UNIVERSITÀ DEGLI STUDI DI PAVIA
DOTTORATO DI RICERCA IN FISICA - XXXVI CICLO

PRECISION PHENOMENOLOGY FOR NUCLEON FEMTOGRAPHY

MATTEO CERUTTI

Submitted to the Graduate School of Physics in partial fulfillment of the requirements for the degree of

DOTTORE DI RICERCA IN FISICA
DOCTOR OF PHYLOSOPHY IN PHYSICS

at the

University of Pavia

Supervisors: Dr. MARCO RADICI

Prof. ALESSANDRO BACCHETTA

Cover:

Precision Phenomenology for Nucleon Femtography

Matteo Cerutti

PhD Thesis - University of Pavia

Pavia, Italy, January 2023

A mamma e papà,

Contents

Chapter 1

Introduction

Nucleon femtography is a rapidly expanding topic in the domain of subnuclear physics. Its goal is to unveil the complex multi-dimensional distributions of partons (quark and gluons) inside nucleons. In other words, it seeks to provide multi-dimensional maps of the internal structure of protons and neutrons, which are the building blocks of matter, and they represent more than 99% of the visible mass in the visible Universe.

Through nucleon femtography, we can deepen our understanding of the strong interaction and of Quantum Chromodynamics (QCD), leading to valuable insights on some fundamental aspects such as confinement. It is clear that nucleon femtography can gain more and more precision in understanding the internal dynamics of quarks and gluons inside nucleons as experimental techniques and theoretical frameworks evolve with time.

This thesis focuses on the phenomenology of unpolarized transverse-momentum-dependent (TMD) distributions, where the need for precision is very important. Unpolarized TMDs are a generalization of collinear partonic densities, and provide three-dimensional maps of the momentum distribution of unpolarized partons within unpolarized hadrons. They can be accessed in various experimental measurements of particle scatterings in the Deep-Inelastic Scattering (DIS) kinematic regime.

The quest for precision in nucleon phenomenology emerges from the complicated interplay of theoretical calculations and experimental measurements. In fact, parton distributions such as TMDs provide information on the still poorly known regime of non-perturbative QCD, and can be extracted only through a fit to the available experimental data, based on theoretically justified factorization theorems. Therefore, it is crucial to achieve a high level of precision in order to obtain accurate interpretations of experimental data.

In this thesis, we firstly review the procedure of extracting unpolarized quark TMDs in the proton. Then, we illustrate the most recent work simultaneous extraction of TMD Parton Distribution Functions (PDFs) and TMD Fragmentation Functions (FFs), the latter providing information about the fragmentation of a quark into a detected hadron.

The work presented in this thesis represents at the moment the state-of-art of global extractions of unpolarized TMDs in the literature [1]. In fact, it is the most recent and sophisticated simultaneous extraction of TMD PDFs and TMD FFs. The level of sophistication of a global analysis is given by two ingredients: the number of experimental data sets included in the fit, determining how widespread is the information that we can get from experimental measurements; the degree of accuracy in the theoretical calculations, indicating how many terms we include in the perturbative expansion of calculable ingredients. In our analysis, we consider the most recent experimental data for Semi-Inclusive DIS (SIDIS) and Drell–Yan processes, including very precise measurements at the LHC. We compute theoretical predictions through the `NangaParbat`[1] computational tool for TMD studies, developed by the MAP Collaboration, at the current highest perturbative accuracy, namely at next-to-next-to-next-to-leading logarithmic accuracy (N^3LL).

This global analysis is at the cutting edge because it addresses for the first time the normalization issue of SIDIS measurements, that occurs when the theoretical accuracy increases beyond the leading order. This issue has been known in the literature since a decade, but it was overlooked in the past global analyses. Here, we make an important step forward in the identification of the source of this issue, and we propose a theory-based operative solution to overcome it. Because of this, we are able to reach an extremely good agreement between theoretical predictions and the majority of experimental data in both the shape and normalization, making an important breakthrough in unpolarized TMD phenomenology.

After discussing proton TMDs, we present a determination of unpolarized quark TMD PDFs in the pion [2]. In fact, it is interesting to understand the internal dynamics of partons inside the pion because it is the simplest particle in the hadronic spectrum, and it is the Goldstone boson of chiral symmetry breaking in the Standard Model (SM). Here, we perform an extraction of unpolarized pion TMDs through a fit of pion-nucleus Drell–Yan measurements. We include the largest number of available experimental data, and we make theoretical calculations at the current highest accuracy. Therefore, this new analysis is a significant step forward in the tomography of the structure of pions, which is far less known than the one of protons.

To complete the discussion on TMD extraction, we examine some example of variations in the general setup of the above phenomenological analyses. We show that in our fitting framework there are hints that TMDs do not significantly depend on their collinear input, as expected. Then, we study for the first time the compatibility among the experimental data sets included in a global fit of TMDs. We collect evidences that very precise high-energy Drell–Yan data have to be treated carefully.

In the final part of this thesis, we show how the precision reached by collinear PDF extractions can lead to the investigation of BSM physics in

---

[1]The `NangaParbat` code is available at https://github.com/MapCollaboration/NangaParbat

the QCD sector. Indeed, reaching high levels of precision in nucleon phenomenology not only plays a key role in improving our knowledge of the internal structure of hadrons, but has also implications in broader areas of particle physics. In fact, accurate determinations of partonic distributions can enter as fundamental ingredients in various calculations, including Standard Model predictions of cross sections and event generators.

As a consequence, parton densities play a crucial role in Beyond-Standard-Model (BSM) physics. In fact, they enter the determination of different SM parameters in the Electroweak sector, such as the W-boson mass m_W, or the weak mixing angle $\sin\theta_W$. Deviations from SM expectations can provide indirect evidence for BSM physics.

In fact, we demonstrate for the first time that we can identify BSM effects in the internal structure of protons. In particular, we investigate the possibility of strong parity violation in the proton's structure, providing the first phenomenological estimate of strong parity-violating PDFs, and showing that a high-energy photon may interact with more left-handed than right-handed quarks in an unpolarized proton [3]. Such phenomenological investigation is the first one ever attempted in this area of research, and it could open the door for future studies of the so-called "strong CP" problem, which is expected to be deeply connected to the explanation of the matter-antimatter imbalance in the Universe.

1.1 The internal structure of hadrons

The ultimate goal of hadronic physics is to unveil and understand all the details of the dynamics of the constituents of hadrons dictated by the strong interaction.

The strong interaction is much more elusive than the electroweak one. In fact, some fundamental particles such as electron and neutrino were discovered in the first half of the XX century, while the concept of fundamental building blocks of particles under the effect of strong interactions has been postulated only in the 1960s.

In 1964, Gell–Mann [4] and Zweig [5], independently, explained many properties of the multitude of hadronic particles observed in the mass spectrum by hypothesizing that they are made of spin-$\frac{1}{2}$ elementary particles named "quarks".[2] Thus, hadrons are bound states of a given combination of quarks or antiquarks, their antiparticles. Baryons are made by three quarks, while mesons by a quark-antiquark pair.

In the late 1964, Greenberg proposed to introduce a further quantum number to solve the issue of quarks of identical quantum states coexisting inside

---

[2]The term "quark" was chosen by Gell–Mann, taking inspiration from James Joyce's tale *Finnegans Wake*. Zweig proposed the name "ace" for the same fundamental particle, but Gell-Mann's terminology came to prominence once the concept of quark had been commonly accepted.

some hadrons (*e.g.*, the Δ^{++} particle) without violating the Pauli exclusion principle [6]. This new quantum number was later named by Gell–Mann "color charge", or simply "color".

It became immediately evident that strong interactions are strictly related to the property of color. In particular, this new quantum number led to the introduction of 8 new particles, named "gluons", which mediate the interactions between quarks and antiquarks, similarly to photons for electromagnetic interactions. Moreover, Nature allows to measure only hadrons, namely colorless bound states of quarks and/or antiquarks, whereas it is not possible to observe quarks as free particles. Such phenomenon is known as *color confinement*, and there is still no rigorous mathematical proof of it.

Due to this behavior of Nature, there was a long debate in the community of physicist about whether quarks are physical entities or a mathematical abstractions used to explain concepts that were not fully understood at the time. The only way to determine the actual existence of quarks was to investigate phenomena appearing in the internal structure of hadrons, being the direct consequence of their internal dynamics. Therefore, it seemed natural to shed light on the inside of protons by hitting them with high-energy electron beams, which act as probes of the internal structure of the target.

The quest for proving the existence of quarks culminated in the late 1960s with deep inelastic scattering experiments at the Stanford Linear Accelerator Center (SLAC) [7]. The observed scattering patterns are consistent with the existence of point-like constituents within protons and neutrons [8, 9]. This discovery confirmed the existence of quarks, and gave credit to Gell-Mann and Zweig's theoretical postulations.

In 1969, Feynman proposed the so-called *parton model* [10] to justify one of the main observation of the SLAC measurements, namely that the objects describing the internal structure of hadrons depend only on a combination of the squared momentum transfer Q^2 and the energy transfer ν, and not on the two separately [11]. This property was later called Bjorken scaling. In the parton model, this phenomenon was explained by assuming that the leptons in the beam collide against almost free point-like particles, the *partons*, inside the nucleons. Partons were later identified with quarks and gluons because experiments proved that they have the same quantum numbers.

All these observations and models were condensed into a quantum field theory that can include all the properties of quark dynamics. Such a theory, called Quantum Chromodynamics (QCD, in short) was formulated in the 1970s. It is a non-Abelian $SU(3)$ gauge theory that exhibits the property of *asymptotic freedom*: interactions between quarks become asymptotically weaker as the energy increase [12, 13]. This means that partons behave as point-like particles at high energy (short distance), while they are confined in hadron states at low energy (large distance). Thus, asymptotic freedom allows for the introduction of perturbative techniques in QCD calculations at high energy. This is the region where the strong coupling constant α_s is approximately small, and it is

often referred to as *perturbative QCD* (*pQCD*) region.

The first process to which QCD was applied is the DIS, aiming to provide an explanation of the Bjorken scaling. To do this, it is necessary to isolate the short-distance physics in the DIS cross section. This ensures that perturbative techniques can be properly applied. Thanks to the mathematical tool of operator product expansion (OPE) [14], it was shown that the lowest-order result in pQCD reproduces Bjorken scaling, which is, however, broken at higher orders. In fact, corrections with large logarithms of the momentum transfer spoil the perturbative treatment, and they have to be resummed through some evolution equations.

The deviation of Bjorken scaling was later confirmed by several experiments such as the NA2 experiment of the European Muon Collaboration (EMC) [15, 16] and the NA4 experiment of the Bologna–CERN–Dubna–Munich–Saclay (BCDMS) Collaboration [17, 18] (for a more recent review, see Ref. [19]).

In 1977, in a milestone work Altarelli and Parisi proposed a new way to describe partons and their interactions [20]. They introduced for the first time the concept of parton densities, called Parton Distribution Functions (PDFs) [21, 22], and fragmentation functions (FFs) [23, 24] and they described how PDFs and FFs change with the energy scale Q^2. These new ingredients can be defined only in the deep inelastic regime and at all orders of α_s. When it was realized that Dokshitzer [25], Gribov and Lipatov [26] derived independently the same equations of Altarelli and Parisi, QCD evolution were finally named DGLAP, after the initial of all their authors. Thanks to DGLAP equations, theoretical calculation can be compared to experimental data at any value of the energy scale Q^2 (at least, in the deep inelastic regime).

In order to perform QCD predictions for an experimental observable, one must consider that the cross section contains both short- and long-distance physics; therefore, it is not directly calculable in pQCD. However, there are some classes of processes for which it is possible to formally separate long-distance from short-distance behaviors in a systematic way. In other words, there are factorization theorems that allow the application of perturbative techniques to some key processes involving hadrons (denoted with h), such as DIS process $lh \to l'X$ (where l is a lepton and X means unobserved hadrons), Drell–Yan process $h_1 h_2 \to l^+ l^- X$, or single hadron production from electron-positron annihilation, $e^+ e^- \to hX$. In the 1980s, Collins, Soper and Sterman explicitly demonstrated factorization theorems for these processes [27].

In high-energy processes, it is customary to assume that for a hadron moving in a certain dominant direction, usually named "longitudinal", all partons move approximately collinear with it, sharing a fraction x of its longitudinal momentum. Analogously, a hadron detected in a final state is assumed to move almost collinear with the corresponding fragmenting parton, sharing a fraction z of its (longitudinal) momentum. In this collinear framework, factorization theorems have been demonstrated for all process involving hadrons in the initial or final state, or both. Thanks to collinear factorization, it is possi-

ble to express experimental observables in terms of perturbative subprocesses (at the partonic level) and universal parton distributions (collinear PDFs) or fragmentation functions (collinear FFs) [28].

Nowadays, the phenomenological studies of PDFs have reached a very high level of sophistication [29–31] and also make use of deep concepts of Machine Learning (ML) in computational frameworks. Despite a similar level of sophistication, FFs are less known than PDFs [32, 33], at least for unpolarized distributions. All polarized distributions are less known.

1.2 Transverse-Momentum-Dependent distributions

Parton distributions in the collinear framework can be seen as one-dimensional maps in momentum space of the internal structure of a specific hadron. In fact, at leading order they represent the probability density of finding a quark (or a gluon) inside a hadron carrying a fraction x of its momentum in the longitudinal direction. However, this interpretation is not so straightforward when QCD corrections are introduced [34].

Although the level of agreement with different experimental measurements has reached an extremely good quality, there are experimental evidences that cannot be interpreted within the collinear framework [35–38]. Along with azimuthal single-spin asymmetries (SSA) [39–42], the violation of the Lam-Tung sum rule [43–46], namely an observed sizeable asymmetry in the azimuthal distribution of Drell–Yan lepton pairs not explained in pQCD, indicates that the annihilating partons have an intrinsic transverse momentum, not induced by QCD radiation. Perhaps, the most renowned example is the EMC experiment at CERN [47], where for the first time it became evident that the spin of the proton is only partially given by the helicities of quarks. Since few years, we know that the helicity of gluons does not saturate the spin sum rule, although with large errors. Therefore, the question is if the remaining contribution is due to the orbital motion of partons inside the proton itself (for a recent review, see Ref. [48]).

These open questions led to the conclusion that partons inside hadrons have non-negligible intrinsic momentum in the transverse plane with respect to the hadron's longitudinal direction. This kind of momenta are integrated out in collinear framework because they are considered as much smaller than the longitudinal (dominant) momentum. However, Nature shows that they have to be accounted for a complete description of hadrons, because they produce experimentally observable effects.

Without integrating out intrinsic transverse momenta, parton distributions acquire a dependence also on the transverse momentum $\boldsymbol{k}_\perp$ of a quark with respect to the hadron direction of motion. This picture is illustrated by Transverse-Momentum-Dependent PDFs (TMD PDFs), which represent three-dimensional generalizations of PDFs in collinear framework: they are 3-D maps in momentum space of partons inside hadrons (see, *e.g.*, Refs. [49–51]).

$N\backslash q$	U	L	T
U	f_1		
L		g_1	
T			h_1

$N\backslash q$	U	L	T
U	f_1		$h_1^\perp$
L		g_{1L}	$h_{1L}^\perp$
T	$f_{1T}^\perp$	g_{1T}	$h_1, h_{1T}^\perp$

Table 1.1: Leading twist parton densities. Left panel: collinear PDFs depending on the longitudinal momentum fraction x. Right panel: TMD PDFs depending on the longitudinal momentum fraction x and the transverse momentum $\boldsymbol{k}_\perp$. The letters U, L, T stand for unpolarized, longitudinally polarized and transversely polarized nucleons (rows) and quarks (columns), respectively.

Similarly to partonic distributions, TMD FFs depend also on the transverse momentum $\boldsymbol{P}_\perp$ of the detected hadron with respect to the fragmenting parton direction of motion, and they can be seen as a 3-dim picture of the hadronization process (see, *e.g.*, Refs. [52, 53]).

At twist 2, namely if we consider only leading-power contributions in the $1/Q$ expansion of factorization theorems, there are only 3 quark PDFs in collinear framework; they are listed in the left panel of Tab. 1.1, where the letters U, L, T, denote unpolarized, longitudinally and transversely polarized particles, respectively. In the case of quark TMD PDFs, there are more possibilities of correlations between spin and quark transverse momentum. The 8 possible quark TMDs are listed in the right panel of Tab. 1.1 [51, 54–57]. In this thesis, we focus on the extraction of unpolarized quark TMD PDFs, which are the most studied in TMD phenomenology. Also some of the polarized TMDs have been extracted in the literature either in the parton model approach [58–62] or in the TMD factorization approach [63–67]. Similar decompositions of the fragmentation correlator can be worked out also for collinear FFs and TMD FFs, as indicated in Tab. 1.2. The majority of the investigations in TMD phenomenology are addressed to unpolarized quark TMD FFs, and very few analyses have been performed for the polarized ones [59, 60, 67–69]

$h\backslash q$	U	L	T
U	D_1		
L		G_1	
T			H_1

$h\backslash q$	U	L	T
U	D_1		$H_1^\perp$
L		G_{1L}	$H_{1L}^\perp$
T	$D_{1T}^\perp$	G_{1T}	$H_1, H_{1T}^\perp$

Table 1.2: Leading twist fragmentation functions. Left panel: collinear FFs depending on the longitudinal momentum fraction z. Right panel: TMD FFs depending on the longitudinal momentum fraction z and the transverse momentum $\boldsymbol{P}_\perp$. The letters U, L, T stand for unpolarized, longitudinally polarized and transversely polarized hadron (rows) and fragmenting quarks (columns), respectively.

The foundation of the present formalism for TMDs can be traced back to the work of Collins, Soper and Sterman, who built a systematic framework to

make theoretical prediction for TMD-sensitive observables [27, 70–72]. They provided fundamental insights on TMD evolution equations, on the separation between perturbative and nonperturbative components of TMD distributions, and on the methods to demonstrate factorization theorems, de facto opening the door for future phenomenological analyses. Indeed, TMDs can be only partially determined by means of well-established perturbative methods that take into account collinear and soft radiation to all orders. Since perturbative calculations in QCD become unreliable for values of transverse momentum close to the Landau pole (Λ_{QCD}), nonperturbative components of TMDs have to be introduced, and they have to be determined through fits to experimental data.

In this thesis, we will focus on the unpolarized TMDs $f_1(x, \boldsymbol{k}_\perp^2)$ and $D_1(z, \boldsymbol{P}_\perp^2)$. Because of factorization theorems [73], the information encoded in TMDs can be accessed through a fit of experimental data for Drell–Yan (DY) process, semi-inclusive DIS (SIDIS), or back-to-back hadron production from electron-positron annihilation. In particular, the DY cross section can be written as a convolution of two TMD PDFs, the cross section of e^+e^- annihilations as a convolution of two TMD FFs, and the SIDIS one as the convolution of one TMD PDF and one TMD FF. Therefore, the availability of measurements of different processes in different experiments makes it possible to test factorization theorems and to extract TMD PDFs and FFs through so-called global fits.

In the past, several works studied the nonperturbative transverse momentum components of parton distributions in Drell–Yan or in SIDIS, even without mentioning the concept of TMDs (see, *e.g.*, Refs. [74–77]). In more recent works, the extraction of TMDs have been performed through fits of DY [78–83] or SIDIS data [84, 85], or both [86–88]. Moreover, TMDs can be determined in the so-called parton branching approach [89–92] by solving evolution equation with an iterative method similar to parton shower, but including the transverse momentum of partons.

Alternatively, a complementary approach to calculate nonperturbative quantities from first principles is given by large-scale numerical simulations of lattice QCD. These methods have been successfully used to study quark masses [93, 94], hadron masses [95], decay constants and form factors [96, 97], CKM matrix elements [98, 99], the strong coupling constant [100], and many more. An overview of recent lattice QCD results is given in Ref. [101]. There are very recent attempts to calculate partonic distributions on the lattice in both collinear [102] and TMD [103] frameworks (see Ref. [104], for a recent review), but the computational methods require much effort, and further steps have to be made to improve the reliability of the results, particularly to calculate TMD PDFs as a function of the partonic transverse momentum.

In this thesis, we discuss the extraction of unpolarized quark TMDs from a global fit that represents the state of the art in the current literature.

A precise knowledge of TMDs is important not only to understand the

confined motion of quarks and gluons inside the hadron, but also to improve the reliability of predictions involving TMDs. In fact, the perturbative part of TMDs may be dominant in the domain of high-energy colliders, but also the nonperturbative components become relevant when experimental uncertainties are extremely small, as for experimental data at the LHC (see, *e.g.*, Ref. [105]). Moreover, a precise determination of unpolarized TMDs plays a key role also in the extraction of polarized TMDs. In fact, the experimental measurements sensitive to polarized physics are usually released as single-spin or double-spin asymmetries, which contain the unpolarized cross section in the denominator.

Looking at future developments in hadronic physics, important opportunities to explore the 3D internal structure of hadrons could be offered by the (possible) so-called JLab22 upgrade [106], an improvement in energy and luminosity of the CEBAF accelerator at the Jefferson Lab (JLab), and by the Electron-Ion Collider (EIC) [107], a new facility which is going to be built at the Brookhaven National Laboratory (BNL). Therefore, in the near future it is of fundamental importance to produce precise and reliable theoretical predictions in both the perturbative and nonperturbative domains, especially in view of the foreseen luminosity that could be achieved.

1.3 Effects of new physics through precise nucleon phenomenology

Nowadays, there are several works that make use of the available information on the internal structure of the proton to identify Beyond-Standard-Model (BSM) effects. Such effects are not directly related to the internal structure of the proton itself, but understanding the latter is needed to recognize the former. For instance, the determination of the running weak mixing angle $\sin\theta_W$ at high energy scales has been performed by studying the Drell–Yan process at large invariant mass of the final lepton pair [108–112]. In this context, an accurate knowledge of PDFs in collinear framework is fundamental because they are needed to perform the calculation of the experimental observable. Moreover, QCD is important in the determination of hadronic contributions to the muon anomalous magnetic moment $g-2$ [113, 114]. Also, an accurate data analysis for the extraction of the W-boson mass should consider flavor dependence of partonic intrinsic transverse momentum [105]. In other words, since hadronic contributions can be crucial to understand the discrepancies between experimental data and purely perturbative EW calculations, hadronic physics plays a key role as an input to BSM physics searches in the EW sector.

Since QCD is less under control than EW physics, investigations of BSM effects in the QCD sector are more difficult. In this thesis, we show that the knowledge of some aspects of QCD has reached such a precision that it is possible to estimate effects of BSM physics even inside the structure of nucleons.

The presence of operators that violate parity (P) or charge-parity (CP) symmetries is admitted in the QCD Lagrangian. In the last decades, a particular operator with these features, the so-called "θ-term", has been studied. This BSM term is usually not included in the QCD Lagrangian because it should give rise to a nonzero electric dipole moment for the neutron (nEDM), while experiments put very small upper limits on it [115]. Hence, an upper limit to the value of θ is set to 10^{-12}. However, there are no theoretical reasons that prevent it to be of order 1. This is the so-called "strong CP problem". A possible explanation can be found in the theory of Peccei and Quinn [116].

Addressing the strong CP problem could offer a solution to one of the biggest questions of cosmological physics, namely the baryon asymmetry observed in the Universe. The presence of CP violation emerges as a good candidate to account for the detected excess of matter over antimatter in the Universe. In fact, in the scenario where matter and antimatter are in perfect equilibrium, the universe would be made up only by radiation, without any matter. The existence of matter imposes that CP symmetry must be somehow broken. However, weak CP violation is not sufficient to explain the observed imbalance. A theoretical model that includes strong CP violation, as the one by Peccei and Quinn [116], could provide an explanation of the observed imbalance, if confirmed by experiment.

In this thesis, we will not investigate the origin of strong parity violation. We will assume that the QCD Lagrangian can contain parity-violating terms. Then, we explore the consequences in DIS experiments. We show that new collinear PDFs appear in the partonic correlation function, which produce new structure functions in the DIS cross section. A similar discussion for parity-violating di-hadron fragmentation functions was presented in a pioneering work in Ref. [117], and more recently for PDFs and FFs in Refs. [118, 119].

The main innovation of the study reported in this thesis is the phenomenological estimate of the size of the newly introduced parity-violating PDFs through a fit of the available experimental data.

Identifying signals of BSM physics is surely much simpler in the EW sector than in QCD. Indeed, the internal structure of nucleons can be investigated only through nonperturbative physics, and is affected by several power-suppressed corrections, which can originate effects that may be wrongly connected to new physics. In this thesis, these corrections are somewhat under control, namely they are included in the analysis or they are estimated much smaller than the required accuracy.

The phenomenological estimate of parity violation reported in this thesis is the fist one in the literature, and could possibly open a new door for the investigation of new physics beyond the Standard Model. This is an example of how QCD phenomenology can be used to study BSM physics, if the precision of experimental data and theoretical predictions is sufficiently high.

1.4   Outline of the thesis

The main topics covered in this thesis are:

❍ the simultaneous extraction of unpolarized quark TMD PDFs in the proton and TMD FFs through a global fit of available experimental data; the work is conventionally named MAP22 [1]

❍ the extraction of unpolarized quark TMD PDFs in the pion through a fit of available experimental data; the work is conventionally named PionMAP22 [2]

❍ the phenomenological estimate of strong parity violation in the internal structure of the proton [3].

The author of this thesis was one of the main developers of the `NangaParbat` code, which is the computational tool for TMD studies employed to perform the analyses presented in this thesis, and was one of the main authors of the papers on which this thesis is based.

This dissertation is organized as follows. In Ch. 2, the theoretical framework of TMDs is discussed, addressing the factorized expressions of the experimental observables of Drell–Yan process and Semi-Inclusive DIS (SIDIS), which are the two main sources of information on TMDs. Then, we discuss the theoretical ingredients needed to build a TMD distribution, namely TMD evolution and its matching onto collinear distributions, resummation of large logarithms, the logarithmic ordering, and its nonperturbative content. Finally, we address the normalization issue of SIDIS observables when increasing the perturbative accuracy.

The phenomenological analysis of the MAP22 global fit is illustrated in Ch. 3. We present the global data set included in the analysis, the criteria introduced to restrict the analysis in a reliable region where TMD factorization holds, and we discuss how experimental and theoretical uncertainties are treated. Then, after briefly outlining the parameterization introduced for the nonperturbative parts of TMDs, we show the results of the MAP22 global fit. We discuss the global χ^2 as an estimator of the quality of the fit, we show the results for the fitted parameters, we display the TMD PDFs and FFs in different kinematic regions, plus some other ingredients, such as the average squared transverse momenta and the Collins–Soper kernel, a fundamental ingredient in TMD evolution. Also, we discuss variations of the baseline fit configuration, namely variations of its perturbative accuracy, of the choice of the prescription to introduce the nonperturbative TMD content, and of the cut imposed on the transverse momentum to restrict the analysis in TMD region. Finally, we estimate the impact on the TMD uncertainty bands of new data from the JLab22 upgrade and the EIC machine.

In Ch. 4, we present the phenomenological analysis of the PionMAP22 fit. Again, we describe the data set included in the analysis, the criteria to

restrict the analysis in TMD region, and the treatment of experimental and theoretical uncertainties. We discuss the choice for the nonperturbative part of TMD PDFs in the pion, the results of the fit, and we display the extracted pion TMDs, making also a comparison with the MAP22 proton ones. Finally, we perform theoretical predictions for upcoming data from the COMPASS Collaboration.

Ch. 5 is devoted to some investigation on possible modifications of the usual setup of our TMD phenomenological analyses. Specifically, we discuss a possible modification on the theoretical input of collinear PDFs in order to get information about their role in the extraction of TMD distributions (through a fit of experimental data). After that, we present a study on the identification of possible inconsistencies among different data sets included in a TMD analysis. Based on previous studies in PDFs determinations, we show how problematic measurements can be identified and, eventually, removed from the baseline data set through objective criteria. We collect evidences that this procedure could produce some benefits also in TMD phenomenology.

The first phenomenological estimate of strong parity violation signals in the internal structure of nucleons is discussed in Ch. 6. We illustrate the formalism of parity-violating (PV) PDFs, their contribution to DIS structure functions, and the experimental observable sensitive to this newly introduced effects. We describe our first model for PV PDFs. We discuss the available experimental data introduced in the phenomenological analysis and, finally, we present the results.

Finally, a brief overview of the results illustrated in this thesis and an outlook on possible future developments are reported in Ch. 7.

Chapter 2

Factorized framework for TMD unpolarized observables

As already mentioned in Ch. 1, the possibility to extract information from different processes about nonperturbative objects describing the internal structure of hadrons, such as TMDs or collinear parton densities, is a key feature of QCD. This allow us to draw a global picture of the effects caused by the internal dynamics of partons inside hadrons.

In this chapter, we describe the theoretical framework used to make predictions for the experimental processes that are included in global analyses of unpolarized TMDs, namely the Drell–Yan (DY) and Semi-Inclusive DIS (SIDIS) processes.

In hadron tomography, the Drell-Yan (DY) process is of fundamental importance to get clean information on the transverse-momentum content of hadrons, since it can be described in terms of TMD PDFs.

At the same time, the SIDIS process has been proven to be a fundamental tool to investigate several aspects of perturbative and nonperturbative QCD, which sometimes cannot be accessed through the Drell-Yan process, such as the framgmentation of colored quarks into colorless hadrons. This kind of process is important to perform a simultaneous analysis of TMD PDFs and TMD Fragmentation Functions (TMD FFs).

In Sec. 2.1, we give an overview of the TMD factorization formula for the different Drell-Yan obsevables measured by different experimental collaborations, which are included in recent global analyses. In Sec. 2.2, we discuss the TMD factorization formula for SIDIS obsevables measured by different experimental collaborations. Then, the formalism of TMD distribution is introduced in Sec. 2.3, including their perturbative and nonperturbative contents and their evolution equations. Finally, in Sec. 2.4 the problem related to the normalization of SIDIS observable is discussed. In particular, the formalism of Refs.[1, 2, 81] is taken into consideration.
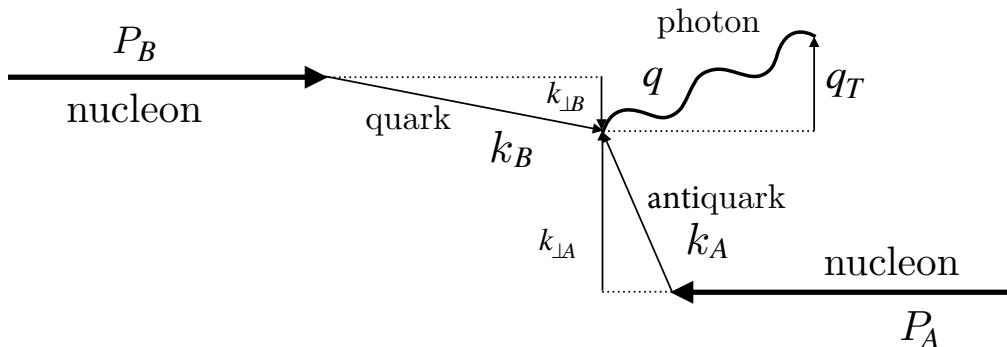
2.1 Drell–Yan observables

Figure 2.1: Diagram describing the relevant momenta involved in a DY process in the center-of-mass (CM) reference frame. A virtual photon with momentum q is produced by the annihilation of a quark and an antiquark with momenta k_B and k_A, respectively, inside two protons with momenta P_B and P_A (defining the reference axis). The quark (antiquark) has a transverse momentum $\boldsymbol{k}_{\perp B}$ ($\boldsymbol{k}_{\perp A}$), which is not measured. The total measured transverse momentum of the photon (or the final-state lepton pair) is $\boldsymbol{q}_T$. At large values of q^2, the longitudinal components are all much larger than the transeverse components. In this regime, $\boldsymbol{q}_T = \boldsymbol{k}_{\perp A} + \boldsymbol{k}_{\perp B}$.

In Drell–Yan process, two hadrons h_A and h_B with four-momenta P_A and P_B, respectively, collide with center-of-mass energy squared $s = (P_A + P_B)^2$. In the final state, a pair of lepton $\ell^-(l')$ and an antilepton $\ell^+(l)$ is measured, while all other hadrons X composing the final state are not detected. We neglect the lepton masses, since they are only a small fraction of the hadrons ones. The process can be represented as follows (see Fig. 2.1 for a graphical representation):

$$h_A(P_A) + h_B(P_B) \longrightarrow \ell^+(l) + \ell^-(l') + X \ . \tag{2.1}$$

The Drell–Yan process was discussed for the first time in Ref. [120], where the authors considered only that the result of the collision is the production of a virtual photon γ^* with four-momentum q and invariant mass $Q = \sqrt{q^2}$, which eventually decays in a lepton-antilepton pair with four-momenta l and l' constrained by the momentum conservation rule, $q = l + l'$.

With the advent of the high-energy physics colliders, it has been possible to produce also a Z^0-boson from the collision of two hadrons. In the literature, this kind of process is usually named "Z-boson production". Since the measured final-state is still a lepton-antilepton pair, we consider it as same as Drell–Yan process throughout this thesis.

We consider the center-of-mass (CM) frame as the frame of reference, where we identify the light-cone basis formed by the four-vectors P_A and P_B. Hence, we identify the longitudinal direction $\hat{z}$ as the axis of the hadronic collision: one of the two hadronic four-momenta P_A and P_B can be chosen as a reference direction. The plane of the collision $\hat{x}\hat{z}$ can be identified by the vectors of the hadron and the virtual boson, being the $\hat{y}$-axis chosen accordingly (for a discussion on different choices for the frame of reference see Ref. [121]). In this frame of reference, $\boldsymbol{q}_T$ is the photon transverse momentum with respect to the hadrons' direction of motion.

We define the absolute value of the transverse momentum and the rapidity of the vector boson or, equivalently, of the lepton-antilepton pair as

$$|\boldsymbol{q}_T| = \sqrt{q_x^2 + q_y^2}, \qquad y = \frac{1}{2} \ln\left(\frac{q_0 + q_z}{q_0 - q_z}\right). \qquad (2.2)$$

In order to extract unpolarized TMDs, we need to compare experimental data to the theoretical predictions of the cross section of this process integrated over the azimuthal angle of the intermediate vector boson (for a detailed diagram of the process, see Ref. [122]), and with two unpolarized hadrons in the initial state. Namely, the object of interest is the transverse-momentum spectrum of the vector boson. Such an observable can be written in terms of two structure functions , F_{UU}^1 and F_{UU}^2 [121, 123]. The structure function F_{UU}^2 is suppressed in the kinematic limit where TMD factorization holds, namely where $M^2 \ll Q^2$ (with M the mass of the incoming hadrons) and $\boldsymbol{q}_T^2 \ll Q^2$. In fact, F_{UU}^2 is a twist-4 object, and it provides a correction of the cross section at the order $O\left(\frac{M^2}{Q^2}\right)$. It can be written in terms of the $F_{UU}^{\cos 2\phi}$ structure function[1], according to the Lam–Tung sum rule [124], which is exact in collinear framework at $O(\alpha_S)$, while it is violated at higher order [125]. On the experimental side, violations of the Lam–Tung sum rule have been observed due to unexpectedly large $\cos 2\phi$ modulation of the cross section [43, 44, 46]. Therefore, the F_{UU}^2 structure function can be described in terms of the $F_{UU}^{\cos 2\phi}$ one only at $O(\alpha_S)$ in the large-$|\boldsymbol{q}_T|$ region, where the most of its integral is expected to be situated. In the complementary region, it is suppressed as a twist-4 object, and it is expected to give a small contribution (few %) to the DY cross section, at least at large Q [123].

At leading twist and in the $|\boldsymbol{q}_T| \ll Q$ region, the expression of the cross section is

$$\frac{d\sigma^{\mathrm{DY/Z}}}{d|\boldsymbol{q}_T|\, dy\, dQ} = \frac{16\pi^2 \alpha^2}{9Q^3} |\boldsymbol{q}_T|\, \mathcal{P}\, F_{UU}^1\big(x_A, x_B, |\boldsymbol{q}_T|, Q\big) + O\left(\frac{|\boldsymbol{q}_T|}{Q}\right) + O\left(\frac{M^2}{Q^2}\right),$$
$$(2.3)$$

where α is the (running) electromagnetic coupling, x_A and x_B are the longitudinal momentum fractions of the quark or antiquark acting in the collision at

---

[1]The $F_{UU}^{\cos 2\phi}$ structure function determines the $\cos 2\phi$ modulation of the unpolarized DY cross section.

the partonic level, and $\mathcal{P}$ is the phase space factor.

Such a phase-space reduction factor is included to account for potential cuts on the final-state leptons kinematics, which turns out to give a relevant impact when high-precision data are taken into account (see, *e.g.*, a recent analysis in Ref. [126]). It is important to note that an additional parity-violating term contributes to the cross section in presence of cuts on single-lepton variables [127]. This term can be originated either by the EW sector or the QCD sector, as will be discussed in Ch. 6. However, in Ref. [81] it was shown that this contribution is negligible in the experimental conditions we are taking into consideration.

The structure function F_{UU}^1 contributes to the cross section at every point of the phase-space, but the theoretical formalism that can be used to build its expression depends on the considered kinematic region. At low transverse momentum $\boldsymbol{q}_T^2 \ll Q^2$, TMD factorization allow us to express the structure function as a convolution over the partonic transverse momenta of two TMD PDFs:

$$
\begin{aligned}
F_{UU}^1 & \big(x_A, x_B, |\boldsymbol{q}_T|, Q\big) \\
& = x_A\, x_B\, \mathcal{H}^{\mathrm{DY}}(Q, \mu) \sum_{a=q,\bar{q}} c_a(Q^2) \int d^2 \boldsymbol{k}_{\perp A}\, d^2 \boldsymbol{k}_{\perp B}\, f_1^a(x_A, \boldsymbol{k}_{\perp A}^2; \mu, \zeta_A) \quad (2.4) \\
& \quad f_1^{\bar{a}}(x_B, \boldsymbol{k}_{\perp B}^2; \mu, \zeta_B) \delta^{(2)}(\boldsymbol{k}_{\perp A} + \boldsymbol{k}_{\perp B} - \boldsymbol{q}_T).
\end{aligned}
$$

In the above equation, $\mathcal{H}^{\mathrm{DY}}$ is commonly named "hard factor". In perturbative QCD, it can be computed order by order in the strong coupling α_s and is equal to 1 at leading order. The results shown in this thesis are obtained following the definition of Ref. [128]. The expression at $O(\alpha_s)$ of $\mathcal{H}^{\mathrm{DY}}$ is given by

$$
\mathcal{H}^{\mathrm{DY}}(Q, \mu)\bigg|_{O(\alpha_s)} = 1 + \frac{\alpha_s}{4\pi} C_F \left(-16 + \frac{2\pi^2}{3} + 6 \ln\left(\frac{Q^2}{\mu^2}\right) - 2 \ln^2\left(\frac{Q^2}{\mu^2}\right) \right). \quad (2.5)
$$

This function encodes the virtual part of the hard scattering at the partonic level and depends on the hard scale Q and on the renormalization scale μ.

The summation over a in Eq. (2.4) runs over the active quarks and anti-quarks at the scale Q, and the coefficients $c_a(Q^2)$ are the related electroweak charges given by

$$
c_a(Q^2) = e_a^2 - 2 e_a V_a V_\ell\, \chi_1(Q^2) + (V_\ell^2 + A_\ell^2)(V_a^2 + A_a^2)\, \chi_2(Q^2) , \quad (2.6)
$$

with

$$
\chi_1(Q^2) = \frac{1}{4\sin^2\theta_W \cos^2\theta_W} \frac{Q^2(Q^2 - M_Z^2)}{(Q^2 - M_Z^2)^2 + M_Z^2 \Gamma_Z^2} , \quad (2.7)
$$

$$
\chi_2(Q^2) = \frac{1}{16\sin^4\theta_W \cos^4\theta_W} \frac{Q^4}{(Q^2 - M_Z^2)^2 + M_Z^2 \Gamma_Z^2} , \quad (2.8)
$$

where the constants e_a, V_a, and A_a represent the electric, vector, and axial charges of the flavor a, respectively. Moreover, V_ℓ and A_ℓ are the vector and axial charges of the lepton ℓ; $\sin\theta_W$ is the weak mixing angle; M_Z and Γ_Z are mass and width of the Z boson.

In Eq. (2.4), the unpolarized TMD PDFs of a given quark are denoted by f_1^a. In particular, the first line of Eq. (2.4) displays the convolution of two TMD PDFs f_1^a and $f_1^{\bar a}$ of the hadrons h_1 and h_2, respectively. It describes the annihilation of a quark (antiquark), with longitudinal momentum fraction $x_A = Qe^y/\sqrt{s}$ and transverse momentum $\mathbf{k}_{\perp A}$, with the corresponding antiquark (quark), with longitudinal momentum fraction $x_B = Qe^{-y}/\sqrt{s}$ and transverse momentum $\mathbf{k}_{\perp B}$. The presence of the delta-function $\delta^{(2)}(\mathbf{k}_{\perp A}+\mathbf{k}_{\perp B}-\mathbf{q}_T)$ guarantees the momentum conservation in the annihilation. It is useful to underline that TMD factorization requires the transverse momentum of the photon $\mathbf{q}_T^2$ to be much smaller than Q^2, while the intrinsic transverse momenta of partons $\mathbf{k}_{\perp A}$ and $\mathbf{k}_{\perp B}$ can be of any size allowed by momentum conservation.

Since TMD distributions arise from a double-factorization procedure [73, 129], they depend on two factorization scales (for more details, see Sec. 2.3.1). One scale is the standard renormalization scale μ of ultraviolet (UV) singularities and another is the renormalization scale ζ connected to rapidity divergences. These two scales are equally important for the computation of the cross section, and they should be treated separately. The scale μ can assume an arbitrary value, because it is introduced by the theoretical formalism and it cannot be measured. Throughout this thesis, we set it equal to Q (the energy scale of the process). More complex is the situation for the rapidity scales of the two TMDs, which must obey the relation $\zeta_A\zeta_B = Q^4$ [73]. For convenience, we set $\zeta_A = \zeta_B = \mu^2 = Q^2$.

In order to avoid the convolution of transverse momenta, it is convenient to rewrite Eq. (2.4) in the so-called $|\mathbf{b}_T|$-space. It is the conjugate space to transverse momenta $\mathbf{k}_\perp$. In order to write the formalism in $|\mathbf{b}_T|$-space, we need to introduce the Fourier transform of each TMD. In this thesis, we consider the following definition of the Fourier transform of the TMD PDFs:[2]

$$\hat{f}_1^a\big(x,\mathbf{b}_T^2;\mu,\zeta\big) = \int d^2\mathbf{k}_\perp\, e^{i\mathbf{b}_T\cdot\mathbf{k}_\perp}\, f_1^a\big(x,\mathbf{k}_\perp^2;\mu,\zeta\big)$$
$$= 2\pi \int_0^\infty d|\mathbf{k}_\perp|\,|\mathbf{k}_\perp|J_0(|\mathbf{b}_T||\mathbf{k}_\perp|)\, f_1^a\big(x,\mathbf{k}_\perp^2;\mu,\zeta\big). \tag{2.9}$$

In Eq. (2.9), the function J_0 is the 0-th order Bessel function of the first kind and is given by the following integral expression

$$J_0(x) = \frac{1}{2\pi}\int_0^{2\pi} d\theta e^{ix\cos\theta}\,. \tag{2.10}$$

---

[2]The $|\mathbf{b}_T|$-dependent function $\hat{f}_1^a$ is the Fourier transform of the actual TMD f_1^a. However, for semplicity, henceforth we will refer to $\hat{f}_1^a$ as to TMD.

By inserting Eq. (2.9) into the cross section in Eq. (2.4), we finally get the expression of the F_{UU}^1 structure function in terms of the ingredients in $|\boldsymbol{b}_T|$-space

$$
\begin{aligned}
&F_{UU}^1\big(x_A, x_B, |\boldsymbol{q}_T|, Q\big) \\
&= \frac{x_A x_B}{2\pi}\, \mathcal{H}^{\mathrm{DY}}(Q, \mu) \sum_a c_a(Q^2) \\
&\int_0^{+\infty} d|\boldsymbol{b}_T|\,|\boldsymbol{b}_T|\,J_0\big(|\boldsymbol{b}_T||\boldsymbol{q}_T|\big)\hat{f}_1^a(x_A, \boldsymbol{b}_T^2; \mu, \zeta_A)\hat{f}_1^{\bar{a}}(x_B, \boldsymbol{b}_T^2; \mu, \zeta_B).
\end{aligned}
\tag{2.11}
$$

This is the formula for the F_{UU}^1 structure function actually implemented in the `NangaParbat` computational tool for the analyses reported in this thesis. In this formula, some ingredients can be calculated fully perturbatively, while others require the introduction of nonperturbative contributions. The former will be discussed in Sec. 2.3.2, the latter in Sec. 2.3.4.

We note that in Ref. [87] the variable $|\boldsymbol{b}_T|$ was replaced by ξ_T. Such a choice was made to avoid confusion with other definitions. In fact, in the literature of Generalized Parton Distributions (GPDs) the symbol $\boldsymbol{b}_T$ is typically introduced for the impact parameter. In this thesis, we decided to use $|\boldsymbol{b}_T|$ instead of ξ_T as it is more common in TMD, q_T-resummation and SCET (Soft Collinear Effective Field Theory) literature, by keeping in mind that this is the Fourier conjugate variable of $\boldsymbol{q}_T$, and not the impact parameter. Finally, we stress that the expression of the structure function in $|\boldsymbol{b}_T|$-space could be different depending on the definition of the Fourier transform.[3] However, its expression in $\boldsymbol{k}_\perp$-space must be always the same as in Eq. (2.4).

As will be discussed in Sec. 3.1 and summarized in Tab. 3.1, the observable generally provided by the experimental collaborations for DY process is the differential cross section (normalized to the total) with respect to the transverse momentum $|\boldsymbol{q}_T|$. By considering each bin delimited by the initial (i) and final (f) values of kinematical variables, the experimental values for the above mentioned observables are compared with the following theoretical quantity:

$$
\mathcal{O}_{\mathrm{DY}}^{th}(|\boldsymbol{q}_T|_{i,f}, y_{i,f}, Q_{i,f}) = \fint_{|\boldsymbol{q}_T|_i}^{|\boldsymbol{q}_T|_f} d|\boldsymbol{q}_T| \int_{y_i}^{y_f} dy \int_{Q_i}^{Q_f} dQ\, \frac{d\sigma^{\mathrm{DY/Z}}}{d|\boldsymbol{q}_T|\, dy\, dQ}\,, \tag{2.12}
$$

where the $\fint$ symbol represents the integral divided by the width of the kinematic bin where the integration takes place:

$$
\fint_{|\boldsymbol{q}_T|_i}^{|\boldsymbol{q}_T|_f} d|\boldsymbol{q}_T| = \frac{1}{|\boldsymbol{q}_T|_f - |\boldsymbol{q}_T|_i} \int_{|\boldsymbol{q}_T|_i}^{|\boldsymbol{q}_T|_f} d|\boldsymbol{q}_T|\,. \tag{2.13}
$$

Consequently, Eq. (2.12) corresponds to the cross section in Eq. (2.3) integrated over rapidity and invariant mass of the exchanged boson and aver-

---

[3]Notice that in Ref. [87] the Fourier transform was defined with an extra $1/(2\pi)$ factor.

aged over its transverse momentum. When the experimental collaboration released only the normalized differential cross section, the final observable is obtained by dividing both sides of Eq. (2.12) by the appropriate fiducial cross section, namely the fully integrated cross section, which is computed through the `DYNNLO` code [130, 131].[4]

However, this is the case only when experimental data are measured in facilities where both the initial-state hadrons of the DY process are accelerated, namely only for data sets released by experimental collaborations working at high-energy colliders (Tevatron, RHIC and LHC). In the case of low-energy fixed-target experiments, where only one of the two hadrons is accelerated, the measured cross section is usually written as

$$E \frac{d\sigma^{\mathrm{DY}}}{d^3\boldsymbol{q}} = \frac{1}{2\pi \, |\boldsymbol{q}_T|} \frac{d\sigma^{\mathrm{DY}}}{d|\boldsymbol{q}_T| \, dy},$$ (2.14)

where E and $\boldsymbol{q}$ are the energy and the three-momentum of the photon , respectively. From Eq. (2.14), in principle the experimental value in a given kinematic bin needs to be compared to the following theoretical quantity:

$$\mathcal{O}_{\mathrm{DY}}^{th}(|\boldsymbol{q}_T|_{i,f}, y_{i,f}, Q_{i,f}) = \oint_{|\boldsymbol{q}_T|_i}^{|\boldsymbol{q}_T|_f} d|\boldsymbol{q}_T| \oint_{y_i}^{y_f} dy \int_{Q_i}^{Q_f} dQ \, \frac{1}{2\pi \, |\boldsymbol{q}_T|} \frac{d\sigma^{\mathrm{DY}}}{d|\boldsymbol{q}_T| \, dy \, dQ} \, .$$ (2.15)

However, since all the fixed-target experiments considered in the phenomenological analyses do not provide bins of $|\boldsymbol{q}_T|$ but just the average transverse momentum values $|\overline{\boldsymbol{q}_T}|$, the integration over $|\boldsymbol{q}_T|$ is not considered. Moreover, some experiments provide only the average value $\overline{y}$ for the rapidity. Accordingly, the theoretical quantity considered for those experiments reads

$$\mathcal{O}_{\mathrm{DY}}^{th}(|\overline{\boldsymbol{q}_T}|, \overline{y}, Q_{i,f}) = \frac{1}{2\pi \, |\overline{\boldsymbol{q}_T}|} \int_{Q_i}^{Q_f} dQ \, \frac{d\sigma^{\mathrm{DY}}}{d|\boldsymbol{q}_T| \, dy \, dQ} \bigg|_{y=\overline{y}, \, |\boldsymbol{q}_T|=|\overline{\boldsymbol{q}_T}|} \, .$$ (2.16)

Some low-energy fixed-target experiments (see Tab. 3.1) use the variable "Feynman x" x_F in place of the rapidity y. For the variable x_F the following relations with other kinematic variables hold:

$$y(x_F, Q) = \sinh^{-1}\left(\frac{\sqrt{s}}{Q} \frac{x_F}{2}\right), \quad x_A = \sqrt{\frac{Q^2}{s} + \frac{x_F^2}{4}} + \frac{x_F}{2}, \quad x_B = x_A - x_F \, .$$ (2.17)

---

[4]See https://www.physik.uzh.ch/en/groups/grazzini/research/Tools.html

Using Eq. (2.17), one obtains

$$E \frac{d\sigma^{\text{DY}}}{d^3\boldsymbol{q}} = \frac{2E}{\pi\sqrt{s}} \frac{d\sigma^{\text{DY}}}{d\boldsymbol{q}_T^2\, dx_F}\,. \tag{2.18}$$

When experimental collaborations provide bins in x_F and average transverse momentum values $|\overline{\boldsymbol{q}_T}|$, the experimental values are compared to the following theoretical quantity:

$$\mathcal{O}_{\text{DY}}^{th}(|\overline{\boldsymbol{q}_T}|, x_{Fi,f}, Q_{i,f}) \tag{2.19}$$

$$= \int_{Q_i}^{Q_f} dQ \fint_{x_{Fi}}^{x_{Ff}} dx_F \frac{2E}{\pi\sqrt{s}} \frac{d\sigma^{\text{DY}}}{d\boldsymbol{q}_T^2\, dx_F\, dQ}\bigg|_{|\boldsymbol{q}_T|=|\overline{\boldsymbol{q}_T}|} \tag{2.20}$$

$$\approx \frac{\overline{Q}\cosh(\overline{y})}{\pi\,|\overline{\boldsymbol{q}_T}|\,\sqrt{s}\,(x_{Ff}-x_{Fi})} \int_{y(x_{Fi},\overline{Q})}^{y(x_{Ff},\overline{Q})} dy \int_{Q_i}^{Q_f} dQ \frac{d\sigma^{\text{DY}}}{d|\boldsymbol{q}_T|\, dy\, dQ}\bigg|_{|\boldsymbol{q}_T|=|\overline{\boldsymbol{q}_T}|}\,,$$

where

$$\overline{Q} = (Q_i + Q_f)/2\,, \qquad \overline{y} = [y(x_{Fi},\overline{Q}) + y(x_{Ff},\overline{Q})]/2\,. \tag{2.21}$$

We stress that Eq. (2.19) is an approximation of the actual experimental observable, since we replaced y and Q with $\overline{y}$ and $\overline{Q}$ in the prefactor in front of the cross section and pull it out of the integral. It is a convenient choice dictated by the structure of `NangaParbat`, the computational tool for TMD studies developed by the MAP Collaboration [1].

This last case happens when experimental collaborations provides average values for both transverse momentum and x_F. In our work, such data are compared against the following theoretical quantity:

$$\mathcal{O}_{\text{DY}}^{th}(|\overline{\boldsymbol{q}_T}|, \overline{x_F}, Q_{i,f}) \approx \frac{\overline{Q}\cosh(\overline{y})}{\pi\,|\overline{\boldsymbol{q}_T}|\,\sqrt{s}} \int_{Q_i}^{Q_f} dQ \frac{d\sigma^{\text{DY}}}{d|\boldsymbol{q}_T|\, dy\, dQ}\bigg|_{|\boldsymbol{q}_T|=|\overline{\boldsymbol{q}_T}|,\,y=\overline{y}}\,, \tag{2.22}$$

where, in this case, $\overline{y} = y(\overline{x_F},\overline{Q})$ with $\overline{x_F} = (x_{Fi} + x_{Ff})/2$.

2.2 SIDIS observables

In semi-inclusive DIS, a lepton ℓ with momentum l scatters off a hadron target N with four momentum P mass M. In the final state, a single hadron h (a pion or a kaon, generally) with four-momentum P_h is measured, while all other hadrons composing the final state are not detected. The mass of the lepton can be neglected, since it is only a small fraction of the hadron's one. The process can be represented as follows:

$$\ell(l) + N(P) \to \ell(l') + h(P_h) + X\,. \tag{2.23}$$

Following the standard notation, X contains all the products of the process that are not detected.

Generally, the available experimental data used for studying the internal structure of nucleons come from scattering on protons and light nuclei, such as deuterium and helium. In the analyses performed by the MAP collaboration, the scattering on a nucleus is approximated as a scattering on an isoscalar combination of protons and neutrons. In fact, since the knowledge of nuclear collinear parton densities is not as refined as for the nucleon (for a review, see Ref. [132]), there are only few recent exploratory attempts to introduce nuclear modifications to TMD distributions [133, 134].

For convenience, we define the (space-like) four-momentum transfer, namely the four-momentum of the virtual exchanged photon, as $q = l - l'$, and the invariant momentum transfer as $Q^2 \equiv -q^2 > 0$, and the following standard SIDIS variables [55, 135]:

$$x = \frac{Q^2}{2\,P\cdot q}\,, \qquad y = \frac{P\cdot q}{P\cdot l}\,, \qquad z = \frac{P\cdot P_h}{P\cdot q}\,, \qquad \gamma = \frac{2Mx}{Q}\,. \qquad (2.24)$$

In the above equation, x is the Bjorken scaling variable, which defines the fraction of the initial hadron momentum carried by the struck parton, y is the energy fraction that is transferred by the lepton to the hadron. Therefore, both x and y lie between 0 and 1, and provide information about the degree of inelasticity of the process (being 1 the value for elastic scattering process). In Eq. (2.24), the variable z is the fraction of the fragmenting parton momentum carried by the final-state hadron.

As shown in Fig. 2.2, we will follow the definitions and notations for transverse momenta discussed in Ref. [136, 137]. Before going into the details of a specific frame of reference, we define the transverse part P_{hT} of the final-state hadron momentum P_h as orthogonal with respect to the momenta P and q. Moreover, we define the transverse part q_T of the photon momentum q as orthogonal with respect to the momenta P and P_h.

We can make several choices for the frame of reference. If we choose the light-cone basis formed by the vectors P and q, we find a class of frames of reference where $\boldsymbol{P}_{hT}$ is defined as the hadron transverse momentum with respect to the photon's direction of motion. As a consequence, the four-momentum P_{hT} has only transverse components and the frame-independent relation $|\boldsymbol{P}_{hT}|^2 = -P_{hT}^2$ holds. Instead, if we choose light-cone basis formed by P and P_h, we obtain a class of frames of reference where $\boldsymbol{q}_T$ is defined as the photon transverse momentum with respect to the hadron's direction of motion. As a consequence, the four-momentum q_T has only transverse components and the frame-independent relation $|\boldsymbol{q}_T|^2 = -q_T^2$ holds. The two momenta are related by a Lorentz boost [138, 139]

$$q_T^\mu = -\frac{P_{hT}^\mu}{z} - 2x\frac{|\boldsymbol{q}_T|^2}{Q^2}P^\mu\,. \qquad (2.25)$$
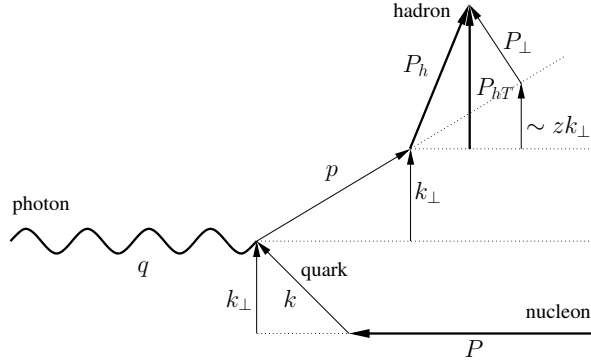
Figure 2.2: Diagram describing the relevant momenta involved in a SIDIS event in the Breit (nucleon-photon) frame. A virtual photon with momentum q (defining the reference axis) strikes a parton with momentum k inside a nucleon with momentum P. The parton has a transverse momentum $\boldsymbol{k}_\perp$ (not measured). The struck parton with momentum $p = k + q$ fragments into a hadron with momentum P_h, which acquires a further transverse momentum $\boldsymbol{P}_\perp$ (not measured) with respect to the fragmenting quark axis. The total measured transverse-momentum of the final hadron is $\boldsymbol{P}_{hT}$. When Q^2 is very large, the longitudinal components are all much larger than the transverse components. In this regime, $\boldsymbol{P}_{hT} \approx z\boldsymbol{k}_\perp + \boldsymbol{P}_\perp$.

In the kinematic region where the invariant mass of the photon is much larger then the target and hadron masses ($M^2, M_h^2 \ll Q^2$) and $\boldsymbol{q}_T^2, \boldsymbol{P}_{hT}^2 \ll Q^2$, Eq. (2.25) reduces to

$$\boldsymbol{q}_T \approx -\frac{\boldsymbol{P}_{hT}}{z} \,. \tag{2.26}$$

We stress that Eq. (2.26) is valid up to power corrections of $O(|\boldsymbol{q}_T|/Q)$. Power corrections can give rise to the modification of some kinematic variables (kinematic power corrections), like the term proportional to $\boldsymbol{q}_T^2/Q^2$ in Eq. (2.25), or they can introduce the contribution of higher-twist parton densities (dynamical power corrections). In this thesis, we will neglect any power corrections that should vanish in the considered region, apart from some modifications to the normalization of the SIDIS observables (that could be seen as the effect of power corrections, see Sec. 2.4 for more details).

The unpolarized differential cross section for the SIDIS process can be written in terms of two structure functions, $F_{UU,T}$ and $F_{UU,L}$ [55] as follows[5]

$$\frac{d\sigma^{\text{SIDIS}}}{dx dz d|\boldsymbol{q}_T| dQ} = \frac{8\pi^2 \alpha^2 z^2 |\boldsymbol{q}_T|}{x\,Q^3} \frac{y^2}{1 - \epsilon} \left[F_{UU,T}\big(x, z, |\boldsymbol{q}_T|, Q\big) + \epsilon F_{UU,L}\big(x, z, |\boldsymbol{q}_T|, Q\big) \right], \tag{2.27}$$

where α is the QED coupling constant, y is defined in Eq. (2.24), and ϵ is the

---

[5]We note that in Ref. [1] there is a typographical error of a factor 2 in this formula.

ratio of longitudinal and transverse photon flux

$$\epsilon = \frac{1 - y - \frac{1}{4}\gamma^2}{1 - y + \frac{1}{2}y^2 + \frac{1}{4}\gamma^2 y^2}, \qquad (2.28)$$

where $\gamma = \dfrac{2Mx}{Q}$. As for the notations, the subscripts of the structure functions $F_{XY,Z}$ refer to the lepton (X), the target (Y), and the intermediate photon (Z) polarizations. The structure function $F_{UU,L}$ is formally a twist-4 ingredient [138], and is related to the photon longitudinal polarization as same as the structure function F_{UU}^2 in the DY process. In this case, the information on $F_{UU,L}$ is really poor. While moderately accurate measurements of this contribution have been made in inclusive DIS, there are currently no measurements of that quantity in SIDIS. In the limit considered here, it is expected to be suppressed [138]. However, there very recent investigations indicating that the $F_{UU,L}$ structure function might not vanish at small $|\boldsymbol{q}_T|$, and its size might be even up to the 30% of the $F_{UU,T}$ [106]. In the analyses presented in this thesis, we neglect it, and we leave this topic for a future study.

Neglecting target mass corrections $\mathcal{O}(\gamma)$ and $\mathcal{O}(\gamma^2)$, which are a source of kinematic power corrections, the y-dependent prefactor in Eq. (2.27) can be approximated as

$$\frac{y^2}{(1 - \epsilon)} \approx Y_+ = 1 + \left(1 - \frac{Q^2}{xs}\right)^2. \qquad (2.29)$$

As for Drell-Yan, the SIDIS structure functions can be written in terms of TMD PDFs and TMD Fragmentation Functions FFs by relying on factorization theorems [73, 140]. Therefore, at low transverse momenta $|\boldsymbol{q}_T| = |\boldsymbol{P}_{hT}|/z \ll Q$, the structure function $F_{UU,T}$ can be defined as follows

$$F_{UU,T}\big(x, z, |\boldsymbol{q}_T|, Q\big) = x\,\mathcal{H}^{SIDIS}(Q, \mu) \sum_{a=q,\bar{q}} e_a^2$$

$$\times \int d^2\boldsymbol{k}_\perp \int \frac{d^2\boldsymbol{P}_\perp}{z^2} f_1^a(x, \boldsymbol{k}_\perp^2; \mu, \zeta_A)\, D_1^{a \to h}(z, \boldsymbol{P}_\perp^2; \mu, \zeta_B)\delta^{(2)}\Big(\boldsymbol{k}_\perp + \frac{\boldsymbol{P}_\perp}{z} + \boldsymbol{q}_T\Big)$$

$$+ Y_{UU,T}(Q^2, \boldsymbol{P}_{hT}^2) + O\left(\frac{M^2}{Q^2}\right).$$

$$(2.30)$$

Here, the sum runs over quarks and antiquarks a. The variable $\boldsymbol{k}_\perp$ is the transverse momentum of the struck quark with respect to the target-nucleon axis, whereas $\boldsymbol{P}_\perp$ is the transverse momentum of the produced hadron h with respect to the fragmenting quark axis (see Fig. 2.2). As for the DY case, we stress that TMD factorization requires the transverse momentum of the photon $\boldsymbol{q}_T^2$ to be much smaller than Q^2, while the intrinsic transverse momenta of the initial-state parton $\boldsymbol{k}_\perp$ and the final-state hadron $\boldsymbol{P}_\perp$ can be of any size allowed by momentum conservation (imposed by the delta-function). The hard factor

$\mathcal{H}^{\mathrm{SIDIS}}$ describes the scattering at the partonic level. It can be computed order by order in the strong coupling α_s, and is equal to 1 at leading order. In the present work, we follow the definition of Ref. [128]. The expression at $O(\alpha_s)$ of $\mathcal{H}^{\mathrm{SIDIS}}$ is given by

$$\mathcal{H}^{\mathrm{SIDIS}}(Q,\mu)\Big|_{O(\alpha_s)} = 1 + \frac{\alpha_s}{4\pi}C_F\left(-16+\frac{\pi^2}{3}+6\ln\left(\frac{Q^2}{\mu^2}\right)-2\ln^2\left(\frac{Q^2}{\mu^2}\right)\right). \quad (2.31)$$

In Eq. (2.30), the function $D_1^{a\to h}(z,\boldsymbol{P}_\perp^2;\mu,\zeta_B)$ is the TMD FF for an unpolarized parton of flavor a that fragments in an unpolarized (measured) hadron h. As already mentioned in Sec. 2.1, since these distributions arise from a double-factorization procedure [73, 129], both TMD PDFs and TMD FFs depend on two factorization scales.

The function $Y_{UU,T}(Q^2,\boldsymbol{P}_{hT}^2)$ is the so-called Y-term. It was introduced for the first time in Ref. [72]. Such term ensures a matching to the perturbative calculations at large values of transverse momentum. In fact, the cross section expression in Eq. (2.27) is valid in any region of transverse momentum. In the region at low transverse momenta $|\boldsymbol{q}_T| \ll Q$, it is determined by Eq. 2.30 (W-term, in literature) and by the Y-term in the complementary region $|\boldsymbol{q}_T| \gtrsim Q$. Indeed, in this last region, the TMD factorization breaks down and the collinear factorization becomes the appropriate framework to describe the cross section [138].

In Eq. (2.27), the electroweak couplings are given only by the square of the electric charges e_a^2 because we limit to the energy region $Q^2 \ll M_Z^2$, where the contribution of the Z-boson exchange and the γZ-interference terms can be neglected. This approximation is reasonable for the experimental observables included in TMD global analyses, which lay at low values of Q (smaller than 10 GeV).

As for the Drell-Yan case, it is convenient to rewrite Eq. (2.30) in $|\boldsymbol{b}_T|$-space in order to avoid the convolution of transverse momenta

$$F_{UU,T}\big(x,z,|\boldsymbol{q}_T|,Q\big) = \frac{x}{2\pi}\,\mathcal{H}^{SIDIS}(Q,\mu)\sum_{a=q,\bar{q}}e_a^2$$
$$\int_0^{+\infty} d|\boldsymbol{b}_T||\boldsymbol{b}_T|J_0\big(|\boldsymbol{b}_T||\boldsymbol{q}_T|\big)\hat{f}_1^a(x,\boldsymbol{b}_T^2;\mu,\zeta_A)\,\hat{D}_1^{a\to h}(z,\boldsymbol{b}_T^2;\mu,\zeta_B)\,, \quad (2.32)$$

where the Fourier transform of the TMD PDF is defined in Eq. (2.9), while for the TMD FF is defined as

$$\hat{D}_1^{a\to h}\big(z,\boldsymbol{b}_T^2;\mu,\zeta\big) = \int \frac{d^2\boldsymbol{P}_\perp}{z^2}\,e^{-i\boldsymbol{b}_T\cdot\boldsymbol{P}_\perp/z}\,D_1^a\big(z,\boldsymbol{P}_\perp^2;\mu,\zeta\big)$$
$$= 2\pi\int_0^\infty \frac{d|\boldsymbol{P}_\perp|}{z^2}\,|\boldsymbol{P}_\perp|J_0(|\boldsymbol{b}_T||\boldsymbol{P}_\perp|/z)\,D_1^a\big(z,\boldsymbol{P}_\perp^2;\mu,\zeta\big)\,. \quad (2.33)$$

The structure of TMD PDFs and FFs will be discussed in detail in Sec. 2.3.

As already mentioned for DY in Sec. 2.1, some ingredients of Eq. (2.32) can be calculated fully perturbatively, while other ones require the introduction of nonperturbative techniques. The former ones will be discussed in Sec. 2.3.2, the latter ones in Sec. 2.3.4.

