UNIVERSITY OF IOANNINA
SCHOOL OF NATURAL SCIENCES
PHYSICS DEPARTMENT

Multijet cross section measurements up to the multiplicity of five jets with data from the CMS experiment at the LHC in proton - proton collisions at the center of mass energy of $13\ TeV$**.**

Argyro Ziaka

MSc under the supervision of Professor Panagiotis Kokkas

Ioannina, 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΦΥΣΙΚΗΣ

Μετρήσεις ενεργών διατομών παραγωγής πιδάκων σωματίων πολλαπλότητας εώς πέντε με δεδομένα του πειράματος CMS του LHC σε αλληλεπιδράσιες πρωτονίου - πρωτονίου με ενέργεια στο κέντρο μάζας $13\ TeV$**.**

Αργυρώ Ζιάκα

Μεταπτυχιακή εργασία υπό την επίβλεψη του καθηγητή: Παναγιώτη Κόκκα

Ιωάννινα, 2023

Ευχαριστίες

Πολλά είναι τα άτομα που συνείσφεραν, ο καθένας με τον δικό του ιδιαίτερο τρόπο, ώστε να έλθει εις πέρας η παρούσα διπλωματική εργασία. Κατ΄ αρχήν, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου Παναγιώτη Κόκκα, για την εποπτεία που παρείχε, την καθοδήγηδη και την αστείρευτη βοήθεια και στήριξη. Επίσης, ευχαριστώ τον κύριο Ιωάννη Παπαδόπουλο, για την πολύτιμη βοήθειά του όσον αφορά την υλικοτεχνική υποδομή, αλλά και τον ρόλο που έπαιξε κατά την πρώτη μου επαφή με την C++ και το ROOT. Βαθιά είναι και η ευγνωμοσύνη μου για τους Πάρη Γιαννείο και Πολυδάμα Κοσμόγλου, οι οποίοι υπήρξαν συνεργάτες, αρωγοί και φίλοι. Ήδη από την πρώτη στιγμή άφιξής μου στον τομέα της Φυσικής Υψηλών Ενεργειών, μέχρι και σήμερα, βρισκόταν δίπλα μου ανελλιπώς, να με συμβουλεύουν και να με διδάσκουν. Πέραν όμως από τον επαγγελματικό τομέα, καθοριστικό ρόλο έπαιξε και η συνεισφορά της οικογενείας μου, συγκεκριμένα της μητέρας μου, Άννας, και των αδελφών μου, Δημήτρη, Ζαχαρούλα και Μαρία, μέσω της ανιδιοτελούς στήριξης που μου προσέφεραν, όλα τα χρόνια σπουδών μου. Ιδιαίτερης αναφοράς χρήζει και η κολλητή μου φίλη, Ελένη, η οποία υπήρξε ανελλιπώς στο πλευρό μου τα τελευταία 10 χρόνια, παρέχοντάς μου συμβουλές και στήριξη. Τέλος, ευχαριστώ από καρδιάς τους κοντινούς μου ανθρώπους από το ευρύτερο οικογενειακό περιβάλλον καθώς και τους στενούς μου φίλους που, εκούσια είτε ακούσια, και ο καθένας με τον τρόπο του, με βοήθησαν να ανταπεξέλθω.

Acknowledgements

Apart from the people that I have already mentioned, I would like to express my deep gratitude and appreciation to Patrick L.S. Connor, whose contribution to the development of the current analysis has been substantial. Not only did he develop the overall analysis framework, but also he has always been reachable and willing to selflessly offer his assistance.

Περίληψη

Στα πλαίσια της παρούσας διπλωματικής εργασίας πραγματοποείται η μέτρηση των ενεργών διατομών παραγωγής πιδάκων σωματίων πολλαπλότητας εώς πέντε. Η μέτρηση βασίζεται σε δεδομένα του πειράματος CMS του LHC σε αλληλεπιδράσεις πρωτονίου - πρωτονίου με ενέργεια στο κέντρο μάζας $\sqrt{s} = 13\ TeV$ αντιστοιχούν σε ολοκληρωμένη λαμπρότητα $59.82\ fm^{-1}$. Τόσο οι ενεργές διατομές, όσο και οι αντίστοιχοι λόγοι τους, συγκρίνονται με γεννήτορες Monte Carlo που περιλαμβάνουν τις διαδικασίες του καταιγισμού παρτονίων, της αδρανοποίησης και των πολυπαρτονικών αλληλεπιδράσεων. Οι λόγοι των ενεργών διατομών είναι ευαίσθητοι στην σταθερά των ισχυρών αλληλεπιδράσεων και επιτρέπουν τον υπολογισμό της.

Abstract

A measurement of the inclusive differential multijet cross sections with multiplicity up to five and their ratios is presented. The analysis is based on data from proton-proton collisions collected by the CMS experiment at LHC with a centre-of-mass energy of 13 TeV and correspond to an integrated luminosity of 59.82 fm^{-1}.The cross section measurements, as well as their ratios, are compared with several Monte Carlo event generators that can successfully simulate the processes of parton shower, hadronisation and multiparton interactions. The multijet cross section ratios are sensitive to the strong coupling constant and enable its determination.

Contents

Chapter 1

Standard Model of Particle Physics

Particle Physics describes through its methods, as well as its content, the reduction of complicated phenomena to simpler principles and structures that can be utilised in order to develop a theoretical description in a much deeper level. Particle Physics is mainly occupied by the description of the characteristics and properties of the elementary particles along with their interactions.

The scientific knowledge that came up by a weary and a long lasting collaborative effort is summed up to the so-called Standard Model (SM), a mathematical theory that successfully describes the electromagnetic, the weak and the strong nuclear interactions, that comprise three out of the four kinds of the fundamental forces that rule nature, the fourth being gravity. The above interactions, manifested via what are considered today as elementary particles of matter, i.e. leptons and quarks, are described by a basic symmetry principle, upon which the SM is based, called "gauge symmetry", that is established by two theories: the Quantum Theory and the Theory of Special Relativity. Gravity, however, is not incorporated to the SM and is described by the Theory of General Relativity [1] [2].

1.1 Elementary Particles and Interactions

The elementary particles, i.e. the particles that lack inner structure, are separated in two categories: matter particles, that make up matter, and force mediators, the carriers of the fundamental forces.

Matter particles are divided into leptons and quarks. Both kinds of particles are fermions with spin 1/2 and do not contain internal structure in distances as small as $\Delta x \approx 10^{-18}$ m, distances that modern accelerators, including the LHC, are able to study. Leptons are further subdivided

into three generations and are subject to electromagnetic interactions. Note that only charged leptons take part in such interactions. There are 6 leptons classified in these three generations according to their charge Q and their lepton number L. These six leptons are the electron (e), the electron neutrino (ν_e), the muon (μ), the muon neutrino (ν_μ), the tau (τ) and the tau neutrino (ν_τ). The electron, muon and tau have charge -1, with mass between a couple hundred keV and a few GeV, while their respective neutrinos are neutrally charged and exhibit tiny mass. Taking into consideration their anti-particles, there are 12 leptons in total. Thus, to each lepton corresponds an anti-lepton with the same mass, but opposite charge and lepton number. In a similar manner, there are 3 generations of quarks and 6 flavors: up (u), down (d), strange (s), charm (c), top (t) and bottom (b), with electric charge $+2/3$ or $-1/3$, and their mass spectrum ranges from a few MeV to some hundreds of GeV. They can take part in electromagnetic, weak and strong interactions. Quarks are also characterised by another quantum number called color charge. Each quark can exhibit one of three colors (red, green, blue), that are combined in order to form colorless objects. Therefore, there are $3 \times 6 = 18$ quarks; including their anti-quarks, 36, and thus, the total number of matter particles sums up to 48 fermions.

The interactions among matter particles are carried out by the exchange of virtual quanta, each being linked to one of the fundamental forces. This way, the virtual particles convey the forces among the matter particles, and can be characterised as the quanta of each field. These virtual quanta are the force mediators, being vector bosons, i.e. bosons with spin 1. The electromagnetic force is being mediated by the photon (γ) which is massless and electrically neutral, the weak force by electrically charged $W^\pm$ and neutral Z, which mass is around 80 GeV and 91 GeV respectively. The above bosons combined, compose the mediators of the electroweak interactions, i.e. the electromagnetic interactions (QED) and the Quantum Theory of weak interactions. The strong force is mediated by eight massless and electrically neutral gluons g, that compose Quantum Chromodynamics (QCD), and similarly to quarks, also carry color charge, thus, they are able to self-interact. The number of force mediators is 12 bosons totally.

The SM is completed with the addition of the Higgs boson, a scalar boson with spin 0 and mass around 125 GeV. This particle is predicted by the Brout-Englert-Higgs Mechanism. The Higgs boson was discovered by the CMS and ATLAS experiments at CERN in 2012, and it is associated with the Higgs field that prevails the universe, providing the $W^\pm$ and Z bosons, as well as the fermions, with mass. In total, there are 61 elementary particles in the SM discovered so far.

The theoretical framework composing the Standard Model, as well as Particle Physics, is Quantum Field Theory (QFT). This theory is based on the concept that every particle is associated with a field and all interactions result from exchange in particles. QFT is a combination of Quantum Theory and Special Relativity, describing the elementary particles and their interactions through mathematical formulations.

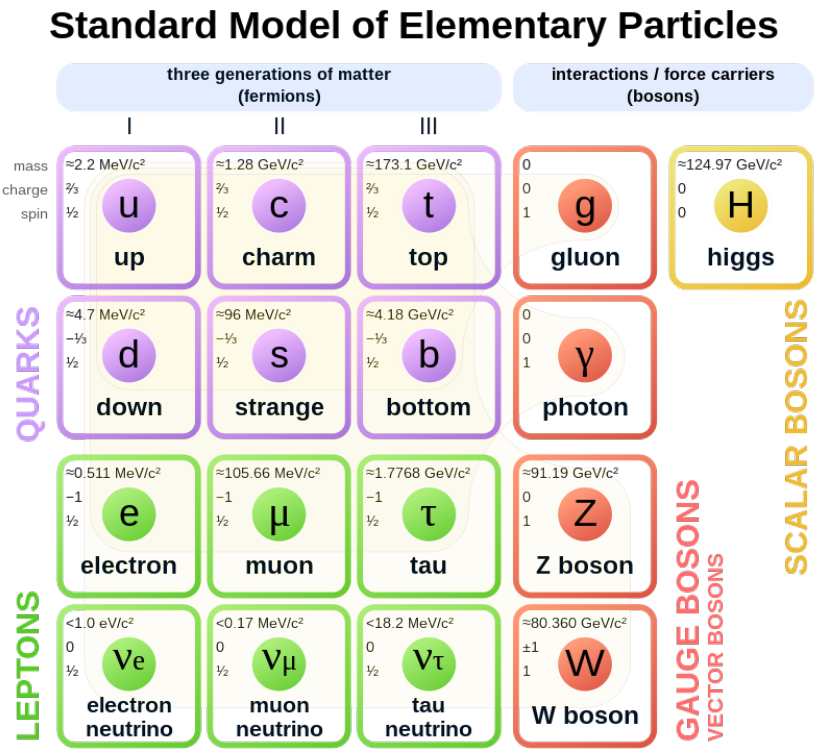


Figure 1.1: *The elementary particles of the Standard Model* [3]

1.2 Lagrangian Formulation and Gauge Theories

When studying classical, but also quantum fields, the use of a Lagrangian formalism is needed, with the equations of motion deriving from the least-action principle, where the action is described as a function of a Lagrangian. In classical particle mechanics the dynamics of a system are condensed into a Lagrangian function L:

$$L = T - V \tag{1.1}$$

where T: is the kinetic energy and V: the potential of the system. The Lagrangian is a function of the q coordinates and their time derivatives:

$$L = L\left(q(t),\ q(\dot{t})\right) \tag{1.2}$$

The advantage of the Lagrangian formulation is that selecting a specific system of coordinates is not mandatory. The above Lagrangian can be used to define another quantity, the action S:

$$S\left[q\right] = \int_{t_1}^{t_2} L\left(q(t),\ q(\dot{t})\right) dt \tag{1.3}$$

with S being a function of $q(t)$. The least-action principle declares that the actual path a particle follows between two points $q(t_1)$ and $q(t_2)$ is the one that minimises the action, i.e. is the path of least action. By considering a variation of the form $q(t) \to q'(t) = q(t) + \delta q(t)$ and by demanding stationary action ($\delta S = 0$), one can produce the Euler-Lagrange equations of motion:

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = 0 \tag{1.4}$$

In a similar manner, the Lagrangian formulation for any field theory (classical or quantum) becomes a function of a Lagrangian density $\mathcal{L}$:

$$\mathcal{L} = \mathcal{L}\left(\phi,\ \partial_\mu \phi\right) \tag{1.5}$$

where the Lagrangian density containing ϕ represents any field $\phi(x)$. This way, one will no longer calculate the position of localised particles as a function of time; on the contrary they will calculate fields ϕ as functions of position and time. The derivatives are defined as:

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \left(\frac{1}{c}\frac{\partial}{\partial t},\ \overrightarrow{\nabla}\right), \quad \partial^\mu \equiv \frac{\partial}{\partial x_\mu} = \left(\frac{1}{c}\frac{\partial}{\partial t},\ -\overrightarrow{\nabla}\right) \tag{1.6}$$

where x^μ and x_μ are the position 4-vectors:

$$x^\mu = \left(x^0, x^1, x^2, x^3\right), \quad x_\mu = \left(x^0, -x^1, -x^2, -x^3\right) \tag{1.7}$$

with $x^0 = ct$ being the time-like component. In this case, the action takes the following form:

$$S[\phi] = \int dt \int d^3x \mathcal{L}\left(\phi,\ \partial_\mu \phi\right) = \int d^4x \mathcal{L}\left(\phi,\ \partial_\mu \phi\right) \tag{1.8}$$

Again, by imposing the least-action principle, the equations of motion (Euler-Lagrange) for the field ϕ can be derived:

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} = 0 \tag{1.9}$$

Among others, there are three Lagrangians that produce the Euler-Lagrange equations: one for a free scalar ($spin = 0$), one for a spinor ($spin = 1/2$) and one for a vector ($spin = 1$) field.

1. **spin=0:** Suppose there is a real scalar field ϕ with mass m. This system is described by a Lagrangian density of the following form:

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \phi\right)\left(\partial^\mu \phi\right) - \frac{1}{2}m^2\phi^2 \tag{1.10}$$

The Euler-Lagrange equation that corresponds to this Lagrangian leads to the field equation of motion or Klein-Gordon equation:

$$\partial_\mu \partial^\mu \phi + m^2 \phi = 0 \tag{1.11}$$

2. **spin=1/2:** Assume that a spinor field ψ of mass m exists; then it is described by the Lagrangian density:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi \tag{1.12}$$

where $\bar{\psi}$ is the adjoint spinor: $\bar{\psi} = \psi^\dagger \gamma^0$ and $\gamma^\mu, \mu = 0, 1, 2, 3$ are the 4×4 Dirac matrices. ψ and $\bar{\psi}$ are being treated as separate fields and are determined in such a way, so that $\psi\psi^\dagger$ is relativistically invariant. By using the Euler-Lagrange equation, one ends up with the Dirac equation (of motion):

$$i\left(\partial_\mu \bar{\psi}\right)\gamma^\mu + m\bar{\psi} = 0 \tag{1.13}$$

3. **spin=1:** For a vector field A_μ with mass $m = 0$ (such as a photon) that is produced by a source j^μ, the Lagrangian density is:

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} - j^\mu A_\mu \tag{1.14}$$

In the equation above $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is the electromagnetic field tensor, that combines the electric and magnetic field. This time, the Euler-Lagrange equation gives:

$$\partial_\mu F^{\mu\nu} = j^\nu \tag{1.15}$$

which correspond to the Maxwell equations.
For a massive vector field V_μ the Lagrangian is:

$$\mathcal{L} = -\frac{1}{4} V^{\mu\nu} V_{\mu\nu} + \frac{1}{2} m^2 V_\mu V^\mu \tag{1.16}$$

from which the Proca equations can be derived:

$$\partial_\mu V^{\mu\nu} + m^2 V^\nu = 0 \tag{1.17}$$

The Lagrangian formulation plays an important role when it comes to studying symmetries. Symmetries are intertwined with conservation laws and thus, conserved quantities. This relationship is summed up in one of the most fundamental theorems of physics, the Noether Theorem,

put forward by Emmy Noether. The Noether Theorem states that for any continuous symmetry of the action S, there always exists a corresponding conserved quantity called current J_μ, which satisfies $\partial_\mu J_\mu = 0$. As a result, the existence of a symmetry means that the action will remain invariant under a symmetry transformation. Respectively, the Lagrangian will be invariant under a similar transformation of symmetry.
Gauge transformations are crucial in formulating a quantum theory. When the Lagrangian of a complex scalar field ϕ remains invariant under transformations of the form:

$$\phi \to \phi' = e^{+i\alpha}\phi, \quad \phi^* \to \phi'^* = e^{-i\alpha}\phi \tag{1.18}$$

where ϕ: is the field and ϕ^*: its complex conjugate, then these transformations are called global gauge transformations, because the parameter α is a constant, which means that they occur concurrently. There are also local gauge transformations, where the parameter α is a function of a time-space point $x^\mu : \alpha(x^\mu)$:

$$\phi \to \phi' = e^{+i\alpha(x)}\phi, \quad \phi^* \to \phi'^* = e^{-i\alpha(x)}\phi \tag{1.19}$$

When acting with a local gauge transformation, the Lagrangian is no longer invariant. However, the Lagrangian describing the interactions of the field ϕ with a vector field A^μ is still invariant under local gauge transformations. This constitutes a unique gauge transformation regarding local ones: $U = e^i\alpha(x)$, with $\alpha(x)$ being a parameter. This can be considered as 1×1 unitary matrices, which elements belong to the $U(1)$ group and therefore, the above transformation is a local gauge transformation, with the respective symmetry $U(1)$ being a gauge symmetry that is also associated with the preservation of the leptonic and baryon numbers. In general, most symmetries in physics are described by group theory. The most common groups in particle physics are the groups of unitary $n \times n$ matrices that compose the unitary group $U(n)$. By further restricting these unitary matrices into having determinant 1, leads to the special unitary group $SU(n)$. The number of parameters and generators of the group $U(n)$ is n^2, with rate n, whereas for the $SU(n)$ group, the respective number of parameters and generators is $n^2 - 1$, with rate $n - 1$.
In particle physics, internal symmetries, meaning symmetries that are not related to the external space-time, are being thoroughly used. There are three symmetries capable to describe the fundamental interactions of the known particles. The first of these symmetries is the $U(1)$ gauge symmetry, not necessarily equivalent to the $U(1)$ group of the electromagnetic interactions. The corresponding gauge field is B_μ, not associated with the photon A_μ, and the quantum number is not the electric charge, but the hypercharge Y and thus, it can be written as $U_Y(1)$. The second gauge symmetry is $SU(2)$, related to the weak interactions, and its quantum number is the *weak isotopic spin* T. The way the elementary particles behave in regard to the weak

isotopic spin is related to parity violation in weak interactions. The $SU(2)$ symmetry has the $W^i_\mu, i = 1, 2, 3$ gauge fields, it is associated to the weak interactions and takes the form: $SU_L(2)$. The product $SU_L(2) \otimes U_Y(1)$ describes the electroweak interactions, it was introduced for the first time in 1961 by Glashow and later supplemented in 1967 by Weinberg and in 1968 by Salam, who incorporated the spontaneous symmetry breaking and the Higgs Mechanism, so that the gauge and matter fields attain mass, whenever needed. The third symmetry is $SU(3)$, related to the strong interactions. Its respective quantum number is the color in regard only with quarks, each comprising of three colors, forming the fundamental representation with respect to $SU(3)$. The gauge fields of $SU(3)$ are the eight gluons $G^i_\mu, i = 1, 2, ..., 8$ and the corresponding gauge theory is Quantum Chromodynamics (QCD), written as $SU_C(3)$. Therefore, the gauge theories representing the SM is the product $SU_C(3) \otimes SU_L(2) \otimes U_Y(1)$ [1, 4].

1.3 Electroweak interactions and Higgs mechanism

The standard Electroweak theory is prescribed by the $SU_L(2) \otimes U_Y(1)$ gauge symmetry and in this frame, couplings with the fermions are mediated by the $W^\pm, Z$ vector bosons and the Higgs particle. For this purpose, the Lagrangian is split into two parts, both written down as dictated by the $SU_L(2) \otimes U_Y(1)$ gauge symmetry and renormalisability. The Higgs vacuum expectation value (VEV) provokes the spontaneous symmetry breaking, this way producing massive fermions and non vanishing vector bosons. Thus, by separating the Higgs term of the Lagrangian: [5, 6]

$$\mathcal{L} = \mathcal{L}_{gauge} + \mathcal{L}_{Higgs} \tag{1.20}$$

1.3.1 The Gauge Sector

The part $\mathcal{L}_{gauge}$ of the Lagrangian in Equation 1.20 that involves bosons and fermions is given by:

$$\mathcal{L}_{gauge} = -\frac{1}{4}\sum_{A=3}^{3} F^A_{\mu\nu}F^{A\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \bar{\psi}_L i\gamma^\mu D_\mu \psi_L + \bar{\psi}_R i\gamma^\mu D_\mu \psi_R \tag{1.21}$$

where $B_{\mu\nu}$ and $F_{\mu\nu}$ are the gauge antisymmetric tensors constructed out of B_μ and W^A_μ respectively. More specifically, $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ with the B_μ gauge field being associated with the $U(1)$ group and $F^A_{\mu\nu} = \partial_\mu W^A_\nu - \partial_\nu W^A_\mu - g\epsilon_{ABC}W^B_\mu W^C_\nu$ where W^A_μ corresponds to the three $SU(2)$ generators, ϵ_{ABC} are the group structure constants that for $SU(2)$ identify as the totally antisymmetric Levi-Civita tensor and g is the $SU(2)$ gauge coupling.

The fermion fields are given by their left- and right-handed components:

$$\psi_{L,R} = \left[\left(1 \mp \gamma_5\right)/2\right]\psi, \qquad \bar{\psi}_{L,R} = \bar{\psi}\left[\left(1 \pm \gamma_5\right)/2\right] \tag{1.22}$$

with γ_5 the Dirac matrix, used as the operator of chirality: $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. Any Dirac state ψ can be extracted by combining left- and right-handed chiral states of a particle and can be projected using projector operators of the form:

$$P_\pm = \left(1 \pm \gamma_5\right)/2 \tag{1.23}$$

that satisfy the relations: $P_\pm P_\pm = P_\pm, P_\pm P_\mp = 0, P_+ + P_- = 1$.
Since the standard EW theory is based on chirality, the left-handed ψ_L and right-handed ψ_R particle states exhibit different behavior under the gauge group, permitting parity and charge conjugation non conservation in EW interactions. Therefore, all fermion mass terms of the form $\bar{\psi}_L\psi_R + h.c.$ are not allowed in the symmetric limit. Particularly, in the Minimal Standard Model (MSM: i.e. the model that only includes all observed particles plus a single Higgs doublet), all left-handed particle states ψ_L are doublets of $SU(2)$ and all right-handed particle states ψ_R are singlets.
When there are no mass terms, only vector and axial vector interactions in the Lagrangian are observed, with the property of not mixing ψ_L and ψ_R. Fermion and $W^\pm, Z$ masses are introduced through the spontaneous symmetry breaking. The covariant derivatives $D_\mu\psi_{L,R}$ in Equation 1.21 are given as:

$$D_\mu\psi_{L,R} = \left[\partial_\mu + ig\sum_{A=1}^{3} t^A_{L,R}W^A_\mu + ig'\frac{1}{2}Y_{L,R}B_\mu\right]\psi_{L,R} \tag{1.24}$$

where $t^A_{L,R}$ is the generator of the $SU(2)$ symmetry in the reducible $\psi_{L,R}$ representations that satisfies the commutation relation of the form:

$$\left[t^A_L, T^B_L\right] = i\epsilon_{ABC}t^C_L, \qquad \left[t^A_R, T^B_R\right] = i\epsilon_{ABC}t^C_R \tag{1.25}$$

Similarly, $Y_{L,R}$, called hypercharge, is the $U(1)$ generator and it is related to the electric charge generator in units of e:

$$Q = t^3_L + \frac{1}{2}Y_L = t^3_R + \frac{1}{2}Y_R \tag{1.26}$$

The Lagrangian of Equation 1.21 is capable of producing all fermion couplings of the gauge bosons. The charged W_μ fields are described by $W^{1,2}_\mu$, whereas the Z_μ boson and the photon A_μ are produced from combinations of W^3_μ and B_μ. The Charged Coupling, i.e. interactions of the

$W^{\pm}$ bosons, can be derived taking the $W_{\mu}^{1,2}$ terms in Equations 1.20 and 1.24:

$$g\left(t^1 W_{\mu}^1 + t^2 W_{\mu}^2\right) = \left\{\left[\left(t^1 + it^2\right)/\sqrt{2}\right]\left[\left(W_{\mu}^1 - iW_{\mu}^2\right)/\sqrt{2}\right] + h.c.\right\} \tag{1.27}$$

$$= g\left\{\left[\left(t^+ W_{\mu}^-\right)/\sqrt{2}\right] + h.c.\right\} \tag{1.28}$$

where $t^{\pm} = t^1 \pm it^2$ and $W^{\pm} = \left(W^1 \pm iW^2\right)/\sqrt{2}$. By applying the generic relation to L and R, one obtains:

$$V_{\bar{\psi}\psi W} = g\bar{\psi}\gamma_{\mu}\left[(t_L^+/\sqrt{2})(1-\gamma_5)/2 + (t_R^+/\sqrt{2})(1+\gamma_5)/2\right]\psi W_{\mu}^- + h.c. \tag{1.29}$$

In the Neutral Current (NC) sector, that corresponds to the interactions of the Z boson and the photon, the photon A_{μ} and the mediator Z_{μ} of the weak NC are orthogonal and normalised linear combinations of B_{μ} and W_{μ}^3:

$$A_{\mu} = B_{\mu}\cos\theta_W + W_{\mu}^3 \sin\theta_W, \qquad Z_{\mu} = -B_{\mu}\sin\theta_W + W_{\mu}^3\cos\theta_W \tag{1.30}$$

The above two equations define the weak mixing angle θ_W. The W_{μ}^3 and B_{μ} terms are written as:

$$W_{\mu}^3 = A_{\mu}\sin\theta_W + Z_{\mu}\cos\theta_W, \qquad B_{\mu} = A_{\mu}\cos\theta_W - Z_{\mu}\sin\theta_W \tag{1.31}$$

Utilising the Equations 1.20, 1.24 and also the charge matrix Q, one takes:

$$gt^3 W_{\mu}^3 + g'Y/2B_{\mu} = \left[gt^3\sin\theta_W + g'(Q-t^3)\cos\theta_W\right]A_{\mu} + \left[gt^3\cos\theta_W - g'(Q-t^3)\sin\theta_W\right]Z_{\mu} \tag{1.32}$$

Equal couplings to left and right fermions with a strength equivalent to the electric charge characterises the photon and thus:

$$g\sin\theta_W = g'\cos\theta_W = e \quad \longrightarrow \quad \tan\theta_W = g/g' \tag{1.33}$$

Once θ_W is fixed by the photon couplings, the Z couplings can be derived:

$$V_{\bar{\psi}\psi Z} = \frac{g}{2\cos\theta_W}\bar{\psi}\gamma_{\mu}\left[t_L^3(1-\gamma_5) + t_R^3(1+\gamma_5) - 2Q\sin^2\theta_W\right]\psi Z^{\mu} \tag{1.34}$$

with $V_{\bar{\psi}\psi Z}$ being the notation for the vertex. By specialising in the case of doublet fields, i.e. four fermion interactions, the three-level relation of g can be obtained with the Fermi coupling

constant G_F precisely measured from muon decays:

$$\frac{G_F}{2} = \frac{g^2}{8m_W^2} \tag{1.35}$$

By recalling that $g \sin\theta_W = e$, the relation can be associated to the fine-structure constant of QED:

$$m_W \sin\theta_W = \left(\frac{\pi\alpha}{\sqrt{2}G_F}\right)^{1/2} \simeq 37.2802GeV \tag{1.36}$$

where $\alpha \equiv e^2/4\pi = 1/137.036$. In the above relations, m_W is the mass of the W boson. In order to derive the effective four-fermion interactions that are equivalent, at low energies, to the CC and NC couplings given by the Equations 1.29 and 1.34, the $W^{\pm}$ and Z masses are relatively large and are being provided by the $\mathcal{L}_{Higgs}$. As for the gauge bosons self-interactions, these can be derived from the $F_{\mu\nu}$ term of the $\mathcal{L}_{gauge}$ by using Equation 1.30 and $W^{\pm} = (W^1 \pm iW^2)/\sqrt{2}$. Defining a three-gauge-boson vertex, one obtains ($V \equiv \gamma, Z$):

$$V_{W^-W^+V} = ig_{W^-W^+V}\left[g_{\mu\nu}(p-q)_\lambda + g_{\mu\lambda}(r-p)_\nu + g_{\nu\lambda}(q-r)_\mu\right] \tag{1.37}$$

with

$$g_{W^-W^+\gamma} = g\sin\theta_W = e, \quad g_{W^-W^+Z} = g\cos\theta_W \tag{1.38}$$

The photon coupling to W is fixed by the electric charge, as imposed by QED gauge invariance. The ZWW coupling is larger by a $\cot\theta_W$ factor. This form of the triple gauge vertex is very special: in general, there could be departures from the above SM expression, even restricting us to Lorentz invariant, electromagnetic-gauge symmetric and C and P conserving couplings. The triple gauge vertices have been experimentally tested in the past, mainly at LEP and at the Tevatron, now also at the LHC.
The quadratic gauge coupling is proportional to $g^2\epsilon_{ABC}W^BW^C$
$\epsilon_{ADE}W^DW^E$, so in the term with $A = 3$ there are three charged W's. In a similar manner, for $A = 1$ or 2 there are 2 charged W's and 2 W^3's, with each W_3 being a combination of γ and Z according to Equation 1.31. Then, the quadratic vertex can take the form:

$$V_{WWVV} = ig_{WWVV}\left[2g_{\mu\nu}g_{\lambda\rho} - g_{\mu\lambda}g_{\nu\rho} - g_{\mu\rho}g_{\nu\lambda}\right] \tag{1.39}$$

where μ and ν refer to the W^+W^+ pair in the $4W$ vertex and to VV in the case of $WWVV$, whereas, $g_{WWWW} = g^2$, $g_{WW\gamma\gamma} = -e^2$, $g_{WW\gamma Z} = -eg\cos\theta_W$, $g_{WWZZ} = -g^2\cos^{\theta_W}$ The quadratic coupling, since it is quadratic to g, hence small, has not been directly tested so far. [5,6]

1.3.2 The Higgs Sector

In the SM, the spontaneous breaking of the electroweak $SU(2)\times U(1)$ symmetry is responsible for giving mass to the $W^{\pm}$ and Z bosons, as well as fermions, rendering the electroweak interactions to be short-ranged. Such a Lagrangian, which corresponds to the second term of Equation 1.20, is formulated as:

$$\mathcal{L}_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi^\dagger\phi) - \bar{\psi}_L\Gamma\psi_R\phi - \bar{\psi}_R\Gamma^\dagger\psi_L\phi^\dagger \tag{1.40}$$

where ϕ is a column vector that includes all Higgs fields:

$$\phi = \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix} \tag{1.41}$$

with ϕ^+, ϕ^0 being self-interacting doublets with 4 degrees of freedom and hypercharge $Y = 1$, transforming as a reducible representation of the gauge group. The quantities Γ, which take into consideration all coupling constants, are matrices that make the Yukawa couplings invariant under the Lorentz and gauge groups. Without loss of generality, Γ can be taken to be independent of γ_5. The potential $V(\phi\phi^\dagger)$ is symmetric under $SU(2)\otimes U(1)$ and, in order for the theory to be renormalisable, it contains quadratic, at most, terms of ϕ:

$$V(\phi\phi^\dagger) = -\mu^2\phi\phi^\dagger + \frac{1}{2}\lambda(\phi\phi^\dagger)^2 \tag{1.42}$$

where μ^2 and $\lambda > 0$ are real constants. The parameter λ must necessarily be positive, so that the potential will have a stable vacuum state [1, 6].
Spontaneous symmetry breaking is induced when the minimum of V is obtained for non-vanishing ϕ values. This minimum is a classical analogue of the quantum mechanical vacuum state and they are both states of minimum energy. When the quadratic term of the potential is negative, then the neutral component of the scalar doublet receives a vacuum expectation value (VEV) v:

$$\langle 0|\phi|0\rangle = v = \begin{pmatrix}0\\ v\end{pmatrix} \neq 0 \tag{1.43}$$

that causes spontaneous breaking of the SM symmetry $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ to $SU(3)_C \otimes U(1)_{cm}$. When $\mu^2 > 0$, the minimum of the potential is at zero: $|\phi| = \sqrt{\phi^\dagger\phi} = 0$. The total minimum of the theory is given when $\mu^2 < 0$, corresponding to the "Mexican Hat" potential, illustrated at Figure 1.2, as:

$$\phi^\dagger\phi = \frac{-\mu^2}{2\lambda} = \frac{v}{\sqrt{2}} \tag{1.44}$$

and it will determine the ground state, whereas the spontaneous symmetry breaking indicates the existence of a local or/and extensive symmetry in the system that does not abide by the ground state.

For the case of $\mu^2 < 0$, there is an infinite set of equivalent minima, that lie around a circle of radius $\sqrt{-\mu^2/2\lambda}$ and by choosing one of them, the rotational symmetry breaks spontaneously [6, 7]. Out of the 4 generators of $SU(2)_L \otimes U(1)_Y$, three will spontaneously break, implying that they cause transformations of the ground state and indicate the existence of three massless Goldstone bosons related to the 3 out of the 4 degrees of freedom of the Higgs field. The latter forms couplings with the W_μ and B_μ gauge fields of the electroweak symmetry, through the covariant determinant that appears at the kinetic term of $\mathcal{L}_{Higgs}$ (Equation 1.40):

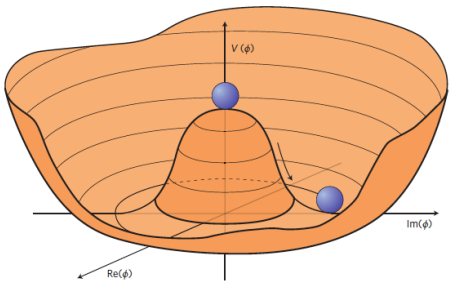


Figure 1.2: *Higgs potential for* $\mu^2 < 0$ *(Mexican hat)* [7]

$$D_\mu = \partial_\mu + ig\sum_{A=1}^{3} t^A W_\mu^A + ig'(Y/2)B_\mu \tag{1.45}$$

where t^A and $Y/2$ are the $SU(2)\otimes U(1)$ generators in the reducible representation spanned by ϕ. To the gauge boson masses contribute not only doublets, but all non-singlet Higgs representations. The VEV, when introducing a scalar field H, can be replaced by:

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v+H \end{pmatrix} \tag{1.46}$$

Then, the potential unavoidably takes the following form:

$$V = -\frac{v^2\mu^2}{2} + \mu^2 H^2 + \frac{\mu^2}{\sqrt{2}v}H^3 + \frac{\mu^2}{8v^2}H^4 \tag{1.47}$$

Therefore, by combining Equation 1.30, with the determinant of Equation 1.45 and $W^\pm = (W^1 \pm iW^2)/\sqrt{2}$ the Higgs Lagrangian turns out to be:

$$\mathcal{L}_{Higgs} = \frac{1}{2}(\partial_\mu H)(\partial^\mu H) + \frac{1}{4}g^2v^2W_\mu^+W^-\mu + \frac{1}{4}(g'^2+g^2)v^2Z_\mu Z^\mu - \lambda v^2H^2 \tag{1.48}$$

$$+cubic\ interaction\ terms\ +\ quatric\ interaction\ terms\ +\ const. \tag{1.49}$$

where the first term is the kinetic term of the H field. The neutral and two charged massless Goldstone degrees of freedom mix with the gauge fields that correspond to the broken generators of the $SU_L(2)\otimes U_Y(1)$ symmetry and transform into the longitudinal components of the $W^\pm$ and

Z bosons. This way, these bosons obtain mass, which is given by the second and third term of the $\mathcal{L}_{Higgs}$ [5, 6]:

$$m_W^2 = \frac{g^2 v^2}{4}, \quad m_Z^2 = \frac{(g'^2 + g^2)v^2}{4} = \frac{m_W^2}{\cos^2 \theta_W} \tag{1.50}$$

The fourth generator does not break, since it is related to the conserved $U(1)_{cm}$ symmetry and the corresponding gauge field, the photon, remains massless. In a similar manner, the 8-color gauge bosons, the gluons, that are intertwined with the conserved $SU(3)_C$ symmetry with 8 generators that do not break, also remain massless, but they are confined within hadrons and mesons, as a result of the QCD asymptotic behavior. Therefore, out of the four initial degrees of freedom of the Higgs field, two are absorbed by the $W^{\pm}$ gauge bosons, one by the Z boson and the remaining degree of freedom is taken up by the Higgs boson H, that was introduced by P. Higgs. The Higgs boson is neutral in electromagnetic interactions and transforms as a singlet in $SU(3)_C$; as a result, it does not couple in tree level with the massless photons and gluons [8]. The Higgs boson is a neutral particle with zero spin ($spin = 0$) and even CP; it can be considered as the mediator of the Higgs field, that prevails the universe. Every particle that interacts with it gains mass, with the heavier the particle, the stronger the interaction. Particles like the photon, do not interact with the Higgs boson and therefore, remain massless. The mass of the Higgs boson is calculated by the relation:

$$m_H = \sqrt{2\lambda} v \tag{1.51}$$

where λ is a self-coupling parameter in $V(\phi)$. The Higgs field expectation value v is calculated as:

$$v = (\sqrt{2} G_F)^{-1/2} \tag{1.52}$$

and depends on the Fermi coupling parameter G_F, that is determined by muon decay measurements. The term λ is a free parameter of the SM and thus, there can be no a priori prediction for the Higgs boson mass. Moreover, when it comes to the sign of the mass parameter: $\mu^2 = -\lambda v^2$, it has to be negative for electroweak symmetry breaking to occur. Experimentally, the Higgs boson was discovered on July 4, 2012 at the LHC by the ATLAS [9] and CMS [10] experiments at mass $m_H = 125.10 \pm 0.14$ GeV, from which the values $\lambda \approx 0.13$ and $|m| = 88.4$ GeV can be deducted.

