University of California
Santa Barbara

Search for R-parity violating supersymmetry at the 13 TeV LHC

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy

in

Physics

by

Rohan Bhandari

Committee in charge:

Professor David Stuart, Chair
Professor Harry Nelson
Professor Nathaniel Craig

June 2018

The Dissertation of Rohan Bhandari is approved.

Professor Harry Nelson

Professor Nathaniel Craig

Professor David Stuart, Committee Chair

May 2018

Search for R-parity violating supersymmetry at the 13 TeV LHC

Dedication here

Acknowledgements

Acknowledgements Here.

Curriculum Vitæ
Rohan Bhandari

Education

20XX	Ph.D. in Physics (Expected), University of California, Santa Barbara.
20XX	M.A. in Physics, University of California, Santa Barbara.
20XX	etc

Publications

Publications.

Abstract

Search for R-parity violating supersymmetry at the 13 TeV LHC

by

Rohan Bhandari

Abstract text.

Contents

List of Figures

List of Tables

Chapter 0

Introduction for Non-technical Readers

0.1 Section Title

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Part I

Theoretical Context and Motivations

Chapter 1

Introduction

The standard model (SM) of particle phyiscs is one of the most successfull theories in all of science. It has unified the electromagnetic, strong, and weak forces together into one framework, describing the interactions of electrons, gluons, quarks, and the other fundamental particles of the unvierse.

1.1   Permissions and Attributions

1. The content of chapter 2 and appendix A is the result of a collaboration with Alice and Bob, and has previously appeared in the (Journal) (paper citation). It is reproduced here with the permission of (Institution): `http://`.

Chapter 2

Theory

2.1 Section Title

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

2.2 Section Title
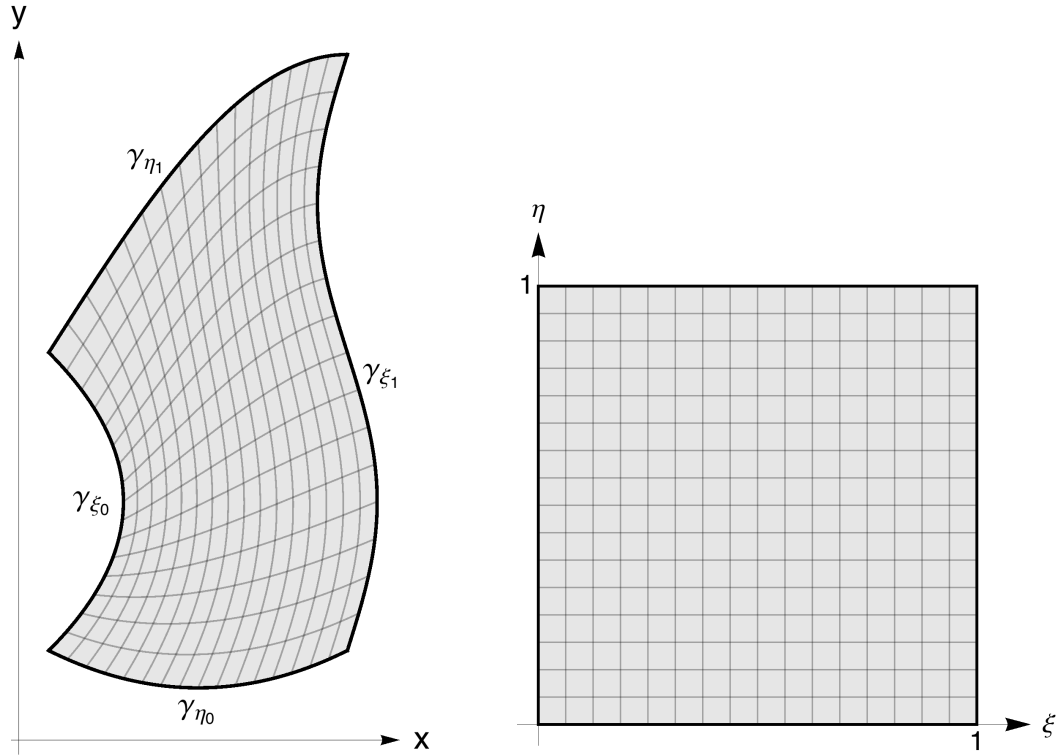
Figure 2.1: Figure Captions.

Part II

Experimental Apparatus

Chapter 3

Experimental Apparatus

While the Large Hadron Collider was approved over 20 years ago, much of its design was influenced by the needs of searches for physics beyond the Standard Model (BSM). The two most important properties of a collider, for a BSM search, are its center-of-mass energy and its (instantaneous) luminosity, both of which were designed to be higher than any previous experiment. The designed center-of-mass energy ($\sqrt{s}$) of 14 TeV allows for the production of particles heavier than ever previously, while the designed luminosity of 10^{34} cm^{-2}s^{-1} allows BSM searches to probe very rare processes.

In the same way, the design of the Compact Muon Solenoid (CMS) detector reflects the needs of BSM searches. In particular, to fully search the uncovered parameter space of new physics an all-purpose, hermitic detector that can precisely measure a variety of particles and reliably determine the MET in an event is needed to cover the many (un)theorized new physics models.

This chapter summarizes in more detail the major features of both the LHC and CMS. A complete description of both can be found in Refs.[INCLUDE].
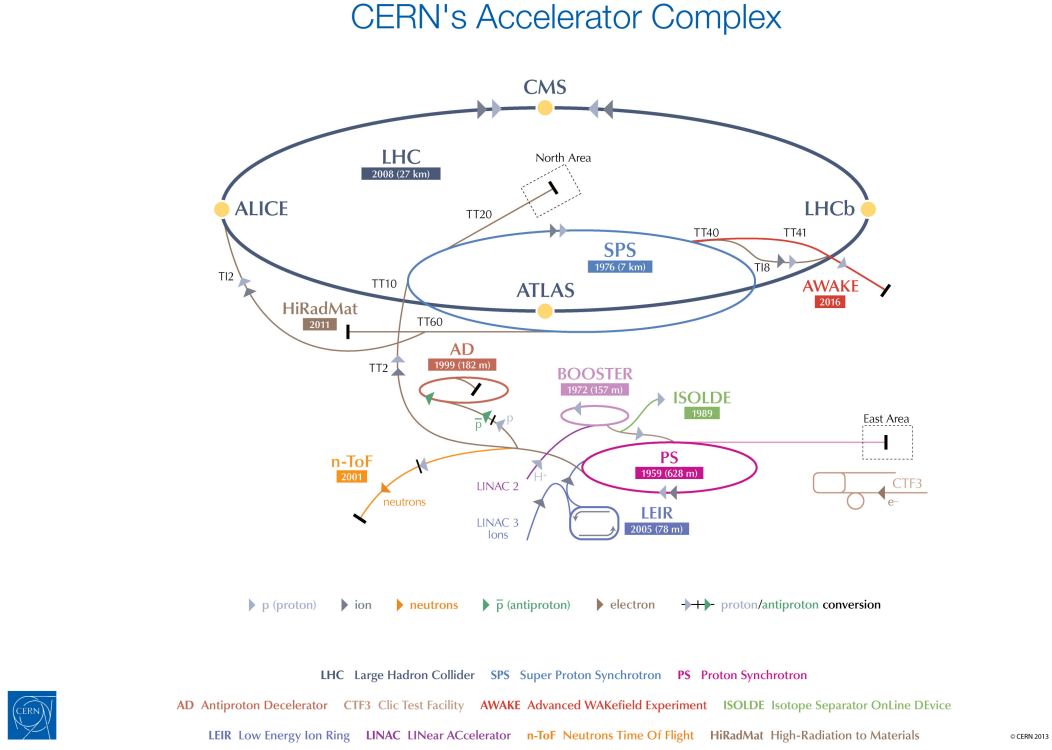
Figure 3.1: A schematic of the CERN LHC accelerator complex. [1]

3.1   The Large Hadron Collider

The LHC since 2015 has been colliding protons together at $\sqrt{s} = 13$ TeV, slightly below the designed specifications but still at an unsurpassed energy. In order to reach this center-of-mass energy, the LHC uses a large accelerator complex consisting of a sucession of many smaller particle accelerators, which is necessary to produce protons and bring them up to a speed such that they can be injected into the LHC ring. A diagram of the CERN accelerator complex is shown in Figure 3.1.

This process first begins with a simple bottle of hydrogen gas, which are ionized by an electric field to produce the needed protons. These resulting protons are then fed into LINAC 2, the first accelerator in the chain, which accelerates them up to 50 MeV,

creating a beam of protons. The proton beam is then passed succesively to the Proton Synchtron Booster, Proton Synchotron, and Super Proton Synchtron, where the beam reaches energies of 1.4 GeV, 25 GeV and 450 GeV, respectively. At the Proton Synchotron, the beams are additionally split into "bunches", each consisting of $O(10^{11})$ protons and separated in time by 25 ns. Finally, the protons can be injected into the two beam pipes of the LHC, each circulating in opposite directions. These beams continue to be accelerated until they reach their final energy of 6.5 TeV, allowing for collisions at $\sqrt{s} = 13$ TeV. At this point, the proton beams are focused and fine-tuned at several stages in order to increase the luminosity. In 2016, the LHC was able to collide protons with an instantaneous luminosity of 1.4×10^{34} cm^{-2}s^{-1}, exceeding its designed specification, and deliver a record-high integrated luminosity of 41.07 fb^{-1}.

These improvements from previous generations of colliders greatly increase the reach of searches for new, heavy particles. Since the center-of-mass energy of the actual colliding partons ($\sqrt{\hat{s}}$) is typically much less than the overall center-of-mass energy, raising the collider's energy can greatly increase the production cross-section of heavy particles, especially of those around the TeV scale. For example, Figure 3.2, which depicts the ratio of parton luminosities at $\sqrt{s} = 13$ TeV and 8 TeV as a function of the characteristic mass scale of the event, shows that a 2000 GeV gluino will be produced through gg scattering processes $\sim$15$\times$ more often with less than a doubling of the collider energy. Increasing a collider's energy, however, is not always a practical option, involving new technologies, expensive upgrades, or even a new collider. When this is the case, the best alternative to continue to probe rare processes is to simply take more data, more quickly, which a high luminosity collider like the LHC allows for.
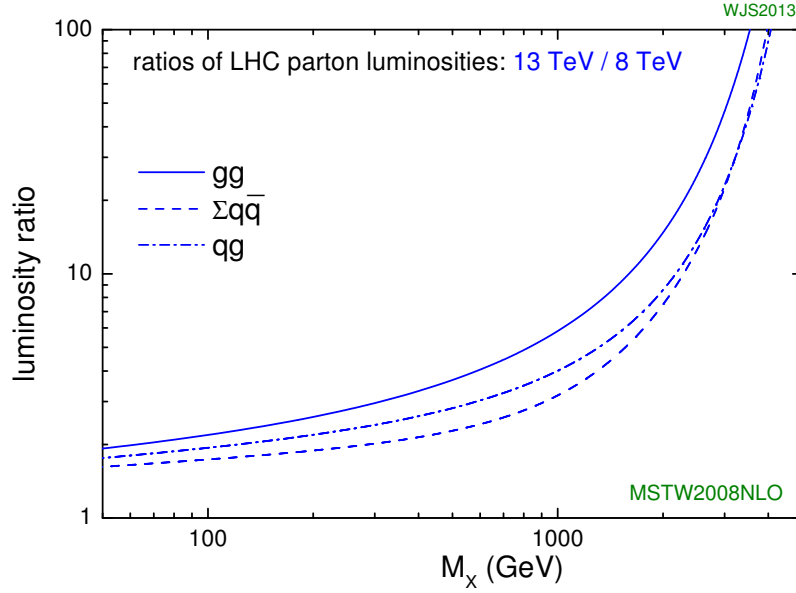
Figure 3.2: Ratio of parton luminosities at $\sqrt{s} = 13$ and 8 TeV. [2]

3.2 Compact Muon Solenoid

Along the tunnels of the LHC, below Cessy, France, sits the CMS detector where the proton-proton collisions are recorded. The overall shape of the detector is cylindrical with a length of 21.6 m and radius of 7.3 m, while weighing roughly 14,000 metric tons. The CMS detector is sometimes called a cylindrical onion, as this shape is constructed through layers of specialized detectors, each designed to provide precise measurements for a particular particle type. Peeling back the layers from the outside-in, the first sub-detector is the muon system. Next is a superconducting solenoid of 6 m internal diameter that produces a magnetic field of 3.8 T, and perhaps most importantly provides the "S" in CMS. Placed within the solenoidal magnet, is the rest of the CMS detector, namely the Hadronical Calorimeter (HCAL), Electromagnetic Calorimeter (ECAL), and a silicon tracker. The design of fitting most of the detector components within the solenoid is
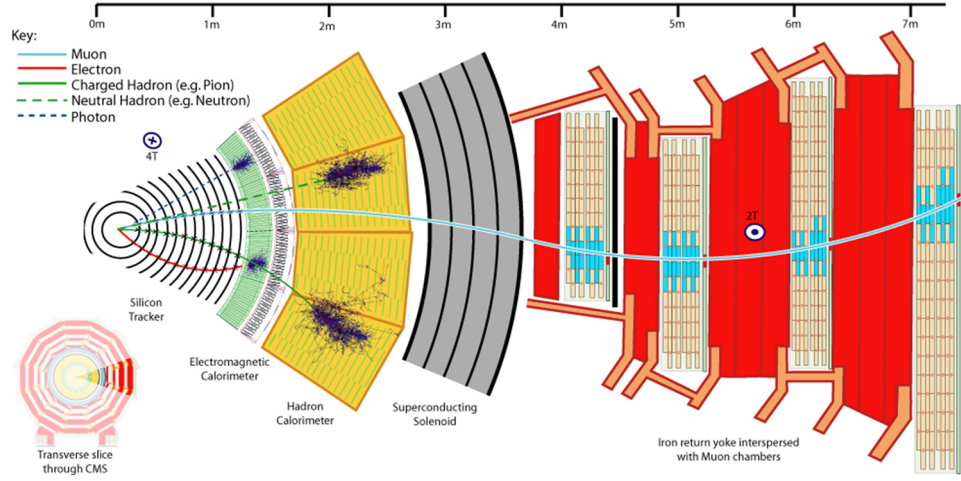
Figure 3.3: A diagram showing the various sub-detectors of the CMS detector. [3]

responsible for the "C" in CMS. A diagram of the layout of the CMS detector can be seen in Figure 3.3.

At the center of the silicon tracker is Interaction Point 5, the beam crossing which provides the proton-proton collisions to the CMS detector, and is the nominal origin of CMS's coordinate system. The x-axis is defined to point towards the center of the LHC ring and the y-axis is defined to point up towards the surface, both of which are transverse to the proton beam. The z-axis points along the beamline with the positive direction given by the right-hand rule relative to the x- and y-axes. Due to CMS detector shape, it is often useful to convert the cartesian coordinates to a cylindrical coordinate system. In this system, the azimuthal angle, ϕ, is measured from the x-axis in the xy-plane, and the polar angle, θ, is measured from the z-axis. The polar angle, however, is often replaced by psuedorapidity, defined as $\eta = -\ln(\theta/2)$. Thus, any point in the CMS coordinate system can be represented by (z, η, ϕ). A diagram showing both the cartesian and cylindrical coordinate systems can be seen in Figure 3.4.

The remainder of this section briefly describes the mean features of the various CMS sub-detectors.

11
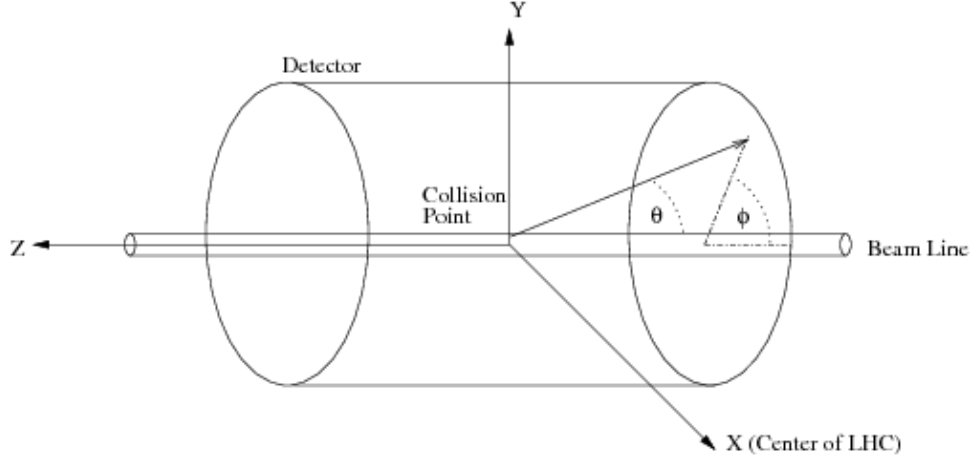
Figure 3.4: A diagram of the cartesian and cylindrical coordinate systems used by CMS. [4]

3.2.1   Inner Tracking System

The tracking system is used for precise measurements of the trajectories of charged particles, as well as reconstruction of seconday vertices. As the tracking system is the closest subdetector to the interaction point, it faces a very large particle flux rate and so must be able to provide both high granularity and fast response, as well be able to survive operating in those conditions with ane expected lifetime of 10 years. At the same time, these features must be balanced with minimizing the amount of material in order to reduce unwanted interactions with the detector, such as multiple scattering, photon conversion, and nuclear interactions. These requirements lead to a tracking system composed entirely of silicon technology.

The CMS tracking system is actually composed of two parts. The first is the pixel detector, which surrounds the interaction point, and is composed of 3 barrel layers at radii between 4.4 and 10.2 cm and 2 endcap layers that extend the acceptance up to $|\eta| < 2.5$. In total, the pixel detector covers an area of roughly 1 m^2 with 66 million pixels and achieves a resolution of roughly 10 and 20 microns in the directions transverse and logitudinal to the beam line, respectively.
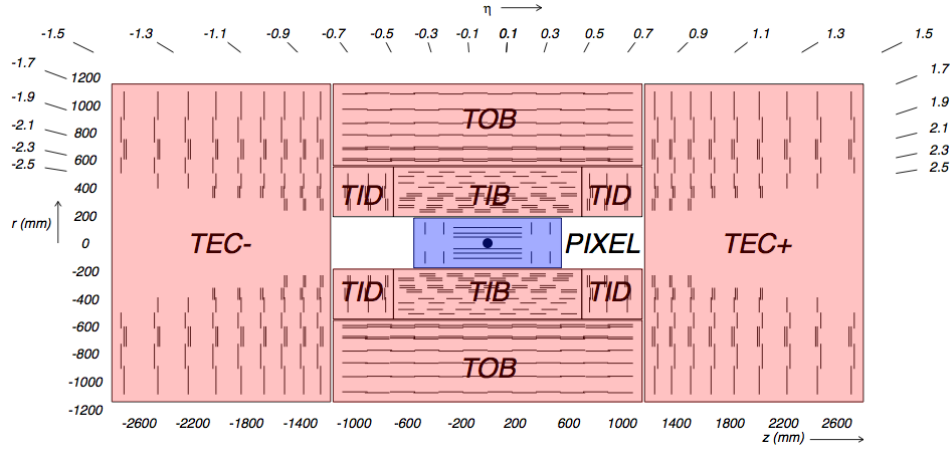
Figure 3.5: Layout of the CMS tracking system, showing both the pixel detector (blue) and the strips detector (red). [5]

The second part of the tracking system is the strips detector which sits just outside the pixel detector. The strips detector is composed of 4 parts: the tracker inner barrel (TIB), tracker inner disks (TID), tracker outer barrel (TOB), and the tracker endcaps (TEC). The TIB and TID extend up to 55 cm in radius and are composed of 4 and 3 layers, respectively, while the TOB, which surronds the TIB and TID, extends out to 116 cm and is composed of 6 layers. Lastly, the TEC which sits next to the other strip detector components, covers a radius of 22.5 to 113.5 cm and is composed of 9 disks. In total, the strips detector covers an area of 198 m^2 with 9.3 million strips. A layout of the tracking system including the pixel and strips detector is shown in Figure 3.5.

