Chapter 1

High-energy nuclear physics

1.1 Quantum Chromodynamics

In the mid-20$^{\text{th}}$ century, the realm of particle physics underwent a transformative phase, marked by the discovery of a seemingly endless variety of subatomic particles. This era witnessed the unveiling of numerous hadrons, which left physicists with the necessity of developing a framework that could describe the behaviour of these particles and their interactions. This led to the development of the static quark model, which emerged in the 1960s as a groundbreaking conceptual framework to categorise the various observed particles. Developed independently by Murray Gell-Mann[1] and George Zweig[2, 3], this model postulated the existence of fundamental constituents of hadrons called quarks, which, in order to reflect the experimental findings, had to be fermions (to describe baryons with spin 1/2 and 3/2) with fractional electric charge. The quark model beautifully explained the organization of hadrons in terms of three quarks (u, d, and s), leading to the development of a more structured and coherent classification of particles.

Despite the phenomenological success of the static quark model, it had two problems: it introduced particles with fractional charge, which had never been observed before, and, most importantly, it gave rise to a violation of the Fermi-Dirac statistics. The Δ^{++}, Δ^{-}, and Ω^{-} baryons, in fact, have symmetric orbital, spin and flavour wavefunctions, which defied the Pauli exclusion principle that should have implied antisymmetric wavefunctions for these particles.

To resolve these inconsistencies, a new degree of freedom, the *colour*, was introduced. Hadrons wavefunctions were assumed to be totally antisymmetric in colour quantum numbers, effectively implementing the Pauli exclusion principle.

The simplest model of colour would be to assign quarks to the fundamental representation of a global $SU(3)$ symmetry. Each quark now carries a colour index: q_i, where $i = 1, 2, 3$, and transforms under the fundamental (3) representation of $SU(3)$, while antiquarks, $\bar{q}_i$, transform in the $\bar{3}$ representation. Introducing the totally antisymmetric tensor ε^{ijk}, possible compositions of quarks that give rise to colour singlets are

$$\bar{q}^i q_i, \qquad \varepsilon^{ijk} q_i q_j q_k, \qquad \varepsilon^{ijk} \bar{q}_i \bar{q}_j \bar{q}_k,$$

which are the quarks compositions of mesons, baryons, and antibaryons, respectively.

One of the tests supporting the existence of colour and fractional electric charge came in the form of the ratio R, of the e^+e^- total hadronic cross-section to the cross-section of a pair of muons produced from the same annihilation process. The virtual photon emitted in the annihilation can produce all electrically charged pairs of particles and antiparticles, as shown in Fig. 1.1.

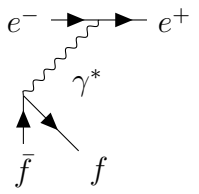


Figure 1.1: e^+e^- annihilation to a pair of fermions

The ratio R is given by

$$R = \frac{\sigma(e^+e^- \to hadrons)}{\sigma(e^+e^- \to \mu^+\mu^-)} = N_c \sum_f Q_f^2 \ ,$$

where N_c represents the number of existing colours and Q_f is the electric charge of the quark flavour f. Notably, this ratio is dependent on the energy of the center-of-mass system and encompasses all possible quark flavors that can be produced by the virtual photon at that specific energy level. The experimental data for R (shown in Fig. 1.2) exhibited a remarkable agreement with the predictions of the three-color model, thereby providing compelling evidence for the existence of color and fractional electric charge of quarks.

The final step that propelled the development of Quantum Chromodynamics (QCD) as a comprehensive theory of the strong force was the insight into the mechanism that ensured all hadron wavefunctions to be color singlets. This emerged from the discovery of asymptotic freedom, a phenomenon observed in deep-inelastic scattering experiments. Non-Abelian gauge theories, often referred to as Yang-Mills theories, were identified as having this unique characteristic. This realization led to the formulation of QCD by elevating the global color $SU(3)$ symmetry to a local one, allowing the 8 quanta of the $SU(3)$ gauge field, called *gluons*, to mediate the strong force, successfully describing the confinement and behavior of quarks and gluons within hadrons.

The QCD Lagrangian density can be written as

$$\mathcal{L}_{QCD} = -\frac{1}{4} F^a_{\mu\nu} F^{\mu\nu}_a + \sum_f \bar{q}^i_f (i\gamma^\mu (\mathcal{D}_\mu)_{ij} - m_f \delta_{ij}) q^j_f \ , \tag{1.1}$$

where $F^a_{\mu\nu}$ is the field strength tensor defined in terms of the gluon field A^a_μ and the $SU(3)$ structure constant f^{abc}:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^{abc} A^b_\mu A^c_\nu \tag{1.2}$$

and $(\mathcal{D}_\mu)_{ij}$ is the covariant derivative:

$$(\mathcal{D}_\mu)_{ij} = \partial_\mu \delta_{ij} - i g_s (t^a)_{ij} A^a_\mu \ ,$$

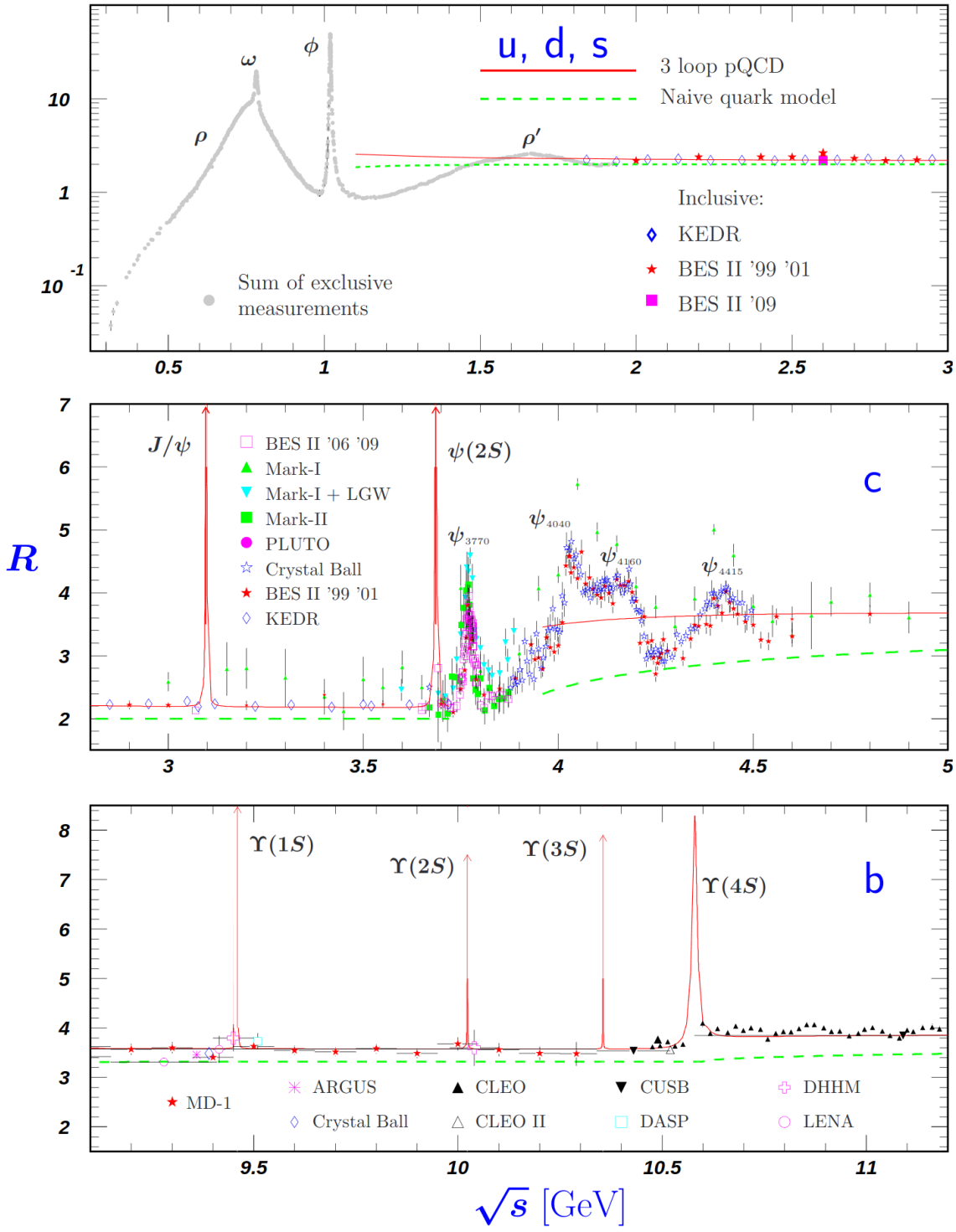


Figure 1.2: R as a function of $\sqrt{s}$ in the light-flavor, charm, and beauty threshold regions taken from [4]. The green curve is a naive quark-parton model prediction, while the red one is a 3-loops perturbative QCD prediction. Breit-Wigner parameterizations of J/ψ, ψ(2S), and Υ(nS), n = 1,2,3,4 are also shown

with t^a being one of the generators of the $SU(3)$ representation.

The last term in Eq. 1.2 is peculiar to non-Abelian theories, and gives rise to triplet and quartic gluon self-interactions illustrated in Fig. 1.3. g_s is a coupling parameter related to the coupling constant α_s, which determines the strength of the interaction between the coloured particles.

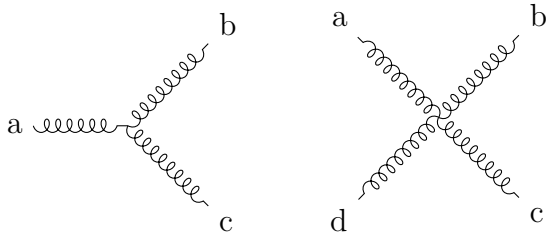


Figure 1.3: Feynman diagrams for gluons self-interactions

The second term of Eq. 1.1 describes the interactions between quarks and gluons, sketched in Fig. 1.4, and contains the mass term for the fermions. It is noteworthy to observe that the interaction between quarks and gluons is diagonal in flavor, meaning that the strong interaction conserves the flavor of quarks. In contrast, colour mixing is allowed within the framework of QCD.

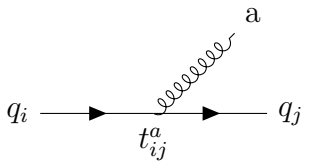


Figure 1.4: Feynman diagram for quark-gluon interaction

1.1.1 Running coupling constant

If one considers a dimensionless physical observable, denoted in the following as R, which solely depends on a single energy scale, Q, one might naturally expect that R would maintain a constant value, independent of the specific energy scale chosen. However, this does not hold true when loop diagrams are studied: the necessity of renormalisation introduces a new energy scale denoted as μ. This scale, known as the renormalisation scale, is the point at which the subtraction of the ultraviolet divergences is carried out. Critically, μ is an arbitrary parameter and, as such, is non-physical. Consequently, R becomes dependent on the ratio Q^2/μ^2 and the renormalised coupling $\alpha_s = g_s^2/4\pi$: $R = R\left(\frac{Q^2}{\mu^2}, \alpha_s\right)$. The μ independence of R (which is an essential requirement given μ's arbitrariness) can be expressed as

$$\mu^2 \frac{\mathrm{d}R\left(\frac{Q^2}{\mu^2}, \alpha_s\right)}{\mathrm{d}\mu^2} = \mu^2 \left[\frac{\partial}{\partial \mu^2} + \frac{\partial \alpha_s}{\partial \mu^2}\frac{\partial}{\partial \alpha_s}\right] R\left(\frac{Q^2}{\mu^2}, \alpha_s\right) = 0 \ , \tag{1.3}$$

a fundamental equation known as the renormalisation group equation. This equation is exactly true in the case of a prediction that considers all perturbative orders. If

one limits the expansion at a fixed order α_s^N, then a dependence of R from μ is observed at the α_s^{N+1} order.
Solving Eq. 1.3 requires the introduction of the concept of the running coupling $\alpha_s(Q^2)$, which evolves as a function of Q. By introducing

$$t \equiv \log(Q^2/\mu^2), \qquad \beta(\alpha_s) \equiv \mu^2 \frac{\mathrm{d}\alpha_s}{\mu^2} \quad ,$$

Eq. 1.3 can be written as

$$\left(-\frac{\partial}{\partial t} + \beta(\alpha_s)\frac{\partial}{\partial \alpha_s}\right) R(e^t, \alpha_s) = 0$$

This first-order partial differential equation can be solved by defining a new function: the running coupling $\alpha_s(Q^2)$

$$t = \log(Q^2/\mu^2) \equiv \int_{\alpha_s}^{\alpha_s(Q^2)} \frac{\mathrm{d}x}{\beta(x)}, \quad \text{with } \alpha_s = \alpha_s(\mu^2) \quad . \tag{1.4}$$

By differentiating Eq. 1.4 with respect to t and α_s, one gets:

$$\beta(\alpha_s(Q^2)) = \frac{\partial \alpha_s(Q^2)}{\partial t}, \quad \frac{\mathrm{d}\alpha_s(Q^2)}{\mathrm{d}\alpha_s} = \frac{\beta(\alpha_s(Q^2))}{\beta(\alpha_s)} \quad . \tag{1.5}$$

It results from this last set of equations that $R(1, \alpha_s(Q^2))$ satisfies Eq. 1.3; hence, the running coupling constant has absorbed the μ scale dependence of R. As a consequence, the knowledge of $R(1, \alpha_s)$, which can be evaluated in fixed-order perturbation theory, allows knowing the dependence of R from Q^2, the physical scale at which the coupling is gauged, by simply substituting $\alpha_s \to \alpha_s(Q^2)$.

The β function

The running of the coupling constant is determined by the $\beta(\alpha_s)$ function, which is evaluated from loop corrections to the bare vertices of the theory. As of the time of the writing of this Thesis, the β function has been evaluated up to 5 loops[5]. In Fig. 1.5, the 1-loop Feynman diagrams contributing to the β function evaluation are reported.

By limiting the calculations at the first order in the perturbative expansion, one gets:

$$\beta(\alpha_s) = -\alpha_s^2 \frac{11\mathrm{N_c} - 2\mathrm{N_f}}{12\pi} + \mathcal{O}(\alpha_s^3) \equiv -\alpha_s^2 \beta_0 + \mathcal{O}(\alpha_s^3) \quad , \tag{1.6}$$

where $\mathrm{N_c}$ is the number of colours (3), while $\mathrm{N_f}$ is the number of quark flavours which can be considered massless at the physical scale Q^2 at which the coupling is being measured. From Eqs. 1.6 and 1.5, one can extract the Q^2 dependency of the running coupling constant:

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \alpha_s(\mu^2)\beta_0 \log(Q^2/\mu^2)} \ , \tag{1.7}$$

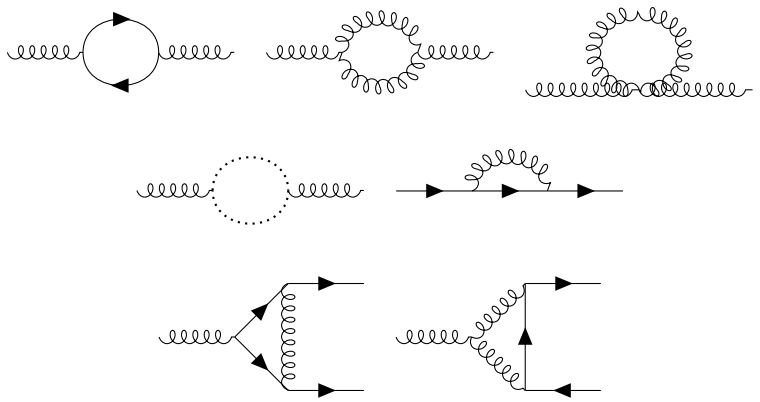


Figure 1.5: 1-loop Feynman diagrams contributing to the β function evaluation

Notably, since β_0 is positive in a 6 quark-flavours framework, the strong coupling constant exhibits a monotonic decreasing trend as a function of Q^2. This behaviour differs from the one of the electromagnetic coupling constant, which increases with the energy scale due to the screening effect of vacuum polarisation. For QCD, the running of the coupling constant is a direct consequence of the non-Abelian nature of the theory, allowing for gluon self-interactions, which give rise to an anti-screening effect. The idea is that the emission of virtual gluons by static colour sources causes their colour charges to 'leak out' into the surrounding vacuum. Since the interaction between distributions of charges is weaker than the one between point-like charges when the distributions overlap, the effective coupling constant decreases at short distances. This behaviour is known as asymptotic freedom, a key feature of QCD that allows for the perturbative expansion of the theory at high energy scales, where the strong coupling constant is small. At the same time, the running of the coupling constant implies that the theory is non-perturbative at low energy scales, and phenomenological models are required to describe the strong interaction in this regime.

Instead of using the renormalisation scale μ as a free parameter, one can use the running coupling constant to define a physical scale, Λ_{QCD}, which is the energy scale at which the coupling constant would diverge, if extrapolated outside the perturbative regime. Using Eq. 1.7, one can write

$$\alpha_s(\Lambda_{QCD}) = \frac{1}{\beta_0 \log(Q^2/\Lambda^2_{QCD})} \quad .$$

The value of Λ_{QCD} is determined by the specific definition being used. However, to obtain the value of the coupling constant measured at $Q^2 = M_Z^2$, an approximate value of Λ_{QCD} of around 200 MeV can be used.

Measurements of the running of the coupling constant at different values of Q are illustrated in Fig. 1.6 and compared to the theoretical prediction at 5 loops. The agreement between the experimental data and the theoretical prediction is remarkable, confirming the validity of the QCD framework at high energy scales.

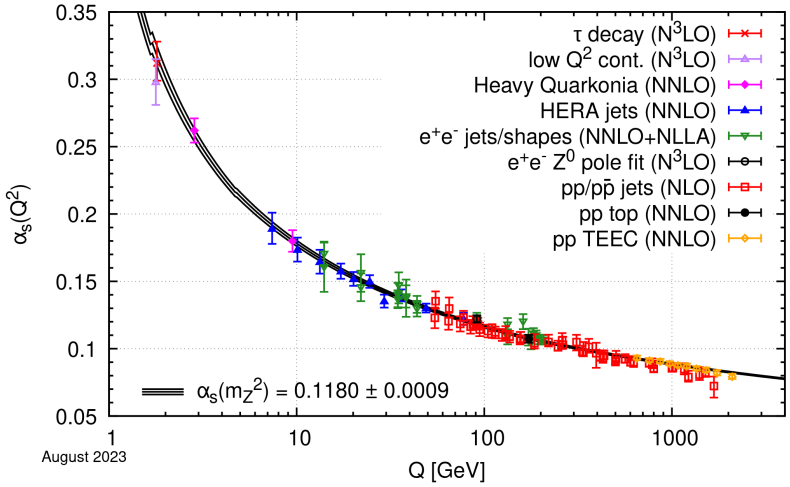


Figure 1.6: Summary of measurements of α_s as a function of the energy scale Q, compared to the running of the coupling computed at five loops, taking as an input the current PDG average, $\alpha_s(M_Z^2) = 0.1180 \pm 0.0009$ GeV/c^2. Taken from [4]

1.2 Confinement

The concept of confinement is one of the most intriguing aspects of QCD. It is the phenomenon by which quarks and gluons are never observed as free particles, but are always confined within colour-neutral hadrons. The confinement of quarks and gluons is a direct consequence of the non-Abelian nature of the theory, which, as described in the previous Section, is characterised by an increase of the strong coupling constant at low energy scales. The confinement of quarks and gluons is a non-perturbative effect, and despite extensive research, a comprehensive theoretical description of confinement in QCD remains elusive. Phenomenological models like the MIT bag model have been proposed, but a full comprehension of confinement is still lacking. Lattice QCD simulations are the most successful approach to study the non-perturbative regime of the theory, and they have provided a wealth of information on the properties of hadrons and the strong interaction at low energy scales.

1.2.1 MIT bag model

The MIT bag model [6] is a phenomenological model of confinement, which describes hadrons as bound states of quarks and gluons confined within a finite volume, called the bag. The model was developed in the 1970s by A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf, and it has been widely used to study the properties of hadrons and the strong interaction. In the MIT bag model, N non-interacting massless fermions are confined within a spherical cavity of radius R, which is the bag radius. The confinement arises from a balance between pressure due to the kinetic energy of the fermions inside the bag and an ad hoc external pressure, which is introduced to confine the fermions within the bag. The fermions are described by the Dirac equation for massless fermions:

$$i\gamma^\mu \partial_\mu \psi = 0 \quad ,$$

where ψ is the fermion field, and γ^μ are the Dirac matrices. The solution to the Dirac equation is given in terms of the spherical Bessel functions of the zeroth and first order, $j_0(p_0 r)$ and $j_1(p_0 r)$:

$$\psi = \mathcal{N} e^{-ip_0 t} \begin{pmatrix} j_0(p_0 r)\chi^+ \\ \vec{\sigma} \cdot \hat{r} j_1(p_0 r)\chi^- \end{pmatrix} \quad ,$$

where p_0 is the energy of the fermion, χ^+ and χ^- are the two components of the fermion four dimentional spinor ψ, and $\vec{\sigma}$ are the Pauli matrices. The colour flux at a point r inside the bag is given by

$$j^\mu_{ab}(r) = \bar{\psi}_a(r)\gamma^\mu \psi_b(r) \quad ,$$

where a and b are the colour indices of the fermions. If the quantum numbers are not to be lost through the surface of the bag, which is the definition of confinement, then:

$$n_\mu j^\mu_{ab}(r) = \bar{\psi}_a(r)\gamma \cdot n \psi_b(r) = 0$$

on the surface, where n_μ is a unit space-like vector perpendicular to the surface. Using the gamma properties, $(i\gamma \cdot n)^2 = 1$, so that by assuming that $i\gamma \cdot n = +1$, the boundary condition on the surface of the bag is given by

$$\bar{\psi}(R)\psi(R) = 0 \quad ,$$

leading to the solution of the Dirac equation in the bag:

$$\left[j_0\left(p_0 R\right)\right]^2 - \left[j_1\left(p_0 R\right)\right]^2 = 0 \quad ,$$

with solution $p_0 R = 2.04$. The total energy inside the bag is given by

$$E = \frac{2.04N}{R}(\hbar c) + \frac{4\pi}{3} R^3 B \quad ,$$

where the first term is the kinetic energy of the fermions, and the second term is the energy due to the presence of an external pressure B which keeps the fermions confined in the bag. The bag pressure is a phenomenological parameter of the model, and it is introduced to confine the fermions within the bag. It can be extracted by minimising the energy of the system with respect to the bag radius R, which yields $B = 234$ MeV/fm^3 for a baryon with $R = 0.8$ fm.

1.2.2 Lattice QCD

Lattice QCD is a numerical technique used to study the non-perturbative regime of QCD. The method is based on the discretisation of space-time on a four-dimensional lattice, and the evaluation of the path integral of the theory using Monte Carlo methods, i.e. by sampling possible configurations of the quark and gluon fields according to the probability distribution given by the QCD Lagrangian. The lattice spacing

is a parameter of the method, and allows one to avoid the ultraviolet divergences of the theory, which are typical in perturbative QCD, by introducing a cutoff on the momenta of the quark and gluon fields.

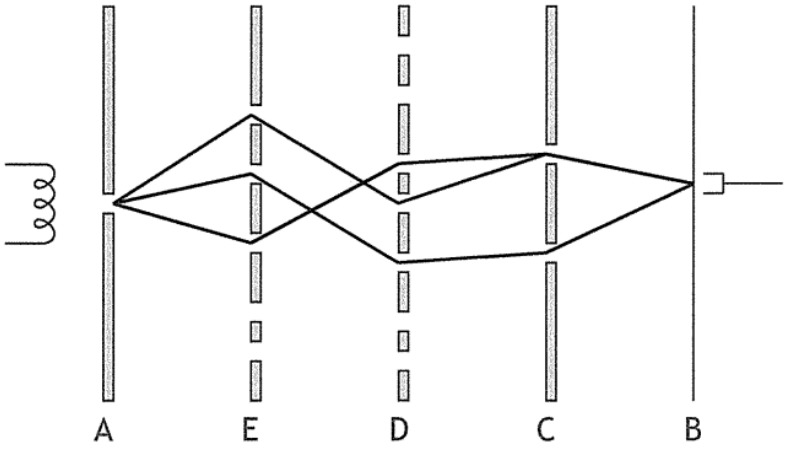


Figure 1.7: Feynman introduction to path integrals. Here, a particle emitted from a source at x_a is detected at x_b. A finite number of screens, each with a finite number of holes, is placed between the source and the detector. The probability amplitude for the particle to hit the detector is given by the sum of the probabilities of moving from the source to the detector through all possible paths. By adding an infinite amount of screens with an infinite number of holes, and by also considering the time at which the particle passes through each screen, the sum becomes an integral over all possible paths, called a *path integral*.

The Lattice QCD simulations are based on the path integral formalism of quantum field theory [7], developed by R. Feynman in the 1940s. The path integral provides a natural extension to quantum mechanics of the least action principle of classical mechanics, and it allows one to calculate the probability amplitude of a particle to move from one point to another in space-time, considering the evolution of the system over all possible paths. The transition amplitude from the state (x_a, t_a) to the state (x_b, t_b) is given by

$$A\left[(x_a, t_a) \to (x_b, t_b)\right] = \langle x_b, t_b | e^{-iH(t_b - t_a)} | x_a, t_a \rangle = \sum_{\text{paths}} e^{iS[x(t)]} \quad , \tag{1.8}$$

where H is the Hamiltonian of the system, $S[x(t)]$ is the action of the system for a given path $x(t)$, and the sum is over all possible paths from (x_a, t_a) to (x_b, t_b). By taking the continuum limit on space-time, one obtains an integration over all the possible space-time paths of the system:

$$\sum_{\text{paths}} e^{iS[x(t)]} \to \int_{x_a}^{x_b} \left[\mathcal{D}x(t)\right] e^{iS[x(t)]} \quad , \tag{1.9}$$

where the right-hand side term is a functional integral over all possible paths. It is interesting to note that by combining Eqs. 1.8 and 1.9, a quantity resembling the

partition function of a statistical system is obtained:

$$\mathcal{Z} = \sum_{x_a} \langle x_a, t_a | e^{\beta H} | x_a, t_a \rangle \quad .$$

It is possible to express the partition function in terms of a path integral by applying a Wick rotation to the time variable, $t \to -i\tau$, with $\tau_a = 0 \leq \tau \leq \tau_b = \beta$ and considering the Euclidean action in place of the Minkowskian one, $S_E = iS$. Furthermore, since the state at τ_a is the same as the one at τ_b in the partition function definition, a periodic boundary condition is imposed: $x(\tau_a) = x(\tau_b)$. With these considerations, the partition function can be expressed as

$$\mathcal{Z} = \int [\mathcal{D}x(\tau)]\, e^{-S_E[x(\tau)]} \quad .$$

This formalism, which was here developed for a single particle, can be extended to a quantum field theory, and in particular to QCD.

Lattice QCD simulations are computationally intensive, and they require large supercomputers to perform the calculations. To limit the computational costs, calculations are often performed at larger up and down quark masses than in nature, drastically reducing the number of virtual quark-antiquark loops that have to be taken into account. Because of the employed Monte Carlo approach, only a finite number of configurations can be considered, leading to statistical uncertainties in the lattice QCD results. In order to obtain physical results, several limits have to be taken: i. the continuum limit, i.e. the extrapolation of the lattice spacing to zero, ii. the infinite-volume limit, i.e. the extrapolation of the lattice size to infinity, and iii. the physical quark-mass limit, i.e. the extrapolation to physical quark masses. Many present-day lattice calculations are already performed directly at, or very close to, the physical values of the quark masses, so that the latter extrapolation becomes less of an issue.

The results of the lattice QCD simulations are in good agreement with the experimental data as shown in Fig. 1.8 for the spectrum of hadrons obtained from lattice QCD simulations, taken from [8], compared to the experimental data.

1.3 Quark Gluon Plasma

The concept of deconfinement refers to the transition from a confined state to a state where quarks and gluons are no longer confined within colour-neutral hadrons, but can move in a larger volume. As modelled by the MIT bag model, non-perturbative QCD effects can be described in terms of an external pressure, which confines quarks and gluons within a finite volume. If the external pressure is overcome by the pressure due to the kinetic energy of the quarks and gluons, then the hadrons constituents are no longer confined, and a transition to a state called Quark-Gluon Plasma (QGP) occurs. Lattice QCD calculations are used to understand the properties of the QGP, and they predict that a strongly interacting system with zero net baryon density evolves smoothly from a confined (hadronic) towards a deconfined (quarks and gluons) state when its temperature is increased up to ~ 155 MeV (1.8×10^{12} K) [9, 10], reaching energy densities of ~ 1 GeV/fm^3.

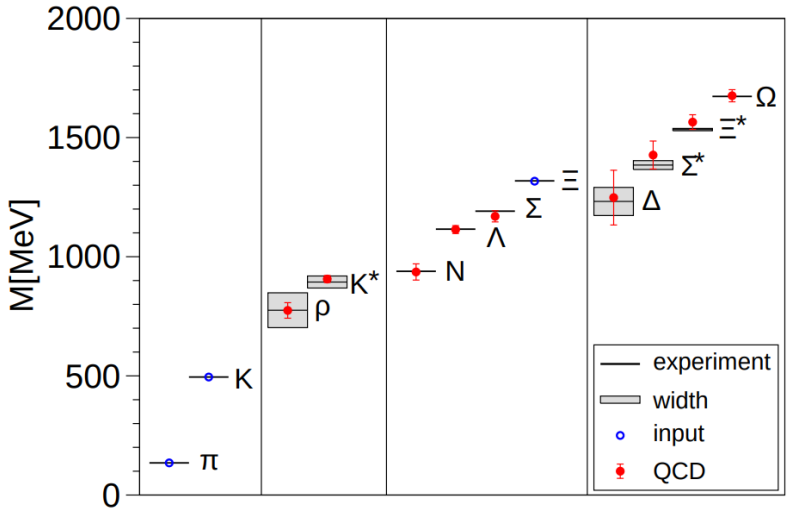


Figure 1.8: The light hadron spectrum of QCD. Horizontal lines and bands are the experimental values with their decay widths. Lattice QCD results [8] are shown by solid circles. Vertical error bars represent the combined statistical and systematic error estimates. π, K and Ξ have no error bars, because they are used to set the light quark mass, the strange quark mass, and the overall scale, respectively.

It is believed that the Universe underwent a phase of deconfinement in the early stages of its evolution, a few microseconds after the Big Bang. Direct observation of the primordial QGP (i.e. that created just after the Big Bang) would provide a wealth of information on the early Universe; however, the Universe experienced a phase in which electrons were not bound to nuclei (electromagnetic plasma), making it opaque to electromagnetic radiation, and denying us the possibility of directly observing the primordial QGP. Once the Universe cooled enough (3000 K) to allow electrons to bind to nuclei, the electromagnetic radiation decoupled with a black body spectrum of around 3000 K. Since then, as the Universe expanded, this electromagnetic radiation has redshifted to a temperature of around 2.7 K, and is denoted as the Cosmic Microwave Background (CBM). The CMB is the oldest light in the Universe and provides a snapshot of the Universe when it was 300'000 years old, long after the QGP had already cooled down. Hence, the only way to study the QGP is by recreating it in laboratories, through the collision of heavy ions at high energies. In the past decades, several experiments [12, 13, 14, 15] have been carried out to study the properties of this state of matter, and the results have provided valuable insights into the properties of the strongly-interacting matter.

1.3.1 High-energy heavy-ion collisions

Heavy-ion collisions are the most suitable environment to study the properties of the QGP. In these collisions, two heavy ions, such as lead or gold nuclei, are accelerated to ultra-relativistic energies and made to collide head-on. Given the large amount of energy deposited in the collision, the system reaches high energy densities of around 16 GeV/fm^3 after 1 fm/c [17], which allows for the production of the QGP. As nuclei are objects made of many nucleons (i.e. protons and neutrons), the interaction

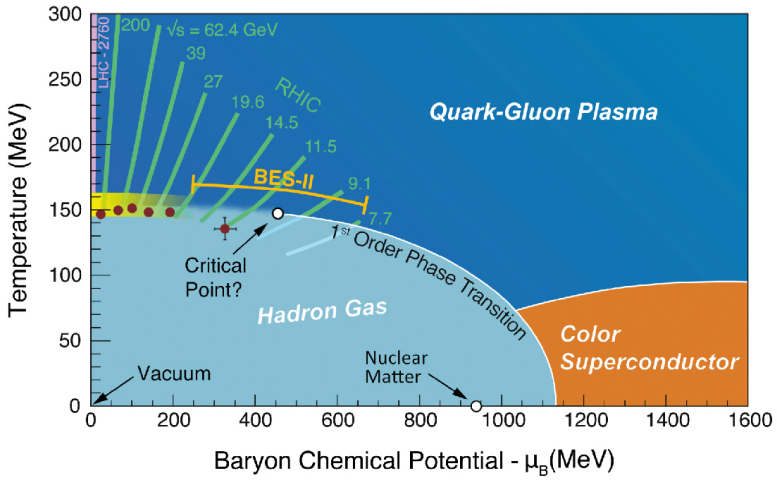


Figure 1.9: QCD phase diagram in the temperature-baryon chemical potential plane. Taken from [11].

volume (called *fireball*) is larger and longer-lived than in proton-proton collisions, allowing the usage of thermodynamics and fluidodynamics to describe the system.

It is possible to distinguish between two different collision regimes. When the centre-of-mass energy per nucleon pair ($\sqrt{s_{\mathrm{NN}}}$) is below a few GeV, the nucleons are stopped in the collision as they lose energy and momentum. Due to conserved currents, the quantum numbers of the initial state are preserved, so that, for example, the net baryon production (i.e. the difference between the number of baryons and antibaryons) is positive. This regime is called *stopping regime*. When the $\sqrt{s_{\mathrm{NN}}}$ increases, the initial state baryon number is carried away by the receding nucleons, and the net baryon number in the fireball is zero. This regime is denoted as *Bjorken regime*, or *transparency regime*.

The collision of two nuclei is a rather complex process, with a space-time evolution that can be divided into several stages, as depicted in Fig. 1.10. They can be studied by measuring different final-state observables that are correlated to such stages. One of the first descriptions of the fireball evolution was given by Bjorken [18], who proposed a simple model to describe the expansion of the system in the longitudinal direction. In this model, it is assumed that there exists a central-plateau structure in the inclusive particle productions as a function of the rapidity variable, which is defined as

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

where E is the energy and p_z is the longitudinal momentum of the particle. In other terms, the Bjorken model assumes that the system is boost-invariant, i.e. independent of the longitudinal velocity. According to this model, the evolution of the system is described by the following stages: i. collision of the two nuclei, ii. pre-equilibrium, iii. QGP formation, iv. hadronisation, and v. freeze-out.