It has become clear by now that physical mass is provided by spontaneous symmetry breaking and translate to Yukawa interactions with ϕ: $\mathcal{L}_{int} = -\alpha_f \bar{\psi}_f \psi_f \phi$, where the index f signifies the quarks or leptons. Then, the fermions of the SM gain mass through the renormalisable interactions with the Higgs field, the Yukawa interactions, that abide by the SM symmetries,

but produce fermion mass once the electroweak symmetry breaks, which is characterised by the following Lagrangian:

$$\mathcal{L}_{Yukawa} = -\hat{h}_{d_{ij}}\bar{q}_{L_i}\phi d_{R_j} - \hat{h}_{u_{ij}}\bar{q}_{L_i}\tilde{\phi}u_{R_j} - \hat{h}_{l_{ij}}\bar{l}_{L_i}\phi e_{R_j} + h.c. \tag{1.53}$$

In the Lagrangian above, $\tilde{\phi} = i\sigma_2\phi^*$ and $q_L(l_L)$ and $u_R, d_R(e_R)$ are the quarks (leptons) from the $SU(2)_L$ doublets and singlets, while each $\hat{h}_{X_{ij}}$ is parameterised by a 3×3 matrix. The mass terms that refer to the neutrinos are left out, but they can be included in a way similar to the up quarks, if right-handed neutrinos supplemented the SM. Neutrinos can also gain Majorana mass via non-renormalisable interactions with the Higgs field in 5 dimensions.
Once the Higgs field obtains a VEV and after a rotation of the fermion mass eigenstate, that also diagonalises the Higgs-fermions interactions, all fermions gain mass:

$$m_{f_i} = \frac{h_{f_i}v}{\sqrt{2}} \tag{1.54}$$

The indices $i, j = 1, 2, 3$ refer to the three generations of the up and down quarks and all charged leptons. Note that the mechanism of spontaneous symmetry breaking does not explain any underlying reasons why there is such a diversity in fermion masses, a phenomenon called flavor hierarchy [11].

1.4 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the widely accepted theory of the strong interactions that binds quarks into nucleons and in turn nucleons into nuclei. Very much like the Standard Model of electromagnetic and weak interactions, QCD is a quantum field theory (QFT) for massless vector particles whose dynamics is governed by a Lagrangian that is invariant under a non-abelian gauge symmetry [12]. The most fundamental tenet of QCD is that hadronic matter is made of quarks. The idea of quarks arose from the need to have a physical manifestation for the $SU(3)$ of flavour $SU(3)_f$ observed in the spectrum of the lowest-mass mesons and baryons. The observed baryons are interpreted as three-quark states. The quark constituents of the baryon necessarily have half-integral spin, so that the spin of the low-mass baryons is accounted for. Then, the quarks in the $spin = \frac{3}{2}$ states will be a symmetrical state of space, spin and $SU(3)_f$ degrees of freedom. However, the Fermi-Dirac statistics require that the wave function be totally antisymmetric. To resolve this dilemma the color degree of freedom was introduced: a color index i with three possible values (red, green, blue corresponding to the values $i = 1, 2, 3$) that each quark carries. The baryon wave functions are totally antisymmetric in this new index. Only color singlet states can exist in nature in order for the states to not be proliferated due to the introduction of the

color degree of freedom. [13] The only stable color singlets are quark–antiquark pairs, mesons, and three-quark states, baryons. In high energy reactions, like deep inelastic scattering, the quark and gluon constituents of hadrons act as quasi-free particles, partons [14].
Particles can be roughly classified on the basis of whether the strong force acts on them. Particles on which the strong force acts are called hadrons. Particles on which the strong force does not act, leaving the weak force and the electromagnetic force as the dominant forces, are called leptons. Protons (p), neutrons (n) and pions (π) are hadrons; electrons (e) and neutrinos (ν_e) are leptons. Further distinction can be made among the hadrons because some of them are bosons (mesons); the pion (π) is an example. The other hadrons are fermions (baryons); the proton (p) is an example [15].

Table 1.1: *Quantum numbers of quarks* [11]

	u	d	s	c	b	t
electric charge (Q)	+2/3	-1/3	-1/3	+2/3	-1/3	+2/3
isospin (I)	1/2	1/2	0	0	0	0
z-isospin (I_z)	-1/2	+1/2	0	0	0	0
starngeness (S)	0	0	-1	0	0	0
charm (C)	0	0	0	1	0	0
bottomness (B)	0	0	0	0	-1	0
topness (T)	0	0	0	0	0	1

Quarks make up nucleons; the baryons are bound states of three quarks whereas the mesons are composed of a quark and an antiquark bound together. For three flavors of quarks $q = u, d$ or s, there are 9 possible $q\bar{q}$ combinations of mesons. For baryons, the qqq combinations produce 27 outcomes, but baryons made of top quarks are not expected to exist because of the top quark's short lifetime. Baryons and mesons make up a wider category of particles, hadrons; all hadrons are postulated to be colorless. Mesons and baryons are characterised by the charge Q, the baryon number B ($+\frac{1}{3}$ for quarks and $-\frac{1}{3}$ for anti-quarks), the lepton number L, that are conserved additive quantum numbers, and also the total spin J that is calculated by the orbital angular momentum l and spin s as: $|l - s| \leq J \leq |l + s|$. So, the proton p, a uud bound state, identifies as a baryon, since by the additive quantum numbers of the two kinds of quarks, u and d, $B = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$ and not a lepton: $L = 0$ and has charge $Q = \frac{2}{3} + \frac{2}{3} - \frac{1}{3} = 1$. Similarly, π^+ is a $u\bar{d}$ state and thus a meson: $B = L = 0$ with charge $Q = \frac{2}{3} - \left(-\frac{1}{3}\right) = 1$. Parity P, the reflection of coordinates about the origin, is defined as $(-1)^{l+1}$ and charge conjugation C is given as $(-1)^{l+s}$, deduced by interchanging the positions and the spins of quarks. Strong and electromagnetic interactions are invariant under both P and C, whereas weak interactions do not respect these symmetries. However, to a good approximation, weak interactions are invariant under the product transformation CP [16].

All baryons have $B = 1$ baryon number (of course $B = -1$ for anti-baryons), but $B = 0$ for photons, leptons and mesons. Consequently, if baryon number is conserved in all physical processes, then the proton, being the lightest baryon, should not decay [17]. Note, however, that the mass of the proton, neutron or any other baryon is not the sum of the masses of the constituent quarks. For example, the total mass of the three quarks in a proton is only 20 MeV/c^2, woefully less than the proton's mass of 938.3 MeV/c^2. Nearly all of the protons mass is due to the internal energies of *(1)* the quark motion and *(2)* the fields that bind the quarks together [15]. Baryons are put into categories based on their isospin (I), and quark content. Therefore, there are six groups of baryons: Nucleons (N), Deltas (Δ), Lambdas (Λ), Sigmas (Σ), Xis (Ξ) and Omegas (Ω). The rules of their classification consider the up (u), down (d) and strange (s) quarks to be light and the charm (c), bottom (b) and top (t) quarks to be heavy [18].

- Baryons made up by any combination of three u and/or d quarks are Ns $I = 1/2$ or Δ ($I = 3/2$) baryons.
- Baryons containing two u and/or d quarks are Λ baryons ($I = 0$) or Σ baryons ($I = 1$). If the third quark is heavy, its identity is given by a subscript.
- Baryons containing one u or d quark are Ω baryons ($I = 0$), and subscripts indicate any heavy quark content.

For example, according to these rules, Λ_c^+ contains a c quark and some combination of two u and/or d quarks. The lightest baryons are the proton, composed by uud quarks (938 MeV/c^2), followed by the neutron made up by udd quarks(940 MeV/c^2) and they are both Nucleons.
Similarly, mesons are classified into groups according to their isospin (I), total angular momentum (J), parity (P), C-parity (C), whenever applied, and their quark content. Pseudoscalar and pseudovector mesons have 0^{-+} and 1^{++} respectively. This applies to a state with $l = 0$. For an excited state with $l = 1$ their J^{PC} quantum numbers can be 0^{++} and thus a scalar, 1^{++}, a vector, or 2^{++}, corresponding to a tensor. Flavorless mesons are those made of a pair of quark and anti-quark of the same flavor; consequently, all their quantum numbers are zero: $S = 0, C = 0, B = 0, T = 0$. The lightest known mesons are the pions (π); pions are pseudoscalars (0^{-+}) and they can be either neutral, generated by the superposition $\frac{1}{2}\left(d\bar{d} - u\bar{u}\right)$ (135 MeV/c^2) or charged π^+ (π^-) composed by a $u\bar{d}$ $\bar{u}d$ pair with mass 140 MeV/c^2 [11].

1.4.1 The QCD Lagrangian

Since QCD is a renormalisable theory based on the $SU(3)$ group with color triplet quark matter fields, the QCD Lagrangian density is defined as:

$$\mathcal{L}_{QCD} = \sum_{q=1}^{n_f} \bar{\psi}_i^q \left[i\gamma^\mu D_\mu - m_q \right]_{ij} \psi_j^q - \frac{1}{4} \sum_{A=1}^{8} F^{A\mu\nu} F_{\mu\nu}^A \tag{1.55}$$

where ψ_i^q denotes a quark field with (fundamental) color index i, $\psi^q = (\psi_{qR}, \psi_{qG}, \psi_{qB})^T$ and flavor f, γ^μ is the Dirac matrix that expresses the vectorial nature of the strong interactions, with μ being a Lorentz vector index, D_μ is the covariant derivative in QCD given as:

$$D_\mu = \partial_\mu - i g_s t^\alpha A_\mu^\alpha \tag{1.56}$$

with g_s the strong coupling, related to α_S by $g_s^2 = a\pi\alpha_S$, which determines the strength of the interaction between colored quanta, A_μ^α are the gluon fields with color index α and t^α are the $SU(3)$ group generators in the triplet representation of quarks, proportional to the hermitian and traceless Gell-Mann matrices of $SU(3)$ and thus, they are 3×3 matrices themselves. These generators are just the $SU(3)$ analogs of the Pauli matrices in $SU(2)$. By convention, the constant of proportionality is normally taken to be $t^\alpha = \frac{1}{2}\lambda^\alpha$, where λ^α are the Gell-Mann matrices [5,19]. The second part of the $\mathcal{L}_{QCD}$ (Equation 1.55) is purely gluonic, with the gluon field tensor $F_{\mu\nu}^A$ given by:

$$F_{\mu\nu}^A = \partial_\mu A_\nu^\alpha - \partial_\nu A_\mu^\alpha - g_s f_{ABC} A_\mu^B A_\nu^C, \tag{1.57}$$

where the f_{ABC} are the structure constants of $SU(3)$, defined through the commutators of the t^A matrices: $\left[t^A, t^B\right] = i f_{ABC} t^C$. The term $g_s f_{ABC} A_\mu^B A_\nu^C$ does not have a QED analogue and is responsible for the self-interactions of gluons through three- and four-point vertices [20].
Typically, color is not measured in the final state - instead all possible incoming colors are averaged and all possible outgoing ones are summed, whereby QCD scattering amplitudes (squared) in practice always contain sums over quark fields contracted with Gell-Mann matrices. These contractions in turn produce traces which yield the color factors that are associated to each QCD process, and which basically count the number of "paths through color space" that the process at hand can take [19]. By convention, the normalisation of the $SU(N_c)$ matrices is chosen to be:

$$\mathrm{Tr}\, t^A t^B = T_R \delta^{AB}, \qquad T_R = \frac{1}{2} \tag{1.58}$$

where T_R is associated with the color factor for a gluon split to a quark-antiquark pair. With this choice, the color matrices obey the following relations:

$$\sum_A t^A_{ab}\ t^A_{bc} = C_F \delta_{ac}, \qquad C_F = \frac{N_c^2 - 1}{2N_c} \tag{1.59}$$

$$\mathrm{Tr}\, T^C T^D = \sum_{A,B} f^{ABC} f^{ABD} = C_A \delta^{CD}, \quad C_A = N_c \tag{1.60}$$

The color factor C_F is related to gluon emission from a quark, while C_A is associated with gluon emission from a gluon. For the specific case of $SU(3)$ ($N_c = 3$), the values of these two color factors are [13]:

$$C_F = \frac{4}{3}, \qquad C_A = 3 \tag{1.61}$$

The dynamics of quarks and gluons as embedded on the QCD Lagrangian lead to a number of reverberations. Firstly, the addition of the color degree of freedom does not lead to any associations with new hadrons. The explanation lays in the dynamics mediated by the gluons. At very high energies (subnuclear dimensions) the strong force becomes small but on the contrary, it increases with distance; this property is called asymptotic freedom. Consequently, only color singlet entities with no net color charge can be observed as free particles. This property is in accordance with experiments where only colorless mesons and baryons that have integer multiples of the electric charge are detected. However, colored quarks with fractional electric charges and other colored objects like gluons are confined in subnuclear dimensions. This property of QCD is called confinement [21].

Perturbative QCD relies on the idea of an order-by-order expansion in a small coupling $\alpha_S = g^2/4\pi \ll 1$. In this frame, some given observable f can be given as:

$$f = f_1 \alpha_S + f_2 \alpha_S^2 + f_3 \alpha_S^3 + ... \tag{1.62}$$

where only the first one or two terms of this series can be calculated, with the understanding that remaining ones should be small. In principle, the coefficients f_i of the above series can be calculated through the use of Feynman diagramatic techniques [20].

On the other hand, confinement is a property dictating that no isolated colored charge can exist, but can only be manifested as color singlet particles. For example, the potential between a quark and an antiquark has been studied on the lattice. Lattice QCD is an approach that involves discretising space-time and considering the values of the quark and gluon fields in all the vertices/edges of the resulting 4-dimensional lattice (with imaginary time). Through suitable Monte Carlo sampling over all possible field configurations, one can essentially determine the

relative likelihood of different field configurations, and this provides a solution to QCD. The potential of the example at hand has a Coulomb part at short distances and a linearly rising term at long distances:

$$V_{q\bar{q}} \approx C_F \left[\frac{\alpha_S(r)}{r} + ... + \sigma r \right] \tag{1.63}$$

where in $SU(3)$ $C_F = 3/4$, r is the relative distance between the quark and anti-quark and σ is a parameter. The linearly rising term in the potential makes it energetically impossible to separate a $q\bar{q}$ pair. If the pair is created at one space-time point, for example in e^+e^- annihilation, and then the quark and the antiquark start moving away from each other in the center of mass frame, it soon becomes energetically favorable to create additional pairs, smoothly distributed in rapidity between the two leading charges, which neutralise color and allow the final state to be reorganised into two jets of colorless hadrons, that communicate in the central region by a number of hadrons with small energy [5, 20].

1.4.2 QCD Feynman diagrams

The Feynman rules are defined by the action operator:

$$S = i \int \mathcal{L} \; d^4x \tag{1.64}$$

that provides the phase of the transition amplitudes. The Lagrangian density in Equation 1.64 can be separated into a free piece $\mathcal{L}_0$, that usually contains all the bilinear terms of the fields, and an interaction piece $\mathcal{L}_I$, which contains all the rest terms:

$$S = S_0 + S_I, \quad S_0 = i \int d^4x \; \mathcal{L}_0(x), \quad S_I = i \int d^4x \; \mathcal{L}_I(x) \tag{1.65}$$

The S_0 term will produce the inverse propagators whereas the Feynman rules for the interacting part of the theory, that is treated as perturbation, derives from S_I. However, perturbation theory cannot be applied unless a gauge fixing term is introduced. However, making a choice of gauge is necessary, when defining the propagator for the gluon field:

$$\mathcal{L}_{gauge-fixing} = -\frac{1}{2\lambda}(\partial^\mu A^\alpha_\mu)^2 \tag{1.66}$$

This choice of gauge will fix the class of covariant gauges with gauge parameter λ. In a non-Abelian theory, such as QCD, a ghost Lagrangian $\mathcal{L}_{ghost}$ must supplement this covariant fixing term, due to non-physical degrees of freedom that are being introduced and propagate freely. These are cancelled off by including into the theory a non-physical set of fields, the

ghosts, which are scalars but obey the Fermi statistics. Feynman rules exist for ghosts too and in every diagram with a closed loop of internal gluons containing only triple-gluon vertices, a diagram with the gluons in the loop replaced by ghosts must be added. The ghost Lagrangian is given as:

$$\mathcal{L}_{ghost} = \partial_\mu \eta^{\alpha\dagger}(D^\mu_{\alpha\beta}\eta^\beta) \tag{1.67}$$

Here, η^α is a complex scalar field that complies with the Fermi statistics. As mentioned, the ghost fields cancel non-physical degrees of freedom which would otherwise propagate in covariant gauges. Therefore, the QCD Lagrangian becomes:

$$\mathcal{L}_{QCD} = \sum_{q=1}^{n_f} \bar{\psi}_i^q \left[i\gamma^\mu D_\mu - m_q\right]_{ij} \psi_j^q - \frac{1}{4}\sum_{A=1}^{8} F^{A\mu\nu}F^A_{\mu\nu} - \frac{1}{2\lambda}(\partial^\mu A^\alpha_\mu)^2 + \partial_\mu \eta^{\alpha\dagger}(D^\mu_{\alpha\beta}\eta^\beta) \tag{1.68}$$

Just as in QED, the Feynman rules can be read off from the action. Since only the gluon propagator is gauge dependent, one gets [1]:

1. The quark propagator:

$$\delta^{ab}\frac{i}{\not{p} - m_q} = \delta^{ab} i\frac{\not{p} + m_q}{p^2 - m_q} \tag{1.69}$$

   where $\not{p} = \gamma^\mu p_\mu$, p is the momentum of the quark, m_q its mass and δ^{ab} is the Kronecker delta function. The respective ghost propagator is $\frac{i}{k^2}\delta^{ab}$, where k stands for the momentum of the ghost particle.

2. The gluon propagator:

$$\delta^{ab} i\frac{-g_{\mu\nu}}{p^2} \tag{1.70}$$

   where $g_{\mu\nu}$ is the metric tensor with diagonal elements $(+1, -1, -1, -1)$ and the rest are zero.

3. The quark-gluon vertex:

$$-ig_s t^a_{bc}\gamma_\mu \tag{1.71}$$

   which depends on the strong coupling g_s and also contains the t^A color factor of the $SU(3)$ group. These propagators are diagonal in color and the vertex for a gluon of color A to

[1] The diagrams were made by [22]

scatter a quark of color b to a quark of colour c contains contains $(t^A)_{bc}$. The quark–gluon vertex is flavor-independent. The factor of a ghost quark-gluon vertex is $g_s(p+k)^\mu f^{abc}$

4. The three-gluon vertices:

$$-g_s f^{abc}\left[(p-q)^\lambda g^{\mu\nu} + (q-r)^\mu g^{\nu\lambda + (r-p)^\nu g^{\lambda\mu}}\right] \tag{1.72}$$

where f^{abc} is the $SU(3)$ group structure constant, p, q, r are the gluon momenta, μ, ν, λ are Lorentz indices and a, b, c are color indices. The momenta are all in-going: $p+q+r=0$.

5. The four-gluon vertices:

$$-ig_s^2\left[f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\sigma} - g^{\mu\sigma}g^{\nu\rho}) + f^{ace}f^{bde}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\sigma}g^{\nu\rho}) + f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma})\right] \tag{1.73}$$

where g_s, f^{abc} and $g^{\mu\nu}$ are the same as before [13, 14, 23, 24].

Feynman diagrams are used to make very precise calculations of the probability of any given process. Since a given process can be represented by many possible Feynman diagrams, the contributions of every possible diagram must be entered into the calculation of the total probability that a particular process will occur. Using the Feynman diagrams one can compute the matrix element $\mathcal{M}$ that is analogous to the amplitude f of producing a final state, which is immediately related to the cross section (σ), and also decay rates (Γ), quantities that can be measured.

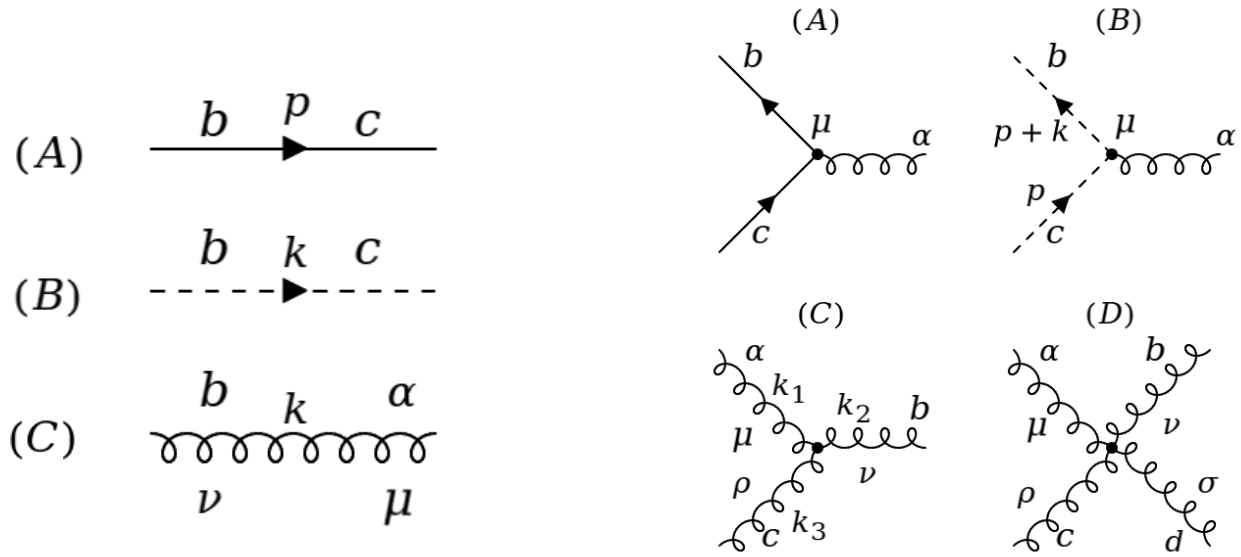


Figure 1.3: *Schematic representation of (A) quark, (B) ghost and (C) gluon propagators (left) and Feynman diagrams of (A) quark-gluon, (B) ghost-gluon, (C) three-gluon and (D) four-gluon vertices (right).*

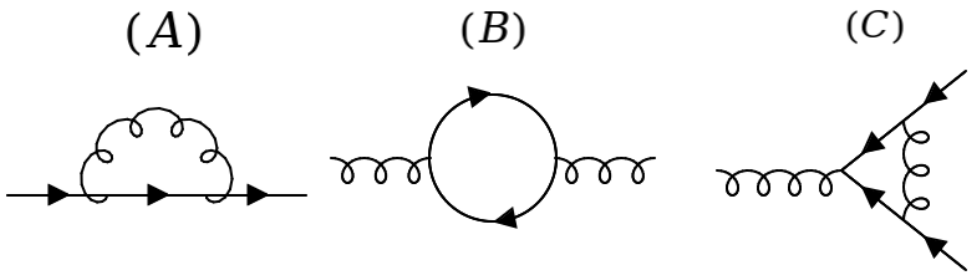


Figure 1.4: *Diagrams of one loop contributions to (A) the quark self-energy, (B) the gluon self-energy and (C) the quark-gluon vertex.*

1.4.3 Renormalisation and strong coupling α_S

Calculations in QCD are extremely demanding, with the size of the coupling constant, although perturbative, even at large values of the exchanged momentum Q^2 being such that the perturbative expansion will converge rather slowly. In order to obtain satisfactory accuracy, several orders of Perturbation Theory are required and the calculations become more and more complex with the order of the approximation. Additionally, evaluating large class of higher-order diagrams lead to results that are by default ill-defined, such as infinities. Moreover, the powers of coupling constants, that is, the order in the perturbative expansion, increase with the number of loops. Examples of diagrams accounting for contributions to the quark self-energy, the gluon self-energy and the quark-gluon vertex at one loop are shown in Figure 1.4.

The momentum integrals of loop corrections to propagators diverge for large loop momenta. For the theory to be meaningful, these divergences need to be regularised. One way to do this is by applying the concept of $D = 4 - 2\epsilon$ dimensions where the loop integrals converge. Here, a regulator or cut-off of the divergences Λ is introduced. In other words, the considered Lagrangian will describe the interactions of fields in D-dimensions. Quantities that appear in the Lagrangian, such as mass and coupling constants, will absorb the would-be divergences and all theory parameters consistent with its symmetries will adjust as functions of the cut-off. After renormalisation, the regulator is removed again ($\Lambda \to \infty,\ \epsilon \to 0$). In physical terms, through renormalisation, the effects of quantum fluctuations with momenta larger than the scale of a given problem can be absorbed into the parameters of the theory [21, 23, 25].

A renormalisation scheme has to be chosen, which will prescribe which part of the loop corrections the renormalised quantities absorb. A scheme that is frequently used is the Minimal Subtraction $\bar{MS}$ scheme that simplifies higher-order calculations. In this scheme, all quantities and masses depend on the renormalisation scale μ. When renormnalising, due to the change in dimensions, an arbitrary renormalisation scale μ needs to be introduced, in order to keep consistent dimensions and units to all quantities. Thus, the strong coupling $\alpha_S = g_s^2/4\pi$ will now depend on the scale μ at which it is commonly evaluated, through g_s, that will remain dimensionless. This dependence

can be expressed in terms of renormalisation group equations (RGE):

$$\mu^2 \frac{\partial \alpha_S(\mu^2)}{\partial \mu^2} = \beta(\alpha_S(\mu^2)) \ , \quad \beta(\alpha_S) = -\alpha_S^2(b_0 + b_1\alpha_S + b_2\alpha_S^2 + ...) \tag{1.74}$$

where the β function is negative, at least when α_S is small, a fact that constitutes the origin of asymptotic freedom, stating that the coupling becomes weaker at high momentum scales (short distances), i.e., the theory almost becomes a free theory, in which quarks and gluons don't interact. Conversely, at low momentum scales the coupling grows strong, causing quarks and gluons to be tightly bound into hadrons (infrared slavery). The β function depends on the b_i coefficients:

$$b_0 = \frac{33 - 2n_f}{12\pi}, \quad b_1 = \frac{153 - 19n_f}{24\pi^2}, \quad b_2 = \frac{77139 - 15099n_f + 325n_f^2}{3456\pi^3} \tag{1.75}$$

In the above equation, n_f is the the number of quark flavors whose mass is lower than μ. When neglecting higher orders, except for b_0, and also ignoring the fact that the number of "light" flavors n_f depends on μ, the solution of the renormalisation group equation for α_S at some scale Q as a function of its value at the renormalisation scale μ takes the form:

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu^2)}{1 + b_0 \ln(Q^2/\mu^2)\alpha_S(\mu^2)} = \frac{1}{b_0 \ln Q^2/\Lambda^2} \tag{1.76}$$

As dictated by Equation 1.76, the strong coupling can be expressed in terms of the non-perturbative constant Λ (or Λ_{QCD}), a scale at which the coupling becomes infinite at leading order. Perturbation theory is valid only when $\alpha_S(\mu^2) \ll 1$, which corresponds to scales $\mu \gg \Lambda$. Since Λ is a non-perturbative quantity, it cannot be theoretically defined; its value will depend on which b_i coefficients were used for a given value of $\alpha_S(Q^2)$, the renormalisation scheme, the number of flavors assumed, etc. Measurements, however, have determined its order of magnitude, $\Lambda \approx 200MeV$, a value closely related to the hadron masses.
The strength of a QCD interaction for a process involving momentum transfer Q is given by $\alpha_S(\mu)$ with $\mu \sim Q$. One can measure the strength of an interaction at various scales; one of these reference scales is the Z-boson mass ($Q^2 = M_Z^2$) [14, 20], with the standard mean value in 2020 (PDG) being $\alpha_S(M_Z^2) = 0.1770 \pm 0.0001$ [11].

1.4.4 Cross section predictions for *pp* collisions

Physics at hadron colliders cover a wide energy range and phenomena, including regions where quantitative theoretical calculations are weary. Thus, a deeper understanding of the dynamics between the proton constituents and in turn, their interactions with other composite objects via

the strong interaction is needed. For this purpose, the Parton Model was proposed by Bjorken and Feynman, in order to describe deep inelastic processes. It is based on the conception of asymptotic freedom which assumes that quarks or gluons (partons) inside a hadron do not interact with other partons, they are therefore "free" particles.

According to the above, form factors $f(x, Q^2)$ for each parton are being thoroughly used to describe the inner structure of protons; these factors are called Parton Distributions Functions (PDF), where x denotes the fraction of the proton momentum a parton carries and Q^2 stands for the momentum transfer of the interaction. In other words, these PDFs describe the distribution of the proton energy among the different quarks and gluons. All interactions for which perturbation theory cannot be applied due to the nature of processes that take place in colored composite objects are summed in PDFs. A hadron comprises of three main quarks, the valence quarks, that determine its charge and flavor. The inner structure of hadrons, however, is much more complicated, since they are governed by a strong gluon field due to the QCD interactions among quarks, which creates a "sea" of quark and anti-quark pairs. This results in all hadrons to be made of quarks, anti-quarks and also gluons, with any of these partons being potentially able to interact with an incoming particle. Ultimately, the computation of the cross section for a hadron-hadron process can be separated into two parts: a component that corresponds to the hard process and can be calculated on perturbative level in orders of the strong coupling α_S and the non-perturbative regime, described by the parton model and and is encapsulated in the PDFs. These two schemes are distinguished by introducing a non-physical parameter, the factorisation scale μ_F [27].

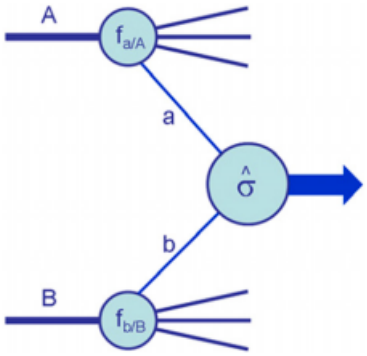


Figure 1.5: *Sketch of hadron-hadron scattering* [26].

By applying the factorisation theorem, the cross section of producing an arbitrary $n-$parton final state from the interaction of two hadrons h_1 and h_2 in processes of the form: $h_1 h_2 \to n + X$ can be written as:

$$\sigma_{2\to n} = \sum_{\alpha,b} \int_0^1 dx_\alpha dx_b \int f_{\alpha/h_1}(x_\alpha,\ \mu_F) f_{b/h_2}(x_b,\ \mu_F) d\hat{\sigma}_{\alpha b\to n}(\mu_f,\ \mu_r) \tag{1.77}$$

$$= \sigma_{\alpha,b} \int_0^1 dx_\alpha dx_b \int f_{\alpha/h_1}(x_\alpha,\ \mu_F) f_{b/h_2}(x_b,\ \mu_F) \frac{1}{2\hat{s}} \int d\Phi |\mathcal{M}_{ab\to n}|^2(\Phi_n; \mu_f,\ \mu_r) \tag{1.78}$$

where h_1 and h_2 are the incoming hadrons, $f_{\alpha/h_1}, f_{b/h_2}$ are their respective PDFs and $\hat{\sigma}_{\alpha b\to n}$ is the parton level cross section distribution, that is given by the corresponding matrix element squared $|\mathcal{M}_{ab\to n}|^2$, averaged over initial-state spin and color degrees of freedom and integrated over the

available n-parton final-state phase space Φ_n. The parton flux $\hat{s}$ is : $1/(2\hat{s}) = 1/(2x_\alpha x_b s)$, where s is the hadronic center-of-mass energy. The first integral is calculated over all possible momentum fractions $x(\alpha), x(b)$ of each of the partons inside the two hadrons. As a consequence of the factorisation of possible parton emissions into a soft and collinear part, below μ_F , and a hard part, above μ_F , the latter are explicitly described by the matrix element $\mathcal{M}_{\alpha b\to n}$. At higher μ_F, more of the momentum of the hadrons is given to gluons and sea quarks with relatively low values of the momentum fraction x. Conversely, the population of partons that contain a large fraction of the parent hadron's momentum will decrease.
The matrix element is calculated by a summation over the Feynman diagrams, denoted as:

$$\mathcal{M}_{ab\to n} = \sum_i \mathcal{F}^{(i)}_{ab\to n} \tag{1.79}$$

and the phase-space element is given by [28]:

$$d\Phi_n = \prod_{i=1}^{n} \frac{d^3p_i}{(2\pi^3)} \frac{1}{2E_i} \cdot (2\pi)^4 \delta^{(4)} \left(p_\alpha + p_b - \sum_{i=1}^{n} p_i \right) \tag{1.80}$$

The PDFs are treated as being universal, i.e., they can be determined by one set of processes and used to predict the cross section for any process using the factorised form of Equation 1.77. However, PDFs cannot be obtained perturbatively, so they are fit from deep inelastic scattering measurements. Their evolution can be obtained as a function of Q^2 from the DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) equations. The PDFs inside a proton at next-to-next-to-leading order in α_S (NNLO) in the MSHT analysis for Q=100 GeV are shown in Figure 1.6.

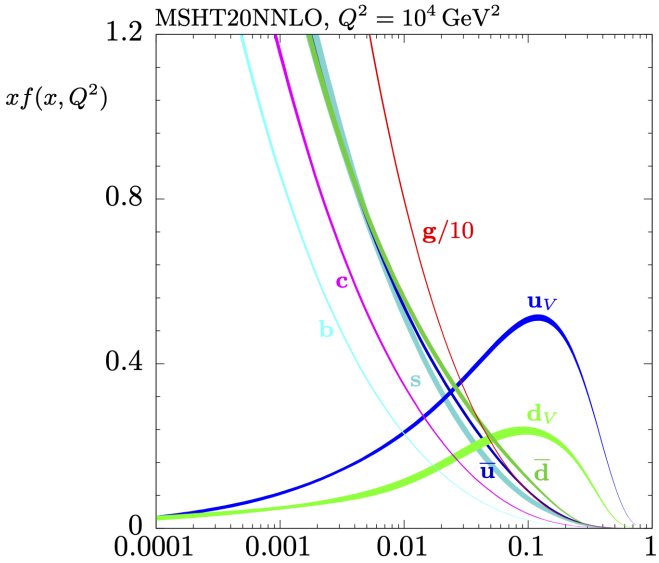


Figure 1.6: *The MSHT Parton Distribution Function for $Q = 100$ GeV* [29].

When a final parton is produced, a bunch of collimated partons and hadrons collinear to the initial parton is formed. This fragmentation and collinear showering arises from the iterative gluon radiation and quark-antiquark production since both are colorful particles and the strong coupling α_S increases while it moves away from the proton local color field. The collection of the collimated final state hadrons is called a "jet" [27].

Bibliography

[1] K. Vagionakis. *Particle Plysics: An Introduction to the Basic Structure of Matter*. NTUA Press, 2013.

[2] David Griffiths. *Introduction to Elementary Particles 2nd Edition*. John Wiley & Sons, 2008.

[3] Standard Model of particle physics. https://en.wikipedia.org/wiki/Standard_Model.

[4] M. Robinson. *Symmetry and the Standard Model*. Springer-Verlag New York, 211.

[5] Stephen Myers and Herwig Schopper. Particle Physics Reference Library, 2019. https://doi.org/10.1007/978-3-030-38207-0.

[6] G. Altarelli. Collider Physics within the Standard Model: a Primer, 2013. https://arxiv.org/abs/1303.2842.

[7] John Ellis. Higgs Physics, 2013. https://cds.cern.ch/record/1638469.

[8] Yosef Nir. Higgs and Electroweak: One Higgs, three discoveries, 2020. https://cerncourier.com/a/one-higgs-three-discoveries/.

[9] G. Aad et al [ATLAS Collaboration]. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, Vol. 716, 2012. https://doi.org/10.1016/j.physletb.2012.08.020.

[10] S. Chatrchyan et al [CMS Collaboration]. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, Vol. 716, 2012. https://doi.org/10.1016/j.physletb.2012.08.021.

[11] P. A. Zyla et al [Particle Data Group]. Review of Particle Physics. Progress of Theoretical and Experimental Physics, 2020. https://doi.org/10.1093/ptep/ptaa104.

[12] Benjamin Grinstein. Quantum Chromodynamics: High Energy Experiments and Theory. *Physics Today*, 57, 2004. https://doi.org/10.1063/1.1801872.

[13] R.K. Ellis, W.J. Stirling, and B.R. Webber. *QCD and Collider Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 1996.

[14] Michael H. Seymour. Quantum Chromodynamics. 2009. https://doi.org/10.48550/arXiv.hep-ph/0505192.

[15] David Halliday, Robert Resnick, and Jearl Walker. *Fundamentals of Physics, 12th Edition*. John Wiley & Sons, 2021.

[16] Francis Halzen and Alan D. Martin. *Quarks and Leptons: An Introductary Course in Modern Particle Physics*. John Wiley & Sons, 1984.

[17] A. Das and T. Ferbel. *Introduction to Nuclear and Particle Physics, Second Edition*. World Scientific Publishing, 2005.

[18] S.S.M. Wong. *Introductory Nuclear Physics, 2nd Edition*. Wiley & Sons, 1998.

[19] Peter Skands. Introduction to QCD. 2017. https://doi.org/10.1142/9789814525220_0008.

[20] G.P. Salam. Elements of QCD for Hadron Colliders. 2020. https://doi.org/10.23730/CYRSP-2020-005.1.

[21] Ian Brock and Thomas Schorner-Sadenius. *Physics at the Terrascale*. John Wiley & Sons, 2011.

[22] Online Feynman diagram maker. https://www.aidansean.com/feynman/.

[23] Michelangelo L. Mangano. Introduction to QCD. In *1998 European School of High-Energy Physics*, 1998.

[24] QCD Feynman Rules. https://web2.ph.utexas.edu/~vadim/Classes/2019s-qft/qcdfr.pdf.

[25] John Collins. *Foundations of Perturbative QCD, Vol. 32*. Cambridge University Press, 2011.

[26] J.M. Campbell, J.W. Huston, and W.J. Stirling. Hard interactions of quarks and gluons: a primer for lhc physics. *Rep. Prog. Phys. 70*, 2007. https://doi.org/10.1088/0034-4885/70/1/R02.

[27] Luis Pascual Domínguez. *Studies of photon isolation in a large-pileup environment and search for axion-like particles decaying into two photons with the ATLAS detector at the LHC*. PhD Thesis, 2021.

[28] John Campbell, Joey Huston, and Frank Krauss. *The Black Book of Quantum Chromodynamics: A Primer for the LHC Era*. Oxford University Press, 2018.

[29] UCL-London's Global University MSHT Parton Distribution Functions, 2023. https://www.hep.ucl.ac.uk/msht/.

Chapter 2

CERN and LHC

2.1 European Laboratory for Particle Physics

The resolution concerning the establishment of CERN, an acronym of Conseil Européen pour la Recherche Nucléaire, was discussed in December 1951 during an intergovernmental meeting of UNESCO in Paris. In June 1953, the final draft of the CERN Convention was signed by 12 European countries, including Greece, while its facilities, which are located in the French-Swiss border, started operating in 1954. Since then, its main research focuses, not so much on the study of atomic nuclei, as on Particle Physics, which studies the nature and interactions of elementary particles, since nowadays, knowledge around matter extends way beyond the nucleus. For this reason, the laboratory operating under the auspices of CERN is often referred to as the European Laboratory for Particle Physics [1].

Figure 2.1: *CERN's convention* [1].

Today, CERN constitutes one of the largest centers of nuclear research. It consists of 23 member states, almost 2500 staff, who participate in the design, construction and operation of the facilities, the preparation and the conduction of the experiments, as well as the analysis of the collected data, a process in which 12200 scientists of 110 different nationalities, working in 70 institutions from around the world, take part [2].