The available experimental data for the SIDIS process were released by the COMPASS [141] and HERMES [142] Collaborations. The provided observable is the *hadron multiplicity*, namely the differential number of hadrons produced per corresponding inclusive DIS event. In terms of cross sections, the multiplicities can be defined as

$$
M(x, z, |\boldsymbol{P}_{hT}|, Q) \;=\; \frac{d\sigma^{\text{SIDIS}}}{dx\, dz\, d|\boldsymbol{P}_{hT}|\, dQ} \bigg/ \frac{d\sigma^{\text{DIS}}}{dx\, dQ} \,. \tag{2.34}
$$

Here, the $d\sigma^{\text{SIDIS}}$ is the differential cross section for the SIDIS process, and the $d\sigma^{\text{DIS}}$ is the corresponding for inclusive DIS. The former is defined as in Eq. (2.27), the latter reads

$$
\begin{aligned}
\frac{d\sigma^{\text{DIS}}}{dx\, dQ} &= \frac{8\pi\,\alpha^2}{x\,Q^3} \frac{y^2}{2(1-\epsilon)} \left[F_T(x, Q^2) + \epsilon F_L(x, Q^2) \right] \\
&\approx \frac{4\pi\,\alpha^2}{x\,Q^3} \left[Y_+ F_2(x, Q^2) - y^2 F_L(x, Q^2) \right],
\end{aligned} \tag{2.35}
$$

where the approximation $F_2 \approx F_T + F_L$ is justified because we neglect target mass corrections. Following the standard DIS conventions, only the photon polarization is explicitly written (T, L) in the subscript of the structure functions. The structure function F_L in the DIS cross section is not an higher-twist correction such as the $F_{UU,L}$ structure function in the SIDIS process. However, it is different from zero starting only from next-to-leading order in α_s. Therefore, at the perturbative order considered in Ref. [1], the structure function F_L cannot be neglected, at variance with, *e.g.*, Refs. [84, 87].

In order to make the comparison against the experimental values in each kinematic bin, we calculate the multiplicity with the `NangaParbat` computational tool by *separately* averaging the numerator and denominator of the multiplicity in Eq. (2.34) over the respective kinematics. For the numerator the computed expression is

$$
\begin{aligned}
\mathcal{O}^{th}_{\text{SIDIS}}&(x_{i,f}, z_{i,f}, |\boldsymbol{P}_{hT}|_{i,f}, Q_{i,f}) \\
&= \fint_{Q_i}^{Q_f} dQ \fint_{x_i}^{x_f} dx \fint_{z_i}^{z_f} dz \fint_{|\boldsymbol{P}_{hT}|_i}^{|\boldsymbol{P}_{hT}|_f} d|\boldsymbol{P}_{hT}| \; \frac{d\sigma^{\text{SIDIS}}}{dx\, dz\, d|\boldsymbol{P}_{hT}|\, dQ} \,,
\end{aligned} \tag{2.36}
$$

while for the inclusive DIS the expression is the following

$$
\mathcal{O}^{th}_{\text{DIS}}(x_{i,f}, Q_{i,f}) = \fint_{Q_i}^{Q_f} dQ \fint_{x_i}^{x_f} dx \; \frac{d\sigma^{\text{DIS}}}{dx\, dQ} \,. \tag{2.37}
$$

Finally, the computed multiplicity is given by

$$\mathcal{M}^{th}(x_{i,f}, z_{i,f}, |\boldsymbol{P}_{hT}|_{i,f}, Q_{i,f}) = \frac{\mathcal{O}_{\text{SIDIS}}^{th}(x_{i,f}, z_{i,f}, |\boldsymbol{P}_{hT}|_{i,f}, Q_{i,f})}{\mathcal{O}_{\text{DIS}}^{th}(x_{i,f}, Q_{i,f})} \ . \tag{2.38}$$

The COMPASS collaboration provides multiplicities in bins of $\boldsymbol{P}_{hT}^2$, whereas the HERMES collaboration in bins of $|\boldsymbol{P}_{hT}|$ (see also Tab. 3.3). In both cases, the numerator of the observable can be calculated as in Eq. (2.36), but in the COMPASS case the average is on $\boldsymbol{P}_{hT}^2$. Moreover, a cut on the invariant mass of the hadronic final states $W^2 = (P + q)^2$, through the relation

$$M_N^2 = W^2 - \frac{1-x}{x}Q^2 \ , \tag{2.39}$$

is introduced by both collaborations (see Tab. 3.3), which makes the upper integration limit x_f in the integral over the kinematic variable x a Q-dependent quantity.

2.3 TMD distributions

In this section, we discuss all the ingredients needed to express a TMD distribution at the energy scale Q (see also Refs. [1, 73, 81]). We stress that the actual TMD is defined in $|\boldsymbol{k}_\perp|$-space, while its Fourier transform depends on the variable $\boldsymbol{b}_T^2$ (the conjugate variable of the transverse momentum). In this thesis, we discuss the full expression of the TMD distribution in $|\boldsymbol{b}_T|$-space. Then, one can recover the expression of the actual TMD distribution simply by taking its Fourier anti-transform.

2.3.1 TMD evolution and matching

In the following, we will discuss formalism for the evolution and the matching for TMD distributions, which is equivalent for both TMD PDFs and TMD FFs.

As already mentioned in Secs. 2.1-2.2, TMD distributions acquire a dependence on two scales, μ and ζ, due to the removal of the ultraviolet and rapidity divergences at the level of the operator definition [73, 129, 143].

In Quantum Field Theory, the dependence on the renormalization scale of a renormalized object is dictated by a Renormalization Group (RG) equation. Therefore, in the case of TMD distributions, we deal with two evolution equations

$$\frac{\partial \ln \hat{f}_1}{\partial \ln \mu} = \gamma(\mu, \zeta) \ , \qquad\qquad \frac{\partial \ln \hat{f}_1}{\partial \ln \sqrt{\zeta}} = K(|\boldsymbol{b}_T|, \mu) \ , \tag{2.40}$$

where γ is the anomalous dimension of the RG evolution in the scale μ, and

K is the so-called "Collins-Soper kernel", namely the anomalous dimension of the rapidity RG (RRG) evolution in $\sqrt{\zeta}$ [70]. Notice that, for brevity, we have dropped the flavour index a and $\bar{a}$, being the evolution equations equal for each flavor. Moreover, since in this section we will only discuss the dependence of $\hat{f}_1$ on the scales μ and ζ, we will also temporarily drop the dependence on x and $\boldsymbol{b}_T$.

A key feature of the Collins-Soper kernel K is that is not finite, and is renormalized by adding a counterterm that generates an additive anomalous dimension. Therefore, the rapidity anomalous dimension $K(|\boldsymbol{b}_T|, \mu)$ obeys its own RG equation

$$\frac{\partial K}{\partial \ln \mu} = -\gamma_K\big(\alpha_s(\mu)\big) \,, \tag{2.41}$$

where γ_K is the so-called cusp anomalous dimension. Since the cross derivatives of $\hat{f}_1$ must be equivalent, a RG equation arises also for the UV anomalous dimension

$$\frac{\partial \gamma}{\partial \ln \sqrt{\zeta}} = -\gamma_K\big(\alpha_s(\mu)\big) \,. \tag{2.42}$$

Fixing the boundary condition to the point $\zeta = \mu^2$, the solution of this differential equation is

$$\gamma(\mu, \zeta) = \gamma_F\big(\alpha_s(\mu)\big) - \gamma_K\big(\alpha_s(\mu)\big) \ln \frac{\sqrt{\zeta}}{\mu} \,, \tag{2.43}$$

where $\gamma_F(\alpha_s(\mu)) \equiv \gamma(\mu, \mu^2)$. By assuming that the TMD $\hat{f}_1$ is known at some starting scales μ_i and ζ_i, the solution of the evolution equations in Eq. (2.40) reads

$$\hat{f}_1(\mu_f, \zeta_f) = R\big[(\mu_f, \zeta_f) \leftarrow (\mu_i, \zeta_i)\big] \hat{f}_1(\mu_i, \zeta_i) \,, \tag{2.44}$$

where the so-called Sudakov form factor R provides the evolution of $\hat{f}_1$. It results

$$R\big[(\mu_f, \zeta_f) \leftarrow (\mu_i, \zeta_i)\big]$$
$$= \exp\left\{ K(\mu_i) \ln \frac{\sqrt{\zeta_f}}{\sqrt{\zeta_i}} + \int_{\mu_i}^{\mu_f} \frac{d\mu'}{\mu'} \gamma(\mu', \zeta_f) \right\} \,. \tag{2.45}$$

We note that Eq. (2.45) can be implemented in various ways [144–147]. In this work, we follow the standard approach described in [73], namely the Collins-Soper-Sterman (CSS) approach. Moreover, the calculation of all the ingredients involved in Eq. (2.45) is performed by adopting a fully numerical approach.

The expression of di Sudakov form factor in Eq. (2.45) can be obtained as follows. The application of the RG evolution equation on $\hat{f}_1(\mu_f, \zeta_f)$ gives:

$$\hat{f}_1(\mu_f, \zeta_f) = \exp\left[\int_{\mu_i}^{\mu_f} \frac{d\mu'}{\mu'} \gamma(\mu', \zeta_f) \right\} \hat{f}_1(\mu_i, \zeta_f) \,. \tag{2.46}$$

The factor $F(\mu_i, \zeta_f)$ can then rewritten using the RRG evolution equation as follows:

$$\hat{f}_1(\mu_f, \zeta_f) = \exp\left\{ K(\mu_i) \ln \frac{\sqrt{\zeta_f}}{\sqrt{\zeta_i}} + \int_{\mu_i}^{\mu_f} \frac{d\mu'}{\mu'} \gamma(\mu', \zeta_f) \right\} \hat{f}_1(\mu_i, \zeta_i), \qquad (2.47)$$

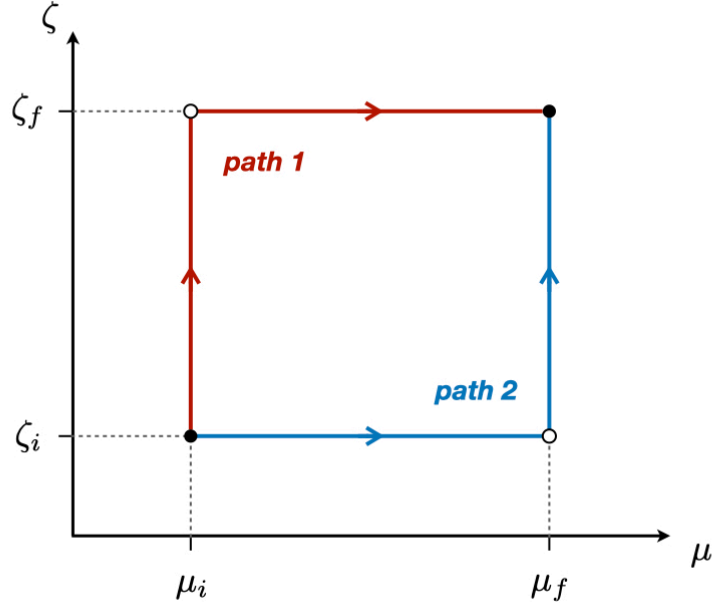which is the same result as in Eq. (2.45).

Figure 2.3: Two equivalent paths of RG and RRG evolution: red lines show the path 1 defined in Eq. (2.47); blue lines show the path 2 defined in Eq. (2.49).

Therefore, Eq. (2.45) describes the evolution of the TMD distribution at the initial scales $\hat{f}_1(\mu_i, \zeta_i)$ first in the ζ direction at fixed $\mu = \mu_i$ and then in the μ direction at fixed $\zeta = \zeta_f$ (path 1 in Fig. 2.3). However, it is easy to verify that exchanging the order of the two evolution operators (path 2 in Fig. 2.3) leads to the exact same result. Indeed, the application of the RRG evolution equation on $\hat{f}_1(\mu_f, \zeta_f)$ gives:

$$\hat{f}_1(\mu_f, \zeta_f) = \exp\left\{ K(\mu_f) \ln \frac{\sqrt{\zeta_f}}{\sqrt{\zeta_i}} \right\} \hat{f}_1(\mu_f, \zeta_i). \qquad (2.48)$$

In this case, the factor $F(\mu_f, \zeta_i)$ can then rewritten using the RG evolution equation as follows:

$$\hat{f}_1(\mu_f, \zeta_f) = \exp\left\{ K(\mu_f) \ln \frac{\sqrt{\zeta_f}}{\sqrt{\zeta_i}} + \int_{\mu_i}^{\mu_f} \frac{d\mu'}{\mu'} \gamma(\mu', \zeta_i) \right\} \hat{f}_1(\mu_i, \zeta_i), \qquad (2.49)$$

We can express the Collins–Soper kernel $K(\mu_f)$ by solving Eq. (2.41) as

$$K(\mu_f) = K(\mu_i) - \int_{\mu_i}^{\mu_f} \frac{d\mu'}{\mu'} \gamma_K(\alpha_s(\mu')) \,. \qquad (2.50)$$

Finally, by introducing Eq. (2.50) in Eq. (2.49) it is easy to see that Eq. (2.49) is exactly equivalent to Eq. (2.48). In this way, we demonstrated that the Sudakov form factor and, thus, TMD evolution do not depend on the chosen path to connect the point (μ_i, ζ_i) to the point (μ_f, ζ_f) in the μ-ζ plane. In particular, the following relation holds:

$$
\begin{aligned}
R\left[(\mu_f, \zeta_f) \leftarrow (\mu_i, \zeta_i)\right] &= R\left[(\mu_f, \zeta_f) \leftarrow (\mu_i, \zeta_f)\right] R\left[(\mu_i, \zeta_f) \leftarrow (\mu_i, \zeta_i)\right] \\
&= R\left[(\mu_f, \zeta_f) \leftarrow (\mu_f, \zeta_i)\right] R\left[(\mu_f, \zeta_i) \leftarrow (\mu_i, \zeta_i)\right] \,.
\end{aligned}
\qquad (2.51)
$$

This is true for any path connecting the point (μ_i, ζ_i) to the point (μ_f, ζ_f), and is a direct consequence of imposing that the cross derivatives of $\hat{f}_1$ are equivalent at each perturbative order.

An important property of TMDs is that at small values of $|\boldsymbol{b}_T|$ they can be matched onto their collinear counterparts ($f_1(x)$ for unpolarized TMD PDFs, $D_1(z)$ for unpolarized TMD FFs). Reinstating for clarity the x and $\boldsymbol{b}_T$ dependence, we can write the TMD PDF at the initial scales μ_i and ζ_i at $|\boldsymbol{b}_T| \ll 1$ as [73]

$$\hat{f}_1^a(x, \boldsymbol{b}_T^2; \mu_i, \zeta_i) = \sum_{b=q,\bar{q},g} \int_x^1 \frac{dy}{y} C_{ab}(y, \boldsymbol{b}_T^2; \mu_i, \zeta_i) f_1^b\left(\frac{x}{y}; \mu_i\right) \,, \qquad (2.52)$$

where $f_1^b\left(\frac{x}{y}; \mu_i\right)$ is the PDF for the parton b in collinear framework, and C_{ab} are the so-called matching coefficients. Such coefficient functions are fully perturbatively calculable and are currently known up to N^{3}LO, *i.e.* $O(\alpha_s^3)$. Then, Eq. (2.52) can be written in a more compact way as follows

$$\hat{f}_1(x, \boldsymbol{b}_T^2; \mu_i, \zeta_i) = \left[C \otimes f_1\right](x, \boldsymbol{b}_T^2; \mu_i, \zeta_i) \,, \qquad (2.53)$$

where the symbol $\otimes$ denotes the Mellin convolution, implying the sum over flavors (and including the gluon). The matching function C_{ab} for each pair of flavors a, b is included in the function C, which is a matrix in flavor-space multiplying a column vector of collinear PDFs. In this notation, the actual evolved TMD PDF becomes

$$\hat{f}_1(x, \boldsymbol{b}_T^2; \mu_f, \zeta_f) = R\left[\boldsymbol{b}_T^2; (\mu_f, \zeta_f) \leftarrow (\mu_i, \zeta_i)\right]\left[C \otimes f_1\right](x, \boldsymbol{b}_T^2; \mu_i, \zeta_i) \,. \qquad (2.54)$$

Both the matching and the evolution are affected by nonperturbative effects that become relevant in the large-$|\boldsymbol{b}_T|$ region. Usually, in order to take into account these effects, a phenomenological nonperturbative function f_{NP} is introduced. In the CSS approach [73], the TMD in $|\boldsymbol{b}_T|$-space gets a multiplicative

correction that is diagonal in the flavor space (it does not mix different flavors). In addition, the nonperturbative content of the TMD is smoothly damped away at small-$|\boldsymbol{b}_T|$ due to the introduction of the so-called b_*-prescription. All these topics are discussed in Sec. 2.3.4.

In order to calculate evolved TMDs at the energy scale of a given experimental measurement, we need to choose the values of both the initial and final pair of scales, (μ_i, ζ_i) and (μ_f, ζ_f) in Eq. (2.54). It can be shown that in the $\overline{\text{MS}}$ renormalization scheme there exist a particular scale $\mu_b(|\boldsymbol{b}_T|)$

$$\mu_b(|\boldsymbol{b}_T|) = \frac{2e^{-\gamma_E}}{|\boldsymbol{b}_T|} \,, \tag{2.55}$$

with γ_E the Euler constant, that prevents the appearance of large logarithms $\ln\left(\frac{\mu^2 \boldsymbol{b}_T^2}{4e^{-2\gamma_E}}\right)$ in the perturbative expansion of both the Collins-Soper kernel K and the matching coefficients C at the initial scales by choosing $\mu_i = \sqrt{\zeta_i} = \mu_b$.

The final value of the UV renormalization scale μ must be equal to the one inserted in the hard factor $\mathcal{H}$ (see Eqs. (2.11)-(2.32)). If we look at the expression of the hard factor (for both SIDIS and DY processes), it can be seen that we can avoid large logarithms $\ln(\mu/Q)$ by setting the final scale μ_f of the order of Q: we choose $\mu_f = Q$. The rapidity scales ζ_A and ζ_B are fixed by the relation $\zeta_A \zeta_B = Q^4$. Therefore, a standard choice is $\zeta_{A,f} = \zeta_{B,f} = \zeta_f = Q^2$. However, any choice that fulfills this relation is equivalent at the level of the evolution. In fact, in Eq. 2.45 it can be seen that the evolution factors R included in the definition of two TMDs combine such that the result depends on the product $\zeta_{A,f} \zeta_{B,f}$.

2.3.2 TMD perturbative content

In this section, we discuss the expression of the perturbative ingredients needed to build the structure functions F_{UU}^1 and $F_{UU,T}$ for DY and SIDIS processes, respectively.[6]

Firstly, we consider the hard factor $\mathcal{H}$. Up to order $O(\alpha_s^2)$, its perturbative expansion is given by

$$\mathcal{H}(Q, Q) = 1 + \sum_{n=1}^{2} \left(\frac{\alpha_s(Q)}{4\pi}\right)^n H^{(n)} \,. \tag{2.56}$$

The coefficients $H^{(n)}$ for DY process can be read off from, *e.g.*, Ref. [148]. If we consider contributions beyond two-loop accurancy, the hard factor acquires a non-trivial flavour structure (see, *e.g.*, Ref. [128]). As a consequence, $\mathcal{H}$ should in principle be moved inside the flavour sum in Eq. (2.11) (or Eq. (2.32), for SIDIS process). However, in this thesis we will only show phenomenological analyses that do not take into account corrections beyond $O(\alpha_s^2)$.

---

[6]See Eqs. (2.11)-(2.32)

Next, we consider the matching coefficients C introduced in Eq. (2.52) to express the TMD distribution in the small-$|\boldsymbol{b}_T|$ region. Their perturbative expansion at the initial scale is

$$C_{ab}(x, \boldsymbol{b}_T^2; \mu_b, \mu_b^2) = \delta_{ij}\delta(1-x) + \sum_{n=1}^{\infty} \left(\frac{\alpha_s(\mu_b)}{4\pi}\right)^n C_{ab}^{(n)}(x) , \qquad (2.57)$$

where the coefficient functions $C_{ab}^{(n)}(x)$ up to two-loop accuracy have been already calculated in Refs. [149, 150]. The calculation of the $O(\alpha_s^3)$ corrections to the quark matching functions for TMD PDFs can be found in Ref. [151].

The perturbative expansions of the anomalous dimensions K, γ_F, and γ_K in the Sudakov form factor in Eq. (2.45), read, respectively,

$$
\begin{aligned}
K(\mu_b) &= \sum_{n=0}^{\infty} \left(\frac{\alpha_s(\mu_b)}{4\pi}\right)^{n+1} K^{(n)} , \\[2mm]
\gamma_F(\alpha_s(\mu)) &= \sum_{n=0}^{\infty} \left(\frac{\alpha_s(\mu)}{4\pi}\right)^{n+1} \gamma_F^{(n)} , \\[2mm]
\gamma_K(\alpha_s(\mu)) &= \sum_{n=0}^{\infty} \left(\frac{\alpha_s(\mu)}{4\pi}\right)^{n+1} \gamma_K^{(n)} .
\end{aligned}
\qquad (2.58)
$$

The coefficients $K^{(n)}$ are listed up to $n=2$ in Ref. [128] and up to $n=3$ in Ref. [150]. Due to a different definition of the CS kernel K, they differ by a factor -2. Also the coefficients $\gamma_F^{(n)}$ are given in Refs. [128, 150] up to $n=2$, and they differ by a minus sign due to a different definition of the anomalous dimension. Finally, the coefficients $\gamma_K^{(n)}$ were originally computed in Ref. [152] and are also given in Refs. [128, 150] up to $n=2$, where they differ by a factor 2. The coefficient $\gamma_K^{(3)}$ has been recently computed in Refs. [153–155].

2.3.3 TMD logarithmic accuracy

In this section, we discuss how to consistently combine the perturbative ingredients in Eqs. (2.56)-(2.58) for the computation of the full expression of the structure functions in Eqs. (2.11)-(2.32) (see also Refs. [156, 157]).In particular, we use the same convention adopted in Ref. [81] for the definition of the logarithmic ordering.

As it is well known, TMD factorization provides resummation of large logarithms of $Q/|\boldsymbol{q}_T|$ or, equivalently, of Q/μ_b. The resummation is implemented in the Sudakov form factor R in Eq. (2.45), whose perturbative expansion reads

$$R = 1 + \sum_{n=1}^{\infty} \left(\frac{\alpha_s(Q)}{4\pi}\right)^n \sum_{k=1}^{2n} L^k R^{(n,k)} , \qquad (2.59)$$

with

$$L = \ln \frac{Q^2}{\mu_b^2} \ . \tag{2.60}$$

Since the inner sum runs up to $2n$, the double-logarithmic nature of the resummation can be clearly seen in Eq. (2.59). This structure can be traced back to the evolution equations in Eq. (2.40) that perform the resummation of two different categories of logarithms. However, our particular choice of the scales ($\mu_i = \sqrt{\zeta_i} = \mu_b$ and $\mu_f = \sqrt{\zeta_f} = Q$) makes the two categories coincide, leading to the introduction of up to two logarithms for each power of α_s. Hence, Eq. (2.59) must include all powers of α_s if the scales are such that $\alpha_s L^2 \gtrsim 1$.

In order to explicitly define the logarithmic ordering, we can rearrange the expansion in Eq. (2.59) as follows

$$R = 1 + \sum_{k=0}^{\infty} R_{\mathrm{N}^k\mathrm{LL}} \ , \tag{2.61}$$

with

$$R_{\mathrm{N}^k\mathrm{LL}} = \sum_{n=1+[k/2]}^{\infty} \left(\frac{\alpha_s(Q)}{4\pi} \right)^n L^{2n-k} R^{(n,2n-k)} \ , \tag{2.62}$$

where $[k/2]$ is defined as the integer part of $k/2$. According to this definition, the leading-logarithmic (LL) approximation is given by the term $k = 0$ in Eq.(2.61), the next-to-leading-logarithmic (NLL) approximation by the term $k = 1$, and so on. If we multiply the expression of $R_{\mathrm{N}^k\mathrm{LL}}$ by a general power p of α_s, we obtain

$$\left(\frac{\alpha_s(Q)}{4\pi} \right)^p R_{\mathrm{N}^k\mathrm{LL}}$$
$$= \sum_{m=1+[(k+2p)/2]}^{\infty} \left(\frac{\alpha_s(Q)}{4\pi} \right)^m L^{2m-(k+2p)} R^{(m-p,2m-(k+2p))} \leftrightarrow R_{\mathrm{N}^{k+2p}\mathrm{LL}} \ , \tag{2.63}$$

where the symbol $\leftrightarrow$ means that the left- and right-hand sides have the same logarithmic accuracy. This is a crucial point in the discussion. In fact, the Sudakov form factor in the cross section can be multiplied by some power of α_s deriving from the perturbative expression of hard factor $\mathcal{H}$ and/or the matching functions C. Consequently, equation (2.63) states that, at the cross section level, the inclusion of an additional power of α_s in the perturbative expansion of $\mathcal{H}$ and/or C leads to a contribution higher by two orders with respect to the leading term in the logarithmic expansion. For example, at LL and NLL accuracy the functions $\mathcal{H}$ and C have to be computed at $O(1)$, while at NNLL and N³LL they need to include the $\mathcal{O}(\alpha_s)$ corrections, and so on. This logarithmic counting is illustrated in the left panel of Fig. 2.4: the diagonal bands represent the terms included in each $R_{\mathrm{N}^k\mathrm{LL}}$, with $\mathcal{H}^{(n)}$ the perturbative
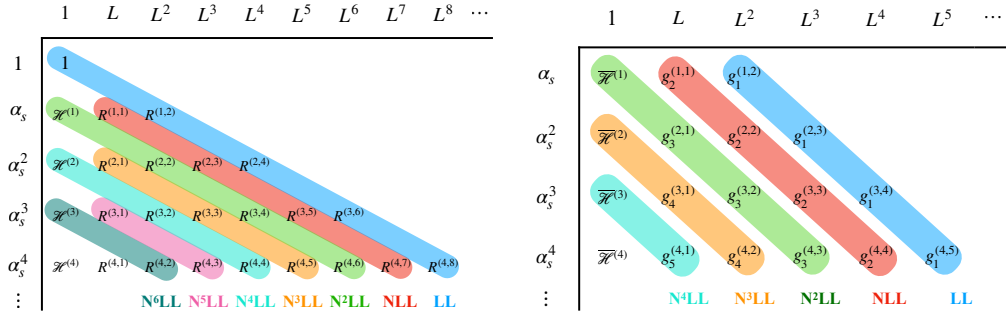
Figure 2.4: Graphical representation of logarithmic countings: in the left panel the counting is done at the level of the cross section, in the right panel at the level of the logarithm of the cross section.

coefficients of either $\mathcal{H}$ or C or a combination of the two.

Generally, the counting discussed above applies to any process whose amplitude factorises in the appropriate limit, such as in TMD factorization. However, if we consider the specific case of DY, also the phase space for the emission of n real particles in $|\boldsymbol{b}_T|$ space factorizes (see, *e.g.*, Ref. [158]). This two aspects allow one to *exponentiate* soft-collinear emissions such that the Sudakov form factor can be written in the following general form (see, *e.g.*, Ref. [159])[7]

$$R = \exp\left[\frac{1}{2}Lg^{(1)}(\alpha_s L) + \frac{1}{2}g^{(2)}(\alpha_s L) + \frac{1}{2}\alpha_s g^{(3)}(\alpha_s L) + \ldots\right], \qquad (2.64)$$

where the functions $g^{(i)}$ are defined with the boundary condition $g^{(i)}(0) = 0$. As compared to the general counting in Eq. (2.59), this exponentiation relates all the terms in Eq. (2.59) of the type $\alpha_s^n L^m$ with $n+1 < m \leq 2n$ to the lower-order terms. In Eq. (2.64), the logarithmic counting is performed at the level of the argument of the exponential. In this view, the terms $Lg^{(1)}$, $g^{(2)}$, $\alpha_s g^{(3)}$, etc., are responsible for the resummation of the LL contributions $\alpha_s^n L^{n+1}$, the NLL contributions $\alpha_s^n L^n$, the NNLL contributions $\alpha_s^n L^{n-1}$, etc. This counting is driven by the condition $\alpha_s L \gtrsim 1$, which extends the validity of the resummed result (truncated at a given level: NLL, NNLL, etc.) to larger values of the divergent L (for smaller values of q_T).

In Ref. [148], it has been shown that the logarithmic counting applied to the argument of the exponential is equivalent to consider the logarithm of the cross section. Indeed, neglecting for simplicity the contribution of the matching

---

[7]The factors $1/2$ in the argument of the exponential are justified by the fact that each of the two TMDs involved in the cross section of a given process (DY or SIDIS) contains an evolution factor R.

coefficients, for the DY cross section we have

$$\ln\left(\frac{d\sigma}{dQdydq_T}\right) \propto \ln\mathcal{H} + Lg^{(1)} + g^{(2)} + \alpha_s g^{(3)} + \ldots \qquad (2.65)$$

The logarithm of $\mathcal{H}$ can be expanded as

$$\ln(1 + \alpha_s H^{(1)} + \alpha_s^2 H^{(2)}) = \alpha_s H^{(1)} + \alpha_s^2\left(H^{(2)} - \frac{H^{(1)2}}{2}\right) + \mathcal{O}(\alpha_s^3)\,. \qquad (2.66)$$

In this context, the first term $\alpha_s H^{(1)}$ contributes to the tower $\alpha_s^n L^{n-1}$, namely the NNLL contribution. The second term $\alpha_s^2\left(H^{(2)} - H^{(1)2}/2\right)$ contributes to the $\alpha_s^n L^{n-2}$ tower, thus to the N^{3}LL contribution. The same counting applies to the matching coefficients C. Therefore, this means that the inclusion of $\mathcal{O}(\alpha_s)$ contributions in $\mathcal{H}$ and C implies the introduction of NNLL corrections, while $\mathcal{O}(\alpha_s^2)$ contributions in $\mathcal{H}$ and C contribute to N^{3}LL accuracy, and so on. In the right panel of Fig. 2.4 , a graphical representation of this counting is reported. As for the left panel, the bands represent the logarithmic towers, $\overline{\mathcal{H}}^{(n)}$ are the appropriate coefficients of the expansion of either $\ln H$ or $\ln C$ or a combination of the two, while the functions $g^{(i)}$ are replaced by the symbol $g_i^{(n,m)}$ (being n the order of α_s and m the power of the resummed L. This logarithmic counting has been used in several works (see, *e.g.*, Refs. [76, 148, 160, 161]). In this thesis, we will simply denote this counting with the acronyms NLL, NNLL, and so on, and we will refer to it as to "standard counting".

In literature, a slightly different convention has also been widely used for the logarithmic ordering (see, *e.g.*, Refs. [156, 162–165]). By expanding the Sudakov form factor (2.64) and multiplying it by the expansion of the hard function in Eq. (2.56), the result for the DY cross section is

$$\frac{d\sigma}{dQdydq_T} \propto 1 + Lg^{(1)} + g^{(2)} + H^{(1)}\alpha_s Lg^{(1)} + \ldots\,, \qquad (2.67)$$

where in the r.h.s. the combination of the first-order terms $\alpha_s H^{(1)}$ and $Lg^{(1)}$ in both expansions is reported. From the previous discussion, this term has the same form $\alpha_s^n L^n$ as $g^{(2)}$.

Then one can state that in order to reach NLL accuracy [162] we need to include not only the $g^{(2)}$ contribution but also the $H^{(1)}$ one. The same argument can be applied to all orders: at any given logarithmic accuracy, it prescribes to include one more order in the perturbative expansion of H (and/or C) with respect to the standard counting. We will refer to this counting as the "primed counting", and we will denote it as NLL′, NNLL′, and so on. This convention for logarithmic counting and for "standard counting" is the same up to subleading terms.

The apparent contradiction between the standard and primed countings is in fact due to subleading contributions. It can be resolved by observing

that the first term of the perturbative expansion of $\alpha_s L g^{(1)}$ is proportional to $\alpha_s^2 L^2$. When considering the general expansion of the cross section given in Eqs. (2.61)-(2.63), the term proportional to $\alpha_s^2 L^2$ is of the form $\alpha_s^n L^{2n-2}$ and, thus, it belongs to the NNLL tower. This is formally subleading with respect to the NLL accuracy determined by the $g^{(2)}$ term in the exponent.

In Tab. 2.1 we summarize the perturbative accuracy for each perturbative ingredient to be used for a consistent computation of the cross section in Eq. (2.3) for both the "standard" and the "primed" countings. The numbers in Tab. 2.1 give the power of α_s at which the corresponding quantity is to be computed, while the last column reports the corresponding accuracy in computing the evolution of the strong coupling constant α_s. In the logarithmic counting reported in Ref. [166], α_s is evolved at one loop less than the cusp anomalous dimensions because the running coupling RG equation resums single logs, and, therefore, the β function can be taken at the same order as the non-cusp anomalous dimension. For consistency, in the MAP Collaboration analyses we take α_s from the LHAPDF grid of the PDF set we use. In the

Accuracy	H and C	K and γ_F	γ_K	PDF and α_s evolution
LL	0	-	1	-
NLL	0	1	2	LO
NLL$'$	1	1	2	NLO
NNLL	1	2	3	NLO
NNLL$'$	2	2	3	NNLO
N^{3}LL	2	3	4	NNLO
N^{3}LL'	3	3	4	NNNLO
N^{4}LL	3	4	5	NNNLO

Table 2.1: Truncation order in the expansions of Eqs. (2.56)-(2.58) for the two logarithmic countings considered in this work (see text). The last column lists the order used for the collinear PDFs determination and the evolution of α_s.

literature, different conventions are used in some cases to identify the level of perturbative accuracy. In Refs. [82, 83, 88], the authors do not make use of the definitions we described in Tab. 2.1, also because they evaluate the double-scale evolution of TMDs in the so-called "ζ-prescription", which is a different approach from the standard CSS formalism.

2.3.4 Nonperturbative content of TMDs

In Sec. 2.3.1, we showed that μ_b can be identified as a natural choice for the initial scales μ_i and $\sqrt{\zeta_i}$. Therefore, the strong coupling $\alpha_s(\mu)$ in the perturbative expansion of K (see first line of Eq. (2.58)) and C (see Eq. (2.57)) must be computed at the initial scale μ_b. Consequently, the value of $\alpha_s(\mu_b)$ may strongly increase for large values of b_T, and eventually diverge when μ_b

reaches the Landau pole at Λ_{QCD}. In other words, the TMD formula defined in Sec. 2.3.1 is perturbatively meaningful only at low values of $|\boldsymbol{b}_T|$, such that the scale μ_b is sufficiently far from the Landau pole Λ_{QCD}. However, since the upper bound of the integral over $|\boldsymbol{b}_T|$ in Eqs. (2.11)-(2.32) is infinity, the TMD factorized expression of the cross section does require accessing large values of b_T. Regularizing this behavior is arbitrary (see, *e.g.*, Refs. [75, 167–169]). We report here some possible choices, starting from the "b_*-prescription" proposed by Collins in Ref. [71]. Such prescription aims to maximize the use of perturbation theory for small $|\boldsymbol{b}_T|$ and, at the same time, combine nonperturbative information. Therefore, an *arbitrary* parameter $b_{\max}$ is introduced as the maximum value of $|\boldsymbol{b}_T|$ at which perturbation theory is considered reliable. Hence, $b_{\max}$ must be such that

$$\alpha_s(\mu_b(b_{\max})) = \alpha_s\left(\frac{2e^{-\gamma_E}}{b_{\max}}\right) \ll 1 \ . \tag{2.68}$$

In this way, a minimum value for $\mu_b(|\boldsymbol{b}_T|)$ is defined through the introduction of the function $b_*(\boldsymbol{b}_T^2)$, which saturates at $b_{\max}$ in the large-$|\boldsymbol{b}_T|$ region. A common choice for the function $b_*(\boldsymbol{b}_T^2)$, originally proposed in Ref. [71], is

$$b_*^{\text{Collins}}(\boldsymbol{b}_T^2) = \sqrt{\frac{\boldsymbol{b}_T^2}{1 + \frac{\boldsymbol{b}_T^2}{b_{\max}^2}}} \ . \tag{2.69}$$

This effectively allows one to avoid the Landau pole in the evaluation of $\alpha_s(\mu_{b_*})$. In recent years, another expression has been proposed (see Ref. [170])

$$b_*^{\text{Bacchetta}}(\boldsymbol{b}_T^2) = b_{\max}[1 - \exp\left(-\boldsymbol{b}_T^4/b_{\max}^4\right)]^{\frac{1}{4}} \ , \tag{2.70}$$

which leads to a sharper transition between the perturbative and the non-perturbative regions. In Fig. 2.5, the comparison between the two functions is displayed. While the small-$|\boldsymbol{b}_T|$ limit of the two functions is exactly the same, they approach the maximum value $b_{\max}$ with two different rates. As a consequence of its functional form, the function $b_*^{\text{Collins}}(\boldsymbol{b}_T^2)$ approaches the nonperturbative region in a milder way than $b_*^{\text{Bacchetta}}(\boldsymbol{b}_T^2)$, and it deviates from the behavior of $|\boldsymbol{b}_T|$ earlier.

Moreover, the $b_*(\boldsymbol{b}_T^2)$ function can be chosen to modify not only the large distance region $|\boldsymbol{b}_T| \gtrsim 1/\Lambda_{QCD}$ (as already shown), but also the very short distance region $|\boldsymbol{b}_T| \ll 1/\Lambda_{QCD}$, as extensively discussed in Ref. [171]. Such further modification is not strictly mandatory, namely it is not introduced to preserve the theory from some inconsistency (especially when considering only small values of $|\boldsymbol{q}_T|$). However, it is shown to be necessary for low-Q phenomenological studies, where both the ratios $|\boldsymbol{q}_T|/Q$ and M/Q are not sufficiently small that TMD and collinear factorizations are simultaneously valid [171]. Moreover, such modification is useful to preserve the physical meaning of the integral in Eq. 2.45. In fact, if we choose $\mu_i = \mu_b$, than $\mu_i \to \infty$
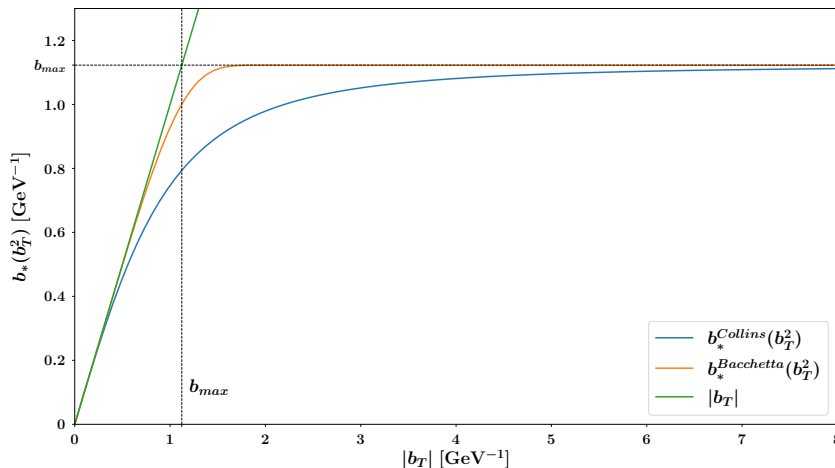
Figure 2.5: Comparison between the definition of the $b_*(\boldsymbol{b}_T^2)$ functions in Eqs. (2.69)-(2.70) by fixing $b_{\max} = 2e^{-\gamma_E}$ (with γ_E the Euler constant).

at small values of $|\boldsymbol{b}_T|$, and the lower bound of the integral becomes larger than the upper one ($\mu_i > \mu_f = Q$). The need of having a regular behavior at $|\boldsymbol{b}_T| \ll 1/\Lambda_{QCD}$ was already discussed in the literature, and the modified-log prescription was the first attempt to address this issue (see, *e.g.*, Ref. [159]). The idea of limiting the lower limit $\mu_i < \mu_f$ of the integral in the Sudakov form factor (see Eq. (2.45)) was discussed for the first time in Ref. [172], where the authors introduced a new scale μ_b' which does not become larger than Q as $|\boldsymbol{b}_T| \to 0$.

This new feature gives a physical interpretation of the cross section integrated in $|\boldsymbol{q}_T|$, being the lowest-order term equivalent to the lowest-order collinear result [171]. To this end, the cutoff value $b_{\min}$ is introduced in the "modified" b_*-prescription. We can satisfy all these requirements by replacing the variable $|\boldsymbol{b}_T|$ with a monotonic function $\bar{b}_*(\boldsymbol{b}_T^2)$ that must asymptotically behave as follows

$$
\begin{aligned}
\bar{b}_*(\boldsymbol{b}_T^2) &\to b_{\min} \quad \text{for} \quad |\boldsymbol{b}_T| \to 0 \,, \\
\bar{b}_*(\boldsymbol{b}_T^2) &\to b_{\max} \quad \text{for} \quad |\boldsymbol{b}_T| \to \infty \,.
\end{aligned}
\tag{2.71}
$$

According to the boundaries defined above, the proposed functional form in Ref [171] is given by

$$
\bar{b}_*^{\text{Collins}}(\boldsymbol{b}_T^2) = \sqrt{\frac{\boldsymbol{b}_T^2 + b_0^2/(C_5\mu)^2}{1 + \boldsymbol{b}_T^2/b_{\max}^2 + b_0^2/(C_5^2\mu^2 b_{\max}^2)}} \,,
\tag{2.72}
$$

where $b_0 = b_{\max}$, C_5 is of $\mathcal{O}(1)$ (we choose $C_5 = 1$), and μ is the energy scale. The functional form clearly takes inspiration from Eq. (2.69). By taking the
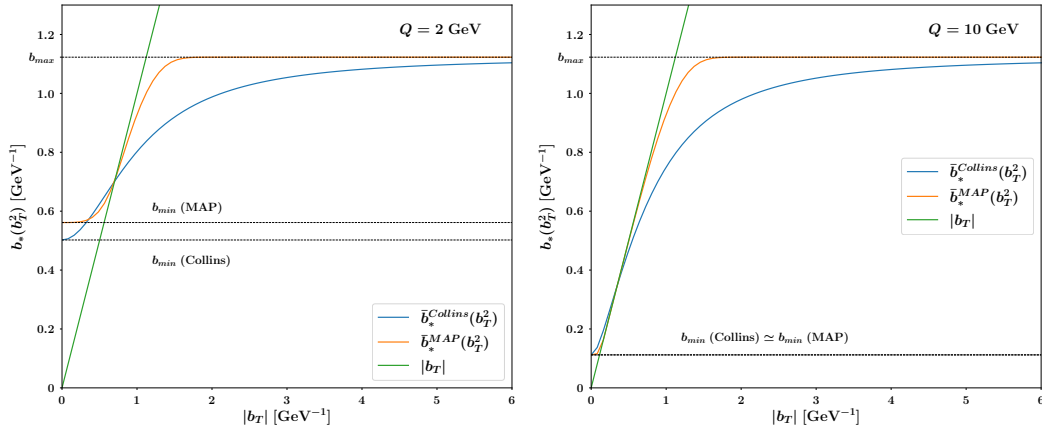
Figure 2.6: Graphical representation of two definition of the function $\bar{b}_*(\boldsymbol{b}_T^2)$ in the "modified" b_* prescription of Eqs. (2.72) and (2.74): in the left panel the functions are evaluated at $\mu = 2$ GeV, in the right panel at $\mu = 10$ GeV.

limit of this expression for $|\boldsymbol{b}_T| \to 0$ we obtain

$$\bar{b}_*^{\text{Collins}}(0) = \frac{b_{\max}}{\mu} \sqrt{\frac{1}{1 + 1/\mu^2}} \equiv b_{\min}^{\text{Collins}} . \tag{2.73}$$

In this thesis, we adopt for $\bar{b}_*(\boldsymbol{b}_T^2)$ the same functional form chosen by the MAP Collaboration (see Refs. [1, 2, 81, 87]) that guarantees a smoother and more rapid convergence towards the asymptotic limits:

$$\bar{b}_*^{\text{MAP}}(\boldsymbol{b}_T^2) = b_{\max} \left(\frac{1 - \exp\left(-\frac{\boldsymbol{b}_T^4}{b_{\max}^4}\right)}{1 - \exp\left(-\frac{\boldsymbol{b}_T^4}{b_{\min}^4}\right)} \right)^{\frac{1}{4}} , \tag{2.74}$$

with

$$b_{\max} = 2e^{-\gamma_E} \text{ GeV}^{-1} \approx 1.123 \text{ GeV}^{-1} , \qquad b_{\min}^{\text{MAP}} = 2e^{-\gamma_E}/\mu . \tag{2.75}$$

In Fig. 2.6, the comparison between the two functions $\bar{b}_*^{\text{Collins}}(\boldsymbol{b}_T^2)$ and $\bar{b}_*^{\text{MAP}}(\boldsymbol{b}_T^2)$ is displayed. As can be seen, the small-$|\boldsymbol{b}_T|$ behaviour is different at small values of the energy scale μ, and also the two cutoff values $b_{\min}^{\text{Collins}}$ in Eq. (2.73) and $b_{\min}^{\text{MAP}}$ in Eq. (2.75) are distinct (left panel). When the value of the energy scale μ is large enough, $b_{\min}^{\text{Collins}}$ can be approximated to $b_{\min}^{\text{MAP}}$ in Eq. (2.75) (right panel). In the phenomenological results reported in this thesis, we make use of the expression in Eq. (2.74) to account for the b_* prescription. The impact of the modification of the b_* prescription on a phenomenological analysis will be estimated in Sec. 3.3.2.

The saturation of the function $b_*(\boldsymbol{b}_T^2)$ to a cutoff value at large-$|\boldsymbol{b}_T|$ has also the effect of introducing power corrections which scale like $(\Lambda_{\text{QCD}}/|\boldsymbol{q}_T|)^k$ [167],

with $k > 0$. Such corrections need to be accounted for by a nonperturbative function in the region $|\boldsymbol{q}_T| \simeq \Lambda_{\text{QCD}}$. The nonperturbative function can be effectively defined by rewriting the TMD $\hat{f}_1$ as

$$
\begin{aligned}
\hat{f}_1(x, \boldsymbol{b}_T^2; \mu, \zeta) &= \left[\frac{\hat{f}_1(x, \boldsymbol{b}_T^2; \mu, \zeta)}{\hat{f}_1(x, b_*(\boldsymbol{b}_T^2); \mu, \zeta)} \right] \hat{f}_1(x, b_*(\boldsymbol{b}_T^2); \mu, \zeta) \\
&\equiv f_{\text{NP}}(x, \boldsymbol{b}_T^2, \zeta) \hat{f}_1(x, b_*(\boldsymbol{b}_T^2); \mu, \zeta) .
\end{aligned}
\tag{2.76}
$$

The advantage of such separation is that, due to the saturation of $b_*(\boldsymbol{b}_T^2)$ for large values of $|\boldsymbol{b}_T|$, the function $\hat{f}_1(x, b_*(\boldsymbol{b}_T^2), \mu, \zeta)$ is kept in the perturbative region. Instead, the non-perturbative ingredients are confined into f_{NP}. This new function must be modeled with a set of parameters that have to be determined through a fit to the available experimental data. The choice of such model is guided by some properties of f_{NP} that can be derived from Eq. (2.76). First of all, we can see that f_{NP} does not depend on the renormalization scale μ. Indeed, by using Eqs. (2.44) and (2.45) and by imposing $\mu_i = \sqrt{\zeta_i} = \mu_b$, we find

$$
\begin{aligned}
f_{\text{NP}}(x, \boldsymbol{b}_T^2, \zeta) &= \frac{\hat{f}_1(x, \boldsymbol{b}_T^2; \mu, \zeta)}{\hat{f}_1(x, b_*(\boldsymbol{b}_T^2); \mu, \zeta)} \\
&= \frac{\hat{f}_1(x, \boldsymbol{b}_T^2; \mu_b, \mu_b^2)}{\hat{f}_1(x, b_*(\boldsymbol{b}_T^2); \mu_{b_*}, \mu_{b_*}^2)} \exp\left\{ K(\mu_b) \ln \frac{\sqrt{\zeta}}{\mu_b} - K(\mu_{b_*}) \ln \frac{\sqrt{\zeta}}{\mu_{b_*}} \right\} \\
&\quad \exp\left\{ \int_{\mu_b}^{\mu_{b_*}} \frac{d\mu'}{\mu'} \left[\gamma_F(\alpha_s(\mu')) - \gamma_K(\alpha_s(\mu')) \ln \frac{\sqrt{\zeta}}{\mu'} \right] \right\} ,
\end{aligned}
\tag{2.77}
$$

where $\mu_{b_*} \equiv \mu_b(b_*(\boldsymbol{b}_T^2))$. It is clear that the dependence on μ cancels in the ratio. Moreover, μ_{b_*} saturates to some minimal value $\mu_b(b_{\text{max}})$ for large values of $|\boldsymbol{b}_T|$, while μ_b becomes increasingly small and approaches to 0. As a consequence of this deviation between μ_{b_*} and μ_b, as well as between $\sqrt{\zeta}$ and μ_b in the exponential in Eq. (2.77), the function f_{NP} tends to be suppressed.[8] In the small-$|\boldsymbol{b}_T|$ region, b_* approaches b_{min} while $|\boldsymbol{b}_T|$ approaches to 0. Thus, looking at Eq. (2.75), we can see that μ_{b_*} saturates to Q while μ_b becomes larger and larger. In this limit, we have [171]

$$
f_{\text{NP}} \xrightarrow[|\boldsymbol{b}_T| \to 0]{} 1 + \mathcal{O}\left(\frac{1}{Q^p} \right) ,
\tag{2.78}
$$

where p is some positive number. However, we can neglect the power suppressed contribution because we assume that they are small in the kinematic re-

---

[8]It can be shown that the argument of both the exponential functions in Eq. (2.77) tends to $-\infty$ as $|\boldsymbol{b}_T| \to +\infty$.

gion addressed our investigations. In this way, $f_{\rm NP} \to 1$ for $|\boldsymbol{b}_T| \to 0$. We stress that the "artificial" separation between perturbative and non-perturbative components of a TMD is *arbitrary* and depends on the particular choice of the function $b_*(\boldsymbol{b}_T^2)$ (or in general on the prescription used to avoid the Landau pole). For any choice, only the combination in Eq. (2.76) is meaningful. In Sec. 3.3.2, we will show the impact of modifying the b_* prescription in the shape of the TMD distribution as extracted in phenomenological analyses.

Another way to design the full TMD distribution has been recently proposed in Ref. [173]. In this "hadron structure oriented" approach, the focus is put on preserving a connection to the TMD parton model interpretation when combining perturbative and nonperturbative transverse momentum in TMD factorization.

2.4 Normalization factors for SIDIS multiplicities

In this section, we address the normalization issue that occurs in the description of SIDIS multiplicities when pushing the theoretical formalism to high accuracy. This issue was pointed out for the first time in phenomenological studies in Refs. [86, 174] and, recently, in Ref. [1], whereas it was not observed in the global analysis reported in Ref. [88]. Therefore, the normalization of SIDIS multiplicity is an issue that the TMD community is still discussing, and it could be related to general properties of the TMD formalism. In this thesis, we report the approach introduced in Ref. [1] to fix this problem.

We start from pointing out some known outcomes regarding previous analyses which included SIDIS multiplicities. In Refs. [84, 85], it was shown that a good description of both normalization and shape of HERMES SIDIS multiplicities can be achieved still at the level of parton model. This means that the experimental data can be described to a good extent at the lowest perturbative order (LO) with a simple Gaussian ansatz for the nonperturbative part of TMDs and without TMD evolution. As a consequence, it is straightforward to expect that the TMD formalism at NLL is able to get better or similar results, since it only implies the introduction of the double-scale evolution through the Sudakov form factor. Indeed, a first attempt to compare theoretical calculations at NLL and experimental data of HERMES multiplicities was provided in the analysis of Refs. [175], where no issue on the normalization of such data was observed. Moreover, within TMD factorization in Ref. [87] it was possible to succesfully reproduce the normalization and shape of HERMES SIDIS multiplicities and the shape of the available COMPASS multiplicities at NLL accuracy. More recently, a new release of COMPASS data was published [141]. Then, in Ref. [176] it was demonstrated that the TMD distributions extracted from Ref. [87] are able to correctly reproduce normalization and shape of the new data. However, when going beyond NLL in the perturbative accuracy (to NLL' or higher), the TMD formula severely underestimates the measure-

ments [86, 174, 176] by nearly constant factors in each bin.[9]

As it will be shown in Sec. 3.3.1, the authors of Ref. [1] confirm that an excellent description of both normalization and shape of the SIDIS multiplicities can be achieved at NLL accuracy, whereas the N^2LL results are much smaller. The quality of the description at N^3LL is similar, with the theoretical predictions still far from the NLL ones and, consequently, from the experimental data. At average kinematics of the COMPASS measurements, the ratios of multiplicities M_{NLL}/M_{N^2LL} and M_{NLL}/M_{N^3LL} can even exceed a factor 2.

By looking at Tab. 2.1, one can identify the order in the strong coupling that is needed to reach the consistent accuracy in the resummation of logarithmic divergences. It is clear that an appreciable modification on the normalization of the predictions at two different accuracies can be obtained only with a significant variation of the hard factor, the matching coefficients, or the collinear input (or all of them). In our analysis, the major difference in the predictions between different logarithmic orders is due to the perturbative corrections to the hard factor in Eq. (2.30) [128]. We report here the expression of the SIDIS hard factor $\mathcal{H}^{SIDIS}(Q, \mu)$[10] at NLO with the standard choice $\mu = Q$

$$\mathcal{H}^{\text{SIDIS}}(Q, Q) = 1 + \frac{\alpha_s(Q)}{4\pi} C_F \left(-16 + \frac{\pi^2}{3} \right). \tag{2.79}$$

We can immediately see that the structure function in Eq. (2.30), at $Q = 2$ GeV with $\alpha_s \approx 0.3$, is reduced to about 60% of its original value just by introducing $\mathcal{O}(\alpha_s)$ corrections. This modification is not compensated by a similar suppression in the DIS cross section in the denominator of Eq. (2.34). In fact, the differences between the LO and NLO expressions of the inclusive DIS cross section are typically below 5% and the NLO results are actually larger than the LO ones. A similar result was obtained in Ref. [88], where it was shown that the SIDIS cross section in TMD factorization drops down going beyond NLL. However, in Ref. [88] SIDIS experimental data are nicely described in both shape and normalization at N^3LL, even if with a very restrictive cut on the TMD region.

Following Ref. [88], one could argue that the NLL expression would overshoot the data by a factor 1.5 at least, while N^2LL and N^3LL ones should reproduce them. However, as already pointed out at the beginning of this section, there are several past works that have shown a good agreement with data even using a parton-model approach. Moreover, we stress that the structure function in Eq. (2.30) integrated over $\boldsymbol{q}_T$ corresponds to the value at $|\boldsymbol{b}_T| = 0$ of its Fourier transform (see Eq.Eq. 2.9). Since we introduced a b_* prescription

---

[9]Note that tensions between the TMD cross sections and the associated measurements exist also at large transverse momentum in SIDIS [177], DY [178], and electron-positron annihilation into two hadrons [179].

[10]In the literature, different definitions for the hard factor have been computed, but they are compensated by different definitions of the matching coefficients C in Eq.(2.52). Here we follow the definition of Ref. [128].

with a cutoff at $b_{\min} = 2e^{-\gamma_E}/\mu$, the value of the NLL TMD formula at $|\boldsymbol{b}_T| = 0$ is equivalent to the LO expression of the collinear SIDIS structure function, which is independent of the TMD nonperturbative parameters. Therefore, we can express the integral of the structure function in Eq. (2.30) as follows[11]

$$
\begin{aligned}
\int d^2\boldsymbol{q}_T F_{UU,T} &= x \sum_a e_a^2 \, \mathcal{H}(Q,Q) \left(\hat{f}_1^a(x, \boldsymbol{b}_T^2; Q) \hat{D}_1^{a\to h}(z, \boldsymbol{b}_T^2; Q) \right) \bigg|_{|\boldsymbol{b}_T|=0} \\
&\overset{\mathrm{NLL}}{=} x \sum_a e_a^2 \, f_1^a(x; Q) \, D_1^{a\to h}(z; Q).
\end{aligned} \tag{2.80}
$$

It is commonly known that LO predictions for SIDIS process in collinear framework give a reasonable description of the experimental data and, if anything, they seem to be lower (and not larger) than the experimental data [142, 180]. Consequently, the integral of the NLL expression of the structure function included in our analysis must be in good agreement with the data, and errors in the normalization should not occur at this level of logarithmic accuracy.

Given all these reason, namely that the NLL predictions describe SIDIS data well and the $\mathrm{N}^2\mathrm{LL}$ and $\mathrm{N}^3\mathrm{LL}$ calculations are below by a constant factor, we propose a modification of the normalization to restore the good agreement with data obtained at NLL. This issue was already extensively discussed in Ref. [176], which we take as inspiration for our proposed solution. Since the TMD factorization is valid only at low $|\boldsymbol{q}_T|$, we observe that the integral of the TMD formula should reproduce only part of the full collinear cross section. In fact, when integrating over the full transverse-momentum space, terms at large $|\boldsymbol{q}_T|$ are missing. Following the discussion above, the only exception is the order $\mathcal{O}(\alpha_s^0)$ in $b_{\min}$ prescription, since the contributions from gluon radiation at high transverse momentum are not present at this order.

At higher orders, the integral of the cross section in the TMD region (*i.e.*, the integral of the so-called W term in the language of Ref. [73]), is much smaller than the corresponding collinear cross section at $\mathrm{N}^2\mathrm{LL}$ or higher accuracies in the kinematics of fixed-target SIDIS experiments (namely, at low values of the energy scale Q). In order to recover the full collinear cross section, one should include all the terms in the fixed-order calculation that are not included in the TMD resummed expression. This is the so-called Y term in the conventions of Ref. [73]

In principle, the Y term should be negligible in the low-$|\boldsymbol{q}_T|$ region, where the W term should be dominant and should reproduce the normalization of experimental data. However, this is not always the case at low Q: the Y term is finite but relatively large, even at $|\boldsymbol{q}_T| = 0$ [86].

The integral of the W term (*i.e.*, the integral of Eq. (2.27)), with our $b_{\min}$

---

[11]This argument is true for any b_* prescription with $b_{\min} = 2e^{-\gamma_E}/\mu$. Note that in the absence of a $b_{\min}$ prescription, the integral of the structure function would vanish.

prescription, at order $\mathcal{O}\alpha_s$, is given by

$$\int d^2\boldsymbol{q}_T\, W\bigg|_{\mathcal{O}(\alpha_s)} = \sigma_0 \frac{\alpha_s}{4\pi} \sum_q e_q^2 \Big[D_1^{q\to h} \otimes C_{\mathrm{TMD}}^{qq} \otimes f_1^q$$
$$+ D_1^{q\to h} \otimes C_{\mathrm{TMD}}^{qg} \otimes f_1^g \qquad (2.81)$$
$$+ D_1^{g\to h} \otimes C_{\mathrm{TMD}}^{gq} \otimes f_1^q \Big](x, z, Q),$$

where

$$\sigma_0 = \frac{4\pi^2\,\alpha^2\,z^2\,|\boldsymbol{q}_T|}{x\,Q^3} Y_+. \qquad (2.82)$$

We define the double convolution over both x and z as

$$\Big[D_1^{a\to h} \otimes C^{ab} \otimes f_1^b \Big](x, z, Q)$$
$$= \frac{1}{z^2} \int_x^1 \frac{dx'}{x'} \int_z^1 \frac{dz'}{z'} D_1^{a\to h}(z'; Q) C^{ab}\left(\frac{x}{x'}, \frac{z}{z'}\right) f_1^b(x'; Q). \qquad (2.83)$$

The C_{TMD}^{ab} coefficients of Eq. (2.81) are given by:

$$C_{\mathrm{TMD}}^{qq}(x, z) = 2C_F\bigg(-8\delta(1-x)\delta(1-z) + \delta(1-x)\Big[2L_2(z) + (1-z)\Big]$$
$$+ \delta(1-z)(1-x) \bigg), \qquad (2.84)$$

$$C_{\mathrm{TMD}}^{gq}(x, z) = 2C_F\bigg(P_{gq}(z)\delta(1-x)\ln\left(z(1-z)\right) + z\delta(1-x) \bigg), \qquad (2.85)$$

$$C_{\mathrm{TMD}}^{qg}(x, z) = 2T_F\bigg(\delta(1-z)2x(1-x) \bigg). \qquad (2.86)$$

In the above expressions L_2 is the abbreviation for the logarithmic function

$$L_2(\xi) \equiv \frac{1+\xi^2}{1-\xi}\ln\xi\,. \qquad (2.87)$$

Moreover, here and in the following, the functions P_{qq}, P_{gq} and P_{qg} are the standard Altarelli-Parisi splitting functions. The C_{TMD}^{ab} coefficients correspond to the combination of the matching coefficients C^{ab} from the OPE expression at low-$|\boldsymbol{b}_T|$ of the TMD PDFs and FFs (see Sec. 2.3.1).

The W term integrated over transverse momentum $|\boldsymbol{q}_T|$ in Eq. (2.81) should be compared to the collinear expression of the cross section at the same order

(see, *e.g.*, Ref. [181])

$$
\begin{aligned}
\frac{d\sigma^{\text{SIDIS}}}{dx dQ dz}\bigg|_{\mathcal{O}(\alpha_s)} = \sigma_0 \frac{\alpha_s}{4\pi} \sum_q e_q^2 \bigg\{ &\Big[D_1^{q\to h} \otimes C_1^{qq} \otimes f_1^q \\
&+ D_1^{q\to h} \otimes C_1^{qg} \otimes f_1^g \\
&+ D_1^{g\to h} \otimes C_1^{gq} \otimes f_1^q \Big](x,z,Q) \\
+ \frac{1-y}{1+(1-y)^2} &\Big[D_1^{q\to h} \otimes C_L^{qq} \otimes f_1^q \\
&+ D_1^{q\to h} \otimes C_L^{qg} \otimes f_1^g \\
&+ D_1^{g\to h} \otimes C_L^{gq} \otimes f_1^q \Big](x,z,Q) \bigg\}.
\end{aligned}
$$
(2.88)

The QCD coefficients C_1^{ab} and C_L^{ab} can be perturbatively calculated. The former can be written as

$$
C_1^{ab}(x,z;Q,\mu) = C_{\text{nomix}}^{ab}(x,z,Q,\mu) + C_{\text{mix}}^{ab}(x,z),
$$
(2.89)

where we define the coefficient C_{nomix} as the sum of all those terms that contain either a $\delta(1-x)$ or a $\delta(1-z)$, or both. We note that not all these terms are present in C_{TMD}. This definition holds at all orders in α_s. It can be shown that the C_L matching coefficients, instead, only contain "mixed" contributions [176].

We report here the explicit expression of the C_{nomix}^{ab} coefficients:

$$
\begin{aligned}
C_{\text{nomix}}^{qq}(x,z;Q,\mu) = C_{\text{TMD}}^{qq}(x,z) &\\
+ 2C_F \bigg[\delta(1-x) &\bigg(P_{qq}(z) \ln \frac{Q^2}{\mu^2} + L_1(z) - L_2(z) \bigg) \\
+ \delta(1-z) &\bigg(P_{qq}(x) \ln \frac{Q^2}{\mu^2} + L_1(x) - L_2(x) \bigg) \bigg],
\end{aligned}
$$
(2.90)

$$
C_{\text{nomix}}^{gq}(x,z;Q,\mu) = C_{\text{TMD}}^{gq}(x,z) + 2C_F \delta(1-x) P_{gq}(z) \ln \frac{Q^2}{\mu^2},
$$
(2.91)

$$
C_{\text{nomix}}^{qg}(x,z;Q,\mu) = C_{\text{TMD}}^{qg}(x,z) + 2T_F \delta(1-z) P_{qg}(x) \ln \left(\frac{Q^2}{\mu^2} \frac{1-x}{x} \right).
$$
(2.92)

Here, L_1 is the abbreviations for the following logarithmic function:

$$
L_1(\xi) \equiv \left(1 + \xi^2\right) \left(\frac{\ln(1-\xi)}{1-\xi} \right)_+.
$$
(2.93)

The expressions for the "mixed" coefficients C^{ab}_{mixed}, are given by:

$$
\begin{aligned}
C^{qq}_{\text{mixed}}(x, z; Q, \mu) = 2C_F \Bigg[2 \frac{1}{(1-x)_+} \frac{1}{(1-z)_+} \\
- \frac{1+z}{(1-x)_+} - \frac{1+x}{(1-z)_+} + 2(1+xz) \Bigg],
\end{aligned}
\tag{2.94}
$$

$$
C^{gq}_{\text{mixed}}(x, z; Q, \mu) = 2C_F \left[P_{gq}(z) \frac{1}{(1-x)_+} + 2(1+x-xz) - \frac{1+x}{z} \right], \quad (2.95)
$$

$$
C^{qg}_{\text{mixed}}(x, z; Q, \mu) = 2T_F P_{qg}(x) \left[\frac{1}{(1-z)_+} + \frac{1}{z} - 2 \right]. \tag{2.96}
$$

In the above expressions μ denotes the factorization scale for PDFs or FFs, $T_F = 3$.

In order to find a solution of the normalization issue of TMD predictions for SIDIS multiplicities, we take into account only the contribution of all the "nonmixed" terms C_{nomix} to build the collinear cross section. We consider this choice reasonable because the factorizable nature of C_{nomix} allows for a redefinition of each individual TMD. Hence, we define

$$
\begin{aligned}
\frac{d\sigma^{\text{nomix}}}{dx dQ dz} \bigg|_{\mathcal{O}(\alpha_s)} = \sigma_0 \frac{\alpha_s}{4\pi} \sum_q e_q^2 \Big[& D_1^{q \to h} \otimes C^{qq}_{\text{nomix}} \otimes f_1^q \\
& + D_1^{q \to h} \otimes C^{qg}_{\text{nomix}} \otimes f_1^g \\
& + D_1^{g \to h} \otimes C^{gq}_{\text{nomix}} \otimes f_1^q \Big] (x, z, Q),
\end{aligned}
\tag{2.97}
$$

and similarly for higher orders. Then, we introduce the following normalization factor:

$$
\omega(x, z, Q) = \frac{d\sigma^{\text{nomix}}}{dx \, dz \, dQ} \bigg/ \int d^2 \boldsymbol{q}_T \, W. \tag{2.98}
$$

It is important to stress that this normalization factor depends only on the choice of the collinear PDFs and FFs sets, is independent of the parametrization of the TMD nonperturbative part (see Eq. (2.88)), and, thus, can be precomputed before performing a new fit.

At NLL, the normalization prefactor is 1 by definition with the b_{min} cutoff in b_* prescription. By increasing the logarithmic accuracy beyond NLL, the value of the prefactor exceeds one and guarantees that the integral of the TMD part of the cross section reproduces most of the collinear cross section, as suggested by the data. On the contrary, without the inclusion of the normalization factor, the integral of the TMD formula for the cross section would be too small, requiring a compensation from the high-transverse-momentum tail, which is not observed in the experimental data.

As a consequence of the introduction of our normalization prefactor, the

theoretical expression for the SIDIS cross section in Eq. (2.27) becomes

$$\frac{d\sigma_\omega^{\text{SIDIS}}}{dx\,dz\,d|\boldsymbol{q}_T|\,dQ} = \omega(x,z,Q)\,\frac{d\sigma^{\text{SIDIS}}}{dx\,dz\,d|\boldsymbol{q}_T|\,dQ}\,. \tag{2.99}$$

The impact of the normalization factor defined in Eq. (2.98) will be addressed in detail in Sec. 3.2.2.

