair production in pp collisions at  $\sqrt{s} = 13$  tev in final states with large missing transverse momentum and two higgs bosons decaying via  $h \rightarrow b\bar{b}$ . *Physical Review D*, 97(3), Feb 2018.
- [131] ATLAS Collaboration. Search for pair production of higgsinos in final states with at least three  $b$ -tagged jets in  $\sqrt{s} = 13$  TeV  $pp$  collisions using the ATLAS detector. *Phys. Rev. D*, 98:092002. 34 p, Jun 2018. Submitted to Phys.Rev.
- [132] Albert M Sirunyan et al. Search for Physics Beyond the Standard Model in Events with High-Momentum Higgs Bosons and Missing Transverse Momentum in Proton-Proton Collisions at 13 TeV. *Phys. Rev. Lett.*, 120(24):241801, 2018.
- [133] A. M. Sirunyan et al. Combined search for electroweak production of charginos and neutralinos in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP*, 03:160, 2018.
- [134] CMS Collaboration. The Phase-2 Upgrade of the CMS Level-1 Trigger. CMS Technical Design Report CERN-LHCC-2020-004, CMS-TDR-021, CERN, Geneva, Apr 2020.
- [135] Klaus Hanke, Julie Coupard, Heiko Damerou, Anne Funken, Brennan Goddard, Alessandra Lombardi, Django Manglunki, Simon Mataguez, Malika Meddahi, Bettina Mikulec, Giovanni Rumolo, Richard Scrivens, Elena Shaposhnikova, and Maurizio Vretenar. The LHC Injectors Upgrade (LIU) Project at CERN: Proton Injector Chain. page WEPVA036. 4 p, 2017.

- [136] CMS Collaboration. High-Luminosity Large Hadron Collider (HL-LHC): Technical design report. CMS Technical Design Report CERN-2020-010, Oct 2020.
- [137] R Garoby, G Bellodi, F Gerigk, K Hanke, A Lombardi, M Pasini, C Rossi, E Sargsyan, and M Vretenar. Linac4, a New Injector for the CERN PS Booster. Linac4, the future Injector of the CERN complex of Proton accelerators. page 4 p, Jul 2006. revised version submitted on 2006-08-25 16:44:41.
- [138] Francois-Xavier Nuiry, Oliver Aberle, Maxime Bergeret, Alessandro Bertarelli, Nicolo Biancacci, Roderik Bruce, Marco Calviani, Federico Carra, Alessandro Dallochio, Luca Gentini, Simone Gilardoni, Ricardo Illan Fiastre, Inigo Lamas Garcia, Alessandro Masi, Antonio Perillo-Marccone, Stefano Pianese, Stefano Redaelli, Emilien Rigutto, and Benoit Salvant. Design and Prototyping of New CERN Collimators in the Framework of the LHC Injector Upgrade (LIU) Project and the High-Luminosity (HL-LHC) Project. page MOPAB006. 4 p, 2017.
- [139] Elena Shaposhnikova, Julie Coupard, Heiko Damerau, Anne Funken, Simone Gilardoni, Brennan Goddard, Klaus Hanke, Lelyzaveta Kobzeva, Alessandra Lombardi, Django Manglunki, Simon Mataguez, Malika Meddahi, Bettina Mikulec, Giovanni Rumolo, Richard Scrivens, and Maurizio Vretenar. LHC Injectors Upgrade (LIU) Project at CERN. page MOPOY059. 4 p, 2016.
- [140] Malika Meddahi, Julie Coupard, Heiko Damerau, Anne Funken, Simone Gilardoni, Brennan Goddard, Klaus Hanke, Lelyzaveta Kobzeva, Alessandra Lombardi, Django Manglunki, Simon Mataguez, Bettina Mikulec, Giovanni Rumolo, Elena Shaposhnikova, and Maurizio Vretenar. LHC Injectors Upgrade (LIU) Project at CERN. page THPF093. 4 p, 2015.
- [141] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup Per Particle Identification. JHEP, 10:059, 2014.
- [142] Marcel Andre Schneider. The Data Quality Monitoring Software for the CMS experiment at the LHC: past, present and future. Conference Report CMS-CR-2018-228, CERN, Geneva, Sep 2018.
- [143] L Tuura, A Meyer, I Segoni, and G Della Ricca. CMS data quality monitoring: Systems and experiences. J. Phys.: Conf. Ser., 219:072020. 9 p, 2010.