The experiments conducted inside the CERN facilities focus on the study of the fundamental components of matter, the elementary particles, in order to understand the mechanisms that rule

their interactions, but also the fundamental laws of nature. This can be accomplished through collisions of the subatomic particles that are accelerated by colliders until their velocity reaches almost the speed of light. In order for this to happen, particle accelerators and detectors are used. The accelerators impel beams of particles to higher energies and force them to collide with each other or on stable targets, whereas the detectors note and record the results of these collisions [3].

The accelerator's complex of CERN consists of consecutive accelerators, with each having higher energy than the previous one and accelerates every particle beam to a certain energy, before the beams are lead to the next accelerator [4]. The first system is LINAC1, it produces the 160 MeV protons, that are injected Proton Synchrotron Booster (PSB). Then, these protons are accelerated to 2 GeV and are transferred to Proton Synchrotron (PS), where they remain until their enegry reaches 26 GeV. Super Proton Synchrotron is used in order to further increase their energy to 450 GeV, before they are moved, with a few minute gap, to the main ring of LHC (Large Hadron Collider), where their energy becomes 6.5 TeV, providing 13 TeV as total energy of collision. The protons move in opposite direction inside the accelerator, until their speed becomes 0.999999990 times the speed of light in vacuum ($c = 3 \times 10^8 m/s$) until they finally collide with each other. There are four areas inside the LHC where the main detectors (CMS, ATLAS, ALICE and LHCb) are located. Later in this thesis, an overview of the CMS experiment will be presented, on which the current analysis is based [5].

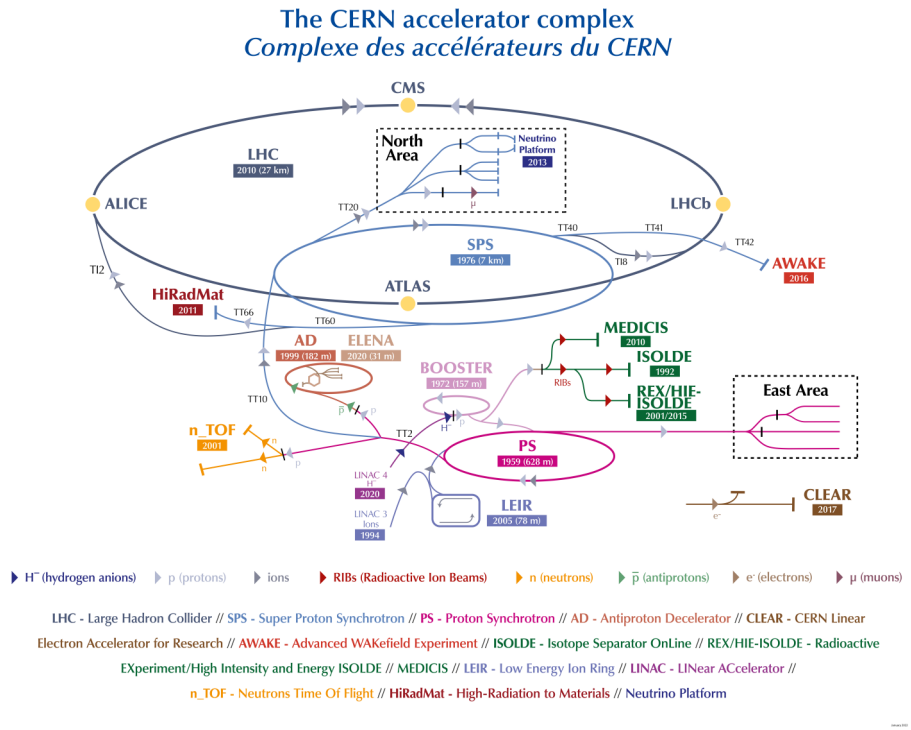


Figure 2.2: *The CERN accelerator complex* [5].

2.2 Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) the greatest and most powerful accelerator, was set for the first time in operation in 10 September 2008 and still remains the most recent addition to the accelerators complex of CERN. The LHC is located 175 m underground and consists of a ring of superconducting magnets, 27 km in length, as well as numerous accelerating structures, in order to increase the energy of the particles that move across the accelerator.
The main purpose of the function of LHC is to provide answers to questions regarding the Standard Model, that still remain unanswered. The Standard Model – a theory that was developed in the early '70s and describes the elementary particles and their interactions – has predicted a wide range of physical phenomena. However, the Standard Model is incomplete. The experiments conducted inside the LHC aim to answer inquiries concerning the origin of mass: the SM fails to explain it and the reason why some particles are massive, whereas other are massless. Theoretical physicists Robert Brout, Françoils Englert and Peter Higgs suggested the Brout-Englert-Higgs Mechanism, that provides mass to particles, in cases when they interact with the Higgs Potential that prevails the universe. Particles that strongly interact with the Higgs potential are heavier than those who interact lightly. The Higgs Boson, carrier of the Higgs potential, was discovered at CERN in 2012, thus confirming the Higgs Mechanism. However, the study of all its properties and decay modes is still pending. Another question that needs answer concerns the existence of supersymmetry, since the Standard Model is unable to describe a theory that unifies all the fundamental forces, and it lays on the fact that manufacturing a theory for gravity, similar to the one that describes the other forces, is a particularly difficult task. Furthermore, the experiments that are executed in the LHC aspire to answer questions that have to do with the particles and the phenomena responsible for both dark matter and energy, the reason why the amount of matter in the universe exceeds that of anti-matter, even though they seem to have been produced in equal parts during the Big Bang, and lastly, the production of the particles that constitute matter by quark-gluon plasma [5].

2.2.1 The course of LHC

As it was mentioned before, LHC is the most powerful particle accelerator that was constructed by CERN at the same tunnel, 27 km in perimeter and 175 m deep, that housed the Large Electron-Positron Collider (LEP) [6]. The proposal and approval of the CMS and ATLAS experiments lasted from 1993 until 1997 and a few years later the ALICE and LHCb experiments were approved [7].
The first trial run at the LHC was carried out on 10 September 2008. On 18 September of the same year, however, a malfunction of the cryogenic system resulted in the temperature of

the magnets to increase above the established temperature of operation, which is near absolute zero. The accelerator started to operate again on 20 November 2009, when the first successful collision of proton beams was accomplished, at 900 GeV center of mass energy. A few days later, on 30 November 2009, LHC surpassed the Tevatron accelerator at Fermi National Accelerator Laboratory (Fermilab), as the most powerful particle accelerator, by accelerating particles at 1.18 TeV. On 19 and 30 March 2010, successful collisions of proton beams were carried out at 3.5 and 7 TeV center of mass energy, respectively.

On 5 April 2012, the first proton collision was conducted at center of mass energy 8 TeV, while on 4 July 2012 it was announced that a new particle had been discovered by the CMS and ATLAS experiments, that had properties similar to those of the Higgs boson, at a mass range of 125 GeV. The Higgs boson, as it is described by the Standard Model, constitutes the manifestation of the Brout-Englert-Higgs Mechanism that provides mass to the $W^{\pm}$ and Z bosons, as well as the quarks and leptons. Of course, other types of Higgs bosons have been predicted by theories that extend beyond the Standard Model. On 8 October 2013, François Englert and Peter Higgs were awarded the Nobel Prize «for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider».

On 14 February 2013, LHC was deactivated in order to be upgraded and to function again at its full energy, 13 TeV (at center of mass) and at increased luminosity. The accelerator was set in operation again on April 2015 [9].

Another important accomplishment was the first observation of a very rare decay of the particle B_s^0 into to muons: $B_s^0 \to \mu^+\mu^-$. Since then, these types of decay has been examined thoroughly, because they can make way to theories beyond the Standard Model, such as Supersymmetry.

On April 2017, the LHC began accelerating particles again, after a 17-week break. The main goal was to reach integrated luminosity of 45 fm^{-1} by the end of 2017. On July 2017, at the EPS Conference on High Energy Physics in Venice, the observation of Ξ_{cc}^{++} by LHCb was announced, a particle that consists of two charm and one up quark, of mass 3621 MeV. The fact that its mass was almost four times greater than that of

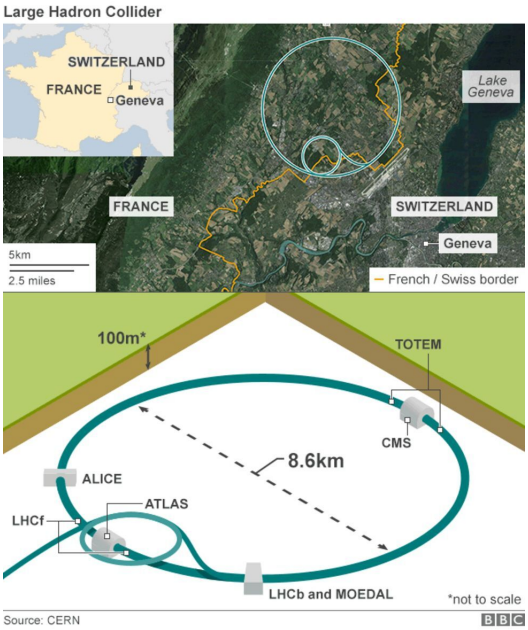


Figure 2.3: *Aerial view and graphic illustration of the LHC* [8].

the heaviest known baryon, the proton, is due to its two charm quarks. On 28 October 2018, the last proton collision Run 2 was completed, until the beginning of Run 3, after a long period of upgrades [10].
On December 2018, the second period of upgrades and maintenance started, in order to reach the maximum possible center of mass energy of 14 TeV, which was scheduled to finish by the end of 2021 or the beginning of 2022. The main purpose of this upgrade was the installation of the High Luminosity Large Hadron Collider (HL-LHC), that will have greater luminosity and is scheduled to start operating on 2027. The HL-LHC will give scientists the opportunity to further study the already known mechanisms, such as the Higgs boson, but also to observe new rare phenomena. For instance, at the HL-LHC at least 15 million Higgs bosons will be produced, compared to 2017, when its production was calculated to 3 million occurrences of Higgs boson [11].
In 2022 the LHC Run3 started, after a vast programme of works completed during the Long Shutdown 2 (LS2). Protons now collide at higher energies (13.6 TeV compared to 13 TeV) and with higher luminosities (containing up to 1.8×10^{11} protons per bunch compared to $1.3 - 1.4 \times 10^{11}$) than Run2. This third experimental phase runs until the end of 2025. A few weeks after the start of Run3, several records were already reached. Some of them regard a peak luminosity at 2.5×10^{34}, pileup (almost simultaneous collision points) > 100 and stored energy per beam almost 400 MJ. The current schedule foresees the Long Shutdown 3 (LS3) to start in 2026 and to last for three years [10].

2.2.2 LHC Characteristics

LHC is made up of a 27 km ring that includes numerous superconducting magnets in order to increase the energy of the inserted protons. The maximum value of the collision energy that the protons can acquire is 14 TeV at the center of mass. Until now, LHC produced proton collisions at energy 13 TeV (6.5 TeV per beam). During Run 3, that began on July 2022, after the upgrade that started in December 2019, collisions of the maximum energy take place, at 13.6 TeV.
The number of events N_{event} that is produced by the collisions carried out inside the LHC is given by:

$$N_{event} = \mathcal{L}\sigma_{event} \tag{2.1}$$

where σ_{event}: is the cross section of the event that is under examination and $\mathcal{L}$: is the luminosity. The luminosity of LHC depends only on parameters of the beam and therefore can be written as a Gaussian distribution:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi\epsilon_n \beta^*} F \tag{2.2}$$

where N_b: is the number of particles per bunch, n_b: the number of bunches per beam, f_{rev}: the rotation frequency, γ_r: a relativistic factor, ϵ_n: the emissions across the beam, β^*: the β function at the collision point and lastly, F: represents a geometric factor that describes the reduction of luminosity caused by the angle between the scattering beams at the interaction point [12].
Luminosity is one of the most important parameters of an accelerator. It corresponds to the number of collisions that can be produced per surface units and per time units:

$$\mathcal{L} = \frac{1}{\sigma}\frac{dN}{dt} \tag{2.3}$$

The maximum value of luminosity that can be achieved at LHC is 10^{34} cm^{-2}s^{-1} at 14 TeV, a value that corresponds to 10^{34} collisions per second. During Run2, though, the peak luminosity reached 2×10^{34} cm^{-2}s^{-1}. Every now and again, LHC is being upgraded in order to increase, not only its energy, but also its luminosity. A direct way to accomplish that is to focus the beam exactly at the scattering point by reducing the parameter β^*, which demands redesigning the optics at the Interaction Regions (IR) and replacing the quadrapolic magnets, that are responsible for the final focus of the beam. The stage at which these magnets need to be replaced partially depends on their time of life and their resilience to high doses of radiation. Another way of raising the luminosity of LHC includes the increase of the number of bunches or the number of protons per bunch. However, there is a limit to the extent that these parameters can be altered, such as the distance between the beams and their long range interactions among them, the electron clouds, the repercussions of focusing the beams, pile up of events etc.
Given the fact that during the experiment the luminosity does not remain constant, since the total number of protons keeps reducing, the integrated luminosity is of great importance and it's given as the integral of the given luminosity in time:

$$\mathcal{L} = \int \mathcal{L}\, dx \tag{2.4}$$

The integrated luminosity is measured as the inverse cross sections, which is b^{-1}. In the chart seen below (Figure 2.4) the integrated luminosity at the CMS and ATLAS experiments is represented for the various stages of the operation of LHC (LHC Runs) [13]. In this case, the integrated luminosity $\mathcal{L}_{int}$ is given by the following formula:

$$\mathcal{L}_{int} = \mathcal{L}_0 \tau_{\mathcal{L}} \left[1 - e^{\frac{-T_{run}}{\tau_{\mathcal{L}}}}\right] \tag{2.5}$$

where $\mathcal{L}_0$: is the initial luminosity, $\tau_{\mathcal{L}}$: the life time of luminosity and T_{run}: the duration of the experiment. Supposing that the experiment is being conducted for 200 days a year and the life time of luminosity is 15 hours, the best duration of each round is 12 or 5.5 hours, for 7 and 15 hours mean time of completion respectively, the maximum total integrated luminosity per year

ranges between 80 fb^{-1} and 120 fb^{-1}, depending on the time of completion.

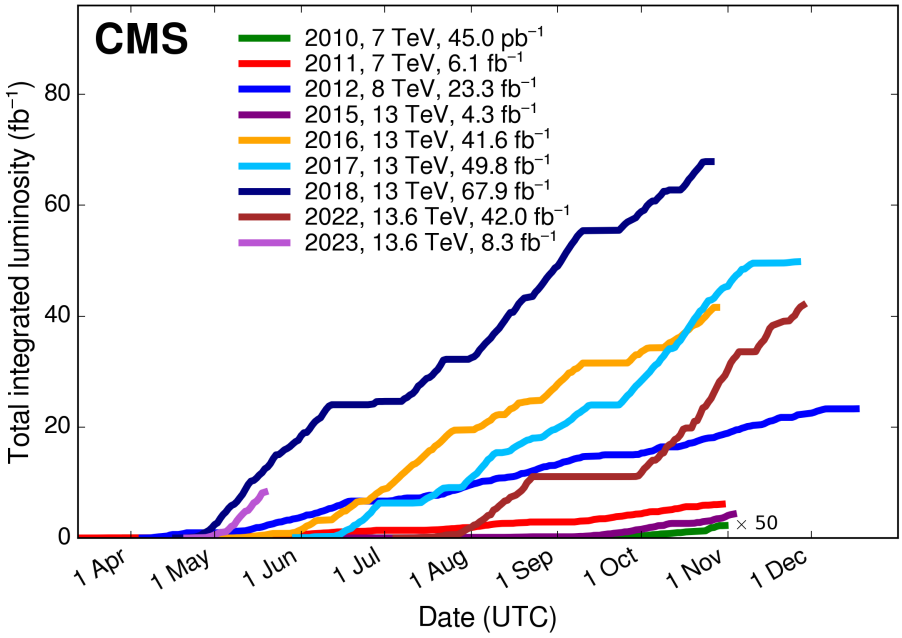


Figure 2.4: *The total integrated luminosity delivered by the LHC and recorded from the CMS experiment during Run 1, Run 2 and begining of Run 3* [13].

In the interior of the accelerator, the beams of protons accelerate to high energies and their speed reaches almost the speed of light, until they collide with each other. The beams move in opposite directions inside different tubes in complete vacuum. Their acceleration is done by superconducting magnets using coils made of $NbTi$. The magnet system of the LHC is cooled down by super-fluid He, that keeps the temperature at 1.9 K, and is inside a magnetic field of 8.33 T [12]. Thousands of magnets of different shapes and sizes are being used in order to guide the beams inside the accelerator. These include 1232 dipoles, 15 m in length, that turn the beams of charged particles, and 392 quadrapole magnets, $5-7$ m each, which use lays entirely in focusing the beams. Another type of magnets that is being utilised, forces the protons to come extremely close to each other, right before a collision occurs, in order to increase the probability of scattering. At the LHC hexapoles and octapoles are also being used [14].

In order for the beams to remain in the interior of the accelerator, to accelerate and to maintain their maximum energy, RF cavities are being used, with oscillation frequency at 400.8 MHz. The RF cavities impose a total of 16 MeV potential, 2 MeV each. During acceleration, the energy provided to the beams is relatively small, 275 kW per beam, while the energy that is required for the beams to remain in orbit is much greater [12]. The main purpose of the RF cavities is to increase the initial energy of 450 GeV of the protons, to 6.5 TeV, almost 14 times

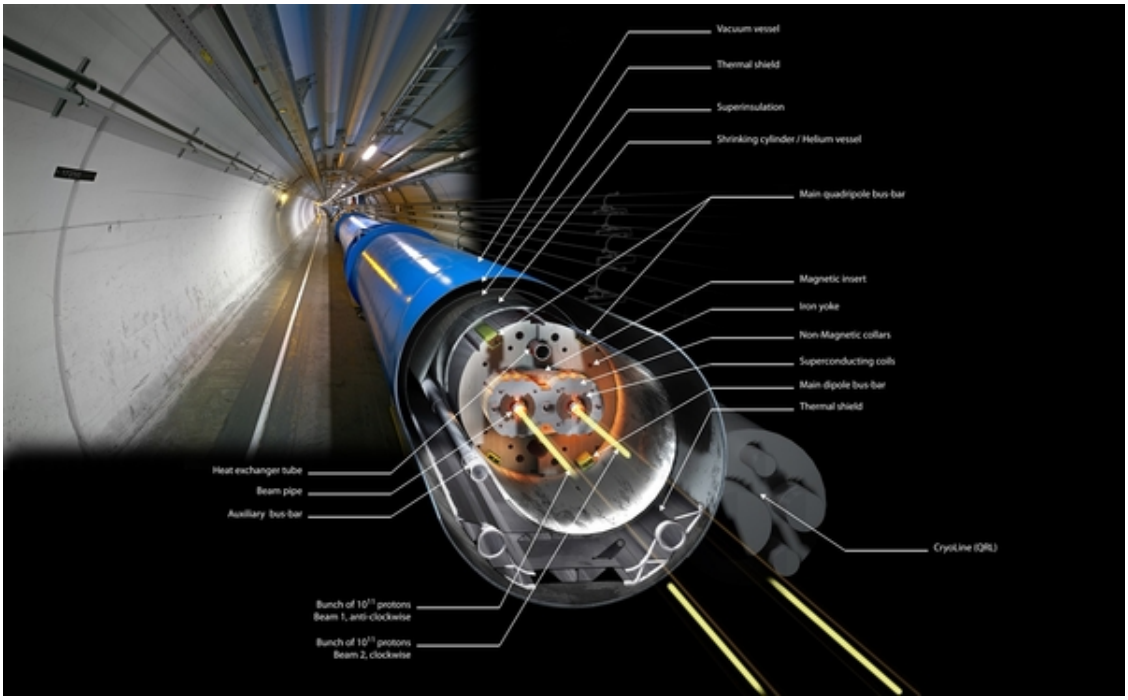


Figure 2.5: *Bisection of LHC* [14].

the energy they have when inserted to the accelerator tube. The maximum energy is reached in approximately 20 minutes, with the bunches to have passed by the cavities more than 10 million times. Since the field of the RF cavities oscillates, i.e. constantly changes direction, at given speed, the moment when the particles arrive at the cavities is important. When the beam reaches the required energy, the protons with the correct energy that arrive at the appropriate moment will accelerate no more. On the contrary, protons with slightly different energy that don't arrive at the ideal moment, will accelerate or decelerate in order to achieve the energy required. This way, the protons don't move inconsistently, but in accordance with each other, and remain in bunches. At the LHC there are 2808 bunches, whereas in the beginning of the experiment, the number of protons per bunch is 1.2×10^{11}.

Apart from the RF cavities, CERN manufactures the so-called crab cavities, that are intended for the successor of LHC, the High Luminosity LHC. These cavities will provide transverse momentum to the particles, in order to lead them as they approach the interaction point [15].

LHC has three different vacuum systems: an insulation vacuum for the magnets, an insulation vacuum that eases the distribution of the Helium, in order to thermally isolate the electromagnets, that are kept at 1.9 K, and when there are no major leaks, has stable pressure at 10^{-6} mbar, 10^{-6} Torr or 10^{-9} atm. The third vacuum system is inside the tube, along with the beam, to prevent any collisions of the protons with the air molecules. The terms of the last vacuum system concern the life time of the beam and the background of the experiment. The terms of the cryogenic temperatures are expressed as gas densities, as a function of the hydrogen density, taking into account the ionisaion cross section of any kind of gas. The equivalent hydrogen gas

densities must remain at low values, less than 10^{15} H_2 m^3, so that the required life time of the beam is ensured to be 100 hours. At the interaction regions, the densities are less than 10^{13} H_2 m^3, so that the background is minimised. At the regions of the vacuum system that are at room temperature, pressure must remain at 10^{-10} mbar, 10^{-10} Torr or 10^{-13} atm [12].

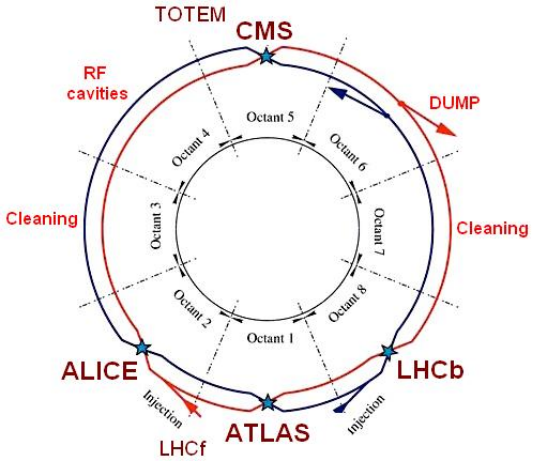


Figure 2.6: *LHC is not a perfect circle, but it made of eight arcs and eight 'insertions'. The magnets that turn the beams are placed in the arcs, 154 in each. In the diagram the injection points, the RF cavities, the dump and cleaning points are depicted* [16].

Bibliography

[1] CERN official website – Our history. https://home.cern/about/who-we-are/our-history.

[2] CERN official website – Our people. https://home.cern/about/who-we-are/our-people.

[3] CERN official website – About. https://home.cern/about.

[4] CERN official website – The Large Hadron Collider. https://home.cern/science/accelerators/large-hadron-collider.

[5] CERN official website – Facts and figures about the LHC. https://home.cern/resources/faqs/facts-and-figures-about-lhc.

[6] Brittanica – Large Hadron Collider. https://www.britannica.com/technology/Large-Hadron-Collider.

[7] Chris Llewellyn Smith. Genesis of the Large Hadron Collider. *Phil. Trans. R. Soc. A.*, 373, 2015. http://doi.org/10.1098/rsta.2014.0037.

[8] Jonathan Webb. Large Hadron Collider restarts after two-year rebuild. *BBC News*, 2015. https://www.bbc.com/news/science-environment-32160755.

[9] CERN oficcial website – The Higgs Boson. https://home.cern/science/physics/higgs-boson.

[10] Taking a closer look at LHC. https://lhc-closer.es/taking_a_closer_look_at_lhc/1.home/taking_a_closer_look_at_lhc/1.home.

[11] CERN official website – High-Luminosity LHC. https://home.cern/science/accelerators/high-luminosity-lhc.

[12] Lyndon Evans and Philip Bryant. The CERN Large Hadron Collider: Accelerator and Experiments. *New J. Phys*, 9, 2007. https://dx.doi.org/10.1088/1367-2630/9/9/335.

[13] CMS Luminosity – Public Results. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[14] UK Research and Innovation – What happens at the LHC. https://aka.stfc.ukri.org/research/particle-physics-and-particle-astrophysics/large-hadron-collider/what-happens-at-the-lhc/.

[15] CERN official website – Accelerating: Radiofrequency cavities. https://home.cern/science/engineering/accelerating-radiofrequency-cavities.

[16] Taking a closer look at the LHC – LHC layout. https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_layout.

Chapter 3

The CMS experiment

The Compact Muon Solenoid (CMS) is a general purpose detector, constructed for studying particles and phenomena that derive from high energy proton collisions that take place inside the LHC. It has a lengthy experimental program, that extends from the study of the Standard Model, including the Higgs boson, to the search of extra dimensions and particles that could constitute dark matter.

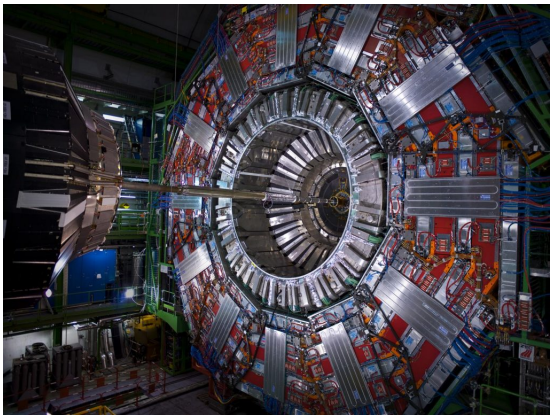


Figure 3.1: *The CMS detector* [1].

The CMS detector was built around a huge solenoid magnet of cylindrical spiral shape made of superconducting coils, that produces a magnetic field of 4 T, almost 100000 times that of the Earth. The magnetic field is being confined by a steel yoke, that forms the main part of the detector.

In contrast to other detectors of the LHC experiments, it is built in 15 sections on ground level, before it is lead underground, in a cave near Cessy, France, where it is reassembled. The entire detector is 21 m long, 15 m wide and its mass comes up to almost 21.5 t.

The CMS experiment is one of the biggest scientific collaborations, employing 5000 physicists, engineers, technical personnel, students and staff from 200 institutes in 50 countries, as a 2019 research indicates [2].

At the interaction point, the proton beams collide with a rate of 10 billion collisions a second (10^9 collisions/sec), but due to the trigger system, an event selection is being carried out and as a result the total number of events available for analysis is reduced to 1000 events/sec. The products of the collisions are recorded by the various levels of the CMS detectors.

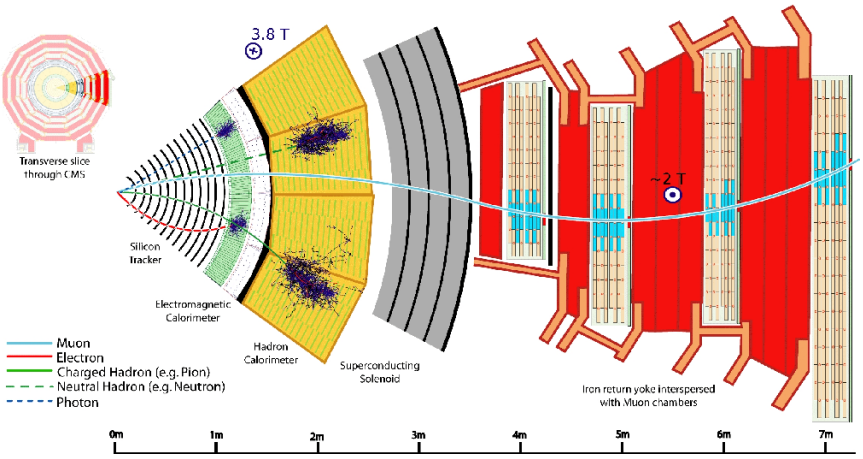
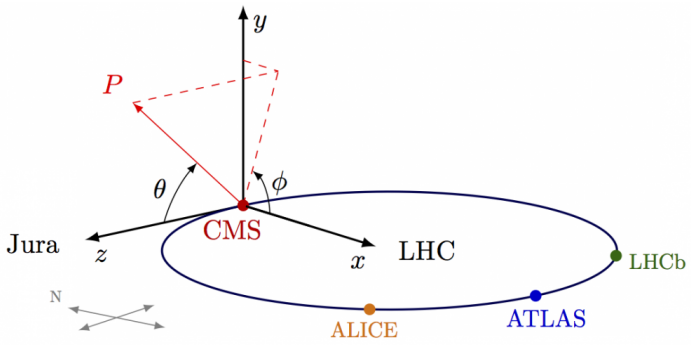
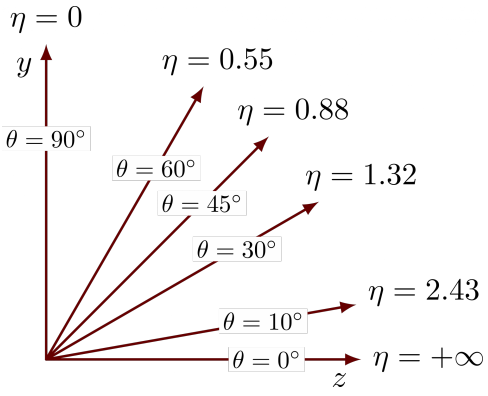


Figure 3.2: *Graphic illustration of the CMS detecting systems* [3].

The CMS experiment uses a right-handed coordinate system, the beginning of which lies at the collision point. The direction of the x axis is radial towards the center of the LHC ring, the y axis is upward and vertical to the LHC plane, whereas the z axis and the azimuthial angle (ϕ) are being measured from the positive side of the x axis on the $x-y$ plane. The radius r indicates the distance from the z axis, while the pseudo-rapidity is defined as [4]:

$$\eta = -\ln\left[\frac{tan\theta}{2}\right] \tag{3.1}$$





(b) Pseudorapidity vs angle.

Figure 3.3: *The CMS coordinate system (left) and The correspondence between pseudo-rapidity (η) and the polar angle (θ)* [5].

The detecting system of CMS, which function lays on the measurement of the energy and momentum of photons, electrons and muons, consists of the Tracking System (Pixel Detector and Silicon Strip Tracker), the Electromagnetic Calorimeter, the Hadron Calorimeter, the Supercon-

ducting Solenoid Magnet and the Muon Chambers, in order of increasing radius. Their set-up is shown on the Figure 3.4, while their function is described in the following units.

3.1 Tracking System

The Tracking System of CMS has been designed in order to detect the various signatures of the particles that are produced by the protons' scattering inside the LHC. This is achieved by identifying and measuring muons, electrons, photons and jets in a wide energy range.
The momentum and track resolution is performed by the 4 T magnetic field, provided by the solenoid magnet. The Tracking System expands 115 cm in radius and is 270 cm long towards each side of the interaction point. The corresponding time and resolution of the electrical set-up is relatively small and this is the reason why this type of detecting system is ideal for the experiment, that has a large particle flow. The two Tracking Systems of CMS are the Pixel Detector and the Silicon Strip Tracker [6].

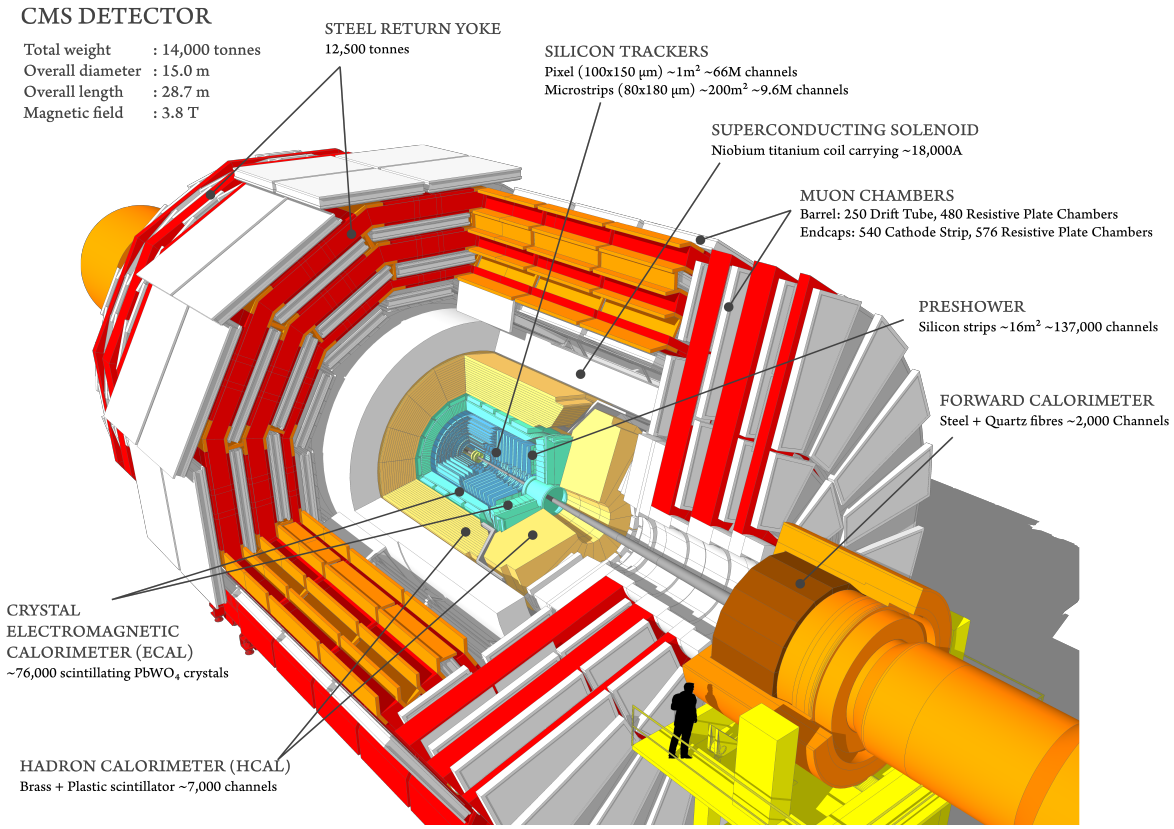


Figure 3.4: *Graphic illustration of the CMS detector* [7].

Pixel Detector

The Pixel Detector is made up of approximately 66 million pixels, that allow it to detect the trajectory of the particles that emerge from the proton collisions, especially those that have small time of life [8]. It is the detector closest to the beam pipe and it consists of three barrels (BPix) and two endcap disks (FPix). The barrels are 53 cm long, they are positioned at radius 4.4, 7.7 and 10.2 cm each and they include 48 million pixels in total, covering an area of 0.78 m^2. Likewise, the endcaps extend to $6 - 15$ cm in radius and are placed at distance $z = \pm 34.5$ cm and $z = \pm 46.5$ cm on each side. They consist of 18 million pixels and occupy a surface of 0.28 m^2. The size of each pixel is $100 - 150$ μm The arrangement of the barrels combined with the endcaps provide 3 points of the particles' track with $15 - 20$ μm deviation and cover almost the entire pseudo-rapidity range ($-2.5 \leq \eta \leq 2.5$).
The sensors of the pixels adopt the n-on-n technology according to which, the pixels consist of high dose n-implants that are place on top of a high resistance n-sublayer. The pn conjunction is positioned on the back side of the sensor, that is surrounded by a protective ring. The p-stop technique is being used for the endcaps and the p-spray for the barrels, in order to provide insulation between the pixels. Both kinds of sensors show high resilience to radiation [9].

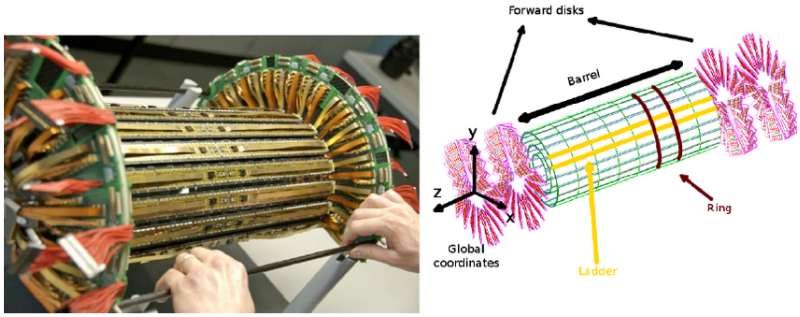


Figure 3.5: *Actual illustration and drawing of the pixel barrel and disks* [10]

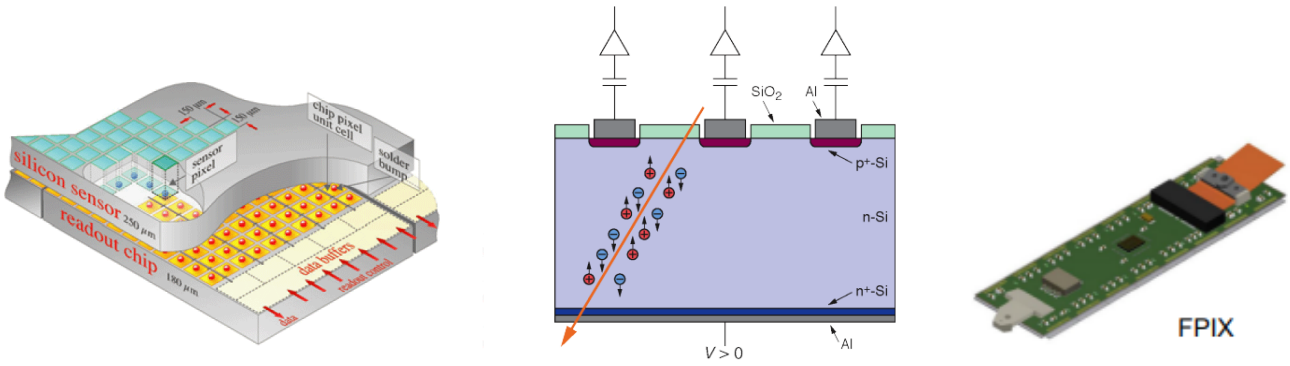


Figure 3.6: *A CMS silicon pixel detector (left)* [11], *Graphic illustration of the pixel function (center)* [12] *and Drawing of a pixel detector module used in FPIX (right) [13].*

Being so close to the interaction point, the number of particles that go through the pixel detector is huge; the rate of the particles that are being detected at 8 cm distance from the beam is estimated to be 10 million particles per cm^2 per second. The detector is able to decompose and reconstruct all the traces the particles leave behind. When a charged particle passes through the detector, it provides enough energy for the electron to blast off the silicon atoms, and thus creating electron-hole pairs. Each pixel makes use of electric current to collect these charges as electric signal at their surface. A piece of silicon is attached to each pixel tile (read-out chips or ROCs) that reads the signal and uses a tiny welding point, the so-called bump bonding technique, to amplify the signal. By knowing which pixels have received a particle, the extraction of its trajectory is possible. The detector contains two dimensional pixels and is made up by multiple layers and therefore, the result is the construction of a three dimensional image.
Given the large number of pixels, their power needs to be relatively small. Each pixel has 50 μW in power, with their total summing up to 3.6 kW. In order for the detector not to overheat, the pixels are positioned on top of pipes, with cooling liquid C_6F_{14} running through them [14].

