Extension of the Model Unspecific Search in CMS to Final States with Jets using 2012 Data

von
Andreas A. E. Albert

Masterarbeit in Physik

vorgelegt der

Fakultät für Mathematik, Informatik und Naturwissenschaften
der Rheinisch-Westfälischen Technischen Hochschule Aachen

im
November 2015

angefertigt am
III. Physikalischen Institut (IIIA)

bei
Prof. Dr. Th. Hebbeker

Zweitgutachter
Prof. Dr. Ch. Wiebusch

Abstract

The Model Unspecific Search in CMS (MUSiC) is presented. In MUSiC, proton-proton collision events are sorted into event classes based on final state topology. For each event class, distributions of kinematic variables in data and Monte Carlo simulation of Standard Model physics are automatically compared. The search is explicitly not optimised for any specific model of New Physics to avoid missing phenomena in unexpected topologies. The analysis strategy is applied to the full data set collected at $\sqrt{s} = 8\,\text{TeV}$ by CMS in 2012 corresponding to an integrated luminosity of $19.5\,\text{fb}^{-1}$. Special emphasis is placed on jet triggered events, which have not previously been considered in the MUSiC analysis. The results of the analysis are presented and discrepancies between data and simulation are studied. A simulation-based sensitivity study is performed to demonstrate the analysis performance for a potential Dark Matter phenomenom.

Zusammenfassung

Die Modell-unabhängige Suche in CMS (MUSiC) wird vorgestellt. In MUSiC werden Proton-Proton Kollisionsereignisse nach ihrem Endzustand klassifiziert. In jeder Ereignisklasse werden die Verteilungen bestimmter kinematischer Variablen in Daten und Monte Carlo Simulation des Standardmodells automatisch verglichen. Um zu vermeiden, dass unerwartete Phänomene übersehen werden, wird die Suchmethode explizit nicht auf ein bestimmtes Modell von Physik jenseits des Standardmodells optimiert. Die Analysestrategie wird auf den kompletten Datensatz von CMS aus dem Jahr 2012 bei $\sqrt{s} = 8\,\text{TeV}$ angewandt, der einer integrierten Luminosität von $19.5\,\text{fb}^{-1}$ enspricht. Erstmals werden auch Ereignisse ausgewertet, die aufgrund der Präsenz hadronischer Jets aufgezeichnet wurden. In der Diskussion der Ergebnisse wird ein besonderer Schwerpunkt auf Ereignisklassen mit solchen Ereignissen gelegt und Abweichungen zwischen Simulation und Daten werden untersucht. Abschließend wird eine simulationsbasierte Sensitivitätsstudie durchgeführt, die demonstriert, wie ein mögliches Signal Dunkler Materie von MUSiC gefunden werden könnte.

Contents

List of Figures

1 Introduction

The first years of running at the Large Hadron Collider (LHC) are often told as the story of finding the Higgs boson. Its discovery is a monument not only to the technical ingenuity that made it possible experimentally, but also to the powerful influence of theoretical physics.

Decades before the discovery, the Higgs mechanism had been proposed to solve a glaring deficiency of the theoretical understanding, which at the time only described massless elementary particles. The Higgs mechanism became such a promising solution that vast experimental efforts were undertaken to find its signature, the Higgs boson. From conception to execution, the experimental search was supported by the work of theorists. The experiments were designed with the Higgs boson's decay signatures in mind. Exploiting internal consistency of the electroweak theory allowed experimentalists to constrain the possible range of the Higgs boson mass with ever increasing precision. In this case, reliance on the guidance of theorists was well-advised.

Of course, there has been and still is a plethora of open questions beside the Higgs mechanism. For many of these questions, theoretical solutions have been proposed, which again allow us to work out experimental signatures to look for. However, we should not forget that this approach has a fundamental blind spot: If our searches for New Physics are based on theorists' ideas, we will not find what no theorist is thinking of. To ensure that no clues are missed, a complementary analysis strategy is needed that is independent of theoretical guidance. Such a model independent search is implemented in the *Model Unspecific Search in CMS* or MUSiC.

This document is structured in five chapters. The first chapter introduces the Standard Model, the CMS experiment and the search strategy adopted in MUSiC. In the second chapter, the details of the analysis, such as event selection and object definition, are presented. Dedicated studies performed in preparation of the analysis are discussed in the third chapter. The fourth chapter gives the results of the analysis of the full CMS data set recorded in 2012. Finally, the results and conclusions of the analysis are summarised in the fifth chaper.

1.1 Theory

In this section, a brief overview of relevant theoretical concepts is given. The text is based on information from the excellent introductory accounts of [1–3] and the very dense reviews in [4].

1.1.1 The Standard Model

The modern understanding of elementary particle physics is summarised in the *Standard Model* (SM). The SM describes the properties of all known elementary particles and their interactions via the electromagnetic, weak and strong forces. It is formulated using the theoretical set of tools of quantum field theory (QFT). As the name suggests, QFT models quantum mechanical phenomena using fields. Analogously to the quantization of the quantum mechanical oscillator, field excitations are quantized and interpreted as particles. The mathematical implementation of QFT makes intensive use of variational calculus, which allows the complete information about the physical properties of a system to be encoded in a Lagrangian density[1], from which observables are constructed mathematically.

Formally, the SM is a $U(1)\otimes SU(2)\otimes SU(3)$ gauge theory composed of the Glashow-Salam-Weinberg model ($U(1)\otimes SU(2)$ part), which gives a unified description of the electromagnetic and weak forces and Quantum Chromodynamics (QCD), the theory of the strong interaction ($SU(3)$). Gauge theories introduce vector fields to enforce invariance of the theory with respect to local gauge transformations

[1] Because there is little risk of confusion, the Lagrangian density is often just called "Lagrangian".

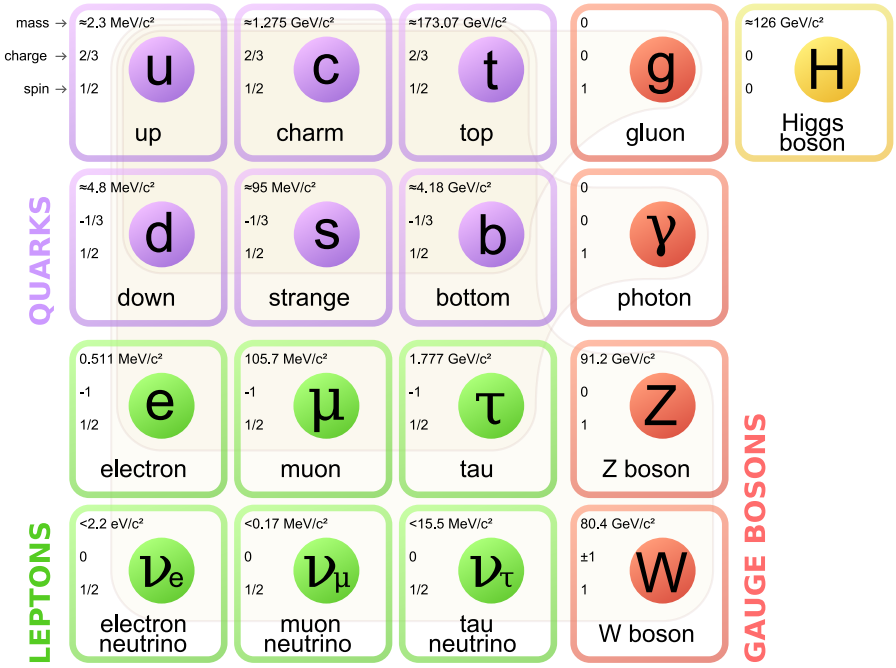


Figure 1.1: The particle content of the Standard Model. Fermions (leptons and quarks) come in three pairs called *generations*. Each generation has the same configuration of electromagnetic charge, but vastly different particle masses. The four gauge bosons (gluons, photons, W and Z bosons) serve as mediator particles for the strong, electromagnetic and weak interactions. The Higgs boson is the most recent addition to the set of experimentally observed particles. The light overlayed shading highlights the groups of fermions interacting with a given boson. Taken from [5].

in their respective group. The quantized excitations of the vector fields give rise to the physical gauge bosons. For each group generator, one boson is introduced: the photon ($U(1)$), the $W^{\pm}$ and Z bosons ($SU(2)$) and eight gluons ($SU(3)$). The $SU(2)$ and $SU(3)$ theories are non-Abelian: Their generators do not commute, which causes boson-boson interactions.

The electroweak model is explicitly parity violating. The left-chiral fermion fields in each generation form $SU(2)$-doublets, while the right-chiral fields transform as singlets.

To incorporate the non-zero boson and fermion masses, a scalar complex field, which transforms as an $SU(2)$-doublet, and a corresponding potential are introduced. By chosing the potential in a way that causes the scalar field to acquire a non-zero vacuum expectation value ("Mexican hat"-shape), mass terms for fermions[2] and the $W^{\pm}$ and Z bosons are generated. Because the Lagrangian is still locally $SU(2)$ symmetric, while the vacuum state of the universe is not, the $SU(2)$ symmetry is said to be "broken spontaneously".

An overview of the particle content of the SM is shown in fig. 1.1.

1.1.2 Perturbation Theory

Once the Lagrangian of a theory is known, it is possible to derive the equations of motion, which are however generally not analytically solvable. In order to be able to perform cross-section calculations, the methods of perturbation theory are used. An expression for the desired observable is derived as a series expansion in terms of the coupling strength of the interaction. If the coupling is sufficiently small, the series converges quickly and its leading terms will already give reliable results. This is generally true for the electroweak but not for the strong interaction, because the energy dependence of α_S prevents convergence for small momentum transfers (cf. sec. 1.1.3).

[2]Right-chiral neutrinos are explicitly not included in the model, thus preventing neutrino mass terms from appearing.

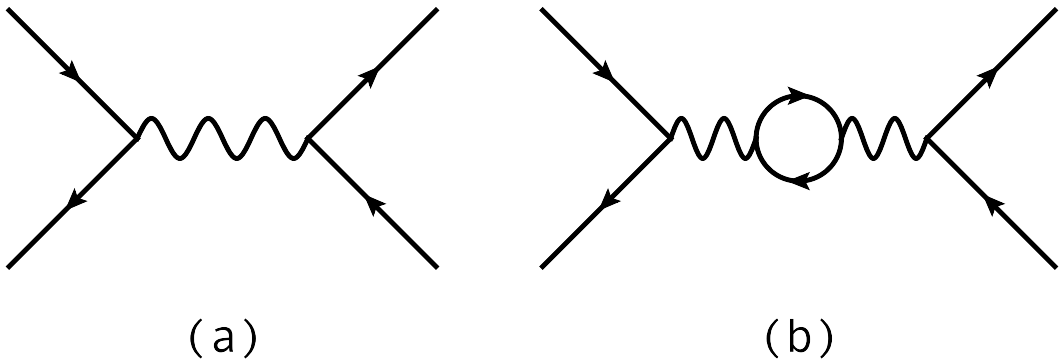


Figure 1.2: Feynman graphs of s-channel fermion-fermion scattering. Straight lines represent fermions, wavy lines bosons. The arrow direction of the fermion lines differentiates between incoming and outgoing fermions, which are equivalent to outgoing and incoming anti-fermions, respectively. While both diagrams have the same initial and final states, the intermediate propagator differs. The influence of (b) on the process cross-section is not included in an LO calculation. (b) is also an example of the kind of loop diagram that makes renormalisation necessary.

The accuracy of perturbative calculations is classified by the expansion order to which it is performed. A *leading-order* (LO) calculation takes into account only terms proportional to the power of the coupling which give the first non-vanishing contribution to the observable. Higher order calculations are referred to as *next-to-leading* order (NLO), next-to-next-to-leading order (NNLO) and so on.

The terms of the perturbative series expansion are commonly represented using Feynman diagrams. An example of two diagrams representing two terms of the power series describing fermion-antifermion scattering in the s-channel is given in fig. 1.2.

1.1.3 Renormalisation

An important feature in the construction of quantum field theories is renormalisation [1, 2]. In calculating amplitudes of Feynman diagrams with loops (e.g. fig. 1.2 (b)), momentum integration in the range $(0, \infty)$ needs to be performed, causing some of the integrals to diverge. The divergences are a result of the implicit extrapolation of SM behavior to energy scales beyond its validity. In order to obtain physical calculation results at experimentally accessible energy ranges, the integration is cut off at an unphysical scale Λ and *counter terms* are added to the Lagrangian to prevent the results from depending on Λ. The additional terms effectively shift interaction couplings and particle masses from the "bare" input parameters to the physically observable values.

Renormalisation introduces an energy scale dependence of the interaction strengths, which can be calculated in the *renormalisation group* framework as a function of the *renormalisation scale* μ_R. The μ_R dependence is referred to as *running coupling*. The fine-structure constant α, which defines the coupling in the electromagnetic interaction, increases with increasing energy scale μ_R, which is equivalent to decreasing length scale l.

$$-\frac{d\alpha}{d\mu_R} = \frac{d\alpha}{dl} < 0$$

This behavior is the reason for the familiar Maxwellian fall-off of the Coulomb field with increasing distance from the central charge. For the strong coupling constant α_S, the behavior is reversed.

$$-\frac{d\alpha_S}{d\mu_R} = \frac{d\alpha_S}{dl} > 0$$

The increasing strength of the strong interaction with increasing length scales leads to *quark confinement*, which prevents the observation of single quarks and forces the creation of color neutral

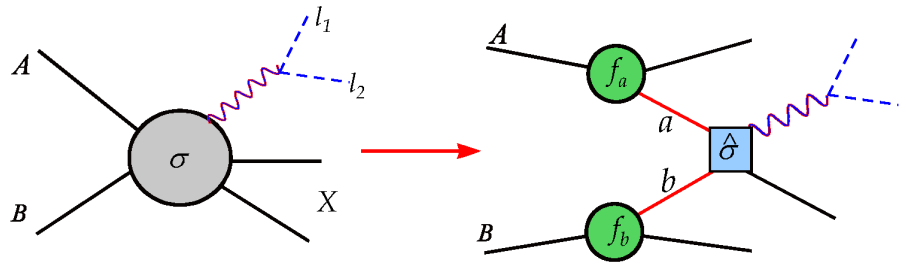


Figure 1.3: Schematic showing the factorisation of hadron-hadron cross-section calculations. A and B are incoming hadrons, a and b are their partons and X is an arbitrary final state. The use of parton distribution functions f allows to write the hadron-hadron cross-section in terms of the calculable parton-parton cross-section $\hat{\sigma}$. Image taken from [6].

hadrons[3]. The increase of α_S towards low μ_R creates a regime of *non-perturbative* QCD, where the perturbation series does not converge and thus perturbative calculations fail. In this regime, which defines the behavior of hadronic showers, modeling can only be performed by Monte Carlo methods.

While the energy scale dependence can be shown to cancel in the full perturbative series, a residual scale dependence is present in fixed-order calculations. The uncertainty associated with the choice of renormalisation scale is discussed in sec. 4.5.3.

1.1.4 Factorisation

In energy ranges above $\approx 1\,\text{GeV}$, perturbative methods allow for the calculation of parton level cross-sections, i.e. for the example case of incoming u and d quarks and final state X, the cross-section $\sigma_{ud\to X}$ can be calculated. This is not sufficient for cross-section calculations in proton-proton collisions, because the momentum distribution inside the proton needs to be taken into account. The parton level cross-sections and the proton-proton cross-sections can be related in a factorised form.

$$\sigma_{pp\to X}(s,\mu_F) = \frac{1}{2}\sum_{i,j}\int_0^1 \mathrm{d}x_1 \int_0^1 \mathrm{d}x_2 \;\; f_i(x_1,\mu_F) f_j(x_2,\mu_F) \;\; \sigma_{ij\to X}(x_1 x_2 s, \alpha_S(\mu_F))$$

The summation over i and j runs over parton flavors, $x_{1,2}$ are the fractions of proton momentum carried by the partons participating in the interaction. s is the center-of-mass energy of the proton-proton system, sx_1x_2 that of the parton-parton system. $f_i(x,\mu)$ is called parton distribution function and describes the probability to find a parton of flavor i at the momentum fraction x and effective scale μ in the proton. μ_F is the factorisation scale.

The factorisation approach disentangles the non-perturbative low-energy interactions of partons inside the proton from the perturbative high-energy "hard" scattering process. The non-perturbative parts are collected in the parton distribution functions, which are process independent and can be determined experimentally. While their dependence on x is not calculable, evolution in μ_F with respect to a given scale μ is possible in the DGLAP framework.

Similar to the renormalisation scale μ_R, the factorisation scale μ_F is an unphysical parameter which is commonly chosen to be $\mu_F \approx |Q|$, where Q is the momentum transfer in the scattering process. However, the parameter choice is not unique and introduces a source of uncertainty into the calculation (cf. 4.5.3).

1.1.5 Defects of the Standard Model

While the SM performs very well in describing wide ranges of experimental observations, it is known to be incomplete. Important features of the universe, such as the observed imbalance between matter and antimatter, gravitational interactions and the presence of Dark Matter are not described within

[3] An exception to this rule is the top quark, which is sufficiently heavy to decay before hadronising.

the SM. Additionally, the recent discovery of an SM-like Higgs boson at a mass of $\approx 125\,\text{GeV}$ raises the *hierarchy problem*. If no BSM phenomena are present up to the Planck scale ($\mathcal{O}(10^{19}\,\text{GeV})$), where gravitation becomes relevant at sub-atomic length scales, radiative corrections to the Higgs field strength are much larger than its observed value. The SM can then only produce the observed Higgs boson mass if its internal parameters are chosen to cancel the radiative corrections over more than ten orders of magnitude. This necessity of extreme *fine-tuning* is commonly considered to be a sign of an incomplete understanding of physics above the electroweak scale ($\mathcal{O}(10^2\,\text{GeV})$). Finally, particle masses are not calculable from first principles within the SM, which raises the question whether a more fundamental theory of nature would be able to predict the observed particle masses from a smaller set of parameters. Although incomplete, this list of open questions already motivates many of the recent searches for physics beyond the Standard Model (BSM).

1.2 The CMS Experiment

The Compact Muon Solenoid (CMS) [7] is a multi-purpose detector located at the Large Hadron Collider (LHC) [8] at CERN in Geneva. The LHC is a circular collider with a circumference of 27 km and provides proton-proton and heavy-ion collisions. In its running period in 2012 it delivered proton-proton collisions at a center-of-mass energy of 8 TeV and instantaneous luminosities of up to 7.7 Hz/nb [9].

In the collider context, cylindrical coordinates are used: with the origin in the nominal interaction point of the proton beams, the z-axis is defined to be in the beam direction. r is the radial distance to the beam axis, the azimuth angle ϕ is measured relative to the connecting vector between the centers of detector and accelerator ring. A commonly used additional coordinate is the pseudo-rapidity η.

$$\eta = -\log(\tan(\theta/2))$$

θ is the polar angle with respect to the beam direction. Thus, the direction perpendicular to the beam axis is denoted by $\eta = 0$, while the z-direction corresponds to $|\eta| \to \infty$.

An expansive description of CMS is provided in [7]. In the following a summary is given of detector layout (sec. 1.2.1), triggering (sec. 1.2.2) and particle reconstruction (sec. 1.2.3).

1.2.1 Detector

At the heart of CMS is the eponymous magnet, which is a superconducting solenoid coil with an operating temperature of 4.5 K, providing a central magnetic field of 3.8 T. The coil houses a tracking system as well as electromagnetic and hadronic calorimeters (ECAL and HCAL, respectively). Following the magnet's symmetry, the detector sub-systems are divided into a cylindrical *barrel* and two planar *endcap* parts. While the exact border depends on the individual detector components, the more central barrel region generally extends to $|\eta| \approx 1.5$, while the endcap region encompasses the range $1.5 < |\eta| < 3.0$.

The tracking system is silicon based and consists of an inner pixel detector surrounded by a larger volume of silicon strip layers. The pixel detector provides a very high granularity close to the beam, yielding a single point resolution of $\approx 20\,\mu\text{m}$ for each traversing charged particle. The surrounding strip detector records additional hits with a single point resolution in the range $\approx 50 - 300\,\mu\text{m}$. The tracker extends up to $|\eta| = 3.0$. The tracking system allows for charge assignment, vertex measurements for τ- and b-tagging and is the main source for sensitivity to the momentum of charged particles in CMS.

Surrounding the tracking system is the ECAL, which consists of scintillating lead-tungstate (PbWO4) crystals. Lead tungstate is especially suited as calorimeter material due to its large mass density, short radiation length and sufficient radiation hardness. Test beam measurements found the ECAL energy resolution to be better than 1 % for electrons with energies above about 20 GeV [7]. The ECAL is the only part of CMS that is directly sensitive to photons. The ECAL barrel part ends at $|\eta| = 1.479$, the endcap instrumentation begins at $|\eta| = 1.653$, with an uninstrumented gap in between.

The ECAL is surrounded by a sampling HCAL, which uses brass as an absorber and plastic scintillator as the active material. It is less granular than the ECAL but provides an energy measurement

for charged as well as neutral hadrons. Because the material budget is limited by the coil radius, an additional HCAL component is positioned directly outside the coil in the central detector region. It uses the coil material as an absorber and increases the probability of hadronic showers to be fully contained within the detector volume.

The outermost part of the CMS detector is the muon system, which employs gaseous detectors for muon tracking, momentum measurement and triggering. In the barrel region, there are four layers positioned in and around the iron return yoke comprising a total of 250 drift tubes (DTs) and 172 000 sensitive wires. Due to the relatively low rates of incoming particles, the chambers can be used for tracking. In the endcap region, muons are detected using 468 Cathode Strip Chambers (CSCs), which are designed to withstand the larger particle hit rates of up to $1\text{kHz}/\text{cm}^2$ compared to the barrel region. DTs and CSCs provide position and momentum measurements of muons and are in principle usable for triggering. As an independent source of online triggering information, Resistive Plate Chambers (RPCs) are placed close to the DTs in six (three) layers in the barrel (endcap) region. RPCs are chosen for their fast response time, which allows to clearly assign trigger signals to LHC bunch crossings.

1.2.2 Triggering

In the 2012 data taking period at the LHC, bunches of 10^{11} protons were collided every 50 ns. On average, there are ≈ 25 proton-proton interactions per bunch-crossing, which corresponds to an interaction rate of 0.5 GHz [10]. While it is possible to read out the CMS detector modules at such a rate, the sheer amount of information prohibits indiscriminate storage. This makes necessary a system of filtering, which for each collision quickly decides if it is to be stored or discarded. This filtering is called *triggering*. The CMS trigger system is divided into two steps: The Level-1 trigger and the High-Level trigger.

1.2.2.1 Level 1 Trigger

The Level-1 (L1) Trigger [10] is a fast, hardware based triggering system. It is designed as a first rough selector to reduce the rate of incoming events by a factor large enough to permit more complex analysis in subsequent steps. At the L1 stage the detector information from calorimeters and muon system is available. This prominently excludes tracking information, since track reconstruction is rather complex and thus too slow. A large number of different Level-1 trigger paths is available, covering the signatures of different objects at different scales. The individual L1 triggers are used as *seeds* for more complex High-Level trigger paths, i.e. the High-Level paths require a certain L1 trigger to have fired. The L1 trigger reduces the event rate to less than 100 kHz [10], which are then analysed further.

1.2.2.2 High-Level Trigger

The High-Level Trigger (HLT) [11] is a software based triggering stage. Using the events preselected by the L1 trigger, it is designed to reduce the event rate such that storage becomes possible. At this triggering stage, the complete detector information is available and track reconstruction is performed. Also, the software used in later offline analysis is run in simplified versions at this online stage. This includes a variant of the Particle Flow algorithm, which combines signatures left in different detector sub-modules by the same particle. The close similarity between on- and offline reconstruction permits a close correspondence of triggering and offline event selection and only small fractions of events have to be discarded at later analysis stages. While ensuring satisfactory performance, the software run at the HLT level is $\mathcal{O}(10^2 - 10^3)$ faster compared to the offline stage. An average event rate of 400 Hz passes the HLT and is stored offline [10].

1.2.3 Particle Reconstruction

The many detector sub-modules deliver electronic signals which on their own give little information about the collision event at hand. In order to learn about the observed interaction, particle reconstruction must be performed: A large range of signals must be collected and condensed to the point where

particle signatures become recognizable. This task is closely related to object selection (cf. section 2.2), which is largely analysis dependent. Here, an overview over general reconstruction is given.

1.2.3.1 Particle Flow

In CMS, a large part of particle reconstruction is performed by the Particle Flow (PF) algorithm. The aim of the algorithm is to combine the measurements from several sub-detectors to optimize the resulting reconstruction performance. The reconstruction is performed in three steps: (i) clustering of energy deposits in the calorimeters, (ii) linking of tracks, ECAL and HCAL clusters into so-called blocks and (iii) identification of particle candidates within blocks. The procedure is described extensively in Refs. [12, 13]. A concise, qualitative account is given here.

Clustering Clustering combines energy deposits in adjacent calorimeter units (cells in the ECAL, towers in the HCAL). Starting from the units with locally maximal energy deposits (*seeds*), adjacent cells are added as long as they satisfy certain criteria, such as having detected a minimal amount of energy. The goal is to separate chunks from different sources.

Linking Linking aims to group together deposits in different sub-systems that stem from the same particle. In a first step, tracks measured in the tracker are extrapolated to the calorimeters and muon chambers. If an extrapolated track coincides with the location of an energy cluster or a track in the muon system, a link is established. To link clusters without corresponding tracks, a "cluster envelope" is defined, which allows for a rough correspondence between HCAL and ECAL deposits to be established. Sets of linked deposits are called blocks.

Particle Identification The heart of Particle Flow lies in the identification of particles[4], which is performed in an intuitive manner. The blocks are scanned for known particle signatures. In descending order of signature clarity, particles are identified and removed from the block, and the scanning is repeated. Clarity here means how sure we can be that a signature corresponds to a certain particle. E.g., the clearest signature is considered to be that of a *global muon*, i.e. a muon with matching track in tracker and muon system. Since matching means geometric matching as well as energetic compatibility and since only a small fraction of particles will cause a signal in the muon chambers, it is rather unlikely to find such a signature by chance. Thus, global muons are identified and removed in the first iteration. In the following iterations, signatures for electrons, charged hadrons, photons and neutral hadrons are checked until no blocks are left. In all three of the steps above, additional criteria and sub-methods are used to define a robust algorithm that optimally uses the advantages of different sub-detectors. E.g. when combining energy clusters from different parts of the detector, it is wise to give higher priority to the directional information gathered from the more granular sub-detector. Finally, the output of the PF algorithm is simply a list of candidates.

1.2.3.2 Jet clustering

Stable[5] particles such as electrons and muons can be reconstructed directly. This is not true for gluons and quarks, because confinement introduces the additional layers of parton showering and hadronisation that prevent the observation of single partons. After showering and hadronisation, there is a number of resulting final state particles for every outgoing gluon or quark from the hard interaction (cf. sec. 1.1 and 2.4.1). Because of the stochastic nature of the process, many final configurations are possible.

In order to be able to deduce information about the outgoing partons from the resulting final state particles, *jets* are introduced. A jet is a set of Particle Flow candidates that is grouped together by a clustering algorithm. The algorithm considers all PF candidates in the event and combines them into sets according to criteria like geometric and energetic separation between candidates. In the ideal case, all particles resulting from the hadronisation of one parton end up in the same jet which allows

[4] In Particle Flow terminology, the particles are referred to as *candidates*.

[5] In the context of CMS, all particles that are not expected to decay within the detector are considered stable. E.g. muons decay with a lifetime $\gamma c\tau > 660\,\mathrm{m}$ [4]. At typical LHC energies, muon decays will happen well outside the confines of the detector, which makes the muon a stable particle.

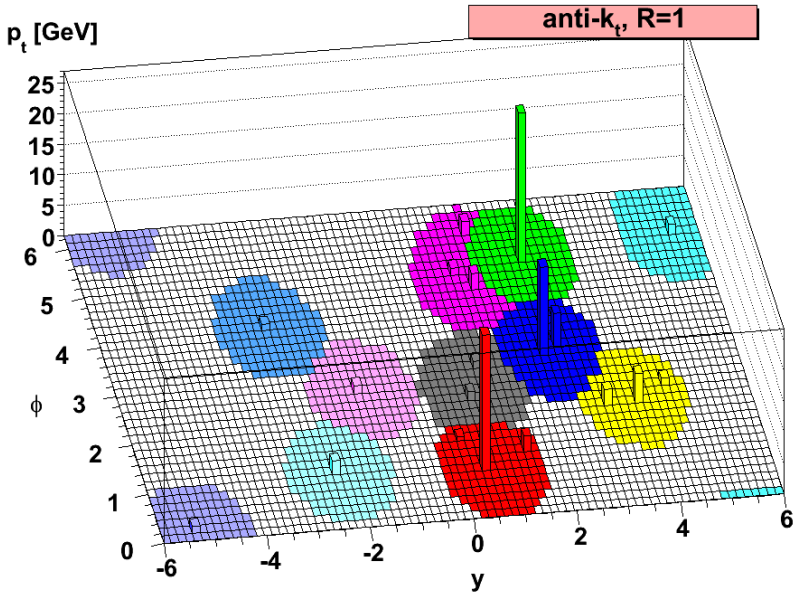


Figure 1.4: Clustering behavior of the `Anti-`k_t algorithm. In the y-ϕ-plane, a smaller number of high-momentum particles (large z-coordinate) is clustered with a large number of low-momentum particles (small z). Resulting jet shapes are highlighted in different colors. For sufficiently isolated high-momentum particles, the algorithm simply places a conical jet around the particle. Where this construction would lead to overlapping jets, the shapes are cropped depending on the momentum ratio between jets. Image taken from [14].

us to identify jet and parton properties. For example, an intuitive definition of a jet algorithm would simply determine the highest-energetic object in the event as a jet seed. Then, all objects inside a geometric cone of certain radius around the seed are combined into a jet and the collected particles are removed from the event. The process is repeated until no objects remain.

When choosing a jet algorithm definition, it is important to ensure that jet behavior is calculable so that theoretical predictions can be made, e.g. for differential cross-sections. The two most important aspects for calculability are collinear and infrared safety [15]. A jet algorithm is said to be collinear safe if the resulting jet does not change if one of the input objects is replaced by two daughter objects sharing the parent object's energy and momentum, even if the angle between the two daughters' momenta is small. This behavior is realised in gluon splitting and thus occurs quite frequently in hadronisation. An algorithm is infrared safe if additional low-momentum particles (such as commonly present from the underlying collision event) do not influence the result. The above example of a naive jet algorithm is evidently collinear unsafe. If the highest energetic particle is replaced by two particles each carrying half the total energy, a different initial seed may be chosen, which may radically change the result of the clustering process.

A common choice of jet clustering algorithm is the `Anti-`k_t algorithm [14]. It is a special case of a general definition of sequential recombination algorithms. Sequential recombination algorithms perform iterative steps of merging groups of objects, without requiring one group to be finished before others are considered. Anti-k_t operates by defining a measure of distance $d_{ij} = d_{ji}$ between two groups of objects i and j as well as a "beam distance" d_{iB}, which only depends on the properties of object group i. While the two-group distance d_{ij} acts as an ordering quantity for the combination of groups i and j, d_{iB} acts as a stopping criterion for combinations involving group i. In the beginning, every single object in the event (in the case of MUSiC "objects" means PF candidates such as muons, electrons, etc.) is alone in its own, separate group. In each iteration of the algorithm the minimal

values $d_{ij,min}$ and $d_{iB,min}$ are calculated. If $d_{ij,min} < d_{iB,min}$, i.e. the minimal distance between two groups is smaller than the minimal beam distance of all groups, the groups i and j are merged. If on the other hand $d_{ij,min} > d_{iB,min}$, i.e. the overall minimal distance is between group i and beam, the group i is called a jet and removed from the event. If there are still groups left in the event, the next iteration begins. The distance measures are defined as

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \tag{1.1}$$

$$d_{iB} = k_{ti}^{2p}, \tag{1.2}$$

$$\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2} \tag{1.3}$$

Where y_i and ϕ_i are the rapidity[6] and azimuth angle of object i, respectively. The parameter p quantifies the relative behavior of the energy (first factor in d_{ij}) and geometric (second factor) scales. For `Anti-`k_t $p = -1$. Choosing $p = 0$ and $p = 1$ yields, respectively, the distance measures defined in the k_t- and *Cambridge/Aachen* algorithms. R is a continuous parameter that may be chosen freely. An in-depth discussion of the algorithm's behavior is presented in [14], some important aspects are illustrated in fig. 1.4. For isolated high-momentum objects surrounded by objects with sufficiently lower momentum, conical jets are realised. When multiple jets overlap and no single momentum dominates, the resulting shapes are cones cropped according to the momentum balance between different objects. Thus, although the clustering algorithm seems much more complex than our simplistic example from above, the resulting jets still largely conform to the intuitive expectation. The CMS reconstruction uses the FastJet [16] implementation of the `Anti-`k_t algorithm.

1.3 MUSiC

1.3.1 Previous Results

The MUSiC analysis concept has been in continual development since before CMS data-taking began. The idea of performing a comprehensive comparison of collider data and SM simulation dates back to the L3 experiment at LEP [17, 18], which was later applied to data from D0 experiment at the Tevatron collider [19, 20]. A first detailed study of the physics potential of MUSiC was finished in 2009 [21, 22]. The analysis concept was subsequently applied to 7 TeV CMS data collected in 2010 and 2011 [23, 24]. The current state of the analysis is in large parts based on the framework used for the 7 TeV analysis, which is applied to collision events with leptons in 2012 data in [25]. While extensive simulation studies were performed to ensure the analysis performance, no BSM phenomena were observed. A similar analysis is performed at the LHC's other large multi-purpose experiment ATLAS [26, 27].

This document aims to extend the scope of the analysis to final states with jets and the identification of jets with bottom quarks ("b tagging "). The application of b tagging to the MUSiC analysis was previously studied in simulation [28] and CMS data collected in 2010 [29].

1.3.2 Method

The aim of the MUSiC is to avoid any bias arising from a self-inflicted restriction of the field of view. To accomplish a maximally broad view, the event selection is not designed according to sensitivity criteria or a pre-determined notion of interest in certain topologies. Instead, criteria based on reconstruction quality are used to select events and identify objects. Data is analyzed in two steps: Classification and Scanning.

1.3.2.1 Classification

Since no final state selection is applied, a vast number of event topologies is examined in MUSiC. To analyze these very different events in a fully inclusive way would be largely unproductive, since

[6] Rapidity is defined as $y = tanh^{-1}(p_z/E)$. For particles with $p \gg m$, it coincides with the pseudo-rapidity η (cf. sec. 1.2).

features of individual physical processes would be masked by the sheer number of events. Therefore, event classes are introduced and events are sorted into them based on their final state, i.e. which objects are reconstructed in the detector. Three types of event classes are considered:

- **Exclusive Event Classes** contain events with a final state that exactly equals the name of the event class. E.g. the 2μ exclusive event class will only contain events with two muons in the final state. An event with two muons and one jet would not enter this class. Thus, exclusive event classes implicitly veto any additional objects passing MUSiC identification criteria.

- **Inclusive Event Classes** do not employ a veto on additional objects. Every event is included in any inclusive event class defined by any subset of its final state particles. Therefore, inclusive classes are not disjoint sets of events. E.g., the event with two muons and a jet will enter the 2μ inclusive class as well as the 2μ+1jet, 1μ+1jet, 2μ and 1μ inclusive classes.

- **Jet Inclusive Event Classes** are a step between exclusive and inclusive classes. While additional jets are permitted in these classes, any other objects are vetoed. Due to initial and final state radiation, each physical process is not only visible in its *pure* final state (e.g. Z boson production in dilepton final states), but also in final states with added jets (e.g. dilepton and one or multiple jets). Jet inclusive classes take this into account and collect all jet multiplicities of a certain final state. They also serve as a tool for comparisons with the results of dedicated analyses which often do not use strict jet vetoes to enhance the selection efficiency for a desired process.

The necessary event and object selection is described in sec. 2.1. The thus separated event topologies allow for discrimination between different processes. The classification is performed in its entirety for data and simulated events.

1.3.2.2 Scanning

The large number of event classes and distributions produced in MUSiC requires an automated procedure to assess the degree of agreement of simulation and data. For a number of kinematic distributions per class, a local p-value calculation is performed to quantify the (dis-)agreement in any bin region in a distribution. By iterating and performing the calculation in every connected bin region of a distribution, the region with the most significant deviation between data and simulation, i.e. the smallest local p-value, is determined. This region is called the Region of Interest (RoI).

Different distributions, even of the same event class, will in general have different numbers of connected bin regions, simply because they have different numbers of total bins. The more bin regions are present in a distribution, the more likely it is to observe a significant deviation just by chance. This phenomenon is called the "Look-Elsewhere-Effect" (LEE). A pseudo-experiment driven method to take it into account is described in sec. 1.3.2.5.

The considered kinematic distributions are detailed in section 1.3.2.3. A pragmatic description of the p-value calculation is given in section 1.3.2.4. A more in-depth treatment including implementation details of the scanning procedure may be found in [24].

1.3.2.3 Kinematic distributions

For each event class, three kinematic distributions are evaluated:

1. The scalar sum of transverse momenta $\Sigma|p_T|$

$$\Sigma|p_T| = \sum_i |p_{T,i}| \tag{1.4}$$

The left hand-side of the equation is a symbolic short-hand that will be used throughout this document. For exclusive classes, the momenta of all particles in the final state are summed over. In inclusive classes, only the particles defining the event class name are included. If this choice

is not unambiguous (e.g. an event with three muons contributing to the 2 muon inclusive class), the particles with the highest p_T are used. $\Sigma|p_T|$ is not sensitive to the angular distribution of reconstructed final state particles and gives a reliable measure of the total energy content of the event.

2. The missing transverse energy E_T^{miss}.

$$E_T^{\text{miss}} = \left|\vec{E}_T^{\text{miss}}\right| = \left|-\sum_i \vec{p}_{T,i}\right| \tag{1.5}$$

if it is significantly different from zero in an event. Because the incoming partons only carry significant momentum in beam direction, conservation of momentum dictates that the vectorial sum of particle momenta in the plane transverse to the beam vanishes. However, due to particles such as neutrinos escaping the experiment undetected, as well as particle momenta being mismeasured, this constraint is not necessarily fulfilled by the measured momenta. E_T^{miss} is thus a measure of the momentum imbalance in an event. It is directly sensitive to the production of additional undetected particles, such as Dark Matter.

3. The invariant mass M_{inv}

$$M_{\text{inv}} = \sqrt{\left(\sum_i E_i\right)^2 + \left(\sum_i \vec{p}_i\right)^2} \tag{1.6}$$

again of the particles defining the event class name. If the class includes E_T^{miss}, the transverse mass $M_T$is used instead.

$$M_T = \sqrt{\left(\sum_i E_{T,i}\right)^2 + \left(\sum_i \vec{p}_{T,i}\right)^2} \tag{1.7}$$

For convenience, both quantities are referred to as invariant mass M_{inv}. The invariant mass is useful to detect resonant decays of heavy particles into the reconstructed final state particles, which is a common scenario in proposed models for beyond the Standard Model (BSM) physics. Also, it offers some sensitivity to the angular kinematics of the final state.

The choice of kinematic distributions is motivated by the model independence of the analysis. All quantities must be well-defined and interpretable over large ranges of different final states and physics processes. This requirement excludes many of the quantities considered in dedicated analyses, which are specifically tailored either to a certain final state, a certain physical phenomenon or both. An example of this would be the angular separation between the two leading jets in an event [30]. While it is well-defined for all final states with at least two jets, it is primarily of interest for QCD phenomena and thus not a very attractive quantity in final states dominated by W boson production.

1.3.2.4 p-value

MUSiC uses a p-value calculation to quantify the agreement between data and simulation. Generally, a p-value for some discrete test statistic T is defined as the probability of finding a more extreme value of T than the measured one.

$$p = \sum_{\text{More extreme T}} P(T) \tag{1.8}$$

where $P(T)$ is the probability density function (PDF) of T. In MUSiC, the difference between observed events N_{data} and SM prediction N_{SM} in any given region is used as a test statistic. To model the probability density of this statistic, two components are needed:

- **Systematic uncertainties** lead to a shift of the expected mean number of events in a region. This shift is modeled by a Gaussian PDF, which is truncated at zero and re-normalized.

- **Statistical fluctuation** of the event count around the expected mean is modeled by a Poissonian PDF centered on the expected mean.

The total PDF is described by the convolution of the Gaussian and Poissonian terms. The total p-value is calculated as:

$$p_{\text{data}} = \begin{cases} \sum_{i=N_{\text{data}}}^{\infty} C \cdot \int_0^{\infty} \mathrm{d}\lambda \, \exp\left(-\frac{(\lambda - N_{\text{SM}})^2}{2\,\sigma_{\text{SM}}^2}\right) \cdot \frac{e^{-\lambda}\,\lambda^i}{i!} & \text{if } N_{\text{data}} \geq N_{\text{SM}} \\ \sum_{i=0}^{N_{\text{data}}} \underbrace{C}_{\text{normalisation}} \cdot \int_0^{\infty} \mathrm{d}\lambda \, \underbrace{\exp\left(-\frac{(\lambda - N_{\text{SM}})^2}{2\,\sigma_{\text{SM}}^2}\right)}_{\text{systematics}} \cdot \underbrace{\frac{e^{-\lambda}\,\lambda^i}{i!}}_{\text{statistics}} & \text{if } N_{\text{data}} < N_{\text{SM}} \end{cases} \tag{1.9}$$

For any given region, three input numbers are needed: The expected number of events in the region N_{SM} and its uncertainty σ_{SM}, which are determined from Monte Carlo simulation, and N_{data}, the observed number of events in the region. The integral implements the convolution between the partial PDFs, while the sum is that of eq. 1.8. The two cases differentiate between observed excesses ($N_{\text{data}} \geq N_{\text{SM}}$) and deficits ($N_{\text{data}} < N_{\text{SM}}$), which leave the PDF unchanged but reverse the notion of "more extreme deviations" and thus the direction of summation.

1.3.2.5 Look-Elsewhere Effect

To account for the Look-Elsewhere effect (LEE), the p-value, which gives the probability to find a more extreme deviation than the observed one *in a given connected bin region*, is converted to the probability to find such a deviation *in a given distribution*, which is called $\tilde{p}$. In other words, one needs to calculate a p-value using the previously calculated p-value as a test statistic. Unfortunately, the PDF one would need to calculate an expression like eq. 1.9 is not known beforehand. It must be calculated using pseudo-experiments. The pseudo-experiment method is performed separately for each distribution in each event class and based solely on the SM simulation, i.e. the data do not play a role except in providing p_{data}.

A pseudo-experiment is generated by randomising ("dicing") the bin contents of the SM prediction in a distribution according to their uncertainties. Each bin is diced two times: First using a Gaussian PDF centered on the nominal SM bin content to model systematic uncertainties and secondly using a Poissonian PDF centered on the systematics-shifted bin content to model statistic fluctuation. In the systematic shifting, correlations are taken into account. For example, the uncertainty associated with the cross-section of a simulation sample has a correlated effect in all bins. Thus, it is diced once and varied by the same amount in all bins for one pseudo-experiment. In contrast, the uncertainty due to the finite number of generated simulation events is not correlated between different bins and diced separately for each bin.

The pseudo-data obtained by this method are scanned for deviations from the SM simulation using the same process as in data, which yields a RoI and corresponding p-value for each pseudo-experiment.

Up to 10^5 pseudo-experiments are generated per distribution. For each pseudo-experiment, the calculated p-value is saved. This collection of p-values then serves to approximate the PDF of the p-value as a test statistic under the SM hypothesis. The $\tilde{p}$-value is calculated as the ratio of the number of pseudo-experiments with a smaller p-value than that in data and the total number of pseudo-experiments.

$$\tilde{p} = \frac{N_{\text{pseudo}}(p_{\text{pseudo}} < p_{\text{data}})}{N_{\text{pseudo, all}}}$$

The set of p-values is used to calculate a $\tilde{p}$-value for each pseudo-experiment. The $\tilde{p}$-values from pseudo-experiments are used to provide an expected shape of the distribution in data.

In order to save computing time and thus enable the large number of 10^5 pseudo-experiments per considered class, stopping criteria are introduced. The calculation of $\tilde{p}$ for a class is skipped entirely if:

- The class contains no observed events and the expected number of events is less than one or compatible with zero within three times its uncertainty.
- The p-value for the class is greater than 0.3.

In these classes, it is not necessary to calculate a $\tilde{p}$-value to determine that no significant deviation is present. Furthermore, the generation of pseudo-experiments is stopped after $2 \cdot N_{100}$ rounds, where N_{100} is the number of rounds it took to generate 100 pseudo-experiments with $p < p_{\text{data}}$. This criterion is expected to only lead to a significantly early termination for event classes with $\tilde{p} > 0.1$, which do not require very high precision in $\tilde{p}$.

1.3.3 Presentation of Results

MUSiC results are presented in three ways:

- For every class and each distribution, a plot comparing data and simulation is generated (e.g. fig. 3.6). Simulated events are shown in colored histograms, grey hatched areas represent simulation uncertainties and black points show data. The $\tilde{p}$-value is given and blue dashed lines indicate the RoI. From the plot, the SM processes contributing to the event class can be identified and the kinematic features of the distribution are visible. This is useful if a class is to be considered in detail, e.g. for cross-checking with a dedicated analysis.
- The large number of classes and distribution of course prevents the analyst from manually looking at every single distribution plot. A condensed view of the complete set of results is available in the form of an *integral scan*, in which the shape information of distributions is neglected and each individual distribution is treated as a single-bin counting experiment. While this method of course removes any shape sensitivity, it allows to directly communicate the total content of each event class, which is sensitive to the total cross-section.
- Finally, an ultimately condensed view of the results is available in the distribution of the $\tilde{p}$-values of all classes[7] (e.g. fig. 4.3). While many small discrepancies with large $\tilde{p}$-values are expected, only a small number of event classes with significant deviations is expected. By comparing the distribution of $\tilde{p}$ in data and simulation, a de facto meta-analysis of all event classes is performed and the "agreement of agreement" can be gauged. Here, MUSiC takes full advantage of its large number of final states.

[7] Exclusive, inclusive and jet inclusive event classes are treated separately, because only the exclusive event classes are statistically independent.

2 Analysis

2.1 Event Selection

2.1.1 Data Streams

The MUSiC analysis uses the complete dataset recorded by the CMS experiment in the 2012 running period at $\sqrt{s} = 8\,\text{TeV}$. The baseline of events with satisfactory experimental conditions, i.e. without fatal accelerator or detector problems, are obtained from the 'golden json'[1] of the 22Jan2013ReReco reconstruction campaign [31]. Depending on the fired triggers in an event, it is sorted into one of multiple subsets called data streams. MUSiC uses a total of five data streams, which are listed in table 2.1. The data streams are not mutually exclusive: Events may pass multiple triggers and thus enter multiple data streams. To avoid double counting, the streams are cleaned against one another, i.e. events already present in one stream are rejected in all others.

2.1.2 Triggers

From the baseline of all events contained in the data streams, MUSiC selects events based on multiple triggers, which are listed in table 2.2. In addition to the single and double electron and muon triggers used in previous iterations of the analysis, the single-jet trigger `PFJet320` is used. It requires an online reconstructed Particle Flow jet with at least 320 GeV of transverse momentum. All of the triggers used in MUSiC were run unprescaled[2] during the complete 2012 data-taking. The `PFJet320` trigger fired with a rate of about 10 Hz at an instantaneous luminosity of $\mathcal{L} = 7\,\text{Hz/nb}$ [10].

[1] The full name of the 'golden json' is `Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt`.

[2] A *prescale* is introduced to artificially reduce the event rate of a trigger and reduces the effectively recorded integrated luminosity for this trigger.

Table 2.1: Data sets used as inputs to this analysis. The data were reconstructed in the `22Jan2013ReReco` campaign and include single and double lepton data streams as well as a jet data stream from all four experimental periods (*runs* A-D) of 2012. For the double muon data stream, parked data sets are used for all runs except A.

/SingleMu/Run2012A-22Jan2013-v1	/DoubleMu/Run2012A-22Jan2013-v1
/SingleMu/Run2012B-22Jan2013-v1	/DoubleMuParked/Run2012B-22Jan2013-v1
/SingleMu/Run2012C-22Jan2013-v1	/DoubleMuParked/Run2012C-22Jan2013-v1
/SingleMu/Run2012D-22Jan2013-v1	/DoubleMuParked/Run2012D-22Jan2013-v1
/SingleElectron/Run2012A-22Jan2013-v1	/DoubleElectron/Run2012A-22Jan2013-v1
/SingleElectron/Run2012B-22Jan2013-v1	/DoubleElectron/Run2012B-22Jan2013-v1
/SingleElectron/Run2012C-22Jan2013-v1	/DoubleElectron/Run2012C-22Jan2013-v1
/SingleElectron/Run2012D-22Jan2013-v1	/DoubleElectron/Run2012D-22Jan2013-v1
/Jet/Run2012A-22Jan2013-v1	
/JetHT/Run2012B-22Jan2013-v1	
/JetHT/Run2012C-22Jan2013-v1	
/JetHT/Run2012D-22Jan2013-v1	

Table 2.2: Trigger paths used in MUSiC and their official names. For each of the paths, requirements on the reconstructed final state objects are given. These ensure that the reconstructed event contains a particle that would nominally pass the trigger system. Momentum thresholds are chosen to avoid the turn-on effects in triggering efficiencies. For information on the trigger naming conventions, please refer to the sec. 2.1.2.

Object	**Name**	**Additional MUSiC Requirements**
Single Muon	HLT_IsoMu24_eta2p1	One isolated muon with $p_T > 25\,\text{GeV}$
Double Muon	HLT_Mu17_Mu8	Two isolated muons each with $p_T > 20\,\text{GeV}$
Single Electron	HLT_Ele80_CaloIdVT_TrkIdT	One isolated electron with $p_T > 100\,\text{GeV}$
Double Electron	HLT_Ele17_CaloIdT_CaloIsoVL_ TrkIdVL_TrkIsoVL_ Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL	Two isolated electrons each with $p_T > 25\,\text{GeV}$
Single Jet	HLT_PFJet320	One jet with $p_T > 400\,\text{GeV}$

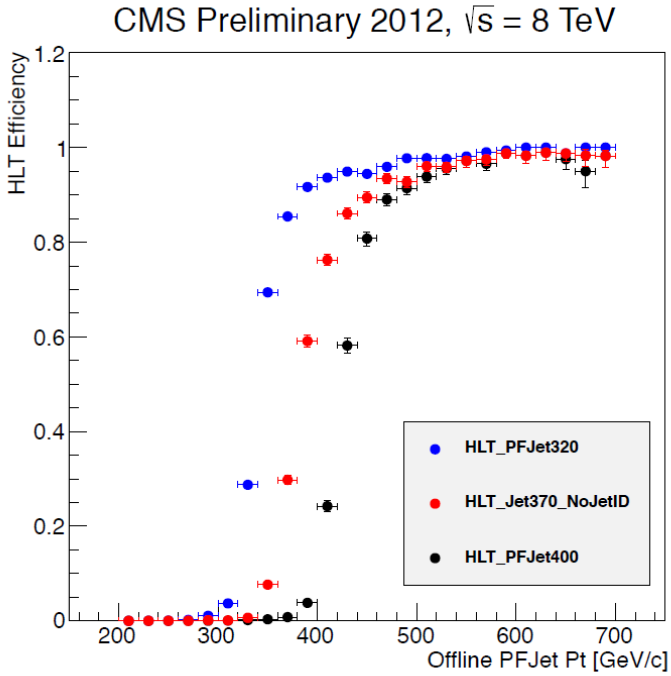


Figure 2.1: Efficiency curves for single-jet triggers. MUSiC employs the PFJet320 trigger (blue points). As a function of the offline reconstructed p_T the curves are shifted to the right due to additional jet energy corrections not applied at the HLT level. Events where the leading reconstructed jet has a transverse momentum of less than 400 GeV are discarded to avoid the steep efficiency slope. Image taken from [10].

Table 2.3: Geometric and p_T-acceptance for object identification in MUSiC.

Object	$p_{T,\min}$(GeV)	$\lvert\eta\rvert_{\max}$
Muon	25	2.1
Electron	25	2.5
Photon	25	1.442
Jet	50	2.4
B jet	50	2.4
E_T^{miss}	50	/

For each triggered event, it is required that the offline reconstructed final state passes the nominal trigger requirements. E.g. for events selected by the `PFJet320` trigger, an offline reconstructed jet with sufficient p_T is required for the event to be accepted. In addition to the p_T-thresholds applied at HLT level, stricter p_T-requirements are applied offline. This serves as a protection against trigger efficiency turn-on effects. The importance of the additional p_T criterion is illustrated by the efficiency curves for single PFJet triggers as shown in fig. 2.1. Because online and offline reconstructed p_T values may differ, the efficiency curve has finite slope. Events in this region are rejected to avoid an amplification of minor p_T mis-measurement.

The offline triggering requirements for all triggers are listed in tab. 2.2. In each trigger path name, the triggering object is given (e.g. "Mu" for muon), immediately followed by the online p_T-requirement. E.g. "Mu24" is short for "muon with $p_T > 24\,\text{GeV}$". The value immediately following the "eta" code word gives the maximum $|\eta|$-value for particles to be considered. For the electron triggers, the identification method is specified: Loose (L), Tight (T) or very tight (VT) working points of the calorimeter (Calo) or tracker (Trk) based identification (Id) and isolation (Iso) requirements are applied [32]. The "PF" code word in the jet trigger name indicates that a Particle Flow jet is required.

2.1.3 Event Filters

To ensure experimental purity of the data, additional event rejection is performed using event filters. These include rejection of events with mis-calibration issues, e.g. in a number of runs, the HCAL calibration via an injected laser beam caused fake energy deposits to be recorded. Events with this kind of method-related problems are rejected based on centrally available lists of affected events. Additionally, noise rejection filters are used to reject events where instrumentation noise strongly influences the recorded data. Prominent sources include sudden energy spikes in isolated HCAL towers, which are reconstructed as electrically neutral hadronic activity, thus causing fake jets and E_T^{miss}. All of the event filters recommended by the Jet/MET (JME) POG [33, 34] are employed. In addition, the Particle Based Noise Rejection Filter (PBNR) is introduced. It is motivated and discussed in section 3.1.

2.2 Object Selection

In order to enable the final state dependent sorting of events into classes, object identification criteria are defined. Six objects are defined: Muons, electrons, photons, jets, b jets and E_T^{miss}. For each kind of object, dedicated final state independent identification criteria are employed, which are discussed in the following. Additional p_T- and η-requirements are given in tab. 2.3.

2.2.1 Muons

Muon identification is split for low- and high-p_T muons. For $p_T < 200\,\text{GeV}$, the tight working point of the baseline selection [35] is applied, which focuses on track quality criteria. For $p_T \geq 200\,\text{GeV}$, the high-p_T muon selection [36] is applied, which uses a dedicated track reconstruction. Some common criteria are shared between the two categories. The exact criteria are listed in table 2.4.

Table 2.4: Muon ID criteria. *Common* criteria are applied for all muons, while *baseline* and high-p_T criteria are split depending on the muon p_T.

Common	
≥ 1 pixel hit	
≥ 5 tracker layers with hits	
≥ 2 muon stations	
≥ 1 muon chamber hit in global track fit	
$I_{\text{PF}} < 0.12$	
Baseline ($p_T < 200\,\text{GeV}$)	**High-p_T** ($p_T \geq 200\,\text{GeV}$)
PF muon	Dedicated track reconstruction
track fit $\chi^2/\text{ndf} < 10$	$\sigma_{p_T}/p_T < 0.3$
transverse IP $< 2\,\text{mm}$	
longitudinal IP $< 5\,\text{mm}$	

Muon isolation is evaluated via the *PF combined relative isolation* requirement on I_{PF}, which considers energy deposits by PF candidates in a cone of $\Delta R < 0.4$ around the muon.

$$I_{\text{PF}} = \frac{\sum \left(E_T(\text{CHPV}) + E_T(\text{NH}) + E_T(\text{PH}) - 0.5 \cdot E_T(\text{CH}, \not{PV}) \right)}{p_{T,\mu}} \tag{2.1}$$

The deposits from charged hadrons from the primary vertex (CHPV), neutral hadrons (NH) and photons (PH) are summed. Half of the contribution from charged hadrons not associated to the primary vertex (CH,$\not{PV}$) is subtracted as an estimate of the fraction of neutral hadron energy stemming from pile-up. The factor of 0.5 accounts for the ratio of charged and neutral components commonly found in jets.

2.2.2 Electrons

Electrons are identified by the HEEP ID v4.1 criterion [37, 38], which is designed for improved performance with high-p_T electrons, while still performing well at lower transverse momenta. Starting from ECAL driven electron candidates (as opposed to the tracker driven PF reconstruction), requirements on angular distance of track and ECAL supercluster ($|\Delta\eta_{in}|$, $|\Delta\phi_{in}|$) are applied. The shape of the electromagnetic shower is controlled via the H/E and $\sigma_{i\eta i\eta}$ variables. H/E is the ratio ratio of HCAL energy deposits close to electron to the electron's ECAL deposit, while $\sigma_{i\eta i\eta}$ measures the the η-spread of the electron's ECAL energy deposit in units of ECAL crystals. Ratios of the super cluster energy contained in η-ϕ-strips around the seed crystal $E^{n_\eta \times n_\phi}$ are used to select showers that are narrow in η. To account for the influence of pile-up, the average pile-up energy density per unit angular area ρ is measured for each event. With increasing ρ, the isolation criteria are loosened.