3.2.2   Electromagnetic Calorimeter

The primary purpose of the electromagnetic calorimeter (ECAL) is to measure the energy of electrons and photons. The ECAL is a hermetic, hemogenous detector made up of a barrel part, convering the $|\eta| < 1.479$ region, and two endcap parts that covers $1.566 < |\eta| < 3.0$. Both the barrel and endcap sections are comprised of lead tungstate

13

($PbWO_4$) crystals with 61,200 in the barrel and 7,324 in each of the endcaps. The use of the $PbWO_4$ crystals was motivated by their high density, short radiation length, small Molière radius, and radiation hardness, all of which allow for a fine granularity, radiation resistant, compact calorimeter.

The lead tungstate crystals act as scintillators, which produce an amount of light that is proportional to the energy of an incident particle. This light is then converted to an electrical signal by silicon photodectors (avalanche photodetectors in the barrel and vacuum hototriodes in the endcaps), which is used for the final energy measurement. The resulting resolution on the energy measurements is given by

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C \tag{3.1}$$

where S is the stochastic term, N the noise term, C the constant term, and E is in units of GeV. Typical values for S, N, and C, as measured in electron beam tests, are 2.8%, 12%, and 0.30%, repectively.

In addition to the ECAL barrel and endcaps is a preshower detector, which sits in front of the endcaps, convering $1.653 < |\eta| < 2.6$. The main purpose of the preshower detector is to identify neutral pions by improving the granularity, so as to be able resolve photon pairs from the decay of high energy pions that otherwise would be mis-measured as single photons. The preshower detector also provides improved position resolution for electrons and photons and helps identify electrons from minimium ionizing particles. The full layout of the ECAL is shown in Figure 3.6.

3.2.3   Hadronic Calorimeter

The primary purpose of the hadronic calorimeter (HCAL) is to measure the energy of hadrons, which can pass through the ECAL as they primarilly interact through the

14
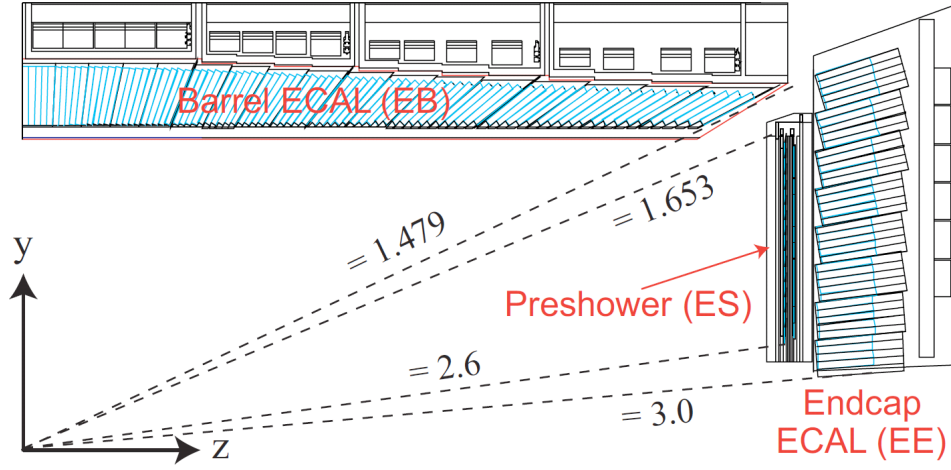
Figure 3.6: A cross section of the ECAL, showing its geometry and layout. [6]

strong force. The HCAL is a sampling calorimeter made up of either brass, iron, or steel absorbers and uses plastic scintillator tiles as the sampling material, which measures the energy of hadrons through scintillation, similarly to the ECAL: As a hadron reaches through the HCAL, it interacts with one of the absorber layers, which results into a "shower" of particles that produces light in the scintillator tiles as the resulting particles pass through the sampling layers. These light pulses are converted to electrical signals by optical fibers, which when summed have an amplitue proportional to the hadron's energy.

The HCAL is separated into 4 components: the hadron barrel (HB), hadron endcap (HE), hadron forward (HF), and hadron outer (HO), the layouts of which are shown in Figure 3.7. The HB and HE completely surround the ECAL and were designed to minimize any cracks between the two subdetectors with the HB covering $|\eta| < 1.3$ and the HE covering the rest up to $|\eta| = 3$. Both components function as sampling calorimteres with alternating absorber and sampling layers. In the HB, the first and last layers are made of steel while the 14 other absorber layers are made of brass, while the HE is made
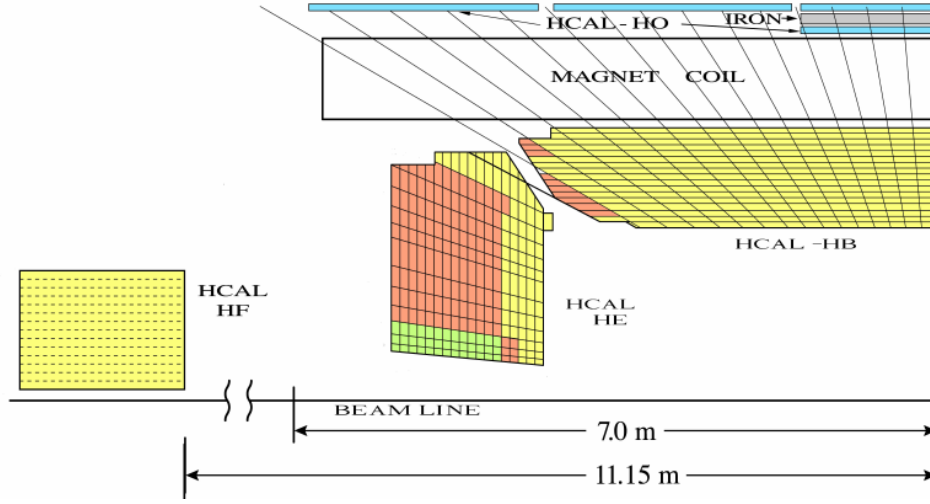
Figure 3.7: The layout and geometry of a quarter of the HCAL detector. [7]

up of 18 brass absorber layers. For both components, there are sampling layers made of plastic scintillator tiles interspersed between each of the absorber layers.

The HF is used to measure the energy of the forward most hadrons in the pseudo-rapidity range of $3.0 < |\eta| < 5.0$. At this forward position, the HF faces extradorinary levels of particle flux and had to be designed to handle this radiation. Due to this constraint, the HF uses quartz fibers instead of plastic scintillator tiles as its active medium, as the quartz fiber are more radiation hard. The HF uses both long fibers, which run the full depth (165 cm) of the detector, and short fibers, which begin 22 cm from the front end of the detector. This geometry allows the HF to provide depth information of the energy deposits, which helps to identify electron and photons from hadrons, as the former tends to deposit most of its energy in the first depth, while the latter deposits its energy more equally between the two depths. These fibers are embedded into the steel structure of the HF, which also acts as the absorber.

Lastly is the HO, whose main purpose is to act as a "tail catcher". Due to the geometrical constraint that the HCAL fit within the CMS solenoid, the HB does not

16
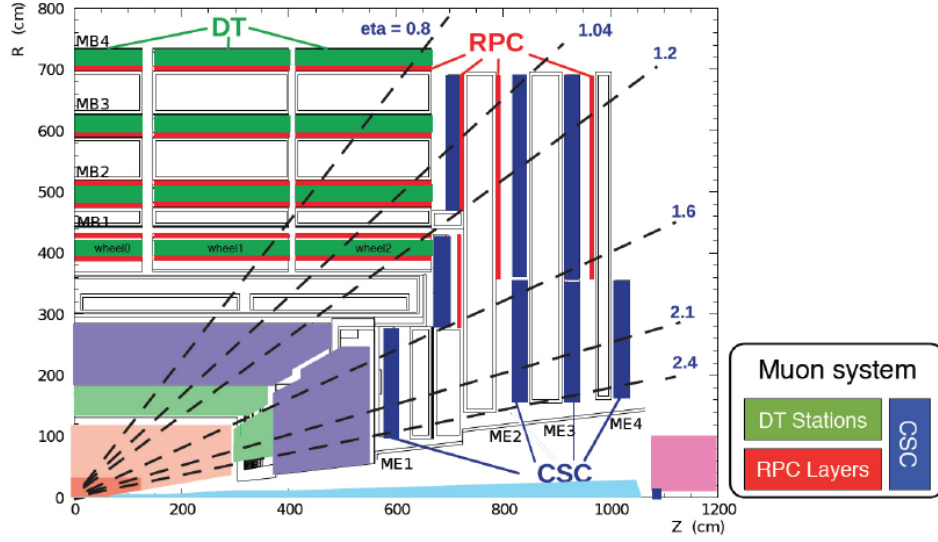
Figure 3.8: The layout of the muon system within the CMS detector. [8]

have enough material in the central η region to adequately contain hadron showers. So to provide extra sampling layers, the HO sits just beyond the solenoid and has 1 to 2 scintillator layers and uses the magnet as an extra absorber layer. At this position, the HO is able to identify late starting showers and measure the amount of energy that is deposited past the HB.

3.2.4   Muon System

Muons, as implied by the "M" in CMS, are a central focus of the CMS detector, and the responsibiilty of identifying muons with high purity and providing precise momenta measurements falls to the muon system. To do this, the muon system is composed of three types of gaseous detectors, motivated by the need to cover a large area and varying radiation environments. Figure 3.8 shows the layout of the muon system within the CMS detector.

In the barrel region, $|\eta| < 1.2$, drift tube chambers (DTs) are used as the neutron-

induced background is small, the muon rate is small, and the magnetic field is uniform. The DTs are organized into 4 stations, with three of the stations containing 8 chambers that measure position in the $r - \phi$ plane and 4 that measure position in the z-direction. The last station only contains the 8 $r - \phi$ measuring chambers. In the endcap regions of CMS, $0.9 < |\eta| < 2.4$, the expected muon and background rates are higher and the magnetic field is large and non-uniform, both of which preclude the use of DTs. Instead, the muon system endcaps are instrumented with cathode strip chambers (CSCs) that have a high response time, fine segmentation, and higher radiation resistance The CSCs have 4 stations in each endcap with chambers that are aligned perpendicular to the beam line and are able to provide measurements in the $r - phi$ plane and z-direction, along with the beam crossing time of a muon.

Both the DTs and CSC are capable of providing high efficiency and pure muon p_{T} triggers, independent of the rest of the detector. But in order to further improve this, particularly at the full LHC luminosity, another complementary trigger system consisting of Resisitive Plate Chambers (RPCs) was added to both the the barrel and endcap regions ($|\eta|$¡1.6). The RPCs are double-gap chambers that operate with a fast response and good time resolution. The spatial resolution, however, is coarser than the DTs or CSCs, though the extra hits in the RPC still help resolve ambiguities when making tracks. There are a total of 6 RPC layers in the barrel muon system, which help improver triggers for low p_{T} muons, and 3 layers in each of the endcaps that help reduce background and improve the time and p_{T} resolution of muons.

3.2.5   Trigger System

The high instantaneous luminosity of the LHC provides many techincal challenges for the data aquisition system (DAQ), with proton-proton collissions occuring every 25 ns,

corresponding to a frequency of 40 MHz. At this collision rate, it is unfeasible to process and store the data for each event. In order to reduce the rate, a two-stage trigger system is used to select only the most "interesting" events for processing.

The first stage is the Level-1 (L1) trigger system, which has approximately only 4 μs to decide whether or not an event should be further processed. In order to operate at this timescale, the L1 trigger uses only coarse-grained information from the CMS calorimeters and the muon system.

For the calorimeter set of data, the L1 first generates trigger primitives by looking for lar for large energy deposits in the calorimter. These trigger primitives are then passed to the Regional Calorimeter Trigger (RCT), which uses this information to determine electron/photon candidates and transverse energy sums per calorimeter region. In addition, the RCT also calculates information relevant for detecting minimally ionizing particles, vetoing tau leptons, and muon isolation. Lastly, the Global Calorimeter Trigger (GCT) uses the information from the RCT to construct jets and calculate the event-level transverse energy and missing transverse energy, along with the final isolated and non-isolated electron/photon candidates.

For the muon portion of the L1 trigger system, the DTs and CSCs both compute local trigger informationwhich consists of two- and three-dimensional track segments, respectively. This information is then passed to a join DT-CSC track finder, which connects these segments into a full candidate track. At the same time, the RPC constructs a separate, independent set of track candidates. Both sets of candidate tracks are sent to the Global Muon trigger (GMT), which also takes in the relavent information from the RCT), to construct muon candidates.

Lastly, the candidate particles and event-level information from the GCT and GMT are sent to the Global Trigger, which takes this information and checks to see if certain criteria are met. If so, a L1 Accept (L1A) is generated, which signals for the event to
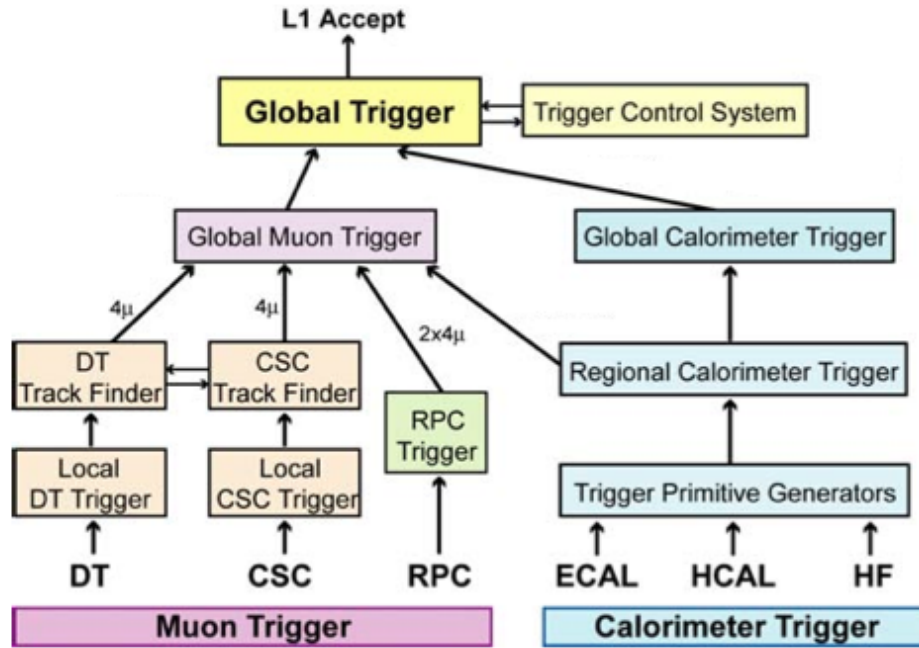
Figure 3.9: Flowchart depicting the generation of a L1 Accept. [9]

be fully read out. This process, shown in Figure 3.9, reduces the full readout rate to at most 100 kHz.

On the generation of an L1A, all the CMS subsystems read out their buffered data corresponding to the L1A event to the an event builder. The data the event builder receives is both more complete and at a finer resolution, allowing it to construct more complex quantities before sending it to the High Level Trigger (HLT), the second stage of the trigger system.

The HLT software is run on a processor farm that reconstructs events in greater detail to decide whether or not they should be kept. This framework is flexible as both the HLT software and the processor farm can be updated to meet changing experimental needs. As such the exact criteria used by the HLT in its decision varies with time, but generally involves thresholding the p_T and/or multiplicity of particles along with

event-level quantities, such as H_T. At the end of this process, the trigger rate is at approximately 100 Hz.

Chapter 4

Particle Reconstruction and Identification

4.1 Tracks

Track reconstruction with the CMS detector faces many challenges, as at each bunch crossing $\sim O(10^3)$ particles are expected to pass through the CMS tracking system, all of which must be reconstructed in time to be inputted to the HLT. This constraint makes it immensely challenging to attain high track-finding efficiency, while minimzing the number of fake tracks.

The first step of track reconstruction is to reconsutrct "hits" in a process called "local reconstruction". In this step, signals in both the pixel and strip channels that are above some zero-suppression threshold are clustered together into hits, where the cluster positions and uncertainties are then estimated.

Next, track are reconstructed from these hits in order to provide estimates for the momentum and position of charged particles associated with the track. This is done using specialized software based off of Kalman filters known as Combinatorial Track Finder (CTF). In order to reduce the combinatorial complexity of the problem, the CTF track reconstruction is peformed six times. Each iteration attempts to reconstruct the

most easily-identifiable tracks, e.g. high-p_T tracks, and then removes the hits associated with those tracks. This helps simplify the track reconstruction in the following iterations.

Each iteration consists of four steps:

1. A seed is generated from a few (generally 2 or 3) hits. This seed provides an initial estimate of the trajectory and uncertainties associated with thet track.

2. A Kalman-based track finder is used to extrapolate seed trajectories along their expected paths. Additional hits that are compatabile with a path are assigned to that track candidate.

3. A track-fitting module uses a Kalaman filter and smoother to provide estimates of the trajectory parameters for each track.

4. A track selection sets quality thresholds and discards tracks that fail the specified criteria.

A detailed description of track reconstruction can be found in Reference [13].

4.1.1   Vertices

An essential part of event reconstruction is identifying which particles were produced at parton-parton interaction vertices (primary vertices) and which were produced at secondary decay vertices afterwards (secondary vertices). The process for selecting primary vertices consists of three steps: track selection, track clustering, and track fitting.

The track selection criteria is chosen in order to select tracks that are consistent with being produced promptly in the primary interaction region. Tracks are required to have a lot transverse impact parameter relative to the beam spot, a certain number of hits in the pixel and strip detectors, and good fit quality when fitting to the trajectory.

No requirement is placed on the p_{T} of tracks, in order to ensure a high reconstruction efficiency, even for minimum-bias events.

After the track selection, a deterministing annealing clustering algorithm is used in order to group together tracks that appear to originate from the same vertex [14]. Here the selected tracks are originally all assigned to the same vertex and are then slowly divided into multiple vertices. This process continues until reaching a cutoff defined by balancing the risk of incorrectly splitting vertices and the resolving power of the algorithm.