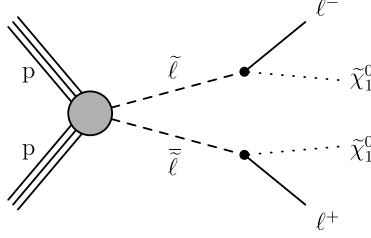


Figure A.1: Direct slepton pair production, with each slepton decaying into a lepton and  $\tilde{\chi}_1^0$ , which we take as the LSP.

## Appendix A

### Additional interpretations for the EWK Combination

#### A.1 Additional models

For the electroweakino search, we consider a process of slepton pair production where each slepton decays into a lepton and the lightest neutralino  $\tilde{\chi}_1^0$ , which we take as the LSP. The final state is expected to include two opposite-sign (OS), same-flavor (SF) leptons (electrons or muons). The process is shown in Fig. A.1. We assume any other sparticle is decoupled and cannot be included in the process.

For the neutralino and chargino production that are assumed to produce two bosons (W, Z, or H) through subsequent SUSY decays, the second set of models involves the production of  $\tilde{\chi}_1^0$  and  $\tilde{\chi}_2^0$ , as shown Figure A.2. The  $\tilde{\chi}_1^\pm$  is assumed to always decay into a W boson and a  $\tilde{\chi}_1^0$  (taken to be the LSP), while we consider three decay modes for  $\tilde{\chi}_2^0$ :  $\tilde{\chi}_2^0 \rightarrow Z\tilde{\chi}_1^0$  (WZ topology),  $\tilde{\chi}_2^0 \rightarrow H\tilde{\chi}_1^0$  (WH topology), and 50% of each of these two decays (mixed topology). The production cross sections are computed in the limit of mass-degenerate winos  $\tilde{\chi}_1^\pm$  and  $\tilde{\chi}_2^0$ , and a light bino-like  $\tilde{\chi}_1^0$ , since we

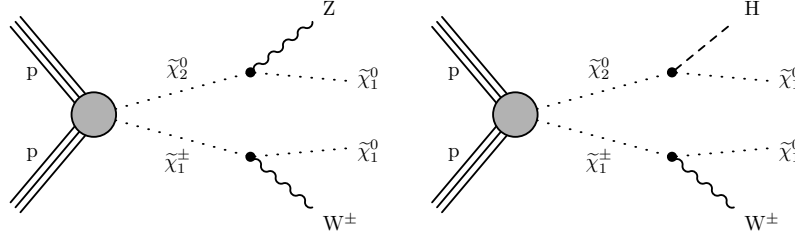


Figure A.2: Production of  $\tilde{\chi}_1^\pm$  and  $\tilde{\chi}_2^0$ , with the  $\tilde{\chi}_1^\pm$  decaying to a W boson and the LSP,  $\tilde{\chi}_1^0$ , and the  $\tilde{\chi}_2^0$  decaying to either a Z boson and the  $\tilde{\chi}_1^0$  (left) or a H boson and the  $\tilde{\chi}_1^0$  (right).

assume that all other sparticles are heavy and decoupled.

