RWTH-Aachen

Master's Thesis in Physics

Search for new Physics with the MUSiC Algorithm in 2018 CMS Data

Yannik Kaiser

presented to

The Faculty of Mathematics, Computer Science and Natural Sciences
at RWTH Aachen University

III. Institute of Physics A

supervised by

Prof. Dr. Thomas Hebbeker

Second Examiner

Prof. Dr. Martin Erdmann

Aachen, 20th December 2023

Contents

1. Abstract

The present work provides an initial evaluation of the Model Unspecific Search in CMS (MUSiC) based on data from CMS collected in 2018. The proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ correspond to an integrated luminosity of $59.8\,\text{fb}^{-1}$. MUSiC adopts an innovative approach by impartially and model-independently seeking deviations between measured data and theoretical predictions without any specific model-centric focus.
The analysis relies on events containing at least one high-energy electron or muon, utilizing Monte Carlo (MC) simulations for the Standard Model predictions. Notably, data is categorized into event classes based on their final states and the agreement with predictions is assessed through the kinematic distributions M_{inv}, $\sum |\vec{p}_T|$ and p_T^{miss}.
The search spans the entire distributions, encompassing all contiguous regions, with the Look-Elsewhere Effect (LEE) corrected to compute a global p-value. The resulting most significant deviations are scrutinized in detail and the overall agreement between measurement and prediction is qualitatively discussed. Despite rigorous analyses, no definitive evidence for new physics has been found. Nevertheless, the outcomes of this study provide a solid basis for future investigations.

2. The Standard Model

2.1. Introduction

The Standard Model (SM) in elementary particle physics is proving to be a highly successful theory, as evidenced by extensive experimental confirmations [1–3]. This model encompasses extensive knowledge of the fundamental particles that make up the universe and their complex interactions, while incorporating the principles of special relativity and quantum mechanics. It is important to note that the Standard Model in its current form does not include the gravitational force as an inherent component, and thus it remains as the only fundamental force not yet explained within the SM [3, 4]. In the Standard Model, all fundamental physical processes are described by the interactions of 17 elementary particles. These particles are systematically divided into matter and exchange particles, while their interactions are subdivided into the domains of strong, weak and electromagnetic forces (Fig.2.1) [3, 5, 6].

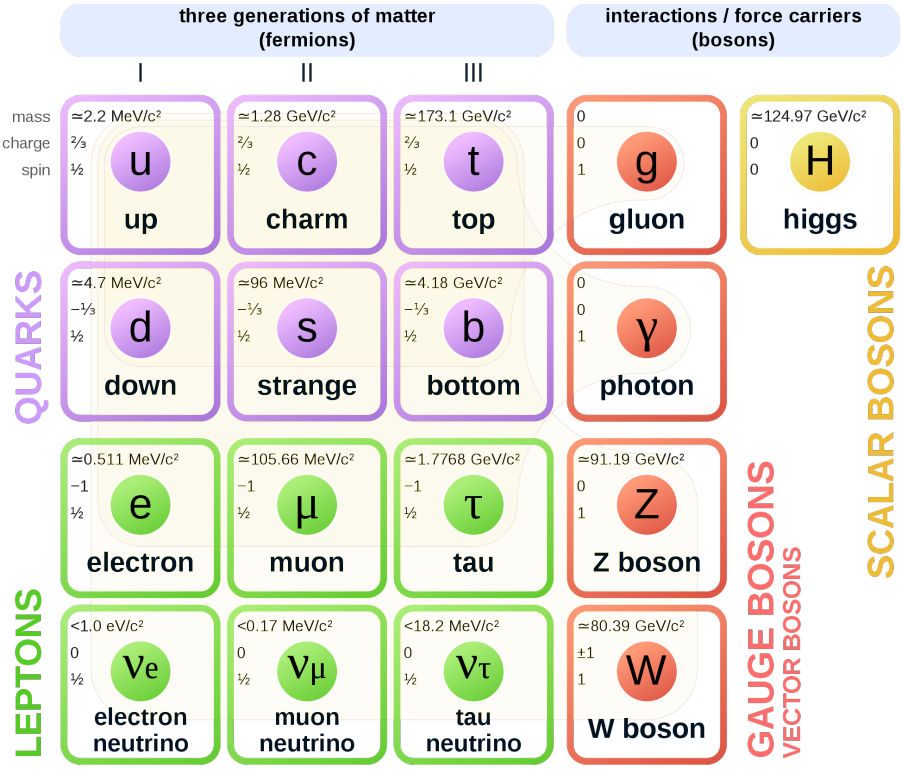


Fig. 2.1.: Standard Model overview, featuring twelve fundamental fermions and five fundamental bosons. Fermions are divided into quarks (purple) and leptons (green) across three generations. Bosons are categorized as gauge/vector bosons (red) and scalar bosons (yellow) (modified) [7].

2.1.1. Matter Particles

In the category of the matter particles are 12 of the 17 particles which are also called fermions because of their half-integer spin. Each particle is considered a point-like and indivisible object described by its mass and a set of quantum numbers. These quantum numbers include the spin, weak isospin, electric charge, and color charge of the particle. The Fermi-Dirac statistics describes the behavior of these particles and says that they must obey Pauli's exclusion principle. The Pauli principle states that no particle can occupy the same state, characterized by its quantum numbers, as another particle in the same system. These fermions are divided into quarks and leptons, which are distinguished into three generations based on their different masses [1, 3]. The matter particles of the second and third generation are thereby heavier than those of the first and are very unstable, whereby they decay after a certain time to particles of the first generation [2]. This also explains why the visible matter of our universe consists primarily of up/down quarks and electrons. If you take a closer look at quarks, they are the particles that are able to form atomic nuclei, which means they are able to interact strongly. Each quark has an electric charge of either $+\frac{2}{3}$ or $-\frac{1}{3}$ (Fig.2.1) and never exists in isolated form. In fact, only bound states, called hadrons, can have stable quarks. Hadrons consist of either two quarks, also known as mesons, or three quarks, known as baryons. Leptons, on the other hand, cannot form atomic nuclei but are able to exist on their own [8–10]. They exist thereby both in charged form (e, μ, τ) and (electrically) neutral form (ν_e, ν_μ, ν_τ) [3, 5]. The latter are the neutrinos, which both interact very rarely and are taken as massless in the Standard Model. Despite the Standard Model assumption that neutrinos are massless, numerous experiments indicate oscillations between the three generations of neutrinos [11, 12]. These oscillations would only be possible if the neutrinos possessed a non-vanishing mass. Experiments, such as solar, atmospheric, reactor, and accelerator neutrino experiments, have provided evidence for the existence of a mass for neutrinos. Despite these indications, the actual mass of neutrinos has not yet been accurately determined. Indeed, only upper limits on the individual neutrino masses can be derived from specific experiments ($m_{\nu_e} < 1\,\mathrm{eV}$, $m_{\nu_\mu} < 0.19\,\mathrm{MeV}$ and $m_{\nu_\tau} < 18.2\,\mathrm{MeV}$) [1, 3, 13].

2.1.2. Exchange particles & Interactions

Besides the matter particles, there are five other particles called bosons (gluon, photon, W boson, Z boson and the Higgs boson). They are divided according to their spin into four vector bosons (spin 1) and one scalar boson (spin 0). The vector bosons are associated with the fundamental forces because they are the exchange particles that act as carriers [3, 14]. The fundamental forces include the electromagnetic force, the strong force, the weak force and the gravitational force. The gravitational force is not considered in the Standard Model of particle physics. Although it plays a very important role in phenomena on large scales, with primarily neutrally charged objects, it is negligibly small compared to the other 3 forces in the size range of individual particles.

Electromagnetic interaction

The photon (γ) is the exchange particle of the electromagnetic force. In this interaction, both quarks and the leptons e, μ and τ can participate by reacting with each other either by annihilation, radiation or absorption of a photon. In this process, the uncharged photons only couple to charged particles. The effect of this force is not limited to a range and it is also responsible for most of the known reactions in our world, because it is e.g. the primary acting force, which is responsible for the structure and the properties of atoms.

Weak interaction

The second fundamental force is the weak force, which is not limited in its range, just like the electromagnetic force, but loses strength rapidly with increasing distance, which makes it perceptible up to a range of $\sim 10^{-17}\,\mathrm{m}$ [6]. The cause for this are the exchange particles of the weak force, the Z^0 boson and the $W^{\pm}$ bosons. With their heavy masses of $80.4\,\mathrm{GeV}$ ($W^{\pm}$) and $91.2\,\mathrm{GeV}$ (Z^0), they are very unstable and have very short lifetimes on the order of $10^{-25}\,\mathrm{s}$. Here, the $W^{\pm}$ bosons are charged, while the Z^0 boson is neutral. As described later (Subsec. 2.2.2) one is additionally able to unify the weak force and the electromagnetic force in a common unbound electroweak force with the carriers γ, Z^0 and $W^{\pm}$.

In 1956, the Wu experiment [15] found that parity, a fundamental symmetry in physics, is not preserved in the weak interaction. Parity refers to the behavior of a physical system when the spatial coordinates are reversed. Conservation of parity means that the physical processes remain unchanged, while a violation of parity indicates that the processes depend on the spatial orientation. A crucial role in this phenomenon is played by $W^{\pm}$ bosons, which interact exclusively with fermions, which have left-handed chirality, and with anti-fermions, which have right-handed chirality. Chirality is closely related to helicity, which describes the orientation of spin relative to the direction of motion. In general, chirality and helicity are distinct from each other, except when dealing with massless particles or the ultra-relativistic region. In this case chirality and helicity are equal. For particles with low energy and large mass, superpositions of left-handed and right-handed helicity states can be observed, since there are reference frames in which the direction of flight is opposite. However, neutrinos have the special property of being massless in the theory of the Standard Model, which makes chirality coincident with helicity in their case. Since right-handed neutrinos do not participate in the weak interaction and the weak interaction is generally the only form of interaction in which neutrinos participate, right-handed neutrinos and left-handed anti-neutrinos do not interact with ordinary matter [3, 10].

Strong interaction

The strong interaction is the interaction between quarks that explains the cohesion of hadrons. The range of the strong interaction is on the order of an atomic nucleus (~10^{-15} m) [6]. This is due to the exchange particles of the strong interaction, the gluons (g). These have their own color/color charge, which is the strong force equivalent of the electric charge. In this way, the gluons are able to couple with their own carriers without losing strength, but instead gaining potential as the distance increases. This also leads to the fact that no color-charged particle can appear individually [3, 8]. There are altogether eight different gluons, which are assigned to the color changes red, blue, green and their combinations.

2.2. Gauge Groups

The interactions between particles and fields, especially in quantum field theory, can be described mathematically using gauge theories. A key idea of these theories is the use of symmetries, i.e. mathematical transformations that ensure that the physical laws in a system remain unchanged.
The so-called gauge symmetry or gauge invariance is a special kind of symmetry in gauge theories. It states that physical results remain invariant when certain mathematical gauge transformations (U) are applied to fields (ψ). This allows different mathematical representations of the same physical state to be equivalent (Eq.2.1) [8].

$$\psi(x) \rightarrow \psi'(x) = U\psi(x) \tag{2.1}$$

We can start to understand the local gauge invariance within the Standard Model with the Euler-Lagrange equation (Eq.2.2).

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \Phi_i)} \right) = \frac{\partial \mathcal{L}}{\partial \Phi_i} \quad (i = 1, 2, 3, ...) \tag{2.2}$$

Where $\mathcal{L}$ is a function of the fields Φ and their derivatives. If one now demands the local gauge invariance, this leads to additional terms in $\mathcal{L}$ which describes the interactions of particles by means of the boson fields [16].

2.2.1. Quantum Electrodynamics (QED)

In QED, one starts with the Dirac Lagrangian of a free massive fermion (Eq.2.3).

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi \tag{2.3}$$

In which ψ stands for the wave function of the particles, γ^μ for the gamma matrix and m for the mass. Next, one introduces the required gauge invariance (Eq.2.1). In QED, the gauge invariance is based on the $U(1)$ symmetry, leading to the emergence of the photon as the mediator of the electromagnetic interaction. With $U = e^{i\Theta(x)}$, this gives an equation for $\mathcal{L}$ (Eq.2.4):

$$\mathcal{L} \rightarrow \mathcal{L}' = \mathcal{L} - \bar{\psi}\gamma^\mu\psi(\partial_\mu\Theta) \tag{2.4}$$

To preserve the invariance of the equation under this transformation, one extends the equation by a vector field A_μ, which here stands for the forces between charges (Eq.2.5 & Eq.2.6).

$$\mathcal{L} = (i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi) - (q\bar{\psi}\gamma^\mu\psi)A_\mu \tag{2.5}$$

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu - \frac{1}{q}\partial_\mu\Theta(x) \tag{2.6}$$

with q as charge. By adding the vector field A_μ, one more term must be added to describe the free vector particles. This is done by adding the Proca-Lagrangian equation (Eq.2.7):

$$\mathcal{L}_{Proca} = -\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{8\pi}m_A^2A^\nu A_\nu \tag{2.7}$$

with the electromagnetic field tensor $F^{\mu\nu} = \partial_\mu A^\nu - \partial^\nu A^\mu$ and the mass of the field m_A. Since the second term of Eq.2.7 is not invariant under the transformation of Eq.2.6, consequently $m_A = 0$ must be to maintain invariance. Thus, the final Langrange equation of QED (also called Langrange density) (Eq.2.8) is obtained [10, 16]:

$$\mathcal{L}_{U(1)} = \underbrace{i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi}_{\text{lepton propagation}} - \underbrace{\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu}}_{\text{photon propagation}} - \underbrace{(q\bar{\psi}\gamma^\mu\psi)A_\mu}_{\text{lepton-photon interaction}} \tag{2.8}$$

2.2.2. Electroweak unification

When one speaks of electroweak unification, one is speaking of the theory in particle physics that unifies the electromagnetic interaction and the weak interaction. This unification is mediated by mediator bosons, including the W^+ and W^- bosons with masses of about 80.4 GeV and the neutral Z^0 boson with mass of about 91.2 GeV.
The weak interaction is called *weak* because it has a limited strength compared to the electromagnetic interaction. Also, the mediator bosons are massive, with a very short lifetime of about 10^{-25} s and have a small coupling constant. At higher energies, the effective cross sections of

this interaction increase [10].
The solution to unify the electromagnetic and weak interactions was worked out by Sheldon Lee Glashow [17], Abdus Salam [18] and Steven Weinberg [19]. They resorted to the group structure of $SU(2) \otimes U(1)$ and postulated four massless fields: W^1 & W^2 (two charged fields) and W^3 & B (two neutral fields). These fields do not exist in nature, but mix with each other (Eq.2.9).

$$W^{\pm} = \sqrt{\frac{1}{2\pi}} \left(W^1 \mp iW^2\right)$$

$$Z^0 = -B \sin\left(\Theta_W\right) + W^3 \cos\left(\Theta_W\right) \tag{2.9}$$

$$\gamma = B \cos\left(\Theta_W\right) + W^3 \sin\left(\Theta_W\right)$$

With Θ_W as the mixing angle, experimentally determined as $\sin\left(\Theta_W\right) = 0.2312$ [20]. The $W^{\pm}$ bosons have a central role in the weak interaction as they mediate the charge transformation in the weak interaction. They are unique in that they can change the *flavor* of a particle, couple exclusively to left-handed particles and carry the weak isospin as their charge. This fact leads to the violation of parity, which means that the interaction between left- and right-handed particles is different. The Z boson, on the other hand, represents the neutral component of the electroweak interaction and is called the *neutral current*. It carries the hypercharge Y, which is related to the electric charge Q and the isospin I by the Gell-Mann-Nishijima formula (Eq.2.10) and couples to both left-handed and right-handed particles, but with different strengths [8, 10, 16].

$$Q = I^3 + \frac{1}{2}Y \tag{2.10}$$

With I^3 as the third component of the weak isospin. However, electroweak unification cannot explain the masses of the Z^0 boson and the $W^{\pm}$ bosons determined by experiments. The explanation lies in the Higgs mechanism (Subsec.2.3), which generates the masses of these bosons by interactions with the Higgs field.

2.2.3. Quantum Chromodynamics (QCD)

Quantum Chromo Dynamic, or QCD, is the theory that describes the strong interaction between quarks and gluons, which are the building blocks of hadrons such as protons and neutrons. This strong interaction is mediated by the color charge. There are three different color charges: Red, Blue and Green, which are represented in mathematical expressions by 3 spinors (ψ_r, ψ_b, ψ_g) in a component vector (Eq.2.11) [16].

$$\psi = \begin{pmatrix} \psi_r \\ \psi_b \\ \psi_g \end{pmatrix} \tag{2.11}$$

A crucial feature of QCD is the local gauge invariance, which is called SU(3) transformation. This gauge invariance leads to the description of the spinors representing the color charges and the existence of 8 gauge fields called gluons. Gluons are massless and carry color charges themselves (1 color & 1 anti-colour). Thus they can interact with quarks as well as with themselves. A consequence of QCD is the phenomenon of confinement. This means that single quarks never appear isolated, but always in bound states, like mesons (quark-antiquark pairs) or baryons, like protons and neutrons, which on the other hand are combinations of three quarks in a color-neutral state. At large distances between quarks, the strong interaction remains constant and leads to the formation of these color-neutral states. As published in July 2015 [21], QCD also allows for more exotic states such as pentaquarks, where five quarks (all three colors plus one color and its anti-color) occur in a color-neutral state. The strength of the strong interaction is described by the so-called coupling constant (α_S), which decreases with increasing momentum transfer intensity $|Q^2|$.

$$\alpha_S(Q^2) = \frac{12\pi}{(33 - 2N_f) \ln\left(\frac{Q^2}{\Lambda^2}\right)} \tag{2.12}$$

$$\Lambda^2 = q_0^2 \exp\left(-\frac{1}{A\alpha_S(q_0^2)}\right), \qquad A = \frac{33 - 2N_f}{12\pi} \tag{2.13}$$

with charge q_0, N_f = number of flavours and Λ as scale parameter, which till today cannot be calculated from QCD but has to be determined experimentally [22]. This phenomenon, known as asymptotic freedom, means that at very small distances ($r \ll R_{proton}$, $|Q^2| \gg 0$) the interaction between quarks becomes weak, and the quarks behave almost like free particles [8, 10, 16].

2.3. Higgs

The Higgs mechanism explains how fundamental elementary particles acquire their mass. This theoretical concept was first proposed in 1964 by the physicists Peter Higgs [23], François Englert, Robert Brout [24], T.W.B. Kibble, Carl R. Hagen, and Gerald Guralnik [25] and received its experimental confirmation in 2012 by the CMS and ATLAS collaborations at the Large Hadron Collider (LHC) [26, 27]. A fundamental problem associated with the electroweak

interaction is the fact that the W and Z bosons responsible for this interaction have mass, whereas the original theory is based on the principle of gauge invariance, according to which these bosons should be massless. This contradiction is resolved by the Higgs mechanism. The Higgs mechanism ensures that the local gauge invariance can be restored despite a mass of the bosons in the electroweak interaction. To explain the mass of the W and Z bosons, a new field, the Higgs field, is integrated into the Lagrangian (Eq.2.14).

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \psi)^*(\partial^\mu \psi) + \frac{1}{2}\mu^2(\psi^*\psi) - \frac{1}{4}\lambda^2(\psi^*\psi)^2 \tag{2.14}$$

with $\psi = \psi_1 + i\psi_2$ a complex field, λ a real constant and μ as complex constant. This Higgs field pervades the space and interacts with the elementary particles. The interaction between fermions and the Higgs field takes place via the so-called Yukawa coupling, whose strength is proportional to the mass of the particles (Fig.2.2) [6, 14, 16].

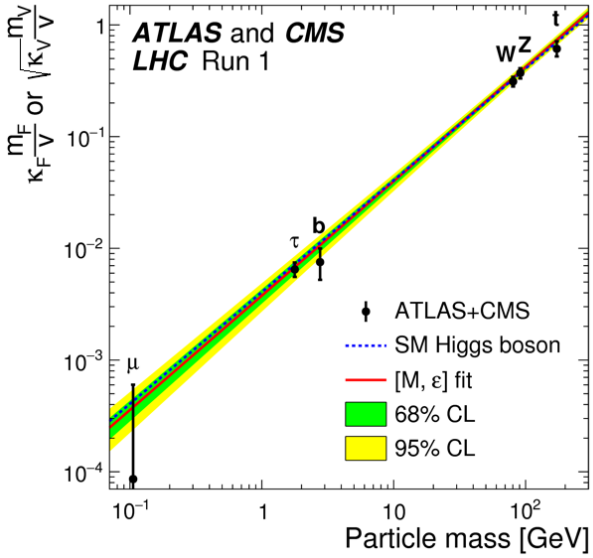


Fig. 2.2: Measured couplings of the Higgs boson to various particles. Discrepancies between the measured values and the anticipated outcomes in the Standard Model (m_F/v or m_v/v for fermions and bosons, with v = 246 GeV denoting the vacuum expectation value of the Higgs field) are expressed through the parameters κ_F or $\sqrt{\kappa_v}$, both assuming a value of one in accordance with the Standard Model (modified) [28].

One aspect of the Higgs mechanism is that it produces spontaneous symmetry breaking by generating an asymmetry in the ground state of the potential V (Eq.2.15) in the Higgs field ψ at $\psi = \pm\frac{\mu}{\lambda}$ when $\mu > 0$ [10].

$$V(\psi) = \frac{1}{2}\mu^2\psi^2 + \frac{1}{4}\lambda\psi^4 \tag{2.15}$$

This asymmetry ensures that the electroweak symmetry in the Lagrangian of the theory is effectively restored as the W and Z bosons acquire mass. Their mass is also proportional to their coupling with the Higgs field. The Higgs particle itself, created in the course of interactions with the Higgs field, was detected in 2012 by the CMS and ATLAS experiments at the LHC. The mass of the Higgs particle is nowadays defined as about 125.10 ± 0.14 GeV [20]. This groundbreaking discovery not only confirmed the existence of the Higgs mechanism, but also

illustrated the central role of the Higgs field in the creation of masses for elementary particles.

2.4. Parton density functions

Parton Density Functions (PDFs) are essential components of modern particle physics and are used to describe the distribution of momentum within protons. They indicate the probability with which a parton p (be it a quark or a gluon) carries a certain fraction x of the total momentum of a proton in its rest frame when a momentum transfer Q occurs. The exact mathematical form of the parton density functions cannot be derived in advance. Therefore, one makes use of empirically fitted smooth functions obtained from experimental data, e.g. from experiments such as HERA. These experimentally determined data form the basis for the modeling of the Parton Density Functions and enable a realistic simulation of scattering processes, including the calculation of effective cross sections and the consideration of kinematics in hard scattering processes. It is not possible to predict in advance which specific partons will participate in a given interaction when two protons collide. Therefore, models and simulations must be used to predict the probability and nature of the interactions. A key aspect is the scale dependence of the Parton Density Functions, which is made possible by the DGLAP equations (Eq.2.16).

$$\sigma(PP \to X) = \sum_{i,j} \int_0^1 \int_0^1 dx_1\, dx_2 PDF(p_i, x_1, Q^2) PDF(p_j, x_1, Q^2) \sigma(p_i p_j \to X) \qquad (2.16)$$

with the partonic effective cross sections $\sigma(p_i p_j \to X)$ for different parton combinations (p_i and p_j). However, the Q-value requires that the PDFs be known at a different reference $Q_0 \neq Q$. These equations allow the PDF contributions to be measured at a particular scale and the results extrapolated to other energy scales. The application of the DGLAP equations, which describe the development of the parton densities on higher energy scales, plays a crucial role in understanding the PDFs at different energies and using them for precise calculations in different scattering processes. Moreover, it is fundamental to understand the complex dynamics of partons in protons and to analyze their contributions to scattering processes [8, 10, 13].

2.5. Shortcomings of the Standard Model

While the Standard Model has achieved notable successes in describing fundamental particle interactions, it remains incomplete, leaving several unresolved questions and unexplained phenomena. Notably absent from its purview is the fourth fundamental interaction, gravity. Attempts to unify all known forces into a single theoretical framework have been hampered. Although successful in uniting electromagnetic and weak interactions through the $\mathrm{SU(2)_L \times U(1)_Y}$ gauge groups, the strong interaction described by the $\mathrm{SU(3)_C}$ gauge group and gravity have yet

to be seamlessly integrated [3, 8].

The concept of a *Grand Unified Theory* (GUT) (Fig. 2.3), encapsulating the electromagnetic, weak, and strong interactions through gauge groups, remains an overarching objective. The SU(5) gauge group represents the smallest unifying group, offering predictions that approximate observed values but fall short of complete concordance.

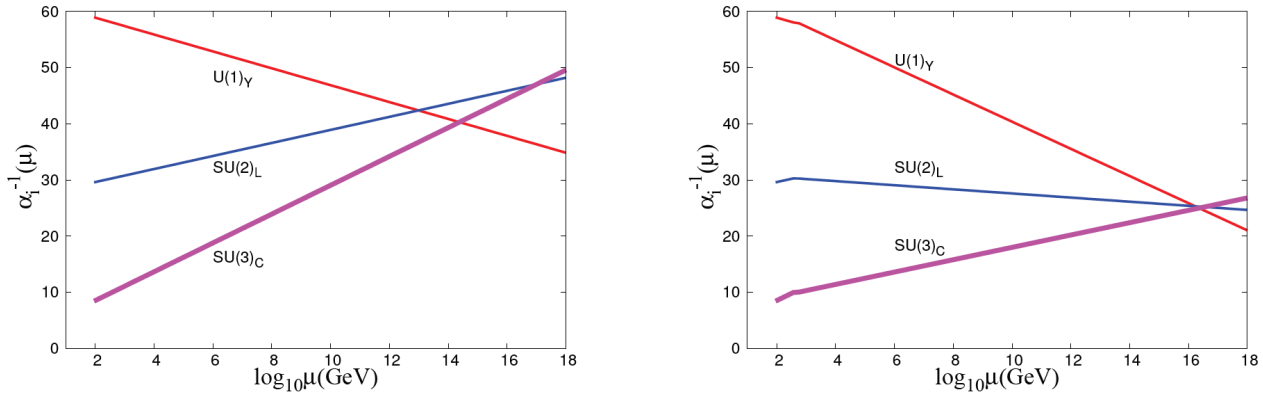


Fig. 2.3.: Variation of the three fine structure constants, denoted as $\alpha_i^{-1}(\mu) = \frac{g_i^2(\mu)}{4\pi}$. On the left, the depiction corresponds to the Standard Model, while on the right, it reflects the Standard Model augmented by the Supersymmetric (SUSY) extension, specifically within the framework of the Grand Unified Theory (GUT) idea of the Minimal Supersymmetric Standard Model (MSSM) [29].

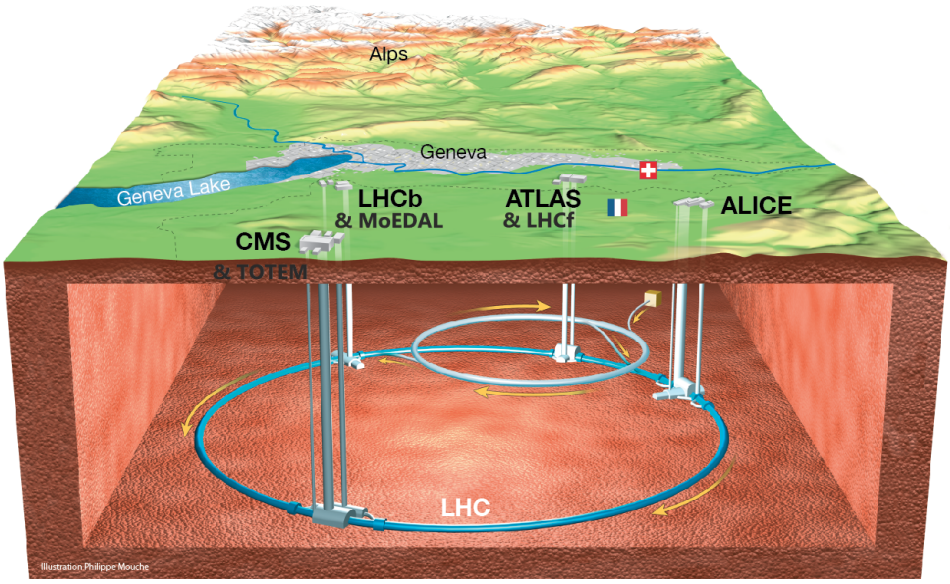
Further complexities arise in the understanding of the universe's composition. While the SM effectively elucidates the matter component, constituting approx. 5% of the universe, dark matter (approx. 26%) and dark energy (approx. 69%) defy explanation within the SM framework. Proposed theoretical models await experimental confirmation, presenting an ongoing challenge in particle physics.

Additional enigmas persist, such as the unexplained number of fermions and their mass hierarchy. The SM fails to account for the mass of neutrinos and the exclusive existence of left-handed neutrinos. Moreover, the SM relies on nineteen free parameters, the origins and values of which remain elusive. The unresolved nature of these parameters underscores the continued quest for a more comprehensive theoretical framework in particle physics.

3. Experimental Setup

3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) is a particle accelerator that stretches over a distance of 27 kilometers and is located at the European Organization for Nuclear Research (CERN) in Geneva (Fig.3.1). This makes the LHC the world's largest particle accelerator, with the highest center-of-mass energy ever achieved. The LHC began its operation in 2008, performing both proton-proton collisions and heavy ion collisions. The individual particles are accelerated to speeds close to the speed of light and reach a center-of-mass energy of up to $\sqrt{s} = 13.6\,\text{TeV}$ in proton-proton collisions. One of the greatest successes of the LHC so far is the discovery of the Higgs boson in 2012 [26, 27].



5

Fig. 3.1.: Overall view of the LHC (modified) [30]

The particle accelerator itself is built deep underground, about 50-170 m below the surface, to be protected from cosmic radiation. Inside the LHC's tubes, which have a diameter of 5.6 cm, is an extremely high vacuum pressure of up to 10^{-13} bar to prevent unwanted collisions with air molecules. In addition, the system has 1232 dipole magnets, 392 quadrupole magnets, 688 sextupole magnets, and 16 octupole magnets, which use a magnetic field of up to 8.33 T to keep

particle beams focused and on their circular course. To achieve such magnetic field strengths, the superconducting magnets are cooled down to 1.5 K [14]. Due to its size and the synchrotron nature of the LHC, particles must first be accelerated to a minimum entry velocity. This process takes place via a sequence of pre-accelerators (Fig.3.2).

