Inclusive searches for supersymmetry at $\sqrt{s} = 13$ TeV using razor kinematic variables, and data scouting using the CMS trigger system

Thesis by
Dustin James Anderson

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2018
Defended 27 April 2018

Dustin James Anderson
ORCID: 0000-0001-8173-3182

ACKNOWLEDGEMENTS

First, I thank my advisor Maria Spiropulu, who oversaw my journey from particle physics beginner to CERN veteran. Her vision for the Caltech CMS group, and her nonstop efforts to realize it, made the work documented in this thesis possible.

Second, I thank the others who helped make my Ph.D. experience meaningful and worthwhile. I am deeply grateful to the following people:

- Cristián Peña, for being a great friend and role model, both at work and outside it.
- Javier Duarte, for helping me learn the ropes of HEP, for getting me interested in statistics, and for trying to get me interested in soccer.
- Yi Chen, for inspiring me in my first year by pulling off amazing feats and then kindly explaining them.
- My fellow grad students at CERN: Thong Nguyen, Jay Lawhorn, Zhicai Zhang, and Jiajing Mao. You made life there much more colorful. Special thanks to Thong for our quality Slack discussions.
- Adi Bornheim, for being a source of wisdom, sanity, and warmth, and for occasional help with car trouble.
- Si Xie, for masterminding our search program and for helping me move in the right direction in the face of uncertainty.
- Maurizio Pierini, for being my mentor at CERN, and for showing me how to navigate a very confusing world and make a real impact on it.
- Jean-Roch Vlimant, for always having bigger ideas than me, bravely pursuing them, and often convincing me to follow.
- Others in the group who took the time to help me when I needed it: Emanuele Di Marco, Josh Bendavid, Artur Apresyan, and Dorian Kcira. Thanks also to Harvey Newman for many useful discussions.
- The Friday night dinner and D&D group: Tony Bartolotta, Jason Pollack, Kevin Barkett, Jeremy Brouillet, Jon Blackman, and Dan Brooks. You've all been amazing friends and have made the Ph.D. process a lot less lonely.

- Shao Min Tan; I'm glad we were there to support each other at CERN.
- Tristan McKinney and Belinda Pang (and Cristian again), for the best lunch conversations anyone could ask for.
- My parents Ann and Ken Anderson. Sorry that I'm always thousands of miles away, and thank you for everything.
- My brothers Adam and Alex, and my sister Kelly: you're the best siblings I've ever had. Keep it up.
- My girlfriend Jiejing, for helping me find much-needed balance these past several months. Your wonderful optimism and perspective have made such a difference.

ABSTRACT

We present two searches for supersymmetric particles using proton-proton collision data collected by the CMS experiment at $\sqrt{s} = 13$ TeV. The searches use razor kinematic variables for signal discrimination and target the pair production of heavy gluinos and squarks in R-parity conserving supersymmetry. The first search is performed on 2.3 fb^{-1} of data collected by CMS in 2015. Two complete, independent background predictions are made: one based on fits using a parameterized functional form, and the other based on Monte Carlo simulation corrected via control samples in data. The second search is an expanded version of the first search, and is performed using the Monte Carlo-based background prediction method on 35.9 fb^{-1} of data collected in 2016. Both searches obtain results compatible with standard model background expectations. The null results are interpreted as limits on the masses and cross sections of gluinos, squarks, and higgsinos in the context of simplified models of supersymmetry. We discuss the outlook for the fit-based search strategy and explore how the technique of gaussian process regression may be useful as a tool to combat the challenges of this analysis methodology.

We also describe a new paradigm for trigger-level collider data analysis, which we refer to as data scouting. In this paradigm, searches for new physics are performed using event information reconstructed within the experiment's trigger software. This circumvents traditional event rate constraints, such as disk space and the latency of offline reconstruction. We provide details on the implementation of a framework for data scouting in the CMS High-Level Trigger system and its successful use in Run II of the LHC. We discuss the impact of scouting on the physics program of CMS and demonstrate that it enables searches for new physics that would not otherwise be possible due to trigger constraints, such as hadronic resonance searches at low mass and searches for leptonic decays of dark photons.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Dustin Anderson. "Data Scouting in CMS". In: *PoS* ICHEP2016 (2016), p. 190.
D.A. designed and implemented the data scouting framework described here with input from CMS trigger coordination and guidance from the Physics Performance and Dataset group.

[2] *Searches for dijet resonances in pp collisions at* $\sqrt{s}$ = 13 TeV *using data collected in 2016.* Tech. rep. CMS-PAS-EXO-16-056. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256873.
D.A. developed data processing and analysis code for this search, and defined the trigger strategy for event selection and efficiency measurements in the low-mass search.

[3] Vardan Khachatryan et al. "Inclusive search for supersymmetry using razor variables in pp collisions at sqrt(s) = 13 TeV". In: *Phys. Rev.* D95.1 (2017), p. 012003. DOI: 10.1103/PhysRevD.95.012003. arXiv: 1609.07658 [hep-ex].
D.A. was one of the leading contributors to this search. He carried out the background prediction using the Monte Carlo-based strategy, contributed significantly to the fit-based background prediction, and performed the statistical interpretation of the search.

[4] D. Anderson et al. "On timing properties of LYSO-based calorimeters". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 794 (2015), pp. 7–14. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2015.04.013. URL: http://www.sciencedirect.com/science/article/pii/S0168900215004829.
D.A. participated in the test beam experiments to characterize LYSO crystal calorimeters, and assisted in the analysis of the data.

[5] Farrell, Steven et al. "The HEP.TrkX Project: deep neural networks for HL-LHC online and offline tracking". In: *EPJ Web Conf.* 150 (2017), p. 00003. DOI: 10.1051/epjconf/201715000003. URL: https://doi.org/10.1051/epjconf/201715000003.
D.A. designed, implemented, and tested the neural network models for track parameter prediction, with input from others in the HEP.TrkX collaboration.

[6] D. Anderson, J.-R. Vlimant, and M. Spiropulu. "An MPI-Based Python Framework for Distributed Training with Keras". In: *ArXiv e-prints* (Dec. 2017). arXiv: 1712.05878 [cs.DC].
D.A. wrote the *mpi_learn* library described in this paper, and assisted in the benchmarking of the software on the Cooley supercomputing cluster.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Part I:

CMS and the Search for Supersymmetry

> All my means are sane, my motive and my object mad.
>
> Captain Ahab, in Herman Melville, *Moby-Dick*

Chapter 1

INTRODUCTION

The standard model (SM) of particle physics encodes much of humanity's understanding of how the universe operates at the submicroscopic level. It is formulated as a relativistic quantum field theory that specifies the set of fundamental particle types in the universe, their properties such as mass and spin, and their interactions via the electromagnetic, weak nuclear, and strong nuclear forces. From the SM framework one can calculate predictions for particle decays and lifetimes, scattering cross sections, resonance masses, and other quantities observable in physics experiments.

A major effort in the particle physics community for the past decades has been to measure physical phenomena with high precision and compare the measurements with the SM's predictions to attempt to discover the limits of the theory's applicability. One component of this effort has been the construction of increasingly powerful particle accelerators that can probe the interactions of particles at higher and higher energies. The SM's numerical predictions have stood up to scrutiny at energies spanning many orders of magnitude [1]. The observation by the Compact Muon Solenoid (CMS) [2] and ATLAS [3] collaborations of a resonance compatible with the Higgs boson, postulated to be related to the mechanism that generates the masses of the fermions and vector bosons, was a recent striking success for the SM.

The searches described in this thesis are motivated by extensions of the SM that incorporate supersymmetry (SUSY), a hypothetical symmetry of nature that would manifest itself in the existence of new heavy particles not contained in the current SM. The existence of SUSY would address theoretical concerns about the SM's plausibility as a model of reality. In particular, it may resolve what is known as the *hierarchy problem*, which is the observation that quantum corrections to the Higgs boson mass ought to push its value much higher than what we measure it to be. SUSY also provides a particle candidate for the dark matter (DM) that appears to exist throughout the universe and interact weakly or not at all with ordinary matter. A SM with SUSY may also exhibit unification of the electromagnetic, strong, and weak nuclear force coupling strengths at high energy.

In its first run, the Large Hadron Collider (LHC) delivered proton-proton collisions at center-of-mass energies of 7 and 8 TeV. Collisions recorded by the CMS and ATLAS detectors were analyzed and used to search for evidence of SUSY. No significant deviations from the SM predictions were found, and the searches statistically excluded some regions of the SUSY parameter space [4].

In this thesis, we present searches for SUSY in CMS data collected in the second run of the LHC, at the machine's current maximum energy of 13 TeV. The searches are carried out using kinematic quantities known as razor variables that are designed to facilitate discovery and characterization of SUSY. The increase in energy from 8 to 13 TeV makes it possible to search for SUSY particles in a wider mass range and in a larger kinematic phase space than in LHC Run I. Furthermore, the higher luminosity achieved by the LHC in 2016 yielded a dataset roughly twice as large as that collected in Run I. Thus, the searches presented here have significantly higher sensitivity to a variety of SUSY processes than those carried out at 7 and 8 TeV.

The increased LHC energy and luminosity present challenges related to data processing that constrain the types and sensitivities of searches that can be conducted. The CMS trigger system is limited in the volume of event data that it can record, and further constraints are imposed by the capacity of the prompt event reconstruction system. We present an alternative paradigm for collider data analysis, known as data scouting, that can alleviate these issues by exploiting the reconstruction algorithms available at the trigger level. Data scouting has become a mainstay of the CMS data collection framework, and it is receiving increasing attention from analysts as a way to perform searches that would otherwise be impossible due to trigger constraints.

This thesis is organized into four parts. Part I provides background on theoretical and experimental aspects of SUSY searches. In Chapter 2, we give a brief description of the LHC, the CMS detector, and the associated trigger and data processing systems. In Chapter 3, we give an overview of supersymmetry and summarize the previous searches carried out at the LHC.

Part II describes our implementation in CMS of data scouting, a framework for trigger-based data analysis. In Chapter 4 we give details of the scouting trigger algorithms deployed between 2015 and 2017. In Chapter 5 we discuss the past and expected future physics impact of data scouting.

In Part III we present the results of the searches for SUSY conducted on 2015

and 2016 CMS data using razor variables. We present two methods of SM background prediction, one based on fits to a smooth function shape, the other based on Monte Carlo simulation (MC), and describe their evolution over two years of CMS data collection. Chapters 6 and 7 describe respectively the fit-based and MC-based searches carried out on the 2015 CMS dataset. These methods provide predictions for the SM background under two completely different sets of statistical assumptions; this dual approach is unique among CMS SUSY searches. Chapter 8 describes the MC-based search conducted on the 2016 dataset, with modifications appropriate to a dataset an order of magnitude larger than in 2015. Finally, Chapter 9 describes the shortcomings of the fit-based background prediction when applied to the 2016 dataset and discusses in some detail how one might overcome them using new statistical techniques.

In Part IV we conclude by summarizing the current state of SUSY in light of the analyses that have been performed, and describe the outlook for SUSY searches during the rest of LHC Run II and beyond. We also provide three appendices describing other work carried out concurrently with this research. Appendix A describes experimental studies of the timing properties of scintillating crystals and photodetectors. This effort is motivated by the upcoming high-luminosity upgrade of the LHC, where time-of-flight information may be used to mitigate the effects of pileup in pp collisions. Appendix B describes the development of neural network-based algorithms for particle tracking. These early-stage studies form part of a research program aiming to use modern machine learning techniques to improve on the tracking algorithms currently used at the LHC. Finally, Appendix C describes a software framework written to facilitate distributed training of machine learning models in a supercomputing cluster setting. We demonstrate the performance of the framework and discuss its role in the research and development of new algorithms for high energy physics.

Chapter 2

PHYSICS WITH THE CMS DETECTOR

When protons in the LHC collide within the CMS detector, high-energy particles are produced that carry information about the underlying collision event. To convert this information into a form suitable for analysis, the CMS Collaboration deploys an ensemble of detector hardware, computing infrastructure, and data processing software. In this chapter we describe the elements of this ensemble, beginning with the LHC itself and ending with the reconstructed event data that serves as the starting point for all CMS physics measurements and searches.

2.1 Proton-proton collisions at the LHC

The LHC is a circular proton-proton collider, operated by CERN and located near Geneva, Switzerland. It delivers *pp* collisions at center-of-mass energies ranging up to 13 TeV at an average rate as large as 32 MHz. The main purpose of the LHC is to provide the data needed to probe the mechanism of electroweak symmetry breaking and to search for evidence of particles and interactions not present in the SM.

The LHC occupies the 27-kilometer tunnel that previously housed the LEP collider. A sequence of machines in the CERN accelerator complex, shown schematically in Figure 2.1, brings protons up to the full beam energy of 6.5 TeV in several stages. The protons originate from hydrogen gas, which is ionized using an electric field. They then pass through the Linac 2 accelerator, the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS), which accelerate them to 50 MeV, 1.4 GeV, 25 GeV, and 450 GeV, respectively. Finally, the protons enter the LHC itself, where their energies are increased to 6.5 TeV before they are brought into collision. The two LHC beams circulate in opposite directions and are made to collide at four locations in the LHC ring. The CMS, ATLAS, LHCb, and ALICE particle detectors are located at these points. A detailed description of the LHC design can be found in [5].

LHC operations from 2009 to 2012 are referred to as 'Run I' of the machine. Run I was conducted at collision energies of up to 8 TeV and provided the data used by CMS and ATLAS to establish the existence of a Higgs-like boson. In 2015, after a planned two-year shutdown, the LHC began 'Run II.' The collision energy was

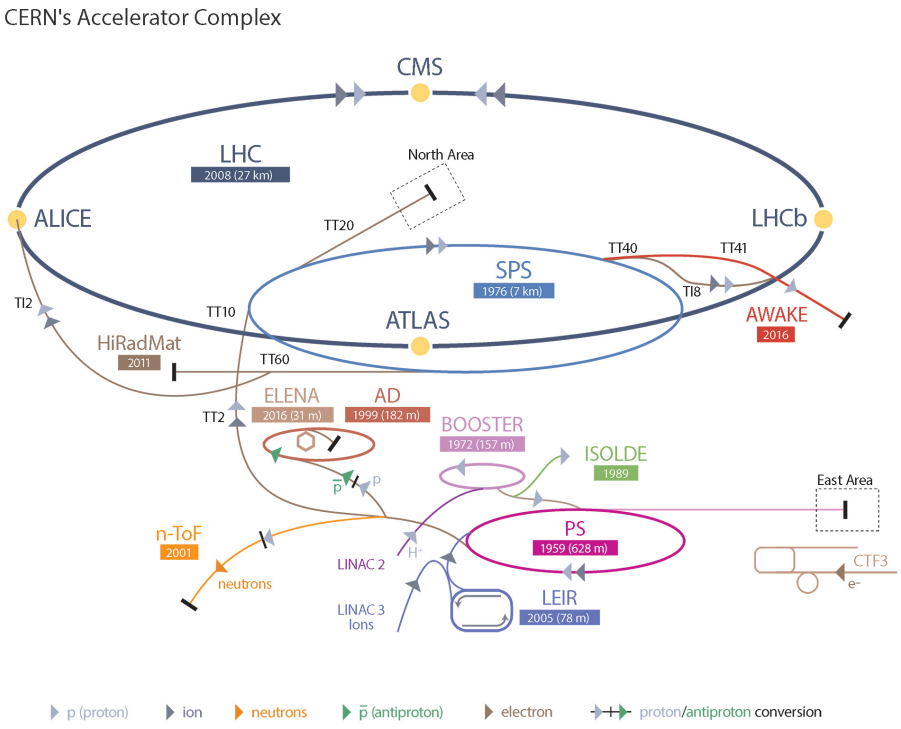


Figure 2.1: The CERN accelerator complex [6].

increased to 13 TeV, the maximum achieved to date. The minimum time interval between collision events was halved from 50ns to 25ns early in the 2015 data-taking period.

Pileup collisions

An aspect of the LHC particularly relevant for CMS data analysis is that each beam is divided into discrete *bunches*, each consisting of up to 1.15×10^{11} protons. The beams may each contain up to 2808 bunches. In each LHC collision event, two bunches pass through each other, producing some number of individual pp interactions. The average number of collisions per bunch crossing is

$$N_{avg} = \frac{\mathcal{L}\sigma_{inel}}{f}, \tag{2.1}$$

where $\mathcal{L}$ is the instantaneous luminosity, σ_{inel} is the inelastic pp scattering cross section, and f is the frequency of bunch crossings. The value of σ_{inel} at 13 TeV has been measured by CMS [7] and ATLAS [8]; the nominal value obtained by CMS is 71.3 mb. At $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}and $f = 31.6$ MHz, this yields a mean of 23

collisions per bunch crossing. In practice, the number of colliding bunches and the instantaneous luminosity vary from run to run.

Each collision event therefore consists of many superposed particle interactions, which must be disentangled from one another during analysis. In general there will be at most one interaction of interest per event. The others, referred to as *pileup* interactions, contribute a diffuse background of low transverse momentum particles to the event. The presence of pileup poses challenges for event reconstruction and analysis because it adds additional spurious particles to every event, adding noise and slowing down the event reconstruction process.

LHC performance in Run II

In 2015, the first year of Run II, the LHC delivered collisions to CMS at instantaneous luminosities ranging up to 5×10^{33} cm^{-2}s^{-1} [9]. The total integrated luminosity, defined as

$$\mathcal{L}_{int} = \int \mathcal{L}(t)dt, \tag{2.2}$$

was 4.2 fb^{-1}, of which 2.3 fb^{-1} was recorded by CMS and certified good for physics analysis. The average number of pileup collisions per event was 14.

In 2016, the LHC reached its design luminosity of 10^{34} cm^{-2}s^{-1} and ran at that level or higher for a large fraction of the year. The luminosity delivered to CMS that year was 40.8 fb^{-1}, an order of magnitude larger than that delivered in 2015. Of this, 35.9 fb^{-1} was certified good for physics analysis. The increase in luminosity brought an increase in the number of pileup collisions, with an estimated 27 pp interactions per bunch crossing.

As of this writing, CMS has recently ended its 2017 pp run. The LHC delivered 51.0 fb^{-1} to CMS, of which 41.8 fb^{-1} were certified good for physics. Analysis of this dataset is still in progress by the collaboration.

A chart illustrating the increase of the integrated luminosity over each year's run period is shown in Figure 2.2. Values for both the Run I and Run II data-taking periods are shown.

2.2 Description of the CMS detector

The CMS detector is one of two general-purpose particle detectors associated with the LHC (the other is ATLAS). A detailed description of CMS can be found else-

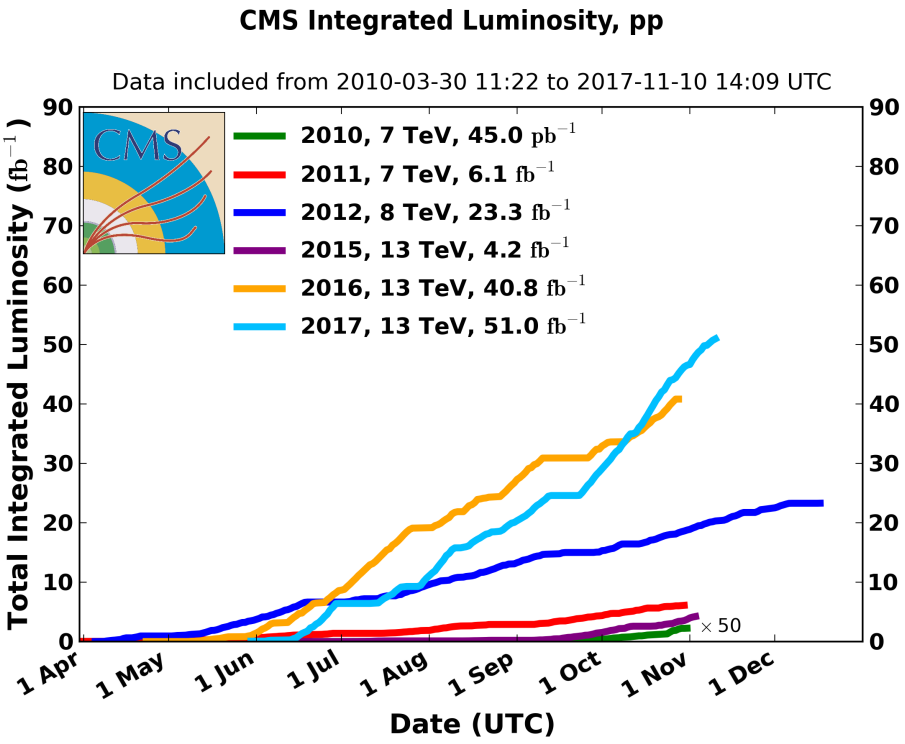


Figure 2.2: Integrated luminosity delivered to CMS in each year of LHC running [9].

where [10], but it will be useful to briefly describe each component of the detector here.

CMS has a cylindrical structure, with the LHC beam pipe passing along its axis. Its central feature is a 3.8T superconducting solenoid, inside which are subdetectors dedicated to measuring specific types of particles. The magnetic field from the solenoid causes the paths of charged particles to bend, allowing accurate estimation of their momenta during reconstruction. Each major subdetector consists of a cylindrical 'barrel' component and two 'endcaps' that are transverse to the beam, giving nearly hermetic enclosure of the collision region. A diagram of CMS is shown in Figure 2.3. In Figure 2.4 is a schematic illustrating how each particle type interacts with the detector.

CMS coordinate system

A cylindrical coordinate system is used to describe locations in the CMS detector. The origin of the coordinates is at the center of the detector, with the z-axis pointing along the beam and the x-axis pointing radially inward toward the center of the

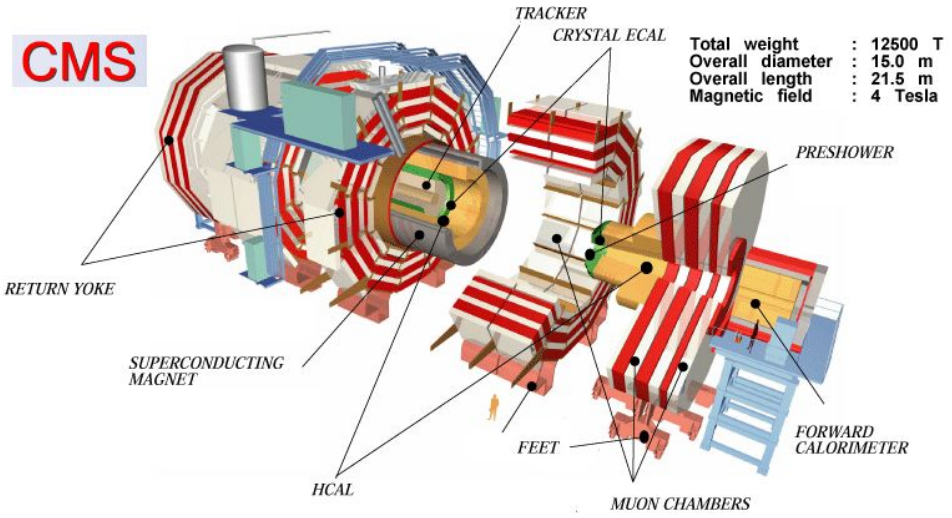


Figure 2.3: A diagram of the CMS detector [11].

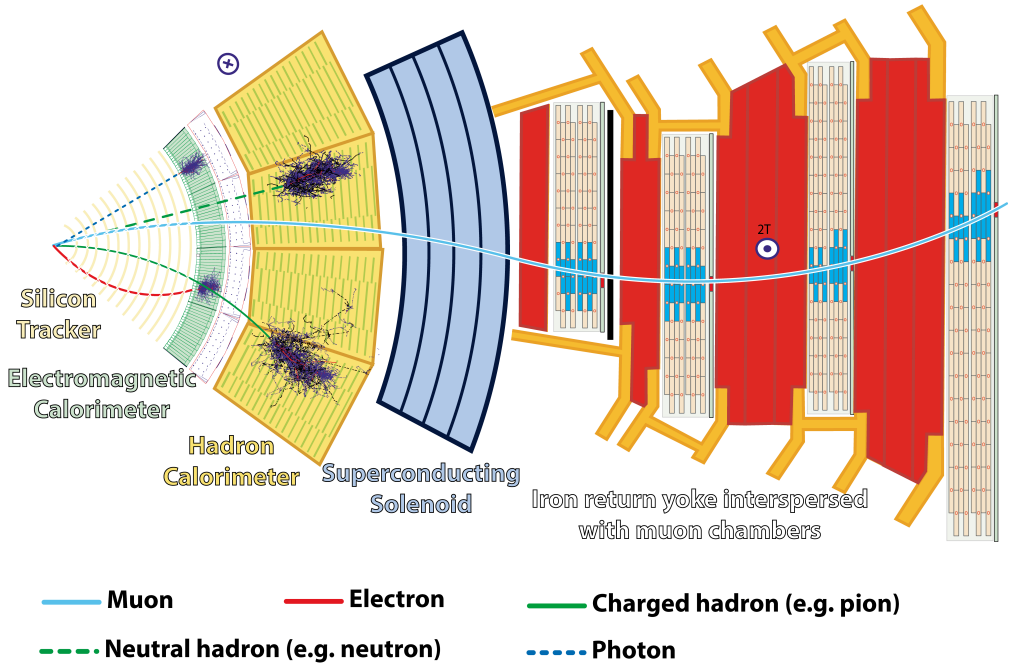


Figure 2.4: A diagram showing a slice of the CMS detector, illustrating how various particle types interact with the detector. Photons (dashed blue line) pass through the tracker and deposit their energy in the ECAL. Electrons (red line) leave tracks in the tracker before depositing their energy in the ECAL. Hadrons (green lines) create showers in the HCAL, leaving tracks in the tracker if charged. Muons (solid blue line) pass through all detector layers, leaving tracks in the inner tracker and the muon system [12].

LHC. The three coordinates used for spatial locations are the radial distance ρ =

$\sqrt{x^2 + y^2}$, the azimuthal angle ϕ, and the pseudorapidity η, which is defined as

$$\eta = -\log\left(\tan(\theta/2)\right), \tag{2.3}$$

where θ is the polar angle made with the z-axis. The choice of η as a coordinate stems from the fact that the η interval between two points is roughly invariant under Lorentz boosts along the beam direction. The compositeness of the proton implies that the total z-momentum in any given pp collision is unknown (the colliding partons carry an indeterminate fraction of the protons' energies), which makes it convenient to parameterize events in a z-boost-invariant way.

Particle four-momenta are usually specified by their component p_T in the transverse plane, their directional coordinates η and ϕ, and their energy E.

The inner tracker

The subdetector closest to the collision point is the silicon tracking system, which is designed to record the passage of charged particles and provide precise information about their directions and momenta. The tracker has an outer radius of nearly 110 cm and covers the range $|\eta| < 2.5$.

The inner part of the tracker is made of silicon pixels each measuring 100×150 μm^2. There are three layers of pixels in the barrel and two in the endcaps; in total the tracker contains 66 million pixels. The average occupancy per pixel in one LHC bunch crossing is approximately 10^{-4}, low enough to allow reconstruction of tracks and vertices in the LHC's high-pileup environment.

Outside the pixel tracker is the strip tracker, which consists of 9.6 million silicon strip sensors. The strip tracker is organized into four regions (Inner Barrel, Outer Barrel, Endcap, and Inner Disks) with sensors of varying sizes appropriate to the particle flux at each location.

The spatial hit resolution of the pixel tracker is measured to be 10-20 μm, and that of the strip tracker is measured to be 23-52 μm. The layout of the silicon tracker is illustrated in Figure 2.5.

The electromagnetic calorimeter

Surrounding the tracker is the electromagnetic calorimeter (ECAL), which is made of monolithic crystals of lead tungstate ($PbWO_4$). The ECAL barrel contains 61200

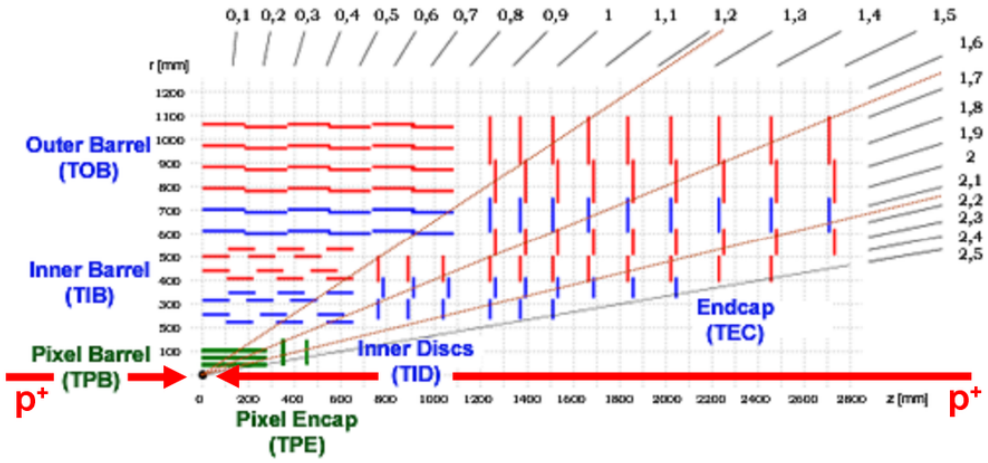


Figure 2.5: Diagram of the CMS inner tracker. The collision point is shown in the lower left. The pixel tracker is shown in green, and the strip tracker is shown in red and blue [13].

crystals and covers $|\eta| < 1.479$; the two endcaps each contain 7324 crystals and cover $1.479 < |\eta| < 3.0$. The crystals in the barrel and in the endcap are coupled respectively to silicon avalanche photodiodes (APDs) and vacuum phototriodes (VPTs) for readout. In front of each ECAL endcap is a preshower device, consisting of two planes of silicon strip detectors placed behind lead absorbers. A diagram of the ECAL is shown in Figure 2.6.

When electrons and photons pass through the ECAL, their energy is converted into scintillation light. The $PbWO_4$ crystals in the barrel and in the endcap have depths corresponding to 25.8 and 24.7 radiation lengths ($X_0 = 0.89$ cm) respectively. The width of each crystal is comparable to the Moliere radius of $PbWO_4$ (2.2 cm), so the electromagnetic shower is contained within a relatively small number of crystals. The scintillation light produces APD/VPT pulses that carry information about the energy of the incident particles.

Large radiation doses cause the formation of color centers in the ECAL crystals that reduce their transparency to scintillation light. This effect leads to a change in each crystal's response over time; the response decreases as radiation is absorbed, and slowly increases when no collisions are occurring due to a natural annealing process in the crystals [15]. A monitoring system is used to track the transparency of each crystal throughout each run period. This is accomplished by firing laser light into the crystals and measuring their response during gaps in the LHC bunch train, which are spaced 90 μs apart. The response change over the course of a

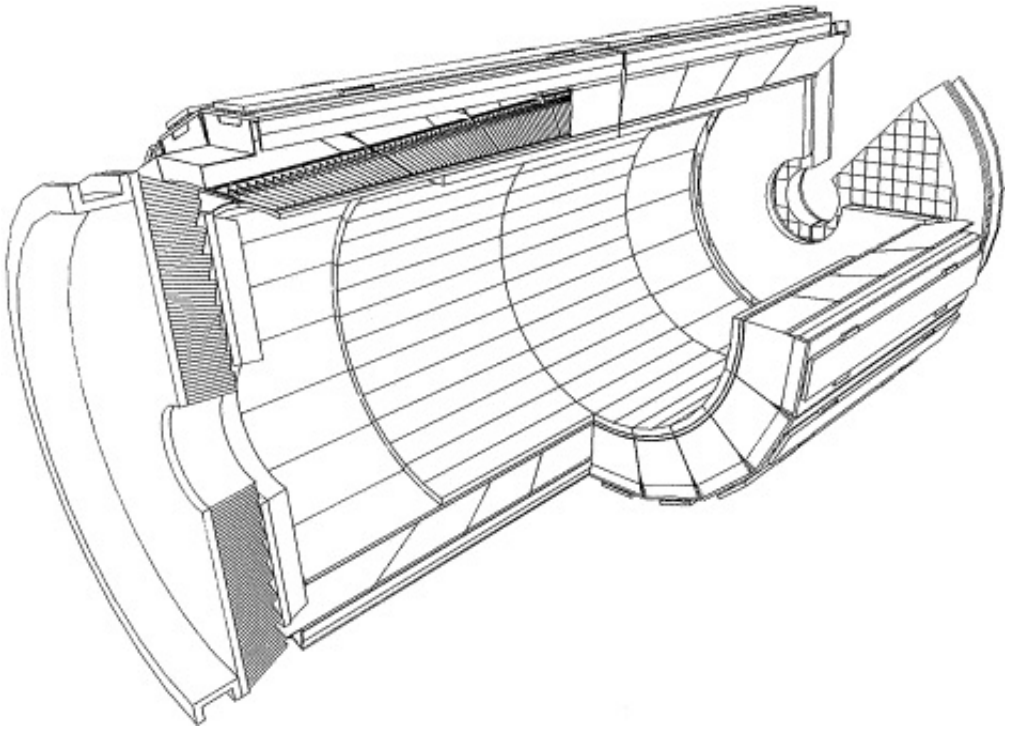


Figure 2.6: Cutaway view of the CMS ECAL [14].

year is dramatic (see Figure 2.7) and this continuous monitoring and calibration is needed to guarantee robust measurements of photon and electron energies. The laser monitoring system is designed to measure the transparency of every crystal once every 30 minutes with a precision of 0.1%.

The hadron calorimeter

Between the ECAL and the CMS solenoid is the hadron calorimeter (HCAL), which is made of alternating layers of brass absorber and plastic scintillator tiles. The purposes of the HCAL are to measure the energies of incident hadrons and to provide fully hermetic coverage for accurate measurement of the total transverse momentum in each event. The volume devoted to the scintillator tiles is made as small as possible in order to maximize the amount of absorber: the scintillator tiles are 3.7 mm thick and are sandwiched between 5 cm brass plates. The plastic tiles are read out via wavelength-shifting (WLS) fibers, which are spliced to clear fibers that channel the light to photodetectors.

The barrel of the HCAL consists of 2304 'towers' covering the range $|\eta| < 1.4$. In the region $|\eta| < 1.26$ the calorimetry is complemented by additional scintillator tiles located outside the CMS solenoid; this 'hadron outer' (HO) calorimetry

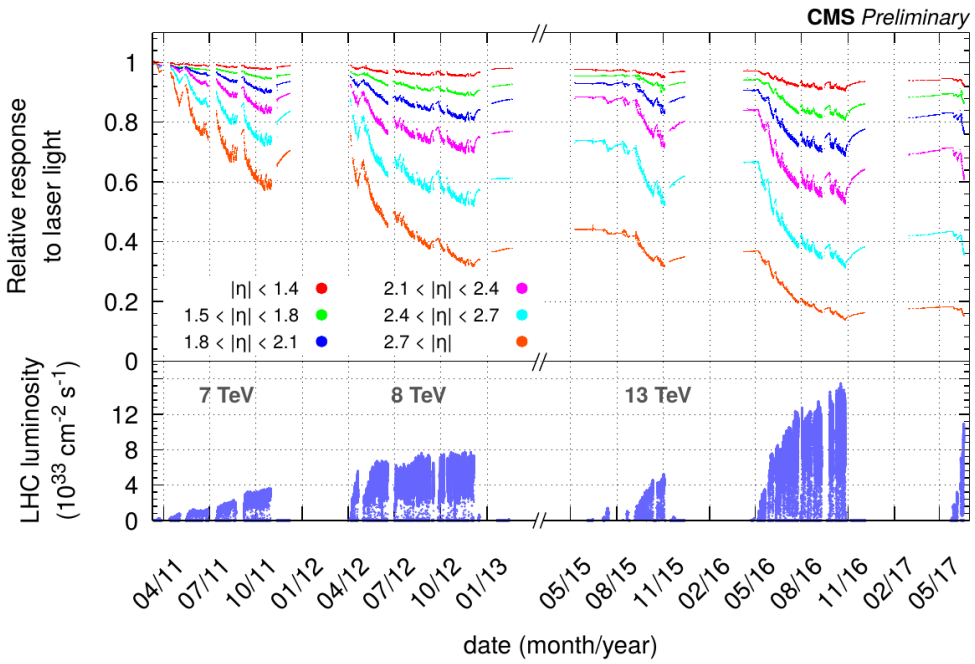


Figure 2.7: Change in response of the CMS ECAL crystals over time. The transparency is shown for crystals at different values of η [16].

increases the instrumented region in the barrel to cover more than 10 nuclear interaction lengths. The HCAL endcaps, comprising 2304 detector towers, cover the range $1.3 < |\eta| < 3.0$. Finally, a forward hadron calorimeter (HF) extends the coverage out to $|\eta| = 5.0$. The HF consists of an ensemble of grooved steel plates into which quartz fibers are inserted. Cherenkov light emitted in the fibers is sent to photomultipliers, which measure the energy in a total of 1800 readout channels. The components of the HCAL are indicated in Figure 2.8.

The muon system

High-energy muons are generally not stopped by the CMS tracker or calorimeters, and dedicated detectors are placed outside the solenoid to identify them. The muon system contains three different types of gaseous detectors: drift tube chambers (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs). DTs cover the barrel region out to $|\eta| = 1.2$; they are deployed in four layers, interleaved with the iron return yoke of the magnet. CSCs form the endcap of the muon system and cover the region $0.9 < |\eta| < 2.4$. RPCs are coupled to each DT in the barrel region, and additional RPCs are placed in the endcap out to $|\eta| = 1.6$. RPCs have coarser position resolution but finer time resolution than DTs and CSCs; they

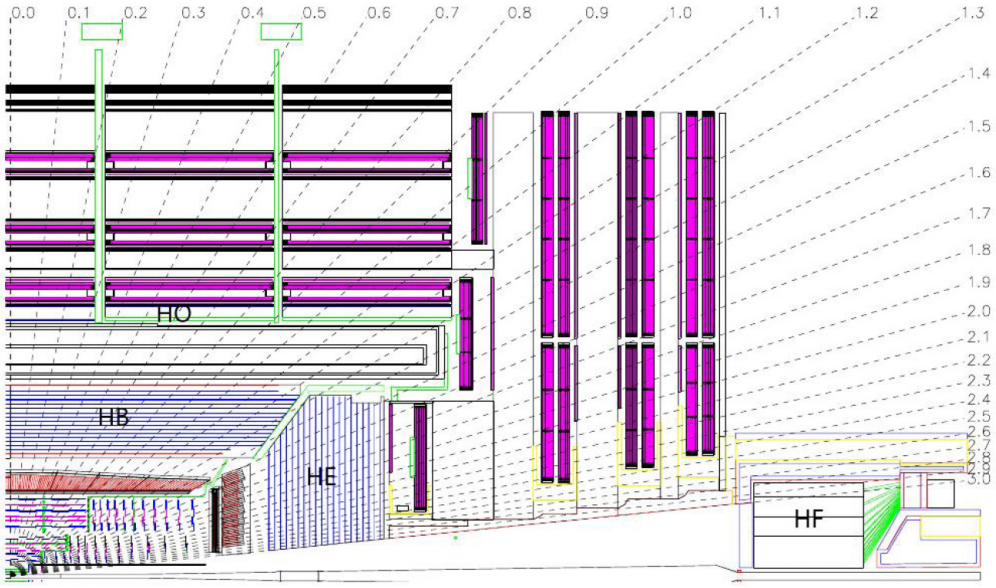


Figure 2.8: Diagram of the CMS detector layout, with the barrel (HB), endcap (HE), outer (HO), and forward (HF) components of the HCAL labeled. The purple stripes in the outer regions of the detector indicate the placement of the muon chambers [17].

are useful in ensuring that muon hits are assigned to the correct bunch crossing. The layout of the muon system is indicated in Figure 2.8.

For muons with p_T lower than 200-300 GeV, scattering in the inner detector material limits the momentum resolution of the muon system. Thus, measurement of muon momentum is performed mainly by the inner tracker. For very high-p_T muons, however, the effect of scattering is small and the momentum measurement benefits from the incorporation of muon chamber hit information.

2.3 The trigger system

The rate of collisions at the LHC is far too high for every collision event to be saved and reconstructed. Instead, CMS uses a trigger system to select events of interest [18]. The trigger system operates in real time during data-taking and must satisfy strict constraints on event processing time.

The CMS trigger system is divided into two stages, the Level-1 (L1) Trigger and the High-Level Trigger (HLT). Each trigger stage identifies events that are likely to be useful for physics analysis. Selected events are passed to the next stage of processing, while unselected events are discarded without being saved to disk. Starting from the full LHC collision rate (~10^9 Hz) the L1 Trigger system selects approx-

imately 100 kHz of events. The HLT further reduces this to ~1 kHz of physics events, which are saved to disk for offline processing.

The L1 Trigger

The L1 Trigger is implemented in hardware using custom ASIC chips and FPGA cards. Within a latency of 4 μs, it must receive the event data from the detector, perform basic reconstruction of physics objects, and make a trigger decision. Due to the limits of the CMS readout electronics, the L1 output event rate must be no more than 100 kHz.

A diagram of the L1 event processing logic is shown in Figure 2.9. Physics object reconstruction is performed using data from the calorimeters (ECAL and HCAL) and from the muon chambers (DTs, CSCs, and RPCs). Calorimeter data is aggregated by a two-stage system consisting of a Regional Calorimeter Trigger (RCT) and a Global Calorimeter Trigger (GCT). The output of the GCT is a set of candidate electron, photon, and jet objects. Data from the muon chambers is processed by the Global Muon Trigger (GMT), which outputs a set of muon candidates.

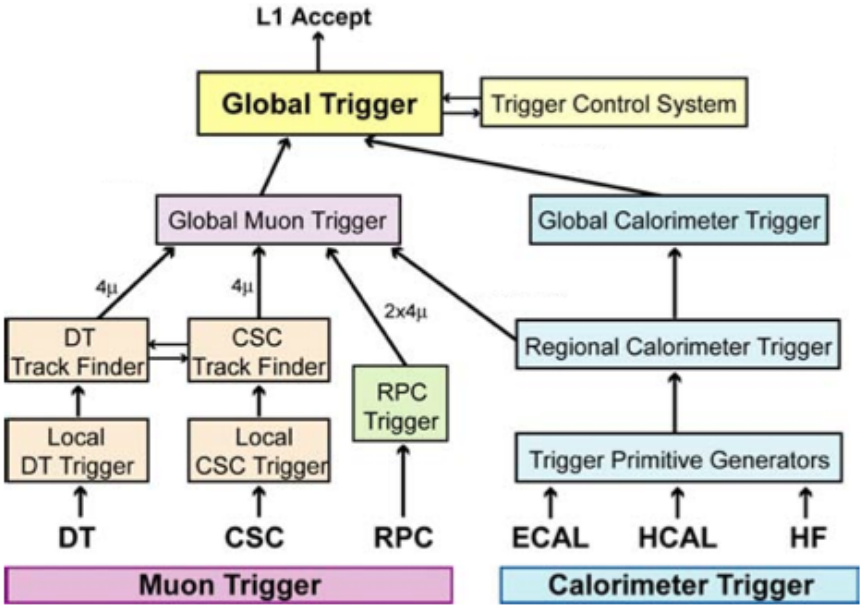


Figure 2.9: Diagram of the L1 trigger [19].

The final trigger decision is made by the L1 Global Trigger (GT), which receives as input the physics objects reconstructed by the GCT and the GMT. The GT consists

of a number of separate trigger paths, which compare the multiplicity and momenta of reconstructed physics objects against a set of predefined selection cuts.

The L1 system was upgraded in 2016 to feature more modern FPGAs and high-bandwidth optical links for communication between trigger cards [20]. Where previously the calorimeter hit data was coarse-grained before being processed by the RCT, the upgraded system allows the calorimeter triggers access to the full ECAL and HCAL granularity. This improvement allows physics objects to be reconstructed at the L1 with higher resolution, yielding more performant triggers that have lower cut thresholds yet still respect the 100 kHz rate limit. The upgrade also significantly increases the power and flexibility of the GT. It relaxes the hard limit of 128 trigger paths that existed previously, and enables more complex trigger logic involving, for example, the invariant masses of reconstructed particles.

The HLT

The HLT is implemented as software running on a processor farm at LHC Point 5. It consists of a few hundred trigger paths, each selecting for a particular physics signature. The paths execute in parallel for each physics event selected by the L1 Trigger, and an event is accepted if it meets the requirements of any path.

HLT paths are divided into modules, each of which is either a 'producer' (which performs some reconstruction task) or a 'filter' (which applies a cut and rejects the event if it does not pass). An HLT path can in principle execute arbitrary code when making its trigger decision, using the raw CMS event data as input. The main constraint is the total compute time needed to process an event. Modified versions of the offline particle reconstruction algorithms described in Section 2.4, including the particle flow algorithm, have been written for use at the HLT. These algorithms yield physics objects similar to those built by the full reconstruction software, but trade accuracy for speed where necessary.

Some HLT paths are *prescaled*, meaning that they are blind to a certain fraction of events. A prescaled trigger is assigned a prescale factor that is an integer $N > 1$, and the trigger only runs on one in every N events entering the HLT. Prescaled triggers are used to perform measurements that do not require the full statistics of the data, in cases where the unprescaled path would have prohibitively high rate.

To keep up with the rate of events selected by the L1 Trigger, the average time to run the full suite of HLT paths should not exceed a certain threshold, which is on

the order of hundreds of milliseconds. The exact CPU time budget varies from year to year as the HLT farm is upgraded with new machines. The distribution of event times is long-tailed: most events can be accepted or rejected very quickly, while a few require more intensive processing before the decision can be made. This long-tailed distribution is shown in Figure 2.10. Also shown is the mean execution time of the menu as a function of the instantaneous luminosity. In case the luminosity is too high for the full menu to run within the time constraint, the HLT can be dynamically switched to a more restricted menu where some paths are prescaled or disabled.

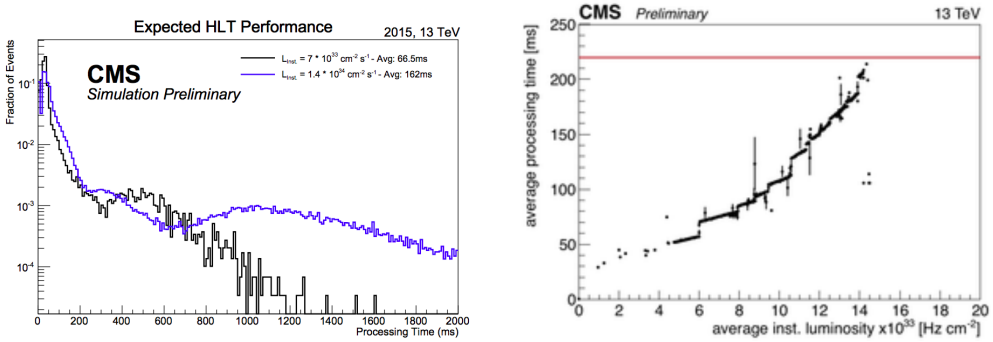


Figure 2.10: Left: distribution of simulated HLT event processing times for two different instantaneous luminosity scenarios at 13 TeV [21]. Right: average HLT event processing time as a function of instantaneous luminosity in 2016 [22]. The red line represents the limit imposed by the HLT farm.

The most time-intensive task performed at the HLT is track and vertex reconstruction. The HLT particle flow algorithm makes use of an iterative track reconstruction algorithm similar to that used offline. Some iterations of the tracking sequence are omitted to save time; this limits the HLT's ability to reconstruct displaced tracks, such as those from hadrons that undergo a nuclear interaction within the tracker material. The combinatorics of track seed extension are also alleviated by retaining fewer track candidates at each layer (in the offline algorithm, up to five candidates are considered at each layer).

A simpler track reconstruction procedure is also available that only considers hits in the pixel tracker, ignoring the strip tracker. These 'pixel tracks' are used as seeds for the full tracking algorithm, and as input to the HLT vertex identification algorithm. Some applications make use of an even faster vertexing algorithm ('fast primary vertex finding') that identifies vertices compatible with selected high-p_T jets by projecting pixel hits along the jet axis [23]. This algorithm reduces the running time of some HLT paths by a factor of 4 to 6.

Streams and primary datasets

HLT paths are organized into *streams*, which are groups of paths sharing a similar purpose and having the same output event content. For example, 'physics' streams contain HLT paths selecting events for offline physics analysis. Events entering these streams are marked for prompt reconstruction, and their full raw data is sent to the CMS Tier-0 computing center for storage and further processing.

Other data streams contain HLT paths used for specialized tasks, such as detector calibration. These tasks may not require the full raw event data, and therefore some parts of the raw data may be discarded for events in these streams in order to save storage space. For example, a dedicated data stream is used for the calibration of the ECAL using π^0 decays to photons [24]. Only ECAL data from crystals near π^0 candidates is needed for this calibration task, so the raw data from all other parts of the detector is discarded for events in this stream. The raw CMS data is formed from the output of ~600 Front-End Driver (FED) boards, each associated with a particular subdetector. Each FED's data is a modular unit that can be individually saved or discarded, which facilitates the process of reducing the event size in these streams.

Streams are further subdivided into *primary datasets* (PDs). All PDs in a stream are processed in the same way and have the same event content; the subdivision is mainly to avoid excessively large file sizes.

2.4 Physics object reconstruction

Events in the main physics streams are immediately sent offline for reconstruction. This is referred to as *prompt reconstruction* and is performed centrally at the CMS Tier-0 computing center located at CERN.

The particle flow algorithm

The central feature of the reconstruction software is the particle flow (PF) algorithm [25], which takes as input the digital readout of all subdetectors within CMS and produces a complete description of the event in terms of 'PF candidates', each meant to correspond to a single particle produced in the collision. A PF candidate is characterized by its 4-momentum, its particle type, and additional information specific to its particle type. Seven types of particles are included in the event description, each described briefly in the following subsections.

In the first stage of the PF algorithm, the data from each subdetector is preprocessed, yielding *PF clusters* in the calorimeters, and tracks in the inner tracker and muon system. These elements are used as building blocks in the particle reconstruction procedure.

Charged particle tracks are reconstructed in the inner tracker via a combinatorial Kalman filter (KF) algorithm [26]. Using as input a collection of seeds reconstructed in the pixel tracker, the KF constructs tracks one detector layer at a time, updating the measured track parameters according to the hit positions in each layer. The entire KF algorithm is iterated several times with different settings. In each iteration, the hits used to make tracks in the previous iterations are masked. Each iteration is configured to target a particular type of track: the first iteration targets high-p_T tracks originating from the collision point, while subsequent iterations are optimized for low-p_T tracks, displaced tracks, and tracks inside high-p_T jets.

In the ECAL and HCAL, cells with recorded energy larger than a seed threshold are considered *cluster seeds*. Clusters are grown iteratively from seeds by adding cells sharing at least a corner with an existing cluster and having energy greater than twice the noise level. An expectation-maximization algorithm is then used to fit a mixture of Gaussian distributions to each cluster, with the number of mixture components equal to the number of cluster seeds. The location and height of each component of the Gaussian mixture are taken to be the position and energy, respectively, of a PF cluster. PF cluster energies are calibrated using simulation to ensure that they faithfully reproduce the energies of the underlying electromagnetic or hadronic showers.

In the ECAL, groups of clusters localized in η and extended in ϕ are combined together into *superclusters*. The superclustering procedure is meant to group together energy deposits originating from bremsstrahlung from an incident particle. In the reconstruction of electrons and photons, each particle candidate is associated with a single supercluster.

After the clustering stage, a *link algorithm* is used to group together PF clusters and tracks in different subdetectors that are consistent with being created by a single particle. For example, an ECAL cluster may be linked to an HCAL cluster that is in the same location in η and ϕ. After the link algorithm runs, a particle type is assigned to each object according to what clusters and tracks are associated with it.

Muons

Muon reconstruction is performed by matching tracks in the inner tracker with hits in the muon chambers. This matching can be performed 'inside out,' i.e. by extrapolating inner tracks to the muon chambers and looking for compatible tracks there; or 'outside in' (vice versa). Muons reconstructed in the former way are called *tracker muons* and those reconstructed in the latter way are called *global muons*. Both types of muon are used in CMS data analysis. Selecting tracker muons that are not also global muons recovers efficiency for muons with $p_T < 10$ GeV, which may not have well-reconstructed tracks in the muon system due to multiple scattering in the magnet yoke.

Hadrons that are not fully stopped in the HCAL may leave hits in the muon system. These *punch-through* hadrons are prevented from being reconstructed as muons by examination of the energies of ECAL and HCAL hits assigned to the muon candidate.

Electrons and photons

Electrons are identified as energy deposits in the ECAL that are compatible with reconstructed tracks in the tracker. Tracks pointing at an ECAL supercluster are refit using a Gaussian sum filter (GSF) algorithm [27] optimized for identifying electron tracks. The algorithm is a nonlinear variant of the KF. Rather than assuming a Gaussian distribution for the particle energy loss at each layer, it uses an approximation to the Bethe-Heitler distribution, constructed using a mixture of Gaussian distributions. Tracks reconstructed with the GSF recover hits due to bremsstrahlung that may be missed by the KF algorithm.

The electron object produced by the PF algorithm consists of a GSF track and its associated ECAL supercluster. Superclusters with no associated track are reconstructed as photons.

Charged and neutral hadrons

Energy deposits in the HCAL generally originate from energetic hadrons. HCAL deposits with associated tracks are reconstructed as charged hadrons (protons, charged pions, and charged kaons), while those without tracks are reconstructed as neutral

hadrons (neutrons and neutral kaons). If an HCAL deposit has an associated track but the momentum of the track is significantly smaller than the energy of the HCAL deposit, reconstruction proceeds under the hypothesis of a neutral particle overlapping with a charged hadron. In the rare event that the track momentum is significantly larger than the HCAL cluster energy, the particle may be reconstructed as a muon or the track may be treated as misreconstructed and discarded.

HF photons and hadrons

Due to the limited η coverage of the tracker and the ECAL, particles with $|\eta| > 3.0$ must be reconstructed using the forward HCAL (HF) only. The HF contains quartz fibers of two different lengths, placed such that the longer fibers span the entire HF while the shorter ones only occupy the rear part. This layout allows each energy deposit to be roughly decomposed into electromagnetic (EM) and hadronic components, with the EM component registering mainly in the longer fibers. EM and hadronic hits in the HF are classified by the PF algorithm as forward photons and hadrons respectively, and are included in the list of PF candidates without further calibration.

Primary and secondary vertices

The reconstructed tracks in each event are used to estimate the number and locations of the collision vertices. Vertexing is performed using a deterministic annealing (DA) algorithm [26]. Vertices located along the beamline (originating from *pp* interactions) are referred to as *primary vertices*. The resolution of the z position of the reconstructed primary vertices depends on the number of tracks in the event; for events with more than 50 reconstructed tracks, it is better than 25 μm. The primary vertex having tracks summing to the largest total transverse momentum in the event is considered to be the vertex of interest, and all others are considered pileup vertices.

Secondary vertices are the decay locations of particles produced in the initial *pp* collision that live long enough to move away from the beam before decaying. The presence of secondary vertices is useful in identifying such particles and is used in the *b*-tagging procedure described later.

Jet identification

Collision events in CMS consist primarily of quantum chromodynamic (QCD) *jets*, which are collimated sprays of hadrons produced when a high energy quark or gluon is knocked free of a proton. Jets are reconstructed by way of the *anti-k_T* clustering algorithm [28], which takes as input the collection of PF candidates in an event. The algorithm sequentially groups together PF candidates near one another in the two-dimensional space of η and ϕ according to the metric

$$d_{ij} = \min\left(k_{ti}^{-2}, k_{tj}^{-2}\right)\frac{\Delta R_{ij}^2}{R^2}, \tag{2.4}$$

where k_{ti} is the transverse momentum of the ith particle, $R > 0$ is a cone size parameter, and

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + \Delta\phi_{ij}^2 \tag{2.5}$$

is the distance between the two particles in the plane of η and ϕ. The distance $\Delta\phi_{ij}$ is the distance in ϕ between the two particles, taking into account the periodicity in ϕ. Particles with $\Delta R_{ij} > R$ are never clustered together, so each jet has a maximum radius of R.

The output of the anti-k_T algorithm is a collection of jet objects, each represented by its 4-momentum. Figure 2.11 shows an example event with the jets identified by the anti-k_T algorithm highlighted. The standard jets used for CMS physics analysis in Run II use anti-k_T with $R = 0.4$. Larger jets with $R = 0.8$ are used in analyses involving high-p_T massive particles (such as W, Z, or Higgs bosons or top quarks) decaying hadronically. The energies of the jets are corrected via a multi-stage procedure that compensates for any miscalibration of their raw values [29].

Pileup mitigation

Pileup interactions pollute every CMS physics event with additional particles not originating from the *pp* interaction of interest. There may be entire jets from pileup interactions as well as spurious particles within jets. Charged pileup particles can be identified by associating their tracks to a primary vertex. To reduce their effect on the jet reconstruction procedure, all charged PF candidates with tracks consistent with pileup vertices are discarded before the jet clustering algorithm is run. This procedure is known as *charged hadron subtraction* (CHS). The application of CHS removes approximately 70% of the charged pileup contribution to jet transverse momentum [29].

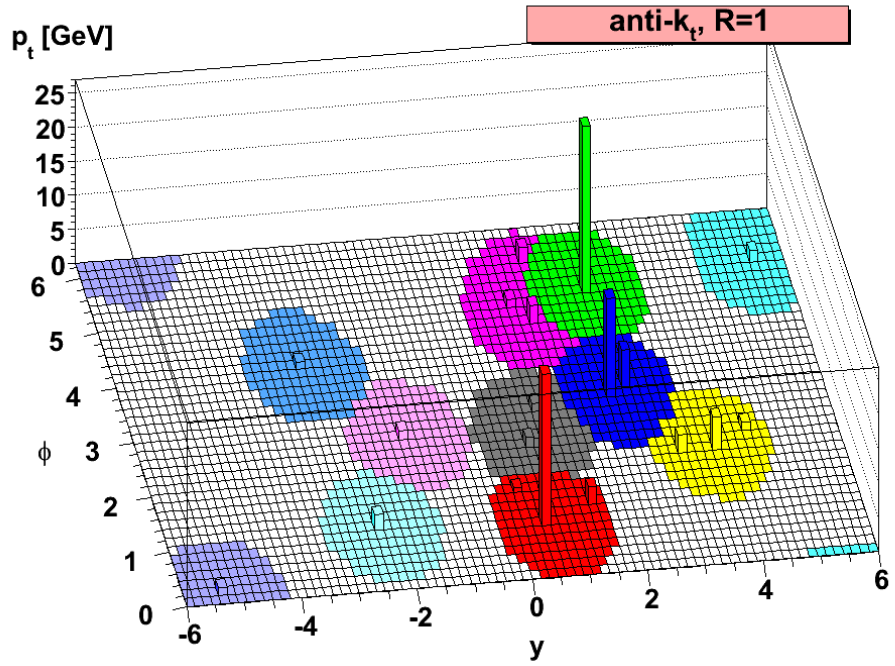


Figure 2.11: A simple parton-level event clustered with the anti-k_T algorithm. The cell heights indicate particle momenta, and the colored areas indicate the jets returned by the algorithm [28].

Neutral particles from pileup cannot be identified using track information. Their effect is estimated with the *jet areas* method, which corrects each jet's momentum via the formula

$$p_T^{corr} = p_T^{jet} - \rho * A^{jet}, \tag{2.6}$$

where p_T^{jet} is the nominal jet momentum, ρ is an estimate of the median p_T density in the $\eta - \phi$ plane, and A^{jet} is an effective jet area. The jet areas are calibrated to accurately subtract the pileup contribution to jet momenta.

Jet b-tagging

Many physics processes of interest are marked by the presence of b quarks in the final state. Jets originating from b quarks ('b-jets') have composition different from other jets, containing B-mesons that can live long enough to leave secondary decay vertices a few millimeters from the beam axis.

The Combined Secondary Vertex (CSV) algorithm is a classifier that estimates the likelihood that a jet is a b-jet. The classifier combines several input variables and

outputs a discriminator value between 0 and 1, with higher values indicating higher confidence that the jet is a b-jet. The input variables include the number of reconstructed secondary vertices, secondary vertex position and mass, and the number of tracks and their characteristics such as their distance of closest approach to the beam axis.

CMS analyses on Run II data make use of an improved version of the original CSV algorithm, dubbed CSVv2. Improvements from the previous version of the algorithm include a new secondary vertex reconstruction algorithm, a larger number of input variables, and the use of an artificial neural network for discrimination instead of a simple likelihood ratio [23]. For jets with $p_T > 20$ GeV, the 'medium' working point for the CSVv2 algorithm has approximately a 63% chance of identifying a b-jet. The mistag probability depends on the flavor of the quark that originated the jet. The chance of tagging a c-quark jet as a b-jet is 12%, and that of tagging a u, d, or s quark is 0.9%.

Missing transverse energy

The total (vector) momentum in the transverse plane of CMS in each pp collision is typically approximately zero. We define the missing transverse momentum (interchangeably called missing transverse energy) as the negative vector sum $\vec{p}_T^{miss}$ of all reconstructed PF candidate momenta in an event:

$$\vec{p}_T^{miss} = -\sum_{i=0}^{N_{PF}} \vec{p_{Ti}}. \tag{2.7}$$

The magnitude of $\vec{p}_T^{miss}$ is indicated by E_T^{miss}. Deviations of E_T^{miss} from zero are suggestive of either mismeasurement of particle or jet momenta or the presence of particles that escaped the detector without being recorded. In the latter case, the escaping particles might be neutrinos, or they might be exotic particles such as those from SUSY. The E_T^{miss} is hence often used as a search variable in CMS SUSY analyses.

As described above, the momentum of each jet in the event is corrected for pileup contamination and other effects. These jet energy corrections are propagated to the $\vec{p}_T^{miss}$:

$$\vec{p}_T^{miss} \rightarrow \vec{p}_T^{miss} - \sum_{jets} (\vec{p}_{T,jet}^{corr} - \vec{p}_{T,jet}). \tag{2.8}$$

This correction (referred to in CMS as the Type-1 $\vec{p}_T^{miss}$ correction) allows the measured $\vec{p}_T^{miss}$ to benefit from the calibrations derived for jets.

The resolution of the measured E_T^{miss} depends on the number of primary vertices in the event. For low numbers of vertices the resolution is 10-15 GeV. Each additional pileup vertex contributes 3.3-3.6 GeV of smearing in quadrature to the measured E_T^{miss} [30].

2.5 MC simulation

Physics analyses in CMS usually make heavy use of MC simulation to predict the amount of SM background and potential new physics signal present in the data. This is accomplished via a three-stage procedure.

First, the matrix elements for the physics process of interest are estimated via direct calculation of Feynman diagrams. Samples are then drawn from the resulting differential distribution; each sample is an independent MC event. This procedure is performed by software such as MadGraph [31].

Second, a program such as Pythia [32, 33] is used to perform QCD showering and hadronization of free quarks and gluons. A matching procedure [34] is used to avoid double-counting of jets from initial state radiation (ISR). The parameters of the shower algorithm are tuned to achieve good empirical correspondence with data.

Third, the interaction of each particle with a detailed model of the CMS detector is simulated. CMS uses two different detector simulation frameworks. GEANT4 [35] is used for full detector simulation, and is the choice for most MC samples. A simplified 'fastsim' framework [36] has also been developed for situations in which large numbers of events must be simulated quickly. It is used mainly in the simulation of SUSY simplified models, where hundreds of MC samples, each with a different combination of SUSY particle masses, must be generated.

Chapter 3

SUPERSYMMETRY AND SEARCHES AT THE LHC

We begin this chapter by giving a short overview of the SM and the hierarchy problem. Next we describe supersymmetry and discuss its experimental signatures at hadron colliders. We then describe the paradigm for SUSY searches at the LHC, and motivate the searches described in Part III in the context of previous SUSY searches performed on data from Run I.

3.1 Brief overview of the standard model

The standard model of particle physics contains three classes of particles:

- Force-carrying particles, which are bosons with spin one,
- Matter particles, which are fermions with spin one half,
- The Higgs particle, which is a boson with spin zero.

In the SM, the electromagnetic, weak nuclear, and strong nuclear (QCD) forces arise as a consequence of the gauge symmetry that underlies the theory. The symmetry is governed by the group $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. Each particle is associated with a relativistic quantum field and participates in some or all of the fundamental interactions. The particle content of the SM is illustrated in Figure 3.1.

Gauge bosons

The internal $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry of the SM is a gauge symmetry, meaning that the theory is invariant under arbitrary group transformations at each point of spacetime (gauge transformations). To enforce this invariance, it is necessary to include spin-one bosonic fields in the theory that transform under the adjoint representation of the group [37]. The particles associated with these bosonic fields are the force-carrying particles of the SM. In particular, each factor of the SM symmetry group corresponds to a set of fields as follows:

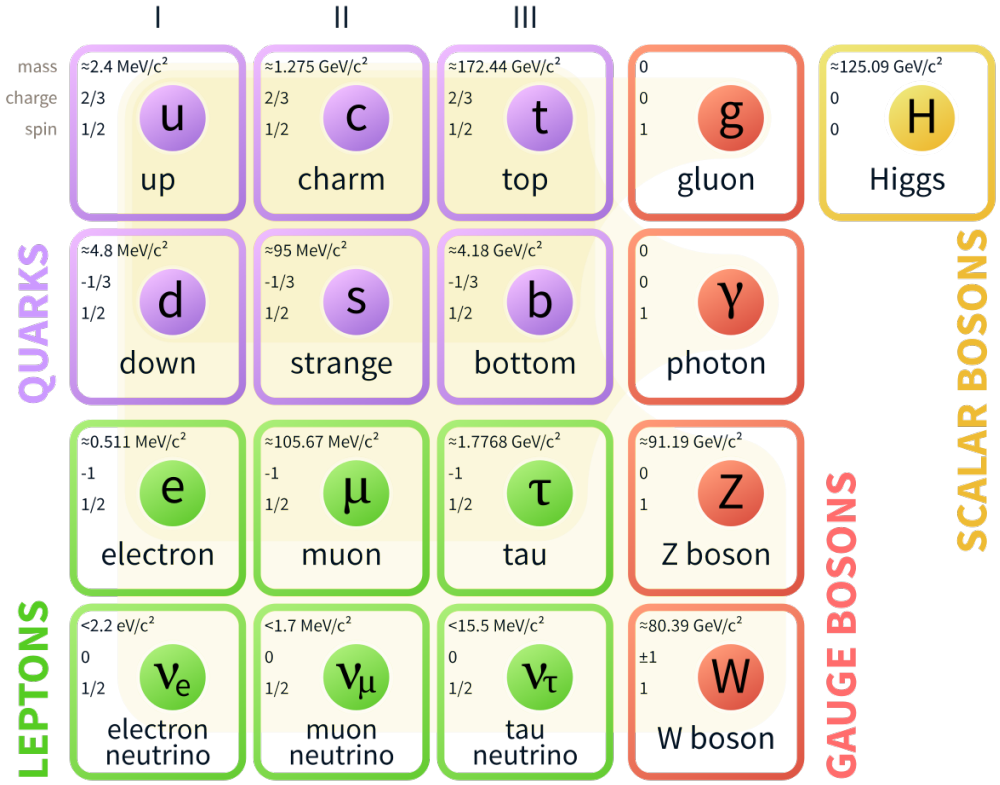


Figure 3.1: The particle content of the SM.

- The eight *gluons* G^{α}_{μ} are associated with the $SU(3)_C$ factor. They are the carriers of the strong nuclear force.
- The three bosons W^{a}_{μ} are associated with $SU(2)_L$.
- The boson B_{μ} is associated with $U(1)_Y$.

None of these *gauge bosons* have explicit mass terms in the SM Lagrangian, as these would not respect the gauge symmetry. Through the process of electroweak symmetry breaking described below, the bosons W^{a}_{μ} and B_{μ} mix to form the $W^{\pm}$ and Z bosons, the carriers of the weak nuclear force; and the photon γ, the carrier of the electromagnetic force. This process also imbues the $W^{\pm}$ and Z bosons with nonzero mass.

Matter fields

The fermions in the SM fall into two categories: quarks, which interact via the strong nuclear force, and leptons, which do not. Both quarks and leptons come in

three 'generations.' The first generation contains the particles that dominate the interactions of ordinary matter: the up and down (u and d) quarks, the electron e, and the electron neutrino ν_e. The second generation contains the charm and strange (c and s) quarks, the muon μ, and the muon neutrino ν_μ. The third generation contains the top and bottom (t and b) quarks, the τ lepton, and its associated neutrino ν_τ.

Fermions are represented in the theory by left- and right-handed Weyl spinors [38]. The $\mathrm{SU}(2)_L$ interaction in the SM is chiral: the left-handed fermions fall into $\mathrm{SU}(2)_L$ doublets while the right-handed ones are singlets. Each fermion has both a left- and a right-handed component, except for the neutrinos, which in the SM are left-handed only. Experimental observation of neutrino flavor oscillations implies that neutrinos have nonzero mass, and therefore that right-handed neutrinos must exist [39]. However, the form that right-handed neutrinos should take in the SM is not yet known, and we currently speak of the SM as a theory with massless neutrinos.

The Higgs field

The SM contains a fundamental scalar, the Higgs field Φ, which governs the mechanism of electroweak symmetry breaking. It is a complex $\mathrm{SU}(2)_L$ doublet,

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{3.1}$$

with a quartic potential,

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi\right)^2, \tag{3.2}$$

governed by two parameters, μ and λ. The Higgs field is charged under the $\mathrm{U}(1)_Y$ gauge group.

The SM Lagrangian

The full Lagrangian of the SM can be broken into the following terms:

$$\mathcal{L}_{SM} = \mathcal{L}_{gauge} + \mathcal{L}_{fermion} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}. \tag{3.3}$$

The terms $\mathcal{L}_{gauge}$ and $\mathcal{L}_{fermion}$ contain the kinetic energy terms for the gauge fields and for the quarks and leptons. For brevity we will not write them out here. The

term $\mathcal{L}_{Higgs}$ contains the kinetic and potential terms for the Higgs field:

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi), \tag{3.4}$$

where D_μ is the gauge covariant derivative:

$$D_\mu \Phi = \left(\partial_\mu - i g_2 \frac{\sigma_a}{2} W_\mu^a - i g_1 \frac{1}{2} B_\mu \right) \Phi, \tag{3.5}$$

with g_1 and g_2 the coupling constants of the $\mathrm{U}(1)_Y$ and $\mathrm{SU}(2)_L$ interactions, respectively.

The last term in $\mathcal{L}_{SM}$ contains the couplings between the Higgs field and the fermion fields [40]:

$$\mathcal{L}_{Yukawa} = -\hat{y}_{d_{ij}} \bar{q}_{L_i} \Phi d_{R_j} - \hat{y}_{u_{ij}} \bar{q}_{L_i} \tilde{\Phi} u_{R_j} - \hat{y}_{l_{ij}} \bar{l}_{L_i} \Phi e_{R_j} + h.c. \tag{3.6}$$

Here q_{L_i} and l_{L_i} denote the left-handed quark and lepton doublets; d_{R_j}, u_{R_j}, and e_{R_j} denote the right-handed down-type quarks, up-type quarks, and leptons, respectively; and $\tilde{\Phi} = i\sigma_2 \Phi^*$. The Yukawa couplings $\hat{y}_{d_{ij}}$, $\hat{y}_{u_{ij}}$, and $\hat{y}_{l_{ij}}$ describe the interaction strength between the Higgs field and the down-type quarks, up-type quarks, and leptons, respectively. The indices i and j index the three generations of fermions.

Electroweak symmetry breaking

As formulated, the SM is a theory of massless particles (besides the Higgs, which has an explicit mass term), and it is expressed in terms of bosons W_μ^a and B_μ that are not observed in nature. The observed bosons, and the masses of the fermions, arise via the process of electroweak symmetry breaking (EWSB) [41, 42], which is governed by the dynamics of the Higgs field.

EWSB occurs if the quadratic term in the Higgs potential (Eq. 3.2) is negative. In this case the minimum of the potential is not at zero: the field will have a nonzero vacuum expectation value (VEV), which we may express up to an arbitrary field redefinition as

$$\langle 0 | \, \Phi \, | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{3.7}$$

with $v = \sqrt{\mu^2/\lambda}$. This vacuum state breaks the $\mathrm{SU}(2)_L$ symmetry that previously allowed arbitrary rotations in the space of ϕ^+ and ϕ^0. The field Φ still has an arbitrary phase, so a U(1) symmetry (called $\mathrm{U}(1)_{EM}$) remains.

This 'spontaneous breaking' of $SU(2)_L$ to $U(1)_{EM}$ gives rise to the electroweak $W^{\pm}$ and Z bosons as follows. The squared covariant derivative in Eq. 3.4 yields terms of the form

$$(D_\mu \Phi)^\dagger (D^\mu \Phi) = \left(\frac{g_2^2 v^2}{4}\right) W_\mu^+ W^{\mu -} + \frac{1}{2}\left(\frac{\left(g_2^2 + g_1^2\right) v^2}{4}\right) Z_\mu Z^\mu + \dots \tag{3.8}$$

where the $W^{\pm}$ and Z fields are defined as linear combinations of the $SU(2)_L$ and $U(1)_Y$ gauge bosons:

$$W_\mu^\pm = \frac{W_\mu^1 \mp i W_\mu^2}{\sqrt{2}}, \quad Z_\mu = \frac{g_2 W_\mu^3 - g_1 B_\mu}{\sqrt{g_1^2 + g_2^2}}. \tag{3.9}$$

The terms shown in Eq. 3.8 act as mass terms for these bosons. The physical W and Z bosons therefore have masses:

$$m_W^2 = \frac{g_2^2 v^2}{4}, \quad m_Z^2 = \frac{\left(g_2^2 + g_1^2\right) v^2}{4} \equiv \frac{m_W^2}{\cos^2 \theta_W}, \tag{3.10}$$

where θ_W is called the weak mixing angle. We also define the photon field,

$$A_\mu = \frac{g_1 W_\mu^3 + g_2 B_\mu}{\sqrt{g_1^2 + g_2^2}}. \tag{3.11}$$

No mass term for this field appears in Eq. 3.8, and the photon remains massless.

After EWSB, excitations of the Higgs field around its vacuum value can be expressed as

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{3.12}$$

The excitation $H(x)$ is a physical particle, the Higgs boson, with mass $m_H = \sqrt{2\lambda} v$.

The fermion masses arise from the Yukawa interactions (Eq. 3.6). To see this, we replace Φ by v in each of the Yukawa terms. After making a field redefinition to rotate the fermions into the mass eigenstate basis (such that $\hat{y}_{f_{ij}} \rightarrow y_{f_i} \delta_{ij}$ for the up/down type quarks and the leptons), Eq. 3.6 becomes

$$-\mathcal{L}_{Yukawa} = m_{d_i} \bar{d}_{L_i} d_{R_i} + m_{u_i} \bar{u}_{L_i} u_{R_i} + m_{l_i} \bar{l}_{L_i} l_{R_i} + h.c., \tag{3.13}$$

summed over $i = 1, 2, 3$, where the fermion masses are $m_{f_i} = y_{f_i} v/\sqrt{2}$. Since there are no right-handed neutrinos in the theory, no neutrino mass term appears and the neutrinos remain massless.

The Yukawa terms also give rise to interactions between the fermions and the physical Higgs boson. These interactions have the same form as Eq. 3.13 except that they feature the Higgs boson H in place of the VEV v. The form of these interactions implies that the strength of the Higgs boson's interaction with each fermion is proportional to the fermion's mass.

The hierarchy problem

The Yukawa interaction between the Higgs particle and the SM fermions leads to quantum corrections to the Higgs mass m_H. The leading correction is from the one-loop Feynman diagram shown on the left side of Figure 3.2, which contributes [43]

$$\Delta m_H^2 = -\frac{\left|y_f\right|^2}{8\pi^2}\Lambda_{UV}^2 + \dots \tag{3.14}$$

for each fermion f in the theory. Here Λ_{UV} is an ultraviolet cutoff regulating the loop integral calculation; it represents a hypothetical scale at which new physics beyond the SM becomes relevant. If the SM holds at energies all the way up to the Planck mass $M_P \sim 10^{18}$ GeV, then $\Lambda_{UV} \sim M_P$ and the correction Δm_H is many orders of magnitude larger than the observed Higgs mass.

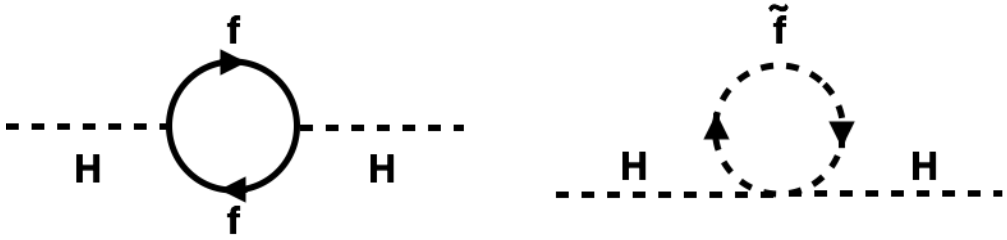


Figure 3.2: Left: one of the Feynman diagrams that generate large corrections to the Higgs boson mass in the SM. Right: the diagram in the MSSM that cancels the quadratic part of the Higgs mass correction from the left diagram.

It is possible that the 'bare' Higgs mass m_H and the correction Δm_H delicately cancel so as to yield the observed Higgs mass of 125 GeV. This cancellation requires these two quantities to be 'fine-tuned' such that their values, each of order 10^{18}, only differ by ~ 100 GeV. This situation conflicts with the idea of *naturalness*, which is the suggestion that physics at low energy scales (here the electroweak scale) should not be finely sensitive to the details of physics at much higher energy scales (here the Planck scale) [44].

This *hierarchy problem* for the Higgs mass is alleviated if there are additional scalar fields S that couple to the Higgs through Lagrangian terms of the form $-\lambda_S |H|^2 |S|^2$. In this situation, the Higgs mass receives an additional correction from the diagram shown on the right side of Fig. 3.2:

$$\Delta m_H^2 = \frac{\lambda_S}{16\pi^2} \Lambda_{UV}^2 + \dots \tag{3.15}$$

This correction has opposite sign from the one in Eq. 3.14. If the couplings λ_S are related in an appropriate way to the Higgs couplings to fermions, the quadratic part of the correction to the Higgs mass cancels. Supersymmetry, to which we turn now, provides a set of fields that alleviate the hierarchy problem in this way.

3.2 Theory of supersymmetry

Supersymmetry is a hypothetical symmetry of nature that relates bosonic and fermionic fields to one another. A supersymmetry transformation is generated by an operator Q that converts a fermionic state into a bosonic one, and vice versa. Schematically,

$$Q\,|\text{boson}\rangle = |\text{fermion}\rangle\,, \quad Q\,|\text{fermion}\rangle = |\text{boson}\rangle\,. \tag{3.16}$$

The operator Q is a Lorentz spinor. It is restricted to satisfy (anti-)commutation relations of the (very schematic) form [43, 45]:

$$\{Q, Q^\dagger\} = P^\mu, \tag{3.17}$$

$$\{Q, Q\} = \{Q^\dagger, Q^\dagger\} = 0, \tag{3.18}$$

$$[P^\mu, Q] = [P^\mu, Q^\dagger] = 0, \tag{3.19}$$

where P^μ is the momentum operator.

We see from Eq. 3.17 that Q mixes spacetime symmetries and internal symmetries in a nontrivial way: applying a combination of Q and $Q^\dagger$ to a state results in a spacetime translation (which are generated by P^μ). This distinguishes Q from the generators of the $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry group of the SM, which commute with all elements of the Poincaré group (as required by the Coleman-Mandula theorem [46] for symmetries generated by scalar operators). A theory with a symmetry generated by a fermionic operator with the form of Q is said to be *supersymmetric*.

In a supersymmetric theory, particles fall into *supermultiplets* that contain bosons and fermions related to one another by supersymmetry transformations. We consider two types of supermultiplets:

- *Chiral supermultiplets* contain one Weyl fermion and one complex scalar.
- *Vector supermultiplets* contain one Weyl fermion and one spin-1 vector boson.

Chiral and vector supermultiplets each have two fermionic and two bosonic degrees of freedom – Weyl fermions and (massless) vector bosons each have two helicity states, and complex scalars have real and imaginary components. The particles in a supermultiplet have the same mass and transform in the same representation of the SM gauge group.

3.3 SUSY in the SM

To extend the SM to be a supersymmetric theory, it is first necessary to place all of the known SM particles into supermultiplets. It is not possible for any of the known particles to be superpartners of one another [47], so this involves doubling the particle content of the SM, introducing a new superpartner for each SM particle. The SM gauge bosons, having spin one, fall into vector supermultiplets and have spin-$1/2$ superpartners, called *gauginos*. The SM fermions fall into chiral supermultiplets and have scalar superpartners, called *sfermions*.

The SM Higgs boson is a scalar and hence must be part of a chiral supermultiplet. In fact, the structure of the SUSY Lagrangian requires that the theory contain at least *two* Higgs chiral supermultiplets. One, denoted H_u, gives masses to the up-type quarks, and the other, denoted H_d, gives masses to the down-type quarks and to the charged leptons [43]. After EWSB, there are a total of five scalar Higgs particles: three neutral (two CP-even and one CP-odd) and two charged (one positive and one negative). One of the CP-even neutral scalars, denoted h^0, has very similar interactions to those of the SM Higgs boson. While the masses of the other four Higgs scalars can be arbitrarily large, there is an upper bound on the mass of h^0. At tree level it cannot exceed the mass of the Z boson, but higher-order loop corrections push the bound up to ~140 GeV, depending on the masses of the top superpartners [48]. The Higgs-like particle discovered by CMS and ATLAS in Run I of the LHC is therefore consistent with h^0.

Placing the known SM particles into supermultiplets and adding one additional Higgs supermultiplet are the minimal changes required to formulate the SM in a SUSY framework. This Minimal Supersymmetric Standard Model (MSSM) forms the basis of much of the LHC SUSY search program.

SUSY breaking

Adding SUSY to the SM solves the hierarchy problem for the Higgs boson mass via the addition of two scalar degrees of freedom for every SM fermion. The couplings have the appropriate values such that the quadratic contribution from Eq. 3.14 is exactly cancelled by that from Eq. 3.15. Moreover, no MSSM scalar's mass suffers from a quadratic divergence at any order in perturbation theory [43].

However, if nature were fully supersymmetric, all of the superpartners would have the same masses and couplings as their SM counterparts, and many of them would have been discovered already. We know therefore that if SUSY exists, it is inexact or spontaneously broken. Terms that break SUSY (i.e. that introduce differences between particles and their superpartners) can be introduced into the MSSM Lagrangian without reintroducing the hierarchy problem into the theory. Such terms are called *soft* SUSY-breaking terms and are characterized by positive mass dimension. They can take the following general forms [49, 50]:

- Bilinear and trilinear couplings for the scalar superpartners
- Masses for the scalar superpartners
- Masses for the gauginos

This set of soft SUSY-breaking terms is sufficient for all of the MSSM superpartners to have masses different from their SM counterparts. Introducing these terms adds a large number of parameters to the MSSM; the full Lagrangian with soft SUSY-breaking terms included has over 100 free parameters that are not present in the SM [43]. Simplifying assumptions are often made to reduce the number of parameters to a manageable level for interpretation of collider searches [51].

Electroweak symmetry breaking and the μ parameter

Before SUSY is broken, the minimum of the Higgs potential is at 0 and there is no EWSB. The nature of EWSB in the MSSM depends on the SUSY-breaking parameters in the theory. If electroweak symmetry is broken, the neutral components of H_u and H_d obtain VEVs, denoted v_u and v_d respectively. We parameterize the ratio of the VEVs as

$$\tan\beta = v_u/v_d. \tag{3.20}$$

The minimization condition for the Higgs potential in the MSSM is [52]:

$$\frac{1}{2}m_Z^2 = \frac{m_{H_d}^2 - m_{H_u}^2 \tan^2\beta}{\tan^2\beta - 1} - \mu^2, \tag{3.21}$$

where m_{H_D} and m_{H_u} are soft SUSY-breaking parameters. The parameter μ is part of the MSSM Lagrangian and is unrelated to the breaking of SUSY. To avoid large fine-tuning between the two terms of Eq. 3.21, μ should have a value near the SUSY-breaking scale. Variants of the MSSM, such as the NMSSM [53], dynamically generate μ from a mechanism related to SUSY breaking, thus ensuring that these scales are close to one another.

R-parity

Unlike in the SM, the Lagrangian of the MSSM contains terms that explicitly violate the conservation of lepton and baryon number. Unless there is a mechanism to suppress these terms, the interactions allowed by the MSSM would not respect the observed limits on, e.g., the lifetime of the proton. One way to prevent this from happening is to introduce a quantity called R-parity:

$$P_R = (-1)^{3(B-L)+2s}, \tag{3.22}$$

and postulate that it is conserved in the MSSM. Here B and L are the baryon and lepton number of a particle and s is its spin. All of the particles in the SM have even R-parity, while all of their superpartners have odd R-parity.

The effect of enforcing R-parity conservation in the MSSM is to remove the L- and B-violating terms from the Lagrangian. It also has a striking implication for collider phenomenology. If R-parity is conserved, then every interaction vertex in the theory contains an even number of particles with $P_R = -1$. This implies that:

- Supersymmetric particles must be produced in pairs in collider events, and
- The lightest supersymmetric particle (abbreviated LSP) is absolutely stable, and every SUSY particle will eventually decay to a final state consisting of SM particles and LSPs.

The LSP also provides a particle candidate for the dark matter whose existence in the universe is suggested by astrophysical observations [54].

Theories that violate R-parity are sometimes considered in collider searches for SUSY, but in general R-parity conservation is assumed to hold.

Summary of the MSSM superpartners

Here we briefly describe each of the classes of superpartner particles that exist in the MSSM, and discuss the interactions of each.

Neutralinos and charginos

The superpartners of the SM electroweak bosons ('binos' and 'winos'), and those of the Higgs scalars ('higgsinos'), mix with one another to form four neutral and two charged mass eigenstates. The neutral states ('neutralinos'), denoted $\tilde{\chi}_1^0$, $\tilde{\chi}_2^0$, $\tilde{\chi}_3^0$, and $\tilde{\chi}_4^0$, are mixtures of the superpartners of the SM B_μ and W_μ^0 and the neutral Higgs scalars. The charged states ('charginos'), denoted $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^\pm$, are mixtures of the superpartners of the W_μ^1, the W_μ^2, and the charged Higgs scalars. The physical neutralino masses are obtained by diagonalizing the matrix [43]

$$M_{\tilde{N}} = \begin{pmatrix} M_1 & 0 & -c_\beta s_W m_Z & s_\beta s_W m_Z \\ 0 & M_2 & c_\beta c_W m_Z & -s_\beta c_W m_Z \\ -c_\beta s_W m_Z & c_\beta c_W m_Z & 0 & -\mu \\ s_\beta s_W m_Z & -s_\beta c_W m_Z & -\mu & 0 \end{pmatrix}, \tag{3.23}$$

where M_1 and M_2 are SUSY breaking mass parameters, c_W and s_W denote $\cos\theta_W$ and $\sin\theta_W$, and c_β and s_β denote $\cos\beta$ and $\sin\beta$. The chargino masses are obtained by diagonalizing the block matrix

$$M_{\tilde{C}} = \begin{pmatrix} 0 & X^T \\ X & 0 \end{pmatrix}, \quad X = \begin{pmatrix} M_2 & \sqrt{2} s_\beta m_W \\ \sqrt{2} c_\beta m_W & \mu \end{pmatrix}. \tag{3.24}$$

The decays of neutralinos and charginos most relevant for searches for hadronically produced SUSY at the LHC are illustrated in Figure 3.3. A neutralino can undergo a decay to a lighter neutralino and a neutral gauge or Higgs boson. A chargino can decay to a neutralino and a charged W or Higgs boson. Neutralinos and charginos can both decay to a fermion-sfermion pair.

Gluinos

The gluino is the color-octet superpartner of the SM gluon. Being charged under $\mathrm{SU}(3)_c$, it cannot mix with the other superpartners.

Gluinos decay via the squark-quark-gluino interaction vertex illustrated on the left side of Figure 3.4. If the gluino is heavier than at least one squark, it will undergo

the decay $\tilde{g} \to \tilde{q}q$. If it is lighter than the squarks, it will decay through off-shell squarks, e.g. $\tilde{g} \to q\bar{q}\tilde{\chi}_1^0$.

Being strongly interacting, gluinos may be produced with significant cross section at the LHC. Example diagrams illustrating gluino pair production are shown in the top row of Figure 3.5.

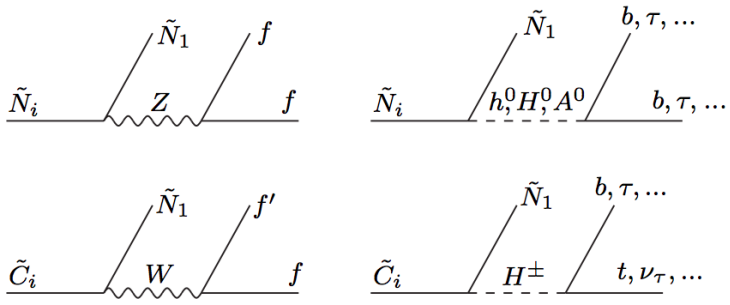


Figure 3.3: Two-body decays of MSSM neutralinos and charginos to gauge or Higgs bosons. Decays to a fermion-sfermion pair (not shown) are also possible [43].

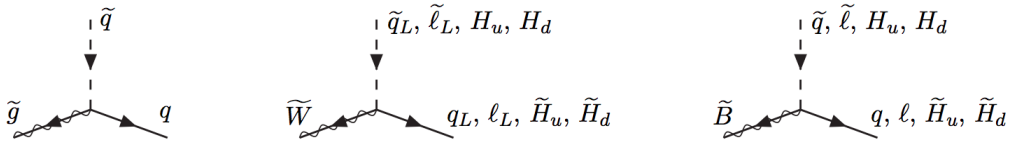


Figure 3.4: Feynman diagrams for MSSM three-point gluino (left), wino (center), and bino (right) couplings to fermions [43].

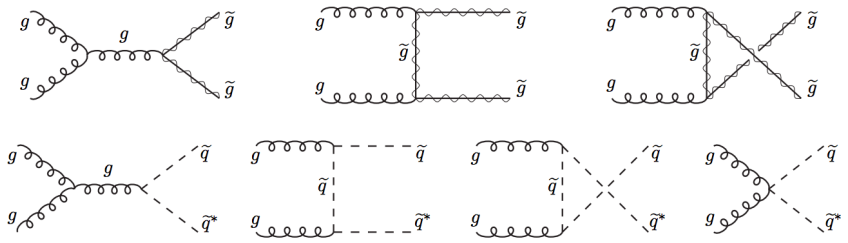


Figure 3.5: Diagrams illustrating gluino (top row) and squark (bottom row) production from two colliding gluons [43].

Squarks and sleptons

The scalar superpartners of the SM quarks are called *squarks*, and those of the SM leptons and neutrinos are called *sleptons* and *sneutrinos*. Their names are formed

by appending '*s*' to the SM particle name: *stop*, *sbottom*, and so on. There are two squarks or sleptons for each SM quark and charged lepton, which correspond to the left- and right-handed field components; and one sneutrino for each SM neutrino.

In principle there can be arbitrary mixing of the fields in each sfermion family: the down-type squarks, the up-type squarks, the sleptons, and the sneutrinos. It is often assumed that the SUSY-breaking MSSM parameters that mediate this mixing are small or zero [43]. Mixing within the third generation of squarks and sleptons does arise, however, as a result of the large Yukawa couplings for the third generation. Thus, e.g., the physical stops $\tilde{t}_1$ and $\tilde{t}_2$ may be mixtures of the left- and right-handed stop states $\tilde{t}_L$ and $\tilde{t}_R$. The Yukawa couplings also affect the running of the sfermion masses under the renormalization group (RG) equations, tending to give the third-generation sfermions different (usually lower) masses than those of the first and second generation.

Strong production of squarks from colliding gluons is illustrated in the bottom row of Figure 3.5. The squarks interact with the gluino as shown in Figure 3.4; they can therefore decay to a quark and a gluino if this is kinematically allowed. A squark can also decay to a quark and a neutralino or chargino; this is the primary decay mode if the squark is lighter than the gluino.

3.4 Natural SUSY and simplified signal models

The vastness of the MSSM parameter space requires that we make simplifying assumptions to tractably interpret LHC data in the context of SUSY. Some of the typical assumptions underlying LHC SUSY searches are:

- R-parity is conserved, and the neutralino $\tilde{\chi}_1^0$ is the LSP.
- The lightest chargino $\tilde{\chi}_1^\pm$ is close in mass to the LSP (it is usually taken to be 5 GeV heavier). Another neutralino, $\tilde{\chi}_2^0$, may also be close in mass. (Small neutralino and chargino mass splittings arise when μ is small, $\mu \ll M_1, M_2$ [55].)
- The gluino is much heavier than the neutralinos and charginos (this is motivated by its behavior under the RG, which is driven by the strong coupling constant [43]).

Certain SUSY parameter regions are also favored by naturalness considerations: the breaking of SUSY should not reintroduce an unacceptable level of fine-tuning into

the calculation of the h^0 mass. The 'acceptable' amount of fine-tuning is somewhat subjective, but some general considerations apply [56]:

- The Higgs mass increases with μ, which also controls the masses of the higgsinos. Thus, the higgsinos should not be too heavy.
- The dominant radiative correction to the Higgs potential is logarithmic in the stop masses [55]. Thus, both stops should be relatively light (as should the left-handed sbottoms).
- The stop mass is subject to loop corrections whose value involves the gluino mass, which thus influences the Higgs mass at two-loop level. Therefore the gluino should not be too heavy.

These constraints can also be understood in the context of Eq. 3.21: naturalness suggests that each term on the right-hand side should be roughly of order m_Z^2, and the soft breaking parameters m_{H_d} and m_{H_u} receive corrections from the stop and gluino masses at one and two loops, respectively. The authors of [52] suggest that higgsino masses below 500 GeV, third-generation squark masses below 1.5 TeV, and a gluino mass below 3 TeV, are appropriate expectations for a natural SUSY scenario.

Typical signatures of natural SUSY scenarios of the above type at the LHC include large E_T^{miss} from stable LSPs that escape the detector, high jet multiplicity due to gluino and/or squark decays, and b-tagged jets from SUSY particles decaying to third-generation quarks. In the case of decays to top quarks, the final state may also contain high-p_T leptons. Cascade decays of neutralinos and charginos to the LSP may yield additional energetic jets, leptons, or electroweak bosons if the mass splitting between SUSY states is large enough.