Once the track clustering is completed, an "adaptive vertex fitter" is used to determine the 3D-position of vertices with at least two tracks [15], in which tracks corresponding to a vertex are each assigned a probability of correctly belonging to the vertex. The weighted sum of these probabilities is then used in the fitting algorithm to determine primary vertices.

In this process, many more than one primary vertex are reconstructed due to multiple parton-parton interactions in an event. There is, however, usually only one primary vertex of interest in the event, corresponding to the primary vertex with the highest sum of track p_{T}^2. This primary vertex is commonly referred to as *the* primary vertex.

4.2   Calorimeter Clusters

Energy deposits in the various CMS calorimeters are clustered together to form "calorimeter clusters". The purpose of the calorimeter clusters are to aid in, detecting and measuring the energy and direction of stable neutral particles, separating these neutral particles from the energy deposits of charged hadrons, reconstructing and identifying electrons, along with their corresponding bremsstrahlung photons, and measuring the energy of charged hadrons with low-quality track parameters

The calorimeter clustering is done in three steps. First, cluster seeds are identified as cells with an energy both larger than some threshold and larger than the energy of neighboring cells. Next, these cluster seeds are formed into "topological clusters" by iteratively merging together neighboring cells that have signigicant energy deposits. In this process, topological clusters can merge such that a cluster contains multiple cluster seeds. Lastly, the energy in a topological cluster is distributed among its seeds through a Gaussian-mixture model that results in the final calorimeter clusters.

This clustering is performed separately in each subdetector calorimeter, including the ECAL preshower detector, for which its two layers are treated independently. There is no clustering performed in the HF, as each cell's short- and long-fibers measure the electromagnetic- and hadronic-energy components, as described in Sub-section 3.2.3. These components directly give rise to "HF EM" and "HF HAD" clusters.

4.3   Particle Flow

The particle-flow (PF) algorithm is a holistic approach towards event reconstruction. It combines the basic information of subdetectors, i.e. the tracks and clusters defined above, to identify each final-state particle and reconstruct their corresponding properties. By correlating measurements from the tracker and calorimeter, the PF algorithm is able to provide improved energy and momentum measurements. A complete, detailed presentation of the particle-flow algorithm is given in [16, 17, 18].

4.3.1   Linking

As a particle transverses through the CMS detector, it is expected to generate multiple input elements to the PF algorithm. Thus, the first step of reconstructing particles is to *link* the various PF elements stemming from different subdetectors together. This is

done by defining a "distance" between two linked elements, where the closer the distance the more probable it is the two elements correspond to the same particle. The linking algorithm then creates *PF blocks* by associating directly or indirectly linked elements together. The exact criteria used to link elements together and define their distance depends on the type of elements being considered and are listed below.

A link between a track and calorimeter cluster is established by extrapolating the track trajectory through the ECAL and HCAL, up to a depth where energy deposits are most expected. If the extrapolated track falls within the area of the calorimeter cluster, the two elements are linked and the link distance is defined as the separation between their positions in the (η, ϕ) plane. In the case multiple clusters are linked to the same track only the link with the smallest distance is kept.

To link tracks with clusters from potential bremsstrahlung photons, tangents to the track are extrapolated to the ECAL. If a tangent line falls within the cluster, the track and cluster are linked with the η-ϕ separation used as the link distance.

Calorimeter clusters in the HCAL, ECAL, and preshower detector are linked together when the position of a cluster in a more granular calorimeter (preshower or ECAL) is within the boundaries of a cluster with less granularity (ECAL or HCAL). The distance between these two clusters is defined as either the η-ϕ or x-y separation for HCAL-ECAL and ECAL-preshower links, respectively. In the case where multiple HCAL(ECAL) clusters are linked to the same ECAL(preshower) cluster, only the link with the smallest distance is used.

Links between a track and a standalone-muon track, defined as a track segment constructed from hits in the muon system, are established when a global fit to the two tracks has an acceptable fit quality. The link distance in this case is defined as the χ^2 of the fit, only the link with the smallest χ^2 is retained when there are multiple links to the same standalone-muon track. The resulting links are called "global muons".

4.3.2   PF Reconstruction and Identification

Once the PF blocks have been constructed, the PF algorithm is applied to reconstruct and indentify a set of particles from each PF block. This algorithm proceeds sequentially to hierarchichally reconstruct particles as described below.

First, PF muons are formed by global muons whose momentum is compatible with that determined by only using the tracker. The corresponding tracks are then removed from the PF block.

Next, PF electrons are identified by using information from the inner tracker and calorimters. Electron candidates in a PF block are seeded by tracks with links to ECAL clusters. These tracks are then re-extrapolated to the ECAL, and if a tracks is found to be compatible with ECAL energy deposits and consistent with an electron, the track and clusters are labelled a PF electron and are removed from the PF block.

The remaining elements in the PF block are used to form charged hadrons, photons, neutral hadrons, and, in rare cases, additional muons, PF elements are identified as one of these particle-types by comparing the track momentum to the linked cluster energies. The following scenarios define the identification process:

- If the total cluster energy is much smaller than the track momentum, the excess track momentum is labeled as a muon or fake track. This occurs for less than 0.03% of the tracks used in the algorithm.

- If the total cluster energy agrees within the uncertainty of the track momentum, a PF charged hadron is formed. The PF charged hadron is then assigned a mass equal to that of a charged pion and a momentum based on a fit of the tracker and calorimeter measurements.

- If the total cluster energy is significantly larger than the track momentum and

the excess is greater than the total ECAL energy, than the track is considered a PF charged hadron, as described above, the excess ECAL energy is labeled a PF photon, and the remaining energy is assigned to a PF neutral hadron. The excess ECAL energy is preferentially given to photons over neutral hadrons, as typically photons account for 25% of the energy of a jet, while neutral hadrons only account for 3%.

- If the total cluster energy is significantly larger than the track momentum and the excess is less than the total ECAL energy, the track is considered a PF charged hadron and the excess calorimeter energy is assigned as a PF photon.

- If there are ECAL or HCAL clusters without any linked tracks, the deposits are respectively treated as PF photons and PF neutral hadrons.

4.4   Leptons

As the identification criteria for selecting PF electrons and PF muons are weak, these object serve as candidate particles. To increase the purity of true electrons(muon), more stringent criteria must be passed for a PF electron(muon) to be considered an analysis-level electron(muon).

4.4.1   Electrons

The electron candidates are required to have $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 2.5$, and to satisfy identification criteria [19] designed to remove hadrons misidentified as electrons, photon conversions, and electrons from heavy-flavor hadron decays. This criteria is shown in Table 4.1, where $\sigma_{i\eta i\eta}$ is a variable based on the width of the electron shower shape,

28

Criteria	Barrel requirement	Endcap requirement
$\sigma_{i\eta i\eta}$	< 0.0101	< 0.0283
$\Delta\eta(\text{cluster}, \text{track})$	< 0.0103	< 0.07333
$\Delta\phi(\text{cluster}, \text{track})$	< 0.0336	< 0.114
$E_{hadronic}/E_{electromagnetic}$	< 0.876	< 0.0678
$\frac{1}{E} - \frac{1}{p}$ [GeV^{-1}]	< 0.0174	< 0.0898
$|d_0|$ [mm]	< 0.0118	< 0.0739
$|d_z|$ [mm]	< 0.373	< 0.602
Missing hits	≤ 2	≤ 1
Pass photon conversion	True	True

Table 4.1: Identification criteria that a PF electron must pass in order to be considered an analysis-level electron.

and d_0 and d_z are the transverse and logitudinal impact parameters, respectively, of the associated electron track.

Additionally, to preferentially select electrons that originate in the decay of W and Z bosons, electrons are required to be isolated from other PF candidates. The relative isolation of a particle I^{rel} is quantified using an optimized version of the mini-isolation variable I_{mini}. Mini-isolation is computed as the scalar sum of the p_T of charged hadrons from the PV, neutral hadrons, and photons that are within a cone of radius $R^{\text{mini-iso}}$ surrounding the electron momentum vector $\vec{p}$ in η-ϕ space [20]. The cone radius $R^{\text{mini-iso}}$ varies with $1/p_T$ according to

$$
R^{\text{mini-iso}} = \begin{cases} 0.2, & p_T \leq 50\text{GeV} \\ 10 \text{ GeV}/p_T, & 50 < p_T \leq 200 \text{ GeV} \\ 0.05, & p_T > 200 \text{ GeV}. \end{cases} \tag{4.1}
$$

The p_T-dependent cone size reduces the rate of accidental overlaps between the electron and jets in high-multiplicity or highly Lorentz-boosted events, particularly overlaps
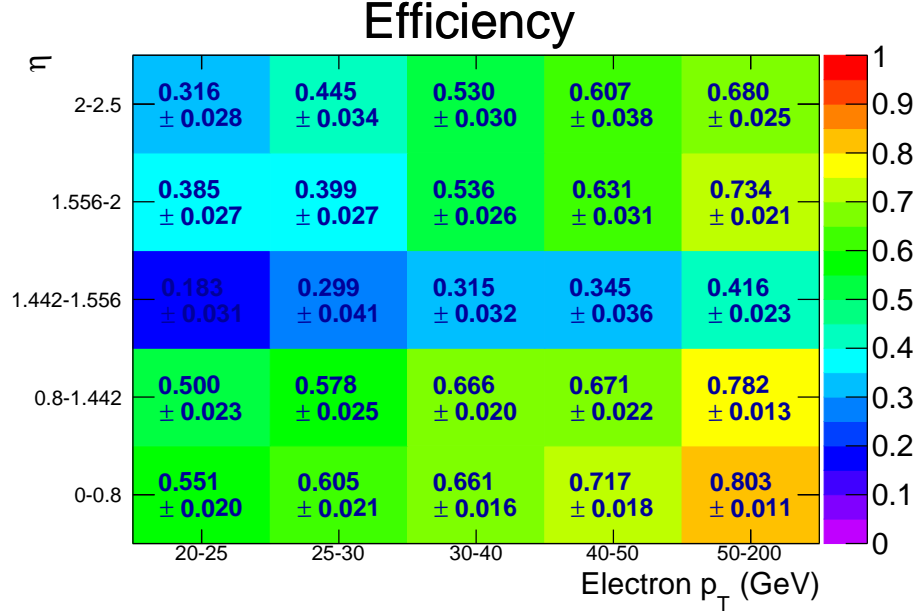
Figure 4.1: The efficiency to select an analysis-level electron as a function of p_T and η.

between bottom quark jets and leptons originating from a boosted top quark. Relative isolation is computed as $I^{\mathrm{rel}} = I_{\mathrm{mini}}/p_T$ after subtraction of the average contribution from additional proton-proton collisions in the same bunch-crossing (pileup). To be considered isolated, electrons must satisfy $I^{\mathrm{rel}} < 0.1$.

The combined efficiency for the electron reconstruction, identification, and isolation requirements, shown in Figure 4.1, is about 50% at p_T of 20 GeV, increasing to 65% at 50 GeV, and reaching a plateau of 80% above 200 GeV [21].

4.4.2   Muons

The muon candidates are required to have $p_T > 20$ GeV, $|\eta| < 2.4$, and to satisfy identification criteria [22] to select a high purity muon sample. This criteria is shown in Table 4.2, where d_0 and d_z are the transverse and logitudinal impact parameters, respectively, of the associated muon track. Analogously to electron candidates, muon

Criteria	Requirement
Is PF muon	True
Fraction of valid tracker hits	> 0.8
AND	
$|d_0|$ [mm]	< 2
$|d_z|$ [mm]	< 5
Is global muon	True
Normalized global-track χ^2	< 3
χ^2 of tracker-standalone position match	< 12
Track-kink χ^2	< 20
Segment compatibility	> 0.303
OR	
Segment compatibility	> 0.451

Table 4.2: Identification criteria that a PF muon must pass in order to be considered an analysis-level muon.

candidates are required to satisfy $I^{\mathrm{rel}} < 0.2$, where the looser threshold is to account for purity differences between electrons and muons.

The efficiency for reconstructing muons, shown in Figure **??** is about 70% at p_{T} of 20 GeV, increasing to 80% at 50 GeV, and reaching a plateau of 95% for $p_{\mathrm{T}} >$ 200 GeV [23].

4.5   Jets

When a quark or gluon is produced at the LHC, due to color confinement, they quickly hadronize and produce a collimated "spray" of particles, called a jet, which is the direct detector observable. The parameter of interest, however, is the momentum of the inital parton before hadronization. Thus, clustering the constituent jet particles in a way that accurately reconstructs the inital parton momentum is essential. Events, however, often contain multiple jets with each jet typically composed of some $\sim$ 10-100 particles that are incident on many detector channels across a large area. This makes the problem of jet clustering non-trivial and an important aspect of object reconstruction.
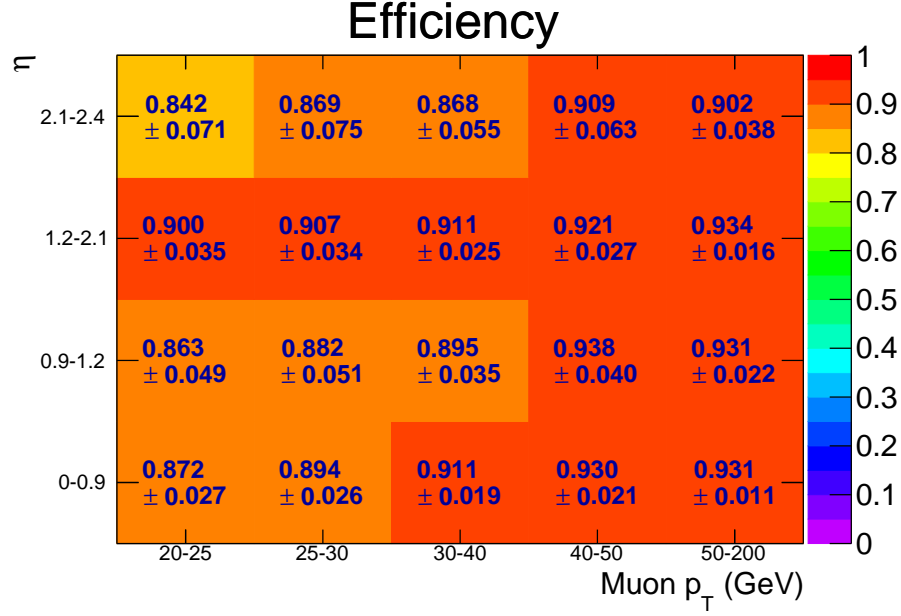
Figure 4.2: The efficiency to select an analysis-level muon as a function of p_{T} and η.

4.5.1   Clustering

Jets are not physically-defined objects, and instead it is the clustering rules that define what a jet is. The common class of clustering algorithms used are sequential recombination algorithms, which work by defining a distance measure between pairs of particles, typicaly based on energy and spatial-location, and then combining the closest pair of particles. This process proceeds sequentially and ends when some threshold condition is met.

In particular, the distance measures, d_{ij}, which is the distance between two entities (particle or psuedojet), and d_{iB}, which is the distance between an entity and the beam, are often defined to be

$$d_{ij} = \min(p_{\mathrm{T}i}^{2p}, p_{\mathrm{T}j}^{2p})\frac{\Delta R_{ij}^2}{R^2}, \tag{4.2}$$

$$d_{iB} = p_{\mathrm{T}i}^{2p}, \tag{4.3}$$

where $\Delta R_{ij}^2 = \Delta \eta_{ij}^2 + \Delta \phi_{ij}^2$, R is the jet radius parameter that sets the scale of the jet's size, and p which is a parameter of the algorithm. The clustering procedure then proceeds by finding the smallest of the distances, and if it is a d_{ij}, the two entities i and j are merged, while if it is a d_{iB}, entity i is called a jet and is removed from the clustering list.

The parameter p determines the clustering order of the algorithm and thus different choices for the value of p lead to different clustering algorithms. For $p = 1$, the clustering follows the k_{T} algorithm [24], which prioritizes clustering softer particles first. Setting $p = 0$, reproduces the Cambridge-Aachen algorithm [25], which uses energy-independent clustering, relying only on the spatial distances between particles. Lastly, choosing $p = -1$, gives the anti-k_{T} algorithm [26], which favors using the harder particles as the jet seeds and then clustering around them. A more detailed disucssion of jets, including a comparison of these and other clustering algorithms can be found in Reference [27].

4.5.2   Selection

The jets used in this dissertation are constructed by clustering PF candidates with the anti-k_{T} algorithm and $R = 0.4$, using the FastJet package [28]. To reduce the effect of pileup on the jet clustering and energy measurements, a process called "Charged Hadron Subtraction" is applied, where PF charged hadrons that do not originate from the PV are not included in the jet clustering. In addtion, to remove the neutral energy component of pileup, the contribution from PF neutral hadrons produced by pileup is estimated based on the area of a jet and the energy density of the event and subtracted from the jet [29]. Lastly, to prevent double-counting, any jets which contain a PF candidate identified as

Criteria	Requirement
Number of constituents	> 1
Charged multiplicity	> 0
Neutral electromagnetic fraction	< 0.99
Neutral hadron fraction	< 0.99
Charged electromagnetic fraction	< 0.99
Chaged hadron fraction	> 0

Table 4.3: Identification criteria that a jet candidate must pass in order to be considered an analysis-level jet.

an analysis-level electron or muon are removed from the jet collection.

Finally, to be considered an analysis-level jet, the jets must have $p_{\mathrm{T}} > 30$ GeV, $|\eta| \leq 2.4$, and must pass a set of loose identification requirements [30, 31] to surpress, for example, calorimeter noise. These requirements are shown in Table 4.3. The resulting jets are considered to be "small-R" jets.

4.5.3   b-tagging

Jets that are formed through b-quark hadronization have many unique properties that allow them to be differentiated from jets formed through other quarks or gluons. This abiilty to "b-tag" jets is a very useful tool for determining what physics processes occured in an event, as b quarks are associated with specific physics process, such as top quark decays. At the same time, many SUSY scenarios, where naturalness considerations motivate light third-generation squarks, result in either the direct or indirect production of b quarks through the decay of SUSY particles. Thus, b-tagging jets is not only often a crucial component to reducing backgrounds from other processes where no b quarks are expected, but also a powerful selector for potential signal events.

This analysis uses the Combined Secondary Vertex v2 (CSVv2) algorithm [32, 10], which utilizes the long lifetimes, large masses, high-momentum daughter particles, and frequent semi-leptonic decays typical of b hadrons to b-tag jets. As the b quark can only

decay to an up or charm quark through highly Cabibbo surpressed weak interactions, b hadrons tend to have long lifetimes, typically on the order of 150 ps. Because of this, b mesons can travel a few mm to a cm from the PV before decaying and producing displaced tracks from which a secondary vertex can be reconstructed. In addition, due to the relatively large b-quark mass, b mesons tend to be heavy, which leads to both large secondary vertex masses and daughter particles with a hard momentum spectrum. Lastly, the weak decay of the b quark results in an associated electron or muon in about 20% of decays. The presence of these soft, nonisolated leptons provides an additional marker for the presence of a b-quark.