Because of instrumentation differences in barrel and endcap, some criteria are modified in the two regions. The barrel-endcap overlap region $1.442 < |\eta| < 1.56$ is excluded.

2.2.3 Photons

Photons are identified using the tight working point of the *Simple Cut Based Photon ID 2012* [39]. The identification criteria are based on the calorimeter shower shape in lateral ($\sigma_{i\eta i\eta}$) and longitudinal (H/E) directions (cf. sec. 2.2.2), isolation from additional photons and neutral and charged hadrons, and a dedicated veto against electrons. The exact criteria are listed in table 2.6. Isolation criteria are evaluated in a cone with radius $R = 0.3$ and ρ-corrected for pile-up, i.e. the average energy density due to pile-up is subtracted before evaluating the isolation. Only photons reconstructed in the barrel region are used.

Table 2.5: Electron ID criteria of the HEEP ID v4.1.

Variable	**Barrel** ($\lvert\eta\rvert < 1.442$)	**Endcap** ($1.56 < \lvert\eta\rvert < 2.5$)
E_T	$> 35\,\text{GeV}$	
Reconstruction is ECAL driven	yes	
$\lvert\Delta\eta_{in}\rvert$	< 0.005	< 0.007
$\lvert\Delta\phi_{in}\rvert$	< 0.06	
H/E	< 0.05	
$\sigma_{i\eta i\eta}$	/	< 0.03
Super cluster energy ratios	$E^{2x5}/E^{5x5} > 0.94$ or $E^{1x5}/E^{5x5} > 0.83$	/
ECAL isolation + Had. Depth-1 Isolation	$< 2\,\text{GeV} + 0.03 \cdot E_T + 0.28\rho$	if $E_T < 50\,\text{GeV}$: $< 2.5\,\text{GeV} + 0.28 \cdot \rho$ else: $< 2.5\,\text{GeV} + 0.03 \cdot (E_T - 50\,\text{GeV}) + 0.28 \cdot \rho$
Track isolation: Track p_T	$< 5\,\text{GeV}$	
Hits lost in inner layer	< 2	
d_{xy}	$< 0.02\,\text{cm}$	$< 0.05\,\text{cm}$

Table 2.6: Photon selection requirements of the *Simple Cut Based Photon ID 2012* [39].

Requirements
Conversion safe electron veto
Single Tower $H/E < 0.05$
$\sigma_{i\eta i\eta} < 0.011$

PF isolation (ρ-corrected)	**Requirement**
Charged Hadron	$< 0.7\,\text{GeV}$
Neutral Hadron	$< 0.4\,\text{GeV} + 0.04 \cdot p_T^\gamma$
Photons	$< 0.5\,\text{GeV} + 0.005 \cdot p_T^\gamma$

2.2.4 Jets

Jets are clustered from PF candidates by the `Anti-`k_t algorithm with a distance parameter of $R = 0.5$ ("AK5PF jets") and identified using the *Particle Flow Tight Jet ID* ("PFTight ID") [40]. The PFTight ID is based on jet composition quantities. In addition to requiring at least two constituents to a jet, thresholds on the fractions of the total jet energy carried by a specific type of PF candidate are imposed. It is required that at least one of the constituents is charged, i.e. has a track associated with it. These requirements are designed to reject jets that are not the result of parton hadronization processes. The requirements are summarized in table 2.7. The relevance of jet energy fractions for noise rejection purposes is discussed in section 3.1. *Charged Hadron Subtraction* [41] is used: By identifying charged hadron tracks stemming from pile-up vertices and subtracting them, jets become more resilient against the effects of pile-up.

The energy reconstruction of jets ("Jet Energy Scale" or JES) is calibrated in four steps [42, 43]:

Table 2.7: Jet selection requirements of the PFTight Jet ID.

Jet Energy Fractions	Requirement
Neutral Hadron	< 0.9
Charged Hadron	> 0.0
Neutral Electromagnetic	> 0.9
Charged Electromagnetic	< 0.9
Muon	< 0.8

Number of constituents	Requirement
Total	> 1
Charged	> 0

- The **Offset Correction** counteracts a constant overestimation of jet energy due to pile-up interactions. The average momentum density per y-ϕ-area due to pile-up is measured. Together with the corresponding area covered by each jet, the additional energy due to pile-up is estimated and subtracted per jet. This step is applied to data and simulation.

- The simulation based **MC Correction** removes most of the p_T- and η-dependence of the jet response. Using a matching between reconstructed and generator level jets, an average correction factor $\left\langle \frac{p_T^{\text{gen}}}{p_T^{\text{rec}}} \right\rangle$ is determined in bins of p_T and η and applied to data and simulation.

- The **Absolute Residual Correction** is only applied to data and corrects the differences in the η-dependence of the jet response compared to simulation. The residual η-dependence is determined from back-to-back dijet events, which are balanced in p_T. Additional recoiling particles may cause one of the two jets to be located in the central $|\eta| < 1.3$ region, while the other jet is located in a more forward region. The difference of reconstructed energies of the two jets is used to calibrate the response in forward $|\eta|$ regions relative to the well-understood central region.

- The **Relative Residual Correction** is also only applied to data and removes any residual p_T-dependence of the jet response. Events with well-reconstructed leptonic decays of Z bosons recoiling against jets are selected. Since no intrinsic E_T^{miss} is present in this process, the measured E_T^{miss} is due to response differences between boson and jet. This allows a calibration of the jet response p_T-linearity relative to the precisely known boson response.

Correction values based on $11\,\text{fb}^{-1}$ of data at $\sqrt{s} = 8\,\text{TeV}$ are used [44, 45]. Calibration uncertainties are of the order of some percent, their treatment is discussed in sec. 2.4.4.3.

2.2.5 B Jets

B jets are identified by the *Combined Secondary Vertex* (CSV) algorithm [46]. A jet is identified as a b jet if the CSV discriminator value is larger than 0.679, which is referred to as the *medium working point* (CSVM) [47]. Reconstruction, calibration and uncertainty handling are performed as for jets. For classification, b jets are treated as a separate object group from jets, i.e. an event with two jets, one of which is a b jet, will enter the 1jet+1b exclusive, but not the 2jet exclusive class. B tagging is discussed in section 2.3.

2.2.6 E_T^{miss}

E_T^{miss} is calculated at the PF candidate level ("PFMet") and no identification criteria are required beside a minimum p_T threshold. E_T^{miss} is corrected for the influence of pile-up by applying *Charged Hadron Subtraction* ("Type-0" correction) (cf. sec. 2.2.4). The effect of the jet energy scale is taken

Table 2.8: Particle collections and criteria used for cleaning. The cleaning is performed by comparing all particles from one collection with those from another on a per-particle basis. If two particles are found to be overlapping, the one with the less reliable signature is removed.

Collections	Overlap criterion	Removal of
μ vs. μ	$\Delta R < 0.4$	μ with lower probability of track fit
e vs. e	$\Delta R < 0.4$ + shared supercluster/track	lower energetic e
e vs. μ	$\Delta R < 0.4$	e
γ vs. γ	$\Delta R < 0.4$ + shared supercluster	lower energetic γ
jet vs. all	$\Delta R < 0.5$	jet

into account ("Type-1" correction) and a ϕ-modulation due to an xy-shift of particle momenta is corrected ("xy-shift" correction). All corrections are detailed in [48].

2.2.7 Object Cleaning

The particle definitions used for reconstruction allow for a single deposit in the detector to be assigned to multiple particles. To avoid double-counting, an overlap removal procedure ("cleaning") is applied. Object collections are compared on a per-object basis and an object type dependent overlap criterion is evaluated. If the criterion is fulfilled, i.e. two objects are found to be overlapping, one of them is removed. The collections cleaned against each other and the respective criteria are listed in table 2.8. Jets are cleaned against all other collections because the `Anti-`k_t algorithm will reconstruct a jet around any sufficiently isolated PF candidate, thus introducing overlap with any other kind of object.

2.3 B Tagging

2.3.1 Motivation

Jets do not trivially contain information about the kind of parton they were initiated by. Thus, processes that look very different in terms of Feynman diagrams may still result in indistinguishable final states in terms of jets. To counteract this loss of information, *tagging* procedures are introduced. The aim of tagging is to exploit knowledge about hadronisation in order to make an educated guess about the flavor of a jet.

An established tagging technique is b tagging, i.e. the identification of bottom flavor jets (b jets). B jets are relevant in many contexts. For example, top quark decays almost always result in b quark production, which makes b tagging an important tool to select events with top quarks. The relatively high mass of the bottom quark also makes it a decay candidate in searches for any BSM process with Yukawa-like couplings (e.g. Higgs $\rightarrow$ bb).

2.3.2 Method

B tagging relies on secondary vertices (SV) and track impact parameters (IP)[3]. Because hadronisation is governed by the strong interaction, quark flavor numbers are conserved. This means that if there is a single b-quark leaving the hard interaction, there will be a baryon containing b quarks in the resulting shower. With his knowledge, the problem of identifying b jets reduces to the problem of identifying baryons with b quarks in hadron showers. B mesons have characteristic lifetimes of about $c\tau \approx 0.5\,\text{mm}$ [4], which is below the geometric radius of the tracker and resolvable by the track reconstruction. Thus, B mesons decay inside the tracker, giving rise to additional tracks. Because these tracks do not stem from the primary vertex, they have a finite impact parameter and form a secondary vertex.

[3]The impact parameter is defined as the distance of closest approach between track and primary vertex, see fig. 2.3.

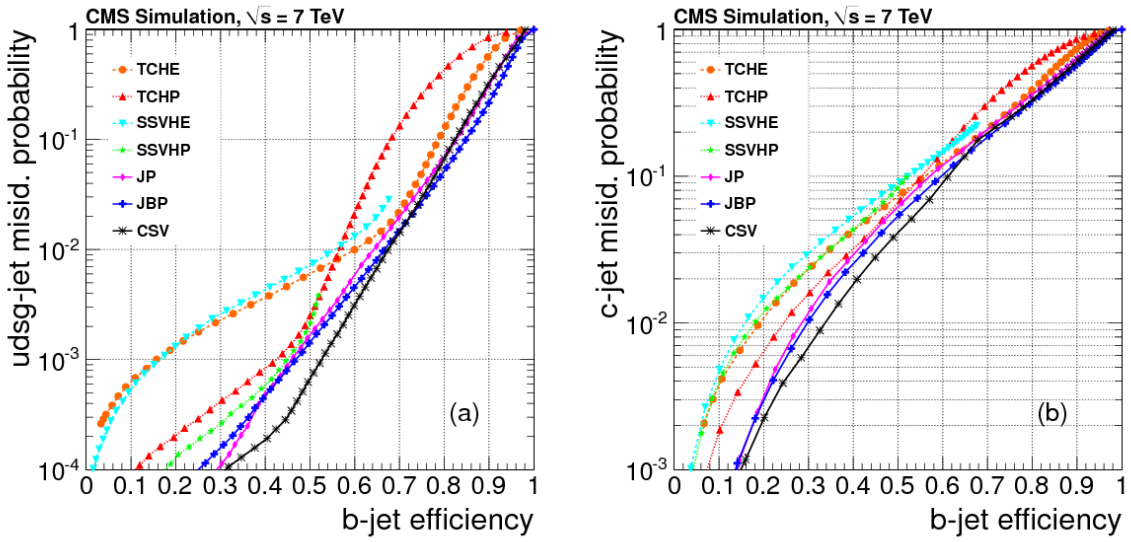


Figure 2.2: b tagging mis-identification probabilities for (a) light and (b) charm quark jets as a function of the tagging efficiency for different tagging algorithms [50]. The curves are obtained from jets with $p_T > 60\,\text{GeV}$ in simulated multijet events. The CSV algorithm shows the most favorable behavior, i.e. the best efficiency-to-mistag ratio of all taggers in the relevant efficiency region around $\approx 70\,\%$.

Vertex information is however not conclusive. Since particle decays are probabilistic processes, there is a probability distribution of the decay times of the B mesons. This distribution overlaps with those of other mesons. Especially the charmed D mesons with lifetimes of the order $c\tau \approx 100\,\mu\text{m}$ may result in similar SV and IP configurations as B mesons, while other mesons such as K ($c\tau \approx 1\,\text{cm} - 10\,\text{m}$) and $\pi^{\pm}$ ($\pi^{\pm}$: $c\tau \approx 8\,\text{m}$) do not yield measurable IP/SV information [4].

Secondary vertex and IP information is combined with other kinematic properties of the jet, such as vertex mass[4], and used as input to different available b tagging algorithms. The tagging algorithms map the input quantities to a single output number, which may then be used to define an identification threshold. Depending on how the threshold in chosen, tagging efficiencies and mis-tagging rates vary, where in general higher efficiencies go along with increased mis-tagging rates. Intuitively, high efficiencies and low mis-tagging rates are favorable.
In this analysis, the Combined Secondary Vertex (CSV) algorithm [46] is used, which extensive study in data and simulation has shown to be effective [49–51] (fig. 2.2). The medium working point (CSVM) of the CSV algorithm is used (cf. sec. 2.2.5).

2.3.3 B Tagging Scale Factors

The SM prediction in this analysis is based on simulation and is it necessary to ensure that the tagging criterion performs equally well in data and simulation. To that effect, tagging rate scale factors are used. The scale factors are applied in two steps:

1. FastSim tagging rates are corrected to match FullSim. The determination of the scale factors for this step is discussed below.

2. All simulation samples (FullSim and corrected FastSim) are corrected to match the tagging rates in data. The scale factors for this step are taken from a dedicated CMS analysis [51, 53], which measures the b tagging efficiency and mis-tagging rates. The tagging efficiency is determined from template fits of simulated distributions for each jet flavor to measured distributions. Variables are chosen that are known to be distributed differently for different jets flavors to

[4]The vertex mass is defined as the combined invariant mass of all tracks associated to the vertex.

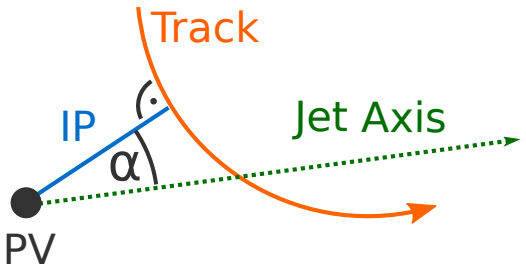


Figure 2.3: Illustration of the impact parameter (IP) and its sign. The IP is the distance of closest approach between a track and the primary vertex (PV). It has a positive (negative) sign for $\alpha < \pi/2$ ($\alpha > \pi/2$). Adapted from [52].

enhance sensitivity. The measurement of mis-tagging rates makes use of the symmetry of the mis-tagging rate with respect to the IP sign (cf. sec. 2.3.2). By comparing the tagging rates for jets with positive and negative IPs, the mis-tagging rates can be inferred.

The used method for applying scale factors is discussed in sec. 2.3.3.1, the determination of tagging rates and scale factors in simulation is described in sec. 2.3.3.2.

2.3.3.1 Applying Scale Factors

Scale factors are applied using a method recommended by the b tagging *Physics Object Group*[5] [54]. In order to correct the tagging rate ϵ_i in a sample i to be equal to the tagging rate ϵ_j in another sample j, a scale factor $SF = \epsilon_i/\epsilon_j$ is used. If $SF < 1$ (the tagging rate in sample i is too high), a fraction of the jets tagged in i by is artificially de-tagged.

$$f_{SF<1} = \frac{\text{Number jets to de-tag}}{\text{Number of tagged jets}} = \frac{\epsilon_i - \epsilon_j}{\epsilon_i} = 1 - SF \tag{2.2}$$

If $SF > 1$ (the tagging rate in i is too low), a fraction of jets not tagged in i are additionally tagged.

$$f_{SF>1} = \frac{\text{Number jets to tag additionally}}{\text{Number of untagged jets}} = \frac{\epsilon_j - \epsilon_i}{1 - \epsilon_i} = \frac{1 - SF}{1 - 1/\epsilon_i} \tag{2.3}$$

The dependence of $f_{SF>1}$ on ϵ_i cannot be completely replaced by the SF variable. It is therefore necessary to know not only the scale factors SF, but also the tagging rate ϵ_i of the sample that is to be corrected, i.e. the FastSim and FullSim simulation samples. The same method is used for the b tagging efficiency as well as the mis-tagging rates for other flavors. The tagging correction is performed stochastically for each jet by generating a uniformly distributed random number $0 < x < 1$ and triggering the appropriate correction if $x < f$.

2.3.3.2 Determining Rates and Scale Factors in Simulation

In simulation, the tagging rates may easily be obtained from generator information. The generator level particles (after parton showering and hadronisation) are clustered into jets in the same way as reconstructed particles. A matching is performed between generator level jets and reconstructed jets based on their ΔR separation[6]. Reconstructed jets are assigned to one of three categories ("flavors") according to the parton that initiated the generator jet[7]: **l** (light) for up, down and strange quarks as well as gluons; **c** for charm quarks and **b** for bottom quarks. Using the number N_i^{tag} of jets of flavor

[5] *Physics Object Groups* (POGs) are groups of analysts performing dedicated studies of particle identification methods in CMS. They provide recommendations for use by other analysts.

[6] Approximately $1 - 2\,\%$ of reconstructed jets are found to not have a matching generator level jet. These jets are disregarded for the tagging rate determination.

[7] The *algorithmic* flavor definition is used, which also considers a jet to be of bottom or charm flavor if it contains bottom or charm quarks from gluon splitting [55].

i tagged as b jets by the tagging criterion and N_i, the number of all jets of flavor i, the tagging rate can be calculated.

$$\epsilon_{\mathrm{i}} = \frac{N_{\mathrm{tag,i}}}{N_{\mathrm{all,i}}} \tag{2.4}$$

This method is applied for the CSVM criterion (cf. sec. 2.2.5) in bins of p_T of the reconstructed jets and separately for FullSim and FastSim samples. The resulting tagging rates are shown in fig. 2.4, where the same simulation samples are used as in the main analysis (cf. sec. 2.4.1). Also shown are the scale factors used to correct tagging rates in FastSim samples to the rates determined for FullSim samples, which are calculated by dividing the FullSim tagging rate by the FastSim rate in each p_T bin. After applying both correction steps (FastSim to FullSim and simulation to data), tagging rates in FastSim and FullSim agree well, indicating that the correction method is effective.

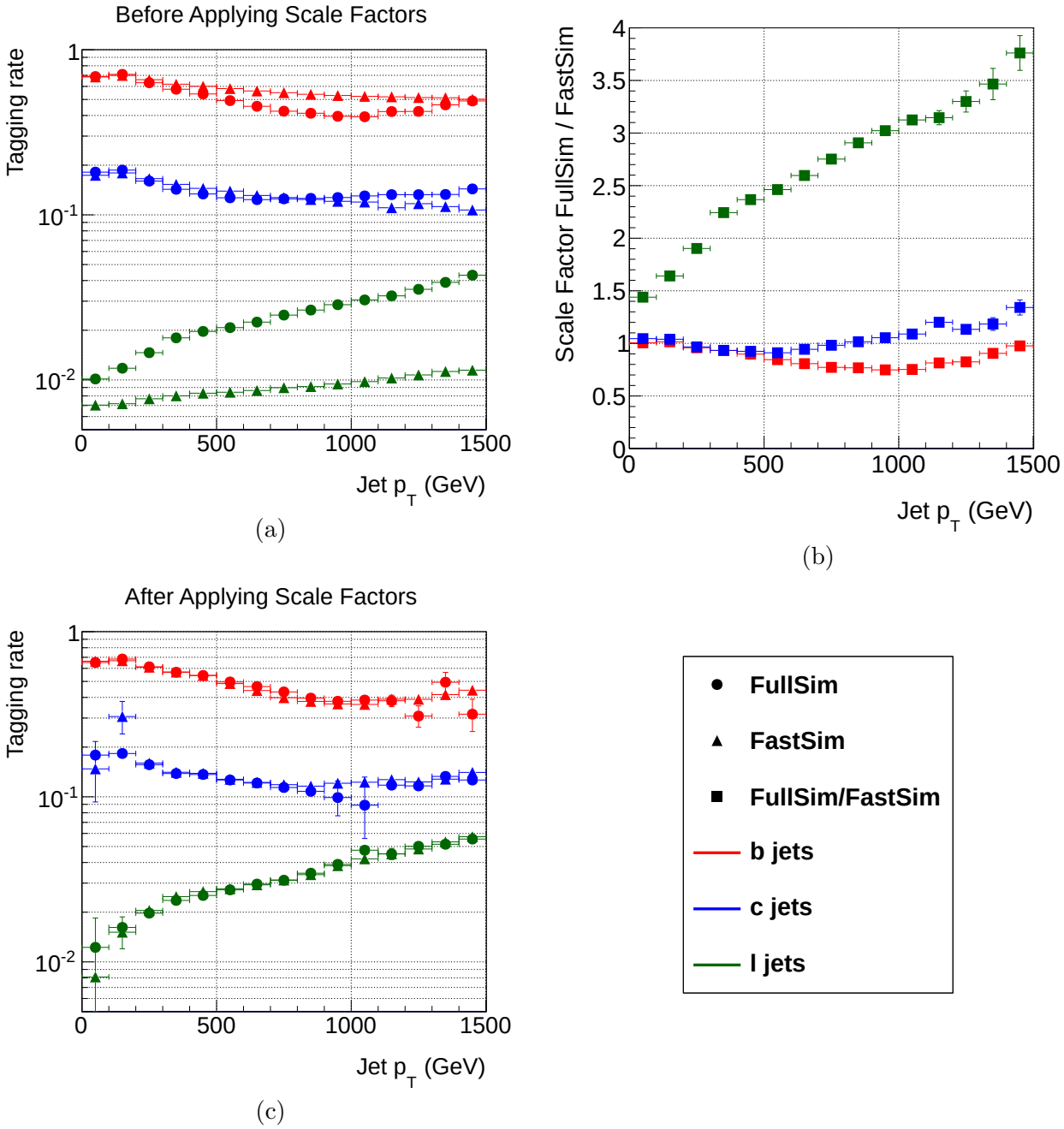


Figure 2.4: Uncorrected CSVM tagging rates in FullSim and FastSim (a) for bottom/b, charm/c and light/l flavor jets. Scale factors (b) are derived by dividing the rate in FullSim by that in FastSim in each bin. After application of both FastSim to FullSim and simulation to data scale factors, the tagging rates in FullSim and FastSim are found to agree (c), demonstrating the effectiveness of the method. Jet flavor is assigned based on generator information. The rightmost bins include overflow, uncertainties are statistical only.

2.4 Standard Model Simulation

MUSiC aims at a full view of the CMS data and a complete simulation of Standard Model physics is necessary. Fig. 2.5 shows the results of CMS cross-section measurements of many of the SM processes relevant to MUSiC with the exception of QCD jet production. In this subset alone, cross-sections have been precisely measured and found to agree to theoretical predictions over a range of eight orders of magnitude[8]. The processes can be roughly categorised in three groups: Single- and multi-boson production, top quark production and processes involving Higgs bosons. For each of the boson and top quark processes, simulation samples are used. Higgs boson related processes are not included at this time. The exact composition of the list of used samples is discussed in the following.

2.4.1 Monte Carlo Samples

Production of Monte Carlo simulation samples is generally performed in four steps:

1. **Event Generation:** Event generation is the process of producing parton-level interaction events. This is a description of Feynman-diagram-like processes with incoming partons and outgoing partons and additional stable or unstable particles. In this analysis, samples produced in LO with the event generators Pythia6 [58] and Madgraph [59] and in NLO with Powheg [60–62] are used[9]. Where available, k-factors are used to scale the process cross-section to higher perturbative orders. The momentum distribution inside the proton is modeled using

[8] Including the inclusive jet production cross-section [57] extends this range to just over nine order of magnitude.

[9] For simplicity, samples are referred to by their generator, i.e. "Madgraph sample" means "a sample produced by event generator Madgraph, for which parton showering and hadronisation were performed by Pythia6 ".

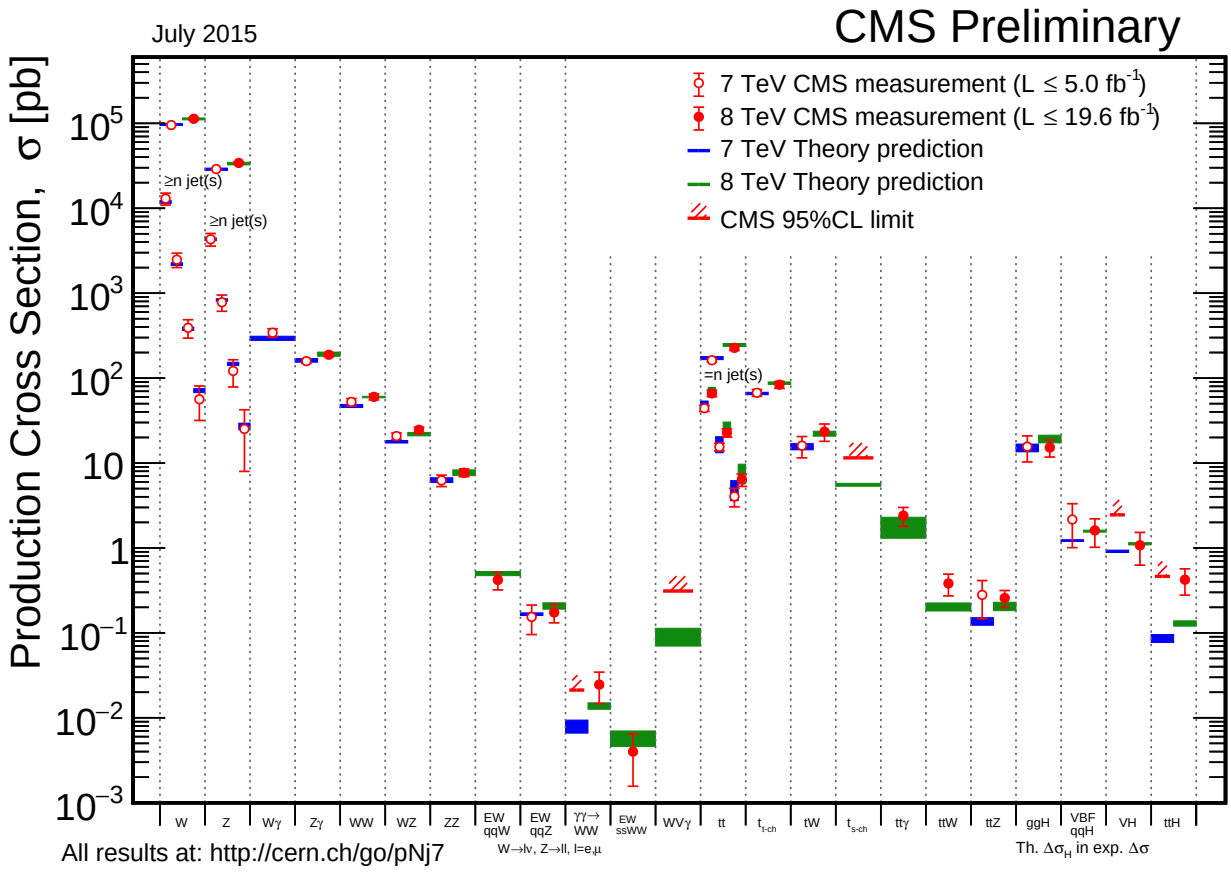


Figure 2.5: Overview of SM production cross-section measurements performed by CMS [56].

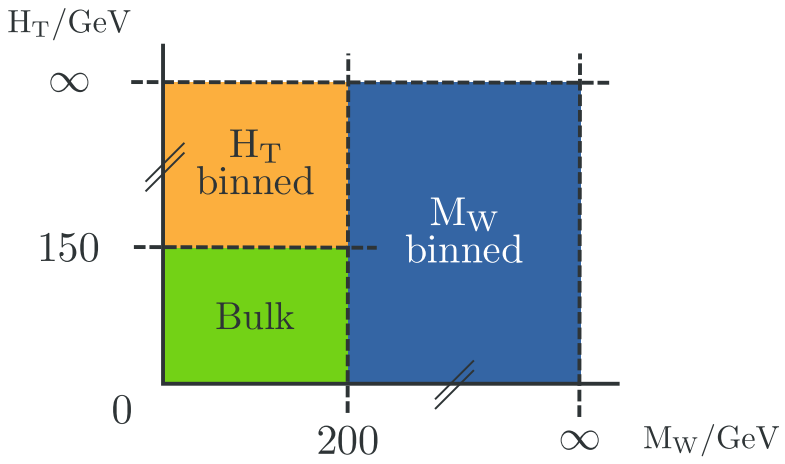


Figure 2.6: Composition of the W+jets SM simulation. An unbinned *bulk* sample is relied upon for relatively on-shell bosons and moderate jet p_T-values. Phase space tails are modeled using samples binned in M_W and H_T.

the LO CTEQ6L1 PDF set [63] in PYTHIA6 and MADGRAPH samples and the NLO CT10 set [64] in POWHEG samples.

2. **Parton showering:** Incoming and outgoing partons from the primary interaction may radiate gluons, which may in turn radiate further gluons or split into quark-antiquark pairs, giving rise to parton showers. Parton showering is modeled by PYTHIA6. It is formulated as an evolution process from the high energy scale of the primary interaction to low energy scales, at which non-perturbative effects set in [65].

3. **Hadronisation:** Partons resulting from showering form hadrons. This process is governed by non-perturbative QCD and modeling is entirely empirical [65]. The hadronisation process is simulated by PYTHIA6, which uses the Lund string model [66]. In this model, partons are bound by a force linearly increasing with distance. Once the potential energy reaches a critical point, production of a quark-antiquark pair is energetically favorable. This process repeats until only color neutral hadrons are left, which may subsequently decay and give rise to additional leptons and photons.

4. **Detector simulation:** Modern particle detectors such as CMS are complicated machines and a simulation of the interaction of the physical particles with the detector material is necessary. Unless otherwise stated, the detector simulation is performed by GEANT [67], which is an object-oriented framework for the simulation of particle-matter interactions ("Full simulation" or "FULLSIM "). For some privately produced samples a fast alternative to GEANT is used ("Fast simulation" or "FASTSIM ", cf. sec 2.4.2).

The used simulation samples used are briefly summarised in the following. A concise overview is given in table 2.9, the unabridged list of samples including detailed cross-section and production information may be found in appendix A.

2.4.1.1 Boson Production

Bosonic processes give large contributions to many event classes containing leptons. Single, double and triple boson processes are considered.

- The **Drell-Yan** samples model the production of leptons mediated by on- and off-shell Z bosons and photons. Different samples are available depending on the final state leptons. The bulk of the distributions is covered by M_{ll}- and $p_{T,Z}$-binned MADGRAPH samples. To enhance the simulation statistics for very high M_{ll}, additional PYTHIA6 samples are used for $M_{ll} > 600\,\text{GeV}$.

Although not strictly considered a Drell-Yan process, Z mediated neutrino production is also grouped in this section. The process is modeled by MADGRAPH samples binned in $p_{T,Z}$.

- **W boson production** is especially important because it gives sizable contributions to a large range of final states with jets, leptons and with or without E_T^{miss}. This makes W production relevant for many dedicated analyses, resulting in a large number of available simulation samples from generators POWHEG, PYTHIA6 and MADGRAPH, all with different binnings and thus different but overlapping phase space volumes. There is no straight-forward method for combining these samples to form a comprehensive description of the full SM. In the work of [25], a combination of three LO MADGRAPH samples has been found to perform well: An unbinned *bulk* sample, a H_T-binned[10] sample and a privately produced M_W-binned FASTSIM sample. The sample composition in the H_T-M_W plane is shown in fig. 2.6. Kinematically constant NNLO cross-section k-factors calculated by `FEWZ` [68, 69] are used.

- **Photon production** is different from other bosonic processes because photons are stable. Photon processes thus contribute to fewer classes than the diversely decaying W and Z bosons. However, associated production of photons and jets has a large total cross-section of $\mathcal{O}(\mu b)$, causing significant contributions in hadronically triggered classes. Single and double photon production are modeled by PYTHIA6 samples binned in the photon p_T in the center-of-mass frame of the interaction. The single photon cross-sections are scaled to NLO using k factors determined in the Higgs boson search effort [70].

- **Multi-boson production** involving Z and W bosons can - due to the multiple decay modes of each boson and the large number of combinations thereof - contribute to virtually any final state. The cross-sections span a range of $\mathcal{O}(0.1-10\,\text{pb})$. Most of the samples in this category are produced using MADGRAPH, with a smaller number of individual samples produced by POWHEG or PYTHIA6. For roughly half of the processes, NLO k-factors are available.

2.4.1.2 Top Quark Production

At the LHC, top quarks are predominantly produced in pairs via the strong interaction with additional smaller contributions of weakly produced single top quarks. Both single and double top quark production are modeled using POWHEG. NNLO k-factors are available for all top quark pair production samples [71, 72], whereas some single top quark sample cross-sections are only known at LO [68, 73]. Samples for top quark production with associated bosons and the rare quadruple top quark production process and are provided by MADGRAPH and in part scaled to NLO [68, 74, 75]. The cross-sections vary from $\mathcal{O}(1\,\text{fb})$ for quadruple top up to $\mathcal{O}(100\,\text{pb})$ for top pair production. Top quarks almost exclusively decay to bottom quarks and W bosons, causing a large number of possible final states often involving b jets.

2.4.1.3 Multijet Production in QCD

QCD multijet production is modeled by a total of six samples, all of which are produced in LO using PYTHIA6. The samples are binned in the p_T of the leading generated jet and each cover the full p_T-range of this analysis. Jets are required to carry a minimum transverse energy $E_T > 7\,\text{GeV}$. In addition to the bulk sample, enriched samples are used, which employ a filtering mechanism on generator level to enhance certain sub-processes of jet production.

- The **BCtoE** filter selects events with jets containing electrons produced in the decays of bottom and charmed mesons.

- The **EM** filter selects events with isolated photons, electrons and charged kaons and pions.

- The **Mu5(15)** filters select events containing muons with a p_T of at least 5(15) GeV.

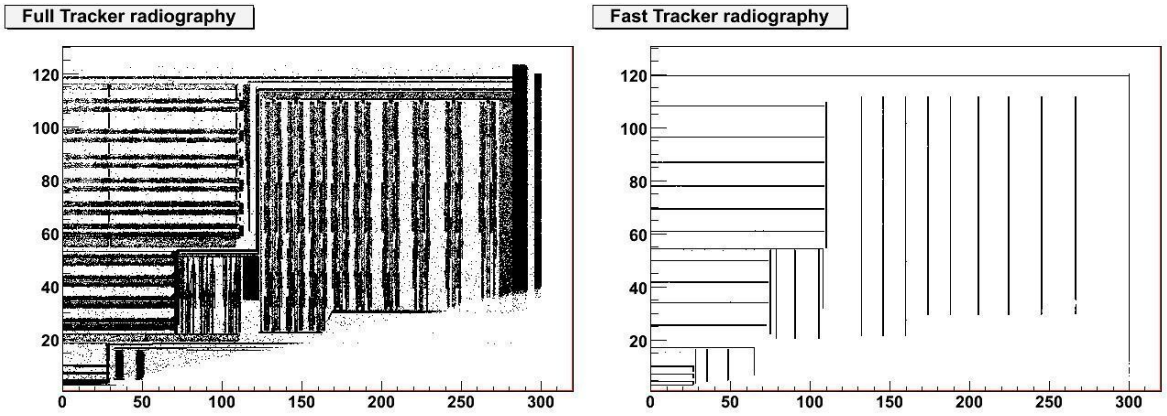
[10] H_T is the scalar sum of jet transverse momenta.

Table 2.9: Summary of processes in SM simulation. For each process generator, binning variable, number of events and cross-section are given. The generators are MADGRAPH (MG), PYTHIA6 (P6) and POWHEG (PH). Parton showering and hadronisation are always performed by PYTHIA6. The symbol V is used to refer to both W and Z bosons [11]. For more information about the processes, please refer to the text. A full list of used samples is presented in appendix A.

Process	**Generator**	**Binning**	**# Events** (10^6)
Diphoton	P6	p_T	10
Photon + jets	P6	p_T	3
Drell-Yan $\rightarrow ll$	MG	M	60
Drell-Yan $\rightarrow ll$	MG	$p_{T,Z}$	38
Drell-Yan $\rightarrow \mu\mu$	P6	M_{ll}	2
Drell-Yan $\rightarrow ee$	P6	M_{ll}	0.5
Z+jets $\rightarrow \nu\nu$	MG	$p_{T,Z}$	30
W+jets	MG	/	76.1
W+jets	MG	H_T	53.1
W+jets	MG	M_{W}	1.1
QCD	P6	p_T	90
QCD BCToE	P6	p_T	8
QCD EM	P6	p_T	180
QCD Mu5	P6	p_T	70
QCD Mu15	P6	p_T	30
QCD B	P6	p_T	35
Single Top (s-channel)	PH	/	0.3
Single Top (t-channel)	PH	/	4
Single Top (assoc. W)	PH	/	0.8
TBZToLL	MG	/	0.1
$t\bar{t}$	PH	/	6.2
$t\bar{t}$	PH	$M_{t\bar{t}}$	4.0
$t\bar{t}+\gamma$	MG	/	1.0
$t\bar{t}$W	MG	/	0.1
$t\bar{t}$WW	MG	/	0.2
$t\bar{t}$Z	MG	/	0.2
$t\bar{t}t\bar{t}$	MG	/	0.1
WW	P6	/	10.0
ZZ	PH	/	15.4
ZZ	MG	/	2.7
WZ	P6	/	7.9
$gg \rightarrow$ VV	P6	/	1.0
Vγ	MG	/	10.0
VVV	MG	/	0.8
W$\gamma\gamma$	MG	/	2.0

- The **B** filter selects events containing a generator level b-quark.

A privately produced b-enriched sample is used. Its production and properties are described in sec. 2.4.2.1. Since all of the enhanced processes also occur in the bulk as well as the other enriched samples, an inverse filtering is performed on analysis level. In each sample, all events passing one of the dedicated filters used in any of the other samples are discarded.



(a) (b)

Figure 2.7: Radiography of the tracker geometry used in (a) FullSim and (b) FastSim. While FullSim realistically simulates the tracker layers as volumes of finite thickness, FastSim approximates them as thin layers and neglects passive material [76].

2.4.2 FastSim

FastSim is a scheme for the time-efficient simulation of the interaction of particles with the CMS detector using a simplified detector geometry and parametrisation of material interactions [76]. It is a fast alternative to the Geant-based FullSim procedure, which simulates low-level material effects for all sub-detectors, cable ducts and dead material. The reconstruction sequence remains as it is performed for data and FullSim with the important exception of track finding. A full description of the FastSim procedure and its differences to FullSim is given in ref. [76]. Here, an instructive example of those differences is given: the simulation of the tracking system.

The CMS silicon tracking systems consists of active material (a large number of plates of a certain thickness), and some passive materials (such as cables). In FullSim, this complex structure is included. Fig. 2.7 (a) shows a "radiography" of the simulated tracker, which means that each black dot in the image corresponds to the origin of a photon from Bremsstrahlung. Since Bremsstrahlung only occurs when a particle traverses a medium, the point density reflects the simulated material budget. In this view, the "fuzziness" of the plates clearly indicates their inhomogeneities and spatial extent. In contrast, fig. 2.7 (b) shows an analogous image from FastSim. Here, all of the "fuzziness" is removed. The plates are represented by thin lines of material without any inhomogeneities, particle hits are not simulated as complex interactions, but simply placed where simulated tracks intersect the thin material layers. No energy deposit and measured charge information in the layers is available and the passive material budget is neglected.

Once all hits are simulated, reconstruction is performed. In FullSim, the usual tracking algorithms are run, which attempt to fit particle tracks to match the recorded hits. Because the number of hits in an event is high on average, this process takes a lot of computation time, which is one of the major contributions to the FullSim runtime. In FastSim, the tracking algorithms are modified to use generator-level information about the particle tracks as seeds. The algorithm simply collects all tracker hits that lie reasonably close to the extrapolated track of a simulated particle. Large amounts of computation time are saved because the algorithm never explores the large numbers of combinatoric combinations of hits. However, it also prevents the reconstruction of "fake" tracks, i.e. tracks that arise due to combinatorics rather than physics. While the methods used to speed up FastSim might seem unorthodox and overly simple, studies of its performance have found it to give reasonable results [76]. Significant deviations in the performance of FastSim and FullSim are observed in tracking-related phenomena such as b tagging. This necessitates the use of additional scale factors to correct the FastSim result (cf. 2.3.3).

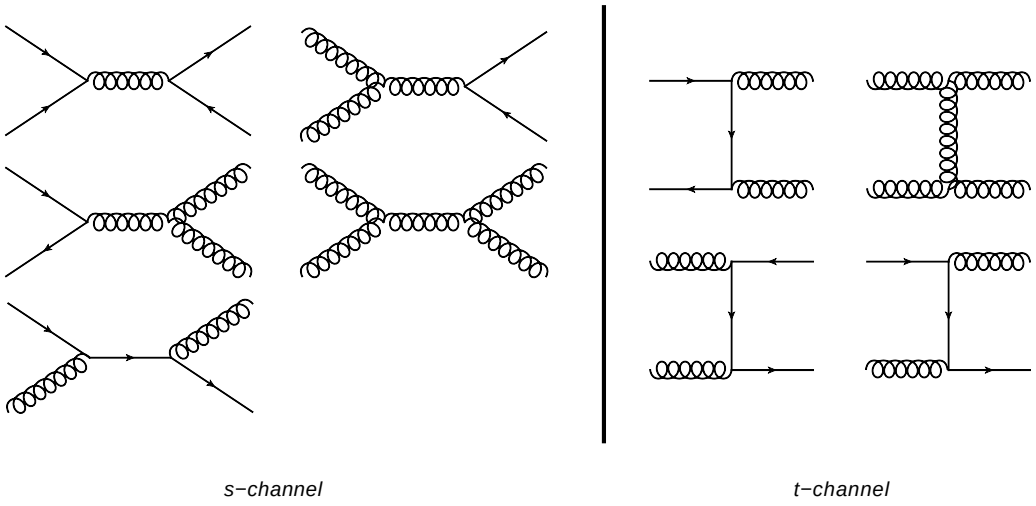


Figure 2.8: Feynman diagrams contributing to QCD jet production as simulated in Pythia6. Fermion lines represent quarks with the exception of top quarks. All leading order scattering processes with two initial and two final state partons are calculated in the s- and t-channels. The graphs are identical for the enriched and non-enriched samples. Enrichment is accomplished via filtering after the generation step.

2.4.2.1 QCD b-Enriched FastSim sample

The centrally produced QCD b enriched sample (a sample where the production of a generator level b quark is required, cf. sec. 2.4.1.3) is mainly used by analysts interested in b quark physics phenomena, which take place in the transverse momentum range of $\mathcal{O}(10\,\text{GeV})$. Thus, more than 95% of the events in the sample are produced requiring $p_T < 150\,\text{GeV}$ for the leading jet in the event, which results in exactly zero of these events passing the selection of this analysis (cf. sec. 2.1.2). Even for the "high-p_T" bin starting at 150 GeV, only a minuscule fraction in the tail of this sample will be selected. This lack of simulated events yields a poor modeling quality for classes dominated by direct QCD production of b quarks, such as the 2b exclusive class (fig. 2.9 (f)). Instead of a continuous distribution, only a small number of spikes is present in this class. No shape information can be deduced and a meaningful comparison between data and simulation is not possible.

To allow for an assessment of these b quark dominated classes, an additional b enriched QCD sample is produced using FastSim. The event generation is performed using the Pythia6 interface to CMSSW version 5_3_14, which also performs the FastSim simulation and reconstruction. For the event generation, the Pythia6 `Z2star`-tune [77] is used. The used configuration file is given in appendix B. All Pythia6 settings are chosen to be identical to the settings used in the central production with the exception of the kinematic filtering configuration. As in central production, the sample is produced in bins of the p_T of the leading jet in the event. In addition to the exclusive bins with upper limits 30, 50 and 150 GeV and the inclusive bin with lower limit 150 GeV, inclusive bins with lower edges 300, 500 and 700 GeV are introduced. Each binned sample is "cut off" at the threshold of the following sample on analysis level. While the original central CMS production reported a computation time of ≈ 150s per event, the FastSim production requires ≈ 15s per event. After filtering, about 35×10^6 events are present in the additional bins. A filtering efficiency of about 12 % is observed, which is similar to the centrally produced sample.

Validation is performed to ensure that the newly produced sample reasonably agrees with the FullSim sample. Without applying final state requirements, the distributions of p_T, η and ϕ of single jets and H_T, the scalar sum of transverse momenta of all jets in an event, are determined (fig. 2.9 (a-d)). For these spectra, very good agreement is observed over the ranges where the FullSim sample has sufficient event statistics. In the p_T and H_T spectra, the increased energetic reach of

the FastSim sample is evident: While the FullSim sample statistics end in the region $p_T = 1\,\text{TeV}$ ($H_T = 2.0 - 2.5\,\text{TeV}$), the FastSim sample extends up to 2.5 TeV (5 TeV), doubling the accessible energy range. The importance of this increase is easily motivated by considering the event counts, i.e. the histogram y-axis. At the end of the statistics reach of the FullSim sample, the simulation prediction is $\mathcal{O}(100\,\text{events}/50\text{GeV})$, which means hundreds of events in and above this energy scale are expected, with no simulated events to compare to. The statistics reach of the FastSim sample, on the other hand, extends to expected event rates $\mathcal{O}(0.1\,\text{events}/50\text{GeV})$. Thus, the sample statistics only become limiting in a tail region where at most single events are expected per bin.
Although both samples use events generated by identical configurations of Pythia6, a significant difference in the shape of the distribution of the CSV discriminator can be observed (fig. 2.9 (e)). This difference is expected due to the tracking differences discussed above and corrected for using scale factors (cf. section 2.3.3).

An inherent caveat of the validation process is limited number of available FullSim events: A comparison is only possible where FullSim makes meaningful predictions, i.e. regions where the additional FastSim sample would not have been needed in the first place. This is impressively illustrated in the distribution of H_T with a typical MUSiC final state selection applied: Fig. 2.9 (f) shows the H_T distribution in the 2b excl. final state. Here, no final state dependent validation is possible, because the FullSim sample does not provide any shape information.

2.4.3 Monte Carlo Weighting

Simulated events are weighted to account for multiple effects. A correct normalisation according to sample cross-section and recorded integrated luminosity in data is enforced by introducing a normalisation weight w_{norm} for each simulation sample. Estimates of the influence of higher-order perturbative corrections to process cross-sections are implemented via flat, i.e. kinematically constant, k-factors, which are not available for all samples.

$$w_{\text{norm}} = \frac{N_{\text{MC}}}{k \times \sigma \times \mathcal{L}_{\text{int}}}$$

To account for differences in the distribution $P(N_{vertex})$ of the number of vertices per bunch crossing in data and simulation, a pile-up weight w_{PU} is introduced. The applied weight exclusively depends on the number of vertices in the event and fixes the distribution $P(N_{vertex})$ in simulation to that in data. This recommended approach is documented in [78, 79].

$$w_{\text{PU}}(N_{\text{vertex}}) = \frac{P_{\text{Data}}(N_{\text{vertex}})}{P_{\text{MC}}(N_{\text{vertex}})}$$

Depending on the generator and settings chosen for the production of a sample, generator weights w_{Gen} are available. Generator weights are used to implement differences in cross-section between different matrix elements used in the same generation step.
The total weight applied to an event is the product of the individual weights.

$$w_{\text{total}} = w_{\text{norm}} \times w_{\text{PU}} \times w_{Gen} \tag{2.5}$$

2.4.4 Systematic Uncertainties

Systematic uncertainties are treated in the handling of simulation samples.

2.4.4.1 Simulation normalisation

The normalisation for each process depends on the calculated cross-section and the recorded integrated luminosity. Cross-sections are calculated perturbatively and are thus subject to higher-order corrections, the effect of which is estimated using an overall relative uncertainty depending on the perturbative order of the calculation. The large LO uncertainty is mainly relevant for QCD samples and motivated in sec. 4.5.3.

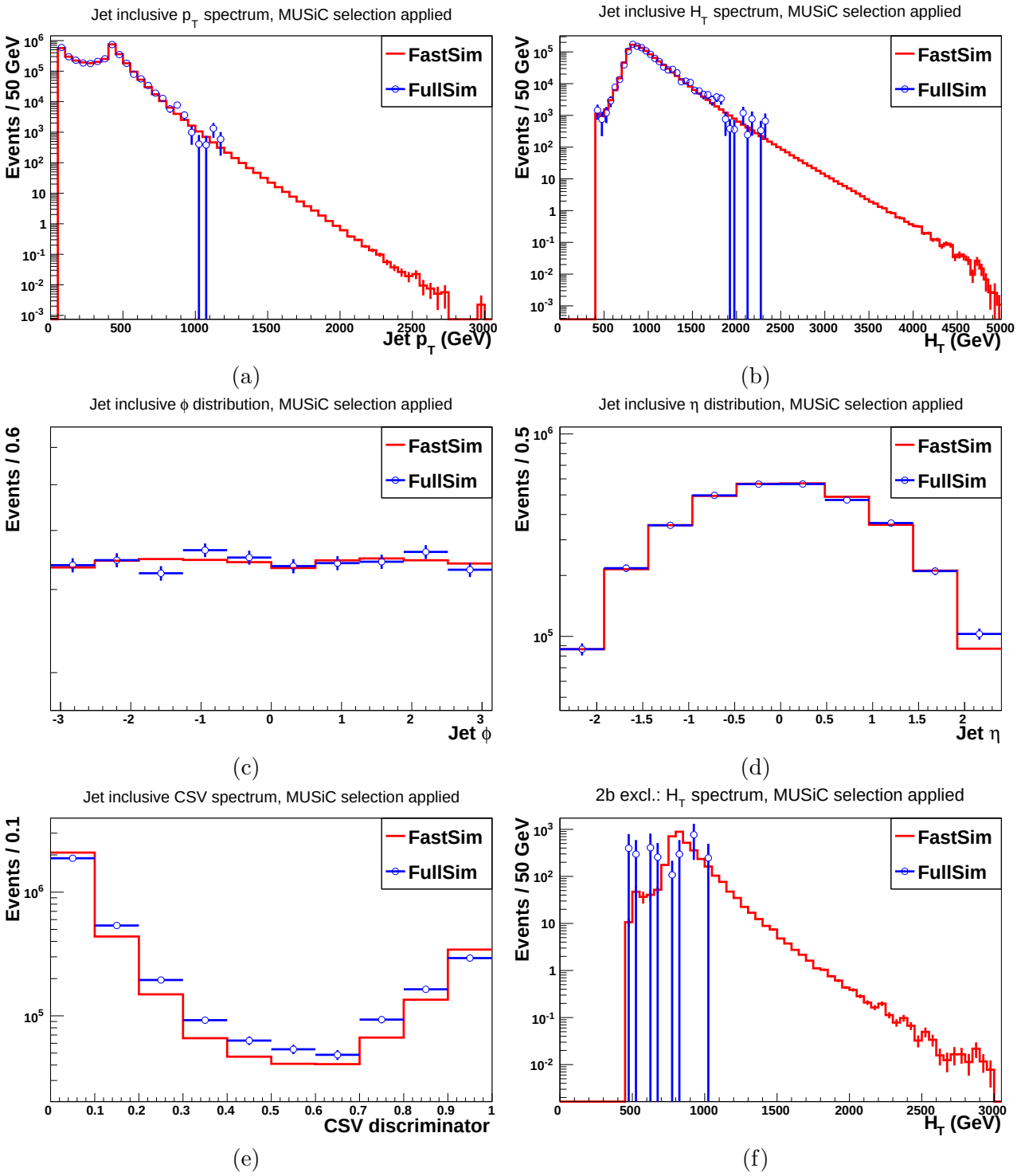


Figure 2.9: Comparison of the b enriched QCD samples in FullSim and FastSim in jet triggered events. Jet p_T (a), event H_T (b), jet ϕ (c) and η (d) distributions show good agreement. The p_T and H_T variables show the increased energetic reach of the FastSim sample. The CSV output discriminator (e) shows a shape difference due to the differences in tracking simulation. In (a-e), no difference is made between jets and b jets. When a MUSiC final state selection like in the 2b exclusive class is applied (f), validation is prevented by the limited FullSim statistics.

Perturbative order	Cross-section uncertainty
LO	50 %
NLO	10 %
NNLL	10 %
NNLO	5 %

The integrated luminosity is determined experimentally and its uncertainty is estimated to be 2.5 % [9], which directly translates to an identical uncertainty on the simulation normalisation.

2.4.4.2 Efficiencies and Fake Rates

In principle, the exact values of identification efficiencies and mis-identification rates are not of primary importance for this analysis as long as they agree in data and simulation. To account for discrepancies, flat relative systematic uncertainties on event weights are introduced. For electrons, muons and photons, the differences between data and simulation are small and no corrections are applied. A normalisation uncertainty is conservatively estimated from the efficiency differences found in dedicated analyses.

For b jets, efficiency differences can be substantial, especially for the case of FastSim, which makes corrections necessary (cf. sec. 2.3.3). The uncertainty of the efficiency scale factor $SF = \epsilon_{\text{data}}/\epsilon_{\text{simulation}}$ is used.

Object	Efficiency uncertainty	Reference
Muon	1 %	[80]
Electron	2 %	[81]
Photon	3 %	[39]
B Jet	5 %	[51]

The uncertainties are added in a fully correlated way for objects of the same type and treated as uncorrelated for objects of different types.

2.4.4.3 Jet Energy Scale

The jet energy calibration is performed experimentally (cf. sec. 2.2.4) and thus influenced by multiple sources of uncertainty, which are provided by the JetMet POG [33, 44, 45]. The uncertainty is nearly uniform in η within the tracker coverage range $|\eta| < 2.5$ and generally decreases with p_T, ranging from 3 % at 30 GeV to 1 % at 1 TeV.
The classification process is performed once each for the JES central value and upper and lower confidence interval bounds. The shifted distributions of each kinematic variable in each event class are subtracted and the difference is used as a systematic uncertainty.

2.4.4.4 Parton Distribution Functions

The PDFs are determined experimentally by fitting a model functions to special collider observables. For each fit parameter, a confidence interval is obtained. The PDFs are made available not only for the parameter central values but also with the parameters shifted up and down within their uncertainties. This yields a larger number of PDFs (a PDF *set*). In principle, the uncertainty could be propagated by performing the full MC production for each of the members of a PDF set, which is however, prohibitively time-consuming in practice. Instead, the *PDF4LHC* [82, 83] reweighting scheme is used. Similar to the way the JES uncertainty is treated, the classification process is performed not only for the best fit PDF, but also the shifted PDFs. The influence of shifted PDFs is implemented via an additional event weight that depends on generator information about the incoming partons: the parton flavor, the fraction x of proton momentum carried by the parton and the scale Q of the interaction. A kinematic distribution is obtained for each of the members of the PDF set. The PDF uncertainty is then obtained by subtracting the shifted distributions.

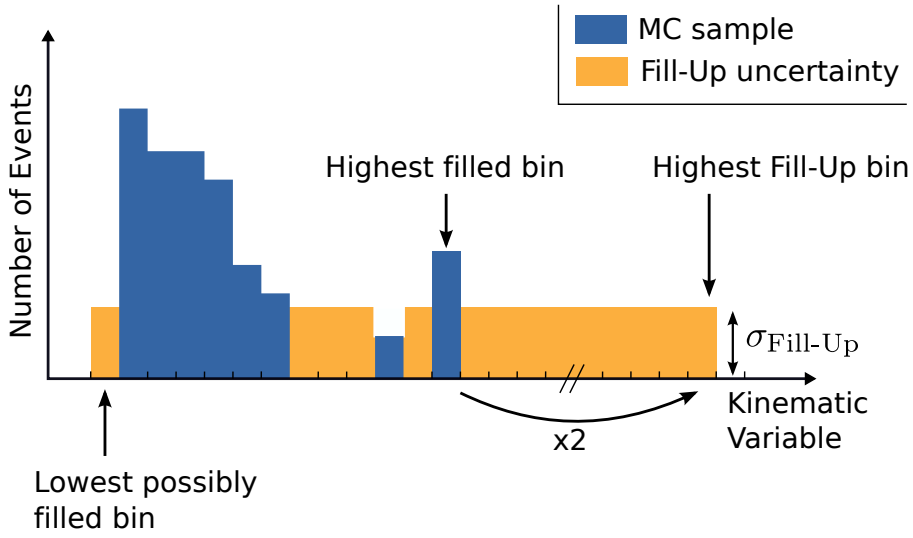


Figure 2.10: Schematic of the Fill-Up procedure. Between the lowest possibly filled bin and the bin containing twice the central value of the highest filled bin, all empty bins are assigned an uncertainty $\sigma_{\text{Fill-Up}}$.

The handling of PDF sets is implemented in the LHAPDF package [84], version 6.1.3 of which is used.

2.4.4.5 Low Statistics Uncertainty

The number of events in a simulation sample can be characterised by the effective luminosity $\mathcal{L}_{\text{eff}} = N_{\text{Events}}/\sigma_{\text{Process}}$. To minimize the influence of single simulation events, $\mathcal{L}_{\text{eff}} > \mathcal{L}_{\text{data}}$ is favorable. Otherwise, any measure of deviation between data and simulation will become sensitive to binning effects: If e.g. an expected mean event yield of one event per bin is modeled by a sample with insufficient $\mathcal{L}_{\text{eff}}$, a bin may just by chance contain no simulated event and a single (or more) data events. Since all uncertainties are only calculated based on event characteristics, no uncertainty is assigned to such a bin, which in turn prevents the calculation of a p-value. This raises a problem although the mean event density per unit kinematic quantity is modeled correctly.

