

2021

## Proton Spin Structure from Simultaneous Monte Carlo Global QCD Analysis

Yiyu Zhou  
*William & Mary*

Follow this and additional works at: <https://scholarworks.wm.edu/etd>



Part of the [Quantum Physics Commons](#)

---

### Recommended Citation

Zhou, Yiyu, "Proton Spin Structure from Simultaneous Monte Carlo Global QCD Analysis" (2021).  
*Dissertations, Theses, and Masters Projects*. William & Mary. Paper 1627407586.  
<https://doi.org/10.21220/9K5N-4T96>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact [scholarworks@wm.edu](mailto:scholarworks@wm.edu).

Proton Spin Structure from Simultaneous Monte Carlo  
Global QCD Analysis

Yiyu Zhou

Xingping, Shaanxi, P. R. China

Master of Science, William & Mary, 2017  
Bachelor of Science, Xi'an Jiaotong University, 2015

A Dissertation presented to the Graduate Faculty of  
The College of William & Mary in Candidacy for the Degree of  
Doctor of Philosophy

Department of Physics

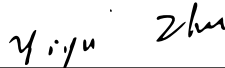
College of William & Mary  
July 2021

©2021  
Yiyu Zhou  
All rights reserved.

# APPROVAL PAGE

This Dissertation is submitted in partial fulfillment of  
the requirements for the degree of

Doctor of Philosophy



---

Yiyu Zhou

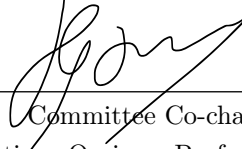
Approved by the Committee July 2021



---

Committee Co-chair

Wally Melnitchouk, Adjunct Professor, Physics  
College of William & Mary



---

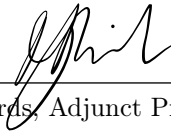
Committee Co-chair

Konstantinos Orginos, Professor, Physics  
College of William & Mary



---

Todd Averett, Professor, Physics  
College of William & Mary



---

David Richards, Adjunct Professor, Physics  
College of William & Mary



---

Chueng Ji, Professor, Physics  
North Carolina State University

## ABSTRACT

Despite the great effort and achievements made towards understanding proton spin structure in the past few decades, a complete picture is still elusive. Parton distribution functions (PDFs), which in quantum chromodynamics (QCD) encode the momentum and helicity distributions of quarks and gluons inside a proton, provide the means by which to quantify the proton structure information. Being inherently nonperturbative, PDFs have to be extracted from unpolarized and polarized lepton-hadron and hadron-hadron scattering data. In particular, experiments that measure unpolarized and polarized jet observables can provide insight into the momentum and helicity distributions of gluons, which have generally been more difficult to determine reliably than those of quarks.

In the past, extraction of the spin-averaged and spin-dependent (or helicity) PDFs has been performed in separate analyses. In this thesis, we perform the first simultaneous extraction of both types of quantities from deep-inelastic scattering (DIS), Drell-Yan and single jet observables, within the Monte Carlo global QCD analysis framework developed by the Jefferson Lab Angular Momentum (JAM) Collaboration. The results from this work indicate that the gluon helicity distributions depend rather strongly on the theory assumptions on which the global analysis is based, which calls for the need of measurements with higher precision.

As an application of the new simultaneous JAM analysis, we perform an impact study for future Electron-Ion Collider (EIC) data with parity-conserving and parity-violating polarization asymmetries on quark and gluon helicity distributions in the proton. The extrapolation of structure functions from the current data is studied for the first time in the context of the impact study. Theory assumptions, such as SU(2) and SU(3) flavor symmetries, are also studied to give a more thorough understanding of the impact of EIC pseudodata on proton spin structure.

# TABLE OF CONTENTS

Acknowledgments . . . . .	iv
Dedication . . . . .	v
List of Tables . . . . .	vi
List of Figures . . . . .	vii
CHAPTER . . . . .	1
1 Introduction . . . . .	2
1.1 Perturbative QCD and factorization . . . . .	3
1.2 History of proton spin structure . . . . .	6
1.3 Outline . . . . .	9
2 High energy scattering observables . . . . .	10
2.1 Deep-inelastic scattering . . . . .	10
2.1.1 Unpolarized DIS . . . . .	11
2.1.2 Polarized DIS . . . . .	14
2.1.3 EIC observables . . . . .	18
2.2 Drell-Yan process . . . . .	21
2.3 Jets in proton-proton collisions . . . . .	23
2.3.1 Jets in unpolarized collisions . . . . .	24
2.3.2 Jets in polarized collisions . . . . .	26
2.3.3 Jet algorithms . . . . .	28
2.4 Observables in Mellin space . . . . .	30
3 Parton distribution functions . . . . .	35

3.1	Parameterization	36
3.2	DGLAP evolution	40
3.3	Positivity constraints	41
4	Aspects of fitting	45
4.1	Bayesian approach	46
4.1.1	Likelihood function	47
4.1.2	Multi-step strategy	48
4.2	Monte Carlo technique	50
4.3	Optimization	51
4.4	Experimental data	52
4.4.1	Unpolarized datasets	52
4.4.2	Polarized datasets	54
4.4.3	EIC observables	55
5	Results from JAM global QCD analysis	62
5.1	Analysis of spin-averaged PDFs	63
5.2	Analysis of spin-dependent PDFs	74
5.2.1	Quality of fits	75
5.2.2	Spin-dependent PDFs and theory inputs	80
5.2.3	Fits with JAM17 $g_A$ and $a_8$	86
5.3	Analysis of jet observables	88
5.3.1	Renormalization and factorization scales	88
5.3.2	Cuts on $p_T$	90
5.4	Extraction of helicity basis PDFs	92
5.5	EIC impact study	97
5.5.1	Constraints from $A_{LL}$ pseudodata	98

5.5.2 Constraints from $A_{UL}$ pseudodata	103
6 Conclusion and outlook	106
6.1 Summary of results	107
6.2 Future of global QCD analysis	111
<b>APPENDIX A</b>	
Partonic jet tree level diagrams	113
Bibliography	116
Vita	126



## ACKNOWLEDGMENTS

First and foremost, I would like to thank my former advisor Joshua Erlich, who had been very patient with the steps I made, and very supportive regarding the course and research. When I found out that I was not suited for research in high energy theory, he also kindly helped me to find new directions. I will be always grateful for his advising and kindness.

Then, I would like to thank my advisor Wally Melnitchouk, for his fully hearted support, both academically and personally. As an advisor, he not only mentored me doing research, but also gave me confidence when I have doubts. The experience working with Wally is encouraging and insightful, which I will always value.

I would also like to thank Nobuo Sato for his patient tutorials and helpful suggestions during my time as a graduate student. What I can learn from him is beyond knowledge and I will always appreciate the opportunity working with him.

Thanks to all other members of the JAM Collaboration who have helped make this work possible, and to the members of my defense committee for their feedback and suggestions.

Thanks to my friends and family, especially my parents, for their support during my studies. Especially, my mother has always been the greatest support to me in the hardest times.

It has been three years since I saw my whole family last time, the town where I was born, raised and taught, has changed so much that many places in my childhood memories are no longer there. However, I still can not wait to see with my family and take a walk to the places where I used to play as a child. Nor could I wait to give a hug to our dog, who may have been wondering where I was all these years.

To my parents, Shangjun and Xia.

## LIST OF TABLES

3.1	Parameter ranges for the spin-averaged PDFs	38
3.2	Parameter ranges for the spin-dependent PDFs	40
4.1	$\chi^2$ for the EIC baseline fits	58
4.2	Estimation of uncorrelated systematic uncertainties for $A_{UL}$	61
5.1	Steps in JAM global QCD analysis for unpolarized fits	63
5.2	$\chi^2$ for JAM unpolarized fits	64
5.3	Scenarios for theory inputs in extracting the spin-dependent PDFs	74
5.4	$\chi^2$ for JAM polarized fits	76
5.5	$\chi^2$ for the two $\Delta g$ solutions with positivity constraints	83
5.6	Truncated moment $\int \Delta g$ integrated from 0.01	85
5.7	Truncated moment $\int \Delta g$ integrated from 0.05	85
5.8	$\chi^2$ for unpolarized jet observables with varying $\mu_{R,F}$	88
5.9	$\chi^2$ for polarized jet observables with varying $\mu_{R,F}$	89
5.10	$\chi^2$ for unpolarized jet observables with and without $p_T$ cut	91
5.11	Scenarios for the EIC impact study	98

## LIST OF FIGURES

2.1 DIS	11
2.2 Drell-Yan process	21
2.3 Collinear enhancement in jet production.	24
2.4 Hadronic jets	24
2.5 Mellin contour for the integration in complex plane	32
4.1 Multi-step strategy	49
4.2 Kinematics of unpolarized datasets	53
4.3 Kinematics of polarized datasets	55
4.4 Kinematics of EIC pseudodata	57
4.5 Projected $A_{LL}$ and $A_{UL}$ at EIC kinematics	60
5.1 Fit to proton $F_2$	65
5.2 Fit to deuteron $F_2$	66
5.3 Fit to reduced cross section at HERA with $\sqrt{s} = 318$ GeV	67
5.4 Fit to reduced cross section at HERA with other $\sqrt{s}$ values	68
5.5 Fit to DY data at E866/NuSea with $pp$ and $pd$ collisions	69
5.6 Fit to unpolarized jet observables from Tevatron and RHIC	70
5.7 Spin-averaged PDFs before and after inclusion of jets	71
5.8 Spin-averaged PDFs before and after inclusion of STAR jets	72
5.9 Comparison of spin-averaged PDFs with other groups	73
5.10 Fit to proton $A_1$ and $A_{  }$	77
5.11 Fit to deuteron $A_1$ and $A_{  }$	78
5.12 Fit to neutron $A_1^n$ and ${}^3\text{He}$ $A_1^h$ and $A_{  }^h$	79
5.13 Fit to polarized jet observables from STAR and PHENIX	80
5.14 Spin-dependent quark PDFs fitted with varying theory inputs	81
5.15 Spin-dependent gluon PDF fitted with varying theory inputs	82
5.16 Moments of spin-dependent PDFs with varying theory inputs	84
5.17 Spin-dependent PDFs with different SU(2/3) values	87
5.18 Spin-averaged PDFs with $A_1$ $\mu_{R,F}$	89

5.19 Spin-dependent PDFs with $A_1 \mu_{R,F}$	90
5.20 Spin-averaged PDFs with and without $p_T$ cut	91
5.21 Helicity basis from simultaneous fits	92
5.22 Demonstration of ROC and AUC	94
5.23 AUC plot for the helicity basis	95
5.24 Impact of EIC projected $A_{LL}^p$ data on $g_1^p$	99
5.25 Impact of EIC projected $A_{LL}^p$ and $A_{LL}^d$ data on $g_1^n$	100
5.26 Uncertainty ratios for truncated moments with and without EIC $A_{LL}^p$	101
5.27 Uncertainty ratios for truncated moments with and without EIC $A_{LL}^n$	102
5.28 Uncertainty ratios for truncated moments with and without EIC $A_{UL}^p$	104
A.1 Partonic jets ( $qq, qq'$ and $q\bar{q}'$ channels)	114
A.2 Partonic jets ( $q\bar{q}$ channel)	114
A.3 Partonic jets ( $qg$ channel)	115
A.4 Partonic jets ( $gg$ channel)	115

PROTON SPIN STRUCTURE FROM SIMULTANEOUS MONTE CARLO  
GLOBAL QCD ANALYSIS

# CHAPTER 1

## Introduction

Ever since the discovery of the nucleus by Rutherford, a great amount of effort has been put into understanding the basic constituents of our universe. Experiments quickly revealed that the nucleus consists of protons and neutrons, or *nucleons*, which led to the discovery of the strong force that binds the positively charged protons within the tiny space of a nucleus, overcoming the large repulsion caused by the electromagnetic force. Particle scattering experiments in the following decades with higher energies unraveled the substructure of nucleons, revealing that quarks, bound by gluons that mediate the strong force, form the protons and neutrons. Studies of the strong force and bound states of quarks followed, and this great endeavor led to quantum chromodynamics (QCD), a theory that can in principle describe the dynamics of quarks bound by gluons, and the formation of nuclei.

## 1.1 Perturbative QCD and factorization

QCD describes the dynamics of quarks and gluons (referred to as *partons*), however, the applicability of the description is limited by the energy scales, which leads to the topic of this section. First, we need to look at two basic concepts in QCD, asymptotic freedom and confinement. The quark-gluon coupling constant  $g_s$ , or, as more commonly used in QCD, the strong coupling  $\alpha_s \equiv g_s^2/(4\pi)$ , is the key to understanding the two concepts.  $\alpha_s$  depends on the number of active flavors  $N_f$ , and in its lowest order can be written as

$$\alpha_s \approx \frac{12\pi}{(11C_A - 4N_f T_R) \ln(\mu_R^2/\Lambda^2)}, \quad (1.1)$$

where  $C_A$  and  $T_R$  are SU(3) color factors [1]. Scale dependence arises from the renormalization scale  $\mu_R$ , which is associated with the regularization of ultraviolet divergences, and from  $\Lambda$ , which is a QCD constant of integration.

In particle scattering processes,  $\mu_R$  is usually set to be proportional to the momentum transfer,  $Q$ . In the region where  $\mu_R \sim Q \gg \Lambda$ , the strong coupling  $\alpha_s$  becomes small, and quarks can be treated as free particles inside the nucleon, which is known as asymptotic freedom. This forms the basis of perturbative QCD (pQCD), in which hard scattering processes can be described in terms of systematic expansions in powers of  $\alpha_s$ .

In the region where  $Q$  is comparable to  $\Lambda$ ,  $\alpha_s$  becomes large and perturbation theory is no longer applicable. The quarks are also confined by the large coupling, and this leads to the second concept which is confinement, the reason why quarks are not observed directly. This region is called the nonperturbative region, and has been studied by the community with great endeavor, including this work.



To access the inner structure of the nucleon, high energy lepton-nucleon and nucleon-nucleon collisions are measured at particle accelerator facilities. However, because detectors can only detect bound state hadrons, rather than quarks, inevitably the computation of cross section receives contributions from the nonperturbative (*i.e.* low- $Q$ ) part.

Factorization allows us to calculate cross sections by separating the contributions from perturbative and nonperturbative regions [2]. To illustrate the idea, we take deep-inelastic scattering (DIS) as an example. In DIS, a lepton is scattered from a nucleus and when the momentum transfer  $Q$  from the lepton to the nucleus is large, the cross section, at leading power in  $1/Q^2$ , can be approximated as a convolution of a hard scattering cross section,  $d\hat{\sigma}_{f_i}^{\text{DIS}}$ , and a soft, nonperturbative function,  $f_i$ ,

$$d\sigma^{\text{DIS}}(x, Q^2) = \sum_i \int_x^1 \frac{d\xi}{\xi} f_i\left(\frac{x}{\xi}, Q^2\right) d\hat{\sigma}_{f_i}^{\text{DIS}}(\xi, Q^2). \quad (1.2)$$

Here  $d\hat{\sigma}_{f_i}^{\text{DIS}}$  is interpreted as the cross section for the incoming lepton scattering off a parton of flavor  $i$  inside the nucleon, *i.e.*, the partonic cross section, and can be calculated perturbatively to a given order in  $\alpha_s$ . The nonperturbative function  $f_i(x)$  is interpreted as the probability of parton  $i$  carrying a fraction  $x$  of the nucleon's longitudinal momentum, and is often referred to as the parton distribution function (PDF). The total cross section is then obtained by integrating over all momentum fractions and summing over all flavors.

Cross sections for other high energy processes, such as Drell-Yan (production of lepton pairs from the annihilation of quarks and anti-quarks from two colliding hadrons, abbreviated as ‘‘DY’’ below) and single jet (collimated high-energy particles produced from proton-proton

or proton-antiproton collisions) production, can be factorized in a similar manner,

$$\begin{aligned} d\sigma^{\text{DY}}(x_F, Q^2) &= \sum_{i,j} \int dx_i \int dx_j f_i(x_i, Q^2) f_j(x_j, Q^2) d\hat{\sigma}_{ij \rightarrow \bar{l}l}^{\text{DY}}(x_i, x_j, Q^2, \dots), \\ d\sigma^{\text{jets}}(p_T, \eta) &= \sum_{i,j} \int dx_i \int dx_j f_i(x_i, Q^2) f_j(x_j, Q^2) d\hat{\sigma}_{ij \rightarrow \text{jet}X}^{\text{jets}}(x_i, x_j, p_T, \dots), \end{aligned} \quad (1.3)$$

where  $x_i$  and  $x_j$  are the momentum fractions carried by partons of flavor  $i$  and  $j$  in the two colliding hadrons,  $\eta$  and  $p_T$  are the rapidity and transverse momentum of the detected jet, respectively, and  $d\hat{\sigma}_{ij \rightarrow \bar{l}l}^{\text{DY}}$  and  $d\hat{\sigma}_{ij \rightarrow \text{jet}X}^{\text{jets}}$  are the partonic cross section for the Drell-Yan and single jet production processes. Although the partonic cross sections  $d\hat{\sigma}$  are different for different processes, the PDFs  $f_i(x, Q^2)$  are universal, which enables a global analysis of world data to be performed in order to extract a universal solution for the PDF.

Factorization and universality are very useful in extracting the nonperturbative PDFs from experimental data, and significant progress has been made during the past few years [3–6]. Apart from momentum distributions, efforts have been made to determine the spin structure of the proton by extracting spin-dependent PDFs ( $\Delta f$ ) [7, 8]. Importantly, however, momentum distributions (*i.e.* spin-averaged PDFs) and spin-dependent distributions have not yet been studied simultaneously, despite the fact that they are different linear combinations of same quantities, with  $f = f^\uparrow + f^\downarrow$  and  $\Delta f = f^\uparrow - f^\downarrow$ , where  $f^\uparrow$  ( $f^\downarrow$ ) is the parton distributions with spin aligned (anti-aligned) with the proton spin. In this thesis, a *simultaneous* global analysis of spin-averaged and spin-dependent PDFs is presented for the first time.

It is worth mentioning that, although this thesis focuses on the extraction of the proton PDFs, the neutron PDFs can be obtained via isospin symmetry, *i.e.*,  $u^p = d^n$ ,  $\bar{u}^p = \bar{d}^n$ , where the superscripts  $p$  and  $n$  denote the proton and neutron, respectively.

## 1.2 History of proton spin structure

The first precision measurement of proton spin structure performed by the European Muon Collaboration (EMC) [9] with polarized DIS found that the net contribution from quarks to the proton spin at the scale  $Q_{\text{EMC}}^2 = 10.7 \text{ GeV}^2$  was  $\Delta\Sigma(Q_{\text{EMC}}^2) \sim 0.1$ , where  $\Delta\Sigma$  is the first moment of quark helicity distributions. The result was very different from the prediction made by Ellis and Jaffe [10], which is  $\Delta\Sigma(Q_{\text{EMC}}^2) \sim 0.6$ , by assuming zero contribution from strange quark polarization. The question then remains: what is carrying the rest of the proton spin? The discovery that quarks carry such a small fraction of the proton spin was originally termed the “proton spin crisis”, and more recently has been known as the “proton spin puzzle”.

In attempting to explain the EMC result, Hatsuda and Zahed [11] found that with a consistent treatment of the U(1) axial anomaly, the quark spin  $\Delta\Sigma$  observes a large cancellation, and thus gives a better agreement with the aforementioned EMC measurement. Altarelli and Ross [12] attributed the cancellation to the gluon anomaly and argued that the EMC measured quantity should be interpreted as  $\Delta\Sigma - \alpha_s \Delta G / (2\pi)$ , with  $\Delta G$  the first moment of the gluon helicity distribution. Fritzsche [13] later interpreted the small value of the quark spin as a gluon helicity term by considering the nucleon matrix element of the axial baryonic current in the chiral limit. Many other theoretical interpretations were proposed, summaries of which can be found in the review articles in Refs. [14–16].

Following the EMC measurement, the Spin Muon Collaboration (SMC) extracted for the first time the  $x$ -dependent spin distributions for up and down valence quarks, as well as the nonstrange sea quarks using semi-inclusive DIS (SIDIS) measurements [17, 18]. Later, the high-precision SIDIS measurements at HERMES [19] and COMPASS [20, 21] for  $\pi^\pm$  and  $K^\pm$  production from deuteron and/or hydrogen targets, together with single-inclusive

$e^+e^-$  annihilation (SIA) and polarized DIS data, were used in the JAM17 analysis [8] for the first simultaneous determination of the fragmentation functions and spin-dependent PDFs. In particular, the strange spin distribution,  $\Delta s^+ = \Delta s + \Delta \bar{s}$ , was found to be consistent with zero, and thus could not account for the total proton spin budget.

Knowing that the quark spin is not the whole story of the proton spin, measurements of observables that are sensitive to gluon spin then followed. One such observable is jet production in polarized proton-proton or proton-antiproton collisions. Such measurements were performed at the Relativistic Heavy Ion Collider (RHIC) by the STAR [22–26] and PHENIX [27] collaborations. They are expected to be an ideal source of information on the gluon spin contribution  $\Delta G$  because the gluon PDF enters at lowest order in the jet production cross section. However, to date no compelling evidence has been found in global QCD analyses to suggest that gluons make up the remainder of the proton spin after the quark contribution is accounted for [7, 28].

What is missing from the above discussion, on the other hand, is the orbital angular momentum of quarks and gluons. Although other sum rules like Ji’s sum rule [29] are available, according to Jaffe and Manohar [30], the proton spin can be decomposed as

$$\frac{1}{2} = \frac{1}{2}\Delta\Sigma(\mu) + \Delta G(\mu) + L_{Q+G}(\mu), \quad (1.4)$$

where  $\Delta\Sigma$  and  $\Delta G$  are quark + antiquark helicity and gluon helicity, respectively, and  $L_{Q+G}$  is the parton orbital angular momentum. Note that all terms depend on the renormalization scale,  $\mu$ . Interest in the proton spin has therefore expanded to include the orbital angular momentum of quarks and gluons that may make up a portion of the proton spin. The orbital angular momentum is encoded in the transverse momentum dependent (TMD)

distributions which, as the name suggests, depend on the quark transverse momentum  $k_T$ , as well as in the generalized parton distributions (GPDs) that depend on the transverse spatial distance  $\mathbf{b}$ . Since the extraction of TMDs and GPDs from experimental data in global QCD analyses is still in its infancy, information on these distributions is usually supplemented by calculations from lattice QCD [31, 32].

Presently, as the detailed physics plans for the Electron-Ion Collider (EIC) are being assessed [33], more opportunities are being opened up for studying the proton spin decomposition. With its high luminosity, high beam polarization, and unprecedented kinematic coverage, the EIC can not only access more observables, but also with smaller statistical uncertainties. This will allow more precise determination of helicity PDFs, as well as TMD and GPDs. Combining with the collinear helicity PDFs, the transverse momentum and spatial distributions will help us complete the picture of the proton spin decomposition.

The focus of this thesis will be on the collinear distribution functions, namely the spin-averaged and spin-dependent PDFs. Accessing such distributions requires global analysis of world data, such as DIS, Drell-Yan and jet production, and in the case of  $\Delta f$ , polarized DIS and jets are needed. In particular, we will be interested in extracting the gluon helicity PDF  $\Delta g$  from jet production data in polarized hadron-hadron collisions. Although the impact of polarized jet data on  $\Delta g$  has been studied previously by several groups, such as DSSV [7, 28], a number of theoretical assumptions have been made in these analyses, which could lead to biased results and need to be revisited. In particular, in this work various theory assumptions, such as SU(2) and SU(3) flavor symmetry, positivity constraints (requiring that  $|\Delta f| \leq |f|$  for each flavor) are tested and their biases are studied. Moreover, we also present in this work the first simultaneous extraction of spin-averaged and spin-dependent PDFs. With its consistent treatment of both types of PDFs, we will argue that a simultaneous analysis represents a new paradigm that will define the

future of PDFs extraction.

### 1.3 Outline

The organization of the rest of this dissertation is as follows: in Chapter [2](#), we will describe the theory of the observables used in this work, including unpolarized and polarized DIS, parity-conserving and parity-violating polarization asymmetries from the EIC, unpolarized Drell-Yan, and jet production in unpolarized and polarized hadron-hadron collisions. Then, in Chapter [3](#), basic concepts of parton distribution functions are elaborated. For example, the Mellin transform is applied to PDFs in order to make the PDF scale evolution more efficient. PDF sum rules are also introduced in this chapter, which offer additional theory constraints on the PDF determination. In Chapter [4](#), the fitting process is outlined, including Bayesian inference, Monte Carlo sampling, as well as the treatment of experimental data sets. We also describe various ways employed to increase the speed of the fits. The results of the fits are presented in Chapter [5](#), which is divided into five parts. The first part shows the result of the analysis of the spin-averaged PDFs, while the second part focuses on the spin-dependent PDFs and their dependence upon theory assumptions, especially for  $\Delta g$ . After a brief complementary study on jet observables in the third part, the fourth part shows the results from the simultaneous analysis of spin-averaged and spin-dependent PDFs. The last part uses the PDFs extracted from the existing data to perform an impact study of future parity-conserving and parity-violating polarization asymmetries planned to be measured at the EIC. Chapter [6](#) summarizes the results and discusses future extensions of the PDF analysis reported in this thesis.

# CHAPTER 2

## High energy scattering observables

The theoretical basis of global QCD analysis is collinear factorization, whose application to DIS, Drell-Yan and jet observables will be discussed in this chapter. We will also describe the Mellin transformation and the practical benefit of utilizing this method in global QCD analysis.

### 2.1 Deep-inelastic scattering

Deep-inelastic scattering is the scattering of a lepton from a hadron, with the hadron being a fixed target or a beam colliding with the lepton. This section is divided into two parts, the first part is focused on the unpolarized DIS observables, while the second part will extend the topic to polarized DIS observables.

### 2.1.1 Unpolarized DIS

Consider a high-energy lepton scattering from a hadron, with the momenta of incoming and outgoing leptons labeled by  $k^\mu$  and  $k'^\mu$ , respectively, the momentum of the struck hadron (assumed hereafter to be a proton) labeled by  $P^\mu$ , and the momentum transfer given by  $q^\mu \equiv k^\mu - k'^\mu$ , as shown in Fig. [2.1](#).

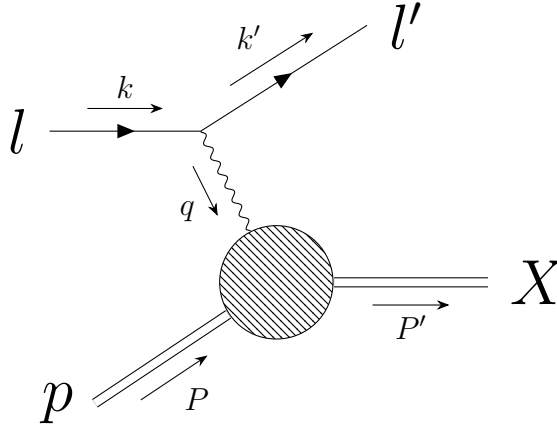


FIG. 2.1: Deep-inelastic charged lepton-hadron scattering with a single photon exchange, with  $p$  denoting the proton, and  $l$  and  $X$  representing the lepton and undetected particles, respectively. The momentum  $q$  is spacelike.

The differential cross section can be written as a contraction of a symmetric (S) leptonic tensor  $L_{\mu\nu}^S$  and a symmetric hadronic tensor  $W_S^{\mu\nu}$ ,

$$\frac{d^2\sigma}{dx dy} = \frac{2\pi y\alpha^2}{Q^4} L_{\mu\nu}^S W_S^{\mu\nu}, \quad (2.1)$$

where  $\alpha = e^2/(4\pi)$  is the electromagnetic fine structure constant, the variables  $x$  and  $y$  correspond to the Bjorken scaling variable and inelasticity, respectively, defined as

$$x \equiv \frac{Q^2}{2P \cdot q} = \frac{Q^2}{2M\nu}, \quad y \equiv \frac{P \cdot q}{P \cdot k} = \frac{\nu}{E}, \quad (2.2)$$



with  $Q^2 \equiv -q^2$  the four-momentum transfer squared and  $\nu \equiv P \cdot q/M = E - E'$  the energy transfer, using the momenta defined in the proton rest frame

$$\begin{aligned}
P^\mu &= (M, 0, 0, 0), \\
k^\mu &= (E, 0, 0, E), \\
k'^\mu &= (E', E' \sin \theta \cos \phi, E' \sin \theta \sin \phi, E' \cos \theta),
\end{aligned} \tag{2.3}$$

where  $M$  is the mass of the proton, and  $E$  and  $E'$  are the energies of the incoming and outgoing leptons, respectively, which are taken to be massless. For the process in Fig. [2.1](#), the symmetric leptonic tensor with single photon exchange can be determined completely within QED

$$L_{\mu\nu}^S = 2(k_\mu k'_\nu + k_\nu k'_\mu - g_{\mu\nu} k \cdot k'). \tag{2.4}$$

The hadronic tensor that describes the interaction between the photon and the nucleon cannot at present be computed from first principles. However, from electromagnetic current conservation  $q_\mu W_S^{\mu\nu} = 0$  and parity conservation, we can write

$$W_S^{\mu\nu} = \left( \frac{q^\mu q^\nu}{q^2} - g^{\mu\nu} \right) F_1(x, Q^2) + \left( P^\mu + \frac{q^\mu}{2x} \right) \left( P^\nu + \frac{q^\nu}{2x} \right) \frac{F_2(x, Q^2)}{P \cdot q}, \tag{2.5}$$

where  $F_1(x, Q^2)$  and  $F_2(x, Q^2)$  are the structure functions that characterize the structure of hadrons as “seen” by the virtual photon. The structure functions can be expressed by a

series expansion in powers of  $1/Q^2$ . For  $F_1$ , we have

$$\begin{aligned} F_1(x, Q^2) &= \frac{1}{2} \sum_q e_q^2 ((C_{1,q}^{\text{DIS}} \otimes q^+)(x, Q^2) + (C_{1,g}^{\text{DIS}} \otimes g)(x, Q^2)) + \mathcal{O}\left(\frac{1}{Q^2}\right), \\ F_2(x, Q^2) &= x \sum_q e_q^2 ((C_{2,q}^{\text{DIS}} \otimes q^+)(x, Q^2) + (C_{2,g}^{\text{DIS}} \otimes g)(x, Q^2)) + \mathcal{O}\left(\frac{1}{Q^2}\right), \end{aligned} \quad (2.6)$$

where  $e_q$  is the electric charge of quark with flavor  $q$ ,  $C_{1,q}^{\text{DIS}}$  and  $C_{1,g}^{\text{DIS}}$  ( $C_{2,q}^{\text{DIS}}$  and  $C_{2,g}^{\text{DIS}}$ ) are the hard scattering coefficients for  $F_1$  ( $F_2$ ) of DIS, and  $q^+ = q + \bar{q}$ . Both the hard scattering coefficients and the PDFs depend on a renormalization scale  $\mu_R$  that regulates ultraviolet divergences. The scale is usually set to the hard scale as in Eq. (2.6), but can be varied to estimate the theoretical uncertainty.

The symbol  $\otimes$  in Eq. (2.6) represents the convolution integral

$$[C \otimes f](x, Q^2) = \int_x^1 C(\xi, Q^2) f\left(\frac{x}{\xi}, Q^2\right) \frac{d\xi}{\xi}. \quad (2.7)$$

Notice that Eq. (2.6) is true to all orders of perturbative QCD, because the expansion in orders of  $\alpha_S$  is implicit for the hard scattering coefficients,

$$C_{1,q}^{\text{DIS}}(x, Q^2) = C_{1,q}^{\text{DIS},(0)}(x, Q^2) + \frac{\alpha_S(Q^2)}{4\pi} C_{1,q}^{\text{DIS},(1)}(x, Q^2) + \mathcal{O}(\alpha_S^2), \quad (2.8)$$

and similarly for  $C_{1,g}^{\text{DIS}}$ ,  $C_{2,q}^{\text{DIS}}$  and  $C_{2,g}^{\text{DIS}}$ .

At leading order, the structure functions are related to the PDFs by [\[34\]](#)

$$F_2(x) = 2xF_1(x) = x \left( \frac{4u^+(x)}{9} + \frac{d^+(x) + s^+(x)}{9} \right), \quad (2.9)$$

when considering only the three light flavors. The gluon sensitivity, however, appears only at higher orders, which is therefore suppressed by the strong coupling constant.

For the unpolarized DIS process, we include proton and/or deuteron  $F_2$  measurements from BCDMS [\[35\]](#), SLAC [\[36\]](#), and NMC [\[37\]](#), as well as  $F_2^d/F_2^p$  measurements from NMC [\[38\]](#), with cuts  $W^2 = M^2 + Q^2(1-x)/x > 10 \text{ GeV}^2$  and  $Q^2 > m_c^2$ , where  $m_c \approx 1.27 \text{ GeV}$  is the mass of charm quark. With the same cuts, we also include the reduced proton neutral current and charged current cross sections from the combined H1 and ZEUS analysis from HERA [\[39\]](#).

### 2.1.2 Polarized DIS

For polarized DIS interactions, the differential cross sections can be written in a similar manner as a contraction of an antisymmetric (A) leptonic tensor  $L_{\mu\nu}^A$  and an antisymmetric hadronic tensor  $W_A^{\mu\nu}$

$$\frac{d^2\Delta\sigma}{dx dy} = \frac{2\pi y\alpha^2}{Q^4} L_{\mu\nu}^A W_A^{\mu\nu}, \quad (2.10)$$

with kinematic variables defined as in Section [2.1.1](#). The antisymmetric leptonic tensor in

this case is given by

$$L_{\mu\nu}^A = -2i\lambda\epsilon_{\mu\nu\alpha\beta}k^\alpha k'^\beta, \quad (2.11)$$

where  $\epsilon_{\mu\nu\alpha\beta}$  is the antisymmetric Levi-Civita tensor, and  $\lambda = \pm 1$  represents the helicity of the incoming lepton.

The antisymmetric hadronic tensor  $W_A^{\mu\nu}$  can be written in terms of the polarized structures represented by  $g_1(x, Q^2)$  and  $g_2(x, Q^2)$

$$W_A^{\mu\nu} = i\epsilon^{\mu\nu\alpha\beta}\frac{q_\alpha S_\beta}{P \cdot q}g_1(x, Q^2) + i\epsilon^{\mu\nu\alpha\beta}\frac{q_\alpha}{P \cdot q}\left(S_\beta - P_\beta\frac{S \cdot q}{P \cdot q}\right)g_2(x, Q^2), \quad (2.12)$$

where  $S_\beta$  is the nucleon spin four-vector, and we follow the convention that  $S^2 = -M^2$  and  $S \cdot P = 0$ , and for the polarized structure functions, we have

$$g_1(x, Q^2) = \sum_q \frac{e_q^2}{2}((\Delta C_{1,q}^{\text{DIS}} \otimes \Delta q^+)(x, Q^2) + 2(\Delta C_{1,g}^{\text{DIS}} \otimes \Delta g)(x, Q^2)) + \mathcal{O}\left(\frac{1}{Q}\right), \quad (2.13)$$

where  $\Delta C_{1,q}^{\text{DIS}}$  and  $\Delta C_{1,g}^{\text{DIS}}$  are the hard scattering coefficients for polarized DIS, and  $\Delta q^+ = \Delta q + \Delta \bar{q}$ . Notice that Eq. (2.13) is true to all orders of perturbative QCD as well, because the expansion of  $\alpha_S$  is implicit in the hard scattering coefficients

$$\Delta C_{1,q}^{\text{DIS}}(x, Q^2) = \Delta C_{1,q}^{\text{DIS},(0)}(x, Q^2) + \frac{\alpha_S(Q^2)}{4\pi}\Delta C_{1,q}^{\text{DIS},(1)}(x, Q^2) + \mathcal{O}(\alpha_S^2), \quad (2.14)$$

and similar for  $\Delta C_{1,g}^{\text{DIS}}$ . At leading order, the structure functions are simplified to

$$\begin{aligned} g_1(x, Q^2) &= \frac{1}{2} \left( \frac{4\Delta u^+(x)}{9} + \frac{\Delta d^+(x) + \Delta s^+(x)}{9} \right), \\ g_2(x, Q^2) &= -g_1(x, Q^2) + \int_x^1 g_1(z, Q^2) \frac{1}{z} dz + \mathcal{O}\left(\frac{1}{Q}\right), \end{aligned} \quad (2.15)$$

where  $g_2$  is given by the Wandzura-Wilczek relation [40]. Again, there is no sensitivity to the polarized gluon distribution  $\Delta g$  at leading order from polarized DIS.

In experiments with polarized beams or targets, the quantity that is usually measured is the polarization asymmetry, which is a ratio of spin-dependent cross sections to spin-averaged cross sections. For the case of polarized DIS, the asymmetry can be expressed in terms of the spherical angles  $\theta$  and  $\phi$  that describe the target polarization relative to the virtual photon three momentum  $\vec{q}$ ,

$$A = \frac{\sigma^\uparrow - \sigma^\downarrow}{\sigma^\uparrow + \sigma^\downarrow} = \frac{\sqrt{1 - \epsilon^2} A_1 \cos \theta + \sqrt{2\epsilon(1 - \epsilon)} A_2 \sin \theta \cos \phi}{1 + \epsilon R}, \quad (2.16)$$

where  $\sigma^{\uparrow\downarrow} \equiv d^2\sigma^{\uparrow\downarrow}/dx dy$  are the cross sections in Eq. (2.10) with lepton helicity  $\lambda(\uparrow) = +1$  or  $\lambda(\downarrow) = -1$ ,  $\sigma^\uparrow - \sigma^\downarrow = d^2\Delta\sigma/(dx dy)$  in (2.10) and  $\sigma^\uparrow + \sigma^\downarrow = d^2\sigma/(dx dy)$  in (2.1). The variables  $\epsilon$  and  $R$  are defined as

$$\epsilon = \frac{2(1 - y) - \frac{1}{2}\gamma^2 y^2}{1 + (1 - y)^2 + \frac{1}{2}\gamma^2 y^2}, \quad R = \frac{(1 + \gamma^2)F_2 - 2xF_1}{2xF_1}, \quad (2.17)$$

where  $\gamma^2 = 4M^2 x^2/Q^2$ , which is small in Bjorken limit where  $Q^2 \rightarrow \infty$  with  $x$  fixed. The variable  $R$  is the ratio of longitudinal to transverse photon absorption cross sections and  $\epsilon$  is the ratio of longitudinal to transverse virtual photon polarizations.