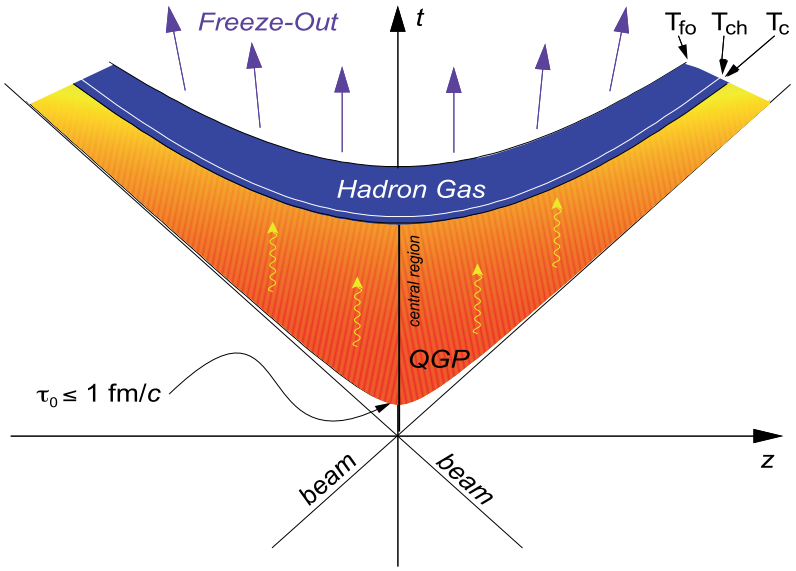


Figure 1.10: Space-time evolution of a heavy-ion collision. Taken from [16].

Collision

The two nuclei are accelerated to ultra-relativistic energies and are thus Lorentz-contracted in the direction of motion. The collision takes place in a very short time $\tau_{\text{coll}} = 2R/\gamma$, where R is the nucleus radius and γ the Lorentz factor. At LHC energies, the nuclei crossing time is of ~ 0.005 fm/c, which is much smaller than the time scale of the strong interaction $\tau_{\text{strong}} \sim 1/\Lambda_{QCD} \sim 1$ fm/c. Particles produced in the collision through parton interactions mediated by the strong force are thus created once the colliding nuclei have already passed through each other and moved away from the interaction region. During the collision, a large amount of energy is deposited in the interaction region, leading to the formation of a fireball with high energy densities.

Pre-equilibrium

After τ_{coll}, secondary particles are produced from the energy deposited in the collision. In the Bjorken regime, it is possible to evaluate the energy density of the system as a function of time by measuring the transverse energy of the particles produced at midrapidity, where the net baryon density is zero. The energy density is given by

$$\varepsilon_{\text{Bjorken}} = \frac{\langle m_{\text{T}} \rangle}{\tau A} \left. \frac{\mathrm{d}N}{\mathrm{d}y} \right|_{y=0} = \frac{1}{\tau A} \left\langle \left. \frac{\mathrm{d}E_{\text{T}}}{\mathrm{d}y} \right\rangle \right|_{y=0} ,$$

where A is the transverse area collision region, $\langle m_{\text{T}} \rangle$ is the mean transverse mass, defined as $m_{\text{T}} = \sqrt{m^2 + p_{\text{T}}^2}$, N is the number of secondary particles, and $\langle \mathrm{d}E_{\text{T}}/\mathrm{d}y \rangle$ is the mean transverse energy density. It is interesting to evaluate the energy densities reached at the time of particle formation, which can be estimated using Heisemberg's uncertainty principle: $\tau = \hbar/m_{\text{T}}$. The mean transverse mass of the particles produced in the collision at the proper time of formation τ_{f} can be evaluated using

the approximate formula:

$$\langle m_{\mathrm{T}} \rangle = \frac{\frac{\mathrm{d}E_{\mathrm{T}}(\tau_{\mathrm{f}})}{\mathrm{d}y}}{\frac{\mathrm{d}N(\tau_{\mathrm{f}})}{\mathrm{d}y}} \sim \left. \frac{\frac{\mathrm{d}E_{\mathrm{T}}}{\mathrm{d}\eta}}{\frac{\mathrm{d}N}{\mathrm{d}\eta}} \right|_{\text{final state}} ,$$

which leads to a formation time of ~ 0.35 fm/c at RHIC [19], to which corresponds an energy density of ~ 15 GeV/fm^3.

Quark-gluon plasma formation

The system of produced particles reaches thermal equilibrium through multiple scatterings at a proper time τ_{eq}, and the QGP is formed. Its evolution can be described using relativistic hydrodynamics, which can be used to predict τ_{eq}. Studies at RHIC allow to constrain τ_{eq} in the range of $0.6 < \tau_{\mathrm{eq}} < 1$ fm/c, corresponding to energy densities of $5.4 < \varepsilon_{\mathrm{Bjorken}}(\tau_{\mathrm{eq}}) < 9$ GeV/fm^3, consistent with what is expected for QGP formation.

Hadronisation

The deconfined system expands and cools down, until the temperature decreases below the pseudo-critical value for the transition crossover. The QGP undergoes a phase transition and hadronises, producing an expanding gas of colour-neutral particles. At LHC energies, this is expected to happen ~ 10 fm/c [20] after the QGP formation. This transition is associated with a sharp decrease in the entropy density of the system.

Freeze-out

During the hadron gas expansion, the particle density decreases to a point where the inelastic interactions cease. The moment in which this happens is called *chemical freeze-out*, and takes place at around $T_{\mathrm{chem}} \sim 160$ MeV [21]. The chemical abundances of the hadron gas cannot vary anymore, although elastic interactions, which change the momentum spectrum of the produced particles still occur. Once the temperature decreases below $T_{\mathrm{therm}} \sim 130$ MeV, the elastic interactions cease as well, and this is referred to as *thermal freeze-out*. The particles then keep expanding without interactions (*free-straming*), and are detected by the experimental apparatus.

1.3.2 Hadron species abundances

Measurements of the multiplicity of identified hadron species, i.e., the chemical composition of the system, allow for the investigation of various properties of the system at the time of chemical freeze-out. The relative abundances of different hadron species produced in hadronic collisions are found to be well described by the *Statistical Hadronisation Model* (SHM) [22]. The SHM is based on the assumption that the fireball created in an ultra-relativistic heavy-ion collision is in thermal and chemical equilibrium at the time of chemical freeze-out, which is assumed to be characterised by the same temperature for all the hadronic species.

A system is in thermal equilibrium when the distribution of particle velocities follows a Maxwell-Boltzmann distribution, which depends solely on the temperature of the system. Chemical equilibrium is achieved when the multiplicities of the different hadron species can be described in terms of their mass and the temperature of the system. The hypothesis of thermal and chemical equilibrium is postulated, and allows the use of statistical mechanics to describe the system by defining a partition function that characterises the statistical properties of a system in equilibrium. The validity of this assumption is then verified a posteriori by comparing the model predictions with the experimental data. The equilibrium at the chemical freeze-out is purely statistical, i.e. the particles' abundances are determined through the principle of maximum entropy, which suggests that hadronisation fills the phase space in the most probable configuration. Furthermore, the SHM does not take into account the dynamics of the system, but only its final state at the time of the chemical freeze-out, without any assumption on the presence of a partonic phase.

The system, which can be described with a Yukawa approach in terms of a hadron gas interacting through the exchange of resonances, is described in the SHM picture in terms of a non-interacting gas of hadrons and resonances. This framework was found to be consistent with the equation of state from lattice QCD for temperatures below the pseudo-critical temperature of the QCD phase transition [9].

The ensemble used to describe the system is the grand canonical ensemble, which describes a system exchanging energy and particles with the environment. In a grand canonical ensemble, the energy and the quantum numbers are conserved on average, but not exactly and locally as in a micro canonical ensemble. The fireball produced in an ultra-relativistic heavy-ion collision can be described in terms of small clusters that exchange energy and particles with the rest of the fireball. Smaller collision systems such as pp and e^+e^- collisions, where fewer particles are produced, are described by the canonical ensemble, where the quantum numbers are conserved exactly and locally, while the energy is only conserved on average.

To reduce the amount of calculations, the SHM typically focuses on the lightest hadron species, which are the most abundant in the system. This approach excludes heavy-flavour hadrons, such as mesons with mass $M \gtrsim 1.8$ GeV/c^2 and baryons with mass $M \gtrsim 2$ GeV/c^2. In the examined mass range the hadron spectrum is well understood, and decay channels are precisely known. However, this limits the validity of the model to $T \lesssim 190$ MeV, as at higher temperatures the heavy-flavour hadrons and resonances start to contribute significantly to the particle multiplicities. Nevertheless, the description of the system in terms of a hadron gas would not be valid for temperatures above the pseudo-critical temperature of the QCD phase transition (~ 155 MeV), where the system is expected to be in a deconfined state.

The density of the hadron species i in the system can be described from the partition function of the system and is given by

$$n_i = \frac{1}{V}\frac{\partial(TZ_{\mathrm{i}}^{\mathrm{GC}})}{\partial\mu_{\mathrm{i}}} = \frac{g_{\mathrm{i}}T^2}{2\pi^2}\sum_{\mathrm{k}=1}^{\infty}\frac{(\pm1)^{\mathrm{k}+1}}{k}\lambda_{\mathrm{i}}^{\mathrm{k}}m_{\mathrm{i}}^2K_2\left(\frac{\mathrm{k}m_{\mathrm{i}}}{T}\right) \quad ,$$

where V is the volume of the system, g_{i}, $Z_{\mathrm{i}}^{\mathrm{GC}}$, $\lambda_{\mathrm{i}} = e^{\mu_{\mathrm{i}}/T}$ and m_{i} are the degeneracy factor, the grand canonical partition function, the fugacity, and the mass of the hadron species i, respectively. K_2 is the modified Bessel function of the second kind.

The chemical potential μ_{i} of the hadron species i is related to the conservation (on average, on a large volume) of the quantum numbers of the system. It is given by $\mu = \sum_{\mathrm{j}} \mu_{Q_{\mathrm{j}}} Q_{\mathrm{j}}$, where Q_{j} are the conserved quantum numbers and $\mu_{Q_{\mathrm{j}}}$ are the chemical potentials, which guarantee the conservation of the conserved charges. μ_{i} is thus the energy needed to add a particle with quantum numbers Q_{j} to the system.

The obtained multiplicities of the hadron species are then corrected with: i. the feed-down from decays of short-lived particles, which cannot be detected directly; ii. short-distance repulsions of hadrons, modelled with an excluded volume correction by assigning each hadron a volume $V = 4\frac{4}{3}\pi R^3$, where R corresponds to the hard-core radius measured in nucleon-nucleon collisions, ~ 0.3 fm; iii. the width of the resonances, by adding an integration over the mass using a Breit-Wigner distribution; iv. a suppression factor on the strangeness production γ_{s}, which is taken into consideration as strange quarks might not reach chemical equilibrium due to their larger mass than up and down quarks. This last correction factor is found to be compatible with 1 at the energies of the SPS, RICH, and LHC, hinting at a strangeness equilibrium in the system.

Of the five free parameters of the SHM (T, μ_B, μ_s, μ_{I_3} and V), two can be fixed through conservation of the initial state quantum number. Particle abundances are typically fitted in terms of the temperature T, the baryon chemical potential μ_B and the volume of the system V.

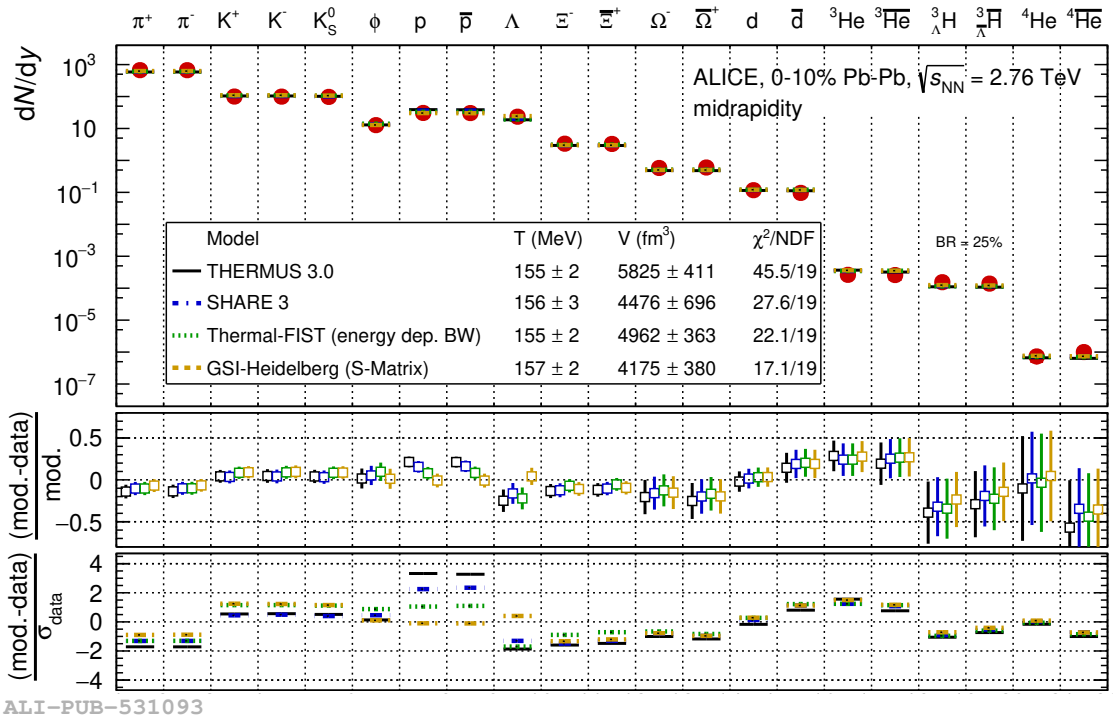


Figure 1.11: Multiplicity per unit of rapidity of different hadron species and light nuclei measured by ALICE compared to SHM fits from THERMUS [23], SHARE [24], Thermal-FIST [25], and GSI-Heidelberg [26]. Differences between the model calculations and the measured yields are shown in the bottom panels. Taken from [12].

Figure 1.11 illustrates the yields of different identified hadron species measured by the ALICE Collaboration in the 10% most central Pb–Pb at $\sqrt{s_{\rm NN}} = 5.02$ TeV. They are compared to predictions from various implementations of the SHM, including THERMUS [23], GSI-Heidelberg [26], SHARE [24] and Thermal-FIST [25]. These models make different assumptions regarding the equilibrium condition and/or conservation laws at the chemical freeze-out stage. The SHM is capable of describing the different particles' abundances, which are charcaterised by a large range of 9 orders of magnitude. The yields of exotic hadron species, such as hyperons and light nuclei are precisely described too. This confirms the assumption that the system is in thermal and chemical equilibrium at the time of the chemical freeze-out, and that the hadronisation process is well described by a statistical model.

The freeze-out temperature is found to be of $\sim$ 155 MeV, very close to that predicted by lattice QCD calculations for QCD phase transition. Most recent results from the ALICE Collaboration [27] show that the measured electric charge and baryon chemical potentials at midrapidity at the LHC energies are compatible with 0, indicating that the system created in Pb–Pb collisions at the LHC is on average baryon-free and electrically neutral.

1.3.3 Radial flow

After the chemical freeze-out, inelastic interactions between the hadrons do not take place anymore. However, elastic interactions can still occur, leading to a modification of the momentum distribution of the particles. By studying the transverse momentum spectra of the particles produced in the collision, it is possible to extract information on the produced system.

For a stationary thermal source with a temperature T, the Lorentz-invariant momentum distribution of the particles is given by

$$E\frac{\mathrm{d}N}{\mathrm{d}^3p} = \frac{\mathrm{d}N}{p_{\rm T}\mathrm{d}p_{\rm T}\mathrm{d}\varphi\mathrm{d}y} = \frac{g_{\rm i}V}{(2\pi)^3}E\frac{1}{e^{(E-\mu_{\rm i})/T}\pm 1} \quad ,$$

where E is the energy of the particle, $g_{\rm i}$ and $\mu_{\rm i}$ are the degeneracy factor and chemical potential of the particles of species i, respectively, V is the volume of the source, and the + sign is for fermions while the − sign is for bosons. By using the relation

$$\frac{\mathrm{d}N}{p_{\rm T}\mathrm{d}p_{\rm T}} = \frac{\mathrm{d}N}{m_{\rm T}\mathrm{d}m_{\rm T}} \quad ,$$

and by assuming that $e^{(E-\mu_{\rm i})/T} \gg 1$, the transverse momentum spectrum of particles can be obtained by integrating over the azimuthal angle φ and the rapidity y, leading to

$$\frac{\mathrm{d}N}{m_{\rm T}\mathrm{d}m_{\rm T}} = \frac{g_{\rm i}V}{(2\pi)^2}m_{\rm T}e^{\mu_{\rm i}/T}K_1\left(\frac{m_{\rm T}}{T}\right) \xrightarrow[m_{\rm T}\gg T]{} V'\sqrt{m_{\rm T}}e^{-m_{\rm T}/T} \quad . \tag{1.10}$$

From Eq. 1.10 emerges that for a stationary thermal particle source with a temperature T, the transverse mass spectrum of the particles follows an exponential distribution with a slope parameter T, and is independent of the particle species. This is known as the $m_{\rm T}$-scaling of the transverse mass spectra, and is observed in pp and small-system collisions at low $\sqrt{s}$, with a slope parameter of $T \sim 167$ MeV.

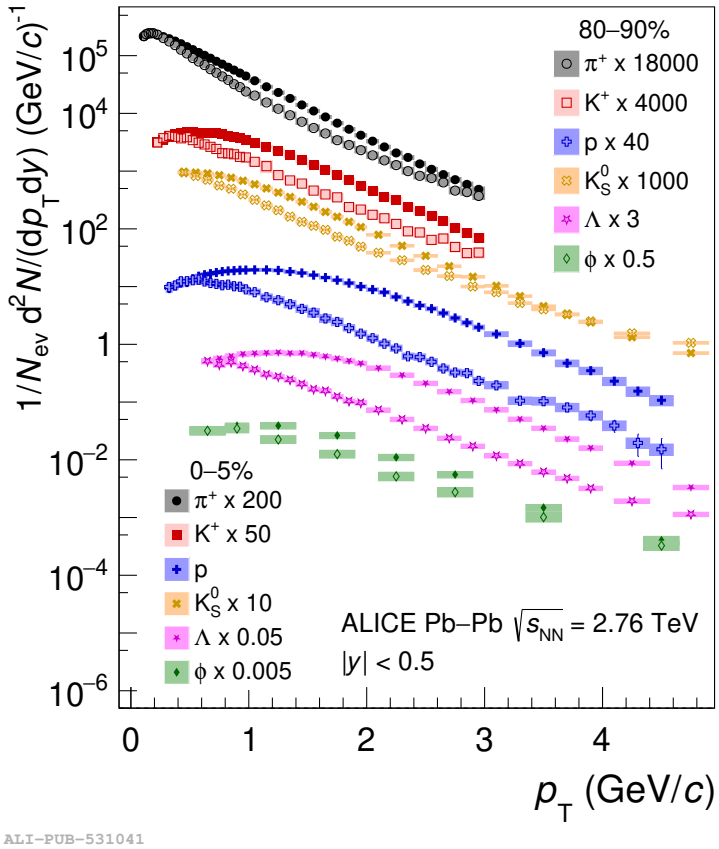


Figure 1.12: Transverse momentum distributions of π^+, K^+, p, K^0_s, Λ, and ϕ mesons for the 0–5% and 80–90% centrality intervals in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV measured by the ALICE experiment. The data points are scaled by various factors for better visibility. Taken from [12].

However, in heavy-ion collisions, this description is not valid, as the QGP, which is a thermalised system of deconfined quarks and gluons, has a thermal pressure. The pressure difference between the QGP and the surrounding vacuum leads to a collective expansion of the fireball, which is called *radial flow*. This causes the particles to be pushed in the transverse direction, causing a modification of the transverse momentum spectra of the particles, which is characterised by a shift of the spectra towards higher p_T values, with a more pronounced effect for heavier particles, as shown in Fig. 1.12. The radial flow can also be defined as a correlation between the velocity of an element of the system and its space-time position, which is superimposed on the random thermal motion of the particles.

The radial flow can be described by using phenomenological models, such as the blast-wave model [29], which assumes that the particles are emitted from a thermal source with a collective velocity field. It relies on the Cooper-Frye prescription [30], which assumes an instantaneous freeze-out of the particles in the radial direction, to calculate the transverse momentum spectra of the particles in terms of the temperature of the source, the collective velocity field, and the transverse flow profile.

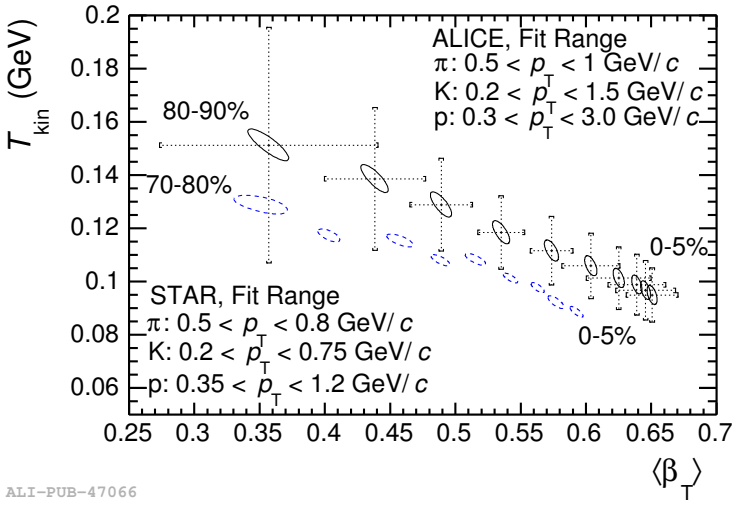


Figure 1.13: Results of blast-wave fits obtained with the ALICE experiment at the LHC, compared to similar fits at RHIC energies, for different centrality intervals. Taken from [28]

Results from SPS, RHIC, and LHC experiments show that the transverse momentum spectra of the particles produced in heavy-ion collisions can be well described by the blast-wave model, with a temperature of the source of around 110–120 MeV, almost independent of $\sqrt{s_{\mathrm{NN}}}$, and a collective velocity field of around 0.5–0.6 c in the most central collisions, with a growing trend with $\sqrt{s_{\mathrm{NN}}}$, as shown in Fig. 1.13. Although the blast-wave model provides a good description of the transverse momentum spectra of the particles produced in heavy-ion collisions, it is important to note that the parameters wich are evaluated with this model are extracted by fitting the m_{T} spectra of the particles, and could not be directly related to the properties of the QGP. Hence, it is important to study whether more sophisticated models, such as hydrodynamics, can provide a more detailed description of the space-time evolution of the system.

Hydrodynamics-based models rely on energy and momentum conservation laws, which need to be expressed in a relativistic form due to the relativistic nature of the system:

$$\partial_\mu T^{\mu\nu} = 0 \quad , \text{ with } \quad T^{\mu\nu} = (\varepsilon + p)u^\mu u^\nu - p g^{\mu\nu} \quad ,$$

where $T^{\mu\nu}$ is the energy-momentum tensor, ε is the energy density, p is the pressure, u^μ is the four-velocity of the fluid element, and $g^{\mu\nu}$ is the metric tensor. An additional condition comes from the conservation of the baryon number, which is expressed as

$$\partial_\mu (n_B u^\mu) = 0 \quad ,$$

where n_B is the baryon density. A further equation is needed to close the system of equations, and is provided by the equation of state of the QGP, which relates the pressure to the energy density and the baryon density. As the fireball is in a non-equilibrium state in the early stages of the collision, the hydrodynamics equations need an initial condition to describe the system. This is usually provided by the

Glauber [31] or T$_{\mathrm{R}}$ENTo [32] models, which describe the collision of the two nuclei as a superposition of nucleon-nucleon collisions, and provide the initial energy density profile of the system. Models based on the Colour Glass Condensate framework such as IP-Glasma [33], or Monte Carlo simulations based on the description of partonic showers such as AMPT [34] can also be used to provide the initial conditions for the hydrodynamics models.
The hydrodynamics calculations provide a good description of the transverse momentum spectra of the particles produced in heavy-ion collisions, and the emerging parameters are similar to those obtained with the blast-wave model. In addition, hydrodynamics calculations and their extensions, such as viscous hydrodynamics, can be used to study other observables, such as the elliptic flow, which is a measure of the anisotropy of the particle emission in the transverse plane.

1.3.4 High-p_{T} hadrons and jet quenching

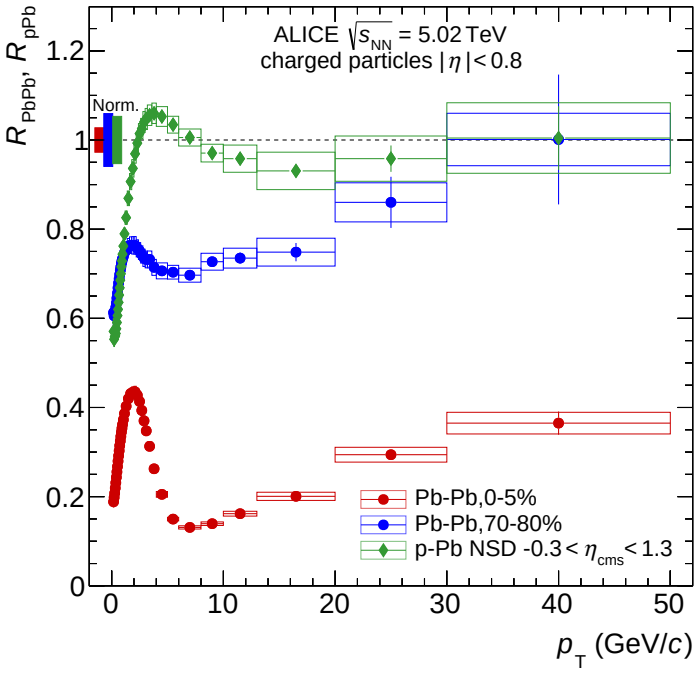


Figure 1.14: Nuclear modification factors measured by ALICE in central $(0-5\%)$ and peripheral $(70-80\%)$ Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 13$ TeV and in p–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 13$ TeV. Taken from [35].

High-p_{T} partons are produced in hard-scattering processes, i.e. those with a high momentum transfer. As such, they are typically produced in the early stages of the collision and experience the whole fireball evolution. They can therefore be used to probe the properties of the earliest stages of the QGP. The Glauber model predicts the production cross-section of hard-scattering processes to scale with the number of binary nucleon-nucleon collisions. To test this prediction, the nuclear

modification factor R_{AA} is defined as the ratio of the p_{T}-differential hadron yield in heavy-ion collisions to the one in proton-proton collisions, scaled by the mean number of binary nucleon-nucleon collisions $\langle N_{\mathrm{coll}} \rangle$:

$$R_{\mathrm{AA}} = \frac{\mathrm{d}N_{\mathrm{AA}}/\mathrm{d}p_{\mathrm{T}}}{\langle N_{\mathrm{coll}} \rangle \ \mathrm{d}N_{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}} \quad .$$

As the p_{T} increases, the production mechanism becomes harder, and it is expected that the nuclear modification factor approaches unity, namely *binary scaling*. However, experimental results [35] show that the R_{AA} measured in the most central heavy-ion collisions exhibits a suppression at low p_{T}, a peak at intermediate p_{T} (~ 2 GeV/c), and a significant suppression by a factor of ~ 5 at high p_{T}, as shown in Fig. 1.14. The suppression at low p_{T} and the peak at intermediate p_{T} are attributed to initial-state effects, such as the modification of parton distribution functions in the nucleus [36], and the Cronin effect [37]. These effects can be studied in small colliding systems [38], where the production of a QGP is not expected, or through the measurement of the R_{AA} for colour-neutral particles, such as photons [39], which are not affected by the presence of a deconfined medium. The high-p_{T} suppression of R_{AA} increases with collision centrality and is related to the formation of the QGP. Partons traversing the medium lose energy through elastic scatterings with the QGP constituents and via gluon radiation, which is the dominant process for high-energy partons. This energy loss can be described in the BDMPS formalism [40], which assumes that the radiated gluon becomes de-coherent from the emitting parton through multiple soft scatterings with the medium constituents. The energy loss of the partons is quantified as

$$\Delta E = \frac{1}{4} \alpha_s C_{\mathrm{R}} \hat{q} L^2 \quad ,$$

Where α_s is the strong coupling constant, C_{R} is the Casimir factor, which is 3 for gluon-gluon couplings and 4/3 for quark-gluon interactions, $\hat{q}$ is the transport coefficient, and L is the path length of the parton in the medium. The transport coefficient is related to the energy-density of the medium, as $\hat{q} \propto \varepsilon^{3/4}$, so that energy loss measurements can be used to infer the properties of the QGP. Typical energy losses of high-p_{T} partons traversing the whole QGP are of the order of 40 GeV, which is a significant fraction of the parton energy. The energy loss of the partons leads to a suppression of the high-p_{T} hadron spectra, which consequently leads to a suppression of the nuclear modification factor at high p_{T}.

At leading order, partons are produced in back-to-back pairs, forming what is known as di-jets. The parton shower generated by the jet pointing in the direction of the medium is expected to lose more energy than the one pointing in the opposite direction, leading to what is known as *jet quenching*. These interactions can be studied by measuring two-particle azimuthal distribution correlations of the produced hadrons. For each event, the particle with the highest p_{T} above a certain threshold is selected, and the azimuthal angle of other high-p_{T} particles is measured with respect to the first one. For proton-proton and proton-nucleus collisions, the azimuthal distribution presents two peaks: the *near-side peak*, which is found at $\phi = 0$ and is due to the hadrons produced in the same jet as the trigger particle,

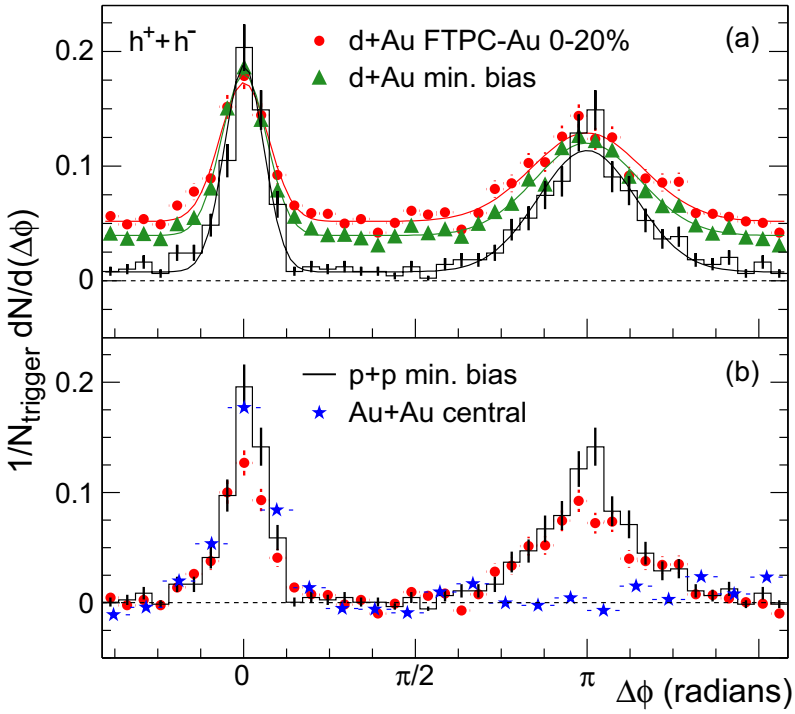


Figure 1.15: Top: Two-particle azimuthal distributions for minimum bias and central d–Au collisions, and for proton-proton collisions at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV measured by the STAR Collaboration. Bottom: Comparison of pedestals-subtracted two-particle azimuthal distributions for central d–Au collisions to those seen in proton-proton and central Au–Au collisions. Taken from [41].

and the *away-side peak*, which is found at $\phi = \pi$ and is due to the hadrons produced in the opposite jet. In heavy-ion collisions, the away-side peak is suppressed due to the energy loss in the QGP, as shown in Fig. 1.15.

In addition, the QGP can also affect the hadronisation process. In small colliding systems, such as pp and p–Pb collisions, hadronisation occurs via fragmentation, where the colour string holding together a quark-antiquark pair breaks, producing another quark-antiquark pair. This process leads to the production of a collimated jet of hadrons with p_{T} lower than the one of the first quark-antiquark pair. When the QGP is produced, another hadronisation mechanism becomes available, as the QGP is a thermalised system: the coalescence (or recombination) mechanism. Low p_{T} partons close in the velocity-space phase space can bind together to form a higher-p_{T} hadron. This enhances the production of baryons, as the p_{T}-distribution of partons inside the QGP follows a rapidly decreasing trend.

1.3.5 Strangenness enhancement

One of the first proposed signatures of the QGP formation is the strangeness enhancement [42], which refers to the observation of an increased production of strange hadrons in heavy-ion collisions compared to proton-proton collisions. At a microscopic level, the enhancement of strange hadrons is a direct consequence of the high temperatures reached in the collisions, which allow for the thermal production of

strange quarks and antiquarks. Strange quarks are not present as valence quarks in the initial state, and can be produced in hard $2 \to 2$ scatterings ($gg \to s\overline{s}$, $q\overline{q} \to s\overline{s}$), or via gluon splitting ($g \to s\overline{s}$) during the evolution of the system. While these processes are dominant for the production of strange hadrons with high p_{T}, at low p_{T} the production of strangeness is dominated by non-perturbative processes. At a macroscopic level, the enhancement of the production of strange hadrons can be explained using the statistical approach of the SHM. The description of the hadrons production in smaller colliding systems, such as $\mathrm{e^+e^-}$ and pp collisions, requires the use of a canonical ensemble, which is used to describe a system that exchanges only energy with a reservoir, and requires the exact (local) conservation of quantum numbers. This reduces the phase space available for particle production [43], leading to a suppression of strange quark production as compared to the grand canonical ensemble, called *canonical suppression*.

The strangeness enhancement in heavy-ion collisions has been observed in several experiments, such as the STAR experiment at RHIC [44] and the ALICE experiment at the LHC [45]. In addition, an enhancement of strange hadrons has been observed in small systems, such as pp [46] and p-Pb [47, 48] collisions, in high-multiplicity events, and follows a hierarchy determined by the hadron strangeness. This observation is intriguing, as QGP production is not expected in such systems. Furthermore, several effects such as azimuthal correlations and mass-dependent hardening of p_{T} distributions, which are typically associated with QGP formation in nuclear collisions, have been observed in these systems. Could small droplets of QGP be formed in these collisions? Could other mechanisms explain the observed effects? These are still open questions in the field of high-energy nuclear physics, and they are the subjects of ongoing research.

Figure 1.16 shows the p_{T}-integrated yield ratios of strange ($\mathrm{K^0_s}$, Λ) and multi-strange ($\Xi^{\pm}, \Omega^{\pm}$) hadrons to pions ($\pi^+ + \pi^-$) as a function of the charged particle multiplicity density, $\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle$, measured in $|y| < 0.5$ in different collision systems (pp, p–Pb and Pb–Pb) using the ALICE experiment at the LHC [46]. The data show a clear enhancement of strange to non-strange hadron production with increasing multiplicity. At the highest multiplicities, the yield ratios saturate at values that are compatible with what is measured in Pb-Pb collisions, where QGP is formed.

