Strange jets

Master's thesis
University of Turku
Theoretical physics
2024
B.Sc. Emilia Veikkola
Examiners:
Prof. Mikko Voutilainen
Dos. Iiro Vilja

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using Turnitin Originality Check service.

Funded/Co-funded by the European Union (ERC, JEC4HL-LHC, 101043975). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

UNIVERSITY OF TURKU
Department of Physics and Astronomy

Veikkola, Emilia Strange jets

Master's thesis, 109 pp.
Theoretical Physics
September 2024

This thesis presents an in-depth study of strange jets, with a focus on their properties and the determination of their jet energy scale (JES). Using data from the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC), this work addresses the challenges involved in identifying and analyzing strange jets, particularly in relation to other jet flavors. The study is organized into two main parts:

First, a Monte Carlo-based study examines strange jets' particle composition, angular distribution, and jet response. A significant finding is the higher proportion of neutral kaons in strange jets, particularly with high transverse momentum (p_T), which distinguishes them from other jet flavors. The study also reveals that the leading particles in strange jets carry a substantial fraction of the jet's total energy.

Second, a data-based analysis explores the determination of the jet energy scale for strange jets relative to light-flavor jets. Event selection is performed through charm tag, which is based on the theoretical assumption that in W boson decays, the presence of a charm jet indicates a likely accompanying strange jet. Scale factors are derived to verify the differing rates of up/down and charm/strange pairs. By combining multiple observables, such as mass and p_T difference, the strange jet JES is quantitatively estimated. The analysis also investigates potential sources of bias, such as helicity effects and p_T-dependent charm tagging efficiencies.

This work provides a proof-of-principle for determining the strange jet energy scale and outlines possible future directions for strange jet tagging techniques. Furthermore, it demonstrates a clear difference in particle content between strange jets and up/down jets and provides a comprehensive literature review on strange jets.

Keywords: CERN, LHC, CMS, strange quark, strange jet, jet energy scale

Contents

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Professor Mikko Voutilainen for giving me the opportunity to join his research group. Mikko introduced me to experimental research in high-energy physics during the Helsinki Institute of Physics (HIP) summer student program two years ago, which inspired me to pursue a career in particle physics.

I am also thankful to Docent Iiro Vilja for being my co-supervisor and for his guidance in theoretical physics. I would like to extend my gratitude to everyone in the CMS Experiment group at the Helsinki Institute of Physics. Your support and assistance have been crucial to the completion of this thesis. In particular, I especially want to thank M.Sc. Mikael Myllymäki for his invaluable assistance and guidance.

Finally, I am endlessly grateful to my family and friends for their continuous support and unconditional love. I could not have achieved this without you. My eternal gratitude goes to my partner, Matias, for his unwavering support in all my endeavors. Your countless hours of assistance and encouragement have been invaluable. Thank you ♡.

1 Introduction

The Standard Model of particle physics represents the most detailed and reliable framework we currently have for describing the fundamental particles, such as quarks, leptons, and bosons, and the forces that govern their interactions, the electromagnetic, weak, and strong forces. Over several decades, this model has been rigorously tested through countless experiments, from particle accelerators to cosmic observations, and has consistently provided accurate predictions that match experimental results with remarkable precision.

Despite its success, the Standard Model is widely recognized as an incomplete theory. While it effectively describes the behavior of particles within a certain energy scale, it does not account for all aspects of the universe. For instance, it fails to incorporate gravity, and it does not explain the nature of dark matter or dark energy, which together make up a most of the universe's mass-energy content. Moreover, the Standard Model leaves unanswered questions, such as the matter-antimatter asymmetry and the exact mechanism behind electroweak symmetry breaking.

These limitations have motivated ongoing research efforts to push the boundaries of the Standard Model by testing its predictions and by searching for phenomena that it cannot explain. Such efforts include probing the conditions of the early universe, exploring the behavior of particles at extreme energies in particle colliders, and searching for new particles or forces that could extend or modify the current model. These endeavors are important for deepening our understanding of the universe and for potentially uncovering a more complete theory that unifies all known forces and particles under a single, coherent framework.

A key area of investigation within the Standard Model is the study of the strong force, one of the four fundamental forces that govern our universe, described by Quantum Chromodynamics (QCD). QCD explains how quarks and gluons, the components of protons and neutrons, interact through the strong force. Central to QCD

is color confinement, which ensures quarks and gluons are never found in isolation but always combine to form color-neutral particles. Jets, which are clusters of particles produced by high-energy collisions in particle physics experiments, provide a way to study these interactions. Jets are formed through complex processes of fragmentation and hadronization, where quarks and gluons break apart and recombine, adhering to the principle of color confinement, to create the streams of particles that we detect. By examining jets, it is possible to gain a deeper understanding of the strong force and the intricate dynamics of the particles involved.

However, studying jets presents significant challenges like the complexity of the processes that lead to jet formation. Another challenge is the precise measurement and reconstruction of jets in experimental data. Due to the high multiplicity of particles produced in collisions, distinguishing specific jets from other jets or other types of events, such as underlying event or pile-up effects, can be difficult.

This thesis aims to address these challenges by conducting a comprehensive study of strange jets, jets containing strange quarks, using data from the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) at CERN. The research presented here delves into the properties of strange jets and studies their jet energy scale, with a particular emphasis on using a charm tagger to identify strange jets in W-boson decays. The thesis includes an overview of the theoretical background in Chapter 2, the experimental setup in Chapter 3, the event reconstruction techniques Chapter 4, a comprehensive literature review in Chapter 5, methods used in this thesis in Chapter 6 and the detailed results of this analysis, focusing on the identification and the jet energy scale of strange jets are presented in Chapters 7 and 8. Finally, Chapter 9 concludes the findings and discusses their potential implications for future research. This thesis used ChatGPT-4 in accordance with University of Turku's AI guidelines[1], with details of the usage explained in Appendix E.

[1] https://intranet.utu.fi/fi/Yksikko/opintohallinto/Documents/Vararehtorin%20linjaus%20TekoÃďly%203_2023.pdf

2 Theoretical background

In this chapter, the fundamental concepts related to the constituents of the universe are introduced. The discussion begins with an examination of the conventions in high-energy physics, followed by a comprehensive overview of the Standard Model and Quantum Chromodynamics. The metastability of the electroweak vacuum and principles of jet formation are then explored. The chapter concludes with a section on particle strangeness, setting the stage for the experimental discussions that follow.

2.1 Conventions in high energy particle physics

A common practice in physics is to adopt units that are both meaningful and helpful to simplify calculations. In high-energy particle physics, standardized conventions are universally used to enhance transparent communication and ensure consistency across diverse experiments and analyses. These conventions include the use of standardized units of measurement, coordinate systems, and the representation of physical quantities.

2.1.1 Natural units

In high-energy particle physics, a widely used system is natural units, where the speed of light and the reduced Planck constant are set to unity, $c = \hbar = 1$. This selection leads to a unification of units in terms of energy. Energy values are specified in units of electronvolt (eV), which represents the energy gained by an electron when accelerated across a 1-volt potential difference. Energy, momentum, and mass can be quantified in units of eV, while time and length are measured in eV^{-1} and thus

$$[\text{length}]^{-1} = [\text{time}]^{-1} = [\text{energy}] = [\text{mass}] = [\text{momentum}] = \text{eV}, \tag{1}$$

where the electron volt is approximately $1.602 \cdot 10^{-19}$ J.

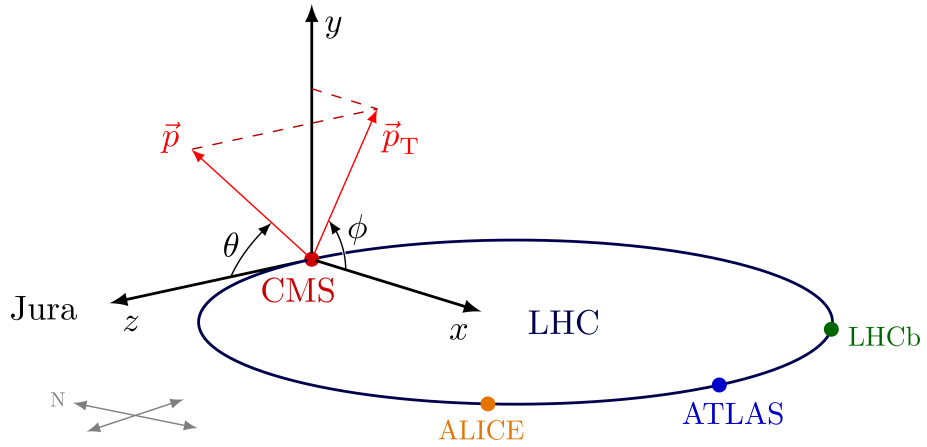


Figure 1: The CMS coordinate system [1].

The energies involved in high-energy particle collisions fall within the magnitudes of GeV and TeV, while the rest masses of observed elementary particles range from MeV to GeV. Consequently, in high-energy physics, it is common to treat the rest masses of particles as negligible compared to their energies.

2.1.2 Coordinate systems

In collider experiments, such as those at the Large Hadron Collider (LHC), a right-handed Cartesian coordinate system is used. In this system, the origin is set at the interaction point inside the center of the detector. The positive x-axis points towards the center of the LHC ring, the positive y-axis extends vertically upwards from the ring, and the positive z-axis is defined by the right-handed coordinate system to point in the anti-clockwise direction when viewed from above. This ensures consistency with the symmetry of the proton beams moving in opposite directions. The (x, y) plane, perpendicular to the beam, is referred to as the transverse plane. Figure 1 illustrates this coordinate system. The point labeled "IP" marks the interaction point where the beams intersect and protons collide.

In addition to the Cartesian coordinates, spherical polar coordinates (r, θ, ϕ) are also commonly used in high-energy physics to describe particle trajectories. In this

system, r represents the radial distance from the origin, θ is the polar angle relative to the z-axis, and ϕ denotes the azimuthal angle in the x-y plane.

Moreover, two other important variables are introduced: rapidity (y) and pseudorapidity (η). Rapidity is particularly useful because, unlike the polar angle θ, rapidity is Lorentz invariant under boosts along the z-axis. This property is advantageous in collider physics because the center-of-mass system of two colliding quarks or gluons inside a proton is often moving along the z-axis at high speed. Calculations can be performed in the center-of-mass coordinate system and then easily transferred to the detector frame using a Lorentz boost.

Rapidity is defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{2}$$

where p_z is the momentum along the z-axis and E is the energy.

Pseudorapidity (η), on the other hand, serves as a useful approximation for rapidity when the particle mass is negligible compared to its energy $m \ll E$, because it can be simply calculated from the polar angle θ alone. Pseudorapidity is defined as:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right). \tag{3}$$

When $m \ll E$ (or equivalently $m \ll |\vec{p}|$), the approximation $y \approx \eta$ applies.

2.1.3 Fundamental variables in high energy physics

In high-energy physics, the four-momentum vector is typically parameterized by the particle's energy E, the magnitude of its three-momentum $|\vec{p}|$, and the directional angles θ and ϕ in spherical coordinates. However, in high-energy collider physics, rapidity y or pseudorapidity η are often preferred over the polar angle θ due to

their Lorentz invariance properties. This leads to a four-momentum expressed as $p^{\mu} = (E, p_T, \eta, \phi)$, where p_T is the transverse momentum.

Additionally, the angular separation between particles or jets in the ϕ-η plane is measured using the distance ΔR, defined as:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}.$$

It is worth noting that ΔR can be calculated using either pseudorapidity η or rapidity y, depending on the context. In many analyses, pseudorapidity is used for simplicity, as it is easier to calculate directly from the geometry of the detector. However, in more precise jet clustering algorithms, such as those used in FastJet (e.g., anti-k_T), the more exact rapidity y is often employed instead of η for calculating ΔR.

In the context of deep inelastic scattering (DIS), an important variable is the Bjorken scaling variable x, commonly referred to as Bjorken x. This dimensionless quantity represents the fraction of the momentum of a hadron carried by one of its constituent partons (such as a quark or gluon) and is defined as

$$x = \frac{Q^2}{2p \cdot q}, \tag{4}$$

where Q^2 is the negative square of the four-momentum transfer, p is the four-momentum of the hadron, and q is the four-momentum transfer in the scattering process.

2.2 The Standard Model of Particle Physics

The *Standard Model* of particle physics (SM) stands as the cornerstone of our current understanding of the universe, offering a precise and foundational theory that describes the characteristics of known elementary particles and their interactions. These interactions are described by three fundamental forces: strong, weak, and electromagnetic forces. While the Standard Model has been validated through ex-

periments, it is not without limitations, which motivates the exploration of new avenues in physics.

2.2.1 Fundamental particles and interactions

The SM categorizes all known elementary particles into two main groups: fermions and bosons. Fermions are the building blocks of matter, while bosons mediate the fundamental forces. Fermions, which have a spin of $\frac{1}{2}$, are divided into quarks and leptons, each coming in three generations: quarks: up (u), down (d); charm (c), strange (s); top (t), bottom (b); and leptons: electron (e), electron neutrino (ν_e); muon (μ), muon neutrino (ν_μ); tau (τ), tau neutrino (ν_τ). Quarks combine to form hadrons, such as protons and neutrons, through the strong interaction. Leptons, however, do not interact via the strong interaction.

Bosons are further divided into vector and scalar bosons. Vector bosons, characterized by spin-1, are the force carriers that mediate the fundamental interactions: the photon (γ) mediates the electromagnetic force, the $W^{\pm}$ and Z^0 bosons mediate the weak force, and the gluon (g) mediates the strong force. Scalar bosons, on the other hand, are represented by the single spin-0 Higgs boson (H), which is responsible for giving mass to other elementary particles through the Higgs mechanism. This does not necessarily apply to neutrinos, which may acquire their mass through mechanisms such as the see-saw mechanism. For composite particles like protons, only a small fraction (approximately 0.5%) of their mass originates from the Higgs mechanism, with the majority of their mass arising from the dynamics of the strong interaction. The SM encompasses all observed elementary particles and their interactions, excluding gravity. The SM particles are presented in Figure 2.

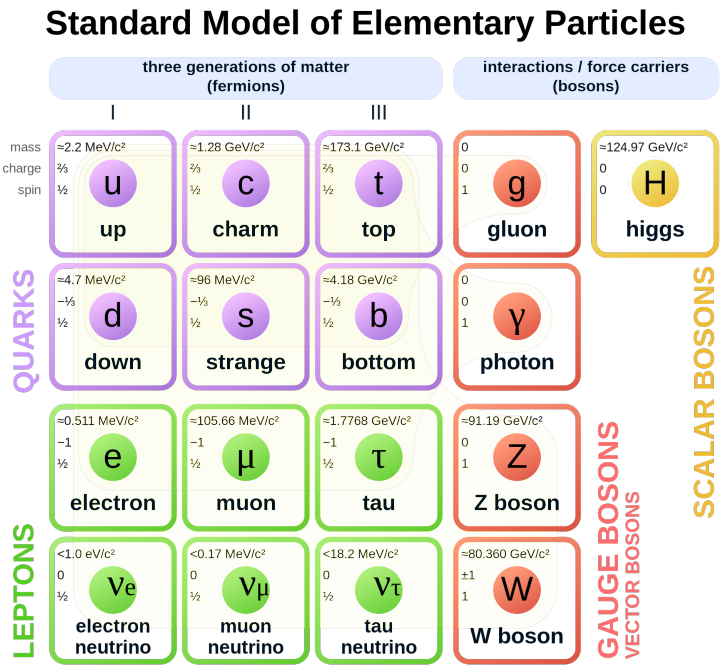


Figure 2: The complete set of elementary particles in the Standard Model, along with their respective mass, charge, and spin characteristics [2].

2.2.2 Symmetries and conservation laws

The SM has several symmetries, which give rise to various conservation laws that dictate the behavior of particles and interactions.

The gauge symmetries of the SM are described by the group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The $SU(3)_C$ component corresponds to Quantum Chromodynamics (QCD), the theory of the strong interaction, which acts on quarks and gluons through the exchange of color charge. The $SU(2)_L$ and $U(1)_Y$ components govern the electroweak interaction, where $SU(2)_L$ relates to the weak isospin, and $U(1)_Y$ corresponds to the hypercharge. After electroweak symmetry breaking, $SU(2)_L$ and $U(1)_Y$ combine to form the electromagnetic interaction governed by $U(1)_{\text{em}}$. These gauge symmetries are local, meaning that the transformations can vary at different points in spacetime.

An important symmetry is the conservation of flavor in strong interactions. In QCD, the strong interaction conserves quark flavors, meaning that for any quark

produced, a corresponding antiquark of the same flavor must also be produced. For example, if a strange quark (s) is produced, an antistrange quark ($\bar{s}$) is also produced. Even if a strange quark originates from the parton distribution functions (PDFs) of the proton, the net antistrange quark will still be present in the proton remnants. This principle of flavor conservation is a key motivation behind the Pythia string model [3]. The string model treats the formation of hadrons as the result of stretching color strings between quark-antiquark pairs, ensuring that flavor conservation is maintained throughout the hadronization process.

However, the weak interaction, mediated by the W boson, breaks flavor symmetry. While the weak interaction still conserves quark and lepton numbers separately, it allows for processes that change quark flavors. For example, the decay processes $W^+ \to c\bar{s}$ and $W^- \to \bar{c}s$ demonstrate the violation of flavor symmetry by the weak interaction.

In addition, the SM incorporates several symmetries that give rise to important conservation laws, such as baryon number conservation in particle interactions. However, baryon number is not an exact symmetry, and early universe processes like sphalerons in non-perturbative QCD could violate this conservation. Sphalerons were particularly important during possible first-order phase transitions as the universe cooled, where their ability to violate baryon number conservation, combined with CP violation, potentially explains the observed matter-antimatter asymmetry.

Both the weak and strong interactions conserve baryon number, which can result in the transfer of baryon number from a proton to the resulting jets during particle interactions. In strange jets, the leading particle is more likely to be an s-baryon, particularly at lower transverse momenta (p_T), where this baryon number transfer occurs more easily.

2.2.3 Theoretical framework

The Standard Model (SM) describes the interactions among elementary particles through three fundamental forces: the electromagnetic, weak, and strong forces. These interactions are governed by quantum field theories—Quantum Electrodynamics (QED) for the electromagnetic force, Quantum Chromodynamics (QCD) for the strong force, and the Electroweak Theory, which unifies the electromagnetic and weak forces under a common framework. This Chapter is mostly based on References [4–6].

This theoretical framework is articulated through the language of Quantum Field Theory (QFT), specifically as a gauge theory based on the symmetry group:

$$SU(3)_C \times SU(2)_L \times U(1)_Y,$$

which ensures local gauge invariance for each of these interactions. The corresponding gauge bosons—gluons for the strong force, $W^{\pm}$ and Z^0 bosons for the weak force, and the photon for the electromagnetic force—mediate these fundamental interactions.

Electroweak symmetry breaking and Higgs mechanism

Electroweak symmetry breaking (EWSB) is a process in which the Higgs field acquires a non-zero vacuum expectation value, reducing the $SU(2)_L \times U(1)_Y$ symmetry to the $U(1)_{\text{em}}$ symmetry of electromagnetism. This symmetry breaking gives mass to the W and Z bosons while keeping the photon massless. The Yukawa interactions, facilitated by this mechanism, also provide mass to fermions, making EWSB crucial to the SM's structure.

Quantum Electrodynamics

Quantum Electrodynamics (QED) describes electromagnetic interactions within the SM, based on the $U(1)_{\text{em}}$ symmetry. In QED, charged particles interact through the

exchange of photons, which are the gauge bosons of the electromagnetic force.

Quantum Chromodynamics

Quantum Chromodynamics (QCD) describes the strong interaction, responsible for binding quarks and gluons together within hadrons, such as protons and neutrons. QCD is a non-Abelian gauge theory based on the symmetry group $SU(3)_C$, where quarks carry color charges (red, green, blue) and gluons, the force carriers, facilitate the exchange of these charges. Two key properties of QCD are color confinement and asymptotic freedom.

Color confinement

Color confinement ensures that quarks and gluons are never found in isolation but are always confined within color-neutral hadrons. As quarks are separated, the strong force, governed by the strong coupling constant α_s, increases, making it energetically favorable to produce quark-antiquark pairs rather than allowing quarks to exist freely. This phenomenon arises from the non-Abelian nature of QCD, where gluons themselves carry color charge and interact with each other.

Asymptotic freedom

Conversely, asymptotic freedom refers to the behavior of quarks and gluons at high energy scales or short distances, where α_s decreases, causing them to behave almost like free particles. This property is crucial in understanding high-energy processes, such as the production of quark-antiquark pairs in particle collisions.

2.2.4 Challenges and extensions

Despite its successes, the Standard Model (SM) faces significant challenges. It does not incorporate gravity, making it incompatible with General Relativity, and it fails

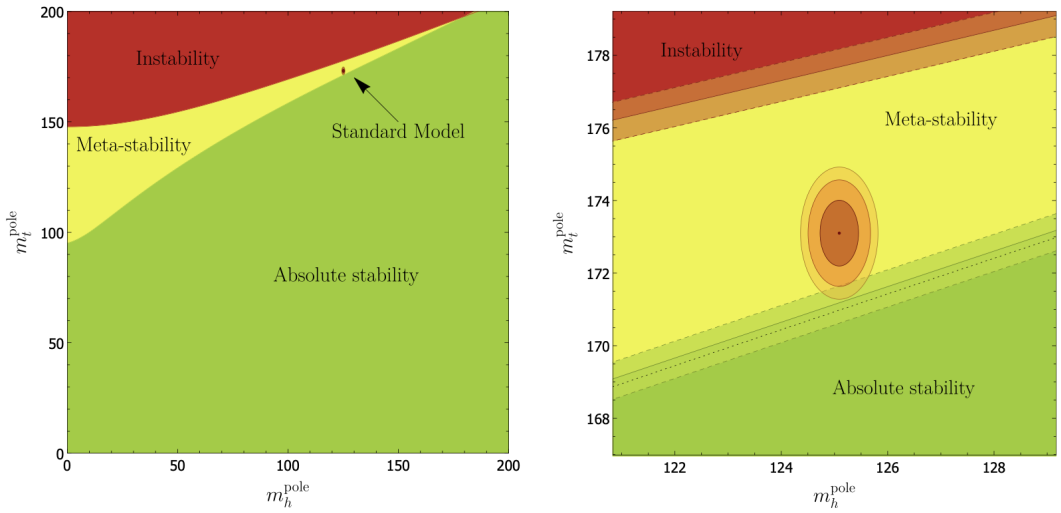


Figure 3: The stability of the EW vacuum as a function of the pole masses of the Higgs boson $m_{\rm h}^{\rm pole}$ and the top quark $m_{\rm t}^{\rm pole}$ [7].

to account for dark matter and dark energy, which make up most of the universe's mass-energy content. The SM also struggles with the hierarchy problem, the strong CP problem, and does not fully explain the origin of neutrino masses and oscillations.

Additionally, the matter-antimatter asymmetry observed in the universe remains unresolved within the SM. These issues highlight the need for new physics beyond the SM, leading to investigations into theories like supersymmetry, extra dimensions, and string theory.

2.3 Metastability of the electroweak vacuum

The electroweak vacuum, the ground state of the Higgs field, is crucial in the SM of particle physics. This vacuum is responsible for the Higgs mechanism, which gives mass to fundamental particles by breaking the electroweak symmetry. However, recent theoretical developments suggest that this vacuum may not be absolutely stable, but rather metastable [7].

Metastability implies that the current SM vacuum is not the lowest possible

energy state. While it is stable under normal conditions, there could exist a deeper, more stable vacuum state at higher energy levels. This scenario is largely dependent on the masses of the Higgs boson and the top quark, as these parameters influence the shape of the Higgs potential. The observed masses of the Higgs boson $m_{\text{Higgs}}^{\text{pole}} \approx 125$ GeV and the top quark $m_{\text{top}}^{\text{pole}} \approx 173$ GeV suggest that the electroweak vacuum is metastable, implying that our universe might exist in a metastable state [8, 9].

If the electroweak vacuum is metastable, there is a non-zero probability that it could transition to this true vacuum state through quantum tunneling. However, calculations suggest that such a transition is extremely unlikely to occur within the lifetime of the universe, making the vacuum effectively stable for all practical purposes [8].

2.4 Jet formation

Jets are collimated streams of particles produced in high-energy particle collisions, resulting from the hadronization of quarks and gluons. The formation of jets begins when a high-energy quark or gluon is produced in a particle collision. This quark or gluon then undergoes parton showering, a process where it radiates additional quarks and gluons through a series of emissions. During parton showering, further discussed in Chapter 3.3.1, which involves multiple branching processes, each parton splits into two or more lower-energy partons, creating a cascade.

Due to the property of color confinement, the partons produced in the showering process cannot exist as free particles. They undergo hadronization, a non-perturbative QCD process, where quarks and gluons combine to form hadrons, such as mesons and baryons. The resulting hadrons create a jet—a narrow cone of particles moving in roughly the same direction. An illustration of a jet is shown in Figure 4.

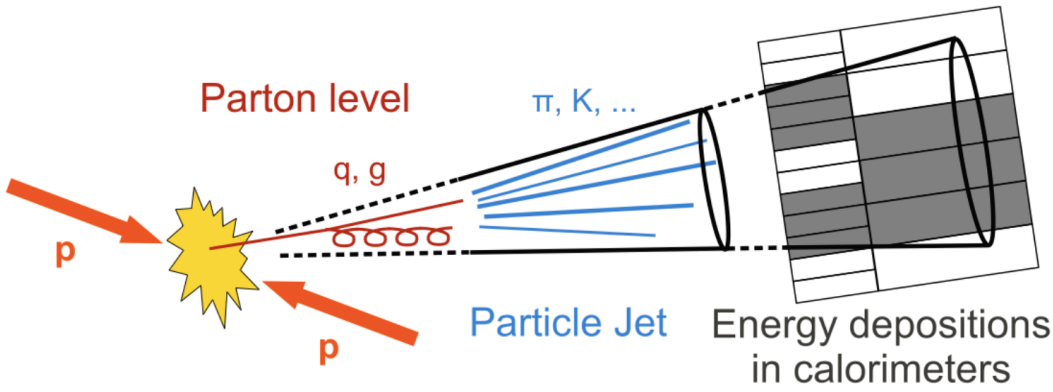


Figure 4: A sketch of jet formation. The red lines represent the final state quark or gluon, which form a hadron shower, the jet, shown as blue lines [10].

2.5 Strangeness

In 1953, Murray Gell-Mann [11] postulated the existence of an undiscovered property to explain why certain particles, like kaons and hyperons, had unusually long lifetimes compared to other particles. These particles were found to be produced via the strong interaction but decayed via the weak interaction. This proposed property was named *strangeness*.

Strangeness was ascribed numerical values of 0, +1, or -1 based on empirical observations. It maintains conservation across all reactions except those mediated by the weak interaction.