In reality, not all simulation samples have sufficiently large $\mathcal{L}_{\text{eff}}$. To prevent problems due to insufficient numbers of simulated events, a *Fill-Up* procedure is introduced, which is designed to estimate the influence of the simulation of additional events. It is performed separately for each kinematic distribution and each physics process modeled in simulation. A schematic view of the procedure is given in fig. 2.10. Three steps are performed:

1. The lowest possibly filled bin is determined. In the $\Sigma|p_T|$-distribution, it is defined by the sum of the p_T thresholds of the contributing objects. E.g. in the the 3jet excl. class, the lowest possibly filled bin begins at $\Sigma|p_T| = (400 + 50 + 50)\,\text{GeV} = 550\,\text{GeV}$, i.e. the sum of the trigger threshold for the leading jet and the identification thresholds for the two non-leading jets. For the E_T^{miss}-distribution, the lowest bin begins at the E_T^{miss} object identification threshold of 50 GeV (cf. sec. 2.2.6), in the M_{inv}-distribution, it begins at zero. This bin may or may not actually be filled.

2. The highest actually filled bin is determined.

3. All bins which are located between the lowest possibly filled bin and the bin containing twice the central value of the highest actually filled and contain no simulated events receive an additional uncertainty $\sigma_{\text{Fill-Up}}$. The bin content is not changed.

If there are no empty bins before the highest filled bin, i.e. there are no "holes" in the simulated spectrum, the uncertainty is only added for the bins after the highest filled bin. The choice of $\sigma_{\text{Fill-Up}}$

is motivated by the hypothetical addition of one more unweighted event to the MC sample, which may have "just been missed" by the limited of number of produced events. The uncertainty is thus chosen to be normalized to the mean sample weight $\langle w_{\text{total}} \rangle$ (cf. eq. 2.5) over the whole distribution.

$$\sqrt{\sum_{\text{Empty bins}} \sigma^2_{\text{Fill-Up}}} = \langle w_{\text{total}} \rangle$$

The uncertainty is chosen to be flat, i.e. independent of the kinematic variable, which directly yields an expression for the uncertainty value.

$$\sigma_{\text{Fill-Up}} = \frac{\langle w_{\text{total}} \rangle}{\sqrt{N_{\text{empty bins}}}}$$

Because the uncertainty is designed to deal with empty bins, it is one-sided in the upward direction.

2.4.5 Jet Resolution Smearing

In measurements of jet p_T resolution, differences between data and simulation are observed and scale factors $c = \frac{\sigma_{Data}}{\sigma_{MC}}$ are calculated [55]. The resolution measurement is based on dijet events and events with a Z boson and recoil jet, which are intrinsically balanced in p_T, the distribution of the difference between the measured p_T-values can be used to determine the experimental resolution. Generally, the simulated resolution is better than in data, with scale factors between 1.0 and 1.3. To correct for this, the resolution of simulated jets is artificially degraded using a centrally recommended method [85]: For reconstructed jets with a matched generator jet, the difference between reconstructed p_T and generated p_T is increased by a factor c.

$$p_{T,\text{rec}} \to \max\left(0,\, p_{T,\text{gen}} + c \cdot \left(p_{T,\text{rec}} - p_{T,\text{gen}}\right)\right)$$

If a reconstructed jet has no matched generator jet, no information about $p_{T,\text{gen}}$ is available. In this case, $p_{T,\text{rec}}$ is smeared by a Gaussian centered on $\mu = 1$ with width $\sigma = \sqrt{c^2 - 1} \cdot \sigma_{MC}$, where σ_{MC} is the width of the $\frac{p_{T,\text{rec}} - p_{T,\text{gen}}}{p_{T,\text{gen}}}$ distribution measured in simulation (cf. sec 3.2).

$$p_{T,\text{rec}} \to p_{T,\text{rec}} \cdot \text{Gauss}(\,\mu,\, \sigma\,)$$

2.4.6 Binning

Variable bin sizes are determined separately for each kinematic distribution of each event class. The lowest bin in each distribution begins at zero. Starting from a minimal bin width of 10 GeV, its width is increased in steps of 10 GeV until the bin width is larger than the resolution of the corresponding kinematic variable at the bin center.

$$\text{bin width} \geq \sigma(\text{bin center}) \tag{2.6}$$

The process is repeated for the subsequent bins until the upper boundary of 8 TeV is reached. All bin widths are integer multiples of 10 GeV.

Resolution functions for individual objects are used as input for the kinematic variable resolution. The object resolution functions are are described in sec. 2.4.6.1, the kinematic variable resolution functions are given in sec. 2.4.6.2.

2.4.6.1 Resolution for Objects

For each physics object, a parametrised resolution function is used. For a given event class, the object resolution functions serve as input to the resolution functions of the kinematic variables.

The (b-)jet p_T resolution as a function of p_T is obtained from a dedicated CMS analysis [55].

$$\sigma_{\text{jet/b}}(p_T) = \sqrt{9\,\text{GeV}^2 + 0.72\,\text{GeV} \cdot p_T + 1.8 \times 10^{-3} \cdot p_T^2} \tag{2.7}$$

The electron p_T resolution functions are approximated by substituting $p_T = E_T$ in energy dependent resolution functions [7]. The photon resolution is taken to be approximately equal to the electron resolution.

$$\sigma_{e/\gamma}(p_T) = \sqrt{1.44\,\mathrm{GeV}^2 + 7.84 \times 10^{-4}\,\mathrm{GeV} \cdot p_T + 9 \times 10^{-6} \cdot p_T^2} \tag{2.8}$$

The E_T^{miss} resolution depends on $\Sigma|p_T|$ of all particles used to calculated it [86], which is approximated by the $\Sigma|p_T|$ variable calculated in the corresponding event class.

$$\sigma_{E_T^{\mathrm{miss}}}(\Sigma|p_T|) = 1.78\,\mathrm{GeV} + 0.63\sqrt{\mathrm{GeV}} \cdot \sqrt{\Sigma|p_T|} \tag{2.9}$$

2.4.6.2 Resolution for Kinematic Variables

As a basis for binning, resolution functions for the kinematic variables are necessary that only depend on the event class final state and the kinematic variable itself.

The $\Sigma|p_T|$-resolution is estimated by adding the p_T-resolutions of the final state objects in quadrature. The single-object resolution is evaluated at $\langle p_T \rangle = \Sigma|p_T|/N_{\mathrm{objects}}$.

$$\sigma_{\Sigma|p_T|}(\Sigma|p_T|) = \sqrt{\sum_{\text{object type i}} N_i \cdot \sigma_i^2\left(\langle p_T \rangle\right)} \tag{2.10}$$

σ_i is the resolution of object type i, N_i is the number of objects of type i in the event class name.

The M_{inv}-resolution cannot be determined in a final state independent way. It is approximated by evaluating the $\Sigma|p_T|$-resolution at $\Sigma|p_T| = M_{\mathrm{inv}}$. The approximation is exact for the simplified final state topology of objects which are equidistant in ϕ and are located at $\eta = 0$.

$$\sigma_{M_{\mathrm{inv}}}(M_{\mathrm{inv}}) = \sigma_{\Sigma|p_T|}(\Sigma|p_T| = M_{\mathrm{inv}}) \tag{2.11}$$

The E_T^{miss} resolution is obtained from eq. 2.9. It is independent of the resolution functions of the other objects in the event class. In order to be able to parametrise it as a function of E_T^{miss}, the approximation $E_T^{\mathrm{miss}} \approx \Sigma|p_T|$ is used, which is conservative because it can only overestimate the resolution.

$$\sigma_{E_T^{\mathrm{miss}}}(E_T^{\mathrm{miss}}) = \sigma_{E_T^{\mathrm{miss}}}(\Sigma|p_T| = E_T^{\mathrm{miss}}) \tag{2.12}$$

3 Dedicated studies

3.1 Jet Identification and Noise Rejection

The choice of jet identification criteria is motivated by a need for sufficient noise rejection. A benchmark for noise rejection and E_T^{miss} performance is the 1jet+E_T^{miss} event class. It is especially sensitive to E_T^{miss} induced by instrumentation effects. Noise spikes in the HCAL, for example, may cause the Particle Flow algorithm to reconstruct a neutral hadron, which is then used in jet clustering and may give rise to an additional, unphysical jet. This noise is not present in simulation and event filters are used to reject events in data with noise from known sources (cf. sec. 2.1.3).

Fig. 3.1 shows the E_T^{miss}-distribution of the 1jet+E_T^{miss} class with the *PFJet Loose* jet identification criterion[1] applied after all recommended filters. The simulated distribution consists of two regimes: a QCD dominated low-E_T^{miss} region and a high-E_T^{miss} region containing contributions from W and Z boson decays involving neutrinos. From 400 GeV up to the very high E_T^{miss}-tails, a significant excess is observed in data. In this section, the origin of this excess and its relation to the applied jet ID criteria is studied. The findings are used to motivate a tightening of the jet identification requirements and an event rejection criterion.

3.1.1 Jet composition

Sources of anomalous E_T^{miss} are studied by the MET POG [87] and event filters are implemented to remove events contaminated by known sources of noise. However, noise filters have a finite efficiency, which causes them to miss a certain fraction of noise events. Thus, even after the application of these filters, residual high-E_T^{miss} events are observed [86]. These events are found to be removed by rejecting events containing jets not fulfilling tight ID criteria. Jet ID criteria impose restrictions on jet composition.

The composition of a jet can be accessed via jet energy fractions, which reflect the fraction of jet energy resulting from each kind of elementary object. The jets used in this analysis are clustered from Particle Flow candidates and offer five energy fractions, one for each type of Particle Flow candidate[2].

- Neutral Hadron Fraction (NHF)
- Charged Hadron Fraction (CHF)
- Neutral Electromagnetic Fraction (NEF)
- Charged Electromagnetic Fraction (CEF)
- Muon Fraction (MUF)

In the rapidity range $|\eta| < 2.5$, the five fractions add up to one, leaving only four of them independent. Because the applied object overlap removal procedure favors muons, only a negligible number of jets has a muon fraction significantly different from zero. Intuitively, "real" jets from hadronisation show a different clustering behavior in these quantities than jets caused by noise. In the example case of an HCAL noise spike giving rise to a reconstructed jet, the jet will carry large NHF. An important

[1] The PFJet Loose criterion differs from the PFJet Tight criterion introduced in sec. 2.2.4 by relaxed requirements on the Neutral Hadron, Neutral Electromagnetic and Charged Electromagnetic energy fractions, which are required to be smaller than 0.99 instead of 0.9 [40]. Since the fractions are per definition smaller than one, the *loose* requirements are indeed very inclusive.

[2] For convenience, we just refer to "fractions" like "Neutral hadron fraction", which strictly speaking are called "energy fractions", as in "Neutral hadron energy fraction".

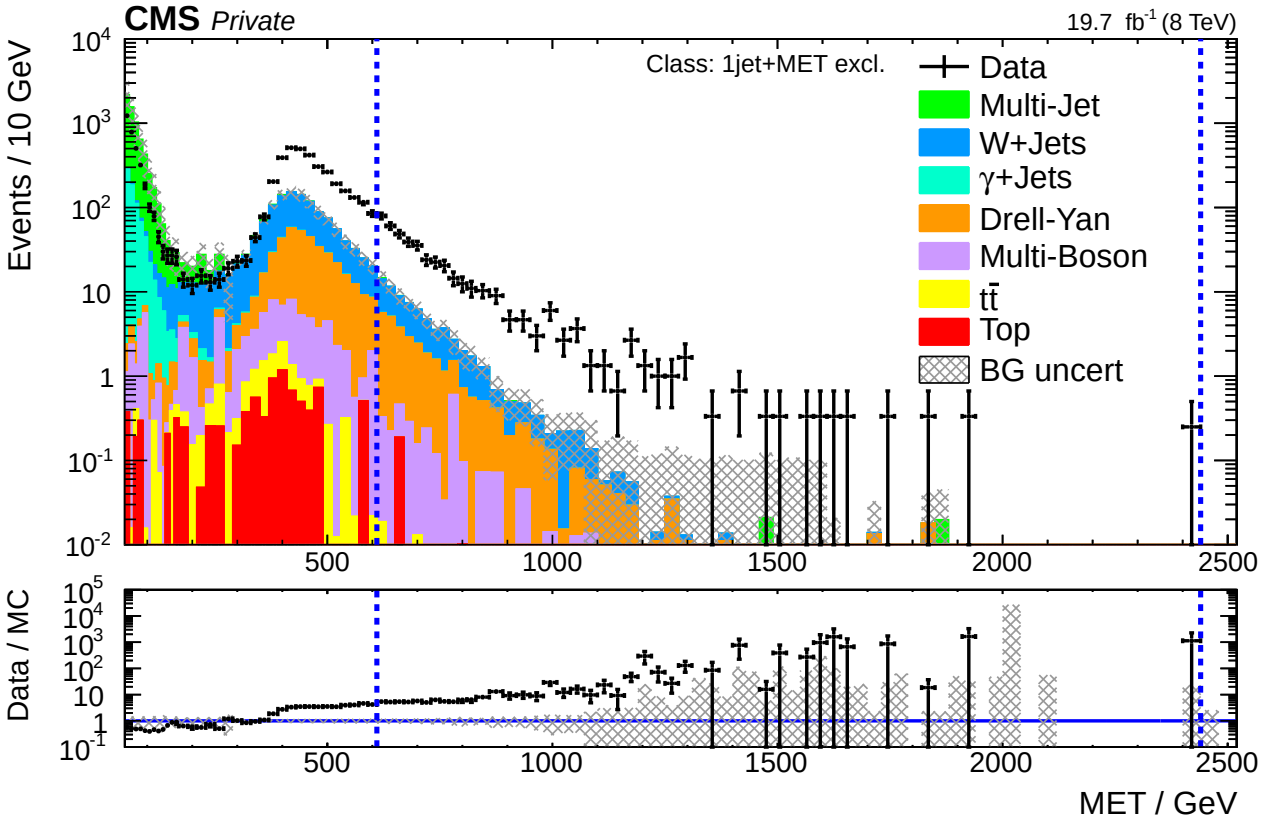


Figure 3.1: E_T^{miss}-distribution of the 1jet+E_T^{miss} event class with the loose PFJet ID applied. The low-E_T^{miss} region ($< 250\,\text{GeV}$) is dominated by QCD events, which do not cause intrinsic E_T^{miss}. Large E_T^{miss} is produced in events containing W and Z bosons and a recoil jet. Beginning at 400 GeV a large excess of data compared to simulation is visible. The onset of the additional data component coincides with the jet trigger threshold.

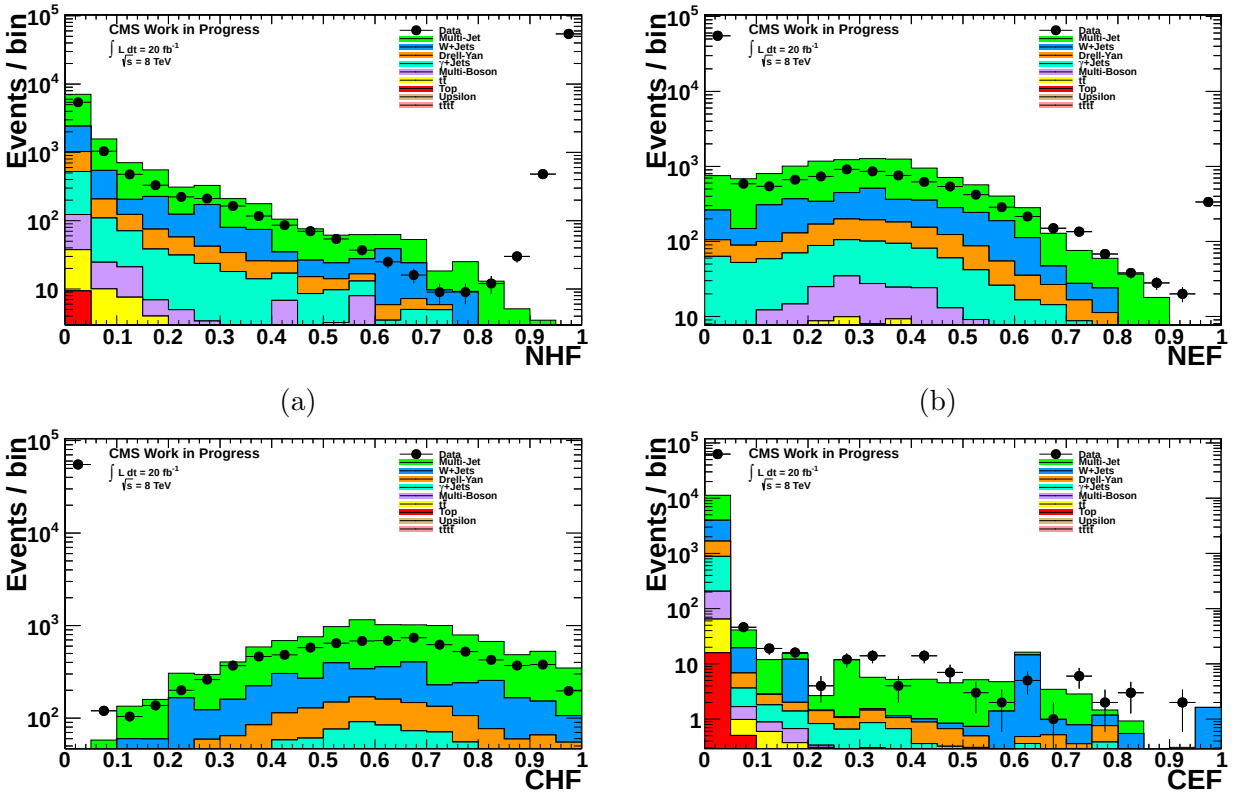
property of energy composition as a jet quality indicator is its independence from the characteristics of the primary physics interaction. The energy composition is a result of hadronisation, which is governed by the characteristics of non-perturbative QCD. Because of factorisation (cf. 1.1.4), the primary interaction can influence the jet composition only by determining the outgoing parton flavor and momentum, the effects of which are moderate in strength and well understood [55].

3.1.2 Energy fractions in Monojet + E_T^{miss}

Fig. 3.2 shows the distributions of the jet energy fractions for the jet in 1jet+E_T^{miss} events. Good agreement of data and simulation is observed over most of the range, with exceptions at NHF > 0.9 and NEF > 0.9. A very similar deviation is observed in the CMS analysis of the monojet+E_T^{miss} final state [88, 89] before strict jet selection criteria are applied. Good agreement is also observed there for the jet energy fractions with the exception of regions close to zero and one, respectively. However, not all of the E_T^{miss} event filters are used in the monojet analysis, which leaves unclear what fraction of the deviation would be removed by other filters.

Following [90], Events with a jet with $p_T > 30\,\text{GeV}$ and $NHF > 0.9$ or $NEF > 0.9$ are selected for further investigation. These events and the corresponding jets are referred to as "tagged".

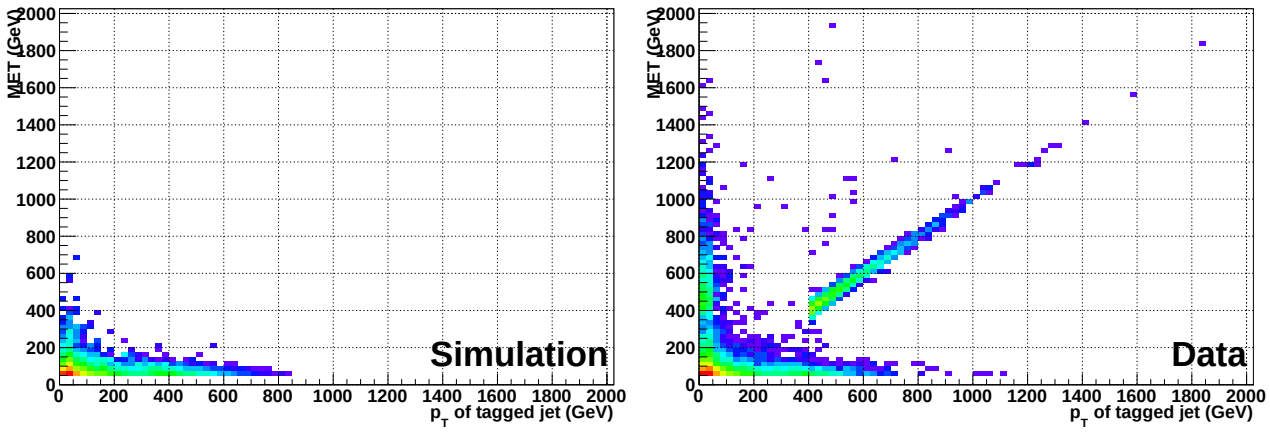
Fig. 3.3 shows the correlation between the p_T of the tagged jet and E_T^{miss} in the event for all tagged events containing $E_T^{\text{miss}} > 50\,\text{GeV}$. Both data and simulation have a continuous distribution close to the origin with tails leading to the high-p_T and high-E_T^{miss} region. However, in data, there are additional high-E_T^{miss} tails, and a striking feature of exactly correlated p_T and E_T^{miss}, which is not present in simulation. The exact correspondence of the two quantities indicates that E_T^{miss} is



(a) (b)

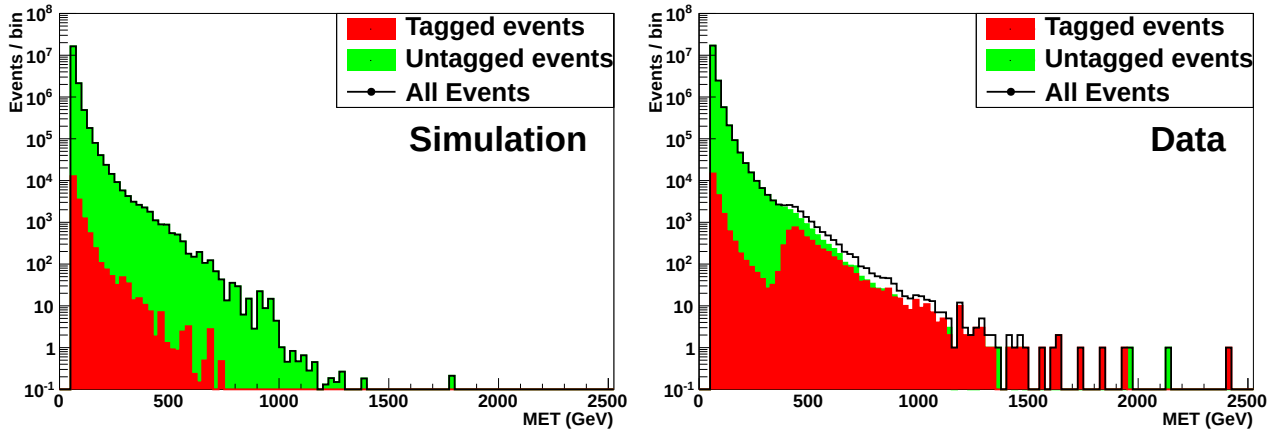
(c) (d)

Figure 3.2: Neutral Hadron (a), Neutral Electromagnetic (b), Charged Hadron (c) and Charged Electromagnetic (d) energy fractions of jets fulfilling the PFJet Loose ID in the Monojet+E_T^{miss} final state. All fractions show acceptable shape agreement over a large range of the distribution. However, at large NHF and NEF values, a significant excess is observed in data. Because the sum of the fractions cannot exceed one, an excess is also visible for low values of NEF, CHF and CEF.



(a) (b)

Figure 3.3: Distribution of E_T^{miss} in events where at least one jet has $p_T > 30\,\text{GeV}$ and $NHF < 0.9$ or $NEF > 0.9$ in dependence of p_T of the respective jet for (a) simulation and (b) data. In addition to a similar clustering behavior at low E_T^{miss}, data show significant high-E_T^{miss} tails.



(a) (b)

Figure 3.4: Distribution of E_T^{miss} in tagged and untagged events in (a) simulation and (b) data. No final state requirement is imposed, except for the presence of $E_T^{\text{miss}} > 50\,\text{GeV}$. In simulation the spectrum of tagged events closely follows the same shape as that of the untagged events and presents a minor contribution over the complete range. In data, the tagged spectrum has an additional spike feature around $E_T^{\text{miss}} \approx 500\,\text{GeV}$, close to the trigger threshold for single jets, and contributes significantly to the total distribution.

predominantly calculated from the jet p_T in these events. This happens when the jet constituents are the only reconstructed Particle Flow candidates or at least collectively carry the majority of momentum in the event. In this case the formula for E_T^{miss} (cf. eq. 1.5) reduces to the negative of the jet p_T.

The E_T^{miss}-distributions in data and simulation are shown in fig. 3.4. In simulation, the spectra of tagged and untagged events do not exhibit characteristic differences in shape. Over the complete range of the distribution, tagged events make up only a small fraction ($\mathcal{O}(0.1\,\%)$) of the overall spectrum. In data, this behavior is observed only for the low-E_T^{miss} region up to $E_T^{\text{miss}} \approx 400\,\text{GeV}$. The high-$E_T^{\text{miss}}$ tails are dominated by tagged events causing high-E_T^{miss} excesses compared to simulation. The onset of the excess above 400 GeV, the trigger threshold for single jets, is consistent with E_T^{miss} caused predominantly by single jets.

A strong indication for noise contamination in the tagged events is the distribution of the azimuth angle ϕ of the E_T^{miss} vector. Because of the rotational symmetry of the incoming beam configuration, ϕ is expected to be uniformly distributed, with any deviation being caused by detector effects, such as local detector inefficiencies. In the analysis of the 2012 data set, significant ϕ-modulation was observed and a heuristic correction method was derived [91], which is also applied in this analysis (cf. sec. 2.2.6). The *xy-shift* method parametrises the average x- and y-components of the E_T^{miss} vector as a function of the number of reconstructed vertices in an event and subtracts them. After the correction, residual modulation on the level of 5% is observed in events with Z bosons decaying to muons [91]. Intensive simulation studies have been performed to identify the effects of different sources of modulation such as local detector inefficiencies and the location of the beam crossing [92]. While it could be shown that modulation amplitudes and phases vary for events with objects reconstructed with different detector components[3], no comprehensive correction method has been found. In the simulation samples described in sec. 2.4.1, modulation sources known previously to sample production, such as detector alignment and inefficiencies, are included. However, not all sources are simulated and the simulated sources may change after sample production. Data and simulation show similar modulation

[3] In an event where most of the energy deposits are contained in the ECAL, the modulation will in general be different than in an event where most energy is deposited in the HCAL. While most events are not primarily reconstructed with a single detector component, the energy distribution in the components will vary.

amplitudes but different phases, resulting in deviations of the order of the modulation amplitude even after applying corrections [91].

Fig. 3.5 shows the distribution of the azimuth angle of the E_T^{miss}-vector both with and without the monojet final state selection in both tagged and untagged events. The color coding indicates the contributing physics processes: One of the dominant sources of E_T^{miss} in SM events is W-boson production in association with jets. In leptonic decays of W-bosons, neutrinos are created, causing intrinsic E_T^{miss}. Tagged events in simulation are dominated by photon production and EM-enriched QCD, i.e. QCD events with jets containing electrons. The presence of electrons and photons in these events causes them to sometimes pass the tagging requirement on the EM fraction. The spectra in fig. 3.5 not only differ in SM composition but also in their level of ϕ-modulation. With respect to the underlying flat plateau, the untagged spectra show features on the order of $5-10\,\%$ in agreement with the modulation amplitude present in simulation. The modulation in simulation arises from events with generator-level E_T^{miss} of less than 50 GeV, which are susceptible to E_T^{miss} caused by reconstruction effects. An increase in the reconstructed E_T^{miss} identification threshold to e.g. 100 GeV would significantly reduce the modulation strength.

In tagged events in the monojet+E_T^{miss} final state, sharp structures are observed reaching up to four times the height of the plateau, which is a strong indicator for noise contamination. Without the monojet+E_T^{miss} final state requirement, the modulation is much less pronounced even in tagged events, although some of the peaking structures are still discernible. The final state dependence of the modulation strength indicates that noise-induced events are predominantly present in the monojet+E_T^{miss} final state.

When one compares fig. 3.1 to Dark Matter signal shapes (cf. fig. 4.35), the similarity is apparent, raising the question whether a BSM phenomenon is observed here. Beside the previously studied evidence of a noise-like nature of the excess, a BSM explanation is made unlikely by the extreme mismatch between the excess size and isolation. Most strikingly, it is isolated and completely contained in a very small region of jet composition. As hadronisation and parton showering are statistical processes governed by perturbative QCD, it is very hard to imagine a BSM mechanism causing such an effect. The excess is also isolated in one final state, without any overlap with neighbouring distributions. As is studied in sec. 4.7, an excess of this size in the monojet+E_T^{miss} channel would also give considerable contributions in other multijet+E_T^{miss} event classes. Because the branching into multiple final states is driven by initial state radiation, it is also a general feature that is independent of the BSM phenomenology. In conclusion, the excess in the 1jet+E_T^{miss} event class does not only "look like noise", it also does *not* "look like a signal", which in combination makes the BSM hypothesis very unlikely in this case.

To remove noise-like jets, the PFTight jet ID is employed (cf. sec. 2.2.4). This leads to a de facto removal of tagged events from the Monojet + E_T^{miss} class because the only remaining object is E_T^{miss} and no E_T^{miss} exclusive class is considered. However, tagged events with additional final state particles will still contribute to some event class. Since these events contain in part sizable E_T^{miss} contributions caused by noise, a noise contamination in other classes is to be expected. Although this may not cause large deviations if the respective classes are sufficiently populated by noise-free events, a rejection scheme is employed. In a manner similar to [86, 90], tagged events are rejected altogether. This filtering based on individual jet characteristics is referred to as *Particle Based Noise Rejection* (PBNR). A total number of 7×10^5 additional data events, i.e. events rejected by no other event filter, is rejected by this criterion. This corresponds to less than 1 % of the total number of events passing the complete selection. The resulting E_T^{miss}-distribution of the 1jet+E_T^{miss} class is shown in fig. 3.6.

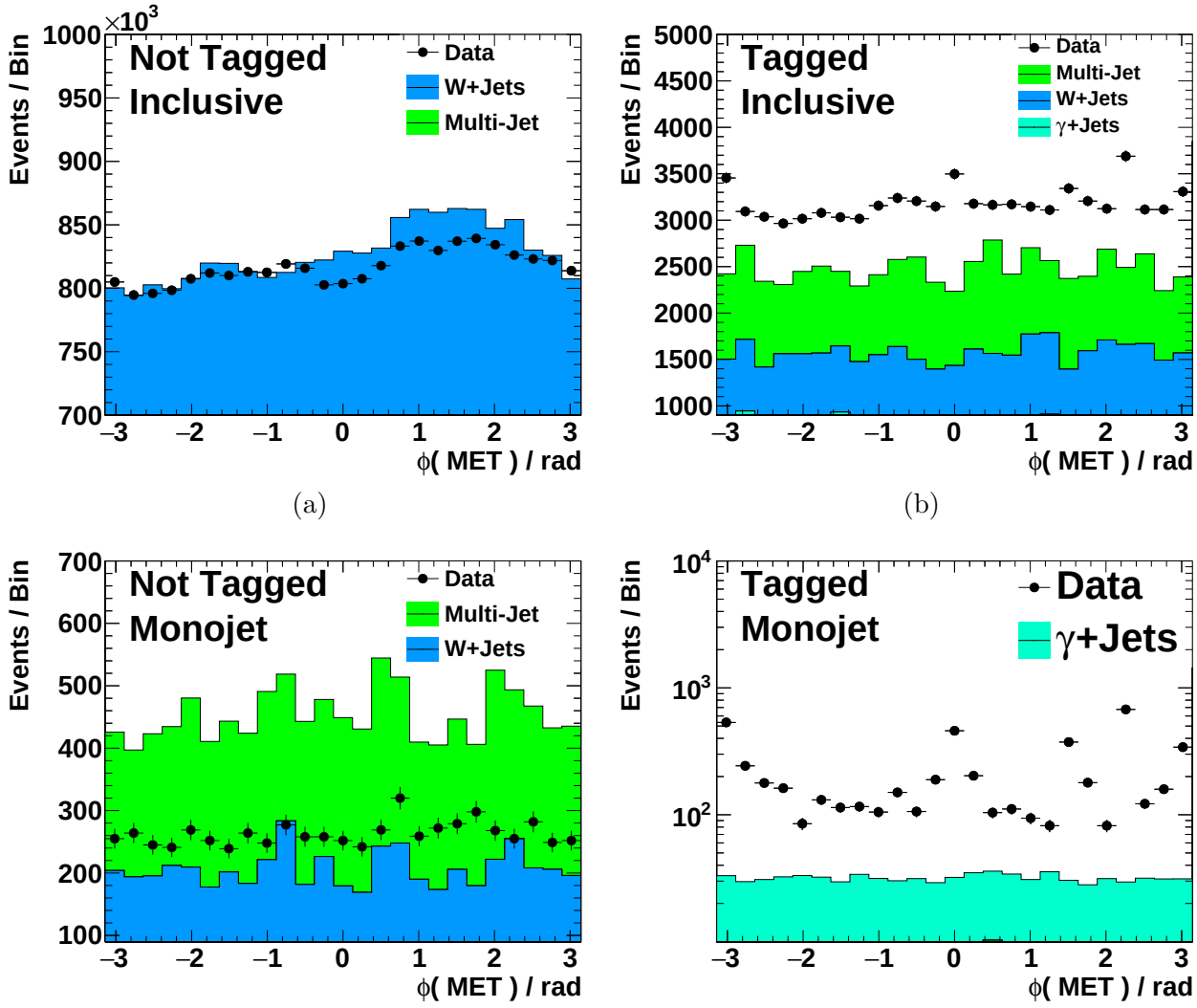


Figure 3.5: ϕ-distributions of the E_T^{miss} vector in events with $E_T^{\text{miss}} > 50\,\text{GeV}$ (top row) and events in the 1jet+E_T^{miss} final state (bottom row). The plots on the left (right) show untagged (tagged) events. Events are said to be tagged if they contain at least one jet with $p_T > 30\,\text{GeV}$, NHF> 0.9 and NEF> 0.9. Note the logarithmic scale in (d). Simulation sample composition is as in the main analysis (cf. sec. 2.4.1). While all spectra show some amount of ϕ-modulation, especially pronounced features are visible in the tagged spectra.

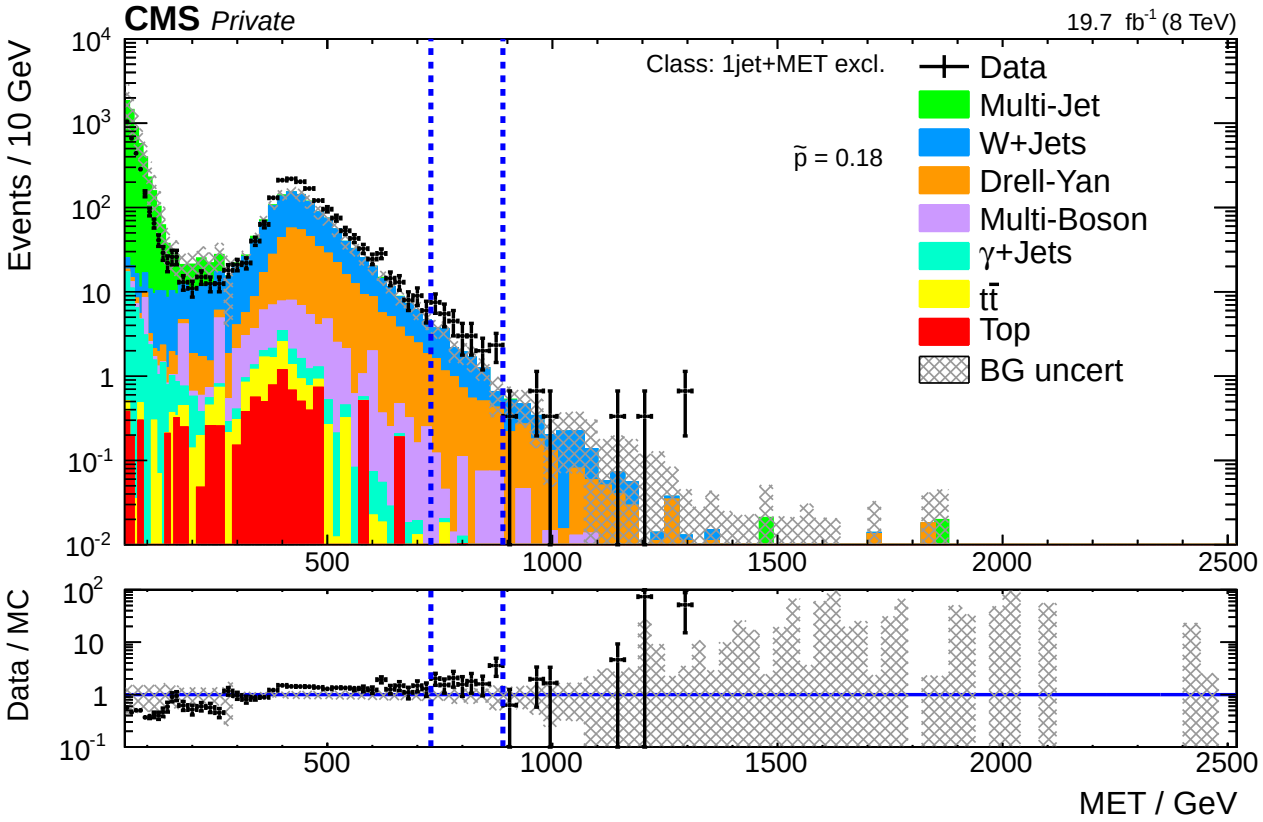


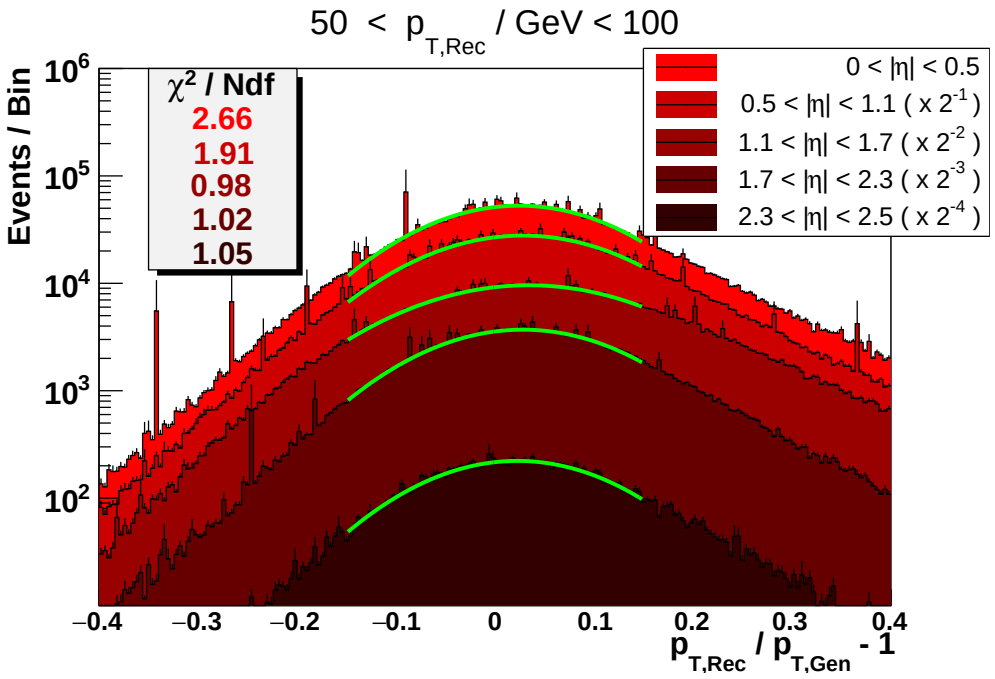
Figure 3.6: E_T^{miss}-distribution of the 1jet+E_T^{miss} event class with the tight PFJet ID and Particle Based Noise Rejection (PBNR) applied. While the low-E_T^{miss} region is unchanged with respect to fig. 3.1, the high-E_T^{miss} tail is significantly reduced in data.

3.2 Jet resolution measurement

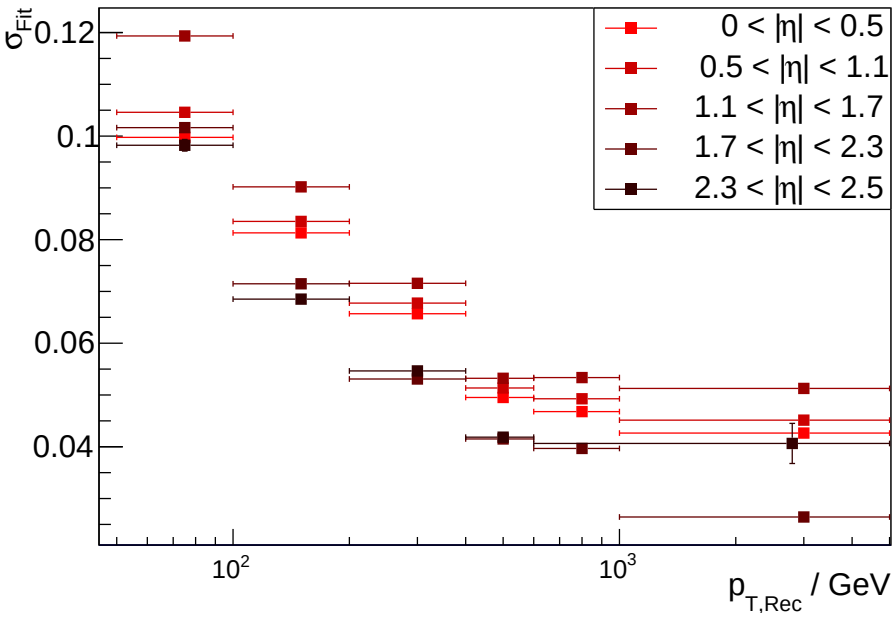
In order to perform jet resolution corrections (cf. sec 2.4.5), σ_{MC}, the core width[4] of the distribution of $p_{T,rec}/p_{T,gen}$[5] in simulation, must be determined. For this purpose, all pairs of matched generator and reconstructed jets are selected and $p_{T,rec}/p_{T,gen}$ is recorded. The distributions are fitted using Gaussian functions over a range of approximately two σ and the width is taken from the fit result. Events are selected in the same way as in the main analysis and no final state requirements are employed. To account for the p_T- and $|\eta|$-dependence of the jet resolution, the measurement is performed in bins of both variables. The edges of the $|\eta|$-bins are chosen in accord with [55], which provides the used data-simulation scale factors (cf. sec. 2.4.5). Example distributions are shown together with their respective fits in fig. 3.7 (a) for the p_T-bin with $50 < p_{T,Rec} < 100\,\text{GeV}$. Reasonable fit quality is observed. Fig. 3.7 (b) shows the widths from all fits. With increasing jet p_T, the resolution improves due to the reduced influence of low energy pile-up particles and the improving calorimeter resolution. A somewhat smaller modulation with $|\eta|$ is observed with the unusual effect that the core width at first worsens with increasing $|\eta|$ but then improves even over the values at central rapidity. This is due to the binning in p_T and $|\eta|$: In a given p_T bin, the average total jet momentum will be larger for higher $|\eta|$, which results in an improved calorimeter resolution. It was checked that the effect can be reversed by binning in the momentum p instead of the transverse momentum p_T.

[4] The distribution is expected to be only Gaussian at its core, i.e. close to the main peak. The tails, i.e. regions farther away from the peak, are usually not Gaussian. Only the width of the main peak is of interest for our correction method. Thus only the central *core width* is measured.

[5] *Rec* and *gen* refer to *reconstructed* and *generated* values.



(a)



(b)

Figure 3.7: Example distributions of $p_{T,rec}/p_{T,gen} - 1$ (a) with fitted Gaussian functions (green solid lines) and the resulting Gaussian widths (b). The distributions are binned in $|\eta|$ and $p_{T,rec}$. Distributions for all $|\eta|$-bins for $50 < p_{T,rec} < 100\,\text{GeV}$ are shown. Fitting is performed over the core range of approximately two σ to avoid the non-Gaussian tails. Fit quality is found to be satisfactory. Uncertainties are statistical only. In the last $|\eta|$-bin, the two highest p_T bins are merged because of diminishing event statistics.

4 Results

The analysis is applied to the full data set of $\mathcal{L}_{int} = 19.5\,\text{fb}^{-1}$ recorded by CMS in 2012 at $\sqrt{s} = 8\,\text{TeV}$. A total[1] of 1340 exclusive event classes is observed containing a grand total of 9.9×10^7 data and 11.1×10^7 expected events. Of the observed events, 1.1×10^7 events are selected by jet triggering, resulting in 229 event classes without leptons. Of all event classes, 960 contain at least one b jet. In the inclusive classes, the same total set of events is distributed among 1610 classes, the contents of which overlap. Of all exclusive event classes, 727 contain no data, and 706 of these classes contain simulation that is consistent with zero observed events within the simulation uncertainty. For the inclusive event classes, these numbers are 965 and 875, respectively. The opposite case, where an event class contains data but no simulation is not observed.

The increase of the number of event classes due to jet triggering is only modest by itself, but enhanced by the addition of b-quarks. Because the differences between identification and triggering criteria are relatively small for leptons, jet triggering will hardly contribute new events to classes such as 2μ+1jet, since the events contributing to such a class will very likely be contained in the muon data stream. Significant numbers of events will only be added in classes which do not contain any other triggering particles, i.e. are only made up of non-triggering objects. Without b tagging, only E_T^{miss} and photons are non-triggering objects, resulting in only 54 jet triggered classes. By adding b-tagged jets, more combinations are made possible, resulting in the 229 exclusive event classes quoted above. The largest part of new event classes is initiated by the addition of b jets, which give rise to new event classes independent of jet triggering. Nearly all classes previously containing some jet multiplicity are split into different classes with different multiplicities of tagged and non-tagged jets, yielding the substantial increase by 960 exclusive classes.

In sec. 4.1 the results of an *Integral Scan* are presented (cf. sec. 1.3.3). In this scan mode, every event class is treated as an independent single-bin counting experiment and the total integrated SM expectation is compared to the total observed event yield. This view is suitable to give an overview over a large number of classes and is sensitive to total process cross-sections.

In sec. 4.2 and 4.3, the results of kinematic distribution scans are discussed. Here, the distribution shape is taken into account and the most significant connected bin region (RoI) is found, for which a LEE-corrected $\tilde{p}$-value is calculated (cf. sec. 1.3.2.5). This scan method gives sensitivity to smaller local deviations that may not be visible in the integral scan. Single classes with significant deviations are discussed in this section.

Interesting effects which are spread over multiple event classes and event class types are discussed in sec. 4.4. This section is focused the combination of different types of evidence generated by MUSiC.

Because of this document's focus on the use of jet triggered classes and b tagging, sec. 4.5 and 4.6 are dedicated to the discussion of the kinematic distributions of jet and b classes, independent of their significance.

4.1 Integral Scan of Exclusive and Inclusive Event Classes

The event classes showing the most significant deviations in integral scans of the exclusive and inclusive event classes are summarised in fig. 4.1 and 4.2, respectively. Two interesting patterns are observed in both scans: The dilepton+E_T^{miss} (+X) final states and a large number of event classes with low numbers of contributing events are among the most significantly deviating.

[1] All event classes where there is at least one (unweighted) event found in data or simulation are counted in the total number.

The presence of the dilepton+E_T^{miss} final states is related to problems in the modeling of experimental effects on the E_T^{miss}-distribution. Because the problem manifests itself in a region populated with large numbers of events, a significant normalisation deviation is produced. This is discussed in sec. 4.4.1.

In most classes where significantly less than one event is expected from simulation, the relative uncertainty of the expected value is large because only a small number of simulation events contribute per physics process. The SM expectation in these classes is thus consistent with zero observed events, but already one or two observed events will yield significant p-values. In the set of low-statistics classes, the choice of selecting the most significant p-values is thus biased in favor of excesses. A simple way of testing this mechanism is by considering all exclusive event classes with small simulation predictions ($N_{\text{SM}} < 1$) and integrating over them. There are 805 such event classes, of which 97 contain at least one data event and 708 do not. In total, approximately 90 events are expected, while 130 are observed, indicating that the prevalence of excesses is not only caused by the selection bias. Of course, the exact numbers depend on the arbitrary choice of what to consider a low-statistics class, i.e. what threshold to choose for N_{SM} and generally the data/simulation-ratio approaches one for higher thresholds. This effect is likely enhanced as an unfortunate side-effect of the introduction of b tagging, which further splits up high-multiplicity final states and thus creates more event classes with low event counts. The issue may also be related to multiplicity modeling. Exact single jet multiplicities, e.g $N_{\text{jets}} = 6$, are commonly observed to be insufficiently modeled in simulation, which generally underestimates the high-multiplicity cross-section [93–95] and thus causes apparent excesses in data. This effect is independent of the low-statistics issue, but often coincides with it because of the low cross-sections for high final state multiplicities. An improvement of this issue may be achieved by simulation software such as Sherpa and MadGraph5_aMC@NLO [96], which provide consistent parton shower matching to NLO matrix element calculations.

In summary, no clear signs of BSM phenomena are observed. An eye-catching surplus of excesses in low-statistics classes is likely due to a combination of selection bias and insufficient modeling of final states with high jet multiplicities. Significant deviations in the dilepton+E_T^{miss} are likely due to insufficient modeling of detector effects.

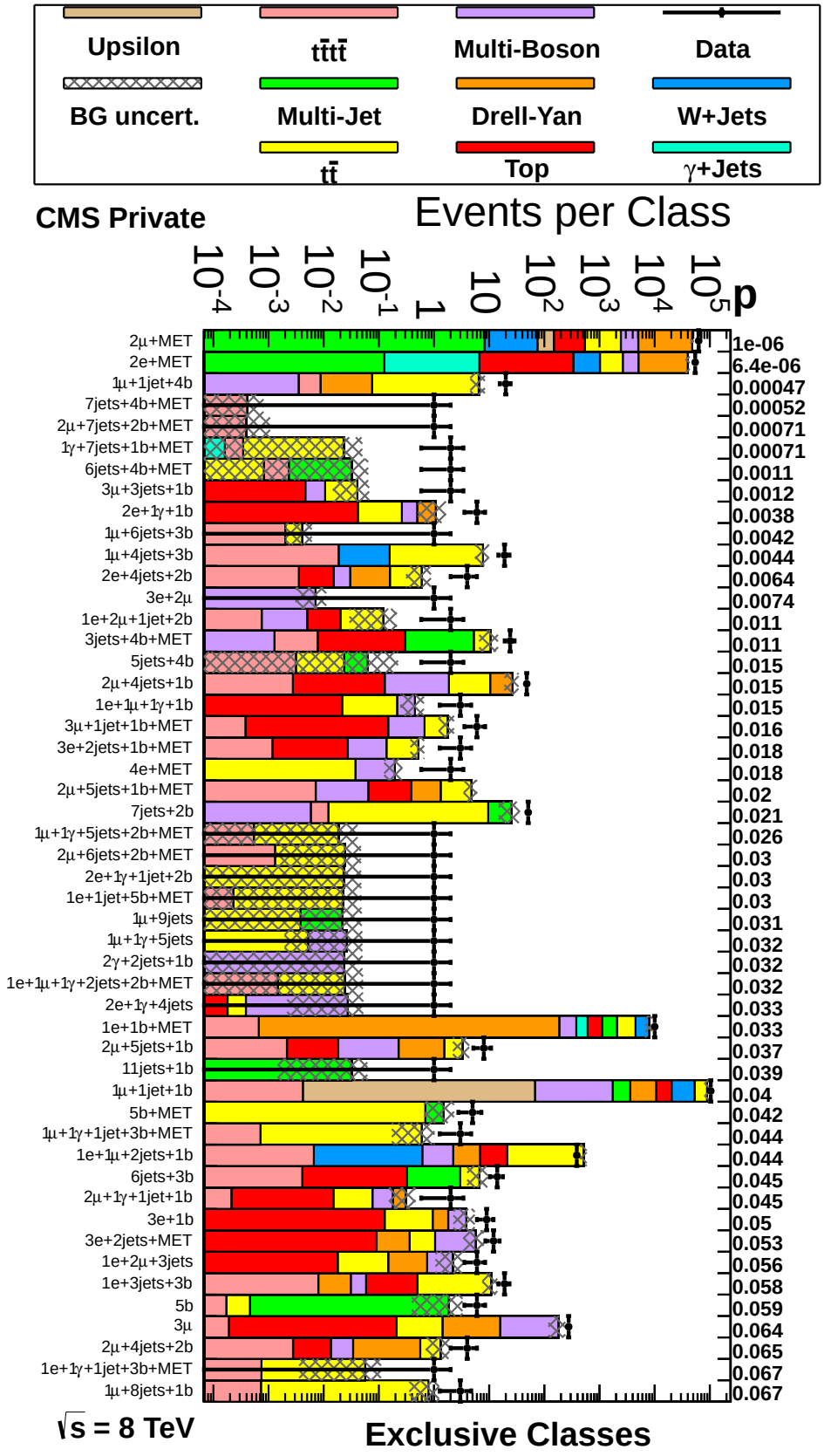


Figure 4.1: Overview of the 50 exclusive classes with the most significant p-values in an integral scan. Distribution shape information is disregarded, every class is considered as a single-bin counting experiment. In the graphic, each line corresponds to one event class, with the bar indicating the SM expectation in the class and black the point indicating the observed number of events. Color coding of SM simulation is performed as in other MUSiC plots. The p-value calculated for each class is given to the right of each line, classes are ordered by significance. Because each class has the same number of bins $N_{\text{bin}} = 1$ in this view, p- and $\tilde{p}$-values are numerically identical (cf. 1.3.2.5).

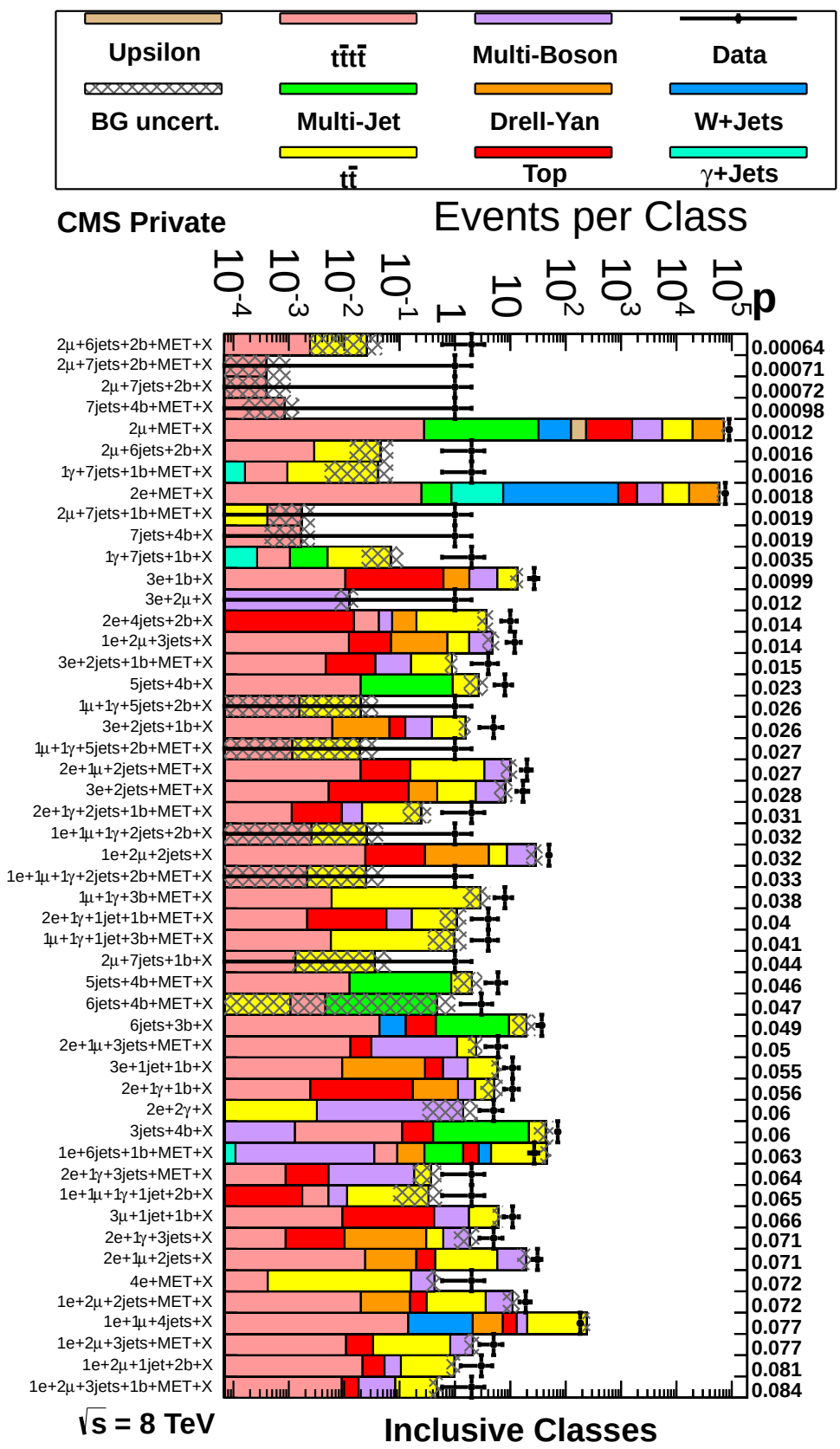


Figure 4.2: Overview of the 50 inclusive classes with the most significant p-values in an integral scan. The plot structure is explained in fig. 4.1.