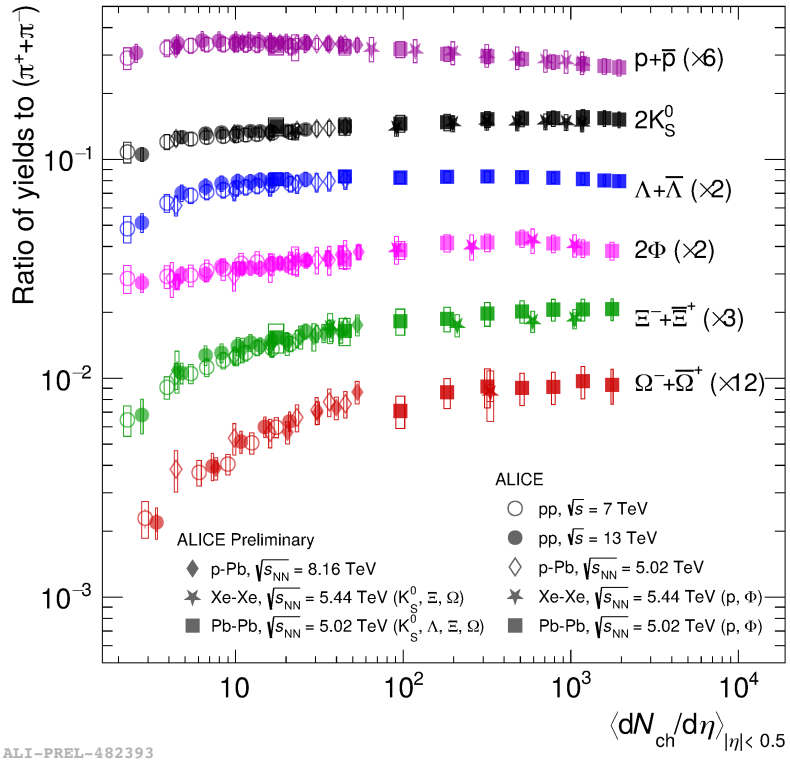


Figure 1.16: p_{T}–integrated yield ratios of strange ($\mathrm{K}^0_{\mathrm{s}}$, Λ) and multi-strange ($\Xi^{\pm}, \Omega^{\pm}$) hadrons to pions ($\pi^+ + \pi^-$) as a function of $\langle \mathrm{d}N_{\mathrm{ch}}/\mathrm{d}\eta \rangle$ measured in pp, p–Pb and Pb–Pb collisions at midrapidity($|y| < 0.5$). Taken from [49].

Chapter 2

Open heavy-flavour production in proton-proton collisions

PEN HEAVY-FLAVOUR hadrons, composed of a heavy quark (charm or beauty) along with lighter quarks, are exclusively formed in high-momentum transfer processes due to the large masses of approximately 1.3 GeV/c^2 and 4.2 GeV/c^2 for charm and beauty quarks, respectively. As a result, they are created in the early stages of the collision, and their production cross-section in the partonic interaction can be evaluated perturbatively using QCD. Studying the production of open heavy-flavour hadrons in proton-proton (pp) collisions not only provides a crucial test of the perturbative QCD framework, but also allows to set constraints on models. Furthermore, measurements in proton-proton collisions, where the production of a deconfined medium is not expected due to the lower energy densities reached, are necessary ingredients for the study of heavy-ion collisions, where the properties of the QGP can be investigated.

2.1 Factorisation theorems

The production of open heavy-flavour hadrons in proton-proton collisions can be described using the factorisation theorems [50], which allow for the separation of short-distance, perturbative behaviour from long-distance, non-perturbative phenomena. The total production cross-section can be expressed as

$$\sigma_{\mathrm{pp}}^{\mathrm{H}} = \sum_{\mathrm{a,b=g,q,\overline{q}}} \int \mathrm{d}x_1 \mathrm{d}x_2 f_{\mathrm{a/A}}(x_1, \mu_F^2) f_{\mathrm{b/B}}(x_2, \mu_F^2) \hat{\sigma}_{\mathrm{ab\to c}}(x_1, x_2, \mu_F^2, \mu_R^2) D_{\mathrm{c\to H}}(z, \mu_F^2) \quad ,$$

i.e., the convolution of: i. the Parton Distribution Functions (PDFs) $f_{\mathrm{a/A}}(x_1, \mu_F^2)$ and $f_{\mathrm{b/B}}(x_2, \mu_F^2)$, describing the initial-state probability of finding a parton a in the proton A carrying a fraction x_1 of the proton's momentum, and a parton b in the proton B carrying a fraction x_2 of its momentum, respectively; ii. the hard partonic scattering cross-section $\hat{\sigma}_{\mathrm{ab\to c}}(x_1, x_2, \mu_F^2, \mu_R^2)$, defining the probability of producing the final state c from the collision of partons a and b; and iii. the Fragmentation Functions (FFs) $D_{\mathrm{c\to H}}(z, \mu_F^2)$, which describe the probability of a parton of type

c fragmenting into a heavy-flavour hadron H with a momentum fraction z. While the PDFs and FFs are non-perturbative quantities, parametrised from experimental data and regarded as universal across different processes, the hard partonic scattering cross-section can be perturbatively calculated using QCD, but needs specific evaluations for each process.

Factorisation theorems have been widely used to describe the production of open heavy-flavour hadrons in proton-proton collisions, and have proven to be successful in modeling experimental data. Figure 2.1 shows the production cross-section of prompt and non-prompt D^0-mesons in proton-proton collisions at $\sqrt{s} = 13$ TeV measured at midrapidity ($|y| < 0.5$) as a function of the transverse momentum by the ALICE experiment [51], compared to FONLL perturbative QCD predictions [52]. The term *prompt* refers to charm-hadrons directly produced in the hadronisation of a charm quark or through the strong decay of a directly produced excited charm-hadron or charmonium state, while *non-prompt* charm hadrons are produced in the decay of a hadron containing a beauty quark. The FONLL predictions are in good agreement with the non-prompt D^0-meson production cross-section, whereas the prompt contribution lies at the upper edge of the theoretical uncertainty band, albeit being described within the uncertainties. Similar trends are observed in the production of other open heavy-flavour hadrons across different experimental facilities, such as the Tevatron, RHIC, and LHC.

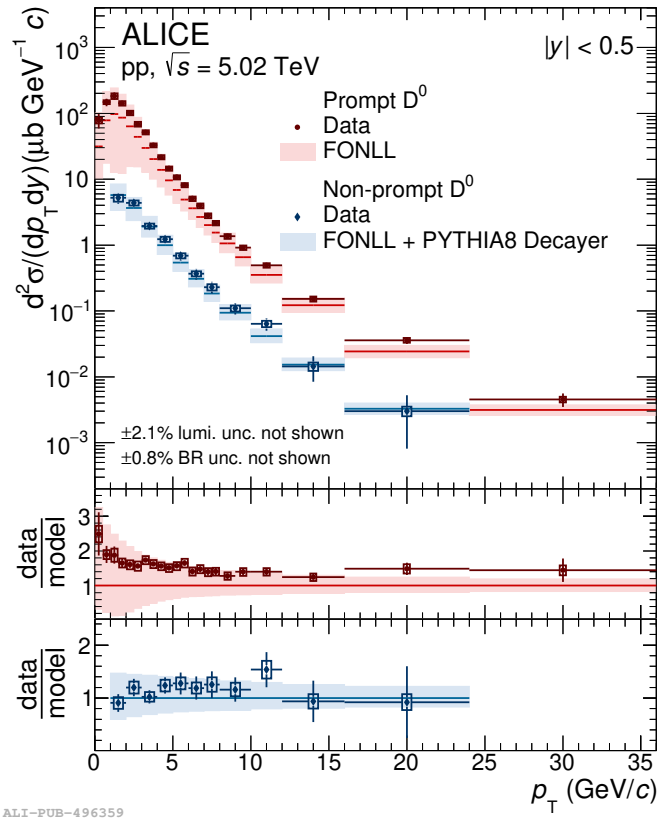


Figure 2.1: p_{T}-differential production cross-section of prompt and non-prompt D^0-mesons [51] compared to predictions obtained with FONLL calculations [52] combined with PYTHIA 8 [53] for the $\mathrm{H_b} \rightarrow \mathrm{D}^0 + \mathrm{X}$ decay kinematics.

2.1.1 Parton Distribution Functions

Deep inelastic scattering

The PDFs are non-perturbative quantities describing the probability of finding a parton carrying a fraction x of the proton's momentum in the initial state of a process. The first experimental evidence revealing the partonic structure of the proton emerged from deep inelastic scattering experiments carried out at the Stanford Linear Accelerator Center (SLAC) in the 1960s [54], where an electron was scattered off a proton, and the transferred momentum q was measured. The cross-section for deep inelastic scattering can be defined in terms of the Lorentz invariant variables $Q^2 = -q^2$ and $x = \frac{Q^2}{2P\cdot q}$, yielding

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x\mathrm{d}Q^2} = \frac{4\pi\alpha^2}{xQ^4}\left[(1-y)\,F_2(x,Q^2) - xy^2F_1(x,Q^2)\right] \quad ,$$

where $y = Q^2/(sx)$, $s = (P+p_\mathrm{e})^2$ denotes the centre-of-mass energy of the electron-proton system, and the structure functions $F_1(x,Q^2)$ and $F_2(x,Q^2)$ represent an extension of the form factors for elastic scattering. The first measurements of high-energy inclusive inelastic scattering experiments were performed using a 20 GeV linear accelerator at SLAC, and showed that the structure functions $F_1(x,Q^2)$ and $F_2(x,Q^2)$ were independent of Q^2 at fixed x within the studied $1 < Q^2 < 10\ \mathrm{GeV}/c^2$ range. This was in contrast with the behavior observed for the proton elastic form factors, where a decrease of two orders of magnitude was observed within the same Q^2 interval. This independence of the structure functions from Q^2 in deep inelastic scatterings was predicted by Bjorken in 1968 for $Q^2 \to \infty$ [55], and is known as *Bjorken scaling*. A physical interpretation of this phenomenon arrived just one year later, in 1969, with Feynman's parton model [56], which described the interaction in terms of elastic scattering of the probe off a point-like constituent (parton) within the proton. This model explains the scale-invariance property of the proton structure functions, as the scattering centres are assumed to be structure-less. In this picture, the Bjorken variable x acquires a new interpretation as the fraction of the proton momentum carried by the struck parton. The parton model also offers a straightforward definition of the structure functions in terms of the parton distribution functions $f_\mathrm{a}(x)$:

$$F_2(x,Q^2) = \sum_\mathrm{a} e_\mathrm{a}^2 x f_\mathrm{a}(x) \quad ,$$

where the sum is over partons with electric charge e_a, and f_a are unknown, but universal functions for a given hadron, describing the probability of finding a parton of type a with a fraction x of the proton's momentum.

To explore the spin properties of the partons, the structure functions F_1 and F_2 were studied at different centre-of-mass energies. By investigating the relationship between the two structure functions, it was established that the partons have spin 1/2, as the Callan-Gross relation [57], which holds true for point-like Dirac particles, was found to be satisfied:

$$F_2(x,Q^2) = 2xF_1(x,Q^2) \quad .$$

In the next years, it became clear that additional constituents within the proton carry momentum but lack electric or weak charge, as the so-called momentum sum rule was not saturated by the measured PDFs in electron and neutrino scatterings. This missing momentum was attributed to gluons, which were discovered in the 1970s and are the field quanta of the strong force.

Bjorken scaling violation

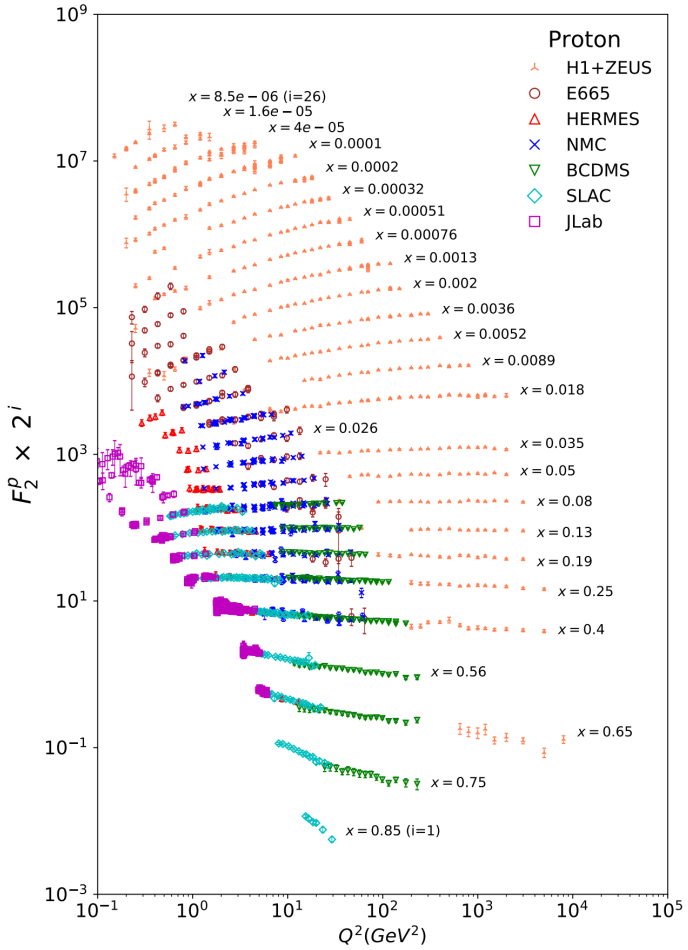


Figure 2.2: The proton structure function F_2^p measured in electromagnetic scattering of electrons and positrons on protons, and for electrons/positrons and muons on a fixed target [4].

By the late 1970s, measurements of the structure functions at larger Q^2 values taken at CERN and DESY revealed that Bjorken scaling was violated, i.e., the structure functions were not Q^2 independent. Figure 2.2 shows measurements of the proton structure functions $F_2(x, Q^2)$ as a function of Q^2 for various values of x taken from different experiments [4]. It is clear from the plot that structure functions present an increasing trend as a function of Q^2 at low x, and a decreasing trend as a function of Q^2 at high x.

The parton model fails to explain this behaviour, as it relies on the assumption that the transferred energy is sufficiently large to neglect the proton and its

constituents' masses, as well as the interactions among partons. In particular, the partons' transverse momentum with respect to the proton momentum is neglected. The key to understanding Bjorken scaling violation comes from QCD and the realization that the parton's transverse momentum is not necessarily restricted to be small. A quark, for instance, can emit a gluon and acquire large transverse momentum k_T with a probability proportional to $\alpha_s \mathrm{d}k_T/k_T^2$ at large k_T. The integral extends up to the kinematic limit $k_T \sim Q^2$, giving rise to contributions proportional to $\alpha_s \log Q^2$, which break scaling. The evolution of PDFs with Q^2 from a parametrisation at a given Q_0^2 can be perturbatively described using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [58, 59, 60], which require the introduction of a new arbitrary scale, at which the factorisation of the non-perturbative processes happens: the factorisation scale μ_F. There exists a wide range of PDF parametrisations, such as the NNPDF [61], CTEQ [62], and MMHT [63], which are determined from global fits to a wide range of experimental data, including deep inelastic scattering, Drell-Yan, and jet production.

2.1.2 Partonic cross-section

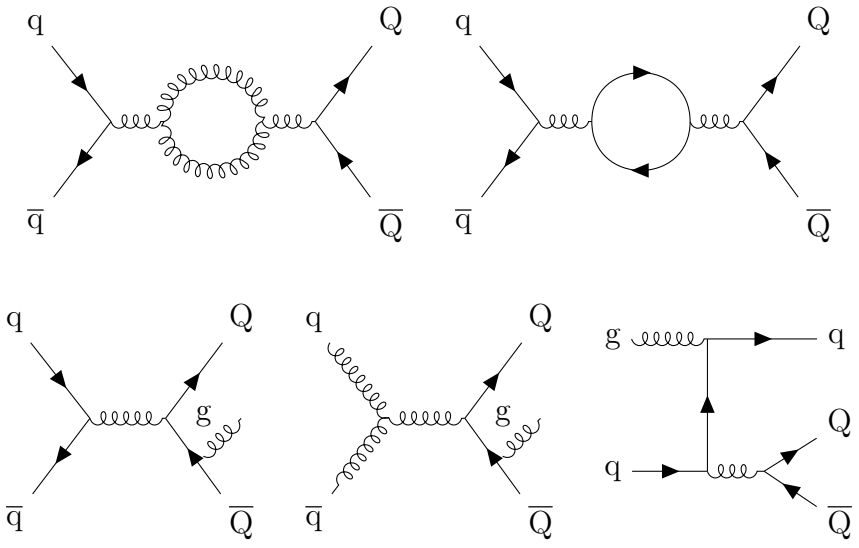


Figure 2.3: Feynman diagrams contributing to the first order corrections of the heavy-flavour production cross-section calculations.

Because of their large masses, heavy quarks can only be produced through hard-scattering processes, characterised by momentum transfers of the order of $Q^2 \geq 4m_{\mathrm{b,c}}^2$. In this regime, the strong coupling constant is significantly smaller than unity, allowing for the perturbative calculation of the heavy quark production cross-section from partonic scattering using QCD. While predictions at next-to-next-to-next-to-leading order ($\mathrm{N^3LO}$) are available for certain processes, such as Higgs production [64, 65], the current state-of-the-art calculations for heavy quark production are at next-to-leading order (NLO) with all-order resummation to next-to-leading logarithmic (NLL) accuracy in the limit where the p_{T} of a heavy quark

is much larger than its mass [52]. The contributions arising at the NLO include 1-loop virtual corrections to the Born process and real emission of a gluon or a quark-antiquark pair, and are depicted in Fig 2.3.

2.1.3 Fragmentation Functions

Quarks and gluons produced in hard-scattering processes ultimately give rise to colourless observable hadrons. The associated process, known as hadronisation, is non-perturbative and is described by the Fragmentation Functions (FFs) $D_{\mathrm{c}\to\mathrm{H}}(z, \mu_F^2)$, which provide the probability of a parton of type c fragmenting into a hadron H with a momentum fraction z. FFs are typically determined from experimental data, usually by analyzing the final-state hadrons produced in electron-positron collisions where the initial momenta are well-known. These FFs are then applied in the evaluation of cross-sections in other colliding systems, assuming that the relevant hadronisation processes are "universal", i.e., independent of the collision energy and system. Many FFs have been determined from global fits to data, such as the NNFF1.1h [66], DSS [67], and KKP [68] parametrisations, and typically differ in the data sets used for the fit, the treatment of the data, and the functional form of the FFs. Similarly to PDFs, FFs also evolve with the energy scale of the interaction, and this evolution is described perturbatively by the DGLAP equations.

2.2 Hadronisation: from macroscopic to microscopic descriptions

Despite being a very powerful tool for describing heavy-flavour hadron production, the factorisation theorem approach is not the only method used to describe the production of hadrons in high-energy collisions. Over the years, microscopic models have been developed to describe the hadronisation process and are typically implemented in Monte Carlo event generators. The standard approach for describing complex event topologies begins with a matrix-element calculation for the production of a few well-separated partons, followed by the application of a parton shower.

The *parton shower* provides an approximate perturbative treatment of QCD dynamics, which is then combined with a non-perturbative model for the hadronisation process at a certain infrared cut-off scale, typically taken to be of the order of 1 GeV. The basic idea of the parton shower relies on the Sudakov form factor [69], which expresses the probability of a parton not radiating another parton in a given phase space region. If a parton does radiate, the newly-produced parton becomes the source of a new cascade, continuing until the parton shower terminates at the hadronisation scale. At this point, partons are allowed to fragment into hadrons through hadronisation models.

Although several hadronisation models have been developed, each implementing a different approach to the description of this non-perturbative process, a common feature is the hypothesis of local parton-hadron duality [70], which states that the flow of momentum and quantum numbers at the hadron level tends to follow the flow established at the parton level.

2.2.1 Independent fragmentation

The description of the hadronisation process using FFs relies on the assumption of *independent fragmentation*, i.e., the probability of a parton fragmenting into a hadron is considered independent of the other partons produced in the same collision. In the original scheme proposed by Field and Feynman [71], the fragmenting quark combines with an antiquark from a $q\bar{q}$ pair produced from the vacuum to create a meson with energy fraction z. The remaining quark, with energy fraction $(1 - z)$, fragments in the same way, continuing until an energy cut-off is reached. The distribution of z is the fragmentation function. The assumption of independent fragmentation is valid in the case of low-multiplicity e^+e^- collision events, where the number of produced partons is small and no hadronic remnants are present.

2.2.2 String model

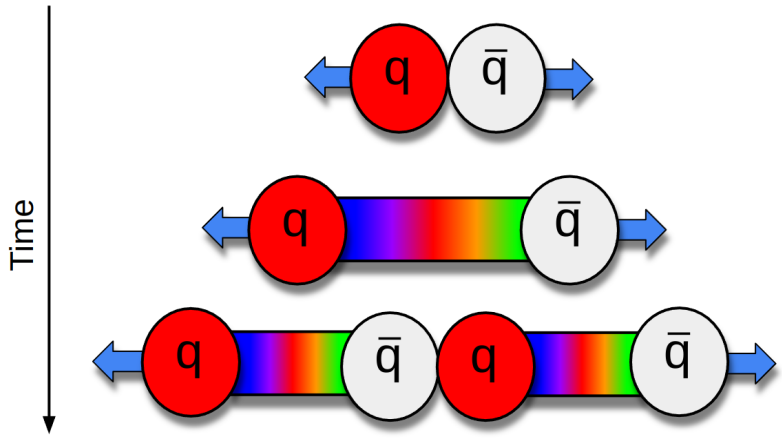


Figure 2.4: Schematic representation of the Lund string model hadronisation process.

The most widely used model for the description of the hadronisation process is the Lund string model [72], employed in the PYTHIA event generator [73]. In this model, the strong force between quarks and gluons is modelled in terms of a colour string with energy given by the Cornell potential [74],

$$V(r) = -\frac{A(r)}{r} + \kappa r \quad .$$

Since the linear term is dominant, the Cornell potential is typically approximated as $V(r) = \kappa r$, with $\kappa \sim 1$ GeV/fm. This implies that a constant force is exerted between the quarks, leading to a linear increase in the potential energy with the distance between the quarks. When back-to-back quarks are produced in a hard-scattering process, they move apart, causing the string to stretch and accumulate energy. When the energy stored in the string becomes large enough, it becomes energetically convenient to materialise a quark-antiquark pair from the colour flux

tube, making the string break. The probability of string breaking is given by

$$P \propto \exp\left(-\frac{\pi m_{\mathrm{T,q}}^2}{\kappa}\right) = \exp\left(-\frac{\pi m_{\mathrm{q}}^2}{\kappa}\right) \exp\left(-\frac{\pi p_{\mathrm{T,q}}^2}{\kappa}\right) \quad .$$

The mass dependence of this equation leads to a gaussian suppression factor that limits the probability for the colour field to produce quark-antiquark pairs for heavier flavours. The $\mathrm{s\overline{s}}$ and $\mathrm{c\overline{c}}$ pair production probability can be estimated with respect to that for $\mathrm{u\overline{u}}$ and $\mathrm{d\overline{d}}$ pairs [75], yielding

$$\mathrm{u : d : s : c} \sim 1 : 1 : 1/3 : 10^{-11} \quad .$$

The soft production of $\mathrm{c\overline{c}}$ results significantly suppressed, by a factor of 10^{-11}, compared to $\mathrm{u\overline{u}}$ pairs. Such quarks are therefore almost exclusively produced in hard-scattering processes, quantitatively confirming the qualitative discussion in Sec. 2.1.2 on the feasibility of using perturbative QCD to describe the production of heavy-flavour quarks.

Non-independent fragmentation and multiple parton interactions

The fragmentation model described above shares many similarities with the independent fragmentation model discussed in Sec. 2.2.1, with the main difference being the pictorial description of interactions with a colour string connecting the quarks. However, the gluon radiation has been neglected until now. A significant difference emerges once the gluon bremstrahlung is considered [76].

In this framework, gluons are represented as carrying both a colour and an anticolour charge. Therefore, in a three-parton system ($\mathrm{q\overline{q}g}$), the quark is colour-connected to the anticolour index of the gluon, and the colour index of the gluon is connected to the antiquark. The interaction between the quark-antiquark pair is suppressed by a factor $1/\mathrm{N_c}$, where $\mathrm{N_c}$ is the number of colours, and can be neglected in a leading-colour approximation ($\mathrm{N_c} \to \infty$). The presence of the gluon produces a corner, or "kink", on the string. The Lorentz boost of a string causes the hadrons it forms through its breaking to go preferentially in the direction of its motion. Most hadrons that the q-g string segment produces will go between the quark and the gluon, while hadrons from the $\overline{\mathrm{q}}$-g string will go between the gluon and the antiquark. Only very few hadrons will go between the quark and the antiquark. This behaviour leads to an angular distribution of hadrons in $\mathrm{e^+e^-}$ three-jet final states which differs from that predicted by independent fragmentation and is found to be in better agreement with experimental results.

At leading-colour approximation, in the partonic final states each quark is colour-connected to a single other parton in the event. Since gluons carry both a colour and an anticolour charge, they are connected to two other partons. With this aprroximation, the Lund model provides a good description of measurements in $\mathrm{e^+e^-}$ collisions, where a parton-rich environment is not produced. However, recent results on the production of heavy-flavour baryons in proton-proton and proton-lead collisions at the LHC [77, **?**] show that a description of the hadronisation process based on independent fragmentation is not sufficient to describe the data, as it significantly underestimates the baryon production.

In hadronic collisions, one must consider that to achieve a comprehensive description of a given process, a description of coloured initial-state partons and their associated remnants should be taken into account, as they hadronise and may potentially interact with each other. Furthermore, new insights on the underlying event and soft-physics processes occurring in a hadronic collision suggest that such events are dominated by Multiple Parton Interactions (MPI), where two or more distinct hard-parton interactions occur in a single hadron-hadron collision. Phase-space overlaps among the partons produced in the MPIs become more likely as the number of partons in the collision increases, and their non-independent hadronisation should be considered for a more complete description of the process.

New models based on colour-reconnection beyond the leading-colour approximation [78] have been developed, where the leading-colour connections produced in the partonic showers are rearranged to form new subleading topologies that would have been present in a full-colour treatment. In addition, new colour junctions allow for an enhanced production of baryons, to account for the observed enhancement in the production of baryons in proton-proton and proton-lead collisions.

2.2.3 Cluster model

A different approach to the description of the hadronisation process is the cluster model [79], which is implemented in the HERWIG event generator [80]. It is based on the *preconfinement* of colour [81], which arises from the observation that by following the colour structure of the parton shower and studying the colour-singlet pairs of colour-connected quark-antiquark states, one finds that they tend to end up close in phase space. This suggests that quarks and gluons produced in this evolution become organised in clusters of colour singlets with finite masses, i.e., the mass distribution of these clusters is independent of the hard scattering process and its centre-of-mass energy. The confinement can then convert these singlets of small mass into hadrons.

The first step of the cluster hadronization model is to non-perturbatively split the gluons left at the end of the parton shower into quark-antiquark pairs. Each gluon is allowed to decay into any of the accessible quark flavours with a probability given by the available phase space for the decay.

Then, the hadronisation process takes place. The key idea is that because the cluster mass spectrum is both universal and steepling falling at high masses, the clusters can be regarded as highly excited hadron resonances and decayed, according to phase space, into the observed hadrons. Before the actual cluster decays, a few heavier clusters are split into lighter clusters (*cluster fission*), for a more reasonable agreement with experimental results. A cluster is split into two clusters if the mass, M, is such that

$$M^{\mathrm{Cl_{pow}}} \geq {\mathrm{Cl_{max}}}^{\mathrm{Cl_{pow}}} + (m_1 + m_2)^{\mathrm{Cl_{pow}}} \quad ,$$

where $\mathrm{Cl_{max}}$ and $\mathrm{Cl_{pow}}$ are parameters of the model, and $m_{1,2}$ are the masses of the constituent partons of the cluster. For clusters that need to be split, a $\mathrm{q\overline{q}}$ pair is produced from the vacuum. Only up, down and strange quarks are chosen with probabilities given by other model parameters. Once a $\mathrm{q}, \overline{\mathrm{q}}$ pair is produced, the

cluster is decayed into two new clusters with one of the original partons in each cluster.

Finally, the cluster is decayed into a pair of hadrons. For a cluster of a given flavour $(q_1, \overline{q}_2)$, a quark-antiquark or diquark-antidiquark pair $(q\overline{q})$ is extracted from the vacuum and a pair of hadrons with flavours $(q_1, \overline{q})$ and $(q\overline{q}_2)$ is formed. The hadrons are selected from all the possible hadrons with the appropriate flavour based on the available phase space, spin and flavour of the hadrons. As a consequence, heavier hadrons are suppressed, leading to a natural description of the baryon and strangeness suppression. The cluster model was found to describe data reasonably well, with far fewer parameters than the string model [82].

2.2.4 Coalescence model

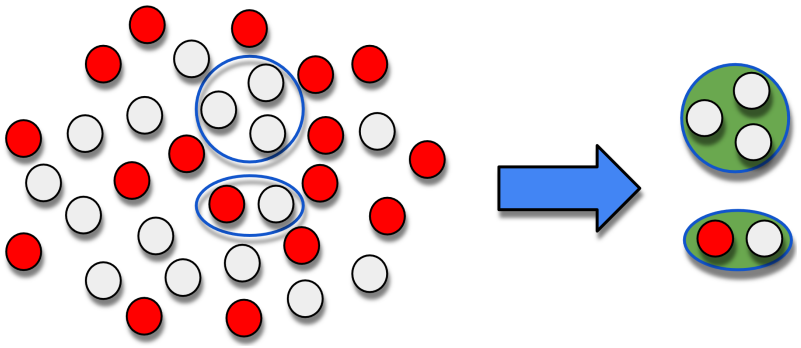


Figure 2.5: Representation of hadron production via recombination.

As outlined in Sec. 1.3.4, the hadronisation process can be modified in the presence of a strongly-interacting deconfined medium. A novel mechanism for the production of hadrons, the recombination (also called coalescence), can take place in a thermalised system such as the QGP. When recombination occurs, low p_{T} partons close in the velocity-space phase space can bind together to form a higher-p_{T} hadron. This could lead to a depletion of the low-p_{T} hadron yield, and an enhancement of the intermediate-p_{T} hadron yield. This process modifies the production of hadrons in heavy-ion collisions, where the large amount of produced partons increases the probability of recombination. For this reason, the description of fragmentation functions derived from e^+e^- collisions, where coalescence does not play a role, becomes inadequate.

Only models implementing the coalescence mechanism can effectively describe the production of hadrons in heavy-ion collisions. Intriguingly, models implementing recombination that successfully describe the production of hadrons in these collisions, fail in doing so when this process is de-activated, as shown in Fig. 2.6.

New experimental measurements in pp and p–Pb collisions at the LHC [46, 84, 85, 86, 87, 88] provided a wealth of results suggesting that certain phenomena that are observed in nuclear collisions, such as strangeness enhancement and flow, may also occur in smaller collision systems. These effects are typically related to the presence of collective behaviours, thus raising the question of whether the coalescence hadronisation mechanism could also play a role in small collision systems. A hint

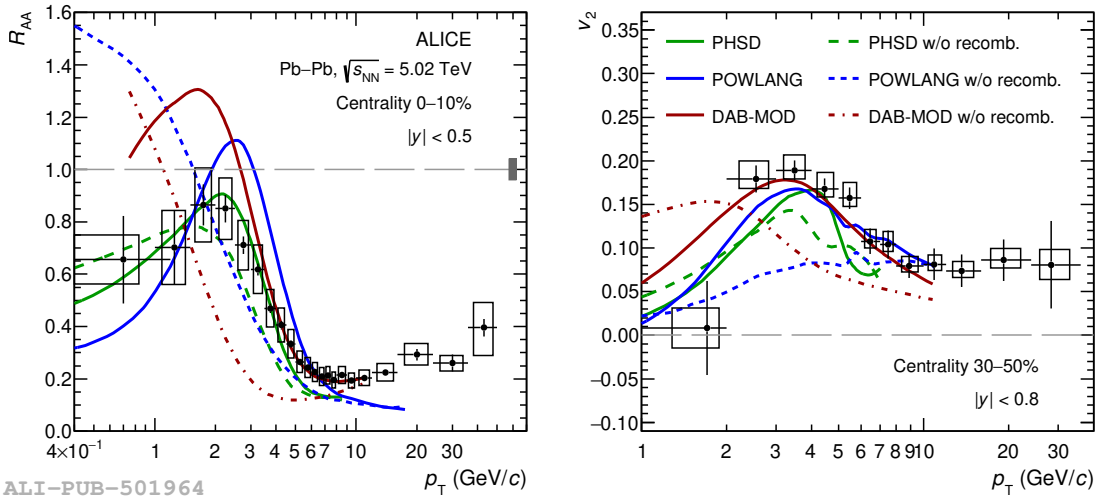


Figure 2.6: Prompt D-meson R_{AA} in the 0–10% centrality class (left panel) and v_2 in the 30–50% centrality class (right panel) compared with predictions obtained with and without including hadronisation via recombination. Taken from [83].

that this could be the case comes from the observation that the production of charm baryons in pp and p–Pb collisions [77, 85] is significantly larger than expected from models based on a fragmentation tuned on $\mathrm{e^+e^-}$ collisions.

Various models have been developed in order to describe the observed charm-baryon enhancement in proton-proton and proton-lead collisions with the assumption of a recombination mechanism. These include the Quark (re-)Combination Mechanism [89] model, where coalescence between a charm quark and equal-velocity light quarks from fragmentation takes place, and thermal weights are applied to account for relative production of scalar and vector mesons. In addition, the assumption of equilibrium allows for a macroscopic description of the system using a hydrodynamic appoach: the Catania coalescence model [90] describes a thermalised system of u, d, s quarks, and gluons, where the charm quark can hadronise via either fragmentation or coalescence with light quarks from the bulk, while, the POWLANG model [91] predicts the formation of a small, deconfined and expanding fireball in proton-proton collisions, where charm quarks are subject to rescattering and hadronization, and can recombine with light quarks as in heavy-ion collisions. Each of these models provides a different hadronization mechanism in proton-proton collisions compared to $\mathrm{e^+e^-}$ ones, and independent hadronization is no longer assumed.

2.2.5 Core-corona model

Core-corona models [92] offer an intermediate approach between the two presented above. In these models, the fireball is divided into a hihg-density core part, treated with hydrodynamics and heavy-ion freeze-out models, and a corona part, described using the Lund string model. This framework is based on the assumption that small QGP droplets may only be produced once the density is high enough. The EPOS

Monte Carlo event generator [93] implements such a picture, and provides a very competitive description of data.

2.2.6 Statistical hadronisation model

The last model presented in this section exploits the idea of a thermalised system, which is by definition in thermal equilibrium. A macroscopic approach to hadronisation, based on a statistical model, can be used to describe the charm-hadron production in hadronic collisions. The principles of thermodynamics can be applied to the hadronisation process, where the hadron yields are determined by the hadron's mass and the temperature and baryon chemical potential of the system. As described in Sec. 1.3.2, this approach has been successfully adopted to describe the light and strange hadron production in heavy-ion collisions and in smaller systems. However, the description of heavy-flavour hadron production and the baryon enhancement in this framework requires the inclusion of a strong feed-down from an augmented set of excited charm- or beauty-baryon states, not included in the PDG [4]. Recent model developments [94, 95] have extended the Statistical Hadronisation Model (SHM) to include a set of 18 Λ_c's, 42 Σ_c's, 62 Ξ_c's, and 34 Ω_c's up to a mass of 3.5 GeV, predicted by the Relativistic Quark Model [96]. Such models provide a good description of the measured Λ_c^+/D^0 production yield ratio and the p_T spectra of D mesons and Λ_c^+ at midrapidity.