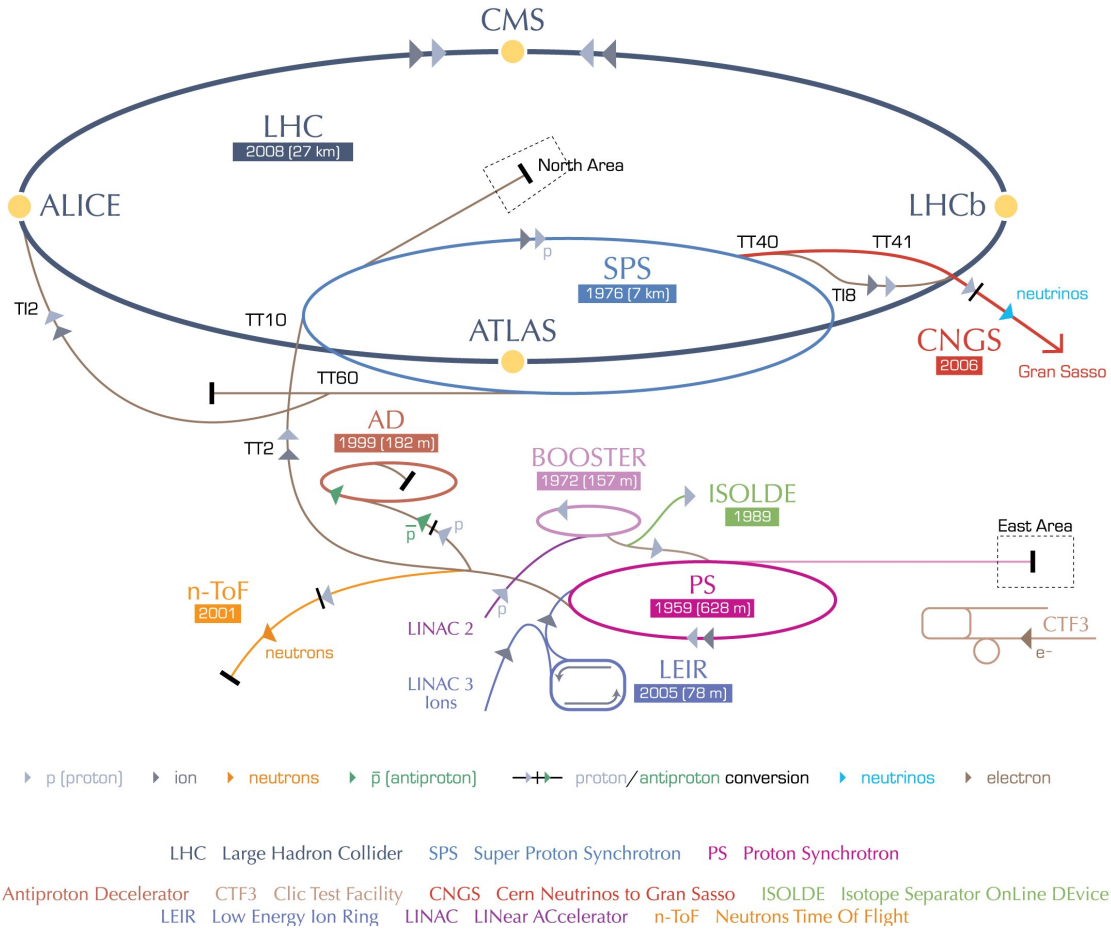


Fig. 3.2.: Illustrative representation of the CERN accelerator complex, depicting the Large Hadron Collider (LHC) in conjunction with its antecedent accelerators, namely LINAC2, PSB, PS and SPS, alongside additional accelerative structures. (modified) [31]

In the following, we focus exclusively on the acceleration process of the protons. First the protons are grouped into bunches and reach an energy of 50 GeV thanks to the linear accelerator 2 (LINAC 2). They then pass through the proton synchrotron booster (PSB), the proton synchrotron (PS) and the super proton synchrotron (SPS). After passing through the SPS, the bunches have a collimated length of around 1.35 ns, which brings them to an energy of around 450 GeV. After the distance between the bunches is reduced to 50 ns, they are fed into the LHC. Here, each bunch eventually reaches its final energy of 6.8 TeV [1, 3, 6].
The main goal of the LHC is to explore fundamental particles and their interactions in the highest detail and to search for phenomena beyond the Standard Model. To accomplish these tasks, four experiments are set up at the LHC [1, 3, 5, 32]:

- The **ALICE** experiment (A Large Ion Collider Experiment) was designed to study heavy ion collisions simulating conditions that existed immediately after the Big Bang.
- **ATLAS** (A Toroidal LHC Apparatus) and **CMS** (Compact Muon Solenoid) are versatile detectors that analyze proton and heavy ion collisions at extremely high energies. They encompass a broad physics programme that deals with the intricacies of the Standard Model, including the Higgs boson, while also exploring phenomena beyond the Standard Model, such as additional space dimensions and particles, that could potentially account for the existence of dark matter.
- The **LHCb** experiment (Large Hadron Collider beauty) focuses on studying the properties of b-hadrons during proton-proton collisions. Its goal is to gain a deeper understanding of the asymmetry between matter and antimatter in the universe.

3.2. Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is a central component of the Large Hadron Collider's extensive detector ensemble. The researchers involved in the studies and analyses of the CMS detector are members of the CMS Collaboration. The CMS experimental apparatus is shaped like a cylinder with a diameter of about 15 m, a length of 21 m, and a mass of 14.000 t. Also impressive is the ability to generate a magnetic field of up to 3.8 T [3, 5, 6, 33]. For a detailed and comprehensive account of the structural and operational intricacies of the CMS detector, see the authoritative reference in [34].

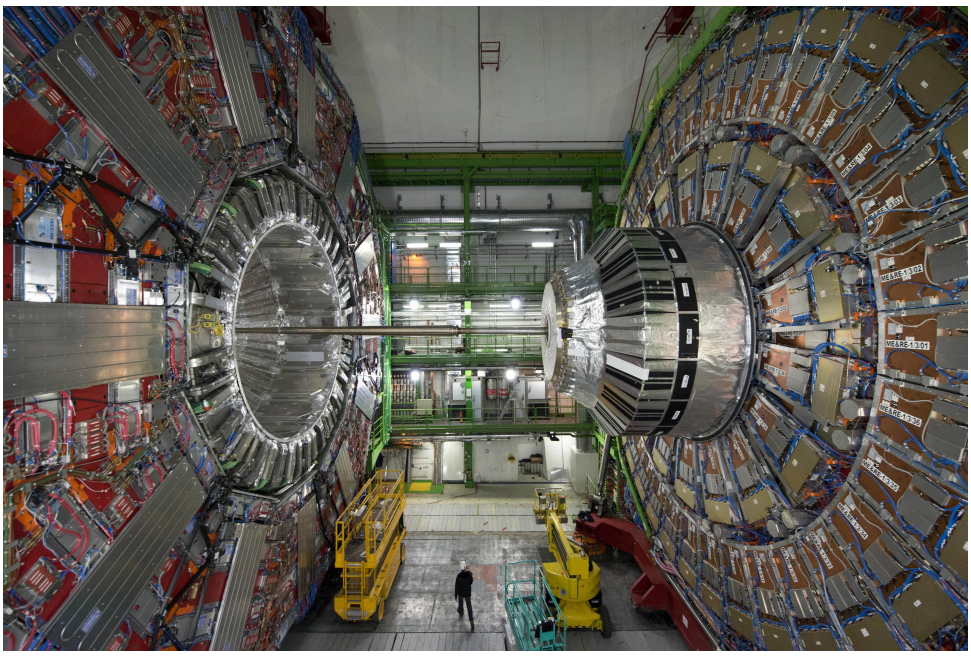
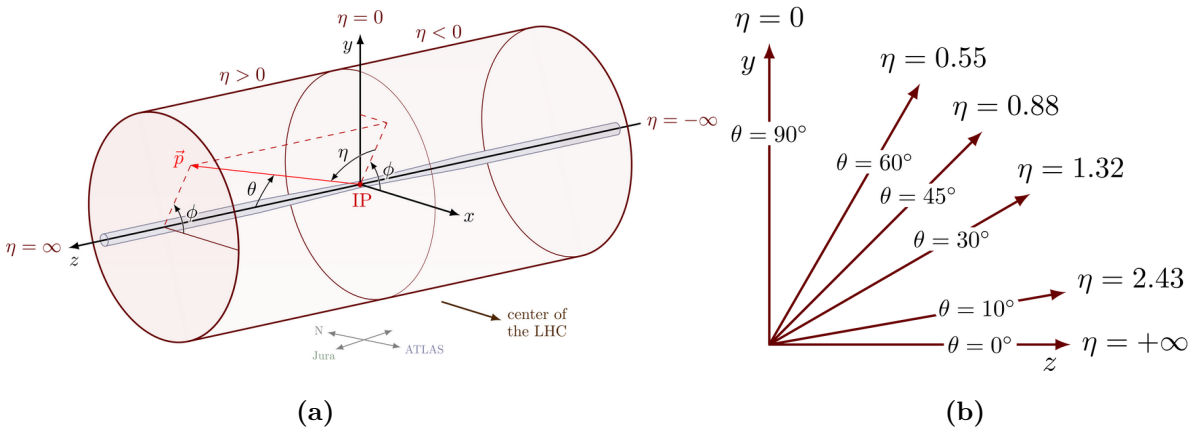


Fig. 3.3.: An overview of the Compact Muon Solenoid, featuring a human figure at its center to provide a reference scale [33].

3.2.1. Coordinate System of CMS

In the context of position determination in the CMS detector, a special coordinate system is used. This coordinate system has its origin precisely in the center of the particle collisions and allows an exact spatial localization of the events. The axes are aligned as follows: The x-axis points in the direction of the center of the LHC, the y-axis runs vertically upwards, while the z-axis is aligned along the particle beam (Fig.3.4).



(a) **(b)**

Fig. 3.4.: Graphical representation of the coordinate system of the Compact Muon Solenoid (CMS) in the Large Hadron Collider (LHC). **Left:** Representation of the relationships between the pseudorapidity η, the Cartesian coordinates (x, y, z) and the polar angles (θ, ϕ). **Right:** Equivalent representation of typical pseudorapidity values with the corresponding polar angles [35, 36].

In this coordinate system the radial distance r is similar in its function to the polar coordinate system. In contrast, no use is made of the angle Θ. Instead, the pseudorapidity $\eta = -\ln(\tan(\Theta/2))$ is used, along with the azimuth angle Φ, which is in the xy-plane. This choice brings the advantage that in the high energy range (E$\gg$m) the pseudorapidity is equal to the particle rapidity [1, 3, 14].

3.2.2. The CMS Detector

The detector itself is divided into two disjoint sectors: the rotationally symmetric barrel region and the two end caps. These sectors are defined in terms of their pseudorapidity $|\eta|$ (e.g. for the ECAL (subseq.3.2.2) barrel region: $|\eta| < 1.4442$, endcap region: $1.566 < |\eta| < 2.5$) [3, 14]. Inside the CMS detector, various subdetectors are positioned to detect the broadest possible spectrum of elementary particles and precisely measure their fundamental properties such as momentum, mass, charge and energy. In addition, a two-stage trigger system is used (Subsec.3.2.3). The subdetectors of CMS listed here are shown according to their radially symmetric arrangement from the inside to the outside (Fig.3.5).

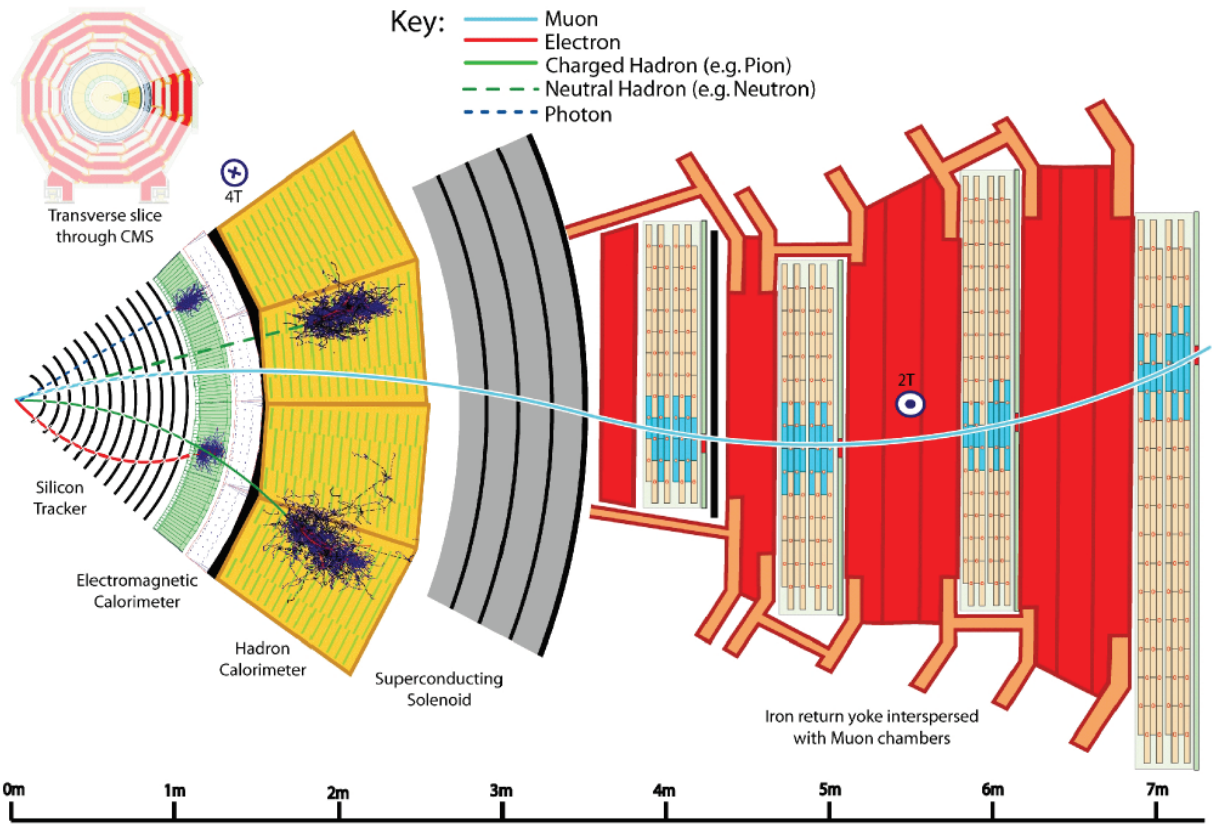


Fig. 3.5.: Comprehensive description of the various sub-detectors within the Compact Muon Solenoid, together with example particles that can be measured [37].

Silicon tracker

The innermost section of the CMS detector contains silicon pixel and silicon strip detectors [38] with 1440 pixel modules, a pixel size of 100 μm x 150 μm (65 million pixels), and a cell size of 10 cm x 80 mm for the silicon strip (15.200 modules with a total of 10 million silicon strips), which together act as silicon trackers (Fig.3.6). They are used to measure the tracks and momenta of charged particles up to $|\eta| < 2.5$ with an accuracy of 10 μm. Their design and operation allow precise detection of the high particle flux density near the interaction point while providing excellent spatial resolution. The silicon trackers consist of a highly granular structure to meet the requirements. Here, silicon detectors were chosen due to their suitability for module formation as well as their fast response time. To minimize radiation damage, the trackers are operated at a temperature of $-20\,^{\circ}$C. The efficiency of the tracking system in measuring the properties of the particles is 90% for electrons, 98% for muons and varies between 85%-95% for hadrons (depending on their energy). The pixel and strip detectors in the silicon tracker are divided into three subsystems [13, 16, 39]:

- **Pixel Detector:** This part forms the innermost region of the tracker and consists of four 2D layers extending at distances of 3, 7, 11, and 16 cm from the interaction point. This arrangement allows for three-dimensional reconstruction of the tracks. As charged

particles traverse the silicon, they lose energy and if a sufficient amount of energy is deposited, electrons will be ejected from the silicon atoms. These electrons are collected by an applied voltage and converted into an electrical signal. Each pixel has a readout chip that amplifies this electrical signal.

- **Tracker Inner Barrel (TIB) and Tracker Inner Disk (TID):** For larger radii of more than 20 cm, the particle flux decreases, so silicon microstrips are used here instead of pixels. The TIB consists of four concentric layers of microstrips with widths between 80 and 180 μm, which function similarly to the silicon pixels. In addition, there are two TIDs, one at each end of the TIB. These consist of three small disks each, which allow reconstruction of tracks with higher pseudorapidity (η).

- **Tracker Outer Barrel (TOB) and Tracker EndCaps (TEC):** These surround the TIB and TID and contribute to further trace capture. The TOB consists of six concentric layers of silicon microstrips. The TECs at the end of each TOB layer terminate the tracker system and allow reconstruction of tracks with higher pseudorapidity, similar to the TIDs.

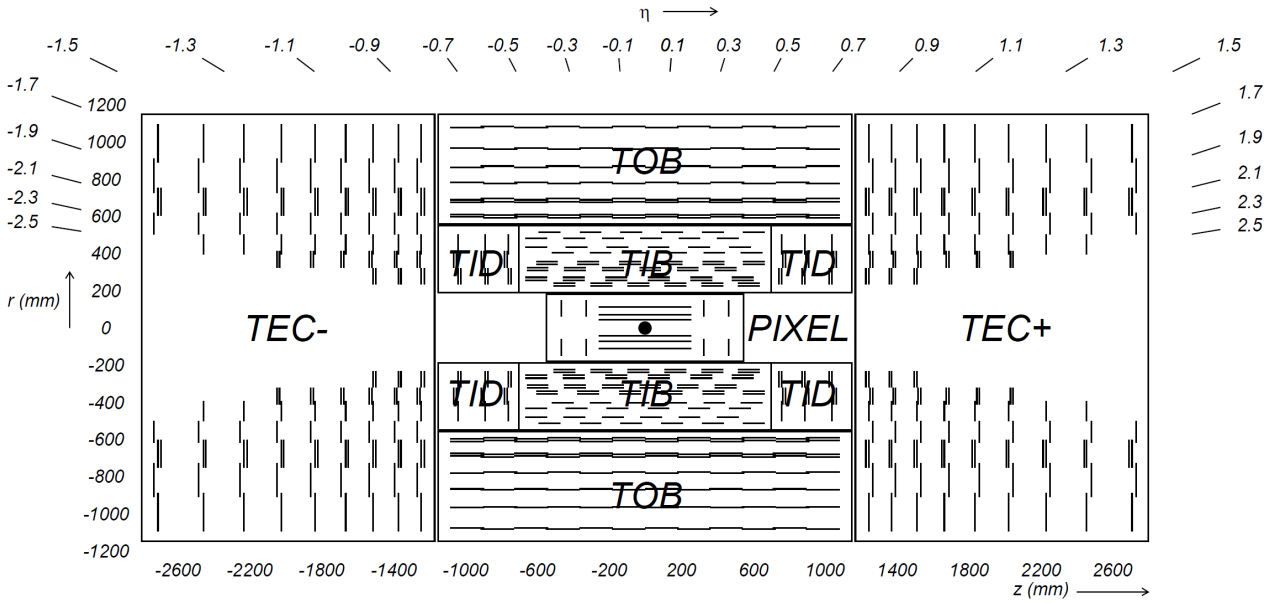


Fig. 3.6.: A schematic illustration of the tracking system encompassing the interaction point within the CMS detector. Key components include the PIXEL subsystem enclosed by the Tracker Inner Barrel (TIB) and Tracker Inner Disks (TID), all enveloped by the Tracker Outer Barrel (TOB) and Tracker EndCaps (TEC±) [34].

Electromagnetic calorimeter

The Electromagnetic Calorimeter (ECAL) [40] is used to measure the energy of electromagnetically interacting particles, such as electrons and photons, with a precision of about ±1%. It is located about 1.3 m from the collision point and surrounds the silicon detector. The ECAL consists of a total of 80.000 $PbWO_4$ crystals, which were selected for their small radiation

length of only 0.89 cm and their small Moliére radius [41]. These properties allows the ECAL to efficiently detect the interactions of electrons and photons. A majority of the crystals are located in the barrel region of the detector, while the crystals in the end caps are divided into two halves and arranged in an array of 5x5 groups called supercrystals. This arrangement allows precise and effective detection of energy over a wide pseudorapidity range of up to $|\eta| < 3$. The crystals of the ECAL have the property of scintillating at a temperature of 18 °C with a light output of about 4.5 γ/MeV. Since this property is temperature dependent, to ensure stability of the ECAL temperature within a narrow range of 0.05 °C, the crystals are carefully cooled with water. During the short period of 25 ns between bunch crossings, the ECAL crystals emit about 85% of the generated light (Bremsstrahlung), which creates electron/positron pairs, resulting in a shower (Fig.3.5) of particles where the electrons release all its energy in the ECAL. This rapid light emission thereby reduces the possibility of cross-contamination between events. The amount of light emitted by the crystals is then detected by photodetectors attached to the back of each crystal and converted into an electrical signal. The energy resolution of the ECAL for electromagnetic showers depends on the energy of the incident particles and can be quantified by appropriate parameterization (Eq.3.1) [16].

$$\left(\frac{\sigma_E}{E}\right)^2 = \underbrace{\left(\frac{2.8\%}{\sqrt{E/\mathrm{GeV}}}\right)^2}_{\text{stochastic effect}} + \underbrace{\left(\frac{12\%}{E/\mathrm{GeV}}\right)^2}_{\substack{\text{electronic,}\\ \text{digitization}\\ \text{and pile-up noise}}} + \underbrace{(0.30\%)^2}_{\substack{\text{leakage of energy}\\ \text{from the crystals}}} \tag{3.1}$$

with E as energy deported in the ECAL [13, 16, 34, 39].

Hadronic calorimeter

The Hadron Calorimeter (HCAL) [42] operates in analogy to the Electromagnetic Calorimeter and plays a central role in the detection and analysis of hadron reactions in the CMS detector. In the HCAL, hadrons enter the brass section of the calorimeter and induce cascades of secondary particles. These particle showers are precisely detected in the plastic scintillator (Fig.3.5).
The measurements in the HCAL span a wide pseudorapidity range. In the central region they extend to $|\eta| < 1.3$, while in the end-cap regions measurements can be made up to $|\eta| < 3$. Considering the components along the beam direction, the measurement capabilities extend even to $|\eta| < 5$. The energy resolution of the combination of ECAL and HCAL was parameterized by measurements as follows (Eq.3.2) [16]:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{100\%}{\sqrt{E/\mathrm{GeV}}}\right)^2 + (4.5\%)^2 \tag{3.2}$$

Superconducting Magnet

The next layer of the CMS consists of a cylindrical superconducting magnet [43] that generates a magnetic field of 3.8 T. This magnetic field has the ability to affect the trajectory of charged particles. This deflection allows the silicon tracker to measure both the charge polarity and momentum of the particles. Furthermore, the iron yokes play a crucial role as they are arranged in three layers and close the magnetic flux. They also act as a filter for the muon chambers, allowing only muons and neutrinos to pass through the iron yokes.

Muon Chambers

The outermost and also largest layer of the CMS detector is formed by the muon chambers [44], which were deliberately placed in the outermost layer of the detector. This results from the fact that muons have a much higher mass compared to electrons and therefore they are less susceptible to the magnetic field generated by the superconducting magnet. Consequently, muons can pass through all the inner subsystems and eventually leave the solenoid magnet. Together with neutrinos, muons are the only particles that can penetrate the heavy iron yokes of the CMS detector.
The muon chambers are equipped in two regions:

- **Barrel region**: Drift chambers (DTs) and resistive plate chambers (RPCs) are implemented in the barrel region, which is in the pseudorapidity range $|\eta| < 1.2$. This choice of detectors is justified due to the comparatively low muon rate in this region. The drift chambers have a length of 2 m and have a rectangular cross section of 14 x 42 mm^2. They are filled with a special gas mixture and exhibit low neutron-induced background. The magnetic field in this region is nearly homogeneous. The drift chambers operate on the principle that charged muons interact with the gas mixture, ionizing some atoms and producing free electrons. These electrons drift along the field lines of the magnetic field to the anode wire in the center of the drift chamber, causing the charge to be measured as current spikes. The resulting spatial resolution is about 100 μm.

- **Endcap region**: Cathode strip chambers (CSCs) in combination with resistive plate chambers (RPCs) are installed in the endcap region, which spans the $0.9 < |\eta| < 2.4$ pseudorapidity range and has the highest muon rate. The CSCs are arranged in trapezoidal chamber geometries and are gaseous detectors with multiple perpendicular wires and cathode strips that can be read out separately. Unlike DTs, CSCs can be operated in inhomogeneous magnetic fields. This feature makes them particularly suitable for handling high muon rates, where a single CSC can provide trigger primitives.

In addition, there is an overlap region between the barrel and endcap regions ($0.9 < |\eta| < 1.2$) that serves as a link between the detector types, allowing seamless and consistent detection of muon trajectories across the entire pseudorapidity range.

Resistive Plate Chambers (RPCs), incorporated in both the barrel and endcap regions, complement detection in both regions. Each RPC consists of two electrode surfaces with a signal readout layer on top of them. They also operate on the principle of gas-based detection, but in avalanche mode. As a result, they have lower spatial resolution but better temporal resolution, which makes them particularly suitable for assigning tracks to bunch crossings.
Together with the tracking system, these muon chambers provide high resolution for measuring the transverse momentum of muons. In the barrel region, for $20 < p_T < 100\,\text{GeV}$, the resolution is higher than 2%, while for muons in the end caps with $p_T < 1\,\text{TeV}$ it is still below 10%. This allows precise measurements of muons in a broad momentum range and reconstruction of the path of the muons based on the measurement positions (Fig.3.5)[13, 16, 39, 45].

3.2.3. Trigger and Data acquisition systems

The CMS experiment relies on handling an enormous amount of data generated by collisions with a bunch crossing of 25 ns and a frequency of 40 MHz. Due to this high data generation rate, it would be impossible to store all generated events unfiltered. Therefore, a sophisticated trigger system is crucial to reduce the amount of data by at least a factor of 10^6 before storing it for further analysis.
The trigger system is divided into two main components: the Level 1 (L1) and the High-Level Trigger (HLT). The L1 trigger system, which is hardware-based, plays a key role in handling the enormous data rate. It uses field programmable gate array (FPGA) and application-specific integrated circuits (ASIC) technology and is capable of throw out events at a maximum rate of 100 kHz, due to the bandwidth limitation of the detector readout process. This hardware component of the trigger system receives information from the detector's calorimeter and muon system. This information includes not only the energy, but also the quality of the measured objects. The L1 Global Trigger processes this information to make an informed decision whether to accept or reject an event. The decision relies on reconstructing basic track segments derived from both the cluster and muon system, or by examining clusters of localized energy deposits within the calorimeters, specifically referred to as trigger primitives. This information is matched with predefined patterns to obtain an initial estimate of the transverse momentum (p_T) in a period of about 1 μs. When the L1 system accepts an event, it is relayed to the HLT at a maximum rate of 100 kHz.
In contrast, the HLT trigger system is software-based and has a computational farm comprising about 26.000 commercial cores. This software component is responsible for unpacking the raw data and then performs a detailed physical object reconstruction. During this process, quality criteria are applied that vary depending on the specific trigger path. When these criteria are met, the event in that path is marked as triggered and stored offline at an average rate of 400 Hz [10, 13, 16].

4. MUSiC

MUSiC (Model Unspecific Search in CMS) is a general model independent approach to the search for new physics, complementary to dedicated analyses for particular models and in specific final states. MUSiC is an unbiased approach without the input of any specific BSM model, relying purely on the predictions of the SM while also taking into account relevant experimental effects [46]. In the CMS experiment, the conventional practice entails the pursuit of new physics by scrutinizing distinct, previously pre-filtered final states resulting from proton-proton collisions, with analyses custom-crafted to align with specific BSM models under investigation. This tailored analytical approach boasts the advantage of heightened sensitivity to the unique characteristics of the particular BSM model(s) in question. However, these dedicated analyses, often carried out at various loci, confront inherent limitations arising from practical constraints, including computational resources and human expertise. Consequently, only a finite number of such analyses can be feasibly conducted, leaving a multitude of unexplored BSM models uncharted. Furthermore, this focused methodology tends to confine investigations to a select subset of final states, those most pertinent to the targeted BSM model(s), potentially overlooking compelling phenomena concealed within alternative final-state configurations. Enter the MUSiC analysis, designed to circumvent these constraints. MUSiC endeavors to categorize the myriad events stemming from hundreds of diverse final states, abstaining from the imposition of model-specific filters. Instead, it systematically assesses each of these final states in search of discrepancies with the predictions outlined in the Standard Model of particle physics. This alternative approach seeks to uncover hitherto concealed phenomena that might elude detection through traditional means. It is acknowledged that this strategy comes at the expense of some sensitivity in comparison to dedicated analyses, which are optimized for specific BSM scenarios. Nonetheless, MUSiC's distinctive merit lies in its ability to pinpoint regions and final states where experimental data deviate from SM expectations. These deviations, in turn, may signify the presence of new physics phenomena, warranting in-depth scrutiny through dedicated analyses to unravel their underlying nature.

4.1. Monte Carlo Event Generation

Monte Carlo event generators are indispensable tools for calculating cross sections of physical processes and for simulating detailed event kinematics. These generators use pseudorandom number generators to produce events according to theoretical predictions. These are used to distinguish measurements of possible new physics from Standard Model processes and to

describe expected outcomes for new physics. The simulation of an event in a hadron collider goes through several successive steps:

- **Hard scattering process:** In this first step, the basis for the event is laid by simulating basic interactions between partons.
- **Parton emission:** The hard scattering process is followed by the emission of partons consisting of quarks and gluons. These emitted particles are also subject to interactions, which refines the event kinematics.
- **Hadronization:** In this phase, the emitted partons are converted into hadronic particles. This process is based on non-perturbative QCD effects and leads to the formation of hadronic jets.
- **Detector simulation:** The final step is to simulate the interaction of the generated particles with the detectors in the experiment. This allows the comparison of the data recorded by the detector with the simulated events to evaluate the experimental data.

Depending on the generator, the cross section is calculated either at leading order (LO) or next to leading order (NLO), taking into account additional vertex contributions that depend on the particular generator configuration. If the theoretical effective cross section for particular samples is known more accurately than the Monte Carlo generator calculations, a correction can be made using a correction factor (k-factor) (Eq.4.1). This factor takes into account higher order corrections for the production cross section.

$$k = \frac{\sigma_{NLO}}{\sigma_{LO}} \quad \text{or} \quad k = \frac{\sigma_{NNLO}}{\sigma_{LO}} \tag{4.1}$$

The same hard scattering process can result in different jet multiplicities in the final state due to effects such as initial state radiation. Some generators, like MadGraph, can simulate additional jets at the matrix element level, while others, like Pythia8, add additional jets and rebalance the event to ensure conservation of energy and momentum [10, 16].

4.2. Datasets and simulated samples

The analysis presented in this thesis uses the datasets acquired by the CMS experiment in the year 2018, derived from proton-proton collisions at a center-of-mass energy of 13 TeV and an integrated luminosity of $59.8\,\mathrm{fb}^{-1}$.

In the MUSiC analysis, a comparative study is undertaken, wherein previously recorded data originating from CMS at the LHC is juxtaposed with Monte Carlo simulations (Sec.4.1) rooted

in the Standard Model. The generation of Monte Carlo events is facilitated through the utilization of event generators such as PYTHIA 8.212 [47], MADGRAPH5 aMC@NLO version 2.6.5 [48–50], POWHEG v2 [51–62] and SHERPA 2.2 [63]. These generators collectively serve to emulate processes pertinent to the Standard Model [46]. The datasets involved in this analysis encompass a comprehensive repository of detector information, which is stored as physical entities distributed across non-local segments of the computing grid. The standardized data format employed herein is the MiniAOD format, chosen for its efficiency in preserving all the requisite information for physics analyses. Nevertheless, it is worth noting that the information contained within the MiniAOD data surpasses the specific requirements of the MUSiC analysis. Consequently, a preliminary processing stage referred to as *skimming* is implemented. During this skimming process, both the MiniAOD files containing measurement data and those housing MC simulations are condensed to encompass solely the data elements germane to MUSiC. Subsequently, this refined data is archived in the *.pxlio* (Physics extension Library) format [64]. This process has been greatly accelerated during the existence of MUSiC and now it takes less than a month to convert all data and MC samples into the new *.pxlio* format.

