Soft QCD parameter tuning using feed-forward neural networks

Michael Dessena

19th April 2022

Abstract

——INSERIRE ABSTRACT——

Contents

Chapter 1

Introduction

The **standard model of particle** is the theory that describe all the elementary particles, the electroweak [1] and the strong interaction

Chapter 2

Hadron-Hadron Scattering

In a high energy proton-proton collision we can have either soft or hard processes. Most of the time the hard processes are accompanied by soft interaction, occurring along the hadron interaction. While, the hard processes as the Higgs boson production, the high p_T jet production are well understood using perturbation theory, the soft processes as the underlying event, the hadronization are not so well understood. In fact, these processes are described by non perturbative QCD.

This chapter gives a theoretical introduction on the QCD factorization theorem, the fixed order calculation for the perturbative approach and on the all orders approaches (e.g. the parton shower).

2.1 QCD factorization theorem

The fundamental idea for the description of a hadron-hadron collision is given by the fact that hadrons are made-up from partons that, at some high energy scale, can be seen as free. So the hadron-hadron interaction can be seen as an interaction between two free partons, given a sufficient high energy scale for the interaction. This idea was developed in the framework of the deep inelastic scattering, and the Bjorken scaling observation confirms it [2]. The generalization of this concept leads to the QCD factorization theorem.

The factorization theorem was introduced first by Drell and Yan [3]. The hadron-hadron scattering is described in terms of partons extending the formalism used for deep inelastic scattering

$$\sigma_{AB} = \int dx_a \, dx_b \, f_{a/A}(x_a) \, f_{b/B}(x_b) \, \hat{\sigma}_{ab \to X} \quad . \tag{2.1}$$

Where X is a partonic/leptonic state and $a\,(b)$ a quark or an antiquark in the hadron $A\,(B)$. This is valid in the "scaling" limit:

$$s \longrightarrow \infty \, , \qquad \frac{M_X}{\sqrt{s}} = \text{finite} \quad ; \tag{2.2}$$

where the center-of-mass energy grows to infinity but with a fixed ratio of the X state invariant mass and the center-of-mass energy.

The problem arises from the perturbative corrections from real and virtual gluon emission, in particular from the colinear gluon emissions. These contributions take

to a logarithmic divergence (spoil the convergence of the perturbative expansion). These dependencies can be absorbed by the parton distribution functions (DGLAP equations). This results in the violation of the scaling

$$f_{a/A}(x_a) \longrightarrow f_{a/A}(x_a, Q^2) \quad , \tag{2.3}$$

now, the parton distribution function depend on the momentum scale Q^2 of the interaction.

So, we can rewrite the factorization theorem in Eq. 2.1 as:

$$\sigma_{AB} = \int dx_a \, dx_b \, f_{a/A}(x_a, Q^2) \, f_{b/B}(x_b, Q^2) \, \hat{\sigma}_{ab \to X} \quad , \tag{2.4}$$

at this point the *finite* corrections to the leading-logarithmic cross section in the perturbative expansion are specific for each process (not universal, they are process dependent). This leads in the Eq. 2.4 to the α_s series:

$$\sigma_{AB} = \int dx_a \, dx_b \, f_{a/A}(x_a, \mu_F^2) \, f_{b/B}(x_b, \mu_F^2) \, \left[\hat{\sigma}_0 + \alpha_s(\mu_R^2)\hat{\sigma}_1 + \ldots \right] \quad . \tag{2.5}$$

In Eq. 2.5 two scales enter the formula:

- The *factorization scale* μ_F: this scale separates long- and short- distance physics, this scale is related to the resolution with which the hadron is being probed.

- The *renormalization scale* μ_R: the scale at which is evaluated the strong coupling constant α_s. The dependence of α_s on the renormalization scale is related to different effects such as the vacuum polarization, the quark self-energy, the vertex corrections, and the gluon loop corrections to the elementary three-gluon and four-gluon couplings.

The higher-order corrections to the cross section in Eq. 2.5 are important because they lead to an improvement in the cross section prediction, gradually reducing the dependencies on μ_R and μ_F. In the absence of an all order prediction, a choice for the two scales have to be taken. Typically, the scales are assumed to be equal: in the Drell-Yan process the standard choice is $\mu_F = \mu_R = M$, with M the lepton pair mass [4]; other cases are the invariant masses of Z-boson and top quark or the jet transverse energy to study [4] the production cross sections for Z-bosons, top quarks and large E_T jets.

The parton distribution functions used in the hard scattering are solution to the DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) equation [5, 6, 7, 8]

$$\mu_F^2 \frac{\partial f_{i/H}(x, \mu_F^2)}{\partial \mu_F^2} = \sum_j \frac{\alpha_s(\mu_F^2)}{2\pi} \int_x^1 \frac{dz}{z} \, P_{i \to j}(z) \, f_{j/p}\left(\frac{x}{z}, \mu_F^2\right) \quad . \tag{2.6}$$

Where $P_{i \to j}$ are the splitting functions: they are the probability to have a parton of type i that becomes, by the emission of a quark or a gluon, a parton j, carrying fraction z of the momentum of parton i.

The splitting functions have perturbative expansions:

$$P_{i \to j}(x, \alpha_s) = P_{i \to j}^{(0)}(x) + \frac{\alpha_s}{2\pi} P_{i \to j}^{(1)}(x) + \ldots \quad . \tag{2.7}$$
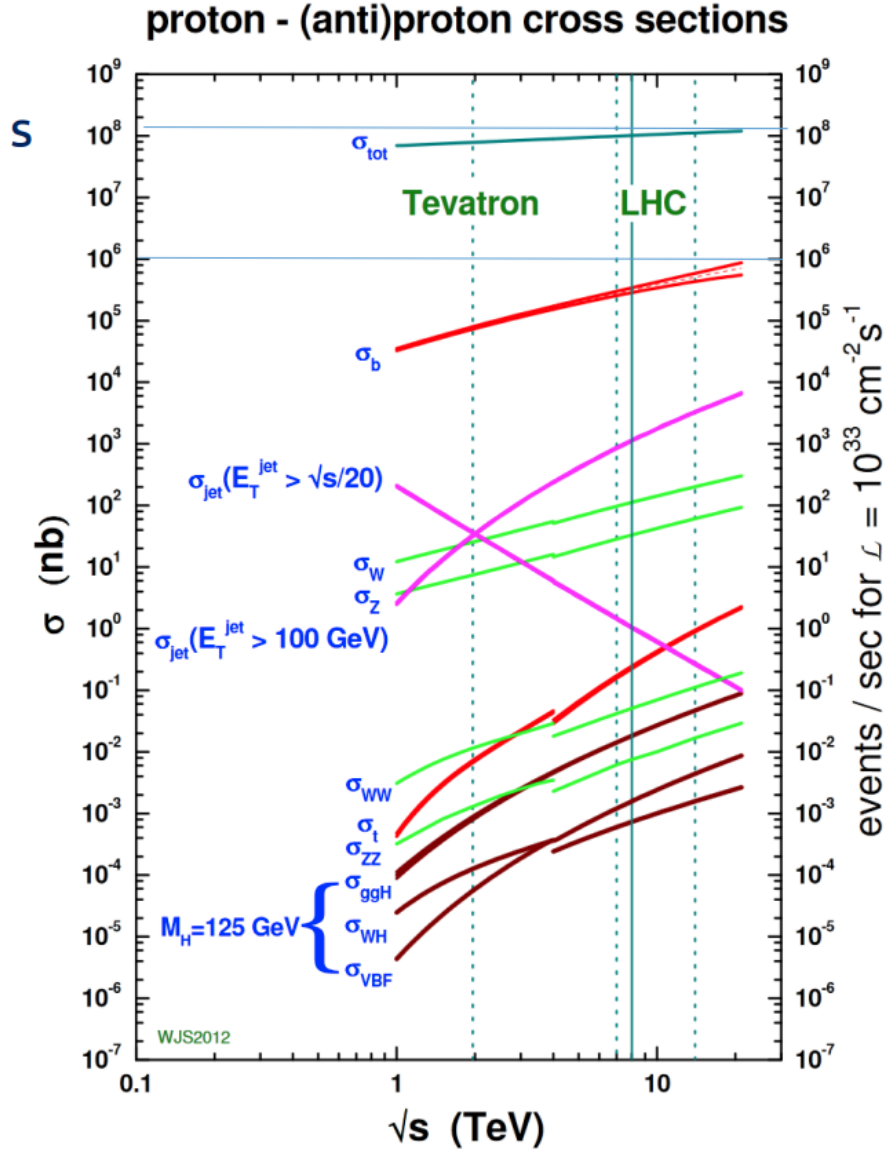
Figure 2.1: Next-to-leading order cross sections at Tevatron and LHC colliders energies (the splitting are at the transition between $p\bar{p}$ and pp cross section). Figure from [9]

This procedure has been used to calculate Standard Model cross section in $p\bar{p}$ and pp scattering respectively at Tevatron and LHC energies as shown in Fig. 2.1.

The parton distribution functions dependence on Q^2 can be derived theoretically via the DGLAP equations. While, the x dependence is given fitting the deep-inelastic and other hard-scattering processes experimental data. The experimental coverage in (x, Q^2)-plane is shown in Fig. 2.2 where is also underlined the relationship between (x, Q^2) and kinematic variables in Drell-Yan processes for a final state with invariant mass M and rapidity y is shown, further details in section 2 of [4]. Assuming that the factorization scale Q is equal to M (The reference center of mass energy is 13 TeV), for two incoming particles with four-momentum respectively p_1

and p_2 the relations with y and M are:

$$p_1^\mu = \frac{\sqrt{s}}{2}(x_1, 0, 0, x_1)$$
$$p_2^\mu = \frac{\sqrt{s}}{2}(x_2, 0, 0, -x_2)$$
$$\implies \quad x_1 = \frac{M}{\sqrt{s}}e^y \quad x_2 = \frac{M}{\sqrt{s}}e^{-y} \quad , \quad \text{(2.8)}$$

where $s = (p_1^\mu + p_2^\mu)^2$. For example the figure shows that the production of a final state with invariant mass $M = 100$ GeV and rapidity $y = 2$ is given by the interaction of two hadrons with $x_1 \approx 0.05$ and $x_2 \approx 0.001$ with $Q^2 = 10^4$ GeV^2
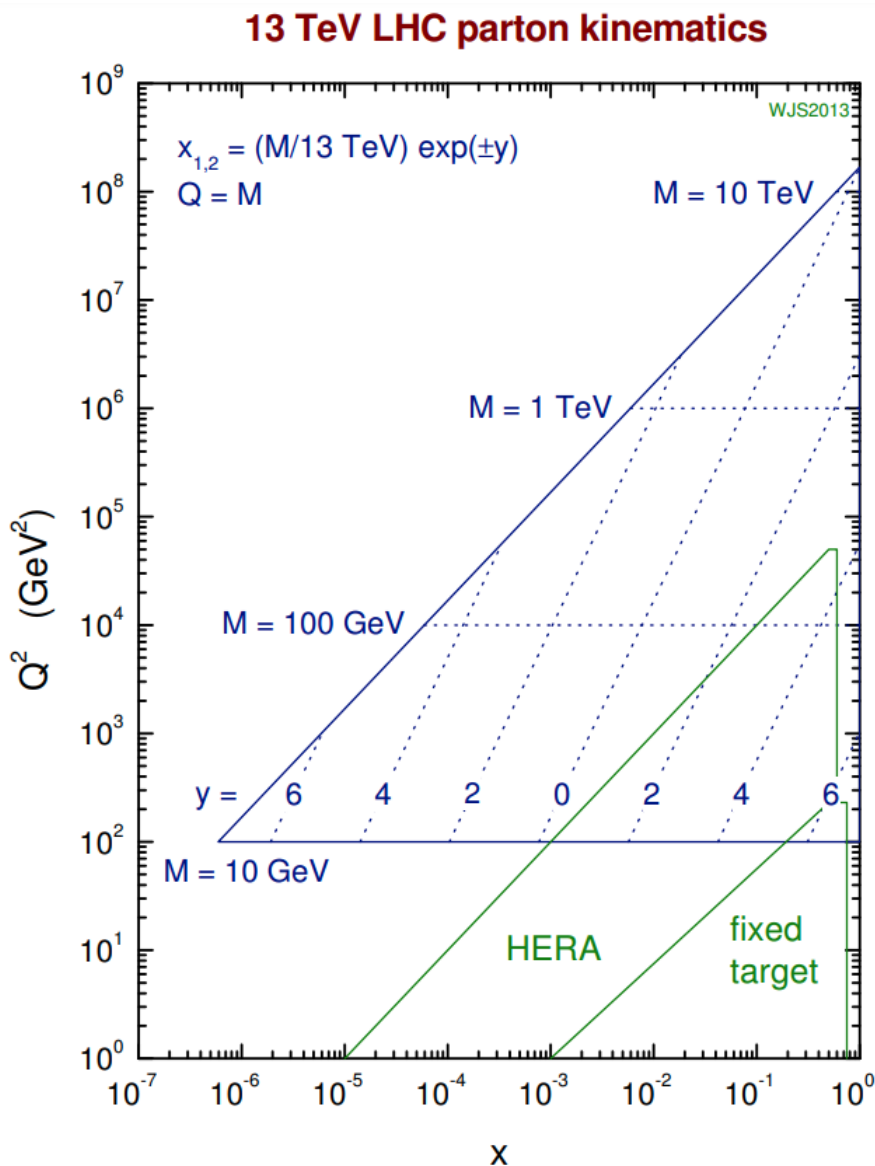

Figure 2.2: Graphical representation of the parton (x, Q^2) variables coverage of different experiments. For LHC these x and Q^2 are related to the kinematic variables y and M. Figure from [9]

2.2 Partonic cross section

Partonic cross section is one of the fundamental ingredients in our recipe for the description of hadron-hadron interactions. This can be calculated in a perturbative series in α_s from QCD first principles using quantum field theory.

The calculation at the leading order (LO) is performed evaluating all the possible tree-level Feynman diagrams for every process. Than the calculation proceeds computing the squared matrix element and integrating over the available phase space (analytically or numerically).

At this point, we can already encounter some divergence that have to be avoided imposing restrictions on the phase space.

2.2.1 Higher order calculations

The LO calculation can describe broad feature of a particular process and provide a first estimation of its cross section; anyway, in many cases this is insufficient.

The main source of uncertainty derives from the LO dependence on the unphysical renormalization and factorization scales. Some process may contribute only when going beyond the first approximation, and some divergence can be resummed.

To the next-to-leading order (NLO) calculation requires all the Feynman diagrams that take an extra α_s. This contribution can arise in two different ways:

- Virtual: internal lines (loops);

- Real: external lines (real particles).

Virtual corrections contains infrared divergences, arising from the integral on the loop circulating momentum, that cancel against infrared singularities given by collinear emissions or soft emissions [10, 11, 12].

A common strategy for the renormalization is dimensional regularization: it consists into performing the calculation in a $D = 4 - 2\epsilon$-dimensional space ($\epsilon < 0$), in that way the singularities appear as single and double poles in ϵ. Than, the limit $\epsilon \to 0$ is taken after the divergences have cancelled.

This NLO calculation with regularization allows to extend the treatment to zero transverse momentum.

The importance of higher order calculations can be seen with the following example. In a Z boson production:

1) **LO**: the Z is produced without transverse momentum (p_T), anything can recoil against the Z for momentum conservation (Fig. 2.3a).

2) **NLO**: the Z acquire a finite p_T, in this case the Z boson p_T is balanced by a single parton/gluon (Fig. 2.3b).

3) **NNLO**: the Z p_T can be balanced by two jets (Fig. 2.3c).