In many cases, interpretation of an LHC SUSY search is carried out using *simplified SUSY models*, in which two or three SUSY states are accessible and the rest are treated as decoupled. The masses of the SUSY particles in the model are treated as free parameters, and cross-section limits or discovery significances are derived as functions of the masses. The simplified models are not intended to represent realistic SUSY spectra; rather they provide a convenient low-dimensional framework for understanding the consequences of a search.

The searches in this thesis are primarily oriented toward studying strongly produced SUSY particles (gluinos and squarks). When interpreting these searches in terms of

specific simplified models, we consider the following scenarios (with standardized names explained in [57]):

- **T1tttt**, **T1bbbb**, **T1qqqq**: pair production of gluinos, with each gluino undergoing a three-body decay to $t\bar{t}\tilde{\chi}_1^0$ (for T1tttt), $b\bar{b}\tilde{\chi}_1^0$ (for T1bbbb), or $q\bar{q}\tilde{\chi}_1^0$ (for T1qqqq), with q a first- or second-generation quark.

- **T1ttbb**: pair production of gluinos with specific branching ratios to any combination of $t\bar{t}\tilde{\chi}_1^0$, $b\bar{b}\tilde{\chi}_1^0$, and $t\bar{b}\tilde{\chi}_1^-$ / $b\bar{t}\tilde{\chi}_1^+$. Interpretation in this model can be performed for different values of the decay branching ratios.

- **T5ttcc**: pair production of gluinos, with each gluino decaying to a top squark and a top quark. The stop mass is taken to be very close to the LSP mass, so that its dominant decay is to $c\tilde{\chi}_1^0$ instead of $t\tilde{\chi}_1^0$. This model is of interest in searches that target boosted event topologies because the top quarks can receive a significant fraction of the energy from the gluino decay.

- **T2tt**, **T2bb**, **T2qq**: pair production of stops, sbottoms, or light squarks, respectively, each decaying to an LSP and a top, bottom, or light quark as appropriate.

In simulation of SUSY simplified models, a particular production and decay chain is specified for the simulated events. Thus, e.g., in the T1tttt model, all MC simulated events have $pp \to \tilde{g}\tilde{g}$ even though direct electroweak production of LSPs is also possible under the model spectrum. Feynman diagrams for the decays considered in the listed models are shown in Figure 3.6.

3.5 SUSY searches in LHC Run I

Searches for SUSY in a variety of final states were carried out on the Run I CMS [58–72] and ATLAS [73–83] datasets. No conclusive evidence of the supersymmetric partner of any SM particle was observed. Limits on squark and gluino production cross sections were computed in a number of simplified model scenarios. The most stringent constraints come from the 2012 run, in which CMS and ATLAS each collected and analyzed 20 fb^{-1} of data at 8 TeV. Figures 3.7 and 3.8 summarize the sparticle mass regions excluded by CMS at the 95% confidence level in the simplified models considered.

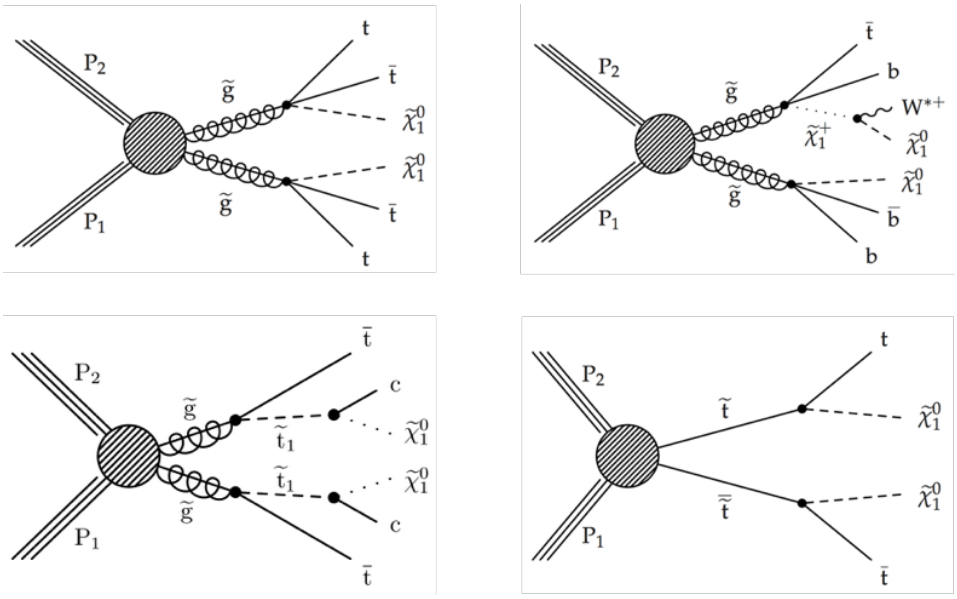


Figure 3.6: Diagrams illustrating the decay topologies for the simplified SUSY models T1tttt (upper left), T1ttbb (upper right), T5ttcc (lower left), and T2tt (lower right). The models T1bbbb and T1qqqq can be obtained from the T1tttt diagram by replacing all top quarks with bottom or light quarks, respectively. The models T2bb and T2qq can be obtained from the T2tt diagram in the same way (the stops should also be replaced with sbottoms or light squarks). The T1ttbb model also includes all other combinations of gluino decays to $t\bar{t}\tilde{\chi}_1^0$, $b\bar{b}\tilde{\chi}_1^0$, $t\bar{b}\tilde{\chi}_1^-$, and $b\bar{t}\tilde{\chi}_1^+$, in addition to the decay shown in the diagram.

In the aftermath of LHC Run I, work was done to summarize the impact of the searches on the natural SUSY parameter space. In [52], the authors randomly sample SUSY parameter points from a six-dimensional subspace under the constraints that the mass of h^0 is 125 GeV and the LSP is the lightest neutralino. Assuming that the higgsinos, third-generation squarks, and gluino are the only accessible states, they conclude that scenarios with $m_{stop1} < 230$ GeV or $m_{gluino} < 440$ GeV are clearly excluded, with $m_{stop1} < 300$ GeV or $m_{gluino} < 1040$ GeV excluded if the LSP is lighter than 200 GeV. In [51], CMS SUSY search results were analyzed in a Bayesian framework in the context of the 'phenomenological MSSM', or pMSSM. A similar bound (500 GeV) on the gluino mass was found, and a bound of 300 GeV was placed on the lightest colored SUSY particle. LSP masses below 300 GeV were also found to be strongly disfavored, though not ruled out. In [84], a meta-analysis of the 8 TeV CMS and ATLAS searches was performed. The distribution of p-values obtained across all searches was found to be consistent with a lack of new physics, though a deficit of search regions having observed counts below the SM expectation was noted.

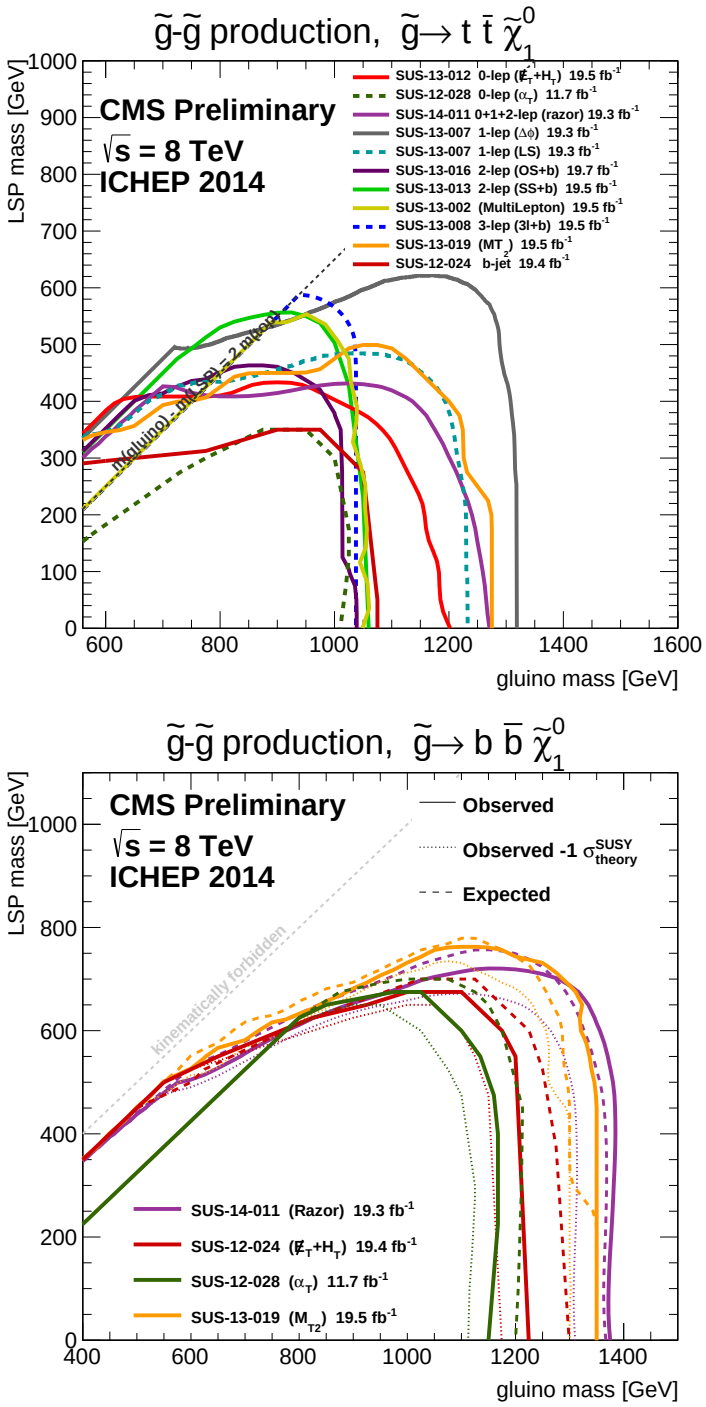


Figure 3.7: Limits on the T1tttt (top) and T1bbbb (bottom) simplified models at the 95% confidence level, computed from the results of CMS searches using the dataset from the 2012 LHC run. Limits are set in the two-dimensional plane of the gluino mass (x-axis) and the LSP mass (y-axis). The result of the search using razor variables is indicated by the pink curve in each plot [4].

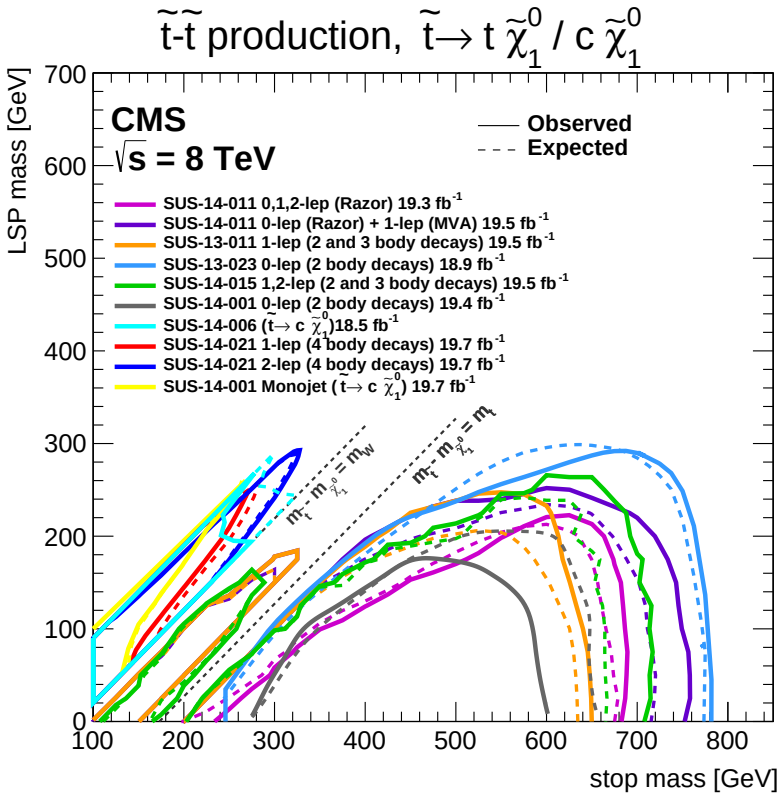


Figure 3.8: Limits on the T2tt simplified model at the 95% confidence level, computed from the results of CMS searches using the dataset from the 2012 LHC run. Limits are set in the two-dimensional plane of the stop mass (x-axis) and the LSP mass (y-axis). The result of the search using razor variables is indicated by the pink curve [4].

The null result from the 8 TeV SUSY searches has led some to revisit the assumptions of the naturalness paradigm. The authors of [85] formalize the fine-tuning argument in a probabilistic framework, and use it to argue that the stop mass may easily be above the LHC reach without compromising naturalness. Other theoretical proposals such as split SUSY [86, 87], stealth SUSY [88], and R-parity violating SUSY have also received increased attention. These have been the focus of searches in CMS that target distinctive signatures such as long-lived particles or large lepton multiplicities [89–91]. Other efforts have centered on 'compressed' SUSY scenarios in which the mass splitting between SUSY states is small. Limits on compressed models are generally much weaker than on those with large mass splittings [92].

3.6 SUSY at 13 TeV

The increase of the LHC energy from 8 to 13 TeV makes it possible to produce SUSY states of higher mass than before, and it grants a larger phase space to production of states that were already accessible at 8 TeV. The ratio of LHC parton luminosities at 13 and 8 TeV are shown in Figure 3.9 as a function of center-of-mass energy. This ratio increases rapidly at high energies. For gluinos produced via gluon-gluon interactions at 1.5-2 TeV, the cross section at 13 TeV is an order of magnitude higher than at 8 TeV. This indicates that searches for heavy gluinos with 13 TeV data should be able to improve existing limits even with a very small dataset. We thus tailor our 2015 razor SUSY search (Chapters 6 and 7) to emphasize gluino signals, before expanding to also focus on squark signals in the larger 2016 dataset (Chapter 8).

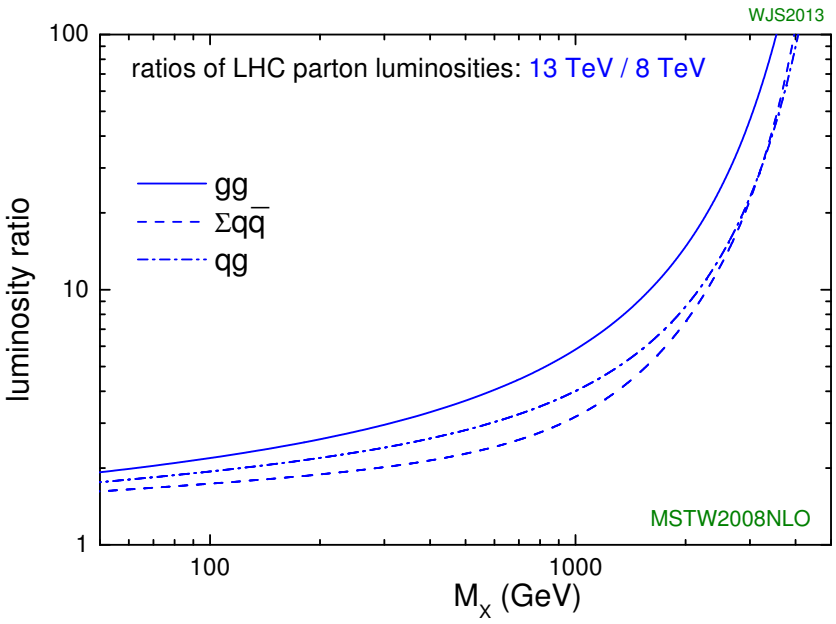


Figure 3.9: Ratio of LHC parton luminosities at 13 and 8 TeV for gg, qg, and qq interactions [93].

3.7 Razor kinematic variables

In this thesis we search for supersymmetric particles using *razor variables*, which are kinematic quantities developed specifically to identify SUSY-like signals of new

physics [94]. We describe here the derivation of the razor variables and discuss their utility in SUSY searches.

To begin, we consider pair production of heavy squarks at the LHC. Each squark decays to a quark, which is observed as a jet; and the LSP, which escapes without being detected. This is the situation in, e.g., the T2bb model shown in Figure 3.6. Let the squark and LSP masses be m_Q and m_χ, and assume the quarks to be massless. We have access to measurements of the lab-frame momenta of the jets, denoted p_1 and p_2, and to $\vec{p}_T^{\,miss}$, but not to the LSP momenta. We consider the problem of how to estimate the SUSY particle masses in this situation of imperfect information.

In the center-of-mass reference frame of each produced squark, the visible quark has momentum

$$q_i = \frac{M_\Delta}{2}\{1, \hat{u}_i\}, \tag{3.25}$$

where $i = 1, 2$, $\hat{u}_i$ are unit 3-vectors, and

$$M_\Delta = \frac{m_Q^2 - m_\chi^2}{m_Q} \tag{3.26}$$

carries information about the masses of the SUSY particles. If we knew what Lorentz boosts to perform to go from the lab frame to the squark rest frames, we could apply those boosts to $p_{1,2}$, recover $q_{1,2}$, and thus measure M_Δ.

Due to the missing kinematic information (the LSP momenta), the true boosts cannot be recovered exactly. We instead find approximations to the correct boosts by relying on the fact that the rest-frame energy of quark 1 equals that of quark 2 (both have value $M_\Delta/2$). Consider a boost in the z-direction, $\vec{\beta}_L$, which is applied to both p_1 and p_2; and a transverse boost, $\pm\vec{\beta}_R$, which is applied in opposite directions to the two particles. This is illustrated in Figure 3.10. The longitudinal boost $\vec{\beta}_L$ takes us from the lab frame into an approximation of the center-of-mass frame of the pp collision. The subsequent transverse boosts $\pm\vec{\beta}_R$ take us from the collision frame to the approximate rest frames of the produced squarks. Let γ_L and γ_R be the Lorentz factors associated with $\vec{\beta}_L$ and $\vec{\beta}_R$.

We obtain the values of the boosts $\vec{\beta}_L$ and $\vec{\beta}_R$ as follows. Let unprimed, primed, and double-primed variables indicate quantities measured in the lab frame, the approximate collision frame, and the approximate squark rest frames, respectively, and let E_i, p_{iz}, and $\vec{p}_{iT}$ represent the observed energies, longitudinal momenta, and transverse momenta of the jets. We have

$$E''_{1,2} = \gamma_R(E'_{1,2} \mp \vec{\beta}_R \cdot \vec{p}'_{1,2}) \tag{3.27}$$

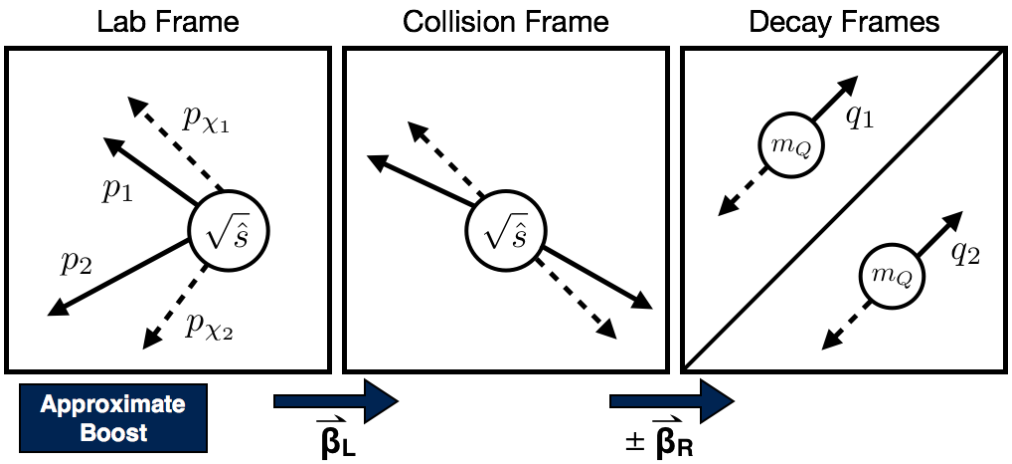


Figure 3.10: Diagram illustrating the reference frames considered in the derivation of M_R. Left: the lab frame, with observed jet momenta p_1 and p_2, and unobserved LSP momenta p_{χ_1} and p_{χ_2}. Center: the collision center-of-mass frame, which can be reached by a longitudinal boost that we approximate by $\vec{\beta}_L$. Right: the individual squark decay frames, where true quark momenta q_1 and q_2 are each proportional to $M_\Delta/2$. The boosts $\pm\vec{\beta}_R$ approximate the true boosts from the collision frame to the squark decay frames.

and

$$E'_{1,2} = \gamma_L(E_{1,2} - \vec{\beta}_L \cdot \vec{p}_{1,2}). \tag{3.28}$$

The equal energy condition, $E''_1 = E''_2$, can be expressed in terms of the collision-frame quantities as

$$E'_1 - E'_2 = \vec{\beta}_R \cdot \left(\vec{p}'_{1T} + \vec{p}'_{2T}\right). \tag{3.29}$$

Noting that the boost $\vec{\beta}_L$ is along z, we have $\vec{p}'_{iT} = \vec{p}_{iT}$. We write Eq. 3.29 in terms of lab-frame quantities and find an expression for the magnitude of the transverse boost:

$$\left|\vec{\beta}_R\right| = \frac{\gamma_L\left((E_1 - E_2) - |\vec{\beta}_L|(p_{1z} - p_{2z})\right)}{|\vec{p}_{1T} + \vec{p}_{2T}|\cos\phi_R}, \tag{3.30}$$

where ϕ_R is the azimuthal angle between $\vec{p}_{1T} + \vec{p}_{2T}$ and the boost $\vec{\beta}_R$.

We define the razor variable M_R by

$$M_R = \gamma_R(E''_1 + E''_2). \tag{3.31}$$

M_R is an estimator for the SUSY-sensitive quantity M_Δ. Because we have imperfect information about the event kinematics, two additional constraints must be imposed in order to uniquely specify the boosts $\vec{\beta}_{L,R}$ and hence the value of M_R. Having in mind the context of massive SUSY particles produced near threshold,

we first choose to make the transverse boost $\vec{\beta}_R$ as small as possible. This choice corresponds to $\cos\phi_R = 1$, i.e., a boost along the direction of $\vec{p}_{1T} + \vec{p}_{2T}$.

For the second constraint, we choose $\vec{\beta}_L$ so that M_R is extremized. This is satisfied by

$$\vec{\beta}_L = \frac{p_{1z} + p_{2z}}{E_1 + E_2}\hat{z}. \tag{3.32}$$

After expressing M_R in terms of lab-frame quantities and using Eqs. 3.30 and 3.32 for the boosts, we obtain the simple formula

$$M_R = \sqrt{(E_1 + E_2)^2 - (p_{z1} + p_{z2})^2}, \tag{3.33}$$

which is equal to the transverse energy in the event.

The second razor variable, R^2, is obtained by considering a different estimator for M_Δ as follows:

$$\begin{aligned} M_\Delta &\approx \sqrt{m_Q^2 - m_\chi^2} \\ &= \sqrt{(E_{1,2} + E_{\chi_{1,2}})^2 - (\vec{p}_{1,2} + \vec{p}_{\chi_{1,2}})^2 - m_\chi^2} \\ &= \sqrt{2E_{\chi_{1,2}}E_{1,2} - 2\vec{p}_{\chi_{1,2}} \cdot \vec{p}_{1,2}} \\ &= \sqrt{(E_{\chi_1}E_1 - \vec{p}_{\chi_1} \cdot \vec{p}_1) + (E_{\chi_2}E_2 - \vec{p}_{\chi_2} \cdot \vec{p}_2)}, \end{aligned} \tag{3.34}$$

where E_{χ_i} and $\vec{p}_{\chi_i}$ are the energies and momenta of the LSPs. The second and third lines are true for the decay products of either squark; in the fourth line we average over the two squarks. The approximation in the first line assumes that the LSP is light compared with the squark.

To evaluate this expression, we must estimate the (unknown) LSP momenta. We do this via the rough approximation that the LSPs are transverse and each carry half of the missing energy: $E_{\chi_1} = E_{\chi_2} = E_T^{miss}/2$ and $\vec{p}_{\chi_1} = \vec{p}_{\chi_2} = \vec{p}_T^{miss}/2$. The above expression then becomes

$$M_T^R \equiv \sqrt{\frac{E_T^{miss}}{2}(E_1 + E_2) - \frac{\vec{p}_T^{miss}}{2} \cdot (\vec{p}_{T1} + \vec{p}_{T2})}. \tag{3.35}$$

We have neglected the longitudinal LSP momenta, so the variable M_T^R is less than M_Δ, with a kinematic endpoint at that value.

We define R^2 as the dimensionless ratio

$$R^2 \equiv \left(\frac{M_T^R}{M_R}\right)^2. \tag{3.36}$$

For signal events, M_R and M_T^R both approximate M_Δ (each using different kinematic information from the event) and R^2 will be approximately 1/4. Events with little E_T^{miss} or fake E_T^{miss} from mismeasured jets have smaller values of R^2.

To search for SUSY using the razor variables, we select events with M_R near the mass scale of the signal under consideration, and apply a cut on R^2 to reduce the background from events whose kinematics are not consistent with the signal. Figure 3.11 shows typical M_R and R^2 distributions for SUSY signal events.

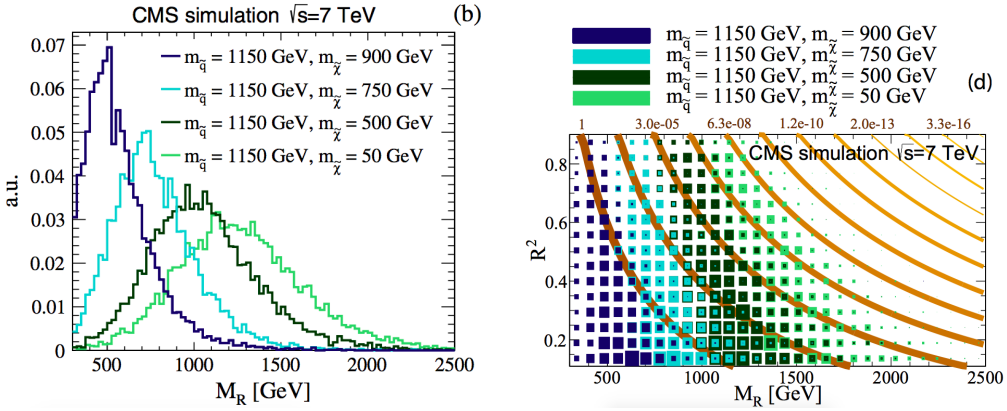


Figure 3.11: Left: M_R distributions for several squark production scenarios. Right: two-dimensional distributions of M_R and R^2 for the same signal models [95].

The above derivations assume that there are only two visible jets in the event. If there are more than two jets (or leptons), we enforce a two-jet event topology by clustering the observed physics objects (jets and leptons) into two 'megajets' (see Figure 3.12). Each object is assigned to one megajet or the other, and the four-momentum of the megajet is taken to be the sum of the four-momenta of the objects assigned to it. Assignment of objects to megajets is performed in such a way as to minimize the sum of the squared megajet masses,

$$m_{j1}^2 + m_{j2}^2. \tag{3.37}$$

This condition encourages objects near one another in η and ϕ to be assigned to the same megajet.

An inclusive SUSY search using razor variables was carried out on the full 8 TeV CMS dataset [71, 72]. It yielded limits on the masses of third-generation squarks and gluinos that are among the best obtained with the Run I data (see Figures 3.7 and 3.8).

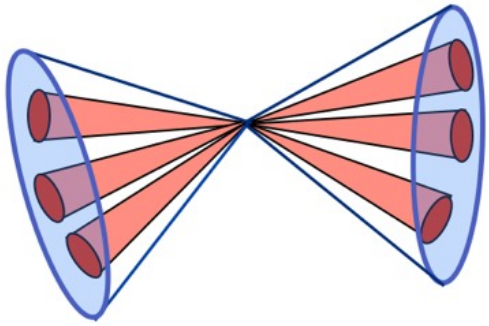


Figure 3.12: Diagram illustrating the clustering of jets into megajets for the calculation of M_R and R^2.

Part II:

Towards Trigger-Based Searches: Data Scouting

The beauty of the moment is a beauty sadly lost.

Belle and Sebastian, "We are the Sleepyheads"

Chapter 4

DATA SCOUTING TRIGGER DEVELOPMENT

In this chapter we describe *data scouting*, a paradigm for collider data analysis based on trigger-level event reconstruction. Data scouting complements the traditional analysis paradigm in which events are selected by the trigger and are sent offline to undergo an expensive reconstruction procedure. By taking advantage of the reconstruction that already takes place at the HLT level, we were able to use scouting to dramatically increase the number of CMS physics events stored for analysis while having negligible impact on available computing resources.

After motivating data scouting and describing the history of its use in CMS, we provide details on a framework for scouting that was designed and deployed for data taking during CMS Run II, and give information on the datasets collected using the framework. In the next chapter, we describe the impact of the scouting framework on the CMS physics program.

4.1 Motivation: limits to CMS event processing

The standard CMS framework for triggering on events and reconstructing them offline was described in Chapter 2. There are four primary constraints on the number of events that can be recorded using this procedure:

1. The data acquisition system (DAQ) of CMS has finite bandwidth. Restrictions on the data volume are imposed by the size of the temporary raw data storage at LHC Point 5, and by the bandwidth of the link between Point 5 and the CMS computing center at the main CERN site [96].

2. The prompt reconstruction system must be able to reconstruct all selected events in pseudo-real time without significant backlog [97]. It is desired that all physics data be reconstructed and available within 48 hours of being collected [98]. This constraint can be partially avoided using data parking, discussed in Section 4.5, but this requires a significant investment of computing resources to reconstruct the data later and also puts stress on the DAQ bandwidth.

3. The total amount of storage space for data is limited. As more data is taken, datasets must increasingly be kept in long-term storage on magnetic tape instead of on disk, which slows data access significantly. The cost of purchasing tape and disk storage must be considered when deciding how much data to record.

4. The average trigger decision at the HLT must be made within a few hundred milliseconds, as described in Section 2.3.

As a consequence of these constraints, CMS records events for physics analysis at a maximum rate of approximately 1 kHz, orders of magnitude more slowly than the rate (~30 MHz) at which pp collisions occur in the detector. It would be counterproductive to record every single collision event, because most do not contain 'interesting' physics and would not be used by any analysis. Nevertheless, the requirements of the trigger impose significant constraints on current searches and SM measurements. These constraints become more onerous over time as LHC luminosity rises.

For example, many searches for new physics in hadronic final states rely on H_T triggers, which select events based on the value of

$$H_T \equiv \sum_{i=0}^{N_{\text{jets}}} |\vec{p}_{iT}| \,, \tag{4.1}$$

where $\vec{p}_{iT}$ is the transverse momentum of the ith reconstructed jet. The H_T variable is a proxy for the amount of hadronic activity in the event and is often used as a SUSY search variable. In Run II, the loosest H_T triggers in the HLT menu select events having $H_T > 800$ or 900 GeV. Events with H_T below the threshold cannot be recorded unless they present some other feature of interest. Many SUSY searches, for example, rely on triggers that select events with large values of both H_T and E_T^{miss}. The additional requirement of E_T^{miss} allows the H_T threshold to be lowered while keeping the rate under control.

The razor SUSY searches discussed in Part III are also constrained by trigger considerations: the thresholds of the hadronic razor triggers described in Section 6.4 are driven by HLT rate and timing restrictions.

4.2 The data scouting paradigm

As mentioned in Section 2.3, the HLT performs reconstruction algorithms similar to those used offline. This includes a version of the full PF algorithm and its components: track finding, clustering of calorimeter hits, and muon, electron, photon, and hadron identification. The good performance of the physics objects produced by these algorithms suggests a new strategy for analyzing CMS data:

1. Reconstruct events at the HLT level during data taking by running the PF algorithm or other reconstruction routine.
2. Apply a loose selection on the reconstructed physics objects. For each event passing the loose selection, save the trigger-reconstructed physics objects to disk.
3. (Optionally) discard the raw detector data.
4. Perform searches for new physics, or SM measurements, using the saved HLT-level events.

We refer to this strategy as data scouting.

In the data scouting paradigm, trigger-based algorithms are used as the basis for event interpretation. This approach neatly avoids three out of the four constraints listed above that restrict the number of events that can be recorded:

1. While the full raw data for a CMS event is of order 1 MB in size [97], the physics objects reconstructed by the HLT can be represented using only a few kB of memory. Trigger-reconstructed events can therefore be recorded at rates of several kHz and still not occupy more DAQ bandwidth than a single ordinary HLT path (of which there are hundreds in the menu).
2. The reconstruction is carried out at trigger level, which entirely removes the need for prompt reconstruction for the selected events. This means that the limit on event processing time from the offline reconstruction system is not relevant.
3. If the raw data is discarded, storage space is only needed for the reduced dataset consisting of the reconstructed trigger objects. This requires a negligible amount of disk resources compared to storing full events.

Data scouting is implemented via HLT paths that place events into dedicated scouting data streams (see Section 4.4). The output of these streams bypasses the prompt reconstruction system, instead packing the reconstructed trigger objects into a special scouting format, which is saved to disk.

The HLT CPU resource constraint

The fourth constraint listed above, from the limited processing power of the HLT farm, is not eliminated in the data scouting picture. Naively this constraint is even more severe for scouting because we rely on the HLT to perform the event reconstruction. However, the design of the HLT makes it possible to save a very large number of data scouting events without any further investment of CPU resources at all.

The key to this is the way HLT paths are structured. A typical path will perform several stages of reconstruction in sequence, with each stage more detailed than the last (see Figure 4.1). After each stage of the reconstruction, a cut is applied, and failing events are rejected with no further processing. After the last reconstruction sequence has been run, a final selection is made to determine if the event passes the trigger. Thus, even events that do not pass the trigger may pass through several stages of reconstruction before being rejected. The purpose of this design is to remove unwanted events as early as possible, to avoid wasting resources.

Figure 4.1: Typical HLT path structure. Producers perform physics object reconstruction using the raw event data as input. Filters reject the event if it does not meet specified requirements. A real HLT path may proceed through many stages of production and filtering operations before its final trigger decision is made (only two such stages are depicted here).

The HLT software ensures that physics objects reconstructed by one trigger path are made available for use by all other paths. For example, if one path runs the HLT version of the PF algorithm (which is very expensive, taking on the order of 1s to run), other paths may use the reconstructed PF objects in their trigger decision for that event without running the algorithm again.

These facts make it possible for data scouting to be conducted ‘in the shadow’ of existing trigger algorithms. A data scouting trigger can take physics objects

reconstructed in intermediate or final steps of a different trigger path and write them directly to disk. Because most HLT paths have a tight cut at the end to keep the rate low, large numbers of events are reconstructed at the HLT but not selected for offline analysis. It is possible to save scouting events at a rate of several kHz using only existing physics objects from events that are rejected by other paths.

4.3 History of data scouting in CMS

The idea of data scouting was conceived in Run I of CMS, and the technique was used to perform searches for exotic resonances decaying to dijets. Dijet resonance searches tend to be severely constrained by trigger requirements: events with two back-to-back jets are ubiquitous at hadron colliders and a trigger that recorded all of them would have prohibitively high rate.

In CMS, dijet searches traditionally use H_T triggers, which were discussed in Section 4.1. The H_T trigger threshold rises with increasing LHC energy and luminosity, and this implies that the lowest resonance mass the search can probe is pushed higher and higher over time.

This challenge is illustrated by the green and light blue curves in Figure 4.2, which indicate exclusion limits from the LHC experiments on production of a hypothetical Z' resonance decaying to quarks. The LHC sets state-of-the-art limits at high Z' mass, but at lower masses the limits from standard offline dijet resonance searches are nonexistent (for CMS) or worse than those from the Tevatron (for ATLAS, who use prescaled trigger paths to extend their dijet search to lower masses). The ATLAS limit (light blue curve) provides an explicit demonstration of the worsening of the limits due to trigger restrictions. Their limits at lower masses rely on data from triggers with increasingly high prescales, which are needed in order to keep the rate low. The trigger with the lowest threshold (selecting jets with p_T between 59 and 99 GeV) has a prescale factor of 460000 [99].

Data scouting provides dijet searches with relief from the rising H_T trigger rates. This was first demonstrated in 2011, when a data scouting trigger path was deployed in CMS for a few hours of data taking at 7 TeV. The trigger had a L1 requirement of $H_T > 100$ GeV, and it selected all events having either $H_T > 350$ GeV or dijet mass greater than 400 GeV at the HLT level. The only data recorded for each event was the collection of jets reconstructed with the HLT PF algorithm. These few hours' worth of data, collected at the very end of the 2011 run period and corresponding to 0.13 $\text{fb}^{-1}$of integrated luminosity, were sufficient to improve existing limits on

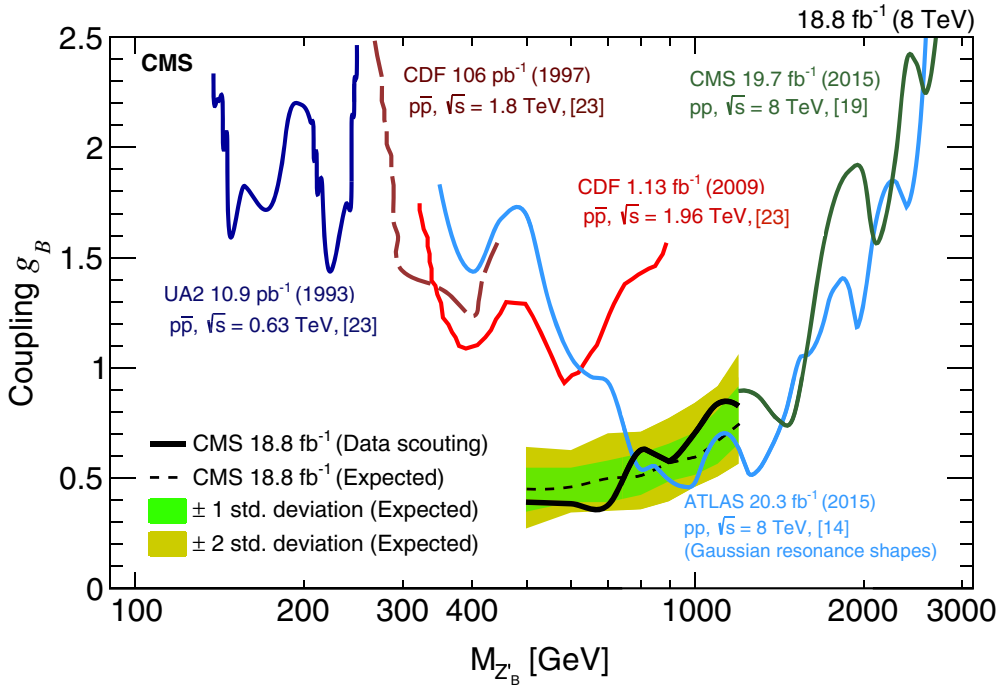


Figure 4.2: Limits on the coupling of a hypothetical leptophobic Z' resonance as a function of the resonance mass, with results from a variety of hadron collider experiments. The green and light blue curves are the results from the CMS and ATLAS 2012 dijet resonance searches, respectively. The black curve and the green and yellow bands indicate the observed and expected limits from the CMS low-mass resonance search using data scouting [100].

a number of resonance models in the range 0.6-0.9 TeV [101]. The dijet mass spectrum observed in this search is displayed in the top part of Fig. 4.3.

After this success, a second scouting trigger was designed and deployed at the HLT for most of the 2012 CMS data taking period. The trigger required $H_T > 250$ GeV and had a maximum rate of 1 kHz. This rate was too high for the PF algorithm to be run for every event, so the trigger instead reconstructed and saved calorimeter jets ('calo jets'), which are clustered directly from energy deposits in the ECAL and HCAL. Calo jets require negligible HLT resources to reconstruct, and at high momentum their mass resolution is adequate despite the lack of tracking information. A dijet search was again conducted using the scouting dataset, and the resulting limits on massive Z' resonances were the best to date between 500 and 800 GeV. These limits are indicated by the black line in Fig. 4.2. The dijet mass spectrum from this search is shown in the bottom part of Fig. 4.3.

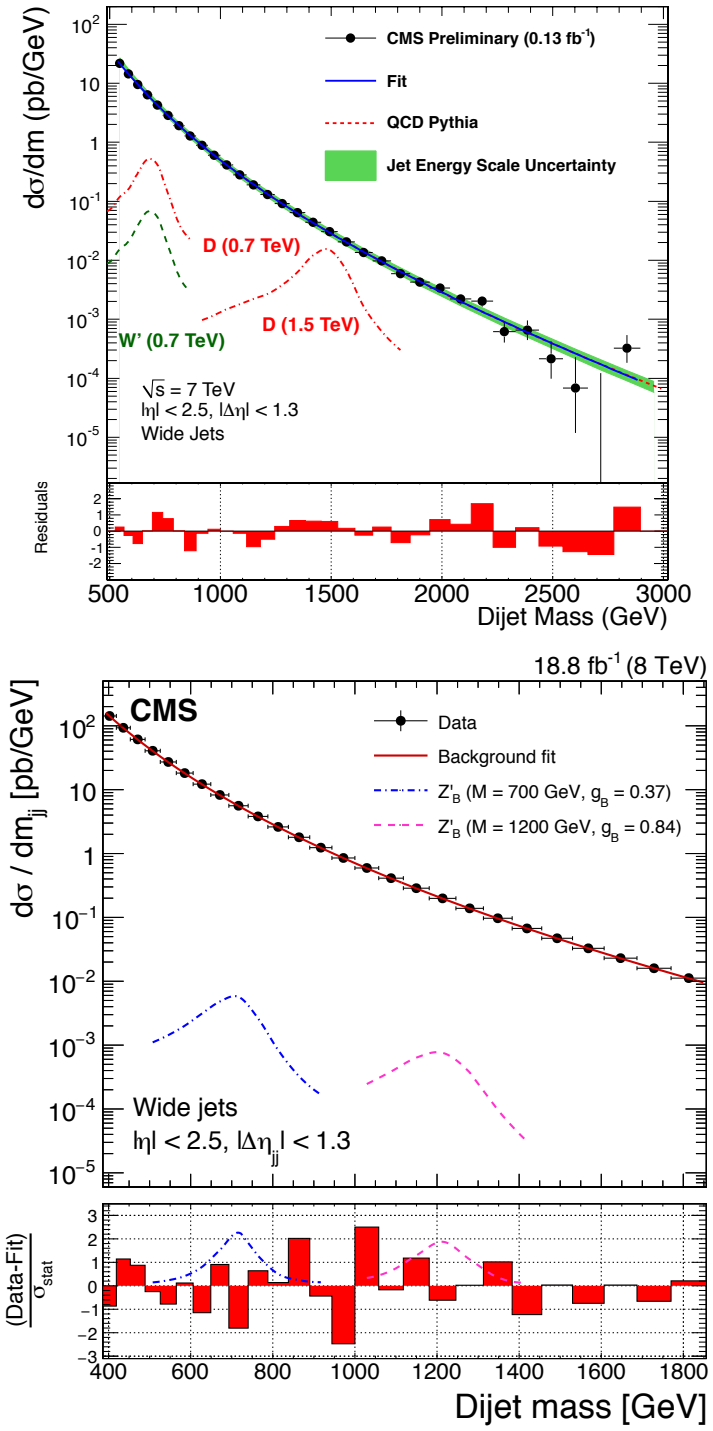


Figure 4.3: Dijet mass spectra obtained in CMS searches using data scouting in 2011 [101] (top) and 2012 [100] (bottom). For each spectrum, a fit to a predefined functional form is displayed, along with the significance of the data's deviation from the fit (bottom panel of each plot).

4.4 Design of a multipurpose scouting framework for Run II

For Run II of CMS it was desired to build a more comprehensive software framework for data scouting that would be enabling for a variety of physics analyses. The main components of such a framework are:

- HLT data streams containing 'scouting trigger paths' that define the events to be selected for data scouting
- A common data format for the scouting events that is lightweight yet holds all necessary information reconstructed by the HLT
- A mechanism for monitoring the quality of scouting data and comparing with offline reconstructed data.

In the following subsections we will give details on each of these aspects for the framework that was implemented and deployed in CMS starting in 2015.

Scouting data formats

The experiments with data scouting in Run I suggest the following strategy for collecting as much scouting data as possible:

1. Run the HLT PF algorithm and record PF jets and particle candidates at the maximum attainable rate;
2. For events in excess of this rate, reconstruct and record calo jets instead.

This dual strategy was adopted for Run II. Because the event information provided by the PF algorithm is different from (and much larger than) that provided by calo jet reconstruction, two scouting data formats were designed, one for each flavor of reconstruction. The data formats are denoted the *calo-scouting* and *PF-scouting* event formats.

The data formats are designed to be as lightweight as possible. They consist of C++ built-in data types (floats, integers, bools, and vectors) wrapped in classes compatible with the CMS event processing software. This ensures that future versions of the CMS software and the ROOT library can read the event format, which is important for the future use and preservation of these datasets. The event content is also modular, so that object collections can be omitted if they were not reconstructed in the event.

Calo-scouting event format

The event format for scouting with calo jets is depicted on the left side of Figure 4.4. It consists of the following components:

- Calo jets objects, which hold the momentum vectors of the reconstructed calo jets as well as auxiliary information, such as the fraction of the jet energy contained in the ECAL and in the HCAL.
- The magnitude and angle of $\vec{p}_T^{miss}$, computed using calorimeter-level quantities.
- The value of ρ, the median p_T density in the event, computed using calorimeter-level information.
- Primary vertex objects, each holding the coordinates and associated uncertainties of a reconstructed primary vertex.

Events with $H_T > 250$ GeV saved in this format have an average size of approximately 1.5 kB.

In the Run I data scouting analysis using calo jets, there was no dedicated scouting event format: the CMS software objects corresponding to the reconstructed calo jets were saved with no reformatting. The average event size in that analysis was 10 kB [100]. The much smaller event size under the new Run II framework indicates that repacking the jets into dedicated scouting jet objects significantly decreases the storage overhead.

PF-scouting event format

The event format for scouting with PF objects is shown on the right side of Figure 4.4. It consists of the following components:

- The collection of PF candidates with $p_T > 0.6$ GeV. Each PF candidate is represented by its 4-momentum and an integer indicating its particle type.
- PF jet objects, which hold the momenta and identification variables of jets clustered from the PF candidates.

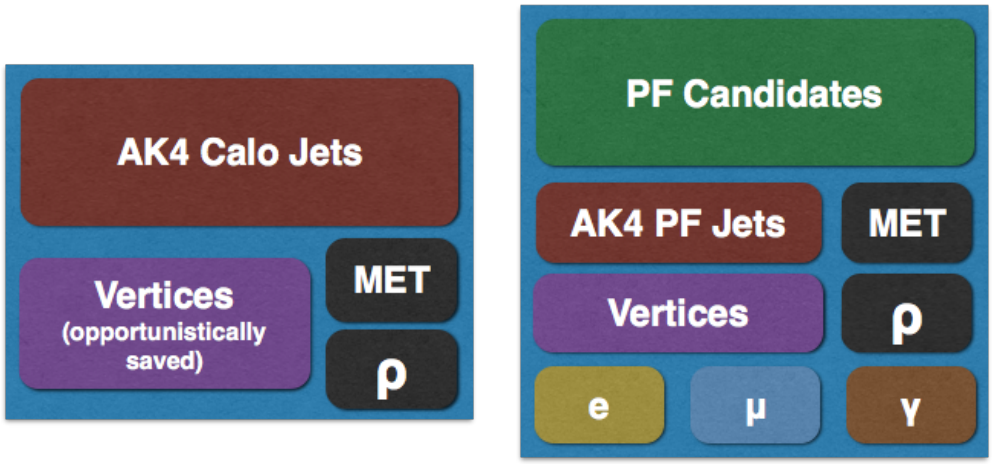


Figure 4.4: Schematics of the calo-scouting (left) and PF-scouting (right) event formats. 'MET' refers to $\vec{p}_T^{miss}$, ρ is a measure of the median p_T density in the event, and 'AK4' indicates the anti-k_T algorithm with $R = 0.4$.

- Primary vertex objects, identical to those included in the calo-scouting event content.
- The magnitude and angle of $\vec{p}_T^{miss}$, and the value of ρ.
- Electron, muon, and photon objects, each associated with a 4-momentum vector and a collection of identification variables.

The typical PF-scouting event size is 10 kB, most of which is occupied by the PF candidate objects. The inclusion of the PF candidates allows for more complex analysis strategies involving, for example, jet substructure variables computed using the constituents of the jets.

HLT data streams for scouting

Two scouting data streams were implemented in the HLT software, referred to as the PF-scouting and calo-scouting streams. Each stream has its own set of primary datasets and trigger paths, and its own output data format as discussed above.

Scouting trigger paths have the same basic structure as standard HLT paths (see Figure 4.1), except that after the last stage of event reconstruction there is no final event selection filter. Instead, the reconstructed physics objects are passed to a

software module that packs them into one of the special scouting formats. The objects are then written to disk in files grouped by PD.

The hadronic scouting triggers designed and deployed in 2015 were inspired by the scouting triggers used in Run I. They select events based on the value of calorimeter-level H_T, using H_T thresholds significantly lower than those of the paths in the standard HLT menu. The triggers are seeded by L1 triggers that select events based on the value of H_T reconstructed at the L1. The thresholds of the H_T scouting triggers are illustrated in Figure 4.5, and more information about the choice of thresholds is provided in the following subsections.

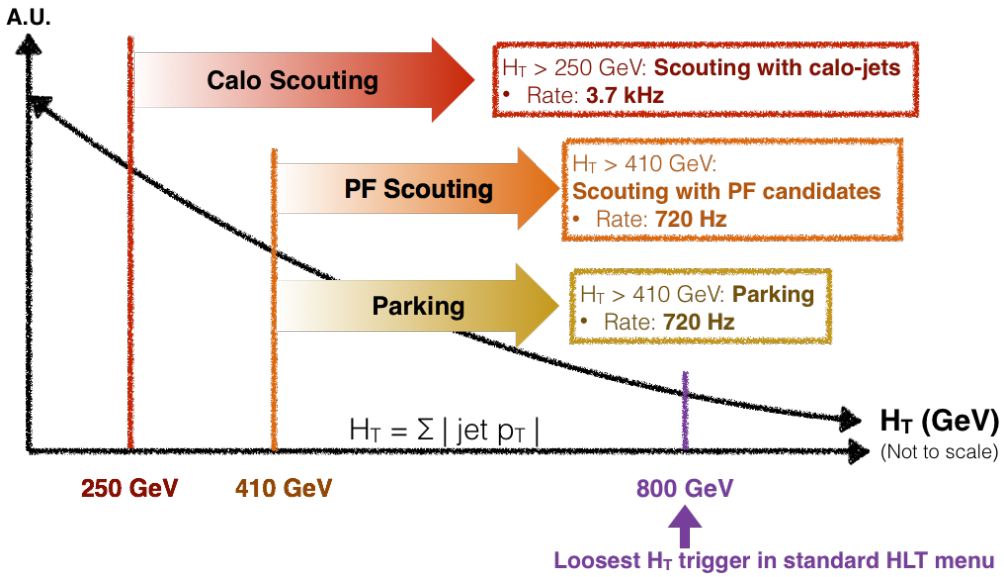


Figure 4.5: Diagram illustrating the H_T thresholds of the hadronic scouting and parking triggers deployed in Run II. Note that the thresholds indicated for the PF-scouting and parking triggers correspond to the triggers used in 2016; in 2015 the thresholds were set to 450 GeV. The H_T trigger in the standard HLT menu (indicated in purple for comparison with the scouting triggers) was tightened from 800 GeV to 900 GeV in 2016.

An important idea in trigger design is that of a trigger *turn-on* curve, which quantifies the fraction of desired events that the trigger successfully selects. Turn-on curves will be shown later in the context of a dijet resonance search with data scouting. Additional triggers are included in the scouting data streams to facilitate the measurement of turn-on curves; these are paths with looser selection requirements that provide a baseline for the efficiency measurement. The triggers used to measure turn-on curves are prescaled by large factors to avoid excessive event rates.

The PF-scouting stream

The list of trigger paths in the PF-scouting stream is provided in Table 4.1. The rate at which these paths select events is low enough that the PF algorithm can be run for each event. The reconstructed PF objects are used to produce output events in the PF-scouting format described above. In addition to the standard PF sequence, additional reconstruction sequences are run to compute isolation sums and other needed identification variables for muons, electrons, and photons.

The triggers in the PF-scouting stream each come in two versions: one that performs b-tag reconstruction and one that does not. The two triggers in each pair have identical event selection requirements. The b-tagging version of each trigger runs the CSVv2 algorithm (see Section 2.4) to predict whether each reconstructed jet is a b-jet. This functionality is placed in a separate path for a technical reason: the HLT b-tag reconstruction sequence contains filters that reject some events (specifically those that have no jets or no reconstructed vertices). By putting the b-tagging sequence outside the main trigger path, we avoid spuriously rejecting these events. The rates in Table 4.1 refer to the version of each path that does not run b-tagging; the rate of the b-tagging version is generally slightly lower due to the filters associated with the b-tag sequence.

To facilitate the measurement of trigger turn-on curves for the scouting triggers, additional paths are included that have the same L1 trigger requirements as the main scouting paths, but no HLT requirement at all. These L1-only triggers provide a baseline for measuring trigger efficiencies. That is, they allow the calculation of the trigger efficiency as

$$\epsilon_{HLT} = \frac{\text{\# passing L1 and HLT}}{\text{\# passing L1}}. \tag{4.2}$$

This measurement is typically performed in bins of a kinematic variable of interest, such as H_T or the invariant mass of the two leading jets. Because of the extremely high rate of the L1 seeds, the L1-only scouting triggers are prescaled. The prescale factor is chosen to achieve a rate of order 10 Hz per path.

In fact, the triggers in this stream have event selection thresholds so low that the L1 trigger paths seeding them may not be fully efficient for the selected events. That is, there are events that would be selected by the scouting HLT paths but which are prematurely rejected at the L1 stage. To quantify this effect, an additional minimum-bias trigger path is included that has a pass-through L1 seed (which is

heavily prescaled but accepts every event it sees). For this trigger, an extremely loose selection is applied at the HLT level that demands only the presence of a single jet in the event. This trigger can be used in conjunction with the L1-only triggers just described to obtain a measurement of the L1 seed efficiencies:

$$\epsilon_{L1} = \frac{\text{\# passing L1 and min-bias}}{\text{\# passing min-bias}}. \tag{4.3}$$

The PF-scouting stream also contains a trigger that selects events having two muons, with no requirement on jet activity in the event. The muons are required to have $p_T > 3$ GeV and to have di-muon invariant mass $m_{\mu\mu} > 10$ GeV. A L1-only muon trigger path, and a zero-bias path accepting every event it sees, are added to the stream to assist with efficiency measurements for this trigger. The $\mu\mu$ scouting trigger was implemented as a proof-of-concept of a non-hadronic scouting trigger. A new di-muon scouting trigger was developed for the 2017 run, motivated by the use case of a search for dark photons. This trigger is described in Section 4.7.

Trigger	Rate in 2015 (Hz)	Rate in 2016 (Hz)	Rate in 2017 (Hz)	Notes
$H_T >$ 410 GeV	-	750	740	Deployed in 2016
$H_T >$ 450 GeV	160	500	-	Removed in 2017
$m_{\mu\mu} >$ 10 GeV	200	530	-	Removed in 2017
L1 H_T	7	8	40	
L1 di-muon	23	4	10	
Min-bias	3	5	30	
Zero-bias	16	9	10	

Table 4.1: List of paths in the PF-scouting stream, with typical rates in the 2015, 2016, and 2017 LHC runs. Rates are computed for instantaneous luminosities of 5×10^{33}, 1×10^{34}, and 1.5×10^{34} cm^{-2}s^{-1} respectively for 2015, 2016, and 2017. The rates for the L1-only and min/zero-bias paths are controlled by prescale factors that change from year to year.

The calo-scouting stream

The list of trigger paths in the calo-scouting stream is provided in Table 4.2. Triggers in this stream reconstruct calo jets, which are used to produce output events in the calo-scouting format described above. Because it is virtually costless to reconstruct calo jets, the rates of these triggers can be much higher than those of the PF-scouting triggers.

The calo-scouting triggers do not perform primary vertex reconstruction. However, primary vertices can be saved as part of the calo-scouting event content if they are

available in the event. When another HLT path runs primary vertex reconstruction (usually as part of some other physics reconstruction sequence), the vertex objects are saved in the output scouting event. Having vertex information available in a fraction of the saved events can be useful in analysis. For example, it can be used to monitor changes in physics object traits as a function of the number of primary vertices.

Calo jets are reconstructed without using information from the CMS tracker, so b-tag information cannot be obtained for them with the usual HLT CSV sequence. Instead, a modified version of the b-tagging sequence is deployed. This method begins by performing the fast primary vertex reconstruction described in Section 2.3. Regional tracking is then performed using only hits consistent with the identified primary vertex and with the locations of the 8 highest-p_T jets in the event. This collection of tracks is used as input to the CSVv2 algorithm for b-jet identification [23]. As in the PF-scouting triggers, the b-tagging sequences in calo-scouting need to be placed in separate paths in order to avoid unnecessarily rejecting events.

Regional tracking can also be used to reject calo jets originating from pileup or noise. Pixel tracks are reconstructed inside of selected calo jets, and the variable

$$r_{\text{tracks}} \equiv \frac{\sum_{\text{tracks}} p_{T,\,\text{track}}}{p_{T,\,\text{jet}}}, \tag{4.4}$$

i.e., the ratio of the track momentum to the jet momentum, is computed. The distribution of this ratio is shown in Figure 4.6 for signal and pileup jets. Using r_{tracks} allows one to reject 75% of pileup jets while retaining 95% of signal jets [102]. This variable is saved in the calo-scouting events for use by analyzers.

Pixel track reconstruction and calo jet b-tagging were not included in the scouting framework for the 2015 LHC run. They were added in 2016, after the successful use of the 2015 calo-scouting data by the dijet resonance search team (described in the next chapter).

As for the PF-scouting stream, the calo-scouting stream contains additional triggers to assist in the measurement of trigger turn-on curves. These are exact copies of the L1-only and minimum-bias trigger paths described in the previous section, except that they reconstruct calo jets instead of running the full PF sequence. Because they are the same as those listed in Table 4.1, they are omitted from Table 4.2.

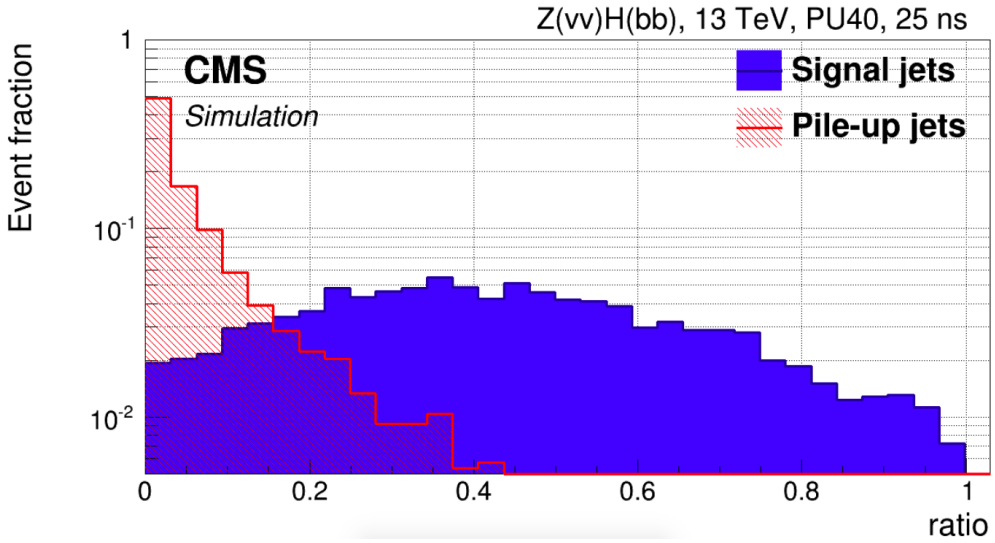


Figure 4.6: Distribution of the quantity r_{tracks} for signal (blue) and pileup (red) jets [102].

Trigger	Rate in 2015 (Hz)	Rate in 2016 (Hz)	Rate in 2017 (Hz)	Notes
$H_T > 250$ GeV	1500	4000	3900	
Di-muon trigger	-	-	1900	Deployed in 2017

Table 4.2: List of paths in the calo-scouting stream, with typical rates in the 2015, 2016, and 2017 LHC runs. Duplicates of triggers in the PF-scouting stream (Table 4.1) are not shown. Rates are computed for instantaneous luminosities of 5×10^{33}, 1×10^{34}, and 1.5×10^{34} cm^{-2}s^{-1} respectively for 2015, 2016, and 2017.

Scouting data monitoring

A third stream is implemented at the HLT with the purpose of monitoring the quality of the scouting triggers and events. This monitoring stream is a hybrid of a standard physics stream and a data scouting stream. Selected events undergo standard prompt reconstruction and are also saved in the scouting format. The dual output allows for side-by-side comparison of the scouting physics objects with standard reconstructed objects. It can be used to calibrate the energies and identification efficiencies of the trigger objects, and also to identify any problems with the stream configurations or the event packing software.

The scouting monitor stream contains copies of all of the triggers in the PF-scouting and calo-scouting streams. Because the output undergoes prompt reconstruction, the rate of the stream must be kept low. Prescales on the triggers in the scouting monitor stream are chosen in order to keep the total rate of the stream below about 30 Hz. The stream can then be used to study the scouting data without having a

major impact on the HLT data volume.

4.5 Data parking and its relationship with scouting

The term *data parking* refers to the practice of selecting events at the HLT and immediately moving them to tape storage, skipping prompt reconstruction. Events selected in this way ('parked data') remain on tape until there are sufficient free computing resources to reconstruct them.

Data parking allows more than the standard 1 kHz of physics events to be recorded, because the rate is not constrained by the limited capacity of the prompt reconstruction system. The achievable rate of data parking is constrained by the bandwidth of the CMS DAQ, by the amount of tape storage space, and by the practicality of eventually bringing the data out for reconstruction.

Data parking was a central aspect of the CMS data taking strategy in 2012. A total of 300-350 Hz of data was collected using parking triggers [98]. The downsides of parking became apparent: the parked datasets were large and took many months to reconstruct fully with the resources available. By the time the reconstructed data was ready, most analyses of 2012 data had already been published and attention had moved to preparation for Run II.

Despite its disadvantages, data parking fits naturally into the data scouting paradigm, especially in the context of searches for new physics. A main drawback of data scouting is the discarding of the raw data and the reliance entirely on HLT-reconstructed physics objects, which may suffer more from detector noise and miscalibration than the standard PF objects. If a search for new physics is performed on scouting data and returns a positive result, it may be difficult to find out whether the supposed new physics is real or the result of noise that affected the HLT reconstruction in some way. Parking the raw events selected by the data scouting triggers removes this handicap: if a physics result with the scouting data is called into question, the parked data can be brought from tape to disk and reconstructed. The analysis can then be performed again using the parked data to confirm or disconfirm the result. In case the scouting analysis returns an unambiguous result, the parked data can stay on tape forever, sparing the resources that would be needed to reconstruct it.

This strategy was deployed in CMS in 2015 and 2016 to complement the data scouting triggers. It was estimated that the DAQ could handle 600 Hz of parked data safely. A suite of parking triggers, described in Table 4.3, was created, along with a

new data stream to hold them. These triggers select the same events as the triggers in the PF-scouting stream, dividing them among a number of parked datasets. The H_T parking triggers each select a slice of the H_T spectrum. Each of these slices is sent to a separate parked dataset. The stream is designed in this segmented way for three reasons:

1. To prevent any one parked dataset from being too large,
2. To allow individual slices in H_T to be reconstructed if needed,
3. To allow some slices in H_T to be disabled without shutting down the whole stream if the rate of parking becomes too large.

Trigger	Rate in 2015 (Hz)	Rate in 2016 (Hz)	Notes
$410 < H_T < 430$ GeV	-	130	Disabled in July 2016
$430 < H_T < 450$ GeV	-	100	
$450 < H_T < 470$ GeV	25	83	
$470 < H_T < 500$ GeV	30	95	
$500 < H_T < 550$ GeV	35	110	
$550 < H_T < 650$ GeV	35	110	
$H_T > 650$ GeV	30	92	

Table 4.3: List of H_T parking triggers, with typical rates in the 2015 and 2016 LHC runs. Rates are computed for instantaneous luminosities of 5×10^{33} (2015) and 1×10^{34} cm^{-2}s^{-1}(2016). The triggers with $410 < H_T < 450$ GeV were not present during the 2015 run. In 2015 the $m_{\mu\mu} > 10$ PF-scouting trigger was also used for parking.

4.6 Scouting and parking evolution in 2015-2017

The scouting framework was deployed successfully in 2015. The H_T PF- and calo-scouting triggers each collected 1.8 fb^{-1} worth of data that year. The event sizes were as expected, and data volumes placed no significant strain on the HLT resources.

In 2016, the scouting framework was deployed again with various improvements. The calo-scouting event content was enhanced with b-tagging and pixel track reconstruction, as discussed above. The threshold of the H_T PF-scouting trigger was decreased from 450 to 410 GeV, and analogous triggers were added to the parking stream. In 2016 scouting data accounted for approximately 1 percent of the total volume of data recorded by CMS. The rate of events recorded went as high as 6 kHz, many times that of the standard HLT physics streams.

The rate of data parking had to be slowed during the second half of 2016 due to the unexpectedly high luminosity provided by the LHC. This was accomplished by disabling the parking trigger selecting events with H_T between 410 and 430 GeV, the lowest part of the segmented range.

The major changes for the scouting program in 2017 were the deployment of the new di-muon trigger (described in Section 4.7), and the removal of the segmented H_T parking triggers that existed in previous years. The parking triggers were replaced by prescaled versions of the scouting triggers, which recorded 10% of events to tape for monitoring purposes. Additionally, new L1 seeds were added to the hadronic scouting paths in response to the rising threshold of the L1 H_T seed. The new L1 seeds include triggers on the p_T of the leading jet in the event and on the invariant mass of the two leading jets. The latter was made possible by the upgrade of the L1 trigger infrastructure, which allowed paths with more complex logic to be implemented in the L1 Global Trigger (see Section 2.3).

Estimated and actual trigger rates

MC-based rate measurements

The scouting framework was designed in late 2014 and early 2015, before CMS had ever collected data at its maximum energy of 13 TeV. This made it necessary to estimate the rates of the proposed scouting and parking triggers using MC simulation. A set of datasets of simulated QCD multijet events were created for this purpose.

A difficulty of using MC for rate measurements is the enormous range of energies obtained in LHC collisions. The lowest-energy events are overwhelmingly common but have very little chance of even passing the L1 trigger. Thus, simulating pp events completely at random is unlikely to lead to adequate statistics for high-energy collision events, which are the events of interest for the trigger study. This is addressed by simulating several QCD samples, each with prescribed upper and lower bounds on the amount of hadronic activity. Together, the samples cover the entire range of event energies in a non-overlapping way. The number of events in each sample is roughly the same, so samples with large hadronic activity have more events relative to their cross section than those with low activity. This approach leads to small statistical uncertainties on the estimated trigger rates.

The proposed triggers are emulated using the HLT software to obtain a trigger de-

cision for each simulated event. The rate of each trigger is estimated as

$$\text{rate} = \sum_{i=0}^{N_{\text{samples}}} \mathcal{L} \times \sigma_i \times \frac{\text{\# selected events in sample } i}{\text{\# events in sample } i}, \tag{4.5}$$

where σ_i is the cross-section associated with the ith QCD MC sample. The statistical uncertainty on the rate measurement is simply given by the Poisson uncertainty on the number of counts.

The actual hadronic scouting trigger rates recorded in 2015 (listed in Tables 4.1 and 4.2) were lower than anticipated based on the simulation by approximately a factor of two. This indicates a possible miscalibration of the MC simulation for QCD events. The effect was observed for standard HLT paths as well, such as those used to select hadronic events for the inclusive razor analysis (see Part III).

Data-driven rate measurements

Preparation of the data scouting streams for the 2016 CMS run was made easier by the availability of the 2015 data. In particular, CMS recorded a limited number of high-pileup runs in 2015. By measuring trigger rates in runs with varying amounts of pileup, one can predict the rates in 2016 data, which was predicted to (and did) have significantly higher pileup than 2015 data.

Data-driven trigger rates can be obtained using unbiased samples of HLT events. These are collected by a prescaled HLT path that automatically accepts every event coming from the L1 trigger. After emulating the trigger decision for the scouting triggers on each event in this sample, the rate for each path is computed as

$$\text{rate} = \frac{\text{\# passing trigger}}{\text{\# in sample}} \times \text{rate of unbiased trigger} \times \text{prescale of unbiased trigger}. \tag{4.6}$$

The rate must then be scaled to the number of colliding bunches anticipated during data taking (the high-pileup runs in 2015 were taken with fewer bunches than were expected in 2016).

Figure 4.7 shows the rates estimated for the 2015 $H_T > 450$ GeV trigger and for emulated versions of the trigger with lower thresholds, including the $H_T > 410$ GeV trigger that was eventually deployed in 2016. For the $H_T > 450$ GeV path, rates were taken directly from the CMS central monitoring system. For the emulated triggers, they were computed using Eq. 4.6. The rates scale roughly linearly with the

number of pileup collisions. To determine the average number of pileup collisions in a given LHC run period, one can use the relation

$$N_{\text{pileup}} = \frac{\mathcal{L}\sigma_{inel}}{N_{\text{bunches}} f}, \tag{4.7}$$

where σ_{inel} is the inelastic pp cross section, N_{bunches} is the number of colliding bunches per LHC beam, and f is the beam revolution frequency (11 kHz).

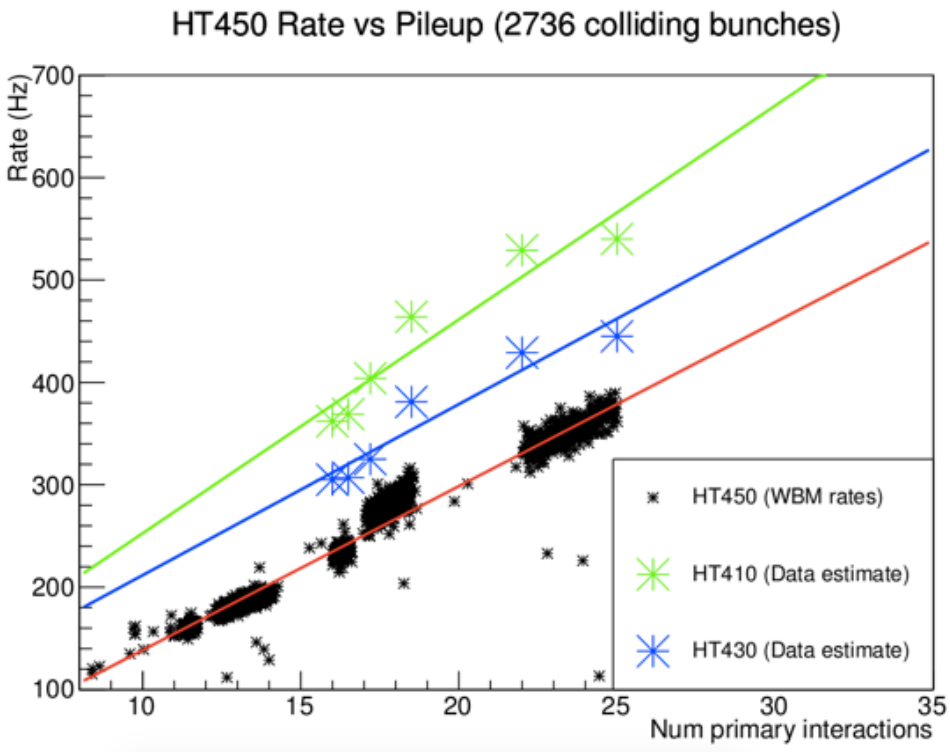


Figure 4.7: H_T scouting trigger rates scaled to 2736 colliding bunches, for different average numbers of pileup interactions. Rates for the $H_T > 450$ GeV trigger are taken directly from 2015 CMS data, while those for $H_T > 430$ GeV and $H_T > 410$ GeV triggers are obtained by emulating the triggers on data. Linear extrapolations to higher numbers of pileup interactions are shown for each trigger.

The data-driven trigger rate measurements accurately predicted the observed 2016 rates listed in Table 4.1.

Trigger timing measurements

In Section 4.2 it was argued that data scouting could be used to collect significant quantities of data without much impact on HLT CPU resources. This was checked explicitly using computers at CERN dedicated to benchmarking the timing performance of HLT paths [21].

Trigger timing can be measured using the unbiased samples of HLT events discussed above. The HLT software can be outfitted with profiling modules that determine the time taken by each producer and filter in the path of interest. To obtain the average trigger time per event, one emulates the trigger on the unbiased HLT data sample with these profiling modules enabled. The time can be measured for the path running either by itself or in the context of the full HLT menu. In the latter case, the measurement indicates the average amount of time added to the full menu. To avoid placing strain on HLT resources, it is desirable for a new path to add no more than a few milliseconds of processing time to the menu.

Table 4.4 shows the estimated CPU time for H_T scouting triggers with different thresholds, as measured on 2015 data. The values indicated that the trigger threshold could safely be lowered from its 2015 value of 450 GeV to 410 GeV without adding signficant time to the menu. It was suggested to lower the threshold even further, but this was not done, for three reasons:

- The upgrade of the L1 trigger was in progress and the precise impact of the new L1 on the HLT menu time was unknown,
- The HLT menu time depends significantly on the amount of pileup, and detailed timing estimates were not obtained for high-pileup scenarios.
- It was desired to keep the trigger threshold for PF-scouting synchronized with that for parking.

This choice turned out to be appropriate given the very high pileup levels attained in 2016 and 2017. In 2017, at 1.5×10^{34} cm^{-2}s^{-1}, the PF-scouting H_T path was estimated to contribute around 3 ms to the running time of the HLT.

H_T threshold (GeV)	Total path time (ms)	Estimated time added to menu (ms)
410	5.4	0.59
420	5.3	0.57
430	5.2	0.48
440	5.0	0.47
450	4.6	0.45

Table 4.4: Trigger timing measurements for H_T PF-scouting paths using 2015 data. Measurements are performed on events with an average of 21-24 pileup interactions. The time added to the full menu is estimated using an early version of the 2016 trigger menu.

The calo-scouting H_T triggers, which cheaply reconstruct calorimeter jets, require a negligible investment of CPU time. The addition of the regional pixel tracking and b-tagging, which was done for the 2016 run, added 1-2 ms of processing time on average per event. The impact of this change on the total HLT timing was negligible, because many other HLT paths were already running the sequences in question.

4.7 Design of a new di-muon scouting trigger for 2017

The success of the data scouting framework in 2015 and 2016 prompted interest from data analysts seeking to expand the range of new physics searches possible with CMS data. One possibility that was pursued was to use data scouting to collect events with two muon candidates and to perform a search for dark photons decaying to muons. More information on the physics case for this is included in the next chapter.

To build a scouting trigger for a dark photon search, the di-muon trigger used in 2016 had to be modified significantly. The changes were intended to loosen the kinematic selection on the muons, enabling muon pairs with low p_T, low invariant mass, and possible displacement from the beamline to be selected. The major changes were:

- The L1 requirement was loosened substantially. The 2016 muon scouting trigger's L1 seed imposed p_T thresholds on both leading and subleading muons. It was supplemented in 2017 with a large number of new L1 seeds, including some with no p_T requirements; these instead require the muons to be close together in η and to lie within the CMS barrel. Some L1 seeds place an opposite-charge requirement on the muon pair.
- The di-muon invariant mass cut was removed.
- Quality requirements on the di-muon vertex were removed, enabling the selection of muon pairs from displaced decays. This required the standard HLT muon reconstruction sequences to be modified to remove or loosen several internal filters.

The muon objects used in the scouting data format were modified to include more detailed information about the muon's associated track (namely the values and uncertainties of the five parameters defining the track). Additionally, scouting vertex

objects were outfitted with a more complete set of uncertainty information. Secondary vertices from displaced muon decays were reconstructed by the trigger path and saved in the scouting event content. These additional variables are needed in the context of the dark photon search to select well-reconstructed muon pairs.

Finally, to handle the increased rate resulting from loosening the L1 requirement and removing the di-muon mass cut, the modified trigger path was moved from the PF-scouting stream into the calo-scouting stream. The PF reconstruction sequence was removed from the trigger path, and the calo-scouting stream was reconfigured to save HLT muon objects.

The di-muon invariant mass spectrum for events selected by the new trigger in 2017 is shown in Figure 4.8. Thanks to the removal of the mass cut and the L1 p_T cuts, the mass spectrum extends far below 1 GeV. The known resonances in the $m_{\mu\mu}$ spectrum can be seen clearly by eye.

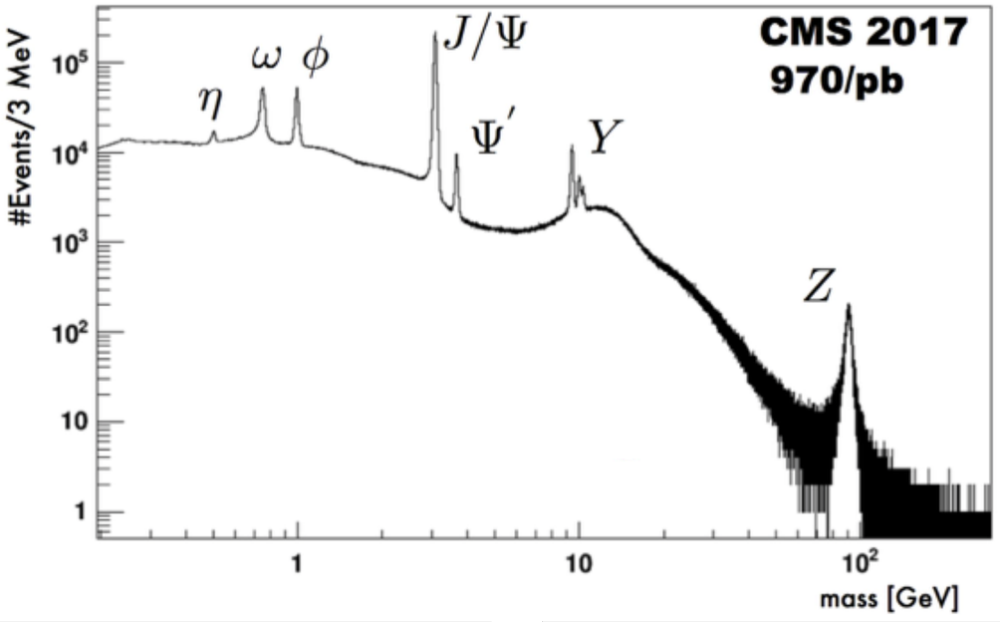


Figure 4.8: Di-muon invariant mass spectrum from events selected by the muon scouting trigger deployed in 2017, with many known resonances labeled.

Chapter 5

IMPACT OF DATA SCOUTING

In this chapter we discuss the analysis of the data collected with the scouting framework in 2015 and 2016, look ahead to the analyses in progress on 2017 scouting data, and discuss future directions for scouting.

5.1 Dijet resonance searches with Run II scouting data

The first search for new physics conducted with the Run II data scouting framework was a low-mass dijet resonance search. The search was carried out on the first 12.9 fb^{-1} of data collected in 2016 [103]. A follow-up analysis extended the results to include a total of 27 fb^{-1} [104]. The search was performed using the same strategy as past CMS dijet searches, using a fit to a smooth parameterized functional form. The latest iteration of the search, performed on the 2017 CMS dataset, is in preparation as of this writing.

Trigger turn-on measurements

In the 2016 dataset, the available triggers for the dijet resonance search are the calo-scouting trigger, which selects events with $H_T > 250$ GeV; and the H_T PF-scouting trigger, which requires $H_T > 410$ GeV.

The efficiency of the $H_T > 250$ GeV path is measured as a function of the dijet mass, yielding a turn-on curve for the trigger. It is computed using the dedicated L1-only and minimum-bias trigger paths, as described in Section 4.4. The turn-on curve is illustrated in Figure 5.1.

The PF-scouting trigger efficiency is measured in the same way. It is found that the trigger becomes fully efficient for dijet masses between 600 and 700 GeV.

Based on the measured trigger efficiencies, it is decided to perform the 2016 dijet search using the calo-scouting dataset collected using the $H_T > 250$ GeV trigger path. This choice yields sensitivity to signal masses in the range 0.6-1.6 TeV. Using calo jets instead of PF jets is not found to have any significant negative impact on the analysis sensitivity.

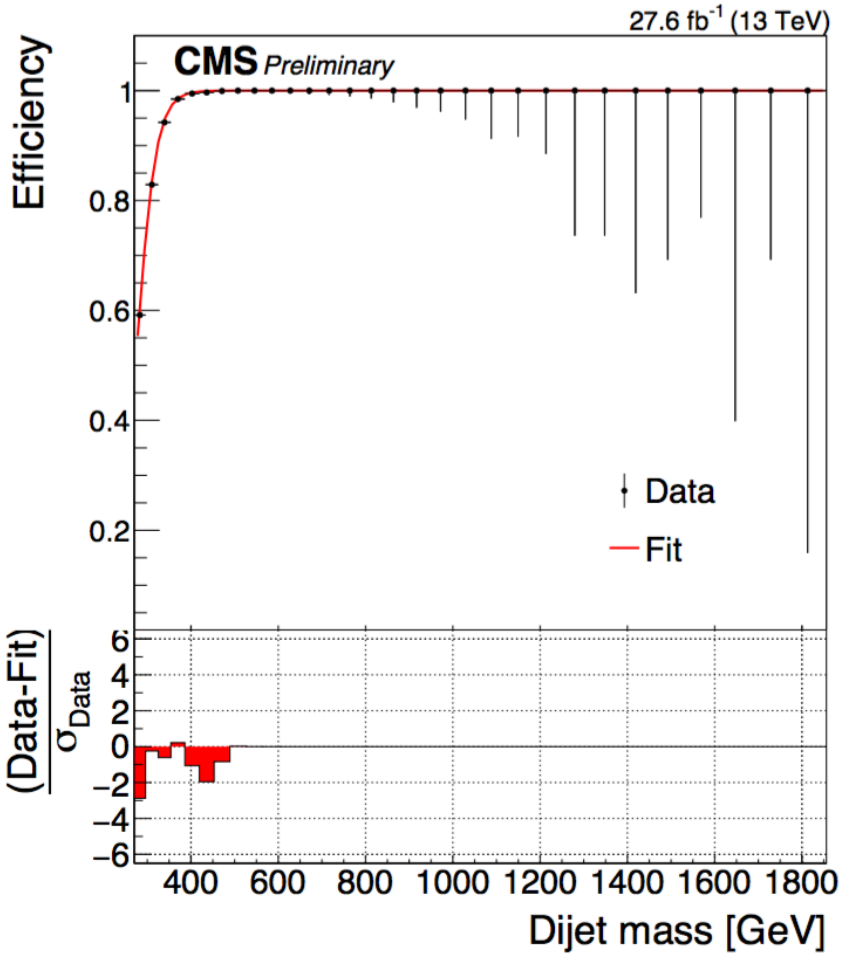


Figure 5.1: Efficiency of the $H_T > 250$ GeV calo-scouting trigger on 2016 data. The efficiency is measured with respect to the dijet invariant mass and is calculated using dedicated L1-only and minimum-bias scouting trigger paths [105].

For the 2017 search, the scouting trigger efficiency is measured using a different technique. A trigger is added to the scouting monitor stream that selects events having a muon with $p_T > 50$ GeV. The set of events collected by this path is used as an unbiased reference sample to measure the H_T trigger efficiency. This method is more convenient than measuring the efficiency using looser H_T triggers, because there is no confounding effect from the turn-on of the L1 trigger. The turn-on curve from the 2017 measurement is shown in Figure 5.2 for the calo-scouting trigger. The trigger reaches 99% efficiency at a dijet mass of 350 GeV.

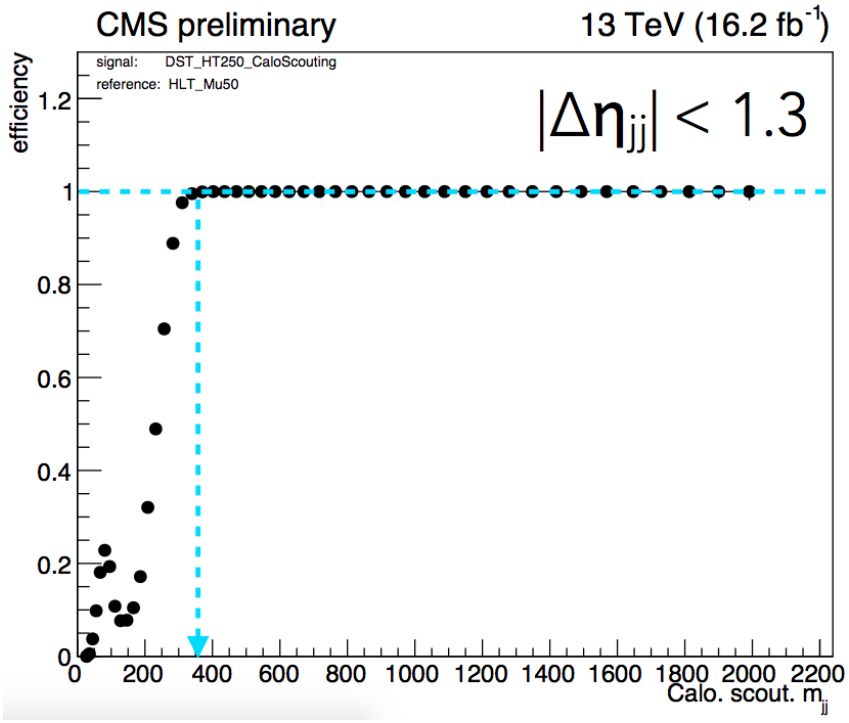


Figure 5.2: Efficiency of the $H_T > 250$ GeV calo-scouting trigger on 2017 data. The efficiency is measured with respect to the dijet invariant mass and is calculated with reference to a sample of events collected with a muon trigger [106].

Comparison of HLT reconstructed objects with standard physics objects

The scouting monitor dataset described in Section 4.4 makes it convenient to directly compare the momenta of HLT calo jets in the scouting data with those of PF jets reconstructed offline. The percent difference between the p_T of an HLT calo jet and that of the corresponding PF jet, measured in bins of HLT jet p_T, is shown on the left-hand side of Figure 5.3. The bias is no larger than 4% and decreases with increasing jet p_T.

The resolution of the dijet invariant mass is measured for HLT calo jets and compared with that for offline reconstructed PF jets. The measured resolution values are displayed on the right-hand side of Figure 5.3. It is seen that the resolution for HLT calo jets is 1-2% worse than that of offline PF jets.

Results and impact

The dijet mass spectrum obtained using the full 2016 scouting dataset is shown in Figure 5.4. The parametric functional form fits the background well, and no

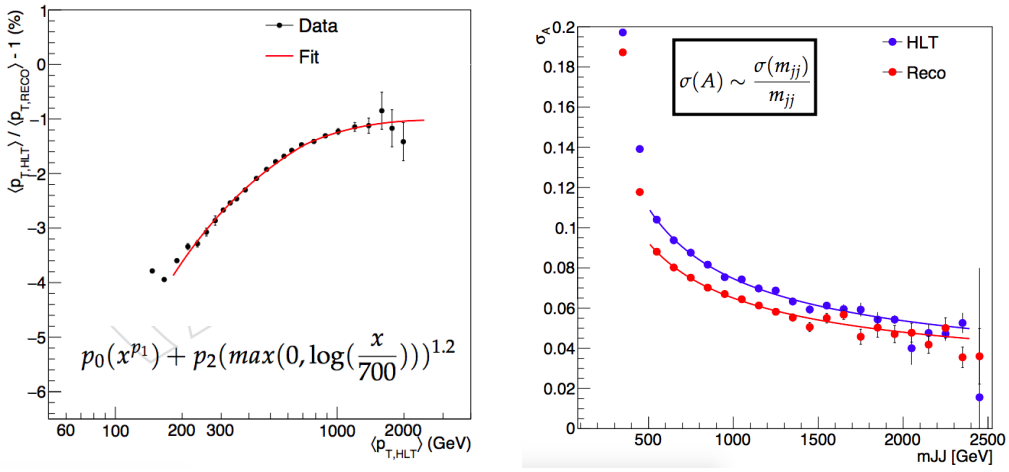


Figure 5.3: Left: average percent difference between HLT calo jet p_T and offline reconstructed PF jet p_T, measured in bins of HLT calo jet p_T. The bias is parameterized with the smooth functional form shown in red. Right: resolution of the dijet mass in HLT (blue) and offline reconstructed (red) events [106].

excess over the smooth background shape is observed. Limits are set on a variety of theoretical models for new physics using the results of the search. Among these are new limits on the production of a Z' resonance decaying to quarks, which improve on those shown in Figure 4.2. The limit is shown as a function of the Z' mass in Figure 5.5.

Historical aside: dijet scouting and the 750 GeV diphoton excess

Interest in hadronic resonance searches below 1 TeV was spurred in late 2015 by the joint announcement by ATLAS and CMS of excesses in the diphoton mass spectrum [107, 108]. ATLAS and CMS observed local excesses with significances of 3.9 and 2.6 standard deviations, respectively, at a diphoton mass of approximately 750 GeV. The global significances of the excesses were 2.1 and 1.2 standard deviations, respectively. Despite the low significance of the CMS excess, the announcement generated a large amount of attention from the theory community.

It was noted [109, 110] that the 8 TeV dijet search using data scouting [100] placed important constraints on the production of a new state at 750 GeV (see Figure 5.6). This is because a strongly-produced resonance should generically exhibit decays to final states with jets. Data scouting therefore drew interest as a way to investigate the possible new particle (see Figure 5.7). If the diphoton excess were confirmed on

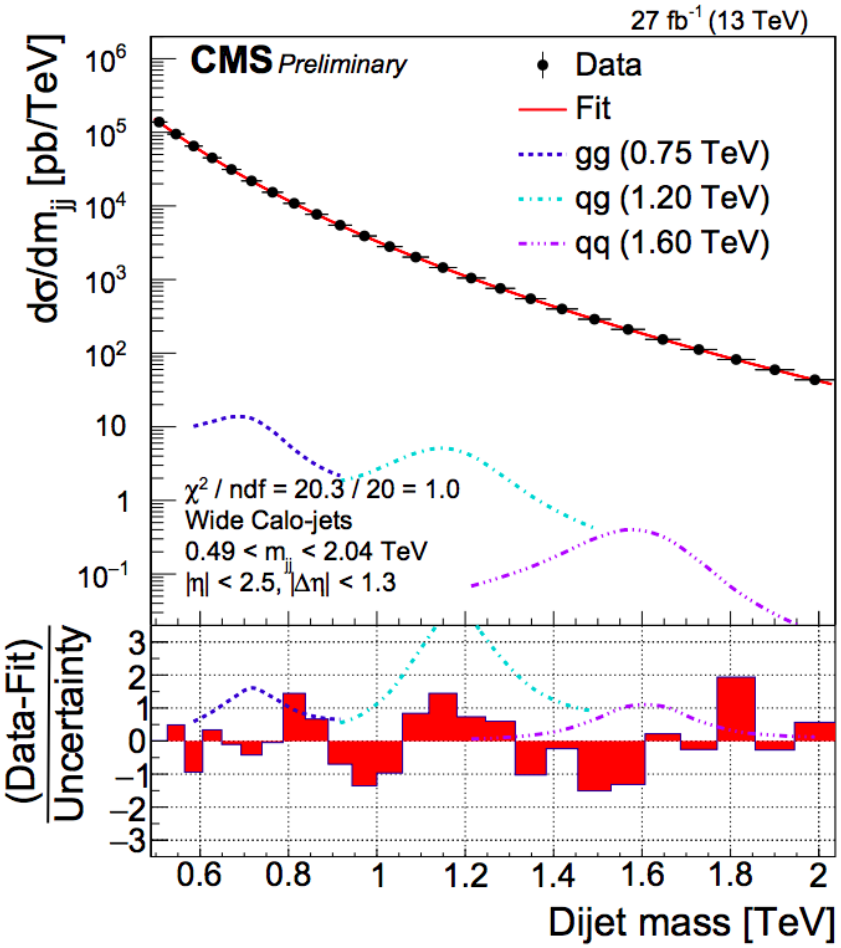


Figure 5.4: Dijet mass spectrum obtained using scouting data collected in 2016, with the fit to a parameterized functional background shape overlaid in red. The bottom panel shows the significance of the difference between the data and the fit in each bin [104].

the 2016 dataset, the discovery could be corroborated in the hadronic decay channel using the scouting dataset.

The CMS and ATLAS diphoton resonance searches were repeated with the data collected in the first half of the 2016 LHC run [112, 113]. Both searches returned null results, as did the dijet scouting search, which was made public at the same time [103]. This indicates that the excesses in the 2015 data had merely been statistical fluctuations. However, the incident highlights the role that data scouting can play in searching for signatures of new physics that are not accessible through standard physics analyses.

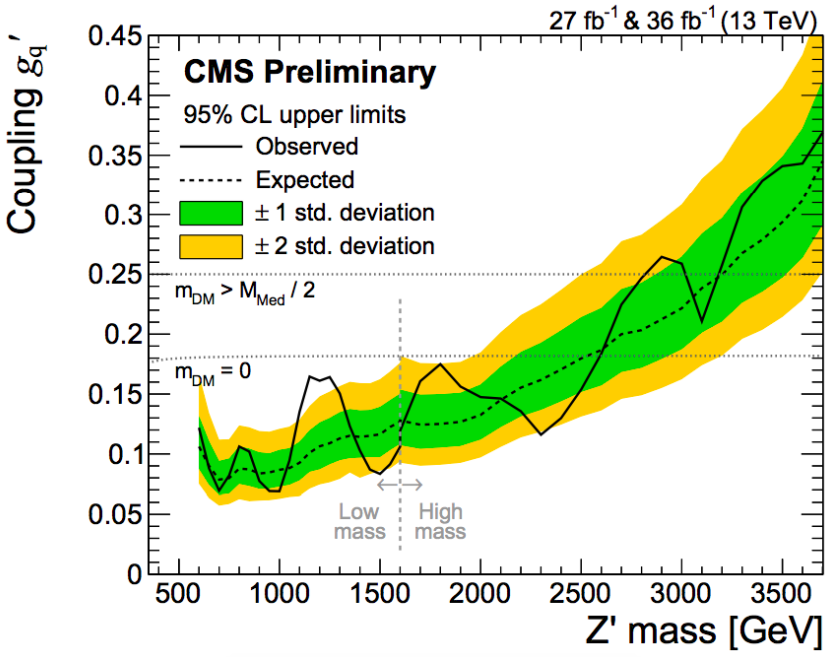


Figure 5.5: Limits on the coupling of a leptophobic Z' resonance decaying to quarks, computed using the results of the 2016 CMS dijet resonance search. In the region to the left of the dashed gray line, the limit is computed using the dataset collected with the calo-scouting trigger [104].