## A.2 Search for an oppositely charged same-flavor lepton pair

The search described in Section 7.1.2 has additional signal regions regarding the production of sleptons. In this case, events are required to have a leading lepton  $p_T$  greater than 50 GeV. Events are then required to have a dilepton invariant mass that is not compatible with a Z boson,  $m_{T2}(\ell\ell) > 110$  GeV. Finally, events are classified according to their jet multiplicity: all events without jets are used to build the jet-less category while events with at least one jet are required to have  $p_T^{\ell 2}/p_T^{j1} > 1.2$ , to reject events with large hadronic activity. For the signal extraction, different categories are also built based on the  $p_T^{\text{miss}}$  of the event.

The Drell-Yan background is estimated by considering the data event yield in a selection of events that replicates that of the signal region, but requiring the invariant mass of the lepton pair to be consistent with a Z boson. Then, this contribution is extrapolated to the signal region by means of a transfer factor obtained from a sample enriched in Drell-Yan events.

## A.3 Interpretations

Table A.1 summarizes which searches are used to place exclusion limits for all topologies considered in the full Combination.

The results for the neutralino-chargino production for the WZ topology, the WH topology, and the mixed topology with 50% branching fraction to each of WZ and WH are shown in Fig-



| Search                | Signal topology |    |    |    |    |  | sleptons |
|-----------------------|-----------------|----|----|----|----|--|----------|
|                       | WZ              | WH | ZZ | ZH | HH |  |          |
| 1 $\ell$ 2b           |                 | ✓  |    |    |    |  |          |
| 4b                    |                 |    |    |    | ✓  |  |          |
| 2 $\ell$ on-Z         | ✓               |    | ✓  | ✓  |    |  |          |
| 2 $\ell$ non-resonant |                 |    |    |    |    |  | ✓        |
| 2 $\ell$ soft         | ✓               |    |    |    |    |  | ✓        |
| > 3 $\ell$            | ✓               | ✓  | ✓  | ✓  | ✓  |  |          |
| Hadr. WX              | ✓               | ✓  |    |    |    |  |          |

Table A.1: Summary of all experimental searches considered in the combination and the signal topologies for which each search is used in the combined results.