The polarization of the proton target can be adjusted to be parallel ( $\uparrow$ ) or perpendicular ( $\Rightarrow$ ) to the lepton beam, so that the asymmetry in Eq. (2.16) can be separated into longitudinal and transverse spin asymmetries,

$$\begin{aligned} A_{\parallel} &= \frac{\sigma^{\downarrow\uparrow} - \sigma^{\uparrow\uparrow}}{\sigma^{\downarrow\uparrow} + \sigma^{\uparrow\uparrow}} = D(A_1 + \eta A_2), \\ A_{\perp} &= \frac{\sigma^{\downarrow\Rightarrow} - \sigma^{\uparrow\Rightarrow}}{\sigma^{\downarrow\Rightarrow} + \sigma^{\uparrow\Rightarrow}} = d(A_2 - \zeta A_1), \end{aligned} \quad (2.18)$$

where the photon polarization factors  $D$ ,  $d$ ,  $\eta$  and  $\zeta$  are

$$\begin{aligned} D &= \frac{y(2-y)(2+\gamma^2 y)}{2(1+\gamma^2)y^2 + (4(1-y) - \gamma^2 y^2)(1+R)}, \\ d &= \frac{\sqrt{4(1-y) - \gamma^2 y^2}}{2-y} D, \\ \eta &= \gamma \frac{4(1-y) - \gamma^2 y^2}{(2-y)(2+\gamma^2 y)}, \quad \zeta = \gamma \frac{2-y}{2+\gamma^2 y}, \end{aligned} \quad (2.19)$$

and the virtual photoproduction asymmetries  $A_1$  and  $A_2$  that appear in Eqs. (2.16) and (2.18) can be expressed by spin-averaged and spin-dependent structure functions

$$A_1 = \frac{g_1 - \gamma^2 g_2}{F_1}, \quad A_2 = \gamma \frac{g_1 + g_2}{F_1}. \quad (2.20)$$

Due to the suppression of  $\gamma$  when  $Q^2 \gg M^2$ , we will only fit  $A_1$  or  $A_{\parallel}$  measurements. Specifically, we include fixed target polarized DIS proton, deuteron and helium  $A_1$  and/or  $A_{\parallel}$  measurements from EMC [41], SMC [42, 43], COMPASS [44–46], SLAC [47–52], and HERMES [53, 54], with the same cuts on  $Q^2$  and  $W^2$  as for unpolarized DIS data sets, as in Refs. [8, 55]. Details of the observables from different collaborations are listed in Table 5.4.

### 2.1.3 EIC observables

As the world's first polarized lepton-hadron (and lepton-nucleus) collider, the Electron-Ion Collider (EIC) will explore uncharted territory in spin physics [56], extending the kinematic coverage in both Bjorken- $x$  and momentum exchange  $Q^2$ . While the kinematic coverage will be discussed in detail later in Section 4.4.3, this section will focus on the observables from the impact study [57], namely the parity-conserving ( $A_{LL}$ ) and parity-violating ( $A_{UL}$ ) polarization asymmetries can be studied at the EIC.

The parity-conserving polarization asymmetry is defined in Eq. (2.18) as  $A_{\parallel}$ , and for EIC impact study, it is named after the polarization directions as the double longitudinal spin asymmetry  $A_{LL}$ .

At typical EIC kinematics, one can take  $M^2 \ll Q^2$ , in which case  $\gamma^2 = 4M^2x^2/Q^2 \rightarrow 0$  and  $A_{LL}$  simplifies to

$$A_{LL} = \frac{y(2-y)}{y^2 + 2(1-y)(1+R)} \frac{g_1}{F_1}, \quad (2.21)$$

which is proportional to  $g_1$ . According to Eq. (2.15), the leading order contribution from individual quark flavors  $\Delta q^+ = \Delta q + \Delta \bar{q}$  is proportional to the square of their charges, so that proton measurements mostly determine the  $\Delta u^+$  distribution. In order to constrain the  $\Delta d^+$  flavor one needs other hadrons, such as deuterons or  $^3\text{He}$  nuclei, which are more sensitive to the down quark helicity distribution via the neutron structure function  $g_1^n$ , which at leading order is given by

$$g_1^n \approx \frac{4\Delta d^+ + \Delta u^+ + \Delta s^+}{18}. \quad (2.22)$$

Another novel observable that can be studied at the EIC is the parity-violating asymmetry that involves the scattering of unpolarized leptons from longitudinally polarized hadrons,

$$A_{UL} = \frac{\sigma^{\uparrow} - \sigma^{\downarrow}}{\sigma^{\uparrow} + \sigma^{\downarrow}}, \quad (2.23)$$

where  $\uparrow$  ( $\downarrow$ ) denotes the spin of the hadron along (opposite to) the beam direction. In this asymmetry the parity-conserving contributions from photon exchange and the vector-vector part of  $Z$ -boson exchange cancel exactly in the numerator, leaving the dominant contribution from the interference of photon and the axial-vector part of  $Z$ -boson exchange [58]. Neglecting the diagonal  $Z$ -exchange contributions and taking  $M^2 \ll Q^2$ , the parity-violating asymmetry can be written in terms of the spin-dependent interference  $\gamma Z$  structure functions  $g_{1,5}^{\gamma Z}$  [59],

$$A_{UL} = \frac{G_F x Q^2}{2\sqrt{2}\pi\alpha} \left( \frac{g_A^e Y^- g_1^{\gamma Z} + g_V^e Y^+ g_5^{\gamma Z}}{xy^2 F_1 + (1-y) F_2} \right), \quad (2.24)$$

where  $G_F$  is the Fermi constant,  $\alpha$  is the fine structure constant,  $g_V^e = -\frac{1}{2} + 2\sin^2\theta_W$  and  $g_A^e = -\frac{1}{2}$  are the vector and axial-vector couplings of the electron to the  $Z$  boson, and  $\theta_W$  is the weak mixing angle. The kinematic factors in the numerator of (2.24) are given by  $Y^\pm = 1 \pm (1-y)^2$ .

At leading twist, the polarized  $\gamma Z$  interference structure functions can be written in terms



of the helicity PDFs as [59],

$$\begin{aligned}
g_1^{\gamma Z}(x, Q^2) &= \sum_q e_q g_V^q ((\Delta C_{1,q} \otimes \Delta q^+)(x, Q^2) + 2(\Delta C_{1,q} \otimes \Delta g)(x, Q^2)), \\
g_5^{\gamma Z}(x, Q^2) &= \sum_q e_q g_A^q (\Delta C_{5,q} \otimes \Delta q^-)(x, Q^2),
\end{aligned}
\tag{2.25}$$

where  $\Delta q^- = \Delta q - \Delta \bar{q}$  is the difference of quark and antiquark helicity PDFs, and the weak vector and axial-vector quark couplings are  $g_V^{u,c} = \frac{1}{2} - \frac{4}{3} \sin^2 \theta_W$ ,  $g_V^{d,s} = -\frac{1}{2} + \frac{2}{3} \sin^2 \theta_W$ , and  $g_A^{u,c} = \frac{1}{2} = -g_A^{d,s}$ , respectively. The contribution from the  $g_5^{\gamma Z}$  structure function to  $A_{UL}$  is suppressed by the factor  $g_V^e$  ( $\approx -4\%$ ) and is generally negligible in the  $x \lesssim 10^{-2}$  region. The  $g_1^{\gamma Z}$  structure function thus provides an independent linear combination of helicity PDFs, which, when combined with the electromagnetic  $g_1$  structure function, can allow cleaner flavor separation. For illustration, taking  $\sin^2 \theta_W \approx \frac{1}{4}$  we can write  $g_1^{\gamma Z}$  for the proton in the leading order approximation in terms of the quark singlet combination,

$$g_1^{\gamma Z,p}(x, Q^2) \approx \frac{1}{9} \Delta \Sigma(x, Q^2) = \sum_q \Delta q^+.
\tag{2.26}$$

In particular, this combination of helicity PDFs involves the  $s$  quark on equal footing with the  $u$  and  $d$  quarks, making parity-violating DIS an exciting process for extracting the strange helicity distribution, to which existing polarized fixed target data have little sensitivity [8]. Furthermore, as a purely inclusive process, parity-violating DIS provides constraints on the flavor separation of PDFs, that are independent of semi-inclusive DIS (SIDIS) observables which rely on flavor tagging [60, 61], and thus allow for new opportunities to test and validate the universality of PDFs.

## 2.2 Drell-Yan process

The Drell-Yan process [62] occurs in the collision of two hadrons (assumed as protons hereafter), where a quark from one proton annihilates with an antiquark from the other proton, creating a virtual photon or  $Z$  boson, which later decays into a detected lepton-antilepton pair. The leading order diagram is shown in Fig. 2.2

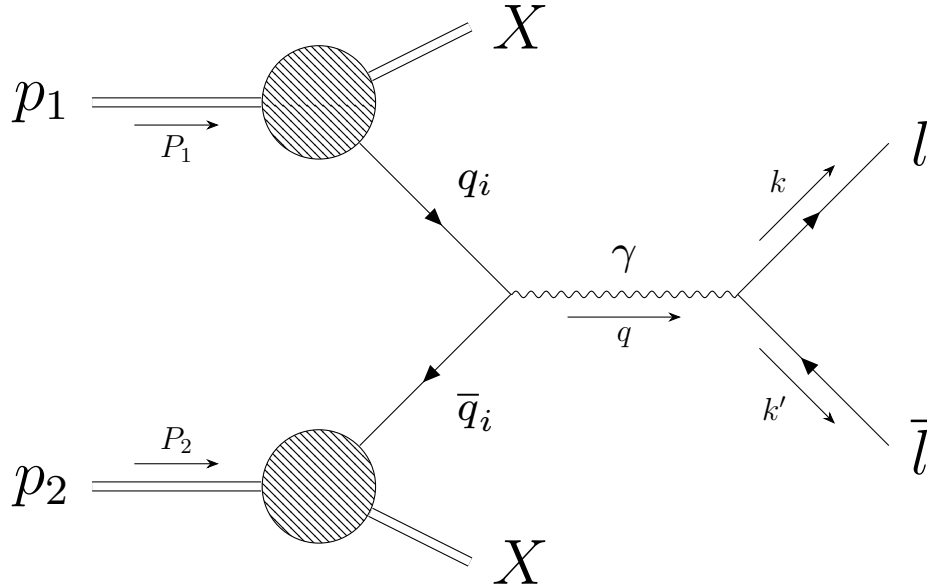


FIG. 2.2: The Drell-Yan process:  $p_1$  and  $p_2$  represent the two colliding protons,  $q_i$  and  $\bar{q}_i$  represent a quark and antiquark of the same flavor  $i$ ,  $l$  and  $\bar{l}$  represent the produced lepton and antilepton, and  $X$  represents undetected particles. The momentum  $q$  is timelike.

Assuming the protons carry momentum  $P_1$  and  $P_2$ , respectively, the quark and antiquark will carry momentum  $p_a \equiv x_a P_1$  and  $p_b \equiv x_b P_2$ , with  $x_a$  and  $x_b$  being the longitudinal momentum fractions carried by the quark and antiquark, respectively. With  $Q^2 \equiv q^2$  and  $x \equiv Q^2/S = Q^2/(P_1 + P_2)^2$ , the differential cross section can be written generally as [63]

$$\sigma = \sum_q \int_0^1 dx_a \int_0^1 dx_b \int_0^1 dz f_q(x_a, Q^2) f_{\bar{q}}(x_b, Q^2) \hat{\sigma}_{q\bar{q} \rightarrow l\bar{l}}(z, Q^2) \delta(x - zx_a x_b), \quad (2.27)$$

where  $\hat{\sigma}_{q\bar{q}\rightarrow l\bar{l}}$  is the partonic level differential cross section,  $f_q$  and  $f_{\bar{q}}$  are the PDFs of the annihilating quark and antiquark from the two protons, respectively, and the sum goes over all quark flavors.

The kinematic variables that are usually used to describe Drell-Yan events are Feynman- $x$  ( $x_F$ ) and the dilepton mass ( $M$ ), defined as

$$x_F \equiv x_a - x_b, \quad M^2 = Q^2 = x_a x_b S, \quad (2.28)$$

where we assume the massless limit of  $P_1^2 = P_2^2 = 0$ . At leading order, the cross section is given by [\[64\]](#)

$$\frac{M^3 d^2\sigma}{dM dx_F} = \frac{8\pi\alpha^2}{9(x_a + x_b)} \sum_q e_q^2 x_a x_b (f_q(x_a) f_{\bar{q}}(x_b) + f_{\bar{q}}(x_a) f_q(x_b)), \quad (2.29)$$

where the sum goes over all quark flavors, and  $e_q$  is the corresponding electric charge. When  $x_a \gg x_b$ , the Drell-Yan process is a very good probe of the  $\bar{d}/\bar{u}$  asymmetry. This can be seen by considering the leading order cross sections for proton-proton ( $pp$ ) and proton-neutron ( $pd$ ) collisions,

$$\begin{aligned} \sigma^{pp} &\propto \frac{4}{9} u(x_a) \bar{u}(x_b) + \frac{1}{9} d(x_a) \bar{d}(x_b), \\ \sigma^{pn} &\propto \frac{4}{9} u(x_a) \bar{d}(x_b) + \frac{1}{9} d(x_a) \bar{u}(x_b), \end{aligned} \quad (2.30)$$

which are simplified in the region  $0.1 \lesssim x \lesssim 0.5$ , where empirically  $u(x) \gg \bar{u}(x)$ ,  $d(x) \gg \bar{d}(x)$  and the strange distribution is negligible compared to the up and down

PDFs. Therefore, for the ratio one has

$$\frac{\sigma^{pd}}{2\sigma^{pp}} \approx \frac{\sigma^{pp} + \sigma^{pn}}{2\sigma^{pp}} \approx \frac{1}{2} \frac{1 + \frac{1}{4} \frac{d(x_a)}{u(x_a)}}{1 + \frac{1}{4} \frac{d(x_a)}{u(x_a)} \frac{\bar{d}(x_b)}{\bar{u}(x_b)}} \left( 1 + \frac{\bar{d}(x_b)}{\bar{u}(x_b)} \right), \quad (2.31)$$

and since  $d(x) \ll 4u(x)$ , one recovers the simple relation

$$\frac{\sigma^{pd}}{2\sigma^{pp}} \approx \frac{1}{2} \left( 1 + \frac{\bar{d}(x_2)}{\bar{u}(x_2)} \right). \quad (2.32)$$

In our analysis, we include  $pp$  and  $pd$  Drell-Yan data from the Fermilab E866 experiment [65], with a cut of  $Q^2 > 36 \text{ GeV}^2$ .

## 2.3 Jets in proton-proton collisions

A jet is a collimated spray of hadrons detected in high-energy particle reactions, and is a manifestation of collinear enhancement. To see this, consider the illustration in Fig. 2.3, where the propagator of the gluon line is

$$\frac{1}{(p_1 + p_2)^2} = \frac{1}{2E_1 E_2 (1 - \cos \theta)} \quad (2.33)$$

in the massless limit,  $p_1^2 = p_2^2 = 0$ , where  $p_1 = (E_1, \vec{p}_1)$ ,  $p_2 = (E_2, \vec{p}_2)$ , and  $\theta$  is the angle between  $\vec{p}_1$  and  $\vec{p}_2$ .

By having  $\cos \theta$  approaching 1, the amplitude for collinear emissions is enhanced. At the detector level, the collinear enhancement is manifested in the form of a collimated spray of energetic particles, known as a jet.

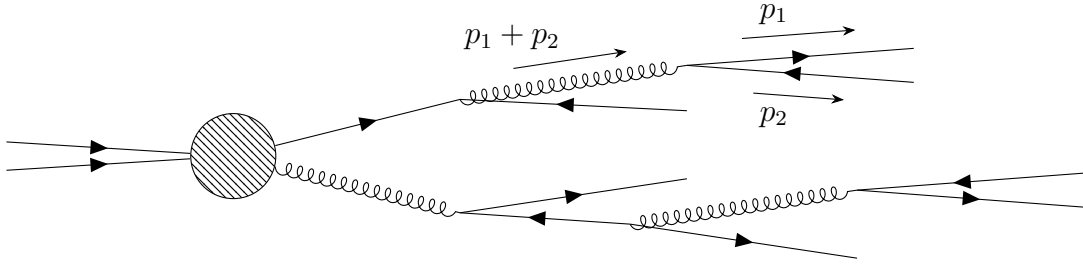


FIG. 2.3: Collinear enhancement in jet production.

### 2.3.1 Jets in unpolarized collisions

Jet production occurs when two protons (or other hadrons) collide, with a parton from one proton interacting with a parton from the other proton, producing quarks and/or gluons, which later hadronize into a collimated spray of energetic particles that are detected. Illustrated in Fig. 2.4 is the process for single jet production, where  $\hat{\sigma}_{ab}$  represents the partonic interactions which will be given in more detail in Appendix A.

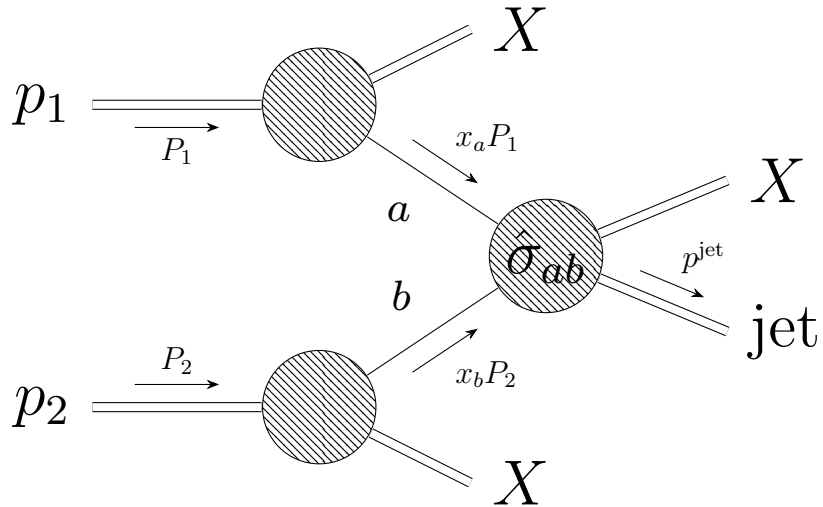


FIG. 2.4: Single jet production,  $\mathbf{p}_1$  and  $\mathbf{p}_2$  represent the two colliding protons,  $\hat{\sigma}_{ab}$  in the blob stands for the partonic level interactions, jet represents the detected jet and  $X$  represents undetected particles.

Assuming the protons carry momentum  $P_1$  and  $P_2$ , respectively, the partons from the two protons will carry momentum  $p_a \equiv x_a P_1$  and  $p_b \equiv x_b P_2$ , with  $x_a$  and  $x_b$  being the

longitudinal momentum fractions carried by the two partons  $a$  and  $b$ , respectively. The detected jet would carry momentum  $p^{\text{jet}}$ , with its transverse momentum being  $p_T$  and pseudorapidity  $\eta = -\ln(\tan \frac{\theta}{2})$ , where  $\theta$  is the angle between the proton beams and detected jet.

The differential cross section can be written as

$$\frac{d\sigma}{dp_T d\eta} = \sum_{a,b} \int dx_a \int dx_b f_a(x_a, \mu_F) f_b(x_b, \mu_F) d\hat{\sigma}_{ab \rightarrow \text{jet}X}(x_a, x_b, p^{\text{jet}}, \mu_{F,R}; r), \quad (2.34)$$

where  $\mu_F$  and  $\mu_R$  are the factorization and renormalization scales that are used in regulating the infrared and ultraviolet divergences, running of strong coupling constant  $\alpha_s$  and evolution of PDFs. They are usually chosen to be  $p_T$ , but can be varied to estimate theoretical uncertainties. The functions  $f_a$  and  $f_b$  are PDFs for partons  $a$  and  $b$  respectively. Finally,  $d\hat{\sigma}$  is the partonic differential cross section for inclusive single jet production, it depends on parton momenta  $p_a \equiv x_a P_1$  and  $p_b \equiv x_b P_2$ , jet momentum  $p^{\text{jet}}$ , factorization and renormalization scales  $\mu_F$  and  $\mu_R$ , and jet radius  $r$  which will be discussed in Sec. [2.3.3](#) below. Further manipulations of the expression [\(2.34\)](#) will be discussed later in Sec. [2.3.2](#).

According to Figs. [A.3](#) and [A.4](#), the single jet production process is sensitive not only to quark distributions, but also to gluon distributions at tree level. This makes the jet observable a very good probe of the gluon structure of the nucleon.

For unpolarized single jet production, we include measurements of single jet cross sections in  $p\bar{p}$  collisions from D0 [\[66\]](#) and CDF [\[67\]](#) at Fermilab, and in  $pp$  collisions from STAR [\[22\]](#) at RHIC.

### 2.3.2 Jets in polarized collisions

The double longitudinal spin asymmetry  $A_{LL}^{\text{jet}}$  for single inclusive jet production with polarized  $pp$  collisions is defined as

$$A_{LL}^{\text{jet}} = \frac{\sigma^{++} - \sigma^{+-}}{\sigma^{++} + \sigma^{+-}} \equiv \frac{d\Delta\sigma}{d\sigma}, \quad (2.35)$$

where  $\sigma^{++}$  and  $\sigma^{+-}$  are the inclusive jet cross sections when the two colliding proton beams have equal and opposite helicities, respectively. The numerator in  $A_{LL}^{\text{jet}}$  can be written in factorized form in terms of the perturbative hard-scattering cross section  $d\Delta\hat{\sigma}_{ab\rightarrow\text{jet}X}$  and spin-dependent PDFs  $\Delta f(x, \mu_F)$ ,

$$d\Delta\sigma = \sum_{a,b} \int dx_a \int dx_b \Delta f_a(x_a, \mu_F^2) \Delta f_b(x_b, \mu_F^2) d\Delta\hat{\sigma}_{ab\rightarrow\text{jet}X}(x_a, x_b, p^{\text{jet}}, \mu_{F,R}^2; r), \quad (2.36)$$

where the sum goes over all the partonic reactions  $a + b \rightarrow \text{jet} + X$  that contribute, and the variables  $x_a, x_b, p^{\text{jet}}, \mu_R, \mu_F$  and  $r$  are the same as in Eq. (2.34).

The measurements are typically given differential in the jet transverse momentum  $p_T$  and jet pseudorapidity  $\eta$ , introducing the kinematic variables

$$V = 1 - \frac{p_T e^\eta}{\sqrt{S}}, \quad W = \frac{p_T^2}{SV(1-V)}, \quad (2.37)$$

where  $S \equiv (P_1 + P_2)^2$  is the hadronic center of mass energy squared. Correspondingly, the

partonic level kinematic variables are

$$\begin{aligned}
s &\equiv (p_a + p_b)^2 \equiv (x_a P_1 + x_b P_2)^2, \\
t &\equiv (p_a - p^{\text{jet}})^2 \equiv (x_a P_1 - p^{\text{jet}})^2, \\
u &\equiv (p_b - p^{\text{jet}})^2 \equiv (x_b P_2 - p^{\text{jet}})^2, \\
v &\equiv 1 + \frac{t}{s}, \quad w \equiv \frac{-u}{s+t},
\end{aligned} \tag{2.38}$$

where  $x_a$  and  $x_b$  can be expressed as

$$x_a = \frac{VW}{vw}, \quad x_b = \frac{1-V}{1-v}. \tag{2.39}$$

With change of variables [68], one can therefore write Eq. (2.36) as

$$\begin{aligned}
\frac{d\Delta\sigma}{dp_T d\eta} &= \frac{2p_T}{S} \sum_{a,b} \int_{VW}^V \frac{dv}{v(1-v)} \int_{VW/v}^1 \frac{dw}{w} \Delta f_a(x_a, \mu_F) \Delta f_b(x_b, \mu_F) \\
&\times \left( \frac{d\Delta\hat{\sigma}_{ab \rightarrow \text{jet}X}^{(0)}(s, v)}{dv} \delta(1-w) + \frac{\alpha(\mu_R)}{\pi} \frac{d\Delta\hat{\sigma}_{ab \rightarrow \text{jet}X}^{(1)}(s, v, w, \mu_F, \mu_R; r)}{dv dw} \right).
\end{aligned} \tag{2.40}$$

The denominator of the asymmetry in Eq. (2.35), which is nothing but the unpolarized single jet production cross section, can be manipulated in a similar manner, with  $\Delta f(x, \mu_F^2) \rightarrow f(x, \mu_F^2)$ . Since in  $A_{LL}^{\text{jet}}$  the numerator and denominator depend on the spin-dependent and spin-averaged PDFs, respectively, in principle we should extract  $f(x, \mu_F^2)$  and  $\Delta f(x, \mu_F^2)$  simultaneously, and this will be discussed later in Section 5.4.



We integrated the Fortran code provided in Ref. [68] for the calculation of jet cross sections in unpolarized and polarized collisions, with the double counting in final state partons corrected [69, 70].

For the fit we include polarized jet production data in polarized  $pp$  collisions from the STAR [22–26] and PHENIX [27] collaborations at RHIC. We do not include the data presented in Ref. [71], which has been superseded by an updated analysis in Ref. [23].

### 2.3.3 Jet algorithms

In particle detectors, events are recorded as  $p_T$  values in the  $\eta$ – $\phi$  plane, where  $\eta$  and  $\phi$  are the pseudorapidity and azimuthal angles of the detected particles. It is then important to find the actual jet observation among the thousands of overserved events. Because a jet usually contains many hadronized particles, jet algorithms are designed to cluster together the particles that belong to a jet. Although many types of algorithms exist, they can be classified into seeded and seedless algorithms.

A seeded algorithm, like the midpoint cone algorithm [72], consists of two main parts. The first part involves constructing a list of proto-jets, with the following steps:

1. start with a list of events ordered by  $p_T$ , from high to low;
2. create an empty list of proto-jets;
3. from the events outside of proto-jets, use the one with the highest  $p_T$  as the seed;
4. compute the distances  $r_i$  between the seed and all the events (numbered by  $i$ );
5. collect all the events that have a distance  $r_i < R$ ;
6. recompute the jet axis of collected events and use it as the new seed;
7. repeat from step 4 until a stable jet axis is obtained;
8. declare the collected events as a proto-jet and put into the proto-jet list;

9. repeat from step 3 until the all the events are in at least one proto-jet.

The next part would be to check if there are any proto-jets that overlap with each other. If the answer is no, then one can claim all proto-jets as jets and the algorithm is finished. If the answer is yes, then one proceeds with the steps:

1. compute the overlapping energy of the two proto-jets;
2. test whether it exceeds a threshold fraction of the smaller proto-jet;
  - (a) if yes, then merge the two proto-jets;
  - (b) if no, then split the overlapping events by their distances to the two proto-jets' axes;
3. repeat from step 1 until no proto-jets overlap;
4. declare all proto-jets as jets, and the algorithm is finished.

The distance  $r_i$  between the seed and  $i$ th event is defined as  $r_i = \sqrt{(\eta_i - \eta_s)^2 + (\phi_i - \phi_s)^2}$ , where  $\eta_{i,s}$  and  $\phi_{i,s}$  are the pseudorapidities and azimuthal angles of the  $i$ th event (or seed). Parameter  $R$  is usually chosen to be of order 1, and the threshold fraction is usually chosen to be 50%.

The midpoint cone algorithm is infrared safe (up to NLO) and collinear safe, meaning that the jets do not change when adding a soft particle or substituting one particle with two collinear particles. The midpoint cone algorithm was used by the D0 [66] and STAR [22, 23] collaborations. The PHENIX [27] measurement, however, employed a cone jet algorithm without splitting or merging of overlapped proto-jets.

Seedless algorithms, such as  $k_T$  [73] and anti- $k_T$  [74] algorithms, are recursive algorithms and can be described by the following:

1. start with a list of events;
2. compute the distance between the beam ( $B$ ) and each event  $i$  as  $d_{iB} = E_{T,i}^{2p}$ , defined by the transverse energy  $E_{T,i}$  of event  $i$  raised to the power of  $2p = \pm 2$ ;
3. for each pair of events  $i$  and  $j$ , also compute  $d_{ij} = \min(E_{T,i}^{2p}, E_{T,j}^{2p})r_{ij}^2/R^2$ ;
4. find the smallest among  $d_{iB}$  and  $d_{ij}$ , label it as  $d_{\min}$ ;
5. if  $d_{\min}$  is a  $d_{ij}$ , merge events  $i$  and  $j$  into a new event;
6. if  $d_{\min}$  is a  $d_{iB}$ , claim event  $i$  a jet and remove from list;
7. repeat from step 2 until there are no more events.

In a similar manner as for the midpoint cone algorithm, here  $r_{ij}$  is defined by  $r_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$ ,  $R$  is usually chosen to be of order 1, and  $p = \pm 1$  for the  $k_T$  and anti- $k_T$  algorithms, respectively. Also, the transverse energy  $E_{T,i}$  is defined as  $E_{T,i} \equiv E_i \sin \theta_i$ , with  $E_i$  being the total energy of event  $i$  and  $\theta_i$  the angle between event  $i$  and hadron beams.

The  $k_T$  algorithm tends to cluster low energy events first and is used by CDF [67]. In contrast, the anti- $k_T$  algorithm tends to cluster high energy events first and is used by STAR [24–26]. Both algorithms are infrared and collinear safe.

## 2.4 Observables in Mellin space

Since PDFs are extracted numerically from experimental data through global QCD analysis involving significant computational resources, it is beneficial to improve the speed of evaluating cross sections by reducing the computational complexity (which we will also discuss in Section 3.2). In this section we introduce the Mellin transform, a method that can help us achieve faster numerical computations [75]. The Mellin transform of a function

$f(x)$  is defined as

$$\mathbf{f}(N) = \int_0^1 f(x)x^{N-1} dx, \quad (2.41)$$

where  $N$  is a complex number. When  $N$  is an integer, in particular,  $\mathbf{f}(N)$  is called the  $N$ th Mellin moment of  $f(x)$ .

An observable  $\mathcal{O}(x, Q^2)$  that is defined by a single convolution integral in  $x$  space, such as in Eqs. (2.6) and (2.13), becomes a simple product in Mellin space,

$$\mathcal{O}(N, Q^2) = \sum_i \mathbf{H}_i(N, Q^2) \mathbf{f}_i(N, Q^2), \quad (2.42)$$

where  $\mathbf{H}_i$  is the generic hard scattering coefficient and  $\mathbf{f}_i$  is the nonperturbative function, both in Mellin space, the sum goes over all parton flavors.

To compare with the experimental data,  $\mathcal{O}(N, Q^2)$  has to be converted back to  $x$  space with the inverse Mellin transform, which can be performed by a contour integral in the complex plane,

$$\mathcal{O}(x, Q^2) = \frac{1}{2\pi i} \oint dN x^{-N} \mathcal{O}(N, Q^2). \quad (2.43)$$

In practice, the contour is taken as the lines shown in Fig. 2.5. The upper part of the contour is defined as  $N = c + ze^{i\phi}$ , where  $c$  is on the real axis and is fixed to the right of the rightmost pole of  $\mathcal{O}(N, Q^2)$ , and  $\phi$  is usually chosen to be  $3\pi/4$  for the integral to quickly converge, parameter  $z$  is a positive real number. The lower part of the contour is

a mirror reflection of the upper part with respect to the real axis.

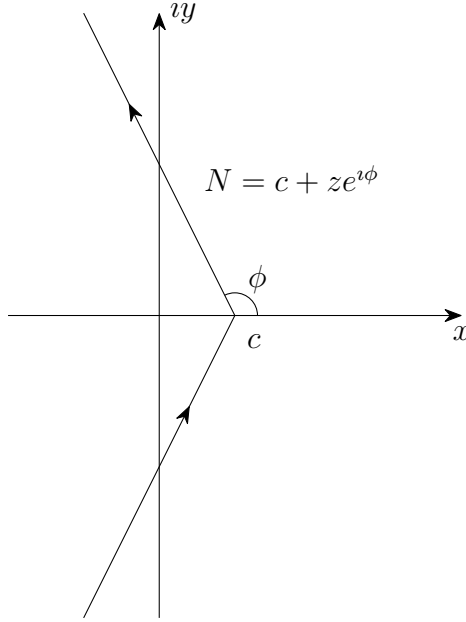


FIG. 2.5: Mellin contour for the integration in complex plane.

The integral can then be expressed in terms of the variable  $z$ ,

$$\mathcal{O}(x, Q^2) = \frac{1}{\pi} \int_0^\infty \text{Im} \{ e^{i\phi} x^{-N} \mathcal{O}(N, Q^2) \} dz, \quad (2.44)$$

where the symmetry with respect to the real axis is applied. To evaluate the integral over  $z$ , numerically we use the Gaussian quadrature sum

$$\begin{aligned} \mathcal{O}(x, Q^2) &= \frac{1}{\pi} \sum_l w_l \text{Im} \{ e^{i\phi} x^{-N_l} \mathcal{O}(N_l, Q^2) \} \\ &= \frac{1}{\pi} \sum_l w_l \text{Im} \left\{ e^{i\phi} x^{-N_l} \sum_i \mathbf{H}_i(N_l, Q^2) \mathbf{f}_i(N_l, Q^2) \right\}, \end{aligned} \quad (2.45)$$

where  $w_l$  is the Gaussian weight for the  $l$ th Gaussian node  $z_l$  and  $N_l = c + z_l e^{i\phi}$ . The advantage of this method is that the computationally expensive coefficients  $\mathbf{H}_i(N_l, Q^2)$  can be pre-calculated and stored in the so-called ‘‘Mellin tables’’. Later, the Mellin tables can be loaded and summed with the PDFs (in Mellin space) to quickly produce the cross sections (in Mellin space).

For the scenarios where there are two variables in  $x$  space (Eq. (2.27)), we would use the double Mellin method [75]. The Mellin transform for an observable  $\mathcal{O}(x_1, x_2, Q^2)$  that convolutes with two nonperturbative functions  $f_1(x)$  and  $f_2(x)$  is given by

$$\mathcal{O}(N, M, Q^2) = \sum_{i,j} \mathbf{H}_{i,j}(N, M, Q^2) \mathbf{f}_{1,i}(N, Q^2) \mathbf{f}_{2,j}(M, Q^2), \quad (2.46)$$

where  $\mathbf{H}_{i,j}$  is the double Mellin version of the hard scattering coefficient. Returning back to  $x$  space follows naturally with the double Mellin inversion

$$\mathcal{O}(x_1, x_2, Q^2) = \frac{1}{(2\pi i)^2} \oint \! \! \! \oint dN dM x_1^{-N} x_2^{-M} \mathcal{O}(N, M, Q^2). \quad (2.47)$$

Analogously, by defining  $N = c_n + z_n e^{i\phi_n}$  and  $M = c_m + z_m e^{i\phi_m}$ , and performing a single Mellin inversion twice, we have

$$\begin{aligned} \mathcal{O}(x_1, x_2, Q^2) = & -\frac{1}{2\pi^2} \operatorname{Re} \left[ \int_0^\infty dz_n \int_0^\infty dz_m x_1^{-N} \right. \\ & \left. \times \left( e^{i(\phi_n + \phi_m)} x_2^{-M} \mathcal{O}(N, M, Q^2) - e^{i(\phi_n - \phi_m)} x_2^{-M^*} \mathcal{O}(N, M^*, Q^2) \right) \right]. \end{aligned} \quad (2.48)$$

As in the single Mellin scenario, we evaluate the integral with Gaussian quadrature,

$$\begin{aligned} \mathcal{O}(x_1, x_2, Q^2) = & -\frac{1}{2\pi^2} \operatorname{Re} \left[ \sum_i w_i \sum_j w_j x_1^{-N_i} \right. \\ & \left. \times \left( e^{i(\phi_n + \phi_m)} x_2^{-M_j} \mathcal{O}(N_i, M_j, Q^2) - e^{i(\phi_n - \phi_m)} x_2^{-M_j^*} \mathcal{O}(N_i, M_j^*, Q^2) \right) \right], \end{aligned} \quad (2.49)$$

where the inner sum over  $j$  is taken first, and  $w_i$  and  $w_j$  are the Gaussian weights. Typically,  $N$  and  $M$  take the same contour, so that  $c_n = c_m$  and  $\phi_n = \phi_m$ .

The Mellin transform is also useful in evaluating DGLAP evolution, which will be discussed in Section [3.2](#).

# CHAPTER 3

## Parton distribution functions

In the previous chapters, we saw that the high-energy scattering observables in QCD can be described by hard scattering coefficients and soft nonperturbative functions, with the soft nonperturbative functions having to be extracted through global QCD analysis. In this chapter, we will discuss how to parameterize those soft nonperturbative functions, namely the PDFs. A good parameterization of PDFs is a balance between flexibility and complexity, meaning the PDFs are sufficiently flexible to describe the data, but also not too complex that the computation takes too long.

This chapter will also discuss the scale dependence of the PDFs, using the DGLAP evolution equations. Some of the theoretical constraints on PDFs will also be presented, including number and momentum sum rules for spin-averaged PDFs, SU(2) and SU(3) flavor symmetries for spin-dependent PDFs, and positivity constraints.



### 3.1 Parameterization

For the parameterization of the spin-averaged and spin-dependent PDFs we use the standard form with a generic template function at the input scale  $\mu_0$ ,

$$T(x, \mu_0; \mathbf{a}) = \frac{a_0}{\mathcal{N}} x^{a_1} (1-x)^{a_2} (1 + a_3 \sqrt{x} + a_4 x), \quad (3.1)$$

where  $\mathbf{a} = \{a_0, \dots, a_4\}$  denotes the set of parameters to be fitted, and the normalization constant  $\mathcal{N}$  is defined by Euler beta functions

$$\mathcal{N} = B(a_1 + n, a_2 + 1) + a_3 B\left(a_1 + n + \frac{1}{2}, a_2 + 1\right) + a_4 B(a_1 + n + 1, a_2 + 1), \quad (3.2)$$

which is the  $n$ th moment (defined by Eq. (2.41)) of the bare function. This form is chosen to maximally decorrelate the overall normalization parameters from the parameters that determine the shape of the PDFs in  $x$ .