Another important benefit of performing a NLO calculation is the the reduction of the dependence on the unphysical renormalization (μ_R) and factorization (μ_F) scales. It is proven that higher order calculations of observables calculated to order α_s^n are dependent on the unphysical scales only at order higher than α_s^{n+1} [4]. The
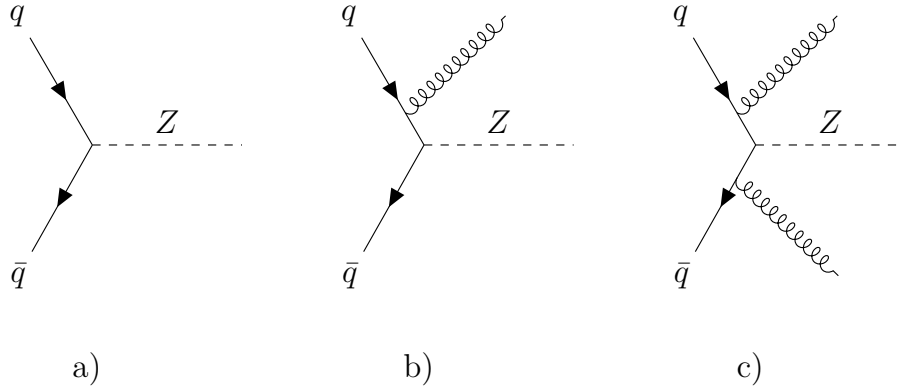
Figure 2.3: Feynman diagrams for the Z production by annihilation of a quark and an antiquark at LO (a), NLO (b), NNLO (c). At LO the Z can only be produced with a $p_T = 0$ for the conservation of the momentum.

range of predictions corresponding to different scale choices is usually attributed to *theoretical uncertainties*, this is shown in Fig. 2.4, where the uncertainties reduce from the LO calculation to the NLO and even more to the NNLO.
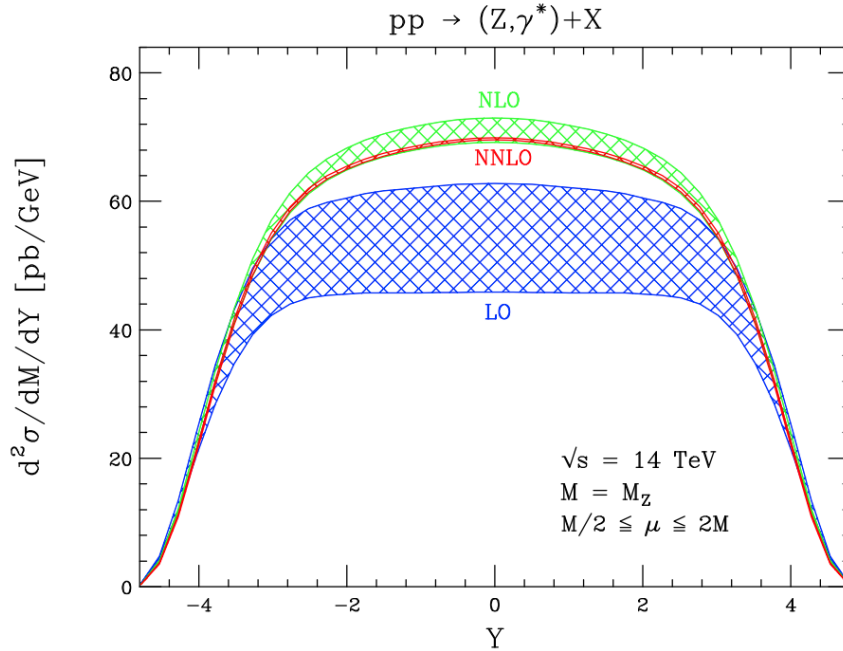
Figure 2.4: The rapidity distribution predictions at LO (blue) NLO (green) and NNLO (red) for the Z production at the center of mass energy $\sqrt{s} = 14$ TeV. The band width is related to the uncertainties. Going from LO to NLO there is a increase in the cross section prediction and a reduction on the scales uncertainties, the NNLO prediction is in the NLO error band width but there is a further increase in the precision of the prediction. Figure from [4], section 6.

2.3 Parton Showers

A different approach, instead than calculating order by order in the perturbative expansion, is the use of an *all-order* approach to describe the phenomena observed at high-energy colliders.

Different all-order approaches exist such as resummation and parton showers. Resummation is based on the observation that in many quantities the smallness of the expansion coefficients α_s is violated by large logarithmic enhancements. This takes the dominant contribution from each order and "resums" them by means of an evolution equation. The main problem in QCD is related to the fact that lot of quantities have correction of the form $\alpha_s^n \log^k(Q_i/Q_j)$ where Q_i and Q_j are two different energies scales, for example:

– Renormalization and factorization scales logs: $\alpha_s^n \log^n(Q^2/\mu_f)$

Various methods to perform this resummation exist.

An example is the Z production p_T spectrum shown in Fig. 2.5: here the comparison between experimental CDF data and theoretical predictions is shown: in the low p_T region the *all-order* approach regularizes the divergence of the fixed order calculation and describes the data better.
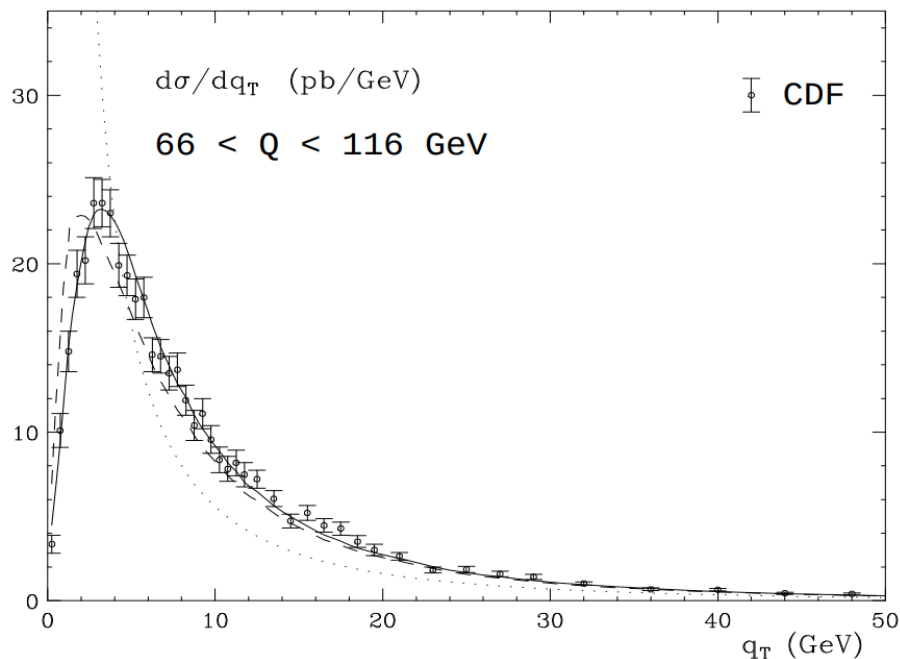
Figure 2.5: CDF data on Z production cross section at Tevatron collider, CDF experiment, the predictions from fixed order calculation (dotted) with resummation (dashed), and with the inclusion of power corrections (solid) are compared. Figure take from [13]

An other *all-order* approach is parton showers: it is implemented in different programs, as PYTHIA [14], HERWIG [15] and SHERPA [16]. This starts from few parton arising from hard interaction and than these are related to partons to a lower energy scale close to Λ_{QCD} using the DGLAP evolution equation formalism. The solution to this equation can be written using a Sudakov form factor arising from the probability of no gluon emission in the evolution from higher scale to lower

scale.

In the parton showering process, additionally to the kinematic variable (momentum fraction z and an azimuthal angle ϕ) and flavours of the partons, an evolution variable t is generated. PYTHIA8 use as evolution variable the squared of the relative transverse momentum of the two partons in the splitting (p_T^2). Different choices are made in HERWIG and SHERPA.

As mentioned before, the shower evolution is based on the standard (LO) DGLAP splitting kernels P(z) described here:

$$P_{q \to qg}(z) = C_F \frac{1+z^2}{1-z} \quad ; \tag{2.9}$$

$$P_{g \to gg}(z) = C_A \frac{(1-z(1-z))^2}{z(1-z)} \quad ; \tag{2.10}$$

$$P_{q \to q\bar{q}}(z) = T_R(z^2 + (1-z)^2) \quad ; \tag{2.11}$$

where $C_F = \frac{4}{3}$, $C_A = N_C = 3$ and $T_R = \frac{1}{2}$, each contribution is multiplied by N_f if summing over all contributing quark flavours.

Both Initial State Radiation (ISR) and Final State Radiation (FSR) algorithms are based on these splitting kernels. The respective probabilities of emitting radiation as one moves in the decreasing evolution variable sequence are:

$$FSR: \qquad \frac{d\mathcal{P}_{FSR}}{dp_T^2} = \frac{1}{p_T^2} \int \frac{dz}{z} \frac{\alpha_s}{2\pi} P(z) \quad ; \tag{2.12}$$

$$ISR: \qquad \frac{d\mathcal{P}_{ISR}}{dp_T^2} = \frac{1}{p_T^2} \int \frac{dz}{z} \frac{\alpha_s}{2\pi} P(z) \frac{f'(x/z, p_T^2)}{f(x, p_T^2)} \quad . \tag{2.13}$$

We can write-out our Sudakov form factor by using the two probability in Eq. 2.12 and Eq. 2.13, as

$$\Delta(p_T^2) = \exp\left(- \int_{p_{T0}}^{p_T'} \frac{d\mathcal{P}_{PS}}{dp_T^2} \, dp_T \right) \qquad \text{with} \quad PS = ISR, \ FSR \quad . \tag{2.14}$$

The Sudakov form factor give the probability of a parton to evolve from an harder scale to a softer scale without emitting a parton harder than some resolution scale. The introduction of the Sudakov form factor resums all the effects from the soft and collinear gluon emission. For more details and some plots of different Sudakov form factor values see section 3.5 of [4].

2.3.1 Merging parton showers and matrix element calculations

What we have now is: regions dominated by soft and collinear gluon emissions are described very well by parton showers approach; on the other hand, regions where partons are energetic and widely separated are well described by matrix element calculations. So, the best approach would be to combine the two different description, This would require an universal formalism for parton showers and matrix element calculations. This universal formalism was created in 2001 and it is call "Les Houches Accord" [17]. In order to combine the two approach some care must be taken: there

is the risk of double counting. There are different technique that prevent this risk for example CKKW [18] is used to combine LO matrix element calculations and parton shower.

A more best way is to combine NLO matrix element calculation with parton showers this is done by Frixione, Nason, Webber in the MC@NLO framework [19, 20, 21, 22]. In this scenario the risk of double counting is given by the fact that at NLO one emission is made real, than the progress of the parton shower give a double counting between real and virtual emission as shown in Fig. 2.6.
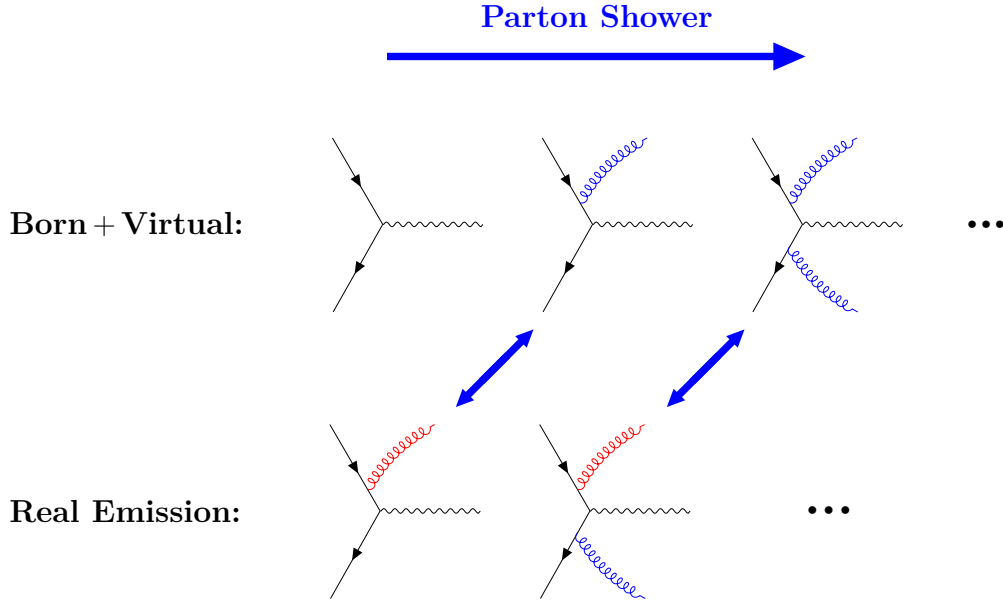

Figure 2.6: The FxFx margin scheme have to avoid this double counting. Feynman diagrams that can lead to a double counting are grouped with the violet arrow, the blue emission are related to the parton shower while the red ones to the NLO process.

2.4 Parton distribution functions

The last ingredient in our recipe is the knowledge of the quark and gluon distributions inside the two hadrons that undergo to the scattering. We have already seen that these quantities are depending on the virtuality of the interaction.
The information on the quark distribution inside a hadron $f_{q/p}(x, Q^2)$ arises from lepton-hadron DIS experiments, from lepton-pair production in hadron-hadron collisions (Drell-Yan processes) and jet measurements to study gluon distribution $f_{g/p}(x, Q^2)$. All these quantities are the experimental input in order to evaluate the PDF inside the hadron while the Q-evolution can be described by DGLAP equation. The evolution of the PDF can be run either with a NLO or with a NNLO calculations.

The kinematic region covered by experiments is shown in Fig. 2.2. At very low x and Q^2 the DGLAP evolution is believed to be no longer applicable and a BFKL (Balitsky-Fadin-Kuraev-Lipatov) [23, 24] description must be used; anyway, this has not any experimental evidence so the DGLAP approach is used as default in all the PDF analysis.
A lot of processes are available for the PDFs evaluation and a lot of PDF set have

been generated, as an example Fig. 2.7 shows the NNPDF3.1 set [25] at NNLO for a virtuality $Q^2 = 10$ GeV^2 (left) and $Q^2 = 10^4$ GeV^2 (right). Note that the gluon
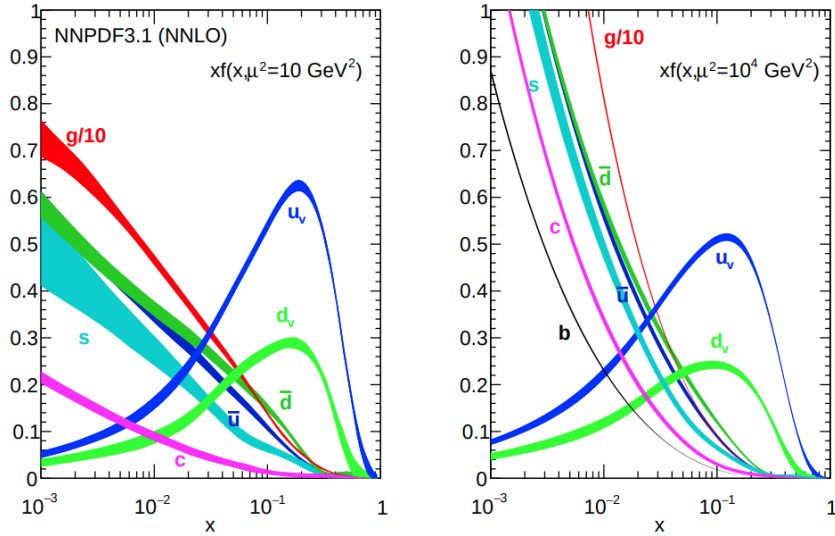

Figure 2.7: The NNPDF3.1 NNLO PDFs set, evaluated at $Q^2 = 10$ GeV^2 (left) and $Q^2 = 10^4$ GeV^2 (right). At low x the contribution from the gluons is the dominating one while at higher x the dominant contribution is from the valence quarks. The PDF Q evolution shows that when our proton is probed to higher Q^2 the resolution increase (higher Q correspond to smaller distance resolution) and so we have a bigger contribution from the sea quarks at low x values.

contribution have been scaled of a factor 10: in fact, in the low x region, $x < 0.01$, the gluon contribution is the dominating one, while at high x value the valence quarks dominate the PDF.

In Fig. 2.7 we can also see that with increasing virtuality (Q^2) at low x the density of the sea quarks increases: this is related to the fact that our hadrons are probed at higher energy and the probe resolution is proportional to the energy.

$$\text{Resolution} \sim \frac{\hbar}{Q} \quad . \tag{2.15}$$

So, when probed at higher energy, the hadrons appear denser than when are probed at lower energy. This, as will be discussed in the next chapter, is related to the higher number of interaction between partons in a single hadron-hadron collision. The phenomenon of having more than one interaction between partons in a single hadron-hadron collision is called *multiple parton interaction*: this concept will be discussed in the next chapter.

2.5 A real proton-proton collision

We have understood that the complexity in the description of a proton-proton collision arises from composite nature of the protons. In this chapter we discussed the importance of the QCD factorization theorem that help us in the calculation of the hadronic cross section with the convolution between the partonic cross section and

the PDF. We have discussed the importance of the parton shower algorithm where a set of partons are evolved in a more complex final state by emissions in the initial and final states.