The CSVv2 algorithm exploits variables based on information about secondary vertices, their associated displaced tracks, and the presence of soft leptons to accurately tag b-quark jets. A selection of the variables with high discrimination are listed below:

- The significance of the flight distance in the transverse plane.

- The number of SV

- The SV mass

- The number of tracks associated with the SV

- The ratio of the transverse momentum of the SV tracks and the transverse momentum of the jet

- The 3D impact parameter of soft leptons associated with the jet

These variables are then fed into a multilayer perceptron with one hidden layer then outputs a score between 0 and 1, indicating the likelihood the jet is a b-quark jet. The distribution of CSVv2 disciminator values for different flavor jets is shown in Figure 4.3 [10].
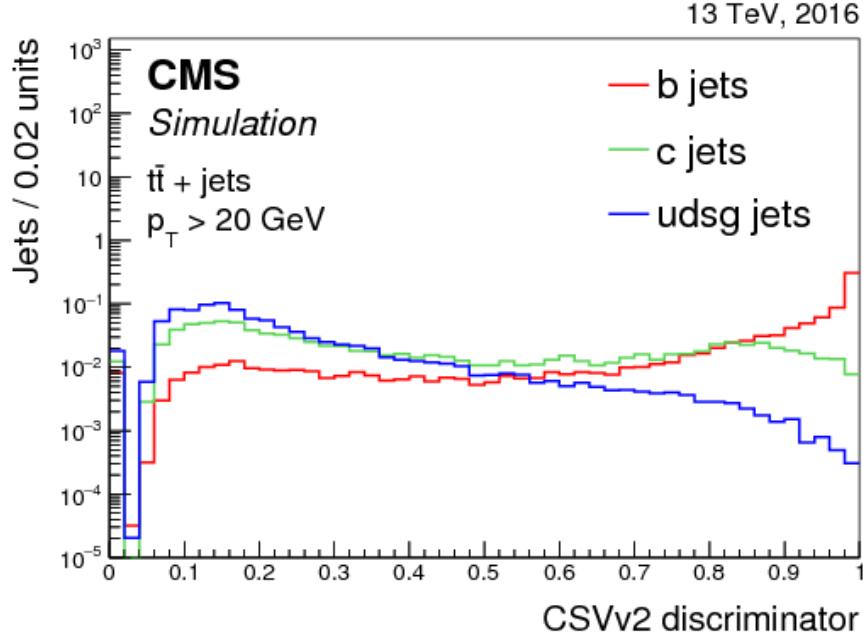
Figure 4.3: The distribution of the CSVv2 discriminator values for jets of different flavors. Jets are selected from $t\bar{t}$ events and required to have $p_\mathrm{T} > 20$ GeV [10].

A threshold score for b-tagging a jet is chosen such that the mis-tag rate for light-flavor jets is 1%. This corresponds to a mis-tag rate for c-quark jets of 13–15% (11–13%) in the barrel (endcap) and a tagging efficiency for b jets of 60–67% (51–57%) in the barrel (endcap) for jets with p_T between 30-50 GeV. The tagging efficiency increases with p_T before decreasing to $\approx 50\%$ for jets above 150 GeV. The b-tagging efficiency as a function of jet p_T is shown in Figure 4.4.

4.6   Large-radius Jets

While the distance parameter of small-R jets is optimized for clustering the hadronization products of a single parton, it is often useful to exploit information of physical processes on scale larger than a single jet, such as top quark or W boson decays. One way to capture this information is to cluster jets with a large distance parameter, which encodes
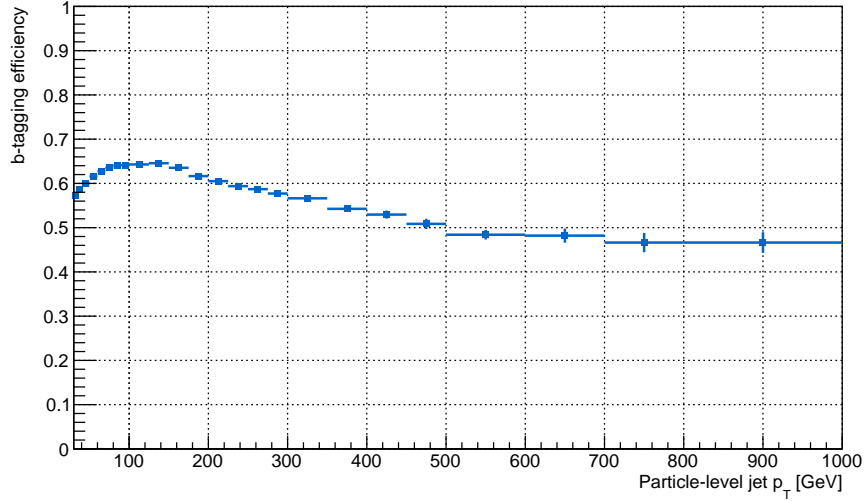
Figure 4.4: The efficiency of the CSVv2 algorithm as a function of jet p_{T} at the working point used in this analysis.

the momentum, angular, and multiplicity information of the partons contained in the jet.

In particular, this analysis constructs these "large-R" jets by clustering "small-R" jets with a distance parameter of $R = 1.2$. Due to the relatively small number of small-R jets in an event ($\lesssim 10$), the construction of these large-R jets is insensitive to the clustering algorithm, and the anti-k_{T} algorithm is used. While these large-R jets could be constructed by performing the clustering at the PF candidate level, no significant improvement in performance was noted. Thus, clustering small-R jets were chosen for the following practical considerations. First, the FASTJET implementation of the anti-k_{T} algorithm has complexity $\mathcal{O}(n \log n)$ [33], which results in a speed-up of on the order of 100x when clustering small-R jets ($\lesssim 10$ objects) compared to PF candidates (~ 1000 objects). Secondly, large-R jets clustered from PF candidates would require the computation of new energy-measurement and pileup-removal calibrations for jets of this specific radius. Small-R jets, however, already incorporate standardized calibrations for $R = 0.4$ jets. Thus, by clustering the calibrated small-R jets, the large-R jets already

37

incorporate corrections without the need of any additional development.

In addition to small-R jets, the jets associated with selected leptons are included in the formation of large-R jets in order to capture as much of the kinematic information as possible. For example, this helps reduce the difference between large-R jets formed by clustering hadronic and (semi-)leptonic decays, as by including the lepton, the only information difference between the two scenarios is due to the undetected neutrino(s).

This technique of clustering small-R jets into large-R jets has been used previously by both the ATLAS and CMS collaboration, e.g. References [34, 35].

Part III

Data and Simulation

Chapter 5

Data samples and simulation

5.1 Data

The dataset used in this search corresponds to 35.9 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV collected by the CMS detector over the year 2016. This is a subset of the 40.8 fb^{-1} delivered by the LHC and selected to correspond to when all sub-detectors were fully-operational. A plot of the cumulative delivered and recorded integrated luminosity by the LHC and CMS, respectively, is shown in 5.1.

5.2 Monte Carlo Simulation

Simulated samples are used to model both SM and BSM physics processes, and are extremely useful in the design and execution of new physics searches. In this particular analysis the Monte Carlo simulations (MC) are used to in the following ways:

- Design and optimization of the analysis strategy

- Validation of (signal-plus-)background prediction methods

- Study of processes with impure and/or statisticallly small control regions

- Commissioning and understanding of collected data
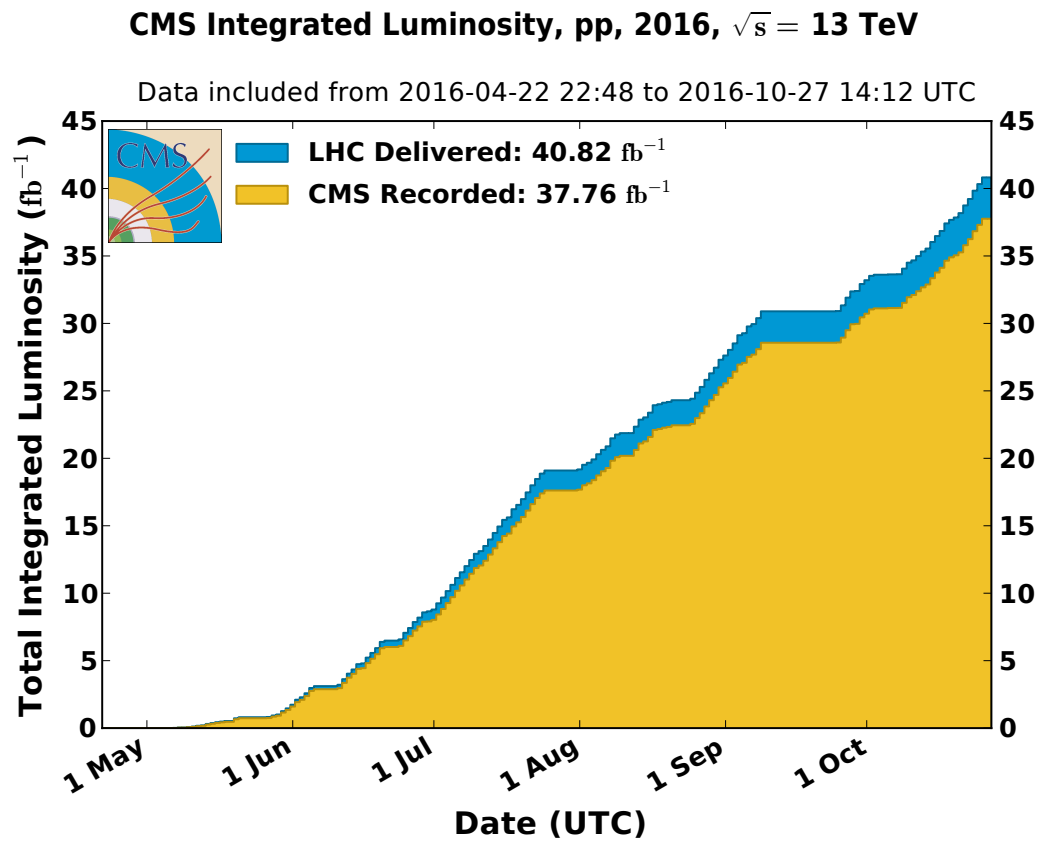
Figure 5.1: Delivered and recorded integrated luminosity by the LHC and CMS, respectively, over 2016. [11]

- Modelling BSM physics processes that may or may not exist

MadGraph5_aMC@NLO 2.2.2 is used in leading-order mode [36, 37] to generate the $t\bar{t}$, W + jets, quantum chromodynamics multijet (QCD), and Drell–Yan background processes with extra partons. Comparison to a Powheg 2.0 [38, 39, 40] sample generated at next-to-leading order (NLO) shows that the NLO effects do not have a significant impact. The $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}t\bar{t}$, and t-channel single top quark production backgrounds are generated with MadGraph5_aMC@NLO 2.2.2 in NLO mode [41], while the tW, $\bar{t}W$, and s-channel single top quark processes are generated with Powheg 2.0. The $t\bar{t}$, W + jets, and QCD samples are generated with up to 2, 4, 2 extra partons, respectively. All samples are generated using a top quark mass of 172.5 GeV and with the NNPDF3.0 set of parton distribution functions (PDF) [42]. For the fragmentation and showering of partons, the generated samples are interfaced with Pythia 8.205 [43] and use the CUETP8M1 tune to describe the underlying event [44]. All samples use the highest precision cross sections available [45, 46, 47, 48, 49, 50, 51]. The detector response is simulated with Geant4 [52]. Simulated samples are processed through the same reconstruction algorithms as the data.

The signal samples are generated with up to two extra partons in leading-order mode and dynamic factorization and renormalization scales by MadGraph5_aMC@NLO 2.2.2. The same fragmentation, parton showering, simulation, and event reconstruction procedure as for the background samples is used. The samples are normalized to NLO + next-to-leading logarithmic cross sections [12].

Chapter 6

Event Selection

6.1 Baseline selection

One of the main challenges for a SUSY search is that the ratio of SM events to SUSY events is (ROUGHLY) 10 billion to 1. To surmount this problem, it is paramount to develop highly efficient signal-to-background discriminators. Luckilly, SUSY signatures typically have characteristics unlike most SM processes. In particular, for the T1tbs process, events are expected to have a large number of jets, many of which are b-quark jets, resulting in large amount of hadronic energy. Additionally, the mass scale of the event is expected to be larger than most SM events due to the high masses of the gluinos (i.e. $\tilde{1}$ TeV). These features are used to construct the "baseline selection", defined as a set of requirements that events must pass in order to be included in the analysis. Here the baseline selection is defined as $N_{\mathrm{leps}} = 1$, $H_{\mathrm{T}} > 1200$ GeV, $M_{\mathrm{J}} > 500$ GeV, $N_{\mathrm{jets}} \geq 4$, and $N_{\mathrm{b}} \geq 1$. Figure **??** shows the "N-1" distributions of these variables, which are plots showing the 1D distribution of a variable with the baseline selection applied, except for the requirement corresponding to the plotted variable. Figure **??** shows a "cutflow" table, which depicts the expected yields for each process as each requirement of the baseline selection is indiviually applied. Note that this analysis explicitly requires exactly 1 lepton (defined as a muon or electron). As can be seen in 0-lepton bin of the N-1 plot of N_{leps},

there is still significant amounts of QCD production compared to the expected signal yield. Requiring exactly 1 lepton reduces the background by XX%, while only reducing the signal by YY% compared to being inclusive in N_{leps}. An additional benefit of this selection is that the SM backround is dominated by a single process ($t\bar{t}$), which reduces the complexity of the background prediction. Including additional N_{leps} regions is being investigated for future iterations of the analysis. A final note of interest is that there is no requirement on the $E_{\mathrm{T}}^{\mathrm{miss}}$, making this analysis sensitive to BSM models other than RPV SUSY that produce either little or no $E_{\mathrm{T}}^{\mathrm{miss}}$ in an event.

A final requirement for the baseline selection is that events must pass a series of filters designed to remvoe poorly reconstructed events. These standard filters remove events with noise in the HCAL or ECAL, beam halo effects, jets that fail to pass quality criteria, and events with zero good primary vertices.

6.2   Trigger

In order to select events in data that pass the baseline selection, events are required to either have an online-H_{T} of at least 900 GeV or at least one jet with online-p_{T} above 450 GeV. Figure **??** shows the performance of the trigger during the first XX fb^{-1} (Runs B-G) and the remaining YY fb^{-1} (Run H). During Runs B-G the trigger plateaus at YYYY GeV with 100% efficiency, during Run H, however, a bug in the trigger implementation caused very high-p_{T} jets to not be included in the online-H_{T} calculation, which resulted in a significantly plateau efficiency of YY%. In order to recover the lost efficiency, a trigger requiring at least one jet with online-p_{T} above 450 GeV is "OR"'d with the H_{T} trigger With this addition, the trigger effieciencies for background and signal are ¿99% for events with H_{T} ¿1200 GeV, as shown in Figure **??**.

44

Figure 6.1: Illustration depicting the $(N_{\mathrm{jets}}, M_{\mathrm{J}})$ binning after the baseline selection, with control and signal region bins denoted by "CR" and "SR", respectively.

6.3   Analysis Binning

After the baseline selection, the background is dominated by $t\bar{t}$ events with small contributions from W + jets and QCD production. There are additional rare background processes, jointly noted as "Other", with tiny, but non-zero contributions that arise from single top quark, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$, $t\bar{t}t\bar{t}$, and Drell-Yan production.

In order to further increase the signal-to-background ratio, as well as create background-dominated control regions, the analysis region is binned with respect to N_{jets} and M_{J}. The N_{jets} bins are defined as $4 \leq N_{\mathrm{jets}} \leq 5$, $6 \leq N_{\mathrm{jets}} \leq 7$, and $N_{\mathrm{jets}} \geq 8$. Each N_{jets} bin is further split into bins of $500 < M_{\mathrm{J}} \leq 800$ GeV, $800 < M_{\mathrm{J}} \leq 1000$ GeV, and $M_{\mathrm{J}} > 1000$ GeV, with the exception of the $4 \leq N_{\mathrm{jets}} \leq 5$ bin for which the two highest M_{J} bins are combined due to the limited data sample size in the $M_{\mathrm{J}} \geq 1000$ GeV bin. A diagram representing this binning is shown in Figure 6.1 The low-N_{jets}, low-M_{J} bins are expected to be background-dominated and are used as control regions for constraining sytematics and for validating the prediction methodology, while the high-N_{jets}, high-M_{J} bins.

Within each N_{jets} and M_{J} bins, the N_{b} distribution is examined for evidence of new physics and is separated into bins of $N_{\mathrm{b}} = 1$, 2, 3, and ≥ 4. The two lowest N_{b} bins are used to provide constraints on the background normalizations and systematic uncer-tainites, while the higher N_{b} bins are the most sensitive to potential signals due to its larger signal-to-background ratios.

In total, this analysis has 8 kinematic regions–3 control and 5 signal regions with four N_{b} bins per kinematic region. The simulated N_{b} distribution for the SM background processes and a signal model with $m_{\tilde{g}} = 1600$ GeV for each kinematic region is shown in

Figure **??** and the corresponding yields are given in Table **??**.

Part IV

The Search

Chapter 7

Background Prediction

7.1 Overview

This analysis seeks to find evidence of new physics by searching for deviations from the SM in the N_b distribution. In order to do this, it is essential to be able to robustly and accurately predict both the normalization and shape of the N_b distribution. To obtain these predictions, a global maximum-likelihood fit is performed. This fit is carried out both for a background-only hypothesis and for signal-plus-background hypotheses, in which a signal contribution is extracted in addition to the contributions of SM background processes. The model is constructed using the poisson probabilities of the bin contents of the N_b distribution for all N_{jets}, M_J regions, while systematic uncertainties are applied as nuisance parameters.

As the kinematic tails of the N_{jets} and M_J variables are difficult to model reliably, the $t\bar{t}$ and QCD normalizations are individually allowed to (almost) freely vary in each (N_{jets}, M_J) bin. The $t\bar{t}$ normalizations are constrained in each bin by the low-N_b bins, while the QCD normalizations are constrained by control regions with no identified leptons ($N_{leps} = 0$). The overall W + jets normalization is determined from data and is allowed to vary across N_{jets} bins by amounts measured using a kinematically similar Z + jets sample, while the normalization of Other is largely taken from simulation, as its contribution is

48

small in the regions considered. Further details on the measurement of the normalizations are given in the following sections.

Once the SM backround processes are normalized accordingly, further corrections to the N_b shape are relatively small. The nominal N_b shape prediction for each process is taken from simulation with data-to-simulation correction factors (SFs) applied for the tagging efficiency of heavy- and light-flavor jets [32, 53]. This shape is allowed to vary in order to assess the impact of mismodeling of relevant parameters, such as the rate of gluon splitting to $b\bar{b}$ and the b-tagging SFs. The appropriate ranges for these parameters are determined based on measurements in dedicated control samples and then constrained by a simultaneous fit across all bins of N_{jets} and M_J in a correlated manner. A detailed discussion of these variations and their measurements is given in **??**.