For the spin-averaged PDFs, we parameterize the valence distributions  $u_v = u - \bar{u}$  and  $d_v = d - \bar{d}$  and the gluon distribution  $g$  each using a single template function from (3.1). For the sea quark and antiquark distributions, in order to allow greater flexibility we use two template functions, each parameterized independently,

$$q(x, \mu_0) = \mathcal{S}(x, \mu_0) + q_0(x, \mu_0), \quad (3.3)$$

with a flavor-independent term  $\mathcal{S}$  that is more singular (more negative  $a_1$  values) and

thus dominates the  $x \rightarrow 0$  behavior, and a flavor-dependent part  $q_0$  that allows for a nonperturbative origin of the sea (with less negative  $a_1$  value). This form is used for the antiquark  $\bar{u}$  and  $\bar{d}$  PDFs, as well as the strange and antistrange distributions.

The parameter  $a_0$  of the gluon PDF is fixed by the momentum sum rule, while the  $a_0$  for the  $u_v$ ,  $d_v$ , and  $s_0$  are fixed by the valence number sum rules,

$$\int_0^1 x(g(x) + u^+(x) + d^+(x) + s^+(x)) dx = 1, \quad (3.4)$$

$$\int_0^1 u_v(x) dx = 2, \quad \int_0^1 d_v(x) dx = 1, \quad \int_0^1 (s_0(x) - \bar{s}_0(x)) dx = 0,$$

where  $q^+ = q_v + 2\bar{q}$  for  $q = u, d$  and  $s^+ = s + \bar{s}$ . The value of  $n$  is chosen to be 2 for the normalization  $\mathcal{N}$  in Eq. (3.2), so that  $a_0$  corresponds to the second moments which are used in the momentum sum rule (3.4). The parameters  $a_3$  and  $a_4$  are fixed to be zero for the  $s_0$ ,  $\bar{s}_0$  and  $\mathcal{S}$  components, but are free to vary for the valence, gluon, and other sea quark distributions.

The parameter ranges for the spin-averaged PDFs are summarized in Table 3.1. Constraints are not set for parameters  $a_0$  of  $u_v$ ,  $d_v$  and  $g$  because they are fixed by the number or momentum sum rules (3.4), while  $a_0$  is set to be greater than 0 for other flavors to ensure that PDFs are generally positive. The first full moments are required to be finite by setting  $a_1 > -1$  for the flavors that are used in the number sum rules, which are all the flavors except  $\mathcal{S}$  and  $g$ . This limit is further enforced to  $a_1 > -0.5$  for  $u_v$ ,  $d_v$ ,  $s_0$  and  $\bar{s}_0$  so that they dominate the high- $x$  region, leaving the-low  $x$  region to the universal sea distribution  $\mathcal{S}$ , which is more singular. The gluon distribution, however, is not bounded by the number sum rule, and therefore a more generous lower limit of  $-1.9$  is set for  $a_1$  to ensure its second full moment still exists for the momentum sum rule. The  $a_1$  parameter

TABLE 3.1: Parameter ranges for the spin-averaged PDFs

	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$
$g$	$(-\infty, \infty)$	$[-1.9, 1]$	$[0, 10]$	$[-10, 10]$	$[-10, 10]$
$u_v$	$(-\infty, \infty)$	$[-0.5, 1]$	$[0, 10]$	$[-10, 10]$	$[-10, 10]$
$d_v$	$(-\infty, \infty)$	$[-0.5, 1]$	$[0, 10]$	$[-10, 10]$	$[-10, 10]$
$\bar{u}_0$	$[0, 1]$	$[-1, 1]$	$[0, 10]$	$[-10, 10]$	$[-10, 10]$
$\bar{d}_0$	$[0, 1]$	$[-1, 1]$	$[0, 10]$	$[-10, 10]$	$[-10, 10]$
$s_0$	$[0, 1]$	$[-0.5, 1]$	$[0, 10]$		
$\bar{s}_0$	$[0, 1]$	$[-0.5, 1]$	$[0, 10]$		
$\mathcal{S}$	$[0, 1]$	$[-1.9, -1]$	$[0, 10]$		

of the universal sea distribution  $\mathcal{S}$  is set within  $-1.9$  and  $-1$  to ensure its dominance in the low- $x$  region. The lower limits for  $a_2$  of all flavors are set to 0 to avoid divergences at  $x = 1$ . The remaining limits are set to be large enough so that there is no edge effect.

For the spin-dependent PDFs, the same template (3.1) is used, with the exception that  $n$  is chosen to be 1 for the normalization  $\mathcal{N}$  in Eq. (3.2) so that  $a_0$  maximally corresponds to the first full moments that are used in the SU(2) and/or SU(3) flavor symmetries.

The helicity valence flavors  $\Delta u_v = \Delta u - \Delta \bar{u}$  and  $\Delta d_v = \Delta d - \Delta \bar{d}$  are parameterized with a single shape, while the helicity gluon and sea/antiquark distributions are parameterized each with two shapes for more flexibility

$$\begin{aligned} \Delta \bar{u} = \Delta \bar{d} = \Delta \bar{s} = \Delta s = \Delta s_0 + \Delta \mathcal{S}_1, \\ \Delta g = \Delta g_0 + \Delta \mathcal{S}_2, \end{aligned} \tag{3.5}$$

where we set the sea distributions equal (since there is no sensitivity to individual flavors).

The  $a_3$  and  $a_4$  parameters are fixed to zero for all flavors, with all other parameters free.

The  $a_0$  normalization parameter for the valence  $\Delta u_v$  and/or  $\Delta d_v$  PDFs are fixed by fitting to the triplet and octet axial vector charges  $g_A$  and  $a_8$  with uncertainties taken into account, using the first moments of  $\Delta q^+ = \Delta q_v + 2\Delta\bar{q}$  for  $\Delta q = \Delta u, \Delta d$  and  $\Delta s^+ = \Delta s + \Delta\bar{s}$ ,

$$\begin{aligned} \int_0^1 (\Delta u^+(x, Q^2) - \Delta d^+(x, Q^2)) dx &= g_A, \\ \int_0^1 (\Delta u^+(x, Q^2) + \Delta d^+(x, Q^2) - 2\Delta s^+(x, Q^2)) dx &= a_8. \end{aligned} \tag{3.6}$$

The axial charges are given by the standard values  $g_A = 1.269(3)$  and  $a_8 = 0.586(31)$  [76], or  $g_A = 1.24(4)$  and  $a_8 = 0.46(21)$  extracted by fitting spin-dependent PDFs and fragmentation functions (FFs) with polarized inclusive and semi-inclusive DIS, and single-inclusive  $e^+e^-$  annihilation data simultaneously in JAM17 [8]. Unlike the spin-averaged scenario, where the sum rules are imposed by strict relations, the sum rules in (3.6) are restricted by including the central values and uncertainties of  $g_A$  and  $a_8$  as data points.

The parameter ranges for the spin-dependent PDFs are summarized in Table 3.2. Since helicity PDFs are differences of parton distributions, they are not bound to be positive, and therefore the range of  $a_0$  is set to be  $[-10, 10]$  for all flavors. To ensure that the first full moments are finite in the sum rules (3.6), the lower limit of  $a_1$  is set to be  $-0.99$ . The lower limits of  $a_1$  are further enforced to be  $-0.5$  for  $\Delta g$ ,  $\Delta u_v$ ,  $\Delta d_v$  and  $\Delta s_0$ , making sure that  $\Delta\mathcal{S}_1$  and  $\Delta\mathcal{S}_2$  will dominate the low- $x$  region. The lower limits for  $a_2$  for all flavors are set to 0 to avoid divergences at  $x = 1$ . The remaining limits are set to be large enough so that there are no edge effects.

Both the spin-averaged and spin-dependent PDFs are parameterized at the input scale  $\mu_0^2 = (1.27)^2 \text{ GeV}^2$ , which is slightly below the charm mass ( $m_c = 1.28 \text{ GeV}$ ) that we adopt from the PDG [77].

TABLE 3.2: Parameter ranges for the spin-dependent PDFs

	$a_0$	$a_1$	$a_2$
$\Delta g$	$[-10, 10]$	$[-0.5, 2]$	$[0, 10]$
$\Delta u_v$	$[-10, 10]$	$[-0.5, 2]$	$[0, 10]$
$\Delta d_v$	$[-10, 10]$	$[-0.5, 2]$	$[0, 10]$
$\Delta s_0$	$[-10, 10]$	$[-0.5, 2]$	$[0, 10]$
$\Delta \mathcal{S}_1$	$[-10, 10]$	$[-0.99, 2]$	$[0, 10]$
$\Delta \mathcal{S}_2$	$[-10, 10]$	$[-0.99, 2]$	$[0, 10]$

### 3.2 DGLAP evolution

The spin-averaged and spin-dependent PDFs are parameterized at the input scale  $\mu_0^2$ , and in order to evolve them to the experimental scales, we use the DGLAP evolution equations [78]. In  $x$  space, the evolution equations for the spin-averaged quark and gluon PDFs are given by

$$\begin{aligned} \frac{dq_i(x, \mu^2)}{d \ln(\mu^2)} &= \frac{\alpha_s(\mu^2)}{2\pi} \int_x^1 \frac{dy}{y} \left( q_i(y, \mu^2) P_{qq} \left( \frac{x}{y} \right) + g(y, \mu^2) P_{qg} \left( \frac{x}{y} \right) \right), \\ \frac{dg(x, \mu^2)}{d \ln(\mu^2)} &= \frac{\alpha_s(\mu^2)}{2\pi} \int_x^1 \frac{dy}{y} \left( g(y, \mu^2) P_{gg} \left( \frac{x}{y} \right) + \left( \sum_i q_i^+(y, \mu^2) \right) P_{gq} \left( \frac{x}{y} \right) \right), \end{aligned} \quad (3.7)$$

where  $P_{qq}$ ,  $P_{qg}$ ,  $P_{gq}$  and  $P_{gg}$  are the quark to quark, gluon to quark, quark to gluon, and gluon to gluon splitting functions, respectively, and  $q_i$ ,  $\bar{q}_i$  and  $g$  are the PDFs for quark of flavor  $i$ , antiquark of flavor  $i$ , and the gluon, and finally  $q_i^+ = q_i + \bar{q}_i$ . Using the definition of Mellin convolution (2.7), the above equations can be more compactly written as

$$\begin{aligned} \frac{dq_i(x, \mu^2)}{d \ln(\mu^2)} &= \frac{\alpha_s(\mu^2)}{2\pi} (q_i \otimes P_{qq} + g \otimes P_{qg}), \\ \frac{dg(x, \mu^2)}{d \ln(\mu^2)} &= \frac{\alpha_s(\mu^2)}{2\pi} \left( g \otimes P_{gg} + \sum_i q_i^+ \otimes P_{gq} \right). \end{aligned} \quad (3.8)$$

Solving these convoluted integro-differential equations can be highly nontrivial and computationally expensive [79–81]. By transforming to the Mellin space, however, the convolution is simplified to products,

$$\begin{aligned}\frac{d\mathbf{q}_i(N, \mu^2)}{d \ln(\mu^2)} &= \frac{\alpha_s(\mu^2)}{2\pi} (\mathbf{q}_i \mathbf{P}_{qq} + \mathbf{g} \mathbf{P}_{qg}), \\ \frac{d\mathbf{g}(N, \mu^2)}{d \ln(\mu^2)} &= \frac{\alpha_s(\mu^2)}{2\pi} \left( \mathbf{g} \mathbf{P}_{gg} + \sum_i (\mathbf{q}_i + \bar{\mathbf{q}}_i) \mathbf{P}_{gq} \right),\end{aligned}\tag{3.9}$$

and can be solved by the strategy presented in Ref. [82].

Similarly for the spin-dependent PDFs, we have in Mellin space the DGLAP equations

$$\begin{aligned}\frac{d\Delta\mathbf{q}_i(N, \mu^2)}{d \ln(\mu^2)} &= \frac{\alpha_s(\mu^2)}{2\pi} (\Delta\mathbf{q}_i \Delta\mathbf{P}_{qq} + \Delta\mathbf{g} \Delta\mathbf{P}_{qg}), \\ \frac{d\Delta\mathbf{g}(N, \mu^2)}{d \ln(\mu^2)} &= \frac{\alpha_s(\mu^2)}{2\pi} \left( \Delta\mathbf{g} \Delta\mathbf{P}_{gg} + \sum_i (\Delta\mathbf{q}_i + \Delta\bar{\mathbf{q}}_i) \Delta\mathbf{P}_{gq} \right),\end{aligned}\tag{3.10}$$

which can also be solved as discussed in Ref. [82]. The heavy quark mass thresholds for the evolution of the PDFs and  $\alpha_s$  within the zero-mass variable-flavor-number-scheme (ZM-VFNS) [83] are chosen from the PDG values  $m_c = 1.28$  GeV and  $m_b = 4.18$  GeV in the  $\overline{\text{MS}}$  scheme [77].

The splitting functions in leading order and next-to-leading order are given in Ref. [84].

### 3.3 Positivity constraints

One of the important theory assumptions that will be tested are the positivity constraints. In the naïve parton model, the spin-averaged PDFs are positive because they are inter-

preted as the sum of the number densities of partons with parallel ( $f^\uparrow$ ) and antiparallel ( $f^\downarrow$ ) polarizations with respect to the parent hadron (proton). The spin-dependent PDFs, being the difference between  $f^\uparrow$  and  $f^\downarrow$ , should be bounded by the corresponding spin-averaged distributions [85].

However, the naïve parton model can only be recovered in the asymptotic limit. At any finite scale  $Q^2$ , it receives non-negligible perturbative corrections. Therefore, a more reliable positivity bound should only be derived from the requirement that any measurable physical cross sections have to be positive.

A brief derivation of the positivity constraints will be presented here; for a more detailed description, see Ref. [85]. In the Bjorken limit, the photoproduction asymmetry  $A_1$  for a transverse virtual photon and a longitudinally polarized nucleon, as in Eq. (2.20), is

$$A_1(x, Q^2) = \frac{\sigma_{1/2} - \sigma_{3/2}}{\sigma_{1/2} + \sigma_{3/2}} = \frac{g_1(x, Q^2) - \gamma^2 g_2(x, Q^2)}{F_1(x, Q^2)} \approx \frac{g_1(x, Q^2)}{F_1(x, Q^2)}, \quad (3.11)$$

where  $g_2$  is neglected due to the presence of  $\gamma^2 = 4M^2 x^2 / Q^2$  when  $M^2 \ll Q^2$ , and the subscripts for  $\sigma$  denote the total angular momentum of the photon-nucleon pair along the incoming lepton's direction. The absolute value of the numerator therefore has to be bound by that of the denominator because both  $\sigma_{1/2}$  and  $\sigma_{3/2}$  are observables, which then gives us

$$|g_1(x, Q^2)| \leq |F_1(x, Q^2)|. \quad (3.12)$$

At leading order, we know from Eqs. (2.9) and (2.15) that

$$\begin{aligned} F_1(x, Q^2) &= \frac{1}{2} \sum_i e_i^2 (f_i + \bar{f}_i)(x, Q^2), \\ g_1(x, Q^2) &= \frac{1}{2} \sum_i e_i^2 (\Delta f_i + \Delta \bar{f}_i)(x, Q^2), \end{aligned} \tag{3.13}$$

where  $i$  goes over all quark flavors. Therefore we have

$$\left| \sum_i e_i^2 (\Delta f_i + \Delta \bar{f}_i)(x, Q^2) \right| \leq \left| \sum_i e_i^2 (f_i + \bar{f}_i)(x, Q^2) \right|. \tag{3.14}$$

Furthermore, this bound has to be satisfied for any choice of target (hence for any quark and antiquark combination), as well as for charged-current scattering. This relation must therefore be true for any quark flavors at leading order,

$$|\Delta f_i(x, Q^2)| \leq |f_i(x, Q^2)|, \quad |\Delta \bar{f}_i(x, Q^2)| \leq |\bar{f}_i(x, Q^2)|. \tag{3.15}$$

The positivity bound for gluon distributions can be similarly derived from the inclusive Higgs production in gluon-gluon scattering [85], and gives

$$|\Delta g(x, Q^2)| \leq |g(x, Q^2)|. \tag{3.16}$$

Two caveats for the positivity constraints (3.15) and (3.16) follow. First, the bounds are only true at leading order, while at next-to-leading and higher orders, more careful treatments are needed. Second, the constraints are derived from physical observables,



which means for the kinematic regions that are not covered by available data, they are less justified to be used for constraining the helicity PDFs.

This is why we would like to test and confront the justification of positivity constraints. Before doing that, however, we need to discuss how to impose the constraints. After some exploration, we found that the optimal numerical implementation of positivity constraints is as follows:

1. compute  $f_i(x, \mu_0^2)$  and  $\Delta f_i(x, \mu_0^2)$  for all  $x \in X$  and all flavors  $i$ ;
2. calculate  $f_i^\uparrow(x, \mu_0^2)$  and  $f_i^\downarrow(x, \mu_0^2)$  at the above values;
3. if a value is negative for any flavor at any  $x$ , it is added to two arrays distinguishing  $f^\uparrow$  and  $f^\downarrow$ ;
4. the arrays are multiplied by a factor of 150 and returned as residuals.

Here,  $\mu_0^2 = (1.27)^2 \text{ GeV}^2$  is the input scale, and  $X$  is an array consisting of the points 0.005, 0.01, and 100 more points evenly distributed between 0.1 and 0.9. Denser points are chosen at high  $x$  for the array  $X$  because positivity bounds tend to be violated in this region. The multiplicative factor of 150 is used to make sure the violation to positivity is not too small such that the fitter will ignore it. The returned residuals will be minimized in the same way as for the residuals from cross sections.

Positivity is imposed only at the input scale  $\mu_0$  because once positivity is satisfied at a lower scale, it will be preserved by the DGLAP equations to higher scales [85].

# CHAPTER 4

## Aspects of fitting

In the previous chapters, we have seen that high-energy scattering observables can be described in QCD by hard scattering coefficients and soft nonperturbative functions, with the soft nonperturbative functions extracted from experimental data through global QCD analysis. In this chapter, the discussion will focus on the practical aspects of global QCD analysis, including optimization of fit parameters, estimation of uncertainties, enhancement of computational speed, and the treatment of experimental observables.

The methodology that will be elaborated in this chapter is the Monte Carlo Bayesian approach developed by the JAM collaboration. Starting with the spin PDF analysis known as JAM15 [55], the JAM collaboration developed an iterative Monte Carlo (IMC) procedure for analysing both leading and higher twist quark helicity distributions using Jefferson Lab and global DIS data. Following this, the JAM16 [60] analysis extracted for the first time fragmentation functions (FFs) from single-inclusive  $e^+e^-$  annihilation (SIA) using Monte Carlo methods. Later in the JAM17 [8] study, spin-dependent PDFs and FFs were for the

first time fitted simultaneously, with emphasis on the impact of semi-inclusive DIS (SIDIS) data on sea quark helicity. With the JAM19 [61] and JAM20-SIDIS [86] global analyses, spin-averaged PDFs and FFs were fitted simultaneously, focusing on the impact of  $\pi^\pm$  and  $K^\pm$  production and SIDIS data. The present work features the reliable extraction of the gluon helicity via the inclusion of unpolarized and polarized single jet production data from the Tevatron and RHIC, and is the first-ever simultaneous analysis of spin-averaged and spin-dependent PDFs.

## 4.1 Bayesian approach

The key objective of a global QCD analysis is to obtain an accurate description of the experimental data within the given model. In statistics, this is equivalent to finding the probability  $\rho$  of obtaining different sets of model parameters  $\mathbf{a}$  in the fits given the knowledge of data, written in Bayes' formula

$$\rho(\mathbf{a}|\text{data}) = \frac{1}{Z} \mathcal{L}(\text{data}|\mathbf{a})\pi(\mathbf{a}), \quad (4.1)$$

where  $Z$  is the “evidence” parameter and acts as the normalization factor

$$Z = \int \mathcal{L}(\text{data}|\mathbf{a})\pi(\mathbf{a}) d^n a, \quad (4.2)$$

and  $n$  is the dimension of  $\mathbf{a}$ . The function  $\pi(\mathbf{a})$  is the prior probability distribution, and  $\mathcal{L}(\text{data}|\mathbf{a})$  is the Gaussian likelihood function. The prior probability distribution  $\pi$ , as its name suggests, is the distribution of the parameters before knowledge of the data. The likelihood function, on the other hand, encodes the probability that the data can be

inferred from the prior distribution. In the context of JAM analysis, the prior distribution is initially set to a flat distribution to avoid bias in the parameter choices. At the same time, boundaries are set so that the parameters do not cross into unphysical regions.

The rest of this section will discuss the likelihood function that is used for optimizing the parameters, and the multi-step strategy that we develop for increasing the efficiency of fits.

### 4.1.1 Likelihood function

The Gaussian likelihood function is defined in terms of the  $\chi^2$  function as

$$\mathcal{L}(\mathbf{a}, \text{data}) = \exp\left(-\frac{1}{2}\chi^2(\mathbf{a}, \text{data})\right). \quad (4.3)$$

For uncorrelated uncertainties  $\alpha_i$  only, the  $\chi^2$  function can be defined as

$$\chi^2(\mathbf{a}, \text{data}) = \sum_{e,i} \left(\frac{d_{e,i} - t_{e,i}(\mathbf{a})}{\alpha_{e,i}}\right)^2, \quad (4.4)$$

where  $d_{e,i}$  is the  $i$ th data value from experiment  $e$ , and  $\alpha_{e,i}$  and  $t_{e,i}$  are its corresponding uncorrelated uncertainty and theoretical prediction, respectively. However, many experiments provide elaborate analysis of data with correlated uncertainties and normalization

uncertainty, which requires the  $\chi^2$  function to be generalized as

$$\chi^2(\mathbf{a}, \text{data}) = \sum_e \left( \sum_i \left( \frac{d_{e,i} - \sum_k r_{e,k} \beta_{e,k,i} - t_{e,i}(\mathbf{a})/N_e}{\alpha_{e,i}} \right)^2 + \sum_k r_{e,k}^2 + \left( \frac{1 - N_e}{\delta \mathcal{N}_e} \right)^2 \right). \quad (4.5)$$

While  $d_{e,i}$ ,  $\alpha_{e,i}$  and  $t_{e,i}$  are as in Eq. (4.4), the correlated uncertainties  $\beta_{e,k,i}$  and normalization parameter  $N_e$  modify the theory by additive and multiplicative shifts, respectively. The normalization parameter  $N_e$  is usually associated with the uncertainty in measuring the luminosity. The parameters  $r_{e,k}$  control the amount of additive shifts of the data introduced by the point-by-point correlated systematic uncertainties  $\beta_{e,k,i}$  of type  $k$ . While the values of  $r_{e,k}$  can be fitted, in practice they are computed analytically *via*  $\partial\chi^2(\mathbf{a}, \text{data})/\partial r_{e,k} = 0$  [87], which reduces the number of fitting parameters and allows more efficient computation of the  $\chi^2$ . Gaussian penalties for  $r_{e,k}$  are included with  $\sum_k r_{e,k}^2$ , and for  $N_e$  we use a penalty with the quoted experimental normalization uncertainty  $\delta \mathcal{N}_e$ .

### 4.1.2 Multi-step strategy

In the JAM global QCD analysis, a “multi-step strategy” is implemented in order to efficiently scan the parameter space by sequentially including a subset of the datasets in each step. The minimization algorithm starts with prior samples (initial guess parameters) drawn from a flat distribution within the physical region of the parameters, and uses the flat distribution to obtain the posteriors for the selected datasets. These posteriors are then used as priors for the next iteration, where the data from the previous step are augmented with addition of new datasets. This procedure is repeated until all the datasets have been included. This strategy is designed to increase the efficiency of scanning the parameter

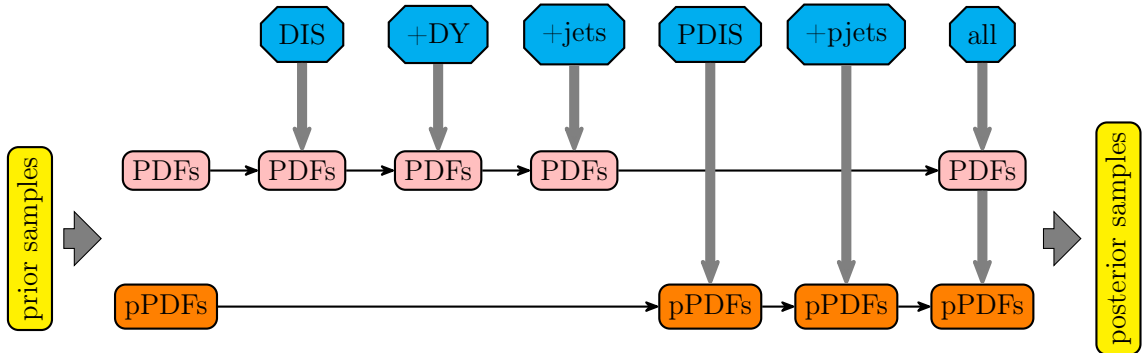


FIG. 4.1: Illustration of multi-step strategy, where “pPDFs” means the polarized PDFs.

space by avoiding feeding too many datasets into the fitter at once.

While different approaches can be employed in designing which datasets are utilized at each step, our strategy is based on first including datasets that have less numerical complexity, and systematically adding those that are more numerically expensive. At the same time, in the early steps, we include only the datasets that capture the gross features of the underlying physics, and restrict the flexibility of the PDFs by only fitting with the parameters  $a_0$ ,  $a_1$  and  $a_2$  in the template (3.1). In later steps, those datasets that require finer details such as flavor separation will be included, and the parameters  $a_3$  and  $a_4$  released, if applicable.

Employing this strategy, in this work we fit first the spin-averaged PDFs using fixed target DIS data, after which HERA collider DIS data are included, followed by Drell-Yan, and inclusive jet production in hadronic collisions from the Tevatron and RHIC. At this point, the parameters of the spin-averaged PDFs are fixed, and spin-dependent PDFs are then fitted using first the polarized DIS data, then RHIC jet data from polarized  $pp$  scattering. At the final step, both spin-averaged and spin-dependent PDFs are released and fitted simultaneously with all the data from previous steps. A summary of the multi-step strategy is shown in Fig. 4.1.

## 4.2 Monte Carlo technique

The prior samples in Fig. 4.1 are drawn by Monte Carlo random sampling, respecting the parameter ranges summarized in Tables 3.1 and 3.2. Starting from  $N$  randomly drawn points in the parameter hyperspace, we propagate them through multiple steps to find the best description of the observables by minimizing  $\chi^2$ .

For each  $\chi^2$  minimization of the  $N$  samples, the data are distorted by adding a Gaussian noise within the quoted uncorrelated uncertainties,

$$\tilde{\sigma}_i = \sigma_i + R_i \alpha_i, \quad (4.6)$$

where  $\sigma_i$  is the  $i$ th experimental data point,  $\alpha_i$  is its corresponding quadrature sum of all uncorrelated uncertainties, and  $R_i$  is a random number distributed with probability density  $e^{-R_i^2/2}/\sqrt{2\pi}$ . The confidence region of the parameters is estimated by drawing the posterior distribution from the aforementioned Monte Carlo random samples (prior distribution), using the Gaussian distorted data.

From the ensemble of parameters  $\{\mathbf{a}\}$  of dimension  $N$  drawn from the posterior distribution, one can compute the expectation values and variances for any generic observable  $\mathcal{O}$  (either a PDF at a given  $x$  and  $Q^2$ , or a cross section computed as a function of PDFs),

$$\begin{aligned} \mathrm{E}(\mathcal{O}) &= \frac{1}{N} \sum_{k=1}^N \mathcal{O}(\mathbf{a}_k), \\ \mathrm{V}(\mathcal{O}) &= \frac{1}{N} \sum_{k=1}^N (\mathcal{O}(\mathbf{a}_k) - \mathrm{E}(\mathcal{O}))^2, \end{aligned} \quad (4.7)$$

where the variance gives the  $1\sigma$  confidence interval for the observable  $\mathcal{O}$ . The Bayesian “master formulas” (4.7) provide the most robust determination of PDF uncertainties available within the global QCD analysis paradigm, without the need for introduction of additional *ad hoc* prescriptions, such as tolerance factors, which are sometimes employed in single-fit analyses to account for tensions between datasets [88].

### 4.3 Optimization

In order to perform the fit efficiently, we employ an additional method to maximally optimize the fitting process, as described in the following.

Based on the observation that the cross section values calculated at a sufficient number of points can be interpolated to find the values at the uncalculated points, we implement the interpolation.

Firstly, a grid is chosen carefully such that it covers the experimental kinematics with sufficient density, while still having a number  $\lesssim 50\%$  of the experimental points. Then the observables are only computed at the points in that grid and the computed values are used to interpolate the actual experimental values.

For unpolarized DIS, DY and polarized DIS, we sample kinematic points in the  $\ln x$ - $\ln Q^2$  plane by distributing them relatively evenly, computing the cross sections at those sampled points during each fit, and using interpolation to determine the cross sections at the experimental kinematic points. We also make sure that the experimental kinematics are contained within the sampled grid to avoid complications with extrapolation.

As for unpolarized jet observables, however, the cross sections can extend many ( $\sim 8$ ) orders of magnitude and change rapidly for Tevatron kinematics [66, 67], so that interpo-



lating the whole kinematic plane is impractical. Instead, we compute the cross sections at every other  $p_T$  point within each rapidity bin, and interpolate the remaining points. Again, to avoid complications of extrapolation, we make sure that the lowest and highest  $p_T$  points are always computed. For polarized jet observables, we use the same method as for unpolarized jets to interpolate  $A_{LL}$ .

The interpolation reduces the computation time by roughly 70% for DIS and DY and 40% for unpolarized and polarized jets, while only introducing errors that are one to several orders of magnitude smaller than the experimental uncorrelated uncertainties.

## 4.4 Experimental data

In this section, we will discuss the treatment of experimental data, including kinematic cuts and uncertainties.

### 4.4.1 Unpolarized datasets

The kinematic coverage of unpolarized datasets is shown in Fig. [4.2](#).

For unpolarized inclusive DIS, we apply the cuts  $W^2 > 10 \text{ GeV}^2$  and  $Q^2 > m_c^2$ , where  $m_c \approx 1.27 \text{ GeV}$  is the mass of charm quark, in order avoid power corrections, as well as offshell effects which are expected to be important only at high  $x$ . With the above cuts, we include proton and/or deuteron  $F_2$  measurements from BCDMS [\[35\]](#), SLAC [\[36\]](#), and NMC [\[37\]](#), as well as  $F_2^d/F_2^p$  measurement from NMC [\[38\]](#). With the same cuts, we also include the reduced proton neutral current and charged current cross sections from the combined H1 and ZEUS analysis at HERA [\[39\]](#).

The treatment of the correlated uncertainties for DIS datasets is straightforward. The

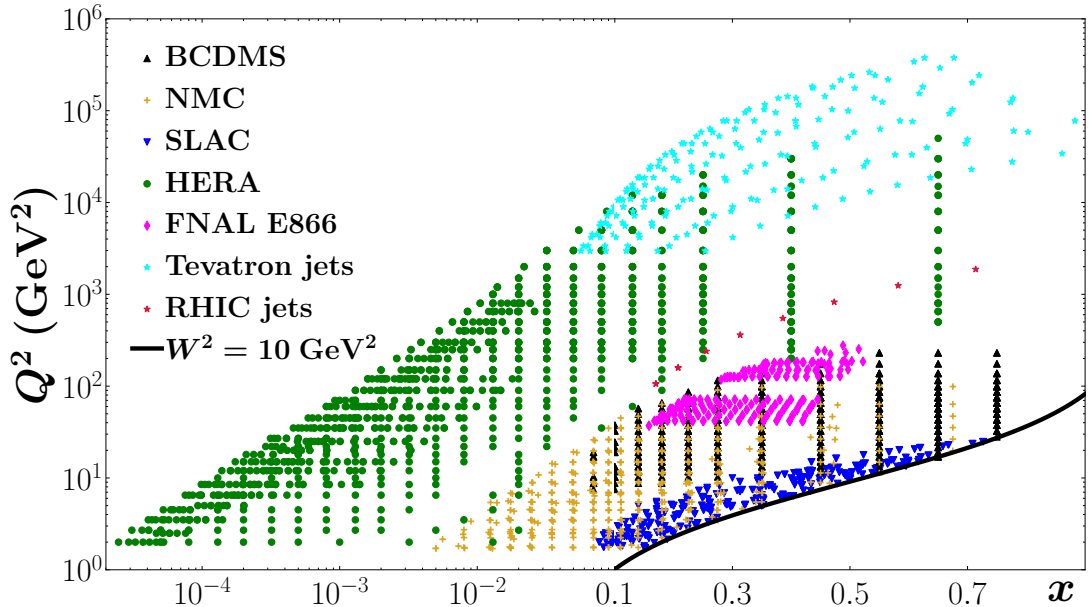


FIG. 4.2: Kinematic coverage of the unpolarized datasets used in this analysis, including unpolarized fixed-target DIS from BCDMS (black upward triangles), NMC (golden  $\times$ ) and SLAC (blue downward triangles), DIS collider data from HERA (green solid circles), DY data from  $pp$  and  $pd$  collisions at Fermilab (magenta diamonds), and jet production in unpolarized  $p\bar{p}$  scattering at the Tevatron (cyan stars) and  $pp$  scattering at RHIC (crimson stars). The variable  $x$  here denotes Bjorken- $x$  for DIS and Feynman- $x$  for DY and jet production, while the scale  $Q^2$  represents the four-momentum transfer squared for DIS and DY, and transverse momentum squared for jets. The solid curve represents the boundary for the maximum  $x$  attainable at fixed  $W^2 = M^2 + Q^2(1-x)/x = 10 \text{ GeV}^2$ .

fixed target data also provides normalization uncertainties, which are included in the fits as in Eq. (4.5).

As for DY data, a cut of  $Q^2 > 36 \text{ GeV}^2$  is applied to avoid the data points that have been found to contradict DIS data, as discussed in Ref. [89]. For the fit, we include  $pp$  and  $pd$  Drell-Yan data from the Fermilab E866 experiment [65], which has the  $J/\psi$ ,  $\psi'$  and  $\Upsilon$  resonance families already removed using dimuon mass cuts.

For jet production in unpolarized collisions, the data included are from D0 [66] and CDF [67] collaborations at Fermilab, and STAR [22] at RHIC.

The treatment of correlated, uncorrelated, and normalization uncertainties is as illustrated in Eq. (4.5). However, the CDF data provide uncertainties that are only correlated within each rapidity ( $\eta$ ) bin. To account for this, we split such uncertainties into multiple columns, with each column only having nonzero entries for one of the  $\eta$  bins.

Apart from the uncertainties, both CDF and D0 provide parton-to-hadron correction factors  $C_{p \rightarrow h}$  which are obtained from Monte Carlo simulations [67]. These factors correct the perturbative calculation, which is at the partonic level, to the hadronic level, and are applied after the evaluation of the cross sections.

The RHIC unpolarized jet data were taken in different years (2003 and 2004), and are therefore included as different datasets, with names STAR 2003 and STAR 2004.

#### 4.4.2 Polarized datasets

The kinematic coverage of polarized datasets is shown in Fig. 4.3.

For polarized inclusive DIS, we apply the same cuts  $W^2 > 10 \text{ GeV}^2$  and  $Q^2 > m_c^2$  as for the unpolarized case. With the above cuts, we include fixed-target polarized DIS proton, deuteron and helium  $A_1$  measurements from EMC [41], SMC [42, 43], COMPASS [44–46], SLAC [47–52], and HERMES [53, 54].

For the jet polarization asymmetries  $A_{LL}$ , the data included are from the STAR [22–26] and PHENIX [27] collaborations at RHIC.

As for the unpolarized data, the general treatment of correlated, uncorrelated, and normalization uncertainties is illustrated in Eq. (4.5). For the STAR data, we use from [23] the knowledge that an uncertainty in the relative luminosity measurement usually results in a shift of the  $A_{LL}$  data by an additive constant (fully correlated systemic uncertainty),

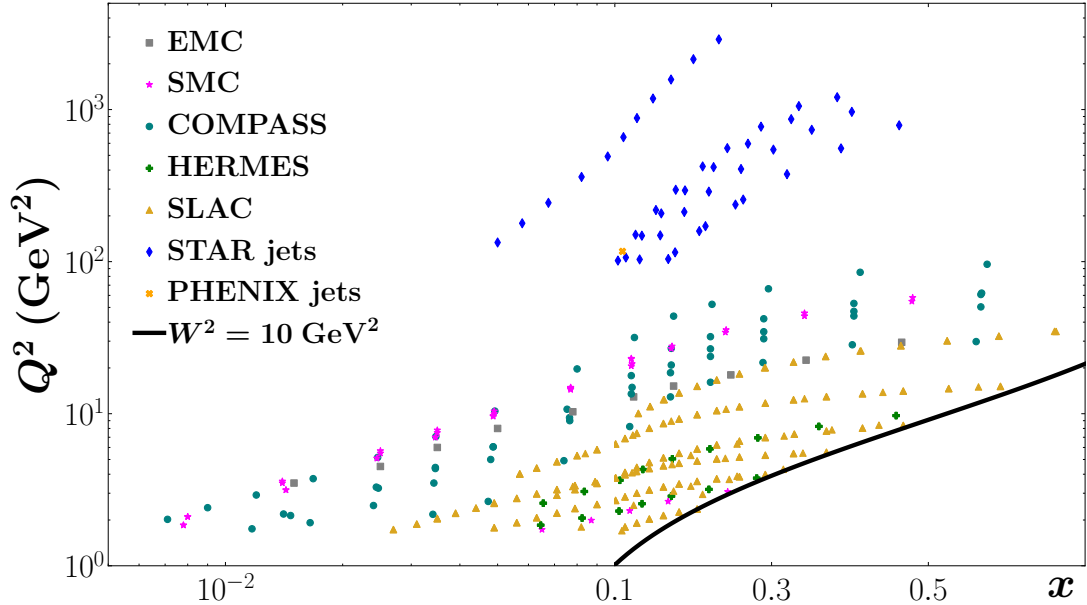


FIG. 4.3: Kinematic coverage of the polarized datasets used in this analysis, including spin-dependent fixed-target DIS from EMC (gray solid squares), SMC (magenta stars), COMPASS (teal solid circles), HERMES (green  $\times$ ) and SLAC (golden upward triangles), and jet production in polarized  $pp$  scattering from STAR (blue diamonds) and PHENIX (orange  $\times$ ). The variable  $x$  here denotes Bjorken- $x$  for DIS and Feynman- $x$  for jet production in  $pp$  collisions, while the scale  $Q^2$  represents the four-momentum transfer squared for DIS and transverse momentum squared for jets. The solid curve represents the boundary for the maximum  $x$  attainable at fixed  $W^2 = M^2 + Q^2(1 - x)/x = 10 \text{ GeV}^2$ .

and an uncertainty in the measurement of the polarization magnitude scales the magnitude of the  $A_{LL}$  data (normalization uncertainty). We also treat the uncertainties that are only weakly correlated as uncorrelated, so as not to underestimate the uncertainties, in particular, for the STAR 2012 [23], 2015 [24] and 2021 [26] data.