All these processes are important in the description of a real proton-proton collision but also the partons that are left unscattered are non-color singlet and can contribute to the complex final state observed in the experiments, and additionally, as mentioned before, nothing prevent additional partons scatterings from taking place and growing more and more the complexity of the partons final state.

Than, another problem is related to the not-well-understood hadronization process. Hadronization is not known from first principles and different models have been implemented in different programs: the *cluster fragmentation model* implemented in HERWIG and the *string fragmentation model* in PYTHIA simulated this process of hadronization where the set of final-state partons is transformed into a set of hadrons. All these processes are schematically shown in Fig. 2.8.
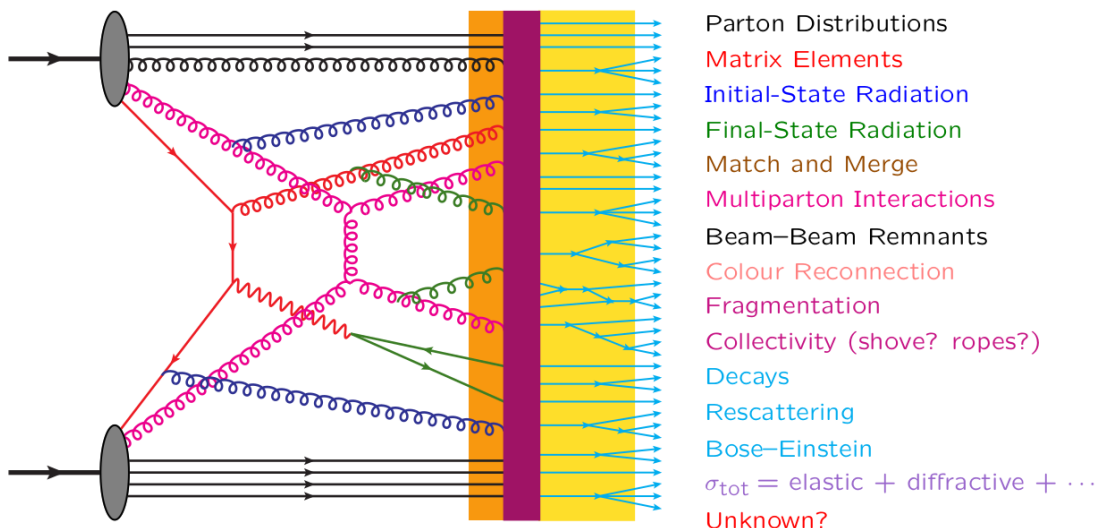

Figure 2.8: A schematic representation for a *pp* collision. Reading the image from left to right one can have an idea on the evolution of the system. The two incoming hadrons enter the scattering from the left side, the read line indicate the main hard scattering and the fuchsia one the second parton scattering (MPI) each interaction is associated with initial (blue) and final (green) state radiation, the unscattered partons (black lines) re-enter the color reconnection and hadronization processes. Than the new formed hadrons (lightblue) can undergo to different decays.

Next chapter is going to describe the PYTHIA Monte Carlo generator in more details. PYTHIA introduces different free parameters that need to be tuned with experimental data from Tevatron and LHC. The tune methods are described in Chapter 5 along with the description of some already existing tune for the underlying event in proton-proton collision.

Chapter 3

Multiple Parton Interactions

By the fact that the hadrons are viewed as "bunches" of partons, it is likely that in the same hadron-hadron collision different couples of partons can undergo to a scattering. This phenomenon is known as *Multiple Parton Interactions* (MPI) and it is related to the composite nature of the incoming hadrons.

So, it is clear that at some level these MPI have to exist, and they can become important in the description of the event; they can change the color topology of the colliding system as a whole.

In this scenario it is important to have a good understanding of the phenomenon. The aim of this section is to describe the basic concepts that are use to simulate MPI, for example in `Pythia` Monte Carlo event generator; than some focus is given to the free parameters we are going to tune in the following sections.

3.1 Basic Concepts

The main hypothesis of the multiple interactions models is that: the QCD factorization theorem is true not only for the hard process but also for the other partons scatters.

So, we can write:

$$\frac{d\sigma_{\text{int}}}{dp_\perp} = \sum_{i,j,k,l} \int dx_1 \int dx_2 \int d\hat{t} \, f_i(x_1, Q^2) f_j(x_2, Q^2) \frac{d\hat{\sigma}_{ij \to kl}}{d\hat{t}} \delta \left(p_\perp^2 - \frac{\hat{t}\hat{u}}{\hat{s}} \right) \quad . \quad (3.1)$$

This represent the interaction differential cross section for the hadron-hadron collisions, where $\frac{d\hat{\sigma}_{ij \to kl}}{d\hat{t}}$ is the differential cross section for QCD hard $2 \to 2$ processes, this processes are the one reported in Table 3.1.

In Eq. 3.1 we used the Mandelstam variables associated to the partonic system:

$$\hat{s} = (p_1 + p_2)^2 = (p_3 + p_4)^2 = x_1 x_2 s \tag{3.2}$$
$$\hat{t} = (p_1 - p_3)^2 = (p_2 - p_4)^2 \tag{3.3}$$
$$\hat{u} = (p_1 - p_4)^2 = (p_2 - p_3)^2 \tag{3.4}$$

where p_1, p_2 are the four-momenta of the incoming partons and p_3, p_4 the four-momenta of the outgoing partons.

Note that in Eq. 3.1 the jet cross section is twice as large $\sigma_{\text{jet}} = 2\sigma_{\text{int}}$, because at first approximation each interaction gives rise to two jets.

We assume also that the "hardness" of processes is defined by the p_T scale of the interaction ($Q^2 = p_T^2$).

As you can see from the formulae in Table 3.1 at small scattering angles, for $t \to 0$, the t-channel gluon exchange processes $qq' \to qq'$, $qg \to qg$ and $gg \to gg$ dominate the full matrix element. For scatterings that are soft relative to $\hat{s}$, $|\hat{t}| \ll \hat{s}$, we can approximate $|\hat{t}|$ as:

$$p_T^2 = \frac{\hat{t}\hat{u}}{\hat{s}} = \frac{\hat{t}(-\hat{s}-\hat{t})}{\hat{s}} \approx |\hat{t}| \quad , \tag{3.5}$$

in this limit, the only differences between quark and gluon cross sections are the color factors

$$\hat{\sigma}_{gg} : \hat{\sigma}_{qg} : \hat{\sigma}_{qq} = \frac{9}{4} : 1 : \frac{4}{9} \quad . \tag{3.6}$$

So, the Eq. 3.1 can be rewritten like

$$\frac{d\sigma_{int}}{dp_T^2} \approx \int \int \frac{dx_1}{x_1} \frac{dx_2}{x_2} F(x_1, p_T^2) F(x_2, p_T^2) \frac{d\hat{\sigma}_{2\to2}}{dp_T^2} \quad , \tag{3.7}$$

whit:

$$\frac{d\hat{\sigma}_{2\to2}}{dp_T^2} = \frac{8\pi\alpha_s^2(p_T^2)}{9p_T^4} \quad ; \tag{3.8}$$

$$F(x, Q^2) = \sum_q \left(x\, q(x, Q^2) + x\, \bar{q}(x, Q^2) \right) + \frac{9}{4} x\, g(x, Q^2) \quad . \tag{3.9}$$

Now, we can integrate the Eq. 3.7:

$$\sigma_{int}(p_{T\,min}) = \int_{p_{T\,min}^2}^{(\sqrt{\hat{s}}/2)^2} \frac{d\hat{\sigma}_{2\to2}}{dp_T^2}\, dp_T^2 \propto \frac{1}{p_{T\,min}^2} \xrightarrow{p_{T\,min}\to0} \infty \tag{3.10}$$

The total cross section is divergent in the limit $p_T \to 0$, this divergence is shown in Fig. 3.1. Due to this divergence the total interaction cross section at some p_T scale can exceed the total proton-proton cross section.

To understand this paradox should be noted that the interaction cross section described in Eq. 3.10 is related to the interaction probability between two partons and counts the number of interactions, while the total proton-proton cross section σ_{pp} counts the number of events. For example, an event (a proton-proton collision) in which two partons interact counts once in the total cross section and twice in the interaction cross section.

So, the ratio between this two quantities is perfectly allowed to be larger than unity, we can write it down as:

$$\frac{\sigma_{int}(p_{T\,min})}{\sigma_{tot}} = \langle n \rangle (p_{T\,min}) \quad . \tag{3.11}$$

Furthermore, we have to consider the *screening effect*: in fact the incoming hadrons are color singlet objects. Therefore, when the p_T of an exchanged gluon is small, and so the associated wavelength large, this gluon cannot longer resolve the color charges and the effective coupling is decreased, this screening set a cutoff in the divergence. The screening effect is schematically shown in Fig. 3.2.
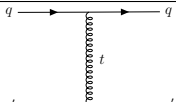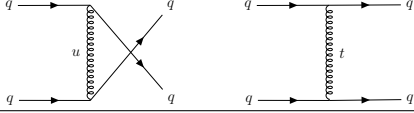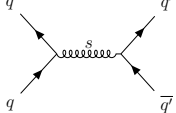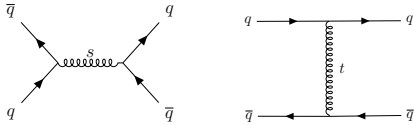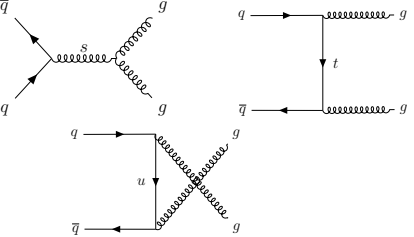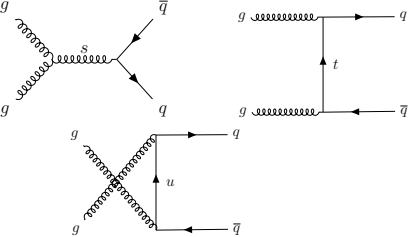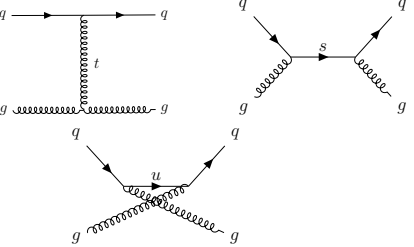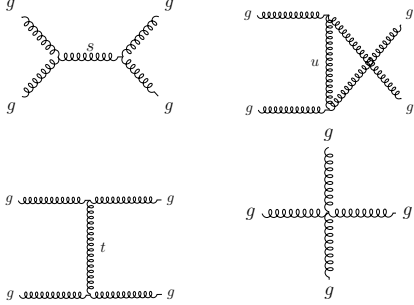
Process	Amplitude	$\sum |\mathcal{M}|^2/(4\pi\alpha_s)^2$
$qq' \to qq'$		$\dfrac{4}{9}\dfrac{s^2+u^2}{t^2}$
$qq \to qq$		$\dfrac{4}{9}\dfrac{s^2+u^2}{t^2} + \dfrac{4}{9}\dfrac{s^2+t^2}{u^2} - \dfrac{8}{27}\dfrac{s^2}{tu}$
$q\bar{q} \to q'\bar{q}'$		$\dfrac{4}{9}\dfrac{t^2+u^2}{s^2}$
$q\bar{q} \to q\bar{q}$		$\dfrac{4}{9}\dfrac{s^2+u^2}{t^2} + \dfrac{4}{9}\dfrac{t^2+u^2}{s^2} - \dfrac{8}{27}\dfrac{u^2}{st}$
$q\bar{q} \to gg$		$\dfrac{32}{27}\dfrac{t^2+u^2}{tu} - \dfrac{8}{3}\dfrac{t^2+u^2}{s^2}$
$gg \to q\bar{q}$		$\dfrac{1}{6}\dfrac{t^2+u^2}{tu} - \dfrac{3}{8}\dfrac{t^2+u^2}{s^2}$
$qg \to qg$		$-\dfrac{4}{9}\dfrac{s^2+u^2}{su} + \dfrac{s^2+u^2}{t^2}$
$gg \to gg$		$\dfrac{9}{2}\left(3 - \dfrac{tu}{s^2} - \dfrac{su}{t^2} - \dfrac{st}{u^2}\right)$

Table 3.1: Parton-Parton differential cross sections ($2 \to 2$ QCD process), can beh calculated in pQCD evaluating the matrix element for each process involving quark, antiquark and gluon. Table from [26]
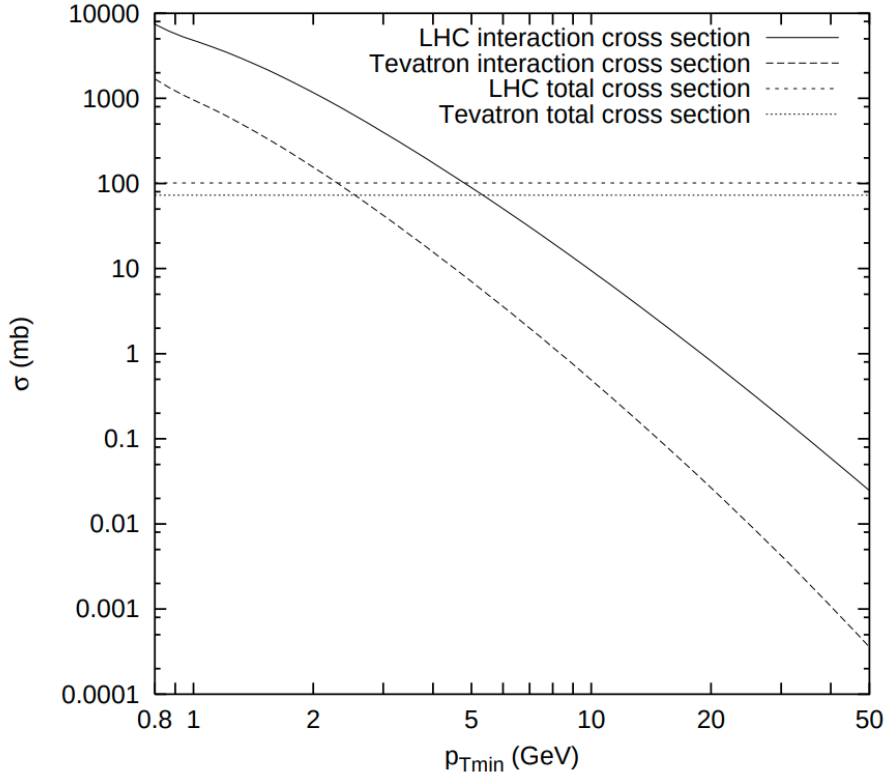
Figure 3.1: This figure shows the interaction cross section (σ_{int}) at Tevatron ($p\bar{p}$, $\sqrt{s} = 1.8$ TeV) and at LHC (pp, $\sqrt{s} = 14$ TeV). The flat lines are the corresponding values for the total cross section. The interaction cross section that arise from Eq. 3.10 is divergent for $p_{T\,min} \rightarrow 0$ in reality a dumping of this divergence is expected due to the color screening effect.

(a) Screened colour charges (b) Resolved colour charges
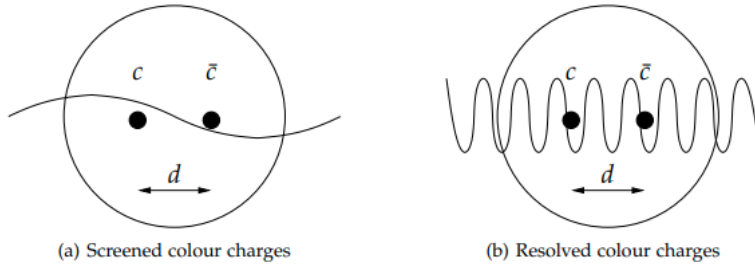
Figure 3.2: A picture of the screening effect. The left figure show two color charge that are not resolved by the gluon in fact the wave length is greater than the spatial separation, d, of the two charges. So resolution of the probe is not enough to discriminate the color charges. While on the right figure the two charges are very well distinct.

This cutoff is associated with color screening distance i.e. the average size of the region within the compensation of color charge occurs. This cutoff is then introduced in the factor as

$$\frac{\alpha_s^2 \left(p_{T0}^2 + p_T^2\right)}{\alpha_s^2 \left(p_T^2\right)} \frac{p_T^4}{\left(p_{T0}^2 + p_T^2\right)^2} \quad . \tag{3.12}$$

This factor contains the phenomenological regularization of the divergence, with the factor p_{T0} that have to be tuned to data.