2.A Appendix - Analytic expression of perturbative ingredients

In this section, we report the analytical expression of the perturbative ingredients needed to build the structure functions F_{UU}^1 and $F_{UU,T}$ for DY and SIDIS processes in TMD factorization at N^2LL accuracy (see Tab. 2.1). We follow the definition of Ref. [128] for the hard factor (see Sec. 2.1 and Sec 2.2) and the matching coefficients (see Sec. 2.3.1). The NLO expression of the hard factor $\mathcal{H}^{\text{SIDIS}}(Q,Q)$ for SIDIS process is reported in Eq. (2.79). The one for DY process reads

$$\mathcal{H}^{\text{DY}}(Q,Q) = 1 + \frac{\alpha_s(Q)}{4\pi}C_F\left(-16 + \frac{7\pi^2}{3}\right), \tag{2.100}$$

where $C_F = 4/3$. The NLO expressions of matching coefficients $C_{ab}^{\text{PDF}}(x,b,\mu,\zeta)$ for TMD PDFs (see Eq. (2.52)) are given by

$$C_{qq}^{\text{PDF}}(x,b_*,\mu_{b_*},\mu_{b_*}^2) = \delta(1-x)$$
$$+ \frac{\alpha_s(\mu_{b_*})}{4\pi}C_F\left(-\frac{\pi^2}{6}\delta(1-x) + 2(1-x)\right), \tag{2.101}$$

$$C_{qg}^{\text{PDF}}(x,b_*,\mu_{b_*},\mu_{b_*}^2) = \frac{\alpha_s(\mu_{b_*})}{4\pi}\left(2x(1-x)\right), \tag{2.102}$$

$$C_{qq'}^{\text{PDF}}(x,b_*,\mu_{b_*},\mu_{b_*}^2) = 0, \tag{2.103}$$
$$C_{q\bar{q}}^{\text{PDF}}(x,b_*,\mu_{b_*},\mu_{b_*}^2) = 0, \tag{2.104}$$
$$C_{q\bar{q}'}^{\text{PDF}}(x,b_*,\mu_{b_*},\mu_{b_*}^2) = 0, \tag{2.105}$$

where μ_{b_*} has been defined in Sec. 2.3.4. The NLO expressions of the ones for TMD FFs are given by

$$C_{qq}^{\text{FF}}(z,b_*,\mu_{b_*},\mu_{b_*}^2) = \delta(1-z)$$
$$+ \frac{\alpha_s(\mu_{b_*})}{4\pi}C_F\frac{1}{z^2}\left(4p_{qq}(z)\ln z - \frac{\pi^2}{6}\delta(1-z) + 2(1-z)\right), \tag{2.106}$$

$$C_{qg}^{\text{FF}}(z,b_*,\mu_{b_*},\mu_{b_*}^2) = \frac{\alpha_s(\mu_{b_*})}{4\pi}\frac{1}{z^2}\left(2p_{qg}(z)\ln z + 2z(1-z)\right), \tag{2.107}$$

$$C_{qq'}^{\text{FF}}(z,b_*,\mu_{b_*},\mu_{b_*}^2) = 0, \tag{2.108}$$
$$C_{q\bar{q}}^{\text{FF}}(z,b_*,\mu_{b_*},\mu_{b_*}^2) = 0, \tag{2.109}$$
$$C_{q\bar{q}'}^{\text{FF}}(z,b_*,\mu_{b_*},\mu_{b_*}^2) = 0, \tag{2.110}$$

where the functions $p_{qq}(z)$ and $p_{gq}(z)$ are related to the one-loop DGLAP kernels and are defined as [150]

$$p_{qq}(z) = \frac{1 + x^2}{1 - x}, \tag{2.111}$$

$$p_{qg}(z) = 1 - 2z(1 - z). \tag{2.112}$$

Also, the perturbative expressions at NNLO of the anomalous dimensions K and γ_F read

$$
\begin{aligned}
K(|\boldsymbol{b}_T|, \mu) = {} & -8C_F \frac{\alpha_s(\mu)}{4\pi} \boldsymbol{L}_\mu \\
& + \left(\frac{\alpha_s(\mu)}{4\pi}\right)^2 8C_F \left[\left(\frac{2}{3}n_f - \frac{11}{3}C_A\right)\boldsymbol{L}_\mu^2 \right. \\
& \qquad\qquad + \left(-\frac{67}{9}C_A + \frac{\pi^2}{3}C_A + \frac{10}{9}n_f\right)\boldsymbol{L}_\mu \\
& \qquad\qquad \left. + \left(\frac{7}{2}\zeta_3 - \frac{101}{27}\right)C_A + \frac{14}{27}n_f\right],
\end{aligned}
\tag{2.113}
$$

$$
\begin{aligned}
\gamma_F(\mu) = {} & \frac{\alpha_s(\mu)}{4\pi} 6C_F \\
& + \left(\frac{\alpha_s(\mu)}{4\pi}\right)^2 \left[C_F^2(3 - 4\pi^3 + 48\zeta_3) + C_F C_A\left(\frac{961}{27} + \frac{11\pi^2}{3} - 52\zeta_3\right)\right. \\
& \qquad\qquad \left. + C_F n_f\left(-\frac{130}{27} - \frac{2\pi^2}{3}\right)\right],
\end{aligned}
\tag{2.114}
$$

where n_f is the number of active flavors, $C_A = 3$, ζ_3 is the value of the Riemann Zeta function $\zeta(t)$ at $t = 3$, and

$$\boldsymbol{L}_\mu = \ln\left(\frac{\mu|\boldsymbol{b}_T|}{2e^{-\gamma_E}}\right). \tag{2.115}$$

Finally, the N³LO expression of the anomalous dimension γ_K is given by

$$
\begin{aligned}
\gamma_K(\mu) = {} & 8C_F \frac{\alpha_s(\mu)}{4\pi} \\
& + \left(\frac{\alpha_s(\mu)}{4\pi}\right)^2 \left[C_A C_F\left(\frac{536}{9} - \frac{8\pi^2}{3}\right) - \frac{80}{9}C_F n_f\right] \\
& + \left(\frac{\alpha_s(\mu)}{4\pi}\right)^3 \left[-\frac{32}{27}C_F n_f^2 + C_A C_F n_f\left(-\frac{1672}{27} + \frac{160\pi^2}{27} - \frac{224\zeta_3}{3}\right)\right. \\
& \qquad\qquad + C_A^2 C_F\left(\frac{980}{3} - \frac{1072\pi^2}{27} + \frac{88\pi^4}{45} + \frac{176\zeta_3}{3}\right) \\
& \qquad\qquad \left. + C_F^2 n_f\left(-\frac{220}{3} + 64\zeta_3\right)\right]
\end{aligned}
\tag{2.116}
$$

Chapter 3

Extraction of unpolarized proton TMDs

In this chapter, we report the most recent simultaneous extraction of unpolarized quark TMD PDFs and FFs using more than two thousand data points from several experiments for two different kinds of processes: Semi-Inclusive Deep Inelastic Scattering (SIDIS) and production of Drell–Yan (DY) lepton pairs, significantly improving the previous analysis by the MAP Collaboration [87]. This new analysis is an important step towards understanding the 3D structure of protons at the best possible degree of precision. In fact, this is currently the most sophisticated global analysis in TMD phenomenology.

The level of sophistication of a TMD extraction essentially depends on two ingredients: the amount of analyzed data from different processes, namely how general is the experimental information from which TMDs are extracted, and the perturbative accuracy of the theoretical formalism.

The process of TMD extraction relies on TMD factorization theorems, which determine a precise framework for defining and extracting these quantities, and establish their universality and evolution equations. In this context, the precision of the calculations is defined by the maximum order reached in the resummation of large logarithmic terms, which in turn determines the appropriate powers of the strong coupling constant, α_s, to be included into the perturbative expansion of quantities involved in the calculations [73, 81, 164, 182].

A fundamental aspect in TMD extractions is also the combination of analyzed data coming from different processes. Only in this case, we can speak of a global fit, meaning that it tests the universality of parton distributions by combining information from different sources. The cross section of SIDIS process is built by two kind of TMD distributions: the TMD PDFs, describing how partons are arranged in the nucleon, and TMD FFs, describing how a parton fragments into a detected final-state hadron. Independent information about TMD PDFs can be obtained also from DY measurements, since there are no identified hadrons in the final state. Instead, the knowledge of TMD FFs would be significantly improved by using data from electron-positron annihila-

tions into two almost back-to-back hadrons [170]. Unfortunately, this data are presently not available. However, measurements for the inclusive production of single hadrons have been recently made available by the BELLE Collaboration [183]. The issue with this data is that transverse momenta need to be defined with respect to the thrust axis. Rigorous factorization theorems for this process have been discussed from different approaches only recently (see, e.g., Refs. [184–187] and references therein). A first extraction of TMD FFs from BELLE data in a non-standard approach has been reported in Ref. [188]. At the moment, this kind of observable cannot be incorporated in standard unpolarized TMD fitting frameworks. Therefore, for unpolarized TMD extractions we currently talk about a global fit when data from SIDIS and DY processes are included.

In the last decade, several extractions of TMDs have been presented [78–82, 84–88, 175, 189]. Most of them have some limitations: they are either performed in a parton-model framework without QCD corrections (and TMD evolution), or they do not consider the complete set of available data, or do not perform a full global fit. TMDs are also studied in a different framework, the so-called parton-branching approach (see Refs. [89–91]).

Before the present analysis, henceforth named MAP22, there are only two works that have successfully reached the stage of performing a comprehensive TMD global fit by combining SIDIS and DY data: the above mentioned extraction of Ref. [87], henceforth named PV17, and the extraction of Ref. [88], henceforth named SV19. As already discussed in Sec. 2.4, the PV17 extraction reached only the NLL accuracy, it was based on the calculation at mean kinematics in each bin of the included experimental observables, and did not manage to describe the normalization of all data sets (COMPASS and Tevatron data). In the MAP22 analysis, we improved the accuracy of the theoretical calculations to what we will refer to as N^3LL^-. In fact, at the time of this work there were no NNLO collinear FFs, which are needed to reach the full N^3LL accuracy (see Tab. 2.1).[1] In addition to the improvement in the perturbative accuracy, this analysis includes many measurements published after the PV17 extraction.

Instead, the SV19 global fit was performed at the same perturbative accuracy and included approximately the same data sets as MAP22. One significant difference is in the selection of the experimental points to be included in the global fit. Precisely, we include a larger number of experimental data points. Moreover, the implementation of TMD evolution (see Sec. 2.3.1), the choice of functional forms of the nonperturbative components, and the control of the SIDIS normalization (see the discussion in Sec. 2.4) are crucial differences between MAP22 and SV19.

As already discussed in Sec. 2.4, the description of the normalization of SIDIS data from fixed-target experiments at moderate to low scales presents

---

[1]During the completion of this work, further efforts were made to include part of the NNLO corrections [33, 190, 191] in the extraction of FFs.

considerable difficulties in our analysis. To address this issue, we choose to adjust the theoretical predictions with the normalization prefactor defined in Eq. (2.98). With this procedure, we set the normalization of our calculations before the fit, in a way that is independent of the fitting parameters.

The baseline fit presented in this chapter is performed at N^3LL^-, including 2031 data points and obtaining $\chi^2/N_{dat} = 1.06$. We also discuss variations of this baseline fit by changing the theoretical accuracy, the selected data, and the b_* prescription in the theoretical calculations.

The chapter is organized as follows. In Sec. 3.1, we describe the global data set included in the MAP22 analysis and how experimental data have been selected. In Sec. 3.2 we present the results of our global fit. In Sec. 3.3, we discuss the possible variations of the baseline setup. Finally, Sec. 3.4 shows the estimate of the impact of pseudodata from the EIC and from a possible JLab 22 GeV upgrade.

3.1 Data selection

In this section, we describe the experimental data included in the MAP22 [1] simultaneous extraction of unpolarized quark TMD PDFs in the proton and unpolarized TMD FFs. We consider a very large number of data sets for both DY lepton pair production and SIDIS processes, for the observables discussed in Secs. 2.1-2.2. The kinematic coverage in the x-Q^2 plane spanned by these experimental sets is displayed in Fig. 3.1.

The data sets included in the analysis presented in this chapter are:

- DY di-muon production from the collision of a proton beam with an energy of 200, 300 and 400 GeV on copper and lead fixed targets from E288 ($\sqrt{s} = 19.4, 23.8, 27.4$ GeV, respectively) [192];

- DY di-muon production from the collision of a proton beam with an energy of 800 GeV on a copper fixed target from E605 ($\sqrt{s} = 38.8$ GeV) [193];

- DY di-muon production from the collision of a proton beam with an energy of 800 GeV on a ^2H fixed target from E772 ($\sqrt{s} = 38.8$ GeV) [194];

- $Z \to e^+e^-$ distribution from CDF experiment of Tevatron Run I [195] and Run II [196] ($\sqrt{s} = 1.8, 1.96$ TeV, respectively);

- $Z \to e^+e^-$ distribution from CDF experiment of Tevatron Run I [197] and Run II [198] ($\sqrt{s} = 1.8, 1.96$ TeV, respectively);

- $Z \to \mu^+\mu^-$ distribution from D0 Run II [199];

- forward Z-production data from the LHCb experiment at 7 [199], 8 [200], and 13 [201] TeV;

- Z-production data from the CMS experiment at 7 [202], 8 [203] and 13 [204] TeV;

- Z-production data differential in rapidity from the ATLAS experiment at 7 [202], 8 [205] and 13 [206] TeV;

- off-peak (low- and high-mass) DY data from the ATLAS experiment at 8 TeV [205],

- preliminary Z-production data from the STAR experiment at 510 GeV;

- DY di-muon production from the PHENIX Collaboration [207];

- Unpolarized SIDIS multiplicites on hydrogen (p) or deuterium (D) gas targets from HERMES [142] experiment ($\sqrt{s} = 7.26$ GeV);

- Unpolarized SIDIS multiplicites on $6^L iD$ target from COMPASS [141] experiment ($\sqrt{s} = 17.325$ GeV).

The majority of data sets included in this work was already taken into consideration in the global fit of SIDIS and DY data in Ref. [87] and in the fit of DY data discussed in Ref. [81]. For more details, we refer the reader to those papers.
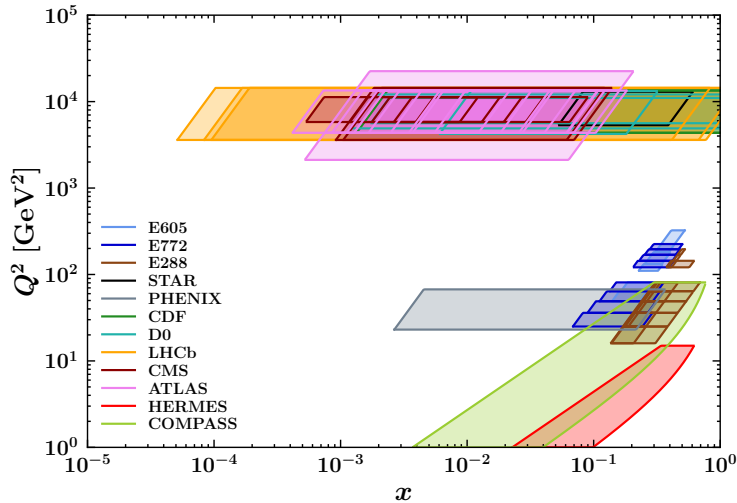

Figure 3.1: The x vs. Q^2 coverage spanned by the experimental data considered in the MAP22 analysis (see also Tab. 3.1 and Tab. 3.3).

3.1.1 Drell–Yan data set

Since our analysis is based on the applicability of TMD factorization, which is valid only in the region $|\boldsymbol{q}_T| \ll Q$, we impose the following cut on the transverse

momentum of the intermediate boson for the DY data set

$$|\boldsymbol{q}_T| < 0.2\,Q\,, \tag{3.1}$$

which is comparable to the choices of Refs. [81–83, 88], sometimes even more restrictive. We note that in Refs [82, 83, 88] a further cut on the ratio between experimental and theoretical uncertainties on high-energy data (particularly, for the ATLAS data set) has been included, and the E772 data sets below 9 GeV were not considered.

Experiment	N_{dat}	Observable	$\sqrt{s}$ [GeV]	Q [GeV]	y or x_F	Ref.
E605	50	$Ed^3\sigma/d^3\boldsymbol{q}$	38.8	7 - 18	$x_F = 0.1$	[193]
E772	53	$Ed^3\sigma/d^3\boldsymbol{q}$	38.8	5 - 15	$0.1 < x_F < 0.3$	[194]
E288 200 GeV	30	$Ed^3\sigma/d^3\boldsymbol{q}$	19.4	4 - 9	$y = 0.40$	[192]
E288 300 GeV	39	$Ed^3\sigma/d^3\boldsymbol{q}$	23.8	4 - 12	$y = 0.21$	[192]
E288 400 GeV	61	$Ed^3\sigma/d^3\boldsymbol{q}$	27.4	5 - 14	$y = 0.03$	[192]
STAR 510	7	$d\sigma/d|\boldsymbol{q}_T|$	510	73 - 114	$|y| < 1$	-
PHENIX200	2	$d\sigma/d|\boldsymbol{q}_T|$	200	4.8 - 8.2	$1.2 < y < 2.2$	[207]
CDF Run I	25	$d\sigma/d|\boldsymbol{q}_T|$	1800	66 - 116	Inclusive	[195]
CDF Run II	26	$d\sigma/d|\boldsymbol{q}_T|$	1960	66 - 116	Inclusive	[196]
D0 Run I	12	$d\sigma/d|\boldsymbol{q}_T|$	1800	75 - 105	Inclusive	[197]
D0 Run II	5	$(1/\sigma)d\sigma/d|\boldsymbol{q}_T|$	1960	70 - 110	Inclusive	[198]
D0 Run II (μ)	3	$(1/\sigma)d\sigma/d|\boldsymbol{q}_T|$	1960	65 - 115	$|y| < 1.7$	[199]
LHCb 7 TeV	7	$d\sigma/d|\boldsymbol{q}_T|$	7000	60 - 120	$2 < y < 4.5$	[200]
LHCb 8 TeV	7	$d\sigma/d|\boldsymbol{q}_T|$	8000	60 - 120	$2 < y < 4.5$	[201]
LHCb 13 TeV	7	$d\sigma/d|\boldsymbol{q}_T|$	13000	60 - 120	$2 < y < 4.5$	[208]
CMS 7 TeV	4	$(1/\sigma)d\sigma/d|\boldsymbol{q}_T|$	7000	60 - 120	$|y| < 2.1$	[202]
CMS 8 TeV	4	$(1/\sigma)d\sigma/d|\boldsymbol{q}_T|$	8000	60 - 120	$|y| < 2.1$	[203]
CMS 13 TeV	70	$d\sigma/d|\boldsymbol{q}_T|$	13000	76 - 106	$|y| < 2.4$ (5 bins)	[204]
ATLAS 7 TeV	18	$(1/\sigma)d\sigma/d|\boldsymbol{q}_T|$	7000	66 - 116	$|y| < 2.4$ (3 bins)	[209]
ATLAS 8 TeV on-peak	36	$(1/\sigma)d\sigma/d|\boldsymbol{q}_T|$	8000	66 - 116	$|y| < 2.4$ (6 bins)	[205]
ATLAS 8 TeV off-peak	4 / 8	$(1/\sigma)d\sigma/d|\boldsymbol{q}_T|$	8000	46 - 66 / 116 - 150	$|y| < 2.4$	[205]
ATLAS 13 TeV	6	$(1/\sigma)d\sigma/d|\boldsymbol{q}_T|$	13000	66 - 116	$|y| < 2.5$	[206]
Total	484					

Table 3.1: Breakdown of the DY data sets considered in the MAP22 analysis. For each data set, the table includes information on: the number of data points (N_{dat}) that survive the nominal cut on $|\boldsymbol{q}_T|$ (see Eq. (3.1)), the measured observable, the center-of-mass energy $\sqrt{s}$, the range(s) in invariant mass Q, the angular variable (either y or x_F), and the published reference (when available). The total number of DY data points amounts to 484.

In Tab. 3.1, we summarize all the DY experimental data sets included in

our analysis. Note that the experimental observable is given within a fiducial region for some sets. In this case, kinematic cuts on transverse momentum $p_{T\ell}$ and pseudo–rapidity η_ℓ of the single final-state leptons are enforced in the experimental analysis. This implementation in the `NangaParbat` computational tool has been extensively discussed in Ref. [81]. The second column of Tab. 3.1 reports the number of data points (N_{dat}) for each experiment after imposing the kinematic cut in Eq. (3.1).

As already pointed out in Sec. 2.1, the measured cross sections has been released in different forms by the experimental collaborations. In Tab. 3.1, we can see that some of them are normalized to the total (fiducial) cross section while others are not. For the data sets where the total cross section σ is required, we calculated it through the code `DYNNLO` [130, 131]. The fiducial cross section has been consistently calculated at the same perturbative order of the differential cross section. To be precise, it is computed at NLO for NNLL accuracy, and NNLO for N³LL accuracy. The values of the total cross sections at different orders are reported in Tab. 3.2. They have been calculated using the MMHT2014 set of collinear PDFs, according to choices that will be discussed in Sec. 3.1.3.

Experiment		LO [pb]	NLO [pb]	NNLO [pb]
D0 Run II		170.332	242.077	253.573
D0 Run II (μ)		100.765	119.002	124.675
CMS 7 TeV		291.977	384.569	398.853
CMS 8 TeV		340.132	456.337	473.411
ATLAS 7 TeV	$\|y\| < 1$	196.457	251.296	253.781
	$1 < \|y\| < 2$	135.511	181.267	181.466
	$2 < \|y\| < 2.4$	12.568	17.091	17.104
ATLAS 8 TeV on-peak	$\|y\| < 0.4$	89.531	113.650	116.766
	$0.4 < \|y\| < 0.8$	89.120	112.853	115.738
	$0.8 < \|y\| < 1.2$	85.499	109.800	112.457
	$1.2 < \|y\| < 1.6$	69.018	91.884	95.187
	$1.6 < \|y\| < 2$	43.597	59.114	62.127
	$2 < \|y\| < 2.4$	14.398	19.574	20.937
ATLAS 8 TeV off-peak	46 GeV $< Q <$ 66 GeV	15.199	14.449	14.368
	116 GeV $< Q <$ 150 GeV	3.805	5.317	5.521
ATLAS 13 TeV		-	694.3	707.3

Table 3.2: Total (fiducial) cross sections computed with `DYNNLO` [130, 131] using the central member of the `MMHT2014` collinear PDF sets [210]. It is required for the computation of the normalised differential cross sections at the different perturbative orders.

The total number of DY data points included in the analysis presented in this chapter is 484. Note that for E605, E772 and E288 at 400 GeV we have excluded those bins in Q around the Υ resonance peak ($Q \simeq 9.5$ GeV).

3.1.2 SIDIS data set

For the SIDIS process, the identification of the region of validity of the TMD formalism is not a trivial task. In fact, one of the issues about determining such region is that the available experimental data cover the region at small values of Q, where even the validity of collinear factorization can be challenged. Hence, the definition of the TMD region may be revised as new data are released and the theoretical description is improved, as discussed in Refs. [135, 211, 212].

In SIDIS process, different cuts on kinematics have to be applied. A first cut on the virtuality Q of the exchanged photon is needed to fulfill the condition $Q \gg \Lambda_{\mathrm{QCD}}$. In fact, this is the region where perturbation theory and the collinear factorization theorem are applicable. After imposing this kind of cut, the corrections $O(M^2/Q^2)$ (see the definition of the SIDIS structure function in Eq. (2.30)) can be neglected.[2] In order to fullfil this requirement, the energy scale Q^2 of the SIDIS experimental data included in the MAP22 analysis is required to be larger than 1.4 GeV^2, which is comparable to the cuts implemented in other studies of this process, even in collinear framework [32, 33].

A further cut on the variable z is required to restrict the analysis to the current fragmentation region. This is the region where the detected hadron in the final state is produced from the fragmentation of the active parton in the hard scattering. In this region, the experimental observable can be interpreted in terms of parton distribution and fragmentation functions. The kinematic region at small-z is named "target fragmentation region", and can be investigated through the formalism of fracture functions (see, *e.g.*, [213]). Instead, the region at large-z is governed by exclusive physics processes, and can be studied through Generalized Parton Distributions (GPDs) [214]. We restrict our global analysis in the current fragmentation region by requiring $z > 0.2$, and we avoid contributions from exclusive processes by considering only bins at $z < 0.7$. The lower limit of this cut is the same introduced in different fits of fragmentation functions in collinear framework [32, 33]. The upper limit is slightly more restrictive than in Ref. [87] in order to exclude contributions from exclusive channels and to focus on a region where the extracted collinear fragmentation functions have small relative uncertainties.

Finally, our baseline choice for the cut on transverse momentum is

$$|\boldsymbol{P}_{hT}| < \min\big[\min[c_1\,Q, c_2\,zQ] + c_3\,\mathrm{GeV},\ zQ\big]\,, \qquad (3.2)$$

with fixed parameters $c_1 = 0.2$, $c_2 = 0.5$ and $c_3 = 0.3$. Note that this choice is more restrictive than a similar one made in Ref. [87], but much less restrictive than the one made in Ref. [88]. It allows us to include in the analysis many data points with $|\boldsymbol{P}_{hT}| \ll Q$, but also with $0.2\,Q < |\boldsymbol{q}_T| < Q$. In Sec. 3.3, some variations of the baseline SIDIS cut in Eq. (3.2) will be discussed to give phenomenological support to our choice.

For what concerns the data sets included in the present analysis, the main

---

[2]This means that we can neglect target-mass corrections and higher twist contributions.

difference with Ref. [87] is that we considered the new release of COMPASS data [141]. In this new data analysis that superseeds the previous one, the vector–boson contributions to the fragmentation of observed hadrons have been subtracted. In the case of the HERMES data set, we consistently selected the vector–meson–subtracted data set (`.vmsub` set). Moreover, we chose to include the `zxpt-3D`-binning for HERMES multiplicities, since it provides a finer binning in $|\boldsymbol{P}_{hT}|$. The breakdown of the entire SIDIS data set included in the present analysis is reported in Tab. 3.3.

Experiment	N_{dat}	Observable	Channels	Q [GeV]	x	z	Ref.
HERMES	344	$M(x, z, |\boldsymbol{P}_{hT}|, Q)$	$p \to \pi^+$ $p \to \pi^-$ $p \to K^+$ $p \to K^-$ $d \to \pi^+$ $d \to \pi^-$ $d \to K^+$ $d \to K^-$	$1 - \sqrt{15}$	$0.023 - 0.6$ (6 bins)	$0.1 - 1.1$ (8 bins)	[142]
COMPASS	1203	$M(x, z, \boldsymbol{P}_{hT}^2, Q)$	$d \to h^+$ $d \to h^-$	$1 - 9$ (5 bins)	$0.003 - 0.4$ (8 bins)	$0.2 - 0.8$ (4 bins)	[141]
Total	1547						

Table 3.3: Breakdown of the SIDIS data sets included in the MAP22 analysis. For each data set, the table includes information on: the number of data points (N_{dat}) surviving the nominal cut on $|\boldsymbol{P}_{hT}|$, the measured observable, the SIDIS fragmentation channel, the range(s) in photon invariant mass Q, the ranges in the kinematic variables x and z, and the public reference. The total number of SIDIS data points amounts to 1547.

In the second column of Tab. 3.3, for each data set the number of experimental data (N_{dat}) that respect the aforementioned kinematic cuts is shown. The total number of SIDIS experimental data points included in the analysis illustrated in this chapter is 1547.

In conclusion, the total number of DY and SIDIS data points surviving our kinematic cuts is 2031.

3.1.3 Error treatment

For the considered experimental measurements, a set of systematic and statistical uncertainties is associated to each data point. As already discussed in Ref. [81], it is important to properly treat the experimental uncertainties in order to obtain a reliable result of a global fit and, consequently, of the extracted TMDs. In the analysis described in this chapter, we choose to consider systematic uncertainties as fully correlated only if it is explicitly specified in

the corresponding published paper. Instead, the statistical uncertainties are always considered as uncorrelated.

In order to properly treat fully correlated uncertainties, the iterative t_0-prescription [215] was introduced in Ref. [81]. This prescription usually plays a significant role to avoid the underestimation of the predictions caused by the so–called D'Agostini bias [216]. At variance with Ref. [81], in the analysis presented in this chapter we do not make use of such prescription because our experimental data do not show any sensitivity to it. In fact, after performing the fit with and without the t_0-prescription, we observed that our analysis is not affected by the D'Agostini bias. Therefore, we see no reason to introduce the t_0-prescription in the computation of the χ^2.

In addition to the systematic experimental uncertainties, one has to carefully treat all the possible sources of systematic theoretical errors. A first source of error is given by the intrinsic uncertainty of the collinear parton distributions (fragmentation functions) needed to build the TMD PDFs (FFs) to be extracted by our global analysis (see Eq. (2.52)). In our case, we choose the PDFs set MMHT2014 [210] and the FFs set DSS. These choices are comparable with the ones in Refs. [81, 88]. Since the HERMES collaboration provides SIDIS multiplicities for identified charged pions and kaons separately (see Tab. 3.3), we consider DSS14 [217] for $\pi^\pm$ and DSS17 [32] for $K^\pm$. According to the nature of the PDF and FF set used in the MAP22 analysis, their uncertainties are computed using the Hessian method [217–219] and are included in the χ^2 calculation as systematic errors. It is difficult to evaluate the degree of correlation of such theoretical errors between the bins in $|\boldsymbol{q}_T|$ of a given data set. A crude estimate could be obtained by making predictions of a given experimental observable with different replicas of the considered collinear input set. This procedure has to be independent of the nonperturbative model. Therefore, all the predictions are calculated with a parameterization that is kept fixed for all the replicas. Since the MMHT2014 is composed of Hessian members, we do the exercise with the NNPDF3.1 [220] set of PDFs at NNLO, which is composed by 100 Monte Carlo (MC) replicas. In the left panel of Fig. 3.2, we display the ratio between the predictions of a subset of replicas and the replica 0 (the mean over the ensemble) of the NNPDF3.1 for the first $|y|$-bin of the ATLAS 7 TeV data set. In the right panel, we show the correlation matrix for each $|\boldsymbol{q}_T|$-bin of the considered replicas, identified by mathix labels.

We observe that PDF uncertainties are significantly correlated across bins, but not always 100% correlated. The same outcome holds also for collinear FFs. In order to account for this correlation, we decomposed the corresponding errors into a fully correlated part (80% of the total), while we treated the remaining 60% as uncorrelated.[3]

Moreover, a collinear set of FFs for unidentified charged hadrons is needed

---

[3]Notice that, with this decomposition, the sum in quadrature of correlated and uncorrelated parts reproduce the original errors.
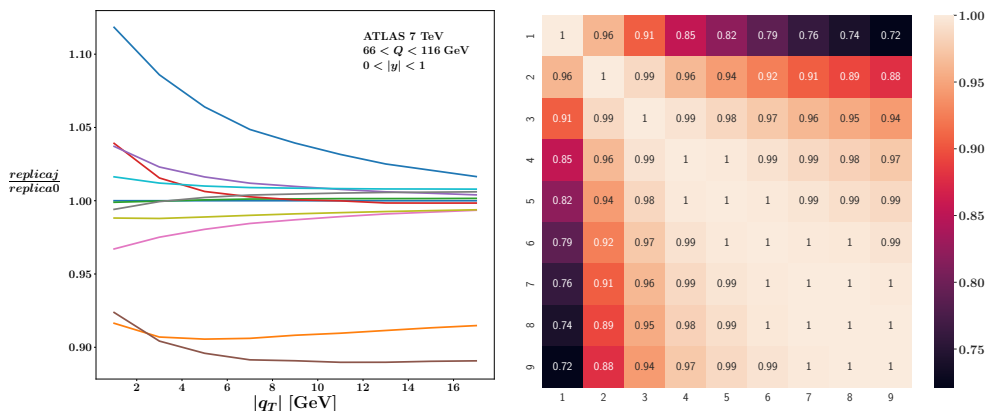
Figure 3.2: Left panel: ratio between the predictions of the DY differential cross section for one of the ATLAS data sets of a given replica of the NNPDF3.1 collinear PDF set and the mean replica (member 0). Right panel: correlation matrix of the PDF replicas in each $|\boldsymbol{q}_T|$-bin.

to reproduce the SIDIS observables measured by COMPASS. The most recent extraction of collinear FFs for unidentified charged hadron is the DSS07 [221] set. However, such extraction does not provide an estimate of the uncertainties. Therefore, we computed the COMPASS multiplicities by using the sum of the DSS14 and DSS17 sets for pions and kaons.[4] The related error is calculated by propagating to the unidentified charged hadron multiplicity the Hessian errors associated to each of the two hadronic components.

A second source of theoretical uncertainties is given by the choice of a specific set for the collinear distributions. In fact, as already pointed out in Ref. [88], the choice of a given set may have a sizeable impact on the final result. In the MAP22 analysis, we did non consider other collinear sets due to the hard computational effort needed. We discuss this topic in Ch. 5. Similarly, we leave for furure studies the investigation of other sources of theoretical uncertainties such as higher-twist corrections and the choice of perturbative scales. The effects of the introduction of TMD flavor dependence will be addressed in future studies.

3.2 Phenomenological results

In this section, we will show the phenomenological results of the simultaneous extraction of proton TMD PDFs and TMD FFs through the global fit of the SIDIS and DY experimental data reported in Tabs. 3.1-3.3. In order to determine the accuracy of this analysis, we use the convention for the logarithmic

---

[4]According to Ref. [88], we assumed that the yield due to other hadronic species, such as protons, antiprotons, Λ, etc., is negligible as compared to the sum of kaons and pions. As discussed in Ref. [222], the contribution to the total yield due to hadrons heavier that pions and kaons is indeed marginal (about 5%).

ordering discussed in Sec. 2.3.3. At the time of this work, the full N³LL accuracy could not be achieved, since sets of fragmentation functions in collinear framework at NNLO were not available. Recently, two analysis of collinear FFs have been discussed in Refs. [33, 190], making a simultaneous fit of TMD PDFs and FFs at full N³LL possible. This study is left for a future work.

3.2.1 Nonperturbative models of TMDs

As already discussed in Sec. 2.3.4, a nonperturbative function has to be introduced in the expression of a TMD distribution to account for the power corrections induced by the b_* prescription. Early attempts to describe unpolarized observables where made in parton model-like approximation to TMDs (see Ref. [166] and references therein). Here, we illustrate the choices of Ref. [1] for f_{NP} (and D_{NP} for TMD FFs) in Eq. (2.77). We stress that both the Collins–Soper kernel K [143] and the OPE in Eq. (2.52) need to be modified to account for nonperturbative effects. For the Collins–Soper kernel K, this results in a nonperturbative correction term, named $g_K(\boldsymbol{b}_T^2)$. For such new ingredient, we choose a specific functional form:

$$K(|\boldsymbol{b}_T|, \mu_{b_*}) = K(b_*, \mu_{b_*}) + g_K(\boldsymbol{b}_T^2), \qquad g_K(\boldsymbol{b}_T^2) = -g_2^2 \frac{\boldsymbol{b}_T^2}{2}, \qquad (3.3)$$

where g_2 is a free parameter. As commonly known, this corrections add to TMD evolution a nonperturbative term that scales like $(\zeta_f/Q_0^2)^{g_K/2}$, where Q_0 is an arbitrary scale at which the nonperturbative model is parametrized. We set $Q_0 = 1$ GeV. Since at small $|\boldsymbol{b}_T|$ the perturbative calculation of the CS kernel $K(b_*, \mu_{b_*})$ must not be affected by this new nonperturbative term, the function g_K needs to vanish in the limit $|\boldsymbol{b}_T| \to 0$. The choice of the quadratic $|\boldsymbol{b}_T|$ dependence for the $g_K(\boldsymbol{b}_T^2)$ function takes inspiration from previous works focused either in the phenomenology [74, 77, 87, 88, 223, 224] or in the theory side [143].

The nonperturbative corrections to the OPE can be parameterized by a multiplicative function that generally depends on x or z and $\boldsymbol{b}_T$. The net result of the inclusion of the nonperturbative corrections into the evolved TMD PDF reads:

$$
\begin{aligned}
\hat{f}_1^a(x, \boldsymbol{b}_T^2; \mu_f, \zeta_f) = & [C \otimes f_1](x, b_*; \mu_{b_*}, \mu_{b_*}^2) \exp\left\{ \int_{\mu_{b_*}}^{\mu_f} \frac{d\mu}{\mu} \gamma(\mu, \zeta_f) \right\} \\
& \times \left(\frac{\zeta_f}{\mu_{b_*}^2} \right)^{K(b_*, \mu_{b_*})/2} f_{1\,NP}(x, \boldsymbol{b}_T^2; \zeta_f, Q_0),
\end{aligned}
\qquad (3.4)
$$

and the same holds for the TMD FF, where the nonperturbative function $D_{1\,NP}(z, \boldsymbol{b}_T^2; \zeta, Q_0)$ is introduced. In our definition, the $f_{1\,NP}$ and $D_{1\,NP}$ factors include both the correction to the evolution provided by the g_K function of Eq. (3.3) and the correction to the associated OPE. Following the requirements

discussed in Sec. 2.3.4, we define $f_{1\,NP}$ as

$$
f_{1\,NP}(x, \boldsymbol{b}_T^2; \zeta, Q_0) = \left[\frac{\zeta}{Q_0^2}\right]^{g_K(\boldsymbol{b}_T^2)/2} \frac{1}{g_1(x) + \lambda^2\, g_{1B}^2(x) + \lambda_2^2\, g_{1C}(x)}
$$
$$
\times \left\{ g_1(x)\, e^{-g_1(x)\frac{\boldsymbol{b}_T^2}{4}} + \lambda^2\, g_{1B}^2(x) \left[1 - g_{1B}(x)\frac{\boldsymbol{b}_T^2}{4}\right] e^{-g_{1B}(x)\frac{\boldsymbol{b}_T^2}{4}} \right.
$$
$$
\left. + \lambda_2^2\, g_{1C}(x)\, e^{-g_{1C}(x)\frac{\boldsymbol{b}_T^2}{4}} \right\},
$$
(3.5)

and for the TMD FF the $D_{1\,NP}$ is defined as

$$
D_{1\,NP}(z, \boldsymbol{b}_T^2; \zeta, Q_0) = \left[\frac{\zeta}{Q_0^2}\right]^{g_K(\boldsymbol{b}_T^2)/2} \frac{1}{g_3(z) + \frac{\lambda_F}{z^2}\, g_{3B}^2(z)}
$$
$$
\times \left\{ g_3(z) e^{-g_3(z)\frac{\boldsymbol{b}_T^2}{4z^2}} + \frac{\lambda_F}{z^2} g_{3B}^2(z) \left[1 - g_{3B}(z)\frac{\boldsymbol{b}_T^2}{4z^2}\right] e^{-g_{3B}(z)\frac{\boldsymbol{b}_T^2}{4z^2}} \right\}.
$$
(3.6)

Notice that λ, λ_2, λ_F are free parameters. We stress that the nonperturbative factors $f_{1\,NP}$, $D_{1\,NP} \to 1$ for $|\boldsymbol{b}_T| \to 0$. This is coherent with the definition of the Fourier transform of Eq.(2.9). Consequently, the integral of the nonperturbative model in $|\boldsymbol{k}_\perp|$-space is equal to 1: it affects only the shape in transverse-momentum and not the normalization. The $g_i(x)$ functions describe the widths of the distributions. According to model calculations (see Ref. [225] and references therein) and more generally to Lorentz-invariance constraints on the proton light-front wave functions (see, *e.g.*, the discussion in Ref. [226]), we assume that these widths depend on the variable x and z as

$$
g_{\{1,1B,1C\}}(x) = N_{\{1,1B,1C\}} \frac{x^{\sigma_{\{1,2,3\}}}(1-x)^{\alpha_{\{1,2,3\}}^2}}{\hat{x}^{\sigma_{\{1,2,3\}}}(1-\hat{x})^{\alpha_{\{1,2,3\}}^2}}, \tag{3.7}
$$

$$
g_{\{3,3B\}}(z) = N_{\{3,3B\}} \frac{(z^{\beta_{\{1,2\}}} + \delta_{\{1,2\}}^2)(1-z)^{\gamma_{\{1,2\}}^2}}{(\hat{z}^{\beta_{\{1,2\}}} + \delta_{\{1,2\}}^2)(1-\hat{z})^{\gamma_{\{1,2\}}^2}}, \tag{3.8}
$$

where $\hat{x} = 0.1$, $\hat{z} = 0.5$ and N_i, σ_i, α_i, β_i, δ_i, γ_i are free parameters.

In total, the models considered in the global fit of this analysis involves 21 free parameters to be fitted to more than two-thousand experimental data. They are grouped as follows: 11 are related to the nonperturbative part of the TMD PDF (λ, λ_2, N_i, σ_i, α_i in Eqs. (3.5), (3.7)), 9 for the nonperturbative part of the TMD FF (λ_F, N_i, β_i, δ_i, γ_i in Eqs. (3.6), (3.8)), and one associated to the nonperturbative part of the Collins–Soper kernel (g_2 in Eq. (3.3)).

We stress that the functional forms in Eqs. (3.5)-(3.8) are phenomenological ansätze, namely are arbitrary. However, they naturally correspond to the Fourier transform of the sum of a Gaussian, a weighted-Gaussian (multiplied

by $\boldsymbol{k}_\perp^2$) and, in the case of the TMD PDFs, of a third Gaussian. Therefore, they are positive definite for all values of the transverse momentum ($|\boldsymbol{k}_\perp|$ for TMD PDFs and $|\boldsymbol{P}_\perp|$ for TMD FFs).[5]

In order to avoid negative contributions in $|\boldsymbol{k}_\perp|$-space, the parameters λ and λ_2 in Eq. (3.5) are squared. In the case of TMD FFs, we do not find the necessity to square the parameter λ_F in Eq. (3.6) because the fit always favors positive values for this parameter. As can be seen in Eqs. (3.7), (3.8), the widths of the Gaussians and the weighted-Gaussian are $x-$ (or $z-$) dependent and vanish as x (or z) approaches 1. The choice is inspired by early models introduced in global fits of PDFs (or FFs) in collinear framework. In principle, the parameters of our phenomenological model may depend on the partonic flavor q. However, in throughout this thesis they are assumed to be flavor-independent, namely the nonperturbative transverse-momentum shape of TMDs is assumed to be the same for each flavor. The investigation of the sensitivity of the experimental data on the partonic flavor is considered as a future extension of this work. We stress that the fact that we can obtain a good description of the experimental data sets included in the analysis described in this chapter does not exclude the possibility of performing a successful fit with flavor-dependence, which is actually expected according to model calculations [227–232], lattice QCD analyses [233], and also if QED corrections are taken into account [139, 234]. The possibility of the inclusion of flavor dependence on the phenomenological analysis of SIDIS data was studied in Ref. [84]. Recently, flavor-dependent parameterizations have been tested on Drell–Yan data sets in Refs. [82, 83].

Our choice of the functional form of the whole nonperturbative function takes also inspiration by model calculations of TMD PDFs (see, e.g., [140, 225, 227, 235–239]) and TMD FFs (see, e.g, [231, 240]). Many of these models give a justification for the existence of terms that behave as Gaussians and weighted Gaussians. The details of the functional dependence predicted by the models are related to the correlation between the transverse momentum of the quarks and their spin. In the case of fragmentaton functions, a different role can be played by different production channels (*e.g.*, direct production from the fragmentation of the active quark or the production through the decay of hadronic resonances).

3.2.2 Fit quality

In this section, we discuss the quality of the MAP22 baseline fit at $\mathrm{N^3LL^-}$ imposing the cuts on kinematics as discussed in Sec. 3.1. The propagation of the fit errors is performed through the so-called bootstrap method, which consists in fitting an ensemble of Monte Carlo (MC) replicas of the experimental data (see Ref. [87] for more details). We choose the number of replicas by requiring

---

[5]Note, however, that when evolved at the scale Q a TMD distribution can become negative at large values of transverse momentum due to evolution effects.) .

that the mean and standard deviation of the ensemble accurately reproduces the original data points. In this case, we find that 250 replicas are sufficient.

Clearly, the most complete statistical information about the extracted TMDs is given by the full ensemble of 250 replicas. However, it is useful to define a single representative replica over the full set in order to quantify the agreement between the fitted theoretical predictions and the experimental data. In order to estimate the quality of our fit, we find that the most appropriate indicator is the the χ^2 value of the best fit to the experimental unreplicated data (χ_0^2). We refer to this fit as the "central replica". The ensemble of replicas is therefore employed only to identify the errors associated to the prediction for the experimental observable and the extracted TMD distributions, as required when applying a general bootstrap technique.

Other possible indicators of the quality of the fit are the average of the χ^2 over all replicas, indicated as $\langle \chi^2 \rangle$, as well as the χ^2 of the mean replica, indicated as χ_m^2 and computed as the χ^2 of the average over the full ensemble of TMD replicas [81]. In a stable fit, these values should be very close to each other.

In Fig. 3.3, we show the χ^2 values per number of data N_{dat} (χ^2/N_{dat}) of the 250 MC replicas of the MAP22 global fit. As expected, the shape of the histogram closely resembles a χ^2 distribution. The observation that the distribution exhibits a peak at a value extremely close to 1 serves as an indication of the good quality of the our baseline fit.
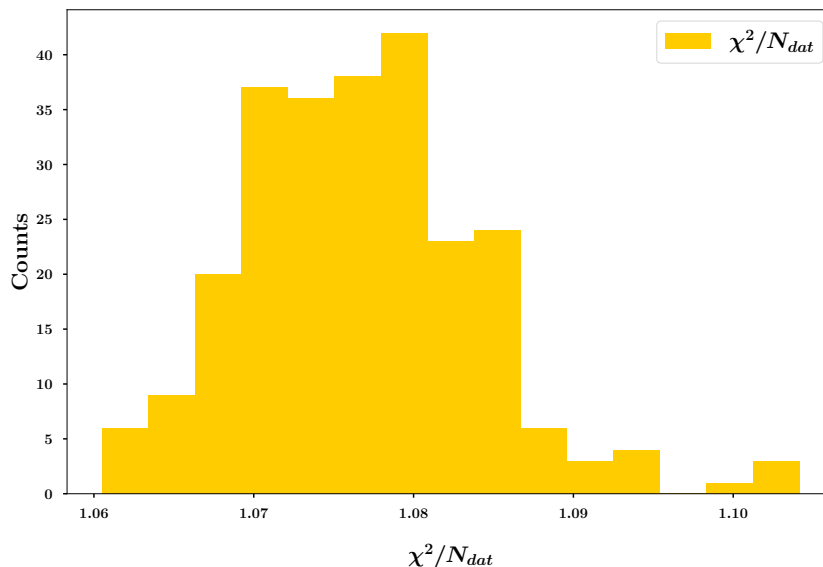

Figure 3.3: Statistical distribution of the χ^2 values of the 250 Monte Carlo replicas of the MAP22 global fit.

In Tabs. 3.4-3.5, we report the breakdown of the χ_0^2 values normalized to the number of data points (N_{dat}) for DY and SIDIS data sets.

As already discussed in Ref. [81], in the presence of correlated uncertainties

		N^{3}LL$^-$		
Data set	N_{dat}	χ^2_D	χ^2_λ	χ^2_0
CDF Run I	25	0.45	0.09	0.54
CDF Run II	26	0.995	0.004	1.0
D0 Run I	12	0.67	0.01	0.68
D0 Run II	5	0.89	0.21	1.10
D0 Run II (μ)	3	3.96	0.28	4.2
Tevatron total	71	0.87	0.06	0.93
LHCb 7 TeV	7	1.24	0.49	1.73
LHCb 8 TeV	7	0.78	0.36	1.14
LHCb 13 TeV	7	1.42	0.06	1.48
LHCb total	21	1.15	0.3	1.45
ATLAS 7 TeV	18	6.43	0.92	7.35
ATLAS 8 TeV	48	3.7	0.32	4.02
ATLAS 13 TeV	6	5.9	0.5	6.4
ATLAS total	72	4.56	0.48	5.05
CMS 7 TeV	4	2.21	0.10	2.31
CMS 8 TeV	4	1.938	0.001	1.94
CMS 13 TeV	70	0.36	0.02	0.37
CMS total	78	0.53	0.02	0.55
PHENIX 200	2	2.21	0.88	3.08
STAR 510	7	1.05	0.10	1.15
DY collider total	251	1.86	0.2	2.06
E288 200 GeV	30	0.35	0.19	0.54
E288 300 GeV	39	0.33	0.09	0.42
E288 400 GeV	61	0.5	0.11	0.61
E772	53	1.52	1.03	2.56
E605	50	1.26	0.44	1.7
DY fixed-target total	233	0.85	0.4	1.24
HERMES	344	0.48	0.23	0.71
COMPASS total	1203	0.62	0.3	0.92
SIDIS total	1547	0.59	0.28	0.87
Total	**2031**	**0.77**	**0.29**	**1.06**

Table 3.4: Breakdown of the values of χ^2 normalized to the number of data points N_{dat} that survive the kinematic cuts for all data sets considered in our baseline fit. The χ^2_D refers to uncorrelated uncertainties, χ^2_λ is the penalty term due to correlated uncertainties (see Eq. (3.9)), χ^2_0 is the sum of χ^2_D and χ^2_λ. All χ^2 values refer to the central replica (see text).

on different bins of the experimental data the total χ^2 can be expressed as the

		N³LL⁻		
Data set	N_{dat}	χ^2_D	χ^2_λ	χ^2_0
HERMES ($p \to \pi^+$)	45	0.86	0.42	1.28
HERMES ($p \to \pi^-$)	45	0.61	0.31	0.92
HERMES ($p \to K^+$)	45	0.49	0.04	0.53
HERMES ($p \to K^-$)	37	0.18	0.13	0.31
HERMES ($d \to \pi^+$)	41	0.68	0.45	1.13
HERMES ($d \to \pi^-$)	45	0.63	0.35	0.97
HERMES ($d \to K^+$)	45	0.2	0.02	0.22
HERMES ($d \to K^-$)	41	0.14	0.08	0.22
HERMES total	344	0.48	0.23	0.71
COMPASS ($d \to h^+$)	602	0.55	0.31	0.86
COMPASS ($d \to h^-$)	601	0.68	0.3	0.98
COMPASS total	1203	0.62	0.3	0.92
SIDIS total	1547	0.59	0.28	0.87

Table 3.5: Breakdown of the values of χ^2 normalized to the number of data points N_{dat} that survive the kinematic cuts for SIDIS data sets considered in our baseline fit. Same notations as in previous table.

sum of two separate contributions

$$\chi^2 = \sum_i^N \left(\frac{d_i - \bar{t}_i}{\sigma_i} \right)^2 + \chi^2_\lambda = \chi^2_D + \chi^2_\lambda , \qquad (3.9)$$

where the so-called partial chi-squared χ^2_D is given by the standard formula for N experimental data points d_i and all possible sources of uncorrelated (including statistical) uncertainties $\sigma_i^2 = \sigma^2_{i,\text{stat}} + \sigma^2_{i,\text{uncor}}$, but involving shifted theoretical predictions $\bar{t}_i$ due to the correlated uncertainties,

$$\bar{t}_i = t_i + \sum_{\alpha=1}^k \lambda_\alpha \, \sigma^{(\alpha)}_{i,\text{corr}} , \qquad (3.10)$$

where $\sigma^{(\alpha)}_{i,\text{corr}}$ is the value of the α-th source of correlated uncertainty associated to the i-th experimental data point and λ_α is the so-called nuisance parameter. In Eq. (3.9), the additional term χ^2_λ is a penalty contribution due to the presence of correlated uncertainties. It is completely determined by the nuisance parameters,

$$\chi^2_\lambda = \sum_{\alpha=1}^k \lambda_\alpha^2 . \qquad (3.11)$$

An interesting feature of the nuisance parameters λ_α is that their optimal value

can be calculated by minimizing the total χ^2 in Eq. (3.9) with respect to them.

We stress that this is only an alternative expression of the χ^2, and it can be demonstrated that Eq. (3.9) is equivalent to the general definition involving the covariance matrix [241]. In the following, we will display shifted predictions in Eq. (3.10) for all observables included in the fit because they provide a better visualization of the agreement between theory predictions and experimental data.

As can be seen in Tab. 3.4, the global χ_0^2 of the MAP22 baseline fit value is 1.06, indicating that the description of the included global data set is very good.[6] This result suggests that the formalism involved in MAP22 fit is able to simultaneously describe experimental data coming from two different processes over a wide kinematic range. This is a phenomenological demonstration of the universality of TMDs. As can be seen in Tabs. 3.1-3.3, the data set at low-energy scale is composed by fixed-target DY experiments and SIDIS observables, while the high-energy data set comes from collider experiments at the LHC and Tevatron.

Although the penalty term χ_λ^2 is expected to be of modest size, we notice that sometimes it is more than 25% of the total χ_0^2. This is due to the fact that the size of the shifts produced by correlated uncertainties is often large. The value of $\chi_\lambda^2/N_{\mathrm{dat}}$ obtained in the MAP22 analysis is larger than the one in the PV19 [81], mainly because the theoretical uncertainties related to collinear PDFs and FFs are considered here as 80% correlated (see Sec. 3.1), while they were considered as fully uncorrelated in the PV19 fit.

From Tab. 3.5, we can see that for HERMES and COMPASS multiplicities the values of $\chi_0^2/N_{\mathrm{dat}}$ are almost always smaller than 1, indicating that these data sets, which represent about 75% of the total number of data points considered in the MAP22 analysis (1547 over 2031) are very well described.

We note that the largest contribution to the χ^2 for HERMES multiplicities comes from the channel where a positive pion (π^+) is detected in the final state. This happens for both proton and deuteron targets, and it is consistent with the results of both Refs. [84, 87] and [88].

In the case of identified charged kaons in the final state, the χ^2 is lower than the one of pions. Such result is a consequence of the large statistical errors affecting the measurement of kaon multiplicities and by the large theoretical error due to the uncertainties in the extraction of kaon collinear FFs (sometimes of the order of 50%).

The comparison between theoretical results for the SIDIS multiplicities of Eq. (2.34) and HERMES data for the production of charged pions and kaons off a deuteron target is shown in Fig. 3.4. A specific x-bin is displayed in each column, while a specific final-state channel in each row. The results are reported as functions of the transverse momentum $|\boldsymbol{P}_{hT}|$ of the measured final-state hadron. For a better visualization of different representative z-bins,

---

[6]The χ^2 related to the mean replica and the average over the full ensemble are, respectively, $\chi_m^2 = 1.07$ and $\langle\chi^2\rangle = 1.08 \pm 0.01$.

points with different markers and colors are displayed. For the same reason, different offsets are introduced as indicated in the plot legend. The theoretical results of the MAP22 fit are represented as light blue rectangles with size corresponding to the 68% Confidence-Level (C.L.) band.[7]

We note that the central column, corresponding to the $0.2 < x < 0.35$ bin, does not include the magenta points for the highest $0.6 < z < 0.8$ bin in the K^- channel (second row) because they are excluded by the kinematic cut in z. Moreover, we can see that in all panels there are only three green points for the lowest z bin, while four for the other bins (orange and magenta points). This is caused by the z-dependent cut on transverse momentum in Eq. (3.2), which leads to the exclusion of a larger number of $|\boldsymbol{q}_T|$-bins at lower values of z.

In Fig. 3.5, we display the HERMES multiplicities with same conventions and notation as in Fig. 3.4 but off a proton target. We remark that for the negative kaon channel the kinematic cuts have a more drastic effect. In fact, the magenta points for the $|\boldsymbol{P}_{hT}|$-distributions at the largest z-bin are excluded for the two largest x-bins considered (central and rightmost panels of second row from top).

In Fig. 3.6, the result of the MAP22 fit for the COMPASS SIDIS multiplicities for the production of unidentified negatively charged hadrons off a deuteron target is shown. Each panel, which corresponds to a specific Q and x bin, displays the multiplicity as a function of $\boldsymbol{P}_{hT}^2/Q^2$ on a logarithmic scale. As for the plots of HERMES multiplicities, different markers and colors are introduced to display the data points of different representative z bins, as indicated in the plot legend. Again, the light-blue rectangles correspond to the 68% CL theoretical results. As already discussed in Sec. 3.1, we remark that the results for unidentified negatively charged hadrons h^- are obtained by simply adding the results for negatively charged pions and kaons, $h^- \sim \pi^- + K^-$.

We note that the quality of the MAP22 fit is very nice for almost all bins, which is also denoted by small χ^2 values in Tab. 3.5. The agreement worsens for the lowest Q-bin ($1.3 < Q < 1.73$ GeV), particularly for $x \gtrsim 0.02$. This could be due to the fact that such data set can be considered at the boundary of both the TMD and the collinear factorization regions, at least according to the kinematical cuts introduced in Sec. 3.3. Also in this case, the uncertainties on the theoretical predictions for the largest z-bin included in the fit ($0.6 < z < 0.8$) are significantly larger than for the others because of much larger uncertainties in the collinear FFs. Finally, if we look at the panels in Fig. 3.6 from top to bottom, we note that the lowest z-bin distributions ($0.2 < z < 0.3$) are included in the fit only for the largest Q-bins. This is a consequence of the kinematic cut in Eq. (3.2).

In Fig. 3.7 the COMPASS multiplicities with same conventions and notation as in Fig. 3.6, but for unidentified positively charged hadrons h^+, are reported. Again, theoretical results at 68% C.L. are displayed as light-blue

---

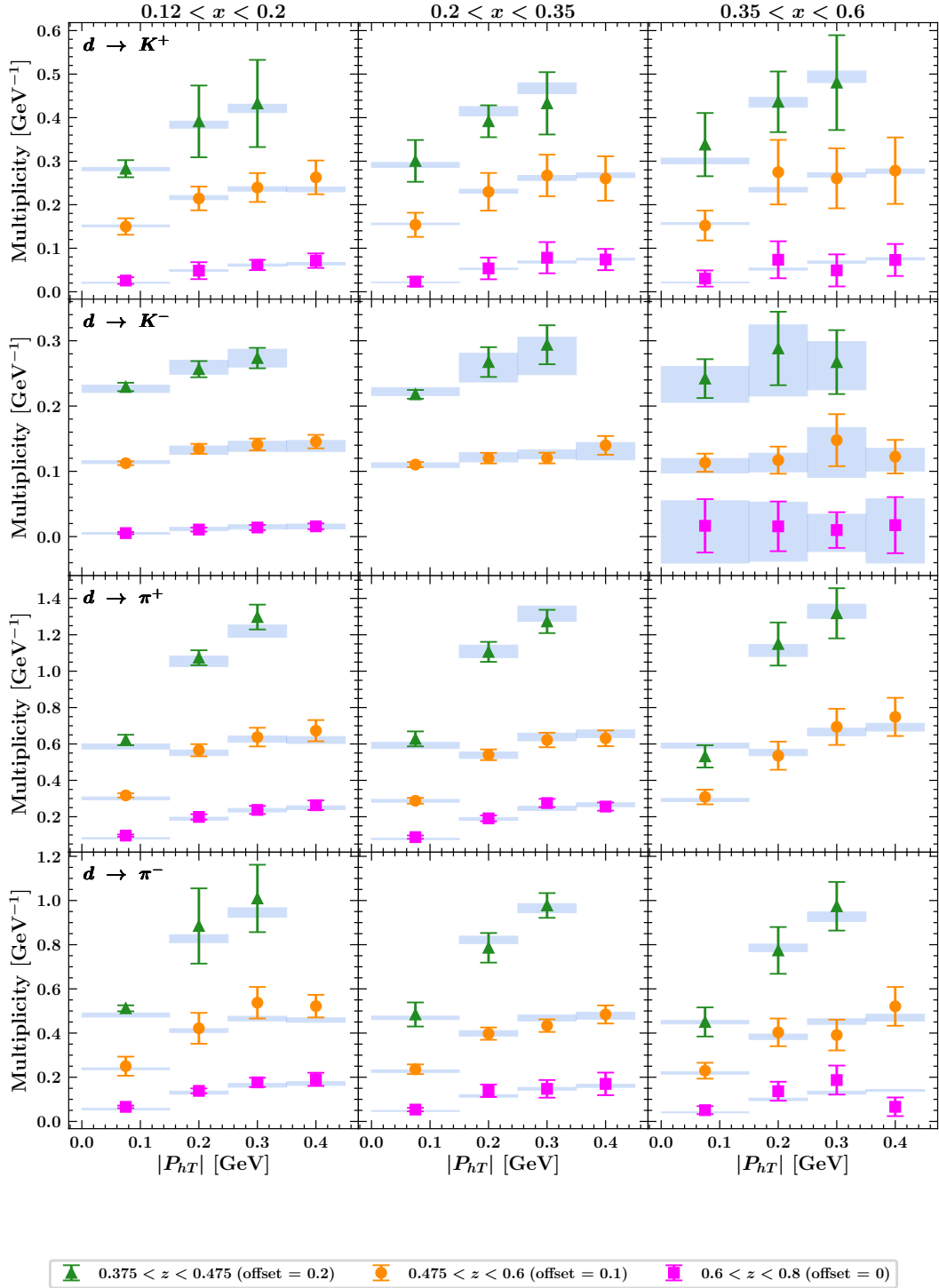[7]The 68% C.L. band is built by excluding the largest and the smallest 16% of the replicas.

Figure 3.4: Comparison between data for the HERMES multiplicities and theoretical predictions (rectangles with size proportional to the uncertainty at the 68% C.L.) for the production of charged pions and kaons off a deuteron target for different x and z bins as a function of the transverse momentum $|\boldsymbol{P}_{hT}|$ of the final-state hadron. For better visualization, each z bin is shifted by the indicated offset.
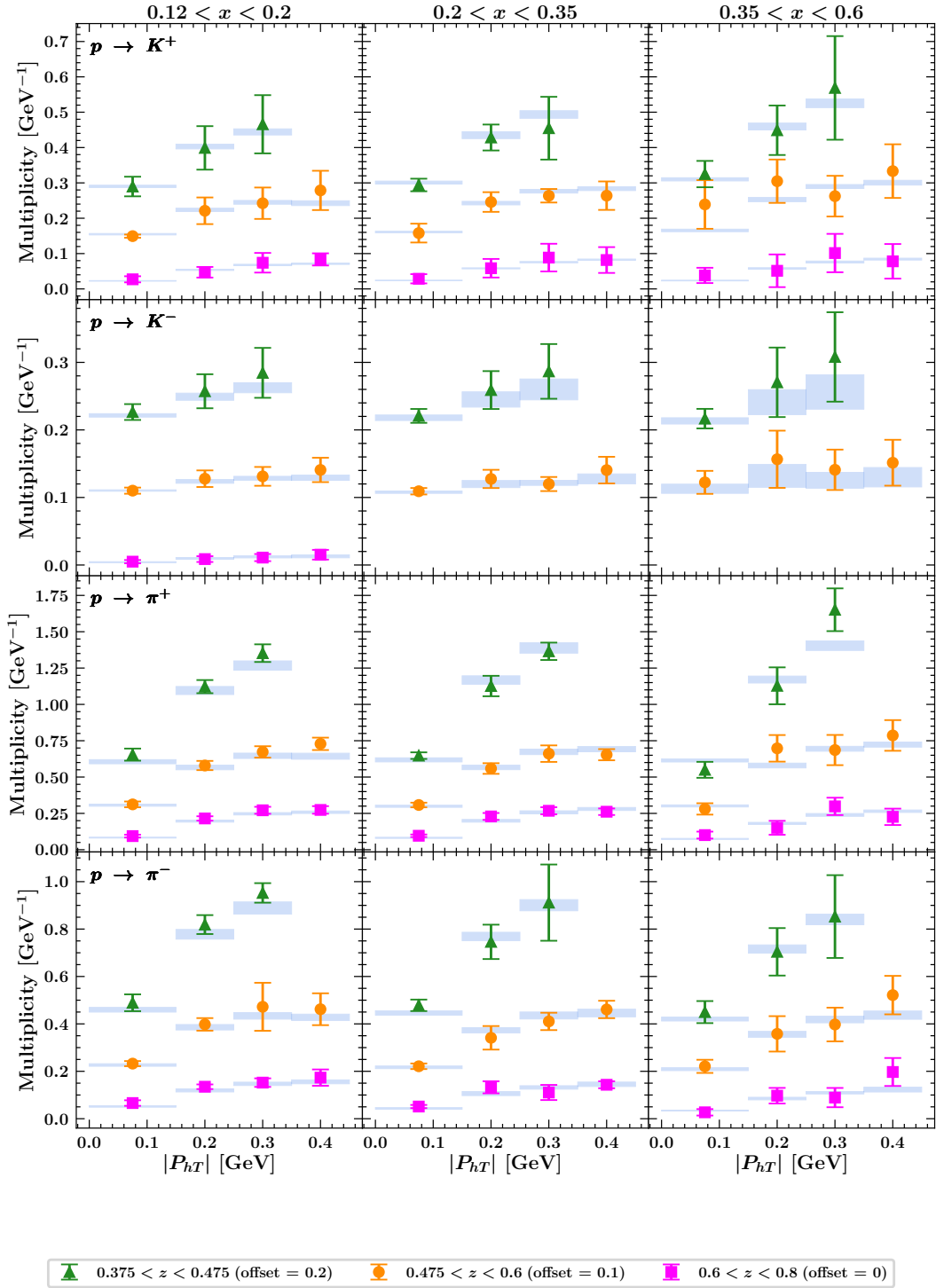
67

Figure 3.5: Same conventions and notation as in previous figure but for charged pions and kaons off proton target.

Figure 3.6: Comparison between data for the COMPASS multiplicities and theoretical predictions rectangles with size proportional to the uncertainty at the 68% C.L. for the production of negatively charged hadrons off a deuteron target. For each Q, x bin, the multiplicities are displayed as functions of $\boldsymbol{P}_{hT}^2/Q^2$ for different z bins surviving kinematic cuts, as indicated in the legend.

Figure 3.7: Same conventions and notation as in previous figure but for uniden-
tified positively charged hadrons off deuteron target.

rectangles, and are obtained as the sum of the results for positively charged pions and kaons, $h^+ \sim \pi^+ + K^+$. Similar comments as for Fig. 3.6 can be made about the agreement between data and theory.

As can be seen in Tab. 3.1, DY data represents approximately 25% of the full set of analyzed data. From Tab. 3.4, we can see that most of low-energy DY data from fixed-target experiments (E605, E288, E772), but also from STAR, can be fitted with good accuracy (low χ^2 values), much better than high-energy DY data from collider experiments like, *e.g.*, at the LHC. As already discussed in Ref. [81], this difference in the level of agreement of the description most likely arises from the larger (systematic) errors affecting low-energy DY data and the big impact of the uncertainties on collinear PDFs at these kinematics.

We point out that the quality of our fit for the ATLAS data sets is poor (see Tab. 3.4). In particular, the value of the χ^2 becomes larger and larger for low-rapidity bins of both ATLAS 7 TeV and ATLAS 8 TeV data sets, the worst case being at $|y| < 1$ for ATLAS 7 TeV. Different effects could be responsible for this result. In fact, since the experimental observable is a differential cross section normalized over the total one, systematic errors cancel in the ratio, leading to extremely precise measurements (errors of the order of 0.1-1 ‰). Therefore, fitting this experimental data is very difficult, because even small theoretical effects could give significant contributions to the χ^2 value. Moreover, it is known that different implementations of phase-space cuts on the final-state leptons could lead to modifications in both the shape and the normalization of the theoretical observable (see, *e.g.*, Refs. [126, 242, 243]). We leave this issue for a future study. In any case, we stress that, at variance with Ref. [88], we obtain our result without excluding any extra points on top of the ones exceeding the maximum value of $|\boldsymbol{q}_T|/Q$ in Eq. (3.1). In Ref. [88], a more restrictive cut was implemented only for the most precise ATLAS experimental data. Surprisingly, we observe that this issue at low rapidity is not present for the CMS data set at 13 TeV, which is splitted in different rapidity bins as the ATLAS set.

It is important to comment the results of our baseline fit for those data sets that were not included in the previous PV19 analysis of Ref. [81] (see Sec. 3.1.1). For the E772 data set, the quality of the description is good only for data points above the peak of the Υ resonance. For $Q < 9$ GeV, the description of the experimental data worsens. At variance with Ref. [88], the $Q < 9$ GeV bins are still included in our analysis because there is no evident motivation to exclude them. The new CMS data set at $\sqrt{s} = 13$ TeV is very nicely described, even better than the CMS sets at smaller center-of-mass energies $\sqrt{s} = 7, 8$ TeV. This is probably due to the fact that the CMS 13 TeV data set is binned in rapidity. It is important to fit this new data sets because they extend the kinematic coverage in our analysis.

In order to better visualize the quality of our fit, the comparison between experimental data and theoretical results for a representative selection of the

DY data set is reported in Figs. 3.8-3.11. In the upper panels of each plot, we display the $|\boldsymbol{q}_T|$-differential cross section, while in the lower panels we show the ratio between data and theory. As for the SIDIS case, the light-blue bands are the theoretical results at 68% C.L.

In Fig. 3.8, the DY cross section for the E288 data set at beam energy $E_{beam} = 200$ GeV is displayed for different bins of the energy scale Q. Since the experimental collaboration did not release the $|\boldsymbol{q}_T|$-bin limits for DY fixed-target observables, we calculate the cross section at mean values of $|\boldsymbol{q}_T|$ (see discussion in Sec. 2.1). Hence, we display the 68% C.L. uncertainty of the fit as a continuous band rather than a series of rectangles. We notice that the uncertainty band is larger for lower Q bins. This trend is caused by the presence of larger correlated systematic uncertainties for smaller invariant masses of the lepton pair.

Fig. 3.9 displays our fit results for the cross section for DY in $p\bar{p}$ collisions at the Tevatron. Black data points in the left panel correspond to the results of Run I of the CDF experiment, while in the right panel the results for Run II are reported. The lower panels show the corresponding ratio between experimental data and theoretical results. The latter are displayed as light-blue rectangles, each one corresponding to the integral of the cross section within the corresponding bin limits. The size of the rectangle is given by the 68% C.L. accuracy. We note that the quality of the fit for CDF data is comparable to the one in Ref. [81].