#### 4.4.3 EIC observables

The expected kinematic coverage of the EIC observables, compared with the available polarized world data, is shown in Fig. 4.4. As the world's first polarized lepton-hadron (and lepton-nucleus) collider, the EIC will extend the kinematic coverage in Bjorken- $x$  down to  $x \approx 10^{-4}$ , and in  $Q^2$  up to  $Q^2 \approx 10^3 \text{ GeV}^2$ . The measurement of the polarization asym-

metry  $A_{LL}$  at this facility will give us access to the  $g_1$  structure function at unprecedented low values of  $x$ , and thus reduce uncertainties in spin-dependent PDFs at small parton momentum fractions. Furthermore, the wider  $Q^2$  coverage will allow scaling violations in the  $g_1$  structure function to be determined more precisely, from which improved constraints can be derived on the spin-dependent gluon distribution.

Furthermore, access to polarized deuteron and  $^3\text{He}$  beams will allow separation of the helicity into individual quark flavors, and significantly reduce the uncertainties on the total helicity carried by quarks,  $\Delta\Sigma$ , compared to proton data alone, which are mostly sensitive to the  $u$  quark polarization.

With its high luminosity and hadron polarization compared to HERA [56], the EIC will also be able to access entirely new observables, such as the parity-violating asymmetry  $A_{UL}$ . This observable, as discussed earlier in Section 2.1.3, can provide new linear combinations of helicity PDFs that would enhance the flavor separation capabilities, in particular for strange quark helicity distribution [57].

### Baseline PDFs

The EIC data, unlike other available world data, have to be simulated from a baseline of known PDFs. Therefore to begin with, we first need to obtain the unpolarized structure functions  $F_1$  and  $F_2$  which appear in the denominators of the polarization asymmetries in Eqs. (2.21) and (2.24). To this end, we follow steps 1 to 5 in Table 4.1, where the multi-step strategy described in Section 4.1.2 is employed. For DIS fixed-target and HERA data, cuts  $W^2 = M^2 + Q^2(1-x)/x > 10 \text{ GeV}^2$  and  $Q^2 > m_c^2$  are applied; for DY a cut  $Q^2 > 36 \text{ GeV}^2$  is applied; and for single jet production data a jet transverse momentum cut  $p_T > 10 \text{ GeV}$  is used. From the fitted spin-averaged PDFs we calculate the  $F_1$  and  $F_2$  structure functions, which are then kept fixed throughout the rest of the analysis of the

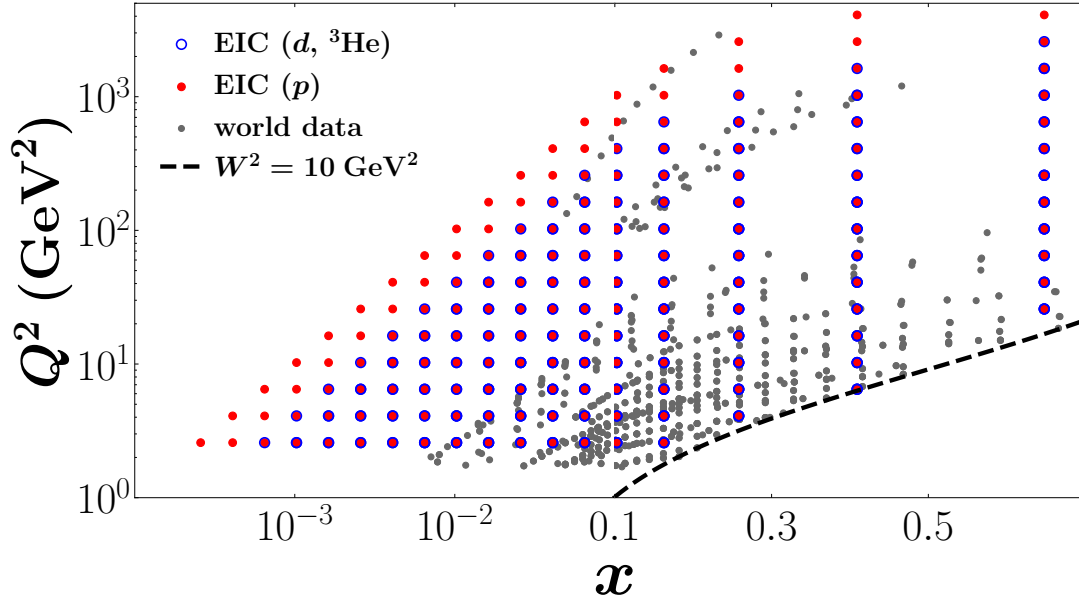


FIG. 4.4: Kinematic coverage of data sets used in the EIC impact study [57], including the available world data (gray solid circles) from Fig. 4.3, and projected EIC data for polarized protons (red solid circles), deuterons and  $^3\text{He}$  (blue open circles). The variable  $x$  here denotes Bjorken- $x$  for DIS and Feynman- $x$  for jet production in  $pp$  collisions, while the scale  $Q^2$  represents the four-momentum transfer squared for DIS and transverse momentum squared for jets. The dashed curve represents the boundary for the maximum  $x$  attainable at fixed  $W^2 = M^2 + Q^2(1-x)/x = 10 \text{ GeV}^2$ . (Figure from [57].)

spin-dependent data for the EIC impact study.

To obtain a baseline for the helicity PDFs, we follow steps 6 to 7 in Table 4.1. Same cuts on  $Q^2$  and  $W^2$  for the polarized DIS datasets, and on jet transverse momentum  $p_T$  for the polarized jet datasets, are applied as for the unpolarized case. For the SU(2) and SU(3) flavor symmetries, we consider two scenarios for  $A_{LL}$  and  $A_{UL}$  by imposing only the SU(2) constraint, and also the SU(3) constraint in Eq. (3.6), with  $g_A = 1.269 \pm 0.003$  and  $a_8 = 0.586 \pm 0.031$  being the triplet and octet axial vector charges, respectively [76]. The baseline spin-averaged and spin-dependent PDFs determined from the global fit to the existing data are then used to simulate the impact of the observables at the EIC, assuming various scenarios for the theoretical inputs.

step	process	particle	$N_{\text{dat}}$	PDF	$\chi^2/N_{\text{dat}}$	
					SU(2)	+SU(3)
1	DIS fixed target [35–38]	$p, d$	1495	$f$	1.13	
2	+ DIS HERA [39]	$p$	1185	$f$	1.30	
3	+ DY [65]	$pp, pd$	250	$f$	1.05	
4	+ Tevatron jet [66, 67]	$p\bar{p}$	186	$f$	0.96	
5	+ RHIC jet [22]	$pp$	10	$f$	1.24	
6	pDIS [41–54]	$p, n, d, {}^3\text{He}$	365	$\Delta f$	0.93	0.93
7	+ polarized jet [22–25, 27]	$pp$	45	$\Delta f$	0.75	0.76
<b>Total</b>			<b>3552</b>		<b>1.15</b>	<b>1.15</b>

TABLE 4.1: Summary of  $\chi^2$  values and number of data points  $N_{\text{dat}}$  for the various processes used in obtaining the baseline for the EIC impact study. The  $\chi^2$  shown for the unpolarized fits (steps 1 to 5) are acquired at the final step (step 5), similarly for the polarized fits (steps 6 and 7), where the  $\chi^2$  at step 7 are shown in the table. The two scenarios, SU(2) only and plus SU(3), are available, of course, for the polarized fits only. The  $\chi^2$  is 0.024 for fitting  $g_A$  in the SU(2) only scenario, and are 0.000 and 0.009 for fitting  $g_A$  and  $a_8$  respectively in the plus SU(3) scenario. The STAR 2015 single jet production data [26] is not included because the impact study is done before this data is available to public.

### Estimation of statistical and systematic uncertainties

The absolute statistical uncertainties for the polarization asymmetries are determined according to

$$\delta A \approx \frac{1}{\sqrt{\mathcal{L}\sigma_{\text{unp}}}}, \quad (4.8)$$

where  $\mathcal{L}$  is the estimated integrated luminosity for the specific process, and  $\sigma_{\text{unp}}$  is the integrated unpolarized cross section in each  $(x, Q^2)$  bin. This approximation is valid as long as the asymmetries are  $\ll 1$ . Assuming that the cross sections can be considered

constant in each  $(x, Q^2)$  bin, the integrated unpolarized cross section can be written as

$$\sigma_{\text{unp}} = \int_{\text{bin}} dx dQ^2 (\sigma^{\uparrow\uparrow} + \sigma^{\downarrow\uparrow}) \approx \Delta x \Delta Q^2 \frac{8\pi\alpha^2}{Q^2 s x^2} \left( xy F_1 + \frac{1-y}{y} F_2 \right), \quad (4.9)$$

where  $\Delta x$  and  $\Delta Q^2$  are the intervals of the bins.

In the case of the  $A_{LL}$  asymmetry, we assume a 2% uncorrelated systematic uncertainty from the pion background, independent of the region of kinematics [90].

Since the predictions for  $A_{LL}$  are based on the extrapolation of existing measurements that are only available for  $x \gtrsim 0.01$ , we consider three possible scenarios, which we denote by “low”, “mid” and “high”, to better assess the effect of extrapolation on the EIC pseudodata impact. The high and low pseudodata sets are generated by shifting the values of  $A_{LL}$  in the unmeasured region by  $\pm 1\sigma$  CL, estimated from existing helicity PDF uncertainties, while the mid set is generated using the central predictions. For each dataset, the uncertainties are shifted in the same way as the observables.

For the  $A_{LL}$  asymmetry, we consider the center of mass energies  $\sqrt{s} = 29, 45, 63$  and 141 GeV for a proton beam with an assumed integrated luminosity of  $100 \text{ fb}^{-1}$ , while for deuteron and  $^3\text{He}$  beams we include  $\sqrt{s} = 29, 66$  and 89 GeV and assume  $10 \text{ fb}^{-1}$  of integrated luminosity. Projected  $A_{LL}$  data and their uncertainties are shown in Fig. 4.5 for the mid scenario for proton, deuteron and  $^3\text{He}$  beams. For the high (low) case that are not shown in the figure, the small- $x$  region of the asymmetry will be shifted slightly upwards (downwards) by  $\pm 1\sigma$  CL. The systematic uncertainties follow the shape of the asymmetry, since they are estimated as a flat 2% error. The statistical uncertainties are similar for the high and low cases, as  $\sigma_{\text{unp}}$  in Eq. (4.8) is well-constrained down to  $x \sim 10^{-4}$ .



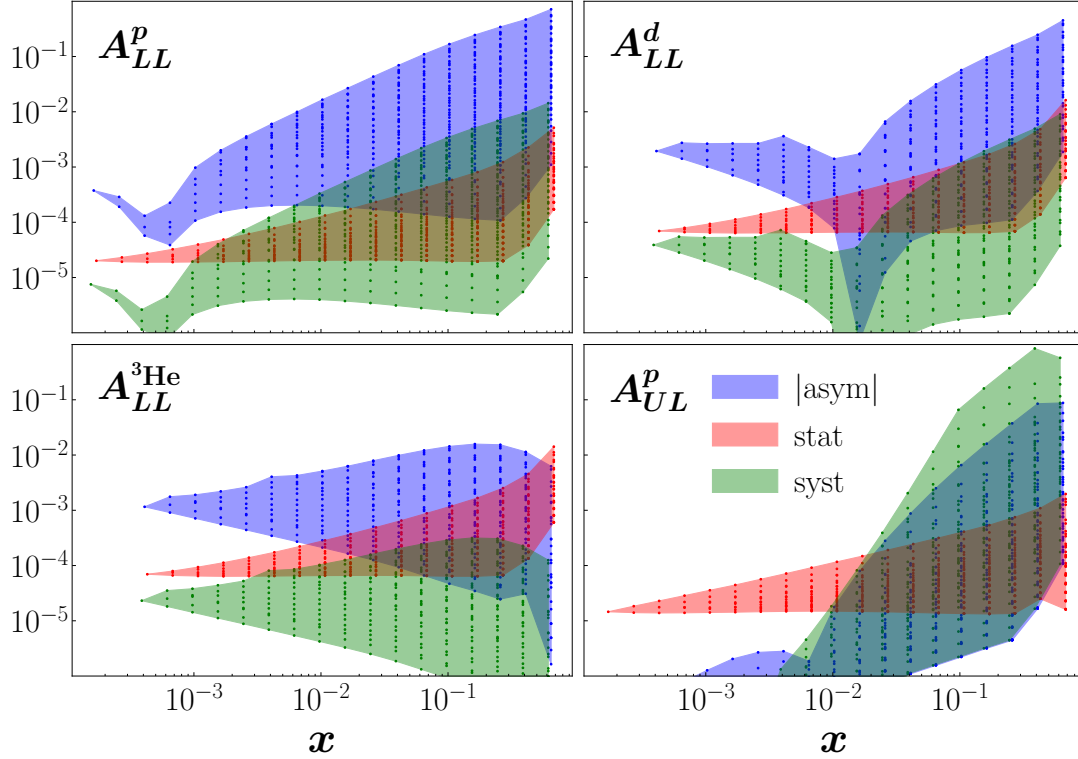


FIG. 4.5: Simulated absolute values of the longitudinal double-spin asymmetry  $A_{LL}$  for proton, deuteron, and  $^3\text{He}$  beams and the parity-violating asymmetry  $A_{UL}$  with a proton beam at the EIC (blue bands), using the “mid” predictions with both SU(2) and SU(3) assumptions from the baseline PDFs, along with estimated statistical (red bands) and uncorrelated systematic (green bands) uncertainties. (Figure from [57].)

For the parity-violating  $A_{UL}$  asymmetry, we use the values given in Table 4.2 for the predicted systematic uncertainties from the pion background, which are dependent on the electron beam energy  $E$  and pseudo-rapidity  $\eta = \ln(x\sqrt{s}/Q)$ . We consider the low, mid and high scenarios, as for the  $A_{LL}$  asymmetry, and include only proton beam data at center of mass energies  $\sqrt{s} = 29, 45, 63$  and  $141$  GeV, with an assumed integrated luminosity of  $100 \text{ fb}^{-1}$ . The absolute values of the proton parity-violating asymmetry and the corresponding uncorrelated statistical and systematic errors are shown in the lower right panel of Fig. 4.5. Note that estimates of correlated systematic uncertainties are not included in the EIC impact study. Potential overall normalization errors will not affect the

analysis, as the pseudodata are generated using the baseline PDFs described in previous section.

$\Delta\eta$	$E = 18 \text{ GeV}$	$E = 10 \text{ GeV}$	$E = 5 \text{ GeV}$
$(-3.5, -2.0)$	0.02	$10^{-3}$	$10^{-5}$
$(-2.0, -1.0)$	0.8	0.4	0.1
$(-1.0, 0.0)$	1	8	5
$(0.0, 1.0)$	10	10	10

TABLE 4.2: Relative uncorrelated systematic uncertainties for  $A_{UL}$  from the pion background, for various electron beam energies  $E$  and pseudo-rapidity intervals  $\Delta\eta$  [90].

# CHAPTER 5

## Results from JAM global QCD analysis

Using the Bayesian approach and Monte Carlo framework, we have successfully performed the first simultaneous extraction of spin-averaged and spin-dependent PDFs from available deep-inelastic scattering, Drell-Yan and single jet observables. First a fit with unpolarized data from DIS, DY and jet data was performed, providing the foundation for further analysis of polarized observables, which are mostly asymmetries and involve unpolarized cross sections in their denominators. Along with fitting unpolarized data, we have carefully examined the impact that unpolarized jet data can have on the PDFs, where the RHIC unpolarized jet data [22] were included for the first time ever in a global QCD analysis. Following this, we analyzed the impact that various theory inputs, including SU(2) and SU(3) flavor symmetries and positivity constraints, can have on the helicity distributions of quarks and gluons. Finally, a series of simultaneous analyses were performed with the first consistent extraction of PDFs in a helicity basis.

As an application of the above results, we also performed an EIC impact study using the

baseline PDFs from the JAM analysis. By critically examining the low- $x$  extrapolation of helicity PDFs in the unmeasured region, we assessed the impact that the EIC parity-conversing and parity-violating asymmetries can have on our current knowledge of helicity PDFs with adequately minimized bias.

## 5.1 Analysis of spin-averaged PDFs

Applying the multi-step strategy illustrated in Fig. 4.1 for the unpolarized fits we developed a set of 5 steps that fit our practical needs (Table 5.1). Firstly, we start with DIS fixed target data, which carves out a rough shape for the PDFs. In the second step, the DIS collider data from HERA are added to further constrain the uncertainties of the PDFs, particularly at low  $x$ . The third step then adds DY data that provide constraints on the flavor separation of PDFs. Lastly, we include the jet data in two steps (first Tevatron and then RHIC jets), because as the first global QCD fit to include the RHIC unpolarized jet data, we would like to determine what impact the RHIC unpolarized jet data have on the PDFs.

The resulting  $\chi_{\text{dat}}^2 \equiv \chi^2/N_{\text{dat}}$  ( $N_{\text{dat}}$  is the number of data points) values for the unpolarized world data are listed in Table 5.2. Overall, the fit gives a reasonable  $\chi_{\text{dat}}^2$  of 1.18 for a total of 3,130 data points. The agreement of the fit with each of the experimental datasets

step	DIS fixed target	DIS HERA	DY	Tevatron jets	RHIC jets
1	✓				
2	✓	✓			
3	✓	✓	✓		
4	✓	✓	✓	✓	
5	✓	✓	✓	✓	✓

TABLE 5.1: Steps in JAM global QCD analysis for unpolarized fits

is further illustrated in Figs. 5.1 to 5.4 for DIS, Fig. 5.5 for DY and Fig. 5.6 for jet observables. Afterwards, we will discuss the spin-dependent PDFs as well as the impact of jet observables.

experiment	reference	observable	reaction	$N_{\text{dat}}$	$\chi^2/N_{\text{dat}}$
NMC	37	$F_2$	$\mu p$	273	1.68
NMC	38	$F_2^d/F_2^p$	$\mu p, \mu d$	174	0.97
SLAC	36	$F_2$	$e^- p$	218	0.96
SLAC	36	$F_2$	$e^- d$	228	0.74
BCDMS	35	$F_2$	$\mu p$	348	1.16
BCDMS	35	$F_2$	$\mu d$	254	1.10
HERA (318 GeV)	39	$\sigma_r^{\text{NC}}$	$e^+ p$	402	1.52
HERA (318 GeV)	39	$\sigma_r^{\text{NC}}$	$e^- p$	159	1.63
HERA (318 GeV)	39	$\sigma_r^{\text{CC}}$	$e^+ p$	39	1.34
HERA (318 GeV)	39	$\sigma_r^{\text{CC}}$	$e^- p$	42	1.02
HERA (300 GeV)	39	$\sigma_r^{\text{NC}}$	$e^+ p$	75	1.12
HERA (251 GeV)	39	$\sigma_r^{\text{NC}}$	$e^+ p$	259	1.01
HERA (225 GeV)	39	$\sigma_r^{\text{NC}}$	$e^+ p$	209	1.09
E866/NuSea	65	$M^3 d^2\sigma / dM dx_F$	$pp$	121	1.15
E866/NuSea	65	$M^3 d^2\sigma / dM dx_F$	$pd$	129	0.90
D0	66	$d^2\sigma / d\eta dp_T$	$p\bar{p}$	110	0.89
CDF	67	$d^2\sigma / d\eta dp_T$	$p\bar{p}$	76	1.11
STAR (2003)	22	$d^2\sigma / 2\pi d\eta dp_T$	$pp$	5	4.94
STAR (2004)	22	$d^2\sigma / 2\pi d\eta dp_T$	$pp$	9	1.05
<b>total</b>				<b>3,130</b>	<b>1.18</b>

TABLE 5.2: Inclusive DIS, DY and single jet production data used in this analysis, indicating the observables fitted, the reactions, number of data points in each experiment ( $N_{\text{dat}}$ ) and respective  $\chi^2/N_{\text{dat}}$  values. The  $\chi^2$  are computed by Eq. (4.5) using the average of theory predictions from all the replicas. The quantities in the parentheses of HERA experiments are the values of  $\sqrt{s}$ , center of mass energies. Observables  $\sigma_r^{\text{NC}}$  and  $\sigma_r^{\text{CC}}$  represent the reduced neutral-current and charged-current cross sections respectively. STAR (2003/2004) means the unpolarized single jet production data taken by STAR collaboration at RHIC in the year 2003/2004.

The  $\chi_{\text{dat}}^2$  values shown in Table 5.2, as well as the fits to data in Figs. 5.1 to 5.6 below, are obtained from the last step (fifth step) of the unpolarized fit.

Following the steps in Table 5.1, unpolarized DIS fixed target data are the first to be fitted, which include the proton  $F_2$  and the deuteron to proton ratio  $F_2^d/F_2^p$  data from NMC [37], as well as the proton and deuteron  $F_2$  data from SLAC [36] and BCDMS [35].

Shown in Fig. 5.1 are the fits to proton  $F_2$  data from NMC, SLAC and BCDMS. Overall, they can be described quite well in the JAM global fit, with  $\chi_{\text{dat}}^2$  of less than or slightly above 1. Only the NMC proton  $F_2$  dataset shows a slightly worse agreement of  $\chi_{\text{dat}}^2 \approx 1.6$ . This lack of good agreement for NMC is also found in other global QCD analyses, such as in the CJ15 [91] and NNPDF3.1 [92] fits.

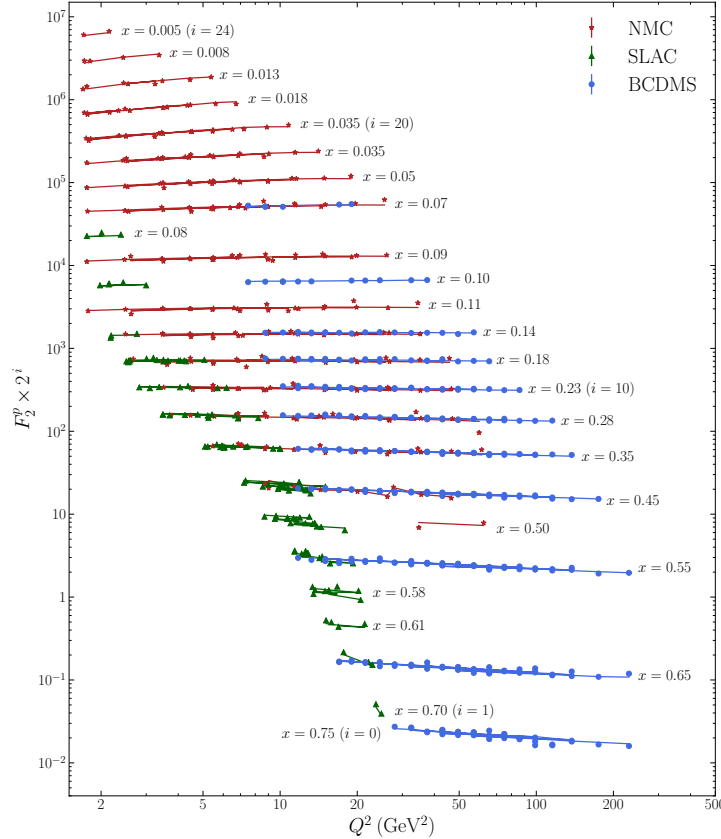


FIG. 5.1: Proton  $F_2$  from the NMC [37] (red stars), SLAC [36] (green upward triangles) and BCDMS [35] (blue solid circles) collaborations. Data points at different  $x$  are multiplied by factors of  $2^i$  to be displayed separately. The data are compared with  $F_2^p$  from the JAM global analysis (solid curves of corresponding colors) that are obtained from the average of all Monte Carlo replicas.

The first step also includes the deuteron  $F_2$  data from SLAC [36] and BCDMS [35], and deuteron to proton ratio  $F_2^d/F_2^p$  data from NMC [38], which are shown in Fig. 5.2. These data are also well described by the JAM global fits, with all  $\chi_{\text{dat}}^2$  less than or slightly above 1.

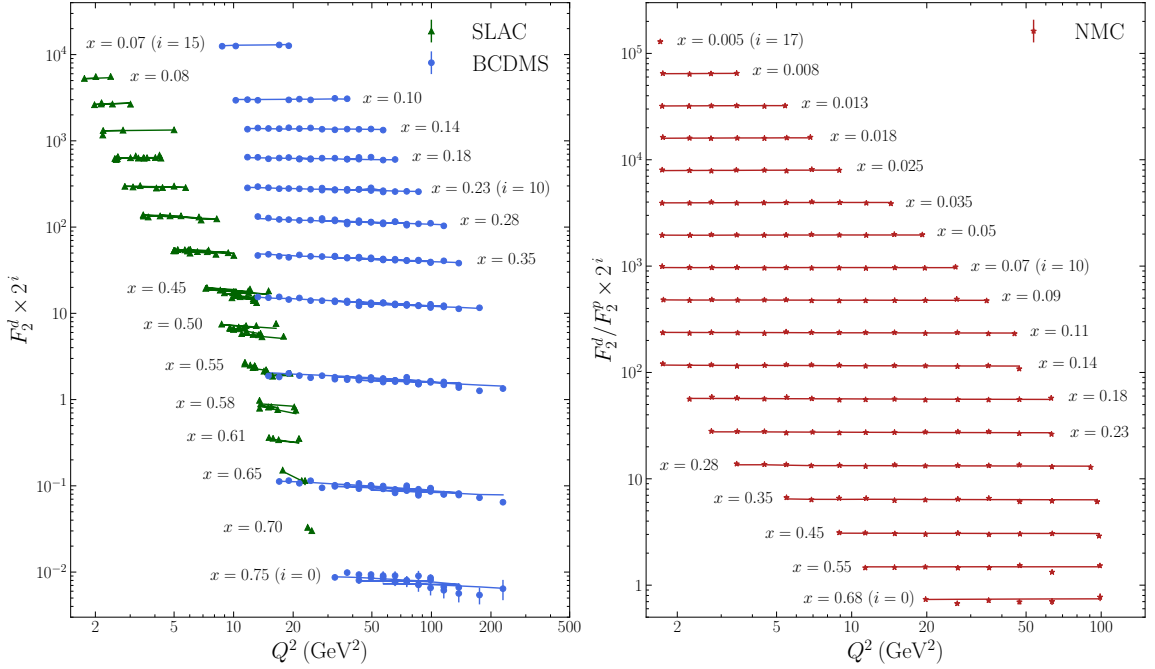


FIG. 5.2: Deuteron  $F_2$  from the SLAC [36] (green upward triangles, left panel) and BCDMS [35] (blue solid circles, left panel), and  $F_2^d/F_2^p$  from NMC [37] (red stars, right panel) collaborations. Data points at different  $x$  are multiplied by factors of  $2^i$  to be displayed separately. The data are compared with  $F_2$  (or  $F_2^d/F_2^p$ ) from the JAM global analysis (solid curves of corresponding colors) that are obtained from the average of all Monte Carlo replicas.

Having acquired an overall good fit in the first step, we proceed to the second step where the HERA collider data [39] are considered.

Shown in Fig. 5.3 are the fits to the reduced neutral-current and charged-current cross sections  $\sigma_r$  from HERA at center of mass energy  $\sqrt{s} = 318$  GeV. Overall, a slightly worse agreement is found for the fit, with  $\chi_{\text{dat}}^2$  roughly between 1.3 and 1.6. The reduced charged-current cross section from  $e^-p$  collisions, on the other hand, exhibits a better agreement,

with  $\chi_{\text{dat}}^2$  only slightly above 1.

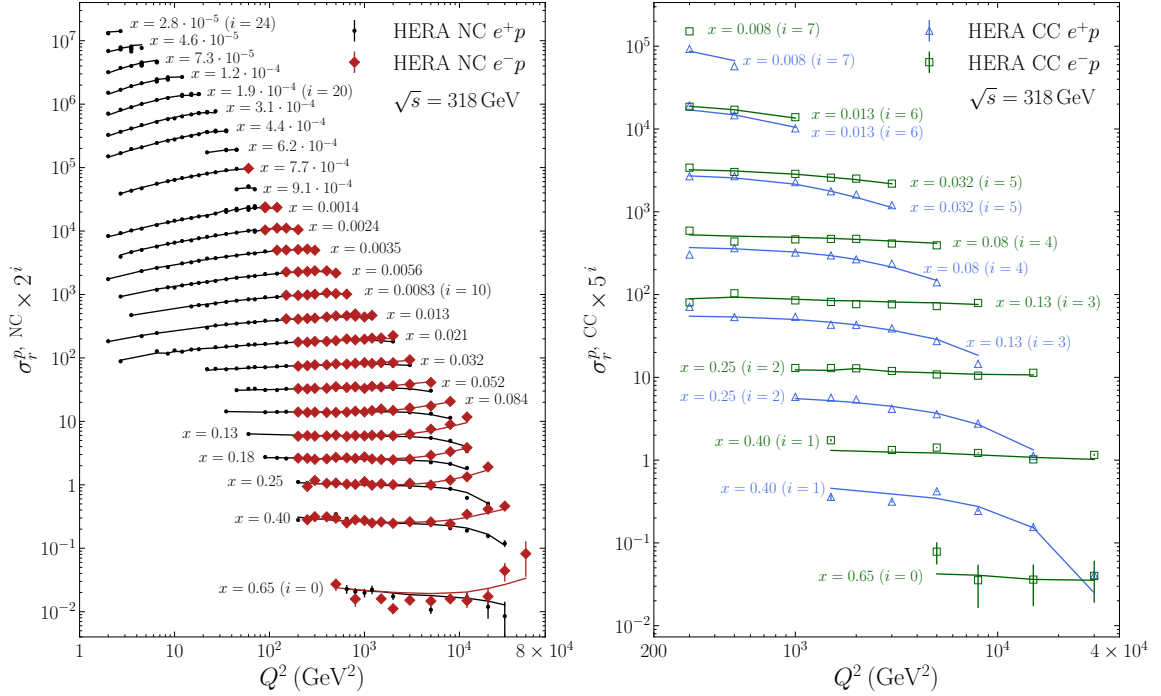


FIG. 5.3: The reduced neutral-current (left panel) and charged-current (right panel) cross sections  $\sigma_r$  from HERA [39] at the center of mass energy  $\sqrt{s} = 318$  GeV. Shown on the left panel are the reduced neutral-current cross sections from  $e^+p$  (black solid dots) and  $e^-p$  (red diamonds) collisions. On the right panel, the reduced charged-current cross sections from  $e^+p$  (blue hollow triangles) and  $e^-p$  (green hollow squares) collisions are shown. Data points at different  $x$  are multiplied by factors ( $2^i$  and  $5^i$  for the left and right panels respectively) to be displayed separately. The data are compared with fits from the JAM global analysis (solid curves of corresponding colors) that are obtained from the average of all Monte Carlo replicas.

Besides the data at center of mass energy  $\sqrt{s} = 318$  GeV, the second step also contains measurements of the reduced neutral-current  $e^+p$  cross sections at  $\sqrt{s}$  equal to 225, 251 and 300 GeV by HERA [39], which are shown in Fig. 5.4. An overall good agreement with the JAM global fits is found in this case, with  $\chi_{\text{dat}}^2$  only slightly above 1.



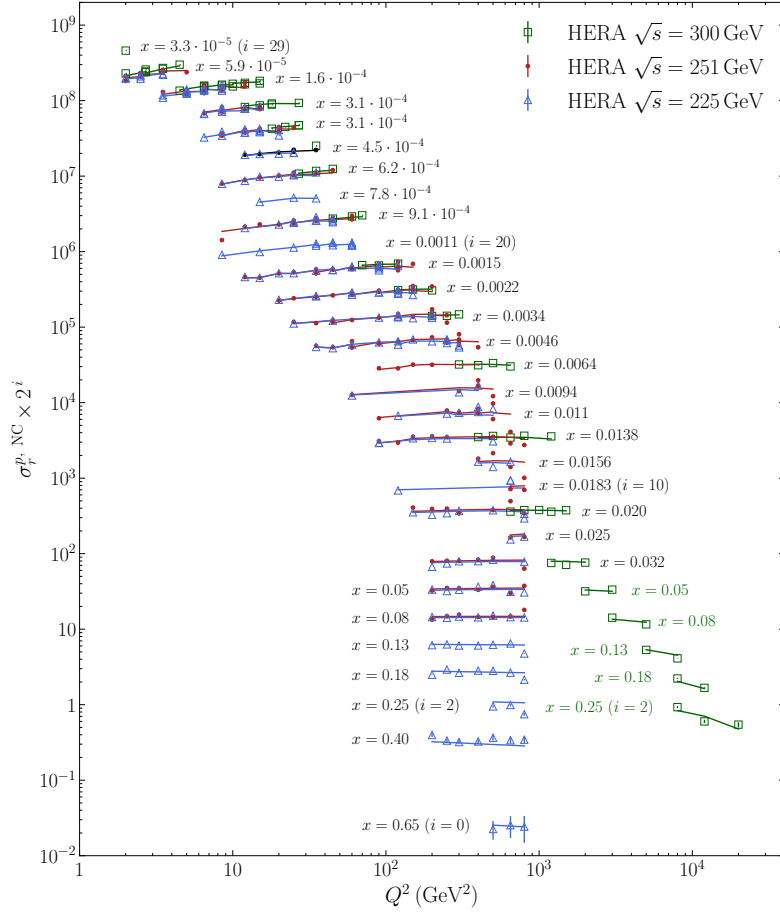


FIG. 5.4: The reduced neutral-current cross section  $\sigma_{\tau}^p$  from HERA [39]  $e^+p$  collisions, with center of mass energies  $\sqrt{s}$  equal to 300 GeV (green hollow squares), 251 GeV (red dots) and 225 GeV (blue hollow triangles). Data points at different  $x$  are multiplied by factors of  $2^i$  to be displayed separately. The data are compared with fits from the JAM global analysis (solid curves of corresponding colors) that are obtained from the average of all Monte Carlo replicas.

In the third step, we add DY cross section data from E866/NuSea [65], which are shown in Fig. 5.5. The JAM global fits are in good agreement with the DY data as well, with  $\chi_{\text{dat}}^2$  less than or slightly above 1.

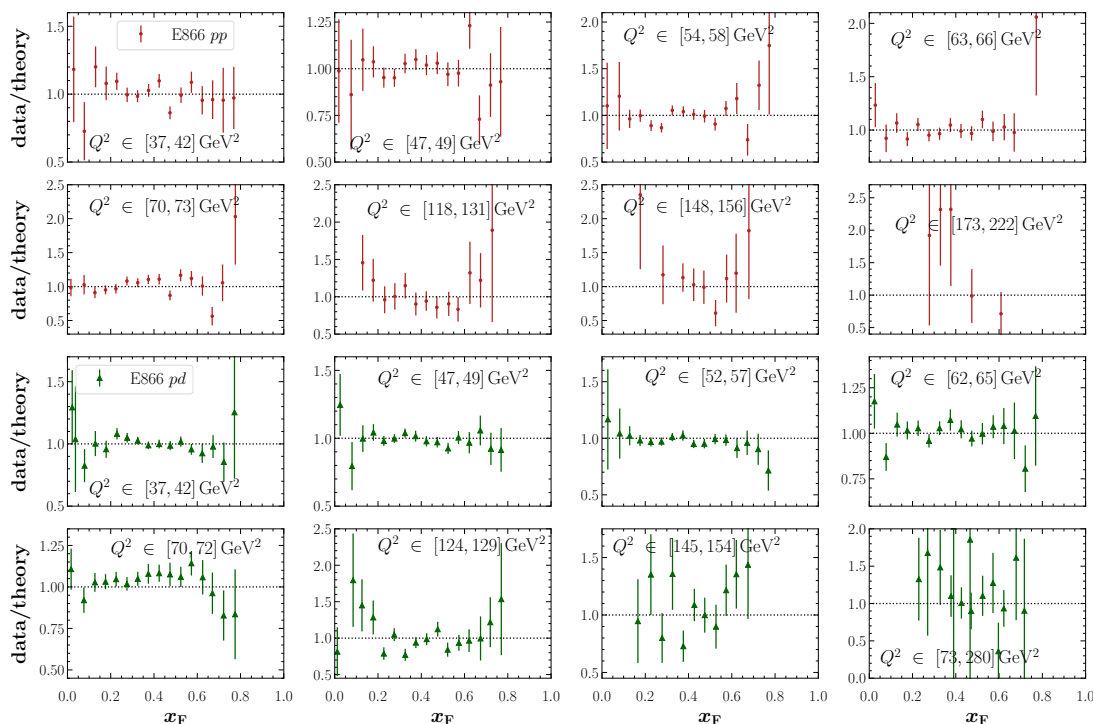


FIG. 5.5: Ratio of data over theory for DY cross sections from E866/NuSea [65], with  $pp$  (red solid dots) and  $pd$  (green upward triangles) collisions. The theory values are obtained from the average of all Monte Carlo replicas, and the error bars indicate the range of the ratio spanned by the uncorrelated uncertainties of the data. The  $Q^2$  ranges (in  $\text{GeV}^2$ ) are indicated for each panel.

One thing that needs to be mentioned is that, after the inclusion of the DY lepton-pair production data, we find that it becomes necessary to release the parameters  $a_3$  and  $a_4$  (see the template function in Eq. (3.1)) for the parametrized distributions  $g$ ,  $u_v$ ,  $d_v$ ,  $\bar{u}_0$  and  $\bar{d}_0$  (see Table 3.1). This improves the  $\chi_{\text{dat}}^2$  for the DY  $pp$  collision data from around 1.3 to 1.1, and in particular for  $pd$  collision data from around 2.3 to 0.9, indicating the sensitivity of the DY data to certain finer details of the PDFs, especially via the  $\bar{u}$  and  $\bar{d}$  distributions.