The concept of strangeness was integral to the development of the quark model [12, 13]. In this model, particles are composed of quarks, which carry fractional electric charges and specific quantum numbers, including strangeness. For instance, strange quarks (denoted as s) have a strangeness of -1, while anti-strange quarks ($\bar{s}$) have a strangeness of $+1$. All other elementary particles have a strangeness of $S = 0$, and the strangeness of composite particles is the sum of the strangeness of their constituent quarks.

Strangeness is particularly important in the study of hadrons. Mesons, which

are hadrons composed of a quark and an antiquark, and baryons, composed of three quarks, both exhibit strangeness. Examples of strange mesons include kaons (K), which contain either a strange quark or an anti-strange quark. Examples of strange baryons include the lambda (Λ) and sigma (Σ) particles, which contain one or more strange quarks.

The conservation of strangeness in strong and electromagnetic interactions, but not in weak interactions, is a key feature in understanding the behavior and lifetimes of various particles. For example, the relatively long lifetime of kaons compared to other mesons is due to the fact that their decay involves the weak interaction, where strangeness is not conserved [11].

3 Experimental research and simulation

This chapter begins with an overview of the Large Hadron Collider (LHC) and its pre-accelerators. It will be followed by a comprehensive review of the Compact Muon Solenoid (CMS) experiment, detailing the physical detector and its subdetector configurations. Finally, the reconstruction of physical objects and the simulation of collision events from the detector signals are explored. This chapter is mostly based on references [14–16].

3.1 The Large Hadron Collider at CERN

The Large Hadron Collider (LHC), spanning 26.7 kilometers in circumference, is currently the world's largest and most powerful particle accelerator. This two-ring superconducting hadron accelerator and collider is located about 100 meters underground, at the border between France and Switzerland near Geneva. The location of the LHC ring is depicted in Figure 5. Operated by the European Organization for Nuclear Research (CERN), the LHC aims to explore fundamental particle interactions by accelerating and colliding beams of protons and heavy ions at unprecedented energies. The tunnel housing the LHC was originally constructed in the 1980s for the CERN Large Electron-Positron Collider (LEP), which operated until 2000. It features eight straight sections and eight arcs, situated between 45 and 170 meters below the surface on a plane inclined at 1.4% towards Lake Léman. Since commencing operations in 2008, the LHC has operated in cycles, alternating between run periods—when particle collisions are recorded—and long shutdowns (LS) for maintenance and upgrades. The LHC is expected to operate until 2041, with its primary objective being to answer key questions about the nature of matter and the forces governing particle interactions, significantly contributing to our understanding of the universe.

The LHC is designed with specific interaction points, where particle beams inter-

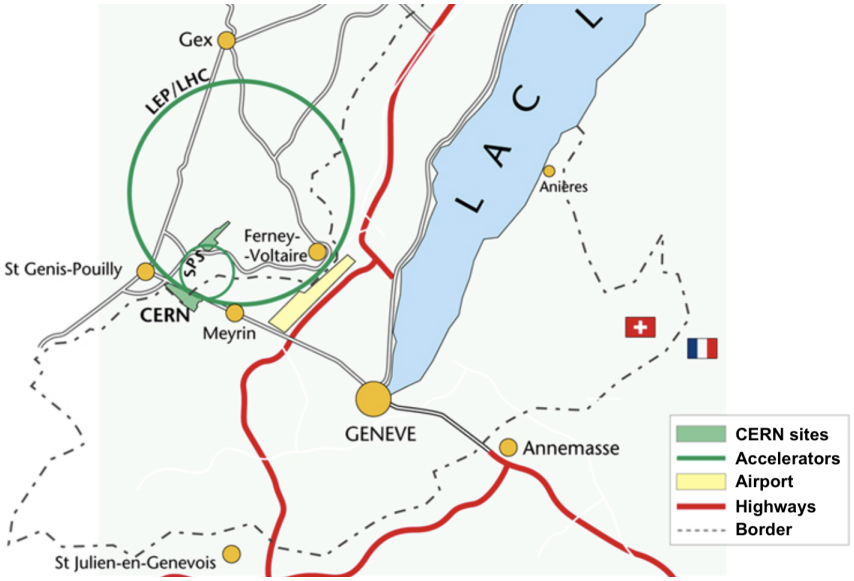


Figure 5: The Large Hadron Collider beneath the border between France and Switzerland near Geneva [17].

sect, while otherwise remaining isolated in two nearly adjacent beam pipes. An 8 T magnetic field directs the beams, which travel in opposite directions at 99.999999% of the speed of light. This magnetic field is produced by superconducting niobium-titanium (NbTi) coils, maintained at a temperature of 1.8 K using superfluid helium.

The process begins with the extraction of protons from hydrogen gas in a duoplasmatron, a metal cylinder into which the hydrogen gas is introduced. An electric discharge breaks the hydrogen molecules into atoms, which are then ionized to form protons. These protons are initially accelerated by the linear accelerator, LINAC 2, to an energy level of 50 MeV. The protons are then transferred to the Proton Synchrotron Booster (PSB), which increases their energy to 1.4 GeV. Next, the protons enter the Proton Synchrotron (PS), where their energy is boosted to 25 GeV. In the final stage before entering the LHC, the Super Proton Synchrotron (SPS) accelerates the protons to 450 GeV. Once the protons reach this energy, they are injected into the LHC ring, where they are further accelerated to their final collision energies: up to a center-of-mass energy of $\sqrt{s} = 7$ TeV in 2010 and 8 TeV in 2011-2012 during

Figure 6: The LHC and its surroundings. Illustration by Max Degtyarev [18].

Run 1, $\sqrt{s} = 13$ TeV in Run 2, and $\sqrt{s} = 13.6$ TeV in the ongoing Run 3. The integrated luminosities were 27.4 fb^{-1} for Run 1, 150.78 fb^{-1} for Run 2, and 158.17 fb^{-1} (so far) for Run 3 [19].

The protons inside the LHC are grouped into bunches to enhance the collision probability, with each bunch containing approximately 10^{11} protons. The bunches are accelerated using 16 superconducting radiofrequency (RF) cavities housed within cryomodules. These cavities generate oscillating electromagnetic fields that increase the energy of the protons to the desired operational level. The RF cavities control the spacing between the bunches, ensuring that they are separated by 25 ns intervals as they are directed around the circular accelerator ring. As the bunches travel within the accelerator, they move in opposite directions through the beam pipes, which are kept in an ultra-high vacuum environment. Despite only a small fraction of protons colliding at each interaction point, the large number of protons in each bunch and the high frequency of revolutions result in a significant number of collisions. These collisions, which occur at the interaction points, generate subatomic particles that are subsequently detected and analyzed for further study. The accumulation of

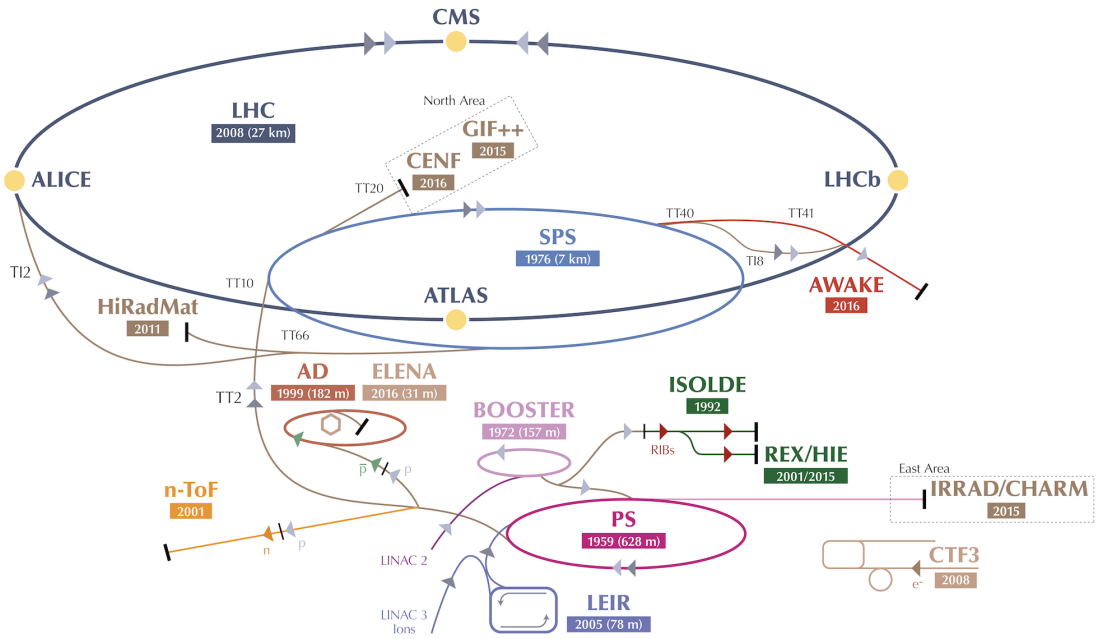


Figure 7: The LHC ring, its pre-accelerators and the main experiments. Trajectories of the accelerated particles illustrated. [20]

detectable proton-proton interactions during each bunch crossing is referred to as pileup.

The LHC hosts four major experiments, each designed to investigate different aspects of particle physics. The A Toroidal LHC ApparatuS (ATLAS) and the Compact Muon Solenoid (CMS) are general-purpose detectors that were pivotal in the discovery of the Higgs boson in 2012. The Large Hadron Collider beauty (LHCb) experiment focuses on understanding the differences between matter and antimatter by studying particles containing b quarks. The A Large Ion Collider Experiment (ALICE) is specialized in studying the properties of quark-gluon plasma, a state of matter thought to have existed just after the Big Bang. The LHC ring and the main experiments are illustrated in Figure 7.

3.2 The CMS experiment

The Compact Muon Solenoid (CMS) experiment is one of the primary experiments at the Large Hadron Collider (LHC) at CERN, located at Point 5. It uses the CMS detector to study particles resulting from proton-proton, lead-lead, and proton-lead collisions. This chapter provides a comprehensive overview of the CMS detector's components and their functions, alongside the analysis methods used to reconstruct collision events.

3.2.1 Detector design

The CMS detector is a cylindrical, multilayered system with a diameter of 15 meters and a length of 21.6 meters, designed to study particles from various collision types. The detector features several specialized components, including silicon trackers, an electromagnetic calorimeter (ECAL), a hadronic calorimeter (HCAL), and muon detectors, all housed within distinct regions known as the barrel and endcaps. Central to the CMS design is a superconducting solenoid magnet, which enables precise tracking and momentum measurement of particles across both the barrel and endcap regions. The design and components of the CMS are illustrated in Figures 8 and 9.

The barrel region, covering a pseudorapidity range of approximately $|\eta| < 1.3$, surrounds the interaction point and provides coverage along the central axis of the detector. The endcaps, positioned at both ends of the barrel, extend the coverage to the forward and backward regions with a pseudorapidity range of $1.3 < |\eta| < 3.0$. Furthermore, the hadronic forward (HF) detector, which covers $3.0 < |\eta| < 5.2$, plays a critical role in controlling additional jets and enhancing the precision of missing transverse energy (MET) measurements.

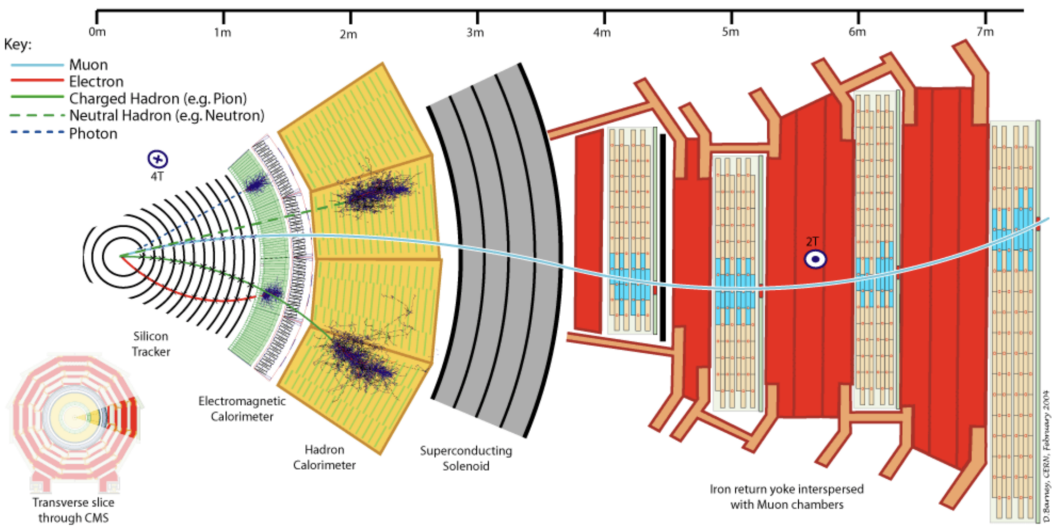


Figure 8: The transverse slice of the CMS detector with particle trajectories [21].

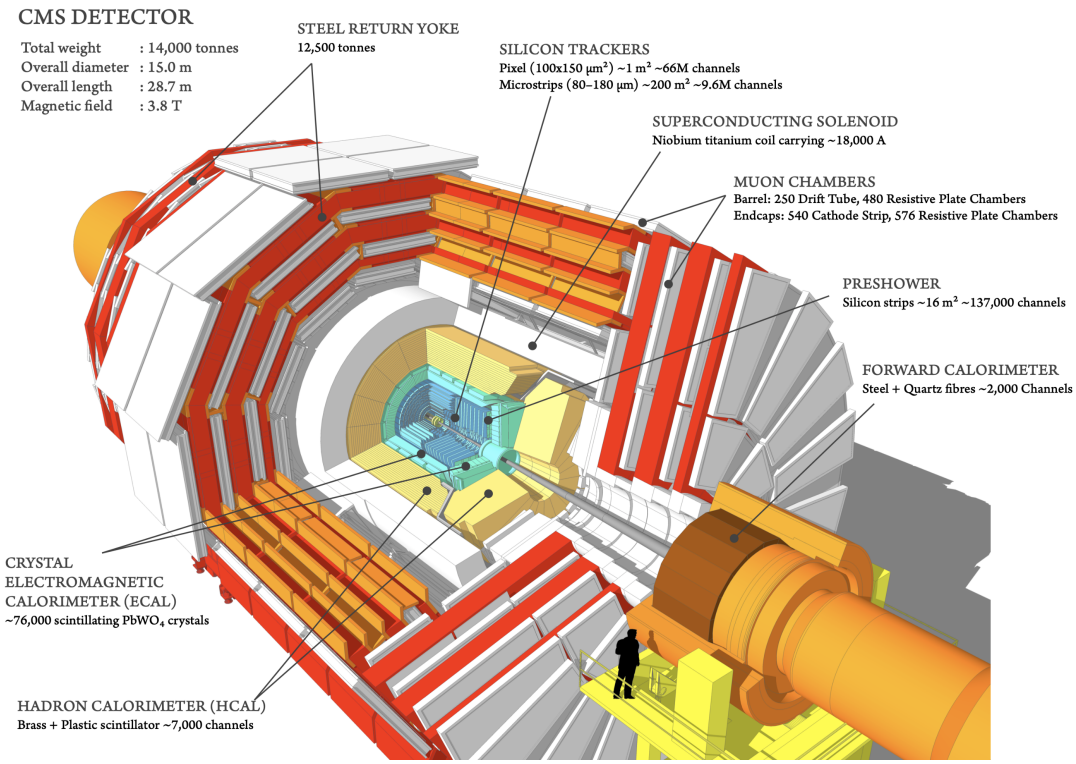


Figure 9: The different parts and layers of the CMS detector [22].

Superconducting solenoid magnet

At the heart of the CMS detector is a powerful superconducting solenoid magnet, which generates a magnetic field of 3.8 T. This magnetic field is crucial for bending the paths of charged particles, allowing for precise momentum measurements. The magnet's design ensures that the magnetic field is both strong and uniform, providing optimal conditions for particle tracking.

An important feature of the CMS magnet, particularly relevant for Particle Flow (PF) reconstruction, is its ability to effectively separate charged particles from neutral particles within dense jet cores. This capability significantly enhances particle identification and energy reconstruction, proving especially beneficial for strange particle tagging through the identification of neutral hadrons, such as K_L^0 and K_S^0. In contrast, the ATLAS detector, with its weaker magnetic field, is less effective in this regard.

Silicon tracker

The innermost component of the CMS detector is the silicon tracker, which consists of a series of pixel and strip detectors. The silicon tracker is responsible for precise vertex and track reconstruction, allowing for accurate determination of particle trajectories.

The innermost layer of the silicon tracker, known as the inner tracker, is positioned closest to the collision point. It detects charged particles using silicon pixel detectors and silicon strip detectors. When charged particles such as electrons, muons, pions, and protons traverse the silicon detectors, they ionize the silicon, creating electron-hole pairs. The resulting ionization signal allows for the precise reconstruction of particle trajectories. The inner tracker also measures the momentum of these charged particles by analyzing the curvature of their paths in the magnetic field generated by the solenoid.

Beyond the inner tracker, the outer tracker comprises additional layers of silicon strip detectors. These layers extend the tracking volume, providing further precision in measuring particle trajectories and enhancing the overall accuracy of the CMS detector. The outer tracker plays a crucial role in preserving the resolution of track reconstruction at larger radii from the collision point, which is particularly beneficial for identifying particles that decay later in their flight, such as K_S^0 mesons, which often decay within the outer tracker volume.

Electromagnetic Calorimeter (ECAL)

Surrounding the silicon tracker is the electromagnetic calorimeter (ECAL), which is designed primarily to measure the energy of photons and electrons. The ECAL is constructed from lead tungstate ($PbWO_4$) crystals, selected for their excellent energy resolution and fast response time. When photons or electrons enter the ECAL, they initiate electromagnetic showers, generating cascades of secondary particles. These particles interact with the crystals, producing scintillation light proportional to the energy of the initial particle. The combination of energy absorption and scintillation allows for precise energy measurements. This subdetector plays a vital role in identifying electrons and photons and measuring their energies with high precision.

A key feature of the ECAL is its high granularity, with each ECAL tower consisting of a 5x5 array of crystals. This granularity is especially beneficial for separating photons, such as those from $\pi^0 \to \gamma\gamma$ decays, from neutral hadrons like K_L^0. Furthermore, the ECAL is not only optimized for photons and electrons but also plays a role in measuring the energy of hadrons that shower early, with approximately two-thirds of hadrons depositing about half of their energy in the ECAL. This is particularly relevant for K_L^0 identification.

Hadronic Calorimeter (HCAL)

The hadronic calorimeter (HCAL) surrounds the ECAL and is responsible for detecting hadrons. The HCAL is composed of layers of dense absorber materials interspersed with scintillating tiles. As hadrons enter the HCAL, they interact with the atomic nuclei in the absorber material, generating hadronic showers. These interactions, which primarily consume energy rather than release it, result in a less efficient response compared to electromagnetic calorimeters. Due to the infrequent collisions with the small nuclear targets, the HCAL needs to be substantially thicker (over 2 meters, compared to 22 cm for the ECAL) to adequately capture these interactions. The secondary particles from these interactions produce scintillation light in the tiles, which is then detected and quantified.

Muon detectors

The outermost layer of the CMS detector comprises the muon chambers, which are essential for identifying and reconstructing muons. Muons, being minimally ionizing particles, can penetrate through the preceding layers of the detector. They are detected using gas-ionization detectors such as drift tubes, cathode strip chambers, and resistive plate chambers. As muons pass through these detectors, they ionize the gas, generating a signal that allows for their trajectory and momentum to be precisely measured. While muons are typically well-tracked using the inner pixel and silicon tracker, the muon chambers are crucial for identifying these particles and are particularly important for tracking the highest energy muons. Additionally, the muon system can be used to identify charm jets through semileptonic decays into muons, as demonstrated in [23]. This system complements the inner tracker data, ensuring accurate identification and momentum measurement of muons.

3.2.2 Trigger system

The CMS detector is equipped with an efficient trigger system designed to select potentially interesting collision events from the millions occurring every second due to 25 ns bunch crossing rate. It operates in two main stages: the Level-1 (L1) trigger and the High-Level Trigger (HLT). These stages work in tandem to reduce the data flow from the original 40 MHz to approximately 1 kHz for storage and detailed analysis. Given the high collision rates at the LHC, it is crucial to quickly and accurately identify events of interest while discarding less significant ones.

Level-1 (L1) Trigger

The Level-1 (L1) trigger is the first stage of the CMS trigger system, implemented using custom-designed hardware processors. Operating at the full collision rate of 40 MHz, the L1 trigger must make decisions within a few microseconds. It relies on coarse-grained data from the calorimeters and muon detectors to perform rapid event selection. The L1 trigger identifies high-energy particles, such as electrons, photons, muons, and jets, or large missing transverse energy, and forwards only the most promising events to the next stage. This process reduces the event rate from 40 MHz to approximately 100 kHz, achieving a reduction factor of more than 100.

High-Level Trigger (HLT)

Following the L1 trigger, the High-Level Trigger (HLT) further refines the event selection process using software algorithms. The HLT receives events at a rate of about 100 kHz and reduces this to around 1 kHz for permanent storage. Although the full event information is theoretically accessible at this stage, the reconstruction algorithms must operate much faster than those used offline. As a result, the HLT often performs only partial or local reconstruction, particularly for the tracker, while still applying sophisticated selection criteria to identify complex event topologies and

rare phenomena that might be missed by the L1 trigger.

3.3 Simulation

In high-energy physics, bridging theoretical predictions with experimental measurements is crucial. This connection is facilitated through the integrated processes of event generation and detector simulation, which enable the interpretation of vast amounts of complex data and accurately model detector responses. By simulating particle collisions and their interactions within detectors, precise calibration, alignment, and optimization of experimental setups can be achieved. Simulations are indispensable for identifying discrepancies that may indicate new physics beyond the Standard Model, estimating backgrounds and systematic uncertainties, and enhancing the sensitivity of searches for new phenomena. They also validate analysis methods before applying them to actual data, ensuring reliability and mitigating biases. Additionally, simulations are essential in designing and optimizing detectors, predicting how novel phenomena might appear in the data, and translating raw data into a deeper understanding of underlying physics processes.

In the context of the CMS experiment, simulation plays a crucial role by providing tools to understand detector performance, predicting collision outcomes, and interpreting data. This involves both event simulation, which models the physics of particle collisions, and detector simulation, which models the CMS detector's response to those particles. This Chapter provides an overview of these simulation processes and their importance in the CMS experiment. This Chapter is based on the detailed methodologies described in [24].

3.3.1 Event simulation

Event simulation is a significant part of the CMS experiment, allowing to generate theoretical predictions of collision events based on the Standard Model and various

Beyond Standard Model (BSM) theories. The process involves several stages:

Monte Carlo generators

Event simulation begins with Monte Carlo (MC) generators, which produce simulated collision events. Popular MC generators used in CMS include PYTHIA [25], HERWIG [26], SHERPA [27], and MADGRAPH [28]. These generators model the initial hard scattering process, parton showering, hadronization, and underlying events, including contributions from multiparton interactions (MPIs). MPIs are particularly relevant for understanding the production of additional charm and strange quarks, contributing to the overall complexity of the event.

Parton distribution functions (PDFs)

The initial state of colliding protons is described by Parton Distribution Functions (PDFs), which provide the probability distributions of quarks and gluons inside the protons [29]. PDFs are crucial inputs to the MC generators, affecting the kinematics and dynamics of the simulated events. The choice of PDFs can significantly influence the predicted cross-sections and event rates, thereby impacting the overall simulation results. In these PDFs, strange and charm quarks are parameterized as sea quarks, reflecting their presence in the proton due to quantum fluctuations. An important open question is the potential asymmetry between strange and anti-strange quarks, which could have significant implications for certain processes.

Hard scattering

The event simulation process begins with the hard scattering phase, which refers to the initial high-energy collision between partons (quarks and gluons) within the protons. This stage is modeled by Monte Carlo (MC) generators such as MADGRAPH [28], which calculate the matrix elements for the fundamental interactions based on

QCD and electroweak theory. The hard scattering process determines the primary products of the collision, setting the stage for subsequent particle interactions and decays.

Parton showering

Following the hard scattering event, parton showering occurs, which is a crucial stage in the event simulation process. During parton showering, the scattered partons emit additional radiation, resulting in a cascade of lower-energy partons. This stage involves two key processes: Initial State Radiation (ISR) and Final State Radiation (FSR).

ISR occurs before the hard scattering event, where the incoming partons (from the colliding protons) emit gluons as they approach the collision point. This alters the momentum of the partons involved in the hard scattering and must be accounted for in the simulation to accurately model the kinematics of the event. FSR, on the other hand, takes place after the hard scattering event, where the outgoing partons (produced by the hard scattering) emit additional gluons. These gluons can further split into quark-antiquark pairs, including strange and charm quarks in addition to the lighter up and down quarks. This splitting is an important mechanism for producing strange quarks, complementing their presence as sea quarks in the PDFs. However, the ratio of strange (s) to up/down (ud) splitting in the parton shower is not well understood, as it depends on quark masses, which are not well defined for colored particles.

Both ISR and FSR are modeled by generators like PYTHIA and HERWIG [30], which simulate the evolution of the parton system, including changes in momentum and angular distributions. The accurate simulation of ISR and FSR is essential for predicting the observable particle distributions and ensuring consistency with experimental data.

Hadronization

The partons produced in the showering process do not exist as free particles but instead undergo hadronization, where they combine to form color-neutral hadrons such as protons, neutrons, pions, and kaons. This process is modeled using phenomenological approaches like the Lund string model in PYTHIA [3] or the cluster model in HERWIG [31]. Hadronization is a non-perturbative process, meaning it cannot be directly calculated from first principles in QCD, but must instead be modeled based on experimental data. The resulting hadrons are the particles that are detected by the CMS detector.

Underlying event

In addition to the primary interactions, the underlying event encompasses all the additional soft interactions occurring in a proton-proton collision. These include multiple parton interactions (MPI), remnants from the colliding protons, and initial and final state radiation. The underlying event contributes to the overall event topology and must be accurately modeled to correctly simulate the energy flow and particle multiplicities observed in the CMS detector. The modeling of the underlying event is integrated into the MC simulation and is tuned using data from real collisions to ensure an accurate representation. For CMS during Run 2, Pythia tunes CUETP8M1 [32] and CP5 [32] were used, with CP5 continuing to be used in Run 3. These tunes are specifically optimized to improve the accuracy of the underlying event simulation.

Pile-up simulation

To accurately reflect real LHC conditions, pile-up simulation is incorporated. Pile-up refers to additional proton-proton interactions occurring in the same bunch crossing as the primary event of interest. On average, Run 2 data had about 20-30 inter-

actions per bunch crossing, while Run 3 data sees around 40-60 interactions. Each pile-up interaction typically produces about 2-3 particles, with approximately 70% of them reconstructed as vertices.

From a simulation perspective, pile-up is modeled using Pythia-simulated minimum bias events drawn from a large library of approximately one billion events. These events are processed using two main methods. Classical mixing involves adding individual pile-up events directly to the primary collision event during the simulation. This approach allows for high precision since the pile-up conditions can be closely matched to real data-taking scenarios, though it is computationally intensive as each event must be processed separately. Alternatively, premixing combines pile-up events into a single "premixed" event before adding it to the primary collision event, allowing for faster simulation at the expense of some precision. Both methods are crucial for accurately replicating the impact of pile-up on detector signals and event reconstruction [33].