2.2.7 Baryon enhancement: Λ_c^+/D^0 ratio

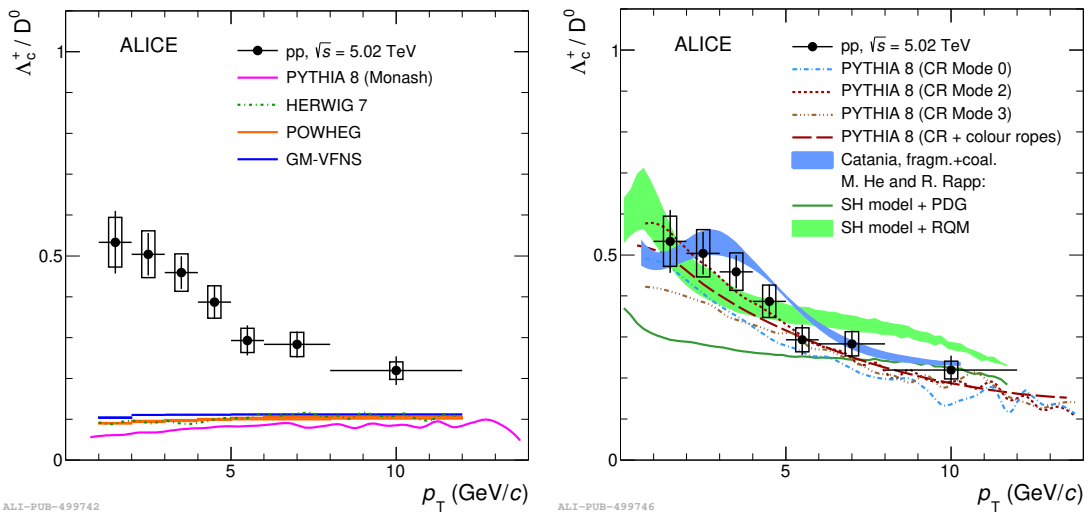


Figure 2.7: Λ_c^+/D^0 production yield ratio measured at $\sqrt{s} = 13$ TeV by the ALICE Collaboration as a function of p_T, compared to theoretical predictions. Figure taken from Ref. [84].

The hadronisation process can be experimentally studied through the measurement of production yield ratios of hadrons. Since the initial state of the collision and the charm production cross section is the same for all charm hadrons, the

measurement of the production yield ratio of different charm hadrons can be expressed, using the factorisation theorem, as the ratio of the fragmentation functions of the hadrons, which describe the hadronisation process. Fig. 2.7 shows the Λ_c^+/D^0 baryon-to-meson production yield ratio measured at $\sqrt{s} = 13$ TeV by the ALICE Collaboration as a function of p_T, compared to theoretical predictions.

In the left panel, the data are compared to i. PYTHIA 8 with Monash tune [97]; ii. HERWIG 7 [98]; iii. POWHEG [99] NLO pQCD calculations, matched with PYTHIA 6 to generate the parton shower; iv. General-Mass Variable-Flavour-Number Scheme [100] (GM-VFNS) NLO pQCD calculation with next-to-leading-log resummation. All these models implement fragmentation processes tuned on results of charm production measurements is e^+e^- collisions, and predict an almost p_T-independent Λ_c^+/D^0 ratio of around 0.1, significantly underestimating the measured values by a factor of about 7 at low p_T.

In the right panel, the measurements are compared to models that include mechanisms to enhance baryon production. They include i. PYTHIA 8 simulations with colour-reconnection beyond the leading-colour approximation [78] (three colour reconnection modes (0,2,3), which apply different constraints on the allowed reconnection are considered); ii. PYTHIA 8 with colour-reconnection plus rope hadronisation [101] where colour charges can act coherently to form a rope, increasing the effective string tension; iii. Catania coalescence model [90], where a QGP is formed in pp collisions and hadronisation occurs through both recombination and fragmentation; iv. SHM [22] where the underlying charm baryon spectrum is either taken from the PDG [4], or augmented to include additional excited baryon states, which have not yet been observed but are predicted by the RQM [96]. These models are capable of describing both the magnitude and the p_T dependence of the Λ_c^+/D^0 ratio, suggesting that the hadronisation process in proton-proton collisions might differ from that in e^+e^- collisions.

The influence of the surrounding environment in the hadronisation of the charm quark could also potentially explain the observed strangeness enhancement in proton-proton collisions when compared to e^+e^- collisions. Enhanced production of charm-strange hadrons may result from the coexistence of numerous strange quarks (which could be thermally produced in an environment with $T \gg m_s$) within the same region as the charm quark, thereby increasing the probability of recombination between them. This phenomenon can be studied through the measurement of the production yield ratio of charm-strange hadrons to non-strange charm hadrons, such as the D_s^+-meson to D^+-meson ratio, which constitutes the primary focus of this Thesis.

Chapter 3

A Large Ion Collider Experiment

ALICE (A Large Ion Collider Experiment) is a general-purpose detector at the CERN LHC [102]. It was conceived and built to focus on the study of QCD and heavy-ion collisions. It is designed to address the physics of strongly interacting matter and the quark-gluon plasma at extreme values of energy density and temperature in nucleus-nucleus collisions. Nonetheless, its physics programme also includes proton-proton and proton-nucleus collisions, which provide reference data for the heavy-ion programme and address a number of specific strong-interaction topics for which ALICE is complementary to the other LHC detectors. During the LHC Long Shutdown 2, between 2019 and 2021, ALICE underwent a major upgrade to improve its capabilities to probe the QGP with heavy-flavour quarks, and to enable completely new measurements of the thermal emission of dielectron pairs [103].

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a two-ring superconducting hadron accelerator and collider based at CERN [104], near Geneva, in the border between Switzerland and France. The LHC is the world's largest circular particle accelerator with a 26.7-kilometer circumference housed in a 3.8-meter-wide tunnel buried 50 to 175 meters underground, capable of accelerating protons up to energies of $\sqrt{s} = 14$ TeV and heavy ions ($^{208}\mathrm{Pb}^{82+}$) up to a centre-of-mass energy per nucleon pair of $\sqrt{s_{\mathrm{NN}}} = 5.5$ TeV.

The LHC, which was used to host the Large Electron-Positron Collider [105] (LEP), is the final stage of a chain of accelerators that provide protons and heavy ions with sufficient energy to be injected into the LHC, illustrated in Fig. 3.1. Protons are firstly collected by stripping electrons from a gaseous hydrogen source with an electric field and then grouped into bunches. The Linear accelerator 4 (Linac 4), the first linear accelerator in the LHC injection chain, accelerates negative hydrogen ions to 160 MeV. The ions are then stripped of their two electrons, and protons are subsequently injected into the Proton Synchrotron Booster (PSB), which accelerates them to 1.4 GeV, and then into the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), reaching 25 GeV and 450 GeV, respectively. The protons are finally injected in the LHC, where they are accelerated up to a maximum energy of

7 TeV. In the LHC, the beams are forced to follow the circular path of the ring by strong magnetic fields (~ 8 T) produced by superconducting dipole magnets. Other magnets are used to steer and focus the beams in the eight different interaction points.

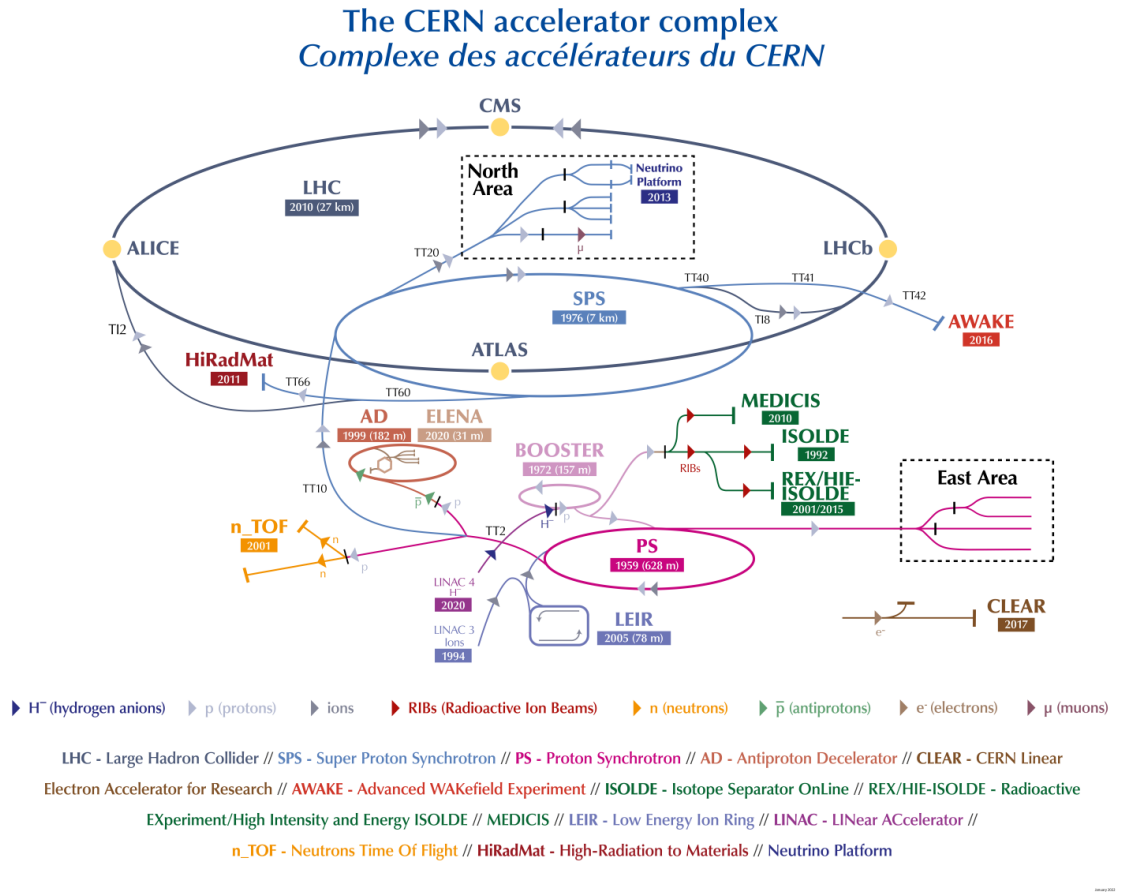


Figure 3.1: The Large Hadron Collider (LHC) accelerator complex at CERN. Figure taken from Ref. [106].

Heavy-ion acceleration follows a similar path. Pb ions (Pb^{25+} – Pb^{28+}) are extracted in the Electron-Cyclotron Resonance (ECR), and are focused and accelerated in a Radio-Frequency Quadrupole (RFQ) using only electric fields. Pb^{28+} ions with momentum of 2.5 keV/c are selected with a spectrometer, and are accelerated to 4.2 MeV/c in the Linear accelerator 3 (Linac 3). A first 0.5 µm copper stripping foil is then used to increase the charge state of the ions to Pb^{53+}. The Low Energy Ion Ring (LEIR) groups the ions into buches and accelerates them to 72 MeV. The PSB and PS then accelerate the ions to 95.4 MeV and 5.09 GeV. After the PS, a copper stripper (~ 1 mm) is used to further ionise the Pb ions to Pb^{82+}, which are then accelerated to 158 GeV in the SPS. The ions are finally injected into the LHC, where they are accelerated to 2.76 TeV per nucleon.

The LHC can collide protons with protons, protons with heavy ions, and heavy ions with heavy ions. The LHC has four main interaction points, where the four main detectors are located: A Large Ion Collider Experiment (ALICE), A Toroidal LHC

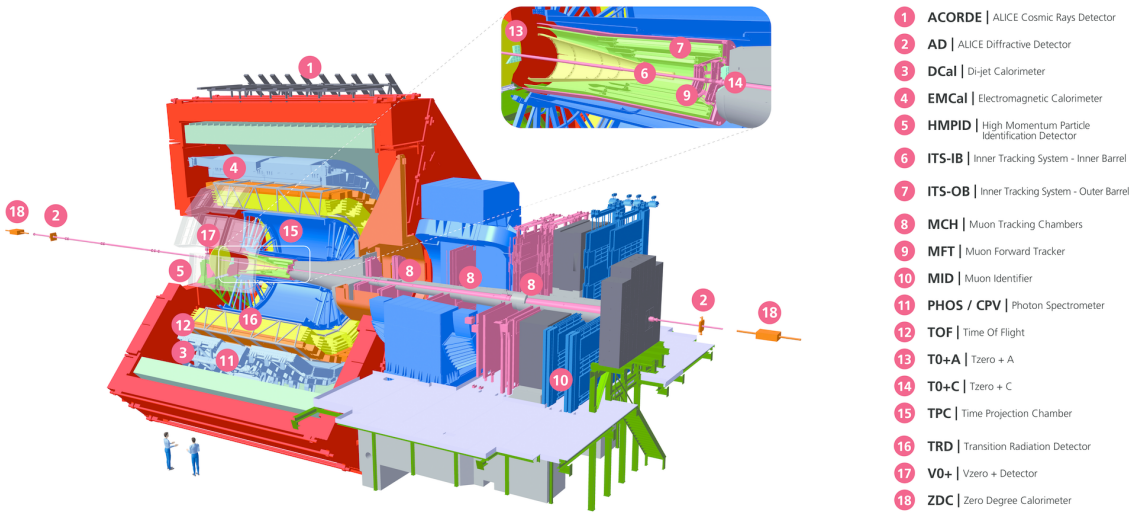


Figure 3.2: The ALICE experimental apparatus. The top right panel shows a zoom of the ITS, T0-A, T0-C and MFT detectors. Figure from ALICE figure repository [49].

ApparatuS (ATLAS), Compact Muon Solenoid (CMS), and LHC-beauty (LHCb). The four experiments were conceived and built to address different physics topics: ATLAS and CMS are general-purpose detectors designed to study the Higgs boson, which they discovered in 2012 [107, 108] and to search for physics beyond the Standard Model; LHCb is dedicated to the measurement of beauty quarks as proxy for CP violation studies, and to the study of matter-antimatter asymmetry. ALICE is the only detector at the LHC that is dedicated to the study of QCD rather than the electroweak sector of the Standard Model. A more detailed description of the ALICE detector is given in the following sections.

3.2 The ALICE experiment

The physics programme of ALICE revolves around the study of the properties of strongly interacting matter at the extreme values of energy density and temperature reached in ultra-relativistic heavy-ion collisions. This unique environment poses several challenges, the most stringent one being the need to carry out measurements in a very high multiplicity environment. Originally, estimates for the charged particle multiplicity density at mid-rapidity in central Pb–Pb collisions ranged from $\mathrm{d}N/\mathrm{d}\eta = 2000$ up to almost $\mathrm{d}N/\mathrm{d}\eta = 8000$. Detectors with high granularity, fast readout capabilities, and high radiation hardness were therefore needed. The ALICE detector was designed to meet these requirements, and to provide a comprehensive set of measurements to study the properties of the QGP. The ALICE apparatus is shown in Fig. 3.2. It is based on a central barrel, covering full azimuth ($0 < \varphi < 2\pi$) and pseudorapidity region $|\eta| < 0.9$, and a forward muon system with a dipole magnet providing a total bending power of $B\rho = 3$ Tm, covering full azimuth and pseudorapidity interval $-4.0 < \eta < -2.5$.

The ALICE coordinate system is a right-handed system with the z-axis pointing

along the beam direction, in the direction away from the muon arm, the y-axis pointing vertically up, and the x-axis pointing horizontally towards the center of the LHC. The nominal interaction point is the origin of the coordinate system. The two sides of the detector along the beam axis are referred to as the C-side, where the muon arm is positioned, and the A-side, where the FV0 is positioned. The polar angle θ is defined with respect to the z-direction, while the azimuthal angle φ increases counter-clockwise starting from the x-axis towards the CMS side.

The central barrel is enclosed in the L3 solenoid, which has an internal length of 12.1 m and a radius of 5.75 m, providing a magnetic field of 0.5 T. The L3 experiment was a detector at the Large Electron-Positron Collider (LEP) at CERN, which was built in the cavern where ALICE is now located, operating from 1989 to 2000. The central barrel detector system is designed for efficient tracking in the high track-density environment of heavy-ion collisions, covering transverse momenta from $\sim$ 100 MeV/c to $\sim$ 100 GeV/c with excellent hadron and electron identification capabilities. It is also capable of the reconstruction of primary and secondary vertices with a resolution $\sim 40\,\mu$m. This allows for precise measurements of ground-state heavy-flavour hadrons, which typically decay at a distance of $\sim 100\,\mu$m from the primary vertex. The L3 magnet hosts the Inner Tracking System (ITS), the Time Projection Chamber (TPC), the Transition Radiation Detector (TRD), the Time-Of-Flight (TOF) detector, the High-Momentum Particle Identification Detector (HMPID), and the Electromagnetic Calorimeter (EMCal). The central barrel also includes forward detectors such as the Muon Forward Tracker (MFT) and the Fast Interaction Trigger (FIT). The muon arm covers the forward pseudorapidity range $-4.0 < \eta < -2.5$ and consists of absorbers, a large dipole magnet, planes of triggering Resistive Plate Chambers (RPC) and tracking multiwire proportional chambers (MWPC).

In the following sections, the main components of the ALICE experiment used for the measurements presented in this Thesis are described in detail, ordered from the innermost layer to the outermos detector, following the path of a particle produced in a hadronic collision at mid-rapidity. The upgrades of the ALICE detectors during the LHC Long Shutdown 2 will be highlighted.

3.2.1 ALICE upgrades overview

The ALICE detector underwent a major upgrade during the LHC Long Shutdown 2, between 2019 and 2021, which renewed the experiment, now referred to as ALICE 2. The upgrade aimed at significantly improving the capabilities of ALICE to probe the QGP with heavy-flavour quarks, and enabling new measurements of the thermal emission of dielectron pairs. To achieve these goals, the ALICE detector was improved by enhancing the tracking capabilities of the central barrel detectors and increasing the readout rate of the detectors to collect larger data samples.

The Inner Tracking System (ITS) was replaced with a new detector, the ITS 2, a thinner and lighter detector with the first layer closer to the interaction point, which improves the pointing resolution by a factor of 3 in the transverse direction and a factor 6 in the longitudinal direction, as shown in Fig 3.3.

The upgrade strategy for the enhanced readout rate is based on the LHC plans

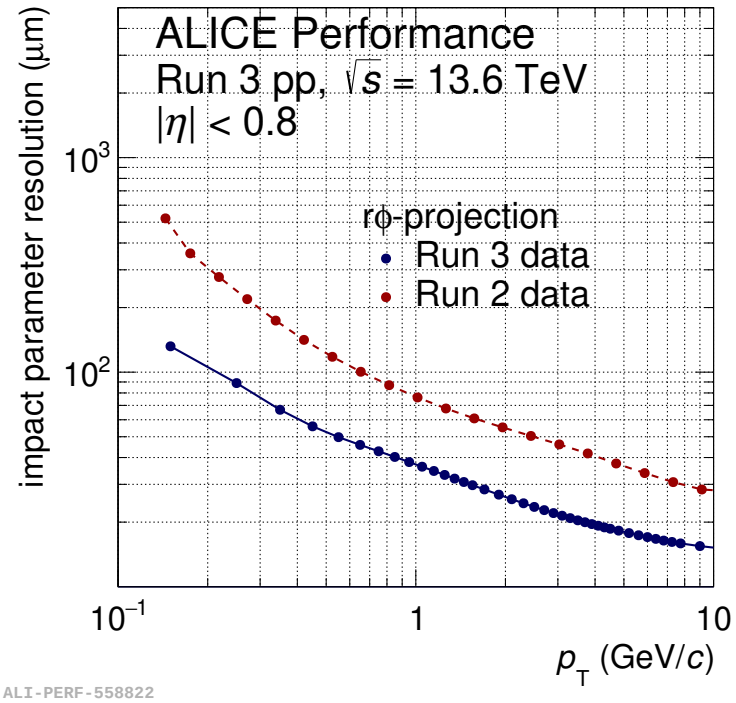


Figure 3.3: Impact parameter resolution in $r\phi$ as a function of p_{T} in pp collisions at $\sqrt{s} = 13.6$ TeV from Run 3 data compared with the same quantity measured in collisions at $\sqrt{s} = 13$ TeV from Run 2 data. Figure taken from ALICE figure repository [49].

to increase the luminosity of Pb–Pb collisions progressively after the LHC Long Shutdown 2, eventually reaching an interaction rate of about 50 kHz (from less than 1 kHz during LHC Run 1 and 2 data-taking periods), i.e. instantaneous luminosity of $\mathcal{L} = 6 \times 10^{27}$ cm^{-2} s^{-1}. At these high interaction rates, each TPC drift time period of $\sim 100\,\mu$s will contain on average 5 Pb–Pb events. It was therefore decided to use a continuous, untriggered readout strategy. With the MWPC used in ALICE 1 for the TPC readout, the ion backflow into the drift region had to be suppressed by active gating, limiting the readout rate to about 700 Hz for Pb–Pb collisions. This is overcome in ALICE 2 by using a readout based on Gas Electron Multiplier [109] (GEM) foils, which reduce the ion backflow and resulting space charge in the TPC to a level that can be corrected for while operating the detector with Pb–Pb interaction rates up to 50 kHz.

To synchronise the continuous data stream across all readout and processing branches, the data stream is divided into time frames (TF) of nominal length of 128 LHC orbits (~ 11 ms). Each TF is subdivided into heartbeat frames (HBF) with a length corresponding to an orbit of $\sim 89.4\,\mu$s. The detector data are time stamped with a precision of an LHC bunch crossing of 25 ns.

A new software framework has been developed to allow for distributed and efficient processing of this unprecedented amount of data. Because of the continuous readout in Run 3, the vertex-to-track association is no longer unambiguous. There-

fore, in place of the analysis data model used in Runs 1 and 2, based on the hierarchical structure of the "event content", the analysis data model for Run 3 is based on a columnar data format, where collisions and tracks are represented as separate tables, connected by an index. This maps naturally on a "flattened" structure-of-arrays data format. Hence, the columnar data format provided by Apache Arrow [110] was chosen. The new Online-Offline ($\mathcal{O}^2$) framework ensures a unified and coherent computing environment from data taking up to analysis. Throughout all stages of the data processing, the timeframe represents the minimal processing unit.

The result of asynchronous reconstruction is the Analysis Object Data (AOD) format, with the best calibration available. The content of an AOD timeframe is kept contiguously in shared memory allowing efficient application of parallel execution and pipelining. AODs are stored as ROOT [111] trees, exploiting their columnar storage format, and are then used to perform physics analyses.

3.2.2 Inner Tracking System

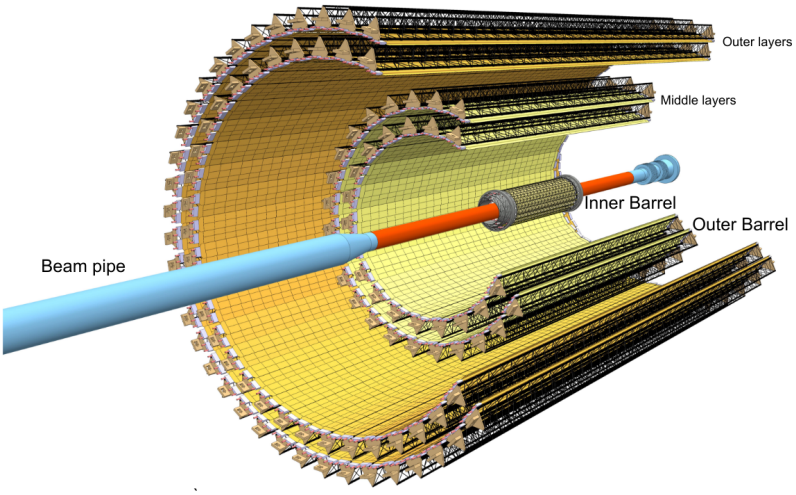


Figure 3.4: Schematic view of the ALICE ITS-2. Figure taken from Ref. [103].

The Inner Tracking System (ITS) is the innermost detector of the ALICE apparatus, and is designed to provide precise tracking and vertexing capabilities in the high-multiplicity environment of heavy-ion collisions. The ITS upgrade is one of the key improvements of the ALICE 2 detector. The ITS 1 was a six-layer silicon detector, with two layers of Silicon Pixel Detectors (SPD), two layers of Silicon Drift Detectors (SDD), and two layers of Silicon Strip Detectors (SSD). The ITS 1 has been replaced by the ITS 2, which is fully constructed with ALPIDE [112] chips, i.e., Monolithic Active Pixel Sensors [113] (MAPS) implemented in a 180 nm CMOS technology for imaging sensors provided by TowerJazz [114].

MAPS are an evolution of the hybrid pixel sensors which have been widely used in the past by several experiments [102, 115, 116, 117]. Hybrid pixel sensors are made of an active layer, typically made of silicon, and a readout layer, connected to the active layer with bump-bonds. On top of being expensive and complex, the bump-bonding operation introduces additional material in the detector, therefore increasing the

material budget of the detector. MAPS, on the other hand, are monolithic sensors, where the active layer and the front-end electronics are integrated into the same silicon wafer. This significantly reduces the material budget of the detector. The ALPIDE chip is a 15 mm×30 mm chip, with a pixel size of 29.24 µm × 26.88 µm, and a total of 512×1024 pixels. Each pixel cell contains a sensing diode, a front-end amplifier, a shaping stage, a discriminator, and a digital section. The digital section includes a multi-event buffer with three hit storage registers and a pixel mask register. At the time of writing this Thesis, the ITS 2 is the largest-scale application of MAPS in a high-energy physics experiment.

In addition, the ITS 2 brought a plethora of improvements: i. a new beryllium beam pipe with an outer radius reduced from 28 mm to 18 mm allows for an innermost detector layer closer to the interaction point, from 39 mm to 22.4 mm; ii. an increased granularity for all layers, which are now silicon pixel detectors with a cell size of 29.24 µm × 26.88 µm; iii. the number of layers for the inner barrel was increased from two to three, raising the total number of layers from six to seven; iv. the material budget was reduced to 0.36% X_0 (1.10% X_0) per layer for the innermost (outer) layers. A schematic view of the ITS 2 is presented in Fig. 3.4, while the main layout parameters are summarised in Table 3.1.

Table 3.1: Main layout parameters of the new ITS2. Taken from Ref. [103].

Layer no.	Average radius (mm)	Stave length (mm)	No. of staves	No. of HICs/ stave	Total no. of chips
0	23	271	12	1	108
1	31	271	16	1	144
2	39	271	20	1	180
3	196	844	24	8	2688
4	245	844	30	8	3360
5	344	1478	42	14	8232
6	393	1478	48	14	9408

3.2.3 Time Projection Chamber

The Time Projection Chamber is the main tracking detector of the ALICE experiment, and is designed to provide precise p_{T} measurements and particle identification via $\mathrm{d}E/\mathrm{d}x$ of charged particles. The TPC is a large cylindrical detector with a length and outer diameter of about 5 m, resulting in a volume of 88 m^3 filled with a gas mixture of Ne-CO_2-N_2 (90-10-5). It covers a symmetric pseudorapidity interval around midrapidity ($|\eta| < 0.9$) at full azimuth. The field cage has a high-voltage (100 kV) electrode in its center, which divides the active volume into two halves. The inner diameter of the central field cage drum is 114 cm, providing the necessary space for the installation of the ITS. Each of the two endplates houses 18 inner and outer readout chambers (IROCs and OROCs), which are arranged into pairs to form 18 equal azimuthal sectors.

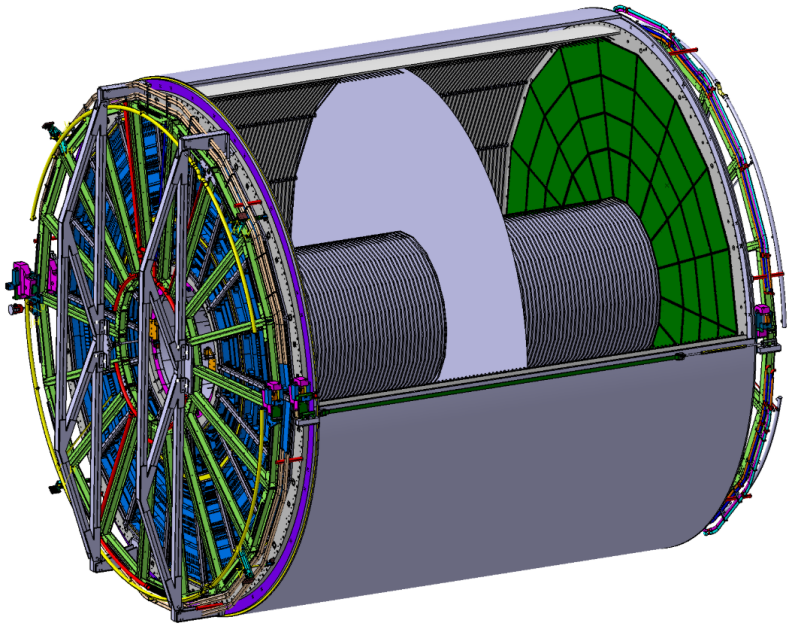


Figure 3.5: Schematic view of the ALICE TPC. Figure taken from Ref. [103].

During Runs 1 and 2, the readout chambers were based on MWPCs, which have to be operated with an active ion gating grid in order to prevent the ions produced in the amplification region from reaching the active drift volume, which could lead to space-charge distortions. The ambitious physics program of the ALICE experiment for Run 3 requires the elimination of the intrinsic trigger rate limitation to about 3 kHz of the original MWPC-based TPC, imposed by the operation of the active ion gating grid.

Operating the TPC at a collision rate of 50 kHz implies that on average five collision events pile up within the TPC readout time window of about $\sim 100\,\mu s$. This excludes triggered operation and defines the need for continuous readout, requiring the exploitation of gas amplification techniques capable of providing sufficient ion blocking without an active gate. At the same time, the readout system must ensure that the dE/dx resolution of the TPC is preserved.

GEMs represent a valid solution to overcome the limitations of MWPCs. Sufficient ion blocking can be achieved by stacking four GEM foils using standard (S, 140 µm) and large (LP, 250 µm) hole pitch in an S-LP-LP-S configuration and by adjusting the gain share among the four layers. The latter optimisation allows for efficiently blocking the ions, most of which are produced in the last amplification step, i.e., in the layer closest to the readout pad. A careful choice of hole patterns is made to avoid the accidental alignment of holes in subsequent layers.

Thanks to this design, the detector is capable of satisfying the requirements for Run 3 operation, with an ion backflow below 2% and a local energy resolution at the ^{55}Fe-peak below 14%.

3.2.4 Time-of-Flight

The TOF detector is a large area detector covering the central pseudo-rapidity region ($|\eta| < 0.9$), extending from an inner radius of 370 cm to an external one of 399 cm. Its main purpose is the particle identification in the intermediate momentum range, achieving a π/K and K/p separation better than 3 times the time-of-flight resolution below about 2.5 GeV/c for pions, and up to 4 GeV/c for protons. The TOF detector is built with a modular structure corresponding to 18 sectors in φ and five modules along the beam direction. It is based on Multi-gap Resistive Plate Chambers (MRPCs) [118], which are capable of providing a time resolution of about 40 ps [102]. The gas mixture used in the MRPCs is $C_2H_2F_4(90\%)$, i-$C_4H_{10}(10\%)$, $SF_6(5\%)$, which proved to be a good solution as no significant ageing effects were observed after 3.5 times the dose foreseen in the first 10 years of operation.

Together with the ITS and TPC, the TOF is the detector in charge of providing event-by-event identification of large samples of pions, kaons, and protons. The mass of a given particle is estimated from the measurement of its time-of-flight:

$$m = p \cdot \sqrt{\left(\frac{t_{\text{flight}}}{L}\right)^2 - 1} \quad , \tag{3.1}$$

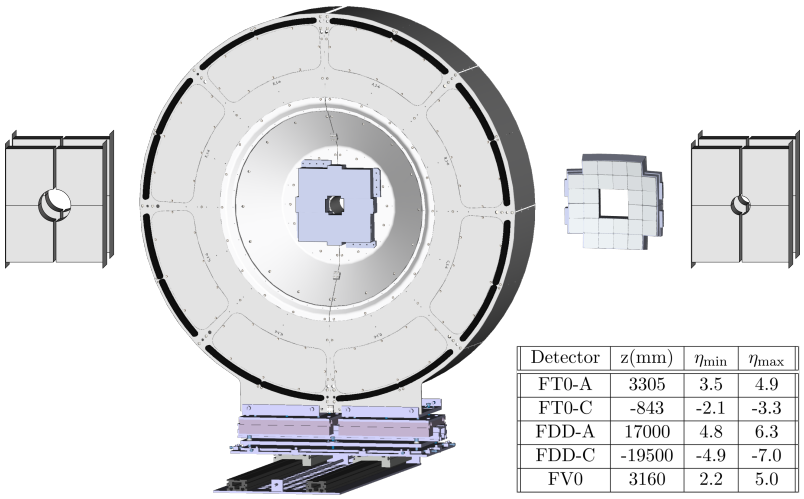
where p is the particle's momentum, measured from the curvature of its trajectory, t_{flight} is the time-of-flight, and L is the track length. The time-of-flight is evaluated as the difference between the time of arrival of the particle at the TOF (t_{hit}) and the time of the collision (t_0), estimated using a timing signal from the FT0 detector: $t_{\text{flight}} = t_{\text{hit}} - t_0$.

3.2.5 Fast Interaction Trigger

The Fast Interaction Trigger (FIT) serves as the main forward trigger, luminometer, and interaction-time detector. It also provides an initial indication of the vertex position and determines the multiplicity, centrality, and reaction plane of heavy-ion collisions. The FIT consists of five distinct detector stations, positioned at different locations along the beam line: the FT0 (-A and -C), the FV0, and the FDD (-A and -C). An illustration of the FIT is shown in Fig. 3.6.

FT0

The FT0 is made of two arrays of quartz Cherenkov radiators, FT0-A and FT0-C, coupled to MicroChannel Plate-based photomultipliers (MCP). Its main task is to determine the vertex position with an accuracy of a few centimeters, and to provide the interaction time with the precision needed by the TOF system for event-by-event particle identification. An excellent time resolution (< 50 ps) is therefore required. The FT0-A is located at 3.3 m from the nominal interaction point, while the FT0-C is only 84 cm from the interaction point. To provide a reliable indication of the time of a collision, the FT0-C has a convex shape, to ensure that each of the 112 quartz radiators is positioned at a distance of 84 cm from the nominal interaction point. On the contrary, the FT0-A is characterised by a planar geometry, made



Detector	z(mm)	η_{min}	η_{max}
FT0-A	3305	3.5	4.9
FT0-C	-843	-2.1	-3.3
FDD-A	17000	4.8	6.3
FDD-C	-19500	-4.9	-7.0
FV0	3160	2.2	5.0

Figure 3.6: View of the FIT detectors illustrating the relative sizes of each component. From left to right, the FDD-A, FT0-A, FV0, FT0-C, and FDD-C are shown. The FT0-A and FV0 systems have a common mechanical support: the former is the small quadrangular structure in the centre of the large, circular FV0 support. The inset table lists the distance from the interaction point and the pseudorapidity coverage for each component. Figure taken from Ref. [103].

of 96 quartz radiators. To achieve the best possible timing resolution, the signal path from each MCP anode to the front-end electronics has the same length. The intrinsic time resolution of each quadrant is $\sigma_t \sim 13$ ps

FV0

The FV0 is a large, segmented scintillator disk. It achieves a single MIP time resolution of about 200 ps with a very uniform response across the entire detection surface thanks to its light collection scheme, shown in Fig. 3.7. The active element of FV0 is a 4 cm-thick plastic scintillator divided into five concentric rings of equal pseudorapidity coverage. The outer diameter of the largest ring is 144 cm and the inner diameter of the smallest is 8 cm. The four inner rings are subdivided into eight sectors of 45 degrees each, while the outermost ring, due to its large area, has 16 sectors. A grid of equal-length fibers is connected to the back side of the scintillator, where they are attached to the photomultiplier tubes (PMTs). Each sector is read out by an independent PMT to provide a minimum bias and multiplicity trigger, when coupled to the FT0 information.