In Fig. 3.10, we show the results of the fit for the DY cross section in pp collisions at the LHC. From left to right, the black data points refer to the measurements released by the CMS Collaboration at increasing center-of-mass energy $\sqrt{s} = 7$, 8, 13 TeV. We remark that for the experimental data sets at $\sqrt{s} = 7$, 8 TeV, the results are normalized to the fiducial cross section. As in previous figures, the lower panels display the ratio of experimental data to the 68% C.L. theoretical results. For $\sqrt{s} = 7$, 8 TeV, the quality of the fit is comparable to the one in Ref. [81]. For the new data set at $\sqrt{s} = 13$ TeV, the agreement in the displayed rapidity bin is excellent, and remains very good also for higher values for the rapidity.

Fig. 3.11 reports the comparison between the theoretical results for the DY cross section in pp collisions normalized to the fiducial cross section for the ATLAS data at $\sqrt{s} = 7$ TeV. Three representative bins at increasing rapidity $|y|$ are shown from the left to the right panel. The leftmost one corresponds to the worst described experimental data set in our global analysis, with $\chi^2/N_{\mathrm{data}} = 13.5$. We notice that our fitted predictions correctly reproduce the behaviour of the data points, and the presence of extremely small experimental errors leads to a large contribution to the χ^2 calculation. As in previous figures, the lower part of each panel displays the ratio of experimental data to the 68% C.L. theoretical results. We note that the quality of the fit increases at more forward rapidities (from left to right). The same trend is observed at $\sqrt{s} = 8$ TeV, but not for CMS at 13 TeV.
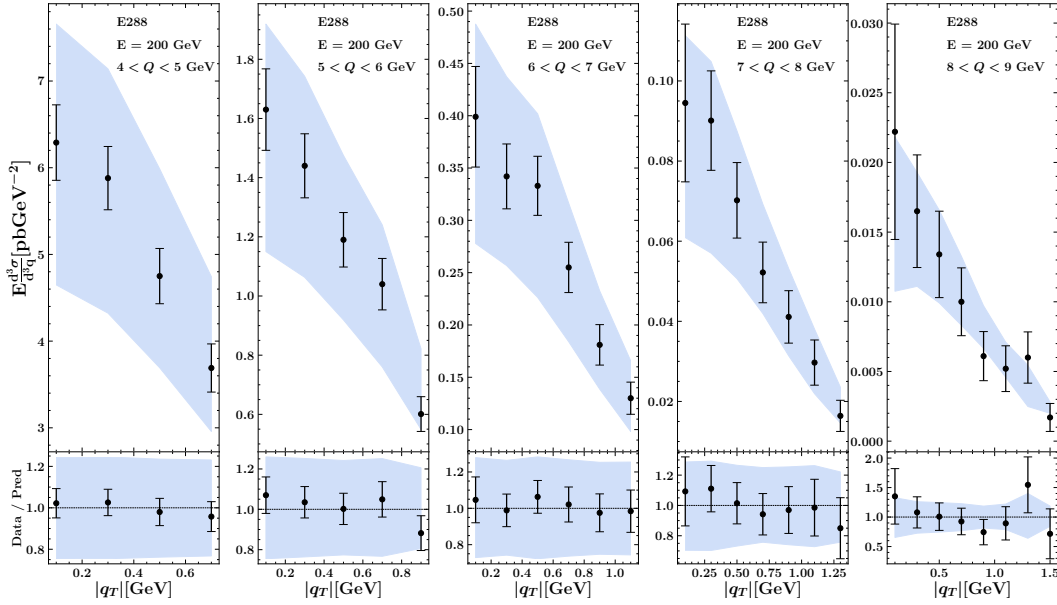
Figure 3.8: Upper panel: comparison between data and theoretical predictions for the DY cross section differential in $|\boldsymbol{q}_T|$ for the E288 data set at $E_{beam} = 200$ GeV for different Q bins; uncertainty bands correspond to the 68% CL. Lower panel: ratio between experimental data and theoretical cross section. For clarity, the whole band is displayed because for fixed target experiments we calculate the q_T-spectra at the mean variable.
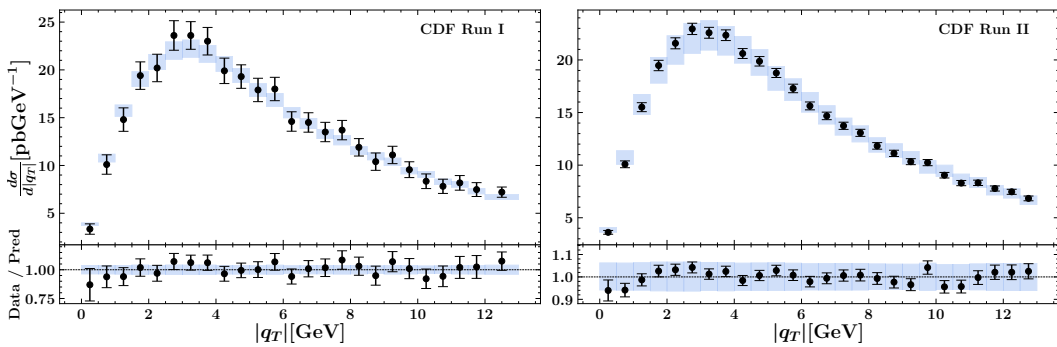

Figure 3.9: Upper panels: comparison between experimental data and theoretical predictions for the cross section differential in $|\boldsymbol{q}_T|$ for Z bosons produced in $p\bar{p}$ collisions at the Tevatron from CDF Run I (left panel) and run II (right panel); uncertainty bands correspond to the 68% C.L. Lower panel: ratio between experimental data and theoretical results.
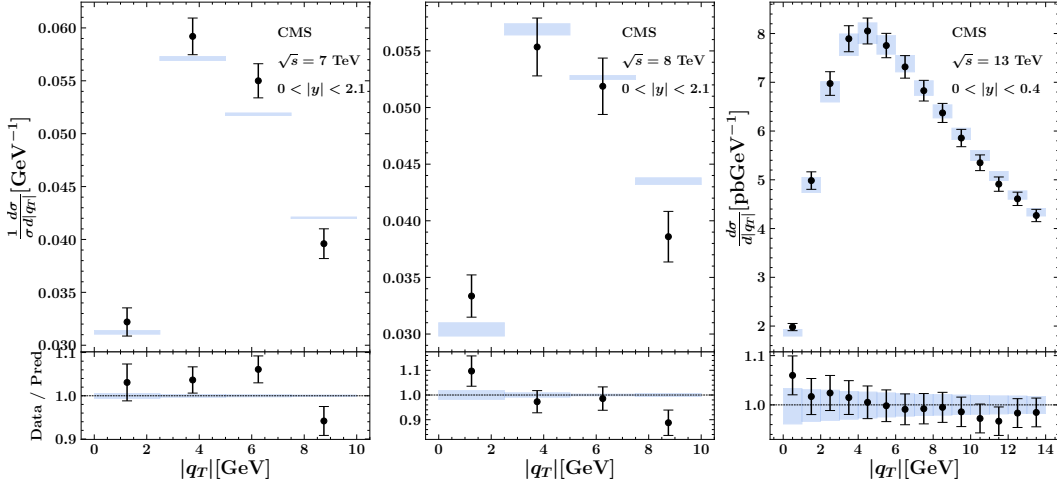
Figure 3.10: Same as in previous figure but for Z boson production in pp collisions measured by the CMS Collaboration. From left to right: increasing $\sqrt{s} = 7, 8, 13$ TeV, respectively. For $\sqrt{s} = 7, 8$ TeV, the results are normalized to the fiducial cross section.
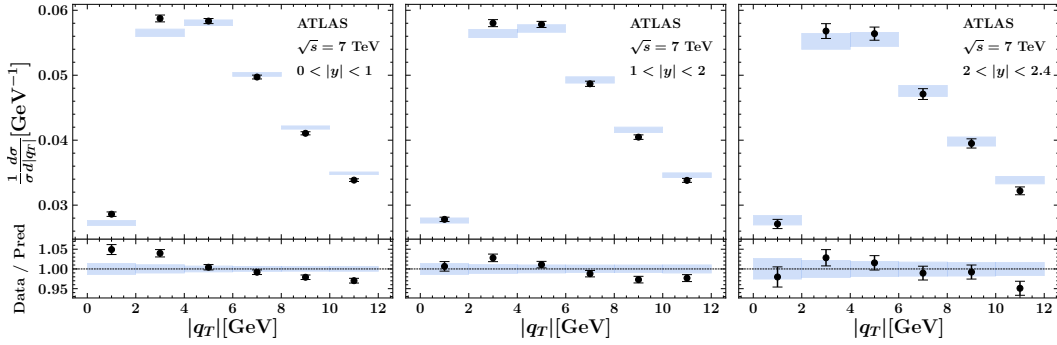

Figure 3.11: Same as in the left and central panels of previous figure, but for ATLAS kinematics at $\sqrt{s} = 7$ TeV. From left to right, results at increasing rapidity.

3.2.3 Extracted TMD distributions

In this section, we consider the TMD distributions extracted from our baseline fit at the N^3LL^- accuracy. Tab. 3.6 reports the full list of our 21 fitting parameters, along with their average values and standard deviations. Most of the parameters are well determined. Only the parameter γ_2 is found to be compatible with zero.

The λ parameter measures the relative weight of the weighted-Gaussian with respect to the first Gaussian in the nonperturbative part of the TMD PDF (see Eq. (3.5)). Since the value of this parameter is close to 2 and not compatible with zero, the contribution of the weighted-Gaussian component has a significant role in obtaining a good fit of the data. In Eq. (3.5), the parameter λ_2 measures the relative weight of the third Gaussian with respect

Parameter	Average over replicas
g_2 [GeV]	0.248 ± 0.008
N_1 [GeV^2]	0.316 ± 0.025
α_1	1.29 ± 0.19
σ_1	0.68 ± 0.13
λ [GeV^{-1}]	1.82 ± 0.29
N_{1B} [GeV^2]	0.134 ± 0.017
σ_2	0.455 ± 0.050
α_2	4.27 ± 0.31
λ_2 [GeV^{-1}]	0.0215 ± 0.0058
N_{1C} [GeV^2]	0.0130 ± 0.0069
σ_3	12.71 ± 0.21
α_3	4.27 ± 0.13
N_3 [GeV^2]	0.0055 ± 0.0006
β_1	10.23 ± 0.29
δ_1	0.0094 ± 0.0012
γ_1	1.406 ± 0.084
λ_F [GeV^{-2}]	0.078 ± 0.011
N_{3B} [GeV^2]	0.2167 ± 0.0055
β_2	4.17 ± 0.13
δ_2	0.167 ± 0.006
γ_2	0.0007 ± 0.0110

Table 3.6: Average and standard deviation over the Monte Carlo replicas of the free parameters fitted to the data.

to the first one; this parameter is small but not compatible with zero, which means that also this component of the TMD PDF is necessary to reach a good description of experimental data.

As already discussed in Sec. 3.2.1, the parameterization of the nonperturbative part of TMD FFs in Eq. (3.6) contains just the combination of a Gaussian and a weighted Gaussian because this is sufficient to describe the data in an accurate way. The λ_F parameter measures the relative weight of the two components; its value is close to 0.1, indicating that the contribution of the weighted Gaussian is small. However, being its value incompatible with zero, it has non-trivial consequences on the tail of the TMD FF, as we will show below. We tried also to include a third Gaussian similar to the TMD PDF case, but the quality of the fit did not increase.

The g_2 parameter is a key ingredient in the parameterization of the nonperturbative part of a TMD, since it is related to the Collins–Soper (CS) kernel (see Sec. 2.3.1). A more detailed discussion on the CS kernel will be addressed below. A similar parameter was used in the analysis of Ref. [87]. We observe

that the value obtained in the present global fit is smaller by almost a factor 4 with respect to Ref. [87]. This apparent incompatibility may be caused by the higher theoretical accuracy of the present analysis and by the role of the very precise high-energy Drell–Yan measurements, which may be responsible for the very small standard deviation of g_2.
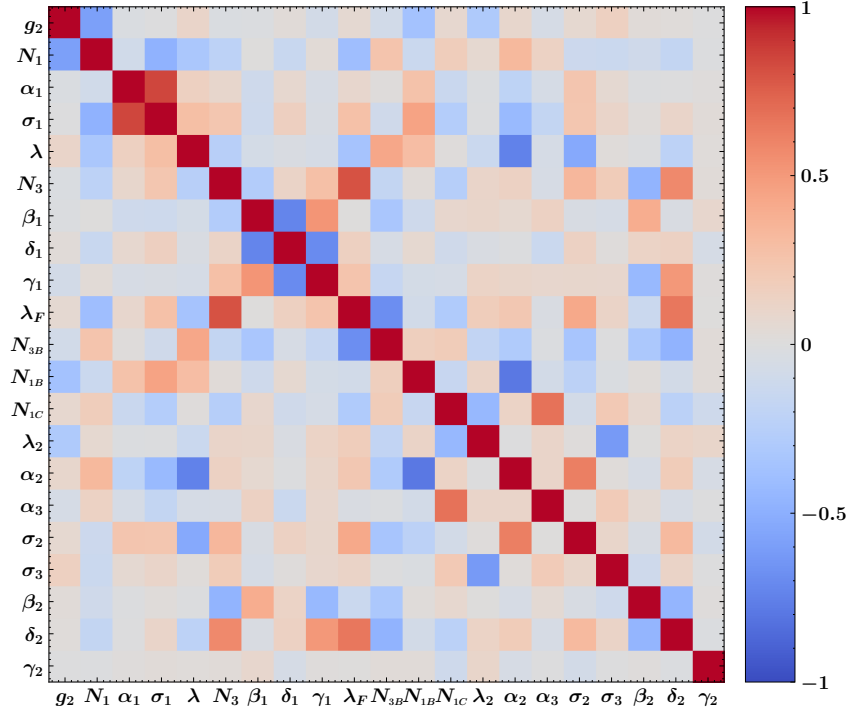

Figure 3.12: Graphical representation of the correlation matrix for the fitted parameters.

Fig. 3.12 displays a representation of the correlations among the 21 fitting parameters. Looking at the provided color code in the legend, we realize that the off-diagonal elements are generally small, except for certain (anti-)correlations observed among the parameters β_1, δ_1 and γ_1. These particular parameters govern the z–dependent width of the Gaussians in the TMD FF (see Eqs. (3.6)- (3.8)). The absence of significant correlations overall indicates that the chosen parametrization is appropriate and effectively captures the non-perturbative content of TMDs.

A visual representation of the uncertainty on the determination of the non-perturbative parameters is reported in Fig. 3.13, where the distribution of the values of a subset of parameters is shown.

The shapes of the up-quark TMD PDFs in the proton at $\mu = \sqrt{\zeta} = Q = 2$ and 10 GeV are summarized in Fig. 3.14 as a function of $|\mathbf{k}_\perp|$ and x. The surfaces refer to the central replica, no uncertainty band is shown for the readability of the figure. From this 3D-plot, we can better identify the relation between the normalization of the TMDs and the collinear PDFs at a specific value of the variable x.
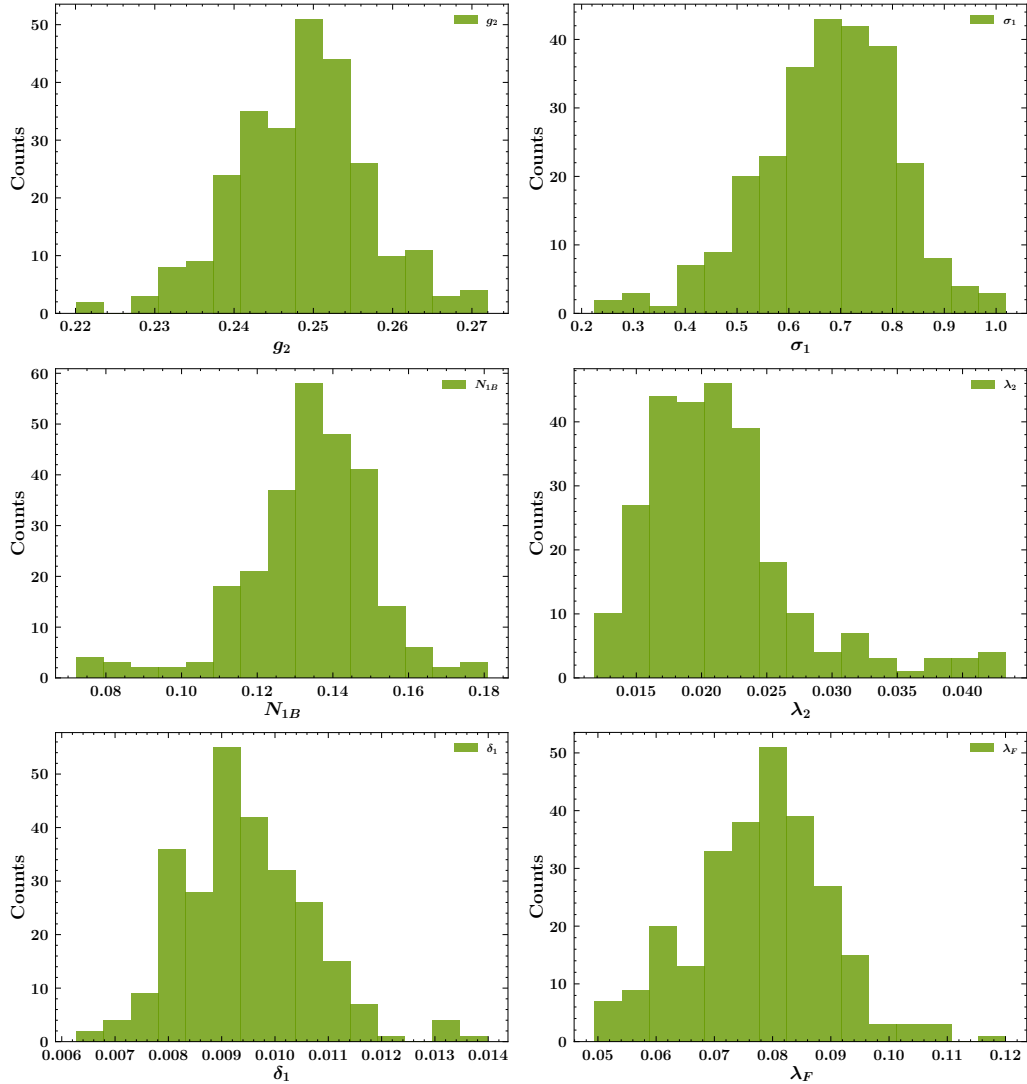
Figure 3.13: Distribution of the 250 replicas for a selection of the MAP22 free parameters.

A more detailed view is offered in Fig. 3.15. Here, we show the unpolarized TMD PDF for the up quark in the proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV (left panel) and 10 GeV (right panels) as a function of the quark transverse momentum $|\boldsymbol{k}_\perp|$ for three different values of x, namely $x = 0.001$, 0.01, and 0.1. In order to better visualize the uncertainties on the extracted TMDs, we draw also the 68% C.L. error bands, built by excluding the largest and the smallest 16% of the replicas for each value of $|\boldsymbol{k}_\perp|$.

From Figs. 3.14-3.15, we note that the TMD PDFs extracted from this analysis are wider at intermediate $x = 0.01$, but have also a high tail at $x = 0.001$. As already mentioned, a significant role is played by the presence of the weighted Gaussian. This may be a sign of separate contributions from different quark flavors and/or from different spin configurations in the shape
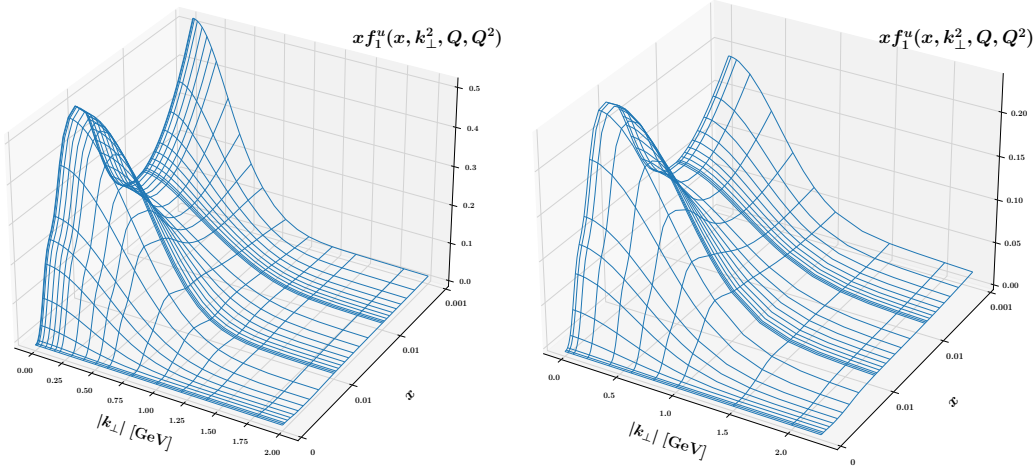
Figure 3.14: The TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV (left panel) and 10 GeV (right panel) as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$ and x. Results for the central replica, uncertainty bands are not shown for better readability.
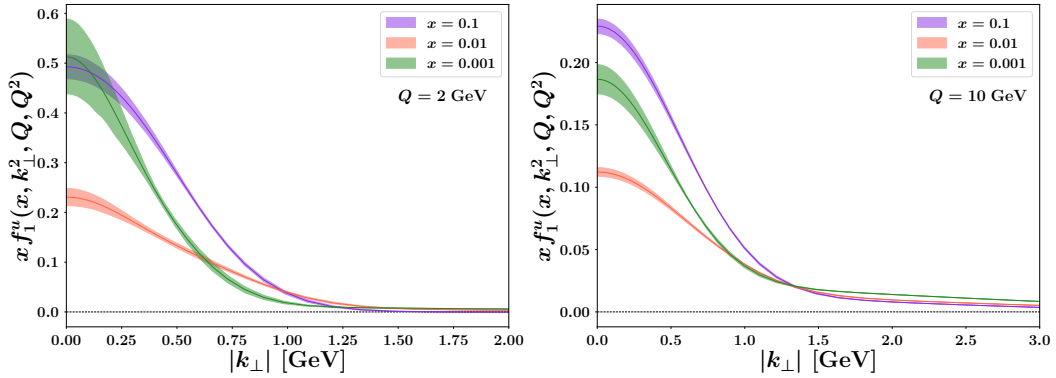

Figure 3.15: The TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV (left panel) and 10 GeV (right panel) as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$ for $x = 0.001, 0.01$ and 0.1. The uncertainty bands represent the 68% C.L.

of the TMDs (see Sec. 2.3).

It is worth noticing that the largest error band happens at $x = 0.001$ in both left and right panels, particularly at low $|\boldsymbol{k}_\perp|$. This is caused by the lack of of experimental data in that kinematic region (see Fig. 3.1), which is covered only by some DY data. An important role in reaching a better description of the TMD PDFs at low x will be played by future data from the Electron-Ion Collider (EIC) [107, 244]. The impact of new experimental data on the MAP22 extraction will be discussed in Sec. 3.4.

In Fig. 3.16, we simultaneously display in the same panel the 68% C.L. uncertainty bands (colored bands) and the spread of the full ensemble of replicas

(light blue lines) of the extracted unpolarized TMD PDFs for the up quark in the proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV (upper panels) and 10 GeV (lower panels) as a function of the quark transverse momentum $|\boldsymbol{k}_\perp|$ for the usual three different values of x. As for Fig. 3.15, we can see that the region where our TMD extraction shows the largest uncertainties is at small values of x. Moreover, Fig. 3.16 shows that whole set of replicas is not very spread, which is a signal of the accuracy level of the MAP22 extraction.
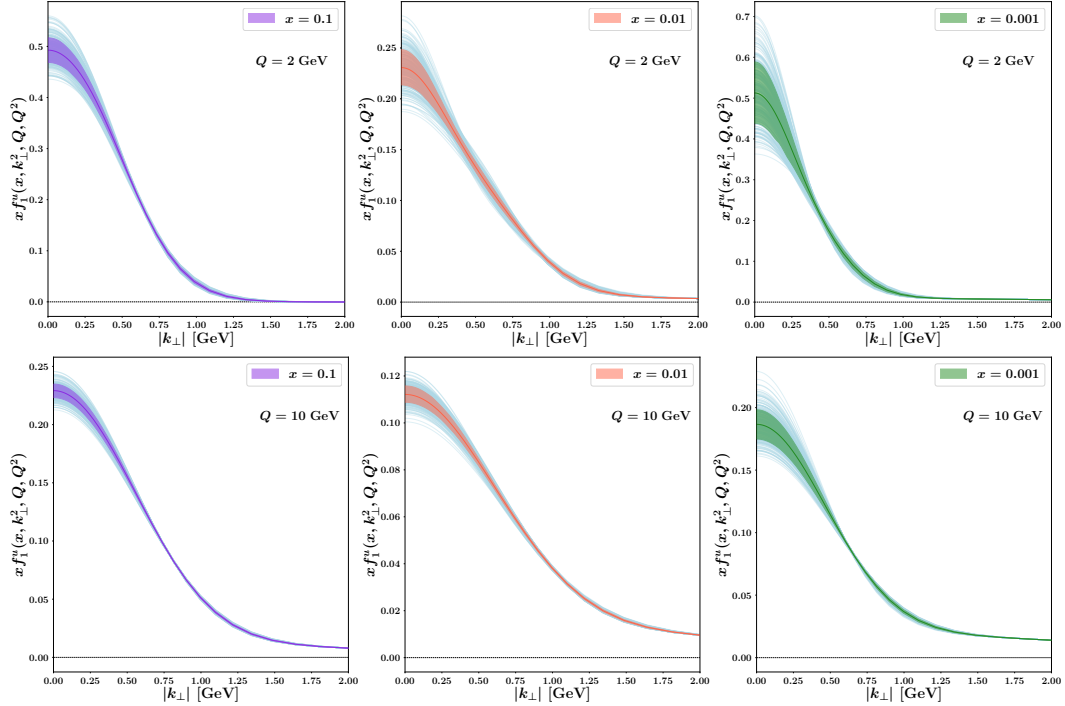

Figure 3.16: The TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV (upper panels) and 10 GeV (lower panels) as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$ for $x = 0.001$ (right panel), 0.01 (central panel) and 0.1 (left panel). The light blue lines depict each replica of the full ensemble, while the 68% C.L. is represented by the colored bands.

Notice that in Fig. 3.16 the values in the x-axis of the upper and lower panels are the same. This indicates that the TMD PDFs at larger values of the energy scale are broader, as expected from the evolution.

In Fig. 3.17, the TMD FF for the up quark fragmenting into a π^+ is displayed at $\mu = \sqrt{\zeta} = Q = 2$ GeV (left panel) and 10 GeV (right panel) as a function of the pion transverse momentum $|\boldsymbol{P}_\perp|$ for two different values of $z = 0.3$ and 0.6. As in previous figures, the uncertainty bands correspond to the 68% C.L. In both the left and right panels, a distinct additional structure becomes apparent at intermediate transverse momenta $|\boldsymbol{P}_\perp|$, particularly at $z = 0.3$. This peculiar feature arises as a consequence of the presence of the weighted Gaussian described in Eq. (3.6). It may be related to the contribution of secondary channels in the fragmentation of an up quark into a π^+, such

as the fragmentation into a Δ particle, which eventually decays into a pion. Further investigations on this topic are needed, and the inclusion of data from electron-positron annihilations would significantly improve our understanding of this feature.
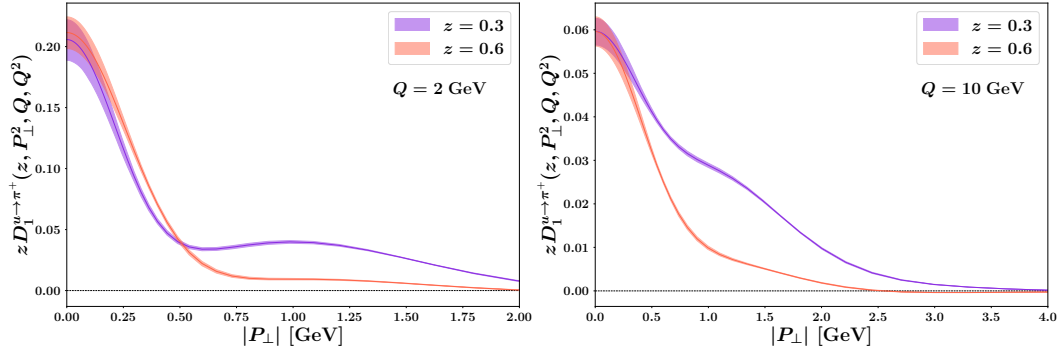

Figure 3.17: The TMD FF for an up quark fragmenting into a π^+ at $\mu = \sqrt{\zeta} = Q = 2$ GeV (left panel) and 10 GeV (right panel) as a function of the hadron transverse momentum $|\boldsymbol{P}_\perp|$ for $z = 0.3$ and 0.6. The uncertainty bands represent the 68% C.L.

In Fig. 3.18, we simultaneously display in the same panel the 68% C.L. uncertainty bands (colored bands) and the spread of the full ensemble of replicas (light blue lines) of the extracted unpolarized TMD FF for the up quark fragmenting into a π^+ at $\mu = \sqrt{\zeta} = Q = 2$ GeV (upper panels) and 10 GeV (lower panels) as a function of the pion transverse momentum $|\boldsymbol{P}_\perp|$ for the above mentioned two different values of z.

It is important to highlight that the displyed error bands in Figs. 3.15-3.18 represent the uncertainty on the fitted parameters (see Eqs. (3.5)-(3.6)). These uncertainties are determined by propagating the error on the extraction of the collinear parton distribution functions (PDFs) and fragmentation functions (FFs), as discussed in Sec. 3.1.3. However, since the fitting procedure is performed by using the central set of collinear distributions, all TMD replicas are forced to have the same integral in $\boldsymbol{k}_\perp$ (i.e., their values at $|\boldsymbol{b}_T| = 0$ are identical). Consequently, the plots in Figs. 3.15-3.18 only partially include the error associated to the collinear distributions.

A refined estimation of the uncertainty on TMD extractions could be achieved by considering a set of collinear distributions composed by an ensemble of MC replicas. Each of them should be associated to a single MC replica of the experimental data built to estimate the error on the nonperturbative parameters of TMDs. In this way, each replica of the fitted parameters would be linked to a definite replica of the collinear distribution, accounting for the uncertainty in both the shape and the normalization of a TMD distribution.
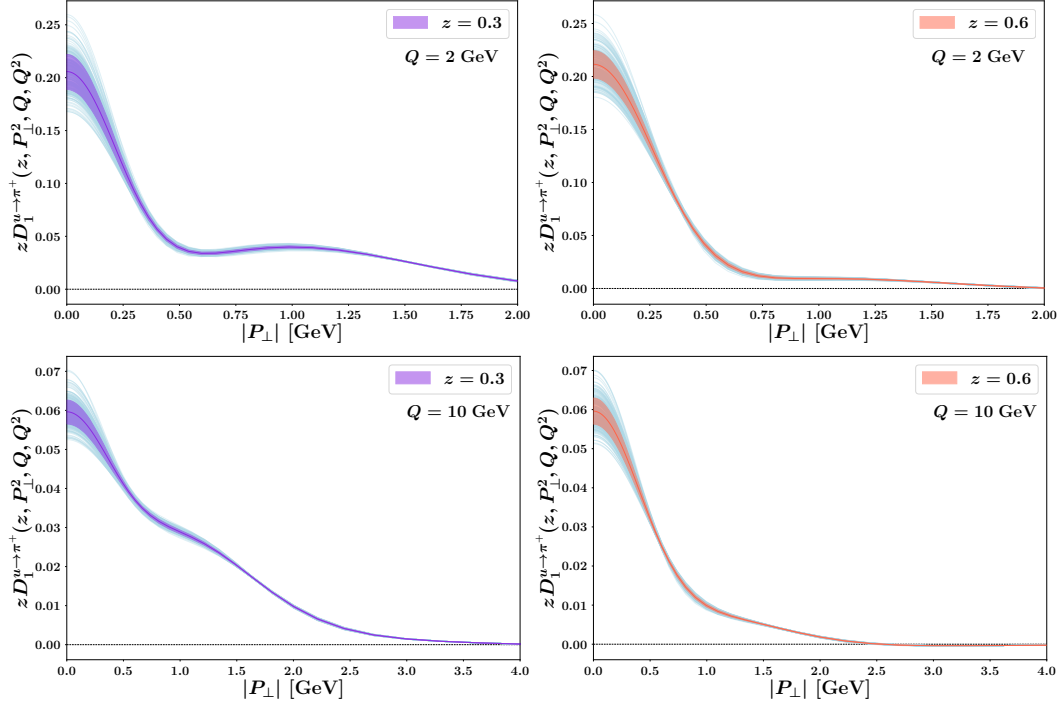
Figure 3.18: The TMD FF for an up quark fragmenting into a π^+ at $\mu = \sqrt{\zeta} = Q = 2$ GeV (upper panels) and 10 GeV (lower panels) as a function of the hadron transverse momentum $|\boldsymbol{P}_\perp|$ for $z = 0.6$ (right panel) and 0.3 (left panel). The light blue lines depict each replica of the full ensemble, while the 68% C.L. is represented by the colored bands.

3.2.4 Average squared transverse momenta

In order to better visualize the width of the extracted TMDs, we choose to provide the average squared momenta $\langle \boldsymbol{k}_\perp^2 \rangle (x, Q)$ (for TMD PDFs), $\langle \boldsymbol{P}_\perp^2 \rangle (z, Q)$ (for TMD FFs), which are calculated with the Bessel weighting technique (see Refs. [245, 246] for a detailed discussion).

In the case of the TMD PDF for a quark q in the proton at $\mu = \sqrt{\zeta} = Q$, the average squared transeverse momentum $\langle \boldsymbol{k}_\perp^2 \rangle (x, Q)$ is defined as follows [245, 246]:

$$\langle \boldsymbol{k}_\perp^2 \rangle^q (x, Q) = \frac{\int d^2 \boldsymbol{k}_\perp \, \boldsymbol{k}_\perp^2 \, f_1^q(x, \boldsymbol{k}_\perp^2, Q, Q^2)}{\int d^2 \boldsymbol{k}_\perp \, f_1^q(x, \boldsymbol{k}_\perp^2, Q, Q^2)} = \frac{2M^2 \, \hat{f}_1^{q\,(1)}(x, \boldsymbol{b}_T^2, Q, Q^2)}{\hat{f}_1^q(x, \boldsymbol{b}_T^2, Q, Q^2)} \Bigg|_{|\boldsymbol{b}_T|=0},$$
$$(3.12)$$

where the Fourier transform $\hat{f}_1^q$ of the TMD PDF has been defined in Eq. (2.9),

and the first Bessel moment of the TMD PDF $\hat{f}_1^{q\,(1)}$ is defined as [245]:

$$
\begin{aligned}
\hat{f}_1^{q\,(1)}(x, \boldsymbol{b}_T^2, Q, Q^2) &= \frac{2\pi}{M^2} \int_0^{+\infty} d|\boldsymbol{k}_\perp| \, \frac{\boldsymbol{k}_\perp^2}{|\boldsymbol{b}_T|} \, J_1\big(|\boldsymbol{k}_\perp||\boldsymbol{b}_T|\big) \, f_1^q(x, \boldsymbol{k}_\perp^2, Q, Q^2) \\
&= -\frac{2}{M^2} \frac{\partial}{\partial \boldsymbol{b}_T^2} \hat{f}_1^q(x, \boldsymbol{b}_T^2, Q, Q^2) \, .
\end{aligned}
$$

(3.13)

However, we need to shift the value of $|\boldsymbol{b}_T|$ from 0 to a value well inside the nonperturbative region [246]. This is required to obtain meaningful values for the average squared transverse momenta, *i.e.*, finite, positive, and not dominated by the perturbative tails of the TMDs. In this way, the Bessel functions $J_{0,1}$ have the effect of taming the power-law behavior of the TMD at large transverse momentum making the involved integrals to converge. It is clear that the choice of the specific value for $|\boldsymbol{b}_T|$ is arbitrary, but should not have a significant effect on the associated numerical values. We choose $|\boldsymbol{b}_T| = 1.5\, b_{\max}$, which guarantees that the average squared transverse momenta are positive in the x, Q values considered in the fit. Accordingly, Eq. (3.13) becomes:

$$
\langle \boldsymbol{k}_\perp^2 \rangle_r^q(x, Q) = \frac{2M^2 \, \hat{f}_1^{q\,(1)}(x, \boldsymbol{b}_T^2, Q, Q^2)}{\hat{f}_1^q(x, \boldsymbol{b}_T^2, Q, Q^2)} \bigg|_{|\boldsymbol{b}_T|=1.5\, b_{\max}} \, ,
$$

(3.14)

where the subscript r stands for *regularized*. We have numerically checked that the results are equivalent (up to numerical accuracy) when choosing either the integral or differential expressions in Eq. (3.13).

A similar arguments apply to the *regularized* average squared transverse momentum produced during the fragmentation of the quark q into the final state observed hadron h [139, 245, 246]:

$$
\langle \boldsymbol{P}_\perp^2 \rangle_r^{q \to h}(z, Q) = \frac{2\, z^2\, M_h^2\, \hat{D}_1^{q \to h\,(1)}(z, \boldsymbol{b}_T^2, Q, Q^2)}{\hat{D}_1^{q \to h}(z, \boldsymbol{b}_T^2, Q, Q^2)} \bigg|_{|\boldsymbol{b}_T|=1.5\, b_{\max}} \, ,
$$

(3.15)

where the Fourier transform $\hat{D}_1^{q \to h}$ of the TMD FF is defined in Eq. (2.33) and the first Bessel moment of the TMD FF $\hat{D}_1^{q \to h\,(1)}$ is defined as [139]:

$$
\begin{aligned}
\hat{D}_1^{q \to h\,(1)}&(z, \boldsymbol{b}_T^2, Q, Q^2) \\
&= \frac{2\pi}{M_h^2} \int_0^{+\infty} \frac{d|\boldsymbol{P}_\perp|}{z} \frac{|\boldsymbol{P}_\perp|}{z} \frac{|\boldsymbol{P}_\perp|}{z|\boldsymbol{b}_T|} \, J_1\big(|\boldsymbol{b}_T||\boldsymbol{P}_\perp|/z\big) \, D_1^{q \to h}(z, \boldsymbol{P}_\perp^2, Q, Q^2) \\
&= -\frac{2}{M_h^2} \frac{\partial}{\partial \boldsymbol{b}_T^2} \hat{D}_1^{q \to h}(z, \boldsymbol{b}_T^2, Q, Q^2) \, .
\end{aligned}
$$

(3.16)

In Fig. 3.19, we show the scatter plot of $\langle \boldsymbol{k}_\perp^2 \rangle_r^u$ (up quark contribution) at $x = 0.1$ vs. $\langle \boldsymbol{P}_\perp^2 \rangle_r^{u \to \pi^+}$ (commonly known as "favored" fragmentation channel)
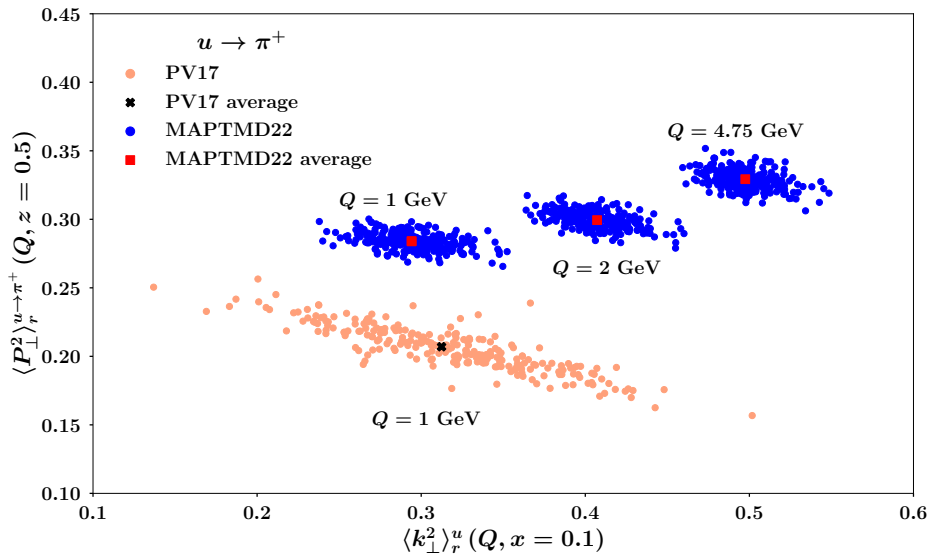
Figure 3.19: Scatter plot of average squared transverse momenta for the TMD PDF of the up quark at $x = 0.1$ and for the TMD FF of the $u \to \pi^+$ fragmentation at $z = 0.5$. Orange circles for the PV17 analysis [87] at NLL at $Q = 1$ GeV; the black cross represents the average. Blue circles for this analysis (MAPTMD22) at N^3LL^- and at $Q = 1, 2, 4.75$ GeV; the red squares represent the average values for each considered Q value.

at $z = 0.5$. The blue circles (denoted by MAPTMD22) are the results related to each single replica, while the red square corresponds to the average over all replicas for the N^3LL^- analysis. In order to show the evolution of the average transverse momenta with the energy scale, three different values of $Q = 1, 2,$ 4.75 GeV are reported in the plot. In Fig. 3.19, also the results for the PV17 replicas [87] at NLL and $Q = 1$ GeV are displayed, with the black cross being the average over the full ensemble. For the values extracted in the PV17 analysis, it is not necessary to introduce the above mentioned regularization because the involved TMDs at $Q = 1$ reduce to their nonperturbative components. From Fig. 3.19, we observe that the MAP22 extraction shows a much less anti-correlation between $\langle \boldsymbol{k}_\perp^2 \rangle$ and $\langle \boldsymbol{P}_\perp^2 \rangle$ at $Q = 1$ than the PV17 one. A possible explanation of this result may be the inclusion in the MAP22 analysis of very precise DY data from LHC.

It is interesting to see the dependence of the average squared transverse momenta on the energy scale. We can clearly see from Fig. 3.19 that, by increasing the value of Q from 1 GeV to 4.75 GeV, the average squared transverse momenta $\langle \boldsymbol{k}_\perp^2 \rangle$ for the TMD PDF of the up quark becomes larger and larger. This is consistent with the content of Figs. 3.15-3.18, and is physically reasonable. Similarly, the average squared transverse momenta $\langle \boldsymbol{P}_\perp^2 \rangle_r^{u \to \pi^+}$ grows at large values of the energy scale, even if in a milder way.

3.2.5 Collins–Soper kernel

Another interesting topic to be discussed is the Collins–Soper kernel [70, 73], namely the anomalous dimension that drives the evolution of TMDs in terms of the rapidity scale ζ (see Sec. 2.3.1). This crucial ingredient of the TMD formalism has been recently discussed in Refs. [247, 248], and several estimates based on lattice QCD have been recently proposed in Refs. [249–253].

The Collins–Soper kernel, as written in Eq. (3.3), can be decomposed in two parts. The first part is perturbatively calculable at a given N^kLL accuracy, and is computed at b_*:

$$
K(b_*(\mathbf{b}_T^2), \mu) = \sum_{n=0}^{k-1} \left(\frac{\alpha_s(\mu_{b_*})}{4\pi} \right)^{n+1} K^{(n,0)} - \sum_{n=0}^{k} \gamma_K^{(n)} \int_{\mu_{b_*}}^{\mu} \frac{d\mu'}{\mu'} \left(\frac{\alpha_s(\mu')}{4\pi} \right)^{n+1},
$$

(3.17)

where $K^{(n,0)}$ and $\gamma_K^{(n)}$ are the coefficients of the perturbative expansion (see, e.g., Ref. [128]). In the `NangaParbat` computational tool, the integral on the r.h.s. is directly computed numerically, thus providing a fully resummed result.

The second part is usually denoted as g_K, and it cannot be computed in perturbation theory. This is one of the results of our fit. We stress that only the full Collins–Soper kernel can be compared to other works, because the behavior of the g_K function is related to the choice of the b_*-prescription.[8]

Fig. 3.20 displays the Collins–Soper kernel as a function of $|\boldsymbol{b}_T|$ at the scale μ fixed at 2 GeV[9], for the present analysis (MAPTMD22, green band) and for four different analyses in the literature [79, 81, 87, 88]. The solid lines at low $|\boldsymbol{b}_T|$ follow the perturbative result, as expected. For the sake of comparison with the SV19 [88] and SV17 [79] results, the MAP22 [1], PV19 [81] and PV17 [87] curves are obtained by setting $b_{\min} = 0$. Note that the slight differences between the curves are due to the different accuracies of the perturbative calculations: the PV17 analysis was performed at NLL, the SV17 analysis at N^2LL, the PV19, SV19 and MAP22 at N^3LL. As in the other plots, the size of the bands around the solid lines corresponds to one standard deviation of the parameter g_2 around its best-fit value.

Moreover, we observe that the inclusion of the $b_{\min}$ value in the b_* prescription produces a saturation of the curve at low $|\boldsymbol{b}_T|$ starting from $|\boldsymbol{b}_T| \approx 1$ GeV^{-1}. The effect of introducing our choice $b_{\min} = 2e^{-\gamma_E}/\mu \approx 1.123/\mu$ is shown by the dashed curves. As expected, the Collins–Soper kernel saturates to a finite value in this framework. As already discussed in Sec. 2.3.4, we stress that such modification occurs at lower values of $|\boldsymbol{b}_T|$ (and becomes less relevant) as the energy scale increases.

The behavior at high $|\boldsymbol{b}_T|$ is driven by g_K and is different for the various analyses, accordingly to the related parameterization. We observe that the

---

[8]This is true also for the nonperturbative part of the TMD distributions: only the full TMD in Eq. (3.4) can be compared to other extractions.

[9]We keep the renormalization scale μ fixed at 2 GeV only for a better comparison with previous results.
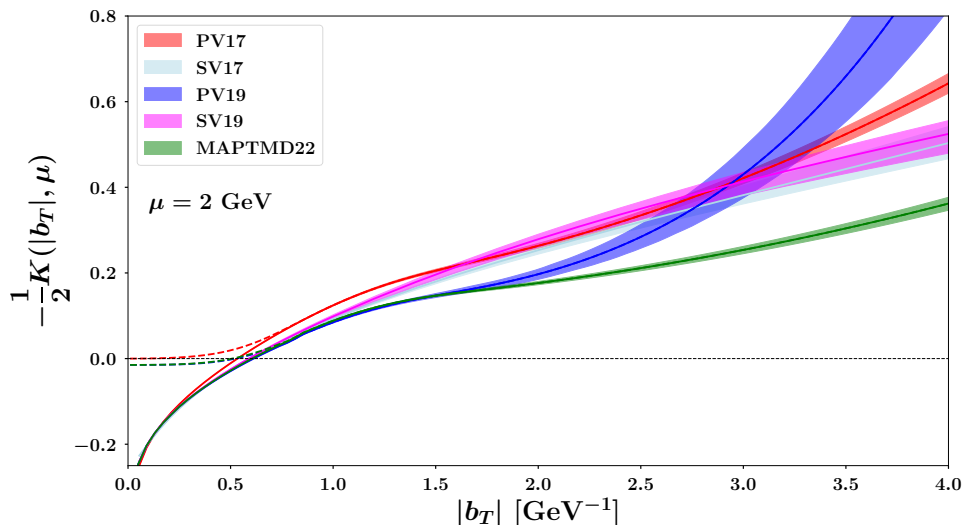
Figure 3.20: The Collins–Soper kernel as a function of $|\boldsymbol{b}_T|$ at the scale $\mu = 2$ GeV from the present analysis (MAPTMD22), compared with the PV17 [87], SV17 [79], PV19 [81], and SV19 [88] analyses (see color codes in the legend). For the MAPTMD22, PV17, and PV19 curves, the uncertainty bands represent the 68% C.L. The corresponding dashed lines show the effect of including the $b_{\min}$-prescription (see text).

curve of the MAP22 analysis is suppressed with respect to the others. This is caused by the fact that the parameter g_2 is smaller than the PV17 one. Further investigations on these differences will be performed in future works, and possible connections with the lattice QCD community will be very fruitful.

3.3 Variations of the fit configurations

In this section, we discuss the results obtained in the framework MAP22 global fit, but by modifying some of the baseline settings. In particular, we perform new fits at NNLL and NLL accuracy. Then, we study the impact of the modification of the b_*-prescription (see Sec. 2.3.4 for details on the theory) on both the shape of the extracted TMDs and some experimental observables. Finally, we study the sensitivity of our analysis with baseline settings to modifications of the cut on transverse momentum imposed to restrict our fit in the TMD region.

3.3.1 Global fit at NNLL and NLL

In the following, we firstly present two different fits at NNLL and NLL accuracy, respectively, comparing them to the baseline one at N^3LL^-.

Data set	N³LL⁻		NNLL		NLL	
	N_{dat}	$\langle\chi^2\rangle \pm \delta\langle\chi^2\rangle$	N_{dat}	$\langle\chi^2\rangle \pm \delta\langle\chi^2\rangle$	N_{dat}	$\langle\chi^2\rangle \pm \delta\langle\chi^2\rangle$
ATLAS	72	5.01 ± 0.26	/	/	/	/
PHENIX 200	2	3.26 ± 0.31	2	0.81 ± 0.11	/	/
STAR 510	7	1.16 ± 0.04	7	0.99 ± 0.03	/	/
Other sets	170	0.83 ± 0.01	170	2.37 ± 0.11	/	/
DY collider	251	2.06 ± 0.07	179	2.3 ± 0.1	/	/
E772	53	2.48 ± 0.12	53	2.05 ± 0.22	/	/
Other sets	180	0.87 ± 0.04	180	0.71 ± 0.04	180	0.81 ± 0.04
DY fixed-target	233	1.24 ± 0.04	233	1.01 ± 0.05	180	0.81 ± 0.04
HERMES	344	0.71 ± 0.04	344	1.1 ± 0.06	344	0.51 ± 0.02
COMPASS	1203	0.95 ± 0.02	1203	0.6 ± 0.06	1203	0.41 ± 0.01
SIDIS	1547	0.89 ± 0.02	1547	0.71 ± 0.05	1547	0.43 ± 0.01
Total	2031	1.08 ± 0.01	1959	0.89 ± 0.01	1727	0.47 ± 0.01

Table 3.7: Comparison of χ^2 values normalised to the number of data points N_{dat} for fits at different perturbative accuracies. The $\langle\chi^2\rangle$ and $\delta\langle\chi^2\rangle$ are the average and standard deviation of the χ^2 values of all replicas.

In the previous sections, we reported the results of the MAP22 baseline fit at N³LL⁻ accuracy (see Tab. 2.1). In Tab. 3.7, we compare them with results obtained at lower accuracy.

As already discussed in Ref. [81], it is crucial to introduce perturbative corrections up to N³LL to obtain a good description of some of the most recent experimental measurements, like at the LHC. However, it is interesting also to extract unpolarized TMDs at lower perturbative orders that enter the definition of the single-spin asymmetries for the extraction of polarized TMDs, where it is not possible to reach the same level of perturbative accuracy.

This is the case, *e.g.*, of the Sivers TMD PDF. In fact, due to the lack of knowledge of the collinear behavior at low $|\boldsymbol{b}_T|$, the computation of the polarized cross section for the Sivers effect presently cannot go beyond the NLL level [64, 65, 254, 255]. Therefore, unpolarized TMDs at the same level of accuracy are required for a consistent extraction.

However, since the most precise LHC data sets require the highest theoretical accuracy, it is possible to obtain a reasonably good fit only by excluding those experimental sets. Specifically, we found that we are able to reach an acceptable good description at NNLL only by removing the full ATLAS data set. As a matter of fact, we observe in Tab. 3.7 that the value of χ^2 in this configuration, namely for the full data set without ATLAS data, is lower than the one of the baseline fit at N³LL⁻.

When changing perturbative accuracy, the differences in the framework are significant. Therefore, we do not expect to find compatible results for the

best fit parameters between the NNLL and N³LL⁻ fits. For instance, we get $\lambda = 12 \pm 10$ GeV⁻¹ and $\lambda_F = 340 \pm 280$ GeV⁻² at NNLL, to be compared to $\lambda = 1.8 \pm 0.3$ GeV⁻¹ and $\lambda_F = 0.08 \pm 0.01$ GeV⁻² at N³LL⁻. We remind the reader that the λ and λ_F parameters determine the relative weight of the weighted Gaussian with respect to the first Gaussian in the non perturbative part of the TMD PDF and FF, respectively (see Eqs (3.5)-(3.6)). They control the size of the DY and SIDIS spectrum at middle to large values of $|\boldsymbol{q}_T|$. Since we obtain large values for such parameters, the weighted Gaussian dominates for both TMD PDF and FF parametrizations in the NNLL analysis. It is clear that this might not be a genuine behavior, and it may be induced by the lack of perturbative corrections in the NNLL fit, which gets compensated by nonperturbative effects.

In Fig. 3.21, we display the 68% C.L. error bands of the unpolarized TMD PDF for the up quark in the proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV and $x = 0.01$ as a function of the quark transverse momentum $|\boldsymbol{k}_\perp|$ for the baseline (violet band) and the NNLL (pink band) accuracies. In the lower plot the two bands are reported as normalized to the central value of N³LL⁻.
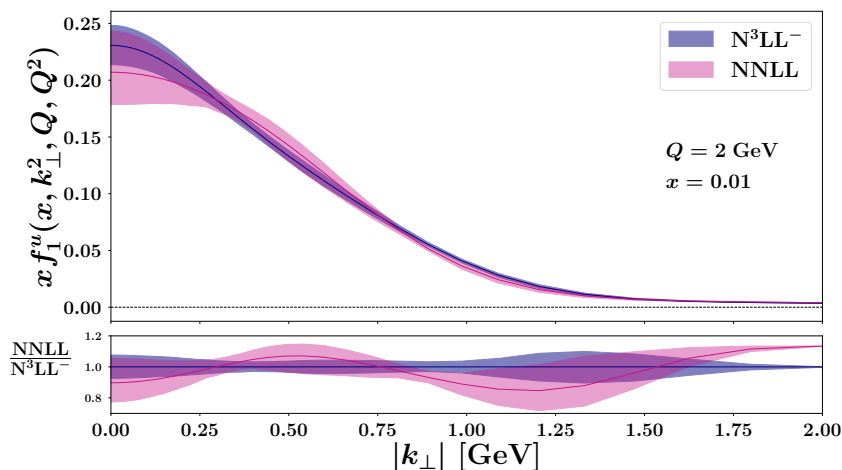

Figure 3.21: Upper plot: the TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV and $x = 0.01$ as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$ for the baseline N³LL⁻ and the NNLL accuracies of the MAP22 global fit. Lower plot: same bands as the upper plot, but normalized to the central value of the N³LL⁻ extraction. The uncertainty bands represent the 68% C.L.

We observe that the NNLL band is compatible with the N³LL⁻ one, indicating that in the shown kinematic region the resulting shape of the TMD PDF is comparable, even considering different global data sets and perturbative accuracies. However, a more relevant contribution of the weighted Gaussian can be seen in the shape of the NNLL curve at intermediate $|\boldsymbol{k}_\perp|$, confirming our discussion on the λ parameter. Note also that the two bands become incompatible at large $|\boldsymbol{k}_\perp|$, where the difference in the perturbative content of the

TMD distribution is expected to be relevant. As already discussed, such nice compatibility is not required nor expected, and in general different results are obtained for other values of x and Q.

We performed also a global fit at NLL accuracy. Similarly to the previous case, we have to exclude other data sets to obtain acceptable description at this order. Hence, we consider only SIDIS data and fixed-target DY data (*i.e.*, E605 and E288).

We point out that with our choice of the b_* prescription the integral of the W-term at NLL is fixed to the SIDIS $|\boldsymbol{q}_T|$-integrated collinear cross section. As a consequence, the value of the normalization prefactor ω in Eq. (2.99) is by construction equal to 1. Moreover, at this accuracy we consistently use the collinear MMHT2014 PDF set at LO and we use the DSS set at NLO for collinear FFs. As shown in Tab. 2.1, at NLL both collinear PDFs and FFs should be the result of an extraction at LO, but we choose FFs at NLO because no recent extractions at LO are currently available.

As can be seen in Tab. 3.7, low χ^2/N_{dat} are obtained for all included data sets. This result may be the consequence of a very flexible functional form fitting a reduced data set. It is interesting to compare such result with the outcome of the PV17 analysis [87]. In that work, the normalization of COMPASS data was fixed by the first bin in $\boldsymbol{P}_{hT}^2$. A new release of COMPASS has been published in the meanwhile, and here we prove that we can obtain an excellent description of the most recent COMPASS data at NLL, without problems of normalization.

In Fig. 3.22, we display the 68% C.L. error bands of the unpolarized TMD PDF for the up quark in the proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV and $x = 0.01$ as a function of the quark transverse momentum $|\boldsymbol{k}_\perp|$ for the baseline (blue band) and the NLL (green band) configurations. In the lower plot is reported the two error bands normalized to the N^3LL^- central value.

We observe that the NLL fit band is not compatible with the N^3LL^- one. As for the comparison with the NNLL fit, we note that the two bands are clearly incompatible at large $|\boldsymbol{k}_\perp|$, where the difference in the perturbative content of the TMD distribution is expected to be relevant. Moreover, we note that the integral of the two TMDs is different, being related to the collinear set of PDFs included in the analysis (MMHT2014 at NNLO for the N^3LL^- fit and MMHT2014 at LO for the NLL one).

Finally, in Fig. 3.23 we show the comparison between the 68% C.L. error bands of the unpolarized TMD PDF for the up quark in the proton for and $x = 0.1$, 0.01 and 0.001 at $\mu = \sqrt{\zeta} = Q = 2$ and 10 GeV as a function of the quark transverse momentum $|\boldsymbol{k}_\perp|$ for the baseline MAP22 fit (blue band) and the two variations at NNLL and NLL accuracy (violet and green bands, respectively).

From Fig. 3.23, we can observe the differences in the results of the three global analyses. In general, they are expected to be different because of the distinct data sets included in the analyses and the different orders in the per-
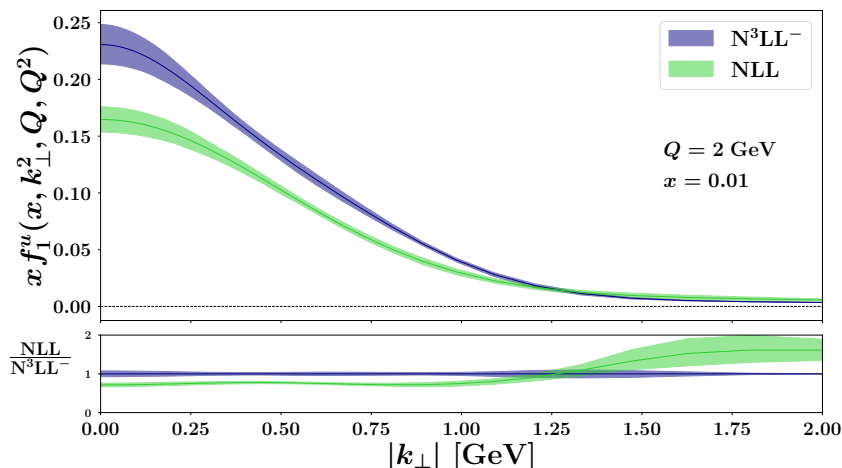
Figure 3.22: Upper plot: the TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV and $x = 0.01$ as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$ for the baseline N^{3}LL$^-$ and the NLL accuracies of the MAP22 global fit. Lower plot: same bands as the upper plot, but normalized to the central value of the N^{3}LL$^-$ extraction. The uncertainty bands represent the 68% C.L.

turbative ingredients. Moreover, since the integral of each TMD PDF is related to the value of the collinear set of PDF used in the extraction, the normalization of the three curves differs by construction according to the value of the MMHT2014 set at LO, NLO and NNLO (see Tab. 2.1).

In Fig. 3.24, the comparison between the the 68% C.L. error bands of the unpolarized TMD PDF for the up quark in the proton at $\mu = \sqrt{\zeta} = Q = 2$ and 10 GeV for $x = 0.1$, 0.01 and 0.001 of the MAP22 global fit at NLL and the PV17 global fit is shown. We note that both the normalization and the shape of the two TMD PDFs are pretty similar at $x = 0.1$, even if they are matched onto two different sets of collinear PDFs.[10] In the other cases, the results are not very compatible. As already discussed, this may be caused by the different data sets included in the analyses[11] and the different choice of collinear PDFs.

3.3.2 Impact of b_*-prescription on unpolarized TMDs

It is interesting to understand also the dependence of the TMD shape when we choose a different b_* prescription (see discussion in Sec. 2.3.4). To this aim, we show the TMDs calculated with different $b_*(|\boldsymbol{b}_T|)$ functions, but keeping the nonperturbative model with the same set of parameters. In fact, it is crucial to use the same parameterization in order to identify the genuine impact of a

---

[10]The MMHT2014 collinear set at LO is used for the MAP22 global fit at NLL, while the GJR08FFNLOE [256] for the PV17 analysis.

[11]We stress that in the PV17 global fit an old version of the COMPASS data set was included, which has been superseded by a new release in recent years.
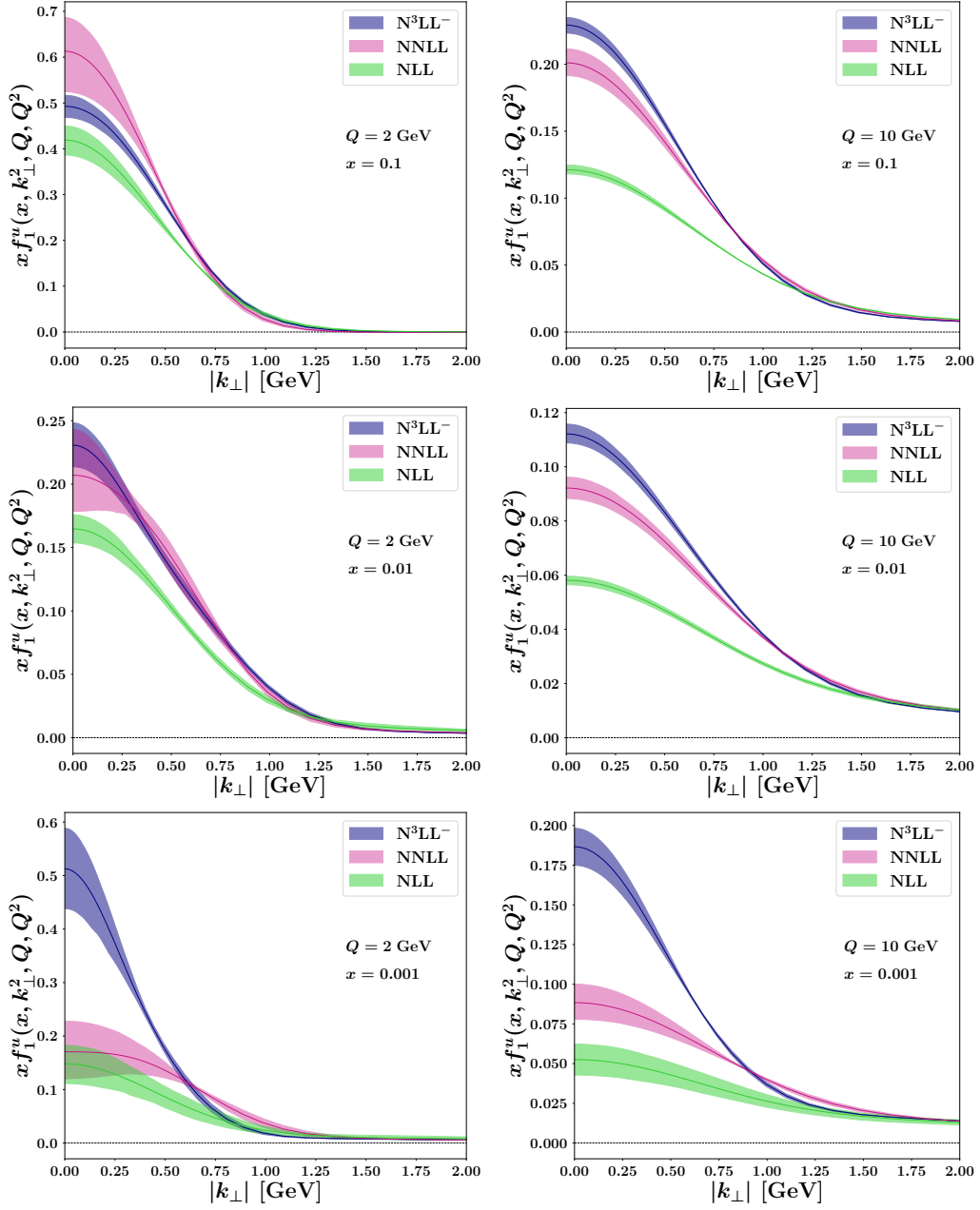
Figure 3.23: The TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV (left panels) and 10 GeV (right panels) and $x = 0.1$, 0.01 and 0.001 (from top to bottom) as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$ for the baseline N^3LL^- (blue band), the NNLL (violet band) the NLL (green band) accuracies in the MAP22 global fit. The uncertainty bands represent the 68% C.L.

particular choice of the b_* prescription before the fitting procedure. We choose
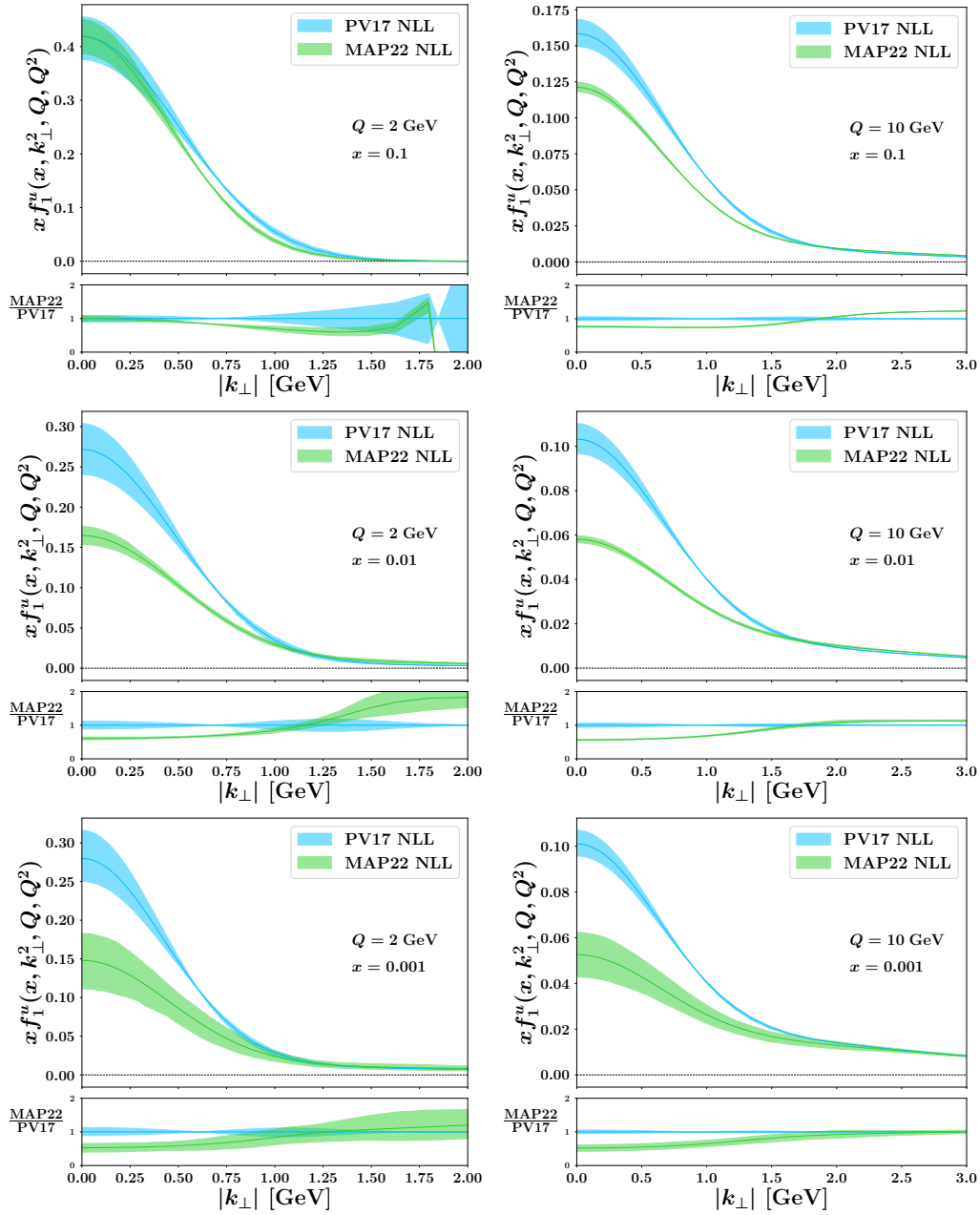
Figure 3.24: The TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV (left column) and 10 GeV (right column) and $x = 0.1$, 0.01 and 0.001 (from top to bottom) as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$. Upper plot: the TMD PDF of the up quark in a proton for the NLL accuracy MAP22 variation (green band) and the PV17 extraction (light blue band). Lower plot: same bands as the upper plot, but normalized to the central value of the PV17 extraction. The uncertainty bands represent the 68% C.L.

the set of parameters of the MAP22 baseline fit.[12] We choose four different

_____

[12]Actually, the choice of the b_* prescription is strictly related to the nonperturbative part of a TMD distribution (see Eq. (2.77) and discussion above). Therefore, a change in the

$b_*(\boldsymbol{b}_T^2)$ functions: $b_*^{Collins}(\boldsymbol{b}_T^2)$ (see Eq. (2.69)), $b_*^{Bacchetta}(\boldsymbol{b}_T^2)$ (see Eq. (2.70)), $\bar{b}_*^{Collins}(\boldsymbol{b}_T^2)$ (see Eq. (2.72)), and $\bar{b}_*^{MAP}(\boldsymbol{b}_T^2)$ (see Eq. (2.74)). The differences among them can be found in Figs. 2.5-2.6.