Impact on dark matter limits

The 2016 dijet search yielded strong limits on simplified models of dark matter, as illustrated in Figure 5.8. The simplified models contain a dark matter particle and a heavy mediator that couples to quarks. Because the dijet search is sensitive to direct production of the mediator in LHC collisions, the limits obtained are relatively insensitive to the mass of the dark matter particle. At mediator masses below 1.6 TeV, the exclusion limits are driven by the data scouting part of the dijet analysis.

5.2 Planned hadronic scouting searches

A number of analyses on the 2016 and 2017 scouting datasets are currently in preparation. These are mainly searches using hadronic events that have no striking features, such as the presence of leptons or photons, that can be easily selected with low-rate HLT paths. In these searches a signal must be identified by building up very large datasets and finding features that emerge in the high-statistics limit. We look briefly at three hadronic searches with scouting data that are anticipated in the near future.

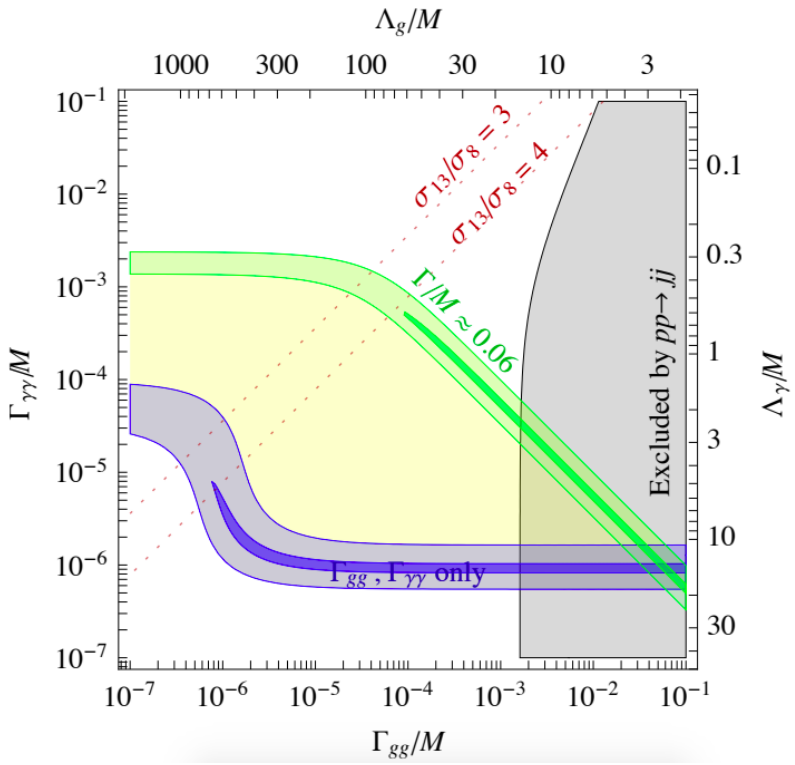


Figure 5.6: Constraints on the decays of a hypothetical resonance with mass 750 GeV based on the CMS and ATLAS data collected in 2015. The x- and y-axes represent the decay widths to gluons and to two photons, respectively. The gray region on the right is excluded by the 8 TeV CMS dijet scouting search [109].

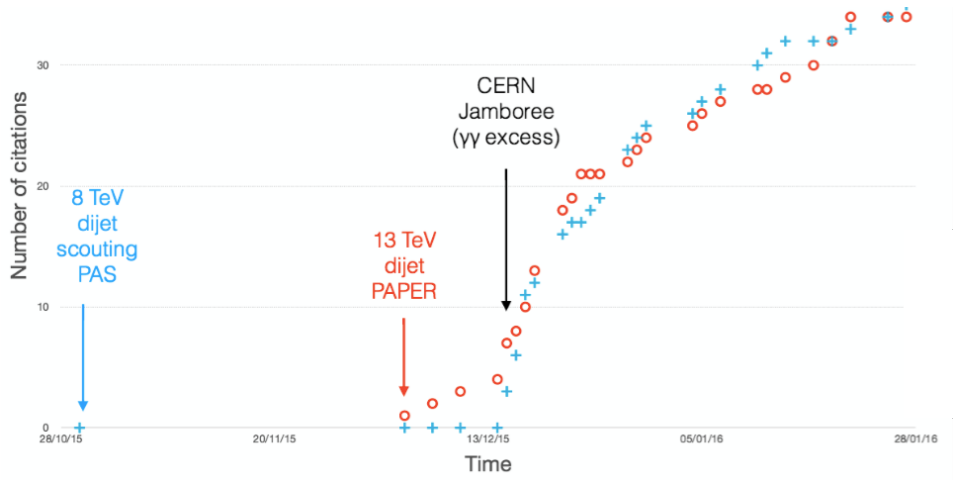


Figure 5.7: Interest in the CMS 8 TeV dijet resonance search with data scouting, as measured by the number of citations of the physics analysis summary (blue crosses) and the published paper (red circles) over time. The black arrow marks the date of the announcement by ATLAS and CMS of the 750 GeV diphoton excess [111].

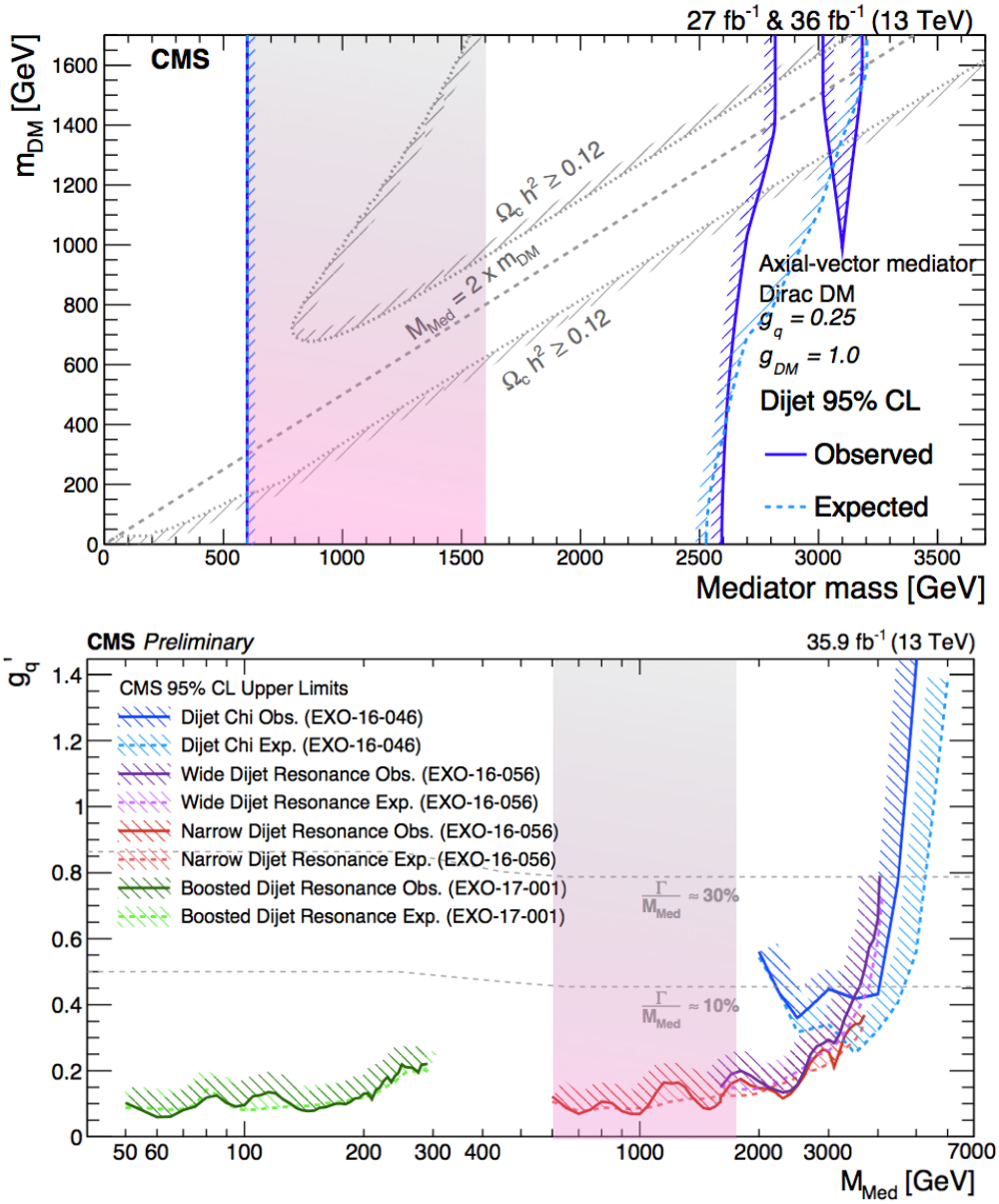


Figure 5.8: Top: 95% confidence level limits from the 2016 CMS dijet analysis on simplified models of dark matter production, expressed as a function of the dark matter and mediator masses for a particular choice of couplings. Bottom: simplified model exclusion as a function of the mediator mass and coupling to quarks (red line). Limits from other CMS searches are also shown. In each plot, the region highlighted in pink represents the limits from the data scouting search [104].

Dijet + ISR search

A notable gap in the dark matter exclusion limits shown in the bottom panel of Figure 5.8 is the region with mediator mass between 300 and 500 GeV. There is hope

to gain sensitivity in this region using events in which the dark matter mediator is produced in association with an ISR jet (see left panel of Figure 5.9). The presence of a third energetic jet in the event (besides the two jets from the resonance decay) allows signal events to pass the trigger more readily. Searching for dijet resonances in events with three or more jets should be feasible with the calo-scouting dataset. There are challenges associated with determining which jet(s) in the event are the ISR jet(s) and which come from the resonance; this must be carefully studied to maximize the analysis sensitivity.

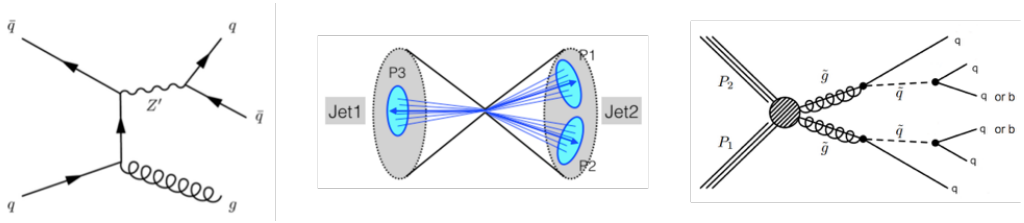


Figure 5.9: Signatures targeted in upcoming hadronic searches using data scouting. Left: production of an exotic Z' boson in association with an ISR jet, with the Z' decaying to two quarks. Center: a tri-jet event in which two jets are very close together geometrically and become merged (image created by Francesco Santanastasio). Right: Pair production of gluinos with *R*-parity violating decay to an all-quark final state.

PF-scouting search with jet substructure

In the case of a dijet resonance produced in association with an energetic ISR jet, the two jets from the resonance may be close together and be clustered into one large jet. Similarly, decays of a tri-jet resonance may result in merged jets (see middle panel of Figure 5.9). These signatures can be investigated using jet substructure variables that quantify the extent to which a jet appears to be the product of multiple merged jets. The green limit curve in the bottom panel of Figure 5.8 is obtained using a search for this boosted dijet signature using standard reconstructed data [114]. The use of the PF-scouting dataset would enable this type of search to be carried out with looser constraints on the jet momenta, thus increasing the range of accessible resonance masses. Initial studies of jet substructure variables using the PF-scouting data have been performed and are seen to produce distributions similar to those obtained in fully reconstructed data.

R-parity violating SUSY models may feature heavy gluinos that decay via a squark into three quarks (see right panel of Figure 5.9). Pair production of gluinos in this model results in six-jet final states in which the jets can be grouped into two three-jet resonances. A search for this signature was performed in Run I [115]. Using the PF-scouting dataset, the analysis can probe a wider range of gluino masses. An interesting feature of the low-mass three-jet analysis is the clear appearance of the mass peak of the hadronically decaying top quark. The top peak appears in the three-jet mass spectrum even in non-*b*-tagged event categories, and is used in the analysis as a candle to calibrate the jet energy scale and resolution.

5.3 Dark photon searches using scouting triggers

The di-muon scouting trigger described in Section 4.7 is being used to perform a search for dark photons using trigger-level data. We briefly describe the motivation and planned strategy for this search.

Motivation for a dark photon search with muon pairs

Some theories of dark matter feature a 'dark electromagnetic' force, governed by a gauge symmetry $\mathrm{U}(1)_D$, that couples only to dark matter particles. The gauge boson Z_D of $\mathrm{U}(1)_D$ is called a *dark photon* [116]. The $\mathrm{U}(1)_D$ symmetry may be spontaneously broken as a result of the dark sector dynamics, leading to dark photons with nonzero mass.

The dark photon mixes with the gauge boson of the SM U(1) interaction, and hence with the SM photon and Z boson. The mixing of Z_D with the SM bosons gives rise to leptonic decays $\mathrm{Z}_D \to \ell\ell$, if this is kinematically allowed. A free parameter ϵ controls the degree of coupling of the dark photon to the photon and the Z. Searches for di-lepton resonances at colliders can be used to constrain ϵ and the dark photon mass m_{Z_D}. A search by LHCb [117] in the di-muon channel recently yielded some of the best limits in this two-dimensional space (Figure 5.10). CMS analyses, particularly Drell-Yan measurements, can be interpreted as limits on dark photons at higher masses (tens or hundreds of GeV) [116]. Very small values of ϵ may yield long-lived dark photons, which can be studied by searching for displaced di-muon resonances.

Data scouting provides a potential avenue to conduct searches for dark photons at lower masses, where the high luminosity of CMS collisions may provide an advantage over LHCb searches. Loosening or removing the standard trigger constraints on the momenta, invariant masses, and vertex position of di-muon pairs may allow CMS to be sensitive to dark photon production at masses as low as 200 MeV.

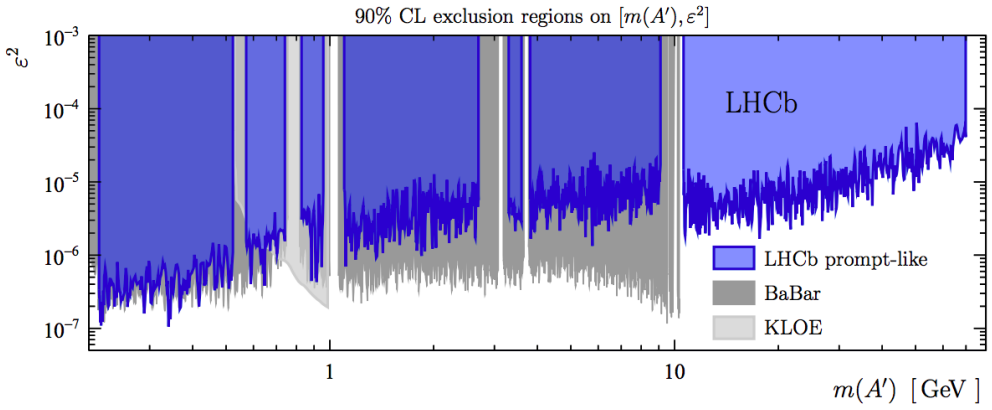


Figure 5.10: Limits on the dark photon mass and the ϵ parameter, computed using the results of the LHCb di-muon search on 2016 data. The best limits to date from other experiments are shown in gray [117].

Planned analysis strategy

The di-muon PF-scouting trigger used in 2016 imposes requirements on the $\mu\mu$ invariant mass (10 GeV) and on the individual muon transverse momenta (11 GeV and 4 GeV, respectively, at the L1). It also requires the muon pair to originate from a vertex along the beamline. Its usefulness in a dark photon search is therefore restricted to the Z_D mass region above about 20 GeV. The modified trigger deployed in 2017 removes the invariant mass and L1 p_T requirements, and is capable of selecting muons from displaced vertices. A p_T requirement of 3 GeV is still imposed at the HLT level, the primary purpose of which is to control the data volume. This requirement could be modified for data taking in 2018 if it is seen to restrict the sensitivity of the analysis.

The main search can be performed as a bump-hunt in the $\mu\mu$ invariant mass spectrum. A displaced search, targeting dark photon masses below 1 GeV and ϵ~10^{-5}, is also planned.

Significant analysis tasks include:

- Obtaining reliable estimates of the trigger efficiency in all search regions and quantify any sculpting of the di-muon mass distribution by the L1 selection cuts. This task is complicated due to the large number of L1 seeds feeding into the trigger and the very heterogeneous set of requirements on the L1 muons.

- Quantifying the HLT muon object performance and tuning the muon identification cuts to optimize each search region.

- Parameterizing the background shape, including the known resonances and the nonresonant component from Drell-Yan events and fake muons. The contribution from fake muons can be estimated using samples of like-sign muon pairs; this is complicated by the L1 trigger selection, which includes seeds with an opposite-sign requirement on the muons. A possible strategy for 2018 is to ensure that each opposite-sign L1 seed has a corresponding same-sign seed (which may be prescaled) so that the same-sign contribution can be estimated without bias [118].

- For the displaced search, characterizing secondary vertices reconstructed from HLT muon objects. This includes understanding the influence of the CMS tracker material on the distribution of secondary vertex positions, and identifing exploitable kinematic features of signal and background events.

5.4 Outlook for the future

As the LHC continues to deliver more data, the marginal benefit of merely repeating traditional searches for new physics will decrease (in the absence of a discovery). Data scouting offers one route to expand the space of searches and measurements that can be performed. By enabling the collection of datasets orders of magnitudes larger than those obtained with the standard HLT menu, it allows analysts to fully make use of the very high luminosity – not just the high energy – of the LHC.

The other LHC experiments have recently adopted strategies similar to data scouting. ATLAS performed a dijet resonance search with 2015 data using a trigger-level analysis, placing limits in the region 450-950 GeV [119]. LHCb's 'Turbo Stream,' deployed at the beginning of Run II, performs detector calibration and analysis preselection at the trigger stage [120], aiming to achieve offline-like reconstruction in real time. These developments indicate a shift towards data scouting-like analysis

paradigms more broadly in HEP. Trigger-based searches have the potential to become a central component of many experiments' trigger strategies as detector data volumes continue to increase in the years to come.

Part III:

Supersymmetry Searches using Razor Variables

What is this sample proving?
Anecdotes cannot say what Time may do.

Joanna Newsom, "Anecdotes"

Chapter 6

FIT-BASED RAZOR SEARCH ON 2015 DATA

Searches for SUSY using razor variables in Run I of CMS employed a background prediction strategy based on fits to a parameterized functional form. This approach capitalizes on the kinematic properties of the razor variables M_R and R^2 introduced in Section 3.7. It does not rely on MC simulation except to validate the search methodology, so it is unaffected by many sources of systematic uncertainty that complicate MC-based searches. It provides a natural complement to existing hadronic and leptonic SUSY searches in CMS.

On the 2015 CMS dataset a razor search for SUSY was carried out on 2.1 fb^{-1} of 13 TeV data with two background predictions: one fit-based and one MC-based. A complete estimate of the SM background, including systematic uncertainties, is performed for both fit and MC approaches. The two background predictions are compared before unblinding the data in the search region, to ensure that they are consistent with one another.

The reason for this dual approach, which we believe is unique in CMS, was a concern that the assumptions of the fit-based background prediction are strong and cannot be fully covered by systematic uncertainties. For more about this issue and how it affected the 2016 inclusive razor search for SUSY, please see Chapter 9. By performing a second background prediction using a completely different set of assumptions (the assumptions that come with a MC-based search), we significantly increase the robustness of the search.

In this chapter we discuss the fit-based background prediction for 2015 data. The next chapter will cover the MC-based prediction and the interpretation of the search in terms of SUSY simplified models.

6.1 Motivation for the fit-based razor analysis

In Section 3.7 we introduced the razor variables M_R and R^2. The M_R variable tends to peak at a particular scale, M_Δ, which is related to the characteristic mass scale of the physics process. The distribution of M_R for $M_R > M_\Delta$ is roughly a falling exponential, as is the distribution of R^2. M_R and R^2 are correlated variables: the

slope of the exponential tail of M_R varies with the value of R^2, and vice-versa. This is illustrated in 7 TeV CMS data in Figure 6.1, where it is seen that the slope of the M_R (R^2) distribution changes roughly linearly with the baseline R^2 (M_R) cut imposed.

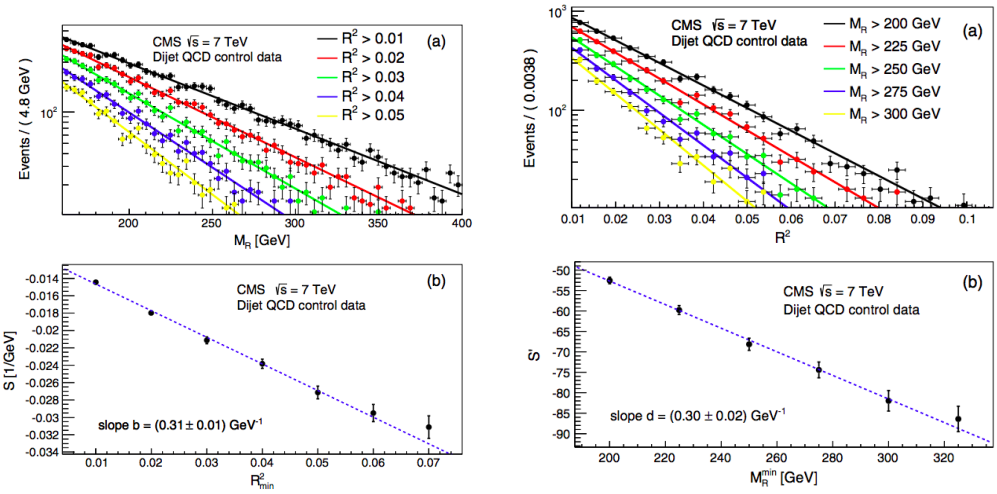


Figure 6.1: Left: distribution of M_R after selecting events with R^2 above various thresholds. Right: distribution of R^2 after selecting events with M_R above various thresholds. All M_R and R^2 distributions are fit with exponential functions. The best-fit exponential slope is shown in the bottom plots as a function of the cut on R^2 (left) or M_R (right) [95].

The fact that the razor variables exhibit this exponential behavior can be exploited to obtain an estimate of the background without resorting to MC simulation. We first identify a parametric family of functions that model the behavior of the falling background accurately; this is detailed in Section 6.2. We then define a signal-sensitive region of the M_R-R^2 plane, and a sideband region in which any SUSY signal is expected to be small compared with the background. We use a fit to data to optimize the background function within the constraints of the specified parametric family. The best-fit function and its uncertainty are used to characterize the SM background in the signal-sensitive region. Excesses in data above the predicted background are a possible sign of new physics.

6.2 Razor fit function

The relationship between M_R and R^2 indicated by Figure 6.1 suggests that we model the SM background with a function of the form

$$f(M_R, R^2) \propto e^{-b(M_R - M_R^0)(R^2 - R_0^2)}, \tag{6.1}$$

where b, M_R^0, and R_0^2 are real-valued parameters.

This describes somewhat adequately the razor variable distributions in data, but does not completely capture the tail behavior for all backgrounds. We add a parameter n to give the function additional freedom to capture the tail shape. The function we use, which more reliably captures the full range of observed background shapes, is

$$f(M_R, R^2) = \left(b\left[(M_R - M_R^0)(R^2 - R_0^2)\right]^{1/n} - 1\right) e^{-bn\left[(M_R - M_R^0)(R^2 - R_0^2)\right]^{1/n}}. \quad (6.2)$$

The prefactor in front of the exponential term is included to make this distribution analytically integrable on rectangles $\left[M_R^{lo}, M_R^{hi}\right] \times \left[R_{lo}^2, R_{hi}^2\right]$ in the M_R-R^2 plane:

$$\begin{aligned}\int_{M_R^{lo}}^{M_R^{hi}} \int_{R_{lo}^2}^{R_{hi}^2} f(M_R, R^2)\, dM_R\, dR^2 = \\ n(bn)^{-n}\left[G(M_R^{lo}, R_{lo}^2) - G(M_R^{hi}, R_{lo}^2) - G(M_R^{lo}, R_{hi}^2) + G(M_R^{hi}, R_{hi}^2)\right],\end{aligned} \quad (6.3)$$

with

$$G(x, y) = \Gamma\left(n, bn\left[(x - M_R^0)(y - R_0^2)\right]^{1/n}\right), \quad (6.4)$$

where Γ is the upper incomplete gamma function.

Binned fits and uncertainty

The functional form Eq. 6.2 is the one used for background modeling in the 2015 razor SUSY search. To ease the numerical computation of the fit result, we bin the data using a predefined grid of n_b rectangles in the M_R-R^2 plane and perform the fit by maximizing the binned Poisson likelihood:

$$\mathcal{L}(\{x_i\}|b, n, M_R^0, R_0^2, N) = \prod_{i=0}^{n_b} \text{Poisson}(x_i \mid \lambda_i(b, n, M_R^0, R_0^2, N)), \quad (6.5)$$

where x_i is the number of observed data counts in bin i. Here we have introduced a new random variable N that represents the mean total number of events in the dataset, summed over all bins. The observed size of the dataset is assumed to be Poisson distributed with mean N. The mean background λ_i in each bin is obtained by normalizing the razor function $f(M_R, R^2)$ to have a total integral of N, and then integrating the normalized function over the bin area using Eq. 6.3. The analytic form of Eq. 6.3 allows the likelihood to be evaluated without the need for numerical integration. Likelihood maximization is performed using the RooFit package [121].

Uncertainty on the fitted function is quantified using the covariance matrix of the fit parameters b, n, M_R^0, R_0^2, and N, which captures the uncertainties on the parameters and the correlations between them. The covariance matrix is computed by RooFit using the curvature of the likelihood function (Eq. 6.5) at its maximum. We use a toy Monte Carlo procedure to convert the uncertainties on the fit parameters into uncertainties on the bin yields λ_i. The following is performed for $j = 0, 1, \ldots, N_{trials}$:

1. Draw a sample $b_j, n_j, M_{R,j}^0, R_{0,j}^2, N_j$ from a multivariate normal distribution, with mean and covariance matrix taken from the output of the RooFit optimization routine;

2. Evaluate $\lambda_{ij} \equiv \lambda_i(b_j, n_j, M_{R,j}^0, R_{0,j}^2, N_j)$ for each analysis bin using the sampled parameters.

This procedure corresponds to repeatedly perturbing the fit function parameters according to their uncertainties, and computing the effect on the predicted bin yields. It returns a set $\left\{\lambda_{ij}\right\}$ of simulated yields in each analysis bin (an example for one bin is shown in Figure 6.2). To assign an uncertainty on the best-fit bin yield λ_k, we compute the width of the smallest interval containing 68% of the simulated values λ_{kj}. This width (symmetrized around the best-fit values λ_k, for simplicity) is illustrated as an uncertainty band on the fit prediction plots shown in the following sections.

A second set of simulated yields, $\left\{x_{ij}\right\}$, is computed by throwing Poisson random variables with means $\left\{\lambda_{ij}\right\}$. This set of yields represents the expected spread of observed event counts in each bin, accouting for both uncertainty on the fit function and the Poisson uncertainty on the number of counts. This set of yields is used to compute the significance of deviations between the observed data and the fit prediction. The deviation is reported as a 'number of sigma,' stylized as $n\sigma$. It is obtained as the gaussian quantile function Φ^{-1} (the inverse of the cumulative distribution function of the normal distribution) applied to the one-sided p-value of the observed data:

$$n\sigma \equiv \Phi^{-1}\left(\frac{\sum_j \mathbb{1}_{x_{ij}<x_i}}{N_{trials}}\right), \tag{6.6}$$

where $\mathbb{1}_{x_{ij}<x_i}$ is an indicator function that is 1 when $x_{ij} < x_i$ and 0 otherwise.

In plots displaying the $n\sigma$ significance in each bin, values are reported in both the sideband region and in the signal-sensitive extrapolation region. It should be kept

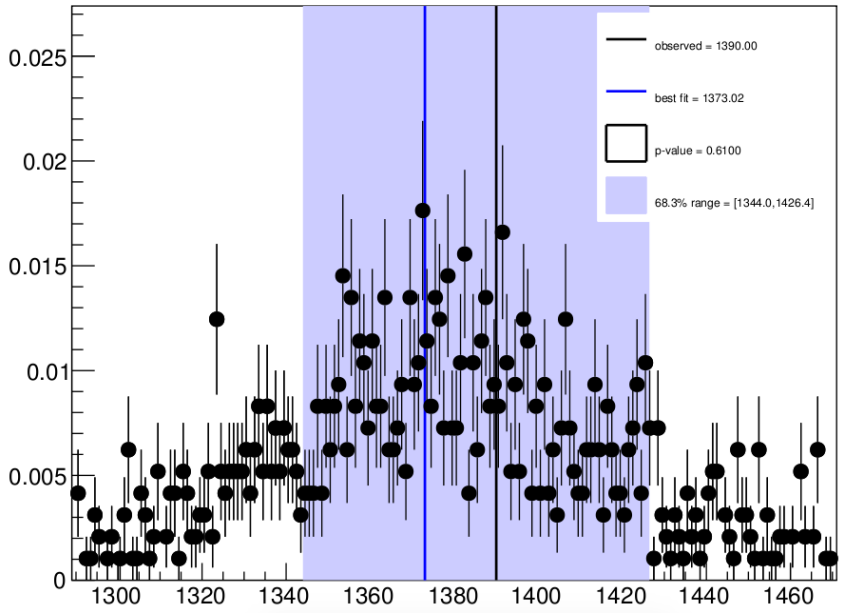


Figure 6.2: Example histogram of simulated yields in one analysis bin, generated from the razor fit result using the toy MC procedure described in the text. The x-axis indicates the bin yield λ_{ij}. The value corresponding to the best-fit bin yield is indicated by the vertical blue line, and the smallest interval containing 68% of the probability mass is highlighted in lighter blue. The vertical black line indicates the number of counts observed in data.

in mind that, for fits performed in the sideband region, the $n\sigma$ values reported for sideband bins are not corrected to account for the dependence of the fit function on the data. They will therefore tend to be smaller in magnitude than those in the extrapolation region.

6.3 Sideband fits and full fits

We perform the razor fit in two different ways. The first method (the 'sideband fit') is used to perform a signal-agnostic search for excesses at high M_R and R^2. The second method (the 'full fit') is used for interpretation of the search in terms of specific signal models.

Sideband fits

To accommodate the possibility of arbitrary signal-like excesses at high M_R and R^2, we perform the sideband fit in an L-shaped region in the M_R-R^2 plane, keeping the rest of the selected region (the extrapolation region) blinded. The sideband region is

defined by the union of two rectangles, one at low M_R and the other at low R^2 (see Figure 6.5). After fitting to the data in the sideband region, the best-fit function and its covariance matrix are used to describe the predicted background shape in the high-M_R, high-R^2 extrapolation region.

The sideband fit is a background-only fit, and does not assume anything about a particular signal except that any signal contamination in the sideband is small compared with the SM background. This makes the search sensitive to a variety of new physics signals, not just the specific simplified SUSY models for which MC is available. After performing the fit in the sideband, we unblind the extrapolation region and look for excesses of any kind above the fitted background.

When comparing the fit-based razor background prediction with the MC-based prediction described in the next chapter, we use the predictions obtained using the sideband fits.

Full fits

To interpret the search in terms of a specific signal model, we remove the distinction between sideband and extrapolation region, and perform a binned signal+background fit in the full M_R-R^2 region under consideration. In this full-region fit the signal shape is determined using a MC template, and the signal strength is treated as a freely floating parameter.

The full fit method is used to perform hypothesis tests to discover or set limits on the production of SUSY signals; it will be discussed further in Chapter 7.

6.4 Razor triggers

The razor variable R^2 is not strongly correlated with E_T^{miss} for most physics processes. Standard SUSY HLT paths select events based on H_T and E_T^{miss}, and are therefore unsuitable for use in this search. Instead, we designed and deployed dedicated razor trigger paths that compute the values of M_R and R^2 at the HLT. These triggers select events if they satisfy the following requirement on M_R and R^2:

$$(M_R + 300) \times (R^2 + 0.25) > \text{threshold}, \tag{6.7}$$

where the selection threshold in 2015 is 240 GeV. The selected region is illustrated in Figure 6.3. The motivation for this hyperbolic selection cut is based on the shape of the M_R-R^2 distribution (see Figure 3.11), and the dependence of the fit function

(Eq. 6.2) on the quantity $(M_R - M_R^0)(R^2 - R_0^2)$. The values of the M_R and R^2 offsets (300 GeV and 0.25) were optimized using a fit to Run I data. In addition to the hyperbolic cut, loose baseline cuts on M_R and R^2 are applied. These cuts have a subdominant effect on the trigger rates and efficiency compared with the hyperbolic cut.

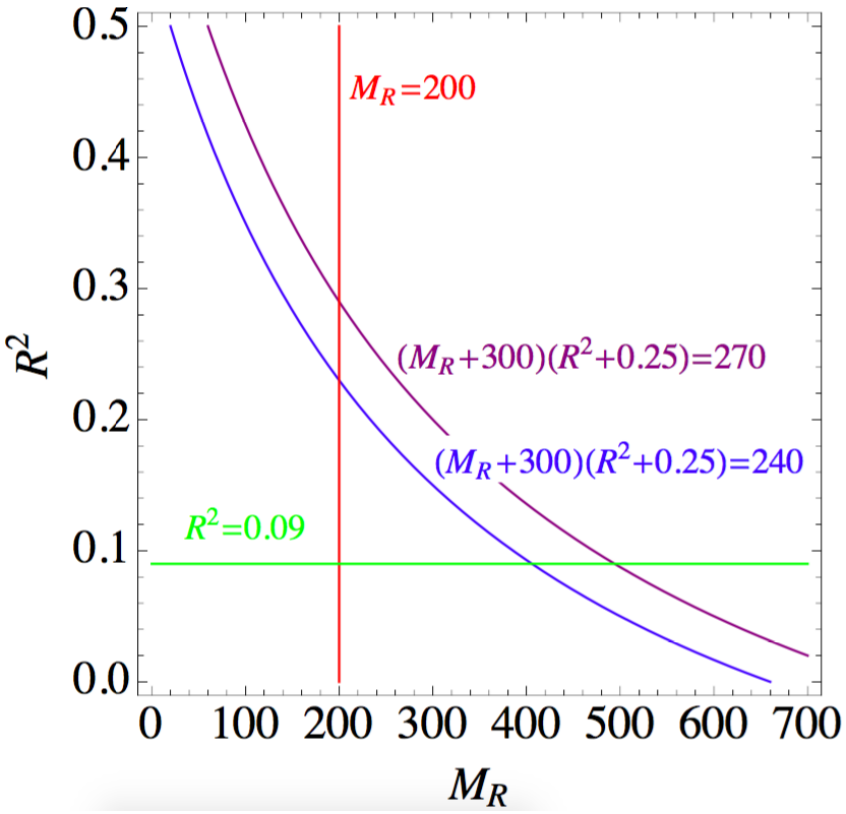


Figure 6.3: Diagram indicating the selection cuts of the hadronic razor dijet and quad-jet triggers used in 2015. Events lying in the upper right of the diagram, with M_R and R^2 satisfying the hyperbolic cut as well as loose rectangular cuts, are selected. Two hyperbolic cuts are shown; the looser cut, $(M_R+ 300)(R^2+ 0.25) >$ 240 GeV, is the one used by the main razor triggers. Backup versions of the paths apply the tighter 270 GeV cut.

For the 2015 razor search, we deployed two flavors of hadronic razor triggers with cuts based on Eq. 6.7. Razor 'dijet' triggers require at least two jets with $p_T > 80$ GeV (in addition to the cuts mentioned above), and razor 'quad-jet' triggers require at least four jets with $p_T > 40$ GeV. An additional 'R^2-only' trigger accepts events with $R^2 > 0.25$, omitting the hyperbolic cut. This trigger is primarily designed to search for signals of dark matter production, which in many models yields a peak around R^2~0.8. We include it in this search to improve upon the effi-

ciencies of the dijet and quad-jet triggers in the high-R^2 region. The R^2-only trigger also requires two jets having $p_T > 80$ GeV.

The calculation of M_R and R^2 at the HLT is performed using PF reconstruction. To prevent the average event processing time (which is dominated by the time taken to run the PF algorithm) from growing too large, a loose preselection is applied before running PF to reject a significant fraction of events. The preselection is based on the values of M_R and R^2 computed using calorimetric jets and E_T^{miss}. It consists of a looser version of the hyperbolic cut. For the trigger with a hyperbolic cut value of 240 GeV, for example, the calorimetric preselection imposes a hyperbolic cut of 200 GeV.

The list of razor triggers is given in Table 6.1. It includes the triggers used for the search as well as 'backup' paths with tighter thresholds, which serve as fallbacks in case the rates of the main triggers become too high.

Trigger cuts	Rate in 2015 (Hz)	Notes
$(M_R+300)(R^2+0.25) > 240$ GeV; two 80-GeV jets	7.7	
$(M_R+300)(R^2+0.25) > 240$ GeV; four 40-GeV jets	1.2	
$R^2 > 0.25$; two 80-GeV jets	0.7	
$(M_R+300)(R^2+0.25) > 270$ GeV; two 80-GeV jets	2.3	Backup path
$(M_R+300)(R^2+0.25) > 270$ GeV; four 40-GeV jets	0.5	Backup path
$R^2 > 0.30$; two 80-GeV jets	0.4	Backup path

Table 6.1: Razor triggers deployed in 2015, with rates estimated from data at 5×10^{33} cm^{-2}s^{-1}.

Our search is performed in the region of high M_R and high R^2, where a variety of SUSY signals may be expected to appear. We measure the efficiency of the razor triggers as a function of M_R and R^2 to identify an appropriate search region. The efficiency measurement is performed using a dataset of single electron events, selected using triggers that have no razor variable requirement. The electron dataset provides an unbiased baseline for the efficiency computation. The efficiency in each bin is estimated as

$$\epsilon = \frac{\text{\# events in electron dataset, passing razor path(s)}}{\text{\# events in electron dataset}}. \tag{6.8}$$

The measured efficiencies are shown in Figure 6.4. We choose to perform the search in the region with $M_R > 500$ GeV and $R^2 > 0.25$, where the triggers are close to 100% efficient.

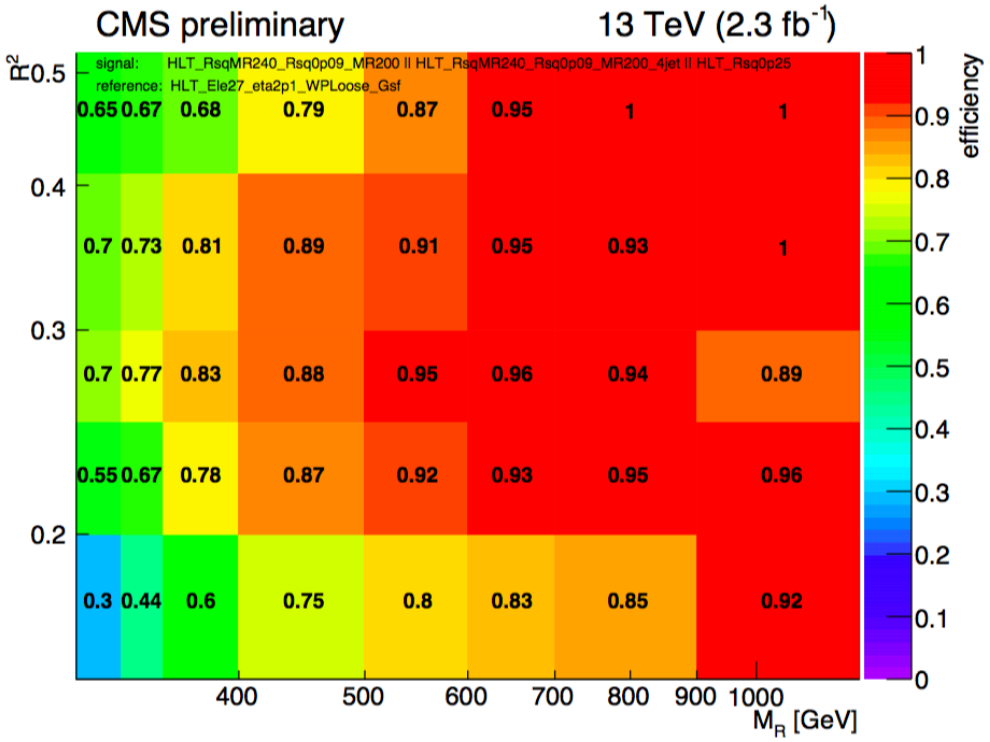


Figure 6.4: Combined efficiency of the 2015 razor trigger suite, measured in data on a sample of events passing inclusive electron trigger paths. The efficiency is computed in bins of M_R and R^2.

6.5 Event selection and categorization

Events for the search are selected based on the number of jets, b-tagged jets, and leptons identified, and on the values of M_R and R^2. The events are sorted into categories in order to enhance the sensitivity to specific types of SUSY signals, and each category is divided into two-dimensional bins in M_R and R^2. The goal of the search is to predict the number of events in each M_R-R^2 bin and compare the prediction with the data.

The razor approach is applicable to events both with and without leptons. We perform the search in a zero-lepton category (denoted 'Multijet'), a one-muon category ('Muon Multijet'), and a one-electron category ('Electron Multijet'). Events with two or more identified leptons are not considered in this search. Leptons are included in the razor megajet clustering procedure in the same way as jets (see Section 3.7). The same parametric family of fit functions (Eq. 6.2) is used for background modeling in the zero- and one-lepton categories. One-lepton and zero-lepton categories have no overlap, so they can easily be statistically combined with each other for analysis interpretation.

Zero-lepton events are required to pass at least one of the razor triggers described in Section 6.4. One-lepton events are required to pass one of a set of inclusive electron or muon triggers, which place no cuts on the razor variables or other jet-based quantities. Noise rejection algorithms are used to reject events consistent with known beam- and detector-related issues [30].

Lepton identification

The analysis makes use of two different sets of selection criteria for identifying leptons. 'Tight' criteria are stringent, with selection efficiencies around 70-75%. 'Veto' criteria are more relaxed, selecting 90-95% of genuine prompt leptons but having higher fake rates. The tight selection criteria are used to select a pure sample of leptons from the decays of W and Z bosons for the one-lepton search categories. The veto criteria are employed to identify and reject leptonic events from the zero-lepton search categories.

Electrons

The tight electron selection criteria consist of cuts on ECAL cluster shape variables, the HCAL/ECAL energy ratio, and the quality of the electron's GSF track [122]. A conversion veto is employed to reject electrons from photons that convert in the tracker. Different cut thresholds are used for electrons in the barrel region and in the endcaps, in order to achieve a similar efficiency in each region. Electron PF candidates with $p_T > 25$ GeV are eligible for the tight selection.

Electrons are also required to be isolated from other electromagnetic and hadronic activity in the event. This is accomplished via a variable called *mini-isolation*. Particle isolation is traditionally computed by defining a cone with a fixed radius, $\Delta R < R_{cone}$ (with ΔR defined as in Eq. 2.5) around the particle, and summing the transverse momenta of the other particles falling within the cone. The isolation is then corrected for the presence of pileup using an effective-area method similar to that described for jets in Section 2.4. The mini-isolation variable is computed in the same way as traditional isolation, except that the cone size changes with the p_T of the electron:

$$R_{cone} = \begin{cases} 0.2, & p_T < 50 \text{ GeV} \\ 10 \text{ GeV}/p_T, & 50 \text{ GeV} < p_T < 200 \text{ GeV} \\ 0.05, & p_T > 200 \text{ GeV}. \end{cases} \tag{6.9}$$

The main purpose of the changing cone size is to accommodate leptons from decays of high-p_T top quarks, which may lie very close to the b-jet from the top decay. It also tends to give a looser isolation requirement in events with large amounts of hadronic activity. For the tight selection, the value of the mini-isolation is required to be less than 10% of the electron's p_T.

The loose 'veto' selection for electrons is based on a multivariate discriminator (MVA) computed using a number of ECAL shower shape and GSF track variables. Compared with a cut-based selection, the MVA has a lower fake rate for a given signal efficiency; this results in fewer hadronic signal events being spuriously rejected. The veto selection is applied to all reconstructed PF electrons with $p_T > 5$ GeV. In addition to the MVA selection, a requirement is imposed on the electron isolation. For electrons with $p_T > 20$ GeV, we require the mini-isolation variable to be no larger than 20% of the electron p_T. For electrons below 20 GeV, we require that the value of the mini-isolation is no more than 5 GeV. The switch to an absolute isolation cut below 20 GeV is seen to significantly increase the efficiency of the isolation requirement.

Muons

Muons are identified based on quality requirements on the reconstructed muon tracks in the inner tracker and muon chambers [123]. We use the 'tight' and 'loose' selections recommended by the muon working group within CMS, and additionally require that the muon track impact parameter (the distance of closest approach to the beamline) be small. The tight selection is applied to all PF muon candidates with $p_T > 20$ GeV, and the loose selection is applied to candidates with $p_T > 5$ GeV.

We assess isolation of muons using the mini-isolation variable. For muons with $p_T > 20$ GeV, we require that the mini-isolation be no greater than 20% of the muon's p_T. For muons with $p_T < 20$ GeV, we switch from a relative cut on isolation to an absolute one, requiring that the absolute isolation sum be no more than 10 GeV.

Tau leptons

Hadronically decaying τ leptons are selected using the loose working point of the hadron-plus-strips algorithm [124], which identifies τ decays to final states with up

to three hadrons. A PF-based isolation cut is also imposed. Events with identified hadronic τ leptons having $p_T > 20$ GeV are rejected from the zero-lepton search category. The hadronic τ veto has an efficiency of about 50%.

Jet and b-jet identification

Jets and b-tagged jets are identified as described in Section 2.4. We use anti-k_T jets with $R = 0.4$, and apply the CHS procedure to reduce pileup. Each jet is required to pass tight selection criteria designed to reject detector noise. Jets passing these criteria and having $p_T > 40$ GeV and $|\eta| < 3.0$ are selected for the analysis. The jet p_T is corrected using JEC factors defined by the CMS jet working group. In MC, jets receive an additional 'residual' correction to compensate for observed differences in jet performance between data and MC simulation. Selected jets are tagged as b-jets based on the 'medium' working point of the CSVv2 discriminator.

Search categories

The analysis search categories are summarized in Table 6.2. The Multijet, Muon Multijet, and Electron Multijet categories are each divided into subcategories with events having 0, 1, 2, and 3 or more b-tagged jets in order to isolate SUSY events with bottom or top quarks in the final state.

Selected jets, electrons, and muons in each event are clustered into two megajets as described in Section 3.7 and used to compute the razor variables M_R and R^2. Baseline cuts on M_R and R^2 are applied. These cuts are tighter for zero-lepton events due to the restrictions imposed by the razor triggers. In the Multijet search region we require $M_R > 500$ GeV and $R^2 > 0.25$, and in the Muon and Electron Multijet search regions we require $M_R > 400$ GeV and $R^2 > 0.15$. We define the L-shaped sideband regions to have widths of 100 GeV in M_R and 0.05 in R^2. The sideband and signal-sensitive regions for hadronic and leptonic analysis categories are depicted in Figure 6.5.

As discussed in Section 3.6, this search is primarily geared toward detection of heavy gluinos, for which the cross section at 13 TeV may be an order of magnitude larger than that at 8 TeV. To improve sensitivity to gluino signals having many jets in the final state, we require $N_{\text{jets}} \geq 4$ in all analysis search regions, where N_{jets} is the number of jets passing the selection requirements.

We require events in the Multijet category to satisfy $\Delta\phi_R < 2.8$, where $\Delta\phi_R$ is the

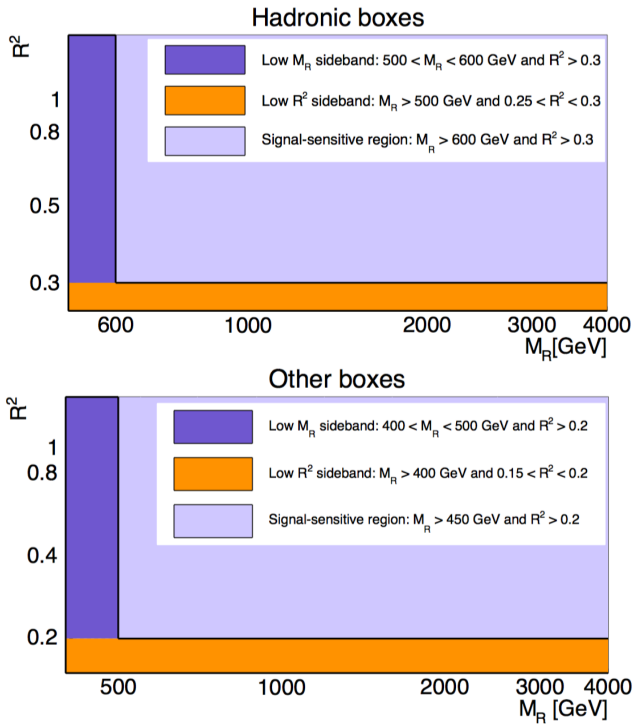


Figure 6.5: Diagram indicating the sideband and extrapolation regions for the hadronic (top) and leptonic (bottom) search categories. The sideband region is the union of the low-M_R and low-R^2 regions indicated. The extrapolation region is the upper-right rectangle in each diagram.

absolute value of the azimuthal angle between the two razor megajets. This requirement reduces the amount of QCD multijet background in the search region. QCD events, which have no intrinsic E_T^{miss}, enter the analysis primarily when the momentum of a jet is badly mismeasured. In such events, the reconstructed $\vec{p}_T^{\,miss}$ vector is aligned with one of the razor hemispheres, and thus, by momentum conservation, the razor hemispheres are back to back in the transverse plane and the value of $\Delta\phi_R$ is close to π.

In the Muon and Electron Multijet categories we require $m_T > 120$ GeV, where m_T is the transverse mass, defined as

$$m_T = \sqrt{\left|\vec{p}_{T,\mathrm{lep}}\right| E_T^{miss} - \vec{p}_{T,\mathrm{lep}} \cdot \vec{p}_T^{\,miss}}, \tag{6.10}$$

where $\vec{p}_{T,\mathrm{lep}}$ is the lepton transverse momentum vector. This cut eliminates a significant fraction of W($\rightarrow \ell\nu$)+jets and single-lepton $t\bar{t}$+jets events, which have a small angle between the lepton and the $\vec{p}_T^{\,miss}$.

The 0, 1, and 2 b-tag subcategories in each search region are fit with independent copies of the razor fit function (Eq. 6.2). The $\geq$3 b-tag subcategory has limited event

Category	Baseline Cut	Sideband	Other Cuts
Multijet	$M_R > 500$ GeV and $R^2 > 0.25$	$M_R < 600$ or $R^2 < 0.3$	$N_{\text{jets}} \geq 4, \Delta\phi_R < 2.8$
Muon Multijet	$M_R > 400$ GeV and $R^2 > 0.15$	$M_R < 500$ or $R^2 < 0.2$	$N_{\text{jets}} \geq 4$, $m_T > 120$ GeV
Electron Multijet	$M_R > 400$ GeV and $R^2 > 0.15$	$M_R < 500$ or $R^2 < 0.2$	$N_{\text{jets}} \geq 4$, $m_T > 120$ GeV

Table 6.2: Search categories used in the 2015 razor analysis. Each search category is divided into 0, 1, 2, and ≥3 b-tag subcategories.

statistics, so it is fit simultaneously with the 2 b-tag category, and shares the values of the parameters b, n, M_R^0, and R_0^2 with it. The consistency of the 2 and ≥ 3 b-tag shapes is compared using a MC simulated event sample. It is seen that the R^2 shape is consistent across the two samples, but that the M_R shape deviates slightly in a manner consistent with linear (see Figure 6.6). We therefore parameterize the background shape in the ≥3 b-tag category as

$$f^{\geq 3b}(M_R, R^2) = \left(1 + m_{M_R}(M_R - M_R^{\text{offset}})\right) f^{2b}(M_R, R^2), \tag{6.11}$$

where the slope and intercept parameters m_{M_R} and M_R^{offset} are allowed to float freely in the fit.

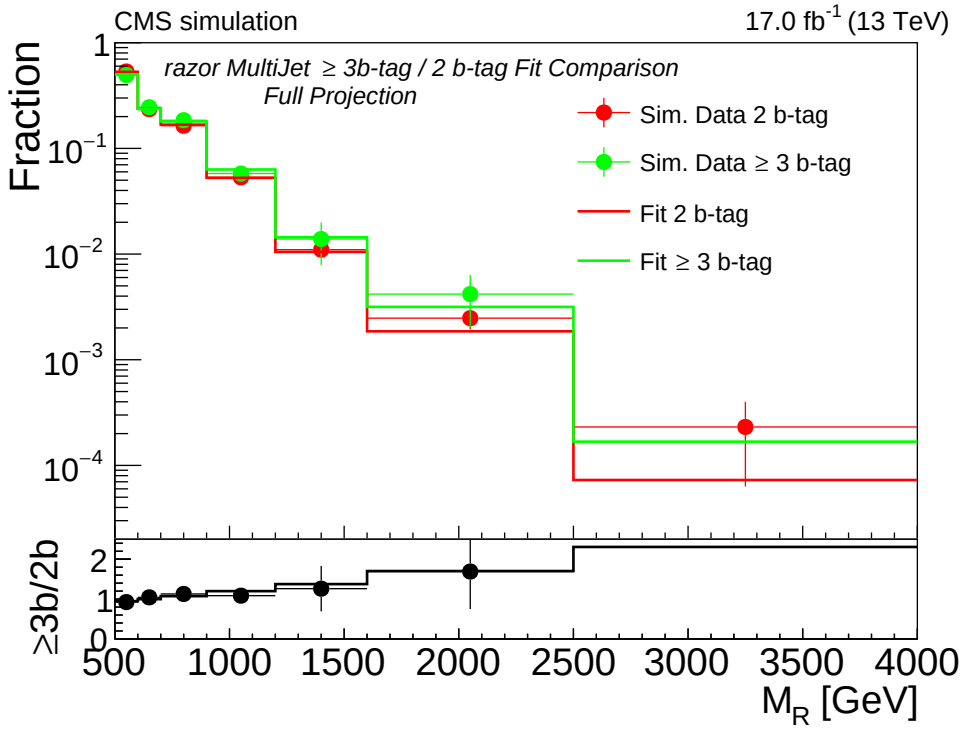


Figure 6.6: Comparison in MC of the fitted razor function shape in the Multijet $2b$ and ≥$3b$ categories. The ratio of the $3b$ fit to the $2b$ fit is plotted in the bottom panel.

The background composition of each search category is illustrated in Figure 6.7. Some notable features are:

- $t\bar{t}$+jets production is the dominant background in the b-tagged leptonic and hadronic event categories;
- $Z(\rightarrow \nu\nu)$+jets production is the dominant background in the Multijet 0 b-tag category;
- $W(\rightarrow \ell\nu)$+jets production contributes significantly in the 0 b-tag leptonic and hadronic event categories;
- QCD multijet production is a small (10-20%) background in the Multijet category for all b-tag multiplicities.

The individual SM backgrounds will be discussed in greater detail in the context of the MC-based background prediction (Chapter 7).

Selection cut optimization

The value of the $\Delta\phi_R$ cut in the Multijet search region, and that of the m_T cut in the Muon and Electron regions, are optimized by considering the expected exclusion limit on a number of SUSY simplified models. Example results are shown in Figure 6.8. We find that the expected limit in the Multijet category is optimized when the $\Delta\phi_R$ cut is 2.8. On the other hand, applying any $\Delta\phi_R$ cut in the Electron and Muon categories is seen to hurt the limit. The signal sensitivity of the one-lepton categories is optimized when the m_T cut is 120 GeV. These conclusions are seen to hold for gluino simplified models having both small and large mass splittings between the gluino and the LSP.

6.6 Validation of the fit

The fit function is validated using a mock dataset consisting of MC events sampled from the major SM background processes (details on the background MC samples are provided in the next chapter). Each physics process is represented in the dataset proportionally to its cross section. The size of the mock dataset corresponds to 17 fb^{-1}, the smallest effective integrated luminosity among the available MC samples.

The MC events are selected and placed into the appropriate search regions, and both sideband and full fits are performed in each region to confirm that the fit function

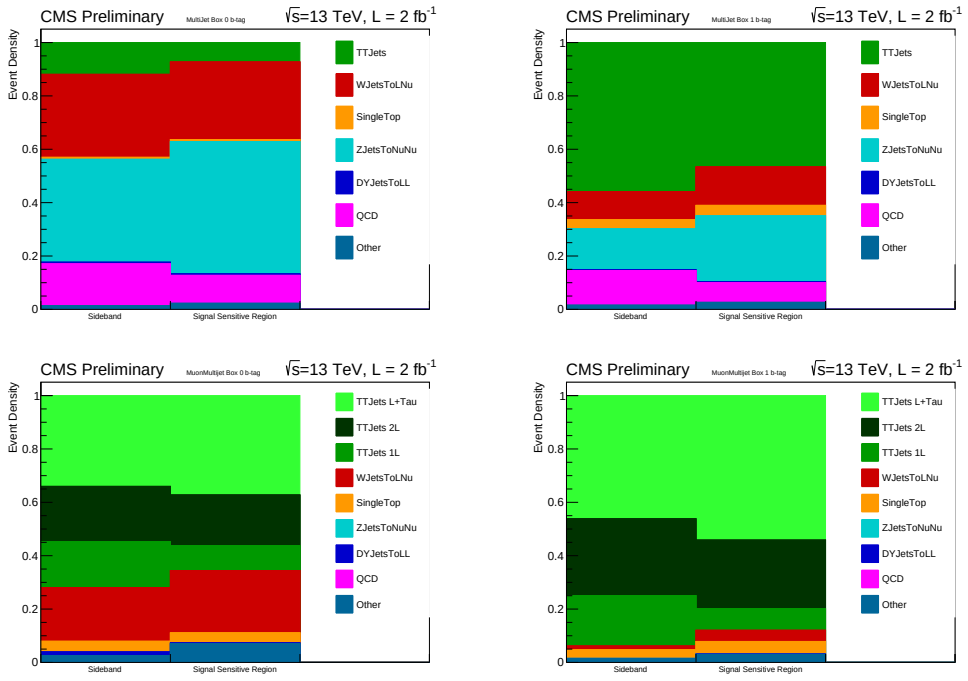


Figure 6.7: Fractional composition of the background in the sideband and extrapolation region of the 0 b-tag (left column) and 1 b-tag (right column) subcategories of the Multijet (top) and Muon Multijet (bottom) search categories. The single top quark, Z($\rightarrow \ell\ell$)+jets, and rare process backgrounds (multiboson and $t\bar{t}$+V) are indicated in addition to the main backgrounds discussed in the text. The categories with 2 and $\geq$ 3 b-tags contain mainly $t\bar{t}$+jets background and are not displayed here.

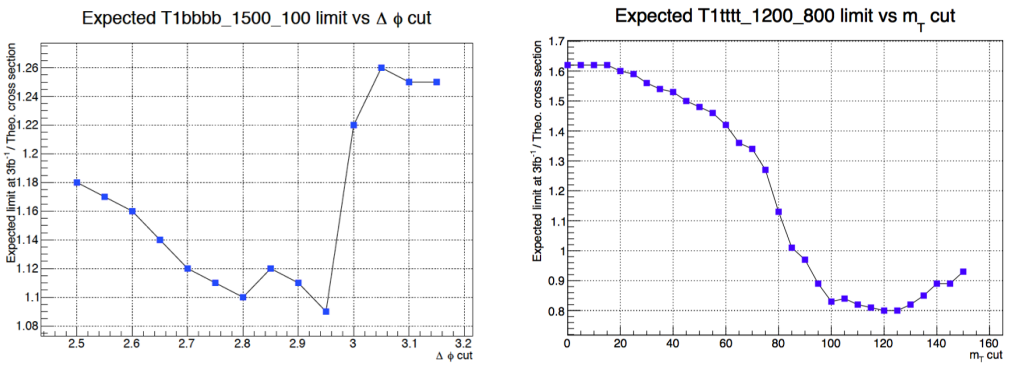


Figure 6.8: Left: expected limit on the T1bbbb model, with gluino and LSP masses set to 1500 GeV and 100 GeV, respectively, as a function of the $\Delta\phi_R$ cut. The y-axis values indicate the excluded cross section divided by the theoretical cross section. Right: expected limit on the T1tttt model, with gluino and LSP masses set to 1200 GeV and 800 GeV, respectively, as a function of the m_T cut.

and its uncertainty adequately describe the background shape. The goodness of the fit is evaluated by comparing the fitted function with the background MC. The

fitted shapes are seen to describe the background within uncertainty in all analysis regions. Example fits to MC are illustrated in Figure 6.9.

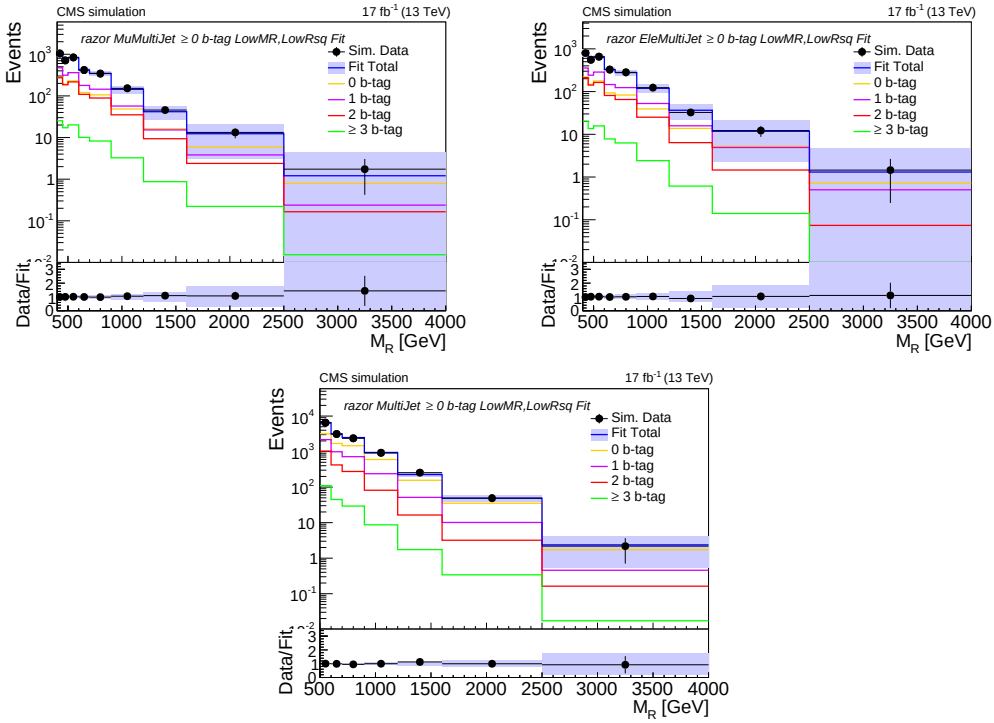


Figure 6.9: Sideband fits to the MC mock dataset in the Muon Multijet (left), Electron Multijet (right), and Multijet (bottom) event categories, projected onto M_R. The yellow, magenta, red, and green lines show the contributions from the 0, 1, 2, and $\geq$3 b-tag fit functions, and the blue line indicates the sum of the four contributions.

To test the robustness of the fit method to variations in the background composition, we vary the fraction of each background physics process up and down by specified amounts and repeat the fit. We test the following variations:

- 30% upward and downward variations of the $t\bar{t}$+jets, W($\rightarrow \ell\nu$)+jets, and Z($\rightarrow \nu\nu$)+jets backgrounds;
- 50% and 100% upward variations of the QCD multijet background;
- 100% upward and 50% downward variations of the rare process backgrounds (multiboson and $t\bar{t}$+V production).

The fit function describes the background well under all of the variations tested, and the fit predictions do not change significantly with the composition of the background.

Signal injection test

We perform a signal injection test to check that the full signal-plus-background fit can detect a signal and accurately extract its strength. To do this, we use the background model obtained from the best fit to the MC simulation to generate a dataset corresponding to 4 fb^{-1} of integrated luminosity, and inject simulated SUSY events with a specified cross section into it. We fit this simulated dataset to estimate the injected signal strength. This procedure is repeated several times to build up a distribution of estimated signal strength values.

The test is performed for several different SUSY models and for a range of different signal cross sections. In Figure 6.10 we show plots of extracted versus injected signal strength, normalized to the theoretical cross section, for two models of gluino production. The fitted cross sections match the injected cross sections closely and do not exhibit bias.

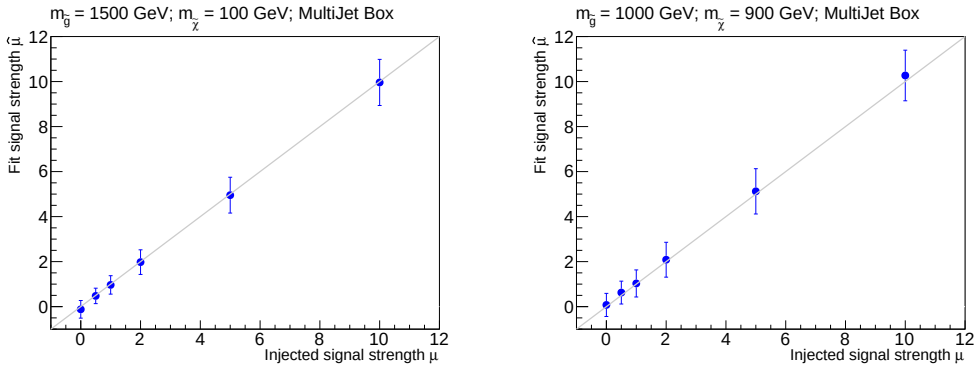


Figure 6.10: Fitted vs. extracted signal strengths for the T1bbbb simplified model with gluino and LSP masses set to 1500 and 100 GeV (left plot), or 1000 and 900 GeV (right plot). The error bars show the standard deviations of the extracted signal strengths over the ensemble of toy experiments.

Sideband fit bias study

We test for bias in the predictions of the sideband fit, which does not have access to the data information in the high M_R and R^2 region. To do this, we conduct many pseudo-experiments using the following procedure:

1. Generate a mock dataset of size 2.1 fb^{-1} by sampling (M_R, R^2) pairs from the best fit function obtained with the full-region fit.

2. Perform both sideband and full fits to the mock dataset, and compute the percent difference between the sideband fit and full fit predicted yields in a large aggregate region: $M_R > 700$ GeV and $R^2 > 0.41$ for the zero-lepton category, and $M_R > 600$ GeV and $R^2 > 0.25$ for the one-lepton categories.

We find that the yields predicted by the sideband fit are on average 5-20% smaller than those from the full fit in all analysis categories. We enlarge the systematic uncertainty on the sideband fit yields to account for this small bias. This has a minimal effect on the search sensitivity; the bias is small compared with the size of the systematic uncertainty on the yield, which varies from 40% to 200% depending on the category.

6.7 Signal region predictions and uncertainties

The sideband fit prediction and the number of observed event counts in each analysis bin are shown in Figures 6.11-6.16. For each analysis category, the results are presented in each b-tag subcategory in an unrolled format that displays the M_R-R^2 plane in a series of one-dimensional slices. The significance $n\sigma$ in each analysis bin is shown in the lower panel of each plot. The data are consistent with the predicted background and no signal-like excess is observed.

While the unblinded data are shown in Figures 6.11-6.16, it should be noted that we did not look at the data in the extrapolation region until we had performed the MC-based background prediction and established that the two methods (fit-based and MC-based) gave results consistent with each other.

We turn to the description of the MC-based background prediction and the final interpretation of the search results in Chapter 7.

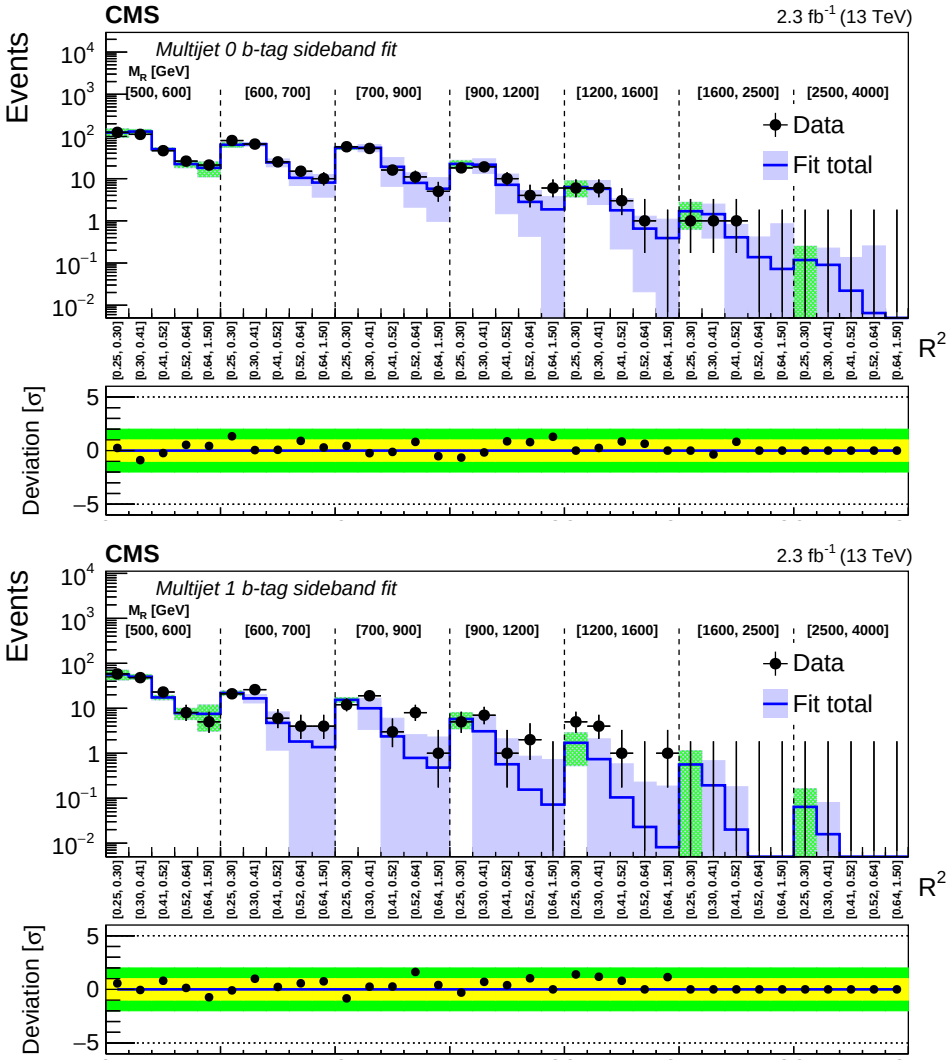


Figure 6.11: Predicted and observed event counts in the Multijet category for the 0 (top) and 1 (bottom) b-tag subcategories. The M_R-R^2 plane is shown in unrolled format; vertical dashed lines denote the boundaries of bins in M_R. In the upper panels, colored bands represent the uncertainty on the fitted function shape. The band is colored green to indicate sideband bins and blue to indicate the extrapolation region. In the lower panels is shown the $n\sigma$ significance of the deviation of the data from the fit. The yellow and green bands indicate the 1- and 2-sigma significance levels.

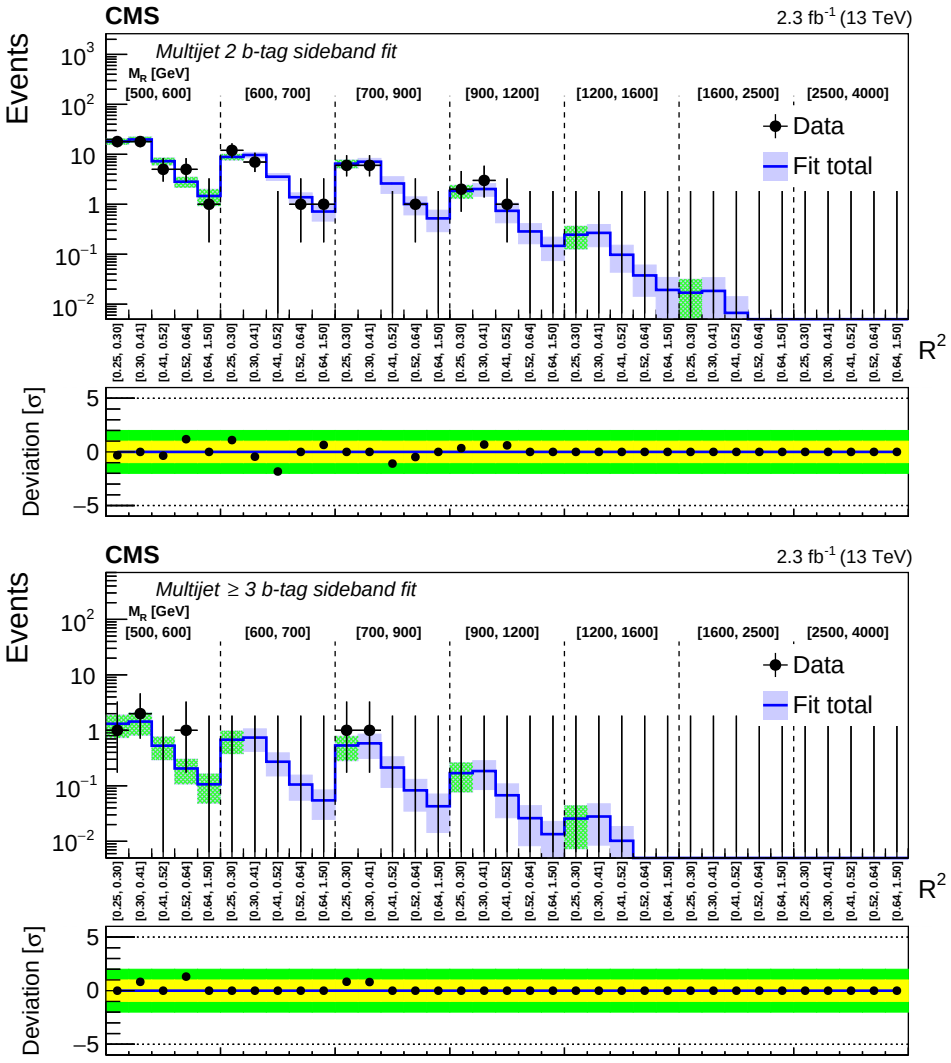


Figure 6.12: Predicted and observed event counts in the Multijet category for the 2 (top) and ≥3 (bottom) *b*-tag subcategories. The layout of the plot is explained in Figure 6.11.

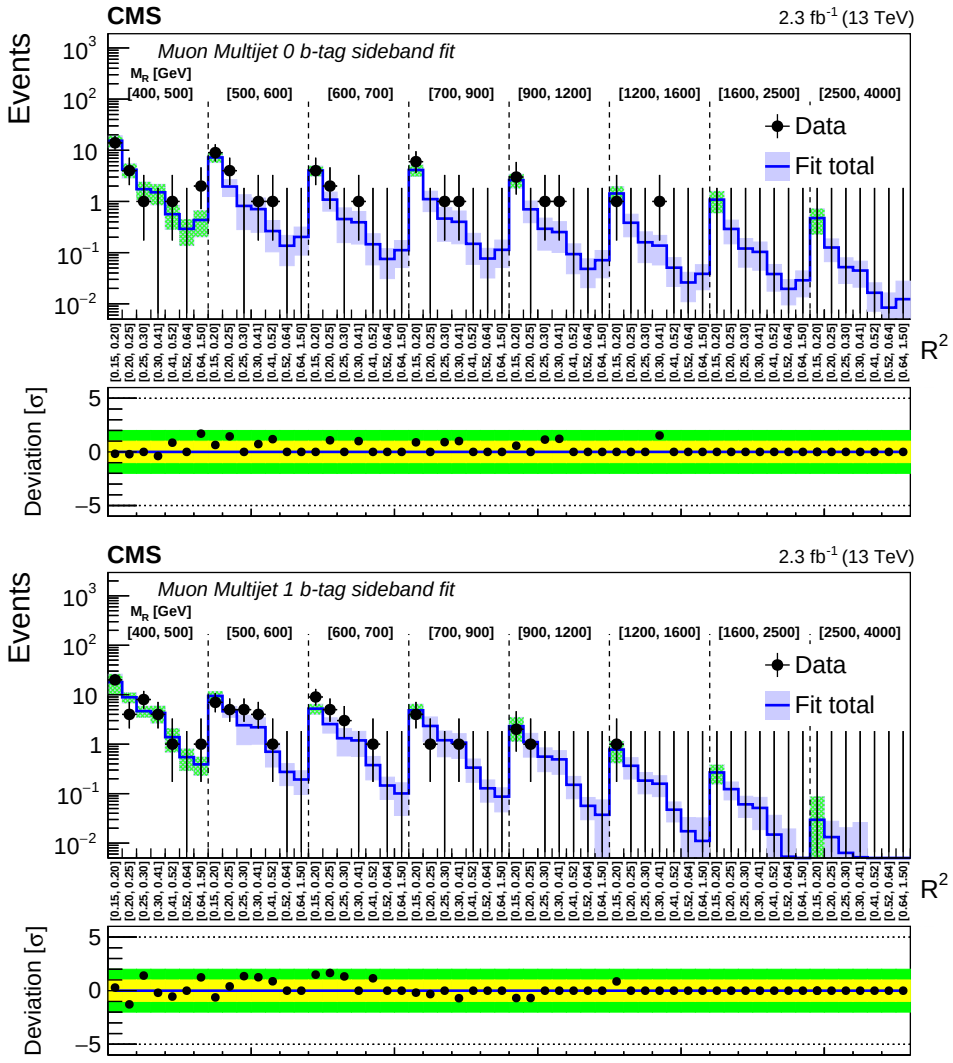


Figure 6.13: Predicted and observed event counts in the Muon Multijet category for the 0 (top) and 1 (bottom) *b*-tag subcategories. The layout of the plot is explained in Figure 6.11.

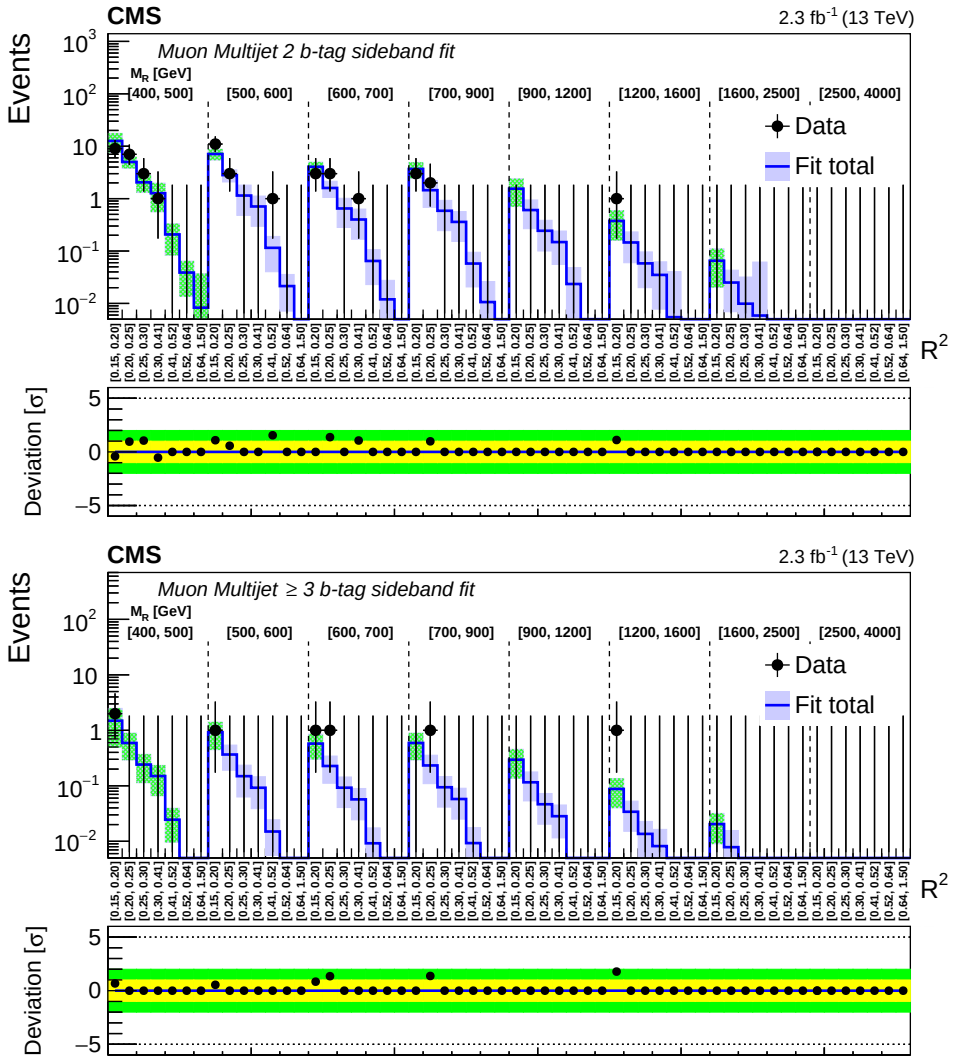


Figure 6.14: Predicted and observed event counts in the Muon Multijet category for the 2 (top) and ≥3 (bottom) *b*-tag subcategories. The layout of the plot is explained in Figure 6.11.

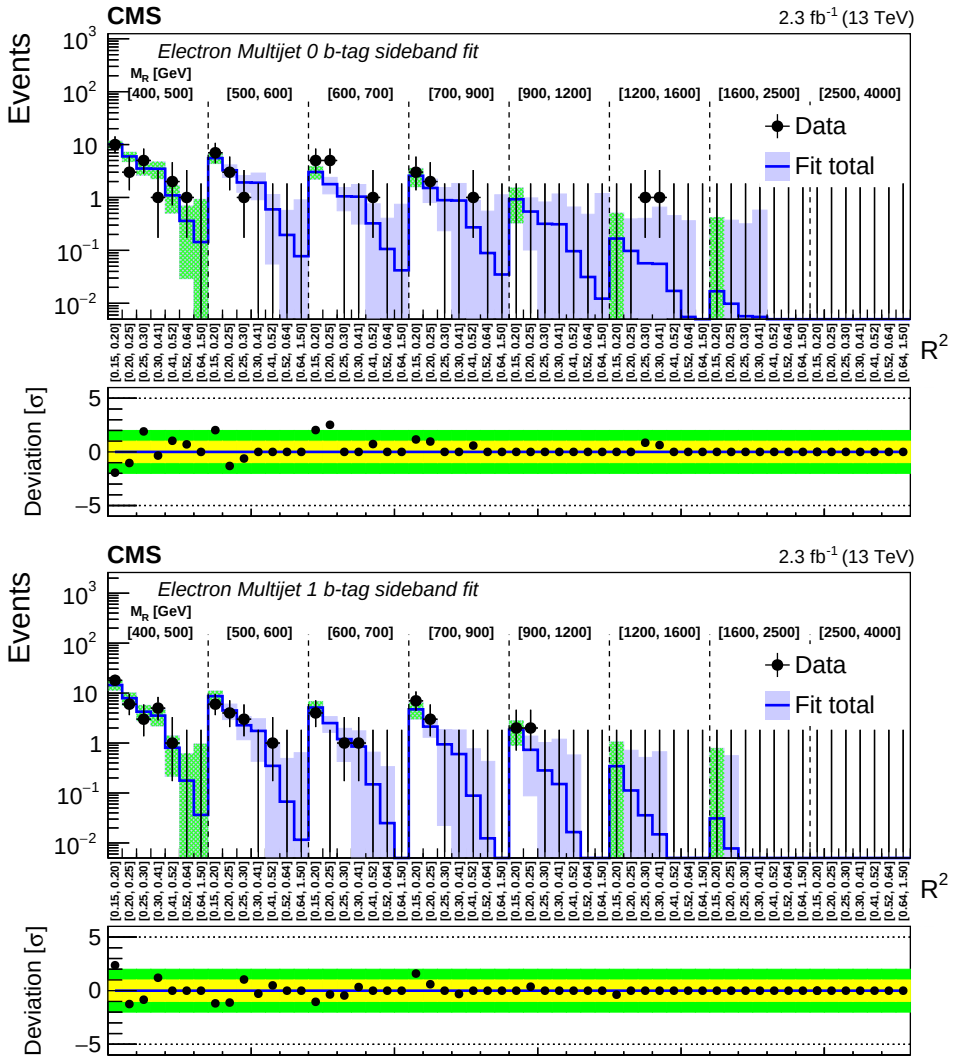


Figure 6.15: Predicted and observed event counts in the Electron Multijet category for the 0 (top) and 1 (bottom) *b*-tag subcategories. The layout of the plot is explained in Figure 6.11.

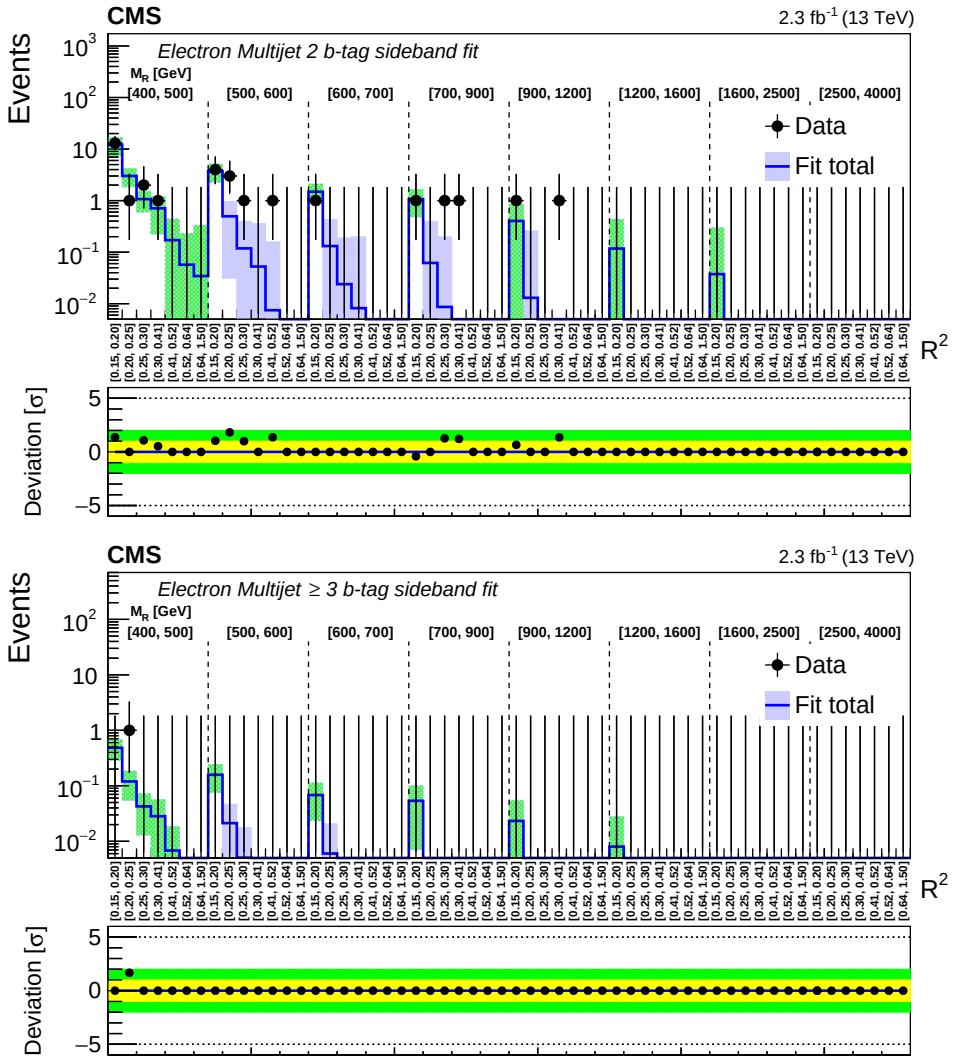


Figure 6.16: Predicted and observed event counts in the Electron Multijet category for the 2 (top) and ≥3 (bottom) *b*-tag subcategories. The layout of the plot is explained in Figure 6.11.

Chapter 7

MONTE CARLO-BASED RAZOR SEARCH ON 2015 DATA

In this chapter we describe the MC-based background prediction strategy used in the 2015 razor search for SUSY. The starting point of this strategy is a collection of MC samples that simulate the relevant SM background processes. The MC simulation can suffer from various deficiencies, which must be identified and corrected using control samples in data. Uncertainties on the method are quantified via closure tests that check various aspects of the background modeling.

The MC-based search uses the same trigger and event selection criteria, analysis search categories, and binning as the fit-based search. Please refer to Chapter 6 for the details on these.

For the 2015 search, the MC-based background prediction is regarded as the primary one, with the fit method serving as a cross-check. The MC-based search has slightly better expected signal sensitivity, especially for 'compressed' SUSY models, which do not exhibit strong peaking behavior in M_R.

7.1 Paradigm for MC-assisted, data-driven search

MC simulations of pp collision events may not perfectly mimic real data. They can suffer from a variety of defects, such as

- Mismodeling of the hadronic recoil in events with radiated jets, due to imperfect parton shower modeling or higher-order QCD effects;
- Mismodeling of the identification probabilities for b-jets, leptons, and other physics objects;
- Absence of detector noise and other data quality issues.

Some MC deficiencies are studied in detail by specialized groups in CMS, who provide recommended procedures to remedy the mismodeling. Usually this consists of reweighting the MC events according to a set of correction factors. Other types of MC deficiencies are analysis specific, showing up in particular phase space regions

or in the distributions of particular kinematic variables. In these cases we must derive our own corrections in a way that does not bias the measurement of a potential SUSY signal.

MC-driven searches in CMS commonly address the challenge of MC mismodeling by defining a *control region* thought to be nearly empty of potential new physics signals while being kinematically similar to the search region. The number of background events in the search region can be predicted as

$$N_{S,\mathrm{Data}} = \left(\frac{N_{S,\mathrm{MC}}}{N_{C,\mathrm{MC}}}\right) \times N_{C,\mathrm{Data}}, \tag{7.1}$$

where $N_{S,\mathrm{Data}}$ and $N_{C,\mathrm{Data}}$ are the number of data events in the search region and control region, and $N_{S,\mathrm{MC}}$ and $N_{C,\mathrm{MC}}$ are the corresponding quantities in MC. The quantity in parentheses is a transfer factor that relates the yield in the search region to that in the control region. The basic assumption of this method is that the MC accurately models the transfer factors (i.e., $N_{S,\mathrm{MC}}/N_{C,\mathrm{MC}} = N_{S,\mathrm{Data}}/N_{C,\mathrm{Data}}$).

Rearranging Eq. 7.1 suggests an equivalent method that we find more interpretable:

$$N_{S,\mathrm{Data}} = \left(\frac{N_{C,\mathrm{Data}}}{N_{C,\mathrm{MC}}}\right) \times N_{S,\mathrm{MC}}. \tag{7.2}$$

The quantity in parentheses is now a correction factor $N_{\mathrm{Data}}/N_{\mathrm{MC}}$ that directly parameterizes the degree of MC mismodeling. Instead of working with the transfer factors N_S/N_C, we use the control region to compute these $N_{\mathrm{Data}}/N_{\mathrm{MC}}$ correction factors. The correction factors are applied to the MC prediction in the search region to arrive at the final background prediction. The assumption of the method (equivalent to the one stated above) is that the mismodeling represented by $N_{\mathrm{Data}}/N_{\mathrm{MC}}$ is the same in the control region and the search region.

We perform the MC-driven background prediction for the razor search using the method suggested by Eq. 7.2 to predict the contributions from the main SM background processes. An exception is the QCD multijet background, for which we develop an entirely data-driven prediction method; this is needed because the available MC samples for this process do not have adequate statistics in the phase space region of interest.

7.2 Primary background processes

The search regions under consideration are dominated by four main SM background processes (see Figure 6.7):

- $t\bar{t}$+jets production, when the W boson from at least one of the top quarks decays leptonically. The escaping neutrino from the W boson decay causes the E_T^{miss} and therefore R^2 to be nonzero. These events enter the one-lepton analysis categories when the lepton is identified, and the zero-lepton categories when it is misidentified or out of acceptance. The events predominantly enter the b-tagged search categories, due to the presence of two b quarks from the top quark decays.

- W+jets production, where the W boson decays leptonically (denoted W($\rightarrow \ell\nu$)+jets). These events mainly populate the search categories with no b-tagged jets.

- Production of Z bosons decaying to neutrinos, denoted Z($\rightarrow \nu\nu$)+jets. The neutrinos from the Z decay create $E_T^{miss}$when the Z recoils off of ISR jets. These events populate the zero-lepton search categories, especially those with no b-tagged jets.

- QCD multijet production, when mismeasurement of one or more jet momenta creates spurious E_T^{miss}.

Control regions are used to derive predictions for each of these main background processes. The search regions additionally have events from single top production, and from DY($\rightarrow \ell\ell$)+jets production where one lepton is misidentified. Other rare background processes include production of two or three W or Z bosons, and production of two top quarks in association with a W or Z boson. The small contributions from these processes to the search regions is predicted using MC simulation.

7.3 MC simulation and reweighting

Simulation of the main SM backgrounds, and of γ+jets and Z($\rightarrow \ell\ell$)+jets production, is performed using the MadGraph 5 package [31]. Multiboson, $t\bar{t}$+V, and s-channel single top quark production are simulated using aMC@NLO, and single top production in the t-channel and in association with a W boson are simulated using Powheg [125]. Simulation using MadGraph is performed at leading order (LO) in QCD; Powheg and aMC@NLO simulation are performed at next-to-leading order (NLO). All MC generators are interfaced with Pythia for fragmentation and parton showering [33]. Events are passed through a GEANT4-based model of CMS [35] for simulation of the detector response. The NNPDF3.0LO and

NNPDF3.0NLO parton distribution functions are used with the LO and NLO event generators, respectively [126].

Event reweighting

MC events are reweighted in order to correct for known differences between the simulation and data. These include:

- Different distributions of the number of pileup vertices;
- Mismodeling of the lepton identification efficiency;
- Mismodeling of the trigger efficiency;
- Mismodeling of the b-tag efficiency and mistag rate.