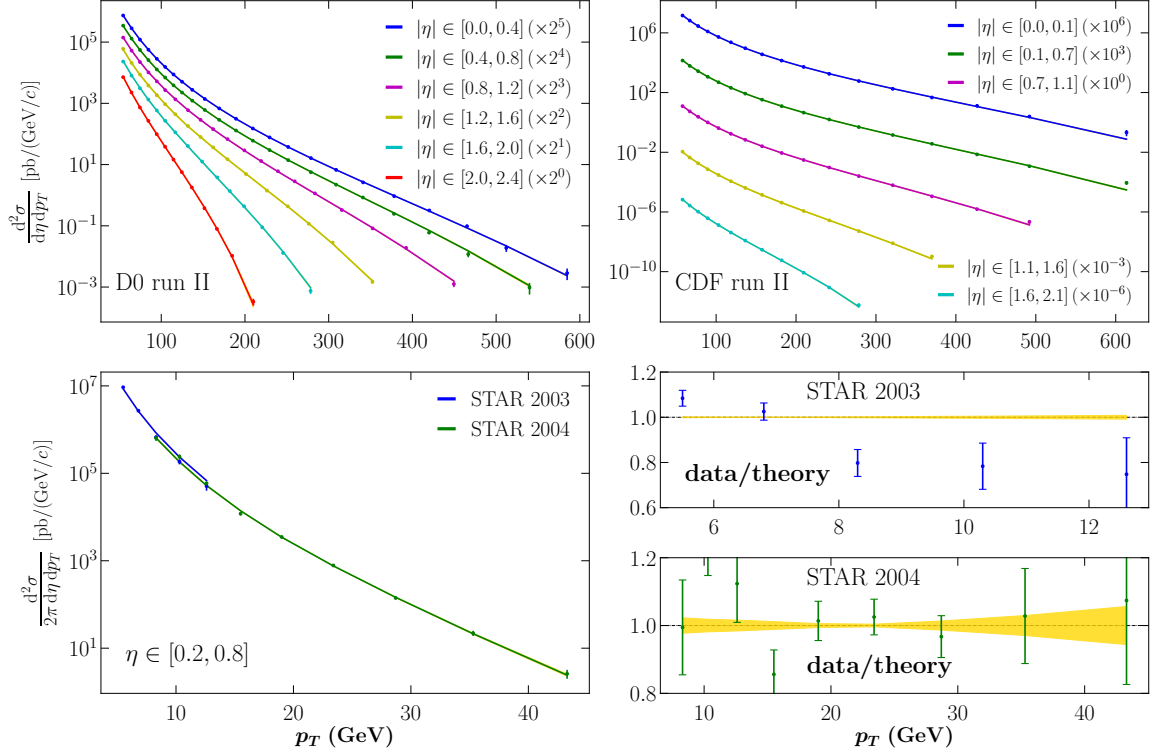


FIG. 5.6: Fit to single jet production cross sections from D0 [66] and CDF [67] at the Tevatron, and STAR [22] at RHIC. Since they have overlapping kinematics, the STAR 2003 and 2004 data are shown in the same panel (bottom left). Different pseudorapidity  $\eta$  bins are marked by colors and scaled by factors for D0 and CDF for clarity. The  $\eta$  bins are obtained using their absolute values for CDF and D0 data, and actual values for STAR. (Note the extra factor  $2\pi$  in the STAR cross section data.) The data (filled circles) are compared with fits from the JAM global analysis (solid curves of corresponding colors) obtained from the average of all Monte Carlo replicas. The theory curves are plotted with yellow bands to show  $1\sigma$  uncertainties, which are barely visible. For STAR 2003 and 2004 data, ratios of data to average theory (filled circles) are shown to illustrate the fit quality, with yellow bands showing the  $1\sigma$  uncertainty of the ratios of theory from each replica and the average theory.

The focus of this thesis is the impact of jet observables, in both unpolarized and polarized collisions, on the spin-averaged and spin-dependent PDFs. For the unpolarized jet observables discussed in this section, the JAM global fit to Tevatron [66, 67] and RHIC [22] jet data is presented in Fig. 5.6.

A good overall agreement with the JAM global fits is found in this case, with  $\chi_{\text{dat}}^2$  only slightly above 1, except for the STAR 2003 data taken at RHIC ( $\chi_{\text{dat}}^2 = 4.94$ ). We find that the significantly larger  $\chi_{\text{dat}}^2$  of the fits to STAR 2003 data is caused by the tension between the lowest- $p_T$  (5.5 and 6.8 GeV) and slightly higher- $p_T$  (8.3 to 12.6 GeV) measurements.

Now let us discuss the impact of jet observables on the spin-averaged PDFs. Given that the PDFs are already well constrained by the DIS and DY data, we do not expect a significant difference between the PDFs before and after the inclusion of jet data, and this is indeed borne out overall in the results shown in Fig. 5.7.

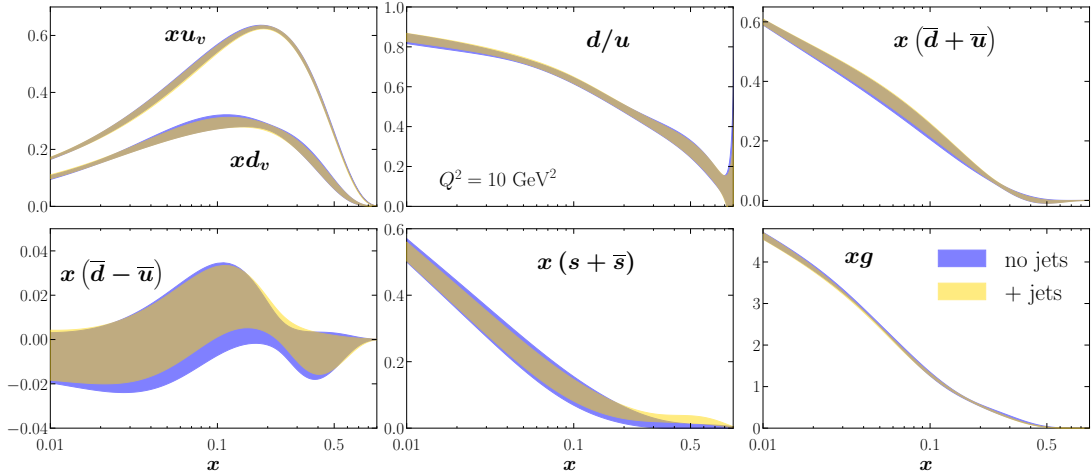


FIG. 5.7: Comparison of spin-averaged PDFs before (blue bands) and after (gold bands) the inclusion of the jet data from Tevatron [66, 67] and RHIC (STAR) [22]. The PDFs are plotted at  $Q^2 = 10 \text{ GeV}^2$ .

In Fig. 5.7, a small reduction of uncertainties in  $\bar{d} - \bar{u}$  and  $s + \bar{s}$  is found. The strangeness  $s + \bar{s}$  is also shifted upward at high  $x$  ( $\gtrsim 0.5$ ). This is due to the indirect effect from the

jet’s gluon sensitivity via the momentum sum rule.

The inclusion of RHIC jet data from unpolarized collisions [22] has no discernible effect on the PDFs obtained from fitting the Tevatron jets [66, 67], as shown in Fig. 5.8.

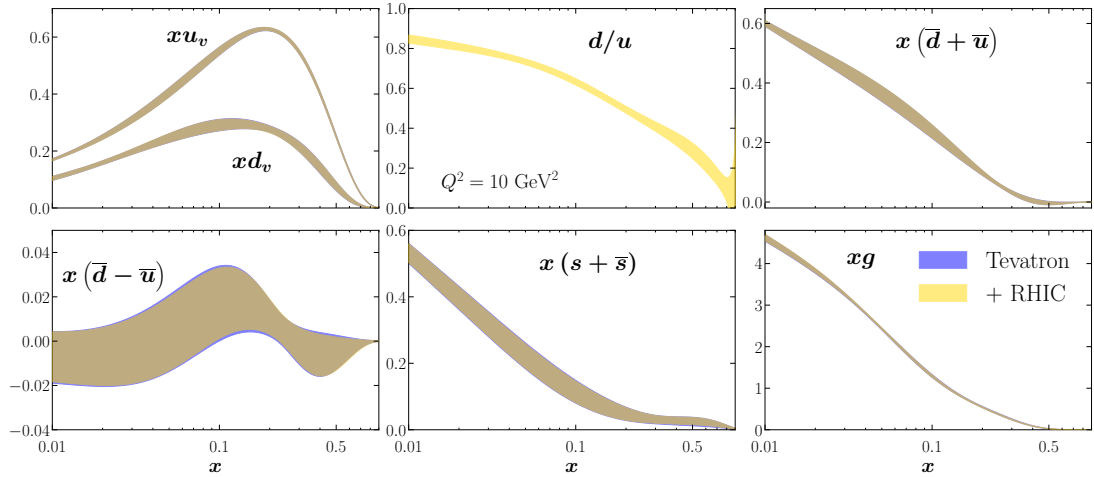


FIG. 5.8: Comparison of spin-averaged PDFs with jet data from the Tevatron [66, 67] only (blue bands) and with the addition of STAR [22] jet data (gold bands). The PDFs are plotted at  $Q^2 = 10 \text{ GeV}^2$ .

Finally, a comparison of the unpolarized PDFs extracted in the present analysis with those of several other global QCD analysis groups is shown in Fig. 5.9. Overall, a good agreement between our PDFs, labelled “JAM”, and those from previous analyses is found for the  $u_v$ ,  $d_v$  and  $g$  distributions.

The light antiquark asymmetry  $\bar{d} - \bar{u}$  of CSKK [93] is different from those of the other groups (including JAM) because CSKK did not include the DY fixed target data [65] that favors a positive asymmetry peaking at  $x \approx 0.1$ . The discrepancy between the CSKK  $d/u$  ratio and those from the other groups (including JAM) can also be accounted for by the absence of DY fixed target data in their analysis.

The JAM  $\bar{d} - \bar{u}$  also has a larger uncertainty and favors smaller values when compared with

the other groups (except CSKK). This may be due to the inclusion in the other groups' analyses (except for JAM19 [61]) of the collider  $W$ -boson production charge asymmetry data that EW sensitive to the light antiquark distributions. Our  $\bar{d} - \bar{u}$  differs from the earlier JAM19 fit mostly because the NMC  $F_2^d/F_2^p$  dataset [38] is corrected in this analysis by an inclusion of a normalization uncertainty, which significantly reduces the magnitude of the asymmetry across most values of  $x$ . On the other hand, our current analysis is not particularly focused on refining the  $\bar{d} - \bar{u}$  asymmetry, and further discussion of this will be left for future unpolarized JAM analyses.

For the strange quark PDF  $s + \bar{s}$ , the JAM19 fit gives a discernibly smaller result compared with those of the other groups (including this analysis). This is understood as being caused by the suppression from the inclusion of semi-inclusive DIS (SIDIS) and single-inclusive  $e^+e$  annihilation (SIA)  $K$  production data, as extensively discussed in Ref. [61]. The upward shift in the strangeness PDF at high  $x$  was discussed earlier in connection with Fig. 5.7.

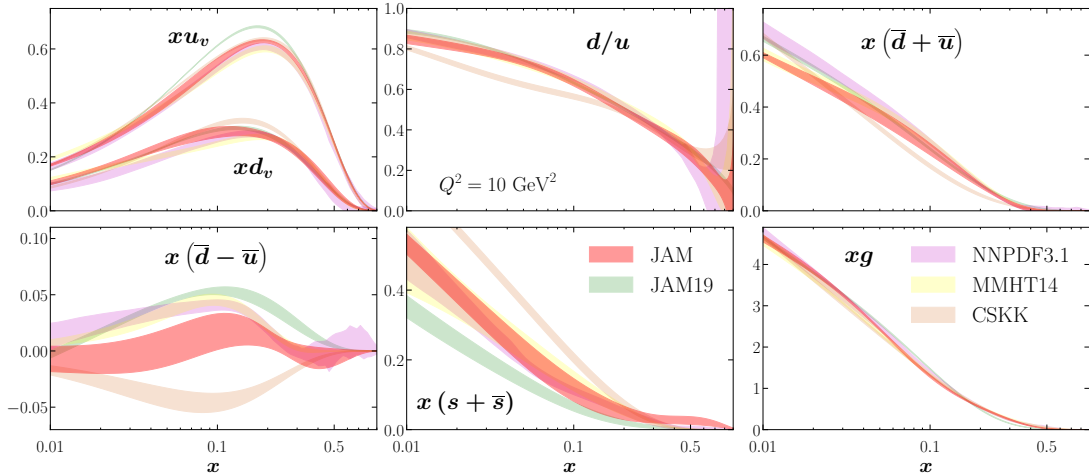


FIG. 5.9: Comparison of spin-averaged PDFs with JAM19 [61], NNPDF3.1 [92], MMHT14 [94] and CSKK [93]. The PDFs are plotted at  $Q^2 = 10 \text{ GeV}^2$ .

## 5.2 Analysis of spin-dependent PDFs

Now that we have a reliable baseline of the spin-averaged PDFs, the next step is to extract the spin-dependent PDFs by fitting the polarized DIS data and single jet production data from polarized  $pp$  collisions. Similar analyses have been performed by the DSSV [7] and NNPDF [95] groups. In those analyses, however, theory inputs including SU(2) and SU(3) flavor symmetries, as well as PDF positivity constraints, have been imposed in order to compensate for the lack of kinematic coverage in both  $x$  and  $Q^2$  when compared with the unpolarized world data (see Figs. 4.2 and 4.3). Given the debate about the role of SU(3) symmetry breaking in the proton spin decomposition [96] and the considerations on the positivity constraints discussed in Section 3.3, it is crucial to understand the roles that these and other theory inputs may play in the extraction of spin-dependent PDFs. To this effect, we consider three scenarios, listed in Table 5.3, which will enable us to explore the impact of the various assumptions on the analysis.

scenarios	SU(2)	SU(3)	positivity constraints
A	✓		
B	✓	✓	
C	✓	✓	✓

TABLE 5.3: Scenarios for various theory inputs in the extraction of spin-dependent PDFs, with the axial charges are given by the standard values  $g_A = 1.269(3)$  and  $a_8 = 0.586(31)$  [76].

In this section we first present results for the  $\chi^2$  values for the polarized world data that are included in this analysis, and discuss the quality of the fits by showing the agreement between the fitted JAM results and the experimental data. Following this, we consider the fitted spin-dependent PDFs for the scenarios A (assuming SU(2) symmetry only), B (assuming both SU(2) and SU(3) symmetry), and C (SU(2) and SU(3) symmetry and positivity of the PDFs), as summarized in Table 5.3. Finally, we present fits using  $g_A$  and

$a_8$  values extracted in the JAM17 analysis [8], as a proxy for the inclusion of polarized semi-inclusive DIS data that constrained the flavor decomposition in that analysis.

### 5.2.1 Quality of fits

The resulting  $\chi_{\text{dat}}^2 \equiv \chi^2/N_{\text{dat}}$  for the polarized world data are listed in Table 5.4, split by the three scenarios of Table 5.3. For the different scenarios, the fits give overall  $\chi_{\text{dat}}^2$  of 0.94 to 0.98 for a total of 426 data points from the polarized DIS and jet data. For scenario B, the agreement of the fit with each experimental datasets is further illustrated in Figs. 5.10 to 5.12 for polarized DIS data, and in Fig. 5.13 for polarized  $pp \rightarrow$  jet data. For the other two scenarios A and C, their fitted values are very similar to those found for scenario B, and will not be shown explicitly, but described by comparing with Figs. 5.10 to 5.13.

For the polarized fits, we follow the same steps as in Fig. 4.1, *i.e.*, fit first the polarized DIS data and then the polarized  $pp \rightarrow$  jet data. Shown in Fig. 5.10 are the fits to proton  $A_1$  from EMC [41], SMC [42, 43] and COMPASS [44, 46], and proton  $A_{\parallel}$  from SLAC [47, 49, 51] and HERMES [54]. Overall, they show a good agreement with the JAM global fit, with  $\chi_{\text{dat}}^2$  of less than or slightly above 1. Only the SMC [43], COMPASS [44] (scenario C only)  $A_1^p$ , and SLAC E155 [51]  $A_{\parallel}^p$  fits show a slightly worse agreement, with  $\chi_{\text{dat}}^2$  between  $\approx 1.2$  and 1.4.

The  $A_1^p$  and  $A_{\parallel}^p$  plot for scenario A has almost exactly same central values as in Fig. 5.10, but with slightly larger error bands at  $x \gtrsim 0.5$ . The same plot for scenario C, in contrast, has slightly lower central values and much narrower error bands for  $x \gtrsim 0.5$ . This is due to the suppression (see Fig. 5.14) of the  $\Delta s^+$  and especially  $\Delta u^+$  PDFs (see the leading order  $g_1$  expression in Eq. (2.15)) in the valence region ( $x \gtrsim 0.5$ ) by the positivity constraints.



experiment	reference	observable	reaction	$N_{\text{dat}}$	$\chi^2/N_{\text{dat}}$		
					A	B	C
EMC	41	$A_1$	$\mu p$	10	0.28	0.28	0.28
SMC	42	$A_1$	$\mu p$	11	0.33	0.30	0.36
SMC	43	$A_1$	$\mu p$	7	1.41	1.41	1.33
SMC	42	$A_1$	$\mu d$	11	1.63	1.62	1.68
SMC	43	$A_1$	$\mu d$	7	0.63	0.63	0.64
COMPASS	44	$A_1$	$\mu p$	11	1.01	1.07	1.25
COMPASS	46	$A_1$	$\mu p$	35	0.91	0.93	0.90
COMPASS	45	$A_1$	$\mu d$	11	0.43	0.46	0.60
SLAC (E80, E130)	47	$A_{\parallel}$	$e^- p$	10	0.76	0.78	0.79
SLAC (E143)	49	$A_{\parallel}$	$e^- p$	39	0.81	0.80	0.82
SLAC (E155)	51	$A_{\parallel}$	$e^- p$	59	1.28	1.28	1.37
SLAC (E143)	49	$A_{\parallel}$	$e^- d$	39	1.08	1.08	1.17
SLAC (E155)	52	$A_{\parallel}$	$e^- d$	59	0.99	0.98	0.96
SLAC (E142)	48	$A_1$	$e^- h$	4	0.91	0.92	0.40
SLAC (E154)	50	$A_{\parallel}$	$e^- h$	15	0.38	0.37	0.66
HERMES	53	$A_1$	$e^+ n$	5	0.13	0.13	0.10
HERMES	54	$A_{\parallel}$	$e^+ p$	16	0.59	0.57	0.59
HERMES	54	$A_{\parallel}$	$e^+ d$	16	1.03	1.02	1.10
STAR (2003)	22	$A_{LL}$	$pp$	6	1.52	1.52	1.51
STAR (2005)	23	$A_{LL}$	$pp$	10	1.12	1.12	1.12
STAR (2006)	23	$A_{LL}$	$pp$	9	0.34	0.35	0.36
STAR (2009)	24	$A_{LL}$	$pp$	22	0.83	0.82	0.93
STAR (2012)	25	$A_{LL}$	$pp$	14	1.59	1.55	1.53
STAR (2015)	26	$A_{LL}$	$pp$	22	1.07	1.10	1.08
PHENIX (2005)	27	$A_{LL}$	$pp$	8	0.60	0.60	0.61
<b>total</b>				<b>456</b>	<b>0.94</b>	<b>0.94</b>	<b>0.98</b>

TABLE 5.4: Polarized inclusive DIS and single jet production data used in this analysis, indicating the observables fitted, the reactions (with  $h$  stands for  ${}^3\text{He}$ ), number of data points in each experiment ( $N_{\text{dat}}$ ) and respective  $\chi^2/N_{\text{dat}}$  values for the scenarios A, B and C (see Table 5.3). The  $\chi^2$  are computed by Eq. (4.5) using the average of theory predictions from all the replicas. Data presented as neutron  $A_1$  are obtained from a  ${}^3\text{He}$  target [53]. The numbers in parentheses for STAR and PHENIX data indicate the years when the data were recorded. The  $\chi^2$  values for the fitted  $g_A$  in Eq. (3.6) are 0.014, 0.001 and 0.070 for scenarios A, B and C, respectively, while the  $\chi^2$  of fitting  $a_8$  in Eq. (3.6) are 0.000 and 2.647 for scenarios B and C, respectively.

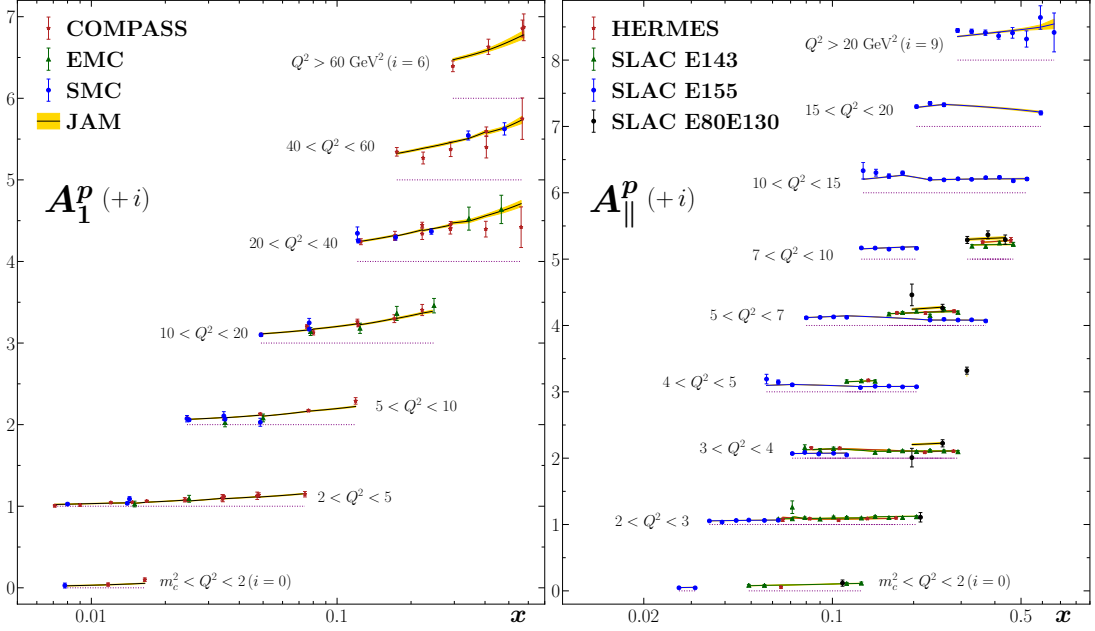


FIG. 5.10: Proton  $A_1$  (left panel) from EMC [41] (green upward triangles), SMC [42, 43] (blue solid circles) and COMPASS [44, 46] (red stars), and proton  $A_{\parallel}$  (right panel) from SLAC [47, 49, 51] (green upward triangles, blue solid circles and black solid circles for E143, E155 and E80E130, respectively) and HERMES [54] (red stars) collaborations. Data points in different  $Q^2$  bins are added by integers  $i$  for clarity, with purple dotted lines indicating the actual zeros. The data are compared with  $A_1^p$  and  $A_{\parallel}^p$  from the JAM global analysis (solid curves with bands indicating  $1\sigma$  uncertainties) that are obtained from the average of all the Monte Carlo replicas.

Next shown in Fig. 5.11 are the fits to the deuteron  $A_1$  from SMC [42, 43] and COMPASS [45], and deuteron  $A_{\parallel}$  from SLAC [49, 52] and HERMES [54]. Overall, they show a good agreement with the JAM global fit, with  $\chi_{\text{dat}}^2$  of less than or slightly above 1. Only the SMC [42]  $A_1^p$  fit shows a slightly worse agreement with  $\chi_{\text{dat}}^2 \approx 1.6$ .

The  $A_1^d$  and  $A_{\parallel}^d$  plot for scenario A has almost exactly same central values as in Fig. 5.11, except that it has slightly larger uncertainty bands for  $x \gtrsim 0.5$ . Unlike the proton case, the same plot for scenario C has central values that are more similar to those in Fig. 5.11 but with smaller error bands. This can be understood from the suppression (see Fig. 5.14) of the  $\Delta d^+$  due to positivity constraints being less dramatic than that of  $\Delta u^+$ .

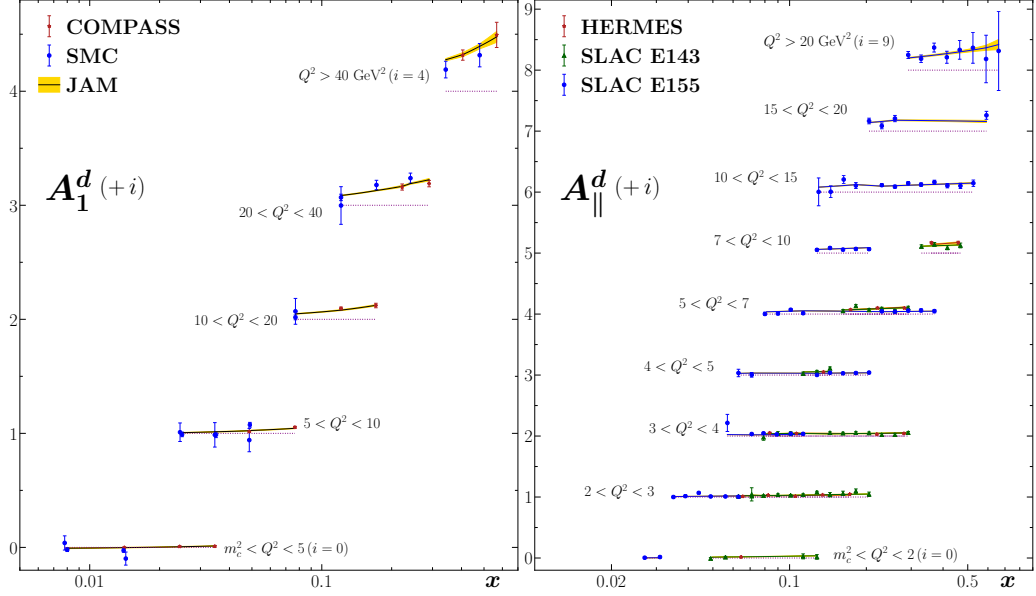


FIG. 5.11: Deuteron  $A_1$  (left panel) from SMC [42, 43] (blue solid circles) and COMPASS [45] (red stars), and deuteron  $A_{\parallel}$  (right panel) from SLAC [49, 52] (green upward triangles and blue solid circles for E143 and E155, respectively) and HERMES [54] (red stars) collaborations. Data points in different  $Q^2$  bins are added by integers  $i$  for clarity, with purple dotted lines indicating the actual zeros. The data are compared with the  $A_1^d$  and  $A_{\parallel}^d$  from the JAM global analysis (solid curves with bands indicating  $1\sigma$  uncertainties) that are obtained from the average of all the Monte Carlo replicas.

Finally, the fits to neutron  $A_1$  from HERMES [53] and  $^3\text{He}$   $A_1$  (left panel) from SLAC E142 [48], along with  $^3\text{He}$   $A_{\parallel}$  from the SLAC E154 [50] experiments are shown in Fig. 5.12. An overall good agreement with the JAM global fit is found with  $\chi_{\text{dat}}^2$  of less than 1.

The  $A_1^n$ ,  $A_1^{^3\text{He}}$  and  $A_{\parallel}^{^3\text{He}}$  results for scenarios A and C are almost identical for both central values and uncertainty bands as in Fig. 5.12.

Now let us change our focus to the jet observables in polarized  $pp$  scattering and their impact on spin-dependent PDFs. Firstly, we present the JAM global fit to the STAR [22–26] and PHENIX [27] jet double longitudinal spin asymmetries  $A_{LL}$  in Fig. 5.13 for scenario B. An overall good agreement with the JAM global fit is found with  $\chi_{\text{dat}}^2$  less than 1. Only the fits to the STAR 2003 [22] and 2012 [25] data show slightly worse agreement, with

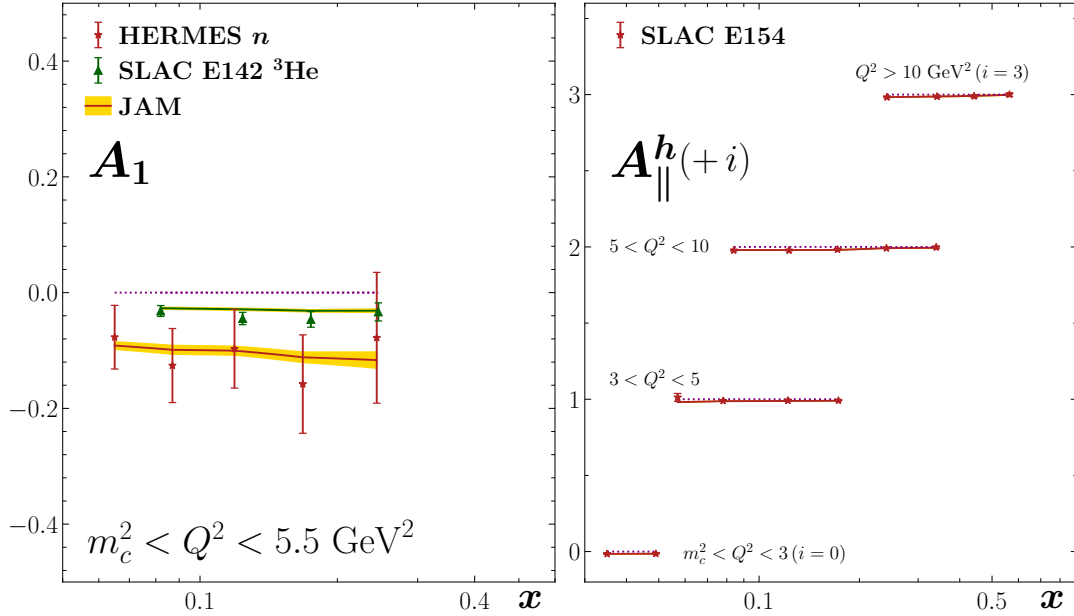


FIG. 5.12: Neutron  $A_1$  from HERMES [53] (red stars with error bars) [left panel] and  ${}^3\text{He}$   $A_1$  from SLAC E142 [48] (green upward triangles) [left panel], and  ${}^3\text{He}$   $A_{\parallel}$  from SLAC E154 [50] (red stars) [right panel]. Data points in different  $Q^2$  bins are added by integers  $i$  for clarity, with purple dotted lines indicating the actual zeros. The data are compared with the  $A_1^n$ ,  $A_1^{3\text{He}}$  and  $A_{\parallel}^{3\text{He}}$  from the JAM global analysis (solid curves with bands indicating  $1\sigma$  uncertainties) that are obtained from the average of all the Monte Carlo replicas.

$\chi_{\text{dat}}^2 \approx 1.5$ , which is likely due to the outliers in the experimental data.

The  $A_{LL}$  plot for scenario A, compared to Fig. 5.13, has slightly higher central values and moderately wider uncertainty bands at  $p_T \gtrsim 30$  GeV, due to the relatively large uncertainties (see Figs. 5.14 and 5.15) of all the helicity PDFs in the absence of SU(3) flavor symmetry. The same plot for scenario C has visibly lower central values when compared to Fig. 5.11, and much narrower error bands for  $p_T \gtrsim 30$  GeV. This is caused by a significant suppression (see Fig. 5.15) of the  $\Delta g$  solution space as a result of the positivity constraints.

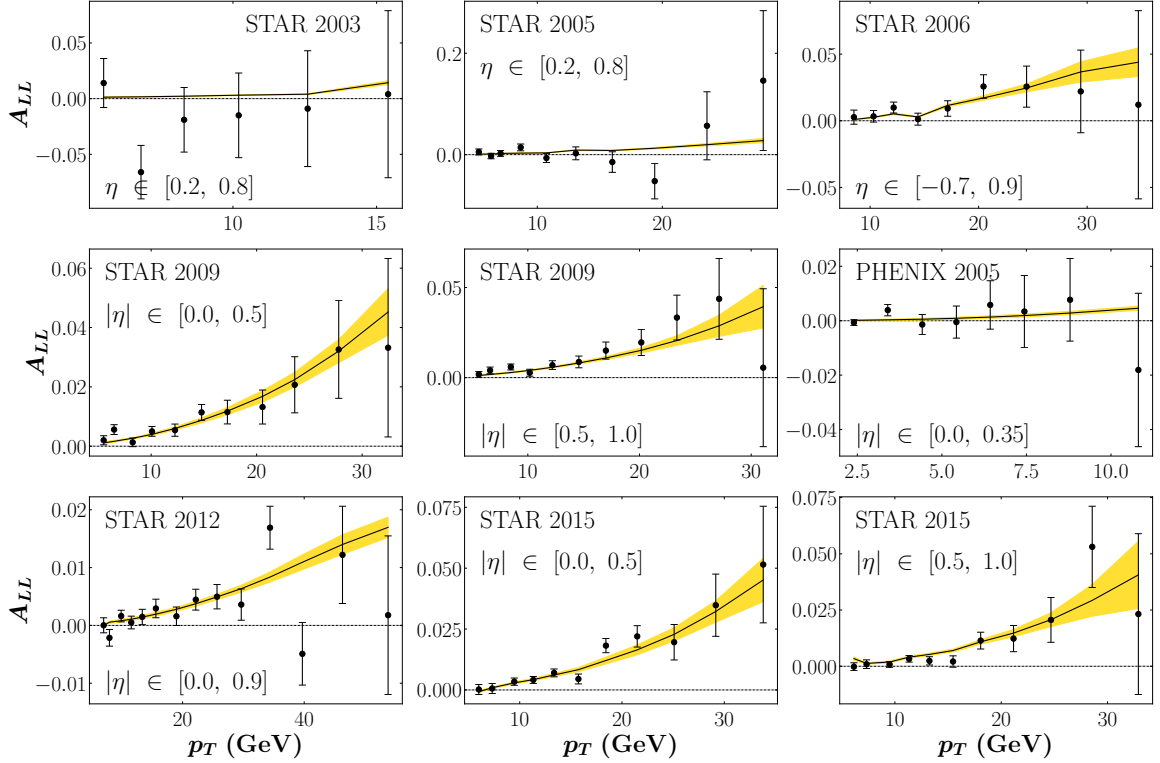


FIG. 5.13: Fit to jet double longitudinal spin asymmetries  $A_{LL}$  measurements from STAR [22–26] and PHENIX [27] collaborations at RHIC. Each subplots show the collaborations, year of data taken and pseudorapidity  $\eta$  or  $|\eta|$  bins. The data (solid dots) are compared with fits from the JAM global analysis (solid curves with bands indicating  $1\sigma$  uncertainties) that are obtained from the average of all Monte Carlo replicas.

## 5.2.2 Spin-dependent PDFs and theory inputs

Knowing that the polarized world data are well fitted, we now focus on the spin-dependent PDFs and how they are affected by the theory inputs. Recall that SU(3) flavor symmetry provides an extra constraint on the  $\Delta s^+$  distribution via Eq. (3.6), while the positivity constraints should mainly affect the helicity distributions at high  $x$ , where the spin-averaged PDFs are closer to zero and will thus exert more suppression on the helicity PDFs.

For the light quark helicity distributions in Fig. 5.14, we indeed find that this matches with the above expectations. All the quark helicity distributions, especially  $\Delta s^+$ , have

significant reduction in their uncertainties from the SU(3) flavor symmetry constraint. The positivity constraints, on the other hand, reshape  $\Delta u^+$ , and in particular  $\Delta d^+$ , discernibly, and further reduce significantly (on top of the reduction from the SU(3) constraint) the uncertainty of  $\Delta s^+$ . The distortion of the shapes of  $\Delta u^+$  and  $\Delta d^+$  comes mainly from their antiquark parts, namely  $\Delta \bar{u}$  and  $\Delta \bar{d}$ . This is because we set  $\Delta \bar{u} = \Delta \bar{d} = \Delta \bar{s} = \Delta s$  back in Eq. (3.5), and the changes in  $\Delta s$  propagate to the  $\Delta u^+$  and  $\Delta d^+$  distributions. The absolute values of  $\Delta d^+$  are smaller than those of  $\Delta u^+$ , and thus the distortion is more discernible for  $\Delta d^+$ .

One point we would like to stress is that a significant bias exists on the strange helicity distribution, since its uncertainty is very strongly constrained by both the SU(3) flavor symmetry and the positivity assumptions.

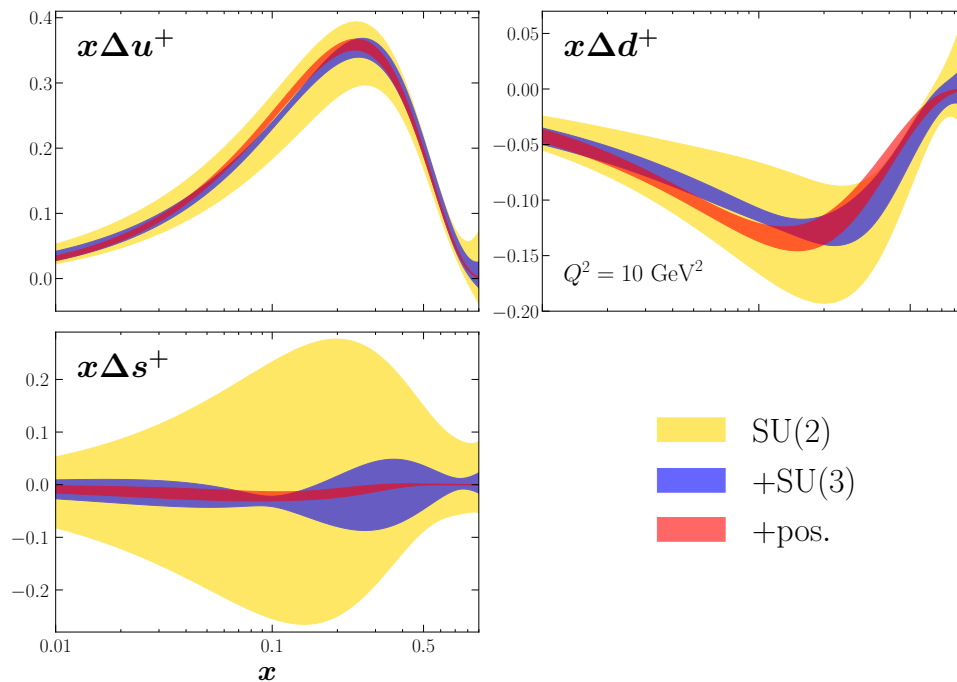


FIG. 5.14: Spin-dependent quark PDF fitted with varying theory inputs as listed in Table 5.3, plotted at  $Q^2 = 10 \text{ GeV}^2$ . The scenarios A, B and C are colored as yellow, blue and red bands, respectively.