Fig. 3.25 shows the comparison between the unpolarized TMD PDF for the up quark in the proton calculated for different choices of the b_* prescription at $\mu = \sqrt{\zeta} = Q = 2$ GeV for $x = 0.1$ (left panel) and 0.01 (right panel).
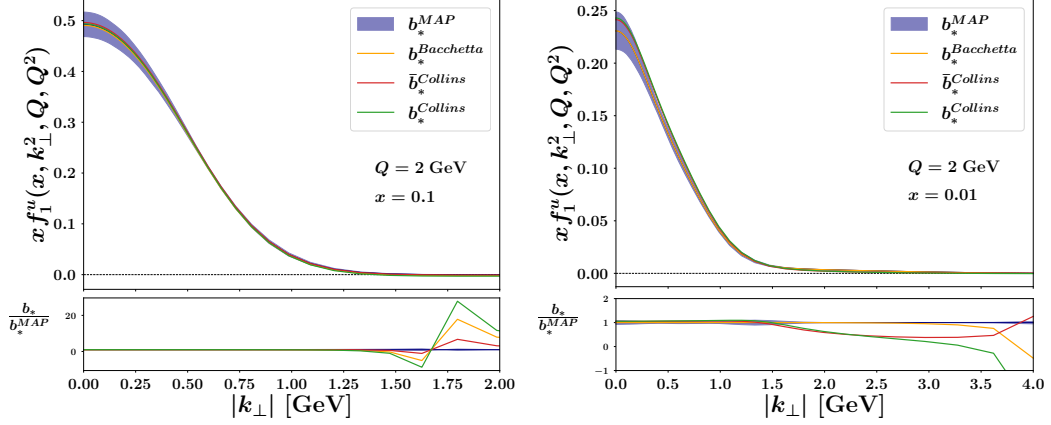

Figure 3.25: The TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 2$ GeV as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$ for x 0.1 (left panel) and 0.01 (right panel) for different b_* prescriptions. Upper plot: TMD PDFs comparison for $b_*^{Collins}(\boldsymbol{b}_T^2)$ (Eq. (2.69)), $b_*^{Bacchetta}(\boldsymbol{b}_T^2)$ (Eq. (2.70)), $\bar{b}_*^{Collins}(\boldsymbol{b}_T^2)$ (Eq. (2.72)), and $\bar{b}_*^{MAP}(\boldsymbol{b}_T^2)$ (Eq. (2.74)); 68% C.L. uncertainty bands are shown only for the MAP22 baseline fit result. Lower plot: ratio between the result for a given $b_*(\boldsymbol{b}_T^2)$ function choice and the $\bar{b}_*^{MAP}(\boldsymbol{b}_T^2)$ one.

Fig. 3.26 shows the same comparison, but at $\mu = \sqrt{\zeta} = Q = 10$ GeV.
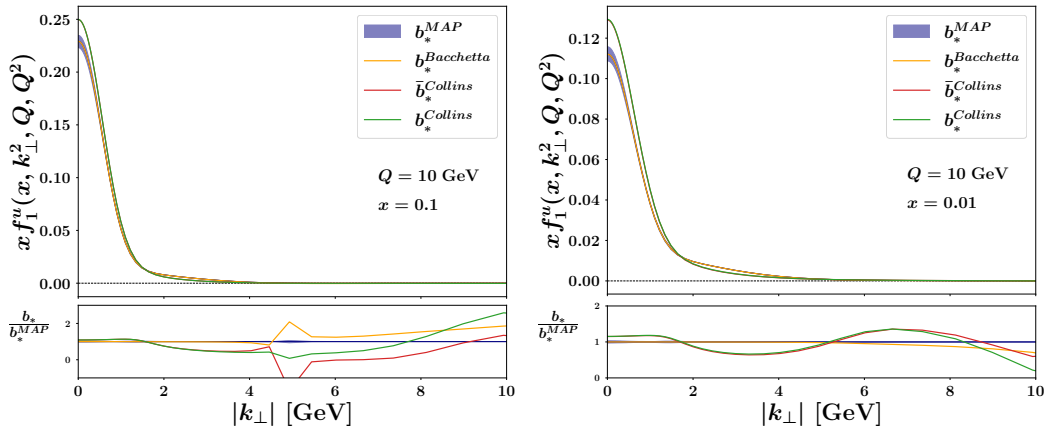

Figure 3.26: Same as in previous figure but at $\mu = \sqrt{\zeta} = Q = 10$ GeV.

$b_*(\boldsymbol{b}_T^2)$ function should require new best fit parameters. However, our aim is not to perform a new TMD extraction, but only to understand the impact of different choices.

In the left panel of Fig. 3.25 we observe that the discrepancy between the curves built with different b_* prescriptions lays in the region at intermediate $|\boldsymbol{k}_\perp|$. This is somehow expected, since the introduction of a b_{min} cutoff should generate different behaviors in the intermediate transverse momentum domain. This is observed also at $x = 0.01$ and for both the panels in Fig. 3.26. The differences between the $\bar{b}_*^{MAP}(\boldsymbol{b}_T^2)$ bands and the $\bar{b}_*^{Collins}(\boldsymbol{b}_T^2)$ can be related to the difference in the value of the b_{min} parameter and the different shapes of the functions. Indeed, the discrepancies that can be seen in Fig. 3.26 may be caused by the different behaviors of the $b_*(\boldsymbol{b}_T^2)$ at both small and large $\boldsymbol{b}_T^2$. It is interesting to note also that at intermediate $|\boldsymbol{k}_\perp|$ the TMDs with $b_*^{Bacchetta}(\boldsymbol{b}_T^2)$ and $\bar{b}_*^{MAP}(\boldsymbol{b}_T^2)$ are slightly larger than the other. This feature is a consequence of the fact that such prescriptions leave more space to perturbative radiation before saturating in both small- and large-$|\boldsymbol{b}_T|$ regions (see discussion in Sec. 2.3.4).

We also perform the fit of the ATLAS data set by considering the $\bar{b}_*^{MAP}(\boldsymbol{b}_T^2)$ prescription of the baseline fit and the $\bar{b}_*^{Collins}(\boldsymbol{b}_T^2)$ one. We report the results in Fig. 3.27, where the comparison of the two configurations are shown for a selection of ATLAS bins. We choose to fit only the ATLAS data set because it is the most precise and problematic one. Therefore, a modification of the b_* prescription used to build the theoretical predictions could be responsible for a visible modification in the agreement with the experimental data.

We note that the peak of the $|\boldsymbol{q}_T|$-spectra is generally shifted at smaller $|\boldsymbol{q}_T|$ values when we consider the $\bar{b}_*^{Collins}(\boldsymbol{b}_T^2)$ prescription. Consequently, the predictions for the first $|\boldsymbol{q}_T|$ bin is generally larger and the intermediate-$|\boldsymbol{q}_T|$ tail is lower than the experimental data. This observation is in agreement with Figs. 3.25-3.26, where we can see that the TMD PDFs with $\bar{b}_*^{Collins}(\boldsymbol{b}_T^2)$ prescription are systematically larger than the baseline ones at small $|\boldsymbol{q}_T|$, while they are slightly smaller at intermediate $|\boldsymbol{q}_T|$.

In conclusion, we find better agreement between the ATLAS experimental data and the theoretical (fitted) predictions built with the choice of the $\bar{b}_*^{MAP}(\boldsymbol{b}_T^2)$ prescription. Such result could be considered as a phenomenological justification for our baseline choice of the b_* prescription.

3.3.3 Cut in $|\mathbf{q}_T|/Q$

The last variation of the MAP22 baseline fit configuration is related to the cut on the experimental data introduced to keep the MAP22 global analysis inside the TMD region. An essential ingredient in any phenomenological analysis of TMDs is the kinematic cut $|\boldsymbol{q}_T|/Q \ll 1$ on the data set to guarantee the applicability of TMD factorization. In Sec 3.1, we extensively discussed the MAP22 default choices for these kinematic cuts. It is of particular interest to determine how the global quality of fit is affected by variations in this cut, as it provides quantitative insight into the range of validity for TMD factorization.

To this purpose, we change the parameters c_1, c_2, and c_3 in Eq. (3.2), which play a role only on the cut for the SIDIS data, while we keep the limit
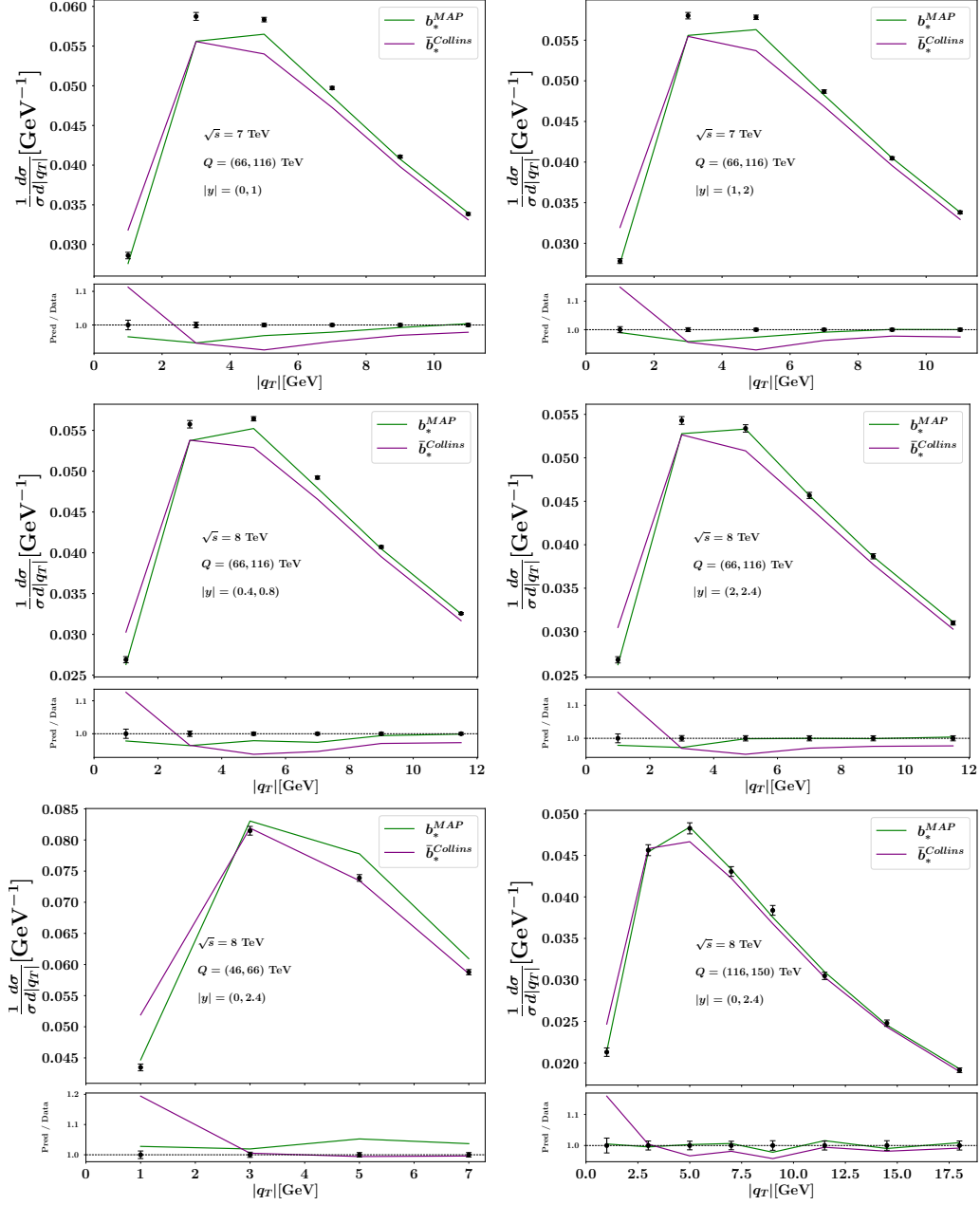
Figure 3.27: Comparison between the experimental data and (fitted) theoretical predictions with $\bar{b}_*^{MAP}(\boldsymbol{b}_T^2)$ (green line) and $\bar{b}_*^{Collins}(\boldsymbol{b}_T^2)$ (purple line) prescriptions for a selection of ATLAS data sets. Lower panel: ratio between theoretical results and experimental data. Fit errors are not shown.

$|\boldsymbol{q}_T|/Q < 0.2$ for the DY data. An analogous study of the $|\boldsymbol{q}_T|/Q$ cut on the DY data has already been performed in Ref. [81].

We consider five different configurations for the cut on SIDIS data:

(a) A first and most conservative cut is performed by fixing the z-independent upper value $|\boldsymbol{q}_T|/Q < 0.4$ which can be obtained by setting $c_1 = c_2 = 0.4$

and $c_3 = 0$ in Eq. (3.2);

(b) A second cut by setting $c_1 = 0.15$, $c_2 = 0.4$ and $c_3 = 0.2$ in Eq. (3.2);

(c) The cut of our baseline fit with $c_1 = 0.2$, $c_2 = 0.5$ and $c_3 = 0.3$;

(d) A fourth cut with $c_1 = 0.2$, $c_2 = 0.6$ and $c_3 = 0.4$ but without imposing $|\boldsymbol{q}_T| < Q$ (*i.e.* removing the outermost "min" in Eq. (3.2));

(e) A fifth cut inspired by the PV17 analysis [87], namely the same as in the previous case but with $c_1 = 0.2$, $c_2 = 0.7$ and $c_3 = 0.5$; this is the least conservative choice.
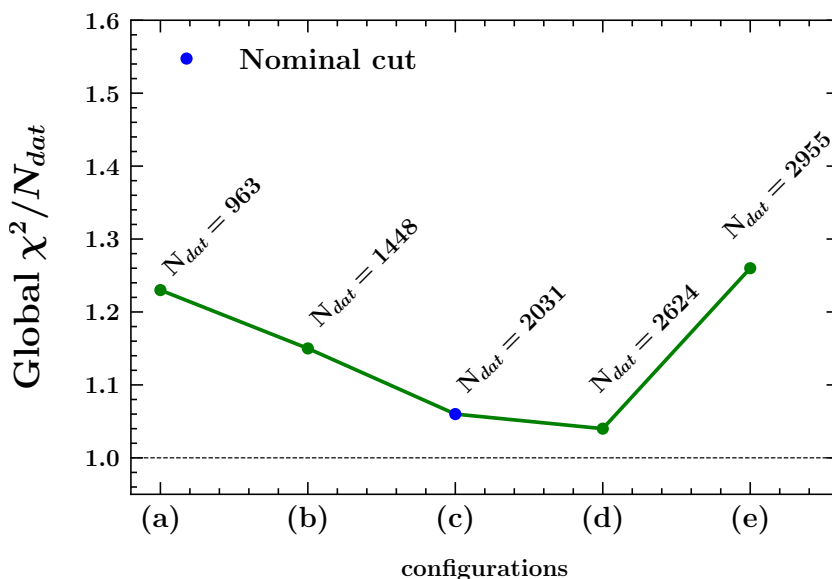

Figure 3.28: Global χ^2/N_{dat} for different configurations of the kinematic cut on SIDIS data sets (see text). The blue point corresponds to the reference cut used in the present baseline fit.

Fig. 3.28 displays the variation of the global χ^2/N_{dat} when taking into account the five configurations described above. We observe that more conservative choices do not necessarily correspond to better χ^2 values. Consequently, we deduce that the TMD formalism is able to describe a bunch of SIDIS data that do not satisfy the formal requirement $|\boldsymbol{q}_T|/Q \ll 1$. In this regard, we notice that the global χ^2/N_{dat} of cut (d) is smaller than the baseline fit, despite including a larger amount of data, some of which at $|\boldsymbol{q}_T| \gtrsim Q$, *i.e.*, well outside the region where TMD factorization is expected to be valid.

The situation is better represented in Fig. 3.29, where we show the comparison between COMPASS data and theoretical predictions from our baseline fit. We choose to plot the SIDIS multiplicity for positively charged hadrons as a function of $|\boldsymbol{P}_{hT}|/Q$ in the bin $1.3 < Q < 1.73$ GeV, $0.02 < x < 0.032$, $0.3 < z < 0.4$. The upper panel of the plot displays the multiplicity while

the lower panel shows the ratio of experimental data to the theoretical predictions. A vertical line separates data points included in the fit (solid circles on the left) from data predicted (empty squares on the right). Remarkably, the agreement remains very good up to $|\boldsymbol{P}_{hT}|/Q \simeq 0.5$, well beyond the value of $|\boldsymbol{P}_{hT}|$ allowed by our baseline cut. We also stress that this behavior is not specific of the considered COMPASS bin but is a general feature also of other bins, as well as of the HERMES data set. In Fig. 3.30, we observe the same feature also for a specific HERMES kinematic bin.[13]
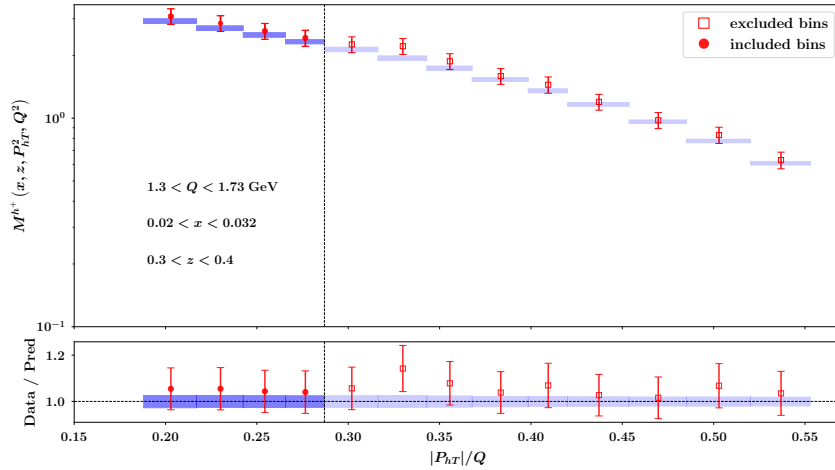

Figure 3.29: Comparison between COMPASS multiplicities and theoretical results for the SIDIS production of unidentified positively charged hadrons off a deuteron target at $1.3 < Q < 1.73$ GeV, $0.02 < x < 0.032$ and $0.3 < z < 0.4$ as a function of $|\boldsymbol{P}_{hT}|/Q$. Upper panel: light-blue rectangles for baseline fit at 68% CL, empty squares for data points not included in the baseline fit. Lower panel: ratio between experimental data and theoretical results. Vertical line separates data included in the fit (left) from predictions (right).

In general, we may conclude that from the MAP22 global analysis it appears that the validity of the TMD formalism, at least in the kinematic region covered by COMPASS and HERMES data sets, seems to extend well beyond the usual cut $|\boldsymbol{q}_T|/Q \ll 1$.

This empirical evidence provides a quantitative reason for our selection of the $|\boldsymbol{q}_T|/Q$ cut in Eq. (3.2) for the baseline fit. It makes clearer why we consistently got χ^2/N_{dat} values close to 1 even when taking into account less stringent cuts. Additionally, it suggests that the applicability of TMD factorization in SIDIS might be better defined in terms of $|\boldsymbol{P}_{hT}|$ rather than $|\boldsymbol{q}_T|$. Further investigations in this particular direction are needed.

---

[13]Notice that the selected bins from the COMPASS and the HERMES data sets are in two different kinematic regions.
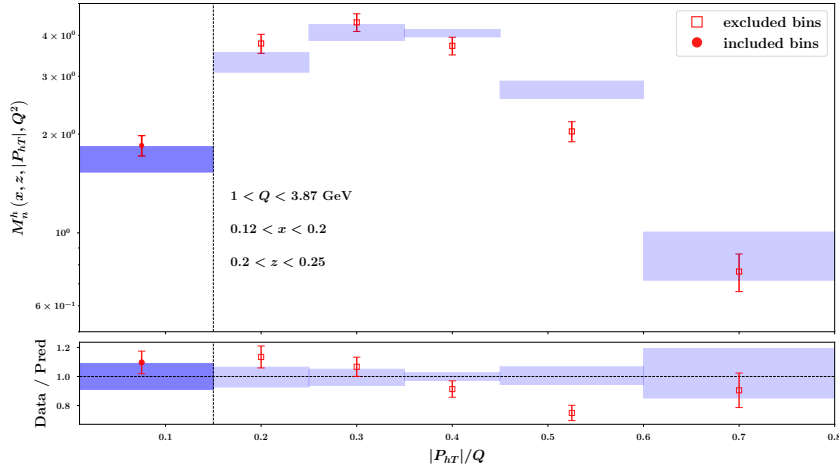
Figure 3.30: Comparison between HERMES multiplicities and theoretical results for the SIDIS production of positively charged pions off a proton target at $1 < Q < 3.87$ GeV, $0.12 < x < 0.2$ and $0.2 < z < 0.25$ as a function of $|\boldsymbol{P}_{hT}|/Q$. Same notations as in previous figure.

3.4 Impact of new data sets

Estimating the impact of new experimental data sets in TMD phenomenology is extremely important for advancing our understanding of fundamental aspects of the theory of strong interactions.

The availability of new sets of experimental data allows us to refine and validate theoretical models and parametrizations. By comparing the predictions of TMD factorization with the outcomes of the latest experiments, we can determine the robustness and accuracy of existing theoretical frameworks. Moreover, the analysis of new data sets enables the exploration of poorly known kinematic regimes and offers the potential to shed light on previously unobserved phenomena.

Precise estimates of the impact of new experimental data are also important to increase precision and reliability of theoretical predictions for future experiments. By quantifying the influence of these data sets, experimental collaborations can optimize designs of a new experiment or a new detector and target specific regions of interest, maximizing the potential of a new machine.

There are two main questions we would like to answer. *Which are the kinematic regions where the new data sets will give the most of information? How much the uncertainties on TMD distributions will be reduced?* To answer these questions, a fundamental role will be played by the interchanges between theoreticians and experimentalists, at least for the first question. In fact, it is the accumulated knowledge by phenomenology that should drive the goals and determine the kinematic domains of new experiments. Hence, it is important to answer these questions before designing a new experiment or a new machine.

We devote this section to the study of the impact of new pseudodata coming

from the EIC machine [107] and the JLab22 upgrade [106] on unpolarized quark TMDs.

There are several procedures that can be followed in order to perform such estimate. In Ref. [107], the sensitivity of different kinematic regions to the parameters of the PV17 [87] extraction of TMD PDFs and TMD FFs is discussed through the calculation of some sensitivity coefficients S, defined as

$$S[\mathcal{O}, f_i] = \frac{\langle \mathcal{O} f_i \rangle - \langle \mathcal{O} \rangle \langle f_i \rangle}{\delta \mathcal{O} \Delta f_i} \, , \tag{3.18}$$

where $\mathcal{O}$, $\delta \mathcal{O}$ are a given experimental observable and its uncertainty, and f_i, Δf_i are a given parameter of a nonperturbative TMD model and its theoretical uncertainty.

Another useful tool for quantitatively estimating the reduction of unpolarized TMD bands is given by the reweighing technique [257, 258], which has been usually used in collinear PDFs framework. However, the outcome of this procedure may be biased by the choice of the reweighing function, and may lead to unreliable results.

The most reliable way to estimate the impact of pseudodata is to repeat a global fit with the same setup, but including them from the very beginning. We choose to follow this procedure. In particular, we generate the values of pseudodata by calculating the theoretical predictions in `NangaParbat` computational tool with the set of parameters of the MAP22 central replica. Then, the associated experimental uncertainties are provided by the experimental collaborations through their Monte Carlo simulations. In this way, the result of the new fit is equivalent to the MAP22 one, while the uncertainties on the extracted TMDs are modified by the introduction of new pseudodata. In particular, such uncertainties are reduced according to the size of the experimental errors of the new pseudodata.

3.4.1   Effects of JLab 22 GeV data on TMDs

In this section, we show the results of the impact on the new experimental data coming from a possible JLab 22 GeV upgrade.[14] In unpolarized TMD phenomenology, JLab22 is expected to extremely improve the experimental accuracy in the measurement of SIDIS cross sections in the moderate- and large-x region, which will help to get refined information on the 3D nucleon structure in the valence domain with greater precision that ever before. In Fig. 3.31, the coverage of the JLab 22 GeV upgrade is illustrated. Only for illustrative purposes, we consider a center-of-mass energy $\sqrt{s} = 7$ GeV, $0.1 < y < 0.9$, and an invariant mass $W^2 < 10$ GeV^2. As already mentioned, we expect that the new experimental data coming from this upgrade will help us to deepen our knowledge on the valence TMD content of the proton. In fact,

---

[14]For more details on the JLab22 upgrade see Ref. [106].

they will provide information about the region $0.1 < x < 0.5$, which is not covered by available SIDIS data sets.
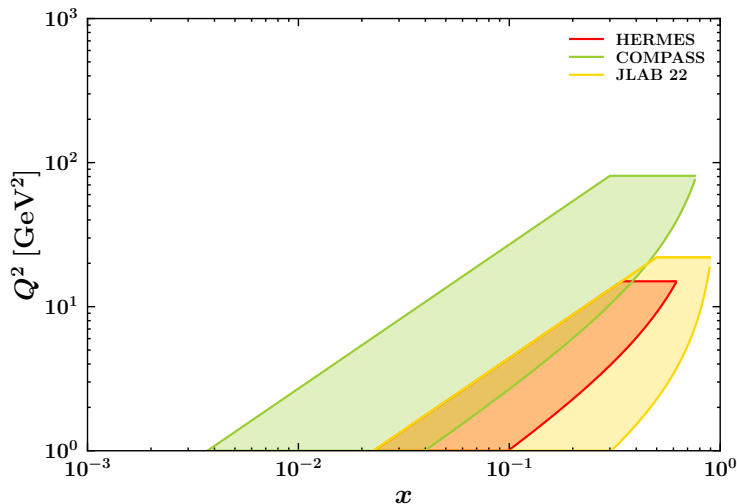

Figure 3.31: The x vs. Q^2 coverage spanned by the SIDIS experimental data considered in the estimate of the impact of JLab 22 GeV upgrade.

From Fig. 3.31, we observe that almost half of the region covered by the JLab22 data is the same of the HERMES and COMPASS data sets. This is important because it allows us to compare and validate the theoretical predictions produced from the knowledge of past experiments. On the contrary, the region beyond $x = 0.1$ - 0.2, particularly at low Q, is so far unexplored.

For the estimate of the impact of the JLab 22 GeV upgrade, two sets of experimental pseudodata were provided by JLab experimentalists, with identified positive or negative pions. Since each set includes more than 30000 experimental bins after imposing the kinematic cuts, we decided to consider only the set with identified positive pions in the final state. Therefore, we added more than 30000 pseudodata to the MAP22 global fit with ~ 2000 data. This unbalance could introduce a bias in the analysis: the fit might be no more sensitive to the actual experimental data.

The impact of the JLab22 data at $Q = 2$ GeV (left panels) and 10 GeV (right panels) and for $x = 0.1$ (first row), 0.01 (second row) and 0.001 (third row) is shown in Fig. 3.32.

As expected, the major reduction of the TMD uncertainty bands provided by the JLab22 pseudodata is in correspondence of the large-x region (upper panels of Fig. 3.32). In such kinematic region, the impact of the new data is huge because of the very small size of the experimental errors and the large amount of pseudodata introduced. In the central and lower panels of Fig. 3.32, we observe that the JLab 22 GeV upgrade may impact also at lower values of x, even if this kinematic region is not covered. This result may be an effect of the bias introduced by the functional form for the parameterization of the TMD nonperturbative part. In fact, it could happen that, due to the limited
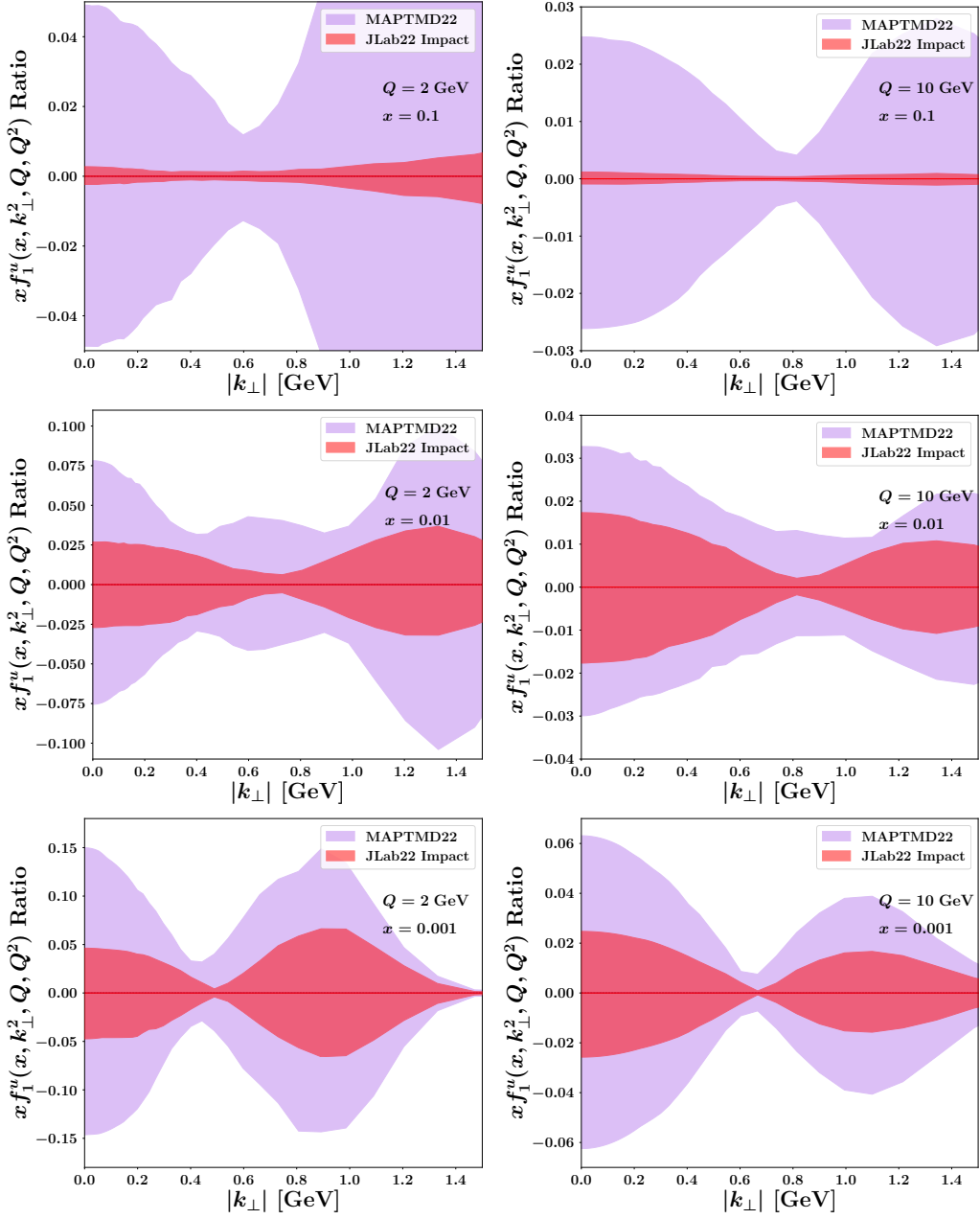
Figure 3.32: Estimate of the impact of the JLab 22 GeV upgrade on the error bands of the unpolarized TMD PDF of the u quark in the proton in $|\boldsymbol{k}_\perp|$ space at Q = 2 GeV (left column) and 10 GeV (right column) for $x = 0.1$, 0.01 and 0.001 (from top to bottom), based on the MAP22 analysis. Purple bands: current 68% C.L. TMD uncertainties from the MAP22 global fit. Red bands: 68% C.L. TMD uncertainties after the inclusion of JLab22 pseudodata.

flexibility of the nonperturbative TMD model, some parameters related to uncovered kinematic regions are fixed by the available experimental data.

In conclusion, we showed that the impact of JLab22 data on reducing quark

unpolarized TMD uncertainties is mainly in the large-x region. In particular, it is estimated to be about two orders of magnitude for $x = 0.1$ starting from the MAP22 global analysis.

3.4.2 Effects of EIC data on TMDs

In this section, we discuss the results on the impact of pseudodata coming from the Electron Ion Collider (EIC).[15]

In Fig. 3.33, the kinematic coverage of the EIC machine is displayed, along with the present SIDIS and DY data sets. Only for illustrative purposes, we consider for the EIC a center-of-mass energy $\sqrt{s} = 140$ GeV, $0.01 < y < 0.95$, according to Ref. [107].
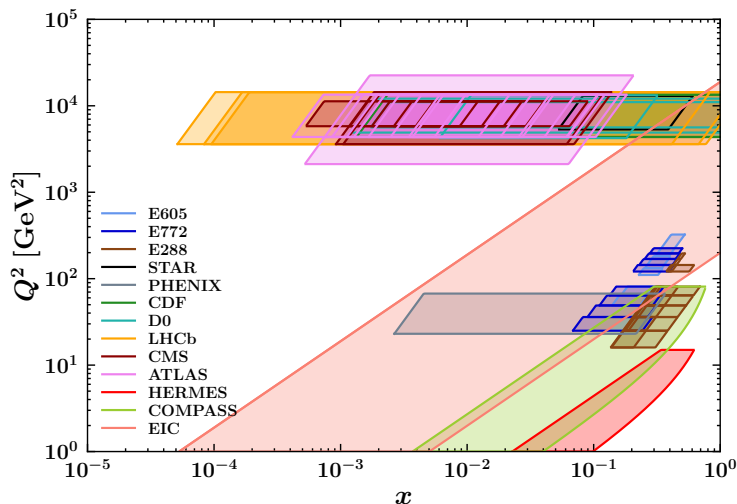
Figure 3.33: The x vs. Q^2 coverage spanned by the experimental data considered in unpolarized TMDs global analyses compared to the one of the EIC.

As can be seen from Fig. 3.33, the EIC machine is expected to provide unique insights on the 3D nucleon structure either in the region at large-x and large-Q or in the one at small- or intermediate-x and small-Q. Therefore, measurements at the EIC energies are expected to fill in the gap between the low-energy fixed-target experiments (both SIDIS and DY) and the high-energy collider machines as the LHC and Tevatron. In fact, the region $10 < Q < 100$ GeV is of particular interest to get information about TMD evolution and the compatibility between SIDIS and DY data.

For the estimate of the impact of the EIC machine, different preliminary sets of pseudodata were available for different energy configurations.[16] Since we got qualitatively similar results for the other energy configurations, we

---

[15]For more details on the EIC machine see Ref. [107].

[16]The EIC pseudodata are based on PYTHIA simulations, and include statistical and systematic errors obtained from a hand-book detector design with moderate particle identification (PID) capabilities.

show the results for the 18x100 GeV case. [17] Moreover, we consider only the pseudodata sets with identified positive pions in the final state. In the end, we were able to add about 3000 pseudodata points to the MAP22 global data set of ~ 2000 data, obtaining a good balance between real experimental data and pseudodata.

In Fig. 3.34, we show the estimate of the impact of EIC data at $Q = 2$ GeV (left panels) and 10 GeV (right panels) and for $x = 0.1$ (first row), 0.01 (second row) and 0.001 (third row).

In the central and the lower panels of Fig. 3.34, we can see that the impact of the EIC pseudodata is significant in the region at small and intermediate x. Moreover, we note that we obtain a nice improvement of the TMD uncertainty bands also at $x = 0.1$ and $Q = 2$ GeV, a kinematic region to which the EIC should not be sensitive. Such result may be related to a limited flexibility of the nonperturbative TMD parameterization. In any case, we expect that this result is an overestimation of the actual reduction of the TMD bands when the real data will become available. However, this tells us how much the availability of these new data sets is of fundamental importance to get a better understanding of the 3D structure of the nucleon. We stress that the importance of getting a better knowledge of the unpolarized structure of nucleons goes beyond the interest in understanding unpolarized physics. In fact, unpolarized TMDs enter in the definition of polarized structure functions and spin asymmetries, and, thus, they play a significant role the determination of polarized effects inside the nucleon.

---

[17]The EIC possible energy configurations for electron-ion collisions are: 5x41 GeV, 5x100 GeV, 10x100 GeV, 18x100 GeV and 18x275 GeV.
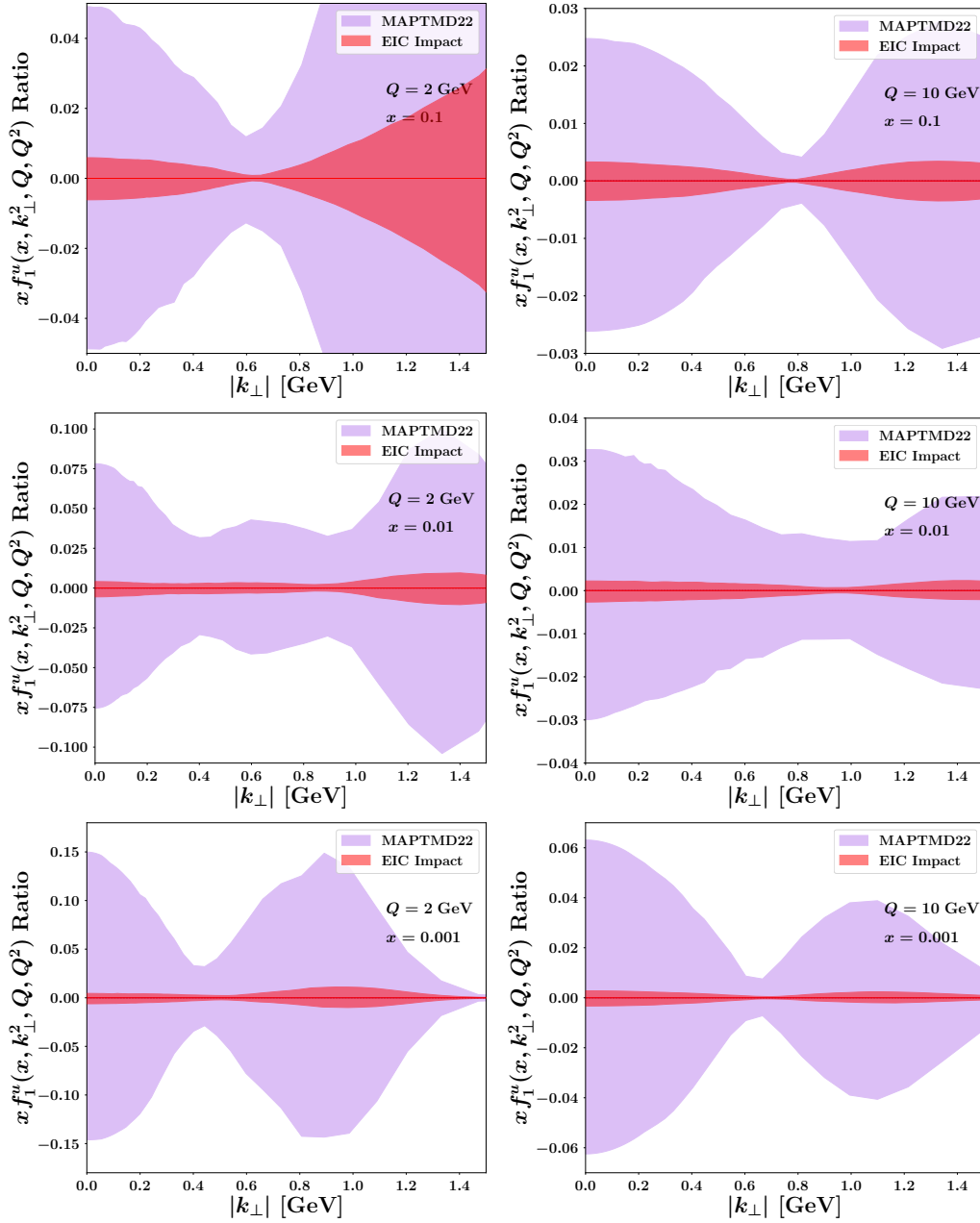
Figure 3.34: Estimate of the impact of the EIC machine on the error bands of the unpolarized TMD PDF of an u quark in the proton in $|\boldsymbol{k}_\perp|$ space at Q = 2 GeV (left column) and 10 GeV (right column) for $x = 0.1$, 0.01 and 0.001 (from top to bottom), based on the MAP22 analysis. Purple bands: 68% C.L. current TMD uncertainties from the MAP22 global fit. Red bands: 68% C.L. TMD uncertainties after the inclusion of EIC pseudodata.

Chapter 4

Phenomenology of pion TMD PDFs

In this chapter, we report on the extraction by the MAP Collaboration of unpolarized pion quark TMDs, henceforth named PionMAP22 [2], through the first ever analysis of the whole set of available data for the DY lepton-pair production in π^--nucleus collisions.

We are interested in the internal structure of the pion because it is the simplest of all hadrons. Moreover, it plays a fundamental role in the Standard Model, being the Goldstone boson of chiral symmetry breaking. Therefore, we find important to investigate its internal structure, which is also responsible for the macroscopic differences between the bound state of a pion and of a nucleon (see, e.g., Ref. [259] for a review).

The one-dimensional momentum distribution of quarks inside a pion can be described in terms of collinear PDFs, and there are various extractions available in the literature starting from the 1990s [260–270]. Moreover, a simultaneous extraction of pion PDFs and TMDs has been recently presented by the Jefferson Lab Angular Momentum (JAM) Collaboration [134].

Despite this extensive literature, the knowledge of the internal structure of pions is much less refined than the one of the proton. Such situation is caused by the lack of experimental data on scattering processes involving pions in different kinematic domains.

Consequently, the knowledge of 3D structure of pions is even more limited. In the last ten years, various model calculations of pion TMDs have been discussed with different approaches [271–278], and only few phenomenological analyses have been published [134, 279, 280]. The extraction of TMD PDFs presented here is an improvement upon the previous ones because it includes the whole set of pion-induced DY data and the theoretical formalism is pushed at the best accuracy currently available.

Since the available measurements are obtained from pion-nucleus collisions, the cross section differential in the lepton-pair transverse momentum can be written in terms of a convolution of a unpolarized proton TMD and a unpolarized pion TMD. We make use of results of the MAP22 global fit discussed in Ch. 3 for proton TMD PDFs, and we extract pion TMDs in the `NangaParbat`

computational tool with the formalism presented in this thesis.

If we compare our analysis to Ref. [280], we include a broader data set and use a different implementation of TMD evolution. As for Ref. [279], our fit includes a broader data set, a better theoretical accuracy, considers the latest extraction of proton TMDs, and consistently uses the same Collins-Soper kernel for both proton and pion TMDs.

The simultaneous analysis of pion PDFs and TMDs in Ref. [134] has been published after the work presented in this thesis. Compared to Ref. [134], we include an even larger number of experimental data points, and we introduce a different functional form for the nonperturbative part of pion TMDs. It is interesting to observe that the proton TMD PDFs from the MAP22 global fit have been used also in Ref. [134], confirming the quality of our extraction.

This chapter is organized as follows. In Sec. 4.1, we describe the data set included in the PionMAP22 analysis and how experimental data have been selected. In Sec. 4.2, we present the results of our fit. Finally, Sec. 4.3 shows the predictions of the PionMAP22 analysis for the kinematic region spanned by the upcoming pion-induced DY data from the COMPASS Collaboration.

4.1 Data selection

In this section, we present the experimental data sets included in the PionMAP22 analysis. The only available sets come from two DY experiments (E615 [281] and E537 [45]) on π^--tungsten collisions. The kinematic coverage of these data sets in the (x,Q^2) plane is illustrated in Fig. 4.1.
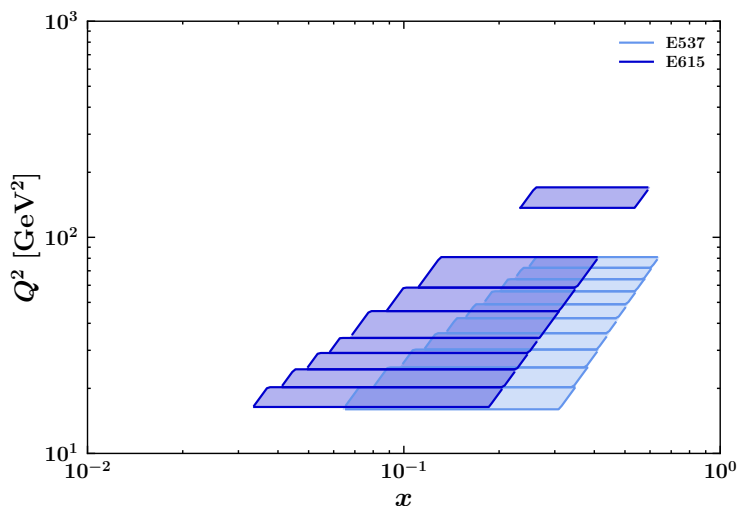

Figure 4.1: The coverage of E615 [281] (blue) and E537 [45] (light blue) data in the (x,Q^2) plane.

From Fig. 4.1, we observe that the kinematic coverage of pion-induced DY data is comparable with the fixed-target DY sets included in the MAP22 global

fit (see Fig. 3.1). As expected by fixed-target DY experiments, they provide information in the kinematic region at small Q and relatively large x.

As already discussed in Sec. 3.1, we need to impose a cut $|\boldsymbol{q}_T| \ll Q$ on the transverse momentum $|\boldsymbol{q}_T|$ to limit our analysis to the TMD region. Hence, we apply to the included data set the following cut:

$$\frac{|\boldsymbol{q}_T|}{Q} < 0.3 + \frac{0.6}{Q} \ . \tag{4.1}$$

This cut differs slightly from the corresponding one introduced in the extraction of proton TMDs (see Eq. (3.1)). Due to the limited number of data sets, we find that this cut represents a balanced compromise between the need to include more data and the requirement to stay within the limits of applicability of TMD factorization.

Moreover, we also exclude those bins of the DY experiment E615 for which $9.00 < Q < 11.70$ GeV to avoid the kinematical region of invariant masses around the Υ resonance. The number of experimental data points (N_{dat}) after the application of these cuts are reported in Table 4.1, together with the definition of the observables and the kinematic regions of each data set.

Experiment	N_{dat}	Observable	$\sqrt{s}$ [GeV]	Q range [GeV]	x_F range
E615	74	$d^2\sigma/dQd|\boldsymbol{q}_T|$	21.8	4.05 - 13.05	0.0 - 1.0
E537	64	$d^2\sigma/dQd\boldsymbol{q}_T^2$	15.3	4.0 - 9.0	-0.1 - 1.0

Table 4.1: For each experiment we indicate the number N_{dat} of data points included in the fit after applying the kinematical cuts, the delivered observable, the center-of-mass energy $\sqrt{s}$, the range in invariant mass Q and the integration range in x_F.

We note that each of the considered data sets is affected by relatively large systematic and statistical uncertainties. In particular, the systematic uncertainties are quite large for both E537 (8%) and E615 (16%). We choose to treat them as fully correlated.

As for the MAP22 analysis, we have to include the systematic uncertainties introduced by the choice of the collinear PDFs. Consistently with Ref. [1], we choose the PDF set MMHT2014 [210] to perform the matching at low-$|\boldsymbol{b}_T|$ of proton TMDs, while for the pion we choose the xFitter20 one [267]. The PDF uncertainties have been estimated by using the Hessian method, and 80% of them has been considered as fully correlated while the remaining 60% as uncorrelated, as already done in the MAP22 global fit. The t_0-prescription is not included in our results, since we found no relevant differences between the fits with and without the t_0-prescription.

4.2 Phenomenological results

In this section, we show the phenomenological results of the extraction of unpolarized quark TMD PDFs in the pion from a fit of all the existing DY data involving pions (see Tab. 4.1). Similarly to the MAP22 analysis, the full N³LL accuracy could not be achieved, because sets of collinear pion PDFs are presently available only at NLO accuracy. Therefore, we extract pion TMDs at N³LL⁻ accuracy.

As already done for the MAP22 global fit, we propagate the errors on the free parameters by fitting an ensemble of 200 MC replicas of experimental data. For consistency with the MAP22 analysis, we used the proton TMD PDFs extracted in that fit and we associate the i–th replica of quark TMD PDFs in the pion to the same replica in the ensemble of quark TMD PDFs in the proton.

4.2.1 Nonperturbative models of TMD PDFs

The cross section of the pion-induced DY process involves the TMD PDFs $\hat{f}_{1p}^a$ and $\hat{f}_{1\pi}^a$ of a quark a in the proton and in the pion, respectively. Actually, the target of the included set is a nucleus (tungsten) and, thus, we should take into account unpolarized quark TMDs in a nucleus. However, in our analysis a nucleus is described as a incoherent sum of free nucleons in first approximation, even though possible nuclear modifications have been recently studied [133]. Nuclear effects in pion TMDs analysis have been introduced for the first time in Ref. [134].

As for the proton, we use the MAP22 global extraction for the unpolarized TMD PDFs in the proton $\hat{f}_{1p}^a$ (see Eq. (3.5)). The parameters of proton TMDs are kept fixed to the values extracted from the MAP22 fit. For consistency, we keep the same $b_*(\boldsymbol{b}_T^2)$ prescription also for pion TMDs, but we parametrize their nonperturbative part f_{1NP}^π as follows:

$$f_{1NP}^\pi(x, \boldsymbol{b}_T^2; \zeta) = e^{-g_{1\pi}(x)\frac{\boldsymbol{b}_T^2}{4}} \left[\frac{\zeta}{Q_0}\right]^{g_K(\boldsymbol{b}_T^2)/2} = e^{-g_{1\pi}(x)\frac{\boldsymbol{b}_T^2}{4}} \left[\frac{\zeta}{Q_0}\right]^{-g_2\frac{\boldsymbol{b}_T^2}{4}}, \qquad (4.2)$$

where Q_0 is an arbitrary starting scale that we choose to be 1 GeV, and g_2 is the parameter related to the nonperturbative part of the Collins-Soper kernel. Consistently, we fix it to the value extracted by the MAP22 fit, since the Collins-Soper kernel is independent of the hadron. The x-dependence of the width $g_{1\pi}$ is given by

$$g_{1\pi}(x) = N_{1\pi} \frac{x^{\sigma_\pi}(1-x)^{\alpha_\pi^2}}{\hat{x}^{\sigma_\pi}(1-\hat{x})^{\alpha_\pi^2}}, \qquad (4.3)$$

with $\hat{x} = 0.1$. In conclusion, only 3 free parameters have to be fitted to data: $N_{1\pi}$, σ_π and α_π.

4.2.2 Fit results

Consistently with the analysis in Ch. 3, we choose the χ^2 value of the best fit to the unfluctuated data, χ_0^2, as the most representative indicator of the quality of our fit. We report in Tab. 4.2 uncorrelated and penalty components of χ_0^2 normalized to the number of data points surviving the kinematical cuts (N_{dat}).

Experiments	N_{dat}	χ_D^2/N_{dat}	$\chi_\lambda^2/N_{\text{dat}}$	χ_0^2/N_{dat}
E537	64	1.00	0.57	1.57
E615	74	0.31	1.22	1.53
Total	138	0.63	0.92	1.55

Table 4.2: The break up of the central replica χ_0^2 into components related to uncorrelated (χ_D^2) and correlated (χ_λ^2) uncertainties, normalized to the number of data points surviving the kinematical cuts (N_{dat}).

In Tab. 4.2, we observe that the value of the total penalty $\chi_\lambda^2/N_{\text{dat}}$ is very large. This result suggest that the comparison between experimental data and theoretical predictions is strongly affected by normalization uncertainties. In fact, the small values of the uncorrelated chi-squared χ_D^2 indicate that our theoretical predictions can nicely describe the shape of the experimental data. The large value of the penalty χ_λ^2 is probably caused by the large correlated systematic uncertainties affecting the experimental data sets ($\sim 16\%$). In fact, the theoretical errors related to pion and proton PDFs uncertainty are not larger than 5-8%. Our result is comparable with Ref. [280], where the same issue was observed for the E615 data set.

We remark that low-energy fixed-target DY data are properly described in global fits of unpolarized proton TMDs (see, *e.g.*, Refs. [1, 80, 81, 87]). Therefore, this issue seems not to be related to the low energy scale of the E615 and E537 experiments. In Ref. [280], it was suggested that the issue could be related to a wrong normalization of the experimental data.

In Fig. 4.2, we show the comparison between the result of our fit (colored band) and experimental data (black points) for a selection of Q bins of the E615 data set. In the upper panels, the differential DY cross section is shown as a function of the transverse momentum $|\boldsymbol{q}_T|$ of the virtual vector boson, in the lower panels the ratio between theory and data is displayed. The uncertainty bands correspond to the 68% C.L.[1] As mentioned above, we note that the shape of the experimental data is very well reproduced. The broad error bands are a consequence of the large correlated systematic errors of the considered data set.

In Fig. 4.3, we illustrate the same kind of comparison as in the previous figure but for the E537 data set. Again, in the upper panels the uncertainty

---

[1]Consistently with the conventions introduced in Ch. 3, the 68% C.L. is built by excluding the largest and smallest 16% of the replicas.
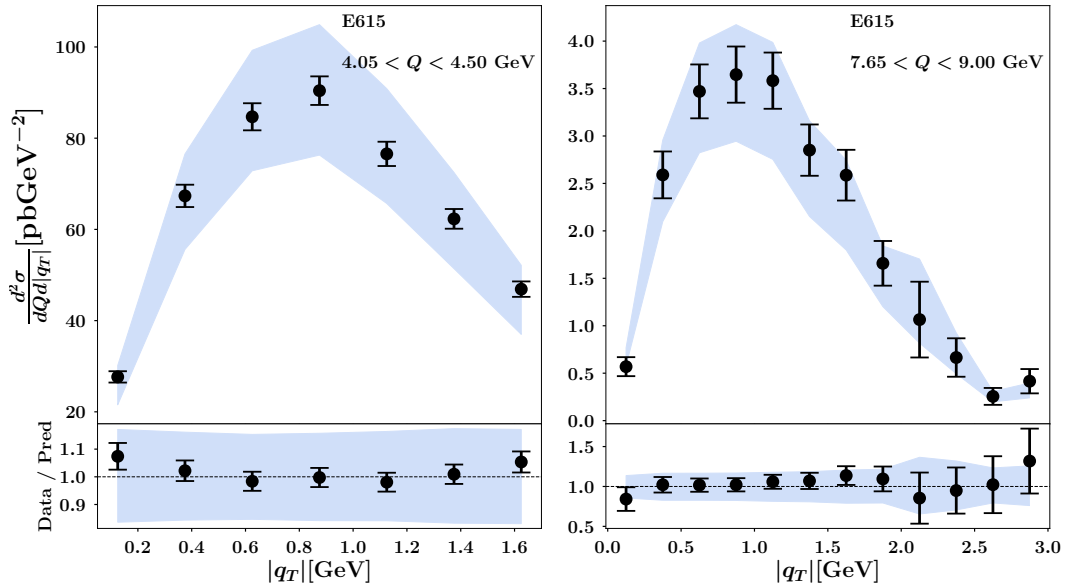
Figure 4.2: Comparison between data (black points) and our fit (colored band) for two different Q bins of the E615 data set. Upper panels: DY cross section differential in $|q_T|$; lower panels: ratio between data and results of the fit. Uncertainty bands correspond to the 68% C.L.

bands at 68% C.L. from our fit are compared to experimental data (black points) for the DY cross section as function of q_T^2 for two different Q bins, while the lower panels display the ratio between fit results and data.

The overall quality of the fit slightly worsens for the E537 data set. Despite a similar value of χ_0^2/N_{dat}, the information we get from the partial contributions to the χ^2 of the E537 data set is different from the E615 one. To be more precise, we observe that the component χ_D^2, related to uncorrelated uncertainties, is three times larger in the E537 case and contributes significantly to χ_0^2. This means that our fit does not appropriately describe the shape of the data points, and is reflected in the inadequate description of the data points at low q_T^2 in Fig. 4.3. This discrepancy may be attributed to the limited flexibility of our parametrization for the nonperturbative part of the quark TMD PDF in the pion. We explored various models for f_{1NP}^π, but no substantial modifications in the final results were observed.

On the other side, the contribution of the penalty term χ_λ^2 is less than half of the E615 one. Indeed, this may be due to the fact that the correlated systematic uncertainties of the E537 data set are much smaller, and it indicates that the normalization problem between theory and data is more under control. Consequently, in Fig. 4.3 the fit has much smaller uncertainty bands.

We also investigated the behavior of our fit when reducing the level of precision in the theoretical predictions. As already seen in the PV19 analysis for proton TMDs, we obtain a nice convergence of the NLL and N^{2}LL results
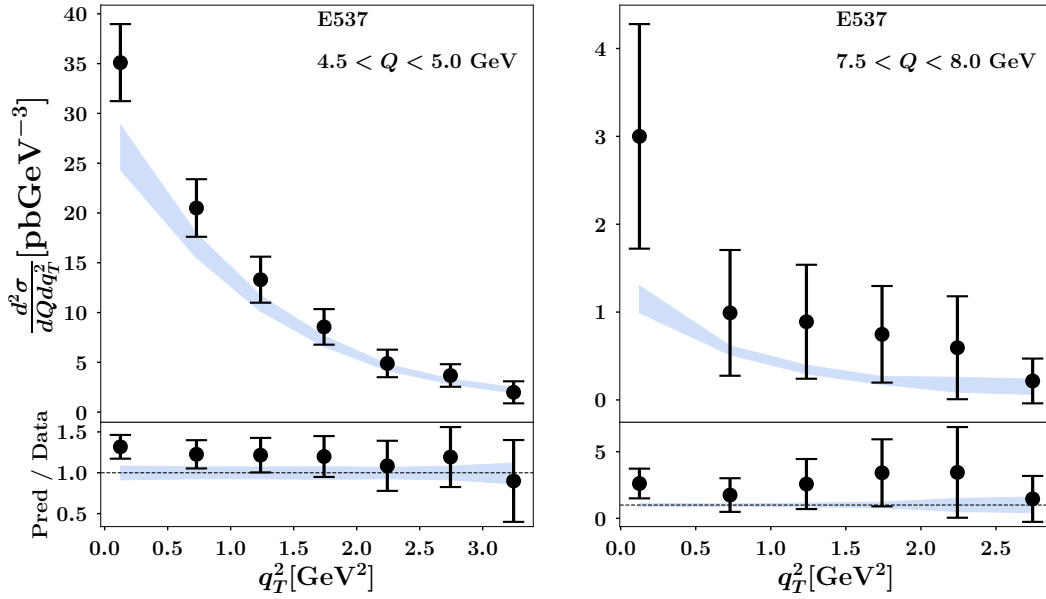
Figure 4.3: Upper panels: comparison between data and theoretical predictions for the DY cross section differential in $\boldsymbol{q}_T^2$ for the E537 data set for different Q bins; uncertainty bands correspond to the 68% C.L. Lower panels: ratio between experimental data and theoretical cross section.

to the N^{3}LL in Tab. 4.2.[2] Moreover, we find that the best fit values of the free parameters remain always within the uncertainty bands of the fit at N^{3}LL$^-$, which is a clear indication of the stability of our results.

4.2.3 Extracted pion TMDs

In this section, we discuss the quark TMD PDFs in the pion extracted from our fit at N^{3}LL$^-$ accuracy. In Tab. 4.3, we report the average values and the standard deviations of the three fitting parameters describing the nonperturbative part of the TMD PDF in Eqs. (4.2) and (4.3). Given their large uncertainties (sometimes of the order of 50%), we note that the parameters are not well constrained. We stress that similar or even worse results have been obtained by testing different ansätze of the nonperturbative parametrization. Therefore, we conclude that the available data set shows limited sensitivity to these degrees of freedom, and additional data are needed to achieve more robust constraints on them.

A visual representation of the uncertainty on the determination of the nonperturbative parameters is reported in Fig. 4.4, where the distribution of 200 replicas is shown.

Moreover, the result of the fit shows strong correlations among the three different parameters. A visual representation of the correlation matrix is dis-

---

[2]The values of the $\chi_0^2/N_{\mathrm{dat}}$ are 2.00 at NLL and 1.72 at N^{2}LL.

Parameter	Average $\pm\sigma$
$N_{1\pi}$ [GeV^2]	0.47 ± 0.12
σ_π	4.50 ± 2.25
α_π	4.40 ± 1.34

Table 4.3: Average and one standard deviation over the MC replicas of the fitting parameters in the nonperturbative part of the quark TMD PDFs in a pion.
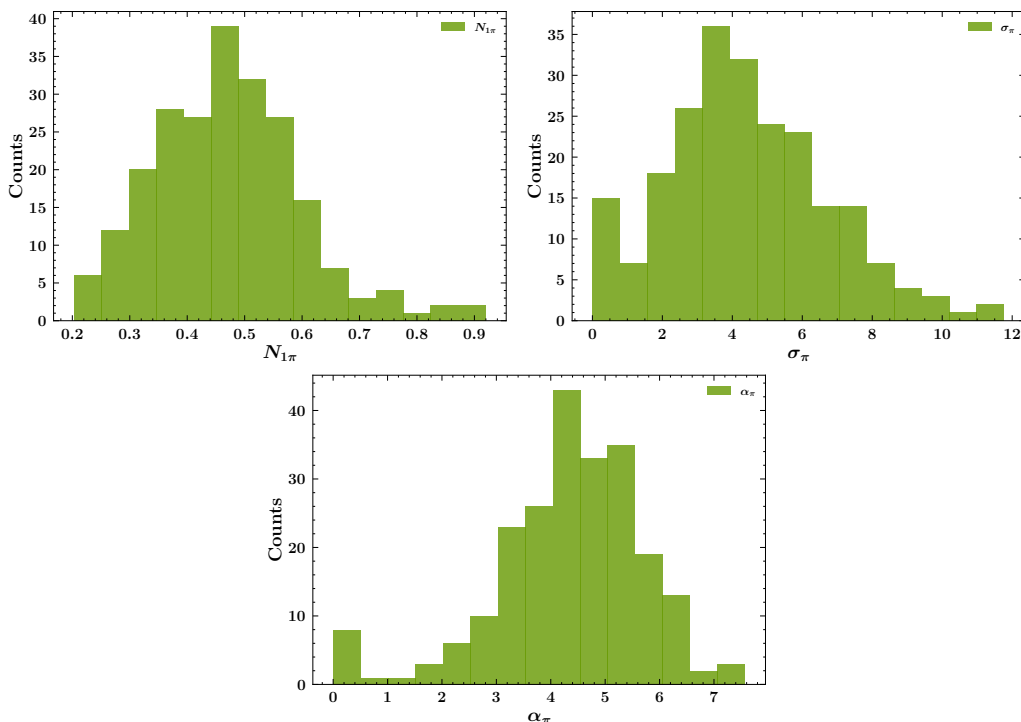
Figure 4.4: Distribution of the 200 replicas of the PionMAP22 free parameters.

played in Fig. 4.5.

We observe that the pion TMD parameters (lower right corner) are strongly correlated or anti-correlated. This could be an indication of a non-optimal parameterization. However, we obtained similar results with different models, and we find that the expression in Eq. (4.2) is reasonable. Strong correlations have been found also in the analyses of Ref. [134, 280]. Once more, this result points toward the need for new experimental data to better constrain the quark TMDs in the pion.

We also note that there is no correlation between the proton TMD parameters and the pion TMD ones. This means that the internal dynamics of quarks inside proton and pions are totally independent.

In Fig. 4.6, we display the unpolarized TMD PDF for a d quark in π^- at $\mu = \sqrt{\zeta} = Q = 5$ GeV (left panel) and 10 GeV (right panel) as a function of
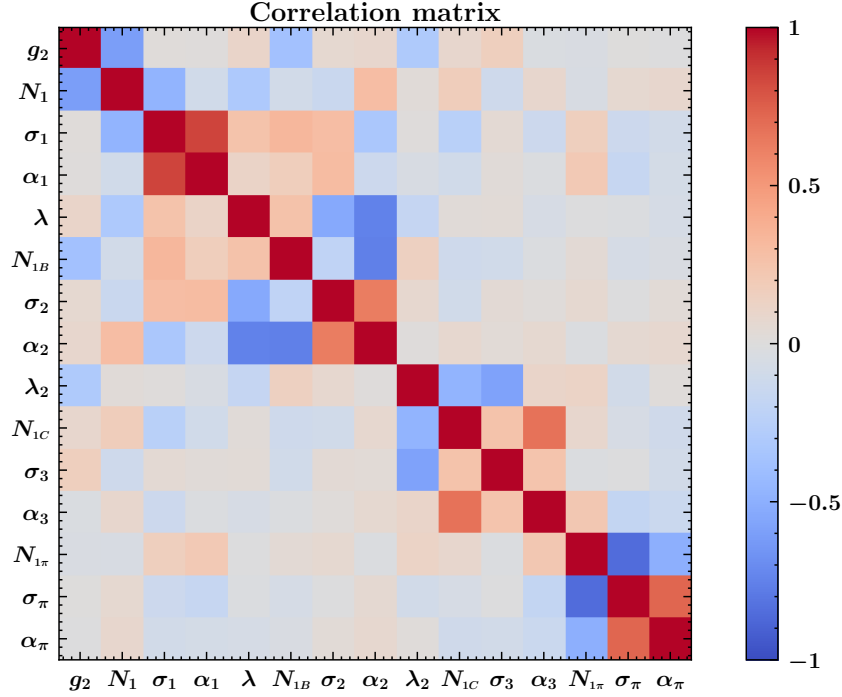
Figure 4.5: Graphical representation of the correlation matrix for the fixed proton TMD parameters and the fitted pion TMD parameters.
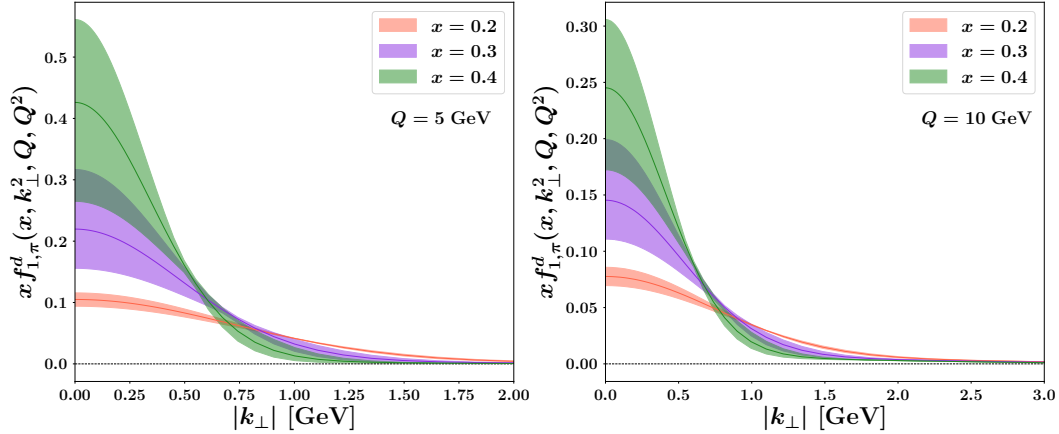
Figure 4.6: The TMD PDF of the down quark in π^- at $\mu = \sqrt{\zeta} = Q = 5$ GeV (left panel) and 10 GeV (right panel) as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$ for $x = 0.2$, 0.3 and 0.4. The uncertainty bands correspond to the 68% C.L.

the quark transverse momentum $|\boldsymbol{k}_\perp|$ for three different values of $x = 0.2$, 0.3, and 0.4. These three values of x have been selected to show the TMD PDF in the region covered by the experimental data (see Fig. 4.1). The error bands correspond to the 68% C.L. We stress that in this plot the uncertainty related to collinear PDFs is only partially accounted, since the the integral in $\boldsymbol{k}_\perp$ of

all TMD replicas is fixed to the central replica of collinear pion PDFs.