3.3.2 Detector simulation

Detector simulation models how the CMS detector responds to the particles produced in collision events. It translates the physical interactions of particles with the detector materials into digital signals that can be analyzed. The key components of detector simulation include:

GEANT4

The primary tool for detector simulation in CMS is GEANT4 [34], a software framework that simulates the passage of particles through matter. GEANT4 models the complex geometry of the CMS detector, including its various subdetectors, and the interactions of particles with the detector materials, such as energy deposition and secondary particle production. This allows for a detailed and accurate reproduction

of the detector's response to different types of particles.

Digitization

The signals produced by particles in the detector are digitized to mimic the electronic readout of the actual CMS hardware. This process converts the continuous energy deposits into discrete digital signals, including hits in the silicon tracker, energy deposits in the calorimeters, and signals in the muon chambers. Digitization is a critical step in ensuring that simulated data can be directly compared with real experimental data.

Trigger simulation

Simulated events must pass through a virtual trigger system that accurately emulates the behavior of the CMS trigger. Unlike in real-time data collection, the trigger simulation does not remove events but allows for the evaluation of trigger efficiency and performance. This process is crucial for determining how well the trigger captures significant events and for understanding its behavior under various conditions. By accurately assessing and emulating the trigger's impact, the simulation ensures that the selection criteria used in data collection are reproduced, maintaining data integrity and volume management.

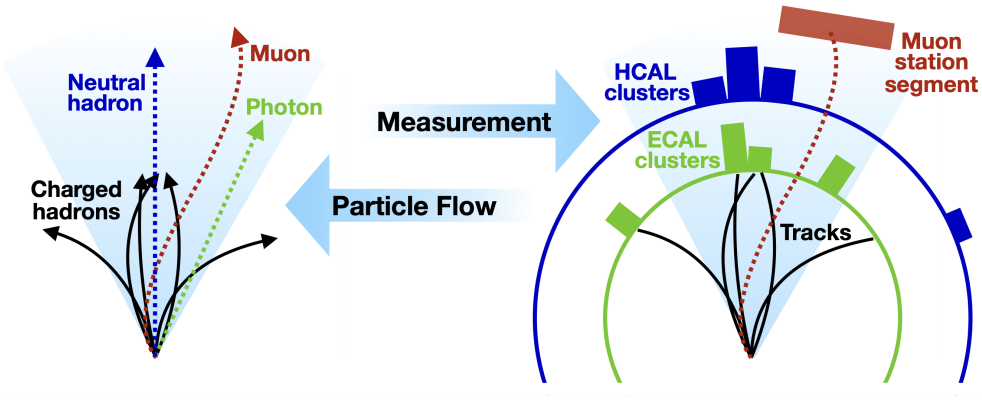


Figure 10: The PF algorithm converts detector signals into particle-level information. Adapted illustration by [37], originally from [38].

4 Event reconstruction

Event reconstruction in the CMS experiment refers to the process of reducing the raw data collected by the CMS detector during particle collisions to determine the characteristics and types of physics objects, such as particles and jets. The aim is to acquire a comprehensive representation of the physics of an event in a collision by distinguishing the types, energies, and trajectories of produced particles for further analysis. The reconstruction process in the CMS experiment consists of multiple stages, with Particle Flow (PF) reconstruction [35] being the final step.

4.1 Particle Flow reconstruction

Particle Flow (PF) reconstruction is an advanced method used in the CMS experiment to achieve reconstruction of all detectable particles produced in proton-proton collisions. By utilizing information from all sub-detectors, the PF algorithm aims to identify and reconstruct individual particles with high precision and efficiency, providing an accurate and detailed understanding of the physics of each collision event. This Chapter is based on References [35, 36].

PF reconstruction algorithm

The PF reconstruction method employs a PF algorithm to utilize the full spectrum of detector information. It relies on three critical components: efficient and pure track reconstruction, a clustering algorithm capable of distinguishing overlapping showers, and an effective linking procedure to connect the deposits of each particle in the sub-detectors.

The PF reconstruction process begins with the formation of energy clusters in the calorimeters, corresponding to the energy deposited by particles in the ECAL and HCAL. Simultaneously, tracks are reconstructed from hits in the silicon tracker, providing the trajectories and momenta of charged particles. The next step involves linking clusters and tracks to identify and classify particles. For example, a track in the silicon tracker pointing to an ECAL cluster is likely to be an electron or photon, while a track pointing to both ECAL and HCAL clusters is likely to be a charged hadron. The algorithm then merges redundant information, ensuring each particle is represented by a single set of measurements, thereby improving the accuracy of the reconstructed properties.

Additionally, the PF algorithm associates reconstructed particles with primary and secondary vertices to determine their origins, which is important for distinguishing particles from the primary collision from those resulting from secondary decays. This approach allows the PF algorithm to achieve increased precision in determining particles' energies, momenta, and trajectories. The particles and their properties reconstructed using the PF algorithm form a set of PF candidates, which serve as the basis for further analysis and interpretation. The PF algorithm is illustrated in Figure 10.

After event reconstruction, physics objects can be tagged to determine their origin, unwanted pileup interactions can be removed, and the objects can be calibrated to ensure an accurate energy scale, making them ready for further analysis.

Particle identification and reconstruction

The PF reconstruction algorithm generates a comprehensive list of candidate particles, classified into electrons, photons, muons, charged hadrons, and neutral hadrons.

Electrons

Electrons, being charged particles, are identified by their curved trajectories in the tracking system, caused by their interaction with the detector's magnetic field, and by the associated energy deposits in the ECAL. The tracks are linked to localized energy deposits in the ECAL, where electrons undergo bremsstrahlung and initiate electromagnetic showers. As they traverse the detector material, electrons often lose energy by emitting bremsstrahlung photons, which can be recovered by the PF algorithm.

Photons

Photons do not leave tracks in the tracking system because they are neutral particles. They deposit their energy in the ECAL through electromagnetic showers. The PF algorithm reconstructs photons by identifying isolated energy clusters in the ECAL without any associated tracks in the tracking system. In the case of a converted photon, where the photon converts into an electron-positron pair before reaching the calorimeter, the algorithm takes into account the tracks from the conversion to correctly identify the photon. A key challenge in photon reconstruction is the overlap when neutral pions (π^0) decay into two photons, which can cause merged electromagnetic showers.

Charged hadrons

Charged hadrons are reconstructed by analyzing their tracks in the silicon tracker, which are then combined with the corresponding energy deposits in the ECAL and

HCAL, with the majority of the energy being deposited in the HCAL due to hadronic interactions. The PF algorithm reconstructs charged hadrons by correlating the track information with the calorimeter energy deposits. The trajectories of the charged particles are determined by the orientation of the track at the vertex intersection point.

Neutral hadrons

Neutral hadrons are reconstructed primarily using calorimeter information, as they do not leave tracks in the silicon tracker. They deposit their energy predominantly in the HCAL through hadronic interactions, although some energy may be deposited in the ECAL if secondary particles, such as photons, are produced from neutral pion (π^0) decays. Neutral hadrons can also be identified by examining excess energy in the ECAL and HCAL clusters. When the total calibrated energy of these clusters exceeds the combined momenta of associated charged particles, this excess may indicate the presence of neutral hadrons. The PF algorithm identifies neutral hadrons by locating energy deposits in the HCAL that are not associated with any track in the tracker.

Muons

Muons are reconstructed by the PF algorithm using information from the silicon tracker and the muon chambers. In the tracker, muons behave like charged particles and leave defined tracks. However, unlike hadrons, they do not interact strongly with the ECAL or HCAL, resulting in minimal energy deposits in these subsystems. Thus, muons are identified by linking the distinctive hits in the muon chambers to the tracks in the silicon tracker.

Missing transverse momentum (MET)

Once all particles and their energies are reconstructed, the PF algorithm calculates the missing transverse energy (*missing transverse energy*) (MET). MET indicates the presence of undetected particles, such as neutrinos, which pass through the sub-detectors without interacting. It is calculated as the negative vector sum of the transverse momenta of all reconstructed particles.

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i}^{N_{\mathrm{particles}}} \vec{p}_{\mathrm{T},i}. \tag{5}$$

Advantages of PF reconstruction

By combining measurements from different sub-detectors, PF reconstruction enhances the precision and efficiency of particle identification, particularly in complex events with high particle multiplicities. This method significantly improves jet energy resolution by accurately accounting for both charged and neutral components. Additionally, PF reconstruction enhances the discrimination between signal and background, leading to more reliable identification of rare processes and new phenomena. Moreover, it effectively mitigates the impact of pileup by distinguishing particles from different vertices, ensuring cleaner event reconstruction.

4.2 Jet clustering

Jet clustering is a technique in high-energy physics, used to identify and reconstruct jets in collision events. These algorithms group particles based on their transverse momentum (p_T) and spatial proximity, which is essential for analyzing the event structure and probing the properties of the Standard Model and beyond. The clustering process relies on various distance measures and criteria to form jets, ensuring accurate reconstruction and interpretation of experimental data. Commonly used jet clustering algorithms include:

Anti-k_t algorithm

The anti-k_t algorithm [39] is a jet clustering method in particle physics, utilized in the CMS experiment at CERN. This algorithm clusters particles into jets by iteratively combining them based on their transverse momentum p_T and their relative distance in the pseudorapidity-azimuthal angle $(\eta - \phi)$ plane.

The algorithm operates by computing two types of distances for each particle:

1. The distance between pairs of particles:

$$d_{ij} = \min\left(k_{t,i}^{2p}, k_{t,j}^{2p}\right) \frac{\Delta_{ij}^2}{R^2}, \tag{6}$$

2. The distance to the beam axis:

$$d_{iB} = k_{t,i}^{2p}, \tag{7}$$

where

$$\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2. \tag{8}$$

Here, $k_{t,i}$ represents the transverse momentum, y_i is the rapidity, and ϕ_i is the azimuthal angle of particle i. For the anti-k_t algorithm, the parameter p, which varies depending on the specific algorithm used, is set to -1.

The clustering procedure involves calculating these distance metrics for all particles. The algorithm begins with the particle having the smallest distance, which could be either d_{ij} (distance to another particle or pseudojet) or d_{iB} (distance to the beam). If d_{ij} is the smallest distance, the two particles are merged into a pseudojet, replacing the original particles. If d_{iB} is the smallest distance, the particle is identified as a jet and removed from the list of particles. This process repeats until no particles remain to be clustered.

An advantage of the anti-k_t algorithm is its focus on hard particles, ensuring that the presence of soft particles does not significantly affect the jet's shape. This results in jets that are more stable and circular in the $\eta - \phi$ plane, facilitating easier interpretation and analysis.

k_t algorithm

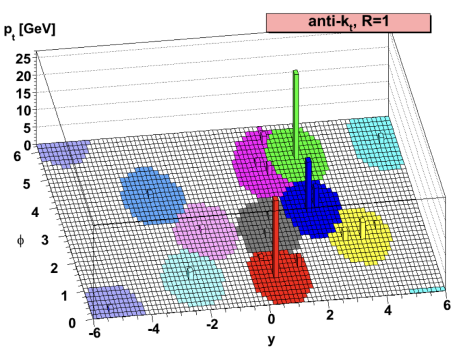
The k_t algorithm [40] is another widely used jet clustering method in particle physics. It utilizes the same distance calculations as the anti-k_t algorithm but differs by setting the exponent p to 1. This change in exponent results in a clustering process that prioritizes the combination of soft particles, leading to the formation of more irregular and softer jets compared to those produced by the anti-k_t algorithm. It is particularly useful in studies of heavy ion collisions, where soft particles are prevalent. However, in the high pile-up environment of the LHC, this characteristic makes the k_t algorithm less convenient, as it can lead to less stable jet shapes and increased sensitivity to soft radiation.

Cambridge/Aachen algorithm

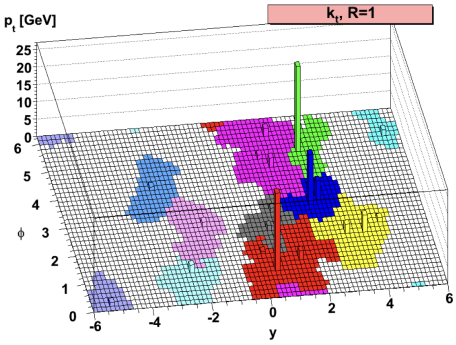
The Cambridge/Aachen algorithm [41] is a jet clustering method recognized for its simplicity and geometric approach. This algorithm clusters particles into jets based solely on their spatial proximity in the pseudorapidity-azimuthal angle $(\eta - \phi)$ plane, using the same distance formula as the k_t and anti-k_t algorithms, but with the exponent p set to 0.

The primary advantage of the Cambridge/Aachen algorithm is its emphasis on geometric distance, resulting in jets that naturally reflect the spatial structure of particle distributions in the $\eta - \phi$ plane. This makes the algorithm particularly useful in jet substructure studies, as it relies solely on angular separation without incorporating particle momentum into the clustering process.

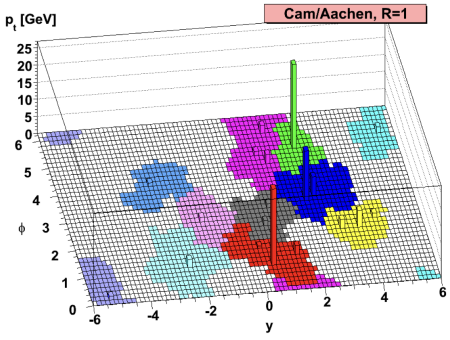
Different clustering algorithms are depicted in Figure 11. The upper left panel shows the anti-k_t algorithm, the upper right panel shows the k_t algorithm, and below them is the Cambridge/Aachen algorithm. Each panel represents the same event, generated by HERWIG [31], and clustered using the respective algorithm.



(a) anti-k_t algorithm



(b) k_t algorithm



(c) Cambridge/Aachen algorithm

Figure 11: Common clustering algorithms in high-energy physics, showing an event generated by Herwig and clustered using different algorithms.

4.3 Pile-up mitigation

Pileup mitigation is a major component of the event reconstruction in the CMS experiment, especially under high-luminosity conditions where multiple proton-proton collisions occur simultaneously within the same bunch crossing. Pileup refers to these overlapping interactions, which can affect the measurement accuracy of various physical quantities.

One widely used method is Charged Hadron Subtraction (CHS) [42]. CHS utilizes tracking data to identify particles originating from pileup after the formation of particle flow candidates and before the jet clustering process. This technique subtracts tracks of charged particle candidates associated with pileup (PU) vertices, leaving neutral particles and charged particle candidates not connected to any PU vertex unaltered. CHS improves the resolution of jets and missing transverse energy (E_T^{miss}) by excluding tracks from pileup interactions.

Pileup Per Particle Identification (PUPPI) [43] is another advanced technique. This algorithm reduces the impact of pileup on observables like jet substructure by utilizing local particle distribution, properties of event pileup, and tracking information. At the particle candidate stage, PUPPI assigns a weight from 0 to 1 to each particle, based on surrounding particles. Particles likely originating from the primary vertex receive a weight of 1. These weights are used to rescale the particle four-momenta, reducing the influence of pileup on observables. This comprehensive weighting process ensures improved accuracy and reliability of the reconstructed observables.

Effective pileup mitigation is essential for a wide range of physics analyses conducted at the CMS. Accurate pileup corrections enhance the precision of jet energy scale and resolution, improve the reconstruction of E_T^{miss}, and increase the overall fidelity of event reconstruction.

4.4 Jet energy calibration

At the CMS experiment, the Jet Energy Correction (JEC) [44] workflow ensures that the measured jet energies are corrected to their true values through a multi-step process. The primary goal of the JEC is to correct the jet energy scale (JES) in both data and simulation, ensuring that the reconstructed jet energy matches the true energy of the originating particles. Additionally, the Jet Energy Resolution (JER) is addressed, primarily in simulations, to ensure that the resolution of jet energies matches what is observed in the data. Together, JES and JER are integral to the JEC workflow, which aims to correct raw jet measurements and ensure reliable data analysis by minimizing systematic uncertainties and providing accurate representations of jet energies.

The JEC workflow begins with the raw reconstructed jets and systematically applies corrections to account for various effects, including pile-up contributions, measured transverse momentum, the jet's spatial position within the detector, and jet flavor. The ultimate goal is to calibrate the jets to the particle-level energy scale, ensuring an accurate representation of their true energy.

Understanding the JES is particularly important as it is a significant source of systematic uncertainties in analyses. JES is crucial for accurately interpreting the transition from particle-level interactions to reconstructed measurements. Accurate calibration and resolution of jets are fundamental for reliable data interpretation and for minimizing uncertainties in high-energy physics experiments. The JEC process at the CMS is depicted in Figure 12. This discussion is based on Reference [44].

4.4.1 Jet energy corrections

Jet energy corrections are essential for ensuring accurate jet measurements. These corrections account for various effects that can distort the measured jet energies, starting with pileup effects and extending to detailed adjustments based on simu-

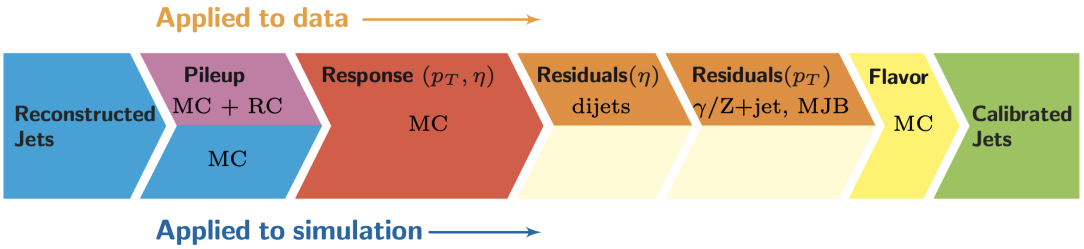


Figure 12: The JEC workflow [44].

lated responses and residual differences between data and simulations. Additionally, flavor-dependent corrections are applied to account for the different responses of the detector to jets originating from various quark flavors.

Pile-up offset corrections

Jet energy corrections start with adjusting for pileup effects, commonly referred to as *pileup offsets* or *L1 corrections*. These corrections address two types of pileup: in-time pileup (IT PU) and out-of-time pileup (OOT PU).

IT PU results from additional collisions within the same event, adding extra tracks and energy in the calorimeters. This type of pileup is addressed using algorithms like CHS and PUPPI [43, 45].

OOT PU, on the other hand, arises because signals from previous collisions can still be affecting the calorimeters while new collisions occur. This is handled by reducing the duration over which the signal is integrated and by using timing and pulse shape techniques to differentiate between signals occurring at the correct time (in-time) and those from previous collisions (out-of-time).

Simulated response corrections

After pileup offset corrections, jets undergo further corrections based on the simulated particle response R_{ptcl}. This response is defined as

$$R_{\text{ptcl}}(\langle p_T \rangle, \eta) = \frac{\langle p_T \rangle}{\langle p_{\text{T, ptcl}} \rangle}[p_{\text{T, ptcl}}, \eta] \quad [44], \tag{9}$$

where $\langle p_T \rangle$ represents the average transverse momentum of the reconstructed jet, $\langle p_{\text{T, ptcl}} \rangle$ is the average transverse momentum at the particle level, and the variables inside the square brackets denote the binning variables: particle-level transverse momentum ($p_{\text{T, ptcl}}$) and reconstructed pseudorapidity (η).

This particle response is then used to obtain simulated response corrections, known as L2L3 corrections. The response correction factor C_{L2L3} is derived from the condition

$$\langle p_{\text{T, ptcl}} \rangle = C_{L2L3} \langle p_{\text{T}} \rangle \quad [46], \tag{10}$$

where $\langle p_{\text{T, ptcl}} \rangle$ is the average particle-level transverse momentum, and $\langle p_{\text{T}} \rangle$ is the average reconstructed transverse momentum. Thus, the response correction factor C_{L2L3} is expressed as

$$C_{L2L3} = \frac{1}{\langle R_{\text{ptcl}}(p_{\text{gen}}, \eta) \rangle} \quad [46]. \tag{11}$$

The L2L3 corrections address the average energy discrepancy between reconstructed jets and particle-level jets. Simulated jet response is illustrated in Figure 13.

Residual corrections for data

Following the application of L2L3 corrections, additional residual corrections (L2L3Res) are applied to address any remaining discrepancies between the jet response observed in actual data and the response simulated by Monte Carlo. Initially, residual corrections are determined using dijet events with low statistical uncertainty, focusing on jet responses across a broad p_T range. To ensure comprehensive

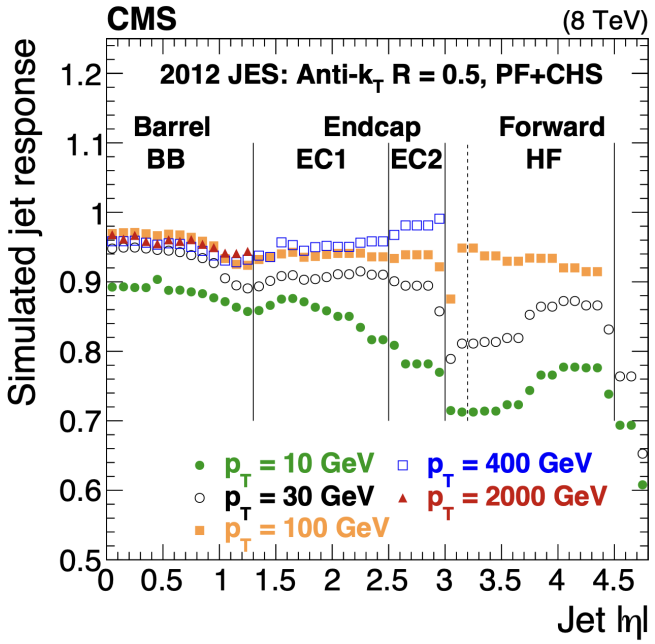


Figure 13: Simulated jet response as a function of pseudorapidity [46].

phase space coverage, a variety of event samples—including dijet, $Z(\rightarrow \mu\mu) + \text{jet}$, $Z(\rightarrow \text{ee}) + \text{jet}$, $\gamma + \text{jet}$, and multijet events—are employed. This diverse sample set is crucial for verifying the magnitude of corrections and for calibration against high-resolution reference objects, such as muons and photons. The barrel region ($|\eta| < 1.3$) is used as a reference for calibrating other regions, ensuring uniformity and consistency across the detector.

Jet flavor corrections

Finally, the last step in the Jet Energy Scale (JES) calculation involves applying flavor corrections. Jets originating from different flavors (u, d, s, c, b, g) exhibit distinct behaviors during hadronization and interaction with the detector, necessitating flavor-dependent correction factors. These factors are derived from simulated samples of jets tagged with truth-level flavor information. Flavor corrections account for differences in how energy and particles are distributed within the jet, as well as the detector's response to these particles.

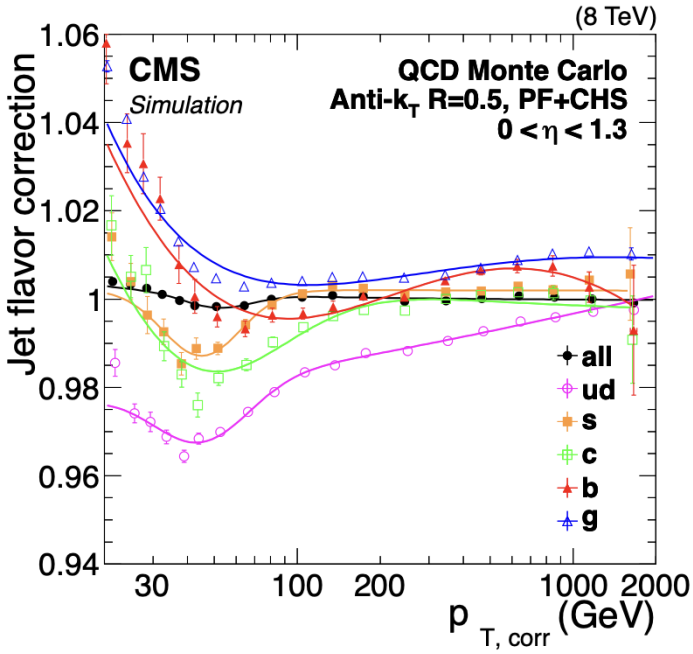


Figure 14: The residual jet-flavor correction factor, which is the inverse of the response, from Run 1, shown as a function of jet p_T. [44].

For instance, strange jets, which contain strange quarks (s), form strange hadrons such as kaons and Sigma baryons (Σ), which introduce additional complexities in jet reconstruction. These strange hadrons often decay into final states that include neutral particles or long-lived particles, making strange jets more challenging to tag and measure accurately compared to jets composed primarily of up, down, or even charm and bottom quarks. Figure 14 illustrates the residual jet-flavor corrections as part of the JES.

4.4.2 Jet energy resolution

The Jet Energy Resolution (JER) is defined as the width of the corrected response distribution for jets. This resolution generally indicates greater variability in the measured energy of jets compared to other physics objects, such as Z bosons. JER essentially reflects how well a detector can measure the energy of a jet. A low JER (narrow width) means that the measured energy is very close to the true energy,

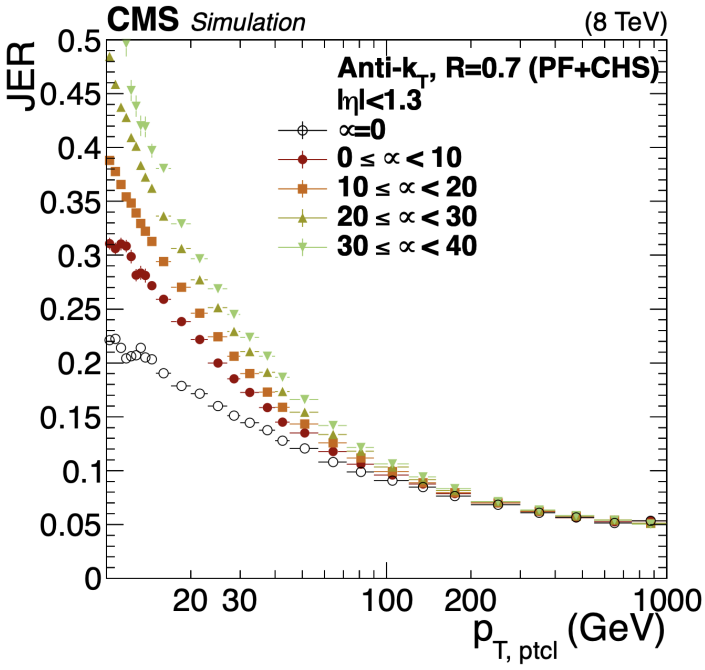


Figure 15: JER as a function of p_T in the barrel region for different levels of pileup (μ). PF+CHS jets with a radius of $R = 0.7$ [46].

with minimal uncertainty. In simulations, the JER is typically obtained by fitting a Gaussian distribution to a response histogram, with the width of the Gaussian providing a straightforward measure of the resolution.