4.3. Object and event selection

Following the initial pre-processing phase, the MUSiC workflow proceeds with object and event selection. This pivotal stage aims to mitigate adverse effects, such as particle misidentification, by imposing specific data constraints. While these constraints may marginally reduce the overall data efficiency, they are essential for achieving a substantial enhancement in data purity. Chosen events for further analysis must feature, at a minimum, either one lepton (e or μ) or one photon (γ) (table 4.1) because events of this kind can be described well by simulations of the Standard Model. Notably, the analysis requirements are marginally more stringent than the trigger thresholds, ensuring the exclusion of any overlapping events in cases where multiple triggers are activated, thus maintaining an accurate count of unique events.

Trigger used	**Trigger level requirement**	**Analysis requirement**
Single muon	1μ with $p_T > 50\,\text{GeV}$	$\geq 1\mu$ with $p_T > 53\,\text{GeV}$
Single electron trigger	$1e$ with $p_T > 115\,\text{GeV}$	$\geq 1e$ with $p_T > 130\,\text{GeV}$
Di-muon trigger	1st μ with $p_T > 17\,\text{GeV}$ 2nd μ with $p_T > 8\,\text{GeV}$	$\geq 2\mu$, each with $p_T > 25\,\text{GeV}$
Di-electron trigger	$2e$ with $p_T > 33\,\text{GeV}$	$\geq 2e$, each with $p_T > 32\,\text{GeV}$
Single photon trigger	1γ with $p_T > 175\,\text{GeV}$	$\geq 1\gamma$ with $p_T > 225\,\text{GeV}$

Table 4.1.: Event selection criteria

The trigger level requirements describe the requirements that CMS prescribes for the trigger mechanism to be triggered and the analysis requirements clarify the criteria defined by MUSiC for the detection and identification of the corresponding particles. Within the scope of MUSiC, the study encompasses a diverse array of objects, including electrons (e), muons (μ), photons

(γ), jets, b-tagged jets and missing transverse momentum, denoted as p_T^{miss} or MET. It is worth noting that no distinction is made based on the charge of the particles when considering this selection. To optimize the efficiency of object selection while minimizing the misidentification rate for each object, additional stringent criteria have been imposed on individual objects (table 4.2).

Object	p_T**[GeV]**	**Pseudorapidity**
Muon	> 25	$\|\eta\| < 2.1$
Electron	> 25	$0 < \|\eta\| < 1.4442$ or $1.566 < \|\eta\| < 2.5$
Photon	> 25	$\|\eta\| < 1.442$
Jet	> 50	$\|\eta\| < 2.4$
b-tagged jet	> 50	$\|\eta\| < 2.4$
MET	> 100	-

Table 4.2.: Object selection criteria

The stipulated selection criteria pertain to the objects utilized in the overall analytical framework. However, specific criteria pertaining to objects associated with the event trigger are delineated (table 4.1). In cases where these criteria concurrently identify an object as two or more particles, a prescribed procedure is invoked. Initially, the particles are arranged in the following order: μ, e, γ and jets. This particular sequence is chosen based on the assumption that it mirrors the descending order of particle purity. In instances where ambiguity persists within this ordered list, preference is given to the object with the highest priority, resulting in the removal of other particles situated in close proximity, typically defined by $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\Phi)^2} = 0.4$ (or $\Delta R = 0.5$ for jets) relative to the focal object within the event.

4.4. Classification

Upon data selection, the classification process is initiated, wherein events are categorized into distinct event classes. To ensure the luminosity normalisation of the observed dataset based on the cross section of the physic process, Monte Carlo events are subjected to appropriate weightings α (Eq.4.2).

$$\alpha = \alpha_{MC} \cdot \alpha_{PU} = \frac{k \cdot \sigma \cdot \int \mathcal{L}\, dt}{N_{MC}} \cdot \alpha_{Gen} \cdot \alpha_{PU} \tag{4.2}$$

with N_{MC} as the number of simulated events, σ as cross section for the respective process, the k-factor k (Eq.4.1) and the integrated luminosity $\int \mathcal{L}\, dt$. If a specific higher-order generator is used in the production process, it is essential to propagate the event weight α_{Gen} originating directly from the generator into the overall scale factor. In addition, the resulting weights for the Monte Carlo dataset α_{MC} is applied alongside the incorporation of the weight α_{PU}

to address the impact of pile-up[1] conditions. Subsequently, both the observed data and MC events undergo further segregation into various event classes, contingent upon the physical entities involved. The classification schema encompasses a range of event classes as in Fig.4.1.

1. **Exclusive classes**: Exclusive classes are defined by events that precisely match the physics objects specified in the event class label, for instance, $1e$+2μ+1jet. Consequently, each event is unambiguously assigned to a single exclusive class.

2. **Inclusive classes**: Inclusive classes encompass events characterized by a core set of selected objects, which may be supplemented with additional objects. A single event can be associated with multiple inclusive classes, e.g. an event like $1e$+2μ+1jet can belong to the inclusive classes such as 2μ+X, $1e$+X, $1e$+1μ+X, and so forth, with the $+X$ indicating the presence of any number of additional objects.

3. **Jet-inclusive classes**: The third event class is the jet-inclusive class, aimed at enhancing the robustness of the MUSiC analysis against radiation effects without altering the fundamental physical processes. This class incorporates the objects of interest in the final state, along with an arbitrary number of jets (e.g., $1e$+2μ+1jet+Njets), including the possibility of 0 extra jets. Notably, due to inaccuracies in simulating events with six or more jets relative to observed data, events featuring six or more jets are categorized under the *+6 jet+Njets* jet-inclusive event classes.

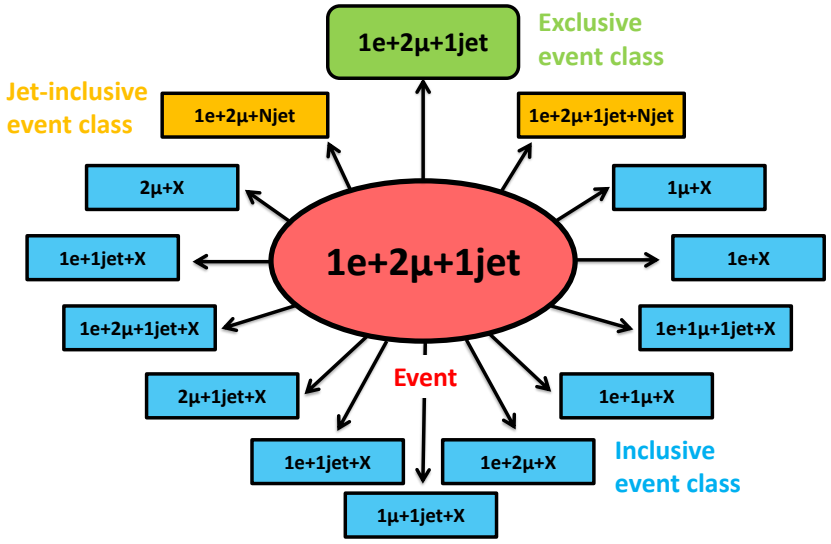


Fig. 4.1.: A MUSiC event classification illustration of a $1e$+2μ+1jet event.

Event classes that exhibit no recorded data and possess an MC event yield, which can be a non integer value, below the threshold of 0.1 are excluded from subsequent analyses. It

[1]In a typical bunch crossing, numerous collisions take place, with only occasional instances leading to a significant interaction intended for triggering due to its potential relevance for subsequent physics analysis. Tracks and energy deposits stemming from extra collisions within the same event (in-time) as well as those originating from preceding or subsequent bunch crossings (out-of-time) are collectively termed pileup.

is worth noting that, in the context of the processes to be elucidated later, the exclusive classes maintain a state of statistical independence and exhibit no inherent correlation with one another, primarily owing to their distinctive compositions. On the other hand, any existing correlations between the inclusive classes have been meticulously considered [46], with the statistical fluctuations of MC simulated events being deemed negligible for the purposes of this analysis.

4.5. The Scan

The analysis proceeds with the critical scan stage. The primary objective of this scan is to conduct a comparative assessment between the observed data and the Standard Model simulation within each event class, with the aim of detecting substantial discrepancies.

4.5.1. Kinematic distributions

Within the scope of the scan, focus is directed towards three key kinematic distributions pertaining to both the CMS event data and the Monte Carlo simulated events. These three distributions have been judiciously selected for their heightened sensitivity to potential Beyond the Standard Model phenomena. Furthermore, the deliberate limitation to just three kinematic distributions serves the dual purpose of ensuring that the analysis remains manageable in complexity and does not overly consume computational resources. The distributions under scrutiny are enumerated as follows:

1. **Sum of transverse momenta (S_T)**: The first kinematic distribution under scrutiny is the sum of transverse momenta, denoted as S_T. For each event class within the analysis, S_T is computed as the scalar sum of the transverse momenta of all N objects explicitly specified in the event class name (Eq.4.3).

$$S_T = \sum_{i}^{N} |\vec{p}_{T,i}| \equiv \sum |\vec{p}_T| \tag{4.3}$$

   Given the prevalence of conjectured heavy particles in the realm of Beyond the Standard Model theories, a closer examination of S_T assumes paramount importance. This stems from the anticipation that, if these heavy particles exist, their presence should manifest in the extreme tails of the S_T distribution.

2. **Combined mass (M)**: The second kinematic distribution under consideration is the

combined mass, typically referred to as the invariant mass (M_{inv}) (Eq.4.4).

$$M_{inv} = \sqrt{\left(\sum_{i}^{N} E_i\right)^2 - \left(\sum_{i}^{N} p_{x,i}\right)^2 - \left(\sum_{i}^{N} p_{y,i}\right)^2 - \left(\sum_{i}^{N} p_{z,i}\right)^2} \tag{4.4}$$

where E_i are the particle energy components and $p_{x,i}$, $p_{y,i}$, $p_{z,i}$ are the momentum components. However, due to the persistent presence of missing transverse momentum, referred to as p_T^{miss} (Eq.4.6), in various event classes, the transverse mass (M_T) is used instead of the invariant mass (Eq.4.5).

$$M_T = \sqrt{\left(\sum_{i}^{N} E_i\right)^2 - \left(\sum_{i}^{N} p_{x,i}\right)^2 - \left(\sum_{i}^{N} p_{y,i}\right)^2} \tag{4.5}$$

This substitution is necessary because the precise measurement of the momentum component in the beam direction is practically unfeasible in these cases, rendering the determination of M_{inv} impractical. A fundamental prerequisite for the applicability of the combined mass in drawing conclusions about massive particles is that the considered event classes must contain a minimum of two objects. This condition ensures the usefulness of the combined mass distribution for characterizing the properties of these particles.

3. **Missing transverse momentum (p_T^{miss} or MET)**: The third kinematic distribution pertains to the missing transverse momentum, denoted as p_T^{miss}. This parameter is defined as the negative vector sum of the transverse momenta of the four-vector (Eq.4.6).

$$p_T^{miss} \equiv | - \sum_{i}^{N} \vec{p}_{T,i}| \tag{4.6}$$

The purpose of p_T^{miss} lies in quantifying the undetected energy within the system, an essential measure ascertained from the transverse momentum not captured by the detectors. Only event classes exhibiting $p_T^{miss} \geq 100\,\text{GeV}$ are included. This threshold has been established considering that lower values of p_T^{miss} primarily arise from anticipated neutrinos in the Standard Model or detector resolution effects.

In the process of establishing kinematic distributions for each event class, a specific selection criterion is applied: only particles explicitly mentioned within the event class are considered to calculate the kinematic variables. For instance, in the case of an event class denoted as 1e+2μ+X, solely the electron and the two muons are taken into account. If a particle is encountered multiple times within a class, such as in a $1e$+X class where a second electron occurs, the particle with the highest transverse momentum (p_T) value is prioritized for inclusion in

the distribution. Another important consideration in constructing these distributions revolves around the choice of bin width. Inadequate bin width selection can give rise to undesirable consequences, impacting both computational efficiency and analysis sensitivity. A larger bin width offers the advantage of reducing the computational time necessary for evaluation, but can lead to a loss of sensitivity in the analysis. Conversely, a smaller bin width is susceptible to being overly influenced by random fluctuations, potentially compromising the analysis.
A judicious choice involves a variable width, guided by the anticipated total detector resolution of all objects in the specific event, with a minimum bin width of 10 GeV serving as the starting point and all subsequent bin widths being integer multiples thereof. To determine an appropriate resolution for each event within the kinematic distributions, several assumptions are made, particularly as the procedure operates at the level of event variables (M_{inv}, $\sum |\vec{p}_T|$ and p_T^{miss}), rather than individual objects within the event. The total detector resolution is contingent upon an event's topology and object content, with estimates available only for individual physics objects based on their transverse momentum or energy. A method is employed to generalize these functions, making variables accessible for each event without requiring knowledge of the individual object's properties. The following paragraphs summarize p_T dependent resolution functions for considered objects, outlining how these functions are translated into bin widths for each event class distribution [1].

For muons, a p_T dependent resolution function (Eq.4.7) is determined from muon simulations.

$$\sigma_\mu(p_T) = \left(0.015 \cdot \frac{p_T^2}{\text{GeV}} + 1.6 \cdot p_T\right) \cdot 10^{-2} \tag{4.7}$$

The transverse energy resolution for electrons and photons is derived from test beam measurements (Eq.4.8) [34].

$$\sigma_{e/\gamma}(E_T) = \sqrt{(0.3 \cdot E_T)^2 + 2.8^2\,\text{GeV} \cdot E_T + (0.12\,\text{GeV})^2} \cdot 10^{-2} \tag{4.8}$$

where E_T denotes the transverse component of the electron or photon energy. A similar approach is taken for the jet transverse momentum resolution, with estimates obtained for PF jets (Eq.4.9) [65].

$$\sigma_{jet}(p_T) = \sqrt{(0.043 \cdot p_T)^2 + 0.85^2\,\text{GeV} \cdot p_T + (3\,\text{GeV})^2} \tag{4.9}$$

The MET resolution has a direct correlation with the transverse momentum of the event and its parameterisation was determined by measuring different events (Eq.4.10) [66].

$$\sigma_{\text{MET}}(p_T) = 1.78\,\text{GeV} + 0.63 \cdot \sqrt{\sum |\vec{p}_T|\,\text{GeV}} \tag{4.10}$$

Subsequently, resolution functions for object transverse momentum or energy are translated into a consistent binning of the three kinematic variables. The width of each bin in the $\sum |\vec{p}_T|$ distribution (Eq.4.11) is determined by assuming the event's total $\sum |\vec{p}_T|$ is evenly distributed over all objects in the event.

$$\sigma_{\sum |\vec{p}_T|} = \sqrt{\sum_{i}^{N} N_i \cdot \sigma_i^2(\langle p_{T,i} \rangle)} \tag{4.11}$$

Using $\langle p_{T,i} \rangle$ to denote the mean transverse momentum assigned to each of the N objects i in an event, where σ_{p_T} represents the resolution of the individual object i, and N_i is the multiplicity of each object type (electron, muon etc.) present in the event.

Assuming objects are evenly distributed in ϕ with $\eta = 0$, the event's $\sum |\vec{p}_T|$ and combined mass are equivalent in this reduced phase-space. The M_T resolution σ_{M_T} is approximated by Eq.4.11.

MET is correlated with $\sum |\vec{p}_T|$, and for genuine MET events, both values are typically of similar magnitude. Simplifying, MET resolution for each bin is estimated by substituting MET for $\sum |\vec{p}_T|$ in Eq.4.11.

Ensuring the overall shape of a distribution is conserved while employing a dynamic bin size is achieved through normalizing the number of events in each bin to the bin width. This approach results in a smooth, continuous representation of a distribution, even when significant changes in bin width occur [1].

4.5.2. p-Value

The p-value represents the probability of observing results as extreme as, or more extreme than, those derived under the assumption of no effect or no difference (null hypothesis). It quantifies the level of agreement between the simulated data and the observed data. To perform this analysis, a hybrid Bayesian-frequentist approach is adopted [67, 68]. To account for statistical fluctuations comprehensively, we use a Poisson distribution (Eq.4.12) to model the probability of obtaining a total of N_{data} observations, considering N_{SM} as the expected number of events from the Standard Model simulations.

$$P(N_{data}) = \frac{e^{-N_{SM}} N_{SM}^{N_{data}}}{N_{data}!} \tag{4.12}$$

This choice allows us to encompass possible extreme scenarios, including both surpluses and deficits concerning the observed data in comparison to the expected values. The Poisson dis-

tributions are subsequently summed (Eq.4.13).

$$p = \begin{cases} \sum_{i=N_{data}}^{\infty} \frac{e^{-N_{SM}} N_{SM}^{i}}{i!} & \text{, if } N_{data} \geq N_{SM} \\ \sum_{i=0}^{N_{data}} \frac{e^{-N_{SM}} N_{SM}^{i}}{i!} & \text{, if } N_{data} < N_{SM} \end{cases} \tag{4.13}$$

The exact mean of the Poisson distribution is not precisely known, therefore we employ a Gaussian distribution centered at N_{data}, with a standard deviation of $\sigma_{SM} = \sqrt{\sigma_{MC,stat}^2 + \sigma_{MC,sys}^2}$ (Eq.4.14).

$$p_{data} = \begin{cases} \sum_{i=N_{data}}^{\infty} C \int_0^{\infty} d\lambda \exp\left(-\frac{(\lambda - N_{SM})^2}{2\sigma_{SM}^2}\right) \frac{e^{-\lambda}\lambda^i}{i!} & \text{, if } N_{data} \geq N_{SM} \\ \sum_{i=0}^{N_{data}} C \int_0^{\infty} d\lambda \exp\left(-\frac{(\lambda - N_{SM})^2}{2\sigma_{SM}^2}\right) \frac{e^{-\lambda}\lambda^i}{i!} & \text{, if } N_{data} < N_{SM} \end{cases} \tag{4.14}$$

Here, C represents the normalisation factor of the Gaussian distribution. All the distributions established in the preceding steps are then partitioned into distinct regions. These regions are defined as contiguous combinations of bins and amount to a total of $N_{bins}(N_{bins}+1)/2$ regions, with N_{bins} representing the number of bins in a given histogram. A p-value is computed for each of these regions (Fig.4.2).

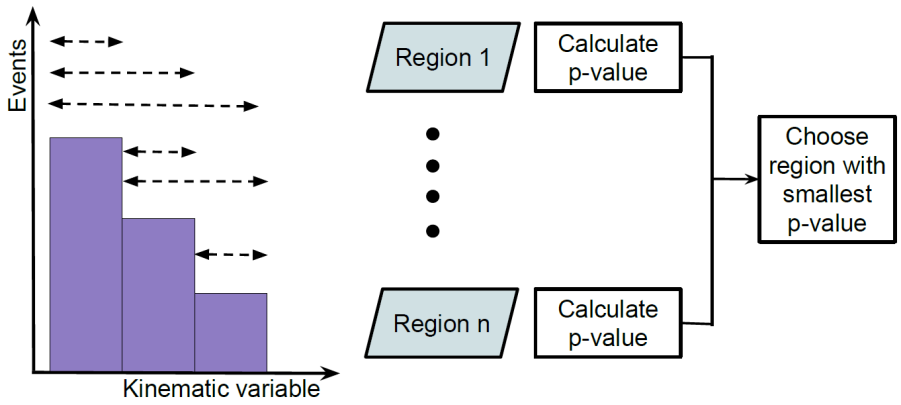


Fig. 4.2.: Schematic depiction outlining the computation of p-values for individual regions and subsequent identification of the region characterized by the minimal p-value.

The region characterized by the most statistically significant departure, as indicated by the smallest p-value (p_{data}), is subsequently identified and designated as the Region of Interest (RoI). The RoI encompasses the domain where the most notable disparities between the measured and simulated data become manifest (Fig.4.3).

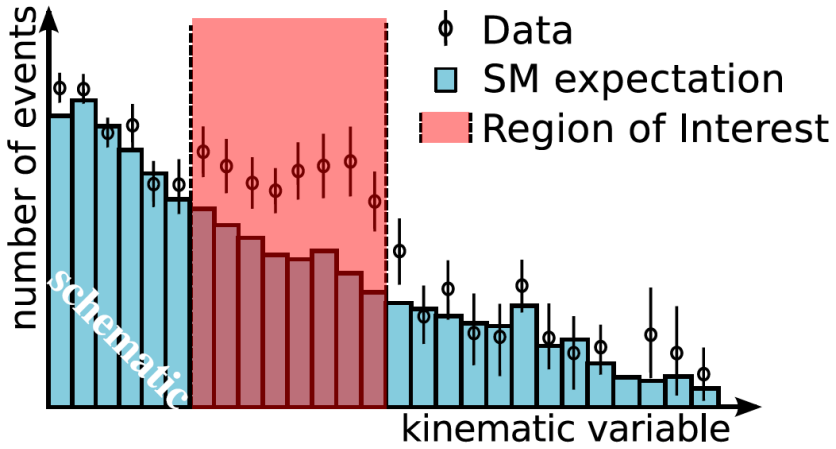


Fig. 4.3.: Visualization depicting the RoI within a kinematic distribution (modified) [5].

To enhance the robustness of our analysis and mitigate potential statistical fluctuations, we have set a minimum threshold for the number of bins within a region. Specifically, three bins are designated as the minimum requirement for S_T and p_T^{miss}, while a single bin suffices for the invariant mass so that narrow resonances can be caught by the algorithm. Furthermore, regions characterized by an insufficient number of simulated events are excluded. This yields a probability distribution for each of the kinetic variables, including S_T, M and p_T^{miss}.

4.5.3. Quality control requirements

The calculation of a p-value necessitates careful consideration of the regions to ensure reliability, computational efficiency and avoidance of ambiguities between multiple regions with the same p-value. Several criteria are applied to veto regions from the region of interest scan [13]:

- **Empty Regions:** Regions without data events, without MC contributions or without contributions from systematic uncertainties are deemed empty. Such regions are excluded from the scan. Additionally, regions where complete contributions are already encompassed within a narrower region are skipped to prioritize the narrowest region with a deviation.

- **Negative Bins:** Monte Carlo simulations at next-to-leading-order precision may yield negative event weights to account for double-counted contributions. Regions with total negative yield indicate overly granular binning and these are vetoed to effectively re-bin the distribution within a larger region with positive total yield.

- **Adaptive Coverage Threshold:** Deviations exceeding 1σ in the (N = number of expected events, σ = corresponding uncertainty) parameter space with small N and total uncertainties surpassing 50% are identified through coverage studies. These regions can result in overly cautious p-value estimations and can bias the calculation of post-trial

probabilities from the pseudo-experiments presented in the next section. A specific functional form (Eq.4.15) is employed to discard such regions, particularly above $N_{SM} > 0.5$, preventing divergence for $N_{SM} \to 0$.

$$\frac{\sigma_{SM}}{N_{SM}} < 1.2 \cdot N_{SM}^{-0.2} \tag{4.15}$$

In addition, a flat veto is applied for relative uncertainties above 50% for $N_{SM} < 5$ to avoid unnecessary restrictions for large N_{SM}, leading to a combined region criterion of (Eq.4.16):

$$\frac{\sigma_{SM}}{N_{SM}} < \max\left(0.5, \min\left(5.0, 1.2 \cdot N_{SM}^{-0.2}\right)\right) \tag{4.16}$$

- **Low Generated Event Count Treatment:** To address limited statistics, regions with a relative statistical uncertainty exceeding 30% are vetoed. This statistical uncertainty requirement is supplemented with a neighbourhood criterion, evaluating the adjacent four bins above and below a region. Relevant processes, constituting 90% of the total yield when ordered by contribution, must be present in the investigated region. The criterion is extended, incorporating an additional check for fluctuations in the neighborhood regions. The relative fraction of a process in the considered region is restricted to fluctuate by no more than ±15% compared to its contribution in the neighborhood regions. This criterion mitigates the impact of spikes in the Monte Carlo simulation, where estimates rely on a small number of events with substantial weights.

4.5.4. Look-Elsewhere Effect and the p̃-value

The p-value, as previously described, allows us to assess localized deviations in the data. Nevertheless, our aim is to perform a comprehensive comparison between the data and simulation on a global scale. This is because an observed deviation may arise as a result of statistical fluctuations. An additional factor known as the *Look-Elsewhere Effect* (LEE) [69] must therefore be taken into account. The Look-Elsewhere Effect takes into account how high the probability is that a deviation will occur that is not a statistical fluctuation. The transition from per-region p-values to a per-distribution p-value, often referred to as a post-trial probability, is essential for comparing deviations observed in various distributions. This transformation accounts for the likelihood of observing deviations of equal or lesser compatibility anywhere within the distribution.
To account for the LEE and thus evaluate the overall distribution of all regions, we introduce a new metric, denoted as the $\tilde{p}$-value (Eq.4.17).

$$\tilde{p} = \frac{N_{pseudo}(p_{min} < p_{min}^{data})}{N_{pseudo}} \tag{4.17}$$

Given the strong correlations among deviations in different regions, an accurate estimation of the $\tilde{p}$-value is paramount. To achieve this, we simulate a series of pseudo-experiments that replicate all the preceding steps. This process involves generating up to 10^5 pseudo-experiments (N_{pseudo}), balancing the need for high sensitivity with computational efficiency. We utilize the smallest p-value (p_{min}^{data}) obtained from the comparison between observed data and Monte Carlo simulations as a reference point. Subsequently, we calculate and count all the locally smallest p-values (p_{min}) from the pseudo-experiments and divide this total by the number of comparisons. This calculation yields the $\tilde{p}$-value, representing the probability of observing a deviation at least as significant as the one observed anywhere in the distribution. In this statistical estimation, there may be cases where no p_{min} values smaller than p_{min}^{data} are found. In such instances, we assign an upper bound for the $\tilde{p}$-value, as the probability of an event, no matter how improbable, can never be exactly zero. In this scenario, for N pseudo-experiments, the upper bound for the $\tilde{p}$-value is $\tilde{p} = \frac{1}{N}$.

4.5.5. Systematic uncertainties

In the generation of pseudo-experiments for an event class distribution, the values assigned to each bin must closely resemble the ensemble of expected values under the null hypothesis, representing a pure simulation scenario. This is achieved by slightly shifting the mean expected values $\langle N_n \rangle$ within each bin using two methods. It's worth noting that this procedure is only applied to bins with both MC expectations and contributions from nuisance parameters. Bins lacking MC expectations or unaffected by uncertainties are exempt from this adjustment. The initial step addresses the representation of systematic uncertainties.

The analysis incorporates estimates of systematic uncertainties from various sources, encompassing integrated luminosity, pileup interactions, total cross sections of Standard Model processes, parton distribution functions (Sec.2.4), energy and momentum scales of objects, reconstruction efficiencies, resolutions, misidentification probabilities and the number of simulated events. Statistical uncertainties due to finite sample size are uncorrelated between bins, while those stemming from other sources are fully correlated across all bins and event classes. Systematic uncertainties, which have a different influence on kinematic properties, are evaluated by variations of these variables, which can lead to certain particles falling out of the selection criteria.

Looking at table 4.3, the integrated luminosity uncertainty is propagated as a normalisation uncertainty. Simulated samples correct for pileup distribution mismatches, with associated uncertainties incorporated into individual event weights. Uncertainties in total cross sections for SM processes are included, assuming full correlation within individual process samples. For processes at LO, a 50% cross-section uncertainty is applied and higher-order calculations involve variations in factorization and renormalisation scales. PDF uncertainties follow the

Source of uncertainty	Typical values
Integrated luminosity	2.5%
Pileup	<5%
Cross sections of SM processes	For processes calculated at LO: 50% For higher-order calculations: varies, <50%
Parton distribution functions	Following PDF4LHC [70] recommendations
Value of α_S	Variations of ± 0.0015 around central value (0.118)
Electron, muon, and photon energy scales	0.15–7.00%
Jet energy scale and resolution	3-5%
Unclustered energy	Varies, typically 0-15 GeV
Reconstruction and identification efficiency	Varies, <10% (relative uncertainty)
Misidentification uncertainties	50%
MC statistical uncertainty	Varies, up to 30%

Table 4.3.: Summary of systematic uncertainties in the analysis [46].

PDF4LHC recommendations [70], treating them as a single, fully correlated source with variations of ± 0.0015 for a central strong coupling value of α_S. Energy and momentum measurement uncertainties for physics objects are estimated by varying kinematic observables within their uncertainties. The muon momentum scale uncertainty is 7% for 1 TeV muons. Electron and photon energy scale uncertainties vary between barrel and endcap regions. Jet energy scale uncertainties, assumed fully correlated, range from 3–5%. Uncertainties in unclustered energy contribute to p_T^{miss} uncertainty. Efficiency related uncertainties, depending on p_T and η, are corrected by scale factors. For b-tagging, scale factors and uncertainties are p_T and flavor-dependent. Misidentification uncertaintie, involve a 50% uncertainty for objects not matched to generated particles. Statistical uncertainties in generated events and total event weights are included due to the limited number of MC events.

All these uncertainties are embodied by a collection of perturbation parameters denoted as i. It is assumed that these nuisance parameters are fully correlated across all bins, reflecting the idea that systematic uncertainties have been segregated to a level where the underlying processes contributing to the uncertainty remain consistent across the entire distribution. The influence of each nuisance parameter is modeled using a Gaussian distribution centered on the mean expectation value for each bin n. To account for the aforementioned correlation effect, a random number κ_i is generated for each perturbation parameter i, following a standard normal distribution. At this juncture, the first adjustment is applied (Eq.4.18):

$$\langle N_{n,shifted} \rangle = \langle N_n \rangle + \sum_{i=1}^{N_i} \kappa_i \cdot \delta_{i,n} \tag{4.18}$$

where $\langle N_{n,shifted} \rangle$ represents the shifted mean value of each bin, N_i signifies the total number of perturbation parameters considered, and $\delta_{i,n}$ denotes the difference between the mean and the upper or lower bound of the confidence interval. In the next phase, $\langle N_{n,shifted} \rangle$ undergoes

further modification. This step is particularly relevant because statistical effects may slightly shift the number of events within a bin. To account for this influence, a Poisson distribution is employed, using $\langle N_{n,shifted} \rangle$ as the mean parameter for smearing. The resultant value represents the final data point utilized in the pseudo-experiment.

4.6. Global overview

To summarize the outcomes of various event classes and kinematic distributions, we present them in the form of histograms. These histograms allow us to compare the distributions of deviations in the event classes with the expected distribution under the Standard Model only hypotheses, as derived from pseudo-experiments. This display format offers the advantage of accommodating different scenarios. It enables the observation of situations where a Beyond Standard Model (BSM) signal exhibits substantial deviations in a few distinct final states, as well as scenarios where a BSM signal manifests as smaller deviations across multiple final states. The $\tilde{p}$ distributions are first computed for all pseudo-experiments, as explained in the preceding analysis steps (Sec.4.5.2 & 4.5.4).