Silicon Strip Tracker

The Pixel Detector is followed by the Silicon Strip Tracker that reaches a radius up to 130 cm. The detectors are made up of 4 different inner barrels (TIB), arranged in shells, and 2 inner endcaps (TID), each containing 3 small disks. The outer barrel (TOB) consists of 6 concentric layers. Lastly, the detector is enclosed by 2 endcaps (TEC). Each endcap contains silicon modules, that are designed specifically to be placed inside the detector [8].
The inner barrels (TIB), cover $z = 70$ cm in space, they are placed at distance 25.5 cm, 33.9 cm, 41.9 cm and 49.8 cm, and deploy silicon sensors, 320 μm wide, and an elevated strip which length ranges between 80 and 120 μm. The first two layers make measurements of $r - \phi$ and $r - z$ coordinates. The solid angle is chosen at 100 mrad and leads to $23 - 34$ μm resolution in the $r - \phi$ direction and 23 μm in the direction of z. The outer barrel (TOB) extends between $|z| = 1.24$ m and $|z| = 2.8$ m from the interaction point, while their distance from the beam axis is 60.8 cm, 69.2 cm, 78.0 cm, 86.8 cm, 96.5 cm and 108 cm. Their resolution is $35 - 52$ μm in the $r - \phi$ direction and 52 μm in the z.
The endcaps (TEC) include 9 disks that lay in the region 120 cm$< |z| <$ 280 cm with 25.5 cm$< r <$ 113.5 cm radius, whereas each inner endcap (TID) contains 3 small disks that fill the gaps between the inner barrels and the endcaps. They are both arranged in rings with the beam axis going through their center and have strips in the same direction as the beam. Some rings are single-sided, while others are double-sided. The width of the TIDs and the three inner TEcs is 320 μm, whereas the rest of the TECs are 500 μm wide.

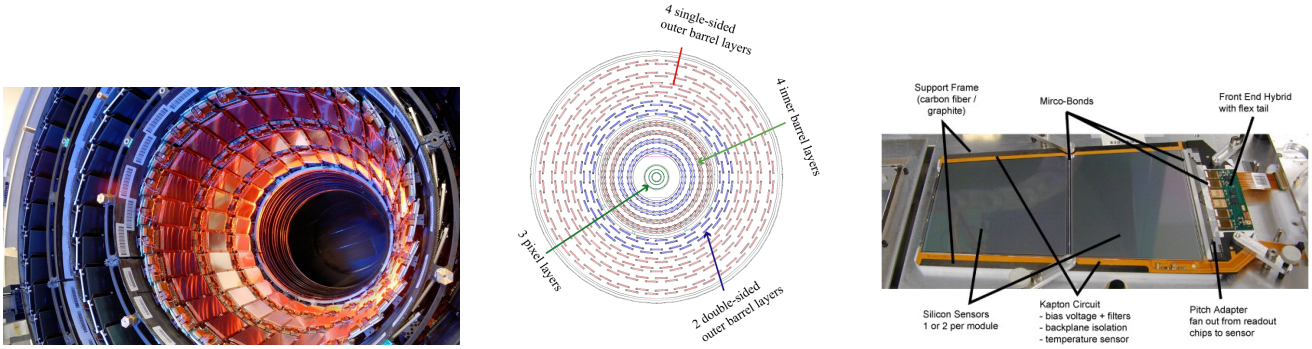


Figure 3.7: *Photo of the CMS silicon strip detectors in the barrel region (TIB) (left)* [15], *disection of the silicon strip barrel (center)* [16] *and mosule (right)* [17].

The elements of this type of detector are single-sided silicon sensors, strip shaped and p-on-n type. The entire detector is made up by 15400 modules arranged in structures made of carbon fibers and houses a temperature controlled tube. The number of the detector strips is 15 million, they cover an area of 198 m^2 and their signal is read by 80000 microchips [8, 9, 18, 19].
Each module consists of three main parts: a number of sensors, the mechanical support structure and the readout electronics. The sensors, which are made of silicon, receive a large number of particles at a relatively small area, due to their nice response and spatial resolution. Their function is similar to that of the Pixel Detector; when a charged particle reaches the material, it detaches an electron from the atoms, that move around inside the electric field that is imposed, giving them a small current pulse that last for a couple ns. The charge in each microstrip is read out and amplified by an Analogue Pipeline Voltage chip (APV25). The APV25 chips co-exist with the electronics that record quantities, such as the temperature, and provide information to match the signal with the collision to which it corresponds. Furthermore, the signals are saved in their memory for a few μs and later they are processed, before they are sent to a laser that turns them into infrared pulses. These pulses are then transmitted via a 100 m optic fiber cord, for their resolution is a radiation-free environment. The tracker uses 4000 such optic fiber parts, thus providing a lightweight and low-power way of transmitting the signals [14].

3.2 Electromagnetic Calorimeter

The Electromagnetic Calorimeter of CMS, or ECAL, is a homogeneous calorimeter made of $PbWO_4$ crystals. These crystals are arranged in a central barrel (EB), covering a pseudo-rapidity range up to $|\eta| \leq 1.48$, whereas the endcaps that shut the barrel (EE), cover a range up to $|\eta| = 3.0$. The crystals are placed 3° from the interaction point.
The cylindrical barrel (EB) is located in 1.29 m radius and consists of 61200 crystals $2.2 \times 2.2 \times 23$

cm^3 divided in 36 supermodules, each weighing 3 t and including 1700 crystals 20° towards the ϕ direction. Each supermodule comprises of 4 modules of which the first contains 500 crystals and the last 3 contain 400 crystals. For simplicity in the construction and assembly, the crystals are grouped in 2×5 arrays that are inside a thin wall (200 μm) in a hive-like structure and compose a submodule. The light that is being produced from the scintilation of the crystals inside the barrels is collected by avalanche photodiodes (APD), two for each crystal, each covering a 5×5 m^2 area.

In a similar manner, the endcaps (EE) are divided into two dees, they cover the barrel in both sides and consist of almost $150002.68 \times 2.68 \times 22$ cm^3 crystals in 36 groups called supercrystals. 268 identical supercrystals in total cover each endcap, with 64 more being used to cover the inside and outside perimeter. Each endcap includes 10746 crystals, corresponding to a total mass of 12.6 t, and kept at temperature -5°C. In the area around the endcaps, the light is detected by vacuum photo-triodes (VTP) that cover a total area of 280 mm^2 [14, 20–22].

Figure 3.8: *Geometric view of one quarter of the ECAL* [20], *crystals of half the Endcap of ECAL (center)* [23] *and lead tungsten crystals (right)* [24].

The typical energy resolution for electrons between 20 and 250 GeV in the CMS ECAL is parameterised as [9]:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{0.12}{E} \oplus 0.30\% \tag{3.2}$$

The preshower detector is positioned in front of the crystals and is made of a lead layer, on top of which silicon strips are placed, less than 2 mm wide. Its endcap part covers a pseudo-rapidity region $1.65 \leq \eta \leq 2.61$ and its main function is to distinguish the neutral pions (π^0) from the photons. The neutral pions are produced from collisions and decay to two low-energy photons in small angle, that can easily be confused with high energy photons produced by decays of the Higgs boson. So, the preshower detector is able to distinguish these two kinds of photons, this way preventing false signals. The preshower of the barrels covers a pseudo-rapidity region up to $|\eta| = 0.9$ and allows the measurement of the angle between the photons with 45 mrad/$\sqrt{E}$ precision in the direction of η. When a photon passes through the lead layer, it causes an electromagnetic shower, that includes electron-positron pairs, detected by the silicon sensors.

This way, the measurement of the energy of the photons becomes possible, whereas the existence of two layers of detectors provides two measurements, enabling the detection of the particle's position.
When high-energy photons are measured by the ECAL, they can be traced back to the collision point. Then, by observing the hits at the preshower and adding the energy that is laid there to the total energy measured by the ECAL, conclusions can be drawn about whether the photon was single or comes from a pair [22].

3.3 Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) measures the energy of hadrons, i.e. the particles that consist of quarks and gluons, such as protons, neutrons and pions. Furthermore, it provides indirect measurements concerning the existance of non-interacting uncharged particles, such as neutrinos. The detection of these particles is important since they indicate the composition of new or supersymmetric particles [25].
The HCAL is in the 1.77 m $< r <$ 2.95 m region, i.e. outside the ECAL and spreads until the inside of the solenoid magnet. It comprises of four different parts: the barrel (HB), the endcap (HE), the forward (HF) and the outer region (HO).
The HB covers pseudo-rapidity region $|\eta| < 1.39$ and is separated into two semi-barrels that consist of 36 wedges in total. Each wedge is deflected by 20° towards the ϕ direction, it is $40-75$ mm wide and is covered with absorbent plates made of brass (70% Cu, 30% Zn). The inner and outer part of the plates is made of steel. There are 17 recesses per wedge, that house 70000 plastic scintillation plates. The plates with the same angular deflection and depth are arranged in megatiles. The width of the first and last layer is 9 mm and 3.7 mm for the rest.
The HE is also made of absorbent brass plates, 79 mm wide, covered in plastic scintillation plates. The inner surface of HE is 4 m away from the interaction point and covers a pseudo-rapidity range $1.30 < |\eta < 3.0$. The hadrons that hit the brass plates of the HB and HE calorimeters cause a hadron shower. Then, they hit the scintillator, resulting in the emission of photons with wave length that corresponds to blue-ultraviolet. The light is transmitted by optic fibers to hybrid photo-diodes (HPDs), where it is being collected. The HPDs are photon detectors especially made for CMS that are able to operate in strong magnetic fields and provide an amplified signal that covers a wide range of energy. The HPD signal, given off in 25 ns intervals, is digitalised by charge integrators and decoders (QIE). The HO calorimeter comprises of one or two layers of scintillators and is placed outside the magnet. It is made up of 5 rings, each having 12 sectors; 6 scintillation disks are arranged in a hive-like structure on top of each sector. The main ring (Ring 0) has two layers of scintillators, 10 mm wide, on every side of an orthogonal segment and

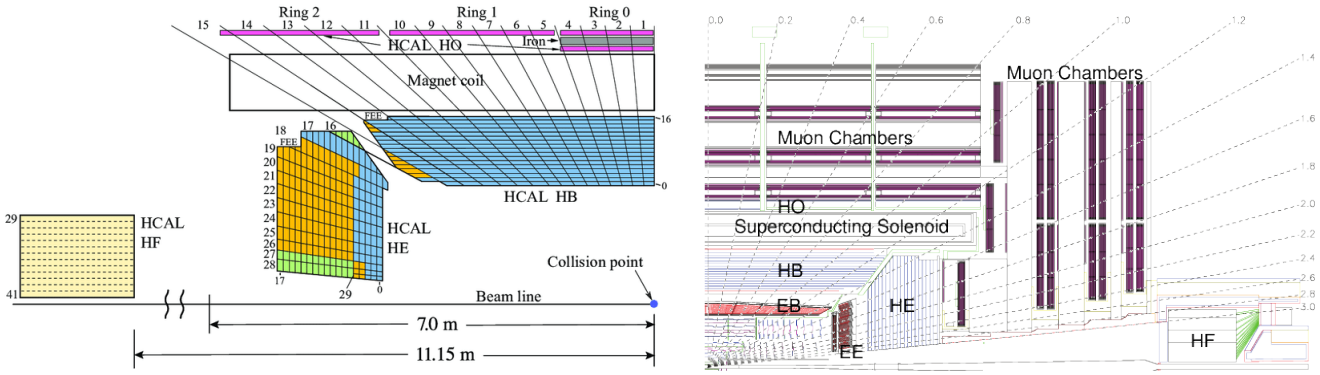


Figure 3.9: *Longitudinal views (r-z) plane of a quarter of the CMS detector showing: the HCAL tower segmentation for the HB, HE and HO detectors (left)* [26] *and he HCAL component (HB, HE, HO, HF) locations, the ECAL (EB, EE) and the muon systems, where the dashed lines represent fixed η values (right)* [27]

are in 3.9 m and 4.1 m radial distance. The rest of the rings consist only of one layer in a radius of 4.1 m. The HO covers a pseudo-rapidity region $|\eta| < 1.26$. The light that comes from the scintillators is transmitted via silicon photo-multipliers (SiPMs), whose signal is then digitalised using QIEs. The use of HO lies in ensuring that there are no energy leaks from the barrel. In the longitudinal plane $(r - z)$ the HB, HE and HO are segmented into η sectors called towers, illustrated in Figure 3.9. In practice, the measured energy deposition of a particle is the sum of the successive layers of tiles or equivalently over the tower. The number of scintillator layers depends on the tower and segment position [9].

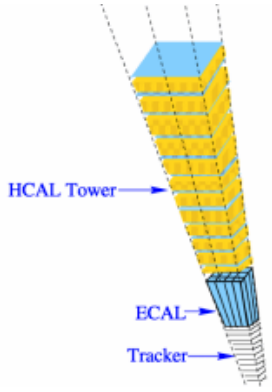


Figure 3.10: *Schematic view of HCAL tower* [9].

Lastly, the two forward calorimeters are placed at each edge of CMS; their purpose is to detect particles that come out of the collision point in small angles as to the direction of the beam. The HFs receive the biggest energy load from the particles that participate in the scattering and thus, they need to be resilient to high levels of radiation. Its measurements play an important role in studying the production of the top quark, as well as the Higgs boson. It is positioned 11.2 m away from the collision point and is made of steel with quartz fiber implants. The shorter fibers (16 cm) measure the total signal that derives from the wave length of the material, whereas the longer fibers (22 cm) measure the deposited energy. The detecting technique that is being enlisted involves the emission of Cherenkov radiation that is produced by secondary particles when they come across the quartz fibers. The light that is being collected by the fibers of the HF is converted to charge by a photo-multiplier tube (PMT) and then the charge is digitalised by the QIE. The pseudo-rapidity measurements lay in the $2.85 < |\eta| < 5.19$ region [25, 28–30].

3.4 Superconducting Solenoid Magnet

The magnet of CMS is the central device around which the experiment is developed, with a magnetic field 4 T, 100000 times more powerful than that of the Earth. Its main activity is to steer the particles that emerge from the high-energy collisions that take place inside the LHC. The greater the momentum of the particle, the harder it is for the magnetic field to bend their trajectory and so, by tracing the latter, a measurement of the momentum is provided. The power of the magnet lays on the fact that a larger field of force can further divert the particles and, in combination with the precise position measurements provided by the tracker and the muon chambers, the momentum measurement even of the most high-energetic particles is enabled.
The magnet of CMS is a solenoid, made of wires, that produce a homogeneous magnetic field when a current goes through them. It is also superconducting, allowing the electric current to flow without resistance and thus, creating a powerful magnetic field. Inside the solenoid the ECAL and HCAL are placed, while the Muon Chambers are interpolated by a 12-sided iron yoke that surrounds the magnet, includes and directs the field. The return yoke is made up of three layers and reach 14 m in diameter. It operates as a filter, allowing only the muons and weakly interacting particles, such as neutrons, to pass through it. This huge magnet takes up most of the material structure of the experiment and needs to be resilient enough, in order to withstand the forces imposed by the magnetic field.

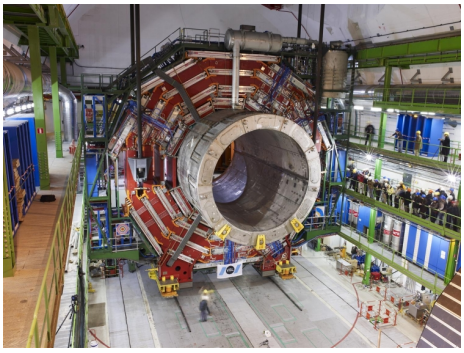


Figure 3.11: *The CMS solenoid magnet* [31].

The magnet is 6.3 m in diameter, 12.5 m long and the energy it encloses is 2.6 GJ at its center. As it was already mentioned before, the flux of the current returns through a return yoke weighing 10000 t, that includes 5 wheels and 2 endcaps, each consisting of 3 disks. Between the three disks, 4 spiral layers are placed, made of enforced superconducting $NbTi$. The total mass of the solenoid comes up to 220 t, whereas the total current that goes through it reaches 19.14 kA [6, 9, 32].

3.5 Muon Chambers

As implied by the name Compact Muon Solenoid, the detection of muons is one of its main responsibilities. The muons are charged particles, akin to electrons and positrons, but are 200 times heavier. Their detection is one of the most powerful tools in the study of the Higgs boson,

mainly via its decay to ZZ and ZZ^*, that in turn result in four charged leptons. If these leptons are muons, then their best mass resolution can be achieved for four particles. Since muons can penetrate metals without interacting for great distances, as opposed to other particles, their course cannot be intercepted by the calorimeters of CMS. Therefore, the muon chambers are placed outside CMS, since they are the only particles that give off signal at this point.
The muon chambers utilises three types of detectors: the Drift Tubes (DT), the Cathode Strips Chambers (CSC) and the Resistive Plate Chambers (RPC).
The 250 DTs that are in the barrels cover a pseudo-rapidity range $0 < |\eta| < 1.3$. The barrel con-

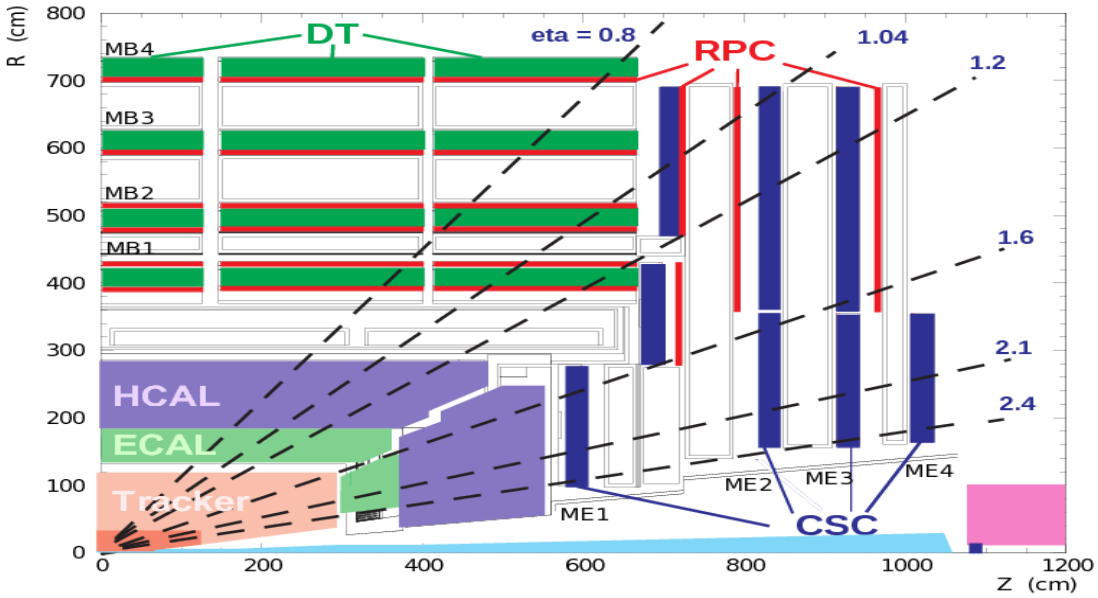


Figure 3.12: *Layout of one quadrant of CMS showing the muon systems* [33].

sists of four stations (MB1-MB4) connected to the return yoke of the magnet; two of them are positioned at the inner and the outer part of the yoke, while the other two are on nooks inside the metal. In total, 60 chambers comprise the inner stations and 70 are used for the outer stations. The chambers include a gas mixture $85\%Ar$ and $15\%CO_2$. When a muon is inserted to the chamber, it causes ionisation of its molecules and the electrons are detected by 172000 channels. The most basic and sensitive part of the DTs are the drift cells, 42×13 mm^2 that are characterised by a drift time of 400 ns. Through the time intervals needed for the detection of the ions the determination of the position of the muons is enabled.
The 540 Cathode Strip Chambers (CSC) are at the endcaps, where the number of muons and the background levels are high and the magnetic field is inhomogenic. They posses small response time, they are resilient to the inhomogenety of the magnetic field and cover a pseudo-rapidity region $0.9 < |\eta| < 2.4$. Each endcap has four trapezoid stations on top of steel disks, perpendicular to the beam. Each disk includes 18 to 36 chambers. The CSCs consist of six levels

(anode wires), each providing a measurement for two position coordinates of the muons. The anode wires are $30 - 50$ μm in diameter and are interlaid by seven cathode pannels (stripes) in outward radial direction. The surface of the largest chambers is 3.4×1.5 m^2 and they contain 50 m^3 gas mixture $50\%CO_2 + 40\%Ar + 10\%CF_4$.

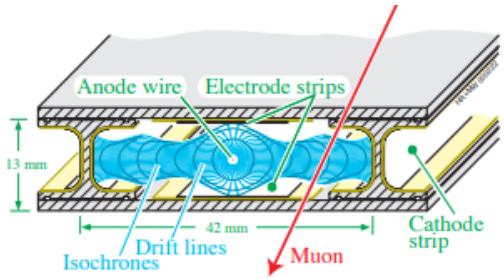


Figure 3.13: *Section of a drift tube cell showing the drift lines and isochrones* [34].

The Resistive Plate Chambers (RPC) are complementary systems that have an excellent time response. Therefore, they provide a definite measurement of the muons that come from every bunch crossing. There are 610 such chambers in total and they are placed on the barrel, as well as the endcaps, providing measurements for a wide pseudo-rapidity region $|\eta| < 1.6$. The RPCs posses double-gap chambers with resistive plates as walls, 2 mm wide, that are separated by $95.2\%C_2H_2F_4$(freon)$+4.5\%i - C_4H_{10} + 0.3\%SF_6$ in gaseous phase. When a particle passes through a RPC, it ionises the gas in both gaps that contain gas. The shower caused by the high electric field will induce an image charge that is read out by the strip detectors. The RPCs are organised in sectors. In the barrel area, there are 4 sectors and 3 in the endcaps [34–36].

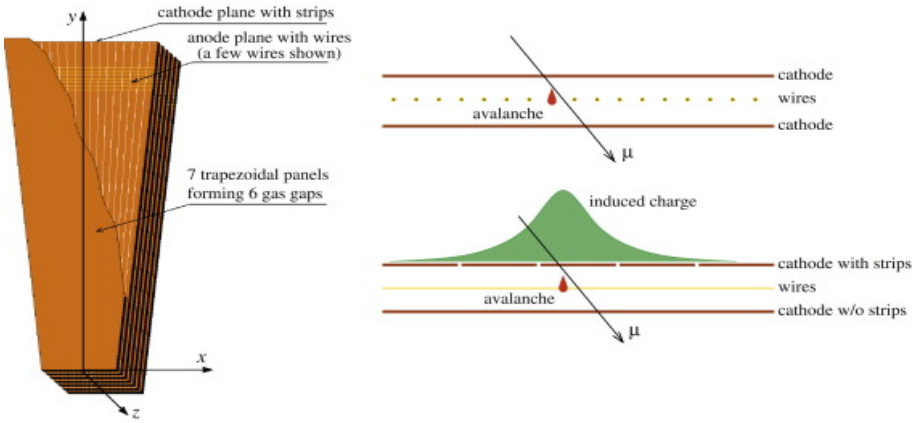


Figure 3.14: *Schematic view of a CMS cathode strip chamber (left) and an illustration of the CSC operation principle (right)* [37].

3.6 Trigger System

During the operation of CMS, 10^9 collisions take place between protons every 25 ns, corresponding to 40 MHz frequency. It can be easily understood that recording and saving this large amount of data can be difficult. Even if these data are saved, they will probably not lead to the discovery of new physics; for example, apart from the head-on collisions, there are also low-energy ones.

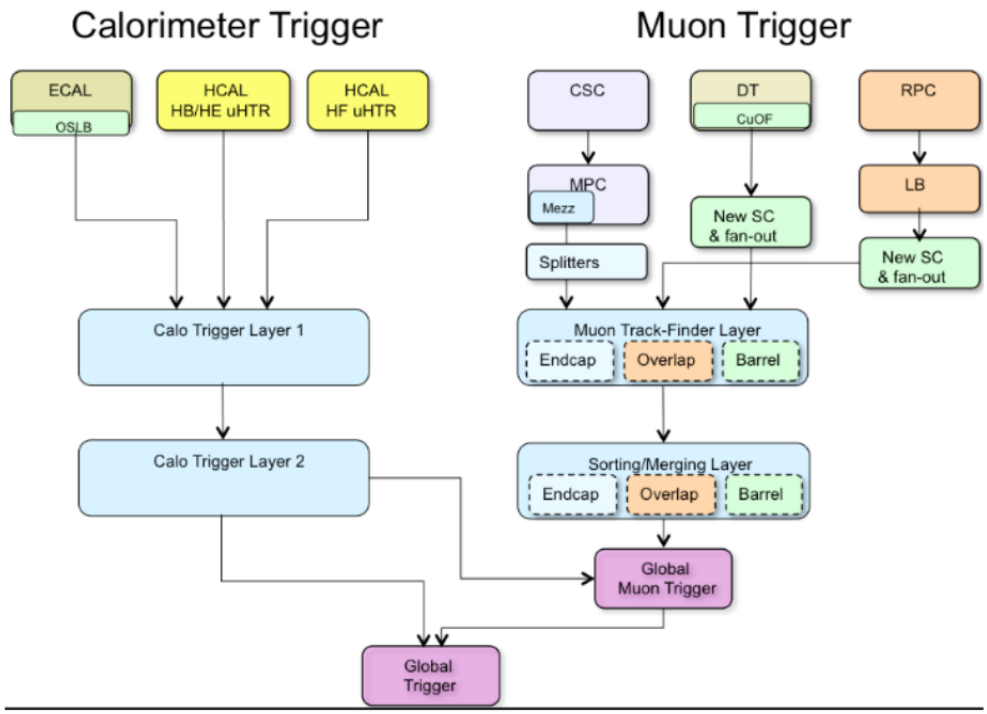


Figure 3.15: *Data-flow chart for the CMS Level-1 trigger upgrade* [39]

The Trigger System selects the most interesting events and therefore reduces the rate to a few hundred events per second. The data are then read-out and saved for analysis in disks. The Trigger System is the first stage when it comes to event selection and comprises of two systems: the Level 1 Trigger (L1T) and the High Level Trigger (HLT).
The L1T consists of especially designed hardware, such as FPGAs, LUTs and ASICs, its exit rate is up to 100 kHz and it receives raw data from the calorimeters and muon chambers. The L1T analyses events that come from all collisions and will determine the ones that will move on to be examined by the HLT, a process that lasts 3.2 μs. Some electronics are housed on the detector, whereas the rest are in an underground control room 90 m away from where the experiment is being conducted [12, 38].

The L1T includes three main subsystems: the L1 Calorimeter Trigger, The L1 Muon Trigger and the L1 Global Trigger. The L1 calorimeter trigger consists of two consecutive layers; Layer 1 receives the energy deposits from the calorimeters (ECAL and HCAL), calibrates, assorts them and sends them to Layer 2, where electrons and jets are being reconstructed. The L1 Muon Trigger comprises of three Muon Track Finders (MFT): the Barrel (BMFT), the Endcap (EMFT) and the Overlap (OMFT) Track Finders, that reconstruct muons from the respective muon detector. All the calibrated and reconstructed objects from the L1 Calorimeter Trigger and the L1 Muon Trigger are then taken to the Global Muon Trigger, where the muons are selected using specially designed algorithms for the trigger. The information from the Layer 2 Calorimeter Trigger and the Global Muon Trigger is transferred to the Global Trigger, that will

make the final decision of the L1T on whether the event will be accepted or not, by generating the L1 Accept (L1A) signal [39].
The HLT [40] includes software embedded in 1000 processors implemented in a computing farm of around 32k CPU cores. This is where the data that have passed through the L1T end up, and after imposing a number of criteria that aim to the selection of events, the total reduction of the event rate comes down to 10^3 Hz. The data are then written up in mass storage at CERN Tier0, where they are processed even further, and are later used in analysis [41].

Bibliography

[1] Fermilab official website – CMS collaboration publishes 1,000th paper. https://news.fnal.gov/2020/06/cms-collaboration-publishes-1000th-paper-2/.

[2] CERN official website – CMS. https://home.cern/science/experiments/cms.

[3] Klaus Rabbertz. *Jet Physics at the LHC: The Strong Force beyond the TeV Scale*. Springer International Publishing, 2017.

[4] The CMS Collaboration. Alignment of the CMS tracker with LHC and cosmic ray data. *JINST*, 9, 2014. https://doi.org/10.1088/1748-0221/9/06/P06009.

[5] LaTeX diagram for polar angle - pseudorapidity correspondence. https://wiki.physik.uzh.ch/cms/latex:tikz.

[6] V. Karimäki [The CMS Collaboration]. The CMS tracker system project: Technical Design Report. 1997.

[7] CERN official website – Detector. https://cms.cern/detector.

[8] CERN official website – Silicon Pixels. https://cms.cern/detector/identifying-tracks/silicon-pixels.

[9] The CMS Collaboration et al. The CMS experiment at the CERN LHC. *JINST*, 3, 2008. https://doi.org/10.1088/1748-0221/3/08/S08004.

[10] Marton Bartok. Simulation of the Dynamic Inefficiency of the CMS Pixel Detector. *JINST*, 10, 2015. https://doi.org/10.1088/1748-0221/10/05/C05006.

[11] CERN official website – Silicon Pixels. https://cms.cern/detector/identifying-tracks/silicon-pixels.

[12] Vito Manzari. Silicon Detectors. 2015.

[13] W. Adam et al. The cms phase-1 pixel detector upgrade. *Journal of Instrumentation*, 16, 2021. https://doi.org/10.1088/1748-0221/16/02/P02027.

[14] CMS Detector. https://cms.cern/book/export/html/1618.

[15] CERN official website – Tracking.

[16] CERN official website – Silicon Strips. https://cms.cern/detector/identifying-tracks/silicon-strips.

[17] Guido H. Dirkes. Construction of the CMS silicon strip tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 581, 2007. https://doi.org/10.1016/j.nima.2007.07.126.

[18] CERN official website – Silicon Strips. https://cms.cern/detector/identifying-tracks/silicon-strips.

[19] The CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. CERN, 2006.

[20] A. Benaglia. The CMS ECAL performance with examples. *JINST*, 9, 2014. https://doi.org/10.1088/1748-0221/9/02/C02008.

[21] Cristina Biino. The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections. *J. Phys.: Conf. Ser.*, 587, 2015. https://doi.org/10.1088/1742-6596/587/1/012001.

[22] The CMS Collaboration. *The CMS electromagnetic calorimeter project : Technical Design Report*. CERN, 1997.

[23] Cms ecal endcaps photographs. https://hepwww.pp.rl.ac.uk/groups/CMSvpt/bestphotos/867/slides/IMGP0329%20mod.htm.

[24] M. Brice. Images of the CMS ECAL Barrel (EB), Nov. 2008. https://cds.cern.ch/record/1431477.

[25] CERN official website - Energy of Hadrons (HCAL). https://cms.cern/node/1202/edit%3Fdestination%3D/node/1202.

[26] The CMS Collaboration. Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s}$ = 13 TeV. *JINST*, 2019. https://www.researchgate.net/publication/336208095_Calibration_of_the_CMS_hadron_calorimeters_using_proton-proton_collision_data_at_sqrts_13_TeV.

[27] Steffen Rocker. Perspektiven zur Beobachtung von Einzel-Top-Quark-Produktion am CMS-Experiment mit Hilfe neuronaler Netze. 2011.

[28] The CMS Collaboration. *The CMS hadron calorimeter project : Technical Design Report*. CERN, 1997.

[29] The CMS Collaboration. Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s}$= 13 TeV. *JINST*, 15 P05002, 2020. https://doi.org/10.1088/1748-0221/15/05/P05002.

[30] The CMS Collaboration. Performance of the cms hadron calorimeter with cosmic ray muons and lhc beam data. *Journal of Instrumentation*, 5, 2010. https://doi.org/10.1088/1748-0221/5/03/T03012.

[31] CERN official website – News. https://cms.cern/news/pay-attention-cms-magnet-powered.

[32] CERN official website – Bending particles. https://cms.cern/detector/bending-particles.

[33] Min Suk Kim. CMS reconstruction improvement for the muon tracking by the RPC chambers. *JINST*, 8, 2012. https://doi.org/10.1088/1748-0221/8/03/T03001.

[34] The CMS Collaboration. The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC. *JINST*, 8, 2014. https://doi.org/10.1088/1748-0221/8/11/P11002.

[35] CERN official website – Detecting Muons. https://cms.cern/detector/detecting-muons.

[36] The CMS Collaboration. CMS, the Compact Muon Solenoid. Muon technical design report. 1996. https://cds.cern.ch/record/1140134.

[37] D. Acosta et al. Efficiency of finding muon track trigger primitives in CMS cathode strip chambers. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 592, 2008. https://doi.org/10.1016/j.nima.2008.03.118.

[38] CMS official website – Triggering and Data Acquisition. https://cms.cern/detector/triggering-and-data-acquisition.

[39] A. Tapper and D. Acosta. CMS Technical Design Report for the Level-1 Trigger Upgrade. 2013. https://cds.cern.ch/record/1556311.

[40] Sergio Cittolin, Attila Racz, and Paris Sphicas. CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project. 2002. https://cds.cern.ch/record/578006.

[41] Laurent Thomas. CMS High Level Trigger performance at 13 TeV. *PoS*, ICHEP2018, 2019. http://cds.cern.ch/record/2703017.

Chapter 4

Jet measurement with CMS

Quantum Chromodynamics (QCD) is the gauge theory that describes the strong interactions between quarks and gluons. Jets, the collimated sprouts of hadron particles, inherit many properties of the original partons that were created in short-distance scatterings. Using perturbative QCD (pQCD) techniques, the production cross section for high transverse momentum (p_T) partons can be calculated. More specifically, predictions for hadron production from proton-proton (pp) collisions require the development of models for parton showering and non-perturbative (NP) corrections, such as hadronisation and underlying event (UE) activity. Fixed-order predictions from pQCD are inadequate for certain observables and kinematic configurations; that is why higher order terms must be accounted for using resummation methods. Studies of jet production can be used to further understand the modeling of strong interactions at short and long distances with great precision.

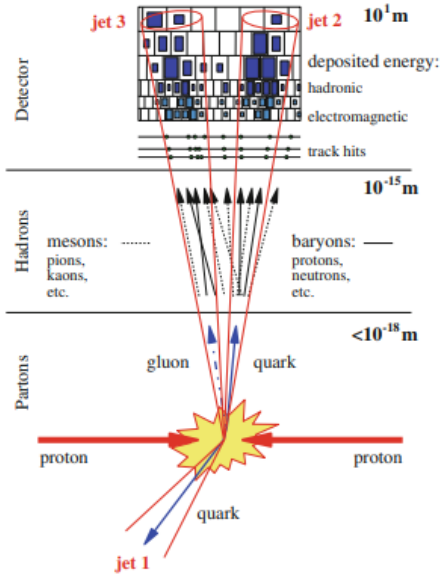


Figure 4.1: *Illustration of a jet to which bundles of partons, hadrons, or detector measurements are grouped together* [1].

One of these studies has to do with inclusive jet cross section for a certain value of $R = 0.8$, that will be discussed later. Properties of jets are modified by hadronisation, a NP process describing the transition of partons into hadrons, whose effects become more evident at larger p_T values. A measurement of the ratios of the inclusive jet cross sections is also included. Due to the fact that many theoretical and experimental uncertainties cancel out in the ratio, better sensitivity to perturbative and NP effects can be achieved [2].

When studying high energy collisions, processes that include the production of quarks and gluons in the final state have to be considered. In the LHC, there is a vast amount of processes involving quarks and/or gluons in the final state. Since protons are collided, a high QCD parton can be radiated form the incoming partons. Furthermore, other particles like the $W^{\pm}$, Z and Higgs bosons can also decay into quarks. Finally, there are decay chains into quarks and gluons.
Overall, the high energy partons produced by a collision appear in the final state as a collimated bunch of hadrons called jets. Conceptually, jets are collimated flows of hadrons and can be seen as proxies to the high-energy quarks and gluons produced in a collision. Such behavior is evident in experiments, where the hadronic final state appears to be aligned towards several directions in the detector. In order to identify the jets in an event, one relies on a jet definition, i.e., a procedure that gives information on how to reconstruct the jets from the set of hadrons in the final state of the collision.
A jet definition can be considered to be made of two essential components. The first is a jet algorithm which includes a set of parameters associated with free knobs in the algorithm. A typical parameter present in almost all jet definitions is the jet radius, which provides a distance in the rapidity-azimuth $(y-\phi)$ plane, above which two particles are considered as no longer part of the same jet. Additionally, a jet definition uses a recombination scheme which specifies how the kinematic properties of the jet are obtained from its constituents [3].

4.1 Jet production in pp collisions

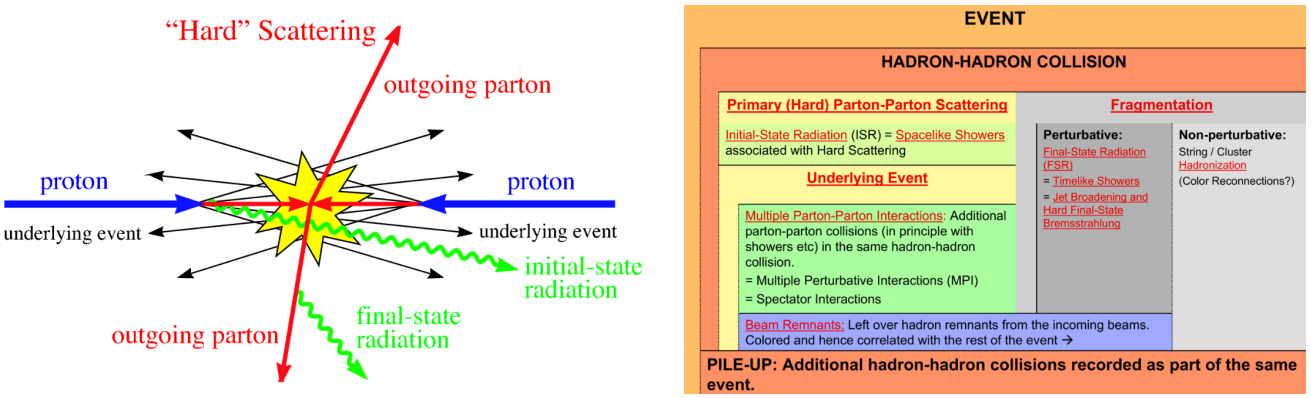


Figure 4.2: *Schematic cartoon of a* $2 \to 2$ *hard-scattering event* [4] *and Dictionary of Hadron Collider Terms* [5].

The interaction of two protons contained in beams includes numerous processes, starting from the long-distance interaction of the incoming protons, then continuing with the short-distance scattering from the incident beam particles and finally resulting in the outgoing particles that

are later measured by the detector. These processes are listed below [1, 5–7].