We notice that in both the left and right panels of Fig. 4.6 the TMD PDF at $x = 0.3$ and 0.4 shows the largest uncertainties, particularly at small values of $|\boldsymbol{k}_\perp|$. This may be due to the limited number of data sets that cover the considered kinematic region. Future data from the COMPASS collaboration are expected to improve the knowledge of pion TMDs in this region of the phase space.

Another interesting aspect that can be noted in Fig 4.6 is that the width of the TMD distributions decreases with increasing longitudinal momentum fraction x. This reflects the fact that at lower values of the fraction of longitudinal momentum carried by the quark there is more phase space for the intrinsic transverse momentum.
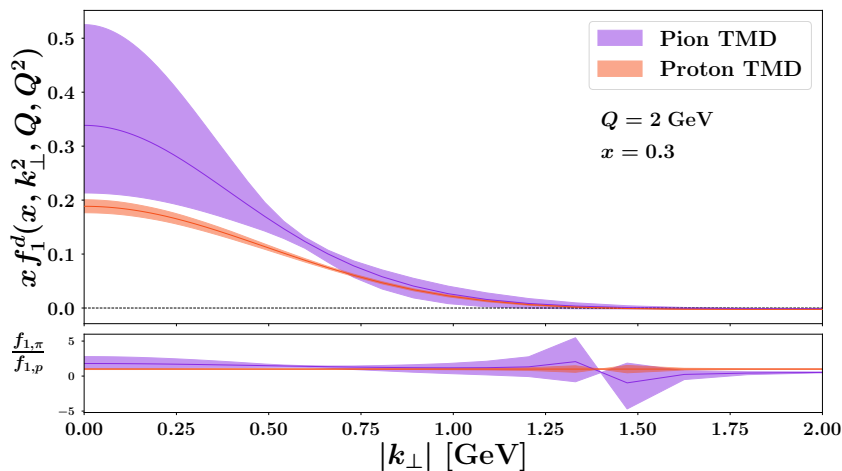
Figure 4.7: Comparison between pion and proton TMDs. Upper plot: the TMD PDF of the up quark in a pion (purple band) and in a proton (red band) at $\mu = \sqrt{\zeta} = Q = 2$ GeV and $x = 0.3$ as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$. We consider the MAP22 [1] extraction for the proton TMD (see Ch. 3). Lower plot: ratio between the pion TMD band and the proton one normalized to the central value of the latter. The uncertainty bands represent the 68% C.L.

The parameter $N_{1\pi}$ of our fit identifies the mean squared transverse momentum of quarks in the pion at $Q = 1$ GeV and $x = 0.1$. It turns out to be $\langle k_\perp^2 \rangle = 0.47 \pm 0.12$, which is larger than the same quantity for the proton (see Fig. 3.19 in Ch. 3). Therefore, the results of our fit favor wider TMDs of quarks in the pion than in the proton.

In Fig. 4.7, we show the comparison between the 68% C.L. error bands of the unpolarized d-quark TMD in a pion and in a proton (MAP22) at $\mu = \sqrt{\zeta} = Q = 2$ GeV and $x = 0.3$, as a function of the quark transverse momentum $|\boldsymbol{k}_\perp|$. As already discussed, the pion TMD is wider that the proton one. In general, the differences between the two distributions may be an effect of different behaviors in x and Q of proton and pion collinear PDFs. A similar result has

been obtained in Ref. [280].

4.3 Predictions for new experiments

The COMPASS Collaboration has recently released data for (un)polarized azimuthal asymmetries in the (polarized) pion–induced Drell–Yan processes [282, 283], and in the near future will probably release also data for the unpolarized cross section differential in $|\boldsymbol{q}_T|$.
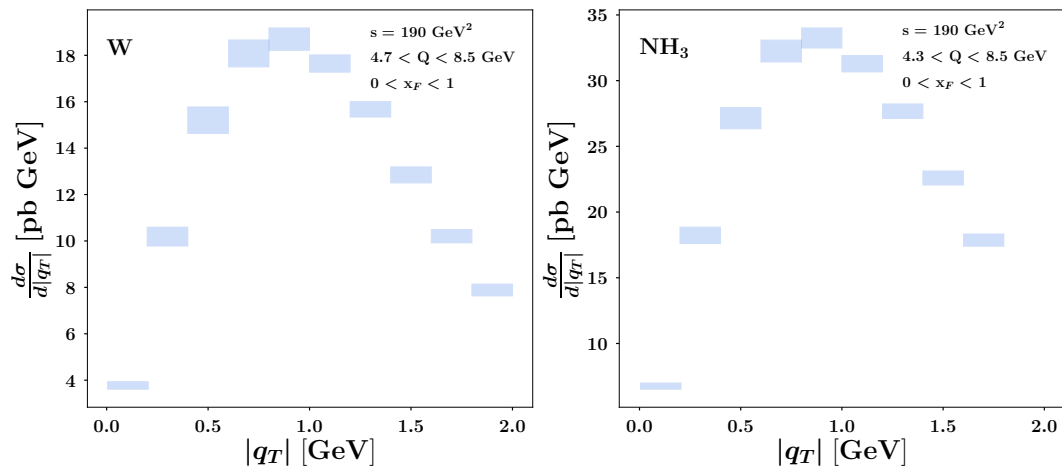
Figure 4.8: Theoretical predictions (based on our fit) for the unpolarized pion–nucleus DY cross section as function of the virtual vector boson transverse momentum $|\boldsymbol{q}_T|$, based on the PionMAP22 fit. Left panel for tungsten (W) nucleus, right panel for ammonia molecule (NH_3). Uncertainty bands correspond to 68% C.L.

In Fig. 4.8, we report the theoretical predictions based on our fit for the unpolarized pion–nucleus DY cross section as function of $|\boldsymbol{q}_T|$. The left panel refers to the tungsten (W) nucleus, while the right panel to the ammonia molecule (NH_3). We explored the same kinematic region of Ref. [283], which is quite similar to the one covered by the data sets included in the PionMAP22 fit (see Fig. 4.1). We can see that the uncertainty bands (68% C.L.) are evidently large, showing that the current information on the internal structure of the pion is not sufficient to make accurate predictions. We hope that the upcoming COMPASS data for pion-induce DY process will help to better constrain the TMD PDFs in the pion, and will shed light on the normalization issue between theory and experimental measurements observed for the E615 data set.

Chapter 5

Variations of the MAP22 TMD analysis

In the previous chapters, we discussed the results of the latest extractions of proton and pion TMDs by the MAP Collaboration [1, 2]. In both the analyses, the same theory and fitting configurations have been considered. In the case of the MAP22 fit of proton TMDs, we discussed some variations on the perturbative accuracy, on the choice of the b_* prescription, and on the cut that defines which data can be considered in the TMD region. In this chapter, we perform two exploratotry studies to assess the impact of more profound modifications in the settings of our TMD phenomenological analyses.

A first possible modification concerns the choice of collinear PDFs, which are a fundamental ingredient of TMD distributions (see, *e.g.*, Eq. (2.54)). In fact, the matching of TMDs onto their collinear counterpart at small $|\boldsymbol{b}_T|$ influences in a crucial way the predictions of the cross section, especially at large values of transverse momentum, but still within the TMD region. Therefore, the results of a TMD extraction are deeply correlated to the collinear PDFs involved in the fit. In principle, collinear and TMD PDFs (or FFs) should be simultaneously extracted in a global fit including data for both collinear and TMD kinematics. However, in the present TMD analysis collinear PDFs (or FFs) are treated as an independent input that we cannot control.

The issue of the choice of collinear PDFs has been already discussed in Refs. [1, 80, 84, 87], and it has been more sistematically studied in Ref. [88]. In this latter work, the authors show how the global quality of their analysis changes for different choices of sets of collinear PDFs, and they find that a modification of the collinear input has a non-negligible impact on the global agreement between theory and experiment. In this chapter, we perform a similar analysis and conclude that in the MAP framework the results associated to different choices of collinear PDFs are comparable.

Another important issue is the compatibility among the experimental inputs of a TMD fit. This study is inspired by a similar analysis reported in Ref. [30], where the authors discuss some selection criteria adopted to con-

117

struct the baseline data set for PDFs extractions. At variance with the previous literature, where very subjective selection criteria have been adopted, Ref. [30] discusses objective criteria to identify an optimal and self-consistent data set for PDF determinations.

In our exploratory study, we discuss the procedure to assess whether an experimental set is or is not to be included in the fitting data set. This procedure relies on the possibility of performing a weighted-fit of problematic experimental sets. In this way, we can point out the existence of possible tensions between selected experimental measurements and the rest of the data set. We focus our attention on ATLAS data, which are described in a poor way in the MAP22 global fit. We prove that they have to be treated carefully, because they deteriorate the value of the global χ^2 of a TMD analysis without having a significant impact on the result of the fit. This is the first study on this issue in TMD phenomenology, and it opens the door to further, more refined investigations on each of the kinematic bins of the ATLAS data set.

In order to estimate the effect of the above changes, we focus on Drell–Yan (DY) data. Moreover, in order to establish a baseline fit for comparison, we first of all perform a fit with the same choices of the MAP22 analysis, but restricted to DY data sets.

The chapter is organized as follows. In Sec. 5.1, we describe a new fit of only DY data with the same setup of the MAP22 global analysis. In Sec. 5.2 we present the results of another fit of the same data set, but by choosing a different collinear PDFs set. In particular, we estimate the impact of the **NNPDF3.1** collinear PDFs set on the extraction of TMD PDFs. Finally, Sec. 5.3 is devoted to a discussion on the determination of possible inconsistencies among experimental measurements included in a TMD analysis.

5.1 Fit of Drell–Yan data

In this section, we discuss the results of a new fit, henceforth named MAP22DY, with the same setup of the MAP22 analysis but using only Drell–Yan (DY) data. In other words, this fit adopts the theory settings discussed in Ch 2, it is based on the methodology described in Ch. 3, and it introduces the kinematic cuts discussed in Sec. 3.1.1. As for the nonperturbative part of the unpolarized proton TMD PDFs, we use the MAP22 phenomenological model $f_{1\,NP}(x, \boldsymbol{b}_T^2; \zeta, Q_0)$ (see Eq. (3.5)). We propagate the errors on the model parameters by fitting an ensemble of 200 Monte Carlo (MC) replicas of the experimental data.

Such fit on DY data is useful to perform investigations on possible modifications of the global settings without undertaking the massive computational effort required by SIDIS data. If benefits are found, we plan to extend the study to the full data set of DY and SIDIS measurements.

5.1.1 Results of the fit

The breakdown of the χ_0^2 values of the central (unfluctuated) replica normalized to the number of data points (N_{dat}) for the MAP22DY fit is reported in Tab. 5.3 for each of the included experimental data sets.

Data set	N_{dat}	MAP22DY fit χ_D^2	χ_λ^2	χ_0^2
CDF Run I	25	0.59	0.06	0.65
CDF Run II	26	1.73	0.02	1.75
D0 Run I	12	0.8	0.02	0.82
D0 Run II	5	1.18	0.49	1.67
D0 Run II (μ)	3	5.16	0.6	5.77
LHCb 7 TeV	7	1.24	0.31	1.55
LHCb 8 TeV	7	0.6	0.13	0.72
LHCb 13 TeV	7	1.27	0.03	1.30
ATLAS 7 TeV	18	3.99	0.71	4.70
ATLAS 8 TeV	48	1.96	0.24	2.20
ATLAS 13 TeV	6	2.84	0.46	3.30
CMS 7 TeV	4	2.09	0.06	2.16
CMS 8 TeV	4	1.53	0.0	1.54
CMS 13 TeV	70	0.15	0.01	0.16
PHENIX 200	2	6.24	1.98	8.21
STAR 510	7	1.2	0.06	1.26
E288 200 GeV	30	0.22	0.38	0.60
E288 300 GeV	39	0.45	0.04	0.49
E288 400 GeV	61	0.46	0.08	0.54
E772	53	1.28	0.7	1.98
E605	50	0.82	0.2	1.02
Total	**484**	**1.04**	**0.22**	**1.26**

Table 5.1: Breakdown of the values of χ_0^2 normalized to the number of data points N_{dat} that survive the kinematic cuts for all the DY data sets considered in the MAP22DY fit. The χ_D^2 refers to uncorrelated uncertainties, χ_λ^2 is the penalty term due to correlated uncertainties (see Eq. (3.9)), χ_0^2 is the sum of χ_D^2 and χ_λ^2.

We obtain a global value of $\chi_0^2/N_{\mathrm{dat}} = 1.26$. Since SIDIS data are excluded by the MAP22DY data set, the $\chi_0^2/N_{\mathrm{dat}}$ value of this analysis is smaller than the partial result of the MAP22 global fit on the Drell–Yan subset ($\chi_0^2/N_{\mathrm{dat}} = 1.67$), as expected.

As already noted in Sec. 3.2.2 for the MAP22 global fit, the $\chi_0^2/N_{\mathrm{dat}}$ obtained in this MAP22DY analysis is larger than the one in PV19 [81], due to the different treatment of theoretical uncertainties related to collinear PDFs.

Indeed, here they are considered as 80% correlated (see Sec. 3.1), while in the PV19 fit they were considered fully uncorrelated.

Although the penalty term χ^2_λ is expected to be of modest size, we notice from Tab. 5.1 that it is sometimes of the same size of the uncorrelated contribution χ^2_D. This is a consequence of the fact that the shifts produced by correlated uncertainties assume often large values.

We observe that the quality of our fit for the ATLAS data sets is poor. In particular, the value of the χ^2_0 increases for low-rapidity bins, the worst case being at $|y| < 1$ for ATLAS 7 TeV, in agreement with the MAP22 results. It is interesting to stress that the low level of agreement between our theoretical predictions and ATLAS data is still present even without the inclusion of SIDIS data sets in the analysis. Such result suggests to investigate the compatibility of ATLAS data with the rest of the DY data set.

We note that also PHENIX and one bin of D0 measurements shows poor agreement with theoretical predictions. However, they represent less then one percent of the DY data set.

5.1.2 TMDs from only Drell–Yan data

In this section, we discuss the TMD distributions extracted from the MAP22DY fit. Tab. 5.2 lists the 12 fitting parameters, along with their average values and standard deviations, resulting from both the MAP22DY (second column) and the baseline MAP22 (third column) analyses.

Parameter	Average over replicas	
	MAP22DY fit	MAP22 fit
g_2 [GeV]	0.24 ± 0.01	0.248 ± 0.008
N_1 [GeV^2]	0.47 ± 0.11	0.316 ± 0.025
α_1	1.8 ± 0.3	1.29 ± 0.19
σ_1	0.79 ± 0.27	0.68 ± 0.13
λ [GeV^{-1}]	2065.4 ± 2064.9	1.82 ± 0.29
N_{1B} [GeV^2]	$(4.38 \pm 0.06) \times 10^{-3}$	0.134 ± 0.017
σ_2	4.12 ± 5.75	0.455 ± 0.050
α_2	8.16 ± 4.92	4.27 ± 0.31
λ_2 [GeV^{-1}]	$(2.26 \pm 0.77) \times 10^{-2}$	$(2.15 \pm 0.58) \times 10^{-2}$
N_{1C} [GeV^2]	$(5.71 \pm 5.71) \times 10^{-3}$	$(1.30 \pm 0.69) \times 10^{-2}$
σ_3	6.09 ± 6.05	12.71 ± 0.21
α_3	3.29 ± 0.75	4.27 ± 0.13

Table 5.2: Average and standard deviation over the Monte Carlo replicas of the free parameters resulting from the MAP22DY (second column) and the MAP22 (third column) fits.

From Tab. 5.2, we observe that the parameters of the MAP22DY fit are

much less constrained than the ones of the MAP22 global fit. Indeed, the 30% of the parameters are compatible with 0 at 1-σ level, and many others are affected by errors of more than 20% of their central values. Such result is due to the lack of SIDIS data in the present analysis. Indeed, the number of SIDIS experimental data points is three times larger than the number of DY ones and, consequently, they play a significant role in constraining the fitting parameters. This suggests that the phenomenological model of $f_{1\,NP}(x, \boldsymbol{b}_T^2; \zeta, Q_0)$ (see Eq. (3.5)) can be properly determined only by simultaneously fitting DY and SIDIS data and might be too flexible for only DY data.

We note that the majority of the fitting parameters of the MAP22DY fit are compatible with the ones of the MAP22 global fit, indicating that the fit to the DY data set is still consistent with the global MAP22 fit. In particular, we note from Tab. 5.2 that the values of the g_2 parameter, which is related to the nonperturbative part of the Collins–Soper (CS) evolution kernel (see Sec. 2.3.1), are nicely compatible between the two fits. This suggests that its value is mainly constrained by high-energy DY measurements, and not by low-energy fixed-target SIDIS data. As for the λ parameter, we notice that its value is much larger in the MAP22DY fit, even if very loosely determined. This may lead to a larger uncertainty of the width of the resulting TMD PDFs.

In Fig. 5.1 we show the comparison between the 68% C.L. error bands of the unpolarized TMD PDF for the up quark in the proton for and $x = 0.1$, 0.01 and 0.001 (top, middle, and bottom rows, respectively) at $\mu = \sqrt{\zeta} = Q = 10$ and 100 GeV (left and right columns, respectively) as a function of the quark transverse momentum $|\boldsymbol{k}_\perp|$ for the MAP22DY fit (pale orange band) and the MAP22 baseline fit (light blue band).

From Fig. 5.1, we can identify the differences in the results of the two analyses. We immediately see that the uncertainties of the TMD PDFs extracted from the MAP22DY fit are much larger than the MAP22 baseline fit because of the lack of SIDIS data. It visualizes the information encoded in Tab. 5.2. Also, we can see that there are no unreasonable features in the shape of MAP22DY TMD PDFs, indicating that the nonperturbative model is not too flexible and does not produce overfitting issues.

It is interesting to observe that the largest differences in the extracted TMDs from the two fits are found at $Q = 10$ GeV and $x = 0.01$, a typical kinematic region covered by SIDIS measurements that obviously play a non-negligible role.

5.2   Introduction of a Monte Carlo set of collinear PDFs

A source of systematic theoretical error in TMD global fits is given by the choice of the set of collinear PDFs (FFs) necessary to build the considered TMD PDFs (FFs). As already pointed out in Sec. 3.1.3, this choice may have
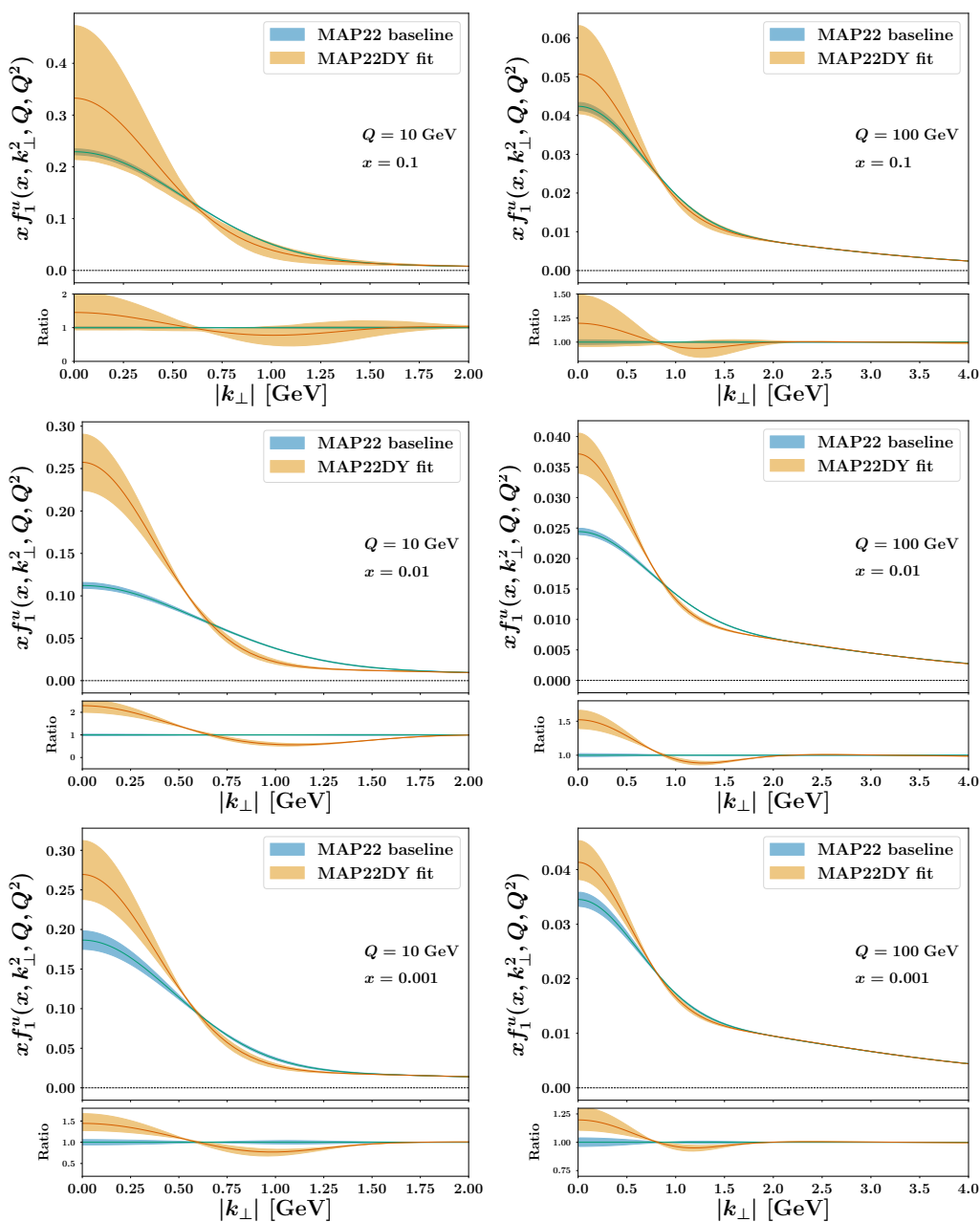
Figure 5.1: The TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 10$ GeV (left column) and 100 GeV (right column) and $x = 0.1, 0.01$ and 0.001 (from top to bottom) as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$. Upper plot: the TMD PDF of the up quark in a proton as extracted from the MAP22DY fit (pale orange band) and the MAP22 baseline fit (light blue band). Lower plot: same bands as the upper plot, but normalized to the central value of the MAP22 baseline fit. The uncertainty bands represent the 68% C.L.

a sizeable impact on the final result.

In this section, we aim to estimate the effect of modifying the choice of collinear PDFs on the extraction of TMD PDFs in the proton. As discussed in Sec. 3.1.3, in the MAP22 [1] global analysis we chose the PDFs set MMHT2014 [210]. In the study discussed in this section, we consider the NNPDF3.1 [220] set. We make this choice because this PDF set has already been considered in other unpolarized TMD analyses [82, 83, 88], and it is at the basis of state-of-the-art collinear FFs extraction (for more details, see Ref. [33]). As for the nonperturbative part of the unpolarized proton TMD PDFs, once more we use the MAP22 model for $f_{1\,NP}(x, \boldsymbol{b}_T^2; \zeta, Q_0)$ of Eq. (3.5).

5.2.1 Fit quality

In this section, we discuss the quality of the new fit, henceforth named MCDY, at full N^3LL accuracy. We impose the same cuts on kinematics as in the MAP22 analysis (for more details, see Sec. 3.1), and we propagate the errors by fitting an ensemble of 100 Monte Carlo (MC) replicas of the experimental data, associating the i–th replica of the collinear PDF set to the same replica of the experimental data. This procedure may allow a better visualization of PDF errors in TMD distributions (see Sec. 5.3.5).

We obtain a global value of $\chi_0^2/N_{\text{dat}} = 1.29$, indicating that the quality of the fit based on the NNPDF3.1 set of collinear PDFs is similar to the one based on the MMHT2014 set. In Tab. 5.3, we report the breakdown of the χ_0^2 values of the central (unfluctuated) replica normalized to the number of data points (N_{dat}). As for the MAP22DY fit, we note from Tab. 5.3 that the penalty term χ_λ^2 is sometimes of the same size of the uncorrelated contribution χ_D^2 (see discussion in Sec. 5.1).

Similarly to the MAP22DY fit, we observe once more that the quality of this new fit for the ATLAS data sets is poor. Therefore, a low level of agreement between our theoretical predictions and ATLAS data is still present even when choosing a more modern MC set of collinear PDFs. This is another indication that specific studies are needed on the compatibility of such experimental data set.

Finally, we note that in the MCDY fit the agreement between D0 (and PHENIX) experimental measurements and theoretical predictions improves significantly.

5.2.2 Extracted TMDs

In this section, we discuss the TMD distributions extracted from the MCDY fit. Tab. 5.4 reports the full list of the 12 fitting parameters, along with their average values and standard deviations, resulting from both the MCDY (second column) and the MAP22DY (third column) analyses.

From Tab. 5.4, we observe that the level of constraint of the parameters fitted with NNPDF3.1 collinear PDFs is comparable to the one obtained with the other choice. This is an indication of the stability of the fitting procedure,

Data set	N_{dat}	MCDY fit		
		χ^2_D	χ^2_λ	χ^2_0
CDF Run I	25	0.49	0.20	0.69
CDF Run II	26	1.41	0.01	1.42
D0 Run I	12	0.82	0.01	0.83
D0 Run II	5	0.83	0.05	0.88
D0 Run II (μ)	3	0.61	0.27	0.88
LHCb 7 TeV	7	2.34	2.48	4.82
LHCb 8 TeV	7	1.01	2.02	3.03
LHCb 13 TeV	7	1.35	0.07	1.42
ATLAS 7 TeV	18	4.51	1.06	5.56
ATLAS 8 TeV	48	3.05	0.66	3.7
ATLAS 13 TeV	6	3.38	1.60	4.98
CMS 7 TeV	4	1.90	0.06	1.96
CMS 8 TeV	4	1.14	0.2	1.34
CMS 13 TeV	70	0.21	0.05	0.25
PHENIX 200	2	3.21	1.40	4.61
STAR 510	7	0.92	0.31	1.23
E288 200 GeV	30	0.16	0.51	0.67
E288 300 GeV	39	0.18	0.29	0.47
E288 400 GeV	61	0.11	0.05	0.17
E772	53	0.77	0.08	0.85
E605	50	0.55	0.22	0.77
Total	**484**	**0.97**	**0.32**	**1.29**

Table 5.3: Breakdown of the values of χ^2_0 normalized to the number of data points N_{dat} that survive the kinematic cuts for all DY data sets. We use the same conventions for the χ^2_0 decomposition as in Tab. 5.1.

since it is reasonable that the constraint on the fitting parameters is determined only by the experimental data included in the analysis. We stress that for each of the two fits the included data set is the same (see Tab. 5.3).

Moreover, we note that the majority of the fitting parameters values is compatible, at list at the 1-σ level, between the two cases. This is a further indication of the stability of the fitting framework. However, any incompatibility can be caused by the effect of the differences between the two collinear PDFs sets, particularly in their shapes in the scale μ_b (see Eq. (2.54)).

In Fig. 5.2 we show the comparison between the 68% C.L. error bands of the unpolarized TMD PDF for the up quark in the proton for $x = 0.1$, 0.01 and 0.001 (top, middle, bottom rows, respectively) at $\mu = \sqrt{\zeta} = Q = 10$ and 100 GeV (left and right columns, respectively) as a function of the quark transverse momentum $|\boldsymbol{k}_\perp|$ for the MAP22DY fit (pale orange bands) and the MCDY fit (green bands).

Parameter	Average over replicas	
	NNPDF3.1 set	MMHT2014 set
g_2 [GeV]	0.16 ± 0.06	0.24 ± 0.01
N_1 [GeV^2]	2.72 ± 0.94	0.47 ± 0.11
α_1	1.36 ± 0.99	1.8 ± 0.3
σ_1	0.88 ± 1.94	0.79 ± 0.27
λ [GeV^{-1}]	25.47 ± 17.23	2065.4 ± 2064.9
N_{1B} [GeV^2]	0.12 ± 0.05	$(4.38 \pm 0.06) \times 10^{-3}$
σ_2	1.36 ± 0.76	4.12 ± 5.75
α_2	1.96 ± 1.80	8.16 ± 4.92
λ_2 [GeV^{-1}]	$(8.21 \pm 8.17) \times 10^{-3}$	$(2.26 \pm 0.77) \times 10^{-2}$
N_{1C} [GeV^2]	$(2.63 \pm 2.63) \times 10^{-5}$	$(5.71 \pm 5.71) \times 10^{-3}$
σ_3	26.8 ± 2.36	6.09 ± 6.05
α_3	4.87 ± 0.9	3.29 ± 0.75

Table 5.4: Average and standard deviation over the Monte Carlo replicas of the free parameters of the MCDY (second column) and the MAP22DY (third column) fits.

From Fig. 5.2, we can appreciate the expected differences in the results of the two analyses. Since the integral of each TMD PDF is related to the value of the collinear set of PDFs used in the extraction, the normalization of the two curves differs by construction. In fact, the integral of each TMD replica of the MAP22DY fit is fixed to the central member of the MMHT2014 collinear PDFs set, because of its Hessian nature. On the contrary, being the NNPDF3.1 set composed by an ensemble of equivalent replicas, the integral of each TMD replica of the MCDY fit is fixed to the corresponding replica of the NNPDF3.1 set, providing a better estimate of the actual uncertainty on TMD distributions.

In Fig. 5.2, we also note that the uncertainty bands of the TMDs extracted in the MCDY fit are larger than the MAP22DY case, except for the $x = 0.1$ case. This effect may reflect the different features of NNPDF3.1 with respect to the older MMHT2014.

It is interesting to observe that the TMD PDFs resulting from the two fits are compatible with each other, indicating that TMD distributions extracted from a phenomenological fit only slightly depend on their collinear input. This is reasonable because the fit should be able to adjust the fitting parameters to get the same features of TMDs that are needed to properly describe the included experimental data set. In conclusion, this study indicates that it is important to consider Monte Carlo sets of collinear PDFs in TMD analyses, because they allow to better identify the uncertainty on the *integral* of the extracted TMDs, whitout strongly affecting their shape.. This finding is somewhat different from the conclusions of Ref. [88].
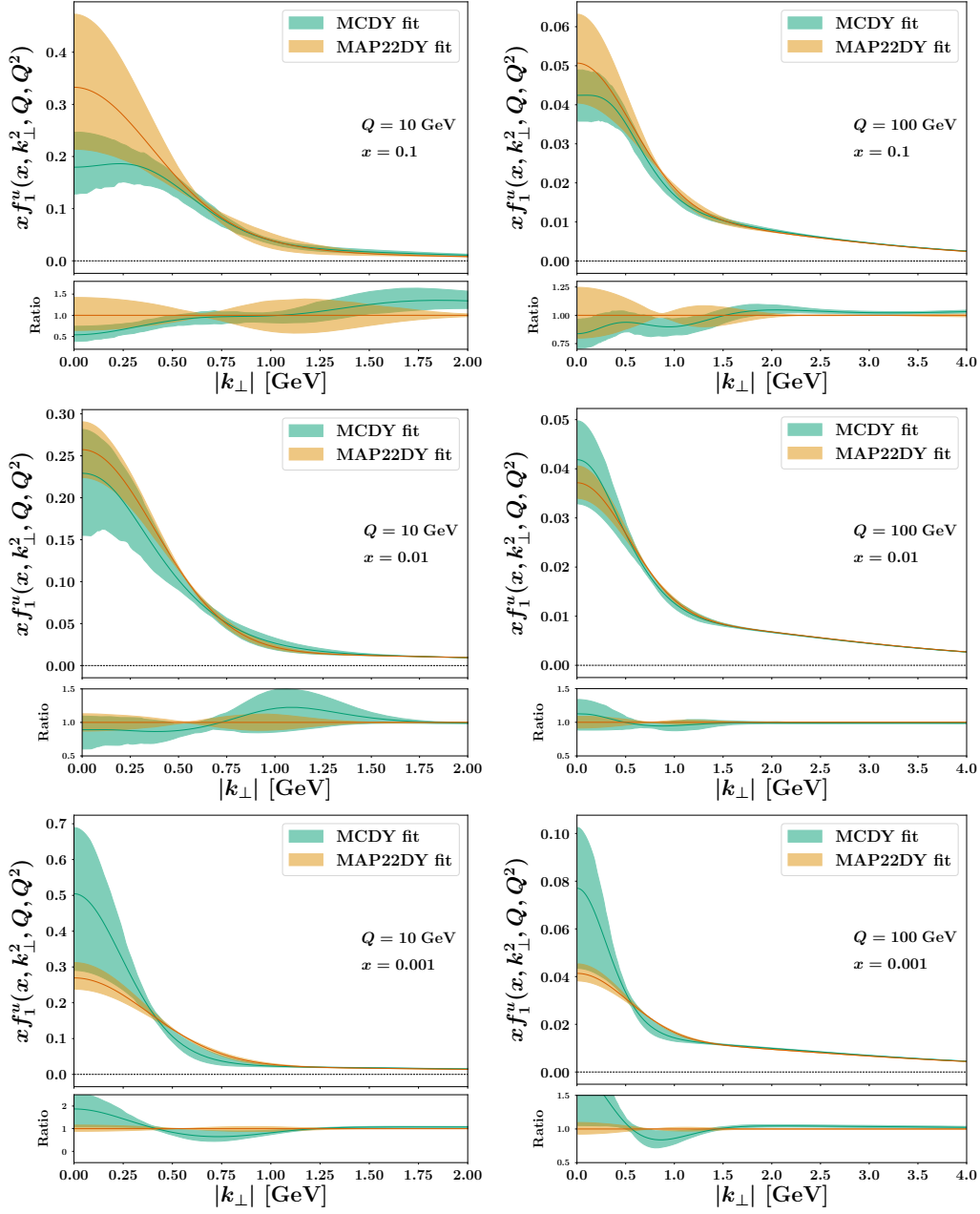
Figure 5.2: The TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 10$ GeV (left column) and 100 GeV (right column) and $x = 0.1$, 0.01 and 0.001 (from top to bottom) as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$. Upper plot: MCDY fit (green band) and the MAP22DY fit (pale orange band). Lower plot: same bands as the upper plot, but normalized to the central value of the MAP22DY fit. The uncertainty bands represent the 68% C.L.

5.3   Determination of a fully consistent data set

In this section, we discuss a possible way to identify a fully consistent data set in a phenomenological analysis, and we investigate the dependence of results on the choice of input data sets. For instance, the DY measurements and the kinematic cuts described in Sec 3.1.1 define an extended data set, and inside it we want to isolate a reduced set that are fully consistent, namely where the incompatibilities between single data sets are removed.

The new "reduced" data set is established through a weighted-fit procedure that was introduced in Ref. [30]. In this procedure, by suitable estimators we first identify data sets that are problematic in terms of fit quality. Then, we perform a dedicated weighted fit for measurements whose estimators exceed some predefined thresholds, and we exclude/keep the considered sets according to the results of this weighted fit.

5.3.1   Selection criteria

We examine the Drell-Yan data set discussed in Sec 3.1.1 with the aim to determine its internal consistency. Indeed, specific measurements may exhibit inconsistencies with the rest of the data set for various reasons: missing systematic uncertainties, underestimated experimental uncertainties, differences in the detailed definition of the reported observable, missing higher-order QCD or electroweak corrections, etc... The goal of this study is not to identify the nature of such inconsistencies, but to single out and exclude the problematic measurements from the baseline data set based on objective criteria. These measurements can be subsequently studied with dedicated fits.

We consider the MCDY fit discussed in Sec. 5.1 as baseline. The number of indicators that can be used to select the problematic data sets is large, quite subjective and potentially biased. In Ref. [30], the authors consider as reasonable estimators the χ^2 per data, the number n_σ of standard deviations of the χ^2 per data point from the unit value, and the stability metric Z [284] that estimates potential inaccuracies affecting the experimental covariance matrix. In our study, we select the full ATLAS data set as potentially problematic because it shows the largest values of χ^2/N_{dat} in both the MAP22 fit (see Tab. 3.4) and in the MAP22DY fit (see Tab. 5.1).[1] The potential inconsistence of this data set is assessed using the weighted fit method, as discussed below. This choice is a reasonable first step to a future more detailed study.

5.3.2   Weighted fit procedure

The weighted fit method relies on determining whether a specific measurement is inconsistent with the global data set by performing a new analysis that pro-

---

[1]Notice that also LHCB and PHENIX data sets may be potentially problematic, given their large values of χ^2/N_{dat} (see Tabs. 3.4-5.1). We leave the investigation of their compatibility for a future study.

vides the best agreement to this measurement. The second step consists in checking if the achieved best agreement does or does not lead to the deterioration in the description of the other data included in the global data set. Such procedure was recently implemented in Ref. [285] in order to study Standard Model parameters through a global PDF fit. Related methods are discussed in Ref. [286].

In the weighted fit discussed here, the selected data set is given a weight that, in the calculation of the χ^2 value, makes it of the same statistical weight as the rest of the global data set. In formulae, by assign the index j to the possible problematic data set, the expression of the χ^2 is modified as

$$\chi^2 = \sum_{i=1}^{N_{\text{exp}}} \chi_i^2 \qquad \longrightarrow \qquad \chi^2 = \sum_{i \neq j}^{N_{\text{exp}}} \chi_i^2 + \omega_j \chi_j^2 , \qquad (5.1)$$

where N_{exp} is the number of experimental data sets included in the fit, and χ_i^2 is the χ^2 value of the data set i. The value of the weight ω_j associated to the selected problematic measurement is usually taken to be

$$\omega_j = \frac{N_{\text{dat}}}{N_{\text{dat}}^{(j)}} , \qquad (5.2)$$

where $N_{\text{dat}}^{(j)}$ is the number of data points in the data set j. For the ATLAS data set, $N_{\text{dat}}^{(j)} = 72$, while for the entire DY data set $N_{\text{dat}} = 412$. Hence, the value of the assigned weight is $\omega_j \simeq 5.5$. In order to assess the stability of our study, we have checked that the results do not change qualitatively by repeating the procedure with two more choices, namely twice and half the nominal value.

Based on the outcomes of the weighted fit, we can identify the following scenarios [30]:

- The value of χ_j^2 does not improve considerably while the χ_i^2 of the rest of the data sets remain essentially unaltered. In this case, we conclude that the data set j shows internal inconsistencies. Further investigations are required to assess their removal from the baseline data set.

- The value of χ_j^2 does not improve considerably and the χ_i^2 of several of other data sets (including those belonging to the same process type of data set j) worsen significantly. In this case, we conclude that the internal inconsistencies of the selected data set cause a distortion of the baseline fit. We remove data set j from the baseline data set.

- The value of χ_j^2 improves considerably and the χ_i^2 of the rest of the data set lays within statistical fluctuations. In this case, we conclude that the data set j carries a small weight over the full set and is not properly fitted. We keep the data set j in the baseline data set.

128

- The value of χ_j^2 improves considerably but the χ_i^2 of several of other data sets (including those belonging to the same process type of data set j) worsen significantly. In this case, we conclude that the selected data set is inconsistent with the full data set. We remove data set j from the baseline data set.

5.3.3 Results of the weighted fit

The outcome of the weighted fit, when applied to the ATLAS data set and the full DY data set, is reported in Tab. 5.5. For a better comparison, we report in Tab. 5.5 the $\chi_0^2/N_{\mathrm{dat}}$ for both the weighted fit (third column) and the unweighted fit (fourth column). For convenience, we consider the MAP22DY fit discussed in Sec. 5.2 as the unweighted fit, but we checked that our results do not depend on this choice: we obtained similar results even considering the full MAP22 set.

Data set	N_{dat}	$\chi_0^2/N_{\mathrm{dat}}$	
		Weighted fit	Unweighted fit
CDF Run I	25	0.70	0.69
CDF Run II	26	1.42	1.42
D0 Run I	12	0.82	0.83
D0 Run II	5	0.92	0.88
D0 Run II (μ)	3	0.81	0.88
LHCb 7 TeV	7	4.86	4.82
LHCb 8 TeV	7	3.04	3.03
LHCb 13 TeV	7	1.41	1.42
CMS 7 TeV	4	1.96	1.96
CMS 8 TeV	4	1.31	1.34
CMS 13 TeV	70	0.26	0.25
PHENIX 200	2	5.00	4.61
STAR 510	7	1.26	1.23
E288 200 GeV	30	0.68	0.67
E288 300 GeV	39	0.48	0.47
E288 400 GeV	61	0.18	0.17
E772	53	0.86	0.85
E605	50	0.81	0.77
ATLAS	72	4.25	4.27
Total	484	2.48	1.29

Table 5.5: Breakdown of the values of χ_0^2 normalized to the number of data points N_{dat} for the weighted (third column) and unweighted (fourth column) fits (see text).

In Tab. 5.5, we can see that the value of the $\chi_0^2/N_{\mathrm{dat}}$ of the ATLAS data

set improves in the weighted fit from 4.27 to 4.25, while the agreement with the rest of the fitted data sets remains almost unchanged, and sometimes it worsens. These results suggest that the ATLAS measurements show "internal" inconsistencies, accordingly to the terminology of Ref. [30]. These inconsistency may be related to the detailed definition of the reported observable.[2] Therefore, further investigations are needed to assess if it must be kept or removed from the baseline data set.

It is important to remark that the global $\chi_0^2/N_{\mathrm{dat}}$ of the weighted fit almost doubles the unweighted one. Hence, the ATLAS data set considerably contributes to the the global χ_0^2. In other words, it leads to a significant deterioration of the estimator of the global agreement between theoretical predictions and experimental data, even if the results of the rest of the data sets is not affected. This represents a significant evidence that ATLAS measurements may be not compatible with the rest of the data set, and may have to be separately treated.

In the following, we will analyze the results of a new fit without the ATLAS data set. This will allow us to better estimate the impact of the discarded measurements.

5.3.4   Results of the fit with a reduced data set

In this section, we report the results of the extraction of unpolarized TMD PDFs at full N³LL accuracy from a fit to the reduced DY data set discussed above, namely without the ATLAS data set. Once more, we choose the same setup of the MCDY fit.

We obtain the value of $\chi_0^2/N_{\mathrm{dat}} = 0.72$. The small value of this figure of merit indicates that the agreement between theoretical predictions and experimental measurements included in the reduced data set is extremely good, even if its optimal value should be 1. However, a χ^2/N_{dat} smaller than 1 may be induced by large experimental errors affecting the considered data set, or by a too flexible model that leads to overfitting. Further insights on this topic will be discussed in Sec. 5.3.5.

In Tab. 5.6, we report the breakdown of the χ_0^2 values of the central (unfluctuated) replica normalized to the number of data points (N_{dat}) for both the fits to the reduced data set (third column) and the baseline data set (fourth column).

In the last row of Tab. 5.6, we list the normalized χ_0^2 of both the reduced (third column) and baseline (fourth column) data sets. We remark that the former is significantly smaller than the latter, even with a comparable number of data points. Moreover, we note that the χ_0^2 of the majority of data sets is slightly smaller than in the baseline case, or at most unchanged. This suggests

---

[2]These data sets include final-state lepton cuts that can produce a disagreement if not properly treated. A new release of ATLAS 8 TeV data sets without these cuts has been announced to be published in the near future [287].

Data set	N_{dat}	$\chi_0^2/N_{\mathrm{dat}}$	
		Reduced data set	Baseline data set
CDF Run I	25	0.66	0.69
CDF Run II	26	1.28	1.42
D0 Run I	12	0.81	0.83
D0 Run II	5	0.91	0.88
D0 Run II (μ)	3	1.26	0.88
LHCb 7 TeV	7	4.87	4.82
LHCb 8 TeV	7	3.12	3.03
LHCb 13 TeV	7	1.63	1.42
CMS 7 TeV	4	1.95	1.96
CMS 8 TeV	4	1.29	1.34
CMS 13 TeV	70	0.26	0.23
PHENIX 200	2	3.19	4.61
STAR 510	7	1.05	1.23
E288 200 GeV	30	0.62	0.67
E288 300 GeV	39	0.41	0.47
E288 400 GeV	61	0.18	0.16
E772	53	0.79	0.85
E605	50	0.58	0.77
ATLAS	72	/	4.27
Total		**0.72 (412)**	**1.29 (484)**

Table 5.6: Breakdown of the values of χ_0^2 normalized to the number of data points N_{dat} for the fits to both the reduced (third column) and baseline (fourth column) DY data sets. Last row: global χ_0^2 normalized to the number of data (reported in parentheses for the two cases).

that the ATLAS set is responsible for a significant increase of the global χ_0^2 (see the values in the last row of Tab. 5.5). Consequently, the agreement between theory and experimental data for the baseline case is generally better than indicated by the value of the global χ_0^2, which is spoiled by the presence of ATLAS data.

Therefore, we conclude that the ATLAS data have to be treated carefully and, when they are included in a global analysis, their important contribution to the value of the global χ^2 may lead to inaccurate assessments on the quality of the fit. Also, including them may imply the use of unnatural flexible models in order to reach a better agreement. We leave this issue for future more detailed studies.

5.3.5 TMD distributions

In Tab. 5.7, we report the full list of the 12 fitting parameters, along with their average values and standard deviations, for the fit to the reduced DY data set discussed in the previous section (second column) and for the MCDY fit discussed in Sec. 5.2.2 (third column).

Parameter	Average over replicas	
	Reduced data set	Baseline data set
g_2 [GeV]	0.15 ± 0.04	0.16 ± 0.06
N_1 [GeV^2]	3.33 ± 1.41	2.72 ± 0.94
α_1	2.39 ± 1.73	1.36 ± 0.99
σ_1	1.87 ± 2.39	0.88 ± 1.94
λ [GeV^{-1}]	21.57 ± 12.47	25.47 ± 17.23
N_{1B} [GeV^2]	0.14 ± 0.05	0.12 ± 0.05
σ_2	1.31 ± 1.15	1.36 ± 0.76
α_2	2.04 ± 0.84	1.96 ± 1.80
λ_2 [GeV^{-1}]	$(4.11 \pm 4.09) \times 10^{-4}$	$(8.21 \pm 8.17) \times 10^{-3}$
N_{1C} [GeV^2]	$(2.89 \pm 2.86) \times 10^{-7}$	$(2.63 \pm 2.63) \times 10^{-5}$
σ_3	35.19 ± 4.37	26.8 ± 2.36
α_3	5.67 ± 0.96	4.87 ± 0.90

Table 5.7: Average and standard deviation over the Monte Carlo replicas of the free parameters fitted to the reduced data set of Sec. 5.3.4 (second column) and to the baseline data set of Sec. 5.2.2 (third column).

From Tab. 5.7, we first remark that the majority of the parameters are affected by large errors. Moreover, the results of both data sets are all compatible at 1-σ level, except for the parameter σ_3. This suggests that the exclusion of ATLAS data from the data set considered in our TMD analyses does not lead to significant modifications of the best fit parameters.

It is interesting to note that the g_2 parameter, which describes the nonperturbative part of TMD evolution (see Sec. 2.3.1), is nicely compatible between the two fits. This indicates that ATLAS data are not of fundamental importance to properly fix the nonperturbative evolution of TMDs. Indeed, there are other experimental measurements at high energies that provide precise information, such as the D0, LHCB and CMS data sets.

We also note that the λ_2 parameter, which is related to the relative weight of the third Gaussian in the parametrization of the TMD PDFs (see Eq. (3.5)), is smaller than in the baseline case by a factor 100. Since it is squared in the formula, this indicates its negligible relevance in the nonperturbative model. Moreover, it shows strong correlations with the other parameters, suggesting that this additional Gaussian may be removed by the fitted parametrization. Further studies on this topic are left for future works.

In Fig. 5.3, we show the comparison between the 68% C.L. error bands of the fit to the reduced data set (dark orange band) and of the baseline fit (green band) for the unpolarized TMD PDF of the up quark in the proton for $x = 0.1$, 0.01 and 0.001 (top, middle, bottom rows, respectively) at $\mu = \sqrt{\zeta} = Q = 10$ and 100 GeV (left and right columns, respectively) as a function of the quark transverse momentum $|\boldsymbol{k}_\perp|$.

In principle, the two TMD PDFs are expected to be somehow different because they are extracted from two different sets of data points.

In agreement with Tab. 5.7, the two uncertainty bands are comparable, except in the low-$|\boldsymbol{k}_\perp|$ region at $Q = 10$ GeV and $x = 0.1$. Such an effect may be due to the fact that the α_2 and λ parameters are more constrained in the fit to the reduced data set (see Tab. 5.7).

We also note that the TMD PDFs extracted from the reduced data set are generally smaller in the low-$|\boldsymbol{k}_\perp|$ region. This is probably the consequence of the smaller contribution of the third Gaussian in the fit (see discussion above).

In conclusion, this study suggests that the inclusion of ATLAS measurements in our TMD analyses considerably spoils the value of the global χ^2 of the fit, while it does not significantly modify the extracted TMD PDFs. The exclusion of ATLAS data points can open up the possibility of obtaining a good fit even with a simpler parametrization for the nonperturbative model of TMDs. Therefore, ATLAS experimental data have to be treated carefully in TMD analyses.
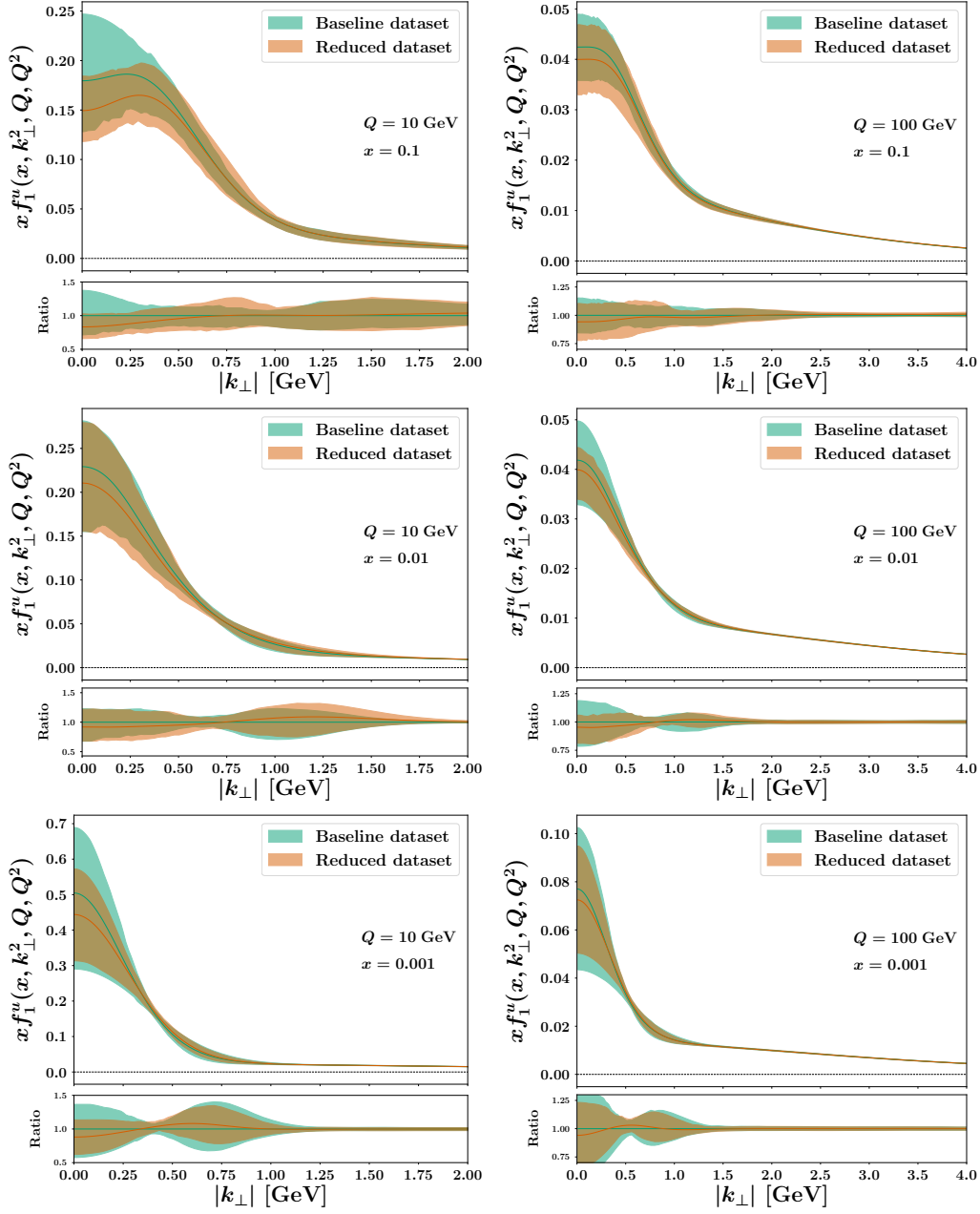
Figure 5.3: The TMD PDF of the up quark in a proton at $\mu = \sqrt{\zeta} = Q = 10$ GeV (left column) and 100 GeV (right column) and $x = 0.1$, 0.01 and 0.001 (from top to bottom) as a function of the partonic transverse momentum $|\boldsymbol{k}_\perp|$. Upper plot: fit to the baseline DY data set (green band) and to the reduced one (dark orange band). Lower plot: same bands as the upper plot, but normalized to the central value of the band of the baseline data set. The uncertainty bands represent the 68% C.L.

5.A Appendix - Comparison of MAP22DY and MCDY parameters

In this section, we report the distribution of each of the 12 fitting parameters obtained from the MAP22DY (red histograms) and the MCDY (green histograms) fits (see Tab. 5.4 for average value and standard deviation). We note that the histograms of the parameters λ and N_{1C} are barely visible for the MCDY case. This is due to the fact that their values are smaller by two order of magnitude than the MAP22DY ones (see Tab. 5.4).
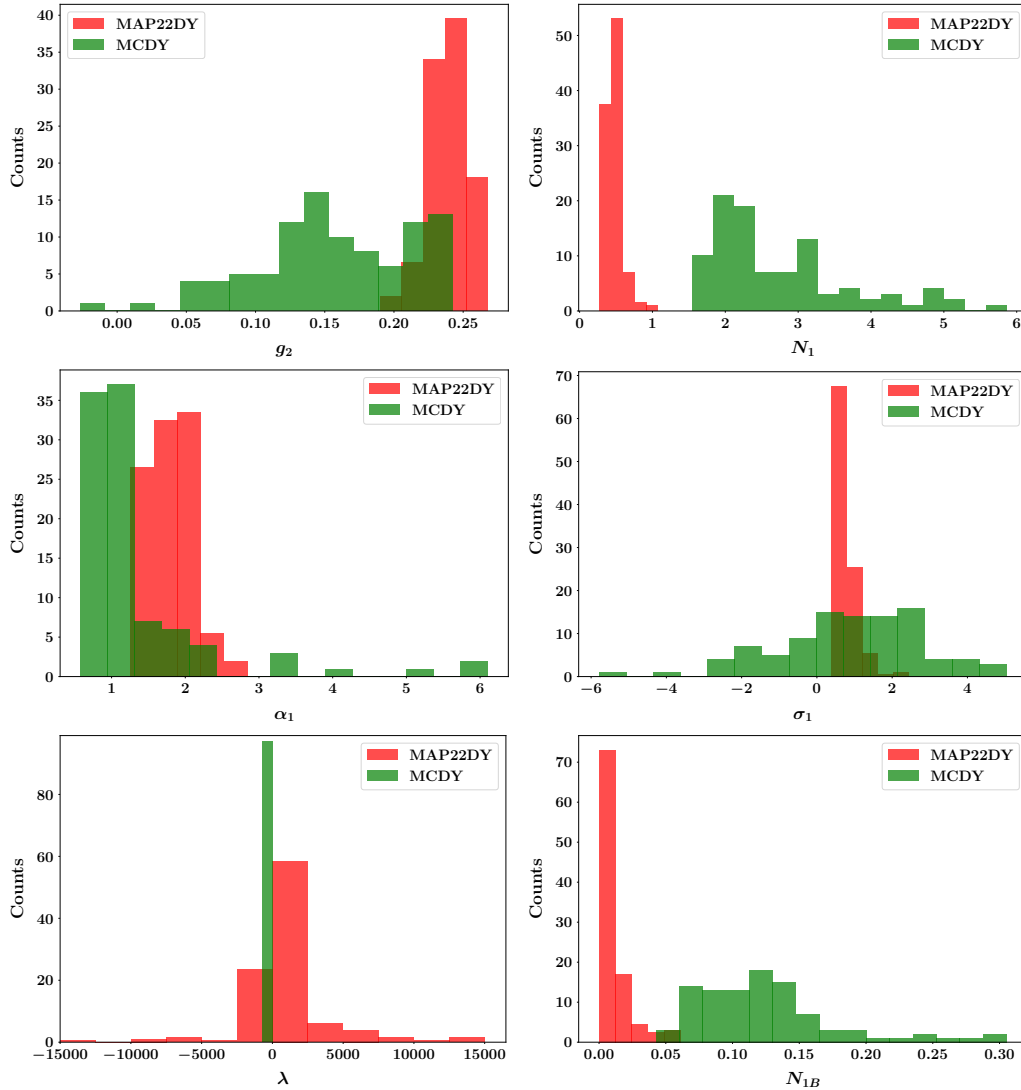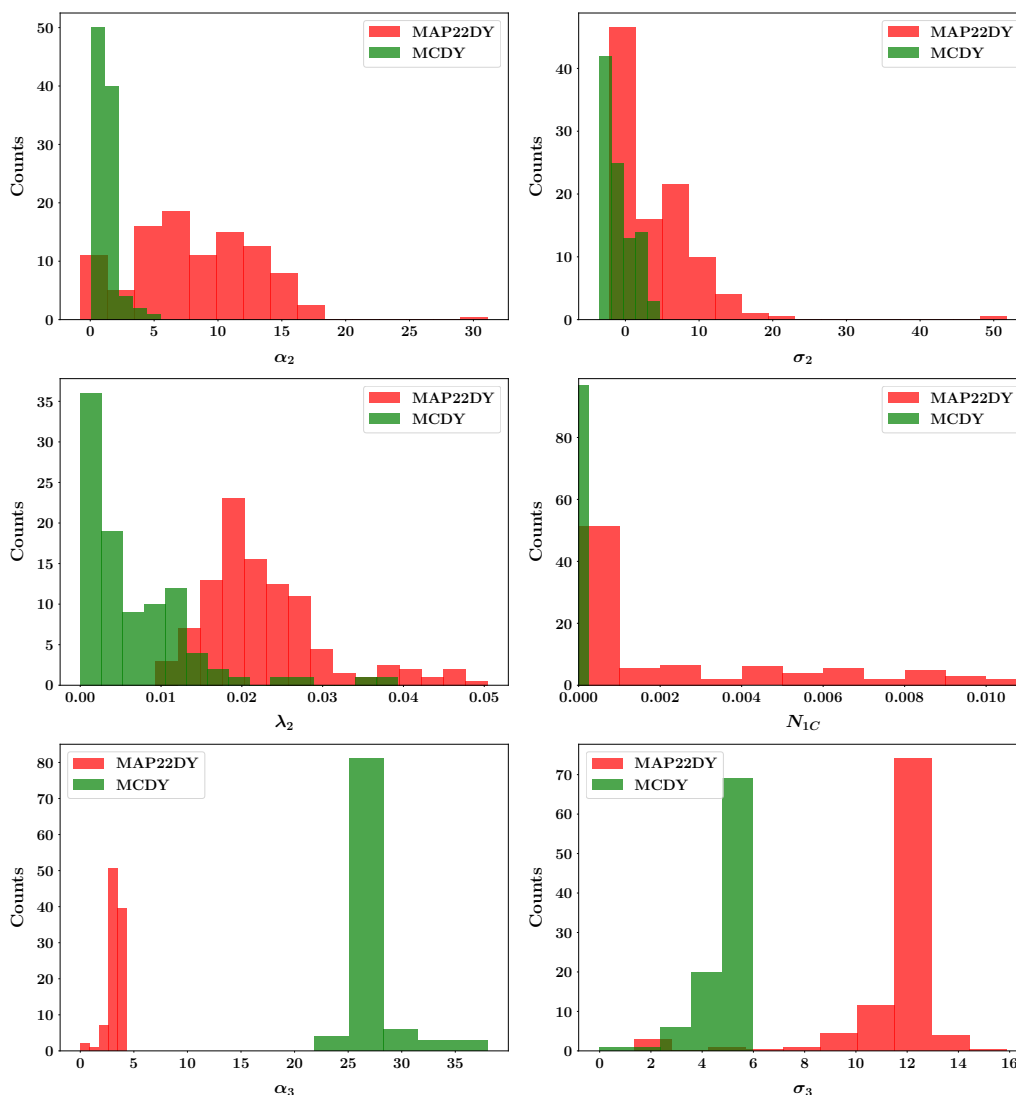
Figure 5.4: Distribution of the replicas for the MAP22DY (red histograms) and MCDY (green histograms) free parameters.

5.B Appendix - Comparison of parameters fitted to baseline and reduced DY data set

In this section, we report the distribution of each of the 12 fitting parameters obtained from the fits to the baseline (green histograms) and the reduced (blue histograms) data sets (see Tab. 5.7 for average value and standard deviation).
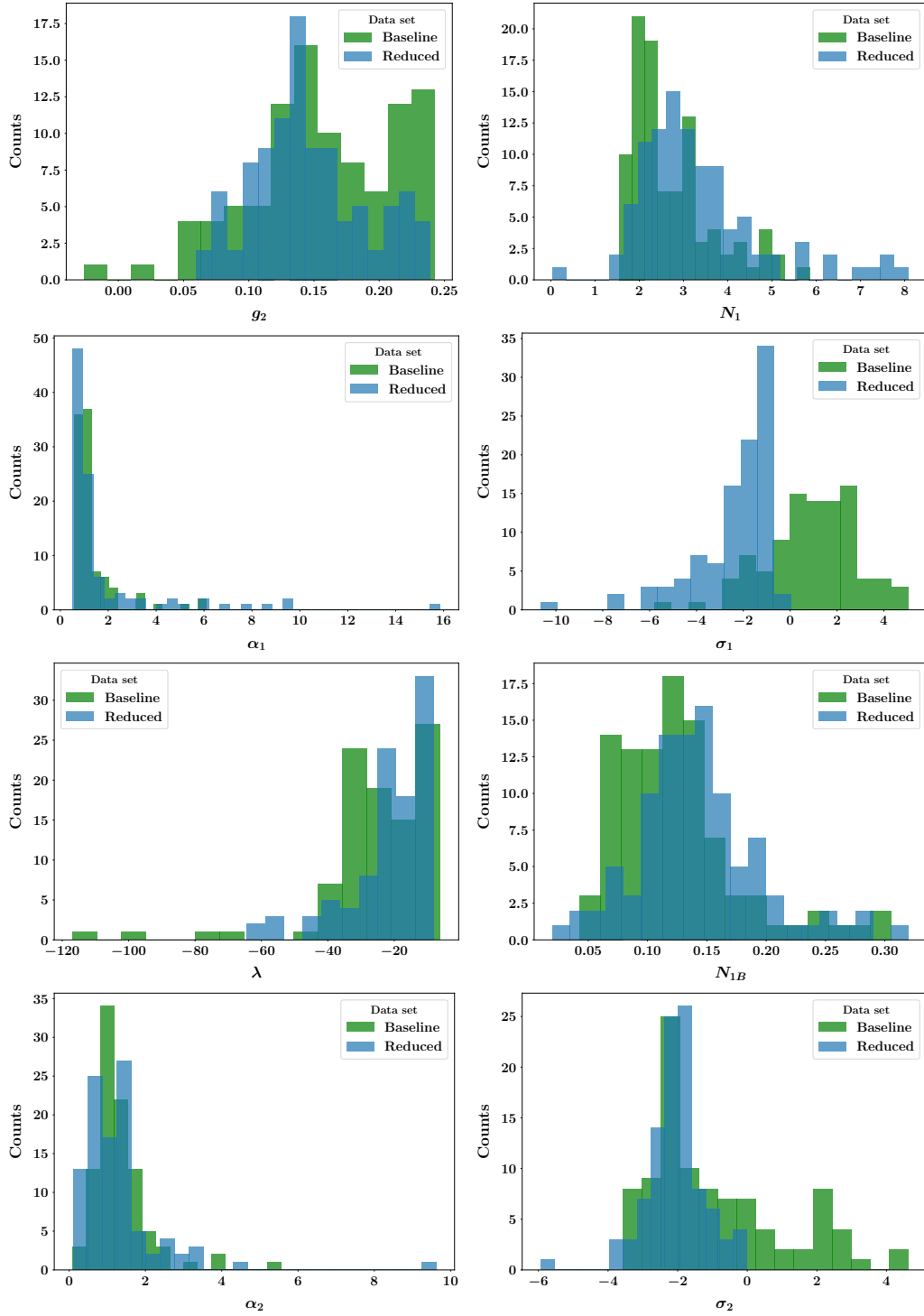
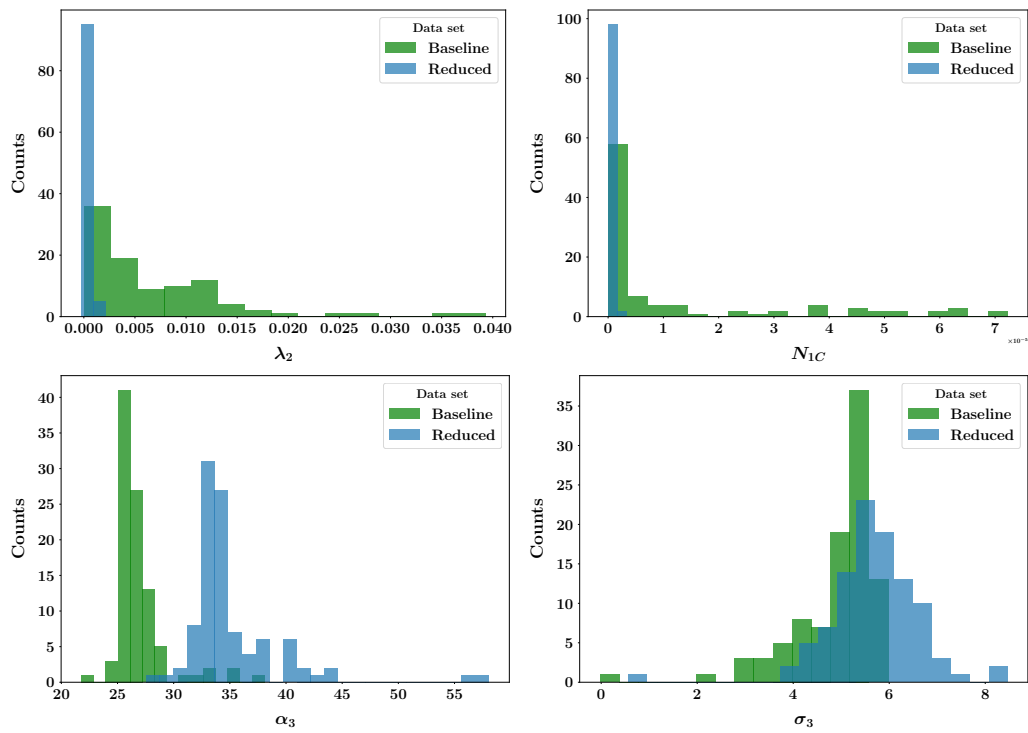Figure 5.5: Distribution of the replicas of the fit to the baseline (green histograms) and reduced (blue histograms) data sets.

Chapter 6

Parity-violating effects in the proton structure

In this chapter, we study another aspect of the internal structure of hadrons. We try to phenomenologically address the following question: *Would the internal structure of the proton be identical in a mirrored world?* Since Quantum ChromoDynamics (QCD) is invariant under parity transformations in the Standard Model (SM), one could argue that the proton structure should remain the same. However, there is no first principle in QCD that guarantees parity (P) invariance from a theoretical point of view. In this chapter, we propose a way to estimate its possible violation, *i.e.*, "strong P violation", in the internal structure of nucleons (without determining its source). We will show that an unpolarized proton may contain more left-handed than right-handed quarks.