4.2 Kinematic Distribution Scan of Exclusive Classes

For each kinematic variable, scan results are presented in the form of a $\tilde{p}$ distribution plot and a table of the 10 event classes exhibiting the most significant deviations between data and simulation. The $\tilde{p}$ distributions of the scans of exclusive classes for $\Sigma|p_T|$, M_{inv} and E_T^{miss} are shown in fig. 4.3, 4.4 and 4.5, respectively. The corresponding tables are tab. 4.1, 4.2 and 4.3. Note that only event classes fulfilling the criteria discussed in sec. 1.3.2.5 are considered in the scan. An excess of highly significant deviations is observed in data with respect to simulation. The classes responsible for this excess are discussed below and reveal likely simulation problems as well as BSM-like excesses in well modeled regions.

Before going into the details of the different groups of significant deviations, consider fig. 4.6, which shows the E_T^{miss}-distributions of the 2e+2jets+E_T^{miss} (a) and 1e+2b+E_T^{miss} (b) exclusive classes. Both classes are dominated by $t\bar{t}$ production and are well modeled over the measured spectrum. However, both distributions have "bumps" in data, i.e. the measured spectrum somewhat deviates from a cleanly falling shape, immediately catching the eye. While modeling problems are often at the root of observed deviations, here it is hard to make such a claim, on the one hand because data look "suspicious" in their own right, on the other hand, because the SM simulation seems to be rather "well-behaved". A possible effect of an insufficient number of simulated $t\bar{t}$ events in this region seems small compared to the observed deviation. These distributions are beautiful examples of the MUSiC algorithm working exactly as intended. Some similar findings are discussed in sec. 4.4.2.

A large group of the most significantly deviating distributions stems from event classes with very low statistics (marked in the table with the comment "low stat."). Some examples of such distributions are shown in fig. 4.7. In these distributions, the low statistics in data and simulation effectively prohibit a comparison of distribution shapes. The RoI and corresponding p-value strongly depends on single simulation events, which by migrating into neighbouring bins may significantly alter the calculation outcome. It is therefore hard to argue for a meaningful interpretation of the p-values. It would thus be favorable to include criteria in the MUSiC framework that prevent the calculation of such meaningless p-values by some kind of rebinning scheme. Such a mechanism has been discussed many times and is currently under development as part of [25]. It should however be noted that two of the "low stat."-classes would be significant even with rebinning, as can be seen from the integral scan: The 1mu+1jet+4b and 1mu+4jet+3b exclusive event classes have integral values of $N_{\text{Data}} = 20$ vs. $N_{\text{SM}} = 6.4 \pm 1.9$ and $N_{\text{Data}} = 19$ vs. $N_{\text{SM}} = 7.7 \pm 2.0$, respectively.

Fig. 4.8 shows an example of a class with relatively low statistics where shape information nevertheless reveals a significant feature in data. The RoI is centered on two events with high masses around 550 GeV, where approximately 0.03 events are expected. When taking into account the deficit of data in the preceding bins, e.g. by integrating the mass range $350 - 800$ GeV, the two observed events stand against an expectation of approximately 0.8 simulated events with a similarly large uncertainty.

Apart from the low-statistics classes, there are four event classes with significant deviations showing especially interesting features:

- $\mathbf{E_T^{miss}}$ distribution of the **1e+2jets+E_T^{miss} excl.** class (fig. 4.9 (a)): This distribution is entirely dominated by W boson production in association with jets. Independent of the SM simulation, data seem to exhibit multiple step-like features. In addition, the W+jets simulation sample systematically overestimates the amount of events at high E_T^{miss}.

- $\mathbf{E_T^{miss}}$ distribution of the **1b+E_T^{miss} excl.** class (fig. 4.9 (b)): This distribution is very similar to that of 1jet+E_T^{miss} (cf. fig. 3.6) with reduced statistics. It shows an underestimate of the high-E_T^{miss} region by the W boson production modeling, the source of which is not known. In addition, a shape is present in simulation at around $E_T^{\text{miss}} = 200$ GeV, which is only mildly pronounced in the 1jet+E_T^{miss} class. In both classes, this shape is mostly composed of events from the b enriched QCD sample (cf. sec. 2.4.1.3), while the non-b-enriched samples contribute only to the falling spectrum in the range $0 < E_T^{\text{miss}} < 100$ GeV. The additional feature is completely removed from both distributions if QCD is modeled using the Madgraph sample also studied in sec. 4.5.2,

which is not split into enriched samples. This observation draws into question the performance of the b enriched FastSim sample. While it performs well in many final states, FastSim may not provide sufficient modeling of reconstruction effects to model the 1jet+E_T^{miss}/1b+E_T^{miss} final state, which is especially sensitive to experimental effects. However, the feature is only present locally, contains a small number of events and shows large uncertainties.

- $\mathbf{M_{inv}}$ distribution of the **1e+1jet+2b excl.** class (fig. 4.10 (a)): In this class, a bump-like structure is found at around $M_{\text{inv}} \approx 1\,\text{TeV}$. The dominating $t\bar{t}$ spectrum contribution seems to be falling significantly more quickly than the data, leading to sizable observed excesses in data. However, in the RoI, it may be arguable that the data are showing - independent of the simulation - a bump-like feature. The possibility of systematic $t\bar{t}$ mismodeling is further discussed in sec. 4.4.2.

- $\mathbf{\Sigma|p_T|}$ distribution of the **1e+4jets excl.** class (fig. 4.10 (b)): In the low-$\Sigma|p_T|$ regions of this distribution, multiple modeling problems collide. The balance issues related to jet triggering raised in sec. 4.5 also endanger modeling quality here. Additionally, QCD simulation samples contribute in a "spiky" way, i.e. neighbouring bins may have unphysically large differences in content because of insufficient numbers of generated events. Although sample statistics are not per se low for the QCD EM enriched sample, which is dominating in this region, the requirement of a reconstructed isolated electron strongly reduces the number of contributing events. In EM enriched QCD events, electrons will only seldomly be sufficiently isolated from other PF candidates to be reconstructed as isolated leptons. A third factor is the influence of the W+jets process, which is not among the most reliably modeled processes.

In summary, the kinematic scans of exclusive event classes reveal a number of interesting features in data and illustrate MUSiC's capability for bump detection. The results also identify insufficient simulation performance for W boson production. The W+jets modeling has been a long-standing source of discrepancies. Because of the large number of available samples, which are binned in different kinematic variables, it is difficult to find a complete and non-overlapping sample composition (cf. sec. 2.4.1). The mis-modeling may be related to higher order perturbative corrections, which in studies of events with single leptons and E_T^{miss} have been found to depend the interaction kinematics [97]. Such an effect would not we covered by the flat k-factor used in this analysis (cf. sec. 2.4.1.1). Future study is warranted by these results.

A smaller modeling issue is observed in the used b enriched FastSim QCD sample, which is however less systematic and only occurs locally. It highlights that while FastSim is useful to fill holes in the available simulation set, it cannot completely replace FullSim samples. However, the satisfactory performance of FastSim in many final states should be considered before discarding it altogether. Compared to the alternatively available FullSim sample, which is so small as to be practically useless, the FastSim sample is still favorable, even if it shows minor defects.

Table 4.1: List of the 10 event classes with the most significant deviations found in the scan of the exclusive $\Sigma|p_T|$-distributions.

Event Class	$N_{\mathbf{Data}}$	$N_{\mathbf{MC}} \pm \sigma_{\mathbf{MC}}$	$\tilde{p}$	**Comment**
2mu+E_T^{miss}	3.1638×10^4	$(2.32 \pm 0.15) \times 10^4$	$< 1.0 \times 10^{-5}$	DY E_T^{miss} issue
2e+E_T^{miss}	2.7121×10^4	$(1.92 \pm 0.15) \times 10^4$	1.10×10^{-4}	DY E_T^{miss} issue
1e+4jet	7.40×10^2	$(4.3 \pm 0.6) \times 10^2$	4.50×10^{-4}	W+jets
1mu+1jet+4b	1.9×10^1	(5.5 ± 1.7)	1.16×10^{-2}	Low stat.
3e+2jet+E_T^{miss}	4	$(2.9 \pm 1.5) \times 10^{-1}$	2.06×10^{-2}	Low stat.
1e+2mu+1b	5	$(5.0 \pm 2.5) \times 10^{-1}$	2.56×10^{-2}	Low stat.
1mu+4jet+3b	9	(1.5 ± 0.7)	2.76×10^{-2}	Low stat.
2gam+2jet	4	$(2.2 \pm 2.4) \times 10^{-1}$	2.77×10^{-2}	Low stat.
1mu+1jet+1b	1.2305×10^4	$(9.0 \pm 0.9) \times 10^3$	3.26×10^{-2}	
2e+1mu+1b+E_T^{miss}	5	$(6.0 \pm 3.3) \times 10^{-1}$	3.29×10^{-2}	Low stat.

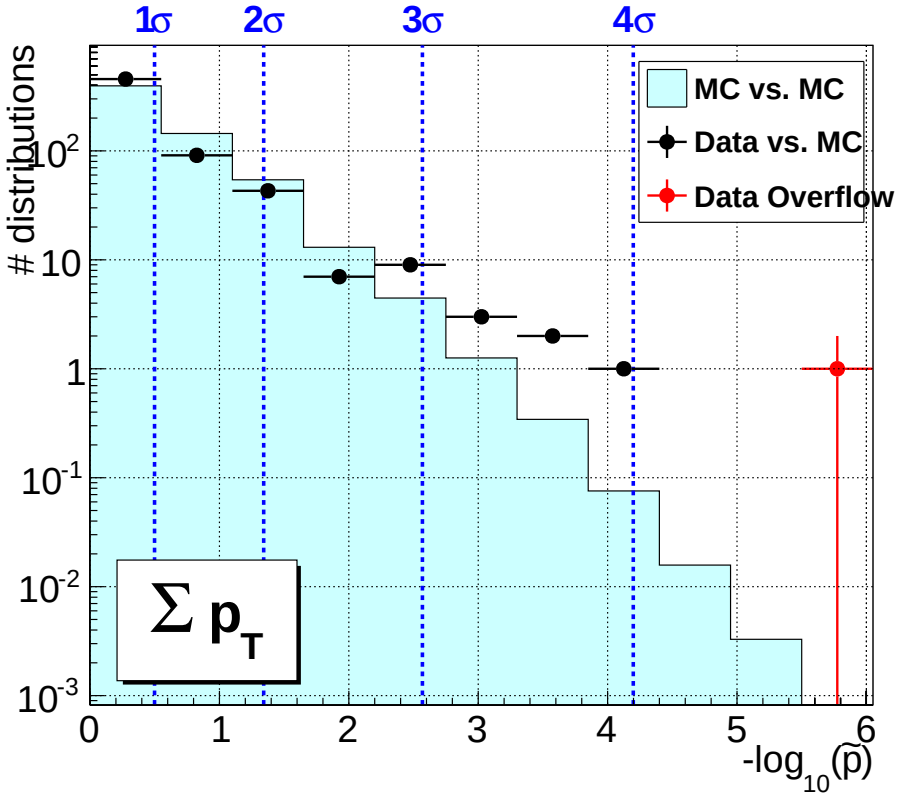


Figure 4.3: Distribution of $\tilde{p}$ in a scan of the $\Sigma|p_T|$-distributions of exclusive event classes. In this axis scaling, significance increases to the right. For reference, the dashed blue lines mark $\tilde{p}$-values that are equivalent to integer number of Gaussian sigmas.

Table 4.2: List of the 10 event classes with the most significant deviations found in the scan of the exclusive M_{inv}-distributions.

Event Class	$N_{\mathbf{Data}}$	$N_{\mathbf{MC}} \pm \sigma_{\mathbf{MC}}$	$\tilde{p}$	Comment
2mu+E_T^{miss}	2.5975×10^4	$(1.88 \pm 0.12) \times 10^4$	$< 1.0 \times 10^{-5}$	DY E_T^{miss} issue
2e+E_T^{miss}	1.2188×10^4	$(8.1 \pm 0.7) \times 10^3$	$< 1.0 \times 10^{-5}$	DY E_T^{miss} issue
1e+1mu+1b	1.1×10^1	(1.0 ± 1.0)	5.88×10^{-3}	cf. sec. 4.4.2
1e+1jet+2b	2.11×10^2	$(1.09 \pm 0.21) \times 10^2$	8.85×10^{-3}	
2mu+1jet	1.663×10^3	$(1.07 \pm 0.15) \times 10^3$	9.83×10^{-3}	DY E_T^{miss} issue
1mu+1jet+4b	5	$(2.9 \pm 2.3) \times 10^{-1}$	1.24×10^{-2}	low stat.
2mu+1b	2.95×10^2	$(1.65 \pm 0.32) \times 10^2$	1.39×10^{-2}	DY E_T^{miss} issue
3e+2jet+E_T^{miss}	4	$(2.8 \pm 1.4) \times 10^{-1}$	1.83×10^{-2}	low stat.
2mu+4jet+1b	5	$(2.1 \pm 1.8) \times 10^{-1}$	1.91×10^{-2}	low stat.
1e+1mu+1gam	2	$(2.7 \pm 2.1) \times 10^{-2}$	2.35×10^{-2}	

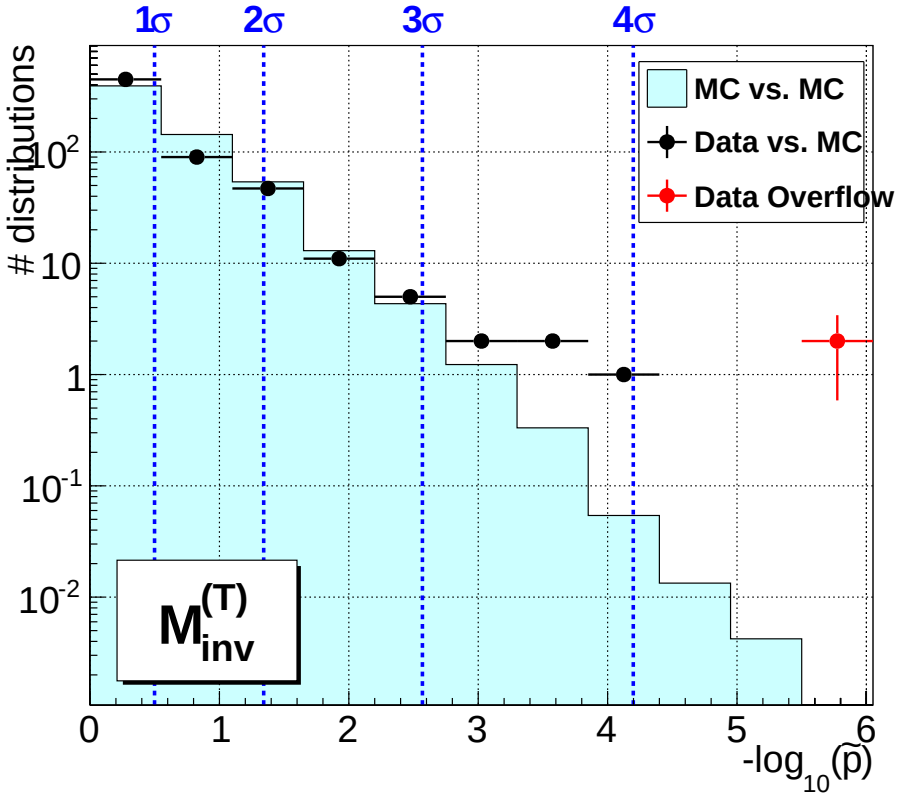


Figure 4.4: Distribution of $\tilde{p}$ in a scan of the M_{inv}-distributions of exclusive event classes.

Table 4.3: List of the 10 event classes with the most significant deviations found in the scan of the exclusive E_T^{miss}-distributions.

Event Class	N_{Data}	$N_{\text{MC}} \pm \sigma_{\text{MC}}$	$\tilde{p}$	**Comment**
2mu+E_T^{miss}	1.4622×10^4	$(1.07 \pm 0.07) \times 10^4$	1.00×10^{-5}	DY E_T^{miss} issue
2e+E_T^{miss}	1.3283×10^4	$(9.3 \pm 0.9) \times 10^3$	1.60×10^{-4}	DY E_T^{miss} issue
1e+2jet+E_T^{miss}	4.7×10^1	$(1.05 \pm 0.11) \times 10^2$	8.30×10^{-3}	
3e+2jet+E_T^{miss}	5	$(5.0 \pm 2.3) \times 10^{-1}$	9.21×10^{-3}	
2e+2jet+E_T^{miss}	1.0×10^1	(2.2 ± 0.8)	3.46×10^{-2}	
2e+1jet+3b+E_T^{miss}	6	(1.1 ± 0.5)	4.13×10^{-2}	
3jet+4b+E_T^{miss}	1.2×10^1	(3.3 ± 1.4)	4.27×10^{-2}	
1b+E_T^{miss}	1.28×10^2	$(7.3 \pm 1.4) \times 10^1$	6.71×10^{-2}	
1e+1jet+1b+E_T^{miss}	2.6×10^1	$(1.06 \pm 0.19) \times 10^1$	7.13×10^{-2}	
1e+2b+E_T^{miss}	3.3×10^1	$(1.55 \pm 0.31) \times 10^1$	7.60×10^{-2}	

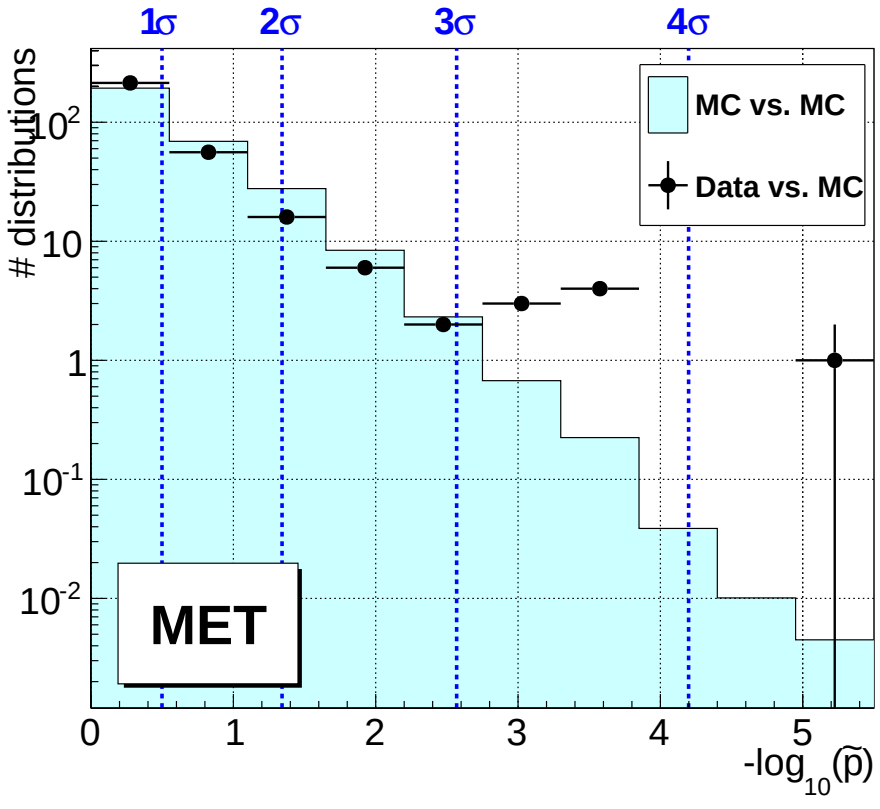
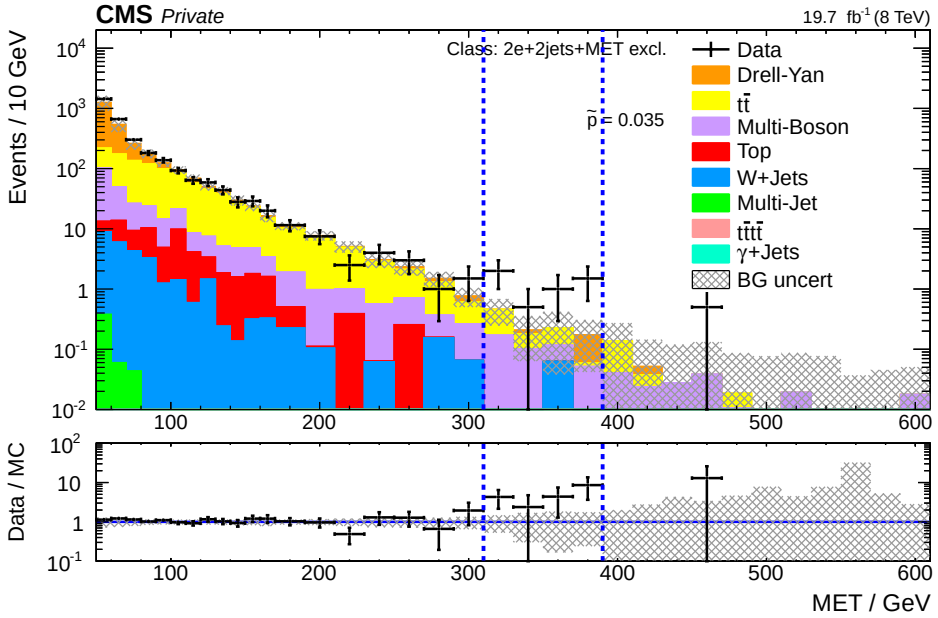
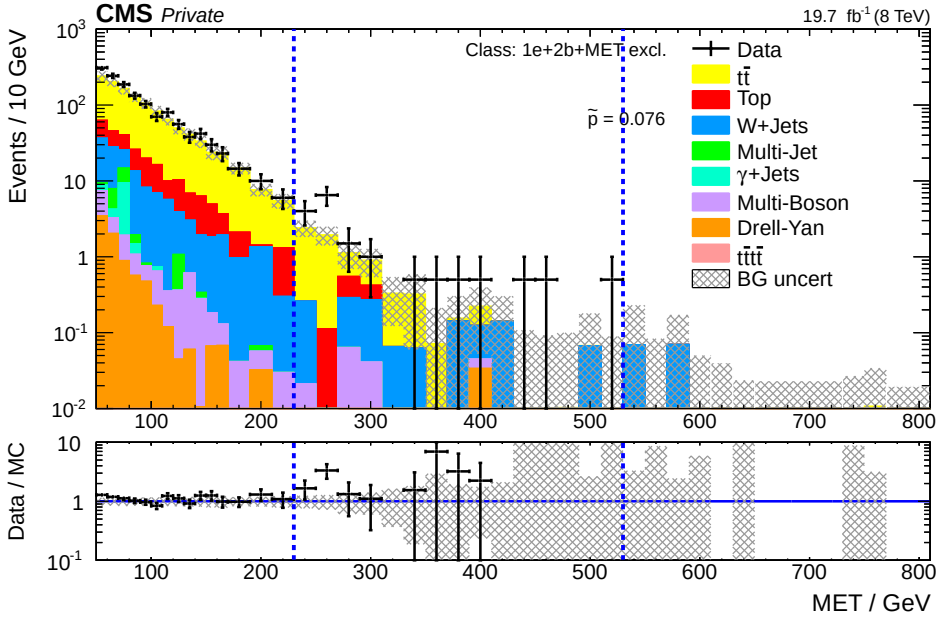


Figure 4.5: Distribution of $\tilde{p}$ in a scan of the E_T^{miss}-distributions of exclusive event classes.

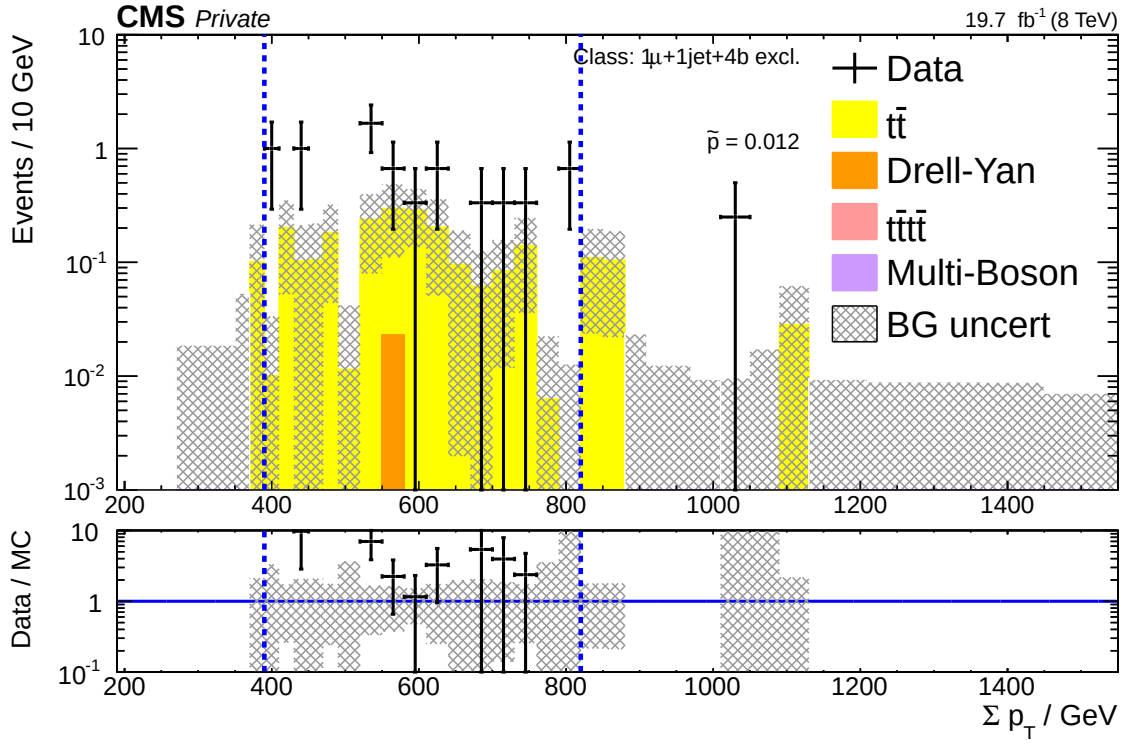


(a)

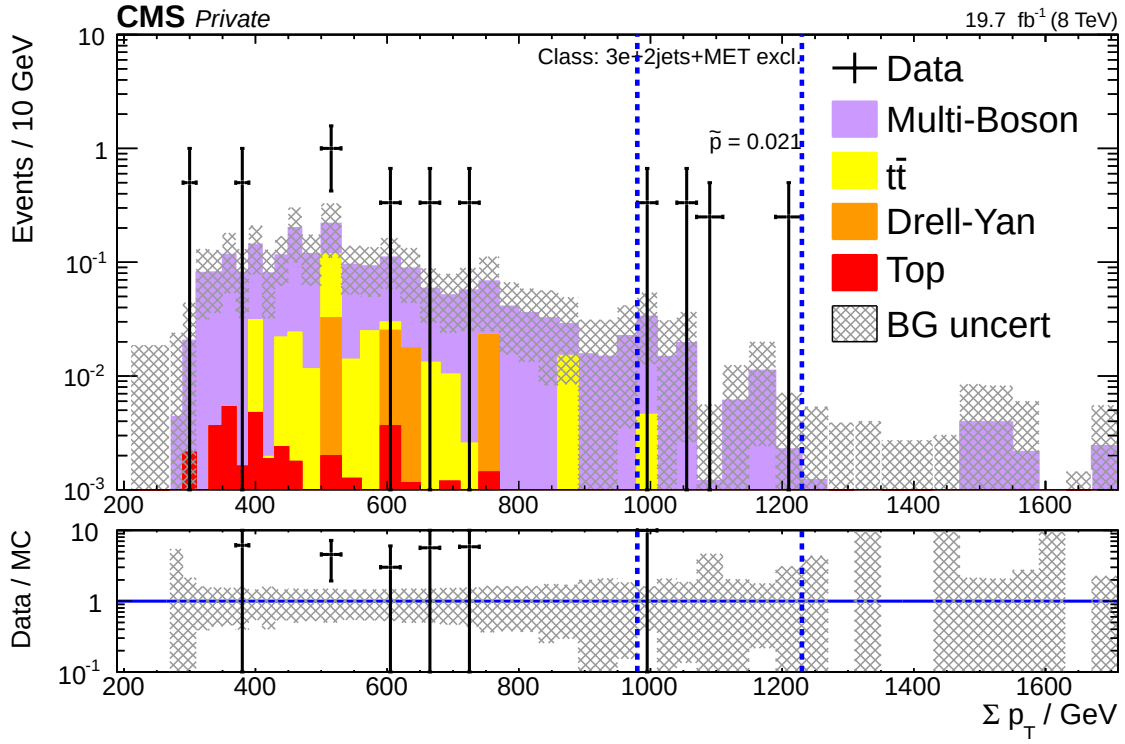


(b)

Figure 4.6: Examples of E_T^{miss}-distributions of exclusive event classes where the RoI-algorithm finds bump-like features in data over well modeled background: 2e+2jets+E_T^{miss} (a) and 1e+2b+E_T^{miss} (b).



(a)



(b)

Figure 4.7: Examples of $\Sigma|p_T|$-distributions of exclusive event classes with low statistics: 1μ+1jet+4b (a) and 3e+2jets+E_T^{miss} (b). In these classes, low statistics prohibit a meaningful shape comparison.

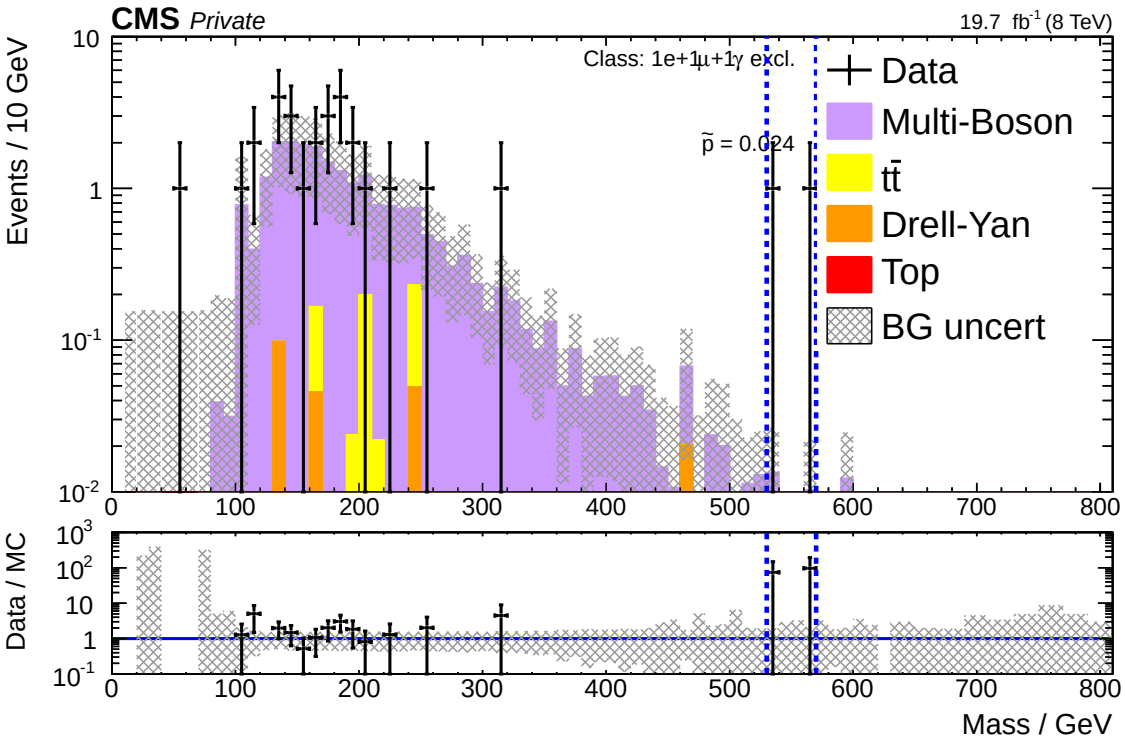
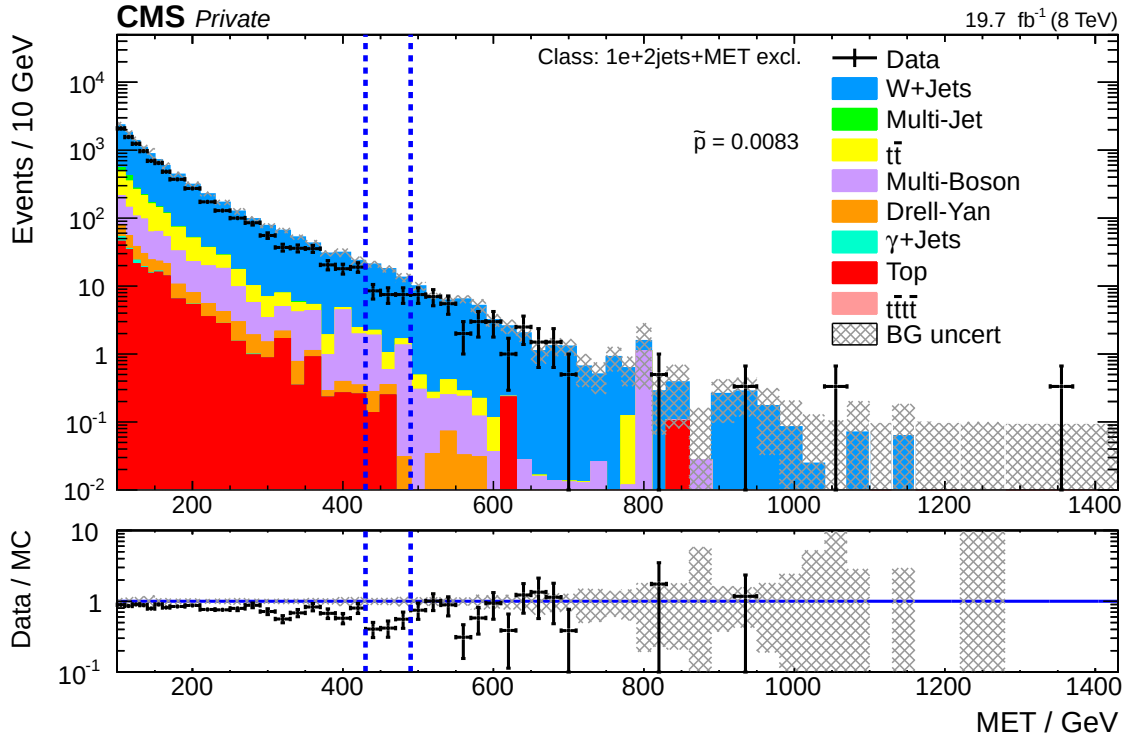
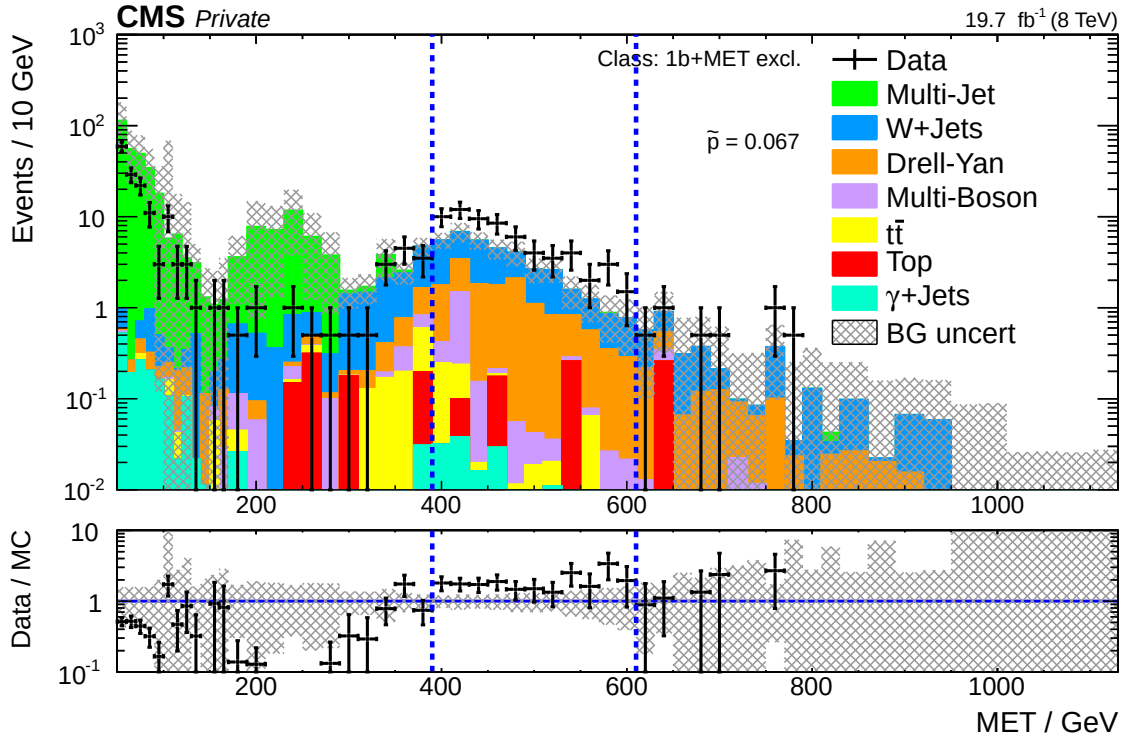


Figure 4.8: An example of a low-statistics distribution still containing some degree of shape information: the M_{inv}-distribution of the 1e+1μ+1γ exclusive class. While the expected number of events is low throughout the distribution, the two outlier events contained in the RoI seem to be well separated from any MC contributions. This piece of information would not be visible in an integral scan of the distribution.

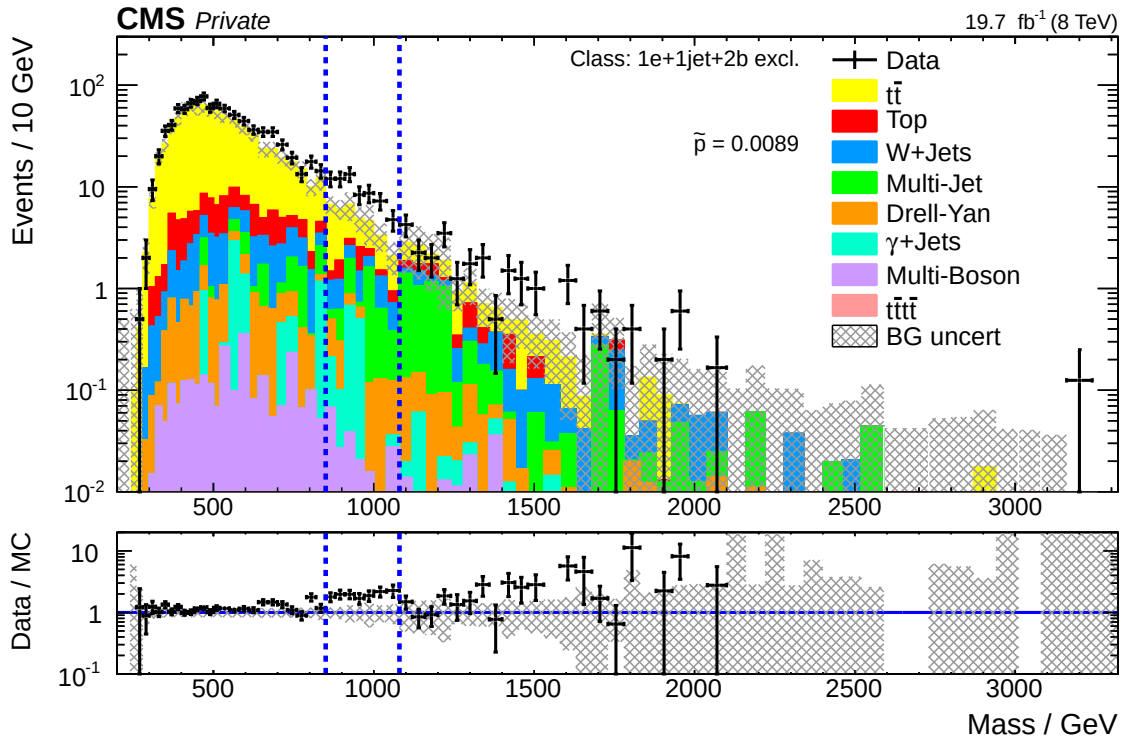


(a)

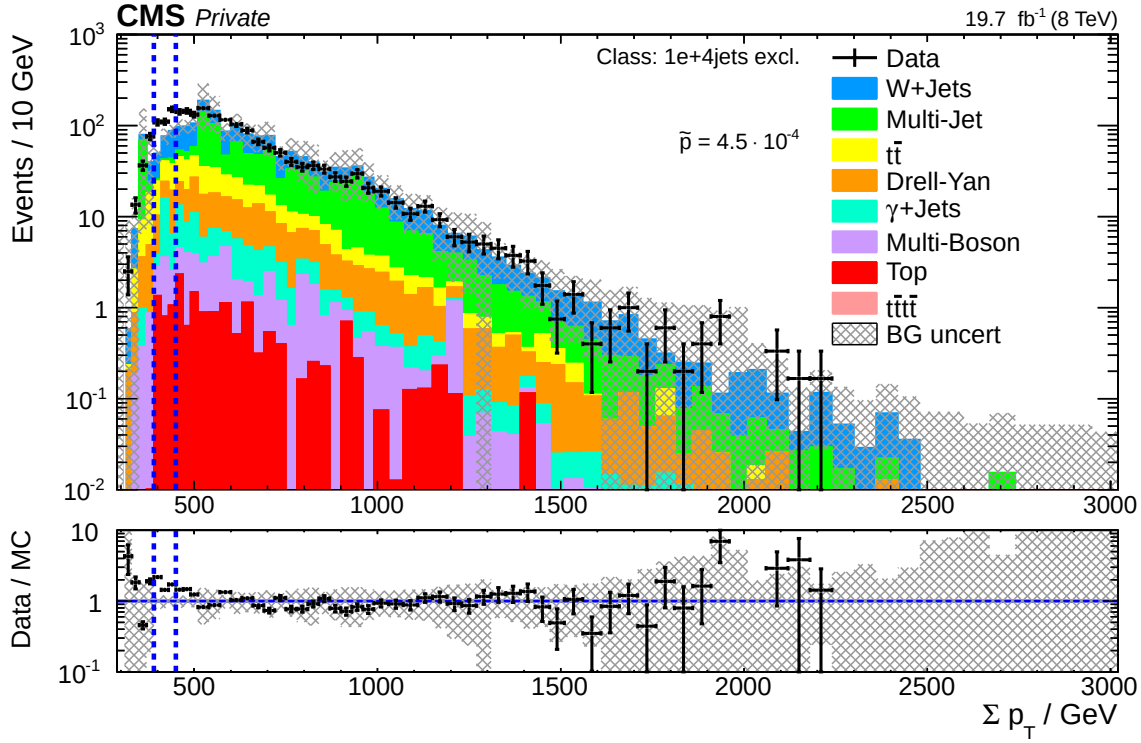


(b)

Figure 4.9: Two of the ten most significantly deviating E_T^{miss}-distributions of exclusive event classes discussed: 1e+2jet+E_T^{miss} (a) and 1b+E_T^{miss} (b).



(a)



(b)

Figure 4.10: Two of the most significantly deviating kinematic distributions of exclusive event classes: M_{inv}-distribution of the 1e+1jet+2b exclusive class (c) and $\Sigma|p_T|$-distribution of the 1e+4jet exclusive class (d).

4.3 Kinematic Distribution Scan of Inclusive Classes

The $\tilde{p}$ distributions of the scans of exclusive classes for $\Sigma|p_T|$, M_{inv} and E_T^{miss} are shown in fig. 4.11, 4.12 and 4.13, respectively. Details of the event classes with the most significant deviations are given in tab. 4.4, 4.5 and 4.6. Note that only event classes fulfilling the criteria discussed in sec. 1.3.2.5 are considered in the scan.

More significant deviations are observed in data than expected from simulation. As in the scans of the exclusive event classes, a number of very low statistics classes are among the most significantly deviating. In the scan of E_T^{miss}-distributions six of the ten most significant classes have very low statistics, with data counts in the RoI that are very close to the distribution integral. A large part of the most significant inclusive classes result from single rare events. The single events in the 2μ+7jet+2b+E_T^{miss} and 7jet+4b final states give rise to a total of six of the most significant deviations in $\Sigma|p_T|$, another six in M_{inv} and four in E_T^{miss}. These rare events are discussed in detail in sec. 4.4.3. A very interesting excess is observed in the M_{inv}-distribution of the 1e+1mu+1b inclusive event class. This excess is discussed in detail in sec. 4.4.2.

Table 4.4: List of the 10 event classes with the most significant deviations found in the scan of the inclusive $\Sigma|p_T|$-distributions.

Event Class	N_{Data}	$N_{\mathrm{MC}} \pm \sigma_{\mathrm{MC}}$	$\tilde{p}$	Comment
2mu+E_T^{miss} +X	5.76930×10^4	$(4.41 \pm 0.29) \times 10^4$	1.70×10^{-4}	DY E_T^{miss} issue
2mu+7jet+2b+X	1	$(0.0 \pm 3.3) \times 10^{-5}$	2.70×10^{-4}	Exotic event
2mu+7jet+2b+E_T^{miss} +X	1	$(0.0 \pm 3.5) \times 10^{-5}$	3.40×10^{-4}	Exotic event
2mu+7jet+1b+E_T^{miss} +X	1	$(0 \pm 6) \times 10^{-5}$	6.50×10^{-4}	Exotic event
7jet+4b+X	1	$(0.0 \pm 1.1) \times 10^{-4}$	9.50×10^{-4}	Exotic event
7jet+4b+E_T^{miss} +X	1	$(2.0 \pm 2.3) \times 10^{-4}$	1.14×10^{-3}	Exotic event
1mu+1gam+3b+X	7	$(5 \pm 4) \times 10^{-1}$	1.48×10^{-3}	Low stat.
2e+E_T^{miss} +X	4.08200×10^4	$(3.04 \pm 0.24) \times 10^4$	1.54×10^{-3}	DY E_T^{miss} issue
2mu+6jet+2b+E_T^{miss} +X	2	$(2.3 \pm 2.5) \times 10^{-2}$	2.73×10^{-3}	Exotic event
1gam+7jet+1b+E_T^{miss} +X	2	$(1.7 \pm 2.5) \times 10^{-2}$	3.88×10^{-3}	Low stat.

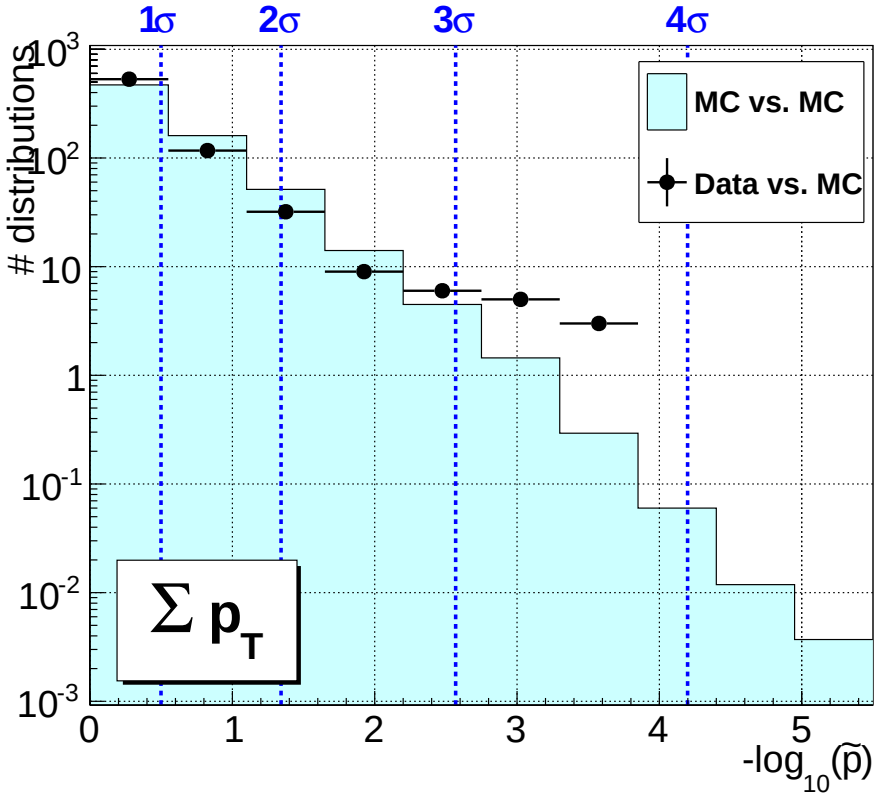


Figure 4.11: Distribution of $\tilde{p}$ in a scan of the $\Sigma|p_T|$-distributions of inclusive event classes.

Table 4.5: List of the 10 event classes with the most significant deviations found in the scan of the inclusive M_{inv}-distributions.

Event Class	$N_{\mathbf{Data}}$	$N_{\mathbf{MC}} \pm \sigma_{\mathbf{MC}}$	$\tilde{p}$	**Comment**
2mu+E_T^{miss} +X	3.2472×10^4	$(2.43 \pm 0.17) \times 10^4$	1.10×10^{-4}	DY E_T^{miss} issue
1e+1mu+1b+X	8.4×10^1	$(3.5 \pm 0.6) \times 10^1$	1.70×10^{-4}	Cf. sec. 4.4.2
2mu+7jet+2b+X	1	$(0.0 \pm 2.4) \times 10^{-5}$	4.90×10^{-4}	Exotic event
2mu+7jet+2b+E_T^{miss} +X	1	$(0.0 \pm 2.8) \times 10^{-5}$	5.10×10^{-4}	Exotic event
7jet+4b+E_T^{miss} +X	1	$(0.0 \pm 1.1) \times 10^{-4}$	7.10×10^{-4}	Exotic event
2mu+7jet+1b+E_T^{miss} +X	1	$(0 \pm 5) \times 10^{-5}$	7.40×10^{-4}	Exotic event
2e+E_T^{miss} +X	3.9173×10^4	$(2.92 \pm 0.24) \times 10^4$	2.41×10^{-3}	DY E_T^{miss} issue
2mu+6jet+2b+E_T^{miss} +X	2	$(2.3 \pm 2.5) \times 10^{-2}$	2.72×10^{-3}	Exotic event
7jet+4b+X	1	$(3.5 \pm 3.4) \times 10^{-4}$	2.74×10^{-3}	Exotic event
1mu+1gam+3jet+X	5	$(0.2 \pm 2.0) \times 10^{-1}$	2.79×10^{-3}	

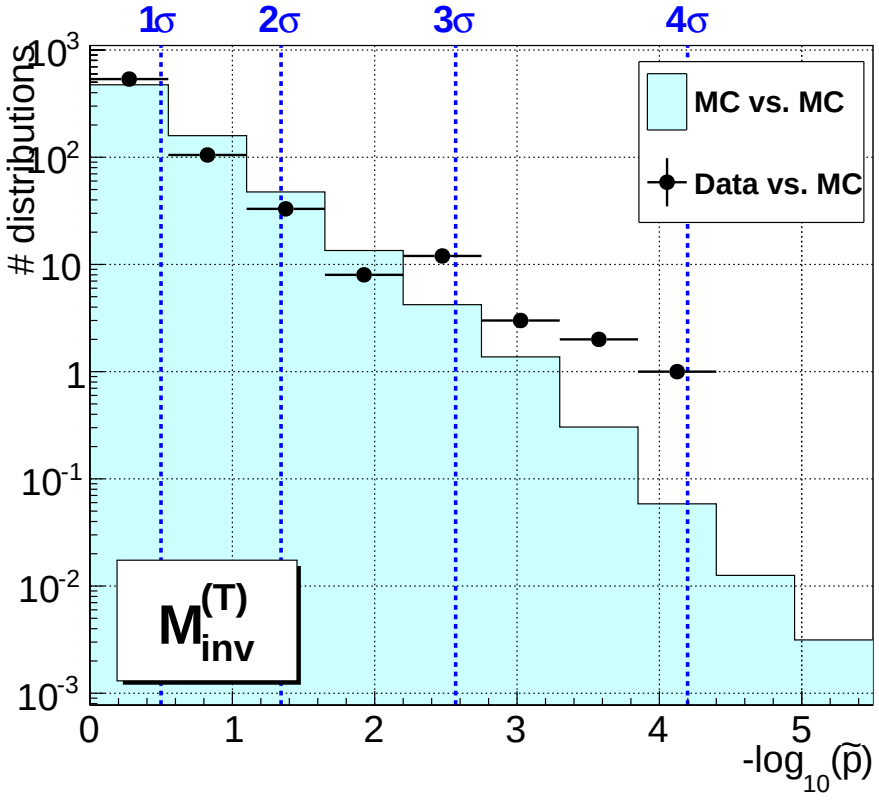


Figure 4.12: Distribution of $\tilde{p}$ in a scan of the M_{inv}-distributions of inclusive event classes.

Table 4.6: List of the 10 event classes with the most significant deviations found in the scan of the inclusive E_T^{miss}-distributions.

Event Class	N_{Data}	$N_{\text{MC}} \pm \sigma_{\text{MC}}$	$\tilde{p}$	Comment
1mu+1gam+3b+E_T^{miss} +X	5	$(0.5 \pm 1.1) \times 10^{-1}$	8.00×10^{-5}	Low stat.
2mu+7jet+2b+E_T^{miss} +X	1	$(0 \pm 5) \times 10^{-5}$	1.80×10^{-4}	Exotic event
2mu+6jet+2b+E_T^{miss} +X	1	$(0.0 \pm 1.0) \times 10^{-4}$	4.60×10^{-4}	Exotic event
7jet+4b+E_T^{miss} +X	1	$(0.0 \pm 1.5) \times 10^{-4}$	5.30×10^{-4}	Exotic event
2mu+7jet+1b+E_T^{miss} +X	1	$(2 \pm 7) \times 10^{-5}$	6.50×10^{-4}	Exotic event
1mu+1gam+5jet+2b+E_T^{miss} +X	1	$(1.3 \pm 2.1) \times 10^{-4}$	7.80×10^{-4}	Low stat.
1gam+7jet+1b+E_T^{miss} +X	2	$(1.7 \pm 2.7) \times 10^{-2}$	1.96×10^{-3}	Low stat.
1e+1mu+1gam+2jet+2b+E_T^{miss} +X	1	$(1 \pm 8) \times 10^{-3}$	5.53×10^{-3}	Low stat.
1mu+1gam+1jet+3b+E_T^{miss} +X	3	$(0.4 \pm 1.2) \times 10^{-1}$	6.34×10^{-3}	Low stat.
2e+1gam+2jet+E_T^{miss} +X	4	$(2.3 \pm 2.9) \times 10^{-1}$	1.68×10^{-2}	Low stat.

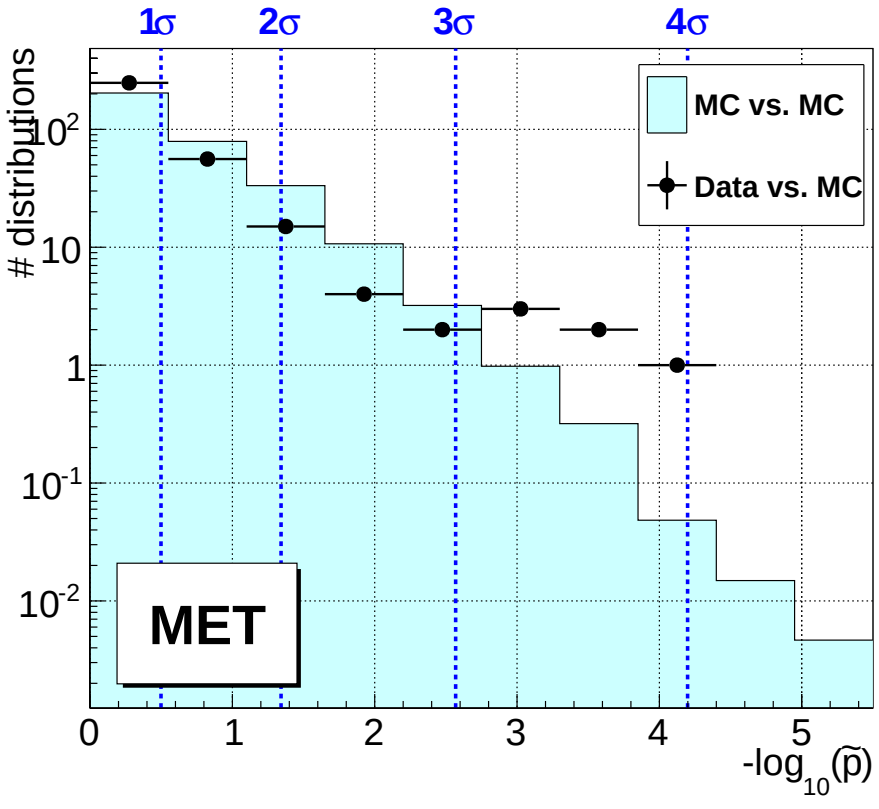


Figure 4.13: Distribution of $\tilde{p}$ in a scan of the E_T^{miss}-distributions of inclusive event classes.