1. **Primary (Hard) Interaction:** They are short-distance parton-parton interactions that include large momentum transfer which results into hard partons and jets. The partonic cross sections are calculable using fixed-order perturbation theory in QCD. The relative probability to find the scattering partons at this step is provided by the parton distribution functions (PDFs), defined by an appropriate factorisation scale. These interactions correspond to $2 \rightarrow 2$ processes, where a parton from each proton (2 partons in total) is subject to the hard process and extracted as outgoing parton. In this kind of process the energy of the jet (in the partonic center-of-mass frame) is closely related to that of the parton in the proton that underwent a hard scattering. Of course, in the final state there can be larger multiplicity.

2. **Initial State Radiation (IRS):** The partons from the primary interaction can emit radiation prior to the short distance scattering yielding the possibility of initial state radiation (ISR). These processes can be calculated probabilistically in terms of summed leading logarithm perturbation theory and are described by the Parton Shower (PS) in the Monte Carlo event generators.

3. **Final State Radiation (FRS):** It is a color radiation step in which new quarks and gluons are added to the state, dominated by partons that have low energy and/or are nearly collinear with the scattered short distance partons.

4. **Underlying Event (UE):** Protons are composite objects and therefore, more than two partons, one from each proton, may interact when two protons scatter, known as Multi-Parton Interaction (MPI) or Douple Parton Scattering (DPS). This process corresponds to the transition between partons and hadrons, and occurs in all high-energy QCD processes. The remnants of the original hadrons are no longer color singlet states, due to the absence of one or more partons, they will interact softly with each other and hadronise, generating an underlying distribution of soft partons. Such a process consists an Underlying Event (UE).

5. **Hadronisation:** The final step in the evolution to long distance states involves a non-perturbative hadronisation process that organises the colored degrees of freedom from the showering and from the softer interactions of other initial state partons into color-singlet hadrons with physical masses. Hadronisation corresponds to the transition between partons and hadrons, and occurs in all high-energy QCD processes. The union of the showering and the hadronisation steps has been labelled as fragmentation, as in fragmentation functions describing the longitudinal distribution of hadrons within final state jets. Fragmentation Functions (FF) $D_{k\rightarrow h}(z, \mu_F)$ are the probabilities of a scattered parton fragmenting into a particular hadron h with fraction z of its momentum. They are universal and are

measured in $e^+ + e^-$ annihilation or semi-inclusive DIS experiments. The non-perturbative hadronisation process yields a collection of ground state hadrons (primarily pions) and resonances, and the resonances then decay into lighter hadrons such as pions.

6. **Pile Up (PU):** Since protons are formed into bunches, multiple pp collisions can take place in a given bunch crossing. Such soft collisions can occur during pile-up from multi-parton interactions at high luminosity. Such interactions can result in additional tracks and energy depositions in the detector calorimeters. The energy measured by the calorimeters can originate either from secondary interactions in the same bunch crossing as the primary (hard) interaction (in-time pile up or IT PU) or from previous or subsequent pp collisions caused by the infinite signal decay time in the calorimeters (out-of-time pile up or OOT PU).

4.2 Jet algorithms

In order to reconstruct a jet and proceed to comparisons between data and theoretical predictions, a precise definition of the jet is required, which takes the form of a jet algorithm. Jet algorithms cluster partons, particles or calorimeter towers based on proximity in coordinate or in momentum space. More specifically, jet algorithms provide a set of rules for grouping particles into jets and they usually include one or more parameters that indicate how close two particles must be so that they belong to the same jet. Additionally, jet algorithms include a recombination scheme that indicates the momentum that should be assigned to the combination of two particles, with the most simple one being the 4-vector sum. For the comparison between experiment and theory to be precise, the description of a hard scatter event a jet algorithm provides should preferably be similar regardless if it is applied at the detector, hadron or parton level. By definition a "good" algorithm will yield similar results whether it is applied to a state with only a few partons (as in NLO perturbation theory), a state with many partons (after the parton shower as simulated in a Monte Carlo), a state with hadrons (as simulated in a Monte Carlo including a model for the hadronisation step and the UE), or applied to the observed tracks and energy deposition in a real detector. Furthermore, the identification of jets needs to be insensitive to the contributions from the simultaneous uncorrelated soft collisions that occur during pile-up at high luminosity. Finally, one should be able to apply the same algorithm at each level in the evolution of the hadronic final state [1, 5].
In practice, a jet algorithm can be thought as a set of mathematical rules, carried out in two steps. The first step operates on the list of 4-vectors, which describes either the perturbative final state, the final-state hadrons in the MC simulated event or the output from the detector, to turn the original list into a set of sublists, one sublist for each jet (plus the beam jets). This

procedure takes place event by event. The second step operates as a guide on how to construct appropriate kinematic quantities from each sublist in order to describe the kinematic properties of the individual jets [5].
As noted, jet algorithms rely on merging nearby objects. They also need to be collinear- and infrared-safe in order to deal with the cancellation of collinear and soft singularities appearing in pQCD. Thus, the result of a jet-clustering procedure should depend neither on the splitting or merging of collinear parton four-vectors nor on the addition of arbitrarily soft partons to the list of objects to be clustered. There are two classes of jet algorithms:

- cone algorithms, that geometrically assign objects to the leading energy-flow objects in an event. They rely on the concept that QCD branching and hadronisation leave the energy flow of the event unchanged, especially the energy flow in a cone.
- sequential recombination algorithms, that repeatedly recombine the closest pair of particles according to some distance measure, usually related to the divergent structure of QCD matrix elements.

At the LHC the method chosen for the purpose of computing efficiency is the anti-k_T clustering algorithm which is implemented in the FastJet package [1, 7]. The FastJet package, written in C++, includes efficient native implementations for all the $2 \to 1$ sequential recombination jet algorithms for pp and e^+e^- collisions, as well as access to 3^{rd} party jet algorithms through a plugin mechanism, including all currently used cone algorithms. It also provides the means to facilitate the manipulation of jet substructure, including some common boosted heavy-object taggers, as well as tools for estimation of PU and UE noise levels, determination of jet areas and subtraction or suppression of noise in jets [8].
The anti-k_T clustering algorithm used at the LHC, belongs to the k_T-algorithms group of the broader category of sequential recombination algorithms. The k_T-algorithm is conceptually simple: two partons, particles, or energies in calorimeter towers are combined if their relative transverse momentum is less than a given value. The momentum space distance is given by:

$$d_{ij} = min\left\{p_{T,i}^{2p}, p_{T,j}^{2p}\right\}\frac{\Delta_{ij}^2}{R^2} \tag{4.1}$$

$$d_{iB} = p_{T,i}^{2p} \tag{4.2}$$

where d_{ij} is the distance between the i and j entities (particles or pseudo-jets), d_{iB} is the distance between the i entity and the beam B, $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is the distance in $y - \phi$ plane between each of partons, p is a parameter that governs the relative power of the energy versus geometrical (Δ_{ij}) scales, but also the sequence of the recombination, and R is a radius parameter that controls the size of the jet [9, 10].

In order to illustrate the clustering process, consider a multi-parton final state. Initially, at the parton level, each parton is considered as a proto-jet. The quantities in Equations 4.1 and 4.2 are computed for every parton and parton pair respectively. If the smallest of these quantities is d_{iB}, then the proto-jet is identified as a jet and it is removed from the proto-jet list. Similarly, if the smallest quantity is d_{ij}, the two proto-jets, i and j, are merged into a single proto-jet by summing their 4-vector components, and the two original entries in the proto-jet list are replaced by this single merged entry. This process is iterated with the corrected proto-jet list until all the proto-jets have become jets. Hence, at the last step the quantity d_{ij} for all pairs of proto-jets is constantly larger than all the distances d_{iB} for the proto-jets individually, i.e., the remaining proto-jets are well separated, and all the latter become jets [5].
The most commonly used values of the parameter p are:

— $p = 1$, for which one recovers the inclusive k_T clustering algorithm, where entities are clustered into jets in order of decreasing transverse momenta. In general, for $p > 0$ the behavior of the jet algorithm with respect to soft radiation is rather similar to that observed for the k_T algorithm, because what matters is the ordering between particles and for finite Δ_{ij} this is maintained for all positive values of p.

— $p = 0$, a case that corresponds to the inclusive Cambridge-Aachen algorithm, where the clustering sequence is based on the proximity of the $y - \phi$ space.

— $p = -1$, refers to the anti-k_t clustering algorithm, used at the LHC. This is the only case that yields cone-like jets. Similar behavior is manifested with respect to soft radiation for all $p < 0$ [9].

The parameter R, that corresponds to the jet size, usually ranges between $0.4 - 0.8$. For small values, non-perturbative corrections from the underlying event (UE) are found to be small, whereas as the cone radius increases, the UE corrections become larger. Also, for small values of R, perturbative $\ln R$ terms are observed, while when the jet size increases, there is a $1/R$ growth that corresponds to hadronic contributions. Thus, as the jet becomes narrow, partons radiated outside of it, i.e., partons which are not recombined with the jet by the chosen jet algorithm, are allowed to become more and more collinear to the emitter, approaching the collinear-singular configuration [11].

4.3 Event and Jet Reconstruction

Events in CMS are reconstructed using the Particle Flow (PF) technique, which reconstructs and identifies single particles through the combination of all subdetector information. The particles

reconstructed with the PF algorithm are jointly referred to as PF candidates; these lay into five categories: photons, electrons (including positrons), muons, charged hadrons, and neutral hadrons.

- The energy of photons is obtained directly from the ECAL measurement, when corrected for zero suppression effects.
- The energy of electrons is determined by combining the track momentum at the primary interaction vertex, the corresponding ECAL cluster energy, and the energy sum of all bremsstrahlung photons associated with the track.
- The energy of muons is obtained from the corresponding track momentum.
- The energy of charged hadrons is determined from a combination of the track momentum and the corresponding ECAL and HCAL energies, corrected for zero-suppression effects, and calibrated for the nonlinear response of the calorimeters.
- The energy of neutral hadrons is obtained from the corresponding calibrated ECAL and HCAL energies.

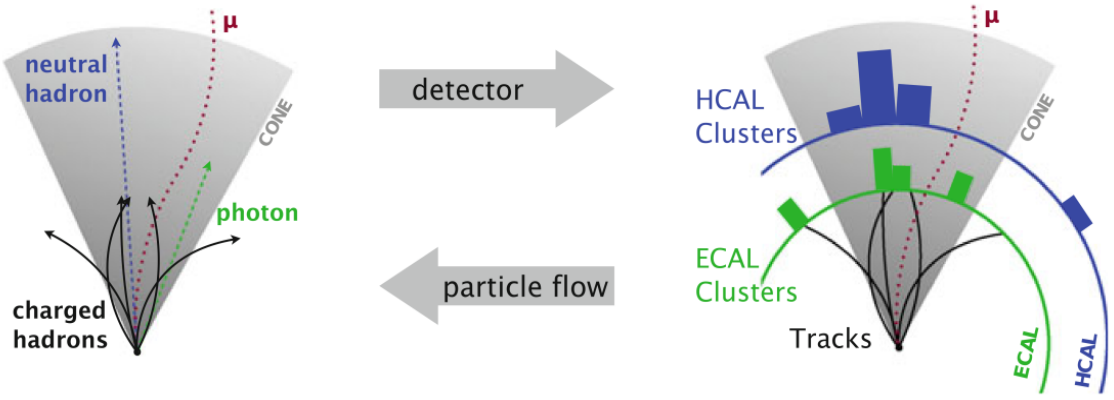


Figure 4.3: *Schematic association of subdetector measurements to physical particle candidates using the PF technique of CMS* [12].

The zero suppression effect is a procedure where only cells with energy above a given threshold are taken into consideration, in order to suppress noise in the calorimeters. In the forward region, energy deposits collected by the HF are considered as electromagnetic or hadronic, depending on the respective energy collected by long and short fibres. The PF candidates are used to derive the physical entities: jets, missing transverse momentum (p_T^{miss}), photons, electrons, muons and taus. Jets are reconstructed by clustering the PF candidates, and the missing transverse momentum p_T^{miss} is the negative vectorial sum of the transverse momenta of all PF candidates reconstructed

in an event. Missing transverse momentum reveals the presence of particles that do not interact with the detector material, e.g. neutrinos [12–14].
A given particle, in general, gives rise to several PF elements in the various CMS subdetectors. Therefore, the reconstruction of a particle proceeds with a link algorithm that connects the PF elements from different subdetectors, that is restricted to the nearest neighbours (tracks, clusters). The quality of the link is quantified by the distance between two linked elements. The link algorithm then produces a PF block of elements associated either by a direct link or by an indirect link through common elements. In each PF block, the muon candidates are reconstructed first, and the corresponding PF elements are removed from it. This is followed by the electron reconstruction, with the aim of collecting also the energy of bremsstrahlung photons. In the same step, energetic and isolated photons are also identified. The remaining elements in the block are then subject to the cross-identification of charged hadrons, neutral hadrons and photons. Secondary particles from nuclear interactions are reconstructed as well [15].
Particles reconstructed using the PF algorithm are being clustered into jets by the anti-k_T clustering algorithm. There are two types of reconstructed jets, depending on how the subdetector information is used: calorimeter jets and PF jets. The calorimeter (CALO) jets are reconstructed from energy deposits in the calorimeter towers alone. A calorimeter tower consists of one or more HCAL cells and the geometrically corresponding ECAL crystals. A 4-momentum is associated with each tower deposit above a certain threshold, assuming zero mass, and taking the direction of the tower position as seen from the interaction point. On the other hand, PF jets are reconstructed by clustering the 4-momentum vectors of PF candidates. The PF jet momentum and spatial resolutions are greatly improved with respect to calorimeter jets. In reconstructing the PF candidate 4-momenta, photons are assumed massless and to the charged hadrons the charged pion mass is assigned. Finally, all stable particles produced by the event generator excluding neutrinos are identified as RefJets [12–14].
The particles produced in PU interactions give rise to additional photons, charged and neutral hadrons. Consequently, reconstructed particles that originate from PU affect jets, E_T^{miss}, the isolation of leptons and the identification of hadronic τ decays. The charged hadrons that are unambiguously associated with the pileup vertices are removed from the list of PF particles used to form physics objects with an algorithm called charged hadron subtractions (CHS). These are called PFchs jets. CHS is able to reduce the number of PU jets by $\approx 85\%$ within the tracker acceptance of the overall pileup contribution.

4.4 Jet Energy Calibration

Like all reconstructed objects, jets need to be calibrated so that the correct jet energy scale (JES) can be assigned to them. This jet energy calibration or correction (JEC) procedure is performed in a series of steps (Figure 4.4), applied to both the data and the samples generated by the MC simulation. The first step accounts for extra energy caused by electronic noise or by PU collisions; other corrections are performed using detailed detector simulations. The third step includes the identification and correction of imperfections in the detector modelling through simulation with data-based methods as a function of jet pseudorapidity and p_T. In the final step, MC-based corrections account for differences in the flavour compositions of signal and calibration samples, if necessary.

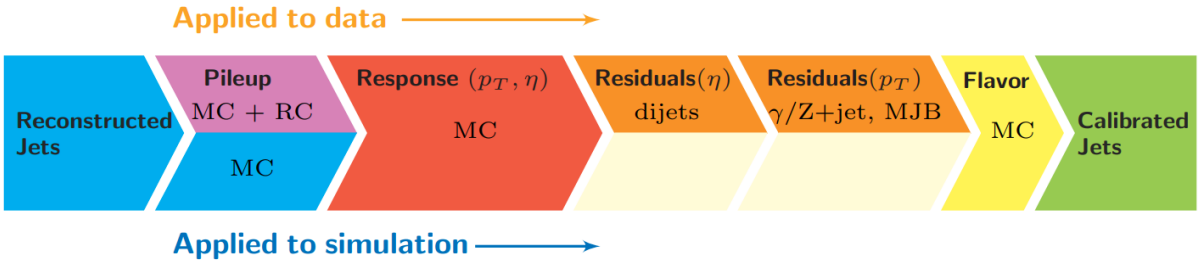


Figure 4.4: *Graphical illustration of the PF algorithm which is used to identify and reconstruct particles at CMS* [13]

Pileup offset corrections. In the first stage, the offset of the OOT PU and IT PU are studied, in order to optimise the subtraction of pileup from the data, with the corrections leading to an improved detector resolution and a more accurate JES. The observables used for monitoring and correcting pileup are p_T, η, the diffuse offset energy density ρ in the event and the jet area A. The OOT PU is mitigated by calorimeter signal processing, and the IT PU by identifying charged particles originating from pileup vertices and removing them with charged-hadron subtraction. The pileup jets are tagged with pileup jet identification (PUJetID) and removed. The remaining diffuse energy from neutral particles and OOT PU is estimated per event. Then, it is subtracted per jet using a calculation of the effective jet area with the extended hybrid jet area method. The dependence of the particle-level PU offset on jet η and p_T for this method is determined from simulation and the data/simulation offset scale factor is determined from zero-bias data and neutrino gun simulation, with the random cone (RC) method.

Simulated response corrections. Once the jets have been corrected for PU offset, the simulated response corrections (R_{ptcl}) are applied. The model that describes the CMS detector simulation contains in detail the detector geometry, data-based alignment and

calibration of the various detector components and emulation of the readout electronics. It is based on the GEANT4 package, that simulates the evolution of the electromagnetic and hadronic showers and their interactions with the materials that compose the detector. Additionally, the fragmentation of the initial quarks and gluons is based on QCD Multijet samples and is simulated using the Pythia6 tune $Z2^*$ event generator. These two components combined provide an accurate and detailed description of the jet response, which is used for the bulk of the JEC. The simulated particle response corrections, R_{ptcl}, account for non-uniformities in detector response which depends on p_T and η is defined as the ratio of arithmetic means of matched reconstructed and particle-level jets transverse momenta:

$$R_{ptcl}(\langle p_T \rangle, \eta) = \frac{\langle p_T \rangle}{\langle p_{T,ptcl} \rangle}[p_{T,ptcl}, \eta] \tag{4.3}$$

in bins of particle level $p_T(p_{T,ptcl})$ and reconstructed η. In the above equation, p_T is the transverse momentum of the reconstructed jets, whereas the notation $[p_{T,ptcl}, \eta]$ denotes the binning variables and the quantities inside the angle brackets indicate the averages within those bins for the variables that are used to parameterise the response.

Residual corrections for data. The evaluation of the detector simulated response does not include the imperfections that are present in the real detector. Furthermore, there are also variations of the jet momenta due to corrections for pileup. Thus, the data must be submitted to residual corrections, so that the latter are accounted for and the detector response remains the same with respect to p_T and η. The residual corrections for the data are first determined with a sample of dijet events with low statistical uncertainty, where the response of jets over a wide range of p_T is corrected relative to one of the jets with $|\eta| < 1.3$. Then, a combination of $Z(\rightarrow \mu\mu)+jet, Z(\rightarrow ee)+jet, \gamma+jet$ and multijet events for jets with $|\eta| < 1.3$ are used for correcting the dependence of the response on p_T. The basic notion is to make use of the the transverse momentum balance, at hard-scattering level, between the jet to be calibrated and a reference object: a jet energy scale different from unity generates imbalance at the reconstructed level.

Jet flavor corrections. For the latest step of the JEC estimation, the quark-initiated jets need to be distinguished from the gluon-initiated/tagged jets. The importance of this step, lays on the fact that the flavor of the particle that initiates the jet determines the composition of the jet particle. Besides, the QCD dijet sample is enriched in gluon jets, while the Z+jet and γ+jet samples are enriched in quark jets, which is why these samples are used for the evaluation of the response differences. Thus, jets resulting from hard gluon radiation without a matching parton have an undefined flavor. Variations in jet fragmentation energy and variations in particle composition of the jet lead to differences in the response for different jet flavors. Softer jet fragmentation results in more particles

outside the detector acceptance. With respect to particle composition, the response is mostly affected by the neutral hadron fraction of the jet. Jets from u and d quarks have the highest response, while those from gluons have the lowest, as a result of gluons fragmenting into the largest number of soft particles. The response values for heavy-flavor jets from c and b quarks are in between those from u/d and g jets due to additional soft particles from heavy-flavor hadron decays compared to u and d jets.

Ultimately, each correction step takes the form of a multiplicative factor C; therefore, the calibrated jet p_T will be calculated as:

$$p_{T,cor} = C \times p_{T,uncor} \tag{4.4}$$

where $p_{T,cor}$ and $p_{T,uncor}$ stand for the corrected, i.e. calibrated, and uncorrected jet transverse momentum respectively. Furthermore, each JEC stage comes with systematic uncertainties. These uncertainties are provided as systematic sources that include correlations across p_T and η. Their sum in quadrature will always correspond to the original uncertainty, and the quadratic sum of all the sources equals the total JEC uncertainty [13].

4.5 Monte Carlo simulations

Monte Carlo (MC) event generators have become an indispensable tool for analysing data from recent collider experiments, since they can be used to simulate the various processes that result from pp collisions. In order to obtain predictions of experimental observables, Monte Carlo simulations are employed for the integration over a final-state phase space. The large and variable dimension of the phase-space d for a n-particle final state, is calculated from the $d = 3n - 4$ dimensions [1], plus flavor and spin labels. Monte Carlo event generators are capable of simulating a wide range of processes expected at the LHC. More specifically, they can extract a signal of new physics from the background of SM processes. Comparisons of their predictions to the data can be used to perform measurements of SM parameters. They also provide realistic input for the design of new experiments, or for new selection or reconstruction procedures within an existing experiment [16].

In general, Monte Carlo techniques are based on numerical techniques that involve the repeated use of computer-generated pseudo-random numbers in order to compute integrals. Due to the large number of dimensions, but also the great number of particles that are produced, the approximate error of the integration is $\sigma \approx 1/\sqrt{N}$, with N denoting the integration points,

[1] Three components of momentum per produced particle, minus four constraints of overall energy-momentum conservation.

regardless of the dimensions d of the problem. Other numerical methods, not based on random number extractions, may suffer from severe computing time penalties as the number of dimensions d increases [17]. Many LHC processes of interest involve large momentum transfers, for example to produce heavy particles or jets with high transverse momenta. Thus the simulation of subprocesses with large invariant momentum transfer is at the core of any simulation of collider events in contemporary experiments through Monte Carlo event generators. At very high scales, the constituent partons of the incoming beams interact to produce a rather small number of energetic outgoing partons, leptons or gauge bosons. Hard scattering subprocesses of the form $ab \to n$ at hadron colliders can be computed in collinear factorisation through the corresponding matrix element squared, $|\mathcal{M}_{ab\to n}|^2$, averaged over initial-state spin and color degrees of freedom. At the lowest scales of the order of 1 GeV, incoming partons are confined in the beams and outgoing partons interact non-perturbatively to form the observed final-state hadrons. These soft processes cannot yet be calculated from first principles but have to be modelled. In the case of final states with low multiplicity, i.e. $2 \to 1, 2 \to 2$ or $2 \to 3$, the calculation is usually performed analytically based on theoretical algebraic expressions and for this reason all MC event generators include such pre-computed matrix elements. However, the matrix elements' numeric calculation for final-state multiplicities of four or larger requires more specialised techniques [18]. This evaluation is usually performed from dedicated matrix-element generators e.g Comix [19] and phase space integrators. Most of the current multi-purpose event generators currently employ leading-order (LO) matrix elements to derive the simulation [16]. However, the arising problem when matching the MC simulation to a next-to-leading order (NLO) matrix element, is that extra hard radiation can be generated from the real correction matrix element as well as from the parton shower which leads to a double counting problem. This is why a subtraction method is utilised in NLO calculations [20].

The simulation starts at the heart of the collision and continues by calculating from perturbation theory the probability distribution of a particular hard scatter, which is the highest momentum transfer process in the event. Simulating the hard process is relatively straightforward because Parton Distribution Functions (PDFs) describe partons coming into the process and lowest order perturbation theory gives a probabilistic distribution of the outgoing partons. The parton shower phase of event generators describes what happens to the incoming and outgoing partons involved in the hard collision. The parton shower can be simulated as a sequential step-by-step process that is formulated as an evolution in momentum transfer scale. Parton showers are based on approximations, namely: *(i)* only strongly ordered emissions are considered and *(ii)* the shower evolution leaves the total cross section invariant. The parton shower evolution starts from the hard process and works downwards to lower and lower momentum scales. As the event evolves downwards in momentum scales, it ultimately reaches the region, at scales of order 1 GeV, where the strong coupling becomes large and perturbation theory breaks down. Therefore, at this scale the perturbative evolution must be terminated and replaced by a non-perturbative

hadronisation model, which takes into account the confinement of a system of partons into hadrons, seen in the detector. When the collision happens, there is a very high probability that there will be other interactions apart from the hard interaction. This gives rise to the underlying event, which is made up of secondary interactions between proton remnants. It produces soft hadrons everywhere in the event, which overlie and contaminate the hard process that was already simulated. The last component of event generation is the fact that many of these hadrons are not stable particles but heavy resonances that then go on to decay. Finally, the simulation of soft QCD physics phenomena and the underlying event contributions is also model-based and includes a wide range of approaches and models included in the simulation and the relevant decay channels [16, 20, 21].
The Monte Carlo event generators used in the current analysis are:

1. **HERWIG++** [22], a general-purpose Monte Carlo event generator for the simulation of hard lepton-lepton and hadron-hadron collisions. It is built upon the HERWIG (Hadrom Emission Reactions With Interfering Gluons) program, and provides a much more flexible structure for further development. HERWIG++ is able to provide a full simulation of high energy collisions, with the matrix element being calculated at Leading Order (LO) for $2 \to 1$ decay and $2 \to 2$ QCD scattering processes. The Parton Shower (PS) is simulated by the angular ordering of successive emissions, while for the Hadronisation, the cluster model is used. The Underlying Event (UE) model of Herwig++ is currently based on the eikonal model [23] and its contribution is obtained from the Multiparton Interactions' (MPI) simulations, tuned to experimental data. Hadron decays are simulated using a matrix element description of the distributions of the decay products, together with spin correlations between the different decays, wherever possible. The tune used in the analysis at hand is the UE-EE-5-CTEQ6L1 or just EE5C [24], which is based on the CTEQ6.1M LO PDF set.

2. **PYTHIA8** [25], a standard tool for the generation of events in high-energy collisions between elementary particles, comprising a coherent set of physics models for the evolution from a few-body hard-scattering process to a complex multiparticle final state. The Matrix Element at LO is calculated for $2 \to 2$ scattering processes, but also Electroweak processes, top production, Higgs processes etc. Analysis of Pythia events can always be performed at the parton or particle level. In the PS simulation, the successive emissions are p_T ordered, whereas the hadronisation mechanism is based on the string model. UE and MPI effects are tuned to experimental data and particle decays are also simulated. The tunes used in the current analysis are: the CUETP8M1 [26] based on NNPDF2.3 LO PDF set, the CUETP8M2T4 [27] based on NNPDF3.0 LO PDF set and the CP5 [28] tune based on NNPDF3.1 NNLO PDF set.

3. **MADGRAPH5** [29, 30], a new version of the MadGraph matrix element generator, written in the Python programming language. This tool automatically generates matrix elements at LO for several decays and $2 \to n$ scattering processes, where n is the number of final-state partons that goes up to 4. Since MADGRAPH5 is not able to simulate the PS, UE and hadronisation, once the events have been generated, they are passed to any shower Monte Carlo program (such as HERWIG or PYTHIA with the MLM method [31]) where partons are perturbatively evolved through the emission of QCD radiation, and eventually turned into physical states (hadronisation).

Bibliography

[1] Tancredi Carli, Klaus Rabbertz, and Steffen Schumann. The Large Hadron Collider: Harvest of Run 1: Studies of Quantum Chromodynamics at the LHC. 2015. https://doi.org/10.1007/978-3-319-15001-7_5.

[2] Cristian Baldenegro. Jet measurements at CMS. *PoS*, ICHEP2020. https://doi.org/10.22323/1.390.0451.

[3] Simone Marzani, Gregory Soyez, and Michael Spannowsky. Looking inside jets: an introduction to jet substructure and boosted-object phenomenology. *Lecture Notes in Physics*, 958, 2022. https://doi.org/10.1007/978-3-030-15709-8.

[4] J.M. Campbell et al. Hard interactions of quarks and gluons: a primer for LHC physics. *Rep. Prog. Phys.*, 70, 2007. https://doi.org/10.1088/0034-4885/70/1/R02.

[5] S.D. Ellis, J. Huston, K. Hatakeyama, P. Loch, and M. Tonnesmann. Jets in hadron–hadron collisions. *Progress in Particle and Nuclear Physics*, 60, 2008. https://doi.org/10.1016/j.ppnp.2007.12.002.

[6] Sisharth K. Prasad, Supriya Das, Sanjay K. Ghosh, Premomoy Ghosh, Sanjib Muhuri, Tapan K. Nayak, and Rajarshi Ray. 'Soft' and 'Hard' Interactions in Proton-Proton Collisions at LHC Energies. *Proc. Indian Natn. Sci. Acad.*, 81, 2014. https://doi.org/10.16943/ptinsa/2015/v81i1/48071.

[7] G.P. Salam. Towards Jetography. *Eur. Phys. J. C*, 67, 2010. https://doi.org/10.1140/epjc/s10052-010-1314-6.

[8] Matteo Cacciari, Gavin P. Salam, , and Gregory Soyez. FastJet User Manual. *Eur. Phys. J. C.*, 72:1896, 2012. https://doi.org/10.1140/epjc/s10052-012-1896-2.

[9] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *J. High Energ. Phys.*, 04(2008)063, 2008. https://doi.org/10.1088/1126-6708/2008/04/063.

[10] J.M. Campbell, J.W. Huston, and W.J. Stirling. Hard Interactions of Quarks and Gluons: a Primer for LHC Physics. *Rept. Prog. Phys.*, 70:89, 2007. https://doi.org/10.1088/0034-4885/70/1/R02.

[11] Mrinal Dasgupta, Lorenzo Magnea, and Gavin P. Salam. Non-perturbative QCD effects in jets at hadron colliders. *J. High Energ. Phys.*, 02(2008)055, 2008. https://doi.org/10.1088/1126-6708/2008/02/055.

[12] Klaus Rabbertz. *Jet Physics at the LHC: The Strong Force beyond the TeV Scale*. Springer International Publishing, 2017.

[13] V. Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12, 2017. https://doi.org/10.1088/1748-0221/12/02/P02014.

[14] A.M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12, 2017. https://doi.org/10.1088/1748-0221/12/10/P10003.

[15] Milos Dordevic. The CMS Particle Flow Algorithm. *EPJ Web Conf.*, 191, 2018. https://cds.cern.ch/record/2678077.

[16] Andy Buckley et al. General-purpose event generators for LHC physics. *Physics Reports*, 504, 2011. https://doi.org/10.1016/j.physrep.2011.03.005.

[17] Luca Lista. *Statistical Methods for Data Analysis in Particle Physics*. Springer International Publishing, 2017.

[18] Stefan Gieseke. Simulation of jets at colliders. *Progress in Particle and Nuclear Physics*, 72, 2013. https://doi.org/10.1016/j.ppnp.2013.04.001.

[19] Tanju Gleisberg and Stefan Hoche. Comix, a new matrix element generator. *J. High Energ. Phys.*, 12(2008)039, 2008. https://doi.org/10.1088/1126-6708/2008/12/039.

[20] Stefan Gieseke. Monte Carlo Event Generators. *Nucl. Phys. B - Proc. Suppl.*, 222-224, 2012. https://doi.org/10.1016/j.nuclphysbps.2012.03.018.

[21] Ian Brock and Thomas Schorner-Sadenius. *Physics at the Terrascale*. John Wiley & Sons, 2011.

[22] M. Bahr, S. Gieseke, and M.A. et al Gigg. Herwig++ physics and manual. *Eur. Phys. J. C*, 58, 2008. https://doi.org/10.1140/epjc/s10052-008-0798-9.

[23] J.M.Butterworth, J.R.Forshaw, and M.H.Seymour. Multiparton Interactions in Photoproduction at HERA. *Z. Phys .C*, 2:637-646, 1996. https://doi.org/10.1007/s002880050286.

[24] M.H. Seymour and A. Siodmok. Constraining MPI models using σ eff and recent Tevatron and LHC Underlying Event data. *J. High Energ. Phys.*, 2013: 113, 2013. https://doi.org/10.1007/JHEP10(2013)113.

[25] Torbjorn Sjostrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191, 2015. https://doi.org/10.1016/j.cpc.2015.01.024.

[26] V. Khachatryan, A.M. Sirunyan, and A. et al. Tumasyan. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76:155(2016), 2016. https://doi.org/10.1140/epjc/s10052-016-3988-x.

[27] The CMS Collaboration. Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV. 2016. https://cds.cern.ch/record/2235192.

[28] A.M. Sirunyan, A. Tumasyan, and W. et al Adam. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80,4, 2020. https://doi.org/10.1140/epjc/s10052-019-7499-4.

[29] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: going beyond. *J. High Energ. Phys.*, 2011, 128, 2011. https://doi.org/10.1007/JHEP06(2011)128.

[30] Fabio Maltoni and Tim Stelzer. MadEvent: automatic event generation with MadGraph. *J. High Energ. Phys.*, 2003. https://doi.org/10.1088/1126-6708/2003/02/027.

[31] Michelangelo L. Mangano, Mauro Moretti, and Roberto Pittau. Multijet matrix elements and shower evolution in hadronic collisions: W$b\bar{b}$+n-jets as a case study. *Nucl. Phys. B*, 632, 2002. https://doi.org/10.1016/S0550-3213(02)00249-3.

Chapter 5

Multijet cross sections and ratios at $\sqrt{s} = 13\ TeV$ with the CMS experiment

The proton collisions at the LHC are viewed as interactions among their constituent partons, (anti-)quarks and gluons. These interactions can be either hard or soft collisions, that correspond to high and low momentum transfer respectively. In the context of perturbative QCD (pQCD), the scattering cross section can be expressed as a sum of strong coupling constant, α_S, terms of increasing power, convoluted with parton momentum distributions inside the proton. The production of two-parton final states is represented by the lowest order α_S^2, whereas the higher order terms α_S^3 and α_S^4 represent the production of three and four partons in the final state [1]. At high momentum transfer Q, i.e. at small distances, the strong coupling α_S becomes weaker. Even though theoretical predictions do not provide any values for α_S, the renormalisation group equation (RGE) described in Equations 1.74 and 1.76, describes the dependence of α_S on the renormalisation scale μ_r and therefore on the momentum transfer Q. Thus, experimental measurements can be utilised to extract α_S at a specific scale; this scale is by convention the $Z-$boson mass M_Z. Then, using the RGE to describe the evolution of the strong coupling, the running of α_S at higher scales Q can be tested.

The distribution of partons with the protons is estimated through the study of inclusive jet production ($p + p \rightarrow jet + X$), where every jet is counted. The inclusive jet cross section as a function of jet p_T and y is one of the most important observables, since it provides essential information on the PDFs and a precise measurement of the strong coupling α_S. There is high probability for the final partons to radiate additional gluons, forming multi-jets in the final state after hadronisation. The numerous events containing multi-jets can be used for various QCD analyses. The multi-jet event cross section σ_{i-jet} that describes a proton collision of the form $pp \rightarrow j\ jets + X$ with $j \geq i$ is proportional to α_S^i.

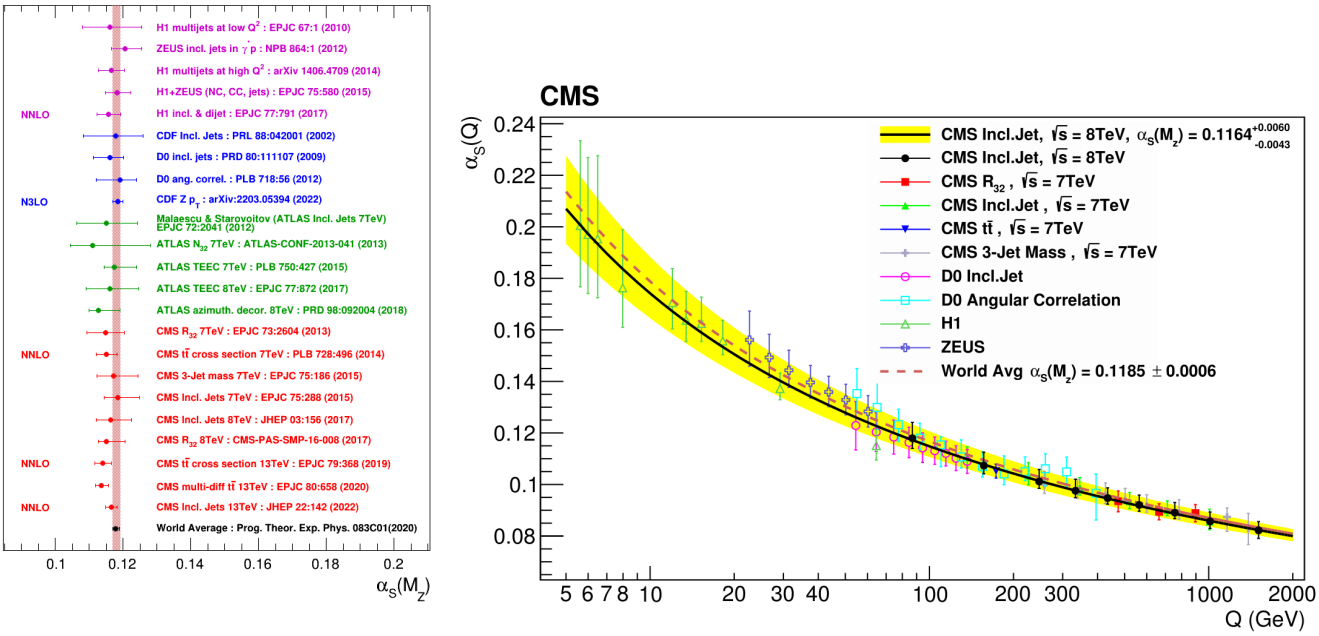


Figure 5.1: *An overview of determinations of the strong coupling constant at the scale of the Z-boson mass from measurements using hadrons (left)* [3] *and the strong coupling constant running α_S (Q) as determined in* [4] *(right).*

More often than not, instead of studying inclusive jet cross sections, it is useful to calculate their ratios, where many theoretical and experimental uncertainties due to luminosity, scale dependence, PDF dependence etc., may cancel between numerator and denominator. The ratios of such cross sections are defined as: $R_{mn} = \frac{\sigma_{m-jet}}{\sigma_{n-jet}}$ with $m > n$, and are proportional to α_S^{m-n}. Hence, these ratios can be used to determine the strong coupling α_S. Previous analyses of the ratio R_{32} as a function of the average transverse momentum, $\langle p_{T1,2} \rangle$, of the two leading p_T jets in the event was performed at $\sqrt{s} = 7$ TeV by the CMS Collaboration and lead to an extraction of $\alpha_S = 0.1148 \pm 0.0055$, where the dominant uncertainty stems from the estimation of higher-order corrections to the next-to-leading order (NLO) prediction [2]. Another measurement of the ratio R_{32} was performed once again by the CMS Collaboration at $\sqrt{s} = 8$ TeV, again, as a function of the average transverse momentum of the two leading jets in the event, $H_{T,2}/2$, giving $\alpha_S(M_Z) = 0.1150 \pm 0.0023$ (all except scale) $^{+0.0050}_{-0.0000}$ (scale) [1].