We estimate the lepton identification and trigger efficiencies on samples of $Z\rightarrow ee$ and $Z\rightarrow \mu\mu$ candidate events using the tag and probe method [127]. An outline of the method is as follows (where $\ell = e$ or μ):

1. Consider each event having an $\ell\ell$ pair with invariant mass consistent with a Z boson decay. At least one of the leptons must pass the desired identification cuts (or trigger requirements, in the case of the trigger efficiency calculation). The passing lepton is called the 'tag' and the other lepton is the 'probe'.
2. Divide the events into 'passing' and 'failing' samples according to whether the probe lepton passes the desired identification cuts. Perform a likelihood fit to the $\ell\ell$ invariant mass distribution in both passing and failing samples to estimate the number of true Z events, N_{pass} and N_{fail}, in each sample.
3. Compute the lepton selection efficiency as $N_{pass}/(N_{pass} + N_{fail})$.

The procedure is carried out on data and the measured efficiencies are compared with those obtained in the MC simulation. The data/MC efficiency ratio is computed in bins of the lepton p_T and η. MC events with leptons are reweighted according to these efficiency factors to correct for the data-MC discrepancy. These corrections are at the level of a few percent or less in all p_T and η regions. Examples of efficiencies measured using the tag and probe procedure are shown in Figure 7.1.

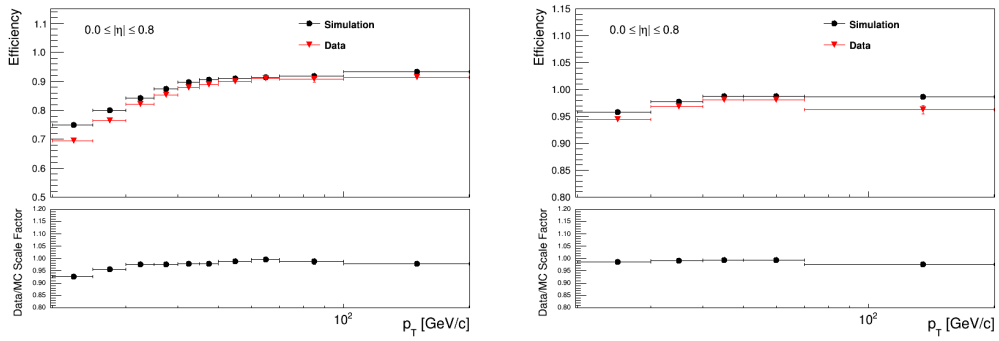


Figure 7.1: Comparisons of the lepton selection efficiency measured in data using Tag and Probe (red points) with that obtained in MC (black points). Left: tight electron identification efficiency in the range $0 < |\eta| < 0.8$, as a function of p_T. The data/MC ratio is shown in the bottom panel. Right: the same, for muons. The data/MC ratio is shown in the bottom panel of each plot.

7.4 Estimation of $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets backgrounds

We now follow the strategy outlined in Section 7.1 to estimate the SM background in the search region. We consider the $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets backgrounds first. The distributions of M_R, R^2, and N_{jets}, which form the basis of event selection and categorization in the search, may be affected by possible mismodeling of the hadronic recoil by the MC. We would like to obtain MC predictions that faithfully model these distributions.

Following the strategy outlined in Section 7.1, we correct these distributions using data control samples that are pure in $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets events. For this we select events having one lepton and a low value of m_T. The m_T cut excludes most potential SUSY signal events from selection and removes any overlap with the one-lepton search regions.

The main assumptions in what follows are that the MC mismodeling of the hadronic recoil is similar at low and high m_T, and that it is uncorrelated with the (mis)identification of a lepton. These assumptions are reasonable because m_T and other variables related to the lepton are properties of the top quark or W boson decay, not the hadronic shower.

One-lepton control region selection

We select events for the one-lepton control region if they pass at least one inclusive single lepton trigger and have the following characteristics:

- At least one electron or muon passing tight identification criteria;
- $30 < m_T < 100$ GeV;
- $E_T^{miss} > 30$ GeV;
- $M_R > 300$ GeV and $R^2 > 0.15$.

The lower cuts on E_T^{miss} and m_T reduce contamination from the QCD multijet background. Note that no explicit cut on the number of jets is applied.

We split this control sample into two subsamples, to isolate the $t\bar{t}$+jets process and the W($\to \ell\nu$)+jets process separately. The events with no b-tagged jets form the W+jets one-lepton control sample. This sample is 80% pure in W($\to \ell\nu$)+jets events, with the dominant background being $t\bar{t}$+jets. The events with b-tagged jets form the $t\bar{t}$ one-lepton control sample. It is 80% pure in $t\bar{t}$+jets events, with W($\to \ell\nu$)+jets and single top events forming most of the background.

Razor variable correction

The selected one-lepton events are divided into bins of M_R and R^2. The data and MC yields are shown in one dimension in Figure 7.2, and in two dimensions in unrolled format in Figure 7.3. The plots indicate that the M_R and R^2 distributions in MC have longer tails in MC than in data – this is the MC deficiency that we wish to correct.

To derive corrections to the $t\bar{t}$+jets MC, we first subtract the estimated number of non-$t\bar{t}$+jets events from the data yield in each bin of the $t\bar{t}$ one-lepton control region. Then we take the ratio of the remaining data counts to the $t\bar{t}$+jets MC predicted yield. Thus, the MC correction factor for bin i is

$$\kappa_{i,t\bar{t}} \equiv \frac{N_{i,\text{Data}} - N_{i,\text{non-}t\bar{t}\,\text{MC}}}{N_{i,t\bar{t}\,\text{MC}}}. \tag{7.3}$$

The computed scale factors and their statistical uncertainties are displayed in the upper part of Figure 7.4. A systematic uncertainty is also assigned on the corrections based on the modeling of the subtracted background processes. This is subdominant compared with the statistical uncertainty; we approximate it as a 20% error on the subtracted contributions $N_{i,\text{non-}t\bar{t}\,\text{MC}}$

After deriving the corrections to the $t\bar{t}$+jets MC, we turn to the W+jets one-lepton control region. We perform the same procedure as above, except that we first correct

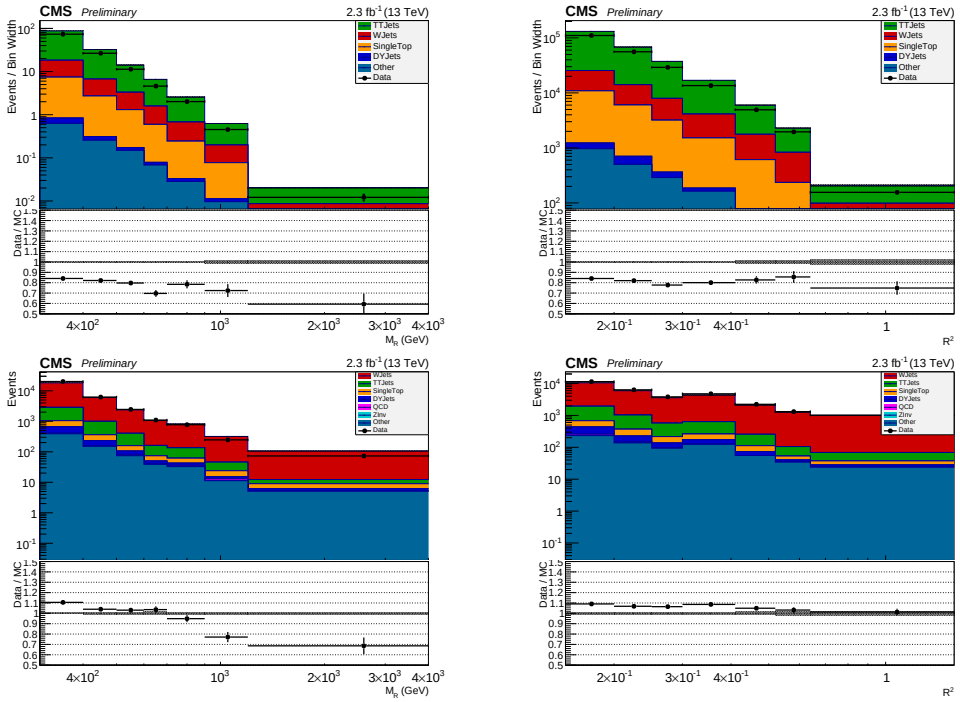


Figure 7.2: Distributions of M_R (left column) and R^2 (right column) in the $t\bar{t}$ (top row) and W+jets (bottom row) one-lepton control samples. The bottom panel in each plot shows the ratio of data to MC. In the W+jets control region, the $t\bar{t}$+jets MC prediction has been corrected using scale factors derived in the $t\bar{t}$ control region.

the estimated amount of $t\bar{t}$+jets contamination using the factors $\kappa_{i,t\bar{t}}$ that were just derived. The scale factors obtained in this way are shown in the bottom part of Figure 7.4.

Jet multiplicity correction

The one-lepton control region is defined with no reference to the number of selected jets in the event. This provides the control region with higher statistics (an order of magnitude more events) than if we imposed the $N_{\text{jets}} \geq 4$ cut used in the search region. However, the corrections to the M_R and R^2 distributions needed for events with four or more jets may be different from those we derive here.

We check explicitly for this possibility by restricting our attention to events with four or more jets and applying the M_R-R^2 corrections derived for the $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets MC. It is seen that an additional flat scale factor of 0.90 ± 0.03 is needed to normalize the MC prediction to the data in this restricted control sample. We take this scale factor as a further correction to the $t\bar{t}$+jets and W($\rightarrow$

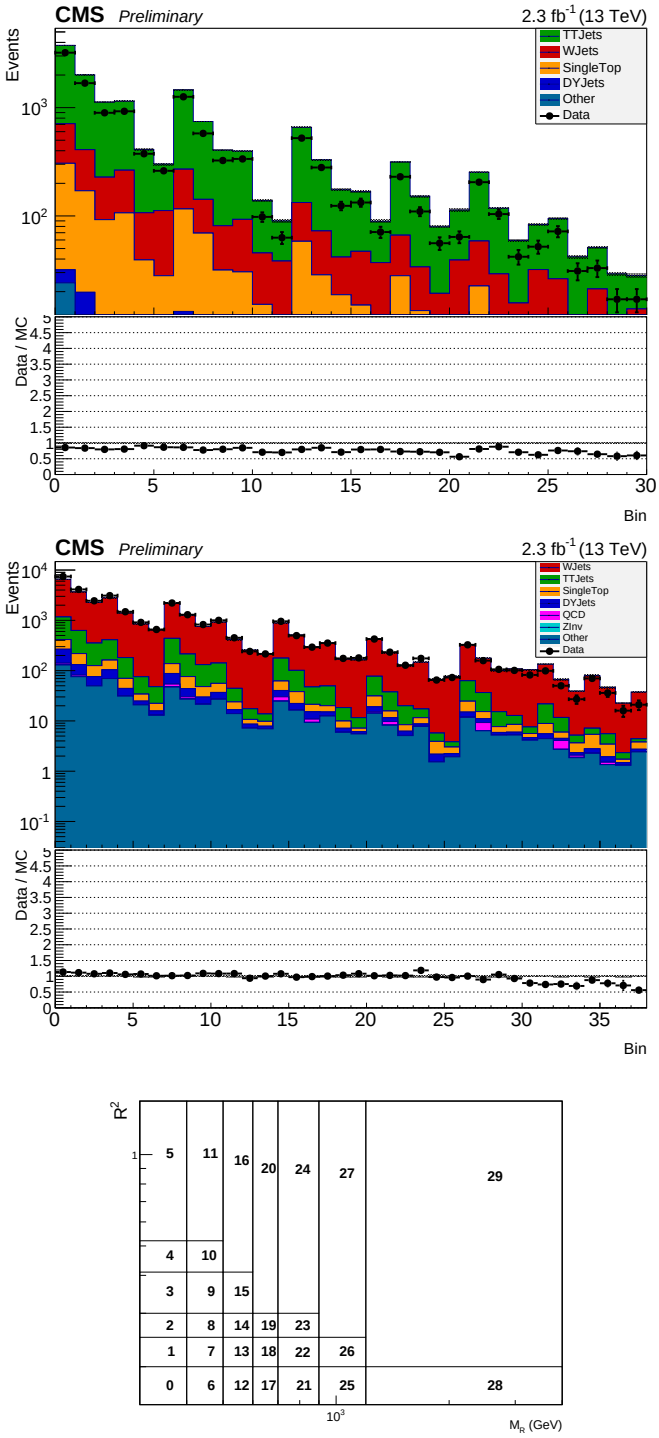


Figure 7.3: Two-dimensional M_R-R^2 distribution in the $t\bar{t}$ (top) and W+jets (middle) one-lepton control regions, displayed in unrolled format using the given bin mapping (bottom). In the W+jets control region, the $t\bar{t}$+jets MC prediction has been corrected using scale factors derived in the $t\bar{t}$ control region.

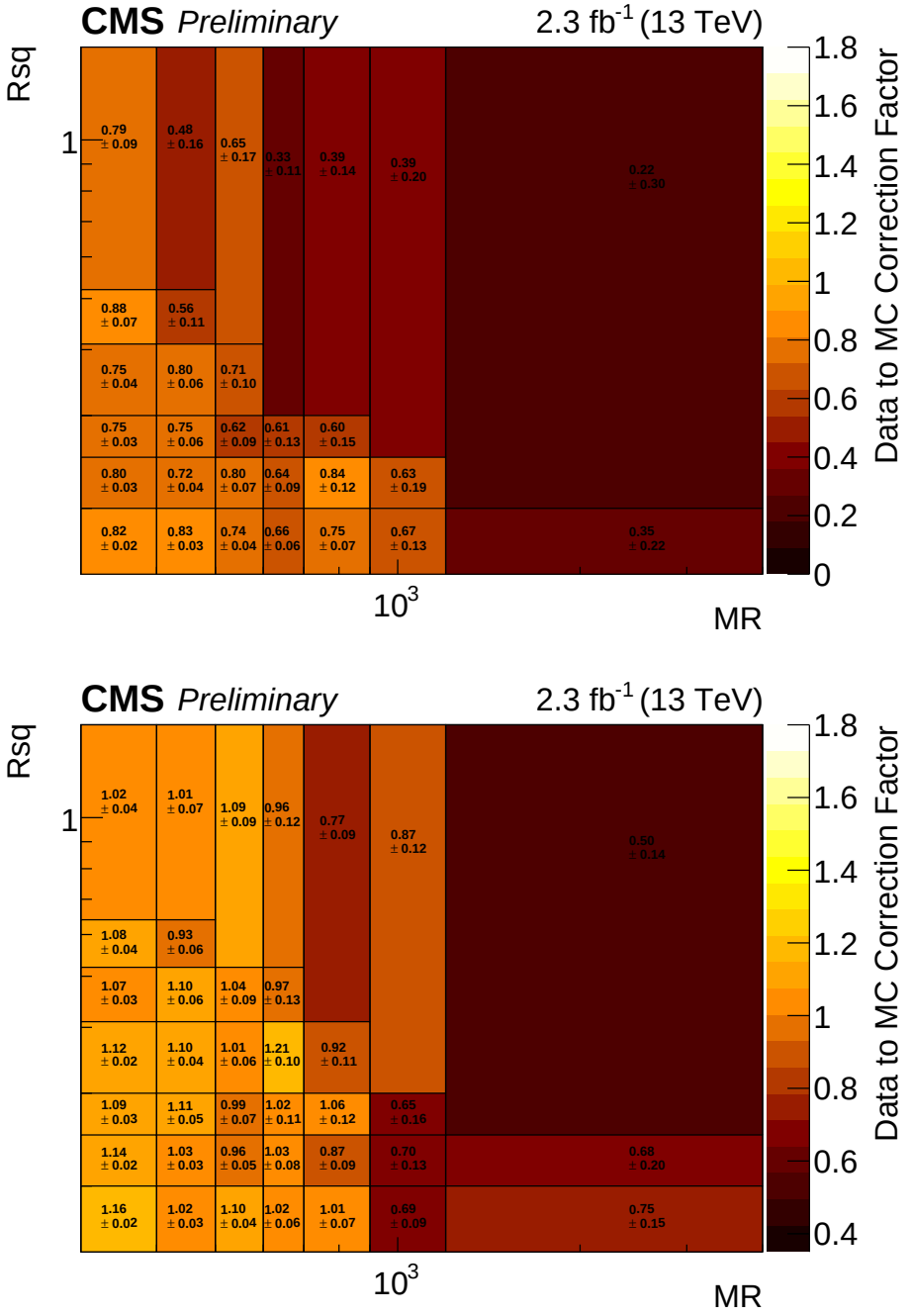


Figure 7.4: MC correction factors derived in bins of M_R and R^2 for the $t\bar{t}$+jets (top) and W($\rightarrow \ell\nu$)+jets (bottom) simulated samples. Uncertainties displayed are statistical.

$\ell\nu$)+jets MC. The M_R-R^2 distribution in the $\geq$ 4 jet region after applying this correction is shown in Figure 7.5. We see no significant mismodeling by the corrected MC, which suggests that factorizing the M_R-R^2 and N_{jets} corrections in this way is

appropriate.

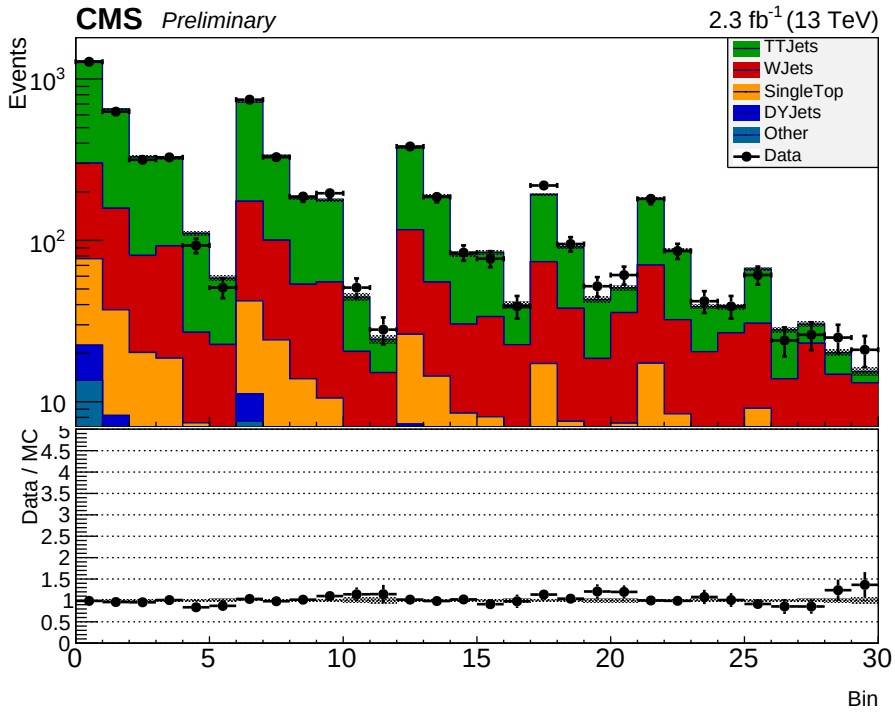


Figure 7.5: Unrolled M_R-R^2 distribution in the one-lepton control region, for events having four or more selected jets, after applying corrections derived in bins of M_R, R^2, and the number of selected jets.

Closure test in one-lepton control region

We perform a more detailed check of the data-MC agreement in the one-lepton control region by examining the one-dimensional M_R and R^2 distributions for each b-tag multiplicity. The M_R distributions are shown in Figure 7.6 for the data and the corrected MC. Discrepancies between data and MC are taken to represent possible MC mismodeling, and systematic uncertainties are assigned to the $t\bar{t}$ and W+jets background predictions according to the level of agreement.

Check of dilepton $t\bar{t}$+jets modeling

The corrections derived in the one-lepton control region for the $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets MC are applied to the MC background prediction in both the zero- and one-lepton analysis search regions. One situation in which this approach could lead to inaccuracies is if the MC mismodeling of dilepton $t\bar{t}$+jets events is significantly different from that of single-lepton $t\bar{t}$+jets events.

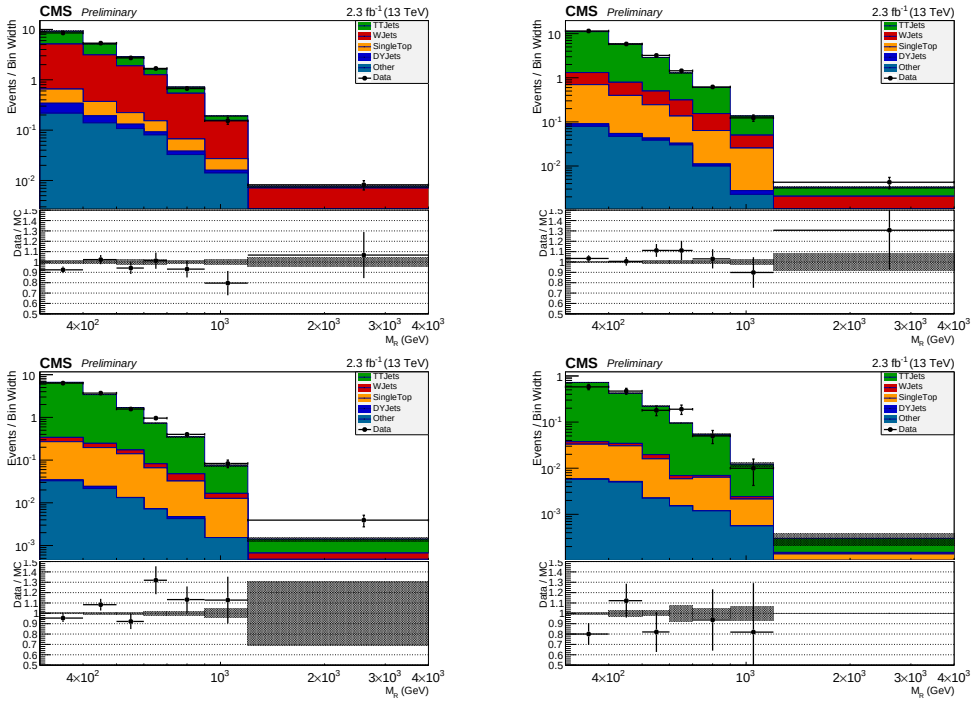


Figure 7.6: M_R distributions for events in the one-lepton control sample with at least four selected jets and 0 (upper left), 1 (upper right), 2 (lower left), and ≥3 (lower right) b-tagged jets. The $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets MC events have been reweighted using the corrections derived for the M_R-R^2 and N_{jets} distributions.

In single-lepton $t\bar{t}$ decays, the $\vec{p}_T^{\,miss}$ and the lepton transverse momentum vector tend to point in the same direction, which results in a small value of m_T (see the definition of m_T in Eq. 6.10). Dilepton $t\bar{t}$ decays tend to have higher values of m_T, because the $\vec{p}_T^{\,miss}$ is the sum of the momentum vectors of two neutrinos and may not align with either lepton. The one-lepton control region requires $m_T < 100$ GeV, while the one-lepton signal regions require $m_T > 120$ GeV. Thus the control region is dominated by single-lepton $t\bar{t}$+jets events, while the one-lepton signal regions have significant contributions from dilepton $t\bar{t}$+jets.

To confirm that the corrections derived for the $t\bar{t}$+jets MC simulation adequately describe dilepton $t\bar{t}$+jets, we select another control sample, this one enriched in dilepton $t\bar{t}$+jets events. This sample contains events passing at least one inclusive single lepton trigger and passing the following selection criteria:

- Two leptons passing tight identification and having $p_T > 30$ GeV;
- Dilepton mass $m_{\ell\ell} > 20$ GeV, and $76 < m_{\ell\ell} < 106$ GeV if the leptons have the same flavor (this suppresses Z decays);

- At least one b-tagged jet;
- $E_T^{miss} > 40$ GeV;
- $M_R > 300$ GeV and $R^2 > 0.15$.

Because our search is restricted to events with zero or one identified lepton, dilepton $t\bar{t}$+jets events entering the analysis are those in which one lepton is misidentified or outside the detector acceptance. We emulate this here by choosing one selected lepton to be 'misidentified' and treated as a jet. We then compute m_T using the other lepton, and apply the search region cut $m_T > 120$ GeV. This procedure is repeated for both leptons, so it is possible for events to enter the control region twice if both leptons pass the m_T cut. Selected $t\bar{t}$+jets MC events are reweighted using the correction factors discussed above.

The agreement between data and MC in this $t\bar{t}$ dilepton control region is assessed both with and without an explicit selection on the number of jets in the event. These results are shown in Figure 7.7 and indicate good agreement between data and the MC prediction. Differences between the data and the MC are treated as systematic uncertainties on the dilepton $t\bar{t}$+jets background in the one-lepton search regions.

Check of lepton veto efficiency modeling

As described in Section 6.5, the lepton veto in the zero-lepton search region identifies electrons and muons with p_T as low as 5 GeV, and taus with p_T as low as 20 GeV. Electron and muon identification efficiencies in MC simulation are corrected using the p_T-dependent scale factors derived using the tag and probe method (see Section 7.3), but these scale factors are not measured below 10 GeV due to the difficulty of selecting a pure sample of Z events at lower lepton momenta. Tag and probe scale factors are not derived at all for tau lepton identification. Mismodeling of the electron or muon veto efficiency below 10 GeV, or of the tau veto efficiency, is therefore a potential risk.

We check for potential mismodeling of this type by selecting a sample of events containing electrons, muons, and taus identified by the veto selection. This sample is defined by the selection requirements of the Multijet search category, except that the lepton veto is inverted so that only events with identified leptons are included. The cut $30 < m_T < 100$ GeV is applied, to remove the overlap with the leptonic search regions. We split the sample into a veto lepton control sample (containing

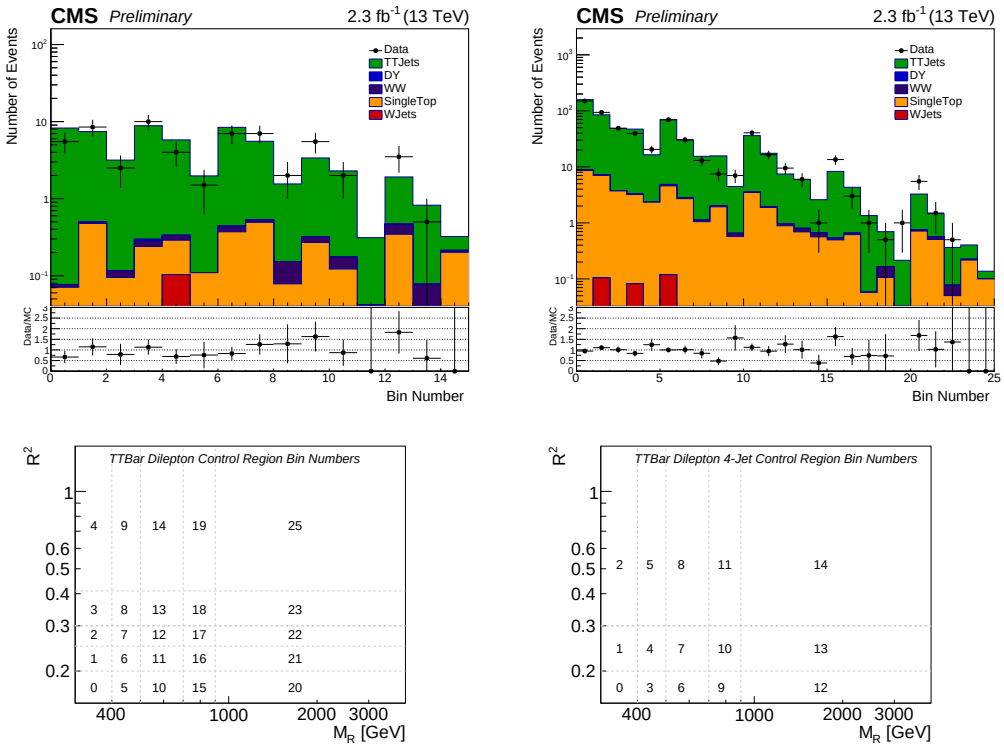


Figure 7.7: Top row: unrolled M_R-R^2 distributions for events in the $t\bar{t}$ dilepton control region having any number of jets (left) or four or more jets (right). The $t\bar{t}$+jets MC has been corrected using the factors derived in the one-lepton control region. Bottom row: bin mapping for the plots in the top row.

events with an identified electron or muon) and a veto tau control sample (containing events with an identified tau).

The agreement between data and MC in these control samples is checked in bins of the lepton (or tau) p_T and η after applying the M_R-R^2 and N_{jets} reweighting to the $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets MC predictions. This comparison is illustrated in Figure 7.8.

We convert discrepancies between data and MC in the control region into appropriate systematic uncertainties on the signal region predictions. To do this, we estimate the number of MC events that would migrate into, or out of, the signal region if the mismodeling were corrected. The estimate is scaled to account for the efficiency of the m_T cut (which is applied in the control region but not the signal region) and the $\Delta\phi_R$ cut (which is applied in the signal region but not the control region):

$$\Delta N_{S,\text{MC}} = \Delta N_{C,\text{MC}} \times \frac{\epsilon_{\Delta\phi_R}}{\epsilon_{m_T}}, \tag{7.4}$$

where $\epsilon_{\Delta\phi_R}$ and ϵ_{m_T} are the efficiencies of the $\Delta\phi_R$ and m_T cuts, measured in bins

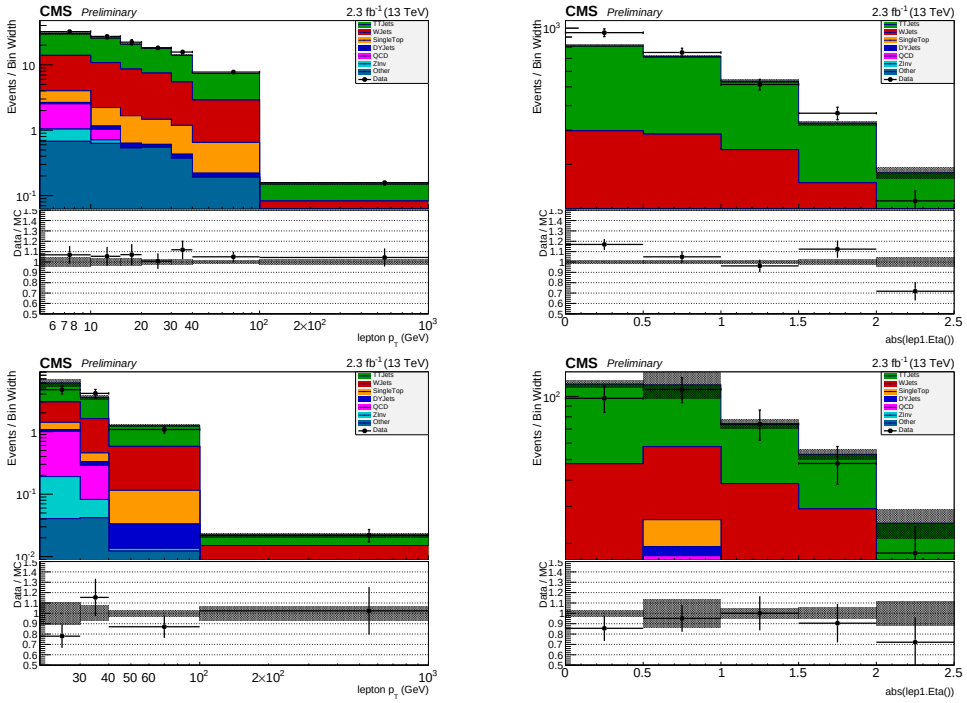


Figure 7.8: Distributions of the lepton p_T (left column) and $|\eta|$ (right column) in the veto lepton (top row) and veto tau (bottom row) control regions. The $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets MC have been corrected using the factors derived in the one-lepton control region.

of lepton p_T.

7.5 Estimation of Z($\rightarrow \nu\nu$)+jets background

The Z($\rightarrow \nu\nu$)+jets background cannot easily be isolated in a dedicated control region, as is done for the $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets processes. The Z decay to neutrinos creates a E_T^{miss} signature in the event, but does not have other features such as leptons that allow it to be conveniently identified. To produce a robust Z($\rightarrow \nu\nu$)+jets background prediction, we exploit this process's kinematic similarity with other physics processes.

Events with dileptonic (ee or $\mu\mu$) Z boson decays are straightforward to select in CMS data, and their kinematics are very similar to those of Z($\rightarrow \nu\nu$)+jets events. We can transform a Z($\rightarrow \ell\ell$)+jets event into a surrogate Z($\rightarrow \nu\nu$)+jets event by pretending that the two leptons from the Z are neutrinos – removing them from the event description, and adding their transverse momentum vectors to the $\vec{p}_T^{miss}$. The main drawback to this approach is that Z($\rightarrow \ell\ell$)+jets events are less common than Z($\rightarrow \nu\nu$)+jets events, with a production cross section three times smaller.

One physics process with a larger cross section that can also act as a surrogate for Z($\to \nu\nu$)+jets is W($\to \ell\nu$)+jets. In W($\to \ell\nu$)+jets events we can add the transverse momentum of the selected lepton to the $\vec{p}_T^{miss}$ and remove it from the event description, simulating an escaping neutrino. Despite the fact that the W boson is charged, the kinematics of W+jets events are similar to those of Z+jets, and the MC simulations of W($\to \ell\nu$)+jets and Z($\to \nu\nu$)+jets may be expected to suffer from similar mismodeling of the hadronic recoil.

Finally, the γ+jets process also exhibits kinematic similarity to Z($\to \nu\nu$)+jets, especially at high boson p_T where the masslessness of the photon is less relevant.

We derive corrections to the Z($\to \nu\nu$)+jets MC, and assign uncertainties on the corrected prediction, using all three of these processes as surrogates. The γ+jets process is used to obtain a primary set of corrections to the Z($\to \nu\nu$)+jets MC. The W($\to \ell\nu$)+jets process provides a second set of corrections, which we use to cross-check the first. The Z($\to \ell\ell$)+jets process, having the highest kinematic similarity with Z($\to \nu\nu$)+jets, is used for a final closure test to check the goodness of the modeling.

Z($\to \nu\nu$)+jets strategy summary

We select three control regions, respectively enriched in γ+jets, W($\to \ell\nu$)+jets, and Z($\to \ell\ell$)+jets events:

1. A sample of events with an identified photon. This is used to derive corrections to the MC in bins of M_R and R^2, and subsequently in N_{jets}.

2. A sample of events with an identified lepton, identical to the W($\to \ell\nu$)+jets-enriched control region studied in Section 7.4 except that the lepton's transverse momentum is added to $\vec{p}_T^{miss}$. This is used to derive a second set of MC corrections. The difference between these and the corrections from γ+jets is taken as a systematic uncertainty on the Z($\to \nu\nu$)+jets prediction.

3. A sample of events with two identified muons or electrons consistent with a Z boson decay. This is used to perform a closure test to confirm that the corrections derived in the other two control regions are appropriate for describing Z events.

γ+jets control region selection

Photons are identified using cuts on ECAL shower-shape variables as well as the HCAL/ECAL energy ratio [128]. Photon candidates that share an energy cluster with a reconstructed electron are vetoed. Additionally, we require the charged-particle isolation sum to be less than 2.5 GeV. For the γ+jets control region we select events having at least one photon satisfying these requirements. The photon's transverse momentum vector is added to the $\vec{p}_T^{miss}$ to emulate a Z boson decaying to neutrinos.

We select events using inclusive single photon triggers, which accept events with a photon having p_T above some threshold. The loosest unprescaled photon trigger has a p_T threshold of 165 GeV; triggers with lower thresholds are prescaled. To obtain a sample that includes photons with lower values of p_T, we divide the range [50 GeV, ∞] in photon p_T into subintervals, and assign one trigger path to select events in each subinterval. In Table 7.1 we list the correspondence between p_T subintervals and photon trigger paths. Each selected event is upsampled by the prescale factor of the trigger used to select it. This yields a smooth p_T spectrum that does not feature kinks from the different trigger prescales. The inclusive photon p_T spectrum is shown in the left panel of Figure 7.9. Here the γ+jets cross section used to normalize the MC has been corrected by the k-factor accounting for higher-order QCD corrections.

Trigger p_T threshold (GeV)	Photon p_T range (GeV)
36	[40, 58]
50	[58, 85]
75	[85, 105]
90	[105, 135]
120	[135, 185]
165	[185, ∞]

Table 7.1: Thresholds of triggers used to select events in different ranges of photon p_T for the γ+jets control sample.

Baseline cuts of $M_R > 400$ GeV and $R^2 > 0.25$ are placed on the selected photon events. The photon p_T spectrum for events passing these cuts is shown in the right panel of Figure 7.9; it is seen that the control sample is dominated by photons with $p_T \gg m_Z$, which are expected to be kinematically similar to Z($\rightarrow \nu\nu$)+jets events.

The physics processes that populate the photon control region are γ+jets, QCD multijet (where a jet is misidentified as a photon), and rare processes including Wγ,

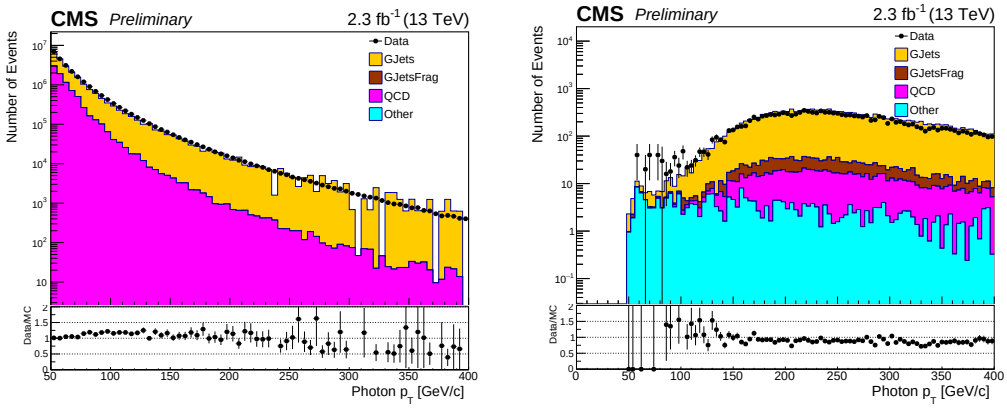


Figure 7.9: Photon p_T distribution obtained in the γ+jets control region before (left) and after (right) applying the baseline cuts $M_R > 400$ GeV and $R^2 > 0.25$. Events in data passing prescaled photon triggers are upsampled to obtain a smooth p_T spectrum.

$Z\gamma$, and $t\bar{t}\gamma$. We model γ+jets and the rare backgrounds using MC simulation. γ+jets events in which the simulated photon has $\Delta R < 0.4$ with respect to any jet are classified as fragmentation photon events; these are considered background and are subtracted from the data yields when computing MC correction factors.

γ+jets purity measurement

We estimate the contribution from QCD jets faking photons using a data-driven fit method. First, we obtain template shapes for the $\sigma_{i\eta i\eta}$ distributions of prompt (i.e., real) photons and of fake photons. The template shape $t_\gamma(\sigma_{i\eta i\eta})$ for prompt photons is obtained from γ+jets MC, and the template shape $t_{QCD}(\sigma_{i\eta i\eta})$ for fake photons is obtained from data by inverting the charged isolation selection cut in the γ+jets control region. We then fit the $\sigma_{i\eta i\eta}$ distribution in data by optimizing the coefficient P_γ in the sum

$$N_\gamma \left[P_\gamma \times t_\gamma(\sigma_{i\eta i\eta}) + (1 - P_\gamma) \times t_{QCD}(\sigma_{i\eta i\eta}) \right], \tag{7.5}$$

where N_γ is the number of events in the distribution. The fitted value of P_γ is the estimated photon purity; the fraction of QCD fakes is $(1 - P_\gamma)$.

The purity fits are performed in bins of M_R and R^2, separately for photons in the barrel region and in the endcaps. Visualizations of the $\sigma_{i\eta i\eta}$ distributions and the best-fit templates are provided in Figure 7.10. The measured purity values are summarized in Figure 7.11. For simplicity, we choose to use $P_\gamma = 0.95 \pm 0.05$ through-

out the whole control region, as all fitted purity values are seen to fall into this range.

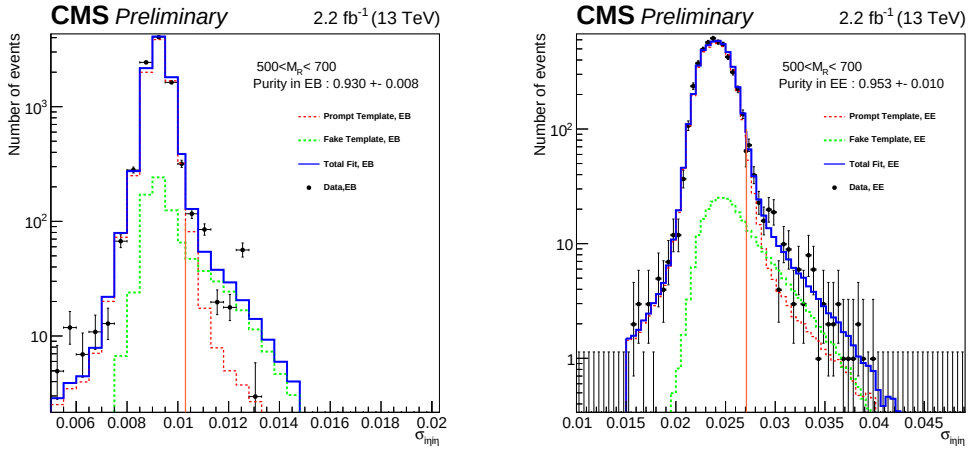


Figure 7.10: Example fits to the $\sigma_{i\eta i\eta}$ distribution in the barrel (left) and endcap (right) to estimate the fraction of prompt photon events. The red and green dashed lines indicate the prompt and fake photon template shapes, and the blue line represents the total fit.

Razor variable and jet multiplicity corrections

The predicted background and observed data in the photon control region are shown in one dimension in Figure 7.12 and in unrolled format in Figure 7.13. Following the procedure of Section 7.4, we subtract the estimated non-γ+jets contributions from the data and compute correction factors by dividing the remaining data yield by the γ+jets prediction. The resulting correction factors are shown in Figure 7.14.

As in the one-lepton control region, we next restrict our attention to the subset of events with $N_{\text{jets}} \geq 4$. We derive an additional flat scale factor to normalize the corrected MC to the data in this restricted control region. The scale factor we extract is 0.87 ± 0.05.

After applying this additional scale factor to the MC, we compare the MC prediction with the data in bins of M_R and $R^2$and observe that the corrected MC models the M_R-R^2 distribution in data well (see Figure 7.15). We then plot the distribution of N_{b-tags} (Figure 7.16). We observe a tendency for the MC to underpredict the yields for high N_{b-tags}, and assign a systematic uncertainty on the Z($\rightarrow \nu\nu$)+jets modeling equal to the magnitude of this discrepancy.

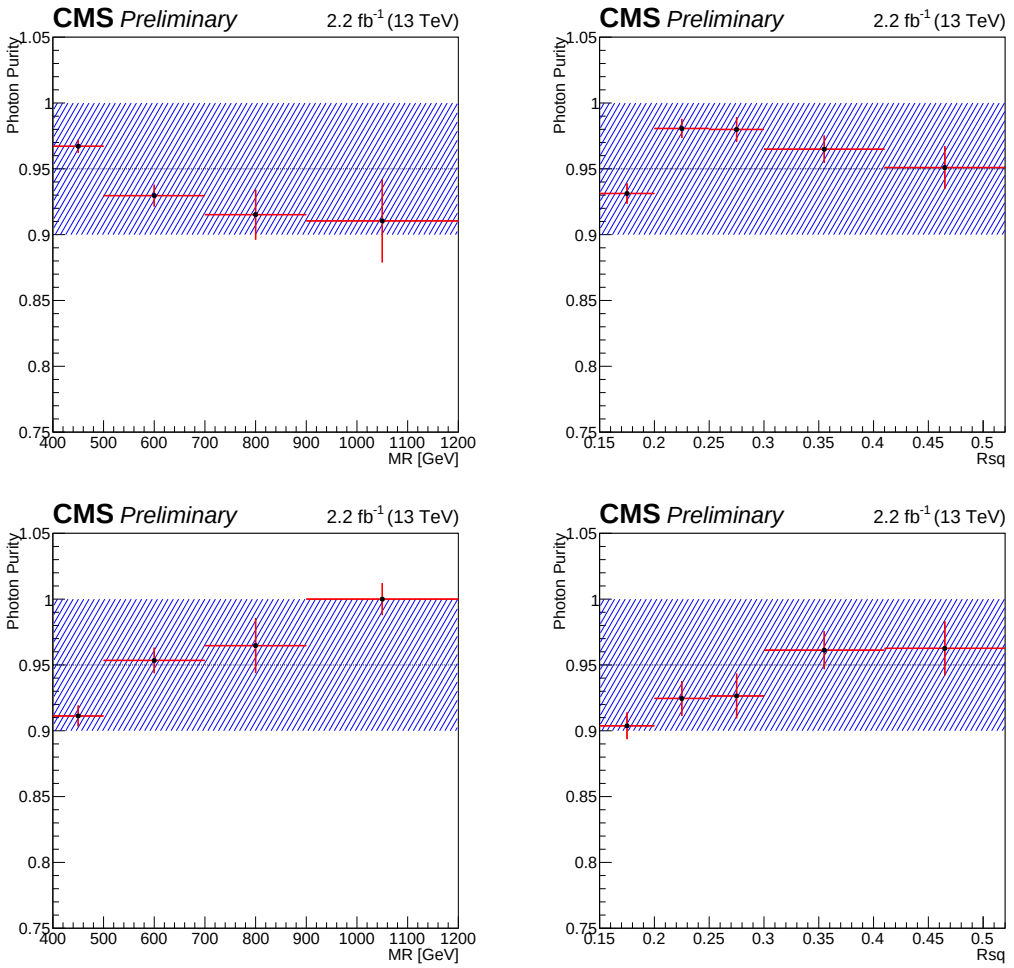


Figure 7.11: Fitted photon purity values in the ECAL barrel (top row) and endcaps (bottom row) as a function of M_R (left column) and R^2 (right column).

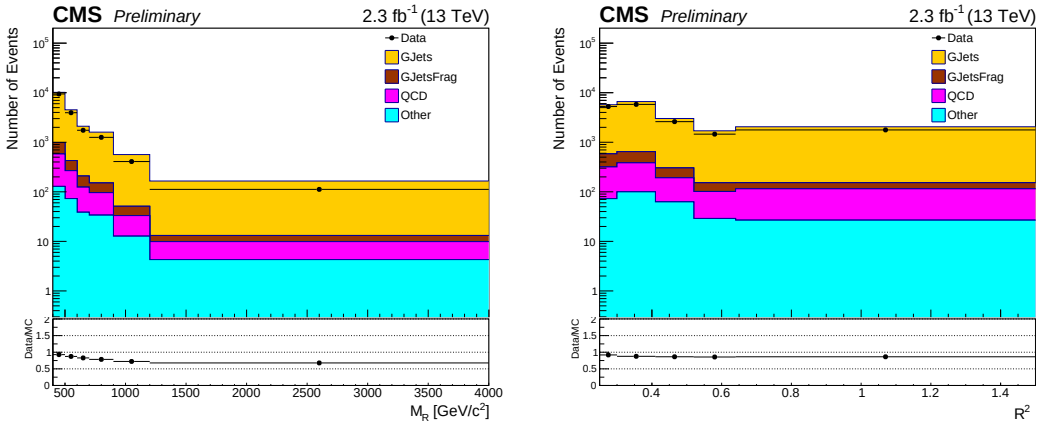
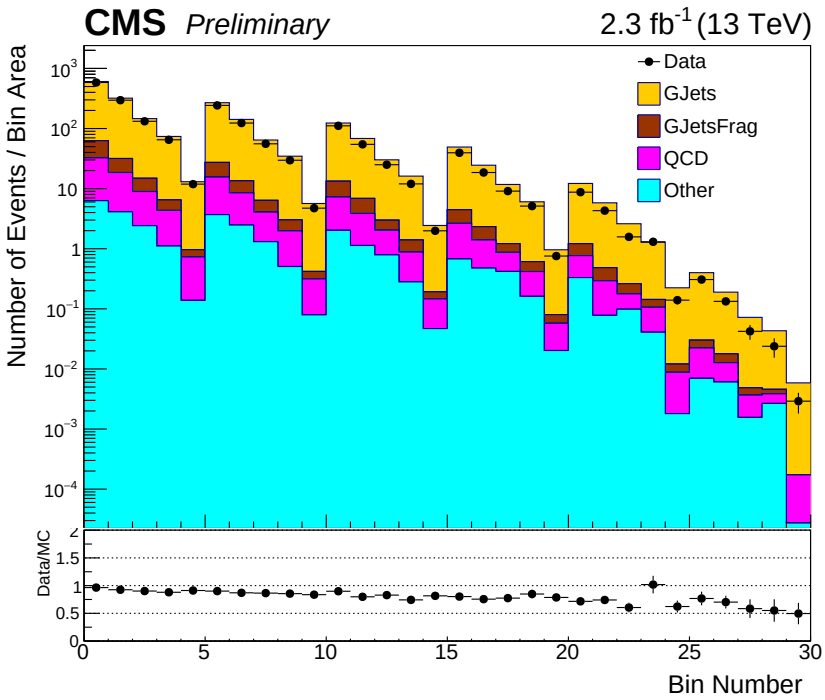


Figure 7.12: Distributions of M_R (left) and R^2 (right) in the γ+jets control region.



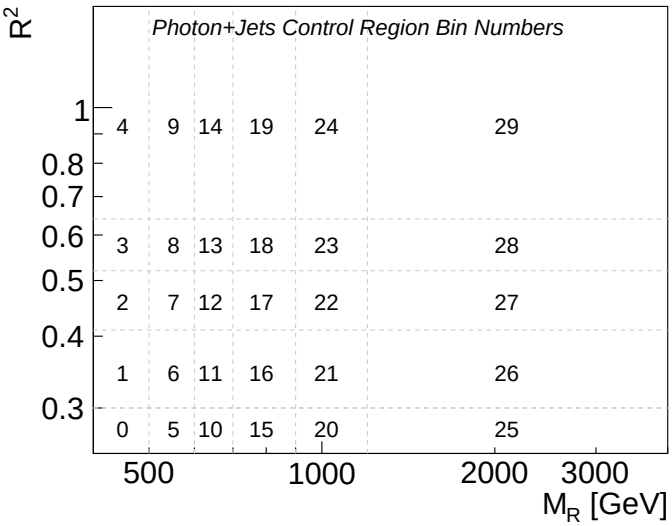


Figure 7.13: Top: M_R-R^2 distribution in the γ+jets control region displayed in unrolled form. Bottom: bin mapping for the unrolled plot.

Cross-check in one-lepton control region

To derive corrections to the Z($\rightarrow \nu\nu$)+jets MC using W($\rightarrow \ell\nu$)+jets events, we use the same one-lepton control region selection as in Section 7.4, except that the lepton's transverse momentum vector is added to $\vec{p}_T^{\,miss}$ for the calculation of R^2. We measure correction factors in bins of M_R and R^2 in the usual way, obtaining the

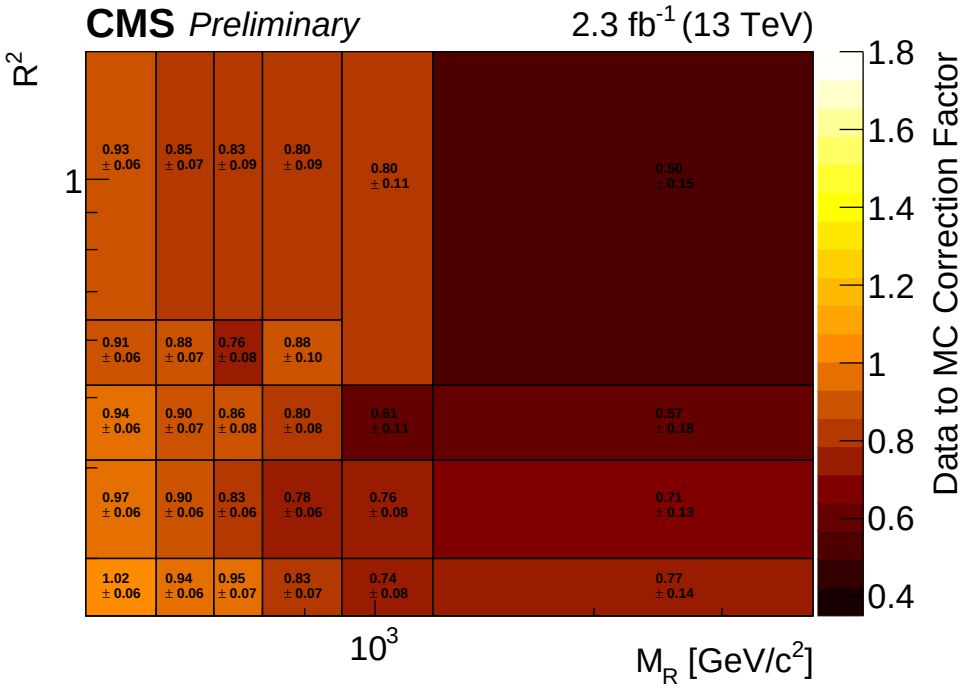


Figure 7.14: Correction factors derived in the γ+jets control region in bins of M_R and R^2, inclusive in the number of selected jets.

values shown in Figure 7.17.

We take the full size of the difference between these scale factors and those obtained from the γ+jets region as a systematic uncertainty on the Z($\rightarrow \nu\nu$)+jets modeling. This uncertainty quantifies how much the MC correction factors depend on which physics process we use to emulate Z($\rightarrow \nu\nu$)+jets. The fractional differences between the scale factors from the two control regions are shown in Figure 7.18. These differences are small, with the largest discrepancy being at the 30% level.

Closure test in two-lepton control region

The two-lepton Z($\rightarrow \ell\ell$)+jets-enriched control region is selected using the following cuts:

- Two electrons or muons passing the tight selection, with the leading lepton satisfying $p_T > 30$ GeV and the subleading lepton satisfying $p_T > 20$ GeV
- $80 < m_{\ell\ell} < 110$ GeV
- No b-tagged jets

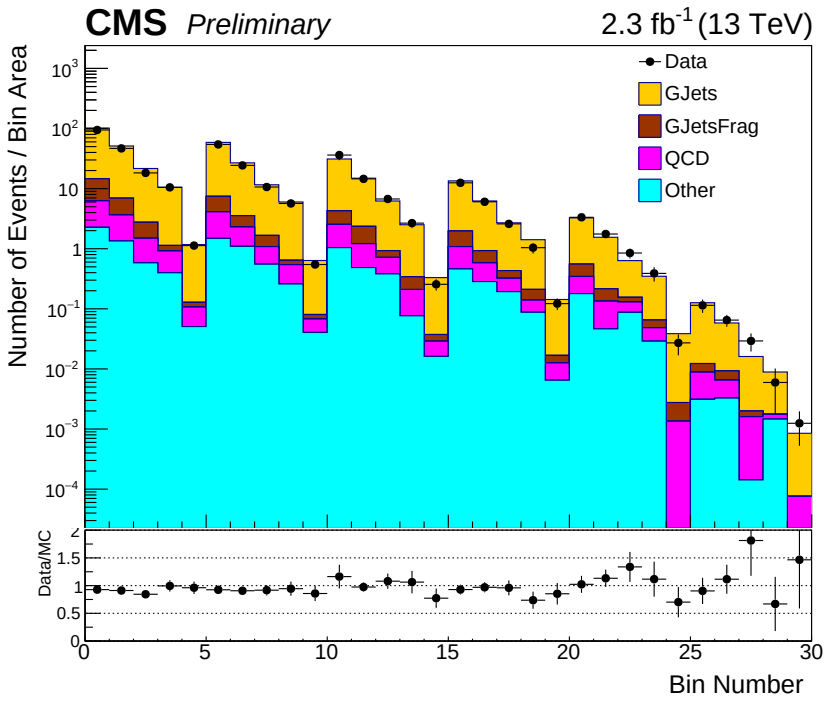


Figure 7.15: Unrolled M_R-R^2 distribution in the γ+jets control region after selecting events with four or more jets and appying the corrections derived in bins of M_R, R^2, and the number of jets.

- At least two jets with $p_T > 80$ GeV
- $M_R > 400$ GeV and $R^2 > 0.25$.

The Z($\rightarrow \ell\ell$)+jets MC is corrected using the M_R-R^2 scale factors derived in the γ+jets control region. The comparison between data and MC is performed both inclusively in N_{jets} and in the region $N_{\text{jets}} \geq 4$. Good agreement is observed in both regions; see Figure 7.19.

7.6 Estimation of QCD multijet background

The term 'QCD event' is a catch-all that refers to events not containing top quarks, W/Z/Higgs bosons, or other particles of interest. QCD events constitute the overwhelming majority of pp collision events at CMS. They typically have little to no true E_T^{miss}; they enter our search when they contain at least one jet whose momentum is badly mismeasured, causing the measured E_T^{miss} to be spuriously high.

An accurate prediction of the QCD multijet background using MC is difficult because of the enormous production cross section for QCD events. The number of

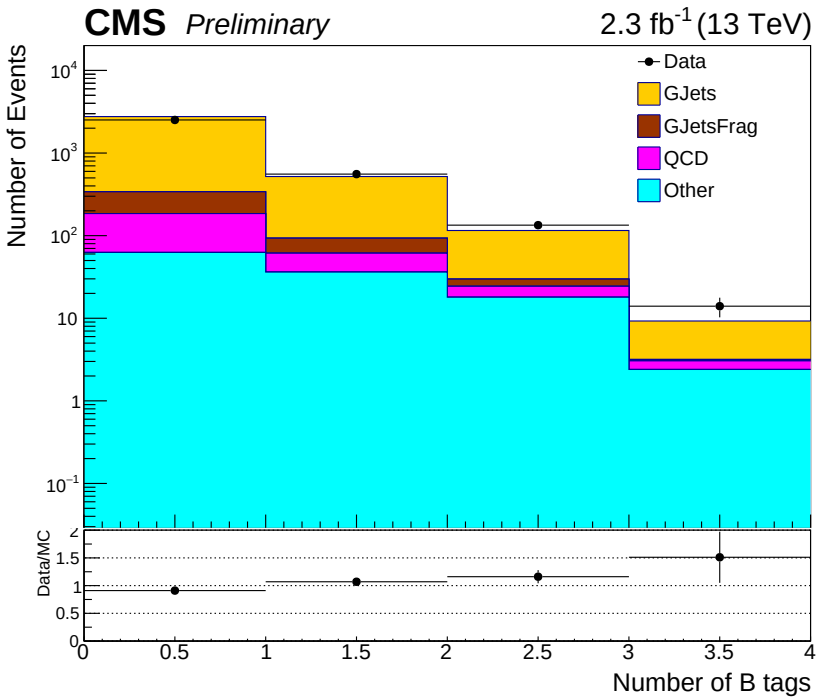


Figure 7.16: Distribution of the number of b-tagged jets in the γ+jets control region after selecting events with four or more jets and appying the corrections derived in bins of M_R, R^2, and the number of jets.

MC events that can be simulated is much smaller than the number of QCD events in data, so MC event weights are high. We instead take a data-driven approach to QCD estimation.

QCD control region selection

In a QCD event in which all of the measured $E_T^{miss}$is due to the mismeasurement of a jet's momentum, the vector $\vec{p}_T^{miss}$ points directly along or opposite the direction of that jet. This usually means that the $\vec{p}_T^{miss}$ aligns in the transverse plane with one of the razor megajets. If this happens, then (by momentum conservation) the two razor megajets must be back to back in the transverse plane. That is, the value of $\Delta\phi_R$ is close to π.

We see in simulation that the region $\Delta\phi_R > 2.8$ is 70-80% pure in QCD events, for M_R-R^2 bins that have a non-negligible amount of QCD in the search region. We use this high $\Delta\phi_R$ region as a control region for the QCD background. The selection for this control region is identical to that of the zero-lepton search region, except that the $\Delta\phi_R$ cut is inverted.

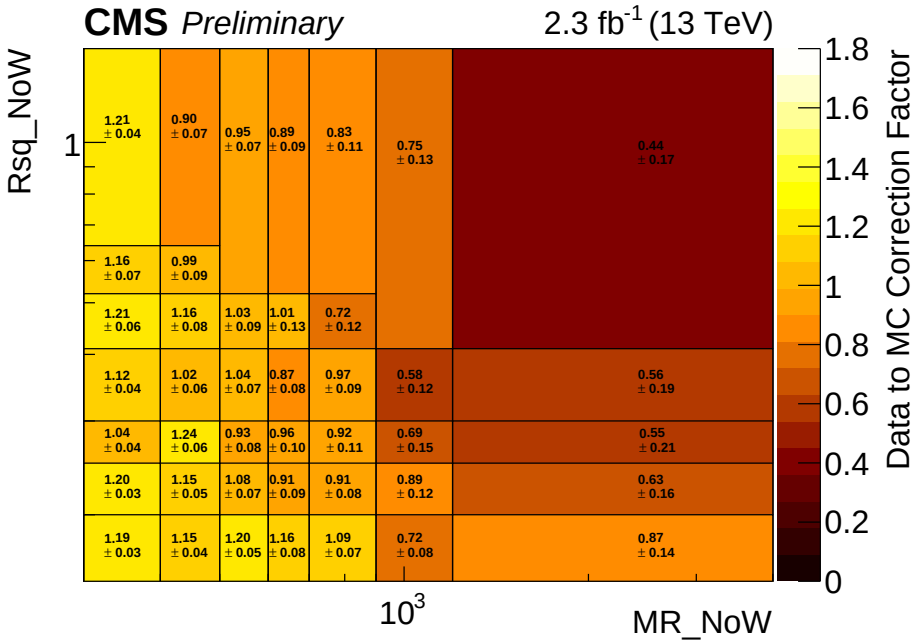


Figure 7.17: Correction factors derived for Z($\to \nu\nu$)+jets in the W($\to \ell\nu$)+jets control region in bins of M_R and R^2, inclusive in the number of selected jets.

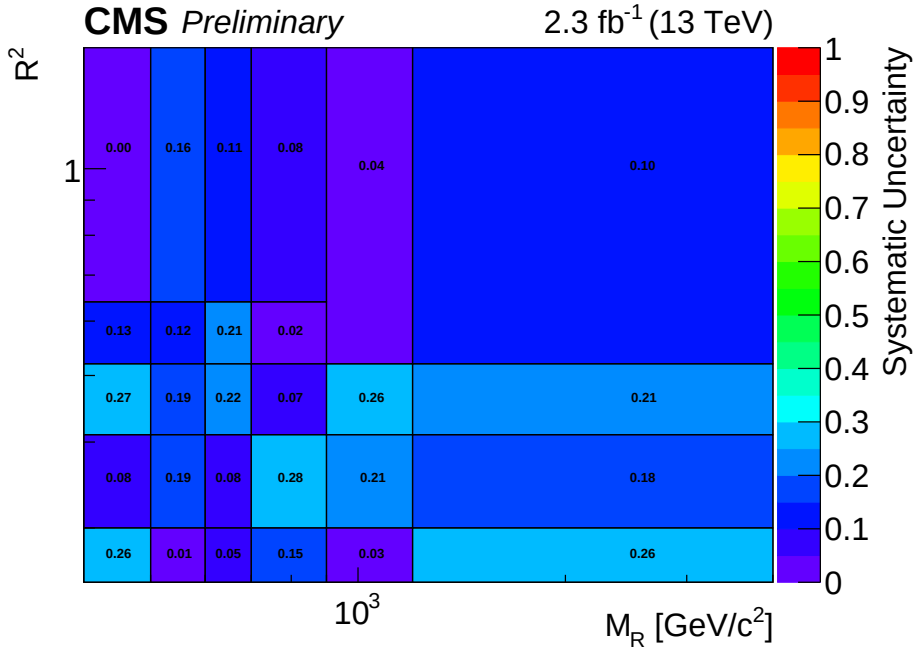


Figure 7.18: Fractional difference between the correction factors derived for Z($\to \nu\nu$)+jets in the γ+jets and W($\to \ell\nu$)+jets control samples in each bin of M_R and R^2.

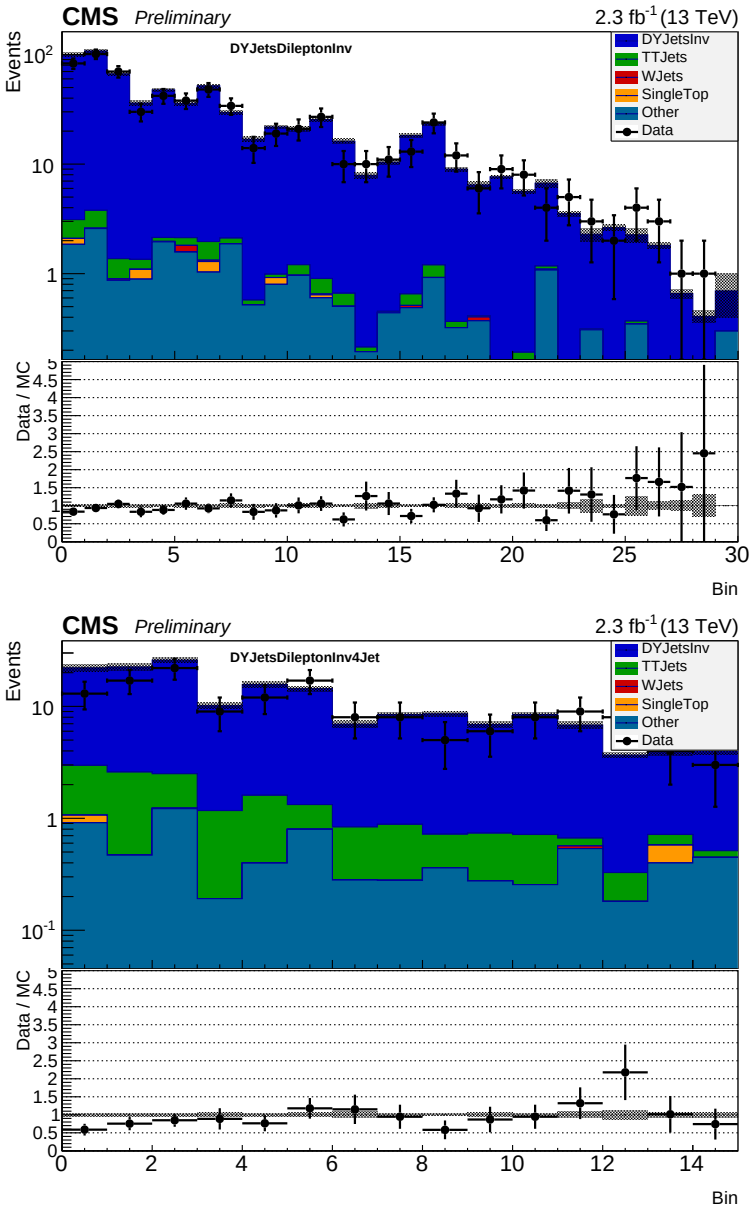


Figure 7.19: Unrolled M_R-R^2 distributions in the $Z \to \ell\ell$ two-lepton control region, before (top) and after (bottom) requiring $N_{\text{jets}} \geq 4$.

To predict the number of QCD events in the search region, we obtain the number of QCD events in the high $\Delta\phi_R$ region and multiply it by the transfer factor ζ defined by

$$\zeta = \frac{N\,(\Delta\phi_R < 2.8)}{N\,(\Delta\phi_R > 2.8)}. \tag{7.6}$$

Modeling of QCD transfer factors

The values of ζ are measured in a low-R^2region in data defined by $0.2 < R^2 < 0.25$. The transfer factors exhibit a falling behavior as a function of M_R that is well described by a power law. We fit the distribution to obtain:

$$\zeta(M_R) = 3.1 \times 10^7 (M_R/\text{GeV})^{-3.1} + 0.062. \tag{7.7}$$

The transfer factors and the fitted function are shown in Figure 7.20.

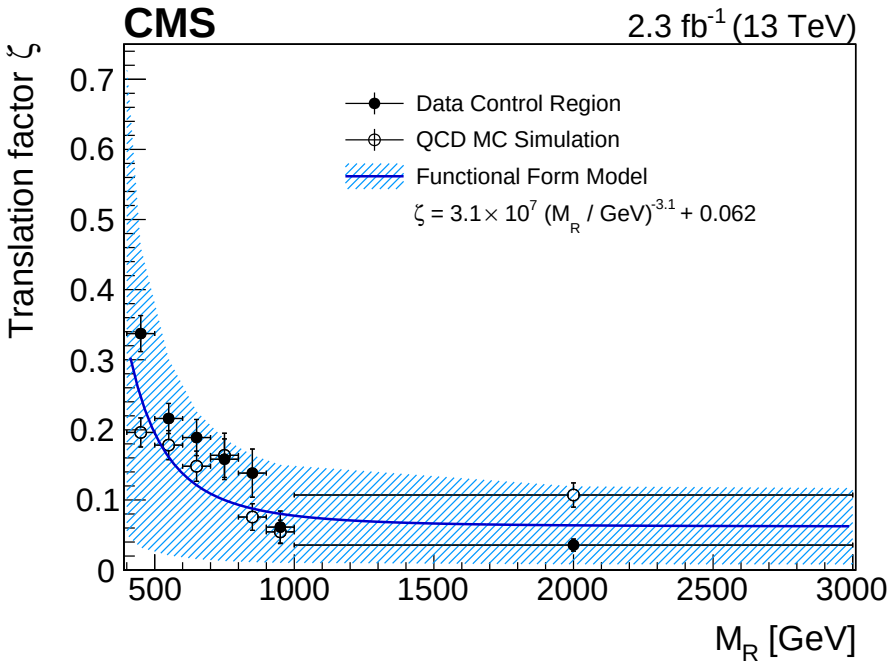


Figure 7.20: QCD transfer factors ζ measured in data (filled circles) and MC (open circles) along with the fitted power law function. The hashed blue band represents the systematic uncertainty on the fit function.

We observe that the value of ζ does not depend significantly on R^2 within the available statistics; see Figure 7.21. We therefore treat it as independent of R^2 and assign a systematic uncertainty of 87% on ζ to cover the spread of the observed values.

7.7 Systematic uncertainties

Systematic uncertainties on the background prediction can be separated into two broad classes. The first class contains uncertainties on our knowledge of specific instrumental and theoretical quantities. For example, the measurement of the lepton identification efficiency is performed on a data sample with limited event statistics,

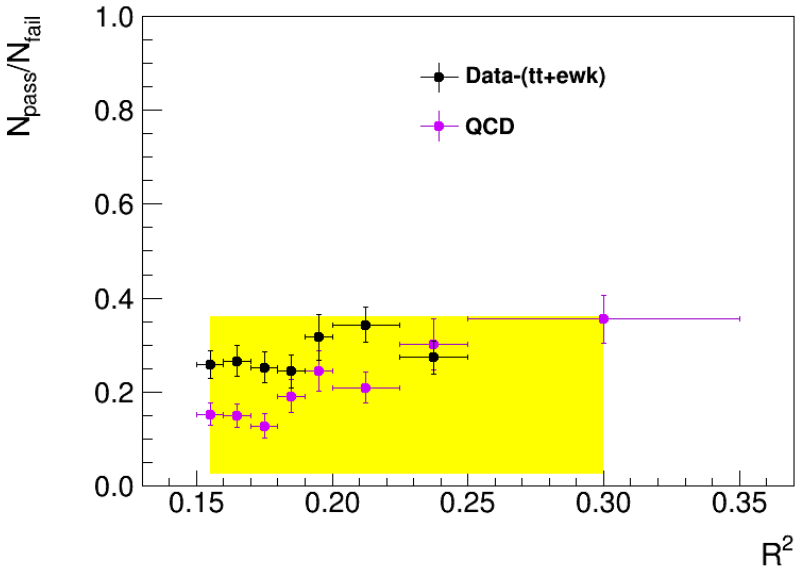


Figure 7.21: QCD transfer factor ζ computed in bins of R^2 in data (black) and MC (pink). The yellow band represents the systematic uncertainty on the transfer factors.

so it has an associated uncertainty. The second class contains uncertainties on other aspects of the background modeling not specifically enumerated. The closure tests performed in each analysis control region are designed to probe for deficiencies in the simulation modeling of the SM background. The results of the closure tests indicate the level of accuracy at which we can control for potential mismodeling and are used to assign systematic uncertainties on each background process. Each uncertainty is described briefly in the following subsections.

Instrumental and theoretical systematics

The uncertainties due to known instrumental and theoretical effects are summarized in Table 7.2. These include uncertainties on the jet energy scale, the lepton and b-jet identification efficiencies, the total integrated luminosity, and the change in the observed yields under variations of the factorization and renormalization scales in the MC simulation.

Uncertainties estimated in analysis control regions

The MC correction factors derived in the one-lepton and photon control regions have associated statistical uncertainties from the limited event sample. These un-

Uncertainty source	Typical uncertainty size
Jet energy scale	2-15%
Lepton momentum scale	7-9%
Lepton selection efficiency	7-8%
b-tagging efficiency	6-15%
b mistag	4-7%
Fac. and renorm. scales	10-25%
Integrated luminosity	2.7%
Monte Carlo statistics	$1/\sqrt{N_{MC}}$

Table 7.2: Summary of instrumental and theoretical sources of uncertainty on the MC-based background prediction.

certainties are displayed as percentages in Figure 7.22. Additionally, the correction factors have associated systematic uncertainties originating from the scale factor calculation procedure, as discussed in Section 7.4. These are small compared with the statistical uncertainty in most bins.

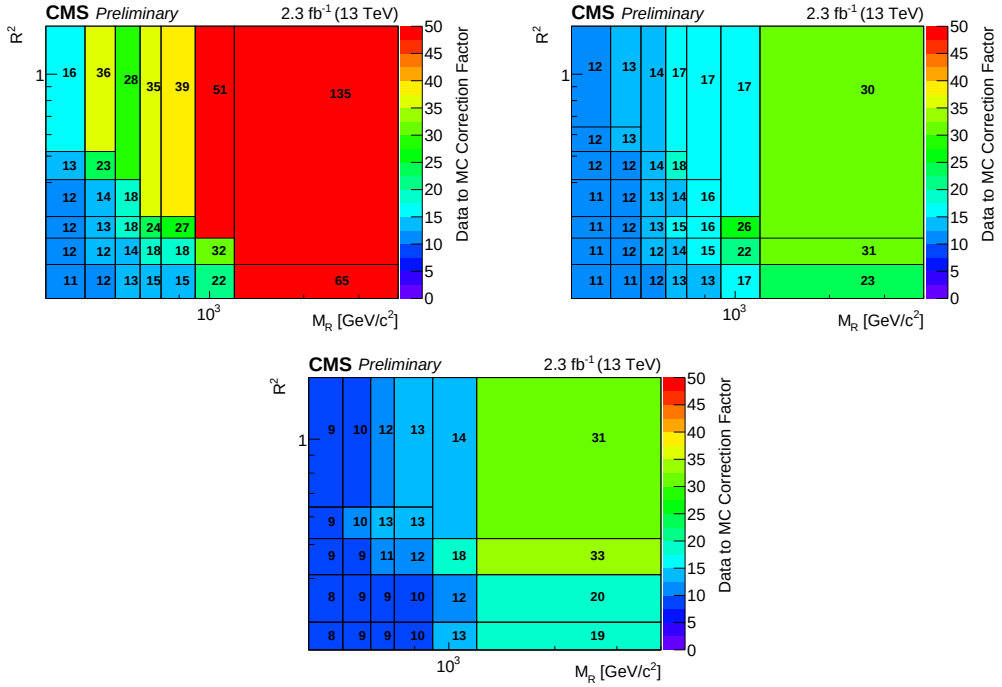


Figure 7.22: Statistical uncertainties on the $t\bar{t}$+jets (left), W($\rightarrow \ell\nu$)+jets (right), and Z($\rightarrow \nu\nu$)+jets (bottom) correction factors, expressed as a percentage.

Each closure test performed in the analysis control regions is used to assess uncertainty on some aspect of the background modeling. The closure tests and the approximate sizes of the resulting uncertainties are listed in Table 7.3. Uncertain-

ties can either be correlated among bins (in which case they are governed by a single nuisance parameter in the likelihood) or uncorrelated (in which case an independent nuisance parameter is assigned for each search bin); we indicate this in the table.

Uncertainty source	Background process(es)	Typical uncertainty size	Correlated?
One-lepton closure test (M_R)	$t\bar{t}$+jets, W($\to \ell\nu$)+jets	5-100%	N
One-lepton closure test (R^2)	$t\bar{t}$+jets, W($\to \ell\nu$)+jets	5-50%	N
$t\bar{t}$ dilepton closure test	2ℓ $t\bar{t}$+jets	0-30%	N
Veto lepton / tau closure tests	$t\bar{t}$+jets, W($\to \ell\nu$)+jets	0-30%	N
γ+jets vs. W+jets scale factors	Z($\to \nu\nu$)+jets	0-30%	Y
QCD systematic	QCD	87%	Y

Table 7.3: Summary of systematic uncertainties estimated in analysis control regions. The rightmost column indicates whether the uncertainty is correlated or uncorrelated from bin to bin.

7.8 Comparison with the fit-based method

This completes the description of the MC-based background prediction for the 2015 razor search. We now compare the MC-based prediction to that obtained using the fit method described in Chapter 6. A bin-by-bin comparison of the two predictions is provided in Figures 7.23-7.25.

In most analysis categories the fit-based and MC-based predictions agree within uncertainty. The most significant discrepancies are observed in the last M_R bin in the Muon Multijet 0 b-tag category, and in the Electron Multijet 2 and $\geq$3 b-tag categories (which are fit simultaneously, as described in Section 6.5). In both cases, the difference is traced to a fluctuation in data in the M_R sideband, which pulls the tail of the fit too high or too low. These deviations reflect the fact that the fit function ansatz for the razor background induces strong correlations among search bin predictions.

We take the good agreement between the fit-based and MC-based predictions as evidence of the robustness of the background modeling. We choose to proceed with the MC-based prediction method for interpretation of the analysis.

7.9 Search region results

The unblinded data and MC-based background prediction in the search region are shown together in unrolled format in Figures 7.26-7.31. No significant excess over the predicted SM background is seen.

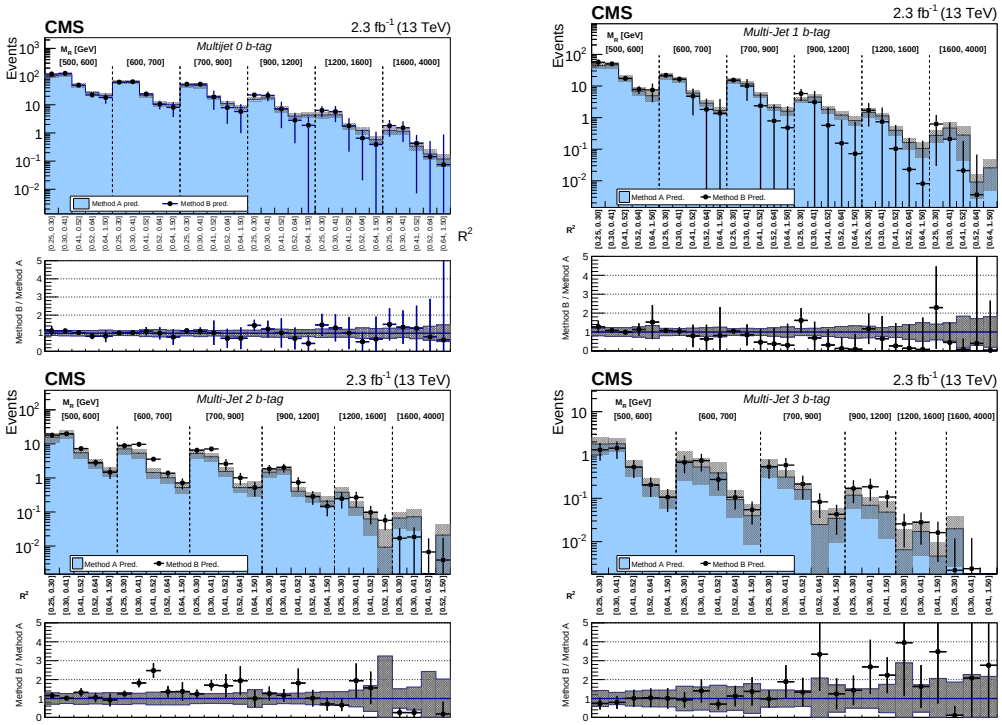


Figure 7.23: Comparison between the MC-based prediction (marked 'Method A') and the fit-based prediction (marked 'Method B') in each b-tag category of the Multijet region. The predictions are shown in unrolled format with dashed vertical lines delineating bins in M_R. The bottom panel of each plot shows the ratio of the two predictions.

7.10 Signal modeling

The search is interpreted in terms of simplified SUSY models as described in Section 3.4. Signal MC samples are generated using MadGraph 5 interfaced with Pythia 8, with detector simulation performed by the CMS Fastsim framework [36]. The Fastsim MC events are reweighted to correct the efficiencies for identifying leptons and b-jets, and to obtain a distribution of the number of pileup vertices similar to that in data. An additional correction is applied to compensate for an observed mismodeling of the ISR jet multiplicity distribution.

Simplified SUSY models

We interpret the null search result as 95% confidence level limits on the pair production of heavy gluinos with decays to quarks and the LSP. We consider two scenarios:

- The gluino decays to the neutralino and two third-generation quarks. This

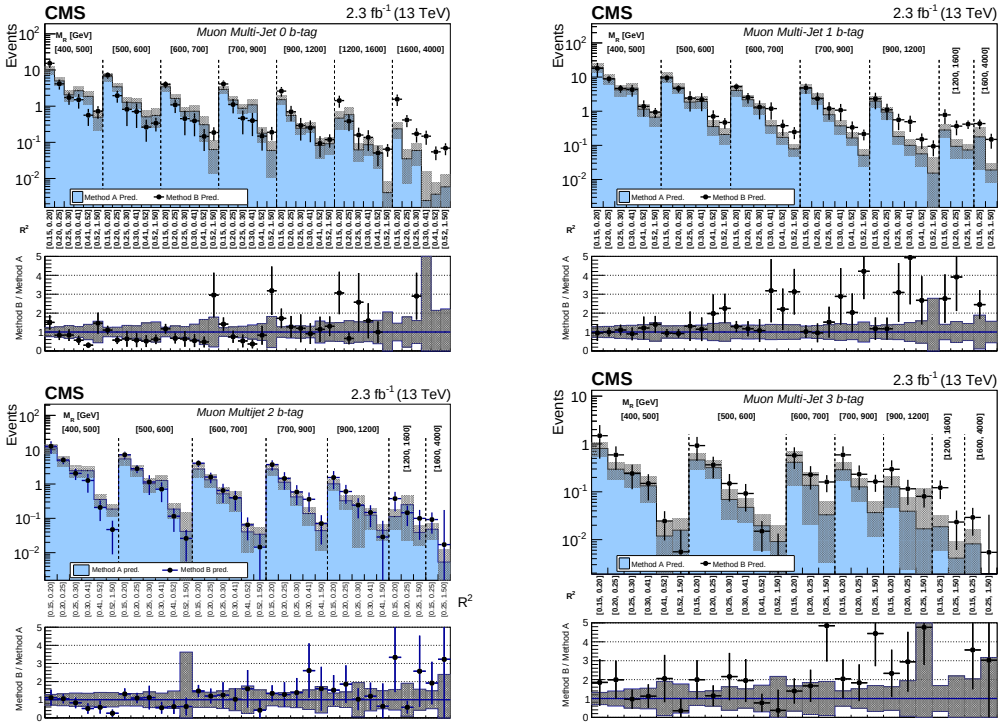


Figure 7.24: Comparison between the MC-based prediction (marked 'Method A') and the fit-based prediction (marked 'Method B') in each b-tag category of the Muon Multijet region. The predictions are shown in unrolled format with dashed vertical lines delineating bins in M_R. The bottom panel of each plot shows the ratio of the two predictions.

scenario encompasses the T1tttt and T1bbbb models, as well as T1ttbb with arbitrary gluino branching ratios to $t\bar{t}\tilde{\chi}_1^0$, $b\bar{b}\tilde{\chi}_1^0$, and $b\bar{t}\tilde{\chi}_1^+/t\bar{b}\tilde{\chi}_1^-$. We compute limits for several choices of branching ratios; the considered values are indicated in the diagram in Figure 7.32. In addition to displaying limits for each of these decay scenarios, we also show a conservative limit that corresponds to the worst-case limit among the considered branching ratios.

- The gluino decays to the neutralino and two first- or second-generation quarks. This corresponds to the T1qqqq simplified model.

The masses of the gluino and the LSP in the simplified models can be specified arbitrarily. Limits on each model are computed for a two-dimensional grid of $(m_{\text{gluino}}, m_{\text{LSP}})$ values, and interpolation is performed to obtain the smooth exclusion contours shown in the exclusion plots. Where relevant, the lightest chargino is taken to be 5 GeV heavier than the LSP. Theoretical cross sections for pair pro-

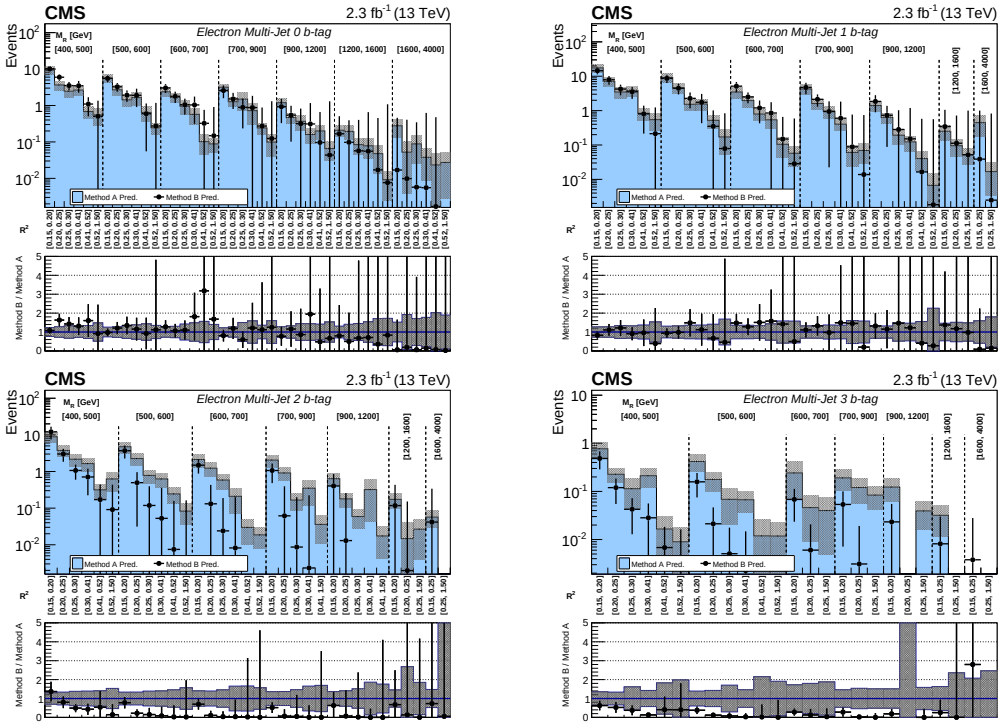


Figure 7.25: Comparison between the MC-based prediction (marked 'Method A') and the fit-based prediction (marked 'Method B') in each *b*-tag category of the Electron Multijet region. The predictions are shown in unrolled format with dashed vertical lines delineating bins in M_R. The bottom panel of each plot shows the ratio of the two predictions.

duction of gluinos are obtained at NLO and next-to-leading logarithmic accuracy from [129].

Signal systematics

The signal MC is subject to the same instrumental and theoretical uncertainties as the background MC (see Table 7.2). In addition, there are uncertainties from the corrections to the Fastsim lepton and *b*-jet identification efficiencies and to the distribution of ISR jet multiplicity. The lepton and *b*-jet uncertainties are typically less than 10%. The ISR jet uncertainty is 15-30% and affects events with hadronic recoil of 400 GeV or higher.

Signal contamination

The control samples used in the MC-based background prediction method are designed to be relatively free of SUSY signal events. This is achieved via the m_T cut

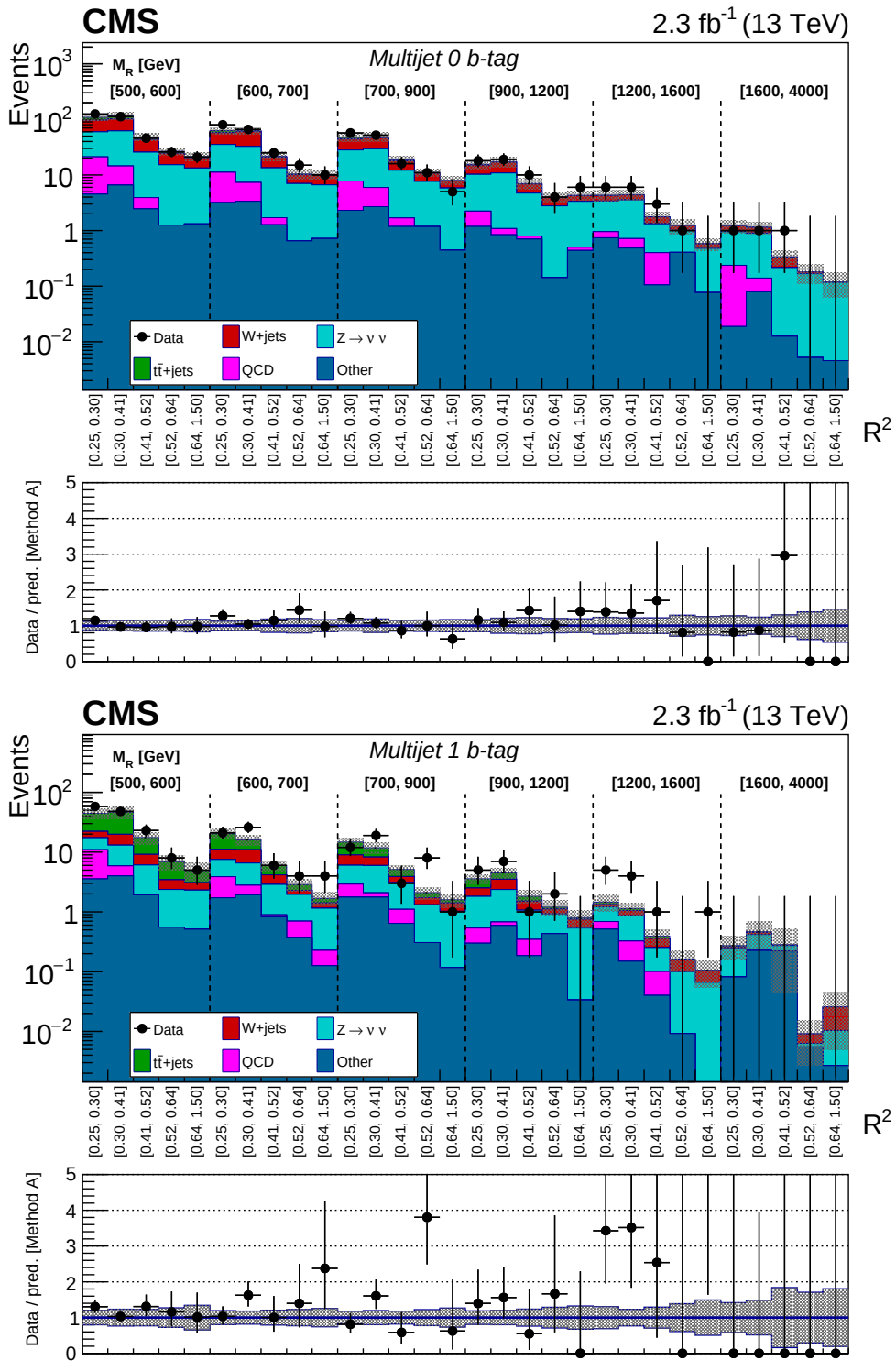


Figure 7.26: Observed data counts in each bin of the Multijet 0 (top) and 1 (bottom) *b*-tag categories, compared with the MC-based background prediction.

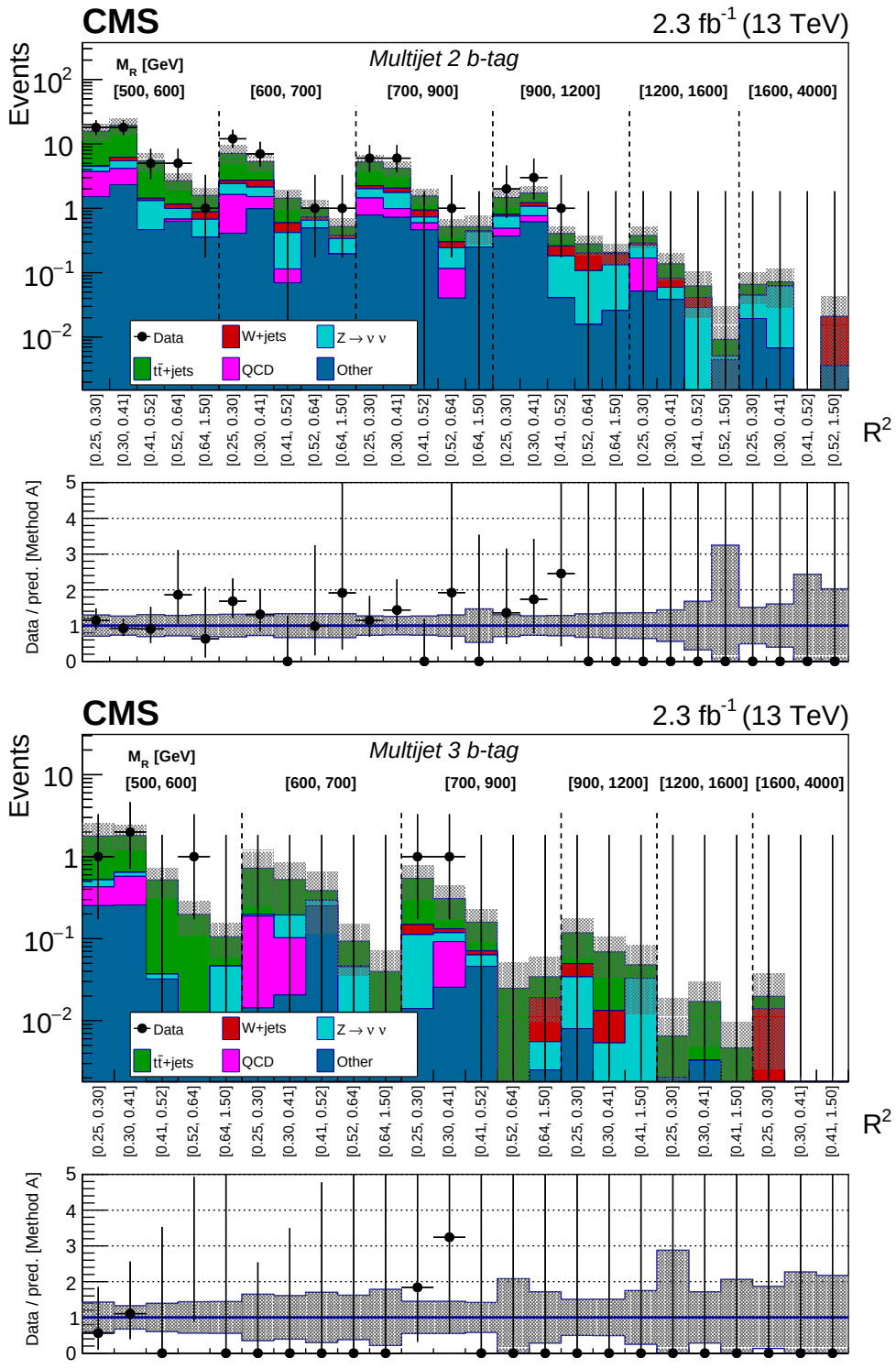


Figure 7.27: Observed data counts in each bin of the Multijet 2 (top) and ≥3 (bottom) *b*-tag categories, compared with the MC-based background prediction.