4.4 Discussion of Selected Final States

4.4.1 Dilepton Classes

In all scanning results, dilepton classes produce the most significant observed deviations between data and simulation. The most striking examples are the 2e+E_T^{miss} and 2μ+E_T^{miss} classes, which are ranked first in deviation significance for the integral scans as well as the scans of all kinematic distributions. In the scan of the M_{inv}-distribution, the 2μ+1jet and 2μ+1b classes are also found to contain large discrepancies[2]. In these classes, the E_T^{miss}-spectrum is not correctly modeled in simulation. Especially the influence of *Unclustered Energy*, i.e. energy deposits not linked to any PF candidate, seems to have an impact in the corresponding regions, possibly also influencing final states such as 2e+1jet, although they do not nominally contain any E_T^{miss}. Uncertainties associated with Unclustered Energy are not treated at this time, but are in preparation for future iterations of the analysis. Discussion of the phenomenon is postponed to the work of [25].

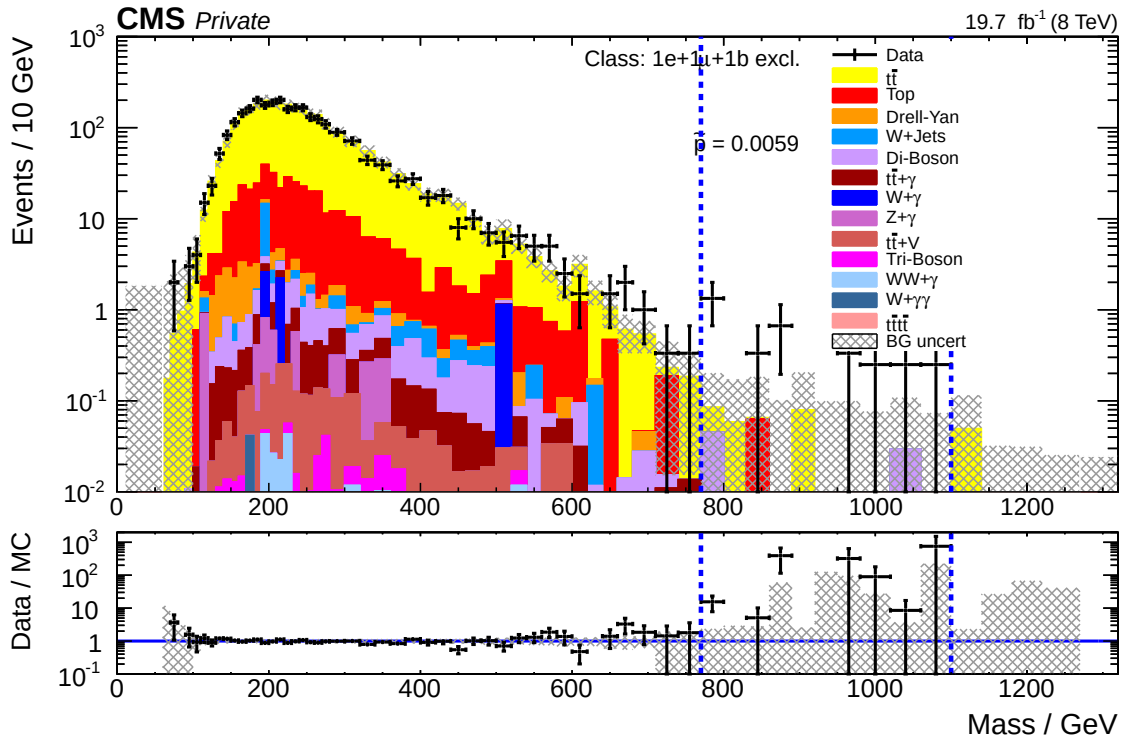
4.4.2 1e+1μ+1b Final State

The 1e+1μ+1b exclusive event class contains the third most significant deviation in the scan of M_{inv}-distributions of exclusive classes. The M_{inv}- and $\Sigma|p_T|$-distributions of this class are shown in fig. 4.14. The final state composition of two leptons and one b jet predominantly favors processes involving top quarks. Over large ranges of both kinematic distributions, shape and normalisation are well modeled. With the exception of the low and high tail regions, where Fill-Up and MC statistics uncertainty are large, the uncertainties are dominated by the 10 % NNLL cross-section uncertainty of the top-pair MC sample and the flat 5 % b-tagging efficiency uncertainty (cf. sec. 2.4.4.1 and 2.4.4.2). In the high-mass region (fig. 4.14 (a)), a significant excess of $N_{\text{Data}} = 11$ data events compared to $N_{\text{SM}} = 1 \pm 1$ expected events is observed. Although the predictive power of the SM simulation in this region is limited by the number of simulated events, the deviation of data and SM behavior is evident due to the bump-like shape of the data, which does not follow the steeply falling shape of the bulk spectrum. The common phenomenon in low-statistics high-energy tails, where simulation and data contributions are scattered across multiple bins, with empty bins in between, which fools the naked eye into suspecting shape differences where there are none, is also not present here: The RoI found by the algorithm is practically identical in content to an overflow bin, below which data and simulation agree well. The simulation contributions at masses higher than the upper boundary of the RoI are negligible compared to the large difference $N_{\text{Data}} - N_{\text{SM}}$. An excess is also observed in the high-energy tail of the $\Sigma|p_T|$-distribution[3] (fig. 4.14 (b)). However, with $N_{\text{Data}} = 3$ and $N_{\text{SM}} = 0.4 \pm 0.4$, the excess is approximately one order of magnitude less significant in $\tilde{p}$. Considering the three close-by bins, where events are expected but not observed, and the not completely negligible SM expectation to the right of the RoI, the "overflow bin"-argumentation cannot plausibly be used.

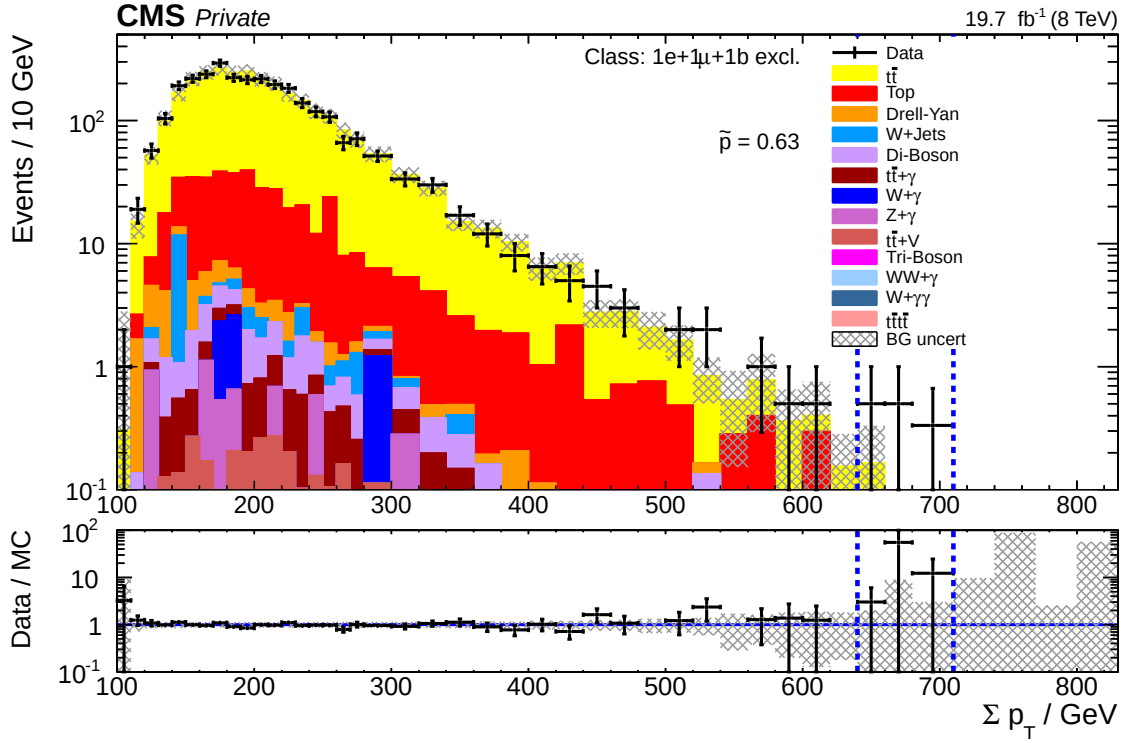
In the inclusive 1e+1μ+1b+X event class, an excess is observed in the same mass region as in the exclusive class with $N_{\text{Data}} = 84$ compared to $N_{\text{SM}} = 35 \pm 6$, yielding $\tilde{p} = 1.7 \times 10^{-4}$, the second most significant deviation in the scan of the M_{inv}-distributions of inclusive classes. In the M_{inv}-distribution (fig. 4.15 (a)), a bump-like structure is visible at approximately 700 GeV. The dominant top quark pair production process seems to provide sufficient simulation statistics to be reliable even in the high-mass tail. Independent of SM simulation, a bump-like discrepancy from a continually falling spectrum shape is evident. It seems counterintuitive that the RoI found by the algorithm does not contain the beginning of the bump, i.e. the three bins just below the lower boundary of the RoI, which show a visible excess in data. However, adding these bins to the RoI increases the SM uncertainty by a larger

[2] The similarity between the deviations in the 2μ + 1jet and 2μ + 1b classes is due to their being dominated by ISR gluon jets. For the content of the RoI as well as the full integral over both classes, the ratio 2μ + 1b/2μ + 1jet is approximately 2 %, which is a reasonable mis-tagging rate for gluon jets. In their dominating SM composition, the two classes are thus not actually different, but rather quasi-randomly chosen subsets of the same physics process. It is thus rather expected than surprising that not only one, but both classes show up in the list of most significant deviations.

[3] It is of course imprecise to refer to the high-$\Sigma|p_T|$ region as "high-energy", but simplifies the terminology.

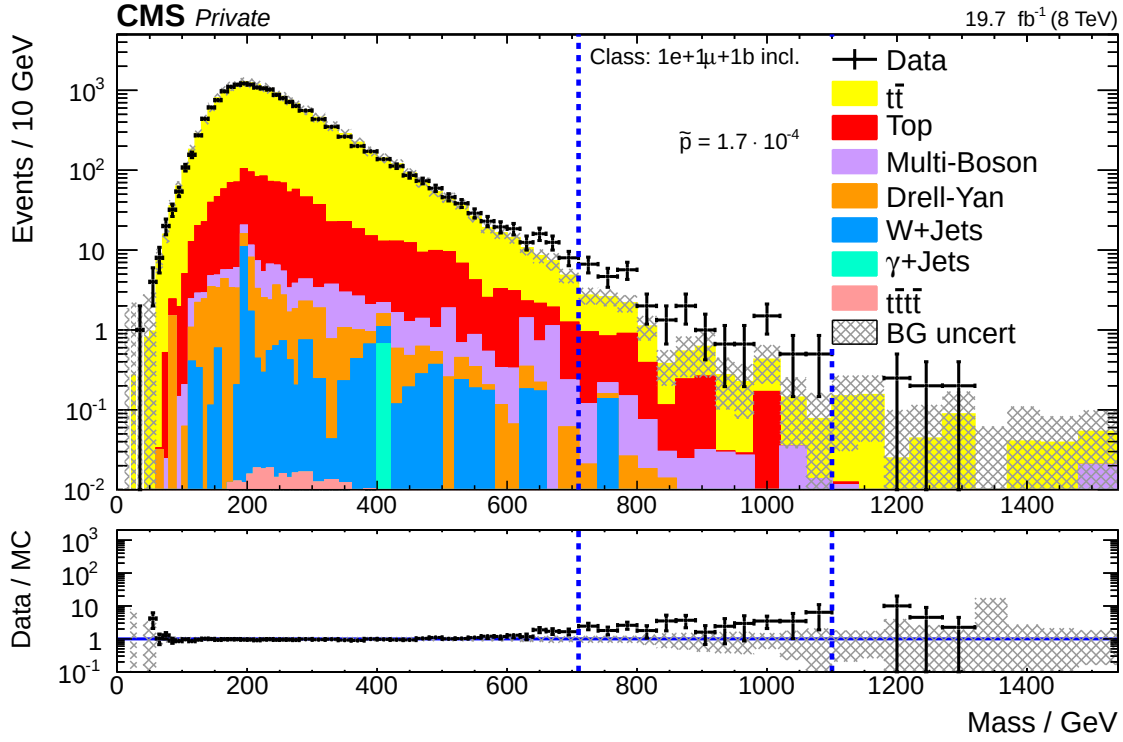


(a)

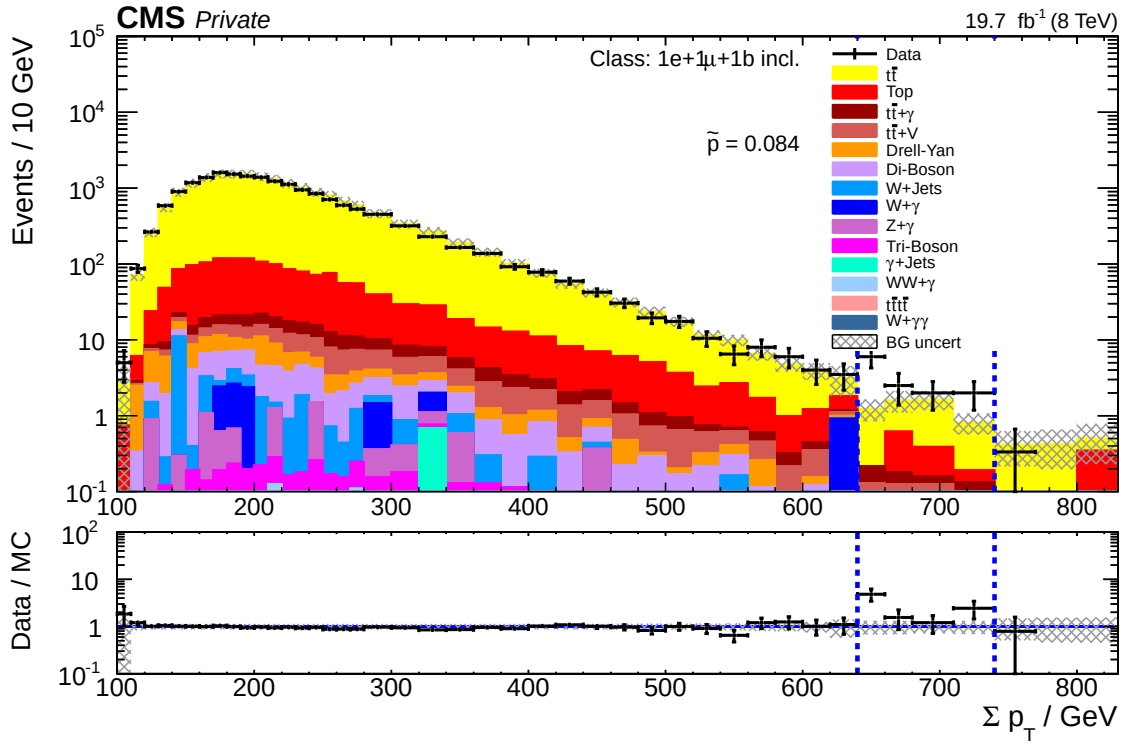


(b)

Figure 4.14: M_{inv}-(a) and $\Sigma|p_T|$-distributions (b) of the 1e+1mu+1b excl. event class. Over large ranges of the spectrum, the SM simulation performs well. In both distributions, the RoI algorithm finds an excess in the tail regions, which tends to be more significant in the M_{inv}-distribution.

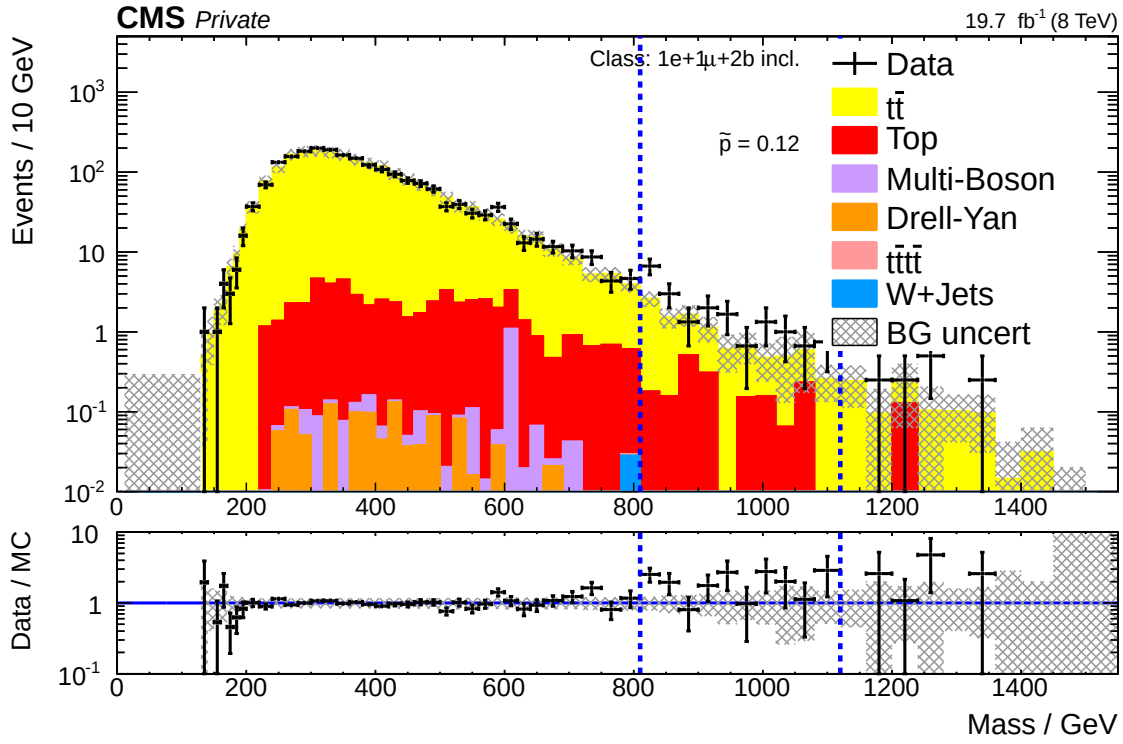


(a)

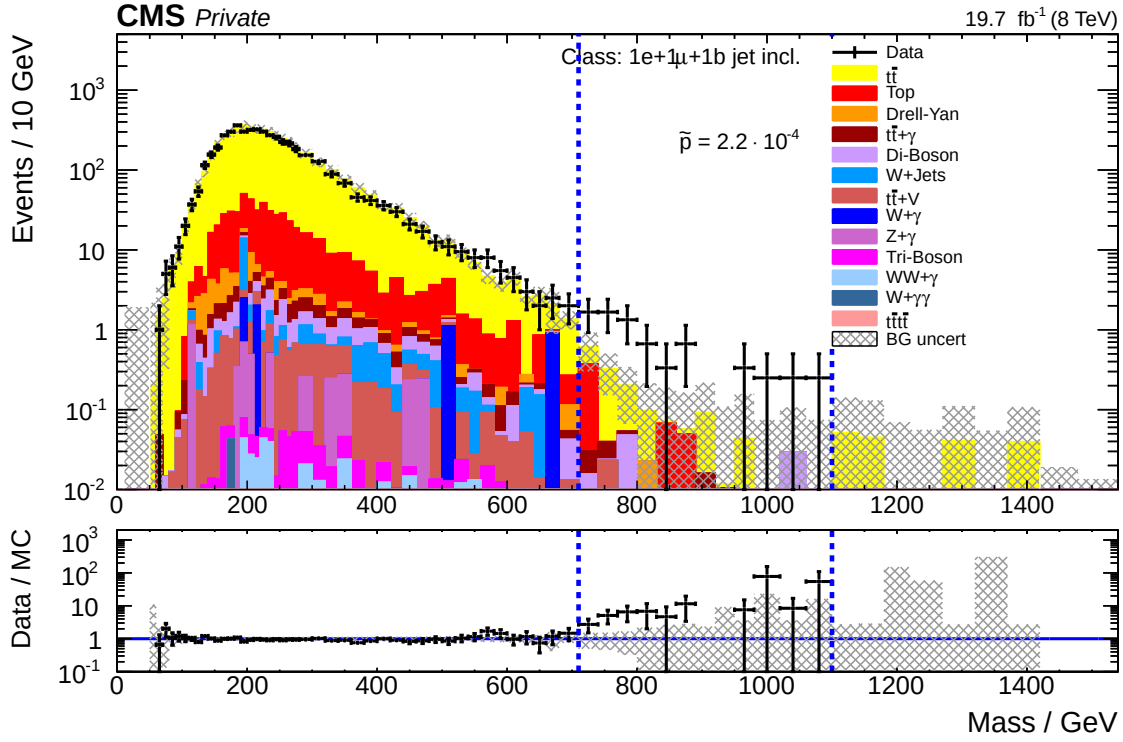


(b)

Figure 4.15: M_{inv}-(a) and $\Sigma|p_T|$-distributions (b) of the 1e+1μ+1b incl. event class. Over large ranges of the spectrum, the SM simulation performs well. As in the corresponding exclusive classes (cf. fig. 4.14), tail excesses are found, with greater significance for the M_{inv}-distribution. In both kinematic variables, the significance is enhanced in the inclusive event classes, where it yields the second most significant $\tilde{p}$-value in the scan of the M_{inv}-distributions of the inclusive classes.



(a)



(b)

Figure 4.16: M_{inv}-distributions of the 1e+1mu+2b incl. (a) and 1e+1mu+1b+Njet (b) classes. These classes contain exclusive subsets of the events in the 1e+1mu+1b event class. The RoI contents are ($N_{\text{Data}} = 58$ vs. $N_{\text{SM}} = 30.5 \pm 6.0$) (a) and ($N_{\text{Data}} = 23$ vs. $N_{\text{SM}} = 4.7 \pm 1.6$) (b), which are in sum roughly equivalent to the RoI content of the 1e+1mu+2b incl. M_{inv}-distribution. The location of the RoI is slightly shifted to higher M_{inv} in (a) because of the additional b jet contributing to the M_{inv} calculation.

amount than the difference between data and simulation, and thus yields less significant p-values.[4] As in the exclusive case, a less significant deviation is also present in the $\Sigma|p_T|$-distribution fig. 4.15 (b), with an observed value of $N_{\text{Data}} = 29$ compared to the expected value of $N_{\text{SM}} = 13.1 \pm 2.6$.

Since the inclusive event class contains events with many different exclusive final states[5], it is not obvious what final states the deviation is observable in. Since all final state subsets of a given event class are also available as inclusive event classes, the origin of the bump can be isolated. The inclusive event classes with added photon or lepton do not contain a bump. The inclusive classes with added jets and b jets are shown in fig. 4.16. In both classes, an excess is found in a mass region similar to that of the 1e+1μ+1b inclusive class. In the 1e+1μ+2b inclusive class, the RoI is shifted to a somewhat higher mass because of the additional b jet contributing to the mass calculation. Although the union of these two classes is not equivalent to the 1e+1μ+1b class, their combined contributions ($N_{\text{Data}} = 58\,(23)$ for 1e+1μ+2b incl. (1e+1μ+1b jet incl.)) are similar to the event count of the excess region in the 1e+1μ+1b+X class. No deviation is present in the 1e+1μ+1jet incl. class, i.e. if no b jet is required.

In analyses of top quark related phenomena, it has been found that the top quark p_T spectrum is not modeled correctly in the available simulation samples because of the significant influence of higher order perturbative corrections [98, 99], which makes a correction procedure necessary [100]. This is most likely not the cause of the discrepancy observed here, since simulation tends to overestimate the top quark spectrum at high p_T, which would cause an observed deficit, but not an excess in the high-mass region.

In summary, intriguing excesses are observed over a number of event classes related to the 1e+1μ+1b final state. The excesses are predominantly visible in inclusive event classes with additional jets and b jets. Whether the excesses are due to statistical fluctuation or are indeed signs of BSM phenomena remains to be determined in the comparison with the data taken by CMS in 2015 and the following years.

4.4.3 Exotic Events

In the sets of the most significantly deviating classes in all scans of distributions and integrals, exotic final states with single-digit numbers of data and simulation events are found. The low event counts prohibit a meaningful comparison of kinematic distribution shapes. The p- and $\tilde{p}$-values produced in the different kinematic distributions vary depending on the coincidence of bins containing SM expectation and the observed event, which is quasi-random. In this section, exotic events contributing to such final states are discussed.

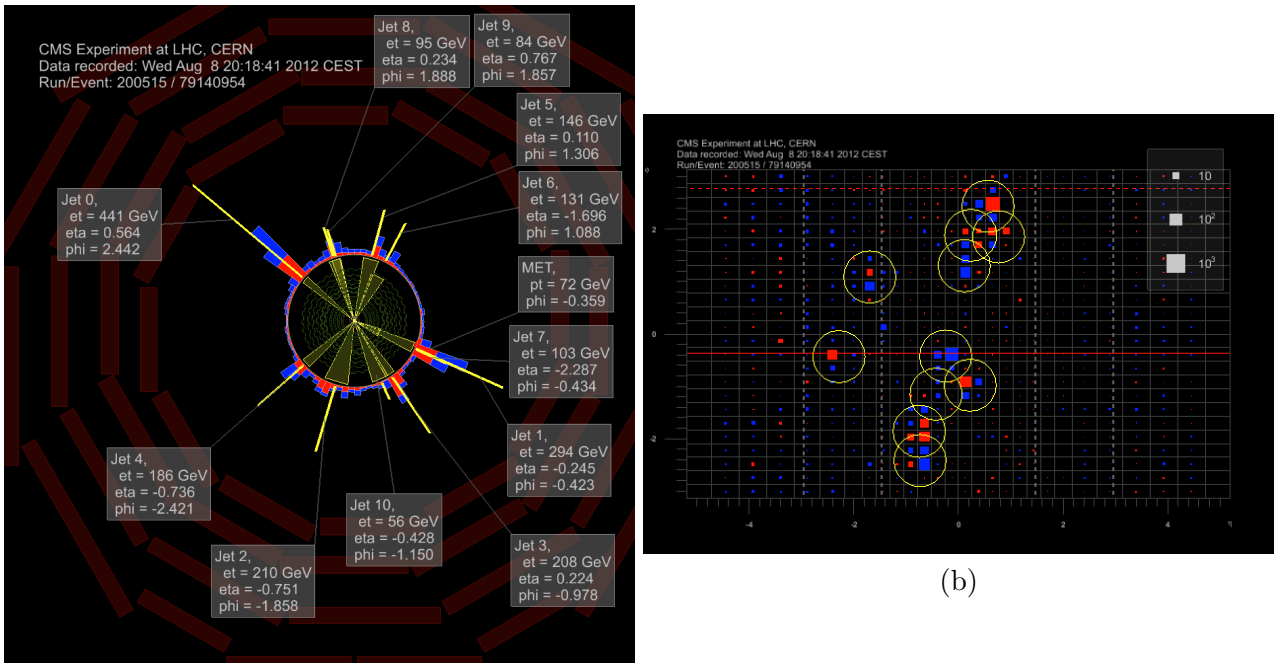
In the 7jet+4b+E_T^{miss} final state the distribution integral yields $N_{\text{Data}} = 1$ and $N_{\text{SM}} = (4\pm3)\times10^{-4}$ and thus an integral p-value of 7×10^{-4}. The kinematic variables for this event are $\Sigma|p_T| = (2.02 \pm 0.06)$ TeV, $M_T = (2.03\pm0.05)$ TeV and $E_T^{\text{miss}} = (55\pm30)$ GeV[6]. The event display is shown in fig. 4.17. The total of eleven b and non-b jets are clustered in three groups. In the central $|\eta| < 2.0$ region, there are two clusters of four and five (b-)jets, which are roughly back-to-back in ϕ and have opposite signs of η. Of both clusters, highest-p_T jets are tagged as b jets. The remaining two jets are located in the barrel-endcap overlap region. Due to the high jet multiplicity, the rather low value of $E_T^{\text{miss}} \approx 55$ GeV most likely results not from prompt undetected particles, but rather from jet mis-measurement and secondary neutrinos from hadron decays.

The 2μ+7jet+2b+E_T^{miss} excl. class has a long-standing status as MUSiC's most exotic lepton triggered final state in 8 TeV data. The distribution integral in this distribution yields $N_{\text{Data}} = 1$

[4] By adding these bins to the RoI, the region contents would become: ($N_{\text{Data}} = 108$ vs. $N_{\text{SM}} = 49.1 \pm 8.5$), ($N_{\text{Data}} = 133$ vs. $N_{\text{SM}} = 63.8 \pm 10.9$), and ($N_{\text{Data}} = 165$ vs. $N_{\text{SM}} = 80.8 \pm 13.1$), for one, two and three added bins, respectively. By considering the naive measure $(N_{\text{Data}} - N_{\text{SM}})/\sigma_{\text{SM}}$, the reduced significance can easily be estimated. The corresponding p-values of 1.6×10^{-6}, 2.5×10^{-6} and 9.5×10^{-7} confirm the validity of the estimate.

[5] The 1e+1μ+1b incl. event class contains all events from the 2e+1μ+1b, 1e+2μ+1b, etc. classes, i.e. all event classes that are constructed by adding one additional object to the class name.

[6] The uncertainty quoted here is the resolution as described in sec. 2.4.6.1.



(a) (b)

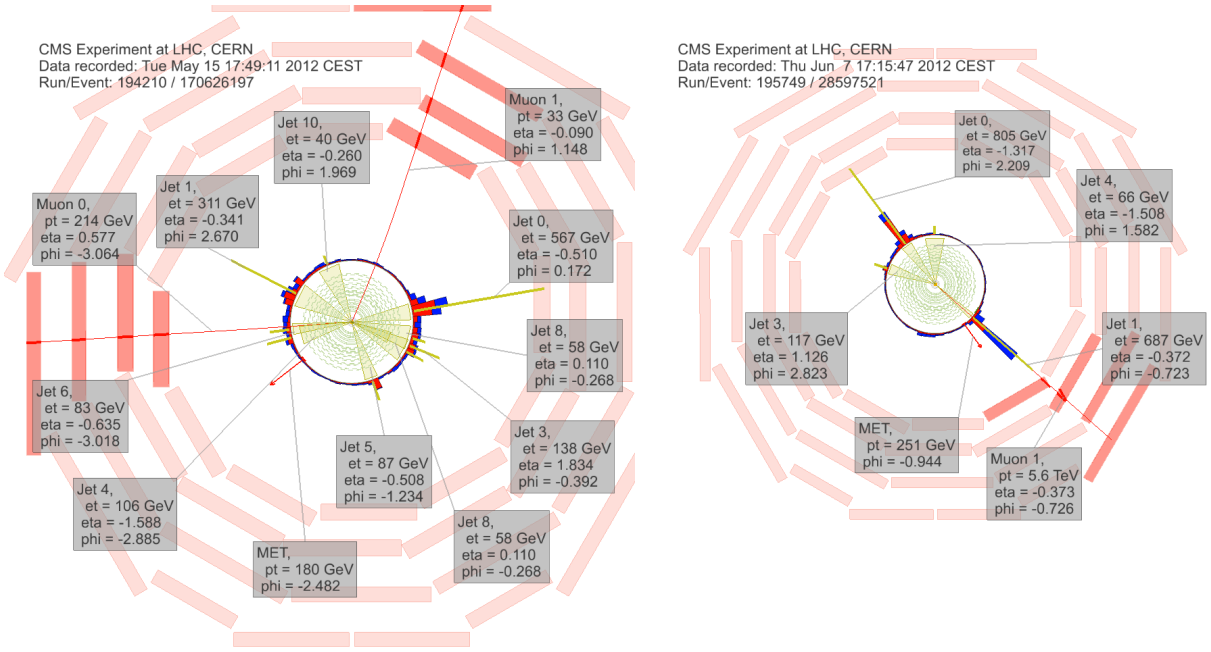
Figure 4.17: Event views of the single data event contributing to the 7jet+4b exclusive event class in the ρ-ϕ-projection (a) and the η-ϕ-plane (b). Jets 0,1, 6 and 10 are b-tagged. Jets 6 and 7 are located in the barrel-endcap overlap region.

and $N_{\mathrm{SM}} = (4 \pm 7) \times 10^{-4}$. The kinematic variables for this event are $\Sigma|p_T| = (1.87 \pm 0.05)\,\mathrm{TeV}$, $M_T = (1.90 \pm 0.05)\,\mathrm{TeV}$ and $E_T^{\mathrm{miss}} = (170 \pm 30)\,\mathrm{GeV}$. A display of the contributing data event is shown in fig. 4.18 (a). In contrast to the extreme final state multiplicity, the values of $\Sigma|p_T| \approx M_{\mathrm{inv}} = 1.9\,\mathrm{GeV}$ are large, but not extremely so. The value of $E_T^{\mathrm{miss}} = 170\,\mathrm{GeV}$ is also in the medium range.

The 1μ inclusive final state is - in contrast to the previously discussed final states - highly populated in data and simulation and not at all exotic. It does not show any significant deviations between data and simulation. However, a comparison between the distribution with and without activated jet triggers (fig. 4.19 (a) and (b), respectively) shows that new events are added in the high-p_T bins of the distribution. This is intriguing because multi-TeV muons are very rare, especially if they pass the object identification requirements but evidently fail the trigger requirements, thus causing the event to be "hidden in the wrong data stream". Closer inspection of the event view of the event in the very-high-p_T bin (fig. 4.18 (b)) shows that in addition to the muon, there are four jets in the event, of which two have comparatively low p_T. The two high-p_T jets are back-to-back in ϕ and have comparable p_T of 687 and 805 GeV, respectively. The jet configuration is on its own similar to a typical dijet production event with additional low-p_T radiation jets. However, the aspect that makes the event interesting is an added muon with reconstructed $p_T \approx (5.6 \pm 2.4)\,\mathrm{TeV}$[7], which is close to the center-of-mass energy of 8 TeV and would thus be among the highest-p_T muons ever found in the CMS data. It also causes the event to be the highest-$\Sigma|p_T|$ event observed in the analysis by a margin of 1.8 TeV. However, the reconstruction of the muon is far from ideal:

- As already indicated by the uncertainty of the muon p_T, the muon p_T resolution in the TeV regime is bad. The momentum reconstruction is heavily reliant on the measurement of the track bending radius, which becomes less precise, the closer a track is to being straight. To combat problems with the reconstruction of high-p_T muons, the *Cocktail*-p_T-reconstruction is used, which combines the information from tracker and muon chambers in an optimised way, thus reducing

[7] The uncertainty corresponds to the bin widths as calculated in sec. 2.4.6



(a) (b)

Figure 4.18: Event displays of the data events contributing to the 2μ+7jet+2b+E_T^{miss} excl.(a) and 1μ+X (b) classes. While the former's rarity is due to the combination of two well reconstructed muons with a high number of nine non-overlapping jets, the latter is extraordinary because of a muon reconstructed with $p_T = 5.6 \pm 2.4\,\text{TeV}$.

the reliance on the tracker measurement. It also takes into account the onset of electromagnetic showering at muon energies above $\approx 900\,\text{GeV}$. In the present case, the combination seems to be at least questionable: The measurement by the tracker alone gives $p_T \approx (0.8 \pm 0.8)\,\text{TeV}$. The tracker measurement thus effectively provides no information, and the resultant p_T-value is entirely derived from muon chamber data. In an unfortunate coincidence, the muon track seems to cross the muon system right in the gap region between adjacent chambers, thus further complicating the measurement. In sum, it cannot be stated with certainty that a high-p_T muon is observed in this event.

- The muon direction of flight is very close to parallel to the axis of one of the two leading jets[8]. Only the very high reconstructed p_T prevents the muon from failing the isolation requirement[9]. A p_T-reduction of 2 % would suffice to cause the muon to be discarded due to insufficient isolation, which is the most likely the reason the event fails the muon trigger requirements. The overlap with the jet can cause two additional complications: *punch-through* and muon production in hadron decays. *Punch-through* is the propagation of jet constituents through the HCAL and magnet coil into the muon system. A jet constituent could thus cause a signal in the muon chambers, which could happen to be consistent with any of the jet constituent tracks and thus give rise to a reconstructed muon. While it is unusual for punch-through to reach more than one muon chamber, it may not be impossible in this case because of the tracks' close proximity to

[8] If jet and muon are so close, how can the jet have a relatively low $p_T \approx 700\,\text{GeV}$? Similarly, how can there be only about $200 - 250\,\text{GeV}$ of E_T^{miss} in the event? For the event to be balanced in p_T, this would require at least 5 TeV of p_T in addition to the muon, thus significantly breaking the center-of-mass barrier. In fact, the answer to both questions is the same: the muon is not a PF muon candidate, which is not required for the high-p_T muon identification (cf. 2.2.1), because the PF algorithm does not handle TeV range muon chamber tracks correctly. It is thus taken into account for neither jet clustering nor the calculation of E_T^{miss}.

[9] The requirement on the PF isolation fraction (cf. 2.2.1) is to be smaller than 0.12, for this muon it is measured to be 0.118.

the muon chamber gaps. An alternative cause would be a muon stemming from a hadron decay. The jet is predominantly composed of charged hadrons (CHF = 95%, cf. 3.1.1). If the majority of the jet energy is carried by a single $\pi^{\pm}$-meson, a high-energetic muon may result from the decay, enter the muon system and cause a "real" muon signal, which then for one of the many aforementioned reasons may be heavily mis-reconstructed.

It speaks to the quality of the detector simulation that this exotic effect seems to be well modeled, as can also be seen from fig. 4.19. Simulated events contributing in the high-p_T region stem from QCD multijet production, which again makes plausible the hypothesis of a jet induced muon signal. To conclude, this extremely high-p_T muon event most likely does not contain a muon from the primary interaction. It however provides an interesting motivation to study its causes in the details of particle reconstruction.

4.5 Multijet Classes

Jet triggered event classes generally give insight into QCD dominated physics processes. This is especially true for multijet exclusive classes, i.e. classes with final states containing jets only. These classes are only weakly influenced by electroweak processes, such as W and Z boson or top quark[10] production, which shape the kinematics of many other final states. While these processes can result in multijet final states, their cross-sections are dwarfed by direct QCD jet production.

In this section, the results for multijet final states are presented. Sec. 4.5.1 gives an overview of the results in different kinematic distributions. In sec. 4.5.2 an intriguing feature of the $\Sigma|p_T|$-distributions is discussed in some more detail.

4.5.1 Kinematic Distributions

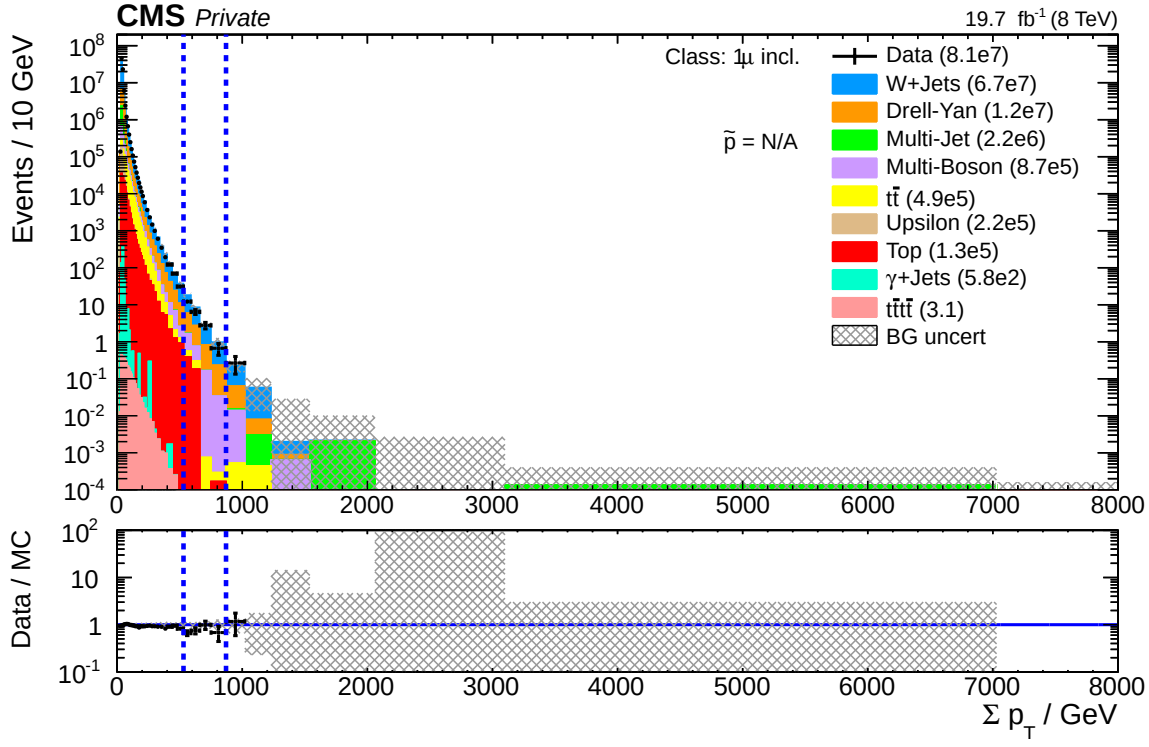
A total of eleven jet exclusive classes are observed, one for each multiplicity from one jet up to eleven jets. Fig. 4.20, 4.21 and 4.22 show the $\Sigma|p_T|$- and M_{inv}-distributions for the 2jet, 4jet and 9jet exclusive classes. Multiple common features strike the eye:

- **Shape**: In QCD production of jets no new production channels become possible with increasing energy and multijet spectra are expected to be simply exponentially falling. While this physical expectation is confirmed for sufficiently large M_{inv} and $\Sigma|p_T|$, a peaking turn-on-like structure is clearly visible in all multiplicities. This feature is a result of the coincidence of single jet triggering and event p_T-balancing and discussed in the following section.

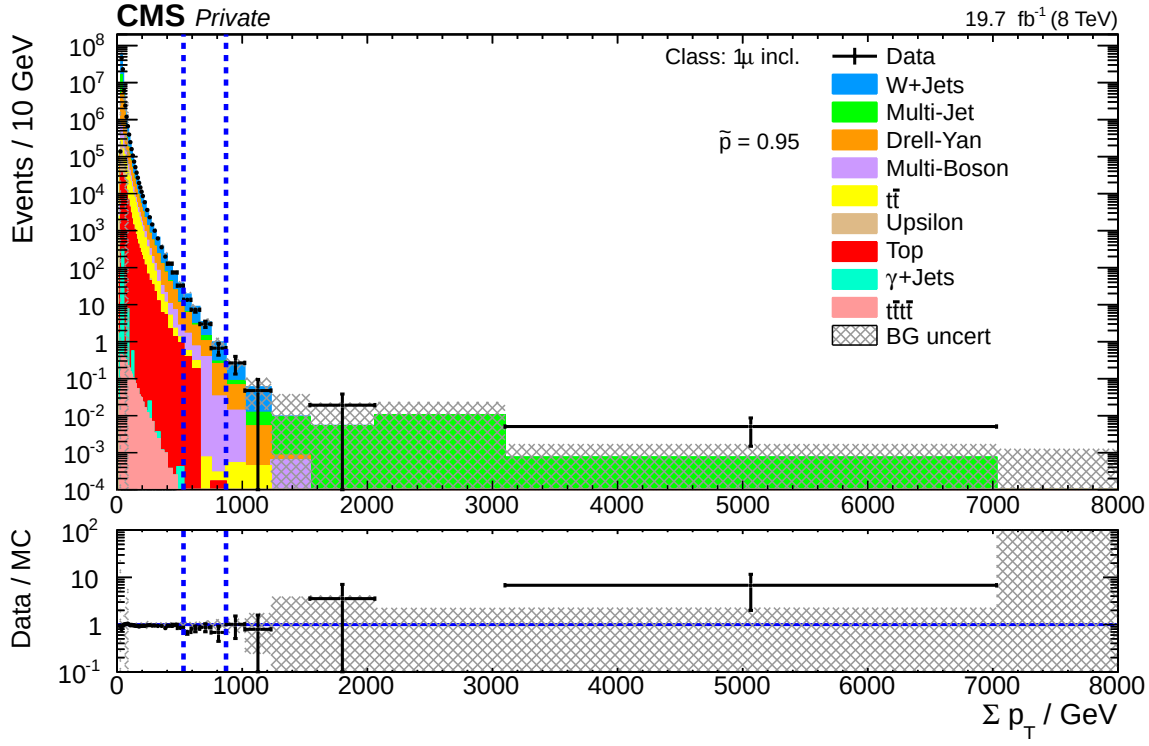
- **Scale**: Remarkable ranges in energy scale are probed with up to $\mathcal{O}(1\,\text{Event/GeV})$ at $M_{\text{inv}} \approx 3.5\,\text{TeV}$, which is a result of the shape of QCD spectra, which are only slowly falling compared to their electroweak counterparts. This allows for a complementary view of the SM, fully utilizing the high energy reach of the LHC.

Because E_T^{miss} is treated as a separate object, the multijet exclusive classes do not provide a E_T^{miss}-distribution and one needs to consult the "neighboring" $n$jet+E_T^{miss} classes. The E_T^{miss}-distributions of the 2jet+E_T^{miss} and 8jet+E_T^{miss} event classes are shown in fig. 4.23. As with the other kinematic distributions, data are remarkably well modeled by the QCD simulation. The dijet E_T^{miss}-distribution highlights the use of different kinematic distributions in a model unspecific search. QCD and W boson induced jet events are very clearly separated into the low- and high-E_T^{miss} regions. In the $\Sigma|p_T|$- and M_{inv}-distributions of the same class (fig. 4.20), the W boson contribution to the bin content is approximately flat at the 1 % level over the whole distribution, which prohibits any sensitivity to this process. By considering the E_T^{miss}-distribution, MUSiC can emulate the kind of event selection requirement a dedicated analysis would use to separate the considered "interesting" process from potentially overwhelming backgrounds without losing any events.

[10] More precisely, only single top quark production is an electroweak process. Top quark pair production is predominantly mediated by the strong interaction. It is nevertheless true that both kinds of top quark production have little influence on multijet final states.



(a)



(b)

Figure 4.19: $\Sigma|p_T|$-distribution of the 1μ incl. class before (a) and after addition of jet triggers (b). In the region of very high p_T, additional events enter the distribution. In these events, a muon candidate passes the object identification criteria but fails the trigger requirements. Especially the rightmost data point in (b) is interesting because of the extreme muon p_T.

Over all spectra, the dominating source of uncertainty is the flat 50 % LO cross-section uncertainty (cf. sec. 2.4.4.1). In the low-statistics regions at the very low and high ends of the distributions, the Fill-Up uncertainty procedure gives significant contributions. Judging the overall very good agreement between data and simulation, it is evident that the assigned cross-section uncertainty is globally rather conservative. However, its size may locally be justified, as e.g. in the low regions of the $\Sigma|p_T|$-distributions, which show discrepancies between data and simulation that may be the result of missing NLO modeling. These deviations are studied more closely in the next section.

4.5.2 Discussion of Low-$\Sigma|p_T|$ Feature

The region of low-$\Sigma|p_T|$ in the dijet exclusive class is populated by imbalanced events: Since triggering is required, one of the two reconstructed jets must have $p_T > 400\,\text{GeV}$. For $\Sigma|p_T|$ to lie in the range of $450 - 800\,\text{GeV}$, a significant p_T-difference between the leading and all sub-leading jets is necessary. The effect is reduced at higher jet multiplicities, where each value of $\Sigma|p_T|$ can be decomposed into more different combinations of single jet p_T values. This phenomenon is intriguing for three reasons:

- By construction, any final state momentum imbalance results in E_T^{miss}, which in total enforces final state balancing. However, no additional particles, including E_T^{miss}, are identified, which means that a significant amount of p_T is lost between the PF candidate level and the final MUSiC object identification. This aspect concerns reconstruction and identification efficiencies and is thus experimental.
- In the dijet class, where the effect is most pronounced, any physical imbalance in the leading jet momenta can only be introduced by NLO effects, where additional jets with lower momentum are produced in addition to the leading jet pair, which would otherwise be exactly balanced. The shape of the imbalance will thus also be sensitive to differences in LO and NLO QCD modeling.
- Dedicated analyses such as the measurement of the dijet cross-section [101] use different (prescaled) triggers for different kinematic regions, so the artificial enriching with imbalanced events due to to a single trigger threshold does not occur and no similar effect is observed.

The balancing of the jets is not only observable in p_T, but also in the angular separation in the transverse plane $\Delta\phi$. The distribution of $\Delta\phi$ in the dijet final state is shown in fig. 4.24 in two $\Sigma|p_T|$-regimes. In the high-$\Sigma|p_T|$ regime, the spectrum is dominated by LO-like dijet events, causing a sharp peak around π. In the low-p_T regime, p_T-imbalance dictated by the triggering requirement also entails a reduced correlation between the jet azimuth angles, resulting in a significantly flatter spectrum. Here, simulation and data seem to agree well.

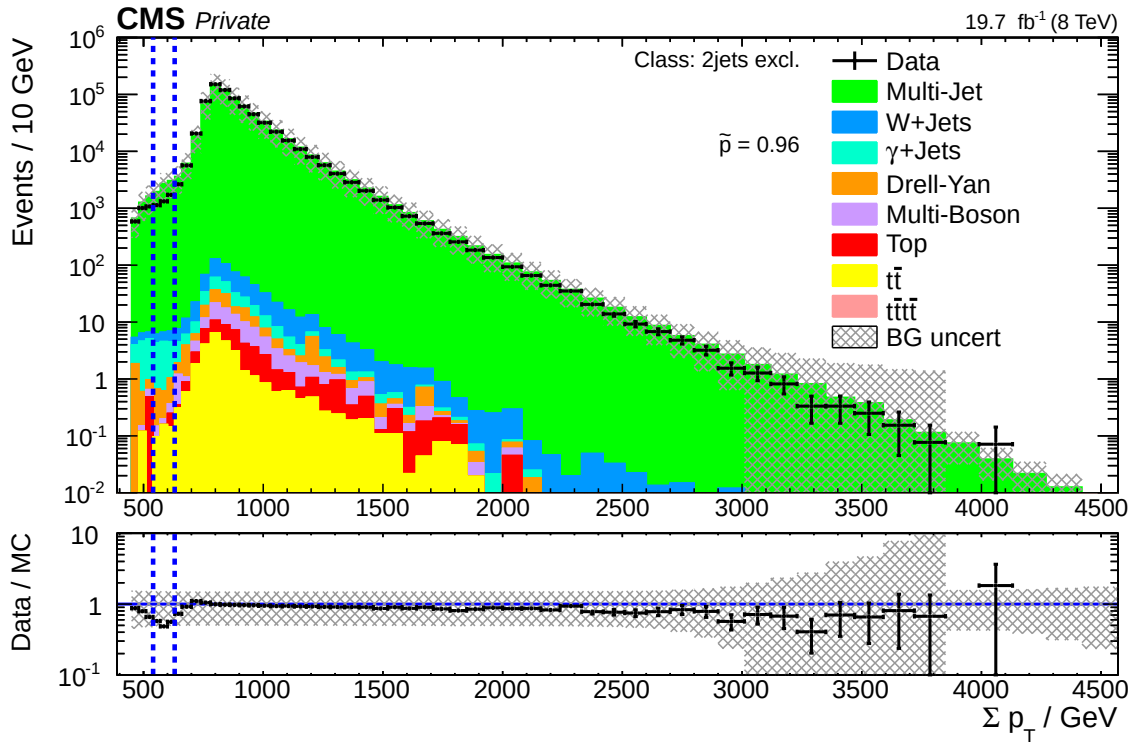
As the $\Sigma|p_T|$-spectrum is the convolution of the two single jet spectra, it obscures individual spectral features. Fig. 4.25 shows the combined single jet spectrum of leading and sub-leading jets in the dijet final state. Analogously to the dijet $\Sigma|p_T|$-spectrum, the single jet spectrum is divided by the trigger turn-on feature at 400 GeV. In this view, the source of the low-$\Sigma|p_T|$ discrepancy becomes apparent: while the high-p_T spectrum is well modeled, simulation and data exhibit a remarkable mismatch in the low-p_T region. A similar behavior is observed in [102].

Concerning the analysis procedure, two possible causes of a deviation are final state selection and jet identification:

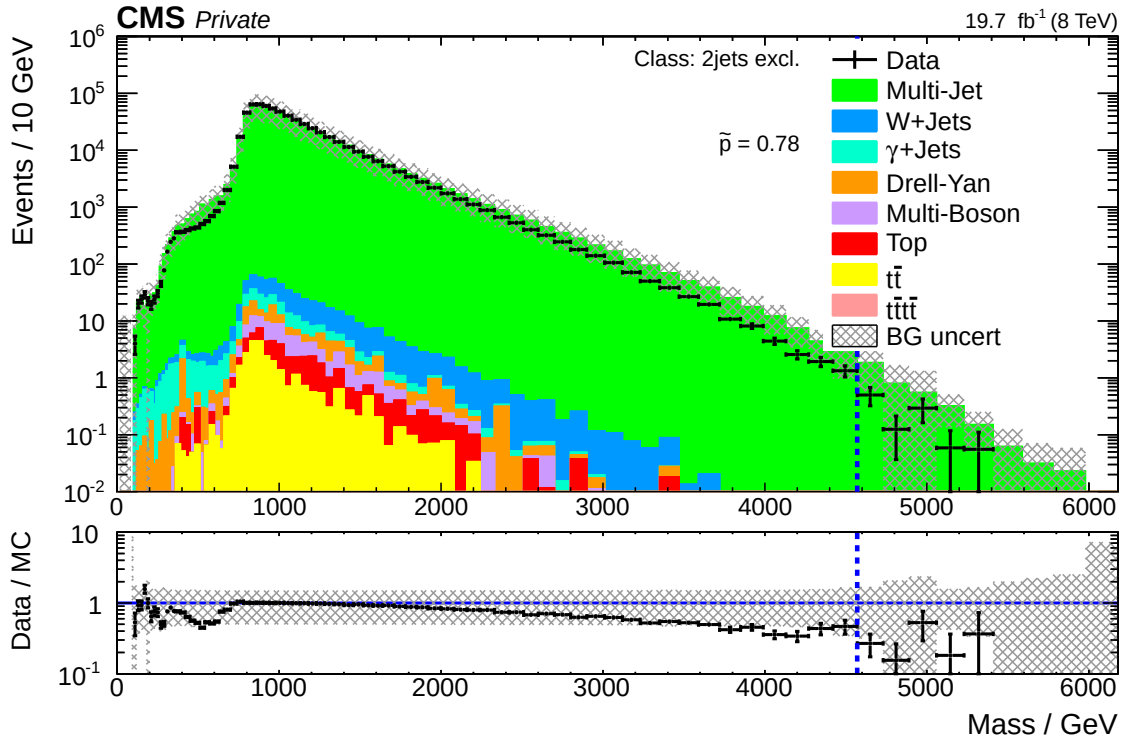
- To gauge the influence of the final state selection, which implicitly vetoes additional jets above $p_T = 50\,\text{GeV}$, but ignores them below, the single jet spectrum can be viewed for all jet final states, i.e. with lepton, photon and E_T^{miss} vetoes applied[11], but without a jet veto[12]. Fig. 4.26

[11] The vetoes effectively separate QCD jet production from other dominant sources such as W boson production, which commonly results in final state leptons.

[12] At first glance, this seems equivalent to the $\Sigma|p_T|$-distribution of the 1jet+X class. However, in the 1jet+X class, only the momentum of the leading jet is considered for $\Sigma|p_T|$. Since events are required to be triggered, the leading jet must always pass the trigger threshold, i.e. $p_T > 400\,\text{GeV}$, which effectively removes the region we are interested in from the distribution.