Now the interaction cross section is smoothly regularized and therefore finite.

To be notice that: parameter p_{T0} do not have to be energy-independent, but since the energy is related with the sensibility of our probe, higher energy is related with the capacity of probing PDF to lower x values (as discussed previously, see Fig. 2.7) and in this low x region the number of partons rapidly increases. So, the partons are closer packed in this regions and as a consequence the color screening distance decrease.

The number of partons is related to x with a power low so, it is likely to have a dependence of the same form for p_{T0} respect to the center-of-mass energy

$$p_{T0}(\sqrt{s}) = p_{T0}^{ref} \left(\frac{\sqrt{s}}{E_{CM}^{ref}} \right)^{E_{CM}^{pow}} .$$ (3.13)

In the next section we are going to discuss how this and other effects fit into the Monte Carlo generator `pythia8`.

3.2 `Pythia8` Monte Carlo events generator

`Pythia8` is a standard tool for the simulation of events in high energy collisions. `Pythia` contains the evolution from a few-body system to a complex multiparticles final state.

3.2.1 Parton Shower

In `Pythia8` all the contributions from Initial State Radiation (ISR), Final State Radiation (FSR) and Multi Parton Interactions (MPI) are interleaved into a single common sequence of decreasing p_T.

The parton shower has been described in the previously section. The solution to the DGLAP equation is given putting Eq. 2.12 and Eq. 2.13 in Eq. 2.14 for the ISR and the FSR by a Sudakov form factor that is related to the no emission probability in the p_T-evolution.

The evolution variables for ISR and FSR are defined starting from the virtuality (Q^2) of the emission:

$$p_T^2 = \begin{cases} (1-z)Q^2 & ISR \\ z(1-z)Q^2 & FSR \end{cases},$$ (3.14)

so, in the two cases: for the FSR $Q_{FSR}^2 = (p^2 - m_0^2)$ in fact a time-like virtuality ($p^2 > 0$) is implicated, while for the ISR we have $Q_{ISR}^2 = (-p^2 + m_0^2)$ with a space-like virtuality ($p^2 < 0$). In both case Q^2 values are positive defined. Than the actual strong of the radiation is set by the choices of $\alpha_{s\,ISR}$ and $\alpha_{s\,FSR}$ values.

3.2.2 Multiple Parton Interactions in `Pythia`

The MPI, as said before, are also generated in a decreasing p_T sequence. So the hardest MPI is generated first. Than we can write the probability for an interaction, i, to occur at a scale p_T using a Sudakov-type expression (as we have done for the

ISR and the FSR):

$$\frac{d\mathcal{P}_{MPI}}{dp_T} = \frac{1}{\sigma_{ND}} \frac{d\sigma_{2\to2}}{dp_T} \ \exp\left(-\int_{p_T}^{p_T\,i-1} \frac{1}{\sigma_{ND}} \frac{d\sigma_{2\to2}}{dp_T'} \, dp_T'\right) \quad . \tag{3.15}$$

3.2.3 Momentum and flavour conservation

One problem is to achieve momentum conservation, so we need to take into account the modification in the PDF by the $i-1$ interaction. To do that in `Pythia` the PDF are rescaled to the remaining available x range, adjusting their normalization.
We need to take into account the momentum fraction x_i removed from the hadron remnant by the $i-th$ interaction. This is done evaluating PDF not at x_i but at a rescaled value

$$x_i' = \frac{x_i}{X} \qquad \text{with} \quad X = 1 - \sum_{j=1}^{i-1} x_j \quad . \tag{3.16}$$

So, using these quantities, we can rewrite our PDFs as:

$$f_i(x,Q^2) \ \longrightarrow \ \frac{1}{X} f_0\left(\frac{x}{X}\right) \quad , \tag{3.17}$$

where f_0 is the original one-parton inclusive PDFs.
Now, requiring also the flavour conservation and taking into account the number of valence and/or sea quarks involved in the preceding MPI. We have now the full forms of the PDFs used for the $i-th$ MPI:

$$f_i(x,Q^2) = \frac{N_{fv}}{N_{fv0}} \frac{1}{X} f_{v0}\left(\frac{x}{X},Q^2\right) + \frac{a}{X} f_{s0}\left(\frac{x}{X},Q^2\right) + \sum_j \frac{1}{X} f_{c_j0}\left(\frac{x}{X},Q^2\right) \quad , \tag{3.18}$$

$$g_i(x,Q^2) = \frac{a}{X} g_0\left(\frac{x}{X},Q^2\right) \quad , \tag{3.19}$$

where:

- $f_i(x,Q^2)$ $(g_i(x,Q^2))$ are the squeezed PDFs for quarks (gluons);

- N_{fv} the number of remaining valence quarks of the given flavour;

- N_{fv0} the number of original valence quarks of the given flavour (for the proton we have $N_u = 2$, $N_d = 1$);

- f_{s0} the sea-quark PDF;

- f_{cj} the companion PDF, this arise from the splitting $g \to q\bar{q}$ and a quark j is kicked out.

The factor a is defined to satisfy the total momentum sum rule.

3.2.4 Impact Parameter Dependence

The simplest hypothesis for the multiple interaction simulation, it is to assume the same initial state for all hadron collisions without dependencies on the impact parameter.

The more realistic scenario it is to include the possibility that each collision could be characterized by a different impact parameter b, where a small b value correspond to a large overlap between the two hadrons this is related to the probability of parton-parton interaction to take place.

To include the impact parameter dependence on the collision, it is necessary to make some assumption on the matter distribution inside the proton. A possibility is to assume a spherically symmetric distribution inside an hadron at rest $\rho(\mathbf{x}) \, d^3x = \rho(r) \, d^3x$. A Gaussian ansatz is the most simple choice but it appears to lead to a narrow multiplicity distribution and a too little pedestal effect. So the coiche is a double Gaussian:

$$\rho(r) \propto \frac{1-\beta}{a_1^3} \exp\left\{-\frac{r^2}{a_1^2}\right\} + \frac{\beta}{a_2^3} \exp\left\{-\frac{r^2}{a_2^2}\right\} \quad, \tag{3.20}$$

where a fraction β of matter is contained in a radius a_2, and the rest in a larger radius a_1.

Now, for a given matter distribution $\rho(r)$, the time-integrated overlap function of the incoming hadrons during the collision is given by:

$$\mathcal{O}(b) = \int dt \int d^3x \, \rho(x,y,z) \, \rho(x+b,y,z+t) \quad. \tag{3.21}$$

Assuming the matter distribution function in Eq. 3.20 and inserting it into Eq. 3.21 we get the following expression.

$$\mathcal{O}(b) \propto \frac{(1-\beta)^2}{2a_1^2} \exp\left\{-\frac{b^2}{2a_1^2}\right\} + \frac{2\beta(1-\beta)}{a_1^2+a_2^2} \exp\left\{-\frac{b^2}{a_1^2+a_2^2}\right\} + \frac{\beta^2}{2a_2^2} \exp\left\{-\frac{b^2}{2a_2^2}\right\} \tag{3.22}$$

this is useful to quantify the effect of overlapping protons. The larger is $\mathcal{O}(b)$ the more probable are parton-parton scatters between the incoming protons $\langle \widetilde{n} \rangle \propto \mathcal{O}(b)$.

So, these assumption change the so-far Poissonian nature of the framework[1]

$$\mathcal{P}(\widetilde{n}) = \langle \widetilde{n} \rangle^{\widetilde{n}} \; \frac{e^{-\langle \widetilde{n} \rangle}}{\widetilde{n}!} \tag{3.23}$$

Now the request that at least one parton interactions in the hadron-hadron collision, ensures that we get a finite total cross section. So the probability that an event is produced by two hadrons scattering with impact parameter b:

$$\mathcal{P}_{\text{int}} = \sum_{\widetilde{n}=1}^{\infty} \mathcal{P}_{\widetilde{n}(b)} = 1 - \mathcal{P}_0 = 1 - e^{-\langle \widetilde{n}(b) \rangle} = 1 - e^{-k\mathcal{O}(b)} \tag{3.24}$$

---

[1] the Poissonian distribution, in Eq. 3.23 describe the probability of having n interactions at each impact parameter. If the matter distribution have an infinite tail (like our in Eq. 3.20) events may be obtained with arbitrarily large b values. This can be a problem for the definition of the total hadron-hadron cross section

So we have now that the number of interaction per event is give by (for impact parameter b):

$$\langle n(b) \rangle = \frac{\langle \widetilde{n}(b) \rangle}{\mathcal{P}_{\text{int}}} \tag{3.25}$$

so, this have modified the probability distribution of interactions number from a Poissonian to a narrower one at each b fixed.

3.2.5 Parton rescattering

It is not necessary that the partons undergoing to a MPI are a different partons couple from the one scattered before. As shown in Fig. 3.3 MPI can also arise when a parton scatters more than once against partons from the other beam, this is call *parton rescattering.*
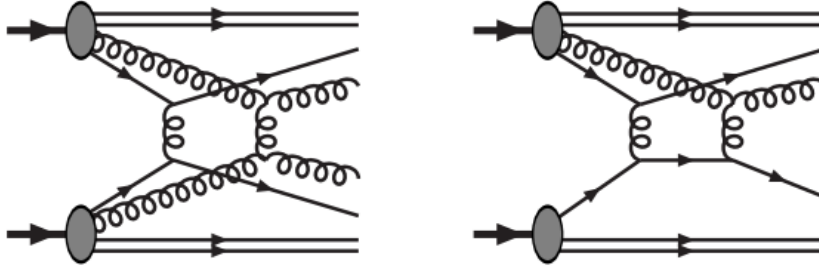

Figure 3.3: Add caption

We can have 3 type of MPI:

1. No ones of the partons that enter in the second scattering undergoes to another scatter before

2. Only one of the two partons have alredy been scattered

3. Both the partons have alredy been scattered before.

The second and the third are the rescatters the overall influence of rescatters in proton-proton interactions was estimated to be small respect to the first case with distinct $2 \rightarrow 2$ scatters.

The simulation of parton rescatters start from the evaluation of the parton density as:

$$f(x, Q^2) \quad \longrightarrow \quad \underbrace{f_{\text{rescaled}}(x, Q^2)}_{\text{hadron remnant}} + \underbrace{\sum_n \delta(x - x_n)}_{\text{scattered parton(s)}} \tag{3.26}$$

The δ take into account the scattered partons that have a fixed momentum fraction x_n. While the hadron remnant is still described by a continuous momentum density, rescaled to achieve momentum conservation:

$$\int_0^1 \left(f_{\text{rescaled}}(x, Q^2) + \sum_n \delta(x - x_n) \right) dx = 1 \tag{3.27}$$

3.2.6 Interleaving of Multiple Interaction and Parton Shower

As discussed above the MPI are simulated in `Pythia` following a decreasing p_T evolution. So we have that the total probability is given from the composition of the various contributions:

$$\frac{d\mathcal{P}}{dp_T} = \left(\frac{d\mathcal{P}_{MPI}}{dp_T} + \sum \frac{d\mathcal{P}_{ISR}}{dp_T} + \sum \frac{d\mathcal{P}_{FSR}}{dp_T} \right) \times$$
$$\times \exp \left\{ - \int_{p_T}^{p_T\, i-1} \left(\frac{d\mathcal{P}_{MPI}}{dp_T'} + \sum \frac{d\mathcal{P}_{ISR}}{dp_T'} + \sum \frac{d\mathcal{P}_{FSR}}{dp_T'} \right) dp_T' \right\} \quad (3.28)$$
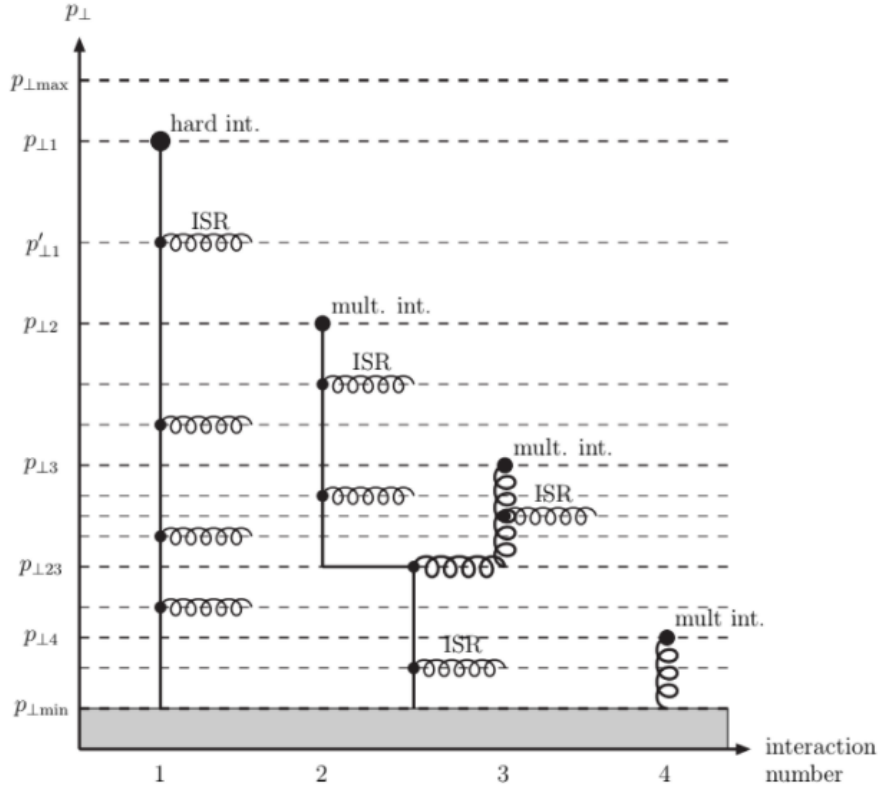
Figure 3.4: add caption

In Fig. 3.4 are shown 4 parton-parton interactions with their associated showers (ISR and FSR). The downward evolution correspond to read the graph from top to bottom. The first hard interaction occur at a scale $p_T = p_{T1}$ while the following ones at lower scales p_{T2}, p_{T3}, p_{T4}. Each interaction is associated with is radiation the first one occurs at $p_T = p'_{T1}$. The scatterings that occur at p_{T2} and p_{T3} are originating from the same mother parton.

This diagram is related to one of the two hadrons. The full event can be illustrated if a similar diagram is drawn for the other hadron and connected to the full black circles.

3.2.7 Beam Beam Remnants and primordial k_T

What is left after that the evolution is end? the evolution in p_T can create an arbitrary complicate final state.

This final state contains contributions from the scattered and unscattered partons that don't enter the p_T evolution. The last are the so called Beam Beam Remnants (BBR). BBR take into account the number of valence quark remaining and the number of sea quark required for the overall flavour conservation.

To ensure momentum conservation the BBR have to take all the remaining longitudinal momentum that is not extracted by the MPI initiators.

Primordial k_T

We have considered only the longitudinal momentum. in a real scenario partons inside the hadrons are fermions so are expected to have a non zero initial transverse momentum arising from the Fermi motion inside the incoming hadrons. This is denoted as "Primordial k_T" this is different from the transverse momentum derived from DGLAP shower evolution or from the hard interaction.

Based on Fermi motion alone, one would expect a value of the primordial k_T is estimate as:

$$k_T \simeq \frac{\hbar}{r_p} \approx \frac{0.2 \text{ GeV} \cdot \text{fm}}{0.7 \text{ fm}} \approx 0.3 \text{ GeV} \qquad (3.29)$$

but for example to reproduce the data for the the p_T distributions of Z bosons produced in hadron–hadron collisions, one need a larger contribution. This phenomenon has not a satisfactory explanation. Until an explanation is found the idea is to consider a effective primordial k_T for the initiators larger than the one in Eq. 3.29.

In `Pythia` the primordial kT is assigned to initiators sampling a Gaussian distributions for p_x and p_y independently with variable width σ

$$\sigma(Q, \widehat{m}) = \frac{Q_{1/2}\,\sigma_{\text{soft}} + Q\,\sigma_{\text{soft}}}{Q_{1/2} + Q} \frac{\widehat{m}}{\widehat{m}_{1/2} + \widehat{m}} \qquad (3.30)$$

Where Q is the hard-process renormalization scale for the main hard process and the p_T scale for subsequent MPI. σ_{soft}, σ_{hard} are the minimum and maximum width, $\widehat{m}$ is the invariant mass, while $Q_{1/2}$ and $\widehat{m}_{1/2}$ the halfway values between the two extremes.