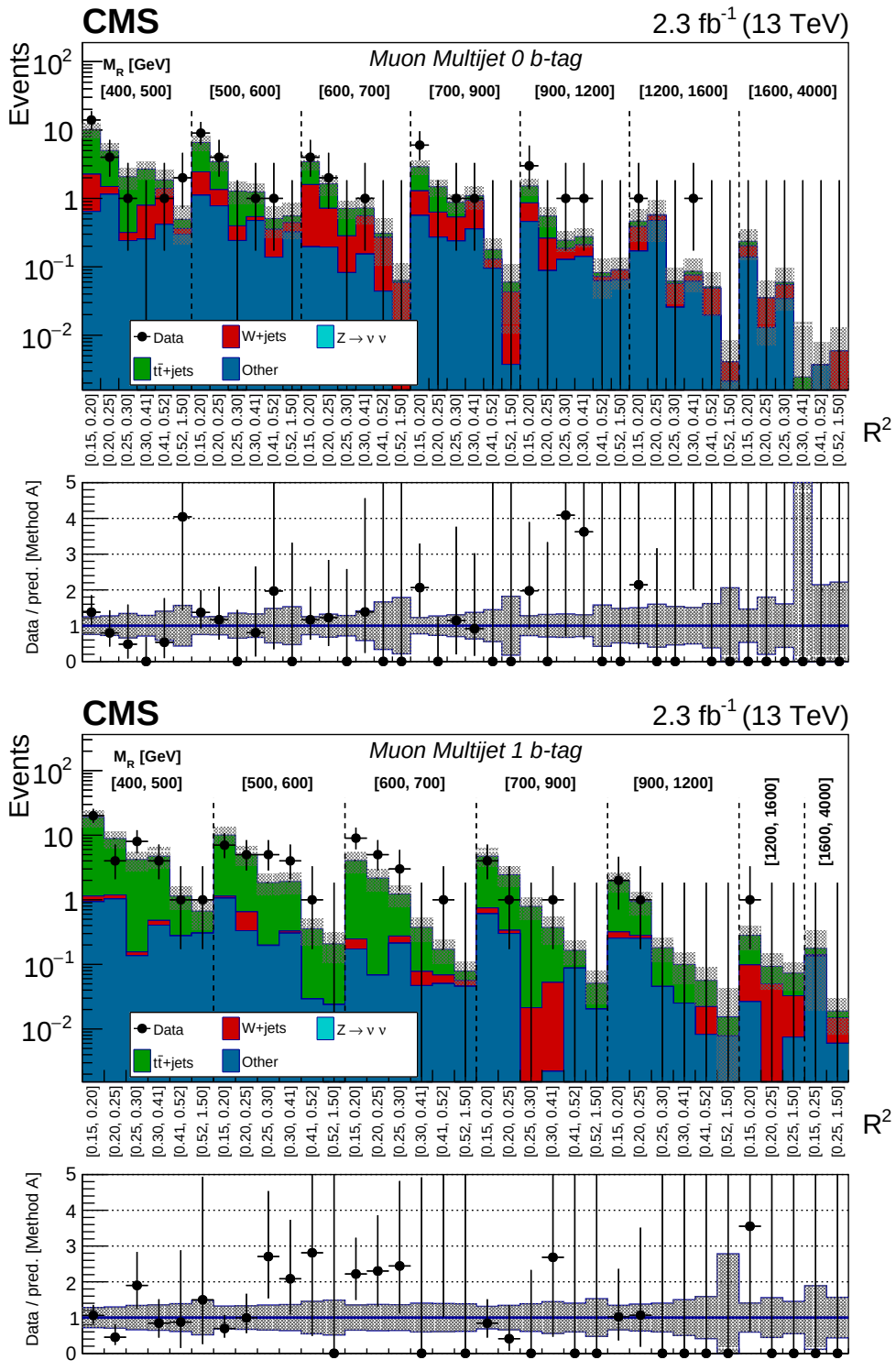


Figure 7.28: Observed data counts in each bin of the Muon Multijet 0 (top) and 1 (bottom) *b*-tag categories, compared with the MC-based background prediction.

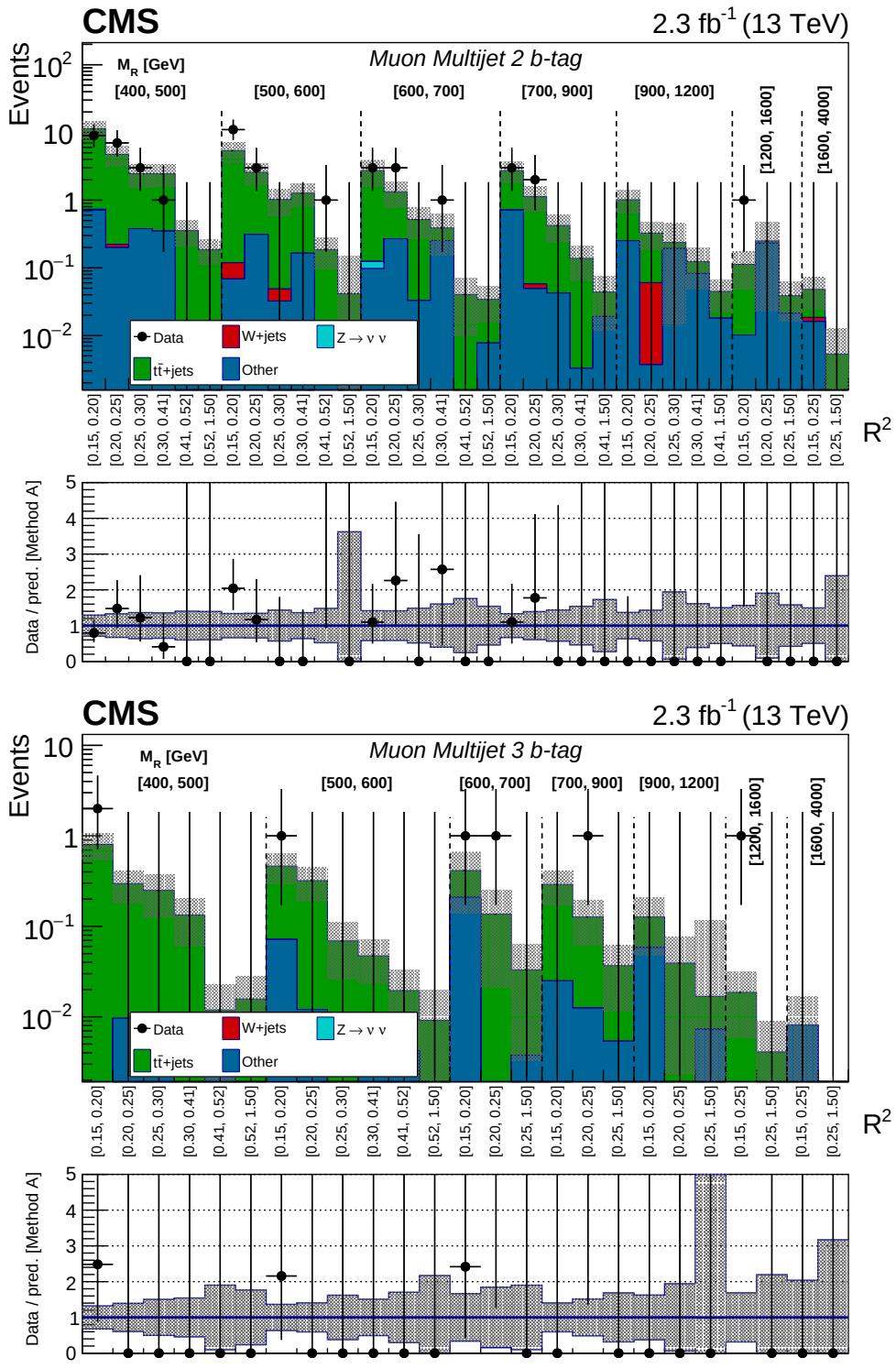


Figure 7.29: Observed data counts in each bin of the Muon Multijet 2 (top) and ≥3 (bottom) *b*-tag categories, compared with the MC-based background prediction.

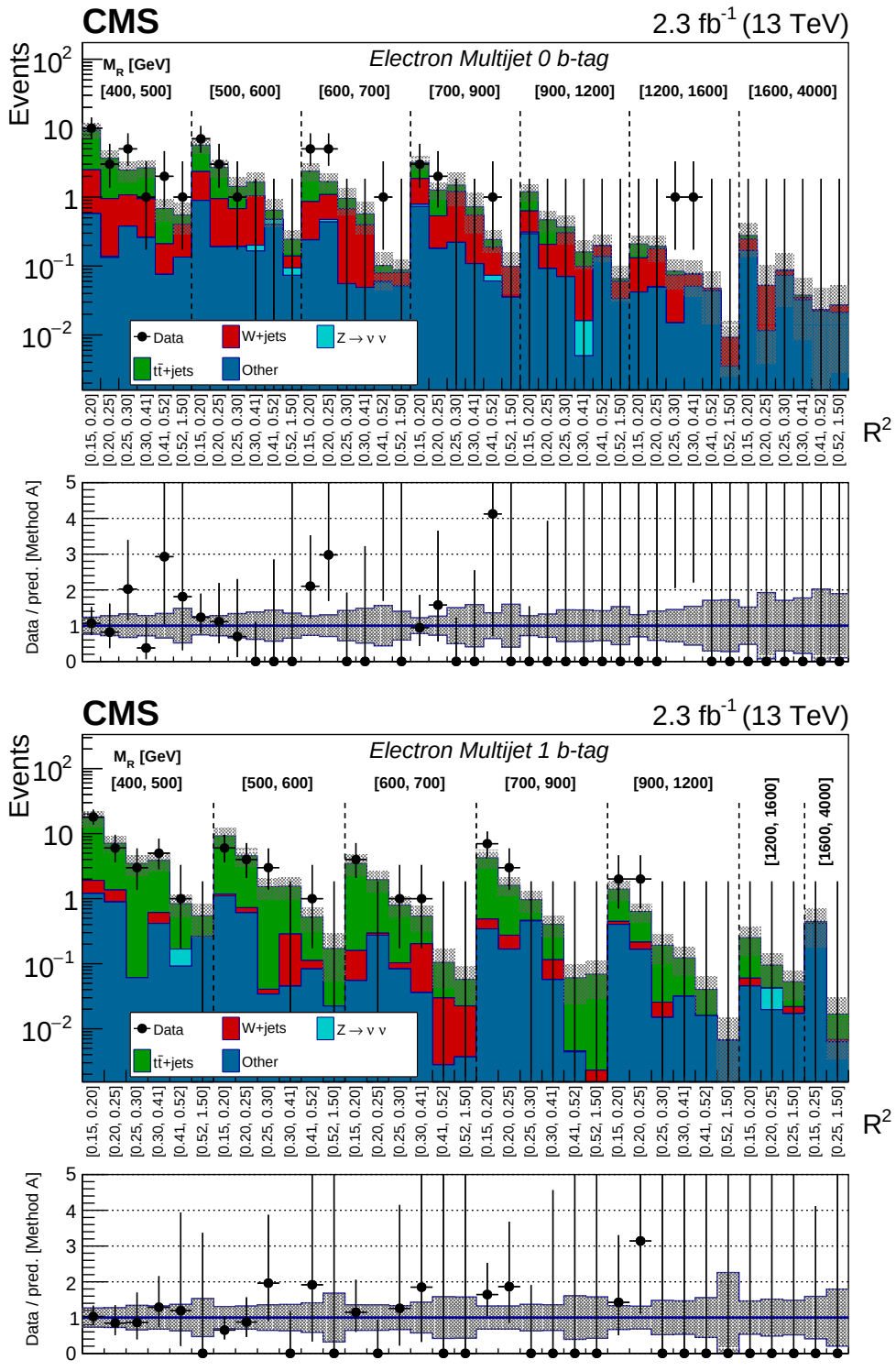


Figure 7.30: Observed data counts in each bin of the Electron Multijet 0 (top) and 1 (bottom) *b*-tag categories, compared with the MC-based background prediction.

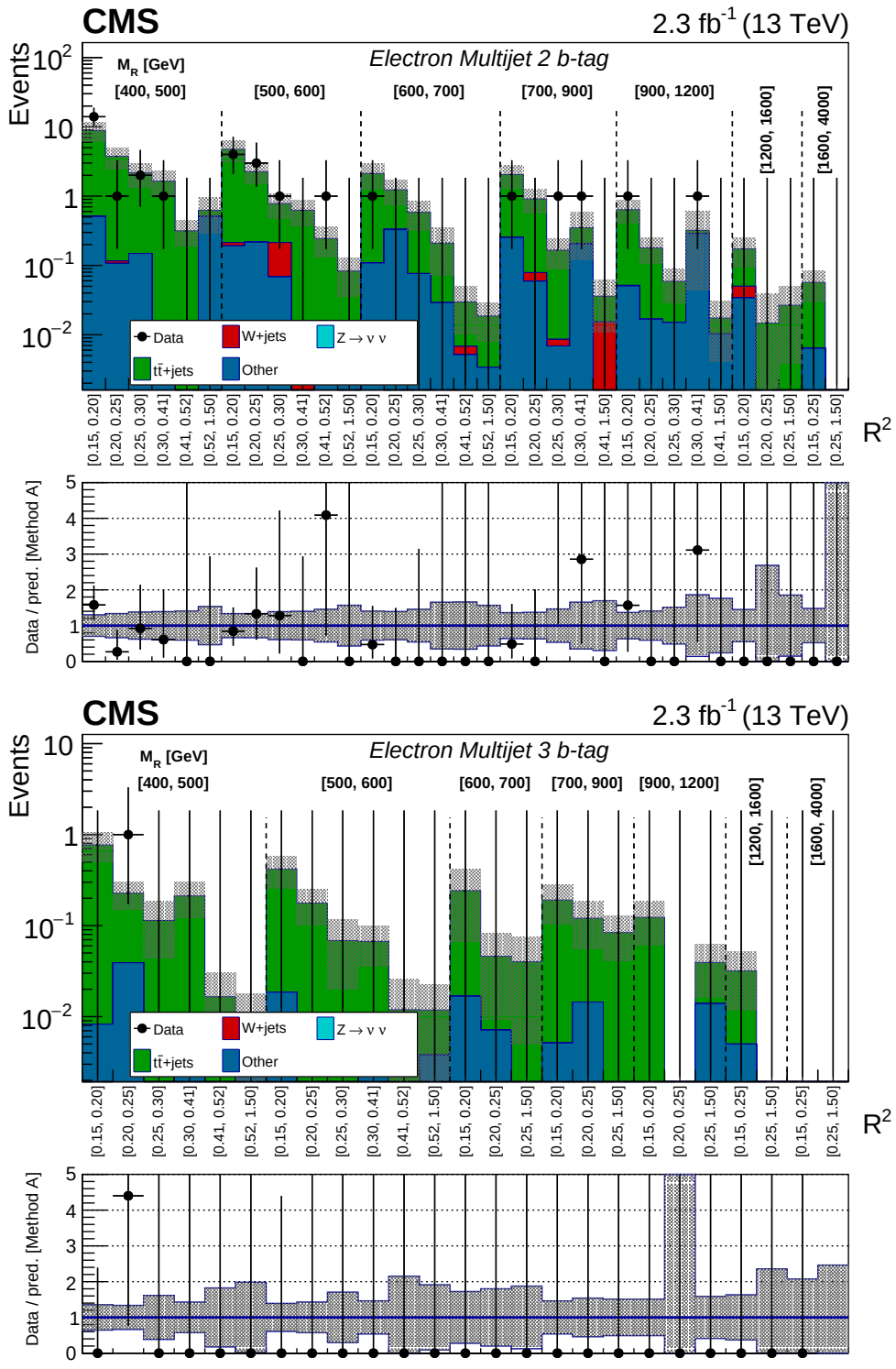


Figure 7.31: Observed data counts in each bin of the Electron Multijet 2 (top) and ≥3 (bottom) *b*-tag categories, compared with the MC-based background prediction.

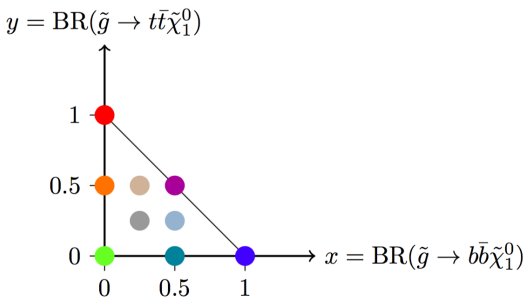


Figure 7.32: Branching ratios of gluinos decaying to third-generation quarks considered in the interpretation of the analysis. The branching ratios $x = \mathrm{BR}(\tilde{g} \to b\bar{b}\tilde{\chi}_1^0)$ and $y = \mathrm{BR}(\tilde{g} \to t\bar{t}\tilde{\chi}_1^0)$ can be freely varied between zero and one; the branching ratio for mixed decays to two quarks and a chargino is $1 - x - y$.

that is applied in all leptonic control regions. Despite this, some SUSY signal events could populate the control regions in non-negligible quantities. Signal contamination in the one-lepton $t\bar{t}$+jets and W($\to \ell\nu$)+jets control regions would increase the number of data events in those samples, which would increase the magnitudes of the MC correction factors derived in these control regions. This in turn would increase the number of predicted $t\bar{t}$+jets and W($\to \ell\nu$)+jets events in the search regions, thus biasing the background estimate.

We account for signal contamination using the *reduced efficiency method*, which works as follows. For each SUSY signal considered, we estimate the effect on the measured MC correction factors due to the presence of signal in the control regions. The impact on the signal region background prediction is, on average,

$$\Delta B = N_{S,\mathrm{MC}} \times \frac{\Delta N_{C,\mathrm{Data}}}{N_{C,\mathrm{MC}}} \tag{7.8}$$

$$= N_{S,\mathrm{MC}} \times \frac{N_{C,\mathrm{SUSY}}}{N_{C,\mathrm{MC}}}, \tag{7.9}$$

where $N_{C,\mathrm{SUSY}}$ is the expected number of SUSY events in control region. We correct the background prediction explicitly by subtracting ΔB from it.

If the SUSY signal cross section is μ times the nominal theoretical cross section, the amount of signal contamination is $\mu\,\Delta B$. The predicted yield in the search region is then

$$(N_{S,\mathrm{SM}} - \mu\,\Delta B) + \mu\,N_{S,\mathrm{SUSY}} \tag{7.10}$$

$$= N_{S,\mathrm{SM}} - \mu(N_{S,\mathrm{SUSY}} - \Delta B). \tag{7.11}$$

In the second line, we have rearranged the terms to show that subtracting $\mu\,\Delta B$ from the background prediction is equivalent to subtracting ΔB from the nominal

SUSY signal yield. This demonstrates that the ΔB correction reduces the effective efficiency of the signal region for SUSY events. In this way, the possibility of signal contamination leads to worsened limits.

7.11 Limit-setting procedure

Limits are computed using the LHC CL_s modified frequentist procedure [130]. The likelihood function takes the form of a product of Poisson distributions, one for each search bin. Systematic uncertainties are quantified using nuisance parameters θ having default values $\tilde{\theta}$. Varying a nuisance parameter from its default value causes the background prediction in one or more bins to change in a manner consistent with the uncertainty. Constraint terms encoding the probability distributions of the nuisances parameters, represented schematically by $p(\tilde{\theta} \mid \theta)$, are added to the likelihood. The full likelihood function is

$$\mathcal{L}(\{x_i\}|\mu,\theta) = \prod_{i=0}^{n_b} \text{Poisson}(x_i \mid \mu,\theta) \times p(\tilde{\theta} \mid \theta). \tag{7.12}$$

Here μ measures the SUSY production cross section σ relative to the theoretical cross section, $\mu = \sigma/\sigma_{theory}$.

The likelihood ratio test statistic,

$$q_\mu \equiv -2 \log \frac{\mathcal{L}(\{x_i\} \mid \mu, \hat{\theta}_\mu)}{\mathcal{L}(\{x_i\} \mid \hat{\mu}, \hat{\theta})}, \tag{7.13}$$

is used to assess the compatibility of the data with the background-only and signal-plus-background hypotheses for a given signal strength [131]. Nuisance parameters associated with systematic uncertainties in the analysis are dealt with by *profiling*; that is, maximizing the likelihood over all possible nuisance parameter values. The likelihood $\mathcal{L}(\{x_i\} \mid \mu, \hat{\theta}_\mu)$ is computed by fixing μ and finding the nuisances parameter values $\hat{\theta}_\mu$ that maximize the likelihood. $\mathcal{L}(\{x_i\} \mid \hat{\mu}, \hat{\theta})$ is computed by maximizing the likelihood over both μ and θ to obtain best-fit values $\hat{\mu}$ and $\hat{\theta}$, under the constraint $0 \leq \hat{\mu} \leq \mu$. The constraint on μ requires the signal strength to be non-negative and enforces that the limit resulting from the test is one-sided. Profiling the nuisance parameters in the test statistic calculation broadens the distribution of q_μ as a function of μ; thus, larger systematic uncertainties decrease the sensitivity of the search by making it more difficult to discriminate between different values of μ.

To perform the hypothesis test for a particular signal strength μ, we obtain the distributions $f(q_\mu \mid \mu)$ and $f(q_\mu \mid 0)$ of the test statistic under the signal-plus-background and background-only hypotheses, respectively. These distributions can be approximated by generating many toy MC datasets, or via the asymptotic formulae given later in this section. We use the asymptotic approach in order to significantly reduce the computing time needed to obtain the limits. A comparison with the toy MC approach is performed for a few signal masses, and the two methods are seen to give similar results.

In a traditional hypothesis test, we would compute the p-value associated with the observed data for the signal-plus-background hypothesis,

$$p_\mu = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu \mid \mu)\, dq_\mu, \tag{7.14}$$

where q_μ^{obs} is the value of the test statistic on the observed dataset. In the CL_s method we additionally need the p-value under the background-only hypothesis,

$$p_0 = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu \mid 0)\, dq_0. \tag{7.15}$$

The quantity CL_s is then computed as

$$CL_s \equiv \frac{p_\mu}{1 - p_0}. \tag{7.16}$$

The value of μ is adjusted iteratively until the value where $CL_s = 0.05$ is found. This value of μ is returned by the algorithm as the upper limit on the SUSY signal strength. Note that, while it is customary to refer to the result of the CL_s procedure as a 95% confidence level limit, the true confidence level is actually higher than this, due to the division by $1 - p_0$ in the calculation of CL_s [130].

In addition to the observed limit obtained using the data, we also report the expected limit, defined as the median signal strength that would be excluded under the assumption that the background-only hypothesis is true. Deviations of the observed limit from the expected limit indicate excesses or deficits in the observed data compared with the predicted background.

Asymptotic test statistic distributions

In the limit of large dataset size, the distribution of the test statistic q_μ for $0 \leq \hat{\mu} \leq \mu$ is [132]

$$q_\mu \approx \frac{(\mu - \hat{\mu})^2}{\sigma^2}, \tag{7.17}$$

and the best-fit signal strength $\hat{\mu}$ is normally distributed with standard deviation σ and mean equal to the true strength μ' of the signal. The approximation in Eq. 7.17, referred to as the *asymptotic approximation* for the test statistic, is obtained by neglecting terms of order $1/\sqrt{N}$, where N is the dataset size.

The width σ can be approximated with the aid of the *Asimov dataset* – a hypothetical dataset in which the contents of each bin are equal to their expected values, and all nuisance parameters have their nominal values. Letting $q_{\mu,A}$ be the value of the test statistic evaluated on the Asimov dataset, and making the same approximation as above, we obtain

$$\sigma^2 \approx \frac{(\mu - \mu')^2}{q_{\mu,A}}. \tag{7.18}$$

Under the asymptotic approximation, the value of CL_s can be expressed analytically as

$$CL_s = \frac{1 - \Phi(\sqrt{q_\mu})}{\Phi(\sqrt{q_{\mu,A}} - \sqrt{q_\mu})}, \tag{7.19}$$

where Φ is the standard gaussian cumulative distribution. This can be used to find the value of μ where $CL_s = 0.05$ without resorting to (often expensive) toy MC simulation. The expected exclusion limit at confidence level α, and its $N\sigma$ error bands, can be computed as

$$\mu_{\mathrm{up}\,+N} = \sigma(\Phi^{-1}(1 - \alpha\Phi(N)) + N). \tag{7.20}$$

Here the median expected limit is obtained by choosing $N = 0$.

For small dataset size, the asymptotic approximation does not hold, and it is found to yield overly optimistic limits [130].

7.12 Limits on simplified gluino model cross sections

In Figure 7.33 we show the expected and observed limits on simplified models of gluinos decaying to third-generation quarks. The left panel shows exclusion curves obtained for the branching ratio values indicated in Figure 7.32. The right panel shows the worst-case limits, which are obtained using the following procedure:

1. For each $(m_{\mathrm{gluino}}, m_{\mathrm{LSP}})$ mass point, find the choice of gluino branching ratios to $t\bar{t}\tilde{\chi}_1^0$ and $b\bar{b}\tilde{\chi}_1^0$ that gives the worst (highest) expected limit on the signal strength;

2. For each mass point, report the observed limit for the choice of branching ratio determined in step 1.

The limits are generally strongest for gluino decays to b quarks, and weakest for decays to top quarks. The decay to $b\bar{t}\tilde{\chi}_1^+/t\bar{b}\tilde{\chi}_1^-$ leads to poor limits at low $\tilde{\chi}_1^0$ mass, where the $\tilde{\chi}_1^0$ has little momentum either in the $\tilde{\chi}_1^\pm$ decay frame or in the lab frame, resulting in low values of E_T^{miss} and R^2. For LSPs with masses of a few hundred GeV, we exclude gluino masses below 1650 GeV assuming decays to bottom quarks, and below 1600 GeV assuming decays to top quarks.

In Figure 7.34 we show the limits on the T1qqqq model. At low LSP mass, we exclude gluino masses below about 1400 GeV.

Limits from the fit-based search

To compute limits using the fit-based background prediction, we perform the full-region fit described in Section 6.3. The fit function parameters b, n, M_R^0, R_0^2, and N are treated as nuisance parameters and allowed to float freely during the calculation of the test statistic. Systematic uncertainties on the signal template shape are included in the same way as in the MC-based procedure.

The fit-based method provides expected limits that are nearly the same as those from the MC-based method. Observed limits from the fit-based method are consistent with the expected limits within statistical uncertainty.

7.13 T2tt and the 'top corridor' region

In addition to considering gluino production, we also use the results of the search to place limits on the T2tt model of stop production, in which the stop decays into a top quark and the LSP. This model requires special care when setting limits, particularly in the region $m_{\text{stop}} \approx m_{\tilde{\chi}_1^0} + m_t$. In this 'top corridor' region, most of the stop's rest energy is converted into the masses of its decay products. Unless the squark is highly boosted, therefore, the LSP will be produced with very little momentum, and consequently it will contribute little to the E_T^{miss}. The signal in this case looks very similar to a $t\bar{t}$ event. In particular, it will have a large presence in the one-lepton $t\bar{t}$ control region. For this type of signal, correct treatment of signal contamination is paramount.

Other issues further complicate the search interpretation for signals in the top corridor. In particular, because the signal is very similar to the $t\bar{t}$+jets background, the limits (or discovery significance) obtained from the search are finely sensitive to mismodeling by the Fastsim simulation framework. Small differences in, for example, the E_T^{miss} spectrum in Fastsim and full simulation events will be seized upon by the limit-setting algorithm and used to discriminate signal from background, even if the difference is an artifact of the simulation. This kind of effect can be corrected by detailed study of the simulation, but it may not go away entirely and may still influence the limit results for signals that look very background-like.

We simulated a T2tt signal in the top corridor with the full CMS simulation and compared it with the corresponding Fastsim sample. In the comparison it was found that the fast simulation sometimes produces spurious high-p_T jets. The presence of these jets cause E_T^{miss} to be severely overmeasured; this creates a long tail in the R^2 distribution. For a signal with little true E_T^{miss}, these mismodeled events drive the search sensitivity and must be carefully removed.

Another issue is that of MC sample statistics. The analysis has low selection efficiency for signal models with little E_T^{miss}, so the signal templates used in limit setting may have spikes due to Poisson fluctuations in the MC. Statistical uncertainty on the MC template is taken into account during limit setting, but nevertheless this kind of fluctuation can artificially improve the limit. Reweighting the signal MC events to have the correct distribution of the number of pileup vertices can accentuate this problem by producing events with large weights.

Additionally, the search sensitivity changes strongly with the value of $m_{\text{stop}} - m_{\tilde{\chi}_1^0} - m_t$, so it is necessary to simulate a large number of signal mass points in the corridor region in order to faithfully portray the exclusion limits there.

Based on these considerations, it is decided that the interpretation of the search for T2tt signals in the top corridor region, especially at low squark mass, is too easily confounded and is not performed for this version of the search. The exclusion limits for T2tt are displayed in Figure 7.34 with the top corridor region blanked out.

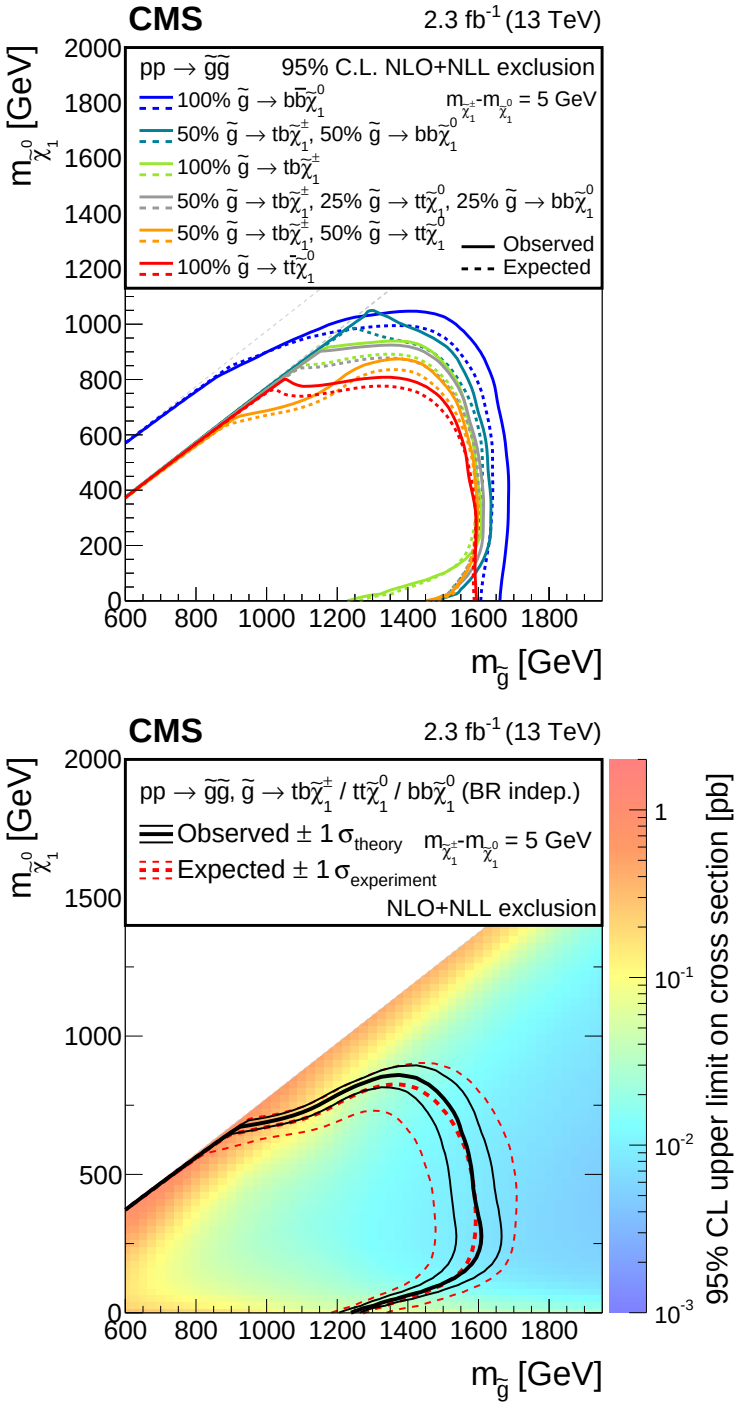


Figure 7.33: Top: expected and observed 95% CL upper limits on the gluino pair production cross section under various assumptions on the branching ratios to third-generation quarks. The two dashed gray lines indicate the locations where $m_{\tilde{g}} - m_{\tilde{\chi}_1^0} = 25$ GeV and 225 GeV, respectively. For $m_{\tilde{g}} - m_{\tilde{\chi}_1^0} < 225$ we consider only the $b\bar{b}\tilde{\chi}_1^0$ decay mode of the gluino. Bottom: the worst-case limit among all considered branching ratios for each choice of gluino and LSP mass.

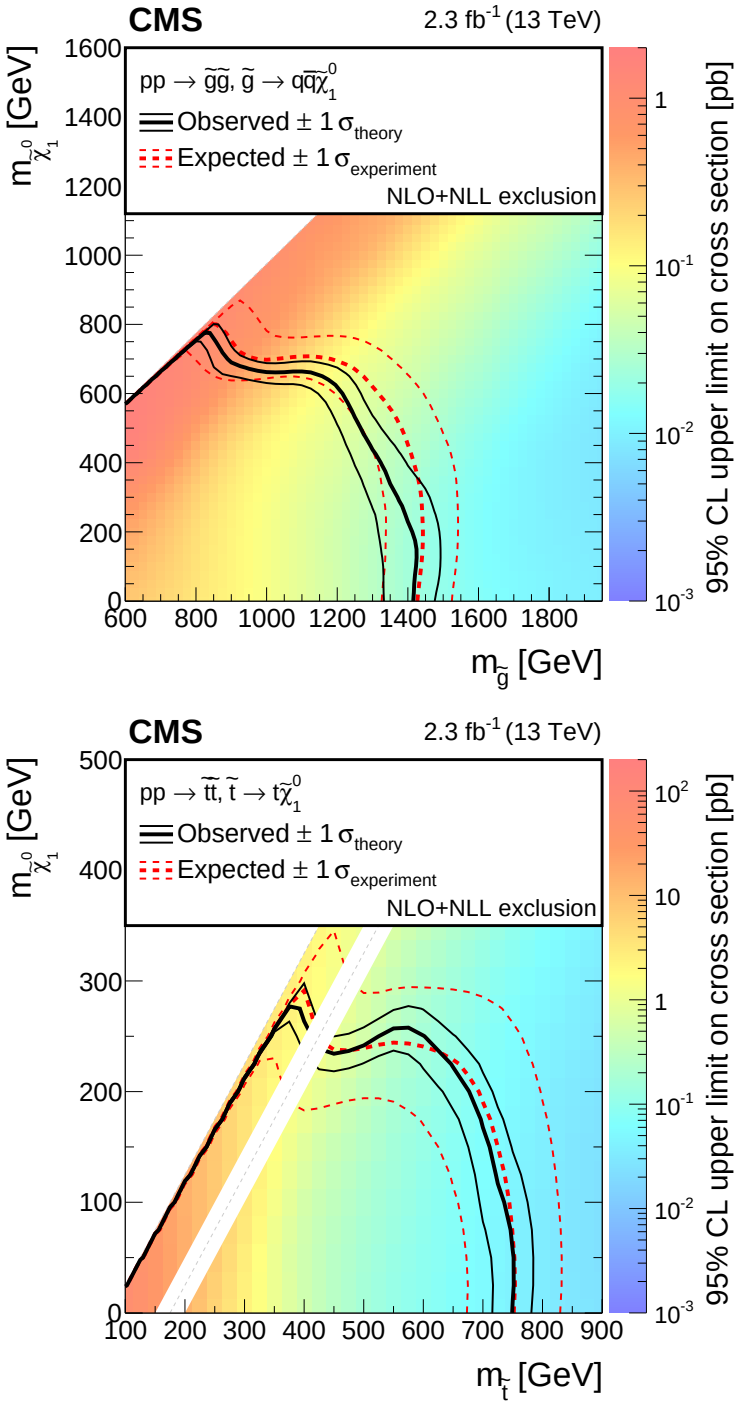


Figure 7.34: Top: expected and observed 95% CL upper limits on the T1qqqq simplified model of gluino production. Bottom: expected and observed limits on the T2tt simplified model of squark production, with the 'top corridor' region blanked out as discussed in Section 7.13.

Chapter 8

MONTE CARLO-BASED RAZOR SEARCH ON 2016 DATA

The MC-based razor search was repeated on the full dataset collected by CMS in 2016. A total of 35.9 fb^{-1} of data was certified good for physics, more than ten times the amount certified in 2015. Certain aspects of the analysis were modified to accommodate this larger dataset and to improve sensitivity to certain SUSY signals. The primary changes are:

- Analysis categories with fewer than four jets are added in order to increase sensitivity to squark production
- Events with seven or more jets are placed in separate search categories, increasing sensitivity to signals with multiple top quarks
- The analysis closure tests are performed in low, medium, and high jet multiplicity categories to accommodate the new search regions
- Additional corrections are performed in the $t\bar{t}$+jets, W($\rightarrow \ell\nu$)+jets, and Z($\rightarrow \nu\nu$)+jets control regions to ensure correct modeling in each b-tag category
- A new QCD background prediction method is introduced

These and other changes are discussed in the following sections. The fit-based background prediction was not used in this version of the analysis; see Chapter 9 for more details.

A second search for SUSY using razor variables, also on 2016 CMS data, was performed concurrently by another group. This analysis considers only events with jets tagged as originating from the hadronic decay of a boosted W boson or top quark. It is an updated version of a similar search performed on Run I data [133]. We statistically combine our results with those of the other group in order to increase the sensitivity to certain SUSY signals. The boosted razor analysis is particularly sensitive to SUSY models with large mass splitting between a heavy particle and its decay products.

8.1 Razor and lepton triggers in 2016

The LHC delivered instantaneous luminosities of 1.34×10^{34} cm^{-2}s^{-1} or higher throughout much of 2016, exceeding its original design luminosity. At this luminosity, the rates of the primary razor triggers used in 2015 were unsustainably high. They were therefore removed from the HLT menu and replaced with the 'backup' paths discussed in Section 6.4 and indicated in Figure 6.3. The backup triggers impose a tighter hyperbolic cut on the razor variables:

$$(M_R + 300) \times (R^2 + 0.25) > 270 \text{ GeV}. \tag{8.1}$$

The efficiency of the tighter razor triggers is measured using the method described in Section 6.4. The measured efficiencies are displayed in Figure 8.1. To achieve close to 100% efficiency in the search region and QCD control region, we tighten the baseline M_R cut in the zero-lepton search regions to 650 GeV.

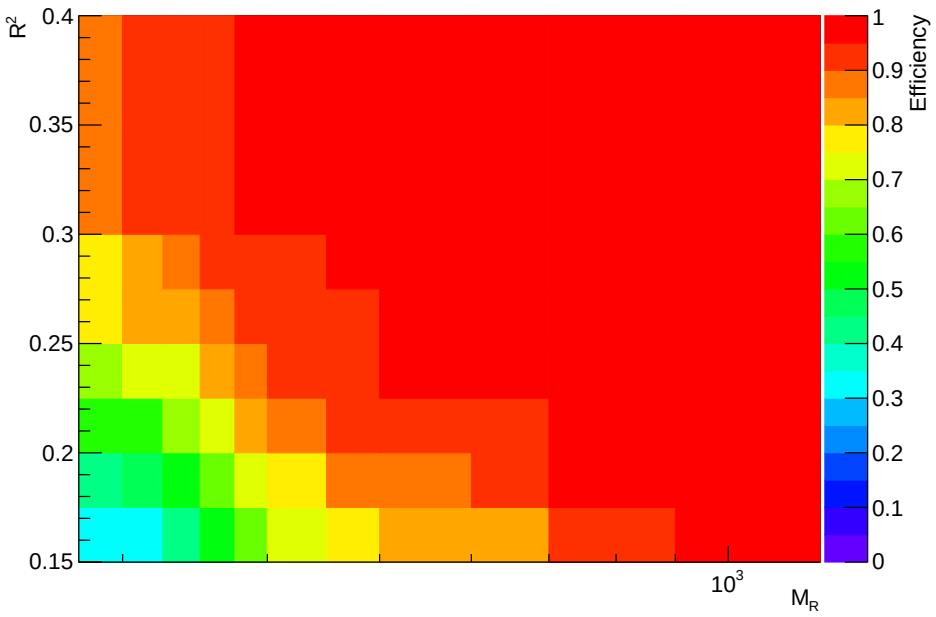


Figure 8.1: Efficiency of the 2016 hadronic razor triggers, displayed in bins of M_R and R^2.

The inclusive single lepton triggers in the 2016 HLT menu have tighter cuts than in 2015 as well. We account for this by increasing the selection thresholds for electrons and muons. In the 2016 search we require tight muons to have $p_T > 25$ GeV and tight electrons to have $p_T > 30$ GeV.

8.2 MC simulation

A new collection of MC samples is used to model the background for the 2016 search. The samples are generated with a pileup distribution more closely matching that measured in 2016 CMS data, and are large enough to provide adequate statistics for a 36 fb^{-1} analysis.

MC samples simulated at NLO are available for some of the leading background processes. These yield better 'out-of-the-box' agreement with data in some cases, which makes the needed corrections smaller. We use $t\bar{t}$+jets samples generated at NLO with Powheg, and W($\rightarrow \ell\nu$)+jets samples generated with aMC@NLO. The NLO $t\bar{t}$+jets MC is reweighted in order to correct for a known mismodeling of the top quark p_T spectrum.

8.3 Study of T1tttt signal sensitivity

The T1tttt simplified SUSY model features pair-produced gluinos that each undergo the decay $\tilde{g} \rightarrow t\bar{t}\tilde{\chi}_1^0$. In the 2015 search, most of the sensitivity to this model comes from the one-lepton search categories, with the zero-lepton categories contributing relatively little to the limit. We investigate in detail the source of sensitivity to this SUSY model and modify the search to achieve increased sensitivity to this type of signal in the hadronic search categories.

Binning in jet multiplicity

The CMS hadronic SUSY analyses with the best sensitivity to T1tttt in 2015 specifically target this model by including search categories with high jet multiplicity [134, 135]. In contrast, the 2015 razor search does not categorize events according to jet multiplicity, except to reject events having fewer than 4 jets.

The jet multiplicity distribution for a T1tttt signal is shown in Figure 8.2 for events passing the baseline razor analysis selection. We find that placing events with $N_{\text{jets}} \geq 7$ into dedicated search categories isolates this signal in a region that has dramatically lower background contamination.

Baseline selection cuts

In Figure 8.3 we show the M_R-R^2 and H_T-E_T^{miss} distributions for the T1tttt signal examined above. We see that the baseline cut in the zero-lepton search categories

($R^2 > 0.3$) rejects a significant fraction of the signal events. The cut in the one-lepton categories ($R^2 > 0.2$) accepts substantially more signal events. Meanwhile, an analysis based on H_T and E_T^{miss} can cut very tightly without rejecting any signal.

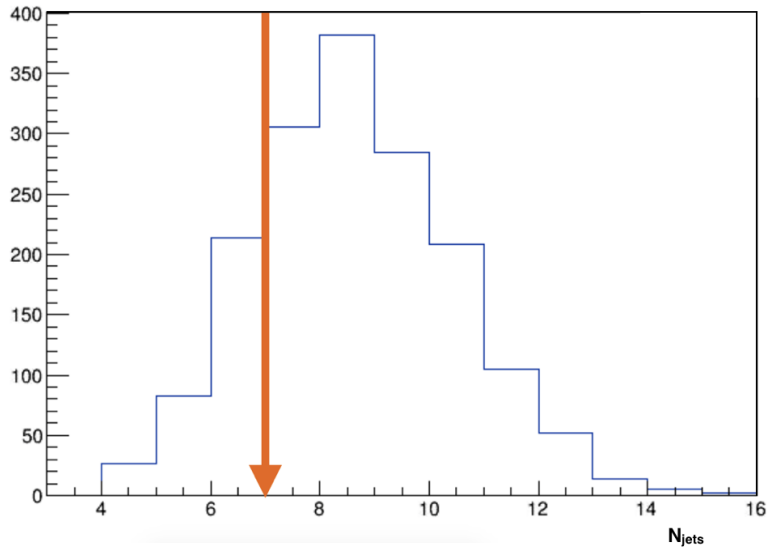


Figure 8.2: Distribution of the number of selected jets for a T1tttt signal with $m_{\text{gluino}} = 1900$ GeV and $m_{\tilde{\chi}_1^0} = 100$ GeV. The orange arrow indicates the chosen cut value of $N_{\text{jets}} \geq 7$.

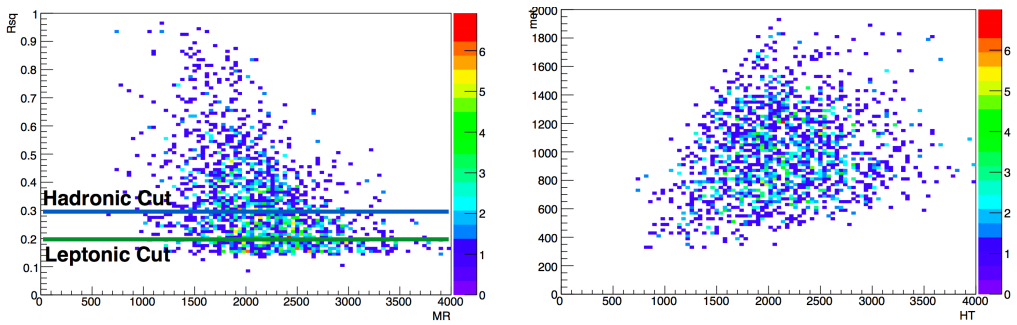


Figure 8.3: Left: distribution of R^2 and M_R for the T1tttt model with $m_{\text{gluino}} = 2000$ GeV and $m_{\tilde{\chi}_1^0} = 100$ GeV. The baseline R^2 cuts for the zero- and one-lepton razor search regions are indicated by horizontal lines. Right: the distribution of H_T and E_T^{miss} for the same model.

Figure 8.4 shows the mean values of R^2 and E_T^{miss} for T1tttt events as a function of the gluino mass, for fixed LSP mass. We see that E_T^{miss} increases roughly linearly with gluino mass, while R^2 increases more slowly at higher masses. This suggests that the R^2 threshold in razor variable searches should be kept low even when considering very high-mass signals.

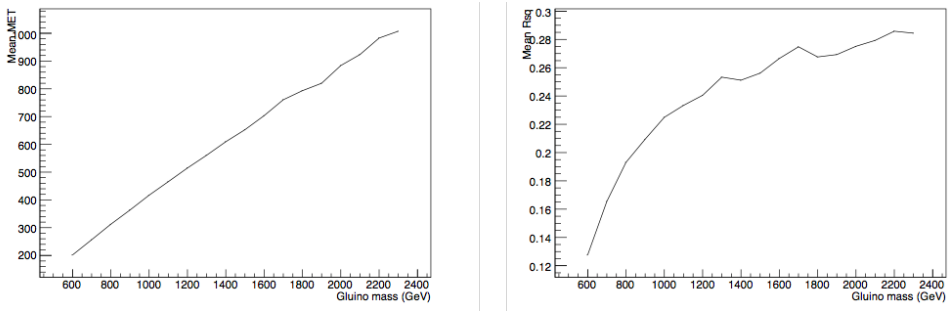


Figure 8.4: Left: mean value of E_T^{miss} for T1tttt events with $m_{\tilde{\chi}_1^0} = 100$ GeV, plotted versus m_{gluino}. Right: mean value of R^2 for the same models.

This examination considers signal events only and does not examine the ratio of signal to background in the search regions. It does not imply anything about the sensitivity of an analysis based on R^2 versus one based on E_T^{miss}; it merely suggests that loosening the baseline R^2 cut in the analysis will increase the number of signal events selected.

8.4 Changes to event categorization

We make two significant changes to the event categorization procedure in light of the results in Section 8.3. First, we place events with seven or more selected jets into dedicated search categories, denoted 'Seven-jet' and 'Lepton Seven-jet.' This adds additional complexity to the background prediction but significantly improves the sensitivity to SUSY decay chains with multiple top quarks. Second, we decrease the baseline R^2 cut in the zero-lepton categories from 0.3 to 0.2 at high M_R. The razor triggers, which cut on the product of M_R and R^2, are fully efficient at $R^2 = 0.2$ for events with $M_R > 1600$ GeV. This is indicated by the turn-on curve shown in Figure 8.5.

In addition to these changes, we also add a zero-lepton search category, the 'Dijet' category, containing events with $2 \leq N_{\text{jets}} \leq 3$. This category is sensitive to production of squarks decaying to bottom quarks or light quarks. To keep the number of search regions to a manageable level, we combine the Muon and Electron Multijet categories into a single 'Lepton Multijet' category.

The full list of search regions considered in the 2016 version of the analysis is given in Table 8.1.

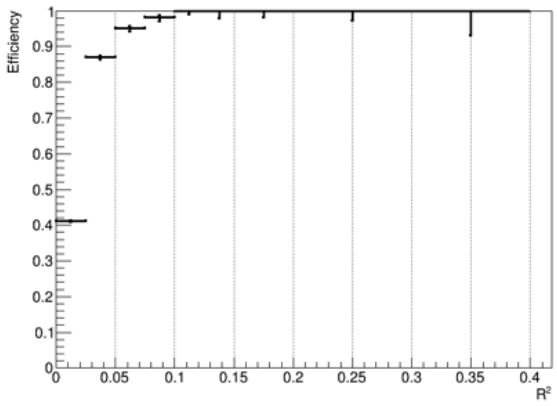


Figure 8.5: Razor trigger efficiency as a function of R^2 for events passing inclusive single electron triggers and having $M_R > 1600$ GeV.

Analysis Category	N_{leptons}	N_{jets}	b-tag Bins
Dijet	0	2-3	0, 1, ≥ 2
Multijet	0	4-6	0, 1, 2, ≥ 3
Seven-jet	0	$\geq$7	0, 1, 2, ≥ 3
Lepton Multijet	1	4-6	0, 1, 2, ≥ 3
Lepton Seven-jet	1	$\geq$7	0, 1, 2, ≥ 3

Table 8.1: List of search categories in the 2016 razor search.

8.5 Combination with boosted razor search

The results of this search are statistically combined with those of a second razor search that targets SUSY events with hadronically decaying boosted W bosons or

top quarks. This 'boosted razor search' has a favorable signal-to-background ratio for certain SUSY signals. We make two changes to our search to ensure that the two analyses can be combined:

- Reject events with W- or top-tagged events from the zero-lepton search categories, ensuring that the two analyses use non-overlapping event samples
- Change the baseline jet selection cuts to $p_T > 30$ GeV and $|\eta| < 2.4$, the values used in the boosted razor search. (In the 2015 search the cuts are $p_T > 40$ GeV, $|\eta| < 3.0$.)

The searches share nuisance parameters for some systematic uncertainties, such as the jet energy scale, but can otherwise be treated as independent.

W- and top-tagging is performed on wide jets clustered using anti-k_T with $R = 0.8$. Jets are tagged if they pass selection criteria on the n-subjettiness variables [136], the soft drop mass [137], and, in the case of top-tagging, the subjet CSV b-tag scores. These criteria assess whether a jet contains subjets consistent with the decay products of a hadronically decaying W or top. Jets with $p_T > 200$ GeV and $p_T > 400$ GeV are considered for W- and top-tagging, respectively.

Rejecting events with W- or top-tagged jets reduces the sensitivity of the search slightly. This sensitivity is recovered in the combination with the boosted razor analysis. In some cases the combined analysis yields stronger limits than what is obtained with our search alone.

8.6 $t\bar{t}$+jets and W($\to \ell\nu$)+jets background predictions

The procedure for correcting the $t\bar{t}$+jets and W($\to \ell\nu$)+jets MC samples is similar to that described in Section 7.4. The one-lepton control sample is selected in the same way as before, and is divided into $t\bar{t}$+jets and W($\to \ell\nu$)+jets-enriched samples. Background contamination is subtracted using the MC estimates, and correction factors are derived in bins of M_R and R^2. The NLO $t\bar{t}$+jets and W($\to \ell\nu$)+jets MC samples feature better modeling of the hadronic recoil than the LO samples, so the needed corrections are generally smaller than they were in the 2015 search. In Figure 8.6 we show the unrolled M_R-R^2 distributions in the $t\bar{t}$ and W+jets one-lepton control samples. In Figure 8.7 we show the correction factors and their statistical uncertainties.

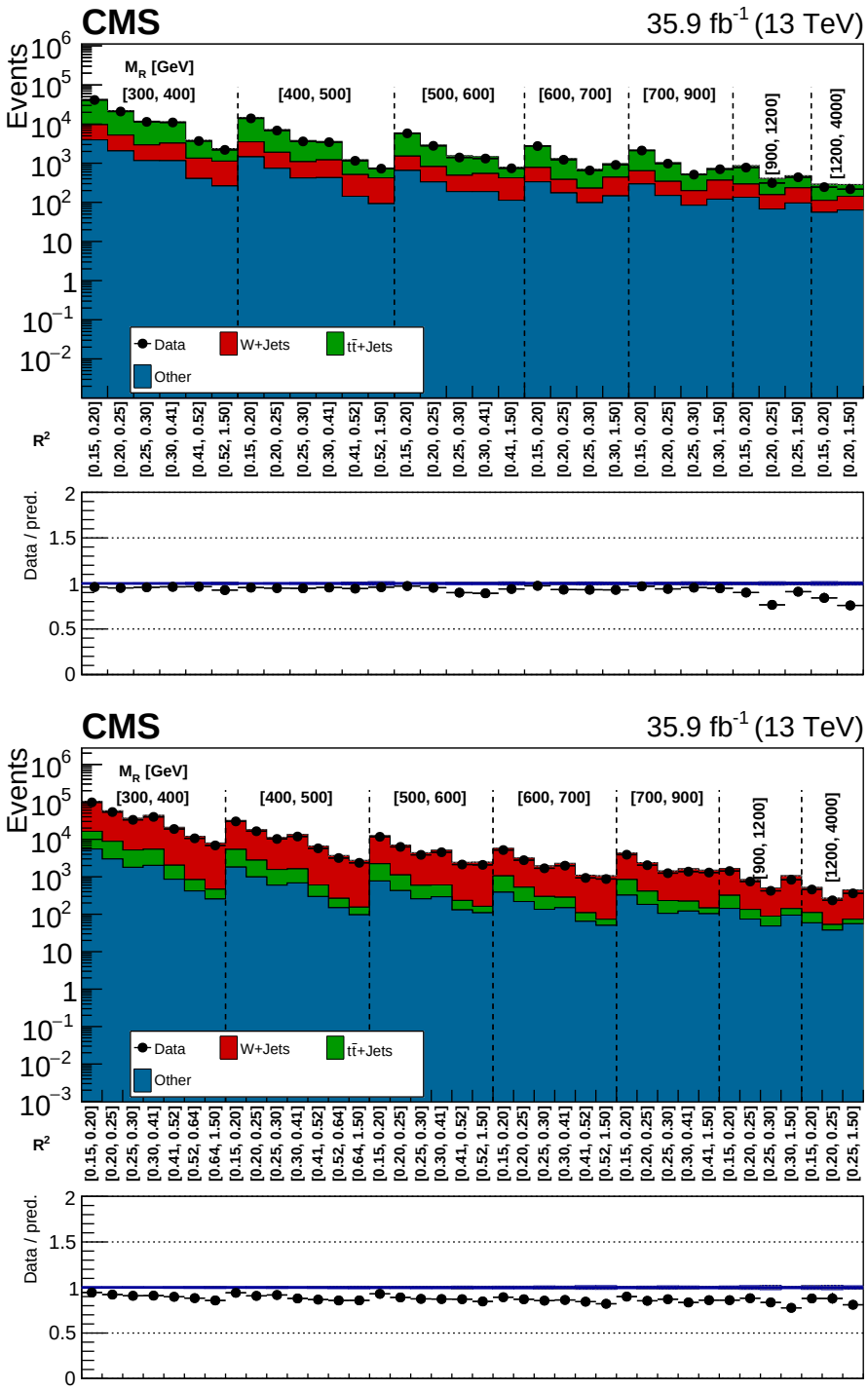


Figure 8.6: Unrolled M_R-R^2 distributions in the $t\bar{t}$ (top) and W+jets (bottom) one-lepton control samples. The $t\bar{t}$+jets simulation prediction in the W+jets control sample has been corrected using the scale factors derived in the $t\bar{t}$ one-lepton sample.

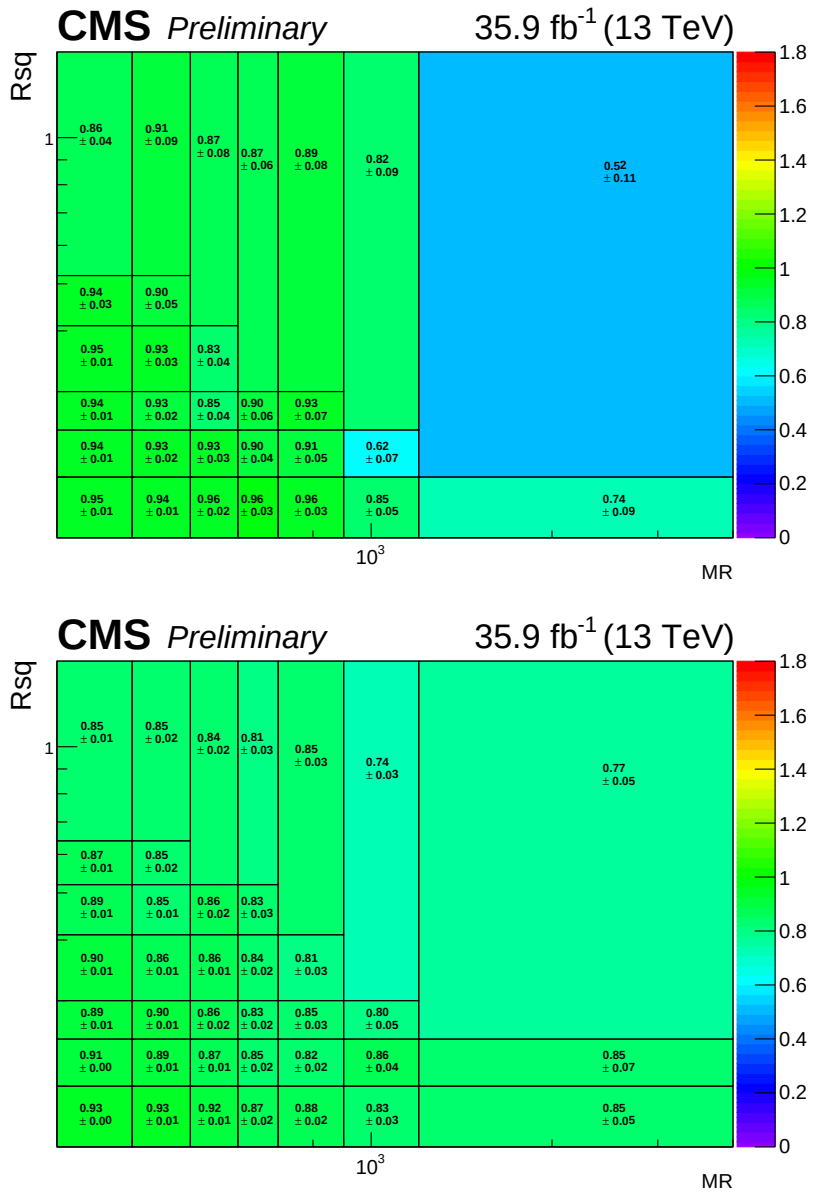


Figure 8.7: Correction factors derived in bins of M_R and R^2 for $t\bar{t}$+jets (top) and W($\rightarrow \ell\nu$)+jets (bottom) MC.

An additional correction is then derived for the N_{jets} spectrum. This is performed in three bins corresponding to the three N_{jets} ranges considered in the search: 2-3 jets, 4-6 jets, and 7 or more jets. The correction is derived separately for the $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets MC. The correction factors are listed in Table 8.2.

Process	N_{jets}	Correction factor
$t\bar{t}$+jets	2-3	1.02 ± 0.01
$t\bar{t}$+jets	4-6	1.02 ± 0.005
$t\bar{t}$+jets	$\geq$7	1.03 ± 0.03
W($\rightarrow \ell\nu$)+jets	2-3	0.96 ± 0.004
W($\rightarrow \ell\nu$)+jets	4-6	0.99 ± 0.01
W($\rightarrow \ell\nu$)+jets	$\geq$7	1.18 ± 0.25
Z($\rightarrow \nu\nu$)+jets	2-3	1.04 ± 0.01
Z($\rightarrow \nu\nu$)+jets	4-6	0.90 ± 0.01
Z($\rightarrow \nu\nu$)+jets	$\geq$7	0.98 ± 0.06

Table 8.2: Corrections binned in N_{jets} for the $t\bar{t}$+jets, W($\rightarrow \ell\nu$)+jets, and Z($\rightarrow \nu\nu$)+jets MC samples. These corrections are applied in addition to the corrections binned in M_R-R^2.

Further M_R corrections and closure test

After applying the above scale factors to the MC, we observe that the M_R distributions in the one-lepton control region for some N_{jets} and N_{b-tags} categories need further correction (see Figure 8.8). Instead of assigning a systematic uncertainty based on observed data-MC discrepancies, as in 2015, we correct these M_R spectra to match the data.

After performing this final set of corrections, we examine the R^2 spectra in each N_{jets} and N_{b-tags} category and observe no further systematic mismodeling. We assign a systematic uncertainty on the $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets backgrounds equal to the uncertainty on the data/MC ratio in each bin of R^2. This uncertainty reflects how accurately we can measure the level of agreement between data and MC in this control region.

Dilepton $t\bar{t}$+jets and lepton veto modeling

The selection of the two-lepton $t\bar{t}$+jets-enriched control sample is the same as described in Section 7.4. The control sample is divided into 2-3, 4-6, and $\geq$ 7 jet categories. Instead of checking the data-MC agreement in the M_R-R^2 plane as in 2015, we perform this check in the one-dimensional M_R spectrum, where a small trend is seen in the ratio of data to MC (see Figure 8.9). We assign a systematic uncertainty on the dilepton $t\bar{t}$+jets prediction equal to the size of this trend. The uncertainty is represented in the likelihood by a single nuisance parameter that interpolates between the M_R shape in MC and that measured in data.

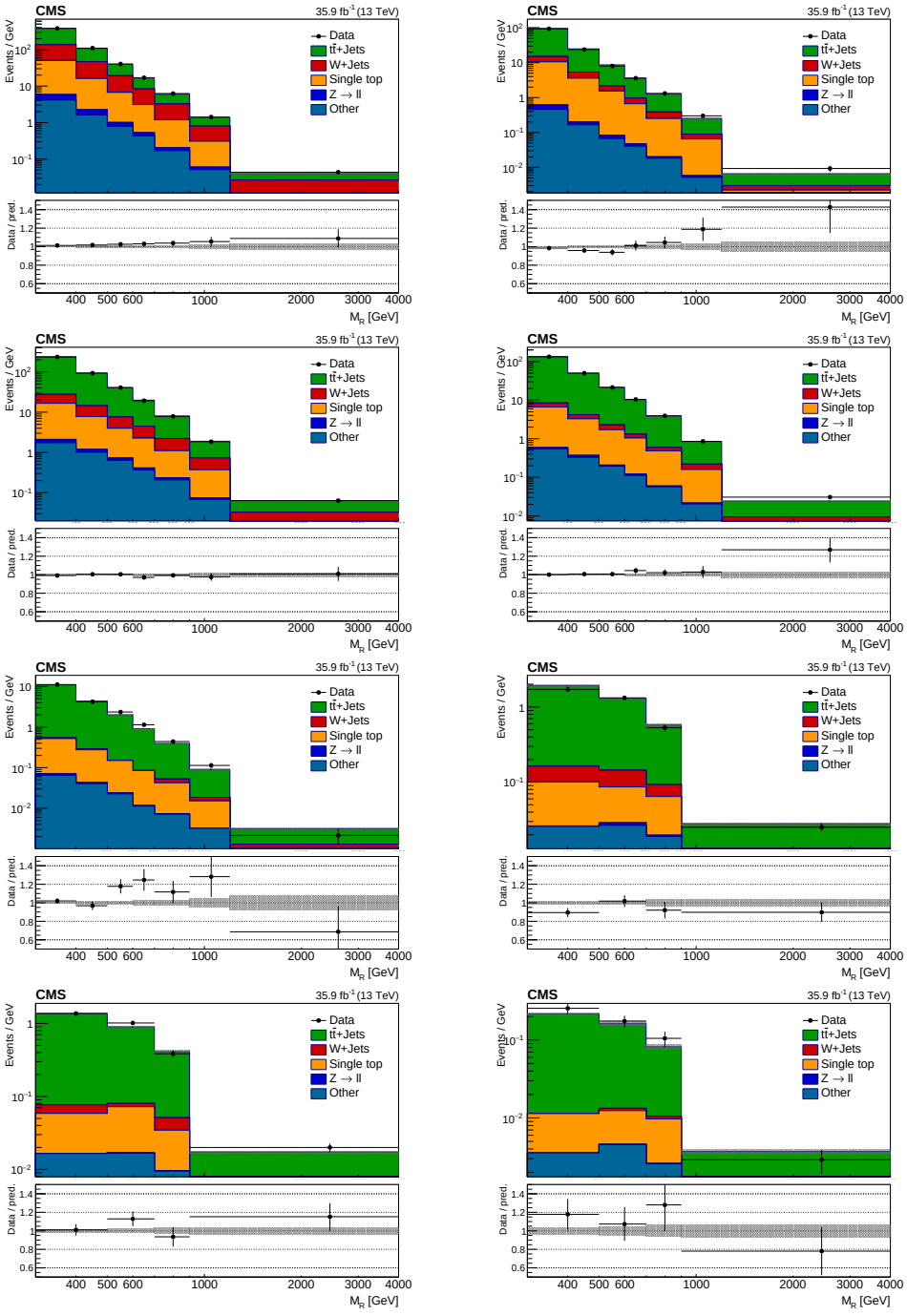


Figure 8.8: M_R distributions for events in the one-lepton control region with different numbers of selected jets and b-tagged jets. The $t\bar{t}$+jets and W($\rightarrow \ell\nu$)+jets MC predictions are reweighted using M_R-R^2 and N_{jets} correction factors. Top row: events with 2-3 jets and 1 (left) or 2 (right) b-tags. Second row: events with 4-6 jets and 1 (left) or 2 (right) b-tags. Third row: (left) events with 4-6 jets and 3 b-tags, (right) events with $\geq$ 7 jets and 1 b-tag. Bottom row: events with $\geq$7 jets and 2 (left) or $\geq$ 3 (right) b-tags.

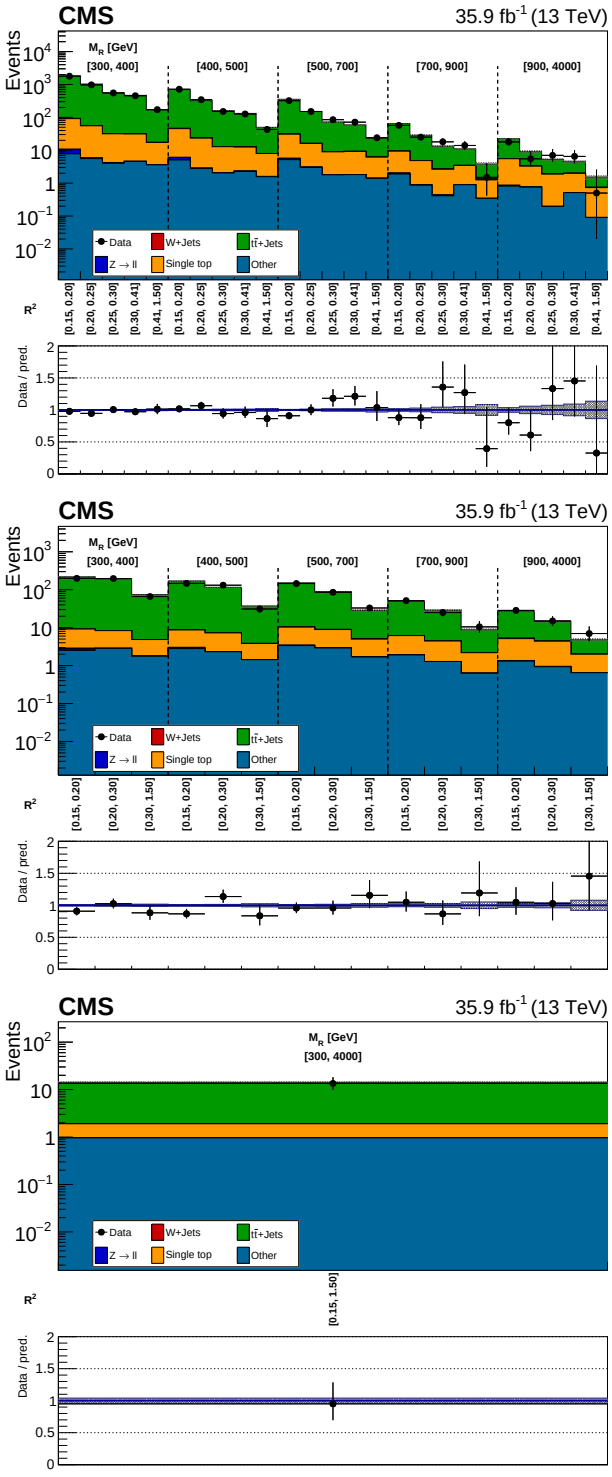


Figure 8.9: Unrolled M_R-R^2 distributions in the $t\bar{t}$ dilepton control sample for events with 2-3 (top), 4-6 (middle), and ≥7 (bottom) selected jets. The $t\bar{t}$+jets MC prediction is reweighted using the correction factors derived in the one-lepton control sample.

The region with 7 or more jets suffers from very low statistics, so the check in this region is performed in a single large bin, in which the overall numbers of data and MC events are compared.

The modeling of the lepton and tau vetos is checked in a control region defined the same way as in Section 7.4. As there, we compare the MC prediction with the data in bins of the lepton p_T and η, and correct for the efficiencies of the $\Delta\phi_R$ and m_T cuts to estimate the effect of any discrepancies on the zero-lepton search region. The check is performed separately for events with 2-3, 4-6, and 7 or more jets. We show the lepton and tau p_T distributions in Figure 8.10. A systematic uncertainty is assigned to cover a small trend in the data/MC ratio in the lepton p_T distribution. Given no significant mismodeling observed in the lepton η distributions, we take the size of the uncertainty on the data/MC ratio as a further systematic representing how accurately we can measure the data-MC agreement.

8.7 Z($\rightarrow \nu\nu$)+jets background prediction

The strategy for the Z($\rightarrow \nu\nu$)+jets background prediction is similar to that described in Section 7.5, with minor changes introduced mainly to support the 2016 analysis strategy. In the following subsections we discuss the changes and show the results of each part of the procedure.

γ+jets control region

In the 2016 analysis we do not use prescaled photon triggers to select events for the γ+jets control region. Instead, we select events passing the loosest unprescaled photon trigger and require the presence of a photon with $p_T > 185$ GeV. This avoids the complication of dealing with trigger prescales and does not have a significant impact on the measured MC correction factors.

As in 2015, the photon purity is estimated using a template fit approach. We change the fit variable from $\sigma_{i\eta i\eta}$ to the charged isolation sum, as this is seen to yield more robust fits. The purity fits are performed in two-dimensional bins of M_R and R^2 for both barrel and endcap photons. Example fits are shown in Figure 8.11. The measured purity values change smoothly with M_R and R^2 (see Figure 8.12); they are observed to vary from 76% to 100% in the barrel and from 63% to 100% in the endcaps.

The unrolled M_R-R^2 distributions in data and MC in the γ+jets control region are

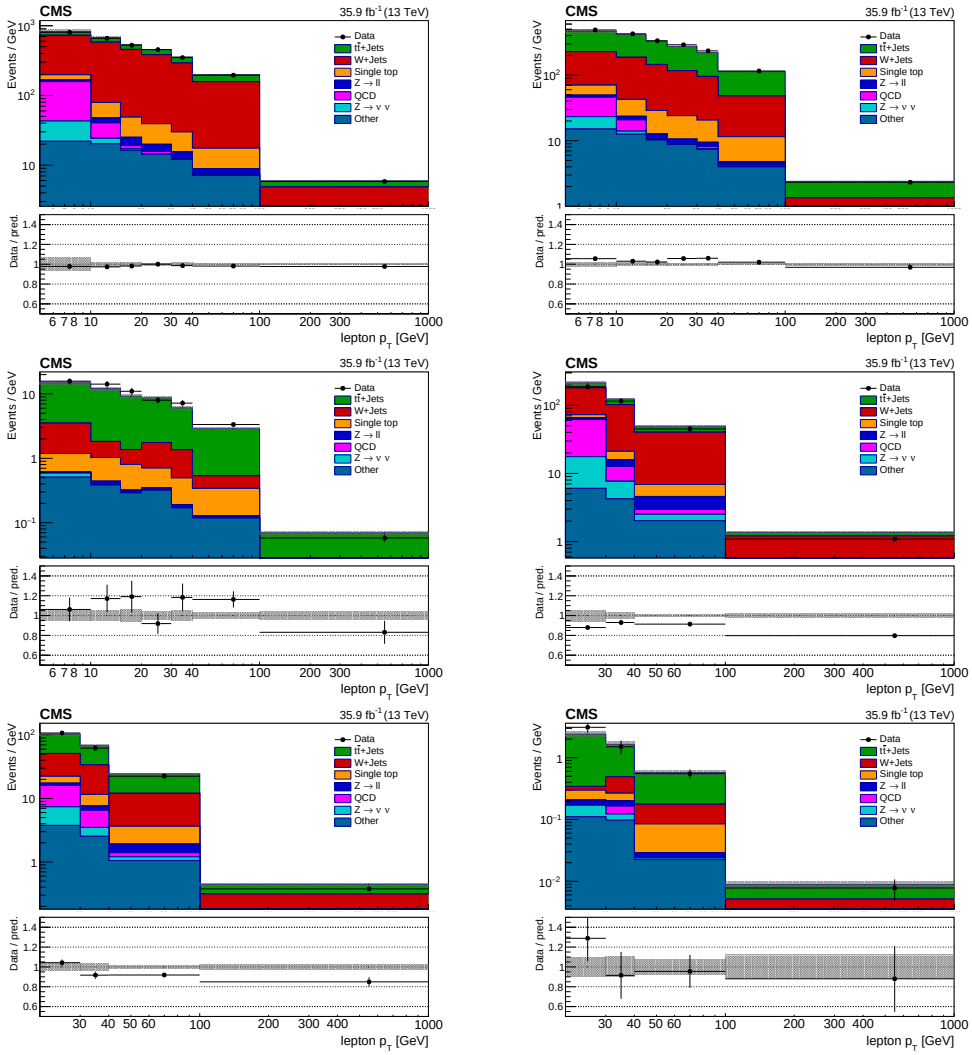


Figure 8.10: Lepton p_T distributions in the veto lepton and tau control samples after the application of all MC correction factors. Top row: veto lepton events with 2-3 (left) and 4-6 (right) jets. Middle row: (left) veto lepton events with ≥7 jets, (right) veto tau events with 2-3 jets. Bottom row: veto tau events with 4-6 (left) and ≥7 (right) jets.

shown in Figure 8.13, and the MC correction factors are shown in Figure 8.14. As in the one-lepton control sample, we derive an additional correction binned in N_{jets}. The needed correction factors are listed in Table 8.2.

We observe that further corrections to the M_R distribution in some N_{jets} and N_{b-tags} categories are needed (see Figure 8.15). As in the one-lepton control sample, we correct the M_R distributions to match the data in each region. After performing these corrections, we examine the R^2 spectra in each N_{jets} and N_{b-tags} category. Observing no further systematic mismodeling by the MC, we take the size of the

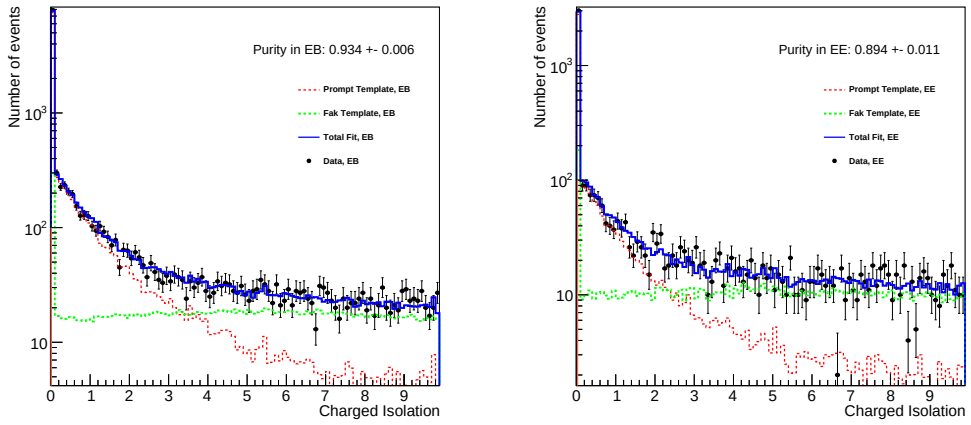


Figure 8.11: Example charged isolation fits used to estimate the purity of γ+jets events in the photon control region, for photons in the barrel (left) and endcap (right) regions.

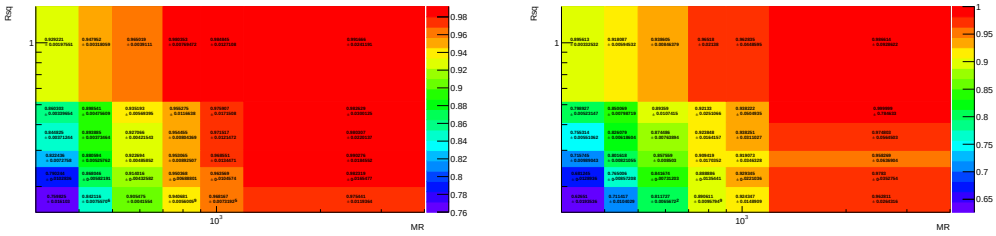


Figure 8.12: Fitted γ+jets purity values in each bin of M_R and R^2 for events with photons in the barrel (left) and endcap (right) regions.

uncertainty on the data/MC ratio in each R^2 bin as a systematic expressing how accurately we can measure the data-MC agreement.

Closure test in two-lepton control region

The cross sections used to normalize the γ+jets MC prediction in this version of the analysis are measured at leading order in QCD. We use the two-lepton DY($\rightarrow \ell\ell$)+jets enriched control sample to obtain the correct normalization for the Z($\rightarrow \nu\nu$)+jets correction factors derived in the γ+jets sample. This is done by applying the M_R-R^2 and N_{jets} correction factors to the DY($\rightarrow \ell\ell$)+jets MC, and determining the additional scale factor needed to normalize the MC to the data. To account for possible variation of the ratio between the Z and γ processes with the number of b-tagged jets, we perform this procedure separately for events with zero, one, and two b-tags. The measured values of the ratio are listed in Table 8.3.

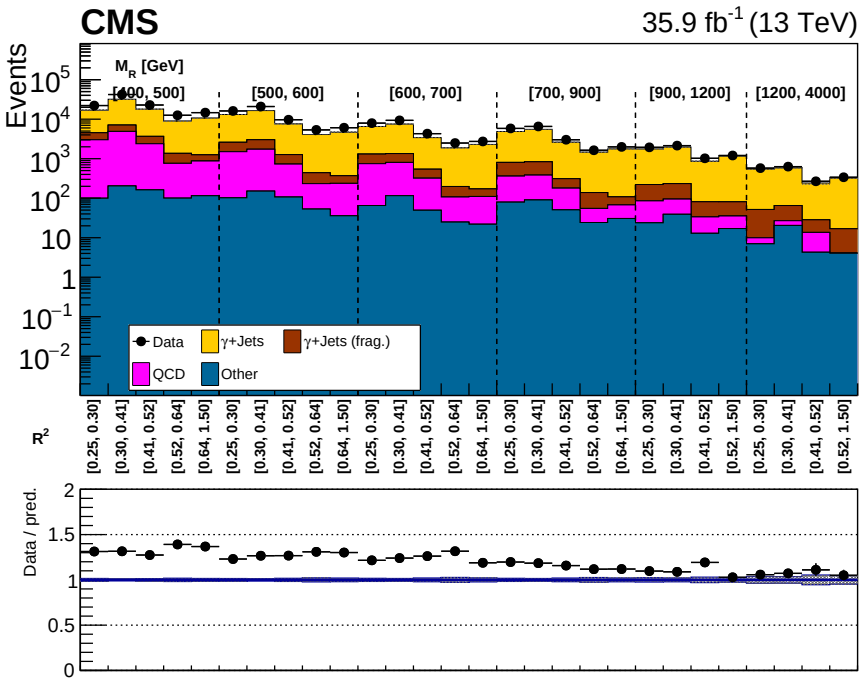


Figure 8.13: Unrolled M_R-R^2 distribution for events in the γ+jets control sample.

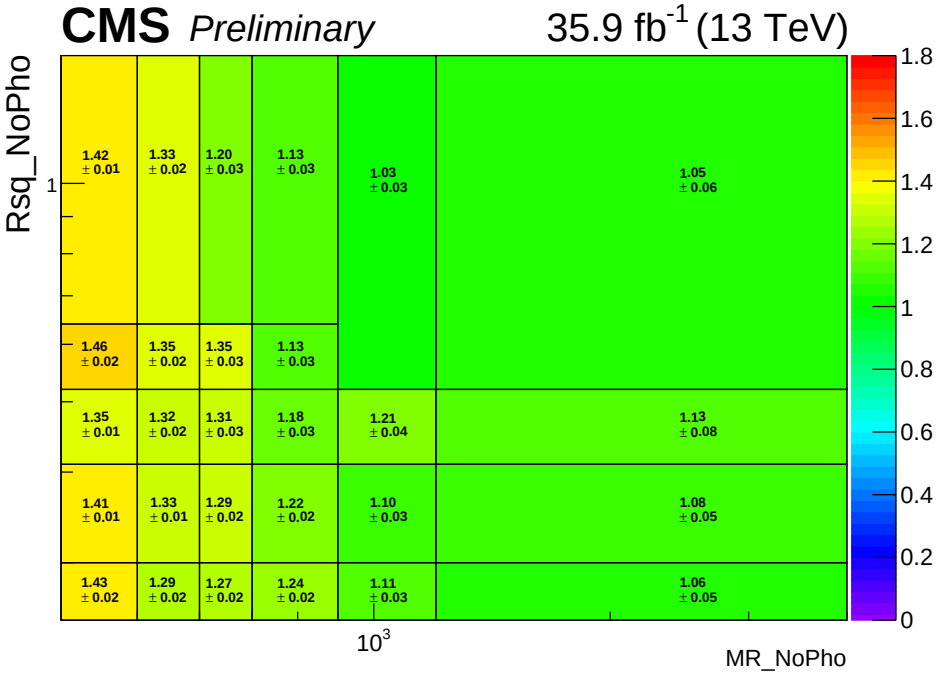


Figure 8.14: Correction factors for Z($\rightarrow \nu\nu$)+jets MC derived in bins of M_R and R^2 in the photon control region.

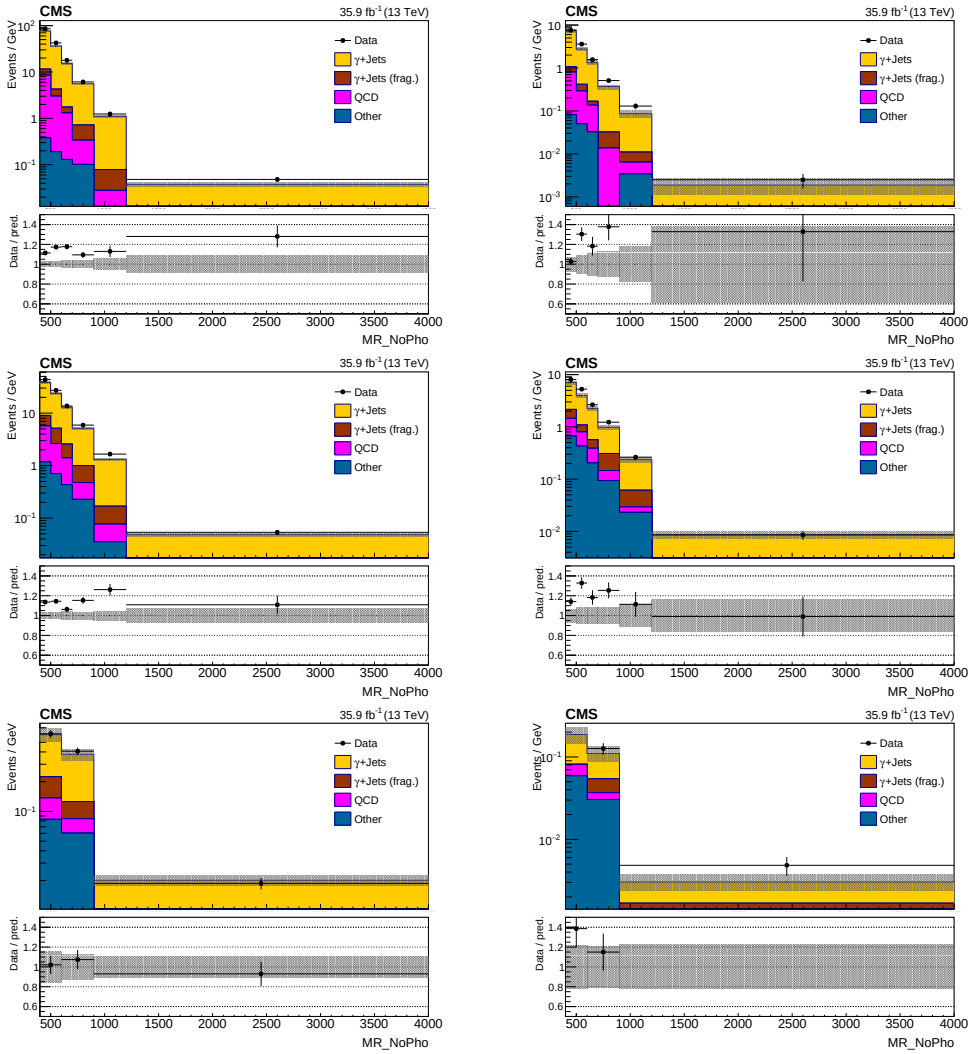


Figure 8.15: M_R distributions for events in the photon control region with different numbers of selected jets and b-tagged jets. The top, middle, and bottom rows correspond to events with 2-3, 4-6, and ≥7 jets, respectively. The left and right columns correspond to events with 1 and 2 b-tagged jets, respectively. The γ+jets MC simulation has been corrected in bins of M_R-R^2 and N_{jets}.

N_{b-tags}	Correction factor
1	0.88 ± 0.03
2	0.74 ± 0.1
≥3	1.35 ± 0.63

Table 8.3: Corrections of the Z($\rightarrow \nu\nu$)+jets scale factors to account for the variation of the Z/γ ratio with b-tag multiplicity. The corrections are measured with respect to events with 0 b-tagged jets.

After normalizing the DY($\rightarrow \ell\ell$)+jets MC to the data, we perform a closure test

to check the modeling of the M_R and R^2 shapes in each N_{jets} category. The one-dimensional M_R distributions are shown in Figure 8.16. We observe small trends in the data/MC ratio in both M_R and R^2; the sizes of these trends are taken as systematic uncertainties on the Z($\to \nu\nu$)+jets modeling. The 7-jet category suffers from very limited statistics, so the closure test there is performed in one large bin to check the overall normalization.

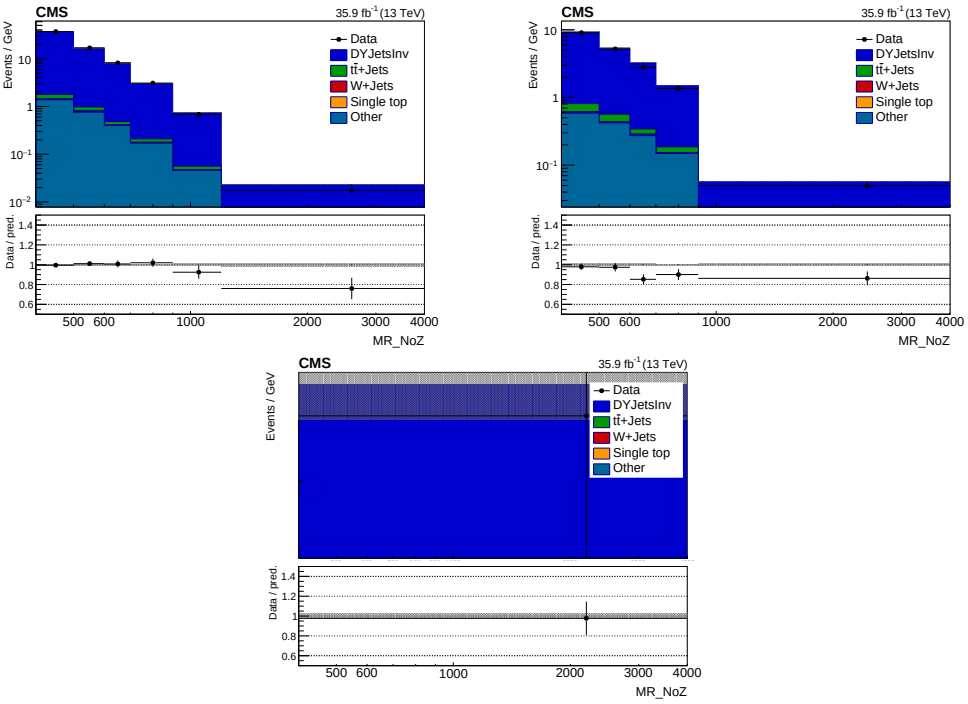


Figure 8.16: M_R distributions in the DY($\to \ell\ell$)+jets dilepton control region, after applying all Z($\to \nu\nu$)+jets MC correction factors to the DY($\to \ell\ell$)+jets MC, for events with 2-3 (left), 4-6 (right), and ≥7 (bottom) selected jets.

Cross-check in one-lepton control region

As in 2015, we compare the M_R-R^2 correction factors derived in the γ+jets control region (after normalizing them using the data in the DY($\to \ell\ell$)+jets control region) with an independent set of correction factors derived in the one-lepton control region. Because the Z($\to \nu\nu$)+jets MC sample used in the search region is generated at LO, we use a LO W($\to \ell\nu$)+jets MC sample generated using MadGraph for this cross-check instead of the NLO sample used in the other analysis regions.

The measured correction factors, and the percentage differences between the correction factors derived using γ+jets and W($\to \ell\nu$)+jets, are shown in Figure 8.17.

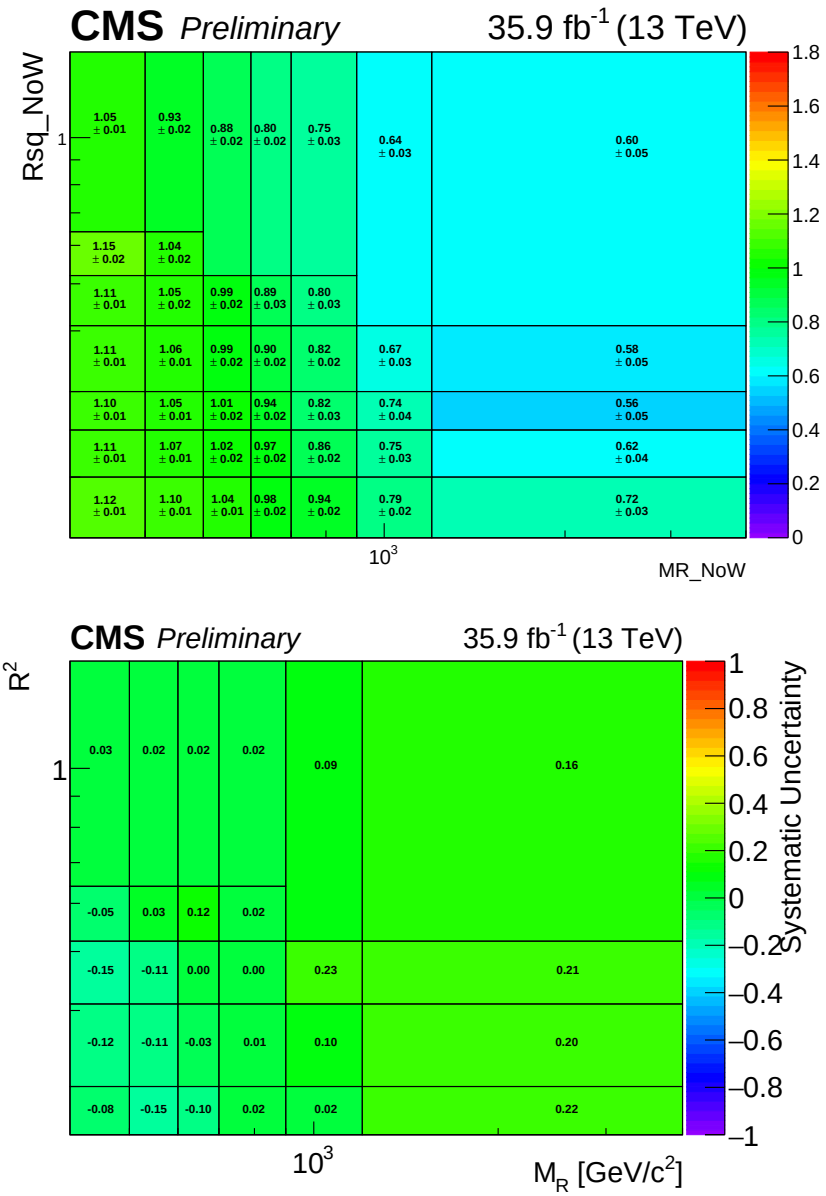


Figure 8.17: : Top: MC correction factors for Z($\rightarrow \nu\nu$)+jets derived in bins of M_R and R^2 using the W+jets one-lepton control sample. Bottom: percentage difference between the scale factors in the top plot and those derived in the photon control sample.