Figure 3.7: Photograph of one half of the FV0 detector. Figure taken from Ref. [103].

FDD

The FDD is made of two similar arrays, FDD-A and FDD-C, surrounding the beam pipe on opposite sides of the nominal interaction point, and consisting of eight rectangular scintillator pads each, arranged into two overlapping layers of four sectors. A quadrant was removed from the innermost corner of each scintillator plate to allow for the passage of the beam pipe. Since the FDD covers a large pseudorapidity interval, and is sensitive to the presence of even a single MIP, it is an ideal system to tag interactions characterised by large rapidity gaps as those from photon-induced ultra-peripheral collisions or diffractive processes.

Chapter 4

D_S^+ and D^+ reconstruction strategy in proton-proton collisions

Due to their mean proper decay lengths ($c\tau$) of 151.2 µm and 309.8 µm respectively [4], D_s^+ and D^+ mesons and their charge conjugates cannot be directly detected, as they typically decay before reaching the ALICE detector. Consequently, their production is inferred through the reconstruction of their decay products. This analysis exploits their hadronic decays into $D_s^+(D^+) \rightarrow \phi\pi^+ \rightarrow K^+K^-\pi^+$ and their charge conjugates, with a branching ratio of $(2.21 \pm 0.06) \times 10^{-2}$ $\left((2.69^{+0.07}_{-0.08}) \times 10^{-3}\right)$ [4]. An additional hadronic decay channel of the D^+ meson with a larger BR of $(9.38 \pm 0.16) \times 10^{-2}$ [4] could be exploited [119]. However, the reconstruction of the two D-mesons species in the same decay channel allows for the cancellation of many of the systematic uncertainties affecting the measurement, leading to a more precise D_s^+/D^+ production yield ratio measurement. The choice of reconstructing D mesons through their decays into a hadronic final state allows for the full reconstruction of the decay topology, therefore providing a more precise measurement. Semileptonic decay measurements are in fact affected by larger uncertainties due to the presence of neutrinos in the final state, which are not detected. In addition, the decay into the resonant $\phi(1020)$ state opens up the possibility of exploiting the invariant mass of the ϕ meson for the selection of the D-meson candidates. A sketch of the decay topology of D_s^+ and D^+ mesons into $\phi\pi^+ \rightarrow K^+K^-\pi^+$ is shown in Fig. 4.1.

D_s^+ and D^+ mesons (and their charge conjugates) are reconstructed in three different steps: i. firstly, charged tracks are reconstructed at midrapidity ($|\eta| < 0.8$) exploiting the ITS and TPC detectors; ii. D_s^+ and D^+ candidates are constructed by combining triplets of tracks with the appropriate charge signs, i.e., (+, –, +) for D_s^+ and D^+ mesons, and (–, +, –) for their antiparticles; iii. finally, the D_s^+ and D^+ candidates are selected by applying a set of topological, kinematical, and Particle-IDentification (PID) selections. Given the large number of tracks produced in a pp collision, the vast majority of the constructed D_s^+ and D^+ candidates are obtained from the combination of uncorrelated tracks, which do not originate from the same decay vertex. This results in a large *combinatorial background*, which has to be suppressed in order to extract the signal of the D mesons.

The spatial resolution capabilities of the ALICE detector described in Chapter 3

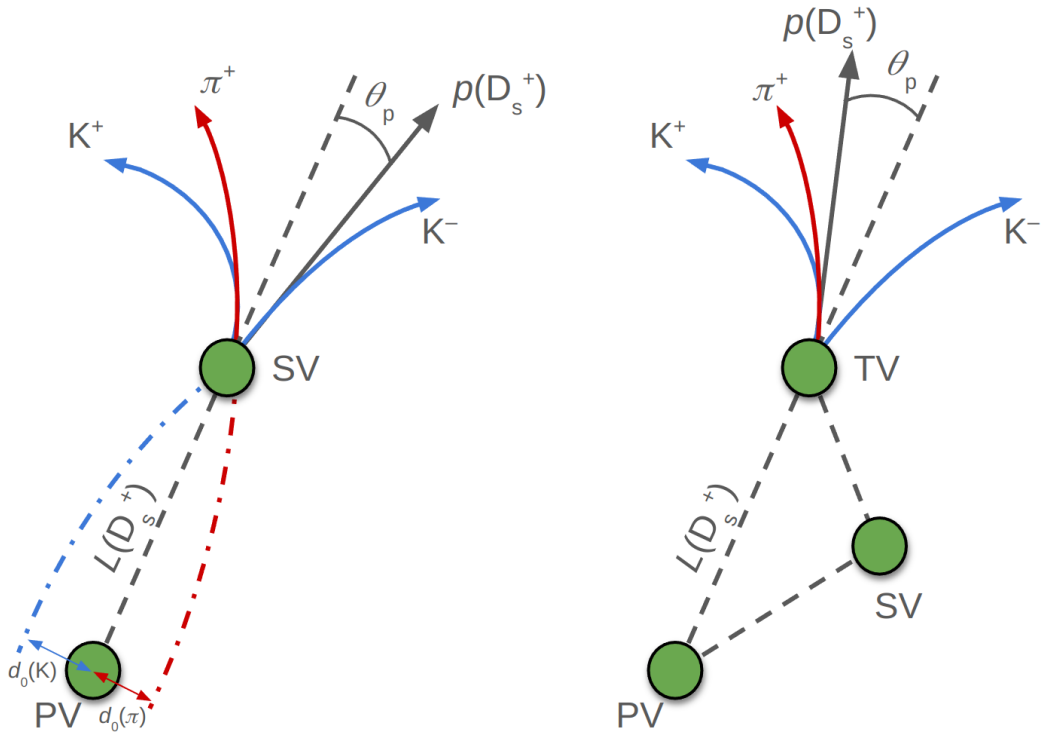


Figure 4.1: Sketch of the decay topology of D_s^+ and D^+ mesons into $\phi\pi^+ \to K^+K^-\pi^+$ for prompt (left) and non-prompt (right) D_s^+ meson.

enable the separation of the secondary decay vertices of D mesons from the primary interaction vertex, which consents the development of an analysis based on the reconstruction and selection of secondary-vertex topologies characterised by relatively large separations from the primary interaction vertex. Furthermore, PID information can be exploited to improve the selection of D mesons and their decay products and reduce the background.

Two distinct categories of D mesons emerge based on their production mechanism: *prompt* and *non-prompt* (also referred to as *feed-down*). Decay vertices of feed-down D mesons are on average more displaced from the primary interaction point with respect to promptly-produced ones, due to the larger mean proper decay length of beauty hadrons ($c\tau \sim 500$ µm [4]) as compared to charm hadrons. Therefore, by exploiting selection criteria based on displaced decay-vertex topologies, it is possible not only to separate D-meson signals from the combinatorial background, but also to discriminate between feed-down and prompt D mesons.

4.1 Data sample and event selection

The analysis reported in this Thesis is performed on a dataset of pp collisions at a centre-of-mass energy of $\sqrt{s} = 13.6$ TeV, collected by the ALICE detector during the 2022 data-taking period. The data sample is collected using a Minimum-Bias trigger (the `Sel8` trigger), which selects events that satisfy the requirement of having a signal coincidence in the FT0-A and FT0-C detectors. Furthermore, since it is observed

that the readout of the TPC (which is performed at the end of each TF) causes a drop in the track reconstruction efficiency, it was decided to exclude the collisions read out during this time interval (*TF border* selection). An additional requirement is imposed on the primary vertex position along the beam axis, $|z_{\mathrm{vtx}}| < 10$ cm, to ensure that the point where protons have collided is located within the region of the detector where the tracking efficiency is optimal. Of the total 59.2 billion collected events, only 54.7 billion have been analysed, corresponding to an integrated luminosity of $\mathcal{L}_{\mathrm{int}} \sim 1\,\mathrm{pb}$. Around 3% of the events are rejected due to the minimum bias trigger, about 1% due to the TF border selection, and 4% of the events are rejected due to the requirement on the z_{vtx} position. A summary of the event selection is shown in Fig. 4.2.

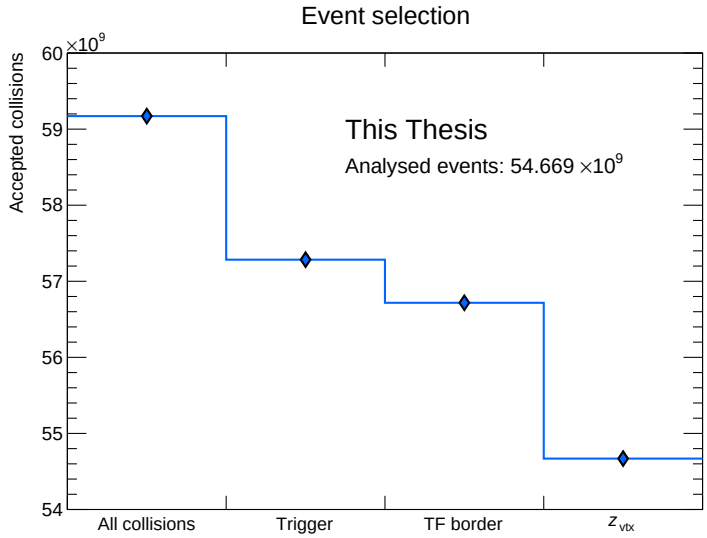


Figure 4.2: Summary of the event selection criteria applied to the data sample.

4.2 D_s^+ and D^+ reconstruction workflow

The reconstruction of D_s^+ and D^+ mesons is a complex process performed in many subsequent steps. As stated in Chapter 3, the reconstructed data are stored in AOD format. This format contains essential information about the reconstructed tracks, which are parameterised at the innermost update point, i.e., the closest point to the primary vertex with a detected track signal. Since the track parametrisation is not the same for each track, as missing hits in the ITS can lead to different radii of the innermost update points, the tracks are propagated to the point of closest approach to the primary vertex, and their parameters are updated accordingly. This procedure is performed in the `track-propagator` workflow, and only tracks passing a set of predefined quality requirements are selected.

Because of continuous readout, there exists the possibility that a single track is found to be compatible with multiple collisions (these tracks are named *ambiguous tracks*). In the AOD files, the ambiguous tracks are only associated with the first

space-time-compatible collision. In order to increase the reconstruction efficiency, the `track-to-collision associator` workflow associates each track with all compatible collisions. This process significantly increases the reconstruction efficiency, especially for 2- or 3-body decaying particles.

For each considered collision, combinations of two or three tracks are built in the `track-index-skim-creator` workflow, which produces a table of track indices for combinations passing a set of loose selection criteria, which depend on the considered hadron species. Candidates for each particle species and their decay vertices are created in the `candidate-creator` workflow, and those passing the more stringent analysis selection criteria are flagged in the `candidate-selector` workflow. The `tree-creator` (`analysis-task`) workflow produces the final output of the analysis, which consists of a ROOT tree (several ROOT histograms) containing the properties of selected candidates.

4.3 $\mathrm{D_s^+}$ and $\mathrm{D^+}$ decay-vertex reconstruction and selection

The decay vertex of the candidate is reconstructed through a minimisation of a χ^2-like quantity, denoted as D:

$$D = \sqrt{\sum_{i=1}^{3} \left[\left(\frac{x_i - x_0}{\sigma_{x_i}} \right)^2 + \left(\frac{y_i - y_0}{\sigma_{y_i}} \right)^2 + \left(\frac{z_i - z_0}{\sigma_{z_i}} \right)^2 \right]} \quad ,$$

where (x_i, y_i, z_i) and $(\sigma_{x_i}, \sigma_{y_i}, \sigma_{z_i})$ represent the position and the uncertainty of the i-th track at the point of closest approach, respectively, while (x_0, y_0, z_0) denotes the position of the reconstructed vertex. The invariant mass and momentum of the $\mathrm{D_s^+}$ and $\mathrm{D^+}$ candidates are computed from the energy and momentum of the measured tracks evaluated at the point of closest approach to the decay vertex. The momentum of the candidate is defined as the sum of the momenta of the three tracks. For the invariant mass computation, the kaon mass is always assigned to the track with opposite charge sign with respect to the D-meson candidate (*opposite-sign track*). For the two *like-sign tracks*, the two pion-kaon mass hypothesis combinations (i.e. $(\mathrm{K^+K^-\pi^+})$ and $(\pi^+\mathrm{K^-K^+})$ for positively charged candidates) are considered.

The vast majority of the created $\mathrm{D_s^+}$- and $\mathrm{D^+}$-meson candidates belong to the large combinatorial background. To increase the signal-over-background ratio and the statistical significance of the measurement, tight selections are required to suppress such candidates. The analyses presented in this Thesis exploit several selection criteria, which can be divided into:

i Track-quality selections

ii Selections based on the decay topology and kinematics

iii Particle identification of the decay products

In the following, the variables used to select D mesons and some of the applied selections are described in more detail.

4.3.1 Track-quality selections

Only tracks that successfully pass strict quality and kinematic requirements are considered eligible for inclusion in the construction of D_s^+- and D^+-meson candidates. In particular, only ITS-TPC tracks with at least 70 (out of a maximum of 159) track-associated space points in the TPC, and a crossed rows (i.e., the total number of hit TPC pad rows) over findable clusters (i.e., pad rows that could potentially be hit, given the track's trajectory) ratio of at least 0.8 are selected. To improve the vertex reconstruction procedure, at least one hit in the 3 innermost layers of the ITS, which compose the ITS inner barrel, was required.

Secondary vertices of D_s^+- and D^+-meson candidates are constructed using tracks having $|\eta| < 0.8$ and $p_T > 0.3$ GeV/c. A track-quality requirement of χ^2 per TPC cluster smaller than 4 has been applied. A cut on the daughter-track transverse impact parameter projection in the transverse plane d_0^{xy} was applied, requiring $d_0^{xy} > 25\,\mu$m for tracks with $p_T < 2$ GeV/c. These selections limit the rapidity acceptance of D mesons, which steeply decreases for $|y| > 0.5$ at low p_T and for $|y| > 0.8$ for $p_T \gtrsim 5$ GeV/c, as shown in Fig. 4.3. The applied track-quality selection criteria are summarised in Table 4.1.

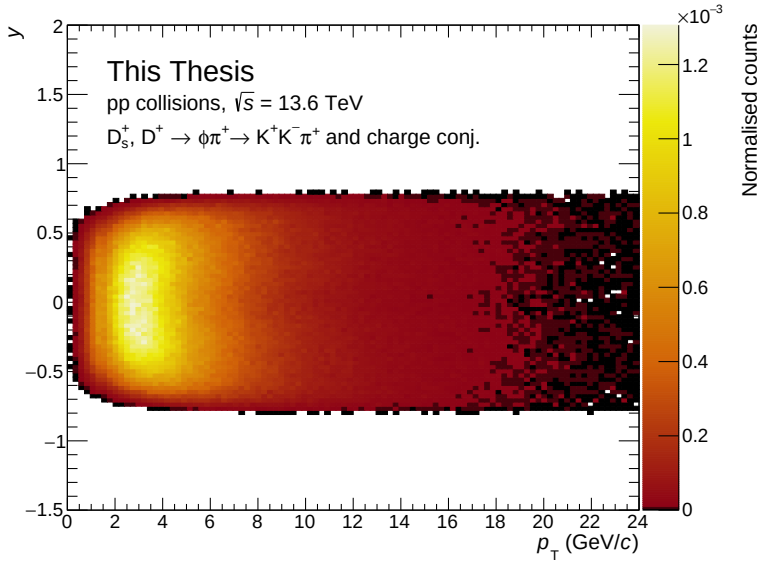


Figure 4.3: Rapidity and transverse momentum distribution of reconstructed D_s^+ and D^+ mesons in pp collisions at $\sqrt{s} = 13.6$ TeV.

4.3.2 Topological selections

D_s^+ and D^+ mesons exhibit a displaced decay vertex topology, which can be used to separate the signal from the uninteresting combinatorial background. Moreover, promptly produced D mesons exhibit different topological features compared to feed-down D mesons, enabling further discrimination between the two production mechanisms. This differentiation potentially offers insights into beauty-quark production through the measurement of open-charm states.

Single-track selection	Value
Number of TPC crossed-rows >	70
$\|\eta\|$ <	0.8
p_T >	0.1 GeV/c
χ^2_{TPC}/TPC clusters <	4
χ^2_{ITS}/ITS clusters <	36
ITS matching	At least 1 cluster in L0, L1, L2

Table 4.1: Applied single-track selection criteria

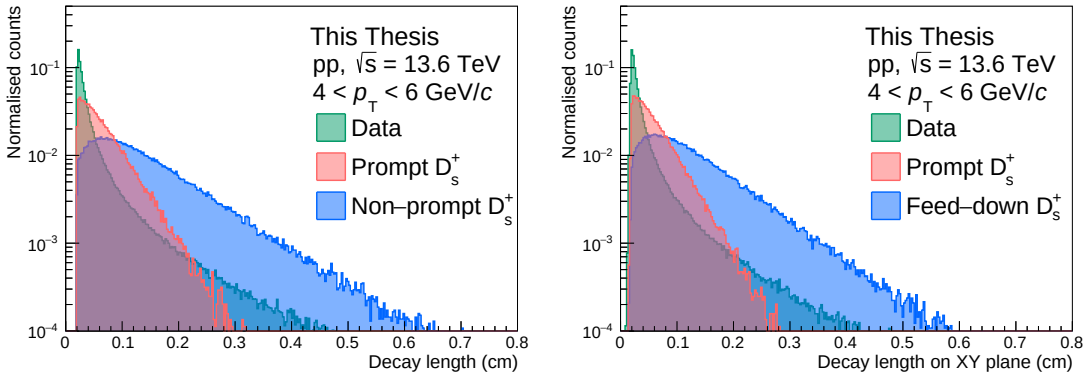


Figure 4.4: Distributions of decay length (left panel) and its projection on the transverse plane (right panel) for D_s^+ mesons in pp collisions at $\sqrt{s} = 13.6$ TeV in the $4 < p_T < 6$ GeV/c interval. The distributions are shown for prompt D_s^+ mesons (blue), non-prompt D_s^+ mesons (orange), and combinatorial background (green). For prompt and non-prompt D_s^+ mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.

The topological selections are tuned as a function of p_T in order to increase the signal-over-background and the statistical significance of the measurement. The different selections are presented herein. The corresponding distributions of variables are shown for the signal, which is divided into prompt and feed-down contributions and obtained from Monte Carlo simulations, as well as for the combinatorial background, which is obtained from real data in an invariant-mass region away from the signal region, denoted as *sidebands* and chosen as $1.7 < M < 1.75$ GeV/c^2 or $2.1 < M < 2.15$ GeV/c^2.

Decay Length

The decay length L is defined as the distance between the primary and secondary vertices. It provides an approximation of the actual decay length of D_s^+ and D^+ mesons, as the particle's curvature resulting from the motion in the presence of a magnetic field is not considered. However, given the small mean proper decay length of ~ 150 (310) µm of D_s^+ (D^+) mesons, this effect can be neglected.

This is one of the most important variables used to distinguish between signal and combinatorial background, since the displaced topology of the signal shifts the decay length distribution towards larger values. Furthermore, since beauty-hadron decay vertices are not reconstructed, the measured decay length of non-prompt D-mesons also accounts for the decay length of their parent hadrons. As a consequence, the resulting decay length distribution for feed-down D-mesons is particularly shifted towards larger values, allowing for an easier separation of this production mechanism from the background. The distribution of the decay length also depends on the D-meson p_{T}, because of the Lorentz boost which contributes to an increase in the distance flown by the hadron measured in the laboratory reference frame. Moreover, since the mean proper decay length of $\mathrm{D^+}$ meson is about twice as large as that of the $\mathrm{D_s^+}$ meson, the decay length distribution of the former is shifted towards larger values. This allows for a possible discrimination of the two D-meson species based on the exclusive exploitation of topological variables.

In addition to the decay length, its projection on the transverse plane can also be used to select $\mathrm{D_s^+}$ or $\mathrm{D^+}$ signal. This variable is particularly useful as it exploits the better resolution of the ALICE experiment in the transverse plane with respect to the longitudinal one. Previous measurements of D mesons also leveraged the normalised decay length, defined as the ratio between the decay length and its uncertainty, to further improve the signal extraction. Although this variable demonstrated a good separation power, it was not used in this analysis, due to possible biases introduced by the uncertainty description in the Monte Carlo simulations.

The distributions of decay length and its projection on the transverse plane are shown in Fig. 4.4 for signal and combinatorial background candidates reconstructed in the $4 < p_{\mathrm{T}} < 6$ GeV/c interval.

Cosine of pointing angle

The pointing angle θ_{p} is defined as the angle between the flight direction of the D meson, obtained from the line connecting the primary and secondary vertices, and the direction of the reconstructed D-meson momentum. This variable can also be defined using only the transverse components of these quantities ($\theta_{\mathrm{p}}^{\mathrm{xy}}$), to exploit the better resolution in the x and y coordinates than in the z coordinate. In an ideal scenario where particles' momentum is perfectly reconstructed, the cosine of the pointing angle for promptly-produced D mesons would be equal to 1, while that of feed-down D mesons would be distributed around 1, but with a tail towards lower values as the flight direction of the D meson is not equal to that of the parent beauty hadron. On the contrary, the pointing angle for combinatorial background could assume any value with the same probability. The pointing angle is therefore particularly useful to separate the signal from the combinatorial background. The distributions of the cosine of the pointing angle and its projection on the transverse plane are shown in Fig. 4.5 for signal and combinatorial background candidates reconstructed in the $4 < p_{\mathrm{T}} < 6$ GeV/c interval. Due to the finite resolution of the tracking detectors, the pointing angle is not perfectly reconstructed, leading to a distribution of the cosine of the pointing angle for prompt $\mathrm{D_s^+}$ that is peaked at 1, but with a tail towards lower values.

Similarly to the decay length, also θ_{p} evolves with the D-meson transverse mo-

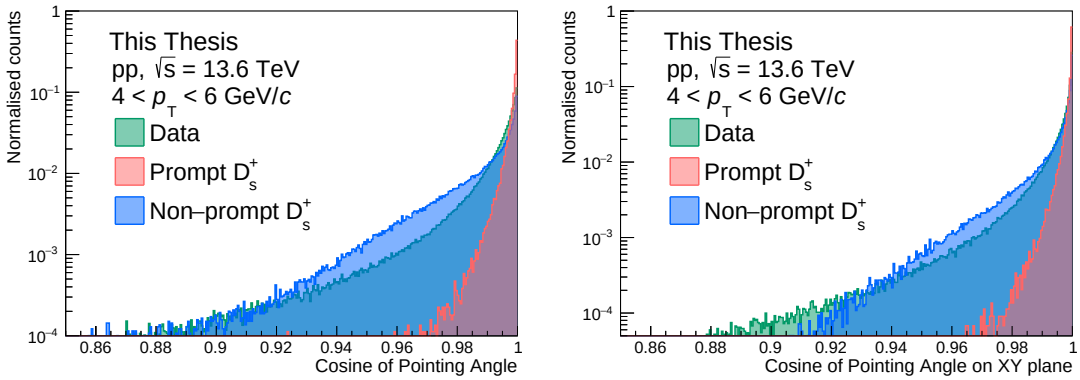


Figure 4.5: Distributions of the cosine of the pointing angle (left panel) and its projection on the transverse plane (right panel) for D_s^+ mesons in pp collisions at $\sqrt{s} = 13.6$ TeV in the $4 < p_T < 6$ GeV/c interval. The distributions are shown for prompt D_s^+ mesons (blue), non-prompt D_s^+ mesons (orange), and combinatorial background (green). For prompt and non-prompt D_s^+ mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.

mentum. Because of the Lorentz boost, the direction of the daughter particles of the D meson is more collimated with the D-meson flight direction at higher p_T. This results in a distribution of the cosine of the pointing angle that is more peaked at 1 for higher p_T values.

Impact parameter in the transverse plane

The projection of the impact parameter on the transverse plane d_0^{xy} is defined as the distance of closest approach between the reconstructed flight line of the D-meson and the primary vertex, projected in the xy plane. It is expected to be very close to zero for promptly-produced D mesons, with any deviation due to the detector resolution. Thanks to the better momentum and vertexing resolution, its distribution becomes narrower at high p_T. It is a powerful variable not only to separate the signal from the combinatorial background, but also to discriminate between prompt and feed-down D mesons, as the latter have a much broader impact parameter distribution, as shown in the left panel of Fig 4.6. In addition, also the projection of the daughter tracks' impact parameter to the primary vertex projected on the transverse plane is used to select the signal. Since the combinatorial background does not have a secondary vertex, many of the tracks used to construct such candidates are produced directly in the primary vertex. This results in a narrow distribution around 0 of the prongs' impact parameters for such candidates. On the contrary, the displaced decay vertex of the signal candidates leads to a broader distribution of the impact parameter for prompt D-meson candidates, and to an even broader one for the non-prompt contribution to the signal, as shown in the right panel of Fig. 4.6.

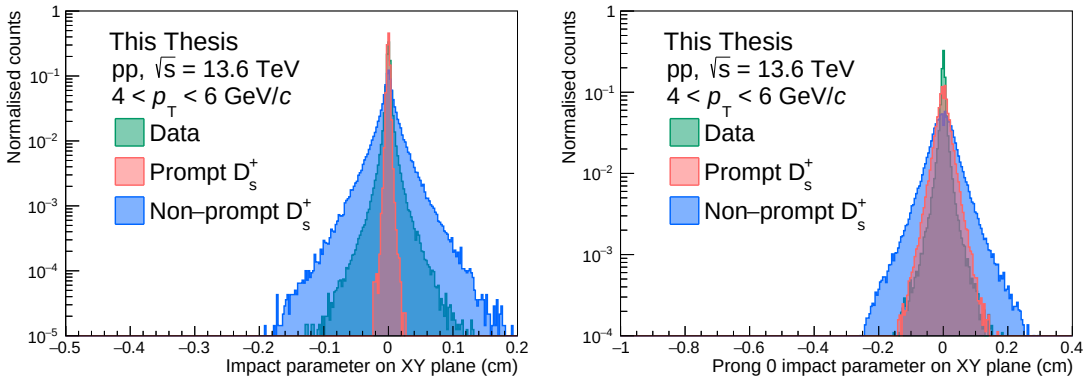


Figure 4.6: Distributions of the projection of the impact parameter on the transverse plane for D_s^+ mesons (left panel) and for one of the opposite-sign daughter tracks (right panel) in pp collisions at $\sqrt{s} = 13.6$ TeV in the $4 < p_T < 6$ GeV/c interval. The distributions are shown for prompt D_s^+ mesons (blue), non-prompt D_s^+ mesons (orange), and combinatorial background (green). For prompt and non-prompt D_s^+ mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.

4.3.3 Kinematic selections

In addition to the selections based on the decay topology, kinematic selections are also applied to increase the signal-over-background ratio and the statistical significance of the measurement.

Difference between reconstructed and PDG mass of the ϕ meson

This selection exploits the production of an intermediate resonant $\phi(1020)$-meson state in the considered decay channel of the D mesons, whose invariant mass of $M_\phi = (1019.461 \pm 0.016)$ MeV/c^2 [4] is known with good precision. The K^+K^- pair produced in the ϕ-meson decay is expected to have an invariant mass very close to the PDG value, as the ϕ resonance is narrow ($\Gamma_\phi = 4.249 \pm 0.013$ MeV/c^2 [4]). Therefore, the difference between the reconstructed and the PDG mass of the ϕ meson $|\Delta M(\mathrm{KK})|$ is expected to be close to zero for signal candidates, while the distribution for combinatorial background is expected to be uniformly distributed.

Moreover, two possible K^+K^- pairs can be built for each triplet of tracks, depending on the mass hypothesis assigned to the like-sign tracks (e.g., for a D_s^+ decay, both the $K^+K^-\pi^+$ and $\pi^+K^-K^+$ mass hypotheses can be considered). Both hypotheses are considered in the reconstruction, and the selection on $|\Delta M(\mathrm{KK})|$ results extremely helpful not only in rejecting the combinatorial background, but also the reflections, i.e., D mesons with wrongly-assigned mass hypothesis on the decay tracks.

The distribution of this variable is shown in the left panel of Fig. 4.7 for signal and combinatorial background candidates reconstructed in the $4 < p_T < 6$ GeV/c interval.

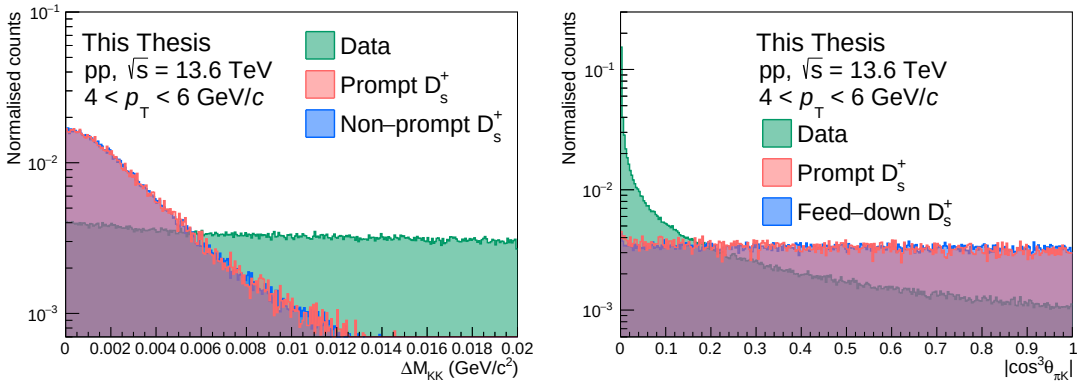


Figure 4.7: Distributions of the difference between the reconstructed and PDG mass of the ϕ meson (left panel) and the cosine cubed of the K-π angle in the KK rest frame (right panel) for D_s^+ mesons in pp collisions at $\sqrt{s} = 13.6$ TeV in the $4 < p_T < 6$ GeV/c interval. The distributions are shown for prompt D_s^+ mesons (blue), non-prompt D_s^+ mesons (orange), and combinatorial background (green). For prompt and non-prompt D_s^+ mesons, the distributions are taken from Monte Carlo simulations, whereas for combinatorial background they are taken from the data sidebands.

Cosine cubed of the K-π angle in the KK rest frame

The decay of a (pseudoscalar) D-meson to a (vector) ϕ-meson and a (pseudoscalar) π^+ final state results in an alignment of the spin of the ϕ meson with respect to the direction of motion of the ϕ relative to the D-meson [120]. As a consequence, the distribution of $\cos(\theta'(\mathrm{K}))$, where $\theta'(\mathrm{K})$ is the angle between one of the kaons and the pion in the $\mathrm{K^+K^-}$ rest frame, follows a $\cos^2(\theta'(\mathrm{K}))$ shape, which in turn implies a flat distribution for the $\cos^3(\theta'(\mathrm{K}))$ variable, in case of signal. In contrast, the combinatorial background has a flat distribution for $\cos(\theta'(\mathrm{K}))$, and its $\cos^3(\theta'(\mathrm{K}))$ distribution peaks at zero. This variable is particularly useful to separate the signal from the combinatorial background, as shown in the right panel of Fig. 4.7.

4.3.4 Particle identification selections

One of the distinctive features of the ALICE detector are the excellent PID capabilities in a wide range of momentum, achieved by combining the information from different detectors.

The ITS was used for its PID information during the LHC Run 1 and Run 2 data-taking periods, when the analogue readout of the SDD and SSD layers allowed for the measurement of the specific energy loss $\mathrm{d}E/\mathrm{d}x$ of charged particles from the charge deposited in the detector. The upgraded ITS 2 is not capable of providing such information, as the readout is now fully digital. Nonetheless, new PID techniques based on ML algorithms exploiting information such as the cluster shape and size are being developed and tested for the ITS 2. Preliminary results show that the ITS 2 PID is capable of achieving a good separation between protons and other particles, as shown in the left panel of Fig. 4.8.

The TPC is the main PID detector in the ALICE experiment, as it provides a

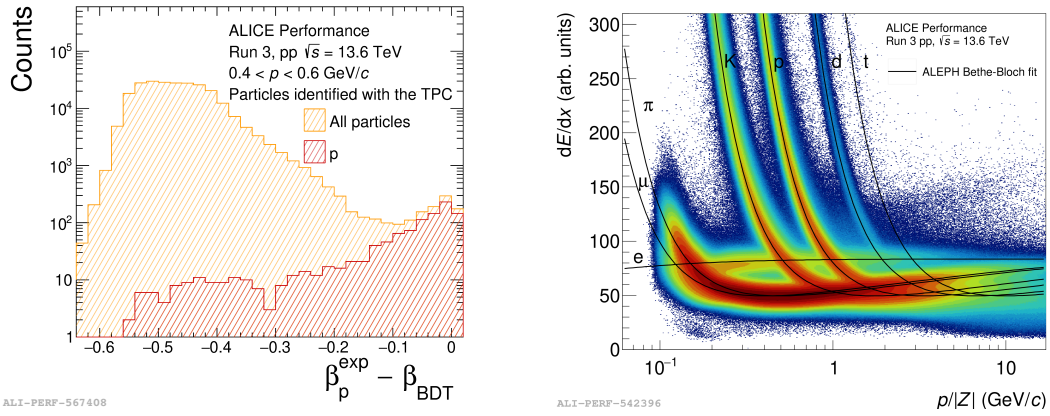


Figure 4.8: Left: ITS 2 particle identification performance in pp collisions at $\sqrt{s} = 13.6$ TeV using a Boosted Decision Tree (BDT) regressor. From the cluster size, the regressor predicts the velocity β (β_{BDT}) of particles. The regressor performance is evaluated for $0.4 < p < 0.6$ GeV/c, through the difference between β_{BDT} and $\beta_{\mathrm{p}}^{\mathrm{exp}} = \frac{p}{\sqrt{p^2+m_{\mathrm{p}}^2}}$. Right: Specific energy loss distribution $\mathrm{d}E/\mathrm{d}x$ in the TPC for pp collisions at $\sqrt{s} = 13.6$ TeV. The curves obtained with a fit using the ALEPH parametrisation of the Bethe-Bloch formula for electrons, muons, pions, kaons, protons, deuterium and tritium are also shown. Figures taken from the ALICE figure repository [49].

simultaneous measure of the specific energy loss via the charge produced in the drift volume and deposited in its 159 pad rows, the charge, and the momentum of each particle traversing the detector gas. A truncated mean of the $\mathrm{d}E/\mathrm{d}x$ samples is calculated discarding the 40% highest-charge clusters, which ensures the removal of the Landau tail of the $\mathrm{d}E/\mathrm{d}x$ distribution caused by δ-rays, which are produced from the ionisation of the gas by the electrons produced in the primary ionisation. The particle identification is based on the comparison of the measured $\mathrm{d}E/\mathrm{d}x$ with the expectation for a specific particle species with a certain momentum p. The expected $\mathrm{d}E/\mathrm{d}x$ is described by the Bethe-Bloch formula, and is typically parametrised with a function originally proposed by the ALEPH collaboration [121]:

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}} \left[P_2 - \beta^{P_4} - \log\left(P_3 + \frac{1}{(\beta\gamma)^{P_5}} \right) \right] \quad .$$

In this parametrisation, β and γ are the particle velocity and Lorentz factor, respectively, and P_{1-5} are the parameters of the fit function. The $\mathrm{d}E/\mathrm{d}x$ distribution in the TPC is shown in the right panel of Fig. 4.8 for pp collisions at $\sqrt{s} = 13.6$ TeV. The curves obtained with a fit using the ALEPH parametrisation of the Bethe-Bloch formula for electrons, muons, pions, kaons, protons, deuterium and tritium are also shown. While in the non-relativistic low momentum region ($p \leq 1$ GeV/c) p, K and π can be separated on a track-by-track basis, at higher momenta, in the relativistic rise region, particles can still be identified on a statistical basis via multi-Gaussian fits. This is possible because the truncated-mean method produces a $\mathrm{d}E/\mathrm{d}x$ peak with a gaussian shape down to three orders of magnitude for long tracks (at least 130 clusters) [122].