5.1 Observables

The inclusive differential multijet cross sections are measured as a function of the average transverse momentum of the leading and sub-leading jets, $H_{T,2}/2 = \frac{1}{2}(p_{T,1} + p_{T,2})$, where $p_{T,1}, p_{T,2}$

denote the transverse momenta of the two leading jets. The inclusive differential jet event cross section is defined as:

$$\frac{d\sigma}{dH_{T,2}/2} = \frac{1}{\mathcal{L}} \frac{N_{event}}{\Delta(H_{T,2}/2)} \tag{5.1}$$

where $\mathcal{L}$ is the effective integrated luminosity, N_{event} is the number of events counted in a $H_{T,2}/2$ bin and $\Delta(H_{T,2}/2)$ is the bin width. The measurements are normalised by the bin width $(dH_{T,2}/2)$ and hence, are reported in units of (pb/GeV). The inclusive differential multijet cross sections are measured for the inclusive 2-, 3-, 4- and 5-jet events and therefore, the number of events N_{event} will stand for the counts of 2-, 3-, 4- and 5-jet events in each bin.
Then, ratios of the form R_{mn} as a function of $H_{T,2}/2$ are studied, which stand for the ratio of the differential cross section of selected inclusive m-jet events to that of inclusive n-jet events in each $H_{T,2}/2$ bin, with $m > n$ and $m \neq n$:

$$R_{mn} = \frac{\left[\frac{d\sigma}{d(H_{T,2}/2)}\right]_{(m-jet)}}{\left[\frac{d\sigma}{d(H_{T,2}/2)}\right]_{(n-jet)}} = \frac{\sigma_{m-jet}}{\sigma_{n_jet}} \propto \alpha_S^{m-n} \tag{5.2}$$

More specifically, the ratios R_{32}, R_{43} and R_{42} are calculated, using the distributions that correspond to the 2-, 3- and 4-jet cross sections. The first two of these ratios are proportional to α_S, whereas the last ratio is proportional to α_S^2.
The measurements are based on the data samples collected by the CMS detector at the CERN Large Hadron Collider during the 2018, for pp collisions with center-of-mass energy of 13 TeV, corresponding to a luminosity of 59.82 fb^{-1}. The inclusive jets sample contains jets within the tracker acceptance $|y| \leq 2.5$ that have transverse momenta $p_T > 150$ GeV.

5.2 Data and Monte Carlo samples

The CMS Computing Model utilises the hierarchy of computing Tiers that forms the Worldwide LHC Computing Grid (WLCG). The WLCG receives and processes the data acquired from the CMS Data Acquisition (DAQ) and Trigger Systems. The first tier center, Tier-0, is devoted to (*i*) making a safe temporary copy and performing prompt first pass reconstruction of the data, containing the raw information from the detector (RAW data) and (*ii*) distributing the FEVT data (RAW and RECO components) to external Tier-1 centres. The next tier center is the Tier-1; the main tasks it carries out are (*i*) the secure and long-term storage of the raw and reconstructed data sent by the Tier-0 center, (*ii*) the running of second-pass reconstruction or other large-scale workflows (Reconstruction Services) and also direct analysis or processing

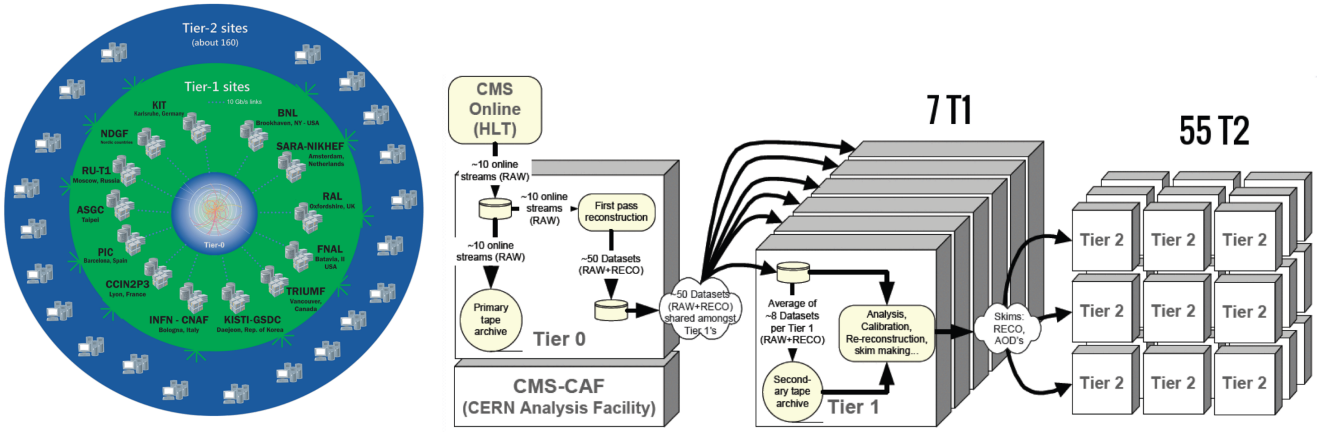


Figure 5.2: *Graphical illustration of the Worldwide LHC Computing Grid (WLCG) tiers (left)* [5] *and the real event data flow in the CMS Computing Model (right)* [6].

in support of analysis elsewhere, such as skimming (reduction of number of events based on specific trigger bits or high level reconstructed quantities), and finally (*iii*) the transfer of the reconstructed data and skims to Tier-2 centers for further analysis. The last tier center is Tier-2; its basic functions include (*i*) fast and detailed Monte Carlo event generation, (*ii*) the medium- or long-term storage of required data samples (AODs, RECO and RAW data), (*iii*) intended for processing in several physics analyses but also calibration, alignment tasks, and detector studies.

The typical data analysis flow in CMS, as well as the other LHC experiments, comprises of several steps of data processing and reduction. The arrangement of the event information from the several steps of the data flow in CMS is based upon the hierarchy of tiers. The available data tiers in CMS are the RAW, RECO and AOD. The RAW contain the detector data after online formating, the L1 trigger result, the result of the HLT selections ("HLT trigger bits"), and some of the higher-level objects created during HLT processing. Their event size is about $1 - 2MB/event$ and thus, they cannot be used for analysis. The RECO include all reconstructed objects derived from RAW data. They contain information from all the reconstruction stages: the reconstructed objects (tracks, vertices, jets, electrons, muons, etc.) but also reconstructed hits and clusters. The reconstructed event format (RECO) is about $3MB/event$, hence they are too heavy to be used for analysis. AOD are derived from the RECO information to provide data for physics analysis

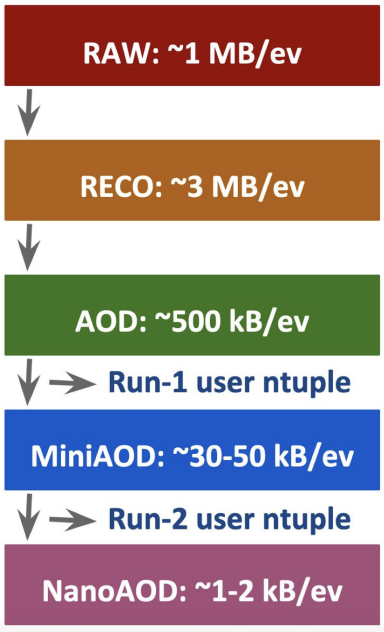


Figure 5.3: *The CMS main data format* [7].

Table 5.1: *Integrated luminosity and the 2018 dataset*

Year	$\mathcal{L}_{int}$ **(fb^{-1})**	**Dataset**
2018	59.82	/JetHT/Run2018A-UL2018_MiniAODv2_GT36-v1/MINIAOD /JetHT/Run2018B-UL2018_MiniAODv2_GT36-v1/MINIAOD /JetHT/Run2018C-UL2018_MiniAODv2_GT36-v1/MINIAOD /JetHT/Run2018D-UL2018_MiniAODv2-v2/MINIAOD

in a convenient, compact format. It will contain a copy of all the high-level physics objects (such as muons, electrons, taus, etc.), plus a summary of the RECO information sufficient to support typical analysis actions such as track refitting, with event size $400 - 500kB/event$ [8]. From multiple copies of AOD spread across many CMS computing sites, intermediate datasets called n-tuples are generated, that have more specialised formats intended for particular analysis needs. For Run 1, CMS physics analysis was based upon the Analysis Object Data (AOD) format but turned out to be too large for it to be used in the same way again. For this reason, the new CMS data flow for Run 2 was a more condensed dataset called Mini-AOD [9] format with event size $40 - 50kB/event$ and the Nano-AOD [10] format, with a rate of with $1 - 2kB/event$, that mimics the typical format of user n-tuples and is expected to be the main tier for Run 3.
The stream of events acquired by CMS is organised in primary datasets, according to the results of the High Level Trigger (HLT) selection. For an event to be recorded by CMS, it has to be accepted by the first level hardware trigger and to satisfy at least one of the composite HLT selections which are commonly referred to as paths. The primary datasets are defined by a set HLT paths, and collect events which have passed at least one of them. Such paths for example are the $DoubleMu$, $JetHT$, $SingleElectron$ etc. The design of the primary datasets is mainly centered around particle candidates reconstructed in the event final state by the HLT, and follows one basic principle: grouping together events with similar physics content [11].The current analysis is based on the $JetHT$ datasets in the Mini-AOD format, where each contains the aggregation of events collected in a different LHC run period (era) during 2018. The complete dataset list used in this analysis corresponding to the 2018 Run is shown in Table5.1.

The data that all physics analyses are based on should be certified as good by the detector and physics objects experts. Hence, they are defined as good when all the sub-detector systems are fully operational and the reconstruction and calibration conditions are optimal. So, physics analyses are only based on the good lumisections from each run, which are specified in the so-called Golden JSON files. In this analysis the official JSON file (recommended by the PdmV group [1]) used is 2018: Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt
Similarly, Monte Carlo simulation samples are arranged in data tiers according to the information

[1]The Physics Data and Monte Carlo Validation (PdmV) group is responsible (among others) for the evaluation of Physics performance and validation of Monte Carlo and data samples coming from (pre-)production campaigns (RelVals), prompt reconstruction, re-reconstructions and skims.

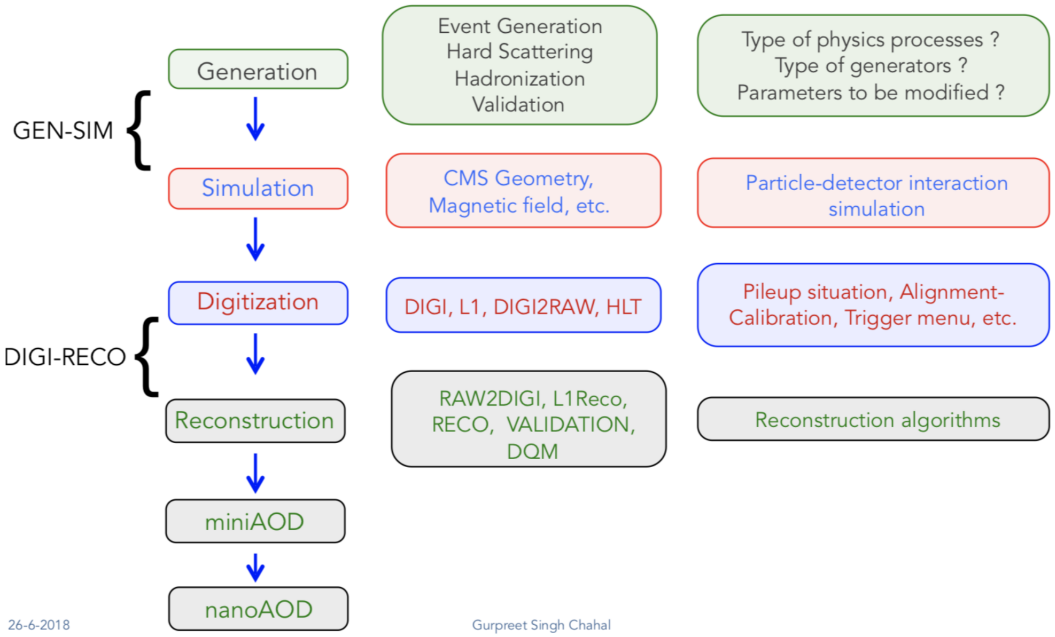


Figure 5.4: *A summary workflow for the production of a Monte Carlo sample, image credits: Gurpreet Chahal* [13]

they contain. The production of a Monte Carlo sample starts with the derivation of events at generator level (GEN), i.e. the manufacture of the four vectors of the particles by simulating the hard scattering, the multiparton interaction, the hadronisation etc, using a Monte Carlo event generator like Pythia8 or Sherpa. This is followed by the simulation (SIM) of the generated particles' interaction with the detector's material, based on the GEANT4 toolkit that is able to simulate the CMS detector. The format of the first Monte Carlo sample is called GEN-SIM, a combination of the two individual formats. When digitised, reconstruction algorithms are applied on the simulated detector signals, including the simulated hits from pileup interactions and the trigger menu information. These events are called DIGI-RECO with the corresponding output format of the Monte Carlo sample being AODSIM. Lastly, the skimmed versions of AODSIM, the Mini-AODSIM and Nano-AODSIM, contain only the necessary information required for physics analyses [12]. The official CMS Monte Carlo samples used in the analysis are shown in Table 5.2.

Table 5.2: *Offcial CMS Monte Carlo samples used in the analysis.*

Year	**MC samples**
2018	1. QCD_Pt_xxtoxx_TuneCP5_13TeV_pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v1/MINIAODSIM 2. QCD_HTxxtoxx_TuneCP5_13TeV-madgraphMLM-pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/MINIAODSIM

5.3 Software, Jet Reconstruction and Event selection

Analysis software

The manipulation of the Data and the Monte Carlo samples required for this analysis was based on the Data Analysis System (or for short simply DAS), a generic framework for jet analysis, currently used by the Standard Model Physics - Hadronic (SMP-HAD) group for several analyses, such as the inclusive jet, the dijet etc. The DAS framework is a ROOT-based [14] environment that includes the CMS Software (CMSSW) [15] and follows the same strategy as the SMPJ framework [16] used by the Standard Model Physics - Jet (SMPJ) group dedicated to the analysis of the data collected during Run 1. All members of the collaboration have access to the code [17] intended for numerous analysis purposes, from the derivation of the n-tuples from the Mini-AOD samples, to the final stages of the analysis. For the current analysis, the recommended version of the CMS Software 12_2_0 for the processing of the Run 2 UltraLegacy (UL) samples was used.

Jet Reconstruction

For the event reconstruction the PFchs algorithm was used. In general, all particles are reconstructed and identified by a Particle Flow (PF) algorithm, which combines the information from the individual subdetectors. As for the jet reconstruction, it was performed by the anti-k$_T$ clustering algorithm, that takes as input the four-vectors of particle candidates reconstructed by the above technique. The radius resolution parameter was set to R=0.8, the default size for "large" jets, a choice made between the available AK4 and AK8 jets in the centrally produced Mini-AOD samples. This distance parameter value allows a direct comparison with former R_{32} analyses, conducted for the Run 1 data.
The Jet Energy Calibration was performed using the recommended Jet Energy Corrections (JEC) provided by the JetMET group [2]. The global tag that specifies the corresponding JEC versions for both Data and MC samples is given in Table 5.3.

Table 5.3: *Jet Energy Corrections Global Tags for Data and MC samples.*

Year	**Data**	**MC**
2018	Summer19UL18_V5	Summer19UL18_V5

[2]The JetMET (JME) Physics Object Group (POG) is responsible for monitoring, reconstructing, calibrating, and providing scale factors and software tools for jets and missing energy in CMS.

Event selection

A primary vertex (PV) is identified by a collection of tracks measured in the tracker with a good fit quality between the hits and compatible with the beam line. The tracks are clustered according to the z-coordinate of their point of closest approach to the beam axis. To select good events, certain event selection cuts are applied to reduce beam induced background, detector level noise and jets arising from fake calorimeter energy deposits. The selection cuts performed on each event are the following:

- Require at least one well reconstructed primary vertex.
- Require the distance (z component) of the PV to be $|z(PV)| < 24$ cm.
- Require the radius in the xy plane of the PV to be $rho < 2$ cm.
- Require the vertex fit number of degrees of freedom $(ndf) > 4$. Thus, at least four tracks must be present in order to perform a valid vertex fit.

where $|z(PV)|$ represents the position of the proton-proton collision along the beam-line and $z = 0$ indicates the center of the CMS detector. As already mentioned, the measurement of the R_{mn} observables is performed on an inclusive jet sample with transverse momentum $p_T > 150$ GeV and rapidity $|y| < 2.5$. More specifically, the event selection is carried out as follows:

- Events with at least two jets with $p_T > 150$ GeV and $|y| < 2.5$ are selected.
- All the rest of the jets having $p_T > 150$ GeV and $|y| < 2.5$ are selected.
- The multiplicity is deduced from the remaining list of jets in the event.

5.4 Monte Carlo sample processing

The Monte Carlo sampled shown in Table 5.2, provided by CMS, are subjected to the following steps of processing, before used in the analysis.

Cross section normalisation

The samples from Pythia8 and Madgraph are split in $\hat{p}_T$ [3] and H_T [4] slices. This is why before combining the different slices, they should be normalised with the corresponding cross section obtained from the cross section database [18].

Corrections on the PU simulation

The Pile Up (PU) effects described in Section 4.1 can also be included in a Monte Carlo sample. But for the correct simulation of the PU, the following two points should be taken into consideration:

- **Removal of overweighted PU events**
  Even though the MC samples are split into slices in order to obtain sufficient statistics over the whole jet p_T phase space, the simulation of the PU is performed independently of the $\hat{p}_T$ and H_T slices. Consequently, the normalisation of each slice with the corresponding cross section results in jets originating from the PU simulation to appear in larger p_T values than configured with the slicing method. For example, a jet with $p_T = 300$ GeV might appear in the 30-50 jet p_T spectrum $\hat{p}_T$ slice, which is not physical. This problem is resolved by ensuring that no low-$\hat{p}_T$ slice can contribute more effectively that any higher $\hat{p}_T$. Thus, events where $(max)\hat{p_T^{PU}} > \hat{p_T^{main}}$, as well as contributions with unphysical weights and contributions to bins with less than 100 entries, are removed.

- **Re-weighting of the simulated PU profile**
  The PU in the simulation is usually overestimated compared to the PU in the experimental data. The amount of PU can be quantified using the Pile-Up (PU) profile, which corresponds to the probability distribution describing the number of interactions per bunch crossing. For the data, the PU is estimated using the pileupCalc.py utility, provided by the Lumi POG[5] and is implemented within the CMSSW. The PU distribution for individual events corresponds to a Poisson distribution with mean value μ, calculated as:

$$\mu = \frac{L_{inst}\sigma_{inel}}{f_{rev}} \tag{5.3}$$

  where L_{inst} is the instantaneous luminosity, σ_{inel} is the total inelastic cross section and f_{rev}

[3] For a $2 \to 2$ LO process like in Pythia8, $\hat{p_T}$ corresponds to the transverse momentum of either of the two outgoing particles of the hard process.

[4] The scalar sum of the jets transverse momenta.

[5] Luminosity Physics Object group, responsible for the determination of the absolute luminosity in CMS.

is the LHC orbit frequency 11246. The recommended value for the Run II cross section is 69.2 mb [19] while the instantaneous luminosity is obtained per lumisection from the JSON files. The L_{inst} is assumed to remain constant within one lumisection, but it may vary for longer time periods leading to a non-Poissonian behaviour. On the other hand, in simulation the PU can be simulated from the consideration of several Poisson distributions which correspond to different PU configurations. Eventually, PU profile inconsistencies between data and MC are faced with the reweighting of the MC PU profile in order to match that of the data.

Jet Energy Calibration

The simulated jets are calibrated with the procedure described in Section 4.4, using the Jet Energy Corrections (JECs) contained in Table. 5.3.

Jet Energy Resolution - Smearing of reconstructed jets

The production of a MC sample incorporates simulating the interaction of generated particles with the CMS detector material. The resolution of the jet energy measurement (JER) in simulation is better than in data. For this reason the reconstructed jets in simulation are smeared, so that their p_T resolution matches the one observed in data. For the smearing of reconstructed jets two different methods are used, both relying on the Scale Factors (SFs) provided by the JetMET group, to account for the data-MC difference [20]

- **Scaling method**
  The 4-momenta of the reconstructed jets are rescaled with the factor:

$$c_{JER} = 1 + (s_{JER} - 1)\frac{p_T^{rec} - p_T^{gen}}{p_T^{rec}} \tag{5.4}$$

  where p_T^{rec} is the reconstructed jet p_T, p_T^{gen} the p_T of the jet at generator (particle) level and s_{JER} is the data-to-simulation core resolution factor provided by the JetMET group. This method is performed under the assumption that for each reconstructed jet, there is a well-matched generated jet, matched with the following criteria:

$$\Delta R < R_{cone}/2, \qquad |p_T^{rec} - p_T^{gen}| < 3\sigma p_T^{rec} \tag{5.5}$$

  where R_{cone} is 0.8, $\Delta R = \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$ and σ is the relative p_T resolution from simulation.

- **Stochastic method**
  The 4-momenta of the reconstructed jets are rescaled with the factor:

$$c_{JER} = 1 + \mathcal{N}(0, \sigma)\sqrt{max(s^2_{JER} - 1, 0)} \tag{5.6}$$

  where the only new variable is $\mathcal{N}(0, \sigma)$ that represents a random number sampled from a Gaussian distribution with 0 mean and σ^2 variance.

According to recommendation from the JetMET group, a "hybrid" method is used here, in the sense that the scaling method is used when the matching is possible, otherwise the stochastic method is applied. The Global Tag (GTs) for the SFs that corresponds to the 2018 Data is Summer19UL18_JRV2.

5.5 Trigger studies

As discussed earlier in Section 3.6, events produced by collisions in CMS are recorded by the Trigger System that is composed of the Level-1 trigger, implemented in the hardware, and the software-based High Level trigger (HLT). The events used in the current analysis were collected with High Level single AK8PFJet (HLT_AK8PFJETX) triggers, that require an event to have at least one PFJet reconstructed with the AK8 algorithm and with a p_T value above a given threshold X (in GeV). For example, HLT_AK8PFJET40 requires a p_T threshold above 40, $p_T > 40$ GeV.
Generally, the rates each trigger records are adjusted due to the restricted trigger bandwidth. In order to reduce the rate, a prescale, that determines the fraction of events that satisfy the conditions of the trigger, is applied. The N prescale denotes the acceptance of 1 out of N events; for example, a prescale of 100 signifies that only 1 in every 100 events satisfying the trigger conditions are recorded. Furthermore, in each LHC fill, the beam intensity decreases with time. So, for the total HLT rate to remain at around 1 kHz and the signal efficiency to be maximised, dynamic prescales (prescale columns) with decreasing values are used. Since both the HLT paths and L1 seeds are prescaled, each event needs to be re-weighted by the L1 times the HLT prescales. The prescale information is used for the data normalisation and it is available in the Mini-AOD samples.
The effective luminosity is the active luminosity times the L1 and HLT prescales, whereas the trigger efficiency ϵ is defined as:

$$\epsilon = \frac{n_T}{n_R} \tag{5.7}$$

where n_T is the number of triggered objects, i.e. the objects that fired the trigger, and n_R is the number of all reconstructed objects, i.e. the objects obtained by an offline reconstruction algorithm [21]. However, the real value at which the trigger starts recording data does not concur with the nominal, where the trigger is 100% efficient. That is why the efficiency for each trigger is calculated and then the point where it becomes 100% efficient is obtained. In this analysis, the evaluation of the trigger efficiencies was performed using the emulation method, except for the HLT_AK8PFJet40, for which the Tag&Probe method was utilised instead. The emulation method is described by the following steps [22]:

- Start with all events that fired a reference trigger. As reference trigger the previous fired trigger is taken (the one with lower threshold).
- Access the L1 and HLT objects.
- Find the number of events that satisfy the L1 and HLT conditions of the path of interest.
- Apply the offline selection cuts and fill the two p_T histograms (the emulated HLT_AK8PF JETX and the reference) for the two set of events.

Then, the efficiency curve for the trigger of interest is obtained by dividing the emulated over the reference histogram. Table 5.4 contains the list of HLT paths used in this analysis, the effective luminosities and the corresponding turn-on points.
Figure 5.5 shows the efficiency curves for said triggers, with the turn-on points being taken at the point when each trigger is 99.5% efficient. In order to select events for this analysis, the phase space has been divided into independent leading jet p_T regions. In each region, only one of the triggers below is used and none of the regions overlaps with another, in order to avoid double-counting.

Table 5.4: *The HLT_AK8PFJet trigger effective luminosities and trigger turn-on points.*

HLT path	eff. lumi (fb^{-1})	Turn-on point (GeV)
HLT_AK8PFJET40	0.0000150598	114
HLT_AK8PFJET60	0.000419033	133
HLT_AK8PFJET80	0.00216941	174
HLT_AK8PFJET140	0.0471297	272
HLT_AK8PFJET200	0.202538	362
HLT_AK8PFJET260	0.465597	395
HLT_AK8PFJET320	1.24011	468
HLT_AK8PFJET400	3.71907	548
HLT_AK8PFJET450	7.38989	592
HLT_AK8PFJET500	59.8166	686

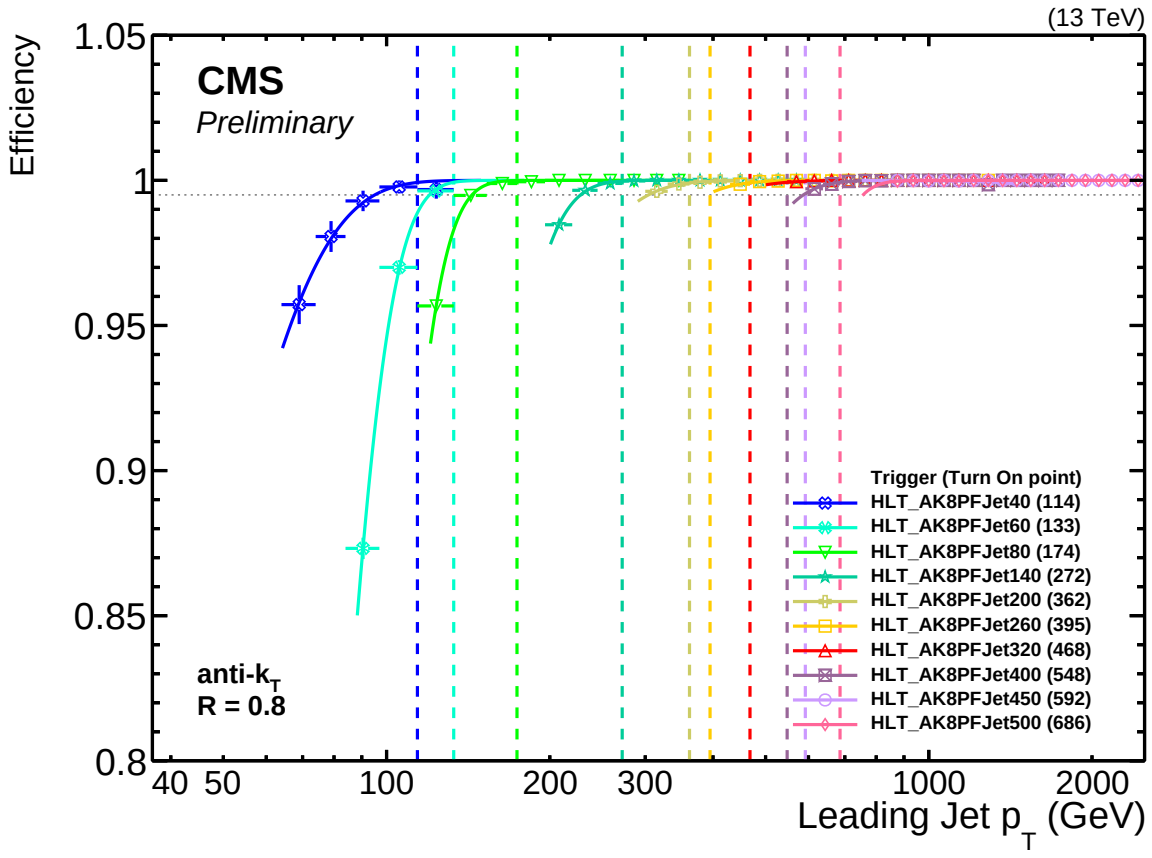


Figure 5.5: *HLT efficiency curves for 2018*

5.6 Detector level measurement

5.6.1 Inclusive differential multi-jet cross section

In this section, the inclusive differential 2-, 3-, 4- and 5-jet event cross sections are presented at detector level, as a function of $H_{T,2}/2$, a factor defined in Section 5.1. The data have been normalised with the total integrated luminosity ($\mathcal{L}_{int}$) and the $H_{T,2}/2$ bin width ($d(H_{T,2}/2)$). Along with the 2018 Data, each histogram contains some official CMS Monte Carlo samples, shown in Table 5.2 from Section 5.2. Plots of the detector level measurement are practical when examining the consistency of the data.

Comparison to Pythia8

Figure 5.6 includes the normalised inclusive differential multi-jet cross sections at detector level and the comparison of the experimental Data to Pythia8 CP5. More specifically, the inclusive 2-, 3-, 4- and 5-jet cross sections are displayed, in red, blue, green and purple respectively, beginning from top to bottom. All distributions exhibit the expected decreasing behavior with respect to $H_{T,2}/2$. As the multiplicity increases, the corresponding distribution is lower than the previous one.

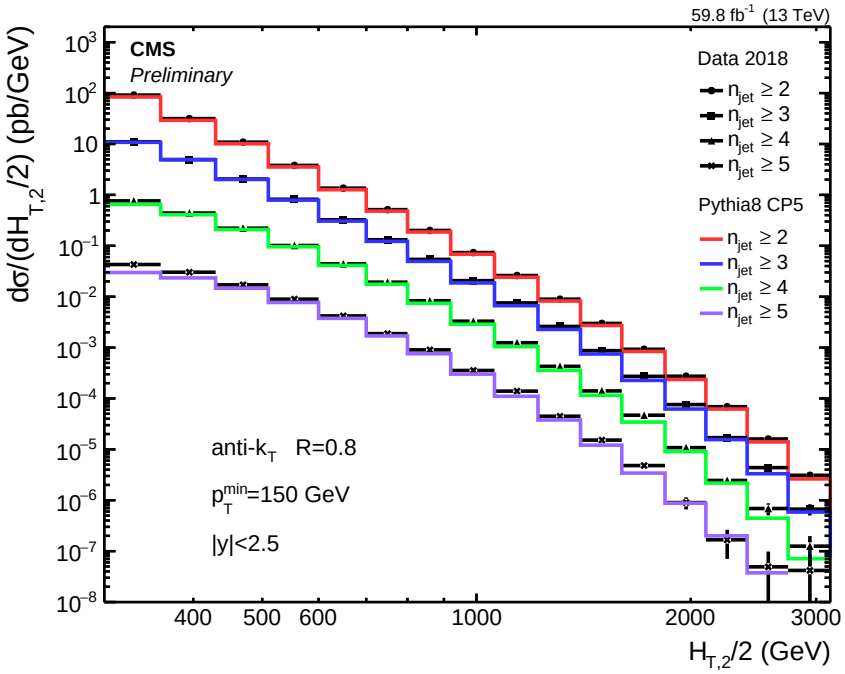


Figure 5.6: *Detector level measurement of the normalised inclusive differential multi-jet cross section in comparison with Pythia8.*

Comparison to Madgraph+Pythia8

Figure 5.7 depicts the normalised inclusive differential multi-jet cross sections at detector level of the experimental data in comparison to Madgraph interfaced with Pythia8, as a function of $H_{T,2}/2$. The measurement displays the expected declining behavior. The multiplicity takes the values $n = 2, 3, 4, 5$ and one can also observe that the smaller the multiplicity, the higher the distribution.

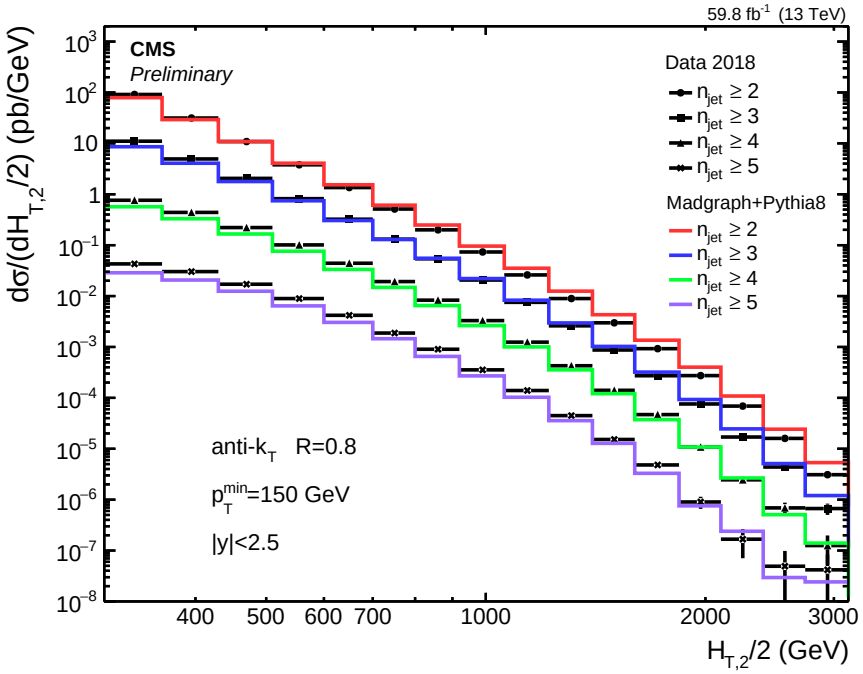


Figure 5.7: *Detector level measurement of the normalised inclusive differential multi-jet cross section in comparison with Madgraph+Pythia8.*

5.6.2 Ratios R_{mn}

Comparison to Pythia8

Figure 5.8 illustrates the detector level measurement of the R_{mn} ratios in contrast with Pythia8. The black points represent the data, whereas the red markers correspond to Pythia8. One can notice that as the order of the ratio becomes higher, the maximum value of the corresponding R_{mn} observable becomes lower. Thus, R_{32} is higher than R_{43} and they are both larger than R_{42}.

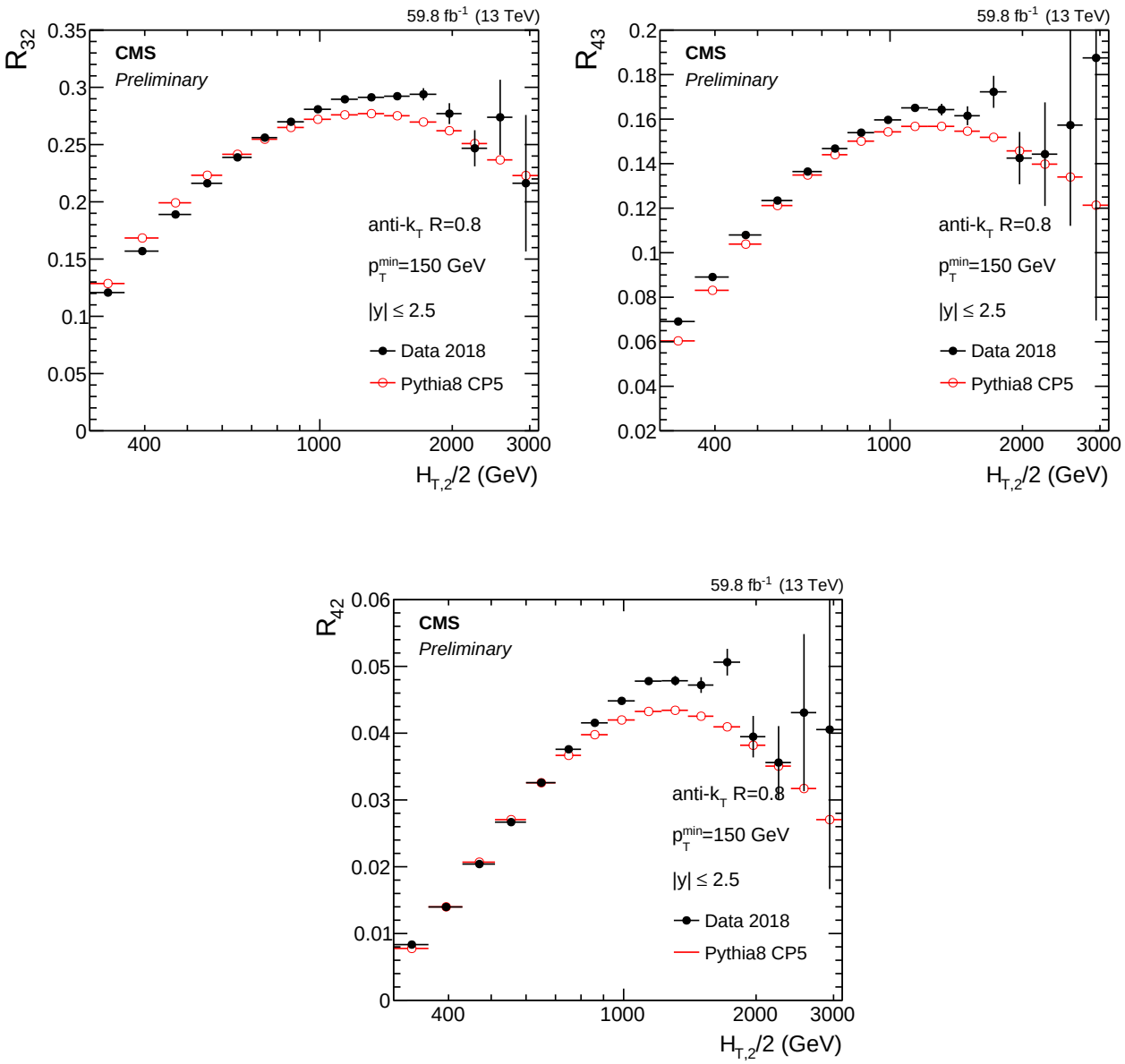


Figure 5.8: *Detector level measurement of R_{32} (top right), R_{43} (top left) and R_{42} (bottom) in contrast to Pythia8.*

Comparison to Madgraph+Pythia8

Figure 5.9 illustrates the detector level measurement of the R_{mn} ratios in contrast with Madgraph+Pythia8. The black points represent the data, whereas the red markers correspond to Madgraph+Pythia8. One can notice that as the order of the ratio becomes higher, the maximum value of the corresponding R_{mn} observable becomes lower, whereas the data comparison to Madgraph+Pythia8 is not as good as that to Pythia8.