3.2.8 Color Reconnection and Hadronization

The last important step at parton level is the color reconnection. Color reconnection is motivated by that MPI leads to different color strings. In the previous step the planar limit of the QCD was assumed where $N_c \to \infty$. Now, moving to real case where $N_c = 3$ all this strings that can be overlapped in physical space can be reconnected. The basic idea is to reconnect strings in order to reduce the total string length; and thus the potential energy.

In PYTHIA8 the reconnection is performed giving to each system a probability to reconnect given by:

$$\mathcal{P}_{\text{rec}} = \frac{p_{T\,\text{rec}}^2}{(p_{T\,\text{rec}}^2 + p_T^2)} \qquad p_{T\,\text{rec}} = R \times p_{T\,0} \qquad (3.31)$$

the `ColorReconnection:range` R is a user-tunable parameter while p_{T0} is the same parameter defined in MPI simulation.

Whit this definition for the probability it is clear that system with low p_T are more likely that can reconnect to other, this idea find is justification in the fact that at low p_T value correspond a larger spatial extension and so these strings have more chance to overlap with other and so to reconnect.

The hadronization takes all these partons (color strings) and transform it in a set of color-neutral hadrons. Hadronization is based solely on the *Lund string fragmentation model* [27, 28].

Lund model basic idea is to break the color line and to reduce the total string length the string is representative of the potential

$$V(r) = -\frac{a}{r} + \kappa r \qquad \text{with} \quad \kappa \approx 1 \text{ GeV/fm} \qquad (3.32)$$

where κ is the string tension. This potential is a combination of an attractive (Coulomb) potential ad a linear potential that phenomenologically include quarks confinement. The linear potential is the dominating part at increasing distance values, so the energy increase with distance.

The simplest case is the one in Fig. 3.5: The $q\bar{q}$ system evolves in space increasing the string length at some point the distance is to large and is convenient to break the string into two strings.
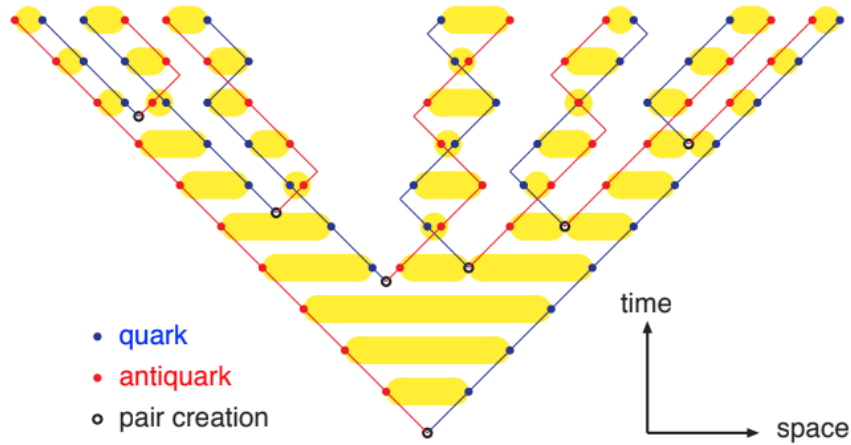
Figure 3.5: ADD

The hadronization step confines the quark into hadrons, than hadrons can undergoes to hadron rescattering and decay. These are the hadrons that are revealed by the detector.

3.3 PYTHIA summary

Tu summarize PYTHIA8 is able to simulate an high energy hadron-hadron collision. The evolution of the system is simulated in a common decreasing-p_T sequence the master formula for the evolution is written in Eq. 3.28.

This evolution start from an hard scale that is the scale of the main parton hard scattering that is described by a fixed ME calculation, PYTHIA can be interfaced with

external frameworks for the ME step as POWHEG and MAD-GRAPH5 AMC@NLO. Then the parton shower is started with the simulation of ISR, FSR and MPI also the parton rescattering is allowed. Once the evolution is ended the hadronization transform the partons in a set of final hadrons these hadron than decay and rescatters against each others before the detection.

All these processes are describe mainly by phenomenological model, due to the not-known-by-first-principles softQCD description. The use of these phenomenological introduces lot of free parameters (some are been pointed out in the previously sections) that have to be tuned with data to give to PYTHIA the ability of simulate real events.

In the next chapter we are going to focus on the study of the underlying event and so on the soft events related to an hard scattering.

Chapter 4

Observable to Study the Underlying Event

The UE are all the processes not associated with the primary hard scatter in an hadron-hadron collision.

All the process described in the previous section: ISR and FSR, MPI, and BBR and their interactions with color exchanges among them, contribute to the Underlying Event (UE) in the proton-proton collision.

The most of the observables to study the UE are sensible only to the sum of these contributions and not to the single ones so a good description of all this process and their interplay is really important.

In this chapter these observables are introduced and described with more details.

4.1 Minimum Bias Measurements and Underlying Event topology

A Minimum Bias (MB) measurement is a collection of inelastic events with a loose event selection. The events are collected requiring the minimal interaction with the detector (the smallest possible bias). The most of the interaction in MB observation are soft, $p_T \lesssim 2$ GeV. The study of the UE require at least one hard scattering ($p_T \gtrsim 2$ GeV) presence, in fact the UE is given by the underlying activity to an hard scatter.

To study the UE the topological structure of an hadron-hadron collision is used. The analysis is performed on an event-by-event basis. In the analysis the direction of the leading object is used to define regions in the $\eta - \phi$ space. Where η is the pseudorapidity defined as $\eta = -\log \tan \left(\frac{\theta}{2} \right)$, while ϕ is the azimuthal angle in the $x - y$ plane. The last one is defined from the direction of the leading object as $\Delta \phi = \phi - \phi_{\text{leading}}$.

The regions classification is shown in Fig. 4.1, we have:

- **Toward region**: the region that contains the leading object, this region contains the most of the particle produced by the hard interaction.

- **Away region**: this region contains the objects that recoil against the leading object, also this region contains mostly the particles produced by the hard
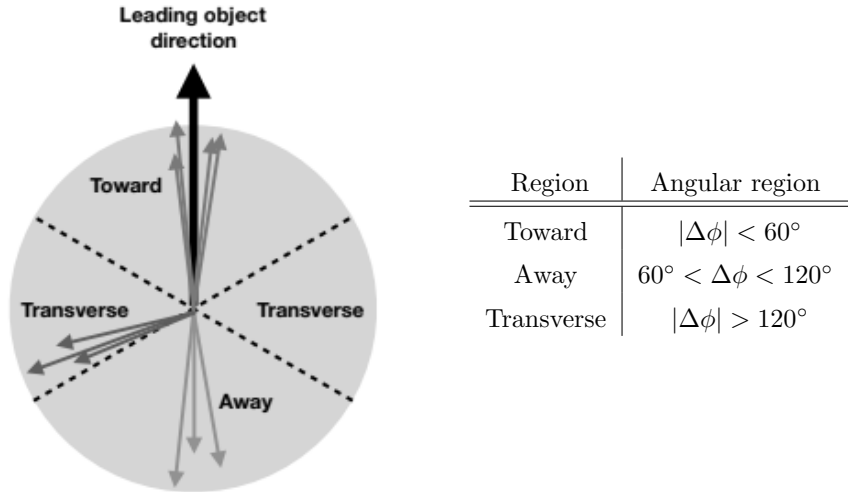
Region	Angular region
Toward	$|\Delta\phi| < 60°$
Away	$60° < \Delta\phi < 120°$
Transverse	$|\Delta\phi| > 120°$

Figure 4.1: This figure shows the four regions for the description of the UE on a event-by-event analysis. the angular values are shown in the table. The regions are defined starting from the leading object direction. The toward and away regions contain most of the contribution from the hard scattering (e.g. in a $t\bar{t}$ production event the two quark t are located in these regions); while, the transverse region are the ones in between the twp other regions, these are the most important for the study of the UE.

interaction.

- **Transverse regions**: the two transverse regions are the most sensitive to UE.

The transverse regions are also separated in:

  – **TransMAX**: This is the transverse region that contains the *maximum* number of charged particles, or scalar p_T sum of charged particles. This regions includes both MPI and hard-process contamination.

  – **TransMIN**: is the transverse region that contains the *minimum* number of charged particles, or scalar-p_T sum of charged particles. This region is the most sensitive to MPI effects.

The leading object definition depend on the type of event under observation. The charged-particle with largest p_T [29], the dilepton system in Drell-Yan observation [30, 31] or $t\bar{t}$ events [32] can all be used as leading object in the analyses for the UE event.
So we are interest in studying the transMAX and transMIN region in particular the observable sensitive to UE in these regions are: the charged-particle density and the charged-particle scalar-p_T sum density in the $\eta - \phi$ space

In an hadron-hadron collision with two jets productions it is observed that also the transverse regions activity increase with energy, as shown in Fig. 4.2, where the different color lines refer to different scales for the collision. Is observed that the away region ($60 \leq |\Delta\phi| \leq 120$) at increasing energy for the collision become broader this is related to the increasing quantity of FSR.
This increment in the transverse regions activity cannot be explained only by the increase of the FSR. To explain this rise one need to attribute activity in the transverse regions to MPI, extra scattering increasing with energy is related to the partons inside the hadrons that become denser packed when probed at higher energies.
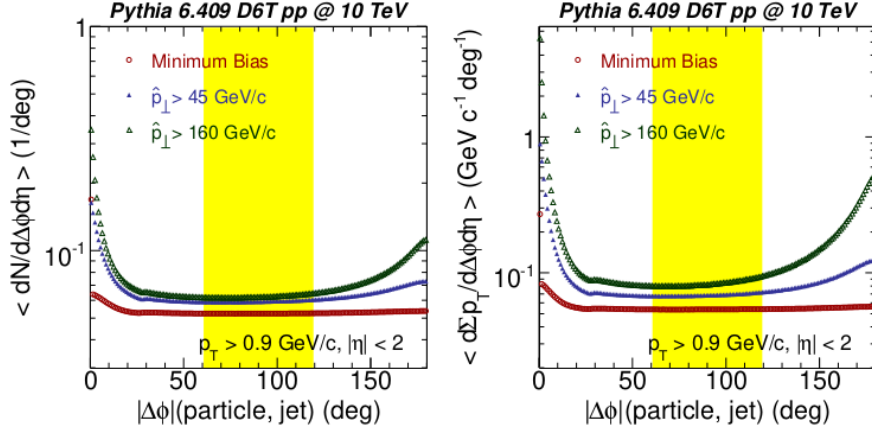
Figure 4.2: A comparison between three different scales for the interaction. The multiplicity of charged particles (left) and the scalar-p_T sum of charged particles (right) are simulated. The activity in the transverse regions increase due two effects: the FSR this is related to the broadening of the away region so some events from the shower end up in the transverse region (yellow band) but this alone can't explain the increment so is required the introduction of the MPI in the description. Figure from chapter 5 of [33].

The evolution of these two quantities in transMAX region in function of the transverse-momentum of the leading object (measurement of the energy scale for the collision) is shown in more detail in Fig. 4.3. The two distribution show a rapid raise at low leading-object p_T than a very low rise start.
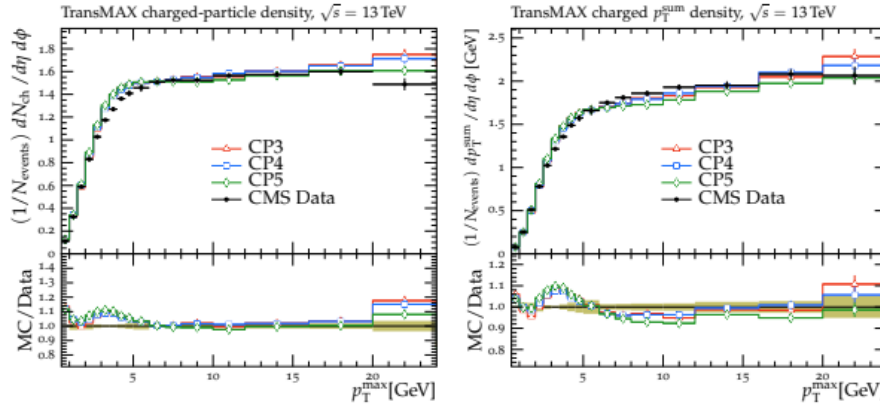

Figure 4.3: The evolution of the charged-particles density (left) and of the scalar-p_T sum density of charged particles (right) as a function of the energy scale for the scattering (leading object p_T), the two distribution increase quite rapidly in the first bins than they saturate ($\approx 6 - 7$ GeV) the scalar-p_T sum density increase a little bit more, for $p_T >$ but very slowly. The black dots are the experimental point and are compared to prediction with CMS PYTHIA tunes: CP3, CP4 and CP5; they are described with more details in the next chapter. Figure from [34].

Now we want to look for the sensitivity of these observables to some PYTHIA8 parameters. In figure Fig. 4.4 is shown the effect of the MPI on booth the distributions shown before. The amount of the MPI is related to the value of p_{T0}: an high value of p_{T0} is related to less MPI (blue line) and an higher value to a major contribution from the the MPI (red line).

Another important observation that was also pointed out in the Chapter 3.1 is
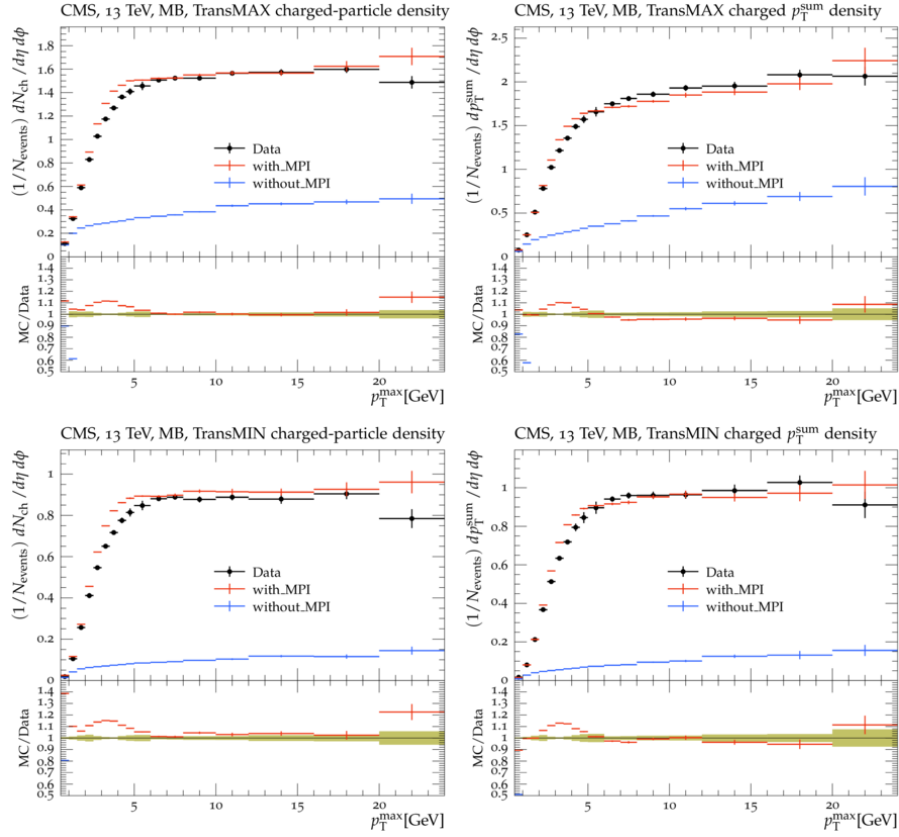
Figure 4.4: This image shows the effect of the MPI in the transMAX and TransMin regions. The contribution of the parton shower alone can't explain the contributions of the underlying event in these two regions (blue line) the introduction of the contribution from the MPI is necessary (red line). The two simulations are compared to the data from [29].

that the amount of activity in the two transverse regions is also dependent on the center-of-mass energy. This evolution is shown in Fig. 4.5 for two different energy. The amount of MPI increase with $\sqrt{s}$ that was expected from the evolution of the PDF with the energy of the collision: the hadrons become denser-packed when probed to higher energies.