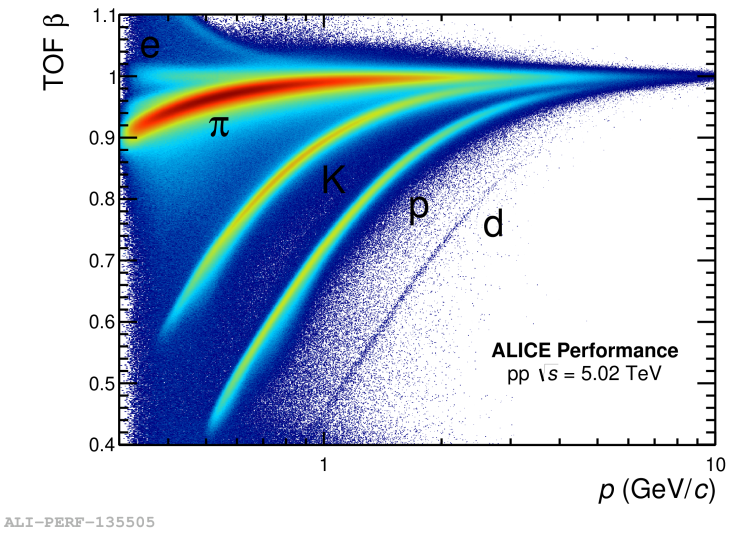


Figure 4.9: β parameter measured with the TOF detector as a function of the momentum p measured by the TPC detector in Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. Figure taken from the ALICE figure repository [49].

The TOF detector is dedicated to PID in the intermediate momentum range, up to 2.5 GeV/c for pions and kaons, and up to 4 GeV/c for protons. In this higher-p_{T} region, the TPC is not able to provide a sufficient separation between the different particle species. The TOF detector is based on the measurement of the time of flight of charged particles between the interaction point and the detector. The start time is provided by the FT0 forward-rapidity detectors, and a total time resolution of the 80 ps is achieved, allowing for a separation of the different particle species based on the time of flight. The TOF PID performance is shown in Fig. 4.9 for Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. The large background is due to tracks that are incorrectly matched to TOF hits in high-multiplicity Pb–Pb collisions.

Other detectors, such as the High Momentum Particle Identification Detector (HMPID) or the Transition Radiation Detector (TRD), complement the PID capabilities of the ALICE experiment. The HMPID is dedicated to the identification of particles with momenta above 2 GeV/c, and is based on the detection of Cherenkov radiation emitted by charged particles traversing a radiator. The TRD is used for the identification of electrons and positrons, and is based on the detection of transition radiation emitted by charged particles traversing a radiator with a change in the dielectric constant. The TRD is particularly useful for the identification of electrons and positrons in the high-multiplicity environment of heavy-ion collisions, where the TPC and TOF detectors are not able to provide a sufficient separation between electrons and pions.

The PID information on the decay tracks can be exploited to improve the background rejection and thereby the signal extraction. For this analysis, only the information from the TPC and TOF detector is used, as the ITS PID capabilities are still under development, and the final state particles of the D mesons are pions and kaons, which can be well identified with the TPC and TOF detectors. A track

is considered compatible with a mass hypothesis (i.e. a hadron species) depending on the difference between the measured signal $S_{\mathrm{meas}}^{\mathrm{TPC,TOF}}$ and the expected signal $S_{\mathrm{exp(\pi,K,p)}}^{\mathrm{TPC,TOF}}$ for the given hypothesis. The measured signal is the truncated mean of the $\mathrm{d}E/\mathrm{d}x$ distribution for the TPC detector, and the time of flight for the TOF detector. The expected signal is extracted from the ALEPH parametrisation of the Bethe-Bloch formula described above for the TPC, and from a parametrisation of the distribution of β as a function of p for the TOF. The difference between the measured and expected signals is then usually normalised by the uncertainty on the measured signal $\sigma_{\mathrm{TPC,TOF}}$:

$$\mathrm{n}\sigma_{\mathrm{\pi,K,p}}^{\mathrm{TPC,TOF}} = \frac{|S_{\mathrm{meas}}^{\mathrm{TPC,TOF}} - S_{\mathrm{exp(\pi,K,p)}}^{\mathrm{TPC,TOF}}|}{\sigma_{\mathrm{TPC,TOF}}} \quad .$$

Tracks are considered if at least one PID signal is available from one of the two detectors. A loose preselection is applied to the candidates based on PID information, requiring the three tracks to have a $\mathrm{n}\sigma^{\mathrm{TPC}} < 5$, or $\mathrm{n}\sigma^{\mathrm{TOF}} < 5$ for the considered mass hypothesis. This selection effectively reduces the amount of reflections in the data sample, as the wrongly-assigned mass hypothesis are often rejected.

Due to a TPC–TOF matching efficiency which is significantly lower than unity, the PID information from the two detectors is typically combined to obtain a single PID variable $\mathrm{n}\sigma_{\mathrm{\pi,K,p}}^{\mathrm{comb}}$, which is used in the final selection. The PID information is combined as

$$\mathrm{n}\sigma_{\mathrm{\pi,K,p}}^{\mathrm{comb}} = \begin{cases} \mathrm{n}\sigma_{\mathrm{\pi,K,p}}^{\mathrm{TOF}} & \text{only TOF information is available} \\ \mathrm{n}\sigma_{\mathrm{\pi,K,p}}^{\mathrm{TPC}} & \text{only TPC information is available} \\ \frac{1}{\sqrt{2}}\sqrt{(\mathrm{n}\sigma_{\mathrm{\pi,K,p}}^{\mathrm{TPC}})^2 + (\mathrm{n}\sigma_{\mathrm{\pi,K,p}}^{\mathrm{TOF}})^2} & \text{otherwise} \end{cases}$$

This reduces the number of features used for the Machine Learning analysis described in the following Chapters, and ensures a proper handling of sparse data.

Chapter 5

D_S^+ and D^+ signal extraction

The main ingredient for the evaluation of the D_s^+/D^+ production-yield ratio is the D_s^+ and D^+ raw yield, i.e., the number of reconstructed D_s^+ and D^+ mesons. Due to the vast amount of combinatorial background and the limited efficiency of about 1%, extracting the raw yield through a candidate counting method is not feasible. Instead, the raw yield is obtained on a statistical basis by fitting the invariant-mass distribution of the D_s^+ and D^+ candidates passing tight selection criteria. To reduce the combinatorial background and enhance the efficiency of D-meson selection, Machine Learning algorithms have been employed. The following sections describe the procedure for the extraction of the raw yield of D_s^+ and D^+ mesons.

5.1 Machine Learning

The term *Machine Learning* (ML) is a broad and versatile concept, encompassing a wide range of algorithms that grant computers the capacity to learn and adapt without being explicitly programmed to do so [123]. A more comprehensive definition characterises ML as the study of algorithms that enhance their performance at a specific task through the accumulation of experience [124]. In recent years, ML techniques have witnessed widespread adoption across diverse fields, with significant impacts realised especially with the emergence of generative models such as GPT [125]. ML algorithms have found extensive applications in the high-energy physics field, primarily for the task of distinguishing interesting signals from the vast background present in particle-collision data. Furthermore, these algorithms have been employed as triggers, aiding in the rapid identification of events of interest, and have also been instrumental in event reconstruction. Notably, ML algorithms were used in the discovery of the Higgs boson [108], one of the most significant achievements in the field of particle physics of the last decades.

5.1.1 Supervised learning

Supervised learning is one of the main branches of machine learning, along with unsupervised and reinforcement learning. Machine learning tasks are usually described in terms of how the machine learning system should process an example, which is a collection of features $\mathbf{x}$ that have been quantitatively measured from some object

or event that one wants the machine learning system to process. In the case of supervised learning, each example is coupled with a corresponding label or target, $\mathbf{y}$. The objective of supervised learning is to learn to predict or infer $\mathbf{y}$ based on the associated features, $\mathbf{x}$, assuming that there exists a functional relationship $\mathbf{y} = f(\mathbf{x})$ between the two. The goal of the ML system is to produce an approximation $\widehat{f}(\mathbf{x})$ of the true function $f(\mathbf{x})$ by minimising a given loss function, which quantifies the discrepancy between the predicted and true labels. Supervised learning problems are further segmented into two distinct sub-categories: classification and regression. In the former, the label $\mathbf{y}$ assumes values from a finite and discrete set of categories, often representing distinct classes or groups. In the latter, the label $\mathbf{y}$ takes the form of one or more continuous variables, necessitating the learning system to deduce a continuous function or mapping between $\mathbf{x}$ and $\mathbf{y}$.

The usage of ML algorithms in classification problems, such as the one presented in this Thesis, allows for the definition of multi-dimensional non-linear decision boundaries, which are not available with traditional selection methods. This is particularly important as it provides more efficient selections and a larger purity of the selected data sample.

The application of a supervised learning algorithm to a dataset involves the following steps: i. the model is trained on a set of labelled data, i.e., the value of $\mathbf{y}$ is known for each example in the training set; ii. the model is tested on a separate set of labelled data, known as the test set, to evaluate its performance; iii. the model is then used to make predictions on new, unseen data.

Training

During the training process, the model learns (i.e., adjusts its internal parameters) to map the input features $\mathbf{x}$ to the corresponding labels $\mathbf{y}$ by minimizing a given loss function. Typically used loss functions include the Mean Squared Error (MSE) for regression tasks and the Cross-Entropy loss [126] for classification tasks. The loss function is minimised through an optimization algorithm, usually stochastic gradient descent [127], which iteratively updates the model parameters to reduce the loss. Since an over-optimisation of the model on the training data can lead to poor generalization on unseen data (the model is said to be *overfitting*), a regularisation term is often added to the loss function to penalise overly complex models. The training process continues until the model reaches a satisfactory level of performance on the training data, or until its performance does not improve further.

Before the final model is trained, hyperparameters tuning is performed to optimise the model's performance. *Hyperparameters* are parameters that are not learned during the training process, but rather define the model's architecture and the training process itself. Hyperparameters tuning is usually performed through a grid search, random search, or with a more efficient bayesian optimisation [128, 129, 130]. Several combinations of hyperparameters are tested on a dedicated labelled dataset: the validation set. Models with different hyperparameter sets are trained with a reduced training phase, and those yielding the best performance are then selected for the final model training.

Testing

After the model has been trained, its performance is evaluated on a dataset that was not used during the training process, known as the test set. Like the training and validation sets, also the test set contains labelled examples. While during the training the model is optimised to minimise the loss function, the test set is used to estimate the model's generalization error, i.e., how well the model performs on unseen data. The model's performance is evaluated using metrics that are specific to the task at hand, such as accuracy for classification tasks, or Mean Squared Error (MSE) for regression tasks. Once the model achieves satisfactory performance on the test set, it is ready to be used for making predictions on unlabelled data.

Cross-validation

With the strategy defined above to optimise the hyperparameters, train the model and validate its performance, the dataset is divided into three subsets: the training set, the validation set, and the test set. When small datasets are involved, this division can lead to a suboptimal model, as the model's performance can be highly dependent on the specific examples in the training, validation, and test sets. Furthermore, this approach limits the amount of data available for training the model, which can lead to poor generalization. To mitigate this issue, cross-validation [131] is often employed. This term refers to a set of techniques that allow for a more robust estimate of the model's generalisation performance by using the entire dataset for training and validation. The most common cross-validation technique is the k-fold cross-validation. It consists in dividing the training sample into k subsets of equal size, called *folds*. Then, the ML algorithm is trained k times, each time using $k-1$ folds as training set, while the remaining fold is used as validation set. The model's performance is then averaged over the k folds to obtain a more robust estimate of this quantity. This operation is repeated for each hyperparameter configuration to be considered. The hyperparameter configuration minimising the loss function is then chosen as the optimal configuration.

5.2 D_s^+ and D^+ selection using Machine Learning

The task of extracting D_s^+ and D^+ signals from the vast combinatorial background is a challenging one, due to the large amount of background compared to signal. It is however an excellent example of classification problem, and ML algorithms can therefore be exploited to enhance the efficiency of the selection.

5.2.1 Data preparation

In order to train a ML model, a labelled dataset with a well-defined set of features is required. The dataset used for training the ML algorithms employed in this Thesis is composed of a number of signal and background examples. To obtain a pure sample of signal candidates, Monte Carlo techniques are used to generate D_s^+ and D^+ mesons. Proton-proton collisions are simulated using the Pythia 8 event generator [73] with colour-reconnection Mode 2 [78], and the generated particles are

propagated through the ALICE detector using the GEANT4 transport simulation toolkit [132]. To enrich the sample of heavy-flavour hadrons, $c\bar{c}$ and $b\bar{b}$ pairs are injected into each simulated event.

Only prompt and non-prompt D_s^+ mesons are used to train the model, as D^+ mesons decay into the same final state as D_s^+ mesons, and selections optimised to reconstruct D_s^+ mesons are also effective for D^+ mesons.

Background candidates are obtained from real data, as MC simulations may not be able to reproduce the complexity of soft processes occurring in the underlying event, or may not be able to model the detector response accurately. Background examples are obtained by selecting candidates from a subsample of the full data sample (corresponding to its 3%) in an invariant-mass region away from both the D_s^+ and D^+ mass peaks, where $1.7 < M < 1.75$ GeV/c^2 or $2.1 < M < 2.15$ GeV/c^2, as shown in Fig. 5.1.

Labels are assigned as a numerical value to each candidate, with 0 indicating a background candidate, 1 a prompt D_s^+ meson, and 2 a non-prompt D_s^+ meson.

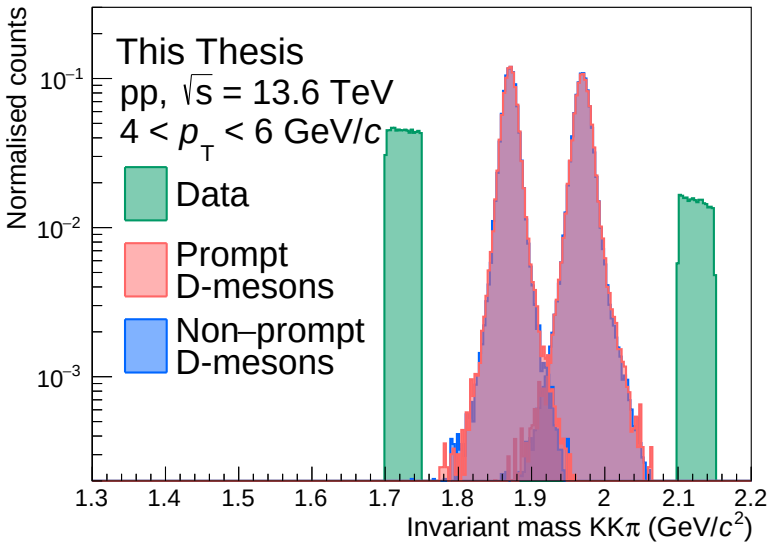


Figure 5.1: Invariant mass distribution of prompt and non-prompt D-mesons (blue and orange, respectively), taken from Monte Carlo simulations, and of the background candidates taken from real data used to train the ML model (green) in the $4 < p_{\mathrm{T}} < 6$ GeV/c interval. Background candidates are selected in the $1.7 < M < 1.75$ GeV/c^2 or $2.1 < M < 2.15$ GeV/c^2 invariant-mass interval.

The dataset is then divided into two different subsamples. The first comprehends 80% of the data, and is used to train the model, while the remaining 20% is used to test its performance. In addition, since the D-meson decay topology can significantly differ depending on the p_{T} of the meson due to different Lorentz boosts, the dataset is divided into several p_{T} intervals, and the model is trained and tested separately for each of them. To achieve a better performance of the ML models, they are trained in broader p_{T} intervals than those used for the analysis, to ensure enough data is available to train a well-performing model. The total number of candidates available for training and testing the model is reported in Table 5.1 for the considered

p_{T} intervals.

Table 5.1: Number of candidates within the p_{T} intervals used to train and test the model.

p_{T} (GeV/c)	Prompt $\mathrm{D_s^+}$	Non-prompt $\mathrm{D_s^+}$	Background
0–1.5	$\sim 4.6 \times 10^3$	$\sim 21 \times 10^3$	$\sim 726 \times 10^3$
1.5–2	$\sim 6.1 \times 10^3$	$\sim 24 \times 10^3$	$\sim 92 \times 10^3$
2–3	$\sim 26 \times 10^3$	$\sim 96 \times 10^3$	$\sim 123 \times 10^3$
3–4	$\sim 34 \times 10^3$	$\sim 124 \times 10^3$	$\sim 114 \times 10^3$
4–5	$\sim 31 \times 10^3$	$\sim 113 \times 10^3$	$\sim 63 \times 10^3$
5–6	$\sim 24 \times 10^3$	$\sim 89 \times 10^3$	$\sim 29 \times 10^3$
6–8	$\sim 32 \times 10^3$	$\sim 115 \times 10^3$	$\sim 22 \times 10^3$
8–12	$\sim 23 \times 10^3$	$\sim 89 \times 10^3$	$\sim 10 \times 10^3$
12–24	$\sim 9.7 \times 10^3$	$\sim 39 \times 10^3$	$\sim 2.6 \times 10^3$

To produce a balanced dataset, the number of candidates in each class is equalised to the number of examples in the minority class. This is achieved by randomly selecting a subset of the majority classes. The balanced dataset is then used to train the model.

The choice of features used to separate signal from background is crucial, as they must be able to discriminate between signal and background candidates, and must be chosen in such a way that no bias is introduced in the final result. The variables used to train the model were introduced in Chapter 4, and are a mix of topological, kinematic, and PID variables. The key idea is to exploit the displaced topology of the D-meson decay, which is a distinctive feature of the signal candidates, the kinematic properties of the D-meson decay, and the PID information of the daughter tracks to discriminate between signal and background candidates. The features used to train the model are reported in Table 5.2. The number in parenthesis after $n\sigma$ indicates the prong number.

The invariant mass of the candidate and its p_{T} are not used to train the model. Exploiting such variables would introduce a bias in the final result, as the model would be trained to select candidates within a specific invariant mass region (that of $\mathrm{D_s^+}$ and $\mathrm{D^+}$ mesons) or p_{T}. This would affect both the selection of the candidates and the p_{T} distribution of the final sample, leading to a biased p_{T}-differential yield. However, some of the variables used to train the model may be correlated with the invariant mass of the candidate, and the ML may learn to discriminate the signal from the background by exploiting this correlation with the $\mathrm{D_s^+}$ meson mass and transverse momentum, rather than the physical properties of the signal and background. To exclude this possibility, the correlation between the features used to train the model is studied. To quantitatively describe the correlation between the variables, the Pearson correlation coefficient ρ is evaluated for each pair of variables. It is defined as the ratio between the covariance of two variables x and y and the product of their standard deviations, $\rho(x, y) = \mathrm{cov}(x, y)/(\sigma_x \sigma_y)$. It expresses the strength and direction of a linear correlation between two variables, ranging from $\rho =$

Table 5.2: Candidate features used to train the ML model.

Variable
$\cos\theta_p$
$\cos\theta_p^{xy}$
Decay length
Decay length XY
Candidate impact parameter XY
$\lvert\cos^3\theta'(\mathrm{K})\rvert$
Prong 0 impact parameter XY
Prong 1 impact parameter XY
Prong 2 impact parameter XY
$n\sigma^{\pi}_{\mathrm{comb}}(0)$
$n\sigma^{\pi}_{\mathrm{comb}}(1)$
$n\sigma^{\pi}_{\mathrm{comb}}(2)$
$n\sigma^{\mathrm{K}}_{\mathrm{comb}}(0)$
$n\sigma^{\mathrm{K}}_{\mathrm{comb}}(1)$
$n\sigma^{\mathrm{K}}_{\mathrm{comb}}(2)$

1 (perfect positive linear correlation) to $\rho = -1$ (perfect negative linear relationship). $\rho = 0$ indicates the absence of linear correlation.

The correlation matrix of the features used to train the model is shown in Fig. 5.2 for the prompt $\mathrm{D_s^+}$, non-prompt $\mathrm{D_s^+}$ and background classes, in the $2 < p_{\mathrm{T}} < 3$ GeV/c. The correlation with the invariant mass and the transverse momentum is also reported. The Pearson coefficient is encoded in the colour of the cell, with red indicating a positive correlation, blue a negative correlation, and grey no correlation. The correlation matrix shows that the variables used to train the model are not correlated with the invariant mass of the candidate, suggesting that a ML model should not modify the invariant-mass distribution of the selected candidates.

Variables carrying the same physical information, such as those related to the candidate decay length, pointing angle, and impact parameter, are strongly correlated among each other, as expected. Different degrees of correlation between the same variable pairs are observed for the different classes. The ML model can exploit these differences to discriminate between the three classes of candidates.

5.2.2 Boosted Decision Trees

Once the training dataset has been composed and the features have been selected, the ML architecture has to be chosen. Several algorithms are available, each with its own strengths and weaknesses. The choice of the algorithm depends on the specific problem to solve, the size of the dataset, and the computational resources available.

Boosted decision trees [133, 134] (BDTs) are a family of machine learning algorithms employed in different fields, including high-energy physics. Their building blocks are decision trees, which are a versatile type of supervised learning algorithm

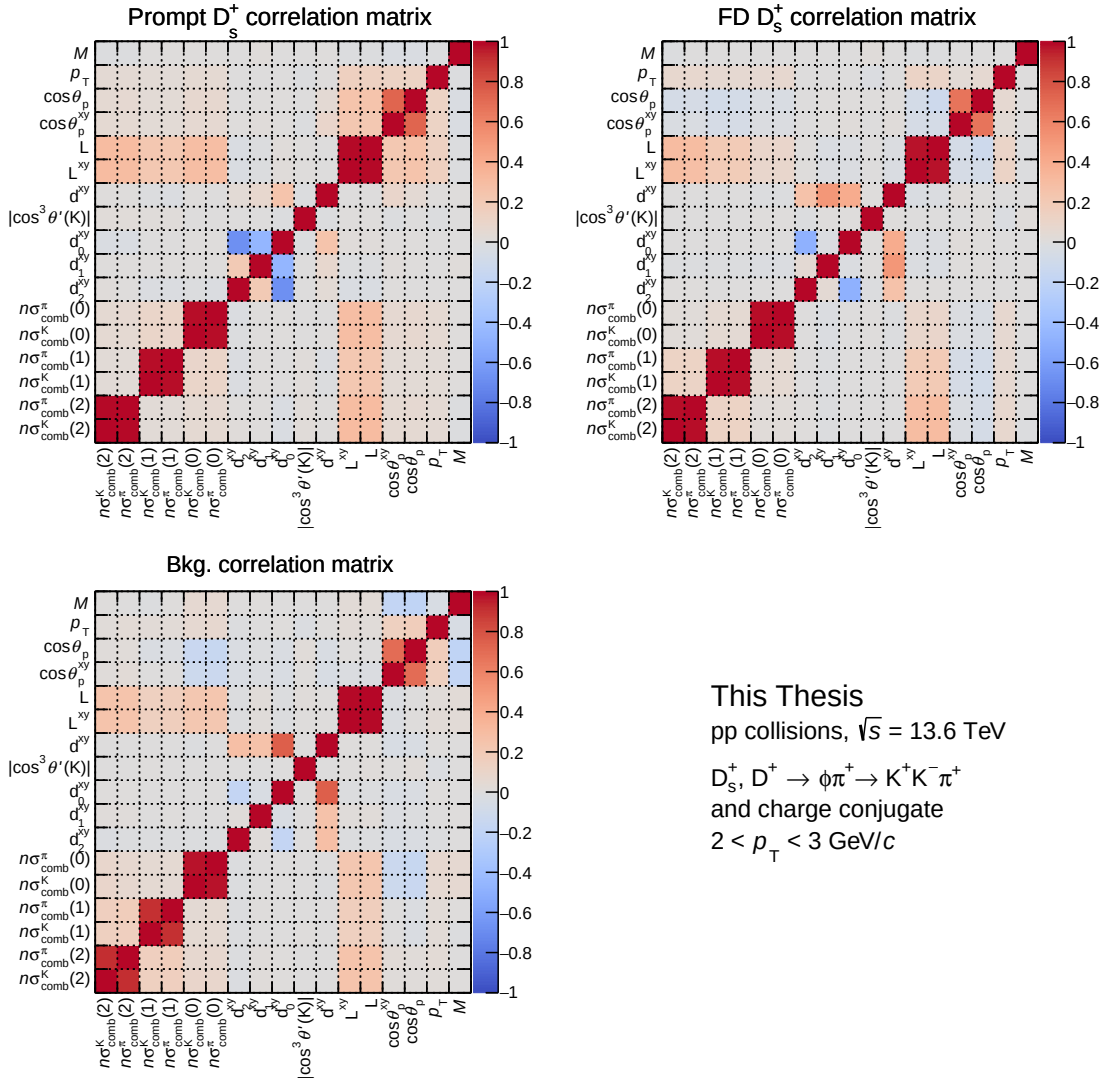


Figure 5.2: Correlation matrix of the features used to train the ML model for prompt D_s^+ (top-left), non-prompt D_s^+ (top-right), and background (bottom-left) candidates in the $2 < p_T < 3$ GeV/c interval. The correlation with the invariant mass and the transverse momentum is also reported. The Pearson coefficient is encoded in the colour of the cell, with red indicating a positive correlation, blue a negative correlation, and grey no linear correlation.

that can be used for both classification and regression tasks. A decision tree is made of many *nodes*, each containing conditions that split the data into two [135] or more [136] children nodes. The first node of the tree, which receives all the data, is called *root* node, while nodes that do not further split the data are called *leaves*, and contain the output of the tree. The model is trained by considering the Gini index, which measures the impurity of the node:

$$G = 1 - \sum_{i=1}^{n} p_i^2 \quad ,$$

where p_i is the fraction of samples in the node that belong to the class i. The Gini index therefore provides an indication of the quality of the split, with $G = 0$ indicating a perfect split. A commonly used algorithm to build *binary* decision trees (i.e., each node contains binary-output conditions, and is split into two children nodes) is the *Classification And Regression Tree* (CART) algorithm [135], which recursively splits the dataset into subsets based on a single feature k and a threshold t_k that minimises the impurity of the subsets (weighted by their size). The cost function that the algorithm tries to minimise is given by

$$J(k, t_k) = \frac{m_{\text{left}}}{m} G_{\text{left}} + \frac{m_{\text{right}}}{m} G_{\text{right}} \quad ,$$

where m_{left} and m_{right} are the number of samples in the left and right nodes, respectively, summing up to the total number of samples m, and G_{left} and G_{right} are the Gini indices of the left and right nodes. The tree is grown until a stopping criterion is met, such as a maximum depth, a minimum number of samples in a node, or a minimum impurity decrease. These are all hyperparameters that can be tuned to optimise the model's performance.

Given their simplicity, decision trees are fairly easy to interpret, and are often called *white-box* models (in contrast to BDTs and neural networks, where the decision-making process is less transparent, therefore called *black-box* models). An additional strength of decision trees is that they require very little data preparation, e.g., they do not require feature scaling or centering, making them a very powerful yet simple tool for data analysis. However, they are prone to overfitting, as they can grow to a large depth, capturing the noise in the training data. To mitigate this issue, their depth can be constrained, but this may lead to a model with limited discrimination power. To build a robust model with a good discrimination power, ensemble methods may be used. Several decision trees can be trained, and the final prediction is obtained by combining the outcome of all the trees.

XGBoost

In this work, the Extrame Gradient Boosting [137] (XGBoost) Boosted Decision Trees (BDT) algorithm is used. It has achieved state-of-the-art results in a number of machine learning and data mining challenges (for example in Ref. [138]). In addition, this algorithm, which is available as an open-source package, can be easily parallelised on CPUs and GPUs [139], thereby reducing the training and application time.

The term *boosting* refers to any ensemble method combining several weak learners into a strong learner. The general idea of most boosting methods is to train many predictors sequentially, each trying to correct its predecessor [140]. The function estimate $\widehat{f}(x)$ is parametrised with an additive functional form:

$$\widehat{f}(x) = \sum_{\mathrm{k}=1}^{\mathrm{M}} \widehat{f}_{\mathrm{k}}(x) \quad ,$$

where M is the number of iterations, $\widehat{f}_0(x)$ is the initial prediction, and $\widehat{f}_{\mathrm{i}}(x)$ is the function increment at the i-th iteration, also called *boost*. To reduce the loss function, a new weak learner, whose functional form is parametrised as $h(x, \theta)$, can be added to the ensemble:

$$\widehat{f}_{\mathrm{t}}(x) \leftarrow \widehat{f}_{\mathrm{t}-1}(x) + \rho_{\mathrm{t}} h(x, \theta_{\mathrm{t}}) \quad .$$

ρ_t is the step size, which is optimised for each iteration t, together with the parameters θ_{t} of the weak learner:

$$(\rho_{\mathrm{t}}, \theta_{\mathrm{t}}) = \arg\min_{\rho,\theta} \sum_{i=1}^{\mathrm{N}} L\left(y_i, \widehat{f}_{\mathrm{t}-1}(x_i) + \rho h(x_i, \theta)\right) \quad ,$$

where L is the loss function, and y_i is the true label of the i-th example. Despite having a well-defined set of equations for minimising the loss function, the optimisation of the parameters is not trivial, as the loss function is non-convex and the search space is high-dimensional. Therefore, the optimisation is usually performed using a gradient-based algorithm [133, 141], where $h(x, \theta_{\mathrm{t}})$ is chosen as the most parallel function to the negative gradient of the loss function with respect to the previous prediction $g_{\mathrm{t}}(x)$:

$$g_{\mathrm{t}}(x) = E_{\mathbf{y}} \left[\frac{\partial L(\mathbf{y}, \widehat{f}_{\mathrm{t}-1}(x))}{\partial \widehat{f}_{\mathrm{t}-1}(x)} \middle| x \right] \quad ,$$

where $E_{\mathbf{y}}$ is the expectation over the true labels. The parameters are then optimised by minimising the difference between the negative gradient and the weak learner prediction:

$$(\rho_{\mathrm{t}}, \theta_{\mathrm{t}}) = \arg\min_{\rho,\theta} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left[-g_{\mathrm{t}} - \rho h(x_{\mathrm{i}}, \theta)\right]^2 \quad .$$

Through the iterative addition of weak learners, the loss function is reduced and the model learns the complex patterns of data. The final prediction is obtained by summing the predictions of all the weak learners. In the XGBoost algorithm, the weak learners are decision trees. The output consists of a numerical score for each class, ranging from 0 to 1 and summing up to unity. Each score represents the confidence of the model in the prediction, which can be interpreted as the probability of the example belonging to that class.

5.2.3 Tuning the model's hyperparameters

The XGBoost algorithm has several hyperparameters [142] that can be tuned to optimise the model's performance. The most important hyperparameters are:

- `eta` or `learning_rate`, which is the step size shrinkage of the gradient descent algorithm. To reduce the risk of overfitting, this factor multiplies the weak-learner prediction ($\rho_t h(x_i, \theta) \rightarrow \texttt{eta} \cdot \rho_t h(x_i, \theta)$), and is usually set to a small value, such as 0.3;
- `max_depth`, which is the maximum depth of the tree. A large depth can lead to overfitting, while a small depth can lead to a model with limited discrimination power;
- `n_estimators`, which defines the number of trees to train. A large number of weak learners can lead to overfitting, while a small number can lead to a model with limited discrimination power. Usually, the number of weak learners is set to around 1000;
- `subsample`, which is the fraction of the training data to be used to train each tree at each iteration;
- `min_child_weight`, which is the minimum sum of instance weight needed in a child. It is related to the purity in a node, and it is used to stop the tree growth;
- `colsample_bytree`, which is the fraction of features to be used to train each tree at each iteration;
- `tree_method`, which defines the algorithm used to build the trees. The `hist` option uses an optimised histogram-based algorithm and is usually the fastest.

The hyperparameters are optimised using the Optuna framework [143], which proved to be a powerful tool thanks to its state-of-the-art algorithms for sampling the hyperparameter space and for efficiently pruning unpromising trials. The Tree-Structured Parzen Estimator [144], is used in this Thesis. It is a Bayesian optimisation algorithm able to explore the hyperparameter space efficiently. The aim of a Bayesian optimisation is to maximise (or minimise, depending on the task) an objective function $f(\mathbf{x})$ by iteratively sampling a bounded hyperparameter space, χ. The algorithm builds a probabilistic model of the objective function, and uses it to decide which hyperparameters to sample next. The model is updated at each iteration, and the hyperparameters that are most likely to improve the model's performance are sampled. The Optuna algorithm is also able to prune unpromising trials, reducing the computational cost of the optimisation. The optimisation is performed using a 5-fold cross-validation, and the hyperparameters that maximise the macro-averaged one-vs-one ROC AUC metric (described in detail in Sec. 5.2.4) are chosen as the optimal configuration. The hyperparameters optimised for the XGBoost model are reported in Table 5.3. An additional hyperparameter, `lambda`, which is the L_2 regularisation term, is also optimised. It helps to prevent overfitting by penalising overly complex models. The optimal hyperparameters are then used to train the model on the full training dataset.

Table 5.3: Optimised hyperparameter configuration for the p_T bins considered in the model training.