7.2 $t\bar{t}$ and QCD Normalizations

The $t\bar{t}$ and QCD normalizations are allowed to float in each (N_{jets}, M_J) bin but with a loose constraint across M_J bins discussed in the following subsection. The largest constraint on the $t\bar{t}$ normalization in each bin is the background-dominated $N_b \leq 2$ bins, while the QCD normalization in each bin is mostly constrained by corresponding bins in a similar 0-lepton kinematic region selected by requiring $N_{\text{leps}} = 0$, $H_T > 1500$ GeV, $M_J > 500$ GeV, $N_{\text{jets}} \geq 6$, and $N_b \geq 1$. The higher H_T requirement compared to the analysis's baseline selection is imposed in order to account for the extra energy in an event carried by the lepton in the $N_{\text{leps}} = 1$ selection, while the higher N_{jets} selection is imposed in order to account for differences in the N_{jets} distribution between the $N_{\text{leps}} = 1$ and $N_{\text{leps}} = 0$ samples. This control sample follows the same kinematic binning as the $N_{\text{leps}} = 1$ regions, except that the N_b distribution in each bin is integrated in N_b for $N_b \geq 1$ and each bin's N_{jets} requirement is increased by 2. A diagram representing the

Figure 7.1: Diagram depicting the $(N_{\text{jets}}, M_{\text{J}})$ binning of the $N_{\text{leps}} = 0$ QCD control region.

binning of the $N_{\text{leps}} = 0$ control sample is shown in Figure 7.1 The QCD contribution in a particular $N_{\text{leps}} = 1$ bin is then constrained by the corresponding $N_{\text{leps}} = 0$ bin. To avoid biasing the normalization measurement, the small contribution of $t\bar{t}$ background to the $N_{\text{leps}} = 0$ control regions is included using the normalization from the corresponding $N_{\text{leps}} = 1$ bins, while contributions from other processes are taken from simulation.

7.2.1 M_{J} Connection

Due to the large freedom of unconstrained normalization parameters, the fit can be sensitive to rare statistical fluctuations and return unphysical normalization values particurlarly in bins dominated by $t\bar{t}$ events. For example, in psuedodata experiments, where data are generated according to the statistical and systematic uncertainties of the pre-fit values, the fit reduced the $t\bar{t}$ contribution in the $N_{\text{jets}} \geq 8$, $M_{\text{J}} > 1000$ GeV bin (where statistical uncertainties are largest) to ~ 0 in about $\sim 1\%$ of the experiments. This can be seen in Figure 7.2 (left) which shows a low tail in the distribution of post-fit $t\bar{t}$ yields in the $N_{\text{jets}} \geq 8$, $M_{\text{J}} > 1000$ GeV bin for 1,000 psuedodata experiments. When yields in a bin have a large fluctuation downwards, the fit must lower the normalization of a process to compensate. The QCD, $W +$ jets, and Other contributions, however, are largely constrained by other data control samples or taken from simulation, and so the fit uses the freedom to adjust the $t\bar{t}$ normalization in order to compensate for the fluctuation, leading to the unphysically small values.

In order to avoid this instability, the normalizations of $t\bar{t}$ and QCD are (independently) connected by log-normal constraints between adjacent M_{J} bins. By correlating the normalizations across M_{J} bins, the fit's sensitivity to large fluctuations in a single
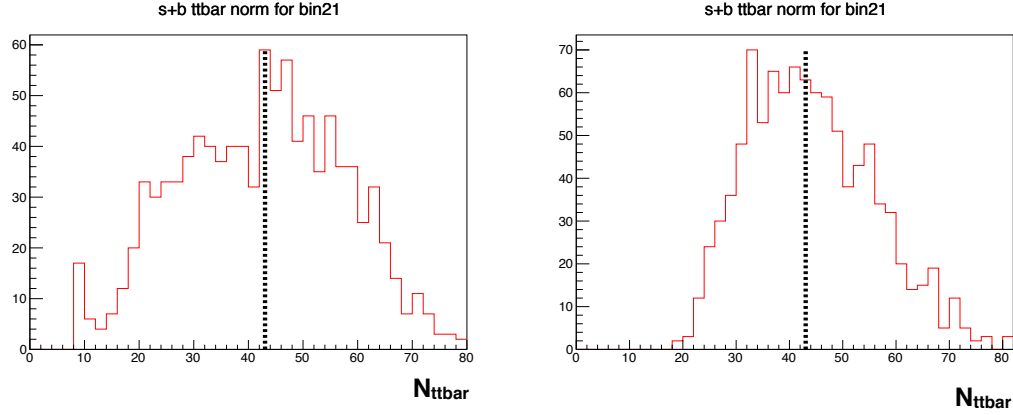
Figure 7.2: Distribution of post-fit yields of $t\bar{t}$ in the $N_{\text{jets}} \geq 8$, $M_{\text{J}} > 1000$ GeV bin for 1,000 psuedodata experiments without (left) and with (right) constraints between adjacent M_{J} bins. The dotted black line indicates the the pre-fit yield.

bin is greatly reduced. The size of these connections is motivated by measurements of the data-to-simulation ratio with $N_{\text{b}} = 1$ events (in order to avoid potential signal contamination) and is particuary chosen to be significantly larger than the uncertainty on the data-to-simulation ratios in order to avoid over-constraining the normalization parameters, while still providing some constraint against unphysical fits. Based on these measurements, shown in Figure 7.3, and criteria, a connection size between adjacent bins of [50%-200%] is chosen.

Figure **??** (right) shows the results of the same 1,000 psuedodata experiments but now with this constraint across M_{J} bins applied. The resulting distribution of post-fit $t\bar{t}$ yields now shows no evidence of unphysical normalizations and appears to be better behaved.

7.3   W + jets **Normalization**

The W + jets background is determined in the fit with one global normalization parameter and two parameters to adjust the bin-to-bin normalization of adjacent N_{jets} bins,
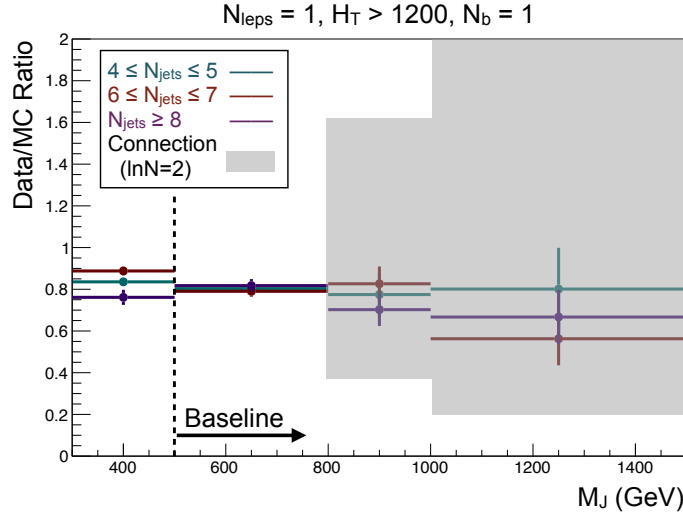
Figure 7.3: Data-to-simulation ratios as a function of M_J for different N_{jets} bins (data points) with a selection of $N_{leps} = 1$, $H_T > 1200$ GeV, and $N_b = 1$ applied. The shaded region corresponds to the size of the M_J connection in each M_J bin.

since the N_{jets} sahpe may not be well-modelled by simulation. The amount the N_{jets} shape may vary is based on the data-to-simulation agreement in a kinematically similar $Z + $ jets sample selected with $N_{leps} = 2$ (ee or $\mu\mu$), $H_T > 1200$ GeV, $M_J > 500$ GeV, $N_b = 1$, and $80 < m_{\ell\ell} < 100$ GeV, where $m_{\ell\ell}$ is the invariant mass of the two leptons. The N_{jets} distribution and data/simulation yields ratio for this sample are shown in Figure 7.4. The resulting variation sizes are 17% between $4 \leq N_{jets} \leq 5$ and $6 \leq N_{jets} \leq 7$ and 62% between $6 \leq N_{jets} \leq 7$ and $N_{jets} \geq 8$. After correcting the N_{jets} spectrum, the residual M_J mismodeling is expected to be small, so no further correction is applied.

7.4   Other Normalization

The nominal normalization for Other is largely taken from simulation, as its contribution is less than 20% in every bin with typical values $\lesssim 5\%$. It is, however, allowed to vary according to statistical and systematic uncertainties.
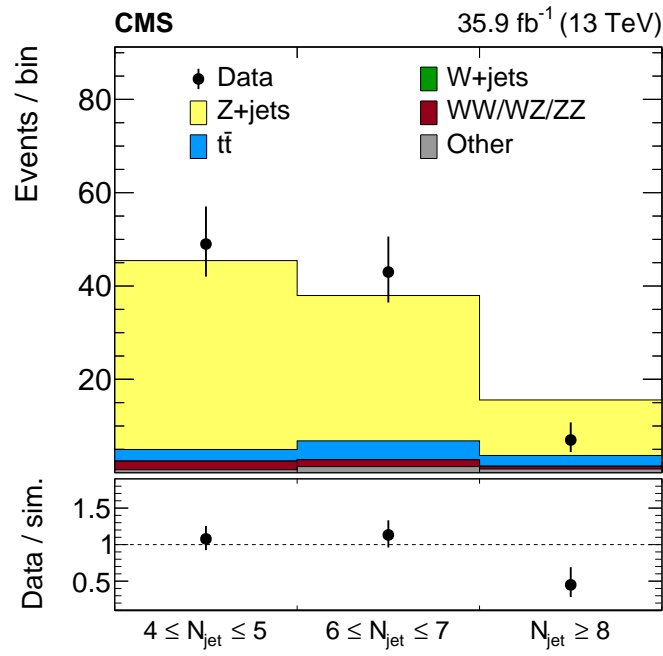
Figure 7.4: Jet multiplicity distribution for data and simulation in a $Z + \text{jets}$ control sample selected by requiring $N_{\text{leps}} = 2$, $H_{\text{T}} > 1200$ GeV, $M_{\text{J}} > 500$ GeV, $N_{\text{b}} = 1$, and $80 < m_{\ell\ell} < 100$ GeV. The total yield from simulation is normalized to the number of events in data. The uncertainty in the ratio of data to simulation yields (lower panel) is statistical only.

Chapter 8

Systematic Uncertainties

The nominal simulated shape of the N_b distribution is allowed to vary by the inclusion of systematic uncertainties. Each uncertainty is incorporated in the fit with template N_b histograms to account for the effects of the systematic variation and a nuisance parameter θ to control the variation amplitude. The nuisance parameters are subject to Gaussian constraints, normalized so that $\theta = 0$ corresponds to the nominal N_b shape and $\theta = \pm 1$ corresponds to ± 1 standard deviation (s.d.) variation of the systematic uncertainty. These uncertainties affect only the N_b shape for $t\bar{t}$, QCD, and W+jets backgrounds, because their normalizations are determined from data, while for the other (subleading) backgrounds the uncertainties affect both the N_b shape and normalization.

8.1 Gluon Splitting

The primary source of systematic uncertainty is on the modelling of the rate of gluon splitting, as events with a gluon splitting to $b\bar{b}$ provide an additional source of b quarks in events. As this process may not be properly simulated, constraining the splitting rate in data is crucial for establishing a robust prediction of the N_b distribution. The dominant contribution of this effect is due to gluons that split specifically to b quark pairs, so the phrase "gluon splitting" will hereafter refer exclusively to gluon splitting to $b\bar{b}$. One way
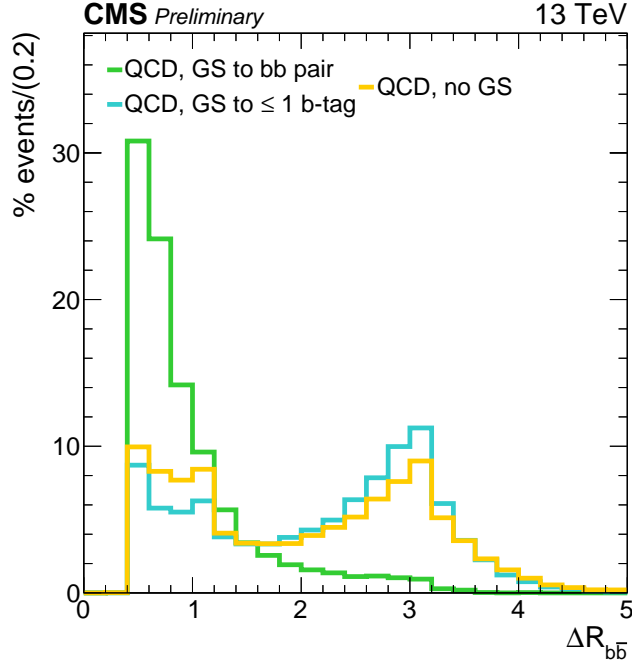
Figure 8.1: The $\Delta R_{b\bar{b}}$ distribution shapes for the three gluon splitting categories: Events with a pair of b-tagged jets resulting from gluon splitting (green), events with a gluon splitting yielding fewer than 2 b-tagged jets (blue), and events without a gluon splitting to $b\bar{b}$. These events are selected by requiring $N_{\mathrm{leps}} = 0$, $H_{\mathrm{T}} > 1500$ GeV, $M_{\mathrm{J}} > 500$ GeV, $N_{\mathrm{jets}} \geq 4$, and $N_{\mathrm{b}} = 2$.

to select a data sample enriched in gluon splitting events is to use the $\Delta R_{b\bar{b}}$ distribution, where $\Delta R_{b\bar{b}}$ is defined as the ΔR between two b-tagged jets, as pairs of b-tagged jets resulting from the same gluon splitting tend to have smaller values of $\Delta R_{b\bar{b}}$ than pairs resulting from hard scatter b-quarks or fakes. This can be seen in Figure 8.1, which shows the dRbb distribution in simulated QCD events with $N_{\mathrm{b}} = 2$ for three important categories: Events that have a correlated pair of b-tagged jets originating from a gluon splitting (green, denoted GSbb) populate the low-$\Delta R_{b\bar{b}}$ region, while events without gluon splitting (yellow, denoted noGS) or where the splitting yields zero one b-tagged jets (blue, denoted GSb) populate the low-and high-$\Delta R_{b\bar{b}}$ regions roughly equally.

Gluon splittings can contribute less than 2 b-tagged jets either because the quarks
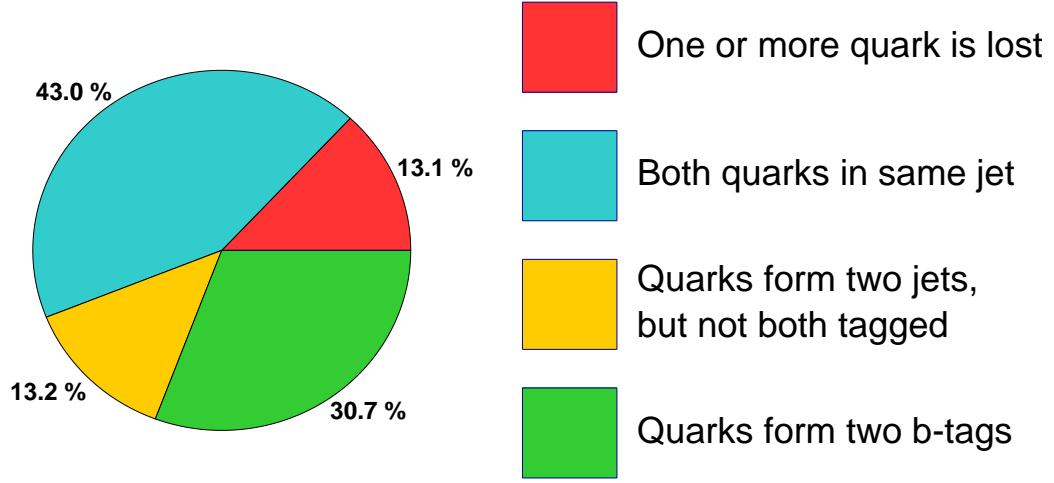
Figure 8.2: The relative fraction of the possible final states that occur from gluon splitting to b$\bar{\text{b}}$ for events satisfying $N_{\text{leps}} = 0$, $H_{\text{T}} > 1500$ GeV, $M_{\text{J}} > 500$ GeV, $N_{\text{jets}} \geq 4$, and $N_{\text{b}} = 2$.

are collimated into a single jet, one of the b-tagged jets is not tagged, or because one of the jets fails to pass the jet selection criteria, typically because it is too soft. The relative fractions of these contributions is shown in Figure 8.2.

The gluon splitting rate is the constrained by fitting the $\Delta R_{\text{b}\bar{\text{b}}}$ distribution by using the difference in shapes of the GSbb, GSb, and noGS categories. This fit varies the normalization of the GSbb and GSb (varied together) and the noGS contributions in order to extract the relative contributions of events with and without a gluon splitting. It is performed in four equal bins in the range of $0 \leq \Delta R_{\text{b}\bar{\text{b}}} < 4.8$ with events selected by requiring $N_{\text{leps}} = 0$, $H_{\text{T}} > 1500$GeV, $N_{\text{b}} = 2$, $N_{\text{jets}} \geq 4$, and $M_{\text{J}} > 500$ GeV as the gluon splitting signal in a $N_{\text{leps}} = 1$ control sample is contaminated by b quarks from the decay of top quarks. Additionally, the $N_{\text{leps}} = 0$ control sample is formed from a subset of the data that is selected to be most stable in the b tagging algorithm performance, since the precision of the $\Delta R_{\text{b}\bar{\text{b}}}$ fit is not limited by the data sample size. This choice isolates the physical effects of gluon splitting from the potential time dependence of the

	Nominal	$M_{\mathrm{J}} > 800$ GeV	$4 \le N_{\mathrm{jets}} \le 5$	$6 \le N_{\mathrm{jets}} \le 7$	$8 \le N_{\mathrm{jets}} \le 9$	$N_{\mathrm{jets}} \ge 10$
GS	0.77 ± 0.09	0.70 ± 0.38	0.80 ± 0.32	0.76 ± 0.14	0.75 ± 0.16	0.95 ± 0.36
No GS	1.21 ± 0.08	1.28 ± 0.35	1.15 ± 0.26	1.22 ± 0.13	1.24 ± 0.15	1.05 ± 0.36

Table 8.1: Gluon splitting weights derived in the nominal fit, a variation with a requirement of $M_{\mathrm{J}} > 800$ GeV, and 4 variations in bins of N_{jets} (with the nominal $M_{\mathrm{J}} > 500$ GeV requirement.)

b tagging performance due to variations in experimental conditions, which are separately incorporated by the b-tag scale factor uncertainties.