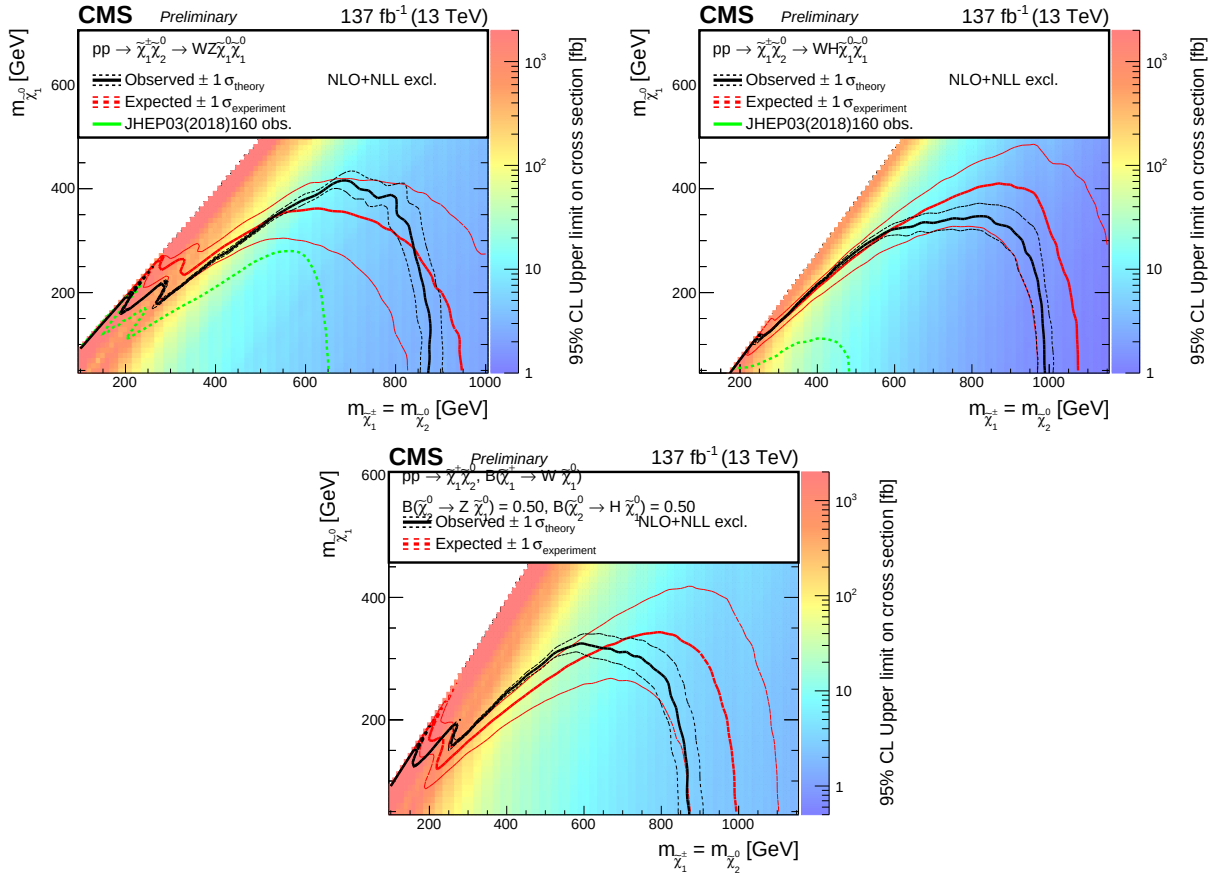


Figure A.3: Results for the neutralino-chargino production for the WZ topology (upper left), the WH topology (upper right), and the mixed topology with 50% branching fraction to each of WZ and WH (bottom).

ure A.3. Figure A.4 shows the analysis with the best exclusion limit at each point in the plane.

The contours for the individual analyses, as well as the combined 95% CL limits, are shown in

Figure A.5.