For real experimental data, measuring the JER requires more sophisticated methods due to the complexities and imperfections present in actual detector environments. Two common techniques are used: the dijet asymmetry method and the γ+jet balance method. Utilizing momentum conservation, the dijet asymmetry method compares the transverse momentum (p_T) of two jets produced in the transverse plane of the dijet system. It relies primarily on the measured kinematics of the dijet events [46]. The γ+jet balance method, on the other hand, utilizes the balance between the photon and the recoiling jet in the transverse plane. This technique uses the photon as a reference object, taking advantage of the accurate transverse momentum (p_T) measurement provided by the ECAL [46]. The JER is illustrated in Figure 15.

5 Literature review of strange quark jets

This Chapter provides a comprehensive review of the literature on strange jets, focusing on their tagging, conservation in parton showers, the strange quark content of a proton, the dead cone effect, and the experimental research on strange quarks.

5.1 Tagging

The tagging process involves analyzing the jet's properties and its constituents to differentiate strange jets from those initiated by other quark flavors. Techniques range from traditional methods based on energy deposition patterns to advanced machine learning algorithms that exploit the complex relationships between particles within a jet.

One approach to tagging strange quark jets is by distinguishing them from jets initiated by down quarks. This can be achieved by examining the different energy deposition patterns in the hadronic and electromagnetic calorimeters. Long-lived neutral kaons (K_L^0), which are more frequently produced in jets originating from strange quarks, primarily deposit their energy in the HCAL. These kaons are stable enough to travel a significant distance before interacting with the HCAL, where they undergo hadronic interactions. In contrast, neutral pions (π^0), more commonly found in jets originating from down quarks, decay almost immediately into photon pairs. These photons initiate electromagnetic showers that predominantly deposit energy in the ECAL. This difference in the decay behavior of neutral particles results in strange quark jets typically exhibiting a higher proportion of neutral hadronic energy, as particles like K_L^0 deposit their energy in the HCAL. On the other hand, down quark jets display a greater fraction of neutral electromagnetic energy due to the prompt decay of neutral pions, which deposit energy in the ECAL [47].

Short-lived neutral kaons (K_S^0), on the other hand, can be identified when they decay in-flight into charged pion pairs ($\pi^{\pm}$) inside the detector's inner tracking sys-

tem. These charged pions leave tracks that can be detected in the silicon tracker, allowing for the reconstruction of the K_S^0 decay. Interestingly, the momentum fraction carried by the short-lived kaons is generally larger in strange jets compared to down jets, providing another critical handle for distinguishing between these types of jets. Despite the identical QCD and electromagnetic interactions shared by strange and down quarks, their differences emerge in the hadronization and subsequent decay processes [48, 49].

Recent advancements in machine learning, specifically Graph Neural Networks (GNNs) and Transformers, have shown promising results in jet tagging for other jet flavors, such as ParticleNet and ParT [50, 51]. ParticleNet is a particle identification algorithm based on graph neural networks, designed to detect hadronic decays of highly Lorentz-boosted top quarks and W, Z, and Higgs bosons, as well as to classify various decay modes such as $Z \to b\bar{b}$, $Z \to c\bar{c}$, and $Z \to q\bar{q}$. ParT, or Particle Transformer, is an advanced neural network architecture that utilizes the transformer model for particle identification and classification tasks in high energy physics. It is particularly effective in distinguishing different types of particles and decay modes by analyzing complex patterns in the data, similar to how transformers are used in natural language processing. These techniques leverage the intricate relationships between particles within a jet to enhance tagging accuracy. Incorporating GNNs and Transformers into the tagging of strange jets could significantly improve the ability to distinguish strange quark jets from other quark jets.

A variety of algorithms exist for tagging strange versus down quark jets, ranging from single whole-jet variables to Boosted Decision Trees (BDTs) and Convolutional Neural Networks (CNNs) utilizing jet images derived from detector subsystem inputs. CNNs, in particular, demonstrate modest yet noticeable improvements in discrimination [47]. This is illustrated in Figure 16.

Tagging strange jets in high-multiplicity environments, such as heavy ion col-

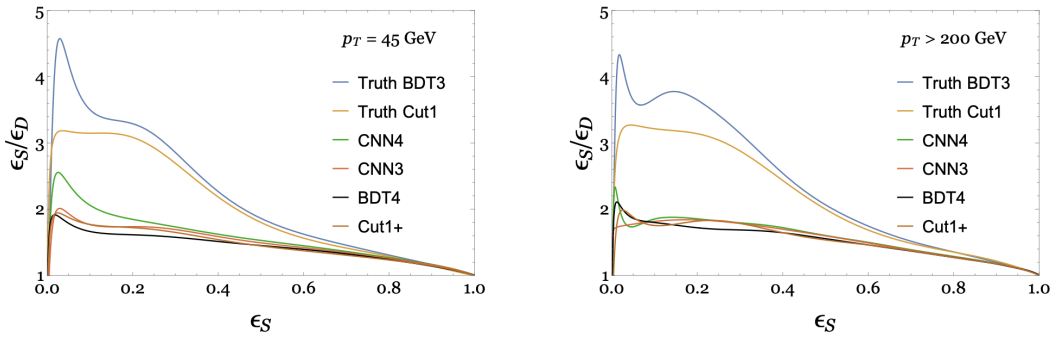


Figure 16: The ratio $\frac{\epsilon_S}{\epsilon_D}$ of strange jets to down jets tagged as strange jets as a function of the tagging efficiency ϵ_S for strange jets from Z-boson decay or QCD-initiated 13 TeV proton-proton collisions with jet $p_T > 200$ GeV [47]

.

lisions or high-multiplicity proton-proton collisions, presents additional challenges due to increased parton shower uncertainties. Addressing these uncertainties requires the use of comprehensive jet substructure information and advanced machine learning techniques.

5.2 Production of strangeness in parton showers

Strangeness is a quantum number assigned to particles based on the presence of strange quarks (s) or anti-strange quarks ($\bar{s}$). In strong interactions, which govern the behavior of quarks and gluons in parton showers, the strangeness quantum number is conserved, meaning that the total net strangeness before and after a strong interaction process remains the same. However, strange quarks are produced during parton showers primarily through the gluon splitting process ($g \to s\bar{s}$). Although the net strangeness does not change when an $s\bar{s}$ pair is produced, the number of strange-quark-containing particles increases, which is often referred to as an increase in "strangeness."

Parton showers describe the process by which high-energy quarks and gluons

evolve into a cascade of lower-energy partons before undergoing hadronization. During this evolution, partons undergo successive branching, splitting into additional partons, including strange quarks. The production of strange quarks during these splittings leads to an overall increase in strangeness in the final hadrons produced after hadronization.

After the parton shower, the partons combine to form hadrons, such as mesons and baryons, in a process known as hadronization or fragmentation. The production of strange hadrons during this phase is influenced by the strangeness content generated during the parton shower.

In thermodynamic systems, such as those created in high-energy heavy-ion collisions, the average energy (or temperature) can be high enough to reduce the suppression typically caused by the mass difference between strange and down quarks. This reduction in suppression facilitates the increased production of strange quarks.

One notable phenomenon observed in proton-proton (pp) and proton-lead (p-Pb) collisions is the enhancement of strange-particle yields relative to pions. This strangeness enhancement indicates that particles with higher strangeness content are produced more abundantly in high-multiplicity collision environments. The increased production of strange particles, such as kaons and hyperons, suggests that the mechanisms responsible for hadronization and parton fragmentation favor the production of strangeness in these scenarios [52].

Recent studies have further elucidated the role of strangeness production in parton showers by analyzing the production rates of particles with varying strangeness content. These analyses reveal that the production rates are significantly higher in events with greater particle multiplicities, indicating that the production of strangeness is closely linked to the overall energy density and dynamics of the collision environment [53].

In elementary collisions, the production of strangeness is typically suppressed

compared to light flavors. However, in larger systems, such as those where a thermodynamic description is applicable, enhanced strangeness production is expected. This is due to the possibility of producing and conserving strangeness over a larger volume. This pattern of strangeness enhancement was notably observed in Pb-Pb collisions at the SPS. It is important to note that a large system size alone does not suffice for strangeness enhancement; a mechanism for long-range communication within the system is also necessary, which might be provided by color deconfinement. The dependence of hyperon enhancements on centrality and energy remains an area that is not yet fully understood [54].

5.3 Strange quark content in the proton

The strange quark content in the proton is a fundamental aspect of QCD and is essential for precise theoretical predictions in particle physics. Strange quarks, being one of the three lightest quarks, are part of the sea quark distribution within the proton.

In QCD, the proton is described as a complex bound state of quarks and gluons. The presence of strange quarks in the proton is understood through parton distribution functions (PDFs), which provide a probabilistic description of the momentum fraction carried by the quarks and gluons. These PDFs are extracted from global fits to experimental data, incorporating measurements from deep inelastic scattering, Drell-Yan processes, and collider experiments [55].

The strange-quark parton distribution function (s-PDF) is a fundamental quantity that characterizes the probability distribution of strange quarks within the proton. According to Flavor SU(3) symmetry, the distributions of the three light sea quarks (up, down, and strange) should be equal. However, strange quarks may experience suppression due to their relatively higher mass [56]. The momentum fraction carried by strange quark and antiquark PDFs constitutes approximately 65% to 80%

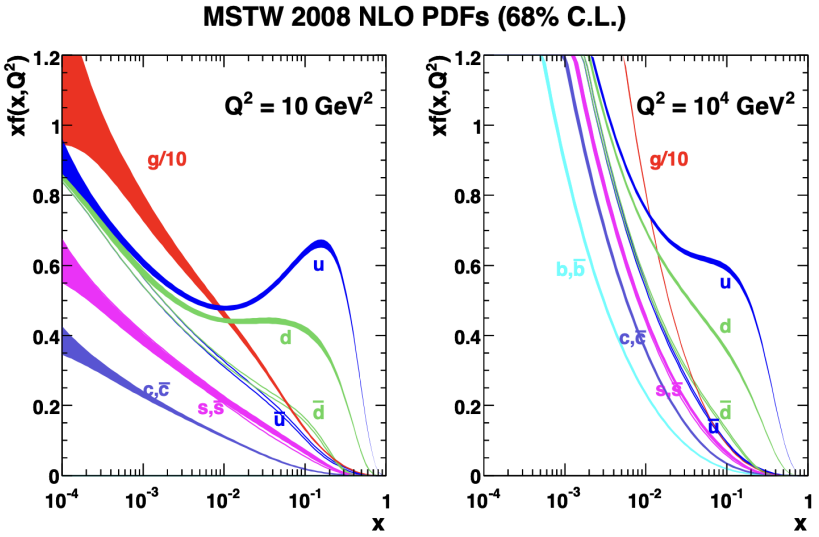


Figure 17: Next-to-leading order calculations of parton distribution functions performed at $Q^2 = 10\,\mathrm{GeV}^2$ and $10^4\,\mathrm{GeV}^2$, where Q^2 is the momentum transfer during proton-probe interactions. [29].

of the momentum fraction carried by the remaining light sea quarks across a broad energy range extending from about 1.6 GeV to 100 GeV. Recent evaluations of the strangeness content in the proton are in general agreement with findings from other modern, comprehensive studies of PDFs [57]. Illustration of PDFs is presented in Figure 17.

The strange parton distribution functions (s-PDFs) can be studied by examining the strange quark sea. Determining the distribution of strange quarks in a particle carries a higher level of uncertainty compared to other quark distributions. Extracting information about the strange sea quark distribution is particularly challenging. This uncertainty is notably significant for the precision of recent measurements, such as the W-boson mass by the ATLAS experiment. Therefore, enhancing the accuracy of the determination of the strange sea is crucial [58]. Recent data from the ATLAS collaboration at the LHC indicate an unexpectedly large presence of strange quarks in the proton's sea, challenging the conventional understanding derived from earlier

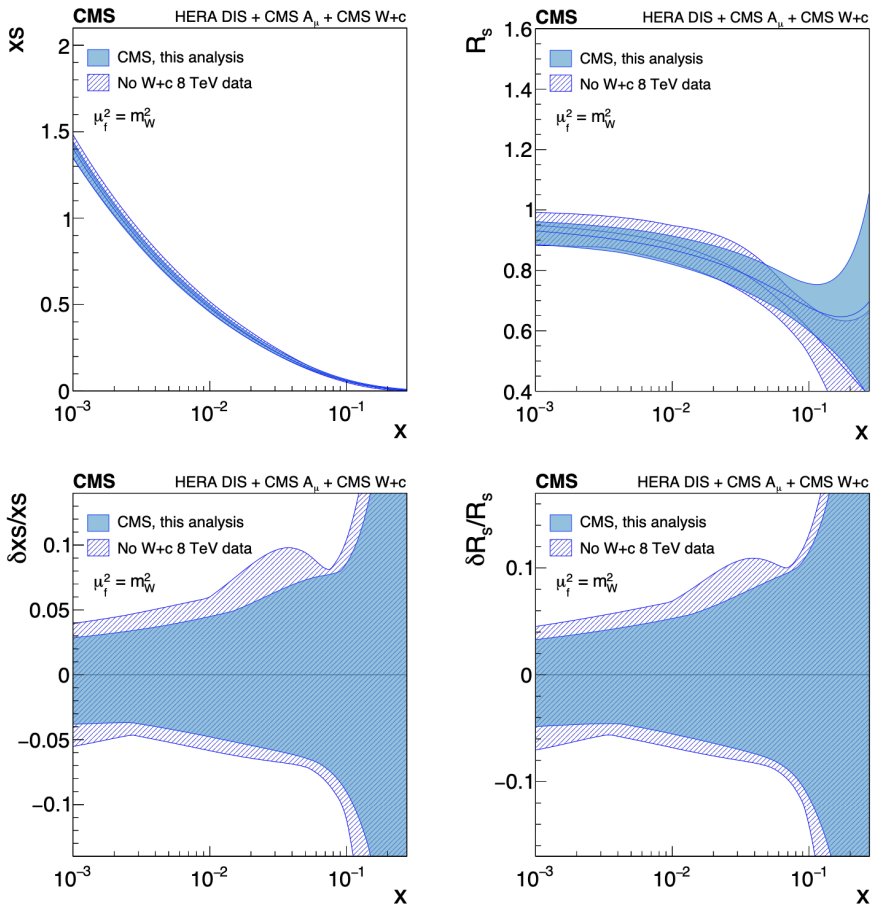


Figure 18: The distribution of strange quarks and the strangeness suppression factor as functions of x at the factorization scale of $\mu^2 = m_W^2$. The lower plots show the corresponding relative total uncertainties [60].

neutrino scattering experiments [59]. The s-PDF is illustrated in Figure 18.

The production of W+c at the LHC has provided a valuable tool for investigating the strange quark distribution within the proton at the energy scale corresponding to the W boson mass. This sensitivity is largely driven by the prominence of the $sg \to W^+ + c$ and $sg \to W^- + \bar{c}$ interactions. A notable feature of W+c production is the opposite electric charges of the W boson and the c quark. This is due to the nature of the primary processes $qg \to W + c$, which consistently produce opposite sign events. On the other hand, background processes such as gluon splitting $q\bar{q}' \to W + g \to W + c\bar{c}$ also result in final states with a W boson and a c quark (or

antiquark) of the opposite sign. However, they also produce a c antiquark (or quark) that carries the same electric charge as the W boson [60].

The strange quark content also has implications for the proton's spin structure. The contribution of strange quarks to the proton's spin, often quantified through the strange quark spin fraction, is a subject of ongoing investigation [61, 62].

Another intriguing subject is the presence of strange matter in compact stars, whether in their hadronic or quark matter phases. The theoretical framework alone does not rule out the possibility of strange quark matter that is completely stable. Nevertheless, the significance of strangeness in the context of compact stars remains uncertain. This uncertainty includes a spectrum of possibilities, ranging from the existence of fully strange stars to scenarios in which strange particles have little to no influence on these dense stellar objects [63].

5.3.1 Strange-antistrange asymmetry

The strange-antistrange asymmetry refers to the difference in the distributions of strange quarks (s) and antistrange quarks ($\bar{s}$) within the proton. While the proton's sea quark distribution includes both strange quarks (s) and their antiparticles ($\bar{s}$), studies suggest a possible asymmetry between them. This asymmetry arises due to non-perturbative effects in QCD, such as meson-baryon fluctuations, where the proton can fluctuate into a kaon and a hyperon state. Such fluctuations would lead to an excess of $\bar{s}$ over s quarks, affecting the overall distribution [61]. This is illustrated in Figure 19.

The asymmetry is primarily driven by the contributions from kaon loops, where kaons are pseudoscalar mesons with a significant strange quark content. These kaon loops introduce differences between the distributions of strange quarks and antiquarks in the proton. The fundamental mechanism leading to the asymmetry involves the virtual dissociation of a proton into a kaon and a hyperon, such as Λ or

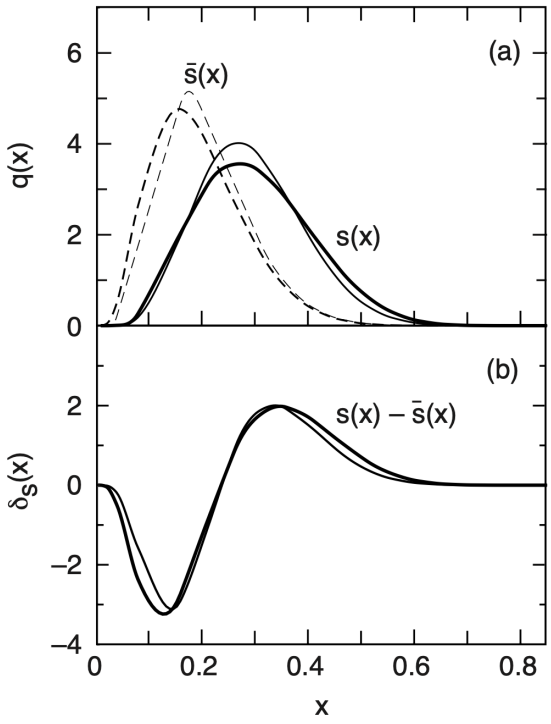


Figure 19: The momentum distributions of strange quarks and antiquarks are modeled using a light-cone meson-baryon fluctuation approach for $K^+\Lambda$ fluctuations. The model compares Gaussian and power-law wavefunctions with parameters: $m_q = 330$ MeV, $m_s = 480$ MeV, $m_D = 600$ MeV, momentum scale $\alpha = 330$ MeV, and power constant $p = 3.5$, using realistic meson and baryon masses [61].

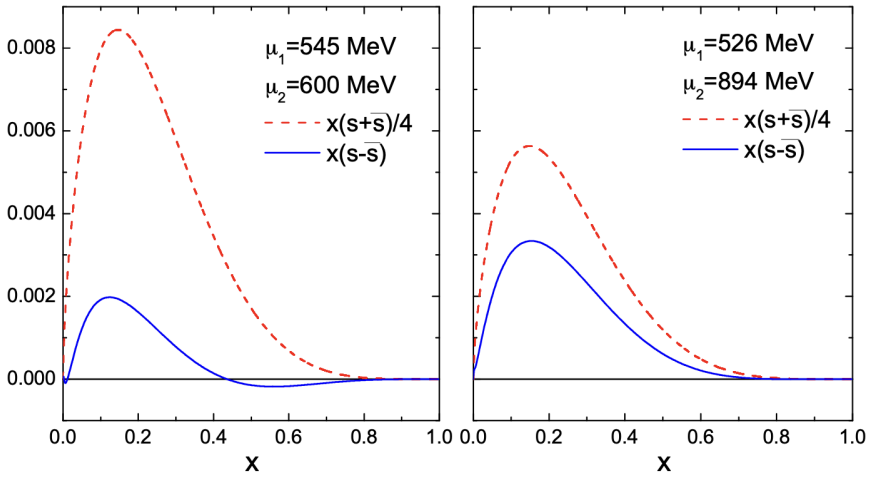


Figure 20: The difference $x(s-\bar{s})$ between the strange and antistrange PDFs, and the total sum $x(s+\bar{s})$, scaled by a factor of $\frac{1}{4}$, derived from kaon loops at $Q^2 = 1\,\mathrm{GeV}^2$. The left panel corresponds to fit parameters $(\mu_1, \mu_2) = (545, 600)\,\mathrm{MeV}$, while the right panel uses $(\mu_1, \mu_2) = (526, 894)\,\mathrm{MeV}$ [64].

Σ, where the strange quark is predominantly associated with the hyperon, and the antistrange quark is linked with the kaon. In particular, while perturbative generation of $s\bar{s}$ pairs through gluon radiation typically produces symmetric distributions, the breaking of chiral SU(3) symmetry can naturally lead to an asymmetry. This is realized through the meson cloud model, where the nucleon is surrounded by a cloud of pseudoscalar mesons like kaons, which interact asymmetrically with the proton's quarks, leading to a nonzero $s - \bar{s}$ asymmetry [64]. This is illustrated in Figure 20.

Experimental research on strange-antistrange asymmetry

Experiments involving neutrino and anti-neutrino scattering on hydrogen and heavy nuclear targets, have suggested a potential charge asymmetry in the strange sea, particularly at low values of Bjorken-x [65]. The parton distribution function (PDF) for strange quarks was determined to be somewhat harder compared to that of antistrange quarks. This difference in hardness was quantified by evaluating the second

moment of the asymmetry, offering a detailed measure of the observed disparity [64]:

$$S^- = \int_0^1 dx\, x\, (s(x) - \bar{s}(x))\,, \tag{12}$$

where x represents the fraction of the nucleon's momentum carried by the strange quark. The asymmetry was found to be approximately $S^- \approx (2 \pm 3) \times 10^{-3}$. Due to the conservation of strangeness, the first moment of the difference between the strange and antistrange quark distributions $(s - \bar{s})$ must be zero [64].

The initial leading-order (LO) analysis of the opposite-sign dimuon production in neutrino-nucleus deep inelastic scattering (DIS), as observed in CCFR [66] and NuTeV [67] experiments suggested a negative asymmetry, quantified as $S^- = (-2.7 \pm 1.3) \times 10^{-3}$, which pointed to a slight imbalance in the distribution of strange quarks and antistrange quarks within the nucleon. This indicated a possible preference for antistrange quarks, resulting in a negative asymmetry.

Subsequently, an analysis [68] incorporating next-to-leading order (NLO) corrections presented a different scenario. This analysis revealed a positive asymmetry value of $S^- = (1.96 \pm 1.43) \times 10^{-3}$ at a momentum transfer squared of $Q^2 = 16\,\mathrm{GeV}^2$. The transition from a negative to a positive asymmetry with the inclusion of NLO effects underscores the critical role of higher-order QCD corrections in accurately interpreting such data. These findings suggest that the strange-antistrange quark asymmetry within the nucleon is more intricate than previously thought, with NLO corrections pointing towards a possible excess of strange quarks over antistrange quarks. This highlights the complex dynamics of quark distributions within the nucleon and the necessity of thorough consideration of both experimental observations and theoretical models.

The strange-antistrange quark asymmetry in the proton reflects complex non-perturbative sea quark dynamics within the framework of QCD. Experimental and theoretical research continues to advance our understanding, although certain as-

pects remain unresolved. Future experiments and improved theoretical models will be essential in providing a more comprehensive picture of strange quark dynamics within the nucleon.

5.4 Dead cone effect and leading particles

The dead cone effect is a phenomenon in QCD that describes the suppression of gluon radiation at small angles relative to a quark's direction of motion. The dead cone is an angular region where gluon bremsstrahlung is suppressed around a heavy quark, illustrated in Figure 21. This effect results in variations in the shower development of light versus heavy quarks, depending on their mass. These differences are evident in both vacuum and quark-gluon plasma environments. The dead cone angle θ_{dc} is defined by:

$$\theta_{dc} = \frac{m_Q}{E_{Radiator}},$$

where m_Q is the mass of the quark, $E_{Radiator}$ is the energy of the quark [69].

The mass of strange quarks introduces the possibility of dead cone effects, potentially influencing the behavior of leading particles. The masses of up and down quarks are relatively small, approximately 2.2 MeV/c^2 for the up quark and 4.7 MeV/c^2 for the down quark. In contrast, the strange quark has a higher mass of approximately 95 MeV/c^2. These mass differences are significant, as they influence the dead cone effect due to the angular suppression of gluon radiation and consequently the characteristics of the parton showers. The dead cone effect suppresses radiation along the quark's momentum direction, resulting in a slightly harder leading particle. However, the dead cone effect is still more prominent in heavier quark jets, such as those involving charm and bottom quarks. The charm quark has a mass of approximately 1.3 GeV/c^2, and the bottom quark has a mass of about 4.2 GeV/c^2.

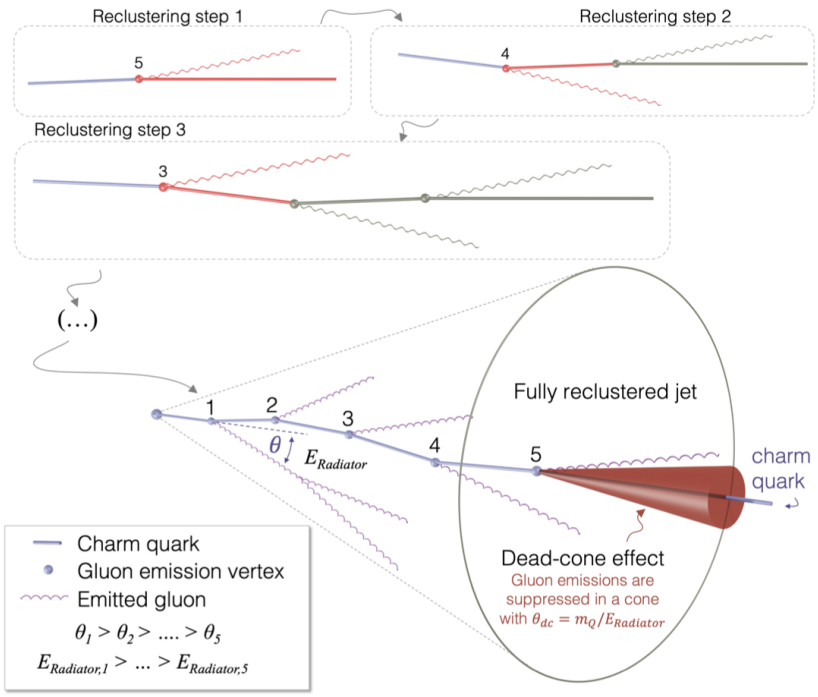


Figure 21: The charm quark's energy ($E_{\mathrm{Radiator},n}$) decreases with each splitting, leading to gluons being emitted at larger angles (θ_n) relative to the quark's direction. The quark's mass (m_Q) remains constant, while gluon emissions are suppressed inside the dead cone, which widens as the quark's energy decreases [70].

5.5 Experimental research on strange quarks

In addition to those already mentioned in this chapter, recent experimental studies have explored various aspects of strange quark behavior, including jet properties, strange gluon plasma, strangeness enhancements, and precise measurements of PDFs [71–73].