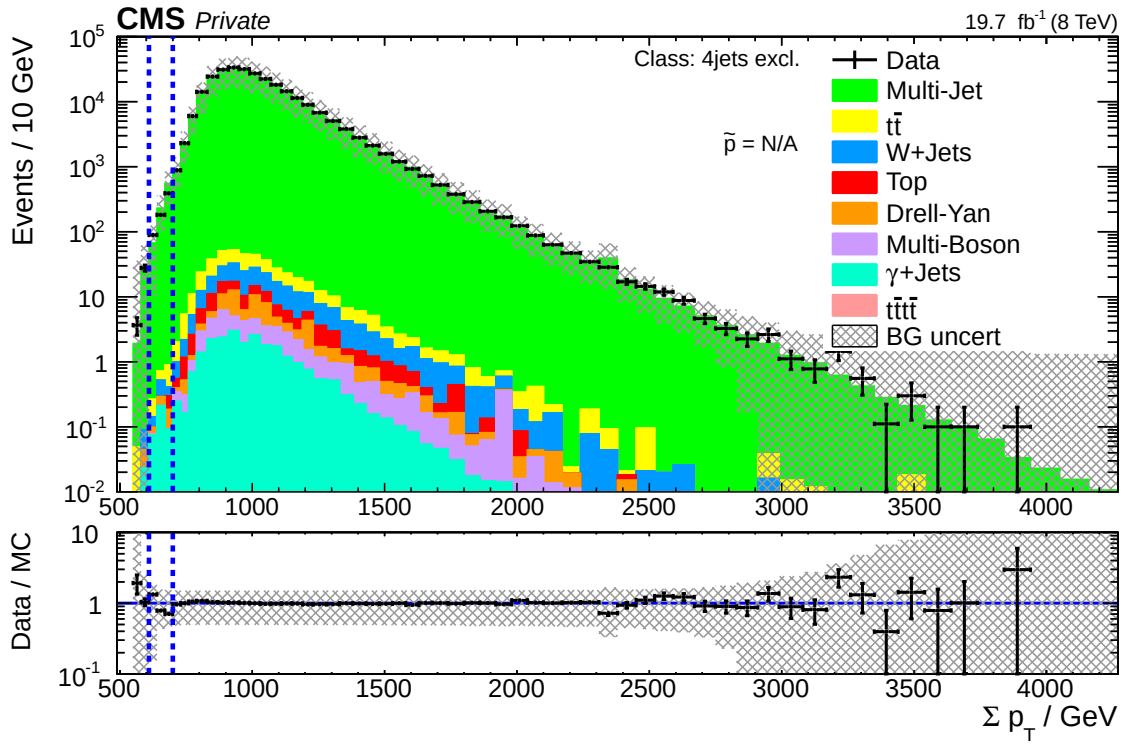
(a)



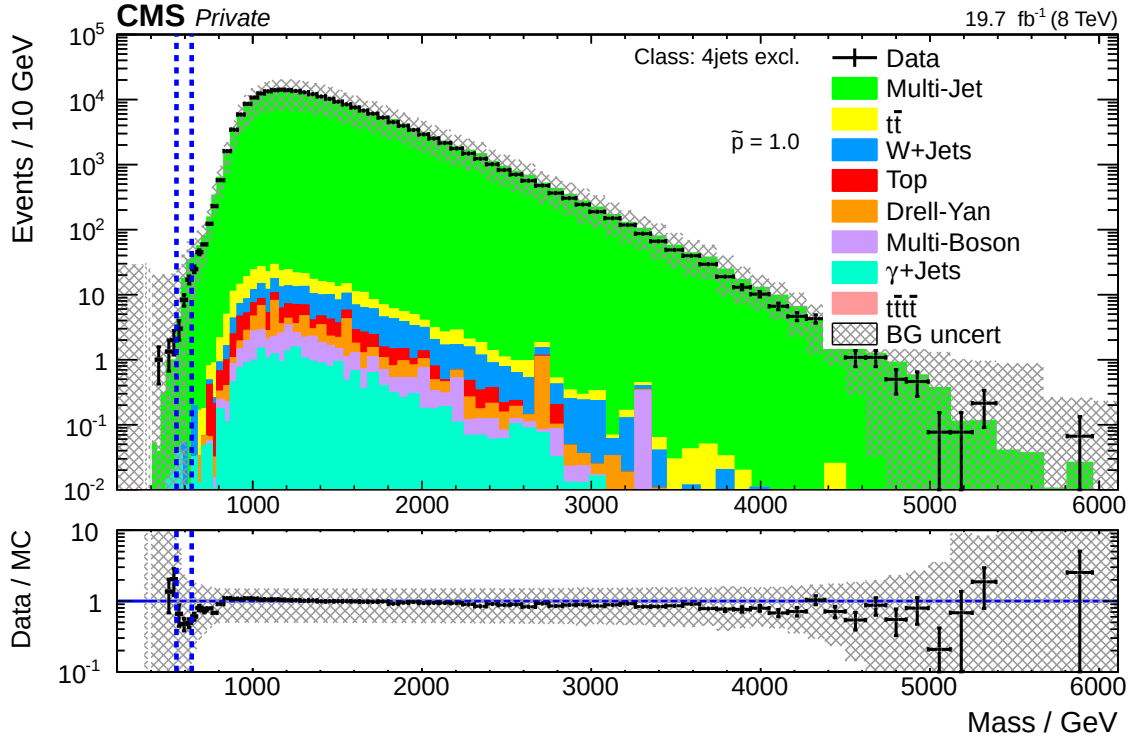
(b)

Figure 4.20: $\Sigma|p_T|$- (a) and M_{inv}-distributions (b) of the 2jet excl. class. Due to its simplicity and abundant production, the dijet final state is among the most popular QCD final states. The sharp peak at $\Sigma|p_T| \approx 800\,\text{GeV}$ is a result of the dominant back-to-back dijet topology, which contains two p_T-balanced jets, and the trigger requirement of $p_T \geq 400\,\text{GeV}$. Over a large $\Sigma|p_T|$-range, the simulation performs well in describing the spectrum. A discrepancy w.r.t. to data at low $\Sigma|p_T|$ is studied in sec. 4.5.2

.

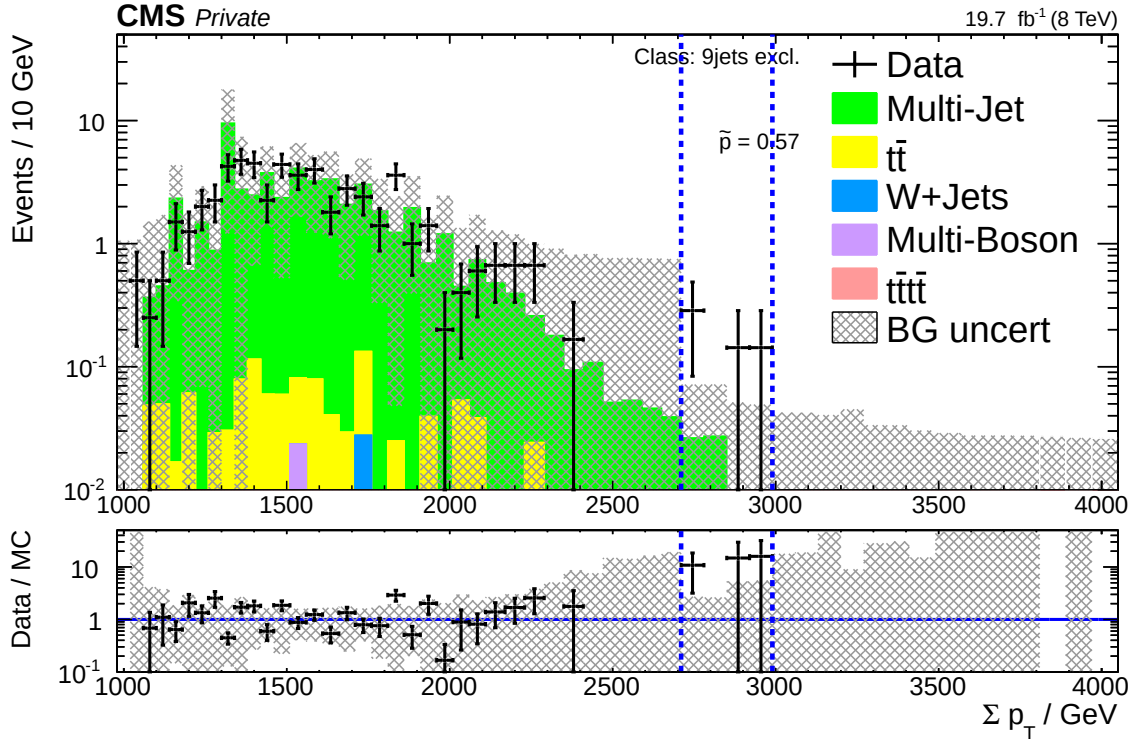


(a)

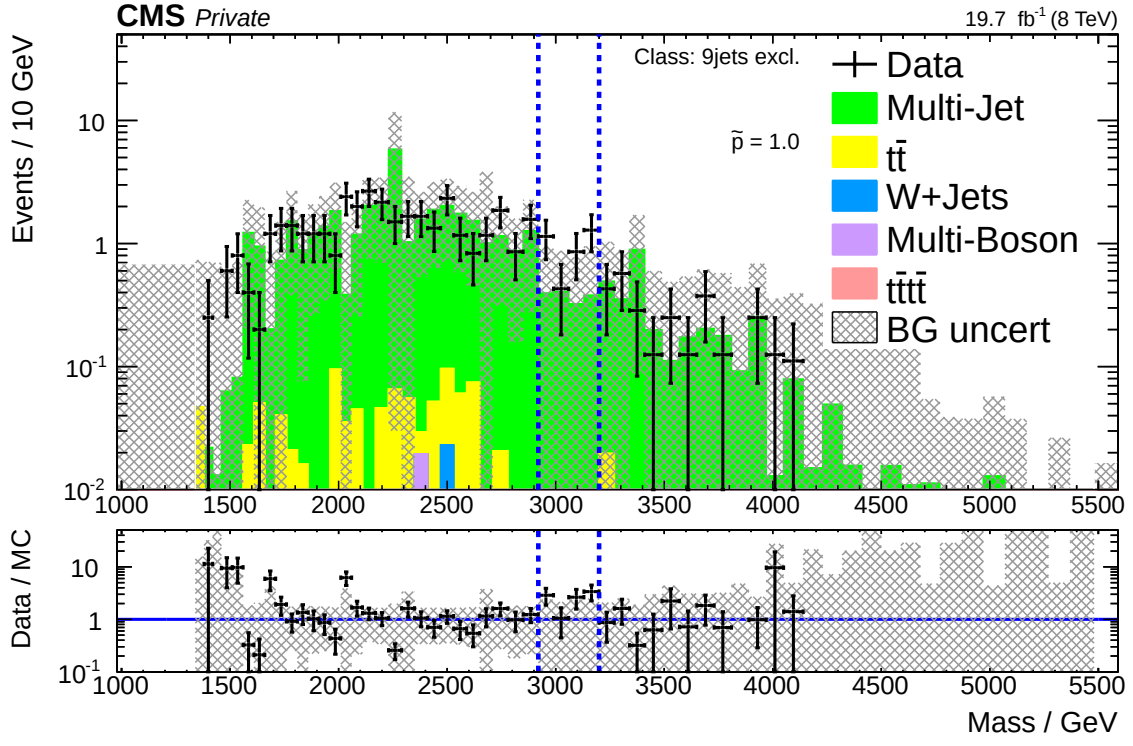


(b)

Figure 4.21: $\Sigma|p_T|$- (a) and M_{inv}-distributions (b) of the 4jet excl. class. At this medium jet multiplicity, the low-$\Sigma|p_T|$ deviations play a reduced role. Simulation of distribution shape and normalisation are in excellent agreement with data over large ranges of their respective kinematic variables.

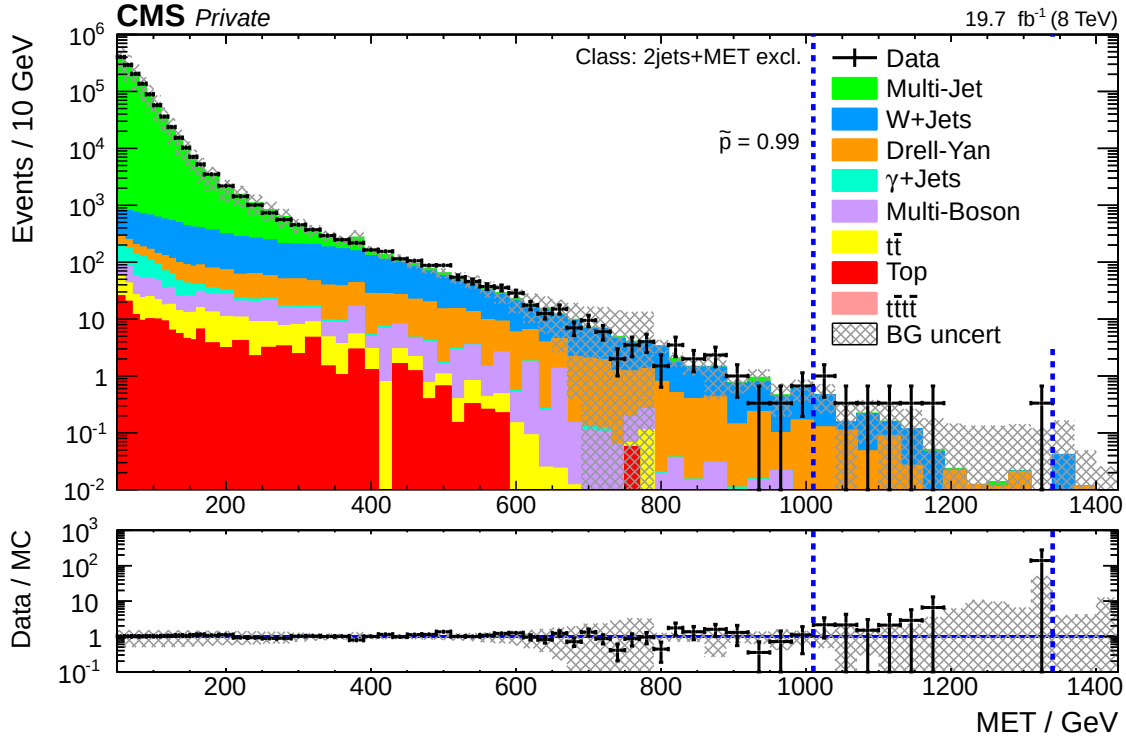


(a)

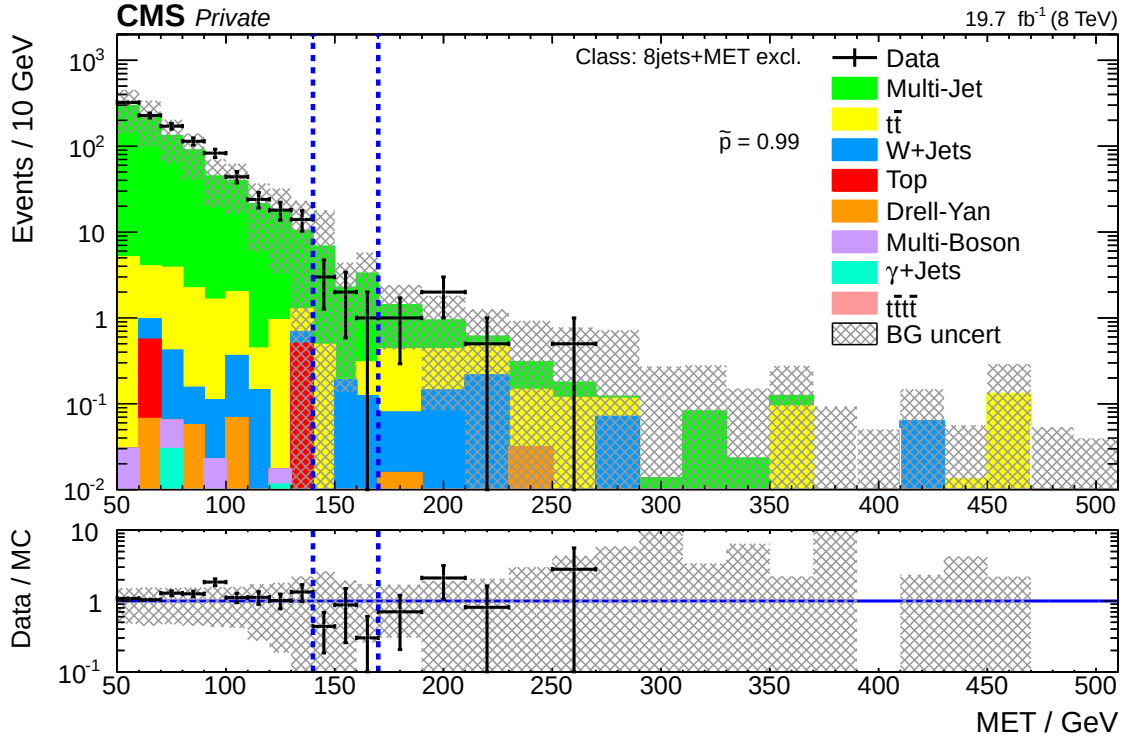


(b)

Figure 4.22: $\Sigma|p_T|$- (a) and M_{inv}-distributions (b) of the 9jet excl. class. Diminishing event statistics cause reduced sensitivity to distribution shape and increased Fill-Up uncertainties. However, even at such high multiplicities, the SM modeling performs remarkably well in describing the total number of events.



(a)



(b)

Figure 4.23: E_T^{miss}-distributions of the 2jet+E_T^{miss} (a) and 8jet+E_T^{miss} exclusive event classes (b). The E_T^{miss}-spectra are well modeled over a range of more than one TeV. In the 2j+E_T^{miss} class, the E_T^{miss}-distribution (a) very well separates QCD jet production from W boson production, giving sensitivity to W decay modeling in a fully hadronic channel.

(a) shows the data / simulation ratio in this final state configuration as a function of the dijet $\Sigma|p_T|$, binned in $|\eta|$. Two trends can be observed: simulation overestimates the contributions at high jet-p_T and the characteristic "s-shaped" low-p_T deviation feature is present in all $|\eta|$-bins independent of the applied vetoes.

- Three types of jet identification criteria are used: p_T- and $|\eta|$- and composition variable requirements. The p_T-requirement of 50 GeV is well above the minimal recommended value of 30 GeV, which can be considered safe. In the context of the discussion in sec. 3.1 it was found that composition variables have next to no influence on the dijet spectrum. The low-$\Sigma|p_T|$-distribution is present with the PFTight as well as PFLoose jet identification criteria (cf. sec. 3.1 and 2.2.4). The presence of the deviation in all $|\eta|$-bins of fig. 4.26 finally excludes an influence of the chosen $|\eta|$-requirement. In sum, the jet identification process is not a likely source of the deviation source.

In combination, this indicates that the deviation is neither caused by the specific final state selection, nor the $|\eta|$- and p_T-ranges used for jet reconstruction. Other causes inherent to the analysis, such as implementation errors in jet resolution smearing, b tagging and application of scale factors have been ruled out by checking the analysis results with these features disabled. An alternative source of deviation is of course the simulation itself. Cross-checking between multiple generators is performed in the following.

4.5.2.1 Comparison with Madgraph and Sherpa

In studies of dijet angular de-correlation, i.e. the deviation of the distribution of the angular separation of dijets from a peak at π, large discrepancies between different event generators at nominally identical settings have been observed [30]. Especially, multijet modeling by MADGRAPH is found to be superior to that of PYTHIA6. To asses the influence of different generators two cross-check samples are used: a MADGRAPH sample of LO QCD available from central CMS production, and a privately produced SHERPA NLO QCD sample.

Madgraph Sample: The HT-binned samples with dataset name `QCD_HT-*-madgraph-pythia6` are used, where the asterisk stands for one of four H_T-bins covering in total the range from 100 GeV to infinity. The `pythia6` part of the dataset name indicates that while MADGRAPH was used for the simulation of the hard interaction, parton showering and hadronisation were simulated by PYTHIA6. Detector simulation was performed by FULLSIM. The CTEQ6L1 PDF set was used (cf. sec. 2.4).

Sherpa Sample: The effects of NLO corrections to the QCD jet production kinematics are assessed by comparing the available LO simulation with NLO results from Monte Carlo event generator SHERPA [103]. For NLO QCD simulation, it is of special importance to correctly combine parton level calculations with parton showering, which are not independent. To this end, SHERPA employs sophisticated shower matching procedures, which preserve the precision of both simulation steps, resulting in increased multijet performance and reduced dependence on the choice of factorisation scale [104].

Using SHERPA 2.1.1 [105], a sample of approximately 15×10^6 multijet events was generated. All $2 \to 2$ parton scatterings are taken into account at NLO accuracy in QCD, matrix element calculation is performed by SHERPA's own AMEGIC matrix element generator. Loop diagrams are calculated by a SHERPA plugin of the loop generator OPENLOOPS [106]. The MC@NLO shower matching prescription [104, 107] is used. The FastJet [16] implementation of the anti-kt algorithm [14] is used for jet clustering on parton level during event generation. To enhance computation efficiency, only events with one jet with $p_T > 300$ GeV and a second jet with $p_T > 20$ GeV are considered further, thus enhancing the relative contribution of events entering the 2jet exclusive event class. While an even more specific selection procedure would promise to further enhance the yield of events in the relevant dijet regime, it may also distort the effects of mis-reconstruction, which are expected to be relevant to the region. The CT10 PDF set was used (cf. sec. 2.4). The used SHERPA configuration file may be found in appendix B. Detector simulation is performed by FASTSIM.

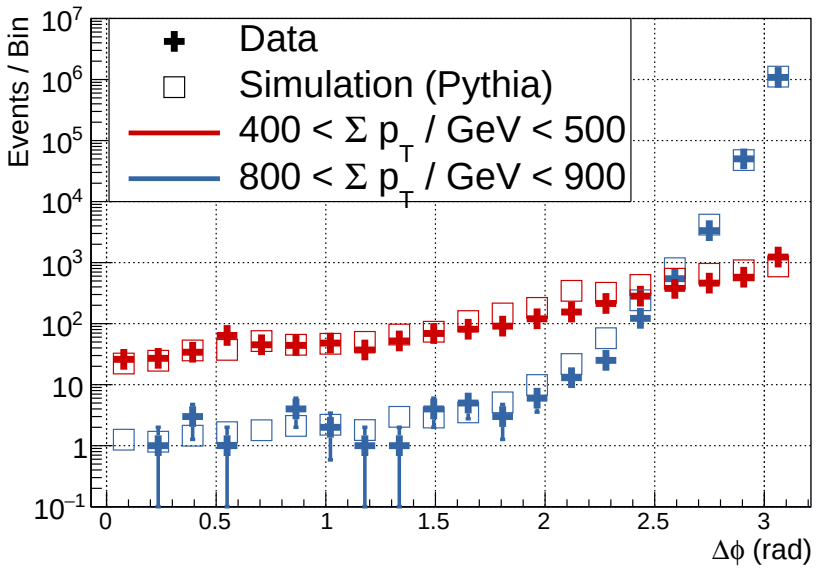


Figure 4.24: Distribution of $\Delta\phi$, the azimuthal angular separation of the two jets in the dijet final state. The distributions are shown for data (solid markers) and simulation (empty markers) in the low- and high-$\Sigma|p_T|$ (red, blue) regions. Error bars indicate statistical uncertainty. For the high-$\Sigma|p_T|$ region, the distribution is strongly peaked at $\Delta\phi \approx \pi$, indicating the dominance of back-to-back dijet events. In the low-p_T region, the distribution is significantly more flat, the jet azimuth angles are less correlated, which hints at a stronger departure from the LO, well-reconstructed dijet topology. The kinematic regions between the two extremes shown here exhibit a steady transition between the low- and high-p_T behavior.

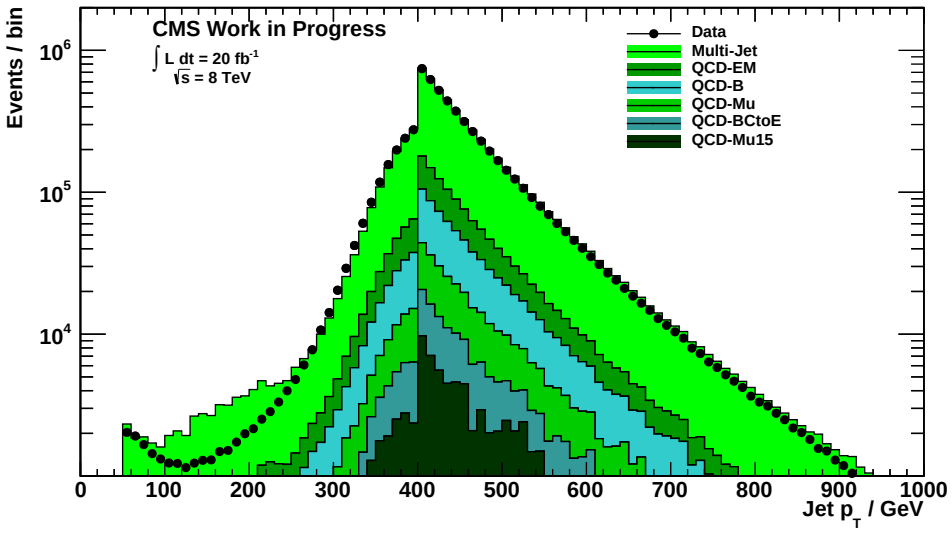
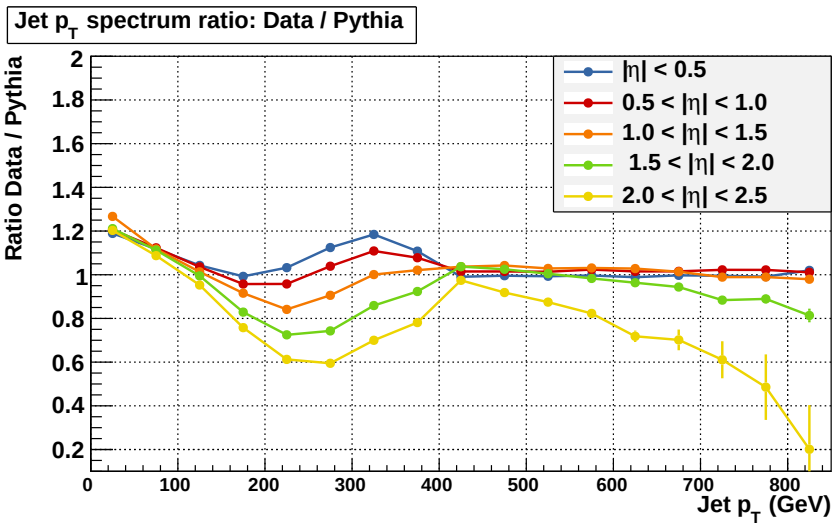
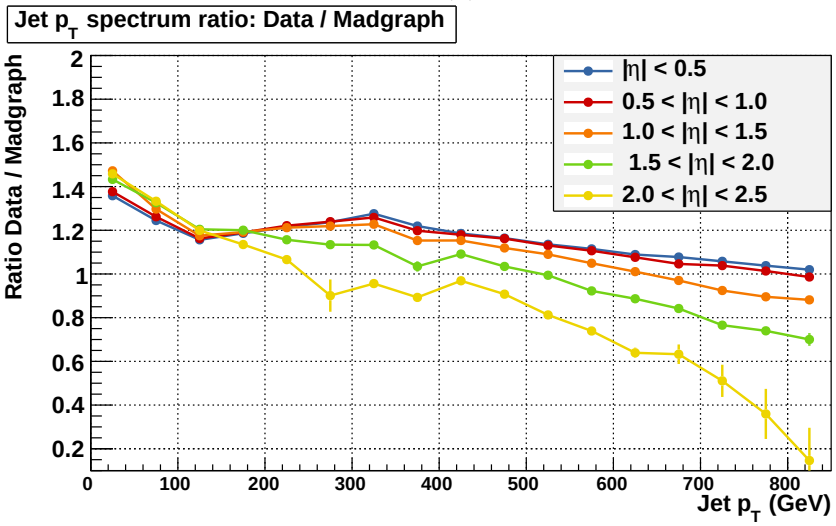


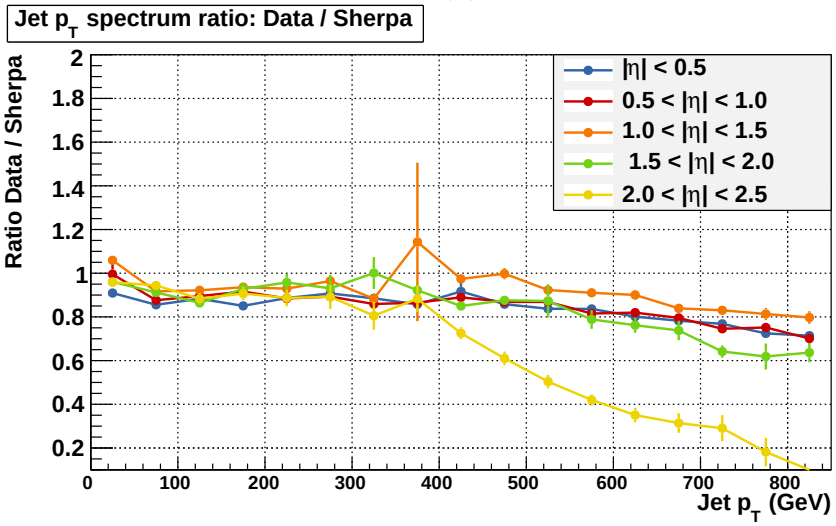
Figure 4.25: Jet-p_T-spectrum of both jets in the dijet final state. Around 400 GeV, the turn-on feature of the `PFJet320` trigger is visible. It is very sharp because of the offline trigger requirement of $p_T > 400$ GeV (cf. sec. 2.1.2 and tab. 2.2). Above the triggering threshold, the jet spectrum is rather well modeled, with small shape discrepancies slowly beginning to appear in the higher p_T-ranges. Below the triggering threshold however, the spectrum shape is strikingly ill-described. The data / simulation ratio of this distribution is shown in fig. 4.26(a).



(a)



(b)



(c)

Figure 4.26: Jet-p_T-spectrum in jet-only final states presented as ratios between data and simulation for event generators Pythia (a), Madgraph (b) and Sherpa (c). Shown uncertainties are statistical only.

Results: The dijet $\Sigma|p_T|$-spectra simulated by MADGRAPH and SHERPA are shown in fig. 4.27 and 4.28, the jet p_T-spectrum data/simulation-ratios are shown in fig. 4.26(b) and (c).

The lower region of MADGRAPH's jet p_T-spectrum ratio (fig. 4.26(b)) exhibits a modulation w.r.t. data of the order of 0.2, roughly half of that of PYTHIA6. However, normalisation aside, MADGRAPH systematically overestimates the cross-section for high-p_T compared to low-p_T jets, resulting in a diminishing ratio towards higher p_T. This may also be related to higher order corrections, which - depending on who calculated them - may lead to a softening [108] or hardening [109] of the jet spectrum. In the dijet $\Sigma|p_T|$-spectrum (fig. 4.27), MADGRAPH seems to reliably model the low-$\Sigma|p_T|$ shape of the distribution. However, the ratio plot shows that the MADGRAPH sample produces an "s-shape" very similar to that in PYTHIA6, which is shifted upwards and thus harder to spot in the absolute plot. While the absolute extent of the deviation seems to be smaller, the effect of the overall "tilting" towards high-p_T jets has to be taken into account, leaving it questionable whether the low-$\Sigma|p_T|$ modeling should be considered superior in MADGRAPH.

SHERPA gives a flat p_T-spectrum ratio w.r.t. data (fig. 4.26 (c))in the low-p_T region, but suffers from the same relative high-p_T cross-section overestimation as MADGRAPH. The dijet $\Sigma|p_T|$-spectrum (fig. 4.28) is well modeled by SHERPA, no deviation is observed below the trigger threshold.

For all generators, a degradation of the jet p_T data/simulation-ratio with $|\eta|$ is observed. As pointed out in [104], this may be related to the choice of renormalisation and factorisation scales, which are commonly set using some functions of the jet transverse momenta for each event. In high-$|\eta|$ topologies, the jet transverse momenta necessarily misrepresent the energy content of the event, leading to mismodeling. Improved performance is claimed for different scale choices, which however introduce discrepancies in the low-$|\eta|$ regime. No comprehensive solution is proposed.
Although dedicated studies have found inclusive jet spectra to agree reasonably well in data and NLO theory [110], the jet inclusive high-p_T cross-section is subject to significant influence of the choice of PDF set. This is also true for high-$|\eta|$ jets, which are the result of large differences in initial state parton momenta, which imply large differences in Bjorken-x. Extreme regions in x tend to be modeled with increased uncertainty with respect to the moderate x regime [63, 64].

In summary, the cause of the low-$\Sigma|p_T|$ deviation can be isolated. After thorough cross-checking, no evidence has been found to suggest a cause in the analysis process. A comparison of different generators shows that the deviation region is susceptible to significant modeling differences. The deviation is found to be present in the MADGRAPH and PYTHIA6 samples, but not in the SHERPA sample. However, it would be rash to pronounce the problem a clear-cut case of "LO vs. NLO". There are caveats to consider: Both the MADGRAPH and PYTHIA6 samples use the PYTHIA6 hadronisation routines, while SHERPA performs its own hadronisation. While it is true that SHERPA performs NLO matching between partons and showers, while PYTHIA6 does not, this is surely not the only difference, leaving it unclear which modeling details are responsible. A further relevant difference is the detector simulation. As has been motivated above, the observed deviation region is sensitive to experimental effects such as mis-reconstruction. It is thus possible that the effect is due to a difference in FULLSIM and FASTSIM, although it is hard to envision a reason for mis-reconstruction modeling to perform better in FASTSIM than in FULLSIM.

As long as the modeling problem cannot be confirmed to be resolved, it may be sensible to exclude the region from the analysis, since it cannot be compared to data in a meaningful way. To remove events with large momentum imbalance, it is not sufficient to increase the jet p_T thresholds. As fig. 4.26 indicates, one would need to set the jet identification p_T-requirement to a threshold similar to the triggering threshold of 400 GeV to remove the region. By doing so, one would also remove large numbers of jets from well-modeled phase space regions in many unrelated final states. It would be more sensible to require jet triggered events to have at least $\Sigma|p_T| \approx 2 \cdot 400\,\text{GeV}$, i.e. twice the triggering threshold. A $\Sigma|p_T|$-threshold is favorable because it more directly relates to the imbalance issue. To avoid formulating a final state dependent $\Sigma|p_T|$-requirement, which would be at odds with the MUSiC philosophy, the p_T-dependent `PFJet320` trigger may be replaced by an H_T-based trigger. In the 2012 data set, the `PFHT650` is the unprescaled H_T-trigger with the lowest threshold.

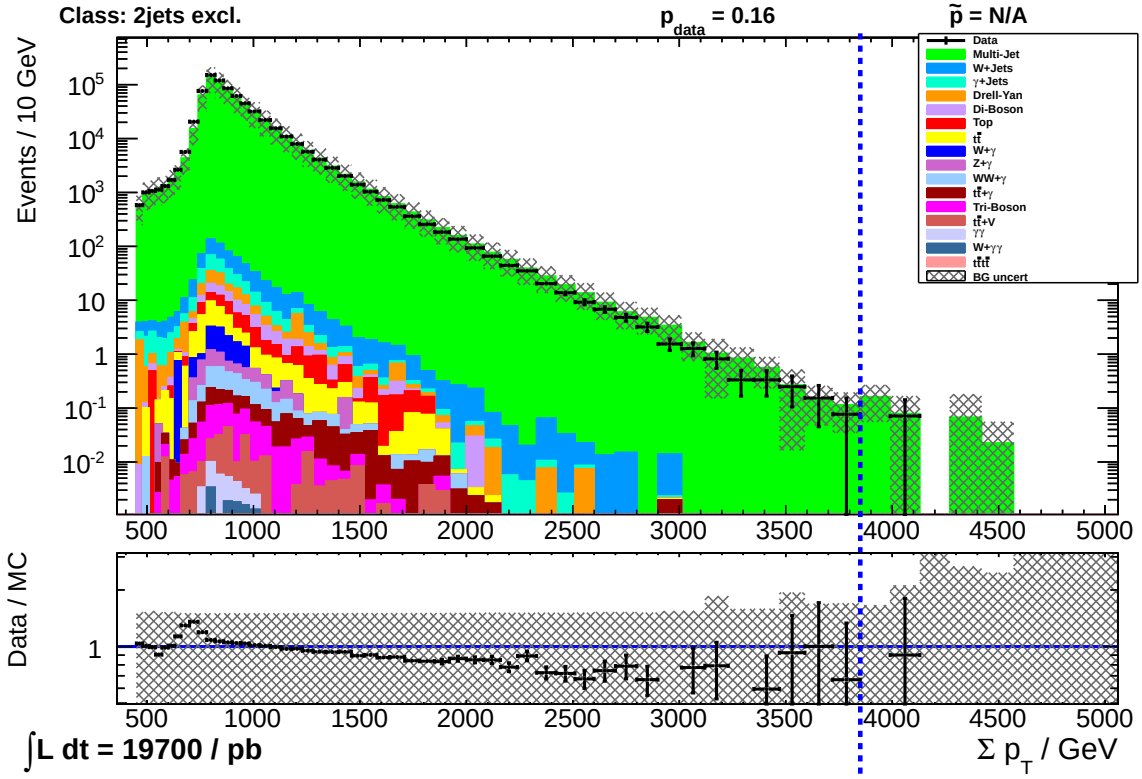


Figure 4.27: $\Sigma|p_T|$-distribution of the 2jet excl. class. The QCD multijet background is modeled using MADGRAPH. Compared to the PYTHIA6 simulation, the discrepancy in the low-$\Sigma|p_T|$ region has a reduced size. Towards higher $\Sigma|p_T|$, Madgraph systematically overestimates the number of observed events.

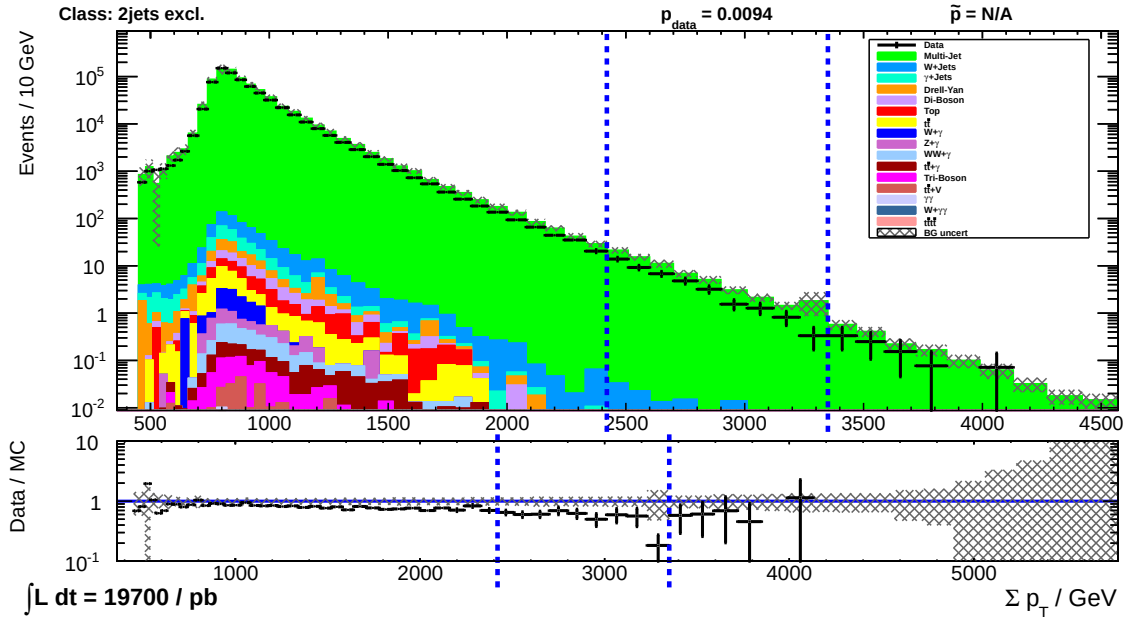


Figure 4.28: $\Sigma|p_T|$-distribution of the 2jet excl. class. The QCD multijet background is modeled by SHERPA. The reduced uncertainties reflect the change to NLO according to sec. 2.4.4.1. No low-p_T discrepancy is visible in the spectrum. In a way similar to MADGRAPH, SHERPA overestimates the number of high-$\Sigma|p_T|$ events.

4.5.3 Normalisation

In the previous sections, the agreement in shape between data and simulation has been discussed. An interesting question is also that of agreement in normalisation. Calculations of NLO corrections to LO cross-sections for jet production commonly find corrections of the order of $15 - 30\,\%$, which are however highly dependent on the arbitrary choice of factorisation and renormalisation scales. A very informative account of the influence of NLO corrections and scale choices on the inclusive three-jet production cross-section is given in [108], which outlines the calculation scheme used to derive cross-section predictions for various CMS jet cross-section measurements [101, 110]. Although "NLO" means $\mathcal{O}(\alpha_S^4)$ for three-jet observables as opposed to $\mathcal{O}(\alpha_S^3)$ for two-jet observables, the results for dijet observables are similar [111, 112]. In the following, figures from the three-jet literature are discussed simply because they are more pleasing to the eye.

The two most informative figures from [108] are reproduced in fig. 4.29, showing the differential inclusive three-jet production cross-section as a function of the leading jet p_T at LO and NLO as well as the dependence of the total inclusive three-jet cross-section on the choice of theory scales. The calculations are performed for $\sqrt{s} = 1.8\,\text{TeV}$[13] using jets clustered on parton level using the Midcone [113] and inclusive k_T algorithms [114] with $p_T > 20\,\text{GeV}$, and MC events with $\Sigma|p_T| > 80\,\text{GeV}$, which are conditions that are reasonably close to experimental practice.

Uncertainties on calculation results due to the ambiguity of scale choices are commonly evaluated by varying the scales μ_R and μ_F (cf. 1.1.3 and 1.1.4) up and down by a multiplicative factor 2. As with the choice of scale, the choice of the of variation is still arbitrary [109]. From the right plot of fig. 4.29 one can read the resulting change in total three-jet cross-section. Here, the great advantage of NLO calculation becomes striking: the NLO result has a much smaller dependence on scale variations than the LO result, which varies by $25\,\% - 50\,\%$ with respect to the chosen initial scale. Also, depending on the scale choice, the LO result may be close to the NLO value or further away from it and lie below or above it. In the left plot of fig. 4.29, scale uncertainties on the differential cross-section are shown as colored bands. Especially the NLO/LO k-factor given in the inset plot is informative: While it is true for that the choice of $\mu = \mu_R = \mu_F = \sum p_T/3$, NLO corrections to the LO prediction are of the order of $20 - 30\,\%$, the k-factor may actually vary between 0.8 and 1.4. The k-factor uncertainty is mainly due to the uncertainty of the LO calculation. In summary, cross-section results from LO QCD calculations are not reliable. Due to sizable uncertainties in the calculations, they may either agree or disagree with the more accurate NLO calculations.

With the uncertainties in mind, it is still surprising to see the remarkable agreement in normalisation between Pythia6 LO simulation and data in most of the multi-(b)-jet event classes. For all multijet classes with jet multiplicities from two to eight, normalisation agrees within $5 - 10\,\%$. This also holds true if one includes b-tagged jets and sums over all (b-)jet classes with identical total multiplicity[14] or if one moves to jet inclusive event classes.
For the Madgraph simulation samples, the normalisation difference to data is of the order of $10 - 20\,\%$ and thus significantly larger. Because of the large associated uncertainties, it is not surprising to see independent simulation samples either agree or disagree with data. Comparisons with dedicated analyses using similar simulation samples are hard to come by as the simulation normalisation is usually simply set "by hand" to match the data. The only analysis that reports the used scaling factor is the dijet resonance search [115], which uses a Pythia6 simulation sample and finds normalisation differences of the order of 30 %. However, the simulation sample used in [115] is produced using internal event reweighting to enforce the presence of sufficient numbers of simulation events in all

[13] The paper fails to include the used center-of-mass energy. The value of 1.8 TeV was obtained by private communication with the author.
[14] E.g. for a total multiplicity of 3, one would have to sum the integrals of the 3jet, 2jet+1b, 1jet+2b and 3b exclusive event classes. This ensures that deviations in normalisation are not "hidden away" by splitting the total set of multijet events into multiple subsets.

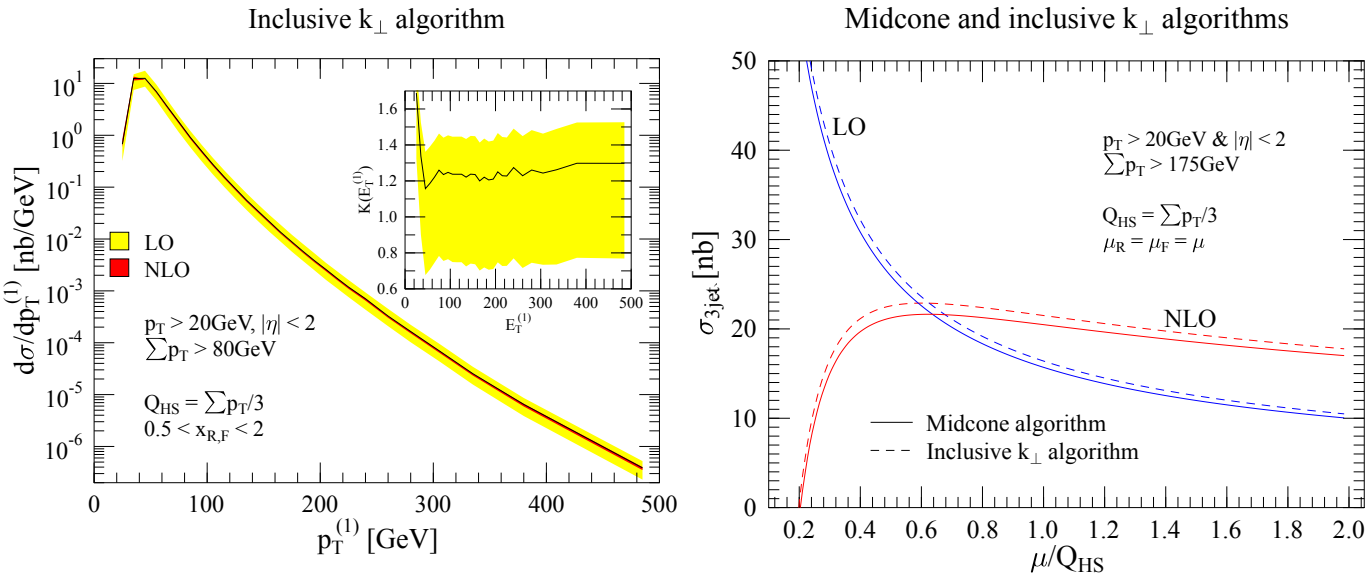


Figure 4.29: Three-jet production cross-section in proton-proton collisions at $\sqrt{s} = 1.8\,\text{TeV}$: Differential cross-section in dependence of the leading jet momentum (left) and total cross-sections depending on the choice of theory scale μ (right), to which renormalisation and factorisation scales are set. The inset in the left plot shows the k-factor, i.e. the ratio NLO/LO. Both images taken from [108].

parts of the accessible phase space[15]. It is not clear what effect the reweighting has on the resulting cross-section.

4.6 Classes with b jets

The use of b tagging as an object identification criterion adds approximately 900 event classes, and thus allows for a more multi-faceted view of SM physics processes. This section is dedicated to the discussion of some of these classes.

B jets have two main sources in the SM: QCD production in multijet topologies on the one, decays of top quarks on the other hand. Since the latter are prone to final state leptons, while the former are not, classes containing b jets are separated into non-leptonic QCD dominated and leptonic top quark dominated subsets.

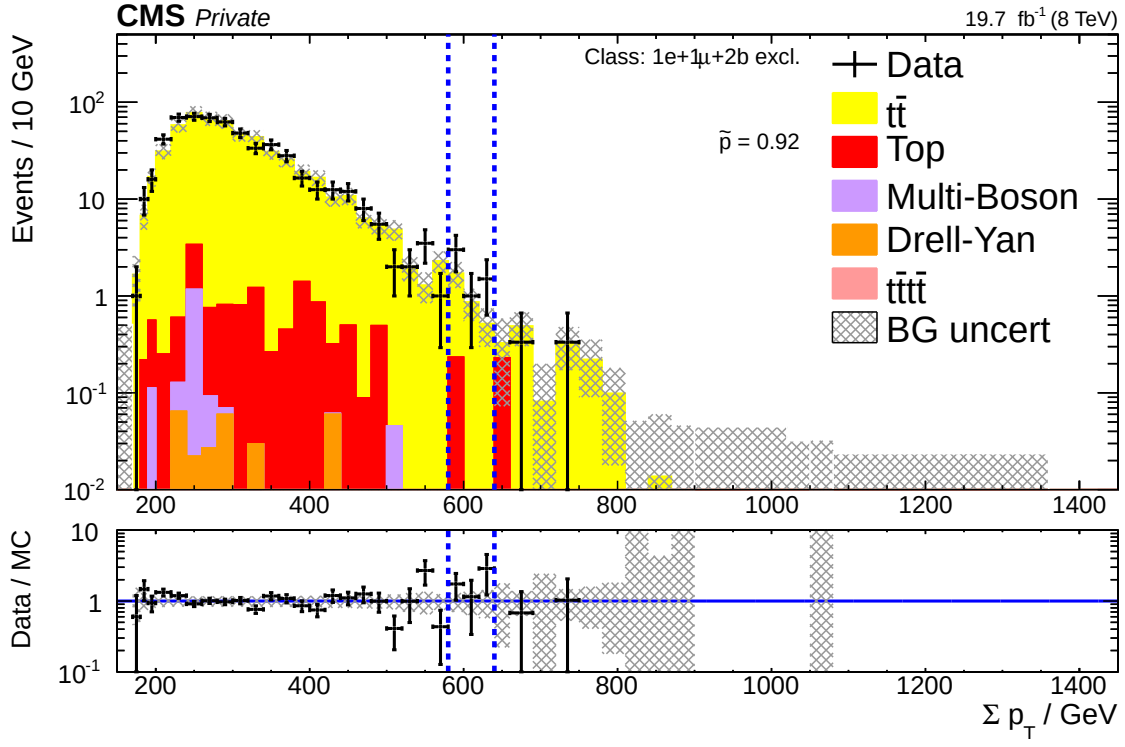
The effect and practical utility of the additional classes is most easily demonstrated with an example from the group of top quark dominated classes. Effectively, b tagging results in the separation of all event classes with finite jet multiplicity into multiple classes with varying tagging multiplicities. The final state with one electron, one muon and two jets, for example, may thus contribute to one of three classes, depending on whether zero, one, or two of the final state jets are tagged as b jets. $\Sigma|p_T|$- and M_{inv}-distributions for the two extreme cases, i.e. zero and two tagged jets, are shown in fig. 4.30 and 4.31. The color-coded physics process information in the distribution plots immediately highlights the effect of the b tagging information: Both event classes are top quark dominated, but the non-top SM contributions are reduced from $\mathcal{O}(10\,\%)$ to $\mathcal{O}(1\,\%)$ by requiring two b jets. This increases the analysis sensitivity to both SM top quark process modeling and BSM top-like phenomena such as fourth generation quarks. This is especially true in the high-$\Sigma|p_T|/M_{\text{inv}}$ regime, where the SM contribution is dominated by non-top processes. Classes dominated by events from processes involving top quarks generally show good agreement between simulation and data. Reliable NNLL and NNLO calculations are available, which results in small background modeling uncertainties and allows for a high precision assessment of SM consistency. A striking example of the enhanced sensitivity as a

[15] In the samples used in this analysis, this function is performed by kinematic binning (cf. sec. 2.4.1).

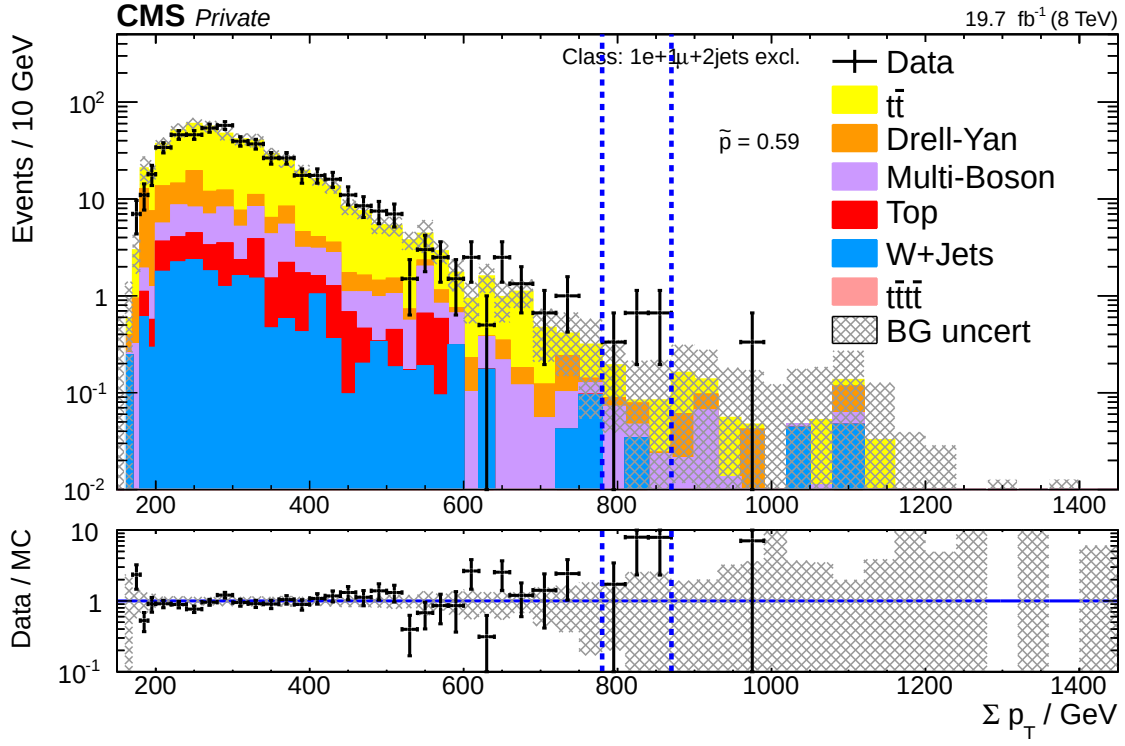
result of small uncertainties is visible in the RoI found in the M_{inv}-distribution of the 1e+1μ+2b event exclusive event class (fig. 4.31 (b)). While not very significant, fluctuations like this one immediately catch the human eye even if no simulation were available because the surrounding data points show a much smaller bin-to-bin variation.

Classes containing b jets but no leptons are very similar to the multijet topologies discussed in sec. 4.5, as the underlying perturbative QCD interactions are symmetric w.r.t. to quark flavor. However, due to finite tagging efficiencies and since a dominant part of light jets is gluon-initiated, b tagging reduces the expected cross-section by a larger factor than a naive $1/N_{\text{quark}}$ per final state b jet. The $\Sigma|p_T|$- and M_{inv}-distributions of the 2b and 3b excl. classes are shown in fig. 4.32 and 4.33. In fact, a close resemblance to the distributions of the multijet distributions is visible. The number of events in the 2b exclusive class is reduced by a factor ≈ 200 w.r.t. the 2jet exclusive class. For low $\Sigma|p_T|$, the 2b exclusive class exhibits a shape mismatch between data and simulation similar to that discussed in 4.5.2. Due to limited simulation sample statistics and the reduced cross-section for the b jet final state, no direct comparison to NLO is possible.

In comparing the kinematic distributions of the 3b exclusive class to those of the 2b exclusive class, two effects are eminent: The number of events in the class rapidly diminishes and normalisation differences between simulation and data become more prominent. With increasing b jet multiplicities, residual efficiency differences between simulation and data are exposed, which are however still covered by combined normalisation uncertainty of process cross-section and b tagging efficiency (cf. sec. 2.4.4.1 and 2.4.4.2).

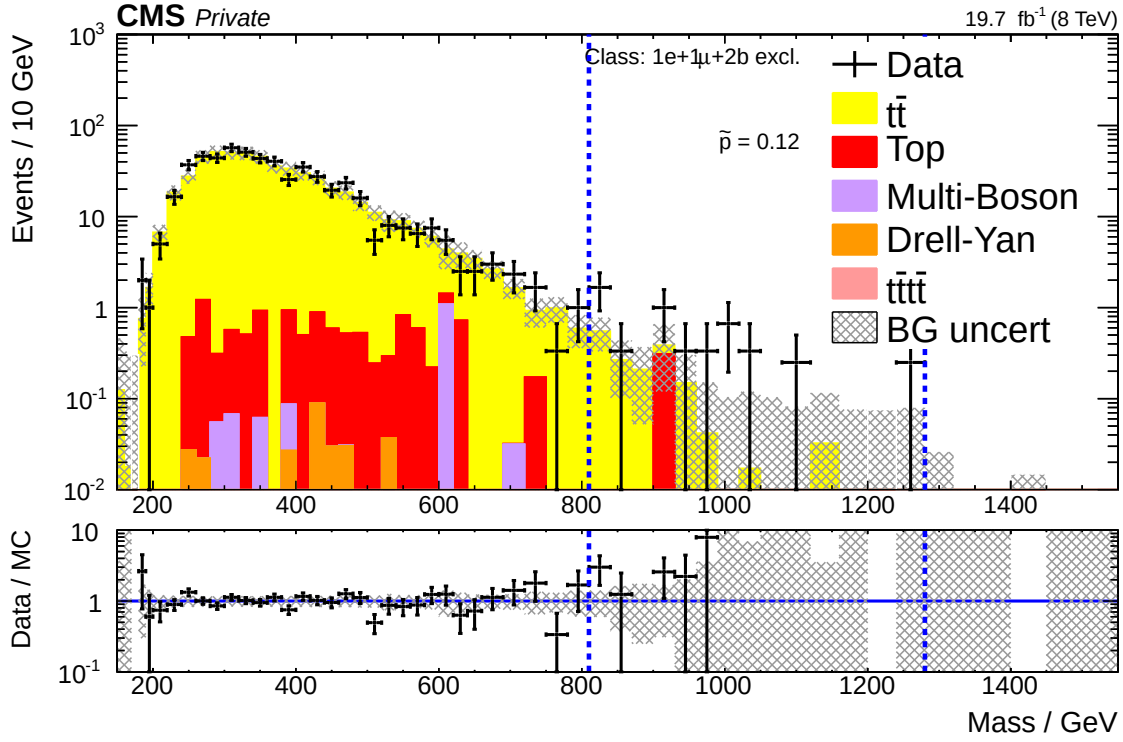


(a)

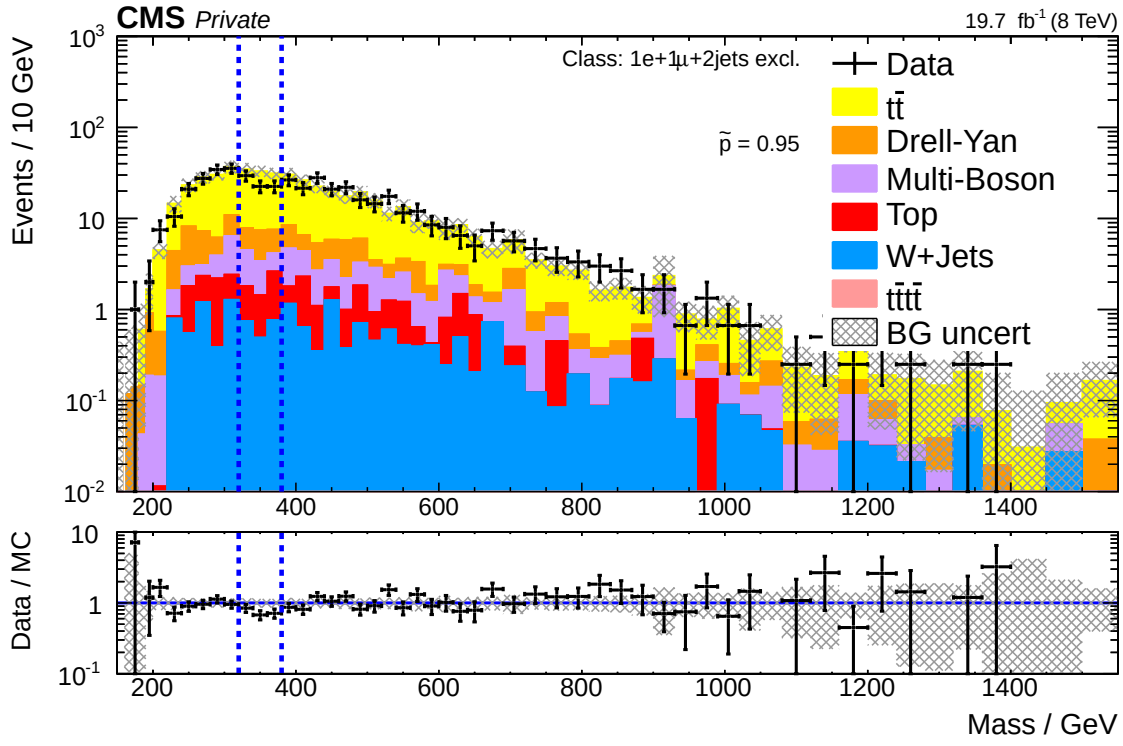


(b)

Figure 4.30: $\Sigma|p_T|$-distributions of the 1e+1μ+2b (a) and 1e+1μ+2jet (b) excl. classes, the final states of which only differ in their b tagging multiplicity. The application of b tagging succeeds in separating top quark processes from others, reducing the non-top contributions from $\mathcal{O}(10\,\%)$ to $\mathcal{O}(1\,\%)$. The restriction to top-related physics also constrains the $\Sigma|p_T|$-region containing SM contributions, and increases sensitivity to possible BSM top-like phenomena at smaller energy scales.