Hyper-parameter	p_T interval (GeV/c)								
	0–1.5	1.5–2	2–3	3–4	4–5	5–6	6–8	8–12	12–24
max_depth	3	3	3	3	3	3	3	3	3
learning_rate	0.04	0.068	0.065	0.10	0.091	0.84	0.070	0.046	0.030
n_estimators	473	339	1352	909	1256	1392	1142	1437	1188
min_child_weight	1	3	10	10	10	9	3	7	5
subsample	0.87	0.95	0.84	0.85	0.95	0.85	0.81	0.94	0.88
colsample_bytree	0.91	0.98	0.90	0.98	0.96	0.95	0.88	0.96	0.89
lambda	8.0×10^{-4}	4.8×10^{-4}	9.1×10^{-4}	1.4×10^{-4}	3.0×10^{-4}	3.2×10^{-4}	1.9×10^{-4}	9.8×10^{-4}	6.7×10^{-4}
tree_method	hist	hist	hist	hist	hist	hist	hist	hist	hist

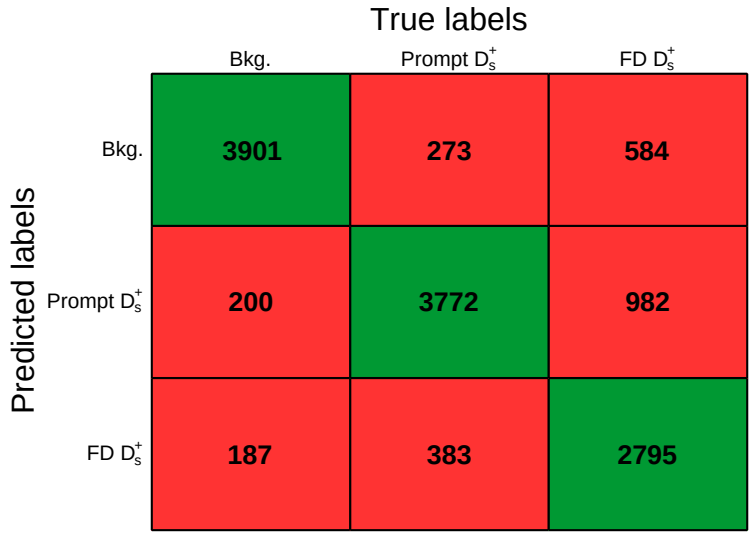


Figure 5.3: Confusion matrix for the BDT model trained in the $2 < p_T < 3$ GeV/c interval. Candidates are classified as the class with the highest score.

5.2.4 Evaluation of the model's performance

After training the model, its performance is evaluated on the test dataset. The model's performance can be assessed using a *confusion matrix*, which summarises the number of examples for a given class (the true label) that are classified by the model as belonging to any of the available classes (the predicted label). A good model should provide a high number of correctly-classified examples (reported on the diagonal of the confusion matrix), and a low number of misclassified examples (off-diagonal elements of the confusion matrix). The confusion matrix also allows an understanding of which classes are more difficult to classify, and which classes are more likely to be confused with each other. An example of a confusion matrix is shown in Fig. 5.3 for the XGBoost model trained on the $2 < p_T < 3$ GeV/c interval.

Despite providing a lot of information on the model's performance, more concise metrics of the model's performance are usually used, for a more direct comparison between different models. In addition, the confusion matrix provides a threshold-dependent measure of the model's performance, as the classification threshold can be varied to increase the number of correctly classified signal candidates at the expense

of the number of correctly classified background candidates, and vice versa.

In binary classification tasks, where only two classes are available (a positive and a negative class), several metrics can be defined from the elements of the confusion matrix. The 2×2 confusion matrix contains four entries: the true positives (TP), which are the number of correctly classified positive candidates, the false positives (FP), which are the number of negative candidates being mistakenly classified as positives, and the analogously defined true negatives (TN) and false negatives (FN). One of the most used tools for binary classifiers is the *Receiver Operating Characteristic* (ROC) curve, which represents the true positive rate (TPR) against the false positive rate (FPR) for different threshold values. The TPR is the fraction of correctly classified positive candidates ($\mathrm{TPR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$), while the FPR is the fraction of incorrectly classified negative candidates ($\mathrm{FPR} = \mathrm{FP}/(\mathrm{FP} + \mathrm{TN})$). If positive candidates are selected as those with a score greater than a certain threshold t, then when $t = 0$ all candidates are classified as positive, and both the TPR and FPR will be equal to 1. On the other hand, if $t = 1$, no candidate is classified as positive, and the TPR and FPR will both equal 0. The ROC *Area Under the Curve* (AUC), is used to measure the model's ability to discriminate between positive and negative candidates, for any given threshold. The ROC AUC ranges from 0 to 1. A random classifier has a ROC AUC of 0.5, while a perfect classifier has a ROC AUC of 1. The ROC AUC is a threshold-independent measure of the model's performance, and is often used to compare different models.

In a multiclass classification task, where more than two classes are available, a generalisation of the ROC curve and the ROC AUC metric is required. In this case, the *One-vs-One* ROC curve can be defined as a plot of the TPR against the FPR for a single pair of classes. The One-vs-One ROC AUC can be averaged to the *macro-averaged* One-vs-One ROC AUC, which is the average of the ROC AUC for each pair of classes and can provide a measurement of the model's ability to discriminate between all the classes.

The One-vs-One ROC curves for the model trained on the $2 < p_{\mathrm{T}} < 3$ GeV/c interval are shown in Fig. 5.4. The ROC AUC is calculated for each class pair, and is reported in the legend. The metric is evaluated on both the training and test sets to test the model's generalisation power. The model's performance is excellent, with a macro-averaged One-vs-One ROC AUC value very close to 1. In addition, little overfitting is observed, as the ROC AUC values for the training and test sets are similar. The model is then used to select D_s^+ and D^+ candidates from the full dataset.

In addition to the ROC AUC, the model's performance can be evaluated by studying the distribution of the probability of belonging to a given class assigned to labelleled candidates. The score distribution for the model trained on the $2 < p_{\mathrm{T}} < 3$ GeV/c interval is shown in Fig. 5.5 for the background, prompt D_s^+, and non-prompt D_s^+ classes. For each class, the score distribution is shown for candidates belonging to the different classes and to both the training and test sets. The distribution of the background score provides interesting information on the model's performance. The score distribution for the background candidates peaks at high values, while the score distribution for the signal candidates (both prompt and non prompt D_s^+) peaks at low values. This highlights that the model has effec-

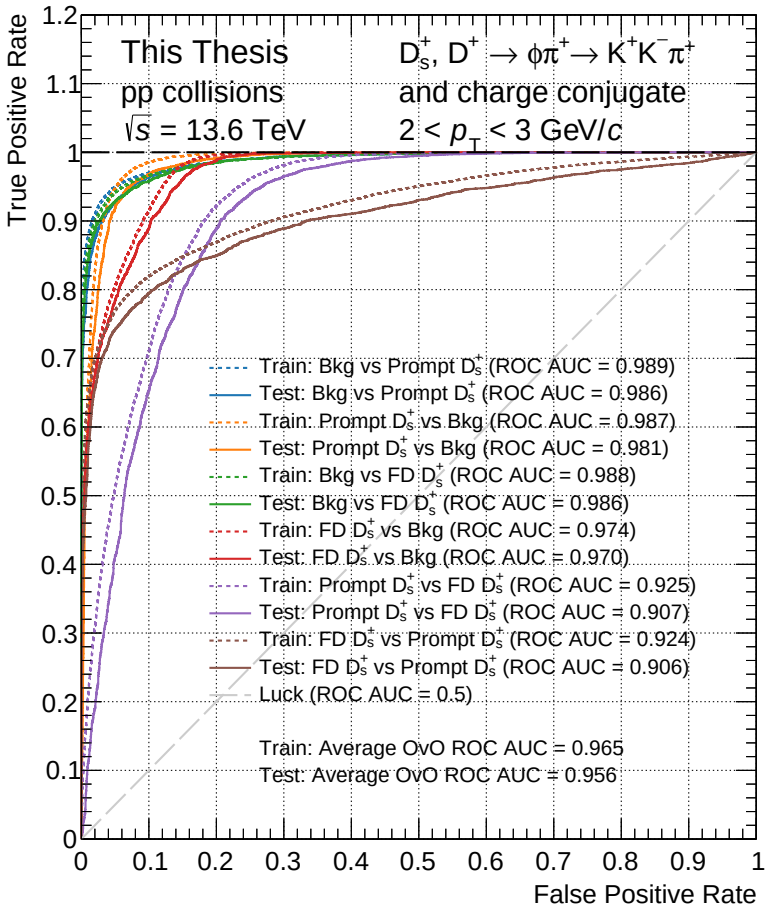


Figure 5.4: ROC curves for the model trained on the $2 < p_T < 3$ GeV/c interval. The One-vs-One ROC AUC metric is calculated for each class pair and reported in the legend.

tively learned to discriminate between signal and background candidates, with good separation power. Furthermore, the score distributions for the training and test sets are fairly similar, indicating that the model generalises well to unseen data. Since non-prompt D_s^+ present a very displaced topology due to the large lifetime of beauty-hadrons, the separation between non-prompt D_s^+ and background is noticeable in the non-prompt D_s^+ score distribution. Finally, as the displacement of prompt candidates lies in between the non-prompt D_s^+ and background, the separation between the three classes is less pronounced in the prompt D_s^+ score distribution, where the prompt D_s^+ distribution does not peak at values around one.

5.2.5 Interpretation of the model's output: Feature importance

The usage of ML algorithms usually provides a better performance in terms of signal-to-background separation, but it also introduces a level of complexity in the selection process. One of the most difficult aspects of using ML models is the interpretation

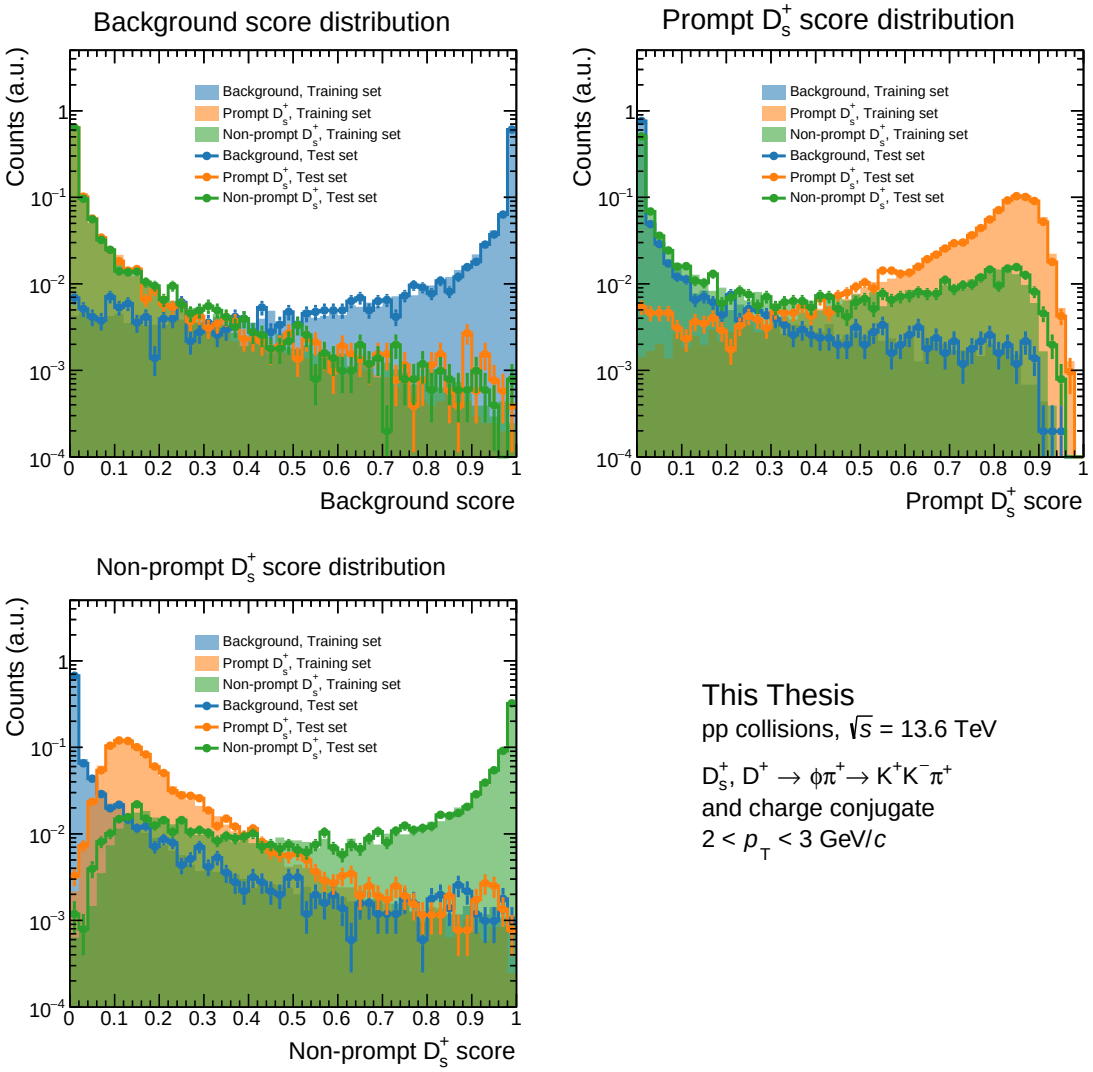


Figure 5.5: Score distribution for the model trained on the $2 < p_{\mathrm{T}} < 3$ GeV/c interval. The score distribution related to the probability of belonging to the background, prompt D_s^+, and non-prompt D_s^+ classes is shown. For each class, the score distribution is shown for candidates belonging to the different classes and to both the training (filled area) and test sets (markers).

of their output. To understand how the model makes its decisions, the feature importance can be studied. This allows the understanding of which features are more important for the model's decision-making process, and the optimisation of the feature selection. In addition, the feature importance can be used to check whether the model is learning on the correct features in terms of the physics of the problem.

One of the most used algorithms for feature importance studies is the SHapley Additive exPlanations [145] (SHAP) algorithm. SHAP is a game-theoretic approach to explain the output of any machine learning model. It is based on the Shapley value [146] from cooperative game theory, which requires retraining the model on all feature subsets $\mathcal{S} \subseteq \mathcal{F}$, where $\mathcal{F}$ is the set of all features. An importance value is assigned to each feature, representing the effect on the model prediction of including that feature. To compute this effect, a model $\widehat{f}_{\mathcal{S}\cup\{i\}}$ is trained with that feature present, and another model $\widehat{f}_{\mathcal{S}}$ is trained with the feature withheld. Then, predictions from the two models are compared on the current input $\widehat{f}_{\mathcal{S}\cup\{i\}}(x) - \widehat{f}_{\mathcal{S}}$. The Shapley values are then computed as the weighted average of all possible differences:

$$\phi_{\mathrm{i}} = \sum_{\mathcal{S}\subseteq\mathcal{F}\setminus\{i\}} \frac{|\mathcal{S}|!(|\mathcal{F}| - |\mathcal{S}| - 1)!}{|\mathcal{F}|!} \left[\widehat{f}_{\mathcal{S}\cup\{i\}}(x) - \widehat{f}_{\mathcal{S}}(x)\right] \quad .$$

Since most models cannot handle arbitrary patterns of missing input values, $\widehat{f}(z_{\mathcal{S}})$ is approximated with $E[\widehat{f}(z)|z_{\mathcal{S}}]$, where $z_{\mathcal{S}}$ is the input missing the features in $\mathcal{S}$. SHAP values therefore explain how to get from the base value $E[\widehat{f}(z)]$ that would be predicted if no features were known to the output $\widehat{f}(x)$.

A beeswarm-style SHAP feature importance plot for the prompt D_s^+ probability predicted by the model trained in the $2 < p_{\mathrm{T}} < 3$ GeV/c interval is shown in Fig. 5.6. The most important features are the cosine of pointing angle, the decay length, the decay length in the XY plane, the cosine cubed of the K-π angle in the KK rest frame, and the PID information on the prong 1. As discussed in Chapter 4, the first three features are related to the displaced topology of D mesons, and are therefore expected to be the most important variables in the model decisions. It is also expected that the prong 1 PID information resulted as the most important PID variable, as this is the opposite sign track, which is always a kaon in the considered decay channel. On the contrary, prongs 0 and 2 could be either kaons or pions, resulting in a lower importance of the PID information for these prongs.

5.2.6 Optimisation of the model selection

Once the model performance has been validated, a set of selection criteria has to be chosen to select the candidates. This is a crucial step of the analysis, as it will define the signal extraction efficiency and the background contamination. Since the model's output consists of a score related to the probability of belonging to each class, and the three probabilities sum up to unity, the selection criteria have a total of two degrees of freedom. A first selection is applied on the maximum probability to be a background candidate, and rejects most of the contamination from the combinatorial background. The second one is applied on the minimum probability

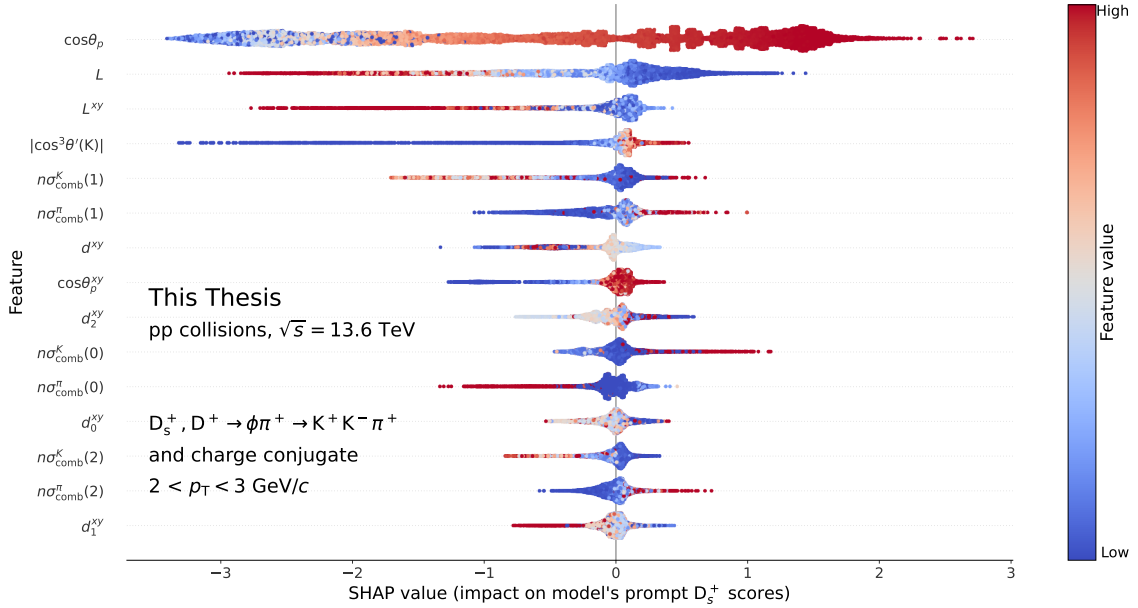


Figure 5.6: SHAP feature importance for the XGBoost model trained on the $2 < p_{\mathrm{T}} < 3$ GeV/c interval.

of being a prompt $\mathrm{D_s^+}$ candidate, and suppresses the signal contribution arising from non-prompt $\mathrm{D_s^+}$ candidates.

A first indication of the optimal selection criteria can be obtained by studying the model's output distributions. A good working point can be chosen as the point where a good separation between the three classes is achieved. For this analysis, however, the statistical significance of the signal is used to define the optimal selection criteria. For each p_{T} interval of the analysis, the signal S and the background B are evaluated by fitting the invariant mass distribution of candidates passing the different ML selections to be considered in the optimisation process. Only a subset of the full dataset, corresponding to $\sim 3\%$ of the available data sample is used in this process. The working point is chosen as the set of selections maximising the statistical significance $S/\sqrt{S+B}$. The efficiency of the selection is checked to ensure that it is kept at sufficiently high levels, to reduce possible biases in the final results due to possible imperfections in the MC description of the data. Usually, the optimisation of the statistical significance is avoided in the optimisation of the selection criteria, as it can lead to a bias in the final results. However, in this case, the optimisation is performed on a very small fraction of the data thanks to the large dataset available, and the bias is expected to be negligible.

A result of the working point optimisation is presented in Fig. 5.7 for the $4.0 < p_{\mathrm{T}} < 4.5$ GeV/c interval. The statistical significance of the signal is shown as a function of the BDT output score threshold for the probability of being a prompt $\mathrm{D_s^+}$ and a background candidate. The chosen set of selection criteria is shown with a green cross.

The optimal selection criteria for each p_{T} interval considered in the analysis are reported in Table 5.4.

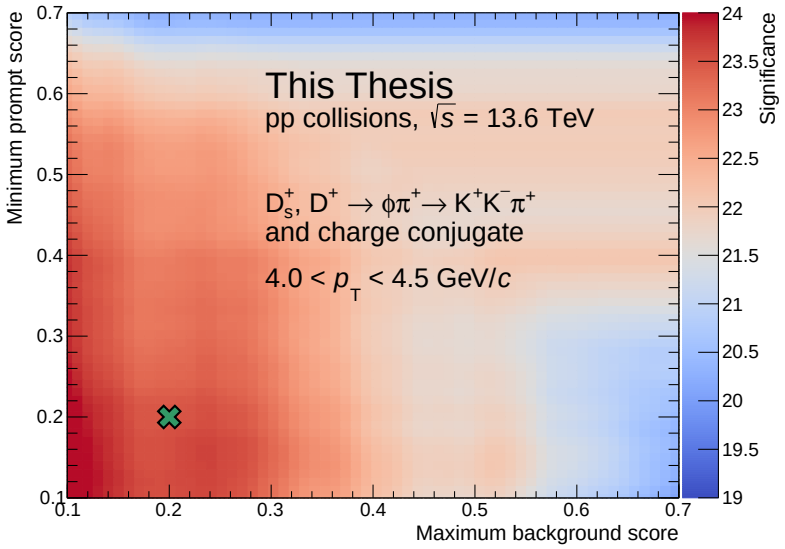


Figure 5.7: Statistical significance of the signal as a function of the selection criteria applied to the model output for the $4.0 < p_{\mathrm{T}} < 4.5$ GeV/c interval. The chosen set of selection criteria is shown with a green cross.

5.3 D_s^+ and D^+ raw-yield extraction

After the working point for the BDT algorithm has been defined, the raw yields of D_s^+ and D^+ mesons are extracted in each p_{T} interval. They are defined as the sum of particles and antiparticles and are measured in the $0.5 < p_{\mathrm{T}} < 24$ GeV/c interval. The raw yield is extracted by fitting the invariant mass distribution of the selected candidates.

Table 5.4: Selection criteria applied to enhance the significance of the D_s^+ meson contribution in the p_{T} bins of the analysis.

p_{T} interval (GeV/c)	Probability to be background <	Probability to be prompt D_s^+ >
0.5−1.0	0.01	0.2
1.0−1.5	0.05	0.2
1.5−2.0	0.15	0.2
2.0−2.5	0.25	0.2
2.5−3.0	0.3	0.2
3.0−3.5	0.2	0.2
3.5−4.0	0.2	0.2
4.0−4.5	0.2	0.2
4.5−5.0	0.2	0.2
5.0−5.5	0.3	0.2
5.5−6.0	0.3	0.25
6−8	0.45	0.2
8−12	0.5	0.2
12−24	0.55	0.2

In several analyses performed by the ALICE Collaboration during the Run 2 data-taking period [51, 147, 148], the raw yield of D_s^+ mesons was extracted by fitting the invariant mass distribution of selected candidates with a probability density function constructed as the sum of a function describing the shape of the combinatorial background (usually an exponential function or a low-order (< 3) polynomial) and of two gaussian distributions to model the D_s^+ and D^+ peaks. The raw yields for the two D-meson species are then obtained by integrating the signal function.

Figure 5.8 shows the fit to the invariant mass distribution of the selected candidates using the approach described above. Due to the concavity-changing shape of the background, the function chosen to describe the background is a third-order polynomial. On top of the previously-unobserved peculiar shape of the background, the fitting function is not able to describe the data accurately between the two peaks, overestimating the data in this invariant mass region.

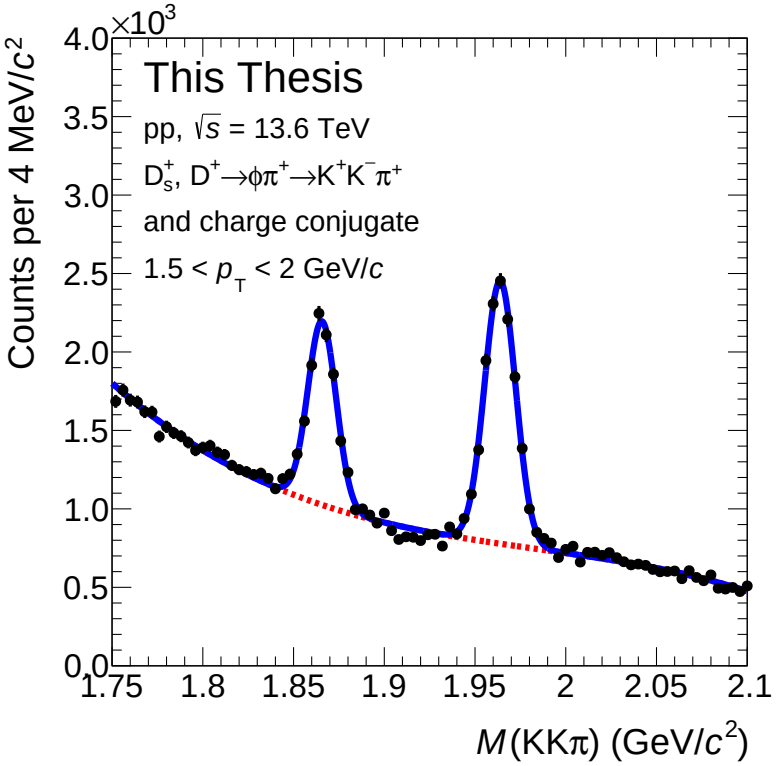


Figure 5.8: Fit to the invariant mass distribution of selected candidates in the $2 < p_{\mathrm{T}} < 3$ GeV/c interval. The fit function is shown as a solid line, while the signal and background components are shown as dashed lines.

These two features can be understood as due to the fact that the background does not solely arise from the combination of independent tracks, i.e. the combinatorial background. Other physics processes can contribute to the contamination of the data sample, giving rise to a *correlated background*. One such process is the decay of D^+ mesons into the $D^+ \to \pi^+K^-\pi^+$ decay channel, where one of the pions is misidentified as a kaon. Despite being suppressed by the applied ML selections, this contribution can give rise to a noticeable contribution due to the large BR of 9.38×10^{-2} [4].

To validate this hypothesis, a simulation of D^+ decays into $D^+ \to \pi^+K^-\pi^+$ was run using PYTHIA 8 [73]. Ten billion D^+ mesons were produced with a uniformly distributed p_T spectrum in the $0 < p_T < 24$ GeV/c interval. The D^+ mesons were then forced to decay into the $D^+ \to \pi^+K^-\pi^+$ decay channel. For each decay, two invariant masses were evaluated, by assigning the kaon mass to one of the two pions, while the correct masses were assigned to the other two prongs. The invariant mass distributions obtained from the simulation are shown in Fig. 5.9 for different p_T intervals. The distribution is characterised by a peak at ~ 2 GeV/c^2, well between the fit range used to extract the raw yield of D_s^+ and D^+ mesons in Fig 5.8. In addition, the invariant mass distribution evolves with the D^+ p_T, with a tail towards higher invariant masses for higher p_T values. This follows naturally from the kinematic properties of the decay.

To account for this contribution, a template fit is included in the fit function. This method involves fitting a distribution using a predefined template whose shape is fixed, and the only adjustable parameter is the normalization of the function. The shape of the template is taken from the same simulation used to train the BDT model, as described in Sec. 5.2.1, where $D^+ \to \pi^+K^-\pi^+$ mesons reconstructed as $D^+ \to K^+K^-\pi^+$ are selected. The distribution is fixed to that obtained before applying any ML models, as it was studied that the applied ML selections do not affect the shape of the correlated background. This allows for reducing statistical fluctuations in the template shape. The final fit function is then constructed as the sum of a parabolic function to describe the combinatorial background, the template function described above for the correlated $D^+ \to \pi^+K^-\pi^+$ background source, and two gaussian functions to describe the D_s^+ and D^+ peaks. The signal parameters (mean, width, normalisation), as well as the parameters of the combinatorial background and the normalisation of the template fit are left free in the fit in the $p_T < 8$ GeV/c interval. Because of the observed momentum resolution worsening at higher p_T, and because of the lower statistics available in the high-p_T intervals, the width of the D^+ signal peak is fixed to that of the D_s^+ divided by a factor 1.2, which is the observed ratio of the peak widths at low p_T, which presents a flat trend with p_T. In these high-p_T intervals, a first fit is performed by keeping both peak widths as free parameters. Then, the D^+ peak width is fixed to that of the D_s^+ divided by 1.2, and the fit is repeated, keeping the remaining parameters free. Additionally, in order to correctly describe the background shape, the fit range is extended to $1.73 < M(KK\pi) < 2.15$ GeV/c^2 from $p_T > 6$ GeV/c, while the narrower invariant mass window $1.75 < M(KK\pi) < 2.1$ GeV/c^2 is fitted at lower p_T. The invariant-mass bin width has been fixed to 2 MeV/c^2. The fits to the invariant mass distributions of selected D-meson candidates are performed using the `flarefly` package [149], which provides a flexible pythonic interface for performing fits.

The fit to the invariant mass distribution of candidates passing the ML selections is shown in Fig. 5.10 for the $2.0 < p_T < 2.5$ GeV/c in the left panel, and for the $5.0 < p_T < 5.5$ GeV/c interval in the right panel. The total fit function is shown as a solid blue line, the signal contributions are shown as filled green and azure areas for D^+ and D_s^+ mesons, respectively, the combinatorial background is represented with a solid red line, while the correlated background is shown as a dashed violet

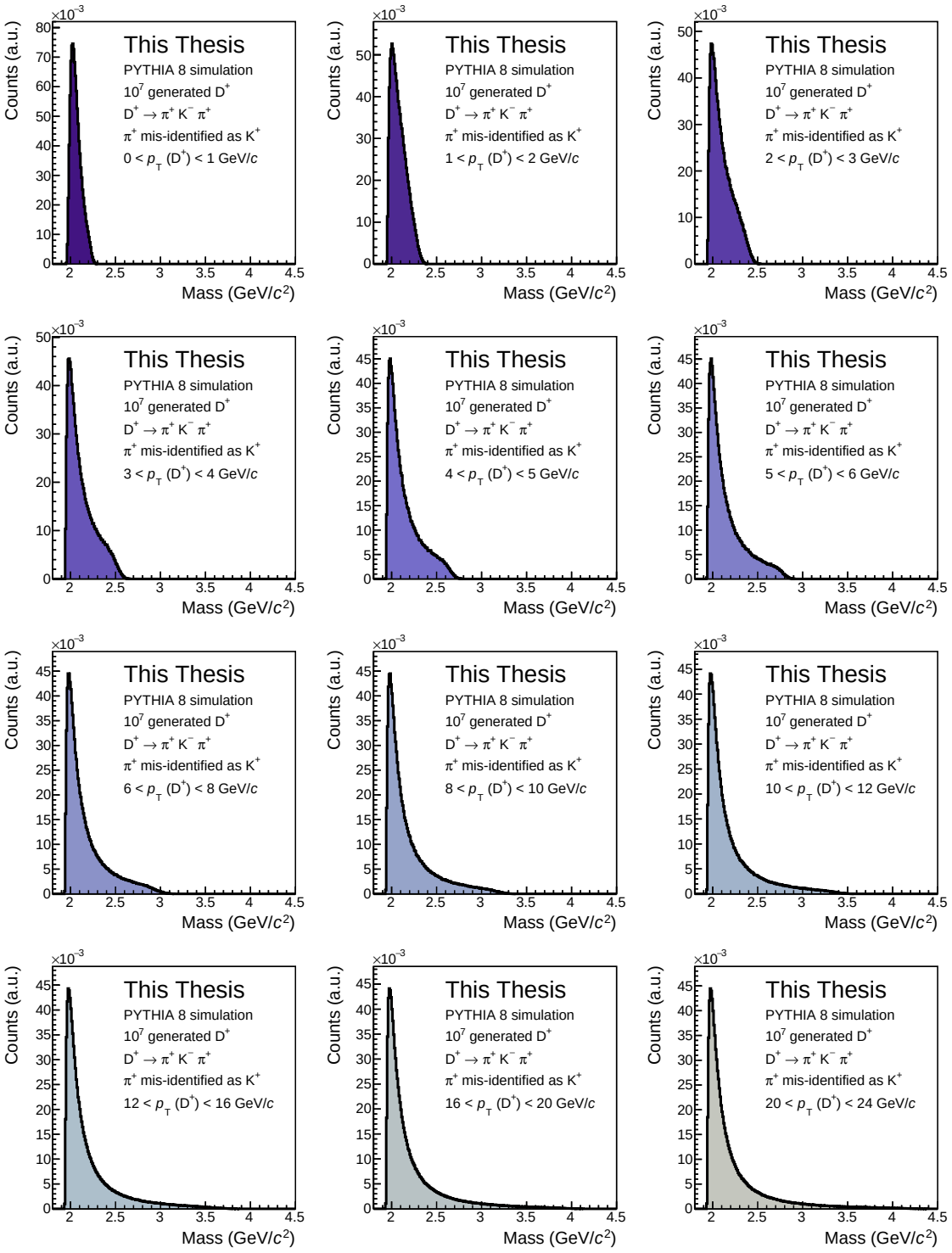


Figure 5.9: Invariant mass distribution simulated decays of D^+ mesons into $D^+ \to \pi^+K^-\pi^+$, where one of the pions produced in the decay is misidentified as a kaon. Distributions for different p_T intervals are shown.

line. The fit is able to describe the data accurately, as can be deduced from the distribution of the difference between the data and the background fit function, shown in the bottom panels for the two p_T bins. Figure 5.10 also shows that the contribution of the correlated $D^+ \to \pi^+K^-\pi^+$ background evolves with p_T, with a larger contribution at lower p_T values. The raw yield of D_s^+ and D^+ mesons is then extracted by integrating the signal function.

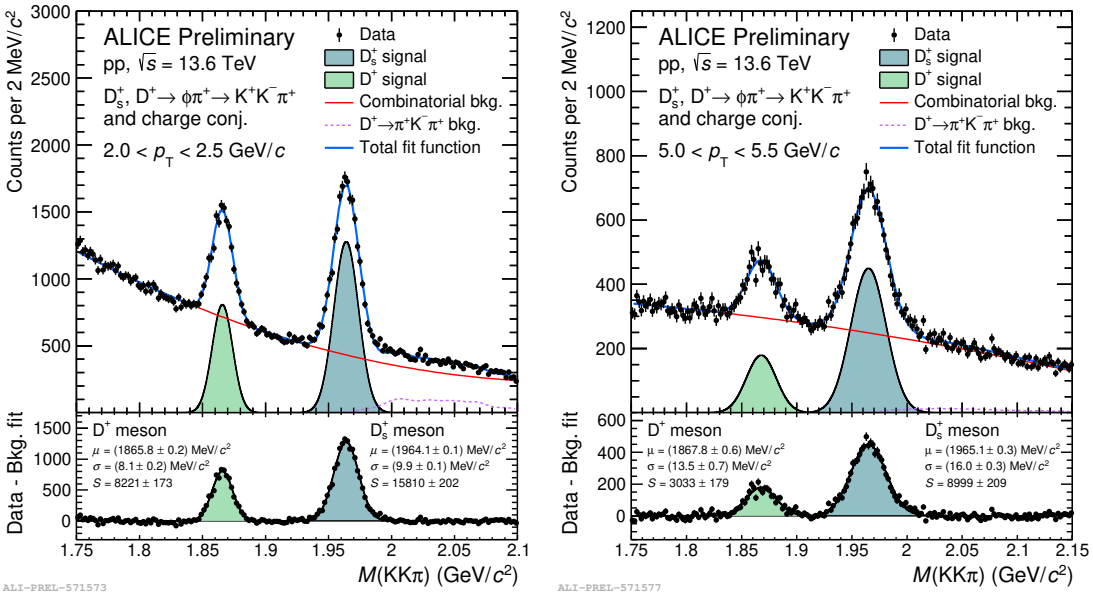


Figure 5.10: Fit to the invariant mass distribution of selected candidate in the 2.0¡p_T ¡2.5 GeV/c (left) and 5.0¡p_T ¡5.5 GeV/c (right) intervals. The total fit function is shown as a solid blue line, while background components are shown as solid red lines (correlated background) and dashed violet lines (dashed violet lines). The signal contributions are shown as filled green and azure areas for D^+ and D_s^+ mesons, respectively. The bottom panels show the distribution of the difference between the data and the background fit function.