8.8 QCD background prediction

As in the 2015 search, the QCD multijet background is predicted using transfer factors ζ that relate the yields in high and low $\Delta\phi_R$ regions. However, in the much larger 2016 dataset it is necessary to model the R^2 dependence of the transfer factors, which was previously neglected. A roughly linear R^2 dependence is observed

in the low-R^2 control region and in QCD MC simulation. We model the transfer factors with a linear function of M_R and R^2:

$$\hat{\zeta}(M_R, R^2) = Q_0 + Q_1\, M_R + Q_2\, R^2 + Q_3\, M_R\, R^2, \tag{8.2}$$

where the parameters Q_i are optimized using the data in the low-R^2 region. The linear fit is performed simultaneously in the rectangular regions $[650\,\text{GeV}, 1600\,\text{GeV}] \times [0.2, 0.3]$ and $[1600\,\text{GeV}, 4000\,\text{GeV}] \times [0.1, 0.2]$ in the space of M_R and R^2. We then use the linear fit to model the QCD transfer factors in the search region by extending the fitted function to higher R^2.

The transfer factor fit is performed separately for the 2-3, 4-6, and $\geq$7 jet analysis regions. Due to the limited statistics of the $\geq$7 jet category, we neglect the M_R-dependence of ζ there and fit only for the R^2-dependence. Measured and fitted transfer factors in the low-R^2 region are illustrated in Figure 8.18.

We observe that the transfer factors are nearly independent of the number of b-tagged jets (see Figure 8.19). To quantify this, we perform the following test. For $N = 0, 1, 2$ and each M_R-R^2 bin in the low-R^2region, we compute the transfer factor ζ twice, once using events with $N_{b-tags} = N$ and once events with $N_{b-tags} \neq N$. We then compute the significance of the difference as

$$\frac{\zeta_N - \zeta_{\neq N}}{\sigma}, \tag{8.3}$$

where σ is the uncertainty on the numerator. The distribution of significances is shown for each value of N in Figure 8.20. The pull distributions are approximately standard normal, with a bias of approximately 0.3σ for $N = 0$. We increase the uncertainty on the measured transfer factors by 30 percent in order to cover the systematic uncertainty on the b-tag independence assumption.

We also observe that the fraction of QCD multijet events at each b-tag multiplicity is independent (within the available statistics) of $\Delta\phi_R$, M_R, and R^2. We take advantage of this fact to increase the statistical precision of the QCD estimate. The number of QCD multijet events in the signal region can be obtained as

$$N_{\text{QCD}}(\text{low}\,\Delta\phi_R) = \alpha_b \times \hat{\zeta}(M_R, R^2) \times N_{\text{QCD}}(\text{high}\,\Delta\phi_R), \tag{8.4}$$

The proportions α_b for $b = 0, 1, 2, \geq 3$ and their uncertainties are computed in the low-R^2 QCD sideband region for each jet multiplicity category.

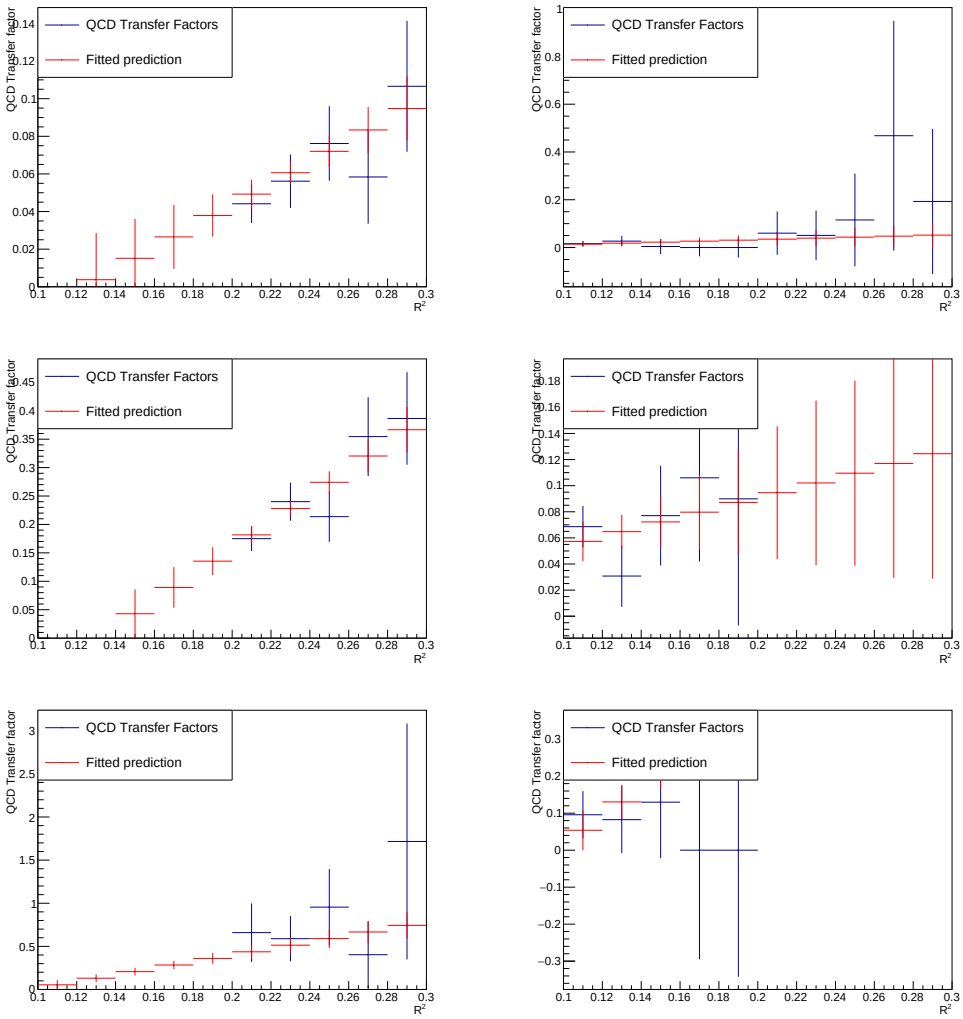


Figure 8.18: One-dimensional slices showing the QCD transfer factors ζ and the fitted ζ values in the low-R^2 sideband region. The top, middle, and bottom rows show the transfer factors for events with 2-3, 4-6, and $\geq$7 jets, respectively. The left column shows the lowest M_R bin, which extends from 650 to 900 GeV for the seven-jet category and from 650 to 800 GeV for the other categories.

Cross-checks of QCD modeling

We perform two checks to ensure that the QCD prediction method describes the transfer factors adequately in the search region. First, we perform the transfer factor fit in QCD MC and compare the fitted transfer factors with those obtained in simulation at high R^2. Because of limited sample statistics in the QCD MC simulation, this test is performed inclusively in the number of selected jets. The results are illustrated in Figure 8.21; they show good agreement between the fit and the measured transfer factors.

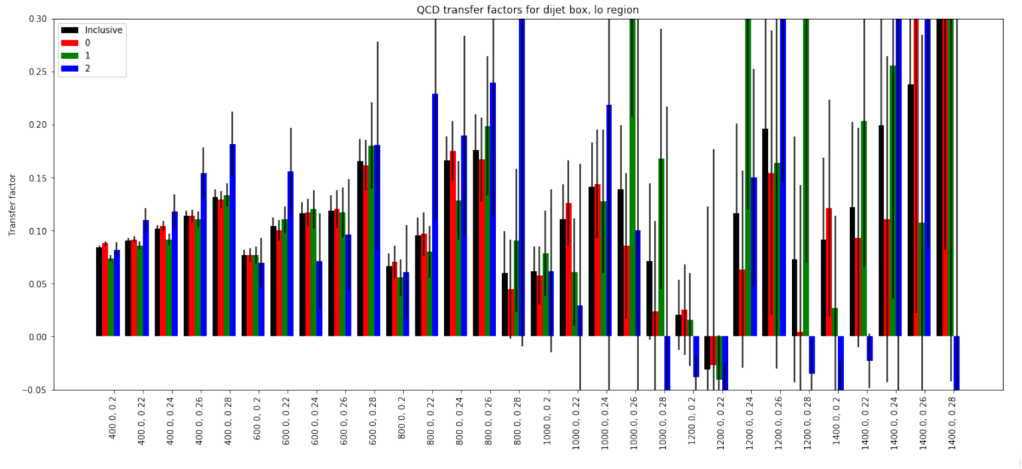


Figure 8.19: Bar plot showing the QCD transfer factors computed in each bin of M_R and R^2 in the 2-3 jet category for events with zero (red), one (green), two (blue), or any number (black) of b-tagged jets. The four colored bars in each group are generally consistent in height with one another, which indicates the independence of the QCD transfer factors of the number of b-tags.

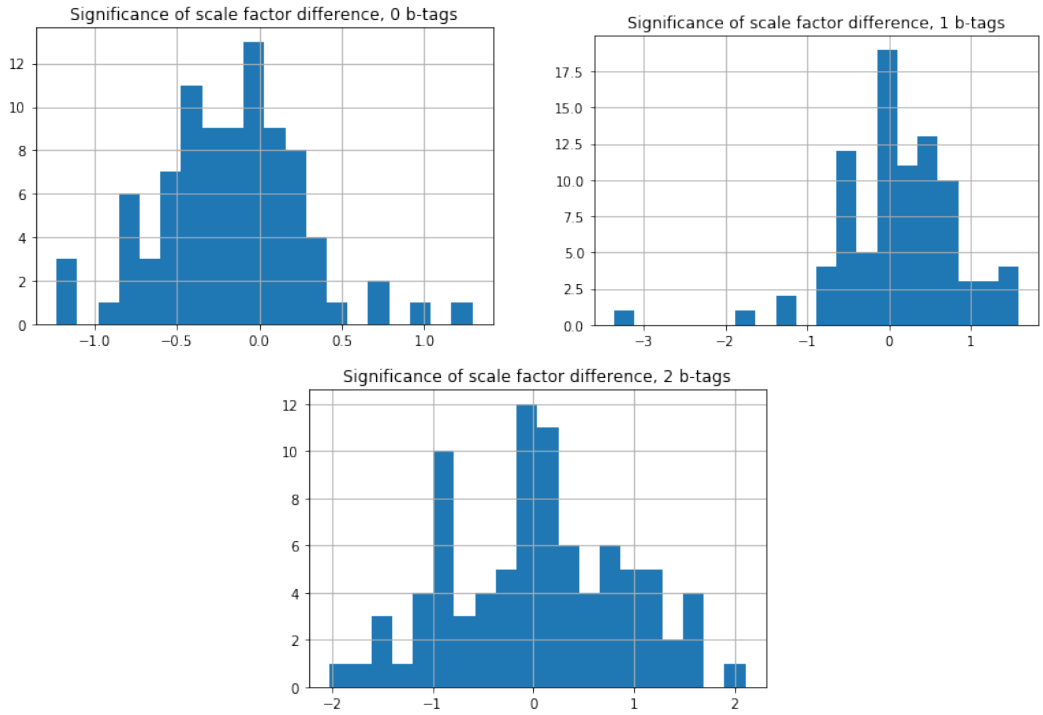


Figure 8.20: Pull distributions indicating the significance of the difference between QCD transfer factors computed using events with different numbers of b-tagged jets.

The region $550 < M_R < 650$ GeV is not part of the analysis search region, but the razor triggers are fully efficient there for $R^2 > 0.3$. For the second cross-check

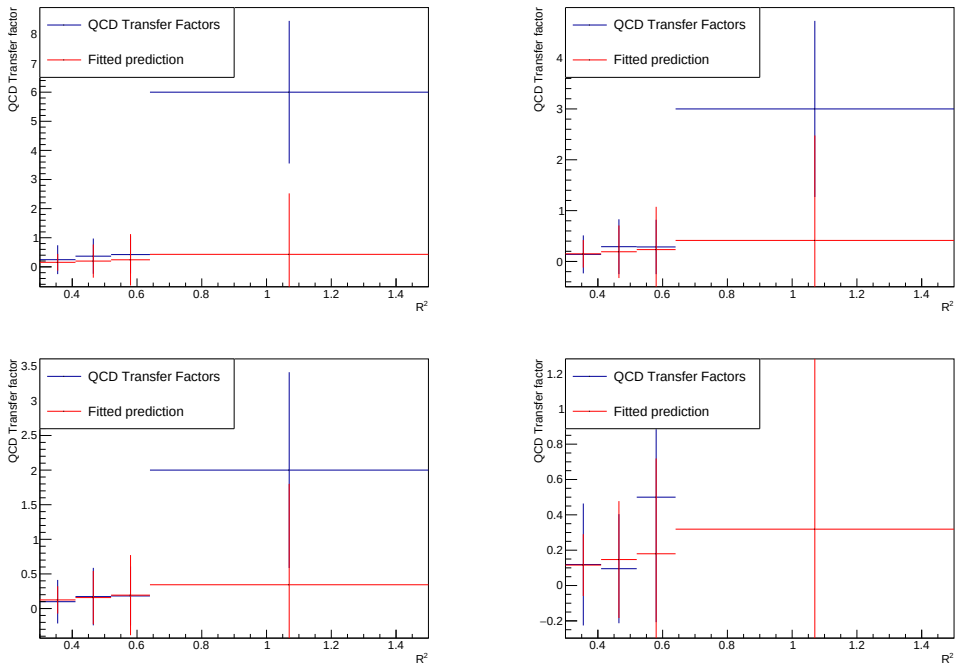


Figure 8.21: Observed and predicted QCD transfer factors plotted in slices of M_R for QCD MC in the high R^2 region. The example slices correspond to the M_R bins 650-800 GeV (upper left), 800-1000 GeV (upper right), 1400-1600 GeV (lower left), and 1600-4000 GeV (lower right).

of the QCD method, we use this region to check the predicted QCD transfer factors. The comparison between the measured and predicted transfer factors is shown in Figure 8.22 for each jet multiplicity category. We observe that the linearity assumption holds in the high-R^2 region and that the fit function describes the transfer factors accurately there.

8.9 Systematic uncertainties on the background prediction

The set of instrumental and theoretical uncertainties considered is similar to that in the 2015 search. These uncertainties are summarized in Table 8.4. The uncertainties on the M_R-R^2 MC correction factors for $t\bar{t}$+jets, W($\rightarrow \ell\nu$)+jets, and Z($\rightarrow \nu\nu$)+jets are shown as percentages in Figure 8.23. Uncertainties estimated based on closure tests in the analysis control regions are listed in Table 8.5.

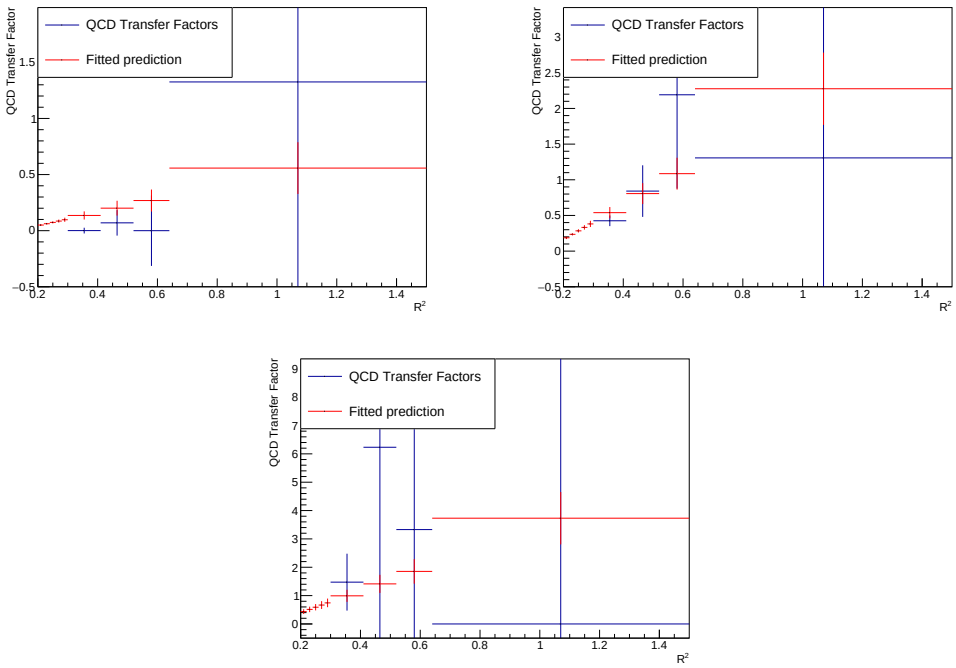


Figure 8.22: Comparison of the QCD transfer factors with the fit-based predictions in the region $550 < M_R < 650$ GeV for the 2-3 (left), 4-6 (right), and ≥7 (bottom) jet categories.

Uncertainty source	Typical uncertainty size
Jet energy scale	6-16%
Lepton selection efficiency	1-2%
b-tagging efficiency	1-7%
b mistag	2-20%
Fac. and renorm. scales	10-25%
Pileup	1-3%
Integrated luminosity	2.6%
Monte Carlo statistics	$1/\sqrt{N_{MC}}$

Table 8.4: Summary of instrumental and theoretical sources of uncertainty on the background prediction.

8.10 Search region results

The unblinded data in the Dijet, Multijet, Lepton Multijet, Seven-jet, and Lepton Seven-jet categories are shown for each b-tag subcategory in Figures 8.24-8.33 along with the MC-based background prediction. The data are consistent with the standard model predictions. We observe small local excesses in the Multijet 1 and ≥3 b-tag regions, and a notable deficit of events in one bin of the Seven-jet 1 b-tag region. These are large enough to have non-negligible impacts on some of the limit

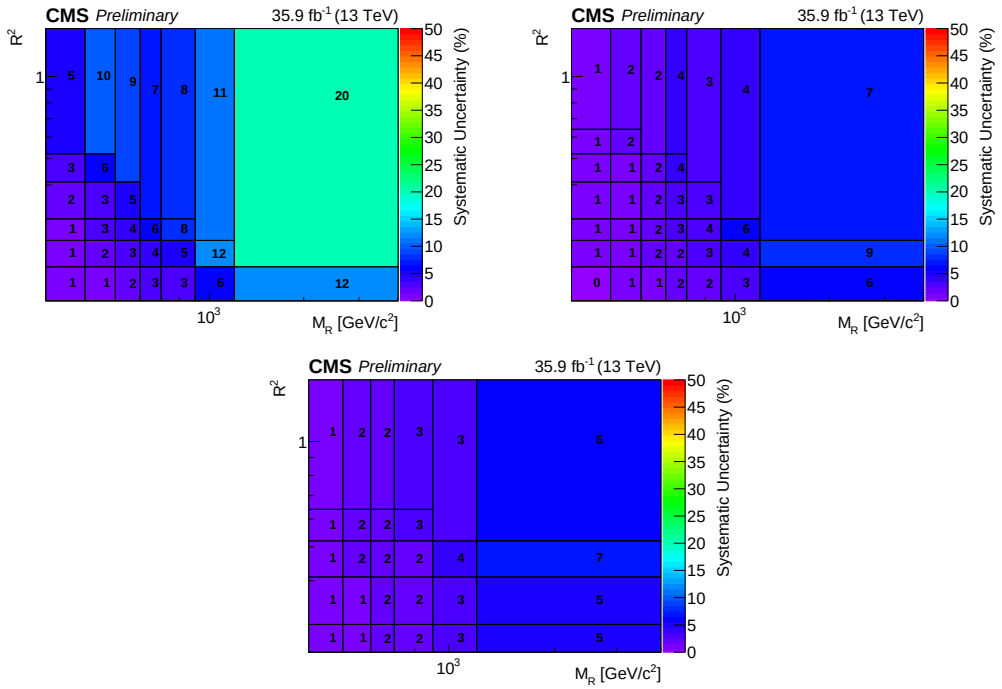


Figure 8.23: Statistical uncertainties on the $t\bar{t}$+jets (left), W($\rightarrow \ell\nu$)+jets (right), and Z($\rightarrow \nu\nu$)+jets (bottom) correction factors, expressed as a percentage.

Uncertainty source	Background process(es)	Typical uncertainty size	Correlated?
One-lepton closure test (R^2)	$t\bar{t}$+jets, W($\rightarrow \ell\nu$)+jets	1-95%	N
$t\bar{t}$ dilepton closure test	2ℓ $t\bar{t}$+jets	1-12%	Y
Veto lepton / tau p_T closure tests	$t\bar{t}$+jets, W($\rightarrow \ell\nu$)+jets	2-50%	Y
Veto lepton / tau $\lvert\eta\rvert$ closure tests	$t\bar{t}$+jets, W($\rightarrow \ell\nu$)+jets	2-40%	N
γ+jets closure test (R^2)	Z($\rightarrow \nu\nu$)+jets	1-40%	N
γ+jets vs. W+jets scale factors	Z($\rightarrow \nu\nu$)+jets	0-23%	Y
DY($\rightarrow \ell\ell$)+jets dilepton closure test	Z($\rightarrow \nu\nu$)+jets	1-25%	Y

Table 8.5: Summary of systematic uncertainties estimated in analysis control regions. The rightmost column indicates whether the uncertainty is correlated or uncorrelated from bin to bin.

plots shown at the end of the chapter, but they are most likely statistical fluctuations and are expected based on the large number of search bins considered.

8.11 SUSY signal treatment

The limit-setting procedure is the same as in the 2015 search. MC simulation of SUSY simplified models is again performed using the CMS Fastsim framework. Some features of the signal simulation require correction or reweighting:

- As in the 2015 search, the Fastsim lepton and b-jet identification efficiencies

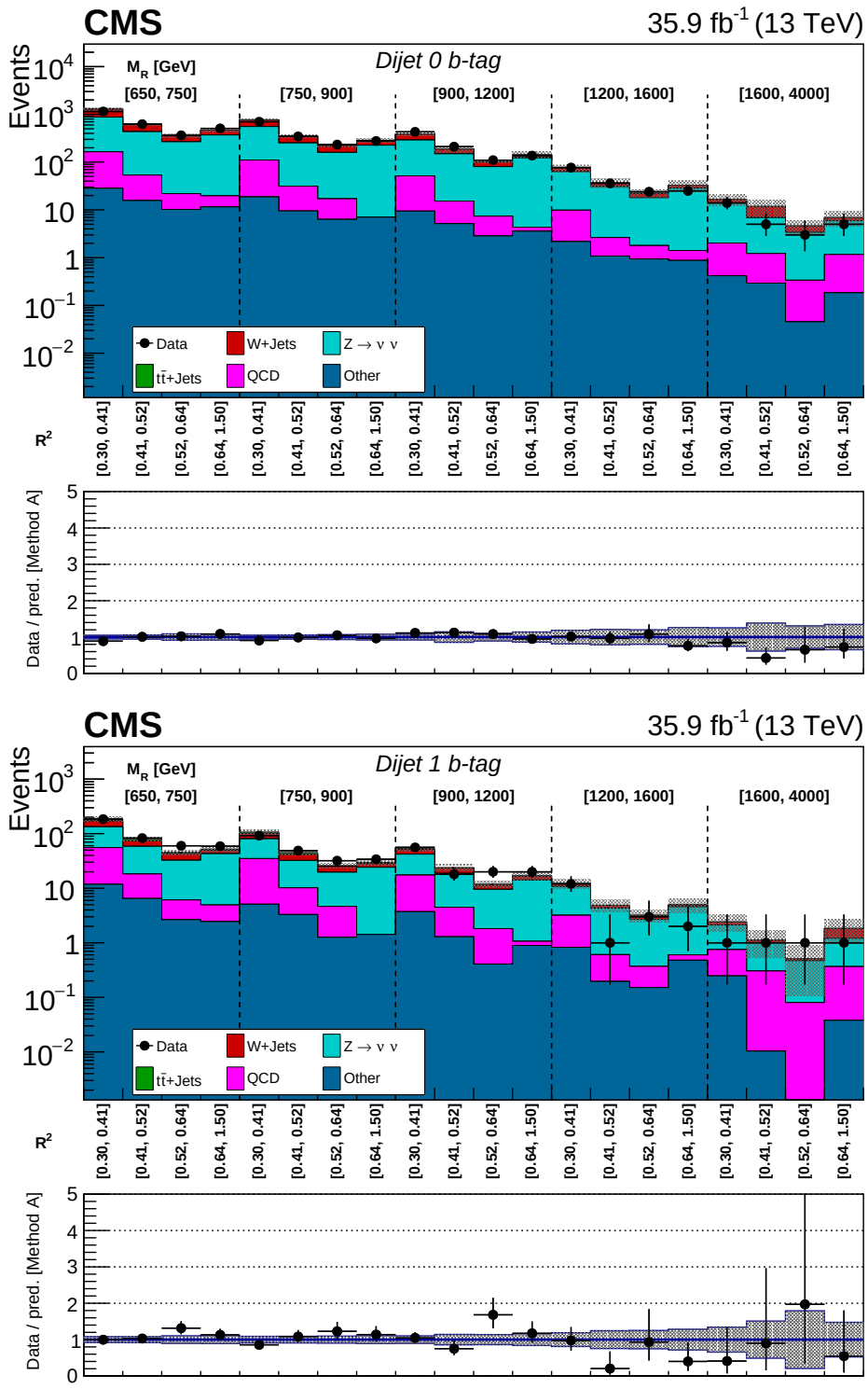


Figure 8.24: Observed data counts in each bin of the Dijet 0 (top) and 1 (bottom) *b*-tag categories, compared with the SM background prediction.

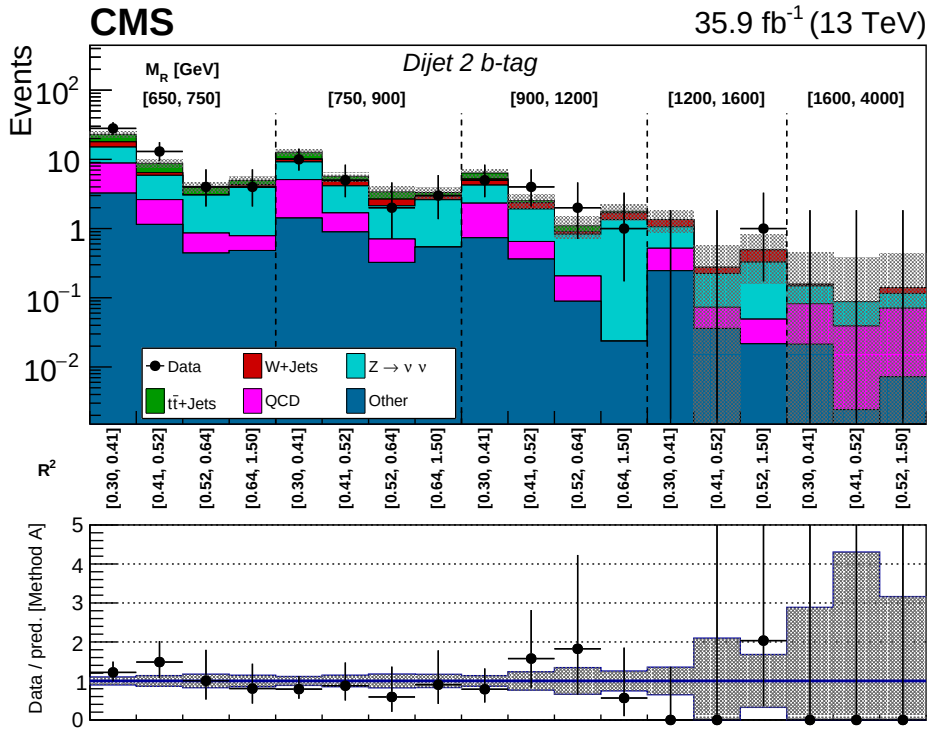


Figure 8.25: Observed data counts in each bin of the Dijet ≥2 *b*-tag category, compared with the SM background prediction.

are corrected by reweighting the MC events with appropriate scale factors.

- The distribution of the number of pileup vertices in the samples is very different from that in data. The standard pileup reweighting procedure would produce a highly nonuniform distribution of event weights, and it is not used for these samples. Instead, we estimate the change in signal acceptance as a function of pileup using a linear fit to the MC. The fit is used to estimate the number of signal events expected given the data pileup distribution. Uncertainty on the event yields from this procedure are propagated to the signal prediction.

- The fast simulation is observed to occasionally produce spurious jets, as discussed in Section 7.13. This causes E_T^{miss} to be mismeasured in a small population of events. Simulated events with jets that do not match generator-level objects, and that have less than 10% of their energy in charged hadrons, are rejected. This veto eliminates a long E_T^{miss} tail that is not present in samples produced with the full CMS simulation.

- We further protect against E_T^{miss} mismodeling by assigning a systematic un-

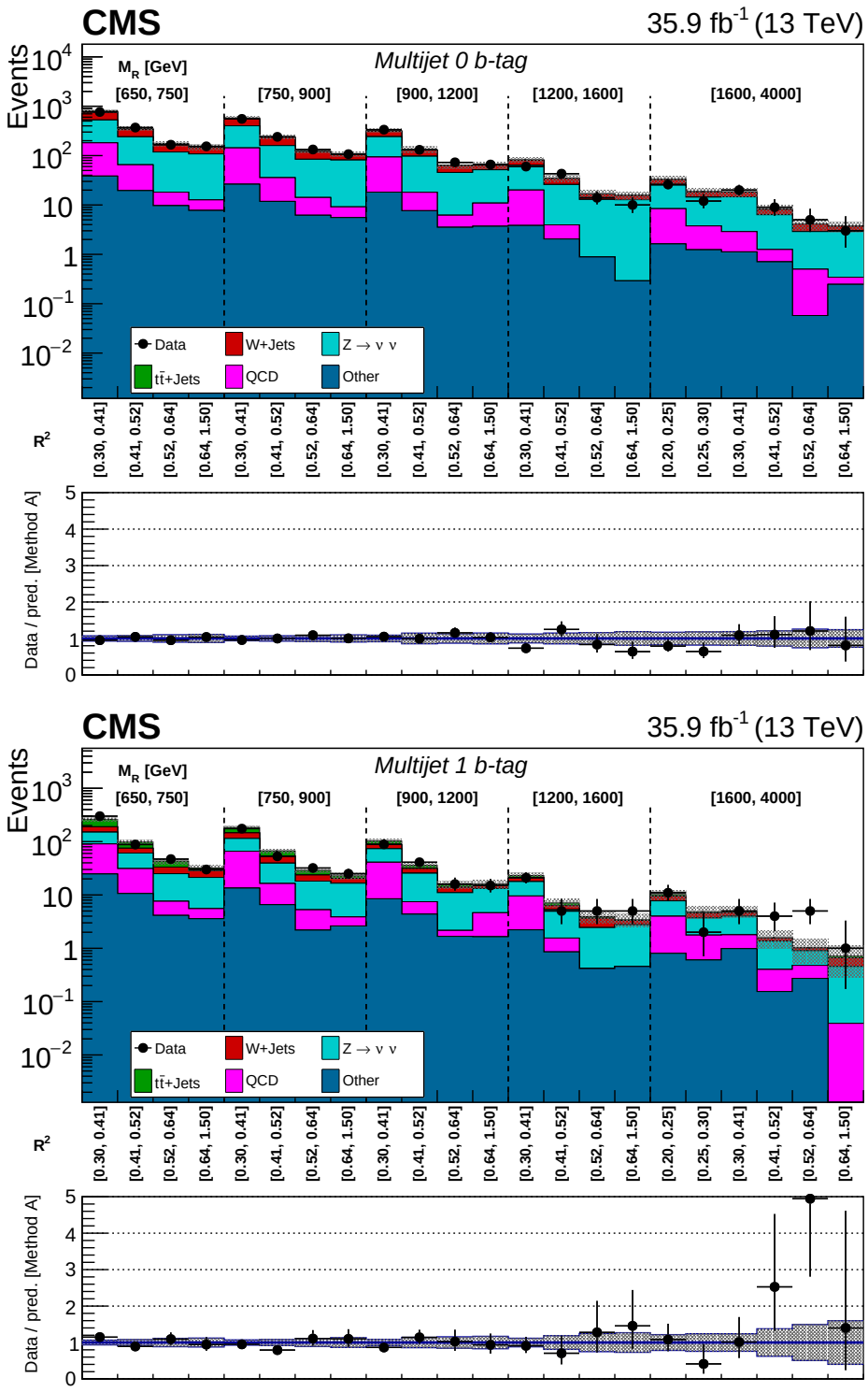


Figure 8.26: Observed data counts in each bin of the Multijet 0 (top) and 1 (bottom) *b*-tag categories, compared with the SM background prediction.

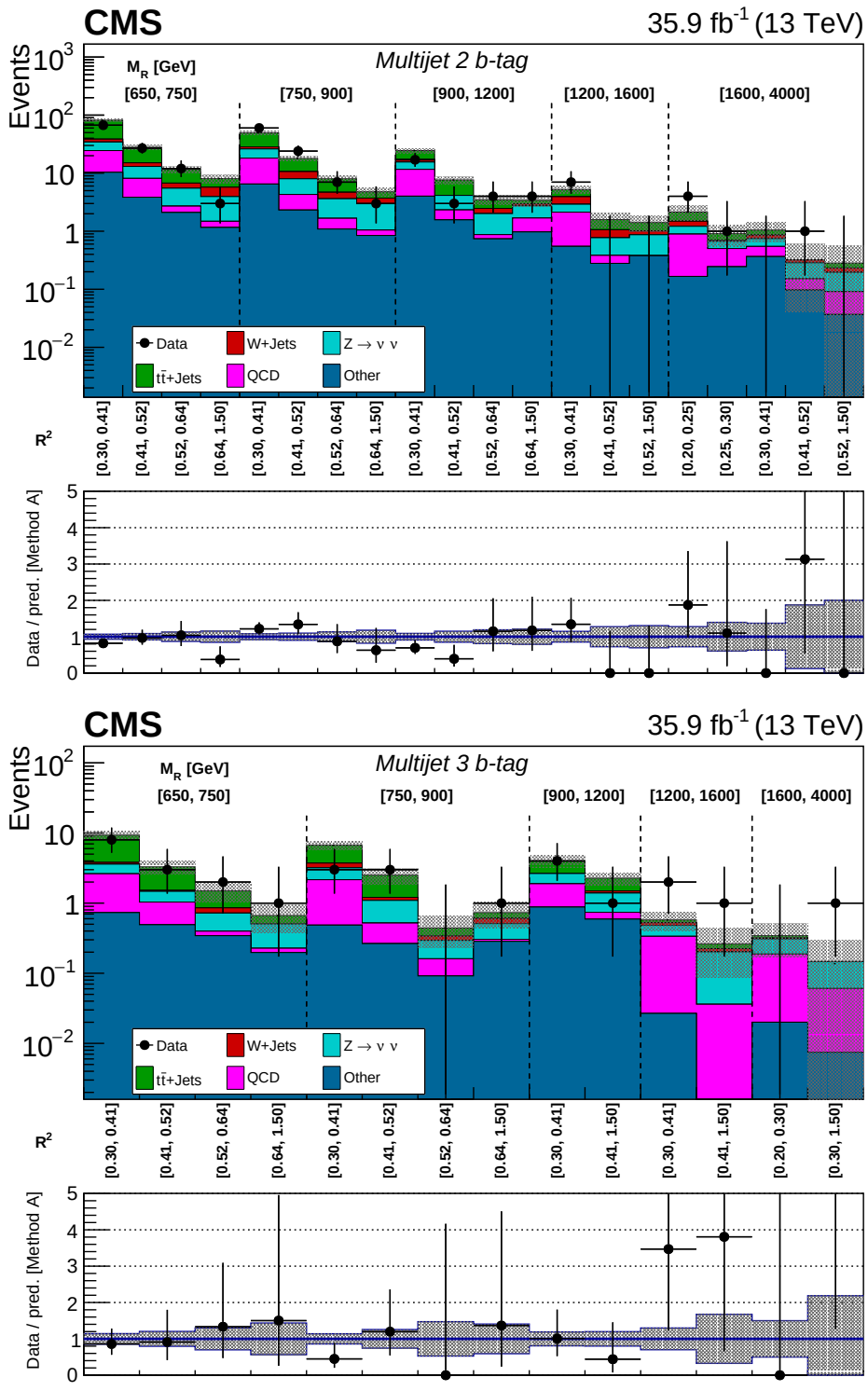


Figure 8.27: Observed data counts in each bin of the Multijet 2 (top) and ≥3 (bottom) *b*-tag categories, compared with the SM background prediction.

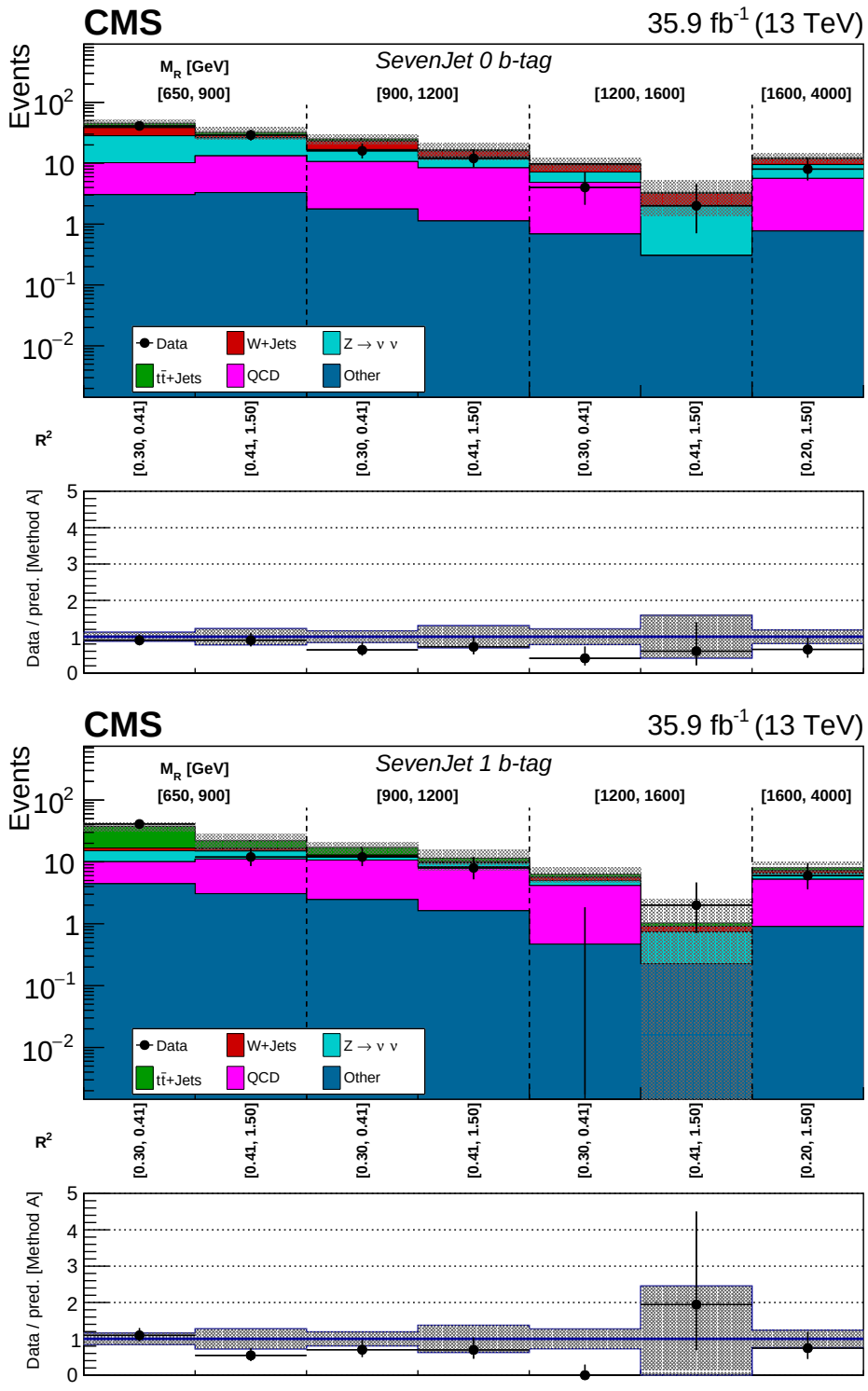


Figure 8.28: Observed data counts in each bin of the Seven-jet 0 (top) and 1 (bottom) *b*-tag categories, compared with the SM background prediction.

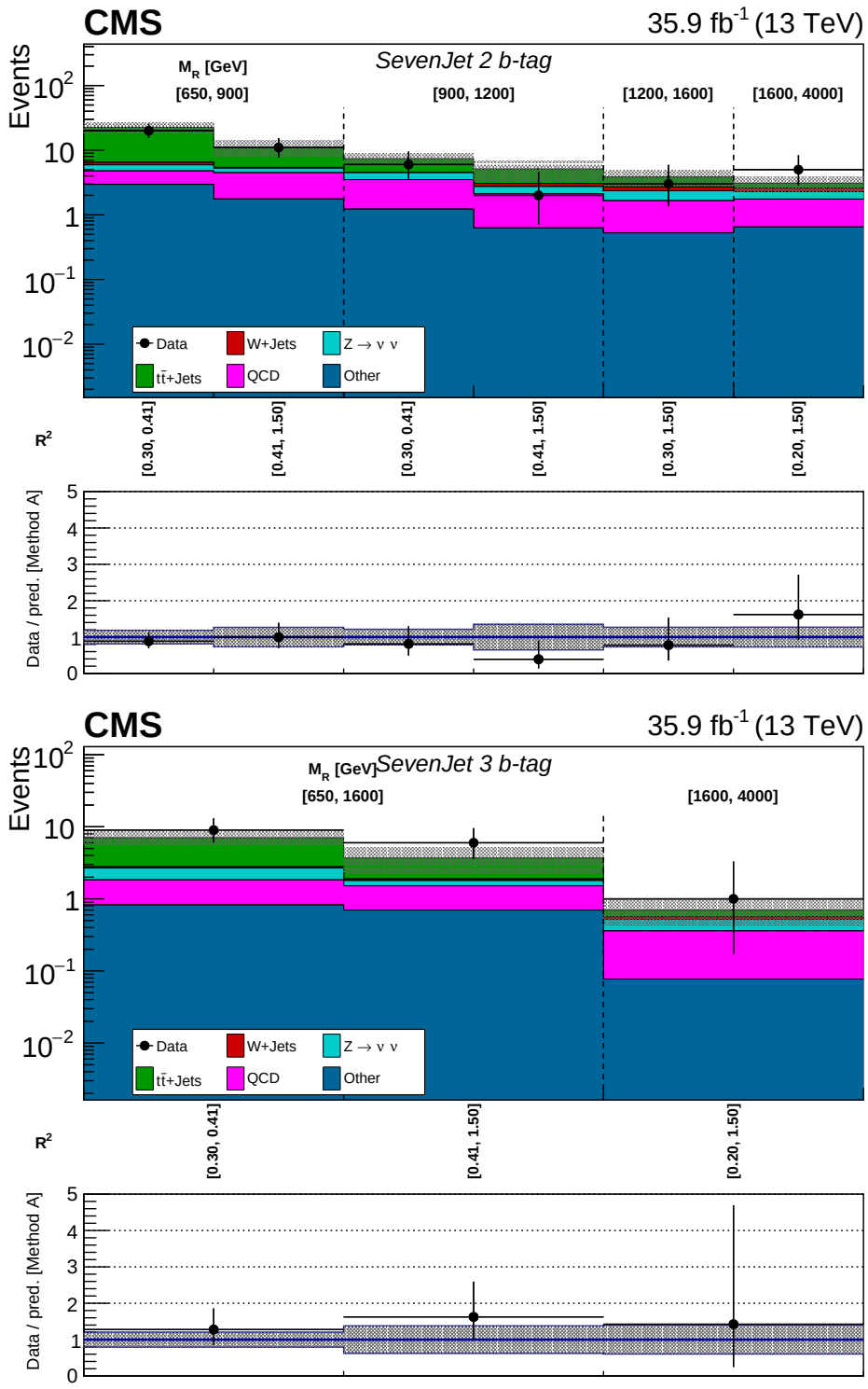


Figure 8.29: Observed data counts in each bin of the Seven-jet 2 (top) and ≥3 (bottom) *b*-tag categories, compared with the SM background prediction.

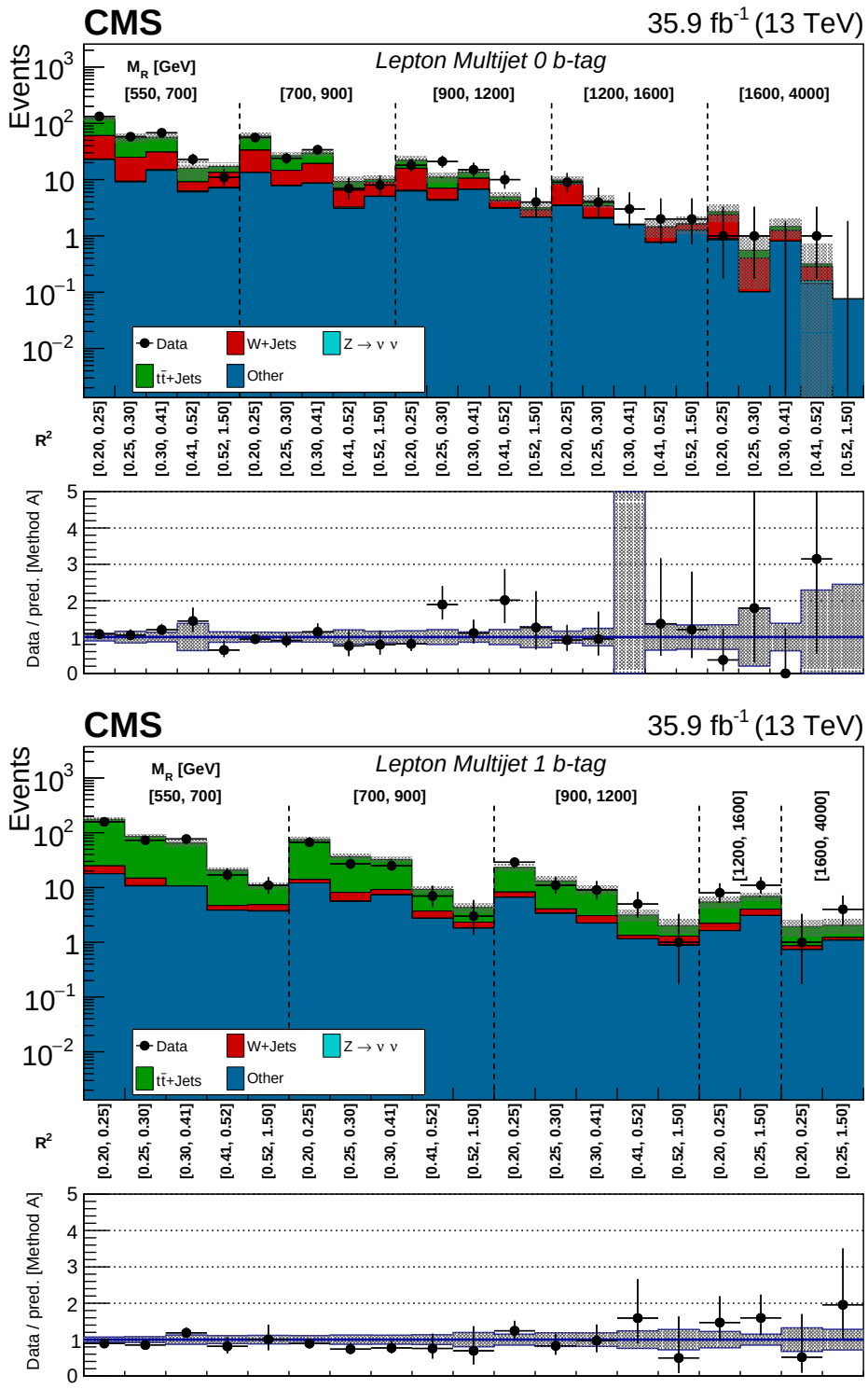


Figure 8.30: Observed data counts in each bin of the Lepton Multijet 0 (top) and 1 (bottom) *b*-tag categories, compared with the SM background prediction.

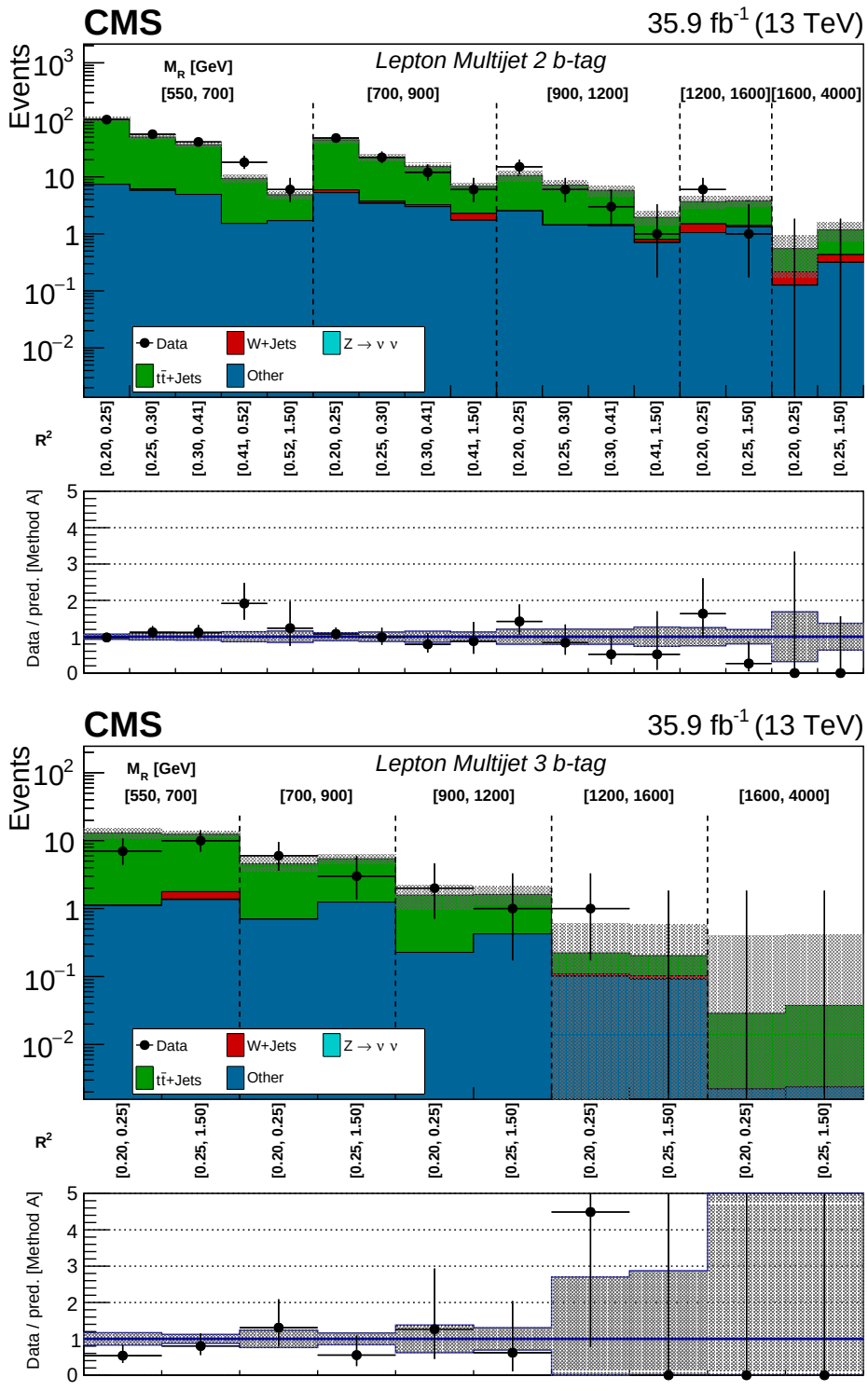


Figure 8.31: Observed data counts in each bin of the Lepton Multijet 2 (top) and ≥3 (bottom) *b*-tag categories, compared with the SM background prediction.

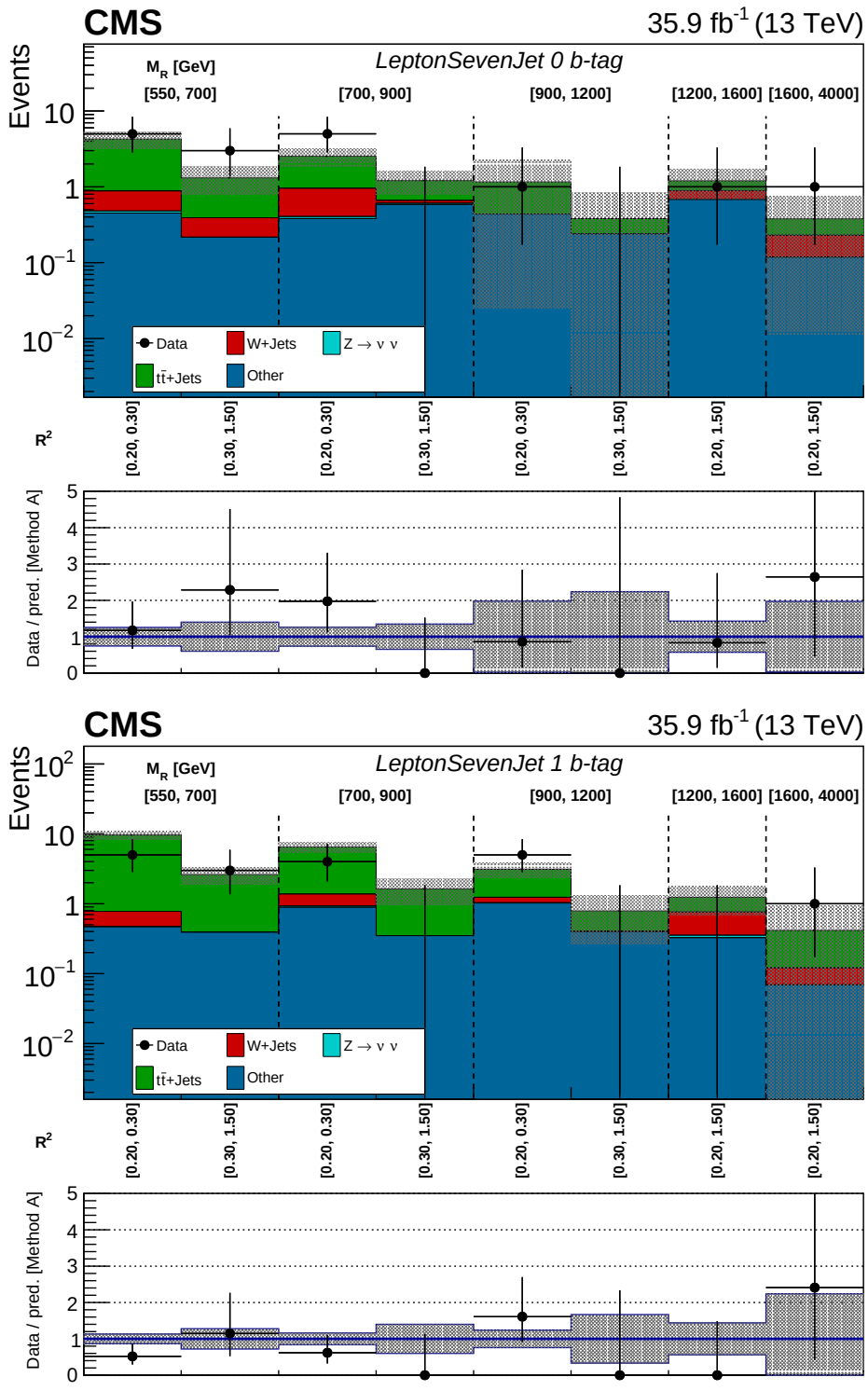


Figure 8.32: Observed data counts in each bin of the Lepton Seven-jet 0 (top) and 1 (bottom) *b*-tag categories, compared with the SM background prediction.

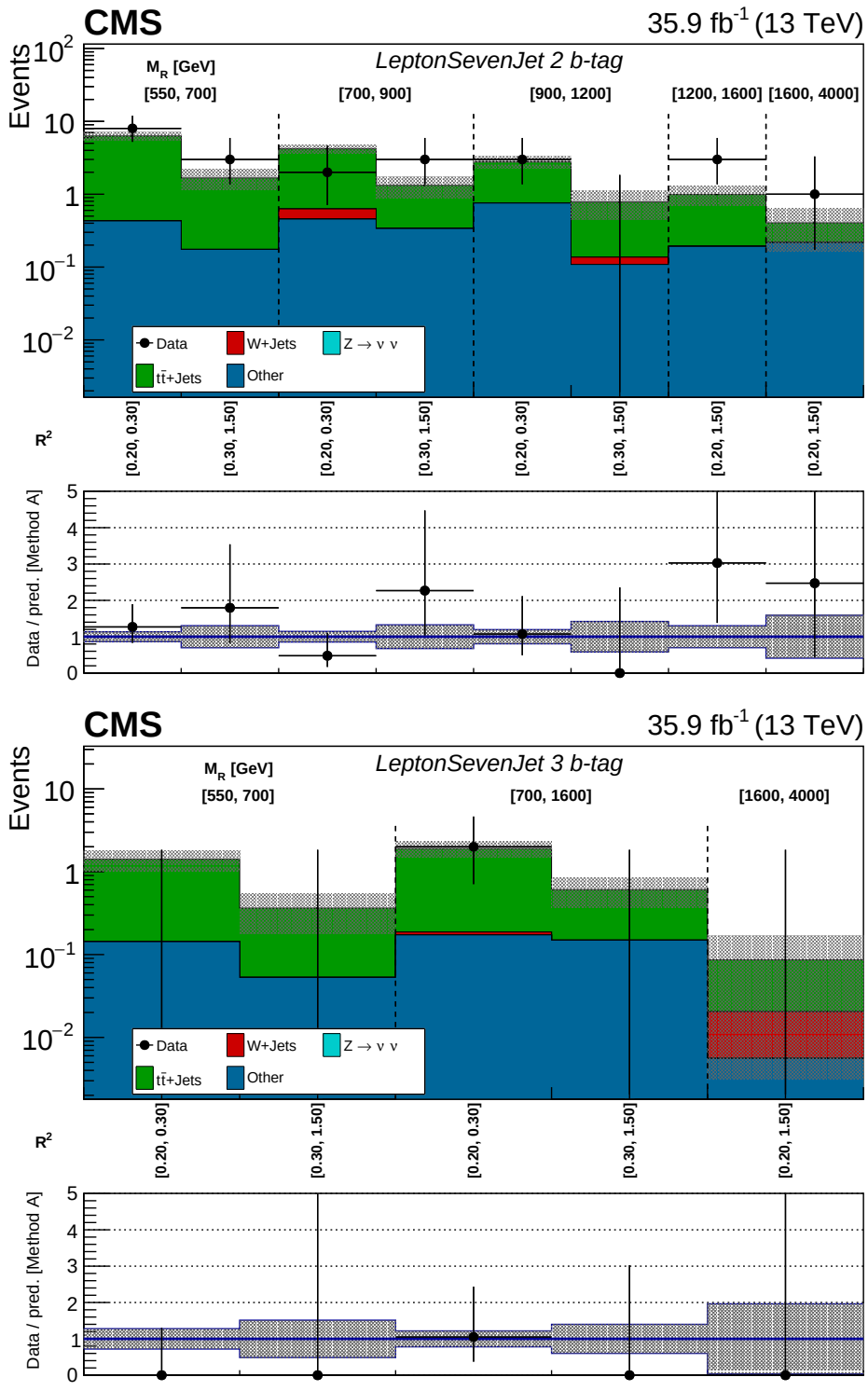


Figure 8.33: Observed data counts in each bin of the Lepton Seven-jet 2 (top) and ≥3 (bottom) *b*-tag categories, compared with the SM background prediction.

certainty on the signal prediction equal to the difference in event yields obtained using the generator-level $\vec{p}_T^{miss}$ vector instead of the reconstructed $\vec{p}_T^{miss}$ in the calculation of R^2.

8.12 Limits on SUSY simplified models

In Figures 8.34-8.37 we show the 95% CL exclusion limits on the simplified models T1tttt, T1bbbb, T1qqqq, T5ttcc, T2tt, T2bb, and T2qq. For the T2tt model, we blank out part of the 'top corridor' region due to the considerations discussed in the previous chapter. For the T2qq model, which features first- or second-generation squarks that decay to a quark and the LSP, we compute the limit both for the case of a single accessible squark state and for the case of eight mass-degenerate squarks (i.e., the SUSY partners of the right- and left-handed up, down, charm, and strange quarks).

Figure 8.38 shows the limits placed by other CMS SUSY searches on the simplified models we consider. We see that the limits placed by the combined razor search are competitive with or slightly better than the other searches for a number of SUSY simplified models.

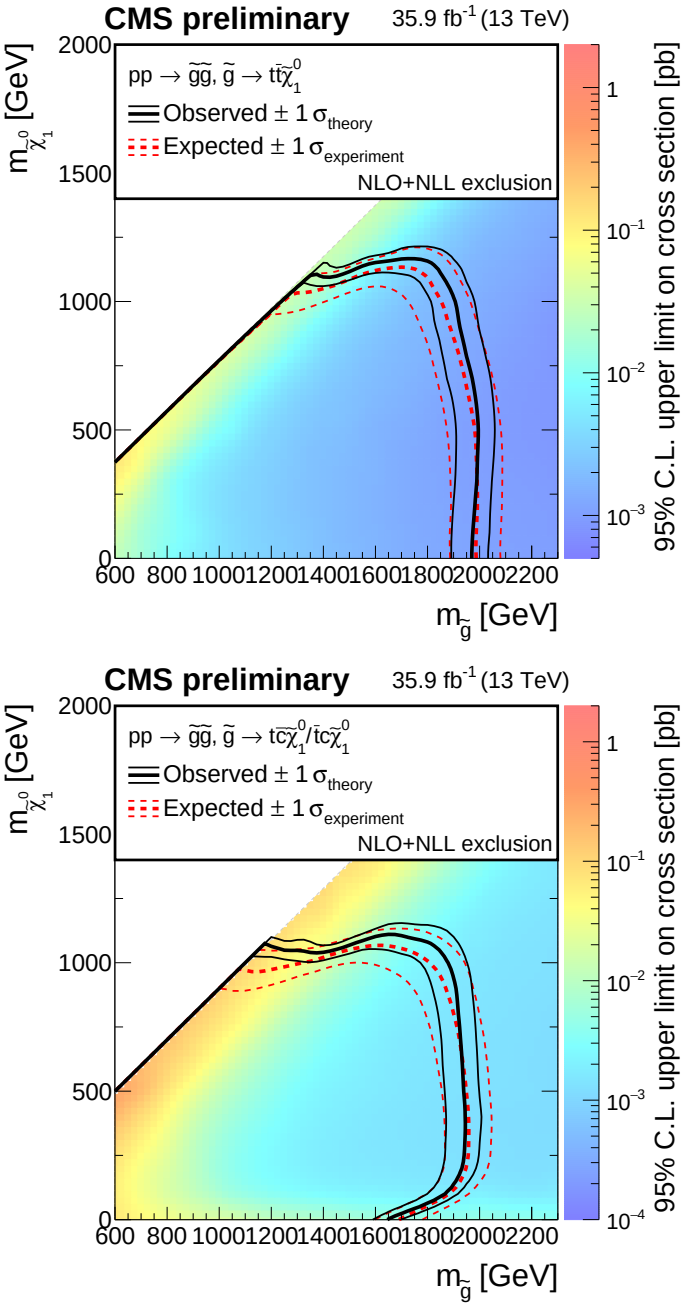


Figure 8.34: Expected and observed 95% CL upper limits on the T1tttt (top) and T5ttcc (bottom) simplified models of gluino production. The limits are obtained by combining the results of our search with those of the boosted razor search.

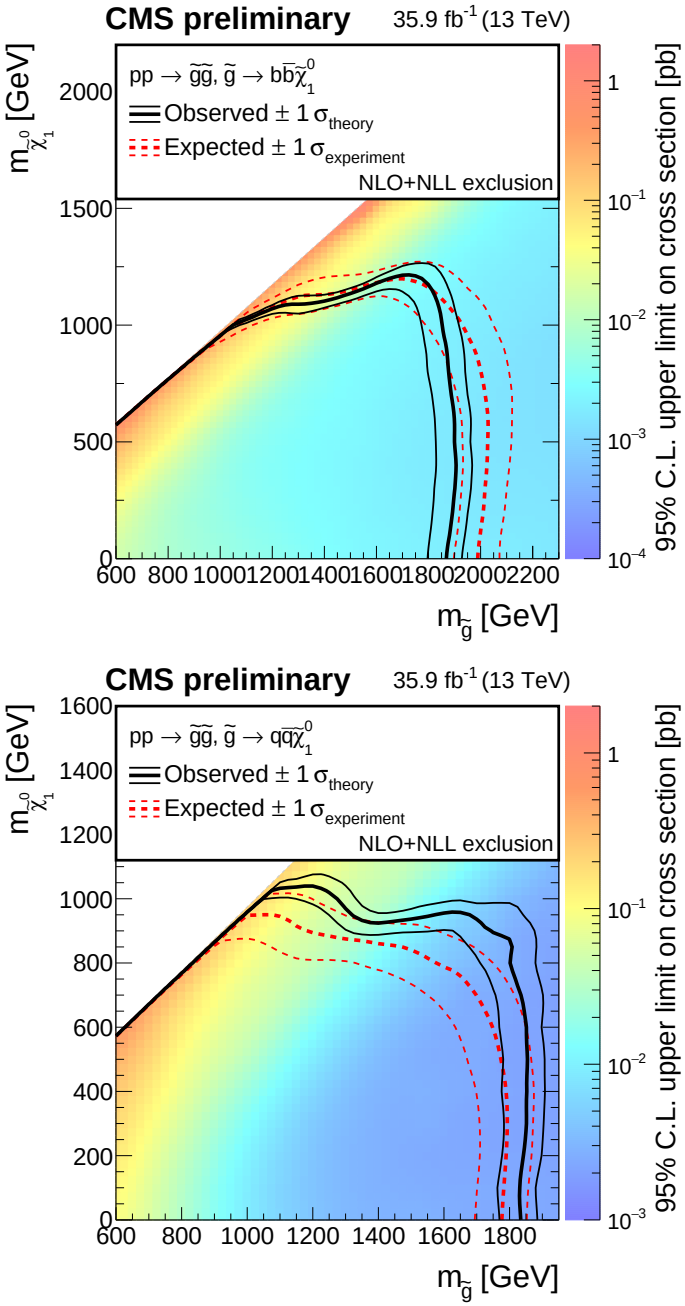


Figure 8.35: Expected and observed 95% CL upper limits on the T1bbbb (top) and T1qqqq (bottom) simplified models of gluino production.

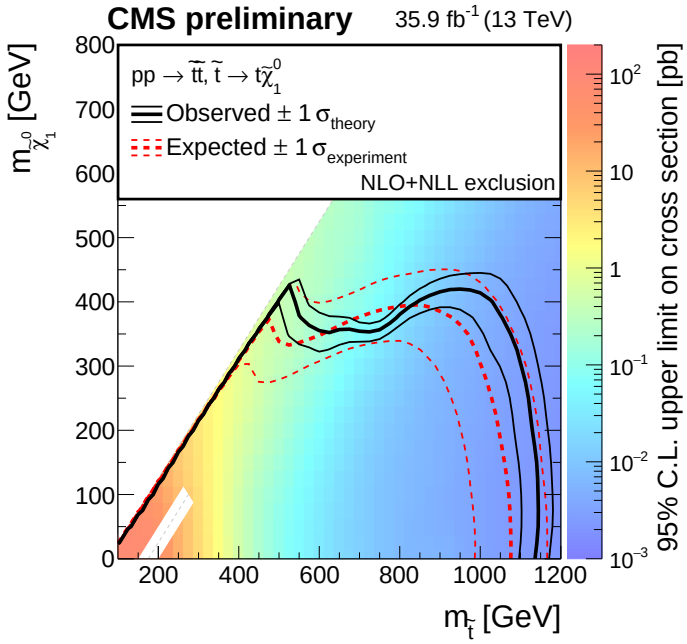


Figure 8.36: Expected and observed 95% CL upper limits on the T2tt simplified models of stop production, with part of the 'top corridor' region blanked out as discussed in Section 7.13. The limits are obtained by combining the results of our search with those of the boosted razor search.

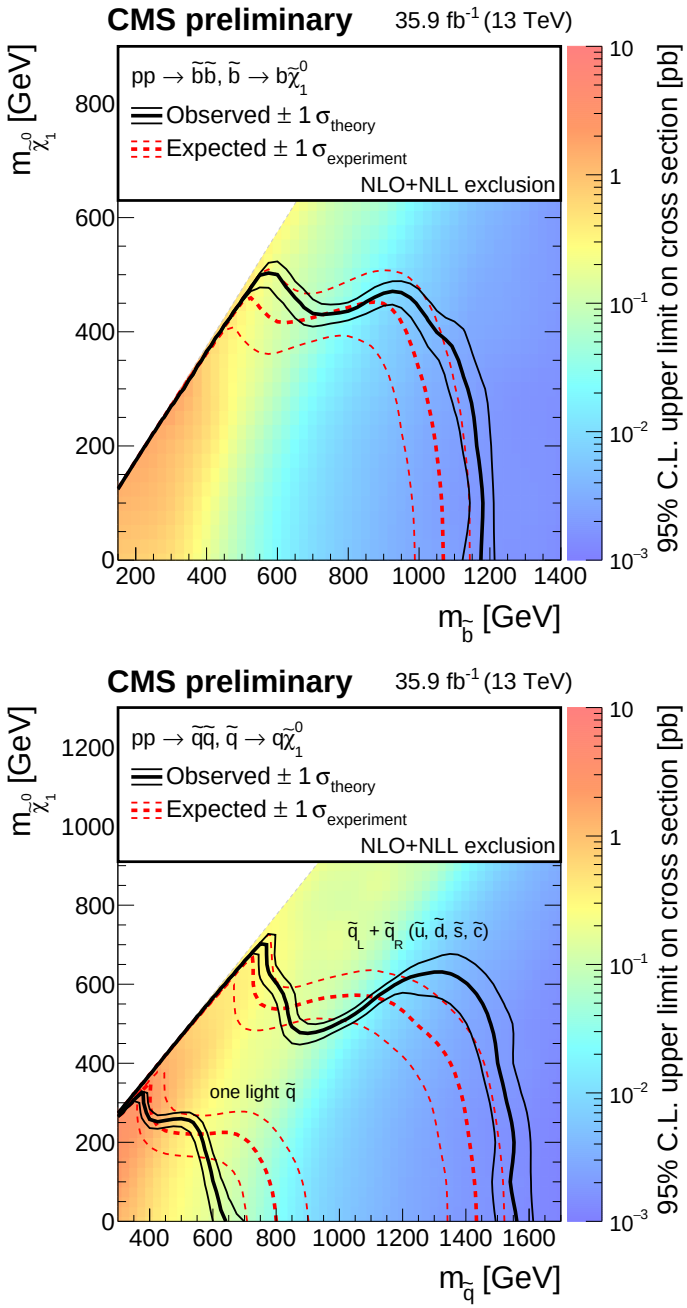


Figure 8.37: Expected and observed 95% CL upper limits on the T2bb (top) and T2qq (bottom) simplified models of squark production. The T2qq plot shows limits for the hypothesis of a single accessible squark state, and for that of eight mass-degenerate squarks.

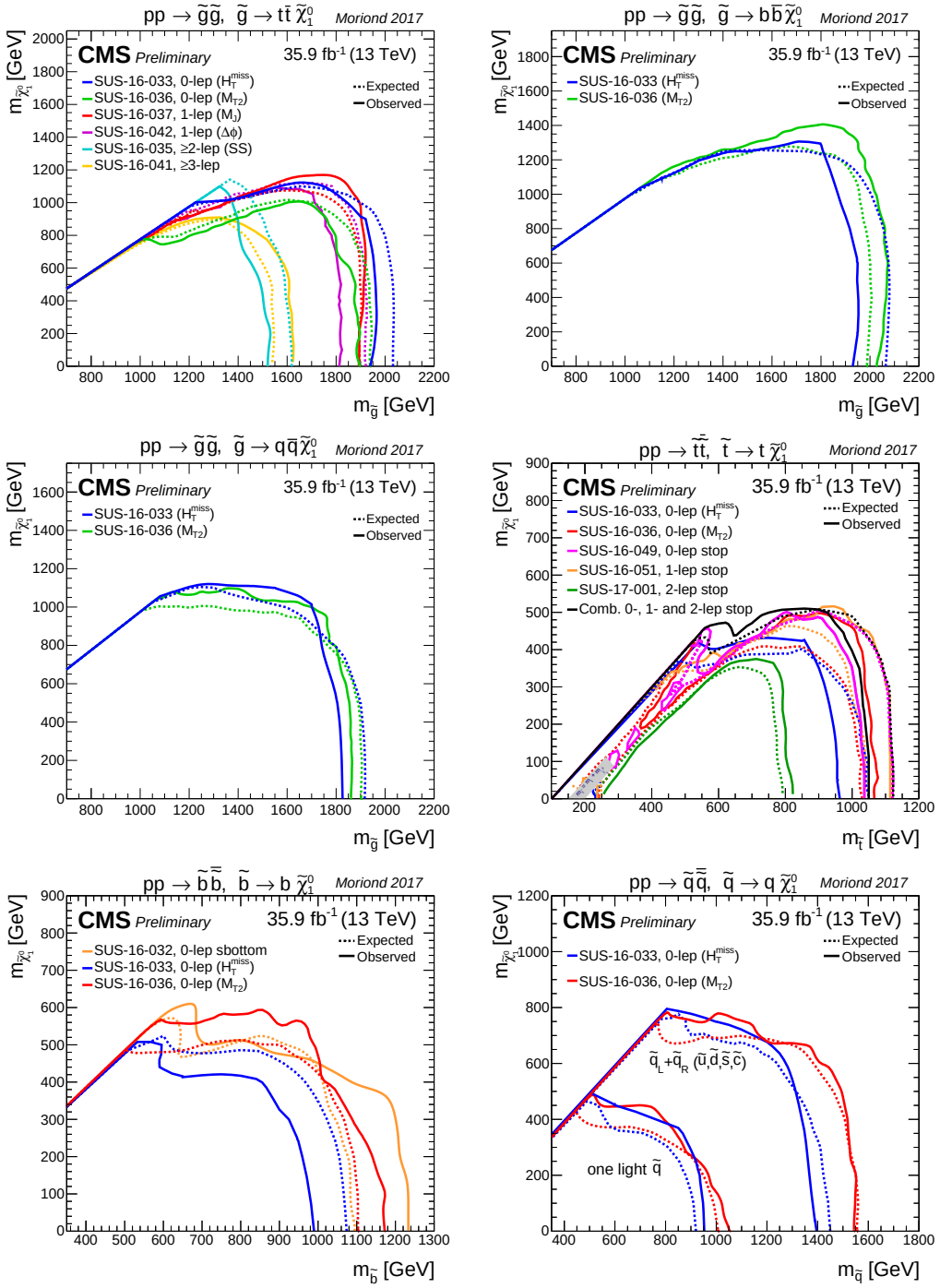


Figure 8.38: Summary of CMS SUSY exclusion limits on simplified models using the 2016 dataset. Top row: T1tttt and T1bbbb. Middle row: T1qqqq and T2tt. Bottom row: T2bb and T2qq [138].

Chapter 9

TOWARDS FUTURE FIT-BASED SEARCHES

The fit-based approach to background prediction for the razor SUSY search is of interest because its modeling assumptions are complementary to those of MC-based searches. However, using a fixed functional form for background estimation imposes a strong hypothesis on the shape of the background that is difficult to quantify using systematic uncertainties. In this chapter we explore the challenges of the fit-based approach in the context of the 2016 CMS dataset. After concluding that the existing analysis strategy is not viable on a dataset of this size, we present an alternative approach to the fit based on gaussian process regression that shows promise as a future direction for this type of search.

9.1 Reparameterizing the razor fit function

The functional form (Eq. 6.2) chosen to represent the background shape in the razor search suffers from nonlinear constraints on the four free parameters b, n, M_R^0, and R_0^2. We repeat the razor function expression here for convenience:

$$f(M_R, R^2) = \left(b\left[(M_R - M_R^0)(R^2 - R_0^2)\right]^{1/n} - 1\right) e^{-bn\left[(M_R - M_R^0)(R^2 - R_0^2)\right]^{1/n}}. \quad (9.1)$$

The requirements on the four parameters are:

- $b > 0$
- $n > 0$
- $M_R^0 < M_{R,\mathrm{min}}$
- $R_0^2 < R_{\mathrm{min}}^2$
- $b[(M_{R,\mathrm{min}} - M_R^0)(R_{\mathrm{min}}^2 - R_0^2)]^{1/n} > 1$

Here $M_{R,\mathrm{min}}$ and R_{min}^2 are the lower bounds of the fit range in the M_R and R^2 directions. If the above constraints are not satisfied, the function cannot be used as a probability density because it is negative or undefined for some values of M_R and R^2. Convergence issues when performing the fits on 2016 data are found to stem

from this problem. During minimization, the point (b, n, M_R^0, R_0^2) may enter a forbidden region of the parameter space. The minimization routine is unable to escape these regions, because the gradient and hessian of the likelihood cannot be computed properly there.

We address the issue by defining a new set of fit parameters:

- $A = M_{R,\text{min}} - M_R^0$
- $B = R^2_{\text{min}} - R_0^2$
- $P = b(AB)^{1/n} - 1$

The fit function can then be expressed in terms of A, B, P, and n as

$$f(X,Y) = \left((P+1) \left[\frac{(X+A)(Y+B)}{AB} \right]^{1/n} - 1 \right) e^{-n(P+1)\left[\frac{(X+A)(Y+B)}{AB}\right]^{1/n}}, \qquad (9.2)$$

with $X = M_R - M_{R,\text{min}}$ and $Y = R^2 - R^2_{\text{min}}$. The function (properly normalized) is a valid probability density for any choices of $A, B, P, n > 0$.

Reparameterizing the function in this way yields much more robust fits, because the minimization procedure is no longer able to push the parameters into the forbidden regions. The positivity constraint on the fit parameters is easily enforced in RooFit.

9.2 The challenge of increasing dataset size

We perform sideband fits with the reparameterized fit function in all of the 2-3 and 4-6 jet search categories. In each region the fit converges and returns a positive-definite covariance matrix for the fit parameters. However, the uncertainties on the fit prediction, which are obtained from the covariance matrix as described in Section 6.2, are too small to adequately describe the spread of the data. This can be seen in the $n\sigma$ distributions in Figure 9.1, which feature a large number of 2-3σ deviations of the data from the fit, even in the sideband region where the fit is performed. Example comparisons between fit-based and MC-based predictions on the 2016 dataset are shown in Figure 9.2. It is seen that the uncertainties on the fit-based prediction are much smaller than those from the MC-based prediction. A similar effect is seen when fitting MC simulated samples with both the sideband and the full fit procedures.

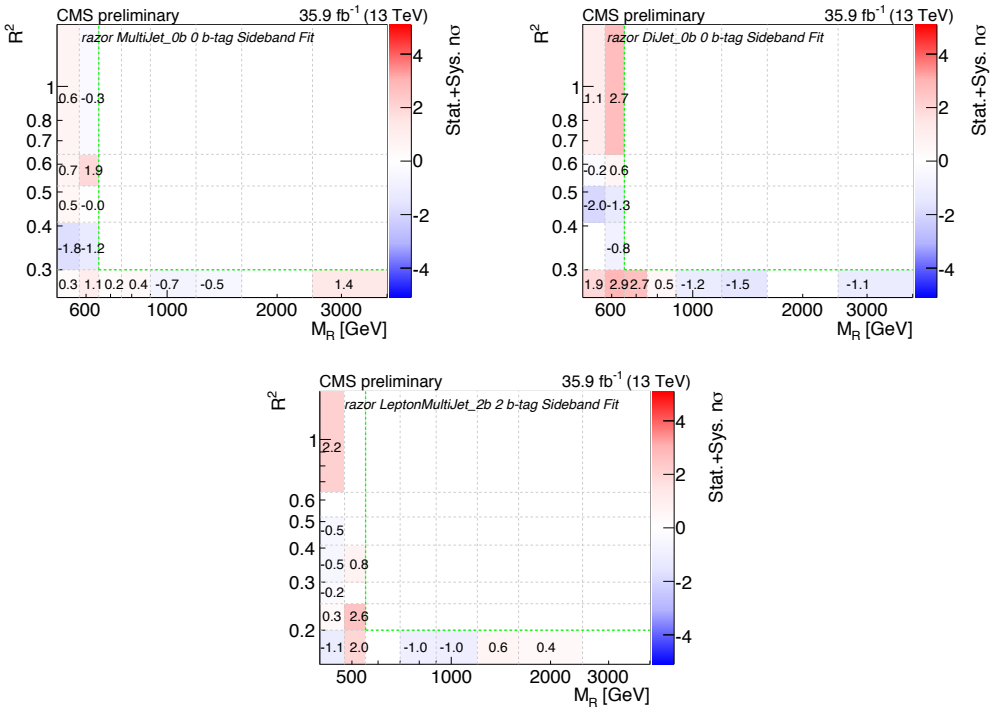


Figure 9.1: Example plots showing the deviations of the data from the razor fit in the sideband region. The Multijet 0 *b*-tag, Dijet 0 *b*-tag, and Lepton Multijet 2 *b*-tag regions are shown in the left, right, and bottom panels, respectively. Each bin is labeled with the $n\sigma$ difference between the data and the sideband fit. The green dotted line delineates the sideband region. Bins in the sideband with no indicated $n\sigma$ value have $|n\sigma| < 0.1$.

The large significance of the deviations reflects bias in the fit model with respect to the 'true' shape that would be obtained in the limit of infinite data. The fit function is chosen based on empirical considerations; it is not expected that any choice of function with a fixed number of parameters will continue to describe the data well as the size of the dataset grows. The uncertainties shown in Figure 9.1 reflect uncertainty on the values of the fit parameters *within the model* but do not express anything about the validity of the model itself.

One could make the fit model more expressive by adding parameters to the functional form. This would decrease the bias of the model, allowing it to capture a wider variety of shapes and increasing the uncertainty on the predictions. However, it is not clear which generalization(s) of the fit function would allow it to better capture the shape in data. Experiments were performed with a number of different functional forms, but no function was seen to adequately describe the data in all analysis categories.

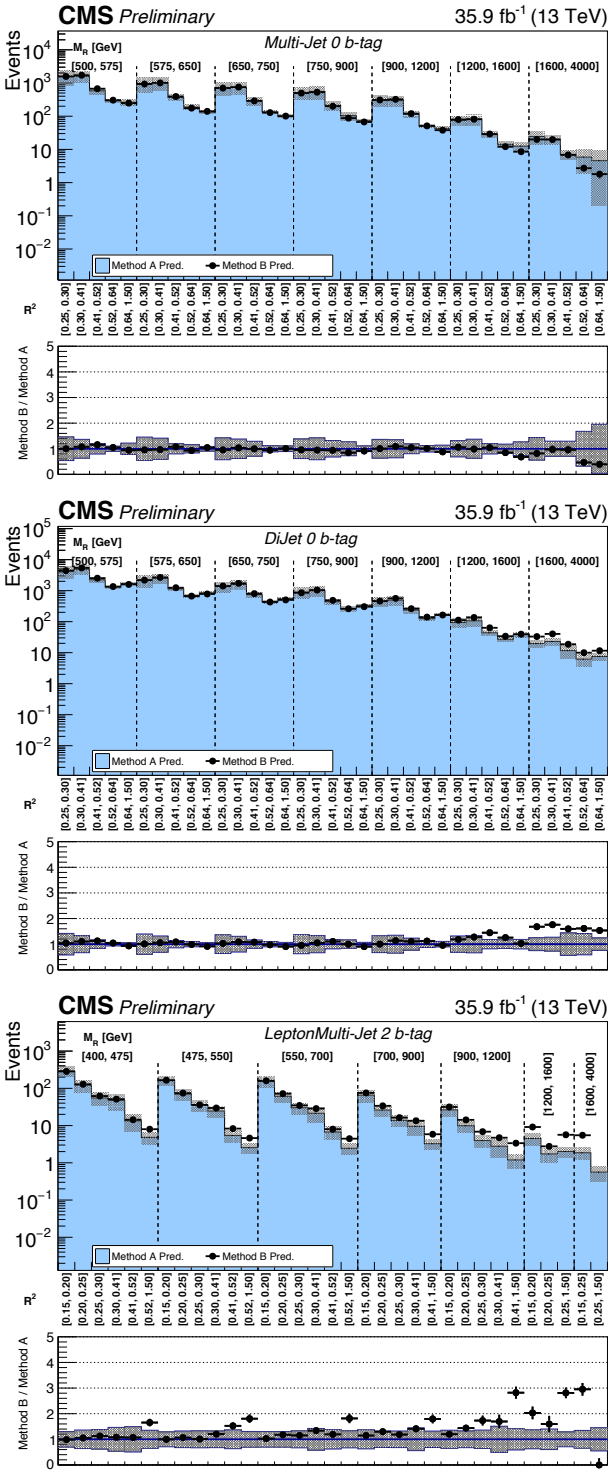


Figure 9.2: Example plots comparing the fit-based and MC-based background predictions in the Multijet 0 *b*-tag (top), Dijet 0 *b*-tag (middle), and Lepton Multijet 2 *b*-tag (bottom) regions.

An additional concern is that the existing fit shape does not describe the data well in the new search regions that have events with 7 or more jets. The M_R distribution in these regions peaks at a higher value than that in other search categories. This M_R peak, and the very low statistics of these event samples, suggest that dedicated fit strategies would have to be developed for these analysis regions.

9.3 Nonparameteric background modeling using gaussian processes

As a future direction for fit-based SUSY searches, we propose and test a background modeling strategy based on the technique of gaussian process (GP) regression [139]. A GP is a stochastic process that provides a prior distribution over all possible background function shapes. It is an infinite dimensional prior: any background shape is theoretically allowed, but some shapes (usually smoothly varying shapes) are preferred over others. The observed data constrain the set of viable fit functions and yield a posterior distribution that quantifies the uncertainty in the fit at each point. Advantages of the GP approach over using a fixed functional family include:

- It is nonparametric and does not require or assume that the true background shape is analytically tractable.
- It handles both small and large dataset sizes without overfitting and without the need to tune the number of model parameters.
- Known information about the covariance structure of the data can be encoded into the model, leading to flexible yet 'physics-aware' background models [140].

GP regression is advocated in [140] as an alternative to the use of 'ad-hoc' background functions like Eq. 9.1. The method is demonstrated using data from an ATLAS dijet resonance search; we will see that fitting in a razor search context requires some modifications to the procedure described there. A similar nonparametric procedure for background modeling is described in [141] in the context of the search for the Higgs particle. GPs have been used as a component of a top quark mass measurement in CMS [142], but otherwise they have not seen much use at the LHC.

Problem setup and data

We consider a dataset $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^T$ of N events, where each event is represented by a vector $\mathbf{x}$ of high-level observables. We focus on the one-dimensional case in which $\mathbf{x}_i$ is the value of M_R in the ith event. The two-dimensional case in which $\mathbf{x}_i$ is an (M_R, R^2) pair is not conceptually different. We take the data to be binned in a fixed, regular grid of B non-overlapping regions. Let $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_B)^T$ be the locations of the bin centers and $\mathbf{y} = (y_1, \ldots, y_B)^T$ be the number of observed data counts in each bin.

Our goal is to model the data with a smooth function $f(\mathbf{x})$, treating each binned observation y_i as a noisy observation of the function value $f(\mathbf{u}_i)$. We use a subscript notation to denote function evaluation at a set of locations, so, e.g., $f_\mathbf{U}$ is shorthand for the vector $(f(\mathbf{u}_1), \ldots f(\mathbf{u}_B))$. The probability distribution $p(\mathbf{y} \mid f_\mathbf{U})$ (referred to as the likelihood or conditional likelihood), which expresses the random jitter of the data around the background function, is assumed known.

For this study we take all object definitions and selection criteria from the 2016 inclusive razor SUSY search. We model the SM background using semileptonic $t\bar{t}$+jets MC generated with Powheg. The background MC events are taken as-is, with no reweighting to correct the pileup distributions, object selection efficiencies, etc. Events are partitioned into boxes according to the inclusive razor scheme.

Gaussian process regression

A GP is a probability distribution over functions $f(\mathbf{x})$, with the property that for any fixed collection of locations $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_M)^T$, the marginal distribution $p(f_\mathbf{Z})$ of the function values at those locations is Gaussian [139]:

$$p(f_\mathbf{Z}) = \text{Normal}(f_\mathbf{Z}\,;\, \mu_\mathbf{Z}, K_{\mathbf{ZZ}}). \tag{9.3}$$

The functions μ and K, which give the mean and covariance of the Gaussian, are called the *mean function* and *covariance function*, respectively. The mean function expresses prior belief over the shape of f, and is typically taken to be zero. The GP's behavior is determined mainly by the choice of covariance function K, which expresses the degree of correlation between function values f at different locations. We write $K_{\mathbf{ZZ}}$ to denote the matrix whose ijth element is $K(\mathbf{z}_i, \mathbf{z}_j)$. The function K can be freely chosen as long as it yields a positive semidefinite covariance matrix for any set of input locations.

The GP can be thought of as an infinite dimensional generalization of the Gaussian distribution: it is an infinite collection of random variables f (indexed by the continuous location variable $\mathbf{x}$) such that any finite subset is distributed as a Gaussian. 'Nearness' of two points $\mathbf{x}_1$ and $\mathbf{x}_2$ is defined by the covariance function. For example, the commonly used squared exponential covariance function,

$$K_{\mathrm{SE}}(\mathbf{x}_1, \mathbf{x}_2) = \alpha e^{-|\mathbf{x}_1 - \mathbf{x}_2|^2 / 2\ell^2}, \tag{9.4}$$

specifies that two function values $f(\mathbf{x}_1)$ and $f(\mathbf{x}_2)$ are highly correlated if $\mathbf{x}_1$ and $\mathbf{x}_2$ are close together. A GP with this covariance favors functions f that are spatially smooth. The characteristic length ℓ and the base correlation strength α are hyperparameters that can take any positive value. The hyperparameters can be treated as fixed or allowed to vary according to a specified prior distribution.

GP regression is usually formulated as a Bayesian inference procedure, with model predictions made using the posterior distribution over background function shapes. In this study we follow the Bayesian approach; however, GPs can be used equally well in a frequentist setting, treating the GP prior $p(f_{\mathbf{U}})$ as a constraint term in the likelihood.

In the Bayesian case, we need to obtain or approximate the posterior distribution over the background function values given the data:

$$p(f_{\mathbf{U}} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid f_{\mathbf{U}}) p(f_{\mathbf{U}})}{p(\mathbf{y})}. \tag{9.5}$$

The numerator in Eq. 9.5 is the product of the likelihood and the prior distribution. The denominator is the marginal likelihood of the observed data $\mathbf{y}$, integrated over all model parameters; it is constant with respect to f.

Gaussian likelihood case

We first consider the case of a Gaussian conditional likelihood,

$$p(\mathbf{y} \mid f_{\mathbf{U}}) = \prod_{i=1}^{B} \mathrm{Normal}(y_i \,;\, f(\mathbf{u}_i), \sigma^2(\mathbf{u}_i)), \tag{9.6}$$

where $\sigma^2(\mathbf{u}_i)$ is the variance of the Gaussian at each location $\mathbf{u}_i$. In this model, the observed bin counts y_i are considered to be random Gaussian smearings of the true function values $f(\mathbf{u}_i)$. We follow [140] and impose Poisson-like noise based on the observed bin counts: $\sigma^2(\mathbf{u}_i) = y_i$.

The Gaussian conditional likelihood is appropriate when the number of data events in each bin is large, so that the Poisson distribution of bin counts can be well approximated by a Gaussian. The reason to make this approximation is that it allows inference for f to be performed analytically. In particular, the product of the GP prior and the Gaussian likelihood in Eq. 9.5 is again a Gaussian. The marginal likelihood $p(\mathbf{y})$ can be computed in closed form:

$$\log p(\mathbf{y}) = -\frac{1}{2}\log\det\Sigma_{\mathbf{UU}} - \frac{1}{2}(\mathbf{y}-\mu_{\mathbf{U}})^T\Sigma_{\mathbf{UU}}^{-1}(\mathbf{y}-\mu_{\mathbf{U}}) - \frac{B}{2}\log 2\pi, \tag{9.7}$$

where $\Sigma_{\mathbf{UU}} = K_{\mathbf{UU}} + \sigma_{\mathbf{U}}^2$ is the covariance matrix with the Gaussian noise folded in.

The marginal likelihood $p(\mathbf{y})$ quantifies the *evidence* for the GP model given the observed data. Maximizing the evidence as a function of the covariance kernel hyperparameters (ℓ and α in the case of the squared exponential covariance) yields an estimate of the most appropriate choice of parameters. Thus, in the Gaussian likelihood case the covariance hyperparameters can be obtained through straightforward numerical optimization of Eq. 9.7.

Predictions and uncertainties for new data points are also available analytically. The mean $\mu_{\mathbf{V}\,|\,\mathbf{y}}$ and variance $\Sigma_{\mathbf{VV}\,|\,\mathbf{y}}$ of the data at new locations $\mathbf{V}$ are given by

$$\mu_{\mathbf{V}\,|\,\mathbf{y}} = \mu_{\mathbf{V}} + K_{\mathbf{VU}}\Sigma_{\mathbf{UU}}^{-1}(\mathbf{y}-\mu_{\mathbf{U}}), \tag{9.8}$$

$$\Sigma_{\mathbf{VV}\,|\,\mathbf{y}} = \Sigma_{\mathbf{VV}} - K_{\mathbf{VU}}\Sigma_{\mathbf{UU}}^{-1}K_{\mathbf{UV}}. \tag{9.9}$$

Because of the matrix inversion, the complexity of these calculations scales as the cube of the number of observed bins, B.