For the gluon distribution  $\Delta g$ , we first notice from Fig. 5.15 that there are two distinct solutions for scenarios A and B, with one being mostly above zero (“positive”) and the other mostly below zero (“negative”). Despite the differences in sign and shape (especially at high  $x$ ), the two  $\Delta g$  solutions can describe the data equally well. This can be seen from Table 5.4 by comparing the  $\chi^2/N_{\text{dat}}$  values for the fits of scenarios A and B (which contain both “positive” and “negative” solutions) with those of scenario C (which does not have the “negative” solution), all of which give  $\chi^2/N_{\text{dat}}$  values below 1. The reason why the “negative” solution can give an equally good description of the data as the “positive” solution is the large cancellation that occurs between the contributions to the jet double longitudinal spin asymmetry from the  $qg$  and  $gg$  channels when  $\Delta g$  is negative.

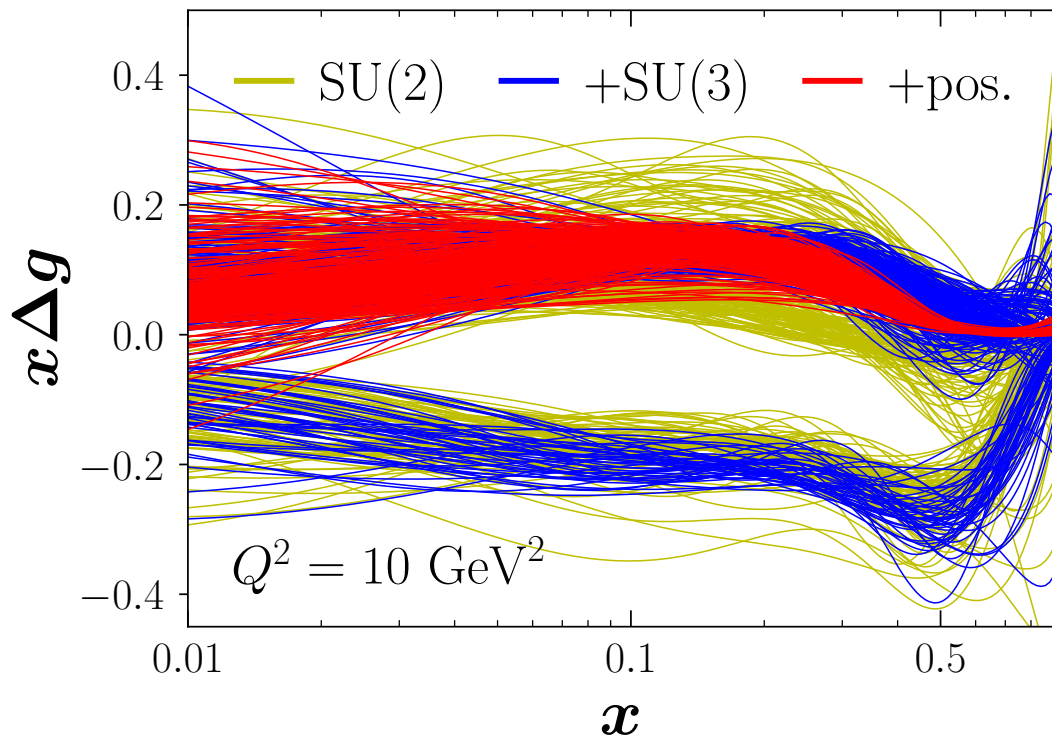


FIG. 5.15: Spin-dependent gluon PDF fitted with varying theory inputs as listed in Table 5.3, plotted at  $Q^2 = 10 \text{ GeV}^2$ . The scenarios A, B and C are colored as yellow, blue and red bands, respectively.

As for the effect of theory assumptions, we find that the addition of SU(3) flavor symmetry has an indirect effect of mildly reducing its uncertainty. However, the positivity constraint, apart from reducing the uncertainty of  $\Delta g$  at high  $x$ , also eliminates the “negative”  $\Delta g$  solution entirely (see Fig. 5.15). This is because, in order to satisfy the positivity constraints, the shape of the “negative” solution becomes distorted and consequently fails to describe the jet  $A_{LL}$  data (see Table 5.5), giving very large  $\chi^2/N_{\text{dat}}$  values. Based on this observation, we remove the “negative”  $\Delta g$  solution from the Monte Carlo replicas for scenario C.

experiment	$N_{\text{dat}}$	$\chi^2/N_{\text{dat}}$	
		“positive”	“negative”
STAR (2003) [22]	6	1.51	1.53
STAR (2005) [23]	10	1.12	0.91
STAR (2006) [23]	9	0.36	2.92
STAR (2009) [24]	22	0.92	7.56
STAR (2012) [25]	14	1.53	3.31
STAR (2015) [26]	22	1.08	11.85
PHENIX (2005) [27]	8	0.61	0.59

TABLE 5.5: Table of  $\chi^2$  for the two  $\Delta g$  solutions for the scenario C, indicating the number of data points in each experiment ( $N_{\text{dat}}$ ) and respective  $\chi^2/N_{\text{dat}}$  values for the “positive” and “negative”  $\Delta g$  solutions.

In addition to the scenarios A, B and C covered in Table 5.3 and shown in Figs. 5.14 and 5.15, we also explored other possibilities, such as SU(2) plus positivity constraints. We found in this case that the resulting spin-dependent PDFs (not shown in this thesis) were very similar to those from the scenario C. Therefore, we find that the order of imposing the SU(3) flavor symmetry and the positivity constraints does not change our observations in connection with Figs. 5.14 and 5.15.

Another way to visualize the impact of theory inputs on the spin-dependent PDFs is to



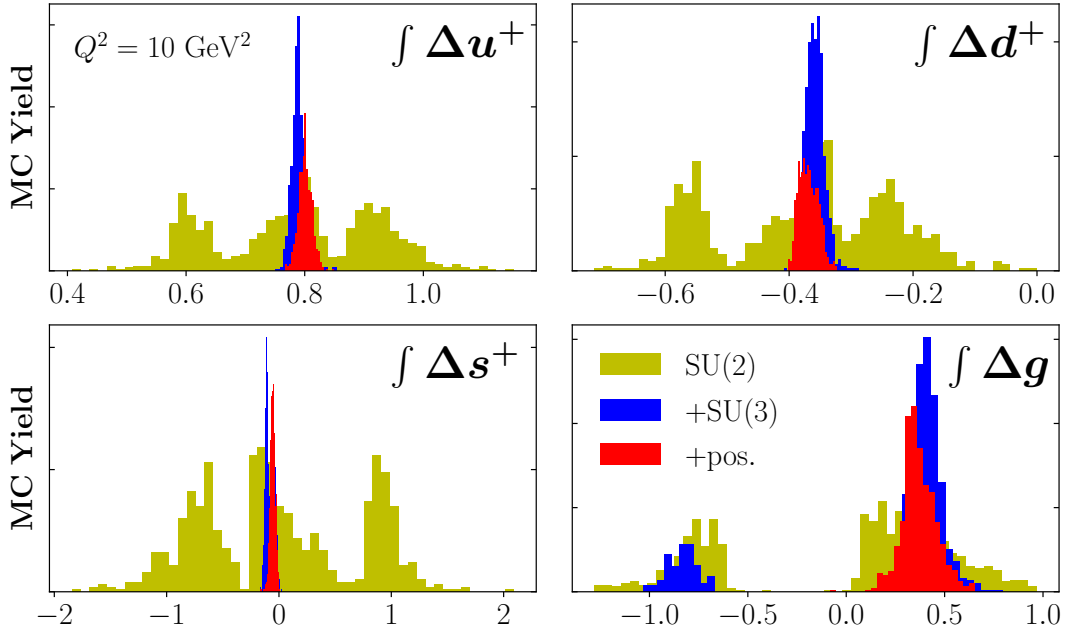


FIG. 5.16: Truncated moments of spin-dependent PDFs that are fitted with different theory inputs as listed in Table 5.3, plotted at  $Q^2 = 10 \text{ GeV}^2$ . The scenarios A, B and C are colored as yellow, blue and red bands, respectively.

plot the histograms of truncated moments, defined as

$$\int \Delta q^+ \equiv \int_{0.01}^1 \Delta q^+ dx, \quad \int \Delta g \equiv \int_{0.01}^1 \Delta g dx, \quad (5.1)$$

where  $\Delta q^+ = \Delta q + \Delta \bar{q}$  for  $u$ ,  $d$  and  $s$ , and 0.01 is chosen to be the lower limit of the integral because it is roughly the smallest  $x$  value to which polarized data have sensitivity (see Fig. 4.3). The reduction of the uncertainties by imposing the SU(3) flavor symmetry is clearly significant for all the light quark distributions (from yellow to blue bands in Fig. 5.16). As for the  $\Delta g$  truncated moments, we can clearly see the two solutions for the scenarios A and B in Fig. 5.16, and the mild reduction of uncertainties after imposing the SU(3) flavor symmetry. The  $\Delta g$  solution that gives negative  $\int \Delta g$  values is then eliminated by the positivity constraints.

The values and uncertainties of the truncated moment  $\int \Delta g$  as defined in Eq. (5.1) are shown in Table 5.6. Putting uncertainties in the  $x \rightarrow 0$  extrapolation aside, these values would enter the proton spin decomposition (1.4) as the  $\Delta G$  term. With the contributions from the “positive” and “negative” solutions differing substantially, the proton spin sum would imply rather different resulting values of the orbital angular momentum  $L_{Q+G}$  for the different cases. Future studies of  $\Delta g$  with data that have more sensitivity, as well as of the extraction of the  $L_{Q+G}$  term from generalized parton distributions, will be needed to more precisely determine the polarization of gluons in the nucleon and the proton spin decomposition.

	A	B	C
“positive”	$0.37 \pm 0.20$	$0.41 \pm 0.08$	$0.37 \pm 0.09$
“negative”	$-0.79 \pm 0.15$	$-0.83 \pm 0.08$	
combined	$0.02 \pm 0.57$	$0.20 \pm 0.47$	

TABLE 5.6: Values and uncertainties of the truncated gluon moment  $\int_{0.01}^1 \Delta g dx$  for the scenarios listed in Table 5.3. For scenarios A and B which contain both the “positive” and “negative”  $\Delta g$  solutions, their values are shown both separately and combined.

To compare with the DSSV14 [7] extracted truncated moment of  $\Delta g$  which is integrated from 0.05 to 1, in Table 5.7 we also present the corresponding JAM extracted values and uncertainties.

	A	B	C
“positive”	$0.21 \pm 0.15$	$0.26 \pm 0.03$	$0.24 \pm 0.03$
“negative”	$-0.57 \pm 0.12$	$-0.60 \pm 0.03$	
combined	$-0.02 \pm 0.39$	$0.12 \pm 0.32$	

TABLE 5.7: Values and uncertainties of the truncated gluon moment  $\int_{0.05}^1 \Delta g dx$  for the scenarios listed in Table 5.3. For scenarios A and B which contain both the “positive” and “negative”  $\Delta g$  solutions, their values are shown both separately and combined. The values shown in this table are to be compared with the widely quoted DSSV14 [7] results that is roughly  $0.2 \pm 0.05$ .

Despite the uncertainties in scenario A, the “positive” solutions from scenarios A, B and C all have a good agreement with the DSSV14 result ( $\approx 0.2 \pm 0.05$ ). The “negative” solution which is eliminated by the positivity constraints, however, gives values that are significantly different from the DSSV14 result in scenarios A and B. This indicates once again that the positivity constraints can introduce significant bias.

### 5.2.3 Fits with JAM17 $g_A$ and $a_8$

As has been previously discussed and shown in Fig. [5.14](#), the strange helicity distribution  $\Delta s^+$  is currently constrained mainly by theory inputs rather than experimental data. In order to constrain  $\Delta s^+$  from experiment, adding polarized semi-inclusive DIS (SIDIS) data in a global QCD analysis would be optimal. To investigate the practicality of various options, we adopt here  $g_A = 1.24(4)$  and  $a_8 = 0.46(21)$  obtained in the JAM17 analysis [\[8\]](#) as a proxy for including the polarized SIDIS data. The JAM17  $g_A$  and  $a_8$  are used for this purpose because rather than imposing SU(2) and SU(3) flavor symmetries with the standard values of  $g_A$  and  $a_8$ , JAM17 extracted the purely data driven  $g_A$  and  $a_8$  values and uncertainties by fitting simultaneously the spin-dependent PDFs and fragmentation functions (FFs) with the polarized inclusive DIS and SIDIS, and  $e^+e^-$  SIA data.

Using the JAM17  $g_A$  and  $a_8$  values, we find no discernible difference in the  $\Delta u^+$ ,  $\Delta d^+$  or  $\Delta g$  PDFs (shown in the right panel of Fig. [5.17](#)) when compared with scenario B in Figs. [5.14](#) and [5.15](#). On the other hand, as shown in the left panel of Fig. [5.17](#), the  $\Delta s^+$  PDF expands its solution space from the less restrictive  $g_A$  and  $a_8$  values and uncertainties. This is expected because the JAM17  $g_A$  is not too different from its standard value and uncertainty, while the JAM17  $a_8$  value and uncertainty are somewhat further away.

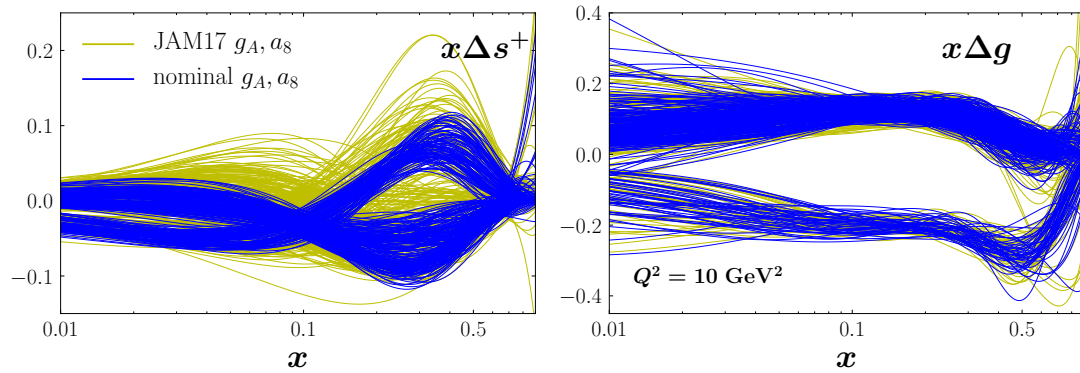


FIG. 5.17: Comparison of spin-dependent strange and gluon distributions  $\Delta s^+$  and  $\Delta g$ , using the standard (or nominal, yellow lines) and JAM17 (blue lines)  $g_A$  and  $a_8$  values, plotted at  $Q^2 = 10 \text{ GeV}^2$ .

## 5.3 Analysis of jet observables

In previous sections, the renormalization and factorization scales  $\mu_R$  and  $\mu_F$  have been set to  $0.5 p_T$ , even though other choices, such as  $\mu_{R,F} = p_T$  and  $\mu_{R,F} = 2 p_T$ , are also possible. Moreover, the tension in the STAR 2003 unpolarized jet data [22] was also only briefly mentioned. In this section, we discuss in more detail the justification for the choices we made in previous sections and show that they are indeed optimal.

### 5.3.1 Renormalization and factorization scales

From a theoretical point of view, the physical cross sections should not depend on the renormalization or factorization scales  $\mu_R$  or  $\mu_F$ . At any fixed, finite order in  $\alpha_s$ , however, the calculation may depend slightly on the choice of  $\mu_R$  and  $\mu_F$ , which should decrease with increasing order of perturbative correction. Therefore, the differences in cross sections from varying  $\mu_R$  and  $\mu_F$  are usually taken as an estimation of theoretical uncertainties. In analyses of jet data, the scales  $\mu_R$  and  $\mu_F$  are generally chosen to be  $0.5 p_T$ ,  $p_T$  or  $2 p_T$  at every data point, with  $p_T$  being the jet transverse momentum at that point. In Table 5.8 we show the fit qualities for the three choices listed above.

	D0	CDF	STAR 2003	STAR 2004
$\mu_{R,F} = 0.5 p_T$	0.89	1.11	4.94	1.05
$\mu_{R,F} = p_T$	0.87	1.14	6.11	1.14
$\mu_{R,F} = 2 p_T$	0.89	1.19	5.42	1.23

TABLE 5.8: Fitted  $\chi^2/N_{\text{dat}}$  values with different  $\mu_{R,F}$  choices, for unpolarized jet observables from D0 [66], CDF [67] and STAR (2003/2004) [22] collaborations, where  $N_{\text{dat}}$  represents the number of data points in each experiment. The  $\chi^2$  values are computed from Eq. (4.5) using the average of theory predictions from all replicas.

Based on the quality of the fits in Table 5.8, the renormalization and factorization scales are chosen to be  $\mu_R = \mu_F = 0.5 p_T$ . As mentioned before, at a finite order in  $\alpha_s$  the

physical cross sections could depend mildly on the choices of  $\mu_R$  and  $\mu_F$ . However, as observed in Fig. 5.18, the difference in the PDFs are very small.

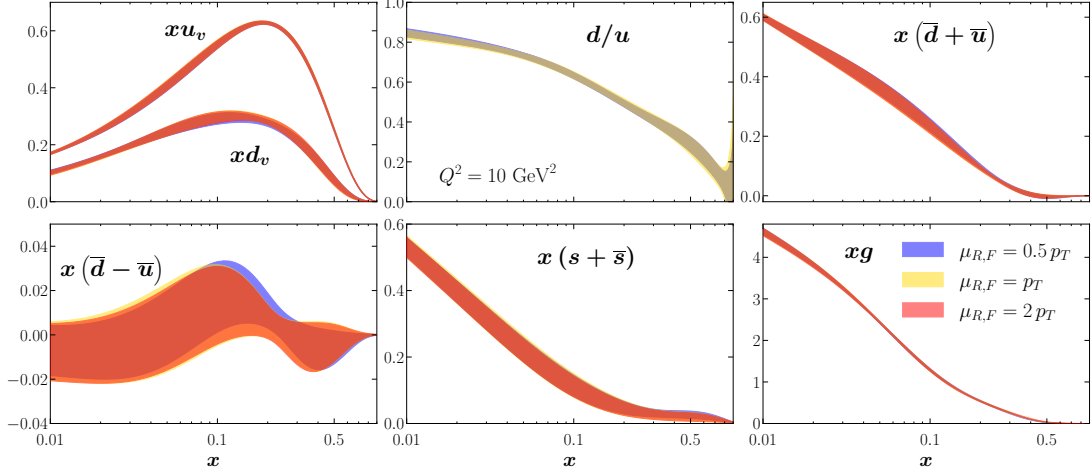


FIG. 5.18: Comparison of spin-averaged PDFs with different choices of  $\mu_R$  and  $\mu_F$ , plotted at  $Q^2 = 10 \text{ GeV}^2$ . The PDFs with choices of  $\mu_{R,F}$  equal to  $0.5 p_T$ ,  $p_T$  and  $2 p_T$  are plotted by blue, yellow and red bands, respectively.

For the jet production data in polarized  $pp$  collisions, varying the renormalization and factorization scales only makes marginal differences in  $\chi^2$  (see Table 5.9). We therefore set  $\mu_R = \mu_F = 0.5 p_T$  for jet production data in both unpolarized and polarized  $pp$  collisions.

	STAR						PHENIX
	2003	2005	2006	2009	2012	2015	
$\mu_{R,F} = 0.5 p_T$	1.52	1.12	0.35	0.82	1.55	1.10	0.60
$\mu_{R,F} = p_T$	1.52	1.14	0.35	0.82	1.55	1.08	0.60
$\mu_{R,F} = 2 p_T$	1.52	1.13	0.36	0.82	1.54	1.10	0.60

TABLE 5.9: Fitted  $\chi^2/N_{\text{dat}}$  values with different  $\mu_{R,F}$  choices, for polarized jet observables from STAR [22–26] and PHENIX [27] collaborations, where  $N_{\text{dat}}$  represents the number of data points in each experiment. The  $\chi^2$  are computed by Eq. (4.5) using the average of theory predictions from all the replicas.

As in the spin-averaged case, the spin-dependent quark and gluon PDFs are mostly insensitive to the variations of  $\mu_R$  and  $\mu_F$ . For  $\mu_{R,F} = 2 p_T$  and  $\mu_{R,F} = 0.5 p_T$ , the uncertainty

bands on the quark PDFs at high  $x$  are larger due to outliers and are therefore not significant.

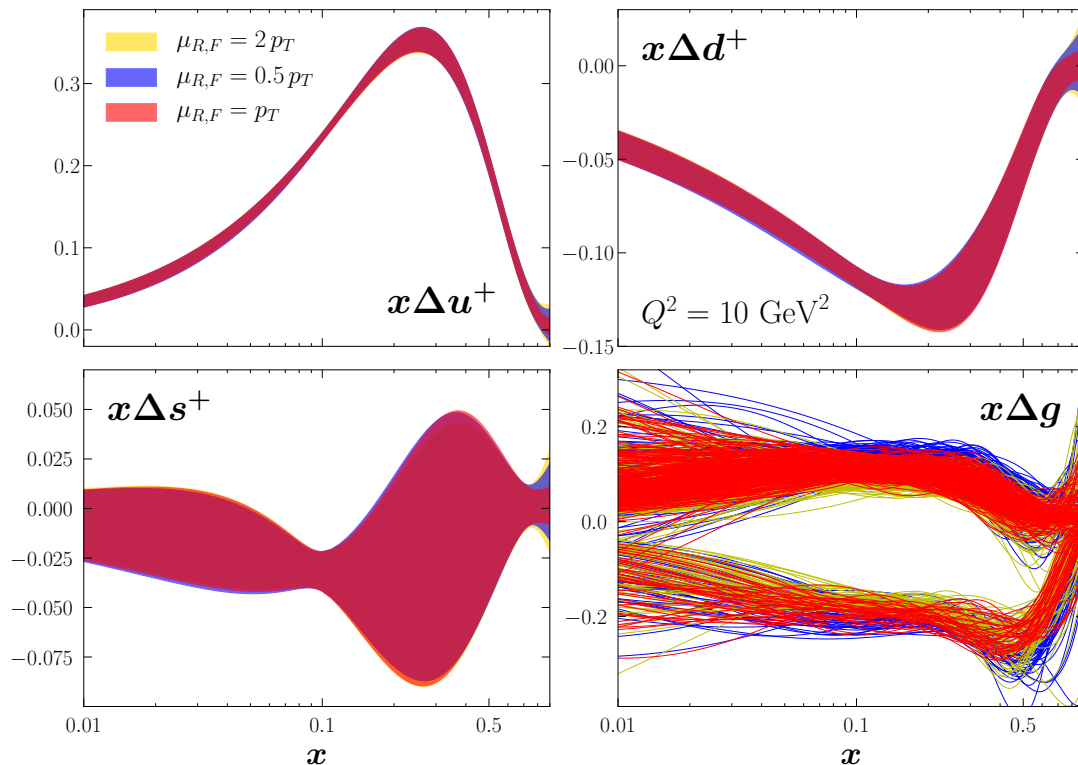


FIG. 5.19: Comparison of spin-dependent PDFs with different choices of  $\mu_R$  and  $\mu_F$ , plotted at  $Q^2 = 10 \text{ GeV}^2$ . The helicity quark distributions ( $\Delta u^+$ ,  $\Delta d^+$  and  $\Delta s^+$ ) with choices of  $\mu_{R,F}$  equal to  $0.5 p_T$ ,  $p_T$  and  $2 p_T$  are plotted by blue, yellow and red bands, respectively, while the helicity gluon distributions are plotted by lines of corresponding colors to show clearly the two solutions.

### 5.3.2 Cuts on $p_T$

As discussed in Section [5.1](#), the tension between the STAR 2003 lowest- $p_T$  (5.5 and 6.8 GeV) and slightly higher- $p_T$  (8.3 to 12.6 GeV) measurements results in a large  $\chi^2_{\text{dat}}$  for the STAR 2003 unpolarized jet dataset [\[22\]](#). To resolve the tension, we place a  $p_T$  cut at 10 GeV, leaving the STAR 2003 data with only the points above 10 GeV. By doing this, the tension can be resolved and the  $\chi^2_{\text{dat}}$  for the STAR 2003 data would improve.

Indeed, a huge improvement in  $\chi_{\text{dat}}^2$  is found for the STAR 2003 data after the inclusion of the  $p_T$  cut, as shown in Table 5.10. The  $\chi_{\text{dat}}^2$  for the D0 [66] and CDF [67] data are almost identical as the lowest  $p_T$  values accessed by D0 and CDF are 54.5 and 58 GeV, respectively. The  $\chi_{\text{dat}}^2$  for the STAR 2004 [22] data is only slightly affected.

	no cut	$p_T > 10$ GeV
D0	0.89	0.89
CDF	1.11	1.10
STAR 2003	4.94	0.02
STAR 2004	1.05	1.19

TABLE 5.10: Fitted  $\chi^2/N_{\text{dat}}$  values with and without the  $p_T$  cut, for unpolarized jet observables from D0 [66], CDF [67] and STAR 2003 and 2004 runs [22]. The  $\chi^2$  values are computed from Eq. (4.5) using the average of theory predictions from all replicas.

Despite the significant reduction in  $\chi_{\text{dat}}^2$  for the STAR 2003 unpolarized jet data, the PDFs are essentially unaffected (see Fig. 5.20). As discussed in connection with Fig. 5.8, since the RHIC unpolarized jet data have no discernible impact on the PDFs, removing 5 points from this dataset as a result of the  $p_T$  cut has negligible effect on the PDFs.

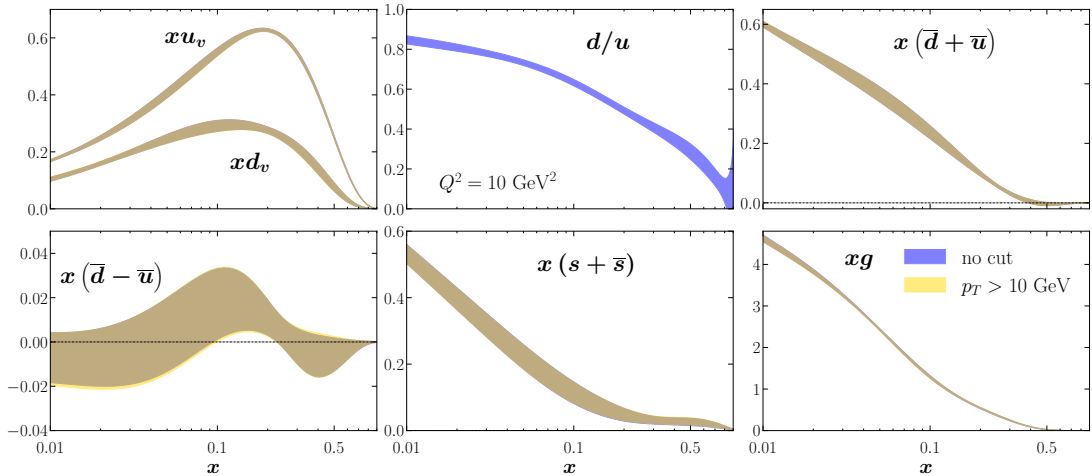


FIG. 5.20: Comparison of spin-averaged PDFs with (yellow bands) and without (blue bands) the  $p_T$  cut, plotted at  $Q^2 = 10$  GeV<sup>2</sup>.



## 5.4 Extraction of helicity basis PDFs

Now that good agreement has been achieved in fitting both the unpolarized and polarized world data (see Tables 5.2 and 5.4), we are in the position to complete the last remaining step in Fig. 4.1, which is the simultaneous extraction of spin-averaged and spin-dependent PDFs from the combined datasets. By performing a simultaneous analysis, we can extract for the first time the helicity dependent PDFs, with spins parallel ( $f^\uparrow$ ) and antiparallel ( $f^\downarrow$ ) to the proton spin, with a consistent treatment of uncertainties. Before showing the results, however, let us first recall that we have set up three different scenarios A, B and C in Table 5.3 in order to study the impact of theory inputs on the extraction of spin-dependent PDFs. Consequently, the extracted helicity basis will be also presented for all three scenarios A, B and C.

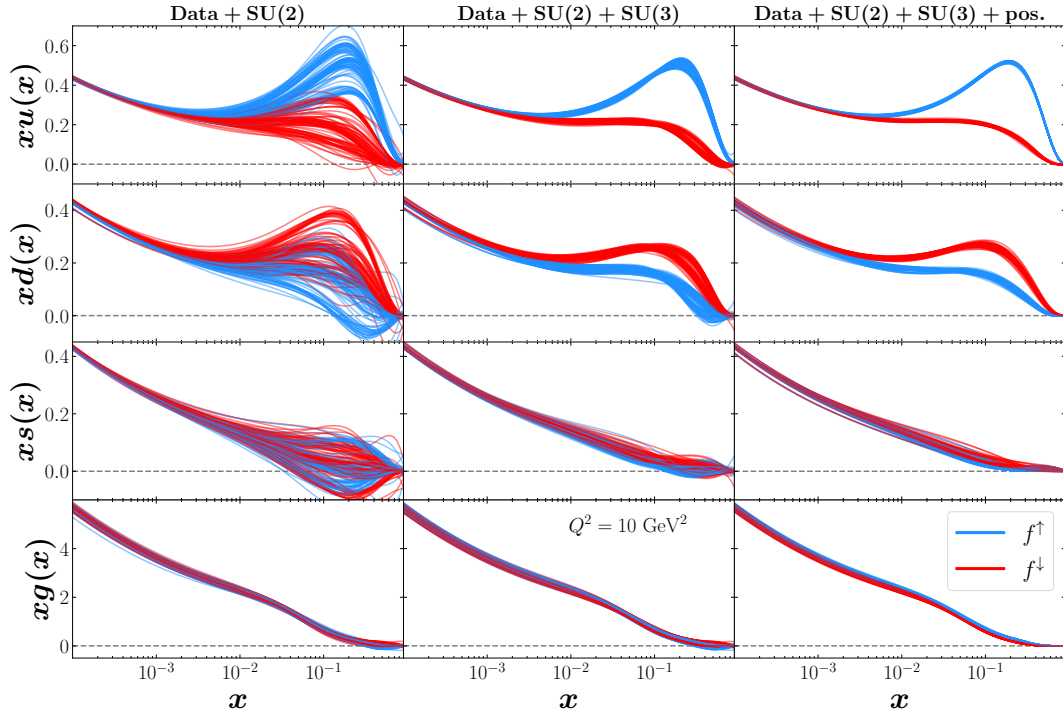


FIG. 5.21: Helicity basis PDFs  $f^\uparrow$  (blue) and  $f^\downarrow$  (red) are plotted for flavors  $u$ ,  $d$ ,  $s$  and  $g$ , respectively from top to bottom row, at  $Q^2 = 10 \text{ GeV}^2$ . Scenarios A, B and C (see Table 5.3) are shown in the left, middle and right columns, respectively.

The helicity basis PDFs  $f^\uparrow$  and  $f^\downarrow$  are presented in Fig. 5.21. It can be immediately inferred from Fig. 5.21 that imposing SU(3) assumption reduces the uncertainties of the helicity basis PDFs for all the light quark flavors, as a result of the extra constraint provided by  $a_8$  in Eq. (3.6). The addition of positivity constraints, on the other hand, suppresses uncertainties mostly in the high- $x$  region, and forces the helicity basis PDFs to remain positive.

As a consequence of the decreasing uncertainties, from the left column to the right in Fig. 5.21 the two helicity basis PDFs  $f^\uparrow$  and  $f^\downarrow$  tend to become more distinguished from each other. Such tendencies, however, are difficult to identify directly in Fig. 5.21, so in order to more quantitatively delineate the different behaviors we will make use of an ‘‘AUC plot’’. Before looking at the specific AUC plot for the helicity PDFs, however, let us demonstrate how an AUC plot works with a simple example.

AUC stands for *area under the curve of ROC*, while ROC means *receiver operating characteristic curve*. In Fig. 5.22, the left column shows two Gaussian distributions  $\mathcal{N}_1$  and  $\mathcal{N}_2$  with different central values  $\mu_1$  and  $\mu_2$ , which need to be discriminated from each other. The right column of Fig. 5.22 plots the cumulative integration values of  $\mathcal{N}_2$  versus those of  $\mathcal{N}_1$ , defined as

$$\int_{-\infty}^x \mathcal{N}_{1,2} dx = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu_{1,2})^2}{2\sigma^2}\right) dx, \quad x \in (-\infty, \infty). \quad (5.2)$$

As shown in the first row of Fig. 5.22, when the two distributions  $\mathcal{N}_1$  and  $\mathcal{N}_2$  overlap entirely with each other, their cumulative integration values are of course identical and therefore give a diagonal line in the  $\int_{-\infty}^x \mathcal{N}_2 dx$  versus  $\int_{-\infty}^x \mathcal{N}_1 dx$  plot (the ROC). In the second row, where the second distribution starts to deviate from the first one, however,

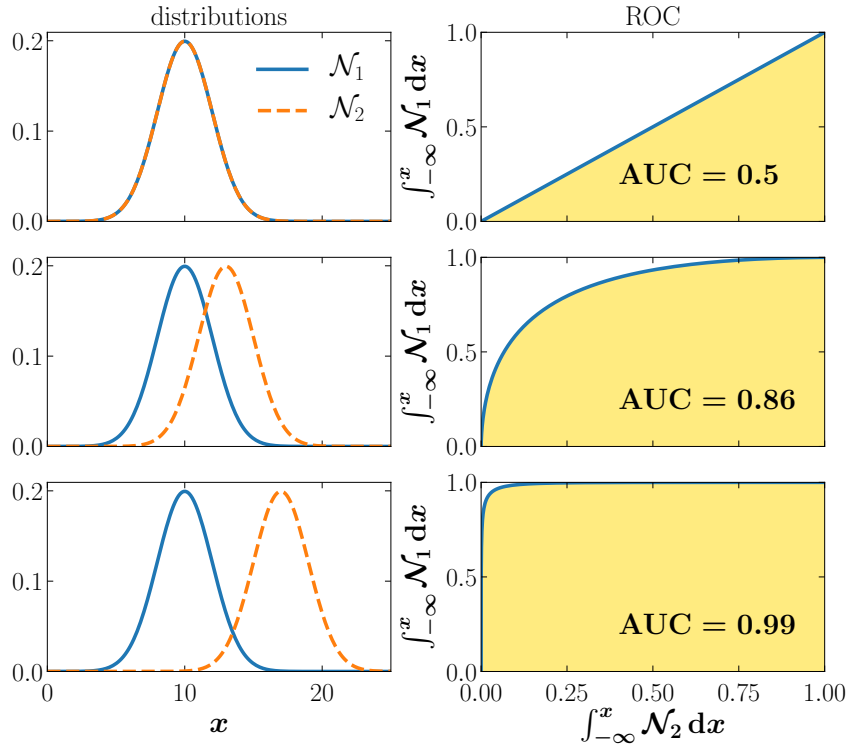


FIG. 5.22: Demonstration of ROC (receiver operating characteristic curve) and AUC (area under the curve of ROC): the left column shows two Gaussian distributions (blue solid and orange dashed curves) needed to be discriminated from each other, and the right column shows the corresponding ROC curves and AUC values.

the ROC is also bent away from the diagonal, because  $\int_{-\infty}^x \mathcal{N}_2 dx$  reaches its maximum value slower than  $\int_{-\infty}^x \mathcal{N}_1 dx$ . In the last row of Fig. 5.22, where the two distributions are almost completely discriminated from each other, the ROC deviates substantially from the diagonal for the same reason.

To quantify the deviation of the ROC from the diagonal, the AUC is calculated along with the plots. When the ROC is curved away from the diagonal, the AUC also deviates from 0.5 and approaches 1. Of course, the ROC could be curved downwards if  $\mathcal{N}_2$  (orange dashed curve) is instead shifted to the left, and in this case the AUC also deviates away from 0.5, but towards 0 (which still indicates a better discrimination). In this sense, the

$1 - \text{AUC}$  is usually used when AUC is less than 0.5, in order to keep the figures intuitive.

For our binary classification problem (discriminating  $f^\uparrow$  from  $f^\downarrow$ ), the AUC plot is made to represent the quality of discrimination as a function of parton momentum fraction  $x$ , and for all scenarios A, B and C, as shown in Fig. 5.23. Following the discussion above, we use the  $1 - \text{AUC}$  when the AUC is less than 0.5.

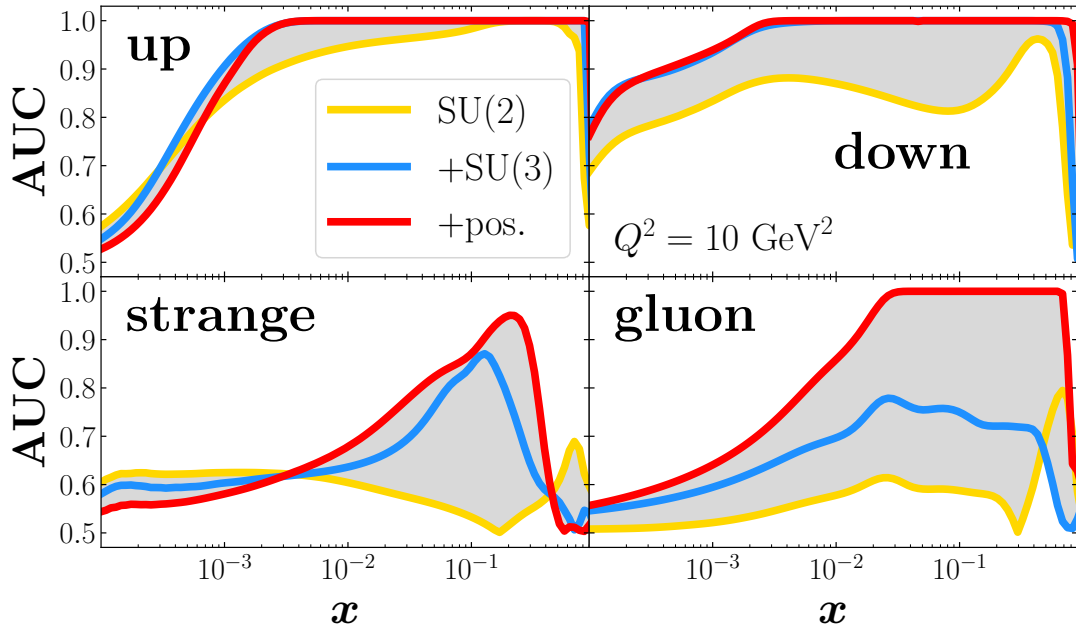


FIG. 5.23: AUC (area under the curve of ROC) is plotted for  $u$ ,  $d$ ,  $s$  and  $g$  at  $Q^2 = 10 \text{ GeV}^2$ . Scenarios A, B and C are colored yellow, blue and red respectively. The more the AUC value approaches 1, the better one can discriminate between  $f^\uparrow$  and  $f^\downarrow$ . On the contrary, the more it approaches 0.5, the harder one can discriminate.