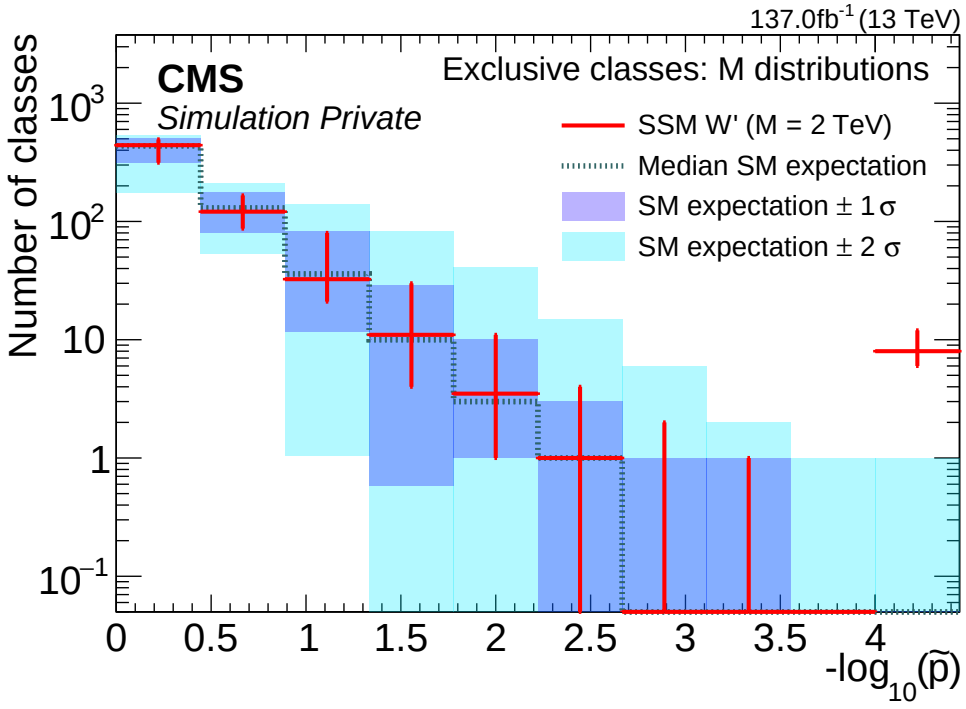


Fig. 4.4.: Exemplary representation of a $\tilde{p}$-value distribution as part of a typical sensitivity study to investigate MUSiC's ability to observe an introduced signal from a hypothetical W′ event on top of MC data. The study explores different classes (final states) within a region of interest scan, as part of a MUSiC analysis [3].

An illustrative example of this analysis is presented in Figure 4.4. In this figure, the number of event classes is plotted against their corresponding $\tilde{p}$-values. For clarity, the $\tilde{p}$-values are

represented in $-\log_{10}(\tilde{p})$ units. The construction of the histogram involves using the median of each pseudo-experiment as the SM reference. To capture the uncertainty in the SM-only hypothesis distribution, one and two sigma deviations from the median are depicted. These deviations are represented by the dark and light blue colored areas surrounding the median and correspond to the regions encompassing 68% ($\pm 1\sigma$) and 95% ($\pm 2\sigma$) of the pseudo-experiments, respectively. Due to the nature of the generation process, the size of the areas above and below the median may vary, as different pseudo-experiments may have a different number of events. The histogram further includes the distribution obtained when scanning for the signal under investigation. In the case of an introduced BSM signal on top of MC data, this distribution is shown as red crosses, which are added on the histograms representing the SM-only hypothesis. In the case of real data, this is not represented by a red cross, but by black dots. A red cross (or black dot) positioned outside of the SM expectation, as seen on the right side of Figure 4.4, is not an overflow bin. Instead it indicates a deviation from the SM that cannot be attributed to mere statistical fluctuations. Such a deviation serves as an indication of a BSM signal. The systematic uncertainties (Sec.4.5.4) may still exert influence on the $\tilde{p}$ distributions employed in the histogram. This influence is observed in the form of a higher number of event classes displaying smaller deviations and appearing in the bins with the most modest discrepancies. To address this, we have extended the uncertainty bands employed in the histogram after conducting pseudo-data studies. This extension allows for an overestimation of uncertainty, encompassing up to 50%. It is important to note that this modification has no bearing on the range where substantial deviations are expected, serving as an indication of potential Beyond Standard Model effects.

5. First Look at the 2018 Dataset

5.1. Region of Interest scan for the 2018 Dataset

This chapter focuses on the core aspects of this work. The main objective was to gain an initial insight into the MUSiC analysis (Ch.4) of the Ultra Legacy 2018 dataset (Run II) of the CMS experiment. It is of interest to find out what progress the current status of the analysis reveals. The results obtained can be summarized as follows:

Overview of the 2018 Dataset

As part of the MUSiC scan run, all three distributions (Invariant Mass, Sum of Transverse Momentum, Missing Transverse Momentum) were analyzed with respect to the three event class groups defined by MUSiC (Exclusive, Inclusive and Jet-Inclusive).

Within the scope of the MUSiC analysis for the Ultra Legacy 2018 Dataset, a total of 225 (274) exclusive classes, 624 (610) inclusive classes, and 407 (100) jet-inclusive classes based on the M_{inv} ($\sum |\vec{p}_T|$) distribution scan were observed, each with at least one data event [1]. Furthermore, within these classes, considering a minimum of 100 GeV in p_T^{miss}, 68 exclusive classes, 339 inclusive classes, and 85 jet-inclusive classes were identified. Subsequently, these classes were also investigated for deviations in the MET distribution.

Two notable event classes, illustrating the efficiency of the MUSiC analysis, are the 2μ+X (Fig.5.1) and the 1μ+2bJet+X (Fig.5.2) event classes from the M_{inv} inclusive scans. In the 2μ+X class (Fig.5.1), a clearly recognisable Z-peak is observed as expected, while in the 1μ+2bJet+X event class (Fig.5.2), there is a clear dominance of $t\bar{t}$ events.

[1] The presented numbers correspond only to classes in which at least one region survived the scan quality control requirements, as described in Subsec.4.5.3

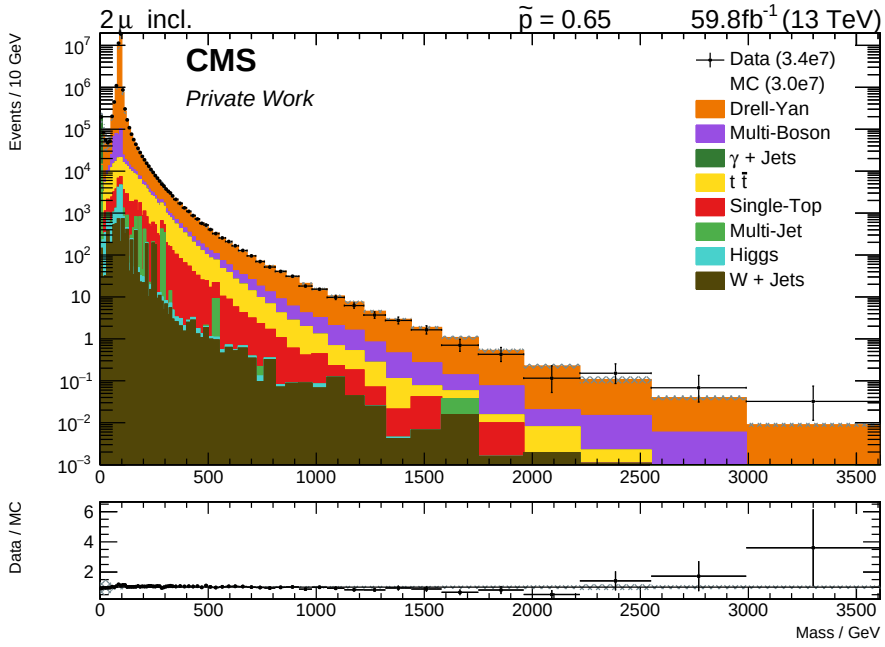


Fig. 5.1.: Distribution of the 2μ+X M_{inv} inclusive event class.

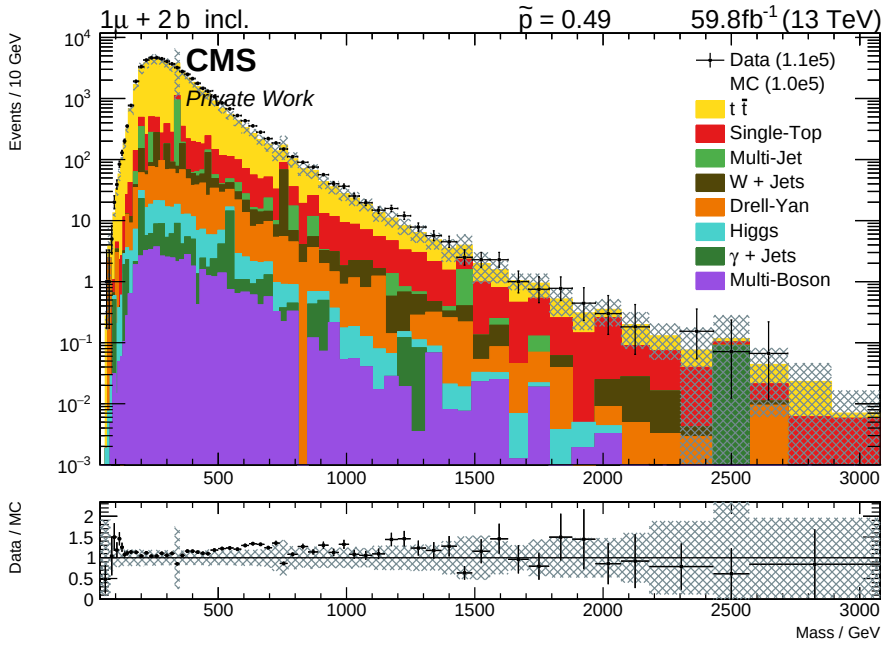


Fig. 5.2.: Distribution of the 1μ+2bJet+X M_{inv} inclusive event class.

The most significant event classes of the 2018 Dataset

The results for the classes with the most significant deviation from the SM prediction within the scan are shown in the tables 5.1-5.9 and the figures 5.3-5.11. The total number of entries is 90, whereby 42 unique classes result when recurring entries and their subtraction are taken into account.

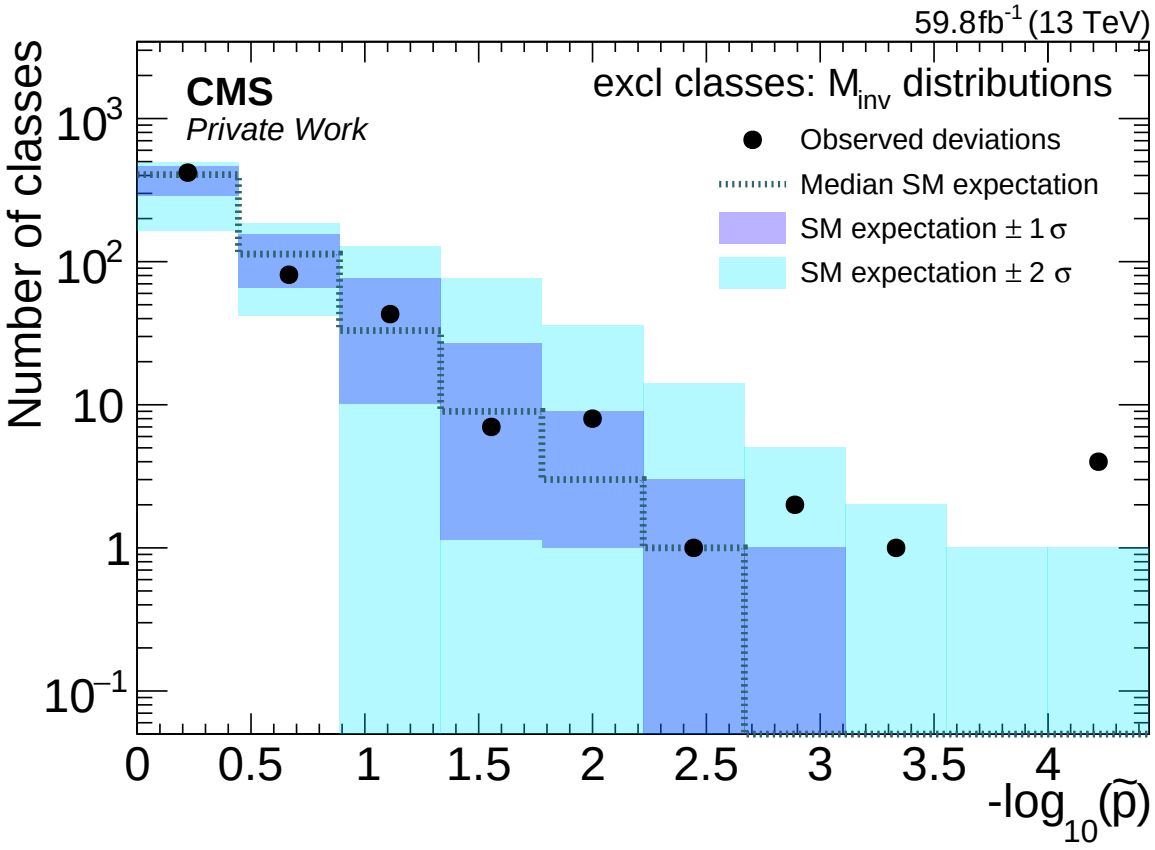


Fig. 5.3.: Distribution of the $\tilde{p}$ values pertaining to the kinematic variable M_{inv} across exclusive event classes based on a RoI scan for the MUSiC analysis. The rightmost black dot belongs to the first 4 event classes in table 5.1.

Event Class	RoI [GeV]	N_{MC}	σ_{MC}	N_{Data}	p	$\tilde{p}$
$2e$+4Jet	750 - 840	210	33	448	10^{-8}	$< 10^{-5}$
$2e$+1Jet+MET	4950 - 5390	0.00019	0.00018	2	$4{\cdot}10^{-8}$	$< 10^{-5}$
2μ+MET	0 - 30	19	12	96	$2.7{\cdot}10^{-7}$	$< 10^{-5}$
4μ	180 - 320	86	4.5	145	$1.2{\cdot}10^{-7}$	0.0001
$1e$+2bJet	740 - 830	44	14	113	$2.9{\cdot}10^{-5}$	0.0005
$1e$+MET	0 - 30	720	150	1360	$2.4{\cdot}10^{-5}$	0.0011
$2e+2\mu$	260 - 300	8.9	0.57	25	$1.2{\cdot}10^{-5}$	0.0016
$1e+2\mu+1\gamma$+1bJet	390 - 1010	0.072	0.035	2	0.0031	0.0051
$1e+1\mu+1\gamma$+2bJet	730 - 970	0.33	0.18	4	0.001	0.011
4μ+3Jet	890 - 1040	0.18	0.095	3	0.0015	0.011

Table 5.1.: Comprehensive examination of the most significant event classes within the Region of Interest (RoI) scan of the exclusive class of the kinematic distribution M_{inv}. The parameter p denotes the median p-value derived from multiple pseudoexperiments, while $\tilde{p}$ signifies the p-value adjusted for the Look-Elsewhere Effect (LEE). The RoI delineates the boundaries of the region of interest, detailed in Section 4.5.4. N_{Data} represents the count of pseudo-data events, N_{MC} corresponds to the anticipated number of events conforming to the Standard Model from Monte Carlo simulations and σ_{MC} as a combination of statistical and systematic uncertainty.

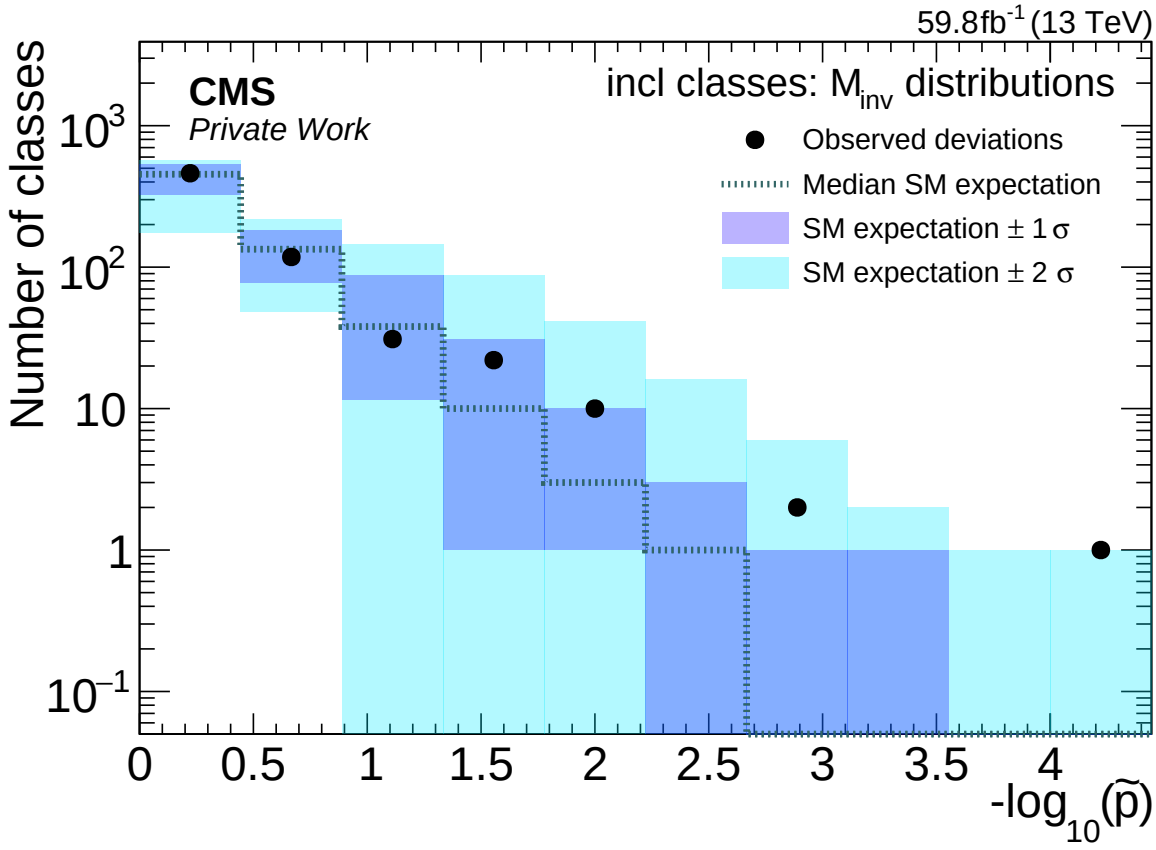


Fig. 5.4.: Distribution of the $\tilde{p}$ values pertaining to the kinematic variable M_{inv} across inclusive event classes based on a RoI scan for the MUSiC analysis. The rightmost black dot belongs to the first event class in table 5.2.

Event Class	**RoI [GeV]**	$\mathbf{N_{MC}}$	$\boldsymbol{\sigma_{MC}}$	$\mathbf{N_{Data}}$	**p**	$\tilde{\mathbf{p}}$
4μ+X	180 - 270	85	4.9	148	$5{\cdot}10^{-8}$	$< 10^{-5}$
$1e$+1γ+X	20 - 50	1900	200	1009	$5.7{\cdot}10^{-6}$	0.0008
$2e$+4Jet+X	750 - 870	430	78	789	$6{\cdot}10^{-6}$	0.001
$1e$+X	70 - 110	50	16	112	0.00066	0.0071
$1e$+3μ+2Jet+X	1380 - 1500	0.045	0.03	2	0.0015	0.0085
1μ+1γ+X	20 - 70	$4.5{\cdot}10^{4}$	$5.9{\cdot}10^{3}$	23577	0.00013	0.009
3μ+5Jet+MET+X	930 - 1050	0.37	0.3	4	0.0028	0.0097
4μ+4Jet+X	2230 - 2410	0.0075	0.007	1	0.0093	0.014
$2e$+1μ+1γ+MET+X	280 - 1260	1.1	0.94	7	0.0041	0.014
$2e$+2μ+X	260 - 310	15	1	32	0.00013	0.015

Table 5.2.: Comprehensive examination of the most significant event classes within the Region of Interest (RoI) scan of the inclusive class of the kinematic distribution M_{inv}. The parameter p denotes the median p-value derived from multiple pseudoexperiments, while $\tilde{p}$ signifies the p-value adjusted for the Look-Elsewhere Effect (LEE). The RoI delineates the boundaries of the region of interest, detailed in Section 4.5.4. N_{Data} represents the count of pseudo-data events, N_{MC} corresponds to the anticipated number of events conforming to the Standard Model from Monte Carlo simulations and σ_{MC} as a combination of statistical and systematic uncertainty.

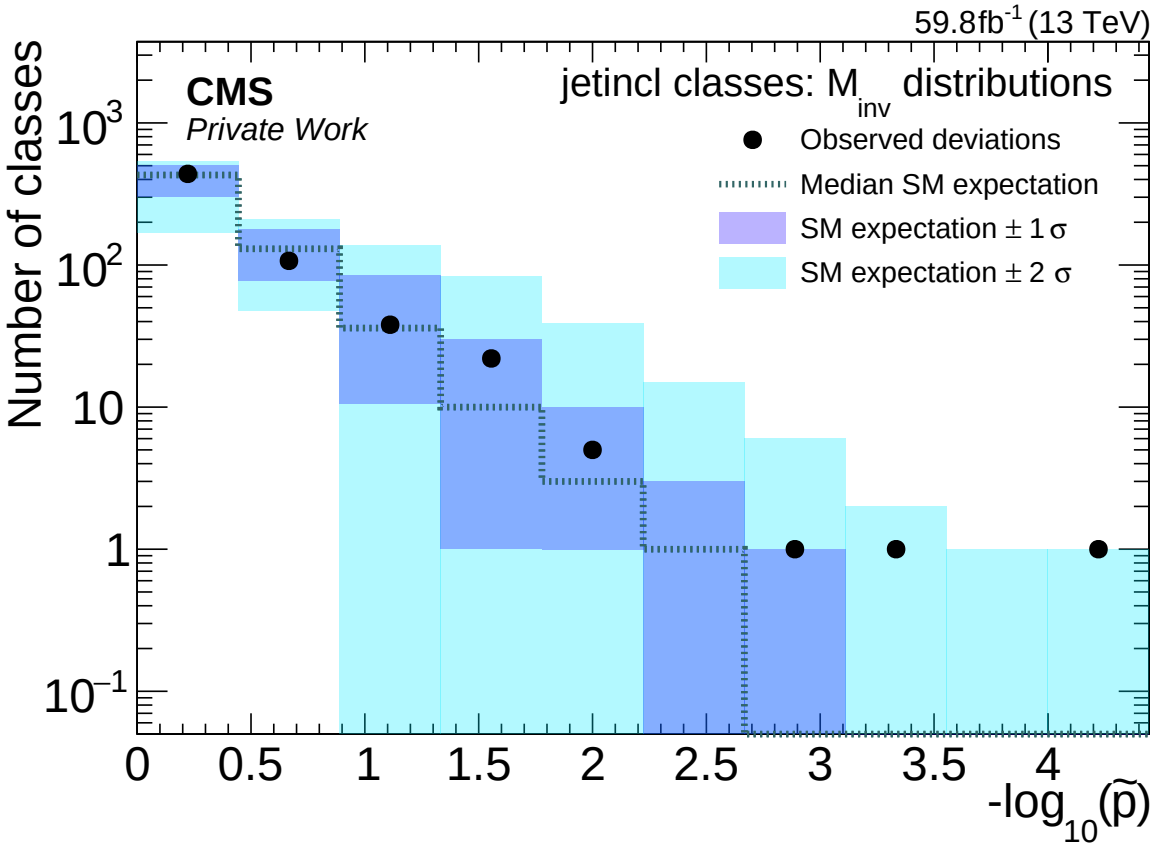


Fig. 5.5.: Distribution of the $\tilde{p}$ values pertaining to the kinematic variable M_{inv} across jet-inclusive event classes based on a RoI scan for the MUSiC analysis. The rightmost black dot belongs to the first event class in table 5.3.

Event Class	RoI [GeV]	N_{MC}	σ_{MC}	N_{Data}	p	$\tilde{p}$
4μ+NJets	170 - 670	140	7.9	223	$1.2{\cdot}10^{-8}$	0.0001
$2e$+4Jet+NJets	750 - 840	250	53	514	$2.7{\cdot}10^{-6}$	0.0005
2μ+3bJet+1Jet+NJets	810 - 900	0.86	0.5	8	$7.9{\cdot}10^{-5}$	0.0017
4μ+3Jet+NJets	830 - 980	0.18	0.1	3	0.0016	0.0077
4μ+4Jet+NJets	2180 - 2350	0.0034	0.0032	1	0.0043	0.0098
$2e$+1μ+1γ+MET+NJets	280 - 1260	0.85	0.78	6	0.0048	0.012
$1e$+2μ+1γ+1bJet+NJets	500 - 980	0.074	0.042	2	0.0035	0.013
$1e$+1μ+1γ+2bJet+NJets	730 - 970	0.52	0.25	5	0.00064	0.014
$2e$+5Jet+NJets	2140 - 2310	13	4.7	40	0.00012	0.018
$2e$+2μ+NJets	260 - 310	14	0.96	31	0.00016	0.018

Table 5.3.: Comprehensive examination of the most significant event classes within the Region of Interest (RoI) scan of the jet-inclusive class of the kinematic distribution M_{inv}. The parameter p denotes the median p-value derived from multiple pseudoexperiments, while $\tilde{p}$ signifies the p-value adjusted for the Look-Elsewhere Effect (LEE). The RoI delineates the boundaries of the region of interest, detailed in Section 4.5.4. N_{Data} represents the count of pseudo-data events, N_{MC} corresponds to the anticipated number of events conforming to the Standard Model from Monte Carlo simulations and σ_{MC} as a combination of statistical and systematic uncertainty.

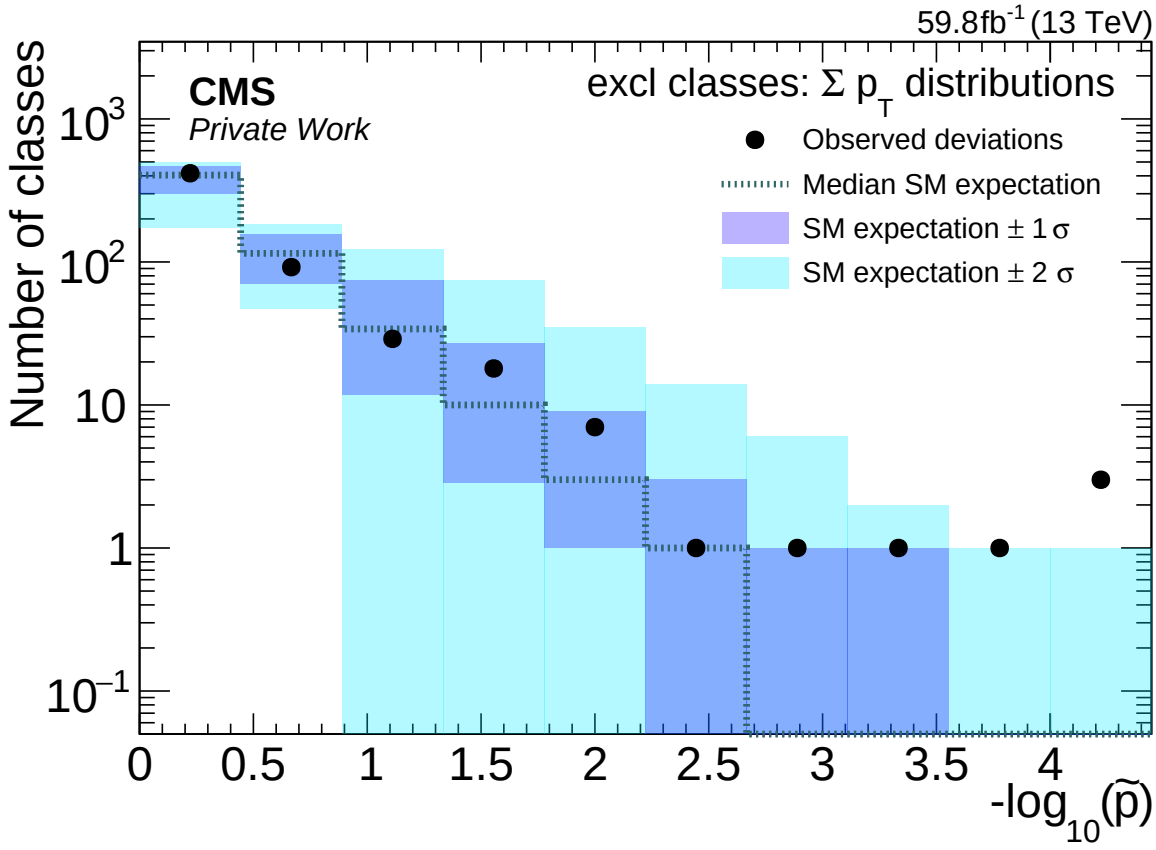


Fig. 5.6.: Distribution of the $\tilde{p}$ values pertaining to the kinematic variable S_T across exclusive event classes based on a RoI scan for the MUSiC analysis. The rightmost black dot belongs to the first 3 event classes in table 5.4.

Event Class	**RoI [GeV]**	$\mathbf{N_{MC}}$	$\boldsymbol{\sigma_{MC}}$	$\mathbf{N_{Data}}$	**p**	$\tilde{\mathbf{p}}$
4μ	140 - 170	21	1.3	55	10^{-8}	$< 10^{-5}$
$2e$+1Jet+MET	4950 - 5280	0.00018	0.00016	2	$3.2{\cdot}10^{-8}$	$< 10^{-5}$
$2e+2\mu$	190 - 420	37	2.4	72	$1.7{\cdot}10^{-6}$	0.0001
1μ	310 - 2060	300	41	501	$5.9{\cdot}10^{-6}$	0.0002
$2e$+1Jet	610 - 690	1700	200	2636	$7.8{\cdot}10^{-6}$	0.0006
4μ+3Jet	550 - 610	0.12	0.048	3	0.00037	0.0011
$2e$+4Jet	900 - 990	150	26	270	$2.6{\cdot}10^{-5}$	0.0027
2μ+1Jet	660 - 750	2300	270	3351	$4.7{\cdot}10^{-5}$	0.0072
4μ+4Jet	1580 - 2030	0.0016	0.0016	1	0.0021	0.0077
$2e$+5Jet	1000 - 1150	43	14	107	$8{\cdot}10^{-5}$	0.0085

Table 5.4.: Comprehensive examination of the most significant event classes within the Region of Interest (RoI) scan of the exclusive class of the kinematic distribution S_T. The parameter p denotes the median p-value derived from multiple pseudoexperiments, while $\tilde{\text{p}}$ signifies the p-value adjusted for the Look-Elsewhere Effect (LEE). The RoI delineates the boundaries of the region of interest, detailed in Section 4.5.4. N_{Data} represents the count of pseudo-data events, N_{MC} corresponds to the anticipated number of events conforming to the Standard Model from Monte Carlo simulations and σ_{MC} as a combination of statistical and systematic uncertainty.