Performance on M_R data

We implement the Gaussian likelihood GP model using the PyTorch numerical computation library, which provides efficient matrix operations and automated derivative calculations. To obtain stable performance for matrix inversion and determinant computations, we make use of the Cholesky decomposition of the covariance matrix:

$$\Sigma = AA^T, \tag{9.10}$$

with A a lower triangular matrix with positive diagonal entries. We compute the log determinant of Σ as twice the sum of the logarithms of A's diagonal elements. The matrix products in Eqs. 9.7-9.9 are evaluated without directly evaluating $\Sigma_{\mathbf{UU}}^{-1}$;

instead, we decompose $\Sigma_{\mathbf{UU}}^{-1}$ as $(A^T)^{-1}A^{-1}$, and then evaluate the left and right halves of the matrix product (e.g. $K_{\mathbf{VU}}(A^T)^{-1}$ and $A^{-1}K_{\mathbf{UV}}$, for the second term in Eq. 9.9) separately using a linear algebra solver before multiplying them together. This improves numerical stability considerably compared with direct inversion of the covariance matrix.

We first test the GP fit model on the Multijet 1 b-tag category, the analysis region with the largest number of $t\bar{t}$+jets MC events. We divide the range $650 < M_R < 1200$ GeV into 50 equal-size bins and fit the binned data with a GP with a squared exponential covariance function. The covariance hyperparameters are optimized using the L-BFGS minimization algorithm [143] provided with PyTorch. Eqs. 9.8 and 9.9 are used to compute and visualize the fitted function values at each location. We also compute the $n\sigma$ significance of the data-fit residual in each bin. To evaluate the method on a region with low event counts, we perform the same procedure using the Dijet 2 b-tag category, using 30 equal-size bins.

The results for the two analysis categories are shown in Figures 9.3 and 9.4. Excellent results are obtained with the GP fit in the Multijet 1 b-tag category; the distribution of $n\sigma$ values is roughly normal and the GP posterior describes the data well. On the other hand, the GP results are very poor in the Dijet 2 b-tag category, particularly in the lowest-count bins.

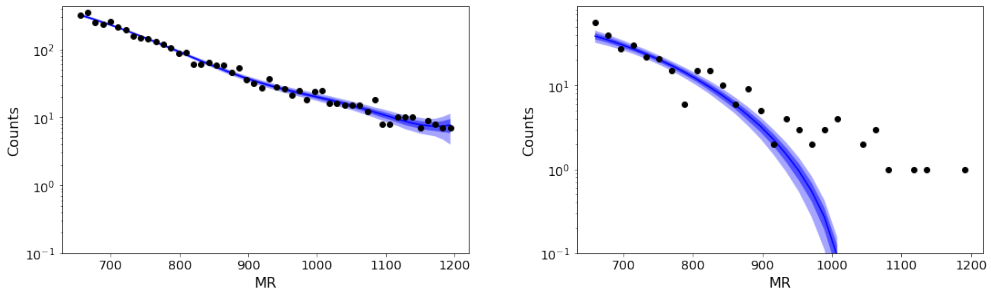


Figure 9.3: Simulated $t\bar{t}$+jets data (black points) and GP fit (blue line) using a Gaussian conditional likelihood in the Multijet 1 b-tag (left) and Dijet 2 b-tag (right) categories. The light and dark bands around the fit function denote 1- and 2-sigma uncertainties on the fitted function values. The failure of the fit in the Dijet 2 b-tag category illustrates that the Gaussian conditional likelihood is not appropriate in the case of low bin counts.

The failure of the GP fit in the low-statistics category is primarily due to the breakdown of the Gaussian approximation to the binned Poisson conditional likelihood. In particular, bins with zero counts (which have $\sigma^2 = 0$) cause the fit to be unstable.

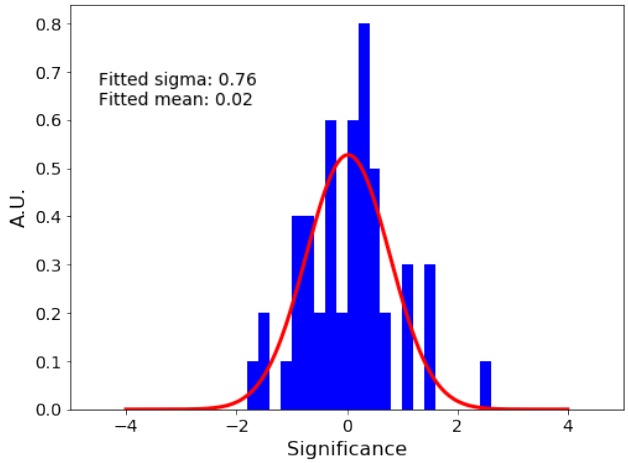


Figure 9.4: Distribution of $n\sigma$ significances in the fitted bins in the Multijet 1 b-tag fit shown in Figure 9.3. The red curve shows a Gaussian fit to the histogram.

Poisson likelihood case

Background modeling in low-count bins can be considerably improved using a GP with a Poisson conditional likelihood:

$$p(\mathbf{y} \mid f_{\mathbf{U}}) = \prod_{i=1}^{B} \text{Poisson}(y_i \,;\, e^{f(\mathbf{u}_i)}). \tag{9.11}$$

The GP prior $p(f_{\mathbf{U}})$ has the same form as above, but now the function f describes the logarithm of the Poisson mean in each analysis bin. For the case of binned data, the Poisson likelihood naturally describes the distribution of observed data counts.

Unfortunately, inference in the Poisson GP model is not analytically tractable, and we must resort to approximate methods. Continuing with the Bayesian approach used above, we focus on modeling the posterior distribution in Eq. 9.5. Two possible strategies for inference in this model are:

1. To use a Monte Carlo algorithm to obtain samples from the posterior, and use the samples to compute quantities of interest (see e.g. [144]).

2. To approximate the posterior distribution as Gaussian, and fit the model by minimizing the Kullback-Leibler divergence between the approximate and true posteriors (this is known as the *variational* approach; see e.g. [145]),

In the following subsections we demonstrate each of these methods on the M_R data and describe their advantages and drawbacks in different use cases.

Inference using Hamiltonian Monte Carlo

Sampling in Bayesian inference problems is often done using Markov Chain Monte Carlo (MCMC) algorithms, which are biased random walks that asymptotically follow the target probability distribution. For this problem we use an MCMC algorithm called Hamiltonian Monte Carlo (HMC) [146] that is suitable for use with highly correlated random variables. HMC is inspired by Hamiltonian particle dynamics: it simulates the evolution of particle trajectories in a system with potential and kinetic energies defined by

$$U(\theta) = -\log p(\theta \,|\, \mathbf{y}), \tag{9.12}$$

$$K(\phi) = \phi^T M^{-1} \phi, \tag{9.13}$$

where θ is a vector of model parameters that we wish to sample, and ϕ is a vector of auxiliary 'momentum' variables conjugate to θ. The mass matrix M is arbitrary, and is taken here to be the identity matrix.

The HMC algorithm proceeds by repeatedly performing the following steps:

1. Sample a new momentum vector ϕ from its probability distribution, $p(\phi) = \text{Normal}(\phi\,;\, 0, M)$.

2. Evolve the values of θ and ϕ forward in time using the following discrete version of Hamilton's equations:

$$\phi \rightarrow \phi - \frac{\epsilon}{2} \frac{\partial U}{\partial \theta}, \tag{9.14}$$

$$\theta \rightarrow \theta + \epsilon M^{-1} \phi, \tag{9.15}$$

$$\phi \rightarrow \phi - \frac{\epsilon}{2} \frac{\partial U}{\partial \theta}, \tag{9.16}$$

   where ϵ is a small step size parameter. One iteration of Eqs. 9.14-9.16 is referred to as a *leapfrog step*. Several leapfrog steps are carried out in sequence; the number of leapfrog steps per HMC iteration is given by a free parameter L.

3. Compute the quantity

$$r = \min\left(1, e^{[U(\theta_i)+K(\phi_i)]-[U(\theta_f)+K(\phi_f)]}\right), \tag{9.17}$$

   where the subscripts i and f denote the values of the variables before and after the application of Step 2, respectively.

4. Accept the HMC update with probability r. If the update is accepted, keep θ at its final value θ_f. Otherwise, reset θ to its old value θ_i. After doing this, output the current value of θ as a sample from $p(\theta \mid \mathbf{y})$.

This procedure generates a sequence of values $[\theta_1, \ldots, \theta_S]$, where S is the number of steps for which the algorithm is run. The algorithm is governed by two free parameters: the step size ϵ and the number of leapfrog steps L. We follow [147] and choose the value of L at each HMC step by sampling an integer uniformly at random between 1 and a maximum $L_{\max}$.

To fit the Poisson GP model, we must find the posterior distributions of the function values $f_{\mathbf{U}}$ and, optionally, the kernel hyperparameters ℓ and α. Inference is easier if we change to the 'whitened' coordinates $g_{\mathbf{U}}$, defined by

$$f_{\mathbf{U}} = A g_{\mathbf{U}}, \tag{9.18}$$

where A is defined as in Eq. 9.10. These coordinates $g_{\mathbf{U}}$ are independent normal variables under the GP prior.

The authors of [144] are able to successfully sample g and the kernel hyperparameters jointly using HMC. We find that this does not work well for our problem, because the posterior is a sharply peaked function of ℓ and α for fixed $g_{\mathbf{U}}$. This issue is investigated in [148], and an alternative sampling algorithm is proposed there to mitigate it. We choose to defer kernel parameter optimization to a later work, and instead fix the parameters to values that work well in practice. We therefore perform HMC using $\theta = g_{\mathbf{U}}$ only.

To choose the HMC parameters ϵ and $L_{\max}$, we employ a Bayesian optimization algorithm that iteratively proposes and tests $(\epsilon, L_{\max})$ pairs. The algorithm attempts to identify the pair $(\epsilon*, L_{\max}*)$ that maximizes an objective function, which we take to be the expected squared jump distance [147], penalized by the number of leapfrog steps:

$$\mathrm{ESJD}(\epsilon, L_{\max}) = \frac{E\left[|\theta_k - \theta_{k-1}|^2\right]}{\sqrt{L_{\max}}}. \tag{9.19}$$

The Bayesian optimization algorithm we use is based on a GP fit of $\log \mathrm{ESJD}$ vs. $(\epsilon, L_{\max})$ and is implemented using the Scikit-Optimize library [149]. We perform this procedure for 30 iterations at the beginning of each HMC run to identify the best parameters ϵ and $L_{\max}$ for the run.

Performance on M_R data

We test the HMC procedure for the Poisson GP on the simulated data in the Multijet 1 b-tag category, using the same binning as above. We fix the squared exponential kernel parameters to $\ell = 200$ GeV and $\alpha = 200$. HMC sampling is initialized from best-fit $g_{\mathbf{U}}$ values obtained using L-BFGS and then run for several thousand iterations. The first half of the HMC samples are discarded as 'warm-up' samples, and the second half are used to estimate the posterior means and variances of each $f(\mathbf{u}_i)$. We also perform the procedure on the Lepton Multijet 2 b-tag category, which has much lower statistics. Here we use $\ell = 1000$ GeV and $\alpha = 200$.

The resulting fits are shown in Figure 9.5. We see that the Poisson GP yields fitted functions that describe the data well in both categories; it works effectively even when there are many low-count and/or empty bins in data. Note that the uncertainty bands drawn in the plots capture the uncertainty on the function values $f_{\mathbf{U}}$ only, and do not include the Poisson uncertainty on the observed data. The distribution of $n\sigma$ values in the Multijet 1 b-tag category is shown in Figure 9.6; the Poisson GP yields a similar $n\sigma$ distribution to the Gaussian GP in this case.

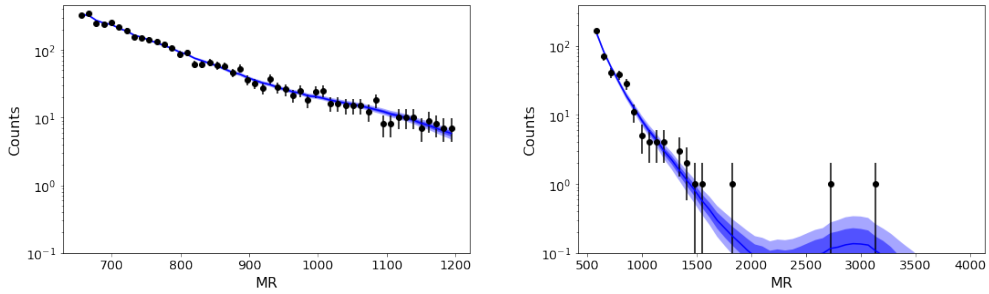


Figure 9.5: Simulated $t\bar{t}$+jets data (black points) and GP fit (blue line) using a Poisson conditional likelihood in the Multijet 1 b-tag (left) and Lepton Multijet 2 b-tag (right) categories. The light and dark bands around the fit function denote 1- and 2-sigma uncertainties on the fitted function values.

Interpolation

A straightforward application of GP regression is to interpolate the values of the fit function in unobserved or blinded bins. Predictions for f at new locations $\mathbf{V}$ are made in a doubly stochastic fashion. For each HMC sample of $g_{\mathbf{U}}$, we compute $f_{\mathbf{U}} = Ag_{\mathbf{U}}$ and compute the mean and covariance of $f_{\mathbf{V}}$ as

$$\mu_{\mathbf{V}\,|\,\mathbf{y}} = \mu_{\mathbf{V}} + K_{\mathbf{VU}}K_{\mathbf{UU}}^{-1}(f_{\mathbf{U}} - \mu_{\mathbf{U}}), \tag{9.20}$$

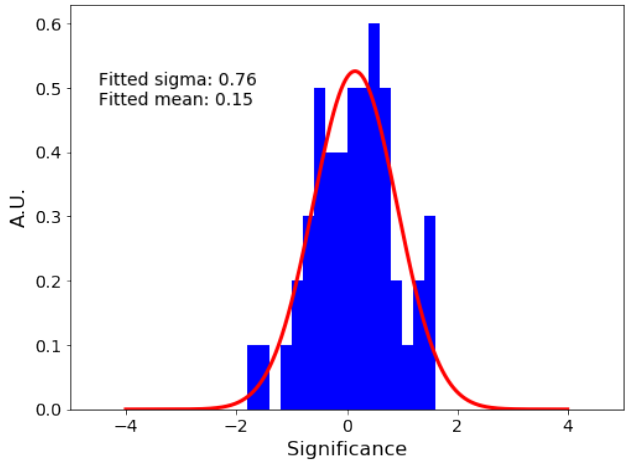


Figure 9.6: Distribution of $n\sigma$ significances in the fitted bins in the Multijet 1 b-tag fit shown in Figure 9.5. The red curve shows a Gaussian fit to the histogram.

$$K_{\mathbf{VV}\,|\,\mathbf{y}} = K_{\mathbf{VV}} - K_{\mathbf{VU}} K_{\mathbf{UU}}^{-1} K_{\mathbf{UV}}. \tag{9.21}$$

We can then randomly draw from a Gaussian with this mean and covariance to obtain a sample of $f_{\mathbf{V}}$. Repeating this for each HMC sample yields a collection of values that constitute a predictive distribution for $f_{\mathbf{V}}$.

The result of this procedure is illustrated in Figure 9.7 for the Multijet 1 b-tag category, using 50 bins in the range $650 < M_R < 1800$. Six bins in the interior of the fit range are blinded and are not used to fit the GP. The predicted function values in the blinded bins are displayed on the plot. We see that the GP smoothly interpolates the missing function values. In the case where a localized signal is expected to appear on top of a smooth background in a particular location, this procedure could be used to predict the background in the search bins, using the surrounding region as a sideband.

Signal extraction

We also investigate the ability of the GP fit to extract a signal of new physics from data. For simplicity, we assume that the signal shape $s_{\mathbf{U}}$ is known exactly. We introduce a parameter ζ that controls the strength of the signal (where $\zeta = 1$ corresponds to the presence of a signal at its theoretical cross section and $\zeta = 0$ corresponds to no signal), and use the conditional likelihood

$$p(y_i \,|\, f(\mathbf{u}_i), \zeta) = \text{Poisson}(y_i \,;\, e^{f(\mathbf{u}_i)} + \zeta \times s(\mathbf{u}_i)). \tag{9.22}$$

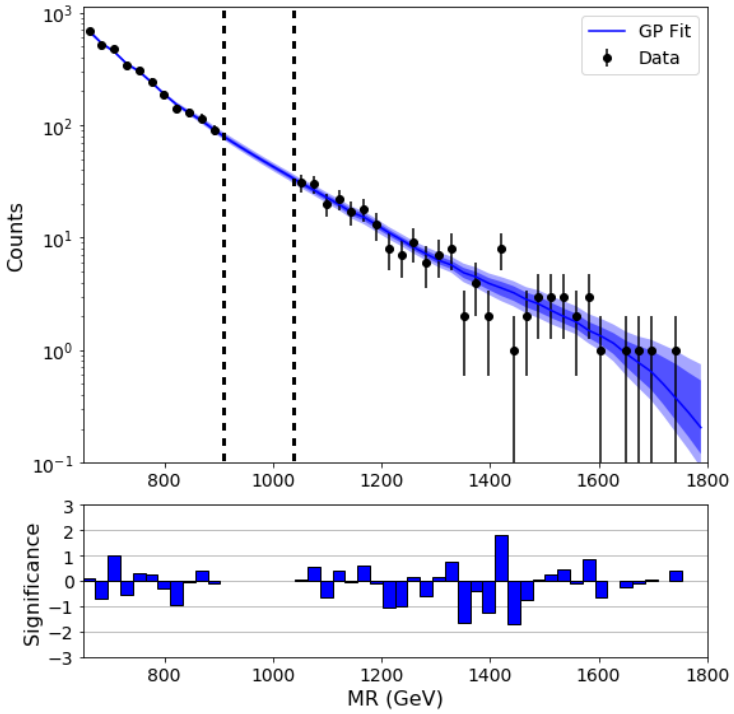


Figure 9.7: GP fit to the simulated data in the Multijet 1 b-tag category, with the six bins between the dashed lines excluded from the fit. The fitted GP is used to predict the fit function values and uncertainties inside the excluded region.

We then run the HMC algorithm with $\theta = \{g_{\mathbf{U}}, \zeta\}$, and use the sampled values of ζ to estimate the distribution of the signal strength given the observed data.

To demonstrate this method, we inject a signal with a narrow Gaussian shape (width 50 GeV) into the data in the Multijet 2 b-tag category, and perform HMC sampling to extract the signal strength. An example signal + background fit using this procedure is illustrated in Figure 9.8. We perform this test many times using a variety of injected signal strengths and observe that the fit is able to extract the true number of signal events with no observable bias (see right side of Figure 9.8).

We then apply this method to a SUSY signal, using a simulated signal sample from the T1tttt model with $m_{\tilde{g}} = 1800$ GeV and $m_{\tilde{\chi}_1^0} = 100$ GeV. The test is performed several times, using new random Poisson draws from the signal template shape in each trial. Results for two different trials are shown in Figure 9.9. We see that the GP is able to extract the true signal in some trials, while in other trials the signal peak is absorbed into the background shape.

To get insight into what is happening in the SUSY signal case, we define a grid of ζ values and compute the best-fit posterior probability of the GP model at each value

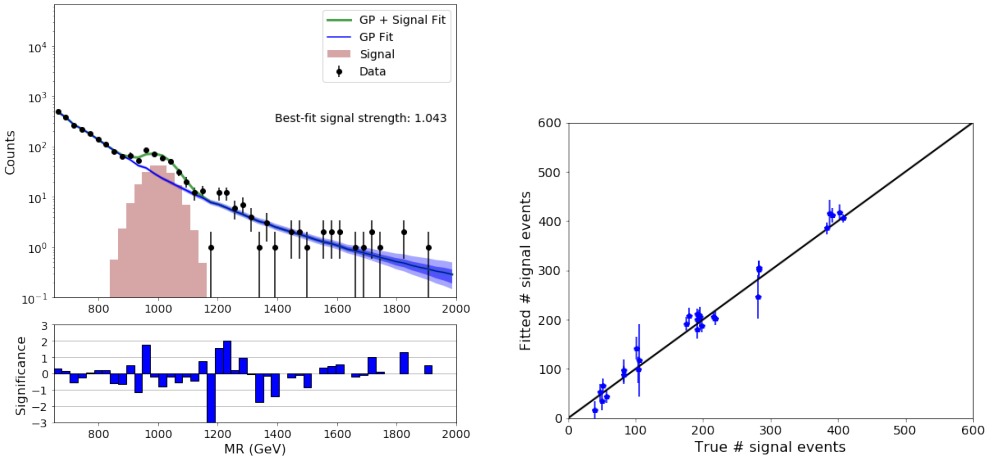


Figure 9.8: Left: example signal + background fit to the Multijet 2 *b*-tag simulated data with a narrow Gaussian signal injected. The black points indicate the binned data (including the injected signal events). The blue curve and shaded blue bands indicate the fitted GP background function and its one- and two-sigma uncertainties, and the green curve shows the sum of the fitted background and fitted signal. The red shaded distribution shows the true shape of the signal. In the bottom plot is shown the $n\sigma$ significance of the data with respect to the fitted signal + background shape in each bin. Right: scatterplot of the fitted number of signal events (with uncertainties obtained from HMC samples) versus true number of signal events across a number of trials with different injected signal strengths.

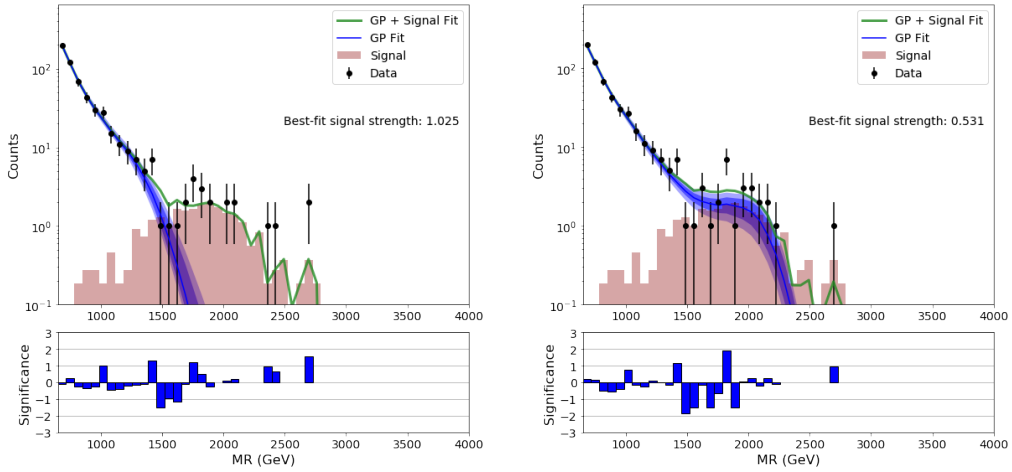


Figure 9.9: The left and right plots show signal + background fits to the simulated data in the Seven-jet 2 *b*-tag category with a SUSY T1tttt signal injected. The plot style is the same as described in Figure 9.8. In the left plot, the fit identifies the signal and assigns it the correct strength. In the right plot (which is a different trial using the same background data and true signal shape), the background function incorrectly absorbs half of the signal events.

of ζ. The posterior probability as a function of ζ is shown in Figure 9.10 for three different realizations of the dataset. We see that, depending on the exact distribution of signal events in the dataset, the likelihood may have a peak at the true value of $\zeta = 1$, or it may not.

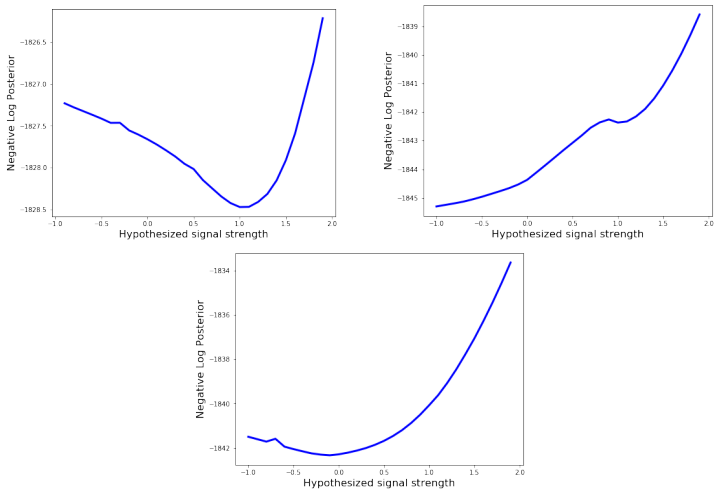


Figure 9.10: The three plots show scans of the minimum negative log posterior probability as a function of the hypothesized signal strength for a GP fit to the Seven-jet 2 b-tag category with a SUSY T1tttt signal injected. Each plot is obtained using a different random realization of the injected signal. In the left plot, the minimum posterior value is obtained at a signal strength of 1. In the right plot, there is a local minimum at 1 but the global minimum is elsewhere. In the bottom plot, there is no global or local minimum at 1.

This behavior can be understood in the context of the nonparametric nature of the GP. The GP can, in principle, fit any functional shape. Signals in the data that match the ansatz of the GP covariance function (in this case, smooth shapes with a characteristic covariance length scale ℓ) will look background-like and can be absorbed into the background shape. On the other hand, signals that do not resemble the background (like the Gaussian signal considered above, which has a width much smaller than ℓ) cannot easily be fit by the GP and will be identified as signal in the signal + background fit.

The message we take from this investigation is that when fitting a signal with a GP, care must be taken to ensure that the signal shape is assigned low probability under the GP prior. This may be accomplished via optimization of the kernel parameters (possibly via Bayesian optimization or a similar procedure), or via a creative choice of covariance function that captures the properties of the background but not the signal.

The GP prior mean function, which we have taken to be zero in the present investigation, could also play a role in improving the GP's signal extraction capability. For example, the current razor function (Eq. 9.1) could be used as the GP mean. This would explicitly encode our knowledge that the background has a falling shape and may reduce the tendency of the GP to swallow a peaking signal. To take this approach we would first require a method for fitting GP hyperparameters, which we were so far unable to do in the Poisson likelihood case. Then the four parameters of the razor fit function could be treated as GP hyperparameters and allowed to vary in the HMC sampling procedure.

Inference using a variational approximation

The above MCMC approach to GP inference is appealing primarily because:

- It asymptotically provides samples from the true (intractable) posterior distribution without placing any assumptions on the posterior shape, and
- It naturally accommodates signal + background fitting as described above.

On the other hand, inference using MCMC is computationally expensive and requires some tuning of the algorithm's parameters (here ϵ and $L_{\max}$). The posterior uncertainties can also be underestimated if the Markov chain does not fully converge to the target distribution.

An alternative to MCMC is the so-called *variational* approach, in which the posterior $p(f_{\mathbf{U}} \,|\, \mathbf{y})$ is approximated using a tractable parametric distribution, $q(f_{\mathbf{U}})$. In GP models this is conveniently carried out in an *inducing point* framework, in which the fit function values are optimized at locations $\mathbf{Z}$ that may or may not coincide with the bin centers $\mathbf{U}$ [145]. We describe this procedure in the next subsection. The use of inducing points can improve the posterior approximation and makes the computation more tractable in the case of very large numbers of observed bins. It also assists in the construction of deep GP models, as described later.

Variational inference in the inducing point framework

We follow [145] in deriving the variational approximation to the GP posterior. We consider the values of the fit function at the bin locations $\mathbf{U}$, as well as at additional locations $\mathbf{Z}$ that can be chosen arbitrarily. To obtain a tractable estimate of the

posterior at $\mathbf{U}$, $p(f_\mathbf{U} \mid \mathbf{Y})$, we first approximate the distribution of inducing function values, $p(f_\mathbf{Z} \mid \mathbf{Y})$, as a multivariate Gaussian $q(f_\mathbf{Z})$:

$$q(f_\mathbf{Z}) = \text{Normal}(f_\mathbf{Z}\,;\,\mathbf{m},\mathbf{S}), \tag{9.23}$$

where $\mathbf{m}$ is an arbitrary mean vector and $\mathbf{S}$ is an arbitrary covariance matrix, to be optimized during the model fit. The approximate distribution of function values at the observed locations, which we call $q(f_\mathbf{U})$, is then obtained by an integral over $\mathbf{Z}$:

$$q(f_\mathbf{U}) = \int p(f_\mathbf{U} \mid f_\mathbf{Z}) q(f_\mathbf{Z})\, df_\mathbf{Z}, \tag{9.24}$$

where $p(f_\mathbf{U} \mid f_\mathbf{Z})$ is Gaussian with mean and covariance given by Eqs. 9.20 and 9.21 respectively. The resulting distribution of $f(u_i)$ for a given i is Gaussian with mean and variance equal to

$$\lambda(u_i) = \mu(u_i) + \boldsymbol{\alpha}(u_i)^T(\mathbf{m} - \mu_\mathbf{Z}) \tag{9.25}$$

and

$$\kappa(u_i) = K(u_i, u_i) - \boldsymbol{\alpha}(u_i)^T(K_{\mathbf{ZZ}} - \mathbf{S})\boldsymbol{\alpha}(u_i), \tag{9.26}$$

respectively, where

$$\boldsymbol{\alpha}(u_i) = K_{\mathbf{ZZ}}^{-1} K_{\mathbf{Z}u_i}. \tag{9.27}$$

To fit the model to the data, we must optimize the values of $\mathbf{m}, \mathbf{S}$, and the locations of the inducing points $\mathbf{Z}$. We choose these parameters to minimize the Kullback-Leibler divergence between the variational posterior $q(f_\mathbf{U}, f_\mathbf{Z})$ and the true posterior,

$$\text{KL}[q \,||\, p] = \mathbb{E}_{q(f_\mathbf{U}, f_\mathbf{Z})}\left[\log \frac{p(f_\mathbf{U}, f_\mathbf{Z} \mid \mathbf{Y})}{q(f_\mathbf{U}, f_\mathbf{Z})}\right]. \tag{9.28}$$

Minimizing Eq. 9.28 is equivalent to maximizing the following objective function:

$$\mathcal{L} = \sum_{i=1}^{B} \mathbb{E}_{q(f(u_i))}\left[\log p(y_i \mid f(u_i))\right] - KL[q(f_\mathbf{Z}) \,||\, p(f_\mathbf{Z})], \tag{9.29}$$

where $p(f_\mathbf{Z})$ is the GP prior for $\mathbf{Z}$. The expectation value in Eq. 9.29 can in general be evaluated using MC sampling. In the Poisson likelihood case, it can be evaluated analytically:

$$\mathbb{E}_{q(f(u_i))}\left[\log p(y_i \mid f(u_i))\right] = y_i \lambda(u_i) - e^{\lambda(u_i) + \kappa(u_i)/2}. \tag{9.30}$$

The KL divergence term $KL[q(f_\mathbf{Z}) \,||\, p(f_\mathbf{Z})]$ can also be computed in closed form. Optimizing the bound Eq. 9.29 yields the best approximation to the true posterior possible within the chosen parametric family.

The objective $\mathcal{L}$ is a lower bound on the marginal likelihood, $p(\mathbf{Y})$, of the GP model. One could therefore try to optimize the covariance hyperparameters (e.g. ℓ and α) via maximization of $\mathcal{L}$, analogously to what was done earlier for the GP with Gaussian conditional likelihood. However, as noted in [145], $\mathcal{L}$ is not a uniformly tight bound on $p(\mathbf{Y})$ for all hyperparameter values, so one should not do this for hyperparameters (such as the SUSY signal strength) for which a rigorous statistical interpretation is required. Because of this, the variational GP model is not appropriate for direct signal + background fitting; it should be used only for interpolation.

Inference for deep GPs

The variational GP framework can be extended in a straightforward way to *deep gaussian processes* (DGP). Here we again follow the presentation in [145]. A DGP consists of L GPs stacked on top of one another, such that the output f^l of the lth GP in the stack becomes the input u^{l+1} to the $(l+1)$st GP. The conditional likelihood of the model is $p(y_i \mid f_i^L)$, where f_i^L is obtained by passing the input point u_i through all L layers of the GP.

The DGP can be seen as mapping its inputs into a latent space ($u_i \to u_i^L$) and applying a standard GP (the final DGP layer) on that space. The covariance between two output function values f_i^L and f_j^L is given by $K^L(u_i^L, u_j^L)$, where K^L is the last layer's covariance function. Since the mapping $u_i \to u_i^L$ is nonparametric, $K^L(u_i^L, u_j^L)$ defines a nonparametric notion of closeness of each pair of input points, governed by DGP layers $1, \ldots, L-1$. By allowing the model to effectively learn this mapping, the DGP alleviates a central challenge in GP modeling, namely the need to manually specify the covariance function.

Inference in a DGP can be performed using a variational approximation similar to that used in the previous section. We define a set of inducing points $\mathbf{Z}^l$ at each layer l, and posit a variational posterior for the function values $f^l_{\mathbf{Z}^l}$ at each layer:

$$q(f^l_{\mathbf{Z}^l}) = \text{Normal}(f^l_{\mathbf{Z}^l}\,;\, \mathbf{m}^l, \mathbf{S}^l), \tag{9.31}$$

where the mean vectors $\mathbf{m}^l$ and covariance matrices $\mathbf{S}^l$ can be chosen freely. To train the model, we find the values of $\mathbf{m}^l$, $\mathbf{S}^l$, and $\mathbf{Z}^l$ for each l such that the objective

$$\mathcal{L} = \sum_{i=1}^{B} \mathbb{E}_{q(f_i^L)} \left[\log p(y_i \mid f_i^L)\right] - \sum_{l=1}^{L} KL\left[q(f^l_{\mathbf{Z}}) \,||\, p(f^l_{\mathbf{Z}})\right] \tag{9.32}$$

is maximized. The expectation in the first term is evaluated via Monte Carlo sampling. This is done by generating a standard Gaussian noise variable ϵ_i^l at each DGP layer and recursively computing

$$\hat{f}_i^l = \lambda(\hat{f}_i^{l-1}) + \epsilon_i^l \sqrt{\kappa(\hat{f}_i^{l-1})}, \tag{9.33}$$

where $\lambda(\hat{f}_i^{l-1})$ and $\kappa(f_i^{l-1})$ are computed using Eqs. 9.25 and 9.26, respectively, and the first-layer inputs $\hat{f}_i^1$ are the locations u_i. This procedure yields samples of f_i^L that can be used to approximate the expectation in Eq. 9.32. Because the stochasticity in the sampling procedure is confined to the variables ϵ_i^l, the computed expectation can be differentiated with respect to $\lambda(\hat{f}_i^{l-1})$ and $\kappa(\hat{f}_i^{l-1})$, and thus, via the chain rule, with respect to the parameters $\mathbf{m}^l$, $\mathbf{S}^l$, and $\mathbf{Z}^l$ (this is the so-called reparameterization trick for differentiating stochastic loss functions [150]). This allows the model to be trained using backpropagation in a similar fashion to an artificial neural network.

Results

We implement the variational GP and the DGP using PyTorch and test them on the M_R data in various event categories. The best results are obtained by initializing $\mathbf{Z}$ ($\mathbf{Z}^l$, for the DGP) to the input data locations $\mathbf{U}$. We also initialize $\mathbf{m}$ ($\mathbf{m}^L$, for the DGP) to the log of the number of observed data counts in each bin, and for the DGP set $\mathbf{m}^l = \mathbf{Z}^l$ for $l < L$ so that each internal layer initially acts as the identity function. Following [145], we set the GP prior mean to the identity function for the internal DGP layers. Finally, we initialize $\mathbf{S}$ and all DGP layer covariances $\mathbf{S}^l$ to a small multiple of the identity.

The techniques described in the previous sections are used to evaluate the necessary matrix products and KL divergences. We use PyTorch's autodifferentiation to compute derivatives of Eqs. 9.29 and 9.32 with respect to the model parameters. We use the Adam optimizer [151] to learn the variational parameters and inducing point positions, as we find it yields much better performance than L-BFGS for this task. Numerical stability is improved greatly by scaling all M_R values down by a factor of 1000 and using an appropriately smaller kernel covariance length ℓ. For the DGP, we use $\ell = 1, \alpha = 100$ for all event categories.

Example variational GP fit results are shown in Figure 9.11, and results using the DGP are shown in Figures 9.12 and 9.13. For the DGP fits we also illustrate the

mapping $u_i \rightarrow u_i^L$ from the input space into the final DGP layer, to show the effective covariance structure that is learned. We see that the DGP tends to 'stretch out' some parts of the input space, decreasing the covariance in those regions, and 'shrink' other parts of the space, increasing the covariance in those regions. It also selectively adds additional uncertainty in some regions, particularly at high M_R where bin counts are very low. The effect of this can be seen by comparing the Multijet 1 b-tag fits in Figures 9.11 (variational GP) and 9.12 (DGP with five layers). The uncertainties in the DGP case are higher, leading to fewer significant deviations between the data and the fitted function.

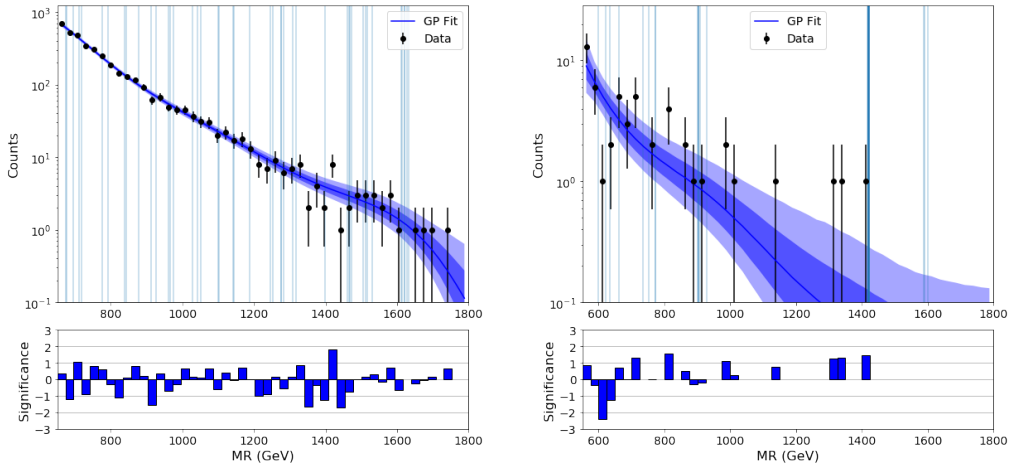


Figure 9.11: Variational GP fits to the Multijet 1 b-tag (left) and Lepton Multijet 3 b-tag (right) razor event categories. The light blue vertical lines in the upper part of each plot indicate the positions of the fitted inducing points.

Outlook

We have investigated the viability of modeling the background with a GP in searches for new physics with low-count binned data. The results indicate that inference for a GP with Poisson conditional likelihood is tractable in this case, and that it yields background function models that adapt flexibly to different dataset sizes.

The signal extraction tests we performed show that a GP is capable of fitting a SUSY signal in some cases, but that the procedure can fail if the signal does not look different enough from the background. The GP therefore does not serve as a drop-in replacement for the existing razor fit function (Eq. 9.1).

On the other hand, the ability of the GP to interpolate between fitted data points is potentially a powerful tool for SUSY searches. For example, the CMS $H \rightarrow \gamma\gamma$

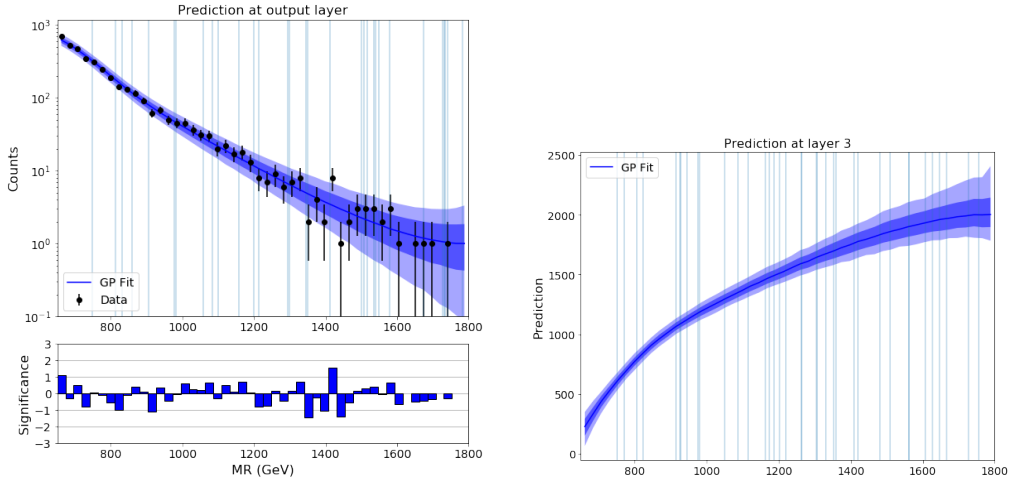


Figure 9.12: Left: Five-layer DGP fit to the Multijet 1 b-tag event category. Right: Visualization of the mapping from the input space to the output of the penultimate DGP layer. The light blue vertical lines in each plot indicate the positions of the fitted inducing points at the last and second-to-last layers, respectively.

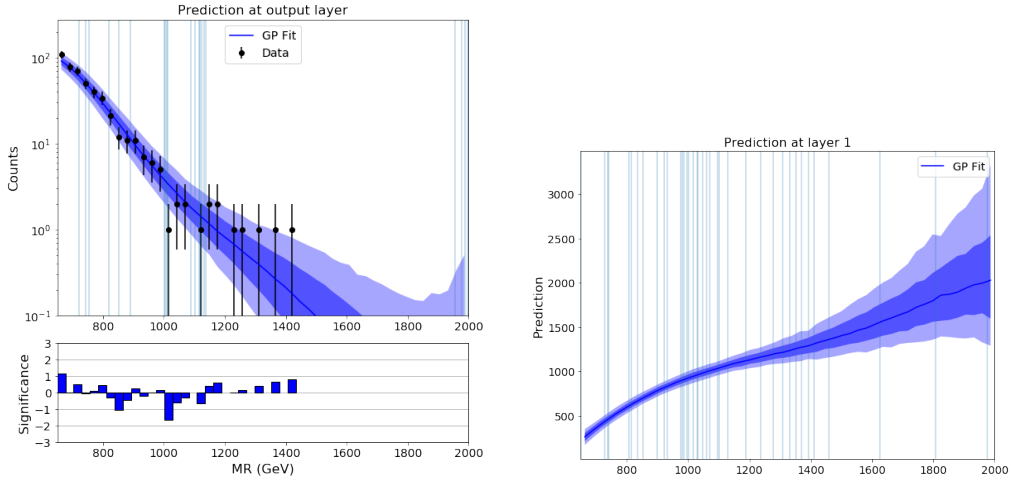


Figure 9.13: Left: Three-layer DGP fit to the Dijet 0 b-tag event category. Right: Visualization of the mapping from the input space to the output of the penultimate DGP layer. The light blue vertical lines in each plot indicate the positions of the fitted inducing points at the last and second-to-last layers, respectively.

tagged razor SUSY search [152] models the $m_{\gamma\gamma}$ spectrum using a number of different ad-hoc functional forms, and must use a complex system of bias tests and information criteria to determine which function to use in each analysis category. The region of $m_{\gamma\gamma}$ around the Higgs mass is blinded and the background fit is performed in the surrounding region. The GP interpolation method described above could provide a principled alternative to the current approach to background prediction and

uncertainty estimation in the blinded bins. The variational GP approach is likely the preferred technique for this use case, because it is more computationally convenient than MCMC. Deep GPs also present a promising direction for background modeling; their flexible quantification of uncertainty may justify the additional complexity they introduce. More broadly, we hope that these initial GP fit studies may inspire or guide the development of other future searches for new physics.

Part IV:

Conclusion

All human wisdom is summed up in two words –
'Wait and Hope.'

Edmond Dantes, in Alexandre Dumas,
The Count of Monte Cristo

Chapter 10

OUTLOOK FOR LHC 2018 AND BEYOND

Neither the searches for supersymmetry detailed in this thesis, nor the other CMS searches summarized in Figure 8.38, provide conclusive evidence for physics beyond the standard model. Limits on the masses of gluinos, squarks, and higgsinos in simplified natural SUSY scenarios are now substantially tighter than those obtained with the Run I dataset. The consequences for naturalness in the broader context of the MSSM and beyond are the subject of current study [153].

The situation will be further clarified by future LHC runs. It is expected that the full CMS Run II dataset (2015-2018) will comprise more than 100 fb^{-1} of integrated luminosity. LHC operations through 2022 are expected to yield a grand total of 300 fb^{-1}. After that, the planned high-luminosity LHC (HL-LHC) will run from approximately 2025 to 2030, yielding up to 3000 fb^{-1} [154]. One analysis of naturalness in the context of the LHC estimates that 3000 fb^{-1} of LHC data will be sufficient to conclusively discover or exclude most natural SUSY scenarios, assuming that the degree of fine-tuning is less than 3%, i.e., if the terms in Eq. 3.21 do not yield cancelations finer than one part in 30 [155]. This indicates that the next several years of LHC running should yield a strong statement about whether naturalness is a useful paradigm for predictions about SUSY.

In the meantime, null results in searches for squarks and gluinos at the LHC have led to increased interest in other SUSY signatures. Searches for direct electroweak production of higgsinos have received more attention. The implications of past searches for scenarios such as RPV and stealth SUSY, which may manifest in lower-E_T^{miss}, higher N_{jets} signal events, have been studied in detail [153].

Other searches for exotic particles at the LHC could yet uncover new physics beyond the SM, possibly unrelated to SUSY. Strategies such as data scouting that use the CMS dataset in novel ways can expand the range of searches that can be performed, increasing the chance of finding a sign of new physics if one exists. Optimal use of the CMS dataset may also necessitate moving beyond the traditional HEP analysis techniques. Recent research in pattern recognition and machine learning has yielded a huge variety of new algorithms for statistical modeling and identification of signals in data. Studying the applicability of these techniques to

HEP problems has become a fruitful area of investigation and has the potential to revolutionize the way we perform data analysis.

We include three appendices to this thesis in which we explore additional topics not specific to CMS. Appendix A studies the potential of calorimetric devices to provide picosecond-level time-of-flight information for incident particles. We include a description of the planned new MIP Timing Detector for the Phase II upgrade of CMS. Appendix B studies new neural network based particle tracking algorithms, which have the potential to modify current paradigms for track reconstruction. Appendix C studies distributed training of machine learning models using large computing clusters. This work offers new tools for large-scale model building to assist in future development of new statistical techniques for event reconstruction and analysis.

Appendix A

PRECISION TIMING CALORIMETRY FOR LHC RUN II

There has been significant recent interest in precision timing as a tool for event reconstruction and analysis at particle colliders. It is particularly attractive as a tool for pileup mitigation in high-occupancy collision events, such as those anticipated at the planned high-luminosity LHC (HL-LHC).

I participated in beam test experiments in the spring of 2014 aimed at characterizing the timing properties of LYSO crystal-based calorimeters for use in high-energy physics applications. This appendix describes these experiments and places them in their context in the development of precision timing detectors in CMS.

A.1 Motivation: the HL-LHC

After Run II (2015-2018) and Run III (2020-2022), the LHC will undergo a long shutdown ('LS3'), during which upgrades will be performed to allow the collider to achieve a luminosity of 5-8 times its current design value. The upgrade is motivated primarily by the diminishing marginal returns of continuing to run the LHC at its current luminosity: after 2020, the running time needed to halve the statistical uncertainty on measurements will be more than ten years [154].

Achieving the HL-LHC's target luminosity entails increasing the number of pileup collisions per bunch crossing to as many as 200 (corresponding to 7.5×10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$). In such a collision environment, it becomes much more difficult to disentangle the primary interaction event from the pileup background. The core CMS reconstruction algorithms, such as PF, degrade signficantly in performance.

Precision timing is a powerful tool for pileup mitigation that has the potential to recover the present-day performance of CMS PF reconstruction under HL-LHC conditions. In a HL-LHC bunch crossing, *pp* interactions will be distributed in time with an RMS spread of 180-200 ps. If every track and calorimeter deposit in the event could be tagged with a timestamp with an uncertainty of ~30 ps, particles with times incompatible with the primary vertex could be flagged as pileup and rejected. This would reduce the effective number of pileup collisions back to the level obtained at the current LHC. Equipping CMS with precision timing capabili-

ties during the Phase II upgrade of the detector during LS3, either by modification of the current detector and readout electronics or by installation of a dedicated sub-detector, would therefore have a strong positive impact on physics performance during the HL-LHC period.

The beam experiments described in the following sections have the aim of benchmarking the timing performance of photodetectors and scintillating crystals for high-energy physics applications. Motivated by the discussion above, we use a time resolution of ~30 ps as a target for particle time-of-flight measurements.

A.2 LYSO crystals and fast photodetectors

In our experiments we consider calorimeters consisting of cerium-doped lutetium yttrium oxyorthosilicate (LYSO) scintillating crystals interfaced with photodetectors. LYSO crystals are desirable for particle physics experiments due to their short radiation length (X_0 = 1.14 cm), large light yield (30,000 scintillation photons per MeV), and high radiation tolerance [156]. They also have a very fast rise time (< 72 ps [157]), making them ideal for precision timing applications.

The photodetectors we use for measurement of the LYSO scintillation photons are microchannel plate photomultiplier tubes (MCP-PMTs) [158]. An MCP-PMT consists of a photocathode, a micro-channel plate (MCP) made up of a large number of micron-scale glass capillaries, and an anode. Photoelectrons emitted by the photocathode impinge on the inner walls of the MCP, where they create a cascade of electrons that strike the anode, yielding a signal. We use MCP-PMTs manufactured by Hamamatsu and Photek [159] in our experiments.

Digital readout of MCP-PMT pulses is performed using a DRS4 waveform digitizer evaluation board [160]. The DRS4 unit has a sampling rate of 5 GHz and provides 1024 samples per readout.

The time resolution of a crystal calorimeter like those we consider can be roughly decomposed into five factors (see Figure A.1), corresponding to jitter from the following processes:

1. The shower development time (the time between the particle's entry into the crystal and its first interaction)

2. The conversion of the incident particle's energy into scintillation light

3. The propagation of the scintillation photons through the crystal to the photodetector

4. Transit of the signal through the photodetector and its output electronics

5. Digitization of the signal by the DRS4 unit

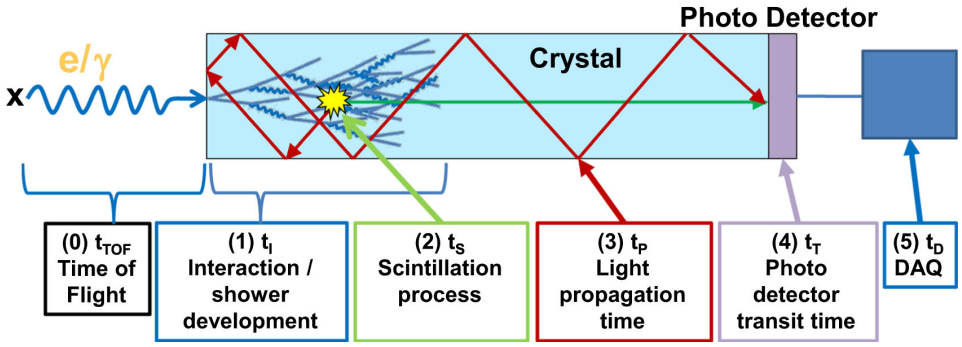


Figure A.1: Cartoon showing the contributions to the timing resolution of a crystal calorimeter for an incident photon or electron.

We characterize the time resolution of LYSO crystal calorimeters by studying the contributions from the first three factors above. The combined time resolution of the MCP-PMTs and the readout electronics is approximately 20 ps [161]. To determine the time resolution of the DRS4 unit, light pulses from a picosecond laser are directed at an MCP-PMT. The output of the MCP-PMT is split and sent to two DRS4 input channels. The difference between the two channel times is measured in many trials. The RMS spread of the time difference, which we take as a measurement of the DRS4's intrinsic time resolution, is about 5 ps.

A.3 Beam test experiments in Spring 2014

Our experiments are performed at the Fermilab Test Beam Facility (FTBF) using beams of electrons with energies ranging from 4 to 32 GeV. Event readout is triggered by a 2×2 mm^2 scintillator counter placed upstream of the experimental setup. Electron identification is performed using a Cherenkov detector provided by the FTBF. All detector elements are placed inside a copper-lined dark box, except for the Cherenkov detector, which is located in another part of the detector hall.

We carry out measurements on two different LYSO calorimeter setups:

1. A $1.7 \times 1.7 \times 1.7$ cm^3 LYSO cube coupled to an MCP-PMT (Figure A.4)

2. A shashlik-style sampling calorimeter cell (shown in Figure A.2) consisting of alternating plates of LYSO and tungsten absorber[1]. Wavelength shifting (WLS) fibers passing lengthwise through the cell transmit scintillation light from the tiles to photodetectors on both ends (see Figure A.7). We additionally investigate a setup in which we attach photodetectors directly to the sides of two adjacent LYSO tiles inside the shashlik cell (Figure A.8), allowing us to measure the time without the jitter from the WLS fibers (though at the cost of lower photostatistics).

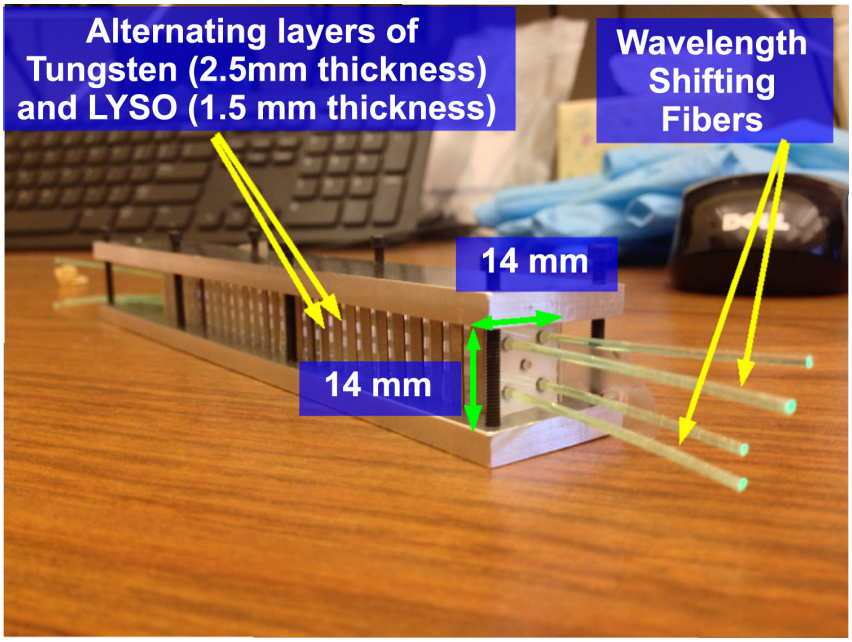


Figure A.2: Photo of the LYSO-tungsten shashlik calorimeter cell with WLS fiber readout. The cell contains 28 LYSO crystal plates and 27 tungsten plates.

In each of these setups we use a standalone MCP-PMT, located upstream of the LYSO calorimeter, to provide a reference for the time-of-flight (TOF) measurement. This device exhibits a sharp signal pulse (shown on the left side of Figure A.3). We fit the peak of the signal pulse with a Gaussian function to extract a reference timestamp. The fit is performed using eight DRS4 samples around the maximum of the pulse; the mean of the fitted Gaussian is assigned as the timestamp t_0.

The LYSO scintillation light produces signal pulses that exhibit a fast rise and a slower decay (see right side of Figure A.3). To extract a timestamp from a scintilla-

[1] A calorimeter with cells of this type was proposed for the Phase-II upgrade of the CMS ECAL endcaps. Its performance was benchmarked here with this application in mind; however, the shashlik design was ultimately not chosen for the ECAL upgrade.

tion pulse, we use a linear constant-fraction fit to the rising edge of the pulse. The fit is performed between 10% and 60% of the maximum pulse height, and the point at which the fitted line reaches 20% of its maximum value is chosen as the pulse time t_1.

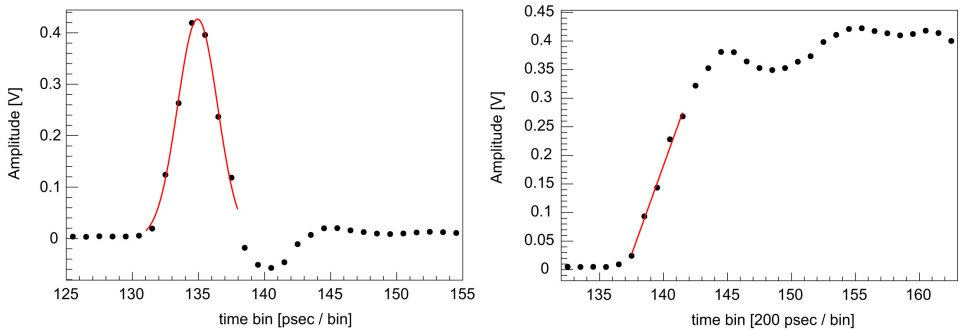


Figure A.3: Left: example Gaussian fit used to assign a timestamp to a pulse from a standalone MCP-PMT. Right: example constant-fraction fit to the rising edge of a LYSO scintillation pulse recorded by an MCP-PMT.

Quality cuts are applied on the reference and scintillation pulses to eliminate noise events. Pulses with amplitude larger than 500 mV saturate the DRS4 input and are not considered. Pulses smaller than 20 mV are rejected due to the difficulty of distinguishing them from noise.

TOF measurements using $1.7 \times 1.7 \times 1.7$ cm^3 LYSO cube

The LYSO cube calorimeter setup is shown in Figure A.4. A Hamamatsu MCP-PMT is optically coupled to the cube of LYSO, and a second MCP-PMT is placed upstream to provide a reference time measurement. Large lead bricks are placed in front of the MCP-PMT on the LYSO to avoid direct hits by stray electromagnetic shower particles. A lead layer corresponding to 4.5 X_0 is placed in front of the LYSO cube to act as a radiator.

The small size of the cube results in incomplete containment of the electromagnetic shower from the incident electron. We measure the energy resolution of the setup by examining the distribution of the LYSO pulse integral, which is proportional to the total collected charge. We obtain the energy peak shown on the left side of Figure A.5 and fit it with a Gaussian function, extracting an energy resolution of 20%.

The TOF measurement is carried out using this setup by extracting the timestamps t_0 and t_1 as described above and obtaining the width of the distribution of $t_1 - t_0$ in

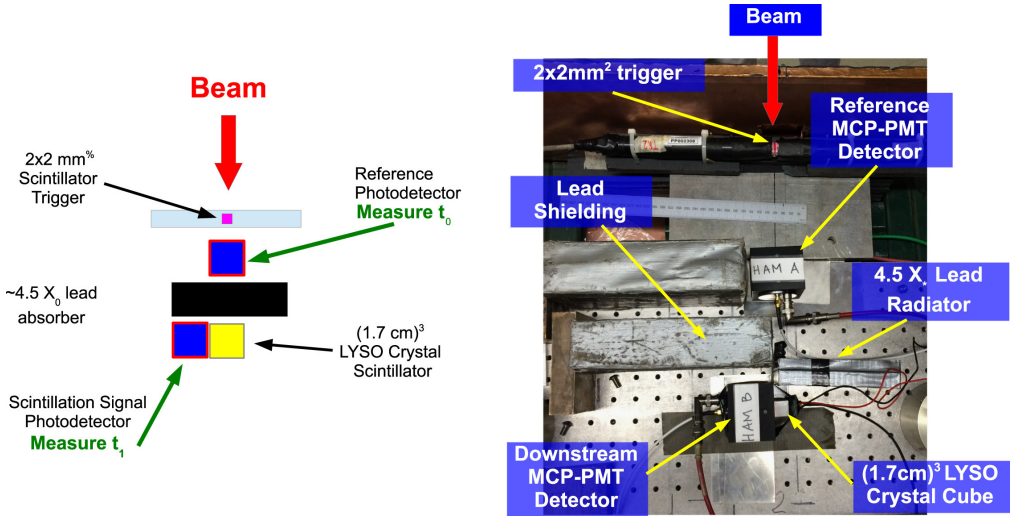


Figure A.4: Left: diagram of the experimental setup for the TOF measurement using a LYSO crystal calorimeter. Right: photo of the setup.

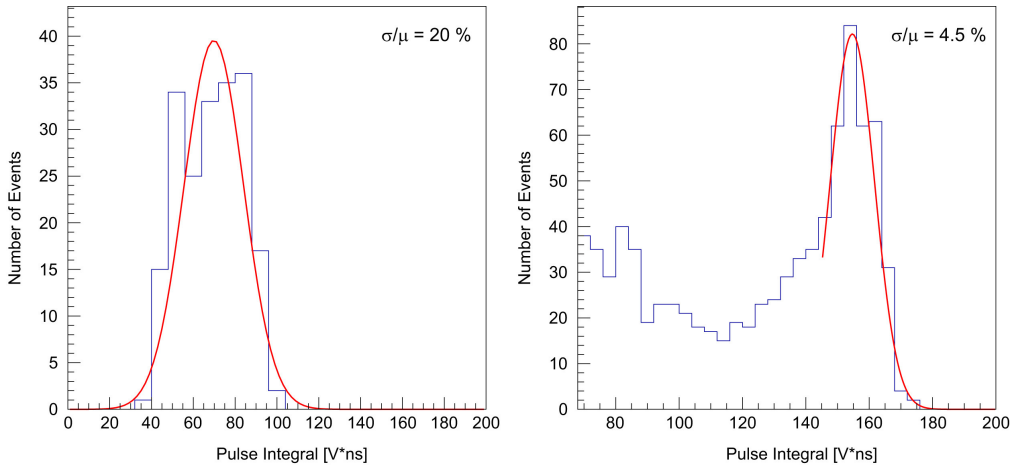


Figure A.5: Histograms of the pulse integral recorded in selected events for the MCP-PMT coupled to the LYSO crystal cube (left) or the LYSO-tungsten shash-lik cell with DSB1 fiber readout (right). In the right plot, background events are included due to misconfiguration of the Cherenkov detector for that run.

a large sample of events. Results for electron beams of different energies are shown in Figure A.6. We find a best time resolution of 34 ps, for electrons with 32 GeV beam energy. The results are plotted as a function of beam energy in the left panel of Figure A.12.

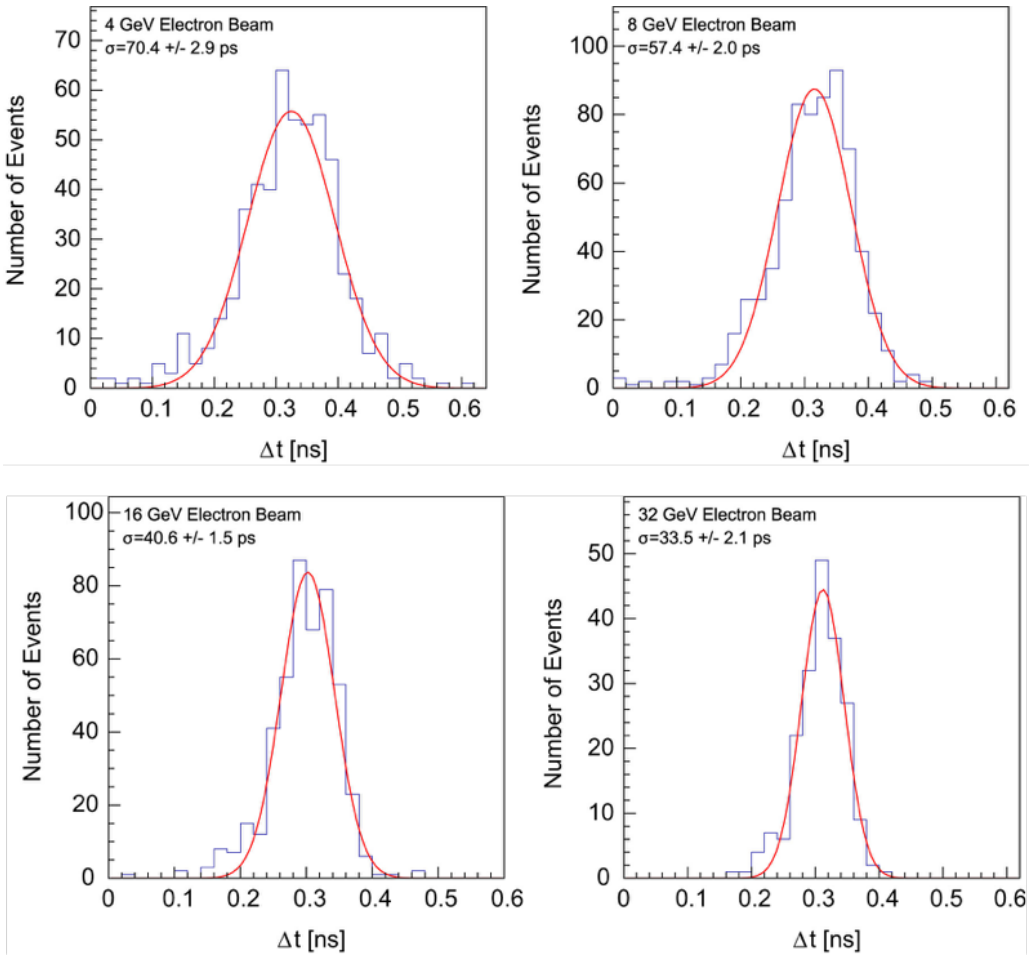


Figure A.6: Distributions of the measured TOF for the LYSO cube calorimeter, with fits to a Gaussian function to estimate the time resolution. Results are shown for electron beams with 4, 8, 16, and 32 GeV energies.

TOF measurements using LYSO-tungsten shashlik cell

The setup for measuring the time resolution of the LYSO-tungsten shashlik calorimeter cell with WLS fiber readout is shown in Figure A.7. Readout of the shashlik cell is tested using both Y11 and DSB1 [162] fibers. The pulse shapes obtained with the two fiber types are compared in Figure A.9. The DSB1 fibers yield a much faster rise time (2.4 ns) than the Y11 fibers (7.1 ns); they therefore provide a better choice for TOF measurements.

For the shashlik cell with DSB1 fiber readout, we obtain the distribution of pulse integrals shown on the right-hand side of Figure A.5. The distribution shown includes background from non-electron particles in the beam, due to an issue with the Cherenkov detector for this run. We obtain an energy resolution of approximately

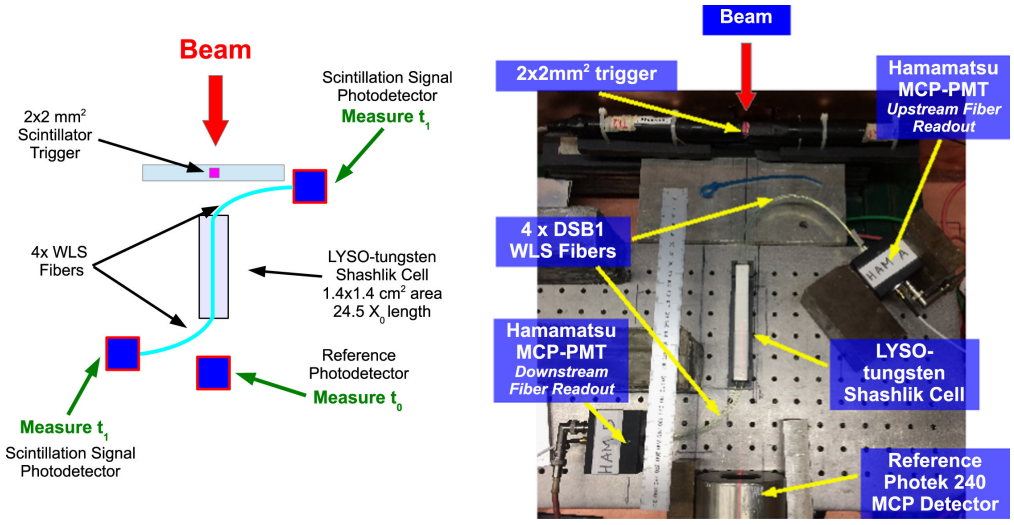


Figure A.7: Left: diagram of the experimental setup for the TOF measurement using the LYSO-tungsten shashlik cell with WLS fiber readout. Right: photo of the setup.

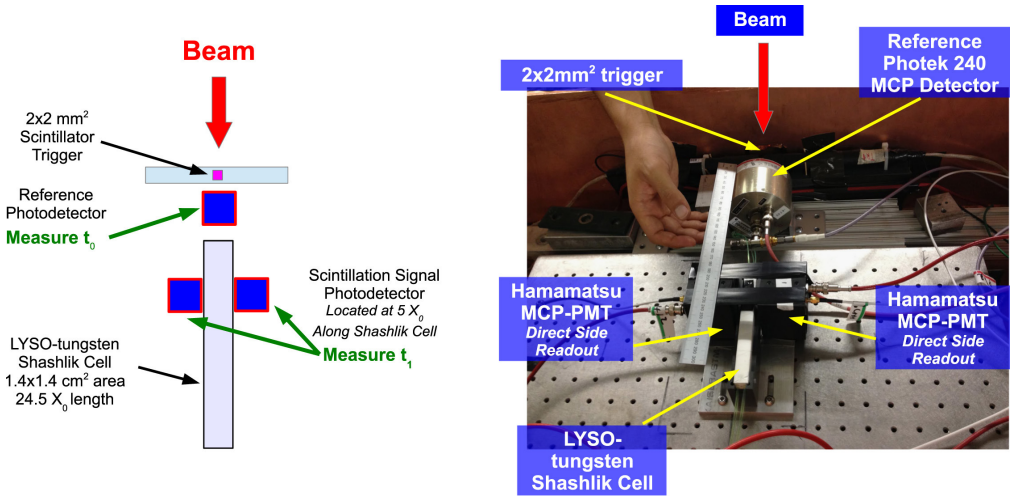


Figure A.8: Left: diagram of the experimental setup for the TOF measurement using the LYSO-tungsten shashlik cell with direct side readout by MCP-PMTs. Right: photo of the setup.

5%.

The TOF measurement is carried out using this setup in the same way as above; results are shown in Figure A.10 and summarized in the middle panel of Figure A.12. The best time resolution we obtain is 104 ps. We see from the energy dependence of the time resolution that the measurement is still in the photostatistics limited regime, so the resolution could be improved if the light collection efficiency were increased.

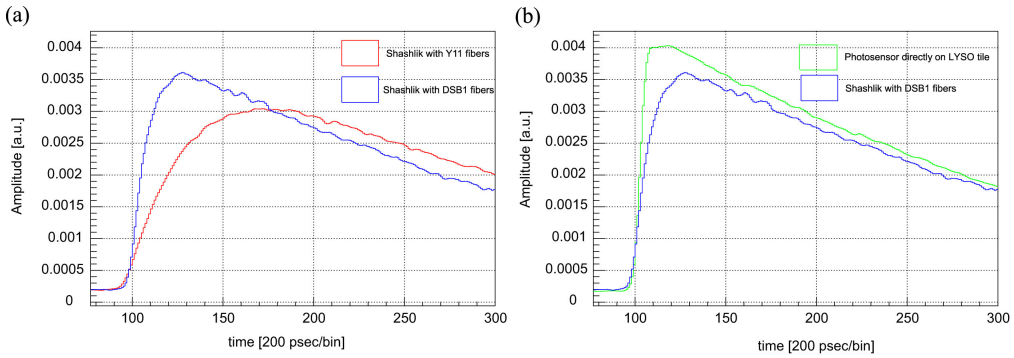


Figure A.9: Left: average MCP-PMT pulse shapes from the LYSO-tungsten shashlik cell read out by Y11 (red) or DSB1 (blue) WLS fibers. Right: comparison of average MCP-PMT pulses from the shashlik cell with DSB1 fiber readout (blue) and with direct optical coupling of MCP-PMTs to two LYSO tiles (green).

We also consider direct readout of two adjacent LYSO tiles in the shashlik cell using MCP-PMTs. This measurement circumvents the WLS fiber readout and the associated time jitter from photon transit through the fibers, at the cost of decreased photostatistics from only reading out two tiles. The setup is illustrated in Figure A.8. The TOF resolutions measured at different beam energies are displayed in Figure A.11 and plotted as a function of energy in the right panel of Figure A.12. We obtain a best time resolution of 55 ps. We see again that the measurement is photostatistics limited; fitting the time resolution with a $1/\sqrt{E}$ term plus a constant, we obtain a constant term of about 30 ps.

A.4 Proposed MIP timing detector for the CMS Phase II upgrade

The studies detailed in the previous section establish that LYSO-based particle detectors can achieve time resolutions on the order of 30 ps, the desired target for HL-LHC applications. Recently, LYSO has been chosen as the active medium for the barrel region of a dedicated MIP timing detector (MTD) for CMS, which will be installed during LS3 as part of the CMS Phase II upgrade. The MTD, which is illustrated in Figure A.13, will consist of a thin layer between the inner tracker and the calorimeters. It comprises a barrel, made of LYSO crystals coupled to silicon photomultipliers (SiPMs) for readout; and two endcaps, made of low-gain silicon avalanche diodes (LGADs). Its goal is to achieve hermetic coverage of the interaction region for $|\eta| < 3.0$ and to provide charged particle timing measurements with ~30 ps resolution. The choice of LYSO and SiPMs for the barrel part of the MTD is motivated by its low cost, the maturity of the technology, and its excellent

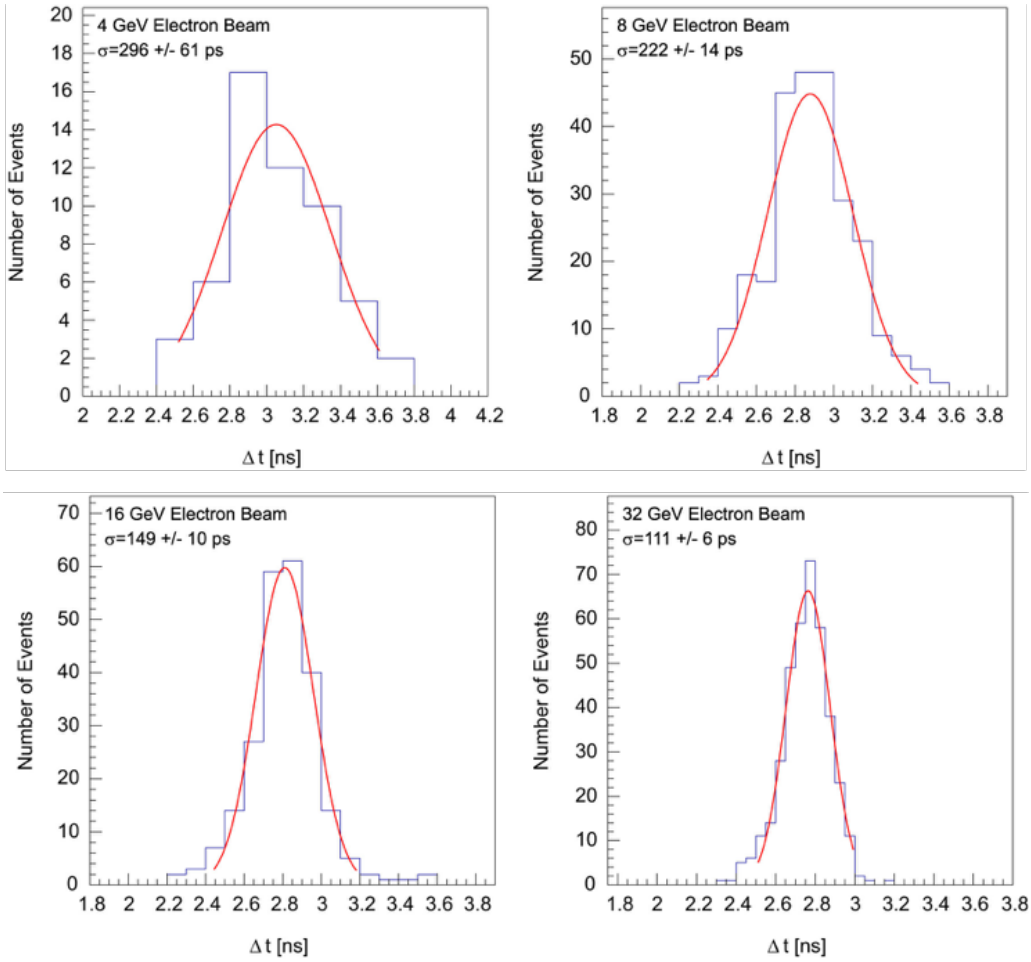


Figure A.10: Distributions of the measured TOF for the LYSO-tungsten shashlik cell with DSB1 fiber readout, with fits to a Gaussian function to estimate the time resolution. Results are shown for electron beams with 4, 8, 16, and 32 GeV energies.

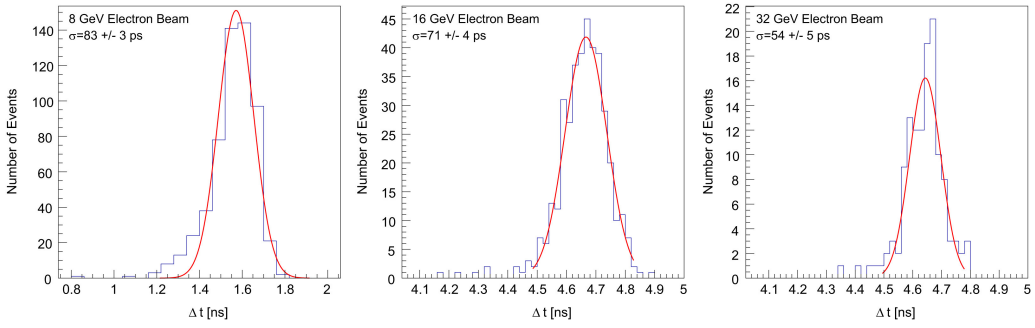


Figure A.11: Distributions of the measured TOF for the LYSO-tungsten shashlik cell direct side readout by MCP-PMTs, with fits to a Gaussian function to estimate the time resolution. From left to right, results are shown for electron beams with 8, 16, and 32 GeV energies.

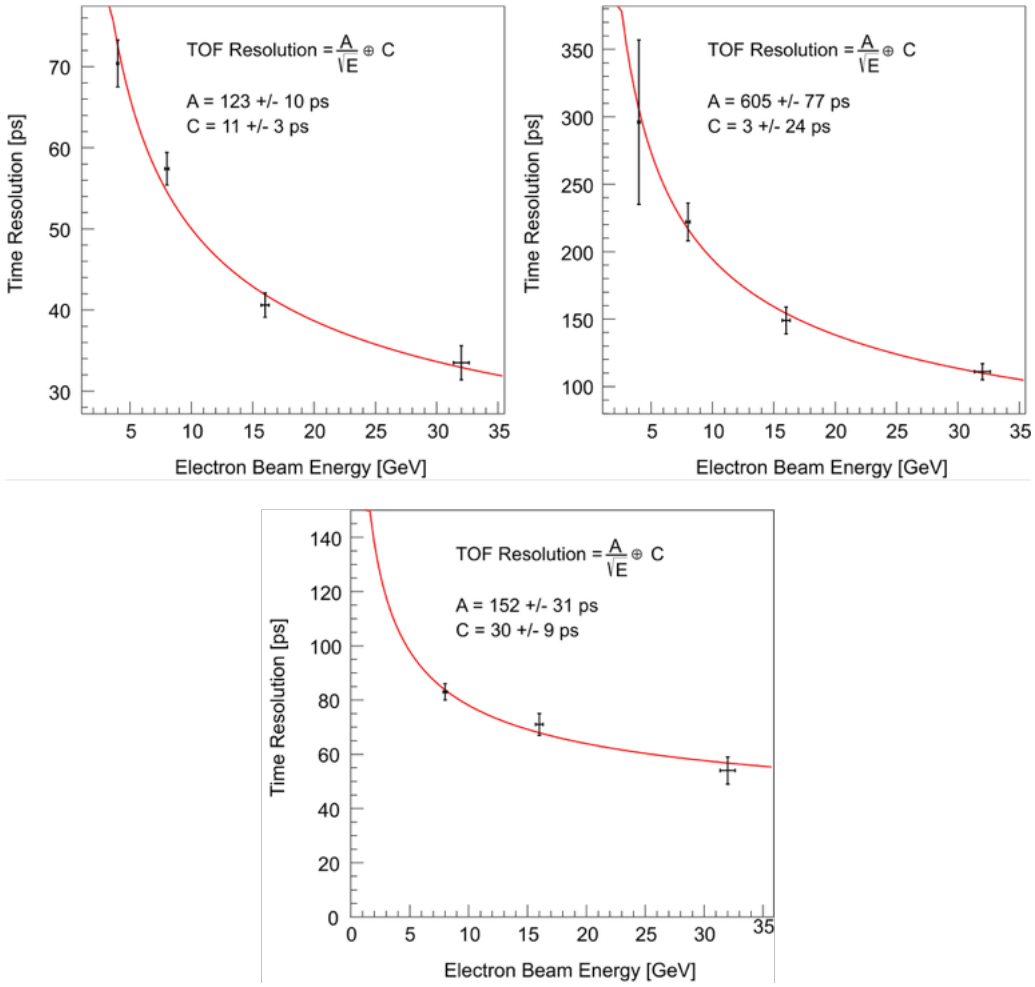


Figure A.12: Timing resolution as a function of electron beam energy for the LYSO cube calorimeter (left), the LYSO-tungsten shashlik cell with DSB1 fiber readout (middle), and the shashlik cell with direct side readout of two tiles (right). The data points in each plot are fit with a $1/\sqrt{E}$ term plus a constant.

timing properties. The choice of LGADs for the endcap timing detector is primarily motivated by their superior radiation hardness.

The Phase II upgrade of the CMS ECAL barrel [163], and the replacement of the ECAL and HCAL endcaps with a new High-Granularity Calorimeter (HGCAL) [164], will increase the CMS detector's ability to measure time information for high-energy showers. However, these detectors alone cannot achieve global event timing for minimum ionizing particles (MIPs) and reconstructed vertices. The planned MTD will deliver this by associating high-precision timestamps to tracks and calorimeter deposits.

The physics impact of a hermetic MIP timing layer has been studied in detail. Per-

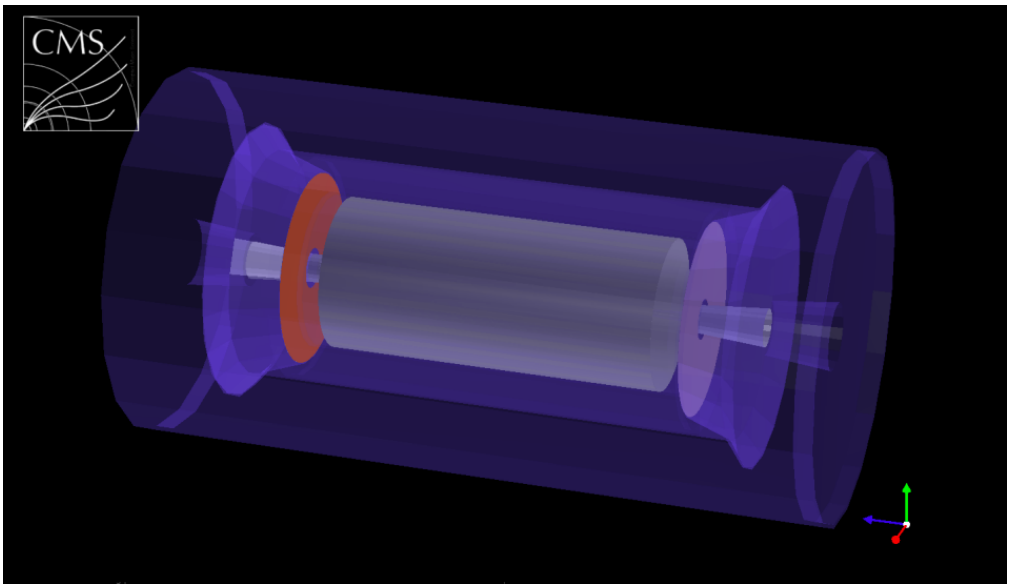


Figure A.13: Diagram of the planned MIP Timing Detector for the CMS Phase II upgrade. The barrel (gray cylinder) is situated between the tracker and the ECAL, and the endcaps (orange discs) are installed on the nose of the endcap calorimeters.

forming primary vertex reconstruction using timing information (via a generalized version of the deterministic annealing algorithm used currently in CMS) reduces the incidence of vertex merging at 200 pileup from 15% to 1%. Vertex timing can be combined with individual particle timestamps to enable vertexing of neutral particles, such as photons in $H \to \gamma\gamma$ events. An illustration of this is given in Figure A.14, which shows the positions and times of the reconstructed vertices in an event along with vertex hypotheses for the two photons. In muon identification, timing can prevent a significant degradation of the muon charged isolation efficiency, as shown in Figure A.15. Improvements from timing are also seen in studies of key HL-LHC physics analyses such as searches for $H \to \tau\tau$ in vector boson fusion events, di-Higgs production, and electroweak SUSY particle production.

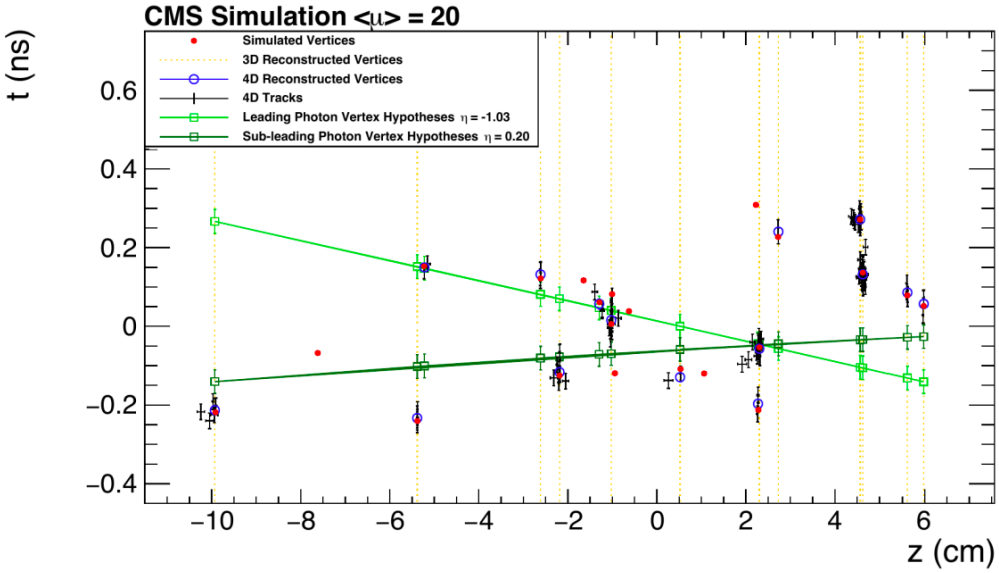


Figure A.14: Diagram illustrating a $H \rightarrow \gamma\gamma$ event analyzed with the assistance of time information. Simulated primary vertices are indicated by red dots. Black markers indicate reconstructed tracks. Yellow dashed lines and blue circles denote vertices reconstructed without and with timing information, respectively. The green lines represent vertex positions and times consistent with the two photons from the Higgs decay. The coincidence of the intersection point of the green lines and a reconstructed 4-D vertex suggests that this is the vertex of the $H \rightarrow \gamma\gamma$ event.

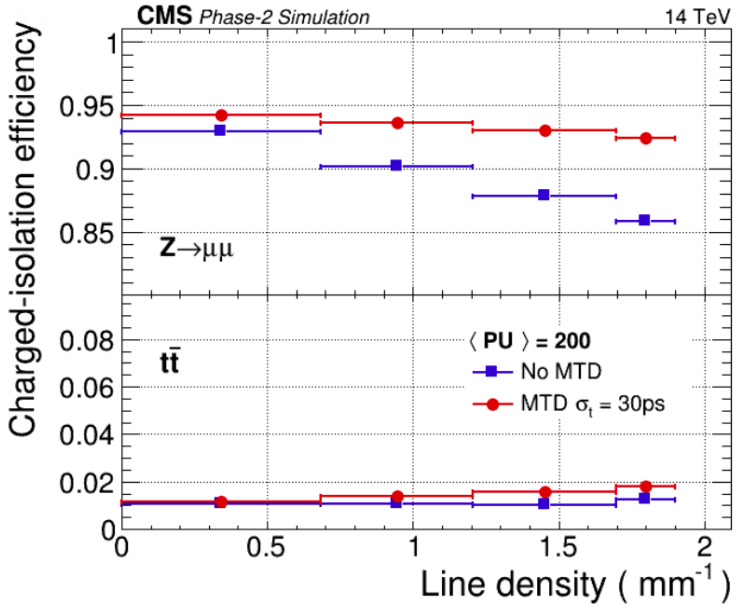


Figure A.15: Efficiency of the charged isolation requirement on reconstructed muon candidates in $Z \rightarrow \mu\mu$ events (top) and in $t\bar{t}$ events with non-prompt muons inside jets (bottom). The line density (plotted on the x-axis) is a proxy for the amount of pileup in the event. The red and blue points indicate the performance with and without time information from the MTD, respectively.

Appendix B

DEVELOPMENT OF NEURAL NETWORK PARTICLE TRACKING ALGORITHMS

In this appendix we discuss progress towards the design of a particle tracking algorithm based on an artificial neural network (NNs). We briefly describe the existing track reconstruction algorithm used in CMS, which is based on a combinatorial Kalman filter, and discuss reasons to seek a new algorithm for use in the future. We then introduce a toy track-finding problem and a NN-based algorithm that reconstructs tracks in that problem using convolutional and recurrent NNs. We conclude by describing possible future directions for this work.

This study was done in the context of the HEP.TrkX DOE pilot project, a one-year project that aims to explore machine learning driven solutions to problems in particle tracking [165].

B.1 Motivation: particle tracking at the HL-LHC

The CMS event reconstruction software identifies particle tracks in the detector using a two-stage process. In the first stage, track *seeds* are constructed from triplets of hits in the pixel tracker. Track seeds are selected based on their geometry, momentum, and perigee. The number of triplets, and hence the number of candidate seeds, scales as the cube of the number of tracker hits. Each selected seed is passed on to the second stage of reconstruction, so the overall performance of the tracking algorithm depends crucially on the ability to reject spurious triplets while maintaining a low false negative rate. The existing seeding procedure reduces the number of triplets by a factor of 10^5.

The second stage of tracking consists of a combinatorial Kalman filter (CKF) algorithm that builds full tracks from seeds [166]. Each seed from the previous stage becomes a *track candidate* whose track parameters are initialized using the seed hit information. A 3-D track is parameterized by its momentum, two angular coordinates, and two impact parameter coordinates – a total of five real numbers. Track candidates are extended in a layerwise manner as follows:

1. The current track parameters and their covariance matrix are used to extend

the track candidate to the next tracker layer. Hits in the next layer that are compatible (within some tolerance) with the track candidate are identified.

2. For each hit identified in step 1, a new track candidate is formed using that hit and all previous hits in the track. (Thus, the track candidate splits into N_{SP} candidates, where N_{SP} is the number of compatible hits in the next layer.)

3. Each new track candidate is refit using a Kalman filter to update the track parameters with the information from the new hit. Low-quality track candidates are identified and are eliminated from subsequent processing steps.

These steps are repeated for each detector layer. The complexity of the algorithm is $O(N_{SP}^2) - O(N_{SP}^3)$, depending on the details of the implementation.

Figure B.1 indicates the steps of the current CMS tracking algorithm and the order-of-magnitude multiplicities expected at each stage at the HL-LHC. The number of compatible space points N_{SP} at each tracker layer increases linearly with LHC luminosity, so it is expected that the time needed to run the existing track reconstruction software will increase quadratically or cubically with the luminosity. With anticipated HL-LHC luminosities of 5-10 times the current LHC design luminosity, the computational requirements of tracking may become prohibitive. One possible way to mitigate this is to rewrite the CKF algorithm to parallelize or vectorize some parts of the procedure [167]. This may reduce the time needed to run the algorithm but does not overcome the inherent quadratic/cubic scaling behavior. A complementary approach is to explore entirely new algorithms for track reconstruction; this is what we pursue here.

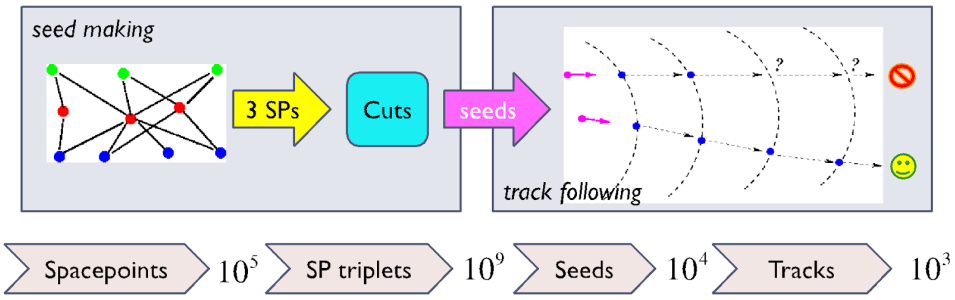


Figure B.1: Cartoon indicating the steps of the current LHC tracking algorithms and the object multiplicities expected at each stage at the HL-LHC.

B.2 A toy particle tracking problem

The full CMS tracker contains millions of channels in a three-dimensional, non-uniform layout. The KF algorithm must contend with fake and missing track hits, nuclear interactions between hadrons and the tracker material, and other effects. When developing new candidate tracking algorithms we find it more convenient to begin by working on a much simpler toy problem that does not contain the full complexity of CMS. After identifying algorithms that perform well in the toy environment, we move to progressively more realistic problems.

The most basic toy problem we consider is a two-dimensional grid of pixels, which are each either 'on' or 'off.' We assume no magnetic field, so that particle tracks are simply straight lines. Tracks are generated by uniformly sampling a start point on the left edge of the square and an end point on the right. There is no simulated noise; it is a perfectly ideal system. An example event generated in this toy framework is shown in Figure B.2.

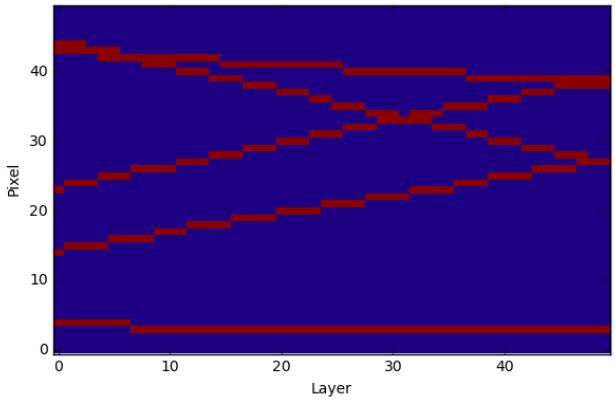


Figure B.2: Example multi-track event generated within the toy tracking framework.

B.3 Neural network architecture

We experiment with tracking algorithms based on neural networks, which have shown impressive ability to extract information from images and sequences of data [168]. NNs are flexible function approximators that consist of alternating sequences of linear transformations (the parameters of which are learned during training) and fixed nonlinear functions. A NN-based tracking algorithm could potentially improve upon the current CMS algorithm's performance in three main ways:

1. Speed up the seed-finding step or decrease the number of spurious track seeds.

2. Improve upon the combinatorial KF, making it more precise and decreasing the number of track candidates that need to be considered.