The extraction of the raw yields is affected by several arbitrary choices, for example, the functional description of the background, the choice of the fit range, and the invariant-mass bin width. Changes in these choices can lead to variations in the extracted raw yields. To estimate the effect of such arbitrary choice in the final observable (the D_s^+/D^+ production-yield ratio), and estimate a systematic uncertainty associated with the raw yield extraction procedure, the fit is repeated several times by varying the fit range, the bin width, and the functional form of the background. In the low p_T region, the minimum mass is varied between 1.71 and 1.75 GeV/c^2, while the maximum mass is varied between 2.13 and 2.19 GeV/c^2. The functions considered to describe the background are a second-order polynomial and an exponential function. At higher p_T, where the signal peaks become broader, the minimum mass is varied between 1.71 and 1.75 GeV/c^2, while the maximum mass is varied between 2.13 and 2.19 GeV/c^2. The bin width is varied between 1 and 4 MeV/c^2 across the studied p_T interval. At $p_T > 8$ GeV/c^2, where the D^+ peak width is fixed to that of the D_s^+ divided by 1.2, the peak width of the D^+

meson is changed by varying the dividing factor by ±10%. At lower p_T, all the fit parameters are left free, as for the central case. For each possible combination of these variations, the signal is extracted for the two D-meson species, and the ratio between the two is calculated. The systematic uncertainty is then defined as the sum in quadrature of the standard deviation of the distribution of the D_s^+/D^+ raw-yield ratio and of the difference between the central raw-yield ratio and the mean of the obtained distribution.

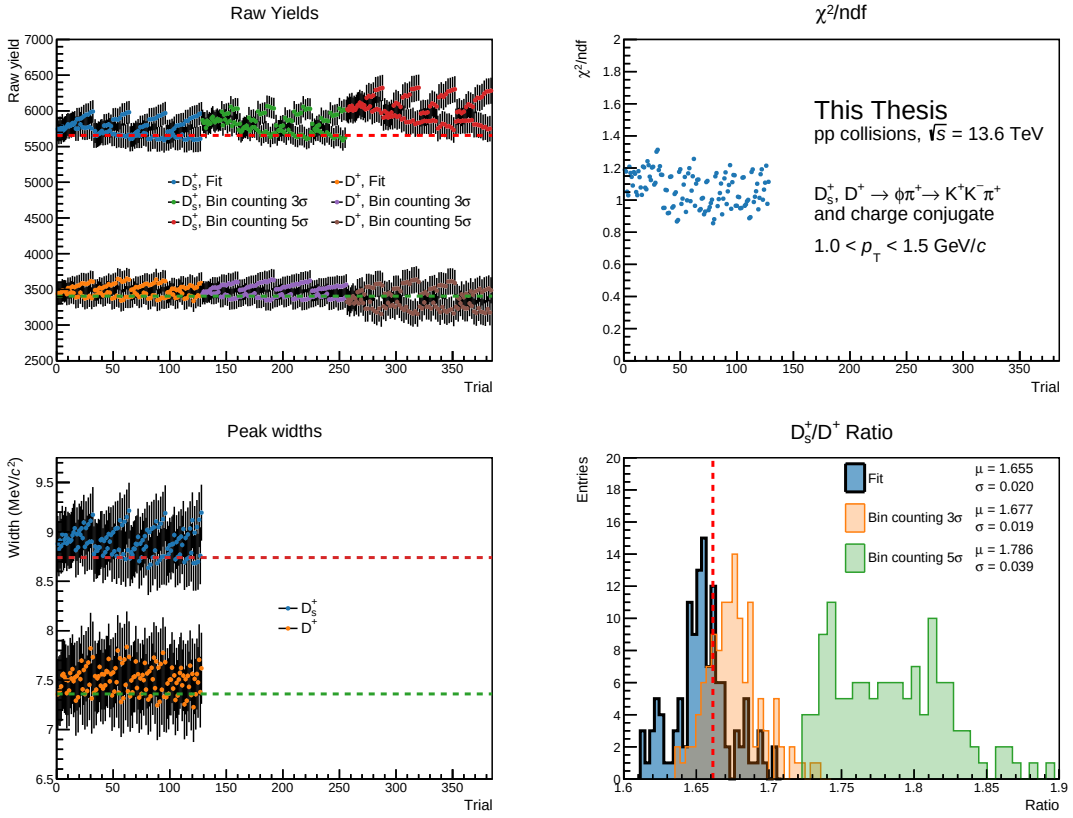


Figure 5.11: Results from the multi-trial approach employed for estimating the systematic uncertainty related to the raw yield extraction in the $1.0 < p_T < 1.5$ GeV/c interval.

The result of this multi-trial approach for the evaluation of the systematic on the raw-yields extraction is shown in Fig. 5.11 for the $1.0 < p_T < 1.5$ GeV/c interval. In the top-left panel, the raw yields extracted from the fit to the invariant mass distribution are reported for the D_s^+ and D^+ mesons for the different trials. As a cross-check, the raw yields are also extracted by summing the counts of a distribution obtained by subtracting the background fit function from the invariant mass distribution of the candidates passing the ML selections. The bin contents are summed within 3 and 5 standard deviations from the peak mean. The extracted raw yields are stable within uncertainty across all the trials and the raw yield definition for the D^+ meson. For the D_s^+ meson, the 5σ bin counting method presents higher raw yields when compared to the ones obtained by integrating the signal function or with a 3σ bin counting method. However, this is considered as related to fluctuation

in the invariant mass distribution rather than a systematic shift of the raw yields due to a different definition of the observable. The raw yields extracted with the default configuration described above are shown as dashed lines. In the top-right panel, the χ^2/ndf of the fit to the invariant mass distribution is shown for the different trials. As a quality check, trials with a χ^2/ndf greater than 2 are discarded. These results illustrate that the fit function is able to describe the data accurately, independently of the configuration of the fit parameters. In the bottom-left panel, the peak widths for D_s^+ and D^+ are reported, and present a stable behaviour across the different trials. The peak widths for the default configuration are also reported as dashed lines. The bottom-right panel shows the distribution of the D_s^+/D^+ raw-yield ratio for the different trials, obtained using the three methods previously introduced. The 5σ bin counting method presents a higher raw-yield ratio when compared to the other two methods, because of the higher D_s^+ yields. The systematic uncertainty is defined as the sum in quadrature of the standard deviation of the distribution of D_s^+/D^+ raw-yield ratio obtained by integrating the signal function and the difference between the central value reported with a dashed red line and the mean of this distribution. This quantity ranges from 1% to 10% of the central D_s^+/D^+ raw-yield ratio, depending on the considered p_T interval. The systematic uncertainty is then assigned after smoothing the p_T dependence of the obtained values. The assigned systematic uncertainty is reported in Table 5.5.

Table 5.5: Systematic uncertainty on the raw-yield extraction for the D_s^+ and D^+ mesons.

p_T (GeV/c)	$\sqrt{\text{RMS}^2 + \Delta^2}$/(central D_s^+/D^+) (%)	Assigned systematic uncertainty (%)
0.5−1	4	3
1−1.5	1	3
1.5−2	2	3
2−2.5	3	3
2.5−3	3	3
3−3.5	3	3
3.5−4	3	3
4−4.5	6	5
4.5−5	7	5
5−5.5	3	5
5.5−6	5	5
6−8	8	8
8−12	9	9
12−24	10	10

Bibliography

[1] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.* **8** (1964) 214–215.

[2] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, pp. 22–101. 2, 1964.

[3] H. Fritzsch and M. Gell-Mann, "Current algebra: Quarks and what else?", *eConf* **C720906V2** (1972) 135–165, arXiv:hep-ph/0208010.

[4] **Particle Data Group** Collaboration, R. L. Workman and Others, "Review of Particle Physics", *PTEP* **2022** (2022) 083C01.

[5] F. Herzog, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, "The five-loop beta function of Yang-Mills theory with fermions", *JHEP* **02** (2017) 090, arXiv:1701.01404 [hep-ph].

[6] K. Johnson, "The M.I.T. Bag Model", *Acta Phys. Polon. B* **6** (1975) 865.

[7] R. P. Feynman, "Space-time approach to non-relativistic quantum mechanics", *Rev. Mod. Phys.* **20** (Apr, 1948) 367–387. https://link.aps.org/doi/10.1103/RevModPhys.20.367.

[8] **BMW** Collaboration, S. Durr *et al.*, "Ab-Initio Determination of Light Hadron Masses", *Science* **322** (2008) 1224–1227, arXiv:0906.3599 [hep-lat].

[9] **HotQCD** Collaboration, A. Bazavov *et al.*, "Equation of state in (2+1)-flavor QCD", *Phys. Rev. D* **90** (2014) 094503, arXiv:1407.6387 [hep-lat].

[10] S. Borsanyi, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, and K. K. Szabo, "Full result for the QCD equation of state with 2+1 flavors", *Phys. Lett. B* **730** (2014) 99–104, arXiv:1309.5258 [hep-lat].

[11] CERN Courier, "Quark-matter fireballs hashed out in Protvino", 2021.

[12] **ALICE** Collaboration, "The ALICE experiment – A journey through QCD", arXiv:2211.04384 [nucl-ex].

[13] **NA38, NA50** Collaboration, M. C. Abreu *et al.*, "Dimuon and charm production in nucleus-nucleus collisions at the CERN SPS", *Eur. Phys. J. C* **14** (2000) 443–455.

[14] **NA50** Collaboration, M. C. Abreu *et al.*, "Anomalous J / psi suppression in Pb - Pb interactions at 158 GeV/c per nucleon", *Phys. Lett. B* **410** (1997) 337–343.

[15] R. Nouicer, "Formation of Dense Partonic Matter in High Energy Heavy-Ion Collisions: Highlights of RHIC Results", in *Advanced Studies Institute on Symmetries and Spin (SPIN-Praha-2008)*. 1, 2009. arXiv:0901.0910 [nucl-ex].

[16] R. Stock, "Relativistic Nucleus-Nucleus Collisions and the QCD Matter Phase Diagram", arXiv:0807.1610 [nucl-ex].

[17] **ALICE** Collaboration, C. Loizides, "Charged-particle multiplicity and transverse energy in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV with ALICE", *J. Phys. G* **38** (2011) 124040, arXiv:1106.6324 [nucl-ex].

[18] J. D. Bjorken, "Highly Relativistic Nucleus-Nucleus Collisions: The Central Rapidity Region", *Phys. Rev. D* **27** (1983) 140–151.

[19] **PHENIX** Collaboration, K. Adcox *et al.*, "Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration", *Nucl. Phys. A* **757** (2005) 184–283, arXiv:nucl-ex/0410003.

[20] **ALICE** Collaboration, K. Aamodt *et al.*, "Two-pion Bose-Einstein correlations in central Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV", *Phys. Lett. B* **696** (2011) 328–337, arXiv:1012.4035 [nucl-ex].

[21] A. Andronic, P. Braun-Munzinger, K. Redlich, and J. Stachel, "Decoding the phase structure of QCD via particle production at high energy", *Nature* **561** (2018) 321–330, arXiv:1710.09425 [nucl-th].

[22] P. Braun-Munzinger, K. Redlich, and J. Stachel, "Particle production in heavy ion collisions", arXiv:nucl-th/0304013.

[23] S. Wheaton and J. Cleymans, "THERMUS: A Thermal model package for ROOT", *Comput. Phys. Commun.* **180** (2009) 84–106, arXiv:hep-ph/0407174.

[24] G. Torrieri, S. Steinke, W. Broniowski, W. Florkowski, J. Letessier, and J. Rafelski, "SHARE: Statistical hadronization with resonances", *Comput. Phys. Commun.* **167** (2005) 229–251, arXiv:nucl-th/0404083.

[25] V. Vovchenko and H. Stoecker, "Thermal-FIST: A package for heavy-ion collisions and hadronic equation of state", *Comput. Phys. Commun.* **244** (2019) 295–310, arXiv:1901.05249 [nucl-th].

[26] A. Andronic, P. Braun-Munzinger, and J. Stachel, "Hadron production in central nucleus-nucleus collisions at chemical freeze-out", *Nucl. Phys. A* **772** (2006) 167–199, arXiv:nucl-th/0511071.

[27] **ALICE** Collaboration, S. Acharya *et al.*, "Measurements of chemical potentials in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", arXiv:2311.13332 [nucl-ex].

[28] **ALICE** Collaboration, B. Abelev *et al.*, "Centrality dependence of π, K, p production in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV", *Phys. Rev. C* **88** (2013) 044910, arXiv:1303.0737 [hep-ex].

[29] E. Schnedermann, J. Sollfrank, and U. W. Heinz, "Thermal phenomenology of hadrons from 200-A/GeV S+S collisions", *Phys. Rev. C* **48** (1993) 2462–2475, arXiv:nucl-th/9307020.

[30] F. Cooper and G. Frye, "Comment on the Single Particle Distribution in the Hydrodynamic and Statistical Thermodynamic Models of Multiparticle Production", *Phys. Rev. D* **10** (1974) 186.

[31] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, "Glauber modeling in high energy nuclear collisions", *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 205–243, arXiv:nucl-ex/0701025.

[32] J. S. Moreland, J. E. Bernhard, and S. A. Bass, "Alternative ansatz to wounded nucleon and binary collision scaling in high-energy nuclear collisions", *Phys. Rev. C* **92** (2015) 011901, arXiv:1412.4708 [nucl-th].

[33] T. Lappi and L. McLerran, "Some features of the glasma", *Nucl. Phys. A* **772** (2006) 200–212, arXiv:hep-ph/0602189.

[34] Z.-W. Lin, C. M. Ko, B.-A. Li, B. Zhang, and S. Pal, "A Multi-phase transport model for relativistic heavy ion collisions", *Phys. Rev. C* **72** (2005) 064901, arXiv:nucl-th/0411110.

[35] **ALICE** Collaboration, S. Acharya *et al.*, "Transverse momentum spectra and nuclear modification factors of charged particles in pp, p-Pb and Pb-Pb collisions at the LHC", *JHEP* **11** (2018) 013, arXiv:1802.09145 [nucl-ex].

[36] M. Arneodo, "Nuclear effects in structure functions", *Phys. Rept.* **240** (1994) 301–393.

[37] B. Z. Kopeliovich, J. Nemchik, A. Schafer, and A. V. Tarasov, "Cronin effect in hadron production off nuclei", *Phys. Rev. Lett.* **88** (2002) 232303, arXiv:hep-ph/0201010.

[38] **PHENIX** Collaboration, U. A. Acharya *et al.*, "Systematic study of nuclear effects in p +Al, p +Au, d +Au, and ^{3}He+Au collisions at $\sqrt{s_{NN}} = 200$ GeV using π^0 production", *Phys. Rev. C* **105** (2022) 064902, arXiv:2111.05756 [nucl-ex].

[39] **PHENIX** Collaboration, S. Afanasiev *et al.*, "Measurement of Direct Photons in Au+Au Collisions at $\sqrt{s_{NN}} = 200$ GeV", *Phys. Rev. Lett.* **109** (2012) 152302, arXiv:1205.5759 [nucl-ex].

[40] R. Baier, Y. L. Dokshitzer, A. H. Mueller, S. Peigne, and D. Schiff, "Radiative energy loss of high-energy quarks and gluons in a finite volume quark - gluon plasma", *Nucl. Phys. B* **483** (1997) 291–320, arXiv:hep-ph/9607355.

[41] **STAR** Collaboration, J. Adams *et al.*, "Evidence from d + Au measurements for final state suppression of high p_{T} hadrons in Au+Au collisions at RHIC", *Phys. Rev. Lett.* **91** (2003) 072304, arXiv:nucl-ex/0306024.

[42] J. Rafelski and B. Muller, "Strangeness Production in the Quark - Gluon Plasma", *Phys. Rev. Lett.* **48** (1982) 1066. [Erratum: Phys.Rev.Lett. 56, 2334 (1986)].

[43] K. Redlich and A. Tounsi, "Strangeness enhancement and energy dependence in heavy ion collisions", *Eur. Phys. J. C* **24** (2002) 589–594, arXiv:hep-ph/0111261.

[44] **STAR** Collaboration, B. I. Abelev *et al.*, "Enhanced strange baryon production in Au + Au collisions compared to p + p at $\sqrt{s_{\mathrm{NN}}} = 200$ GeV", *Phys. Rev. C* **77** (2008) 044908, arXiv:0705.2511 [nucl-ex].

[45] **ALICE** Collaboration, B. B. Abelev *et al.*, "Multi-strange baryon production at mid-rapidity in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV", *Phys. Lett. B* **728** (2014) 216–227, arXiv:1307.5543 [nucl-ex]. [Erratum: Phys.Lett.B 734, 409–410 (2014)].

[46] **ALICE** Collaboration, J. Adam *et al.*, "Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions", *Nature Phys.* **13** (2017) 535–539, arXiv:1606.07424 [nucl-ex].

[47] **ALICE** Collaboration, B. B. Abelev *et al.*, "Multiplicity Dependence of Pion, Kaon, Proton and Lambda Production in p-Pb Collisions at $\sqrt{s_{NN}} =$ 5.02 TeV", *Phys. Lett. B* **728** (2014) 25–38, arXiv:1307.6796 [nucl-ex].

[48] **ALICE** Collaboration, J. Adam *et al.*, "Multi-strange baryon production in p-Pb collisions at $\sqrt{s_{\mathbf{NN}}} = 5.02$ TeV", *Phys. Lett. B* **758** (2016) 389–401, arXiv:1512.07227 [nucl-ex].

[49] **ALICE** Collaboration, "ALICE Figure." https://alice-figure.web.cern.ch/.

[50] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD", *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, arXiv:hep-ph/0409313.

[51] **ALICE** Collaboration, S. Acharya *et al.*, "Measurement of beauty and charm production in pp collisions at $\sqrt{s} = 5.02$ TeV via non-prompt and prompt D mesons", *JHEP* **05** (2021) 220, arXiv:2102.13601 [nucl-ex].

[52] M. Cacciari, M. Greco, and P. Nason, "The p_T spectrum in heavy-flavour hadroproduction.", *JHEP* **05** (1998) 007, arXiv:hep-ph/9803400.

[53] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, arXiv:1410.3012 [hep-ph].

[54] J. I. Friedman and H. W. Kendall, "Deep inelastic electron scattering", *Ann. Rev. Nucl. Part. Sci.* **22** (1972) 203–254.

[55] J. D. Bjorken, "Asymptotic Sum Rules at Infinite Momentum", *Phys. Rev.* **179** (1969) 1547–1553.

[56] R. P. Feynman, "Very high-energy collisions of hadrons", *Phys. Rev. Lett.* **23** (1969) 1415–1417.

[57] C. G. Callan, Jr. and D. J. Gross, "High-energy electroproduction and the constitution of the electric current", *Phys. Rev. Lett.* **22** (1969) 156–159.

[58] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438–450.

[59] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.", *Sov. Phys. JETP* **46** (1977) 641–653.

[60] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys. B* **126** (1977) 298–318.

[61] **NNPDF** Collaboration, R. D. Ball *et al.*, "The path to proton structure at 1% accuracy", *Eur. Phys. J. C* **82** (2022) 428, arXiv:2109.02653 [hep-ph].

[62] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, "New parton distribution functions from a global analysis of quantum chromodynamics", *Phys. Rev. D* **93** (2016) 033006, arXiv:1506.07443 [hep-ph].

[63] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, "Parton distributions in the LHC era: MMHT 2014 PDFs", *Eur. Phys. J. C* **75** (2015) 204, arXiv:1412.3989 [hep-ph].

[64] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, and B. Mistlberger, "Higgs Boson Gluon-Fusion Production in QCD at Three Loops", *Phys. Rev. Lett.* **114** (2015) 212001, arXiv:1503.06056 [hep-ph].

[65] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog, A. Lazopoulos, and B. Mistlberger, "High precision determination of the gluon fusion Higgs boson cross-section at the LHC", *JHEP* **05** (2016) 058, arXiv:1602.00695 [hep-ph].

[66] **NNPDF** Collaboration, V. Bertone, N. P. Hartland, E. R. Nocera, J. Rojo, and L. Rottoli, "Charged hadron fragmentation functions from collider data", *Eur. Phys. J. C* **78** (2018) 651, arXiv:1807.03310 [hep-ph]. [Erratum: Eur.Phys.J.C 84, 155 (2024)].

[67] D. de Florian, R. Sassot, and M. Stratmann, "Global analysis of fragmentation functions for pions and kaons and their uncertainties", *Phys. Rev. D* **75** (2007) 114010, arXiv:hep-ph/0703242.

[68] B. A. Kniehl, G. Kramer, and B. Potter, "Fragmentation functions for pions, kaons, and protons at next-to-leading order", *Nucl. Phys. B* **582** (2000) 514–536, arXiv:hep-ph/0010289.

[69] V. V. Sudakov, "Vertex parts at very high-energies in quantum electrodynamics", *Sov. Phys. JETP* **3** (1956) 65–71.

[70] Y. I. Azimov, Y. L. Dokshitzer, V. A. Khoze, and S. I. Troyan, "Similarity of Parton and Hadron Spectra in QCD Jets", *Z. Phys. C* **27** (1985) 65–72.

[71] R. D. Field and R. P. Feynman, "Quark Elastic Scattering as a Source of High Transverse Momentum Mesons", *Phys. Rev. D* **15** (1977) 2590–2616.

[72] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145.

[73] C. Bierlich *et al.*, "A comprehensive guide to the physics and usage of PYTHIA 8.3", *SciPost Phys. Codeb.* **2022** (2022) 8, arXiv:2203.11601 [hep-ph].

[74] E. Eichten, K. Gottfried, T. Kinoshita, J. B. Kogut, K. D. Lane, and T.-M. Yan, "The Spectrum of Charmonium", *Phys. Rev. Lett.* **34** (1975) 369–372. [Erratum: Phys.Rev.Lett. 36, 1276 (1976)].

[75] S. Ferreres-Solé and T. Sjöstrand, "The space–time structure of hadronization in the Lund model", *Eur. Phys. J. C* **78** (2018) 983, arXiv:1808.04619 [hep-ph].

[76] T. Sjostrand, "Jet Fragmentation of Nearby Partons", *Nucl. Phys. B* **248** (1984) 469–502.

[77] **ALICE** Collaboration, S. Acharya *et al.*, "First measurement of $\Lambda_{\rm c}^{+}$ production down to $p_{\rm T} = 0$ in pp and p–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", *Phys. Rev. C* **107** (2023) 064901, arXiv:2211.14032 [nucl-ex].

[78] J. R. Christiansen and P. Z. Skands, "String Formation Beyond Leading Colour", *JHEP* **08** (2015) 003, arXiv:1505.01681 [hep-ph].

[79] B. R. Webber, "A QCD Model for Jet Fragmentation Including Soft Gluon Interference", *Nucl. Phys. B* **238** (1984) 492–528.

[80] M. Bahr *et al.*, "Herwig++ Physics and Manual", *Eur. Phys. J. C* **58** (2008) 639–707, arXiv:0803.0883 [hep-ph].

[81] D. Amati and G. Veneziano, "Preconfinement as a Property of Perturbative QCD", *Phys. Lett. B* **83** (1979) 87–92.

[82] M. H. Seymour and M. Marx, "Monte Carlo Event Generators", in *69th Scottish Universities Summer School in Physics: LHC Physics*, pp. 287–319. 4, 2013. arXiv:1304.6677 [hep-ph].

[83] **ALICE** Collaboration, S. Acharya *et al.*, "Prompt D^0, D^+, and D^{*+} production in Pb–Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", *JHEP* **01** (2022) 174, arXiv:2110.09420 [nucl-ex].

[84] **ALICE** Collaboration, S. Acharya *et al.*, "Λ_c^+ production in *pp* and in *p*-Pb collisions at $\sqrt{s_{NN}}$=5.02 TeV", *Phys. Rev. C* **104** (2021) 054905, arXiv:2011.06079 [nucl-ex].

[85] **ALICE** Collaboration, S. Acharya *et al.*, "Measurement of the production cross section of prompt $\Xi_{\rm c}^0$ baryons in p−Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", arXiv:2405.14538 [nucl-ex].

[86] **ALICE** Collaboration, S. Acharya *et al.*, "Λ_c^+ Production and Baryon-to-Meson Ratios in pp and p-Pb Collisions at $\sqrt{s_{NN}}$=5.02 TeV at the LHC", *Phys. Rev. Lett.* **127** (2021) 202301, arXiv:2011.06078 [nucl-ex].

[87] **ALICE** Collaboration, S. Acharya *et al.*, "Measurement of the Cross Sections of Ξ_c^0 and Ξ_c^+ Baryons and of the Branching-Fraction Ratio $\mathrm{BR}(\Xi_c^0 \to \Xi^- e^+ \nu_e)/\mathrm{BR}(\Xi_c^0 \to \Xi^- \pi^+)$ in pp collisions at 13 TeV", *Phys. Rev. Lett.* **127** (2021) 272001, arXiv:2105.05187 [nucl-ex].

[88] **CMS** Collaboration, V. Khachatryan *et al.*, "Measurement of long-range near-side two-particle angular correlations in pp collisions at $\sqrt{s}$ =13 TeV", *Phys. Rev. Lett.* **116** (2016) 172302, arXiv:1510.03068 [nucl-ex].

[89] J. Song, H.-h. Li, and F.-l. Shao, "New feature of low p_T charm quark hadronization in *pp* collisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C* **78** (2018) 344, arXiv:1801.09402 [hep-ph].

[90] V. Minissale, S. Plumari, and V. Greco, "Charm hadrons in pp collisions at LHC energy within a coalescence plus fragmentation approach", *Phys. Lett. B* **821** (2021) 136622, arXiv:2012.12001 [hep-ph].

[91] A. Beraudo, A. De Pace, D. Pablos, F. Prino, M. Monteno, and M. Nardi, "Heavy-flavor transport and hadronization in pp collisions", *Phys. Rev. D* **109** (2024) L011501, arXiv:2306.02152 [hep-ph].

[92] K. Werner, "Core-corona separation in ultra-relativistic heavy ion collisions", *Phys. Rev. Lett.* **98** (2007) 152301, arXiv:0704.1270 [nucl-th].

[93] S. Porteboeuf and K. Werner, "Generation of complete events containing very high-p(T) jets", *Eur. Phys. J. C* **62** (2009) 145–150.

[94] M. He and R. Rapp, "Charm-Baryon Production in Proton-Proton Collisions", *Phys. Lett. B* **795** (2019) 117–121, arXiv:1902.08889 [nucl-th].

[95] M. He and R. Rapp, "Bottom Hadrochemistry in High-Energy Hadronic Collisions", *Phys. Rev. Lett.* **131** (2023) 012301, arXiv:2209.13419 [hep-ph].

[96] D. Ebert, R. N. Faustov, and V. O. Galkin, "Spectroscopy and Regge trajectories of heavy baryons in the relativistic quark-diquark picture", *Phys. Rev. D* **84** (2011) 014025, arXiv:1105.0583 [hep-ph].

[97] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", *Eur. Phys. J. C* **74** (2014) 3024, arXiv:1404.5630 [hep-ph].

[98] J. Bellm *et al.*, "Herwig 7.0/Herwig++ 3.0 release note", *Eur. Phys. J. C* **76** (2016) 196, arXiv:1512.01178 [hep-ph].

[99] S. Frixione, P. Nason, and G. Ridolfi, "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *JHEP* **09** (2007) 126, arXiv:0707.3088 [hep-ph].

[100] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, "Collinear subtractions in hadroproduction of heavy quarks", *Eur. Phys. J. C* **41** (2005) 199–212, arXiv:hep-ph/0502194.

[101] C. Bierlich, G. Gustafson, L. Lönnblad, and A. Tarasov, "Effects of Overlapping Strings in pp Collisions", *JHEP* **03** (2015) 148, arXiv:1412.6259 [hep-ph].

[102] **ALICE** Collaboration, K. Aamodt *et al.*, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002.

[103] **ALICE** Collaboration, "ALICE upgrades during the LHC Long Shutdown 2", arXiv:2302.01238 [physics.ins-det].

[104] "LHC Machine", *JINST* **3** (2008) S08001.

[105] S. Myers, "The LEP collider, from design to approval and commissioning",.

[106] E. Lopienska, "The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022",. https://cds.cern.ch/record/2800984. General Photo.

[107] **ATLAS** Collaboration, G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[108] **CMS** Collaboration, S. Chatrchyan *et al.*, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[109] F. Sauli, "GEM: A new concept for electron amplification in gas detectors", *Nucl. Instrum. Meth. A* **386** (1997) 531–534.

[110] Apache Arrow, "Apache Arrow, a cross-language development platform for in-memory analytics." https://arrow.apache.org/.

[111] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", *Nucl. Instrum. Meth. A* **389** (1997) 81–86.

[112] **ALICE** Collaboration, G. Aglieri Rinella, "The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System", *Nucl. Instrum. Meth. A* **845** (2017) 583–587.

[113] W. Snoeys, "CMOS monolithic active pixel sensors for high energy physics", *Nucl. Instrum. Meth. A* **765** (2014) 167–171.

[114] S. Senyukov, J. Baudot, A. Besson, G. Claus, L. Cousin, A. Dorokhov, W. Dulinski, M. Goffe, C. Hu-Guo, and M. Winter, "Charged particle detection performances of CMOS pixel sensors produced in a 0.18νm process with a high resistivity epitaxial layer", *Nucl. Instrum. Meth. A* **730** (2013) 115–118, arXiv:1301.0515 [physics.ins-det].

[115] **CMS** Collaboration, V. Karimäki, "The CMS tracker system project: Technical Design Report",.

[116] G. Aad *et al.*, "ATLAS pixel detector electronics and sensors", *JINST* **3** (2008) P07007.

[117] **LHCb** Collaboration, I. Bediaga, "LHCb VELO Upgrade Technical Design Report",.

[118] Y. Wang and Y. Yu, "Multigap Resistive Plate Chambers for Time of Flight Applications", *Appl. Sciences* **11** (2020) 111.

[119] **ALICE** Collaboration, S. Acharya *et al.*, "Measurement of D-meson production at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C* **77** (2017) 550, arXiv:1702.00766 [hep-ex].

[120] **ZEUS** Collaboration, J. Breitweg *et al.*, "Measurement of inclusive D+-(s) photoproduction at HERA", *Phys. Lett. B* **481** (2000) 213–227, arXiv:hep-ex/0003018.

[121] W. Blum, L. Rolandi, and W. Riegler, *Particle detection with drift chambers*. Particle Acceleration and Detection. 2008.

[122] **ALICE** Collaboration, B. B. Abelev *et al.*, "Performance of the ALICE Experiment at the CERN LHC", *Int. J. Mod. Phys. A* **29** (2014) 1430044, arXiv:1402.4476 [nucl-ex].

[123] A. L. Samuel, "Some studies in machine learning using the game of checkers", *IBM Journal of Research and Development* **3** (1959) 210–229.

[124] T. M. Mitchell, "Machine learning", 1997.

[125] OpenAI, "GPT-4 Technical Report", *arXiv e-prints* (Mar., 2023) arXiv:2303.08774, arXiv:2303.08774 [cs.CL].

[126] A. Mao, M. Mohri, and Y. Zhong, "Cross-entropy loss functions: Theoretical analysis and applications", in *International Conference on Machine Learning*, pp. 23803–23828, PMLR. 2023.

[127] J. Kiefer and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function", *The Annals of Mathematical Statistics* **23** (1952) 462 – 466. https://doi.org/10.1214/aoms/1177729392.

[128] P. I. Frazier, "A tutorial on bayesian optimization", *arXiv preprint arXiv:1807.02811* (2018) .

[129] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms", *Advances in neural information processing systems* **25** (2012) .

[130] J. Mockus, "The bayesian approach to global optimization", in *System Modeling and Optimization: Proceedings of the 10th IFIP Conference New York City, USA, August 31–September 4, 1981*, pp. 473–481, Springer. 2005.

[131] M. Stone, "Cross-validatory choice and assessment of statistical predictions", *Journal of the royal statistical society: Series B (Methodological)* **36** (1974) 111–133.

[132] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4–a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303.

[133] J. H. Friedman, "Greedy function approximation: a gradient boosting machine", *Annals of statistics* (2001) 1189–1232.

[134] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of computer and system sciences* **55** (1997) 119–139.

[135] L. Breiman, *Classification and regression trees*. Routledge, 2017.

[136] J. R. Quinlan, "Induction of decision trees", *Machine learning* **1** (1986) 81–106.

[137] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system", *CoRR* **abs/1603.02754** (2016) , 1603.02754. http://arxiv.org/abs/1603.02754.

[138] B. Kegl, CecileGermain, ChallengeAdmin, ClaireAdam, D. Rousseau, Djabbz, fradav, G. Cowan, Isabelle, and joycenv, "Higgs boson machine learning challenge", 2014. https://kaggle.com/competitions/higgs-boson.

[139] R. Mitchell and E. Frank, "Accelerating the xgboost algorithm using gpu computing", *PeerJ Computer Science* **3** (2017) e127.

[140] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2022.

[141] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial", *Frontiers in neurorobotics* **7** (2013) 21.

[142] XGBoost Documentation, "Xgboost parameters." https://xgboost.readthedocs.io/en/stable/parameter.html.

[143] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework", in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631. 2019.

[144] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization", *Advances in neural information processing systems* **24** (2011) .

[145] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions", *Advances in neural information processing systems* **30** (2017) .

[146] S. Lipovetsky and M. Conklin, "Analysis of regression in game theory approach", *Applied stochastic models in business and industry* **17** (2001) 319–330.

[147] **ALICE** Collaboration, S. Acharya *et al.*, "Charm production and fragmentation fractions at midrapidity in pp collisions at $\sqrt{s} = 13$ TeV", *JHEP* **12** (2023) 086, arXiv:2308.04877 [hep-ex].

[148] **ALICE** Collaboration, S. Acharya *et al.*, "Measurement of prompt D_s^+-meson production and azimuthal anisotropy in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", *Phys. Lett. B* **827** (2022) 136986, arXiv:2110.10006 [nucl-ex].

[149] F. Grosa, S. Politanò, and A. Bigot, "flarefly", Jan., 2023. https://doi.org/10.5281/zenodo.7579657.