Strange quark matter (SQM) is a theoretical state of matter comprising nearly equal quantities of up, down, and strange quarks. Numerous experimental approaches have been employed to search for the potential presence of SQM and to attempt its production in laboratory settings [71]. Studies on strange gluon plasma aim to provide a deeper understanding of the conditions present shortly after the Big Bang [73].

Strangeness enhancement, which refers to the increased production of strange particles, is a key indicator of the phase transition from hadronic matter (the "confined phase") to quark-gluon plasma (QGP) (the "deconfined phase"). This enhancement occurs because gluons in the QGP can split into pairs of strange quarks and antiquarks, significantly increasing the density of strange quarks. This increased production of strange quarks directly leads to a higher yield of multi-strange hadrons during the hadronization process. Strangeness enhancement typically occurs in heavy-ion collisions [74].

5.6 Jet energy scale

The Jet Energy Scale (JES) is a crucial factor in accurate jet measurements, as it directly affects the reconstruction of jet energies in particle physics experiments. In the context of the CMS detector, the JES is influenced by the type of jets being measured, particularly distinguishing between jets formed from calorimeter-based methods (CaloJets) and those reconstructed using the Particle Flow (PF) algorithm.

CaloJets rely primarily on the energy deposited in the calorimeter, and hence

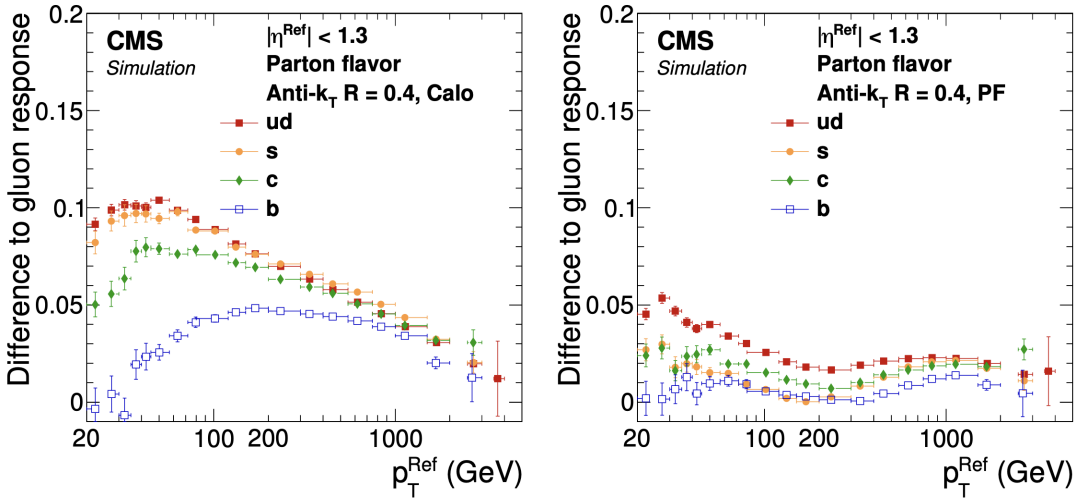


Figure 22: The absolute difference between quark and gluon jets' jet energy responses, plotted as a function of the reference momentum $p_{\mathrm{T}}^{\mathrm{Ref}}$. The left plot is for Calorimeter (Calo) jets, while the right plot is for Particle Flow (PF) jets [36].

show little variation in response between jets originating from different quark flavors, such as strange, up, or down quarks. In contrast, PF jets, which are highly sensitive to the balance between charged and neutral hadrons, can be significantly affected by the presence of strange quarks. This is largely due to the fact that strange quark jets often produce long-lived neutral hadrons, such as K_L^0, which carry significant transverse momentum (p_T) and lead to a reduced response in PF jets compared to jets dominated by up or down quarks.

The differences between Calorimeter Jets (CaloJets) and Particle Flow (PF) jets are further highlighted in Figure 22. The primary difference arises from their sensitivity to the charged versus neutral hadron content within the jet. PF jets are more sensitive to this balance and CaloJets are more influenced by the transverse momentum distribution and angular separation of the hadrons within the jet. This sensitivity makes CaloJets particularly responsive to the overall multiplicity and spatial distribution of particles, factors that also affect JES calibration [36].

5.7 Summary and outlook

Experimental studies on strange quarks have significantly advanced our understanding of QCD and the behavior of matter under extreme conditions. The production and detection of strange hadrons in high-energy collisions provide crucial data for investigating the properties of quark-gluon plasma and the mechanisms of hadronization. Future research could reduce uncertainties in JES, improve the precision of parton distribution functions, and further explore the implications of the dead cone effect in strange quark jets.

6 Methods

This chapter outlines the methods used in the analysis, detailing on the datasets and event selection criteria for studying strange jet properties using MC simulations and the energy scale studies of strange jets based on experimental data.

6.1 Datasets

To investigate the properties of jets, with a specific focus on strange jets compared to other flavors, simulation dataset was used. Additionally, data-based studies focusing on the energy scale of strange jets employed both simulations and experimental datasets. A more detailed description of the datasets used can be found in Appendix A.

- **Run3 QCD PFNano Simulation Dataset (Jet Property Analysis):** A Run3 QCD simulation dataset, generated using `/QCD_PT-15to7000_TuneCP5_13p6TeV_pythia8/Run3Summer23MiniAODv4-castor_130X_mcRun3_2023_realistic_v14-v1/MINIAODSIM`, was used to study the properties of different jets. This dataset includes particle flow (PF) candidates and generator-level particles (GenPartCand) associated with AK4 and AK8 jets. Although the dataset was initially created in MiniAOD format, the PFNano format was selected for its efficient storage and handling of PF candidates and GenPartCands.

- **Run2 Simulation Dataset (Strange Jet Energy Scale):** To study strange jet energy scale, we utilized the `TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8` Run2 simulation dataset. This dataset simulates semileptonic $t\bar{t}$ events, which provide a well-understood environment for strange jet energy scale study along with charm tagging. It is in MiniAOD format and was selected to match the

conditions and format of the corresponding late-2016 data, allowing for direct validation of the energy scale and tagging techniques.

- **Run2 Data (Strange Jet Energy Scale):** The late-2016 data [2], collected by the CMS detector, was used to validate the strange jet energy scale. This dataset includes $t\bar{t}$ events from the `Electron` and `Muon` Primary datasets, with an integrated luminosity of $16.8\,\text{fb}^{-1}$ and an average pileup of 27 interactions per crossing. Semileptonic $t\bar{t}$ events were selected for their clean signature. The use of Run 2 data ensures consistency with the corresponding simulation dataset.

6.2 Event selection: jet property analysis

For the study of jet properties using the Run3 QCD multijet events, the event selection focused on the two leading jets. These jets were filtered to ensure pseudo-rapidities (η) less than 1.3, placing them within the barrel region of the detector. The transverse momenta (p_T) of the jets were required to be within the range of 80 to 100 GeV for the particle-level analysis. For jet-level studies, a more inclusive transverse momentum threshold of $p_T > 10$ GeV was applied.

6.3 Event selection: strange jet energy scale

The signal used to create these datasets was $t\bar{t}$ events where both top quarks decay into a W-boson and a b-jet. One W-boson decays hadronically ($W \to q\bar{q}'$) and the other leptonically ($W \to \ell\nu$). In the hadronic decay, if one jet is tagged as charm, the theory predicts that the other is strange in most cases.

To ensure the events are consistent with $t\bar{t}$ production, a kinematic fit is applied using all final-state products: two light-flavor quark jets, two b-jets (tagged with

[2]Early-2016 data was excluded due to tracking inefficiencies that were difficult to model. The ultra-precise CMS W mass measurement also used the late-2016 dataset.

DeepJet), and one lepton. The fit assigns a likelihood (fitProb), and only events with fitProb > 0.2 are considered.

For jet and lepton selection, the following criteria were applied: An electron or muon with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$ was required. At least four jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.5$, of which exactly two are b-tagged, were selected. The hadronic W-boson candidate was formed from the two hardest non-b-tagged jets, with the requirement that their combined mass is $m_W > 30\,\text{GeV}$. The pair of W-boson and b-jet with the highest probability was selected as the hadronic top candidate. A kinematic fit probability for a $t\bar{t}$ event was applied, requiring fitProb > 0.2. This fit assumes $m_W = 80.4\,\text{GeV}$ and $m_{t,\text{lep}} = m_{t,\text{had}}$ while solving for the neutrino momentum along the beam direction p_Z.

The jets were corrected using centrally-produced MC truth corrections and residual corrections derived from a combination of $Z + \text{jet}$, $\gamma + \text{jet}$, multijet, dijet, and inclusive $W \to qq'$ data. As a result, the W-boson mass between data and MC is well-calibrated, up to known final-state radiation (FSR) biases of approximately 0.5%.

A c-tagging efficiency of 20% and a mis-tagging rate of 1% corresponded to a working point with c-tag > 0.43, used to identify at least one charm-tagged jet. A second working point with c-tag < 0.06 targeted 80% efficiency for identifying non-charm jets, with a 60% mis-tagging rate.

6.4 Analysis Tools

The data analysis for this thesis was conducted using the ROOT framework, which provided essential tools for data handling, statistical analysis, and visualization. The data was structured as `TTree` objects, from which we created e.g. histograms and profiles to visualize different aspects of the analysis. Additionally, we utilized different graphical tools for presenting the results.

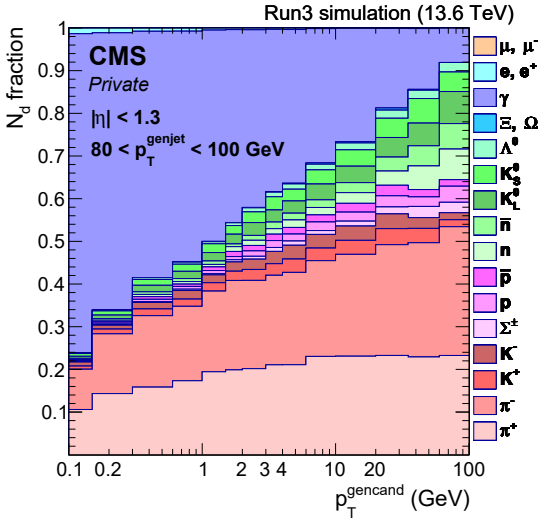
7 Results from simulation

In this chapter, we present the detailed findings from our analysis of strange jets, highlighting the properties that distinguish them from other flavor jets, particularly up and down jets. We quantitatively analyze their particle composition, energy fractions, angular distribution, and jet response. The studies in this Chapter are done using jets drawn from the QCD multijet events with two leading jets in each event. The results are organized to provide a comprehensive overview of the characteristics of strange jets.

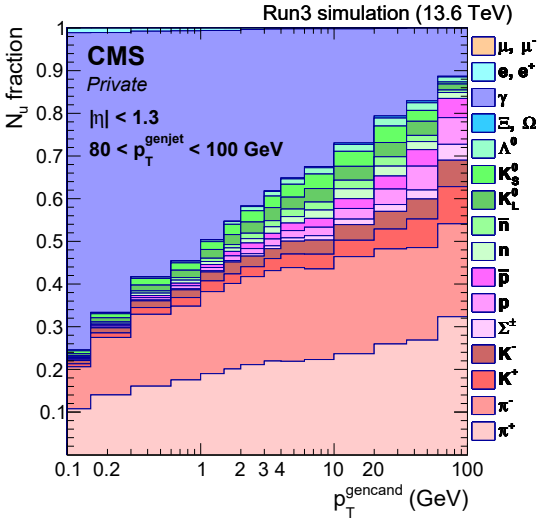
7.1 Particle composition

Particles within strange jets can be categorized by their charge into charged and neutral particles, with neutral particles further subdivided into neutral hadrons and photons. Strange jets differ from other jet flavors on average due to their composition, particularly in the presence of strange hadrons. One fundamental but not entirely unique property is that strange jets must contain at least one strange particle (such as K^+, K^-, K^0_L, K^0_S, or Λ^0). However, other jets can also contain strange quarks from gluon splitting ($g \to s\bar{s}$). Strange jets tend to have a higher proportion of neutral strange hadrons compared to light jets, as particles like K^0_L, K^0_S, and Λ^0 make up a larger fraction of neutral particles than π^0, neutrons, and antineutrons do in light-flavor jets.

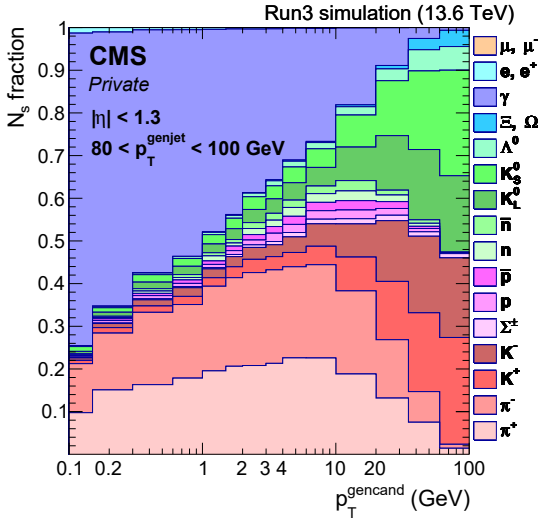
The number fractions of particles in strange, down and up jets as functions of the transverse momentum of the particle (p_T^{gencand}) are illustrated in Figure 23. They consist of the following particles: negatively and positively charged pions (π^-, π^+), photons (γ), which are primarily produced from π^0 decays, protons (p) and anti-protons ($\bar{p}$), neutrons (n) and anti-neutrons ($\bar{n}$), negatively and positively charged kaons (K^-, K^+), short-lived and long-lived neutral kaons (K^0_S, K^0_L), and Lambda baryons (Λ). Additionally, strange baryons such as Sigma (Σ^-, Σ^+), Xi (Ξ), and



(a) Down jets: explicit particle fractions.



(b) Up jets: explicit particle fractions.



(c) Strange jets: explicit particle fractions.

Figure 23: Explicit fractions of different types of particles in down, up, and strange jets. Charged hadrons are shown in shades of red, neutral hadrons in shades of green, and photons and leptons in shades of blue.

Omega (Ω) are also shown, along with muons and anti-muons (μ^-, μ^+) and electrons and positrons (e^-, e^+).

Charged hadrons are primarily charged pions and charged kaons, while neutral hadrons include neutral kaons, neutrons, and antineutrons. This indicates that the increase in neutral hadrons with high p_T in strange jets is primarily due to the rising fraction of neutral kaons, which is a distinguishing feature of strange jets. In contrast, up and down jets contain nearly as many neutrons and antineutrons as neutral kaons.

Particles with low (p_T) in down jets (Figure 23a), in up jets (Figure 23b), and in strange jets (Figure 23c) exhibit similar distributions: over 70% of the particles are photons (γ), more than 20% are charged hadrons, and only a small fraction are neutral hadrons.

As p_T of the particles increase, the fraction of neutral hadrons in strange, down, and up jets also increases, with this rise being much more pronounced in strange jets. Beyond the 20 GeV region, strange jets exhibit a significant increase in neutral hadron content, accompanied by a relative decrease in charged hadrons compared to up and down jets. However, the overall charged hadron content still increases, as neutral pions (π^0) decay into photons, redistributing the original π^0 energy from high to low particle p_T.

In strange jets, particles with high $p_T \approx 100$ GeV consist of approximately 50% neutral hadrons and 50% charged hadrons, with only a small fraction of photons. This observed decrease in the photon fraction may be partly artificial, as the energy of the original π^0 is split between its two decay photons, which places them in lower p_T bins.

In contrast, down and up jets exhibit a different trend with high p_T particles. Particles with $p_T \approx 100$ GeV in up jets consist of about 80% charged hadrons, with a small fraction of neutral hadrons and slightly over 10% photons. This indicates

that high p_T charged hadrons are more common in down and up jets than in strange jets, whereas high p_T neutral hadrons are much more common in strange jets than in down or up jets. Thus, the higher likelihood of finding neutral hadrons in strange jets with higher transverse momentum, compared to up or down jets, suggests that the presence of high p_T^{gencand} neutral hadrons could be a distinguishing feature for identifying a jet as a strange jet.

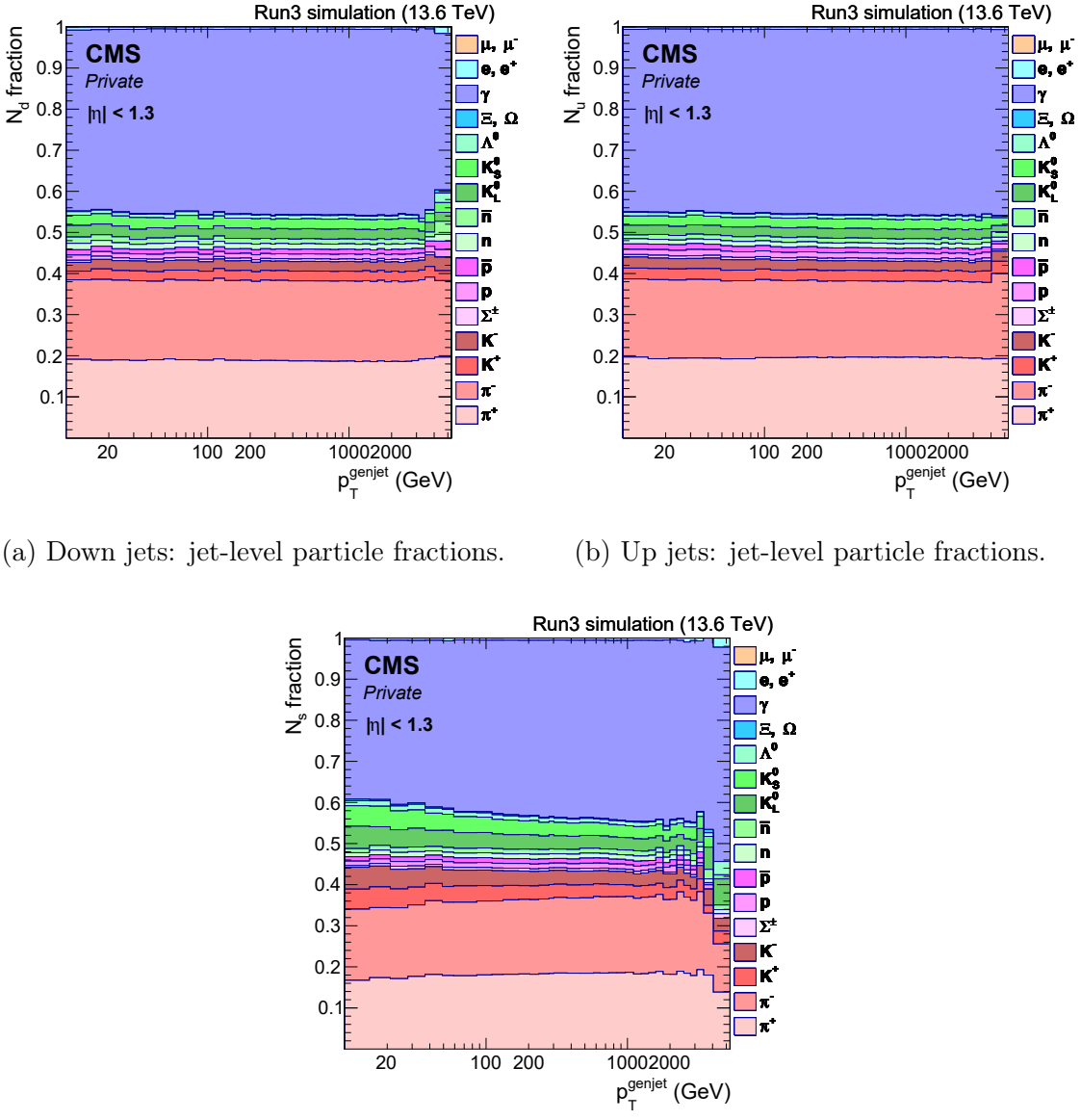
At the jet level, as shown in Figure 24, the behavior of particles does not exhibit the same distinctions observed previously. The differences between strange, down, and up jets become less pronounced, with only a slightly higher fraction of neutral hadrons in strange jets. This suggests that when considering the transverse momentum (p_T) of a jet, there are no significant differences between the various jet types.

7.2 Leading particle analysis and angular distribution

Strange jets are expected to exhibit enhanced kaon production, leading to a higher fraction of leading neutral hadrons compared to jets originating from up, down, or charm quarks, particularly at high transverse momentum (p_T).

Strange quarks are more likely to be the leading partons in strange jets. This is due to the suppression of FSR caused by the strange quark mass (dead cone effect), as well as the fact that in gluon splitting, the leading gluon tends to radiate more FSR before splitting into strange quarks. The key distinguishing feature of strange jets is that the leading particle is often a strange neutral hadron and tends to carry a larger fraction of the jet's total momentum compared to leading particles in light-flavor jets. A leading s-hadron would also be the most critical target for tagging in many scenarios.

Figure 25 illustrates the particle p_T distribution for different types of jets, including all jets and strange jets. The particles are further divided into categories:



(a) Down jets: jet-level particle fractions.

(b) Up jets: jet-level particle fractions.

(c) Strange jets: jet-level particle fractions.

Figure 24: Fractions of different types of particles in down, up, and strange jets. Charged hadrons are shown in shades of red, neutral hadrons in shades of green, and photons and leptons in shades of blue.

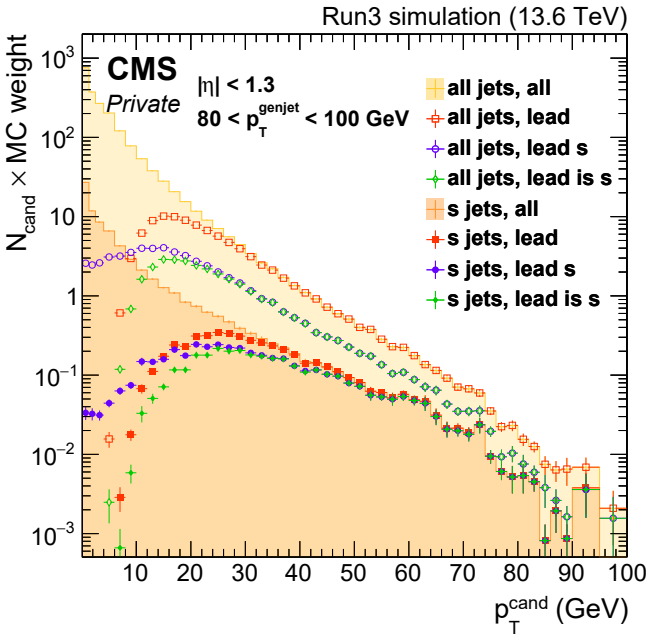
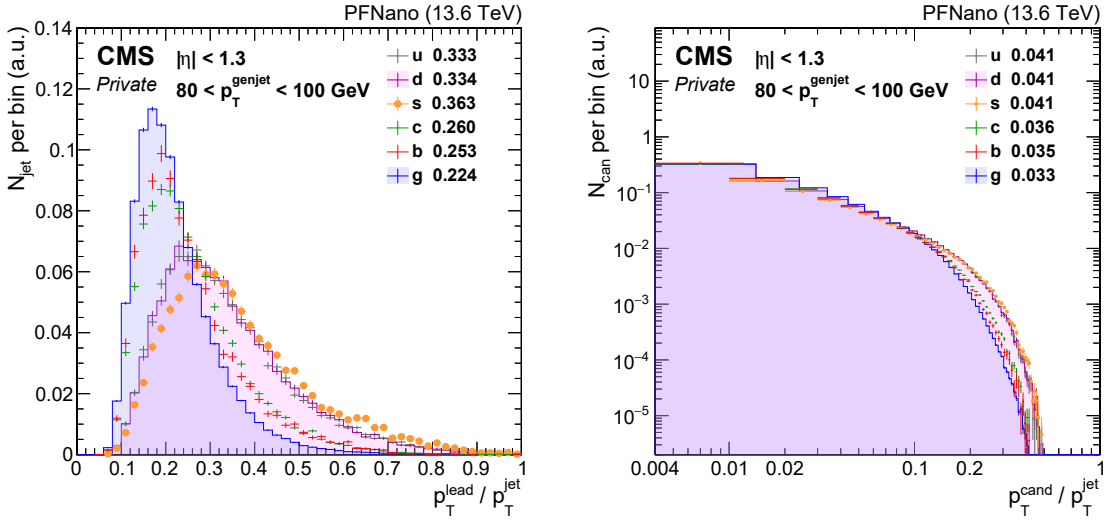


Figure 25: Illustration of all jets, strange jets, and further separation of leading particles and highest momentum strange particles. The figure also highlights which of the leading particles are strange particles.

all particles, leading particles, the highest momentum strange particles, and leading particles that are strange particles.

One of the key features in Figure 25 is that the leading particle p_T distribution, particularly for leading strange particles, falls less steeply in strange jets compared to up and down jets. This allows for better discrimination between strange and light-flavor jets and raises the question if the leading particle p_T distribution in s-jets is different from light flavour jets.

Figure 26a then confirms that the leading particle p_T distribution in strange jets is indeed harder compared to light-flavor jets. It depicts the fraction of the total jet energy carried by the leading particle. In strange jets, the leading particle carries about 36% of the jet's energy, while in up and down jets, this value is approximately 33%. The energy fraction carried by the leading particle is even smaller in heavier jets and gluon jets. This highlights a distinct feature of strange jets: their leading



(a) Fraction of total energy carried by the leading particle.

(b) Fraction of total energy carried by individual particles.

Figure 26: Energy fractions of the leading particle, and energy fractions of individual particles in jets of different flavors.

particle carries a significant proportion of the total jet energy.

Figure 26b illustrates the fraction of energy carried by individual particles relative to the total jet energy. It shows that lighter flavors (d, u, s) exhibit similar behavior, with each particle carrying approximately 4.1 % of the total jet energy on average. In contrast, particles in heavier particles (c, b) or gluons carry about 3.3% to 3.6% of the total energy. This shows that the difference observed for the leading particle in s-jets becomes diluted when considering all particles in the jet.

One expected cause for the harder leading particle in strange jets is the dead cone effect, where the heavier strange quark loses less momentum to FSR. However, the $\pi^0 \to \gamma\gamma$ decay might remove some harder π^0s from the leading particle population. A future analysis that reclusters the photons from π^0 decays could help eliminate this artifact and confirm the role of the dead cone effect more clearly.

Additionally, at $p_T > 20$ GeV, all highest momentum strange particles are leading particles. This indicates that if a strange particle with $p_T > 20$ GeV is observed, it

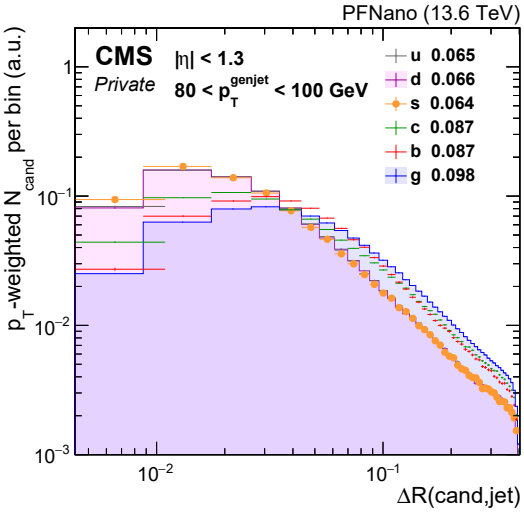


Figure 27: Angular distance (ΔR) between particles in jets of different flavors.

is likely to be the leading particle. This can be seen from the open purple squares and open green diamonds. When focusing on strange jets, we observe that although the leading particle is often strange, this is not always the case. The orange squares indicate that there are strange jets where the leading particle is not a strange particle. However, the distribution largely follows a pattern similar to that observed in the all-jets case.