3. Change the paradigm entirely, for example by combining seed-finding and track extension into a single algorithm.

We pursue the third option first, as it has the greatest potential to produce novel ideas that may be of use later.

We implement an image-recognition NN algorithm that learns to identify tracks in the toy tracker data and produce real numbers corresponding to the parameters of each track. In the 2-D case, in the absence of a magnetic field, two parameters (slope and intercept) are sufficient to define a track. A 3-D generalization of this model could in principle do the same thing for the five track parameters described in the previous section.

The NN is implemented in Keras, a lightweight machine learning library built on the popular Theano and TensorFlow frameworks [169–171]. Keras provides a large number of modular NN building blocks (called 'layers') that can be composed to produce complex NN architectures. We provide brief descriptions of the types of NN layers used in the implementation of the track identification model:

- Input layer: the input to the NN is a two-dimensional matrix of ones and zeros corresponding to the values of the pixels in the toy tracking detector.

- Dense NN layer: takes a fixed-size input vector $\vec{v}_{\text{in}}$, multiplies it by a matrix of weights W, and adds an optional bias term $\vec{b}$:

$$\vec{v}_{\text{out}} = W\vec{v}_{\text{in}} + \vec{b}, \tag{B.1}$$

  where $\vec{v}_{\text{out}}$ is the output of the layer. The values of W and $\vec{b}$ are arbitrary and are learned during network training.

- Convolutional layer: takes a two-dimensional matrix A_{in} as input. Entries in the matrix are treated as pixel intensities in an image. The output matrix is obtained by convolving the input with a fixed $P \times Q$ filter matrix F:

$$A_{\text{out},ij} = \sum_{k=0}^{P} \sum_{l=0}^{Q} F_{kl} A_{\text{in},i+k,j+l}. \tag{B.2}$$

The convolution procedure is illustrated in Figure B.3. The input matrix can be zero-padded with $P - 1$ extra rows and $Q - 1$ extra columns so that the output matrix has the same shape as the input. The elements of F are arbitrary and are learned during network training. The convolutional layer acts as a feature extractor that detects the presence of particular patterns in the input image.

- Activation layer: applies a nonlinear function elementwise to its input. Several choices of activation function are possible; we make use of two different ones in our model:
  - Hyperbolic tangent: $f(x) = \tanh(x)$
  - Rectified linear unit (ReLU): $f(x) = x$ for $x > 0$, $f(x) = 0$ otherwise.

  Activations are typically applied after each dense or convolutional layer to add nonlinearity to the network and restrict the range of the output values.

- Max-pooling layer: coarse-grains an input matrix by a specified factor K. This is done by replacing each $K \times K$ patch of the matrix with a single value equal to the maximum entry in the patch.

- LSTM (long-short-term memory) [172]: a recurrent layer that processes a sequence $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ of input data. The LSTM has an internal state vector $\vec{h}$ and a memory vector $\vec{C}$ that are initialized to zero ($\vec{h}_0 = \vec{C}_0 = \vec{0}$). The memory vector $\vec{C}$ encodes long-term dependencies in the data, enabling the network to 'remember' past inputs longer. For $t = 1, 2, \ldots, k$, the LSTM carries out the following computation:

$$\vec{f}_t \equiv \sigma(W_f\,[\vec{h}_{t-1}, \vec{v}_t] + \vec{b}_f) \tag{B.3}$$

$$\vec{i}_t \equiv \sigma(W_i\,[\vec{h}_{t-1}, \vec{v}_t] + \vec{b}_i) \tag{B.4}$$

$$\vec{o}_t \equiv \sigma(W_o\,[\vec{h}_{t-1}, \vec{v}_t] + \vec{b}_o) \tag{B.5}$$

$$\vec{D}_t \equiv \tanh\Big(W_C\,[\vec{h}_{t-1}, \vec{v}_t] + \vec{b}_C\Big) \tag{B.6}$$

$$\vec{C}_t = \vec{f}_t \times \vec{C}_{t-1} + \vec{i}_t \times \vec{D}_t \tag{B.7}$$

$$\vec{h}_t = \vec{o}_t \times \tanh\Big(\vec{C}_t\Big). \tag{B.8}$$

Here $[\cdot, \cdot]$ denotes vector concatenation and $\times$ denotes elementwise multiplication. The LSTM transformation is parameterized by four matrices

W_f, W_i, W_C, W_o and four bias vectors $\vec{b}_f$, $\vec{b}_i$, $\vec{b}_C$, $\vec{b}_o$, which must be learned during training. The sigmoid function σ is defined by

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{B.9}$$

Sigmoid and tanh functions in the above equations are applied elementwise to their inputs. The LSTM operation is illustrated schematically in Figure B.4. It can be understood as updating the state $\vec{h}$ using information from the input $\vec{v}_t$, with additional context provided by the memory vector $\vec{C}$. The output of the LSTM can be either a fixed-length vector (equal to $\vec{h}_k$) or a sequence of vectors $(\vec{h}_1, \vec{h}_2, \ldots, \vec{h}_k)$, depending on the desired application.

An NN model is implemented as a set of layers that sequentially process the input data. For dense and LSTM layers, the size of the output vector(s) must be chosen. For convolutional layers, the number and shape of the filters must be specified. Each dense, convolutional, or LSTM layer is parameterized by one or more matrices of 'weights,' which are optimized during training in order to minimize a loss function $\mathcal{L}$. The entire network is trained end-to-end, using the backpropagation procedure [173] to compute the gradient of $\mathcal{L}$ with respect to each network parameter. The ADAM optimization algorithm [151], an adaptive version of stochastic gradient descent (SGD), is used to iteratively update the model weights until convergence.

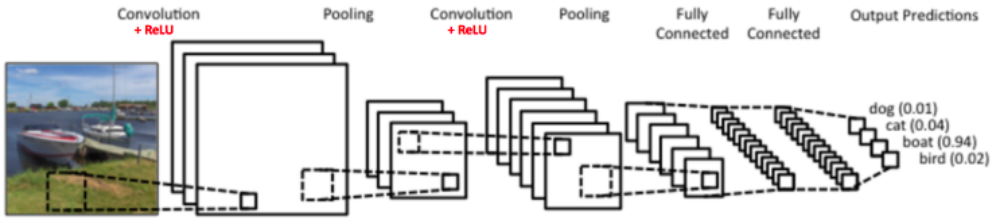


Figure B.3: Diagram illustrating the operations involved in a typical convolutional NN architecture: convolutions, pooling, and application of a dense (fully connected) NN layer [174].

Single-track model

Our NN model follows the common paradigm of alternating convolution and max-pooling layers, which progressively filter and down-sample the input to extract high-level features from it. After the final convolutional layer, the extracted features are passed through a dense layer, which has two outputs representing track

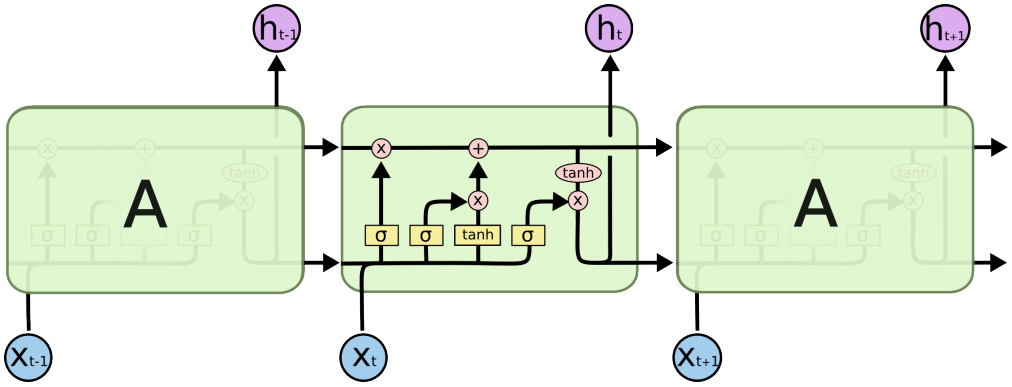


Figure B.4: Diagram of the LSTM operation, with addition, multiplication, and activation functions indicated. The memory state $\vec{C}$ is represented by the black horizontal line passing all the way through the cell [175].

parameters (slope and intercept). A diagram of the NN model is shown in Figure B.5.

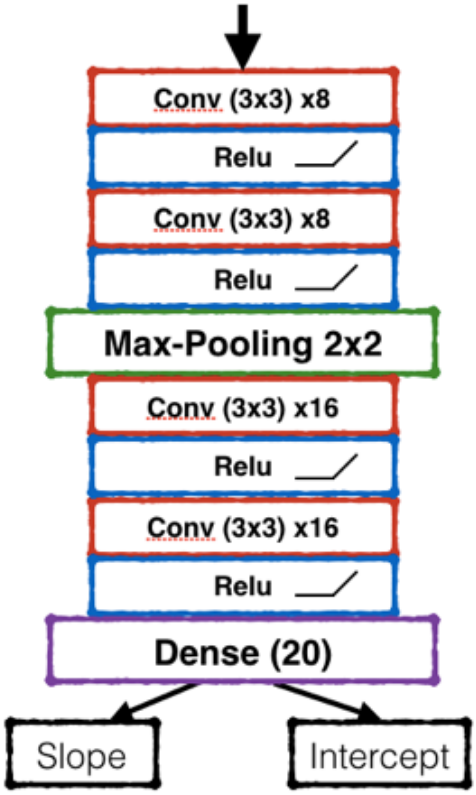


Figure B.5: Diagram of the NN architecture used to identify track parameters in single-track events.

The loss function that we optimize during training is the squared error,

$$\mathcal{L}(\vec{f}(x), \vec{y}) = \sum_{i=1}^{2} \left(\frac{y_i - f_i(x)}{L_i} \right)^2 , \tag{B.10}$$

where $\vec{y}$ and $\vec{f}(x)$ are the vectors of actual and predicted track parameters, respectively, with first and second components corresponding to the intercept and slope parameters. The input x is a matrix representing the readout of the toy detector, and $\vec{L}$ are length scales that determine the relative importance of each track parameter. Here we take $L_1 = 1$ and choose L_2 so that the slope and intercept parameters contribute equally to the loss. In each iteration of ADAM, a minibatch of B training examples is generated, and the loss $\mathcal{L}$ and its gradient are computed and averaged over all examples in the batch. The algorithm uses the gradient of $\mathcal{L}$ to update the model weights.

The single-track model is trained on events with one simulated track each. After training, the model is shown new, randomly generated test events to evaluate its performance. The average squared loss on test events is 0.016. An example event and the model's prediction are shown in Figure B.6.

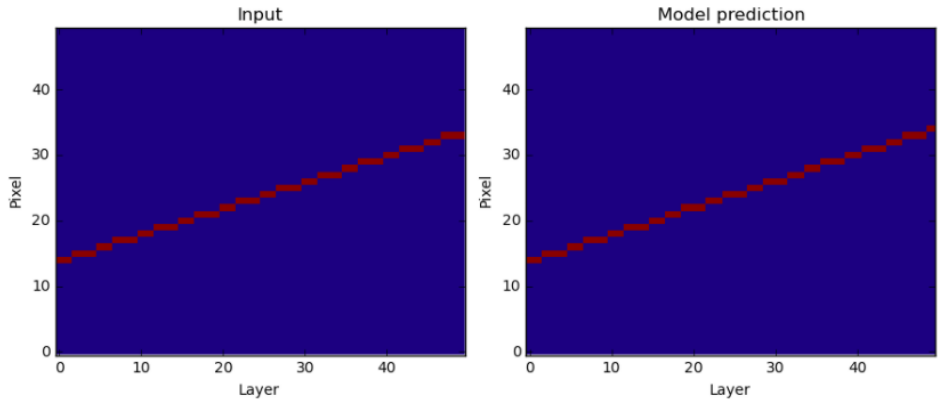


Figure B.6: Left: example single-track event generated in the toy tracker framework. Right: visualization of the model prediction for the track.

To test the robustness of the model to random noise, we retrain it using generated single-track events in which each 'off' pixel in the detector is turned on with probability p. An example event and model output are shown for $p = 0.3$ in Figure B.7. Despite the high level of noise, the model is reliably able to identify the true track.

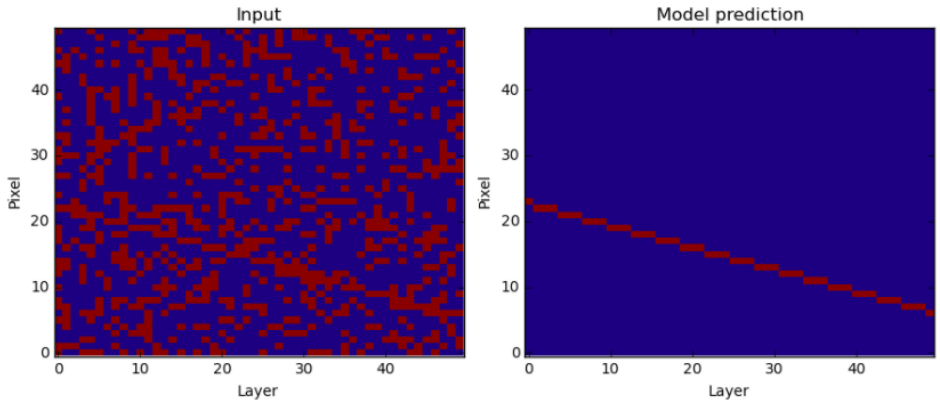


Figure B.7: Left: example single-track event generated in the toy tracker framework, with random noise added. Right: visualization of the model prediction for the track.

Multi-track model

We extend the model to handle events with k tracks by adding a recurrent (LSTM) unit after the final dense layer. The output of the final dense layer is a vector $\vec{v}_{\text{features}}$ of high-level features extracted from the input image. This vector is replicated several times and repeatedly fed into the LSTM: in the notation of Section B.3, we take $\vec{v}_1 = \vec{v}_2 = \cdots = \vec{v}_k = \vec{v}_{\text{features}}$. The LSTM produces a sequence $\vec{f}_1, \vec{f}_2, \ldots \vec{f}_k$ of output vectors, where each $\vec{f}_i$ is an (intercept, slope) pair. Intuitively, the changing memory state $\vec{C}$ of the LSTM can be thought of as a simple attention-like mechanism that focuses on different parts of the input image and remembers which tracks have already been identified. A diagram of the multi-track NN model is shown in Figure B.8. To achieve good performance, we increase the size of the dense layer from 20 to 400 hidden units. The squared error loss function is modified to include a sum over all k tracks in the event:

$$\mathcal{L}(\{\vec{f}_j(x)\}, \{\vec{y}_j\}) = \sum_{j=1}^{k} \sum_{i=1}^{2} \left(\frac{y_{ji} - f_{ji}(x)}{L_i} \right)^2, \tag{B.11}$$

where y_{ji} denotes the ith parameter of track j, and likewise for $f_{ji}(x)$.

We allow the number k of tracks to vary from event to event. The Keras framework constrains the LSTM to produce a fixed-size output tensor, so it always produces track parameters for some maximum number $k_{\max}$ of tracks. However, Keras provides a masking mechanism that we employ to exclude LSTM outputs past the kth track from the loss function computation. The model is trained using a dataset in which the number of tracks in each event is a Poisson random variable with mean 3, capped at a maximum of $k_{\max} = 6$ tracks.

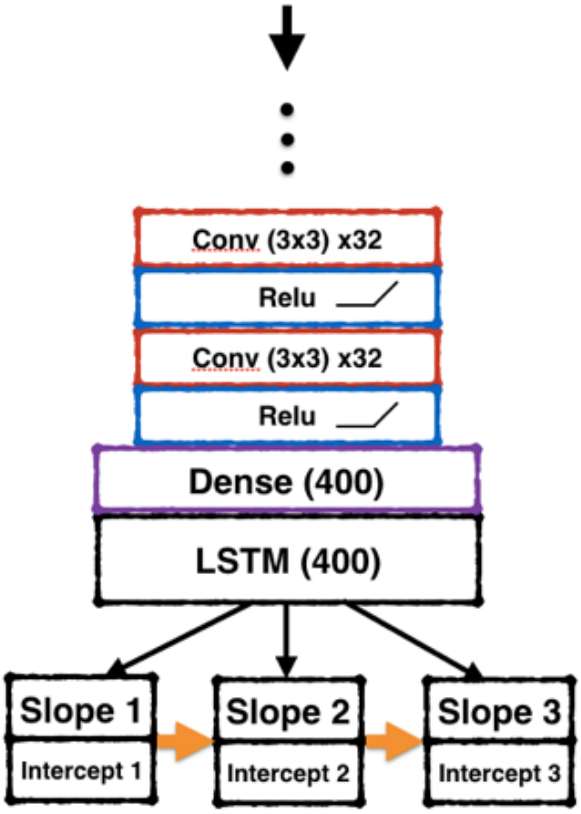


Figure B.8: Diagram of the NN architecture used to identify track parameters in multi-track events. The first few layers (not shown) are similar to those in the single-track model (Figure B.5).

After training on approximately two million toy events, the model achieves an average test loss of 1.6. The output of the model on a few example input events is shown in Figure B.9. We benchmark the model by computing the difference between the generated and reconstructed slope and intercept parameters for a large number of events. The distributions of these quantities are shown in Figure B.10.

Visualizing the convolutional filters

During model training, the image filters in each convolutional NN layer learn to extract specific patterns from the input image. Visualizing these filters can give insight into what the model has learned.

One way to do this for a given filter is to find an input image that activates the filter as much as possible. To do this, we start with an input image consisting of random pixels. We then compute the activation of the filter; in the notation of Section B.3

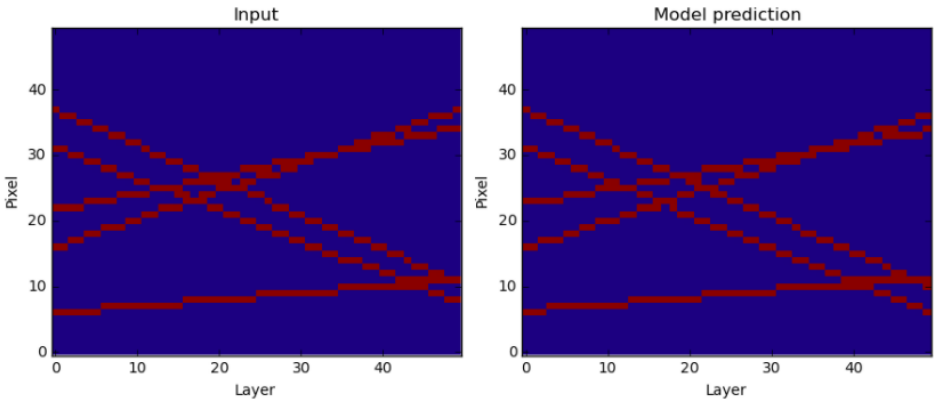


Figure B.9: Left: example multi-track event generated in the toy tracker framework. Right: visualization of the model prediction for the tracks.

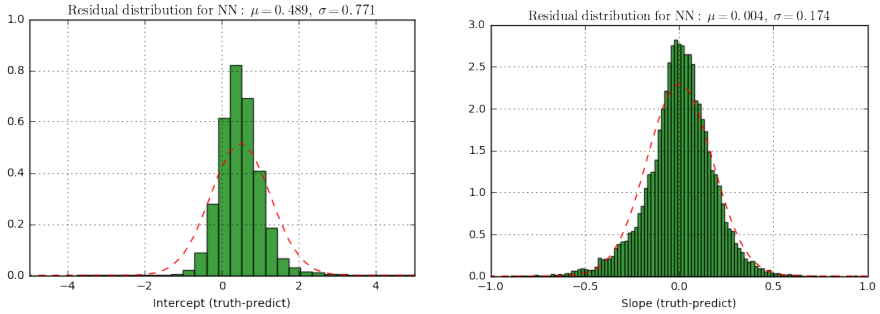


Figure B.10: Difference between actual and predicted intercept (left) and slope (right) parameters in events with up to six tracks for the model shown in Figure B.8 [165].

this is

$$\sum_{i,j} A_{\text{out},ij}. \tag{B.12}$$

We compute the gradient of this quantity with respect to the pixel values of the input image, and perform gradient ascent in this space to find the image that maximizes the activation. This image illustrates the pattern that most strongly triggers the convolutional filter.

Example images generated in this way for some convolutional filters in the NN are shown in Figure B.11. We see several filters that appear to be optimized for finding track intersections, and others which appear to focus on the left edge of the image, where track intercepts can be estimated.

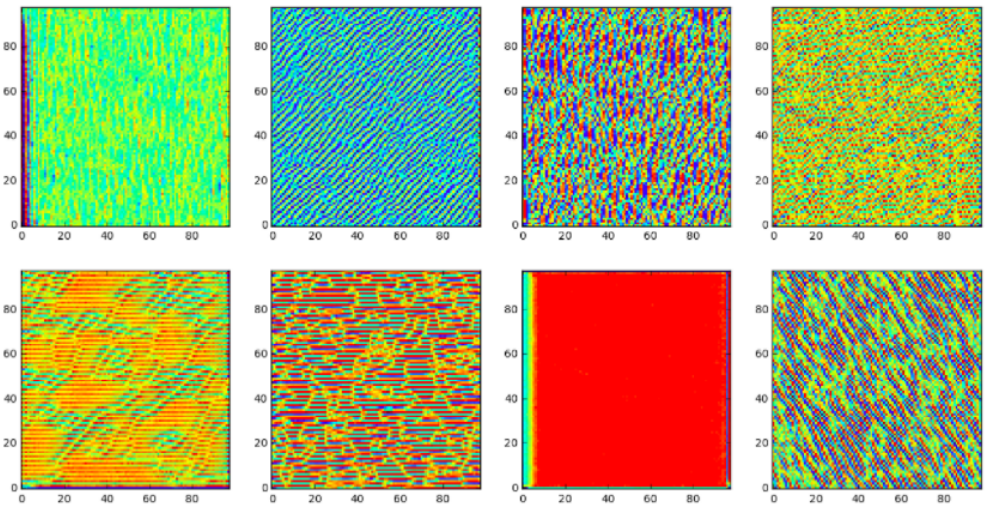


Figure B.11: Images optimized for activation of specific filters in the multi-track NN model.

B.4 Uncertainty quantification

It is important to obtain well-calibrated estimates of the uncertainties on the track parameters predicted by the model. After reviewing existing techniques for quantifying the uncertainty on the outputs of machine learning algorithms, we choose an approach inspired by the photon energy regression used in CMS [176]. In this approach the model simply has an additional output or outputs that represent the uncertainty on its predictions. For the multi-track NN discussed above, this means that the model should have three additional outputs for each track that parameterize the 2×2 covariance matrix Σ of the track parameters. The squared loss function Eq. B.11 is replaced by a gaussian log likelihood:

$$\mathcal{L}(\{\vec{f}_j(x)\},\{\vec{y}_j\}) = \sum_{j=1}^{k} \log\left|\Sigma_j\right| + (\vec{y}_j - \vec{f}_j(x))^T \Sigma_j^{-1} (\vec{y}_j - \vec{f}_j(x)). \tag{B.13}$$

A diagram of the modified model is shown in Figure B.12.

To achieve a low score on this loss function, the model must output track parameters that match the true ones, and also output a covariance matrix that faithfully reflects the accuracy of those track parameters.

Performance

The output of the model on a few sample input events is shown in Figure B.13. For each track identified by the model, we sample from the track parameter covariance

Figure B.12: Diagram of the NN architecture used to identify track parameters and uncertainties in multi-track events.

matrix several times. We plot the samples for each track to demonstrate the model's level of certainty. Greater spread in the sampled track predictions corresponds to larger uncertainty on the track.

To quantify how well-calibrated the uncertainties are, we calculate the distribution of the *Mahalanobis distance*,

$$D_M(\vec{f}(x), \vec{y}) = \sqrt{(\vec{y} - \vec{f}(x))^T \Sigma^{-1} (\vec{y} - \vec{f}(x))}. \tag{B.14}$$

For an n-dimensional normal random variable, D_M^2 has a chi-square distribution with n degrees of freedom. We compare the distribution of Mahalanobis distances with a chi-square using a probability plot, which displays the quantiles of the chi-square distribution on the x-axis and those of the empirical distribution of D_M^2 on the y-axis.

The results are shown in Figure B.14 for single- and multi-track models. For the single-track model we see that the probability plot is nearly linear, albeit with a slope that is different from one. For the multi-track model the plot has a slight curve. Both of these features can be easily fixed through a small deterministic calibration to produce a probability plot with the correct shape (a line with slope 1).

B.5 Challenges and future work

We encounter two main challenges when attempting to scale the above model to larger, more realistic tracking datasets:

- Most trackers used in HEP have non-uniform geometry. For example, the CMS tracker barrel and endcaps have different geometrical layouts and different numbers of pixels in each layer. The model considered above is restricted to rectangular grid-shaped detectors or those on which a grid geometry can be imposed.

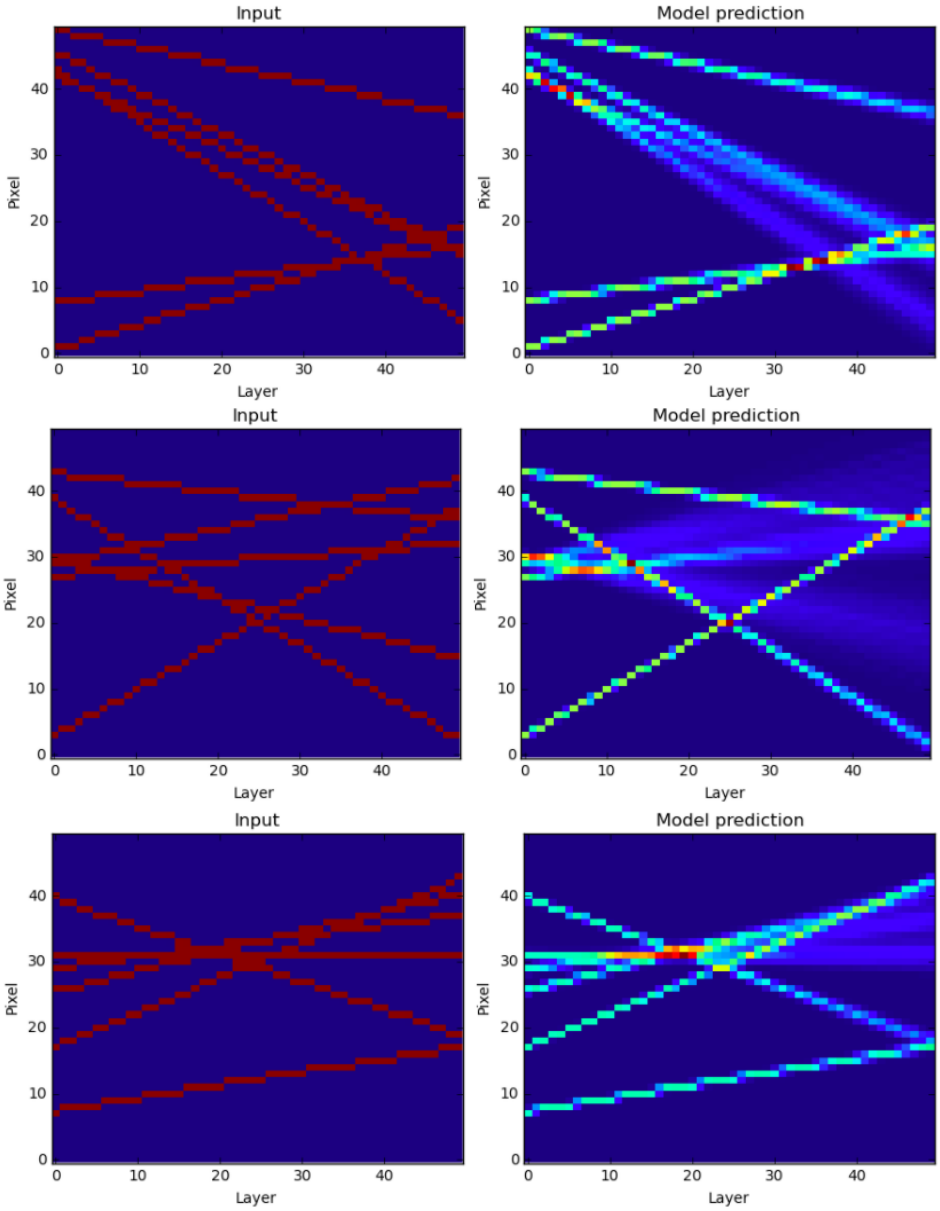


Figure B.13: Left column: example multi-track events generated in the toy tracker framework. Right column: visualization of the model predictions and uncertainties for the tracks.

- As the size of the detector increases, and as the dimensionality increases from two to three, the required NN size and depth increase, which in turn demands more computational resources and longer training times.

To deal with the first issue, it is necessary to build a NN that operates on non-grid input layouts. Recent work on graph convolutional networks may be useful here [177]. The second issue is one of practicality; the space of possible NN archi-

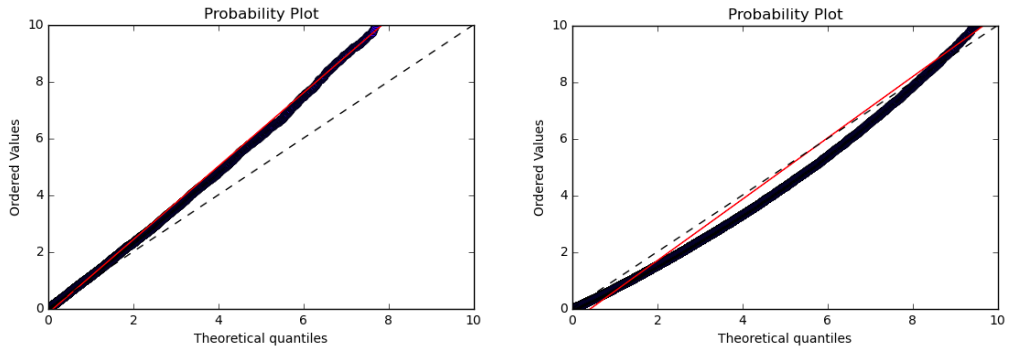


Figure B.14: Probability plots comparing a chi-square distribution with two degrees of freedom with the distribution of the Mahalanobis distance (Eq. B.14) between the predicted and actual track parameters. The x-axis indexes the quantiles of the chi-square distribution with two degrees of freedom. The y-axis shows the corresponding quantiles of the empirical D_M^2 distribution. Left: single-track events. Right: multi-track events.

tectures is enormous and the time needed to find one that works for a problem of given size is large. If the NN is too large, it may not fit in GPU memory, in which case the model needs to be distributed across multiple GPUs. Convolutions in 3-D are considerably more compute-intensive than those in 2-D, and a dedicated 3-D convolution algorithm that takes advantage of the sparsity of the problem is needed.

Several other avenues of investigation are possible and have been studied by others in the HEP.TrkX project [165]. Another approach that has been seen to work well is to encode the tracker hit information as sequences of hit locations, rather than as images. The event size in this case is proportional to the detector hit occupancy, rather than the number of pixels in the detector. LSTMs are used to encode the list of hits into fixed-size vector representations, which can then be used to perform layerwise track extension or to assign hits to track candidates.

Appendix C

DISTRIBUTED MACHINE LEARNING IN A SUPERCOMPUTING CLUSTER SETTING

In this appendix we discuss the large-scale distributed training of neural network models for high-energy physics applications, and provide details on a software framework we implemented for this purpose. The framework, *mpi_learn*, is available online [1]. Its purpose is to increase the speed and scope of NN training to enable rapid prototyping of large models used for research.

C.1 Motivation: large-scale training of neural networks

Neural networks are commonly trained using stochastic gradient descent (SGD) or one of its variants. These algorithms compute gradients of a loss function $\mathcal{L}$ on batches of training data and iteratively update the model parameters $\vec{w}$ using the gradient information. The backpropagation procedure [173] allows efficient computation of the derivative of the loss function with respect to any given model parameter. In the basic SGD algorithm, the weights $\vec{w}$ are updated along the direction of the gradient:

$$\vec{w} \to \vec{w} - \eta \, \nabla_{\vec{w}} \mathcal{L}. \tag{C.1}$$

The learning rate parameter η controls the size of the updates. The above procedure is iterated until a minimum of the loss function is found.

The rate of training is limited by the time needed to compute the gradient of the loss function. Most machine learning software libraries are capable of exploiting the parallel computing capabilities of a graphics processing unit (GPU), which can perform the needed matrix multiplication and addition operations much more quickly than an ordinary CPU. This can dramatically decrease the amount of time needed for model training.

If one has access to several compute nodes with GPUs, or many GPUs on one node, training can be further accelerated using distributed training algorithms. These algorithms, discussed in Section C.3, require processes or nodes to share data with one another. Often one node (the 'master') coordinates the actions of several 'worker'

[1] https://github.com/duanders/mpi_learn

nodes, which perform the gradient computations needed for SGD. Scientific researchers may have access to supercomputing clusters with hundreds of compute nodes, which may be equipped with GPUs. Such systems commonly feature job submission queues built on systems such as SLURM [178]; a single job may request and receive access to many nodes on the supercomputer.

We designed a software framework suitable for submission at supercomputing sites that coordinates distributed NN training across a potentially large number of compute nodes. The framework is implemented using the Message Passing Interface (MPI) [179], which is a commonly used high-level protocol for communication between software processes. The code is written in Python using the mpi4py package [180]. It is built on the popular Keras [169] machine learning library, which offers a modular interface for building and training NNs.

The MPI protocol is agnostic to the underlying hardware, so the same set of function calls can be used for communication between processes on the same node and for communication between nodes. Most supercomputing clusters support MPI and use it as a primary protocol for running jobs.

As the use of NNs becomes more prevalent in the sciences, convenient access to distributed training algorithms becomes increasingly important. We designed the *mpi_learn* library to be easily usable by anyone who can implement a NN model in Keras. It is ideal for the use of non-specialists who have access to high performance computing resources and seek to use them efficiently to train NNs.

C.2 Related work

The *mpi_learn* package was written in the summer of 2016 to address a perceived lack of user-friendly distributed training software for NNs. In particular, we sought an implementation that would easily interface with the Keras library and be suitable for use in a supercomputing cluster context. This package, written within the MPI framework, was developed concurrently with similar work on running distributed training of Keras models with Spark [181].

The Horovod package, released by Uber in late 2017, provides a convenient interface for training Keras models using MPI and fulfills much the same purpose as *mpi_learn* [182]. Very recently, features similar to those provided by our package were added to the PyTorch machine learning framework [183].

We do not claim that the *mpi_learn* framework is better than any other framework.

This package was written for practical reasons in the observed absence of other tools fulfilling the same purpose.

C.3 Distributed training algorithms

The SGD algorithm finds a minimum of the loss function $\mathcal{L}$ by a two-step iterative process. In step one, a 'minibatch' of B events is sampled from the available training dataset. The gradient of the loss function with respect to each model parameter, $\nabla_{\vec{w}}\mathcal{L}$, is computed on the minibatch. In step two, the model parameters are updated according to Eq. C.1. The algorithm terminates when a suitable stopping criteria is met, for example when the loss function's change is within some tolerance of zero.

Variants of SGD are designed to improve the convergence of the algorithm. Often the learning rate is made to decrease according to a fixed schedule or in response to the behavior of the loss. This avoids the situation where the learning rate is too large and the model parameters oscillate around the minimum of the loss forever without reaching the lowest point. Other methods assign a 'momentum' to gradient updates, encouraging each weight update to be in the same direction as the previous one [184]. Adaptive gradient methods such as RMSProp and ADAM assign a different, changing learning rate to each model parameter [151]. They are attractive because they require little to no hyperparameter tuning; however, recent work has pointed out that they sometimes produce models that do not generalize well to new data [185].

SGD and its variants are inherently sequential algorithms: they require computing the gradient on one minibatch of data and updating the model parameters before proceeding to the next minibatch. However, a number of distributed training algorithms have been designed that can efficiently minimize the loss function when data is divided among several compute nodes running in parallel.

Downpour SGD

In the Downpour SGD algorithm [186], one node is designated the 'master' and the others are 'workers' (see Figure C.1). The master node contains a central copy of the NN model to be trained. The worker nodes each have their own copy of the model and a fraction of the available training data.

Training proceeds in an asynchronous manner, with each worker performing the following steps:

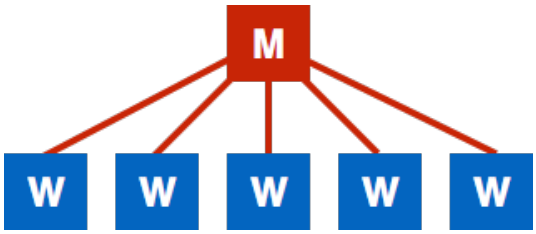


Figure C.1: Training configuration for the distributed learning algorithms considered in this chapter, consisting of a 'master' node and several 'worker' nodes that communicate with it.

1. Sample a minibatch of training data and compute the gradient of the loss function, $\nabla_{\vec{w}}\mathcal{L}$.
2. Send the gradient to the master node.
3. Wait for the master to apply the gradient update and reply with the updated model parameters $\vec{w}$.
4. Update the worker's copy of the model with the new $\vec{w}$, then go back to Step 1.

These steps are shown schematically in Figure C.2. The master node does not perform any gradient computations; it simply receives the gradients computed by the worker nodes and applies them in the order they are received. The master node can perform standard SGD updates, adaptive gradient updates, or any other desired update scheme, using the gradients that it receives.

The advantage of this asynchronous approach to master-worker communication is that performance does not suffer if some nodes are slower than others. Faster nodes will simply send gradients more frequently than slower ones; there is no need to wait for the slow nodes to catch up. This contrasts with synchronous SGD algorithms, in which the master receives gradient information from all worker nodes simultaneously. In this situation, gradient updates occur at the speed of the slowest node in the ensemble.

The Downpour SGD algorithm supports model parallelism, in which the NN model itself is divided into pieces and stored across multiple nodes. The use of multiple master nodes, each handling a different subset of the model parameters, can reduce the communication bottleneck between workers and masters. Unfortunately, model parallelism is not implemented in our current software framework.

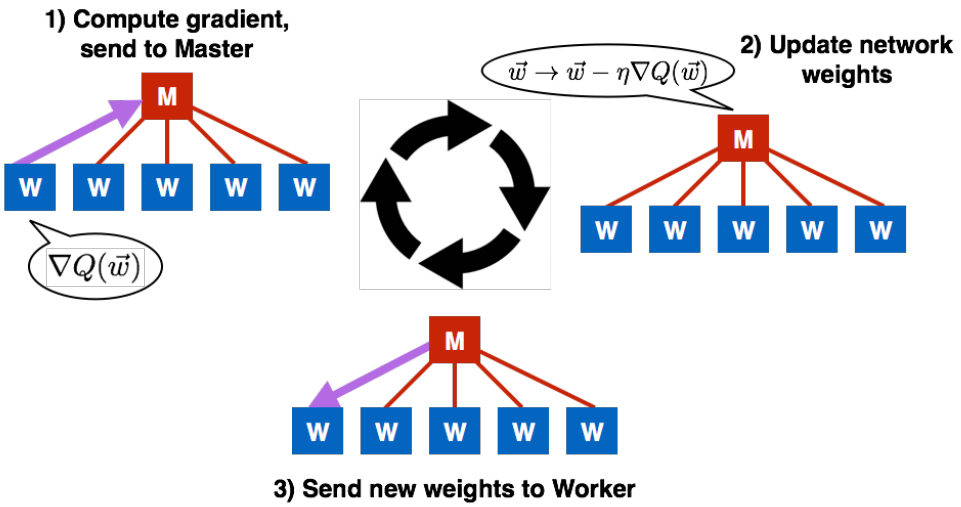


Figure C.2: Schematic representation of the Downpour SGD algorithm.

Elastic Averaging SGD

The Elastic Averaging SGD algorithm also features a master node that exchanges information with many worker nodes. In this algorithm, worker nodes perform SGD (or another sequential learning procedure) independently of one another. Every few iterations, an 'elastic force' pulls the worker's and master's model weights towards one another:

$$\vec{w}_{\text{worker}\,i} \to \vec{w}_{\text{worker}\,i} - \alpha(\vec{w}_{\text{worker}\,i} - \vec{w}_{\text{master}}) \tag{C.2}$$

$$\vec{w}_{\text{master}} \to \vec{w}_{\text{master}} + \alpha(\vec{w}_{\text{worker}\,i} - \vec{w}_{\text{master}}). \tag{C.3}$$

This elastic update is applied asynchronously to one worker node at a time.

A schematic depiction of Elastic Averaging SGD is shown in Figure C.3. The algorithm allows worker nodes more individual freedom to explore the space of model parameters. This way they may discover lower minima of the loss function, leading to better model performance. Elastic Averaging SGD is also communication efficient compared with the Downpour algorithm, because workers do not need to synchronize with the master after processing each minibatch [187].

C.4 An MPI-based training framework

The *mpi_learn* framework implements the Downpour and Elastic Averaging SGD algorithms described in Section C.3. It is written in Python, and is designed for use with Keras. Its features are:

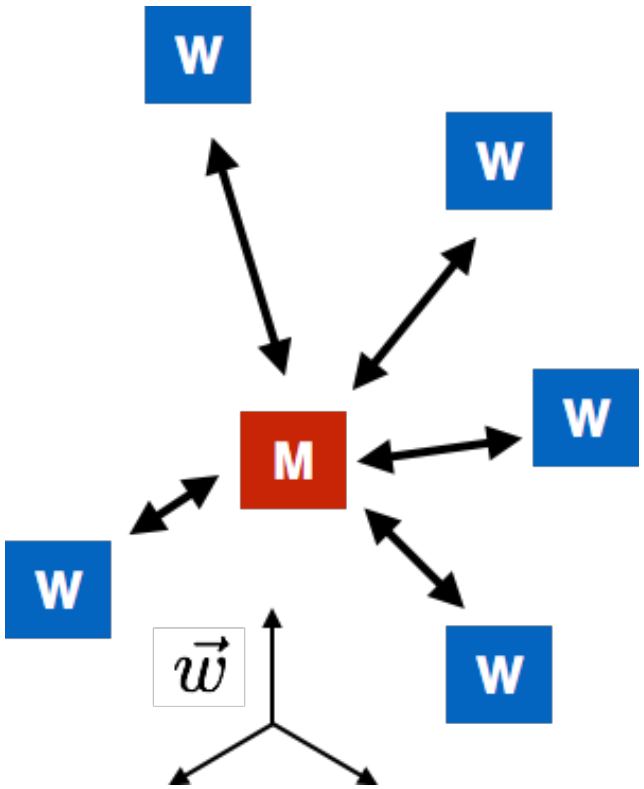


Figure C.3: Schematic representation of the Elastic Averaging SGD algorithm.

- Support for training arbitrary Keras models. The user supplies a Python class containing instructions for building the Keras model, or a JSON file specifying the model architecture.
- Support for several popular SGD weight update algorithms, including ADAM and RMSProp.
- A flexible system for loading training data that supports arbitrary Python data generators.
- Support for most Keras callback functions.

C.5 Benchmarking

We benchmark the performance of the *mpi_learn* library on two systems:

- A Supermicro server with 28 cores and eight NVidia GTX1080 GPUs. Communication between processes is accomplished via shared memory, as all processes are on the same node.

- The ALCF Cooley GPU cluster, with 126 nodes, each having 16 cores and 1 NVidia K80 GPGPU. Nodes are interconnected with FDR Infiniband [188].

The *mpi_learn* framework is used to train a recurrent neural network to classify collision events in CMS. The training dataset is created using the Delphes simulation framework [189]. The model consists of an LSTM network [172] with 20 hidden units, followed by a softmax output over three different categories of collision events. The input data consists of 100 files of 9500 samples each, totaling 50GB. This model takes several hours to train on a node with a single GPU. We use the Downpour SGD algorithm for all experiments described here. The purpose of this study is not to evaluate the performance of the model [190], but rather to evaluate how much faster this model can be trained when multiple GPUs are utilized.

The model is trained several times with various numbers of worker processes, using a batch size of 100 samples. The data in the training set is divided evenly among all workers. Training continues until each worker has processed its training data a fixed number of times (ten, in this case).

Validation of the model's accuracy is performed by the master process using a held-out test set. Validation can be a bottleneck in the training process because it is performed serially; the frequency of validation can be adjusted as needed to minimize its impact on the total training time.

For each training run, the speedup is computed with respect to the time taken by *mpi_learn* using a single worker process. Results are shown in figures C.4 and C.5.

The time needed to train the model with *mpi_learn* and a single worker process is also compared to the training time obtained using Keras alone. The times are similar, indicating that the training overhead from the *mpi_learn* framework itself is small.

For up to 10 worker processes, the speedup is roughly linear with the number of workers. This indicates that the training framework can fully exploit the resources of a multi-GPU node such as the Supermicro server used here.

The speedup deviates from linearity with increasing number of workers. For 60 worker nodes, we observe a speedup of 30 with respect to the nominal training time for this choice of batch size. The deviation from linearity is driven by the time needed for the master process to update the weights of the network and transmit them back to the workers. Because the frequency of weight updates is inversely

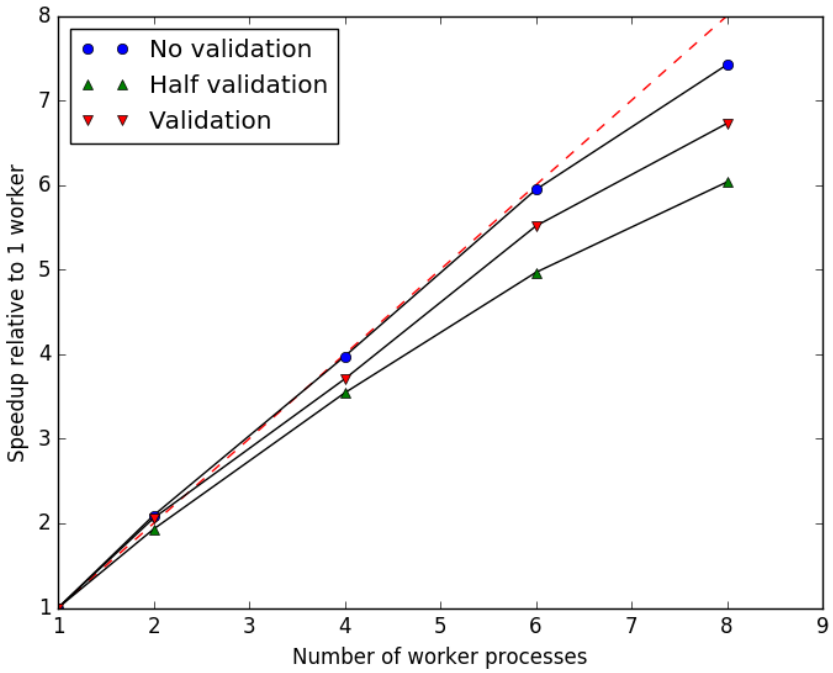


Figure C.4: Training speedup on the Supermicro server with 8 GPUs, as a function of the number of workers used for training, with a batch size of 100 samples. The red diagonal indicates 1:1 speedup.

proportional to the batch size, increasing the batch size can alleviate this bottleneck and speed up the training procedure, as shown in Table C.1 for the example of 20 worker processes. The larger the amount of validation data, the earlier the linear speedup with the number of nodes will break, because of the constant amount of time spent in validation per epoch.

Table C.1: Training speedup obtained with various batch sizes, with respect to a batch size of 100, with 20 workers training the benchmark model.

Batch Size	Speedup
10	0.1
100	1.0
500	3.0
1000	4.1

Model accuracy

As shown in figure C.6, the model accuracy degrades with increased number of workers. This occurs because of the so-called stale gradient issue: in an asynchronous training setting, workers may compute gradient updates using outdated

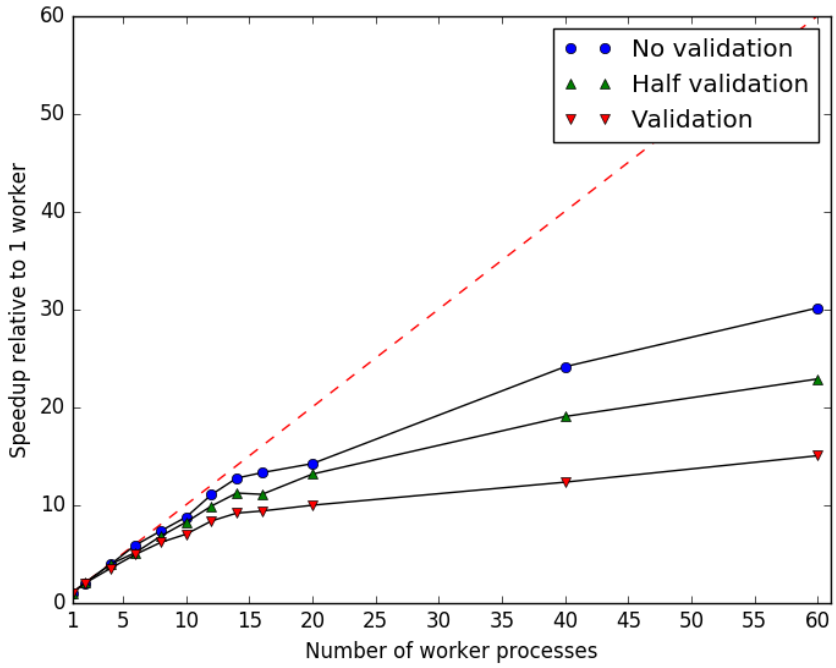


Figure C.5: Training speedup for the benchmark model on the ALCF Cooley cluster with 1 GPU per node, as a function of the number of workers used for training, using a batch size of 100 samples. The red diagonal indicates 1:1 speedup.

model parameters, because the worker's model parameters $\vec{w}_{\text{worker}}$ may not be identical with those of the master. Training on outdated model parameters produces suboptimal gradient updates, which reduces model performance. The issue can be mitigated by a suitable choice of SGD momentum [191], or by moving to an algorithm with synchronous gradient updates.

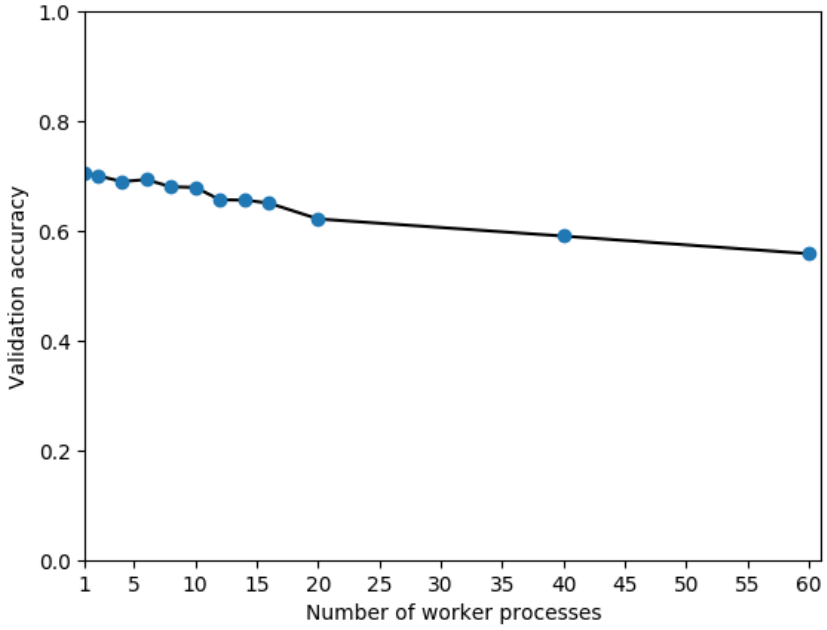


Figure C.6: Model accuracy after 10 training epochs as a function of the number of workers used. The model performance slowly decreases at high worker counts because of workers training on outdated model information.

BIBLIOGRAPHY

[1] CMS Collaboration. *Summary of CMS Cross Section Measurements*. 2013. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined (visited on 01/30/2018).

[2] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021. URL: http://www.sciencedirect.com/science/article/pii/S0370269312008581.

[3] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020. URL: http://www.sciencedirect.com/science/article/pii/S037026931200857X.

[4] CMS Collaboration. *Summary of Comparison Plots in Simplified Models Spectra for the 8TeV Dataset*. 2013. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SUSYSMSSummaryPlots8TeV (visited on 01/30/2018).

[5] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. URL: https://cds.cern.ch/record/782076.

[6] Fabienne Marcastel. "CERN's Accelerator Complex. La chaîne des accélérateurs du CERN". In: (Oct. 2013). General Photo. URL: https://cds.cern.ch/record/1621583.

[7] *Measurement of the inelastic proton-proton cross section at* $\sqrt{s}$ = 13 TeV. Tech. rep. CMS-PAS-FSQ-15-005. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2145896.

[8] M Aaboud et al. "Measurement of the Inelastic Proton-Proton Cross Section at s = 13 TeV with the ATLAS Detector at the LHC". In: 117 (Oct. 2016).

[9] CMS Collaboration. *CMS Luminosity Public Results*. 2010. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (visited on 01/31/2018).

[10] G L Bayatian et al. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. There is an error on cover due to a technical problem for some items. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/922757.

[11] Tommaso Dorigo. "Recent Results of the CMS Experiment". In: *EPJ Web Conf.* 70 (2014), p. 00021. DOI: 10.1051/epjconf/20147000021. arXiv: 1210.6809 [hep-ex].

[12] David Barney. "CMS Detector Slice". CMS Collection. Jan. 2016. URL: https://cds.cern.ch/record/2120661.

[13] V. Halyo, P. LeGresley, and P. Lujan. "Massively Parallel Computing and the Search for Jets and Black Holes at the LHC". In: *Nucl. Instrum. Meth.* A744 (2014), pp. 54–60. DOI: 10.1016/j.nima.2014.01.038. arXiv: 1309.6275 [physics.comp-ph].

[14] B Ille. "The Compact Muon Solenoid (CMS) electromagnetic calorimeter". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 453.1 (2000). Proc. 7th Int. Conf on Instrumentation for colliding Beam Physics, pp. 210–217. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(00)00632-X. URL: http://www.sciencedirect.com/science/article/pii/S016890020000632X.

[15] Adolf Bornheim. "The CMS Ecal Laser Monitoring System". In: *Astroparticle, particle and space physics, detectors and medical physics applications. Proceedings, 9th ICATPP Conference, Como, Italy, October 17-21, 2005*. 2006, pp. 318–322. DOI: 10.1142/9789812773678_0051.

[16] "CMS ECAL Laser monitoring up to 2017, $\pi^0/\eta \to \gamma\gamma$ spectrum and monitoring, ES calibration". In: (July 2017). URL: https://cds.cern.ch/record/2273269.

[17] Ganna Dolinska. "Measurement of the Cross Sections of $t\bar{t}$ Production in the Boosted Regime with the CMS Detector at $\sqrt{s} = 8$ TeV". PhD thesis. Hamburg: Hasylab, DESY, 2016. DOI: 10.3204/PUBDB-2016-03989. URL: http://bib-pubdb1.desy.de/search?cc=Publication+Database&of=hd&p=reportnumber:DESY-THESIS-2016-026.

[18] Vardan Khachatryan et al. "The CMS trigger system". In: *JINST* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[19] CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (2008), S08004. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08004.

[20] A Tapper and Darin Acosta. *CMS Technical Design Report for the Level-1 Trigger Upgrade*. Tech. rep. CERN-LHCC-2013-011. CMS-TDR-12. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch. June 2013. URL: https://cds.cern.ch/record/1556311.

[21] Clint Richardson. "CMS High Level Trigger Timing Measurements". In: *Journal of Physics: Conference Series* 664.8 (2015), p. 082045. URL: http://stacks.iop.org/1742-6596/664/i=8/a=082045.

[22] Mia Tosi. *The CMS trigger in Run 2*. Tech. rep. CMS-CR-2017-340. Geneva: CERN, Oct. 2017. URL: https://cds.cern.ch/record/2290106.

[23] Albert M Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". Submitted to JINST. 2017.

[24] "Electromagnetic calorimeter calibration with 7 TeV data". In: (2010).

[25] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12.10 (2017), P10003. URL: http://stacks.iop.org/1748-0221/12/i=10/a=P10003.

[26] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (2014), P10009. URL: http://stacks.iop.org/1748-0221/9/i=10/a=P10009.

[27] W Adam et al. "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC". In: *Journal of Physics G: Nuclear and Particle Physics* 31.9 (2005), N9. URL: http://stacks.iop.org/0954-3899/31/i=9/a=N01.

[28] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti- k t jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (2008), p. 063. URL: http://stacks.iop.org/1126-6708/2008/i=04/a=063.

[29] Ariel Schwartzman. "Jet energy calibration at the LHC". In: *Int. J. Mod. Phys.* A30.31 (2015), p. 1546002. DOI: 10.1142/S0217751X15460021. arXiv: 1509.05459 [hep-ex].

[30] Vardan Khachatryan et al. "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s}$ = 8 TeV". In: *JINST* 10.02 (2015), P02006. DOI: 10.1088/1748-0221/10/02/P02006. arXiv: 1411.0511 [physics.ins-det].

[31] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[32] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "PYTHIA 6.4 Physics and Manual". In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph].

[33] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[34] Johan Alwall et al. “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”. In: *Eur. Phys. J.* C53 (2008), pp. 473–500. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph].

[35] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[36] Rahmat Rahmat, Rob Kroeger, and Andrea Giammanco. “The Fast Simulation of The CMS Experiment”. In: *Journal of Physics: Conference Series* 396.6 (2012), p. 062016. URL: http://stacks.iop.org/1742-6596/396/i=6/a=062016.

[37] Valery Rubakov and Stephen S. Wilson. *Classical Theory of Gauge Fields*. Princeton University Press, 2002. ISBN: 978-0691059273.

[38] Mark Srednicki. *Quantum Field Theory*. Cambridge University Press, 2007. ISBN: 9780521864497.

[39] I. Gil-Botella. “Neutrino Physics”. In: *Proceedings, 6th CERN - Latin-American School of High-Energy Physics (CLASHEP 2011): Natal, Brazil , March 23 - April 5, 2011*. 2013, pp. 157–205. DOI: 10.5170/CERN-2013-003.157. arXiv: 1504.03551 [hep-ph]. URL: https://inspirehep.net/record/1359452/files/arXiv:1504.03551.pdf.

[40] C. Patrignani et al. “Review of Particle Physics”. In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.

[41] M. Herrero. “The Standard model”. In: *NATO Sci. Ser. C* 534 (1999), pp. 1–59. DOI: 10.1007/978-94-011-4689-0_1. arXiv: hep-ph/9812242 [hep-ph].

[42] S. Dawson. “Introduction to electroweak symmetry breaking”. In: *Proceedings, Summer School in High-energy physics and cosmology: Trieste, Italy, June 29-July 17, 1998*. 1998, pp. 1–83. arXiv: hep-ph/9901280 [hep-ph].

[43] Stephen P. Martin. “A Supersymmetry primer”. In: (1997). [Adv. Ser. Direct. High Energy Phys.18,1(1998)]. DOI: 10.1142/9789812839657_0001,10.1142/9789814307505_0001. arXiv: hep-ph/9709356 [hep-ph].

[44] Porter Williams. “Naturalness, the autonomy of scales, and the 125GeV Higgs”. In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 51 (2015), pp. 82–96. ISSN: 1355-2198. DOI: https://doi.org/10.1016/j.shpsb.2015.05.003. URL: http://www.sciencedirect.com/science/article/pii/S1355219815000416.

[45] Rudolf Haag, Jan T. Lopuszanski, and Martin Sohnius. "All Possible Generators of Supersymmetries of the s Matrix". In: *Nucl. Phys.* B88 (1975), p. 257. DOI: 10.1016/0550-3213(75)90279-5.

[46] Sidney Coleman and Jeffrey Mandula. "All Possible Symmetries of the *S* Matrix". In: *Phys. Rev.* 159 (5 July 1967), pp. 1251–1256. DOI: 10.1103/PhysRev.159.1251. URL: https://link.aps.org/doi/10.1103/PhysRev.159.1251.

[47] Pierre Fayet. "The Supersymmetric Standard Model". In: *Adv. Ser. Direct. High Energy Phys.* 26 (2016), pp. 397–454. DOI: 10.1142/9789814733519_0020. arXiv: 1506.08277 [hep-ph].

[48] Ian J. R. Aitchison. "Supersymmetry and the MSSM: An Elementary introduction". In: (2005). arXiv: hep-ph/0505105 [hep-ph].

[49] D. J. H. Chung et al. "The Soft supersymmetry breaking Lagrangian: Theory and applications". In: *Phys. Rept.* 407 (2005), pp. 1–203. DOI: 10.1016/j.physrep.2004.08.032. arXiv: hep-ph/0312378 [hep-ph].

[50] Csaba Csaki. "The Minimal supersymmetric standard model (MSSM)". In: *Mod. Phys. Lett.* A11 (1996), p. 599. DOI: 10.1142/S021773239600062X. arXiv: hep-ph/9606414 [hep-ph].

[51] Vardan Khachatryan et al. "Phenomenological MSSM interpretation of CMS searches in pp collisions at sqrt(s) = 7 and 8 TeV". In: *JHEP* 10 (2016), p. 129. DOI: 10.1007/JHEP10(2016)129. arXiv: 1606.03577 [hep-ex].

[52] M. Drees and Jong Soo Kim. "Minimal natural supersymmetry after the LHC8". In: *Phys. Rev.* D93.9 (2016), p. 095005. DOI: 10.1103/PhysRevD.93.095005. arXiv: 1511.04461 [hep-ph].

[53] Ulrich Ellwanger, Cyril Hugonie, and Ana M. Teixeira. "The Next-to-Minimal Supersymmetric Standard Model". In: *Phys. Rept.* 496 (2010), pp. 1–77. DOI: 10.1016/j.physrep.2010.07.001. arXiv: 0910.1785 [hep-ph].

[54] Dan Hooper and Tilman Plehn. "Supersymmetric dark matter: How light can the LSP be?" In: *Phys. Lett.* B562 (2003), pp. 18–27. DOI: 10.1016/S0370-2693(03)00548-3. arXiv: hep-ph/0212226 [hep-ph].

[55] Ryuichiro Kitano and Yasunori Nomura. "Supersymmetry, naturalness, and signatures at the LHC". In: *Phys. Rev.* D73 (2006), p. 095004. DOI: 10.1103/PhysRevD.73.095004. arXiv: hep-ph/0602096 [hep-ph].

[56] Michele Papucci, Joshua T. Ruderman, and Andreas Weiler. "Natural SUSY Endures". In: *JHEP* 09 (2012), p. 035. DOI: 10.1007/JHEP09(2012)035. arXiv: 1110.6926 [hep-ph].

[57] Serguei Chatrchyan et al. "Interpretation of Searches for Supersymmetry with simplified Models". In: *Phys. Rev.* D88.5 (2013), p. 052017. DOI: 10.1103/PhysRevD.88.052017. arXiv: 1301.2175 [hep-ex].

[58] Vardan Khachatryan et al. "Search for direct pair production of supersymmetric top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 8TeV$". In: *Eur. Phys. J.* C76.8 (2016), p. 460. DOI: 10.1140/epjc/s10052-016-4292-5. arXiv: 1603.00765 [hep-ex].

[59] Serguei Chatrchyan et al. "Search for top-squark pair production in the single-lepton final state in pp collisions at $\sqrt{s}$ = 8 TeV". In: *Eur. Phys. J.* C73.12 (2013), p. 2677. DOI: 10.1140/epjc/s10052-013-2677-2. arXiv: 1308.1586 [hep-ex].

[60] Vardan Khachatryan et al. "Searches for third-generation squark production in fully hadronic final states in proton-proton collisions at $\sqrt{s}$ = 8 TeV". In: *JHEP* 06 (2015), p. 116. DOI: 10.1007/JHEP06(2015)116. arXiv: 1503.08037 [hep-ex].

[61] Vardan Khachatryan et al. "Search for top squark pair production in compressed-mass-spectrum scenarios in proton-proton collisions at $\sqrt{s}$ = 8 TeV using the α_T variable". In: *Phys. Lett.* B767 (2017), pp. 403–430. DOI: 10.1016/j.physletb.2017.02.007. arXiv: 1605.08993 [hep-ex].

[62] Vardan Khachatryan et al. "Search for direct pair production of scalar top quarks in the single- and dilepton channels in proton-proton collisions at $\sqrt{s}$ = 8 TeV". In: *JHEP* 07 (2016). [Erratum: JHEP09,056(2016)], p. 027. DOI: 10.1007/JHEP07(2016)027,10.1007/JHEP09(2016)056. arXiv: 1602.03169 [hep-ex].

[63] Vardan Khachatryan et al. "Search for supersymmetry in events with soft leptons, low jet multiplicity, and missing transverse energy in proton–proton collisions at $\sqrt{s}$=8 TeV". In: *Phys. Lett.* B759 (2016), pp. 9–35. DOI: 10.1016/j.physletb.2016.05.033. arXiv: 1512.08002 [hep-ex].

[64] Serguei Chatrchyan et al. "Search for gluino mediated bottom- and top-squark production in multijet final states in pp collisions at 8 TeV". In: *Phys. Lett.* B725 (2013), pp. 243–270. DOI: 10.1016/j.physletb.2013.06.058. arXiv: 1305.2390 [hep-ex].

[65] Serguei Chatrchyan et al. "Search for supersymmetry in hadronic final states with missing transverse energy using the variables α_T and b-quark multiplicity in pp collisions at $\sqrt{s}$ = 8 TeV". In: *Eur. Phys. J.* C73.9 (2013), p. 2568. DOI: 10.1140/epjc/s10052-013-2568-6. arXiv: 1303.2985 [hep-ex].

[66] Serguei Chatrchyan et al. "Search for supersymmetry in pp collisions at $\sqrt{s}$=8 TeV in events with a single lepton, large jet multiplicity, and multiple b jets". In: *Phys. Lett.* B733 (2014), pp. 328–353. DOI: 10.1016/j.physletb.2014.04.023. arXiv: 1311.4937 [hep-ex].

[67] Serguei Chatrchyan et al. "Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s}$= 8

TeV". In: *JHEP* 06 (2014), p. 055. DOI: 10.1007/JHEP06(2014)055. arXiv: 1402.4770 [hep-ex].

[68] Serguei Chatrchyan et al. "Search for new physics in events with same-sign dileptons and jets in pp collisions at $\sqrt{s}$ = 8 TeV". In: *JHEP* 01 (2014). [Erratum: JHEP01,014(2015)], p. 163. DOI: 10.1007/JHEP01(2015)014, 10.1007/JHEP01(2014)163. arXiv: 1311.6736.

[69] CMS Collaboration. "Search for supersymmetry in pp collisions at sqrt(s) = 8 Tev in events with two opposite sign leptons, large number of jets, b-tagged jets, and large missing transverse energy." In: (2013).

[70] Vardan Khachatryan et al. "Searches for Supersymmetry using the M_{T2} Variable in Hadronic Events Produced in pp Collisions at 8 TeV". In: *JHEP* 05 (2015), p. 078. DOI: 10.1007/JHEP05(2015)078. arXiv: 1502.04358 [hep-ex].

[71] Vardan Khachatryan et al. "Search for Supersymmetry Using Razor Variables in Events with *b*-Tagged Jets in *pp* Collisions at $\sqrt{s}$ = 8 TeV". In: *Phys. Rev.* D91 (2015), p. 052018. DOI: 10.1103/PhysRevD.91.052018. arXiv: 1502.00300 [hep-ex].

[72] CMS Collaboration. "Exclusion limits on gluino and top-squark pair production in natural SUSY scenarios with inclusive razor and exclusive single-lepton searches at 8 TeV." In: (2014).

[73] Georges Aad et al. "Search for direct top squark pair production in final states with two tau leptons in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector". In: *Eur. Phys. J.* C76.2 (2016), p. 81. DOI: 10.1140/epjc/s10052-016-3897-z. arXiv: 1509.04976 [hep-ex].

[74] Georges Aad et al. "Search for supersymmetry in events containing a same-flavour opposite-sign dilepton pair, jets, and large missing transverse momentum in $\sqrt{s}$ = 8 TeV pp collisions with the ATLAS detector". In: *Eur. Phys. J.* C75.7 (2015). [Erratum: Eur. Phys. J.C75,no.10,463(2015)], p. 318. DOI: 10.1140/epjc/s10052-015-3661-9, 10.1140/epjc/s10052-015-3518-2. arXiv: 1503.03290 [hep-ex].

[75] Georges Aad et al. "Search for massive supersymmetric particles decaying to many jets using the ATLAS detector in *pp* collisions at $\sqrt{s}$ = 8 TeV". In: *Phys. Rev.* D91.11 (2015). [Erratum: Phys. Rev.D93,no.3,039901(2016)], p. 112016. DOI: 10.1103/PhysRevD.93.039901, 10.1103/PhysRevD.91.112016. arXiv: 1502.05686 [hep-ex].

[76] Georges Aad et al. "Search for squarks and gluinos in events with isolated leptons, jets and missing transverse momentum at $\sqrt{s}$ = 8 TeV with the ATLAS detector". In: *JHEP* 04 (2015), p. 116. DOI: 10.1007/JHEP04(2015)116. arXiv: 1501.03555 [hep-ex].

[77] Georges Aad et al. "Search for strong production of supersymmetric particles in final states with missing transverse momentum and at least three b-jets at $\sqrt{s}$= 8 TeV proton-proton collisions with the ATLAS detector". In: *JHEP* 10 (2014), p. 024. DOI: 10.1007/JHEP10(2014)024. arXiv: 1407.0600 [hep-ex].

[78] Georges Aad et al. "Search for supersymmetry in events with large missing transverse momentum, jets, and at least one tau lepton in 20 fb^{-1} of $\sqrt{s}$ = 8 TeV proton-proton collision data with the ATLAS detector". In: *JHEP* 09 (2014), p. 103. DOI: 10.1007/JHEP09(2014)103. arXiv: 1407.0603 [hep-ex].

[79] Georges Aad et al. "Search for top squark pair production in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s}$ =8 TeV pp collisions with the ATLAS detector". In: *JHEP* 11 (2014), p. 118. DOI: 10.1007/JHEP11(2014)118. arXiv: 1407.0583 [hep-ex].

[80] Georges Aad et al. "Search for direct pair production of the top squark in all-hadronic final states in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector". In: *JHEP* 09 (2014), p. 015. DOI: 10.1007/JHEP09(2014)015. arXiv: 1406.1122 [hep-ex].

[81] Georges Aad et al. "Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum using $\sqrt{s}$ = 8 TeV proton–proton collision data". In: *JHEP* 09 (2014), p. 176. DOI: 10.1007/JHEP09(2014)176. arXiv: 1405.7875 [hep-ex].

[82] Georges Aad et al. "Search for direct third-generation squark pair production in final states with missing transverse momentum and two b-jets in $\sqrt{s}$ = 8 TeV pp collisions with the ATLAS detector". In: *JHEP* 10 (2013), p. 189. DOI: 10.1007/JHEP10(2013)189. arXiv: 1308.2631 [hep-ex].

[83] Georges Aad et al. "Search for new phenomena in final states with large jet multiplicities and missing transverse momentum at $\sqrt{s}$=8 TeV proton-proton collisions using the ATLAS experiment". In: *JHEP* 10 (2013). [Erratum: JHEP01,109(2014)], p. 130. DOI: 10.1007/JHEP10(2013)130, 10.1007/JHEP01(2014)109. arXiv: 1308.1841 [hep-ex].

[84] Benjamin Nachman and Tom Rudelius. "A Meta-analysis of the 8 TeV ATLAS and CMS SUSY Searches". In: *JHEP* 02 (2015), p. 004. DOI: 10.1007/JHEP02(2015)004. arXiv: 1410.2270 [hep-ph].

[85] J. Alberto Casas et al. "What is a Natural SUSY scenario?" In: *JHEP* 06 (2015), p. 070. DOI: 10.1007/JHEP06(2015)070. arXiv: 1407.6966 [hep-ph].

[86] Nima Arkani-Hamed and Savas Dimopoulos. "Supersymmetric unification without low energy supersymmetry and signatures for fine-tuning at the LHC". In: *JHEP* 06 (2005), p. 073. DOI: 10.1088/1126-6708/2005/06/073. arXiv: hep-th/0405159 [hep-th].

[87] G. F. Giudice and A. Romanino. "Split supersymmetry". In: *Nucl. Phys.* B699 (2004). [Erratum: Nucl. Phys.B706,487(2005)], pp. 65–89. DOI: 10.1016/j.nuclphysb.2004.11.048,10.1016/j.nuclphysb.2004.08.001. arXiv: hep-ph/0406088 [hep-ph].

[88] JiJi Fan et al. "Stealth Supersymmetry Simplified". In: *JHEP* 07 (2016), p. 016. DOI: 10.1007/JHEP07(2016)016. arXiv: 1512.05781 [hep-ph].

[89] Paul Lujan. "Search for long-lived particles at CMS". In: *Nuclear and Particle Physics Proceedings* 273-275 (2016). 37th International Conference on High Energy Physics (ICHEP), pp. 496–502. ISSN: 2405-6014. DOI: https://doi.org/10.1016/j.nuclphysbps.2015.09.073. URL: http://www.sciencedirect.com/science/article/pii/S2405601415005623.

[90] Fedor Ratnikov. "Search for R-parity violating Supersymmetry using the CMS detector". In: *EPJ Web Conf.* 60 (2013), p. 18012. DOI: 10.1051/epjconf/20136018012. arXiv: 1307.7609 [hep-ex].

[91] Vardan Khachatryan et al. "Search for stealth supersymmetry in events with jets, either photons or leptons, and low missing transverse momentum in pp collisions at 8 TeV". In: *Phys. Lett.* B743 (2015), pp. 503–525. DOI: 10.1016/j.physletb.2015.03.017. arXiv: 1411.7255 [hep-ex].

[92] A. Nelson, P. Tanedo, and D. Whiteson. "Limiting SUSY compressed spectra scenarios". In: *Phys. Rev.* D93.11 (2016), p. 115029. DOI: 10.1103/PhysRevD.93.115029. arXiv: 1509.08485 [hep-ex].

[93] W.J. Stirling. *13/8 TeV LHC luminosity ratios*. URL: http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html (visited on 02/11/2018).

[94] Christopher Rogan. "Kinematical variables towards new dynamics at the LHC". In: (2010). arXiv: 1006.2727 [hep-ph].

[95] Serguei Chatrchyan et al. "Search for supersymmetry with razor variables in pp collisions at $\sqrt{s}$=7 TeV". In: *Phys. Rev.* D90.11 (2014), p. 112001. DOI: 10.1103/PhysRevD.90.112001. arXiv: 1405.3961 [hep-ex].

[96] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. Geneva: CERN, 2002. URL: http://cds.cern.ch/record/578006.

[97] "CMS: The computing project. Technical design report". In: (2005).

[98] "Data Parking and Data Scouting
at the CMS Experiment". In: (Sept. 2012). URL: https://cds.cern.ch/record/1480607.

[99] Georges Aad et al. "Search for new phenomena in the dijet mass distribution using $p-p$ collision data at $\sqrt{s}$ = 8 TeV with the ATLAS detector". In: *Phys. Rev.* D91.5 (2015), p. 052007. DOI: 10.1103/PhysRevD.91.052007. arXiv: 1407.1376 [hep-ex].

[100] Vardan Khachatryan et al. "Search for narrow resonances in dijet final states at $\sqrt{(}s)$ = 8 TeV with the novel CMS technique of data scouting". In: *Phys. Rev. Lett.* 117.3 (2016), p. 031802. DOI: 10.1103/PhysRevLett.117.031802. arXiv: 1604.08907 [hep-ex].

[101] *Search for Narrow Resonances using the Dijet Mass Spectrum in pp Collisions at sqrt s of 7 TeV*. Tech. rep. CMS-PAS-EXO-11-094. Geneva: CERN, 2012. URL: https://cds.cern.ch/record/1461223.

[102] Silvio Donato. "Fast pile-up jet identification at the High Level Trigger of the CMS experiment". In: *LHCC Student Poster Session*. CERN, Geneva, Switzerland, Mar. 2015. URL: https://indico.cern.ch/event/369822/contributions/873711/.

[103] *Search for narrow resonances decaying to dijets in pp collisions at* $\sqrt{s}$ = 13 *TeV using 12.9 fb*$^{-1}$. Tech. rep. CMS-PAS-EXO-16-032. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2205150.

[104] *Searches for dijet resonances in pp collisions at* $\sqrt{s}$ = 13 TeV *using data collected in 2016.* Tech. rep. CMS-PAS-EXO-16-056. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256873.

[105] Magda Diamantopoulou et al. "Searches for dijet resonances in pp collisions at sqrt(s) = 13 TeV using up to 36 fb-1". CMS Internal Note 2017-013.

[106] Javier Duarte. "Dijet Scouting Analysis". CMS internal talk.

[107] *Search for new physics in high mass diphoton events in proton-proton collisions at* $\sqrt{s}$ = 13 *TeV*. Tech. rep. CMS-PAS-EXO-15-004. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2114808.

[108] Morad Aaboud et al. "Search for resonances in diphoton events at $\sqrt{s}$=13 TeV with the ATLAS detector". In: *JHEP* 09 (2016), p. 001. DOI: 10.1007/JHEP09(2016)001. arXiv: 1606.03833 [hep-ex].

[109] Roberto Franceschini et al. "What is the $\gamma\gamma$ resonance at 750 GeV?" In: *JHEP* 03 (2016), p. 144. DOI: 10.1007/JHEP03(2016)144. arXiv: 1512.04933 [hep-ph].

[110] Simon Knapen et al. "Rays of light from the LHC". In: *Phys. Rev.* D93.7 (2016), p. 075020. DOI: 10.1103/PhysRevD.93.075020. arXiv: 1512.04928 [hep-ph].

[111] Francesco Santanastasio. "Data Scouting Interest". CMS internal talk.

[112] *Search for resonant production of high mass photon pairs using* $12.9\,\mathrm{fb}^{-1}$ *of proton-proton collisions at* $\sqrt{s} = 13$ TeV *and combined interpretation of searches at 8 and 13 TeV*. Tech. rep. CMS-PAS-EXO-16-027. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2205245.

[113] *Search for scalar diphoton resonances with 15.4 fb*$^{-1}$ *of data collected at* $\sqrt{s}$*=13 TeV in 2015 and 2016 with the ATLAS detector*. Tech. rep. ATLAS-CONF-2016-059. Geneva: CERN, Aug. 2016. URL: https://cds.cern.ch/record/2206154.

[114] *Search for light vector resonances decaying to a quark pair produced in association with a jet in proton-proton collisions at* $\sqrt{s} = 13$ TeV. Tech. rep. CMS-PAS-EXO-17-001. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2264843.

[115] *Search for light- and heavy-flavor three-jet resonances in multijet final states at 8 TeV*. Tech. rep. CMS-PAS-EXO-12-049. Geneva: CERN, 2013. URL: https://cds.cern.ch/record/1563139.

[116] David Curtin et al. "Illuminating Dark Photons with High-Energy Colliders". In: *JHEP* 02 (2015), p. 157. DOI: 10.1007/JHEP02(2015)157. arXiv: 1412.0018 [hep-ph].

[117] Roel Aaij et al. "Search for Dark Photons Produced in 13 TeV *pp* Collisions". In: *Phys. Rev. Lett.* 120.6 (2018), p. 061801. DOI: 10.1103/PhysRevLett.120.061801. arXiv: 1710.02867 [hep-ex].

[118] Jakob Salfeld. "Dark Photon Search using Dimuon Scouting". CMS internal talk.

[119] *Search for light dijet resonances with the ATLAS detector using a Trigger-Level Analysis in LHC pp collisions at* $\sqrt{s} = 13$ *TeV*. Tech. rep. ATLAS-CONF-2016-030. Geneva: CERN, June 2016. URL: http://cds.cern.ch/record/2161135.

[120] Sean Benson. "The LHCb Turbo Stream". In: (Apr. 2015). URL: https://cds.cern.ch/record/2011573.

[121] Wouter Verkerke and David P. Kirkby. "The RooFit toolkit for data modeling". In: *eConf* C0303241 (2003). [,186(2003)], MOLT007. arXiv: physics/0306116 [physics].

[122] "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at sqrt(s) = 8 TeV". In: *Journal of Instrumentation* 10.06 (2015), P06005. URL: http://stacks.iop.org/1748-0221/10/i=06/a=P06005.

[123] The CMS collaboration. "Performance of CMS muon reconstruction in pp collision events at sqrt(s) = 7 TeV". In: *Journal of Instrumentation* 7.10 (2012), P10002. URL: http://stacks.iop.org/1748-0221/7/i=10/a=P10002.

[124] Vardan Khachatryan et al. “Reconstruction and identification of tau lepton decays to hadrons and tau neutrinos at CMS”. In: *JINST* 11.01 (2016), P01019. DOI: 10.1088/1748-0221/11/01/P01019. arXiv: 1510.07488 [physics.ins-det].

[125] Carlo Oleari. “The POWHEG-BOX”. In: *Nucl. Phys. Proc. Suppl.* 205-206 (2010), pp. 36–41. DOI: 10.1016/j.nuclphysbps.2010.08.016. arXiv: 1007.3893 [hep-ph].

[126] Richard D. Ball et al. “Parton distributions for the LHC run II”. In: *Journal of High Energy Physics* 2015.4 (Apr. 8, 2015), p. 40. ISSN: 1029-8479. DOI: 10.1007/JHEP04(2015)040. URL: https://doi.org/10.1007/JHEP04(2015)040.

[127] “Measuring Electron Efficiencies at CMS with Early Data”. In: (2008).

[128] “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at sqrt(s) = 8 TeV”. In: *Journal of Instrumentation* 10.08 (2015), P08010. URL: http://stacks.iop.org/1748-0221/10/i=08/a=P08010.

[129] Christoph Borschensky et al. “Squark and gluino production cross sections in pp collisions at sqrt(s) = 13, 14, 33, and 100 TeV”. In: *The European Physical Journal C* 74.12 (Dec. 4, 2014), p. 3174. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-014-3174-y. URL: https://doi.org/10.1140/epjc/s10052-014-3174-y.

[130] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, Aug. 2011. URL: https://cds.cern.ch/record/1379837.

[131] Gary J. Feldman and Robert D. Cousins. “A Unified approach to the classical statistical analysis of small signals”. In: *Phys. Rev.* D57 (1998), pp. 3873–3889. DOI: 10.1103/PhysRevD.57.3873. arXiv: physics/9711021 [physics.data-an].

[132] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J.* C71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z. arXiv: 1007.1727 [physics.data-an].

[133] Vardan Khachatryan et al. “Search for supersymmetry in pp collisions at sqrt(s) = 8 TeV in final states with boosted W bosons and b jets using razor variables”. In: *Phys. Rev.* D93.9 (2016), p. 092009. DOI: 10.1103/PhysRevD.93.092009. arXiv: 1602.02917 [hep-ex].

[134] “Search for supersymmetry in the multijet and missing transverse momentum final state in pp collisions at 13 TeV”. In: *Physics Letters B* 758 (2016), pp. 152–180. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2016.05.002. URL: http://www.sciencedirect.com/science/article/pii/S0370269316301496.

[135] Vardan Khachatryan et al. "Search for new physics with the M_{T2} variable in all-jets final states produced in pp collisions at $\sqrt{s}$ = 13 TeV". In: *JHEP* 10 (2016), p. 006. DOI: 10.1007/JHEP10(2016)006. arXiv: 1603.04053 [hep-ex].

[136] Jesse Thaler and Ken Van Tilburg. "Identifying Boosted Objects with N-subjettiness". In: *JHEP* 03 (2011), p. 015. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 [hep-ph].

[137] Andrew J. Larkoski et al. "Soft Drop". In: *JHEP* 05 (2014), p. 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].

[138] CMS Collaboration. *CMS Supersymmetry Physics Results*. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS (visited on 02/27/2018).

[139] David J.C. Mackay. *Introduction to Gaussian Processes*. Dept. of Physics, Cambridge University, UK, 1998.

[140] Meghan Frate et al. "Modeling Smooth Backgrounds and Generic Localized Signals with Gaussian Processes". In: (2017). arXiv: 1709.05681 [physics.data-an].

[141] S. Golchi and R. Lockhart. "A Bayesian Search for the Higgs Particle". In: *ArXiv e-prints* (Jan. 2015). arXiv: 1501.02226 [stat.AP].

[142] *Measurement of the top quark mass in the dileptonic ttbar decay channel using the Mbl, MT2, and MAOS Mblv observables*. Tech. rep. CMS-PAS-TOP-15-008. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2204924.

[143] D. C. Liu and J. Nocedal. "On the Limited Memory BFGS Method for Large Scale Optimization". In: *Math. Program.* 45.3 (Dec. 1989), pp. 503–528. ISSN: 0025-5610. DOI: 10.1007/BF01589116. URL: http://dx.doi.org/10.1007/BF01589116.

[144] J. Hensman et al. "MCMC for Variationally Sparse Gaussian Processes". In: *ArXiv e-prints* (June 2015). arXiv: 1506.04000 [stat.ML].

[145] H. Salimbeni and M. Deisenroth. "Doubly Stochastic Variational Inference for Deep Gaussian Processes". In: *ArXiv e-prints* (May 2017). arXiv: 1705.08933 [stat.ML].

[146] R. M. Neal. "MCMC using Hamiltonian dynamics". In: *ArXiv e-prints* (June 2012). arXiv: 1206.1901 [stat.CO].

[147] z. wang, S. Mohamed, and N. de Freitas. "Adaptive Hamiltonian and Riemann Manifold Monte Carlo Samplers". In: *ArXiv e-prints* (Feb. 2013). arXiv: 1302.6182 [stat.CO].

[148] I. Murray and R. Prescott Adams. "Slice sampling covariance hyperparameters of latent Gaussian models". In: *ArXiv e-prints* (June 2010). arXiv: 1006.0868 [stat.CO].

[149] *Scikit-Optimize*. https://github.com/scikit-optimize/scikit-optimize.

[150] D. P Kingma and M. Welling. "Auto-Encoding Variational Bayes". In: *ArXiv e-prints* (Dec. 2013). arXiv: 1312.6114 [stat.ML].

[151] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980 (2014). arXiv: 1412.6980. URL: http://arxiv.org/abs/1412.6980.

[152] *Search for Excess Higgs production in diphoton final states using the razor variables at* $\sqrt{s}$ = 13 TeV. Tech. rep. CMS-PAS-SUS-16-045. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256437.

[153] Matthew R. Buckley et al. "Cornering Natural SUSY at LHC Run II and Beyond". In: *JHEP* 08 (2017), p. 115. DOI: 10.1007/JHEP08(2017)115. arXiv: 1610.08059 [hep-ph].

[154] G. Apollinari et al. "High Luminosity Large Hadron Collider HL-LHC". In: *CERN Yellow Report* 5 (2015), pp. 1–19. DOI: 10.5170/CERN-2015-005.1. arXiv: 1705.08830 [physics.acc-ph].

[155] Howard Baer et al. "Multichannel assault on natural supersymmetry at the high luminosity LHC". In: *Phys. Rev.* D94.3 (2016), p. 035025. DOI: 10.1103/PhysRevD.94.035025. arXiv: 1604.07438 [hep-ph].

[156] L. Zhang et al. "LSO/LYSO Crystals for Calorimeters in Future HEP Experiments". In: *IEEE Transactions on Nuclear Science* 61.1 (Feb. 2014), pp. 483–488. ISSN: 0018-9499. DOI: 10.1109/TNS.2013.2279993.

[157] S. Seifert et al. "Accurate measurements of the rise and decay times of fast scintillators with solid state photon counters". In: *IEEE Nuclear Science Symposuim Medical Imaging Conference*. Oct. 2010, pp. 1736–1739. DOI: 10.1109/NSSMIC.2010.5874071.

[158] Hamamatsu Photonics K. K. *Photomultiplier Tubes: Basics and Applications*. Hamamatsu Photonics K. K. Electron Tube Division, 2007. URL: http://www.hamamatsu.com/eu/en/community/optical_sensors/articles/photomultiplier_tube_handbook/index.html.

[159] Photek Ltd. *Photomultipliers and Photodiodes*. 2010. URL: http://www.photek.com/pdf/datasheets/detectors/DS006_Photomultipliers.pdf (visited on 03/08/2018).

[160] Stefan Ritt, Roberto Dinapoli, and Ueli Hartmann. "Application of the DRS chip for fast waveform digitizing". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 623.1 (2010). 1st International Conference on Technology and Instrumentation in Particle Physics, pp. 486–488. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2010.03.045. URL: http://www.sciencedirect.com/science/article/pii/S0168900210006091.

[161] A. Ronzhin et al. "Development of a new fast shower maximum detector based on microchannel plates photomultipliers (MCP-PMT) as an active element". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 759 (2014), pp. 65–73. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2014.05.039. URL: http://www.sciencedirect.com/science/article/pii/S016890021400566X.

[162] M. ALBRECHT et al. "SCINTILLATORS AND WAVELENGTH SHIFTERS FOR THE DETECTION OF IONIZING RADIATION". In: *Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications*. WORLD SCIENTIFIC, 2011, pp. 502–511. DOI: 10.1142/9789812702708_0074. eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789812702708_0074. URL: https://www.worldscientific.com/doi/abs/10.1142/9789812702708_0074.

[163] A. Bornheim. "Design studies for the Phase II upgrade of the CMS Barrel Electromagnetic Calorimeter". In: *JINST* 12.03 (2017), p. C03018. DOI: 10.1088/1748-0221/12/03/C03018.

[164] A. M. Magnan. "HGCAL: a High-Granularity Calorimeter for the endcaps of CMS at HL-LHC". In: *JINST* 12.01 (2017), p. C01042. DOI: 10.1088/1748-0221/12/01/C01042.

[165] Farrell, Steven et al. "The HEP.TrkX Project: deep neural networks for HL-LHC online and offline tracking". In: *EPJ Web Conf.* 150 (2017), p. 00003. DOI: 10.1051/epjconf/201715000003. URL: https://doi.org/10.1051/epjconf/201715000003.

[166] Rainer Mankel. "A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 395.2 (1997), pp. 169–184. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(97)00705-5. URL: http://www.sciencedirect.com/science/article/pii/S0168900297007055.

[167] Giuseppe Cerati et al. "Kalman filter tracking on parallel architectures". In: *J. Phys. Conf. Ser.* 898.4 (2017), p. 042051. DOI: 10.1088/1742-6596/898/4/042051. arXiv: 1702.06359 [physics.ins-det].

[168] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *Nature* 521 (May 27, 2015), pages. URL: http://dx.doi.org/10.1038/nature14539.

[169] François Chollet et al. *Keras*. https://github.com/keras-team/keras. 2015.

[170] Theano Development Team. “Theano: A Python framework for fast computation of mathematical expressions”. In: *arXiv e-prints* abs/1605.02688 (May 2016). URL: http://arxiv.org/abs/1605.02688.

[171] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[172] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735. eprint: https://doi.org/10.1162/neco.1997.9.8.1735. URL: https://doi.org/10.1162/neco.1997.9.8.1735.

[173] Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006. ISBN: 978-0-387-31073-2.

[174] Ujjwal Karn. *An Intuitive Explanation of Convolutional Neural Networks*. 2016. URL: https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/ (visited on 03/06/2018).

[175] Christopher Olah. *Understanding LSTM Networks*. 2015. URL: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (visited on 03/06/2018).

[176] Josh Bendavid. “Use of multivariate techniques for calorimeter energy reconstruction”. In: *CALOR 2016*. 2016. URL: %7Bhttps://indico.cern.ch/event/472938/contributions/1150753/attachments/1275329/1891843/calorRegressionMay19-2016.pdf%7D.

[177] T. N. Kipf and M. Welling. “Semi-Supervised Classification with Graph Convolutional Networks”. In: *ArXiv e-prints* (Sept. 2016). arXiv: 1609.02907 [cs.LG].

[178] *Slurm: A Highly Scalable Workload Manager*. URL: https://github.com/SchedMD/slurm (visited on 03/03/2018).

[179] Blaise Barney. *Message Passing Interface (MPI)*. URL: https://computing.llnl.gov/tutorials/mpi/ (visited on 03/03/2018).

[180] Lisandro D. Dalcin et al. “Parallel distributed computing using Python”. In: *Advances in Water Resources* 34.9 (2011). New Computational Methods and Software Tools, pp. 1124–1139. ISSN: 0309-1708. DOI: https://doi.org/10.1016/j.advwatres.2011.04.013. URL: http://www.sciencedirect.com/science/article/pii/S0309170811000777.

[181] Joeri R. Hermans. *Distributed Keras: Distributed Deep Learning with Apache Spark and Keras*. URL: https://github.com/cerndb/dist-keras (visited on 03/03/2018).

[182] Alexander Sergeev and Mike Del Balso. "Horovod: fast and easy distributed deep learning in TensorFlow". In: *arXiv preprint arXiv:1802.05799* (2018).

[183] Adam Paszke et al. "Automatic differentiation in PyTorch". In: (2017).

[184] Sebastian Ruder. "An overview of gradient descent optimization algorithms". In: *CoRR* abs/1609.04747 (2016). arXiv: 1609.04747. URL: http://arxiv.org/abs/1609.04747.

[185] A. C. Wilson et al. "The Marginal Value of Adaptive Gradient Methods in Machine Learning". In: *ArXiv e-prints* (May 2017). arXiv: 1705.08292 [stat.ML].

[186] Jeffrey Dean et al. "Large Scale Distributed Deep Networks". In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. NIPS'12. Lake Tahoe, Nevada: Curran Associates Inc., 2012, pp. 1223–1231. URL: http://dl.acm.org/citation.cfm?id=2999134.2999271.

[187] Sixin Zhang, Anna Choromanska, and Yann LeCun. "Deep learning with Elastic Averaging SGD". In: *CoRR* abs/1412.6651 (2014). arXiv: 1412.6651. URL: http://arxiv.org/abs/1412.6651.

[188] *Visualization Cluster: Cooley*. URL: https://www.alcf.anl.gov/resources-expertise/analytics-visualization.

[189] J. de Favereau et al. "DELPHES 3, A modular framework for fast simulation of a generic collider experiment". In: *JHEP* 02 (2014), p. 057. DOI: 10.1007/JHEP02(2014)057. arXiv: 1307.6346 [hep-ex].

[190] Daniel Weitekamp et al. "Deep topology classifiers for a more efficient trigger selection at the LHC". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (to appear)*. 2017. URL: https://dl4physicalsciences.github.io.

[191] S. Hadjis et al. "Omnivore: An Optimizer for Multi-device Deep Learning on CPUs and GPUs". In: *ArXiv e-prints* (June 2016). arXiv: 1606.04487 [cs.DC].