The identification of strong P violation would have profound consequences that go far beyond the understanding of the internal structure of nucleons. In fact, the violation of the charge-parity (CP) symmetry is the most plausible explanation of the matter-antimatter imbalance observed in the Universe from its early stages. We focus on the QCD domain because it has been proven that CP violation in the SM electroweak sector is not sufficient to justify the size of such imbalance [288–290].

In order to understand this yet unexplained phenomenon, it is necessary to extend the SM by introducing new ingredients. In the last decades, different studies have been devoted to the following gauge invariant and renormalizable term:

$$L_\theta = \frac{\theta g_s^2}{64\pi^2} \epsilon_{\mu\nu\rho\sigma} \hat{F}_a^{\mu\nu} \hat{F}_a^{\rho\sigma} \, , \tag{6.1}$$

where g_s is the strong coupling, $\epsilon_{\mu\nu\rho\sigma}$ the antisymmetric tensor, and $\hat{F}_a^{\mu\nu}$ is the gauge field strnght tensor. This is the so-called "θ-term" of QCD. Since it can be written as a 4-divergence, such term does not add any Feynman rules. Its peculiarity is that it preserves charge (C) symmetry, while it violates parity (P) and time-reversal (T) symmetries. Hence, it violates CP symmetry.

Higher-dimensional operators that violate CP-symmetry have been discussed in the context of SM Effective Field Theories (SMEFT), which are based on the assumption that sizeable low-energy effects can arise from the presence of new physics at significantly larger energy scales (see, *e.g.*, Refs. [291–293]). It has been shown that the first relevant CP-odd operators arise at dimension six, and include lepton and quark electric dipole moments, chromo-electric dipole moments, four-quark operators (see Ref. [294] and references therein).

One of the possible measurable effects of CP violation is the generation of permanent electric dipole moments in various particles (see, e.g., [295] and references therein). However, no experiment has been able to measure any effects of strong CP violation in the last 30 years. In particular, current experimental measurements set an upper bound on the neutron electric dipole moment (nEDM) to be $d_n < 10^{-26}$ e cm, indicating that the coupling parameter of a new BSM CP term in the QCD Lagrangian must be small.

In general, the strong CP problem can be traced back to a broader topic, *i.e.* the justification that strong interactions are invariant under parity transformations. Up to now, no theoretical principle has been found that prevents parity-violating terms in the QCD Lagrangian. Moreover, the experimental evidence that CP violation occurs in the EW sector of the Standard Model could anyway make us wonder why such a violation does not happen in QCD. In fact, in the last thirty years many facilities have measured consistent CP violations in the EW sector. For instance, in March 2019 the LHCb Collaboration announced the discovery of (weak) CP violation in charmed D^0 decays with a level of significance of 5.3 standard deviations [296].

In this chapter, we discuss how to probe the presence of strong P violation through a detailed study of the internal structure of nucleons. We base our discussion on Ref. [3].

If we do not impose strong P invariance, new terms appear in the hadronic tensor for the DIS process, generated by new strong parity-violating (PV) parton densities. These new terms are odd under parity transformation (P-odd), and they can be C-even (CP-odd) or C-odd (CP-even). Similar terms have been discussed in Ref. [117], a pioneering work on the P-odd quark fragmentation in the θ-vacuum, which can give rise to non-trivial dihadron correlation. Also, theory-based models of PDFs and FFs have been recently discussed in Refs. [118, 119].

In this thesis, we discuss the case for collinear PDFs by identifying the new contributions that P-odd PDFs would generate in structure functions. Moreover, we propose for the first time a phenomenological estimate of PV effects through a phenomenolocal fit of the available experimental data.

Our focus is mainly on the DIS process of a longitudinally polarized lepton beam off an unpolarized proton or nuclear target, and we briefly discuss the case of DIS of an unpolarized lepton beam off a polarized target.

In the case of unpolarized target, we show how a genuine new PV contribution in the structure function F_3 appears in the pure photon-exchange channel

when considering longitudinally polarized lepton probes. Being absent in the SM, this contribution would be non-zero only with the presence of a new strong PV PDF.

We estimate the size of such partonic distribution through a fit at the best possible theoretical accuracy of the available experimental data from HERA, SLAC, and Jefferson Lab (JLab) that can be sensitive to this kind of PV effects.

Our phenomenological study, the first one on this topic, could lead to new analyses of DIS measurements with different lepton beam polarizations and charges. New measurements from the JLab 12 GeV program and the EIC machine could shed further light on the presence of experimental effects that could derive from a strong P violation.

This chapter is organized as follows. In Sec. 6.1, we describe the structure functions and the observable which can be affected by the introduction of PV PDFs. In Sec. 6.2, we discuss how structure functions get modified by our ansatz for the new PV parton densities. In Sec. 6.3, we describe the data set and the theoretical choices included in the analysis in Ref. [3]. Finally, in Sec. 6.4 we show the estimate of the size of PV effects in the proton, which is the main result of our analysis.

6.1 Formalism of PV parton densities

In DIS with Neutral Current (NC), the leptonic tensor of a fast-moving lepton with initial (final) 4-momentum k ($k' = k - q$) and helicity λ has three possible channels (γ exchange, $\gamma - Z$ interference and Z exchange) and can be written as

$$L_{\mu\nu}^{(j)} = C^{(j)} L_{\mu\nu}^{(\gamma)} ,$$
$$L_{\mu\nu}^{(\gamma)} = 2 \left[k_\mu k_\nu' + k_\mu' k_\nu - (k \cdot k') g_{\mu\nu} - i\lambda\varepsilon_{\mu\nu\alpha\beta} k^\alpha k'^\beta \right] , \qquad (6.2)$$

where the index j runs over the channels $j = \gamma, \gamma Z, Z$, and the coefficients $C^{(\gamma)} = 1$, $C^{(\gamma Z)} = -(g_V^e - \lambda g_A^e)$, $C^{(Z)} = (g_V^e - \lambda g_A^e)^2$ [297]. The $g_{V,A}^e$ are the electron's neutral weak couplings to the Z. For an anti-lepton, the same formula holds but with the sign of g_A^e flipped.

The hadronic tensor for an unpolarized hadron with momentum P^μ is given by [298]

$$W^{\mu\nu} = \left(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2} \right) F_1 + \frac{\tilde{P}^\mu \tilde{P}^\nu}{P \cdot q} F_2 + i \frac{\varepsilon^{\mu\nu\rho\sigma}}{2(P \cdot q)} P_\rho q_\sigma F_3 , \qquad (6.3)$$

where we introduced the four-vector $\tilde{P}^\mu = P^\mu - q^\mu (P \cdot q)/q^2$.

The cross section can be calculated by contracting the leptonic and hadronic

tensors

$$\frac{d^2\sigma}{dx_B\,dy} = \frac{2\pi y\alpha^2}{Q^4} \sum_{j=\gamma,\,\gamma Z,\,Z} \eta^{(j)}\,L_{\mu\nu}^{(j)}\,W^{\mu\nu} = \frac{2\pi y\alpha^2}{Q^4} \sum_{j=\gamma,\,\gamma Z,\,Z} \eta^{(j)}\,C^{(j)}\,L_{\mu\nu}^{(\gamma)}\,W^{\mu\nu}\,,$$

(6.4)

where α is the fine structure constant, y is the inelasticity of the process, x_B is the Bjorken variable, $Q^2 = -q^2$ is the hard scale of the DIS process, and the prefactors $\eta^{(j)}$ are

$$\eta_\gamma = 1; \qquad \eta_{\gamma Z} = \left(\frac{G_F M_Z^2}{2\sqrt{2}\pi\alpha}\right)\left(\frac{Q^2}{Q^2 + M_Z^2}\right); \qquad \eta_Z = \left(\eta^{\gamma Z}\right)^2\,, \qquad (6.5)$$

with M_Z the mass of the Z boson, and G_F the Fermi coupling constant.

If we neglect the lepton mass, the contraction of the leptonic tensor $L_{\mu\nu}^{(\gamma)}$ of Eq. (6.2) with the hadronic tensor $W^{\mu\nu}$ of Eq. (6.3) gives

$$L_{\mu\nu}^{(\gamma)}\,W^{\mu\nu} = 2\left[Q^2\,F_1^{(\gamma)} + \frac{Q^2}{x_B y^2}\left(1 - y - \frac{1}{4}R^2 y^2\right)F_2^{(\gamma)} - \lambda\frac{Q^2}{2}\frac{2-y}{y}F_3^{(\gamma)}\right]\,,$$

(6.6)

where $R = 2Mx_B/Q$ determines the size of target mass corrections (TMC). By taking into account such TMC, we introduce the relation $2x_B F_1 = (1 + R^2)F_2 - F_L$ into the above contraction and we get

$$L_{\mu\nu}^{(\gamma)}\,W^{\mu\nu} = \frac{Q^2}{x_B y^2}\left[\left(Y_+ + \frac{1}{2}R^2 y^2\right)F_2^{(\gamma)} - y^2\,F_L^{(\gamma)} - \lambda\,Y_-\,x_B F_3^{(\gamma)}\right]\,, \qquad (6.7)$$

with $Y_\pm = 1 \pm (1 - y)^2$.

If we insert the above result into the expression of the cross section in Eq. (6.4), we finally get

$$\frac{d^2\sigma}{dx_B dy} = \frac{2\pi\alpha^2}{x_B y Q^2}\left[\left(Y_+ + \frac{R^2 y^2}{2}\right)\left(F_{2,UU} + \lambda F_{2,LU}\right) - y^2\left(F_{L,UU} + \lambda F_{L,LU}\right)\right.$$

$$\left. - Y_-\left(x_B F_{3,UU} + \lambda x_B F_{3,LU}\right)\right]. \qquad (6.8)$$

This result is equivalent to the standard expression for the NC DIS cross section listed in the PDG review (see also Ref. [297]). For convenience, we explicitly distinguished the terms independent of λ, that involve unpolarized lepton (U) and hadron (U), from the one proportional to λ, that requires a polarized lepton beam (L).

The structure functions involved in the cross section of Eq. (6.8) can be

expressed as [298]

$$F_{2,UU}(x_B, Q^2) = F_2^{(\gamma)} - g_V^e \eta_{\gamma Z} F_2^{(\gamma Z)} + \left(g_V^{e\,2} + g_A^{e\,2}\right)\eta_Z F_2^{(Z)}, \qquad (6.9)$$

$$F_{2,LU}(x_B, Q^2) = g_A^e \eta_{\gamma Z} F_2^{(\gamma Z)} - 2g_V^e g_A^e \eta_Z F_2^{(Z)}, \qquad (6.10)$$

$$F_{L,UU}(x_B, Q^2) = F_L^{(\gamma)} - g_V^e \eta_{\gamma Z} F_L^{(\gamma Z)} + \left(g_V^{e\,2} + g_A^{e\,2}\right)\eta_Z F_L^{(Z)}, \qquad (6.11)$$

$$F_{L,LU}(x_B, Q^2) = g_A^e \eta_{\gamma Z} F_L^{(\gamma Z)} - 2g_V^e g_A^e \eta_Z F_L^{(Z)}, \qquad (6.12)$$

$$F_{3,UU}(x_B, Q^2) = g_A^e \eta_{\gamma Z} F_3^{(\gamma Z)} - 2g_V^e g_A^e \eta_Z F_3^{(Z)}, \qquad (6.13)$$

$$F_{3,LU}(x_B, Q^2) = F_3^{(\gamma)} - g_V^e \eta_{\gamma Z} F_3^{(\gamma Z)} + \left(g_V^{e\,2} + g_A^{e\,2}\right)\eta_Z F_3^{(Z)}. \qquad (6.14)$$

We point out that the sign of terms containing g_A^e in Eqs. (6.10), (6.12),(6.13) must be changed for positron scattering.

As already mentioned, the results presented here are in agreement with standard literature. However, in the SM the contribution to the F_3 structure function that arises from pure γ exchange is zero. It emerges only when introducing the possibility of strong P violation, and it is the new ingredient we estimate in our analysis.

We identify the following parity-violating asymmetry as the observable that is sensitive to this new term:

$$A_{\mathrm{PV}} \equiv \frac{d\sigma(\lambda = 1) - d\sigma(\lambda = -1)}{d\sigma(\lambda = 1) + d\sigma(\lambda = -1)}$$

$$= \frac{\left(Y_+ + \frac{R^2 y^2}{2}\right) F_{2,LU} - y^2 F_{L,LU} - Y_- x_B F_{3,LU}}{\left(Y_+ + \frac{R^2 y^2}{2}\right) F_{2,UU} - y^2 F_{L,UU} - Y_- x_B F_{3,UU}}. \qquad (6.15)$$

In the region $M^2 \ll Q^2$ where collinear factorization holds, the structure functions above can be written in terms of PDFs, arising from the decomposition of the quark correlator. In particular, if we include strong PV terms in the correlation function for unpolarized nucleons, we obtain the following expression for the quark-quark correlator at leading twist,

$$\Phi^q(x, Q^2) = \left\{ f_1^q(x, Q^2) + g_1^{\mathrm{PV}q}(x, Q^2)\gamma_5 \right\} \frac{\slashed{n}_+}{2}, \qquad (6.16)$$

where x is the light-cone momentum fraction of the parent hadron carried by the quark q, and n_+ is a light-like vector pointing to the dominant light-cone "+" direction (neglecting target mass corrections, $x \approx x_B$). The second term is usually ignored in the standard literature of collinear PDFs because it is not allowed by the SM. Such term incorporates the parton distribution g_1^{PV}, which describes the difference in the probability to find right-handed vs. left-handed quarks inside an unpolarized proton. It is a P-odd function, in the sense that it changes its sign under parity transformation; it is also even under charge transformations, hence it is CP-even. Since it originates from the Dirac

structure $\gamma_5 \not{n}_+ = \gamma_5 \gamma^-$, its behavior under QCD evolution is equivalent to the helicity PDF $g_1(x, Q^2)$. Moreover, a connection can be established between the integral of this function and the so-called anapole moment of the proton or the nucleus (see, e.g., Refs. [299–302]), which could be responsible for PV effects measured in electron-proton elastic scattering experiments [303, 304].

We remark that a decomposition similar to the one in Eq. (6.16) can be found also for the transverse-momentum-dependent correlation function, as already discussed in Ref. [119]. Our new function g_1^{PV} corresponds to the integral of the function u_1 in that reference.

If we neglect strong P violation, we can write the structure function F_3 in terms of the unpolarized PDF f_1 [298]. In presence of target mass corrections (see, *e.g.*, Ref. [305] for a recent review), the PDF f_1 must be evaluated at the Nachtmann variable [305]

$$x_N = \frac{2x_B}{1 + \sqrt{1 + R^2}}. \tag{6.17}$$

For convenience, in the following we avoid explicitly writing the arguments of the PDFs. The detailed expression of the structure functions is

$$F_3^{(\gamma)}(x_B, Q^2) = 0, \tag{6.18}$$

$$F_3^{(\gamma Z)}(x_B, Q^2) = \frac{1}{\sqrt{1 + R^2}} \sum_q 2e_q g_A^q f_1^{(q-\bar{q})}, \tag{6.19}$$

$$F_3^{(Z)}(x_B, Q^2) = \frac{1}{\sqrt{1 + R^2}} \sum_q 2g_V^q g_A^q f_1^{(q-\bar{q})}, \tag{6.20}$$

where $f_1^{q-\bar{q}} = f_1^q - f_1^{\bar{q}}$, e_q is the quark charge, $g_{V,A}^q$ are quark's neutral weak couplings to the Z, and R has been defined below Eq. (6.8). The inclusion of strong P violation leads to a modification of the standard results given by the following additional contributions:

$$\Delta F_3^{(\gamma)}(x_B, Q^2) = -\frac{1}{\sqrt{1 + R^2}} \sum_q e_q^2 g_1^{\text{PV}(q+\bar{q})}, \tag{6.21}$$

$$\Delta F_3^{(\gamma Z)}(x_B, Q^2) = -\frac{1}{\sqrt{1 + R^2}} \sum_q 2e_q g_V^q g_1^{\text{PV}(q+\bar{q})}, \tag{6.22}$$

$$\Delta F_3^{(Z)}(x_B, Q^2) = -\frac{1}{\sqrt{1 + R^2}} \sum_q \left(g_V^{q2} + g_A^{q2}\right) g_1^{\text{PV}(q+\bar{q})}. \tag{6.23}$$

We note that also the structure function F_2 gets modified by the following

terms:

$$\Delta F_2^{(\gamma)}(x_B, Q^2) = 0, \tag{6.24}$$

$$\Delta F_2^{(\gamma Z)}(x_B, Q^2) = -\sum_q 2e_q g_A^q x_B g_1^{\mathrm{PV}(q-\bar{q})}, \tag{6.25}$$

$$\Delta F_2^{(Z)}(x_B, Q^2) = -\sum_q 2g_V^q g_A^q x_B g_1^{\mathrm{PV}(q-\bar{q})}. \tag{6.26}$$

The same new contributions have to be introduced also in the structure function F_L.

For sake of completeness, we briefly discuss the case of a polarized nucleon. The correlator becomes

$$\Phi^q(x, Q^2) = \left\{ f_1^q(x, Q^2) + g_1^{\mathrm{PV}q}(x, Q^2)\gamma_5 + S_L\left(g_1^q(x, Q^2)\gamma_5 + f_{1L}^{\mathrm{PV}q}(x, Q^2)\right) \right.$$
$$\left. - \not{S}_T\left(h_1^q(x, Q^2)\gamma_5 - e_{1T}^{\mathrm{PV}q}(x, Q^2)\right) \right\} \frac{\not{n}_+}{2}, \tag{6.27}$$

where $S_{L,T}$ are the longitudinal/transverse polarization of the hadron, respectively. Two new polarized PDFs f_{1L}^{PV} and e_1^{PV} arise by breaking parity invariance. We observe that the PDF f_{1L}^{PV} is P-odd and CP-odd, and should be connected to the electric dipole moment of the proton. Further investigations on this parton density are left for future studies that go beyond this thesis. We observe that our functions f_{1L}^{PV} and e_1^{PV} correspond to the integral of the functions v_{1L} and w_{1T} in Ref. [119], respectively.

6.2 Modeling the unpolarized PV parton density

In this section, we describe our model of the unpolarized PV parton density $g_1^{\mathrm{PV}}(x, Q^2)$ that is included in our phenomenological analysis to estimate the size of PV effects in the nucleons structure.

Currently, there are no models that can generate strong P-violating partonic distributions. As already discussed in Ch. 1, the introduction of the θ-term $\frac{\theta g^2}{32\pi^2}\tilde{G}_{\mu\nu}G^{\mu\nu}$ in the QCD lagrangian[1] would generate too small effects due to the very strong experimental constraints provided by the measurements of the nEDM. Effects of the order of $\theta \sim 10^{-10}$ cannot be identified in the internal structure of nucleons with the present experimental data. Therefore, one should find alternative terms to be added to the QCD lagrangian. Such terms must be Lorentz-invariant, P-odd, CP-odd, and must not break renormalization. If we look at the literature of SM Effective Field Theories (SMEFT), we may find some higher-dimensional P-violating operators which could generate

---

[1]Here, g is the strong interaction coupling constant, $G^{\mu\nu}$ is the full field strength tensor of the gauge field, and $\tilde{G}_{\mu\nu} = \frac{1}{2}\epsilon_{\alpha\beta\mu\nu}G^{\alpha\beta}$.

non-vanishing PV parton densities. Moreover, a model of quark fragmentation in a topologically non-trivial QCD vacuum has been shown to lead to the emergence of P-odd transverse-momentum-dependent fragmentation functions [117]. Finally, PV parton densities can be produced by the inclusion of electroweak corrections to the evolution of PDFs. In particular, contributions from the perturbative radiation of a Z-boson from a quark could be responsible for a P-odd distribution function. Such kind of EW perturbative corrections have never been studied in QCD.[2] However, this effect is expected to be negligible. In fact, the corrections introduced by QED contributions to the DGLAP kernels are of the order 1% [307, 308], and PV contributions arising from EW corrections are expected to be much smaller (by a factor Q^2/M_Z^2).

In order to obtain a first estimate of the new PV PDF g_1^{PV}, we make the simplest possible hypotesis: we assume it is proportional to its parity-even counterpart, *i.e.*, $g_1^{\mathrm{PV}} = a\, g_1$, where a is a fitting parameter. Therefore, the additional strong PV term to be added to the $F_3^{(\gamma)}$ structure function is the following:

$$\Delta F_3^{(\gamma)}(x_B, Q^2) = -\frac{a}{\sqrt{1+R^2}} \sum_q e_q^2 g_1^{(q+\bar{q})}. \tag{6.28}$$

The total additional contributions to the standard expression of the two terms $F_{3,UU}$ and $F_{3,LU}$ are

$$
\begin{aligned}
\Delta F_{3,UU}(x_B, Q^2) = &-\frac{a}{\sqrt{1+R^2}}\left(g_A^e \eta_{\gamma Z} \sum_q 2e_q g_V^q g_1^{(q+\bar{q})} \right. \\
&\left. - 2g_V^e g_A^e \eta_Z \sum_q \left(g_V^{q\,2} + g_A^{q\,2}\right) g_1^{(q+\bar{q})} \right),
\end{aligned}
\tag{6.29}
$$

$$
\begin{aligned}
\Delta F_{3,LU}(x_B, Q^2) = &-\frac{a}{\sqrt{1+R^2}}\left(\sum_q e_q^2 g_1^{(q+\bar{q})} - g_V^e \eta_{\gamma Z} \sum_q 2e_q g_V^q g_1^{(q+\bar{q})} \right. \\
&\left. + \left(g_V^{e\,2} + g_A^{e\,2}\right)\eta_Z \sum_q \left(g_V^{q\,2} + g_A^{q\,2}\right) g_1^{(q+\bar{q})} \right).
\end{aligned}
\tag{6.30}
$$

Moreover, the structure function F_2 becomes

---

[2]These EW corrections due to Z-boson emission should be similar to the corresponding QED corrections due to the emission of a photon [306].

$$\Delta F_{2,UU}(x_B, Q^2) = g_V^e \eta_{\gamma Z}\, a \sum_q 2 e_q g_A^q x_B g_1^{(q-\bar{q})}$$
$$- \left(g_V^{e\,2} + g_A^{e\,2}\right)\eta_Z\, a \sum_q 2 g_V^q g_A^q x_B g_1^{(q-\bar{q})}, \tag{6.31}$$

$$\Delta F_{2,LU}(x_B, Q^2) = -g_A^e \eta_{\gamma Z}\, a \sum_q 2 e_q g_A^q x_B g_1^{(q-\bar{q})} \tag{6.32}$$
$$+ 2 g_V^e g_A^e \eta_Z\, a \sum_q 2 g_V^q g_A^q x_B g_1^{(q-\bar{q})}.$$

To evaluate the size of this effect, we adopt the following approach: we attribute any discrepancy between the SM predictions and the available data entirely to the presence of the new PV PDF, and we estimate the size of the $g_1^{\mathrm{PV}}(x, Q^2)$ distribution needed to be compatible with the available experimental data.

6.3 Experimental data sets

In this section, we illustrate the experimental data set included in the fit of the $g_1^{\mathrm{PV}}(x, Q^2)$ distribution, and we discuss the choices for the inputs to this analysis.

In the analysis described in this chapter, we fit theoretical predictions at NLO accuracy for the electron and positron PV asymmetries in Eq. (6.15) to the experimental data of DIS on an unpolarized proton beams from the HERA Collaboration [309], and of DIS on deuterium target from the PVDIS Collaboration at JLab 6 GeV [310, 311]) and from the SLAC E122 experiment [312]. We list in Tab. 6.1 the number of data points (N_{dat}) along with the kinematics of each experimental set.

Experiment	N_{dat}	Observable	Hadron	$\sqrt{s}$ [GeV]	Q^2 [GeV^2]	y	Ref.
HERA	136	A_{PV} for e^+	proton	319	120 - 30000	0.033 - 0.9	[309]
HERA	138	A_{PV} for e^-	proton	319	120 - 30000	0.033 - 0.9	[309]
PVDIS	2	A_{PV} for e^-	deuterium	4.77	1.085; 1.901	0.20; 0.28	[310]
E122	11	A_{PV} for e^-	deuterium	5.5 - 6.5	0.92 - 1.96	0.15 - 0.36	[312]
Total	287						

Table 6.1: Breakdown of the data sets considered in the analysis presented in this chapter. For each data set, the table includes information on: the number of data points (N_{dat}), the measured observable, the initial-state hadron, the center-of-mass energy $\sqrt{s}$, the covered range(s) in Q^2, the inelasticity y, and the published reference.

We stress that the deuterium is described in first approximation as a incoherent sum of free nucleons, *i.e.* one proton and one neutron. The total number of experimental data included in our fit amounts to 287, with 136 data for positron asymmetry and 151 for electron asymmetry.

We observe that the PVDIS and E122 data sets are characterized by very small values of the energy scale $Q^2 \simeq 1-2$ GeV^2. Therefore, the inclusion of target mass corrections in Eq. (6.8) and Eqs. (6.19)-(6.23) has a sizeable effect for these two data sets, while it is negligible for the HERA measurements at much larger values of Q^2.

Moreover, according to Refs. [310, 311] we introduce the electroweak radiative corrections discussed in Ref. [313] into the C_1 and C_2 coefficients, which involve the lepton and quark axial and vector couplings g_A^q and g_V^q, respectively. In fact, the standard expressions for the u and d components of the above mentioned C coefficients are

$$
\begin{aligned}
C_{1u} &= 2g_A^e g_V^u = 2\left(-\frac{1}{2}\right)\left(\frac{1}{2} - \frac{4}{3}\sin^2\theta_W\right) \\
C_{1d} &= 2g_A^e g_V^d = 2\left(-\frac{1}{2}\right)\left(-\frac{1}{2} + \frac{2}{3}\sin^2\theta_W\right) \\
C_{2u} &= 2g_V^e g_A^u = 2\left(-\frac{1}{2} + 2\sin^2\theta_W\right)\left(\frac{1}{2}\right) \\
C_{2d} &= 2g_V^e g_A^d = 2\left(-\frac{1}{2} + 2\sin^2\theta_W\right)\left(-\frac{1}{2}\right)
\end{aligned}
\tag{6.33}
$$

where θ_W is the weak mixing angle. The same expressions hold for the other quark families. The electroweak radiative corrections in the $\overline{\text{MS}}$ scheme are parametrized in Ref. [313] as

$$
\begin{aligned}
C_{1u}' &= -0.1887 - 0.0011 \times \frac{2}{3}\ln\left(\frac{Q^2}{0.14 \text{ GeV}^2}\right) \\
C_{1d}' &= 0.3419 - 0.0011 \times \frac{-1}{3}\ln\left(\frac{Q^2}{0.14 \text{ GeV}^2}\right) \\
C_{2u}' &= -0.0351 - 0.0009 \times \ln\left(\frac{Q^2}{0.078 \text{ GeV}^2}\right) \\
C_{2d}' &= 0.0248 + 0.0007 \times \ln\left(\frac{Q^2}{0.021 \text{ GeV}^2}\right).
\end{aligned}
\tag{6.34}
$$

In principle, this corrections, which include different effects discussed in Ref. [313], could be considered also at higher Q^2, but they are very small compared to the experimental errors.[3]

In our fit, we need to include a set of collinear unpolarized PDFs and helicity PDFs: we choose the NNPDF4.0 [30] and the NNPDFpol1.1 [315]

---

[3]They are provided by, *e.g.*, the Djangoh event generator [314].

sets, respectively. In order to take into account the error on the PDFs sets, we consider the full ensemble of MC replicas, and we include their uncertainty as a source of theoretical (systematic) error. However, the impact of this kind of uncertainty is very small because it is of order 1% of the experimental errors (or less). Since we perform the calculation of the structure functions F_2, F_L and F_3 at $\mathcal{O}(\alpha_s)$, we consider PDF sets as extracted at NLO.

In order to perform a global extraction of g_1^{PV}, it could be interesting to identify an experimental observable which is sensitive to this new PV distribution function also for DY process. Currently, such observable has never been studied.

6.4 Estimate of the size of PV effects inside nucleons

In this section, we report the results of the fit we performed to estimate the size of the new PV PDF g_1^{PV} in the nucleon structure.

We perform the error propagation through the replica method (see Sec. 3.2). In short, we generate 100 Monte Carlo replicas of the experimental data and we associate each one of them to a single replica of unpolarized and polarized PDFs. The final result is a bootstrap distribution of 100 values for the parameter a (see Eq. (6.28)).

In Tab. 6.2, we report the resulting quality of the fit. We indicate the χ^2 value per number of data points N_{dat} for each of the considered experimental data sets, along with one standard deviation from the full ensemble of replicas. We provide also the values of χ^2/N_{dat} that we obtain with the SM predictions. From the last row of Tab. 6.2, we observe that the mean value of the global

		SM predictions	Our analysis
Data set	N_{dat}	χ^2/N_{dat}	χ^2/N_{dat}
HERA e^+ (p)	136	1.12 ± 0.01	1.12 ± 0.01
HERA e^- (p)	138	0.98 ± 0.01	0.98 ± 0.01
PVDIS e^- (d)	2	0.67 ± 0.12	0.42 ± 0.40
E122 e^- (d)	11	0.97 ± 0.01	0.94 ± 0.02
Total	**287**	**1.042 ± 0.001**	**1.037 ± 0.004**

Table 6.2: Breakdown of the values of χ^2 per number of data points N_{dat} for all data sets considered in our analysis (hadron targets in brackets). The values of χ^2 and uncertainties refer to the mean value and one standard deviation from the ensemble of replicas of the experimental data.

χ^2/N_{dat} is slightly smaller than the SM result. This means that the description of the global data set is improved by our model of PV PDFs. However, we note that the values of χ^2/N_{dat} are all close to or smaller than 1 because of the large experimental errors affecting the majority of the data points. New data

with better precision are required to have more information on the actual size of PV contributions.

In Fig. 6.1, we show the χ^2 per number of data N_{dat} of the 100 MC replicas in the analysis illustrated in this chapter in SM (blue histogram) and after including PV effects (yellow histogram). It is interesting to see that the χ^2 distribution of the 100 replicas is shifted at smaller values after the inclusion of PV effects. This reflects what can be seen in Tab. 6.2, namely that a non-zero PV effect, even if small, is statistically favoured by current experimental data.
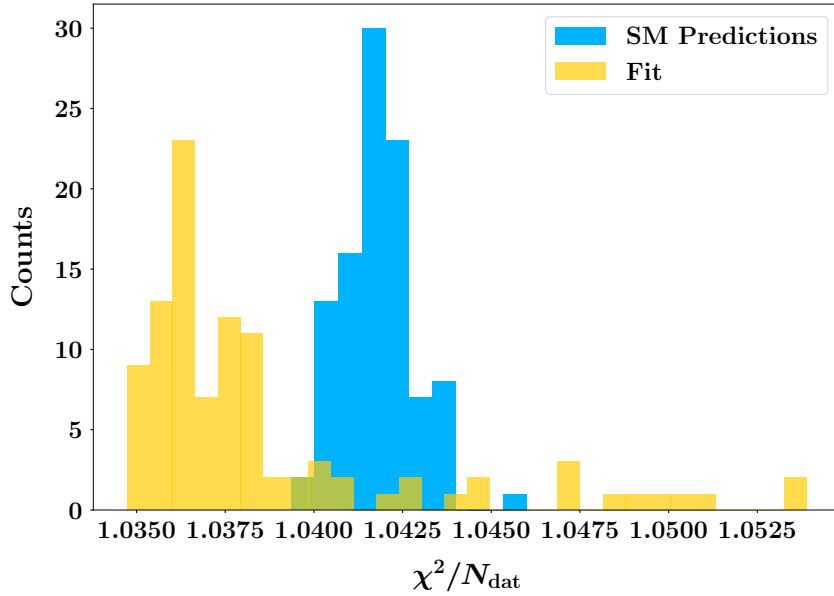

Figure 6.1: Statistical distribution of the χ^2 values of the 100 Monte Carlo replicas of our analysis in SM (blue histogram) and after the inclusion of PV effects (yellow histogram).

The result of the fit for the parameter a is

$$a = \hat{a} \pm \Delta a = (-1.01 \pm 0.66) \cdot 10^{-4} \, , \tag{6.35}$$

where $\hat{a}$ and Δa are the mean value and one standard deviation from the bootstrap set of 100 values, respectively. Our fitting parameter is incompatible with zero at the 1.5 σ level. More precisely, our results indicate that the null hypothesis (no effect) can be rejected with a p-value $= 0.063$: the probability of making an error by rejecting the hypothesis of no parity-violating contributions is 6.3%. It is not negligible but small. This is a phenomenological indication that P violation is statistically favoured by current experimental data with respect to the SM result. Moreover, a negative value of the parameter a indicates that there would be more left-handed quarks than right-handed ones in the proton. In Fig. 6.2, we display the statistical distribution for the parameter a. We observe once more that the left-side tail of the distribution of our fitting
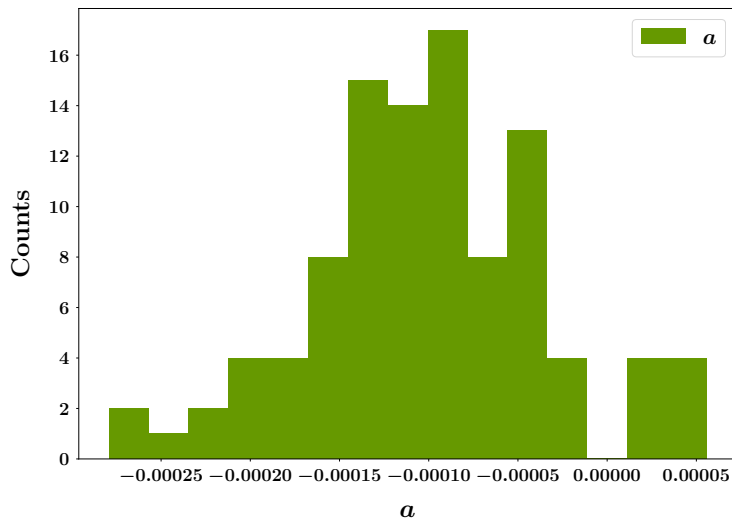
Figure 6.2: Statistical distribution of the 100 replicas of the PV parameter a fitted to experimental data.

parameter exhibits a marginal compatibility with zero. We checked that this result is independent of the chosen set of unpolarized PDFs by performing the fit with different sets. We find no significant differences when using other sets, and each result for a single set is compatible with all the others.

In Fig. 6.3, we display the PV PDF g_1^{PV} for a u quark in the proton at $Q = 2$ GeV as a function of the longitudinal momentum fraction x. In the upper plot, each of the 100 replicas is represented as a colored line, while the light blue band correspond to the 68% C.L.

It is evident that the global distribution of replicas is marginally compatible with a null parton density. On the contrary, a negative g_1^{PV} for an u quark in the proton is clearly favoured. In the lower plot of Fig. 6.3, we show the relative uncertainty of the 68% C.L. band with respect to the mean value of the PV PDF: it is quite large due to the large experimental errors affecting the majority of the data sets included in our analysis.

The upper panel of Fig. 6.4 shows the comparison between theoretical predictions for A_{PV} of Eq. (6.15) (colored bands) and HERA data for inclusive DIS off a proton target of electrons e^- (solid black points) and positrons e^+ (open red points), as a function of Q^2 at the given $x = 0.13$. The central (lower) panel shows the relative difference of data and results of our fit with respect to SM predictions for the e^+ (e^-) asymmetry. The colored uncertainty bands correspond to the 68% C.L. We point out that the small theoretical uncertainty of the PDFs used in this work generates a very narrow with of the displayed uncertainty bands of our theoretical predictions.

As can be seen from Fig. 6.4 and Tab. 6.2, the HERA data sets are nicely described in our framework although the χ^2 for the e^+ asymmetry is slightly worse than the e^- case because of the behaviour of the more precise data points at the lowest Q^2 bins (see central panel). This is the region where the
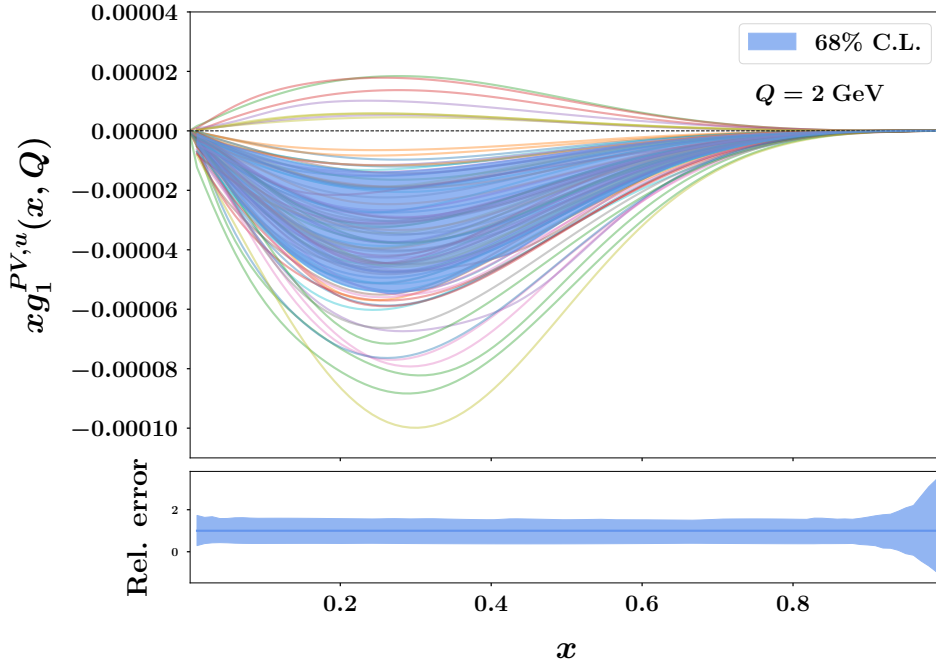
Figure 6.3: Upper plot: PV PDF g_1^{PV} for a u quark in the proton at $Q = 2$ GeV as a function of the longitudinal momentum fraction x. The colored lines represent each of the 100 fitted replicas, while the light blue band the 68% C.L. Lower plot: relative uncertainty for the 68% C.L. band.

contribution of the structure function F_3 in the γ-channel is dominant, which may leave room for improvements of our model in a future work.

It is interesting to consider also the behaviour of our predictions at a fixed energy scale Q as a function of x. In Fig. 6.5, we show the comparison between theoretical predictions for A_{PV} of Eq. (6.15) (colored histogram) and HERA data as a function of x at the given $Q^2 = 800$ GeV^2. The comparison for HERA data with positron beam are displayed in the left panel, while HERA data with electron beam in the right one. The lower panel shows the relative difference of data and results of our fit with respect to SM predictions.

We observe that in both left and right panels the experimental data show a behavior in x which is not reproduced by our theoretical predictions. This may indicate that a better description of the available data can be achieved by a PV PDF g_1^{PV} with a different x dependence. This could motivate a future calculation of a model of these PV effects starting from a new PV structure in the QCD lagrangian. However, we stress that our fit of the experimental HERA data has a good χ^2, also because of their large experimental errors. Moreover, we note that both the χ^2 for e^+ and e^- asymmetries are not modified by introducing PV effects (see Tab. 6.2), indicating that such contributions do not impact the description of HERA data: in fact, the fit is driven by the other experimental data sets with much smaller errors.

In Fig. 6.6, the upper panel shows the comparison between theoretical pre-
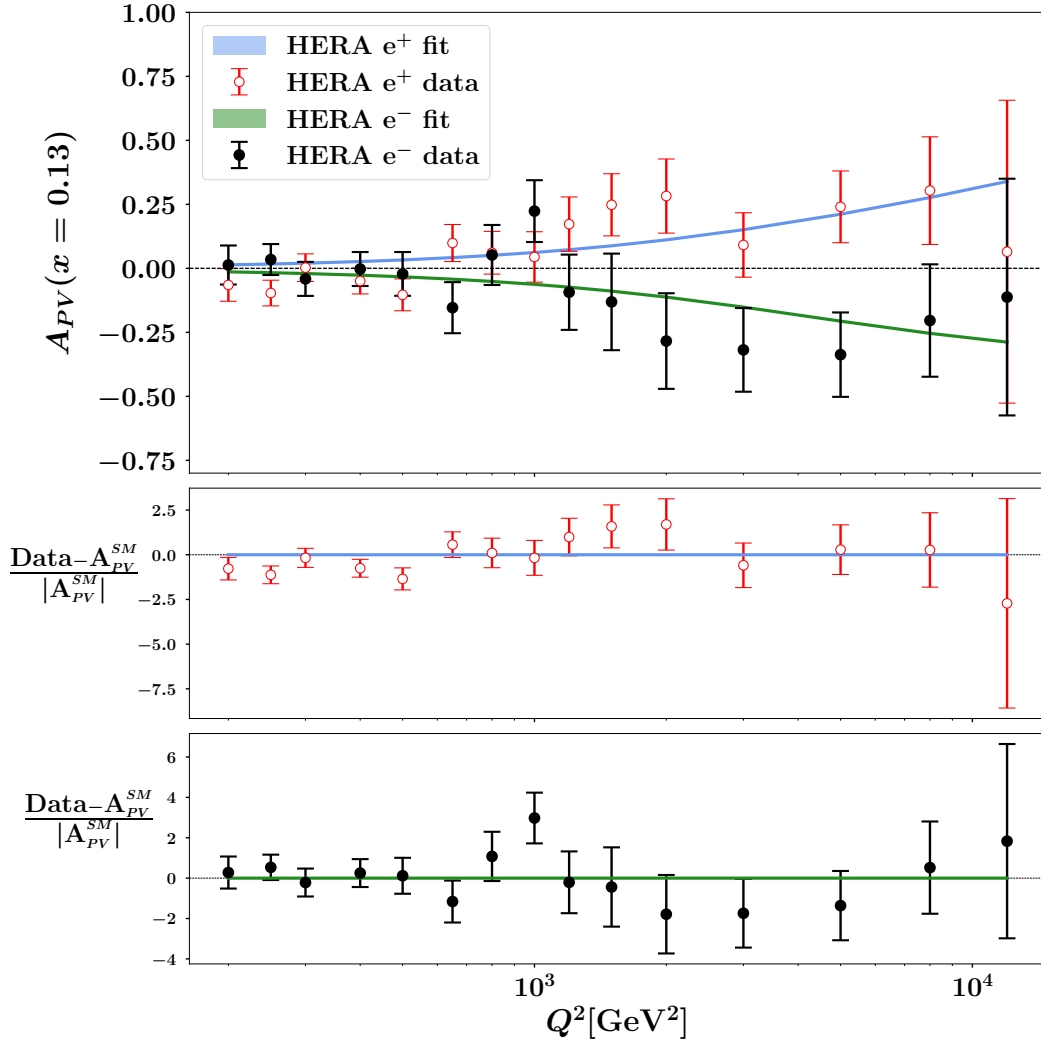
Figure 6.4: Comparison between theoretical predictions and HERA experimental data for e^+p (open red points) and e^-p (solid black points) collisions as a function of Q^2 at $x = 0.13$. Upper panel: PV asymmetry of Eq. (6.15); central panel: relative difference with respect to SM predictions for e^+ asymmetry; lower panel: same for e^- asymmetry. Uncertainty bands correspond to the 68% C.L.

dictions for $A_{\rm PV}$ from an electron beam of Eq. (6.15) and PVDIS data (blue rectangles versus solid points) and E122 data (green rectangles versus open points) as a function of Q^2. The central (lower) panel shows the relative difference of data and results of our fit with respect to SM predictions for the E122 (PVDIS) asymmetry. Similar to Fig. 6.4, the colored uncertainty bands correspond to the 68% C.L. We note that both the experimental data sets are nicely described in our framework, which is reflected in Tab. 6.2 by the systematic improvement of the quality of the fit with respect to the SM framework. Since it is affected by very small uncertainties, this result is particularly relevant
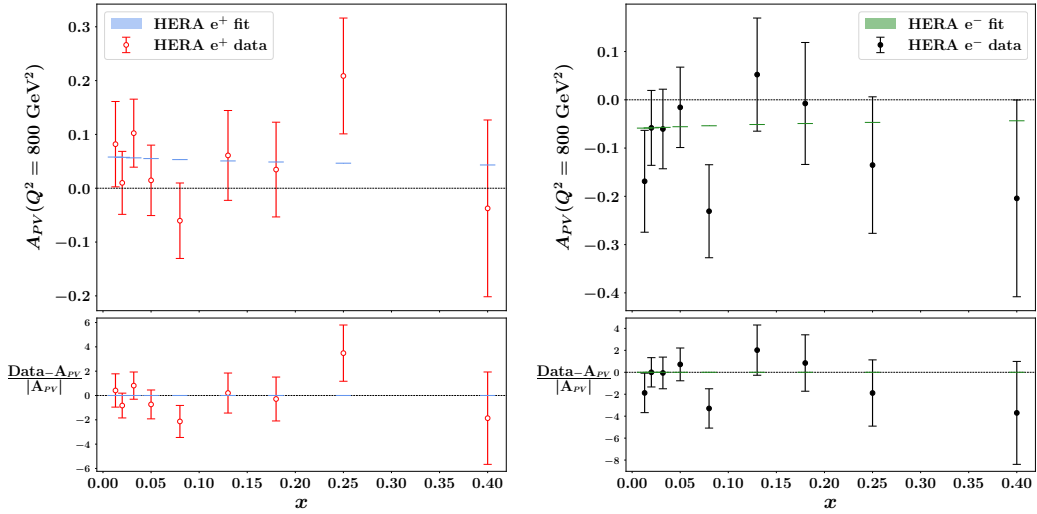
Figure 6.5: Comparison between theoretical predictions (colored histogram) and HERA experimental data for e^+p (left panel) and e^-p (right panel) collisions as a function of x at $Q^2 = 800$ GeV^2. Upper panel: PV asymmetry of Eq. (6.15); Lower panel: relative difference with respect to SM predictions for e^+ asymmetry. Uncertainty bands, unless barely visible, correspond to the 68% C.L.

for the PVDIS data, whose agreement with respect to the SM predictions is significantly improved.

In conclusion, this study shows that the introduction of strong PV contributions to DIS of longitudinally polarized leptons off an unpolarized target leads to a nonvanishing F_3 structure function in the case of pure photon exchange, which is not present in the SM result. The description of existing experimental data on DIS PV asymmetries gets improved by the addition of these new contributions. Their size is phenomenologically estimated to be small, but exhibiting a deviation from zero of about 1.5 σ. More precisely, the probability of making an error by rejecting the hypothesis of no parity-violating contributions is 6.3%, which is not negligible but small and underscores the need of more precise data and more detailed analyses.
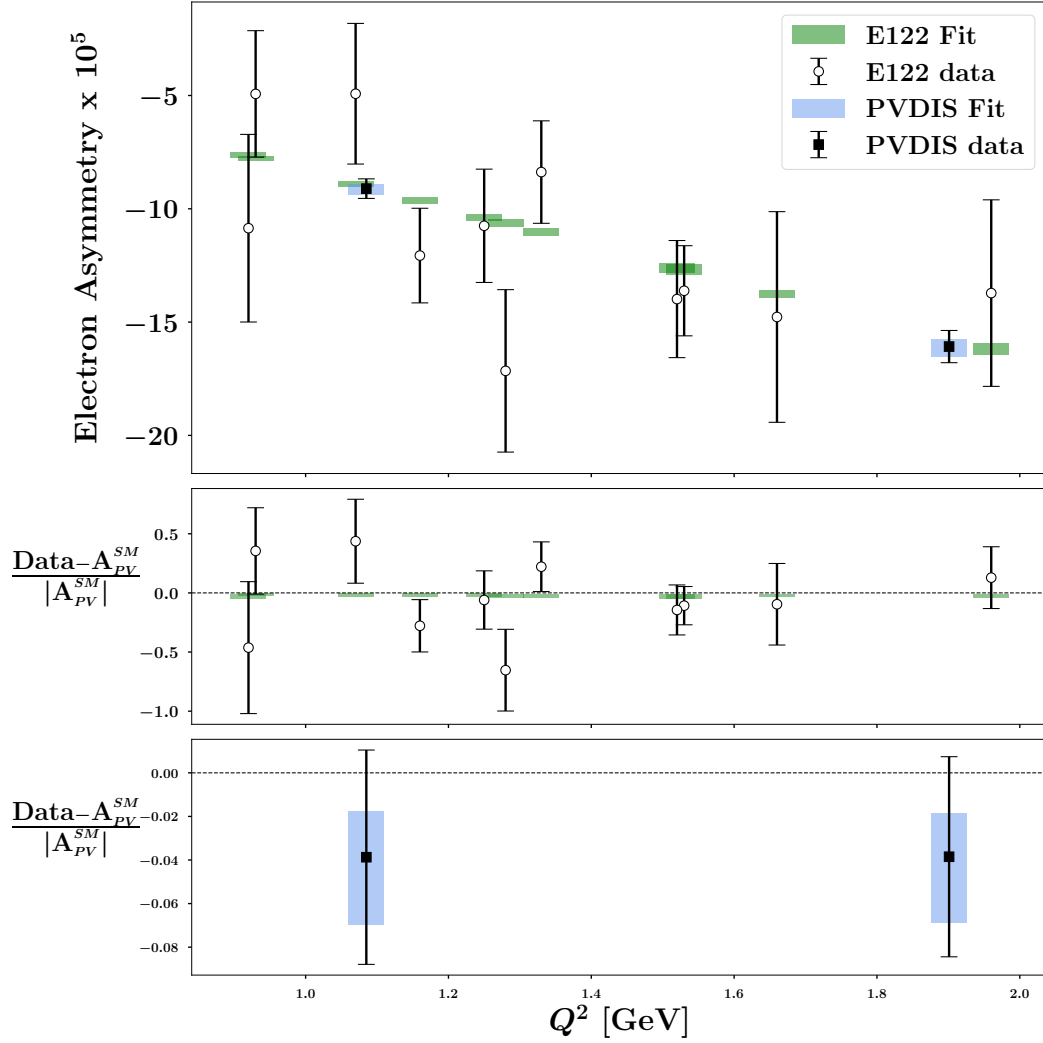
Figure 6.6: Comparison between theoretical predictions and data from PVDIS at JLab 6 GeV (blue recangles versus solid points) and SLAC E122 (green rectangles versus open points) experiments as a function of Q^2. Upper panel: electron e^- asymmetry $A_{\rm PV}$ of Eq. (6.15). Central panel: relative difference with SM predictions for the E122 asymmetry. Lower panel: same for the PVDIS asymmetry. Uncertainty bands correspond to the 68% C.L.

Chapter 7

Conclusions

The main goal of this thesis was to describe the currently most sophisticated exploration of the internal multidimensional structure of the proton, and to illustrate an example of how it is possible to search for Beyond-Standard-Model (BSM) effects on the inner structure of nucleons.

Most of this dissertation has been devoted to the description of the theoretical and phenomenological aspects at the basis of the simultaneous extraction of TMD parton densities (TMD PDFs) and TMD fragmentation functions (TMD FFs) for unpolarized quarks in the proton through fits of Drell–Yan (DY) and semi-inclusive Deep Inelastic Scattering (SIDIS) measurements.

As discussed in Ch. 2, the cross section for these process can be written in terms of TMDs through suitable factorization theorems. Such theorems have been demonstrated only for Drell–Yan, Semi-Inclusive DIS, and back-to-back hadron production in e^+e^- annihilation processes. In Ch. 2, we have focused on the description of the cross sections of different experimental observables for DY and SIDIS measurements that are made available by the various experimental collaborations, and we have examined the perturbative and nonperturbative ingredients of TMDs. In particular, we have illustrated the details of TMD evolution with the energy scale, which is given by a system of two differential equations. In fact, TMDs depend on two renormalization scales: the ultra-violet scale μ and the rapidity scale ζ. We have discussed suitable choices for initial and final scales, and we have examined the perturbative ingredients resulting from this particular choices. We have reviewed the treatment of the resummation of large logarithms appearing in the Sudakov form factor (the TMD evolution operator), and we defined a logarithmic ordering that allows to properly combine perturbative ingredients and, consequently, to identify a specific level of accuracy for the TMD extraction.

Then, we have described the need of a prescription to regularize the behaviour of TMDs at small quark transverse momenta (equivalently, at large values of the conjugated impact variable $|\boldsymbol{b}_T|$). The most common recipe is the so-called b_*-prescription introduced by Collins, Soper, and Sterman (CSS), which freezes the scales to avoid the Landau pole. After reviewing the different

$b_*(\boldsymbol{b}_T^2)$ functions used in the literature, we introduced nonperturbative model-dependent ingredients in the TMD expression: the nonperturbative part of the Collins–Soper kernel, driving the TMD evolution in the rapidity scale, and the intrinsic contribution to the full TMD.

Before discussing a full TMD extraction, we addressed the issue of the normalization of theoretical predictions for the SIDIS multiplicity at high logarithmic accuracy. We have discussed why the formalism is expected to work well at the lowest order and why problems are encountered at higher orders. A rigorous solution of this issue would need the inclusion of higher order contributions, such as power corrections, higher twists and, maybe, the so-called Y-term [73], which are currently not fully known. We identify as a operative solution the introduction of pre-computed normalization factors that are independent of the fit, and are identified by comparing the integral upon transverse momenta of the TMD formula to the corresponding collinear calculation, as explained in Sec. 2.4.

In Ch. 3, we have presented the state-of-the-art by the MAP Collaboration of simultaneous extractions of unpolarized TMD PDFs and TMD FFs from a global fit to presently available experimental data (indicated as MAP22 [1]). We included in the global data set 2031 data points collected by several experiments: 251 data points from DY measurements in collider-mode configuration at the Tevatron, LHC and RHIC, 233 points from fixed-target DY (see Tab. 3.1), and 1547 data points from SIDIS measurements by the HERMES and COMPASS collaborations (see Tab. 3.3).

Theoretical calculations are based on TMD factorization at a perturbative accuracy that is defined as N^3LL^-, namely considering TMD evolution at the next-to-next-to-next leading-log (N^3LL) level, hard factor and matching coefficients at second order in the strong coupling constant ($\mathcal{O}(\alpha_s^2)$), collinear PDFs at next-to-next leading order (NNLO), and collinear FFs at NLO (see Tab. 2.1).

To build the full TMDs, we have combined perturbative ingredients, regularized by means of the b_* prescription defined in Eq. (2.74), and phenomenological nonperturbative models of intrinsic transverse momentum, made by a sum of one or two Gaussians and a weighted Gaussian (see Eqs. (3.5) and (3.6)). In total, we had 21 free parameters to be fitted to 2031 experimental data: 11 for the TMD PDF, 9 for the TMD FF, and one for the Collins–Soper kernel. In this thesis, we have assumed these parameters to be the same for all quark flavors.

In Secs. 3.1.1-3.1.2, we have described the selection criteria for building the baseline data set of experimental measurements considered in the MAP22 analysis. For DY measurements, we have included all data points with final lepton pair transverse momentum $\boldsymbol{q}_T$ such that $|\boldsymbol{q}_T| < 0.2\,Q$, with Q the lepton pair invariant mass. For SIDIS data, we have adopted a more restrictive criterion than in Ref. [87], but less restrictive than in Ref. [88]: our cut includes many data points where the transverse momentum of the detected

hadron is much smaller than the hard scale, *i.e.* $|\boldsymbol{P}_{hT}| \ll Q$, but also with $0.2\,Q < |\boldsymbol{P}_{hT}|/z < Q$ (with z the fractional energy carried by the detected hadron). In total, we included about the 20% of the full SIDIS data set. [1]

The error propagation in the fit is performed with the replica method [87], leading to a final set of 250 TMD replicas. We reached a very nice overall agreement between data and theory, with a $\chi^2/N_{\mathrm{dat}} = 1.06$ for the central replica (defined in Sec. 3.2.2). We found a satisfactory description of all individual data sets, except for a small number of cases, the worst one being ATLAS (see Tab. 3.4 and Fig. 3.11). The obtained TMD PDFs are shown in Figs. 3.14, 3.15 and 3.16, and the TMD FFs in Figs. 3.17 and 3.18. We found that they deviate from a simple Gaussian (especially the FFs) and that the TMD shape changes in a nontrivial way as a function of x (TMD PDF) and z (TMD FF).

We also performed the unpolarized TMD extraction at lower NNLL and NLL accuracy, because they are needed in the theoretical expression of the single-spin asymmetries involving polarized TMDs, that cannot be currently studied at the top level of accuracy. We found that it is possible to simultaneously extract TMD PDFs and TMD FFs even at lower accuracy, but only if we exclude high-energy DY data set. We obtained a comparable or better description of the experimental data than the N^{3}LL result because the most precise high-energy data sets were not considered.

At the end of Ch. 3, we have discussed the effect of modifying the choice for the b_* prescription on TMD PDFs using a specific data set, and we have tested the quality of our fit by exploring other criteria for the $|\boldsymbol{q}_T|/Q$ cut. We demonstrated that our choice of the b_*-prescription among the ones discussed in Ch. 2 leads to the best description of the most precise experimental data sets inclded in the MAP22 fit. Furthermore, we found that we can obtain fits of quality comparable to the baseline result even with less conservative cuts than the nominal one. In particular, we can reach a good description including several points at $|\boldsymbol{q}_T| \gtrsim Q$, opening the door to further discussions on the definition of the range of applicability of TMD factorization for SIDIS observables.

The results of the MAP22 analysis have been used to estimate the impact of pseudodata from the JLab22 upgrade and the EIC machine in reducing unpolarized TMD error bands. To this aim, we have repeated the MAP22 global fit with the same setup, but including pseudodata covering the kinematic region of these future experiments. We obtain that the impact of JLab22 pseudodata is mainly in the large-x region. In particular, we have estimated a reduction of two orders of magnitude in the uncertainty at $x = 0.1$ (starting from the MAP22 global analysis). A similar reduction in the size of TMD PDFs uncertainties has been observed in the region at small and intermediate

---

[1]We succeeded in including about the 40% of the SIDIS data set after imposing the cuts on the kinematic variables z and Q. The majority of them belongs to the kinematic region $0.2\,Q < |\boldsymbol{P}_{hT}|/z < Q$.

x with the inclusion of EIC pseudodata (for a specific choice of its energy configuration).

In Ch. 4, we have presented the extraction by the MAP Collaboration of the TMD PDF for the unpolarized quark in the pion, based for the first time on the fit to the whole set of available experimental measurements [2]. The data set included in the analysis reported in this chapter, which we refer to as PionMAP22, is made of 138 data points for the DY lepton pair production in π^--nucleus collisions from experiments conducted at Fermilab (E615 [281] and E537 [45] collaborations).

Such measurements can be described in TMD formalism in terms of TMD PDFs for unpolarized quarks in the proton and in the pion. For the proton, we have considered TMDs extracted in the MAP22 fit. As for the pion, we have modeled the non perturbative part of the TMD PDF as one single Gaussian with x-dependent width, parametrized by three free parameters. We have achieved a fairly good agreement between experimental data and fitted theoretical predictions, with a global $\chi^2/N_{\mathrm{dat}} = 1.55$. We found that half of its value is contributed by fully correlated experimental errors. The pion TMD PDFs obtained by this analysis (displayed in Fig. 4.6) are affected by large error bands due to the lack of experimental information (compared to the proton case). The results of our fit favor wider TMDs of quarks in the pion than in the proton, in agreement with previous analyses [280].

The results of the PionMAP22 fit were used to produce predictions for unpolarized cross section in pion–nucleus DY collisions at the COMPASS kinematics, which will allow to broaden the experimental information on pion TMDs and, consequently, reduce their error bands.

Chapter 5 has been devoted to the investigation of the impact produced by modifications in some of the settings of our TMD phenomenological analyses.

First of all, in order to establish a baseline result, we have illustrated a new fit of only Drell–Yan data, conventionally indicated as MAP22DY, that has the same setup of the MAP22 global analysis. We obtained that most of the fitted parameters of the MAP22DY fit are compatible with the ones of the MAP22 global fit, despite much less constrained. We considered this result as an indication of the consistency and the stability of our fitting framework.

In addition, we estimated the effect of modifying the choice of collinear PDFs included in the structure of TMD PDFs in the proton. To this aim, we performed a new fit of the DY data set with the same setup as the MAP22DY fit, but changing the input collinear PDFs, replacing the baseline MMHT2014 set with the NNPDF3.1 set [220]. We found that the agreement between data and theory for the fit based on the NNPDF3.1 set is similar to the one based on the MMHT2014 set. Also, TMD PDFs extracted from the two fits are compatible with each other, indicating that they marginally depend on their collinear input, contrary to the findings of Ref. [82].

Furthermore, this study indicates that it is important to consider Monte Carlo sets of collinear PDFs in TMD analyses, because they allow to better

identify the uncertainty on the integral of the extracted TMDs.

In accordance with the MAP22 results, we found it difficult to describe AT-LAS data set: the agreement theoretical predictions and experimental data is poor in both the variations of the fit. This led us to test potential inconsistencies of those data points with the rest of the data set through the weighted-fit method [30]. Artificially increasing the statistical weight of ATLAS data, we found that the agreement between experimental data and fitting curves slightly improves, while the agreement with the rest of the (fitted) data set remains almost unchanged, and sometimes worsens. Moreover, the global χ^2 of the weighted fit is almost twice that of the unweighted one. These results suggest that the ATLAS data may show internal inconsistencies, and that is not fully clear whether they are compatible with the rest of the data set.

The impact of discarding ATLAS data has been estimated by a new fit that excludes them. We have obtained an extremely good agreement between theory and the reduced data set (412 data points against the previous 484), with a $\chi^2/N_{dat} < 1$. The free parameters obtained from fitting the reduced data set are all compatible at 1-σ level with those from the baseline global fit.

To our knowledge, this is the first study of compatibility among specific data sets analyzed in TMD phenomenology. It suggests that ATLAS measurements must be treated carefully: including them in the fit largely increases the global χ^2 without significantly modifying the profile of the extracted TMD PDFs. Moreover, improving the quality of the fit might demand unreasonably complex parameterizations of the nonperturbative part of TMDs.

So far we have discussed the currently most sophisticated studies of the 3-D structure of the proton and the pion. We dedicated the last chapter of this thesis to a phenomenological exploration of BSM effects on the structure of nucleons. Indeed, in Ch. 6 we demonstrated that the level of precision reached by collinear PDF extractions make it possible the investigation BSM physics in the QCD sector. In particular, we have explored the impact of violating QCD parity invariance (strong P violation) on nucleon structure and nuclei [3].

Based on Ref. [3], we wrote for the first time the full contribution of parity-violating (PV) PDFs to the structure functions of neutral-current DIS of longitudinally polarized leptons off unpolarized targets. We found that a new P-odd and CP-even PDF, denoted as g_1^{PV}, contributes to the structure function F_3 in the channel of pure photon exchange, which is absent in the SM result. Moreover, we briefly mentioned the case of DIS of an unpolarized lepton beam off a longitudinally polarized target, which leads to the introduction of another PV PDF, denoted f_{1L}^{PV}, which is P-odd, CP-odd, and is related to the proton electric dipole moment.

We have estimated the size of g_1^{PV} through a fit to existing experimental data from HERA, SLAC, and JLab, which are sensitive to PV effects. As an exploratory model, we have assumed that the PV PDFs are proportional to their parity-even counterpart g_1 and we have fitted the proportionality constant, a. We have shown that a better description of the data can be achieved

with the inclusion of strong PV contributions. We have obtained the value $a = (-1.01 \pm 0.66) \cdot 10^{-4}$, which indicates that there could be more left-handed quarks than right-handed ones in an unpolarized proton.

Finally, we have discussed how this study can be expanded to future analyses of experimental data from DIS processes with different lepton beam polarizations and charges, especially from experiments at the JLab 12 GeV program (or its possible 22 GeV upgrade), and at the future Electron Ion Collider. We remark that this phenomenological study is the first one ever attempted in this field, and it could open the door to future studies of the so-called "strong CP" problem, which is deeply connected to the unexplained matter-antimatter imbalance in the Universe.

7.1 Outlook

The results discussed in this thesis show that the internal momentum-space structure of the nucleon can be studied with a high level of precision. This studies can serve as a starting point for a considerable number of future developments. Here, we discuss which kind of steps can be made towards improving and refining such studies.

The level of sophistication of a TMD extraction essentially depends on two ingredients: the accuracy of the theoretical formalism, and the amount of analyzed data from different processes. On the theory side, we plan to improve the perturbative accuracy of the MAP22 extraction to the full N^3LL. Indeed, a NNLO extraction of collinear FFs has been recently become available [33]. At the same time, we plan to consider Monte Carlo sets of collinear PDFs to obtain a better determination of the actual uncertainty of TMD distributions, as shown in Ch. 5. Moreover, in order to get a more reliable determination of theoretical uncertainties, it is important to introduce the evaluation of the errors induced by scale variations [316]. In fact, they provide an estimate of the theoretical uncertainty due to the truncation of QCD calculations at a specific perturbative order, enhancing precision and refining our understanding of hadron transverse momentum structure.

In order to better evaluate the stability of our extractions, it could be useful to test alternative functional forms of the nonperturbative parts of TMD distributions. When dealing with a lot of experimental data and, consequently, very flexible parametrizations with a large number of free parameters, the introduction of techniques based on Neural Networks may bring benefits. In fact, they can capture complex, non-linear relations among experimental data, offering improved modeling power compared to traditional methods. On the other side, one should be careful with overfitting and with the physical interpretation of the results, above all in the kinematic regions that are barely covered by the experimental information.

In order to fully understand the normalization issue in the description of SIDIS multiplicities at intermediate and low energies, we need to improve

our knowledge on the contributions of power suppressed ingredients, such as kinematic power corrections [88, 317], higher-twist ingredients [106, 318–321] or the role of the Y-term at low transverse momentum [86].

On the experimental side, we expect that in the next years further data will become available, and will play a crucial role in deepening our knowledge of unpolarized TMD PDFs. Some preliminary measurements of azimuthal modulation of the unpolarized SIDIS cross section on a proton target have been recently published [322] by the COMPASS Collaboration, and experimental data on hadron multiplicities will probably be released in the near future. It would be very important to include those data in our analysis because they will allow us to considerably extend our present data set. Also, the 12 GeV physics program at Jefferson Lab [323] will be very important to provide measurements in the large-x and low-Q kinematic region. On a longer term, experimental data coming from the Electron-Ion Collider (EIC) will play a crucial role in better constraining unpolarized TMDs in a wide kinematic range (see Fig. 3.33). A similar impact could be produced by the possible JLab 22 GeV upgrade, but in a complementary kinematic region (see Fig. 3.31).

Moreover, we remark that our MAP22 global extraction is currently missing the independent determination of TMD FFs, which will be possible only with the inclusion of experimental measurements of back-to-back hadron production in e^+e^- annihilation sensitive to transverse relative momenta. This kind of data are not available yet, but they should become available from the BELLE collaboration [183].

Apart from the inclusion of new data sets, in this thesis we set the stage for a refined estimate of the compatibility of ATLAS data with the other data sets. This could lead to a better identification of a self-consistent baseline data set, which will allow for more precise unpolarized TMD extractions.

Another possible extension of the MAP22 analysis is the inclusion of partonic flavor dependence in transverse-momentum dynamics. In fact, state-of-the-art extractions of unpolarized TMDs assume that differences among quark flavors are limited to the dependence on the collinear longitudinal momentum. Even if a very nice agreement between theoretical prediction and present data can be achieved without the inclusion of flavor dependence, we expect different quark flavors to give separate contributions to the overall cross section of different experimental processes, and preliminary studies with state-of-the-art collinear FFs seem to indicate that. The first extraction of flavor-dependent TMD PDFs and TMD FFs is reported in Ref. [84]. Recently, some attempts have been performed with improved sophistication, but only through fits of DY experimental data [82, 83]. Along these lines, the upcoming experimental data on SIDIS processes off proton targets from the COMPASS collaboration will be extremely useful. We remark that an accurate extraction of flavor-dependent unpolarized TMDs has been shown to be necessary for precise measurements at high-energies, such as the determination of the W-boson mass [105].

The availability of unpolarized TMD extractions at a high accuracy provides important benefits also to the extraction of polarized TMDs. The study of these nonperturbative objects will allow to investigate possible correlations between partonic transverse momentum and spin, which can be interpreted as spin-orbit correlations at the partonic level [64]. The most prominent example is the so-called Sivers function, $f_{1T}^\perp$, which represents the number density of unpolarized partons inside a transversely polarized nucleon [58, 63–65]. In the next future, we plan to perform the extraction of $f_{1T}^\perp$ in the `NangaParbat` computational tool for TMD studies.