Next, we explore the key characteristics of jets by examining the angular distance between particles to investigate whether the dead cone effect might result in a reduction in particle density near the jet core or harden the leading particle by suppressing FSR for strange hadrons.

Figure 27 illustrates the angular distance (ΔR) between particles in jets of different flavors. The figure shows that light-flavored quark jets (d, u, s) have smaller average angular distances compared to heavier quark jets and gluon jets. In light-flavored quark jets, particles are generally harder (high p_T) and more concentrated towards the jet center, with no noticeable reduction in density observed in strange jets compared to down or up quark jets. Interestingly, these properties are preserved in the leading particle of these jets.

The expected dead cone radius is approximately $\theta_{dc} = \frac{m_Q}{E_{Radiator}}$, which for strange quarks is roughly $\frac{0.1\,\mathrm{GeV}}{100\,\mathrm{GeV}} \approx 10^{-3}\,\mathrm{rad}$. Since the first bin extends up to $10^{-2}\,\mathrm{rad}$, it is about 10 times wider than the expected dead cone radius.

7.3 Flavour jet response

Finally, we study the different flavour jet response. Here we introduce an application of a scale factor to account for the variation in response between u, d jets and c, s jets used later in Appendix C. Specifically, a scale factor of 0.986 is applied to u and d jets to correct for this difference.

Figure 28 illustrates the response of different flavored jets as a function of generator-level jet momentum. This response highlights the differences in jet energy corrections (JEC) for various jet flavors.Figure 28 also shows that the scale factor originates from the different JECs for u, d jets compared to c, s jets. The 0.986 scaling factor should remain relatively stable across different p_T values. This scaling factor is also important in W mass reconstruction, where the response differences between light-flavor jets (u, d) and strange/charm jets (s, c) can impact the accuracy of the mass measurement.

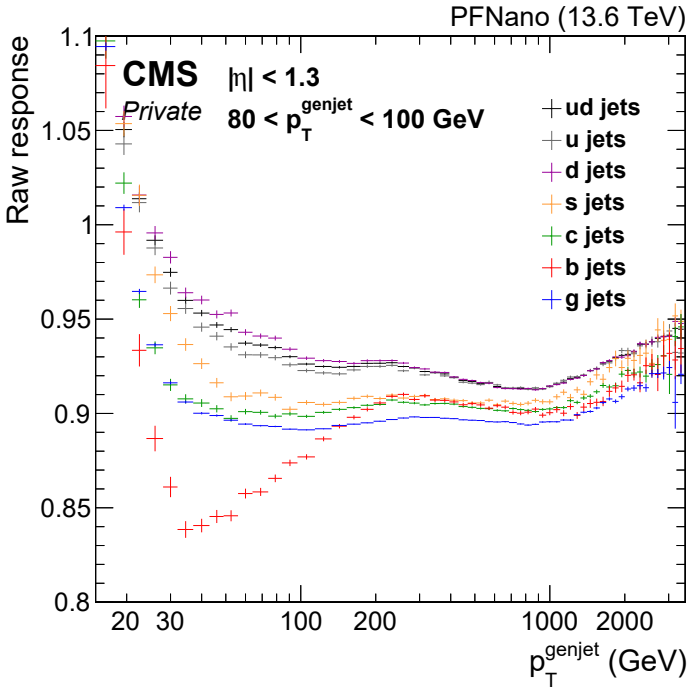


Figure 28: Response for different flavored jets as a function of generator-level jet momentum.

8 Results from data

In this chapter, we introduce a data-based exploration of one key feature of strange jets: the jet energy scale (JES) relative to light quarks. To determine accurate JES, we use ROC curves to identify suitable working points for tagging charm-strange pairs with a charm tagger and up-down pairs with an anti-charm tagger. We derive scale factors to account for b-tagging and c-tagging efficiencies and obtain quantitative estimates of the product and ratio of c-jet and s-jet JES using mass and momentum distributions, from which we extract the s-jet JES with uncertainties. Additionally, we explore potential sources of bias, such as helicity effects that cause up-type quarks to appear harder than down-type quarks, and the p_T-dependent charm tagging efficiency that further amplifies this effect. Finally, we validate the results by comparing scaled MC simulations to experimental data. The studies in this Chapter are done using jets drawn from the 2016 semileptonic $t\bar{t}$ events for MC

simulation and data.

8.1 $W \to cs$ event selection through charm tag

First, we will introduce the method used to tag the charm jets: Instead of directly tagging strange jets, we explore a new strategy to select events where a W-boson decays into a charm and strange quark pair ($W \to c\bar{s}$) and tag the charm jet through secondary vertex identification. This approach allows us to assume that the other jet in the event is likely a strange jet, providing an enriched sample of strange jets. The information is stored in the c tag discriminator. This method of strange jet identification does not rely on directly tagging the strange jet but instead uses the known decay topology of the W-boson.

A key advantage of this approach is that it avoids selection bias on s-jets. Since the strange jet is not explicitly tagged, we can directly assess the performance of strange jet tagging techniques by applying them to the enriched sample and measuring their effectiveness without pre-selection.

8.2 Optimizing charm tag working points through ROC curves

Suitable working points for the used c tag discrimator were determined using Receiver Operating Characteristic (ROC) curves. It is an effective approach for evaluating and comparing the performance of different tagging methods. By analyzing ROC curves, specific c tag discriminator can be derived to differentiate between charm-strange (cs) pairs and other jet pairs, such as up-down (ud) pairs. This Chapter explores the application of ROC curves in jet pair tagging, focusing on identifying the most effective c tag thresholds for accurate jet classification.

Initially, c-tagging (and b-tagging) were used to identify b and c jets. The selected c tag discriminators were used to differentiate cs pairs from ud and other

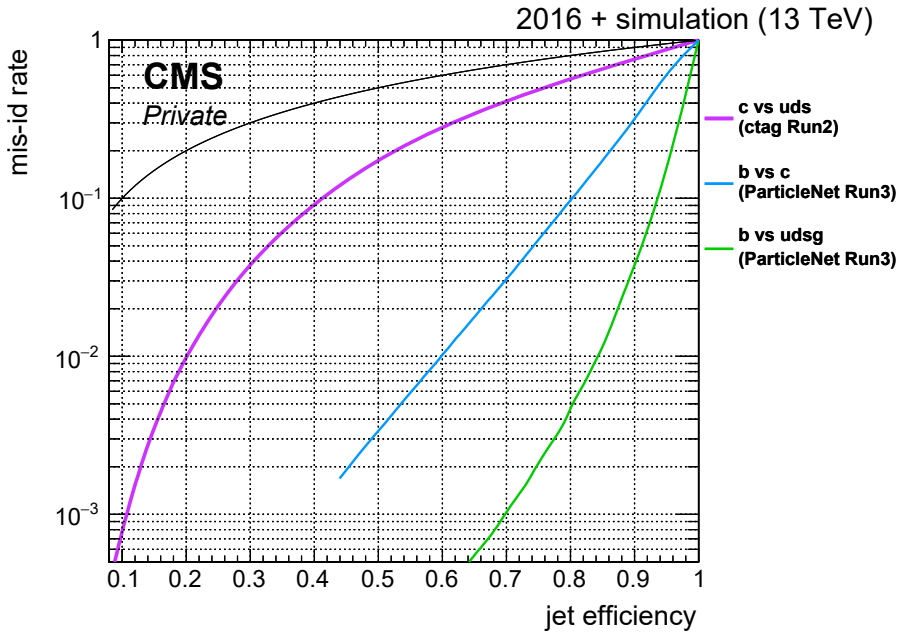


Figure 29: Combined ROC curves for different tagging methods are shown. The purple line represents the ROC curves derived from the Run2 2016 $t\bar{t}$ sample using the c-tag discriminator. The blue and green lines represent the ParticleNet, derived from the Run3 QCD multijet sample for jets with $p_T > 30$ GeV and $|eta| < 2.5$.The black line represents random selection (diagonal).

jet pairs. In addition, DeepFlavour and ParticleNet tagging were applied to further distinguish between light-quark jets using Run3 QCD multijet sample, described in detail in Appendix B.

A ROC curve plots signal efficiency against the background mistag rate, allowing for the determination of optimal c-tag discriminator cut values based on specific efficiency and mistagging requirements. Different ROC curves are illustrated in Figure 29. For this analysis, we used a DeepFlavour binary discriminator defined as $\frac{P(c)}{P(c)+P(uds)+P(g)}$ [75], where $P(c), P(uds), P(g)$ represent the classifier output probabilities for charm, light (u, d, s), and gluon jets.

The c tag working points corresponding to desired mistag rates and efficiencies were derived from the ROC curves. For example, to achieve an c-tagging efficiency

close to 0.20 with a mistag rate of 0.01, the ROC curve (Figure 29) (purple line) yielded an efficiency of 0.20019 and a mistag rate of 0.00986, meeting the target values. The corresponding c tag discriminator cut for this tight working point was found to be > 0.43, which was then used to tag cs pairs in the data. Similarly, to distinguish ud pairs from cs pairs, a c tag discriminator cut < 0.06 was determined for a loose working point, aiming for an efficiency close to 0.80 with a mistag rate around 0.60.

8.3 Charm tag performance and mass distibution of jet pairs

In this Chapter, we study the performance of the charm tag using jet pairs. Figure 30 shows the number fraction of different jet pairs: first for the non-tagged sample (red), and then for the sample where c tag > 0.43 was used to tag the cs pairs (blue) and for sample where c tag > 0.06 was used to tag the ud pairs (green). Tagging significantly increases the purity of the cs sample for further analysis. In the original sample, there are about 50% ud pairs and approximately 40% cs pairs. After tagging, the fractions shift to almost 90% cs and less than 10% ud, making the sample much more cs pure. Here, the performance of c tag discriminator < 0.06 is somewhat less effective giving about 70% of the pairs are ud, with less than 30% being cs pairs. In comparison, in the non-tagged sample, about 50% are ud pairs and slightly more than 40% are cs pairs. The cs sample can be made quite clean, to the extent that a cd combination containing a real c-quark can rise almost to the same level as a wrongly tagged ud pair. Thus, tagging considerably improves the purity of the ud sample.

This analysis also clearly shows that there are relatively more pairs in category x (other pairs), and less ud and cd, but the ud/cs ratios are quite similar. This could indicate that the ud and cs classifications are fairly accurate, but that there is more other pairs (category x) in the data. Alternatively, it might suggest that the data

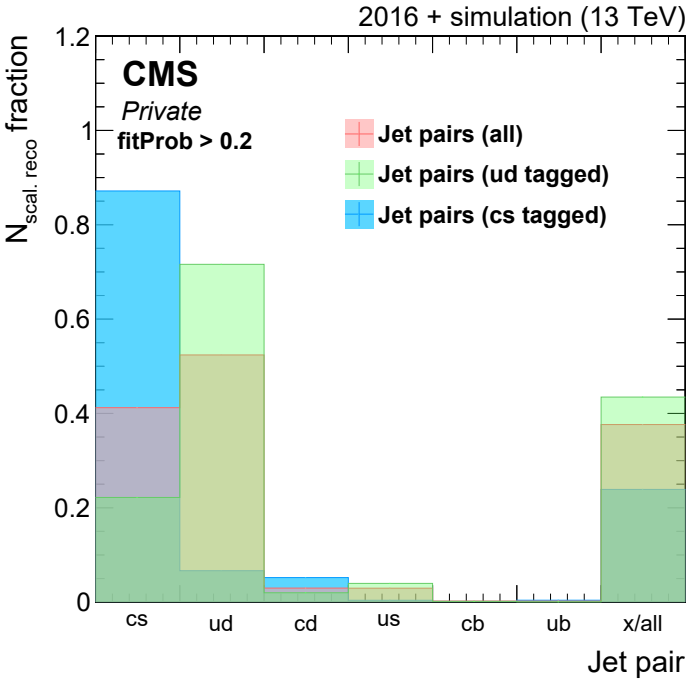


Figure 30: The number fraction of different jet pairs with and without scaling and tagging for cs pairs.

tagging performs slightly worse for both c and anti-c quarks.

In addition, Figure 31 shows the mass distributions for different jet pairs. Here, we see that the mass peak of the ud pairs is slightly shifted to the right, indicating a slightly higher mass than cs pairs. The masses should be the same for both pairs since they both result from W-boson decays. The W-boson mass is invariant and independent of the specific quark flavors involved in the decay. For comparison, the mass peaks at the generator level are also presented. The difference between the reconstructed level and the generator level is evident.

8.4 Extraction of s-jet JES from W-boson mass and momentum difference of CS pairs

In this Chapter we extract the JES for s-jets from mass and momentum difference of cs pairs realtive to ud pairs.

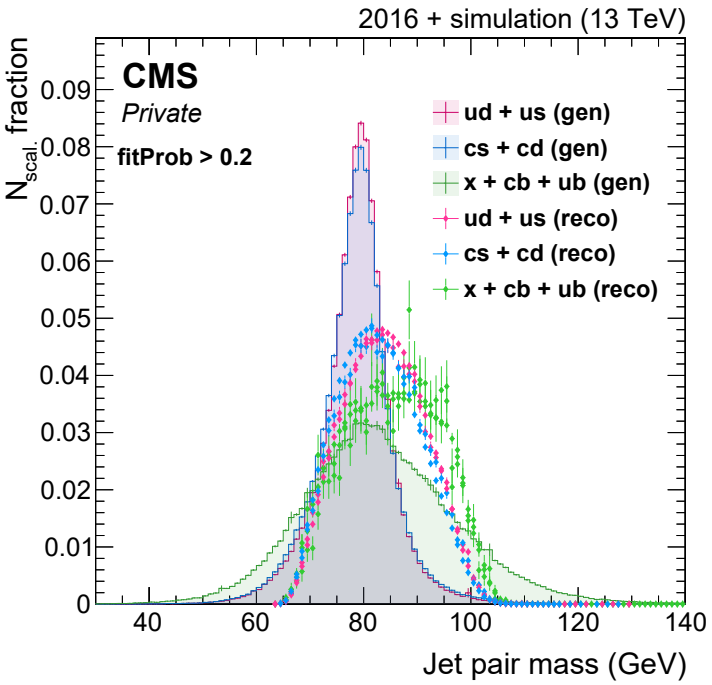
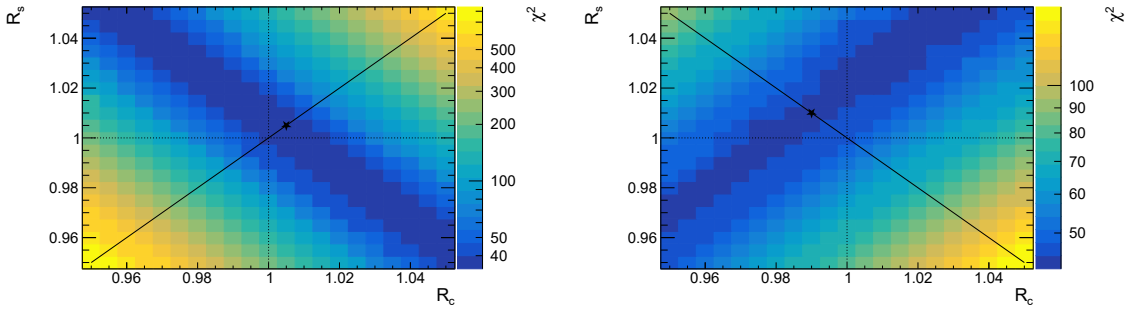


Figure 31: The number fraction of different mass jet pairs at the reconstructed level with scaling.

First, we need to verify the sources of the different ud and cs rates and scale the tagged pairs accordingly. To determine relevant scale factors for MC, we tried different values from 0.7 to 1.3. Even changes of 0.01 in the scale factors demonstrated changes in the chi-squared values. The initial scaling was applied, with a factor of 0.9 when the true flavor was cs and the tagged value was ud or cs. Similarly, a scaling factor of 0.8 was used when the true flavor was ud or x and the tagged value was ud or cs. Other combinations were left without scaling. While these scale factors provided an improvement, in a complete analysis, these efficiencies should be determined directly from the data itself.

After initial scaling, we determined scale factors for the JES. They were obtained through chi-squared (χ^2) minimization of scale factors in mass and momentum distribution for both strange (s) and charm (c) jets. Different scale factors from 0.95 to 1.05 with intervals of 0.005 were tested and plotted to identify the scale factors that



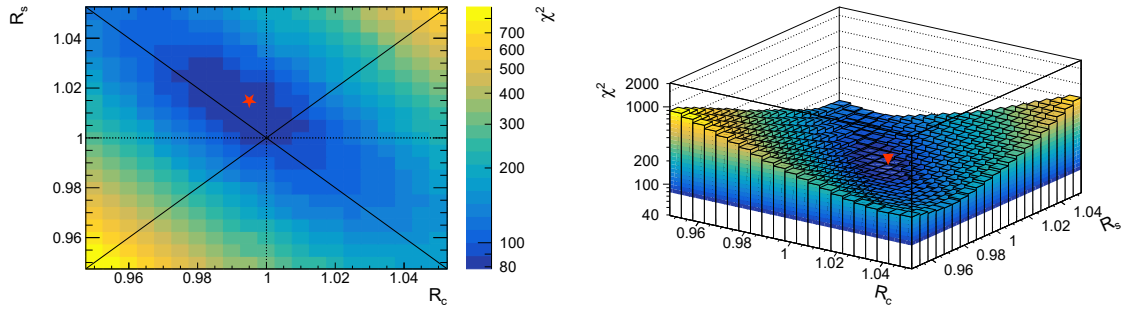
(a) Chi-squared minimization for mass scaling. Here the black line represents $R_s = R_c$ assumption to keep momentum differences unchanged.

(b) Chi-squared minimization for momentum scaling. Here the black line represents $R_s \cdot R_c = 1$ assumption to keep mass unchanged.

Figure 32: Chi-squared minimization plots for mass and momentum scaling.

minimized the χ^2 values for both masses and momenta. This process is illustrated in Figure 32, where the minimizing values are indicated with a star.

The scale factors need to be consistent for both mass and momentum, which is achieved by combining them, as illustrated in Figure 33. The red star marks the combined minimum for mass and momentum scales, determined to be 1.015 for strange jets and 0.995 for charm jets. The chi-squared minimum is 78.56, with 30 degrees of freedom (NDF) for the mass plot and 47 NDF for the p_T plot, resulting in a combined NDF of 77. With the chi-squared minimum being 78.56, this approach effectively determines the correct scale factors.



(a) Combined χ^2 minimization for mass and momentum (colormap).

(b) Combined χ^2 minimization for mass and momentum (3D view).

Figure 33: Combined χ^2 minimization plots for both mass and momentum scale factors.

8.5 Exploration of potential sources of bias from W-boson polarization and charm tagging efficiency

This Chapter explores the potential sources of bias from W-boson polatization and charm tagging efficiency relative to the momentum. Additionally, we discuss inefficiencies in the bottom jet tagging process.

Figure 34 illustrates how often the selected working point with c tag discriminator > 0.43 is satisfied for jets of different true flavors. We observe that the c-tagging efficiency depends strongly on p_T, which biases the relative momentum of charm versus strange jets in the tagged sample. If the c-tagger's turn-on curve is faster or slower in data compared to MC, this will introduce a bias in the relative JES of charm and strange jets.

Figure 35a shows the angular separation in the transverse plane ($\Delta\phi$) for different jets. Up and charm quarks tend to have $\Delta\phi$ values closer to 0 than strange and down quarks, indicating that they align more with the W-boson momentum. This pattern is likely influenced by the polarization of the W-boson produced in top quark decays, quark helicity, and the angular momentum conservation.

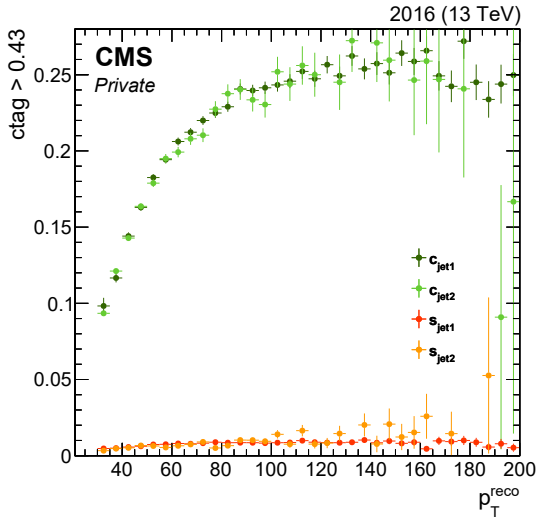
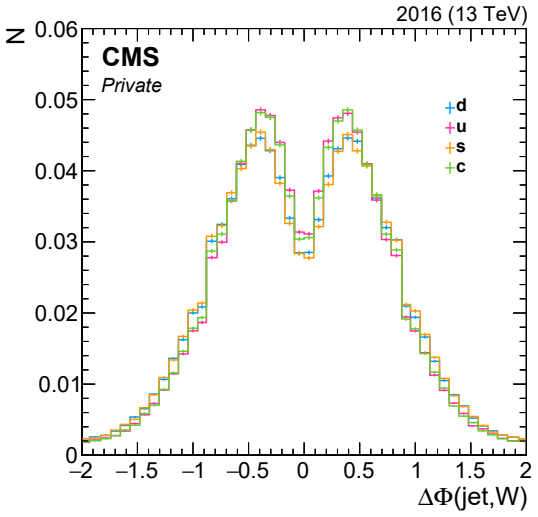
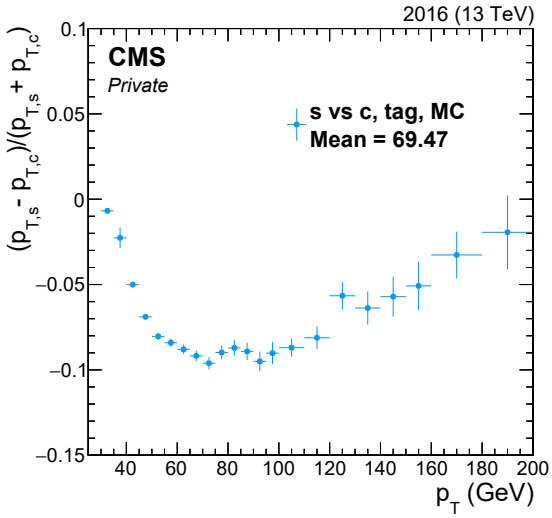


Figure 34: The fraction of jets that are tagged when c tag > 0.43 at the reconstructed level.



(a) The flying angles of different quarks compared to the W boson flight direction at the generator level.



(b) The average transverse momentum of strange relative to charm at the reconstructed level.

Figure 35: The effect of up-type quarks aligning better with the momentum of the W-boson.

Figure 35b shows that the charm has higher transverse momentum than strange, because charm is up-type quark and aligns better with the W boson momentum. It continues more frequently in the direction of the W boson's flight, while the strange quark tends to move backwards. This effect is most pronounced around 80 GeV. The c and u quarks inherit a significant portion of the W boson's momentum primarily due to the conservation of helicity, which dictates that the direction and spin of the particles are aligned in a certain way during the decay process. Additionally, the polarization of the W boson, whether it is longitudinal or transverse, further influences how the momentum is distributed between the decay products, contributing to the observed differences in the momentum inheritance between the c and s jets.

In the W-boson mass rest frame, both jets receive $m_W/2 \approx 40\,\text{GeV}$ momentum and when boosted along the momentum of the W, the jets gain an additional $p_{T,W}/2$. In extreme cases, jet along W direction will receive $p_{T,W}/2 + m_W/2$, while the other going anti-parallel $-p_{T,W}/2 + m_W/2$. If $p_{T,W} \approx m_W \approx 80\,\text{GeV}$, one jet will have $p_T \approx 80\,\text{GeV}$ and the other near 0 GeV. We lose these events because the softer jet has to be above 30 GeV, so this reduces the effect in Figure 35b seen at 80 GeV. This also likely explains the dip at $\Delta\phi \approx 0$ GeV in Figure 35a.

Figure 36 presents a comparative analysis of the p_T distributions across different jet flavors, with a particular focus on charm and strange jets, and their comparison to up and down jets. At the generator level, u and c quarks show similar behaviors, as do s and d quarks, although the latter group exhibits differences compared to u and c quarks. At the reconstruction level, however, charm tagging inefficiency at low p_T introduces a selection bias, causing charm jets to appear harder (higher p_T) and strange jets to appear softer (lower p_T) than they are at the generator level. This bias amplifies the small differences observed at the generator level. The figure includes multiple datasets: strange jets identified by true generator-level flavor in MC simulations, tagged simulations, as well as data, with similar datasets for charm

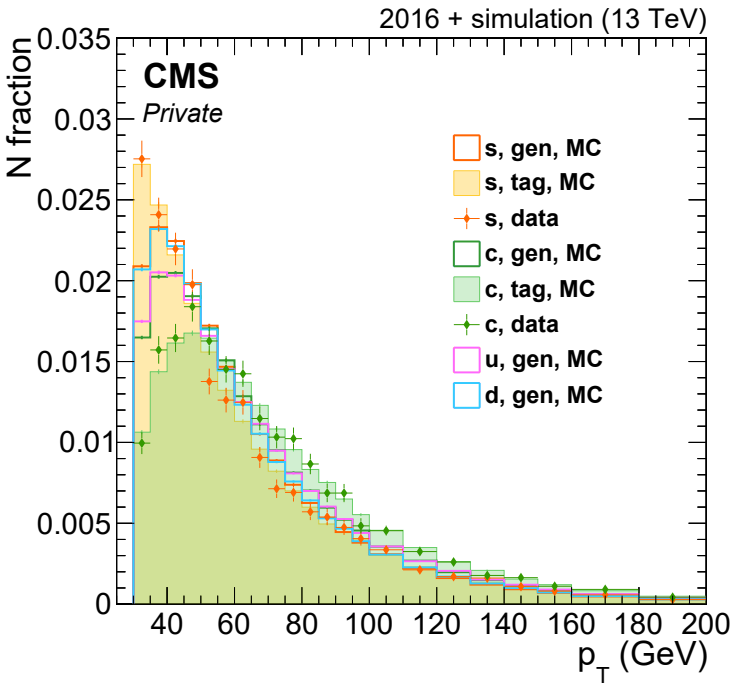


Figure 36: Comparison of p_T distributions between charm and strange quarks.

jets. The strong agreement observed between the simulations and the tagged data across all flavors (d, u, s, c) underscores the effectiveness of our tagging procedure.