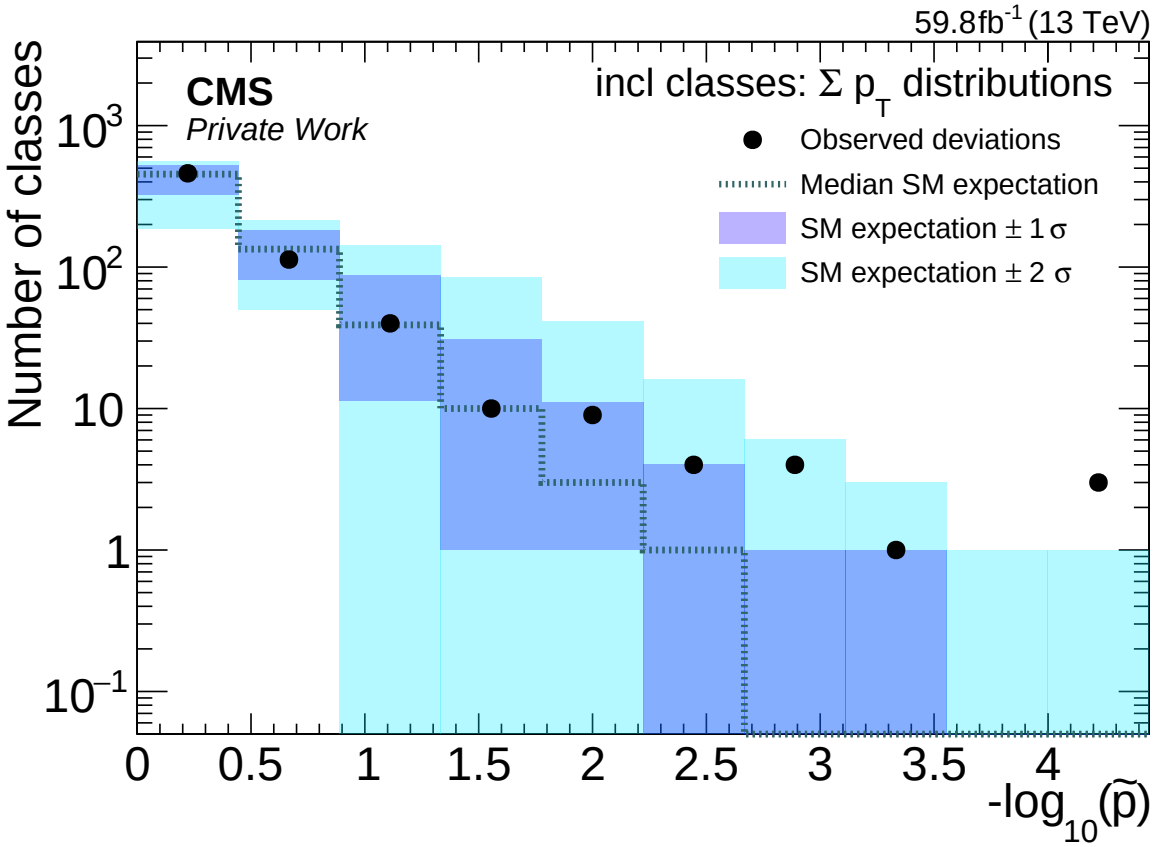


Fig. 5.7.: Distribution of the $\tilde{p}$ values pertaining to the kinematic variable S_T across inclusive event classes based on a RoI scan for the MUSiC analysis. The rightmost black dot belongs to the first 3 event classes in table 5.5.

Event Class	**RoI [GeV]**	$\mathbf{N_{MC}}$	$\boldsymbol{\sigma_{MC}}$	$\mathbf{N_{Data}}$	**p**	$\tilde{\mathbf{p}}$
4μ+X	140 - 170	26	1.8	60	$1.1{\cdot}10^{-7}$	$< 10^{-5}$
$1e$+1μ+3bJet+2Jet+X	760 - 850	0.41	0.4	7	$3.1{\cdot}10^{-5}$	$< 10^{-5}$
4μ+4Jet+X	1460 - 2030	0.0073	0.0063	2	$5.1{\cdot}10^{-5}$	$< 10^{-5}$
2μ+X	300 - 330	$1.2{\cdot}10^{4}$	830	15724	$4.4{\cdot}10^{-6}$	0.0004
$2e$+2μ+X	190 - 280	41	2.8	76	$5.7{\cdot}10^{-6}$	0.0008
$2e$+1Jet+X	570 - 660	6200	660	8954	$1.7{\cdot}10^{-5}$	0.0008
$1e$+1γ+X	80 - 120	4100	410	2331	8.3·10−6	0.0011
$2e$+4Jet+X	900 - 990	250	38	429	$1.4{\cdot}10^{-5}$	0.0018
$2e$+6Jet+X	1330 - 1910	18	8.2	61	$2.9{\cdot}10^{-5}$	0.0028
$2e$+3Jet+X	760 - 850	1200	160	1797	0.00011	0.0051

Table 5.5.: Comprehensive examination of the most significant event classes within the Region of Interest (RoI) scan of the inclusive class of the kinematic distribution S_T. The parameter p denotes the median p-value derived from multiple pseudoexperiments, while $\tilde{p}$ signifies the p-value adjusted for the Look-Elsewhere Effect (LEE). The RoI delineates the boundaries of the region of interest, detailed in Section 4.5.4. N_{Data} represents the count of pseudo-data events, N_{MC} corresponds to the anticipated number of events conforming to the Standard Model from Monte Carlo simulations and σ_{MC} as a combination of statistical and systematic uncertainty.

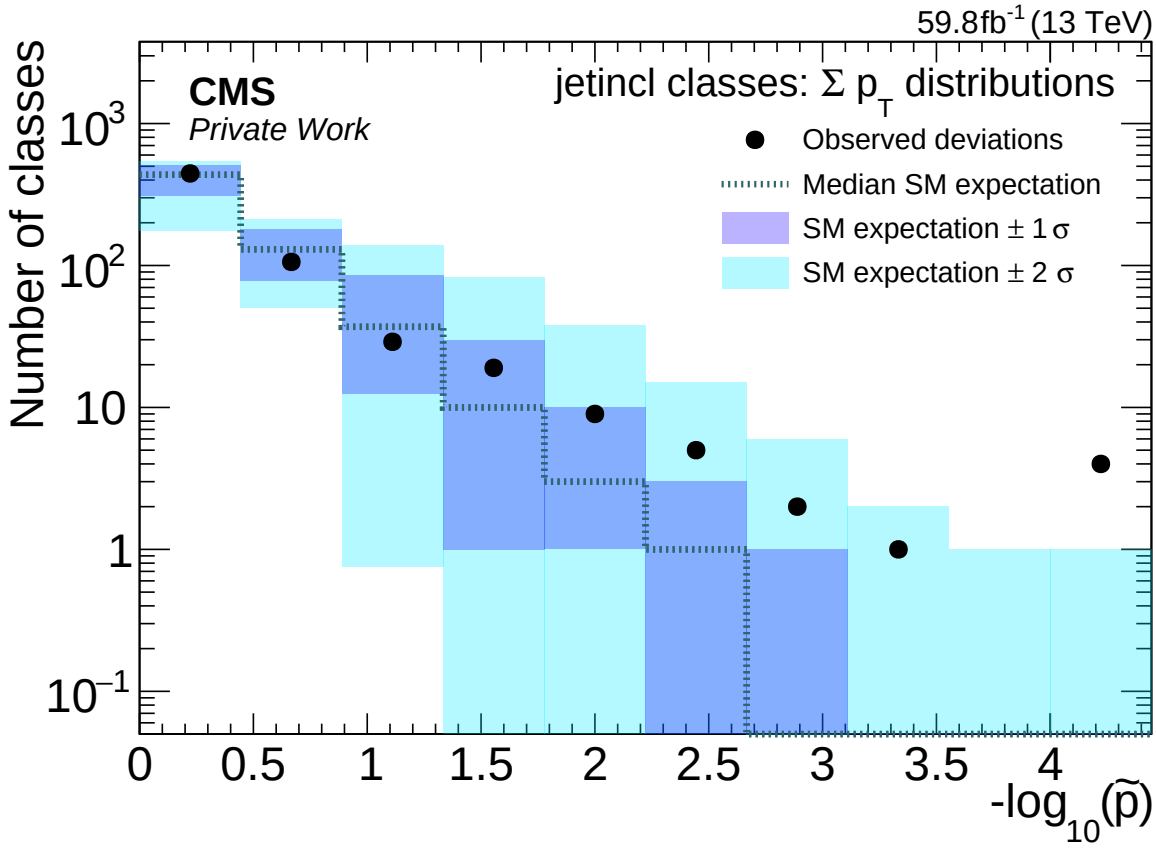


Fig. 5.8.: Distribution of the $\tilde{p}$ values pertaining to the kinematic variable S_T across jet-inclusive event classes based on a RoI scan for the MUSiC analysis. The rightmost black dot belongs to the first 4 event classes in table 5.6.

Event Class	**RoI [GeV]**	$\mathbf{N_{MC}}$	$\boldsymbol{\sigma_{MC}}$	$\mathbf{N_{Data}}$	**p**	$\tilde{\mathbf{p}}$
4μ+NJets	140 - 170	26	1.8	60	$7.5{\cdot}10^{-8}$	$< 10^{-5}$
$2e$+4Jet+NJets	900 - 990	190	31	366	$3{\cdot}10^{-7}$	0.0001
2μ+NJets	320 - 360	9800	680	12902	$5.2{\cdot}10^{-6}$	0.0001
4μ+4Jet+NJets	1460 - 2030	0.0036	0.0032	2	$1.3{\cdot}10^{-5}$	0.0001
$2e$+2μ+NJets	190 - 280	40	2.7	73	$1.2{\cdot}10^{-5}$	0.0007
$2e$+1Jet+NJets	570 - 660	5400	610	7926	$2.4{\cdot}10^{-5}$	0.0009
$2e$+6Jet+NJets	1330 - 1910	11	6.8	47	$3.4{\cdot}10^{-5}$	0.0014
$2e$+5Jet+NJets	1350 - 1690	25	10	81	$9.7{\cdot}10^{-6}$	0.0026
$1e$+1mu+3bJet+2Jet+NJets	760 - 850	0.19	0.18	4	0.00041	0.0026
$1e$+NJets	40 - 70	47	9.9	0	$8.8{\cdot}10^{-7}$	0.0029

Table 5.6.: Comprehensive examination of the most significant event classes within the Region of Interest (RoI) scan of the jet-inclusive class of the kinematic distribution S_T. The parameter p denotes the median p-value derived from multiple pseudoexperiments, while $\tilde{\text{p}}$ signifies the p-value adjusted for the Look-Elsewhere Effect (LEE). The RoI delineates the boundaries of the region of interest, detailed in Section 4.5.4. N_{Data} represents the count of pseudo-data events, N_{MC} corresponds to the anticipated number of events conforming to the Standard Model from Monte Carlo simulations and σ_{MC} as a combination of statistical and systematic uncertainty.

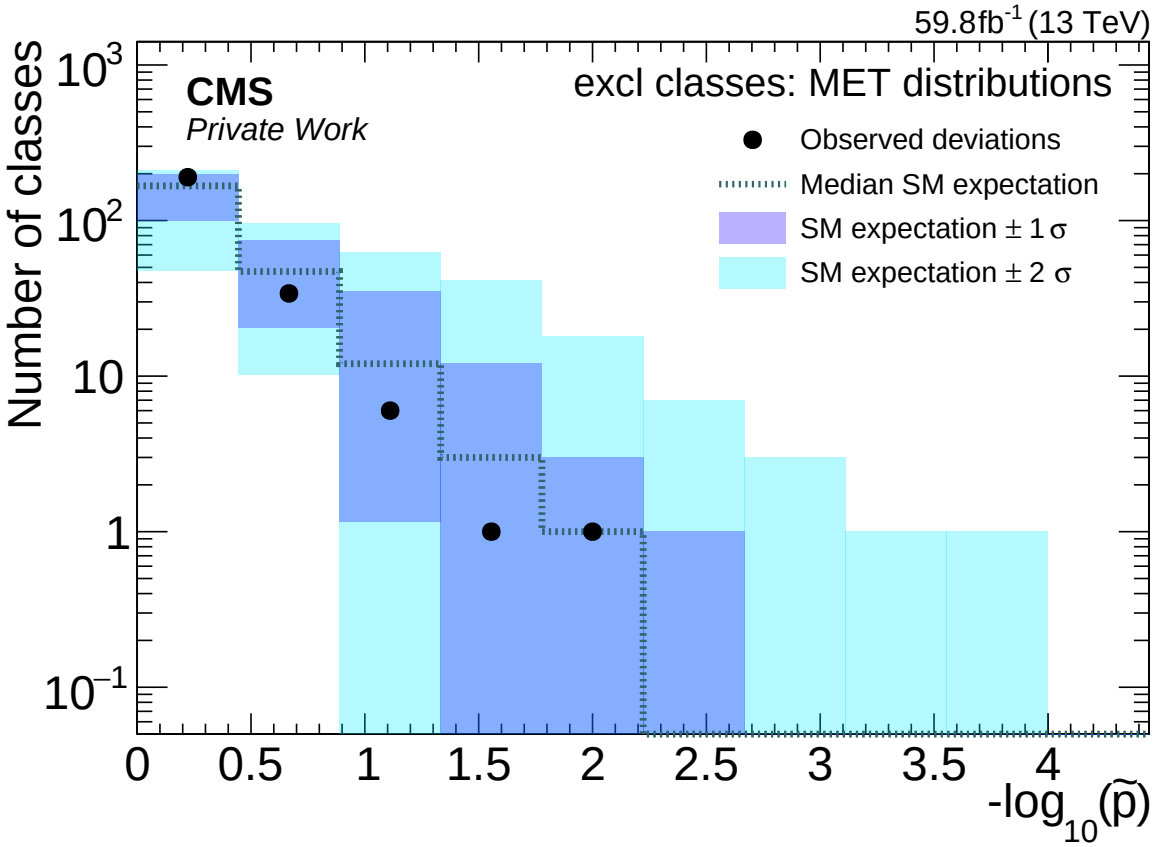


Fig. 5.9.: Distribution of the $\tilde{p}$ values pertaining to the kinematic variable p_T^{miss} across exclusive event classes based on a RoI scan for the MUSiC analysis. The rightmost black dot belongs to the first event class in table 5.7.

Event Class	RoI [GeV]	N_{MC}	σ_{MC}	N_{Data}	p	$\tilde{p}$
$2e$+1μ+1γ+MET	250 - 310	0.0089	0.0084	1	0.011	0.015
3μ+5Jet+MET	100 - 1250	1.7	1.4	7	0.023	0.032
$2e$+3bJet+1Jet+MET	150 - 190	0.19	0.17	2	0.028	0.066
$1e$+3Jet+MET	270 - 330	1100	140	1502	0.0032	0.067
3μ+1bJet+2Jet+MET	110 - 160	2.6	1.5	9	0.016	0.08
$2e$+2bJet+1Jet+MET	90 - 120	39	12	73	0.0086	0.099
1μ+3Jet+MET	330 - 390	1900	240	2552	0.0062	0.11
$1e$+1μ+1γ+1Jet+MET	170 - 250	8.3	3.2	0	0.011	0.12
$1e$+3bJet+2Jet+MET	160 - 250	1.9	1.9	7	0.05	0.14
$2e$+1μ+1γ+1Jet+MET	100 - 530	0.4	0.35	2	0.098	0.14

Table 5.7.: Comprehensive examination of the most significant event classes within the Region of Interest (RoI) scan of the exclusive class of the kinematic distribution p_T^{miss}. The parameter p denotes the median p-value derived from multiple pseudoexperiments, while $\tilde{p}$ signifies the p-value adjusted for the Look-Elsewhere Effect (LEE). The RoI delineates the boundaries of the region of interest, detailed in Section 4.5.4. N_{Data} represents the count of pseudo-data events, N_{MC} corresponds to the anticipated number of events conforming to the Standard Model from Monte Carlo simulations and σ_{MC} as a combination of statistical and systematic uncertainty.

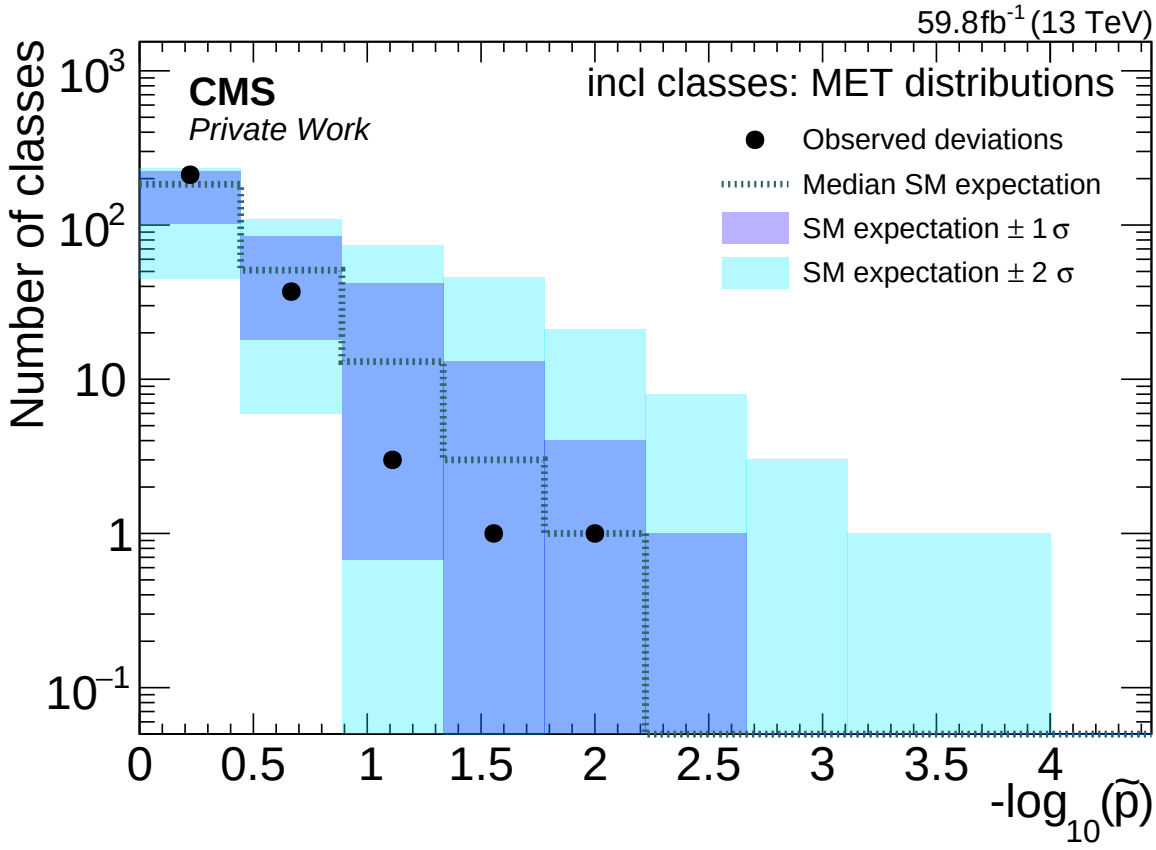


Fig. 5.10.: Distribution of the $\tilde{p}$ values pertaining to the kinematic variable p_T^{miss} across inclusive event classes based on a RoI scan for the MUSiC analysis. The rightmost black dot belongs to the first event class in table 5.8.

Event Class	**RoI [GeV]**	$\mathbf{N_{MC}}$	$\boldsymbol{\sigma_{MC}}$	$\mathbf{N_{Data}}$	**p**	$\tilde{\mathbf{p}}$
$2e$+1μ+1γ+MET+X	100 - 530	1.3	1.1	7	0.0073	0.016
$2e$+1μ+1γ+1Jet+MET+X	100 - 530	0.82	0.58	5	0.0082	0.017
3μ+5Jet+MET+X	100 - 1250	2.4	2	8	0.047	0.054
3μ+4Jet+MET+X	100 - 140	2.4	1.5	9	0.011	0.07
$4e$+1Jet+MET+X	190 - 350	0.058	0.034	1	0.059	0.12
$4e$+MET+X	100 - 770	0.49	0.31	2	0.11	0.13
$3e$+1μ+1bJet+MET+X	140 - 170	0.051	0.037	1	0.055	0.14
$2e$+2bJet+1Jet+MET+X	90 - 120	54	17	97	0.012	0.15
$1e$+2Jet+MET+X	250 - 310	6900	700	8597	0.0095	0.16
$2e$+1μ+4Jet+MET+X	110 - 150	0.89	0.84	4	0.053	0.16

Table 5.8.: Comprehensive examination of the most significant event classes within the Region of Interest (RoI) scan of the inclusive class of the kinematic distribution p_T^{miss}. The parameter p denotes the median p-value derived from multiple pseudoexperiments, while $\tilde{\text{p}}$ signifies the p-value adjusted for the Look-Elsewhere Effect (LEE). The RoI delineates the boundaries of the region of interest, detailed in Section 4.5.4. N_{Data} represents the count of pseudo-data events, N_{MC} corresponds to the anticipated number of events conforming to the Standard Model from Monte Carlo simulations and σ_{MC} as a combination of statistical and systematic uncertainty.

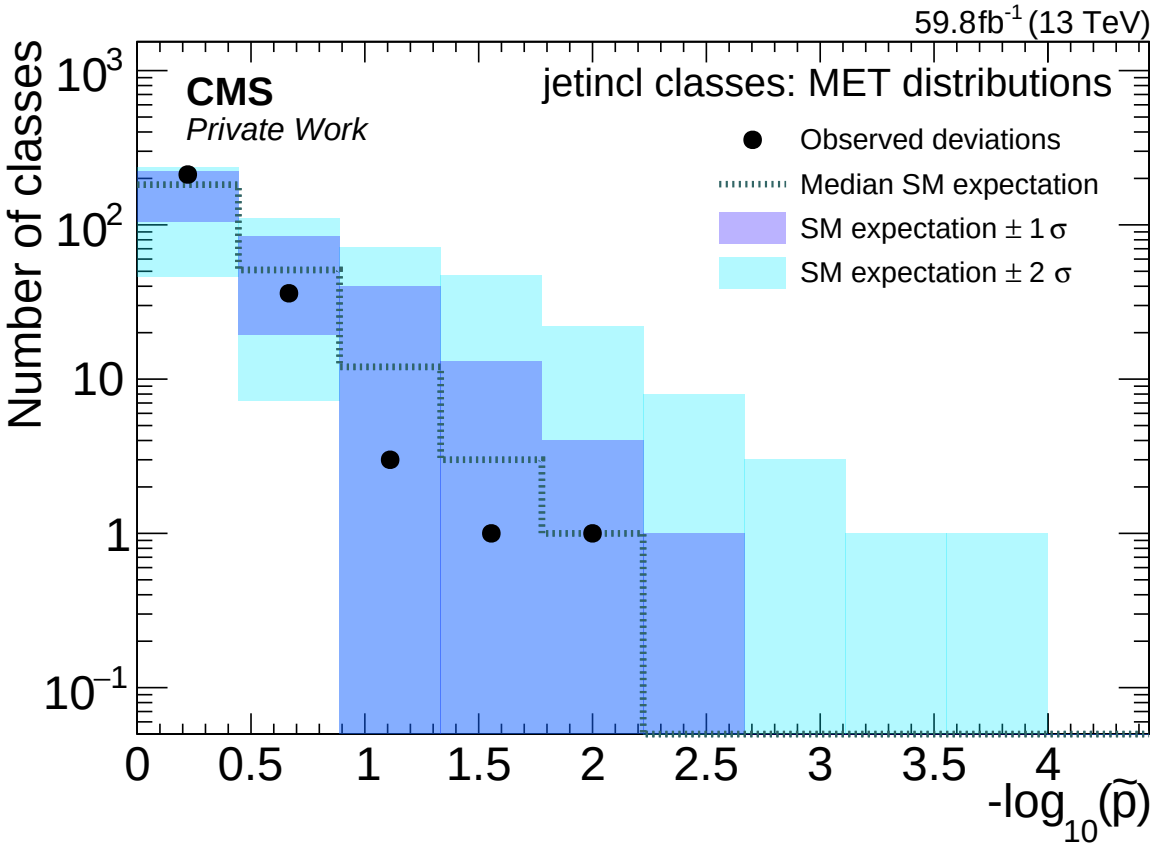


Fig. 5.11.: Distribution of the $\tilde{p}$ values pertaining to the kinematic variable p_T^{miss} across jet-inclusive event classes based on a RoI scan for the MUSiC analysis. The rightmost black dot belongs to the first event class in table 5.9.

Event Class	**RoI [GeV]**	$\mathbf{N_{MC}}$	$\boldsymbol{\sigma_{MC}}$	$\mathbf{N_{Data}}$	**p**	**p̃**
$2e$+1μ+1γ+MET+NJets	160 - 290	0.37	0.36	4	0.0041	0.0081
$2e$+1μ+1γ+1Jet+MET+NJets	100 - 530	0.65	0.49	4	0.015	0.038
$1e$+2Jet+MET+NJets	270 - 330	4000	430	5175	0.0028	0.06
3μ+4Jet+MET+NJets	100 - 190	4	2.2	12	0.017	0.084
$1e$+3Jet+MET+NJets	270 - 350	2200	260	2855	0.0052	0.097
$1e$+1μ+3bJet+1Jet+MET+NJets	140 - 170	0.82	0.63	4	0.031	0.13
1μ+3Jet+MET+NJets	330 - 390	3100	370	4017	0.0093	0.15
$2e$+2bJet+1Jet+MET+NJets	90 - 120	53	16	96	0.012	0.15
3μ+2bJet+MET+NJets	310 - 390	0.046	0.046	1	0.057	0.16
3μ+5Jet+MET+NJets	310 - 390	0.1	0.074	1	0.11	0.16

Table 5.9.: Comprehensive examination of the most significant event classes within the Region of Interest (RoI) scan of the jet-inclusive class of the kinematic distribution p_T^{miss}. The parameter p denotes the median p-value derived from multiple pseudoexperiments, while p̃ signifies the p-value adjusted for the Look-Elsewhere Effect (LEE). The RoI delineates the boundaries of the region of interest, detailed in Section 4.5.4. N_{Data} represents the count of pseudo-data events, N_{MC} corresponds to the anticipated number of events conforming to the Standard Model from Monte Carlo simulations and σ_{MC} as a combination of statistical and systematic uncertainty.

Of the 42 event classes identified, a total of eight event classes were determined for which the $\tilde{p}$ value was so low that they can no longer be explained by statistical fluctuations in the Standard Model with a significance of over 4σ. These eight event classes are as follows:

Event Class	**Total number of appearances outside the SM expectation**
$2e$+4Jet	2
2μ	1
2μ+MET	1
4μ+4Jet	2
$1e$+1μ+3bJet+2Jet	1
4μ	4
$2e$+2μ	1
$2e$+1Jet+MET	2

Note that no event classes outside the SM expectation of MUSiC were detected in the inclusive, jet-inclusive class of invariant mass (M_{inv}) and all missing transverse momentum (p_T^{miss}) scans. All these eight classes were found within the scans of the exclusive class of the Invariant Mass Scan and the Sum of Transverse Momentum (S_T) scans.

A closer look at the results reveals that some event classes were found to be most significant in several scans. This phenomenon can be explained not only by potential effects that could lead to a deviation from the SM, such as the discovery of a new particle, but also by calibration problems or correlations between all observations.

A more detailed analysis of these event classes, which appeared in several top 10 tables of the various scans, quickly reveals that all event classes that repeat six times, together with one of the two event classes that repeat five times, belong to the eight event classes that deviate from the expectation of the Standard Model according to the scan results of MUSiC.

In the following, the eight event classes mentioned, which were found outside the SM expectation, as well as the second event class, which repeats five times across different scans, are therefore examined in more detail.

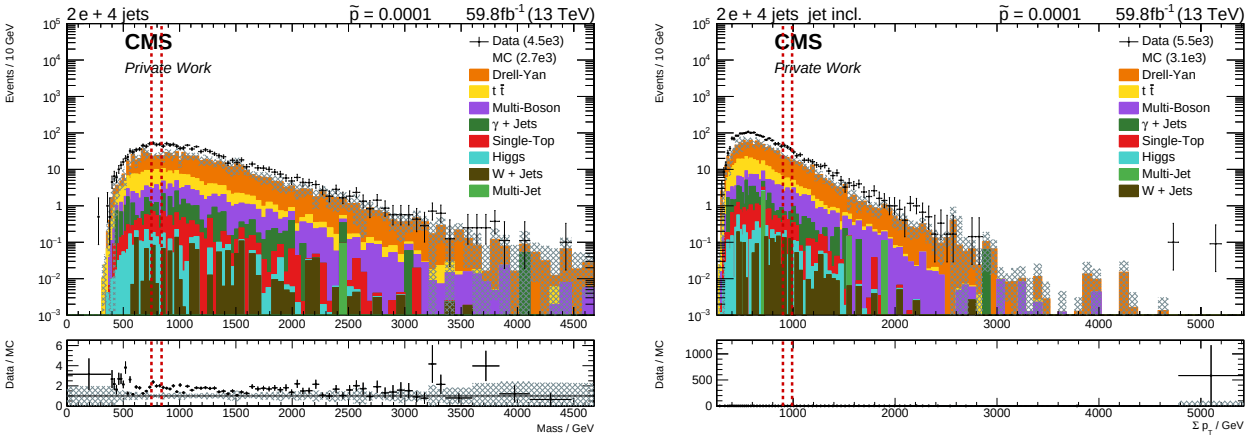
Event Class	**Total number of appearances**
$2e$+4Jet	6
4μ	6
$2e$+2μ	6
4μ+4Jet	5
$2e$+1μ+1γ+MET	5
$2e$+1Jet	3
$2e$+5Jet	3
$2e$+1μ+1γ+1Jet+MET	3
$2e$+2bJet+1Jet+MET	3
3μ+4Jet+MET	3
3μ+5Jet+MET	3
4μ+3Jet	3
$1e$	2
$1e$+1μ+3bJet+2Jet	2
$1e$+1γ	2
$1e$+2Jet+MET	2
$1e$+3Jet+MET	2
$1e$+1μ+1γ+2bJet	2
$1e$+2μ+1γ+1bJet	2
$2e$+6Jet	2
$2e$+1Jet+MET	2
1μ+3Jet+MET	2
2μ	2
$1e$+2bJet	1
$1e$+MET	1
$1e$+3μ+2Jet	1
$1e$+1μ+1γ+1Jet+MET	1
$1e$+3bJet+2Jet+MET	1
$1e$+1μ+3bJet+1Jet+MET	1
$2e$+3Jet	1
$2e$+3bJet+1Jet+MET	1
$2e$+1μ+4Jet+MET	1
$3e$+1μ+1bJet+MET	1
$4e$+1Jet+MET	1
$4e$+MET	1
1μ+1γ	1
1μ	1
2μ+MET	1
2μ+3bJet+1Jet	1
2μ+1Jet	1
3μ+1bJet+2Jet+MET	1
3μ+2bJet+MET	1

Table 5.10.: Tabular representation of all unique event classes identified within the Region of Interest scan of the three kinematic distributions (M_{inv}, S_T and p_T^{miss}), along with the frequency of their occurrences in the respective distributions across different scans.