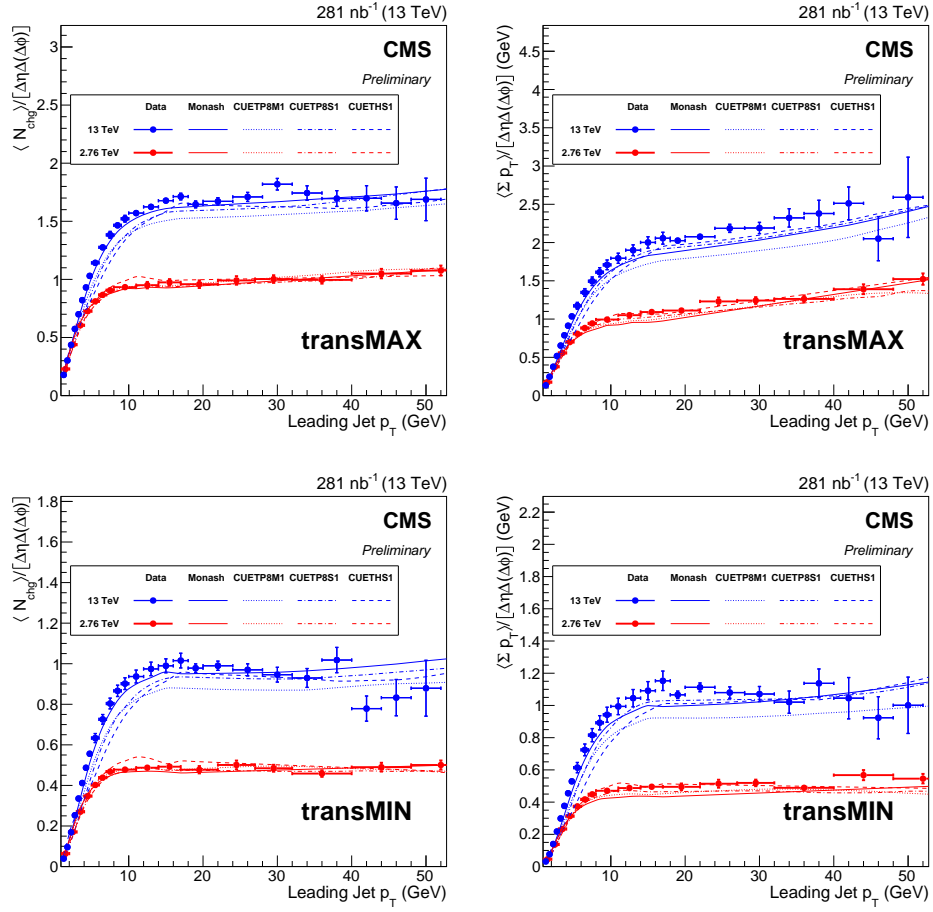**Figure 4.5:** The charged particle density in the transMAX (upper left) and transMIN
(lower left) regions and of the charged particle p_T-sum in the transMAX (upper right)
and the transMIN (lower right) regions evolution as function of the center-of-mass energy
is shown. The red ones are the data for $\sqrt{s} = 2.76$ and the blue ones for $\sqrt{s} = 13$ and
these are compare to different CMS tunes. Figure from [29]

Chapter 5

Tune procedure, CP5 Tune and MCNNTUNES

The study of the underlying event, or more in general of softQCD processes require the use of Monte Carlo generator based on phenomenological model all these models introduce lot of free parameters that must be tuned on experimental data in order to obtain meaningful results. The procedure of estimate the best parameters values is called *tune*.

The tune procedure can be really computational expensive, can require to run the generator a very large amount of times. And usually these MC generator are really expansive in computation time required for a single job.

To tune some parameters the number of jobs you have to run increase with the number of parameter you want to tune. different approaches have been used in the past to tune these MC generator:

1) **Manual tunes**: this approach is based on an optimization of the parameters made by eye. This is absolutely not the best way to tune some parameters, usually it requires a very large time for even semi-reasonable results since the process require a very large number of iterations.

2) **Brute force tunes**: A better way would be to perform a very dense sampling in parameters space and run a generator with every configuration. This is very computational expensive and a scan in a 5 parameters space with 10 division each require 100000 Monte Carlo runs this is really impractical.

3) **Parametrization-based tunes**: An even more better approach is to find a surrogate function to parameterize the response of the MC generator at different values of the parameters to tune and try to study (minimize) this surrogate function instead of the real response of the generator.

5.1 Parametrization-based approach

The last one is the most used the current state-of-art in the tune procedure is to use a polynomial function to fit the response of the generator and minimize it, this approach is implemented in the software PROFESSOR [35].

The first step in the procedure is to fit the response of the generator using a surrogate function simpler to study,

$$h(p) \xrightarrow{\text{parametrization}} \overline{h}(p) \tag{5.1}$$

than a *loss function* $\mathcal{L}(h(p), h_{\text{data}})$ is defined, between the surrogate function and the experimental data, a common choice is the χ^2 function defined as:

$$\mathcal{L}(h(p), h_{\text{data}}) \equiv \chi^2 = \frac{(h(p) - h_{\text{data}})^2}{\sigma} \tag{5.2}$$

Than to find the best parameters estimation this loss function need to be minimized, the set of parameters p_{best} that do this are the best evaluation that our generator can provide for the real values and are our *tune*.

$$p_{\text{best}} = \arg \min_p \mathcal{L}(h(p), h_{\text{data}}) \tag{5.3}$$

In our study instead of the common software PROFESSOR based on the polynomial parameterization we use the machine learning approach implemented in MCNNTUNES software with Feed Forward Neural Networks. MCNNTUNES is a software developed by S. Carazza, S.Aioli and M.Lazzarin [36] based on machine learning library TensorFlow [37]. MCNNTUNES is write in pyhton and it uses neural networks (NN) that are trained to learn the generator behavior to the parameter values.

Let's make a brief introduction on machine learning and in particular on neural networks

5.2 Machine Learning and Neural Networks

Machine learning (ML) is a particular type of Artificial Intelligence it consist in systems that learn automatically by the data that are feed to it and not by the explicit programming of the algorithm. Is clear that the machine learning require the training of the algorithm in order to have a prediction on the problem under analysis.

A particular type of ML is Deep Learning that uses neural networks with more than one layer organized in a hierarchical structure.

5.2.1 Neural Networks - Perceptron

The concept of the neural network was developed in 1958 by Frank Rosenblatt. He introduce the simpler example of NN: the perceptron [38]. A representation of a percepton is shown in Fig. 5.1 the input values are weighted and summed than the weighted-sum is passed to an activation function (step function) an additional offset b can be introduced, the output of the perceptron is than:

$$h(x) = \text{step}(w^T x + b) \tag{5.4}$$

The revolutionary feature of the perceptron was the ability of learning by an adjustment of the weights. But a single perceptron is not enough. An example of
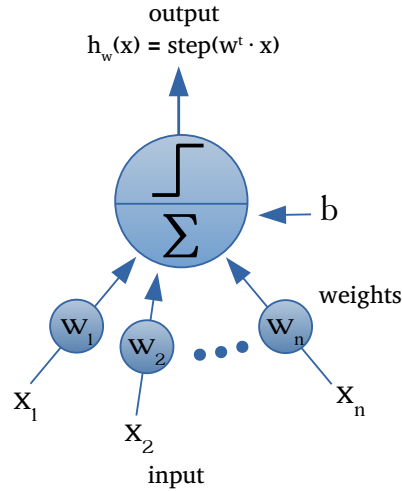
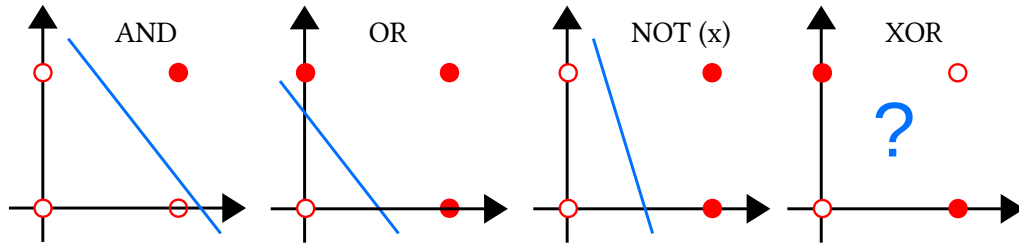**Figure 5.1:** A schematic representation of a perceptron.

Figure 5.2: The figure shows one of the limitations of the perceptron the

limitation in shown in Fig. 5.2 where the impossibility of implement a XOR operation using a perceptron is shown with a graphical explanation. The perceptron is a linear classification algorithm and in the image is draw as a line that set a boundary for the acceptation of the hypothesis.

The structure is very simple with a single unit but is not enough it have a lot of limitation so we have to introduce the concept of NN with more than one unit.

5.2.2 Feed-Forward Neural Networks

different units call "neurons" are linked together to compose an artificial NN. There are different type of NN we are interested in Feed-Forward NNs and fully-connected that are the ones used in MCNNTUNES.

Fig. 5.3 show a schematic view of a NN the basic idea is that the neurons can get some value in input and return a value as output, in fully-connected NN each neurons from a layer are connected to every neurons in the next layer. As we will discuss later one of the biggest feature of the NNs is that the they are universal function approximators [39, 40] the only request is that a sufficient number of hidden layer is available.

The Feed Forward attribute refers to the fact that the NN don't have internal recursion (loops) between neurons.

A common training algorithm for the NN is the back-propagation. A set of Monte Carlo simulation is used to train the NN. The procedure is based on the idea
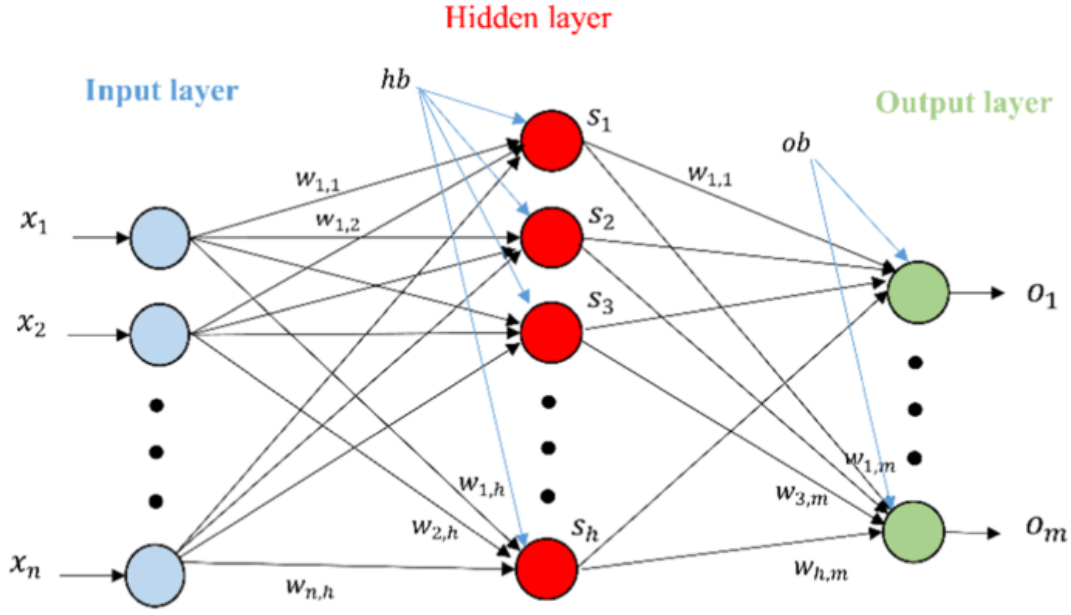
Figure 5.3: A fully-connected feed-forward neural network

of change the weights w_{jk} and parameters ϑ_j in order to minimize a loss function usually defined as the mean squared error (E):

$$E = \frac{1}{2} \sum (h_i(x^j, w_{jk}) - d_i)^2 \tag{5.5}$$

where the h_i are the value in output from the NN and the d_i the real value known from the Monte Carlo truth.

In the back-propagation algorithm the weight and the coefficients are update using the steepest-descent minimization:

$$w_{jk}^{(i+1)} = w_{jk}^{(i)} - \lambda \left(\frac{\partial E}{\partial w_{jk}} \right)^{(i)} \qquad \vartheta_j^{(i+1)} = \vartheta_j^{(i)} - \lambda \left(\frac{\partial E}{\partial \vartheta_j} \right)^{(i)} \tag{5.6}$$

where λ is the learning rate and is a user-tunable free parameter.

The batch is our training set that contains the Monte Carlo simulation that is used to calculate the gradient. The size of the batch used to train the NN is also a free parameters: smaller batch are faster to compute but the direction is not the real one but an approximation while, bigger batch size give a good approximation of the direction of the steepest-descent but can be computational expensive.

The number of evaluation of the entire batch is called epochs and is tunable by the user.

The batch size and the number of epochs are related. A batch of 50 run with a batch size of 10 and a number of epochs of 1000 require 5000 iterations while if we use a batch size of 25 it require only 2000 iterations to run over all the training set.

5.3 Previous Tune for the Underlying Event

In the paper [34] the CMS collaboration presents new PYTHIA8 tunes for the underlying event. The tunes are called CP and a number from 1 to 5 where CP stands
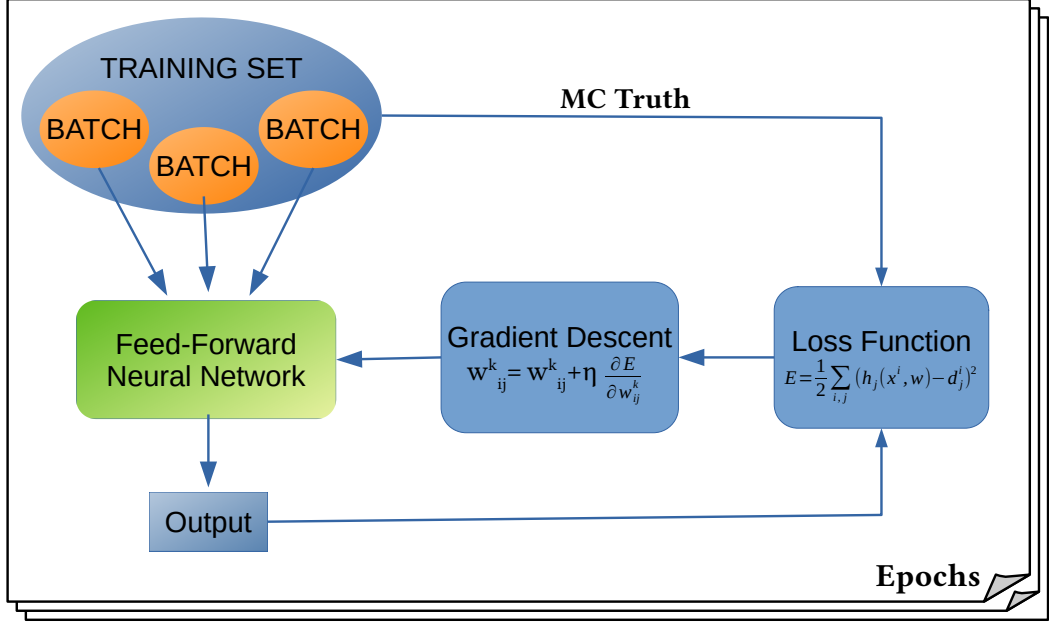
Parameter description	Name in PYTHIA8	Range considered
MPI threshold [GeV], `pT0Ref`, at $\sqrt{s} = \sqrt{s_0}$	`MultipartonInteractions:pT0Ref`	$1.0 - 3.0$
Exponent of $\sqrt{s}$ dependence, ϵ	`MultipartonInteractions:ecmPow`	$0.0 - 0.3$
Matter fraction contained in the core	`MultipartonInteractions:coreFraction`	$0.1 - 0.95$
Radius of the core	`MultipartonInteractions:coreRadius`	$0.1 - 0.8$
Range of color reconnection probability	`ColorReconnection:range`	$1.0 - 9.0$

Table 5.1: This table report the five parameters tuned for the underlying event in CP tunes, the variation ranges used for the sampling are shown in the last column. Table from [34]

for "CMS PYTHIA8". The tunes are performed changing the values of the coupling constant α_s for the ISR, FSR, hard scattering and MPI and the order of the evolution with the value of Q^2 for the interaction. Another difference among these tunes is the choice of the PDF set. The parameter that are tuned are shown in Table 5.1 with the associated variation ranges.

5.3.1 The distributions used

The observables distributions used for the tune are the following one:

- The pseudorapidity distribution of charged hadrons (p, K ,π) measured for an inclusive selection in inelastic proton-proton collisions

- Charged particle density and charged particle scalar p_T^{sum} in TransMIN and TransMAX regions at different $\sqrt{s}$ (1.96 TeV, 7 TeV, 13 TeV)

- The pseudorapidity distributions for single diffractive (SD) and non single diffractive (NSD) events selection

In the next chapter we will describe our tune using `mcnntunes` and in it we used the same distribution listed here.