The lepton+jet $t\bar{t}$ events used for this analysis were assumed to have four leading jets, of which exactly two are tagged as bottom (b-tagged). In this case, both b-jets must be correctly tagged, while the other two jets, representing either charm-strange (cs) or up-down (ud) pairs, are not b-tagged. The charm mistag rate as a b-jet is significantly higher than that for uds jets, often resulting in events where three jets are b-tagged due to charm mistagging, leading to event rejection. Occasionally, a b-tag may fail and a charm jet replaces it, but this typically also leads to event rejection due to inconsistent W-boson and top-quark masses.

However, the presence of more than four jets is common in lepton+jet events, often due to ISR or FSR. These additional jets can sometimes have higher transverse momentum (p_T) than the jets from W-boson decay and may be selected as one of the four leading jets. When the p_T of cs jets is lower than that of ud jets, this

swapping can happen more easily, causing cs jets to be replaced by ISR or FSR jets, thus cs jets may more often be misclassified compared to ud jets.

8.6 Verification of results with data/MC comparisons

Finally, we verify the strange JES with data/MC comparison. Here we introduce mass and momentum distibutions for jet pairs tagged with the previously defined c tag discrimator and we scale them using the efficiency scale factors and JES.

The chi-squared values for mass distribution are illustrated in Figure 37, highlighting the importance of scaling with explicit chi-squared and NDF values. These plots are the sum of the differently tagged samples from the tables 41 and 42 presented in Appendix D. The initial scaling, explained in Chapter 8.4, was applied to verify the sources of the different ud and cs rates, with factors of 0.9 (true flavour cs, tagged as cs or ud) and 0.8 (true flavor ud or x, tagged as cs or ud). Other combinations were left without scaling.

After this initial scaling, an additional scaling was applied based on JES corrections, derived from the chi-squared plot shown in Figure 33a. From that plot, we obtained scale factors of 1.015 for strange jets and 0.995 for charm jets. The combined scaling factors (efficiency + JES) are illustrated in the (eff. + JES) graph. The χ^2 to NDF ratio indicates that the efficiency based scaling improves the agreement between the simulation and the data for both cs and ud category. The JES scaling further improves the match in cs category.

The methods described above can also be applied to study the p_T distributions for different jet pairs. In Figure 38, the differences between simulation, scaled with efficiencies, scaled with efficiencies and JES, and data are further illustrated. These are again the sum of differently tagged samples, presented in Figures 43 and 44. Here, we can again see from the χ^2 to NDF that scaling improves the match in both cs and ud categories.

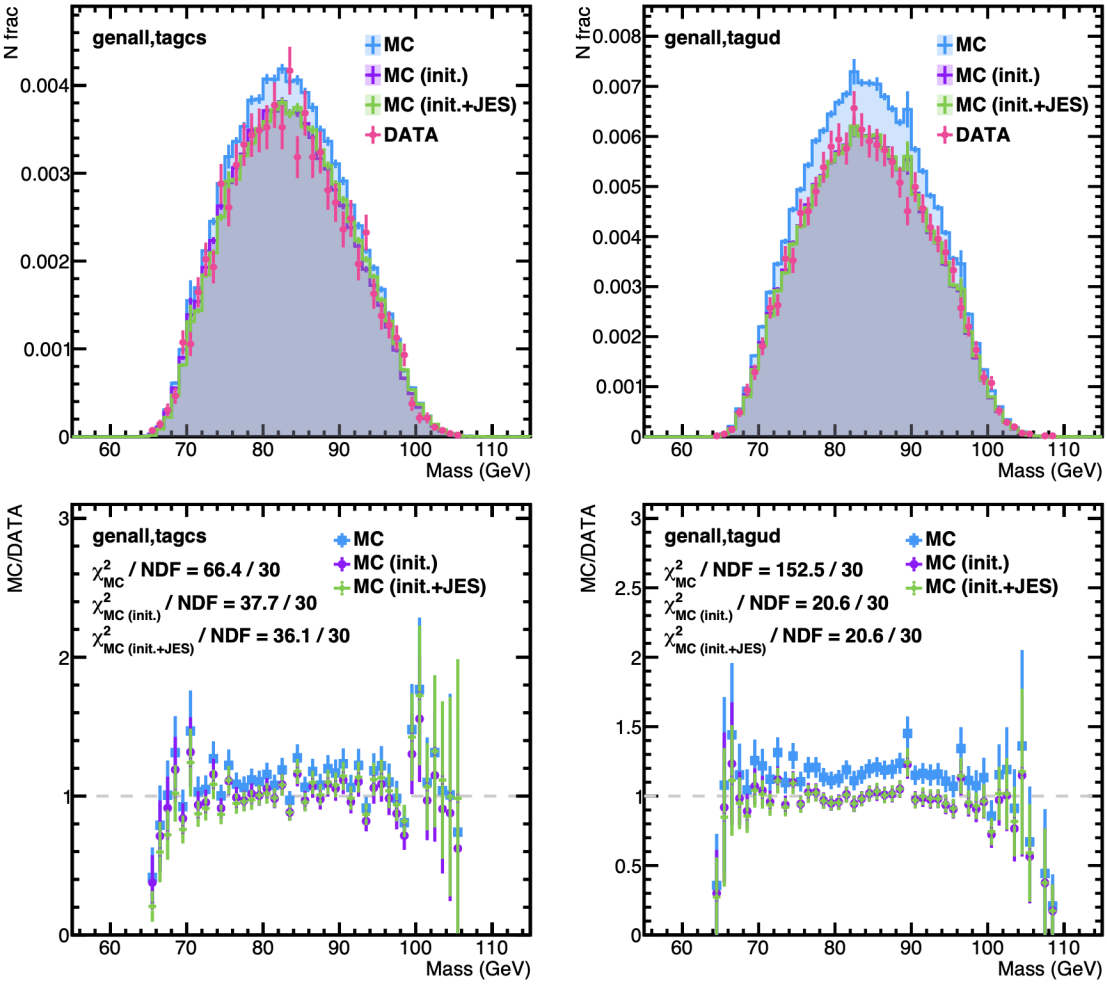


Figure 37: Chi-squared comparison for different mass scaling approaches for tagged as cs and ud -categories.

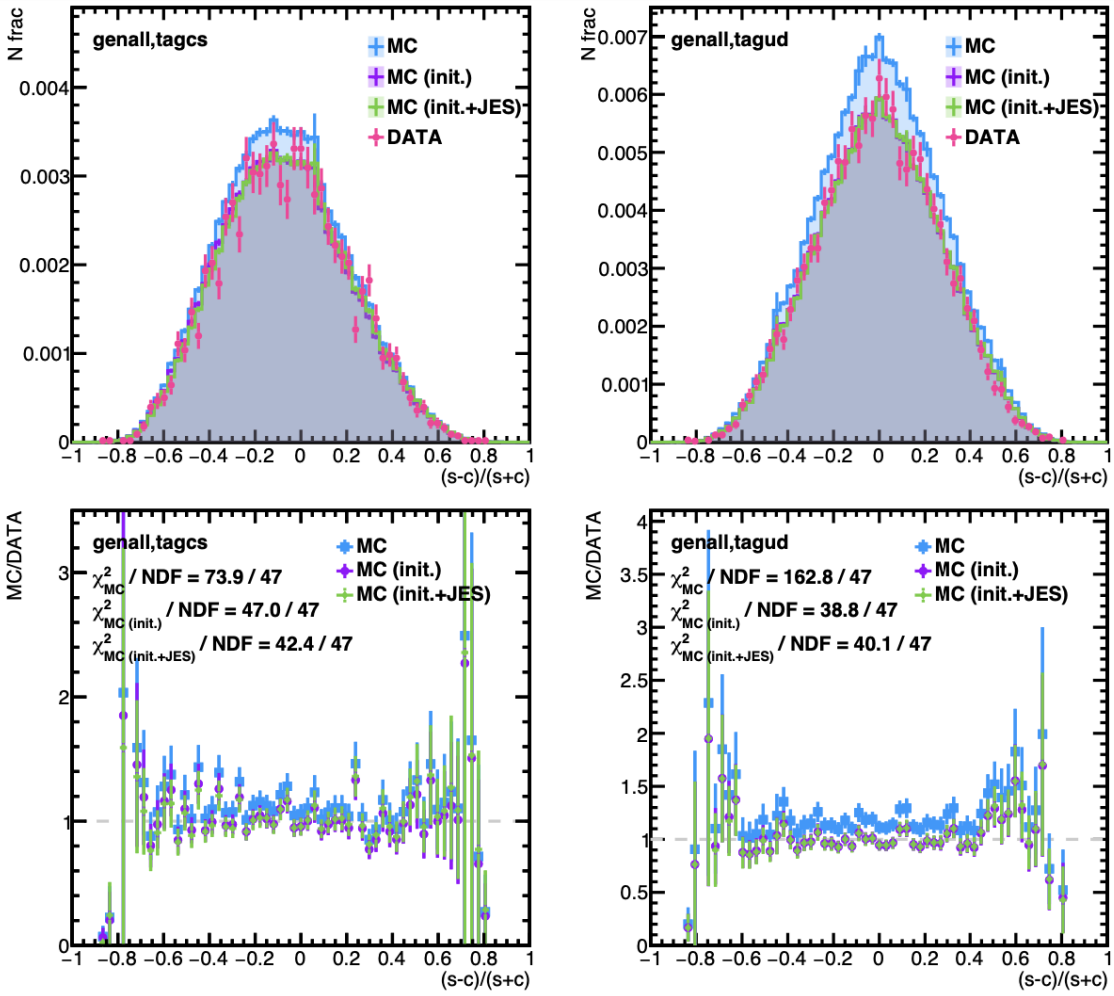


Figure 38: Further illustration of p_T distribution differences and χ^2 comparison for tagged as cs and ud -categories.

9 Conclusions

In this thesis, we explored two major aspects of strange jets, their properties and jet energy scale, through a combination of MC simulations and data-based methods. We also presented a comprehensive literature review on strange jets. These investigations contribute to the understanding of s-jets and their jet energy scale, and, to our knowledge, represent the first all-around study of strange jets.

Firstly, we demonstrated a clear difference in particle content between strange jets and up/down jets. Strange jets have a significantly higher proportion of neutral kaons (from K_L^0 and K_S^0) compared to up and down jets, particularly in leading particles, which carry a high fraction of the jet's transverse momentum (p_T). This characteristic offers a concrete path forward for tagging strange jets.

Secondly, we developed a novel method to determine the jet energy scale for strange (and charm) jets from data, which is important to calibrate all jet flavours. This includes methods to tune charm tagging efficiency, data-based scale factor derivation, and verifying the result using W-boson mass peaks and event rates across different tagging categories.

ROC curves were used to find optimal working points for tagging cs pairs with a charm tagger and ud pairs with an anti-charm tagger. Scale factors were then derived to correct for b-tagging and c-tagging efficiencies, ensuring proper normalization across tagging categories and accurately estimating jet purities. By combining data from multiple tagging categories and observables (mass and p_T difference), we obtained a quantitative estimate of the strange (and charm) jet energy scales, from which the s-jet JES and its uncertainties were extracted. Biases, such as helicity effects and p_T-dependent charm tagging efficiency, were addressed, and the results were verified by comparing scaled MC simulations to experimental data. This work lays the foundation for future studies aimed at improving jet flavor tagging and energy calibration, particularly for strange jets.

A key future application of this work would be the further refinement of the s-jet energy scale with larger dataset and more extensive set of systematic uncertainties, for which this thesis provides a proof-of-principle. A promising avenue for exploration is the use of $W \to cs$ decays as a laboratory to test and improve strange jet tagging techniques, which could lead to significant advancements in strange jet identification. A long-term goal would be the detection of the Higgs boson decaying to an $s\bar{s}$ pair, particularly following the successful establishment of $H \to c\bar{c}$ decays. However, this goal is very challenging due to the extremely small branching ratio and the difficulties involved in tagging strange jets at that level of precision, and it could be out of reach even for the HL-LHC.

A Detailed datasets

This section provides a detailed summary of the datasets used in the analysis. The signal datasets were employed for jet property analysis and the strange jet energy scale, while the background datasets were used only for building a dataset to study the latter. The tables below outline the dataset names, their purpose, and relevant descriptions.

Dataset Name	**Purpose**	**Description**
Run3 QCD PFNano Simulation	Jet property analysis	/QCD_PT-15to7000_TuneCP5 _13p6TeV_pythia8/ Run3Summer23MiniAODv4 -castor_130X_mcRun3_2023 _realistic_v14-v1/ MINIAODSIM
Run2 Simulation	Strange jet tagging	/TTToSemiLeptonic_TuneCP5 _13TeV-powheg-pythia8/ RunIISummer20UL16MiniAODv2- 106X_mcRun2_asymptotic_v17 -v1/MINIAODSIM
Run2 Data (Muon)	Strange jet tagging	/SingleMuon/Run2016F-UL2016 _MiniAODv2-v2/MINIAOD, /SingleMuon/Run2016G-UL2016 _MiniAODv2-v2/MINIAOD, /SingleMuon/Run2016H-UL2016 _MiniAODv2-v2/MINIAOD
Run2 Data (Electron)	Strange jet tagging	/SingleElectron/Run2016F- UL2016_MiniAODv2-v2/MINIAOD, /SingleElectron/Run2016G- UL2016_MiniAODv2-v2/MINIAOD, /SingleElectron/Run2016H- UL2016_MiniAODv2-v2/MINIAOD

Table I: Summary of signal datasets used for jet property analysis and strange jet energy scale studies.

Dataset Name	**Description**
2l2nu	Two-lepton and two-neutrino $t\bar{t}$ events.
allhad	All-hadronic $t\bar{t}$ events.
st_tWt	Single-top tW-channel.
st_tWat	Single-top tW-channel (alternative).
st_tct4f	Single-top t-channel (4-flavor).
st_tcat4f	Single-top t-channel (alternative, 4-flavor).
st_sc4f	Single-top s-channel (4-flavor).
dyjets10	Drell-Yan with jets (mass 10-50 GeV).
dyjets50HT*	Drell-Yan with jets (mass above 50 GeV, with different H_T cuts).
wjetsHT*	$W+$ jets production, with different H_T cuts.
ww	WW production.
wz	WZ production.
zz	ZZ production.
qcdmu*	QCD multijet events in the muon channel.
qcdel*	QCD multijet events in the electron channel.
qcdbctoel*	QCD b/c-to-electron processes.
ttzqq	$t\bar{t}Z \to q\bar{q}$ production.
ttzll	$t\bar{t}Z \to \ell\bar{\ell}$ production.
ttwqq	$t\bar{t}W \to q\bar{q}$ production.
ttwll	$t\bar{t}W \to \ell\bar{\ell}$ production.

Table II: Summary of background datasets used for creating a dataset for strange jet energy scale studies.

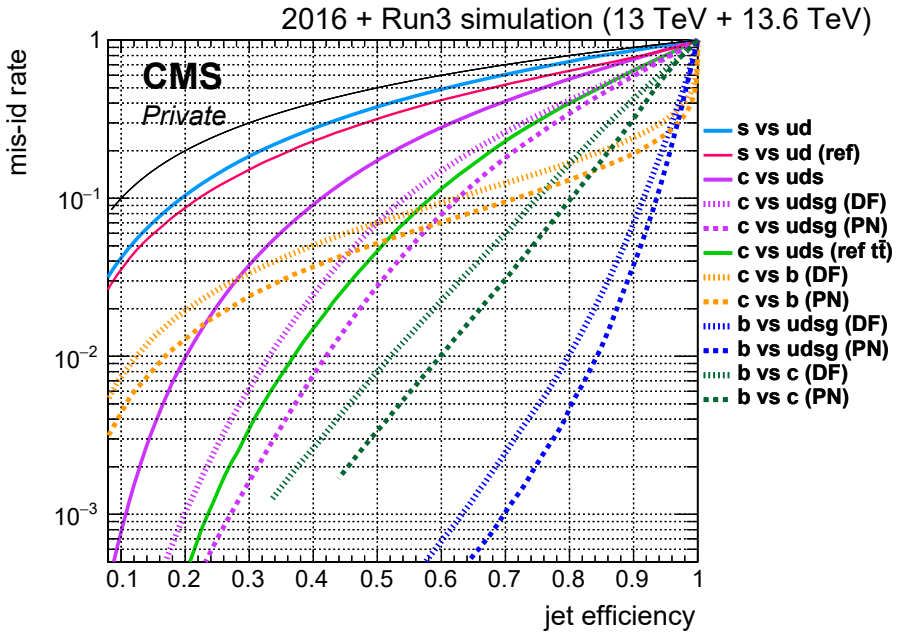


Figure 39: Combined ROC curves for different tagging methods are shown. The blue and purple solid lines represent the ROC curves derived from the Run2 2016 $t\bar{t}$ sample using the c-tag discriminator. The dotted lines correspond to DeepFlavor (DF), and the dashed lines represent ParticleNet (PN), both derived from the Run3 QCD multijet sample for jets with $p_T > 30$ GeV and $|\eta| < 2.5$. The first reference for strange vs. up/down is from [76], and the second for charm vs. up/down/strange is from [75], based on fully hadronic $t\bar{t}$ events. The black line represents random selection (diagonal).

B ROC curves for ParticleNet and DeepFlavour

Figure 39 shows that the PNet discriminates c vs uds better than DeepFlav. However, in contrast, the Run 3 DeepFlavour shows a decline in performance when distinguishing charm jets from uds jets compared to the Run 2 $t\bar{t}$ discrimator [75].

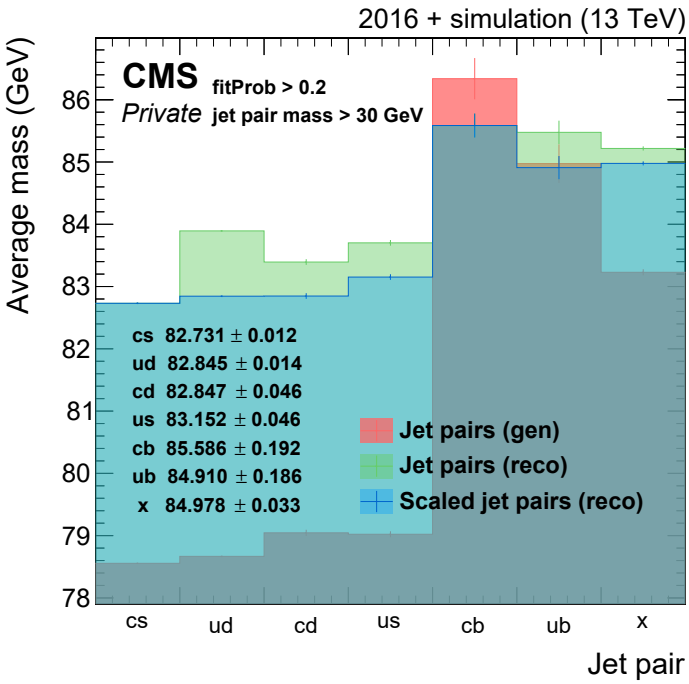


Figure 40: The average masses of different jet pairs in generator and reconstructed level with scaled reco-level values shown explicitly.

C Average masses for different jet pairs

The mass of the ud jet pairs was scaled with a scale factor of 0.986, applied separately to each jet in the ud pair. This scale factor was derived from the relation between responses to account for the variation between u, d jets and c, s jets, explained in Chapter 7.3. This scaling was also applied to u and d jets in other pairs. Figure 40 shows the average masses of different jet pairs at both the reconstructed and generator levels. The 0.986 scale factor is applied specifically to correct the overcalibrated ud-jet response, bringing it closer to the true value. The QCD di-jet sample, from which the centrally-produced jet energy corrections were derived from, contained a large proportion of gluon jets, which lowered the overall response, causing the ud jets to be overcalibrated. Since cs jets have a response between that of ud and gluon jets, their calibration remains relatively accurate and close to the true value.

D Mass and momentum distribution tables

The mass peak distribution as a table is illustrated in Figures 41 and 42. The different categories divided by c tag discriminator are shown in rows, and the true flavors are displayed in columns. The last row is the sum of each column, and the last column in 42 is the sum of each row. The figure shows the values for simulation (MC), scaled with initial scaling, and scaled with initial scaling and JES. The rightmost column also includes data since it has the c tag discriminator information but no true flavors. The momentum distribution is illustrated in Figures 43 and 44. The logic previously mentioned holds here as well, and once again.

In Figures 45 and 46 the chi-squared values for different mass and p_T scaling approaches for categories tagged as x and all are illustrated. This is the sum of the differently tagged samples from the table 42, presented with explicit chi-squared and NDF values.

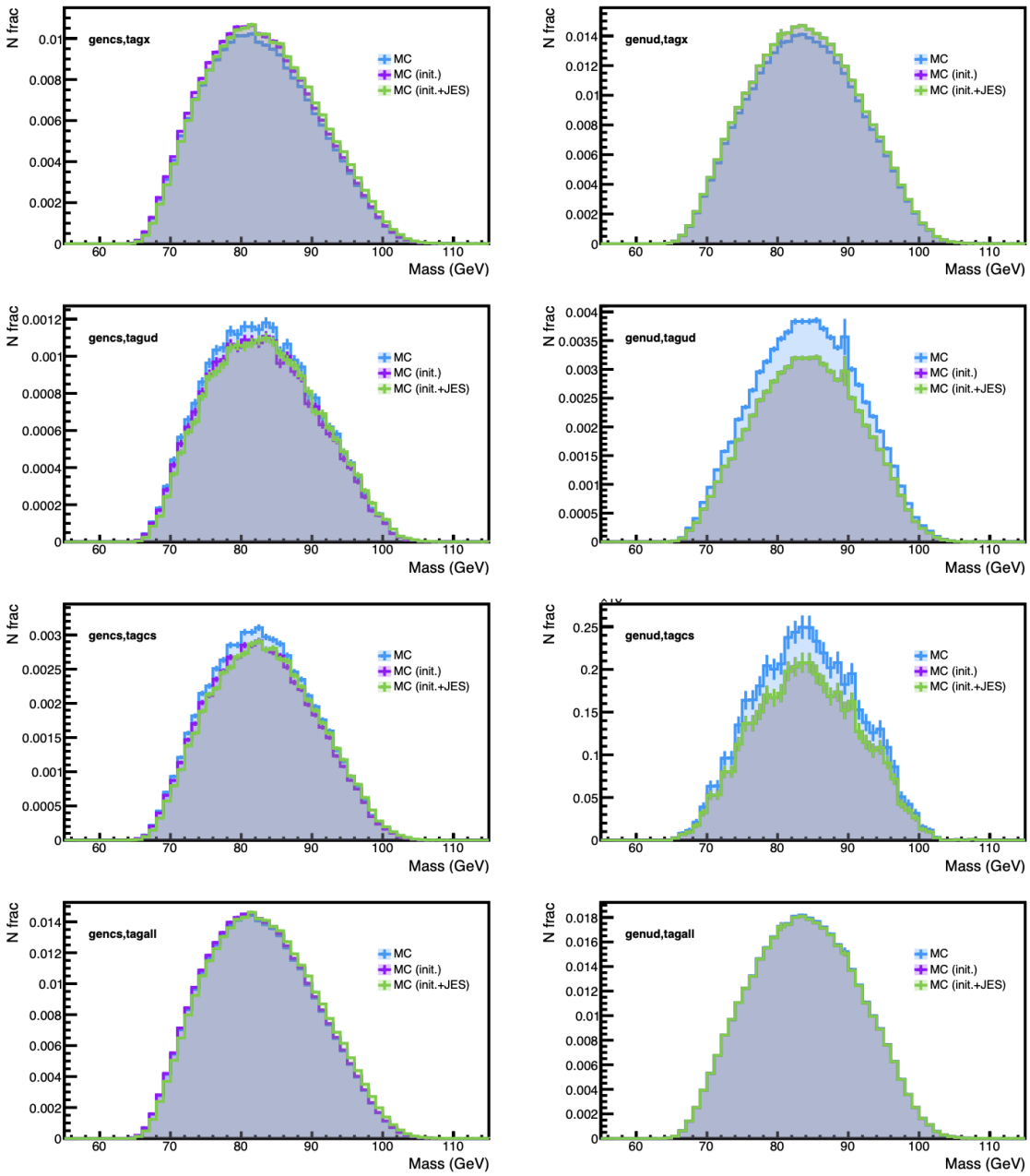


Figure 41: Comparison of simulation and data for mass distributions with different scaling factors applied.

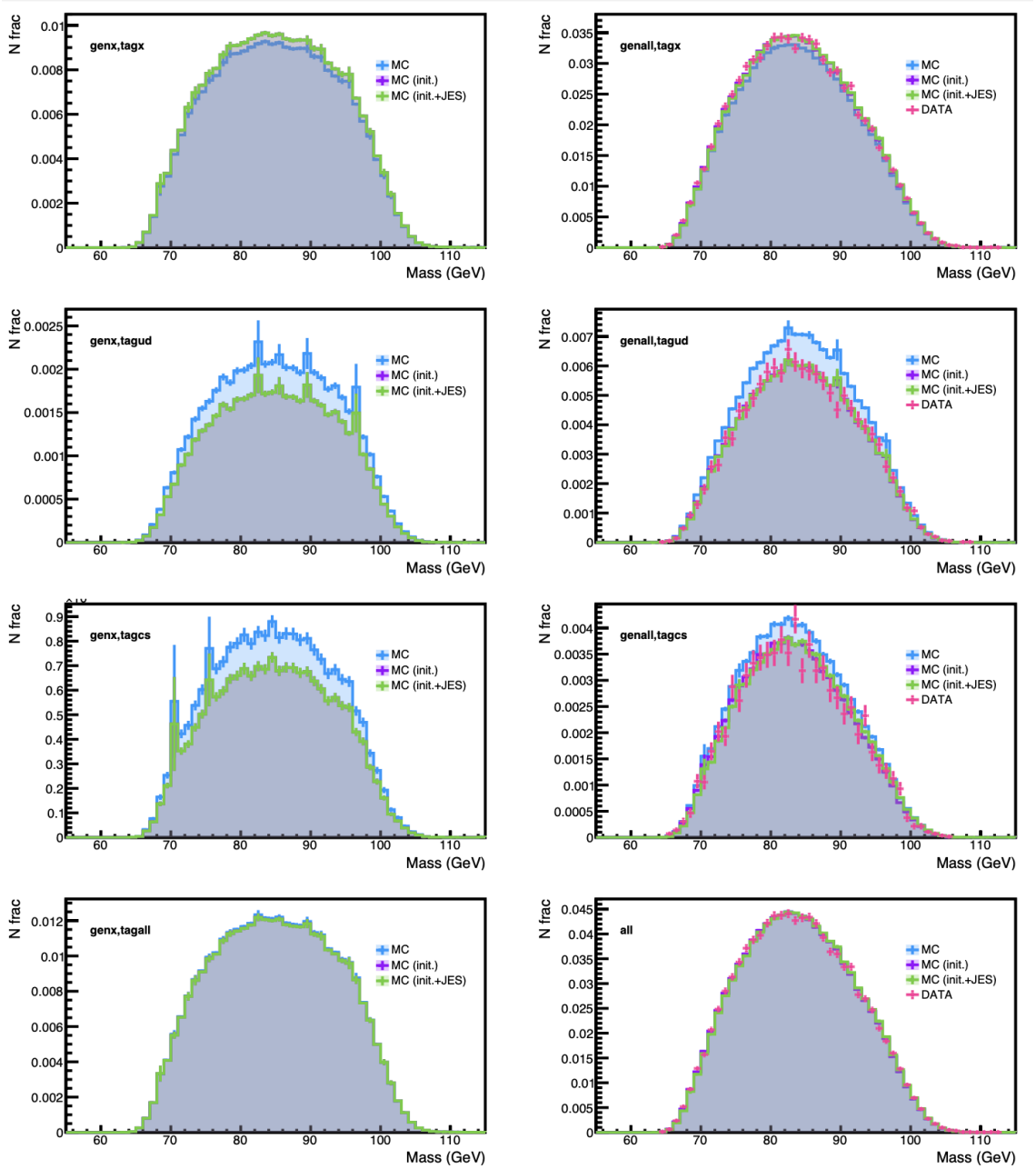


Figure 42: Comparison of simulation and data for mass distributions with different scaling factors applied.