5.2. Event Class studies for the 2018 Dataset

5.2.1. 2e+4Jet

The event class $2e$+4Jet was identified twice in the MUSiC scan as deviating from the expectation of the Standard Model. Once this was observed with a $\tilde{p} = < 10^{-5}$ in the invariant mass exclusive class scan with the region of interest from 750 to 840 GeV and once with a $\tilde{p} = 0.0001$ in the sum of transverse momentum jet-inclusive class scan with the RoI from 900 to 990 GeV. As can be seen from figure 5.12, this process is mainly dominated by Drell-Yan events. Furthermore, both cases show larger deviations ($> 4\sigma$) between the data and the Monte Carlo simulation not only in the RoI, but over the entire Energy range. A comparison of this observation with other $2e$ classes, where the number of jets is varied, indicates that the deviation decreases with a smaller number of jets. This analysis was performed to investigate the influence of the number of jets on the observed deviations. A more thorough examination in further internal analyses revealed this behavior in all MUSiC preliminary analyses of the 2018 data. A preliminary assumption regarding this deviation points to missing ZG samples, which were not yet available as Monte Carlo samples at the time of this work and could therefore only be considered in this analysis with limited statistics. Another possible reason could be the potentially poor jet multiplicity simulation of MC samples. Further comprehensive investigations are currently being carried out to determine the exact cause.



(a) **(b)**

Fig. 5.12.: Distribution of the $2e$+4Jet M_{inv} exclusive event class (left) and S_T jet-inclusive event class (right), with the Region of Interest demarcated by red dashed lines.

5.2.2. 2μ

The event class of 2μ was registered as one of the most significant event classes in the sum of the transverse momentum jet-inclusive scans with a $\tilde{p} = 0.0001$. This class, as shown in Figure 5.13, is again dominated by Drell-Yan events. However, MUSiC has detected a slight elevation in the range of 320 to 360 GeV, where there is a deviation of about 1σ between the

data and the Monte Carlo simulation. Such an effect occasionally occurs in regions where the low p_T trigger transitions to the high p_T trigger (for muons at a threshold of $p_T = 200\,\text{GeV}$) and thus a superposition takes place. To counteract this effect, specific consideration is required, which has not yet been implemented in this case. By implementing this approach, the observed deviation should be eliminated, which would lead to an improved agreement between the data and the MC in the rest of the figure.

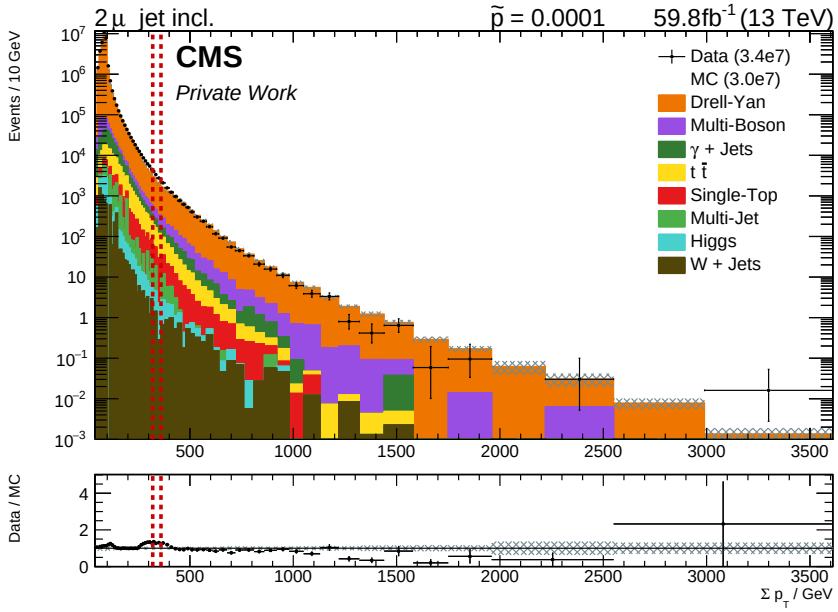


Fig. 5.13.: Distribution of the 2μ S_T jet-inclusive event class, with the Region of Interest demarcated by red dashed lines.

5.2.3. 2μ+MET

In the event class 2μ+MET (Fig.5.14), a deviation of up to 8σ in the range from 0 to 30 GeV with $\tilde{p} = < 10^{-5}$ was determined as part of the invariant mass exclusive scan. A detailed examination of the region of interest, taking into account further MUSiC analyses from 2018 and older MUSiC analyses, shows that the number of Monte Carlo events ($N_{MC} = 19 \pm 12$) in this range is plausible. However, the number of data points ($N_{Data} = 96$) is in comparison unexpectedly large, which leads to the mentioned deviation of up to 8σ. Initial investigations indicate that some of the data points were probably incorrectly reconstructed by MUSiC, resulting in a higher number in this event class than expected. More detailed investigations are currently being carried out to verify this observation. In the remaining area of this event class, there appears to be good agreement between the MC simulations and the data, as well as the expected Drell-Yan dominance.

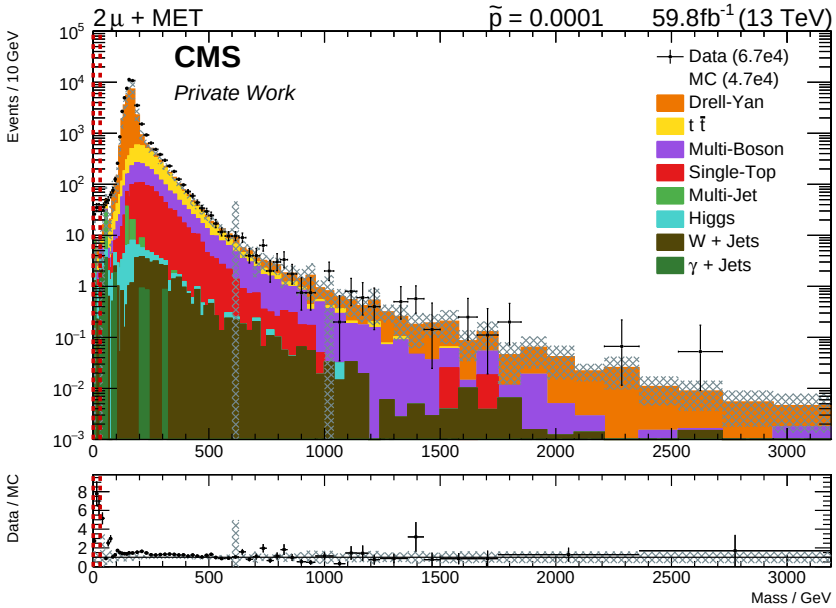
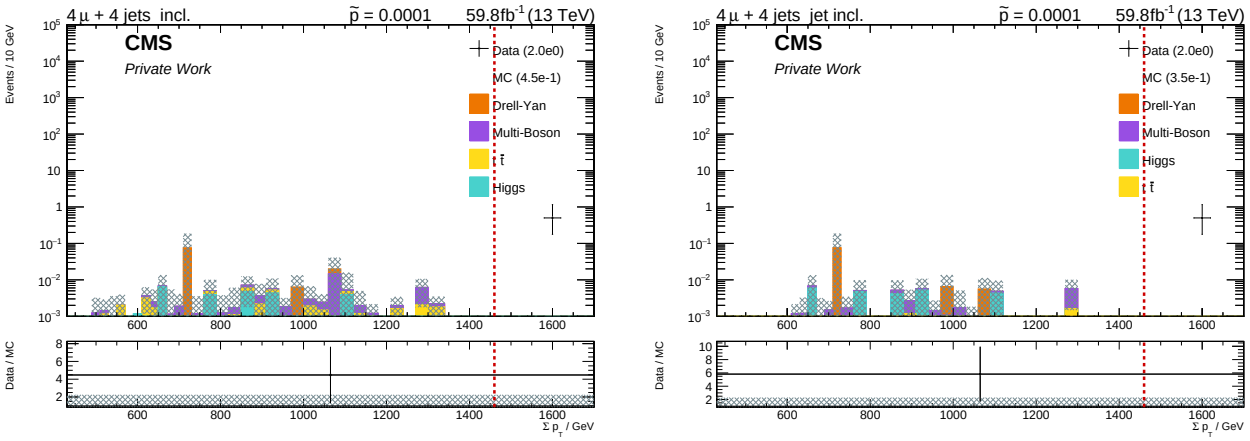


Fig. 5.14.: Distribution of the 2μ+MET M_{inv} exclusive event class, with the Region of Interest demarcated by red dashed lines.

5.2.4. 4μ+4Jet

The event class 4μ+4Jet (Fig.5.15) was identified as one of the most significant event classes in the Sum of Transverse Momentum Inclusive and Sum of Transverse Momentum Jet-Inclusive scans, with a $\tilde{p} = < 10^{-5}$ and $\tilde{p} = 0.0001$, respectively. In both cases, two events were found within the region of interest from 1460 to 2030 GeV, while the Monte Carlo expectations are $N_{MC} = 0.0073 \pm 0.0063$ and $N_{MC} = 0.0036 \pm 0.0032$, respectively. The event class has a comparatively small number of events, with a N_{MC} of 0.45, which are mainly generated by Drell-Yan, multi-boson, $t\bar{t}$ and Higgs events.



(a) **(b)**

Fig. 5.15.: Distribution of the 4μ+4Jet S_T inclusive event class (left) and jet-inclusive event class (right), with the Region of Interest demarcated by red dashed lines.

5.2.5. 2e+1μ+1γ+MET

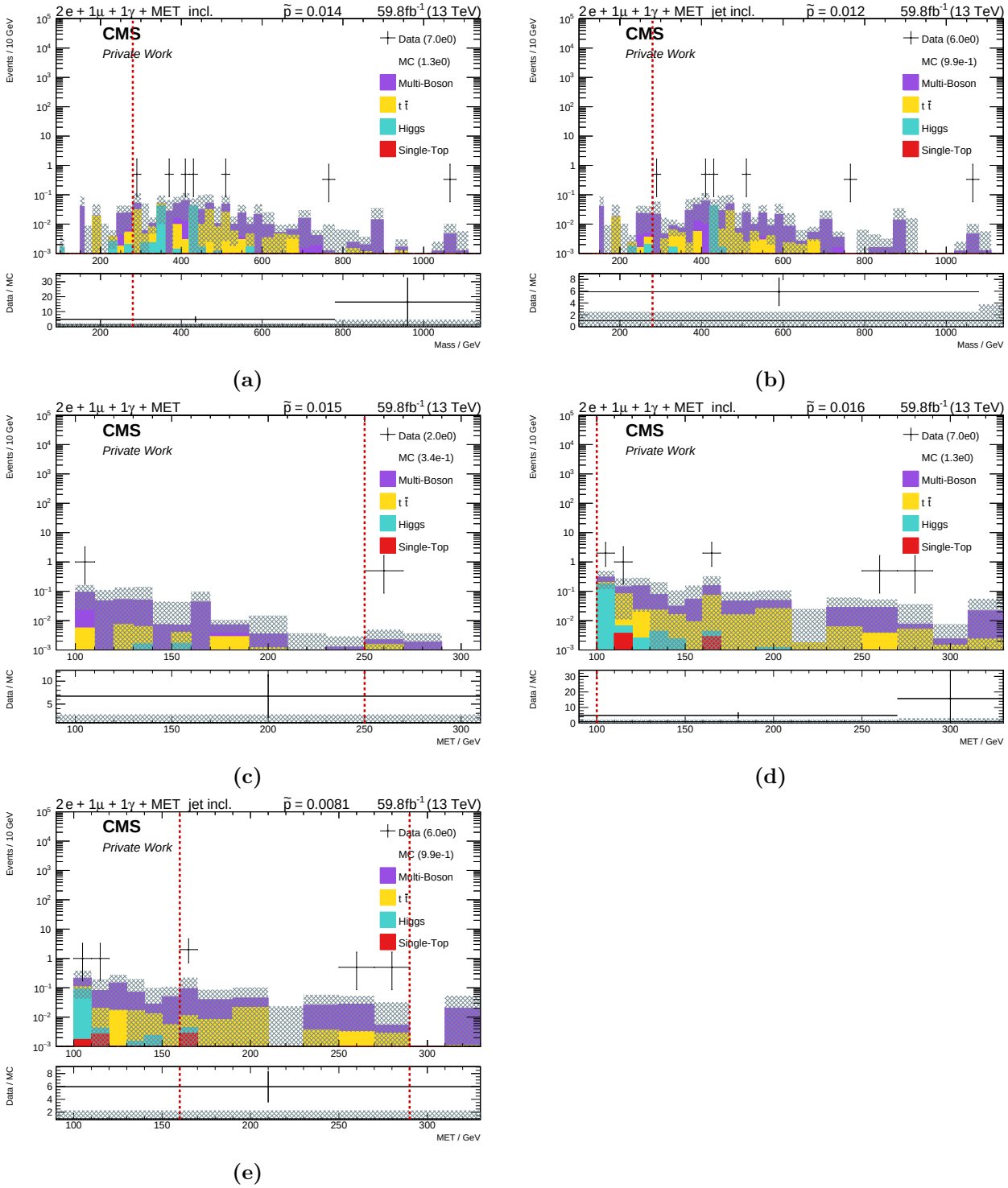


Fig. 5.16.: Distribution of the $2e{+}1\mu{+}1\gamma{+}$MET M_{inv} inclusive event class (5.16a), M_{inv} jet-inclusive event class (5.16b), MET exclusive event class (5.16c), MET inclusive event class (5.16d) and MET jet-inclusive event class (5.16e), with the Region of Interest demarcated by red dashed lines.

The event class $2e{+}1\mu{+}1\gamma{+}$MET was not identified as deviating from the expectations of the Standard Model in any of the conducted scans. However, it appeared in five out of the nine different tables, representing the top 10 most significantly classified event classes in various scans (tables 5.1-5.9). Given the correlated nature of all distributions, such an event class is also of particular interest. The Analysis of the associated distributions (Fig.5.16) shows a dominant presence of Monte Carlo events, especially multi-boson and $t\bar{t}$ events, with a total event count of less than 2 events in each distribution. The regions of interest of all scans contain almost every event of the data in all distributions. Nevertheless, $\tilde{p}$ values of 0.012-0.016 result in less significant deviations. Only the $\tilde{p}$ value of the MET Jet-Inclusive scan falls into a more significant range of 0.0081, but remains within the expectations of the SM.

5.2.6. 1e+1μ+3bJet+2Jet

The event class $1e{+}1\mu{+}$3bJet+2Jet (Fig.5.17) was classified as one of the most significant event classes within the sum of transverse momentum inclusive scans, with a $\tilde{p} = < 10^{-5}$. There are 11 data events within this event class, 7 of which are found in the region of interest of 760 to 850 GeV, while the Monte Carlo expectation is $N_{MC} = 0.41 \pm 0.4$. The event class has a comparatively small number of events, with a N_{MC} of 4.7, which are mainly generated by $t\bar{t}$, Higgs, single top and multi-boson events.

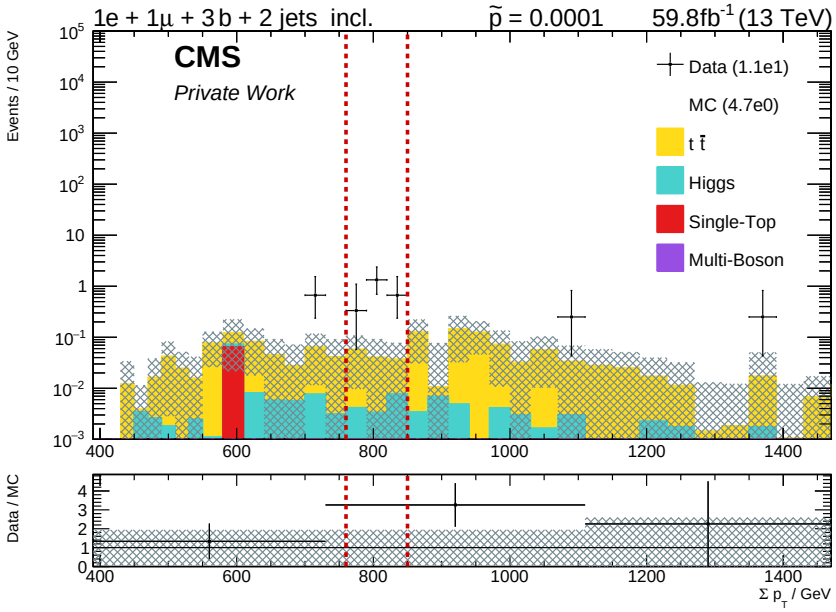


Fig. 5.17.: Distribution of the $1e{+}1\mu{+}$3bJet+2Jet S_T inclusive event class, with the Region of Interest demarcated by red dashed lines.

5.2.7. 4μ

The event class 4μ (Fig.5.18) was found four times within the MUSiC scan as one of the most significant event classes, which show deviations that lie outside the expectations of the Standard

Model. With $\tilde{p} = 0.0001$ in the invariant mass exclusive scan and with $\tilde{p} = < 10^{-5}$ in all sum of transverse momentum scans, this event class shows a high significance.

A closer look at the event class reveals that it is mainly dominated by multi-boson events and that all regions of interest are very close to each other (RoI M_{inv} Excl.: 180-320 GeV, RoI S_T: 140-170 GeV). A deviation of up to about 4σ within the RoI is observed. However, such deviations have already been detected in previous MUSiC analyses of multi-boson dominated event classes and required a more detailed investigation of the k-factor used (Eq.4.1).

Furthermore, the plot of the M_{inv}. Excl. scans shows that although the Higgs peak is correctly predicted by the Monte Carlo data, MUSiC has not detected any event in the data. Since Higgs production is now a well known process, MUSiC should be able to detect it. The exact reason why this is not the case in this instance is not yet clear and is currently being investigated by the rest of the MUSiC team.

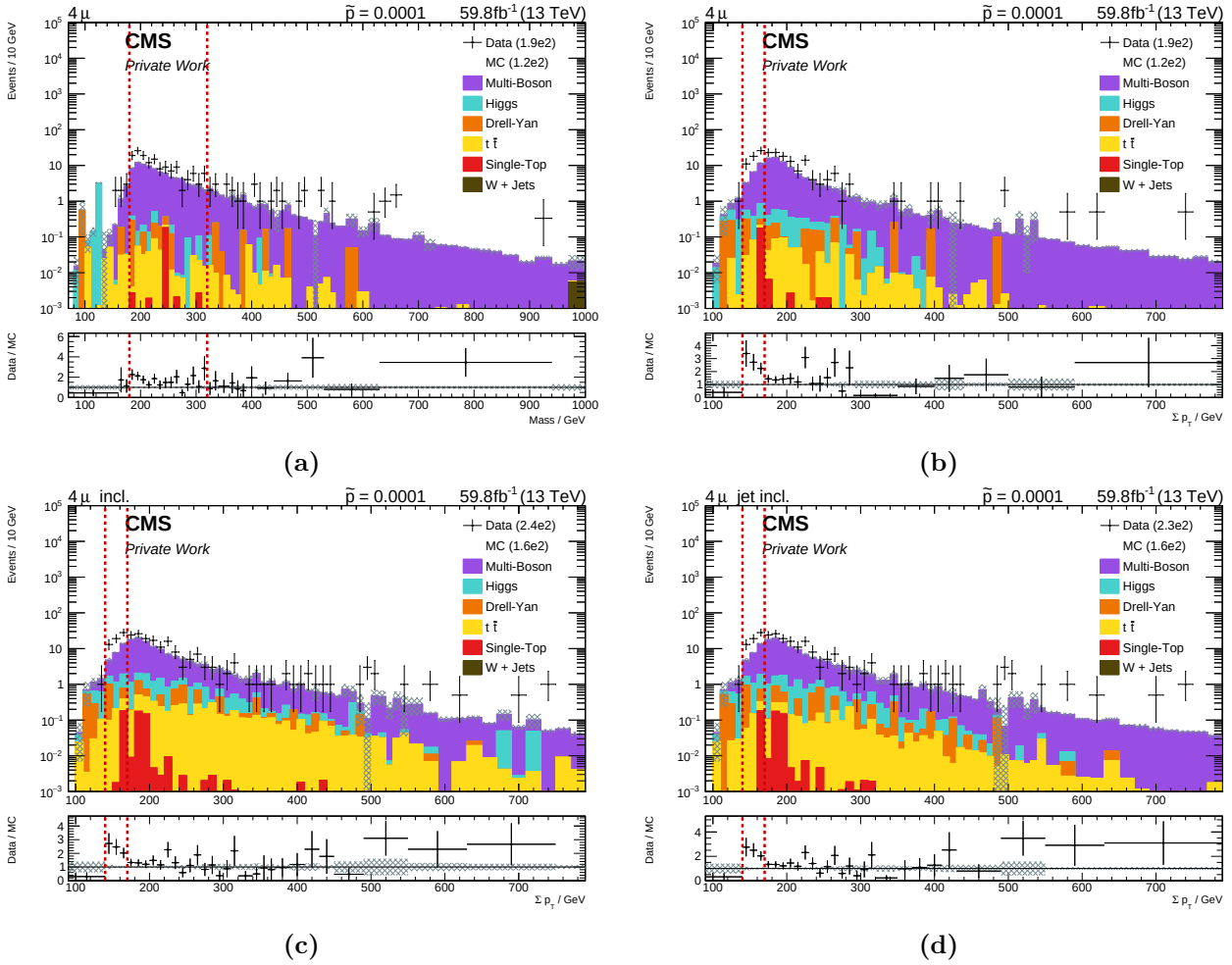


Fig. 5.18.: Distribution of the 4μ M_{inv} exclusive event class (5.18a), S_T exclusive event class (5.18b), S_T inclusive event class (5.18c) and S_T jet-inclusive event class (5.18d), with the Region of Interest demarcated by red dashed lines.

5.2.8. 2e+2μ

The event class $2e{+}2\mu$ (Fig.5.19) was once found among the most significantly classified event classes that lie outside the expectations of the Standard Model, with a $\tilde{p}$ of 0.0001. Similar to the 4μ event class, this is also dominated by multi-boson processes and shows a comparable deviation of up to 4σ in its region of interest at 190-420 GeV. The currently most plausible explanation for this effect probably lies in the same phenomenon that generates the deviations in the 4μ event class.

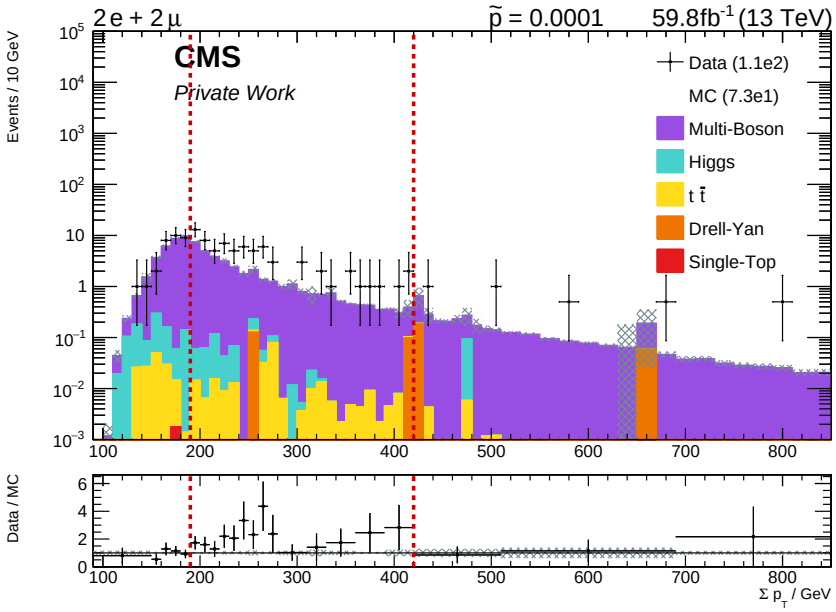
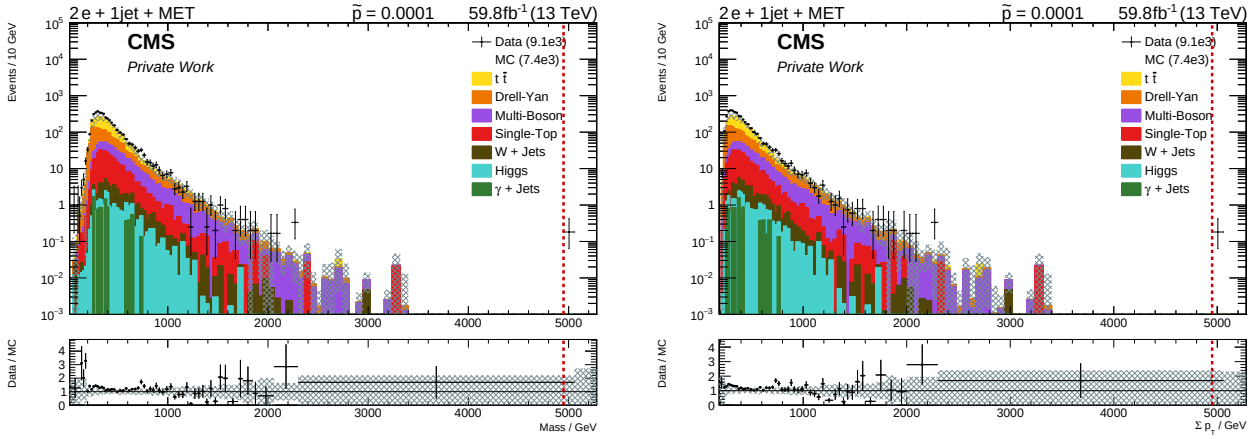


Fig. 5.19.: Distribution of the $2e{+}2\mu$ S_T exclusive event class, with the Region of Interest demarcated by red dashed lines.

5.2.9. 2e+1Jet+MET

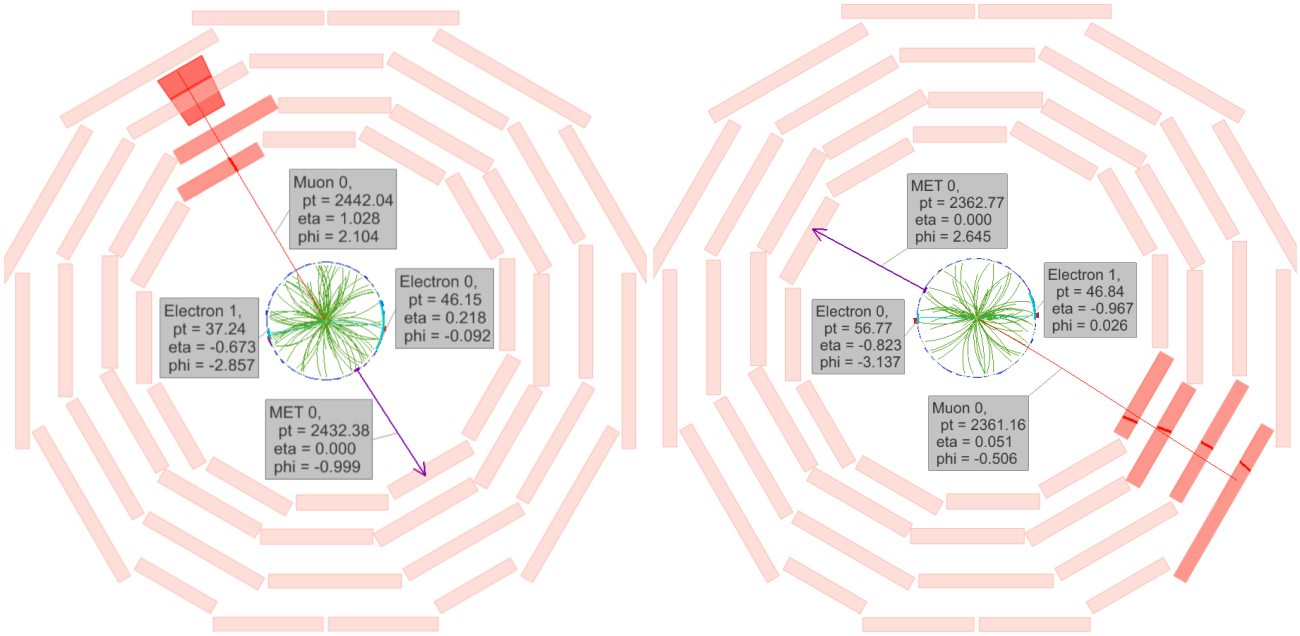
The event class $2e{+}1\text{Jet}{+}\text{MET}$ (Fig.5.20) is the last event class identified by MUSiC as deviating significantly from the expectation of the Standard Model. This class was detected twice, once in the M_{inv} Excl and once in the S_T Excl scan. In both cases, it has a significant $\tilde{p}$ value $< 10^{-5}$. This event class is mainly dominated by $t\bar{t}$ and Drell-Yan events, whereby within the distribution up to a value of approx. 3500 GeV there is a good match with the SM expectation. In both the M_{inv} and the S_T scan in the Exclusive Class at around 5000 GeV, the same two data events were recorded that lie far outside the SM expectation and, according to the Monte Carlo simulations, no further events can be expected. Such outliers have occurred in other event classes and especially in previous MUSiC analyses, but rarely had a significant weighting, so their $\tilde{p}$ values were usually not significant.



(a) **(b)**

Fig. 5.20.: Distribution of the $2e$+1Jet+MET M_{inv} exclusive event class (left) and S_T exclusive event class (right), with the Region of Interest demarcated by red dashed lines.

In order to find the reason for this events and to rule out the possibility that they are Beyond the Standard Model signals, a closer look was therefore taken at these two events in the form of an event display (Fig.5.21).



(a) **(b)**

Fig. 5.21.: Two event displays of the 2 data events that were outside the SM expectation in the event class of the $2e$+1Jet+MET scan, shown in the ρ-ϕ plane of the detector. These events both show two electrons, a muon and missing transverse momentum. The electrons fulfil the prescribed identification criteria and the muons are detected as such by CMS, but incorrectly categorised as a jet by MUSiC.

In the event display of both events classified as significant, the two electrons and the MET part are very clearly recognizable and show the expected properties. What is missing, is the jet,

which has led to these events being classified in the $2e$+1Jet+MET Exclusive Class. Instead, there is a high-energy muon with $p_T \approx 2440\,\text{GeV}$ (Fig.5.21a) and $p_T \approx 2360\,\text{GeV}$ (Fig.5.21b) in both events. For reasons as yet unknown, these muons appear to have been incorrectly reconstructed as jets in MUSiC and therefore caused the outlier signals in the $2e$+1Jet+MET Exclusive Class.
Interestingly, even when considering the event class $2e$+1μ+MET matching the event display, such events are not within the SM expectation. Further investigation is required to find out exactly why this misreconstruction or misclassification occurs.