5.3.2 Pyhtia configuration and the tunes

On the top section of Table 5.2 and Table 5.3 are reported the values for the `pythia8` parameters used in the CP tunes and, on the bottom, the 5 parameters resulting from the tune.

PYTHIA8 parameter	CP1	CP2
PDF Set	NNPDF3.1 LO	NNPDF3.1 LO
$\alpha_S(m_Z)$	0.130	0.130
`SpaceShower:rapidityOrder`	off	off
`MultipartonInteractions:EcmRef` [GeV]	7000	7000
$\alpha_S^{\text{ISR}}(m_Z)$ value/order	0.1365/LO	0.130/LO
$\alpha_S^{\text{FSR}}(m_Z)$ value/order	0.1365/LO	0.130/LO
$\alpha_S^{\text{MPI}}(m_Z)$ value/order	0.130/LO	0.130/LO
$\alpha_S^{\text{ME}}(m_Z)$ value/order	0.130/LO	0.130/LO
`MultipartonInteractions:pT0Ref` [GeV]	2.4	2.3
`MultipartonInteractions:ecmPow`	0.15	0.14
`MultipartonInteractions:coreRadius`	0.54	0.38
`MultipartonInteractions:coreFraction`	0.68	0.33
`ColorReconnection:range`	2.63	2.32

Table 5.2: CP1 and CP2 tunes settings are report here together with the values for the parameters tuned. CP1 and CP2 use a LO PDF set. CP1 α_s is different between matrix element calculation and MPI that use a value of 0.1365 ISR and FSR that instead use 0.130. While, CP2 use the same value for all processes, it is fixed at 0.130. In both cases α_s run with a LO evolution. Table from [34]

PYTHIA8 parameter	CP3	CP4	CP5
PDF Set	NNPDF3.1 NLO	NNPDF3.1 NNLO	NNPDF3.1 NNLO
$\alpha_S(m_Z)$	0.118	0.118	0.118
`SpaceShower:rapidityOrder`	off	off	on
`MultipartonInteractions:EcmRef` [GeV]	7000	7000	7000
$\alpha_S^{\text{ISR}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_S^{\text{FSR}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_S^{\text{MPI}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
$\alpha_S^{\text{ME}}(m_Z)$ value/order	0.118/NLO	0.118/NLO	0.118/NLO
`MultipartonInteractions:pT0Ref` [GeV]	1.52	1.48	1.41
`MultipartonInteractions:ecmPow`	0.02	0.02	0.03
`MultipartonInteractions:coreRadius`	0.54	0.60	0.76
`MultipartonInteractions:coreFraction`	0.39	0.30	0.63
`ColorReconnection:range`	4.73	5.61	5.18

Table 5.3: here are reported CP3, CP4 and CP5 tunes settings, and the results for the tune. The three tunes use an equal α_s value for al the processes, $\alpha_s = 0.118$ running with a NLO evolution. The difference between CP3 and the other two tune is that CP3 use a NLO PDF set while CP4 and CP5 a NNLO one. CP5 ISR emission is also ordered according to rapidity. Table from [34]

Below we are going to focus on CP5 tune. We are going to reproduce this tune using the same settings for `pythia8` but using a different tuning software: `mcnntunes`. In the next section `mcnntunes` working mode is explained.

5.4 MCNNTUNES

mcnntunes [36] is a Shower Monte Carlo generators tuning tool, that implement a tune procedure based on the use of Feed Forward Neural Networks (FFNNs). The advantage of using FFNNs have been described above and is that they are universal function approximators. This remove the polynomial bias present in professor tool.

mcnntunes offers two different operation modes: *PerBin Model* and *Inverse Model*. The first one is based on approach similar to the one in professor but where the response of the generator is parameterized using these FFNNs; the later is a new approach where the NN is trained to infers the parameters value from the experimental values of the bins in the distributions used.

In next two sections we are going to describe the two methods in details.

The two approach have the same starting point that is a sampling of the parameter space (e.g. for the UE analysis we use the parameters space shown in Table 5.1), than the generator is run with every sample. All these MC runs are going to build our dataset that is called *training set*.

5.4.1 Per Bin Model

PerBin Model is a parametrisation-based method. The main idea, as shown in Fig. 5.4 is to build a model (i.e. a neural network) for each bin in order to parameterize the generator output. Each NN takes the parameters values as input and return the bin value as output.

All these NNs are then trained feeding the MC runs from the training set. All the training set are trained using a gradient-based algorithm, as usual for feed forward neural networks, with mean squared error as loss.

Once, the NN is trained, the last step is the tune in which one actually get the best parameters. This step define a surrogate loss function for the tuning problem. In fact, the parameterization step return a model $h^{(i)}(\mathbf{p})$ for each bin, i, and where $\mathbf{p}$ is the vector of the parameters.

Than this surrogate loss function defined as:

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(h^{(i)}(\mathbf{p}) - h_{exp}^{(i)} \right)^2}{\sigma_{(i)}^2} \tag{5.7}$$

need to be minimized in order to evaluate the best estimation for the parameters. In mcnntunes this minimization is performed using the CMA-ES algorithm [41].

So the best estimation for the parameters is the configuration of parameters that minimize this χ^2.

Errors evaluation

The error evaluation for the PerBin Model is given by the definition of a confidence interval using the χ^2 function. In fact, as shown in section 9.6 and 9.7 of [42], for an estimators vector $\hat{h}(\mathbf{p}) = (\hat{h}^{(1)}(\mathbf{p}), \hat{h}^{(2)}(\mathbf{p}), \ldots, \hat{h}^{(n)}(\mathbf{p}))$ for the parameters $\hat{h}(\mathbf{p})$ the probability distribution function and the likelihood (the χ^2 in our case) in limit of
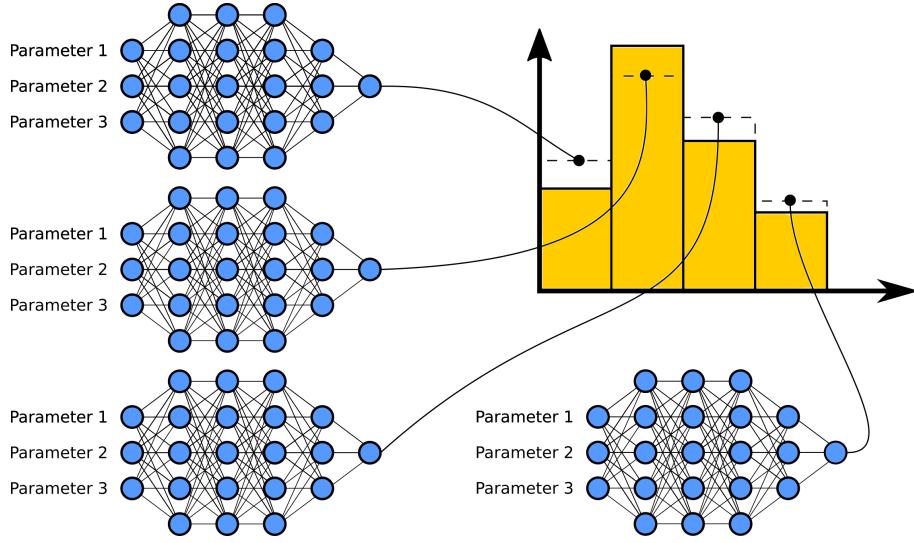
Figure 5.4: Figure from [36]

percentile	Q_γ				
	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
0.683	1.00	2.30	3.53	4.72	5.89
0.90	2.71	4.61	6.25	7.82	9.24
0.95	3.84	5.99	7.82	9.49	11.1
0.99	6.63	9.21	11.3	13.3	15.1

Table 5.4: The table report the values of the quantile Q_γ for different confidence level 0.683 is the row corresponding to the 1σ definition and is the one of our interest.

a large sample are Gaussian distributed. The probability distribution function for the estimators is

$$f(\hat{h}(\mathbf{p})|h(\mathbf{p})) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left[-\frac{1}{2}\left(\hat{h}(\mathbf{p}) - h(\mathbf{p})\right)^T V^{-1}\left(\hat{h}(\mathbf{p}) - h(\mathbf{p})\right)\right] \quad (5.8)$$

where T is the transpose vector and V^{-1} is the inverse covariance matrix. Can be shown that also the likelihood is Gaussian as the probability distribution function. So a changing in the parameter give a calculable variation in the χ^2

$$\frac{\chi^2(\text{c.i.})}{N_{dof}} = \frac{\chi^2_{min}}{N_{dof}} + \frac{Q_\gamma}{N_{dof}} \quad (5.9)$$

The variation is dependent on the number of parameters and on the chosen confidence level ($1\sigma = 0.683$ in our case) and a list of the values is reported in Table 5.4.

Then, the error is defined as the value of the parameters that give a deviation from the minimum value of the χ^2/N_{dof} equal to the Q_γ/N_{dof} value for a confidence level of 0.683, as defined in Eq. 5.9. An example for the evaluation in `mcnntunes` is shown in Fig. 5.5 where the green line is the quantities defined here in Eq. 5.9 and the errors are given by the intersections between this line and the blue line, that is the chi^2/N_{dof} for the different values of the parameter.
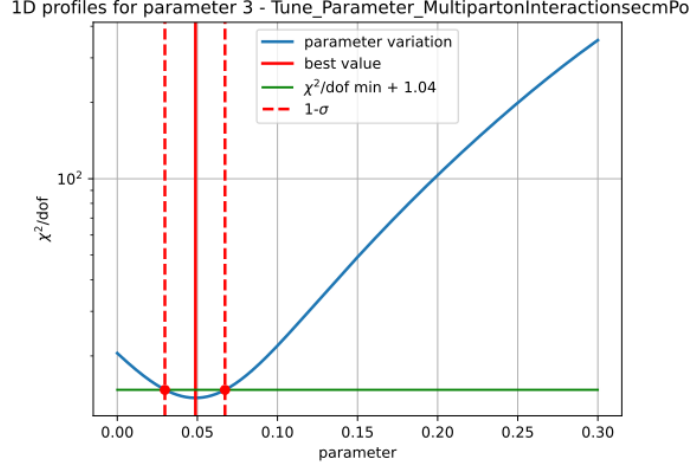
Figure 5.5: This figure shows the error evaluation in `mcnntunes` for the `MultipartonInteractions:ecmPow Pythia8` parameter. The blue line is the value of the $\chi2/N_{dof}$ evaluated for different parameter values. The best estimation is indicated by the vertical red line, while the green line is the quantity in Eq. 5.9. The error is evaluated from the intersection of blue and green lines.

5.4.2 Inverse Model

The Inverse Model is the most innovative tuning procedure introduced by `mcnntunes`. This model contrarily to the PerBin Model takes the histograms bins as input and returns parameters values as output. For the Inverse Model the NN used is only one as shown schematically in Fig. 5.6. What the Inverse Model try to do is to learn the inverted model of the generator. So starting from the observed values the model try to reproduce the parameters values necessary to get the histograms we use as input.

The model is build and then trained with the training set introduced before. Once the model is trained feeding the experimental data to the NN this can try to infer the values of the parameters required to get the output.
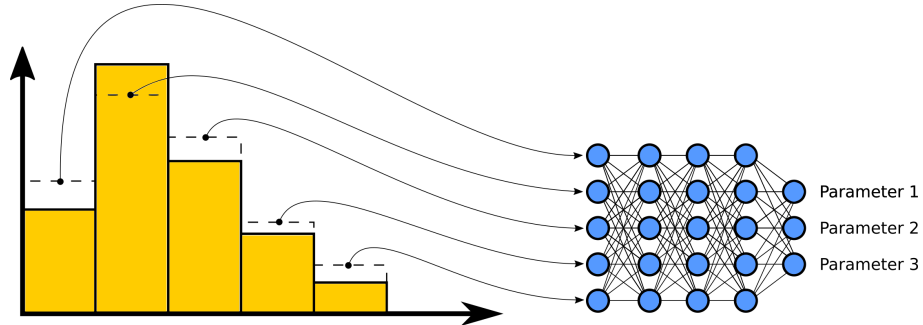

Figure 5.6: Figure from [36]

Errors evaluation

The errors are evaluated in a different way respect to PerBin Model in fact in the Inverse Model there is not a minimization step, and the error is evaluated by a re-

sampling of the experimental data using a *multivariate Gaussian Distribution*, as in Eq. 5.10 with a diagonal covariance matrix that have experimental uncertainties on the main diagonal.

$$f(x_i; h_{\exp}^{(i)}, \sigma_{\exp}^{(i)}) \, = \, \mathcal{N} \cdot \exp\left[-\frac{1}{2} \sum_{j=1}^{N_{bins}} \frac{\left(x_i - h_{\exp}^{(i)}\right)^2}{\sigma_{\exp}^{(i)}{}^2} \right] \tag{5.10}$$

So, a set of histograms is generated, then this is fed to the NN and a distribution of predictions is generated. An example is shown in Fig. 5.7, from this distribution one can compute the error by evaluating the standard deviation.
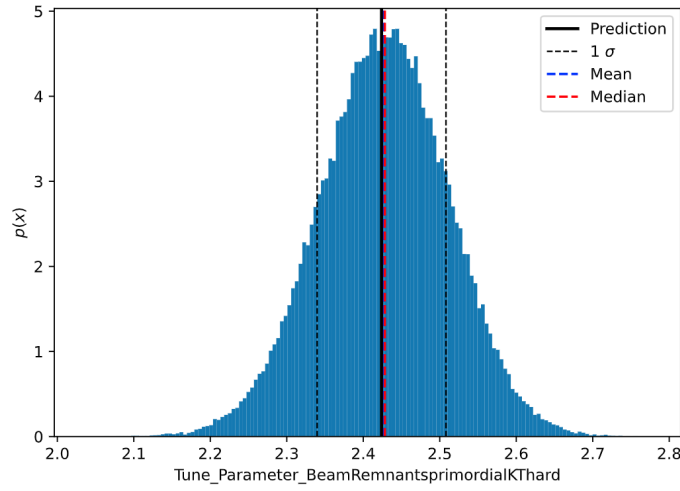

Figure 5.7: Predictions spread for the inverse model after that a Gaussian resample is performed

Note that this is a new method for the tune

Hyperparameters

Chapter 6

Our Tune for the Underlying Event

This section focus on our work in order to reproduce a similar tune to CP5 one, this is done to test the ability of MCNNTUNES of being one valid tool for the tuning of Monte Carlo generator with real data.

6.1   Introduction

We have performed a different tune for the underlying event using the same distributions, shown in Chapter 5.3.1 used in CP5 tune.

Just a quick reminder for the PYTHIA8 settings used in CP5 and in our tune: we employed bot the MCNNTUNES described in the previous chapter.