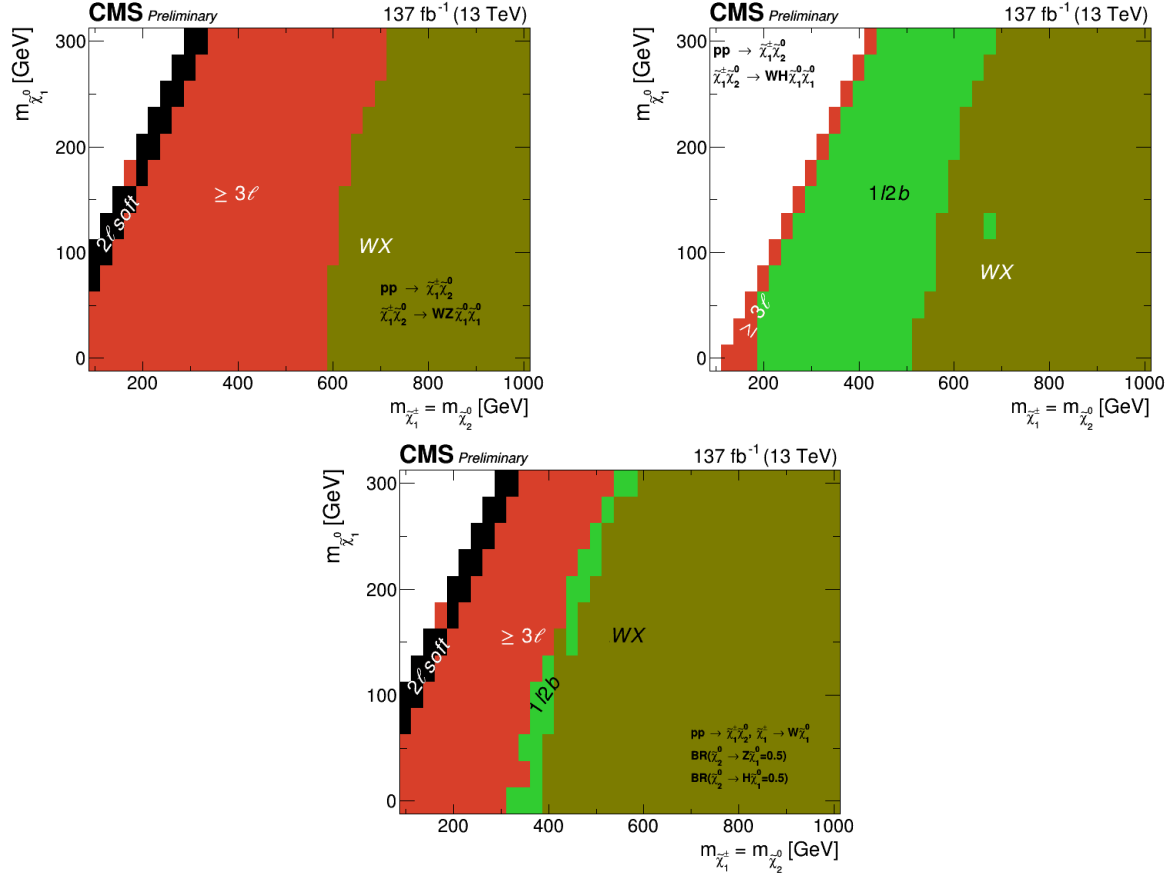


Figure A.4: The analysis with the best exclusion limit at each point in the plane for the WZ topology (upper left), the WH topology (upper right), and the mixed topology with 50% branching fraction to WZ and WH (bottom).