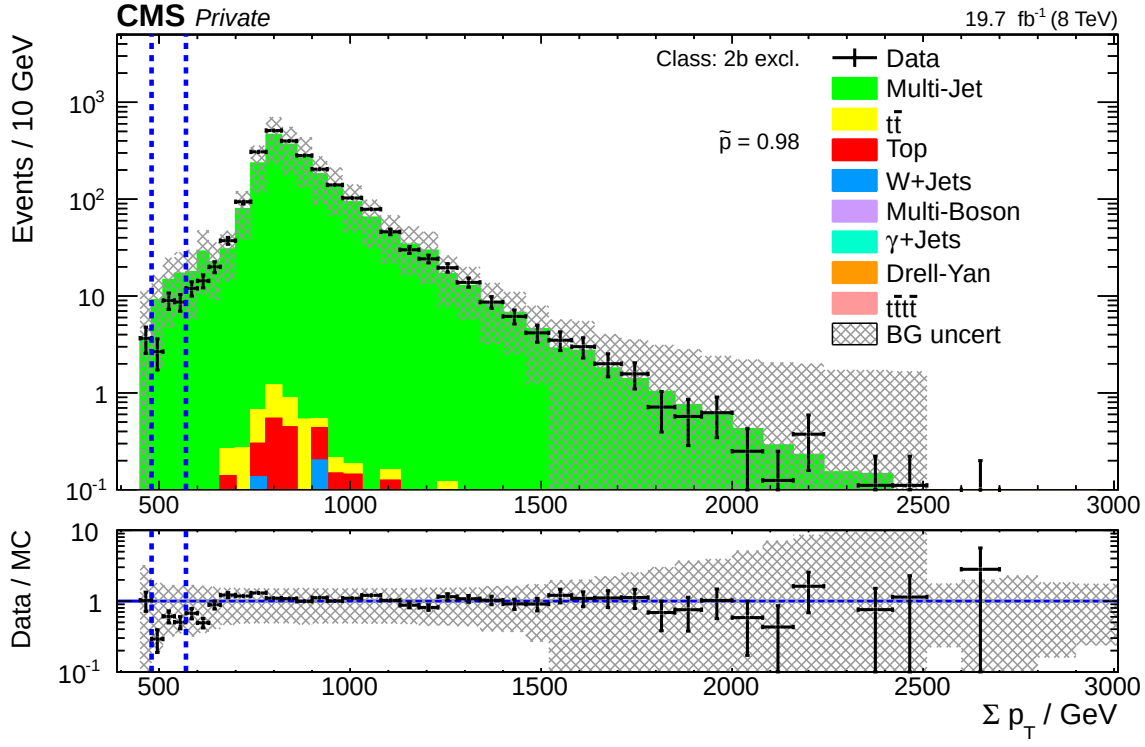


(a)

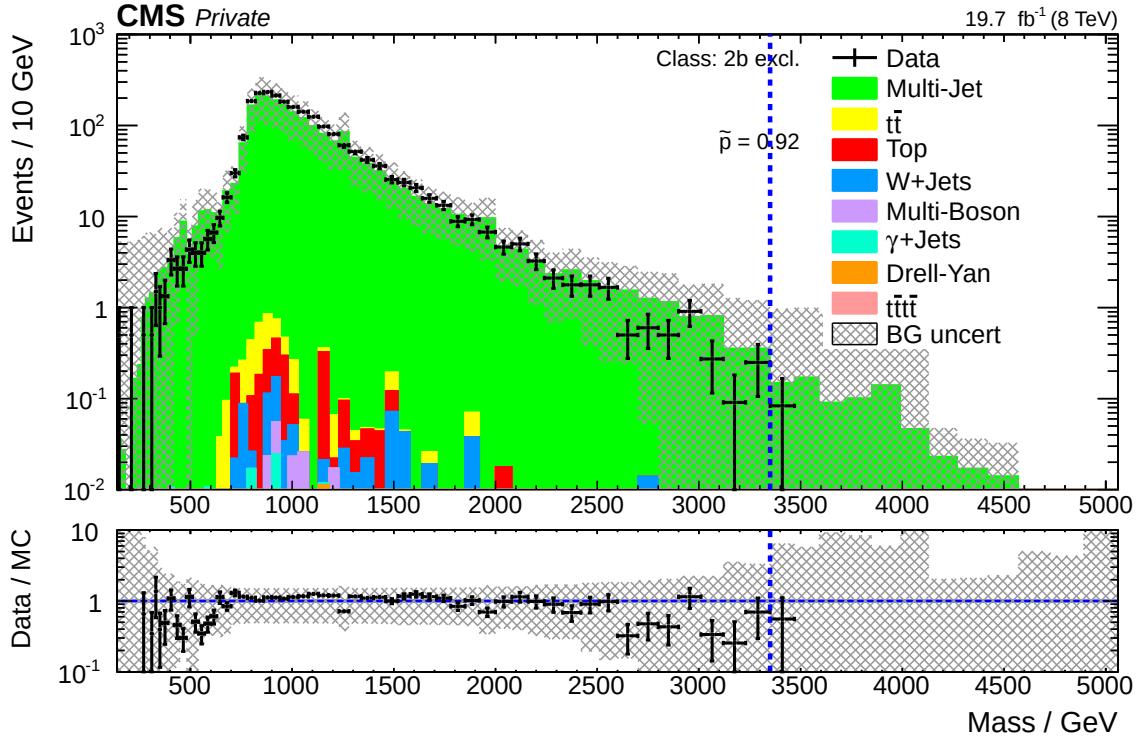


(b)

Figure 4.31: M_{inv}-distributions of the 1e+1μ+2b (a) and 1e+1μ+2jet (b) excl. classes. As with the $\Sigma|p_T|$-distribution, the application of b tagging visibly reduces the SM contributions at high M_{inv}, while hardly reducing the top quark contributions, thus enhancing sensitivity to top-like BSM phenomena. Fig. (b) is also a pleasant example of the MUSiC scan algorithm finding a locally significant region coinciding with what analysts commonly envision statistical fluctuations to look like.

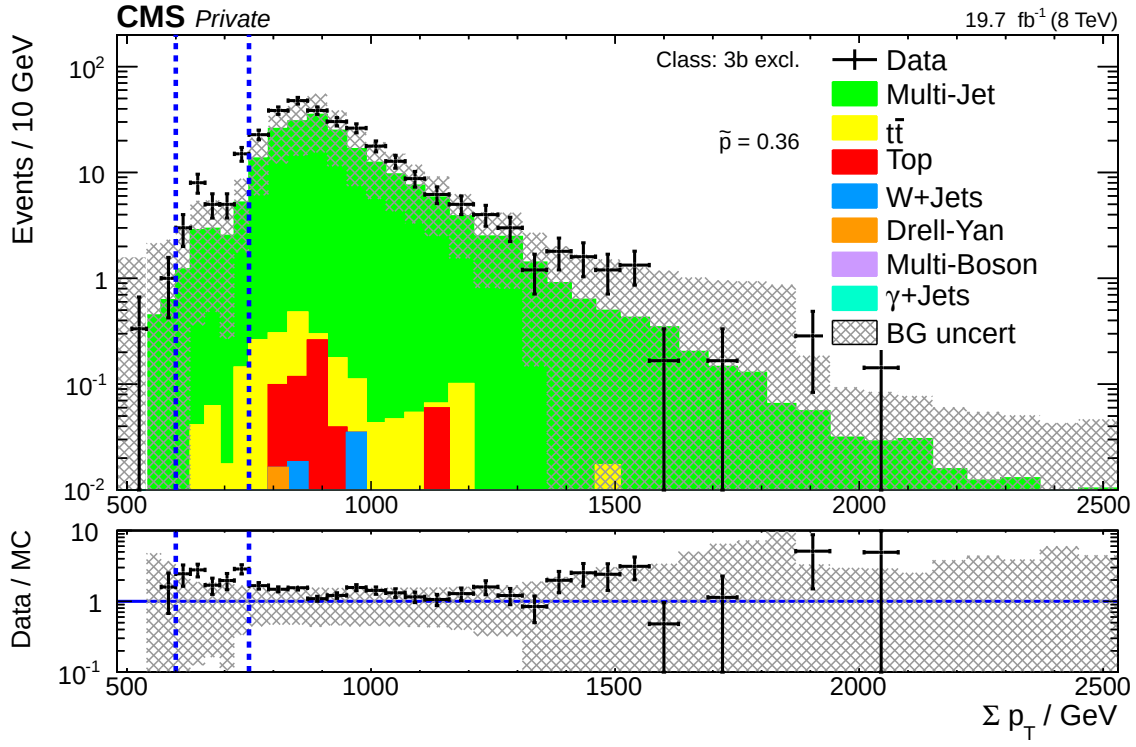


(a)

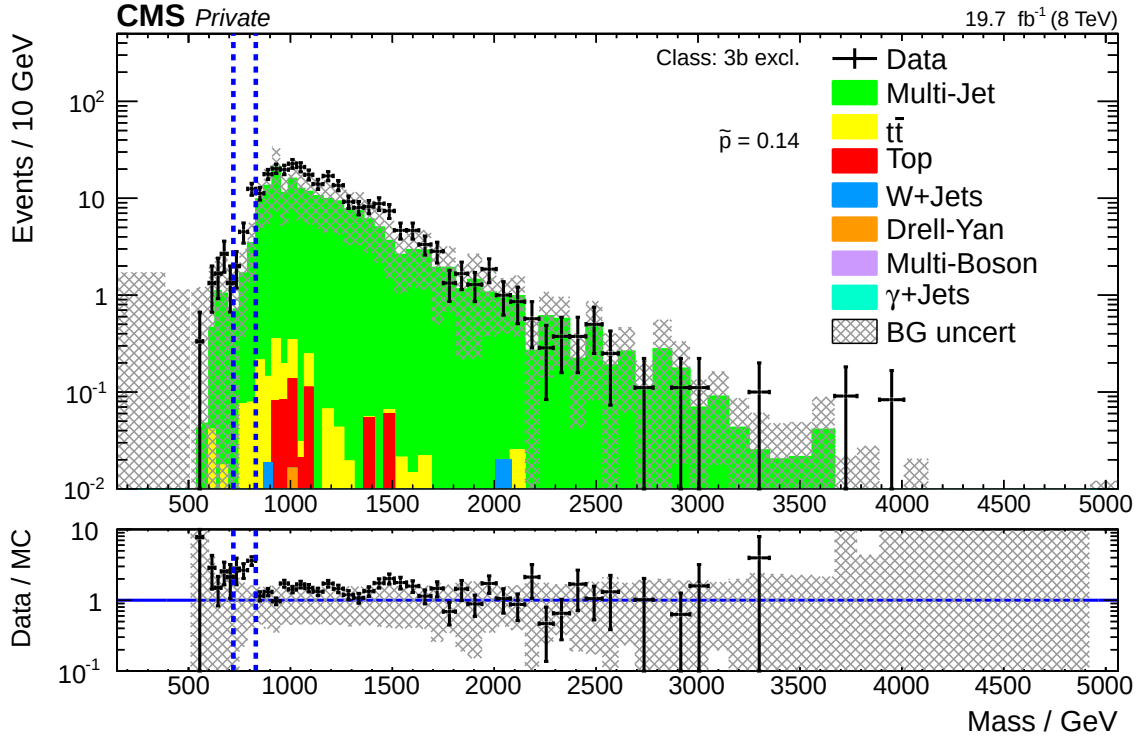


(b)

Figure 4.32: $\Sigma|p_T|$- and M_{inv}-distributions of the 2b excl. event class (a and b, respectively). The spectra are similar to the multijet spectra without tagged b jets. Due to decreased cross-section for b-only jets and limited tagging efficiency, the distribution integral amounts to approximately 0.5 % of that of the 2jet excl. class.



(a)



(b)

Figure 4.33: $\Sigma|p_T|$- and M_{inv}-distributions of the 3b excl. event class (a and b, respectively).

4.7 Sensitivity Study

In order to assert the analysis sensitivity to a certain model, a signal simulation sample may be tested. The sensitivity study is performed in three steps:

- **Simulation:** For a certain signal model, a simulation sample is added to the ensemble of SM simulation samples. If the model contains parameters that can be varied, separate samples may be used for different parameter values.
- **Dicing:** To account for the influence of all simulation uncertainties, the SM+BSM simulation set is randomised ("diced") within its uncertainties 100 times. This is performed in all distributions in the same manner as for the pseudo-experiments used for calculation of $\tilde{p}$. For each signal dicing round, the RoI-algorithm is run to find the most significant deviation between the SM and SM+BSM simulation sets. The median $p_{1/2}$ of the set of 100 calculated p-values is determined and used as a representative value for this signal model in this distribution. The signal shape corresponding to the median p-value is used for plotting.
- **Look-Elsewhere Effect:** As in data, the p-value is converted to a $\tilde{p}$-value via generation of pseudo-experiments (cf. sec. 1.3.2.5) . Up to 10^5 pseudo-experiments are performed. The resulting $\tilde{p}$-values are histogrammed and thus allow for an at-a-glance comparison of the size and number of deviations caused by a potential signal.

4.7.1 Dark Matter

One of the main focuses for BSM searches at the LHC is search for Dark Matter (DM). The existence of DM is inferred from astrophysical phenomena, such as the rotational velocity in galaxies [116]. The radial dependence of the rotational velocity is inconsistent with descriptions based solely on the gravitational effects of the SM matter in the universe. Discrepancies can be accounted for by positing the existence of an additional kind of matter that is not detected by any known telescopic technique and accordingly called Dark Matter. The requirement of "invisibility" imposes upper limits on the coupling to the known elementary forces.

If there is a coupling between DM and SM particles, DM particles can potentially be created in LHC collisions. Due to the limited interactions with SM matter, DM particles and their decay products are not expected to interact with detector elements, thus leaving behind no direct signature. However, if any additional objects are created by initial state radiation, intrinsic E_T^{miss} will be produced, which may then be searched for. This search strategy is employed for different additional particle signatures such as single and double leptons [117, 118], photons [119] or jets [89].

4.7.1.1 Signal Modeling

In [89], a monojet+E_T^{miss} signature is used to search for DM in CMS data at $\sqrt{s} = 8\,\text{TeV}$. An Effective Field Theory (EFT) approach following [120] is used to model DM pair production. In an EFT, the particle mediating between SM matter and DM is assumed to have a high mass compared to the energy scale of the interaction, which gives rise to an effective four-fermion vertex (fig. 4.34). In this case, interaction kinematics are independent of the actual mediator mass and a kinematically flat cross-section suppression factor Λ^{-4} is introduced, which summarizes constants in the cross-section expression. The final state kinematics then only depend on the mass of the DM particle M_χ and the parity properties of the mediator, which is considered to be either a scalar, a vector or an axial-vector particle.

The signal simulation samples used in [89] are centrally available in CMS. An EFT implementation of DM production in MADGRAPH is used to produce separate samples for M_χ between 1 and 700 GeV, for vector and axial-vector couplings and for initial state up- and down-quarks. In all samples, parton showering and hadronisation are performed by PYTHIA6. For this sensitivity study, the 1 GeV (*light*) and 400 GeV (*heavy* DM) mass points with a vector-like mediator are considered. The mediator is assumed to couple to both up and down quarks. The EFT scale Λ is scanned in steps of 100 GeV in the range $500 - 900\,\text{GeV}$.

4.7.1.2 Results

The resulting signal shapes in the 1jet+E_T^{miss} event class are shown in fig. 4.35 and 4.36 for $M_\chi = 1\,\text{GeV}$ and 400 GeV, respectively. The choice of M_χ does not have a significant impact on the signal shape within the sensitivity of this analysis. The distributions are shown for $\Lambda = 600$ and 500 GeV (fig. 4.35 and 4.36, top and bottom, respectively), highlighting the strong influence of Λ on the signal cross-section. An analogous view of the 3jet+E_T^{miss} event class is shown in fig. 4.37 for light DM. The signal branches into this and other multijet+E_T^{miss} event classes mainly because of initial state radiation. While less dominant than in the 1jet+E_T^{miss} event class, the signal is still clearly discernible.

Details of the ten classes showing the most significant median $\tilde{p}$ values for each value of Λ are given in tab. 4.7 and 4.8. The onset of signal sensitivity with decreasing Λ is evident from the composition of most significant event classes. For high Λ, the most significant event classes are essentially randomly comprised, containing various final state objects. With decreasing Λ and thus increasing signal cross-section, final states containing leptons and photons vanish from the list, and their places are taken by multijet+E_T^{miss} classes. The 1jet+E_T^{miss} event class is commonly leading in significance because of its low SM expectation, which is a factor of 10^3 smaller than that of the 2jet+E_T^{miss} class. While the onset of sensitivity is also visible in the given $\tilde{p}$-values, it is best summarized in the *Overflow Fraction* (OF), which, for a given event class, gives the fraction of signal dicing rounds where a value $\tilde{p} < 1 \cdot 10^{-5}$ was calculated. The value of OF thus gives the contribution of a given class to the overflow bin in fig. 4.38 and is somewhat superior to $\tilde{p}$ as it takes into account the statistical fluctuation of the diced signal distributions.

The resulting $\tilde{p}$ distributions are shown in fig. 4.38 for the cases of $M_\chi = 1\,\text{GeV}$ and $M_\chi = 400\,\text{GeV}$ (left and right, respectively) and scales of $\Lambda = 900$, 600, 400 GeV from top to bottom. In this comparative view, the "turning on" of the signal by reduction of the suppression scale is easily discernible. While the SM+BSM $\tilde{p}$ distribution agrees well with the expected SM distribution for a high suppression scale of $\Lambda = 900\,\text{GeV}$, an increasing deviation in the high-significance tail of the distribution is visible with decreasing Λ. Going from top to bottom, the overflow fraction increases by a factor of 10 from 0.4 to 4 for light 0.6 to 2.6 for heavy DM.

The dedicated analysis observes 90 % CL exclusion limits of $\Lambda = 897\,\text{GeV}$ and 779 GeV, for $M_\chi = 1\,\text{GeV}$ and 400 GeV, respectively [89]. A direct comparison is not trivial because MUSiC is not designed to calculate such a number. In terms of sensitivity, the three chosen values of Λ fall in three different categories. While MUSiC is simply not sensitive to the signal in the high-suppression case, the medium- and low-suppression cases highlight the importance of the large number of event classes. For $\Lambda = 600\,\text{GeV}$, many of the sensitive event classes are individually not extremely significant, while only one or two classes exhibit extreme deviations. However, the list of most significantly deviating classes would certainly raise suspicion as there is only a narrow choice of final states listed. In the low-suppression case of $\Lambda = 500\,\text{GeV}$, the coincidence between different event classes becomes less important as many single event classes show extremely significant excesses.

In sum, MUSiC is sensitive to the monojet+E_T^{miss} signature for EFT suppression scales of the order of 0.6 TeV, compared to exclusion limits of up to 0.9 TeV from the dedicated analysis.

Table 4.7: List of the ten event classes with smallest $p_{1/2}$, i.e. the most significant median deviations between SM and SM+Signal, in a scan of the E_T^{miss}-distributions for each $\Lambda = 900,\ 600,\ 500\,\text{GeV}$. The tables are sorted by $\tilde{p}$, which is calculated with respect to $p_{1/2}$. OF denotes the *Overflow Fraction*, i.e. the fraction of signal dicing rounds yielding $\tilde{p} < 1.0 \cdot 10^{-5}$ in a given class. N_{Signal} and N_{SM} are the numbers of SM+Signal and SM events in the RoI in the dicing round corresponding to $p_{1/2}$.

$\Lambda = 900\,\text{GeV}$

No.	Event Class	$p_{1/2}$	$\tilde{p}$	OF(%)	$N_{\text{SM+Signal}}$	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
1	2e+4jets+1b+E_T^{miss}	0.017	0.38	0	1.0	7.4 ± 1.8
2	2μ+2jets+1b+E_T^{miss}	0.015	0.38	0	204.0	297.9 ± 40.4
3	2e+4jets+E_T^{miss}	0.017	0.39	1	27.0	14.8 ± 3.2
4	2e+1μ+E_T^{miss}	0.047	0.39	0	7.0	26.3 ± 11.5
5	2e+3b+E_T^{miss}	0.034	0.4	0	6.0	2.0 ± 0.8
6	1e+1γ+2b+E_T^{miss}	0.045	0.41	0	19.0	9.0 ± 4.1
7	1μ+6jets+E_T^{miss}	0.013	0.41	0	149.0	106.4 ± 15.1
8	1μ+1γ+4jets+1b+E_T^{miss}	0.059	0.41	0	1.0	0.0 ± 0.06
9	3μ+1jet+E_T^{miss}	0.046	0.41	0	1.0	0.05 ± 0.04
10	1e+1μ+4jets+1b+E_T^{miss}	0.02	0.41	1	2.0	0.1 ± 0.2

$\Lambda = 600\,\text{GeV}$

No.	Event Class	$p_{1/2}$	$\tilde{p}$	OF(%)	$N_{\text{SM+Signal}}$	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
1	1jet+E_T^{miss}	$2.91 \cdot 10^{-9}$	$< 10^{-5}$	61	45.0	10.3 ± 2.9
2	1b+E_T^{miss}	$1.59 \cdot 10^{-4}$	0.02	9	106.0	55.7 ± 10.7
3	2jets+E_T^{miss}	$3.33 \cdot 10^{-9}$	0.024	26	49.0	12.3 ± 3.1
4	4jets+E_T^{miss}	$1.93 \cdot 10^{-4}$	0.068	7	12.0	1.6 ± 1.3
5	3jets+E_T^{miss}	$6.71 \cdot 10^{-6}$	0.11	5	19.0	2.8 ± 1.8
6	1jet+1b+E_T^{miss}	$9.38 \cdot 10^{-4}$	0.12	3	3.0	0.0 ± 0.18
7	2b+E_T^{miss}	0.0058	0.18	2	1.0	0.0 ± 0.008
8	2jets+1b+E_T^{miss}	0.0012	0.29	1	31.0	10.5 ± 4.5
9	1μ+1γ+1jet+3b+E_T^{miss}	0.092	0.35	0	1.0	0.03 ± 0.1
10	1e+1μ+5jets+1b+E_T^{miss}	0.032	0.37	0	8.0	3.2 ± 1.1

$\Lambda = 500\,\text{GeV}$

No.	Event Class	$p_{1/2}$	$\tilde{p}$	OF(%)	$N_{\text{SM+Signal}}$	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
1	3jets+E_T^{miss}	$7.11 \cdot 10^{-15}$	$< 10^{-5}$	69	106.0	25.6 ± 6.2
2	2jets+E_T^{miss}	$2.33 \cdot 10^{-21}$	$< 10^{-5}$	76	103.0	15.1 ± 4.7
3	1jet+E_T^{miss}	$2.29 \cdot 10^{-24}$	$< 10^{-5}$	96	95.0	13.2 ± 3.8
4	1b+E_T^{miss}	$5.95 \cdot 10^{-8}$	$2.00 \cdot 10^{-5}$	45	189.0	54.8 ± 11.2
5	4jets+E_T^{miss}	$2.50 \cdot 10^{-8}$	$6.00 \cdot 10^{-5}$	41	64.0	7.5 ± 3.0
6	1jet+1b+E_T^{miss}	$3.00 \cdot 10^{-6}$	0.002	15	7.0	0.2 ± 0.3
7	2b+E_T^{miss}	$5.06 \cdot 10^{-4}$	0.046	08	19.0	3.4 ± 3.1
8	2jets+1b+E_T^{miss}	$1.33 \cdot 10^{-4}$	0.083	07	25.0	8.0 ± 2.6
9	6jets+2b+E_T^{miss}	0.0056	0.15	06	7.0	1.5 ± 0.8
10	3jets+1b+E_T^{miss}	0.0022	0.2	01	28.0	11.0 ± 3.8

Table 4.8: List of the ten event classes with smallest $p_{1/2}$ for $M_\chi = 400\,\text{GeV}$.

	No.	Event Class	$p_{1/2}$	$\tilde{p}$	OF(%)	$N_{\text{SM+Signal}}$	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\Lambda = 900\,\text{GeV}$	1	3jets+E_T^{miss}	$8.31 \cdot 10^{-4}$	0.39	0	2.0	0.0 ± 0.04
	2	1γ+3jets+E_T^{miss}	0.019	0.4	0	2.0	0.0 ± 0.22
	3	1μ+9jets+E_T^{miss}	0.14	0.41	0	2.0	0.7 ± 0.5
	4	1e+1μ+1γ+2jets+E_T^{miss}	0.13	0.41	0	1.0	0.1 ± 0.1
	5	3μ+1jet+1b+E_T^{miss}	0.1	0.42	0	2.0	0.6 ± 0.2
	6	1e+1γ+E_T^{miss}	0.027	0.42	0	6.0	1.5 ± 1.2
	7	1e+1γ+2jets+E_T^{miss}	0.031	0.42	0	24.0	10.8 ± 5.3
	8	1jet+4b+E_T^{miss}	0.039	0.43	0	1.0	9.9 ± 5.0
	9	1e+1μ+3jets+2b+E_T^{miss}	0.024	0.43	0	3.0	0.5 ± 0.3
	10	1b+E_T^{miss}	0.008	0.43	1	4.0	0.3 ± 0.5

	No.	Event Class	$p_{1/2}$	$\tilde{p}$	OF(%)	$N_{\text{SM+Signal}}$	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\Lambda = 600\,\text{GeV}$	1	1jet+E_T^{miss}	$2.34 \cdot 10^{-5}$	0.0044	19	122.0	58.6 ± 12.0
	2	3jets+E_T^{miss}	$1.59 \cdot 10^{-4}$	0.27	1	21.0	5.8 ± 1.9
	3	2jets+E_T^{miss}	$5.37 \cdot 10^{-5}$	0.3	3	19.0	4.2 ± 1.8
	4	1jet+1b+E_T^{miss}	0.0058	0.33	2	5.0	0.2 ± 0.8
	5	4jets+E_T^{miss}	0.0035	0.37	1	27.0	11.7 ± 3.3
	6	2jets+1b+E_T^{miss}	0.002	0.37	1	$1.2 \cdot 10^4$	$(0.0 \pm 4.9) \cdot 10^6$
	7	1jet+2b+E_T^{miss}	0.012	0.37	3	9.0	1.3 ± 2.1
	8	1e+1γ+2b+E_T^{miss}	0.045	0.38	0	14.0	6.3 ± 3.0
	9	1e+4b+E_T^{miss}	0.097	0.38	0	1.0	0.1 ± 0.2
	10	1μ+3jets+1b+E_T^{miss}	0.012	0.4	0	193.0	290.0 ± 40.0

	No.	Event Class	$p_{1/2}$	$\tilde{p}$	OF(%)	$N_{\text{SM+Signal}}$	$N_{\text{SM}} \pm \sigma_{\text{SM}}$
$\Lambda = 500\,\text{GeV}$	1	1jet+E_T^{miss}	$2.06 \cdot 10^{-13}$	$< 1/100000$	78	71.0	7.2 ± 2.4
	2	2jets+E_T^{miss}	$1.02 \cdot 10^{-10}$	0.0035	39	0.0	58.3 ± 9.1
	3	1b+E_T^{miss}	$5.54 \cdot 10^{-5}$	0.0097	13	43.0	16.1 ± 4.5
	4	3jets+E_T^{miss}	$1.24 \cdot 10^{-8}$	0.029	28	69.0	20.5 ± 5.2
	5	4jets+E_T^{miss}	$2.02 \cdot 10^{-4}$	0.07	7	27.0	8.8 ± 2.9
	6	1jet+1b+E_T^{miss}	$3.84 \cdot 10^{-4}$	0.072	4	28.0	9.5 ± 3.5
	7	2jets+1b+E_T^{miss}	$2.77 \cdot 10^{-4}$	0.11	3	9.0	0.04 ± 1.3
	8	6jets+E_T^{miss}	$9.36 \cdot 10^{-4}$	0.14	2	2.0	0.0 ± 0.04
	9	2b+E_T^{miss}	0.0074	0.22	2	522.0	233.7 ± 118.8
	10	5jets+E_T^{miss}	$6.06 \cdot 10^{-4}$	0.24	2	7.0	0.9 ± 0.6

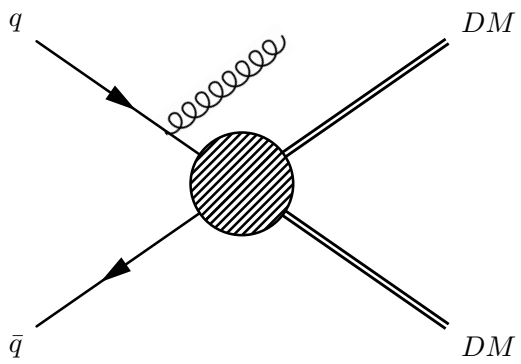


Figure 4.34: Dark Matter pair production via contact interaction. If the mass of the mediating particle is large compared to the LHC energy scale, the details of interaction within the circle become irrelevant. While the Dark Matter particles escape undetected, the ISR jet gives rise a jet+E_T^{miss} signature which may be searched for. The diagram is taken from [89].

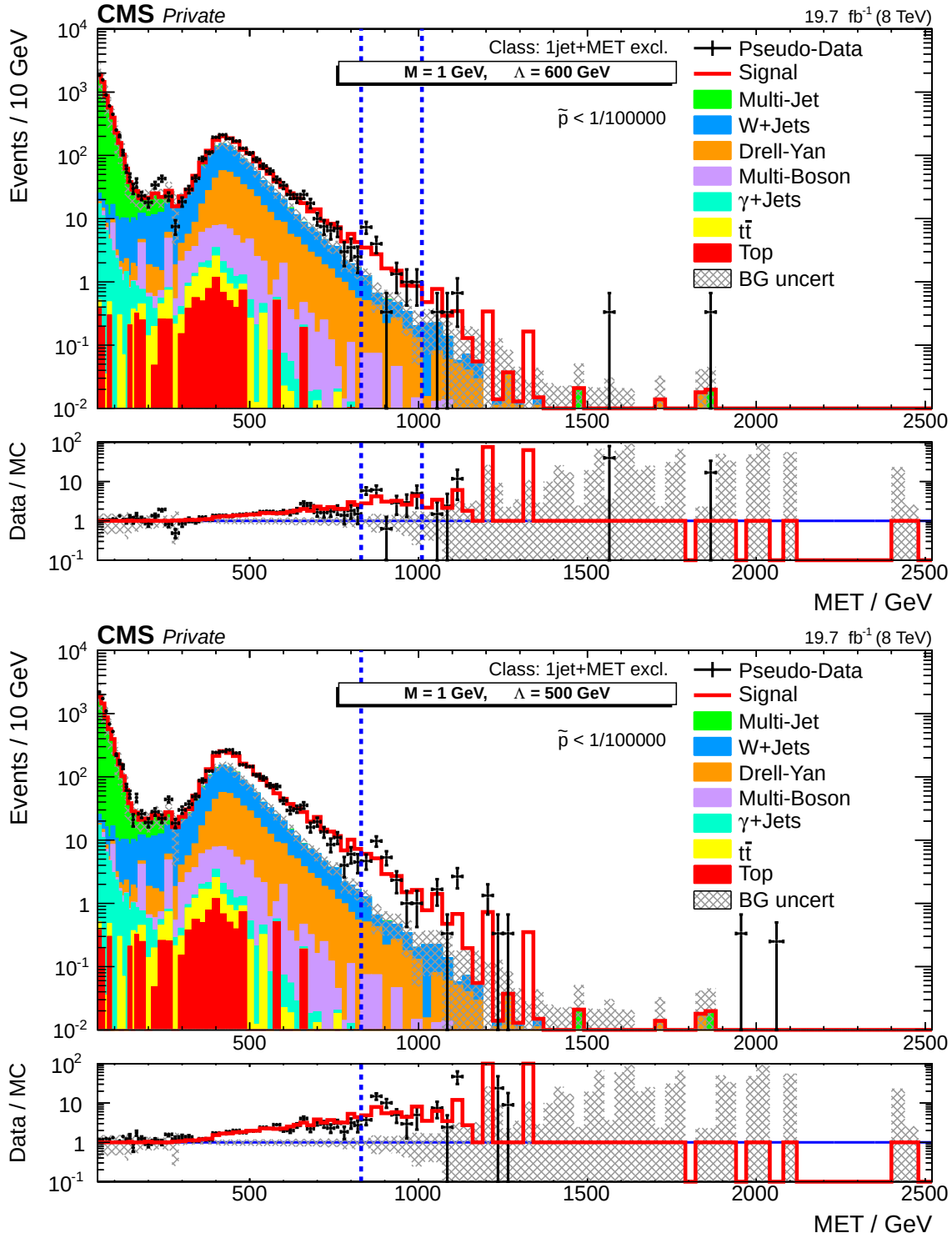


Figure 4.35: E_T^{miss}-distributions of the 1jet+E_T^{miss} event classes with injected DM signal for $\Lambda = 600\,\text{GeV}$ (top) and $500\,\text{GeV}$ (bottom) and constant $M_\chi = 1\,\text{GeV}$. As in runs with data, solid histograms represent the SM simulation, with color coding referring to different physics processes. The red solid line represents the nominal, i.e. non-diced, distribution of the signal on top of the SM "background". The black points represent the median significance outcome of the pseudo-experiments, i.e. the dicing round that gives the median p-value $p_{1/2}$ in this distribution.

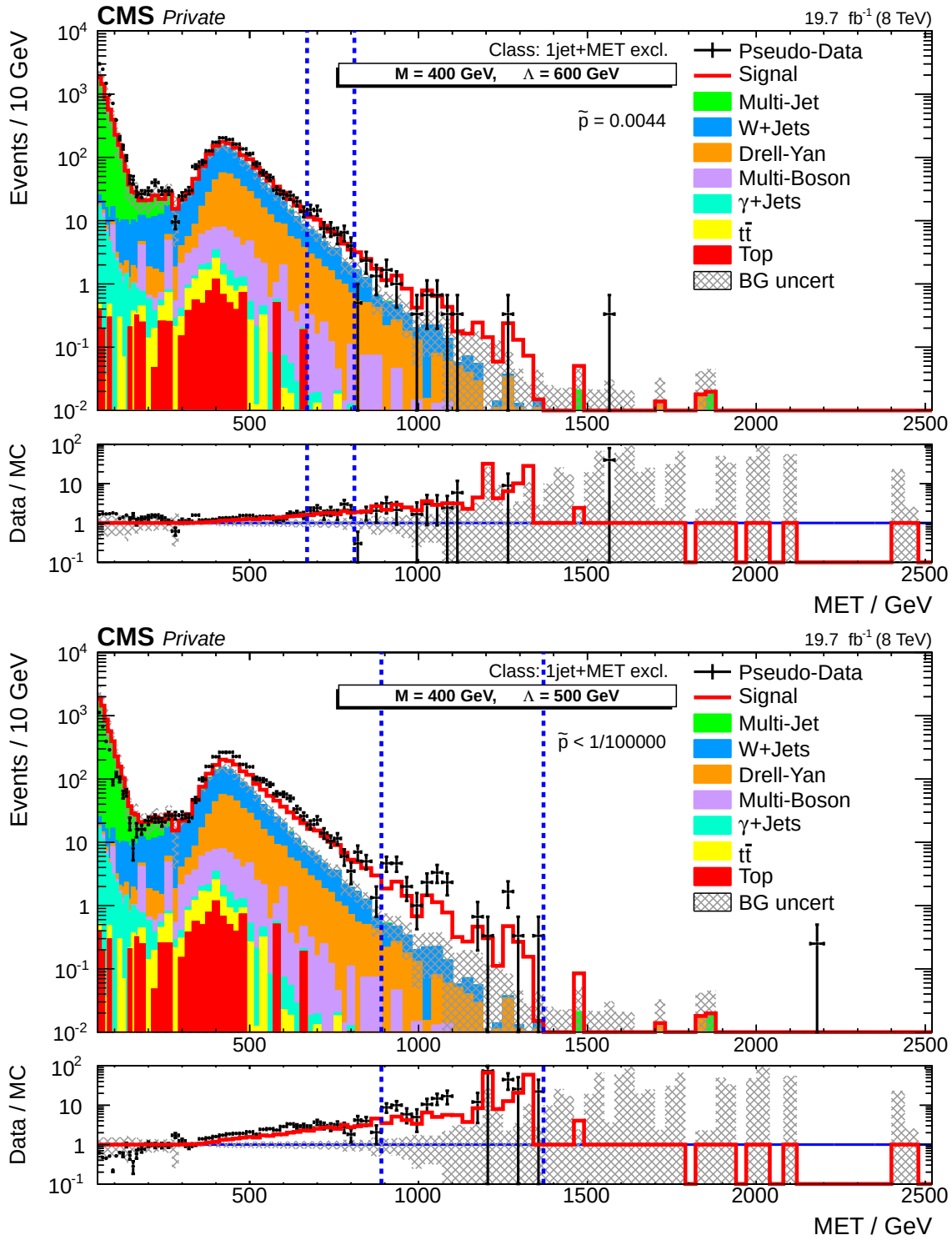


Figure 4.36: E_T^{miss}-distributions of the 1jet+E_T^{miss} event classes with injected DM signal for $\Lambda = 600\,\text{GeV}$ (top) and $500\,\text{GeV}$ (bottom) and constant $M_\chi = 400\,\text{GeV}$. While the signal cross-section is reduced compared with the low-M_χ case (cf. fig. 4.35), the signal shape does not seem to vary strongly.

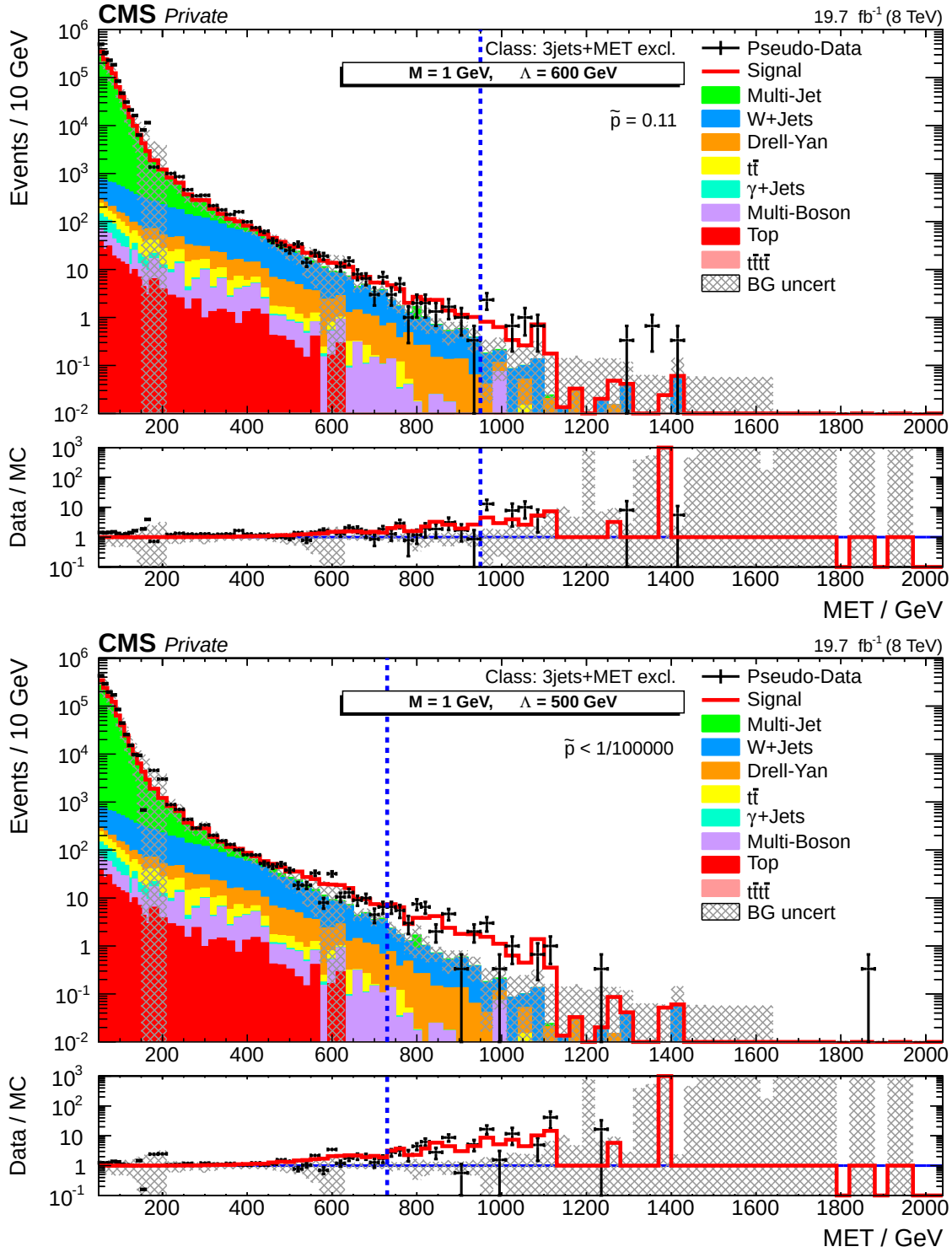


Figure 4.37: E_T^{miss}-distributions of the 3jet+E_T^{miss} event classes with injected DM signal for $\Lambda = 600\,\text{GeV}$ (top) and $500\,\text{GeV}$ (bottom) and constant $M_\chi = 1\,\text{GeV}$.

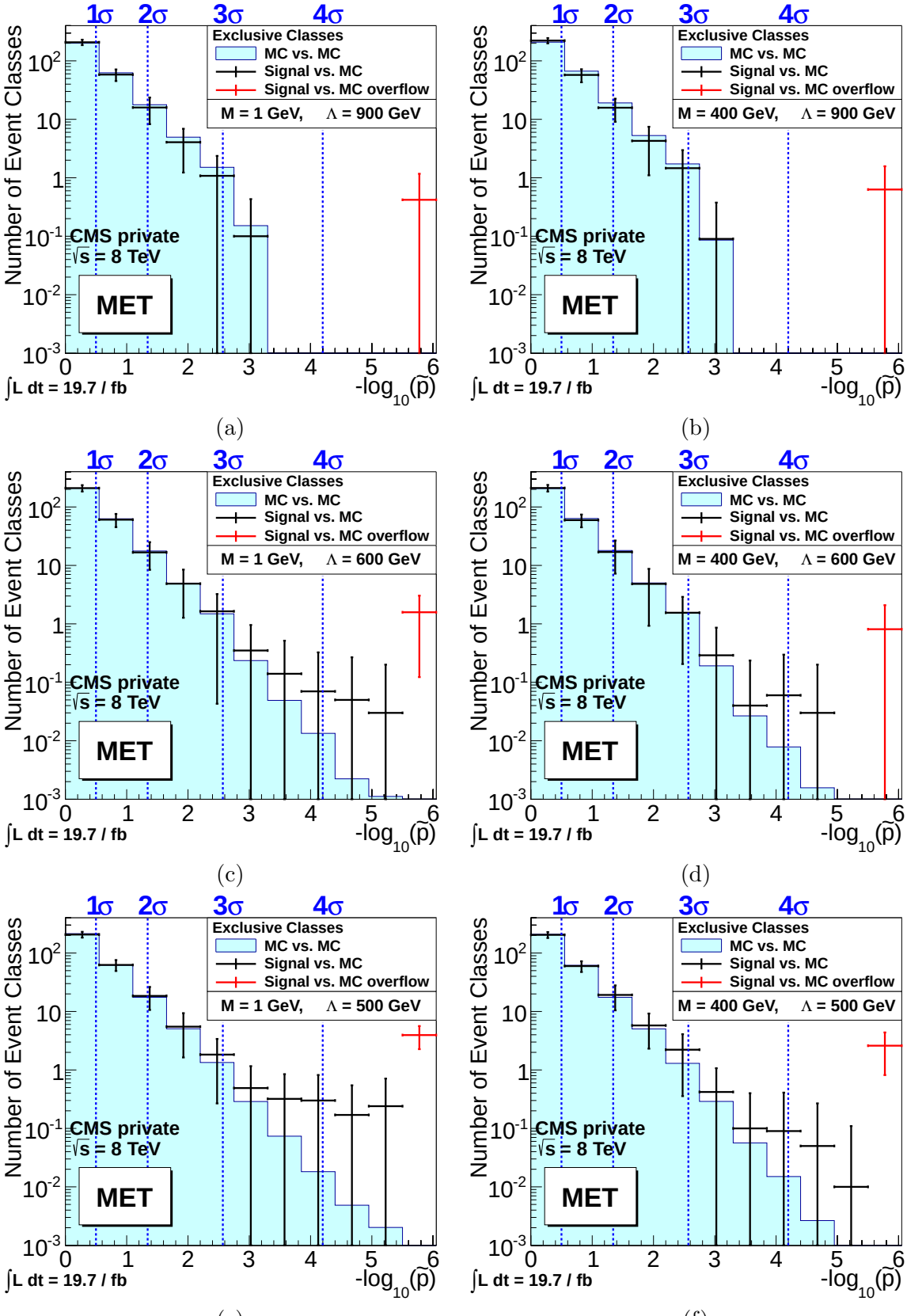


Figure 4.38: Distributions of $\tilde{p}$ for E_T^{miss}-distribution scans with injected DM signal for exclusive (left) and inclusive (right) event classes. From top to bottom, the EFT suppression scale Λ is reduced from initially 900 GeV (a, b) to 600 GeV (c, d) and finally 500 GeV (e, f).

5 Summary and Conclusion

The MUSiC analysis of the complete data set recorded by CMS in 2012 was presented. Integral and kinematic scans of approximately three thousand event classes were performed and good agreement between simulation and data was found in most cases. Small bump-like features were found in multiple final states, highlighting the advantages of the MUSiC approach. Significant excesses were found in the M_{inv}-distributions of the 1e1μ1b+X and related event classes. Comparisons with the CMS data set at $\sqrt{s} = 13\,\text{TeV}$ will reveal if the excesses are signs of physics beyond the Standard Model or statistical fluctuations. In many different event classes, the modeling of W boson production was found to disagree with data. Further studies are needed to understand the cause of the problem and possibly alleviate it.

Jet triggered event classes were included in the analysis for the first time. Jet triggering allows for the analysis of 1.1×10^7 additional data events, which corresponds to approximately 10 % of all events. Jet triggered event classes were found to be well modeled in most of the accessible phase space. No significant deviations were observed. In jet classes, the analysis sensitivity to potential physics beyond the Standard Model is limited by an imperfect simulation of the Standard Model behavior. Due to the potentially substantial impact of higher-order perturbative corrections, the analysis would greatly profit from a consistent set of NLO simulation samples for QCD jet production. To a lesser degree, local disagreement between different simulation programs was found to coincide with disagreement with data. The production of multijet events with significant p_T-imbalances is found to be insufficient, which may also be related to higher-order perturbative corrections. Following the MUSiC philosophy of only analysing well-modeled phase space regions, an exclusion of the concerned regions should be considered.

The identification of jets from bottom quark hadronisation ("b tagging") was successfully introduced to increase the analysis sensitivity to processes producing bottom quarks. Introduction of b tagging nearly triples the number of exclusive event classes and results in a robust separation of Standard Model processes with and without bottom quarks, as well as different sources of bottom quarks. The expected increase of sensitivity is impressively demonstrated by the significant excess found in the 1e1μ1b+X event class, which is not present in the 1e1μ1jet+X class.

The introduction of b tagging revealed available simulation samples to be insufficiently large to model high-momentum b jet final states. To address this, a dedicated simulation sample was produced privately using FastSim. The method was found to perform well in many final states, demonstrating its utility as an alternative, fast source of specialised simulation samples. However, small discrepancies highlighted the temporary nature of the solution. If additional simulation samples are needed in the future and time is not of the essence, it should be considered to request FullSim simulation samples for relevant phase space regions via the central CMS channels.

A sensitivity study was performed to demonstrate the performance of the analysis for a potential Dark Matter signal. Assuming a high-mass mediator between Standard Model matter and Dark Matter, signal simulation produced in an Effective Field Theory framework was used. The signal could be detected in final states with jets and E_T^{miss} up to suppression scales of $\Lambda \approx 0.6\,\text{TeV}$.

Bibliography

[1] T. Lancaster and S. J. Blundell, "Quantum Field Theory for the Gifted Amateur". Oxford Univ. Press, Oxford, 2014. Cited on pages 1 and 3.

[2] A. Zee, "Quantum Field Theory in a Nutshell". Princeton Univ. Press, Princeton, 2010. Cited on pages 1 and 3.

[3] M. E. Peskin and D. V. Schroeder, "An introduction to quantum field theory". Addison-Wesley, Reading, Mass., 1995. Cited on page 1.

[4] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys.* **C38** (2014) 090001. doi:10.1088/1674-1137/38/9/090001. Cited on pages 1, 7, 20, and 21.

[5] MissMJ, "Standard Model of Elementary Particles", 2014. Online, accessed on 02.11.2015. https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg. Cited on page 2.

[6] Wu-Ki Tung, "Bjorken scaling". Online, accessed on 02.11.2015. http://www.scholarpedia.org/article/Bjorken_scaling. Cited on page 4.

[7] CMS Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08004. http://stacks.iop.org/1748-0221/3/i=08/a=S08004. Cited on pages 5 and 36.

[8] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001. doi:10.1088/1748-0221/3/08/S08001. Cited on page 5.

[9] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", CMS-PAS-LUM-13-001, 2013. Cited on pages 5 and 33.

[10] CMS Collaboration, "Commissioning and performance of the L1 and HLT Trigger in 2012 and earlier.", CMS-TRG-12-001, 2015. Cited on pages 6, 14, and 15.

[11] "The CMS high level trigger", *The European Physical Journal C - Particles and Fields* **46** (2006), no. 3, 605–667. http://dx.doi.org/10.1140/epjc/s2006-02495-8. doi:10.1140/epjc/s2006-02495-8. Cited on page 6.

[12] CMS Collaboration, F. Beaudette, "The CMS Particle Flow Algorithm", in *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013)*, pp. 295–304. 2014. arXiv:1401.8155. Cited on page 7.

[13] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS-PAS-PFT-09-001, 2009. Cited on page 7.

[14] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **04** (2008) 063, arXiv:0802.1189. doi:10.1088/1126-6708/2008/04/063. Cited on pages 8, 9, and 77.

[15] CMS Collaboration, "Performance of Jet Algorithms in CMS", CMS-PAS-JME-07-003, CERN. Cited on page 8.

[16] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J.* **C72** (2012) 1896, arXiv:1111.6097. CERN-PH-TH-2011-297. doi:10.1140/epjc/s10052-012-1896-2. Cited on pages 9 and 77.

[17] L3 Collaboration, "The Construction of the L3 Experiment", *Nucl. Instrum. Meth.* **A289** (1990) 35–102. L3-000. doi:10.1016/0168-9002(90)90250-A. Cited on page 9.

[18] Thomas Hebbeker, "A Global Comparison between L3 Data and Standard Model Monte Carlo - a first attempt", 1998. http://web.physik.rwth-aachen.de/~hebbeker/l3note_2305.pdf. Cited on page 9.

[19] D0 Collaboration, "The D0 Detector", *Nucl. Instrum. Meth.* **A338** (1994) 185–253. doi:10.1016/0168-9002(94)91312-9. Cited on page 9.

[20] P. Biallass, "Model Independent Search for Deviations from the Standard Model at the Tevatron: Final States with Missing Energy", Diploma thesis, RWTH Aachen University, 2004. Cited on page 9.

[21] C. Hof, "Implementation of a Model-Independent Search for New Physics with the CMS Detector Exploiting the World-Wide LHC Computing Grid". PhD thesis, RWTH Aachen University, 2009. http://web.physik.rwth-aachen.de/~hebbeker/theses/hof_phd.pdf. Cited on page 9.

[22] CMS Collaboration, "MUSiC: A general search for deviations from Monte Carlo predictions in CMS", *J. Phys. Conf. Ser.* **171** (2009) 012102. CERN-CMS-CR-2009-015. doi:10.1088/1742-6596/171/1/012102. Cited on page 9.

[23] H. Pieta, "MUSiC - A Model Unspecific Search in CMS based on 2010 LHC data". PhD thesis, RWTH Aachen University, 2012. http://web.physik.rwth-aachen.de/~hebbeker/theses/pieta_phd.pdf. Cited on page 9.

[24] P. Papacz, "Model Unspecific Search for new Physics in CMS Based on 2011 Data". PhD thesis, RWTH Aachen University, 2014. http://web.physik.rwth-aachen.de/~hebbeker/theses/papacz_phd.pdf. Cited on pages 9 and 10.

[25] D. Duchardt, "MUSiC @ 8 TeV". PhD thesis, RWTH Aachen University, 2016. In preparation. . Cited on pages 9, 27, 49, and 63.

[26] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003. doi:10.1088/1748-0221/3/08/S08003. Cited on page 9.

[27] F. Cardillo and G. Herten, "A general search for new phenomena with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV". PhD thesis, Freiburg University, 2013. http://cds.cern.ch/record/1955478. Cited on page 9.

[28] E. Dietz-Laursonn, "Model Unspecific Search for New Physics with b-Hadrons in CMS", Diploma thesis, RWTH Aachen University, 2009. Cited on page 9.

[29] T. Esch, "Anwendung von B-Erkennung auf die modellunspezische Suche in CMS", Bachelor thesis, RWTH Aachen University, 2011. Cited on page 9.

[30] CMS Collaboration, "Measurement of dijet azimuthal decorrelations in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", CMS-PAS-SMP-14-015, 2015. Cited on pages 11 and 77.

[31] CMS Collaboration, "Physics Data/MC Validation". Online, accessed on 29.07.2015. https://twiki.cern.ch/twiki/bin/view/CMS/PdmV2012Analysis?rev=95#22_JanReReco. Cited on page 14.

[32] CMS Collaboration, "E/gamma HLT Algorithm work for 2012". Online, accessed on 09.11.2015. https://twiki.cern.ch/twiki/bin/view/CMS/ChangesEGMHLTAlgo2012?rev=14. Cited on page 16.

[33] CMS Collaboration, "JetMET Physics Object Group (JME POG)". Online. https://twiki.cern.ch/twiki/bin/view/CMS/JetMET. Cited on pages 16 and 33.

[34] CMS Collaboration, "MET Optional Filters". Online, accessed on 28.07.2015. https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFilters?rev=66. Cited on page 16.

[35] CMS Collaboration, "Baseline muon selections for Run-I". Online. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId?rev=60#Tight_Muon. Cited on page 16.

[36] CMS Collaboration, "Baseline muon selections for Run-I". Online, accessed on 30.07.2015. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId?rev=60#New_HighPT_Version_recommended. Cited on page 16.