6.2   Per Bin Model results

The training set we use for the PerBin model is composed from approximately 2000 MC runs.
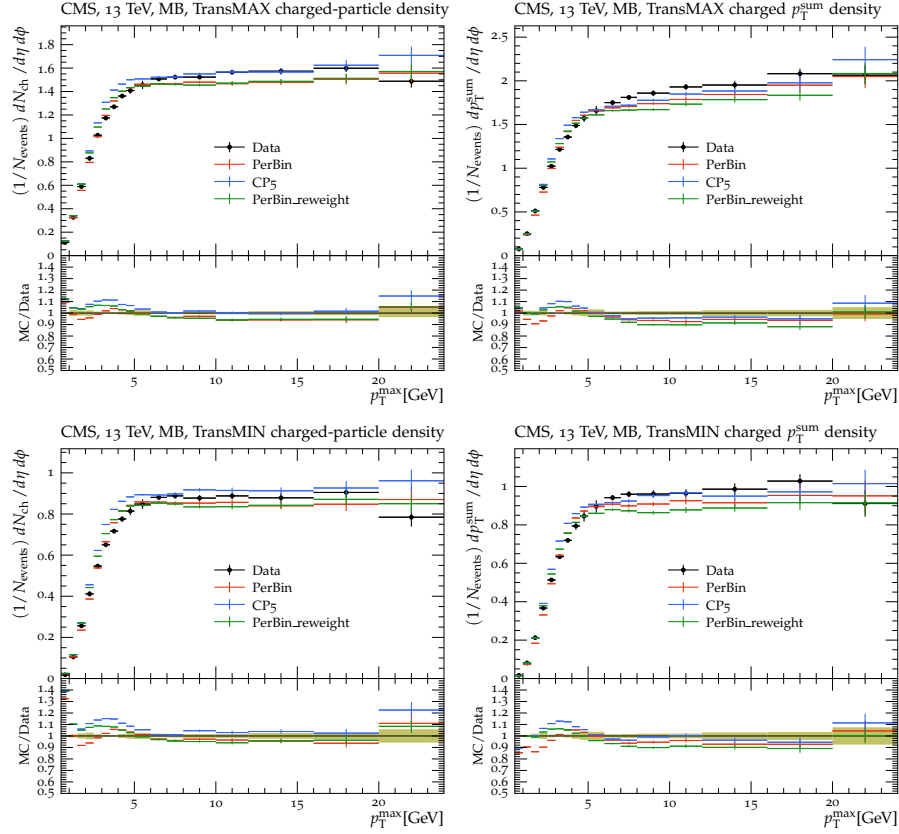
Fig. **??**

Figure 6.1: In this figure the data from the $\sqrt{s} = 13$ TeV CMS analysis [29] that show the transMAX charged particle density (upper left) and the charged p_T-sum density (upper right); the transMIN charged particle density (lower left) and the charged p_T-sum density (lower right) as a function of the transverse momentum of the leading charged particle. The CP5 tune is compared to our tune using the PerBin Model. Our tune (red line) is good as the CP5 (blue line) in describing the data. The first bins are the most important they have a smaller experimental error than the higher p_T data. Also the PerBin model with reweight (green line) seems really good in the description of the data.
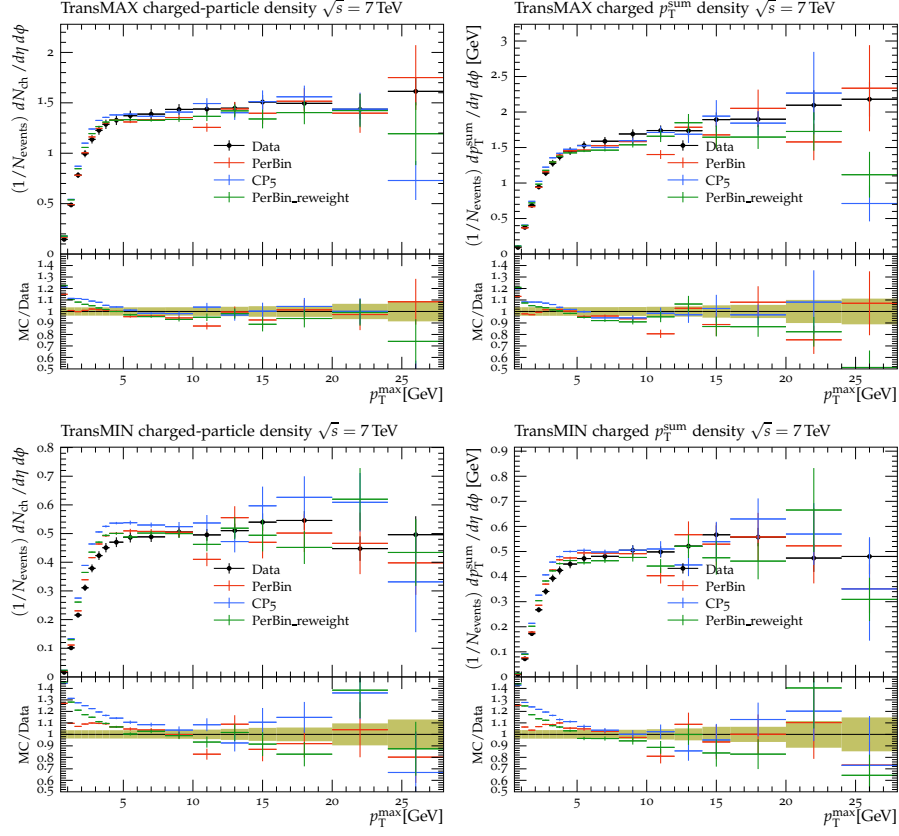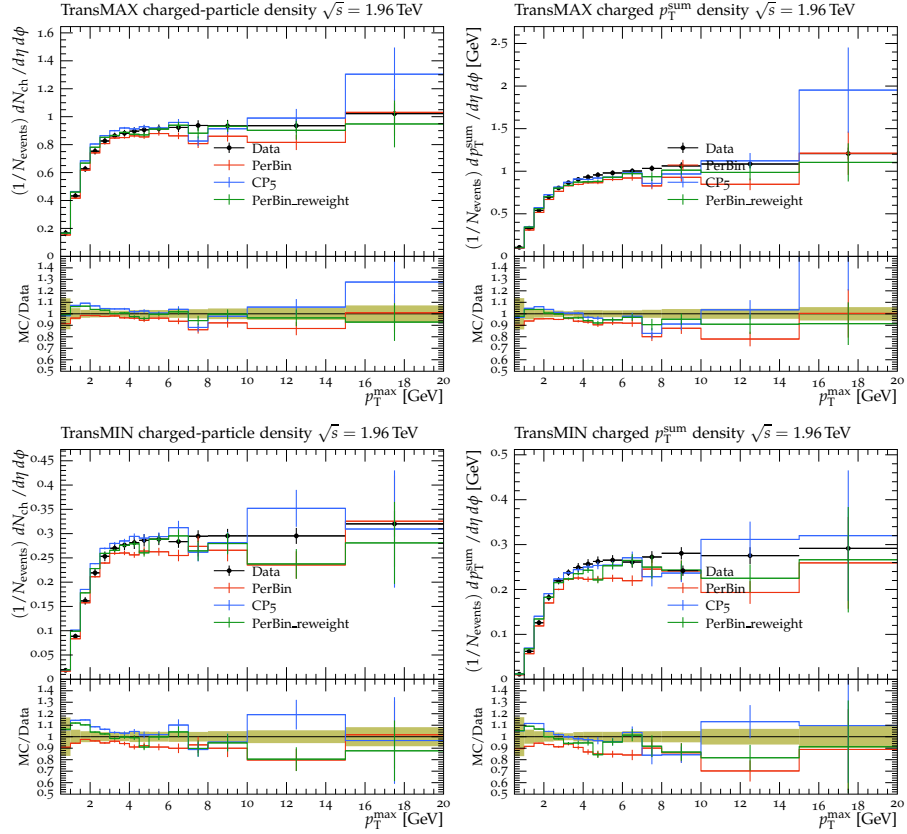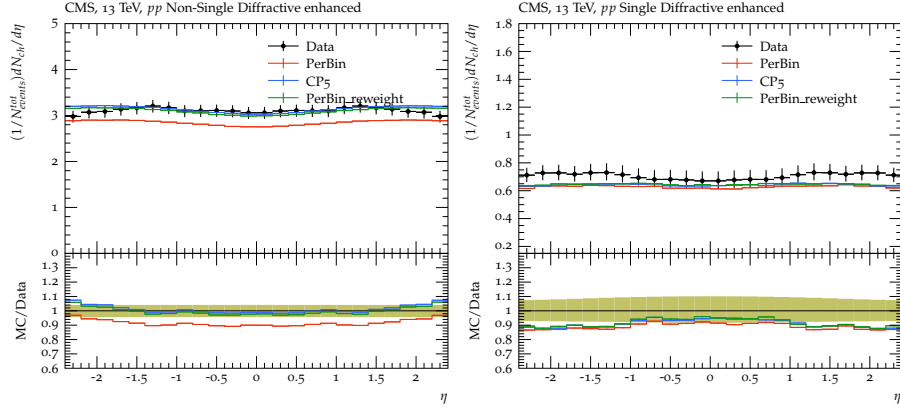
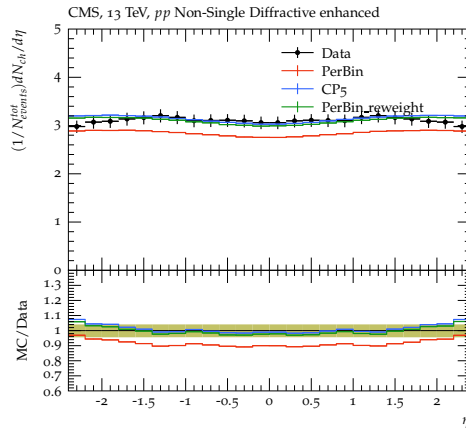**Figure 6.2:** ADD

Figure 6.3: ADD

Figure 6.4: ADD

Figure 6.5: ADD

Bibliography

[1] Steven Weinberg. A model of leptons, Nov 1967. URL https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[2] J. D. Bjorken. Asymptotic Sum Rules at Infinite Momentum. *Phys. Rev.*, 179: 1547–1553, 1969. doi: 10.1103/PhysRev.179.1547.

[3] Sidney D Drell and Tung-Mow Yan. Partons and their applications at high energies, 1971. ISSN 0003-4916. URL https://www.sciencedirect.com/science/article/pii/0003491671900716.

[4] J M Campbell, J W Huston, and W J Stirling. Hard interactions of quarks and gluons: a primer for lhc physics, Dec 2006. ISSN 1361-6633. URL http://dx.doi.org/10.1088/0034-4885/70/1/R02.

[5] L N Lipatov. The parton model and perturbation theory, 1975. URL http://cds.cern.ch/record/400357.

[6] Vladimir Naumovich Gribov and L N Lipatov. Deep inelastic ep scattering in perturbation theory, 1972. URL https://cds.cern.ch/record/427157.

[7] G. Altarelli and G. Parisi. Asymptotic freedom in parton language, 1977. ISSN 0550-3213. URL https://www.sciencedirect.com/science/article/pii/0550321377903844.

[8] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics., 1977.

[9] W.J. Stirling. private communication. URL http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html.

[10] F. Bloch and A. Nordsieck. Note on the radiation field of the electron, Jul 1937. URL https://link.aps.org/doi/10.1103/PhysRev.52.54.

[11] Toichiro Kinoshita. Mass singularities of feynman amplitudes, 1962. URL https://doi.org/10.1063/1.1724268.

[12] T. D. Lee and M. Nauenberg. Degenerate systems and mass singularities, Mar 1964. URL https://link.aps.org/doi/10.1103/PhysRev.133.B1549.

[13] A. Kulesza, G. Sterman, and W. Vogelsang. Electroweak vector boson production in joint resummation, 2002. URL https://arxiv.org/abs/hep-ph/0207148.

[14] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2, Jun 2015. ISSN 0010-4655. URL http://dx.doi.org/10.1016/j.cpc.2015.01.024.

[15] Manuel Bähr, Stefan Gieseke, Martyn A. Gigg, David Grellscheid, Keith Hamilton, Oluseyi Latunde-Dada, Simon Plätzer, Peter Richardson, Michael H. Seymour, Alexander Sherstnev, and Bryan R. Webber. Herwig++ physics and manual, Nov 2008. ISSN 1434-6052. URL http://dx.doi.org/10.1140/epjc/s10052-008-0798-9.

[16] T Gleisberg, S Hoeche, F Krauss, A Schaelicke, S Schumann, and J Winter. Sherpa 1. , a proof-of-concept version, Feb 2004. ISSN 1029-8479. URL http://dx.doi.org/10.1088/1126-6708/2004/02/056.

[17] E. Boos, M. Dobbs, W. Giele, I. Hinchliffe, J. Huston, V. Ilyin, J. Kanzaki, K. Kato, Y. Kurihara, L. Lonnblad, M. Mangano, S. Mrenna, F. Paige, E. Richter-Was, M. Seymour, T. Sjostrand, B. Webber, and D. Zeppenfeld. Generic user process interface for event generators, 2001.

[18] Stefano Catani, Frank Krauss, Bryan R Webber, and Ralf Kuhn. Qcd matrix elements + parton showers. *Journal of High Energy Physics*, 2001(11):063–063, Nov 2001. ISSN 1029-8479. doi: 10.1088/1126-6708/2001/11/063. URL http://dx.doi.org/10.1088/1126-6708/2001/11/063.

[19] Stefano Frixione and Bryan R Webber. Matching nlo qcd computations and parton shower simulations. *Journal of High Energy Physics*, 2002(06):029–029, Jun 2002. ISSN 1029-8479. doi: 10.1088/1126-6708/2002/06/029. URL http://dx.doi.org/10.1088/1126-6708/2002/06/029.

[20] Stefano Frixione, Paolo Nason, and Bryan R Webber. Matching nlo qcd and parton showers in heavy flavour production. *Journal of High Energy Physics*, 2003(08):007–007, Aug 2003. ISSN 1029-8479. doi: 10.1088/1126-6708/2003/08/007. URL http://dx.doi.org/10.1088/1126-6708/2003/08/007.

[21] Stefano Frixione and Bryan R. Webber. The mc@nlo 3.1 event generator, 2005.

[22] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012. doi: 10.1007/JHEP12(2012)061.

[23] Victor S. Fadin. BFKL resummation. *Nucl. Phys. A*, 666:155–164, 2000. doi: 10.1016/S0375-9474(00)00022-1.

[24] I. I. Balitsky and L. N. Lipatov. The Pomeranchuk Singularity in Quantum Chromodynamics. *Sov. J. Nucl. Phys.*, 28:822–829, 1978.

[25] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Patrick Groth-Merrild, Alberto Guffanti, Nathan P. Hartland, Zahari Kassabov, José I. Latorre, Emanuele R. Nocera, Juan Rojo, Luca Rottoli,

Emma Slade, and Maria Ubiali. Parton distributions from high-precision collider data. *The European Physical Journal C*, 77(10), Oct 2017. ISSN 1434-6052. doi: 10.1140/epjc/s10052-017-5199-5. URL http://dx.doi.org/10.1140/epjc/s10052-017-5199-5.

[26] FRANK SIEGERT. Monte-carlo event generation for the lhc, 2010. URL http://etheses.dur.ac.uk/484/.

[27] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2):31–145, 1983. ISSN 0370-1573. doi: https://doi.org/10.1016/0370-1573(83)90080-7. URL https://www.sciencedirect.com/science/article/pii/0370157383900807.

[28] Torbjorn Sjostrand. Jet Fragmentation of Nearby Partons. *Nucl. Phys. B*, 248: 469–502, 1984. doi: 10.1016/0550-3213(84)90607-2.

[29] Underlying Event Measurements with Leading Particles and Jets in pp collisions at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2015. URL https://cds.cern.ch/record/2104473.

[30] Serguei Chatrchyan et al. Measurement of the underlying event in the Drell-Yan process in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 72: 2080, 2012. doi: 10.1140/epjc/s10052-012-2080-4.

[31] A. M. Sirunyan et al. Measurement of the underlying event activity in inclusive Z boson production in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 07:032, 2018. doi: 10.1007/JHEP07(2018)032.

[32] Albert M. Sirunyan et al. Study of the underlying event in top quark pair production in pp collisions at 13 TeV. *Eur. Phys. J. C*, 79(2):123, 2019. doi: 10.1140/epjc/s10052-019-6620-z.

[33] Florian Bechtel. *The underlying event in proton-proton collisions*. PhD thesis, Hamburg U., 2009.

[34] Albert M Sirunyan et al. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80(1):4, 2020. doi: 10.1140/epjc/s10052-019-7499-4.

[35] Andy Buckley, Hendrik Hoeth, Heiko Lacker, Holger Schulz, and Jan Eike von Seggern. Systematic event generator tuning for the LHC. *Eur. Phys. J. C*, 65: 331–357, 2010. doi: 10.1140/epjc/s10052-009-1196-7.

[36] Marco Lazzarin, Simone Alioli, and Stefano Carrazza. MCNNTUNES: Tuning Shower Monte Carlo generators with machine learning. *Comput. Phys. Commun.*, 263:107908, 2021. doi: 10.1016/j.cpc.2021.107908.

[37] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh

Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

[38] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. 1958. URL https://doi.org/10.1037/h0042519.

[39] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(91)90009-T. URL https://www.sciencedirect.com/science/article/pii/089360809190009T.

[40] Moshe Leshno, Vladimir Ya. Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993. ISSN 0893-6080. doi: https://doi.org/10.1016/S0893-6080(05)80131-5. URL https://www.sciencedirect.com/science/article/pii/S0893608005801315.

[41] Nikolaus Hansen. The CMA evolution strategy: A tutorial. *CoRR*, abs/1604.00772, 2016. URL http://arxiv.org/abs/1604.00772.

[42] G. Cowan. *Statistical data analysis*. Oxford University Press, USA, 1998.