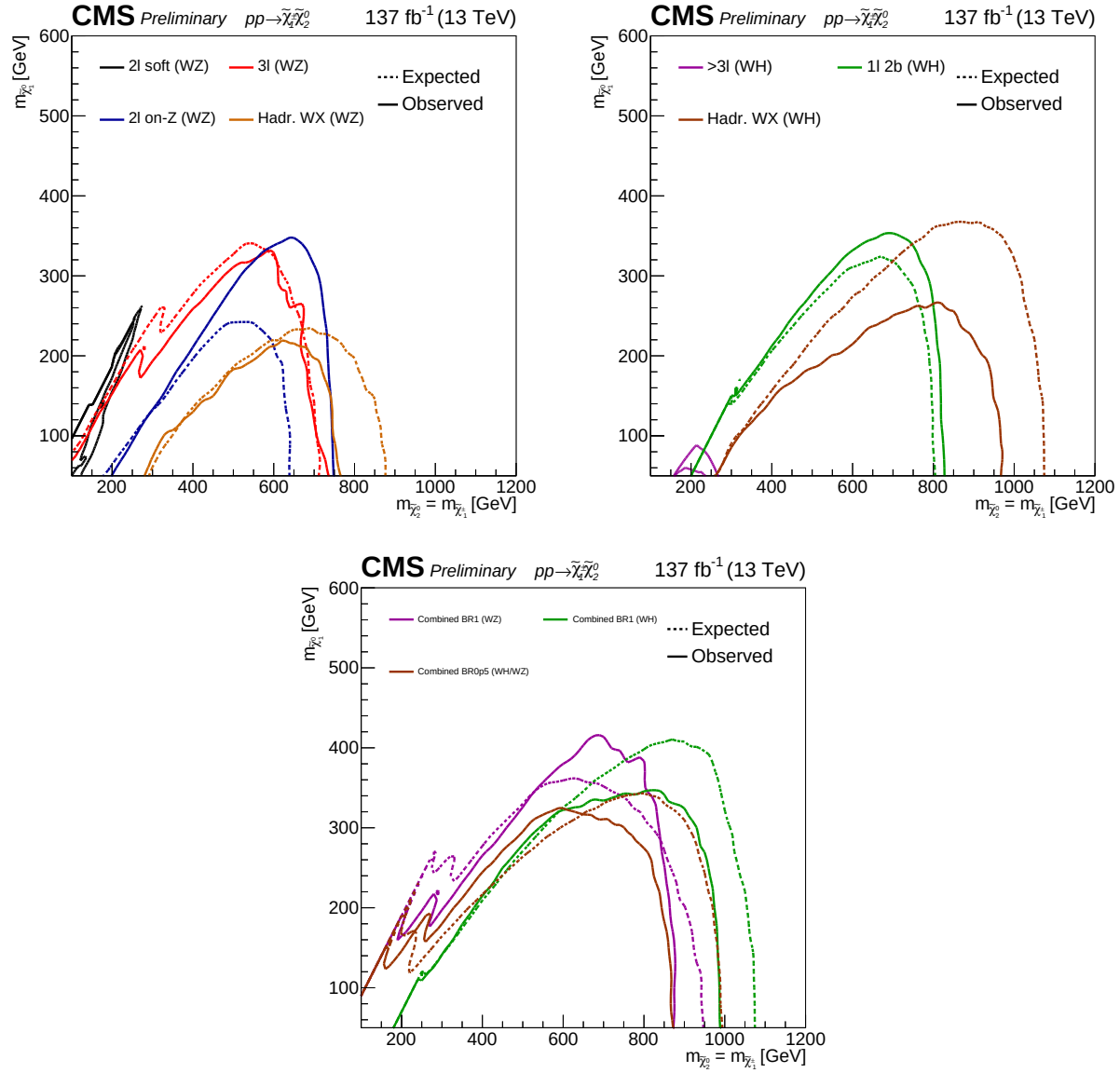


Figure A.5: Exclusion contours at 95% CL for the individual analyses targeting the WZ (upper left) and WH (upper right) topologies. The combined contours for these two topologies, as well as the mixed topology (bottom), are also shown. The decay modes assumed for each contour are given in the legends.

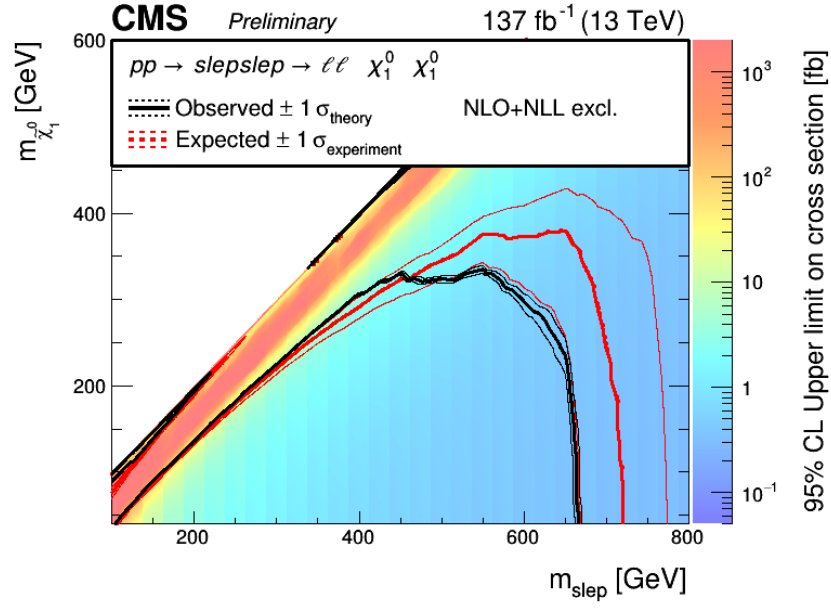


Figure A.6: Observed and expected exclusion limits from direct slepton pair production.

The results of combining the slepton pair production searches of the “ $2\ell$ ” and “ $2\ell$  soft” analyses are shown in Fig. A.6, with an observed (expected) exclusion in the slepton mass of about 665 GeV.