[37] CMS Collaboration, "HEEP Electron ID and isolation". Online, accessed on 28.07.2015. https://twiki.cern.ch/twiki/bin/view/CMS/HEEPElectronID?rev=68. Cited on page 17.

[38] CMS Collaboration, "Exotica High PT Electrons". Online, accessed on 28.07.2015. https://twiki.cern.ch/twiki/bin/view/CMS/HighPtElectrons?rev=54. Cited on page 17.

[39] CMS Collaboration, "Simple Cut Based Photon ID 2012". Online, accessed on 30.07.2015. https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedPhotonID2012?rev=23. Cited on pages 17, 18, and 33.

[40] CMS Collaboration, "Jet ID". Online, accessed on 30.07.2015. https://twiki.cern.ch/twiki/bin/view/CMS/JetID?rev=90#Recommendations%_for_8_TeV_data_a. Cited on pages 18 and 37.

[41] CMS Collaboration, "Pileup Removal Algorithms", CMS-PAS-JME-14-001, CERN, 2014. Cited on page 18.

[42] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, arXiv:1107.4277. CERN-PH-EP-2011-102, CMS-JME-10-011. doi:10.1088/1748-0221/6/11/P11002. Cited on page 18.

[43] CMS Collaboration, "Jet Energy Resolution and Corrections (JERC) Subgroup". Online, accessed on 31.07.2015. https://twiki.cern.ch/twiki/bin/view/CMS/JetEnergyScale?rev=110. Cited on page 18.

[44] CMS Collaboration, "Recommended Jet Energy Corrections and Uncertainties For Data and MC". Online, accessed on 31.07.2015. https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC?rev=80. Cited on pages 19 and 33.

[45] CMS Collaboration, "Status of the 8 TeV Jet Energy Corrections and Uncertainties based on 11 fb^{-1} of data in CMS", CMS-DP-2013-011, May, 2013. Cited on pages 19 and 33.

[46] C. Weiser, "A Combined Secondary Vertex Based B-Tagging Algorithm in CMS", CMS-NOTE-2006-014, CERN, Jan, 2006. Cited on pages 19 and 21.

[47] CMS Collaboration, "Usage of b Tag Objects for 8 TeV Data with the 53X Re-reco ". Online, accessed on 31.07.2015. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation53XReReco?rev=9. Cited on page 19.

[48] CMS Collaboration, "MET Corrections". Online, accessed on 30.07.2015. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis?rev=198#7_7_6_MET_Corrections. Cited on page 20.

[49] CMS Collaboration, "Algorithms for b Jet identification in CMS", CMS-PAS-BTV-09-001, 2009. Cited on page 21.

[50] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, arXiv:1211.4462. CMS-BTV-12-001, CERN-PH-EP-2012-262. doi:10.1088/1748-0221/8/04/P04013. Cited on page 21.

[51] CMS Collaboration, "Performance of b tagging at $\sqrt{s} = 8\,\text{TeV}$ in multijet, ttbar and boosted topology events", CMS-PAS-BTV-13-001, 2013. Cited on pages 21 and 33.

[52] CMS Collaboration, "Mistag Scale Factors for b-tagging in 2012 data", CMS-AN-2013/150, 2013. Cited on page 22.

[53] CMS Collaboration, "b Tag and Vertexing Physics Object Group". Online. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagPOG. Cited on page 21.

[54] CMS Collaboration, "Methods to apply b-tagging efficiency scale factors". Online, accessed on 13.11.2015. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods?rev=25. Cited on page 22.

[55] CMS Collaboration, "Jet Energy Calibration in the 8 TeV pp data", CMS PAPER JME-13-004, 2013. Cited on pages 22, 35, 38, and 43.

[56] CMS Collaboration, "Summaries of CMS cross section measurements". Online, accessed on 09.11.2015. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined?rev=36. Cited on page 25.

[57] CMS Collaboration, "Measurement of the ratio of inclusive jet cross sections using the anti-k_T algorithm with radius parameters R=0.5 and 0.7 in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev.* **D90** (2014), no. 7, 072006, arXiv:1406.0324. CMS-SMP-13-002, CERN-PH-EP-2014-068. doi:10.1103/PhysRevD.90.072006. Cited on page 25.

[58] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175. doi:10.1088/1126-6708/2006/05/026. Cited on page 25.

[59] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph 5 : Going Beyond", *JHEP* **06** (2011) 128, arXiv:1106.0522. doi:10.1007/JHEP06(2011)128. Cited on page 25.

[60] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, arXiv:hep-ph/0409146. doi:10.1088/1126-6708/2004/11/040. Cited on page 25.

[61] 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. doi:10.1088/1126-6708/2007/11/070. Cited on page 25.

[62] S. Alioli, P. Nason, C. Oleari et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, arXiv:1002.2581. doi:10.1007/JHEP06(2010)043. Cited on page 25.

[63] J. Pumplin, D. R. Stump, J. Huston et al., "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis", *J. High Energy Phys.* **07** (Jan, 2002) 012. 44 p. hep-ph/0201195. http://cds.cern.ch/record/534624. Cited on pages 26 and 80.

[64] H.-L. Lai, M. Guzzi, J. Huston et al., "New parton distributions for collider physics", *Phys. Rev.* **D82** (2010) 074024, arXiv:1007.2241. MSUHEP-100707, SMU-HEP-10-10. doi:10.1103/PhysRevD.82.074024. Cited on pages 26 and 80.

[65] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, arXiv:1101.2599. doi:10.1016/j.physrep.2011.03.005. Cited on page 26.

[66] B. Andersson, G. Gustafson, G. Ingelman et al., "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145. LU-TP-83-10. doi:10.1016/0370-1573(83)90080-7. Cited on page 26.

[67] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8. Cited on page 26.

[68] CMS Collaboration, "Standard Model Cross Sections for CMS at 8 TeV", 2013. Online, accessed on 10.11.2015. https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV?rev=30. Cited on page 27.

[69] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to dilepton production in FEWZ", *Phys. Rev.* **D86** (2012) 094034, arXiv:1208.5967. ANL-HEP-PR-12-68. doi:10.1103/PhysRevD.86.094034. Cited on page 27.

[70] CMS Collaboration, "Search for a Higgs boson decaying into two photons in proton-proton collisions recorded by the CMS detector at the LHC.", 2012. Cited on page 27.

[71] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$", *Phys. Rev. Lett.* **110** (2013) 252004, arXiv:1303.6254. doi:10.1103/PhysRevLett.110.252004. Cited on page 27.

[72] CMS Collaboration, "Search for High Mass Resonances Decaying to Electron Pairs at 8 TeV with the Full 2012 dataset", AN-12-415, 2014. Cited on page 27.

[73] N. Kidonakis, "Differential and total cross sections for top pair and single top production", in *Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012)*, pp. 831–834. 2012. arXiv:1205.3453. Cited on page 27.

[74] M. V. Garzelli, A. Kardos, C. G. Papadopoulos et al., "$t\bar{t}\,W^{+-}$ and $t\bar{t}$ Z Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects", *JHEP* **11** (2012) 056, arXiv:1208.2665. doi:10.1007/JHEP11(2012)056. Cited on page 27.

[75] J. M. Campbell and R. K. Ellis, "$t\bar{t}W^{+-}$ production and decay at NLO", *JHEP* **07** (2012) 052, arXiv:1204.5678. doi:10.1007/JHEP07(2012)052. Cited on page 27.

[76] A. Giammanco, "The Fast Simulation of the CMS Experiment", *J. Phys. Conf. Ser.* **513** (2014) 022012. doi:10.1088/1742-6596/513/2/022012. Cited on page 29.

[77] CMS Collaboration, "Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV", *JHEP* **09** (2011) 109, arXiv:1107.0330. CERN-PH-EP-2011-059, CMS-QCD-10-010. doi:10.1007/JHEP09(2011)109. Cited on page 30.

[78] CMS Collaboration, "Pileup Information ". Online, accessed on 10.08.2015. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVPileUpDescription?rev=22. Cited on page 31.

[79] CMS Collaboration, "Pileup Studies". Online, accessed on 10.08.2015. https://twiki.cern.ch/twiki/bin/view/CMS/PileupInformation?rev=32. Cited on page 31.

[80] CMS Collaboration, "Search for physics beyond the standard model in dilepton mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JHEP* **04** (2015) 025, arXiv:1412.6302. CMS-EXO-12-061, CERN-PH-EP-2014-272. doi:10.1007/JHEP04(2015)025. Cited on page 33.

[81] CMS Collaboration, "Measurement of the high energy electron selection efficiency using Drell-Yan events collected in 2012.", CMS AN-2013/174, 2013. Cited on page 33.

[82] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations",, arXiv:1101.0538. Cited on page 33.

[83] Alessandro Vicinii, "Practical implementation of the PDF4LHC recipe". Online, accessed on 10.08.2015. http://www.hep.ucl.ac.uk/pdf4lhc/PDF4LHC_practical_guide.pdf. Cited on page 33.

[84] A. Buckley, J. Ferrando, S. Lloyd et al., "LHAPDF6: parton density access in the LHC precision era", *Eur. Phys. J.* **C75** (2015), no. 3, 132, arXiv:1412.7420. doi:10.1140/epjc/s10052-015-3318-8. Cited on page 34.

[85] CMS Collaboration, "Jet Energy Resolution". Online, accessed on 12.08.2015. https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution?rev=37. Cited on page 35.

[86] CMS Collaboration, "Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV with the CMS Detector", CMS-PAS-JME-12-002, 2012. Cited on pages 36, 37, and 41.

[87] CMS Collaboration, "Missing Transverse Energy (MET) ". Online. https://twiki.cern.ch/twiki/bin/view/CMS/MissingET. Cited on page 37.

[88] CMS Collaboration, "Search for new physics in monojet events in pp collisions at $\sqrt{s} = 8$ TeV", CMS-PAS-EXO-12-048, 2013. Cited on page 38.

[89] CMS Collaboration, "Search for dark matter, extra dimensions, and unparticles in monojet events in proton–proton collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J.* **C75** (2015), no. 5, 235, arXiv:1408.3583. CMS-EXO-12-048, CERN-PH-EP-2014-164. doi:10.1140/epjc/s10052-015-3451-4. Cited on pages 38, 89, 90, and 93.

[90] CMS Collaboration, "Particle-based noise rejection for fully hadronic SUSY analyses", CMS-AN-11-313, 2011. Cited on pages 38 and 41.

[91] CMS Collaboration, "Study of MET Performance in the full 2012 Dataset using Z→ μμ events", CMS AN-2013/233, 2013. Cited on pages 40 and 41.

[92] CMS Collaboration, "Study of Modulation of the ϕ distributions of Missing Transverse Energy at the CMS detector", CMS AN-2014/262, 2014. Cited on page 40.

[93] CMS Collaboration, "W+jets differential cross section measurements at 8 TeV", CMS AN-14-114, 2014. Cited on page 46.

[94] CMS Collaboration, "Differential cross section of jets associated to Z boson in Proton-Proton Collisions at $\sqrt{s} = 8\,\text{TeV}$", CMS AN-2013/049 , 2015. Cited on page 46.

[95] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **02** (2014) 024, arXiv:1312.7582, [Erratum: JHEP02,102(2014)]. CMS-TOP-12-007, CERN-PH-EP-2013-234. doi:10.1007/JHEP02(2014)024, 10.1007/JHEP02(2014)102. Cited on page 46.

[96] J. Alwall, R. Frederix, S. Frixione et al., "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. CERN-PH-TH-2014-064. doi:10.1007/JHEP07(2014)079. Cited on page 46.

[97] CMS Collaboration, "Search for physics beyond the standard model in final states with a lepton and missing transverse energy in proton-proton collisions at sqrt(s) = 8 TeV", *Phys. Rev.* **D91** (2015), no. 9, 092005, arXiv:1408.2745. CMS-EXO-12-060, CERN-PH-EP-2014-176. doi:10.1103/PhysRevD.91.092005. Cited on page 50.

[98] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at 8 TeV", CMS-PAS-TOP-12-027, 2013. Cited on page 67.

[99] CMS Collaboration, "Measurement of the Differential Cross Section for Top Quark Pair Production in pp Collisions at $\sqrt{s} = 8$ TeV",, arXiv:1505.04480. CMS-TOP-12-028, CERN-PH-EP-2015-117. Cited on page 67.

[100] CMS Collaboration, "p_T (top-quark) based reweighting of $t\bar{t}$ MC". Online, accessed on 29.09.2015. https://twiki.cern.ch/twiki/bin/view/CMS/TopPtReweighting?rev=18. Cited on page 67.

[101] CMS Collaboration, "Measurements of differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ with the CMS detector", CMS-PAS-SMP-14-002, 2014. Cited on pages 72 and 82.

[102] CMS Collaboration, "Dijet Azimuthal Decorrelations in pp Collisions at $\sqrt{s} = 8\,\text{TeV}$", CMS AN-2014/119, 2015. Cited on page 72.

[103] T. Gleisberg, S. Hoeche, F. Krauss et al., "Event generation with SHERPA 1.1", *JHEP* **02** (2009) 007, arXiv:0811.4622. doi:10.1088/1126-6708/2009/02/007. Cited on page 77.

[104] S. Hoeche and M. Schonherr, "Uncertainties in next-to-leading order plus parton shower matched simulations of inclusive jet and dijet production", *Phys. Rev.* **D86** (2012) 094042, arXiv:1208.2815. doi:10.1103/PhysRevD.86.094042. Cited on pages 77 and 80.

[105] Sherpa Collaboration, "SHERPA 2.1.1 Documentation",, Online, accessed on 08.09.2015. http://sherpa.hepforge.org/doc/SHERPA-MC-2.1.1.html. Cited on pages 77 and 120.

[106] F. Cascioli, P. Maierhofer, and S. Pozzorini, "Scattering Amplitudes with Open Loops", *Phys. Rev. Lett.* **108** (2012) 111601, arXiv:1111.5206. doi:10.1103/PhysRevLett.108.111601. Cited on page 77.

[107] S. Hoeche, F. Krauss, M. Schonherr et al., "A critical appraisal of NLO+PS matching methods", *JHEP* **09** (2012) 049, arXiv:1111.1220. doi:10.1007/JHEP09(2012)049. Cited on page 77.

[108] Z. Nagy, "Next-to-leading order calculation of three jet observables in hadron hadron collision", *Phys. Rev.* **D68** (2003) 094002, arXiv:hep-ph/0307268. doi:10.1103/PhysRevD.68.094002. Cited on pages 80, 82, and 83.

[109] W. B. Kilgore and W. T. Giele, "Hadronic three jet production at next-to-leading order", in *American Physical Society (APS) Meeting of the Division of Particles and Fields (DPF 99) Los Angeles, California, January 5-9, 1999*. 1999. arXiv:hep-ph/9903361. Cited on pages 80 and 82.

[110] CMS Collaboration, "Measurement of the double-differential inclusive jet cross section at sqrt(s) = 8 TeV with the CMS detector", CMS-PAS-SMP-12-012, 2013. Cited on pages 80 and 82.

[111] W. T. Giele, E. W. N. Glover, and D. A. Kosower, "Higher order corrections to jet cross-sections in hadron colliders", *Nucl. Phys.* **B403** (1993) 633–670, arXiv:hep-ph/9302225. doi:10.1016/0550-3213(93)90365-V. Cited on page 82.

[112] S. D. Ellis, Z. Kunszt, and D. E. Soper, "Two jet production in hadron collisions at order α_S^3 in QCD", *Phys. Rev. Lett.* **69** (1992) 1496–1499. doi:10.1103/PhysRevLett.69.1496. Cited on page 82.

[113] G. C. Blazey et al., "Run II jet physics", in *QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*, pp. 47–77. 2000. arXiv:hep-ex/0005012. Cited on page 82.

[114] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev.* **D48** (1993) 3160–3166, arXiv:hep-ph/9305266. CERN-TH-6860-93. doi:10.1103/PhysRevD.48.3160. Cited on page 82.

[115] CMS Collaboration, "Search for narrow resonances using the dijet mass spectrum in pp collisions at $\sqrt{s}$=8 TeV", *Phys. Rev.* **D87** (2013), no. 11, 114015, arXiv:1302.4794. CMS-EXO-12-016, CERN-PH-EP-2013-015. doi:10.1103/PhysRevD.87.114015. Cited on page 82.

[116] K. Garrett and G. Duda, "Dark Matter: A Primer", *Adv. Astron.* **2011** (2011) 968283, arXiv:1006.2483. doi:10.1155/2011/968283. Cited on page 89.

[117] CMS Collaboration, "Search for dark matter in the mono-lepton channel with pp collision events at center-of-mass energy of 8 TeV ", CMS-PAS-EXO-13-004, CERN, 2013. Cited on page 89.

[118] CMS Collaboration, "Search for dark matter and unparticles produced in association with a Z boson in pp collisions at $\sqrt{s} = 8$ TeV", CMS-PAS-EXO-12-054, 2015. Cited on page 89.

[119] CMS Collaboration, "Search for new phenomena in monophoton final states in proton-proton collisions at $\sqrt{s} = 8$ TeV",, arXiv:1410.8812. CMS-EXO-12-047, CERN-PH-EP-2014-253. Cited on page 89.

[120] P. J. Fox, R. Harnik, J. Kopp et al., "Missing Energy Signatures of Dark Matter at the LHC", *Phys. Rev.* **D85** (2012) 056011, arXiv:1109.4398. doi:10.1103/PhysRevD.85.056011. Cited on page 89.

Danksagung

Wie jeder weiß, ist wissenschaftliche Arbeit vor allem Teamarbeit. Diese Masterarbeit wäre ohne die Hilfe vieler Menschen, die nicht auf der Titelseite stehen, nicht möglich gewesen. In chronologischer Reihenfolge ihres Auftretens möchte ich einigen von ihnen besonders danken.

Zuerst natürlich meinen Eltern, die mich stets ermutigt haben, mein Leben mit Dingen zu verbringen, die mir Freude machen. Auch deshalb musste ich nie daran zweifeln, dass die Physik ein gutes Geschäft für mich ist. Ihnen habe ich auch zu verdanken, dass ich fünf Jahre lang so sorgenfrei studieren konnte.

Lucas Weimer danke ich dafür, dass er mich nie hat vergessen lassen, dass es so viel Spannendes gibt, von dem ich nichts weiß. Ohne ihn hätte ich nie erfahren, dass man als Physiker Feuerwehrchef werden kann, warum im Bergwerk Senf so wichtig ist und dass manchen Leuten ein Marathon einfach zu kurz ist.

Ich danke Thomas Hebbeker, der es mir ermöglicht hat, in seiner Arbeitsgruppe einen Platz zu finden. Ebenso danke ich Arnd Meyer, der alles interessant findet und zu allem Interessantes erzählen kann. Ihrer beider Einfluss hat dieser Arbeit gut getan. Christopher Wiebusch danke ich dafür, dass er sich bereit erklärt hat, das Zweitgutachten für diese Arbeit zu übernehmen.

Natürlich danke ich der Arbeitsgruppe! Herzlich wurde ich auf- und bereits nach Kurzem fachlich ernst genommen. Besonders danken muss ich Debbie Duchardt und Simon Knutzen, ohne die MUSiC für mich ewig ein komplexes Monster geblieben wäre. Analoges gilt für Sebastian Thüer und Sherpa; auch da wäre ich ohne Beratung verloren gewesen. An allen Kollegen schätze ich, dass sie immer wieder bereit sind, ihre Zeit mit den Problemen anderer Leute zu verbringen. Abgesehen von Eurer Hilfsbereitschaft ist es mit Euch aber auch einfach immer witzig.

Schön war's.

Erklärung

Ich versichere, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate kenntlich gemacht habe.

Aachen, den 25.11.2015

Appendices

A Comprehensive List of Simulation Samples

Process	Dataset Name	X-Section /pb	Filter Efficiency	k-Factor (Order)	Number Events	Event Weight	Conditions
$\gamma\gamma$	DiPhotonBorn_Pt-10To25_8TeV-pythia6	270	1.0	1.0 (LO)	$5.0 \cdot 10^5$	11	START53_V7A-v1
	DiPhotonBorn_Pt-10To25_8TeV_ext-pythia6	270	1.0	1.0 (LO)	$1.0 \cdot 10^6$	5.3	START53_V7A-v1
	DiPhotonBorn_Pt-25To250_8TeV-pythia6	25	1.0	1.0 (LO)	$5.0 \cdot 10^5$	1.0	START53_V7A-v1
	DiPhotonBorn_Pt-25To250_8TeV_ext-pythia6	25	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.50	START53_V7A-v1
	DiPhotonBorn_Pt-250ToInf_8TeV-pythia6	0.011	1.0	1.0 (LO)	$5.0 \cdot 10^5$	$4.3 \cdot 10^{-4}$	START53_V7A-v1
	DiPhotonBorn_Pt-250ToInf_8TeV_ext-pythia6	0.011	1.0	1.0 (LO)	$1.0 \cdot 10^6$	$2.1 \cdot 10^{-4}$	START53_V7A-v1
	DiPhotonBox_Pt-10To25_8TeV-pythia6	430	1.0	1.0 (LO)	$5.0 \cdot 10^5$	17	START53_V7A-v1
	DiPhotonBox_Pt-25To250_8TeV-pythia6	16	1.0	1.0 (LO)	$5.0 \cdot 10^5$	0.61	START53_V7A-v1
	DiPhotonBox_Pt-25To250_8TeV_ext-pythia6	16	1.0	1.0 (LO)	$8.0 \cdot 10^5$	0.37	START53_V7A-v1
	DiPhotonBox_Pt-250ToInf_8TeV-pythia6	$3.2 \cdot 10^{-4}$	1.0	1.0 (LO)	$5.0 \cdot 10^5$	$1.3 \cdot 10^{-5}$	START53_V7A-v1
	DiPhotonBox_Pt-250ToInf_8TeV_ext-pythia6	$3.2 \cdot 10^{-4}$	1.0	1.0 (LO)	$9.0 \cdot 10^5$	$6.5 \cdot 10^{-6}$	START53_V7A-v1
γ+Jets	G_Pt-0to15_TuneZ2star_8TeV_pythia6	$9.1 \cdot 10^7$	1.0	1.3 (NLO)	$2.0 \cdot 10^6$	$1.2 \cdot 10^6$	START53_V7A-v1
	G_Pt-15to30_TuneZ2star_8TeV_pythia6	$2.0 \cdot 10^5$	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	2600	START53_V7A-v1
	G_Pt-30to50_TuneZ2star_8TeV_pythia6	$2.0 \cdot 10^4$	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	260	START53_V7A-v1
	G_Pt-50to80_TuneZ2star_8TeV_pythia6	3300	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	43	START53_V7A-v1
	G_Pt-80to120_TuneZ2star_8TeV_pythia6	560	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	7.2	START53_V7A-v1
	G_Pt-120to170_TuneZ2star_8TeV_pythia6	110	1.0	1.3 (NLO)	$2.0 \cdot 10^6$	1.4	START53_V7A-v1
	G_Pt-170to300_TuneZ2star_8TeV_pythia6	30	1.0	1.3 (NLO)	$2.0 \cdot 10^6$	0.39	START53_V7A-v1
	G_Pt-300to470_TuneZ2star_8TeV_pythia6	2.1	1.0	1.3 (NLO)	$2.0 \cdot 10^6$	0.027	START53_V7A-v1
	G_Pt-470to800_TuneZ2star_8TeV_pythia6	0.21	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	0.0027	START53_V7A-v1
	G_Pt-800to1400_TuneZ2star_8TeV_pythia6	0.0071	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	$9.2 \cdot 10^{-5}$	START53_V7A-v1
	G_Pt-1400to1800_TuneZ2star_8TeV_pythia6	$4.5 \cdot 10^{-5}$	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	$5.8 \cdot 10^{-7}$	START53_V7A-v1
	G_Pt-1800_TuneZ2star_8TeV_pythia6	$1.9 \cdot 10^{-6}$	1.0	1.3 (NLO)	$1.0 \cdot 10^6$	$2.5 \cdot 10^{-8}$	START53_V7A-v1
W+γ	WGToLNuG_TuneZ2star_8TeV-madgraph-tauola	460	1.0	1.0 (LO)	$4.0 \cdot 10^6$	1.9	START53_V7A-v1
W+$\gamma\gamma$	LNuGG_enhanced_FSR_8TeV_madgraph	1.8	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.036	START53_V19-v1
	LNuGG_enhanced_ISR_8TeV_madgraph	0.32	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.0063	START53_V19-v1
W+Jets	WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball	$3.0 \cdot 10^4$	1.0	1.21 (NNLO)	$1.0 \cdot 10^7$	39	START53_V7A-v1
	WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball	$3.0 \cdot 10^4$	1.0	1.21 (NNLO)	$5.0 \cdot 10^7$	13	START53_V7A-v2
	WJetsToLNu_M-200_TuneZ2star_8TeV-madgraph	35	0.30	1.21 (NNLO)	$3.0 \cdot 10^5$	0.70	FastSim
	WJetsToLNu_M-500_TuneZ2star_8TeV-madgraph	0.44	0.61	1.21 (NNLO)	$4.0 \cdot 10^5$	0.014	FastSim
	WJetsToLNu_M-1000_TuneZ2star_8TeV-madgraph	1.0	0.011	1.21 (NNLO)	$5.0 \cdot 10^5$	$5.0 \cdot 10^{-4}$	FastSim
	WJetsToLNu_HT-150To200_8TeV-madgraph	235.6	1.0	1.21 (NNLO)	$21.7 \cdot 10^6$	0.21	START53_V7C-v1
	WJetsToLNu_HT-200To250_8TeV-madgraph	90.27	1.0	1.21 (NNLO)	$10.0 \cdot 10^6$	0.18	START53_V7C-v1
	WJetsToLNu_HT-250To300_8TeV-madgraph_v2	48.01	1.0	1.21 (NNLO)	$4.9 \cdot 10^6$	0.11	START53_V7A-v1
	WJetsToLNu_HT-300To400_8TeV-madgraph_v2	38.3	1.0	1.21 (NNLO)	$5.1 \cdot 10^6$	0.15	START53_V7A-v1
	WJetsToLNu_HT-400ToInf_8TeV-madgraph_v2	25.22	1.0	1.21 (NNLO)	$5.0 \cdot 10^6$	0.10	START53_V7A-v1
WW+γ	WWGJets_8TeV-madgraph	0.53	1.0	1.0 (LO)	$2.0 \cdot 10^5$	0.048	START53_V7A-v1
	WWGJets_8TeV-madgraph_v2	0.53	1.0	1.0 (LO)	$3.0 \cdot 10^5$	0.034	START53_V7A-v1
Z+γ	ZGToLLG_8TeV-madgraph	130	1.0	1.0 (LO)	$6.0 \cdot 10^6$	0.40	START53_V7A-v1

Process	Dataset Name	X-Section /pb	Filter Efficiency	k-Factor (Order)	Number Events	Event Weight	Conditions
$t\bar{t}$	TT_CT10_TuneZ2star_8TeV-powheg-tauola	250	1.0	1.0 (NNLO)	$6.0 \cdot 10^6$	0.75	START53_V7A-v1
	TT_CT10_TuneZ2star_8TeV-powheg-tauola	250	1.0	1.0 (NNLO)	$2.0 \cdot 10^7$	0.22	START53_V7A-v2
	TT_Mtt-700to1000_CT10_TuneZ2star_8TeV-powheg-tauola	210	0.074	1.109 (NNLO)	$3.0 \cdot 10^6$	0.11	START53_V7A-v1
	TT_Mtt-1000toInf_CT10_TuneZ2star_8TeV-powheg-tauola	3.0	0.014	1.109 (NNLO)	$1.0 \cdot 10^6$	$7.2 \cdot 10^{-4}$	START53_V7A-v1
$t\bar{t}+\gamma$	TTGJets_8TeV-madgraph	1.4	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.017	START53_V19-v1
$t\bar{t}$+V	TTWJets_8TeV-madgraph	0.21	1.0	1.08 (NLO)	$1.0 \cdot 10^5$	0.023	START53_V7A-v1
	TTWWJets_8TeV-madgraph	0.0020	1.0	1.0 (LO)	$2.0 \cdot 10^5$	$1.8 \cdot 10^{-4}$	START53_V7A-v1
	TTZJets_8TeV-madgraph_v2	0.17	1.0	1.2 (NLO)	$2.0 \cdot 10^5$	0.019	START53_V7A-v1
$t\bar{t}t\bar{t}$	TTTT_TuneZ2star_8TeV-madgraph-tauola	$7.2 \cdot 10^{-4}$	1.0	1.28 (NLO)	$9.0 \cdot 10^4$	$1.8 \cdot 10^{-4}$	START53_V7A-v1
Di-Boson	WW_TuneZ2star_8TeV_pythia6_tauola	34	1.0	1.63 (NLO)	$1.0 \cdot 10^7$	0.11	START53_V7A-v1
	ZZTo2e2mu_8TeV-powheg-pythia6	0.18	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	0.0023	START53_V7A-v1
	ZZTo2e2mu_8TeV_ext-powheg-pythia6	0.18	1.0	1.0 (NLO)	$3.0 \cdot 10^6$	$8.7 \cdot 10^{-4}$	START53_V7C-v1
	ZZTo2mu2tau_8TeV-powheg-pythia6	0.18	1.0	1.0 (NLO)	$8.0 \cdot 10^5$	0.0042	START53_V7A-v1
	ZZTo2mu2tau_8TeV_ext-powheg-pythia6	0.18	1.0	1.0 (NLO)	$3.0 \cdot 10^6$	$8.7 \cdot 10^{-4}$	START53_V7C-v1
	ZZTo2e2tau_8TeV-powheg-pythia6	0.18	1.0	1.0 (NLO)	$8.0 \cdot 10^5$	0.0042	START53_V7A-v1
	ZZTo2e2tau_8TeV_ext-powheg-pythia6	0.18	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	0.0017	START53_V7C-v1
	ZZTo4e_8TeV-powheg-pythia6	0.077	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	0.0010	START53_V7A-v1
	ZZTo4e_8TeV_ext-powheg-pythia6	0.077	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	$7.6 \cdot 10^{-4}$	START53_V7C-v1
	ZZTo4mu_8TeV-powheg-pythia6	0.077	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	0.0010	START53_V7A-v1
	ZZTo4mu_8TeV_ext-powheg-pythia6	0.077	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	$7.6 \cdot 10^{-4}$	START53_V7C-v1
	ZZTo4tau_8TeV-powheg-pythia6	0.077	1.0	1.0 (NLO)	$8.0 \cdot 10^5$	0.0018	START53_V7A-v1
	ZZTo4tau_8TeV_ext-powheg-pythia6	0.077	1.0	1.0 (NLO)	$1.0 \cdot 10^6$	$7.6 \cdot 10^{-4}$	START53_V7C-v1
	ZZJetsTo2L2Nu_TuneZ2star_8TeV-madgraph-tauola	0.28	1.0	1.0 (LO)	$9.0 \cdot 10^5$	0.0058	START53_V7A-v3
	ZZJetsTo2L2Q_TuneZ2star_8TeV-madgraph-tauola	0.91	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.0093	START53_V7A-v1
	ZZJetsTo2Q2Nu_TuneZ2star_8TeV-madgraph-tauloa	0.49	1.0	1.0 (LO)	$8.0 \cdot 10^5$	0.012	START53_V7A-v1
	GluGluToWWTo4L_TuneZ2star_8TeV-gg2ww-pythia6	0.023	1.0	1.0 (LO)	$1.0 \cdot 10^5$	0.0041	START53_V7A-v1
	GluGluToZZTo2L2L_TuneZ2star_8TeV-gg2zz-pythia6	0.012	1.0	1.0 (NLO)	$4.0 \cdot 10^5$	$5.9 \cdot 10^{-4}$	START53_V7A-v1
	GluGluToZZTo4L_8TeV-gg2zz-pythia6	0.0048	1.0	1.0 (NLO)	$5.0 \cdot 10^5$	$1.7 \cdot 10^{-4}$	START53_V7A-v1
	WZJetsTo2L2Q_TuneZ2star_8TeV-madgraph-tauola	1.8	1.0	1.0 (LO)	$3.0 \cdot 10^6$	0.011	START53_V7A-v1
	WZJetsTo2Q2Nu_TuneZ2star_8TeV-madgraph-tauloa	0.70	1.0	1.0 (LO)	$9.0 \cdot 10^5$	0.015	START53_V7A-v1
	WZJetsTo2QLNu_8TeV-madgraph	3.1	1.0	1.0 (LO)	$2.0 \cdot 10^6$	0.021	START53_V7C-v1
	WZJetsTo3LNu_TuneZ2_8TeV-madgraph-tauola	0.87	1.0	1.0 (LO)	$2.0 \cdot 10^6$	0.0085	START53_V7A-v1

Process	Dataset Name	X-Section /pb	Filter Efficiency	k-Factor (Order)	Number Events	Event Weight	Conditions
Drell-Yan	DYJetsToLL_M-10To50_TuneZ2Star_8TeV-madgraph	$1.1 \cdot 10^4$	1.0	1.2 (NNLO)	$3.0 \cdot 10^7$	6.9	START53_V7A-v1
	DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball	3000	1.0	1.2 (NNLO)	$3.0 \cdot 10^7$	2.3	START53_V7A-v1
	DYJetsToLL_PtZ-50To70_TuneZ2star_8TeV-madgraph-tarball	94	1.0	1.0 (LO)	$4.0 \cdot 10^6$	0.37	START53_V7A-v1
	DYJetsToLL_PtZ-50To70_TuneZ2star_8TeV_ext-madgraph-tarball	89	1.0	1.0 (LO)	$1.0 \cdot 10^7$	0.11	START53_V7C-v1
	DYJetsToLL_PtZ-70To100_TuneZ2star_8TeV-madgraph-tarball	52	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.73	START53_V7A-v2
	DYJetsToLL_PtZ-70To100_TuneZ2star_8TeV_ext-madgraph-tarball	53	1.0	1.0 (LO)	$1.0 \cdot 10^7$	0.089	START53_V7C-v1
	DYJetsToLL_PtZ-100_TuneZ2star_8TeV-madgraph	34	1.0	1.0 (LO)	$2.0 \cdot 10^6$	0.25	START53_V7A-v2
	DYJetsToLL_PtZ-100_TuneZ2star_8TeV_ext-madgraph-tarball	33	1.0	1.0 (LO)	$1.0 \cdot 10^7$	0.052	START53_V7C-v1
	DYJetsToLL_PtZ-180_TuneZ2star_8TeV-madgraph-tarball	4.6	1.0	1.0 (LO)	$1.0 \cdot 10^6$	0.058	START53_V7C-v1
	DYToEE_M-500_CT10_TuneZ2star_8TeV-powheg-pythia6	0.044	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0089	START53_V7A-v1
	DYToEE_M-700_CT10_TuneZ2star_8TeV-powheg-pythia6	0.010	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0021	START53_V7A-v1
	DYToEE_M-800_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0055	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0011	START53_V7A-v1
	DYToEE_M-1000_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0018	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$3.6 \cdot 10^{-4}$	START53_V7A-v1
	DYToEE_M-1500_CT10_TuneZ2star_8TeV-powheg-pythia6	$1.7 \cdot 10^{-4}$	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$3.4 \cdot 10^{-5}$	START53_V7A-v1
	DYToEE_M-2000_CT10_TuneZ2star_8TeV-powheg-pythia6	$2.2 \cdot 10^{-5}$	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$4.5 \cdot 10^{-6}$	START53_V7A-v1
	DYToMuMu_M-500_CT10_TuneZ2star_8TeV-powheg-pythia6	0.044	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0089	START53_V7A-v1
	DYToMuMu_M-500_CT10_TuneZ2star_8TeV-powheg-pythia6	0.044	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0089	START53_V7C1-v1
	DYToMuMu_M-500_CT10_TuneZ2star_8TeV-powheg-pythia6	0.044	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0089	START53_V7C2-v1
	DYToMuMu_M-700_CT10_TuneZ2star_8TeV-powheg-pythia6	0.010	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0021	START53_V7A-v1
	DYToMuMu_M-700_CT10_TuneZ2star_8TeV-powheg-pythia6	0.010	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0021	START53_V7C1-v1
	DYToMuMu_M-700_CT10_TuneZ2star_8TeV-powheg-pythia6	0.010	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0021	START53_V7C2-v1
	DYToMuMu_M-800_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0055	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0011	START53_V7A-v1
	DYToMuMu_M-800_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0055	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0011	START53_V7C1-v1
	DYToMuMu_M-800_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0055	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	0.0011	START53_V7C2-v1
	DYToMuMu_M-1000_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0018	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$3.6 \cdot 10^{-4}$	START53_V7A-v1
	DYToMuMu_M-1000_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0018	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$3.6 \cdot 10^{-4}$	START53_V7C1-v1
	DYToMuMu_M-1000_CT10_TuneZ2star_8TeV-powheg-pythia6	0.0018	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$3.6 \cdot 10^{-4}$	START53_V7C2-v1
	DYToMuMu_M-1500_CT10_TuneZ2star_8TeV-powheg-pythia6	$1.7 \cdot 10^{-4}$	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$3.4 \cdot 10^{-5}$	START53_V7A-v1
	DYToMuMu_M-1500_CT10_TuneZ2star_8TeV-powheg-pythia6	$1.7 \cdot 10^{-4}$	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$3.4 \cdot 10^{-5}$	START53_V7C1-v1
	DYToMuMu_M-1500_CT10_TuneZ2star_8TeV-powheg-pythia6	$1.7 \cdot 10^{-4}$	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$3.4 \cdot 10^{-5}$	START53_V7C2-v1
	DYToMuMu_M-2000_CT10_TuneZ2star_8TeV-powheg-pythia6	$2.2 \cdot 10^{-5}$	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$4.5 \cdot 10^{-6}$	START53_V7A-v1
	DYToMuMu_M-2000_CT10_TuneZ2star_8TeV-powheg-pythia6	$2.2 \cdot 10^{-5}$	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$4.5 \cdot 10^{-6}$	START53_V7C1-v1
	DYToMuMu_M-2000_CT10_TuneZ2star_8TeV-powheg-pythia6	$2.2 \cdot 10^{-5}$	1.0	1.024 (NNLO)	$9.0 \cdot 10^4$	$4.5 \cdot 10^{-6}$	START53_V7C2-v1
	ZJetsToNuNu_PtZ-70To100_8TeV	33	1.0	1.0 (LO)	$2.0 \cdot 10^7$	0.029	START53_V7C-v1
	ZJetsToNuNu_PtZ-100_8TeV-madgraph	21	1.0	1.0 (LO)	$1.0 \cdot 10^7$	0.029	START53_V7C-v1

Process	Dataset Name	X-Section /pb	Filter Efficiency	k-Factor (Order)	Number Events	Event Weight	Conditions
Multi-Jet	QCD_Pt-0to5_TuneZ2star_8TeV_pythia6	$5.0 \cdot 10^{10}$	0.98	1.0 (LO)	$9.0 \cdot 10^{5}$	$9.6 \cdot 10^{8}$	START53_V7A-v1
	QCD_Pt-5to15_TuneZ2star_8TeV_pythia6	$4.3 \cdot 10^{10}$	1.0	1.0 (LO)	$1.0 \cdot 10^{6}$	$5.6 \cdot 10^{8}$	START53_V7A-v1
	QCD_Pt-15to30_TuneZ2star_8TeV_pythia6	$9.9 \cdot 10^{8}$	1.0	1.0 (LO)	$9.0 \cdot 10^{6}$	$1.9 \cdot 10^{6}$	START53_V7A-v1
	QCD_Pt-15to30_TuneZ2star_8TeV_pythia6	$9.9 \cdot 10^{8}$	1.0	1.0 (LO)	$3.0 \cdot 10^{5}$	$5.4 \cdot 10^{7}$	START53_V7A-v2
	QCD_Pt-30to50_TuneZ2star_8TeV_pythia6	$6.6 \cdot 10^{7}$	1.0	1.0 (LO)	$4.0 \cdot 10^{5}$	$3.2 \cdot 10^{6}$	START53_V7A-v1
	QCD_Pt-30to50_TuneZ2star_8TeV_pythia6	$6.6 \cdot 10^{7}$	1.0	1.0 (LO)	$6.0 \cdot 10^{6}$	$2.2 \cdot 10^{5}$	START53_V7A-v2
	QCD_Pt-50to80_TuneZ2star_8TeV_pythia6	$8.1 \cdot 10^{6}$	1.0	1.0 (LO)	$3.0 \cdot 10^{5}$	$4.2 \cdot 10^{5}$	START53_V7A-v1
	QCD_Pt-50to80_TuneZ2star_8TeV_pythia6	$8.1 \cdot 10^{6}$	1.0	1.0 (LO)	$5.0 \cdot 10^{6}$	$2.7 \cdot 10^{4}$	START53_V7A-v2
	QCD_Pt-80to120_TuneZ2star_8TeV_pythia6	$1.0 \cdot 10^{6}$	1.0	1.0 (LO)	$8.0 \cdot 10^{5}$	$2.4 \cdot 10^{4}$	START53_V7A-v2
	QCD_Pt-80to120_TuneZ2star_8TeV_pythia6	$1.0 \cdot 10^{6}$	1.0	1.0 (LO)	$5.0 \cdot 10^{6}$	3400	START53_V7A-v3
	QCD_Pt-120to170_TuneZ2star_8TeV_pythia6	$1.6 \cdot 10^{5}$	1.0	1.0 (LO)	$8.0 \cdot 10^{5}$	3500	START53_V7A-v2
	QCD_Pt-120to170_TuneZ2star_8TeV_pythia6	$1.6 \cdot 10^{5}$	1.0	1.0 (LO)	$5.0 \cdot 10^{6}$	510	START53_V7A-v3
	QCD_Pt-170to300_TuneZ2star_8TeV_pythia6	$3.4 \cdot 10^{4}$	1.0	1.0 (LO)	$2.0 \cdot 10^{5}$	2300	START53_V7A-v1
	QCD_Pt-170to300_TuneZ2star_8TeV_pythia6	$3.4 \cdot 10^{4}$	1.0	1.0 (LO)	$5.0 \cdot 10^{6}$	120	START53_V7A-v2
	QCD_Pt-170to300_TuneZ2star_8TeV_pythia6_v2	$3.4 \cdot 10^{4}$	1.0	1.0 (LO)	$1.0 \cdot 10^{7}$	34	START53_V7A-v1
	QCD_Pt-300to470_TuneZ2star_8TeV_pythia6	1800	1.0	1.0 (LO)	$3.0 \cdot 10^{5}$	110	START53_V7A-v1
	QCD_Pt-300to470_TuneZ2star_8TeV_pythia6	1800	1.0	1.0 (LO)	$5.0 \cdot 10^{6}$	5.8	START53_V7A-v2
	QCD_Pt-300to470_TuneZ2star_8TeV_pythia6_v2	1800	1.0	1.0 (LO)	$3.0 \cdot 10^{6}$	9.9	START53_V7A-v1
	QCD_Pt-300to470_TuneZ2star_8TeV_pythia6_v3	1800	1.0	1.0 (LO)	$1.0 \cdot 10^{7}$	1.7	START53_V7A-v1
	QCD_Pt-470to600_TuneZ2star_8TeV_pythia6	110	1.0	1.0 (LO)	$3.0 \cdot 10^{5}$	6.8	START53_V7A-v1
	QCD_Pt-470to600_TuneZ2star_8TeV_pythia6	110	1.0	1.0 (LO)	$3.0 \cdot 10^{6}$	0.56	START53_V7A-v2
	QCD_Pt-600to800_TuneZ2star_8TeV_pythia6	27	1.0	1.0 (LO)	$2.0 \cdot 10^{5}$	2.0	START53_V7A-v1
	QCD_Pt-600to800_TuneZ2star_8TeV_pythia6	27	1.0	1.0 (LO)	$3.0 \cdot 10^{6}$	0.13	START53_V7A-v2
	QCD_Pt-800to1000_TuneZ2star_8TeV_pythia6	3.6	1.0	1.0 (LO)	$3.0 \cdot 10^{5}$	0.22	START53_V7A-v1
	QCD_Pt-800to1000_TuneZ2star_8TeV_pythia6	3.6	1.0	1.0 (LO)	$3.0 \cdot 10^{6}$	0.017	START53_V7A-v2
	QCD_Pt-1000to1400_TuneZ2star_8TeV_pythia6	0.74	1.0	1.0 (LO)	$1.0 \cdot 10^{6}$	0.0074	START53_V7A-v1
	QCD_Pt-1400to1800_TuneZ2star_8TeV_pythia6	0.034	1.0	1.0 (LO)	$2.0 \cdot 10^{6}$	$3.3 \cdot 10^{-4}$	START53_V7A-v1
	QCD_Pt-1800_TuneZ2star_8TeV_pythia6	0.0018	1.0	1.0 (LO)	$9.0 \cdot 10^{5}$	$3.7 \cdot 10^{-5}$	START53_V7A-v1
QCD-BCtoE	QCD_Pt_20_30_BCtoE_TuneZ2star_8TeV_pythia6	$2.9 \cdot 10^{8}$	$5.8 \cdot 10^{-4}$	1.0 (LO)	$1.0 \cdot 10^{6}$	1900	START53_V7A-v1
	QCD_Pt_30_80_BCtoE_TuneZ2star_8TeV_pythia6	$7.4 \cdot 10^{7}$	0.0022	1.0 (LO)	$2.0 \cdot 10^{6}$	1600	START53_V7A-v1
	QCD_Pt_80_170_BCtoE_TuneZ2star_8TeV_pythia6	$1.2 \cdot 10^{6}$	0.011	1.0 (LO)	$1.0 \cdot 10^{6}$	130	START53_V7A-v1
	QCD_Pt_170_250_BCtoE_TuneZ2star_8TeV_pythia6	$3.1 \cdot 10^{4}$	0.020	1.0 (LO)	$1.0 \cdot 10^{6}$	6.4	START53_V7A-v1
	QCD_Pt_250_350_BCtoE_TuneZ2star_8TeV_pythia6	4300	0.024	1.0 (LO)	$2.0 \cdot 10^{6}$	1.0	START53_V7A-v1
	QCD_Pt_350_BCtoE_TuneZ2star_8TeV_pythia6	810	0.029	1.0 (LO)	$1.0 \cdot 10^{6}$	0.24	START53_V7A-v2
QCD-EM	QCD_Pt_20_30_EMEnriched_TuneZ2star_8TeV_pythia6	$2.9 \cdot 10^{8}$	0.010	1.0 (LO)	$3.0 \cdot 10^{7}$	1600	START53_V7A-v1
	QCD_Pt_30_80_EMEnriched_TuneZ2star_8TeV_pythia6	$7.4 \cdot 10^{7}$	0.062	1.0 (LO)	$3.0 \cdot 10^{7}$	2700	START53_V7A-v1
	QCD_Pt_80_170_EMEnriched_TuneZ2star_8TeV_pythia6	$1.2 \cdot 10^{6}$	0.15	1.0 (LO)	$3.0 \cdot 10^{7}$	100	START53_V7A-v1
	QCD_Pt_170_250_EMEnriched_TuneZ2star_8TeV_pythia6	$3.1 \cdot 10^{4}$	0.15	1.0 (LO)	$3.0 \cdot 10^{7}$	2.9	START53_V7A-v1
	QCD_Pt_250_350_EMEnriched_TuneZ2star_8TeV_pythia6	4300	0.13	1.0 (LO)	$3.0 \cdot 10^{7}$	0.32	START53_V7A-v1
	QCD_Pt_350_EMEnriched_TuneZ2star_8TeV_pythia6	810	0.11	1.0 (LO)	$3.0 \cdot 10^{7}$	0.052	START53_V7A-v1

Process	Dataset Name	X-Section /pb	Filter Efficiency	k-Factor (Order)	Number Events	Event Weight	Conditions
QCD-Mu	QCD_Pt-15to20_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	$7.0 \cdot 10^8$	0.0039	1.0 (LO)	$1.0 \cdot 10^6$	$3.1 \cdot 10^4$	START53_V7A-v2
	QCD_Pt-20to30_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	$2.9 \cdot 10^8$	0.0065	1.0 (LO)	$8.0 \cdot 10^6$	4300	START53_V7A-v1
	QCD_Pt-30to50_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	$6.6 \cdot 10^7$	0.012	1.0 (LO)	$9.0 \cdot 10^6$	1700	START53_V7A-v1
	QCD_Pt-50to80_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	$8.1 \cdot 10^6$	0.022	1.0 (LO)	$1.0 \cdot 10^7$	330	START53_V7A-v1
	QCD_Pt-80to120_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	$1.0 \cdot 10^6$	0.040	1.0 (LO)	$9.0 \cdot 10^6$	86	START53_V7A-v1
	QCD_Pt-120to170_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	$1.6 \cdot 10^5$	0.047	1.0 (LO)	$8.0 \cdot 10^6$	17	START53_V7A-v1
	QCD_Pt-170to300_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	$3.4 \cdot 10^4$	0.068	1.0 (LO)	$7.0 \cdot 10^6$	5.9	START53_V7A-v1
	QCD_Pt-300to470_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	1800	0.086	1.0 (LO)	$7.0 \cdot 10^6$	0.38	START53_V7A-v1
	QCD_Pt-470to600_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	120	0.10	1.0 (LO)	$3.0 \cdot 10^6$	0.061	START53_V7A-v1
	QCD_Pt-600to800_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	27	0.100	1.0 (LO)	$4.0 \cdot 10^6$	0.013	START53_V7A-v1
	QCD_Pt-800to1000_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	3.6	0.10	1.0 (LO)	$4.0 \cdot 10^6$	0.0018	START53_V7A-v1
	QCD_Pt-1000_MuEnrichedPt5_TuneZ2star_8TeV_pythia6	0.77	0.11	1.0 (LO)	$3.0 \cdot 10^6$	$4.3 \cdot 10^{-4}$	START53_V7A-v1
QCD-Mu15	QCD_Pt_20_MuEnrichedPt_15_TuneZ2star_8TeV_pythia6	$3.6 \cdot 10^8$	$3.7 \cdot 10^{-4}$	1.0 (LO)	$7.0 \cdot 10^6$	350	START53_V7A-v1
	QCD_Pt_20_MuEnrichedPt_15_TuneZ2star_8TeV_pythia6	$3.6 \cdot 10^8$	$3.7 \cdot 10^{-4}$	1.0 (LO)	$2.0 \cdot 10^7$	120	START53_V7A-v3
Top	T_s-channel_TuneZ2star_8TeV-powheg-tauola	2.8	1.0	1.34 (NNLL)	$2.0 \cdot 10^5$	0.29	START53_V7A-v1
	T_t-channel_TuneZ2star_8TeV-powheg-tauola	47	0.98	1.22 (NNLL)	$3.0 \cdot 10^6$	0.29	START53_V7A-v1
	T_t-channel_TuneZ2star_8TeV-powheg-tauola	47	0.98	1.22 (NNLL)	$2.0 \cdot 10^4$	47	START53_V7A-v2
	T_t-channel_TuneZ2star_8TeV-powheg-tauola	47	0.98	1.22 (NNLL)	$9.0 \cdot 10^4$	11	START53_V7A-v3
	T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	11	1.00	1.04 (NNLL)	$4.0 \cdot 10^5$	0.44	START53_V7A-v1
	Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola	1.6	1.0	1.12 (NNLL)	$1.0 \cdot 10^5$	0.25	START53_V7A-v1
	Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola	25	0.98	1.25 (NNLL)	$1.0 \cdot 10^6$	0.31	START53_V7A-v1
	Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	11	1.00	1.04 (NNLL)	$4.0 \cdot 10^5$	0.44	START53_V7A-v1
	TBZToLL_4F_TuneZ2star_8TeV-madgraph-tauola	0.022	1.0	0.53 (NLO)	$1.0 \cdot 10^5$	0.0015	START53_V7C-v1
Tri-Boson	WWWJets_8TeV-madgraph	0.082	1.0	0.98 (NLO)	$2.0 \cdot 10^5$	0.0072	START53_V7A-v1
	WWZNoGstarJets_8TeV-madgraph	0.063	1.0	0.92 (NLO)	$2.0 \cdot 10^5$	0.0052	START53_V7A-v1
	WZZNoGstarJets_8TeV-madgraph	0.019	1.0	1.02 (NLO)	$2.0 \cdot 10^5$	0.0018	START53_V7A-v1
	ZZZNoGstarJets_8TeV-madgraph	0.0046	1.0	1.2 (NLO)	$2.0 \cdot 10^5$	$5.0 \cdot 10^{-4}$	START53_V7A-v1
Upsilon	Upsilon1SToMuMu_2MuPtEtaFilter_tuneD6T_8TeV-pythia6-evtgen	$9.9 \cdot 10^4$	0.022	1.0 (LO)	$2.0 \cdot 10^6$	22	START53_V7A-v1
	Upsilon2SToMuMu_2MuPtEtaFilter_tuneD6T_8TeV-pythia6-evtgen	$1.3 \cdot 10^5$	0.011	1.0 (LO)	$2.0 \cdot 10^6$	14	START53_V7A-v1
	Upsilon3SToMuMu_2MuPtEtaFilter_tuneD6T_8TeV-pythia6	3800	0.15	1.0 (LO)	$1.0 \cdot 10^6$	5.7	START53_V7A-v1

A Comprehensive List of Simulation Samples

B Monte Carlo Production Configuration Files

B.1 Pythia Configuration for the QCD b Enriched Sample

The Pythia configuration, including comments shown below, was directly taken from the central CMS production. As no changes were made w.r.t. to event generation, no changes to the Pythia configuration was necessary with the exception of the `CKIN(3)` parameter, which defines lower bound of the kinematic binning variable.

```
MSEL=1          ! QCD hight pT processes
CKIN(3)=900.    ! minimum pt hat for hard interactions
MSTU(21)=1      ! Check on possible errors during program execution
MSTJ(22)=2      ! Decay those unstable particles
PARJ(71)=10.    ! for which ctau  10 mm
MSTP(33)=0      ! no K factors in hard cross sections
MSTP(2)=1       ! which order running alphaS
MSTP(51)=10042 ! structure function chosen (external PDF CTEQ6L1)
MSTP(52)=2      ! work with LHAPDF
PARP(82)=1.921 ! pt cutoff for multiparton interactions
PARP(89)=1800. ! sqrts for which PARP82 is set
PARP(90)=0.227 ! Multiple interactions: rescaling power
MSTP(95)=6      ! CR (color reconnection parameters)
PARP(77)=1.016 ! CR
PARP(78)=0.538 ! CR
PARP(80)=0.1    ! Prob. colored parton from BBR
PARP(83)=0.356 ! Multiple interactions: matter distribution parameter
PARP(84)=0.651 ! Multiple interactions: matter distribution parameter
PARP(62)=1.025 ! ISR cutoff
MSTP(91)=1      ! Gaussian primordial kT
PARP(93)=10.0   ! primordial kT-max
MSTP(81)=21     ! multiple parton interactions 1 is Pythia default
MSTP(82)=4      ! Defines the multi-parton model
```

B enrichment is performed via a CMSSW `EDFilter` object, which is to require at least one generator level b quark, identified by its PDG ID (± 5 for b and $\overline{\text{b}}$, respectively).

```
process.bbFilter = cms.EDFilter("MCSingleParticleFilter",
    Status = cms.untracked.vint32(2, 2),
    ParticleID = cms.untracked.vint32(5, -5)
)
```

B.2 Sherpa Configuration for the QCD Sample

The Sherpa configuration file closely follows an example from the Sherpa documentation [105]. The file is separated into multiple sections. The `(run)`-,`(integration)`- and `(mi)`-sections contain technical settings, such as the choice of matrix element generators, beam energies and multi-parton interaction handling scheme.

```
(run){
  % Tags
  FILE:=Myfile
  J1CUT:=300.;
  J2CUT:=20.;
  FSF:=1.;
  RSF:=1.;
  QSF:=1.;

  % general settings
  EVENTS 1M;
  EVENT_MODE = HepMC;
  EVENT_OUTPUT=HepMC_GenEvent[FILE]

  % avoid comix re-init after runcard modification
  WRITE_MAPPING_FILE 3;

  % tags and settings for scale definitions
  SCALES METS
  % tags and settings for ME generators
  LOOPGEN:=OpenLoops;
  ME_SIGNAL_GENERATOR Amegic LOOPGEN;
  EVENT_GENERATION_MODE Weighted;
  RESULT_DIRECTORY res_jJ1CUT_jJ2CUT_ffFSF_rfRSF_qfQSF;

  % model parameters
  MODEL SM;

  % collider setup
  BEAM_1  2212; BEAM_ENERGY_1 4000.0;
  BEAM_2  2212; BEAM_ENERGY_2 4000.0;
}(run)
(integration){
  FINISH_OPTIMIZATION = Off
}(integration)
(mi){
  MI_HANDLER = Amisic;
}(mi)
```

The `(processes)`-section is used to define the hard interaction process to be simulated. Here, the first line ("Process") dictates that $2 \to 2$ parton processes are simulated, where "93" is the shorthand name for partons in Sherpa. In the `(selector)`-section, an event selection scheme is defined that requires the presence of two `Anti-`k_t parton jets with respective p_T requirements.

```
(processes){
  Process 93 93 -> 93 93;
  NLO_QCD_Mode MC@NLO {2};
  Loop_Generator LOOPGEN;
  Order_EW 0;

  Integration_Error 0.05 {2};
  CKKW sqr(20./E_CMS);

  End process;
}(processes)

(selector){
  FastjetFinder  antikt 2  J2CUT  0.0  0.5
  FastjetFinder  antikt 1  J1CUT  0.0  0.5
}(selector)
```