6. Conclusion and Outlook

This research work presents a comprehensive investigation of the Model Unspecific Search in CMS (MUSiC) within the context of exploring potential phenomena beyond the Standard Model. The focus lies on the conception and operational procedures of MUSiC, encompassing the identification criteria and event classification based on their final states. Furthermore, a detailed discussion ensues regarding the statistical analysis to assess the agreement between measured CMS data and predictions of the Standard Model.

A first in-depth scrutiny of the extensive analysis of data recorded by CMS during Run II of the Large Hadron Collider in 2018 at a center of mass energy of $\sqrt{s} = 13\,\text{TeV}$ and an integrated luminosity of $59.8\,\text{fb}^{-1}$ has been undertaken. Despite intensive efforts, there is currently no evidence of new physics beyond the standard model. Possible significant deviations are preliminary attributed to incomplete implementations or show similarities with observations in event classes dominated by Di-Bosons, as in the years 2015 and 2016.

Due to various challenges, including incomplete Monte Carlo simulations, alterations in data structures within samples, and the substantial computational demands, the depth and extent of this work had to be curtailed. This resulted in a limited sensitivity in certain final states, thus preventing a final statement for the analysis of the 2018 dataset at this juncture.

Notwithstanding these limitations, this work provides a fundamental insight into the 2018 dataset and identifies key aspects upon which the MUSiC team, which is conducted by multiple researchers in Aachen and is grounded in the collaborative efforts of the entire CMS collaboration, can build for future analyses.

In my role as the final user of the previous procedure, my task was to adapt the existing MUSiC code to the altered conditions of the analysis, transitioning from the 2016 dataset to the 2018 dataset. This facilitated an initial exploration of the 2018 data, allowing for the identification of preliminary conclusions and issues that may be pertinent for future analyses. The results obtained during this work also serve as a valuable reference for the ongoing upgrade of MUSiC, in which the entire MUSiC team is currently engaged. The definitive deepening of insights from the 2018 dataset and potential indications of new physical phenomena are anticipated once the upgrade is completed, and the final outstanding Monte Carlo productions become available.

7. Acknowledgement

Finally, I would like to express my sincere gratitude to a remarkable group of people who have played pivotal roles in my journey. Prof. Dr. Thomas Hebbeker deserves my deepest gratitude for not only affording me the chance to be part of this exceptional project but also for sharing invaluable glimpses into the future of physics. I am equally indebted to Prof. Dr. Martin Erdmann for graciously taking on the role of the second examiner for my work.
A warm and sincere thank you is reserved for the entire MUSiC team. Working alongside them during my research has been an enriching experience. My appreciation goes out to M.Sc. Lorenzo Vigilante and Dr. Felipe Torres da Silva de Araujo, my mentors, whose unwavering support helped me navigate challenges. I also want to express my gratitude to M.Sc. Ana Andrade, M.Sc. Chinmay Seth, Lucas Karwatzki, B.Sc. Nils Esper, and B.Sc. Patrick Kersten for their outstanding collaboration and support.
I extend additional thanks to Dr. Arnd Meyer, Dr. Sebastian Wiedenbeck, M.Sc. Fabian Nowotny and M.Sc. Valentina Sarkisovi for their invaluable assistance during my time at the Institute 3a.
A special note of appreciation goes to M.Sc. Christoph Domnik and M.Sc. Ronald Oellers, who generously lent their expertise to proofread my work and provided valuable insights.
Lastly, I want to express my deepest gratitude to four extraordinary individuals in my life: my parents, Eric & Anabela Kaiser, my brother Mikel Kaiser and my girlfriend Evelyne Ibe. Their enduring support has been my pillar of strength, enabling me to overcome challenges and reach this significant milestone in my academic journey.

8. Bibliography

1. Duchardt, D. MUSiC: A Model Unspecific Search for New Physics Based on CMS Data at $\sqrt{s} = 8$ TeV. https://web.physik.rwth-aachen.de/user/hebbeker/theses/duchardt_phd.pdf (Aug. 2017).
2. CERN. The Standard Model. https://home.cern/science/physics/standard-model.
3. Kaiser, Y. Sensitivity study of the MUSIC algorithm for LHC run 2. https://web.physik.rwth-aachen.de/user/hebbeker/theses/kaiser_bachelor.pdf (Apr. 2021).
4. Lieb, J. Development of a Fast Search Algorithm for the MUSiC Framework. *Bachelor thesis. RWTH Aachen University, Sept,* 50. https://web.physik.rwth-aachen.de/user/hebbeker/theses/lieb_bachelor.pdf (Sept. 2015).
5. Stroucken, A. *submitter: Sensitivity Studies for the Model Unspecific Search in CMS (MUSiC) at $\sqrt{s} = 13$ TeV* PhD thesis (RWTH Aachen U., Aug. 2016). https://web.physik.rwth-aachen.de/user/hebbeker/theses/stroucken_bachelor.pdf.
6. Rehbein, F. Search for sphalerons in the eμ final state at $\sqrt{s} = 13$ TeV with the CMS experiment. https://web.physik.rwth-aachen.de/user/hebbeker/theses/rehbein_bachelor_corrected.pdf (Aug. 2018).
7. Cush. Standard model of elementary particles: the 12 fundamental fermions and 5 fundamental bosons. https://commons.wikimedia.org/w/index.php?curid=4286964 (Sept. 2019).
8. Pieta, H. MUSiC. A model unspecific search in CMS based on 2010 LHC data. https://web.physik.rwth-aachen.de/user/hebbeker/theses/pieta_phd.pdf (June 2012).
9. Lieb, J. Discovery Potential of a Model Independent Search for New Physics at the LHC. *Preparation. Master thesis. RWTH Aachen University,* 51. https://web.physik.rwth-aachen.de/user/hebbeker/theses/lieb_master.pdf (Aug. 2017).
10. Roemer, J. T. Model Unspecific Search for New Physics with pp Collisions at $\sqrt{s} = 13$ TeV with the CMS Experiment, https://web.physik.rwth-aachen.de/user/hebbeker/theses/roemer_master.pdf (Jan. 2017).
11. Fukuda, Y. *et al.* Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.* **81,** 1562–1567. https://link.aps.org/doi/10.1103/PhysRevLett.81.1562 (8 Aug. 1998).

12. Fukuda, S. *et al.* Tau Neutrinos Favored over Sterile Neutrinos in Atmospheric Muon Neutrino Oscillations. *Phys. Rev. Lett.* **85,** 3999–4003. https://link.aps.org/doi/10.1103/PhysRevLett.85.3999 (19 Nov. 2000).

13. Pook, T. A Model Unspecific Search in CMS (MUSiC) using 13 TeV data. https://web.physik.rwth-aachen.de/user/hebbeker/theses/pook_phd.pdf (Mar. 2022).

14. Brunner, D. Suche nach zusätzlichen Raumdimensionen im Massenspektrum von zwei Elektronen bei $\sqrt{s} = 13$ TeV. https://web.physik.rwth-aachen.de/user/hebbeker/theses/brunner_bachelor.pdf (Aug. 2016).

15. Wu, C. S., Ambler, E., Hayward, R. W., Hoppes, D. D. & Hudson, R. P. Experimental Test of Parity Conservation in Beta Decay. *Phys. Rev.* **105,** 1413–1415. https://link.aps.org/doi/10.1103/PhysRev.105.1413 (4 Feb. 1957).

16. Wiedenbeck, S. Search for New Physics in the e + MET Final State in the 2016 CMS Dataset. https://web.physik.rwth-aachen.de/user/hebbeker/theses/wiedenbeck_master.pdf (Feb. 2018).

17. Glashow, S. L. The renormalizability of vector meson interactions. *Nuclear Physics* **10,** 107–117. ISSN: 0029-5582. https://www.sciencedirect.com/science/article/pii/0029558259901968 (1959).

18. Salam, A. & Ward, J. C. Weak and electromagnetic interactions. *Il Nuovo Cimento (1955-1965)* **11,** 568–577. ISSN: 1827-6121. https://doi.org/10.1007/BF02726525 (4 Feb. 1959).

19. Weinberg, S. A Model of Leptons. *Phys. Rev. Lett.* **19,** 1264–1266. https://link.aps.org/doi/10.1103/PhysRevLett.19.1264 (21 Nov. 1967).

20. Particle Data Group *et al.* Review of particle physics. *Progress of Theoretical and Experimental Physics* **2020,** 083C01. https://pdg.lbl.gov/2020/download/db2020.pdf (2020).

21. Aaij, R. *et al.* Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays. *Phys. Rev. Lett.* **115,** 072001. https://link.aps.org/doi/10.1103/PhysRevLett.115.072001 (7 Aug. 2015).

22. Schmüser, P. *Feynman-Graphen und Eichtheorien für Experimentalphysiker* (Springer, 1988).

23. Higgs, P. W. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.* **13,** 508–509. https://link.aps.org/doi/10.1103/PhysRevLett.13.508 (16 Oct. 1964).

24. Englert, F. & Brout, R. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.* **13,** 321–323. https://link.aps.org/doi/10.1103/PhysRevLett.13.321 (9 Aug. 1964).

25. Guralnik, G. S., Hagen, C. R. & Kibble, T. W. B. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.* **13,** 585–587. https://link.aps.org/doi/10.1103/PhysRevLett.13.585 (20 Nov. 1964).

26. Aad, G. *et al.* Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B* **716,** 1–29. ISSN: 0370-2693. http://dx.doi.org/10.1016/j.physletb.2012.08.020 (Sept. 2012).

27. Chatrchyan, S. *et al.* Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B* **716,** 30–61. ISSN: 0370-2693. http://dx.doi.org/10.1016/j.physletb.2012.08.021 (Sept. 2012).

28. Aad, G. *et al.* Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. https://link.springer.com/content/pdf/10.1007/JHEP08(2016)045.pdf (2016).

29. Parida, M. K. & Samantaray, R. Unification predictions with or without supersymmetry. *The European Physical Journal Special Topics* **229,** 3243–3262. ISSN: 1951-6401. http://dx.doi.org/10.1140/epjst/e2020-000024-x (Dec. 2020).

30. Mouche, P. Overall view of the LHC. Vue d'ensemble du LHC. General Photo. https://cds.cern.ch/record/1708847 (June 2014).

31. Lefèvre, C. *The CERN accelerator complex. Complexe des accélérateurs du CERN* Dec. 2008. https://cds.cern.ch/record/1260465.

32. CERN. Facts and figures about the LHC. https://home.cern/resources/faqs/facts-and-figures-about-lhc.

33. AFP. A view of Compact Muon Solenoid (CMS) Cavern at the European Organization for Nuclear Research (CERN). Photo by Richard Juilliart. https://sciencesprings.wordpress.com/2016/09/06/from-sixth-tone-nobel-winner-says-china-should-not-build-particle-collider/ (Feb. 2015).

34. The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation* **3,** S08004–S08004. https://doi.org/10.1088/1748-0221/3/08/s08004 (Aug. 2008).

35. Neutelings, I. CMS coordinate system. https://tikz.net/axis3d_cms/.

36. Neutelings, I. Pseudorapidity. https://tikz.net/axis2d_pseudorapidity/.

37. Kleine-Limberg, F. Inorganic scintillator calorimeters: CMS. http://www.kip.uni-heidelberg.de/atlas/seminars/WS2009_JC/CMS.pdf (Nov. 2009).

38. Karimäki, V. *et al. The CMS tracker system project: Technical Design Report* https://cds.cern.ch/record/368412 (CERN, Geneva, 1997).

39. Andrade, A. Machine Learning for a Model Unspecific Search in CMS. https://web.physik.rwth-aachen.de/user/hebbeker/theses/andrade_master.pdf (Apr. 2023).

40. *The CMS electromagnetic calorimeter project: Technical Design Report* https://cds.cern.ch/record/349375 (CERN, Geneva, 1997).

41. Particle Data Group *et al.* Atomic and nuclear properties of lead tungstate ($PbWO_4$). https://pdg.lbl.gov/2022/AtomicNuclearProperties/HTML/lead_tungstate.html (2022).

42. *The CMS hadron calorimeter project: Technical Design Report* http://cds.cern.ch/record/357153 (CERN, Geneva, 1997).

43. *The CMS magnet project: Technical Design Report* https://cds.cern.ch/record/331056 (CERN, Geneva, 1997).

44. Layter, J. G. *The CMS muon project: Technical Design Report* https://cds.cern.ch/record/343814 (CERN, Geneva, 1997).

45. Sirunyan, A. *et al.* Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of Instrumentation* **13,** P06015–P06015. ISSN: 1748-0221. http://dx.doi.org/10.1088/1748-0221/13/06/P06015 (June 2018).

46. The CMS Collaboration. MUSiC: A model unspecific search for new physics in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C 81 (2021) 629.* arXiv: 2010.02984 [hep-ex]. https://arxiv.org/abs/2010.02984 (2020).

47. Sjöstrand, T. *et al.* An introduction to PYTHIA 8.2. *Computer Physics Communications* **191,** 159–177. https://doi.org/10.1016%2Fj.cpc.2015.01.024 (June 2015).

48. Alwall, J. *et al.* The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics* **2014.** https://doi.org/10.1007%2Fjhep07%282014%29079 (July 2014).

49. Alwall, J. *et al.* Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *The European Physical Journal C* **53,** 473–500. https://doi.org/10.1140%2Fepjc%2Fs10052-007-0490-5 (Dec. 2007).

50. Frederix, R. & Frixione, S. Merging meets matching in MC@NLO. *Journal of High Energy Physics* **2012.** https://doi.org/10.1007%2Fjhep12%282012%29061 (Dec. 2012).

51. Nason, P. A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms. *Journal of High Energy Physics* **2004,** 040–040. https://doi.org/10.1088%2F1126-6708%2F2004%2F11%2F040 (Nov. 2004).

52. Nason, P. & Zanderighi, G. W^+ W^-, WZ and ZZ production in the powheg-BOX-V2. *The European Physical Journal C* **74.** https://doi.org/10.1140%2Fepjc%2Fs10052-013-2702-5 (Jan. 2014).

53. Campbell, J. M., Ellis, R. K., Nason, P. & Re, E. Top-pair production and decay at NLO matched with parton showers. *Journal of High Energy Physics* **2015.** https://doi.org/10.1007%2Fjhep04%282015%29114 (Apr. 2015).

54. Bagnaschi, E., Degrassi, G., Slavich, P. & Vicini, A. Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM. *Journal of High Energy Physics* **2012.** https://doi.org/10.1007%2Fjhep02%282012%29088 (Feb. 2012).

55. Frixione, S., Nason, P. & Oleari, C. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *Journal of High Energy Physics* **2007,** 070–070. https://doi.org/10.1088%2F1126-6708%2F2007%2F11%2F070 (Nov. 2007).

56. Alioli, S., Nason, P., Oleari, C. & Re, E. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *Journal of High Energy Physics* **2010.** https://doi.org/10.1007%2Fjhep06%282010%29043 (June 2010).

57. Alioli, S., Nason, P., Oleari, C. & Re, E. NLO single-top production matched with shower in POWHEG: s-and t-channel contributions. *Journal of High Energy Physics* **2009,** 111. https://doi.org/10.48550/arXiv.0907.4076 (2009).

58. Alioli, S., Nason, P., Oleari, C. & Re, E. NLO vector-boson production matched with shower in POWHEG. *Journal of High Energy Physics* **2008,** 060–060. https://doi.org/10.1088%2F1126-6708%2F2008%2F07%2F060 (July 2008).

59. Re, E. Single-top Wt-channel production matched with parton showers using the POWHEG method. *The European Physical Journal C* **71.** https://doi.org/10.1140%2Fepjc%2Fs10052-011-1547-z (Feb. 2011).

60. Alioli, S., Nason, P., Oleari, C. & Re, E. NLO Higgs boson production via gluon fusion matched with shower in POWHEG. *Journal of High Energy Physics* **2009,** 002–002. https://doi.org/10.1088%2F1126-6708%2F2009%2F04%2F002 (Apr. 2009).

61. Nason, P. & Oleari, C. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *Journal of High Energy Physics* **2010.** https://doi.org/10.1007%2Fjhep02%282010%29037 (Feb. 2010).

62. Melia, T., Nason, P., Röntsch, R. & Zanderighi, G. W^+ W^-, WZ and ZZ production in the POWHEG BOX. *Journal of High Energy Physics* **2011.** https://doi.org/10.1007%2Fjhep11%282011%29078 (Nov. 2011).

63. Bothmann, E. *et al.* Event generation with Sherpa 2.2. *SciPost Physics* **7.** https://doi.org/10.21468%2Fscipostphys.7.3.034 (Sept. 2019).

64. Bretz, H.-P. *et al.* A development environment for visual physics analysis. *Journal of Instrumentation* **7,** T08005. https://dx.doi.org/10.1088/1748-0221/7/08/T08005 (Aug. 2012).

65. Khachatryan, V. *et al.* Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST* **12,** P02014. arXiv: 1607.03663. http://cds.cern.ch/record/2198719 (2017).

66. The CMS Collaboration. Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV. *Journal of Instrumentation* **10,** P02006. https://dx.doi.org/10.1088/1748-0221/10/02/P02006 (Feb. 2015).

67. Micheas, A. C. & Dey, D. K. Prior and Posterior Predictive P-Values in the One-Sided Location Parameter Testing Problem. *Sankhyā: The Indian Journal of Statistics (2003-2007)* **65,** 158–178. ISSN: 09727671. http://www.jstor.org/stable/25053252 (2023) (2003).

68. Lyons, L. Open statistical issues in Particle Physics. *The Annals of Applied Statistics* **2,** 887–915. https://doi.org/10.1214/08-AOAS163 (2008).

69. Bayer, A. E. & Seljak, U. The look-elsewhere effect from a unified Bayesian and frequentist perspective. *Journal of Cosmology and Astroparticle Physics* **2020,** 009–009. ISSN: 1475-7516. http://dx.doi.org/10.1088/1475-7516/2020/10/009 (Oct. 2020).

70. Butterworth, J. *et al.* PDF4LHC recommendations for LHC Run II. *Journal of Physics G: Nuclear and Particle Physics* **43,** 023001. ISSN: 1361-6471. http://dx.doi.org/10.1088/0954-3899/43/2/023001 (Jan. 2016).

A. Appendix

A.1. List of Monte Carlo Samples

The information regarding the used MC samples can be found in the table provided below.

Processgroup	**Datasetpath**
Drell-Yan	/DYToMuMu_M-120To200_TuneCP5_13TeV-powheg-pythia8
	/DYToMuMu_M-200To400_TuneCP5_13TeV-powheg-pythia8
	/DYToMuMu_M-400To800_TuneCP5_13TeV-powheg-pythia8
	/DYToMuMu_M-800To1400_TuneCP5_13TeV-powheg-pythia8
	/DYToMuMu_M-1400To2300_TuneCP5_13TeV-powheg-pythia8
	/DYToMuMu_M-2300To3500_TuneCP5_13TeV-powheg-pythia8
	/DYToMuMu_M-3500To4500_TuneCP5_13TeV-powheg-pythia8
	/DYToMuMu_M-4500To6000_TuneCP5_13TeV-powheg-pythia8
	/DYToMuMu_M-6000ToInf_TuneCP5_13TeV-powheg-pythia8
	/DYToEE_M-120To200_TuneCP5_13TeV-powheg-pythia8
	/DYToEE_M-200To400_TuneCP5_13TeV-powheg-pythia8
	/DYToEE_M-400To800_TuneCP5_13TeV-powheg-pythia8
	/DYToEE_M-800To1400_TuneCP5_13TeV-powheg-pythia8
	/DYToEE_M-1400To2300_TuneCP5_13TeV-powheg-pythia8
	/DYToEE_M-2300To3500_TuneCP5_13TeV-powheg-pythia8
	/DYToEE_M-3500To4500_TuneCP5_13TeV-powheg-pythia8
	/DYToEE_M-4500To6000_TuneCP5_13TeV-powheg-pythia8
	/DYToEE_M-6000ToInf_TuneCP5_13TeV-powheg-pythia8
	/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/DYJetsToLL_M-10to50_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/DYJetsToLL_LHEFilterPtZ-50To100_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/DYJetsToLL_LHEFilterPtZ-100To250_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/DYJetsToLL_LHEFilterPtZ-250To400_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/DYJetsToLL_LHEFilterPtZ-400To650_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/DYJetsToLL_LHEFilterPtZ-650ToInf_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8
ZToInvisible	/ZJetsToNuNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8
	/ZJetsToNuNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8
	/ZJetsToNuNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8
	/ZJetsToNuNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8
	/ZJetsToNuNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8
	/ZJetsToNuNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8
	/ZJetsToNuNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8

W	/WToTauNu_M-200_TuneCP5_13TeV-pythia8-tauola
	/WToTauNu_M-500_TuneCP5_13TeV-pythia8-tauola
	/WToTauNu_M-1000_TuneCP5_13TeV-pythia8-tauola
	/WToTauNu_M-2000_TuneCP5_13TeV-pythia8-tauola
	/WToTauNu_M-3000_TuneCP5_13TeV-pythia8-tauola
	/WToTauNu_M-4000_TuneCP5_13TeV-pythia8-tauola
	/WToTauNu_M-5000_TuneCP5_13TeV-pythia8-tauola
	/WToMuNu_M-200_TuneCP5_13TeV-pythia8
	/WToMuNu_M-500_TuneCP5_13TeV-pythia8
	/WToMuNu_M-1000_TuneCP5_13TeV-pythia8
	/WToMuNu_M-2000_TuneCP5_13TeV-pythia8
	/WToMuNu_M-3000_TuneCP5_13TeV-pythia8
	/WToMuNu_M-4000_TuneCP5_13TeV-pythia8
	/WToMuNu_M-5000_TuneCP5_13TeV-pythia8
	/WToENu_M-200_TuneCP5_13TeV-pythia8
	/WToENu_M-500_TuneCP5_13TeV-pythia8
	/WToENu_M-1000_TuneCP5_13TeV-pythia8
	/WToENu_M-2000_TuneCP5_13TeV-pythia8
	/WToENu_M-3000_TuneCP5_13TeV-pythia8
	/WToENu_M-4000_TuneCP5_13TeV-pythia8
	/WJetsToLNu_Pt-100To250_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WJetsToLNu_Pt-250To400_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WJetsToLNu_Pt-400To600_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WJetsToLNu_Pt-600ToInf_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-pythia8
Gamma	/GJets_DR-0p4_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8
	/GJets_DR-0p4_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8
	/GJets_DR-0p4_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8
	/GJets_DR-0p4_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8
	/GJets_DR-0p4_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8

TTbar	/TT_Mtt-700to1000_TuneCP5_13TeV-powheg-pythia8
	/TT_Mtt-1000toInf_TuneCP5_13TeV-powheg-pythia8
	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8
	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8
	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8
TTG	/TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8
	/TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8
TTW	/TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8
	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8
TTZ	/TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8
	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8
TTbarTTbar	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8
WG	/WGToLNuG_01J_5f_PtG_500_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WGToLNuG_01J_5f_PtG_130_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WGJets_MonoPhoton_PtG-40to130_TuneCP5_13TeV-madgraph-pythia8
ZZ	/ZZTo4L_TuneCP5_13TeV_powheg_pythia8
	/ZZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8
	/ZZTo4Q_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8
WW	/WWTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WWTo4Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WWTo2L2Nu_MLL_200To600_TuneCP5_13TeV-powheg-pythia8
	/WWTo2L2Nu_MLL_600To1200_TuneCP5_13TeV-powheg-pythia8
	/WWTo2L2Nu_MLL_1200To2500_TuneCP5_13TeV-powheg-pythia8
	/WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8
QCD	/QCD_HT50to100_TuneCP5_13TeV-madgraphMLM-pythia8
	/QCD_HT100to200_TuneCP5_13TeV-madgraphMLM-pythia8
	/QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8
	/QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8
	/QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8
	/QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8
	/QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8
	/QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8
	/QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8

TG	/TGJets_TuneCP5_13TeV-amcatnlo-madspin-pythia8
WZ	/WZTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/WZTo1L3Nu_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8
ZZZ	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8
	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8
HIG	/ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8
	/ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8
	/ggZH_HToBB_ZToQQ_M-125_TuneCP5_13TeV-powheg-pythia8
	/ggZH_HToBB_ZToNuNu_M-125_TuneCP5_13TeV-powheg-pythia8
	/ggZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8
	/ZH_HToBB_ZToQQ_M-125_TuneCP5_13TeV-powheg-pythia8
	/ZH_HToBB_ZToNuNu_M-125_TuneCP5_13TeV-powheg-pythia8
	/ZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8
	/WplusH_HToBB_WToQQ_M-125_TuneCP5_13TeV-powheg-pythia8
	/WplusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8
	/WminusH_HToBB_WToQQ_M-125_TuneCP5_13TeV-powheg-pythia8
	/WminusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8
	/VHToNonbb_M125_TuneCP5_13TeV-amcatnloFXFX_madspin_pythia8
	/VBF_HToZZTo4L_M125_TuneCP5_13TeV_powheg2_JHUGenV7011_pythia8
	/VBF_HToZZTo2L2Nu_M125_TuneCP5_13TeV_powheg2_JHUGenV735_pythia8
	/VBFHToZG_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8
	/VBFHToZG_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8
	/VBFHToWWTo2L2Nu_M-125_TuneCP5_13TeV-powheg-jhugen727-pythia8
	/VBFHToTauTau_M125_TuneCP5_13TeV-powheg-pythia8
	/VBFHToGG_M125_TuneCP5_13TeV-amcatnlo-pythia8
	/VBFHToBB_M-125_TuneCP5_13TeV-powheg-pythia8
	/GluGluHToZZTo4L_M125_TuneCP5_13TeV_powheg2_JHUGenV7011_pythia8
	/GluGluHToZZTo2L2Nu_M125_TuneCP5_13TeV_powheg2_JHUGenV735_pythia8
	/GluGluHToZG_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8
	/GluGluHToZG_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8
	/GluGluHToWWTo2L2Nu_M125_TuneCP5_13TeV_powheg2_JHUGenV714_pythia8

	/GluGluHToTauTau_M125_TuneCP5_13TeV-powheg-pythia8
	/GluGluHToGG_M-125_TuneCP5_13TeV-powheg-pythia8
	/GluGluHToBB_M-125_TuneCP5_13TeV-powheg-pythia8
Top	/ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8
	/ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8
	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8
	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8
	/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8
	/ST_s-channel_4f_hadronicDecays_TuneCP5_13TeV-amcatnlo-pythia8
TTGG	/TTGG_0Jets_TuneCP5_13TeV-amcatnlo-madspin-pythia8
TZQ	/tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-amcatnlo-pythia8
WZZ	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8
GG	/DiPhotonJets_MGG-80toInf_TuneCP5_13TeV-amcatnloFXFX-pythia8
WZG	/WZG_TuneCP5_13TeV-amcatnlo-pythia8
WWW	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8
WWZ	/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8
WWG	/WWG_TuneCP5_13TeV-amcatnlo-pythia8
ZG	/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8
	/ZGTo2NuG_TuneCP5_13TeV-amcatnloFXFX-pythia8

A.2. List of Data Samples

The information regarding the used Data samples can be found in the table provided below.

Processgroup	**Datasetpath**
SingleMuon	/SingleMuon/Run2018A-UL2018_MiniAODv2_GT36-v2/MINIAOD
	/SingleMuon/Run2018B-UL2018_MiniAODv2_GT36-v2/MINIAOD
	/SingleMuon/Run2018C-UL2018_MiniAODv2_GT36-v3/MINIAOD
	/SingleMuon/Run2018D-UL2018_MiniAODv2_GT36-v2/MINIAOD
DoubleMuon	/DoubleMuon/Run2018A-UL2018_MiniAODv2_GT36-v1/MINIAOD
	/DoubleMuon/Run2018B-UL2018_MiniAODv2_GT36-v1/MINIAOD
	/DoubleMuon/Run2018C-UL2018_MiniAODv2_GT36-v1/MINIAOD
	/DoubleMuon/Run2018D-UL2018_MiniAODv2_GT36-v1/MINIAOD
EGamma	/EGamma/Run2018A-UL2018_MiniAODv2_GT36-v1/MINIAOD
	/EGamma/Run2018B-UL2018_MiniAODv2_GT36-v1/MINIAOD
	/EGamma/Run2018C-UL2018_MiniAODv2_GT36-v1/MINIAOD
	/EGamma/Run2018D-UL2018_MiniAODv2_GT36-v2/MINIAOD

A.3. Additional material for some interesting event classes

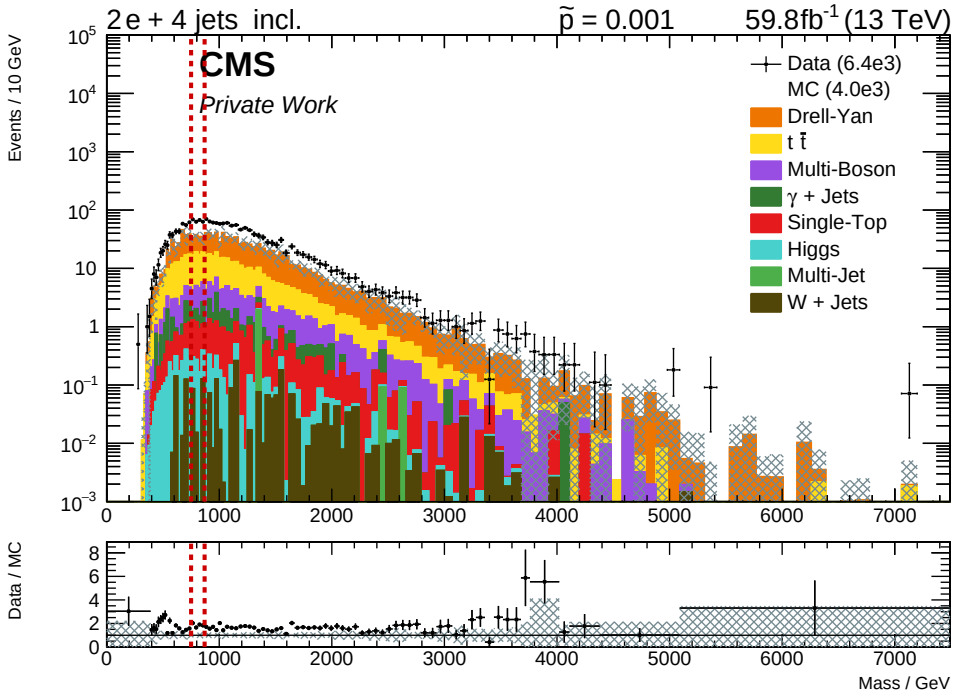


Fig. A.1.: Distribution of the $2e$+4Jet M_{inv} inclusive event class, with the Region of Interest demarcated by a red dashed line.