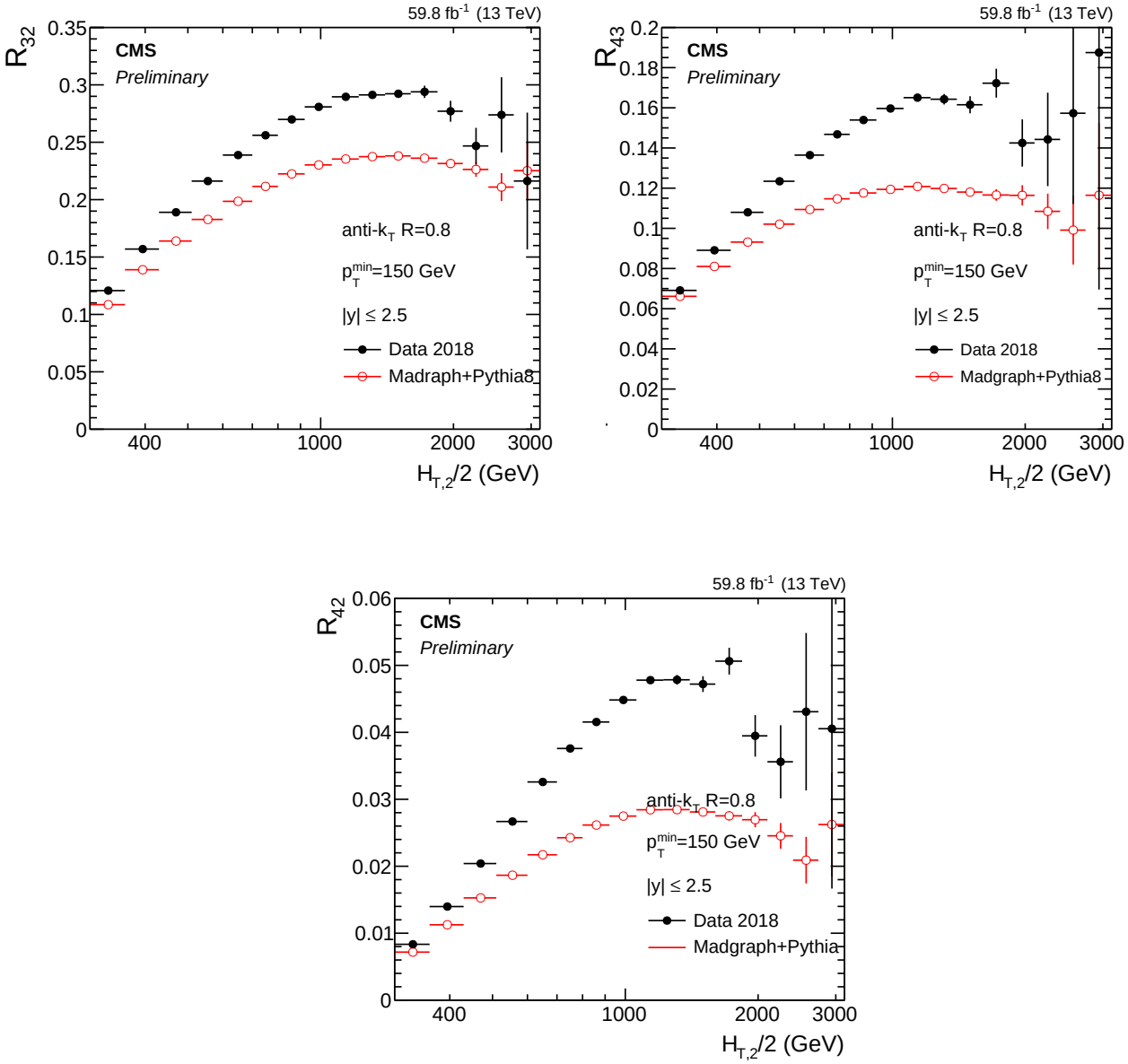


Figure 5.9: *Detector level measurement of R_{32} (top right), R_{43} (top left) and R_{42} (bottom) in comparison to Madgraph+Pythia8.*

5.7 Unfolding

5.7.1 Matrix Inversion method

In high energy physics the measurements are based on counting experiments, with the events being detected and then classified according to their properties. Most analyses, including the present one, are determined from event counts where the event properties are restricted to certain regions in phase space (bins), divided by the integrated luminosity. A set of such observations can be displayed graphically in the form of a binned histogram. The observed event counts n_i in each bin are described by a Poisson distribution with corresponding expectation value μ_i, but only as an estimation of such a distribution. The probability to observe n_i entries in bin i is:

$$P(n_i, \mu_i) = \frac{\mu_i^{n_i} e^{-\mu_i}}{\mu_i!} \tag{5.8}$$

When it comes to detector measurements, the observed event counts are different from the expectation; this derives from three main effects: the detector effects, statistical fluctuations and the background. In the first case, event properties are measured with finite resolution and limited acceptance and thus, the measured value of this property y is different from the true value x. This causes such an event to be reconstructed in the wrong bin, i.e. migrate, or get lost. In the case of statistical fluctuations, it was already mentioned that the measurement provides an estimate of the Poisson parameter μ that describes the number of events. The statistical uncertainty is assigned as the square root of the number of event counts. Finally, events similar to the signal that may have been produced by other processes yield the background and distort the signal by leading to additional contributions to a bin [23].

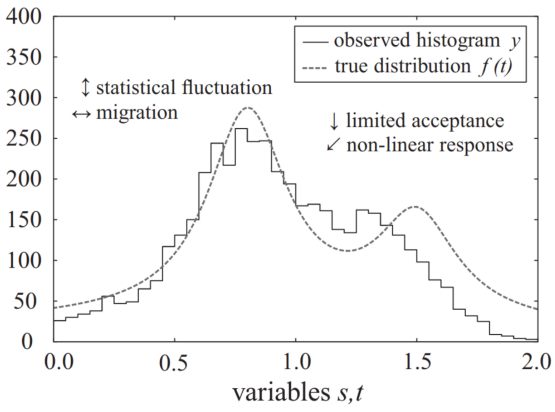


Figure 5.10: *Illustration of the deviations between true distribution $f(t)$ and measured distribution $g(s)$ caused by different sources* [21].

The impact of the effects above on the measured variable s is that their distribution $g(s)$ does not coincide with the distribution $f(t)$ of the true variable t: $g(s) \neq f(t)$. This is illustrated in Figure 5.10. The two distributions are associated by the following equation [21]:

$$\int k(s,t) f(t) dt + b(t) = g(s) \tag{5.9}$$

where $k(s,t)$ is the kernel function, describing the physical measurement process, and $b(s)$ represents the background. The above relation states that the true distribution $f(t)$ is folded with the kernel function $k(s,t)$ and can be estimated by following the unfolding procedure. Unfolding refers to the process of extracting information about the truth content of the measurement bins, given the observed measurements. For histograms with finite number of bins, the distributions $g(s)$ and $f(t)$ are discretised and replaced by the vectors $\mathbf{x}$ and $\mathbf{y}$ with dimensions (number of bins) n and m respectively, with usually $n \neq m$. Moreover, a rectangular matrix $\mathbf{A}$ called response matrix, which parameterises the detector response, takes the place of the kernel function $k(s,t)$, whereas the background distribution $b(s)$ is substituted by the vector $\mathbf{b}$. The above folding equation then takes the form:

$$\mathbf{Ax} + \mathbf{b} = \mathbf{y} \tag{5.10}$$

where $\mathbf{A}$ is the $m \times n$ response matrix that can be interpreted as a density function, $\mathbf{x}$ is the unknown true distribution that needs to be determined, the vector $\mathbf{b}$ describes the expected number of background events and $\mathbf{y}$ is the measured distribution. More specifically, the elements A_{ij} of the response matrix specify the probability to find an event produced (true value) in bin j to be measured (measured value) in bin i [24, 25]:

$$A_{ij} = P(\textit{measured in bin } i | \textit{true value in bin } j) \tag{5.11}$$

In other words, the unfolding refers to using the smeared observations to infer the true physical distribution of the events. This smearing has been caused by the detector effects and the unfolding leads to the unsmearing of the true distribution [26].
The result of the unfolding process is the estimation of the truth distribution $\mathbf{x}$ that allows the direct comparison of the measurement with either theoretical predictions, in which no detection effects are included, or results of other experiments, in which the effects of resolution will in general be different. When considering the case where the number of bins in the true and observed histograms are equal, $m = n$, then the matrix Equation 5.10 can be inverted to yield:

$$\mathbf{x} = \mathbf{A^{-1}}(\mathbf{y} - \mathbf{b}) \tag{5.12}$$

This is also the solution that minimises the χ^2:

$$\chi^2 = (\mathbf{Ax} + \mathbf{b} - \mathbf{y})^{\mathbf{T}}(\mathbf{V})^{-\mathbf{1}}(\mathbf{Ax} + \mathbf{y} - \mathbf{b}) \tag{5.13}$$

where $\mathbf{V}$ is the covariance matrix that describes the statistical covariance among the bins of the measurement. In the case of statistically independent bins the covariance matrix is diagonal [25]. The matrix inversion method can give unbiased results, however they typically suffer from large

bin-to-bin correlations, oscillation patterns and large uncertainties. For this reason, other solving strategies have been developed and used in High Energy Physics analyses, such as the D'Agostini iterative method or Tikhonov regularisation [24]. The software packages for the implementation of such unfolding methods are the ROOUNFOLD [27] and TUNFOLD [28].

5.7.2 Unfolding strategy - Response Matrix

In this analysis the TUNFOLD package is used, combined with the method of the matrix pseudo-inversion. This method relies on the solution of the matrix inversion in Equation 5.13 by using more bins in the measured than in the true (unfolded) distribution, i.e. $m > n$. More precisely, twice as many bins are used in the measured compared to the true distribution: $m = 2n$.
In order to account for the statistical correlations properly, a 2D unfolding of the inclusive differential multi-jet cross section distributions was performed and the ratios R_{mn} at particle level were then calculated from the unfolded distributions. Before the unfolding, the histograms were refilled with the exclusive 2-, 3-, 4- and inclusive 5-jet cross sections at detector level (Figure 5.11), where each event is filled only once.

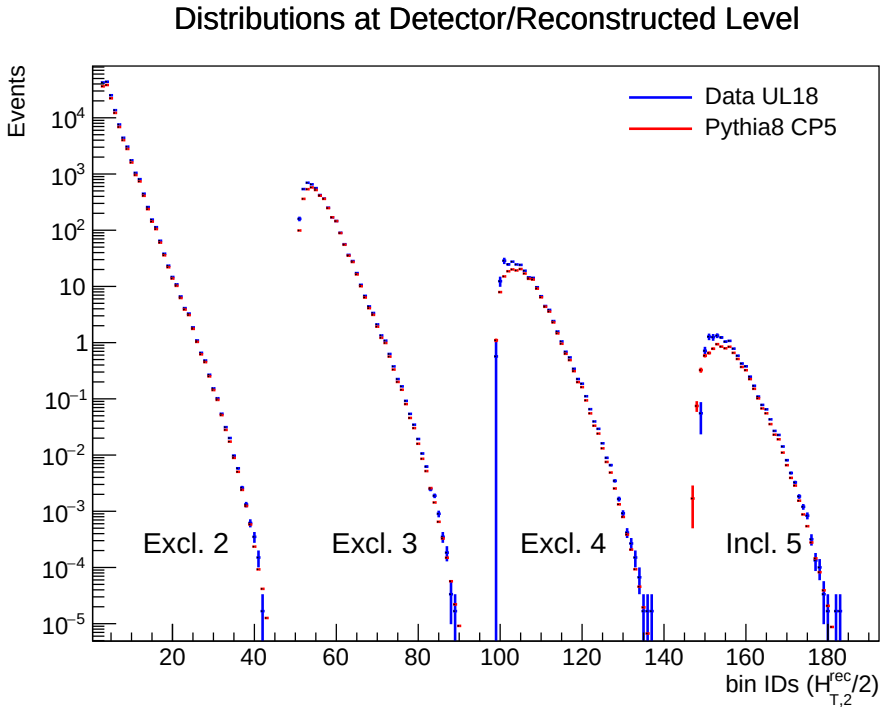


Figure 5.11: *Exclusive 2-, 3- 4- and inclusive 5-jet distribution of the 2018 Data and Pythia8 at detector and reconstructed level respectively.*

This is in contrast to the plots in Chapter 5.5, where all cross section distributions contain inclusive jets and thus, events are filled multiple times. In Figure 5.11, TUnfoldBinning has

been used and therefore, the x axis has switched from the $H_{T,2}/2$ bins to global bin IDs; as a result, it has no longer physical binning. This is done in order to treat the events with a 2D unfolding, where the treatment of correlations is simpler. The reason behind this is that it is easier to pinpoint migrations from one bin to another, but also among distributions, whereas by filling the inclusive jet distributions with different multiplicity, one would be faced with multiple migrations.

The input covariance matrix V, describes that statistical correlations among the distributions with different multiplicity and the $H_{T,2}/2$ bins. These correlations are represented by the dimensionless version of the covariance, included in the correlation matrix, and are calculated as:

$$\rho_{ij} = \frac{V_{ij}}{\sigma_i \sigma_j}, \quad V_{ij} = cov[n_i, n_j] \tag{5.14}$$

where σ_i represents the uncertainty of bin i with $V_{ii} = \sigma_i^2$. Figure 5.12 depicts the correlation matrix for the exclusive 2-, 3-, 4- and inclusive 5-jet cross section distributions, that contains the main diagonal but also off-diagonal elements, and portrays non-negligible correlations among the $H_{T,2}/2$ bins but also small correlations and anti-correlations among distributions with different multiplicity.

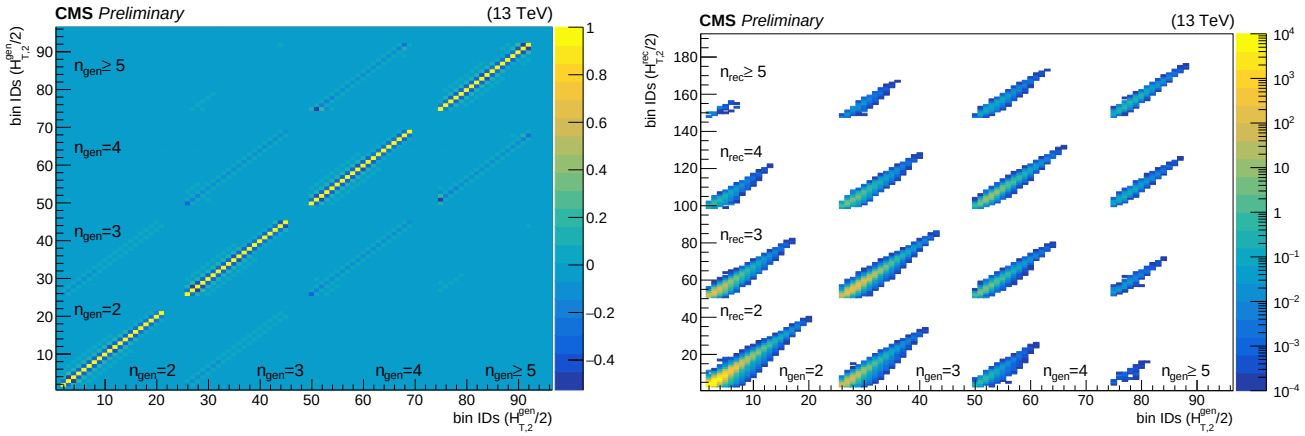


Figure 5.12: *Correlation (left) and Response (right) matrices for the exclusive 2-, 3-, 4- and exclusive 5-jet cross section.*

The Response Matrix (RM) for the inclusive multi-jet cross section distributions was built using the Official CMS Monte Carlo samples shown in Table 5.2. More specifically, the Pythia8 sample was used to extract the main unfolding results, since it has larger statistics. The RM was manufactured on an event-by-event basis by matching the jets between particle (GEN jets) and detector level (PF jets) in the MC sample. The steps of the matching are as follows:

- Loop over the particle-level jets sorted in p_T, starting from the particle-level jet with the highest p_T.
- For a given particle-level jet, define a cone around the jet axis with $R_{matching} = R_{cone}/2$. Here $R_{cone} = 0.8$.
- Try to match the particle-level jet to the highest-p_T detector-level jet with axis inside said cone. The following possible outcomes arise:
  - If the matching is successful and both jets lie within the phase space at both particle and detector level, then they are filled in the RM and are removed from the list of jets. Thus, the RM contains only migrations within the phase space.
  - If the matching is successful, and the detector-level jet is inside the phase space but the particle-level jet is outside of the phase space, then the detector-level jet is considered as fake jet.
  - If the matching is successful and the particle-level jet is within the phase space but the detector-level jet lies outside of the phase space, then the particle-level jet is considered as miss jet.
  - If the matching is unsuccessful, meaning that no detector-level jet can be found inside the cone, the particle-level jet is defined as miss jet.

Finally, after the loop on all particle-level jets has finished, the remaining unmatched detector-level jets are defined as fake jets.
Figure 5.12 illustrates the RM for the exclusive 2-, 3-, 4- and inclusive 5-jet cross sections, with particle level $H_{T,2}/2$ on y axis and detector level $H_{T,2}/2$ on x axis. The particle level spectrum of the cross section distributions of the data is constructed, seen in Figure 5.13, compared to Pythia8 Official CMS MC sample. The x axis is still in global bin IDs and the distributions are later used to extract the inclusive jet cross sections. Then, the particle-level R_{mn} ratios are obtained (Figure 5.14), all drawn on the same x axis, which is in bin IDs, along with the corresponding correlation matrix seen in (Figure 5.14). The correlation matrix includes the R_{mn} ratios at particle level on both the x and y axis; one can deduce that there are strong correlations and anti-correlations among the $H_{T,2}/2$ bins, but also among the R_{mn} ratios.
After obtaining the particle level spectrum, the results of the unfolding can be examined by comparing the experimental detector level data with their unfolded spectrum. The inclusive differential multi-jet cross sections at particle level are extracted from the unfolded distributions of the previous section (Section 5.6). For the inclusive 2-jet cross section, all the distributions were added. In the case of the inclusive 3-jet cross section, the exclusive 3-, 4- and inclusive 5-jet spectra were added, for the inclusive 4-jet cross section the exclusive 4-jet and inclusive

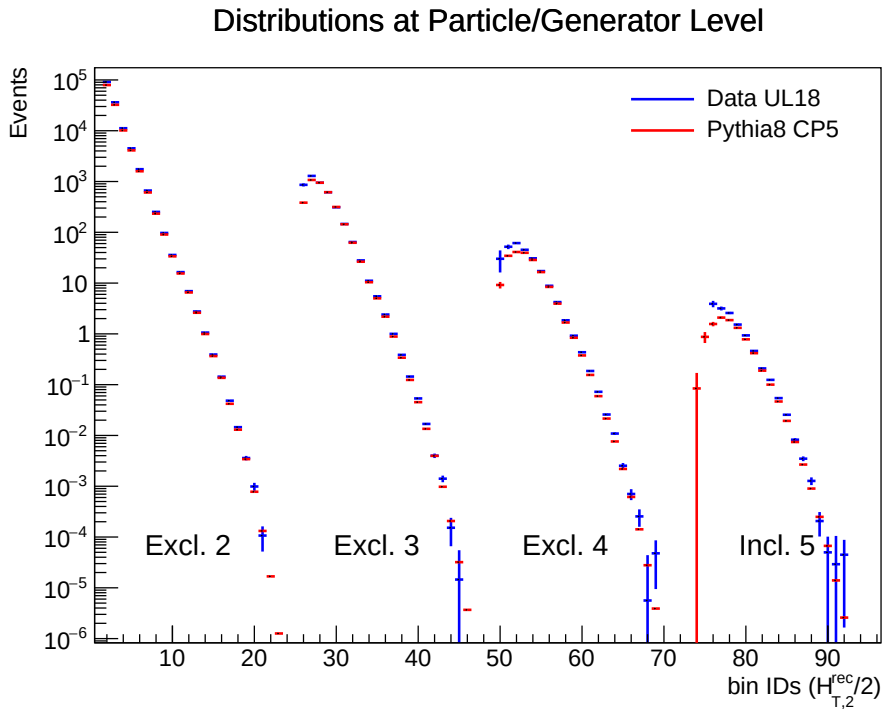


Figure 5.13: *Exclusive 2-, 3- 4- and inclusive 5-jet distribution of the 2018 Data and Pythia8 at particle and generated level respectively.*

5-jet spectra were used. Finally, the inclusive 5-jet cross section distribution at particle level remained as it was. Such a comparison is depicted in Figure 5.15, that includes the data at detector and particle level. At the bottom of each histogram, a ratio of the two distributions is drawn in order to quantify the effect of the unfolding on the measurement. A similar comparison was performed for the R_{mn} observables, seen in Figure 5.16

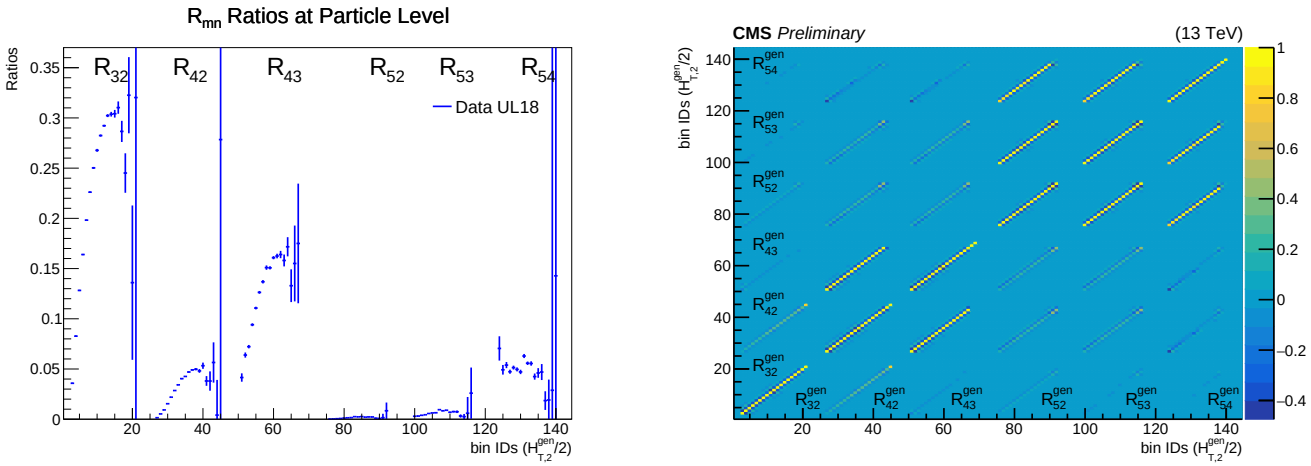


Figure 5.14: *Particle and Generated level distributions for the R_{mn} observables (left) and their correlation matrix (right).*

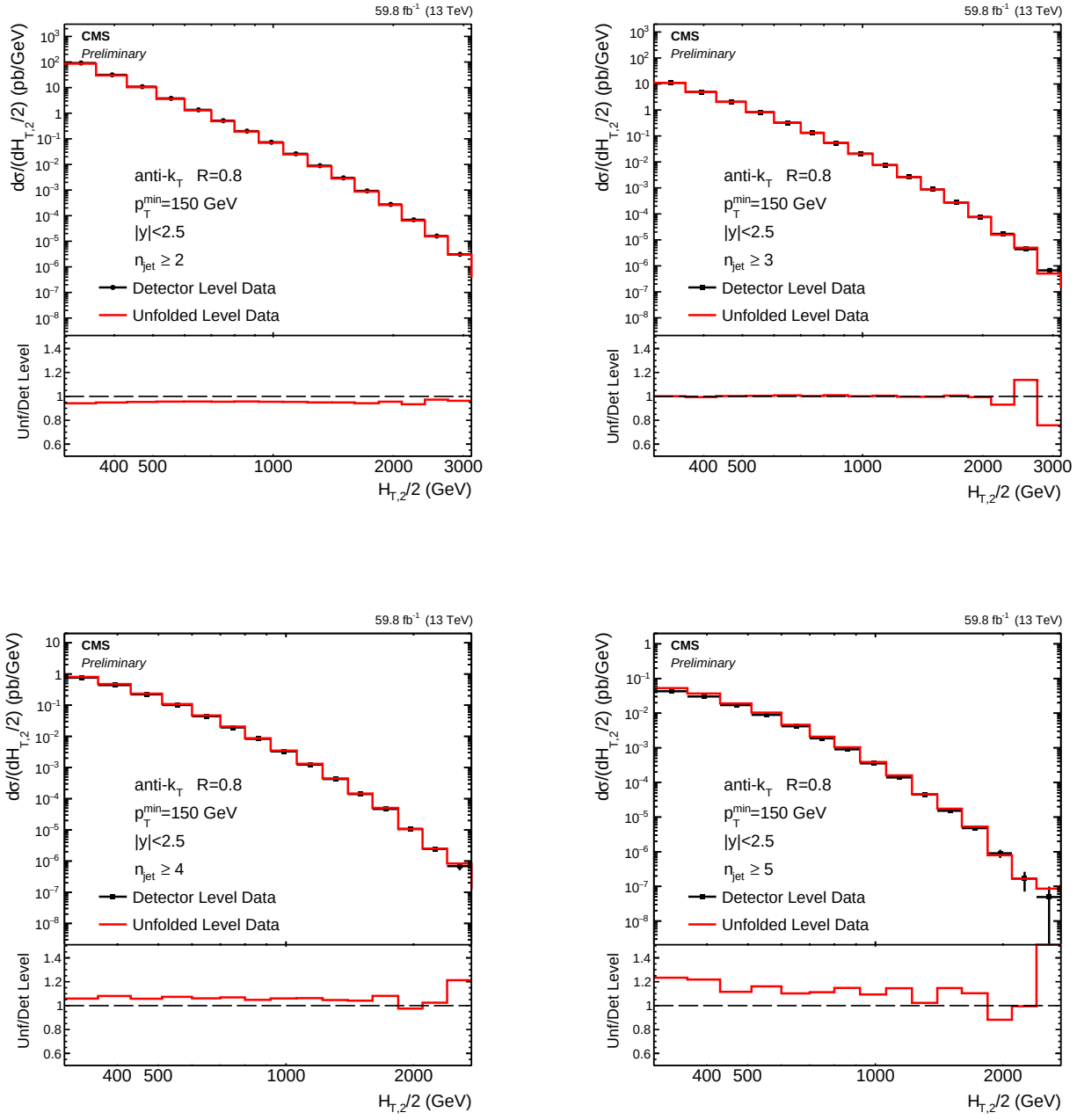


Figure 5.15: *Inclusive 2-jet (top left), 3-jet(top right), 4-jet (bottom left) and 5-jet (bottom right) cross section distributions at detector level in black points and unfolded (particle level) in red line, along with the unfolded over reconstructed ratio.*

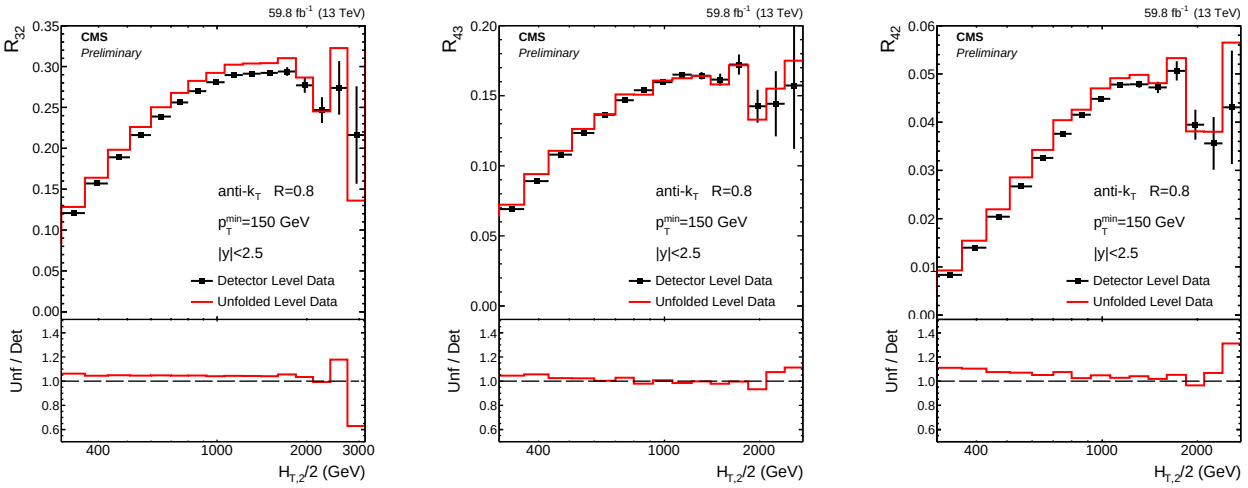


Figure 5.16: *The R_{32} (top left), R_{43}(top right), and R_{42}(bottom) observables at detector level in black points and unfolded (particle level) in red line, along with the unfolded over reconstructed ratios.*

5.8 Systematic Uncertainties

Most measurements of physical quantities in high energy physics and astrophysics involve both a statistical uncertainty and an additional "systematic" uncertainty. Systematic uncertainties arise from uncertainties associated with the nature of the measurement apparatus, assumptions made by the experiment, or the model used to make inferences based on the observed data, etc [29].

Jet Energy Scale (JES) uncertainties

The Jet Energy Calibration procedure introduces many different systematic uncertainty sources. The sensitivity of the measurement to the JEC uncertainty is investigated by varying the jets transverse momenta as:

$$p_T = p_T(1 \pm uncert.source) \tag{5.15}$$

where the uncertainty sources here are provided from the JetMET group together with the Jet Energy Corrections. In order to estimate the total Jet Energy Scale uncertainty for the R_{mn} observables at particle level, the variations of Equation 5.15 are performed at the detector level data and then the unfolding process is repeated for each individual variation. Finally, the differences between the unfolded spectra obtained from the variations and the nominal unfolded spectrum are added in quadrature. The total (relative) JES uncertainty is shown with the violet band in Figures 5.17, 5.18, 5.19.

Jet Energy Resolution (JER) uncertainties

The Jet Energy Resolution (JER) smearing plays a crucial role in the unfolding procedure. The Scale Factors (SFs) reflect the uncertainties in the estimation of such factors and are propagated to the R_{mn} measurement. Hence, for the calculation of the JER uncertainty in the R_{mn} measurement, the smearing procedure is repeated, considering 1σ variation of the SFs. Then the unfolding process is repeated and finally, the differences between the unfolded spectra obtained from the variations and the nominal unfolded spectrum are assigned as JER uncertainty. The total JER uncertainty for the R_{mn} measurement is shown with orange band in Figures 5.17, 5.18, 5.19.

Uncertainties from miss and fake jets

The miss and fake jets derive from detector inefficiencies and background contributions respectively. For the determination of their impact on the unfolded spectrum of the R_{mn} ratios, a 5% normalisation uncertainty was assigned to each miss and fake jet during the unfolding. They are illustrated with the pink band in Figures 5.17, 5.18, 5.19.

Pile Up (PU) reweighting uncertainties

The impact of the PU profile reweighting procedure applied to the MC samples is investigated by considering up and down variations of the reweight factors. Just like in the previous case, the unfolding is repeated for these variations, with the uncertainties being calculated at particle level. In Figures 5.17, 5.18, 5.19 are shown with the yellow band.
The systematic uncertainties in the measurement of the R_{mn} observables that arise from the sources mentioned above are shown in Figures 5.17 5.18 and 5.19, that also include the statistical uncertainty drawn with vertical lines. The total experimental uncertainty for the R_{mn} measurement is calculated from the quadratic sum of systematic and statistical uncertainties. In all cases, the dominant systematic uncertainty is JES, whereas the JER and PU uncertainties have a minor contribution in the total systematic uncertainties.

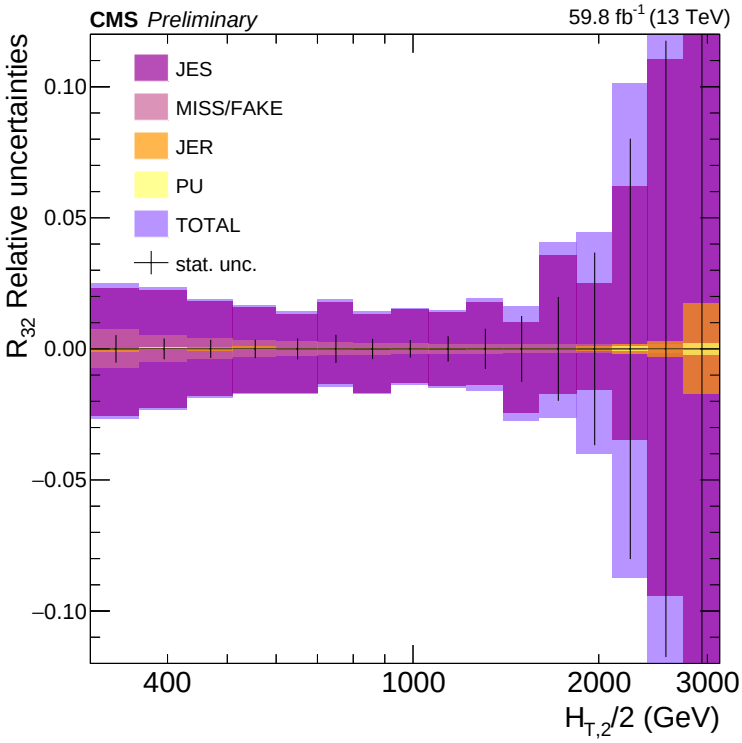


Figure 5.17: *Experimental uncertainties for the R_{32} observable.*

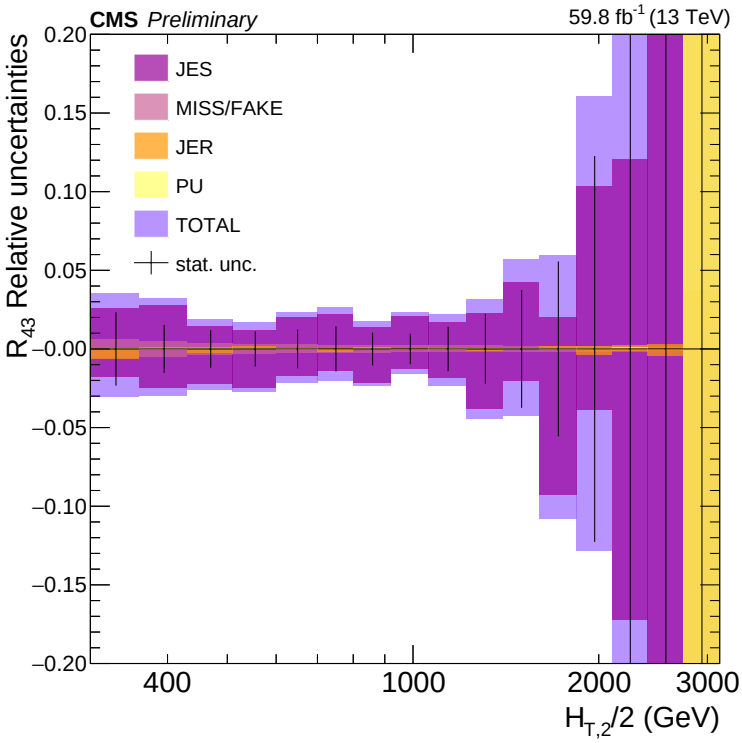


Figure 5.18: *Experimental uncertainties for the R_{43} observable.*

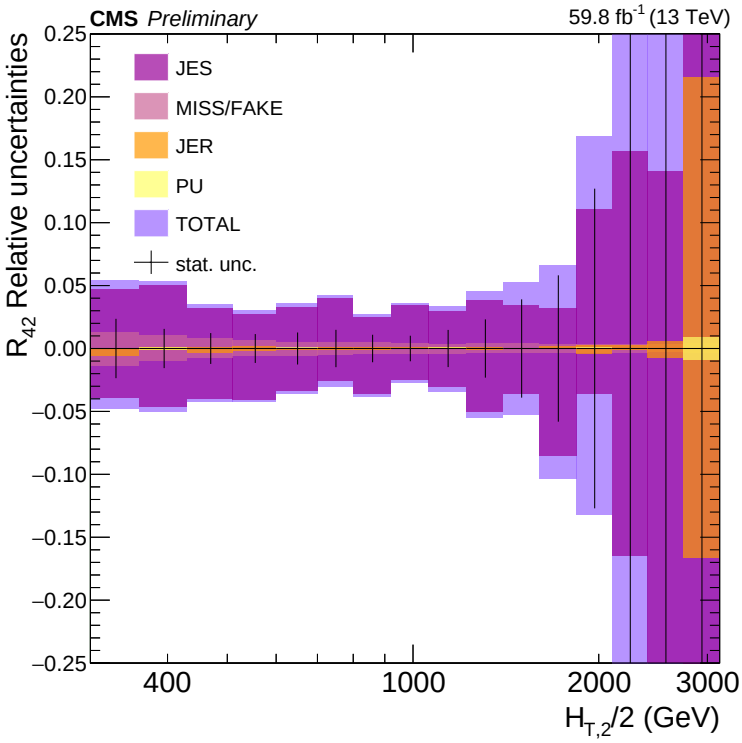


Figure 5.19: *Experimental uncertainties for the R_{42} observable.*

Bibliography

[1] The CMS Collaboration. Determination of the strong coupling constant from the measurement of inclusive multijet event cross sections in pp collisions at $\sqrt{s} = 8$ TeV. 2016. https://inspirehep.net/literature/1514769.

[2] S. Chatrchyan and V. et al. [The CMS Collaboration] Khachatryan. Measurement of the ratio of the inclusive 3-jet cross section to the inclusive 2-jet cross section in pp collisions at $\sqrt{s}$=7 TeV and first determination of the strong coupling constant in the TeV range. *Eur. Phys. J. C*, 73, 2604, 2013.

[3] CMS public physics results combined. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined.

[4] V. Khachatryan et al. [CMS Collaboration]. Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s}$=8 TeV and cross section ratios to 2.76 and 7 TeV. *JHEP*, 03:156, 2017. https://doi.org/10.1007/JHEP03(2017)156.

[5] The Worldwide LHC Computing Grid. https://wlcg.web.cern.ch.

[6] G. L. Bayatyan et al. [CMS Collaboration]. CMS computing: Technical Design Report. Technical design report. 2005. https://cds.cern.ch/record/838359.

[7] D. Piparo (CERN) for O&C and PPD. Offline Data Preparation, Processing Software and Computing, 2021. Offline Data Preparation, Processing Software and Computing.

[8] G.L. Bayatyan, Michel Della Negra, A. Foà Hervé, and Achille Petrilli. CMS computing: Technical Design Report. 2005. https://cds.cern.ch/record/838359.

[9] G. Petrucciani, A. Rizzi, and C. Vuosalo [CMS Collaboration]. Mini-AOD: A New Analysis Data Format for CMS. *J. Phys.: Conf. Ser.*, 664 072052, 2015. https://doi.org/10.1088/1742-6596/664/7/072052.

[10] Andrea Rizzi, Giovanni Petrucciani, and Marco Peruzzi [CMS Collaboration]. A further reduction in CMS event data for analysis: the NANOAOD format. *EPJ Web Conf.*, 2014, 06021, 2019. https://doi.org/10.1051/epjconf/201921406021.