The $\Delta R_{\mathrm{b\bar{b}}}$ fit extracts a weight of 0.77 ± 0.09 for gluon splitting events and a weight of 1.21 ± 0.08 for non-gluon splitting events. The post-fit distributions are shown in Figure 8.3. The GSbb and GSb categories are plotted separately to demonstrate the difference in shapes. The discrepancy in the last bin does not significantly impact the fit because the higher yield bins at lower values of $\Delta R_{\mathrm{b\bar{b}}}$ constrain the fit. The deviations of these weights from unity, summed in quadrature with their post-fit uncertainty, are used to form the ± 1 s.d. variations of the gluon splitting rate nuisance parameter by applying weights of 1 ± 0.25 to gluon splitting events and 1 ∓ 0.22 to non-gluon splitting events in an anti-correlated manner. The fit results are used as a measure of the uncertainty on modelling of the GS rate as opposed to a correction to the central value, since the $\Delta R_{\mathrm{b\bar{b}}}$ variable may not be a perfect proxy for the GS rate. Figure 8.4 shows the effect of the ± 1 s.d. variations on the N_{b} distribution of $t\bar{t}$ for the two most sensitive bins.

In order to test the stability of the fit results and the dependence of the gluon splitting weights across kinematic regions, the $\Delta R_{\mathrm{b\bar{b}}}$ fit is repeated both with a higher M_{J} threshold and with different N_{jets} bins. The resulting weights are shwon in Table ?? and are all consistent with those of the nominal fit.
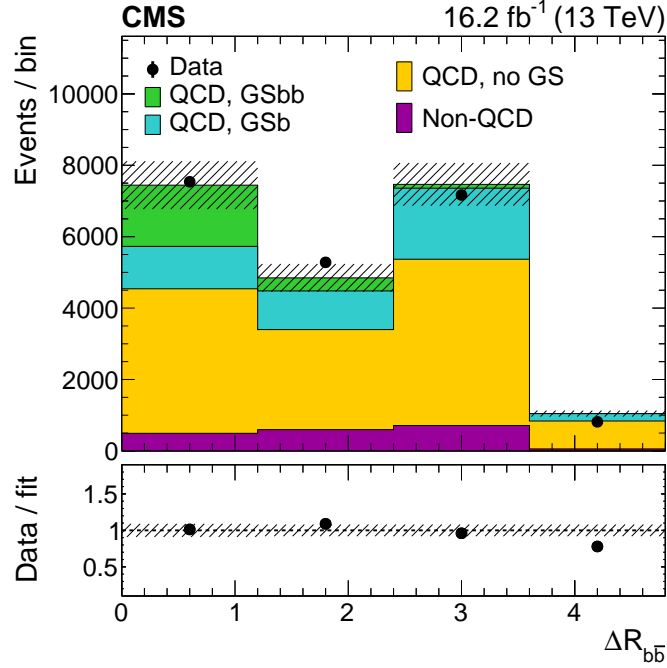
Figure 8.3: Post-fit $\Delta R_{b\bar{b}}$ distributions in a selection with $N_{\text{leps}} = 0$, $H_T > 1500$ GeV, $M_J > 500$ GeV, $N_{\text{jets}} \geq 4$, and $N_b = 2$ with the post-fit uncertainty represented by a hatched band. The ratio of data to simulation yields is shown in the lower panel.
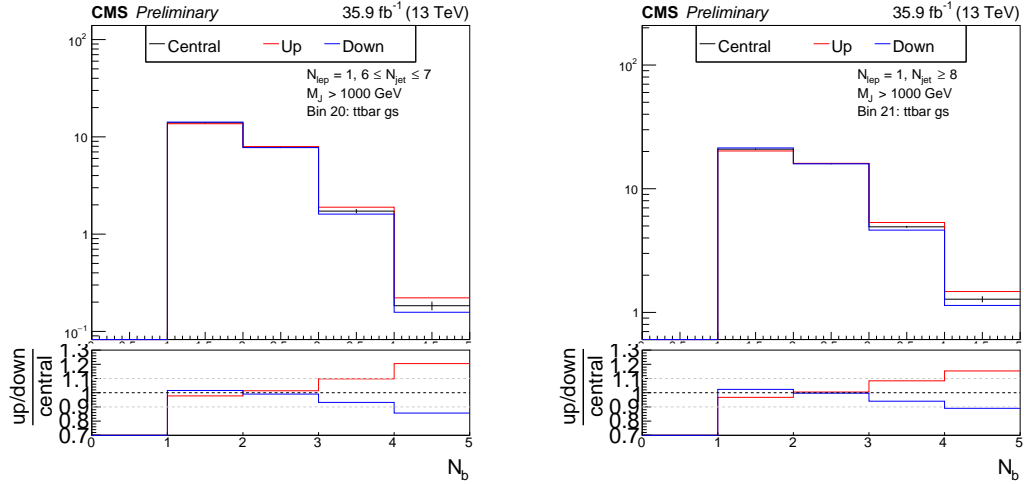
Figure 8.4: Effect of the ± 1 s.d. variations of the gluon splitting rate on the N_b distribution in $t\bar{t}$ normalized to 35.9 fb^{-1} for the two most sensitive bins: ($N_{\text{jets}} \geq 8$, $800 < M_J \leq 1000$ GeV) (left) and ($N_{\text{jets}} \geq 8$, $M_J > 1000$ GeV) (right).

8.2   b-tagging Data-to-simulation Scale Factors

Another significant systematic uncertainty is the uncertainty in the data-to-simulation scale factors (SF) for b tagging efficiency and mistag rates. These scale factors are derived from data in various QCD and $t\bar{t}$ control samples and are binned in jet p_T and jet flavor (light + g, c, and b) [54]. The ± 1 s.d. N_b templates for these scale factors are assessed by varying them according to the uncertainties in their measurements. The effect of these variations on the N_b distribution in $t\bar{t}$ events is shown in Figure ??.

8.3   Additional systematic uncertainties

Other experimental uncertainties are small and include lepton selection efficiency, lepton misidentification rate, jet energy scale, jet energy resolution, and integrated luminosity. The uncertainty associated with lepton selection efficiency is determined by varying the efficiency to select a lepton within its uncertainty determined from data. The N_{leps} distribution for QCD events may not be simulated well because it relies on modeling the tail of the fragmentation function and various detector effects. To account for this, an uncertainty of 20% is assigned to the relative normalization of QCD events in the 0- and 1-lepton bins, which is motivated by data-to-simulation studies of lepton isolation distributions. Jet energy scale uncertainties [55, 56] are assessed by varying the p_T of small-R jets as a function of p_T and η. The uncertainty arising from jet energy resolution [55, 56] is determined by applying an $|\eta|$-dependent factor to the jet p_T to match the jet energy resolution observed in data. The integrated luminosity is varied according to its uncertainty of 2.5% [57], affecting only the backgrounds estimated from simulation. No uncertainty is applied for the amount of pileup as studies have shown its effect to be negligible in this high-H_T selection. The uncertainties due to the limited size
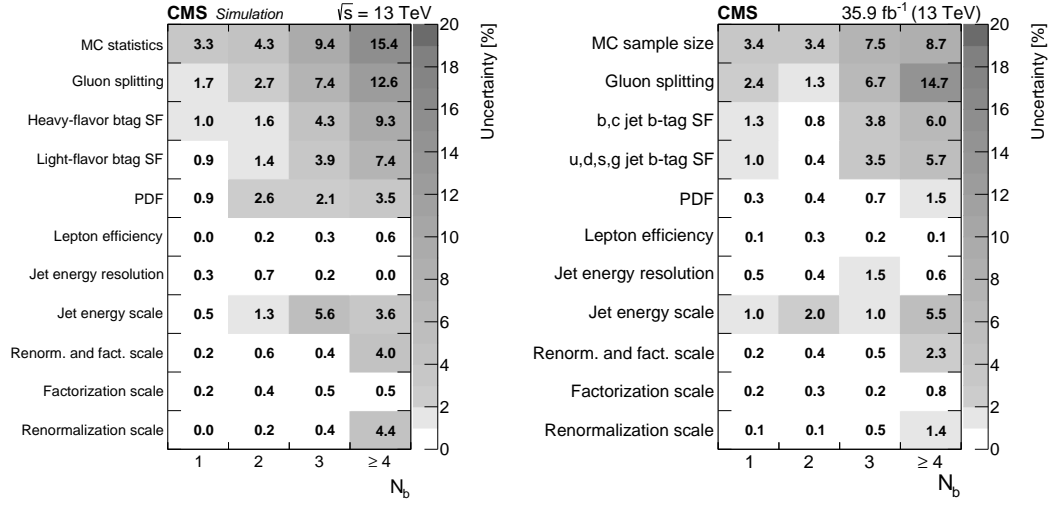
Figure 8.5: Background systematic uncertainties affecting the N_{b} shape (in percent) for the ($N_{\mathrm{jets}} \geq 8$, $500 < M_{\mathrm{J}} \leq 1000$ GeV) (left) and ($N_{\mathrm{jets}} \geq 8$, $M_{\mathrm{J}} > 1000$ GeV) (right) bins. The bottom row shows the total uncertainty for a given N_{b} bin by summing in quadrature all uncertainties. These values are similar for other (N_{jets}, M_{J}) bins.

of simulation samples are incorporated as uncorrelated nuisance parameters in the fit.

Theoretical systematic uncertainties are applied and include independent and correlated variations of the renormalization and factorization scales. Additionally, uncertainties on the PDF are incorporated by considering variations in the NNPDF 3.0 scheme [42]. The size of these uncertainties is typically small as the effect of these variations is largely to modify the cross section of processes, which for the main backgrounds are constrained by data.

The background systematic uncertainties that affect the N_{b} shape are shown in Figure 8.5 for the two most sensitive search bin.

$N_{PV}^{true} \leq 20$	$20 < N_{PV}^{true} \leq 40$	$N_{PV}^{true} > 40$
$8.0 \pm 0.5\%$	$8.1 \pm 0.4\%$	$7.5 \pm 1.5\%$

Table 8.2: The signal efficiency of the most sensitive bin ($N_{\text{jets}} \geq 8$, $M_{\text{J}} > 1000$ GeV) for a 1600 GeV gluino in various bins of the number of truth-level primary vertices.

8.4   Signal Systematics

Several of the systematic uncertainties affecting the signal yield are evaluated in the same way as the background yield. These are the uncertainties due to gluon splitting, lepton selection efficiency, jet energy scale, jet energy resolution, b tagging scale factors, simulation sample size, integrated luminosity, and theoretical uncertainties. All systematic variations affect both the N_{b} shape and normalization, except for the gluon splitting uncertainty, which is taken to affect only the N_{b} shape.

The number of jets from ISR produced in the signal simulation is reweighted based on comparisons between data and simulated $t\bar{t}$ samples. The reweighting factors vary between 0.92 and 0.51 for the number of ISR jets between 1 and ≥ 6. One half of the deviation from unity is taken as the systematic uncertainty in these reweighting factors.

The systematic uncertainties affecting the signal N_{b} shape are shown in Fig. 8.6 (right) for the most sensitive bin in a model with $m_{\tilde{g}} = 1600$ GeV. The dominant signal systematic uncertainties arise from the limited simulation sample size, the b tagging efficiency scale factors, and the ISR modeling. There is no systematic uncertainty taken for pileup reweighting, as the signal efficiency is found to be insensitive to the number of pileup interactions, which is shown in Table 8.2.
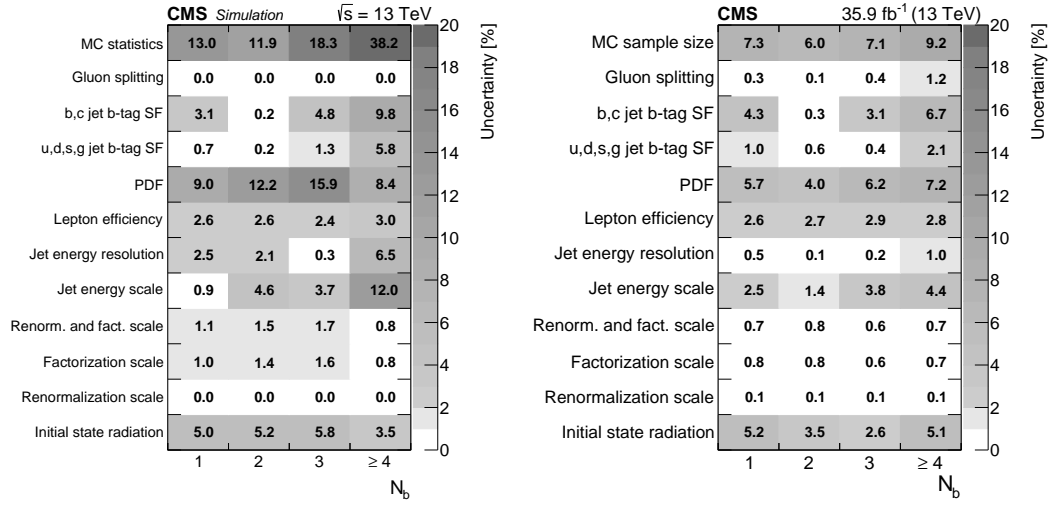
Figure 8.6: Signal systematic uncertainties affecting the N_{b} shape (in percent) for the $(N_{\mathrm{jets}} \geq 8, \, 500 < M_{\mathrm{J}} \leq 1000 \text{ GeV})$ (left) and $(N_{\mathrm{jets}} \geq 8, \, M_{\mathrm{J}} > 1000 \text{ GeV})$ (right) bins. The bottom row shows the total uncertainty for a given N_{b} bin by summing in quadrature all uncertainties. These values are similar for other $(N_{\mathrm{jets}}, \, M_{\mathrm{J}})$ bins.

Chapter 9

Fit Validation

Due to the fit's complexity and many adjustable parameters, it is important to verify the fit model and that behaves as intended.

9.1   Signal Injection Study

Signal injetion studies are a useful way to quantify how well the maximum-likelihood fit can extract a potential signal if it is indeed present. These studies rely on the use of psuedodata experiments. A single experiment consists of generating psuedodata by fluctuating bin yields around their pre-fit values according to their statistical and systematic uncertainties. This pseudodata can then be treated as observations and can be fit with the results examined. As many psuedodata experiments are generated, the collection of observations approximates the distribution of possible observations as defined by the fit model and correspondingly the distribution of post-fit results approximates the distribution of possible post-fit results.

For the signal injection study, 1,000 experiments are generated by fluctuating both the expected background and signal yields (with signal strength = 1) for each gluino mass point. Figure 9.2 shows the median fitted signal strength of the 1,000 experiments for each gluino mass point. For gluino masses between $1000-1700$ GeV the fit shows no evidence of

$m_{\tilde{g}} = 1800$ GeV	$m_{\tilde{g}} = 1900$ GeV	$m_{\tilde{g}} = 2000$ GeV
96%	95%	96%

Table 9.1: Actual coverage probability of the 95% confidence interval of the fit for the mass points with a biased signal extraction.

a bias and has a median extracted signal strength of $\sim$ 1, while for higher gluino masses, the fit tends to under-extract the signal contribution (up to $\sim$ 25% for $m_{\tilde{g}} = 2000$ GeV). These biased mass points correspond to models where the number of signal events is very low. For example, there are only 8.6 events expected for the $m_{\tilde{g}} = 2000$ GeV model, summing oer all analysis bins. This low yield means that gaussian-approximations of the poisson-distributed bin yields used in the fit model are inaccurate, leading to the bias in the fit.

In order to test this hypothesis, additional signal injection studies, each consisting of 1000 experiments, are performed for the $m_{\tilde{g}} = 2000$ GeV mass point, where the injected signal strengths are 1x, 3x, 5x, 10x the nominal cross-section. The resulting median extracted signal is 78%, 92%, 95%, and 98% the injected signal, respectively. These results support the hypothesis, as, with increasing signal strength, the gaussian-approximations become increasingly accurate, allowing for the fit to properly extract the signal contributions. The distributions of fitted signal strengths for these tests are shown in Figure **??**.

No modifications to the fit model are made to correct for this issue. This is because the fit bias only affects mass points that are far above the highest mass (1650 GeV) expected to be excluded by this analysis, and Figure 9.2 shows that the fit bias is much smaller than the precision of the fit for those mass points. Additionally, the coverage of the 95% confidence intervals of the fit is tested using the signal injection experiments and found to be be either correct or slightly conservative, as shown in Table 9.1.
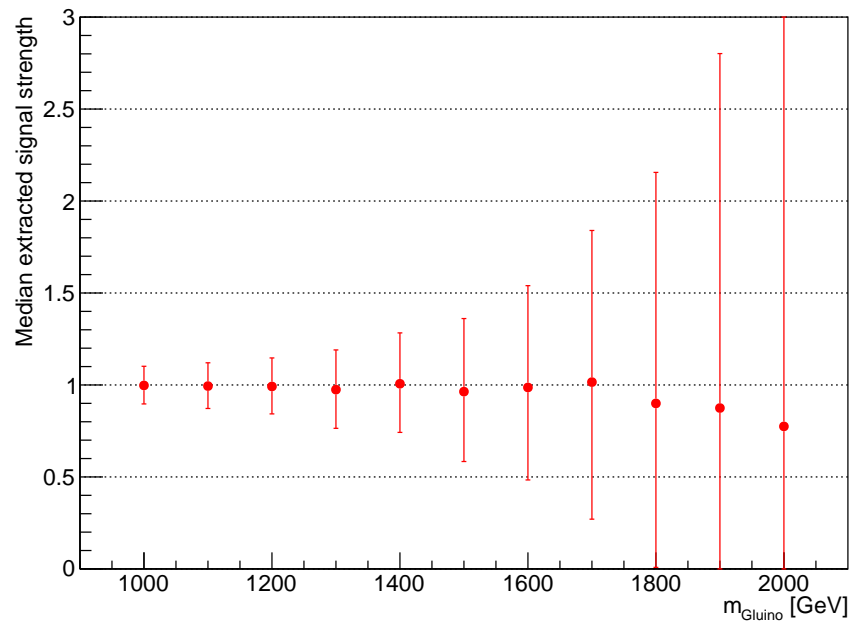
Figure 9.1: Median extracted signal strength of 1,000 psuedodata experiments as a function of gluino mass. The uncertainties drawn are the median upper and lower errors of the fitted signal strengths per mass point.
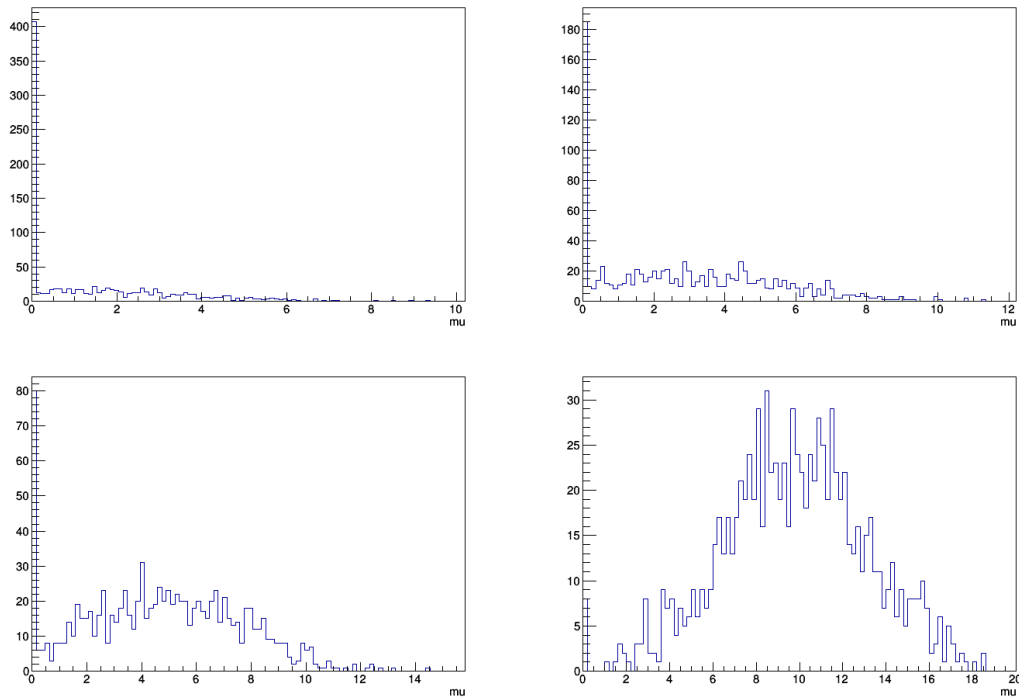
Figure 9.2: Distribution of fitted signal strengths of 1000 psuedodata experiments for a 2000 GeV gluino at 1x (top-left), 3x (top-right), 5x (bottom-left), and 10x (bottom-right) the nominal cross-section. The amount of signal extracted is 78%, 92%, 95%, and 98% the injected signal, respectively.