Of crucial importance is also the study of the interplay between hadron phenomenology and calculation on the lattice. In fact, phenomenology should guide lattice calculations in specific areas of interest, while lattice results should be useful to refine phenomenological models, improving their precision and predictive power. Several efforts have been directed to the comparison between phenomenological TMD extractions and lattice calculations of the Collins–Soper kernel, a fundamental ingredient of TMD evolution (see, *e.g.* Refs [249–253]). Instead, only one preliminary determination of the quark unpolarized TMD PDF has been performed in Ref. [103], indicating that further studies are needed to reach a good agreement between phenomenology and ab-initio calculations on the lattice.

Another crucial development of the topics discussed in this thesis is the implementation in present computational tools for TMD fits of the possibility of performing simultaneous analyses of TMDs and collinear distributions. This has already been done for polarized TMDs, such as the Sivers function, and could be attempted also in the unpolarized case, where the experimental information is larger and more sophisticated experimental techniques are required.

Concerning our exploratory study of possible effects of strong P violation in the internal structure of nucleons described in Ch. 6, it can be improved in many aspects. First of all, a more refined model of the PV PDF g_1^{PV} should be developed, starting from the addition of new PV structures in the QCD lagrangian. Then, it would be very useful to identify new observable that are sensitive to the contribution of such PV PDF. In order to study the stability of our result, it would be interesting to include in our analysis experimental data from Drell-Yan measurements. Moreover, we plan to investigate the possibility of performing an estimate of the size of the PV PDF f_{1L}^{PV}, which is P-odd and CP-odd, and should be connected to the electric dipole moment of the proton. We remark that detecting strong P violation could have implications even beyond the understanding of nucleon structure, potentially shedding light on the unexplained matter-antimatter imbalance in the universe.

In summary, this collection of possible future directions underscores the dynamic scientific activity taking place in hadronic physics, and it points towards the expectation of significant developments in this research field in the future years.

List of publications

- MAP Collaboration: A. Bacchetta, V. Bertone, C. Bissolotti, G. Bozzi, M. Cerutti, F .Piacenza, M. Radici, A. Signori, *Unpolarized Transverse Momentum Distributions from a global fit of Drell-Yan and Semi-Inclusive Deep-Inelastic Scattering data*, JHEP 10 (2022) 127 [arXiv:2206.07598].

- MAP Collaboration: M. Cerutti, L. Rossi, S. Venturini, A. Bacchetta, V. Bertone, C. Bissolotti and M. Radici, *Extraction of Pion Transverse Momentum Distributions from Drell-Yan data*, Phys.Rev.D 107 (2023) 1, 014014, [arXiv:2210.01733]

- A. Bacchetta, M. Cerutti, L. Manna, M. Radici and X. Zheng, *Signals of strong parity violation in deep inelastic scattering*, in publication on Phys. Lett. B [arXiv:2306.04704].

- A. Accardi, P. Achenbach, D. Adhikari, A. Afanasev, C.S. Akondi et al., *Strong Interaction Physics at the Luminosity Frontier with 22 GeV Electrons at Jefferson Lab*, in publication on Nucl. Phys. A [arXiv:2306.09360].

165

Bibliography

[1] MAP collaboration, A. Bacchetta, V. Bertone, C. Bissolotti, G. Bozzi, M. Cerutti, F. Piacenza et al., *Unpolarized transverse momentum distributions from a global fit of Drell-Yan and semi-inclusive deep-inelastic scattering data*, *JHEP* **10** (2022) 127, [2206.07598].

[2] (MAP COLLABORATION) collaboration, M. Cerutti, L. Rossi, S. Venturini, A. Bacchetta, V. Bertone, C. Bissolotti et al., *Extraction of pion transverse momentum distributions from Drell-Yan data*, *Phys. Rev. D* **107** (2023) 014014, [2210.01733].

[3] A. Bacchetta, M. Cerutti, L. Manna, M. Radici and X. Zheng, *Signals of strong parity violation in deep inelastic scattering*, [2306.04704].

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

[5] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking, Volume 1* (1964) .

[6] O. W. Greenberg, *Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons*, *Phys. Rev. Lett.* **13** (1964) 598–602.

[7] SLAC-MIT-CAL TECH collaboration, W. K. H. Panofsky et al., *PROPOSALS FOR INITIAL ELECTRON SCATTERING EXPERIMENTS USING THE SLAC SPECTROMETER FACILITIES*, .

[8] E. D. Bloom et al., *High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees*, *Phys. Rev. Lett.* **23** (1969) 930–934.

[9] M. Breidenbach, J. I. Friedman, H. W. Kendall, E. D. Bloom, D. H. Coward, H. C. DeStaebler et al., *Observed behavior of highly inelastic electron-proton scattering*, *Phys. Rev. Lett.* **23** (1969) 935–939.

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

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

[12] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, *Phys. Rev. Lett.* **30** (1973) 1343–1346.

[13] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, *Phys. Rev. Lett.* **30** (1973) 1346–1349.

[14] K. G. Wilson, *Nonlagrangian models of current algebra*, *Phys. Rev.* **179** (1969) 1499–1512.

[15] EUROPEAN MUON collaboration, J. P. Albanese et al., *The Vertex and Large Angle Detectors of a Spectrometer System for High-energy Muon Physics*, *Nucl. Instrum. Meth.* **212** (1983) 111.

[16] EUROPEAN MUON collaboration, J. J. Aubert et al., *A Detailed Study of the Proton Structure Functions in Deep Inelastic Muon - Proton Scattering*, *Nucl. Phys. B* **259** (1985) 189.

[17] D. Bollini et al., *A High Luminosity Spectrometer for Deep Inelastic Muon Scattering Experiments*, *Nucl. Instrum. Meth.* **204** (1983) 333.

[18] BOLOGNA-CERN-DUBNA-MUNICH-SACLAY collaboration, A. C. Benvenuti et al., *An Upgraded Configuration of a High Luminosity Spectrometer for Deep Inelastic Muon Scattering Experiments*, *Nucl. Instrum. Meth. A* **226** (1984) 330.

[19] G. K. Mallot and R. Voss, *Deep inelastic scattering with the SPS muon beam*, *Adv. Ser. Direct. High Energy Phys.* **23** (2015) 287–311.

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

[21] L. Baulieu and C. Kounnas, *Equivalence of the 'Wilson Expansion' and a Generalized Parton Model*, *Nucl. Phys. B* **141** (1978) 423–431.

[22] J. Kodaira and T. Uematsu, *How Can We Define Properly the Q^2 Dependent Parton Distribution Function in QCD?*, *Nucl. Phys. B* **141** (1978) 497–506.

[23] T. Uematsu, *Q**2 Dependence of Quark and Gluon Fragmentation Functions in a Parton Picture Based on QCD*, *Phys. Lett. B* **79** (1978) 97–102.

[24] J. F. Owens, *On the Q**2 Dependence of Parton Fragmentation Functions*, *Phys. Lett. B* **76** (1978) 85–88.

[25] 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.

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

[27] 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, [hep-ph/0409313].

[28] T. C. Rogers, *An overview of transverse-momentum–dependent factorization and evolution*, Eur. Phys. J. A **52** (2016) 153, [1509.04766].

[29] T.-J. Hou et al., *Progress in the CTEQ-TEA NNLO global QCD analysis*, [1908.11394].

[30] NNPDF collaboration, R. D. Ball et al., *The path to proton structure at 1% accuracy*, Eur. Phys. J. C **82** (2022) 428, [2109.02653].

[31] J. McGowan, T. Cridge, L. A. Harland-Lang and R. S. Thorne, *Approximate N^3LO parton distribution functions with theoretical uncertainties: MSHT20aN^3LO PDFs*, Eur. Phys. J. C **83** (2023) 185, [2207.04739].

[32] D. de Florian, M. Epele, R. J. Hernandez-Pinto, R. Sassot and M. Stratmann, *Parton-to-Kaon Fragmentation Revisited*, Phys. Rev. D **95** (2017) 094019, [1702.06353].

[33] MAP (MULTI-DIMENSIONAL ANALYSES OF PARTONIC DISTRIBUTIONS) collaboration, R. Abdul Khalek, V. Bertone, A. Khoudli and E. R. Nocera, *Pion and kaon fragmentation functions at next-to-next-to-leading order*, Phys. Lett. B **834** (2022) 137456, [2204.10331].

[34] J. C. Collins, *What exactly is a parton density?*, Acta Phys. Polon. B **34** (2003) 3103, [hep-ph/0304122].

[35] G. L. Kane, J. Pumplin and W. Repko, *Transverse Quark Polarization in Large p(T) Reactions, e+ e- Jets, and Leptoproduction: A Test of QCD*, Phys. Rev. Lett. **41** (1978) 1689.

[36] A. Brandenburg, S. J. Brodsky, V. V. Khoze and D. Mueller, *Angular distributions in the Drell-Yan process: A Closer look at higher twist effects*, Phys. Rev. Lett. **73** (1994) 939–942, [hep-ph/9403361].

[37] K. J. Eskola, P. Hoyer, M. Vanttinen and R. Vogt, *Higher twist effects in the Drell-Yan angular distribution*, *Phys. Lett. B* **333** (1994) 526–530, [hep-ph/9404322].

[38] M. Lambertsen and W. Vogelsang, *Drell-Yan lepton angular distributions in perturbative QCD*, *PoS* **QCDEV2016** (2017) 045.

[39] G. Bunce et al., *Lambda0 Hyperon Polarization in Inclusive Production by 300-GeV Protons on Beryllium.*, *Phys. Rev. Lett.* **36** (1976) 1113–1116.

[40] K. J. Heller et al., *Polarization of Lambdas and anti-Lambdas Produced by 400-GeV Protons*, *Phys. Rev. Lett.* **41** (1978) 607.

[41] FNAL-E704 collaboration, D. L. Adams et al., *Analyzing Power in Inclusive π^+ and π^- Production at High χ_F with a 200 GeV Polarized Proton Beam*, *Phys. Lett. B* **264** (1991) 462–466.

[42] STAR collaboration, J. Adams et al., *Cross-sections and transverse single spin asymmetries in forward neutral pion production from proton collisions at s**(1/2) = 200- GeV*, *Phys. Rev. Lett.* **92** (2004) 171801, [hep-ex/0310058].

[43] NA10 collaboration, S. Falciano et al., *Angular Distributions of Muon Pairs Produced by 194-GeV/c Negative Pions*, *Z. Phys. C* **31** (1986) 513.

[44] NA10 collaboration, M. Guanziroli et al., *Angular Distributions of Muon Pairs Produced by Negative Pions on Deuterium and Tungsten*, *Z. Phys. C* **37** (1988) 545.

[45] E. Anassontzis et al., *High mass dimuon production in $\bar{p}n$ and $\pi^- n$ interactions at 125-GeV/c*, *Phys. Rev. D* **38** (1988) 1377.

[46] J. S. Conway et al., *Experimental Study of Muon Pairs Produced by 252-GeV Pions on Tungsten*, *Phys. Rev. D* **39** (1989) 92–122.

[47] EUROPEAN MUON collaboration, J. Ashman et al., *A Measurement of the Spin Asymmetry and Determination of the Structure Function g(1) in Deep Inelastic Muon-Proton Scattering*, *Phys. Lett. B* **206** (1988) 364.

[48] X. Ji, F. Yuan and Y. Zhao, *What we know and what we don't know about the proton spin after 30 years*, *Nature Rev. Phys.* **3** (2021) 27–38, [2009.01291].

[49] M. Diehl, *Introduction to GPDs and TMDs*, *Eur. Phys. J. A* **52** (2016) 149, [1512.01328].

[50] R. Angeles-Martinez et al., *Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects*, Acta Phys. Polon. B **46** (2015) 2501–2534, [`1507.05267`].

[51] P. J. Mulders and R. D. Tangerman, *The Complete tree level result up to order 1/Q for polarized deep inelastic leptoproduction*, Nucl. Phys. B **461** (1996) 197–237, [`hep-ph/9510301`].

[52] A. Metz and A. Vossen, *Parton Fragmentation Functions*, Prog. Part. Nucl. Phys. **91** (2016) 136–202, [`1607.02521`].

[53] M. Anselmino, A. Mukherjee and A. Vossen, *Transverse spin effects in hard semi-inclusive collisions*, Prog. Part. Nucl. Phys. **114** (2020) 103806, [`2001.05415`].

[54] D. Boer and P. J. Mulders, *Time reversal odd distribution functions in leptoproduction*, Phys. Rev. D **57** (1998) 5780–5786, [`hep-ph/9711485`].

[55] A. Bacchetta, M. Diehl, K. Goeke, A. Metz, P. J. Mulders and M. Schlegel, *Semi-inclusive deep inelastic scattering at small transverse momentum*, JHEP **02** (2007) 093, [`hep-ph/0611265`].

[56] A. Metz, M. Schlegel and K. Goeke, *Transverse single spin asymmetries in inclusive deep-inelastic scattering*, Phys. Lett. B **643** (2006) 319–324, [`hep-ph/0610112`].

[57] R. D. Tangerman and P. J. Mulders, *Polarized twist - three distributions g(T) and h(L) and the role of intrinsic transverse momentum*, [`hep-ph/9408305`].

[58] M. Anselmino, M. Boglione, U. D'Alesio, A. Kotzinian, S. Melis, F. Murgia et al., *Sivers Effect for Pion and Kaon Production in Semi-Inclusive Deep Inelastic Scattering*, Eur. Phys. J. A **39** (2009) 89–100, [`0805.2677`].

[59] M. Anselmino, M. Boglione, U. D'Alesio, J. O. Gonzalez Hernandez, S. Melis, F. Murgia et al., *Collins functions for pions from SIDIS and new e^+e^- data: a first glance at their transverse momentum dependence*, Phys. Rev. D **92** (2015) 114023, [`1510.05389`].

[60] H.-W. Lin, W. Melnitchouk, A. Prokudin, N. Sato and H. Shows, *First Monte Carlo Global Analysis of Nucleon Transversity with Lattice QCD Constraints*, Phys. Rev. Lett. **120** (2018) 152502, [`1710.09858`].

[61] C. Lefky and A. Prokudin, *Extraction of the distribution function $h_{1T}^{\perp}$ from experimental data*, Phys. Rev. D **91** (2015) 034010, [`1411.0580`].

[62] M. Boglione, S. Melis and A. Prokudin, *Partonic Transverse Motion in Unpolarized Semi-Inclusive Deep Inelastic Scattering Processes*, *Phys. Rev. D* **84** (2011) 034033, [1106.6177].

[63] M. G. Echevarria, Z.-B. Kang and J. Terry, *Global analysis of the Sivers functions at NLO+NNLL in QCD*, *JHEP* **01** (2021) 126, [2009.10710].

[64] A. Bacchetta, F. Delcarro, C. Pisano and M. Radici, *The 3-dimensional distribution of quarks in momentum space*, *Phys. Lett. B* **827** (2022) 136961, [2004.14278].

[65] M. Bury, A. Prokudin and A. Vladimirov, *Extraction of the Sivers function from SIDIS, Drell-Yan, and $W^{\pm}/Z$ boson production data with TMD evolution*, *JHEP* **05** (2021) 151, [2103.03270].

[66] S. Bhattacharya, Z.-B. Kang, A. Metz, G. Penn and D. Pitonyak, *First global QCD analysis of the TMD g1T from semi-inclusive DIS data*, *Phys. Rev. D* **105** (2022) 034007, [2110.10253].

[67] Z.-B. Kang, A. Prokudin, P. Sun and F. Yuan, *Extraction of Quark Transversity Distribution and Collins Fragmentation Functions with QCD Evolution*, *Phys. Rev. D* **93** (2016) 014009, [1505.05589].

[68] L. Gamberg, Z.-B. Kang, D. Y. Shao, J. Terry and F. Zhao, *Transverse Λ polarization in $e^+ e^-$ collisions*, *Phys. Lett. B* **818** (2021) 136371, [2102.05553].

[69] U. D'Alesio, L. Gamberg, F. Murgia and M. Zaccheddu, *Transverse Λ polarization in e^+e^- processes within a TMD factorization approach and the polarizing fragmentation function*, *JHEP* **12** (2022) 074, [2209.11670].

[70] J. C. Collins and D. E. Soper, *Back-To-Back Jets in QCD*, *Nucl. Phys. B* **193** (1981) 381.

[71] J. C. Collins, D. E. Soper and G. F. Sterman, *Transverse Momentum Distribution in Drell-Yan Pair and W and Z Boson Production*, *Nucl. Phys. B* **250** (1985) 199–224.

[72] J. C. Collins, D. E. Soper and G. F. Sterman, *Factorization for Short Distance Hadron - Hadron Scattering*, *Nucl. Phys. B* **261** (1985) 104–142.

[73] J. Collins, *Foundations of perturbative QCD*, vol. 32. Cambridge University Press, 11, 2013.

[74] C. T. H. Davies, B. R. Webber and W. J. Stirling, *Drell-Yan Cross-Sections at Small Transverse Momentum*, *Nucl. Phys. B* **256** (1985) 413.

[75] J.-w. Qiu and X.-f. Zhang, *Role of the nonperturbative input in QCD resummed Drell-Yan Q_T distributions*, Phys. Rev. D **63** (2001) 114011, [hep-ph/0012348].

[76] T. Becher, M. Neubert and D. Wilhelm, *Electroweak Gauge-Boson Production at Small q_T: Infrared Safety from the Collinear Anomaly*, JHEP **02** (2012) 124, [1109.6027].

[77] R. Meng, F. I. Olness and D. E. Soper, *Semiinclusive deeply inelastic scattering at small q(T)*, Phys. Rev. D **54** (1996) 1919–1935, [hep-ph/9511311].

[78] U. D'Alesio, M. G. Echevarria, S. Melis and I. Scimemi, *Non-perturbative QCD effects in q_T spectra of Drell-Yan and Z-boson production*, JHEP **11** (2014) 098, [1407.3311].

[79] I. Scimemi and A. Vladimirov, *Analysis of vector boson production within TMD factorization*, Eur. Phys. J. C **78** (2018) 89, [1706.01473].

[80] V. Bertone, I. Scimemi and A. Vladimirov, *Extraction of unpolarized quark transverse momentum dependent parton distributions from Drell-Yan/Z-boson production*, JHEP **06** (2019) 028, [1902.08474].

[81] A. Bacchetta, V. Bertone, C. Bissolotti, G. Bozzi, F. Delcarro, F. Piacenza et al., *Transverse-momentum-dependent parton distributions up to N^3LL from Drell-Yan data*, JHEP **07** (2020) 117, [1912.07550].

[82] M. Bury, F. Hautmann, S. Leal-Gomez, I. Scimemi, A. Vladimirov and P. Zurita, *PDF bias and flavor dependence in TMD distributions*, JHEP **10** (2022) 118, [2201.07114].

[83] V. Moos, I. Scimemi, A. Vladimirov and P. Zurita, *Extraction of unpolarized transverse momentum distributions from fit of Drell-Yan data at N^4LL*, [2305.07473].

[84] A. Signori, A. Bacchetta, M. Radici and G. Schnell, *Investigations into the flavor dependence of partonic transverse momentum*, JHEP **11** (2013) 194, [1309.3507].

[85] M. Anselmino, M. Boglione, J. O. Gonzalez Hernandez, S. Melis and A. Prokudin, *Unpolarised Transverse Momentum Dependent Distribution and Fragmentation Functions from SIDIS Multiplicities*, JHEP **04** (2014) 005, [1312.6261].

[86] P. Sun, J. Isaacson, C. P. Yuan and F. Yuan, *Nonperturbative functions for SIDIS and Drell–Yan processes*, Int. J. Mod. Phys. A **33** (2018) 1841006, [1406.3073].

[87] A. Bacchetta, F. Delcarro, C. Pisano, M. Radici and A. Signori, *Extraction of partonic transverse momentum distributions from semi-inclusive deep-inelastic scattering, Drell-Yan and Z-boson production*, *JHEP* **06** (2017) 081, [1703.10157].

[88] I. Scimemi and A. Vladimirov, *Non-perturbative structure of semi-inclusive deep-inelastic and Drell-Yan scattering at small transverse momentum*, *JHEP* **06** (2020) 137, [1912.06532].

[89] A. Bermudez Martinez, P. Connor, H. Jung, A. Lelek, R. Zlebcik, F. Hautmann et al., *Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method*, *Phys. Rev. D* **99** (2019) 074008, [1804.11152].

[90] A. Bermudez Martinez et al., *Production of Z-bosons in the parton branching method*, *Phys. Rev. D* **100** (2019) 074027, [1906.00919].

[91] A. Bermudez Martinez et al., *The transverse momentum spectrum of low mass Drell–Yan production at next-to-leading order in the parton branching method*, *Eur. Phys. J. C* **80** (2020) 598, [2001.06488].

[92] A. B. Martinez, *Transformation of transverse momentum distributions from Parton Branching to Collins-Soper-Sterman framework*, [2307.06704].

[93] ALPHA collaboration, M. Bruno, I. Campos, P. Fritzsch, J. Koponen, C. Pena, D. Preti et al., *Light quark masses in $N_\mathrm{f} = 2 + 1$ lattice QCD with Wilson fermions*, *Eur. Phys. J. C* **80** (2020) 169, [1911.08025].

[94] EXTENDED TWISTED MASS collaboration, C. Alexandrou et al., *Quark masses using twisted-mass fermion gauge ensembles*, *Phys. Rev. D* **104** (2021) 074515, [2104.13408].

[95] R. G. Edwards, J. J. Dudek, D. G. Richards and S. J. Wallace, *Excited state baryon spectroscopy from lattice QCD*, *Phys. Rev. D* **84** (2011) 074508, [1104.5152].

[96] A. Bazavov et al., *B- and D-meson leptonic decay constants from four-flavor lattice QCD*, *Phys. Rev. D* **98** (2018) 074512, [1712.09262].

[97] (HPQCD COLLABORATION)§, HPQCD collaboration, W. G. Parrott, C. Bouchard and C. T. H. , *B→K and D→K form factors from fully relativistic lattice QCD*, *Phys. Rev. D* **107** (2023) 014510, [2207.12468].

[98] FERMILAB LATTICE, MILC collaboration, A. Bazavov et al., *$|V_{us}|$ from $K_{\ell 3}$ decay and four-flavor lattice QCD*, *Phys. Rev. D* **99** (2019) 114509, [1809.02827].

[99] N. Miller et al., *F_K/F_π from Möbius Domain-Wall fermions solved on gradient-flowed HISQ ensembles*, Phys. Rev. D **102** (2020) 034507, [2005.04795].

[100] L. Del Debbio and A. Ramos, *Lattice determinations of the strong coupling*, [2101.04762].

[101] Flavour Lattice Averaging Group (FLAG) collaboration, Y. Aoki et al., *FLAG Review 2021*, Eur. Phys. J. C **82** (2022) 869, [2111.09849].

[102] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen and F. Manigrasso, *Flavor decomposition of the nucleon unpolarized, helicity, and transversity parton distribution functions from lattice QCD simulations*, Phys. Rev. D **104** (2021) 054503, [2106.16065].

[103] LPC collaboration, J.-C. He, M.-H. Chu, J. Hua, X. Ji, A. Schäfer, Y. Su et al., *Unpolarized Transverse-Momentum-Dependent Parton Distributions of the Nucleon from Lattice QCD*, [2211.02340].

[104] M. Constantinou et al., *Parton distributions and lattice-QCD calculations: Toward 3D structure*, Prog. Part. Nucl. Phys. **121** (2021) 103908, [2006.08636].

[105] A. Bacchetta, G. Bozzi, M. Radici, M. Ritzmann and A. Signori, *Effect of Flavor-Dependent Partonic Transverse Momentum on the Determination of the W Boson Mass in Hadronic Collisions*, Phys. Lett. B **788** (2019) 542–545, [1807.02101].

[106] A. Accardi et al., *Strong Interaction Physics at the Luminosity Frontier with 22 GeV Electrons at Jefferson Lab, submitted to Nucl. Phys. A* (6, 2023) , [2306.09360].

[107] R. Abdul Khalek et al., *Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report*, Nucl. Phys. A **1026** (2022) 122447, [2103.05419].

[108] D. Rainwater and T. M. P. Tait, *Testing Grand Unification at the (S)LHC*, Phys. Rev. D **75** (2007) 115014, [hep-ph/0701093].

[109] D. S. M. Alves, J. Galloway, J. T. Ruderman and J. R. Walsh, *Running Electroweak Couplings as a Probe of New Physics*, JHEP **02** (2015) 007, [1410.6810].

[110] M. Farina, G. Panico, D. Pappadopulo, J. T. Ruderman, R. Torre and A. Wulzer, *Energy helps accuracy: electroweak precision tests at hadron colliders*, Phys. Lett. B **772** (2017) 210–215, [1609.08157].

[111] C. Gross, O. Lebedev and J. M. No, *Drell–Yan constraints on new electroweak states: LHC as a pp → l⁺l⁻ precision machine*, Mod. Phys. Lett. A **32** (2017) 1750094, [`1602.03877`].

[112] S. Amoroso, M. Chiesa, C. L. Del Pio, K. Lipka, F. Piccinini, F. Vazzoler et al., *Probing the weak mixing angle at high energies at the LHC and HL-LHC*, [`2302.10782`].

[113] G. Colangelo, M. Hoferichter, A. Nyffeler, M. Passera and P. Stoffer, *Remarks on higher-order hadronic corrections to the muon g−2*, Phys. Lett. B **735** (2014) 90–91, [`1403.7512`].

[114] E. Budassi, C. M. Carloni Calame, C. L. Del Pio and F. Piccinini, *Single π0 production in µe scattering at MUonE*, Phys. Lett. B **829** (2022) 137138, [`2203.01639`].

[115] I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics: A Practical Introduction, Volume 2: Non-Abelian Gauge Theories : QCD and The Electroweak Theory, Fourth Edition.* Taylor & Francis, 2013, 10.1201/9781466513105.

[116] R. D. Peccei and H. R. Quinn, *Some Aspects of Instantons*, Nuovo Cim. A **41** (1977) 309.

[117] Z.-B. Kang and D. E. Kharzeev, *Quark fragmentation in the θ-vacuum*, Phys. Rev. Lett. **106** (2011) 042001, [`1006.2132`].

[118] W. Yang, *Parity odd Fragmentation Functions*, Int. J. Mod. Phys. A **34** (2019) 1950144, [`1906.03427`].

[119] W. Yang, *Parity-odd parton distribution functions from θ-vacuum*, Int. J. Mod. Phys. A **34** (2019) 1950145, [`1907.00129`].

[120] S. D. Drell and T.-M. Yan, *Massive Lepton Pair Production in Hadron-Hadron Collisions at High-Energies*, Phys. Rev. Lett. **25** (1970) 316–320.

[121] S. Arnold, A. Metz and M. Schlegel, *Dilepton production from polarized hadron hadron collisions*, Phys. Rev. D **79** (2009) 034005, [`0809.2262`].

[122] D. Boer, P. J. Mulders and O. V. Teryaev, *Single spin asymmetries from a gluonic background in the Drell-Yan process*, Phys. Rev. D **57** (1998) 3057–3064, [`hep-ph/9710223`].

[123] D. Boer and W. Vogelsang, *Drell-Yan lepton angular distribution at small transverse momentum*, Phys. Rev. D **74** (2006) 014004, [`hep-ph/0604177`].

[124] C. S. Lam and W.-K. Tung, *A Systematic Approach to Inclusive Lepton Pair Production in Hadronic Collisions*, Phys. Rev. D **18** (1978) 2447.

[125] E. Mirkes and J. Ohnemus, *Angular distributions of Drell-Yan lepton pairs at the Tevatron: Order $\alpha - s^2$ corrections and Monte Carlo studies*, Phys. Rev. D **51** (1995) 4891–4904, [hep-ph/9412289].

[126] X. Chen, T. Gehrmann, E. W. N. Glover, A. Huss, P. Monni, E. Re et al., *Third order fiducial predictions for Drell-Yan at the LHC*, [2203.01565].

[127] D. Boer, *Investigating the origins of transverse spin asymmetries at RHIC*, Phys. Rev. D **60** (1999) 014012, [hep-ph/9902255].

[128] J. Collins and T. C. Rogers, *Connecting Different TMD Factorization Formalisms in QCD*, Phys. Rev. D **96** (2017) 054011, [1705.07167].

[129] M. G. Echevarria, A. Idilbi and I. Scimemi, *Factorization Theorem For Drell-Yan At Low q_T And Transverse Momentum Distributions On-The-Light-Cone*, JHEP **07** (2012) 002, [1111.4996].

[130] S. Catani and M. Grazzini, *An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC*, Phys. Rev. Lett. **98** (2007) 222002, [hep-ph/0703012].

[131] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO*, Phys. Rev. Lett. **103** (2009) 082001, [0903.2120].

[132] J. J. Ethier and E. R. Nocera, *Parton Distributions in Nucleons and Nuclei*, Ann. Rev. Nucl. Part. Sci. **70** (2020) 43–76, [2001.07722].

[133] M. Alrashed, D. Anderle, Z.-B. Kang, J. Terry and H. Xing, *Three-dimensional imaging in nuclei*, Phys. Rev. Lett. **129** (2022) 242001, [2107.12401].

[134] Jefferson Lab Angular Momentum (JAM) collaboration, P. C. Barry, L. Gamberg, W. Melnitchouk, E. Moffat, D. Pitonyak, A. Prokudin et al., *Tomography of pions and protons via transverse momentum dependent distributions*, Phys. Rev. D **108** (2023) L091504, [2302.01192].

[135] M. Boglione, A. Dotson, L. Gamberg, S. Gordon, J. O. Gonzalez-Hernandez, A. Prokudin et al., *Mapping the Kinematical Regimes of Semi-Inclusive Deep Inelastic Scattering*, JHEP **10** (2019) 122, [1904.12882].

[136] A. Bacchetta, U. D'Alesio, M. Diehl and C. A. Miller, *Single-spin asymmetries: The Trento conventions*, *Phys. Rev. D* **70** (2004) 117504, [hep-ph/0410050].

[137] D. Boer et al., *Gluons and the quark sea at high energies: Distributions, polarization, tomography*, [1108.1713].

[138] A. Bacchetta, D. Boer, M. Diehl and P. J. Mulders, *Matches and mismatches in the descriptions of semi-inclusive processes at low and high transverse momentum*, *JHEP* **08** (2008) 023, [0803.0227].

[139] A. Bacchetta, G. Bozzi, M. G. Echevarria, C. Pisano, A. Prokudin and M. Radici, *Azimuthal asymmetries in unpolarized SIDIS and Drell-Yan processes: a case study towards TMD factorization at subleading twist*, *Phys. Lett. B* **797** (2019) 134850, [1906.07037].

[140] T. Maji and D. Chakrabarti, *Transverse structure of a proton in a light-front quark-diquark model*, *Phys. Rev. D* **95** (2017) 074009, [1702.04557].

[141] COMPASS collaboration, M. Aghasyan et al., *Transverse-momentum-dependent Multiplicities of Charged Hadrons in Muon-Deuteron Deep Inelastic Scattering*, *Phys. Rev. D* **97** (2018) 032006, [1709.07374].

[142] HERMES collaboration, A. Airapetian et al., *Multiplicities of charged pions and kaons from semi-inclusive deep-inelastic scattering by the proton and the deuteron*, *Phys. Rev. D* **87** (2013) 074029, [1212.5407].

[143] M. Grewal, Z.-B. Kang, J.-W. Qiu and A. Signori, *Predictive power of transverse-momentum-dependent distributions*, *Phys. Rev. D* **101** (2020) 114023, [2003.07453].

[144] J.-y. Chiu, A. Jain, D. Neill and I. Z. Rothstein, *The Rapidity Renormalization Group*, *Phys. Rev. Lett.* **108** (2012) 151601, [1104.0881].

[145] J.-Y. Chiu, A. Jain, D. Neill and I. Z. Rothstein, *A Formalism for the Systematic Treatment of Rapidity Logarithms in Quantum Field Theory*, *JHEP* **05** (2012) 084, [1202.0814].

[146] I. Scimemi and A. Vladimirov, *Systematic analysis of double-scale evolution*, *JHEP* **08** (2018) 003, [1803.11089].

[147] G. Billis, F. J. Tackmann and J. Talbert, *Higher-Order Sudakov Resummation in Coupled Gauge Theories*, *JHEP* **03** (2020) 182, [1907.02971].

[148] W. Bizoń, X. Chen, A. Gehrmann-De Ridder, T. Gehrmann, N. Glover, A. Huss et al., *Fiducial distributions in Higgs and Drell-Yan production at $N^3LL+NNLO$*, JHEP **12** (2018) 132, [1805.05916].

[149] S. Catani, L. Cieri, D. de Florian, G. Ferrera and M. Grazzini, *Vector boson production at hadron colliders: hard-collinear coefficients at the NNLO*, Eur. Phys. J. C **72** (2012) 2195, [1209.0158].

[150] M. G. Echevarria, I. Scimemi and A. Vladimirov, *Unpolarized Transverse Momentum Dependent Parton Distribution and Fragmentation Functions at next-to-next-to-leading order*, JHEP **09** (2016) 004, [1604.07869].

[151] M.-x. Luo, T.-Z. Yang, H. X. Zhu and Y. J. Zhu, *Quark Transverse Parton Distribution at the Next-to-Next-to-Next-to-Leading Order*, Phys. Rev. Lett. **124** (2020) 092001, [1912.05778].

[152] Y. Li and H. X. Zhu, *Bootstrapping Rapidity Anomalous Dimensions for Transverse-Momentum Resummation*, Phys. Rev. Lett. **118** (2017) 022004, [1604.01404].

[153] J. Davies, A. Vogt, B. Ruijl, T. Ueda and J. A. M. Vermaseren, *Large-nf contributions to the four-loop splitting functions in QCD*, Nucl. Phys. B **915** (2017) 335–362, [1610.07477].

[154] S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren and A. Vogt, *Four-Loop Non-Singlet Splitting Functions in the Planar Limit and Beyond*, JHEP **10** (2017) 041, [1707.08315].

[155] S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren and A. Vogt, *On quartic colour factors in splitting functions and the gluon cusp anomalous dimension*, Phys. Lett. B **782** (2018) 627–632, [1805.09638].

[156] I. W. Stewart, F. J. Tackmann, J. R. Walsh and S. Zuberi, *Jet p_T resummation in Higgs production at $NNLL' + NNLO$*, Phys. Rev. D **89** (2014) 054001, [1307.1808].

[157] M. A. Ebert and F. J. Tackmann, *Resummation of Transverse Momentum Distributions in Distribution Space*, JHEP **02** (2017) 110, [1611.08610].

[158] S. Catani, *Higher order QCD corrections in hadron collisions: Soft gluon resummation and exponentiation*, Nucl. Phys. B Proc. Suppl. **54** (1997) 107–113, [hep-ph/9610413].

[159] G. Bozzi, S. Catani, G. Ferrera, D. de Florian and M. Grazzini, *Production of Drell-Yan lepton pairs in hadron collisions: Transverse-momentum resummation at next-to-next-to-leading logarithmic accuracy*, Phys. Lett. B **696** (2011) 207–213, [1007.2351].

[160] T. Becher, M. Neubert and D. Wilhelm, *Higgs-Boson Production at Small Transverse Momentum*, *JHEP* **05** (2013) 110, [`1212.2621`].

[161] A. Banfi, H. McAslan, P. F. Monni and G. Zanderighi, *The two-jet rate in e^+e^- at next-to-next-to-leading-logarithmic order*, *Phys. Rev. Lett.* **117** (2016) 172001, [`1607.03111`].

[162] G. Bozzi, S. Catani, D. de Florian and M. Grazzini, *Transverse-momentum resummation and the spectrum of the Higgs boson at the LHC*, *Nucl. Phys. B* **737** (2006) 73–120, [`hep-ph/0508068`].

[163] S. Catani, L. Cieri, D. de Florian, G. Ferrera and M. Grazzini, *Universality of transverse-momentum resummation and hard factors at the NNLO*, *Nucl. Phys. B* **881** (2014) 414–443, [`1311.1654`].

[164] C. Muselli, S. Forte and G. Ridolfi, *Combined threshold and transverse momentum resummation for inclusive observables*, *JHEP* **03** (2017) 106, [`1701.01464`].

[165] S. Alioli, A. Broggio, S. Kallweit, M. A. Lim and L. Rottoli, *Higgsstrahlung at NNLL'+NNLO matched to parton showers in GENEVA*, *Phys. Rev. D* **100** (2019) 096016, [`1909.02026`].

[166] R. Boussarie et al., *TMD Handbook*, [`2304.03302`].

[167] S. Catani, M. L. Mangano, P. Nason and L. Trentadue, *The Resummation of soft gluons in hadronic collisions*, *Nucl. Phys. B* **478** (1996) 273–310, [`hep-ph/9604351`].

[168] E. Laenen, G. F. Sterman and W. Vogelsang, *Higher order QCD corrections in prompt photon production*, *Phys. Rev. Lett.* **84** (2000) 4296–4299, [`hep-ph/0002078`].

[169] G. Lustermans, J. K. L. Michel, F. J. Tackmann and W. J. Waalewijn, *Joint two-dimensional resummation in q_T and 0-jettiness at NNLL*, *JHEP* **03** (2019) 124, [`1901.03331`].

[170] A. Bacchetta, M. G. Echevarria, P. J. G. Mulders, M. Radici and A. Signori, *Effects of TMD evolution and partonic flavor on e^+e^- annihilation into hadrons*, *JHEP* **11** (2015) 076, [`1508.00402`].

[171] J. Collins, L. Gamberg, A. Prokudin, T. C. Rogers, N. Sato and B. Wang, *Relating Transverse Momentum Dependent and Collinear Factorization Theorems in a Generalized Formalism*, *Phys. Rev. D* **94** (2016) 034014, [`1605.00671`].

[172] D. Boer and W. J. den Dunnen, *TMD evolution and the Higgs transverse momentum distribution*, *Nucl. Phys. B* **886** (2014) 421–435, [`1404.6753`].

[173] J. O. Gonzalez-Hernandez, T. Rainaldi and T. C. Rogers, *Resolution to the problem of consistent large transverse momentum in TMDs*, Phys. Rev. D **107** (2023) 094029, [`2303.04921`].

[174] J. Osvaldo Gonzalez-Hernandez, *Comments on the perturbative and non-perturbative contributions in unpolarized SIDIS*, PoS **DIS2019** (2019) 176.

[175] M. G. Echevarria, A. Idilbi, Z.-B. Kang and I. Vitev, *QCD Evolution of the Sivers Asymmetry*, Phys. Rev. D **89** (2014) 074013, [`1401.5078`].

[176] F. Piacenza, *Perturbative and nonperturbative QCD regimes in transverse-momentum dependent observables*. PhD thesis, Università di Pavia, Pavia U., 2020.

[177] J. O. Gonzalez-Hernandez, T. C. Rogers, N. Sato and B. Wang, *Challenges with Large Transverse Momentum in Semi-Inclusive Deeply Inelastic Scattering*, Phys. Rev. D **98** (2018) 114005, [`1808.04396`].

[178] A. Bacchetta, G. Bozzi, M. Lambertsen, F. Piacenza, J. Steiglechner and W. Vogelsang, *Difficulties in the description of Drell-Yan processes at moderate invariant mass and high transverse momentum*, Phys. Rev. D **100** (2019) 014018, [`1901.06916`].

[179] E. Moffat, T. C. Rogers, N. Sato and A. Signori, *Collinear factorization in wide-angle hadron pair production in e^+e^- annihilation*, Phys. Rev. D **100** (2019) 094014, [`1909.02951`].

[180] 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, [`hep-ph/0703242`].

[181] D. de Florian, M. Stratmann and W. Vogelsang, *QCD analysis of unpolarized and polarized Lambda baryon production in leading and next-to-leading order*, Phys. Rev. D **57** (1998) 5811–5824, [`hep-ph/9711387`].

[182] G. Bozzi, S. Catani, D. de Florian and M. Grazzini, *The q(T) spectrum of the Higgs boson at the LHC in QCD perturbation theory*, Phys. Lett. B **564** (2003) 65–72, [`hep-ph/0302104`].

[183] Belle collaboration, R. Seidl et al., *Transverse momentum dependent production cross sections of charged pions, kaons and protons produced in inclusive e^+e^- annihilation at $\sqrt{s} = 10.58$ GeV*, Phys. Rev. D **99** (2019) 112006, [`1902.01552`].

[184] Z.-B. Kang, D. Y. Shao and F. Zhao, *QCD resummation on single hadron transverse momentum distribution with the thrust axis*, JHEP **12** (2020) 127, [`2007.14425`].

[185] Y. Makris, F. Ringer and W. J. Waalewijn, *Joint thrust and TMD resummation in electron-positron and electron-proton collisions*, *JHEP* **02** (2021) 070, [2009.11871].

[186] M. Boglione and A. Simonelli, *Kinematic regions in the $e^+e^- \to h\,X$ factorized cross section in a 2-jet topology with thrust*, *JHEP* **02** (2022) , [2109.11497].

[187] M. Boglione and A. Simonelli, *Full treatment of the thrust distribution in single inclusive $e^+e^- \to h\,X$ processes*, *JHEP* **09** (2023) 006, [2306.02937].

[188] M. Boglione, J. O. Gonzalez-Hernandez and A. Simonelli, *Transverse momentum dependent fragmentation functions from recent BELLE data*, *Phys. Rev. D* **106** (2022) 074024, [2206.08876].

[189] F. Hautmann, I. Scimemi and A. Vladimirov, *Non-perturbative contributions to vector-boson transverse momentum spectra in hadronic collisions*, *Phys. Lett. B* **806** (2020) 135478, [2002.12810].

[190] I. Borsa, R. Sassot, D. de Florian, M. Stratmann and W. Vogelsang, *Towards a Global QCD Analysis of Fragmentation Functions at Next-to-Next-to-Leading Order Accuracy*, *Phys. Rev. Lett.* **129** (2022) 012002, [2202.05060].

[191] MAP (MULTI-DIMENSIONAL ANALYSES OF PARTONIC DISTRIBUTIONS) collaboration, R. A. Khalek, V. Bertone and E. R. Nocera, *Determination of unpolarized pion fragmentation functions using semi-inclusive deep-inelastic-scattering data*, *Phys. Rev. D* **104** (2021) 034007, [2105.08725].

[192] A. S. Ito et al., *Measurement of the Continuum of Dimuons Produced in High-Energy Proton - Nucleus Collisions*, *Phys. Rev. D* **23** (1981) 604–633.

[193] G. Moreno et al., *Dimuon production in proton - copper collisions at $\sqrt{s} = 38.8\text{-}GeV$*, *Phys. Rev. D* **43** (1991) 2815–2836.

[194] E772 collaboration, P. L. McGaughey et al., *Cross-sections for the production of high mass muon pairs from 800-GeV proton bombardment of H-2*, *Phys. Rev. D* **50** (1994) 3038–3045.

[195] CDF collaboration, T. Affolder et al., *The transverse momentum and total cross section of e^+e^- pairs in the Z boson region from $p\bar{p}$ collisions at $\sqrt{s} = 1.8\ TeV$*, *Phys. Rev. Lett.* **84** (2000) 845–850, [hep-ex/0001021].

[196] CDF collaboration, T. Aaltonen et al., *Transverse momentum cross section of e^+e^- pairs in the Z-boson region from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. D **86** (2012) 052010, [`1207.7138`].

[197] D0 collaboration, B. Abbott et al., *Measurement of the inclusive differential cross section for Z bosons as a function of transverse momentum in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. D **61** (2000) 032004, [`hep-ex/9907009`].

[198] D0 collaboration, V. M. Abazov et al., *Measurement of the shape of the boson transverse momentum distribution in $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^- + X$ events produced at $\sqrt{s}$=1.96-TeV*, Phys. Rev. Lett. **100** (2008) 102002, [`0712.0803`].

[199] D0 collaboration, V. M. Abazov et al., *Measurement of the Normalized $Z/\gamma^* -> \mu^+\mu^-$ Transverse Momentum Distribution in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Lett. B **693** (2010) 522–530, [`1006.0618`].

[200] LHCʙ collaboration, R. Aaij et al., *Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 7$ TeV*, JHEP **08** (2015) 039, [`1505.07024`].

[201] LHCʙ collaboration, R. Aaij et al., *Measurement of forward W and Z boson production in pp collisions at $\sqrt{s} = 8$ TeV*, JHEP **01** (2016) 155, [`1511.08039`].

[202] CMS collaboration, S. Chatrchyan et al., *Measurement of the Rapidity and Transverse Momentum Distributions of Z Bosons in pp Collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. D **85** (2012) 032002, [`1110.4973`].

[203] CMS collaboration, V. Khachatryan et al., *Measurement of the transverse momentum spectra of weak vector bosons produced in proton-proton collisions at $\sqrt{s} = 8$ TeV*, JHEP **02** (2017) 096, [`1606.05864`].

[204] CMS collaboration, A. M. Sirunyan et al., *Measurements of differential Z boson production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **12** (2019) 061, [`1909.04133`].

[205] ATLAS collaboration, G. Aad et al., *Measurement of the transverse momentum and ϕ^*_η distributions of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **76** (2016) 291, [`1512.02192`].

[206] ATLAS collaboration, G. Aad et al., *Measurement of the transverse momentum distribution of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 616, [`1912.02844`].

[207] PHENIX collaboration, C. Aidala et al., *Measurements of $\mu\mu$ pairs from open heavy flavor and Drell-Yan in $p + p$ collisions at $\sqrt{s} = 200$ GeV*, Phys. Rev. D **99** (2019) 072003, [1805.02448].

[208] LHCʙ collaboration, R. Aaij et al., *Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 13$ TeV*, JHEP **09** (2016) 136, [1607.06495].

[209] ATLAS collaboration, G. Aad et al., *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, [1406.3660].

[210] 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, [1412.3989].

[211] M. Boglione, J. Collins, L. Gamberg, J. O. Gonzalez-Hernandez, T. C. Rogers and N. Sato, *Kinematics of Current Region Fragmentation in Semi-Inclusive Deeply Inelastic Scattering*, Phys. Lett. B **766** (2017) 245–253, [1611.10329].

[212] Jefferson Lab Angular Momentum (JAM) collaboration, M. Boglione, M. Diefenthaler, S. Dolan, L. Gamberg, W. Melnitchouk, D. Pitonyak et al., *New tool for kinematic regime estimation in semi-inclusive deep-inelastic scattering*, JHEP **04** (2022) 084, [2201.12197].

[213] K. B. Chen, J. P. Ma and X. B. Tong, *Matching of fracture functions for SIDIS in target fragmentation region*, JHEP **11** (2021) 038, [2108.13582].

[214] M. Diehl, *Generalized parton distributions*, Phys. Rept. **388** (2003) 41–277, [hep-ph/0307382].

[215] NNPDF collaboration, R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo et al., *Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties*, JHEP **05** (2010) 075, [0912.2276].

[216] G. D'Agostini, *On the use of the covariance matrix to fit correlated data*, Nuclear Instrum. and Met. **A346** (1994) 306–311.

[217] D. de Florian, R. Sassot, M. Epele, R. J. Hernández-Pinto and M. Stratmann, *Parton-to-Pion Fragmentation Reloaded*, Phys. Rev. D **91** (2015) 014035, [1410.6027].

[218] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk et al., *Uncertainties of predictions from parton distribution functions. 2. The Hessian method*, Phys. Rev. D **65** (2001) 014013, [hep-ph/0101032].

[219] D. de Florian, R. Sassot, M. Stratmann and W. Vogelsang, *Extraction of Spin-Dependent Parton Densities and Their Uncertainties*, *Phys. Rev. D* **80** (2009) 034030, [0904.3821].

[220] NNPDF collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017) 663, [1706.00428].

[221] D. de Florian, R. Sassot and M. Stratmann, *Global analysis of fragmentation functions for protons and charged hadrons*, *Phys. Rev. D* **76** (2007) 074033, [0707.1506].

[222] 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, [1807.03310].

[223] F. Landry, R. Brock, P. M. Nadolsky and C. P. Yuan, *Tevatron Run-1 Z boson data and Collins-Soper-Sterman resummation formalism*, *Phys. Rev. D* **67** (2003) 073016, [hep-ph/0212159].

[224] R.-b. Meng, F. I. Olness and D. E. Soper, *Semiinclusive deeply inelastic scattering at electron - proton colliders*, *Nucl. Phys. B* **371** (1992) 79–110.

[225] M. Burkardt and B. Pasquini, *Modelling the nucleon structure*, *Eur. Phys. J. A* **52** (2016) 161, [1510.02567].

[226] D. Müller and D. S. Hwang, *The concept of phenomenological light-front wave functions – Regge improved diquark model predictions*, [1407.1655].

[227] A. Bacchetta, F. Conti and M. Radici, *Transverse-momentum distributions in a diquark spectator model*, *Phys. Rev. D* **78** (2008) 074010, [0807.0323].

[228] M. Wakamatsu, *Transverse momentum distributions of quarks in the nucleon from the Chiral Quark Soliton Model*, *Phys. Rev. D* **79** (2009) 094028, [0903.1886].

[229] A. V. Efremov, P. Schweitzer, O. V. Teryaev and P. Zavada, *The relation between TMDs and PDFs in the covariant parton model approach*, *Phys. Rev. D* **83** (2011) 054025, [1012.5296].

[230] C. Bourrely, F. Buccella and J. Soffer, *Semiinclusive DIS cross sections and spin asymmetries in the quantum statistical parton distributions approach*, *Phys. Rev. D* **83** (2011) 074008, [1008.5322].

[231] H. H. Matevosyan, W. Bentz, I. C. Cloet and A. W. Thomas, *Transverse Momentum Dependent Fragmentation and Quark Distribution Functions from the NJL-jet Model, Phys. Rev. D* **85** (2012) 014021, [`1111.1740`].

[232] P. Schweitzer, M. Strikman and C. Weiss, *Intrinsic transverse momentum and parton correlations from dynamical chiral symmetry breaking, JHEP* **01** (2013) 163, [`1210.1267`].

[233] B. U. Musch, P. Hagler, J. W. Negele and A. Schafer, *Exploring quark transverse momentum distributions with lattice QCD, Phys. Rev. D* **83** (2011) 094507, [`1011.1213`].

[234] L. Cieri, G. Ferrera and G. F. R. Sborlini, *Combining QED and QCD transverse-momentum resummation for Z boson production at hadron colliders, JHEP* **08** (2018) 165, [`1805.11948`].

[235] B. Pasquini, S. Cazzaniga and S. Boffi, *Transverse momentum dependent parton distributions in a light-cone quark model, Phys. Rev. D* **78** (2008) 034025, [`0806.2298`].

[236] H. Avakian, A. V. Efremov, P. Schweitzer and F. Yuan, *The transverse momentum dependent distribution functions in the bag model, Phys. Rev. D* **81** (2010) 074035, [`1001.5467`].

[237] T. Gutsche, V. E. Lyubovitskij and I. Schmidt, *Nucleon parton distributions in a light-front quark model, Eur. Phys. J. C* **77** (2017) 86, [`1610.03526`].

[238] R. Alessandro, A. Del Dotto, E. Pace, G. Perna, G. Salmè and S. Scopetta, *Light-front transverse momentum distributions for J=1/2 hadronic systems in valence approximation, Phys. Rev. C* **104** (2021) 065204, [`2107.10187`].

[239] A. I. Signal and F. G. Cao, *Transverse momentum and transverse momentum distributions in the MIT bag model, Phys. Lett. B* **826** (2022) 136898, [`2108.12116`].

[240] A. Bacchetta, L. P. Gamberg, G. R. Goldstein and A. Mukherjee, *Collins fragmentation function for pions and kaons in a spectator model, Phys. Lett. B* **659** (2008) 234–243, [`0707.3372`].

[241] R. D. Ball et al., *Parton Distribution Benchmarking with LHC Data, JHEP* **04** (2013) 125, [`1211.5142`].

[242] L. Buonocore, S. Kallweit, L. Rottoli and M. Wiesemann, *Linear power corrections for two-body kinematics in the qT subtraction formalism, Phys. Lett. B* **829** (2022) 137118, [`2111.13661`].

186

[243] S. Camarda, L. Cieri and G. Ferrera, *Fiducial perturbative power corrections within the $\mathbf{q}_T$ subtraction formalism*, *Eur. Phys. J. C* **82** (2022) 575, [2111.14509].

[244] R. Abdul Khalek et al., *Snowmass 2021 White Paper: Electron Ion Collider for High Energy Physics*, in *2022 Snowmass Summer Study*, 3, 2022. [2203.13199].

[245] D. Boer, L. Gamberg, B. Musch and A. Prokudin, *Bessel-Weighted Asymmetries in Semi Inclusive Deep Inelastic Scattering*, *JHEP* **10** (2011) 021, [1107.5294].

[246] D. Boer, *Average transverse momentum quantities approaching the lightfront*, *Few Body Syst.* **56** (2015) 439–445, [1409.8317].

[247] A. A. Vladimirov, *Self-contained definition of the Collins-Soper kernel*, *Phys. Rev. Lett.* **125** (2020) 192002, [2003.02288].

[248] A. Bermudez Martinez and A. Vladimirov, *Determination of the Collins-Soper kernel from cross-sections ratios*, *Phys. Rev. D* **106** (2022) L091501, [2206.01105].

[249] M. Schlemmer, A. Vladimirov, C. Zimmermann, M. Engelhardt and A. Schäfer, *Determination of the Collins-Soper Kernel from Lattice QCD*, *JHEP* **08** (2021) 004, [2103.16991].

[250] P. Shanahan, M. Wagman and Y. Zhao, *Lattice QCD calculation of the Collins-Soper kernel from quasi-TMDPDFs*, *Phys. Rev. D* **104** (2021) 114502, [2107.11930].

[251] LPC collaboration, M.-H. Chu et al., *Nonperturbative determination of the Collins-Soper kernel from quasitransverse-momentum-dependent wave functions*, *Phys. Rev. D* **106** (2022) 034509, [2204.00200].

[252] LATTICE PARTON (LPC) collaboration, M.-H. Chu et al., *Lattice calculation of the intrinsic soft function and the Collins-Soper kernel*, *JHEP* **08** (2023) 172, [2306.06488].

[253] A. Avkhadiev, P. Shanahan, M. Wagman and Y. Zhao, *Collins-Soper kernel from lattice QCD at the physical pion mass*, 2307.12359.

[254] JEFFERSON LAB ANGULAR MOMENTUM collaboration, J. Cammarota, L. Gamberg, Z.-B. Kang, J. A. Miller, D. Pitonyak, A. Prokudin et al., *Origin of single transverse-spin asymmetries in high-energy collisions*, *Phys. Rev. D* **102** (2020) 054002, [2002.08384].

[255] I. P. Fernando and D. Keller, *Extraction of the Sivers function with deep neural networks*, *Phys. Rev. D* **108** (2023) 054007, [2304.14328].

[256] M. Gluck, P. Jimenez-Delgado and E. Reya, *Dynamical parton distributions of the nucleon and very small-x physics*, *Eur. Phys. J. C* **53** (2008) 355–366, [`0709.0614`].

[257] NNPDF collaboration, R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti et al., *Reweighting NNPDFs: the W lepton asymmetry*, *Nucl. Phys. B* **849** (2011) 112–143, [`1012.0836`].

[258] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti et al., *Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data*, *Nucl. Phys. B* **855** (2012) 608–638, [`1108.1758`].

[259] T. Horn and C. D. Roberts, *The pion: an enigma within the Standard Model*, *J. Phys. G* **43** (2016) 073001, [`1602.04016`].

[260] M. Gluck, E. Reya and A. Vogt, *Pionic parton distributions*, *Z. Phys. C* **53** (1992) 651–656.

[261] P. J. Sutton, A. D. Martin, R. G. Roberts and W. J. Stirling, *Parton distributions for the pion extracted from Drell-Yan and prompt photon experiments*, *Phys. Rev. D* **45** (1992) 2349–2359.

[262] M. Gluck, E. Reya and M. Stratmann, *Mesonic parton densities derived from constituent quark model constraints*, *Eur. Phys. J. C* **2** (1998) 159–163, [`hep-ph/9711369`].

[263] M. Gluck, E. Reya and I. Schienbein, *Pionic parton distributions revisited*, *Eur. Phys. J. C* **10** (1999) 313–317, [`hep-ph/9903288`].

[264] K. Wijesooriya, P. E. Reimer and R. J. Holt, *The pion parton distribution function in the valence region*, *Phys. Rev. C* **72** (2005) 065203, [`nucl-ex/0509012`].

[265] M. Aicher, A. Schafer and W. Vogelsang, *Soft-gluon resummation and the valence parton distribution function of the pion*, *Phys. Rev. Lett.* **105** (2010) 252003, [`1009.2481`].

[266] P. C. Barry, N. Sato, W. Melnitchouk and C.-R. Ji, *First Monte Carlo Global QCD Analysis of Pion Parton Distributions*, *Phys. Rev. Lett.* **121** (2018) 152001, [`1804.01965`].

[267] I. Novikov et al., *Parton Distribution Functions of the Charged Pion Within The xFitter Framework*, *Phys. Rev. D* **102** (2020) 014040, [`2002.02902`].

[268] Jefferson Lab Angular Momentum collaboration, N. Y. Cao, P. C. Barry, N. Sato and W. Melnitchouk, *Towards the*

three-dimensional parton structure of the pion: Integrating transverse momentum data into global QCD analysis, Phys. Rev. D **103** (2021) 114014, [2103.02159].

[269] Jefferson Lab Angular Momentum (JAM) collaboration, P. C. Barry, C.-R. Ji, N. Sato and W. Melnitchouk, *Global QCD Analysis of Pion Parton Distributions with Threshold Resummation*, Phys. Rev. Lett. **127** (2021) 232001, [2108.05822].

[270] C. Bourrely, W.-C. Chang and J.-C. Peng, *Pion Partonic Distributions in the Statistical Model from Pion-induced Drell-Yan and J/Ψ Production Data*, Phys. Rev. D **105** (2022) 076018, [2202.12547].

[271] B. Pasquini and P. Schweitzer, *Pion transverse momentum dependent parton distributions in a light-front constituent approach, and the Boer-Mulders effect in the pion-induced Drell-Yan process*, Phys. Rev. D **90** (2014) 014050, [1406.2056].

[272] S. Noguera and S. Scopetta, *Pion transverse momentum dependent parton distributions in the Nambu and Jona-Lasinio model*, JHEP **11** (2015) 102, [1508.01061].

[273] C. Lorcé, B. Pasquini and P. Schweitzer, *Transverse pion structure beyond leading twist in constituent models*, Eur. Phys. J. C **76** (2016) 415, [1605.00815].

[274] A. Bacchetta, S. Cotogno and B. Pasquini, *The transverse structure of the pion in momentum space inspired by the AdS/QCD correspondence*, Phys. Lett. B **771** (2017) 546–552, [1703.07669].

[275] F. A. Ceccopieri, A. Courtoy, S. Noguera and S. Scopetta, *Pion nucleus Drell–Yan process and parton transverse momentum in the pion*, Eur. Phys. J. C **78** (2018) 644, [1801.07682].

[276] M. Ahmady, C. Mondal and R. Sandapen, *Predicting the light-front holographic TMDs of the pion*, Phys. Rev. D **100** (2019) 054005, [1907.06561].

[277] N. Kaur and H. Dahiya, *Transverse momentum dependent parton distributions of pion in the light-front holographic model*, Int. J. Mod. Phys. A **36** (2021) 2150052, [1908.08657].

[278] C. Shi, K. Bednar, I. C. Cloët and A. Freese, *Spatial and Momentum Imaging of the Pion and Kaon*, Phys. Rev. D **101** (2020) 074014, [2003.03037].

[279] X. Wang, Z. Lu and I. Schmidt, *Transverse momentum spectrum of dilepton pair in the unpolarized $\pi^- N$ Drell-Yan process within TMD factorization*, JHEP **08** (2017) 137, [1707.05207].

[280] A. Vladimirov, *Pion-induced Drell-Yan processes within TMD factorization*, JHEP **10** (2019) 090, [`1907.10356`].

[281] W. J. Stirling and M. R. Whalley, *A Compilation of Drell-Yan cross-sections*, J. Phys. G **19** (1993) D1–D102.

[282] COMPASS collaboration, M. Aghasyan et al., *First measurement of transverse-spin-dependent azimuthal asymmetries in the Drell-Yan process*, Phys. Rev. Lett. **119** (2017) 112002, [`1704.00488`].

[283] Y.-H. Lien and A. Chumakov, *Measurement of Target Spin (in)dependent Asymmetries in Dimuon Production in Pion-Nucleon Collisions at COMPASS*, SciPost Phys. Proc. (2022) 028.

[284] Z. Kassabov, E. R. Nocera and M. Wilson, *Regularising experimental correlations in LHC data: theory and application to a global analysis of parton distributions*, Eur. Phys. J. C **82** (2022) 956, [`2207.00690`].

[285] S. Forte and Z. Kassabov, *Why α_s cannot be determined from hadronic processes without simultaneously determining the parton distributions*, Eur. Phys. J. C **80** (2020) 182, [`2001.04986`].

[286] J. C. Collins and J. Pumplin, *PDF uncertainties: A Strong test of goodness of fit to multiple data sets*, Sci. Cult. Ser. -Phys. **21** (2002) 401–404, [`hep-ph/0106173`].

[287] ATLAS collaboration, G. Aad et al., *A precise measurement of the Z-boson double-differential transverse momentum and rapidity distributions in the full phase space of the decay leptons with the ATLAS experiment at $\sqrt{s} = 8$ TeV*, [`2309.09318`].

[288] P. B. Arnold and L. D. McLerran, *Sphalerons, Small Fluctuations and Baryon Number Violation in Electroweak Theory*, Phys. Rev. D **36** (1987) 581.

[289] P. B. Arnold, D. Son and L. G. Yaffe, *The Hot baryon violation rate is O (alpha-w**5 T**4)*, Phys. Rev. D **55** (1997) 6264–6273, [`hep-ph/9609481`].

[290] T2K collaboration, K. Abe et al., *Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations*, Nature **580** (2020) 339–344, [`1910.03887`].

[291] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, Phys. Rept. **793** (2019) 1–98, [`1706.08945`].

[292] R. Boughezal, F. Petriello and D. Wiegand, *Disentangling Standard Model EFT operators with future low-energy parity-violating electron scattering experiments*, Phys. Rev. D **104** (2021) 016005, [`2104.03979`].

[293] USQCD collaboration, A. S. Kronfeld et al., *Lattice QCD and Particle Physics*, [2207.07641].

[294] R. Alarcon et al., *Electric dipole moments and the search for new physics*, in *Snowmass 2021*, 3, 2022. [2203.08103].

[295] N. Blinov, N. Craig, M. J. Dolan, J. de Vries, P. Draper, I. G. Garcia et al., *Snowmass White Paper: Strong CP Beyond Axion Direct Detection*, in *Snowmass 2021*, 3, 2022. [2203.07218].

[296] LHCʙ collaboration, R. Aaij et al., *Observation of CP Violation in Charm Decays*, Phys. Rev. Lett. **122** (2019) 211803, [1903.08726].

[297] P. Anderson, D. Higinbotham, S. Mantry and X. Zheng, *Complete Formalism of Cross Sections and Asymmetries for Longitudinally and Transversely Polarized Leptons and Hadrons in Deep Inelastic Scattering*, in *30th International Workshop on Deep-Inelastic Scattering and Related Subjects*, 5, 2023. [2306.00097].

[298] Pᴀʀᴛɪᴄʟᴇ Dᴀᴛᴀ Gʀᴏᴜᴘ collaboration, M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001.

[299] V. V. Flambaum, I. B. Khriplovich and O. P. Sushkov, *NUCLEAR ANAPOLE MOMENTS*, Phys. Lett. B **146** (1984) 367–369.

[300] C. S. Wood, S. C. Bennett, D. Cho, B. P. Masterson, J. L. Roberts, C. E. Tanner et al., *Measurement of parity nonconservation and an anapole moment in cesium*, Science **275** (1997) 1759–1763.

[301] S.-L. Zhu, S. J. Puglia, B. R. Holstein and M. J. Ramsey-Musolf, *The Nucleon anapole moment and parity violating e p scattering*, Phys. Rev. D **62** (2000) 033008, [hep-ph/0002252].

[302] R. D. Young, J. Roche, R. D. Carlini and A. W. Thomas, *Extracting nucleon strange and anapole form factors from world data*, Phys. Rev. Lett. **97** (2006) 102002, [nucl-ex/0604010].

[303] SAMPLE collaboration, R. Hasty et al., *Strange magnetism and the anapole structure of the proton*, Science **290** (2000) 2117, [nucl-ex/0102001].

[304] Qᴡᴇᴀᴋ collaboration, D. Androić et al., *Precision measurement of the weak charge of the proton*, Nature **557** (2018) 207–211, [1905.08283].

[305] R. Ruiz et al., *Target mass corrections in lepton-nucleus DIS: theory and applications to nuclear PDFs*, [2301.07715].

[306] H. Spiesberger, *QED radiative corrections for parton distributions*, Phys. Rev. D **52** (1995) 4936–4940, [hep-ph/9412286].

[307] NNPDF collaboration, R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti et al., *Parton distributions with QED corrections*, Nucl. Phys. B **877** (2013) 290–320, [1308.0598].

[308] A. Manohar, P. Nason, G. P. Salam and G. Zanderighi, *How bright is the proton? A precise determination of the photon parton distribution function*, Phys. Rev. Lett. **117** (2016) 242002, [1607.04266].

[309] H1 collaboration, V. Andreev et al., *Determination of electroweak parameters in polarised deep-inelastic scattering at HERA*, Eur. Phys. J. C **78** (2018) 777, [1806.01176].

[310] PVDIS collaboration, D. Wang et al., *Measurement of parity violation in electron–quark scattering*, Nature **506** (2014) 67–70.

[311] D. Wang et al., *Measurement of Parity-Violating Asymmetry in Electron-Deuteron Inelastic Scattering*, Phys. Rev. C **91** (2015) 045506, [1411.3200].

[312] C. Y. Prescott et al., *Further Measurements of Parity Nonconservation in Inelastic electron Scattering*, Phys. Lett. B **84** (1979) 524–528.

[313] J. Erler and S. Su, *The Weak Neutral Current*, Prog. Part. Nucl. Phys. **71** (2013) 119–149, [1303.5522].

[314] E. C. Aschenauer, T. Burton, T. Martini, H. Spiesberger and M. Stratmann, *Prospects for Charged Current Deep-Inelastic Scattering off Polarized Nucleons at a Future Electron-Ion Collider*, Phys. Rev. D **88** (2013) 114025, [1309.5327].

[315] NNPDF collaboration, E. R. Nocera, R. D. Ball, S. Forte, G. Ridolfi and J. Rojo, *A first unbiased global determination of polarized PDFs and their uncertainties*, Nucl. Phys. B **887** (2014) 276–308, [1406.5539].

[316] S. Melis, M. Boglione, J. Gonzalez Hernandez and A. Prokudin, *Theoretical Uncertainties and Dependence on the C_1, C_2, C_3 Parameters in the CSS Formalism in Drell-Yan and SIDIS*, PoS **QCDEV2015** (2015) 038.

[317] A. Vladimirov, *Kinematic power corrections in TMD factorization theorem*, JHEP **12** (2023) 008, [2307.13054].

[318] I. Moult, I. W. Stewart and G. Vita, *Subleading Power Factorization with Radiative Functions*, JHEP **11** (2019) 153, [1905.07411].

[319] L. Gamberg, Z.-B. Kang, D. Y. Shao, J. Terry and F. Zhao, *Transverse-momentum-dependent factorization at next-to-leading power*, [2211.13209].

[320] S. Rodini and A. Vladimirov, *Transverse momentum dependent factorization for SIDIS at next-to-leading power*, [2306.09495].

[321] S.-y. Wei, Y.-k. Song, K.-b. Chen and Z.-t. Liang, *Twist-4 contributions to semi-inclusive deeply inelastic scatterings with polarized beam and target*, Phys. Rev. D **95** (2017) 074017, [1611.08688].

[322] COMPASS collaboration, J. Matousek, *Azimuthal Asymmetries in Unpolarised Semi-Inclusive DIS at COMPASS*, JPS Conf. Proc. **37** (2022) 020104, [2201.05205].

[323] J. Dudek et al., *Physics Opportunities with the 12 GeV Upgrade at Jefferson Lab*, Eur. Phys. J. A **48** (2012) 187, [1208.1244].