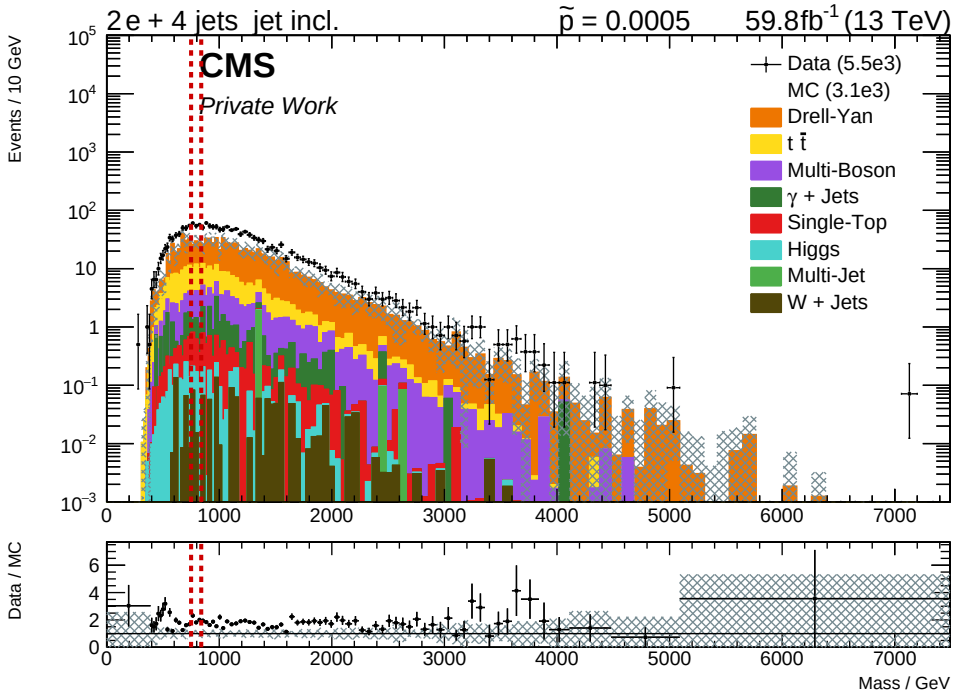


Fig. A.2.: Distribution of the $2e$+4Jet M_{inv} jet-inclusive event class, with the Region of Interest demarcated by a red dashed line.

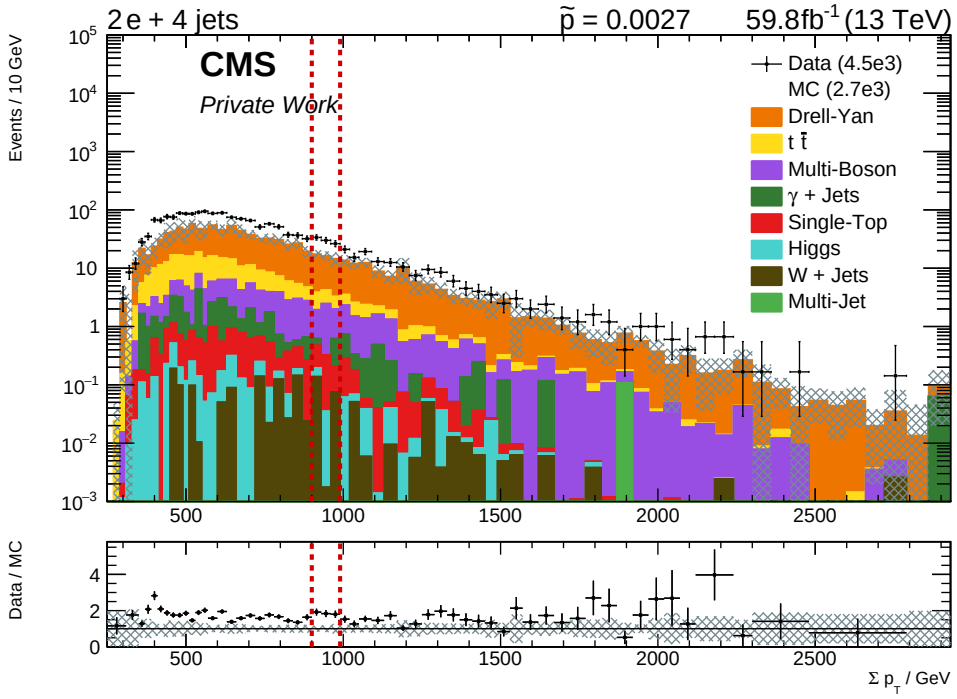


Fig. A.3.: Distribution of the $2e$+4Jet S_T exclusive event class, with the Region of Interest demarcated by a red dashed line.

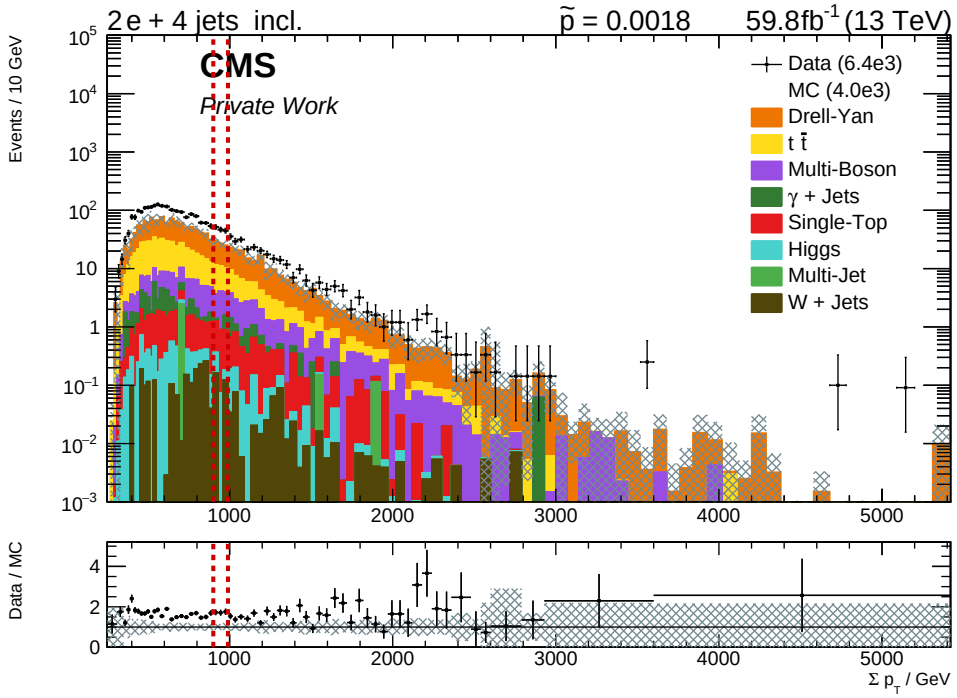


Fig. A.4.: Distribution of the $2e$+4Jet S_T inclusive event class, with the Region of Interest demarcated by a red dashed line.

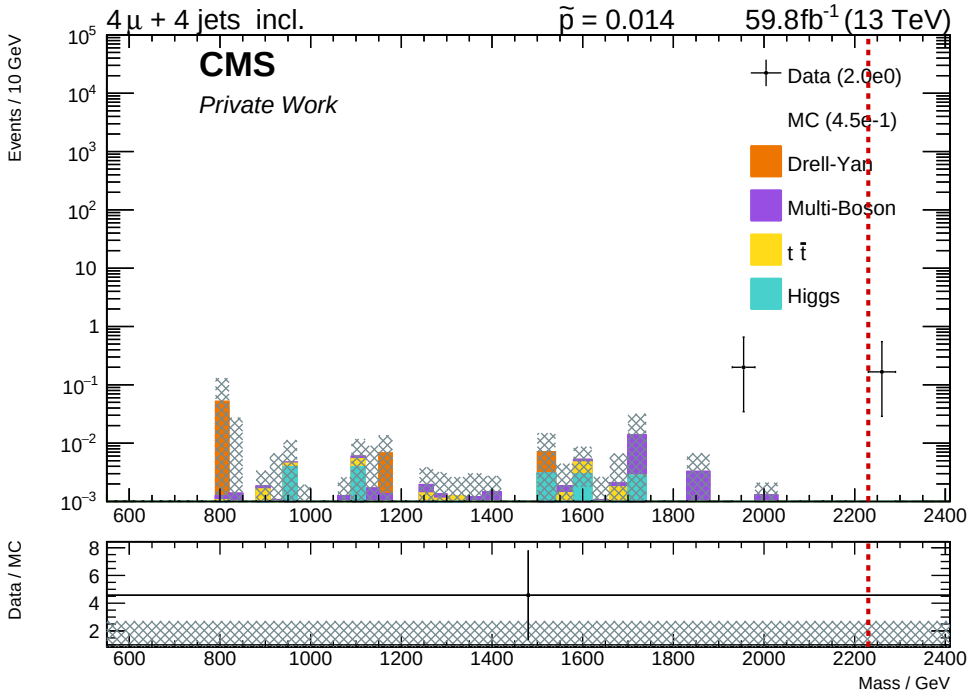


Fig. A.5.: Distribution of the 4μ+4Jet M_{inv} inclusive event class, with the Region of Interest demarcated by a red dashed line.

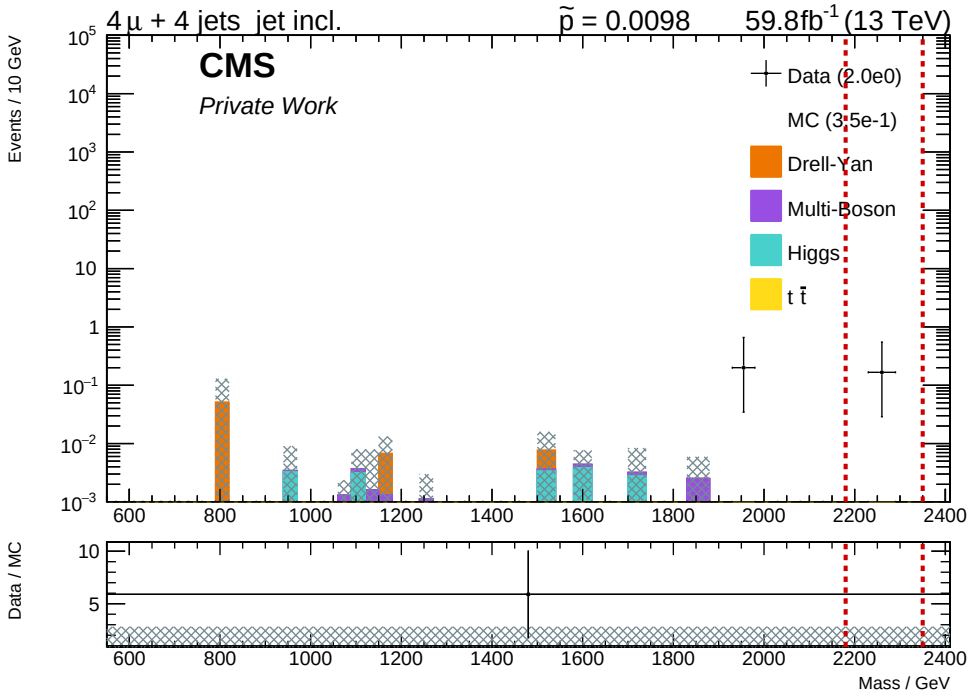


Fig. A.6.: Distribution of the 4μ+4Jet M_{inv} jet-inclusive event class, with the Region of Interest demarcated by a red dashed line.

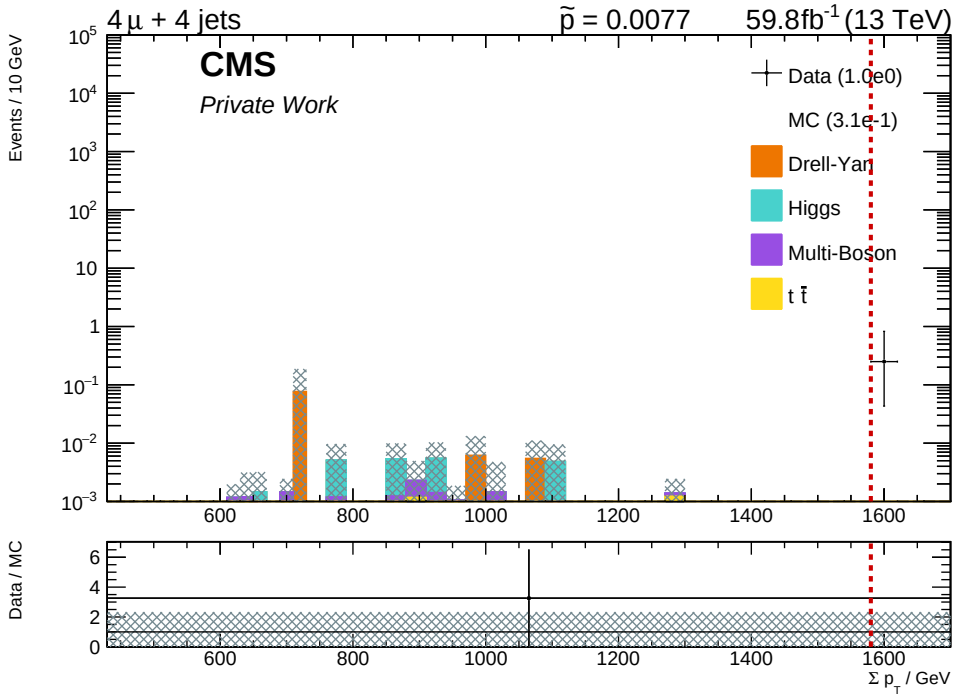


Fig. A.7.: Distribution of the 4μ+4Jet S_T exclusive event class, with the Region of Interest demarcated by a red dashed line.

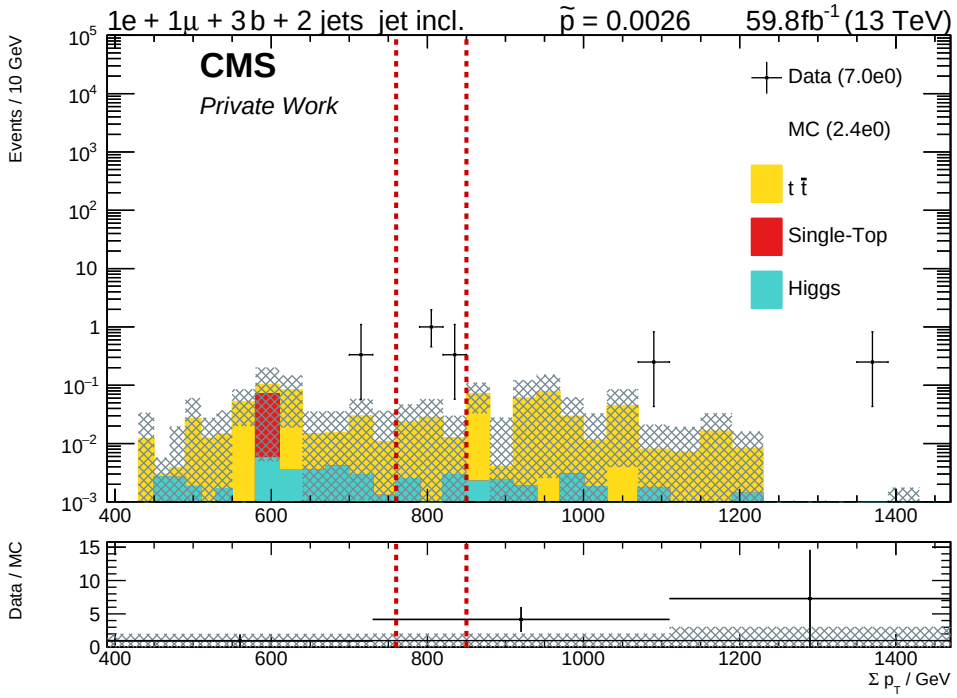


Fig. A.8.: Distribution of the $1e$+1μ+3bJet+2Jet S_T jet-inclusive event class, with the Region of Interest demarcated by a red dashed line.

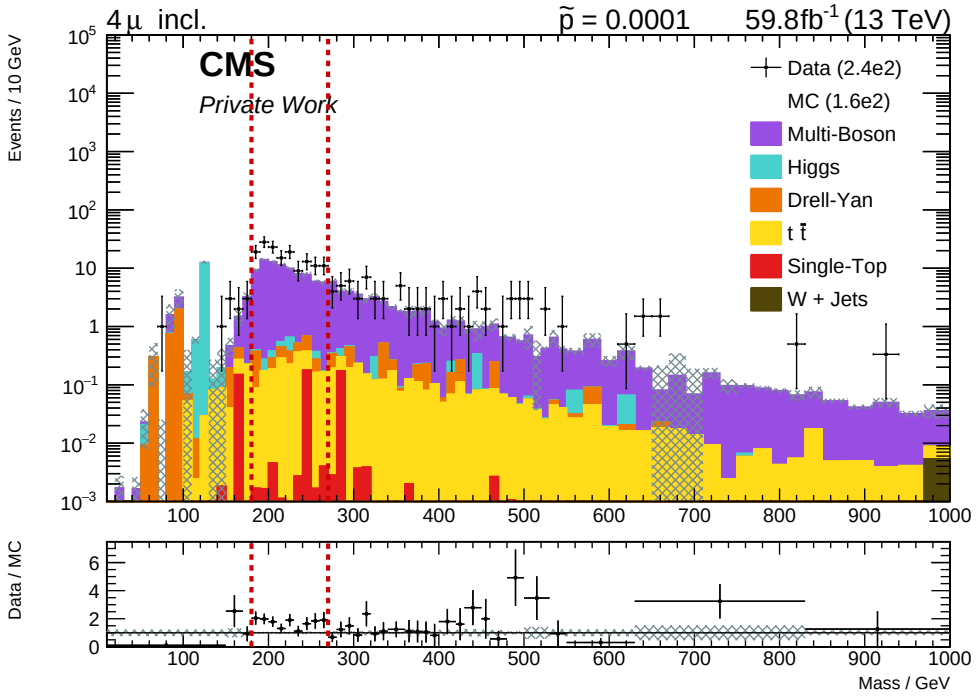


Fig. A.9.: Distribution of the 4μ M_{inv} inclusive event class, with the Region of Interest demarcated by a red dashed line.

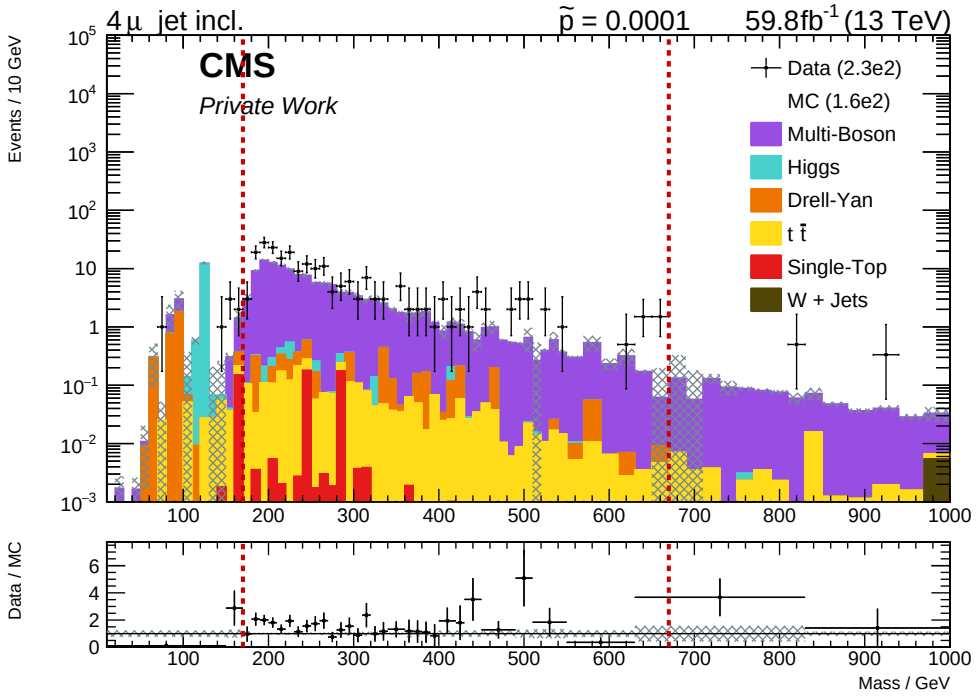


Fig. A.10.: Distribution of the 4μ M_{inv} jet-inclusive event class, with the Region of Interest demarcated by a red dashed line.

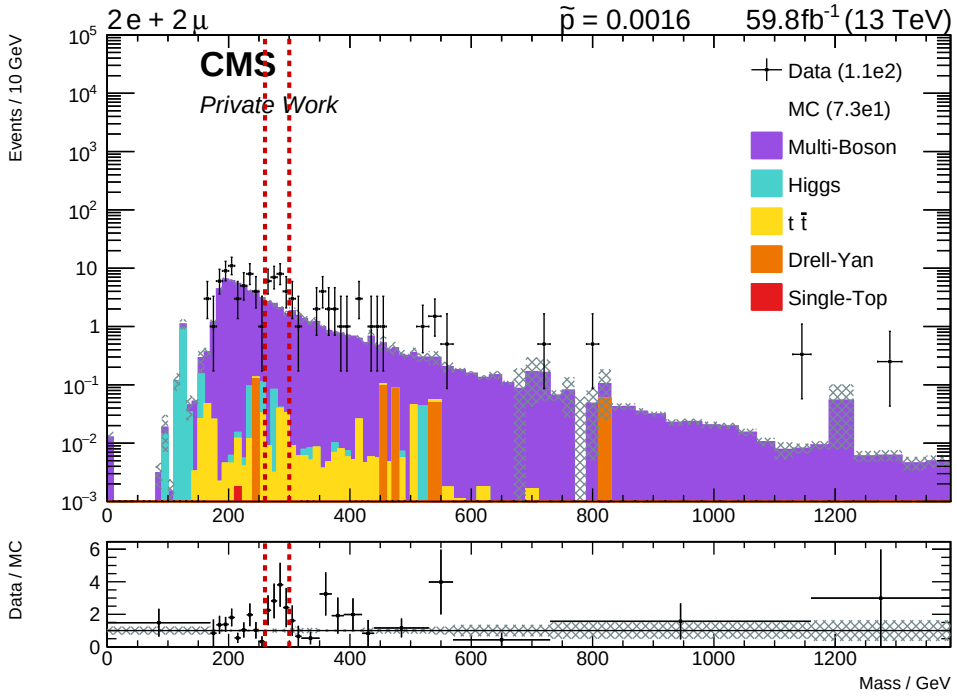


Fig. A.11.: Distribution of the $2e{+}2\mu$ M_{inv} exclusive event class, with the Region of Interest demarcated by a red dashed line.

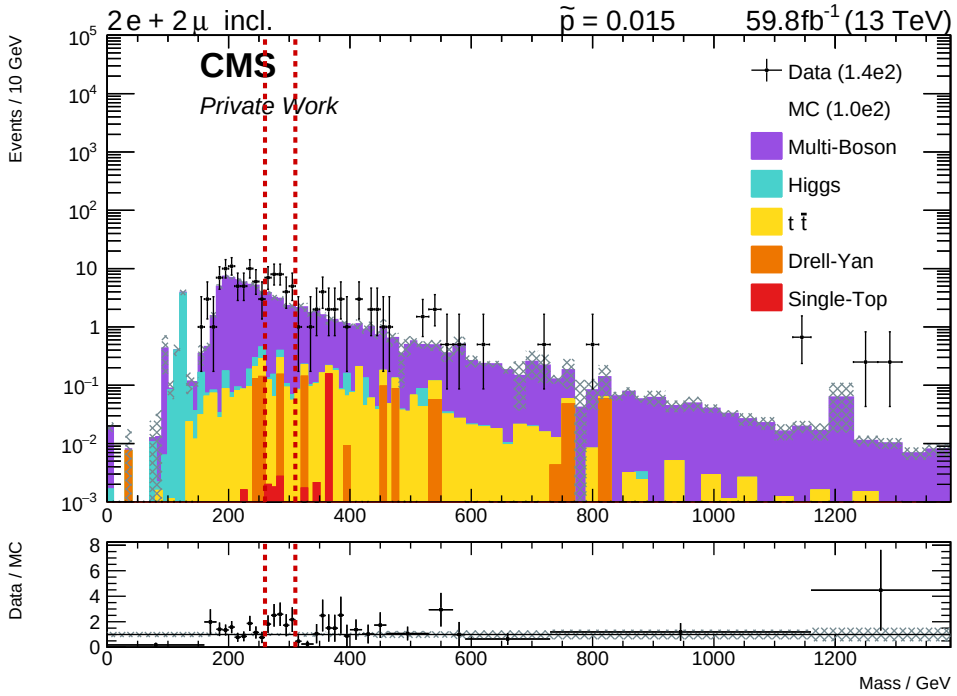


Fig. A.12.: Distribution of the $2e{+}2\mu$ M_{inv} inclusive event class, with the Region of Interest demarcated by a red dashed line.

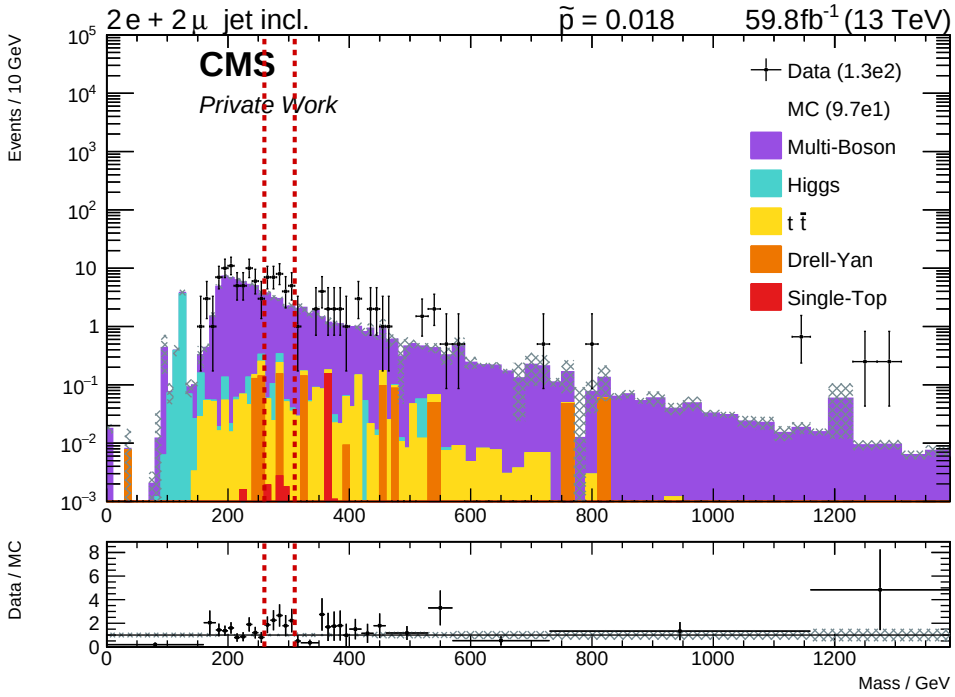


Fig. A.13.: Distribution of the $2e{+}2\mu$ M_{inv} jet-inclusive event class, with the Region of Interest demarcated by a red dashed line.

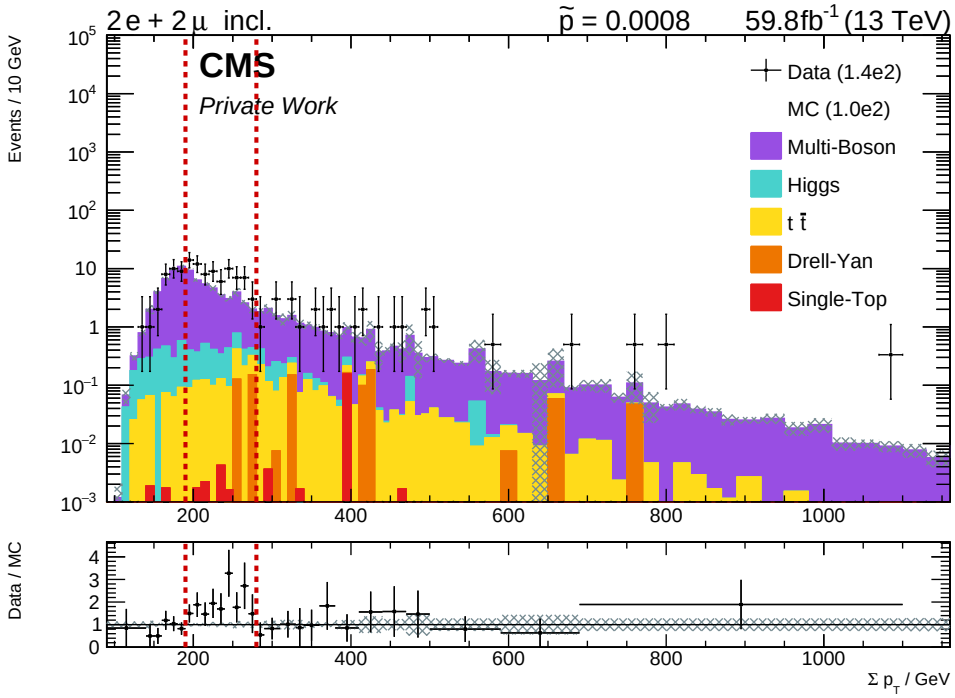


Fig. A.14.: Distribution of the $2e{+}2\mu$ S_T inclusive event class, with the Region of Interest demarcated by a red dashed line.

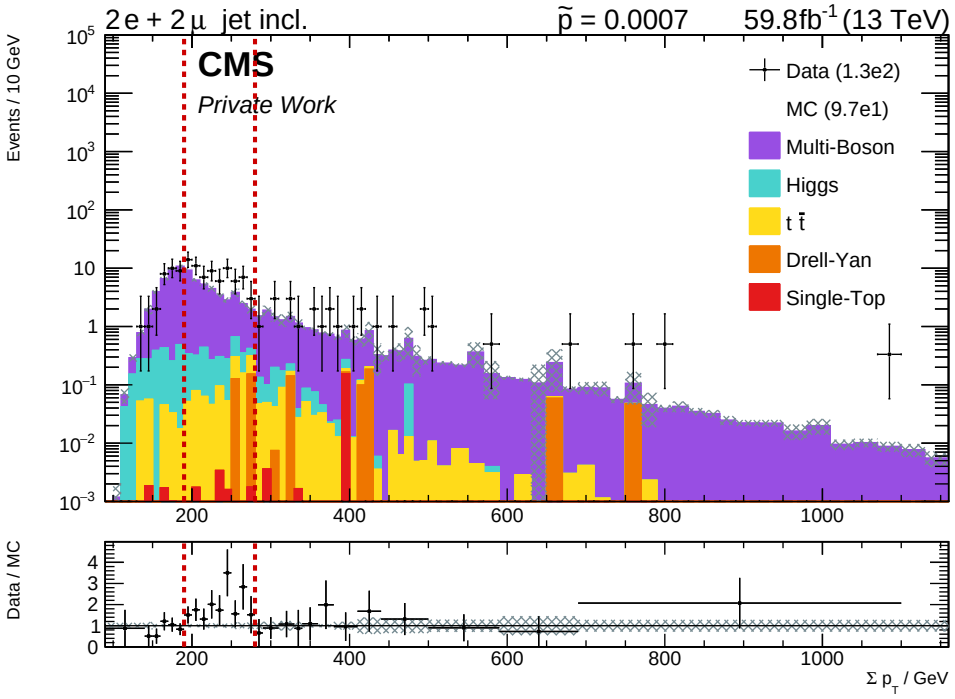


Fig. A.15.: Distribution of the $2e{+}2\mu$ S_T jet-inclusive event class, with the Region of Interest demarcated by a red dashed line.