9.2   Control Region Fit

While the signal injection studies are a useful validation of the fit model, it is important to validate the model using data in order to test for unmodelled effects. To do this, the maximum-likelihood fit is performed with only the low-N_{jets}, low-M_{J} control regions, as defined in Table 6.1. These bins are chosen due to their low-expected signal yields, which avoids signal contamination effects and unblinding the high-expected signal regions in the case further modification of the fit are needed.

The fit, under the background-only hypothesis, is able to model the observed data well, as seen in the post-fit N_{b} distributions shown in Figure 9.3, without needing large adjustments to the nuisance parameters. The change between the pre- and post-fit normalizations of the background processes is shown in Table 9.2, while the pulls of the nuisance parameters corresponding to systematic uncertainties (largely controlling the shape of the N_{b} distribution) are shown in Figure 9.4. Both sets of values appear well-behaved, as the largest change in normalization is less than 50%, with typical values around 10-15%, while the nuisance parameters are all consistent with their pre-fit uncertainties, with most shifted less than 0.05 s.d. The largest pulls correspond to nuisance parameters controlling the gluon splitting rate (gs, +0.42 s.d.), the light-flavor b-tag SFs (btag_udsg, +0.37 s.d.), and the heavy-flavor b-tag SFs (btag_bc, +0.13 s.d.). These nuisances are expected to be shifted up as the observed data is higher than simulation in the tail of the pre-fit N_{b} distributions, as seen in Figure **??**.

Lastly, Table 9.3 compares the post-fit pulls of the background-only and signal-plus-background control region fit. The post-fit pulls between the two fits are fully consistent with each other, as is expected for these signal-poor regions.
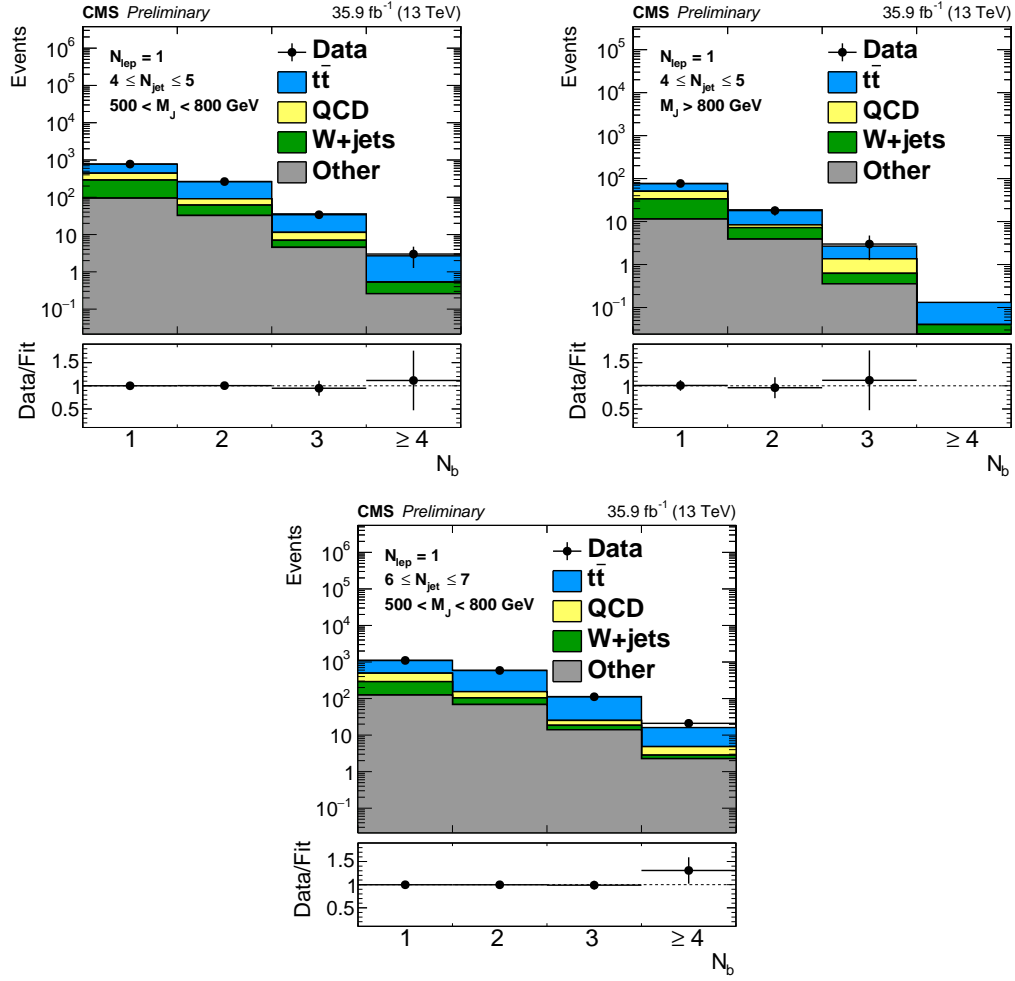
Figure 9.3: Post-fit N_b distributions of the control region fit.
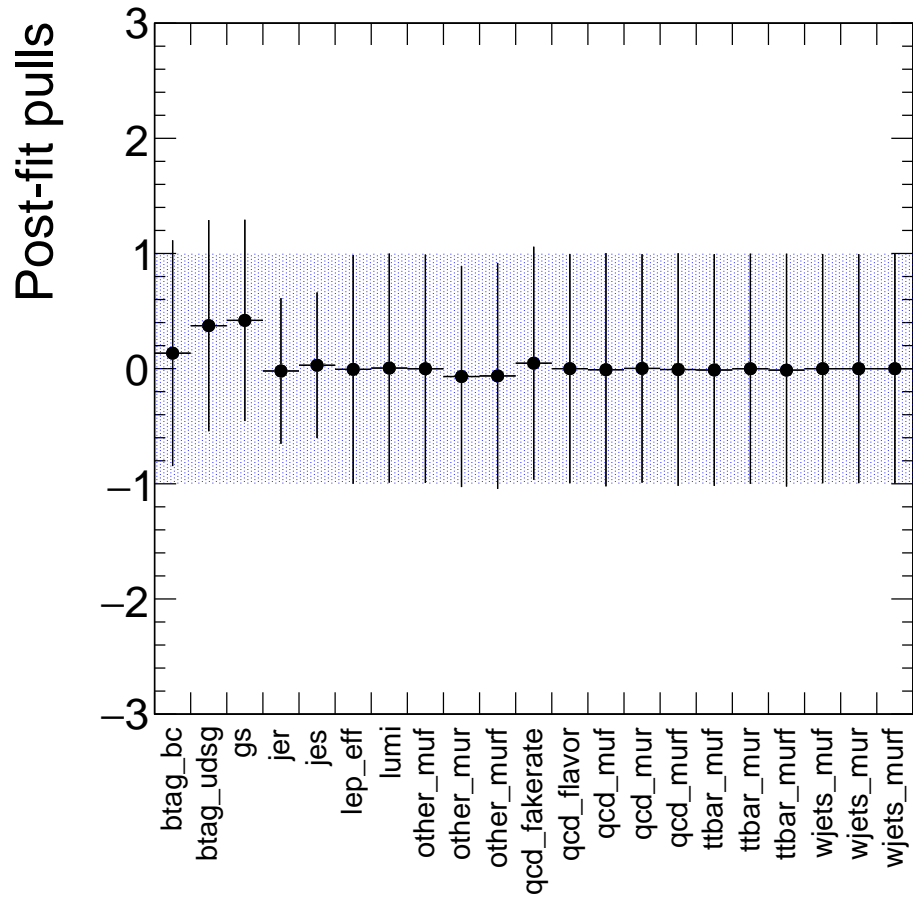
Figure 9.4: Post-fit pulls of the background-only control region fit. The post-fit value of the nuisance parameter is indicated by the data point, while the post-fit uncertainty is shown as a black line and is normalized by the pre-fit uncertainty depicted as the blue band.

Process	Pre-fit Yield	Post-fit Yield (b-only fit)	% change
$4 \leq N_{\mathrm{jets}} \leq 5,\ 500 \leq M_{\mathrm{J}} \leq 800$			
$t\bar{t}$	501.4	533.3 ± 80.7	+6.3
QCD	218.8	186.7 ± 36.8	-14.7
W + jets	400.4	225.5 ± 100.0	-43.7
Other	141.4	131.8 ± 34.5	-6.8
$4 \leq N_{\mathrm{jets}} \leq 5,\ M_{\mathrm{J}} \geq 800$			
$t\bar{t}$	36.9	37.5 ± 13.3	+1.6
QCD	23.1	18.9 ± 4.5	-18.2
W + jets	45.7	25.7 ± 11.4	-43.8
Other	16.8	15.7 ± 3.8	-6.5
$6 \leq N_{\mathrm{jets}} \leq 7,\ 500 \leq M_{\mathrm{J}} \leq 800$			
$t\bar{t}$	1370.4	1148.3 ± 78.0	-16.2
QCD	293.9	262.6 ± 52.1	-10.6
W + jets	367.7	205.6 ± 92.2	-44.1
Other	225.2	209.7 ± 58.5	-6.9

Table 9.2: Table comparing the post-fit normalizations of the control region fit to the pre-fit yields for the various background processes.

Nuisance parameter	Post-fit pull (b-only fit)	Post-fit pull (s + b fit)	$\rho(\theta_i, \mu)$
b,c jet b-tag SF (btag_bc)	$+0.13 \pm 0.98$	$+0.07 \pm 1.05$	-0.18
u,d,s,g jet b-tag SF (btag_udsg)	$+0.37 \pm 0.92$	$+0.28 \pm 0.95$	-0.26
Gluon splitting (gs)	$+0.42 \pm 0.87$	$+0.22 \pm 1.12$	-0.43
Jet energy resolution (jer)	-0.02 ± 0.63	-0.02 ± 0.60	-0.01
Jet energy scale (jes)	$+0.03 \pm 0.63$	$+0.03 \pm 0.61$	-0.03
Lepton efficiency (lep_eff)	-0.01 ± 0.99	-0.01 ± 0.99	$+0.01$
Luminosity (lumi)	$+0.00 \pm 0.99$	$+0.00 \pm 0.99$	-0.01
Fact. scale for other (other_muf)	-0.00 ± 0.99	-0.00 ± 0.99	$+0.00$
Renorm. scale for other (other_mur)	-0.07 ± 0.96	-0.06 ± 1.02	$+0.02$
Renorm. and Fact. scale for other (other_murf)	-0.06 ± 0.98	-0.08 ± 0.96	$+0.01$
QCD fake rate (qcd_fakerate)	$+0.05 \pm 1.01$	$+0.09 \pm 1.14$	$+0.09$
Fact. scale for QCD (qcd_muf)	-0.01 ± 1.01	-0.01 ± 1.01	-0.00
Renorm. scale for QCD (qcd_mur)	$+0.00 \pm 0.99$	$+0.00 \pm 0.99$	-0.00
Renorm. and Fact. scale for QCD (qcd_murf)	-0.01 ± 1.01	-0.01 ± 1.01	-0.00
Fact. scale for $t\bar{t}$ (ttbar_muf)	-0.01 ± 1.01	-0.01 ± 1.00	$+0.00$
Renorm. scale for $t\bar{t}$ (ttbar_mur)	-0.00 ± 1.00	$+0.00 \pm 0.99$	$+0.01$
Renorm. and Fact. scale for $t\bar{t}$ (ttbar_murf)	-0.01 ± 1.01	-0.01 ± 1.00	$+0.01$
Fact. scale for W + jets (wjets_muf)	-0.00 ± 0.99	$+0.00 \pm 0.99$	$+0.00$
Renorm. scale for W + jets (wjets_mur)	-0.00 ± 0.99	-0.00 ± 0.99	-0.00
Renorm. and Fact. scale for W + jets (wjets_murf)	-0.00 ± 1.00	-0.00 ± 1.00	$+0.00$

Table 9.3: Table of post-fit pulls of the background-only and signal-plus-background control region fit. The last column, $\rho(\theta_i, \mu)$, lists the correlation between the corresponding nuisance parameter, θ_i, and the nuisance parameter controlling the signal strength, μ.

71

Chapter 10

Results and Interpretation

10.1   Results

The results of a background-only fit of the observed N_{b} distributions are shown in Figures 10.1 and 10.2. These figures separately show the $N_{\mathrm{leps}} = 1$ control and signal regions, although the fit includes all bins simultaneously. The N_{b} distributions in data are well described by the fit, and examination of the nuisance parameters shows that none of them are significantly changed by the fit. The post-fit yields are presented in Table 10.1.

A signal-plus-background fit is performed for gluino masses ranging from 1000 to 2000 GeV. For all masses, the post-fit N_{b} distribution describes the data well, and the fit extracts at most a small and insignificant signal contribution. For example, with a 1600 GeV gluino, the extracted signal yield relative to the model prediction is $r = 0.18^{+0.41}_{-0.18}$. The change of nuisance parameters by the fit is small and consistent with those of the background-only fit.
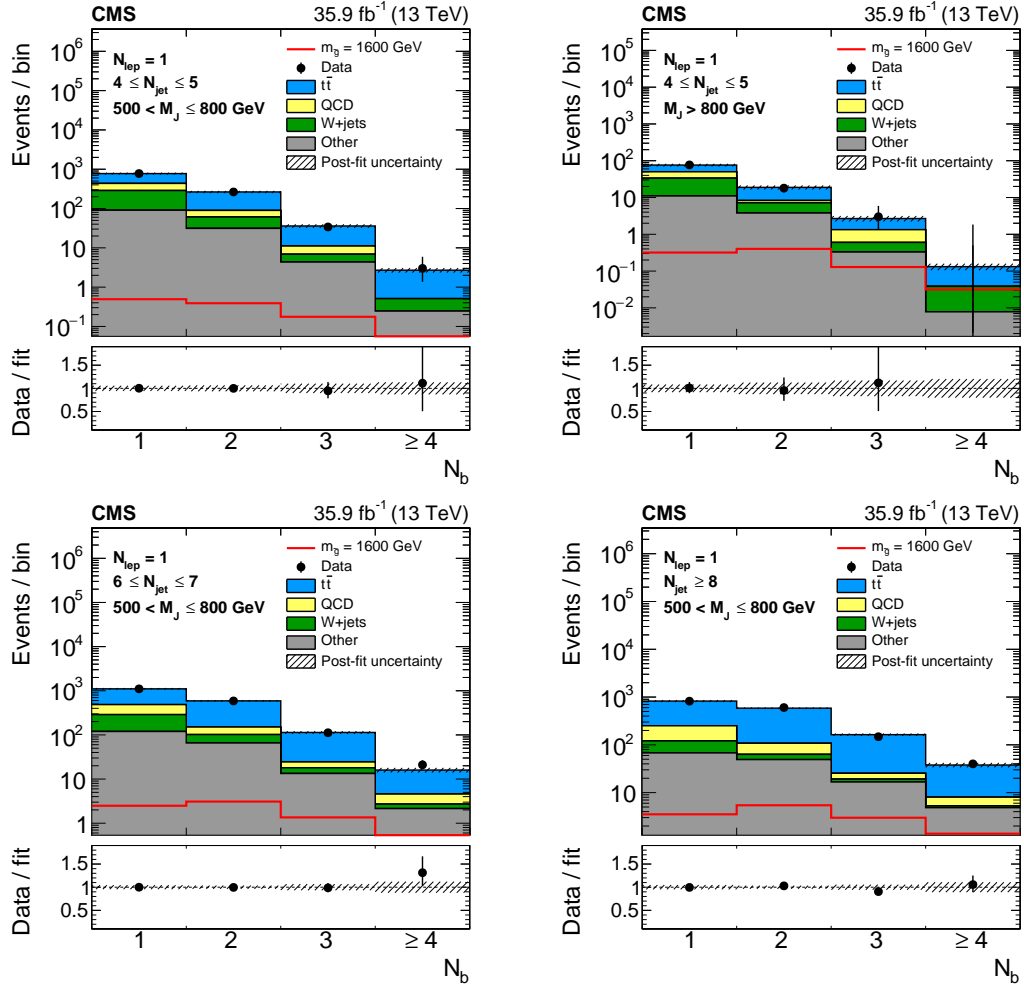
Figure 10.1: Data and the background-only post-fit N_b distribution for bins with low expected signal contribution: $4 \leq N_{\text{jets}} \leq 5$, $500 < M_J \leq 800$ GeV (upper-left), $4 \leq N_{\text{jets}} \leq 5$, $M_J > 800$ GeV (upper-right), $6 \leq N_{\text{jets}} \leq 7$, $500 < M_J \leq 800$ GeV (lower-left), and $N_{\text{jets}} \geq 8$, $500 < M_J \leq 800$ GeV (lower-right). The expected signal distribution is also shown for a gluino mass of 1600 GeV. The ratio of data to post-fit yields is shown in the lower panel. The post-fit uncertainty is depicted as a hatched band.
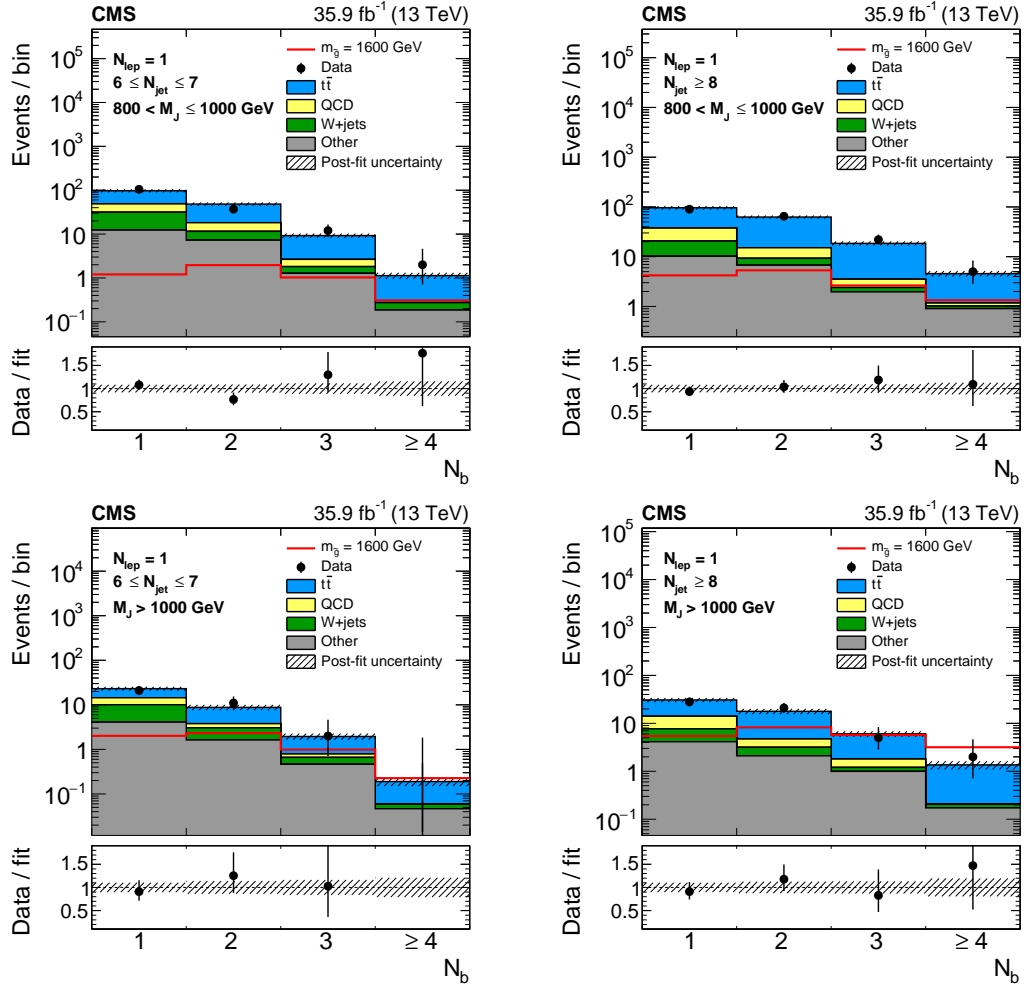
Figure 10.2: Data and the background-only post-fit N_b distribution for bins with large expected signal contribution: $6 \leq N_\text{jets} \leq 7$, $800 < M_\text{J} \leq 1000$ GeV (upper-left), $N_\text{jets} \geq 8$, $800 < M_\text{J} \leq 1000$ GeV (upper-right), $6 \leq N_\text{jets} \leq 7$, $M_\text{J} > 1000$ GeV (lower-left), and $N_\text{jets} \geq 8$, $M_\text{J} > 1000$ GeV (lower-right). The expected signal distribution is also shown for a gluino mass of 1600 GeV. The ratio of data to post-fit yields is shown in the lower panel. The post-fit uncertainty is depicted as a hatched band.