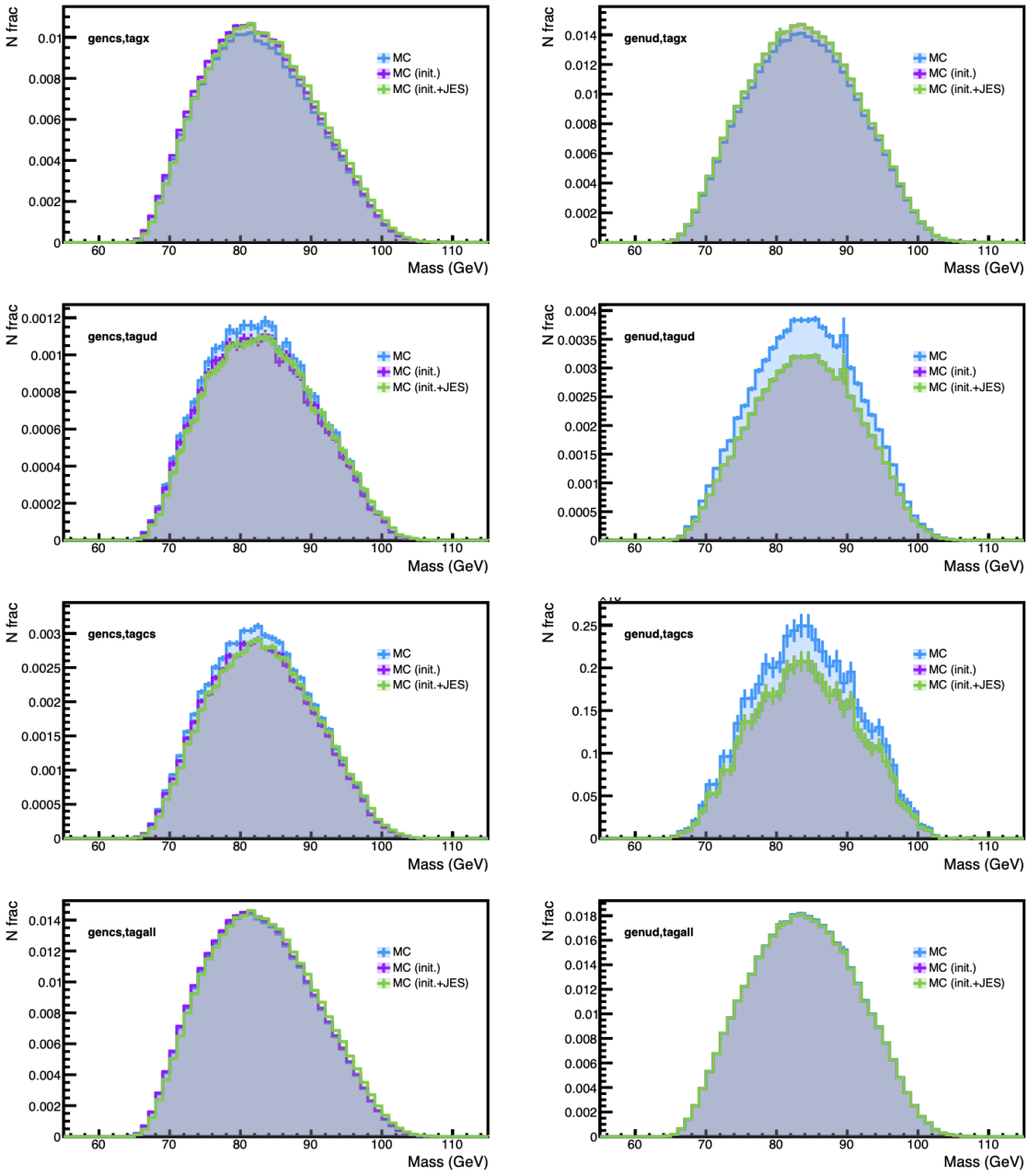


Figure 43: Comparison of p_T distributions between simulation and data with scaling.

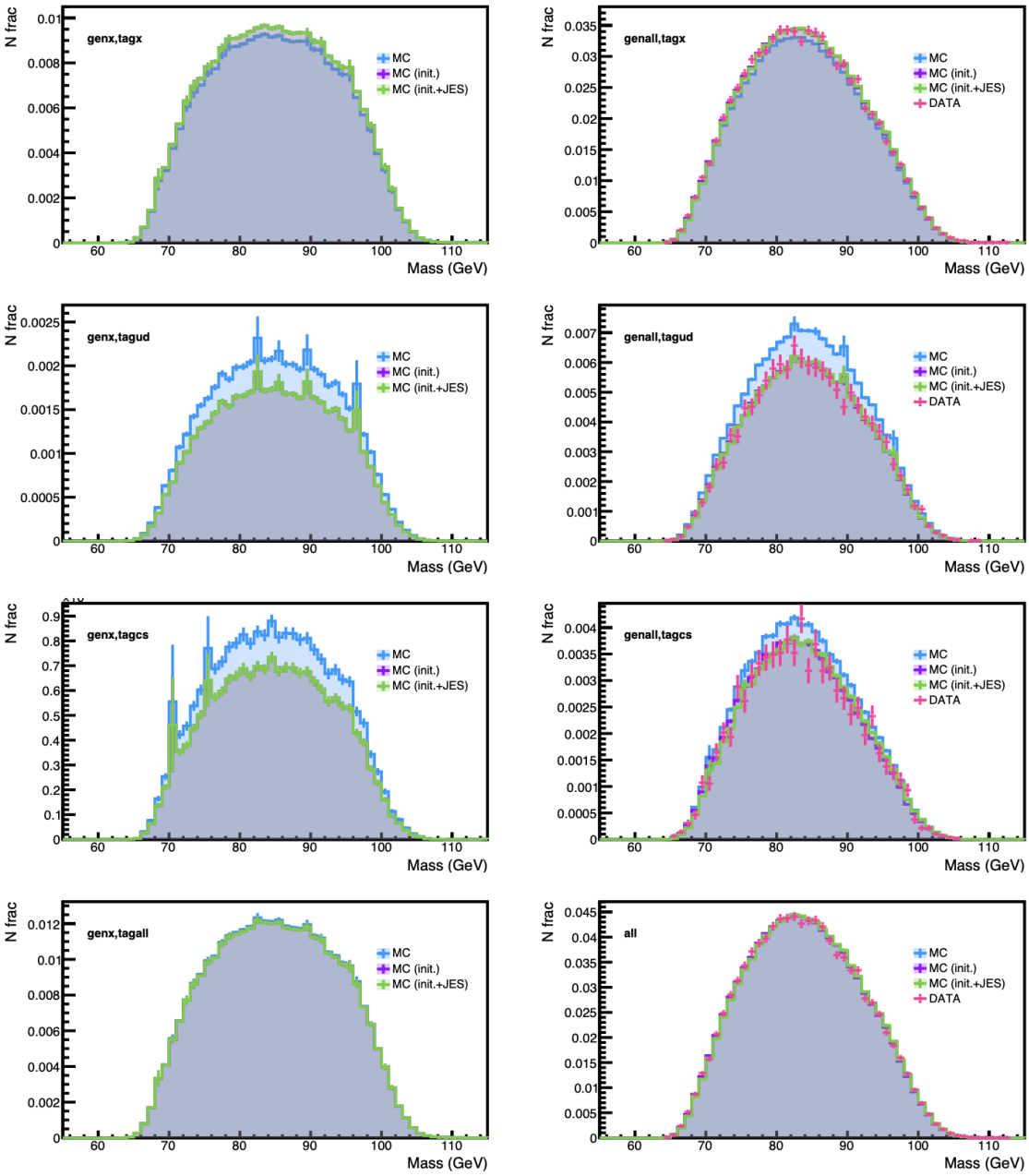


Figure 44: Comparison of p_T distributions between simulation and data with scaling.

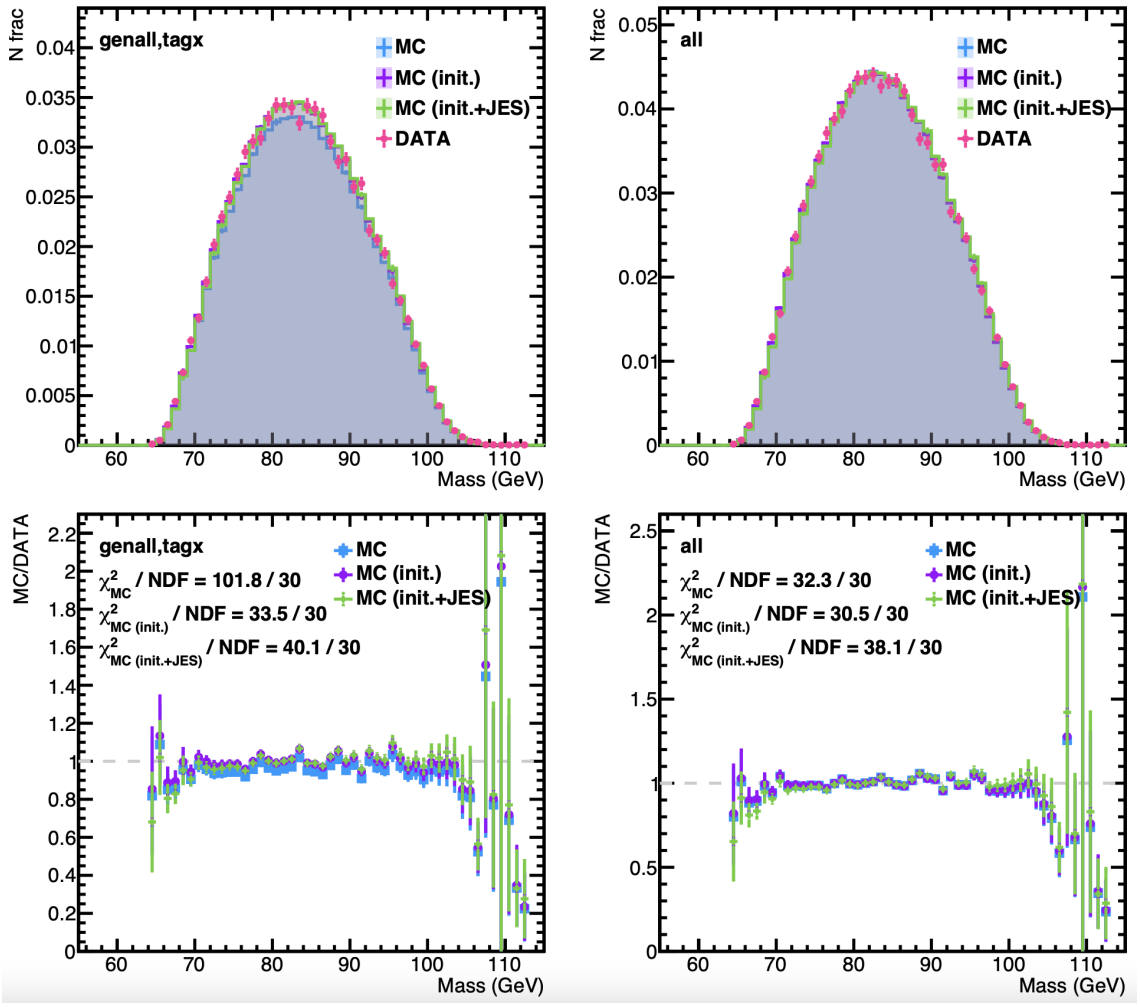


Figure 45: Chi-squared comparison for different mass scaling approaches for tagged as x and all -categories.

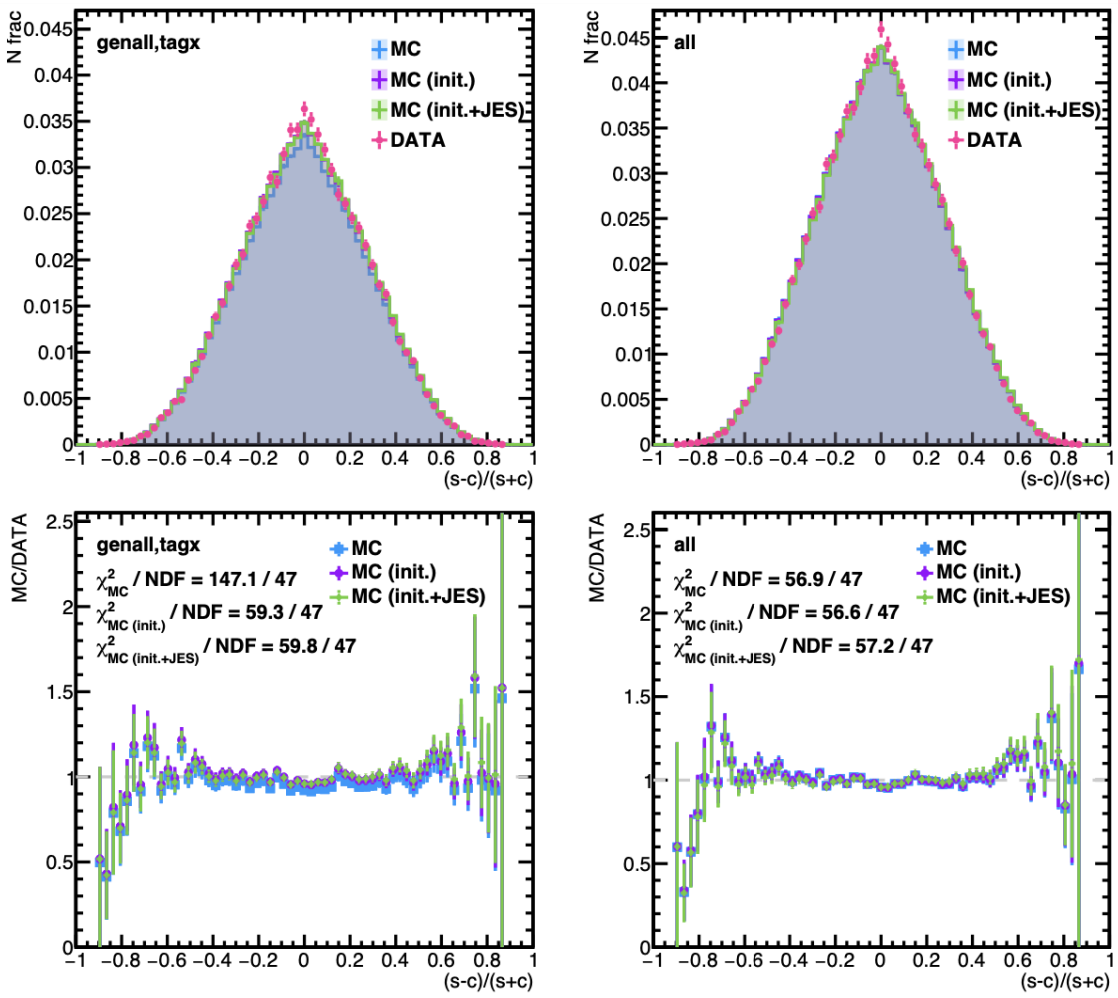


Figure 46: Further illustration of p_T distribution differences and χ^2 comparison for tagged as x and all -categories.

E The usage of AI in this thesis

In this thesis, the large language model ChatGPT-4o, developed by OpenAI, was employed as a tool to assist with the writing process. ChatGPT-4o was used to create initial drafts of sections based on the provided input, such as subsections and specific points to focus on, but these drafts were always subject to editing and verification. While AI tools can generate substantial amounts of text quickly, it was necessary to carefully check and refine the output to ensure that the text was accurate and relevant to the context. Additionally, ChatGPT-4o was used to check spelling and grammar throughout the text, with care taken to preserve the original meaning. This helped improve the clarity and correctness of the language.

References

[1] I. Neutelings. Cms coordinate system. https://tikz.net/axis3d_cms/. [Online; accessed 27-November-2023].

[2] Cush. *Standard Model of Elementary Particles*. Wikimedia Commons, PBS NOVA [1], Fermilab, Office of Science, United States Department of Energy, Particle Data Group. https://upload.wikimedia.org/wikipedia/commons/0/00/Standard_Model_of_Elementary_Particles.svg, 2019. [Online; accessed 12th November, 2023].

[3] T. Sjöstrand et al. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11):852–867, June 2008.

[4] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Westview Press, 1995. Reading, USA: Addison-Wesley (1995) 842 p.

[5] J. Campbell, J. Huston, and F. Krauss. *The Black Book of Quantum Chromodynamics : a Primer for the LHC Era*. Oxford University Press, 2018.

[6] M. D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3 2014.

[7] A. Andreassen et al. Scale-invariant instantons and the complete lifetime of the standard model. *Physical Review D*, 97(5), March 2018.

[8] G. Degrassi et al. Higgs mass and vacuum stability in the standard model at nnlo. *Journal of High Energy Physics*, 2012(8), August 2012.

[9] S. Alekhin et al. The top quark and higgs boson masses and the stability of the electroweak vacuum. *Physics Letters B*, 716(1):214–219, September 2012.

[10] Jets at cms and the determination of their energy scale. CMS Experiment at CERN, 2012. Available online at https://cms.cern/news/jets-cms-and-determination-their-energy-scale.

[11] M. Gell-Mann. Isotopic Spin and New Unstable Particles. *Phys. Rev.*, 92:833–834, 1953.

[12] M. Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8:214–215, 1964.

[13] G. Zweig. An SU(3) model for strong interaction symmetry and its breaking. Version 1. 1 1964.

[14] L. R. Evans and P. Bryant. LHC Machine. *JINST*, 3:S08001, 2008. This report is an abridged version of the LHC Design Report (CERN-2004-003).

[15] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[16] O. S. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[17] Serge D. Map of the Geneva region and of the LHC.. Carte de la region lemanique avec l'emplacemen du LHC. AC Collection. Legacy of AC. Pictures from 1992 to 2002., 1997.

[18] Max Degtyarev. The Large Hadron Collider Beauty Experiment. https://www.behance.net/gallery/67661801/The-incident-at-the-Large-Hadron-Collider-experiment, 2018. [Online; accessed 5-August-2024].

[19] CMS Collaboration. Cms luminosity public results. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, 2024. Accessed: 2024-09-12.

[20] E. Mobs. The CERN accelerator complex. Complexe des accélérateurs du CERN. 2016. General Photo.

[21] S. R. Davis. Interactive Slice of the CMS detector. 2016.

[22] T. Sakuma. Cutaway diagrams of CMS detector. 2019.

[23] The CMS Collaboration. Measurement of the wcq/wq' decay branching fraction ratio in proton-proton collisions at $\sqrt{s} = 13$ tev. CMS Physics Analysis Summary SMP-24-009, CERN, July 2024. Available on the CERN CDS information server.

[24] The CMS Collaboration. *The CMS experiment at the CERN LHC*, volume 3. aug 2008.

[25] C. Bierlich et al. A comprehensive guide to the physics and usage of pythia 8.3, 2022.

[26] J. Bellm et al. Herwig 7.0/herwig++ 3.0 release note. *The European Physical Journal C*, 76(4), April 2016.

[27] T. Gleisberg et al. Event generation with sherpa 1.1. *Journal of High Energy Physics*, 2009(02):007–007, February 2009.

[28] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), July 2014.

[29] A. D. Martin et al. Parton distributions for the lhc. *The European Physical Journal C*, 63(2):189–285, July 2009.

[30] A. Buckley et al. General-purpose event generators for lhc physics. *Physics Reports*, 504(5):145–233, July 2011.

[31] G. Corcella et al. Herwig 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *Journal of High Energy Physics*, 2001(01):010–010, January 2001.

[32] The CMS Collaboration. Extraction and validation of a new set of cms pythia8 tunes from underlying-event measurements. *The European Physical Journal C*, 80, Jan 2020.

[33] D. J. Lange et al. Upgrades for the CMS simulation. *J. Phys. Conf. Ser.*, 608(1):012056, 2015.

[34] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[35] F. Beaudette. The cms particle flow algorithm, 2014.

[36] The CMS Collaboration. Particle-flow reconstruction and global event description with the cms detector. *Journal of Instrumentation*, 12(10):P10003, 2017.

[37] S. H. Laurila. Search for Charged Higgs Bosons Decaying to a Tau Lepton and a Neutrino with the CMS Experiment, 2019.

[38] G. Petrucciani, A. Rizzi, and C. Vuosalo. Mini-aod: A new analysis data format for cms. *Journal of Physics: Conference Series*, 664(7):072052, December 2015.

[39] M. Cacciari et al. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, April 2008.

[40] S. Catani et al. Longitudinally invariant K_t clustering algorithms for hadron hadron collisions. *Nucl. Phys. B*, 406:187–224, 1993.

[41] A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging. Technical report, CERN, Geneva, 2009.

[42] CMS Collaboration. Pileup mitigation at cms in 13 tev data. *Journal of Instrumentation*, 15(09):P09018–P09018, September 2020.

[43] D. Bertolini et al. Pileup per particle identification. *Journal of High Energy Physics*, 2014(10), October 2014.

[44] CMS Collaboration. Jet energy scale and resolution in the cms experiment in pp collisions at 8 tev. *Journal of Instrumentation*, 12(02):P02014–P02014, February 2017.

[45] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the cms tracker. *Journal of Instrumentation*, 9(10):P10009–P10009, October 2014.

[46] The CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002–P11002, November 2011.

[47] Y. Nakai, D. Shih, and S. Thomas. Strange jet tagging, 2020.

[48] A. Albert et al. Strange quark as a probe for new physics in the higgs sector, 2022.

[49] W. Jang et al. Measuring $|V_{ts}|$ directly using strange-quark tagging at the lhc. *Journal of the Korean Physical Society*, 81(5):377–385, August 2022.

[50] H. Qu and L. Gouskos. Jet tagging via particle clouds. *Physical Review D*, 101(5), March 2020.

[51] H. Qu, C. Li, and S. Qian. Particle transformer for jet tagging, 2024.

[52] C. Loizides and A. Morsch. Apparent strangeness enhancement from multiplicity selection in high energy proton-proton collisions, 2021.

[53] P. Cui. Production of strange hadrons in jets and underlying events in pp and p–pb collisions with ALICE. *Physica Scripta*, 97(5):054009, apr 2022.

[54] F. Antinori. Strangeness, charm and beauty in quark matter: SQM 2007 experimental overview. *Journal of Physics G: Nuclear and Particle Physics*, 35(4):044055, mar 2008.

[55] Richard D. Ball et al. Parton distributions for the lhc run ii. *Journal of High Energy Physics*, 2015(4), April 2015.

[56] G. Aad et al. Determination of the strange quark density of the proton from ATLAS measurements of the $W \to \ell\nu$ and $Z \to \ell\ell$ cross sections. *Phys. Rev. Lett.*, 109:012001, 2012.

[57] F. Faura et al. The strangest proton? *The European Physical Journal C*, 80(12), dec 2020.

[58] S. Alekhin, J. Blümlein, and S. Moch. Strange sea determination from collider data. *Physics Letters B*, 777:134–140, feb 2018.

[59] N. Sato et al. Strange quark suppression from a simultaneous monte carlo analysis of parton distributions and fragmentation functions. *Phys. Rev. D*, 101:074020, Apr 2020.

[60] A Tumasyan et al. Measurements of the associated production of a w boson and a charm quark in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$. *The European Physical Journal C*, 82(12):1094, December 2022.

[61] S. J. Brodsky and B. Ma. The quark-antiquark asymmetry of the nucleon sea. *Physics Letters B*, 381(1–3):317–324, July 1996.

[62] C. Bhavesh. Using supernova neutrinos to probe strange spin of proton with juno and theia, 2022.

[63] T. Klähn and D. B. Blaschke. Strange matter in compact stars. *EPJ Web of Conferences*, 171:08001, 2018.

[64] X. G. Wang et al. Strange-quark asymmetry in the proton in chiral effective theory. *Physical Review D*, 94(9), November 2016.

[65] V. Barone et al. A new global analysis of deep inelastic scattering data. *The European Physical Journal C*, 12(2):243–262, January 2000.

[66] Bazarko et al. and A. O. CCFR Collaboration. Determination of the strange quark content of the nucleon from a next-to-leading-order QCD analysis of neutrino charm production. *Zeitschrift für Physik C Particles and Fields*, 65(2):189–198, June 1995.

[67] G. P. Zeller et al. Erratum: "effect of asymmetric strange seas and isospin-violating parton distribution functions on $\sin^2\theta_W$ measured in the nutev experiment" [phys. rev. d 65, 111103 (2002)]. *Phys. Rev. D*, 67:119902, Jun 2003.

[68] D. Mason et al. Measurement of the nucleon strange-antistrange asymmetry at next-to-leading order in qcd from nutev dimuon data. *Phys. Rev. Lett.*, 99:192001, Nov 2007.

[69] L. Cunqueiro et al. Dead-cone searches in heavy-ion collisions using the jet tree. *Physical Review D*, 107(9), May 2023.

[70] ALICE Collaboration. Direct observation of the dead-cone effect in quantum chromodynamics. *Nature*, 605(7910):440–446, May 2022.

[71] R. Klingenberg. Experimental Searches for Strange Quark Matter. *AIP Conference Proceedings*, 644(1):259–269, 11 2002.

[72] S. Bailey et al. Parton distributions from LHC, HERA, tevatron and fixed target data: MSHT20 PDFs. *The European Physical Journal C*, 81(4), apr 2021.

[73] E. Quercigh and J. Rafelski. A strange quark plasma. *Physics World*, 13(10):37, oct 2000.

[74] M. Hanafy et al. Strangeness enhancement at lhc energies using the thermal model and eposlhc event-generator, 2021.

[75] E. Bols et al. Jet flavour classification using deepjet. *Journal of Instrumentation*, 15(12):P12012–P12012, December 2020.

[76] CMS Collaboration. A unified approach for jet tagging in run 3 at $\sqrt{s} = 13.6$ tev in cms. Performance Note CMS DP - 2024/066, CMS, July 2024. Available on CMS information server.