Having explained the meaning of an AUC plot, we next discuss the implications of Fig. 5.23.

Firstly, we note that the discrimination between  $u^\uparrow$  and  $u^\downarrow$  is not improved significantly from the theory inputs, which can be understood since both the spin-averaged and spin-dependent  $u$  distributions are already very well constrained from experiment. The discrimination between  $d^\uparrow$  and  $d^\downarrow$ , on the other hand, receives discernible improvement from the SU(3) flavor symmetry, which results from the reduction of the uncertainty in the

$\Delta d$  PDF (see Fig. 5.14). As for the  $s$ -quark helicity basis PDF, imposing SU(3) flavor symmetry makes a large improvement to their discrimination in the data sensitive region ( $0.01 \lesssim x \lesssim 0.5$ ), resulting from the significant reduction of the uncertainty for  $\Delta s^+$  (see Fig. 5.14).

The positivity constraints make almost no improvement to the discrimination between  $u^\uparrow$  and  $u^\downarrow$  or  $d^\uparrow$  and  $d^\downarrow$  because the helicity basis PDFs of both flavors are already very well separated from each other, and the positivity constraints only reduce their uncertainties. For the  $s$ -quark helicity PDFs, on the other hand, one may expect that given the significant reduction of the uncertainty in  $\Delta s^+$  (see Fig. 5.14) and the less well discriminated  $s^\uparrow$  and  $s^\downarrow$  in scenario B (see Fig. 5.23), the positivity constraints should have discriminated between  $s^\uparrow$  and  $s^\downarrow$  much better. However, because the spin-averaged strange distributions are less well constrained at high  $x$  (see Fig. 5.9), the reduction of the  $\Delta s^+$  uncertainty does not result in better discrimination between  $s^\uparrow$  and  $s^\downarrow$ .

The situation for the gluon distributions has already been discussed extensively in the previous sections; now it is represented by another format in Fig. 5.23. Starting with scenario A, Fig. 5.23 shows that  $g^\uparrow$  and  $g^\downarrow$  can hardly be separated from each other. For scenario B, the reduction of the  $\Delta g$  uncertainties (see Fig. 5.15) improves discernibly the discrimination power. Finally, in scenario C, the discrimination is fully established in the experimentally measured region ( $0.01 \lesssim x \lesssim 0.5$ , see Fig. 4.3), mainly because the positivity constraints eliminate one of the  $\Delta g$  solutions. This brings us back to one of the themes of this work, namely, to study the impact of theory inputs on the extraction of spin-dependent PDFs. What we can infer from the above studies is that the theory inputs, including SU(3) flavor symmetry and positivity constraints, can place significant bias on the extracted parton distributions that are less well constrained by experimental data, namely  $\Delta s$  and  $\Delta g$ .

A final observation that we would like to make from Fig. [5.23](#) is that the quality of the discrimination degrades for all the parton flavors below around  $10^{-2}$  to  $10^{-3}$  as a result of the limited kinematic coverage of the polarized world data. Fortunately, the planned Electron-Ion Collider (EIC) will expand the kinematic coverage for polarized lepton-hadron (and lepton-nucleus) collisions, which should provide stronger experimental constraints on the helicity-dependent distributions. An impact study was conducted recently to assess the improvements in the determination of polarized PDFs (as well as unpolarized PDFs and other types of distributions) from future EIC data, summarized in the *EIC Yellow Report* [33](#). In the next section we will discuss the potential impact of EIC data in more detail.

## 5.5 EIC impact study

As an application of our new JAM analysis, we will present in this section the impact study for future Electron-Ion Collider (EIC) data, and in particular the results of simulations including the parity-conserving  $A_{LL}$  and parity-violating  $A_{UL}$  polarization asymmetry pseudodata and their impact on the quark and gluon helicity distributions. As discussed in Section [4.4.3](#), we consider a total of 6 scenarios for the  $A_{LL}$  and  $A_{UL}$  pseudodata, for each of the low, mid and high extrapolations below  $x \sim 0.01$ , and for both scenarios A and B listed in Table [5.3](#).

In the following we first present results for the constraints on the spin-dependent PDFs from the EIC simulated  $A_{LL}$  asymmetry in Section [5.5.1](#), before discussing the constraints from  $A_{UL}$  in Section [5.5.2](#).

scenario	extrapolation	SU(2)	SU(3)
1	low	✓	
2	mid	✓	
3	high	✓	
4	low	✓	✓
5	mid	✓	✓
6	high	✓	✓

TABLE 5.11: Summary of the 6 scenarios considered in this analysis for the baseline PDFs, with variations of the small- $x$  extrapolation (“low”, “mid”, “high”) and use of standard values  $g_A = 1.269(3)$  and  $a_8 = 0.586(31)$  [76].

### 5.5.1 Constraints from $A_{LL}$ pseudodata

The planned EIC experiments will extend measurements of  $A_{LL}$  down to  $x \approx 2 \times 10^{-4}$ , which is almost 2 orders of magnitude smaller than the range of currently existing data. In estimating the projected uncertainties on the data, a significant extrapolation of the  $g_1$  structure function is therefore necessary into the unmeasured region. The extrapolation uncertainty is illustrated in Fig. 5.24 for the proton  $g_1^p$  structure function at  $Q^2 = 10 \text{ GeV}^2$ , obtained from the JAM baseline results as described in Section 4.4.3. The uncertainty on  $g_1^p$  for  $x \lesssim 10^{-3}$  is quite large, reflecting the absence of constraints from available measurements at low values of  $x$ .

The addition of EIC pseudodata leads to a dramatic reduction of the uncertainties, indicated by the colored bands in Fig. 5.24, which represent extrapolations of  $g_1^p$  according to the  $-1\sigma$  (“low”), central (“mid”), and  $+1\sigma$  (“high”) variations of  $A_{LL}^p$ . The estimated uncertainties in this case are more comparable with the ones in the currently accessible  $x$  region, suggesting the important constraints that can be anticipated from future EIC measurements.

The impact of the EIC  $A_{LL}$  pseudodata on the neutron  $g_1^n$  structure function is illustrated

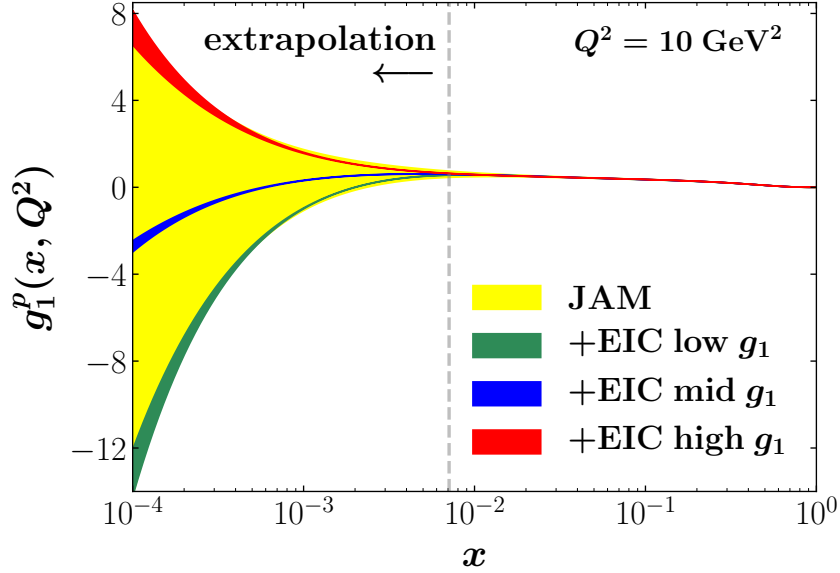


FIG. 5.24: Impact of projected  $A_{LL}^p$  data at EIC kinematics on the proton  $g_1^p$  structure function at  $Q^2 = 10 \text{ GeV}^2$ , with the extrapolated baseline results (yellow band) compared with those including the EIC data for the  $-1\sigma$  (“low”, green band), central (“mid”, blue band), and  $+1\sigma$  (“high”, red band) uncertainties of  $A_{LL}^p$ , for the scenario of imposing both SU(2) and SU(3). The extrapolation region (indicated by the arrow) is to the left of the vertical dashed line at  $x \approx 7 \times 10^{-3}$ . (Figure from Ref. [57].)

in Fig. 5.25 for the central (“mid”) scenario at  $Q^2 = 10 \text{ GeV}^2$ . From the figure one can see that while the proton pseudodata provide some constraints on  $g_1^n$ , further constraints are provided by the deuteron pseudodata, reducing the uncertainties by a factor of 2 – 4 depending on whether SU(3) is imposed or not. The same is observed if  $^3\text{He}$  pseudodata are used instead of deuteron (not shown in the figure). This reduction of uncertainties on  $g_1^n$  is correlated with a reduction of uncertainties on the  $\Delta d$  PDF.

To assess the impact of the EIC pseudodata on the spin carried by quarks and gluons in the proton, it is useful to consider truncated moments of the gluon and quark singlet



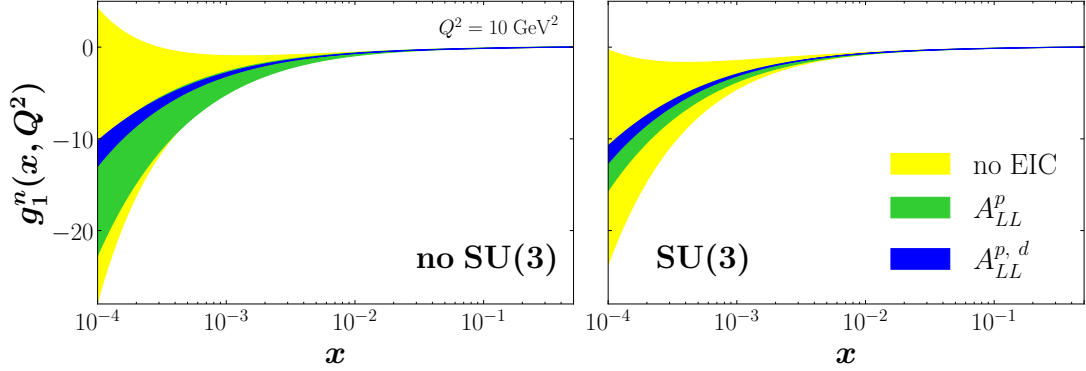


FIG. 5.25: Impact of projected proton  $A_{LL}^p$  and deuteron  $A_{LL}^d$  asymmetry data at EIC kinematics on the neutron  $g_1^n$  structure function at  $Q^2 = 10 \text{ GeV}^2$  for the “mid” scenario. The extrapolated baseline results (yellow bands) are compared with those including EIC proton (green bands) and proton plus deuteron (blue bands) asymmetry pseudodata, for the case of not imposing SU(3) (left panel) and imposing SU(3) (right panel). (Figure from Ref. [57].)

helicity PDFs, defined as

$$\begin{aligned} \Delta G_{\text{trunc}}(Q^2) &= \int_{x_{\text{min}}}^1 \Delta g(x, Q^2) dx, \\ \Delta \Sigma_{\text{trunc}}(Q^2) &= \sum_q \int_{x_{\text{min}}}^1 \Delta q^+(x, Q^2) dx, \end{aligned} \quad (5.3)$$

where the sum extends over the quark flavors  $q = u, d$  and  $s$ , and in the present analysis we take  $x_{\text{min}} = 10^{-4}$ . Comparing the truncated moments and their uncertainties from the fits including the EIC proton  $A_{LL}^p$  pseudodata and those from the baseline set of PDFs, in Fig. [5.26] we show the ratio of uncertainties  $\delta^{\text{EIC}}/\delta$  for both the gluon  $\Delta G_{\text{trunc}}$  and quark singlet  $\Delta \Sigma_{\text{trunc}}$  moment for all the scenarios listed in Table [5.11].

In the most general scenario, where only SU(2) symmetry is imposed via Eq. (3.6), the impact of the  $A_{LL}^p$  EIC pseudodata on  $\Delta G_{\text{trunc}}$  is an  $\approx 60\%$  reduction of the uncertainty relative to the baseline fit uncertainty. For the quark singlet moment  $\Delta \Sigma_{\text{trunc}}$ , on the other

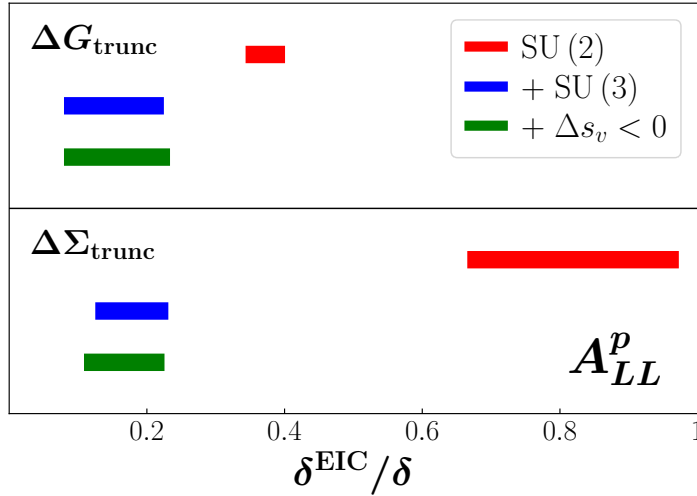


FIG. 5.26: Ratio of uncertainties  $\delta^{\text{EIC}}/\delta$  of the truncated moments of the gluon,  $\Delta G_{\text{trunc}}$  (upper panel), and quark singlet,  $\Delta \Sigma_{\text{trunc}}$  (lower panel), distributions with and without EIC data at  $Q^2 = 10 \text{ GeV}^2$ , for scenarios of imposing only SU(2) (red bands), imposing SU(2) and SU(3) (blue bands), and in addition restricting solutions to ones with negative strangeness in valence region,  $\Delta s_v < 0$  (green bands), using proton  $A_{LL}^p$  EIC data. The ranges of the horizontal bands are obtained using uncertainties from the “low”, “mid” and “high” scenarios in Fig. 5.24. (Figure from Ref. 57.)

hand, there is a much smaller,  $\lesssim 30\%$ , reduction in the uncertainty, which is effectively consistent with no reduction. The ranges of the horizontal bands in Fig. 5.26 are obtained by considering the uncertainties from each of the “low”, “mid” and “high”  $A_{LL}$  scenarios in Fig. 5.24.

The impact of the EIC pseudodata can increase when additional assumptions are made in the analysis. In particular, by imposing SU(3) symmetry via Eq. (3.6) the reduction of uncertainties on  $\Delta G_{\text{trunc}}$  is enhanced from  $\approx 60\%$  to as high as  $80\% - 90\%$ , with an even more dramatic improvement for the quark singlet moment. The reduction of the latter can be understood from the fact that without the SU(3) constraint, both the  $\Delta d^+$  and  $\Delta s^+$  flavors are less well determined, and therefore contribute more to the uncertainty of  $\Delta \Sigma_{\text{trunc}}$ . The gluon distribution  $\Delta g$ , and the corresponding truncated moment  $\Delta G_{\text{trunc}}$ , is

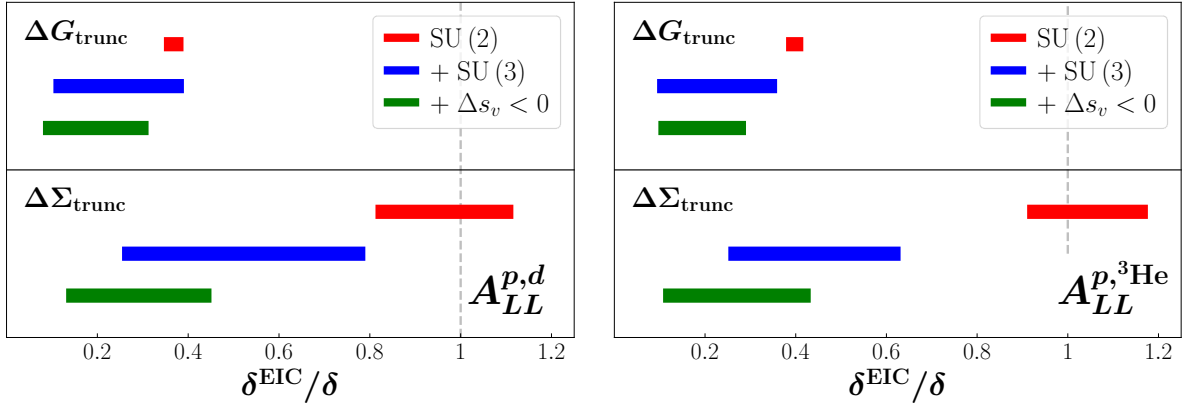


FIG. 5.27: As for Fig. 5.26 but considering the impact of proton  $A_{LL}^p$  and deuteron  $A_{LL}^d$  (left panel), proton  $A_{LL}^p$  and helium  $A_{LL}^{3\text{He}}$  (right panel) EIC pseudodata on the truncated gluon and quark singlet moments. (Figure from Ref. 57.)

less sensitive to SU(3) assumptions, hence the reduction in the uncertainty is more modest.

Note that our Monte Carlo analysis typically contains multiple solutions in parameter space, giving rise to fits with different shapes for poorly constrained distributions, which nevertheless yield essentially identical overall  $\chi^2$  values. This is especially relevant for the strange quark helicity PDF,  $\Delta s^+$ , which can be either positive or negative at intermediate  $x$  values,  $x \sim 0.1 - 0.3$ , depending on whether the fit is constrained by semi-inclusive DIS data or not 8. Typically, solutions with positive strange helicity in the valence region (“ $\Delta s_v > 0$ ”) violate the SU(3) constraint, while the ones with negative strange helicity are more consistent with SU(3). To avoid this violation, we consider in Fig. 5.26 also the scenario of restricting to negative polarized strangeness in the valence region (“ $\Delta s_v < 0$ ”). For the proton  $A_{LL}^p$  pseudodata, however, the removal of the positive strange helicity solutions does not lead to any reduction in the uncertainty, since in this case the positive and negative  $\Delta s_v$  have a very similar effect on  $\Delta\Sigma_{\text{trunc}}$  and its uncertainties.

The effect of inclusion of  $A_{LL}$  pseudodata for polarized deuteron and  $^3\text{He}$  beams is illustrated in Fig. 5.27. Here we observe an even clearer dependence of the impact for  $\Delta\Sigma_{\text{trunc}}$

on the theory assumptions made in the analysis. When only the SU(2) constraint is imposed, no discernible impact on the quark helicity is observed. After imposing SU(3), on the other hand, the impact on  $\Delta\Sigma_{\text{trunc}}$  ranges from 20% – 75%, depending on the low- $x$  extrapolation scenario. If in addition the positive strange helicity solution is removed (due to its relatively large violation of SU(3)), the impact on  $\Delta\Sigma_{\text{trunc}}$  becomes more significant, with 60% – 90% reduction in the uncertainty, and which is also less dependent on the extrapolation.

The impact on the gluon moment  $\Delta G_{\text{trunc}}$  from the inclusion of  $A_{LL}v$  or  $A_{LL}^{3\text{He}}$  pseudodata is similar to the effect of using proton  $A_{LL}^p$  data alone, with  $\approx 60\%$  reduction in uncertainty for the combined  $p + d$  or  $p+^3\text{He}$  analyses. This can be understood from the fact that the gluon contributes to the DIS asymmetry in essentially the same way for  $p$ ,  $d$  or  $^3\text{He}$  beams (appearing only at higher order in  $\alpha_s$ ), so that addition of  $d$  or  $^3\text{He}$  pseudodata does not improve the impact beyond what is already observed for  $p$ . The further addition of SU(3) constraints or removal of  $\Delta s_v > 0$  solutions does not significantly affect the impact on  $\Delta G_{\text{trunc}}$ , since these constraints are largely indirect, with the overall reduction of uncertainties in the range 60% – 90% in either the  $p + d$  or  $p+^3\text{He}$  scenarios.

As we have extensively discussed in Section [5.2.2](#), however, both the SU(3) and  $\Delta s_v < 0$  constraints are less justified than the constraint from SU(2), so that for the scenario that is least biased by theoretical input the impact of EIC  $A_{LL}$  pseudodata is significant only for the gluon truncated moment  $\Delta G_{\text{trunc}}$  and is negligible for  $\Delta\Sigma_{\text{trunc}}$ .

### 5.5.2 Constraints from $A_{UL}$ pseudodata

The impact of the simulation described in Section [4.4.3](#) for the parity-violating proton single-spin asymmetry  $A_{UL}^p$  is shown in Fig. [5.28](#). Interestingly, the situation here is

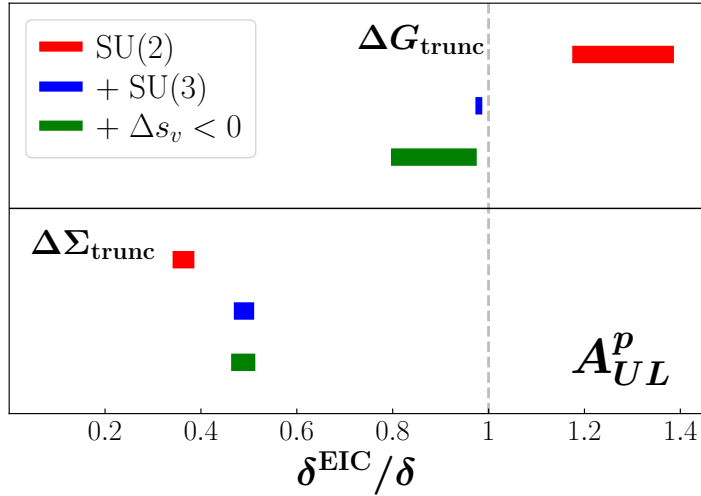


FIG. 5.28: As for Fig. 5.26, but for the proton parity-violating  $A_{UL}^p$  asymmetry. (Figure from Ref. 57.)

somewhat inverted from that found for the  $A_{LL}$  asymmetries in Figs. 5.26 and 5.27. In particular, a strong impact is seen on the quark singlet truncated moment, with  $\approx 50\% - 60\%$  reduction in the uncertainty for all three scenarios considered. This result is in line with the expectation that the  $g_1^{\gamma Z}$  structure function provides the dominant contribution to  $A_{UL}$  (see Eq. 2.24) and weights the different quark flavor contributions approximately equally (see Eq. 2.26). Given that the baseline strange quark helicity distribution has weak constraints from existing data, the new  $A_{UL}$  pseudodata are able to significantly improve the uncertainties on  $\Delta s^+$ , and thus on the quark singlet moment,  $\Delta \Sigma_{\text{trunc}}$ .

On the other hand, no significant improvement is seen for the gluon truncated moment, regardless of the scenario considered. Although the  $g_1^{\gamma Z}$  interference structure function is as sensitive to the gluon distribution as is the electromagnetic  $g_1$  structure function, the relative errors on the parity-violating  $A_{UL}$  asymmetry are much larger than those on existing or projected  $A_{LL}$  data (see Fig. 4.5). It is therefore not surprising that the EIC  $A_{UL}$  data are unable to provide significant new information on the gluon helicity distribution.

In fact, because of statistical fluctuations and the fact that the optimization of  $\chi^2$  is performed on the observables rather than on the PDFs, it is possible in multidimensional fits such as the ones performed here to find an increase in PDF uncertainties in some regions of kinematics with inclusion of additional data [33] (which does not occur at the observable level).

Finally, we note that in the EIC *Yellow Report* [33], the scenario of using SU(2) and SU(3) symmetry constraints from hyperon beta-decay was examined, and was found to have little impact on the quark singlet and gluon moments. In the present, more robust analysis, the addition of a second shape for some of the helicity PDFs and the inclusion of a wider range of solutions for the gluon distribution give rise to an overall less well constrained baseline, and thus to a stronger impact on the quark singlet moment  $\Delta\Sigma_{\text{trunc}}$ .

# CHAPTER 6

## Conclusion and outlook

Although a complete theoretical description of hadron structure is currently very challenging due to the nonperturbative nature of bound state hadrons, factorization of the perturbative and nonperturbative parts in high energy scattering processes allows information on the hadron's quark and gluon dynamics to be extracted from experimental data. The results from global QCD analyses, such as those presented in this work, are therefore extensively used to provide fundamental information on the composition and formation of atomic nuclei. Since the functions that encode the nonperturbative nature of hadrons are, via the QCD factorization theorems, universal, they can be determined from different types of lepton-nucleon and nucleon-nucleon collisions that have sensitivity to various aspects of hadron structure.

Ever since the original EMC result [9] that revealed a surprisingly small spin contribution from the proton's quark constituents, significant efforts have been devoted to understanding the spin structure of the proton. The valence up and down helicity distributions in

the proton have been the first to be constrained using polarized DIS data on protons, deuterons and  $^3\text{He}$  nuclei measured at facilities around the world. Following this, measurements of jet production in polarized  $pp$  collisions were performed to constrain the gluon helicity distribution. By measuring polarized semi-inclusive DIS observables, sea quark polarizations have also been better determined. Moreover, the future Electron-Ion Collider will explore uncharted territory in spin physics, extend the kinematic coverage in both Bjorken- $x$  and momentum exchange  $Q^2$ , and further deepen our understanding of the proton spin structure.

## 6.1 Summary of results

The first part of the results is the analysis of spin-averaged PDFs. Overall, the unpolarized world data are well described by the fitted theoretical predictions from JAM with an overall  $\chi^2_{\text{dat}} = 1.18$ . As the first JAM analysis to include jet production data from unpolarized  $pp$  and  $p\bar{p}$  collisions, and the first ever global QCD analysis that included unpolarized jet data from RHIC [22], we studied their impact on spin-averaged PDFs. We found that the overall jet production data had only slight impact on the already highly constrained spin-averaged PDFs, and that the RHIC unpolarized jet data provided no discernible additional constraints on top of the Tevatron jets [66, 67]. Differences between the JAM PDFs and those from other groups were also found, but were explained as mainly caused by choices of datasets.

With a reliable baseline for the spin-averaged PDFs, we proceeded to the second part of the analysis, which was the extraction of the spin-dependent PDFs from polarized world data. In order to study the potential biases that may originate from theory inputs, we set up three varying scenarios, with SU(2) symmetry only, SU(2) plus SU(3) symmetries, and



SU(2) plus SU(3) plus positivity constraints. A good agreement between the polarized world data and the JAM fitted values was obtained in general, with an overall  $\chi_{\text{dat}}^2$  of 0.94 to 0.98, depending on the various assumptions imposed. The JAM fitted asymmetries plotted with the polarized experiment data also showed a good agreement between the two.

For the spin-dependent PDFs, the uncertainties on the up and down distributions  $\Delta u^+$  and  $\Delta d^+$  were found to be significantly reduced due to the SU(3) flavor symmetry constraint. The positivity constraints then caused a rather smaller reduction of the uncertainties, and a deformation of shapes because we set  $\Delta \bar{u} = \Delta \bar{d} = \Delta \bar{s} = \Delta s$ . For the helicity strange distribution  $\Delta s^+$ , however, significant bias was observed when both the SU(3) flavor symmetry and positivity constraints changed drastically the shape, and reduced the uncertainty of  $\Delta s^+$ . Lastly, the gluon helicity distribution  $\Delta g$  was observed to contain a solution that was mostly above zero (“positive”) and a solution that was mostly below zero (“negative”). While only an indirect effect was found from the SU(3) flavor symmetry, the positivity constraints completely eliminated the “negative” solution. It thus yielded a net positive gluon polarization, which indicated a strong bias acting on the gluon helicity distribution from the positivity constraints.

In addition to using the standard  $g_A$  and  $a_8$  values and uncertainties [76], we also performed a fit using the JAM17 extracted values and uncertainties [8] as a proxy for including polarized SIDIS data. The only noticeable change observed in the spin-dependent PDFs was an expansion of the solution space for  $\Delta s^+$ , which was expected because the SU(3) flavor symmetry was less restrictively enforced by the JAM17 extracted  $a_8$ , that had a larger uncertainty.

The third part of our results was a brief survey of choices made in analyzing the jet

observables. First we examined the choices of factorization and renormalization scales  $\mu_R$  and  $\mu_F$ , where we found that  $\mu_{R,F} = 0.5 p_T$  was an optimal choice for minimizing the  $\chi_{\text{dat}}^2$  for unpolarized jet observables. We then demonstrated that with a  $p_T$  cut at 10 GeV, the tension between the STAR 2003 and 2004 datasets [22] in their overlapping kinematics was resolved. Nevertheless, no perceptible impact on the spin-averaged or spin-dependent PDFs was found due to the choices of factorization and renormalization scales or  $p_T$  cut.

The fourth part of the analysis (and the last part of the JAM analysis on the real world data presented in this work), was the first ever simultaneous analysis of spin-averaged and spin-dependent PDFs. The unpolarized and polarized world data are all well fitted in the simultaneous analysis, with an overall  $\chi_{\text{dat}}^2$  of 1.15 to 1.16, depending on the various assumptions made. This analysis, for the first time ever, extracted quark and gluon PDFs in the helicity basis, using a consistent treatment of uncertainties. In the measured region of the available polarized world data ( $0.01 \lesssim x \lesssim 0.5$ ), the up quark was found to have a net positive polarization, while the down quark had a net negative polarization, both with respect to the proton polarization. For the strange quark, the two helicity basis PDFs  $s^\uparrow$  and  $s^\downarrow$  were less well separated from each other, due to the relatively large uncertainties of both the spin-averaged and spin-dependent strange distributions. The gluon helicity basis PDFs  $g^\uparrow$  and  $g^\downarrow$ , due to the presence of two  $\Delta g$  solutions, were hardly distinguishable from each other within the uncertainties. However, with the “negative” solution removed by the positivity constraints, the two helicity basis PDFs were clearly separated from each other, at the expense of sacrificing the data-driven nature of the global analysis.

From the analyses above, we conclude that the uncertainty bands and shapes of the light quarks helicity basis depend discernibly, or in some cases significantly, on theory inputs, namely the SU(3) flavor symmetry and positivity constraints, both of which can be debatable. Moreover, for gluon distribution, one of the two solutions that are very distinct

from each other was completely ruled out when forcing it to comply with the positivity constraints, despite the fact that it described the data just as well. Since we cannot justify them on purely theoretical grounds, we conclude that significant bias exists when constraints such as SU(3) flavor symmetry and positivity are employed in the extractions of spin-dependent PDFs.

The final part of the results was an application of the JAM analysis to the impact study of projected parity-conserving and parity-violating polarization asymmetries at the EIC. Although previously there have been several studies [97, 98] that assessed the potential impact of the projected DIS data at EIC kinematics on the spin-dependent PDFs, in our analysis [57] the uncertainties on the extrapolation of the projected EIC data into the unmeasured region at low  $x$  was studied for the first time. Moreover, our analysis provided for the first time a complete impact study of the parity-violating polarization asymmetry  $A_{UL}$  in the framework of a global QCD analysis.

For the parity-conserving polarization asymmetry  $A_{LL}$ , the impact of the future EIC data on the gluon truncated moment was found to be significant and only weakly dependent on the theory inputs, while the impact on quark singlet truncated moment was largely influenced by the theory inputs. For the parity-violating polarization asymmetry  $A_{UL}$ , the situation was inverted, with a less biased and more discernible impact found for the quark singlet truncated moment, but a negligible impact observed for the gluon truncated moment. The improvement in the impact on the quark singlet truncated moment stems from the fact that the leading contribution from the  $g_1^{\gamma Z}$  structure function weighted different quark flavors approximately equally. The gluon truncated moment, on the other hand, was not better constrained because the relative errors on  $A_{UL}$  are significantly larger compared with those on existing or projected  $A_{LL}$  data.

## 6.2 Future of global QCD analysis

As more high precision experimental data become available, uncertainties for various sea quark and gluon distributions are expected to decrease, and eventually a refined picture of collinear spin structure of the proton will emerge. With the inclusion of jet production data from polarized  $pp$  scattering, we were able to constrain greatly the helicity gluon distribution. However, improvements can still be made towards constraining the sea quark distributions. Specifically, polarized SIDIS data will help determine the weakly constrained  $\Delta s^+$  PDF in current global QCD analysis. Furthermore, measurements of the longitudinal single-spin asymmetries in  $W^\pm$  production from polarized  $pp$  scattering could provide enhanced sensitivity to the light antiquark helicities,  $\Delta\bar{u}$  and  $\Delta\bar{d}$ . Lastly, with the future EIC formally approved by the U.S. Department of Energy, previously unexplored kinematic regions will be accessible, and a number of new observables will become available. As we have presented in Chapter 5, both the parity-conserving and parity-violating polarization asymmetries at EIC kinematics will help to constrain the helicity PDFs at  $x \lesssim 0.01$ . In the future, all such observables will be included in the JAM analyses.

On the theoretical front, there are several enhancements that are necessary for the description of increasingly precise data. To improve the precision of perturbative calculations, one can include the NNLO corrections to the jet formalism when these become available [99]. To describe DIS and DY observables at high  $x$ , resummation of large logarithms is also necessary [100]. Finally, since in our analyses the zero-mass variable-flavor-number-scheme (ZM-VFNS) [83] is used, special treatment of heavy quark mass effects will be needed to describe heavy quark production. The ACOT renormalization scheme [101] is known to describe well the HERA data at low  $Q^2$ , where corrections proportional to  $M^2/Q^2$  become important. This is particularly useful in determining the  $F_L$  structure function, which vanishes at leading order and is thus more sensitive to higher order corrections.

Apart from the addition of experimental observables and theoretical advances, global QCD analysis also relies on the use of robust methodology. Although the single  $\chi^2$  minimization approach has been successful in extracting nonperturbative structures in the past, such techniques are not well suited for dealing with gigantic parameter spaces or thousands of computationally expensive observable calculations. In this case, Bayesian inference and Monte Carlo sampling have to be used for iteratively improving fits and thoroughly exploring the parameter space. Additionally, using fast numerical calculation methods, such as Mellin space techniques and interpolation, the computational complexity is drastically reduced, and studies of various scenarios such as those explored in this work are possible. Employing these more advanced and robust methods, a reliable set of solutions can be extracted with faithful error estimates. In the future, machine learning can be used to produce the Mellin tables or even predict cross sections from kinematic variables, saving considerable memory and CPU time resources.

The methodology of this analysis is based on the Monte Carlo framework developed in recent years with the JAM15 [55], JAM16 [60] and JAM17 [8] analyses, and the multi-step strategy devised in the JAM19 study [61]. The aim has been to construct a methodology that is robust and reliable in extracting both the spin-averaged and spin-dependent PDFs and estimating their uncertainties. Moreover, this work simultaneously extracted the spin-averaged and spin-dependent PDFs, and its success was a unique test of the universality of PDFs. Although the spin-dependent PDFs are not yet precise enough to receive significant impact from the spin-averaged PDFs, it is certainly only a matter of time that simultaneous analysis will become indispensable in the future once more high precision data become available.

# APPENDIX A

## Partonic jet tree level diagrams

Elaborating on the discussion in Section [2.3.1](#), we present in this appendix the tree level diagrams for the partonic level interactions in single jet production, in particular, the  $\hat{\sigma}_{ab}$  part in Fig. [2.4](#).

There are in total 9 types of partonic interactions at tree level,

$$\begin{aligned} qq &\rightarrow qq, & qq' &\rightarrow qq', & q\bar{q}' &\rightarrow q\bar{q}', \\ q\bar{q} &\rightarrow q'\bar{q}', & q\bar{q} &\rightarrow q\bar{q}, & q\bar{q} &\rightarrow gg, \\ qq &\rightarrow qq, & gg &\rightarrow q\bar{q}, & gg &\rightarrow gg, \end{aligned} \tag{A.1}$$

where  $q$  and  $q'$  denote quarks of different flavors, and  $\bar{q}$  and  $g$  represent antiquark and gluon, respectively. Depending on the incoming partons, these interactions can be classified into 6 different channels, namely, the  $qq$ ,  $qq'$ ,  $q\bar{q}'$ ,  $q\bar{q}$ ,  $qg$  and  $gg$  channels.

Among the different types of interactions, some may contain  $s$ ,  $t$  or  $u$  channels, with  $s$ ,  $t$

and  $u$  being the Mandelstam variables. Therefore, to keep the discussion from becoming lengthy and tedious, we will only show one such channel for each type of interactions for illustration.

In Fig. [A.1](#), we show the interactions from the first row of Eq. [\(A.1\)](#), where only the  $t$  channel is shown.

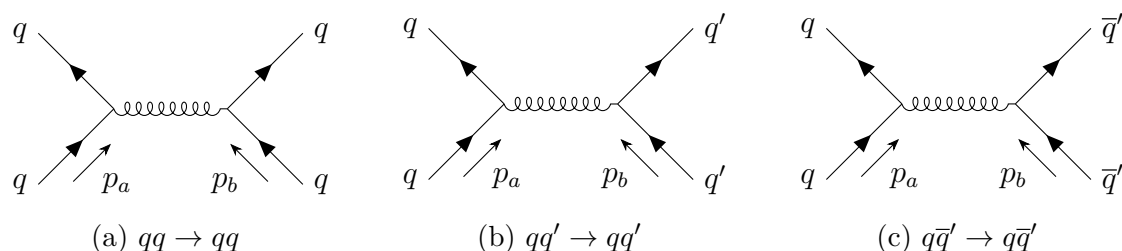


FIG. A.1: Tree level partonic interactions for single jet production, including  $qq \rightarrow qq$ ,  $qq' \rightarrow qq'$  and  $q\bar{q} \rightarrow q\bar{q}$  types, with only the  $t$  channel shown.