| N_b	QCD	$t\bar{t}$	W + jets	Other	All bkg.	Data	Expected $m_{\tilde{g}} = 1600$ GeV
			$4 \leq N_\mathrm{jets} \leq 5$, $500 < M_\mathrm{J} \leq 800$ GeV				
1	148	340	196	91	775 ± 43	777	0.50 ± 0.13
2	29	175	30	31	264 ± 17	264	0.39 ± 0.11
3	4.3	24.8	2.5	4.4	36 ± 4	34	0.18 ± 0.08
≥ 4	0.0	2.2	0.3	0.2	2.7 ± 0.4	3	0.04 ± 0.04
			$4 \leq N_\mathrm{jets} \leq 5$, $M_\mathrm{J} > 800$ GeV				
1	16.5	26.3	22.5	11.0	76 ± 6	77	0.32 ± 0.11
2	1.1	10.6	3.4	3.8	19 ± 2	18	0.40 ± 0.12
3	0.7	1.3	0.3	0.3	2.7 ± 0.5	3	0.13 ± 0.06
≥ 4	0.00	0.09	0.03	0.01	0.13 ± 0.03	0	0.03 ± 0.03
			$6 \leq N_\mathrm{jets} \leq 7$, $500 < M_\mathrm{J} \leq 800$ GeV				
1	197	620	169	120	1106 ± 48	1105	2.5 ± 0.3
2	49	440	36	66	591 ± 21	588	3.1 ± 0.3
3	6.4	89.2	4.6	13.4	114 ± 8	112	1.4 ± 0.2
≥ 4	1.9	11.4	0.6	2.1	16 ± 2	21	0.25 ± 0.09
			$N_\mathrm{jets} \geq 8$, $500 < M_\mathrm{J} \leq 800$ GeV				
1	130	574	53	68	825 ± 38	821	3.5 ± 0.3
2	45	478	14	49	586 ± 20	603	5.4 ± 0.4
3	6.3	138.1	2.5	16.7	164 ± 9	148	3.0 ± 0.3
≥ 4	2.8	29.8	0.4	4.8	38 ± 4	40	1.4 ± 0.2
			$6 \leq N_\mathrm{jets} \leq 7$, $800 < M_\mathrm{J} \leq 1000$ GeV				
1	17.3	48.4	19.2	12.3	97 ± 8	105	1.2 ± 0.2
2	6.6	30.1	4.3	7.3	48 ± 4	37	2.0 ± 0.3
3	0.8	6.6	0.5	1.3	9.3 ± 1.0	12	1.0 ± 0.2
≥ 4	0.0	0.9	0.1	0.2	1.1 ± 0.2	2	0.31 ± 0.09
			$N_\mathrm{jets} \geq 8$, $800 < M_\mathrm{J} \leq 1000$ GeV				
1	17.0	58.7	10.3	10.2	96 ± 8	90	4.2 ± 0.4
2	5.8	47.5	2.5	6.8	63 ± 5	65	5.3 ± 0.4
3	1.1	15.0	0.4	2.0	19 ± 2	22	2.6 ± 0.3
≥ 4	0.2	3.4	0.1	0.9	4.6 ± 0.6	5	1.3 ± 0.2
			$6 \leq N_\mathrm{jets} \leq 7$, $M_\mathrm{J} > 1000$ GeV				
1	4.4	8.7	6.0	4.1	23 ± 2	21	2.0 ± 0.3
2	0.7	5.0	1.4	1.6	8.8 ± 1.2	11	2.3 ± 0.3
3	0.1	1.2	0.2	0.5	1.9 ± 0.3	2	1.0 ± 0.2
≥ 4	0.00	0.13	0.01	0.05	0.19 ± 0.04	0	0.23 ± 0.08
			$N_\mathrm{jets} \geq 8$, $M_\mathrm{J} > 1000$ GeV				
1	6.4	16.7	3.5	4.1	31 ± 3	28	5.4 ± 0.4
2	1.6	13.1	1.1	2.1	18 ± 2	21	8.2 ± 0.5
3	0.6	4.2	0.2	1.0	6.0 ± 0.8	5	5.7 ± 0.4
≥ 4	0.0	1.2	0.0	0.2	1.4 ± 0.3	2	3.2 ± 0.3

Table 10.1: Post-fit yields of the background-only fit, observed data, and expected yields for $m_{\tilde{g}} = 1600$ GeV.
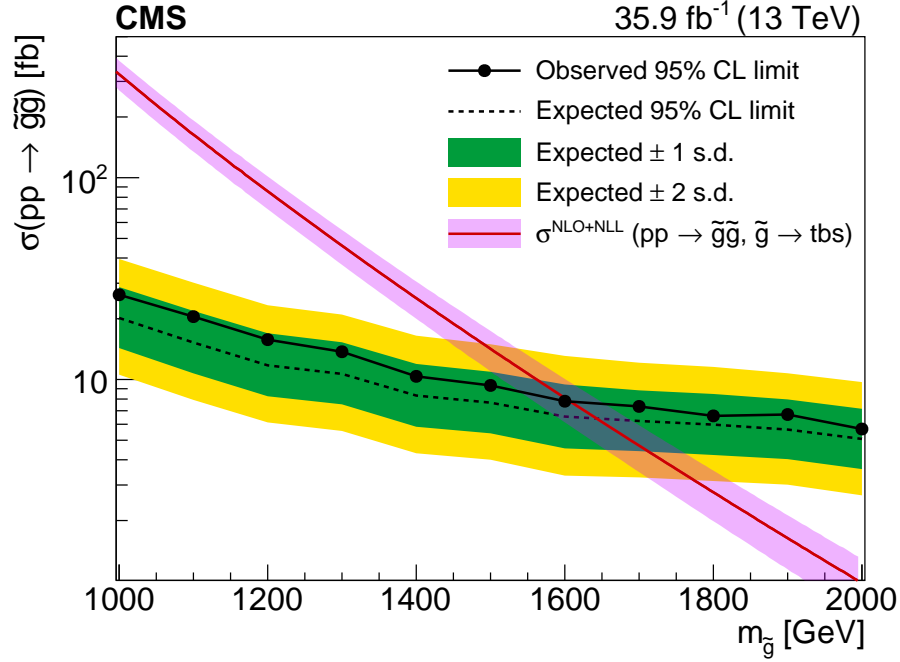
Figure 10.3: Cross section upper limits at 95% CL for a model of gluino pair production with $\tilde{g} \to tbs$ compared to the gluino pair production cross section. The theoretical uncertainties in the cross section are shown as a band around the red line [12]. The expected limits (dashed line) and their ±1 s.d. and ±2 s.d. variations are shown as green and yellow bands, respectively. The observed limit is shown by the solid line with dots.

10.2   Interpretation and Limits

Limits on the signal production cross section are calculated at 95% confidence level (CL) using the asymptotic approximation of the $\mathrm{CL_s}$ criterion [58, 59, 60, 61] and shown in Figure 10.3. Comparing the observed limit to the gluino pair production cross section [12], gluino masses below 1610 GeV are excluded in the benchmark $\tilde{g} \to tbs$ model.

Chapter 11

Conclusions

11.1 Section Title

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Part V

Appendix

Appendix A

Mitagating the HIP Effect

A.1 Section Title

Appendicitis

Appendix B

QCD Flavor Fit

B.1 Section Title

Appendicitis

Bibliography

[1] F. Marcastel, *CERN's Accelerator Complex. La chaîne des accélérateurs du CERN*, *General Photo* (Oct, 2013).

[2] W. J. Stirling, 13/8 TeV *LHC luminosity ratios*, *Private Communication* (2013).

[3] L. R. F. Castillo, *The ATLAS and CMS detectors*, in *The Search and Discovery of the Higgs Boson*, 2053-2571, pp. 4–1 to 4–8. Morgan & Claypool Publishers, 2015.

[4] M. Schott and M. Dunford, *Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV. , Eur. Phys. J. C* **74** (May, 2014) 60 p.

[5] T. Lenzi, *Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors*, Master's thesis, U. Brussels, 2013.

[6] B. Isildak, *Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s} = 7$ TeV*. PhD thesis, Bogazici U., 2011. arXiv:1308.6064.

[7] CMS Collaboration, *Identification and filtering of uncharacteristic noise in the cms hadron calorimeter*, *Journal of Instrumentation* **5** (2010), no. 03 T03014.

[8] P. Paolucci *et. al.*, *CMS Resistive Plate Chamber overview, from the present system to the upgrade phase I*, *Journal of Instrumentation* **8** (2013), no. 04 P04005.

[9] CMS Collaboration, *The cms experiment at the cern lhc*, *Journal of Instrumentation* **3** (2008), no. 08 S08004.

[10] **CMS** Collaboration, A. M. Sirunyan *et. al.*, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *Submitted to: JINST* (2017) [arXiv:1712.0715].

[11] CMS Collaboration, "Public CMS Luminosity Information." https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016_proton_proton_13_TeV_collis, April, 2018. Accessed: April 23, 2018.

[12] C. Borschensky, M. Krämer, A. Kulesza, M. Mangano, S. Padhi, T. Plehn, and X. Portell, *Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33, and 100$TeV*, *Eur. Phys. J. C* **74** (2014) 3174, [arXiv:1407.5066].

[13] **CMS** Collaboration, S. Chatrchyan *et. al.*, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, *JINST* **9** (2014), no. 10 P10009, [arXiv:1405.6569].

[14] K. Rose, *Deterministic annealing for clustering, compression, classification, regression, and related optimization problems*, *Proceedings of the IEEE* **86** (Nov, 1998) 2210–2239.

[15] Früwirth, R and Waltenberger, Wolfgang and Vanlaer, Pascal, *Adaptive Vertex Fitting*, Tech. Rep. CMS-NOTE-2007-008, CERN, Geneva, Mar, 2007.

[16] CMS Collaboration, *Particle-flow reconstruction and global event description with the cms detector*, *Journal of Instrumentation* **12** (2017), no. 10 P10003.

[17] **CMS Collaboration** Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*, Tech. Rep. CMS-PAS-PFT-09-001, CERN, Geneva, Apr, 2009.

[18] **CMS Collaboration** Collaboration, *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*, Tech. Rep. CMS-PAS-PFT-10-001, 2010.

[19] CMS Collaboration, "Cut-based Electron Identification in Run 2." `https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2Archive#Spring15_selection_25ns`, May, 2017. Accessed: May 01, 2018.

[20] K. Rehermann and B. Tweedie, *Efficient identification of boosted semileptonic top quarks at the LHC*, *JHEP* **03** (2011) 059, [arXiv:1007.2221].

[21] **CMS** Collaboration, V. Khachatryan *et. al.*, *Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *JINST* **10** (2015) P06005, [arXiv:1502.0270].

[22] CMS Collaboration, "Baseline muon selections for Run-II." `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2#Medium_Muon`, April, 2018. Accessed: May 01, 2018.

[23] **CMS** Collaboration, S. Chatrchyan *et. al.*, *Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$TeV*, *JINST* **7** (2012) P10002, [arXiv:1206.4071].

[24] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions, Phys. Rev.* **D48** (1993) 3160–3166, [hep-ph/9305266].

[25] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better jet clustering algorithms, JHEP* **08** (1997) 001, [hep-ph/9707323].

[26] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm, JHEP* **04** (2008) 063, [arXiv:0802.1189].

[27] G. P. Salam, *Towards Jetography, Eur. Phys. J.* **C67** (2010) 637–686, [arXiv:0906.1833].

[28] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual, Eur. Phys. J.* **C72** (2012) 1896, [arXiv:1111.6097].

[29] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas, Phys. Lett.* **B659** (2008) 119–126, [arXiv:0707.1378].

[30] CMS Collaboration, "Jet Identification for the 13 TeV data Run2016." https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016, February, 2018. Accessed: May 01, 2018.

[31] T. C. collaboration, *Determination of jet energy calibration and transverse momentum resolution in cms, Journal of Instrumentation* **6** (2011), no. 11 P11002.

[32] **CMS** Collaboration, S. Chatrchyan *et. al.*, *Identification of b-quark jets with the CMS experiment, JINST* **8** (2013) P04013, [arXiv:1211.4462].

[33] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder, Phys. Lett.* **B641** (2006) 57–61, [hep-ph/0512210].

[34] A. Collaboration, *Search for top-squark pair production in final states with one lepton, jets, and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector*, arXiv:1711.1152.

[35] C. Collaboration, *Search for supersymmetry in pp collisions at $\sqrt{s} = 13$ TeV in the single-lepton final state using the sum of masses of large-radius jets, JHEP* **08** (2016) 122, [arXiv:1605.0460].

[36] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, JHEP* **07** (2014) 079, [arXiv:1405.0301].

[37] J. Alwall *et. al.*, *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions, Eur. Phys. J. C* **53** (2008) 473, [arXiv:0706.2569].

[38] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [hep-ph/0409146].

[39] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [arXiv:0709.2092].

[40] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, [arXiv:1002.2581].

[41] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, *JHEP* **12** (2012) 061, [arXiv:1209.6215].

[42] **NNPDF** Collaboration, R. D. Ball *et. al.*, *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [arXiv:1410.8849].

[43] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159.

[44] P. Skands, S. Carrazza, and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 tune*, *Eur. Phys. J. C* **74** (2014) 3024.

[45] M. Czakon, P. Fiedler, and A. Mitov, *Total top-quark pair-production cross section at hadron colliders through* $\mathcal{O}(\alpha_S^4)$, *Phys. Rev. Lett.* **110** (Jun, 2013) 252004.

[46] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, *W Physics at the LHC with FEWZ 2.1*, *Comput. Phys. Commun.* **184** (2013) 208, [arXiv:1201.5896].

[47] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) 111, [arXiv:0907.4076]. [Erratum: DOI10.1007/JHEP02(2010)011].

[48] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J. C* **71** (2011) 1547, [arXiv:1009.2450].

[49] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, *Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons*, *JHEP* **06** (2015) 184, [arXiv:1504.0344].

[50] G. Bevilacqua and M. Worek, *Constraining BSM Physics at the LHC: Four top final states with NLO accuracy in perturbative QCD*, *JHEP* **07** (2012) 111, [arXiv:1206.3064].

[51] Z. Nagy, *Three jet cross-sections in hadron hadron collisions at next-to-leading order*, *Phys. Rev. Lett.* **88** (2002) 122003, [hep-ph/0110315].

[52] **GEANT4** Collaboration, S. Agostinelli *et. al.*, *GEANT4—a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.

[53] **CMS** Collaboration, A. M. Sirunyan *et. al.*, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13TeV." Submitted to JINST, 2017.

[54] **CMS** Collaboration, C. Collaboration, *Performance of heavy flavour identification algorithms in proton-proton collisions at 13TeV at the CMS experiment*, CMS Detector Performance Summary CMS-DP-2017-012, 2017.

[55] **CMS** Collaboration, S. Chatrchyan *et. al.*, *Determination of jet energy calibration and transverse momentum resolution in CMS*, *JINST* **6** (2011) P11002, [arXiv:1107.4277].

[56] **CMS** Collaboration, V. Khachatryan *et. al.*, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8TeV*, *JINST* **12** (2017), no. 02 P02014, [arXiv:1607.0366].

[57] **CMS** Collaboration, GHM *et. al.*, *CMS luminosity measurements for the 2016 data taking period*, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[58] A. L. Read, *Presentation of search results: the* $\mathrm{CL_s}$ *technique*, *J. Phys. G* **28** (2002) 2693.

[59] ATLAS and CMS Collaborations, LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*, Technical Report CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, 2011.

[60] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554. [Erratum: DOI10.1140/epjc/s10052-013-2501-z].

[61] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth. A* **434** (1999) 435, [hep-ex/9902006].