[11] Giovanni Franzoni. Dataset definition for CMS operations and physics analyses. *Nuclear and Particle Physics Proceedings*, 273-275, 2016. https://doi.org/10.1016/j.nuclphysbps.2015.09.144.

[12] Paraskevas Gianneios. *Studies of quantum chromodynamics with jets at the CMS experiment at the LHC*. PhD Thesis, 2022.

[13] Monte Carlo production tools. https://cms-pdmv.gitbook.io/project/.

[14] ROOT data analysis framework. https://root.cern/.

[15] CMS Software. https://github.com/cms-sw/cmssw.

[16] Standard Model Physics – Jet analysis framework. https://github.com/cms-smpj/SMPJ.

[17] DAS framework. https://gitlab.cern.ch/DasAnalysisSystem.

[18] Cross section database. https://cms-gen-dev.cern.ch/xsdb.

[19] Utilities for Accessing Pileup Information for Data. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData(CMSInternalTwikipage).

[20] Jet Energy Resolution. JetEnergyResolution(CMSInternalTwikipage).

[21] Olaf Behnke, Kevin Kroninger, Thomas Schorner-Sadenius, and Gregory Schott. *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*. Wiley-VCH, Weinheim, Germany, 2013.

[22] G. Flouris. *Studies of quantum chromodynamics and measurement of the strong coupling constant with jets from the CMS experiment at the LHC*. PhD Thesis, 2016.

[23] Volker Blobel. Unfolding Methods in Particle Physics. 2011. https://doi.org/10.5170/CERN-2011-006.240.

[24] Stefan Schmitt. Data Unfolding Methods in High Energy Physics. *EPJ Web of Conferences*, 137, 11008, 2017. https://doi.org/10.1051/epjconf/201713711008.

[25] Glen Cowan. *Statistical Data Analysis*. Oxford University Press, 1998.

[26] Mikael Kuusela. *Statistical Issues in Unfolding Methods for High Energy Physics*. MSc Thesis, 2012.

[27] Tim Adye. Unfolding algorithms and tests using RooUnfold. *PHYSTAT 2011*, 313-318, 2011. https://doi.org/10.5170/CERN-2011-006.313.

[28] S. Schmitt. TUnfold, an algorithm for correcting migration effects in high energy physics. *JINST*, 7 T10003, 2012. https://doi.org/10.1088/1748-0221/7/10/T10003.

[29] Pekka Sinervo. Definition and Treatment of Systematic Uncertainties in High Energy Physics and Astrophysics. *PHYSTAT2003*, C030908, 2003. TUAT004.

Chapter 6

Data – MC comparison at particle level

In this section, the inclusive differential 2-, 3-, 4- and 5-jet event cross sections, as well as the R_{mn} observables, are presented at particle level, as a function of $H_{T,2}/2$. For this purpose, the 2018 Data were used, along with the Official CMS Monte Carlo samples and also Pythia8 CUETM1, CUETM2 and Herwig++, generated using the Rivet toolkit. The difference between an official CMS MC sample and one generated by Rivet, is the fact that the former provides information at both reconstructed and generated level, whereas the latter contains information only at generated level. Below every distribution, a ratio of the MC predictions and the Data is drawn, in order to compare the description of the particle level data among the MC samples. More specifically, the Data are drawn in black markers, Pythia8 CP5 in red line, Pythia8 CUETM1 and CUETM2 are illustrated in green and purple line respectively, Madgraph+Pythia8 is depicted in fuchsia and lastly, Herwigg++ is drawn with blue line. Each MC/Data ratio also includes the statistical uncertainties, derived by the data. Statistical uncertainties are the result of stochastic fluctuations arising from the fact that a measurement is based on a finite set of observations. The plots that correspond to the R_{mn} observable, also include the experimental uncertainty band obtained from the quadratic sum of systematic and statistical uncertainties.

6.1 Inclusive differential multi-jet cross section

Figure 6.1 depicts the normalised inclusive differential 2-jet cross section at particle level, with the ratio plot below quantifying the comparison between the experimental Data and MC samples, as a function of $H_{T,2}/2$. The measurement displays the expected declining behavior and it is best described by Pythia8 CUETM2 and CP5.

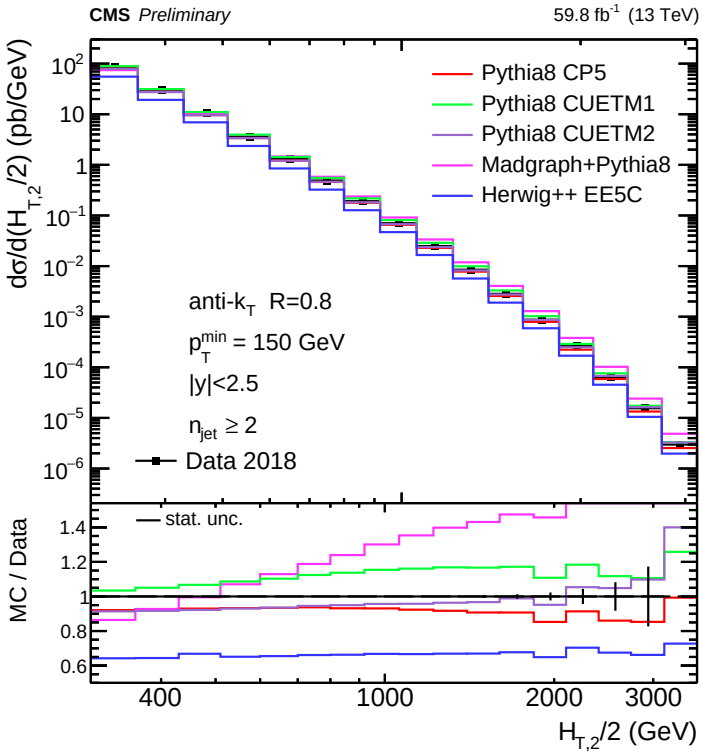


Figure 6.1: *Comparison of the inclusive differential 2-jet cross section between data and Leading Order Monte Carlo predictions at particle level.*

Figure 6.2 portrays the normalised inclusive differential 3-jet cross section at particle level and the comparison of the experimental results to the MC event generators, both with respect to $H_{T,2}/2$. The distribution is lower than the previous and Pythia8 CUETM2 and CP5 provide the best description of the measurement.

Figure 6.3 portrays the normalised inclusive differential 4-jet cross section at particle level along with a quantification of the agreement between the experimental results and the MC event generators, both with respect to $H_{T,2}/2$. The distribution is lower than the inclusive 3-jet cross section and Pythia8 CP5 and CUETM2 provide the best description of the measurement.

Figure 6.4 portrays the normalised inclusive differential 5-jet cross section at particle level along with a quantification of the agreement between the experimental results and the MC event generators, both with respect to $H_{T,2}/2$. The distribution is lower than all the previous inclusive jet cross sections and Pythia8 CP5 along with CUETM2 provide the best description of the measurement. It is worth noting that as the distributions move to higher multiplicity, the agreement of the MC event generators with the data worsens and thus, they fail to describe the measurement with great accuracy.

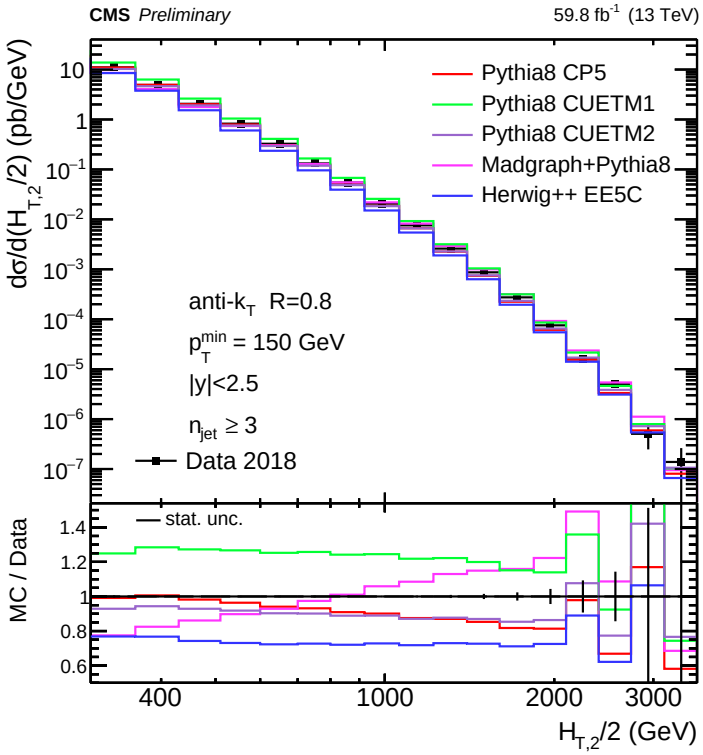


Figure 6.2: *Comparison of the inclusive differential 3-jet cross section between data and Leading Order Monte Carlo predictions at particle level.*

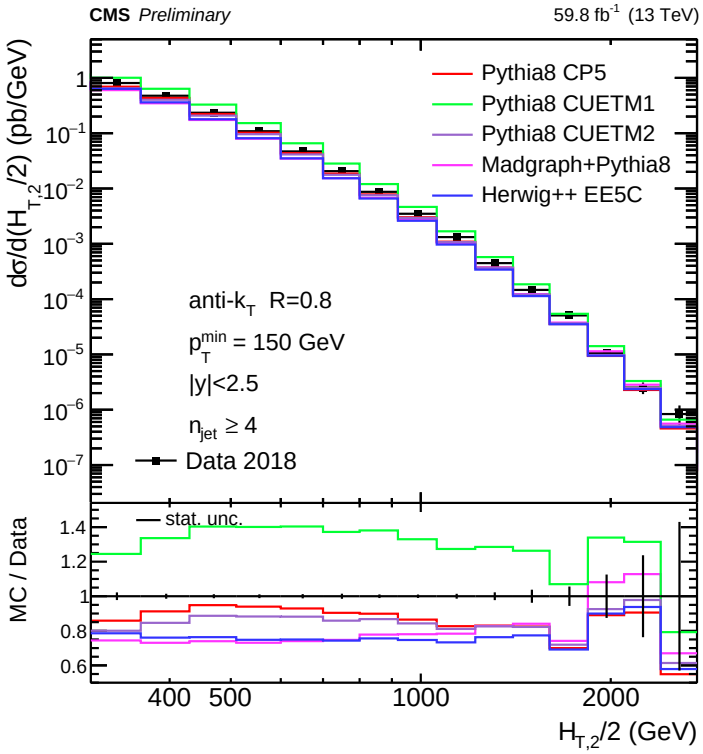


Figure 6.3: *Comparison of the inclusive differential 4-jet cross section between data and Leading Order Monte Carlo predictions at particle level.*

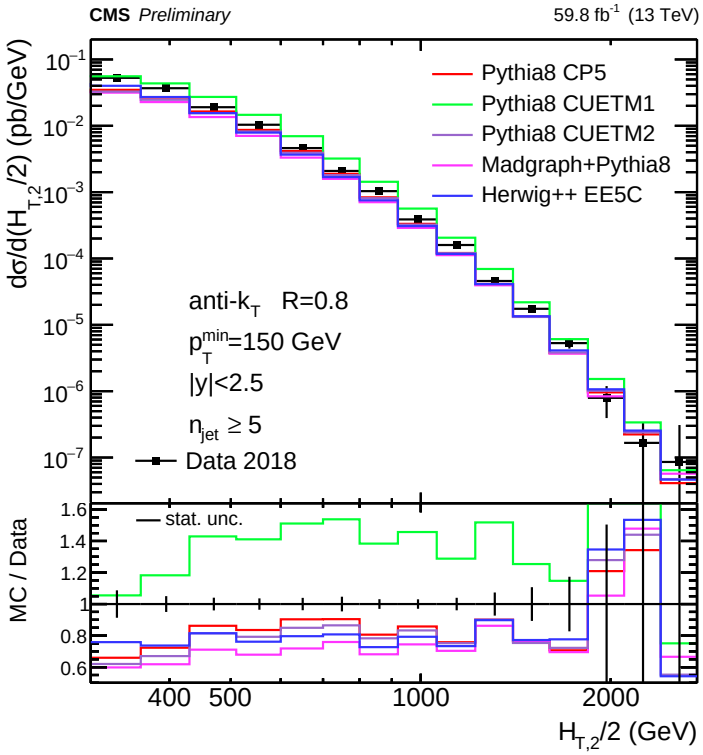


Figure 6.4: *Comparison of the inclusive differential 5-jet cross section between data and Leading Order Monte Carlo predictions at particle level.*

6.2 Ratios R_{mn}

Figure 6.5 represents the R_{mn} observables at particle level in comparison with the MC event generators used in the case of the inclusive multijet cross section distributions, as a function of $H_{T,2}/2$. All R_{mn} observables are best described by Pythia8 CUETM2 (5% for R_{32} and 5-10% for R_{43} and R_{42}) and CP5 ($5-10\%$ for R_{32} and R_{43} and 5% for R_{42}). Pythia8 CP5 and CUETM2 exhibit a similar behavior when it comes to the description of the measurement. The deviation of Pythia8 CUETM2 from CP5 derives from the different α_S values that are used in the parton shower. As it will be discussed in the next chapter, the R_{32} measurement shows nice sensitivity to the strong coupling constant and for this reason, the MC predictions are affected by the parameter used in tuning.

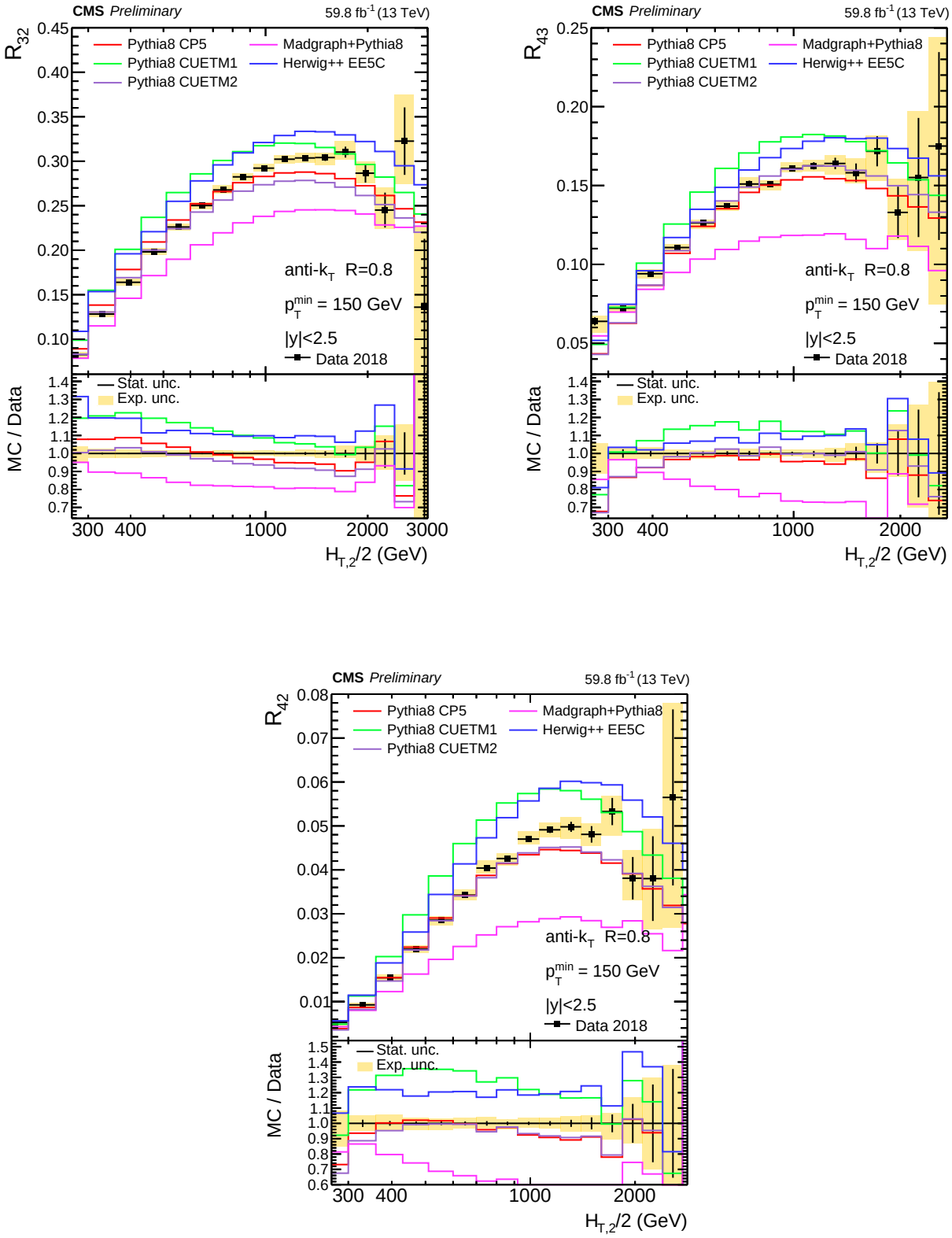


Figure 6.5: *Comparison of the R_{32} (top left), R_{43} (top right) and R_{42} (bottom) observable between data and Leading Order Monte Carlo predictions at particle level.*

Chapter 7

Fixed Order QCD for jet production

One of the important theoretical tools in every analysis of hadronic final states is perturbative quantum chromodynamics (pQCD). This practically means extracting more precise QCD theoretical calculations through higher computations (at least) at the next-to-leading order (NLO) accuracy. These NLO corrections are made of one loop virtual corrections and one gluon/quark real emission. NLO calculations in QCD are used to compute infrared- (IR-) and collinear-safe quantities at the one-loop level. They considerably reduce the uncertainties of theoretical predictions, and experience from e^+e^-, ep, and hadron colliders has shown that they lead to remarkable agreement between theory and data. Jets that have been experimentally measured and theoretically defined must be infrared and collinear safe; therefore, modifying an event by a collinear splitting or adding a soft emission, the set of hard jets that are found in the event, and hence the value of any observable, should remain unchanged. This can be attained by the IRC safe anti-k$_T$ algorithm [1, 2].

In most cases, phenomenological analyses of experimental data use pQCD predictions in fixed order in α_S. The pQCD predictions for a ratio R are then typically computed from the ratio of the corresponding fixed-order predictions for the cross sections in the numerator and the denominator. For consistency, they are both computed at the same order in α_S, i.e. both at leading order (LO), next-to-leading order (NLO) or next-to-next-to leading order (NNLO) in α_S [3].

In the present analysis, the final states of interest comprise of quarks and gluons that ultimately form the jets. More specifically, inclusive n-jet topologies are studied, with n being the minimum number of jets generated in the final state. In the inclusive jet production (or cross section) the measurement contains all present jets, so a n-jet includes n contributions to an inclusive jet observable. The simplest case concerns two incoming partons producing two outgoing partons, i.e. $2 \rightarrow 2$ processes, that forms two jets (dijet production) moving in opposite direction (back-to-back) with zero net transverse momentum. Dijet and three-jet events are widely used in

numerous CMS analyses that require QCD predictions and can be calculated at NLO accuracy by the NLOJET++ program. As for the NLO QCD calculations for up to five jets production at the LHC, these are feasible by combining the SHERPA event generator with other packages that provide one-loop amplitudes, such as NJET. On the other hand, NNLO QCD calculations for dijet production are made through the NNLOJET package, whereas some first steps towards the three-jet production have been recently made.
For the current analysis, fixed order predictions were made for the dijet and three-jet production, based on the FASTNLO framework that is implemented in the NLOJET++ package and enables the fast extraction of the theoretical predictions for various choices of renormalisation (μ_r) and factorisation (μ_f) scales and PDF sets, that are arbitrary. Most commonly, these two scales are set to the same central value, $\mu_r = \mu_f = \mu$ which should be of the order of the "hard scale". The case investigated here is $\mu = H_{T,2}/2$, which is defined in Section 5.1 as the transverse momenta of the two leading jets.
The PDF set used for the fixed-order NLO predictions is shown in Table 7.1. Each PDF set is provided for a wide range of $\alpha_S(M_Z)$, along with the central value.

Table 7.1: *PDF set used in the theory calculations.*

PDF set	**Central value** $\alpha_S(M_Z)$	α_S **range**
NNPDF31_NLO	0.118	0.106 - 0.130

7.1 Next-to-leading order calculations

In order to calculate cross section at next-to-leading order (NLO) accuracy, the Equation 1.77 that describes the hadron-hadron scattering to a n-parton final state that includes the leading order (LO) cross section $\hat{\sigma}$ is customised so that it includes higher orders. Thus, the general structure of a QCD cross section in NLO accuracy is:

$$\sigma = \sigma^{LO} + \sigma^{NLO} \tag{7.1}$$

where the LO cross section σ^{LO} is obtained by integrating the fully exclusive cross section $d\sigma^B$ in the Born approximation over the phase space for the corresponding jet quantity. Supposing that an LO calculation involves m partons with momenta $p_k (k = 1, ..., m)$ in the final state, the LO part, also known as Born Level $\mathcal{B}$ is:

$$\sigma^{LO} = \int_n d\sigma^{\mathcal{B}}, \quad d\sigma^{\mathcal{B}} = d\Phi_m\left(p_k\right) |\mathcal{M}_m(p_k)|^2 F_J^{(m)}(p_k) \tag{7.2}$$

where $d\Phi_m$ and $|\mathcal{M}_m|$ denote the full phase space and the tree-level QCD matrix element to produce m final-state partons respectively; these factors depend on the process. The function $F_J^{(m)}$ defines the physical quantity to be computed, possibly including the experimental cuts. Note that this quantity has to be a jet observable, that is, it has to be infrared and collinear safe: its actual value has to be independent of the number of soft and collinear particles in the final state. Thus, $F_J^{(m+1)} = F_J^{(m)}$ in any case where the $m+1$-parton configuration on the left-hand side is obtained from the m-parton configuration on the right-hand side by adding a soft parton or replacing a parton with a pair of collinear partons carrying the same total momentum.
The NLO part includes contributions from real ($\mathcal{R}$) and virtual ($\mathcal{V}$) corrections. At NLO one has to consider the exclusive cross section $d\sigma^{\mathcal{R}}$ with $m+1$ partons in the final state and the one-loop correction $d\sigma^{\mathcal{V}}$ to the process with m partons in the final state:

$$\sigma^{NLO} = \int_n d\sigma^{NLO} = \int_{m+1} d\sigma^{\mathcal{R}} + \int_m d\sigma^{\mathcal{V}} \tag{7.3}$$

Therefore, the structure of a NLO calculation for the total cross section can be written as:

$$\sigma^{@NLO} = \int d\Phi_{\mathcal{B}} \left[\mathcal{B}_m(\Phi_{\mathcal{B}}; \mu_f, \mu_r) + \mathcal{V}_m(\Phi_{\mathcal{B}}; \mu_f, \mu_r)\right] + \int d\Phi_{\mathcal{R}} \mathcal{R}_m(\Phi_{\mathcal{R}}; \mu_f, \mu_r) \tag{7.4}$$

where each individual part is given by:

$$\mathcal{B}_m(\Phi_{\mathcal{B}}; \mu_f, \mu_r) = \sum_h |\mathcal{M}_m^{(b)}(\Phi_{\mathcal{B}}; h; \mu_f, \mu_r)|^2, \tag{7.5}$$

$$\mathcal{V}_m(\Phi_{\mathcal{B}}; \mu_f, \mu_r) = 2 \sum_h Re\left[\mathcal{M}_m^{(b)}(\Phi_{\mathcal{B}}; h; \mu_f, \mu_r) \mathcal{M}_m^{*(b+1)}(\Phi_{\mathcal{B}}; h; \mu_f, \mu_r)\right], \tag{7.6}$$

$$\mathcal{R}_m(\Phi_{\mathcal{R}}; \mu_f, \mu_r) = 2 \sum_h |\mathcal{M}_{m+1}^{(b+1)}(\Phi_{\mathcal{R}}; h; \mu_f, \mu_r)|^2 \tag{7.7}$$

where b stands for the Born level contribution and $M^{(b)}$ indicates the order of the matrix element. The above calculation of the cross section at NLO accuracy gives rise to ultraviolet and infrared divergences. The former can be handled by carrying out the renormalisation procedure. Thus, one can assume that the virtual cross section is given in terms of the renormalised matrix element and the ultraviolet divergences have been removed. On the other hand, the infrared divergences emerge not solely from the virtual contribution but they are also accompanied by analogous singularities arising from the integration of the real cross section. These singularities originate from the soft ($\epsilon \to 0$) and collinear (parallel to another particle) emissions either in the loop (virtual) or in the additional particle radiated (real). According to the Bloch-Nordsieck (BN) and Kinoshita-Lee-Nauenberg (KLN) theorems these divergences must cancel each other for physically meaningful i.e., infrared safe observables. The two integrals on the right-hand side of Equations 7.3 and 7.4 are separately divergent so that, before any numerical calculation can

be attempted, the separate pieces have to be regularised [4]. In practice, the NLO calculations are currently based on infrared subtraction algorithms such as the Catani-Seymour or dipole substraction method [5]. A similar method is applied on the NLOJET++ calculations on which this analysis is based. The cross section in Equation 7.4 can be rewritten as:

$$\sigma^{@NLO} = \int d\Phi_{\mathcal{B}} \left[\mathcal{B}_m(\Phi_{\mathcal{B}};\mu_f,\mu_r) + \mathcal{V}_m(\Phi_{\mathcal{B}};\mu_f,\mu_r) + \mathcal{I}_m^{(s)}(\Phi_{\mathcal{B}};\mu_f,\mu_r)\right] \tag{7.8}$$

$$+ \int d\Phi_{\mathcal{R}} \left[\mathcal{R}_m(\Phi_{\mathcal{R}};\mu_f,\mu_r) - \mathcal{S}_m(\Phi_{\mathcal{R}};\mu_f,\mu_r)\right] \tag{7.9}$$

where $d\Phi_{\mathcal{B}}$ and $d\Phi_{\mathcal{R}}$ denote Born and real-emission phase space, respectively and $\mathcal{B}, \mathcal{V}, \mathcal{I}, \mathcal{R}$ and $\mathcal{S}$ are the matrix elements for the Born, virtual, integrated subtraction, real emission and real subtraction contribution. The real subtraction term $\mathcal{S}_m$ and and the integrated subtraction term $\mathcal{I}_m^{(S)}$ cancel each other out [6]:

$$0 = \int d\Phi_{\mathcal{B}} \mathcal{I}_m^{(s)}(\Phi_{\mathcal{B}};\mu_f,\mu_r) - \int d\Phi_{\mathcal{R}} \mathcal{S}_m(\Phi_{\mathcal{R}};\mu_f,\mu_r) \tag{7.10}$$

For the R_{32} observable the fixed order NLO predictions were obtained from the NLOJET++ package using the FASTNLO framework. In practice, the calculations were performed separately for numerator's and for the denominator's cross sections of R_{32}, i.e. for the inclusive 2- and 3-jet cross sections. Hence, for the PDF set included in Table 7.1, the parton-level fixed-order predictions for the two cross sections and subsequently for the R_{32} observable were obtained.

7.2 Non-perturbative corrections

The fixed-order theoretical predictions described in the previous section are made only at parton level. For this reason, corrections need to be applied for the non-perturbative (NP) effects of multiple-parton interactions (MPI) and hadronisation (HAD), so that the predictions can be comparable to the experimental data at particle level.
Such non-perturbative effects are estimated using MC event generators, with the corresponding NP corrections being computed as the ratio of the nominal event generation with fully hadronised events, over a sample with MPI and hadronisation switched off. Ultimately, the NP corrections are factors by which the theoretical predictions are multiplied. These factors are defined as said ratios:

$$C_{(LO)}^{NP} = \frac{N_{(LO)}^{PS+HAD+MPI}}{N_{(LO)}^{PS}} \qquad C_{(NLO)}^{NP} = \frac{N_{(NLO)}^{PS+HAD+MPI}}{N_{(NLO)}^{PS}} \tag{7.11}$$

In order to avoid statistical fluctuations that may arise in phase space regions that are less populated, the NP correction factors are parameterised by a simple polynomial function:

$$y = \alpha + b \cdot x^c \tag{7.12}$$

The MC event generators used to calculate such NP corrections are the ones used for the Data-MC comparison at particle level, presented in Chapter 6 (not the Official MC samples) and the respective MC predictions were obtained by the RIVET toolkit. The NP correction factors for the inclusive multijet cross sections are illustrated in Figure 7.1, where Pythia8 CUETM1 is drawn with blue line, Pythia8 CUETM2 with green and Herwig++ is depicted with red line. The NP correction factors exhibit the expected behavior, starting from higher values and then diverging to 1.

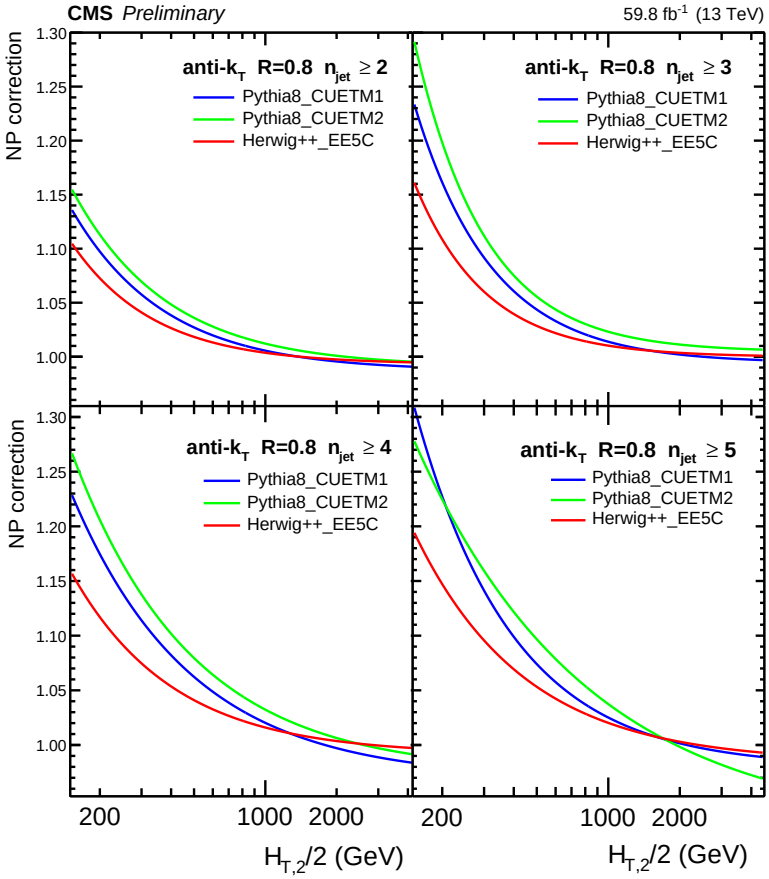


Figure 7.1: *The non-perturbative corrections for the inclusive differential multi-jet cross sections derived from different Monte Carlos and tunes.*

7.3 Sensitivity of R_{32} to α_S

The sensitivity of the numerator and denominator of the R_{32} ratio to the strong coupling constant, for $\mu_r = \mu_f = H_{T,2}/2$, is depicted in Figure 7.2. The PDF chosen for this purpose is the NNPDF31. The FO NLO prediction based on the central $\alpha_S(M_Z)$ value is drawn with a continuous black line. The minimum and maximum values for the PDF set used in the analysis are illustrated with solid red and green lines respectively, whereas all intermediate values are depicted with black dotted lines. One observes that a small change in α_S causes a large change in the inclusive jet cross section.

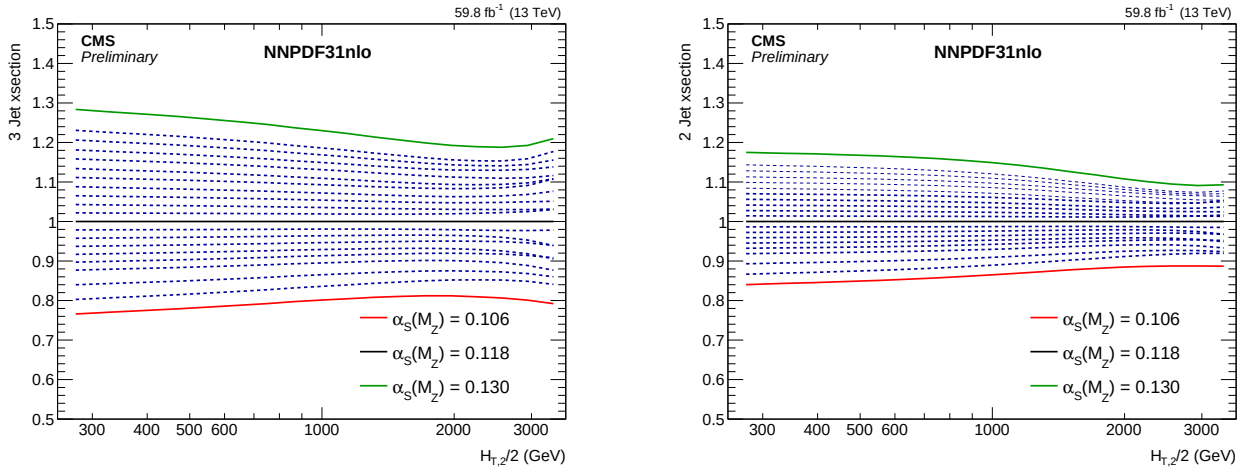


Figure 7.2: *Sensitivity to the strong coupling constant of the numerator (inclusive 3-jet cross section) (left) and the denominator (inclusive 2-jet cross section) (right) of the R_{32} observable.*

In Figure 7.3 illustrates the sensitivity of the R_{32} observable to the strong coupling constant, again for $\mu_r = \mu_f = H_{T,2}/2$. The data are shown with blue markers and include the total experimental uncertainty for the R_{32} measurement, calculated from the quadratic sum of the systematic and statistical uncertainties depicted with the orange band. The solid black line is the central $\alpha_S(M_Z)$ value. The minimum and maximum values of α_S are drawn with continuous red and green lines respectively and the intermittent lines represent the intermediate values. It is evident that a small change in α_S leads to a different prediction for R_{32}; this practically means that the observable exhibits nice sensitivity to $\alpha_S(M_Z)$.

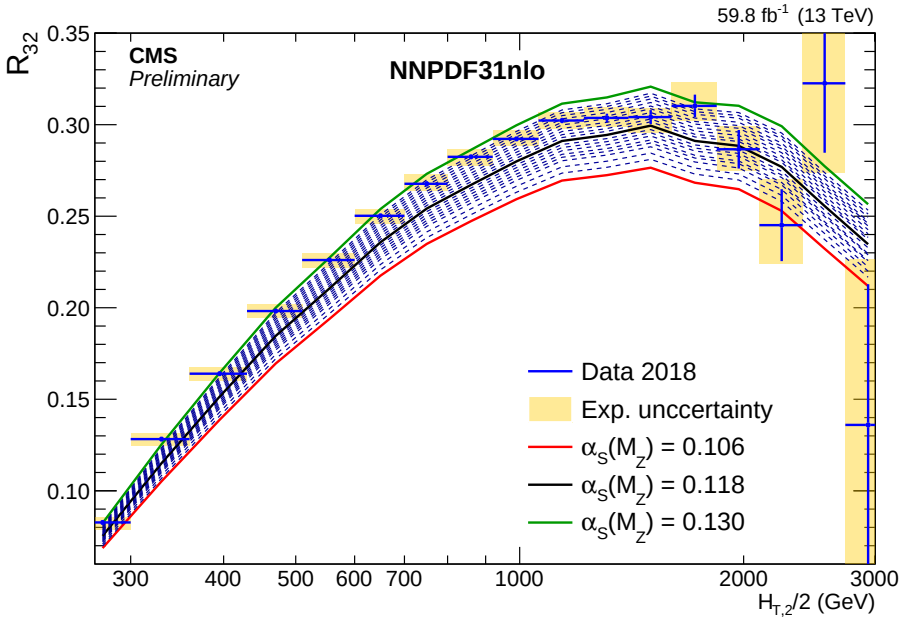


Figure 7.3: *Sensitivity of R_{32} to the strong coupling constant α_S for $\mu_r = \mu_f = H_{T,2}/2$ using the NNPDF31 PDF set.*

Bibliography

[1] Paolo Nason and Bryan Webber. Next-to-Leading-Order Event Generators. *Annual Review of Nuclear and Particle Science*, 62, 2012. https://doi.org/10.1146/annurev-nucl-102711-094928.

[2] G.P. Salam. Towards Jetography. *Eur. Phys. J. C*, 67, 2010. https://doi.org/10.1140/epjc/s10052-010-1314-6.

[3] L. Sawyer, C. Waits, and M. Wobisch. Perturbative QCD predictions in fixed order for cross section ratios. 2021. https://doi.org/10.48550/arXiv.2112.01449.

[4] Stefano Catani and Michael H. Seymour. NLO calculations in QCD: a general algorithm. *Nucl.Phys. - Proc.Suppl.*, 51, 1996. https://doi.org/10.1016/S0920-5632(96)90030-4.

[5] S. Catani and M.H. Seymour. A general algorithm for calculating jet cross sections in NLO QCD. *Nucl. Phys. B*, 485, 1997. https://doi.org/10.1016/S0550-3213(96)00589-5.

[6] S. Hoche, F. Krauss, and M. Schönherr. A critical appraisal of NLO+PS matching methods. *J. High Energ. Phys.*, 2012, 49, 2012. https://doi.org/10.1007/JHEP09(2012)049.

Chapter 8

Conclusions

To summarise, the current analysis deals with two pursuits: the measurement of the normalised inclusive differential 2-, 3-, 4- and 5-jet cross sections as well as the extraction of the R_{32}, R_{43} and R_{42} observables. These measurements were performed as a function of the average transverse momentum of the leading and sub-leading jets, $H_{T,2}/2$. For this purpose, the 2018 CMS dataset collected from pp collisions at $\sqrt{s} = 13\ TeV$ was used, that corresponds to an integrated luminosity $\mathcal{L}_{int} = 59.82\ fb^{-1}$. These measurements were compared with several MC event generators: Pythia8 CP5, CUETM1 and CUETM2, Madgraph+Pythia8 and Herwig++.
In the former case, the inclusive differential multi-jet cross sections are best described by Pythia8 CUETM1, in some cases along with Pythia8 CP5 or CUETM2, however as the multiplicity increases, all MC event generators fail to provide an accurate description.
In the latter case, all R_{mn} ratios are in good agreement with Pythia8 Pythia8 CUETM2 and Pythia8 CP5. In all cases, Madgraph+Pythia8 provides the worst description of the R_{mn} observables. The total experimental uncertainties were also calculated, with the most prominent being the JES uncertainty in all instances.
The NP corrections for the inclusive multi-jet cross sections were also calculated and they exhibit the expected behavior. The sensitivity of the R_{32} observable to the strong coupling constant was examined and compared to the data, through a comparison of the measurement to fixed order predictions at up to NLO accuracy in pQCD using NLOJET ++ within the FASTNLO framework. The experimental measurements along with the theoretical results look promising when it comes to the determination of the strong coupling constant at the scale of the Z boson mass ($\alpha_S(M_Z)$), using the observables like R_{32}.

List of Figures

List of Tables