In Fig. [A.2](#) we show all  $q\bar{q}$  channel interactions, which is the second row of Eq. [\(A.1\)](#), and again, only the  $s$  channel is shown.

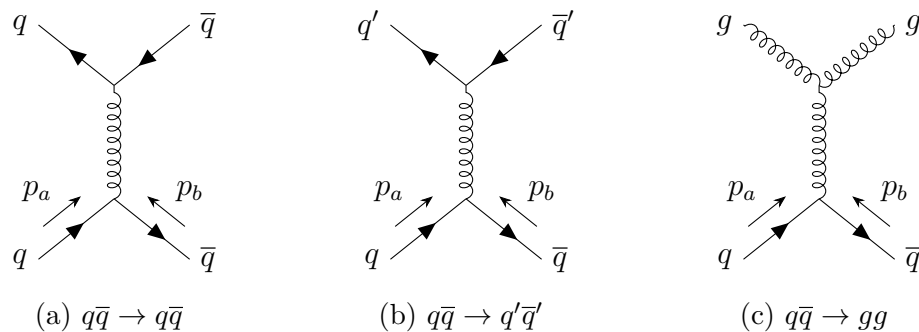


FIG. A.2: Tree level partonic interactions for single jet production, including  $q\bar{q} \rightarrow q\bar{q}$ ,  $q\bar{q} \rightarrow q'\bar{q}$  and  $q\bar{q} \rightarrow gg$  types, with only the  $s$  channel shown.

In Fig. [A.3](#) we have the  $qg \rightarrow qg$  type of interaction. This time two figures are shown, both from the  $t$  channel, but interacting via different types of vertices.

Finally, in Fig. [A.4](#) we have  $gg \rightarrow q\bar{q}$  and  $gg \rightarrow gg$  types of interactions. This time only

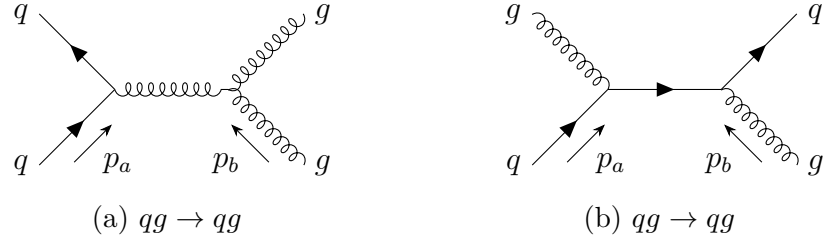


FIG. A.3: Tree level partonic interactions for single jet production, including only  $qg \rightarrow qg$  type, with only the  $t$  channel shown.

the  $t$  channels are shown (which is, however, not relevant for Fig. A.4c).

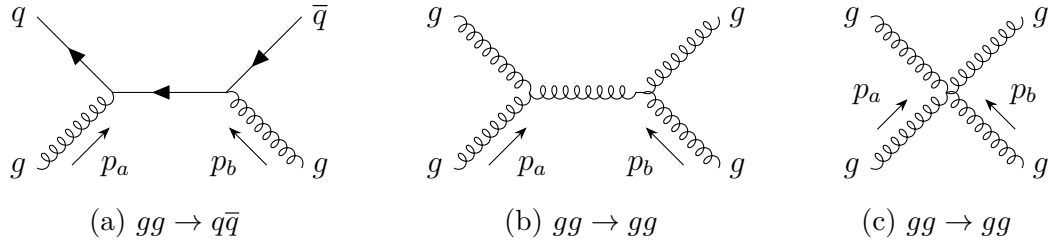


FIG. A.4: Tree level partonic interactions for single jet production, including  $gq \rightarrow q\bar{q}$  and  $gg \rightarrow gg$  types, only  $t$  channel is shown (when applicable).

Interactions with QED couplings do exist, but with the electromagnetic fine structure constant  $\alpha$  much smaller than the strong coupling constant  $\alpha_s$ , their contributions can be neglected.



## Bibliography

- [1] C. Patrignani et al. In: *Chin. Phys. C* 40.10 (2016), p. 100001. DOI: [10.1088/1674-1137/40/10/100001](https://doi.org/10.1088/1674-1137/40/10/100001) (cit. on p. [3](#)).
- [2] John C. Collins, Davison E. Soper, and George F. Sterman. In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: [10.1142/9789814503266\\_0001](https://doi.org/10.1142/9789814503266_0001) (cit. on p. [4](#)).
- [3] P. Jimenez-Delgado, W. Melnitchouk, and J. F. Owens. In: *J. Phys. G* 40 (2013), p. 093102. DOI: [10.1088/0954-3899/40/9/093102](https://doi.org/10.1088/0954-3899/40/9/093102) (cit. on p. [5](#)).
- [4] E. Perez and E. Rizvi. In: *Rept. Prog. Phys.* 76 (2013), p. 046201. DOI: [10.1088/0034-4885/76/4/046201](https://doi.org/10.1088/0034-4885/76/4/046201) (cit. on p. [5](#)).
- [5] Stefano Forte and Graeme Watt. In: *Ann. Rev. Nucl. Part. Sci.* 63 (2013), pp. 291–328. DOI: [10.1146/annurev-nucl-102212-170607](https://doi.org/10.1146/annurev-nucl-102212-170607) (cit. on p. [5](#)).
- [6] Johannes Blumlein. In: *Prog. Part. Nucl. Phys.* 69 (2013), pp. 28–84. DOI: [10.1016/j.pnpnp.2012.09.006](https://doi.org/10.1016/j.pnpnp.2012.09.006) (cit. on p. [5](#)).
- [7] Daniel de Florian, Rodolfo Sassot, et al. In: *Phys. Rev. Lett.* 113.1 (2014), p. 012001. DOI: [10.1103/PhysRevLett.113.012001](https://doi.org/10.1103/PhysRevLett.113.012001) (cit. on pp. [5](#), [7](#), [8](#), [74](#), [85](#)).
- [8] J. J. Ethier, N. Sato, and W. Melnitchouk. In: *Phys. Rev. Lett.* 119.13 (2017), p. 132001. DOI: [10.1103/PhysRevLett.119.132001](https://doi.org/10.1103/PhysRevLett.119.132001) (cit. on pp. [5](#), [7](#), [17](#), [20](#), [39](#), [45](#), [75](#), [86](#), [102](#), [108](#), [112](#)).
- [9] J. Ashman et al. In: *Phys. Lett. B* 206 (1988), p. 364. DOI: [10.1016/0370-2693\(88\)91523-7](https://doi.org/10.1016/0370-2693(88)91523-7) (cit. on pp. [6](#), [106](#)).

- [10] John R. Ellis and Robert L. Jaffe. In: *Phys. Rev. D* 9 (1974). [Erratum: [Phys.Rev.D 10, 1669 \(1974\)](#)], p. 1444. DOI: [10.1103/PhysRevD.9.1444](#) (cit. on p. [6](#)).
- [11] T. Hatsuda and I. Zahed. In: *Phys. Lett. B* 221 (1989), pp. 173–176. DOI: [10.1016/0370-2693\(89\)91493-7](#) (cit. on p. [6](#)).
- [12] Guido Altarelli and Graham G. Ross. In: *Phys. Lett. B* 212 (1988), pp. 391–396. DOI: [10.1016/0370-2693\(88\)91335-4](#) (cit. on p. [6](#)).
- [13] Harald Fritzsch. In: *Phys. Lett. B* 229 (1989), pp. 122–126. DOI: [10.1016/0370-2693\(89\)90168-8](#) (cit. on p. [6](#)).
- [14] M. Anselmino, A. Efremov, and E. Leader. In: *Phys. Rep.* 261 (1995). [Erratum: *Phys.Rept.* 281, 399–400 (1997)], p. 1. DOI: [10.1016/0370-1573\(95\)00011-5](#). arXiv: [hep-ph/9501369](#) (cit. on p. [6](#)).
- [15] Bodo Lampe and Ewald Reya. In: *Phys. Rep.* 332 (2000), p. 1. DOI: [10.1016/S0370-1573\(99\)00100-3](#). arXiv: [hep-ph/9810270](#) (cit. on p. [6](#)).
- [16] Christine A. Aidala, Steven D. Bass, et al. In: *Rev. Mod. Phys.* 85 (2013), p. 655. DOI: [10.1103/RevModPhys.85.655](#). arXiv: [1209.2803 \[hep-ph\]](#) (cit. on p. [6](#)).
- [17] B. Adeva et al. In: *Phys. Lett. B* 369 (1996). Ed. by B. Frois, V. W. Hughes, and N. De Groot, pp. 93–100. DOI: [10.1016/0370-2693\(95\)01584-1](#) (cit. on p. [6](#)).
- [18] B. Adeva et al. In: *Phys. Lett. B* 420 (1998), pp. 180–190. DOI: [10.1016/S0370-2693\(97\)01546-3](#). arXiv: [hep-ex/9711008](#) (cit. on p. [6](#)).
- [19] A. Airapetian et al. In: *Phys. Rev. D* 71 (2005), p. 012003. DOI: [10.1103/PhysRevD.71.012003](#). arXiv: [hep-ex/0407032](#) (cit. on p. [6](#)).
- [20] M. Alekseev et al. In: *Phys. Lett. B* 680 (2009), pp. 217–224. DOI: [10.1016/j.physletb.2009.08.065](#). arXiv: [0905.2828 \[hep-ex\]](#) (cit. on p. [6](#)).
- [21] M. G. Alekseev et al. In: *Phys. Lett. B* 693 (2010), pp. 227–235. DOI: [10.1016/j.physletb.2010.08.034](#). arXiv: [1007.4061 \[hep-ex\]](#) (cit. on p. [6](#)).

- [22] B. I. Abelev et al. In: *Phys. Rev. Lett.* 97 (2006), p. 252001. DOI: [10.1103/PhysRevLett.97.252001](https://doi.org/10.1103/PhysRevLett.97.252001) (cit. on pp. [7](#), [25](#), [28](#), [29](#), [53](#), [54](#), [58](#), [62](#), [64](#), [70](#)–[72](#), [76](#), [78](#), [80](#), [83](#), [88](#)–[91](#), [107](#), [109](#)).
- [23] L. Adamczyk et al. In: *Phys. Rev. D* 86 (2012), p. 032006. DOI: [10.1103/PhysRevD.86.032006](https://doi.org/10.1103/PhysRevD.86.032006) (cit. on pp. [7](#), [28](#), [29](#), [54](#), [55](#), [58](#), [76](#), [78](#), [80](#), [83](#), [89](#)).
- [24] L. Adamczyk et al. In: *Phys. Rev. Lett.* 115.9 (2015), p. 092002. DOI: [10.1103/PhysRevLett.115.092002](https://doi.org/10.1103/PhysRevLett.115.092002) (cit. on pp. [7](#), [28](#), [30](#), [54](#), [55](#), [58](#), [76](#), [78](#), [80](#), [83](#), [89](#)).
- [25] J. Adam et al. In: *Phys. Rev. D* 100.5 (2019), p. 052005. DOI: [10.1103/PhysRevD.100.052005](https://doi.org/10.1103/PhysRevD.100.052005) (cit. on pp. [7](#), [28](#), [30](#), [54](#), [58](#), [76](#), [78](#), [80](#), [83](#), [89](#)).
- [26] Mohamed Abdallah et al. In: *Phys. Rev. D* 103.9 (2021), p. L091103. DOI: [10.1103/PhysRevD.103.L091103](https://doi.org/10.1103/PhysRevD.103.L091103), arXiv: [2103.05571 \[hep-ex\]](https://arxiv.org/abs/2103.05571) (cit. on pp. [7](#), [28](#), [30](#), [54](#), [55](#), [58](#), [76](#), [78](#), [80](#), [83](#), [89](#)).
- [27] A. Adare et al. In: *Phys. Rev. D* 84 (2011), p. 012006. DOI: [10.1103/PhysRevD.84.012006](https://doi.org/10.1103/PhysRevD.84.012006) (cit. on pp. [7](#), [28](#), [29](#), [54](#), [58](#), [76](#), [78](#), [80](#), [83](#), [89](#)).
- [28] Daniel De Florian, Gonzalo Agustín Lucero, et al. In: *Phys. Rev. D* 100.11 (2019), p. 114027. DOI: [10.1103/PhysRevD.100.114027](https://doi.org/10.1103/PhysRevD.100.114027) (cit. on pp. [7](#), [8](#)).
- [29] Xiang-Dong Ji. In: *Phys. Rev. Lett.* 78 (1997), pp. 610–613. DOI: [10.1103/PhysRevLett.78.610](https://doi.org/10.1103/PhysRevLett.78.610), arXiv: [hep-ph/9603249](https://arxiv.org/abs/hep-ph/9603249) (cit. on p. [7](#)).
- [30] R. L. Jaffe and Aneesh Manohar. In: *Nucl. Phys. B* 337 (1990), pp. 509–546. DOI: [10.1016/0550-3213\(90\)90506-9](https://doi.org/10.1016/0550-3213(90)90506-9) (cit. on p. [7](#)).
- [31] C. Alexandrou, M. Constantinou, et al. In: *Phys. Rev. Lett.* 119.14 (2017), p. 142002. DOI: [10.1103/PhysRevLett.119.142002](https://doi.org/10.1103/PhysRevLett.119.142002) (cit. on p. [8](#)).
- [32] Justin Cammarota, Leonard Gamberg, et al. In: *Phys. Rev. D* 102.5 (2020), p. 054002. DOI: [10.1103/PhysRevD.102.054002](https://doi.org/10.1103/PhysRevD.102.054002), arXiv: [2002.08384 \[hep-ph\]](https://arxiv.org/abs/2002.08384) (cit. on p. [8](#)).
- [33] R. Abdul Khalek et al. In: (Mar. 2021). arXiv: [2103.05419 \[physics.ins-det\]](https://arxiv.org/abs/2103.05419) (cit. on pp. [8](#), [97](#), [105](#)).

- [34] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics*. First Edition. Cambridge University Press, 2003. ISBN: 9780511628788 (cit. on p. [14](#)).
- [35] A. C. Benvenuti et al. In: *Phys. Lett. B* 223 (1989), pp. 485–489. DOI: [10.1016/0370-2693\(89\)91637-7](#) (cit. on pp. [14](#), [52](#), [58](#), [64](#)–[66](#)).
- [36] L. W. Whitlow, E. M. Riordan, et al. In: *Phys. Lett. B* 282 (1992), pp. 475–482. DOI: [10.1016/0370-2693\(92\)90672-Q](#) (cit. on pp. [14](#), [52](#), [58](#), [64](#)–[66](#)).
- [37] M. Arneodo et al. In: *Nucl. Phys. B* 483 (1997), pp. 3–43. DOI: [10.1016/S0550-3213\(96\)00538-X](#). arXiv: [hep-ph/9610231](#) (cit. on pp. [14](#), [52](#), [58](#), [64](#)–[66](#)).
- [38] M. Arneodo et al. In: *Nucl. Phys. B* 487 (1997), pp. 3–26. DOI: [10.1016/S0550-3213\(96\)00673-6](#). arXiv: [hep-ex/9611022](#) (cit. on pp. [14](#), [52](#), [58](#), [64](#), [66](#), [73](#)).
- [39] H. Abramowicz et al. In: *Eur. Phys. J. C* 75.12 (2015), p. 580. DOI: [10.1140/epjc/s10052-015-3710-4](#). arXiv: [1506.06042 \[hep-ex\]](#) (cit. on pp. [14](#), [52](#), [58](#), [64](#), [66](#)–[68](#)).
- [40] S. Wandzura and Frank Wilczek. In: *Phys. Lett. B* 72 (1977), pp. 195–198. DOI: [10.1016/0370-2693\(77\)90700-6](#) (cit. on p. [16](#)).
- [41] J. Ashman et al. In: *Nucl. Phys. B* 328 (1989). Ed. by V. W. Hughes and C. Cavata, p. 1. DOI: [10.1016/0550-3213\(89\)90089-8](#) (cit. on pp. [17](#), [54](#), [58](#), [75](#)–[77](#)).
- [42] B. Adeva et al. In: *Phys. Rev. D* 58 (1998), p. 112001. DOI: [10.1103/PhysRevD.58.112001](#) (cit. on pp. [17](#), [54](#), [58](#), [75](#)–[78](#)).
- [43] B. Adeva et al. In: *Phys. Rev. D* 60 (1999). [Erratum: [Phys.Rev.D 62, 079902 \(2000\)](#)], p. 072004. DOI: [10.1103/PhysRevD.60.072004](#) (cit. on pp. [17](#), [54](#), [58](#), [75](#)–[78](#)).
- [44] M. G. Alekseev et al. In: *Phys. Lett. B* 690 (2010), pp. 466–472. DOI: [10.1016/j.physletb.2010.05.069](#). arXiv: [1001.4654 \[hep-ex\]](#) (cit. on pp. [17](#), [54](#), [58](#), [75](#)–[77](#)).

- [45] V. Yu. Alexakhin et al. In: *Phys. Lett. B* 647 (2007), pp. 8–17. DOI: [10.1016/j.physletb.2006.12.076](https://doi.org/10.1016/j.physletb.2006.12.076). arXiv: [hep-ex/0609038](https://arxiv.org/abs/hep-ex/0609038) (cit. on pp. [17](#), [54](#), [58](#), [76](#)–[78](#)).
- [46] C. Adolph et al. In: *Phys. Lett. B* 753 (2016), pp. 18–28. DOI: [10.1016/j.physletb.2015.11.064](https://doi.org/10.1016/j.physletb.2015.11.064). arXiv: [1503.08935 \[hep-ex\]](https://arxiv.org/abs/1503.08935) (cit. on pp. [17](#), [54](#), [58](#), [75](#)–[77](#)).
- [47] Guenter Baum et al. In: *Phys. Rev. Lett.* 51 (1983). Ed. by V. W. Hughes and C. Cavata, p. 1135. DOI: [10.1103/PhysRevLett.51.1135](https://doi.org/10.1103/PhysRevLett.51.1135) (cit. on pp. [17](#), [54](#), [58](#), [75](#)–[77](#)).
- [48] P. L. Anthony et al. In: *Phys. Rev. D* 54 (1996), pp. 6620–6650. DOI: [10.1103/PhysRevD.54.6620](https://doi.org/10.1103/PhysRevD.54.6620). arXiv: [hep-ex/9610007](https://arxiv.org/abs/hep-ex/9610007) (cit. on pp. [17](#), [54](#), [58](#), [76](#), [78](#), [79](#)).
- [49] K. Abe et al. In: *Phys. Rev. D* 58 (1998), p. 112003. DOI: [10.1103/PhysRevD.58.112003](https://doi.org/10.1103/PhysRevD.58.112003). arXiv: [hep-ph/9802357](https://arxiv.org/abs/hep-ph/9802357) (cit. on pp. [17](#), [54](#), [58](#), [75](#)–[78](#)).
- [50] K. Abe et al. In: *Phys. Rev. Lett.* 79 (1997). Ed. by B. Frois, V. W. Hughes, and N. De Groot, pp. 26–30. DOI: [10.1103/PhysRevLett.79.26](https://doi.org/10.1103/PhysRevLett.79.26). arXiv: [hep-ex/9705012](https://arxiv.org/abs/hep-ex/9705012) (cit. on pp. [17](#), [54](#), [58](#), [76](#), [78](#), [79](#)).
- [51] P. L. Anthony et al. In: *Phys. Lett. B* 493 (2000), pp. 19–28. DOI: [10.1016/S0370-2693\(00\)01014-5](https://doi.org/10.1016/S0370-2693(00)01014-5). arXiv: [hep-ph/0007248](https://arxiv.org/abs/hep-ph/0007248) (cit. on pp. [17](#), [54](#), [58](#), [75](#)–[77](#)).
- [52] P. L. Anthony et al. In: *Phys. Lett. B* 463 (1999), pp. 339–345. DOI: [10.1016/S0370-2693\(99\)00940-5](https://doi.org/10.1016/S0370-2693(99)00940-5). arXiv: [hep-ex/9904002](https://arxiv.org/abs/hep-ex/9904002) (cit. on pp. [17](#), [54](#), [58](#), [76](#)–[78](#)).
- [53] K. Ackerstaff et al. In: *Phys. Lett. B* 404 (1997). Ed. by B. Frois, V. W. Hughes, and N. De Groot, pp. 383–389. DOI: [10.1016/S0370-2693\(97\)00611-4](https://doi.org/10.1016/S0370-2693(97)00611-4). arXiv: [hep-ex/9703005](https://arxiv.org/abs/hep-ex/9703005) (cit. on pp. [17](#), [54](#), [58](#), [76](#), [78](#), [79](#)).
- [54] A. Airapetian et al. In: *Phys. Rev. D* 75 (2007), p. 012007. DOI: [10.1103/PhysRevD.75.012007](https://doi.org/10.1103/PhysRevD.75.012007). arXiv: [hep-ex/0609039](https://arxiv.org/abs/hep-ex/0609039) (cit. on pp. [17](#), [54](#), [58](#), [75](#)–[78](#)).
- [55] Nobuo Sato, W. Melnitchouk, et al. In: *Phys. Rev. D* 93.7 (2016), p. 074005. DOI: [10.1103/PhysRevD.93.074005](https://doi.org/10.1103/PhysRevD.93.074005). arXiv: [1601.07782 \[hep-ph\]](https://arxiv.org/abs/1601.07782) (cit. on pp. [17](#), [45](#), [112](#)).

- [56] A. Accardi et al. In: *Eur. Phys. J. A* 52.9 (2016). Ed. by A. Deshpande, Z. E. Meziani, and J. W. Qiu, p. 268. DOI: [10.1140/epja/i2016-16268-9](https://doi.org/10.1140/epja/i2016-16268-9). arXiv: [1212.1701 \[nucl-ex\]](https://arxiv.org/abs/1212.1701) (cit. on pp. [18](#), [56](#)).
- [57] Y. Zhou, C. Cocuzza, et al. In: (May 2021). arXiv: [2105.04434 \[hep-ph\]](https://arxiv.org/abs/2105.04434) (cit. on pp. [18](#), [56](#), [57](#), [60](#), [99](#)–[102](#), [104](#), [110](#)).
- [58] Joshua Allen Magee. “A Measurement of the Parity-Violating Asymmetry in Aluminum and its Contribution to A Measurement of the Proton’s Weak Charge”. William & Mary, 2016. DOI: [10.21220/S22012](https://doi.org/10.21220/S22012) (cit. on p. [19](#)).
- [59] Y. X. Zhao, A. Deshpande, et al. In: *Eur. Phys. J. A* 53.3 (2017), p. 55. DOI: [10.1140/epja/i2017-12245-2](https://doi.org/10.1140/epja/i2017-12245-2). arXiv: [1612.06927 \[nucl-ex\]](https://arxiv.org/abs/1612.06927) (cit. on pp. [19](#), [20](#)).
- [60] Nobuo Sato, J. J. Ethier, et al. In: *Phys. Rev. D* 94.11 (2016), p. 114004. DOI: [10.1103/PhysRevD.94.114004](https://doi.org/10.1103/PhysRevD.94.114004), arXiv: [1609.00899 \[hep-ph\]](https://arxiv.org/abs/1609.00899) (cit. on pp. [20](#), [45](#), [112](#)).
- [61] N. Sato, C. Andres, et al. In: *Phys. Rev. D* 101.7 (2020), p. 074020. DOI: [10.1103/PhysRevD.101.074020](https://doi.org/10.1103/PhysRevD.101.074020). arXiv: [1905.03788 \[hep-ph\]](https://arxiv.org/abs/1905.03788) (cit. on pp. [20](#), [46](#), [73](#), [112](#)).
- [62] S. D. Drell and Tung-Mow Yan. In: *Phys. Rev. Lett.* 25 (1970). [Erratum: [Phys.Rev.Lett. 25, 902 \(1970\)](https://doi.org/10.1103/PhysRevLett.25.902)], pp. 316–320. DOI: [10.1103/PhysRevLett.25.316](https://doi.org/10.1103/PhysRevLett.25.316) (cit. on p. [21](#)).
- [63] Johannes Blumlein and Vajravelu Ravindran. In: *Nucl. Phys. B* 716 (2005), pp. 128–172. DOI: [10.1016/j.nuclphysb.2005.03.003](https://doi.org/10.1016/j.nuclphysb.2005.03.003). arXiv: [hep-ph/0501178](https://arxiv.org/abs/hep-ph/0501178) (cit. on p. [21](#)).
- [64] C. Grosso-Pilcher and M. J. Shochet. In: *Ann. Rev. Nucl. Part. Sci.* 36 (1986), pp. 1–28. DOI: [10.1146/annurev.ns.36.120186.000245](https://doi.org/10.1146/annurev.ns.36.120186.000245) (cit. on p. [22](#)).
- [65] E. A. Hawker et al. In: *Phys. Rev. Lett.* 80 (1998), pp. 3715–3718. DOI: [10.1103/PhysRevLett.80.3715](https://doi.org/10.1103/PhysRevLett.80.3715). arXiv: [hep-ex/9803011](https://arxiv.org/abs/hep-ex/9803011) (cit. on pp. [23](#), [53](#), [58](#), [64](#), [69](#), [72](#)).

- [66] V. M. Abazov et al. In: *Phys. Rev. Lett.* 101 (2008), p. 062001. DOI: [10.1103/PhysRevLett.101.062001](https://doi.org/10.1103/PhysRevLett.101.062001). arXiv: [0802.2400 \[hep-ex\]](https://arxiv.org/abs/0802.2400) (cit. on pp. [25](#), [29](#), [51](#), [53](#), [58](#), [64](#), [70–72](#), [88](#), [91](#), [107](#)).
- [67] A. Abulencia et al. In: *Phys. Rev. D* 75 (2007). [Erratum: [Phys.Rev.D 75, 119901 \(2007\)](https://doi.org/10.1103/PhysRevD.75.119901)], p. 092006. DOI: [10.1103/PhysRevD.75.092006](https://doi.org/10.1103/PhysRevD.75.092006). arXiv: [hep-ex/0701051](https://arxiv.org/abs/hep-ex/0701051) (cit. on pp. [25](#), [30](#), [51](#), [53](#), [54](#), [58](#), [64](#), [70–72](#), [88](#), [91](#), [107](#)).
- [68] B. Jäger, M. Stratmann, and W. Vogelsang. In: *Phys. Rev. D* 70 (2004), p. 034010. DOI: [10.1103/PhysRevD.70.034010](https://doi.org/10.1103/PhysRevD.70.034010). arXiv: [hep-ph/0404057](https://arxiv.org/abs/hep-ph/0404057) (cit. on pp. [27](#), [28](#)).
- [69] Zhong-Bo Kang, Felix Ringer, and Wouter J. Waalewijn. In: *JHEP* 07 (2017), p. 064. DOI: [10.1007/JHEP07\(2017\)064](https://doi.org/10.1007/JHEP07(2017)064). arXiv: [1705.05375 \[hep-ph\]](https://arxiv.org/abs/1705.05375) (cit. on p. [28](#)).
- [70] Werner Vogelsang. *private communication*. 2021 (cit. on p. [28](#)).
- [71] B. I. Abelev et al. In: *Phys. Rev. Lett.* 100 (2008), p. 232003. DOI: [10.1103/PhysRevLett.100.232003](https://doi.org/10.1103/PhysRevLett.100.232003). arXiv: [0710.2048 \[hep-ex\]](https://arxiv.org/abs/0710.2048) (cit. on p. [28](#)).
- [72] Gerald C. Blazey et al. “Run II jet physics”. In: *Physics at Run II: QCD and Weak Boson Physics Workshop: Final General Meeting*. May 2000. arXiv: [hep-ex/0005012](https://arxiv.org/abs/hep-ex/0005012) (cit. on p. [28](#)).
- [73] Stephen D. Ellis and Davison E. Soper. In: *Phys. Rev. D* 48 (1993), pp. 3160–3166. DOI: [10.1103/PhysRevD.48.3160](https://doi.org/10.1103/PhysRevD.48.3160). arXiv: [hep-ph/9305266](https://arxiv.org/abs/hep-ph/9305266) (cit. on p. [29](#)).
- [74] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. In: *JHEP* 04 (2008), p. 063. DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). arXiv: [0802.1189 \[hep-ph\]](https://arxiv.org/abs/0802.1189) (cit. on p. [29](#)).
- [75] Marco Stratmann and Werner Vogelsang. In: *Phys. Rev. D* 64 (2001), p. 114007. DOI: [10.1103/PhysRevD.64.114007](https://doi.org/10.1103/PhysRevD.64.114007). arXiv: [hep-ph/0107064](https://arxiv.org/abs/hep-ph/0107064) (cit. on pp. [30](#), [33](#)).

- [76] P. Jimenez-Delgado, A. Accardi, and W. Melnitchouk. In: *Phys. Rev. D* 89.3 (2014), p. 034025. DOI: [10.1103/PhysRevD.89.034025](https://doi.org/10.1103/PhysRevD.89.034025). arXiv: [1310.3734 \[hep-ph\]](https://arxiv.org/abs/1310.3734) (cit. on pp. [39](#), [57](#), [74](#), [98](#), [108](#)).
- [77] M. Tanabashi et al. In: *Phys. Rev. D* 98.3 (2018), p. 030001. DOI: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001) (cit. on pp. [39](#), [41](#)).
- [78] Guido Altarelli and G. Parisi. In: *Nucl. Phys. B* 126 (1977), pp. 298–318. DOI: [10.1016/0550-3213\(77\)90384-4](https://doi.org/10.1016/0550-3213(77)90384-4) (cit. on p. [40](#)).
- [79] M. Miyama and S. Kumano. In: *Comput. Phys. Commun.* 94 (1996), pp. 185–215. DOI: [10.1016/0010-4655\(96\)00013-6](https://doi.org/10.1016/0010-4655(96)00013-6) arXiv: [hep-ph/9508246](https://arxiv.org/abs/hep-ph/9508246) (cit. on p. [41](#)).
- [80] M. Hirai, S. Kumano, and M. Miyama. In: *Comput. Phys. Commun.* 108 (1998), p. 38. DOI: [10.1016/S0010-4655\(97\)00129-X](https://doi.org/10.1016/S0010-4655(97)00129-X), arXiv: [hep-ph/9707220](https://arxiv.org/abs/hep-ph/9707220) (cit. on p. [41](#)).
- [81] M. Hirai and S. Kumano. In: *Comput. Phys. Commun.* 183 (2012), pp. 1002–1013. DOI: [10.1016/j.cpc.2011.12.022](https://doi.org/10.1016/j.cpc.2011.12.022), arXiv: [1106.1553 \[hep-ph\]](https://arxiv.org/abs/1106.1553) (cit. on p. [41](#)).
- [82] A. Vogt. In: *Comput. Phys. Commun.* 170 (2005), pp. 65–92. DOI: [10.1016/j.cpc.2005.03.103](https://doi.org/10.1016/j.cpc.2005.03.103), arXiv: [hep-ph/0408244](https://arxiv.org/abs/hep-ph/0408244) (cit. on p. [41](#)).
- [83] R. S. Thorne. In: *Phys. Rev. D* 73 (2006), p. 054019. DOI: [10.1103/PhysRevD.73.054019](https://doi.org/10.1103/PhysRevD.73.054019), arXiv: [hep-ph/0601245](https://arxiv.org/abs/hep-ph/0601245) (cit. on pp. [41](#), [111](#)).
- [84] Jacob Ethier. “Proton Spin Structure from Monte Carlo Global QCD Analyses”. PhD thesis. William & Mary, 2018. DOI: [10.21220/s2-f3j1-c374](https://doi.org/10.21220/s2-f3j1-c374) (cit. on p. [41](#)).
- [85] Guido Altarelli, Stefano Forte, and Giovanni Ridolfi. In: *Nucl. Phys. B* 534 (1998), pp. 277–296. DOI: [10.1016/S0550-3213\(98\)00661-0](https://doi.org/10.1016/S0550-3213(98)00661-0), arXiv: [hep-ph/9806345](https://arxiv.org/abs/hep-ph/9806345) (cit. on pp. [42](#)–[44](#)).
- [86] Eric Moffat, Wally Melnitchouk, et al. In: (Jan. 2021). arXiv: [2101.04664 \[hep-ph\]](https://arxiv.org/abs/2101.04664) (cit. on p. [46](#)).



- [87] J. Pumplin, D. R. Stump, et al. In: *JHEP* 07 (2002), p. 012. DOI: [10.1088/1126-6708/2002/07/012](https://doi.org/10.1088/1126-6708/2002/07/012). arXiv: [hep-ph/0201195](https://arxiv.org/abs/hep-ph/0201195) (cit. on p. [48](#)).
- [88] Tie-Jiun Hou et al. In: *Phys. Rev. D* 103.1 (2021), p. 014013. DOI: [10.1103/PhysRevD.103.014013](https://doi.org/10.1103/PhysRevD.103.014013). arXiv: [1912.10053 \[hep-ph\]](https://arxiv.org/abs/1912.10053) (cit. on p. [51](#)).
- [89] Sergey Alekhin, Kirill Melnikov, and Frank Petriello. In: *Phys. Rev. D* 74 (2006), p. 054033. DOI: [10.1103/PhysRevD.74.054033](https://doi.org/10.1103/PhysRevD.74.054033). arXiv: [hep-ph/0606237](https://arxiv.org/abs/hep-ph/0606237) (cit. on p. [53](#)).
- [90] Renee Fatemi. *private communication*. 2020 (cit. on pp. [59](#), [61](#)).
- [91] A. Accardi, L. T. Brady, et al. In: *Phys. Rev. D* 93.11 (2016), p. 114017. DOI: [10.1103/PhysRevD.93.114017](https://doi.org/10.1103/PhysRevD.93.114017). arXiv: [1602.03154 \[hep-ph\]](https://arxiv.org/abs/1602.03154) (cit. on p. [65](#)).
- [92] Richard D. Ball et al. In: *Eur. Phys. J. C* 77.10 (2017), p. 663. DOI: [10.1140/epjc/s10052-017-5199-5](https://doi.org/10.1140/epjc/s10052-017-5199-5). arXiv: [1706.00428 \[hep-ph\]](https://arxiv.org/abs/1706.00428) (cit. on pp. [65](#), [73](#)).
- [93] A. M. Cooper-Sarkar and K. Wichmann. In: *Phys. Rev. D* 98.1 (2018), p. 014027. DOI: [10.1103/PhysRevD.98.014027](https://doi.org/10.1103/PhysRevD.98.014027). arXiv: [1803.00968 \[hep-ex\]](https://arxiv.org/abs/1803.00968) (cit. on pp. [72](#), [73](#)).
- [94] L. A. Harland-Lang, A. D. Martin, et al. In: *Eur. Phys. J. C* 75.5 (2015), p. 204. DOI: [10.1140/epjc/s10052-015-3397-6](https://doi.org/10.1140/epjc/s10052-015-3397-6). arXiv: [1412.3989 \[hep-ph\]](https://arxiv.org/abs/1412.3989) (cit. on p. [73](#)).
- [95] Emanuele R. Nocera, Richard D. Ball, et al. In: *Nucl. Phys. B* 887 (2014), pp. 276–308. DOI: [10.1016/j.nuclphysb.2014.08.008](https://doi.org/10.1016/j.nuclphysb.2014.08.008). arXiv: [1406.5539 \[hep-ph\]](https://arxiv.org/abs/1406.5539) (cit. on p. [74](#)).
- [96] R. L. Jaffe and Aneesh Manohar. In: *Nucl. Phys. B* 337 (1990), pp. 509–546. DOI: [10.1016/0550-3213\(90\)90506-9](https://doi.org/10.1016/0550-3213(90)90506-9) (cit. on p. [74](#)).
- [97] Elke C. Aschenauer, R. Sassot, and M. Stratmann. In: *Phys. Rev. D* 92.9 (2015), p. 094030. DOI: [10.1103/PhysRevD.92.094030](https://doi.org/10.1103/PhysRevD.92.094030). arXiv: [1509.06489 \[hep-ph\]](https://arxiv.org/abs/1509.06489) (cit. on p. [110](#)).

- [98] Ignacio Borsa, Gonzalo Lucero, et al. In: *Phys. Rev. D* 102.9 (2020), p. 094018. DOI: [10.1103/PhysRevD.102.094018](https://doi.org/10.1103/PhysRevD.102.094018). arXiv: [2007.08300 \[hep-ph\]](https://arxiv.org/abs/2007.08300) (cit. on p. [110](#)).
- [99] Daniel de Florian, Patriz Hinderer, et al. In: *Phys. Rev. Lett.* 112 (2014), p. 082001. DOI: [10.1103/PhysRevLett.112.082001](https://doi.org/10.1103/PhysRevLett.112.082001). arXiv: [1310.7192 \[hep-ph\]](https://arxiv.org/abs/1310.7192) (cit. on p. [111](#)).
- [100] M. Dittmar et al. In: (Nov. 2005). arXiv: [hep-ph/0511119](https://arxiv.org/abs/hep-ph/0511119) (cit. on p. [111](#)).
- [101] T. Stavreva, F. I. Olness, et al. In: *Phys. Rev. D* 85 (2012), p. 114014. DOI: [10.1103/PhysRevD.85.114014](https://doi.org/10.1103/PhysRevD.85.114014). arXiv: [1203.0282 \[hep-ph\]](https://arxiv.org/abs/1203.0282) (cit. on p. [111](#)).

## VITA

Yiyu Zhou

Yiyu Zhou was born in 1992 in Xingping, Shaanxi Province, P.R. China. Since his youth, he has always been curious about many things related to science and math. His formal education began in 1999 at Shaanchai Elementary School in Xingping, Shaanxi Province. After completing primary and secondary education in Xingping, he was admitted into the Xi'an Jiaotong University in Xi'an, Shaanxi Province. His major was Biology Engineering at first, but later transferred to the Physics Elite Class in the School of Science. He won the national first class scholarship in 2012 and the undergraduate research scholarship in 2014. Dissatisfied with college-level physics, he entered William & Mary in Fall 2015. In 2017, he joined the [High Energy Theory Group](#) to work with Prof. Joshua Erlich on emergent gravity and supersymmetry. In November 2018, he decided to work on hadron physics with Prof. Wally Melnitchouk at [Jefferson Lab](#) on extracting parton helicity distribution in protons from experimental data. After graduation, he will continue his research in QCD in a joint postdoc position at South China Normal University and UCLA.